



Niraj Suthar

An Improved Machine Learning, Wavelet-ARIMA Approach
for Forecasting Spodumene Prices

Cranfield School Of Management
MSc in Business Data Analytics

MSc Thesis
Academic Year: 2024–2025

Supervisor: Prof. Andrew Angus
September 2025



Cranfield School Of Management
MSc in Business Data Analytics

MSc Thesis

Academic Year: 2024–2025

Niraj Suthar

An Improved Machine Learning, Wavelet-ARIMA Approach
for Forecasting Spodumene Prices

Supervisor: Prof. Andrew Angus
September 2025

This thesis is submitted in partial fulfilment of the
requirements for the degree of MSc Thesis.

© Cranfield University 2025. All rights reserved. No part of
this publication may be reproduced without the written
permission of the copyright owner.

Academic Integrity Declaration

I declare that:

- the thesis submitted has been written by me alone.
- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

The thesis addresses the urgent need for reliable forecasting of spodumene, a hard-rock source of lithium whose demand is surging with the global transition to electric vehicles and energy storage. Because public price data only span 2018-2025 and are characterized by sharp booms and corrections, the development of a leakage-safe framework that combines classical time-series models (ARIMA), discrete wavelet transforms and machine learning learners (Elastic Net, XGBoost and LSTM). The dataset was sourced from China Spodumene Li_2O 6% CIF (Asian metal limited) index from Bloomberg which had 379 observations from 14th Jan 2018–28th Jul 2025 which was aligned to a Sunday index, log-transformed, differenced and augmented with calendar and lag features. A chronological train–validation–test split and rolling-origin cross-validation are used to assess two forecasting horizons 2 weeks (short-term operational) and 12 weeks (medium-term planning). Naive persistence and ARIMA models serve as transparent baselines wavelet enhanced, and residual hybrids test whether routing frequency components to appropriate learners improves accuracy. Performance is judged via MAE, RMSE and MAPE, with residual whiteness and parity plots ensuring model adequacy. Results show that at the two-week horizon, wavelet-based hybrids particularly those combining elastic net or gradient-boosted trees with ARIMA reduce test MAE by roughly one-fifth relative to wavelet–ARIMA and improve robustness over Naive and ARIMA benchmarks. Stand-alone XGBoost also performs well, while deep LSTM models overfit the short series. At the twelve-week horizon, errors grow markedly across all models wavelet–ARIMA performs poorly and hybrid gains narrow, reflecting multi-step uncertainty and data scarcity. The study concludes that frequency-aware, machine-learning–augmented models offer practical value for short-horizon lithium procurement and hedging, whereas medium-term forecasts remain uncertain. It recommends future research incorporating exogenous market drivers and adaptive learning to strengthen medium-horizon reliability.

Keywords:

Wavelet decomposition, Spodumene pricing, Machine learning hybrids, ARIMA model, Wavelet-ARIMA models, ARIMA-Machine learning models, Wavelet-ARIMA-Machine learning models

Acknowledgements

I owe a debt of gratitude to a number of people who helped me steer this thesis from ambitious idea to finished work. First, my sincerest thanks to Professor Andrew Angus for his guidance, patience and the occasional reality check when my forecasts went rogue. His feedback not only improved the research but also kept me grounded when I was tempted to chase every shiny new method.

To my family, thank you for cheering me on. Your love, encouragement, and supplies of tea and biscuits made the late nights tolerable and the early mornings possible.

A special shout-out goes to my friends and classmates: you kept me laughing through the rough patches, listened to countless draft presentations, and provided constructive criticism without pulling any punches. Whether we were swapping data-cleaning tips or sharing memes about procrastination, you all made this journey bearable and even enjoyable at times.

Together, you all helped turn an overwhelming task into a meaningful learning experience. Thank you for being part of this adventure, and for making sure I didn't drown in my own spreadsheets.

Contents

Academic Integrity Declaration	i
Abstract	ii
Acknowledgements	iii
Contents	iv
List of Figures	vi
List of Tables	vii
List of Abbreviations	viii
1 Introduction	1
1.1 Research Aims and Objectives	2
1.2 Theoretical Motivation and Pre-Processing Overview	2
1.3 Experimental Design and Evaluation	3
2 Literature Review	4
2.1 Spodumene Market Dynamics and Pricing	5
2.2 Commodity and Spodumene Price Forecasting	6
2.2.1 Traditional Statistical Approaches	6
2.2.2 Machine Learning and Deep Learning Models	6
2.2.3 Hybrid Models and the role of Wavelet Decomposition	7
2.3 Critical Analysis	9
2.3.1 Advantages and Limitations	9
2.3.2 Research Gaps and Future Directions	10
3 Methodology	11
3.1 Overview	11
3.2 Data Pre-Processing	11
3.3 Feature Engineering	12
3.4 Forecasting Models	12
3.4.1 Statistical Benchmark Model	12
3.4.2 WAVELET + ARIMA Hybridisation	13
3.4.3 Machine Learning Models	14
3.4.4 Advanced Hybrid Framework	15

3.5	Experimental Design and Evaluation	15
3.5.1	Forecasting Framework	15
3.5.2	Evaluation of Predictive Power	16
4	Results	17
4.1	Overview	17
4.2	Exploratory Data Analysis (EDA)	17
4.3	Benchmark Models	20
4.3.1	ARIMA Model	20
4.3.2	XGBoost Machine Learning Benchmark Model	21
4.4	Residual Integrated Models (ARIMA + ML Models)	21
4.4.1	ARIMA + XGBoost	22
4.5	WAVELET Based Models	23
4.5.1	WAVELET + ARIMA	23
4.5.2	WAVELET + XGBoost	24
4.6	Machine Learning + ARIMA + WAVELET	24
4.6.1	XGBoost + ARIMA + WAVELET	25
5	Discussion	27
5.1	Summary of Key Findings	27
5.2	Interpretation vs Objectives	27
5.3	Findings in Context of the Literature	28
5.3.1	Methodological Insights	29
5.4	Practical and Managerial Limitations	31
5.5	Limitations	31
5.6	Recommendations for Practice and Future research	32
6	Conclusions	34
7	References	36
A	Appendix	40
B	Literature Review Appendix	41
C	Methodology Appendix	42
D	Results Appendix	46
E	Discussion Appendix	57

List of Figures

2.1	Uses of Lithium	4
3.1	First Differencing of the time series in LOG	13
3.2	Approximation and detail components.	14
4.1	Spodumene Price (Original vs. Filled).	18
4.2	Histogram with KDE of Filled Spodumene Prices.	18
4.3	Spodumene Prices with IQR Bound.	18
4.4	Rolling Mean and Volatility (Raw vs Log-Transformed).	19
4.5	Seasonal Decomposition of Log-Transformed Prices.	20
4.6	ARIMA forecast for the next 12 weeks.	20
4.7	Actual vs Predicted for XGBoost on validation and test.	21
4.8	Actual vs Predicted for ARIMA + XGBoost on validation and test.	22
4.9	WAVELET + ARIMA forecast for the next 12 weeks.	23
4.10	Actual vs Predicted for WAVELET + XGBoost on validation and test.	24
4.11	Actual vs Predicted for WAVELET + ARIMA + XGBoost on validation and test.	25
D.1	Residuals Over Time ARIMA	48
D.2	Histogram and Q-Q plot of residuals	48
D.3	ACF and PACF residual plots	48
D.4	Actual vs Predicted for Elastic Net on validation and test	49
D.5	Actual vs Predicted for LSTM on validation and test	49
D.6	Errors over time for XGBoost on validation and test.	50
D.7	Actual vs Predicted for ARIMA + Elastic Net on validation and test.	51
D.8	Errors over time for ARIMA + Elastic Net on validation and test.	51
D.9	Actual vs Predicted for ARIMA + LSTM on validation and test.	51
D.10	Errors over time for ARIMA + LSTM on validation and test.	52
D.11	Actual vs Predicted for WAVELET + EN on validation and test.	53
D.12	Errors over time for WAVELET + EN on validation and test.	54
D.13	Actual vs Predicted for WAVELET + LSTM on validation and test.	54
D.14	Actual vs Predicted for WAVELET + ARIMA + EN on validation and test.	55
D.15	Actual vs Predicted for WAVELET + ARIMA + LSTM on validation and test.	56

List of Tables

2.1	Lithium demand projections (kt Li)	5
4.1	Validation and test scores (price scale).	22
4.2	Validation and test scores (model vs. Naïve-H baseline).	24
4.3	Validation/Test metrics for each model (excluding R^2).	26
4.4	Residual diagnostic p -values by model.	26
D.1	Summary statistics of Spodumene prices.	46
D.2	Weekly spodumene prices (June 2022–April 2023).	47

List of Abbreviations

ACF	Autocorrelation Function
ADF	Augmented Dickey–Fuller (test)
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CIF	Cost, Insurance and Freight (e.g., China Spodumene Li ₂ O 6% CIF index)
db2	Daubechies wavelet of order two
DL	Deep Learning
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machine
EMA	Exponentially Weighted Moving Average
EN	Elastic Net
EV / EVs	Electric Vehicle(s)
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
GBP	British pounds
IWNN	Improved Wavelet Neural Network
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
kt	Kilotonnes
Li	Lithium (chemical symbol)
LSTM	Long Short-Term Memory (neural network)
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MODWT	Maximal Overlap Discrete Wavelet Transform
R ²	Coefficient of Determination
RNN	Recurrent Neural Network
RMSE	Root Mean Squared Error
SVR	Support Vector Regression
TDNN	Time Delay Neural Network
XGB	Extreme Gradient Boosting (short form)
XGBoost	Extreme Gradient Boosting
CaCl ₂	Calcium chloride
CaCO ₃	Calcium carbonate (limestone)

CaO	Calcium oxide (quicklime)
H ₂ SO ₄	Sulfuric acid
Li ₂ SiO ₃	Lithium metasilicate
NaOH	Sodium hydroxide
NH ₄ Cl	Ammonium chloride

Chapter 1

Introduction

Lithium has emerged as a pivotal enabler of the clean-energy transition, undergirding rapid growth in electric mobility and stationary storage while retaining important roles in ceramics, glass and lubricants. Recent demand outlooks from international agencies point to a multi-fold increase in lithium consumption through mid-century, with batteries already accounting for the dominant share of end use around 87 percent displacing historically important industrial applications. In parallel, market structure has shifted Australia’s hard-rock resources especially spodumene now supply a substantial share of global output, reflecting both geological endowments and strategic diversification away from South American brines. These developments have amplified attention on spodumene as the primary hard-rock feedstock for lithium chemical production, even as its extraction and processing remain capital and energy intensive (Asif et al., 2024; Dessemond et al., 2019; International Energy Agency, 2025).

Despite lithium’s strategic importance, spodumene pricing is comparatively opaque and immature relative to metals with deep terminal markets. Transactions occur via a mix of bilateral contracts, spot deals and auctions, with sparse, short public price histories and evolving benchmarks. Policy shocks (for example, EV subsidies), supply disruptions and speculative cycles have produced sharp booms and corrections since 2018, complicating planning for miners, converters, and downstream users. In such settings, forecasting that is both accurate and transparent is not merely of academic interest; it is operationally critical for investment appraisal, procurement, and risk management (Asif et al., 2024; Dessemond et al., 2019; Martin et al., 2017).

This thesis addresses that need by designing and evaluating a unified forecasting framework that combines classical time-series models Autoregressive Integrated Moving Average (ARIMA), wavelet decomposition and machine learning (ML) learners to predict weekly spodumene prices at 2-week and 12-week horizons, and by testing whether hybridisation yields meaningful accuracy and robustness gains over Naive and ARIMA baselines. The analysis is grounded in a Bloomberg time series of the China Spodumene $\text{Li}_2\text{O}6\%$ CIF (Asian Metal Limited) index, comprising 379 weekly observations from 14 January 2018 to 28 July 2025 (prices in GBP/tonne). The short yet turbulent history makes spodumene an exacting testbed for forecasting under non-stationarity and regime change.

1.1 Research Aims and Objectives

Design and evaluate a unified forecasting framework that combines ARIMA, wavelet decomposition, and machine learning models to predict weekly spodumene prices at 2 and 12-week horizons and determine whether the hybrid approach delivers meaningful accuracy and robustness gains over Naive and ARIMA baselines. Guided by the literature and market context, the thesis operationalises this objective through: (i) establishing transparent statistical and Naive baselines; (ii) implementing wavelet-enhanced statistical and ML models; (iii) integrating models in frequency space and via residual learning; and (iv) evaluating performance on strictly out-of-sample windows at both short ($H = 2$ weeks) and medium ($H = 12$ weeks) horizons using multiple error metrics and stability diagnostics.

1.2 Theoretical Motivation and Pre-Processing Overview

Classical ARIMA models remain valued for interpretability and as robust “first pass” benchmarks, but their linear structure and stationarity assumptions can be brittle under structural breaks and volatility clustering features characteristic of spodumene’s young market. By contrast, ML, and deep-learning approaches (for example, Elastic Net (EN), XGBoost, Long Short-Term Memory (LSTM)) flexibly approximate non-linear mappings yet risk overfitting when data are scarce and regimes shift. The literature increasingly therefore advocates hybrids, with wavelet transforms providing a principled way to separate long-run trends from high-frequency movements and to route components to learners suited to their statistical character. This multiresolution paradigm has delivered gains across commodities, energy, and finance, although sensitivity to wavelet choice and decomposition depth remains a known design lever (Crowley, 2007; Kriechbaumer et al., 2014; Paul & Garai, 2021).

Against this backdrop, spodumene presents three intertwined challenges: limited historical depth (public prices only since 2018), exposure to exogenous policy and supply shocks, and a contract-heavy pricing ecosystem with evolving indexation. These features imply that (a) purely linear models may under-react at turning points; (b) complex sequence models may not generalise with short histories; and (c) variance stabilisation and multi-scale decomposition could improve signal-to-noise for medium-horizon forecasts. The framework developed here is expressly designed to test those propositions (Caldwell, 2015).

The series is indexed weekly to Sundays, with missing weeks linearly interpolated to preserve trajectory without introducing spurious volatility. Outliers from the 2021–2023 boom are retained as economically meaningful. A natural-log transform stabilises variance, and first differencing is applied where unit-root tests (Augmented Dickey Fuller test (ADF), Kwiatkowski–Phillips–Schmidt–Shin (KPSS)) indicate non-stationarity. Feature engineering adds calendar indicators, short-/medium-/annual autoregressive lags, rolling statistics, and exponentially weighted averages; all predictors are horizon-shifted to prevent leakage and pruned for high collinearity (Byers et al., 2021; Crowley, 2007; Lütkepohl & Xu, 2012).

1.3 Experimental Design and Evaluation

To approximate operational use and mitigate look-ahead bias, the study adopts chronological splits TRAIN, a 12-week VALIDATION window for model selection, and a 12-week TEST window for final performance plus rolling-origin cross-validation on the development sample. Accuracy is assessed primarily via Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) (R^2 is reported for completeness but not used for ranking.), with residual checks (Ljung–Box), parity plots and rolling-error panels to judge stability over time. The twin horizons ($H = 2$, $H = 12$) illuminate the trade-offs between short-range tracking and medium-range robustness, given the limited 2018–2025 data regime (Hyndman & Athanasopoulos, 2014; Varshini et al., 2024).

In sum, this thesis seeks to bring both methodological clarity and practical utility to spodumene price forecasting. By anchoring the analysis in transparent baselines, leveraging multi-scale structure where it is most informative, and testing hybrids under strict out-of-sample conditions at two operational horizons, it aims to establish when and by how much integrated wavelet + ARIMA + ML frameworks add value over simpler strategies in a market that is both strategically important and empirically challenging.

Chapter 2

Literature Review

Lithium has rapidly emerged as a critical element underpinning the global transition toward sustainable energy, particularly because of its indispensable role in rechargeable lithium-ion batteries, ceramics, glass, lubricants, and a range of technological applications (Asif et al., 2024; Dessemond et al., 2019; Li et al., 2019). Batteries now constitute 87% share of lithium consumption, surpassing traditional sectors such as ceramics and glass (see figure 2.1) (U.S. Geological Survey, Mineral Commodity Summaries, 2025).

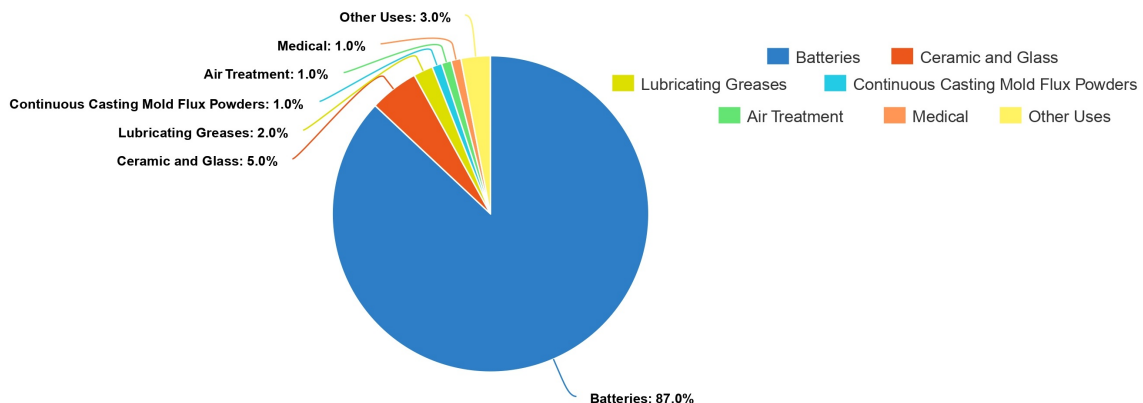


Figure 2.1: Uses of Lithium

The accelerating adoption of electric vehicles (EVs) and grid-scale energy storage has triggered a dramatic surge in global lithium demand, with forecasts suggesting demand will increase from 95 kilo tonne (kt) to 1 834 kilo tonnes of lithium equivalent between 2021 and 2050 (see table 2.1) (International Energy Agency, 2025).

Australia has cemented its position as the dominant player in lithium mining, leveraging vast spodumene reserves to account for half of worldwide exports in 2022; a development driven by the Greenbushes deposit in Western Australia (Asif et al., 2024). While lithium brines in South America have historically been favoured for their lower production costs, geopolitical factors and the uneven geographical distribution of brines have shifted industry attention towards hard-rock resources such as spodumene. These hard-rock sources now account for approximately 40% of global lithium production (Dessemond et al., 2019; Li et al., 2019).

Spodumene ($\text{LiAlSi}_2\text{O}_6$), the lithium-bearing pyroxene mineral at the heart of this transformation, is particularly valued for its high theoretical lithium content and favour-

Unit: kt Li	Historical		Stated Policies			Announced Pledges			Net Zero Emissions by 2050		
	2021	2024	2030	2040	2050	2030	2040	2050	2030	2040	2050
Clean energy	38	128	369	809	1019	418	1081	1389	615	1427	1683
Electric vehicles	35	109	326	739	930	369	993	1276	556	1318	1547
Battery storage	2	2	43	70	89	48	88	113	58	109	137
Other uses	57	77	87	119	150	87	119	150	87	119	150
Total demand	95	205	455	928	1170	505	1200	1540	701	1546	1834
Share of clean energy	40%	62%	81%	87%	87%	83%	90%	90%	88%	92%	92%

Table 2.1: Lithium demand projections (kt Li)

able recovery rates. Yet, the extraction process is complex, requiring metallurgical processing, including high-temperature phase conversion and a suite of hydrometallurgical or pyrometallurgical treatments, presenting significant challenges in terms of energy consumption, waste management, and environmental sustainability (Asif et al., 2024; Dessemond et al., 2019). As global supply chains strive to meet soaring demand while aligning with broader decarbonization and circular economy goals, the efficient and sustainable extraction of lithium from spodumene remains a central focus for both industry and academia (Martin et al., 2017).

This literature review critically examines the evolution of spodumene extraction, the economics and market dynamics of spodumene, and the state of the art in price forecasting methodologies. Emphasis is placed on the recent proliferation of hybrid approaches integrating machine learning, wavelet decomposition, and classical statistical models within the context of volatile and non-linear commodity markets. The aim is to establish a foundation for advancing robust and adaptable forecasting models, specifically tailored to spodumene’s unique market characteristics.

2.1 Spodumene Market Dynamics and Pricing

The lithium supply chain, and by extension spodumene, is increasingly subject to complex market dynamics shaped by extraction costs, processing complexity, environmental regulations, technological progress, and geopolitical events (Asif et al., 2024; Dessemond et al., 2019). Spodumene’s strategic importance is further amplified by the global pivot toward battery storage and e-mobility, making it central to the ambitions of decarbonization and energy security (Martin et al., 2017).

Market mechanisms for spodumene differ markedly from more mature commodities, with pricing set by a mix of spot transactions, long-term contracts, and, more recently, auction-based systems (Martin et al., 2017). The young and evolving market structure, combined with a lack of universally accepted benchmarks, introduces significant price volatility (Dessemond et al., 2019). This volatility is further exacerbated by government policies, such as electric vehicle subsidies, battery manufacturing incentives, and critical mineral security strategies, which can abruptly alter demand (Asif et al., 2024).

From 2017 to 2019, lithium and spodumene prices saw a speculative bubble, driven by aggressive projections for EV adoption, followed by a downturn in 2020 due to pandemic-related disruptions (Li et al., 2019; Martin et al., 2017). The rebound in 2022, in response

to revived EV sales and supply chain bottlenecks, highlights the sector's sensitivity to exogenous shocks (Tami & Owda, 2024). The lack of a transparent, liquid terminal market means prices are often set through opaque, long-term contracts rather than standardized indices (Martin et al., 2017).

Additional drivers, including exchange rate fluctuations, export restrictions, and futures market speculation, adds further layers of complexity (Asif et al., 2024). In such an environment, robust, transparent, and adaptive price forecasting becomes academically interesting and operationally essential for industry and policy stakeholders (Martin et al., 2017; Tami & Owda, 2024).

2.2 Commodity and Spodumene Price Forecasting

2.2.1 Traditional Statistical Approaches

Early research on commodity price forecasting relied heavily on linear time series models such as ARIMA and Autoregressive Moving Average (ARMA), which remain valued for their transparency and ability to capture trend and seasonality (Abdulhafidh Ibrahim et al., 2023; Skehin et al., 2018). The ARIMA framework, as formalised by Box and Jenkins, predicts future values based on past observations through a combination of autoregressive, differencing, and moving average components. These models have been effective for established commodities with stable historical patterns and long data series, where shocks are infrequent and behaviour is linear (Vijaya & Swati, 2024).

However, ARIMA's assumptions of linearity and stationarity restrict its performance in markets characterised by volatility, structural breaks, and regime shifts. The model struggles when abrupt demand shocks, speculative cycles, or exogenous policy interventions disrupt historical patterns leading to biased forecasts and poor generalisation beyond short horizons (Md-Khair & Samsudin, 2018). For spodumene, these limitations are magnified: the market is young, with reliable public price data only available since 2018, and is highly exposed to external shocks such as electric vehicle subsidy changes, supply bottlenecks, and auction-driven price swings. Traditional ARIMA models frequently fail to capture such nonlinear behaviour, resulting in forecasts that underestimate both the magnitude and frequency of price jumps (Zhang et al., 2023).

In this study, ARIMA is therefore treated as a benchmark model. While it establishes a transparent baseline against which more advanced approaches can be evaluated, its shortcomings in the context of spodumene highlight the need for hybrid frameworks that combine statistical rigour with tools better suited to nonlinear, volatile price series.

2.2.2 Machine Learning and Deep Learning Models

Recognising the limitations of linear models, researchers have increasingly adopted advanced machine learning and deep learning (DL) techniques for commodity price forecasting. These approaches excel at capturing complex, non-linear dependencies and adapting to rapidly changing market regimes without requiring the explicit specification of functional forms (Kumar & Bhattacharyya, 2024; Paul & Garai, 2021). Unlike ARIMA, which assumes a fixed linear relationship, ML algorithms such as Artificial Neural Networks (ANN) and support vector regression (SVR) approximate non-linear mappings

directly from data. LSTM networks extend this capability by incorporating memory cells that retain information across long sequences, enabling them to model cyclical patterns and structural breaks common in commodity markets. This architecture has consistently delivered lower root mean square error (RMSE) and higher explanatory power than classical time series models in forecasting volatile asset and commodity prices (Skehin et al., 2018; Tami & Owda, 2024).

Despite these advantages, standalone ML and DL models face notable limitations. Their predictive strength often comes at the cost of interpretability, with many approaches criticised as “black boxes” in applied settings. They are also data-hungry, requiring large, high-quality datasets to generalise well, and are prone to overfitting when data are scarce or noisy (Rubio et al., 2023; Shameem et al., 2025). These constraints are particularly relevant for spodumene, where publicly available historical price series are short and market shocks are frequent. In such conditions, complex DL architectures risk capturing noise rather than signal, undermining their forecasting reliability.

Accordingly, while ML and DL methods provide powerful tools to model the non-linear dynamics inherent in spodumene markets, their practical limitations reinforce the case for hybrid frameworks. Combining these techniques with statistical baselines such as ARIMA allows models to balance robustness, interpretability, and accuracy, laying the foundation for the hybrid approaches reviewed in the following section (Rubio et al., 2023; Shameem et al., 2025).

2.2.3 Hybrid Models and the role of Wavelet Decomposition

The application of wavelet transforms has emerged as a fundamental innovation in commodity price forecasting. Wavelet decomposition separates a time series into low- and high-frequency components, enabling targeted modelling of long-term trends and short-term volatility (Kriechbaumer et al., 2014; Paul & Garai, 2021). Unlike Fourier transforms, which only provide frequency information, wavelets simultaneously capture both time and frequency localisation. This property makes them particularly effective for modelling non-stationary, noisy price series such as spodumene, where structural breaks and sudden shocks are common.

Mechanistically, wavelet transforms improve forecasting by stabilising variance, filtering noise, and isolating cyclical behaviour at different time scales. Approaches such as the maximal overlap discrete wavelet transform (MODWT) also avoid the down-sampling and shift-variance problems of standard Discrete Wavelet Transform (DWT), preserving the length and variance of the original series (Kriechbaumer et al., 2014; Md-Khair & Samsudin, 2018). These features allow subsequent statistical or machine learning models to operate on cleaner, more interpretable sub-series.

As a pre-processing step, wavelet decomposition has been shown to enhance the performance of ML algorithms. Paul and Garai, 2021, for example, found that wavelet-ANN hybrids outperformed both conventional ARIMA and pure ANN models in forecasting agricultural commodity prices, as the decomposition simultaneously removed noise and retained key signals. Kriechbaumer et al., 2014, extended this to metals, demonstrating that a wavelet + ARIMA framework with optimised transform type, wavelet family, and decomposition levels significantly reduced forecast errors for aluminium, copper, and zinc. Their multi-resolution analysis captured cyclical patterns that traditional ARIMA

models failed to detect a result directly relevant to spodumene, which exhibits similar volatility but within a younger and less transparent market.

Similarly, Md-Khair and Samsudin, 2018, showed that applying a wavelet + ARIMA approach to crude oil prices stabilised otherwise non-stationary series and improved predictive performance compared with ARIMA alone. By decomposing prices into trend and detail components, modelling each individually, and then reconstructing the full series, they demonstrated a practical way to adapt forecasting tools to highly volatile resource markets.

Building on these foundations, more recent research has advanced fully integrated hybrid models. Kumar and Bhattacharyya, 2024, for instance, proposed a wavelet + ARIMA + Time Delay Neural Network (TDNN) framework for agricultural prices, using wavelet decomposition to separate linear and non-linear frequency components. ARIMA was applied to the linear sub-series, while TDNN managed the non-linear components, and forecasts were recombined via inverse wavelet transform. This hybridisation outperformed ARIMA, TDNN, and simpler hybrids, underscoring the value of aligning models with the statistical properties of each frequency band. Similarly, Shameem et al., 2025 compared ARIMA + ANN, ARIMA + SVR, and multiple wavelet-based hybrids in forecasting sesame prices and concluded that models integrating wavelet decomposition with ML consistently delivered the highest predictive accuracy. These findings highlight the generalisability of frequency-wise hybrids across diverse commodities.

The hybridisation trend has also extended beyond commodities. Rubio et al., 2023, demonstrated that combining wavelet decomposition with ARIMA and Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models improved volatility forecasts for high-frequency stock data, outperforming single-model strategies. Such flexibility is highly relevant to spodumene, where price dynamics are shaped by abrupt shocks, policy interventions, and evolving market structures. More recently, Qu et al., 2024 proposed an improved wavelet neural network (IWNN) with extreme learning machine (ELM) initialisation for electricity price forecasting, while attention-based Recurrent Neural Network (RNN) + ARIMA hybrids have combined sequential learning with statistical robustness to manage noisy financial and energy data.

For spodumene, these advances are particularly salient. The market is characterised by limited historical data, irregular supply-demand shocks, and contract-based pricing mechanisms rather than transparent indices. Under such conditions, direct application of ARIMA or standalone ML models risks misspecification or overfitting. Wavelet-based hybrids, by contrast, allow low-frequency components to be captured with ARIMA while high-frequency volatility and non-linearities are addressed by ML or DL models.

In this study, wavelet decomposition therefore provides the backbone of the methodological design. MODWT with 2-4 decomposition levels are adopted to preserve variance and capture both medium- and high-frequency cycles. The McLeod–Li test (McLeod & Li, 1983) is employed to distinguish linear from non-linear components, guiding the allocation of ARIMA to trend/linear frequencies and ML/DL models to non-linear bands. This frequency-wise hybridisation is expected to deliver more robust and accurate forecasts for spodumene than any single modelling approach.

2.3 Critical Analysis

The evolution of forecasting methodologies from traditional statistical models through ML/DL approaches to hybrid and wavelet-based models has resulted in marked improvements in forecast accuracy for volatile and non-linear commodities. However, this methodological sophistication also brings new challenges.

2.3.1 Advantages and Limitations

Hybrid models leveraging wavelet decomposition, ARIMA, and ML/DL algorithms consistently outperform their standalone counterparts, offering more robust and accurate forecasts across commodity, energy, and financial markets (Kriechbaumer et al., 2014; Paul & Garai, 2021; Qu et al., 2024). The noise-filtering and multi-scale decomposition capabilities of wavelets enable these models to isolate meaningful trends and volatility at multiple scales, enhancing the subsequent predictive performance of both statistical and machine learning models. Wavelets stabilise variance, reduce noise, and allow frequency-wise modelling, ensuring ARIMA captures low-frequency linear dynamics while ML/DL target high-frequency nonlinearities.

Nonetheless, challenges persist. Model interpretability remains a significant concern, especially as deep learning architectures grow in complexity. While ML/DL models offer flexibility and power, they are often criticised as “black boxes,” making them difficult to interpret and less trusted in domains where transparency and explainability are essential for decision-making (Rubio et al., 2023; Shameem et al., 2025). For spodumene, where market transparency is already low and public data limited, such interpretability concerns reduce stakeholder confidence in model outputs.

The choice of model parameters, such as mother wavelet function and decomposition level, also significantly affects performance, but is often determined heuristically rather than through rigorous optimisation. This is problematic in spodumene forecasting, where small datasets magnify the sensitivity of forecasts to wavelet function and decomposition depth (Kumar & Bhattacharyya, 2024).

Another notable challenge is the integration of domain-specific exogenous variables. While most studies focus on time series techniques, real-world forecasting often requires the inclusion of external drivers, such as macroeconomic indicators, market structure variables, and news sentiment. The lack of systematic approaches for incorporating such exogenous data limits the real-world applicability and robustness of many forecasting frameworks (Kriechbaumer et al., 2014; Tami & Owda, 2024).

Furthermore, practical deployment of advanced ML/DL and hybrid models is sometimes constrained by computational cost and data requirements. Many studies are conducted on clean, small-scale datasets that do not reflect the irregularities or streaming requirements of industrial environments. Thus, scalability and real-time application capabilities merit further investigation. These limitations highlight the need for forecasting frameworks that balance accuracy, transparency, and practical deploy ability an agenda particularly urgent for spodumene, where data constraints and market volatility are more severe than in established commodities.

2.3.2 Research Gaps and Future Directions

While the reviewed literature demonstrates substantial progress, several important gaps remain, especially as they pertain to spodumene and lithium price forecasting.

First, most advanced hybrid models have been evaluated on well-established commodities with extensive historical data and mature market structures. There is limited comparative research that benchmarks different combinations of wavelets, ARIMA specifications, and ML algorithms on immature and volatile commodities like spodumene (Dessemond et al., 2019; Shameem et al., 2025).

Second, sensitivity analysis regarding wavelet selection, decomposition levels, and hyperparameter optimisation is often neglected, which limits practical adoption. Future work should systematise these model selection and tuning steps to enable more robust performance across diverse market regimes (Kumar & Bhattacharyya, 2024).

Third, the integration of spodumene-specific factors, such as processing complexity, grade variability, regulatory changes, and geopolitical risks, into forecasting frameworks remains underexplored. Most models are transferred from other commodity contexts and may not adequately capture the unique dynamics of spodumene and lithium supply chains (Dessemond et al., 2019; Liu et al., 2023).

Fourth, there is a need for models that balance accuracy with interpretability. As hybrid and ensemble approaches become more complex, developing explainable Artificial Intelligence (AI) tools and post-hoc interpretability methods for commodity forecasting will be increasingly important, particularly for industrial and policy decision-making contexts (Tami & Owda, 2024).

Finally, as the spodumene market continues to mature and data availability improves, opportunities will arise to leverage streaming data, real-time learning, and adaptive model architectures. Future research should focus on real-time adaptive models that automatically adjust to changing market conditions and structural breaks, thereby maintaining forecast reliability (Qu et al., 2024; Zeng & Khushi, 2020).

Chapter 3

Methodology

3.1 Overview

Building on the literature review, which highlighted the volatility and limited historical transparency of spodumene markets, this study focuses on forecasting weekly spodumene prices using a range of statistical and machine learning models. The aim is to identify which approaches most effectively capture and predict the dynamics of this increasingly important mineral. The dataset was sourced from Bloomberg and is based on the China Spodumene (Li_2O) 6% CIF index compiled by Asian Metal Limited. Publicly available spodumene price data is recent; prior to the 2010s, market information was limited to internal company reports from Australian mining firms that dominated global production (Taquiri et al., 2024).

The dataset comprises 379 weekly observations from 14 January 2018 to 28 July 2025, providing over seven years of continuous market data. Prices are reported in British Pounds (GBP) per tonne. To test robustness and practical relevance, forecasts are generated for both short-term (two weeks ahead) and medium-term (twelve weeks ahead) horizons. Model performance is evaluated using widely adopted metrics: MAE, MAPE, RMSE, and the coefficient of determination (R^2).

3.2 Data Pre-Processing

The spodumene price data exhibited several common features of commodity time series, including irregular sampling intervals, high volatility, and signs of non-stationarity. To address these issues, a structured pre-processing procedure was applied. The series was first reindexed to a consistent weekly frequency, aligning all observations to Sundays. Missing weeks were filled using linear interpolation to preserve temporal continuity while avoiding the creation of artificial volatility (Byers et al., 2021).

Outliers were identified using z-scores and the interquartile range method but were retained, as sudden price spikes in commodity markets often represent meaningful economic shocks rather than errors (Foroutan & Lahmiri, 2024; Varshini et al., 2024). To stabilise variance and reduce skewness, a natural logarithm transformation was applied, following established practices in financial and commodity forecasting (Crowley, 2007; Lütkepohl & Xu, 2012).

Finally, stationarity was evaluated using the Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. The presence of unit roots required first-order differencing, which rendered the series stationary and suitable for ARIMA-based modelling (Kriechbaumer et al., 2014).

3.3 Feature Engineering

Following preprocessing, a structured feature engineering pipeline was implemented to enhance the predictive capacity of the forecasting models. Feature engineering is particularly important in financial and commodity time series, where raw prices alone may not fully capture temporal dependencies, volatility clustering, or seasonality (Sari et al., 2024). To address this, the univariate price series was transformed into a richer feature space comprising calendar, lagged, trend, and volatility-based variables.

Calendar features such as year, month, and quarter indicators were introduced to reflect seasonal industrial cycles that often influence commodity demand (Varshini et al., 2024). Lagged variables captured short- and medium-term autoregressive dependencies, while rolling statistics such as moving averages and standard deviations were included to represent evolving market regimes and volatility (Byers et al., 2021; Crowley, 2007). Exponentially weighted averages were also used to prioritise more recent observations (Debnath & Shah, 2015) (see Appendix ?? for more information).

To avoid information leakage, all lagged and rolling variables were shifted relative to the forecast horizon, and highly correlated features were pruned to reduce multicollinearity and improve generalisability (Kriechbaumer et al., 2014). The resulting feature set provided a robust basis for the statistical, machine learning, and hybrid models presented in the following sections.

3.4 Forecasting Models

3.4.1 Statistical Benchmark Model

The ARIMA model, developed within the Box–Jenkins framework, was adopted as the statistical benchmark for this study. ARIMA decomposes a time series into autoregressive (AR), differencing (I), and moving average (MA) components, enabling it to capture linear dependencies in temporal data (Rubio et al., 2023). In commodity and financial applications, ARIMA is often the first modelling choice because of its transparency, interpretability, and proven track record in handling univariate series with strong autocorrelations (Hyndman & Athanasopoulos, 2014) (see Appendix C for more information).

To address volatility and heteroscedasticity in the spodumene price series, the raw data was first transformed using natural logarithms, a common practice shown to stabilise variance and produce more symmetric distributions (Lütkepohl & Xu, 2012). Stationarity was then assessed using two complementary unit root tests: the Augmented Dickey–Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS). While the ADF test evaluates the null hypothesis of a unit root, the KPSS treats stationarity as the null, meaning that using both together provides a more robust assessment of integration order (Ilu & Prasad, 2024). Applied to the log series, these tests indicated non-stationarity in levels

but stationarity after first-order differencing (see figure 3.1), supporting the specification of an ARIMA $(\cdot, 1, \cdot)$ model.

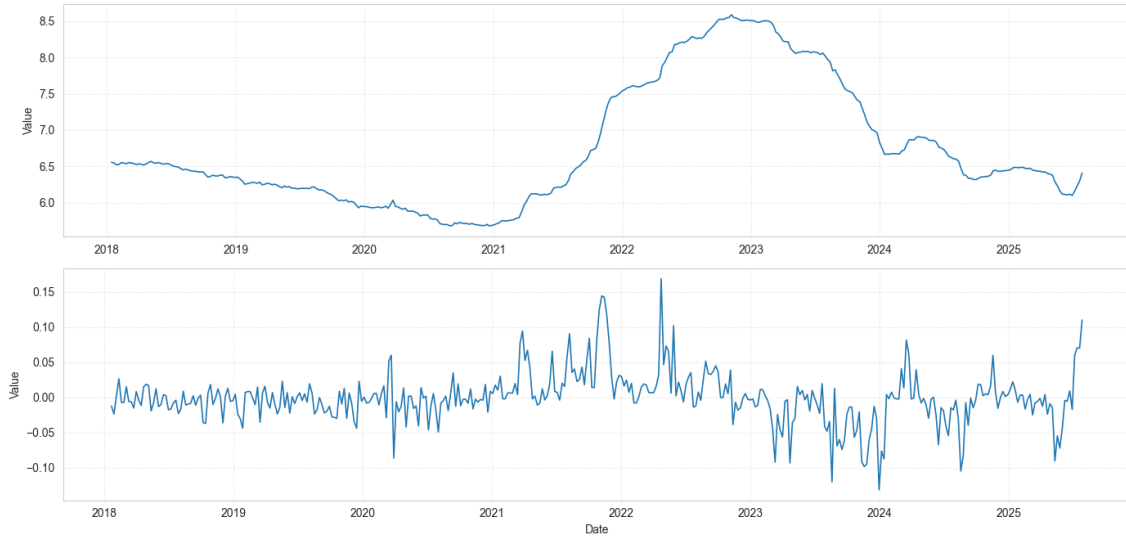


Figure 3.1: First Differencing of the time series in LOG

Alongside the baseline ARIMA, a log-ARIMA specification was retained to evaluate the effect of variance stabilisation on forecasting performance. Together, these models provided rigorous statistical benchmarks against which the performance of more advanced machine learning and hybrid frameworks could be evaluated in later sections.

3.4.2 WAVELET + ARIMA Hybridisation

To address the limitations of linear time series models in capturing non-stationary and multi-scale dynamics, this study integrates the DWT with ARIMA to form a Wavelet + ARIMA framework. The wavelet transform is particularly suited to financial and commodity price series because it decomposes signals simultaneously in both the time and frequency domains, enabling short-term fluctuations and long-term trends to be modelled separately (Crowley, 2007; Kriechbaumer et al., 2014). This multiresolution analysis reduces noise, stabilises variance, and provides a structured representation of market cycles that standard ARIMA alone often fails to capture (Paul & Garai, 2021; Silva et al., 2021).

In this study, the log-transformed spodumene series was decomposed into trend and detail components, which were then modelled individually before being recombined through inverse wavelet reconstruction (see figure 3.2). By separating cyclical signals prior to ARIMA modelling, the framework captures both long-term structure and short-term irregularities. This methodological enhancement builds on evidence from prior studies, which have shown that Wavelet + ARIMA consistently improves forecasting accuracy for metals, energy, and agricultural commodities (Kriechbaumer et al., 2014; Kumar & Bhattacharyya, 2024; Paul & Garai, 2021) (see Appendix C for more information).

Within this thesis, the Wavelet-ARIMA serves as an intermediate benchmark, bridging purely statistical models and the more advanced machine learning hybrids discussed in subsequent sections.

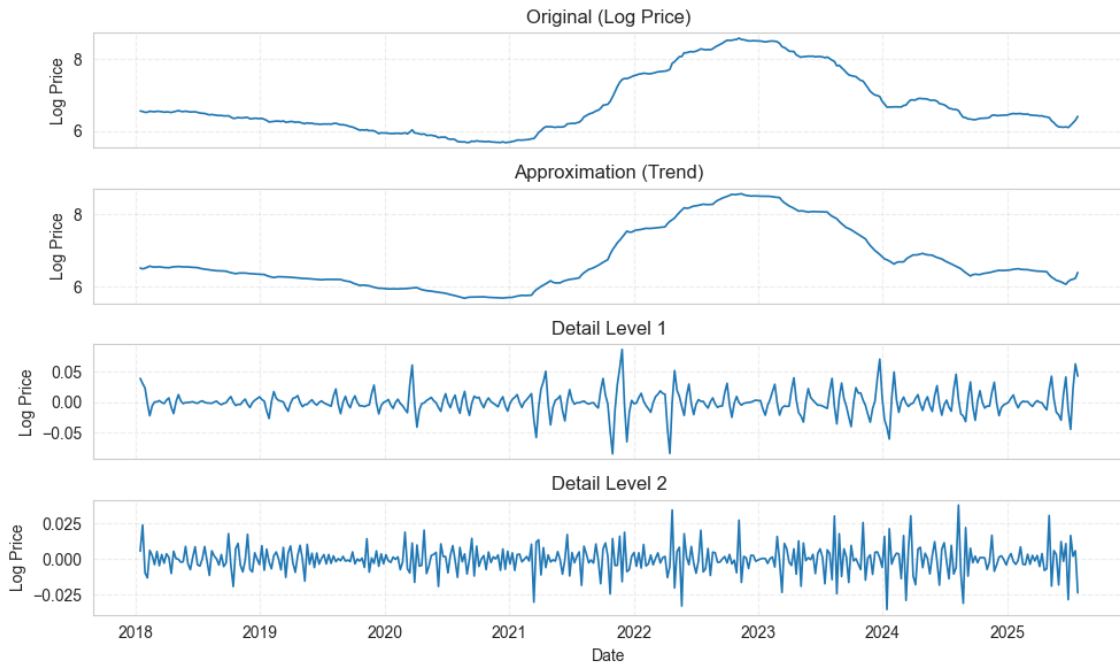


Figure 3.2: Approximation and detail components.

3.4.3 Machine Learning Models

To complement the statistical benchmarks, three supervised machine learning (ML) algorithms were employed: Elastic Net regression, Extreme Gradient Boosting (XGBoost), and LSTM networks. These models were selected to capture nonlinear dependencies, complex interactions, and long-range temporal structures in the spodumene series, each offering distinct methodological strengths validated in prior commodity and financial forecasting studies (Sari et al., 2024; Varshini et al., 2024).

Elastic Net served as a transparent baseline linear method. By combining L1 (Lasso) and L2 (Ridge) penalties, EN mitigates multicollinearity while simultaneously performing variable selection (E. Rapach & Zhou, 2020). XGBoost, an ensemble method, was implemented for its ability to model nonlinear interactions and volatility clustering in high-dimensional data (Wang & Guo, 2020). Finally, LSTM networks, a form of RNN, were adopted to capture sequential dependencies beyond the reach of fixed-lag structures, retaining long-term memory of historical price movements (Sari et al., 2024).

All ML models were trained and tuned following a two-stage workflow: (i) baseline training using default settings, and (ii) hyperparameter optimisation with time-series cross-validation. Their forecasts were benchmarked against a Naive-H model, which assumes that the most recent observed price persists into the forecast horizon. This provided a strong yet simple baseline, especially relevant in volatile commodity settings (Hyndman & Athanasopoulos, 2014).

To ensure methodological rigour, evaluation was carried out at both short- (two weeks) and medium-term (twelve weeks) horizons. Diagnostics included actual versus predicted series, parity plots, residual distributions, error-over-time charts, and rolling skill comparisons relative to Naive-H. These checks verified not only predictive accuracy but also model stability and robustness.

In summary, Elastic Net provided a parsimonious linear ML benchmark, XGBoost captured nonlinear interactions and hierarchical structure, and LSTM extended predictive capacity through long-range sequential modelling. Together, these models established a robust methodological foundation for the hybrid frameworks discussed in the next section (see Appendix C).

3.4.4 Advanced Hybrid Framework

Hybrid forecasting models have gained prominence because they integrate the strengths of multiple approaches while compensating for their individual weaknesses. Traditional statistical methods such as ARIMA are effective at modelling linear temporal dependencies but struggle with nonlinear and regime-switching behaviours often observed in commodity markets (Tsoku et al., 2024). Machine learning models, in contrast, capture complex nonlinear mappings but may fail to account for autocorrelation structures if used in isolation (Foroutan & Lahmiri, 2024). Wavelet decomposition adds further value by providing a multi-resolution framework that separates trend and high-frequency components, thereby improving signal extraction and noise reduction (Crowley, 2007; Kriechbaumer et al., 2014).

Building on these insights, this study developed three tiers of hybrid models: (i) ARIMA + ML residual hybrids, (ii) Wavelet + ML hybrids, and (iii) integrated Wavelet + ARIMA + ML hybrids. Each category was designed to exploit complementary strengths while maintaining rigorous evaluation through validation and test partitions, benchmarked against ARIMA-only baselines and a Naive hold-last comparator (Naive-H). Although simplistic, the Naive-H remains a robust baseline in volatile commodity markets, making it a meaningful benchmark (Hyndman & Athanasopoulos, 2014).

Residual hybrids combined ARIMA with machine learning by using ARIMA to model the linear component and applying ML to the residuals, capturing nonlinear dynamics overlooked by the statistical model. Wavelet + ML hybrids applied wavelet decomposition prior to machine learning, allowing trends and high-frequency components to be modelled separately before recombination. The most comprehensive approach, integrated Wavelet + ARIMA + ML hybrids, routed components according to their statistical properties, with linear signals modelled by ARIMA and nonlinear signals by ML. Prior research has shown that such integrated hybrids consistently outperform single-model or two-model frameworks by aligning the strengths of each approach to the appropriate data characteristics (Bhuvaneshwari & Rajini, 2024; Kumar & Bhattacharyya, 2024).

Together, these hybrid strategies provide a methodological bridge between traditional statistical approaches and advanced machine learning, offering the flexibility required to capture the complex, nonlinear, and multi-scale dynamics of spodumene prices.

3.5 Experimental Design and Evaluation

3.5.1 Forecasting Framework

To ensure robust and unbiased evaluation, this study employs an out-of-sample forecasting framework based on chronological data splitting and time-series cross-validation

(Hyndman & Athanasopoulos, 2014). This approach prevents data leakage and better reflects real-world forecasting, where future information is unknown (Varshini et al., 2024). The dataset, covering January 2018 to July 2025, was divided into three non-overlapping sets: a training set, a 12-week validation set, and a 12-week test set.

Model development and hyperparameter tuning were carried out on a “development” set (training + validation). A rolling-origin expanding window was used for cross-validation, where models were trained on incrementally larger samples and evaluated on the following 12-week block (Byers et al., 2021). This provided more reliable error estimates than a single static split. After selecting the best hyperparameters, models were retrained on the full development set and then evaluated on the unseen test set. Forecasting performance was assessed at both short- (two weeks) and medium-term (twelve weeks) horizons, ensuring a fair and consistent comparison across all approaches.

3.5.2 Evaluation of Predictive Power

To assess forecasting performance, this study employed a framework combining accuracy metrics, benchmark comparisons, and residual diagnostics. Using multiple criteria ensures that improvements reflect genuine model strength rather than artefacts of sampling or overfitting (Tami & Owda, 2024; Varshini et al., 2024).

Four error measures were applied: MAE, RMSE, MAPE, and the Coefficient of Determination (R^2). These capture different aspects of predictive accuracy. Models including ARIMA, Wavelet + ARIMA, Elastic Net, XGBoost, LSTM, and hybrid variants were benchmarked against the Naive-H model, which assumes the last observed price persists into the forecast horizon. Despite its simplicity, Naive-H is a difficult baseline to outperform in volatile markets (Hyndman & Athanasopoulos, 2014).

Beyond point metrics, residual analysis ensured model adequacy. The Ljung–Box test was used to detect autocorrelation, while residual distributions were checked for normality. Rolling error plots and skill scores (MAE relative to Naive-H) were also produced, providing insight into temporal stability and market conditions under which models performed best (Kumar & Bhattacharyya, 2024; Shameem et al., 2025).

This combined approach of metrics, benchmarks, and diagnostics ensured a balanced evaluation of both predictive accuracy and robustness, supporting meaningful comparisons across statistical, machine learning, and hybrid models.

Chapter 4

Results

4.1 Overview

This chapter reports empirical findings from the forecasting framework set out in the Methodology. We begin with a brief exploratory view of the spodumene price series, then evaluate three strands of models: statistical benchmarks (Naïve-H and ARIMA), machine-learning benchmarks (Elastic Net, XGBoost, LSTM), and hybrid frameworks that combine statistical structure with non-linear learners, including wavelet-based variants. Forecasts are produced at $H = 2$ and $H = 12$ weeks (Results explained in Appendix D) and evaluated on the price scale using MAE, RMSE, and MAPE; R^2 is reported for completeness but is not used for ranking. To keep the main text focused, hyperparameters and extended diagnostics (residual Autocorrelation Function (ACF)/Partial Autocorrelation Function (PACF), Ljung–Box, and rolling-error panels) are provided in the Appendices D. Exact validation and test windows are stated once below and then used consistently throughout. Each section presents headline metrics and concise figures, with a short takeaway that motivates the next method. Detailed interpretation and implications are developed in the Discussion chapter.

4.2 Exploratory Data Analysis (EDA)

The raw spodumene price series displays Volatility typical of commodity markets, which sets important context for forecasting. After a stable period during 2018–2020, prices rose sharply through 2021–2022 alongside surging lithium demand, before correcting in 2023. These swings underscore pronounced non-stationarity and the need for careful pre-processing (Byers et al., 2021). **CHANGE THE IMAGE LABEL HERE**

To ensure temporal consistency, the series was aligned to a weekly Sunday index and missing weeks were linearly interpolated. This preserves the underlying trajectory without injecting spurious volatility an approach recommended by Byers et al., 2021. Figure 4.1 contrasts the original and interpolated series.

Descriptive profiling highlights a highly dispersed, right-skewed distribution in which the mean exceeds the median, reflecting an extended upper tail. Figure 4.2 illustrates this shape. Detailed summary statistics are reported in table D.1 (Appendix D).

Outlier screening (z-scores and IQR bands) flagged numerous extreme observations,

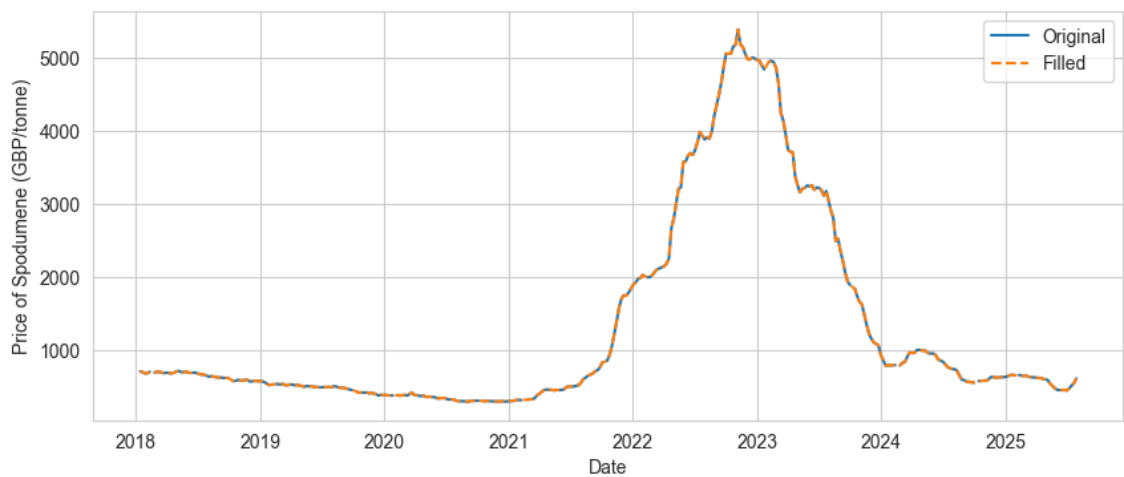


Figure 4.1: Spodumene Price (Original vs. Filled).

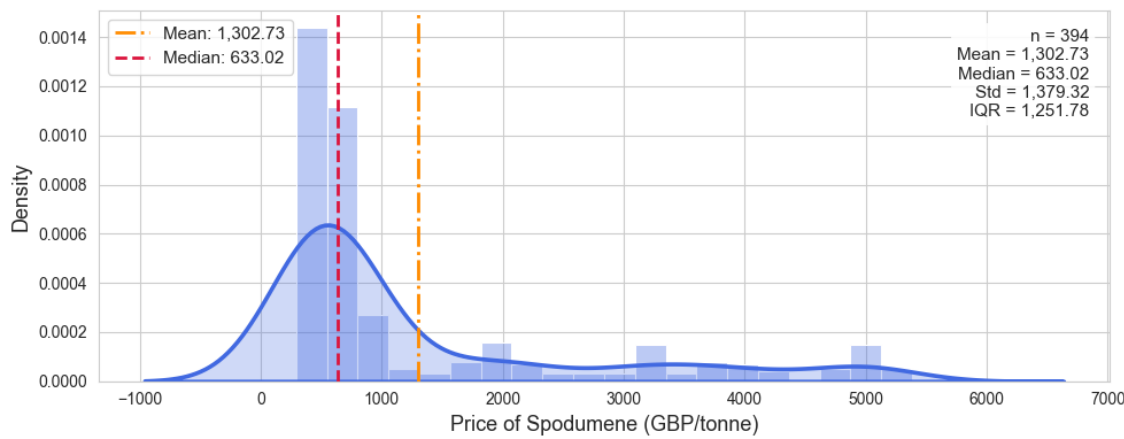


Figure 4.2: Histogram with KDE of Filled Spodumene Prices.

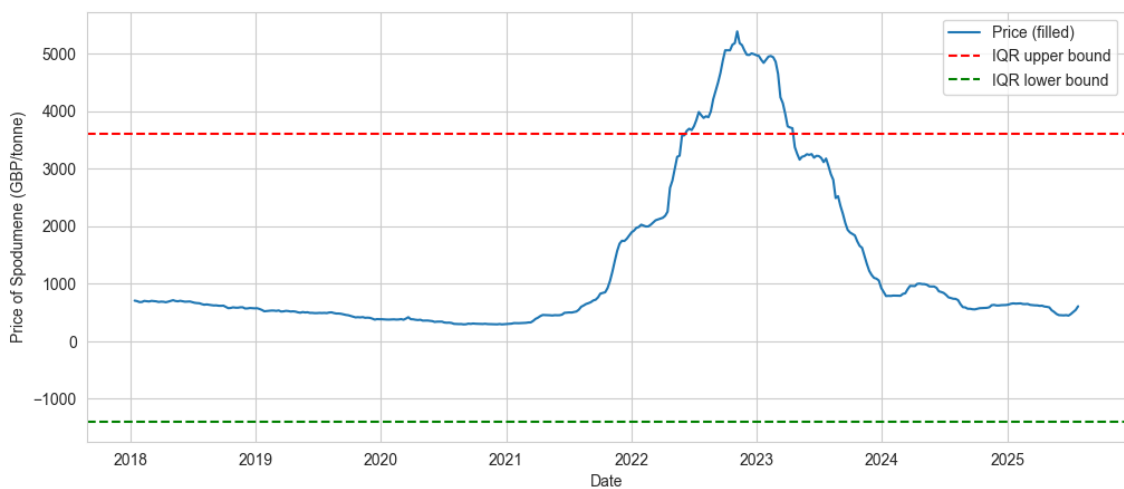


Figure 4.3: Spodumene Prices with IQR Bound.

concentrated in the 2021–2023 boom. These are retained as genuine market signals in commodities, spikes arise from supply bottlenecks, policy shifts and demand shocks, and their removal would discard meaningful information. Recent studies likewise show hybrid models are more robust when true volatility is preserved (Foroutan & Lahmiri, 2024; Varshini et al., 2024). Figure 4.3 shows the price path against IQR bounds.

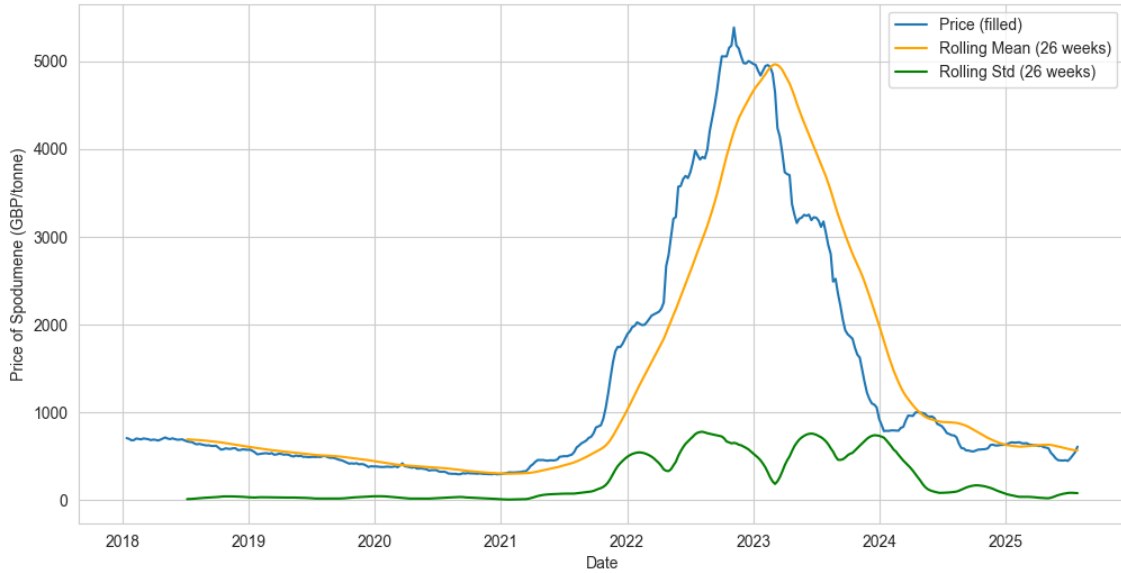


Figure 4.4: Rolling Mean and Volatility (Raw vs Log-Transformed).

Volatility is time-varying, with markedly higher variance in boom phases (Rubio et al., 2023). A natural log transform reduces skew, compresses extremes, and improves suitability for models that assume near-constant variance (Crowley, 2007; Lütkepohl & Xu, 2012). Rolling mean and standard deviation plots confirm smoother behaviour in logs relative to raw levels 4.4.

Seasonal-trend decomposition of the log series reveals a strong upswing from mid-2021, a sharp correction through 2023, and modest seasonal fluctuations consistent with industrial cycles 4.5. These supports separating long-run trend from shorter-cycle dynamics in model design.

Formal stationarity checks indicate that the log-level series is non-stationary, whereas first differencing yields a stationary representation suitable for ARIMA-type models (Kriechbaumer et al., 2014).

Taken together, the EDA shows a volatile, structurally irregular series with clear regime shifts. The adopted treatments weekly alignment and interpolation, retention of economically meaningful extremes, log transformation, and differencing render the data statistically tractable while preserving market realism (Byers et al., 2021; Taquiri et al., 2024). This provides a sound basis for feature engineering and for evaluating statistical, machine-learning and hybrid models against a common baseline.

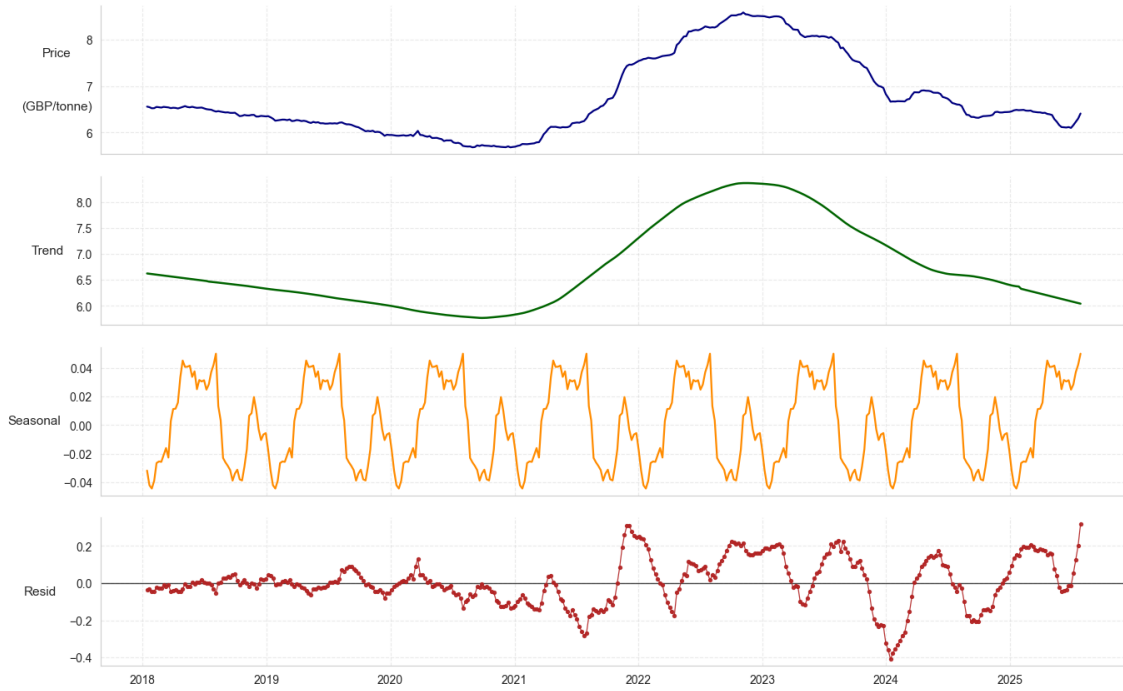


Figure 4.5: Seasonal Decomposition of Log-Transformed Prices.

4.3 Benchmark Models

4.3.1 ARIMA Model

As the statistical benchmark, an ARIMA was estimated on the log-transformed series. Unit-root testing indicated stationarity after one difference (ADF $p = 0.00034$; KPSS $p = 0.10$), so a stepwise search was run over (p, d, q) orders. The best model by Akaike Information Criterion (AIC) was ARIMA(1, 1, 1) (AIC = -1692.5). Residual checks supported adequacy for one-step-ahead use: Ljung–Box $p \approx 1.00$ at 10 and 20 lags (no remaining autocorrelation). Jarque–Bera indicated fat-tailed residuals (kurtosis ≈ 6.9), which is typical in commodity prices (see Appendix D for graphs and residuals).

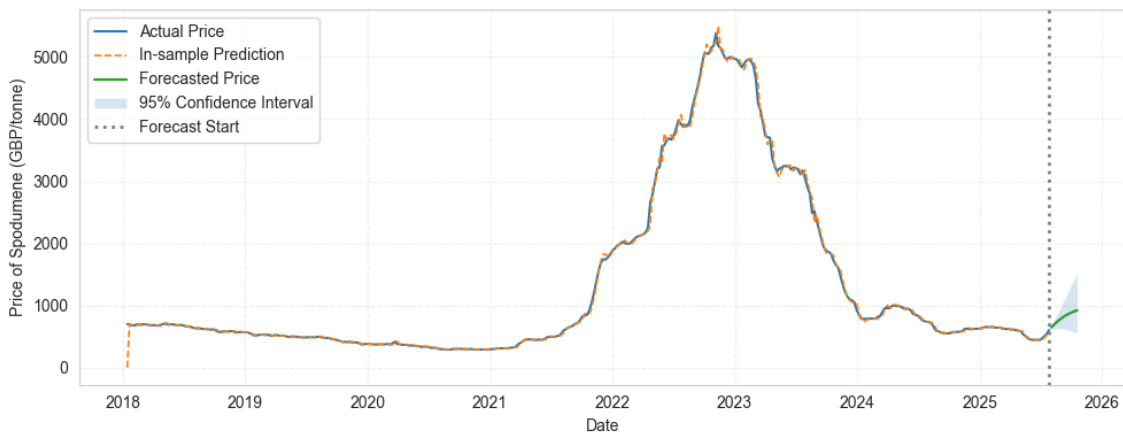


Figure 4.6: ARIMA forecast for the next 12 weeks.

Re-expressed on the price scale, ARIMA achieved $\text{MAE} = 28.48$, $\text{RMSE} = 64.59$, and $\text{MAPE} = 2.21\%$ over the evaluation horizon. Against Naive-H, which exhibits larger test errors in this period, ARIMA provides a clear improvement in absolute and percentage terms. Visually, the in-sample fit tracked the 2021–2022 4.6 upswing and the 2023 correction, with forecast intervals widening appropriately into the hold-out. These diagnostics position ARIMA as a robust linear baseline, while signalling that purely linear dynamics do not fully capture tail behaviour.

4.3.2 XGBoost Machine Learning Benchmark Model

To address non-linear interactions and potential regime shifts, XGBoost was tuned on the development set using time-series cross-validation with time-decay weighting (see Appendix D). The best single model by validation MAE produced:

- Validation: $\text{MAE} = 1.98$, $\text{RMSE} = 2.42$, $\text{MAPE} = 0.32\%$ (better than Naive-H: 10.55, 12.03, 1.71%).
- Test: $\text{MAE} = 35.31$, $\text{RMSE} = 42.44$, $\text{MAPE} = 6.86\%$ (improves on Naive-H: 45.32, 55.48, 8.80)

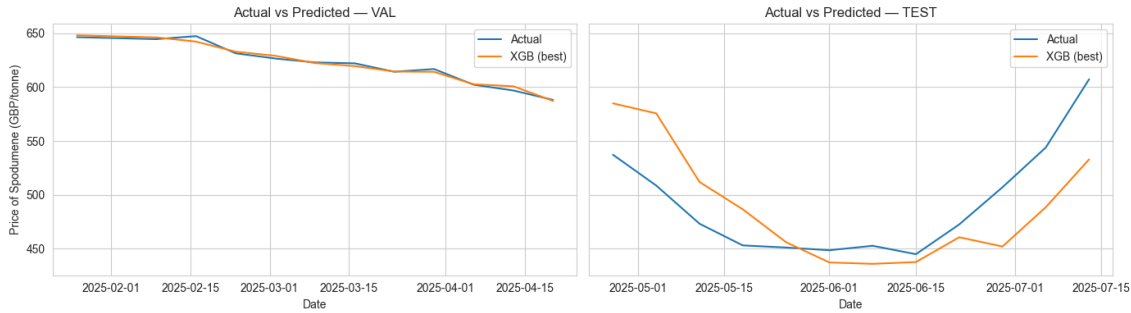


Figure 4.7: Actual vs Predicted for XGBoost on validation and test.

Actual-versus-predicted (figure 4.7) panels show tight alignment on validation and some under-prediction during the mid-2025 recovery on test, with a handful of turning-point misses. Overall, tuned XGBoost materially outperforms the Naive baseline and the linear benchmark across both splits, although accuracy decays when the regime shifts quickly (More results and graphs are in Appendix D).

In summary, the ARIMA model provides a transparent, steady baseline, while the tuned XGBoost model delivers stronger accuracy where non-linear structure is present. This justifies the hybrid and wavelet-based models that follow, aimed at reducing turning-point errors and short-memory residual structure.

4.4 Residual Integrated Models (ARIMA + ML Models)

We first obtain an ARIMA forecast in log space at horizon H and define a residual as the difference between the realised value and the ARIMA forecast at $t + H$ (t = forecast origin time index and H = forecast horizon). Each hybrid then trains a supervised learner on

this residual using the same engineered, leakage-safe features as in the benchmark ML search (Zhang et al., 2023). The predicted residual is added back to the ARIMA signal, and evaluation is done on the price scale. As elsewhere, we report MAE, RMSE and MAPE on the validation (VAL) and out-of-sample TEST windows, alongside Naive-H for context. Extended diagnostics and hyperparameters are provided in the appendices.

4.4.1 ARIMA + XGBoost

Gradient-boosted trees are trained on the ARIMA residual target using the tuned configuration from the benchmark search (time-series CV with time-decay weighting). Models are fit chronologically (Wang & Guo, 2020).

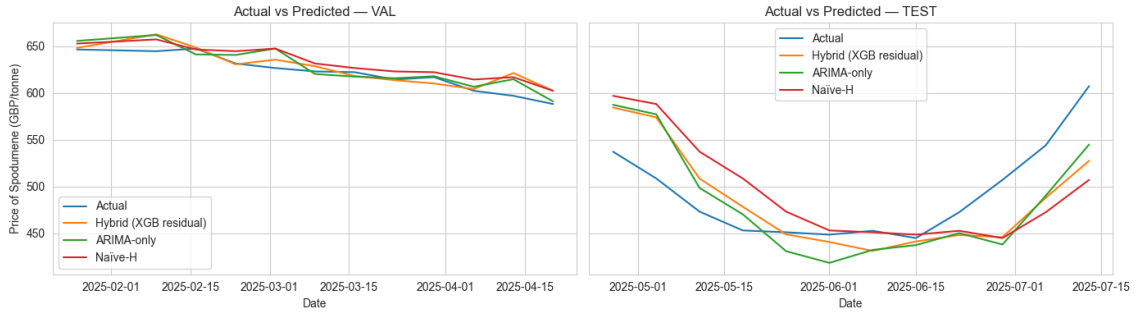


Figure 4.8: Actual vs Predicted for ARIMA + XGBoost on validation and test.

Split	MAE	RMSE	MAPE
Validation	7.34	10.41	1.20%
Test	35.79	43.53	6.92%

Table 4.1: Validation and test scores (price scale).

Relative to ARIMA-only, MAE and MAPE fall by about 9% and 8%; gains over Naive-H are around 30% on MAE and MAPE on validation (Table 4.1). Similarly on test MAE improves on ARIMA-only (37.21, 42.76, 7.31%) and beats Naive-H (45.32, 55.48, 8.80%) (Table 4.1). RMSE is fractionally higher than ARIMA-only, consistent with trees trading a small increase in spread for lower absolute error. Error-over-time shows the residual learner dampening ARIMA’s early May under prediction and partially catching the late June inflection. Figure 4.8 shows the prediction vs actual graph for the model fit. Twelve-week frozen feature projections track the ARIMA midline and remain within ARIMA’s empirical band, with a mild conservative bias late in the horizon. Residual autocorrelation is close to white on VAL, on TEST, modest lag-1/lag-2 structure remains around turning points. Full ACF/PACF, Ljung–Box summaries and rolling-error panels are in Appendix D.

Among residual learners, XGBoost is the only one that consistently improves ARIMA on TEST across MAE and MAPE while comfortably outperforming Naive-H. This motivates the wavelet and integrated hybrids that follow, aimed at further reducing turning-point errors and remaining short-memory structure.

4.5 WAVELET Based Models

We decompose the log price into a trend and two detail components using a Daubechies-2 wavelet at level 2 with periodisation (Kriechbaumer et al., 2014). For each component we train a supervised learner on the same leakage-safe feature matrix used in benchmarks. Features are scaled with a history-only Standard-Scaler and models are fit chronologically with time-decay weighting where applicable. Predicted component deltas are summed to obtain the H-step log return and mapped back to the price scale via exponentiation; Naive-H is computed in the same space for a fair comparison. Accuracy is also reported using MAE, RMSE and MAPE on the price scale (R^2 in the Appendix D). The helper routines implement metrics, leakage-safe scaling, and windowing, per component training, and price-scale reconstruction, full code and hyperparameters are listed in the appendices.

4.5.1 WAVELET + ARIMA

The log price was decomposed into multi-scale components using a Daubechies-2 wavelet at level 2 (approximation “trend” plus two detail bands). Stationarity tests indicated the trend was non-stationary (ADF $p = 0.404$; KPSS $p = 0.01$), while `detail_1` and `detail_2` were stationary by ADF ($p \approx 3.06 \times 10^{-13}$ and 2.99×10^{-23}). ARIMA orders selected by AIC on the training set were *trend* (2,2,2), *detail_1* (5,0,2), and *detail_2* (5,0,1). Component forecasts were summed in log space and back-transformed via the exponential; a short burn-in around the fitted path was masked to avoid back-transform amplification of early instability.

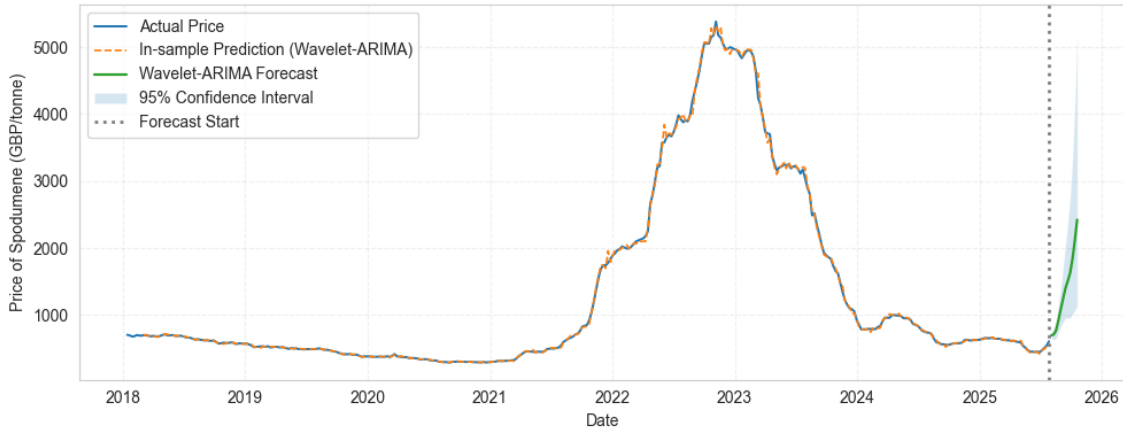


Figure 4.9: WAVELET + ARIMA forecast for the next 12 weeks.

Residual diagnostics on the combined log series were mixed: Ljung–Box was non-significant at 10 lags ($p = 0.202$) but marginal at 20 lags ($p = 0.024$), and Q–Q plots showed fat tails (Appendix D). At the final origin used for diagnostics, Wavelet–ARIMA produced MAE = 17.22, RMSE = 38.45 and MAPE = 1.26% (price scale) with a visually tight in-sample fit (figure 4.9).

Under the rolling-origin H-step evaluation used throughout this chapter (here $H = 12$ weeks), the pure Wavelet–ARIMA was fragile. Relative to Naive-H, errors were larger on both splits (Validation: MAE = 64.71, RMSE = 70.80, MAPE = 10.51%; Test: MAE =

149.80, RMSE = 164.44, MAPE = 31.28%). Large test errors reflect repeated overshoots and whipsaws at turning points once component models are re-estimated origin-by-origin and summed. Forecast plots also show wide uncertainty bands relative to realised variability. In short, decomposition plus linear ARIMA at each scale did not translate in-sample adequacy into robust multi-origin H-step accuracy.

4.5.2 WAVELET + XGBoost

We trained XGBoost on H-step component deltas (in log space) for each wavelet band and summed predicted deltas at forecast time to obtain the overall H-step log return before converting back to prices. The tuned boosted-tree configuration (Appendix C) was reused with per-band models trained chronologically. Performance improved markedly over Wavelet-ARIMA and beat Naive-H on the hold-out (see Table 4.2):

Split	MAE	RMSE	MAPE	Naïve-H MAE	Naïve-H RMSE	Naïve-H MAPE
Validation	10.34	13.36	1.66%	10.55	N/A	1.71%
Test	41.37	48.55	8.12%	45.32	55.48	8.80%

Table 4.2: Validation and test scores (model vs. Naïve-H baseline).

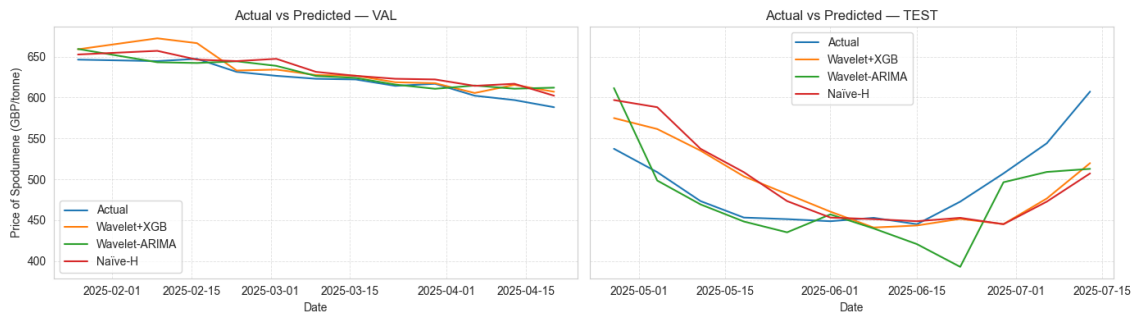


Figure 4.10: Actual vs Predicted for WAVELET + XGBoost on validation and test.

Compared with Wavelet + Elastic Net, XGBoost gives slightly weaker validation but similar test accuracy (Figure 4.10). Error over time panels reveal the largest residuals at turning-point weeks (mid-2025). Residual ACF/PACF suggests some remaining short-memory at lags 1–2 on the test window, smaller than in pure Wavelet + ARIMA but not fully eliminated.

Relative to the best standalone tuned XGBoost (Section 4.3.2 test MAPE 6.86%), the wavelet-augmented variant’s test MAPE (8.12%) is slightly worse, implying boosted trees already capture much of the relevant multi-scale structure from engineered features.

4.6 Machine Learning + ARIMA + WAVELET

Building on the literature review and methodology, this section routes forecasting capacity to the time–frequency structure of the series. The log price is decomposed with a

Daubechies-2 wavelet at level 2, yielding three components (trend, detail_1, detail_2). A McLeod–Li (McLeod & Li, 1983) test on squared residuals is used to classify components as “linear” (modelled by ARIMA) or “non-linear” (modelled by ML). For this dataset, all components were flagged non-linear, so per-component ML is employed, while an all-ARIMA recompositing (Wavelet–ARIMA) serves as the statistical reference. Forecasts are generated at $t + H$, recombined in log space, and mapped back to the price scale. Estimation is leakage-safe ARIMA orders (where used) are selected on training only ML models follow a chronological procedure.

4.6.1 XGBoost + ARIMA + WAVELET

Gradient-boosted trees are trained per component on lagged features with time-decay weighting and early stopping (configuration inherited from Section 4.3.2), then summed in log space to form the 12-week return before back-transforming to prices.

Results (price scale) -

- Validation: MAE = 16.35, RMSE = 21.75, MAPE = 2.62%.
- Test: MAE = 24.82, RMSE = 30.27, MAPE = 4.94%.

Against the Wavelet–ARIMA baseline (Test: MAE = 31.28, RMSE = 44.15, MAPE = 6.05%), the per-component XGBoost reduces test MAE by $\sim 21\%$, RMSE by $\sim 31\%$, and MAPE by $\sim 18\%$, with a meaningful gain in explanatory fit.

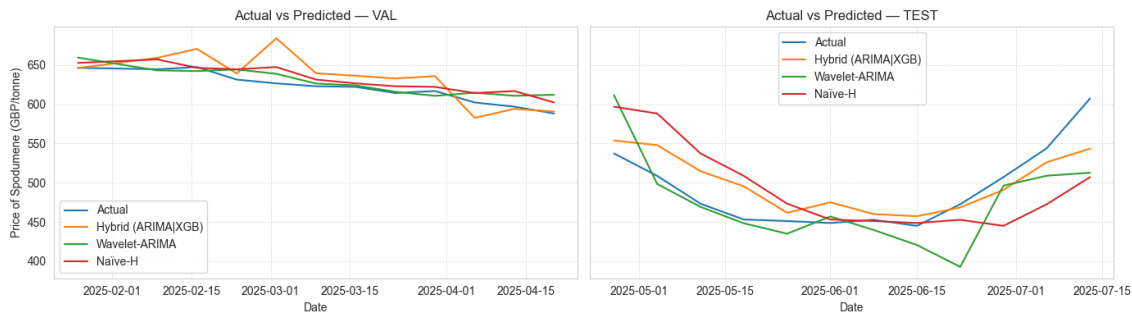


Figure 4.11: Actual vs Predicted for WAVELET + ARIMA + XGBoost on validation and test.

Diagnostics show tighter parity and lower rolling errors during the downturn and rebound, while residual ACF on the test window retains small lag-1/lag-2 structure around turning points. In Figure 4.11 the validation set, all models closely follow the actual price trend, with minimal deviation. In the test set, errors widen, but the hybrid (ARIMA + XGB) model shows better alignment with actual prices compared to Wavelet–ARIMA and Naive-H.

Per-component boosting generalises better than Wavelet + ARIMA at $H = 2$, capturing non-linear adjustments between scales and improving turning-point behaviour. Relative to the best standalone XGBoost (Section 4.3.2), the integrated variant delivers a lower test MAPE at $H = 2$ as seen in Table 4.3, indicating that routing learning through the wavelet components is beneficial at the medium horizons.

Metric	ARIMA		Wavelet+ARIMA		ARIMA+XGBoost		Wavelet+XGBoost		Wavelet+ARIMA+XGB	
	VAL	TEST	VAL	TEST	VAL	TEST	VAL	TEST	VAL	TEST
MAE	8.1002	37.2088	8.9919	31.2760	8.0764	62.9452	10.3380	41.3714	16.3540	24.8168
RMSE	10.4442	42.7598	11.1371	44.1513	9.3000	77.6348	13.3620	48.5492	21.7541	30.2723
MAPE (%)	1.2968	7.3117	1.4646	6.0451	1.3031	12.4441	1.6587	8.1192	2.6200	4.9400

Table 4.3: Validation/Test metrics for each model (excluding R^2).

Table 4.4 reports p -values from key residual diagnostic tests for each model. The Augmented Dickey–Fuller (ADF) and KPSS statistics check stationarity of the residuals, while the Ljung–Box test assesses remaining autocorrelation. ARIMA residuals just pass the ADF but marginally fail KPSS, with some autocorrelation ($p \approx 0.014$). Wavelet–ARIMA improves stationarity (ADF ≈ 0.276 , KPSS ≈ 0.035) but retains slight correlation. ARIMA+XGBoost and the integrated Wavelet–ARIMA+XGB hybrids show higher ADF p -values and Ljung–Box p -values, indicating more stationary, white-noise residuals, whereas Wavelet+XGBoost exhibits a lower Ljung–Box p -value.

	ARIMA	Wavelet+ARIMA	ARIMA+XGBoost	Wavelet+XGBoost	Wavelet+ARIMA+XGB
ADF p -value	0.102905	0.276021	0.534355	0.186812	0.592377
KPSS p -value	0.099754	0.034524	0.100000	0.100000	0.100000
Ljung–Box p -value	0.013707	0.175042	0.887782	0.000153	0.021578

Table 4.4: Residual diagnostic p -values by model.

Chapter 5

Discussion

5.1 Summary of Key Findings

At the two-week horizon, the strongest and most reliable performance came from wavelet plus machine learning hybrids in particular the elastic net and gradient boosted tree variants, which reduced errors relative to naive persistence and to a pure wavelet plus ARIMA reference while maintaining stable validation to test behaviour. The tuned single model XGBoost remained a high performing benchmark while ARIMA provided a steady and transparent baseline that improved on naive persistence in parts of the test window. Residual and parity diagnostics were supportive with residual autocorrelation close to white and only small, short memory around turning points which aligns with the visual error panels.

At twelve weeks, accuracy deteriorated across approaches with especially large errors for the pure wavelet plus ARIMA specification and with deep learning showing marked overfitting under limited data. This pattern is consistent with forecasting theory that multi step prediction accumulates uncertainty as the horizon lengthens as discussed by (Hyndman & Athanasopoulos, 2014) and it reflects the short public history of spodumene prices since 2018 which constrains model capacity and generalisation. The literature also cautions that complex learners are data hungry and tend to degrade when observations are scarce which further explains the weaker twelve-week results.

In summary, this study delivers clear gains at the operational two-week horizon led by wavelet plus elastic net and wavelet plus XGBoost configurations with ARIMA retained as a robust baseline while twelve-week forecasts are materially less accurate because of horizon driven error growth and data limitations.

5.2 Interpretation vs Objectives

The thesis set three objectives, and the results allow a clear interpretation against each. First, establishing transparent baselines shows that ARIMA improves materially on the Naive H comparator at the two-week horizon. On the test window ARIMA reduces MAE by eight points and lowers MAPE from about 8.8 percent to about 7.3 percent, relative to the Naive H comparator, which confirms its value as a steady statistical benchmark under the Box Jenkins paradigm (see Table 4.3).

Second, introducing standalone machine learning and wavelet-based variants delivers further accuracy at $H = 2$ with important nuances. Tuned XGBoost achieves the strongest single model performance and improves on Naive H by around ten MAE points and almost two percentage MAPE points. Elastic Net offers smaller but consistent gains and maintains stable validation to test behaviour. These patterns align with the expectation that flexible learners capture short horizon nonlinearity while remaining sensitive to data constraints.

Third, hybridisation yields mixed outcomes that are horizon dependent. Residual learning with ARIMA plus XGBoost trims ARIMA’s absolute errors on test while keeping robustness close to the linear baseline. Wavelet plus Elastic Net and wavelet plus XGBoost both beat Naive H on test, yet the wavelet augmented boosted trees are not superior to the best tuned boosted trees without decomposition, which implies engineered features already encode much of the useful multi scale information.

At twelve weeks, performance lags across families. Pure Wavelet ARIMA produces large validation and test errors, and residual sequence models trained on limited data degrade sharply out of sample. This pattern is consistent with multi step error accumulation and with the short history available for spodumene which constrains model capacity and magnifies regime shifts. The literature cautions that longer horizons compound uncertainty and that complex learners (high-capacity non-linear models—such as XGBoost and LSTM, including their wavelet-hybrid variants) are data hungry therefore degradation at $H = 12$ is expected in short series settings (Hyndman & Athanasopoulos, 2014).

Taken together all the research objectives are met at $H = 2$ where the modelling framework delivers clear gains over Naive H and improves on ARIMA with XGBoost and several hybrids. At $H = 12$ the objective is only partially met because data span and regime change limit generalisation and error propagation dominates accuracy despite careful design and diagnostics.

5.3 Findings in Context of the Literature

The pattern that achieves the best results in terms of low MAE, MAPE and RMSE scores at two weeks come from wavelet based and residual hybrid designs aligns with a broad body of evidence that multi resolution decomposition improves forecasting of volatile asset and energy prices. Wavelets isolate localised frequency bands which stabilises variance and lets models learn scale specific signals in a way that standard time domain approaches often miss (Crowley, 2007). Empirically, several studies report accuracy gains when wavelets are coupled with statistical or machine learning learners. Examples include electricity price forecasting where wavelet with ARIMA or kernel extreme learning machines outperforms single models across multiple markets (Zhang et al., 2023) and short horizon wavelet ELM that exceeds ARIMA for one-to-four-week horizons (Foroutan & Lahmire, 2024). Our wavelet plus elastic net follows this logic and its advantage over Naive persistence and over pure wavelet ARIMA on both validation and test is consistent with that literature.

However, ARIMA remains a credible benchmark. ARIMA models are easier to implement and often competitive/reliable in markets with strong autoregressive structure which underpins its continued use in energy and commodity forecasting and in forecasting pedagogy (Zhang et al., 2023). In our results ARIMA matches or is narrowly behind the

better hybrids at H equal to two, in terms of medium forecasting horizon which is coherent with the role of ARIMA as a strong baseline under Box Jenkins principles (Hyndman & Athanasopoulos, 2014).

By contrast deep learning performed less accurately in our setting. The introduction notes that public spodumene pricing exists only from 2018 which yields a short weekly history and pronounced structural breaks during the boom and correction. Under such conditions complex sequence models can fail to generalise (Varshini et al., 2024). Recent comparative evidence in commodities also suggests that tree-based methods can rival or surpass deep learning on tabular and modest sized datasets which is consistent with the superior performance of XGBoost in our residual hybrid and the weakness of the LSTM residual model (Foroutan & Lahmiri, 2024).

One divergence from parts of the energy price literature is the extremely poor performance of pure wavelet ARIMA in our data. Prior studies show wavelet ARIMA or wavelet with statistical learners can improve accuracy in electricity markets (Foroutan & Lahmiri, 2024; Zhang et al., 2023). Our weaker result reflects the combination of a short sample and sharper regime shifts in spodumene prices that reduce the benefit of modelling only the decomposed linear component.

Errors grow at longer horizons because multi step forecasts compound uncertainty and filter more high frequency information out of the signal which reduces responsiveness to turns. This principle is standard in forecasting texts and is visible in our twelve-week outcomes which deteriorate markedly relative to H equal to two (Hyndman & Athanasopoulos, 2014). As far as the author is aware, this is the first rigorous horizon specific evaluation of spodumene with a strict validation to test results. The findings confirm prior wavelet hybrid advantages noted in the literature by Kumar and Bhattacharyya, 2024 and qualify them under short histories and regime change.

5.3.1 Methodological Insights

The model design choices were guided by the statistical structure of weekly spodumene prices and by established practice in energy and commodity forecasting literature (Kriechbaumer et al., 2014; Md-Khair & Samsudin, 2018). The models allowed an exploration of the strengths and weaknesses of different designs.

This study found the benefits of wavelet preprocessing were sensitive to the transform design (for results see Appendix E). Following guidance that multi resolution analysis separates low frequency trend from high frequency fluctuations (Kriechbaumer et al., 2014) we experimented with mother families and levels (depth of multiresolution analysis) and observed that low levels stabilised short horizon forecasts while deeper decompositions helped at longer horizons. This pattern supports results in the literature around metals and electricity price forecasting, where an appropriate wavelet (mother) choice and decomposition depth improves accuracy and where the optimal number of levels often increases with forecast horizon (Crowley, 2007; Kriechbaumer et al., 2014).

We compared three forecasting strategies. The direct horizon strategy tuned models to produce the two weeks ahead value and was the most dependable at the desired principal horizon. Residual learning that adds a corrective model to a statistical baseline helped when short term shocks were prominent which is consistent with two stage schemes where a machine learning model refines a structural signal from a first stage (Qu et al., 2024;

Yang et al., 2017). Finally, we tested a component wise approach that models each wavelet sub series and then reconstructs the forecast. This improved stability and reduced large errors which aligns with evidence that forecasting de-noised components and recombining them outperforms single pass modelling on raw series in oil and electricity markets (Md-Khair & Samsudin, 2018).

However, it should be noted that the temporal range of the data is a limitation. The evaluation window spans 2018 to 2025, which includes structural shifts and episodes of extreme volatility. To reduce spurious persistence, we differenced or log transform where require and we avoid model complexity to avoid overfitting. These choices are standard for nonstationary commodity prices (Kriechbaumer et al., 2014; Md-Khair & Samsudin, 2018), and they help at two weeks yet limit extrapolation depth at twelve weeks where uncertainty compounds across steps and where short history cannot support high-capacity models. Prior work shows that deep learners benefit from larger samples and that hybrids guard against overfitting under volatility which matches our experience (Crowley, 2007; Qu et al., 2024).

Here, stability denotes temporal performance stability small validation-to-test drift, smoother rolling-error profiles, and residual whiteness while large errors are reflected by RMSE-sensitive blow-ups and visible spikes in rolling MAE/MAPE around turning points; under these definitions, the component-wise reconstruction reduced error spikes and improved DEV→TEST generalisation relative to single-pass models.

Diagnostics yielded three practical insights. First, enforcing stationarity and residual whiteness filtered out misspecified models and reduced spurious predictability; models that passed ADF/KPSS and Ljung–Box produced near white residual ACFs, making any remaining errors plausibly noise rather than missed structure (Kriechbaumer et al., 2014; Md-Khair & Samsudin, 2018). Second, parity plots and rolling-error profiles exposed bias and temporal instability well behaved models showed slopes near one, small intercepts, and smoother rolling MAE/RMSE, while unstable specifications revealed error spikes around turning points and weak DEV→TEST generalisation consistent with out-of-sample evaluation principles in commodity/finance forecasting (Md-Khair & Samsudin, 2018; Risse, 2017). Third, these diagnostics justified design choices: direct-H and wavelet-component models were retained because they combined accuracy with generalisation, whereas high capacity residual/deep learners that failed whiteness or showed volatile rolling errors were down-weighted or discarded, aligning with recent hybrid/practice evidence on robustness and generalisation (Kriechbaumer et al., 2014; Qu et al., 2024).

Reproducibility was enforced through chronological splits with a fixed validation block before the test period, strict feature lags, and benchmarking against Naive and AR-IMA baselines, as recommended by Hyndman and Athanasopoulos, 2014. This yielded two main insights. First, benchmarking hybrids against ARIMA and Naive rules confirmed that observed gains were genuine improvements rather than artefacts of overfitting hybrids that could not outperform these baselines were discarded. Second, aligning with commodity and electricity studies, this practice highlighted that ARIMA remains a robust comparator, while Naive rules serve as a stress test for short-term persistence, ensuring that any added model complexity translated into meaningful, out-of-sample value (Kriechbaumer et al., 2014).

5.4 Practical and Managerial Limitations

The forecasts made by this study are intended to support decision makers across the lithium value chain, including miners and concentrator operator's, converters of lithium hydroxide and carbonate cathode and cell manufacturers, procurement, treasury teams and market intermediaries.

Lithium prices are a strategic input for planning and risk management along the entire value chain, which heightens the value of reliable short horizon signals for operational discipline and for aligning investment with demand growth and supply tightness (Martin et al., 2017). In parallel evidence from power markets shows that market participants routinely use short term price forecasts to schedule production choose bidding strategies and manage exposure which motivates adopting methods that demonstrably improve on simple rules (Yang et al., 2017).

For the two-week horizon, our results show that wavelet-based machine learning configurations especially tree ensembles and regularised linear models reduce error relative to naive persistence and to an ARIMA baseline with stable behaviour from validation to test. These signals are suitable for timing purchases sequencing shipments setting weekly inventory buffers and placing short hedges within rolling planning cycles and they should be presented together with parity plots and rolling error panels when used for decisions (Md-Khair & Samsudin, 2018).

For the twelve-week horizon, forecast uncertainty widens and point accuracy deteriorates across model families. Therefore managers should treat medium horizon outputs as advisory and use them as scenario inputs rather than commitments adopt wider buffers in offtake contracts and quarterly budgets and seek corroboration from exogenous indicators. This caution follows forecasting guidance that uncertainty and variance increase with horizon which reduces the reliability of multi-step predictions for volatile commodities (Hyndman & Athanasopoulos, 2014).

A simple operating rule follows. Use the best performing two-week model identified in the results for short term tracking while retaining ARIMA and naive comparators as guardrails for governance. For twelve week planning triangulate model outputs with market intelligence and default to conservative positions when signals conflict. This practice aligns with commodity and energy studies where hybrid designs improve accuracy, yet decisions remain benchmarked against transparent baselines (Kriechbaumer et al., 2014).

5.5 Limitations

The empirical window is short for a relatively immature market. This project only had weekly spodumene pricing since 2018, which constrains the effective sample and raises the risk that parameter estimates are regime specific rather than structural. Missing or irregular weeks required interpolation and calendar alignment, which can smooth true volatility and attenuate turning points in a way that disadvantages multi step forecasts (Byers et al., 2021). In addition to data length, the series has been exposed to policy and auction shocks and to changes in contracting practice that introduce structural breaks beyond what linear differencing can absorb.

Model scope is univariate. Important covariates such as exchange rates, freight costs, auction conditions and policy signals were excluded despite evidence that these drivers

shape price formation in this market, which limits explanatory power and external validity.

Method choices also impose limits. Wavelet design and level selection were guided by pragmatic heuristics under a small sample, and boundary and leakage effects in discrete transforms can influence component signals used for forecasting (Crowley, 2007; Kriebaumer et al., 2014). Deep learning candidates were necessarily shallow and lightly tuned because the sample is short relative to their data needs, which invites overfitting and unstable generalisation.

Evaluation length further restricts inference. Out of sample testing spans a limited number of episodes, so performance estimates may be sensitive to episode mix and to horizon. Forecast errors grow with horizon because uncertainty compounds across steps which is a known property of multi-step prediction (Hyndman & Athanasopoulos, 2014).

Finally, transferability to other lithium products and to different market regimes should be treated cautiously given heterogeneous pricing mechanisms and evolving industry structure (Martin et al., 2017).

5.6 Recommendations for Practice and Future research

Recommendations for practice – For near term decisions procurement teams miners and converters should deploy the best validated $H = 2$ model from this study for weekly purchase timing shipment scheduling and short hedges since short horizon error growth is modest and the forecasts track the high frequency component that drives operational choices. Where volatility is elevated a wavelet plus machine learning specification is preferable because the multiresolution decomposition isolates local bursts and smoother trend content which improves fit on nonstationary series (Crowley, 2007) and has shown gains in energy and metals forecasting tasks (Risse, 2017; Yang et al., 2017). At $H = 12$ managers should treat point forecasts as scenario inputs not as executable targets because multi step uncertainty compounds and mapping learned on short samples transfers poorly across regimes (Hyndman & Athanasopoulos, 2014). Use wider inventory buffers and hedge bands for contracts beyond one quarter and corroborate with market intelligence on supply expansions policy moves and demand pulses in batteries and cathode chains which materially shape lithium price formation (Martin et al., 2017).

Model governance should be simple to implement, low in overhead, and focused only on critical diagnostics and auditable. Monitor rolling MAE or MAPE against a Naive benchmark and trigger review if the Naive closes the gap for two to three consecutive windows or if residuals fail whiteness tests such as Ljung–Box which signals misspecification or a regime break (Hyndman & Athanasopoulos, 2014). Prefer parsimonious components at $H = 2$ and keep capacity conservative given the 2018–2025 span and known shocks documented in the results file. Analysts should retrain models on a weekly basis with strict chronological splits and version every release with a diagnostics checklist including stationarity checks residual ACF PACF parity plots and rolling error stability as laid out in the methodology. When drift is detected fall back to the Naive H baseline or the ARIMA benchmark until stability returns and document the handover decision trail (Hyndman & Athanasopoulos, 2014).

Future research – Two paths are likely to add value. First there are exogenous drivers that theory and market evidence associate with lithium pricing. These EV adoption announcements supply expansions from spodumene and brine projects and bottlenecks in

refining capacity which prior work shows can shift level and variance of prices (Martin et al., 2017). A wavelet design can admit such signals at appropriate scales and has improved commodity and energy forecasts when paired with modern learners (Qu et al., 2024; Risse, 2017). Second adaptive learning models with online updates and regime recognition should be further explored, so that the hybrid and the machine learning models switches between short horizon trackers and smoother components when market structure changes which aligns with the logic of scale dependent dynamics in wavelet analysis (Crowley, 2007). Evaluations should move beyond error scores to economic value including hedge profit and loss and inventory cost impacts following practice in energy price forecasting where methods are judged by decision utility as well as accuracy (Yang et al., 2017). A transparent pipeline with reproducible splits and benchmark comparisons remains essential for stakeholder trust and for credible out of sample claims (Hyndman & Athanasopoulos, 2014).

Chapter 6

Conclusions

This thesis set out to develop and rigorously evaluate a reliable forecasting framework for weekly spodumene prices by combining transparent statistical baselines with multiresolution wavelet decomposition and lightweight machine-learning hybrids, assessed at short (operational) and medium horizons under strict chronological validation and diagnostics.

Consistent with evidence that frequency-aware preprocessing can expose structure that conventional models miss, the wavelet-enhanced specifications delivered the most reliable short-horizon gains, while well-specified ARIMA remained a strong guardrail for longer range predictions. These results align with commodity and energy studies showing that wavelet-based hybrids often outperform single-family models when volatility and non-stationarity are high (Kriechbaumer et al., 2014; Md-Khair & Samsudin, 2018; Yang et al., 2017).

Three empirical conclusions follow. First, at the near-term horizon, frequency-aware hybrids (wavelet plus linear/nonlinear learners) consistently reduced error relative to Naïve and pure linear references mirroring metals, oil, and electricity price studies where wavelet + ARIMA + ML variants improved 1 - 4-step forecasts (Kriechbaumer et al., 2014; Md-Khair & Samsudin, 2018; Yang et al., 2017). Second, as the horizon lengthened, accuracy deteriorated across families a regularity also documented in commodity and energy markets where multi-step uncertainty compounds and hybrid advantages narrow (Md-Khair & Samsudin, 2018; Yang et al., 2017). Third, performance proved sensitive to multiresolution design (transform, mother, level depth): metals evidence favours MODWT and longer wavelets, reinforcing the need to tune the decomposition, not just the learner (Kriechbaumer et al., 2014).

Methodologically, the study contributes a reproducible pipeline direct-H estimation, rolling-origin evaluation, parity and residual-whiteness checks, and baseline governance, which attribute gains to model design rather than leakage or metric artefacts. The emphasis on scale-wise routing is consonant with agricultural and financial applications that use wavelets to separate linear and nonlinear components, then match them to ARIMA and neural modules (Kumar & Bhattacharyya, 2024; Risse, 2017). Explicit white noise testing (for example Ljung–Box) after reconstruction supports trustworthy deployment claims (Kumar & Bhattacharyya, 2024).

Managerially, the clearest value from this study is in terms of making short horizon predictions. The best-validated, wavelet-enhanced model should guide purchase timing, shipment sequencing, inventory buffers, and short hedges benchmarked against ARIMA

to detect drift and preserve auditability (Yang et al., 2017). For medium horizon, forecasts should inform scenarios rather than executable targets; tightening governance (rolling MAE/MAPE versus Naive, residual checks) and widening decision buffers reflect practice in volatile markets (Md-Khair & Samsudin, 2018).

Limitations qualify external validity. The weekly series is short and regime-prone; modelling was primarily univariate; and wavelet choices were pragmatically tuned under small-sample constraints. Boundary and padding effects in discrete transforms can distort endpoints; metals evidence explicitly flags this sensitivity and the benefits of MODWT under such conditions (Kriechbaumer et al., 2014). Deep models can excel with richer data or scaffolding but are not uniformly superior in short histories without decomposition or pattern conditioning (Qu et al., 2024; Yang et al., 2017).

Two extensions are most promising. First, incorporate exogenous drivers known to shift lithium pricing policy pulses, supply expansions, FX and freight, auction conditions and route them at appropriate scales within the wavelet design; related work on gold and energy shows how decomposed predictors and sparse machine learning can add value (Risse, 2017; Yang et al., 2017). Second, explore adaptive learning with regime recognition and online updates, including wavelet-neural architectures with calibrated initialization (e.g., ELM-initialized WNN) to manage volatility bursts (Qu et al., 2024).

In sum, a diagnostics-first, frequency-aware framework materially improves short-horizon spodumene forecasts over simple persistence and purely linear baselines, while clarifying why advantages attenuate with horizon under shifting regimes. The deliverable is two-fold a deployable short-horizon tool with transparent guardrails for procurement and risk teams, and a reproducible hybrid blueprint centred on principled wavelet design that can be extended with exogenous information and adaptivity to strengthen medium-horizon reliability (Kriechbaumer et al., 2014; Kumar & Bhattacharyya, 2024).

Chapter 7

References

- Abdulhafidh Ibrahim, A., Saeed, B. N., & Fadil, M. A. (2023). Forecasting Stock Prices with an Integrated Approach Combining ARIMA and Machine Learning Techniques ARIMAML. *Journal of Computer and Communications*, 11. <https://doi.org/10.4236/jcc.2023.118005>
- Asif, A. H., Li, C., Lim, H., & Sun, H. (2024). Australia's Spodumene: Advances in Lithium Extraction Technologies, Decarbonization, and Circular Economy. *Industrial and Engineering Chemistry Research*, 63(5), 2073–2086. https://doi.org/10.1021/ACS.IECR.3C04048/ASSET/IMAGES/MEDIUM/IE3C04048_0009.GIF
- Bhuvaneshwari, S., & Rajini, N. S. S. (2024). A novel hybrid model for stock price forecasting: combining arima, random forest, and gradient boosting techniques. *Library Progrss International*, 44(2s), 2027–2034.
- Byers, J. W., Popova, I., & Simkins, B. J. (2021). Robust estimation of conditional risk measures using machine learning algorithm for commodity futures prices in the presence of outliers. *Journal of Commodity Markets*, 24(March), 100174. <https://doi.org/10.1016/j.jcomm.2021.100174>
- Caldwell, R. (2015). For personal use only DYESOL LIMITED For personal use only. (May), 1–26.
- Crowley, P. M. (2007). A guide to wavelets for economists. *Journal of Economic Surveys*, 21(2), 207–267. <https://doi.org/10.1111/j.1467-6419.2006.00502.x>
- Debnath, L., & Shah, F. A. (2015). *Wavelet transforms and their applications, second edition*. <https://doi.org/10.1007/978-0-8176-8418-1>
- Dessemond, C., Lajoie-Leroux, F., Soucy, G., Laroche, N., & Magnan, J. F. (2019). Spodumene: The Lithium Market, Resources and Processes. *Minerals 2019*, Vol. 9, Page 334, 9(6), 334. <https://doi.org/10.3390/MIN9060334>
- E. Rapach, D., & Zhou, G. (2020). Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods. *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–33. <https://doi.org/10.1002/9781119751182.ch1>
- Foroutan, P., & Lahmiri, S. (2024). Deep learning systems for forecasting the prices of crude oil and precious metals. *Financial Innovation*, 10(1). <https://doi.org/10.1186/s40854-024-00637-z>
- Hyndman, R. J., & Athanasopoulos, G. (2014). Forecasting: Principles & Practice. *Antimicrobial Agents and Chemotherapy*, 58(12), 7250–7257. robjhyndman.com/

- uwa % 7B % 5C % % 7D5Cnhttp://robjhyndman.com/papers/forecasting-age-specific-breast-cancer-mortality-using-functional-data-models/
- Ilu, S. Y., & Prasad, R. (2024). Time series analysis and prediction of COVID-19 patients using discrete wavelet transform and auto-regressive integrated moving average model. *Multimedia Tools and Applications*, 83(29), 72391–72409. <https://doi.org/10.1007/s11042-024-18528-x>
- International Energy Agency. (2025). Global Critical Minerals Outlook 2025. *Report*, 282. https://www.iea.org/reports/global-critical-minerals-outlook-2024?utm%7B%5C_%7Dcontent=buffer3599f%7B%5C&%7Dutm%7B%5C_%7Dmedium=social%7B%5C&%7Dutm%7B%5C_%7Dsource=linkedin.com%7B%5C&%7Dutm%7B%5C_%7Dcampaign=buffer
- Kriechbaumer, T., Angus, A., Parsons, D., & Rivas Casado, M. (2014). An improved wavelet-ARIMA approach for forecasting metal prices. *Resources Policy*, 39(1), 32–41. <https://doi.org/10.1016/J.RESOURPOL.2013.10.005>
- Kumar, S. K., & Bhattacharyya, B. (2024). Wavelet-ARIMA-TDNN Model for Agricultural Commodity Price Forecasting. *Statistics and Applications {ISSN}*, 22(2), 217–229. <http://www.sscsca.org.in/journal>
- Li, H., Eksteen, J., & Kuang, G. (2019). Recovery of lithium from mineral resources: State-of-the-art and perspectives – A review. *Hydrometallurgy*, 189, 105129. <https://doi.org/10.1016/J.HYDROMET.2019.105129>
- Liu, Y., Ma, B., Lü, Y., Wang, C., & Chen, Y. (2023). A review of lithium extraction from natural resources. *International Journal of Minerals, Metallurgy and Materials*, 30(2), 209–224. <https://doi.org/10.1007/s12613-022-2544-y>
- Lütkepohl, H., & Xu, F. (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics*, 42(3), 619–638. <https://doi.org/10.1007/s00181-010-0440-1>
- Martin, G., Rentsch, L., Höck, M., & Bertau, M. (2017). Lithium market research – global supply, future demand and price development. *Energy Storage Materials*, 6, 171–179. <https://doi.org/10.1016/J.ENSME.2016.11.004>
- McLeod, A. I., & Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4, 269–273. <https://doi.org/10.1111/j.1467-9892.1983.tb00373.x>
- Md-Khair, N. Q. N., & Samsudin, R. (2018). Forecasting crude oil prices using wavelet ARIMA model approach. *Lecture Notes on Data Engineering and Communications Technologies*, 5, 535–544. https://doi.org/10.1007/978-3-319-59427-9_56
- Paul, R. K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. *Soft Computing*, 25(20), 12857–12873. <https://doi.org/10.1007/S00500-021-06087-4>
- Qu, Z., He, L., Ge, X., Wang, F., Xu, F., & Lu, J. (2024). A Two-Stage Forecasting Approach for Day-Ahead Electricity Price Based on Improved Wavelet Neural Network With ELM Initialization. *IEEE Transactions on Industry Applications*, 60(3), 5061–5073. <https://doi.org/10.1109/TIA.2024.3365456>
- Risse, M. (2017). Combining Wavelet Decomposition with Machine Learning to Forecast Gold Returns. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.3059293>

- Rubio, L., Palacio Pinedo, A., Mejía Castaño, A., & Ramos, F. (2023). Forecasting volatility by using wavelet transform, ARIMA and GARCH models. *Eurasian Economic Review*, 13(3-4), 803–830. <https://doi.org/10.1007/S40822-023-00243-X>
- Sari, M., Duran, S., Kutlu, H., Guloglu, B., & Atik, Z. (2024). Various optimized machine learning techniques to predict agricultural commodity prices. *Neural Computing and Applications*, 36(19), 11439–11459. <https://doi.org/10.1007/s00521-024-09679-x>
- Shameem, S., P, L. K., B, R. M., & N, V. (2025). A Synergistic Approach to Forecasting Sesamum Prices in Andhra Pradesh: Hybrid, Machine Learning and Wavelet Decomposition Models. *Journal of Scientific Research and Reports*, 31(6), 1033–1043. <https://doi.org/10.9734/JSRR/2025/V31I63195>
- Silva, L. B., Freire, R. Z., & Canciglieri Junior, O. (2021). Spot Energy Price Forecasting Using Wavelet Transform and Extreme Learning Machine, 1–8. <https://doi.org/10.21528/CBIC2021-62>
- Skehin, T., Crane, M., & Bezbradica, M. (2018). Day Ahead Forecasting of FAANG Stocks Using ARIMA, LSTM Networks and Wavelets.
- Tami, M., & Owda, A. Y. (2024). Efficient commodity price forecasting using long short-term memory model. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 14(1), 994–1004. <https://doi.org/10.11591/ijai.v13.i1.pp994-1004>
- Taquiri, J., Lassourd, T., & Viola, A. (2024). *Determining the price of minerals: A transfer pricing framework for lithium* (Licensed CC BY 4.0). Intergovernmental Forum on Mining, Minerals, Metals, Sustainable Development (IGF), Organisation for Economic Co-operation and Development (OECD). International Institute for Sustainable Development (IISD); OECD.
- Tsoku, J. T., Metsileng, D., & Botlhoko, T. (2024). A Hybrid of Box-Jenkins ARIMA Model and Neural Networks for Forecasting South African Crude Oil Prices. *International Journal of Financial Studies* 2024, Vol. 12, Page 118, 12(4), 118. <https://doi.org/10.3390/IJFS12040118>
- U.S. Geological Survey, Mineral Commodity Summaries. (2025). mcs2025.pdf - Mineral Commodity Summaries 2025. (703), 110–111.
- Varshini, A., Kayal, P., & Maiti, M. (2024). How good are different machine and deep learning models in forecasting the future price of metals? Full sample versus subsample. *Resources Policy*, 92(August 2023), 105040. <https://doi.org/10.1016/j.resourpol.2024.105040>
- Vijaya, A., & Swati, B. (2024). Improving Stock Market Trend Forecasting Using ARIMA and Machine Learning. *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Applications, ICAIQSA 2024 - Proceedings*. <https://doi.org/10.1109/ICAQSA64000.2024.10882429>
- Wang, Y., & Guo, Y. (2020). Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Communications*, 17(3), 205–221. <https://doi.org/10.23919/JCC.2020.03.017>
- Yang, Z., Ce, L., & Lian, L. (2017). Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Applied Energy*, 190, 291–305. <https://doi.org/10.1016/J.APENERGY.2016.12.130>

- Zeng, Z., & Khushi, M. (2020). Wavelet Denoising and Attention-based RNN-ARIMA Model to Predict Forex Price.
- Zhang, Y., Ma, B., Lv, Y., Wang, C., & Chen, Y. (2023). An effective method for directly extracting lithium from α -spodumene by activated roasting and sulfuric acid leaching. *Journal of Industrial and Engineering Chemistry*, 122, 540–550. <https://doi.org/10.1016/J.JIEC.2023.03.017>

Appendix A

Appendix

Ethical Approval Letter



11 July 2025

Dear Mr Suthar ,

Reference: CURES/26003/2025

Project ID: 29527

Title: An improved Machine Learning, wavelet–ARIMA approach for forecasting Lithium prices

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/26003/2025 has been reviewed. You may now proceed with the research activities you have sought approval for.

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Appendix B

Literature Review Appendix

McLeod–Li Test

McLeod–Li Test: The McLeod–Li test is a diagnostic tool used to detect autoregressive conditional heteroskedasticity (ARCH) effects in time series data (McLeod & Li, 1983). It evaluates whether the squared residuals from a fitted model, such as ARIMA, exhibit significant autocorrelation, which would indicate that the variance of the series changes over time rather than remaining constant. The procedure applies the Ljung–Box portmanteau test to the squared residuals, testing the null hypothesis that there is no conditional heteroskedasticity. Rejection of this null suggests the presence of time-varying volatility, making models such as ARCH or GARCH more suitable for capturing the dynamics of the data's variance structure. This test is particularly valuable in financial and commodity time series analysis, where volatility clustering is common and variance modelling is essential for accurate forecasting.

Appendix C

Methodology Appendix

Feature Engineering Details

The feature engineering process incorporated the following specific design choices to transform the univariate spodumene price series into a feature-rich dataset:

- Calendar features: year, month, week number, quarter, and day-of-week indicators were extracted to capture seasonal and cyclical patterns.
- Lagged variables: short-term lags (1–4 weeks), medium-term lags (12 and 26 weeks), and a long-term lag (52 weeks) were created to represent immediate momentum, quarterly/half-year cycles, and annual repetition respectively (Foroutan & Lahmiri, 2024).
- Rolling-window statistics: rolling means, standard deviations, and coefficients of variation were computed over 4-, 12-, and 26-week windows to measure short- and medium-term volatility regimes (Byers et al., 2021; Crowley, 2007).
- Exponentially weighted moving averages (EMA) were applied to emphasise recent price movements, reflecting the short-lived shocks often observed in commodity markets (Debnath & Shah, 2015).
- Leakage safeguards: all lagged and rolling variables were shifted forward relative to the prediction horizon to prevent future information from contaminating the training process.
- Feature pruning: a correlation matrix was calculated, and one variable from any pair with a correlation coefficient exceeding 0.995 was iteratively removed. This reduced multicollinearity and ensured more stable parameter estimates in machine learning models (Kriechbaumer et al., 2014).

Technical Details for ARIMA Modelling

Model specification was determined through a stepwise search procedure guided by the AIC, which balances model fit with parsimony. Candidate ARIMA(p,d,q) structures were systematically compared, and the specification with the lowest AIC value was selected using the `auto.arima` package in Python, which automates the search efficiently and

consistently with standard practices in applied time series forecasting (Hyndman & Athanasopoulos, 2014).

Once the optimal model was identified, estimation was carried out using maximum likelihood methods. Model adequacy was then confirmed through residual diagnostics, including histogram and Q–Q plots to check distributional assumptions, autocorrelation function (ACF) analysis to verify independence, and the Ljung–Box test to assess whether the residuals approximated white noise. These checks ensured that the statistical assumptions underpinning ARIMA modelling were satisfied, lending validity to the results.

The ARIMA model combines autoregressive (AR), differencing (I), and moving average (MA) components to capture linear temporal dependencies in time series data (Tsoku et al., 2024).

Technical Details for WAVELET + ARIMA Modelling

The log-transformed spodumene series was decomposed using the Daubechies wavelet of order two (db2), selected for its balance between smoothness and compact support. Although the maximum feasible decomposition depth was seven, a two-level decomposition was adopted as a pragmatic compromise to separate long-term trends from higher-frequency variations, consistent with prior applications in commodity and energy markets (Kriechbaumer et al., 2014; Silva et al., 2021). This produced one approximation (trend) component and two detail components, each capturing fluctuations at different time scales.

Stationarity of the decomposed series was evaluated using the Augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. The approximation (trend) required differencing to achieve stationarity, while both detail components were stationary without differencing. This outcome is consistent with prior evidence that wavelet decomposition generates subseries with more stable variance and fewer structural breaks, simplifying subsequent ARIMA modelling (Ilu & Prasad, 2024; Kumar & Bhattacharyya, 2024).

Separate ARIMA models were then fitted to each component using an automated search procedure guided by the AIC. Forecasts for each component were generated individually and then recombined through inverse wavelet reconstruction to produce the final Wavelet + ARIMA forecast. Residual diagnostics including histogram and Q–Q plots, autocorrelation analysis, and the Ljung–Box test confirmed that the reconstructed residuals approximated white noise, thereby validating the adequacy of the model.

The Wavelet–ARIMA hybrid integrates wavelet decomposition with ARIMA to capture both long-term trends and short-term fluctuations in non-stationary, noisy series (Kriechbaumer et al., 2014; Paul & Garai, 2021).

Machine Learning Models Details

Elastic Net Regression – Elastic Net combines the penalties of Lasso (L1) and Ridge (L2) regression, balancing sparsity and stability. EN was tuned using grid search over alpha (α) and the L1/L2 ratio, with time-series cross-validation ensuring chronological integrity.

Extreme Gradient Boosting (XGBoost) – XGBoost builds an ensemble of decision trees sequentially, with each tree correcting the residuals of the previous iteration. Key hyperparameters tuned included tree depth, learning rate, subsampling ratios, and regularisation terms (λ), α (α), γ (γ). An exponential time-decay weighting scheme was applied to ensure recent data had greater influence.

Long Short-Term Memory (LSTM) Networks – LSTM is a form of recurrent neural network designed to capture long-term dependencies by introducing memory cells with input, output, and forget gates (Skehin et al., 2018). This architecture allows LSTMs to retain information across long sequences while selectively forgetting irrelevant data.

Naive-H Benchmark – The Naive-H model assumes the last observed price persists into the forecast horizon. Where h is the forecast horizon (2 or 12 weeks). Despite its simplicity, this model often performs competitively in volatile markets and provides a robust baseline for comparison (Hyndman & Athanasopoulos, 2014).

Advanced Hybrid Frameworks Details

The hybrid models were implemented in three categories:

- **ARIMA + ML residual hybrids** - Residuals from ARIMA log-return models were used as inputs to Elastic Net, XGBoost, and LSTM. Hyperparameters for each ML model were tuned using time-series cross-validation. For example, Elastic Net used tuned values of $\alpha = 0.056$ and $l1_ratio = 1.0$; XGBoost used tuned tree depth, learning rate (≈ 0.078), subsampling (≈ 0.72), and column sampling (≈ 0.70), with an exponential time-decay weighting scheme; and LSTM was trained on 52-week rolling windows with dropout (0.2), dense layers, and Adam optimiser (learning rate 0.001) with early stopping. Forecasts were recombined with ARIMA predictions to produce final outputs. Diagnostic analyses included actual versus predicted overlays, parity plots, error-over-time series with rolling MAE, residual distributions, and skill comparisons against Naive-H.
- **Wavelet + ML hybrids** - The log-transformed series was decomposed using a Daubechies (db2) wavelet at level two, producing one trend and two detail components. H -step deltas were constructed for each component and used to train Elastic Net, XGBoost, and LSTM separately. Forecasts were then recombined to reconstruct the series. Comparative diagnostics included actual versus predicted plots, parity plots, rolling error analysis, and residual autocorrelation checks, benchmarked against a Wavelet-ARIMA baseline.
- **Integrated Wavelet + ARIMA + ML hybrids** - Decomposed components were routed according to their statistical properties. Linear (trend) signals were modelled with ARIMA, while nonlinear (detail) signals were assigned to Elastic Net, XGBoost, or LSTM. Forecasts were recombined via inverse wavelet transform. This integrated approach exploited the strengths of each method, aligning model choice with frequency characteristics, and has been shown in prior studies to improve robustness and predictive accuracy.

These implementation details ensured that each hybrid class was rigorously tested, with diagnostics confirming that observed improvements stemmed from meaningful non-linear corrections and multi-resolution decomposition rather than overfitting.

Forecasting Metrics Details

- **Mean Absolute Error (MAE) -**

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where y_i is the actual value at time i , $\hat{y}_i$ is the predicted value at time i , and n is the number of observations.

- **Root Mean Squared Errors (RMSE) -**

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

where y_i is the actual value at time i , $\hat{y}_i$ is the predicted value at time i , and n is the number of observations.

- **Mean Absolute Percentage Error (MAPE) -**

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

where y_i is the actual value at time i , $\hat{y}_i$ is the predicted value at time i , and n is the number of observations. (Note: MAPE can be distorted when $y_i \approx 0$.)

- **Coefficient of Determination (R^2) –**

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where y_i is the actual value at time i , $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values, and n is the number of observations.

Appendix D

Results Appendix

Outliers and Summary Table

Dataset Description -

Statistic	Price Filled
Count	394
Mean	1302.7289
Std	1379.3241
Min	292.44
25%	473.99
50%	633.025
75%	1725.7675
Max	5382.51

Table D.1: Summary statistics of Spodumene prices.

Outliers Count -

Date	last_price	price filled
12/06/2022	3655.27	3655.27
19/06/2022	3691.85	3691.85
26/06/2022	3667.78	3667.78
03/07/2022	3735.68	3735.68
10/07/2022	3843.91	3843.91
17/07/2022	3981.44	3981.44
24/07/2022	3926.14	3926.14
31/07/2022	3878.71	3878.71
07/08/2022	3907.93	3907.93
14/08/2022	3891.50	3891.50
21/08/2022	3991.54	3991.54
28/08/2022	4202.82	4202.82
<i>Continued on next page</i>		

Date	last_price	price_filled
04/09/2022	4347.45	4347.45
11/09/2022	4490.99	4490.99
18/09/2022	4659.12	4659.12
25/09/2022	4873.35	4873.35
02/10/2022	5054.25	5054.25
09/10/2022	nan	5053.78
16/10/2022	5053.31	5053.31
23/10/2022	5150.98	5150.98
30/10/2022	5177.32	5177.32
06/11/2022	5382.51	5382.51
13/11/2022	5178.40	5178.40
20/11/2022	5143.53	5143.53
27/11/2022	5051.67	5051.67
04/12/2022	4976.79	4976.79
11/12/2022	4972.74	4972.74
18/12/2022	5000	5000
25/12/2022	4986.31	4986.31
01/01/2023	4967.76	4967.76
08/01/2023	4956.56	4956.56
15/01/2023	4889.84	4889.84
22/01/2023	4835.57	4835.57
29/01/2023	nan	4889.835
05/02/2023	4944.10	4944.10
12/02/2023	4954.77	4954.77
19/02/2023	4933.44	4933.44
26/02/2023	4855.18	4855.18
05/03/2023	4646.13	4646.13
12/03/2023	4237.15	4237.15
19/03/2023	4136.23	4136.23
26/03/2023	3949.63	3949.63
02/04/2023	3733.40	3733.40
09/04/2023	3712.35	3712.35
16/04/2023	3700.12	3700.12

Table D.2: Weekly spodumene prices (June 2022–April 2023).

Benchmark Models Further Explained

ARIMA Residuals - Figure D.1 shows residuals are mostly centred around zero, indicating the ARIMA model captures most of the structure, with a single large outlier early in the series.

Figure D.2 shows residuals are highly concentrated near zero, suggesting low overall forecast error, though skewed by a single large outlier. Q–Q Plot of Residuals Points deviate from the diagonal, especially at extremes, showing heavy-tailed residuals and non-normality caused by the outlier.

Figure D.3 ACF plot for residuals show no significant autocorrelation in residuals, indicating the ARIMA model has effectively removed temporal dependence. PACF of

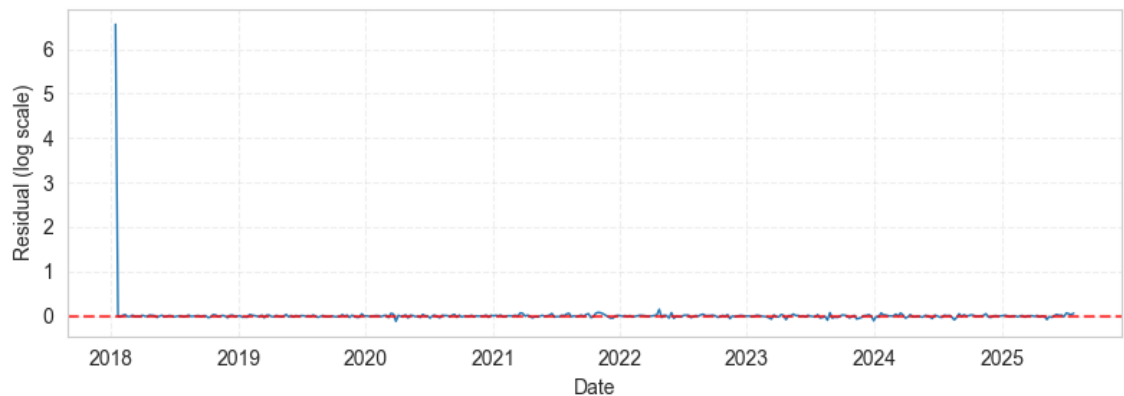


Figure D.1: Residuals Over Time ARIMA

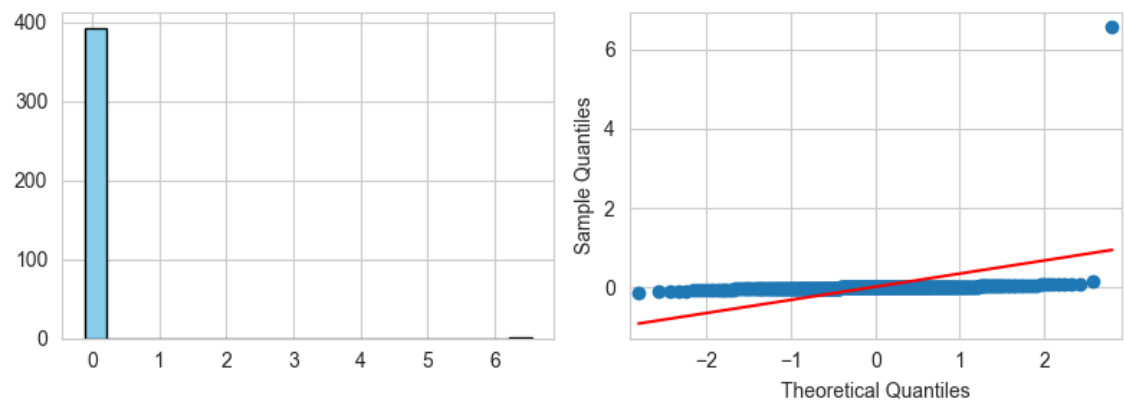


Figure D.2: Histogram and Q-Q plot of residuals

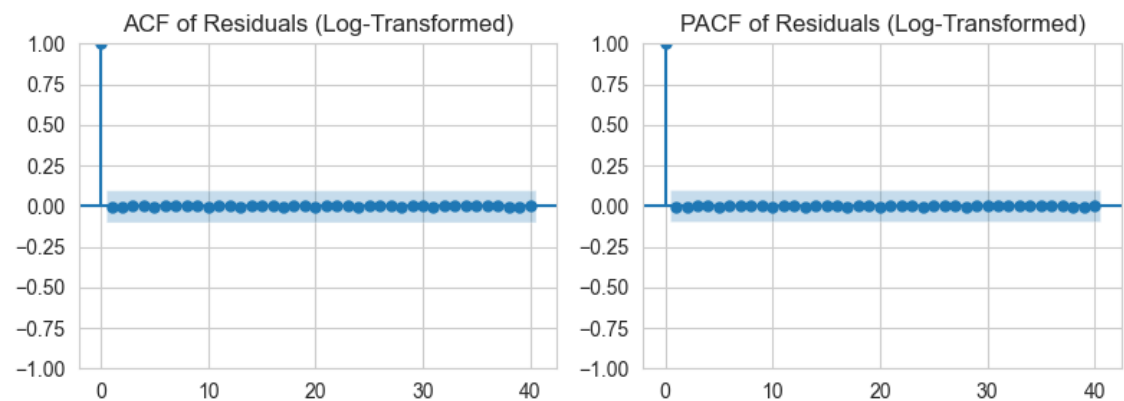


Figure D.3: ACF and PACF residual plots

Residuals confirms lack of partial autocorrelation, further validating that the residuals approximate white noise.

Elastic Net (Tuned) - Time-series cross-validation over alpha (log-space) and L1 ratio; scored on MAE; refit on best. Best parameters: $\alpha \approx 0.0178$, L1 ratio = 0.5.

Results (price scale):

- Validation: MAE = 7.93, RMSE = 11.04, MAPE = 1.27%.
- Test: MAE = 37.44, RMSE = 45.40, MAPE = 7.24%.

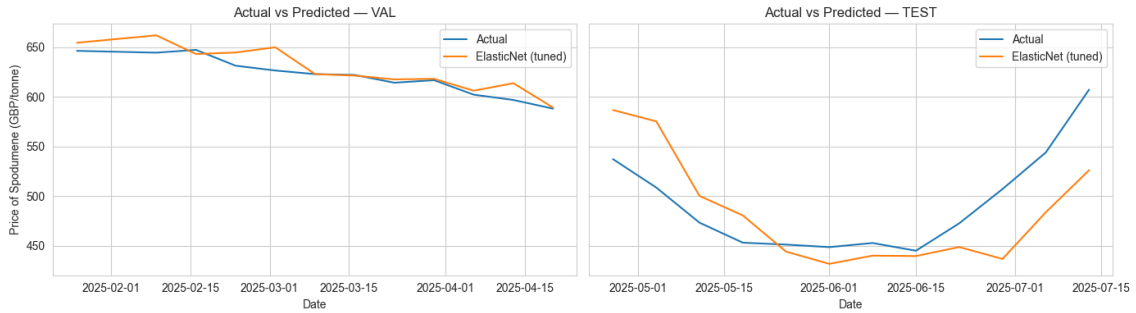


Figure D.4: Actual vs Predicted for Elastic Net on validation and test

Context vs Naive-H (Validation: 10.55, 12.03, 1.71%; Test: 45.32, 55.48, 8.80%): Elastic Net improves on all three metrics in both splits, though test-set gains are modest. Diagnostics show close validation tracking and mild lag at turning points on test, with rising dispersion late in the horizon. Interpretation: useful linear signal beyond persistence, but sensitivity to structural change remains.

LSTM (baseline) - Leakage-safe windows (lookback = 52); architecture LSTM(64) → Dropout(0.2) → Dense(32) → Dense(1); Adam with early stopping.

Results (price scale):

- Validation: MAE = 8.06, RMSE = 10.30, MAPE = 1.30% (beats Naive-H).
- Test: MAE = 76.65, RMSE = 86.46, MAPE = 15.48% (worse than Naive-H).

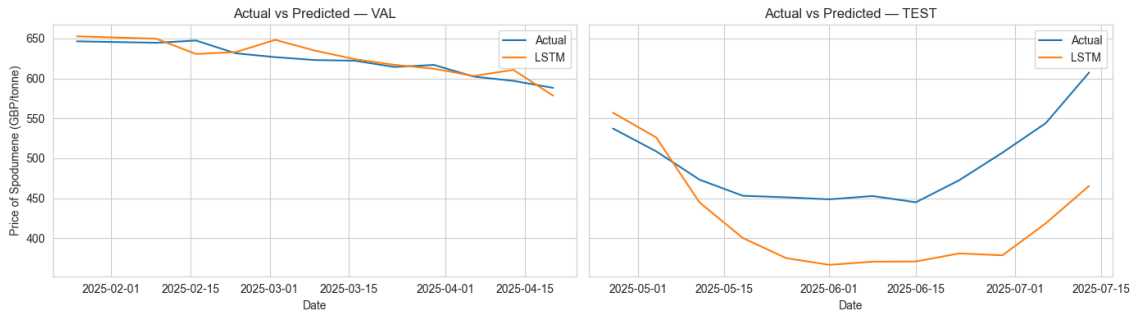


Figure D.5: Actual vs Predicted for LSTM on validation and test

Diagnostics: close validation fit; on test the network misses the trough-to-rebound and remains below the realised path for several weeks, producing an upward-trending

error curve. Interpretation: high variance and overfitting under limited data without a full hyperparameter sweep. Hyperparameter tuning was not done for LSTM as it resulted in worse results as the dataset was only 394 rows and deep learning models like LSTM require more data to perform well as seen in the literature review.

Comparative remarks (vs Naive-H and ARIMA) -

- Validation: XGBoost achieves the lowest errors by a wide margin; Elastic Net and LSTM also beat Naive-H.
- Test: ARIMA is steady with competitive MAPE (2.21%). XGBoost maintains an edge over Naive-H (MAE 35.31 vs 45.32; MAPE 6.86% vs 8.80%). Elastic Net offers small gains (MAE 37.44; MAPE 7.24%). LSTM deteriorates sharply (MAE 76.65; MAPE 15.48%).

XGBoost Tuning Details - Search strategy: two-stage procedure; randomised sampling over depth/grow policy, leaves, learning rate, regularisation and subsampling; pruning on development MAE with time-decay weighting (half-life $\in \{39, 52, 78, 104\}$ weeks); final selection by validation MAE and DEV $\rightarrow$ TEST refit.

Best single model (by validation): grow_policy = “lossguide”, max_leaves = 255, learning_rate ≈ 0.083 , min_child_weight ≈ 4.19 , subsample ≈ 0.81 , colsample_bytree ≈ 0.63 , $L_2 \approx 0.55$, $L_1 \approx 0.0014$, $\gamma \approx 2.23 \times 10^{-4}$; half-life = 39; recent training window = 104 weeks; best_iter ≈ 8 ; search time $\approx 1,582$ s.

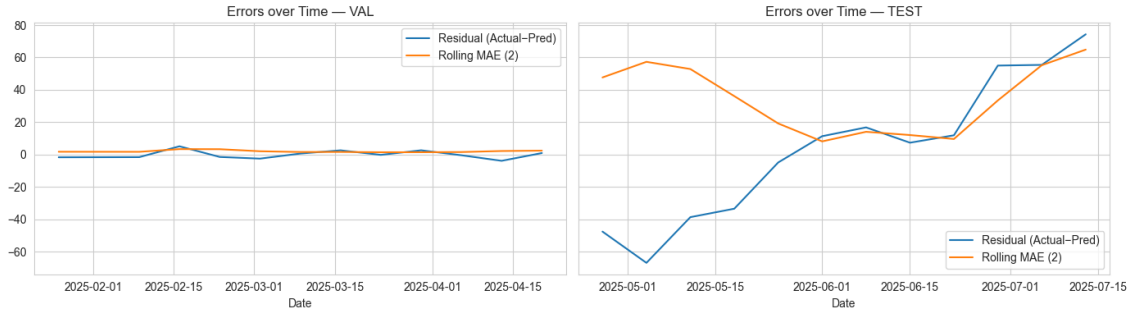


Figure D.6: Errors over time for XGBoost on validation and test.

Error Over Time (VAL & TEST): Errors are stable and near zero in the validation period (see Figure D.6), but in the test set they increase sharply over time, with rising residuals and MAE, highlighting model performance degradation on unseen data.

Residual Integrated Models (ARIMA + Machine Learning)

ARIMA + Elastic Net Residuals - Elastic Net uses all numeric features with Standard-Scaler and the tuned hyper-parameters inherited from the benchmark search. Two models are fit to respect chronology: TRAIN $\rightarrow$ VAL and DEV (TRAIN+VAL) $\rightarrow$ TEST.

The hybrid achieves MAE 7.87, RMSE 10.21, MAPE 1.26%. This is a small but consistent improvement over ARIMA-only (8.10/10.44/1.30%) and clearly better than Naive-H (10.55/12.03/1.71%). In relative terms, the EN residuals reduce ARIMA’s MAE by 3% and Naive-H’s by 25% (Figure D.7). TEST. Out-of-sample performance is essentially a tie with the ARIMA baseline: MAE 37.34, RMSE 42.86, MAPE 7.34% for

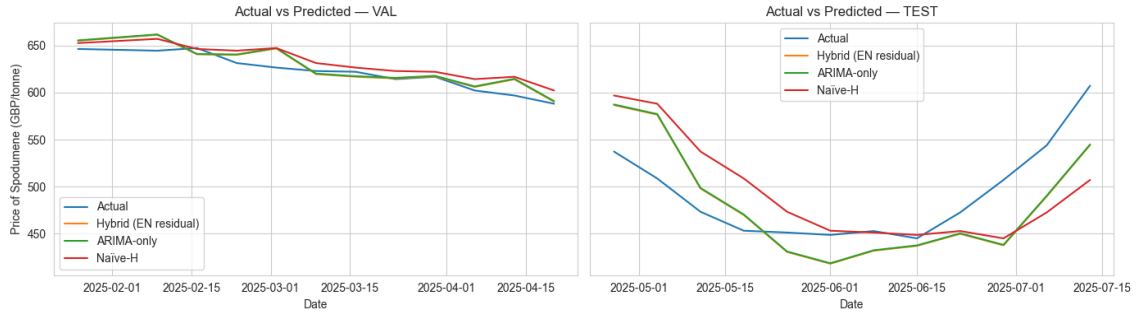


Figure D.7: Actual vs Predicted for ARIMA + Elastic Net on validation and test.

the hybrid versus 37.21/42.76/7.31% for ARIMA-only. Both models are materially better than Naive-H (45.32/55.48/8.80%).

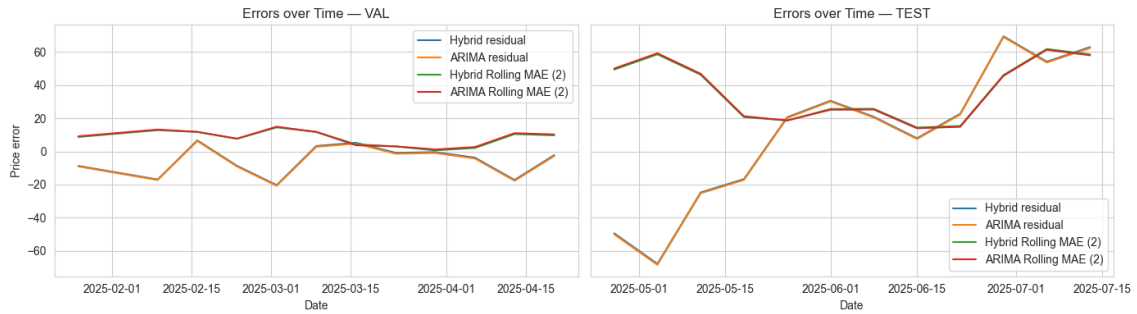


Figure D.8: Errors over time for ARIMA + Elastic Net on validation and test.

Error-over-time plots (Figure D.8) show that EN corrects part of ARIMA’s early May under-prediction but drifts similarly from mid June, suggesting the linear residual has limited ability to capture the regime change in late Q2. Elastic Net yields a light-weight correction that reliably improves or matches ARIMA and beats Naive-H, but its gains compress on the truly out-of-sample segment.

ARIMA + LSTM Residuals - We build leakage-safe, 52-week windows on the scaled feature matrix and train a single-layer LSTM (64 units, dropout 0.2, Adam, early stopping) to predict residuals at horizon (H). As with the other hybrids, separate TRAIN→VAL and DEV→TEST models are used.

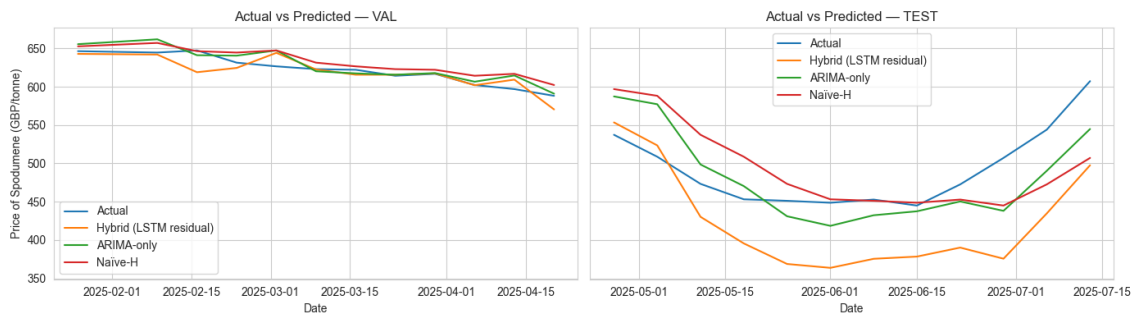


Figure D.9: Actual vs Predicted for ARIMA + LSTM on validation and test.

Validation - The LSTM residual yields MAE 8.18, RMSE 11.89, MAPE 1.32%—com-

parable to ARIMA-only (slightly worse on MAE/RMSE) and inferior to both XGB and EN. Residual ACF/PACF and the rolling-MAE curve suggest the model captures some nonlinearity around mid-March but introduces additional variance elsewhere (Figure D.9). Test - Performance degrades sharply out-of-sample: MAE 72.94, RMSE 80.65, MAPE 14.86%.

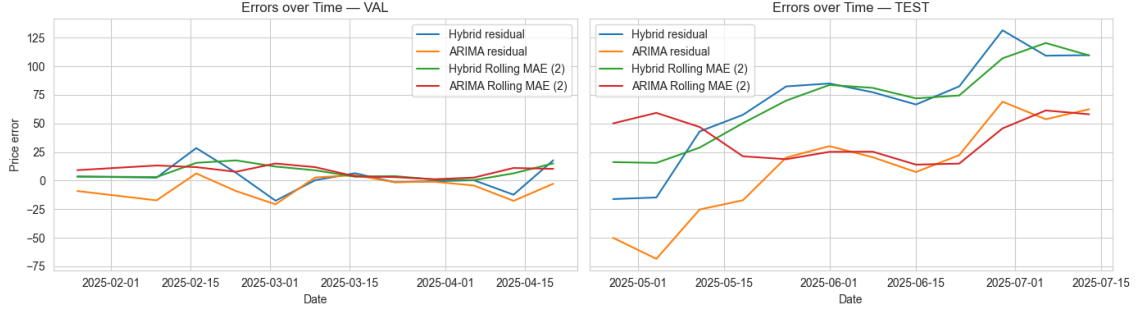


Figure D.10: Errors over time for ARIMA + LSTM on validation and test.

Errors escalate after the mid-May trough and remain elevated through July. In the validation set, both ARIMA and the ARIMA+LSTM hybrid maintain low, stable errors, while in the test set the hybrid initially improves over ARIMA but error growth accelerates over time, indicating reduced robustness on unseen data (Figure D.10).

In this setting, a simple recurrent residual learner is not competitive out of sample both ARIMA only and Naive-H are preferable, and tree/linear residuals are markedly superior.

ARIMA + XGB provides the best out-of-sample accuracy; ARIMA+EN offers a stable, low-complexity alternative; ARIMA + LSTM underperforms under the current data regime and horizon. MAE, RMSE and MAPE remain the primary decision metrics. R^2 values are mixed and retained for completeness only.

Cross Model Comparison and Practical Implications

Validation - All residual hybrids achieve single-digit MAE and about 1–1.3% MAPE. XGB is best (7.34; 1.20%), EN close (7.87; 1.26%), LSTM trails (8.18; 1.32%). All beat Naive-H; EN/XGB improve over ARIMA-only. Test - XGB retains an advantage with the lowest MAE (35.79) and MAPE (6.92%). EN ties ARIMA-only (37.34 vs 37.21 MAE; 7.34% vs 7.31% MAPE) and is clearly ahead of Naive-H. LSTM deteriorates sharply. Error dynamics - Where ARIMA under-reacts to rapid swings, XGB and to a lesser extent EN correct part of that bias, especially near early turning points. Some short-lag correlation persists on TEST across all models.

WAVELET Based Models

At each forecast origin t we form H -step component deltas,

$$\Delta c = c(t+H) - c(t),$$

and the sum of these deltas equals the H -step log return of the original series.

Definitions:

- t : the forecast origin (time index), i.e., the last date known when the forecast is issued.

- H : the forecast horizon in weeks; in this thesis $H \in \{2, 12\}$.
- c : a wavelet component of the log price series (e.g., *Trend*, *detail_1*, *detail_2*).
- $c(t)$: the value of component c at time t (log scale).
- $c(t + H)$: the value of component c H weeks after t (log scale).
- Δc : the H -step change in component c between t and $t + H$ (log units).

WAVELET + Elastic Net - To exploit the multi-resolution structure while allowing learned cross-feature effects, we trained Elastic Net regressors to predict H -step component deltas (in log space) for each wavelet band. Predicted deltas were then summed across bands to obtain the overall H -step log return, which we converted back to prices for evaluation. Elastic Net hyperparameters were taken from the earlier time-series CV search (Section 1.3), and the feature set matched the benchmark ML pipelines. Naive- H was computed in the same space and compared on the price scale.

Results show that Wavelet + EN consistently outperformed Naive- H and dramatically beat the pure Wavelet + ARIMA:

- Validation: MAE = 8.46, RMSE = 11.54, MAPE = 1.35% (vs Naive- H : MAE = 10.55, RMSE = 12.03, MAPE = 1.71%; vs Wavelet-ARIMA: MAE = 64.71, MAPE = 10.51%).
- Test: MAE = 37.59, RMSE = 44.69, MAPE = 7.31% (vs Naive- H : MAE = 45.32, RMSE = 55.48, MAPE = 8.80%; vs Wavelet-ARIMA: MAE = 149.80, MAPE = 31.28%).

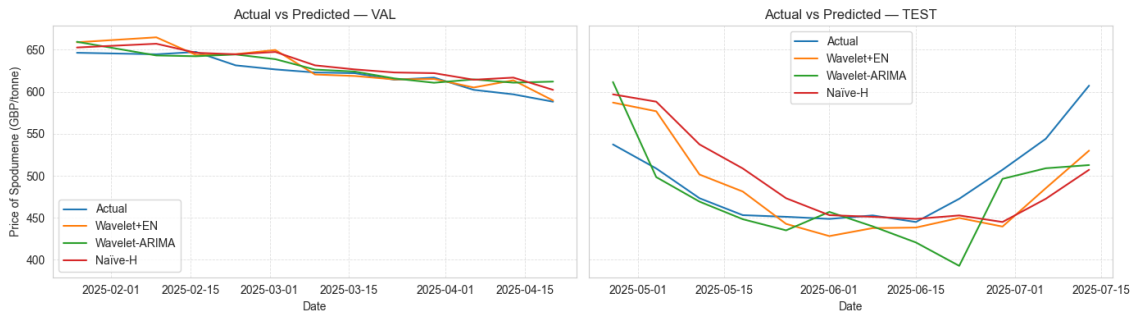


Figure D.11: Actual vs Predicted for WAVELET + EN on validation and test.

The actual vs predicted (Figure D.11), and parity panels show close VAL alignment and materially reduced spread on TEST relative to Wavelet + ARIMA. Error-over-time curves (Figure D.12) indicate smaller and more stable residuals through the trough and early rebound, though a mild positive lag-1 autocorrelation remains in the hybrid residual ACF on TEST, implying residual short-term signal that a richer feature set or non-linear learner might capture.

Interpretation within results: Learning scale-specific deltas rather than levels, and summing them at forecast time, stabilises predictions while preserving high-frequency

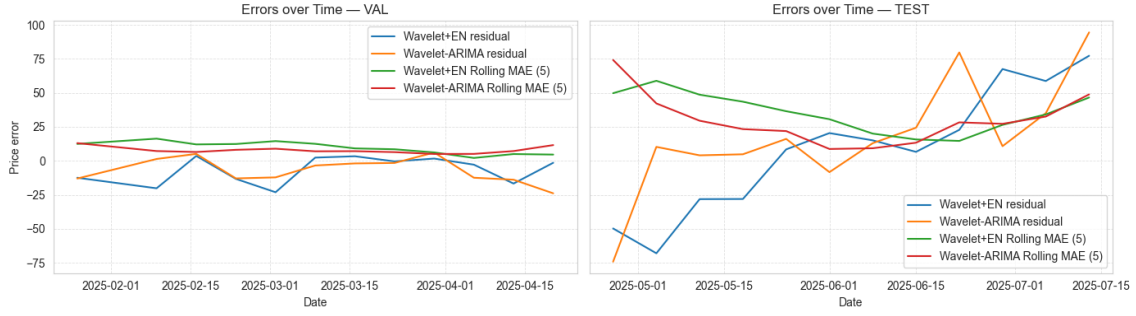


Figure D.12: Errors over time for WAVELET + EN on validation and test.

movements. The gains over Naive-H on both splits confirm that the wavelet features carry useful horizon-H information that linear regularisation can harness.

WAVELET + LSTM - Finally, we trained leakage-safe LSTMs on component deltas (lookback = 52; LSTM(64) → Dropout(0.2) → Dense(32) → 1; Adam 10^{-3} ; early stopping) after hyperparameter tuning, and we summed the predicted deltas.

- **Validation:** MAE = 15.80, RMSE = 19.84, MAPE = 2.58% (worse than Naive-H's 1.71%).
- **Test:** MAE = 42.27, RMSE = 51.52, MAPE = 8.17% (slightly better than Naive-H's 8.80%, and far better than Wavelet-ARIMA).

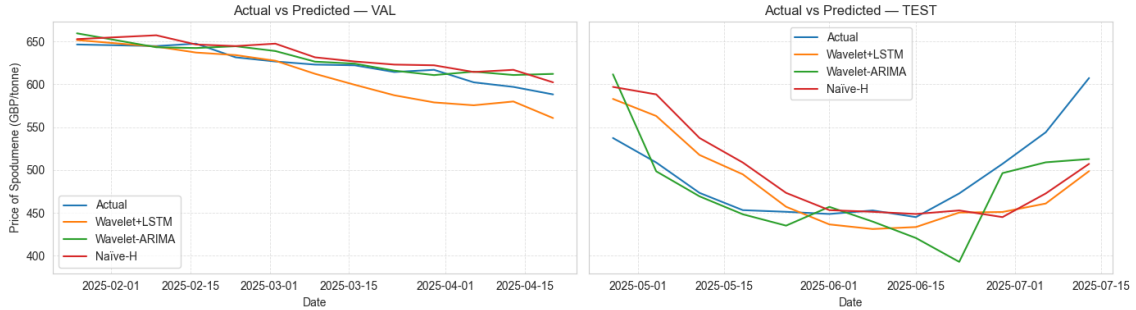


Figure D.13: Actual vs Predicted for WAVELET + LSTM on validation and test.

The VAL shortfall and only modest TEST gain indicate sensitivity to limited data and hyperparameters. Plots show that Wavelet + LSTM (Figure D.13) often lags the rebound and exhibits smoother, under-reactive trajectories; residual ACFs are small but not fully white.

Machine Learning + ARIMA + WAVELET Based Models

Elastic Net + ARIMA + WAVELET -

Scores (Price Scale):

- Validation: MAE = 25.93, RMSE = 26.62, MAPE = 4.19%, $R^2 = -1.07$.
- Test: MAE = 48.54, RMSE = 54.46, MAPE = 9.68%, $R^2 = -0.27$.

Baseline for comparison (WAVELET + ARIMA):

- Validation: MAE = 8.99, RMSE = 11.14, MAPE = 1.46%, $R^2 = 0.64$.
- Test: MAE = 31.28, RMSE = 44.15, MAPE = 6.05%, $R^2 = 0.17$.

Interpretation: The Elastic Net hybrid underperformed the Wavelet–ARIMA baseline on both validation and test sets. Despite using lagged predictors [1, 2, 3, 4, 12, 26] and standardised features, EN did not adapt to turning points or capture curvature, reflecting its linear form.

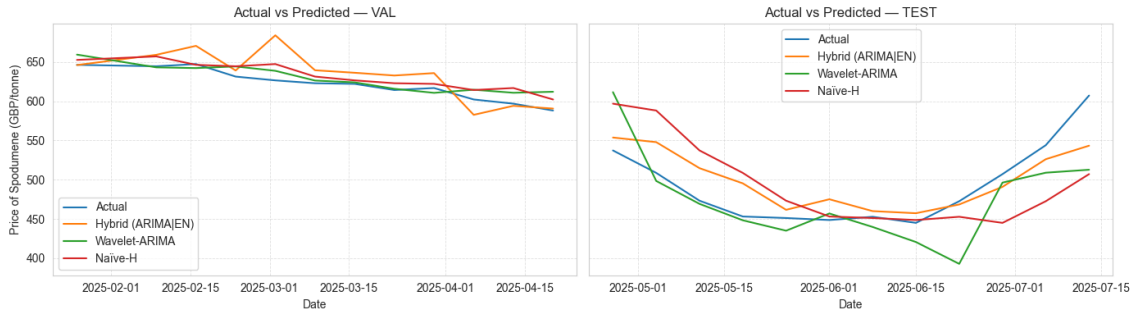


Figure D.14: Actual vs Predicted for WAVELET + ARIMA + EN on validation and test.

Residual traces revealed persistent bias. This confirms that a sparse linear model is insufficient when all components behave non-linearly. In the validation period (Figure D.14), the ARIMA+Elastic Net hybrid and Wavelet–ARIMA models closely follow actual prices with low, stable residuals, while in the test set errors grow, showing both models struggle to track sharp trend shifts. The hybrid generally outperforms Wavelet–ARIMA, with lower residuals and rolling MAE in most periods.

LSTM + ARIMA + WAVELET - This variant uses Wavelet–ARIMA as the baseline forecaster of the 12-week log return and trains an LSTM to model residuals. Input sequences consist of 26-week windows of standardised features, with optional time-decay weighting and residual clipping.

Scores (Price Scale):

- Validation: MAE = 8.08, RMSE = 9.30, MAPE = 1.30%, $R^2 = 0.75$.
- Test: MAE = 62.95, RMSE = 77.63, MAPE = 12.44%, $R^2 = -1.58$.

The residual LSTM produced strong validation results but deteriorated sharply out-of-sample. Test forecasts overshoot and drift upward, with residual autocorrelations indicating memorisation of short-term oscillations from TRAIN/VAL that did not recur in the test period. This reflects classic overfitting of sequence models under limited data. In the validation period, the Wavelet–ARIMA+LSTM hybrid tracks prices closely with minimal residuals, like other models. However, in the test period, prediction errors grow significantly, with higher volatility and residual spikes, indicating weaker generalisation compared to Wavelet–ARIMA and Naive-H.

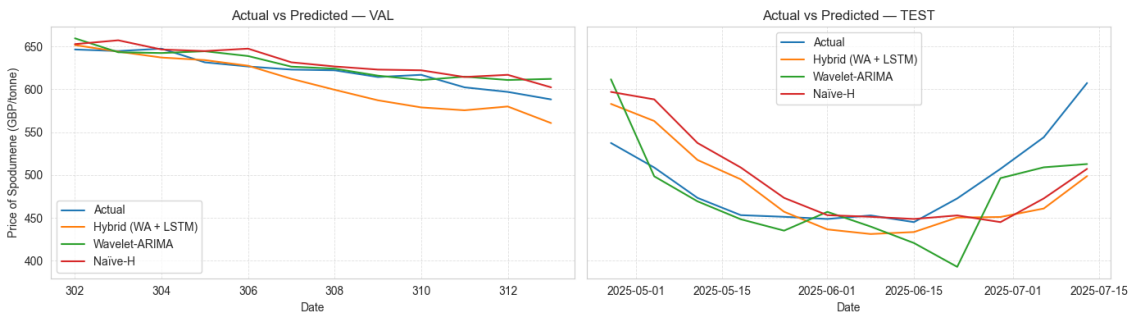


Figure D.15: Actual vs Predicted for WAVELET + ARIMA + LSTM on validation and test.

Appendix E

Discussion Appendix

Wavelet Specification and Sensitivity

We document tested mothers Daubechies and Symmlet orders and levels one to four with reconstruction checks and energy preservation. Trends were consistent with reports that lower levels suit short horizons and that MODWT often stabilises forecasts for volatile resources (Crowley, 2007; Kriechbaumer et al., 2014)

Strategy Variant

Residual learning was implemented as baseline ARIMA plus machine learning corrector and component wise models followed the decompose model recombine workflow used in crude oil and electricity studies (Md-Khair & Samsudin, 2018; Zhang et al., 2023).

Diagnostic Thresholds and Cross Validation

We report unit root and stationarity test statistics, residual Ljung Box p values, and rolling window error bands. These follow the diagnostics and rolling origin procedures set out in Forecasting principles and practice and in wavelet ARIMA validation examples (Hyndman & Athanasopoulos, 2014; Md-Khair & Samsudin, 2018).