How Simple Can Multi-Horizon Forecasting Be?

Abstract

Multi-horizon time-series forecasting requires models to capture heterogeneous temporal dynamics that influence predictions differently across short-, medium-, and long-term horizons. Existing forecasting models often process these dynamics using a uniform modelling strategy or increasingly complex transformer and attention architectures designed to capture dependencies across different temporal scales. However, it remains unclear whether such architectural complexity is necessary to maintain accurate predictions as the forecast horizon fluctuates. To investigates this simple question: how simple can a forecasting model be while remaining effective across short-, medium-, and long-term horizons? We introduce Decomposition–Linear–MLP–Linear (D-LML), a lightweight architecture for direct multi-horizon forecasting for time series data. D-LML decomposes each historical input window into trend, seasonal, and residual components and processes them using two linear mappings and a compact shallow multilayer perceptron before additively reconstructing the future sequence. We evaluate D-LML across multiple datasets and progressively increasing prediction horizons against established forecasting model paradigms. Our experiments showed that D-LML achieves competitive state-of-the-art performances in short, medium and long term across several commonly used datasets, with the superior forecasting accuracy while requiring substantially fewer trainable parameters than high-capacity neural baselines. These findings suggest that carefully structured lightweight models provide a strong alternative for accurate and computationally efficient multi-horizon time-series forecasting.

Keywords: multi-horizon, time series decomposition, linear, MLP, forecasting.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 Introduction

Time Series Forecasting (TSF) for different horizon is an important requirement across many fields. Multi-horizon forecasting allows a model to predict several future time steps simultaneously, offering significant benefits over traditional single-step models that predict only one time point at a time. This approach improves the accuracy of forecasting by capturing the underlying temporal dependencies over multiple horizons, thus reducing errors, especially when predicting peak loads, which are often challenging due to their volatility(Liu & Wang, 2024). Multi-horizon forecasting becomes increasingly challenging as the prediction horizon grows. Short-term forecasts can often exploit strong local dependencies between recent observations and the immediate future. As the forecast horizon increases, however, these local relationships may weaken, while broader temporal structures including trends, periodic behaviour, and longer-term dynamics become increasingly important. Consequently, a model that performs well for short-term prediction may not necessarily maintain the same performance at medium or long horizons.

A multi-horizon prediction model for time series data refers to a forecasting framework where the goal is to predict multiple future time steps ahead (e.g., 1-step, 3-step, 6-step, 12-step forecasts) simultaneously, rather than only the next step. This setting, commonly referred to as multi-horizon forecasting, requires a model to estimate $\hat{\mathbf{Y}}_{t+1:t+H} = [\hat{y}_{t+1}, \dots, \hat{y}_{t+H}]$ from a historical context $\mathbf{X}_{t-L+1:t}$, where L and H denote the historical lookback and prediction horizon, respectively. As H increases, the forecasting problem becomes increasingly challenging because predictive information can shift from recent local dependencies toward broader trends, periodic patterns, and persistent temporal structure. Multi-horizon forecasting is important across a wide range of real-world domains, where decisions often depend on predictions extending over multiple future time steps rather than a single forecast point. Examples include weather prediction across increasing meteorological lead times(Price et al., 2025), electricity and renewable-energy forecasting over operational and planning horizons(Ni et al., 2024; Perera et al., 2022), and financial forecasting of future volatility and asset prices (Beniwal et al., 2024) , traffic flow prediction(Yang et al., 2019), and air pollution estimation (Zhou et al., 2022).This is crucial in practical applications, where decision-making depends on forecasts across different time horizons.

To address these challenges, recent forecasting research has introduced increasingly sophisticated and complicated models based on recurrent networks, attention mechanisms, Transformers, multiscale representations, and deep nonlinear forecasting blocks. Models such as DeepAR, the Temporal Fusion Transformer (TFT), N-BEATS, and TimeMixer demonstrate the potential of expressive architectures for modelling complex temporal dependencies. Nevertheless, more complex model sophistication is generally accompanied by increased parameterization, computational requirements, and implementation complexity. In the literature, it has been observed that models excel at short-term forecasts but degrade in performance for medium and long-term horizons. Recursive strategies suffer from error propagation. Direct approaches are computationally costly and ignore horizon dependencies. In the past years, transformer and attention-based models have been proposed for multi-horizon prediction like TFT, which is complicated and needs heavy programming.

In this paper, we have proposed a simple, less complicated and lightweight model which is composition of linearity and MLP for prediction of multi-horizon values that remain robust

across short-, medium-, and long-term horizons, The aim is to provide a simple multi-horizon prediction model which is stable and efficiently forecasts across all horizons while balancing simplicity, interpretability, and computational efficiency

Our main contributions are summarized as follows:

- We introduce D-LML, a lightweight architecture for direct multi-horizon forecasting that predicts short-, medium-, and long-range future values without recurrent networks, self-attention, Transformers, convolutional blocks, or deeply stacked forecasting modules.
- We develop a unified forecasting framework that can operate on both univariate and multivariate time-series data, allowing the same simple architecture to be evaluated across different data dimensionalities and forecast horizons.
- We systematically investigate whether architectural complexity must increase with forecasting difficulty and horizon length by comparing D-LML with representative linear, MLP-based, recurrent, multiscale, and attention-based forecasting approaches under a common evaluation framework.
- We evaluate forecasting performance jointly with model complexity and parameter efficiency, investigating whether a compact architecture can provide a competitive accuracy–complexity trade-off across short-, medium-, and long-term prediction horizons.

The novelty of this work lies in the development of a simple and computationally lightweight architecture for direct multi-horizon time-series forecasting that can be applied consistently to multivariate data. Overall, this work revisits a basic assumption underlying modern time-series forecasting: that increasingly difficult forecasting horizons should be addressed with increasingly sophisticated models. Rather than asking how much additional complexity can be incorporated into a forecasting architecture, we investigate the complementary question of how much complexity is necessary. Our results suggest that carefully structured lightweight models can provide a strong alternative for multi-horizon forecasting and motivate further investigation of simplicity as an explicit design objective in time-series modelling.

2 Related Work

Recent progress in multi-horizon time-series forecasting has been accompanied by increasingly sophisticated neural architectures. Given the broad scope of the forecasting literature, we focus specifically on methods explicitly designed or evaluated for multi-horizon forecasting, where a model predicts multiple future time steps over different forecast horizons. We distinguish this setting from the broader literature on long-term forecasting, in which models may target distant future values without explicitly investigating forecasting behaviour across multiple horizons. Although numerous methods have been proposed for long-term time-series forecasting, including all such approaches would extend beyond the central research question of this study. We therefore restrict the related-work discussion and empirical comparison to models that explicitly address multi-horizon prediction or report performance across multiple forecasting horizons. This focused selection enables a more direct comparison with D-LML and, more importantly, allows us to investigate how forecasting performance changes as the prediction horizon increases.

Early neural forecasting methods frequently relied on recurrent architectures and autoregressive generation. DeepAR, for example, employs an autoregressive recurrent neural network to model the conditional distribution of future observations and has demonstrated strong probabilistic forecasting performance across collections of related time series (Salinas et al., 2020). However, autoregressive generation is sequential and may propagate prediction errors or uncertainty across increasingly distant horizons.The Temporal Fusion Transformer (TFT) was developed specifically for multi-horizon forecasting and combines recurrent processing, variable selection, gating mechanisms, and interpretable self-attention (Lim et al., 2021). By integrating historical observations, static information, and known future covariates, TFT provides a flexible framework for predicting multiple future time steps. Nevertheless, such flexibility is accompanied by considerable architectural complexity.A substantial body of recent work has focused on increasing model capacity to capture complex dependencies over long forecasting horizons. N-BEATS introduced a deep fully connected architecture consisting of stacked residual blocks that jointly generate backcast and forecast representations (Oreshkin et al., 2019). The model demonstrated that direct multi-step forecasting can be performed effectively without recurrent processing, while its interpretable configuration introduced trend- and seasonality-oriented basis functions. TimeMixer provides another important perspective(Wang et al., 2024). Rather than employing self-attention, TimeMixer uses an MLP-based architecture together with multiscale decomposition and mixing. TimeMixer demonstrates that strong short- and long-term forecasting performance can be obtained without Transformer-based attention. Nevertheless, its multiscale mixing architecture remains substantially more elaborate than a small set of direct linear and shallow nonlinear mappings.

Existing models address this problem in different ways. TFT was explicitly developed for multi-horizon prediction (Lim et al., 2021), while N-BEATS provides direct forecasts of the complete future horizon(Oreshkin et al., 2019) TimesNet evaluates temporal representations across short- and long-term forecasting tasks(Wu et al., 2022), while TimeMixer explicitly investigates forecasting at different temporal scales and across short- and long-term settings (Wang et al., 2024). Collectively, these approaches demonstrate substantial progress in multi-horizon forecasting. However, they also reflect a general tendency toward increasingly sophisticated representation mechanisms, including attention, deep residual structures, frequency-domain transformations, patching, convolutional processing, and multiscale feature extraction. This motivates an important complementary question: how much architectural complexity is necessary to obtain competitive forecasts across different prediction horizons?

Motivated by this question, we propose D-LML (Decomposition–Linear–MLP–Linear) as a deliberately lightweight architecture for direct multi-horizon forecasting. D-LML decomposes each historical input window into trend, seasonal, and residual components and processes them using two linear mappings and a compact shallow MLP. The resulting component forecasts are recombined to directly produce the complete future horizon. The primary objective of D-LML is therefore not to introduce additional representational complexity, but to examine whether a simple decomposition-based architecture can provide a favourable balance between forecasting accuracy, horizon robustness, model complexity, and applicability to multivariate data.

3 Decomposition–Linear–MLP–Linear (D-LML):

Given a time series X containing one or multiple observed variables, the forecasting task aims to utilize a historical sequence of length L to predict the subsequent H future observations. A central challenge in multi-horizon forecasting is to maintain predictive performance as the forecasting distance increases without unnecessarily increasing architectural complexity. To

address this question, we propose D-LML (Decomposition–Linear–MLP–Linear), a lightweight direct multi-horizon forecasting architecture that exploits decomposed temporal structure through simple component-wise predictors. Specifically, D-LML decomposes the historical input into trend, seasonal, and residual components and employs a Linear–MLP–Linear configuration for their respective forecasts. The resulting component forecasts are additively reconstructed to directly produce the complete H-step future sequence. This simple architecture is retained across increasing prediction horizons, enabling a controlled investigation of the relationship between forecast horizon, predictive accuracy, and model complexity. For each historical input window, the model extracts trend, seasonal, and residual components and models them using a Linear–MLP–Linear configuration:

$$\mathbf{X} \rightarrow (\mathbf{T}, \mathbf{S}, \mathbf{R}) \rightarrow \left(f_L(\mathbf{T}), f_{\mathrm{MLP}}(\mathbf{S}), f_L(\mathbf{R})\right) \rightarrow \hat{\mathbf{Y}}_{t+1:t+H}.$$

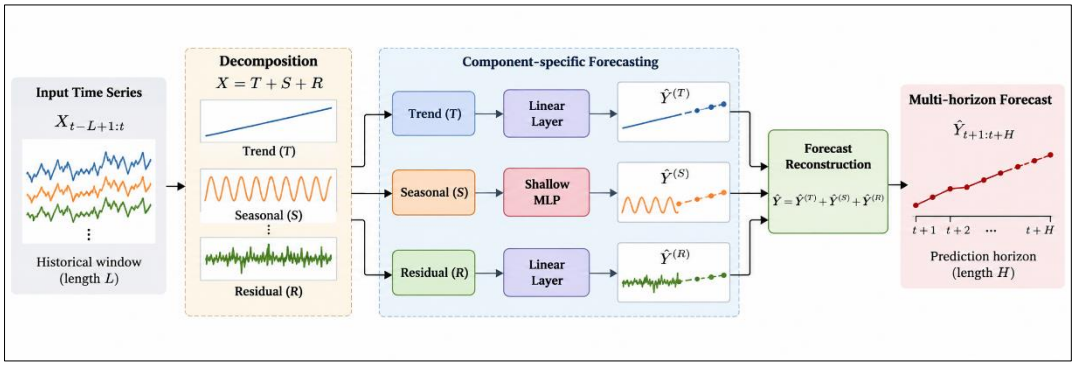


Figure 1: Overall process of D-LML

To investigate whether increasing forecast horizons necessarily require increasing architectural complexity, we introduce D-LML, a lightweight direct multi-step forecasting model based on decomposition and component-wise temporal mappings. Given a historical sequence $\boldsymbol{X} \in \mathbb{R}^{L \times C}$, D-LML first decomposes the input into trend, seasonal, and residual components, such that $\boldsymbol{X} = \boldsymbol{T} + \boldsymbol{S} + \boldsymbol{R}$. The trend and residual components are directly projected from the historical window $L$to the prediction horizon Husing temporal linear layers, whereas the seasonal component is processed by a shallow MLP. Accordingly, $\widehat{\boldsymbol{Y}}^{(T)} = f_L(\boldsymbol{T})$, $\widehat{\boldsymbol{Y}}^{(S)} = f_{\mathrm{MLP}}(\boldsymbol{S})$, and $\widehat{\boldsymbol{Y}}^{(R)} = f_L(\boldsymbol{R})$. The final forecast is reconstructed as $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{Y}}^{(T)} + \widehat{\boldsymbol{Y}}^{(S)} + \widehat{\boldsymbol{Y}}^{(R)}$. D-LML directly generates all Hfuture observations in a single forward pass and therefore avoids recursive prediction and its associated error propagation. The fundamental architecture is maintained across short-, medium-, and long-term forecasting experiments; only the output dimensionality changes according to the required prediction horizon. This deliberately lightweight design allows us to examine whether architectural complexity needs to increase as the forecasting horizon becomes longer. The overall architecture can therefore be summarized as

$$X \xrightarrow{\text{Decomposition}} \{T, S, R\} \xrightarrow{\text{Linear, MLP, Linear}} \{\hat{T}, \hat{S}, \hat{R}\} \xrightarrow{\text{Aggregation}} \hat{Y}.$$

We evaluate D-LML across progressively increasing prediction horizons representing short-, medium-, and long-range forecasting conditions and under both univariate and multivariate settings. Rather than evaluating predictive accuracy alone, we consider the joint relationship between forecast error, forecast horizon, parameter count, and computational complexity. D-LML is compared with representative forecasting paradigms, including linear models, residual MLP architectures, recurrent models, multiscale forecasting methods, and attention-based

architectures. This evaluation is designed to determine not merely whether D-LML can achieve the lowest error in individual settings, but whether a substantially simpler architecture can maintain a competitive accuracy complexity trade-off as the forecasting horizon increases. Detail in supplementary material.

4. Experiments:

We conduct extensive experiments to evaluate the performance and efficiency of D-LML, covering short, medium and long-term forecasting, including 5 real-world benchmarks and 4 baselines. The detailed model and experiment configurations are summarized:

Dataset	Frequency	Fixed lookback L	Historical coverage	Short H_S	Medium H_M	Long H_L
ETTh1	1 hour	336	14 days	24 = 1 day	168 = 1 week	720 = 30 days
Traffic	1 hour	336	14 days	24 = 1 day	168 = 1 week	720 = 30 days
Weather	10 min	1008	7 days	144 = 1 day	432 = 3 days	1008 = 1 week
Exchange-Rate	1 day	90	3 months	7 = 1 week	30 = 1 month	90 = 3 months
ILI	1 week	52	1 year	4 = 4 weeks	12 = 12 weeks	24 = 24 weeks

4.1 Datasets : We evaluate D-LML on widely adopted multivariate time-series forecasting benchmarks spanning different domains, sampling frequencies, sequence lengths, and dimensionalities. The benchmark suite comprises the two Electricity Transformer Temperature datasets (ETTh1), together with Traffic, Weather, Exchange-Rate, and ILI. Details attached in supplementary material. We select the historical lookback window according to the temporal resolution and dominant periodic structure of each dataset, while keeping the lookback fixed across all forecast horizons and competing models within a dataset.

4.2 Baselines :We compare D-LML with representative forecasting models drawn from several major architectural families. Rather than restricting the comparison to a single class of methods, the selected baselines cover linear, MLP-based, recurrent, multiscale, and attention-based approaches. These includes N-BEATS (Oreshkin et al., 2019), DeepAR (Salinas et al., 2020), Temporal Fusion Transformer (TFT) (Lim et al., 2021), and TimeMixer (Wang et al., 2024).

4.3 Forecast horizons. To evaluate model behaviour across different forecasting ranges, we define short-, medium-, and long-term prediction horizons according to the sampling frequency and temporal scale of each dataset. Rather than assigning identical numerical prediction lengths to datasets with different temporal resolutions, each horizon is expressed in terms of its corresponding real-world forecasting duration. For example, the Weather dataset is sampled every 10 minutes; therefore, prediction lengths of $H = 144, 432, and 1008$ correspond to $1-$ $,3-, and 7-day$ forecasts, respectively. This dataset-aware design enables D-LML to be evaluated consistently across short-, medium-, and long-range forecasting scenarios despite differences in sampling resolution. We have used RMSE, MAE, for the comparison of models efficiency.

4.4 Implementation Details: All models are implemented in **PyTorch** and trained separately for each prediction horizon.

4.5 Main Results: Multi-horizon forecasting. As shown in Table 2, D-LML demonstrates strong forecasting performance across datasets with substantially different temporal characteristics, dimensionalities, and prediction horizons. Multi-horizon forecasting performance across six benchmark datasets. Lower RMSE indicates better forecasting performance. Best results are shown in bold. All models are evaluated using the same chronological data partition, input information, and evaluation protocol. Full results including

MAE, sMAPE, R^2, parameter count, training time, and inference time are provided in the Appendix A.

Table 2: Short-term forecasting results. Results are reported at the short-term horizon for each dataset. A lower RMSE or MAE indicates better forecasting performance. **Bold** indicates the best result.

Dataset	Horizon	D-LML RMSE / MAE	TimeMixer RMSE / MAE	TFT RMSE / MAE	DeepAR RMSE / MAE	N-BEATS RMSE / MAE	Lowest RMSE
ILI	4	**175337.5 / 124452.3**	303567.8 / 271757.2	576165.9 / 533012.0	482487.8 / 449972.9	246551.3 / 206648.2	**D-LML**
Exchange	7	**0.00927 / 0.00700**	0.01048 / 0.00809	0.03239 / 0.02676	0.02045 / 0.01551	0.04192 / 0.03298	**D-LML**
ETTh1	24	**1.6942 / 1.2170**	2.0671 / 1.5908	1.7991 / 1.3609	2.0569 / 1.5622	2.8979 / 2.2717	**D-LML**
Weather	144	19.8131 / 14.9866	**15.9816 / 11.5958**	18.6446 / 14.1541	34.0762 / 15.2108	19.6345 / 13.8759	**TimeMixer**
Traffic	24	**0.00757 / 0.00461**	0.00957 / 0.00591	0.01031 / 0.00652	0.01059 / 0.00675	0.01009 / 0.00623	**D-LML**

The short-term results show the clearest advantage for D-LML, which achieves the lowest RMSE on four of the five datasets. D-LML is particularly effective for short-term forecasting, while TimeMixer is its main competitor and is strongest on Weather. TFT is competitive on ETTh1 and Weather but does not lead the short-term RMSE comparison. DeepAR and N-BEATS do not achieve the lowest short-term RMSE on these five datasets.

Table 3: Medium-term forecasting results. Results are reported at the medium-term prediction horizon for each dataset. A lower RMSE or MAE indicates better forecasting performance. **Bold** indicates the best result.

Dataset	Horizon	D-LML RMSE / MAE	TimeMixer RMSE / MAE	TFT RMSE / MAE	DeepAR RMSE / MAE	N-BEATS RMSE / MAE	Lowest RMSE
ILI	12	**229998.1 / 181380.9**	350883.2 / 322786.5	594033.3 / 553350.7	585907.6 / 543428.2	299763.1 / 260236.3	**D-LML**
Exchange	30	**0.01924 / 0.01510**	0.02225 / 0.01693	0.06109 / 0.05347	0.05162 / 0.04219	0.04657 / 0.03787	**D-LML**
ETTh1	168	**2.8053 / 2.2751**	3.5322 / 2.7450	3.0162 / 2.4580	3.1348 / 2.4451	3.5533 / 2.7955	**D-LML**
Weather	432	22.9558 / 17.7810	**17.4429 / 12.0527**	21.1996 / **7.9740**	19.9217 / 13.9332	23.6735 / 19.1794	**TimeMixer**
Traffic	168	0.01012 / 0.00656	0.01072 / 0.00694	**0.01000 / 0.00625**	0.01151 / 0.00773	0.01154 / 0.00782	**TFT**

At the medium horizon, the comparison becomes more dataset-dependent, although D-LML still records the lowest RMSE on three of five datasets: ILI, Exchange Rate, and ETTh1. On ETTh1, its RMSE of 2.8053 remains lower than TFT (3.0162), DeepAR (3.1348), TimeMixer (3.5322), and N-BEATS (3.5533). TimeMixer becomes more competitive on Weather, where it records the lowest RMSE of 17.4429, while TFT achieves the lowest RMSE on Traffic at 0.009999, narrowly ahead of D-LML at 0.010115. D-LML therefore retains strong medium-range performance, but TimeMixer and TFT begin to show advantages on particular datasets. DeepAR improves relatively on Weather but does not obtain the lowest RMSE, while N-BEATS remains behind the leading methods.

Table 4: Long-term forecasting results. Results are reported at the long-term prediction horizon for each dataset. A lower RMSE or MAE indicates better forecasting performance. **Bold** indicates the best result.

Dataset	Horizon	D-LML RMSE / MAE	TimeMixer RMSE / MAE	TFT RMSE / MAE	DeepAR RMSE / MAE	N-BEATS RMSE / MAE	Lowest RMSE
ILI	24	**267529.7 / 224508.7**	462382.1 / 432245.3	589731.0 / 545293.5	589424.7 / 542149.8	351280.9 / 321929.5	**D-LML**
Exchange	90	**0.03802 / 0.03011**	0.05873 / 0.05025	0.11190 / 0.09391	0.05858 / 0.04611	0.06345 / 0.05124	**D-LML**
ETTh1	720	3.5499 / 2.8149	**3.2773 / 2.5654**	3.9800 / 3.2761	4.1121 / 3.2821	4.6998 / 3.8300	**TimeMixer**
Weather	1008	39.2808 / 30.5470	20.8674 / 16.0515	**20.1507 / 15.4338**	26.9660 / 22.7675	34.9819 / 30.2621	**TFT**
Traffic	720	0.01436 / 0.01173	0.01596/ 0.01297	**0.01103 / 0.00741**	0.01314 / 0.00912	0.01303 / 0.00915	**TFT**

Long-term forecasting reveals a stronger division between the methods. D-LML remains strongest on ILI and Exchange Rate, demonstrating that its simple component-wise structure can remain effective over extended horizons on these datasets. However, TimeMixer becomes strongest on ETTh1, reducing RMSE from D-LML's 3.5499 to 3.2773. TFT records the lowest reported RMSE on Weather and Traffic at the longest horizons.This suggests that D-LML's relative advantage is greatest at shorter horizons, while multi-scale or attention-based architectures can become more competitive when very long-range dependencies dominate. Importantly, this is dataset-dependent rather than a uniform deterioration: D-LML still leads ILI and Exchange at their longest evaluated horizons.

Across forecasting ranges, the experiments reveal a clear horizon-dependent pattern. D-LML exhibits its strongest relative performance in short-term forecasting, achieving the lowest RMSE on four of five datasets, and maintains strong performance at medium horizons with three of five RMSE wins. This indicates that decomposition followed by component-specific linear/nonlinear modeling effectively captures local and intermediate temporal dynamics without requiring a highly complex forecasting architecture. At longer horizons, however, performance becomes more dataset dependent. D-LML remains strongest on ILI and Exchange Rate, whereas TimeMixer obtains the lowest RMSE on long-horizon ETTh1 and TFT leads the reported results on Weather and Traffic. The results therefore do not indicate a universal superiority of one architecture across all forecasting scales; rather, they suggest that D-LML is particularly effective at short and medium horizons, while the relative benefits of multi-scale and attention-based modeling become more apparent for some very long-horizon settings. DeepAR and N-BEATS do not achieve the lowest RMSE in the reported comparisons, although their relative performance varies across datasets and horizons.

4.6 Ablation Study.

To verify the effectiveness of the component-specific modeling strategy in D-LML, we conduct ablation experiments by varying the allocation of linear and nonlinear predictors across the decomposed trend, seasonal, and residual components. All variants are parameter-matched, allowing us to isolate the effect of where nonlinear capacity is allocated rather than differences in model size.

Table 5: Ablation study on component-specific nonlinear modeling in D-LML on ETTh1 and Exchange Rate. L and MLP denote linear and nonlinear MLP predictors, respectively. ① represents the proposed D-LML design, where the MLP is assigned to the seasonal component while the trend and residual components are modelled linearly. Bold values indicate the lowest RMSE at each forecasting horizon.

Case	Trend	Seasonal	Residual	ETTh1 H=24	ETTh1 H=168	ETTh1 H=720	Exchange H=7	Exchange H=30	Exchange H=90
① **D-LML**	L	**MLP**	L	**1.6865**	**2.8031**	3.1326	**0.00923**	**0.01855**	0.03995
②	L	L	**MLP**	1.7263	2.8424	3.0965	0.00925	0.01882	0.04159
③	**MLP**	L	L	1.7381	2.9275	**2.9511**	0.01037	0.02050	**0.03892**

From Table 5 , we make the following observations. D-LML (L-MLP-L), which applies linear mappings to the trend and residual components and an MLP to the seasonal component, achieves the best forecasting performance in 4 of the 6 evaluated settings, including all short- and medium-horizon experiments on Exchange Rate and ETTh1. Replacing the seasonal MLP with a linear predictor while assigning the MLP to the residual component (L-L-MLP) generally reduces forecasting accuracy, indicating that additional nonlinear capacity in the residual component provides limited benefit. Similarly, assigning the MLP exclusively to the trend component (MLP-L-L) underperforms D-LML at short and medium horizons, supporting the effectiveness of nonlinear seasonal modelling. Interestingly, MLP-L-L becomes the strongest configuration at the longest horizons (Exchange Rate H=90 and ETTh1 H=720). This suggests that the optimal allocation of nonlinear capacity is horizon-dependent: nonlinear seasonal modelling is particularly effective for short-to-medium horizons, whereas nonlinear trend modelling can become more beneficial for long-term forecasting. Overall, these results demonstrate that temporal components do not require uniform modelling complexity and support D-LML central design principle of selectively allocating nonlinear capacity according to component characteristics.

5 Conclusion

This study introduced D-LML (Decomposition–Linear–MLP–Linear), a simple component-specific framework for multi-horizon time-series forecasting. Across multiple benchmark datasets, D-LML demonstrated particularly strong performance at short- and medium-term horizons, while long-term performance was more dataset dependent. The ablation analysis further showed that selectively assigning nonlinear capacity to different decomposed components can substantially affect forecasting accuracy, with seasonal nonlinear modeling being particularly effective at shorter horizons and trend nonlinear modeling becoming beneficial in some longer-horizon settings. Overall, the findings demonstrate that selective nonlinear modeling combined with simple linear structures can provide an effective alternative to uniformly complex forecasting architectures, offering a promising direction for accurate and efficient multi-horizon forecasting.

6 Ethnics Statement

This study focuses on methodological advances in time-series forecasting using publicly available benchmark datasets. The research does not involve human participants, personal or sensitive information, or interventions that could directly affect individuals or communities. Therefore, we do not identify any significant ethical risks associated with this work.

7 Reproducibility statement

To support reproducibility, we provide detailed information on the datasets, preprocessing procedures, evaluation metrics, model configurations, hyperparameter settings, and experimental protocol in the Appendix A . The implementation code and supplementary materials will also be made publicly available upon publication, enabling researchers to reproduce the reported experiments and further evaluate the proposed D-LML framework.

8 References

Beniwal, M., Singh, A., & Kumar, N. (2024). Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices. *Engineering Applications of Artificial Intelligence*, *129*, 107617. https://doi.org/https://doi.org/10.1016/j.engappai.2023.107617

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748-1764. https://doi.org/https://doi.org/10.1016/j.ijforecast.2021.03.012

Liu, X., & Wang, W. (2024). Deep Time Series Forecasting Models: A Comprehensive Survey. *Mathematics*, *12*(10), 1504. https://www.mdpi.com/2227-7390/12/10/1504

Ni, Z., Zhang, C., Karlsson, M., & Gong, S. (2024). A study of deep learning-based multi-horizon building energy forecasting. *Energy and Buildings*, *303*, 113810. https://doi.org/https://doi.org/10.1016/j.enbuild.2023.113810

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.

Perera, M., De Hoog, J., Bandara, K., & Halgamuge, S. (2022). Multi-resolution, multi-horizon distributed solar PV power forecasting with forecast combinations. *Expert Systems with Applications*, *205*, 117690. https://doi.org/https://doi.org/10.1016/j.eswa.2022.117690

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., & Willson, M. (2025). Probabilistic weather forecasting with machine learning. *Nature*, *637*(8044), 84-90. https://doi.org/10.1038/s41586-024-08252-9

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181-1191. https://doi.org/https://doi.org/10.1016/j.ijforecast.2019.07.001

Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J., & Zhou, J. (2024). Timemixer: Decomposable multiscale mixing for time series forecasting. International conference on learning representations,

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.

Yang, B., Sun, S., Li, J., Lin, X., & Tian, Y. (2019). Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*, *332*, 320-327.

Zhou, Y., An, Y., Huang, W., Chen, C., & You, R. (2022). A combined deep learning and physical modelling method for estimating air pollutants' source location and emission profile in street canyons. *Building and Environment*, *219*, 109246.