Linear vs. Transformer Models for Long-Horizon Exogenous Temperature Forecasting

Ruslan Gokhman, Ph.D. in Mathematics

FACULTY MENTOR: Marian Gidea, Ph.D.

Katz
Katz School of Science and Health

Introduction

Background

- Study: long-horizon **exogenous-only** temperature forecasting
- Models tested: Linear, NLinear, DLinear, Transformer, Informer, Autoformer
- Evaluation: standardized train / validation / test splits

Linear baselines (Linear, NLinear, DLinear) outperform all Transformer-family models. NLinear achieves the best accuracy (MAE = 0.246 / MSE = 0.522) on the official test set.All models have been trained with identical 96-step input/output and exogenous-only features.

Method

- Compare Linear-family models vs Transformer-family models for long-horizon temperature forecasting
- Data & setup: Input/output horizon: 96 steps; input data consists of historical measurements **from other air conditioners.**
- Dataset link: https://storage.googleapis.com/gresearch/smart_buildings_dataset/tabular_data/sb1.zip
- Univariate (temperature only)
- Linear models (LTSF-Linear):
- Vanilla Linear: single linear layer (input → forecast)
- NLinear: normalization via last timestep subtraction
- DLinear: trend + seasonal decomposition before projection
- Transformer models:
  Vanilla Transformer, Informer, Autoformer
- Training & evaluation:
  Framework: PyTorch
  Optimizer: Adam
  Early stopping on validation MAE
  Standardized train / validation / blind test split

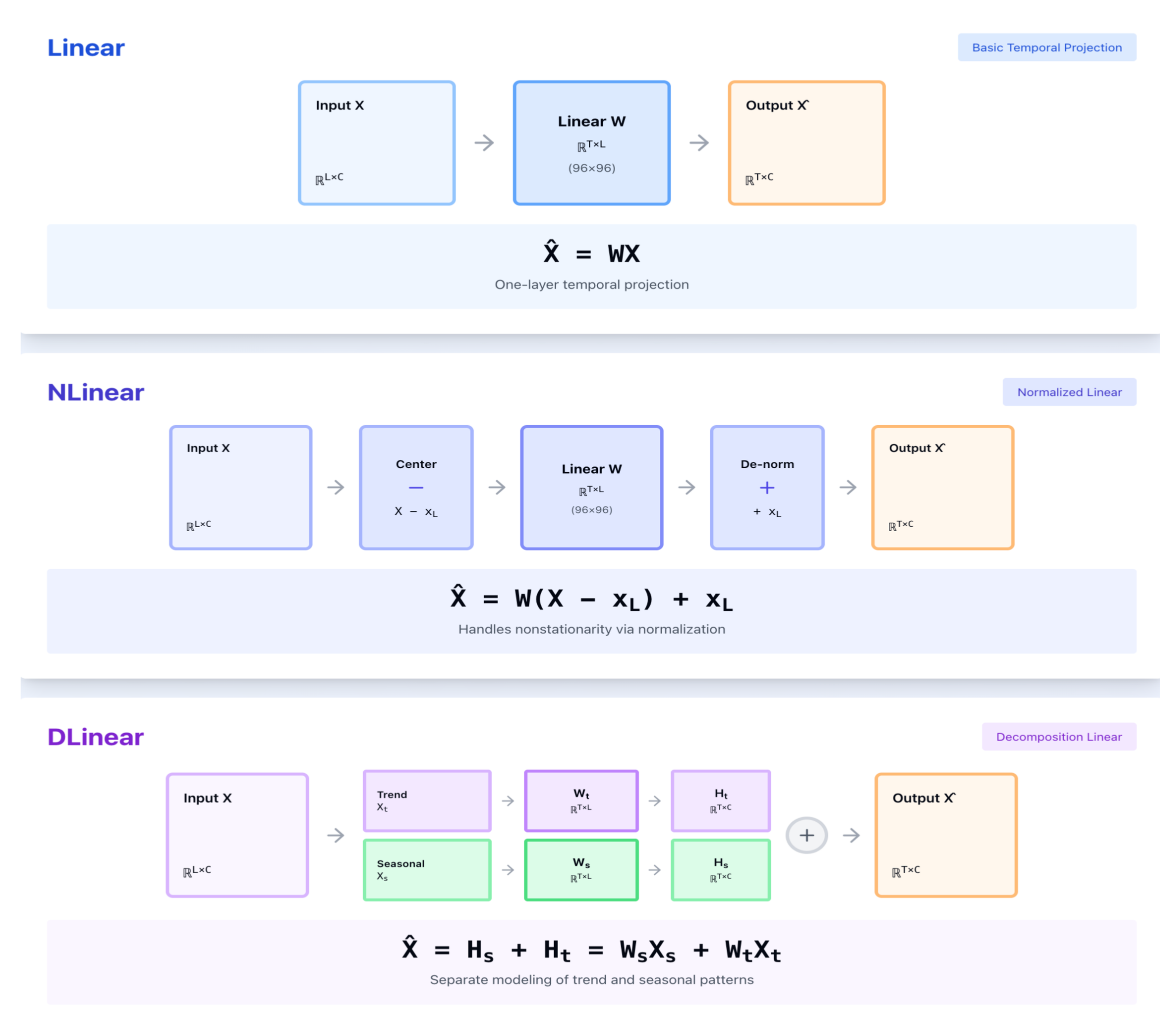
Linear-Based Architecture

Input: past temperature sequence | Output: future horizon (96 steps)

Core idea:
- Direct **linear mapping** from input → forecast
- No attention, recurrence, or deep stacking

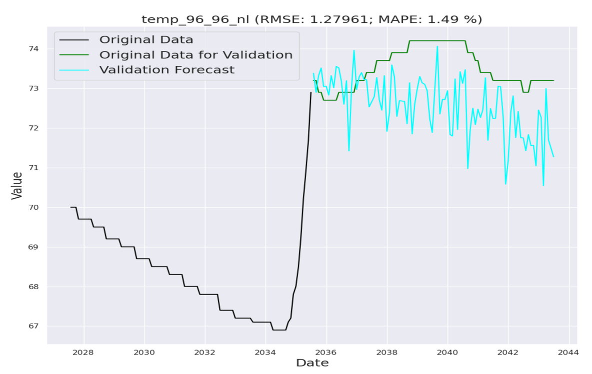
Variants:
- **Linear:** single linear layer
- **NLinear:** subtract last value (normalization)
- **DLinear:** decompose into trend + seasonal, then project

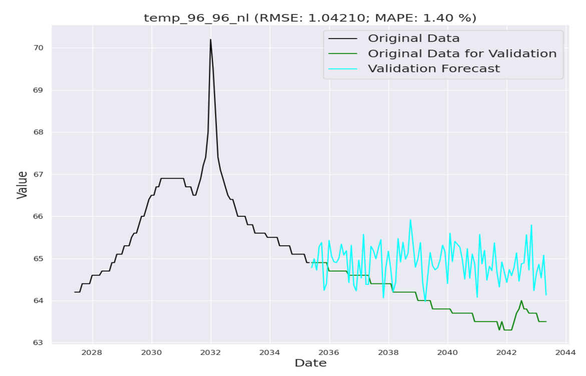
Why it works:
- Preserves temporal structure without overfitting
- Efficient and stable for long-horizon forecasting



Experimental Results

Model	Split	Size	MAE	MSE
NLinear	Training	36,105	–	0.0619
DLinear	Training	36,105	–	0.0660
Linear	Training	36,105	–	0.0679
Transformer	Training	36,105	–	0.0330
Informer	Training	36,105	–	–
Autoformer	Training	36,105	–	0.0844
NLinear	Validation	10,275	0.2529	0.4665
DLinear	Validation	10,275	0.2784	0.4744
Linear	Validation	10,275	0.2843	0.4882
Transformer	Validation	10,275	0.3375	0.5302
Informer	Validation	10,275	–	–
Autoformer	Validation	10,275	0.3493	0.6201
NLinear	**Official Test**	**10,563**	**0.2461**	**0.5220**
DLinear	Official Test	10,563	0.2811	0.5489
Linear	Official Test	10,563	0.2862	0.5634
Transformer	Official Test	10,563	0.8342	1.4371
Informer	Official Test	10,563	0.8342	1.4371
Autoformer	Official Test	10,563	0.3598	0.7368





Conclusions

Finding	Why
Linear models outperform Transformers	Simpler structure reduces overfitting and improves generalization on test data
NLinear achieves best performance	Normalization via last timestep removes level shifts and stabilizes predictions
DLinear performs slightly worse than NLinear	Decomposition helps, but adds extra complexity without clear gain
Vanilla Linear still strong	Direct mapping is robust and generalizes better than high-capacity models
Transformers overfit	High model capacity with limited univariate data leads to poor test performance
Weak inductive bias in Transformers	Lack of built-in trend/seasonality assumptions reduces efficiency
Long-horizon favors linear models	Linear extrapolation is more stable across 96-step forecasting
Autoformer improves but not enough	Decomposition helps, but still less robust than linear baselines
Informer underperforms	Sparse attention sacrifices accuracy for efficiency in simple settings
Key takeaway	Linear models provide the best balance of accuracy, robustness, and interpretability

Future Work

- We plan to explore hybrid architectures that combine linear components with lightweight attention mechanisms to better capture regime shifts in exogenous-only settings.
- Additionally, we will assess model robustness under domain shifts, such as extreme weather events and cross-regional generalization.
- Future research will also incorporate richer exogenous features, including spatial and atmospheric variables, to examine whether linear models maintain their advantage as input dimensionality increases.
- Finally, we aim to develop and release a standardized benchmark for long-horizon exogenous forecasting to enable fair and consistent comparison between linear and Transformer-based models across diverse datasets.

References

Judah Goldfeder, Victoria Dean, Zixin Jiang, Xuezheng Wang, Bing Dong, Hod Lipson, and John Sipple. The smart buildings control suite: A diverse open source benchmark to evaluate and scale hvac control policies for sustainability. arXiv preprint arXiv:2410.03756, 2025. URL https://arxiv.org/abs/2410.03756.
CURE Lab. Ltsf-linear: Linear baselines for long-term time series forecasting. https://github.com/cure-lab/LTSF-Linear, 2022.
Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, 2017.
Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In Advances in Neural Information Processing Systems, 2021.
Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, 2021.