Low-Carbon Forecasting of National Electricity Demand: A Benchmark Across 41 Models, 4 Countries, and 2 Seasons

First Last
Affiliation
email@example.com

First Last
Affiliation
email@example.com

First Last
Affiliation
email@example.com

***Abstract*—We benchmark 41 forecasting models across Denmark, Germany, Hungary, and Spain in summer and winter, tracking carbon emissions with CodeCarbon. We find weak coupling between accuracy and emissions, with many low-emission models achieving state-of-the-art MAE. We provide practical, low-carbon defaults and discuss trade-offs.**

***Index Terms*—Time series forecasting, energy demand, carbon accounting, CodeCarbon, benchmarking, sustainability**

I. Introduction

Electricity demand forecasting supports grid stability and planning. Training modern models can be carbon-intensive. We quantify both accuracy and emissions across diverse model families and seasons.

- Contribution 1: Joint evaluation of accuracy and emissions across 41 models.
- Contribution 2: Public aggregation pipeline and artifacts for reproducibility.
- Contribution 3: Low-carbon defaults with minimal accuracy sacrifice.

II. Related Work

[1], [2]

III. Data

We use 5-year hourly demand data for four European countries with summer and winter subsets. See repository Data/ for CSVs.

IV. Models

We include classical and neural architectures (DLinear, CNN-LSTM, Cycle-LSTM, Transformers including Informer and PatchTST, N-BEATS, Autoformer, TFT, Mamba, and hybrids). Hyperparameters follow repository defaults.

V. Experimental Setup

We run each model across country-season combinations using a Python orchestrator. CodeCarbon tracks per-run emissions (kg CO_2e), energy (kWh), and duration (s).

A. Metrics

We report MAE, RMSE, MSE, R^2, and MAPE. Aggregations are saved to metrics_aggregated.csv and emissions_aggregated.csv.

TABLE I
MAE SUMMARY BY COUNTRY AND SEASON (BEST/WORST/SIZE).

Slice	Best (model)	Best MAE	Worst (model
Denmark Summer	Robust_Improved_Hybrid_Model_v2	5.053	N_Beats_Mod
Denmark Winter	Transformer_Model	9.515	Transformer_
Germany Summer	Cycle_LSTM_Model_v2	9.433	Transformer_
Germany Winter	Cycle_LSTM_Model	11.640	Autoformer_M
Hungary Summer	DLinear_Model	3.979	Mamba_Mod
Hungary Winter	Robust_Improved_Hybrid_Model	4.514	Mamba_Mod
Spain Summer	DLinear_Model_v2	4.162	Transformer_
Spain Winter	DLinear_Model	6.422	Mamba_Mod

B. Reproducibility

VI. Results

A. Accuracy

B. Emissions

Most runs emit $\leq$ 0.100 kg CO_2e; a few outliers (e.g., Transformer v3 variants) reach 0.600 kg to 1.000 kg. See Figure 1.

C. Accuracy–Emissions Trade-off

Pearson ρ(MAE, emissions) $\approx$ 0.112. Many low-emission runs lie on or near the Pareto frontier. See Figure 2.

VII. Discussion

Low-emission families (DLinear, CNN-LSTM, Robust Hybrid, Cycle-LSTM) often match or beat higher-emission models. A soft emission cap of 0.100 kg per training filters dominated configurations without hurting accuracy.

VIII. Recommendations

Default picks by slice: Denmark (Robust Hybrid v2), Germany (Cycle-LSTM v2), Hungary (DLinear/Robust Hybrid), Spain (DLinear v2). Avoid high-emission Transformer v3 for routine runs.

IX. Limitations and Future Work

We observed a Germany-winter count anomaly (N=42). Future work: Pareto front tables per slice, uncertainty intervals, family-level emission summaries.

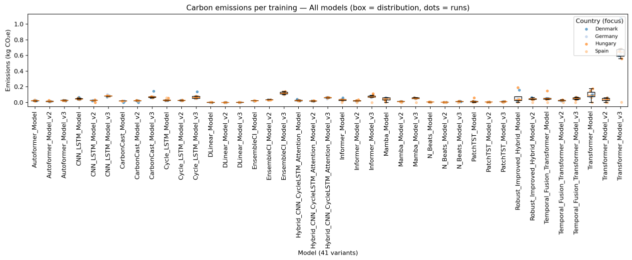


Fig. 1. Emissions distribution across models.

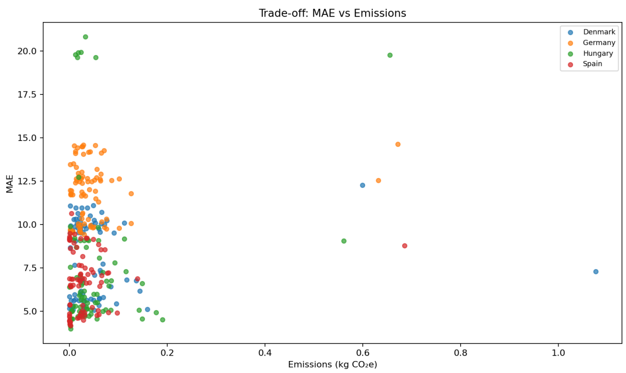


Fig. 2. MAE vs Emissions trade-off.

X. Conclusion

Accuracy and carbon efficiency need not be in tension. Our benchmark and pipeline surface low-carbon choices with competitive MAE.

Acknowledgments

Appendix A
Reproducibility Commands (Windows PowerShell)

Run orchestrations and regenerate benchmark artifacts.

```
# Full run (both seasons, 4 countries, repeat each twice)
python .\main.py --season both --countries DE,DK,ES,HU --repeat 2

# Quick smoke test for DLinear family (summer)
python .\main.py --season summer --filter DLinear --quick --countries DE,DK,ES,HU

# Aggregate and plot
python -m scripts.benchmark

# Regenerate LaTeX tables for the paper
python .\Paper\generate_tables.py
```

TABLE II
Low-carbon leaders (best MAE with emissions ≤ 0.10 kg CO_2e) by slice.

Slice	Model	MAE	Emissions (kg)
Denmark Summer	Robust_Improved_Hybrid_Model_v2	5.053	0.050
Denmark Winter	Transformer_Model	9.515	0.091
Germany Summer	Cycle_LSTM_Model_v2	9.433	0.030
Germany Winter	Cycle_LSTM_Model	11.640	0.025
Hungary Summer	DLinear_Model	3.979	0.002
Hungary Winter	DLinear_Model_v2	4.528	0.002
Spain Summer	DLinear_Model_v2	4.162	0.002
Spain Winter	DLinear_Model	6.422	0.003

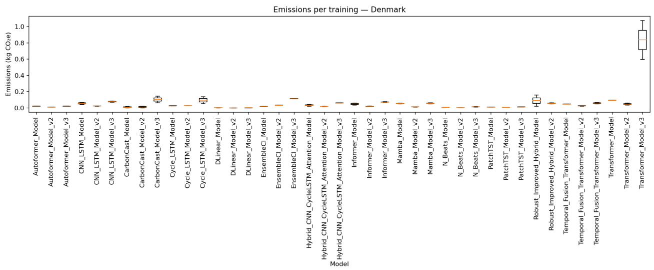
TABLE III
Pareto leaders by slice (minimizing MAE and emissions).

Slice	Model	MAE	Emissions (kg)
Denmark Summer	Robust_Improved_Hybrid_Model_v2	5.053	0.050
Denmark Summer	N_Beats_Model	5.072	0.007
Denmark Summer	N_Beats_Model_v2	5.134	0.004
Denmark Winter	Transformer_Model	9.515	0.091
Denmark Winter	Cycle_LSTM_Model	9.518	0.023
Denmark Winter	DLinear_Model_v2	9.519	0.001
Germany Summer	Cycle_LSTM_Model_v2	9.433	0.030
Germany Summer	PatchTST_Model_v2	9.533	0.007
Germany Summer	N_Beats_Model_v2	9.545	0.003
Germany Winter	Cycle_LSTM_Model	11.649	0.028
Germany Winter	DLinear_Model	11.710	0.002
Germany Winter	PatchTST_Model_v2	11.710	0.007
Hungary Summer	DLinear_Model	3.979	0.002
Hungary Summer	DLinear_Model_v2	4.134	0.002
Hungary Summer	CNN_LSTM_Model_v2	6.398	0.000
Hungary Winter	Robust_Improved_Hybrid_Model	4.514	0.190
Hungary Winter	DLinear_Model_v2	4.528	0.002
Hungary Winter	DLinear_Model_v3	7.535	0.001
Spain Summer	DLinear_Model_v2	4.162	0.002
Spain Summer	Robust_Improved_Hybrid_Model	4.232	0.001
Spain Summer	Robust_Improved_Hybrid_Model_v2	4.286	0.000
Spain Winter	DLinear_Model	6.422	0.003
Spain Winter	DLinear_Model_v2	6.447	0.001
Spain Winter	Transformer_Model_v2	6.869	0.000

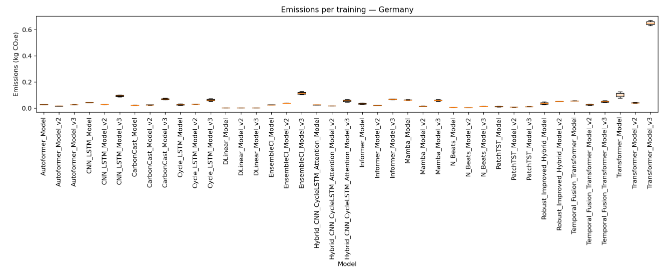
References

[1] A. Lacoste, A. S. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the carbon emissions of machine learning," *arXiv preprint arXiv:1910.09700*, 2019.

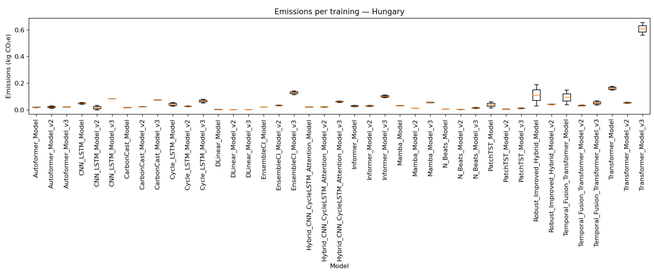
[2] P. Henderson, J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau, "Towards the systematic reporting of the energy and carbon footprints of machine learning," in *International Conference on Machine Learning (ICML) Workshop on Challenges in Deploying and Monitoring Machine Learning Systems*, 2020.



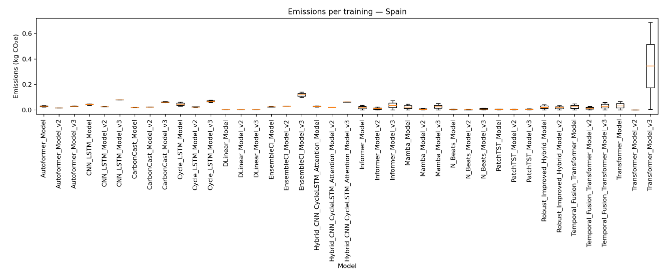
(a) Denmark



(b) Germany

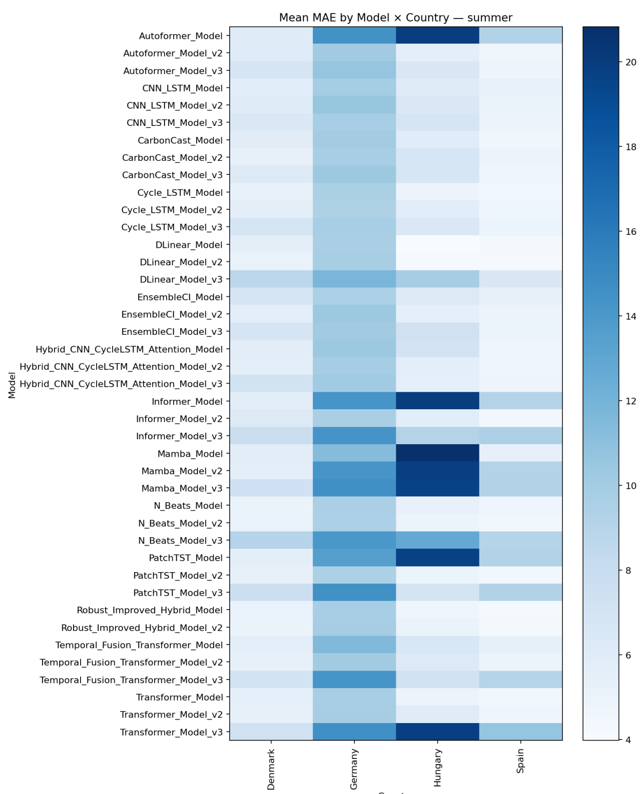


(c) Hungary

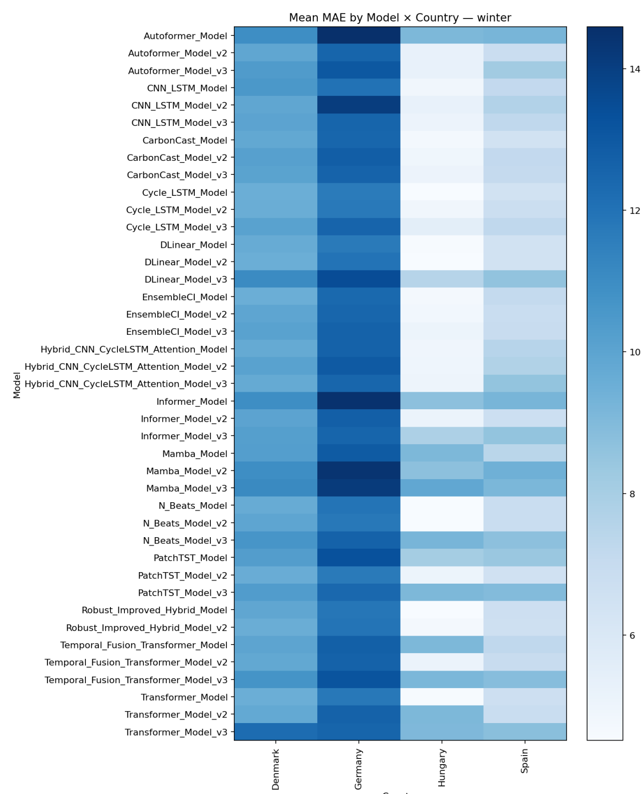


(d) Spain

Fig. 3. Per-country emissions distributions across models.

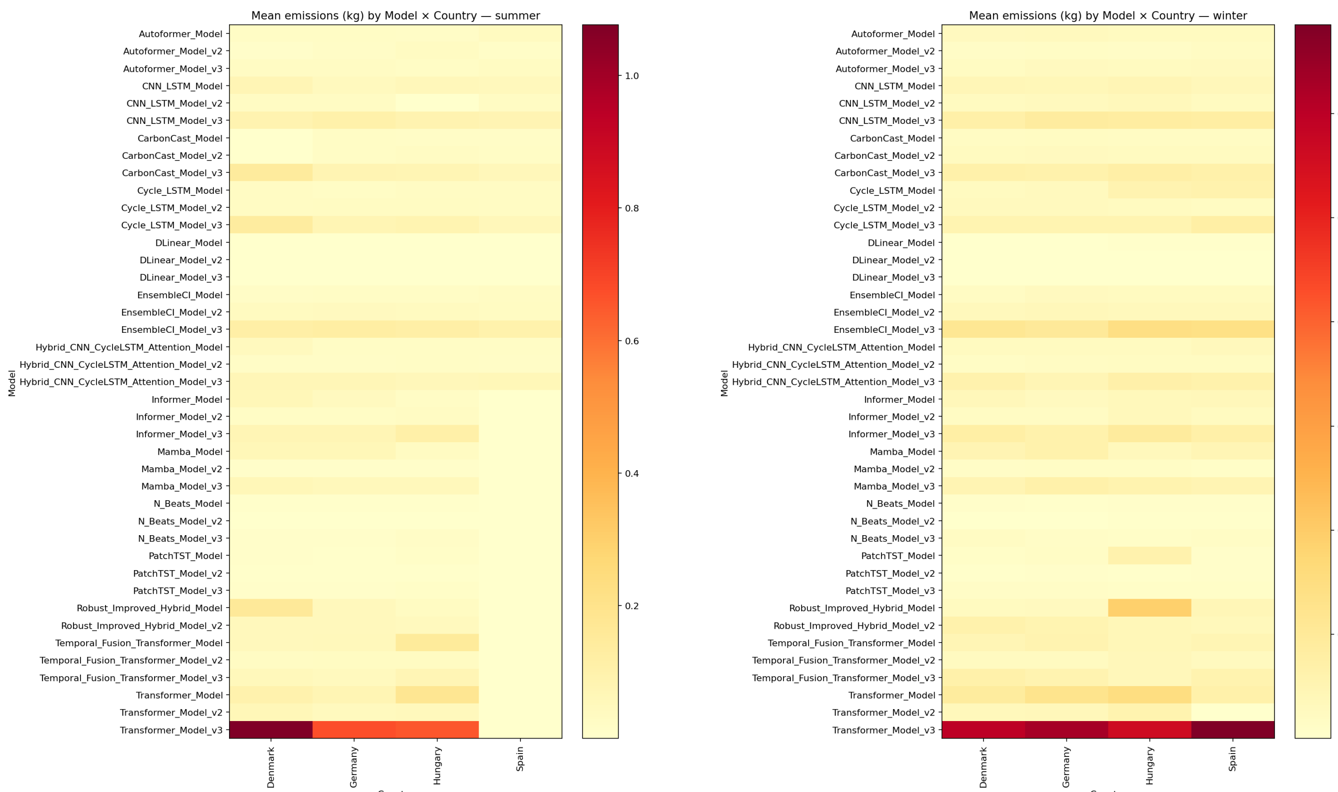


(a) Mean MAE (summer)



(b) Mean MAE (winter)

Fig. 4. Heatmaps of mean MAE by model and country.



(a) Mean emissions (summer)

(b) Mean emissions (winter)

Fig. 5. Heatmaps of mean emissions (kg CO_2e) by model and country.

TABLE IV
TOP-10 MAE MODELS PER SLICE (AVERAGED OVER RUNS).

Slice	Rank	Model	MAE
Denmark Summer	1	Robust_Improved_Hybrid_Model_v2	5.053
Denmark Summer	2	N_Beats_Model	5.072
Denmark Summer	3	Robust_Improved_Hybrid_Model	5.104
Denmark Summer	4	N_Beats_Model_v2	5.134
Denmark Summer	5	DLinear_Model_v2	5.155
Denmark Summer	6	Cycle_LSTM_Model	5.189
Denmark Summer	7	PatchTST_Model_v2	5.312
Denmark Summer	8	Temporal_Fusion_Transformer_Model_v2	5.319
Denmark Summer	9	Transformer_Model_v2	5.339
Denmark Summer	10	CarbonCast_Model_v2	5.343
Denmark Winter	1	Transformer_Model	9.515
Denmark Winter	2	Cycle_LSTM_Model	9.518
Denmark Winter	3	DLinear_Model_v2	9.519
Denmark Winter	4	EnsembleCI_Model	9.528
Denmark Winter	5	Robust_Improved_Hybrid_Model_v2	9.553
Denmark Winter	6	Cycle_LSTM_Model_v2	9.563
Denmark Winter	7	PatchTST_Model_v2	9.601
Denmark Winter	8	DLinear_Model	9.646
Denmark Winter	9	N_Beats_Model	9.654
Denmark Winter	10	Hybrid_CNN_CycleLSTM_Attention_Model	9.714
Germany Summer	1	Cycle_LSTM_Model_v2	9.433
Germany Summer	2	PatchTST_Model_v2	9.533
Germany Summer	3	EnsembleCI_Model	9.545
Germany Summer	4	N_Beats_Model_v2	9.545
Germany Summer	5	Cycle_LSTM_Model	9.590
Germany Summer	6	N_Beats_Model	9.618
Germany Summer	7	DLinear_Model	9.653
Germany Summer	8	Informer_Model_v2	9.671
Germany Summer	9	DLinear_Model_v2	9.710
Germany Summer	10	Transformer_Model	9.743
Germany Winter	1	Cycle_LSTM_Model	11.649
Germany Winter	2	PatchTST_Model_v2	11.710
Germany Winter	3	DLinear_Model	11.710
Germany Winter	4	Cycle_LSTM_Model_v2	11.749
Germany Winter	5	N_Beats_Model_v2	11.763
Germany Winter	6	Transformer_Model	11.777
Germany Winter	7	Robust_Improved_Hybrid_Model	11.869
Germany Winter	8	Robust_Improved_Hybrid_Model_v2	11.926
Germany Winter	9	N_Beats_Model	11.948
Germany Winter	10	DLinear_Model_v2	11.960
Hungary Summer	1	DLinear_Model	3.979
Hungary Summer	2	DLinear_Model_v2	4.134
Hungary Summer	3	Robust_Improved_Hybrid_Model	4.701
Hungary Summer	4	Cycle_LSTM_Model	4.911
Hungary Summer	5	Transformer_Model	4.918
Hungary Summer	6	PatchTST_Model_v2	4.991
Hungary Summer	7	N_Beats_Model_v2	4.993
Hungary Summer	8	Robust_Improved_Hybrid_Model_v2	5.078
Hungary Summer	9	N_Beats_Model	5.272
Hungary Summer	10	EnsembleCI_Model_v2	5.463
Hungary Winter	1	Robust_Improved_Hybrid_Model	4.514
Hungary Winter	2	DLinear_Model_v2	4.528
Hungary Winter	3	N_Beats_Model_v2	4.533
Hungary Winter	4	Cycle_LSTM_Model	4.552
Hungary Winter	5	N_Beats_Model	4.553
Hungary Winter	6	DLinear_Model	4.560
Hungary Winter	7	Transformer_Model	4.561
Hungary Winter	8	CarbonCast_Model	4.686
Hungary Winter	9	Robust_Improved_Hybrid_Model_v2	4.718
Hungary Winter	10	EnsembleCI_Model	4.742
Spain Summer	1	DLinear_Model_v2	4.162
Spain Summer	2	Robust_Improved_Hybrid_Model	4.232
Spain Summer	3	Robust_Improved_Hybrid_Model_v2	4.286
Spain Summer	4	DLinear_Model	4.369
Spain Summer	5	Transformer_Model	4.382
Spain Summer	6	Informer_Model_v2	4.425
Spain Summer	7	N_Beats_Model_v2	4.436
Spain Summer	8	PatchTST_Model_v2	4.461
Spain Summer	9	Cycle_LSTM_Model	4.493
Spain Summer	10	Autoformer_Model_v2	4.605

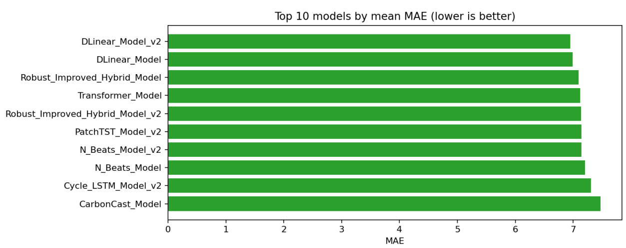
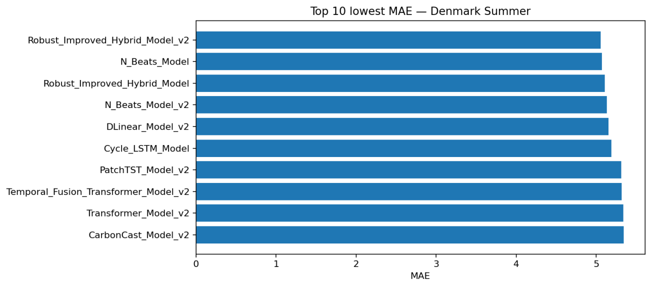
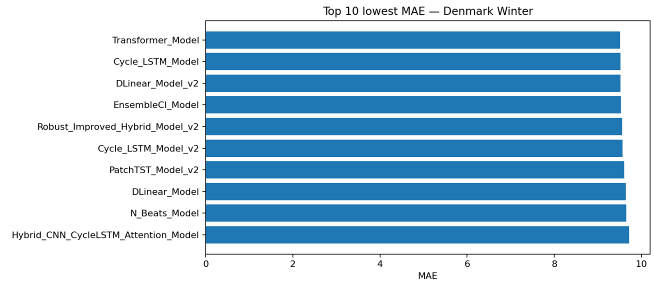


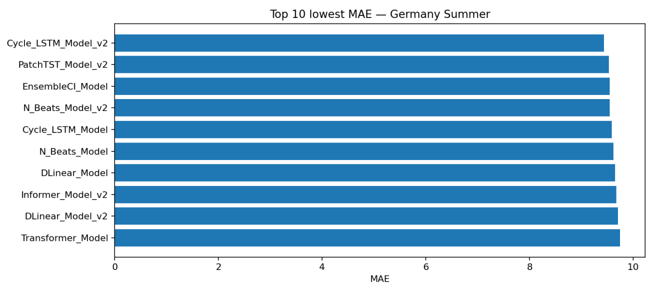
Fig. 6. Global Top-10 MAE leaderboard across slices.



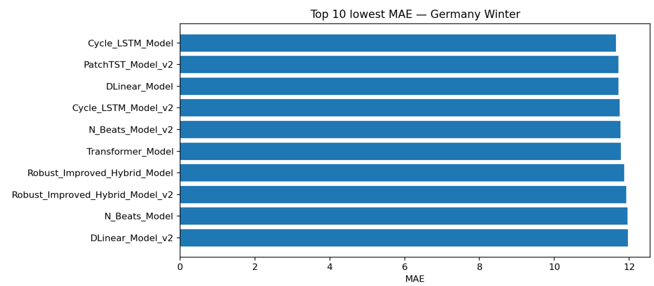
(a) Denmark Summer



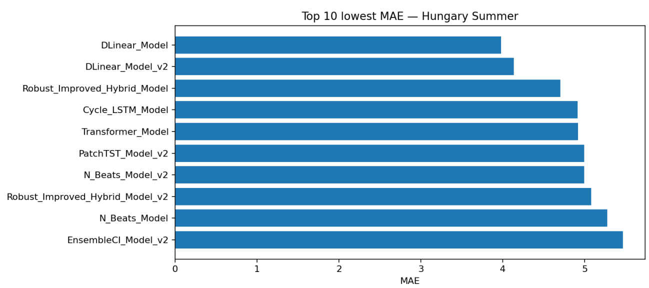
(b) Denmark Winter



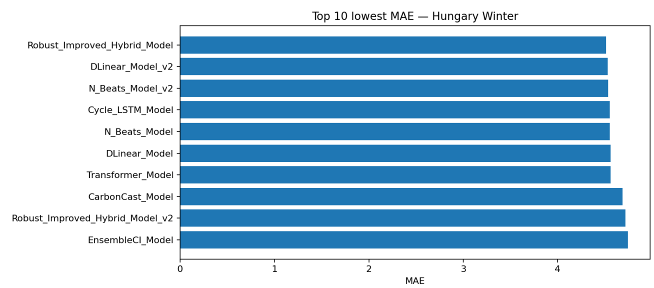
(c) Germany Summer



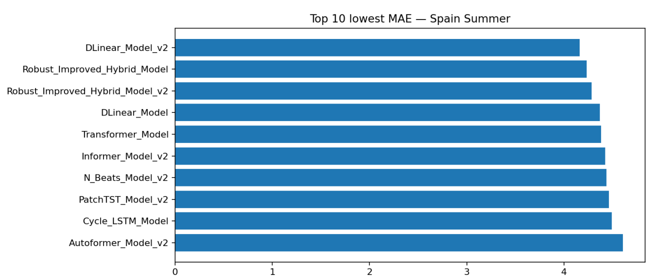
(d) Germany Winter



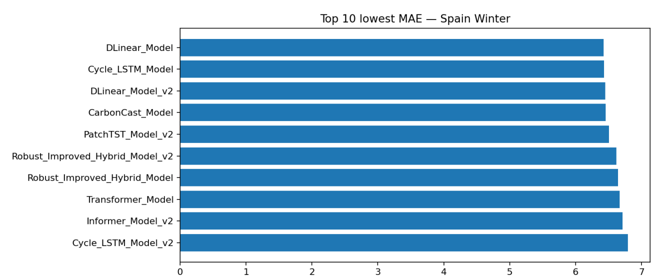
(e) Hungary Summer



(f) Hungary Winter



(g) Spain Summer



(h) Spain Winter

Fig. 7. Per-slice Top-10 MAE leaderboards.