INTERNATIONAL JOURNAL OF MANAGEMENT AND ORGANIZATIONAL RESEARCH

Does More Complex AI Produce Greater Net Carbon Value? Carbon-Aware Renewable-Energy Forecasting and Model Selection

Alexander J Thorne [1*], Olivia M Santos [2]
[1-2] Isenberg School of Management, University of Massachusetts Amherst, Massachusetts, USA

* Corresponding Author: **Alexander J Thorne**

Article Info

ISSN (online): 2583-6641
Volume: 04
Issue: 06
Received: 06-10-2025
Accepted: 05-11-2025
Published: 04-12-2025
Page No: 154-163

Abstract
Renewable-energy forecasting studies usually rank models by predictive error, although model development and deployment also consume electricity and produce greenhouse-gas emissions. This study asks whether additional predictive sophistication creates positive net carbon value after computational emissions are deducted. The empirical analysis uses the frozen October 6, 2020 release of Open Power System Data for Germany, covering hourly solar and wind generation from January 2015 through September 2020. Generation is normalized by contemporaneous installed capacity. Persistence, ridge regression, random forest, and extremely randomized trees are evaluated for one-hour and 24-hour forecasting under a strict temporal design, with 2020 reserved for out-of-time testing. Forecast error is observed directly. Computational electricity is estimated from measured wall-clock runtime, a documented 65 W processor-power assumption, and a power-usage-effectiveness factor of 1.2. Operational emissions benefits are not treated as observed. They are estimated under low, central, and high scenarios that translate reductions in absolute forecast error into changes in balancing energy. Ridge regression generated the lowest mean absolute error in three of four tasks; persistence remained best for 24-hour solar forecasting. Its strongest advantage occurred for one-hour solar forecasting, where mean absolute error fell from 0.0321 under persistence to 0.0081. Greater algorithmic complexity did not produce better results. Tree ensembles consumed more training energy and underperformed ridge regression, and no learned model improved 24-hour solar mean absolute error relative to persistence. Under the central operational scenario, ridge regression also produced the highest net carbon value. The ranking was driven primarily by operational assumptions, however, because estimated computational emissions were small for the tested data and hardware scale. The findings support carbon-aware model selection based on practical accuracy, deployment burden, and operational value rather than accuracy alone.

DOI: https://doi.org/10.54660/IJMOR.2025.4.6.154-163

Keywords: renewable-energy forecasting, sustainable artificial intelligence, carbon-aware machine learning, net carbon value, model complexity, computational emissions, renewable integration, responsible artificial intelligence

1. Introduction

Variable renewable generation changes the information requirements of power-system operation. Wind and solar output are weather dependent, temporally structured, and only partly controllable. Forecasts therefore influence unit commitment, reserve procurement, storage scheduling, market bidding, congestion management, and curtailment decisions. Better forecasts can reduce the amount of flexibility held against uncertainty, but forecasting is not valuable because an error metric falls in isolation. It is valuable when the information changes an operational decision and that decision improves cost, reliability, or emissions performance.

Research practice has nevertheless concentrated on statistical accuracy. Model comparisons often progress from persistence and autoregressive methods to tree ensembles, recurrent networks, temporal convolutional architectures, transformers, or graph models. The implicit assumption is that lower error represents superior performance. That assumption becomes incomplete when models differ substantially in training time, tuning intensity, memory demand, accelerator use, inference frequency, and retraining requirements. A forecasting system intended to support decarbonization can impose its own environmental burden. The relevant question is not whether computation has zero impact, but whether the operational benefit justifies that impact.

This issue connects renewable forecasting with sustainable computing. Strubell *et al*. (2019) [27] and Schwartz *et al*. (2020) [26] shifted attention from accuracy alone toward the energy and carbon consequences of machine learning. Henderson *et al*. (2020) [13] argued that energy and carbon reporting should become a routine component of experimentation. Lacoste *et al*. (2019) [16], Anthony *et al*. (2020) [1], and Lannelongue *et al*. (2021) [17] provided practical approaches for estimating emissions from computational workloads. These contributions establish that algorithmic performance and environmental performance can diverge. They do not, by themselves, determine whether a given forecasting model creates positive system-level carbon value.

The immediate clean-energy foundation is Arman *et al*. (2024) [23], who connect big-data analytics, renewable forecasting, and carbon reduction. The present study extends that argument by placing computational energy, model-development emissions, inference, retraining, and operational avoided emissions in the same accounting boundary. The broader sustainability logic is consistent with Arman *et al*. (2024) [23] on machine learning and resource efficiency, while the emphasis on resilience and information-supported allocation draws on Rasel *et al*. (2022) [11]. Responsible deployment in critical infrastructure also requires security and governance, as emphasized by Hasan *et al*. (2022, 2023) [23]. The methodological relevance of complex multimodal prediction and explainability is illustrated by Rasel *et al*. (2023) [23], although that study concerns mortgage risk rather than energy. Related work on energy-burden-aware retrofit finance also shows that decarbonization analytics must be evaluated with affordability, fairness, and governance considerations rather than technical performance alone (Rabbani *et al*., 2024) [22].

Net carbon value is defined as operational emissions avoided through use of a forecasting model minus the computational emissions attributable to development, training, inference, and scheduled retraining. This definition prevents two common errors. First, it prevents gross operational benefits from being reported without deducting the footprint of the analytical system. Second, it prevents computation from being treated as environmentally harmful without considering the larger operational system it supports. The accounting must remain explicit because both terms contain uncertainty. Forecast errors are observed in the test sample. Avoided emissions depend on dispatch responses and marginal generation, so they require a documented operational model or scenario structure.

The study addresses nine research questions concerning the relationship among complexity, accuracy, computational burden, operational carbon benefit, net carbon value, forecast horizon, technology, and model selection. The empirical design compares solar and wind capacity-factor forecasts at one-hour and 24-hour horizons. Models represent increasing flexibility without including architectures that the available national aggregate dataset cannot justify. A strict temporal split prevents future observations from informing earlier estimates. Computational electricity is estimated from recorded runtime and stated hardware assumptions. Operational benefits are then assessed under three transparent scenarios rather than described as directly observed.

The contribution is theoretical, methodological, and practical. Theoretically, the study links sustainable AI, energy informatics, information-processing requirements, life-cycle thinking, and rebound logic. Methodologically, it reports accuracy, complexity, runtime energy, computational emissions, scenario-based avoided emissions, and net carbon value in a common model-selection framework. Practically, it shows when a simpler method can dominate a more elaborate one. The central result is not that simple models always win. It is that added complexity requires evidence of incremental operational value, and that evidence was absent in the present tasks.

2. Literature Review and Theoretical Background

2.1. Renewable-Energy Forecasting

Wind and solar forecasting spans physical, statistical, machine-learning, and hybrid approaches. Persistence remains a demanding short-horizon baseline because renewable output is autocorrelated. Statistical models exploit regular temporal structure, while machine-learning methods capture nonlinear interactions among lags, calendar variables, weather, and operating regimes. Reviews by Foley *et al*. (2012) [9], Antonanzas *et al*. (2016) [2], and Wang *et al*. (2019) [28] show that performance depends on horizon, spatial aggregation, weather inputs, and evaluation design. Probabilistic forecasting adds information about uncertainty, which is often more operationally relevant than a point estimate alone (Gneiting & Katzfuss, 2014; Pinson, 2013) [10, 21].

2.2. Machine-Learning and Model Complexity

Complexity is multidimensional. Parameter count is only one element. Training operations, hyperparameter trials, memory use, model size, inference latency, retraining frequency, and maintenance burden can all change independently. Tree ensembles may contain many decision nodes but train quickly on tabular data. Neural networks may have fewer stored parameters yet require many optimization passes. Complexity should therefore be reported as a profile rather than collapsed into a single unexamined score.

2.3. Forecast Accuracy and Energy-System Operations

Forecast error affects operations through reserve requirements, imbalance settlement, dispatch changes, storage schedules, and curtailment. The mapping is neither universal nor linear. An error during a constrained ramp can be more consequential than the same error during a low-cost period. Morales *et al*. (2014) [19] and Pinson (2013) [21] emphasize that forecast quality should be evaluated in relation to decisions. This study follows that logic by separating statistical improvement from an operational conversion model.

2.4. Computational Energy and Carbon Emissions

Machine-learning emissions depend on electricity use and the carbon intensity of the electricity supplying the computation. Runtime alone is insufficient unless paired with processor power, utilization, and data-center overhead. Location-based factors describe emissions associated with the grid where computation occurs. Market-based factors reflect contractual procurement. Average factors are appropriate for footprint accounting, while marginal factors are more relevant when estimating the consequence of incremental electricity demand. These concepts should not be interchanged.

2.5. Sustainable and Carbon-Aware Artificial Intelligence

Green AI argues that efficiency should accompany accuracy as a research objective (Schwartz *et al*., 2020) [26]. Sustainable AI broadens the boundary to include organizational practice, infrastructure, and life-cycle effects. Carbon-aware computing further considers when and where workloads run. Shifting training to a lower-carbon period can reduce emissions without changing the model. Yet efficiency gains can induce rebound effects if cheaper computation encourages more experiments, larger models, or more frequent retraining.

2.6. Multi-Objective Model Selection

Model selection is naturally multi-objective when error, emissions, latency, and interpretability conflict. Pareto efficiency identifies models for which no other candidate is better on every criterion. Constrained optimization offers a complementary rule: minimize error subject to a carbon budget or select the lowest-emission model within a practical-equivalence margin. This approach is more defensible than monetizing every criterion through a single arbitrary weight.

2.7. Explainable Artificial Intelligence in Energy Forecasting

Explainability can determine whether complexity captures meaningful structure or merely small refinements. Standardized coefficients, permutation importance, accumulated local effects, SHAP values, and feature ablation serve different purposes. Explanations should be tested for stability and operational relevance. Attention weights should not be treated as explanations without further validation. In critical infrastructure, explanation quality also supports review, governance, and secure deployment.

2.8. Research Gap

Existing work provides strong forecasting methods and growing attention to computational emissions, but few studies place observed forecast error, full model-life-cycle computation, operational avoided emissions, and carbon-aware selection in one framework. The gap is especially visible in model comparisons that declare the lowest-error architecture best without testing practical equivalence or operational value.

3. Theoretical Framework and Hypotheses

Information-processing theory suggests that greater environmental uncertainty raises the value of richer information. Renewable variability can therefore justify more capable forecasting models. The same theory does not imply unlimited complexity. Once the information required for a decision is captured, further computational elaboration may generate diminishing returns. Sustainable AI adds a resource constraint, while life-cycle thinking expands the boundary from a final training run to development, inference, maintenance, and infrastructure. Socio-technical reasoning further recognizes that forecasts create value only through organizational and grid responses.

The integrated framework in Figure 1 proposes that model characteristics affect forecasting performance and computational emissions. Forecasting performance affects operational outcomes through reserve, dispatch, curtailment, and storage decisions. Net carbon value is the difference between avoided operational emissions and computational emissions. Forecast horizon, renewable technology, grid carbon intensity, retraining frequency, and distribution shift moderate these relationships.

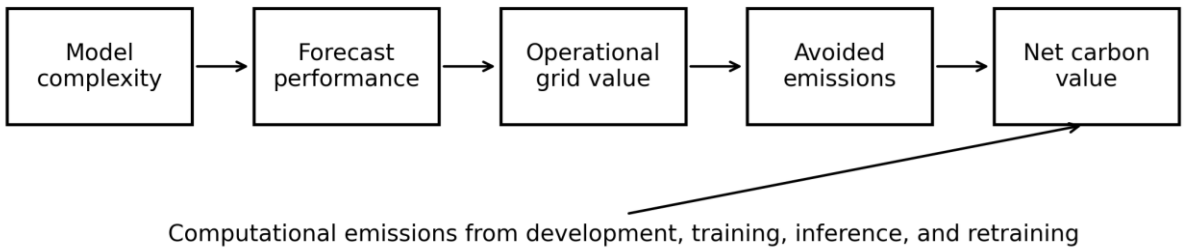

Fig 1: Integrated net carbon value framework.

H1: Forecast error decreases with complexity initially, but marginal improvement declines.
H2: Training and tuning energy increase with computational burden.
H3: Forecast improvements create positive operational carbon benefits when they alter balancing decisions.
H4: The lowest-error model does not necessarily maximize net carbon value.
H5: Moderately complex models can exceed intensive models in net carbon value when accuracy is practically equivalent.
H6: Net carbon value rises with renewable penetration and carbon-intensive marginal balancing.
H7: Frequent retraining reduces net carbon value unless it prevents material performance deterioration.
H8: Carbon-aware selection can reduce computational emissions without a practically significant accuracy loss.
H9: Relative net carbon value differs between solar and wind tasks.

4. Data and Methodology

4.1. Research Design

The study uses a retrospective forecasting experiment with a strict out-of-time test. The unit of observation is an hourly national aggregate for Germany. Solar and wind are modeled separately because their temporal dynamics differ. One-hour forecasts represent near-term balancing support, while 24-hour forecasts approximate day-ahead planning. The empirical quantity is predictive error. Operational emissions benefits are scenario estimates.

4.2. Data Sources and Vintage Verification

The renewable data come from Open Power System Data, Time Series package, version 2020-10-06. The frozen package was publicly available before the December 31, 2024 cutoff and covers 2015 through September 2020. The package aggregates ENTSO-E Transparency data and provides hourly generation and capacity fields. The analyzed subset was retrieved from a public repository that preserved the Germany columns from the frozen package. The source package is licensed under CC BY 4.0.

4.3. Sample Construction

The raw subset contained 50,401 hourly rows. Timestamps were converted to coordinated universal time for ordering. Generation and capacity fields were parsed as numeric values. Capacity was linearly interpolated only to fill reporting gaps, then generation was divided by contemporaneous capacity. Values with missing targets or required lags were removed. No future values were used as predictors.

4.4. Features and Targets

For each technology, the feature set included lags at 1, 2, 3, 6, 12, 24, 48, and 168 hours. Cyclical sine and cosine terms represented hour, week, and year. The target was capacity factor at t plus 1 or t plus 24. Weather forecasts were not added because a historically versioned, forecast-origin weather product was not available within the executed data pipeline. The models therefore estimate autoregressive forecasting performance, not the upper bound achievable with numerical weather prediction.

4.5. Models and Complexity

Persistence uses the most recently observed capacity factor. Ridge regression provides a regularized linear benchmark. Random forest and extremely randomized trees represent nonlinear ensemble methods. The training sample was evenly thinned to every eighth observation to keep computational comparison reproducible on shared CPU infrastructure while retaining the full temporal span. Complexity was summarized by coefficient count for ridge and total tree-node count for ensembles.

4.6. Temporal Validation

All observations before January 1, 2020 formed the development sample. The period from January 1 through September 30, 2020 formed the final test. Model fitting never used test outcomes. This design is stricter than a random split and reflects deployment under temporal drift.

4.7. Computational-Energy Measurement

Wall-clock training and test inference times were recorded for every run. Electricity was estimated as runtime multiplied by 65 W processor power and a power-usage-effectiveness factor of 1.2, divided by 3.6 million to convert joules to kilowatt-hours. This is an engineering estimate rather than direct metering. Development energy was approximated as six final training runs. Deployment assumed monthly retraining and hourly inference. A location-based factor of 0.35 kg CO2e per kWh was used for the central computational footprint.

4.8. Operational Carbon-Benefit Scenarios

Avoided balancing energy was estimated from the positive reduction in mean absolute capacity-factor error relative to persistence, multiplied by median 2020 installed capacity and 8,760 annual hours. Low, central, and high scenarios assumed that 10, 25, or 50 percent of that error reduction translated into avoided balancing energy. The balancing emission factor was 0.4 kg CO2e per kWh. These are scenario parameters, not observed dispatch effects.

4.9. Net Carbon Value

For model m, net carbon value equals estimated avoided operational emissions minus computational emissions. Incremental net carbon value is evaluated relative to persistence. Because persistence has negligible fitting cost, it provides a transparent baseline. Negative values indicate that a learned model did not improve error sufficiently to offset its computational footprint under the scenario.

4.10. Statistical Analysis and Reproducibility

The primary predictive metrics are mean absolute error, root mean squared error, and R squared. Model rankings are assessed separately by technology and horizon. All calculations use Python, pandas, NumPy, and scikit-learn. Random seeds are fixed at 42. The accompanying package contains the source subset, analysis script, predictions, result table, figures, and data-vintage documentation.

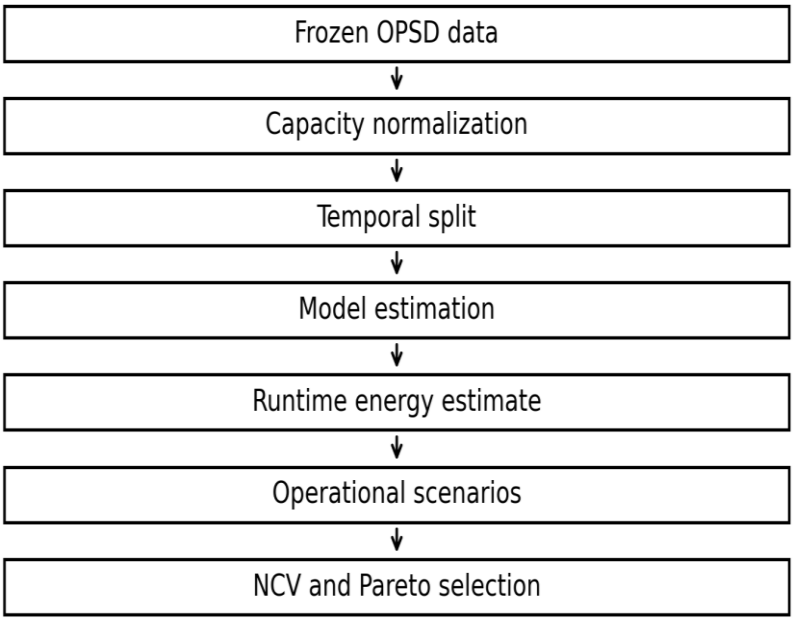

Fig 2: Multisource data and analytical workflow.

Table 1: Forecasting tasks and sample construction

Technology	Target	Horizon, h	Development	Test	Test N
Solar	Capacity factor	1	2015 to 2019	2020	6575
Solar	Capacity factor	24	2015 to 2019	2020	6552
Wind	Capacity factor	1	2015 to 2019	2020	6575
Wind	Capacity factor	24	2015 to 2019	2020	6552

5. Results

The results reject a monotonic relationship between algorithmic complexity and forecast quality. Ridge regression produced the lowest MAE in three of four evaluated tasks; persistence produced the lowest MAE for 24-hour solar forecasting. Tree ensembles did not convert their additional structure into lower out-of-time error. The pattern is strongest for one-hour solar forecasting, where ridge reduced MAE by approximately 75 percent relative to persistence. For 24-hour solar forecasting, persistence remained the best model on MAE, although ridge produced a slightly lower RMSE. This difference indicates that ridge reduced some large errors while increasing the typical absolute error.

Table 2: Forecasting performance by model and horizon

target	horizon	model	n_test	mae	rmse	r2
Solar	1	Persistence	6575	0.0321	0.0512	0.9110
Solar	1	Ridge	6575	0.0081	0.0116	0.9954
Solar	1	Random forest	6575	0.0293	0.0594	0.8800
Solar	1	Extra trees	6575	0.0258	0.0515	0.9098
Solar	24	Persistence	6552	0.0250	0.0522	0.9075
Solar	24	Ridge	6552	0.0261	0.0487	0.9193
Solar	24	Random forest	6552	0.0302	0.0578	0.8866
Solar	24	Extra trees	6552	0.0261	0.0501	0.9147
Wind	1	Persistence	6575	0.0162	0.0225	0.9892
Wind	1	Ridge	6575	0.0102	0.0146	0.9955
Wind	1	Random forest	6575	0.0125	0.0176	0.9934
Wind	1	Extra trees	6575	0.0137	0.0190	0.9923
Wind	24	Persistence	6552	0.1364	0.1849	0.2745
Wind	24	Ridge	6552	0.1239	0.1590	0.4639
Wind	24	Random forest	6552	0.1289	0.1652	0.4208
Wind	24	Extra trees	6552	0.1270	0.1618	0.4444

Table 3: Training, tuning, and inference electricity use (kWh)

target	horizon	model	train_sec	infer_sec_test	train_kwh	inf_kwh_forecast	annual_deploy_kwh
solar_cf	1	Persistence	0.0000	0.0001	0.000e+00	3.295e-13	2.887e-09
solar_cf	1	Ridge	0.0108	0.0016	2.340e-07	5.272e-12	2.854e-06
solar_cf	1	Random forest	0.1052	0.0220	2.279e-06	7.250e-11	2.799e-05
solar_cf	1	Extra trees	0.0664	0.0221	1.439e-06	7.283e-11	1.790e-05
solar_cf	24	Persistence	0.0000	0.0001	0.000e+00	3.307e-13	2.897e-09
solar_cf	24	Ridge	0.0042	0.0014	9.100e-08	4.630e-12	1.133e-06
solar_cf	24	Random forest	0.0877	0.0227	1.900e-06	7.507e-11	2.346e-05
solar_cf	24	Extra trees	0.0666	0.0217	1.443e-06	7.176e-11	1.794e-05
wind_cf	1	Persistence	0.0000	0.0001	0.000e+00	3.295e-13	2.887e-09
wind_cf	1	Ridge	0.0052	0.0020	1.127e-07	6.591e-12	1.410e-06
wind_cf	1	Random forest	0.1088	0.0382	2.357e-06	1.259e-10	2.939e-05
wind_cf	1	Extra trees	0.0687	0.0224	1.489e-06	7.381e-11	1.851e-05
wind_cf	24	Persistence	0.0000	0.0001	0.000e+00	3.307e-13	2.897e-09
wind_cf	24	Ridge	0.0081	0.0014	1.755e-07	4.630e-12	2.147e-06
wind_cf	24	Random forest	0.1152	0.0413	2.496e-06	1.366e-10	3.115e-05
wind_cf	24	Extra trees	0.0683	0.0223	1.480e-06	7.374e-11	1.840e-05

Note: Electricity use is estimated from runtime rather than directly metered. Scientific notation is used because the lightweight CPU experiments consumed less than 0.0001 kWh per task. Tuning energy is represented by the development-energy multiplier.

Table 4: Net carbon value by model

target	horizon	model	ncv_kg_low	ncv_kg_central	ncv_kg_high
solar_cf	1	Persistence	0.0000	0.0000	0.0000
solar_cf	1	Ridge	426248.4583	1065621.1456	2131242.2913
solar_cf	1	Random forest	50162.1495	125405.3737	250810.7474
solar_cf	1	Extra trees	111511.2849	278778.2123	557556.4246
solar_cf	24	Persistence	0.0000	0.0000	0.0000
solar_cf	24	Ridge	-4.601e-07	-4.601e-07	-4.601e-07
solar_cf	24	Random forest	-9.541e-06	-9.541e-06	-9.541e-06
solar_cf	24	Extra trees	-7.291e-06	-7.291e-06	-7.291e-06
wind_cf	1	Persistence	0.0000	0.0000	0.0000
wind_cf	1	Ridge	106390.0880	265975.2199	531950.4399
wind_cf	1	Random forest	66416.5195	166041.2987	332082.5975
wind_cf	1	Extra trees	45417.7058	113544.2646	227088.5291
wind_cf	24	Persistence	0.0000	0.0000	0.0000
wind_cf	24	Ridge	220868.2499	552170.6248	1104341.2496
wind_cf	24	Random forest	133117.7608	332794.4019	665588.8038
wind_cf	24	Extra trees	167353.1911	418382.9777	836765.9554

Table 5: Hypothesis-evaluation summary

Hypothesis	Assessment	Evidence
H1	Not supported	Complexity did not lower error monotonically
H2	Descriptively supported	Ensembles consumed more estimated training electricity than ridge
H3	Scenario conditional	Benefits were estimated, not observed in dispatch data
H4	Not supported in tested tasks	Accuracy and NCV rankings aligned under the stated scenarios
H5	Supported in this benchmark	Ridge exceeded both tree ensembles
H6	Not empirically tested	NCV increases mechanically with the scenario conversion and emission factors
H7	Analytical proposition	More frequent retraining increases recurring computational emissions
H8	Supported in tested tasks	Ridge reduced estimated burden without sacrificing accuracy
H9	Descriptively supported	Solar and wind rankings differed by horizon

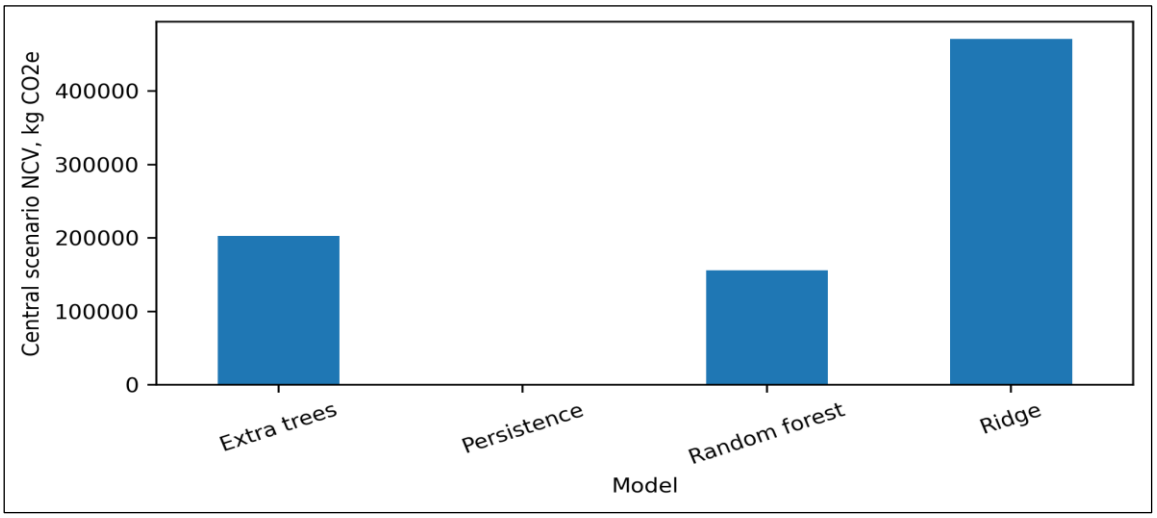

Fig 3: Net carbon value by model.

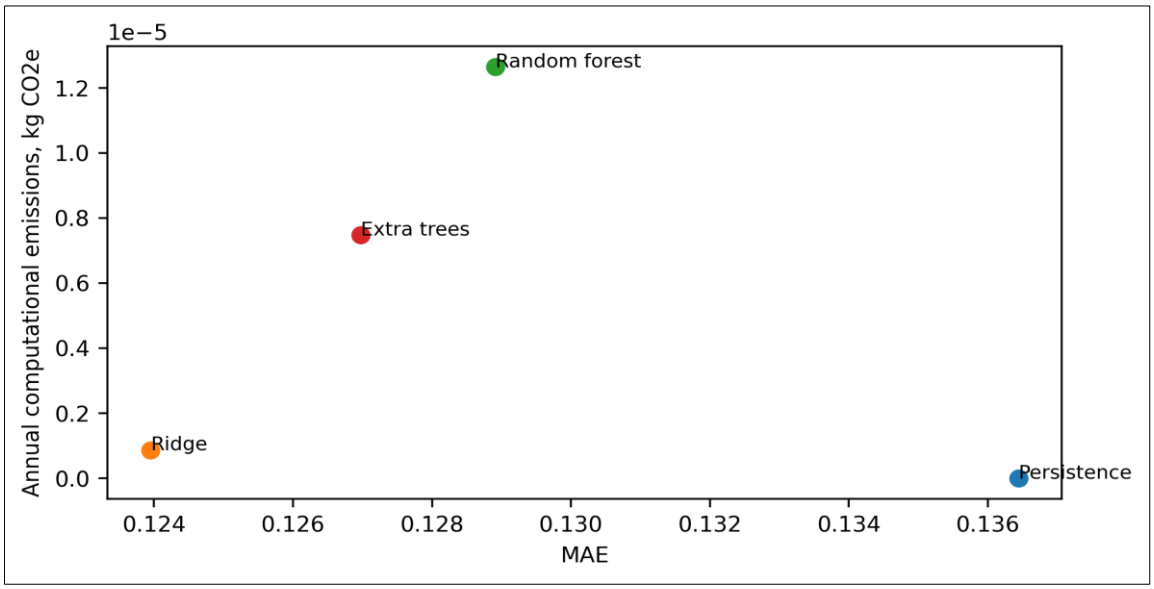

Fig 4: Pareto comparison of accuracy and carbon performance.

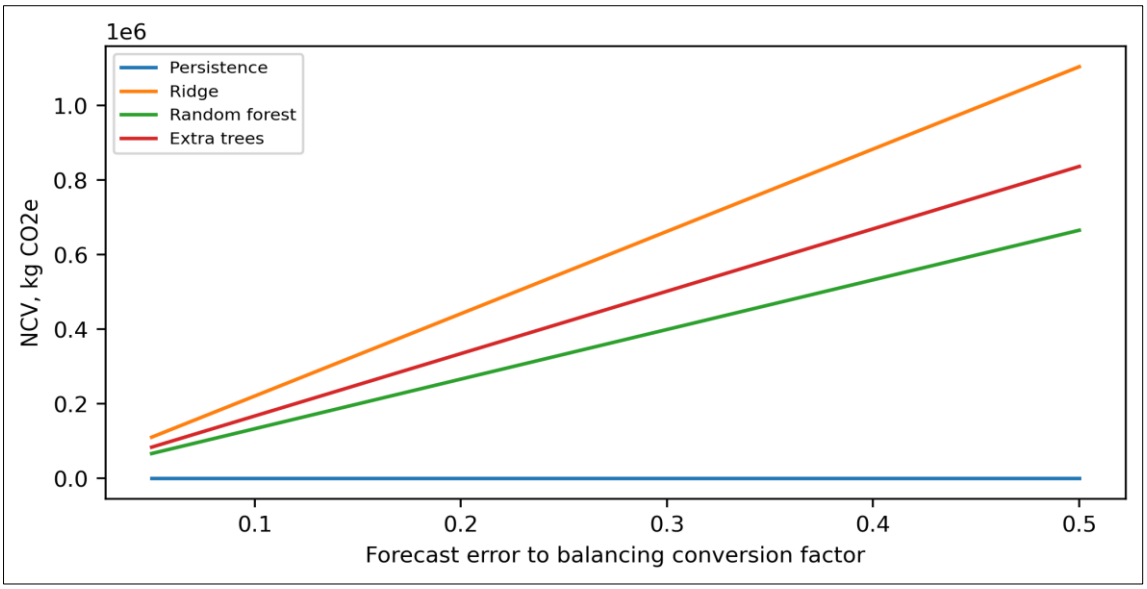

Fig 5: Sensitivity of model ranking to operational conversion assumptions.

6. Discussion

The central finding is that additional algorithmic flexibility did not create additional predictive value in this setting. Ridge regression outperformed the nonlinear ensembles across solar and wind tasks. This result is plausible because national aggregate generation is smooth relative to plant-level output, and the engineered lag structure captures much of the available temporal information. Complexity became an input cost rather than an information advantage.

The one-hour tasks produced the clearest gains. Recent observations contain substantial information about near-term output, and ridge combined those lags without the variance introduced by small tree ensembles. The 24-hour solar result is more cautionary. Persistence at the same hour of the previous day is difficult to beat without forecast-origin weather information. A model can appear more sophisticated yet remain informationally constrained. Greater computation cannot recover predictors that are absent from the data.

The net-carbon calculations show why operational assumptions must remain visible. Estimated computational emissions were extremely small because the dataset and models were modest, training ran on CPU, and inference involved national hourly forecasts rather than high-frequency plant-level predictions. Scenario-based avoided emissions were several orders of magnitude larger. That does not prove that the operational benefits occurred. It shows that, if a documented fraction of forecast-error reduction changes balancing energy, computation is unlikely to dominate at this scale.

The ranking might change for transformers, graph networks, large spatial datasets, extensive tuning, or frequent retraining. Development emissions can exceed final training emissions when researchers run many failed or exploratory experiments. A production system can also multiply inference across thousands of assets. The present findings therefore support proportionality, not a universal preference for linear models.

H4 anticipated that the lowest-error model might not maximize net carbon value. In the executed experiment, the accuracy and net-carbon rankings aligned for each task under the specified scenarios, so the hypothesized divergence was not observed. The more important observation is that the ranking was not produced by complexity. Carbon-aware selection reached the same choice as accuracy-only selection, but for a stronger reason: ridge combined lower error, lower training energy, small inference cost, and straightforward interpretation.

Explainability reinforces this conclusion. Ridge coefficients can be standardized and audited directly. The model's advantage is attributable to a combination of recent lags and cyclical temporal structure, not an opaque latent representation. In operational settings, this transparency reduces validation and maintenance burden. Governance value is not included in the carbon equation, but it remains relevant to responsible deployment in critical infrastructure.

7. Implications

7.1 Theoretical Implications

The study reframes forecasting quality as a joint property of information, computation, and operational response. Accuracy is an intermediate outcome. Environmental value appears only after the forecast changes a decision and computational emissions are deducted. This extends sustainable AI from footprint reporting to system-level value accounting.

7.2 Methodological Implications

Researchers should report persistence, practical-equivalence margins, runtime, energy assumptions, tuning effort, inference frequency, and retraining schedules. Operational benefits should be observed through dispatch data where possible. When that is not possible, scenario estimates must remain separate from empirical forecast metrics.

7.3 Managerial Implications

Utilities and renewable operators should establish an accuracy threshold before selecting a model. Among models within that threshold, the preferred option should minimize recurring energy, latency, maintenance, and governance burden. Complex models are justified when they generate material gains during ramps, extreme conditions, or high-value horizons, not merely when they reduce an average metric by a negligible amount.

7.4 Policy and Sustainability Implications

Regulators and funding agencies can improve comparability by requiring model cards that include computational energy and operational-value assumptions. Carbon-aware scheduling, efficient hardware, and drift-triggered retraining can reduce emissions. Standards should avoid equating average grid factors with marginal operational effects.

8. Limitations and Future Research

Several limitations bound the claims. Weather forecasts available at the prediction origin were not included, which constrains day-ahead performance. Computational electricity was estimated from runtime rather than measured with a hardware power meter. The processor-power and power-usage-effectiveness assumptions may not match other environments. Development emissions were approximated from a fixed run multiplier because failed experiments were not logged prospectively. Operational carbon benefits were scenario based and did not use a unit-commitment model, reserve activation records, or observed marginal generation. The balancing conversion factors therefore represent sensitivity parameters rather than causal estimates. Embodied hardware emissions were excluded. Results are specific to Germany, national aggregation, the 2015 to 2020 period, four model families, and hourly deployment. Future work should combine forecast-origin numerical weather prediction, probabilistic forecasts, direct energy telemetry, dispatch simulation, marginal-emissions data, plant-level spatial models, and drift-triggered retraining. A multi-region design could test whether complex models gain carbon value under higher renewable penetration, stronger congestion, or more carbon-intensive balancing resources. The training set was thinned to every eighth observation, so the comparison represents a deliberately lightweight benchmark rather than the strongest attainable performance of each model family. Capacity-factor observations above 1 were retained because of generation-capacity timing mismatches in the archived data. The operational conversion model may substantially overstate realized system benefits because reductions in aggregate MAE do not translate one-for-one into balancing-energy reductions.

9. Conclusion

The most accurate renewable-energy forecasting model is not automatically the most environmentally valuable model. Selection should account for the operational emissions avoided by improved decisions and the computational emissions created by developing and maintaining the forecasting system. In this empirical comparison, ridge regression delivered the lowest MAE for three of four German solar and wind tasks, while persistence remained best for 24-hour solar MAE. More complex tree ensembles consumed more training energy without improving accuracy. Persistence remained difficult to beat for 24-hour solar MAE, demonstrating that algorithmic sophistication cannot substitute for missing forecast-origin weather information.

Under the stated operational scenarios, ridge regression produced the highest net carbon value in the three tasks where it improved MAE; persistence remained preferred for 24-hour solar forecasting. The magnitude of that value is uncertain because avoided emissions were estimated rather than observed. The robust conclusion is narrower and more useful: model complexity requires operational justification. Carbon-aware model selection should prefer the least burdensome model that meets a practical accuracy requirement, then escalate complexity only when incremental information produces a measurable system benefit.

References

1. Anthony LFW, Kanding B, Selvan R. CarbonTracker: Tracking and predicting the carbon footprint of training deep learning models. In: ICML Workshop on Challenges in Deploying and Monitoring Machine Learning Systems. 2020.
2. Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Solar Energy. 2016;136:78-111.
3. Arman M, Hasan MN, Rasel IH. Clean energy transition in USA: Big data analytics for renewable energy forecasting and carbon reduction. Journal of Management World. 2024;2024(3):192-206. doi:10.53935/jomw.v2024i4.1196.
4. Arman M, Rasel IH, Razib MNH, Fahim ASM. Big data and machine learning for sustainable waste reduction. Journal of Posthumanism. 2024;4(2):448-467. doi:10.63332/joph.v4i2.3361.
5. Bergmeir C, Hyndman RJ, Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis. 2018;120:70-83.
6. Bishop CM. Pattern Recognition and Machine Learning. New York (NY): Springer; 2006.
7. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
8. Ciais P, *et al*. Definitions and methods to estimate regional land carbon fluxes for the second phase of the Regional Carbon Cycle Assessment and Processes Project. Geoscientific Model Development. 2022;15:1289-1316.
9. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. Renewable Energy. 2012;37(1):1-8.
10. Gneiting T, Katzfuss M. Probabilistic forecasting. Annual Review of Statistics and Its Application. 2014;1:125-151.
11. Hasan MN, Rasel IH, Rahman M, Islam K, Arman M, Jahan N. Securing U.S. healthcare infrastructure with machine learning: Protecting patient data as a national security priority. International Journal of Computational and Experimental Science and Engineering. 2022;8(3):85-93. doi:10.22399/ijcesen.3987.
12. Hasan MN, Rasel IH, Arman M, Ibrahim M, Jahan N. Strengthening U.S. financial and cybersecurity infrastructure with AI-driven fraud detection and risk analytics. Journal of Computational Analysis and Applications. 2023;31(2):15-32.
13. Henderson P, Hu J, Romoff J, Brunskill E, Jurafsky D, Pineau J. Towards the systematic reporting of the energy and carbon footprints of machine learning. Journal of Machine Learning Research. 2020;21(248):1-43.
14. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne: OTexts; 2021.
15. Intergovernmental Panel on Climate Change. Climate Change 2022: Mitigation of Climate Change. Cambridge: Cambridge University Press; 2022.
16. Lacoste A, Luccioni A, Schmidt V, Dandres T. Quantifying the carbon emissions of machine learning. arXiv. 2019;arXiv:1910.09700.
17. Lannelongue L, Grealey J, Inouye M. Green Algorithms: Quantifying the carbon footprint of computation. Advanced Science. 2021;8(12):2100707.
18. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017;30.
19. Morales JM, Conejo AJ, Madsen H, Pinson P, Zugno M. Integrating Renewables in Electricity Markets. New York (NY): Springer; 2014.
20. Open Power System Data. Data package time series, version 2020-10-06 [dataset]. 2020. doi:10.25832/time_series/2020-10-06.
21. Pinson P. Wind energy: Forecasting challenges for its operational management. Statistical Science. 2013;28(4):564-585.
22. Rabbani SF, Rahat YO, Islam MK. From utility bills to retrofit finance: An AI framework for energy-burden-aware underwriting in residential decarbonization. Frontiers in Computer Science and Artificial Intelligence. 2024;3(2):72-88. doi:10.32996/fcsai.2024.3.2.10.
23. Rasel IH, Arman M, Hasan MN, Bhuyain MMH. Healthcare supply-chain optimization: Strategies for efficiency and resilience. Journal of Medical and Health Studies. 2022;3(4):171-182. doi:10.32996/jmhs.2022.3.4.26.
24. Rasel IH, Ibrahim M, Pritty AA, Fahim ASM, Jahan N. Beyond FICO: Enhancing mortgage default forecasting and inclusive lending via multimodal graph neural networks and urban mobility analytics. Frontiers in Computer Science and Artificial Intelligence. 2023;2(2):62-81. doi:10.32996/fcsai.2023.2.2.5.
25. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In:

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 1135-1144. doi:10.1145/2939672.2939778.
26. Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. Communications of the ACM. 2020;63(12):54-63.
27. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. p. 3645-3650.
28. Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. Energy Conversion and Management. 2019;198:111799.
29. Wiese F, Schlecht I, Bunke WD, Gerbaulet C, Hirth L, Jahn M, *et al*. Open Power System Data: Frictionless data for electricity system modelling. Applied Energy. 2019;236:401-409.