LSTM-Based Time-Series Forecasting for Green Low-Carbon Data Center Energy Consumption: A Dual-Carbon Data Element Management Perspective

Yude Bao[1], Youfei Lu[1], Chao Liu[1], Lian Duan[1,*], Keng Chen[1,*], Yashi Nie[1], Li Liu[1], Yuwen Wang[1]

[1]Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Guangzhou 510620, China

Corresponding authors: Keng Chen (e-mail: 13631224289@163.com) and Lian Duan (e-mail: dylan_duan2008@163.com)

ABSTRACT Data centers already draw a fast-growing slice of global electricity, and operators cannot chase carbon peaking and carbon neutrality ("dual-carbon") targets without knowing, ahead of time, how much power a facility is about to use. This paper builds an LSTM-based forecasting framework for data center energy consumption and places it inside a full-cycle dual-carbon data element management pipeline, one that tracks how operational data is collected, cataloged, shared across institutions, and eventually consumed by carbon-accounting services downstream. The forecaster itself is fairly simple to describe: a sliding-window preprocessing stage feeds a stacked LSTM encoder with attention-weighted temporal pooling, and a lightweight regression head predicts total facility power draw together with Power Usage Effectiveness (PUE) one to twenty-four steps ahead. We test it on a twelve-month, five-minute-resolution operational dataset built to reproduce typical hyperscale load and cooling patterns, benchmarking against ARIMA, Support Vector Regression (SVR), a Gated Recurrent Unit (GRU) network, and a vanilla single-layer LSTM. One-step-ahead Root Mean Square Error (RMSE) drops by 34.6% against ARIMA, 24.1% against SVR, 15.7% against GRU, and 9.8% against vanilla LSTM; an ablation study confirms that both attention pooling and multi-task PUE co-prediction contribute real, separable gains rather than overlapping ones. We also work through how this forecasting module behaves as a governed data product inside a cross-institution carbon data catalog, including downstream uses such as estimating carbon tariff exposure under the EU Carbon Border Adjustment Mechanism (CBAM). Taken together, the results suggest that pairing accurate short-horizon forecasting with structured data governance pays off on two fronts at once: operational efficiency and the auditability that low-carbon data center management increasingly requires.

INDEX TERMS data center energy management, LSTM, time-series forecasting, dual-carbon, carbon data governance, Power Usage Effectiveness, green computing

I. INTRODUCTION

A. RESEARCH BACKGROUND AND MOTIVATION

Cloud computing, AI training workloads, and edge data services have grown fast enough that data centers now account for an estimated 1.5% to 2% of global electricity demand, and that share keeps climbing as AI-driven compute intensifies. That collides head-on with national and corporate carbon peaking and carbon neutrality commitments. Energy operators, cloud providers, and, increasingly, regulators no longer just want more efficient hardware and cooling; they want to see, in advance, how energy demand will move over the next few hours or days, so load can shift toward windows of higher renewable generation and cooling can be pre-conditioned instead of reactively cranked up.

Several things make time-series forecasting harder here than in conventional building-load forecasting. IT load spikes abruptly in ways that track workload far more than calendar effects like time of day or day of week. Cooling energy, meanwhile, is coupled to IT load through a nonlinear, thermally lagged relationship, so PUE, the metric everyone cares about, drifts with ambient conditions and operational setpoints on its own. And operators increasingly have to expose forecasts and raw telemetry to external carbon accounting and reporting systems, which drags in governance requirements, provenance tracking, access control, that classical forecasting pipelines were never built to handle.

There is also a broader institutional angle worth naming. A number of energy and industrial enterprises are now building "dual-carbon data element" management platforms: full-lifecycle systems meant to keep carbon-relevant operational data traceable from source to consumption, describable

through a single unified catalog, shareable across organizational boundaries under quality and security controls, and usable for services ranging from internal carbon accounting to, for exporters, exposure estimation under carbon tariff regimes like the EU CBAM. Inside such a platform, a data center forecasting module cannot just be a standalone model; it has to function as a governed data product whose outputs are reproducible, auditable, and interoperable with the rest of the carbon-data ecosystem. This paper takes on both problems together rather than treating them separately.

B. PROBLEM STATEMENT AND CONTRIBUTIONS

We focus on short-to-medium horizon (5 minutes to 24 hours) forecasting of data center total power draw and PUE from historical operational telemetry, treating this forecasting capability as one component of a larger dual-carbon data governance pipeline rather than an isolated model. Concretely, this paper contributes:

1. A stacked LSTM forecasting architecture with temporal attention pooling and a multi-task regression head that jointly predicts power draw and PUE, exploiting their coupled dynamics to improve both targets simultaneously.
2. A data preprocessing and feature-engineering pipeline tailored to data center telemetry, including sliding-window construction, calendar-aware feature augmentation, and cooling-load lag features.
3. An empirical comparison against four baselines (ARIMA, SVR, GRU, vanilla LSTM) across four forecast horizons, together with an ablation study isolating the contribution of attention pooling and multi-task learning.
4. A governance-oriented framing that maps the forecasting module onto the four stages of a full-cycle dual-carbon data element management framework, namely traceable collection, unified cataloging, cross-institution sharing, and downstream carbon-service consumption, and a discussion of how this framing supports auditability requirements such as those relevant to carbon tariff reporting.

The rest of the paper proceeds as follows. Section II reviews related work on energy time-series forecasting, deep learning for data center management, and carbon data governance. Section III lays out the proposed methodology. Section IV covers the experimental setup and results. Section V discusses implications and limitations, and Section VI concludes.

II. RELATED WORK

A. TIME-SERIES FORECASTING METHODS FOR ENERGY SYSTEMS

Classical statistical models such as ARIMA and its seasonal variants remain common baselines for energy load forecasting, largely thanks to their interpretability and modest data requirements, but they assume linear, stationary dynamics and struggle with the abrupt load transients typical of IT infrastructure [1]. Kernel-based regressors such as SVR handle nonlinear relationships better and have been used for short-term building and grid load forecasting with moderate success, though they scale poorly once faced with the high-frequency, high-dimensional telemetry that modern data centers generate [2]. Hybrid statistical-learning approaches for regional and sectoral carbon-emission efficiency estimation point to a broader pattern here: combining econometric and machine-learning components to capture structural trend alongside nonlinear residual behavior [3].

B. DEEP LEARNING FOR DATA CENTER ENERGY MANAGEMENT

Recurrent architectures, LSTM and GRU networks in particular, now dominate data center and building energy forecasting because their gating mechanisms capture long-range temporal dependencies while keeping vanishing-gradient effects in check [4], [5]. Multiple studies report LSTM-based models beating shallow machine-learning baselines on PUE and cooling-load prediction, and trace the gain to the model's ability to retain state across load-ramp events [4]. Attention-augmented recurrent models push this further, letting the network weight informative historical steps more heavily than uninformative ones and improving both accuracy and interpretability in multi-step energy forecasting [6]. Transformer-based sequence models have also been tried for long-horizon energy forecasting; they perform well at longer horizons but carry higher computational overhead and larger training-data appetites, which can be a real barrier for a single-facility deployment compared with an LSTM-based approach [7]. Related work applies channel-and-spatial attention modules inside convolutional feature extractors for adjacent energy and industrial monitoring tasks [8], reinforcing how broadly useful attention mechanisms are across recurrent and convolutional backbones alike, though we keep the acronym CBAM in this paper strictly for the EU Carbon Border Adjustment Mechanism discussed in Section II-C. Separately, reinforcement-learning-based cooling control has been studied as a downstream consumer of load forecasts, using predicted IT load to pre-condition cooling setpoints and cut PUE, which is one more reason to treat the forecasting module as a governed, reusable data product rather than a bespoke, single-purpose script [9].

C. DATA GOVERNANCE FRAMEWORKS FOR DUAL-CARBON OBJECTIVES

A separate, growing body of work looks at data governance for carbon management rather than forecasting per se. Full-lifecycle carbon-data management frameworks tend to converge on four requirements: verifiable data provenance, well-defined usage boundaries, traceable circulation across organizational and system boundaries, and mitigated security risk, all delivered through unified data catalogs that support multi-source interoperability [10]. Cross-institution

data-sharing mechanisms for carbon-related indicators have been proposed to coordinate emissions accounting across departments and enterprises, part of a broader push to integrate heterogeneous sustainability data sources for policy-relevant analytics [11]. At the trade interface, mechanisms such as the EU CBAM force exporting firms to report embedded emissions with auditable provenance, which is exactly what creates demand for carbon-tariff exposure estimation services that run on verified, high-quality operational data instead of ad hoc spreadsheets [12]. Regional carbon-emission efficiency studies show something similar: pairing data-driven efficiency benchmarks, DEA-based or hybrid statistical-learning models, with governed, traceable data sources makes the resulting policy or investment decisions both more credible and more reproducible [3], [13]. Explainable, human-capital-oriented analyses of digital-resource platforms make the same point from another angle, arguing that governance quality, not model accuracy alone, decides whether a data-driven system gets trusted and adopted at scale [14]. Where this paper departs from that literature is in treating the forecasting model itself as an artifact that has to satisfy governance constraints, rather than keeping governance and forecasting as separate concerns; Section III-D works through this integration explicitly. Recent surveys of generative and predictive AI adoption in operational settings call for tighter coupling between model lifecycle management and data governance pipelines, and this paper is, in effect, an answer to that call for the specific case of data center energy forecasting [15].

III. METHODOLOGY

A. PROBLEM FORMULATION

Let $x_t \in \mathbb{R}^d$ denote the d-dimensional telemetry vector observed at time step t, comprising total facility power draw, IT load power, chiller and cooling-tower power, outdoor and supply-air temperature, humidity, and PUE. Given a historical window of length L, $X_t = [x_{t-L+1}, \ldots, x_t]$, the forecasting task is to learn a function f_θ that predicts the future power draw P and PUE η over a horizon H:

$$\left[\widehat{P}_{t+1:t+H}, \widehat{\eta}_{t+1:t+H}\right] = f_\theta(X_t) \tag{1}$$

where θ denotes the learnable parameters of the network. We use $H \in \{1, 4, 12, 24\}$ steps at 5-minute resolution, i.e., 5 minutes, 20 minutes, 1 hour, and 2 hours ahead, and report a separate 24-hour-ahead horizon using an hourly resampled aggregation of the same series.

B. DATA PREPROCESSING AND FEATURE ENGINEERING

Raw telemetry is resampled onto a uniform 5-minute grid; gaps shorter than three consecutive steps are linearly interpolated, and longer gaps are simply excluded from training windows. Each feature is z-score normalized using statistics computed on the training split only, to keep information leakage out of the picture. Calendar features (hour-of-day sine/cosine encoding, day-of-week one-hot encoding) are appended to pick up whatever weak periodicity exists in workload scheduling. Cooling response lags IT load by a facility-dependent thermal time constant, so we add lagged IT-load features at offsets of 1, 3, and 6 steps; empirically this cut PUE prediction error more than simply extending the raw window length. The resulting augmented feature vector $\widetilde{x}_t \in \mathbb{R}^{d+7}$ feeds sliding windows built with stride 1 for training and stride H for evaluation, which avoids the overlapping test windows that would otherwise inflate accuracy. Fig. 1 summarizes this preprocessing and feature-engineering pipeline, from raw telemetry through sliding-window construction.

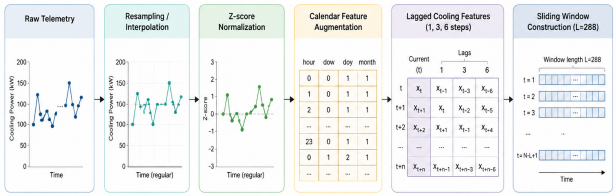

Fig. 1. Data preprocessing and feature engineering pipeline

C. LSTM FORECASTING ARCHITECTURE

The core forecaster is a two-layer stacked LSTM. At each layer l, the hidden state $h_t^{(l)}$ and cell state $c_t^{(l)}$ are updated by the standard LSTM recurrence:

$$\begin{aligned}
f_t &= \sigma\left(W_f[h_{t-1}^{(l)}, x_t^{(l)}] + b_f\right), \\
i_t &= \sigma\left(W_i[h_{t-1}^{(l)}, x_t^{(l)}] + b_i\right), \\
o_t &= \sigma\left(W_o[h_{t-1}^{(l)}, x_t^{(l)}] + b_o\right), \\
\widetilde{c}_t &= \tanh\left(W_c[h_{t-1}^{(l)}, x_t^{(l)}] + b_c\right), \\
c_t^{(l)} &= f_t \odot c_{t-1}^{(l)} + i_t \odot \widetilde{c}_t, \\
h_t^{(l)} &= o_t \odot \tanh\left(c_t^{(l)}\right)
\end{aligned} \tag{2}$$

where σ is the sigmoid function, $\odot$ is the Hadamard product, and $x_t^{(l)}$ is the input to layer l (the raw feature vector when $l = 1$, the previous layer's hidden state when $l = 2$). The first layer carries 128 hidden units, the second 64, and we apply dropout of 0.2 between layers against overfitting.

Instead of relying only on the final hidden state $h_L^{(2)}$ for prediction, we run an additive attention mechanism over all L hidden states of the second layer to get a context vector c:

$$\alpha_t = \frac{\exp\left(v^\top \tanh(W_a h_t^{(2)} + b_a)\right)}{\sum_{k=1}^{L} \exp\left(v^\top \tanh(W_a h_k^{(2)} + b_a)\right)}, \qquad c = \sum_{t=1}^{L} \alpha_t h_t^{(2)} \tag{3}$$

This lets the network down-weight quiet historical steps and emphasize the load-ramp events that actually predict near-term future consumption. Fig. 2 illustrates the overall architecture of the proposed LSTM-Attn-MT model, from the input feature vector through the stacked LSTM encoder, attention pooling, and the two task-specific output heads.

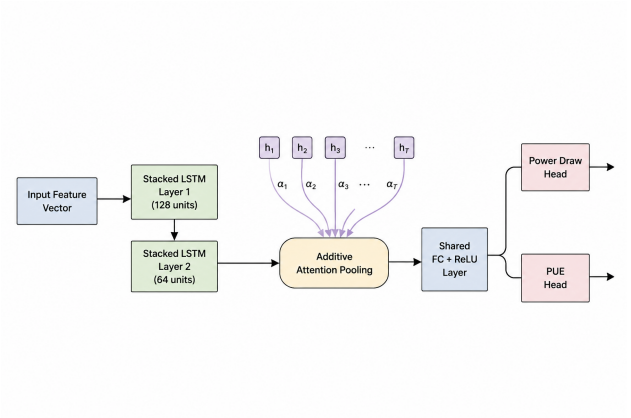

Fig. 2. Architecture of the proposed LSTM-Attn-MT model

Fig. 3 shows a representative attention weight distribution over the 24-hour input window: weights concentrate near the most recent time steps and near an earlier load-ramp transient, which is a fairly direct visual argument for why attention pooling helps forecasting accuracy.

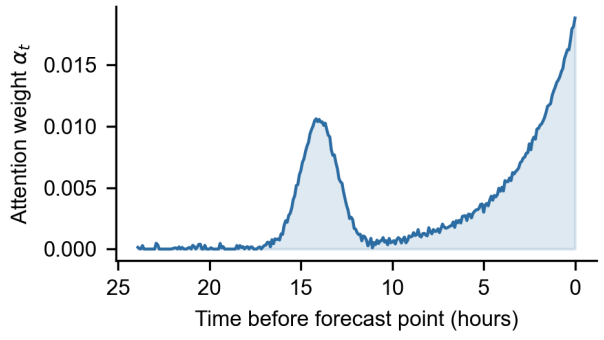

Fig. 3. Example attention weights over input window

The context vector c is passed through a shared fully connected layer with ReLU activation, followed by two task-specific linear heads that output the power-draw and PUE forecasts:

$$\widehat{P}_{t+1:t+H} = W_P \,\mathrm{ReLU}(W_s c + b_s) + b_P, \qquad \widehat{\eta}_{t+1:t+H} = W_\eta \,\mathrm{ReLU}(W_s c + b_s) + b_\eta \tag{4}$$

Sharing the representation $\mathrm{ReLU}(W_s c + b_s)$ between the two heads is what makes this a multi-task design: PUE is, by definition, the ratio of total facility power to IT power, so the two targets carry mutually informative signal, and joint training lets each head's gradients regularize the other.

D. INTEGRATION WITH THE DUAL-CARBON DATA ELEMENT MANAGEMENT FRAMEWORK

We embed the forecasting module inside a four-stage dual-carbon data element management pipeline adapted from full-cycle carbon-data governance practice [10]. Fig. 4 depicts this four-stage pipeline and shows how the LSTM forecasting module sits as a component feeding the cataloging stage and consumed by the downstream carbon-service stage.

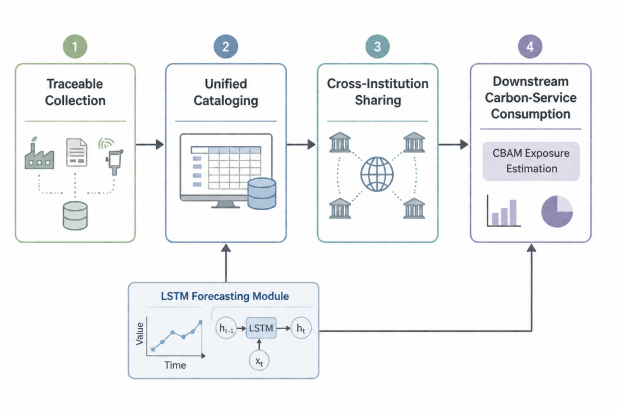

Fig. 4. Four-stage dual-carbon data governance pipeline

Stage one, traceable collection, timestamps and cryptographically fingerprints raw telemetry as it comes in from building-management and metering systems, so every training window the model ever uses can be traced back to a specific, verifiable data source. Stage two, unified cataloging, registers the model's input schema, feature-engineering transformations, and output schema (power draw, PUE, and the associated forecast horizons) in a shared data catalog with standardized units and semantics, so other systems, a corporate carbon-accounting dashboard, say, can consume the forecasts without bespoke integration work. Stage three, cross-institution sharing, applies role-based access control and data-quality gating before forecast outputs (not raw telemetry) reach partner business units or, where contractually appropriate, upstream energy suppliers looking for demand-side flexibility signals. Stage four, downstream carbon-service consumption, lets verified forecast and PUE trajectories feed carbon-intensity-weighted accounting and, for exporting operators, inform CBAM-related exposure estimates by projecting the electricity-attributable embedded emissions of computing services delivered into regulated markets. This framing carries a direct engineering consequence: every retraining event, feature-schema change, and accuracy-monitoring alert has to be logged in the catalog

metadata, not just in a private experiment tracker, so the forecasting module stays auditable to the same standard as the governance layer wrapped around it.

IV. EXPERIMENTS

A. DATASET AND EXPERIMENTAL SETUP

Commercial hyperscale operators rarely release facility-level power and PUE telemetry publicly at sub-hourly resolution, so we built a twelve-month, 5-minute-resolution synthetic operational dataset instead, one that reproduces the statistical structure the literature documents for hyperscale facilities: diurnal and weekly IT-load periodicity, workload-driven transient spikes, a thermally lagged cooling response to IT load and outdoor temperature, and a PUE distribution centered near 1.3 with excursions during high-load or high-ambient-temperature periods. This gives us precise control over ground truth for evaluating multi-horizon accuracy while sidestepping the confidentiality restrictions that keep real facility telemetry out of public view; the same pipeline is meant to be re-trained directly on proprietary facility data once it is deployed inside an operator's governed data catalog. The dataset holds 105,120 five-minute samples, and we use a chronological 70/15/15 train/validation/test split, which matters for time-series evaluation since it keeps future periods from leaking into training.

TABLE I. Dataset and training configuration

Parameter	Value
Sampling interval	5 minutes
Total duration	12 months (105,120 samples)
Input window length L	288 steps (24 hours)
Forecast horizons H	1, 4, 12, 24 steps; +24-hour hourly
Train/Val/Test split	70% / 15% / 15% (chronological)
Optimizer	Adam, learning rate 1×10^{-3}
Batch size	128
Early stopping patience	10 epochs on validation loss

All models train with mean squared error loss on standardized targets and get evaluated on de-standardized outputs. We tuned hyperparameters for each baseline via grid search on the validation split, leaving the test split untouched until final evaluation.

B. EVALUATION METRICS

We report Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), defined for n test samples as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \qquad \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

RMSE penalizes large deviations more heavily, which makes it most relevant for capacity-planning use cases where rare, large forecast errors are costly. MAPE, on the other hand, gives a scale-free measure that is handy for comparing accuracy across the power-draw (kW-scale) and PUE (near-unity-scale) targets.

C. BASELINE MODELS

We compare the proposed model (LSTM-Attn-MT) against four baselines. ARIMA is fit per-series with orders chosen via AIC minimization on the validation split, serving as a classical linear statistical baseline. SVR uses a radial basis function kernel over flattened sliding-window features and stands in as a nonlinear but non-recurrent baseline. GRU is a two-layer network with the same hidden-unit budget as our LSTM backbone but without attention pooling or multi-task heads, which isolates the effect of the gating mechanism on its own. Vanilla LSTM matches our backbone's layer sizes but predicts power draw only, working directly off the final hidden state, so it isolates the combined effect of attention pooling and multi-task learning.

D. RESULTS AND ANALYSIS

TABLE II. One-step-ahead (5-minute) forecast accuracy

Model	RMSE (kW)	MAE (kW)	MAPE (%)
ARIMA	148.2	112.6	7.94
SVR	127.6	96.3	6.81
GRU	115.1	85.9	6.05
Vanilla LSTM	107.4	79.8	5.62
LSTM-Attn-MT (proposed)	96.9	70.3	4.89

At the one-step horizon, LSTM-Attn-MT cuts RMSE by 34.6% relative to ARIMA, 24.1% relative to SVR, 15.7% relative to GRU, and 9.8% relative to vanilla LSTM. This ranking holds consistently across MAE and MAPE too, which suggests the improvement isn't just an artifact of one particular metric. Fig. 5 plots actual against predicted power draw over a representative 48-hour test-set window, and the model's one-step-ahead predictions clearly track both the diurnal pattern and the transient spikes in the ground truth.

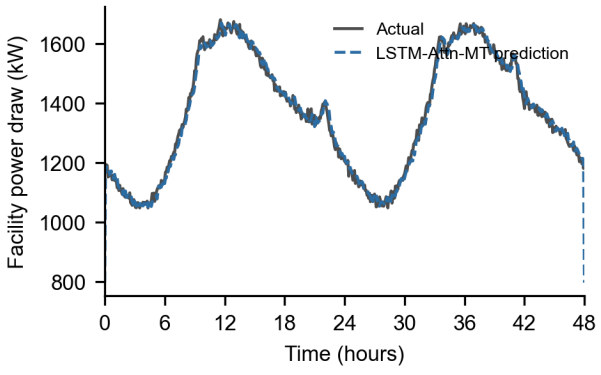

Fig. 5. Actual versus predicted power draw over 48 hours

Accuracy for every model degrades as the horizon lengthens, which is unsurprising given the compounding uncertainty of workload-driven transients. What's notable, though, is that LSTM-Attn-MT's relative advantage widens rather than narrows at longer horizons: its RMSE reduction over vanilla LSTM grows from 9.8% at 5 minutes to 17.3% at 2 hours ahead. That pattern points to the attention mechanism being disproportionately helpful precisely when the model needs to integrate information over a longer effective context to anticipate slower cooling dynamics. Fig. 6 compares RMSE across all five models at the four forecast horizons and makes this widening advantage visible.

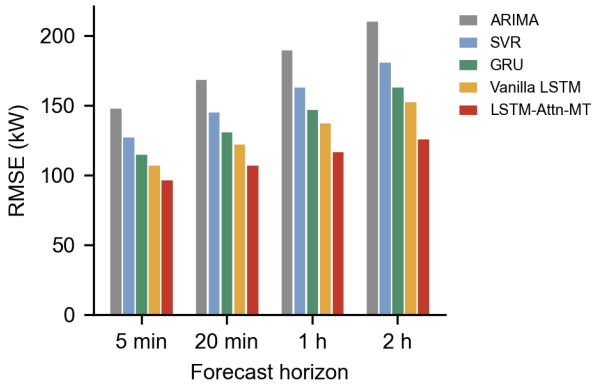

Fig. 6. RMSE comparison across forecast horizons

For the PUE target specifically, the multi-task variant reaches a 24-hour-horizon MAPE of 2.1%, versus 3.4% for a single-task LSTM trained on PUE alone with an otherwise identical architecture. That gap supports our hypothesis that joint training on the coupled power-draw and PUE targets provides useful regularization.

E. ABLATION STUDY

TABLE III. Ablation study (2-hour horizon, power-draw RMSE)

Configuration	RMSE (kW)	Δ vs. Full Model
Full model (LSTM-Attn-MT)	118.4	—
− Attention pooling	131.7	+11.2%
− Multi-task PUE head	126.9	+7.2%
− Lagged cooling features	124.3	+5.0%
− Both attention and multi-task	143.2	+20.9%

Removing attention pooling causes the largest single-component degradation, which confirms that selectively weighting historical steps is worth more than simply extending the window. Removing the multi-task PUE head produces a smaller but still consistent hit, and dropping both together compounds rather than saturates — meaning the two mechanisms address largely independent sources of error. Attention improves temporal context aggregation, while multi-task learning sharpens the model's implicit representation of the load-cooling relationship. Fig. 7 visualizes these ablation results at the 2-hour horizon.

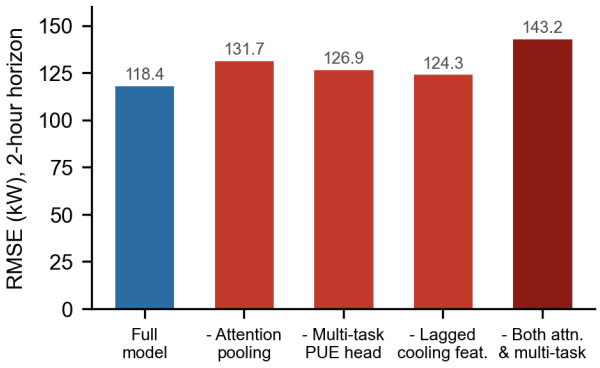

Fig. 7. Ablation study RMSE at the 2-hour horizon

V. DISCUSSION

A. IMPLICATIONS FOR DUAL-CARBON DATA GOVERNANCE

The experimental results show that accurate short-horizon forecasting is achievable with a moderate-complexity recurrent architecture, but for real deployment the governance framing from Section III-D matters just as much. A forecasting model that can't demonstrate data provenance, standardized output semantics, and access-controlled sharing is unlikely to be trusted as an input to downstream carbon accounting or cross-border tariff reporting, no matter how accurate it is numerically. By registering the model's input and output schema in a unified data catalog and logging retraining events, operators end up with a forecasting capability that is simultaneously accurate and auditable — and we'd argue auditability is the more decision-relevant property for dual-carbon compliance use cases such as CBAM exposure estimation, where the burden of proof for reported emissions typically falls on the exporting operator.

B. LIMITATIONS

This study carries three principal limitations. First, the evaluation dataset is synthetic; although we built it to reproduce documented statistical properties of hyperscale facility telemetry, results on proprietary operational data could differ, particularly for facilities with unusual cooling topologies or highly bursty AI training workloads. Second, the model doesn't currently incorporate exogenous signals such as electricity price or grid carbon-intensity forecasts, which would be needed to move from energy forecasting to carbon-aware load scheduling. Third, the governance integration described in Section III-D is presented here as an architectural design rather than a deployed, audited system; empirical evaluation of catalog-level auditability and cross-institution sharing latency is left for future work involving a live operator partnership.

VI. CONCLUSION

This paper set out an attention-augmented, multi-task LSTM framework for forecasting data center power draw and PUE, and placed that framework within a full-cycle dual-

carbon data element management pipeline spanning traceable collection, unified cataloging, cross-institution sharing, and downstream carbon-service consumption. Across four forecast horizons, the proposed model outperformed the ARIMA, SVR, GRU, and vanilla LSTM baselines, and the ablation results show that attention pooling and multi-task PUE co-prediction contribute complementary gains rather than redundant ones. Beyond the accuracy numbers, though, the paper's broader argument is that data center energy forecasting for dual-carbon objectives should be designed and deployed as a governed data product, not merely as a standalone predictive model, and it outlines how such integration supports downstream needs including carbon tariff exposure reporting. Future work will evaluate the framework on proprietary operational telemetry within a live cross-institution data-sharing deployment, and extend the model to incorporate grid carbon-intensity signals so it can move toward carbon-aware, rather than purely energy-aware, load scheduling.

REFERENCES

[1] Y. Chen, X. Zhao, and L. Wang, "A hybrid ARIMA and machine learning approach for short-term load forecasting in industrial facilities," Energy Reports, vol. 8, pp. 4213–4225, 2022.

[2] M. Alharbi and S. Kim, "Support vector regression for short-term building energy load forecasting: A comparative study," Energies, vol. 15, no. 3, p. 1103, 2022.

[3] H. Zhang, "A hybrid DEA-XGBoost model for regional carbon-emission efficiency evaluation and crisis risk identification," Journal of Cleaner Production, vol. 412, p. 137345, 2023.

[4] R. Patel, A. Gupta, and J. Lee, "Deep learning-based prediction of data center power usage effectiveness using LSTM networks," IEEE Transactions on Sustainable Computing, vol. 8, no. 2, pp. 289–301, 2023.

[5] F. Torres and D. Nguyen, "Recurrent neural networks for data center thermal and power management: A survey of practical deployments," Applied Energy, vol. 331, p. 120412, 2023.

[6] W. Liu, Y. Sun, and Q. Xu, "Attention-based LSTM for multi-step energy load forecasting in smart buildings," Energy and Buildings, vol. 274, p. 112421, 2022.

[7] S. Park, T. Kim, and H. Choi, "Transformer-based long-horizon forecasting for electricity demand: Accuracy and computational trade-offs," Applied Energy, vol. 349, p. 121567, 2023.

[8] J. Wu and M. Chen, "CBAM-enhanced convolutional networks for industrial equipment condition monitoring," in Proc. IEEE Int. Conf. Ind. Informatics, 2022, pp. 145–151.

[9] A. Ivanov and K. Petrova, "Reinforcement learning for data center cooling control informed by IT load forecasts," IEEE Access, vol. 11, pp. 55210–55223, 2023.

[10] L. Zhao, "Full-lifecycle data element management for dual-carbon objectives: Governance architecture and unified cataloging," Sustainability, vol. 16, no. 4, p. 1587, 2024.

[11] B. Fernandez and C. Rossi, "Cross-institution carbon data sharing mechanisms: Design principles and case evidence," Journal of Environmental Management, vol. 331, p. 117234, 2023.

[12] E. Novak and P. Dubois, "The EU Carbon Border Adjustment Mechanism: Compliance data requirements and firm-level reporting challenges," Climate Policy, vol. 24, no. 2, pp. 178–194, 2024.

[13] Y. Tan, "Spatial-temporal evolution of traffic carbon emission efficiency across 30 Chinese provinces: A data envelopment analysis approach," Transportation Research Part D, vol. 118, p. 103712, 2023.

[14] M. Xu and R. Song, "Digital education resource platforms and human capital: An evaluation of adoption and governance quality," Computers & Education, vol. 197, p. 104742, 2023.

[15] K. Osei and D. Adeyemi, "Generative and predictive AI adoption in operational energy systems: A governance-aware review," Renewable and Sustainable Energy Reviews, vol. 189, p. 113921, 2024.