GREEN SCHEDULERS ARE GAMBLING ON FORECASTS

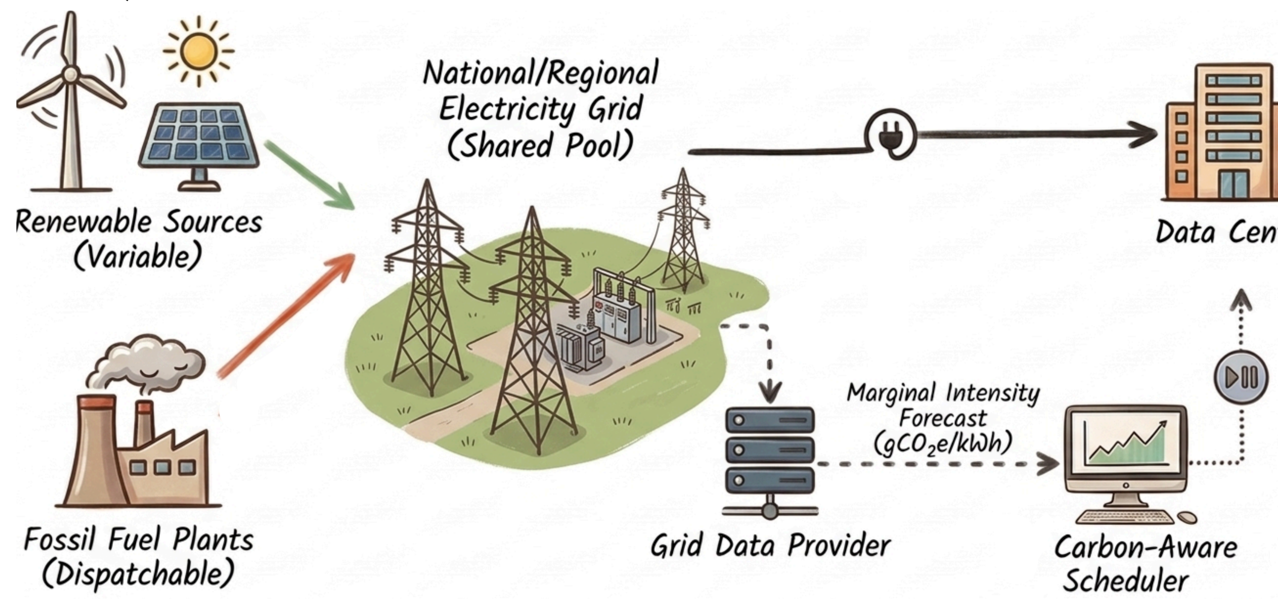
AYUSHI YADAV | MIDDLESEX UNIVERSITY

REDUCING CARBON EMISSIONS BY ENGINEERING THE FIRST SAFETY GATE FOR CARBON-AWARE SCHEDULING.

ENGINEERING SOLUTIONS FOR SDG THEMES

INTRODUCTION

Decarbonising computing infrastructure is a critical and underserved engineering challenge. Major cloud providers including Microsoft, Google, and AWS defer AI and HPC (High Performance Computing) workloads to periods of cleaner grid electricity, targeting SDG 13(Climate Action).



These scheduling decisions rely on carbon intensity forecasts (gCO_2e/kWh), which vary significantly with renewable supply and demand. As shown, a scheduler acts as a control layer between variable renewable sources, fossil fuel plants, and the data centre, using real-time forecasts to shift execution to lower-emission periods.

The current industry standard carbon scheduler, Let's Wait Awhile (Wiesner et al., 2021), scans an 8-hour window and selects the lowest-carbon slot. But forecasts are wrong by an average of 9.79% so when a forecast misleads, a deferred job runs under dirtier conditions than immediate execution. This is a **Carbon Trap**

GAP: **Existing schedulers assume forecasts are always correct, with no mechanism to detect when deferral increases emissions.**

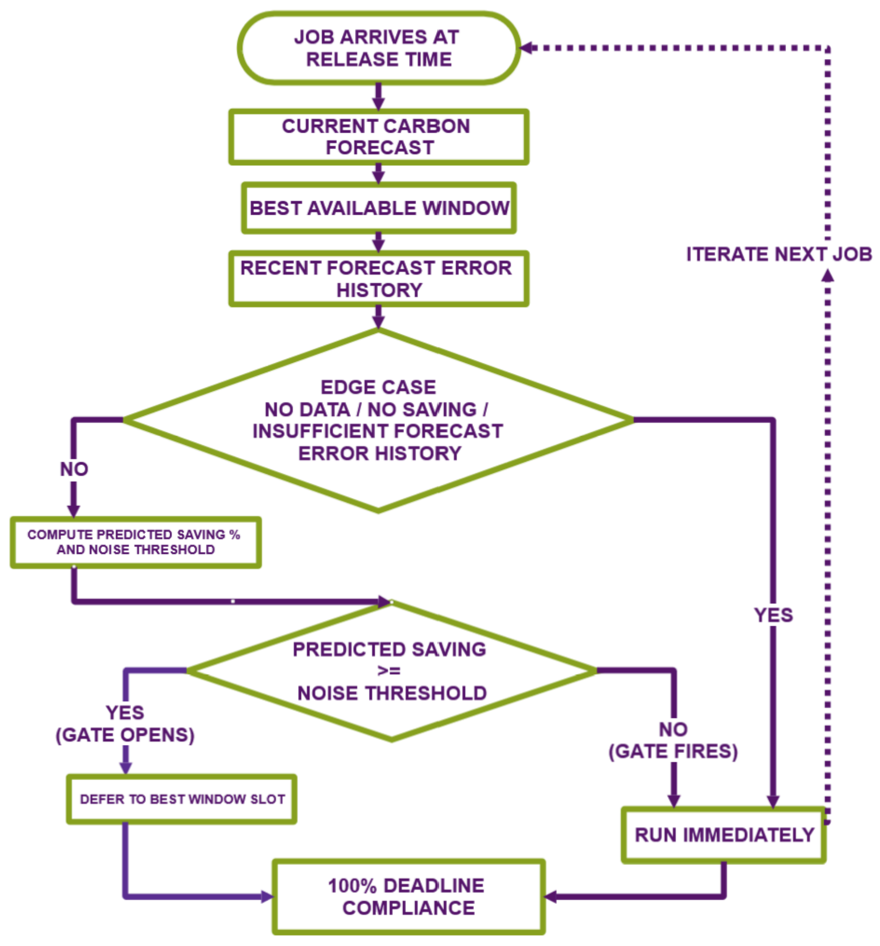
OBJECTIVE

To implement and validate a Reliability Gate that filters out low-confidence scheduling decisions, ensuring net carbon reduction even under high forecast uncertainty. Unlike existing approaches which blindly follow forecasts, **CarbonGuard** balances carbon savings with decision reliability, ensuring scheduling actions are both environmentally beneficial and practically safe.

METHODOLOGY

We replayed 35,927 jobs from ARCHER2 (UK national supercomputer) against 16,893 actual carbon intensity readings from NESO (National Energy System Operator, 2025–2026). This allowed direct comparison of **CarbonGuard** against Let's Wait Awhile under identical real-world conditions across a full year of HPC demand.

The core innovation is the **Reliability Gate** which is a dynamic trust mechanism absent from existing schedulers. When a job arrives, **CarbonGuard** identifies the cleanest available window within its deadline.



Before committing, the Gate performs a real-time Trust Check by measuring the error level that 75% of UK grid forecasts fell within over the previous 24 hours to establish a noise threshold. If the predicted carbon saving exceeds this threshold, the deferral is trustworthy and the gate opens. If not, the job runs immediately, avoiding a potential **Carbon Trap**.

This ensures **CarbonGuard** only takes safe bets, automatically adapting the threshold by tightening in volatile seasons like autumn and winter and relaxing more in summer when forecasts are stable.

KEY FINDINGS

CarbonGuard demonstrates that forecast reliability matters as much as forecast value. The dominant failure mode is not lack of data, but the absence of mechanisms to reason about uncertainty. The **94.8%** reduction in **Carbon Traps**, sustained across 9 sensitivity configurations, confirms this is a robust and generalisable result, not an experimental artefact.

93.3%
CarbonGuard safe decision rate

94.8%
of Let's Wait Awile traps prevented

19x
more reliable than Let's Wait Awile

RESULTS

Through a systematic sweep of 50 candidate sensitivity levels, we identified an optimal "Reliability Sweet Spot" (α=0.25) that maximizes carbon savings while neutralizing the risk of a **Carbon Trap**. Under this calibrated setting, **CarbonGuard** achieved a **15.57%** net carbon saving **(95% CI: 14.76%, 16.45%)**, nearly doubling the 8.42% achieved by the industry standard, Let's Wait Awhile.

Let's Wait Awhile produces **Carbon Trap** events for **13.18%** of all jobs meaning 1 in 8 scheduling decisions actively increases emissions. **CarbonGuard** reduces this to **0.68%**, which is a **94.8%** reduction.

The gate response curve demonstrates adaptive behaviour so as forecast error increases across quintiles, **CarbonGuard** automatically increases its protection threshold, suppressing unsafe deferrals precisely when forecasts are least trustworthy.

Notably, **CarbonGuard** achieves greater carbon savings while making fewer total deferrals, confirming the gate eliminates bad deferrals without blocking good ones. Results remain consistent at **15.70%**, **15.45%**, and **15.52%** across 3 independent seeds, ruling out favourable experimental conditions.

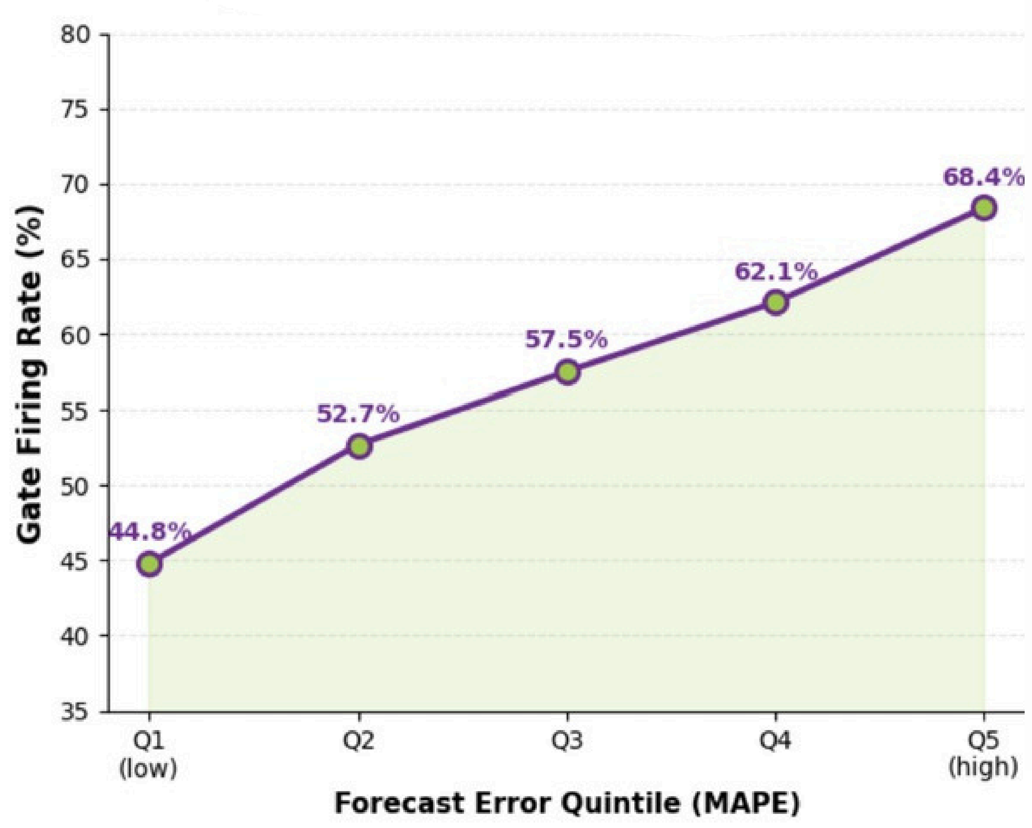


Figure 1:
CarbonGuard increases intervention rate as forecast error rises, preventing unsafe scheduling decisions under high uncertainty.

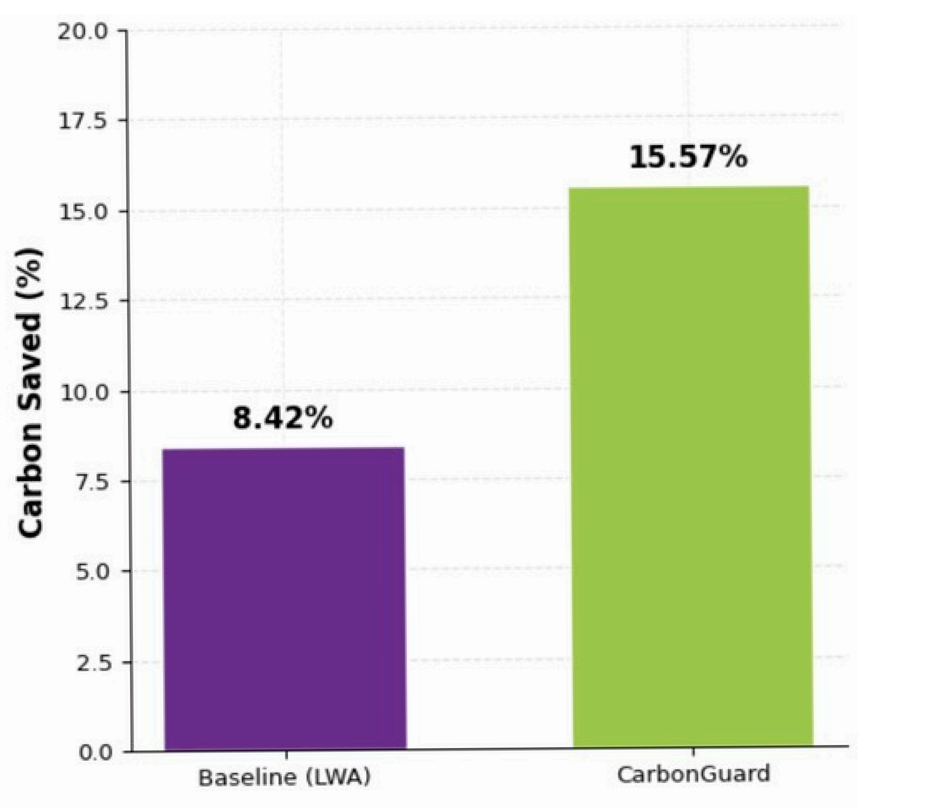


Figure 2:
CarbonGuard achieves significantly higher net carbon savings compared to LWA, confirming that reliability-aware scheduling improves overall system performance.

CONCLUSION

Carbon-aware scheduling is embedded in the infrastructure of the world's largest cloud providers, yet until now no engineering safeguard existed to verify that its decisions were actually reducing emissions. **CarbonGuard** closes that gap. This demonstrates that the primary limitation of existing schedulers is not insufficient data, but the absence of mechanisms to reason about uncertainty. That single intervention eliminates **94.8%** of **Carbon Traps**, nearly doubles net carbon savings, and advances SDG 13 through a practical, deployable engineering solution.

LIMITATIONS & FUTURE WORK

This evaluation focuses only on batch HPC jobs with deadline flexibility hence the approach is less applicable to latency-sensitive workloads. Results are validated on UK grid data (NESO) only so generalisation to other regions with different forecast accuracy profiles requires further validation.

Future work will extend **CarbonGuard** to multi-region grids, explore real-time deployment on live HPC systems, and investigate integration with energy-type disaggregation data to further refine scheduling decisions.

REFERENCES AND ACKNOWLEDGEMENTS

Wiesner, S. et al. (2021) Let's Wait Awhile: How Temporal Shifting Can Reduce Your Carbon Footprint. ACM/IFIP Middleware.

Antici, F. et al. (2023) PM100: A Power Consumption Dataset from a Petascale Supercomputer.

IEEE/ACM Supercomputing (SC). Dodge, J. et al. (2022) Measuring the Carbon Intensity of AI in Cloud Instances.

National Energy System Operator (NESO). UK Carbon Intensity Forecast Data, 2025-2026. Available at: carbonintensity.org.uk

Many thanks to Dr Kelly Androutsopoulos for her supervision and encouragement throughout this research.

Contact: ayushi59599@gmail.com