Uncertainty-Aware Decarbonization for Datacenters

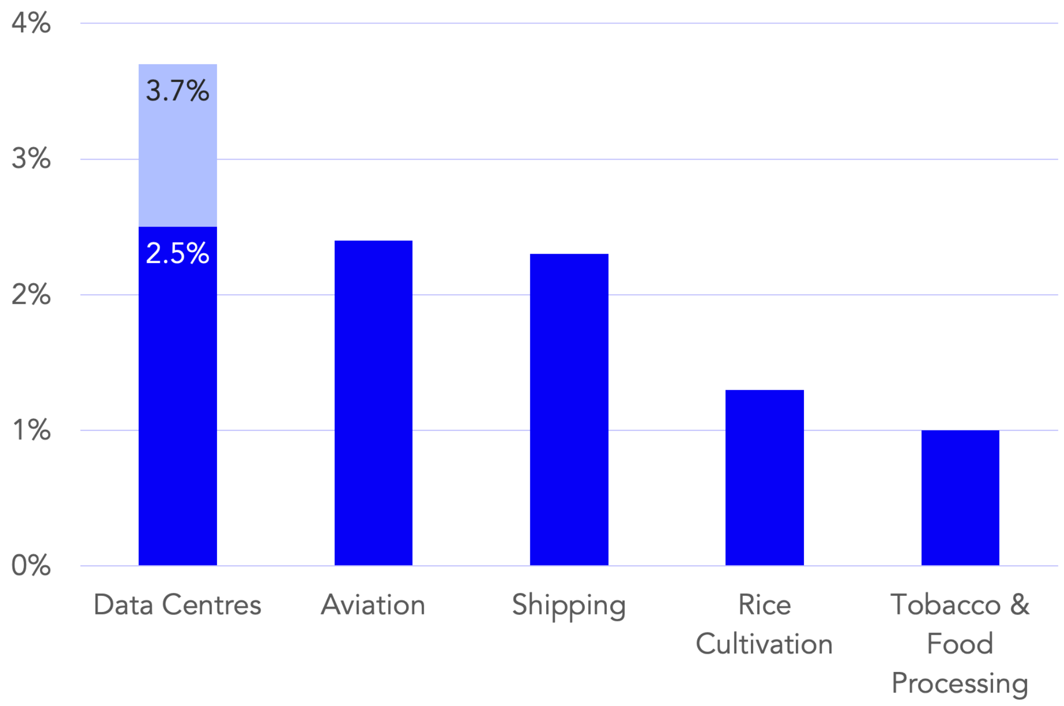
Amy Li[†], Sihang Liu[†], Yi Ding[‡]

[†]University of Waterloo, [‡]Purdue University

Why Datacenter Decarbonization?

Global cloud computing emissions exceed those from commercial aviation

Share of global CO_2 emission generated by sector/category



Source: Climatiq Analysis, The Shift Project, OurWorldinData

climatiq

7/10/24

Load Shifting

DATA CENTERS AND INFRASTRUCTURE

Our data centers now work harder when the sun shines and wind blows

Apr 22, 2020 · 3 min read

Ana Radovanovic
Technical Lead for Carbon-Intelligent Computing

Share

https://blog.google/inside-google/infrastructure/data-centers-work-harder-sun-shines-wind-blows/

7/10/24

Carbon Intensity

Carbon Explorer: A Holistic Framework for Designing Carbon Aware Datacenters

Bilge Acun, acun@meta.com, Meta, USA
Benjamin Lee, leebcc@seas.upenn.edu, University of Pennsylvania, Meta, USA
Fiodar Kazhamiaka, fiodar@stanford.edu, Stanford University, USA
Kiwan Maeng, kwmaeng@meta.com, Meta, USA
Udit Gupta, uditg@meta.com, Harvard University, Meta, USA
Manoj Chakkaravarthy, mchakkar@meta.com, Meta, USA
David Brooks, dbrooks@eecs.harvard.edu, Harvard University, Meta, USA
Carole-Jean Wu, carolejeanwu@meta.com, Meta, USA

Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions

Walid A. Hanafy, University of Massachusetts Amherst, USA
Qianlin Liang, University of Massachusetts Amherst, USA
Noman Bashir, Massachusetts Institute of Technology, USA
Abel Souza, University of Massachusetts Amherst, USA
David Irwin, University of Massachusetts Amherst, USA
Prashant Shenoy, University of Massachusetts Amherst, USA

Ecovisor: A Virtual Energy System for Carbon-Efficient Applications*

Abel Souza, Noman Bashir, Jorge Murillo, Walid Hanafy, Qianlin Liang, David Irwin, and Prashant Shenoy

University of Massachusetts Amherst

1270 IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 38, NO. 2, MARCH 2023

Carbon-Aware Computing for Datacenters

Ana Radovanović, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyue Xiao, Maya Haridasan, Patrick Hung, Nick Care, Saurav Talukdar, Eric Mullen, Kendal Smith, MariEllen Cottman, and Walfredo Cirne

CarbonScaler: Leveraging Cloud Workload Elasticity for Optimizing Carbon-Efficiency

WALID A. HANAFY, University of Massachusetts Amherst, USA
QIANLIN LIANG, University of Massachusetts Amherst, USA
NOMAN BASHIR, University of Massachusetts Amherst, USA
DAVID IRWIN, University of Massachusetts Amherst, USA
PRASHANT SHENOY, University of Massachusetts Amherst, USA

On the Limitations of Carbon-Aware Temporal and Spatial Workload Shifting in the Cloud*

Thanathorn Sukprasert, University of Massachusetts Amherst, USA
Abel Souza, University of Massachusetts Amherst, USA
Noman Bashir, Massachusetts Institute of Technology, USA
David Irwin, University of Massachusetts Amherst, USA
Prashant Shenoy, University of Massachusetts Amherst, USA

Toward Sustainable HPC: Carbon Footprint Estimation and Environmental Implications of HPC Systems

Baolin Li, Northeastern University
Rohan Basu Roy, Northeastern University
Daniel Wang, Northeastern University
Siddharth Samsi, MIT
Vijay Gadepally, MIT
Devesh Tiwari, Northeastern University

(Average) Carbon Intensity

Definition: grams of CO_2eq emitted per kWh of electricity generated.

Existing point prediction methods: ARIMA [1], Neural Networks [2, 3]

What about their uncertainty levels?

1. Neeraj Dhanraj Bokde, Bo Tranberg, and Gorm Bruun Andresen. Short-term co2 emissions forecasting based on decomposition approaches and its impact on electricity market scheduling. Applied Energy, 2021.
2. Diptyaroop Maji, Ramesh K Sitaraman, and Prashant Shenoy. Dacf: day-ahead carbon intensity forecasting of power grids using machine learning. E Energy, 2022.
3. Maji, Diptyaroop, Prashant Shenoy, and Ramesh K. Sitaraman. CarbonCast: multi-day forecasting of grid carbon intensity. BuildSys. 2022.

This Work: Uncertainty Quantification

- Identify and characterize two types of uncertainty
  - Temporal and spatial uncertainty in carbon intensity prediction
- Present an uncertainty quantification method
  - A conformal prediction-based framework
- Provide case studies using real-world production power traces in Scope 2

Uncertainty in Carbon Intensity Prediction

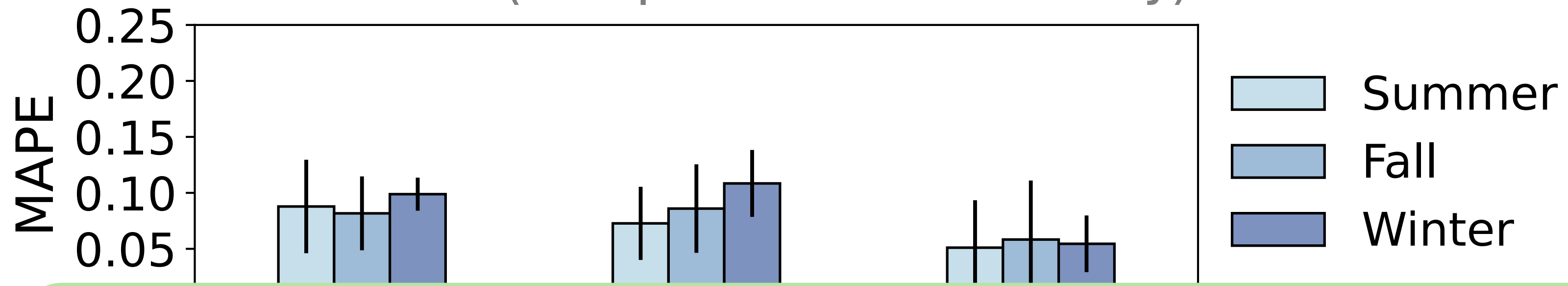
Characterization Setup

- Prediction tool: Pre-trained CarbonCast[1] model
- Test period: June – December 2022
- Regions: CISO (California), ERCO (Texas), and ISNE (New England)

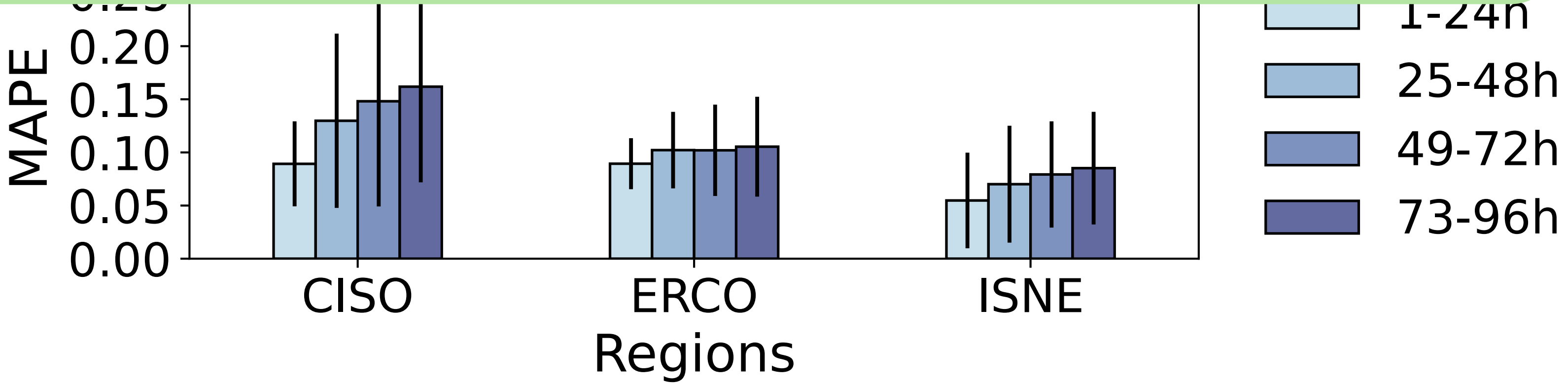
1. Maji, Diptyaroop, Prashant Shenoy, and Ramesh K. Sitaraman. CarbonCast: multi-day forecasting of grid carbon intensity. BuildSys, 2022.

Temporal Uncertainty: Short- & Long-term

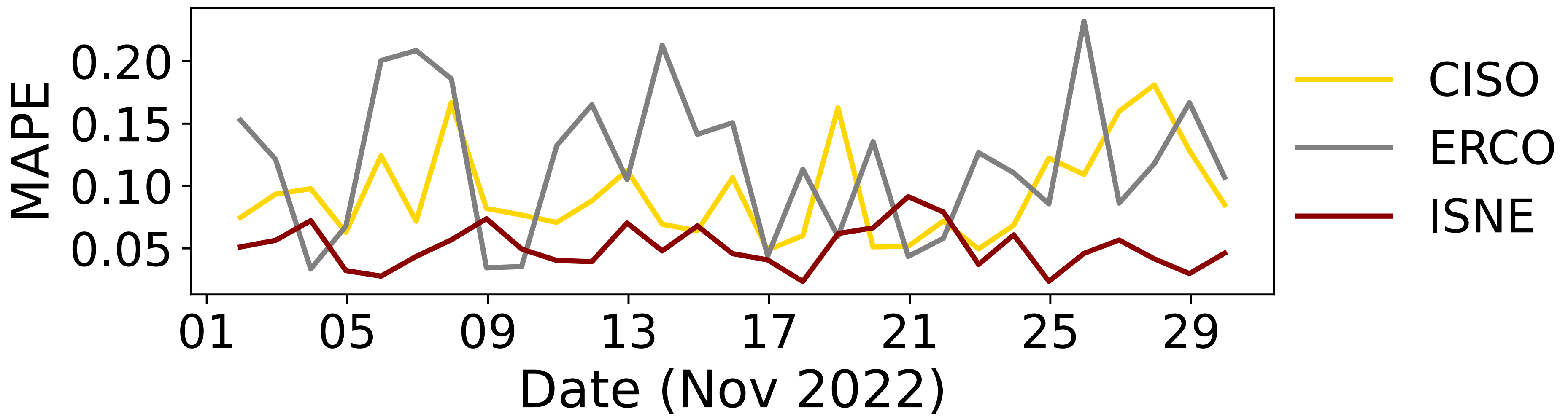
Short-term (24h prediction accuracy)



Addressing temporal uncertainty in carbon-aware scheduling is critical, especially for long-term job planning.



Spatial Uncertainty



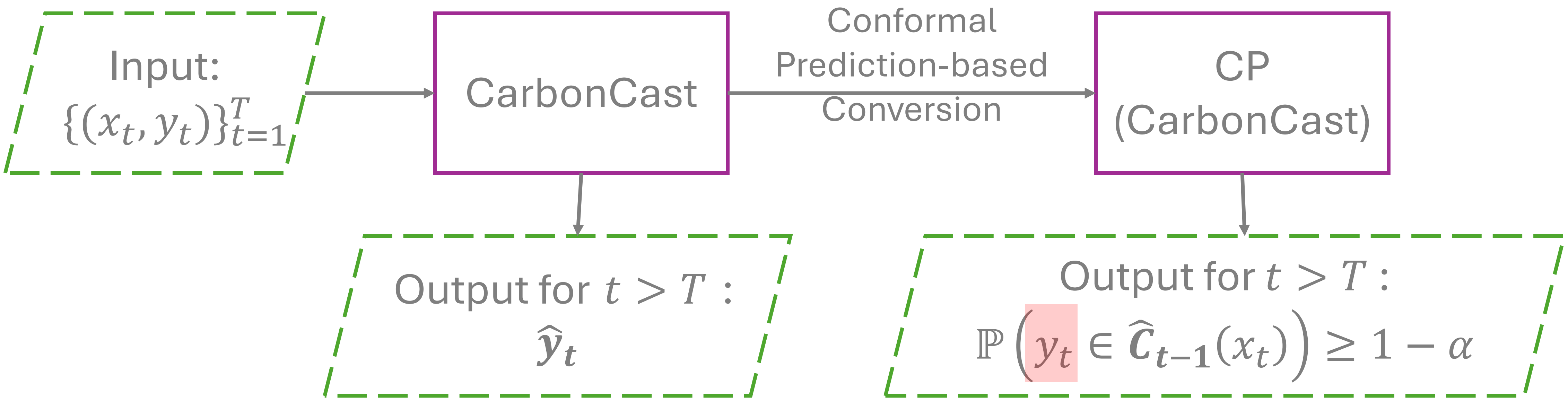
Suppose two datacenters, A and B, locate in different regions. The carbon intensity is predicted to be low in A at a low confidence, and high in B at a high confidence. What should we do?

Uncertainty Quantification

A Conformal Prediction-based Framework

Goal: generate confidence intervals that are guaranteed to contain the **true carbon intensity** with a user-specified probability

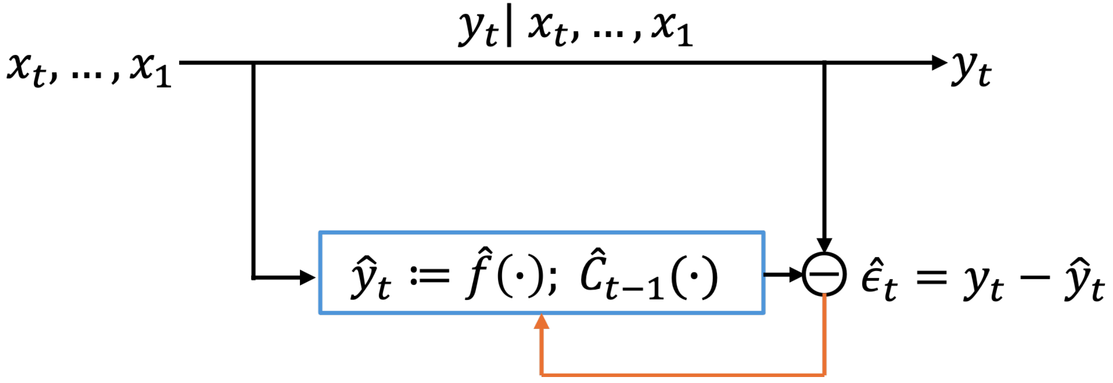
Idea: convert ***any*** algorithm's point predictions into prediction sets



More Highlights on CP-based Framework

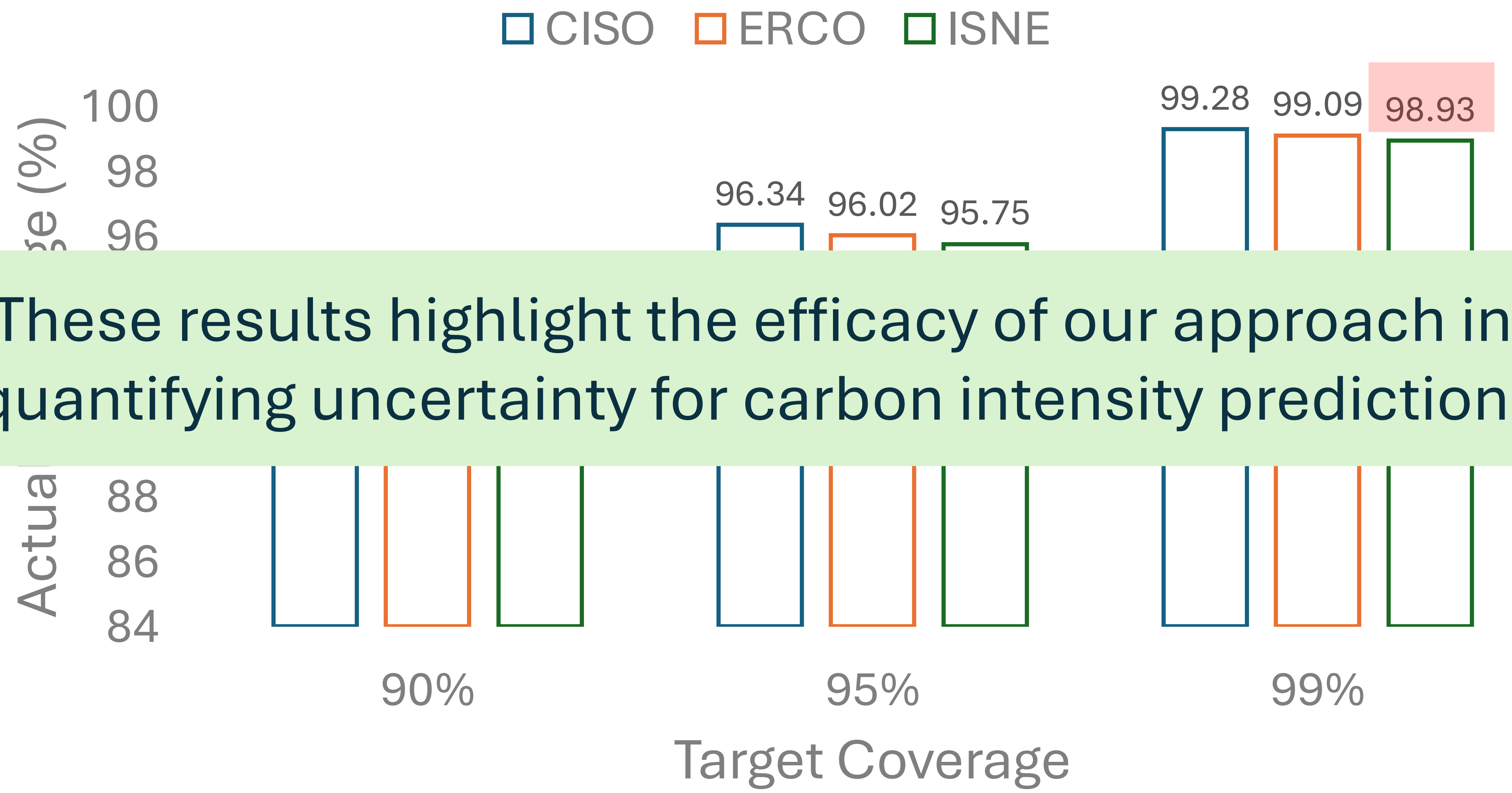
- The CP model may determine that the CarbonCast prediction is highly “non-conformal” and CP will provide a confidence interval that includes the true value but not the CarbonCast prediction.

- To account for the temporal dynamics, we leverage a feedback mechanism to encode the dependencies between time series.



Evaluation 1: Uncertainty Quantification

Main Results



These results highlight the efficacy of our approach in quantifying uncertainty for carbon intensity prediction.

Evaluation 2: Case Studies on Load Shifting

Evaluation Methodology

- Simulation data: Google production power traces[1]
- Load shifting policy: suspend-and-resume[2] (also called WaitAWhile)
  - suspend the work at higher carbon intensity; resume the work at lower carbon intensity.
- Clarification: case studies are only for proof-of-concepts, and cannot demonstrate real system benefits.

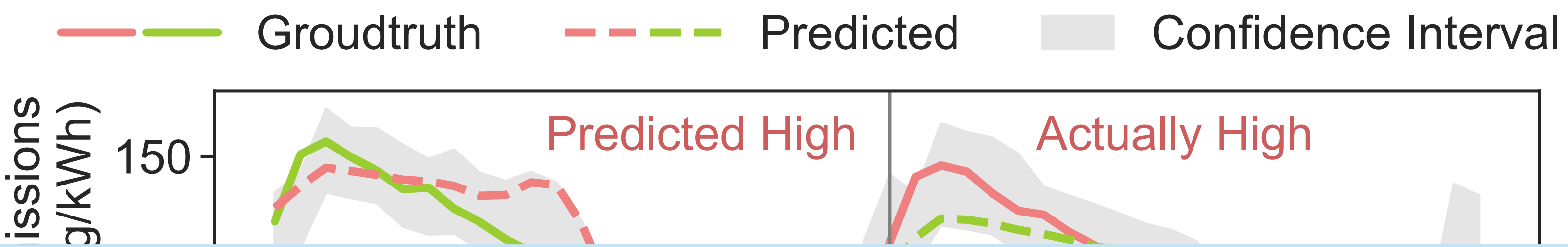
1. Varun Sakalkar, Vasileios Kontorinis, David Landhuis, Shaohong Li, Darren De Ronde, Thomas Blooming, Anand Ramesh, James Kennedy, Christopher Malone, Jimmy Clidaras, and Parthasarathy Ranganathan. Data center power oversubscription with a medium voltage power plane and priority-aware capping. ASPLOS, 2020.
2. Wiesner, Philipp, Ilja Behnke, Dominik Scheinert, Kordian Gontarska, and Lauritz Thamsen. Let's wait awhile: how temporal workload shifting can reduce carbon emissions in the cloud. Middleware, 2021.

Temporal Load Shifting

	CISO	ERCO	ISNE
Misleading Predictions	16.8%	10.6%	13.4%
Increased Emissions	4.3%	6.6%	4.6%

- Misleading Predictions: proportion of days when the predicted carbon intensity for the current day is lower than that of the next day, while in reality, the opposite is true.
- Increased Emissions: proportion of increased carbon emissions if shifting load from the current day to the next day in those cases.

Temporal Load Shifting: A 2-Day Example



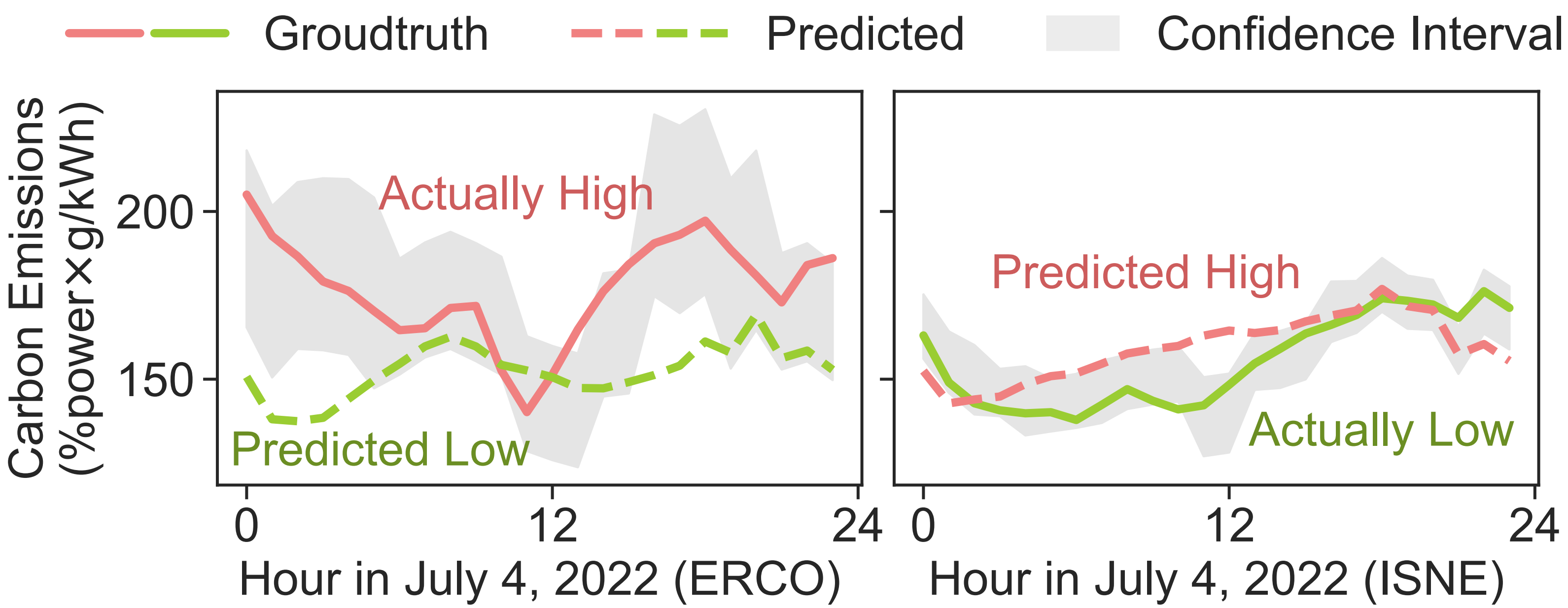
Decision makers should: (1) consider both predicted carbon intensity values and their uncertainty levels, and (2) shift load only when the confidence is sufficiently high.

	Groundtruth	Predicted	Confidence Interval
Day 1	1.00	1.13	[0.83, 1.21]
Day 2	1.05	0.96	[0.84, 1.20]

Spatial Load Shifting

Source	Target	Misleading Predictions	Increased Emissions
CISO	ERCO	5.0%	3.1%
	ISNE	7.8%	5.8%
ERCO	CISO	2.2%	2.7%
	ISNE	5.0%	3.5%
ISNE	CISO	4.5%	4.3%
	ERCO	2.8%	7.3%

Spatial Load Shifting: A 2-Region Example



	Groundtruth	Predicted	Confidence Interval
ERCO	1.00	0.86	[0.86, 1.11]
ISNE	0.87	0.90	[0.83, 0.93]

Yi Ding • You
Assistant Professor of ECE at Purdue University
3w • Edited •

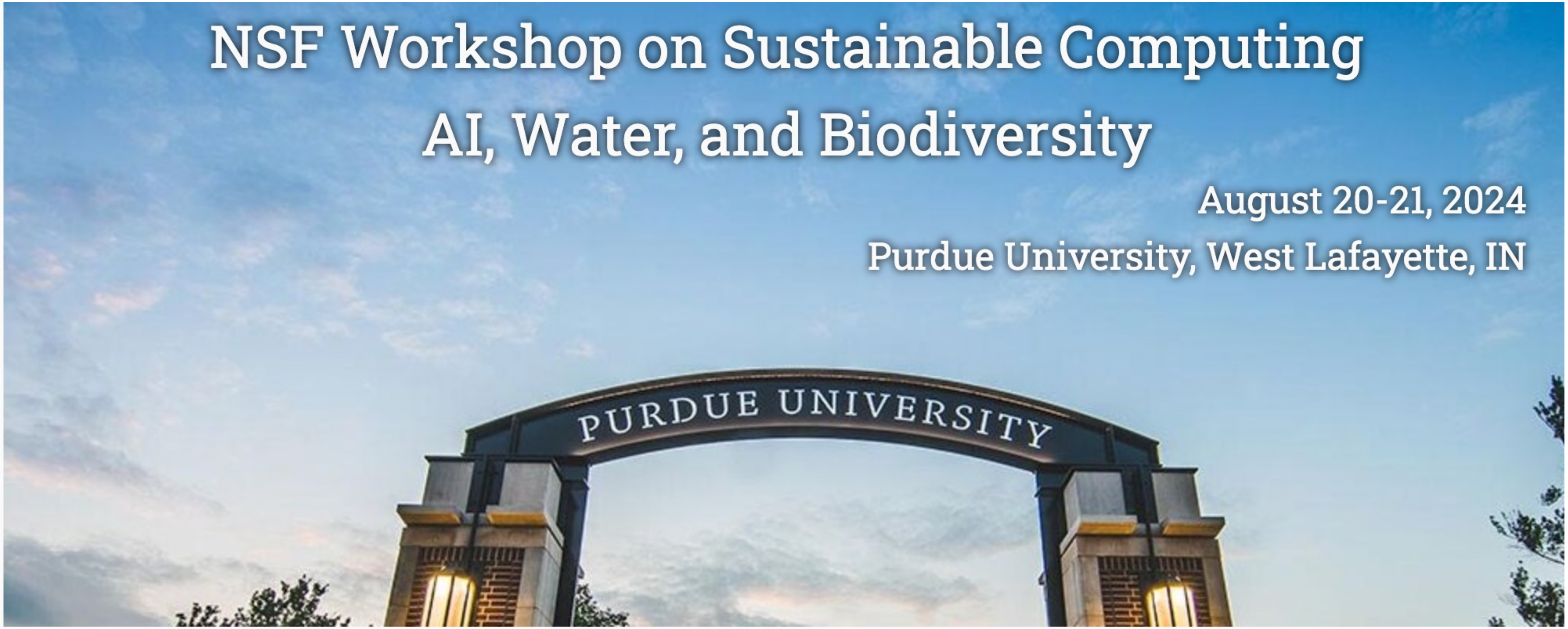
🚀 Exciting News! Applications for NSF Workshop on Sustainable Computing are Now Open! 🚀

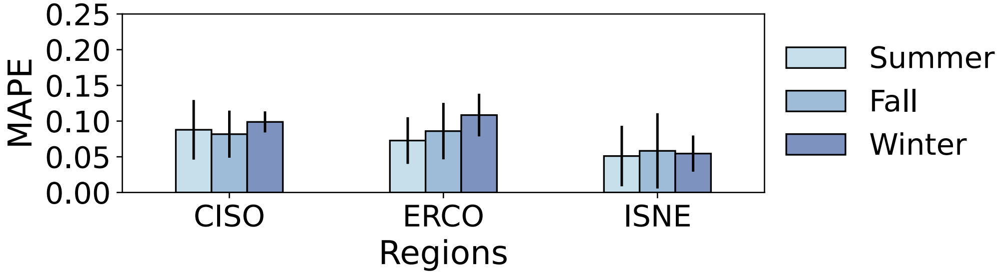
This is your chance to be a part of an incredible sustainable computing community where academia and industry leaders will share insights on AI, water scarcity, and biodiversity.

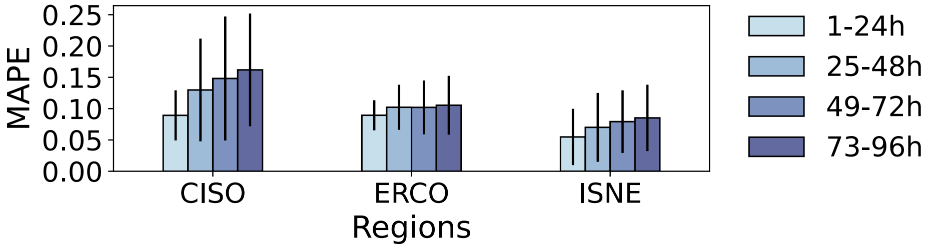
🗓️ Event Date: Aug 20-21, 2024
📍 Location: Purdue University, West Lafayette, IN
🔗 Apply Here: https://lnkd.in/gPqbpByz
⌛ Dates: Apply before July 19, 2024. Acceptance notification by July 22, 2024.
💰 Travel Support: U.S. participants with accepted submissions will be awarded a travel grant of up to $750.

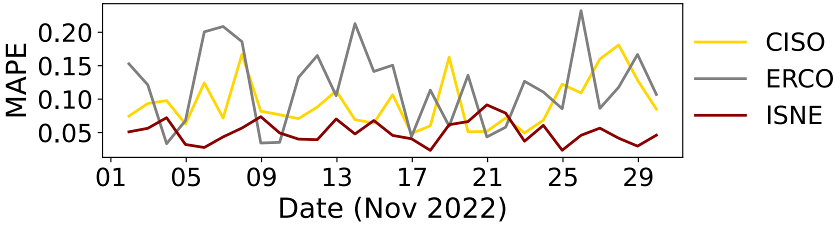
Why Attend?
🌟 Network with academia and industry leaders and peers.
🌟 Extend sustainability metrics beyond carbon to include water and biodiversity.
🌟 Address both positive and negative impacts of AI on these outcomes.

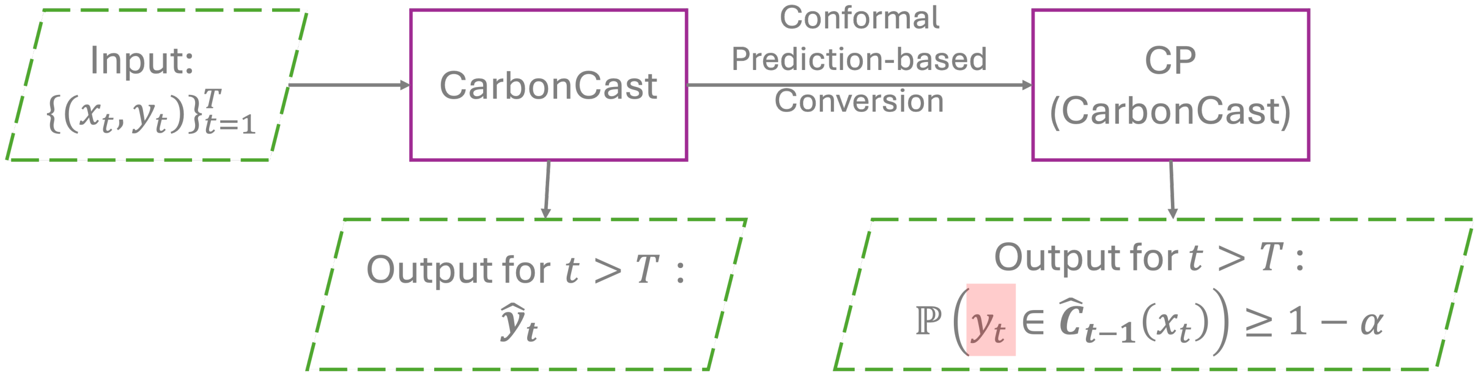
https://nsf-desc-2024.github.io/

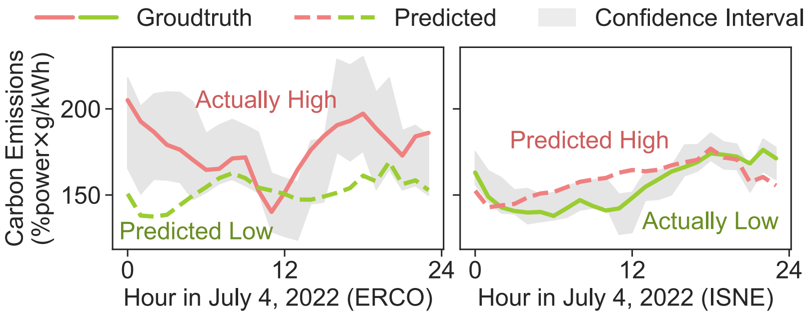


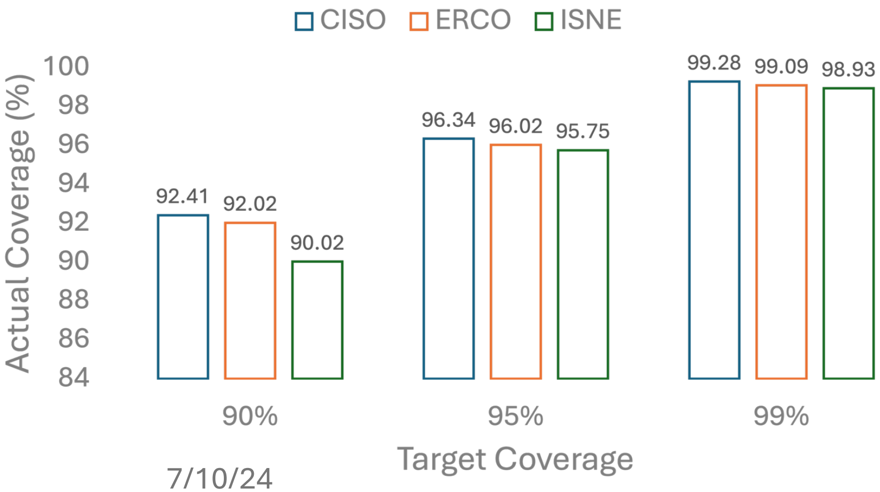












Thanks!

Questions?