Exploring the Explainability Power of AI in Identifying Causal Factors of Air Pollution: An Ozone in Los Angeles Case-Study

Georgia R. Klein, Avani N. Anne, Sarah C. Kavassalis (skavassalis@g.hmc.edu)
Harvey Mudd College, Claremont, CA

HARVEY MUDD COLLEGE

AGU 2023
A21E
2307

Introduction

- "Black-box" Artificial Intelligence (AI) techniques lack explanation of the results.
- Explainable AI (XAI) techniques seek to overcome this limitation but require validation.
- There is extensive research into the factors influencing ozone in Los Angeles, allowing us to evaluate the XAI technique Layerwise Relevance Propagation (Warnecke et al., 2020) on long short-term memory (LSTM) models against known ozone influencing factors.
- Ozone is produced through photochemical reactions involving NO_x and volatile organic compounds. Meteorology can enhance or suppress ozone concentrations (Pusede et al., 2015, Kavassalis & Murpy, 2017)
- LRP has been applied in the medical and energy fields (Lundberg et al., 2018, Erdem & Eken, 2020).

Our aim is to implement LRP XAI on LSTM models trained with ozone time-series data from Los Angeles in order to better understand the quality of LRP results.

Methods

- Data selected from the EPA Air Quality System (AQS)'s Los Angeles N. Main Street station.
- The LSTM was **train**ed on hourly data from **2010-2017**
- The model is **test**ed on hourly data from **2018-2019**.
- Features included are ozone, carbon monoxide, nitrogen dioxide, wind direction, wind speed, outdoor temperature, relative humidity, solar radiation, and ultraviolet radiation.
- LRP for LSTM implementation is sourced from Warnecke et al., 2020, based on the Arras et al., 2017 model.

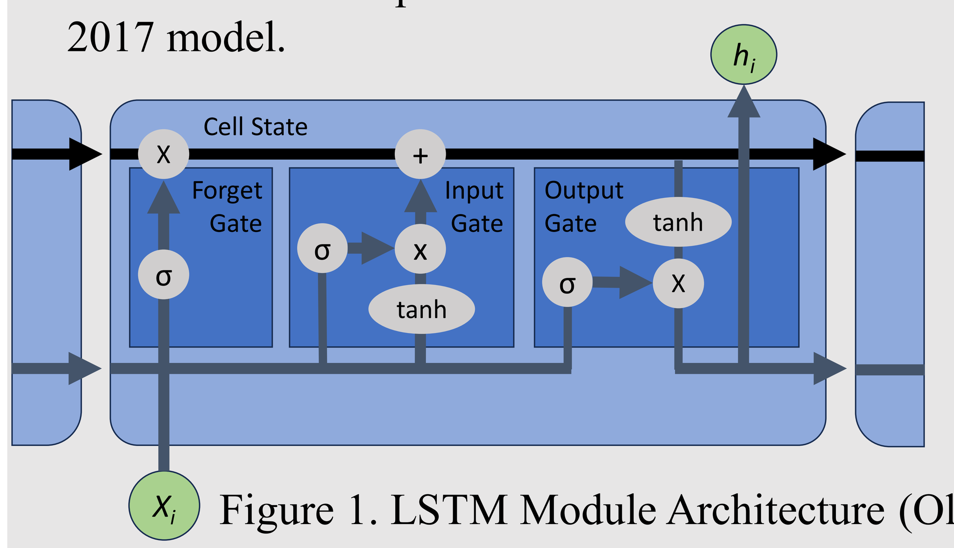


Figure 1. LSTM Module Architecture (Olah, 2015).

LRP calculation of relevance:

- Let the input gate neuron be z_g, and the source (x_t) neuron be z_s. Neuron for the next module is $z_j = z_g \cdot z_s$.
- Calculate relevance of the source provided with the relevance of the neuron in the next module, $R_g \cdot R_s = R_j$.

LSTM Model Results

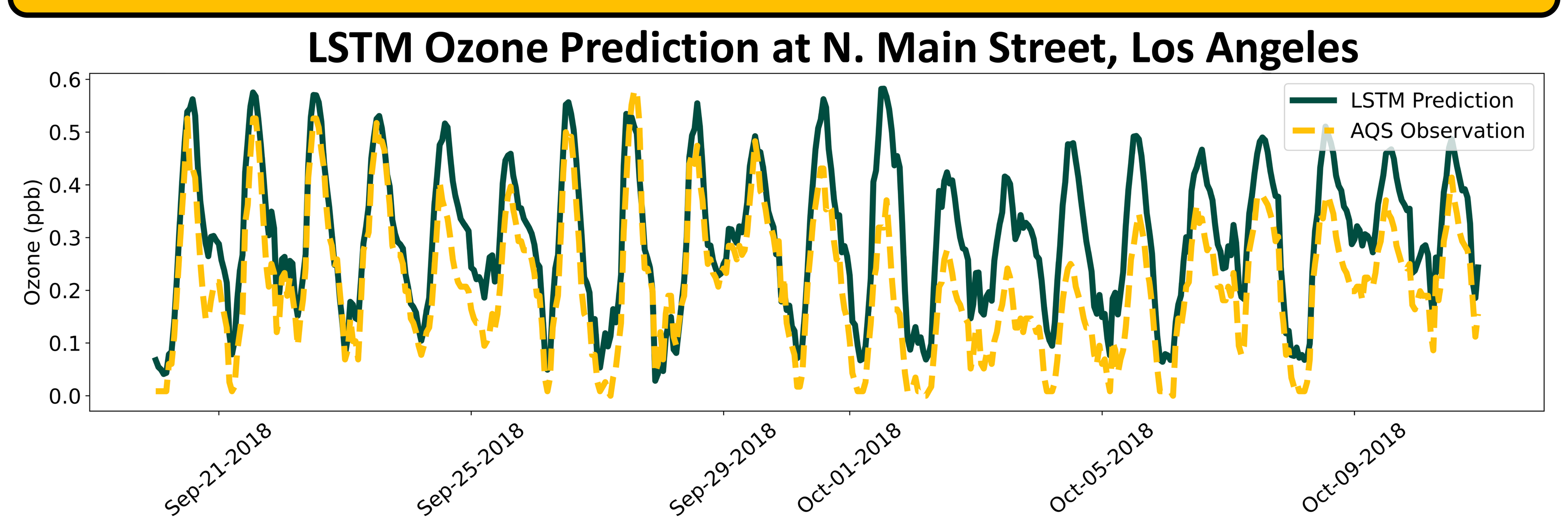


Figure 2. LSTM ozone prediction in comparison to ozone observations for a short example window. R^2 for all two years of test data is 0.71.

- Our LSTM predicts ozone well with an $R^2 = 0.71$.
- Ozone diurnal trends are captured well, as seen in Fig. 2.

If the LSTM predicts ozone well, *why* does it predict the mixing ratios it does?

Can We Interpret Ozone Predictions with Layerwise Relevance Propagation (LRP)?

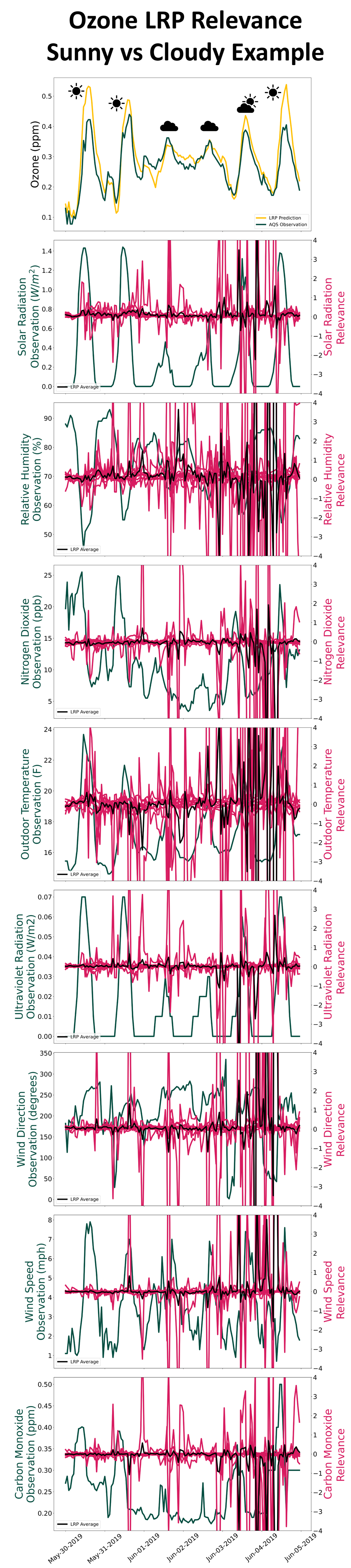


Figure 3. Example time series with 10 member LRP ensemble.

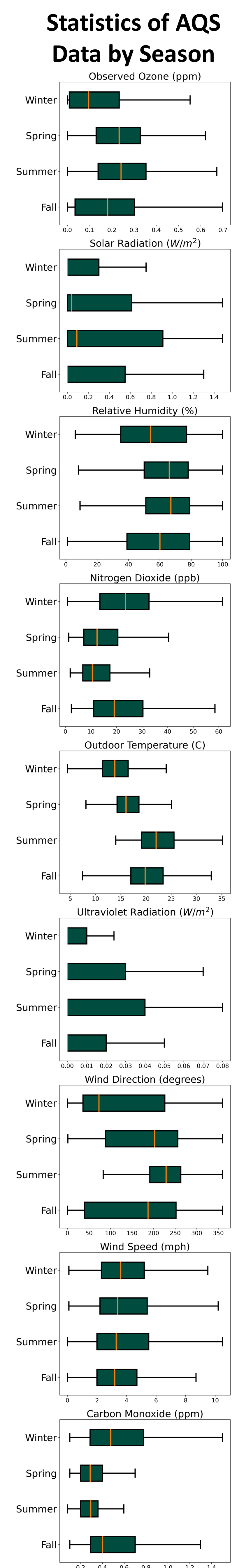
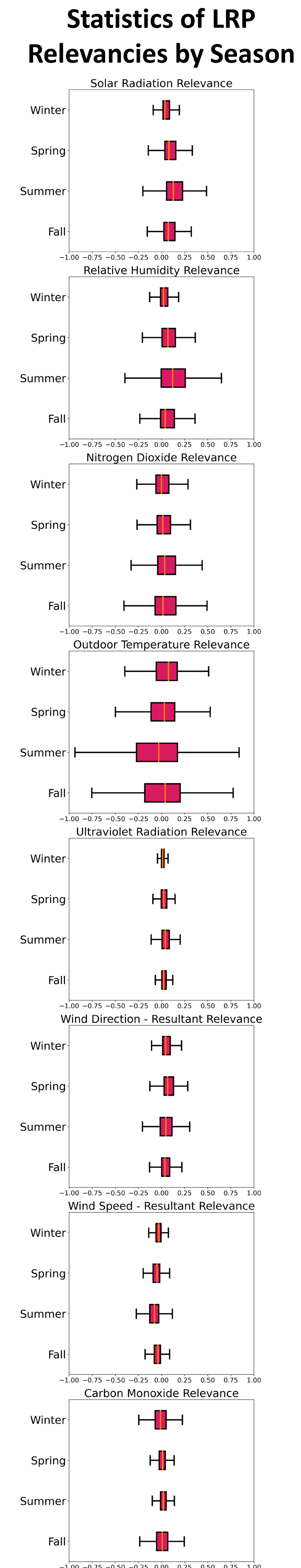


Figure 4. (left) and Figure 5. (below), box covers Q1-Q3 with line at median. Whiskers include data within 1.5x the IQR.



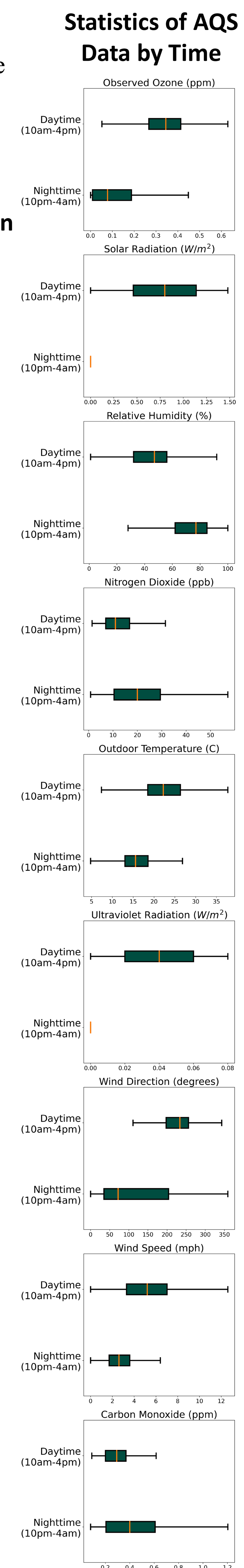
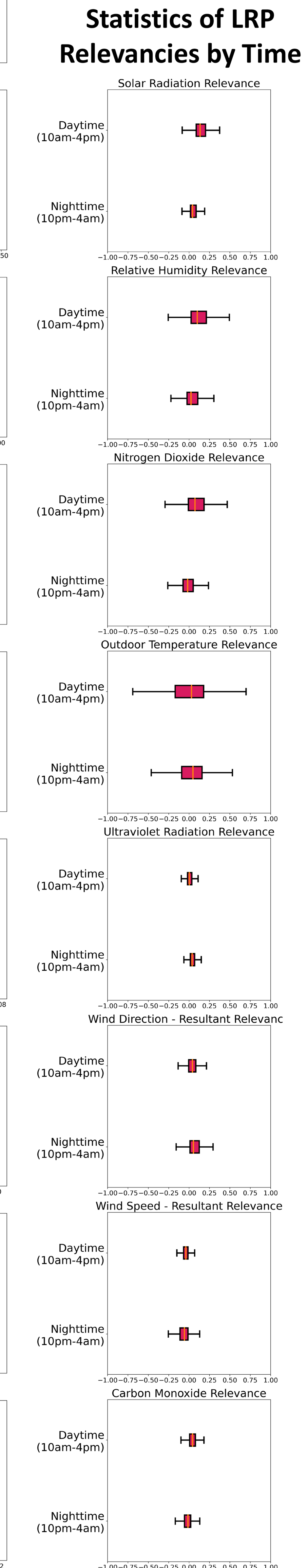


Figure 6. (left) and Figure 7. (below), box covers Q1-Q3 with line at median. Whiskers include data within 1.5x the IQR.



Results and Discussion

- Note: LRP relevancies sum to the LRP ozone predictions. When there is more ozone (summer, daytime), LRP relevancies are higher.
- In Fig. 3, the individual ensemble members for the LRP relevancies are not consistent. For any given ozone prediction, at the hourly scale, LRP does not identify a consistent driver.
- While individual ensemble members don't have predictive power, LRP statistics point to temperature, relative humidity, and NO driving ozone concentrations. This is consistent with chemistry!

LRP may provide meaningful information about main drivers of ozone, but does not seem to provide insights about individual predictions, as seen by the spread in ensemble members for given points in time.

Questions

- How/should LRP be used going forward? Given our findings, what use cases would LRP be beneficial for?
- What assessment is required to explore the features and functionality of LRP, and other XAI techniques, before they are applied?

Sources

1. Arras et al., *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2017.
2. Erdem and Eken, *Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, 2020.
3. Kavassalis and Murphy, *Geophysical Research Letters*, 2017.
4. Lundberg et al., *Nature Biomedical Engineering*, 2018.
5. Olah, 2015, accessed Dec 9th, 2023, http://colah.github.io/posts/2015-08-Understanding-LSTMs/.
6. Pusede et al., *Chemical Reviews*, 2015.
7. Warnecke et al., *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020.

Acknowledgements

Funding for this research was generously provided by Harvey Mudd College through the Program in Interdisciplinary Computation and the Leeds Student Travel Fund. We would like to thank the Harvey Mudd Chemistry Department and Hixon Center for Climate and the Environment for providing space and computing resources for the project and Bruce and Sharon DePriester for their donation to the FICUS lab