A Comparison of Different Machine Learning Methods for Identifying Subseasonal Forecasts of Opportunity Using Metrics of Stratospheric Variability

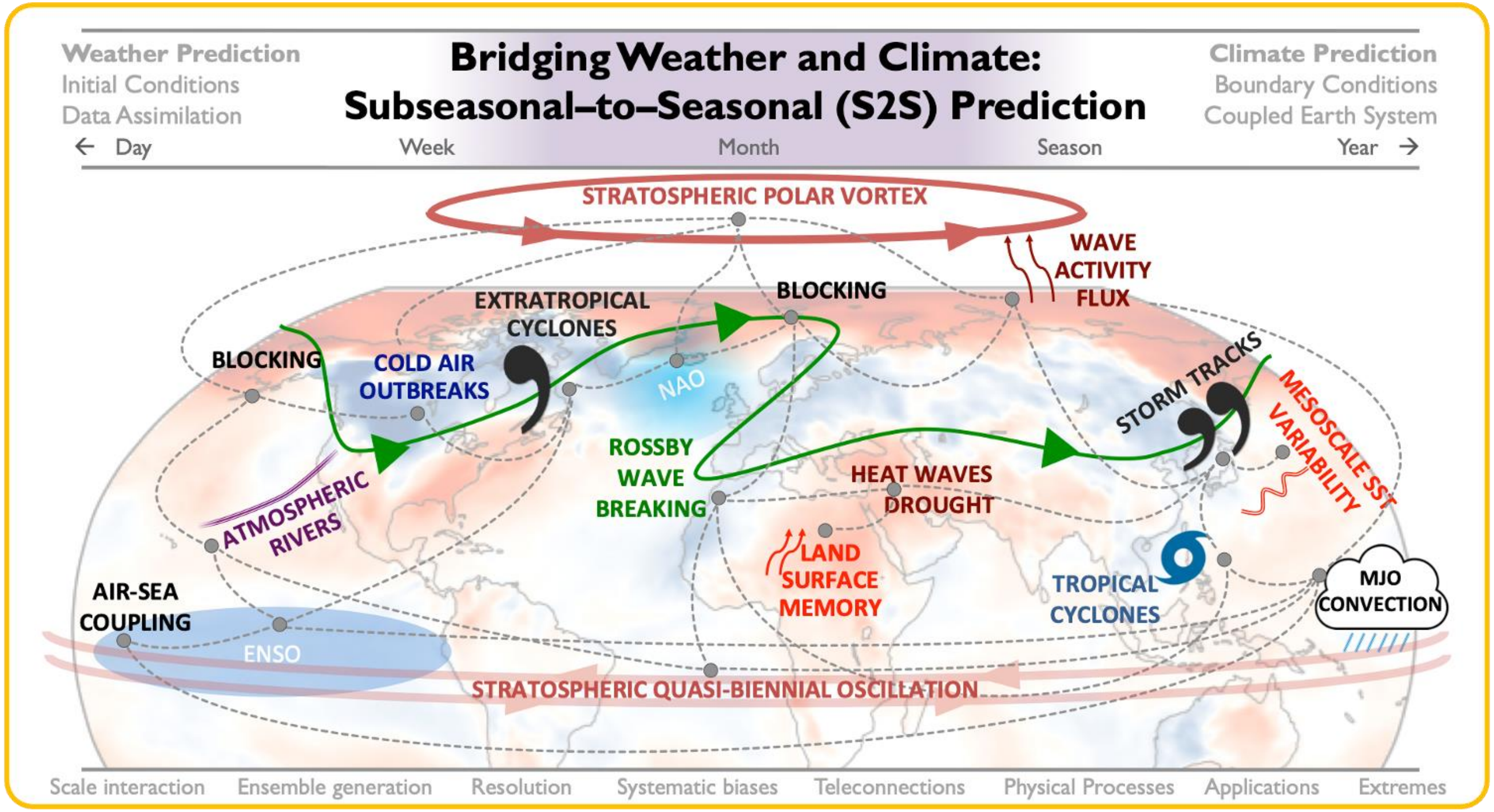
ELENA M. FERNÁNDEZ

29TH CONFERENCE ON PROBABILITY AND STATISTICS

JANUARY 27, 2026

PROJECT SUPPORTED BY #NA23OAR4310383B

Subseasonal Predictions: Sources of Predictability



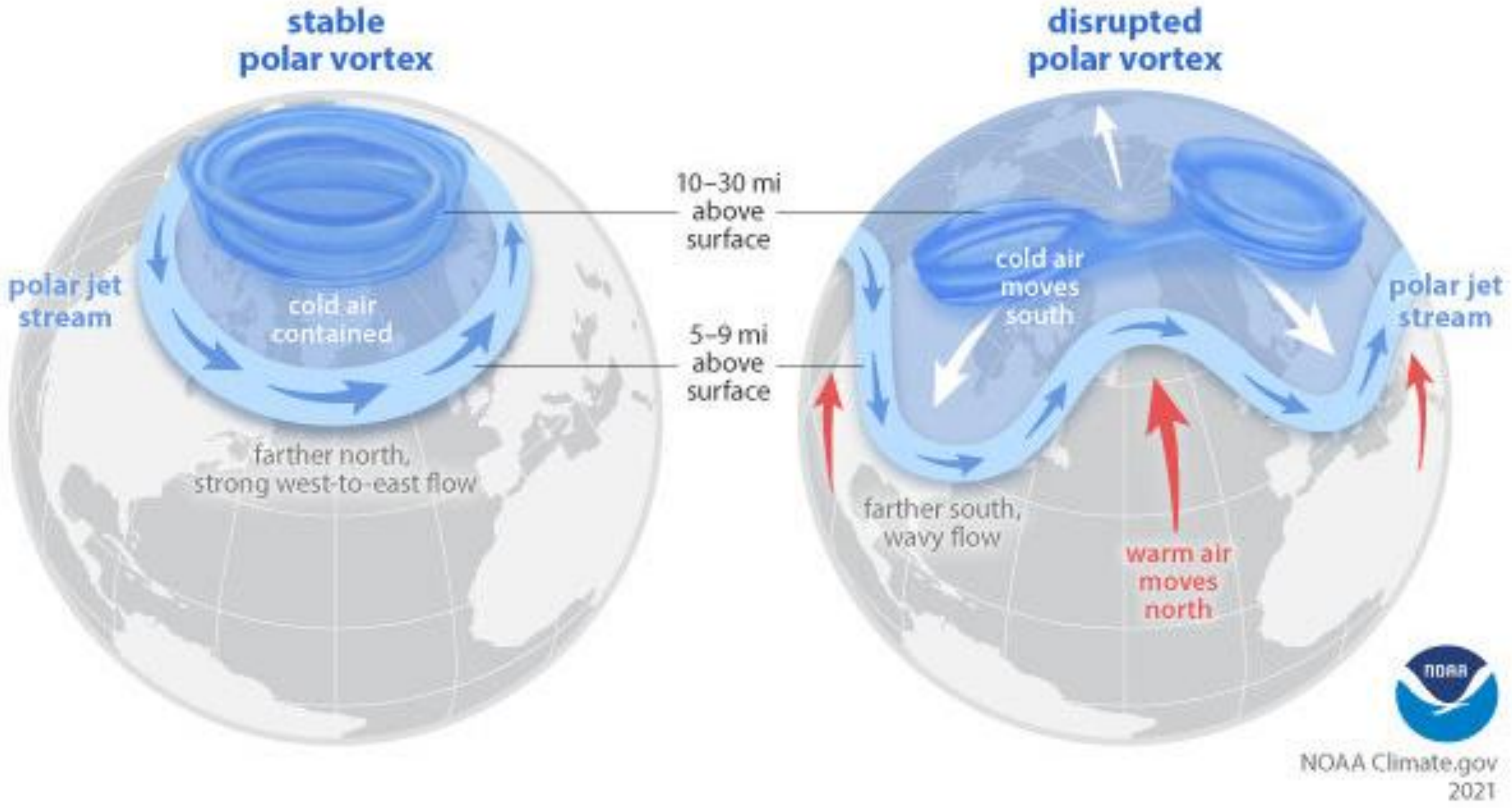
From the Special Collection on S2S Prediction in JGR-Atm and GRL

Lang et al. (2020)

Teleconnections (including anomalous stratospheric conditions) modulate tropospheric flow and can extend predictability beyond the medium-range

Subseasonal Predictions: Stratospheric Polar Vortex

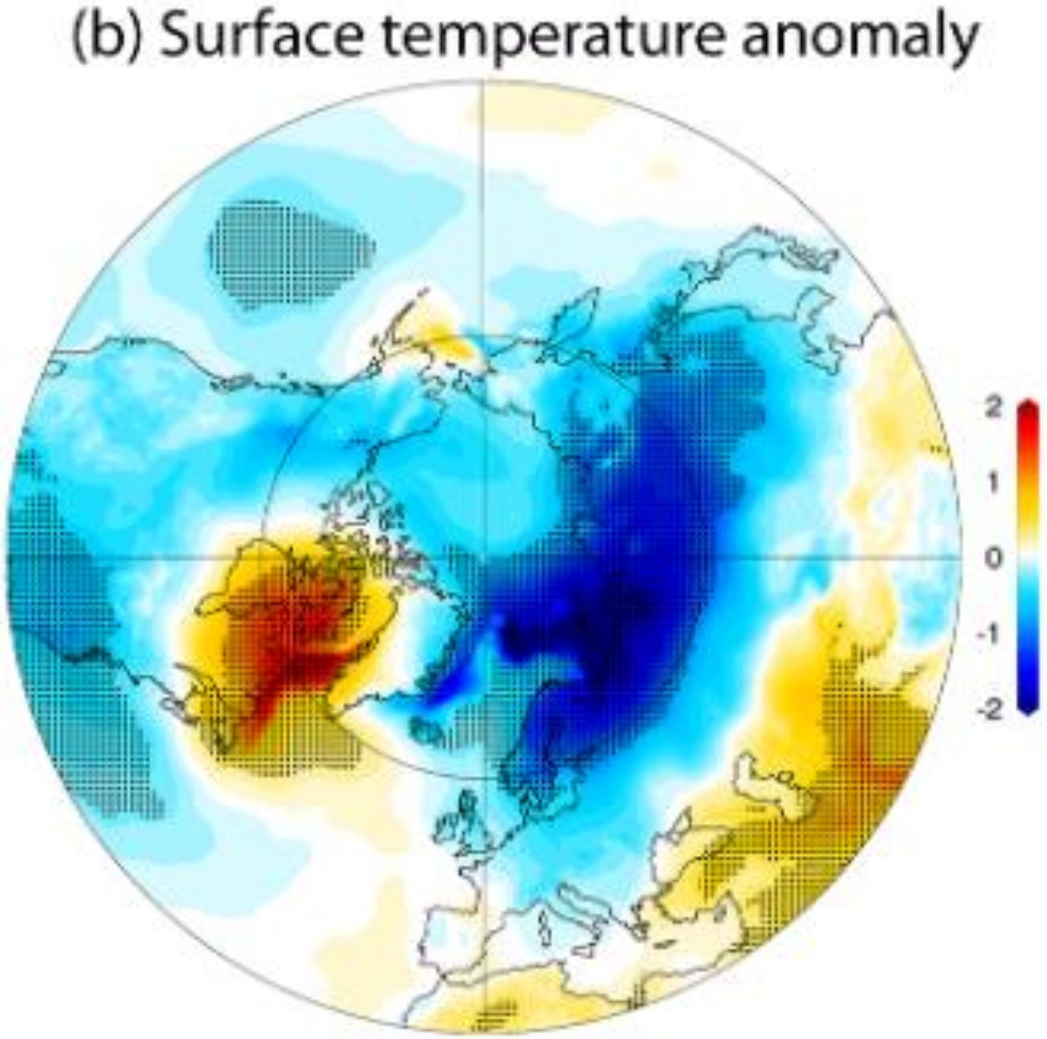
- Stratospheric polar vortex is a band of westerly winds that forms in the stratosphere in the NH winter
- Can be knocked off-balance by Rossby wave breaking activity into the stratosphere



Changes to the stratospheric polar vortex encourage the shifting of the polar jet stream and anomalous tropospheric temperatures

Figure courtesy of NOAA CPC.

Using the Stratosphere to Forecast Tropospheric Extremes



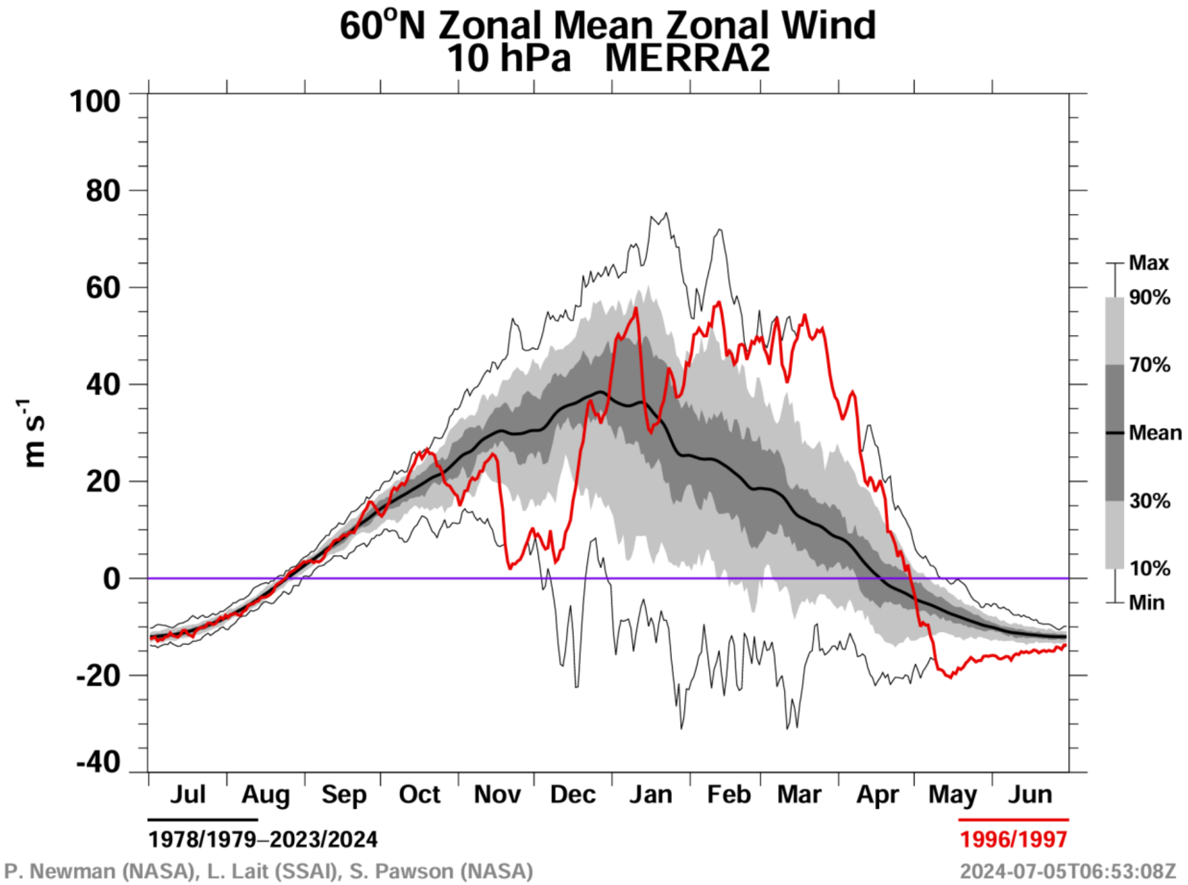
From: Butler et al. (2017)

- Changes to the strength and location of the stratospheric polar vortex serve as a known source of predictability into the troposphere

- Research shows specific conditions of the stratospheric polar vortex can produce enhanced forecast skill of NH temperature extremes (Butler et al. 2017, Tripathi et al. 2015, Keifer et al. 2023)

Using the Stratosphere to Forecast Tropospheric Extremes (cntd.)

- The 10-hPa zonal-mean zonal winds at 60°N are commonly used metric to determine the strength of the stratospheric polar vortex (e.g., Charlton and Polvani 2007)
- Non-zonal metrics defining the stratospheric polar vortex (including its shape) can also be an indicator of tropospheric impacts (e.g., Kretschmer et al. 2018)



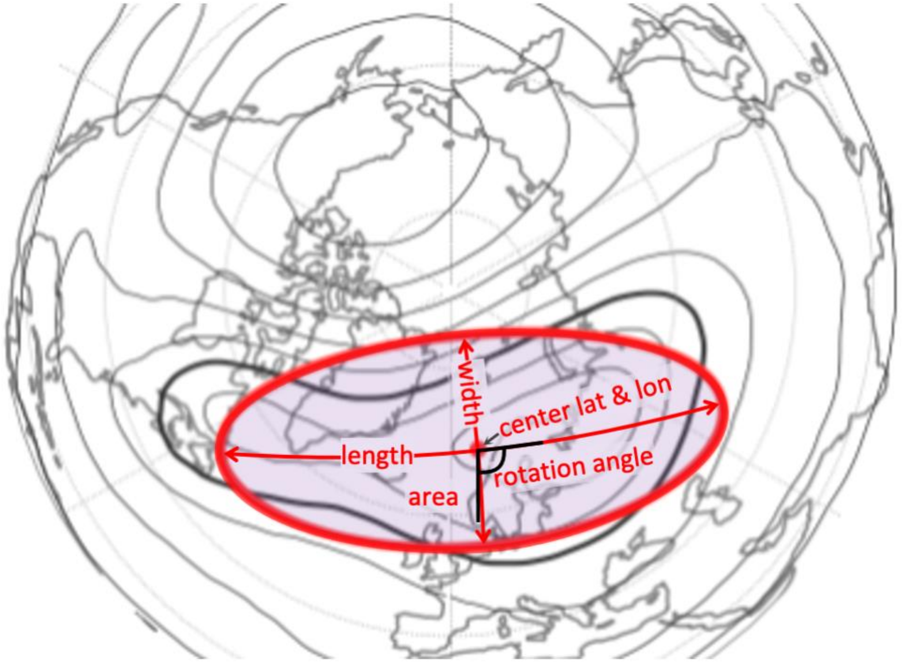
Stratospheric Polar Vortex Ellipse Metrics

Method presented in manuscript Fernández et al. (2026) provides a definition for the daily evolution of the vortex's shape/geometries (*currently Under Revision with the AMS Journal of Applied Meteorology*)

A best-fit ellipse calculation identifies an ellipse representing the 30-km contour of geopotential height at 10hPa, from which non-zonal quantities of the vortex are calculated:

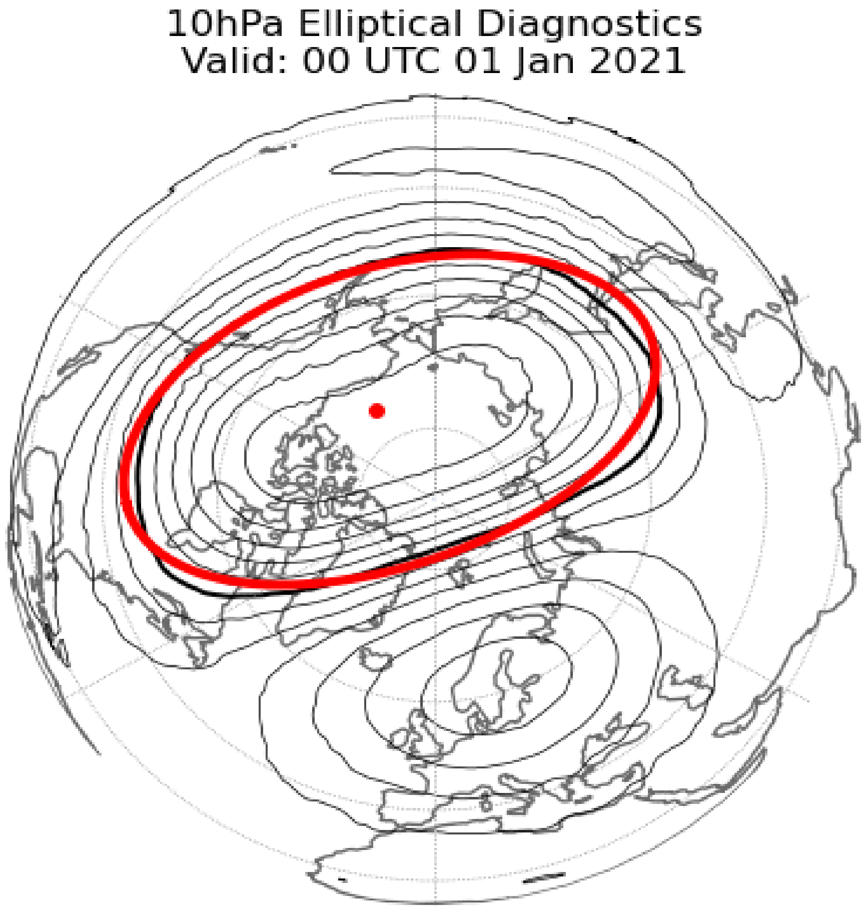
- Center latitude (*cenlat*)
- Center longitude (*cenlon*)
- Vortex ratio (length/width; *rat*)
- Vortex area or size (m^2; *size*)
- Vortex angle, measured between of the long axis with respect to 0˚

10-hPa Geopotential Height and Vortex Ellipse
1800 UTC 1 January 2019



Research Questions

- Is it possible to use the stratospheric polar vortex ellipse metrics for predicting subseasonal temperatures and forecasts of opportunity in the NH?
- Does the inclusion of timeseries information impact the applicability of these ellipse metrics (*i.e., does a LSTM model show any improvement over the baseline RF model*)?
- Are certain ellipse metrics more useful than others?

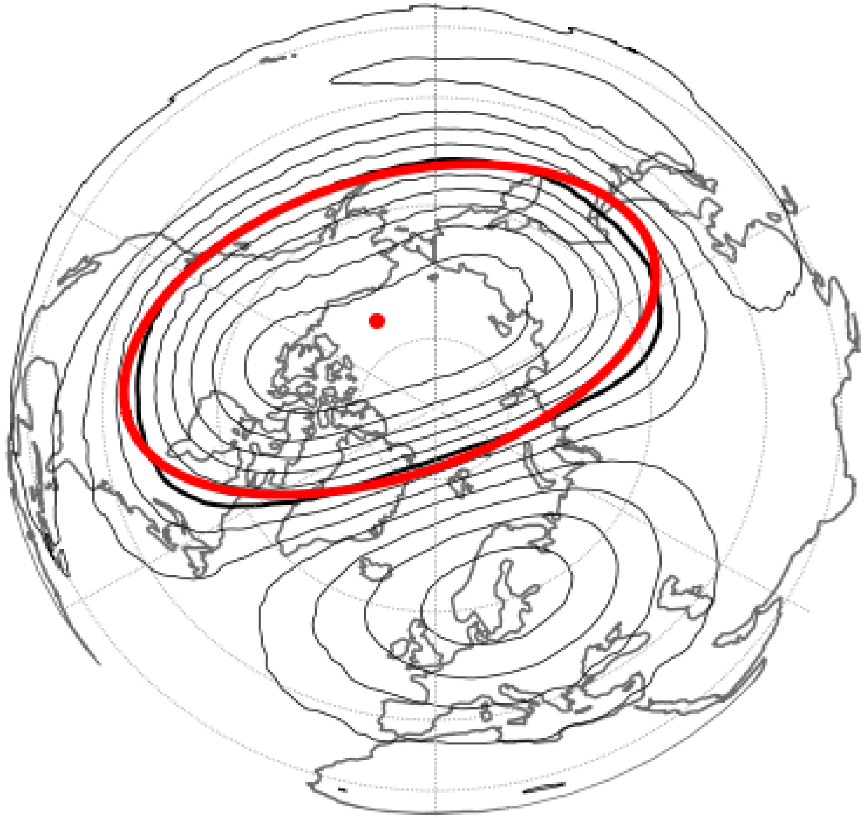


Research Questions

- Is it possible to use the stratospheric polar vortex ellipse metrics for predicting subseasonal temperatures and forecasts of opportunity in the NH?
- Does the inclusion of timeseries information impact the applicability of these ellipse metrics (*i.e., does a LSTM model show any improvement over the baseline RF model*)?
- Are certain ellipse metrics more useful than others?

<u>Goal:</u>

- ✓ Conduct a comparison of two models with the same task of predicting the 14-day lagged temperature outcome for a designated region
- ✓ Determine whether the models identify the same forecasts of opportunity



Data and Methods

All data are daily ERA-5 wintertime values in NH; November through March from 1959/1960 to 2021/2022

Random Forest and Long-Short Term Memory Model Input Features:

- 100 hPa Polar Cap Potential Vorticity area-weighted anomaly → 60-90°N
- 100 hPa N. Atlantic Geopotential Height area-weighted anomaly → 60-80°N, 45-20°W
- 10 hPa ellipse metrics (zonal-mean wind, size, central latitude/ longitude, ratio)

Data and Methods

All data are daily ERA-5 wintertime values in NH; November through March from 1959/1960 to 2021/2022

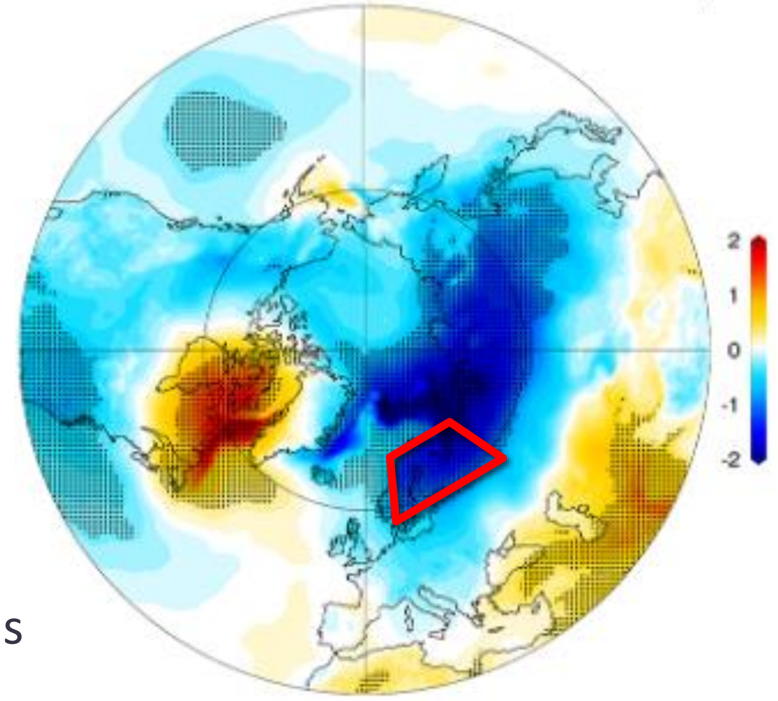
Random Forest and Long-Short Term Memory Model Input Features:

- 100 hPa Polar Cap Potential Vorticity area-weighted anomaly
- 100 hPa N. Atlantic Geopotential Height area-weighted anomaly
- 10 hPa ellipse metrics (zonal-mean wind, size, central latitude/ longitude, ratio)

Target Output for BOTH models:

14-day lagged Positive or Negative temperature anomaly classifications at 1000-hPa over Eurasia (between 10-45°E and 60-75°N)

(b) Surface temperature anomaly



From: Butler et al. (2017)

Training and Testing Data

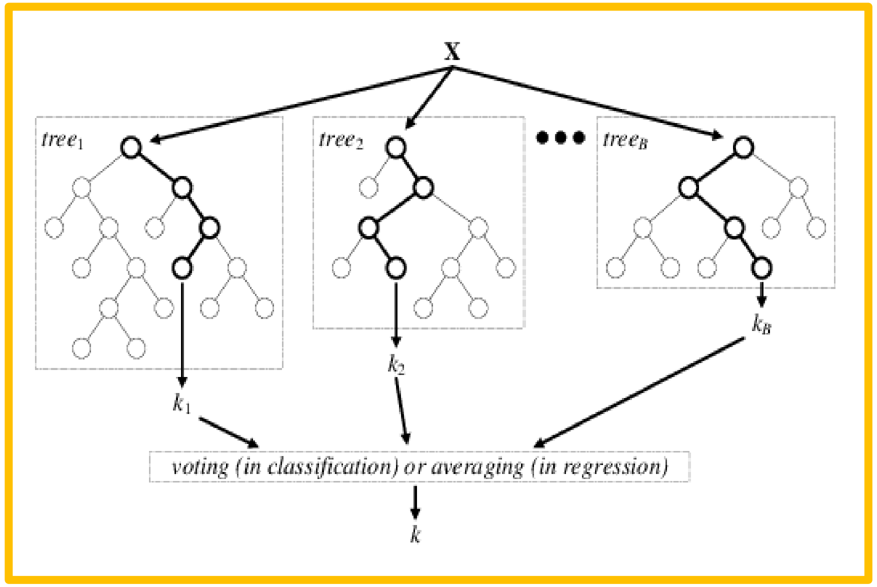
- The input data are max-min normalized after being separated into training and testing components
- Training data represents the first 52 years of data (1959/1960 through 2011/2012)
  - A randomly selected, consecutive 10-year period from within the training data is used for validation
- Testing data is the last 10 years of data (2012/2013 through 2021/2022)
  - ✓ Both models (Random Forest and LSTM) are cross-validated 100 instances
  - ✓ Model performance is evaluated with Brier Skill and F1-Scores
  - ✓ Explainable AI techniques (e.g., SHAP, Forecasts of Opportunity) are used in comparing the interpretability of the models

Random Forest Model

- The Random Forest (RF) model serves as the baseline model in this experiment (Breiman 2001, Hill et al. 2024)
  - Interpretable, not a black-box
  - Easily manage non-linear and complex datasets

- No timeseries information is included (January 1st events predict January 14th temps)

- RFs are reliable for predicting extremes, including European temperatures (Kiefer et al. 2023)



Hyperparameters	Random Forest
Depth	3
Estimators	400

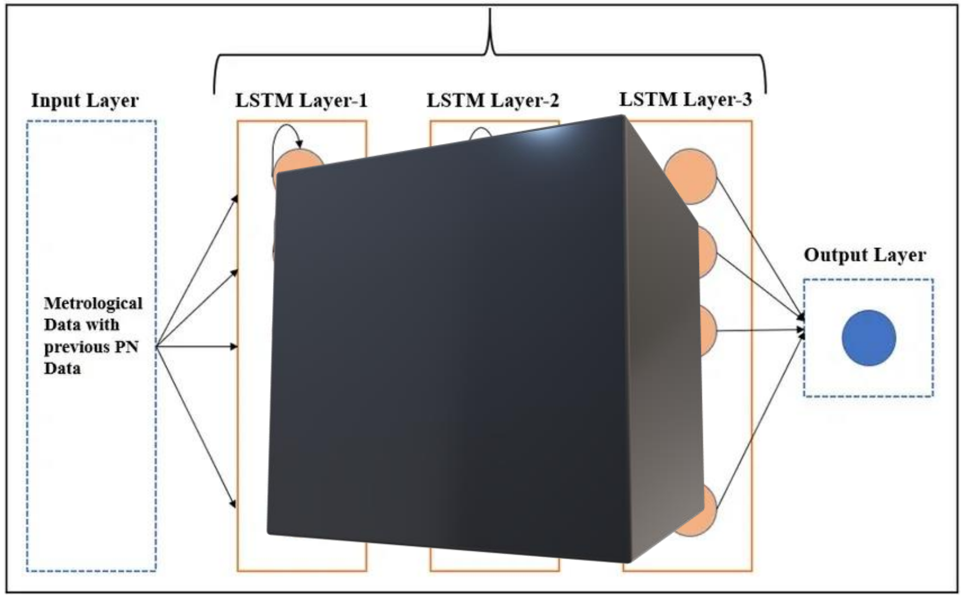
Recurrent Neural Network
(Long Short-Term Memory Model)

- LSTMs are a special type of RNN that uses memory cells to learn and retain information, helping in understanding dependencies across data

- LSTMs are black box models
  - Complex internal structure limits understanding of the model's decision making (Hochreiter and Schmidhuber 1997)

- LSTM model uses a sequential 14-day timeseries for each of the seven input features at every timestep
  - Lagged an additional 14-days to the predicted temperature outcome (maximum 28-day forecast)

- Ridge regularizer selected to help in managing multicollinearity



Hyperparameters	Stacked LSTM
Input Layer	(14,7)
Recurrent Layer 1 Nodes	128
Recurrent Layer 2 Nodes	14
Dense Layer 3	28
Ridge Regularizers	0.03, 0.03, 0.06
Learning Rate	0.0009

Forecasts of Opportunity

- When conducting cross-validation on the model, look at the average ACC across all predictions made
- Compare the average ACC of all predictions with the ACC for predictions with the highest confidence (Mayer and Barnes 2021)

90th Percentile

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Probability Cold	0.45	0.47	0.38	0.41	0.18	0.09	0.25
Probability Warm	0.55	0.53	0.62	0.59	0.82	0.91	0.75

The model is confident, but are these correct?

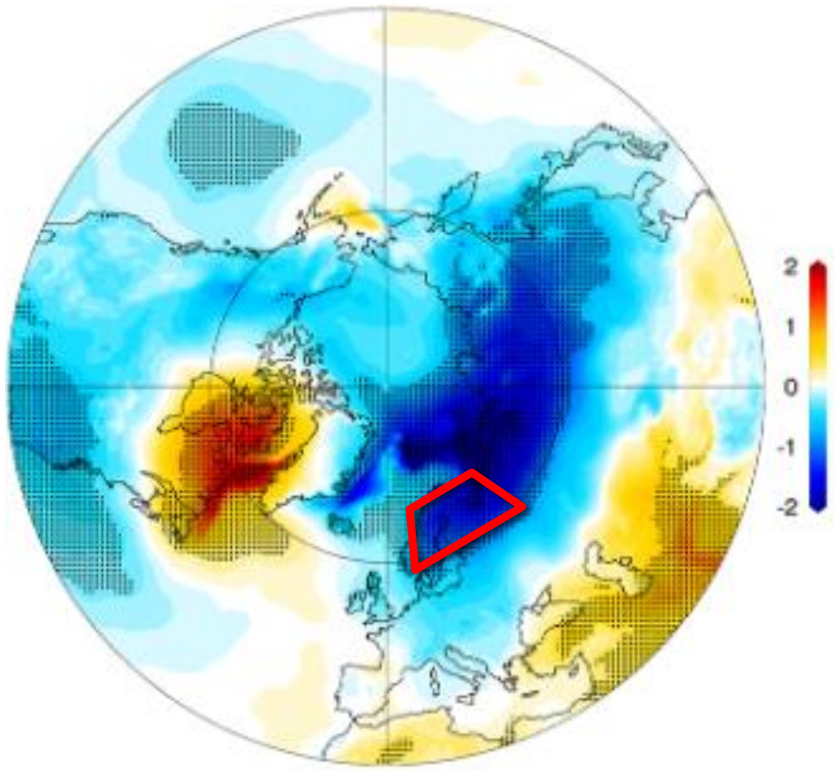
Forecasts of Opportunity

- When conducting cross-validation on the model, look at the average ACC across all predictions made
- Compare the average ACC of all predictions with the ACC for predictions with the highest confidence (Mayer and Barnes 2021)

90th Percentile

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Probability Cold	0.45	0.47	0.38	0.41	0.18	0.09	0.25
Probability Warm	0.55	0.53	0.62	0.59	0.82	0.91	0.75

- ✓ If the overall distribution of ACCs for the higher confidence events is greater than average ACCs, this indicates potential for identifying forecasts of opportunity
- ✓ The model does not have to be statistically skillful or highly accurate for there to be potential for forecasts of opportunity



Results

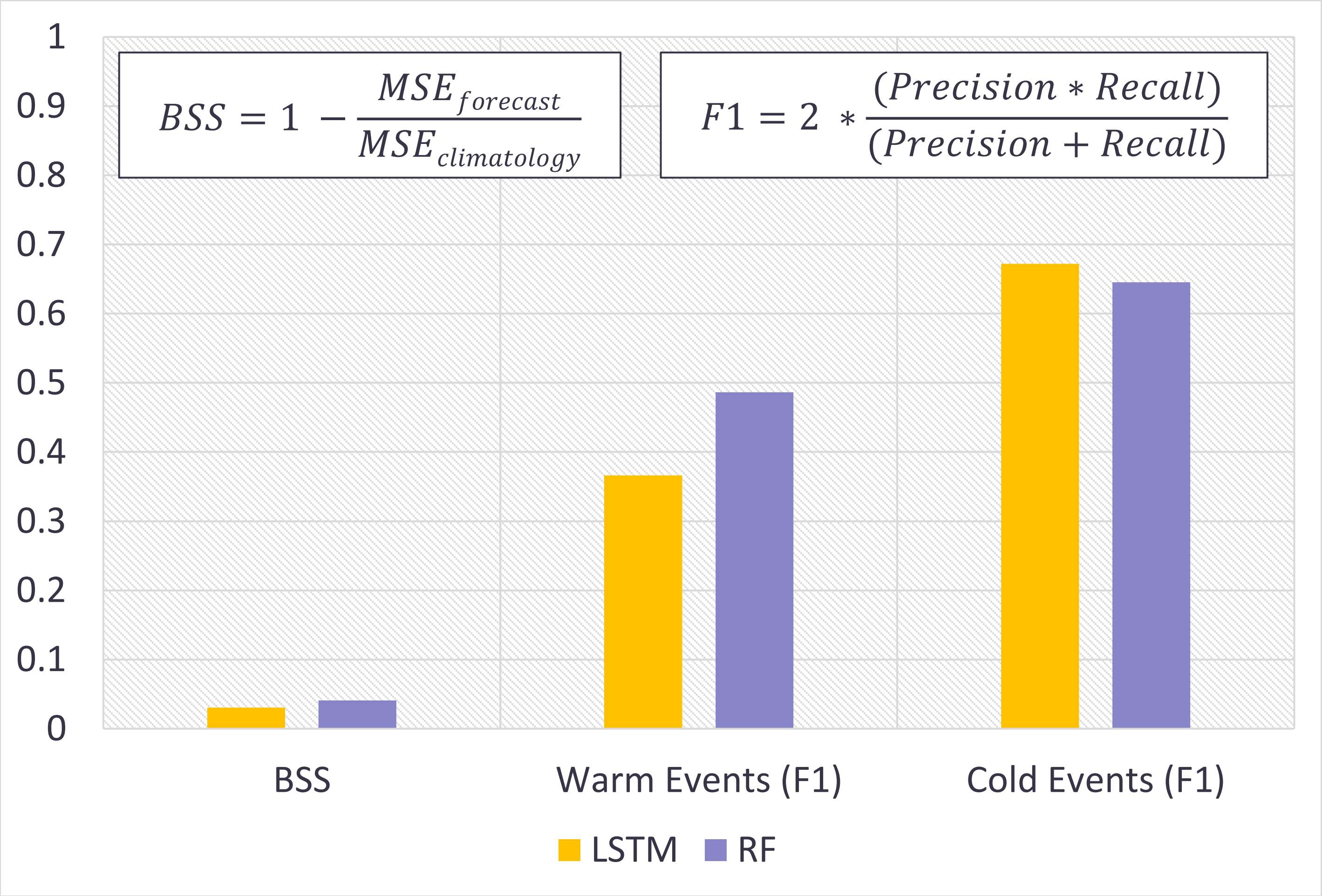
OF A 14-DAY FORECAST FOR

THE DESIGNATED EUROPEAN REGION

Skill Scores of Models

<u>Neither model is very skillful</u>

- BSS only slightly better than climatology (= 0)
- RF model improves over LSTM with F1-Scores near climatology for warm events (~0.5) and slightly “better" for cold events (~0.65)
  - Model can correctly identify and minimizes false alarms for cold events

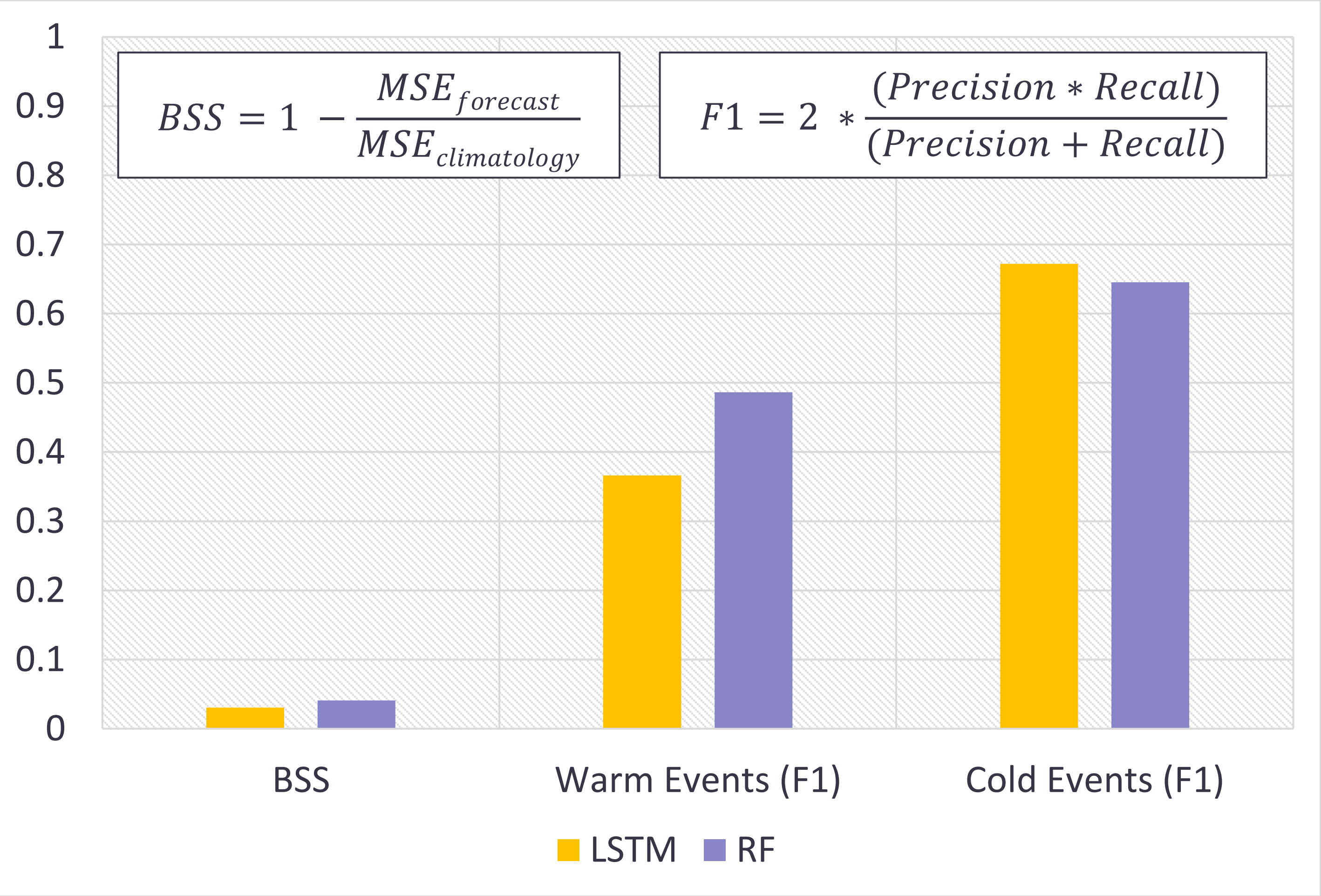


Skill Scores of Models

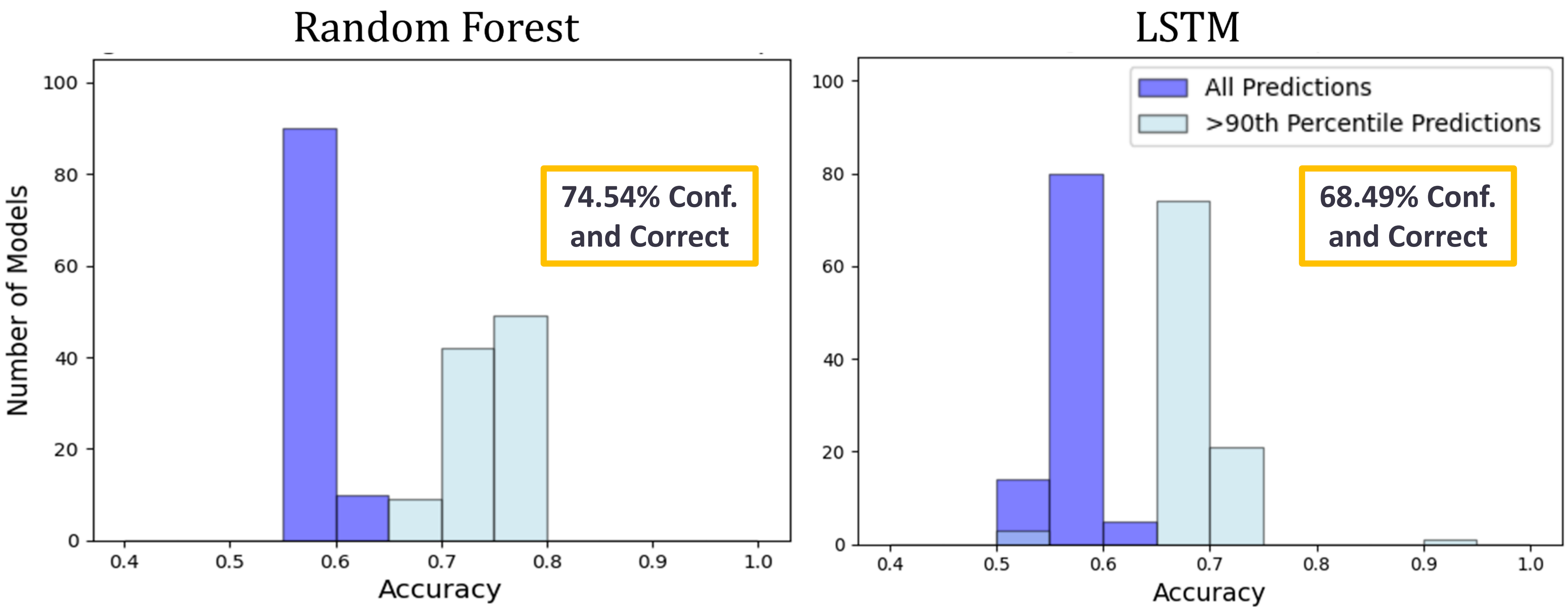
<u>Neither model is very skillful</u>

- BSS only slightly better than climatology (= 0)
- RF model improves over LSTM with F1-Scores near climatology for warm events (~0.5) and slightly “better" for cold events (~0.65)
  - Model can correctly identify and minimizes false alarms for cold events

What information is used by these models?
How do we interpret them?

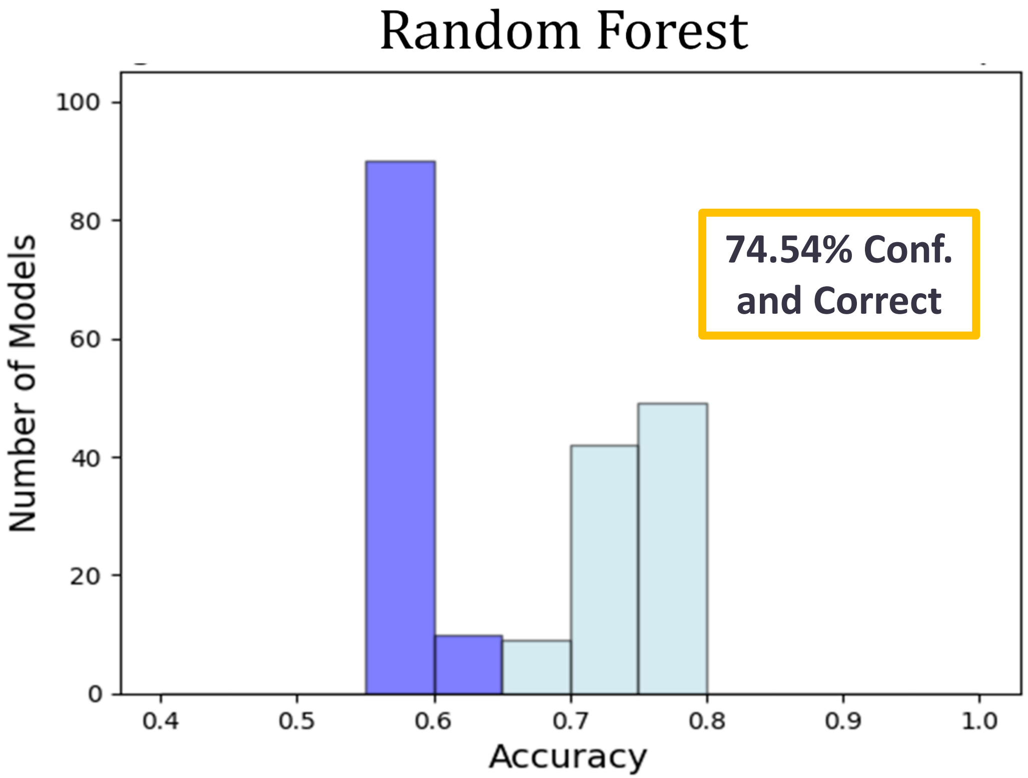


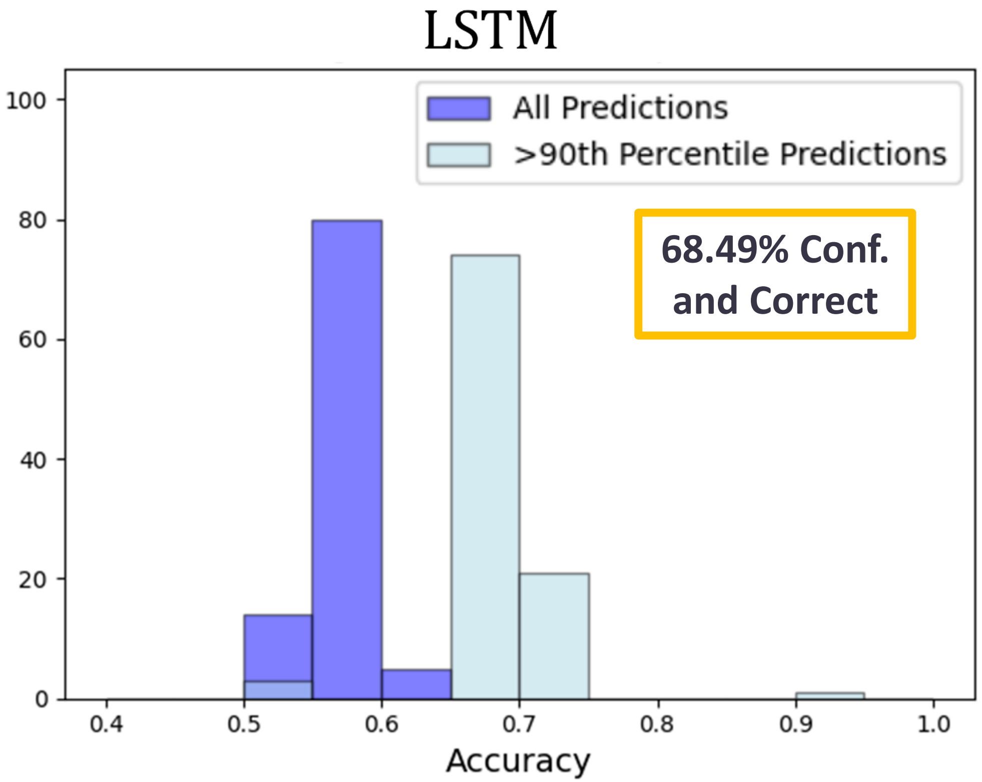
Forecasts of Opportunity



- RF model performs better in its ability to isolate highly confident and correct predictions for the sign of a +14-day temperature outcome in Eurasia
- Timeseries information **does not** impact the subseasonal predictive capabilities of these metrics

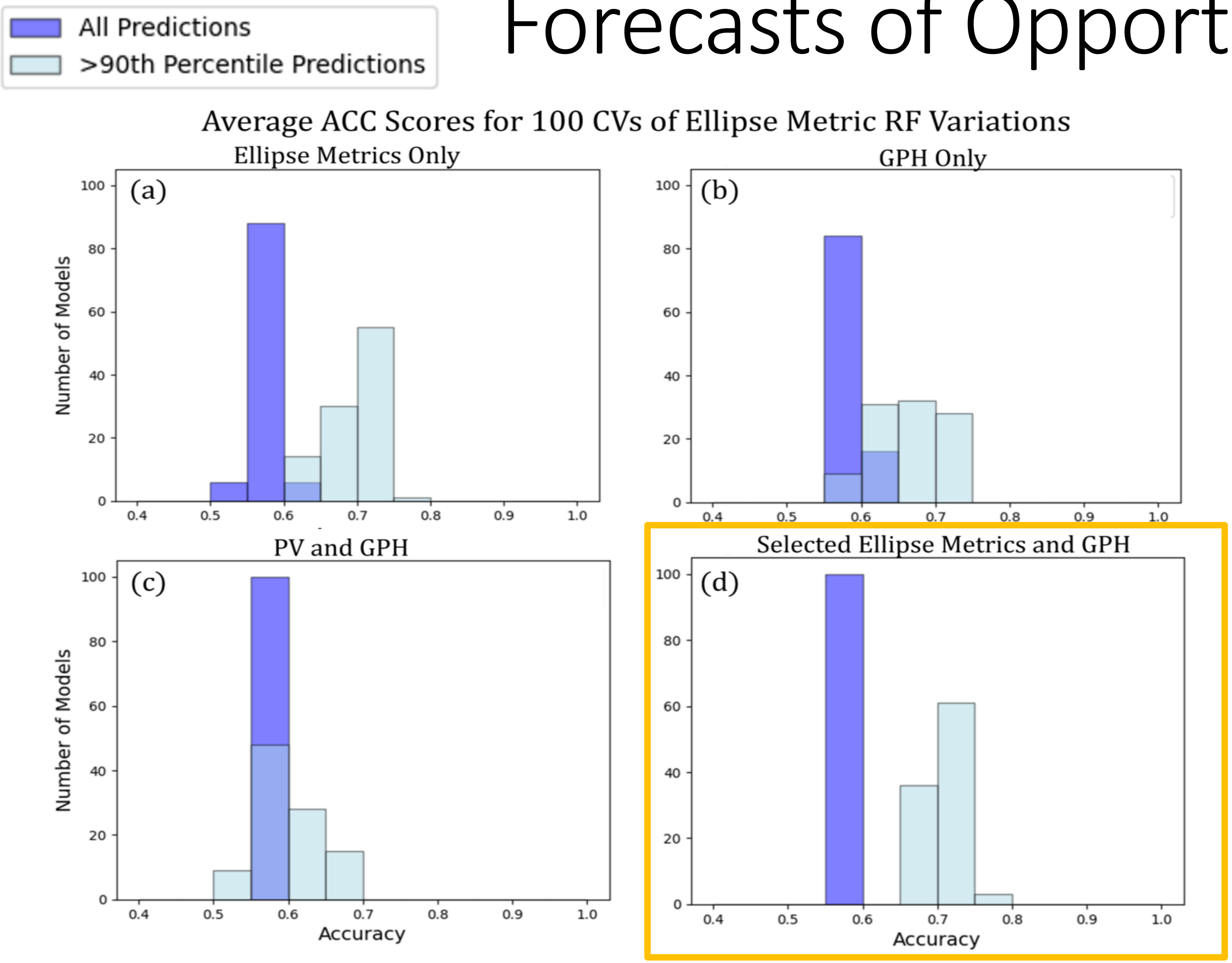
Forecasts of Opportunity





- RF model prioritized in further analysis since the LSTM does not improve the predictions for this application
  - Ease of interpretability compared to the black box nature of the LSTM
  - Higher confidence predictions for opportunistic events

Forecasts of Opportunity

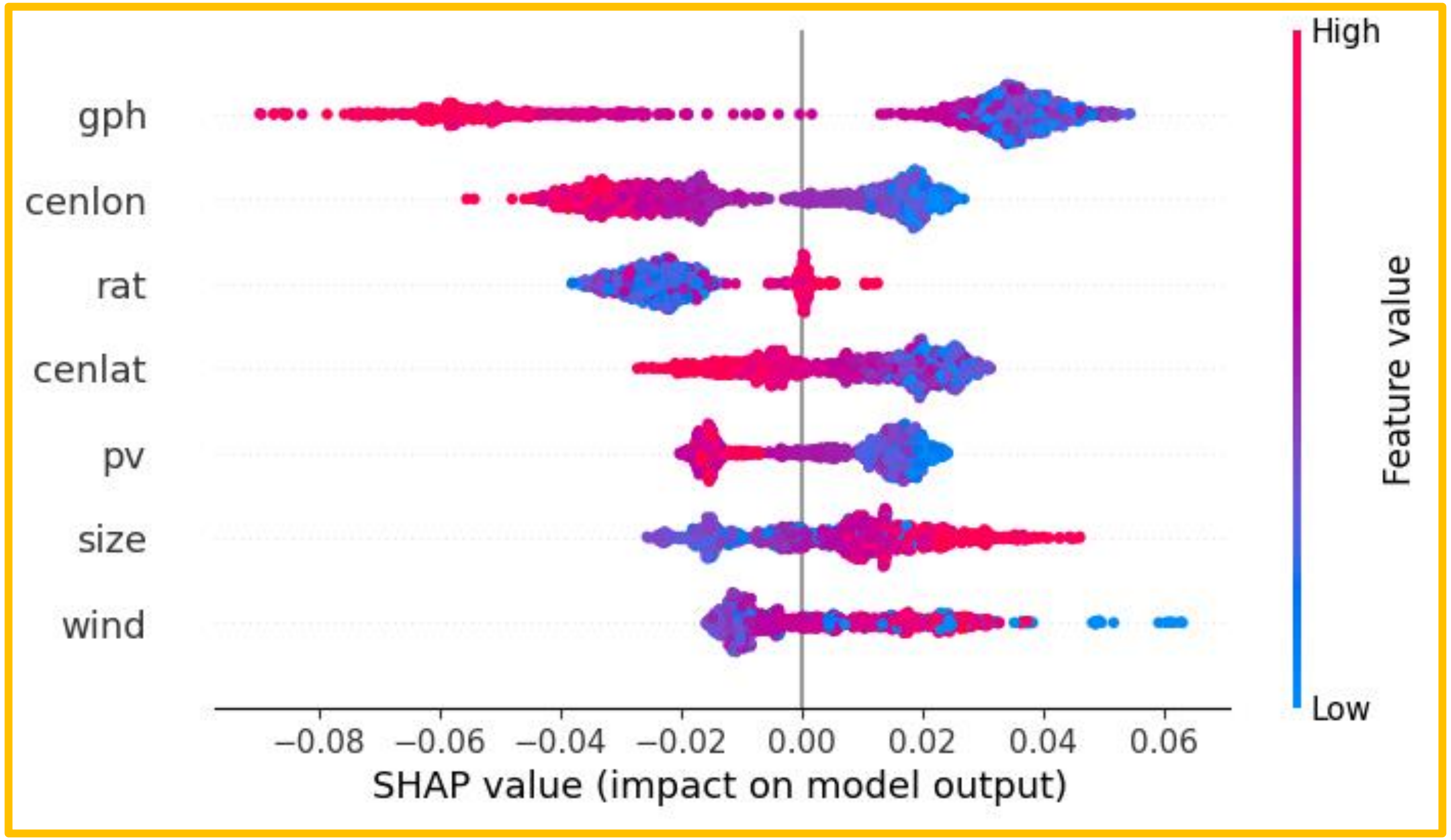


RF model with different combinations of input features

- Higher confidence events persist for combinations containing calculated ellipse metrics despite overall model accuracy remaining consistent
- This supports that there is some prescient information regarding the stratosphere and polar vortex strength, combined with lower stratosphere GPH, that is used in creating higher confidence predictions

Shapley Additive Explanations (SHAP)

SHAP for Negative Temperature Predictions



Explainable AI (XAI) techniques provide insight into the usage of the different input features within the RF

- Features are ranked down the y-axis in order of importance
- Assigned feature value and sign of SHAP value indicate whether positive or negative measures of each input feature are associated with the model's prediction
  - Feature and SHAP value combinations show physical information used by the model

Conclusions

- **Is it possible to use the stratospheric polar vortex ellipse metrics for predicting subseasonal temperatures and forecasts of opportunity in the NH?**

Yes, but not skillfully on average — stratospheric polar vortex ellipse metrics may be used to predict the sign of temperature anomalies at a 14-day lead for the observed European region.

- **Does the inclusion of timeseries information impact the applicability of these ellipse metrics?**

No — the LSTM model does not show any significant improvements in F1/BSS compared to the RF model and lacks in high-confidence predictions/forecasts of opportunity.

- **Are certain ellipse metrics more useful than others?**

Yes — specific stratospheric polar vortex ellipse metrics at 10hPa, coupled with 100hPa GPH in the N. Atlantic, are important for the RF model to present forecasts of opportunity.

Overall, this research highlights that the daily evolution and shape of stratospheric polar vortex provide information that is important for improving subseasonal outlooks of tropospheric temperatures

"Future" Work

- Random Forest model was used to predict the sign of temperature anomalies in two additional regions in the NH and at extended lead times (20 and 30 days)
  Come to see me talk about it this afternoon! – Poster #413

- Random Forest model is also being tested with ellipse metrics calculated using reforecasts from the Unified Forecast System's Subseasonal Prototypes
  - Evaluating the representation of the stratosphere in these physics-based models

- Prepare and defend dissertation – *April 10th, 2026!*

- Find a job – email me if you want to chat!

Email: emfernandez@albany.edu | GitHub: github.com/emf98

Special Thanks to UCAR CPAESS, NOAA WPO, and UAlbany for their support as I finish my degree!

Extra Slides

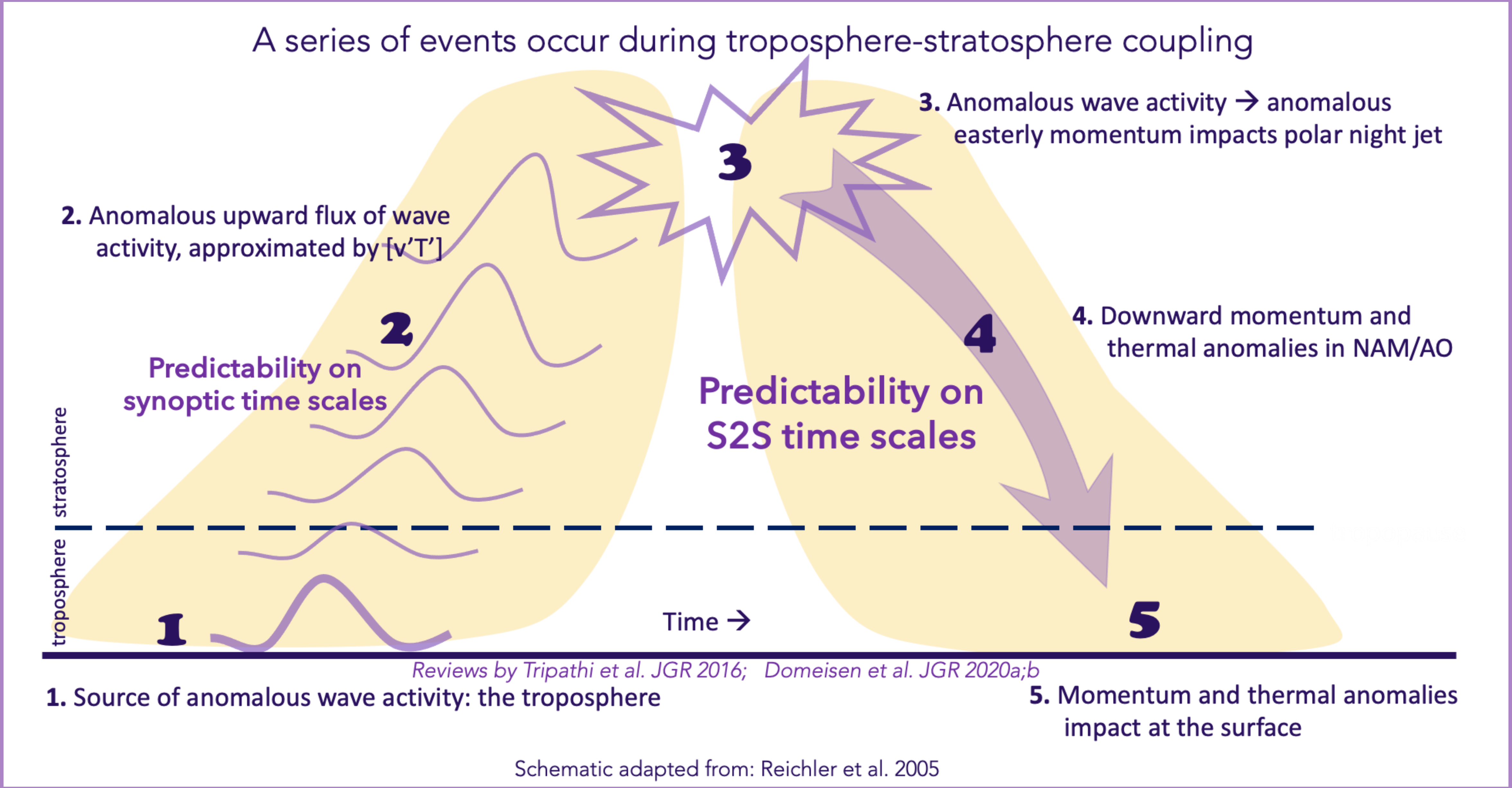
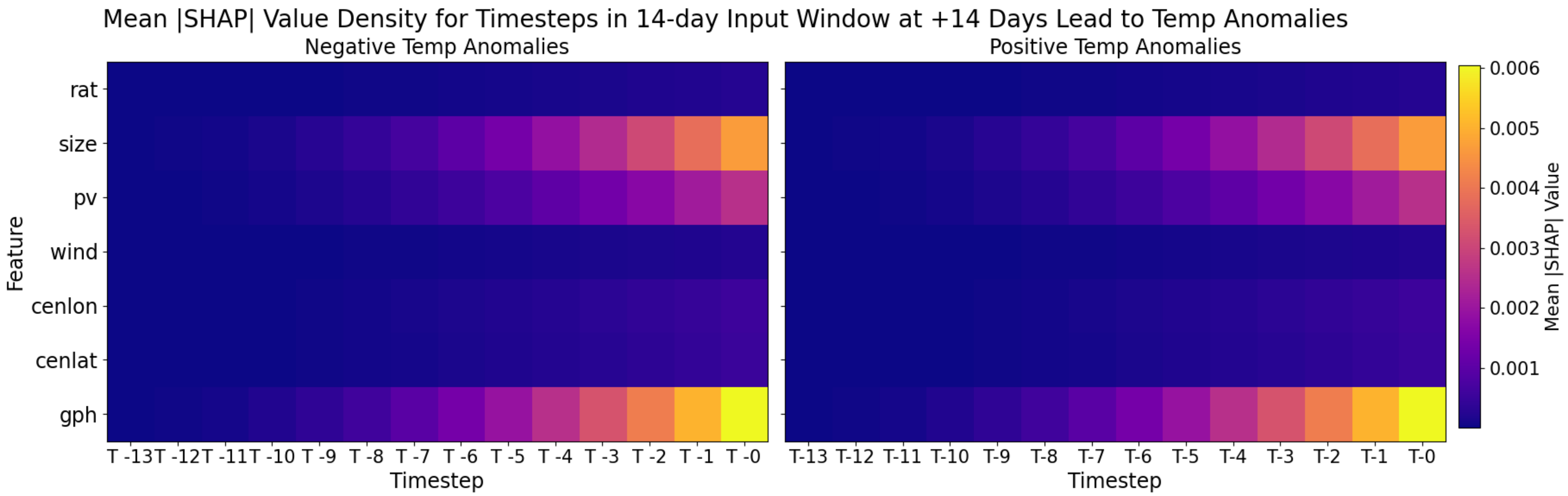


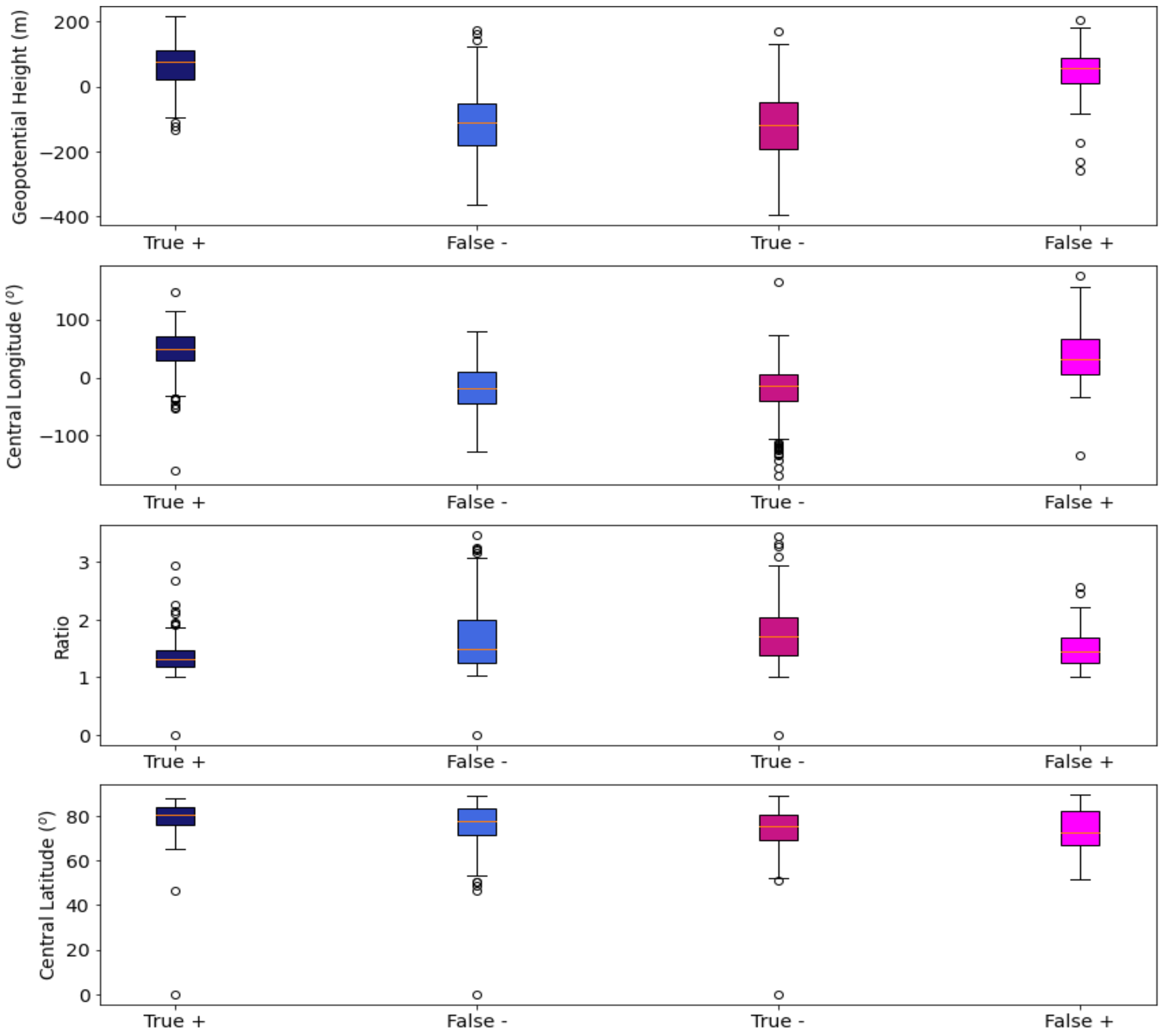
Figure courtesy of Andrea Lopez Lang

Shapley Additive Explanations (SHAP)

LSTM



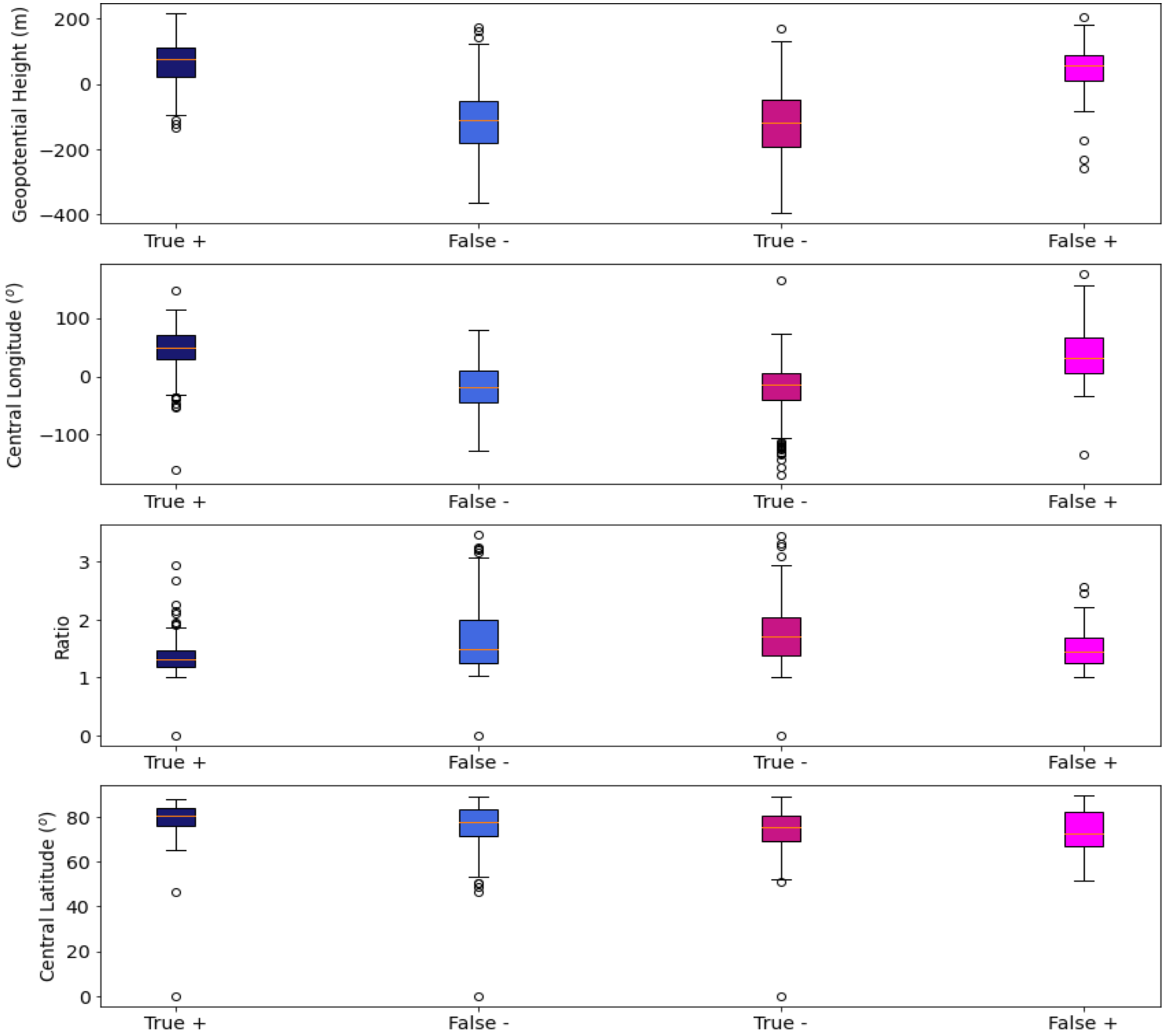
Physical Interpretability: Feature Distributions



Isolated dates corresponding to the four possible outcomes for temperature:

- Positive and Predicted Positive
- Positive and Predicted Negative
- Negative and Predicted Negative
- Negative and Predicted Positive

Physical Interpretability: Feature Distributions



Isolated dates corresponding to the four possible outcomes for temperature:

- Positive and Predicted Positive
- Positive and Predicted Negative
- Negative and Predicted Negative
- Negative and Predicted Positive

Key features:

- ✓ Notable changes in GPH and central longitude distinguishing positive and negative events
  - ✓ + = Higher GPH anomalies and eastward orientation
  - ✓ - = Lower GPH anomalies and westward orientation
- ✓ Larger distribution for ratios, less certainty

Physical Interpretability: 10-hPa GPH Anomalies During Temp Extremes