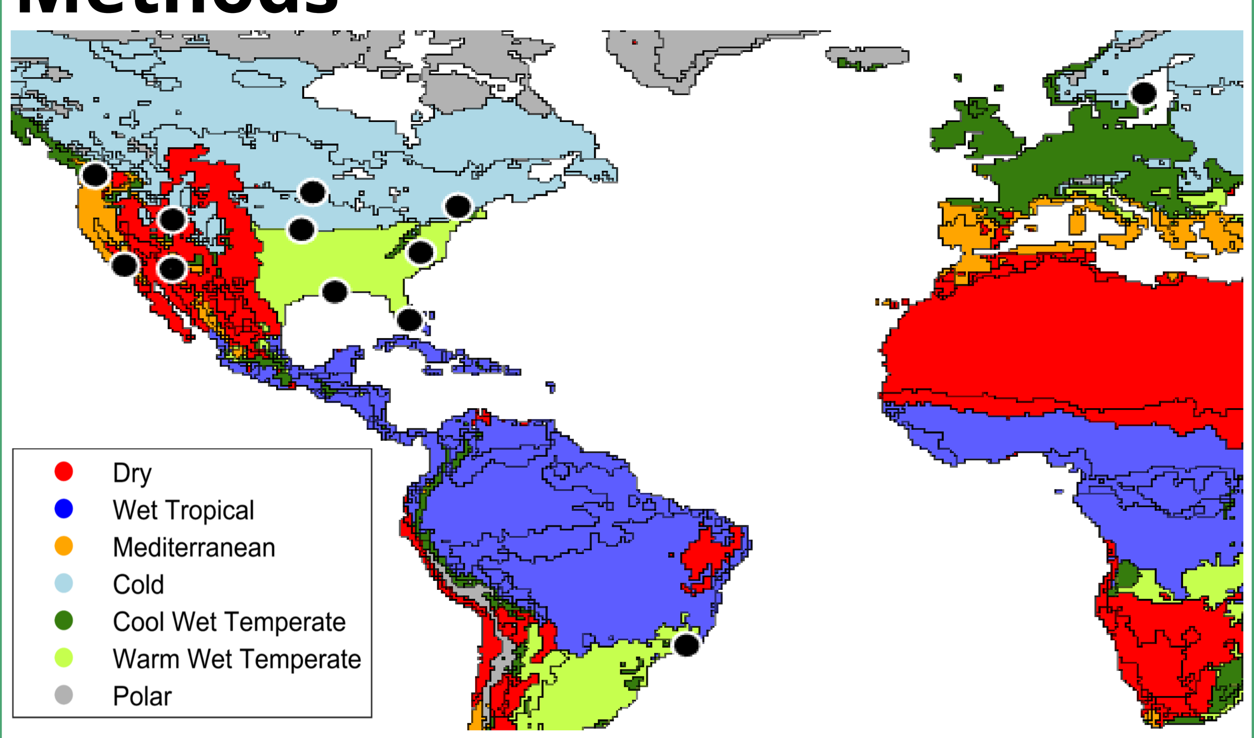
Steps Towards a Simple Neural Network to Improve GFS Temperature and Dew Point Forecasts

Cameron Gilbert[1], Sandra Yuter[1,2], Matthew Miller[1], Warren Lewis[1]
[1]Department of Marine, Earth, and Atmospheric Sciences and the [2]Center for Geospatial Analytics
North Carolina State University, Raleigh, NC

Motivation

NOAA NCEP Global Forecast System (GFS) forecasts have persistent model biases which are in part a function of time of day, season, and climate zone. The usability of GFS forecasts can be improved with post-processing bias corrections. Our goal is to determine the minimal complexity required for a statistical model to achieve reasonable accuracy. As an initial step, we assess the skill of machine learning vs regression to predict time series of temperature and dew point. This minimalist approach uses a small subset of carefully selected variables that relate to weather processes rather than the "kitchen sink" approach that uses large data sets that often include poor quality and irrelevant data.

Methods



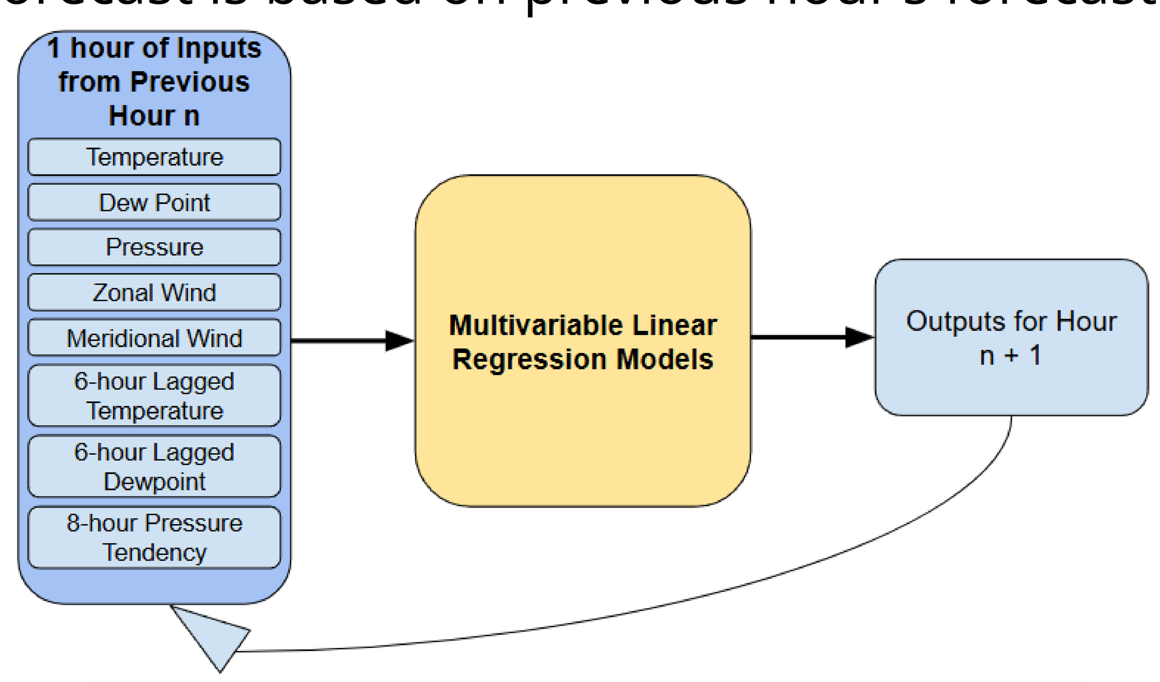
We train and test the models using hourly observations from 12 stations across different climate zones. Each location is handled separately with its own regression and neural network models for temperature prediction and for dewpoint prediction. The training data are from 1/1/2022 through 9/18/2023. The testing data are from 9/19/2023 through 11/9/2024. The skill of the models is assessed by making a time series prediction at each of the 12 locations for 100 random days from the test data set and analyzing the predictions compared to observed conditions. We calculate the median Root Mean Squared Error (RMSE) and bias over each 24-hour forecast period for each station.

Statistical Models

For this analysis there are 12 stations x 4 models (regression for temperature, regression for dewpoint, neural net for temperature, neural net for dewpoint) = 48 total models. Each mode is trained on 10 input variables for their specific station.

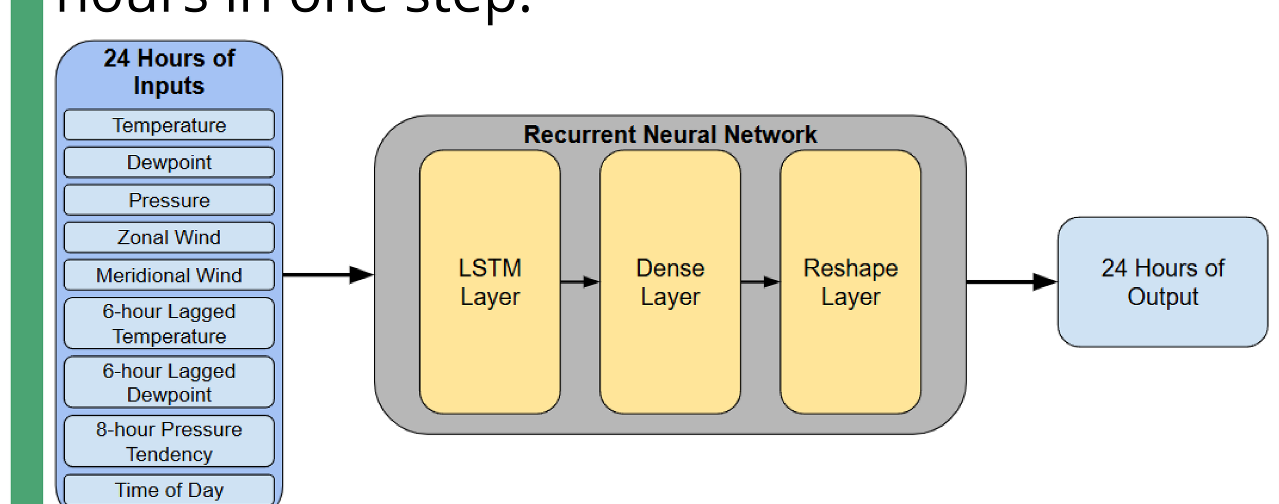
Multivariable Regression

In test mode, input values are from the *previous hour n* and the regression model generates just the next hour's value. This cycle is repeated 24 times to output values for forecast hours *n+1* to *n+24*. Starting with *n+2*, forecast is based on previous hour's forecast.



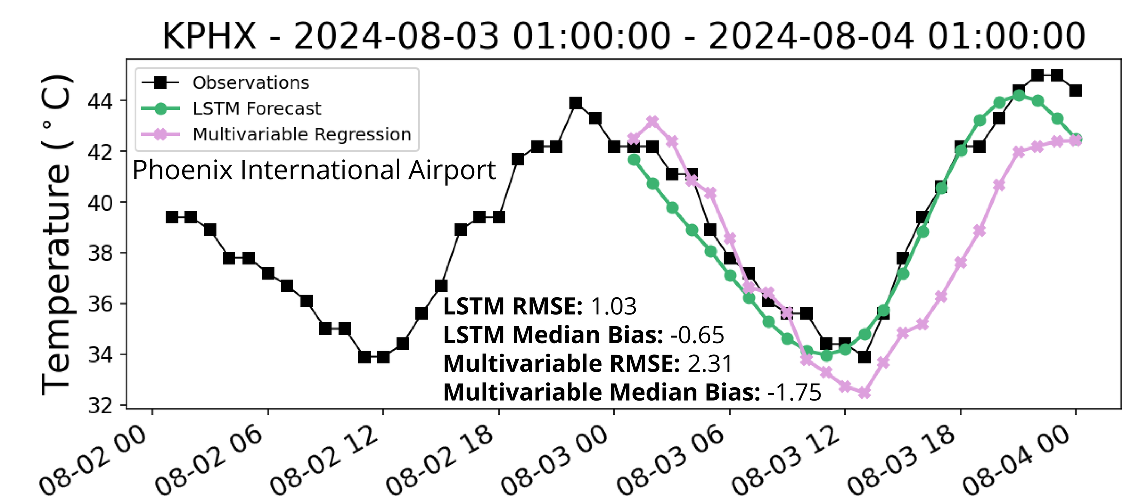
Neural Network

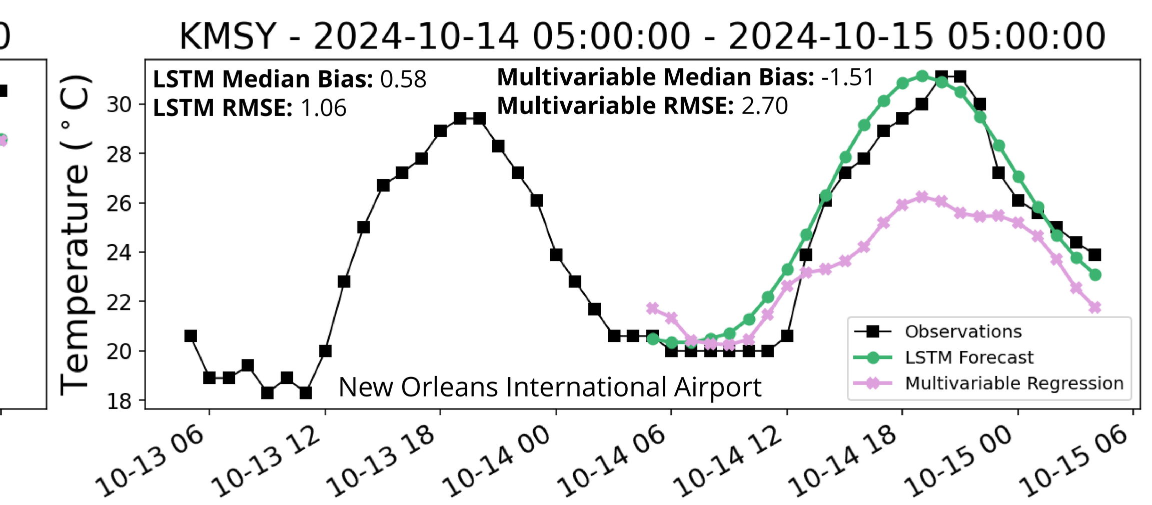
We use a recurrent neural network called a LSTM (Long Short-Term Memory) with three layers. In test mode, this model takes the previous 24 hours of observations as input, and outputs the prediction for the next 24-hours in one step.

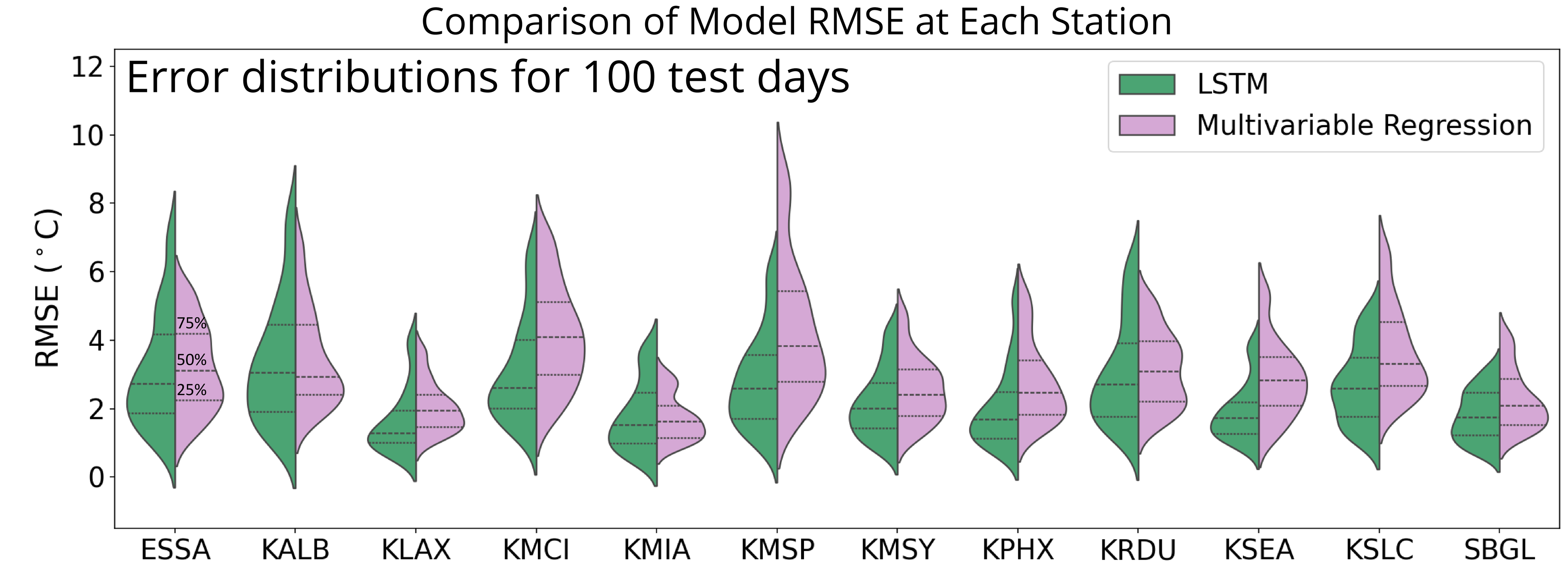


Comparisons of Model Performance

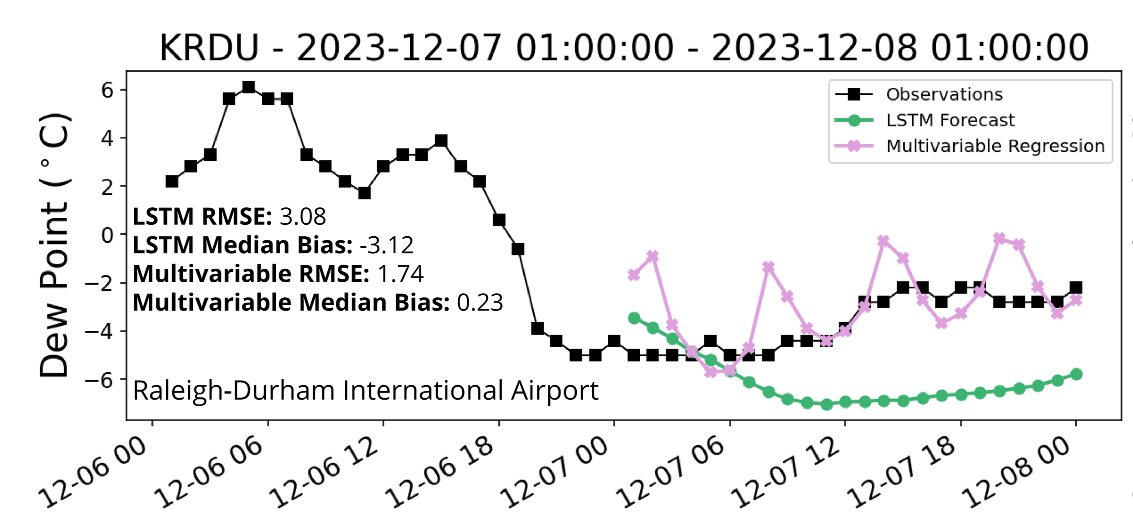
Temperature

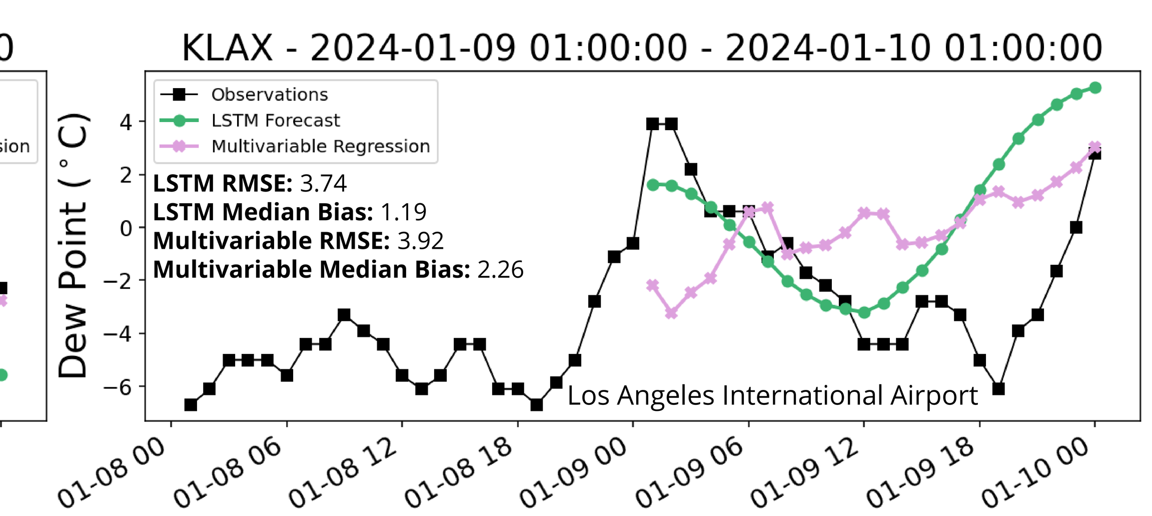


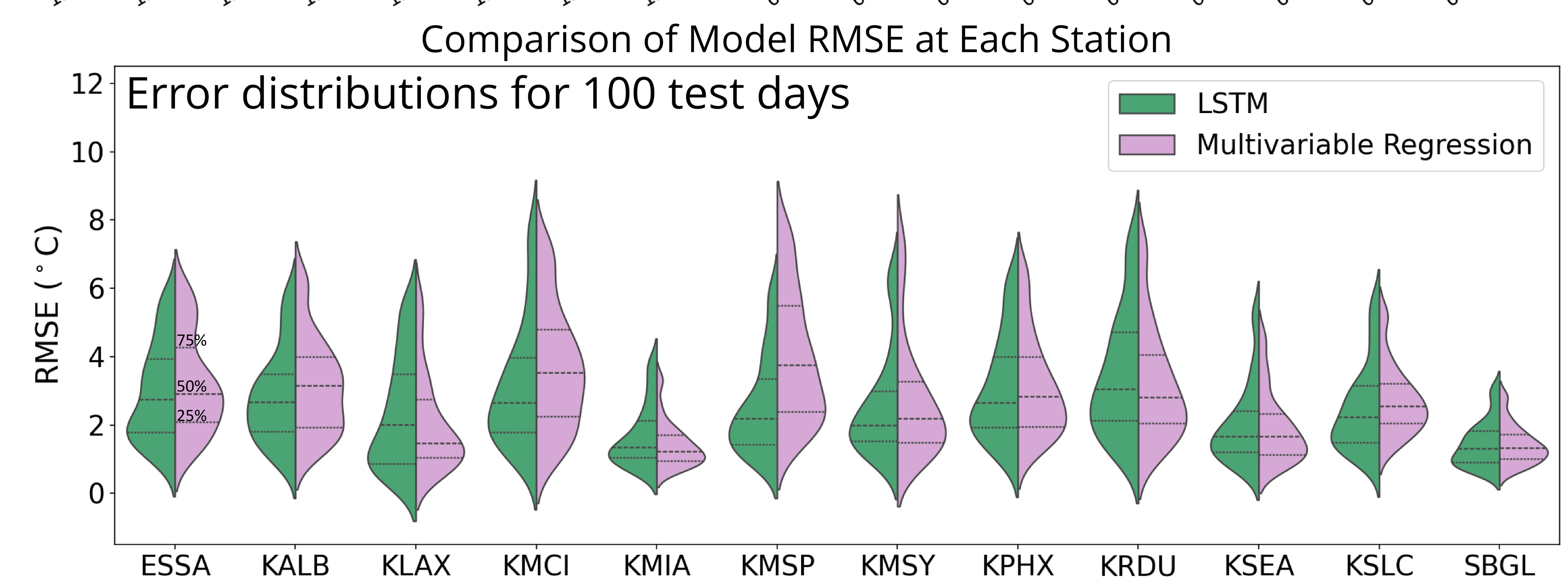




Dewpoint







Summary

Error Statistics (° C)	Multivariable Linear Regression		Recurrent Neural Network (LSTM)	
	Median RMSE	Median Bias	Median RMSE	Median Bias
Temperature Average:	2.81	-0.10	2.19	-0.71
Dew Point Average:	2.45	-0.26	2.21	-0.61

- Overall, for our 12 stations the LSTM model is better than the regression model in terms of RMSE but worse for bias.

Future Work

Test other types of statistical models more suited for time-series forecasting, increase the training data period and number of stations. Train the models to predict model biases.

Acknowledgements
Thank you to Wayne Johnson, Rebecca Moore, McKenzie Sevier, Nathan Faulk, and Logan McLaurin for their feedback. This work is supported by the NCSU Provost Professional Experience Program and ONR N00014-24-1-2216.