Weather and Climate Prediction using Machine Learning

Ananda Badari, Neha Bhatia, Yeojin Chang, Isco Cheung, Sabrina Edwards-Swart
CSE 6242 Georgia Institute of Technology

Background

Why is predicting weather important?

Historically, providing reliable weather predictions to public has been a crucial component in society-operation. Accurate forecasts can help mitigate weather-related losses and maximize planning efficiency in aspects of economy, agriculture and environment. From saving lives in times of natural disaster, making plans for trade and agriculture, or deciding on an outfit in the morning, weather predictions are an essential tool for everyone.

How have people predicted weather in the past?

- **Statistical approaches** based on past observations, as well as numerical models incorporating randomness were some rudimentary methods conducted in the past.
- Advancements in **satellite observation** and **high-performance computers** have revolutionized the quality of weather forecasting and climate modeling.
- In the late 2010s, scientists experimented with **machine learning models** such as regression, SVM, PCA, and different neural network frameworks.
- Most recently, there have been approaches in **Bayesian deep learning** and **uncertainty quantification**, as well as graph neural networks to account for weather forecasting dependence in location.

Problem Statement

Given a historical dataset of various weather features in past 40 years, we **predict temperature and precipitation** in the United States over a one-month time-frame on an hourly scale and display our predictions using a powerful visualization.

How have we improved upon past methods?

1. **Increasing the range of weather prediction to a month**
   Most state-of-the-art models forecast for the upcoming 3-5 days, and we expanded this range to up to 30 days. To do this, we sacrificed spatial resolution to extend the accuracy of hourly predictions into the future.

2. **Utilizing ensemble methods in deep learning**
   Ensemble weather forecasting continues to be the standard approach to Monte Carlo Analysis of dynamical weather systems in physics-based models. This leads to another proposed innovation: use of ensemble stacking with lead forecasting time as a parameter. Generally, this entails training multiple versions of the models, each with different spatial resolutions, as well as models with different architectures, and then performing stacking generalization using a combiner algorithm.

Data Source

- The **WeatherBench** Dataset consists of preprocessed meteorological data from ERA5. ERA5 provides hourly readings of atmospheric, land and oceanic climate variables dating back to 1975. The total size of all provided data is 5.8 TB.
- The data cover the Earth on a 30km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80km.
- This data set includes information on
  - wind direction
  - wind speed
  - temperature
  - geopotential
  - vorticity
  - humidity
  - solar radiation
  - cloud cover
  - precipitation

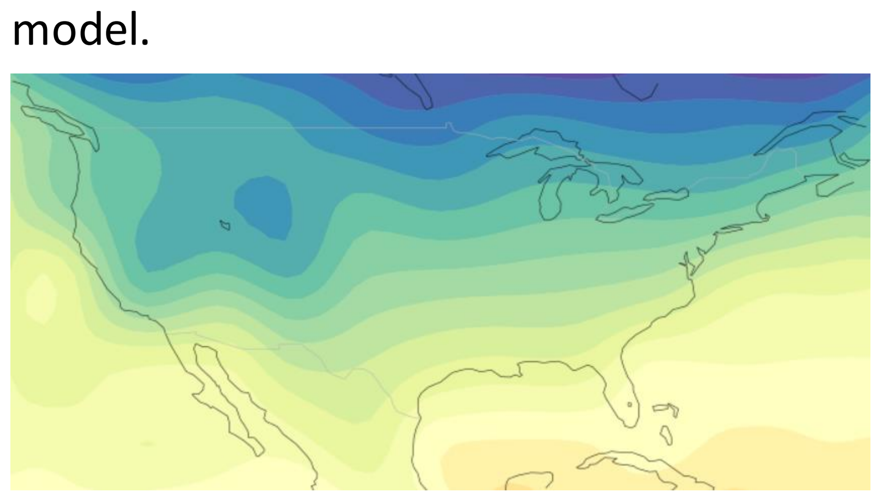
Approaches & Experiments

Data Preprocessing & Baseline Modeling

We used the Python package CliMetLab to simplify and visualize the data.

To provide a baseline for more advanced models, we created simple persistence models, climatology models, and a linear regression model.

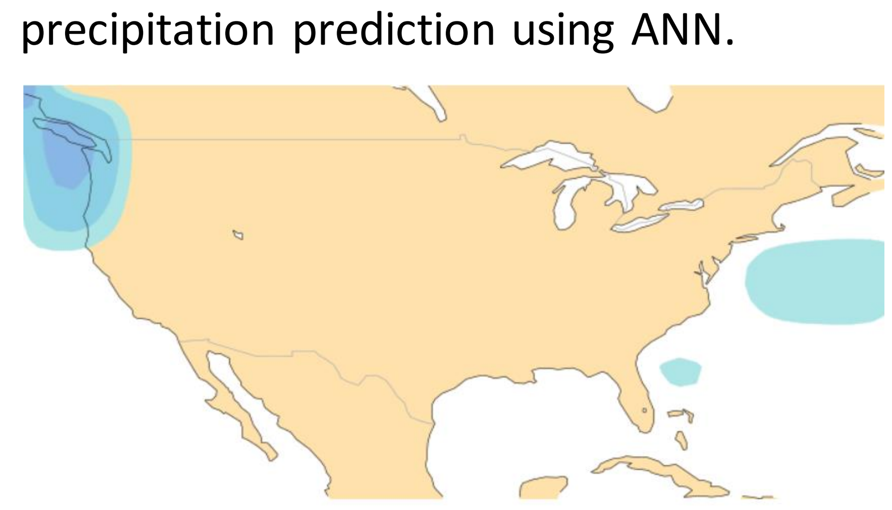
The following is a visualization of a temperature prediction using the climatology model.



Artificial Neural Network

Artificial Neural Network was built, with five hidden layers consisting of eight neurons each and a tan sigmoid transfer function.

The following is a visualization of precipitation prediction using ANN.



ResNet-19 Model

A residual neural network is a type of ANN with hundreds of layers. We implemented this model using Keras in TensorFlow.

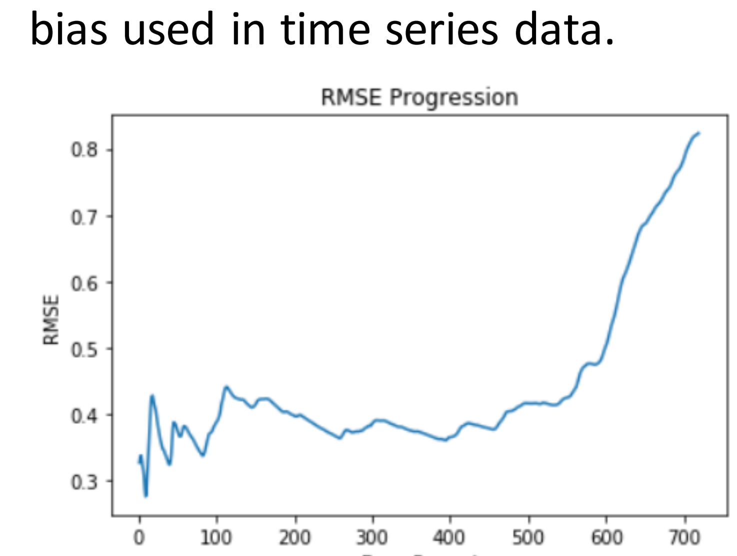
Ensemble Learning

We performed a stacking ensemble over the results of the ANN and ResNet Models as well as the baseline models.

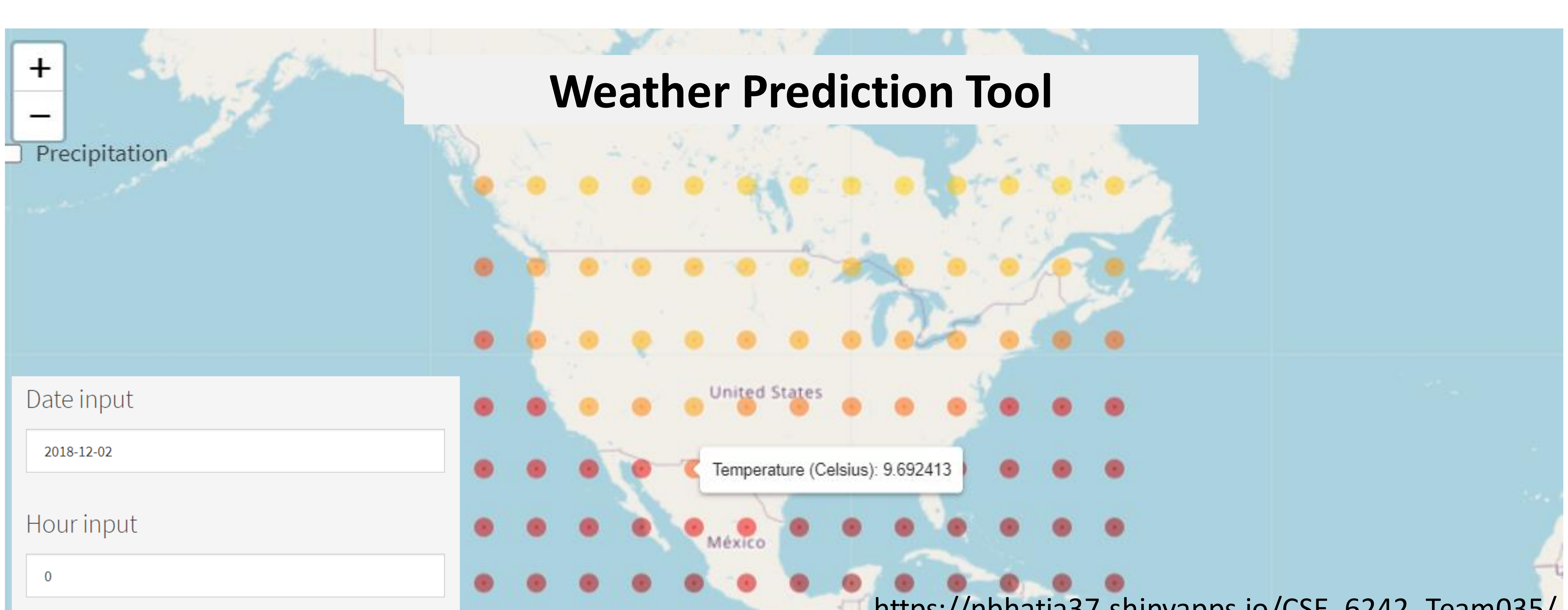
Experiments & Evaluation

We evaluated each of these methods using

- root mean squared error (RMSE), a measure of accuracy
- and tracking signal (TS), a measure of bias used in time series data.



Results & Visualization



https://nbhatia37.shinyapps.io/CSE_6242_Team035/

We built our visualization using R Shiny, which is an R package that facilitates the development of web apps using the R language. Shiny apps can be extended and integrated with CSS, Javascript, and HTML.

- This tool shows the
  - predicted temperature (in degrees Celsius) and
  - precipitation (in cm) for regions in the United States of America during December 2018.
- Users can input a date between December 02, 2018 and December 31, 2018, as well as an hour between 0 and 23.
- Predicted precipitation can be toggled by clicking the 'precipitation' checkbox in the top left corner of the map.
- Hovering your mouse over a circle on the map will show the exact predicted temperature or precipitation for that area.