

Weather and Climate Prediction using Machine Learning

Ananda Badari, Neha Bhatia, Yeojin Chang, Sabrina Edwards-Swart, Jiahong Zhang
Georgia Institute of Technology
North Ave NW, Atlanta, GA 30332
`{abadari3, nbhatia37, chang.yeojin, sjess, jzhang3399}@gatech.edu`

Abstract

Traditional weather prediction has been performed using physics-based models based on dynamical systems. In recent years, the abundance of atmospheric data and computing power gives way to machine learning methods. In this paper, we suggest an ensemble learning approach that combines the predictions of multiple independently trained machine learning models, specifically an Artificial Neural Network and a Deep Residual Convolutional Neural Network to predict temperature and precipitation over a month. Our results can be viewed [here](#).

1. Introduction

Humanity has long tried to predict the weather, as the weather defines the environment around us. The influences and effects brought by weather prediction are not only in choosing clothes or determining to take an umbrella, but also in the various and diverse aspects of development. Peasantry needs weather prediction to deal with the farming arrangement. Environmentalists apply weather prediction to foresee climate changes for global warming control. Meanwhile, weather changing is a crucial consideration for the physical and chemical materials used in infrastructure. Knowing the weather affects everything from trade to agriculture to all of us who check the weather apps on our phones, predicting the weather has an immense impact.

Performing weather forecasting is a unique problem, as data is abundant, but there is a lot of error in standard models, as weather is inherently chaotic and unpredictable. Additionally, as there is a large amount of data, good forecasting requires vast computational power. As a result, our model-building approach will follow more of a proof of concept of some key innovations on standard methods rather than trying to compete with the state-of-the-art. We will also create a platform for visualizing our predictions in an easily interpretable way.

1.1. Problem Statement

Given a historical dataset of various weather features, we predict temperature and precipitation in the United States over a one month time-frame and display our predictions using a useful visualization.

2. Literature Survey

Modeling the Earth's climate is a difficult task, particularly since the factors at work are so complex. The basic concepts behind numerical forecasting and climate modeling have been around for more than a century. The essential ideas of mathematical weather forecasting and climate modeling were developed before modern computing systems were constructed. By the 1950s, most meteorological forecasting centers were using both statistical approaches (predictions based on past observations) and numerical models (unpredictable calculations based on initial conditions). "Analysis methods for numerical weather prediction" presents Bayesian probabilistic arguments for weather prediction [8]. The field of climate science is constantly evolving, and our understanding grows rapidly as complex computer simulations advance. Progress in weather forecasting and climate modeling over the past 50 years has been dramatic, and the field has been completely revolutionized by satellite observation and high-performance computers. We explore applications of such models in "The origins of computer weather prediction and climate modeling" [9].

The importance of visualization is emphasized in "Coupling advanced modeling and visualization to improve high-impact tropical weather prediction", as Shen and Green discusses how NASA hardware was used to better predict severe tropical weather [14], and how visualization was instrumental in conveying this data. Although the practical application of this paper is predicting severe weather using a supercomputer, the data analysis methods used here apply broadly to any big-data weather projects. Broadly speaking, this can be summarized as dividing an area into a grid and using parallel computing (which, while better with more cores, is still possible on nearly all current consumer-grade computers).

grade computers) to process information. Unfortunately, we do not have access to a supercomputer, so this paper will primarily inform our visualizations.

In “Predicting solar generation from Weather forecasts using machine learning,” scientists used machine learning methods to predict the availability of solar power in the future. They used features like seasonality, wind speed, dew point, sky cover, relative humidity, and other weather attributes to measure the solar intensity of certain days. The prediction methods used linear least squares regression and support vector machines with PCA for dimensionality reduction [13]. This paper was published in 2011 used relatively rudimentary methods from today’s standpoint, and we will be expanding such approaches by using Neural Networks.

In “Contrast Pattern Based methods for Visualizing and Predicting Spatiotemporal Events”, researchers developed a new method using contrast patterns (regularities in data). By measuring their growth rate relative to space and time, then clustering appropriately, scientists were able to get more accurate predictions on burglaries, crime, and extreme rainfall events [19]. As weather is also a spatiotemporal event, such approaches could be useful for future work. In “The Quiet Revolution of Numerical Weather Prediction”, the approach is still computationally-intensive forecasting based on physical and chemical processes. This paper provides relevant factors to consider when developing better climatology-affecting features when predicting weather, such as the rising presence of trace gas and aerosols impacting temperature and air quality. The analogy comparing weather prediction to a simulation of the human brain is presented, informing a possible neural network approach [3].

In “Weather - Temperature Pattern Prediction and Anomaly Identification using Artificial Neural Network”, one of the first approaches using neural networks is presented [18]. Here, weather forecasting is done by using historical data, rather than traditional methods that use dynamic systems approaches. The novel method presented here uses Artificial Neural Networks to predict daily temperature patterns. This paper only applied the methodology to daily temperature data, we will be expanding to predict precipitation as well.

The year 2018 brings three papers that innovate on the standard Artificial Neural Networks of the past. “Weather Data Visualization and Analytic Platform” uses satellite imagery and temperature stations to make an air temperature prediction model/visualization for Armenia [2]. This paper also had a special focus on visualization its results. “Visualization of Neural Network Predictions for Weather Forecasting” presents a design to evaluate and visualize the quality of different forecasting models, using a time-dependent analysis of one variable. [12]. “Spatio-temporal Stacked STM for Temperature Prediction in Weather Forecasting”

presents a new architecture using LSTM layers rather than SVMs, ANNs, or MLPs [6]. These focus on predicting solely temperature, while we wish to predict precipitation as well.

The field of weather forecasting has gone through many advances during the 2010s, and “Advances in weather prediction” reviews the progress made during the decade [1]. At the point of publication in 2019, data is more ubiquitous and computing power accessible, but weather still is not reliably predictable over 10 days. “Predictability of Weather and Climate” focuses on the limitations behind forecasting, [7]. Specifically, it’s really hard to predict weather past 10 days in the future due to the chaotic nature of weather. This shows us that we need to look at short-term and longer-term weather prediction differently. We will take this into account by losing spatial-resolution to forecast general temperature and precipitation over longer temporal-resolutions.

The advances were not just in algorithms. In 2020, a new focus on weather data is reflected in papers such as “WeatherBench A Benchmark Data Set for Data-Driven Weather Forecasting” and “Data-driven medium-range weather prediction with a Reset pretrained on climate simulations, A new model for WeatherBench” [10, 11]. These papers propose a benchmark dataset of data processed from the historical ERA5 dataset, for use in Machine Learning models. Also provides evaluation metrics and baseline prediction scores, which is very useful to evaluate our models.

Around this time we see deep learning models outperform standard weather forecasting, specifically in papers like “Long-Term Weather Prediction Based on GA-BP Neural Network” [4]. Using Deep Learning also brought further performance benefits. In “MetNet A Neural Weather Model for Precipitation Forecasting”, MetNet outperforms the at the time state-of-the-art NOAA prediction models by up to 7-8 hours ahead and makes a predictions in seconds as opposed hours [17].

2022 brought more of a focus on different model architectures. “A Survey of Uncertainty Quantification in Machine Learning for Space Weather Prediction” applies Bayesian Deep Learning and Uncertainty Quantification for forecasting, a popular point of failure for deep learning approaches [15], while “Forecasting Global Weather with Graph Neural Networks” uses GNNs with the ERA5 dataset. This adds a graph-style approach for location based weather, as weather in different locations are not independent from each other. Finally, “Temperature forecasting by deep learning methods” presents a RNN using convolutional filters for high-performing temperature forecasting, while “Short-range forecasts of global precipitation using deep learning-augmented numerical weather prediction” is the current state-of-the-art model that uses a cubed-sphere projection with deep learning hybrid model to accurately predict precipitation [5, 16].

3. Methodology

We decided to utilize a novel ensemble learning approach to medium-term (up to 30 days) weather forecasting, using a variety of innovations in our technical approach.

- Data Augmentation with Rotations and Averaging.
- Decreasing spatial resolution for more accurate but generalized predictions longer than 10 days.
- Ensemble Approach with an Artificial Neural Network and a Deep Residual Convolutional Neural Network.
- Intuitive visualization of results and evaluations.

At the beginning of our project, we predicted that our model should perform better than existing models for medium-term weather prediction. At the core of this is the trade-off between spatial resolution and accuracy over time. By reducing the granularity of our data, we intuit that our accuracy will extend farther in to the future.

Since most weather prediction models focus on short term weather prediction- primarily the next several days, with a possible secondary and decreased focus out to 7 to 10 days- our model focusing on medium-term weather prediction should be better than other models for medium-term weather prediction because a machine learning approach to weather prediction has typically only been practiced on short-term prediction models.

Another aspect of classical weather forecasting that has not yet been implemented in the current state of the art deep learning forecasting models is using ensemble methods. 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 predicted 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.

This approach provides many advantages over standard approaches. Firstly, each of the underlying models can be made simpler, as they are part of an ensemble, and can be trained independently of each other. Then, these independently trained models can be combined for different kinds of predictions, for instance a combiner algorithm that weights lower spatial resolution would be better for longer term climatology forecasts, while one for shorter meteorological forecasts could prefer standard approaches.

For our purposes, we will be focused on predicting weather during the last available month of data in the dataset, December 2018. We will be using all previous data as either training or validation data. We are aiming to predict temperature and precipitation for the entirety of December.

3.1. Dataset, Preprocessing, and Augmentation

Our primary dataset is the WeatherBench dataset, which contains preprocessed weather data from the ERA5 dataset, which provides hourly estimates of a large number of climate variables from 1959 to present. We will be using data from the last 10 years. The data contains wind speed, temperature, humidity, cloud cover, solar radiation, precipitation, and vorticity. We will only be looking at the temperature and precipitation features of the data. To help us process the data, we have used TensorFlow. Additionally, we have removed any rows with incomplete data. We have also augmented the data with running window averages for longer term climatology forecasting. In our implementation, we specifically used 2m-temperature and total-precipitation, at the 5.625 degree resolution. The reasons for this were practical, as we do not have the hardware to train models on data in the Terabytes. However, using more or all features of previous data could be an avenue of future work. Additionally, we tried to implement rotations of the original data set [18] and found that performance in almost all cases was bad. This is primarily due to the fact that we are taking time series data and running residual network predictions on not just time-series data, but also location specific data, rather than using visual data in the form of images and performing rotation-equivariant convolutions. Finally, as we set our scope of prediction to be the continental United States, we cropped our data to a bounding box consisting of the continental United States. An example of this boundary can be seen in Figure (3). The final dimension of our processed data was (350633, 7, 13).

Additionally, we had to prepare the data to be used as training data for machine learning models. We defined Data Generator objects to be able to pass this data into our models using batch sizes of 32, and normalizing the data. For a majority of weather forecasting, predictions based on normalized data outperform those using raw data, so we normalized all of our data, keeping track of our population mean and standard deviation, so that we could un-normalize our forecasts. Additionally, when performing training, we elected to shuffle the order of the batches, to get a more representative sample, but did not do so for our validation dataset.

3.2. Ensemble Learning with ANN and ResNet19

We created two models to use as baseline comparisons: a persistence model and a climatological model. A persistence model uses the previous day's weather as the current day's weather, and a climatological model uses averaged historical data to make a prediction based on the area's climate while ignoring recent day-to-day weather. Both of these models- which are versions of how humans instinctively predict weather without using instruments (for example, we expect Seattle will be rainy and Phoenix will be dry,

and we are likely to believe today will be cold if yesterday was cold)- are not expected to be highly accurate but are expected to be more accurate than random, which makes them good choices for baseline models.

As an additional baseline for the Artificial Neural Network (ANN), we created a Convolutional Neural Network (CNN). This allows us to measure whether the additional computational complexity of an ANN (compared to a CNN) creates a meaningful difference. The final model is a Residual Neural Network (ResNet).

Specifically, we build our CNN using 5 fully connected convolutional layers. Because the Earth is periodic in longitude, we used a periodic convolution in the lon-direction. Our Artificial Neural Network was built with five hidden layers consisting of eight neurons each, with a tansigmoid transfer function. For our Residual Neural Network, we created a ResNet-50 model, which chains together 50 residual blocks together. Each of these blocks consist of 2-3 convolutional layers (depending on periodicity), as well as a pooling layer.

For each model (CNN, ANN, and ResNet), we followed a similar high-level procedure. First, we partitioned the data into training, test, and validation portions. Then, we loaded the each segment of the data into a DataGenerator, which produces multiple subsets of each segment of data so that multi-fold training, testing, and validation can be done. Then, the data segments were each run through their respective algorithms, which is broadly summarized as fitting the model with the training and validation data. Specifically, we used data until 2018 as the training data, used 2018 January to November data as validation data, and tested our model on 2018 December data. Then, using the testing data, the Root Mean Square Error (RMSE) was calculated. Finally, the predictions were plotted.

After all of the models were implemented and executed, we created our ensemble learning model. All 5 of our previous models were used as features to train a linear regression model. This method of using a linear or logistic regression model on top of other method's output results is an approach called ensemble by stacking generalization.

Additionally, we created a visualization model in RShiny. The tool takes a date and hour as input and allows the user to decide whether they would like to see forecasts for temperature or precipitation. Then, the tool overlays colored dots on a map of the US and surrounding areas (everywhere that our models did predictions for). The color and border of the dots changes with different time inputs to indicate changes in the data, and hovering over a dot will display its numerical value. This visualization uses our ResNet model.

We also created a **time-lapse** comparing the true weather, CNN model, ANN model, and ResNet model over time.

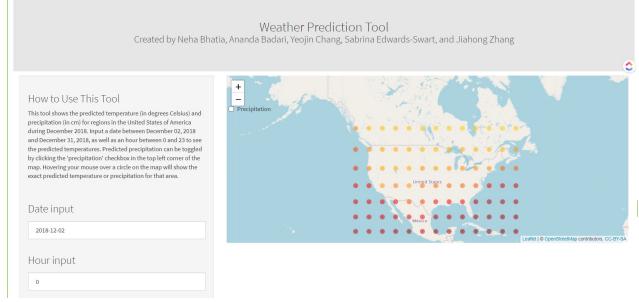


Figure 1. A snapshot of the temperature screen of our team's visualization.

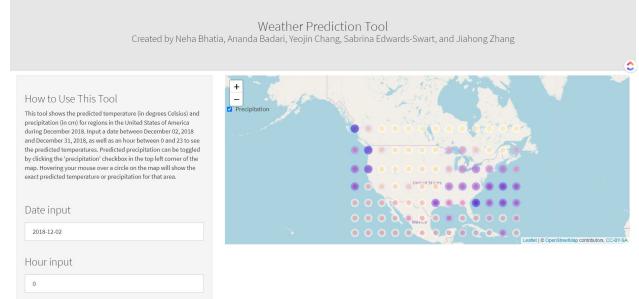


Figure 2. A snapshot of the precipitation screen of our team's visualization.

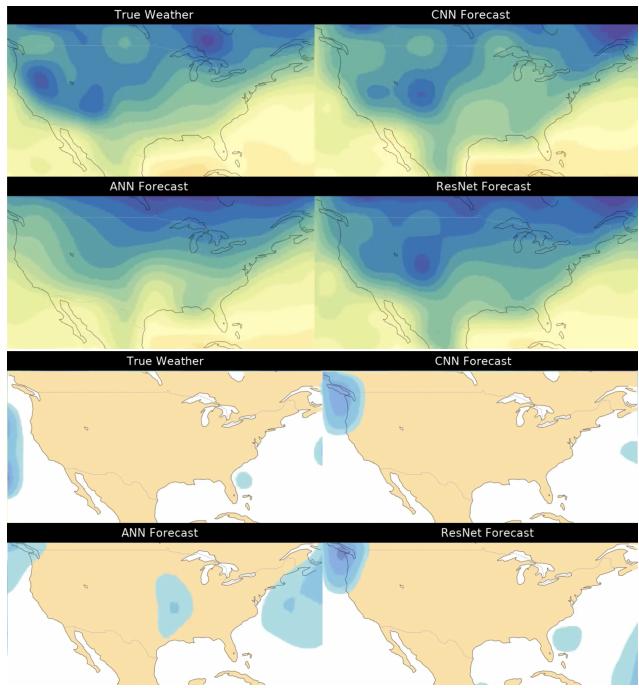


Figure 3. A snapshot of the time lapse our team created.

4. Experiments and Evaluation

For the prediction performance, root mean squared error is used to evaluate between prediction values for temperature and precipitation and the weather records. Further, tracking signal is used for detecting whether bias changes in our model.

4.1. RMSE

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}$$

A common method to evaluate weather prediction models is root mean squared error (RMSE). A smaller RMSE is optimal because it signifies a smaller magnitude of error. It is generally difficult to predict weather after 10 days, so we predicted the RMSE would increase as more predictions are made, and we anticipated a spike in RMSE after 10 days [7].

```
RMSE for persistence temperature predictions: 4.9583805972551
RMSE for climatology temperature predictions: 6.49339788733541
RMSE for linreg temperature predictions: 4.892994529497044
RMSE for ann temperature predictions: 5.7868537805016915
RMSE for resnet temperature predictions: 4.999527295761126
```

Figure 4. Our team's RMSE results for temperature for each of the models.

```
RMSE for persistence precipitation predictions: 0.0005516139134835328
RMSE for climatology precipitation predictions: 0.0006100854097530354
RMSE for linreg precipitation predictions: 0.0005565754581052383
RMSE for ann precipitation predictions: 0.000567829215889841
RMSE for resnet precipitation predictions: 0.0005472448111353846
```

Figure 5. Our team's RMSE results for precipitation for each of the models.

Additionally, we were able to track RMSE over time. The graph plots the RMSE against how far out the predictions are being made (i.e. at $x = 100$, the prediction is being made for 100 days in the future, with the present being defined as the last day the model has training data for). This graph uses data from the ResNet model.

4.2. Tracking Signal

While RMSE captures the magnitude of error, it does not evaluate bias. In order to determine whether our model is consistently making the same type of error (e.g. the model constantly overpredicts temperature), we also kept track of the tracking signal (TS), which is calculated by the following equation.

$$\text{Tracking Signal} = \frac{1}{\text{MAD}} \sum_{t=1}^T (a_t - f_t)^2$$

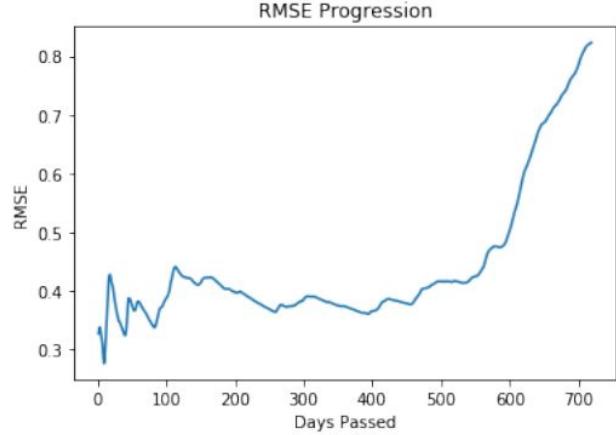


Figure 6. RMSE based on the number of days out of the prediction.

$$\text{MAD} = \frac{1}{n} \sum_{t=1}^T |\hat{y}_t - y_t|$$

A TS of greater than 4 or less than -4 indicates the presence of consistent bias in our predictions, so such TS's would necessitate the revision of a model to be more complex. Since the models we used are highly parameterized, we predicted at the start of our project that we would not get TS values greater than 4 or less than -4.//

```
Tracking signal for climatology temperature predictions: 0.0563525370802115
Tracking signal for persistence temperature predictions: 0.0660315020599293
Tracking signal for linreg temperature predictions: 0.0738529089045817
Tracking signal for resnet temperature predictions: 0.06762291026829927
Tracking signal for ann temperature predictions: 0.061988041274314364
```

Figure 7. Our team's Tracking Signal results for temperature for each of the models.

```
Tracking signal for climatology precipitation predictions: -0.0007930622523
Tracking signal for persistence precipitation predictions: -0.0016770312978
Tracking signal for linreg precipitation predictions: -0.000501888575530287
Tracking signal for cnn precipitation predictions: -0.0023040368649207693
Tracking signal for ann precipitation predictions: -0.0023040368649207693
Tracking signal for resnet precipitation predictions: -0.001587243705893198
```

Figure 8. Our team's Tracking Signal results for precipitation for each of the models.

Since none of our TS values are greater than 4 or less than -4, we were correct in our initial predictions about our TS values, and TS predicts that our models have little consistent biases.

5. Conclusion and Discussion

All team members have contributed a similar amount of effort to this project.

The first thing that we noticed was that the RMSE for both precipitation and temperature for the climatological model were the highest for any model. This indicates to us that a climatological model is inaccurate even when compared to a basic persistence model, so climatological models are not worth pursuing farther in the future.

The linear regression model has a better RMSE for temperature and precipitation than both the CNN and ANN models (though the precipitation values are fairly close), and has a slightly better RMSE for temperature and a slightly worse RMSE for precipitation than the ResNet model (though these values are fairly close). Surprisingly, the same is true for the persistence model, as the RMSE values for temperature and precipitation for persistence and linear regression are fairly close together.

Additionally, in our RMSE progression graph, the RMSE appears to shoot up around 10 days, which is consistent with our predictions for how RMSE progression would look. Subsequently, the RMSE decreases, reaching a local minima at about 30 days, meaning our model was relatively successful at predicting whether out to 30 days compared to weather at 10 days, which achieves part of the goal of this project.

We found that our ability to predict temperature was more accurate than precipitation. This is because temperature has a regular approximate periodicity even on scales of weeks, while precipitation is affected more seasonally. This is best seen in our [comparison video](#), as rain arrives and disappears in ways that seem more random than temperature. We had hoped also to use climatological averages and decreasing spatial resolution to better predict weather, but we found this unfruitful, as climatology was our worst performing model. For weather prediction on the order of days, climatology may as well have been ignored. Persistence, however, was a much better baseline.

Another aspect that we found only limited success with was the ensemble techniques of our paper. Although stacking generalization on the many different models had comparable performance, it was not a non-negligible improvement. In the future, more analysis could be done by taking data at the 1.40625 degree resolution, and then training independent ResNet and ANN models to just predict the temperatures at a certain lat-lon resolution location. Additionally, all input weather features should be used for best performance. Then, we could have an ensemble approach for each location. Another possible future work may be to experiment with different ensemble architectures, to see if there are better ways to combine the predictions. We could, for example, have each model return not only individual predictions but rather return probability distributions.

Generally, improvements could be made in the future with more data, using multimodal predictions with the entire feature set, and considering various ensemble strategies. Even with all of this done, however, there would still be a gap between neural-network predictions and actual weather activities, due to the inherent randomness and chaos of the data.

References

- [1] Richard B. Alley, Kerry A. Emanuel, and Fuqing Zhang. Advances in weather prediction. *Science*, 363(6425):342–344, 2019. 2
- [2] Hrachya Astsatryan, Hayk Grogoryan, Eliza Gyulgyulyan, Anush Hakobyan, Aram Kocharyan, Wahi Narsisian, Vladimir Sahakyan, Yuri Shoukourian, Rita Abrahamyan, Zarmandukht Petrosyan, et al. Weather data visualization and analytical platform. *Scalable Computing: Practice and Experience*, 19(2):79–86, 2018. 2
- [3] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015. 2
- [4] Kening Dou and Xiaoqi Sun. Long-term weather prediction based on ga-bp neural network. In *IOP Conference Series: Earth and Environmental Science*, volume 668, page 012015. IOP Publishing, 2021. 2
- [5] B. Gong, M. Langguth, Y. Ji, A. Mozaffari, S. Stadtler, K. Mache, and M. G. Schultz. Temperature forecasting by deep learning methods. *Geoscientific Model Development Discussions*, 2022:1–35, 2022. 2
- [6] Zahra Karevan and Johan AK Suykens. Spatio-temporal stacked lstm for temperature prediction in weather forecasting. *arXiv preprint arXiv:1811.06341*, 2018. 2
- [7] V Krishnamurthy. Predictability of weather and climate. *Earth and Space Science*, 6(7):1043–1056, 2019. 2, 5
- [8] Andrew C Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986. 1
- [9] Peter Lynch. The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, 227(7):3431–3444, 2008. Predicting weather, climate and extreme events. 1
- [10] Stephan Rasp, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukaina Mouatadid, and Nils Thuerey. Weatherbench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020. e2020MS002203 10.1029/2020MS002203. 2
- [11] Stephan Rasp and Nils Thuerey. Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, 13(2):e2020MS002405, 2021. 2
- [12] Isabelle Roesch and Tobias Günther. Visualization of neural network predictions for weather forecasting. *Computer Graphics Forum*, 38(1):209–220, 2019. 2
- [13] Navin Sharma, Pranshu Sharma, David Irwin, and Prashant Shenoy. Predicting solar generation from weather forecasts using machine learning. In *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 528–533, 2011. 2
- [14] Bo-Wen Shen, Wei-Kuo Tao, and Bryan Green. Coupling advanced modeling and visualization to improve high-impact tropical weather prediction. *Computing in Science Engineering*, 13(5):56–67, Sep. 2011. 1
- [15] Talha Siddique, Md Shaad Mahmud, Amy M. Keesee, Chigomezyo M. Ngwira, and Hyunju Connor. A survey of uncertainty quantification in machine learning for space weather prediction. *Geosciences*, 12(1), 2022. 2
- [16] Manmeet Singh, Nachiketa Acharya, Suryachandra A Rao, Bipin Kumar, Zong-Liang Yang, Dev Niogi, et al. Short-range forecasts of global precipitation using deep learning-augmented numerical weather prediction. *arXiv preprint arXiv:2206.11669*, 2022. 2
- [17] Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*, 2020. 2
- [18] Himani Tyagi, Shweta Suran, and Vishwajeet Pattanaik. Weather - temperature pattern prediction and anomaly identification using artificial neural network. *International Journal of Computer Applications*, 140:15–21, 04 2016. 2, 3
- [19] Dawei Wang. Contrast pattern based methods for visualizing and predicting spatiotemporal events. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1560–1567, Nov 2015. 2