AMSIMP:
Using Machine Learning to Improve Numerical Weather Prediction

An open-source solution that leverages machine learning to improve numerical weather prediction.

Advantages of Using Machine Learning

Weather forecasting has traditionally been done by physical models, which are unstable to perturbations. An attractive alternative is to use neural networks to train simulators from observed data.

Besides training the network, it is less computational intensive than a traditional physical model.

ERA5 Atmospheric Reanalysis Dataset

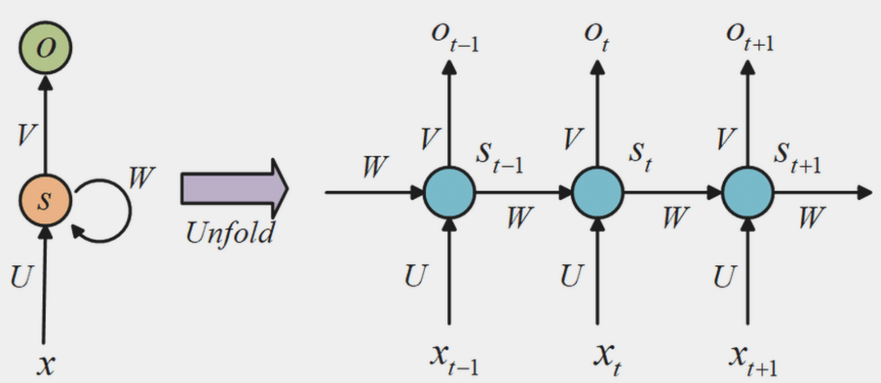
The ERA5 atmospheric reanalysis dataset was used for training, and testing the performance of the neural network architecture. This dataset provides the best guess of the atmospheric state at any point in time .

A spatial resolution of 3° and a temporal resolution of 2 hours was used. In regards to pressure surfaces, 17 vertical pressure levels were chosen

Long Short-Term Memory

The models within the software are built on the long short-term memory architecture. LSTM's were created as a solution to the short-term memory problem found in traditional recurrent neural networks.

They have internal mechanisms called gates that can regulate the flow of information. These gates can learn which data in a sequence is important to keep or throw away.



Open Source

Open source software is software with source code that anyone can inspect, modify, and enhance. Linux, and Chromium are both examples of such software.

The advantages of open source software include:

- Higher quality software
- Regular updates
- Quick fixes.

Results

3.59x FASTER

±1.3°C MAE

Benchmarking

The ECMWF's IFS weather model will act as a comparison, in order to understand the performance of the software. The time required to generate a forecast will also be measured. The error metrics are RMSE, ACC, and MAE.

Analysis of Results

The model is 3.59x faster in comparison to a physical model of a similar resolution. The results are promising; however, the architecture's spatial awareness is lacking, as evident by the current ACC values.

Source of Error

- The use of PCA may have negatively impacted performance.
- Choosing a lower resolution reduced the computational burden during training, however; it may have had a significant impact on the performance.

Looking Forward

- A CNN will be incorporated to improve the model's spatial awareness.
- The model will be trained on a higher resolution dataset.
- It will be trained on a wider variety of atmospheric parameters, precipitation amount being one of them.