

A Machine Learning Workflow for Hurricane Prediction

Albert Kahira^{*†}, Leonardo Bautista Gomez^{*}, Rosa M Badia^{*,†}

^{*}Barcelona Supercomputing Center, Barcelona, Spain

[†]Universitat Politècnica de Catalunya, Barcelona, Spain

E-mail: {albert.kahira, leonardo.bautista, rosa.m.badia}@bsc.es

Keywords—Machine Learning, Deep Neural Networks, Hurricane

I. EXTENDED ABSTRACT

The Atlantic hurricane season runs from June 1st to November causing massive destruction and loss of life. In 2017, 17 named storms hit the Atlantic causing destruction worth an estimated \$316 million and at least 464 fatalities. Meteorologists, by studying previous weather data, predict the expected number of hurricanes in the season. These predictions help authorities prepare for disasters and over the years, better predictions have minimized loss of life and property. However, these predictions rely on human expertise and are often extremely complex due to the thousands of parameters involved and the chaotic nature of weather.

We propose and implement a machine learning model based on deep neural networks to predict the number of hurricanes in the hurricane season. We train the model with more than 100 years of climate data and test it with 5 years. Early results achieve an accuracy of 73% in predicting the number of hurricanes.

A. Background and Motivation

Machine learning has the ability to understand complex models and relationships in data. Recent developments in deep learning models such as Deep Neural Networks (DNN) have led to significant achievements in accuracy [1]. We introduce machine learning model to hurricane prediction to explore the complex relationship between multiple factors such as sea surface temperature, sea level pressure, sea ice cover and wind patterns. We aim to apply a deep learning model to understand the effect of these parameters in the hurricane season and the number of hurricanes. Such insights could significantly improve disaster preparedness and give authorities a better picture of what to expect in the hurricane season.

Previous attempts to use DNN in climate study have been very promising. Liu et al [2] used deep neural networks to detect extreme climate in weather datasets. Zhang et al [3] also used Long Term Short Term memory(LSTM) networks to predict sea surface temperatures. Other studies such as [4] and [5] have also implemented Machine Learning for climate study. However, there has been no studies to predict the number of hurricanes in the hurricane season using Machine Learning.

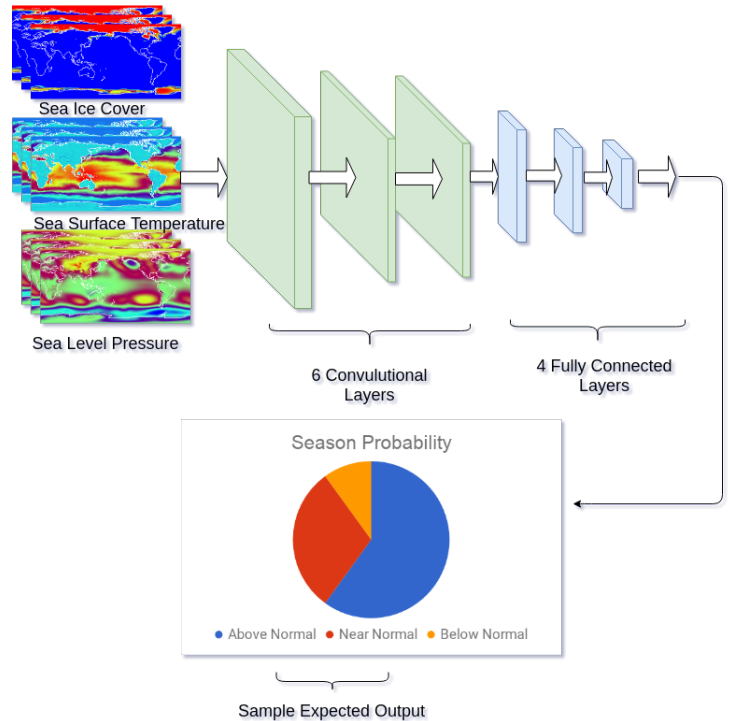


Fig. 1. Work flow

B. Objective

The aim of this study is to introduce Machine Learning to hurricane prediction. Recent scientific advancements have seen geostationary satellites capable of collecting tens of Terabytes of daily data of the weather. On the other side, machine learning models propose efficient techniques to analyse such large data and extract meaningful information. With this large amount of data and the power of high performance computing, machine learning could be an alternative tool for climate study. Furthermore this research aims at introducing approximate computing methods to reduce the computational infrastructure generally required for huge amounts of data.

C. Methodology

Monthly averages of 6 weather variables (sea surface temperature, mean sea level pressure, sea ice cover, 2 metre pressure, U wind speed and V wind speed) from 1901 to 2010 are provided by the earth science department of Barcelona Supercomputing Center. Domain expertise shows that these are

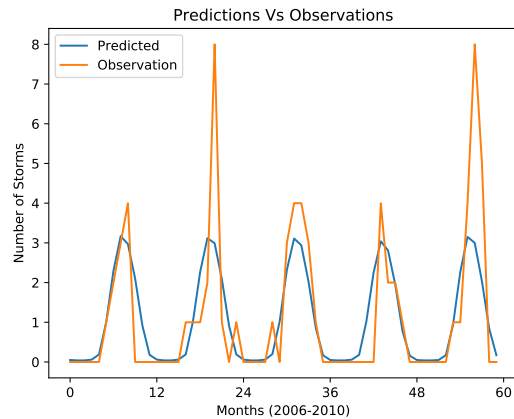


Fig. 2. Predictions for 5 years

the main determinants of the nature and intensity of hurricane season. The total number of named storms for each month in the years 1901 to 210 is also provided which is used as the label for our regression model.

We design a deep learning model with 6 convolutional layers and 4 fully connected layers. Convolutional Neural Networks (CNN) are chosen because of the grid nature of the data. The grid has the shape (160,320) and therefore the input the layer has the shape (160,320,6). There is a Max Pooling after every 2 convolutional layers and a Dropout layer after every fully connected layer. The Convolutional layers have 6 channels, which are the 6 weather variables. The neural network architecture is summarised in figure 1.

We train our model on single node in MareNostrum4 (48 cores) with 1320 training samples split into training, validation and testing. Initial experiments were aimed at finding the correlation between different weather variables and the number of hurricanes. As such, the initial model contained 1 channel. We trained the model with each channel separately and compared the results. Further experiments were aimed at finding the exact part of grid with most effect on the hurricane season, hence, we crop out some parts of the grid and compare the results to those of the original grid.

D. Early Results

Using historical data, we train a DNN to classify the hurricane season based on the number of hurricanes likely to occur. Our initial experiments are aimed at finding the most significant factors in storm formation. Preliminary results show that Sea surface temperature has the highest impact on the prediction of the number of storms. Furthermore, given the average sea surface temperatures in a month, our DNN model is able to predict the number of storms with about 60% accuracy. Figure 2 shows the predictions made by our model for a 5 year period.

Our future goal is to develop a complete end to end work flow to continuously learn weather patterns that affect the hurricane season and accordingly, to make predictions.

E. Conclusion

Early results showed a strong relationship between sea surface temperature and the number of storms. Furthermore, cropping the data grid to eliminate land masses and clean up the data have shown improvements with improvements in accuracy to 73%.

F. Future Work

We plan to implement distributed learning using Py-COMPSs (a programming model and runtime which aims to ease the development of parallel applications for distributed infrastructures, such as Clusters and Clouds) to reduce or eliminate the need to expensive computational infrastructure in climate science.

II. ACKNOWLEDGMENT

The authors would like to thank Dr. Alicia Sanchez Lorente and Dr. Louis Philippe Caron from the Earth Science department of Barcelona Supercomputing Center for providing us with the datasets and their invaluable support in this project.

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 713673.

Albert Kahira has received financial support through the la Caixa INPhINIT Fellowship Grant for Doctoral studies at Spanish Research Centres of Excellence, la Caixa Banking Foundation, Barcelona, Spain.

T

REFERENCES

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.
- [2] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins *et al.*, "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv preprint arXiv:1605.01156*, 2016.
- [3] W. Zhang, L. Han, J. Sun, H. Guo, and J. Dai, "Application of multi-channel 3d-cube successive convolution network for convective storm nowcasting," *arXiv preprint arXiv:1702.04517*, 2017.
- [4] M. B. Richman, L. M. Leslie, H. A. Ramsay, and P. J. Klotzbach, "Reducing tropical cyclone prediction errors using machine learning approaches," *Procedia Computer Science*, vol. 114, pp. 314–323, 2017.
- [5] M. Zhao, I. M. Held, and G. A. Vecchi, "Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of sst anomalies," *Monthly Weather Review*, vol. 138, no. 10, pp. 3858–3868, 2010.



Albert Kahira was born in Lamu, Kenya. He received his BSc degree in Computer Engineering in what is now called Erciyes University and his MSc in Computer Engineering from Abdullah Gul University in Turkey. In 2017 he was awarded the la Caixa INPhINIT Fellowship Grant for Doctoral studies at Spanish Research Centres of Excellence. He is now a research student at Barcelona Supercomputing Center where his research focuses on resilience of machine learning work flows and intelligent distributed systems.