Machine Learning for Extreme Weather Detection

Jakolien van der Meer
(Dated: November 8, 2019)

Objective: The aim of this research is to design a machine learning algorithm that accurately detects and predicts extreme weather events in the Northern Hemisphere, and to perform a comparative study between the CMIP5 climate models and the algorithm.

I. INTRODUCTION

Machine learning is nowadays widely used as a tool to handle large and complex data sets. Climate Science is build around such large data sets, but most climate scientist are still hesitant to use machine learning prediction algorithms because the outcome can not be traced back to a physical model. To change this, in [1] a number of possible applications of machine learning in climate science were proposed, among which the prediction of extreme weather events.

As extreme weather events become more and more common due to climate change as a recent study shows [2], a good prediction of those events is urgently needed. In the aforementioned study, observational weather data from the last 50 years was examined for heat waves affecting at least 20% of the densely populated or agriculturally exploited areas north of the 30th latitude. This data was then coupled with several state-of-the-art climate models and lead to an important conclusion: The extreme heat wave of summer 2018 affecting large areas of the Northern Hemisphere was "virtually certain" due to human-induced climate change. Moreover, the changes that a similar heat wave will reoccur in the coming years are only growing due to the raise in temperature, and were practically zero prior to 2010[1].

Interestingly, the extensive study of [2] does not only show the capability of today's computation power in weather and climate studies, but also its limitations. Namely, 29 CMIP5 models were used with varying parameter inputs and their outcomes were compared to arrive at the accompanying conclusions. The disadvantage of such a propose it that although the uncertainty of the outcome is spread, the anomalies are more difficult to interpret than if using one model. Also the computational efficiency is significantly lower than if using one model for the same task.

Moreover, to avoid having to comb through the newspapers to detect extreme weather events, as was done in [2], automatic detection algorithms can be applied (e.g. [3, 4]). However, this is not straightforward, and the predictability decreases for more extreme events [5]. The use of machine learning would be an obvious choice in this case, since it is designed to recognize patterns in large data sets. The automatic detection of the extreme weather events would also require less computation time, once the algorithm has a large enough quantity of data to learn the concept [6, 7]. If we feed the algorithm with enough historical data a prediction model can then be build by changing the input parameters of the algorithm.

The question remains, how much can we trust the outcome of the resulting algorithm? In other words, can we quantify the uncertainty?

II. RESEARCH QUESTIONS

The following questions are raised:

1. Can we design an algorithm that accurately detects and predicts extreme weather events such as large heat waves and severe storms in the Northern Hemisphere?

2. How does the machine learning algorithm outcome compare to climate model outcomes?

3. How do we quantify the uncertainty of the algorithm?

4. Alternatively, can machine learning be used to combine the outcome of the different global climate models?

5. How much training data do we need to provide the model with to improve its prediction value of extreme weather events?

III. APPROACH

This research will be primarily focused on the added value of machine learning to climate models for detecting and predicting extreme weather events in the Northern hemisphere. Hence the research will be a comparative study between the CMIP5 climate models and a machine learning algorithm that we feed with historical climate data of the last decades. For the last, we will build further on the studies of [6] and [7].

In [6] a semi-supervised bounding box prediction was developed, based on convolutional neural networks to detect extreme weather events. They use a semi-supervised learning algorithm because of the lack of sufficient data on extreme weather events, in combination with exploratory data analysis. The code they developed is

[1] This study looked at the 50 years between 1958 and 2018.

freely available and we plan to use this as a basis for our tool.

Another very interesting topic is studied in [7], where they used computer vision and other machine learning techniques based on linear classifiers to recognize potential severe storms by tracking cloud shapes and movements. They provided the algorithm with more than 50.000 U.S. weather satellite images, so that it could indicate rotational movements in the form of comma-shaped clouds with a 99% accuracy. This lead to a 64% accurate detection of severe storms, such as cyclone formations. This shows that once an algorithm is provided with enough data it can outperform numerical weather prediction models.

For the tool we want to develop, we will start with a feasibility study for cloud shape detection of historical satellite data and detection of large heat waves in the area north of the 30th latitude, by a semi-supervised learning algorithm. After this, we will proceed with the most promising topic and start building the tool with the available data and analyzing the outcome. When the machine learning tool is able to detect certain indicative values of an extreme weather event, we will integrate it with existing numerical prediction models, and analyze the accuracy of the resulting outcomes.

REFERENCES

[1] Rolnick D. et al. (2019) Tackling Climate Change with Machine Learning. arXiv:1906.05433

[2] Vogel, M. M., Zscheischler, J., Wartenburger, R., Dee, D., Seneviratne, S. I. (2019). Concurrent 2018 hot extremes across Northern Hemisphere due to humaninduced climate change. Earth's Future, 7, 692 703. https://doi.org/10.1029/2019EF001189

[3] Ramachandra B., Gadiraju K.K., Vatsavai R. R., Kaiser D. P., Karnowski, T.P. (2016), Detecting Extreme Events in Gridded Climate Data, Procedia Computer Science, 80, 2397 - 2401. https://doi.org/10.1016/j.procs.2016.05.537

[4] Raynaud, L., Touz, B., Arbogast, P. (2018). Detection of Severe Weather Events in a High-Resolution Ensemble Prediction System Using the Extreme Forecast Index (EFI) and Shift of Tails (SOT). Weather and Forecasting, 33 (4), 901-908. 10.1175/WAF-D-17-0183.1

[5] Sterk, A., Holland, M. P. (2018). Extreme value laws and mean squared error growth in dynamical systems. Dynamics and Statistics of the Climate System, 10 (3). DOI: 10.1093/climsys/dzy007

[6] Racah, E., Beckham, C., Maharaj, T. E., Kahou, S., Prabhat, Mr., Pal, C. (2017). ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 3402–3413.

[7] Xinye Zheng, Jianbo Ye, Yukun Chen, Steve Wistar, Jia Li, Jose A. Piedra Fernandez, Michael A. Steinberg, James Z. Wang (2019). Detecting Comma-Shaped Clouds for Severe Weather Forecasting Using Shape and Motion. IEEE Transactions on Geoscience and Remote Sensing, 57 (6): 3788. DOI: 10.1109/TGRS.2018.2887206