HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 3 (2022)

# Predicting The Intensity of Blocking Events Using Machine and Deep Learning

Emir Sahin, Juliana Caulkins and Leya Joykutty

American Heritage School, Rosenberg, TX, USA

## ABSTRACT

Blocking events are high pressure systems that occur in the middle to upper latitudes, diverting the flow of the jet stream and preventing the regular progression of the weather. Blocking events can persist for days to weeks, potentially causing extreme weather events such as droughts or heatwaves. Due to the harmful effects associated with the weather conditions that blocking events can cause, there have been ongoing efforts to forecast them. One such effort includes the Global Ensemble Forecast System (GEFS), which has been reported to underestimate the intensity of stronger blocking events by approximately ten percent. In the current study, one deep learning and three machine learning models were developed to predict the intensity of newly formed blocking events at onset. It was hypothesized that the models would have comparable percent error to the GEFS while using less time and computational resources, that there would be a strong correlation between predicted and actual blocking intensity, and that the deep learning model would have lower error than the machine learning models. The results showed that the models did indeed have comparable error to the GEFS and that there was a statistically significant correlation between predicted and actual blocking intensity for all four models, thus supporting the first two parts of the hypothesis. The third part was not supported since there was no statistically significant difference in error between any of the models, however the deep learning model was noted for not overfitting unlike the machine learning models.

## Statement of Purpose

The purpose of this project is to develop models to predict the intensity of newly formed atmospheric blocking events in the Northern Hemisphere. The independent variable is model type, namely machine learning or deep learning, the dependent variables are mean absolute percent error(MAPE) and mean absolute error(MAE), and the control is the Global Ensemble Forecasting System(GEFS), a preexisting model that can forecast blocking intensity with around 10% error.

## Background

Weather in the Northern Hemisphere is heavily affected by the polar jet stream, a belt of winds that flows from west to east, thus constantly moving weather systems eastward (Atmospheric blocking increase, n.d.). There are two main classifications of weather systems: cyclones and anticyclones. Cyclones have low air pressure and often produce stormy weather whereas anticyclones have high air pressure and tend to produce settled weather (Cyclone and anticyclone, n.d.). Occasionally, a kink will form in the jet stream, resulting in a high-pressure anticyclone remaining stationary in a particular region for an extended period of time, preventing weather systems to the west from passing through that region (Atmospheric blocking increase, n.d). This is known as an atmospheric blocking event and can cause the same weather conditions to persist and intensify in a region for days to weeks (Lupo, 2020). Particularly strong blocks can cause extreme weather events such as the heat wave in the West Coast region of the United States during the summer of 2021, which caused crop losses and put millions at risk of heat stroke (Atmospheric blocking increase, n.d.). Another example of a blocking event with far-reaching implications was the event that occurred over Greenland in the winter of 2009-2010, resulting in blizzards spanning the United States' East Coast that were collectively coined as Snowmageddon (Dunbar (n.d..). Since such extreme conditions can have profoundly negative impacts on society, it is important to be able to predict the intensity of blocking events.

One of the most commonly used data for tracking blocking events involves 500 mb geopotential height. Geopotential height measures the height of the atmosphere at a specific air pressure such as 500 mb. Contour lines on geopotential height maps show troughs in the atmosphere with lower heights that correspond to cyclones and ridges with higher heights that correspond to anticyclones. Thus, if a ridge persists for multiple days, it can be identified as a blocking anticyclone, and the geopotential height data can be used to make predictions. Currently, one of the most accurate methods of forecasting blocking intensity is the Global Ensemble Forecast System (GEFS), which was determined to underestimate the intensity of blocking events by 10% or greater (Lupo, 2020). However, this ensemble system is made up of 21 different forecasts (Global Ensemble Forecast System, 2021) and requires intense computational resources as well as a lot of time to run.

Two much less computationally intensive avenues of computer learning have not yet been applied for predicting blocking intensity: machine learning and deep learning (neural networks). The main difference between these two is that the former requires that spatial features in the data be extracted before the data is passed into the model for predictions whereas the latter is more complex and contains hundreds of mathematical functions called neurons organized into layers in which the model learns to recognize and extract spatial features on its own (Deep Learning vs Machine Learning n.d.). Despite being more advanced than machine learning, however, a neural network deep learning model can still produce results within minutes and is a highly viable option.

## Hypothesis

If machine learning and deep learning models are used to estimate the intensity of blocking events in the Northern Hemisphere, then:

1. Compared to the GEFS (Global Ensemble Forecast System), the models proposed here will have an error comparable to the 10% error in estimating blocking intensity while using significantly less computational power.
2. There will be a statistically significant correlation between predicted and actual blocking intensity for all models.
3. The deep learning model will have less error than machine learning models when predicting blocking intensity.

## Materials

- Tabular data containing 54 years (1968-2021) of atmospheric blocking event data from the Global Climate Change Group at the University of Missouri.
    - Location of onset: Latitude, Longitude
    - Date of onset: Year, month, and day
    - Intensity of blocking event: Blocking Index
- Atmospheric reanalysis dataset containing 43 years (1979-present) of data processed by the NOAA's (National Oceanic and Atmospheric Administration) NCEP (National Centers for Environmental Protection—DOE (US Department of Energy)-Reanalysis 2 model. model. This data comes in the form

of reconstructed 3D grids of the world's atmosphere, including parameters such as **geopotential height**, air pressure, temperature, and wind speed.

- Python and the following Python libraries:
    - Numpy and Pandas: Array and tabular processing
    - Matplotlib and Seaborn: Visualization
    - Scikit-Image: Image processing
    - Scikit-Learn: machine learning
    - Pytorch: Deep learning

# Methods

## Machine Learning Model Procedure

1. The tabular blocking event dataset was downloaded and processed by removing rows of missing data, adding titles to columns based on the website's description of the information contained within each column, removing rows with incorrect date entries, and filtering the data so that it lied between January 1, 1979 and December 31, 2020.
2. The reanalysis data was downloaded and processed to only include data between January 1, 1979 and December 31, 2020.
3. Spatial features within the reanalysis data, which consisted of reconstructed 3D grids of the atmosphere, were extracted using a histogram of oriented gradients (HOG), producing a string of numbers representing those spatial features.
4. The HOG was programmed to identify up to 8 different orientations of edges within the data and split each image of reanalysis data into 16 pixel by 16 pixel cells and 3 cell by 3 cell blocks, normalizing each block with L2-Hys.
5. Both the blocking and reanalysis data were each split into 70% training data and 30% testing data.
6. 3 machine learning models were developed: a support vector regression that used an rbf kernel, a boosted tree regression with 100 estimators, and a random forest regression. An rbf kernel was used in the support vector regression to map the non-linear data used to train the model from a two dimensional space to a higher dimensional space where the model could properly distinguish between data points.
7. The models were run and their mean absolute error (MAE) as well as their mean absolute percent error (MAPE) were automatically calculated during both training and testing.
8. The models' predicted blocking intensities and actual blocking intensities were recorded and used to calculate the absolute errors for each sample pair, as well as the correlation coefficients between predicted and actual intensities.

## Deep Learning Model Procedure

1. The three machine learning models were developed before the deep learning model, so the tabular blocking event dataset as well as the NOAA's reanalysis dataset had already been downloaded and processed.
2. Both the blocking and reanalysis data were split into 70% training data and 30% testing data.

3. A convolutional neural network was built with the following architecture: convolution layer, activation layer, batch normalization layer, and pooling layer repeated three times in that sequence followed by linear layer, activation layer, and batch normalization layer repeated three times in that sequence.
4. The models were run and their mean absolute error (MAE) as well as their mean absolute percent error (MAPE) were automatically calculated during both training and testing.
5. The models' predicted blocking intensities and actual blocking intensities were recorded in an Excel sheet and used to calculate the absolute errors for each sample pair.

## Results

**Table 1.** Mean absolute percent error (MAPE) and mean absolute error (MAE) of the models

| Model | MAPE Train (Blocking Intensity units) | MAPE Test (Blocking Intensity Units) | MAE Train (Blocking Intensity Units) | MAE Test (Blocking Intensity Units) |
|---|---|---|---|---|
| Support Vector Regression | 11.0 | 22.1 | 0.30 | 0.59 |
| Random Forest Regression | 9.8 | 21.5 | 0.26 | 0.57 |
| Boosted Tree Regression | 11.4 | 22.1 | 0.30 | 0.58 |
| CNN (Deep Learning) | 24.1 | 25.0 | 0.64 | 0.64 |

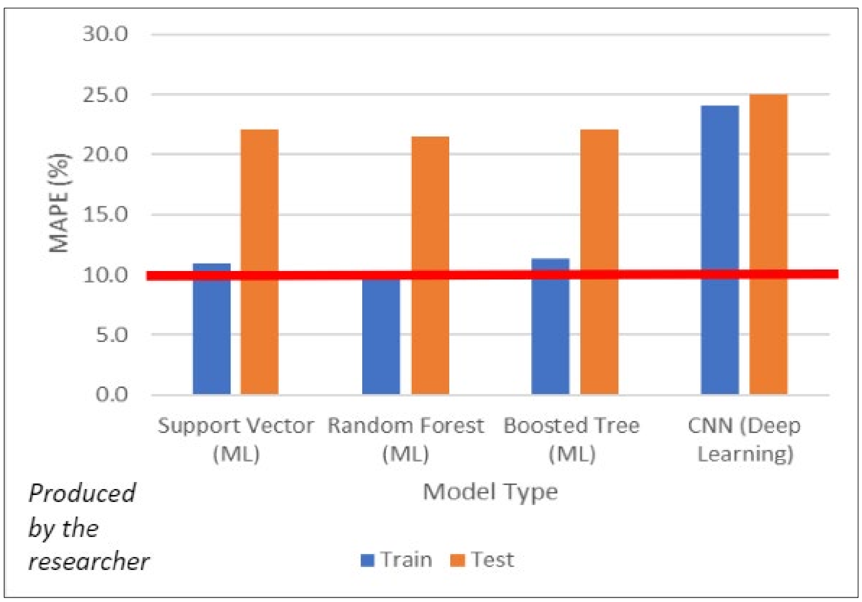


**Figure 1.** MAPE (mean absolute percent error) of the models

Figure 1depicts the MAPE (mean absolute percent error) of each of the models during both training and testing and is based on the results displayed in table 1. The horizontal red line is located at 10% error to represent the error of the Global Ensemble Forecast System (GEFS), and it can be observed that the performances of the models developed in this study are comparable to that of the GEFS.

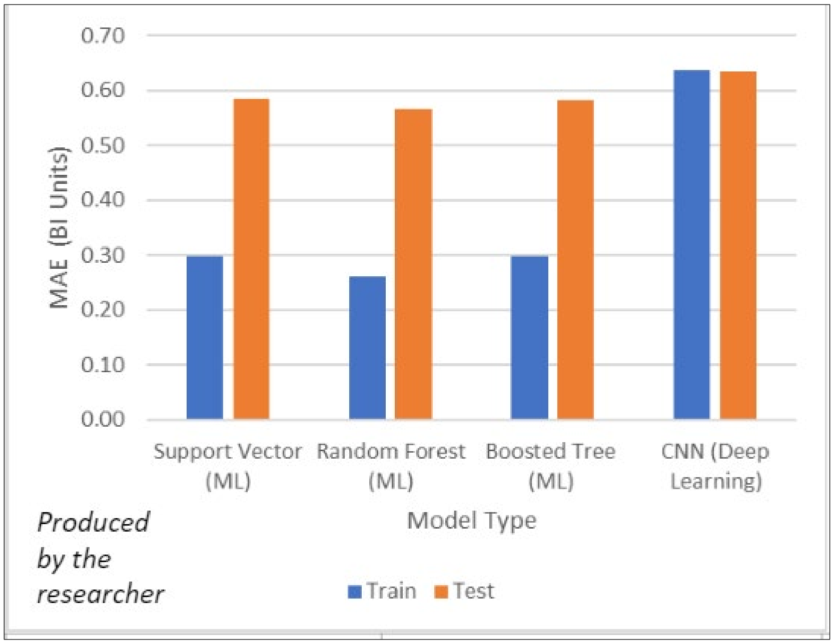


**Figure 2.** MAE (mean absolute error) of the models

Figure 2 shows the mean absolute error (MAE) of the models during both training and testing and is based on the results displayed in table 1. None of the models' errors appear to be significantly different from each other and the ANOVA shown in table 3 confirms this, however the three machine learning models had greater testing error than they did training error, which shows overfitting. Overfitting is when a model memorizes specific features in training data rather than learning overall patterns in the data, causing it to perform worse during testing since it is presented with data it has never seen before, and it should be noted that the deep learning model did not overfit as its training and testing error were nearly identical.

**Table 2.** The correlation between predicted blocking intensity and true blocking intensity during model testing

| Model | Sample Size | Pearson Correlation Coefficient | T-Score | P-Value |
|---|---|---|---|---|
| SVR | 317 | 0.58 | 12.51 | 2.02E-29 ($p<0.0001$) |
| Boosted Tree | 317 | 0.55 | 11.79 | 8.30E-27 ($p<0.00001$) |
| Random Forest | 317 | 0.61 | 13.60 | 1.85E-33 ($p<0.00001$) |
| Deep Learning (CNN) | 317 | 0.58 | 12.76 | 2.34E-30 ($p<0.00001$) |

In table 2, the predictions of the models during testing are compared to the real blocking intensities, and the Pearson correlation coefficients all have significant p-values below 0.00001. Therefore, there is a significant correlation between the predicted and actual blocking intensities, showing that the models' predictions did not occur by chance

**Table 3**. ANOVA test comparing models' absolute error

| source | sum of squares SS | degrees of freedom | mean square MS | F statistic | p-value |
|---|---|---|---|---|---|
| treatment | 0.8899 | 3 | 0.2966 | 1.2945 | **0.2748** |

| | | | | | |
|---|---|---|---|---|---|
| error | 288.7306 | 1260 | 0.2292 | | |
| total | 289.6205 | 1263 | | | |

**Table 4.** Tukey post-hoc analysis

| treatments pair | Tukey HSD Q statistic | Tukey HSD p-value | Tukey HSD inference |
|---|---|---|---|
| SVR vs Boosted Tree | 0.0945 | 0.8999947 | insignificant |
| SVR vs Random Forest | 0.7442 | 0.8999947 | insignificant |
| SVR vs Deep Learning | 1.8977 | 0.5321477 | insignificant |
| Boosted Tree vs Random Forest | 0.6497 | 0.8999947 | insignificant |
| Boosted Tree vs Deep Learning | 1.9922 | 0.4942933 | insignificant |
| Random Forest vs Deep Learning | 2.6419 | 0.2424147 | insignificant |

In tables 3 and 4, an ANOVA test is followed by a Tukey post-hoc analysis to determine whether there was a significant difference in model error between the deep learning and machine learning models. The p-values calculated during both ANOVA and Tukey post-hoc testing were above 0.05, therefore there was no statistically significant difference in error between the models.

## Conclusion

The purpose of this study was to develop three machine learning models (support vector regression, boosted tree regression, and random forest regression) and one deep learning model (convolutional neural network) to predict the intensities of newly formed atmospheric blocking events in the North Hemisphere.

The first part of the hypothesis stated that compared to the GEFS (Global Ensemble Forecast System), the models would have an error comparable to the 10% error in estimating blocking intensity while using significantly less computational power. During testing, the support vector regression model was recorded to have a 22.1% error, the boosted tree regression model was recorded to have 22.1% error, the random forest regression model was recorded to have a 21.5% error, and the deep learning model was recorded to have a 25.0% error. While the GEFS required significant computing power and hours of time to process, the models made in this project had a comparable error to the GEFS while using minimal computing power and producing results within minutes, therefore the first part of the hypothesis was supported.

The second part of the hypothesis stated that there would be a statistically significant correlation between predicted and actual blocking intensity during testing for all models. The predicted and actual blocking intensities during testing for the support vector regression, boosted tree regression, random forest regression, and deep learning models were found to have Pearson correlation coefficients of 0.58, 0.61, 0.55 and 0.58. All four of these correlation coefficients were found to have p-values well below 0.05 and were therefore statistically significant. Thus, the second part of the hypothesis was supported.

The third part of the hypothesis stated that the deep learning model would have lower absolute error during testing than the machine learning models. The mean absolute errors of the support vector regression, boosted tree regression, random forest regression, and deep learning models were found to be 0.59, 0.58, 0.57, and 0.64 respectively. ANOVA testing followed by Tukey post-hoc analysis showed that there was no statistically significant difference ($p>0.05$) between any of the models' errors, therefore the third part of the hypothesis was not supported. However, it should be noted that the three machine learning models had greater testing error than they did training error, which showed overfitting. Overfitting is when a model memorizes specific features in training data rather than learning overall patterns in the data, causing it to perform worse during testing since it is presented with data it has never seen before. While the deep learning model did not outperform the machine learning models in terms of model error, its training and testing error were virtually identical, therefore the deep learning model had the important advantage of not overfitting unlike the other models, leaving potential for future improvement.

## Applications

This experiment can be of practical value for society since atmospheric blocking events have been shown to be linked to extreme weather events such as droughts and heat waves by preventing the regular progression of weather systems through the regions they affect. Due to ongoing climate change in today's world, close attention is being paid to the occurrence and effects of extreme weather, and blocking events are an important part of these phenomena as well. Thus, due to the agricultural, social, and ecological implications of atmospheric blocking events, there have been numerous efforts to study and track them using computer models. Through the development of both machine learning and deep learning models, this project contributes to ongoing efforts to accurately predict the intensity of these events. The errors of the models developed in this study are comparable to that of the Global Ensemble Forecast System (GEFS) while using comparatively little computing power and time to process the data in comparison to the much more powerful computing resources required for a model as advanced as the GEFS, showing that there is still room for improvement to make faster and more efficient models. In addition, unlike the machine learning models, the deep learning model (convolutional neural network) does not overfit, so although the MAE (mean absolute error) and MAPE (mean absolute percent error) of the models were all about the same, the deep learning model is more applicable in the field of predicting blocking intensity since a non-overfitting model has a greater capacity to learn than an overfitting model. Therefore, in addition to showing the plausibility of producing quicker and more efficient models to predict blocking intensity, this study also furthers the understanding in its area by showing why more attention should be given to the development of deep learning models such as convolutional neural networks specifically for blocking intensity prediction.

## Limitations

The machine learning models' results were limited by overfitting, causing them to have greater testing error than training error. It was not expected that despite attempts to find workarounds to this issue, the models would continue to overfit. In addition, the performance of the deep learning model, the convolutional neural network, was limited by the computing resources available for the project. The remote server used to program the models built in this project had very limited random access memory, or RAM, thus limiting the amount of data the neural network could process all at once. Neural networks work best when they are provided with significantly large amounts of data, and although the CNN's performance in terms of its lack of overfitting in this study was certainly noteworthy, it did have the potential to predict blocking intensity with lower error than it ultimately did. However, it should be noted that the amount of memory necessary to improve the deep learning model's performance would still be very little in comparison to the amount of memory needed to run the GEFS, which is an ensemble forecast system consisting of 21 different models.

## Future Research

Future research related to this topic would involve using more powerful computing resources to take advantage of deep learning models' capacity to predict blocking intensity without overfitting by developing a deep learning model that forecasted the intensity of blocking events prior to blocking onset rather than predicting the intensity at onset. Deep learning could be employed to forecast the duration of blocking events as well, as that is another characteristic of these events that is on the forefront of research on this topic. Further expanding upon the possibilities of deep learning for forecasting blocking events, the accuracies of deep learning models other than convolutional neural networks such as recurrent or modular neural networks.

## Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

## References

*Atmospheric blocking increase*. Climate Signals. (n.d.). Retrieved from https://www.climatesignals.org/climate-signals/atmospheric-blocking-increase

*Cyclone and anticyclone*. Air, Cyclones, Anticyclones, and Pressure - JRank Articles. (n.d.). Retrieved from https://science.jrank.org/pages/1927/Cyclone-Anticyclone.html

Deep Learning vs Machine Learning: What's The difference. (n.d.). Retrieved from https://www.computer.org/publications/tech-news/trends/deep-learning-vs-machine-learning-whats-the-difference

Dunbar, B. (n.d.). Stalled Weather Systems More Frequent in Decades of Warmer Atlantic. Retrieved from https://www.nasa.gov/topics/earth/features/blocking-atlantic.html

*Global Ensemble Forecast System*. National Centers for Environmental Information (NCEI). (2021, December 15). Retrieved from https://www.ncei.noaa.gov/products/weather-climate-models/global-ensemble-forecast

Lupo A. R. (2020). Atmospheric blocking events: a review. *Annals of the New York Academy of Sciences*, *1504*(1), 5–24. https://doi.org/10.1111/nyas.14557