Modeling hurricane track development under climate change

Grace Kortum

Introduction

Overview

North Atlantic hurricanes pose a substantial risk to people's safety and wellbeing in the United States. As a result, it is important to investigate how storms will behave in the future and in particular how and why their movement will be affected by climate change. One question that has been studied is whether changes in storm movement due to either internal climate variability or anthropogenic climate change is driven by changes in the genesis location of the storm or changes in the steering flow – how the storm moves after starting in a certain location (Colbert et al., 2013; Colbert et al., 2012). Past results indicate that anthropogenic climate change may lead to an increase in storms that recurve into the open ocean rather than travelling straight westward into North America, driven by a change in steering flow rather than genesis location (Colbert et al., 2013). In addition, Kossin et al. (2010) find that while there has been an increase in the frequency of tropical storms in the North Atlantic, these storms do not make up a substantial portion of the storms that hit land, and the frequency of the storms that make landfall in North America has not been changing. Overall, these results indicate that over the past several decades there have been significant changes in the way that hurricanes develop and move after their genesis.

Hurricane and climate models are used to explain and predict hurricane behavior. In general, these models can be categorized as intensity models which predict storm intensity and wind speeds or track models which predict storm movement (National Hurricane Center and

Central Pacific Hurricane Center, 2019). Additionally, these models are either statistical models, which are based on historical data, or global dynamical models, which are based on geophysical fluid dynamics equations and fundamental properties of the earth system (National Hurricane Center and Central Pacific Hurricane Center, 2019). Most models that are used are global dynamical models, and this is particularly true for modeling storm tracks rather than intensity (National Hurricane Center and Central Pacific Hurricane Center, 2019). While global dynamical models may have higher accuracy, their outputs are difficult to explain because of their very high complexity. Because of this, statistical models can provide a more direct answers about what factors drive certain observed changes in model outputs.

Questions

I seek to better understand how and why hurricane steering flow is changing and how data science models can be used to understand these changes. My two questions are:

1. How well can a simple statistical or empirical model predict hurricane movement from one time period to the next?
2. How has the way that hurricanes move been changing over time due to climate change?

Related work

In addition to the studies by Colbert and Kossin investigating changes in hurricane tracks under climate change that I have already discussed, several studies have used data scientific models to understand hurricane tracks and intensity.

Some past studies have modeled storm evolution as a Markov chain. For example, Jing et al. (2019) apply both regression and a “Markov Environment-Dependent Hurricane Intensity

Model” to model the evolution of tropical cyclone intensity. The “Markov Environment-Dependent Hurricane Intensity Model” considers each time step to be in state associated with a probability distribution of intensity change, and the transition between states is a Markov chain. Masala (2012) also uses a Markov process model to generate simulated hurricane trajectories based on historical data in the North Atlantic and East Pacific from 1950 to 2000.

The CLIPER5 climatology and persistence model is a model that was developed back in 1972 and it relies only on a regression to predict hurricane tracks (National Oceanic and Atmospheric Administration, 2009). This model is no longer used frequently due to the development of more complex models with higher predictive accuracy (National Oceanic and Atmospheric Administration, 2009). Ramirez- Beltran (1996) also employed an autoregression model to predict hurricane tracks. Although this model yielded less accurate predictions than the official National Hurricane Center predictions at the time, it had the benefit of simplicity and only requiring few input parameters (Ramirez- Beltran, 1996).

More recently, neural networks have been employed for hurricane track prediction. Alemany, et al. (2019) employ a recurrent neural network model and obtain results which are in line with the accuracy of the National Hurricane Center’s predictions and yielded good predictions up to 120 hours in the future.

Methods

Data Source

As my primary data source, I used simulated hurricane tracks from HiRAM, the National Oceanic and Atmospheric Administration Geophysical Fluid Dynamics Laboratory (NOAA

GFDL)'s **Hi**gh **R**esolution **A**tmospheric **M**odel. This model is a global dynamical model, and I accessed 8 different ensemble members, each of which included simulated hurricane tracks from the years 1971 to 2019. These ensemble members represent different possible scenarios that could have occurred. They have in common the same governing forces and climate change forcing, but they each have different random noise and initial conditions. The benefit of using this model simulation is that it provides more data than only looking at historical hurricane tracks, and as a result, by having multiple ensemble members it helps to ensure that observed trends over a relatively smaller dataset and time period were not due to randomness, given the high level of noise and variability in climate data.

In addition, I used data on global mean temperature anomalies from the same time period of 1971 to 2019. This data was obtained from NASA's Goddard Institute for Space Studies (NASA, 2020). It is calculated as the degrees Celsius above or below the global mean temperature in 1938, which is set as a benchmark at 0.

Data structuring and engineering

The dataset was originally structured with 4 dimensions. The first is the ensemble member label or 'YENS' (this is an integer from 1 to 8 as there were 8 different ensemble members). Then, each ensemble member has 3 data dimensions. The first is the year. Within each year there is the variable 'ZSTORM1' which represents which storm of the year the data point occurred in. For example, a value of 01would indicate that this was the first storm that occurred this year. Lastly, within each storm there is the variable 'XCOUNT', which indicates a data point's position in the storm. For example, for the first data point measured in any storm, XCOUNT would have a value of 1. The data points in each storm are separated by a period of 6 hours, and they include the parameters of latitude, longitude, day and time, wind speed, and sea level pressure. The variables of interest are summarized in table 1.

variable	Also referred to as…	Units	Relevance
Latitude	lat	degrees	Location of a given point in a storm
Longitude	lon	degrees	Location of a given point in a storm
Within-storm position	XCOUNT	Integer 1-120	Whether a given point is from the beginning or end of a storm. Each position increment is separated by 6 hours.
Global Mean Temperature Anomaly	MeanTemp	Degrees Celsius above or below the 1938 level.	The global mean temperature anomaly of each year in the dataset stands in as a proxy for climate change and change over time as it is generally increasing year over year.
Wind Speed	wind	Meters/second	Wind speed represents the intensity of the storm. In general, it is highest in the middle of the storm track rather than start or end.
Sea Level Pressure	SLP	millibars	Hurricanes are driven by low pressure systems with rising air leading to high winds. Sea level pressure is indicative of the stage and intensity of the storm.
Ensemble	YENS	Integer 1-8	The dataset includes 8 distinct ensemble members, each of which represents a different model realization with different noise and starting conditions.
Time	Year, hour	Years, hours	Dataset spans 1971 to 2019. Within each storm, points are separated by 6 hours.

Table 1: Summary of relevant variables.

Before training any models, the data needed to be restructured to facilitate using autoregressive models. I put the data into a data frame where rows represented each data point sorted chronologically. In addition to the parameters of wind, SLP, latitude, longitude, time, XCOUNT, and temperature anomaly, each row additionally included the latitude, longitude, wind, and SLP from the previous time step as well as the time step before that, given that those two-time steps were part of the same storm as the current time step. To enable including these parameters for every point in the data frame, the first two points of every storm were dropped. Ultimately, this final data frame had a size of 131,857 rows, with each row representing one observation.

Modeling and Analysis

Two different kinds of models were used: a multiple regression model and a neural network model.

For the multiple regression model, 4 similar regressions were run to predict the variables of latitude, longitude, wind, and SLP. The feature space for these regressions included latitude, longitude, wind, and SLP from the two previous time steps (6 hours and 12 hours before), along with global mean temperature anomaly, the position in the storm (XCOUNT), an intercept, and second order interaction terms between the latitude and longitude variables. This model is summarized in equation 1 for latitude as the y variable.

$$\begin{aligned} lat_t = \; & \beta_1 lat_{t-2} + \beta_2 lon_{t-2} + \beta_3 wind_{t-2} + \beta_4 SLP_{t-2} \\ & + \beta_5 lat_{t-1} + \beta_6 lon_{t-1} + \beta_7 wind_{t-1} + \beta_8 SLP_{t-1} \\ & + \beta_9 MeanTemp + \beta_{10} XCOUNT \\ & + \beta_{11} lat_{t-1}\, lon_{t-1} + \beta_{12} lat_{t-1} lon_{t-2} + \beta_{13} lon_{t-1} lat_{t-2} + \beta_{14} lon_{t-2}\, lat_{t-2} \\ & + \beta_0 + \varepsilon \end{aligned} \tag{1}$$

The other regressions for longitude, wind, and SLP as the y variable are exactly analogous to the one in equation 1.

Another set of 4 regressions were run in the same way to predict latitude, longitude, wind, and SLP based on the same feature space, but excluding the global mean temperature anomaly. This was to assess the difference in performance of a predictive model trained with or without global mean temperature anomaly, which represents climate change over this period of time. This model is summarized in equation 2 for latitude as the y variable, with the other regressions for longitude, wind, and SLP as the y variable again being analogous.

$$
\begin{aligned}
lat_t = \quad & \beta_1 lat_{t-2} + \beta_2 lon_{t-2} + \beta_3 wind_{t-2} + \beta_4 SLP_{t-2} \\
& + \beta_5 lat_{t-1} + \beta_6 lon_{t-1} + \beta_7 wind_{t-1} + \beta_8 SLP_{t-1} \\
& + \beta_9 XCOUNT \\
& + \beta_{10} lat_{t-1}\, lon_{t-1} + \beta_{11} lat_{t-1} lon_{t-2} + \beta_{12} lon_{t-1} lat_{t-2} + \beta_{13} lon_{t-2}\, lat_{t-2} \\
& + \beta_0 + \varepsilon
\end{aligned} \tag{2}
$$

One neural network model was trained to predict longitude, wind speed, latitude, and SLP from a feature space including longitude, wind, latitude, and SLP from the previous two timesteps, as well as XCOUNT and the mean global temperature anomaly. The second order interaction terms were not included. In addition, the input variables were all scaled to have a range on (0,1).

The neural network model was built using Keras, and a diagram of its architecture is shown in Fig. 1. The model had 6 dense hidden layers, each having a relu activation function. This activation was chosen because it had a better performance than the sigmoid or tanh activation functions, which were also tried. Different numbers of layers were tested between 2 and 10, and 6 was chosen for yielding the best performance. The size of the layers alternated between 64 and 32 from layer to layer. The model was trained over 20 epochs, with a learning rate of 0.001 and batch size of 32. The loss function was defined to be mean squared error.

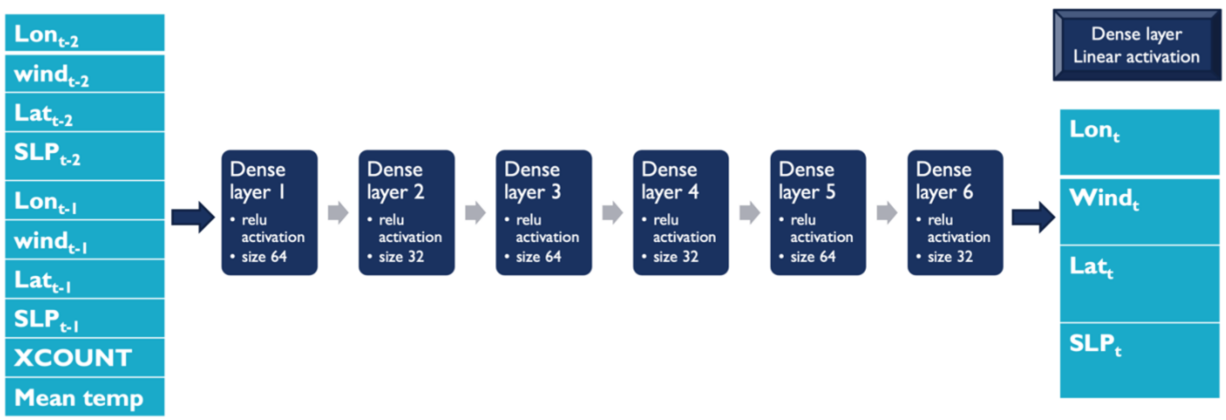

Figure 1: Diagram of the neural network architecture.

For both the multiple regression model and the neural network model, predicted tracks were built out using the models. The predicted tracks were constructed using information about the first two points in each storm in the dataset. The predicted tracks were thus “anchored” at the true genesis location of the storm, and the next data point was predicted using these models. Once the third data point was predicted, the fourth could be predicted based on the second given point and the third predicted point. This method was used to iteratively predict the movement of the storm tracks (given their start location and length).

Because the predicted storm tracks were built iteratively in this way, a new metric was needed to assess their accuracy that would provide an indication of accuracy of prediction beyond just one timestep out. To do this, for every data point in the dataset, the following 10 data points were iteratively predicted. This was done for all the data for the regression model, and for the first 100 storms for the neural network model due to running time constraints. With these 6 to 54 hour ahead predictions (represented by the 10 predicted data points), I quantitatively assessed the average disparity between predicted and actual location for time periods from 6 hours to 54 hours in the future.

Results

Summaries of the 8 regressions are presented in appendices A and B. Notably, the coefficient on global mean temperature anomaly was insignificant in the regressions on latitude and longitude (Table A1, A2). This indicates that there was a minimal effect of climate change on the storm steering flow as represented by this model. However, in the regressions on wind speed and SLP the coefficients on global mean temperature were positive and negative respectively (Table A3, A4), indicating that climate change is associated with higher intensity storms. Because wind and SLP also affect latitude and longitude, this would also lead to a slight change in the steering flow. The coefficients on wind and wind2 in Tables A1 and A2 indicate that wind speed is generally positively correlated with latitude and longitude, indicating higher wind speeds are associated with movement to the north and east. The coefficients on SLP and SLP2 in Tables A1 and A2 indicate that SLP is also generally positively correlated with latitude and longitude, indicating lower SLP is associated with movement to the south and west. Overall, this would indicate that the net effects on steering flow may be small.

The neural network model also trained effectively, with the loss decreasing and plateauing over the 20 epochs (Fig. 2). However, there did appear to be substantial instability, because when the same model was run different times the results were very different.

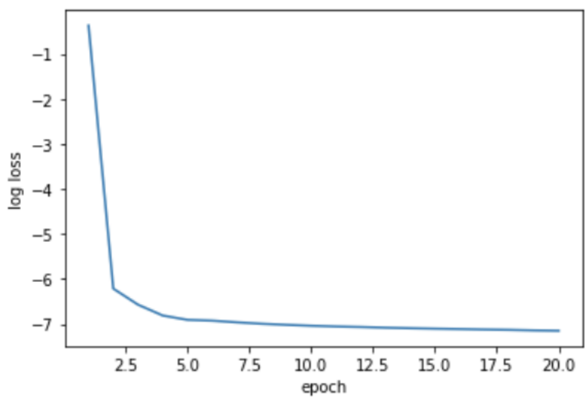

Figure 2: Learning of neural network over 20 epochs. The Y axis shows the natural log of the loss.

Predicted tracks are shown for both the regression and neural network model in Fig. 3. The predicted tracks from the regression model do generally follow the correct pathway, but they are overly smooth compared to the actual tracks. The predicted tracks from the neural network show pathways that look more realistic in shape, but their wind speed predictions are consistently too high.

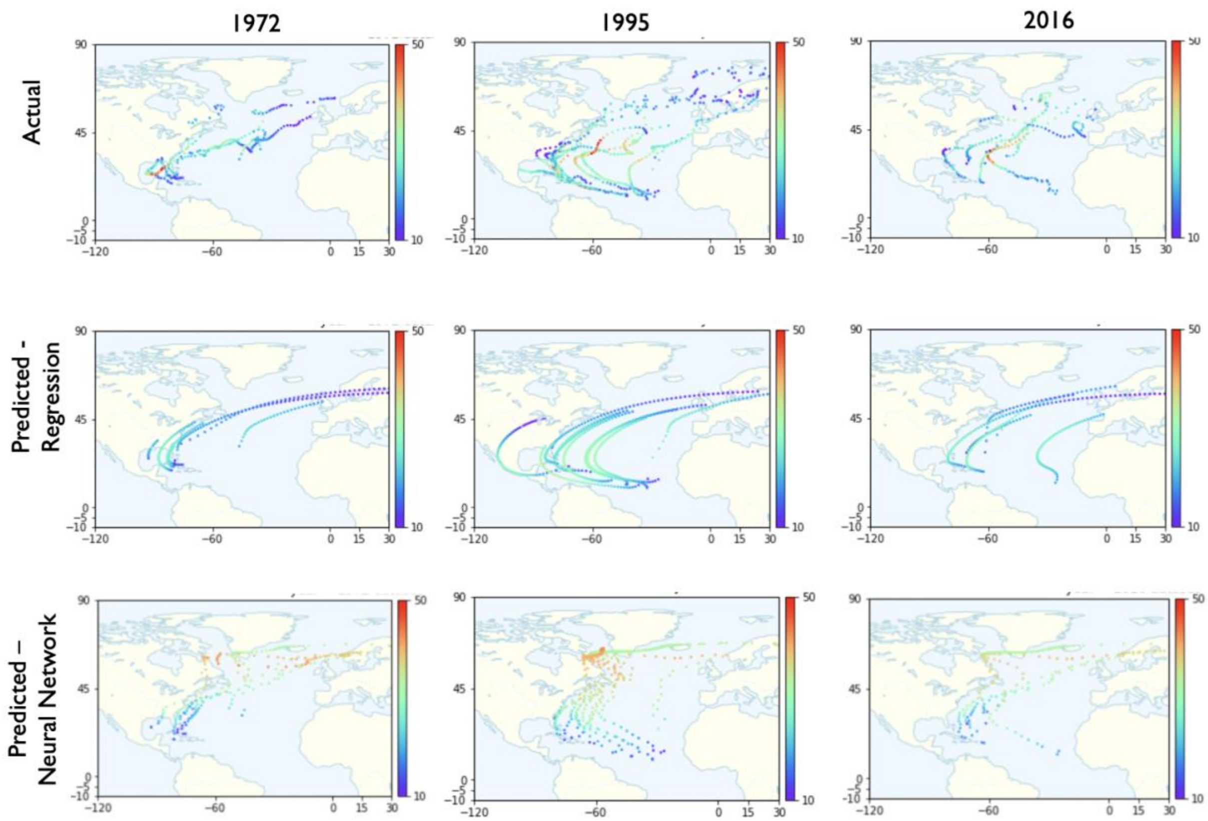

Figure 3: Hurricane tracks from 3 sample years (1972, 1995, and 2016) from the beginning middle and end of ensemble member 1 are shown in the first row. The corresponding predicted tracks from the regression model and from the neural network model are shown in rows 2 and 3 respectively. The color bar represents wind speed in m/s.

As expected, the predictive accuracy of the models decreases over time, with the average difference between the predicted storm and the actual storm increasing from under 100km at 6 hours up to around 700km 54 hours into the future for the regression model (Fig. 4A, B). The predictive accuracy of the regression models including and excluding mean global temperature were the same, as shown by the difference in accuracy between the two in Fig. 4c, which is on average 0. This indicates that including this climate change parameter did not have a big effect on the quality of the model's storm trajectory predictions. The accuracy of the predictions by the neural network model was much lower, with predictions on average over 1000km off of the true location by the first time step 6 hours out (Fig. 4D).

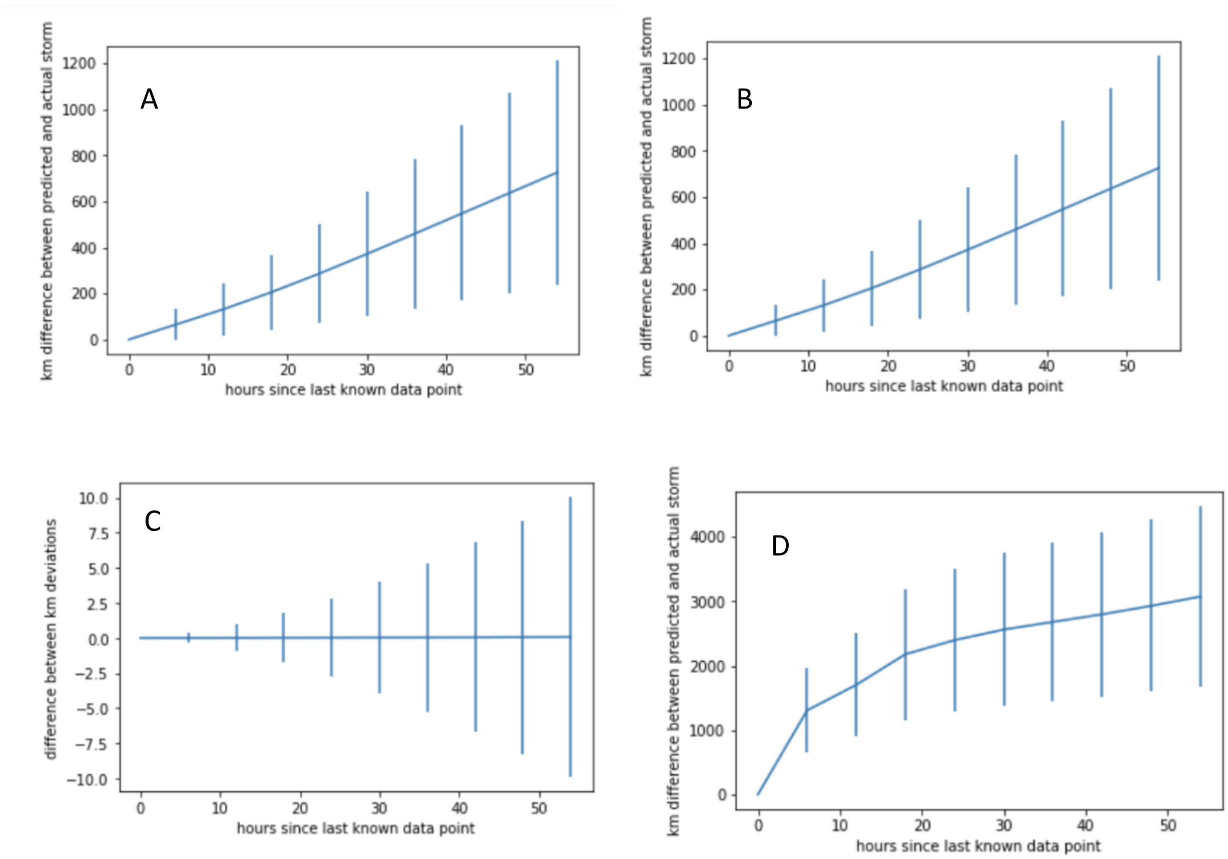

Figure 4: Average model prediction accuracy for a time period of 6 to 54 hours into the future for (A) the regression model including mean global temperature anomaly, (B) the regression model excluding mean global temperature anomaly, (D) the neural network model. (C) shows the difference between (A) and (B). The error bars show 1 standard deviation.

Discussion

Conclusions

These results confirm that a simple multiple regression model can yield reasonable hurricane track predictions up to 6 to 12 hours into the future. Up to 18 hours in the future, the predictions are on average within 200km of the true storm locations. Given that storm diameters range from 100 to 650km, this can still serve as a useful predictor even if the predicted location is not right on target. Although there exist better predictive models than this multiple regression, it has the benefit of simplicity and explainability. The model only requires knowledge of wind speed, SLP, latitude, and longitude for two consecutive points to build a prediction into the future. While global mean temperature anomaly (representing climate change from 1971 to 2019) did significantly affect wind speed and SLP in the regression model, with more intense storms associated with higher temperatures, it did not have a significant impact on the steering flow of the storms or the predictive accuracy of the model. This may indicate that the effect of climate change on steering flow is very small or just that it is not captured in this model.

The neural network model had worse predictive accuracy than the regression model. Given that it appeared to have some instability, this is an indication that it needed more tuning and adjustment to be more effective. Despite the bad predictive accuracy, was that the neural network model produced predicted storm tracks that had a natural appearance that more closely resembled the true storm tracks compared to the regression predictions. Given more improvements to the model, this could be an effective predictor. A next step would then to be to compare the predictions of the neural network when it is run with and without the global mean temperature anomaly variable.

Next steps

The first next step would be to refine the neural network model. In addition, it would be useful to explore the use of a recurrent neural network model as used in Alemany et al. (2019).

The regression model could also be improved by adding more second-degree interaction terms. These could then be reduced back down to fewer dimensions by using a lasso regression model to select the most important features. Experimenting with an AR(1) or AR(3) model instead of just the 2 time step lag model would be useful, as well. Lastly, I would consider modeling using a probabilistic Markov chain model as explored in Masala (2012).

References

Abadi, M., Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Alemany, S., Beltran, J., Perez, A., and Ganzfried, S. 2019. Predicting Hurricane Trajectories Using a Recurrent Neural Network. Association for the Advancement of Artificial Intelligence. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi5zum1vr7tAhX-FVkFHT2vAPkQFjACegQIBBAC&url=https%3A%2F%2Fwww.aaai.org%2Fojs%2Findex.php%2FAAAI%2Farticle%2Fview%2F3819%2F3697&usg=AOvVaw2oKwyAo97C4aWAouwvBsUw

Chollet, F. and others. 2015. Keras. https://keras.io

Colbert, A. J., B. J. Soden, G. A. Vecchi, and B. P. Kirtman, 2013. The Impact of Anthropogenic Climate Change on North Atlantic Tropical Cyclone Tracks. *J. Climate*, 26, 4088–4095, https://doi.org/10.1175/JCLI-D-12-00342.1.

Colbert, A.J. and Soden, B.J. 2012. Climatological Variations in North Atlantic Cyclone Tracks. *J. Climate*, 25, 657–673, https://doi.org/10.1175/JCLI-D-11-00034.1.

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. *Array programming with NumPy*. Nature 585, 357–362 (2020). DOI: 0.1038/s41586-020-2649-2. (Publisher link).

HiRAM (HIgh Resolution Atmospheric Model). 2020. GFDL (Geophysical Fluid Dynamics Laboratory). https://www.gfdl.noaa.gov/hiram/

Hoyer, S. & Hamman, J., 2017. xarray: N-D labeled Arrays and Datasets in Python. Journal of Open Research Software. 5(1), p.10. DOI: http://doi.org/10.5334/jors.148

Hunter, J.D. Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95 (2007), DOI:10.1109/MCSE.2007.55 (publisher link)

Jing, R., and N. Lin, 2019: Tropical Cyclone Intensity Evolution Modeled as a Dependent Hidden Markov Process. *J. Climate*, 32, 7837–7855, https://doi.org/10.1175/JCLI-D-19-0027.1.

Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010. Climate Modulation of North Atlantic Hurricane Tracks. *J. Climate*, 23, 3057–3076, https://doi.org/10.1175/2010JCLI3497.1.

Masala, G. 2012. Hurricane Lifespan Modeling through a Semi-Markov Parametric Approach. J. Forecasting, 32:4. https://doi.org/10.1002/for.2245

McKinney, W. Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56 (2010) (publisher link)

Met Office. 2014. Cartopy. v0.11.2. UK. http://github.com/SciTools/cartopy/archive/v0.11.2.tar.gz

NASA. 2020. Global Climate Change Vital Signs of the Planet: Global Temperature. https://climate.nasa.gov/vital-signs/global-temperature/

National Hurricane Center and Central Pacific Hurricane Center. 2019. NHC Track and Intensity Models. National Oceanic and Atmospheric Administration. https://www.nhc.noaa.gov/modelsummary.shtml

National Oceanic and Atmospheric Administration. 2009. Technical Summary of the National Hurricane Center Track and Intensity Models. https://www.nhc.noaa.gov/pdf/model_summary_20090724.pdf

Pedregosa, F., Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830 (2011) (publisher link)

RAMIREZ-BELTRAN, N.D. 1996. A vector autoregressive model to predict hurricane tracks, International Journal Of Systems Science, 27:1, 1-10, DOI: 10.1080/00207729608929183

Seabold, S. & Perktold, J., 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Appendix A: Regression tables for SLP and wind speed models

Dep. Variable:	Latitude	**R-squared:**	0.998
Model:	OLS	**Adj. R-squared:**	0.998
Method:	Least Squares	**F-statistic:**	2.407e+06
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-1.2576e+05
No. Observations:	131874	**AIC:**	2.515e+05
Df Residuals:	131857	**BIC:**	2.517e+05
Df Model:	16		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.6212	0.367	1.694	0.090	-0.097	1.340
lon2	-0.1454	0.007	-21.297	0.000	-0.159	-0.132
wind2	-0.0023	0.001	-3.053	0.002	-0.004	-0.001
lat2	-0.9953	0.063	-15.900	0.000	-1.118	-0.873
SLP2	-0.0088	0.001	-15.494	0.000	-0.010	-0.008
lon	0.1400	0.007	20.301	0.000	0.127	0.154
wind	0.0103	0.001	13.777	0.000	0.009	0.012
lat	2.0024	0.062	32.254	0.000	1.881	2.124
SLP	0.0090	0.001	15.526	0.000	0.008	0.010
XCOUNT	-0.0010	0.000	-7.468	0.000	-0.001	-0.001
meanTemp	-0.0082	0.006	-1.285	0.199	-0.021	0.004
lat lon	0.0173	0.002	7.089	0.000	0.013	0.022
lat lat2	7.73e-05	1.6e-05	4.837	0.000	4.6e-05	0.000
lat lon2	-0.0185	0.002	-7.423	0.000	-0.023	-0.014
lon lat2	-0.0191	0.002	-7.834	0.000	-0.024	-0.014
lon lon2	1.133e-05	3.53e-06	3.211	0.001	4.41e-06	1.82e-05
lat2 lon2	0.0202	0.002	8.134	0.000	0.015	0.025

Omnibus:	34024.578	**Durbin-Watson:**	2.180
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1315745.846
Skew:	0.529	**Prob(JB):**	0.00
Kurtosis:	18.438	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table A1: Summary of regression results for model run with Latitude as the y variable. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Dep. Variable:	Longitude	**R-squared:**	0.998
Model:	OLS	**Adj. R-squared:**	0.998
Method:	Least Squares	**F-statistic:**	3.423e+06
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-2.0389e+05
No. Observations:	131874	**AIC:**	4.078e+05
Df Residuals:	131857	**BIC:**	4.080e+05
Df Model:	16		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	-9.5426	1.244	-7.669	0.000	-11.981	-7.104
lon2	-0.9179	0.026	-35.952	0.000	-0.968	-0.868
wind2	0.0102	0.001	6.832	0.000	0.007	0.013
lat2	-1.2469	0.127	-9.804	0.000	-1.496	-0.998
SLP2	-0.0046	0.001	-4.241	0.000	-0.007	-0.002
lon	1.9355	0.026	74.527	0.000	1.885	1.986
wind	0.0054	0.001	3.802	0.000	0.003	0.008
lat	1.2426	0.126	9.895	0.000	0.996	1.489
SLP	0.0108	0.001	10.470	0.000	0.009	0.013
XCOUNT	-0.0004	0.000	-1.644	0.100	-0.001	8.19e-05
meanTemp	0.0158	0.012	1.374	0.170	-0.007	0.038
lat lon	0.0274	0.007	3.745	0.000	0.013	0.042
lat lat2	7.995e-06	6.85e-05	0.117	0.907	-0.000	0.000
lat lon2	-0.0319	0.007	-4.299	0.000	-0.046	-0.017
lon lat2	-0.0316	0.007	-4.220	0.000	-0.046	-0.017
lon lon2	-3.021e-05	1.53e-05	-1.968	0.049	-6.03e-05	-1.24e-07
lat2 lon2	0.0361	0.008	4.764	0.000	0.021	0.051

Omnibus:	77413.342	**Durbin-Watson:**	2.247
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	113037081.114
Skew:	1.290	**Prob(JB):**	0.00
Kurtosis:	146.406	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table A2: Summary of regression results for model run with Longitude as the y variable. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Dep. Variable:	SLP	**R-squared:**	0.951
Model:	OLS	**Adj. R-squared:**	0.951
Method:	Least Squares	**F-statistic:**	1.101e+05
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-3.6994e+05
No. Observations:	131874	**AIC:**	7.399e+05
Df Residuals:	131857	**BIC:**	7.401e+05
Df Model:	16		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	34.1887	1.830	18.683	0.000	30.602	37.775
lon2	0.0402	0.022	1.854	0.064	-0.002	0.083
wind2	0.0318	0.005	6.412	0.000	0.022	0.042
lat2	-0.4835	0.152	-3.177	0.001	-0.782	-0.185
SLP2	-0.1254	0.005	-25.239	0.000	-0.135	-0.116
lon	-0.1885	0.022	-8.709	0.000	-0.231	-0.146
wind	-0.0182	0.005	-3.473	0.001	-0.028	-0.008
lat	0.8545	0.150	5.704	0.000	0.561	1.148
SLP	1.1069	0.005	217.351	0.000	1.097	1.117
XCOUNT	0.0161	0.001	24.057	0.000	0.015	0.017
meanTemp	-0.0935	0.040	-2.311	0.021	-0.173	-0.014
lat lon	-0.0365	0.006	-6.573	0.000	-0.047	-0.026
lat lat2	0.0008	6.5e-05	12.040	0.000	0.001	0.001
lat lon2	0.0330	0.006	5.817	0.000	0.022	0.044
lon lat2	0.0369	0.006	6.543	0.000	0.026	0.048
lon lon2	0.0003	1.45e-05	20.514	0.000	0.000	0.000
lat2 lon2	-0.0345	0.006	-6.001	0.000	-0.046	-0.023

Omnibus:	17282.917	**Durbin-Watson:**	2.035
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	90602.205
Skew:	-0.529	**Prob(JB):**	0.00
Kurtosis:	6.920	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table A3: Summary of regression results for model run with SLP as the y variable. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Dep. Variable:	Wind	**R-squared:**	0.858
Model:	OLS	**Adj. R-squared:**	0.858
Method:	Least Squares	**F-statistic:**	3.114e+04
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	03:51:17	**Log-Likelihood:**	-3.4779e+05
No. Observations:	131874	**AIC:**	6.956e+05
Df Residuals:	131857	**BIC:**	6.958e+05
Df Model:	16		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	51.1967	1.682	30.442	0.000	47.900	54.493
lon2	-0.1459	0.020	-7.142	0.000	-0.186	-0.106
wind2	0.2111	0.005	42.359	0.000	0.201	0.221
lat2	-0.4397	0.147	-2.998	0.003	-0.727	-0.152
SLP2	0.1878	0.004	48.033	0.000	0.180	0.195
lon	0.3245	0.021	15.799	0.000	0.284	0.365
wind	0.5838	0.005	106.838	0.000	0.573	0.594
lat	0.1074	0.144	0.744	0.457	-0.175	0.390
SLP	-0.2559	0.004	-63.230	0.000	-0.264	-0.248
XCOUNT	-0.0070	0.001	-11.815	0.000	-0.008	-0.006
meanTemp	0.1241	0.034	3.641	0.000	0.057	0.191
lat lon	-0.0370	0.005	-8.156	0.000	-0.046	-0.028
lat lat2	-0.0010	5.95e-05	-16.533	0.000	-0.001	-0.001
lat lon2	0.0368	0.005	7.902	0.000	0.028	0.046
lon lat2	0.0336	0.005	7.445	0.000	0.025	0.042
lon lon2	-0.0003	1.61e-05	-20.980	0.000	-0.000	-0.000
lat2 lon2	-0.0323	0.005	-6.981	0.000	-0.041	-0.023

Omnibus:	11258.678	**Durbin-Watson:**	2.038
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	67584.957
Skew:	0.161	**Prob(JB):**	0.00
Kurtosis:	6.492	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table A4: Summary of regression results for model run with wind speed as the y variable. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Appendix B: Regression tables for all models excluding global mean temperature

Dep. Variable:	Latitude	**R-squared:**	0.998
Model:	OLS	**Adj. R-squared:**	0.998
Method:	Least Squares	**F-statistic:**	2.565e+06
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-1.2576e+05
No. Observations:	131874	**AIC:**	2.515e+05
Df Residuals:	131858	**BIC:**	2.517e+05
Df Model:	15		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.6206	0.367	1.693	0.091	-0.098	1.339
lon2	-0.1453	0.007	-21.292	0.000	-0.159	-0.132
wind2	-0.0023	0.001	-3.063	0.002	-0.004	-0.001
lat2	-0.9954	0.063	-15.901	0.000	-1.118	-0.873
SLP2	-0.0088	0.001	-15.497	0.000	-0.010	-0.008
lon	0.1400	0.007	20.298	0.000	0.126	0.154
wind	0.0103	0.001	13.768	0.000	0.009	0.012
lat	2.0026	0.062	32.255	0.000	1.881	2.124
SLP	0.0090	0.001	15.525	0.000	0.008	0.010
XCOUNT	-0.0010	0.000	-7.523	0.000	-0.001	-0.001
lat lon	0.0173	0.002	7.090	0.000	0.013	0.022
lat lat2	7.7e-05	1.6e-05	4.817	0.000	4.57e-05	0.000
lat lon2	-0.0185	0.002	-7.425	0.000	-0.023	-0.014
lon lat2	-0.0191	0.002	-7.836	0.000	-0.024	-0.014
lon lon2	1.132e-05	3.53e-06	3.207	0.001	4.4e-06	1.82e-05
lat2 lon2	0.0202	0.002	8.136	0.000	0.015	0.025

Omnibus:	34024.297	**Durbin-Watson:**	2.180
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1315984.143
Skew:	0.529	**Prob(JB):**	0.00
Kurtosis:	18.440	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table B1: Summary of regression results for model run with Latitude as the y variable, excluding global mean temperature from the model. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Dep. Variable:	Longitude	**R-squared:**	0.998
Model:	OLS	**Adj. R-squared:**	0.998
Method:	Least Squares	**F-statistic:**	3.651e+06
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-2.0389e+05
No. Observations:	131874	**AIC:**	4.078e+05
Df Residuals:	131858	**BIC:**	4.080e+05
Df Model:	15		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	-9.5415	1.244	-7.668	0.000	-11.980	-7.103
lon2	-0.9180	0.026	-35.953	0.000	-0.968	-0.868
wind2	0.0103	0.001	6.840	0.000	0.007	0.013
lat2	-1.2466	0.127	-9.803	0.000	-1.496	-0.997
SLP2	-0.0046	0.001	-4.238	0.000	-0.007	-0.002
lon	1.9356	0.026	74.528	0.000	1.885	1.986
wind	0.0055	0.001	3.813	0.000	0.003	0.008
lat	1.2424	0.126	9.894	0.000	0.996	1.489
SLP	0.0108	0.001	10.471	0.000	0.009	0.013
XCOUNT	-0.0004	0.000	-1.594	0.111	-0.001	9.48e-05
lat lon	0.0274	0.007	3.744	0.000	0.013	0.042
lat lat2	8.583e-06	6.85e-05	0.125	0.900	-0.000	0.000
lat lon2	-0.0319	0.007	-4.298	0.000	-0.046	-0.017
lon lat2	-0.0316	0.007	-4.219	0.000	-0.046	-0.017
lon lon2	-3.017e-05	1.53e-05	-1.966	0.049	-6.03e-05	-9.26e-08
lat2 lon2	0.0361	0.008	4.763	0.000	0.021	0.051

Omnibus:	77419.304	**Durbin-Watson:**	2.247
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	113034446.673
Skew:	1.290	**Prob(JB):**	0.00
Kurtosis:	146.404	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table B2: Summary of regression results for model run with Longitude as the y variable, excluding global mean temperature from the model. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Dep. Variable:	SLP	**R-squared:**	0.951
Model:	OLS	**Adj. R-squared:**	0.951
Method:	Least Squares	**F-statistic:**	1.174e+05
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-3.6994e+05
No. Observations:	131874	**AIC:**	7.399e+05
Df Residuals:	131858	**BIC:**	7.401e+05
Df Model:	15		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	34.1821	1.830	18.678	0.000	30.595	37.769
lon2	0.0403	0.022	1.861	0.063	-0.002	0.083
wind2	0.0317	0.005	6.396	0.000	0.022	0.041
lat2	-0.4849	0.152	-3.186	0.001	-0.783	-0.187
SLP2	-0.1254	0.005	-25.243	0.000	-0.135	-0.116
lon	-0.1886	0.022	-8.714	0.000	-0.231	-0.146
wind	-0.0183	0.005	-3.491	0.000	-0.029	-0.008
lat	0.8558	0.150	5.712	0.000	0.562	1.149
SLP	1.1069	0.005	217.353	0.000	1.097	1.117
XCOUNT	0.0160	0.001	23.990	0.000	0.015	0.017
lat lon	-0.0365	0.006	-6.564	0.000	-0.047	-0.026
lat lat2	0.0008	6.5e-05	11.993	0.000	0.001	0.001
lat lon2	0.0329	0.006	5.808	0.000	0.022	0.044
lon lat2	0.0368	0.006	6.534	0.000	0.026	0.048
lon lon2	0.0003	1.45e-05	20.509	0.000	0.000	0.000
lat2 lon2	-0.0345	0.006	-5.992	0.000	-0.046	-0.023

Omnibus:	17297.136	**Durbin-Watson:**	2.035
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	90699.181
Skew:	-0.529	**Prob(JB):**	0.00
Kurtosis:	6.922	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table B3: Summary of regression results for model run with SLP as the y variable, excluding global mean temperature from the model. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.

Dep. Variable:	Wind	**R-squared:**	0.858
Model:	OLS	**Adj. R-squared:**	0.858
Method:	Least Squares	**F-statistic:**	3.320e+04
Date:	Mon, 07 Dec 2020	**Prob (F-statistic):**	0.00
Time:	10:51:17	**Log-Likelihood:**	-3.4779e+05
No. Observations:	131874	**AIC:**	6.956e+05
Df Residuals:	131858	**BIC:**	6.958e+05
Df Model:	15		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	51.2055	1.682	30.448	0.000	47.909	54.502
lon2	-0.1461	0.020	-7.151	0.000	-0.186	-0.106
wind2	0.2112	0.005	42.391	0.000	0.201	0.221
lat2	-0.4378	0.147	-2.985	0.003	-0.725	-0.150
SLP2	0.1878	0.004	48.042	0.000	0.180	0.195
lon	0.3247	0.021	15.803	0.000	0.284	0.365
wind	0.5839	0.005	106.881	0.000	0.573	0.595
lat	0.1056	0.144	0.732	0.464	-0.177	0.388
SLP	-0.2559	0.004	-63.229	0.000	-0.264	-0.248
XCOUNT	-0.0069	0.001	-11.658	0.000	-0.008	-0.006
lat lon	-0.0370	0.005	-8.170	0.000	-0.046	-0.028
lat lat2	-0.0010	5.94e-05	-16.467	0.000	-0.001	-0.001
lat lon2	0.0369	0.005	7.917	0.000	0.028	0.046
lon lat2	0.0337	0.005	7.460	0.000	0.025	0.042
lon lon2	-0.0003	1.61e-05	-20.978	0.000	-0.000	-0.000
lat2 lon2	-0.0324	0.005	-6.997	0.000	-0.042	-0.023

Omnibus:	11269.870	**Durbin-Watson:**	2.038
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	67677.449
Skew:	0.162	**Prob(JB):**	0.00
Kurtosis:	6.495	**Cond. No.**	1.59e+07

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.59e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Table B4: Summary of regression results for model run with wind speed as the y variable, excluding global mean temperature from the model. A subscript of 2 on the variables indicates it is the value from 12 hours prior, whereas the same variables with no subscript are the values from 6 hours prior.