The Attribution of Climate Change on the 2020 Atlantic Hurricane Season Using Counterfactual Analysis and Accumulated Cyclone Energy Prediction

Reetahan Mukhopadhyay, rm3873

May 2, 2022

Abstract

The 2020 North Atlantic Hurricane season set a record for the most named storms in recorded history, at 30. The accumulated cyclone energy (ACE) for the year finished at 180.3725, which classifies the season as "Extremely Active" by the NOAA. Anthropogenic climate change may play a role in impacting the intensity of hurricane seasons, and causing the observation of extremely active seasons such as 2020. In this paper, we aim to study the attribution of anthropogenic climate change on the 2020 Atlantic Hurricane season. We circumvent the unavailability of desired simulated hurricane data through a novel workflow. We build a machine learning model to predict ACE given a set of features affecting hurricane development, secure counterfactual and factual simulation output of these features and apply another machine learning model to perform a time series forecast to overcome the fact none of the simulations extend to 2020, then make counterfactual and factual ACE predictions for 2020 using the first model. From there, we utilize a causal counterfactual framework to determine the probability of necessity, sufficiency, and both of anthropogenic climate change for the 2020 season. Our method yields that climate change is likely to be necessary but not likely to be sufficient to explain the season, but may also be considered sufficient if the question is altered to whether a season like 2020 could have occurred in a time frame prior to the year itself. While this remains a work-in-progress, we have seen encouraging results so far.

1 Introduction

Hurricanes are large rotating storm systems with low pressure centers that typically form over the ocean and deliver heavy rain and strong winds. When they make landfall, these storms can cause significant damage to property, infrastructure, and human life — Hurricane Harvey inflicted an estimated 125 billion USD in damage [1], while the Great Galveston Hurricane took an estimated 8000 – 12000 lives [2]. Given their impact, we may wish to better understand what contributes to the formation and degree of activity of such storms. In the North Atlantic, the official hurricane season runs from June 1 to November 30 each year, but given that there are non-trivial occurrences of storm formation outside this time period, we will consider the activity across the whole year. Hurricane season activity can be measured by several metrics, such as the number of storms, the number of major hurricanes — Category 3 or above on the Saffir-Simpson hurricane scale, number of days with active storms, power dissipation index – which is computed by $10^{-4} \sum_{s_i \in S} \sum_{t \in s_i} v_{max}^3$ where $t \in s_i$ is each observation t taken every 6 hours during a storm s_i's lifetime, and S is all the storms in that year, and v_{max} is the maximal one-minute sustained wind speed during the observation t, in knots. However, since it is commonly used in the literature for being proportional to the storm's kinetic energy, we elect to choose accumulated cyclone energy (ACE) as our metric, which is computed similarly to PDI: $10^{-4} \sum_{s_i \in S} \sum_{t \in s_i} v_{max}^2$.

In order to study the attribution of a forcing such as anthropogenic climate change on the 2020 annual ACE of 180.3725, we use an approach from causal counterfactual theory. Among other tasks, causal inference in general aims to identify the causal effect of some phenomena on others. According to Pearl's causal hierarchy, causal inference can be broken into three levels: association, intervention and counterfactuals — which can be thought of corresponding to the tasks of observing (L_1): how does observing X affect Y, doing (L_2): what will the effect on Y be if I do X, and retrospection (L_3): how would Y turn out in a world where X is different. In our case, we are concerned with L_3, specifically the notion of how the 2020 season may have gone down in a world without anthropogenic climate change (Y_0), and using that to compare against the world we live in, which does experience anthopogenic climate change (Y_1).

We follow a key paper [3] that establishes a causal diagram — a model of variables in a system and arrows indicating their causal relationships — for studying attribution of climate phenomena, as shown in Figure 1. We will assume this simplified model of the system.

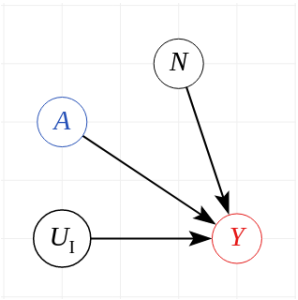

Figure 1: Generic causal diagram for studying the attribution of climate change on climate phenomena. Here, node N refers to anthropogenic forcings — the presence of anthropogenic climate change. Node N refers to natural forcings — the presence of natural forcings. Node U_I refers to internal variability of the climate system, and is regarded as unobserved. Node Y is the climate phenomena under investigation, and the arrows demonstrate all three variables affect Y.

Let Y represent the event of a 2020 ACE being 180.3725 or greater, and X represent the event of anthropogenic climate forcings. [3] establishes the key notion of necessity, sufficiency, and necessity & sufficiency. The probability of necessity PN is set as $PN = P(Y_0 = 0|Y = 1, X = 1)$, which in our scenario refers to the probability that the 2020 season was as active as it was in a counterfactual world without climate change, given that we observe the presence of climate change and the 2020 season actually being active as it was. Essentially, it says that it was necessary for climate change to be present for the season to occur as it did — it could not have occurred that way without it. The probability of sufficiency PS is set as $PS = P(Y_1 = 1|Y = 0, X = 0)$, which in our scenario refers to the probability that the 2020 season was as active as it was in the factual world with climate change, given that we observe no presence of climate change and the 2020 season was not as active as it was. Essentially, it says that it was sufficient for climate change to be present for the season to occur as it did — given only the presence of climate change and no additional causes, the 2020 season would have been active as was. Thus, the probability of being necessary and sufficient PNS, defined as $PNS = P(Y_0 = 0, Y_1 = 1)$, gives a good measure of climate change being the cause of the active season. According to [3], these can definitions can be expressed, with $p_i = P(Y_i = 1)$ for $i \in \{0, 1\}$ as $PN = max(1 - \frac{p_0}{p_1}, 0)$, $PS = max(1 - \frac{1-p_1}{1-p_0}, 0)$, and $PNS = max(p_1 - p_0, 0)$. Additionally we define the notion of risk ratio RR as $RR = \frac{p_1}{p_0}$ which provides the notion of how much likelier we were to observe the 2020 season in the real world compared to the climate change-free world. We can refer to PN,PS,PNS and RR as our attributive quantities. Our ultimate goal is to compute these attributive quantities for our scenario.

Of course, the counterfactual world with climate change is not actually realized, so we cannot simply access the necessary data from it. Given the realism of modern physical climate simulations to the real world, in-silico experimentation can be used to generate this counterfactual world and mine its data. Typical hurricane simulations at high resolution are a very computationally heavy task, usually occurring on high performance computing machines such as supercomputers.

For example, NOAA's Geophysical Fluid Dynamics Laboratory (GFDL) Climate Model 2.5 (CM2.5) Forecast-oriented Low Resolution (FLOR) variant is sometimes used for such simulations. Their user guide states they were developed and tested on the Gaea machine at the Oak Ridge National Laboratory (ORNL), which contains 38 petabytes of memory and 143936 Intel cores [4]. Clearly, devices we have access to are far below that degree of computing power, so performing simulations ourselves was out of the question. The next option would be to simply harness the output of such simulations performed by others who can access these resources. While it may seem expected that such output would be available, after conducting a long and thorough examination ourselves and contacting individuals in the community who may be able to find such data, we were simply unable to locate a single dataset with hurricane simulation output that could be used for our purposes. The reason for no such data being easily available is unknown to us, whether it may simply be that hurricane researchers are not interested in running simulations that can be used to produce simulated seasonal activity metrics, or they have some vested interest to not disclose the raw data from such simulations. Regardless, we decided to cease searching and instead develop a workaround method that will be further discussed in detail in Section 2.

2 Methodology

Let us discuss our developed workflow that overcomes the lack of hurricane simulation data — among other data shortcomings to be discussed — and feeds it into our final counterfactual analysis.

2.1 ACE Prediction

We collect the real monthly ACE data for Atlantic storms from 1960 to 2020 [5], as well as ERA5 Reanalysis data [6] containing information about sea surface temperatures (SST), sea ice concentration (SIC), sea level air pressure, the zonal (UA, east-west) and meriodonal (VA, north-south) components of the wind, over each grid cell in a selected box over the North Atlantic Ocean over the same time period. We also regridded the data from its original resolution to be on the same resolution grid as most of the simulation output — 1.89° latitude by 2.5° longitude, using a bilinear regridding strategy. This gives us a 64 by 41 grid over the ocean, with 5 feature data points reported for each grid cell, for each of the 732 months of data. The five features were selected as they were found to be helpful predictors of hurricane activity [7] [8], and the region is the location of interest for hurricane formation. The predictor dataset and predicted data is split into training and testing. To select a model and hyperparameters, we employ the Fast Library for Automated Machine Learning and Tuning (FLAML) implementation of the automated machine learning framework. This library allows for an efficient search over tree-based (Random Forest, extreme gradient boosting (XGBoost), Lightweight Gradient Boosting(LGBM), etc) models and hyperparameter search using Bayesian hyperparameter tuning. Given the task of regression and root mean squared error (RMSE) as the loss function, we provide the library 15 minutes, which appears to achieve sufficient time after convergence to its current optimal model and set of hyperparameters. We evaluate the model it selects on the testing set and check its RMSE and mean absolute error (MAE). We note that some of the counterfactual ensembles lack data about one of a) pressure, b) both wind components, or c) both SIC and SST. Thus, we found it necessary to train another 3 sets of models with the predictive dataset missing each of those feature sets, otherwise following the same aforementioned procedure.

2.2 Extension of Simulation Data to 2020

Next, we collected 16 counterfactual and 21 factual simulations from [9]. We limit the data to the same region over the Atlantic Ocean as done with the ACE prediction data. The reason we have to collect factual simulation output, despite the fact we live in the "factual world" and have access to "real" data is because the probability distribution we generate in the end is supposed to represent the probability of 2020's ACE taking different values. This should be generated by data points representing 2020's ACE. There is only 1 real world we observe, so we would only

have 1 data point to build this distribution from, which is undesirable — we want more data points to build a distribution. So we need to use simulations from several factual ensembles to represent other realizations of our present world with anthropogenic forcings. Additionally, it allows us for a more fair comparison against the counterfactual ensembles, which of course, are simulations. Each ensemble only had its simulation up to anywhere between December 2014 to December 2019. The target was to collect a 12 month by 64 by 41 by n array, where n is 3 – 5 depending on how many of the features the ensemble had, representing the projected values for the features over the grid for January – December 2020.

In order to generate this, we used a Long Short Term Memory (LSTM)-based time series forecasting method. The method allows for "simultaneous" predictions over all 2624 (64*41) grid cells. We define a LSTM architecture composed of a LSTM layer, a RepeatVector layer - which repeats the input a desired number of times and concatenates to the input, another LSTM layer, and a TimeDistributed layer, which we use to apply a Dense layer to the matrix (input as Atlantic Ocean grid) on each slice of time (each month). For the actual inputs into this architecture, we selected a month as a split point for the simulation into training and testing data. Depending on the number of months in the simulation, we selected a number of months p to "look back" and a number of months q to "forecast". We then generated several sets of "predictor" and "predicted" data: starting from p months into the simulation, we select the past p months to train with and the next q (this was typically 12 or a multiple of 12) months to test with. We iterate one month and repeat this procedure to generate predictor-predicted (X-y) pairs, until we are q months from the end of the simulation. This is done for both the training and testing sets. The architecture is then trained on the training data and then evaluated for the testing set using RMSE and MAE. If the testing results were too low, model parameters — including "look back" and "forecasting" lengths could be adjusted and the model trained until the testing results were desirable. Then we feed the last p months from our testing set (so the last p months of the simulation) and have it generate a forecast for the q months after the end of the current simulation. This gets attached onto the original dataset to form the new dataset for the next iteration, and the whole procedure gets repeated until enough months have been added to reach December 2020. Understanding the potential perils of applying forecasting with data that was forecasted itself, we limited to procedure usually to be ran 2 – 5 times.

2.3 Computing Simulation ACE and Final Attributive Quantities

The final 12 months, representing 2020, are taken from the ensemble to represent its expected counterfactual or factual monthly values for our features. These are read into our models we generated with automated machine learning — each ensemble going to the appropriate model with its matching feature set. This produces the expected monthly ACE for each of these simulations. These are summed to get a yearly prediction of the 2020 ACE for the simulations – counterfactual and factual.

Finally, we read in the predicted counterfactual and factual predicted 2020 ACEs. We fit a Gaussian Kernel Density Estimator (Gaussian KDE) to both datasets, to estimate a probability distribution for both. We computed p_1 by taking the survivor function ($1 - CDF$) of the real 2020 ACE of 180.3725 on the factual distribution, and p_0 by doing the same on the counterfactual distribution. From there, we were able to use the formulae described in Section 1 to compute our attributive statistics PN, PS, PNS and RR as desired. We additionally ran this by fitting a Generalized Extreme Value (GEV) distribution to both datasets, and repeating the same procedure. Description of additional analysis is presented in Section 3.

3 Results and Discussion

In this section we will discuss the results for discussion of the data, the ACE prediction as well as the time series forecasting prior to the final results.

3.1 Data

We opt to discuss the input data into our workflow, other than for the general reason of good practice of inspecting characteristics of your data prior to inputting them into any machine learning workflow, because we made the decision to train and test the model using "real" data from the ERA5 dataset, but we actually apply it to simulation data. The main reason this decision is made is out of a desire for consistency: we have the real-world monthly ACE, but we do not have the simulation ACE, otherwise there would have been no need for this procedure in the first place. To remain consistent, we wish to train our model with real world data to predict the real world ACE. However, we try to justify this decision with the JSD column in Table 1. This refers to the Jensen-Shannon divergence between the Gaussian KDE generated from all the data for that feature in the Reanalysis dataset and the Gaussian KDE generated from all the data for that feature in a selected counterfactual ensemble. In general, the index seems to not be too high (far from 0, as 0 would indicate identical distributions) — providing some validity to our decision.

Additionally, dealing with a spatial dataset, we were interested in possible correlation that may occur across the space for each feature. This gives a sense of how related the conditions in each grid cell may be. We employed using Moran's I index, which is a measure of spatial autocorrelation in a spatial dataset. The results are displayed in the Moran's I column in Table 1, where for each feature we ran the Moran's I test for each month, and then averaged the index across all the months. The results demonstrate an extremely high level of spatial autocorrelation, suggesting a strong correlation between conditions in nearby grid cells. This intuitively makes sense, as in most cases we would expect nearby grid cells to be subjected to generally the same climatological phenomena and physical processes. This initially suggested to us that a Convolutional Neural Net (CNN) may be a good model choice as the convolutions are good for when we have spatial autocorrelation like with image sets, although we later see that a gradient boosted tree was able to handle such a scenario even better in our case.

Feature	JSD	Moran's I
SST	0.21	0.99
Ice Conc.	0.12	0.87
Pressure	0.20	0.96
U-Wind	0.18	0.99
V-Wind	0.07	0.98

Table 1: Data analysis showing the Jensen-Shannon distance for Reanalysis vs counterfactual simulation KDEs generated for each feature, as well as the average Moran's I spatial autocorrelation index for the Reanalysis data over all the months.

3.2 ACE Prediction

We initially built a CNN with an architecture of a pair of 2D convolutional layers, 2D max pooling, another pair of 2D convolutional layers and 2D max pooling, followed by flattening, and four dense layers. The last dense two layers have a softmax and a linear activation, the rest use ReLU. We then attempted AutoML and yielded a LGBM model, and a comparison of the two is shown in Table 2. The choice of automated machine learning is certainly a convenient one from an implementation standpoint, but it also seems to have a proven record of discovering more optimal models. The comparison of predicted and observed ACEs for each testing set month is shown in Figure 2, and direct comparison of the predicted and observed ACEs is provided in Figure 3. Note that many of the months have a zero ACE, as no storms formed in that month — typically those not in proper hurricane season. Through inspection it seems the model is generally good at predicting a month being zero. Suppose we wished to evaluate the model for non-zero months. Eliminating the zero-ACE months from the testing set and re-evaluating drops the RMSE and MAE to 17.13 and 10.90, respectively. The model worsens but maintains a respectable performance, and being able to predict a month as zero is still a valuable characteristic inherently.

Metric	LGBM	CNN
RMSE	12.29	19.63
MAE	6.05	11.97

Table 2: RMSE and MAE comparison for LGBM (via AutoML) and CNN models with optimized hyperparameters

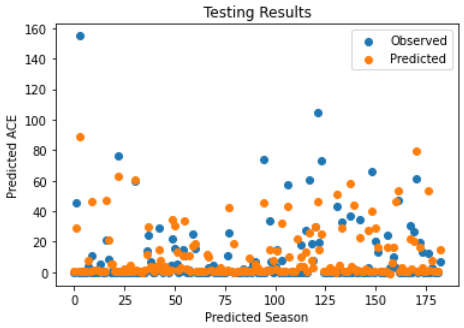

Figure 2: The predicted (orange) and observed (blue) ACE for each month in the testing set under our optimal LGBM model.

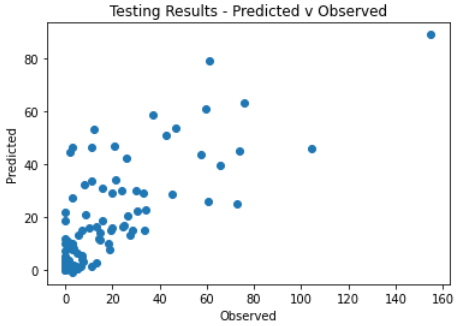

Figure 3: Direct comparison of observed monthly ACE vs predicted monthly ACE for months in testing set under our optimal LGBM model.

Once we saw the LGBM yielded better results on the testing set in regards to RMSE and MAE, we used AutoML to generate models for the other three scenarios: no pressure (P), no SST and SIC, and no UA and VA. The results are provided in Table 3. It seems that dropping the wind features had a far more drastic effect on the results compared to the other features. This begs the question of doing a feature importance analysis, which is provided in Table 4 for the original model. This justifies the significant drop-off from dropping the wind features, and the minimal change from losing pressure, but it would suggest a greater performance decrease from dropping SST.

Metric	No-P	No-UA,VA	No-SST,SIC
RMSE	12.6	17.32	12.56
MAE	6.25	8.24	6.55

Table 3: RMSE and MAE comparison for three generated LGBM models (via AutoML) with features dropped: the first without pressure, the second without both wind components, and the last with no sea temperatures or ice concentrations.

We ran some additional experiments that did not improve overall performance, such as employing Principal Component Analysis (PCA) and Locally Linear Embedding (LLE) to try and perform dimensionality reduction on the grid cell-features, as well as doing a yearly ACE prediction

Feature	Importance Share
SST	0.323
Ice Conc.	0.073
Pressure	0.024
U-Wind	0.282
V-Wind	0.298

Table 4: LGBM Feature Importance Share (individual importance's fraction of sum of all feature impotances) for each feature in the original model

instead of monthly, and isolating the months used to only be June – November, the typical North Atlantic hurricane season.

3.3 Simulation Time Series Forecasting

The need to run a time series forecast was again caused by a lack of available data, with no findings of simulation output containing values for our desired features in 2020. Time series forecasting on a spatial grid is less often a task compared to simple 1-D forecasting, nevertheless we initially tried applying automated machine learning once again, seeing its success on the ACE prediction, using the FLAML library and setting the task to time series forecasting. However, this was being applied to the time series across the months for each individual grid cell of the 2624, for the 3 – 5 features, and that was projected to take over a month to complete. Seeing as most of the initial cells were resolving to Autoregressive Integrated Moving Average (ARIMA) models, we removed AutoML and simply tried an ARIMA fit for each cell. This was still a slow procedure that took almost two days to complete, so when an implementation error was found, re-running the procedure was not a feasible option.

The next option was to try the LSTM-based method as described in Section 2. The R^2, RMSE and MAE for the iteration of this process on a SST in a selected counterfactual ensemble that needed to be extended from December 2017 to December 2020 is provided in Table 5 As we said, there is the risk of continuously predicting on growing augmented data of divulging from reality. Nevertheless, we felt the risk was appropriate. We provide another Jensen-Shannon distance, this time for the 2020 forecasts for each feature against the original data from its counterfactual ensemble, in Table 6. The forecasting for the year expects not to stray far from the rest of the simulation distribution, which at least shows that a substantially unexpected prediction was not made.

Metric	2018	2019	2020
R^2	0.868	0.906	0.902
RMSE	3.46	2.94	3.05
MAE	2.58	2.25	2.34

Table 5: R^2, RMSE, MAE on the "testing" set after each 12 month forecast, to go from 12/2017 - 12/2020, was appended to original dataset and became part of training/testing itself.

Feature	JSD
SST	0.05
Ice Conc.	0.11
Pressure	0.10
U-Wind	0.13
V-Wind	0.25

Table 6: Data analysis showing the Jensen-Shannon distance for counterfactual ensemble vs LSTM 2020 forecast KDEs generated for each feature.

3.4 Final Results

Following the counterfactual and factual simulation ACE predictions, we chose to fit a Gaussian KDE over the data, but also a GEV distribution to keep consistent with the literature [10] [11]. That produced the upper left of Figures 4 and 5. Inspecting the right tail of these plots shows a general greater likelihood for more active seasons under the factual distribution. Checking the real 2020 ACE against the counterfactual distributions allowed us to get our p_0 and p_1 and finally generate our four desired attributive statistics, which we present in Table 7. With PN essentially at 1 in both cases, we have established that anthropogenic climate change is a very necessary cause in explaining the 2020 season. However, PS is middling in both cases, which then leads PNS to do the same. This suggests that climate change may not exactly be a sufficient cause for the season. The risk ratio in both cases is also extremely indicative of the far greater likelihood of the season to occur in the world with climate change.

Measure	Gaussian	GEV
PN	0.999	0.999
PS	0.187	0.269
PNS	0.186	0.269
RR	37727	2254

Table 7: The probabilities of necessity, sufficiency and both as well as the risk ratio, under both a Gaussian KDE fit as well as a GEV fit over the counterfactual and factual simulation data.

The upper right in Figures 4 and 5 demonstrate the increasing necessity of climate change as a factor for more active seasons, while sufficiency seems to decrease (Gaussian) or level off (GEV) with more extremely active possibilities. The more extremely active the season was, it becomes essentially impossible that climate change did not play a role — but it becomes unlikely climate change would be able to be responsible for the event all by itself.

Consider that in our scenario, there is no special meaning about the year 2020 itself, in terms of being a key feature of the problem. Following [3], we can state that we are more interested in if a season as active as 2020 happening in the time since the Industrial Revolution, which is when the effects of anthropogenic climate change became pronounced, instead of the year 2020 itself. So we consider the event of an ACE of 180.3725 or higher occurring at some time t in the period $2021 - \tau \leq t \leq 2020$, and [3] provides a simple reformulation for p_0 and p_1 as $p_x^* = 1 - (1 - p_x)^\tau$. As we can see from the bottoms of Figures 4 and 5, we do not even need to go as far back as the start of the Industrial Revolution — simply giving a $\tau = 15$ year window is enough to show climate change to become a sufficient, and thus a necessary and sufficient cause, of such an active hurricane season, at least once in that timespan.

4 Further Work

Further work is still required on this project. For time constraints, we were only able to collect 16 usable counterfactual ensembles and 21 factual ones — meaning that each of the distributions in Section 3.4 was generated from just that many points, but there are at least several more available that could be used to generate more. Additionally, a few counterfactual and a couple factual ensembles were dropped as their final predicted ACE were either way too high or too low to be reasonable. It has not been concluded whether this signalled an issue with those ensembles, or a general flaw in the workflow. We also note that the results on section 3.4 are contingent on stationarity assumptions about ACE, which are not necessarily true — but as shown by [3] it is possible to overcome this hurdle with additional formulation.

As for additional directions, we could explore PDI or number of major storms instead of ACE as our activity metric. We may also wish to consider additional features for our ACE predictor beyond the five we selected, choosing which may be decided through additional literature review or trial and error. We also note that the [12] also presents a potential opportunity to validate our counterfactual predictions without the hurricane simulations through providing lower and upper bounds. Note that a structural causal model (SCM) is a tuple $< V, U, P(U), \mathcal{F} >$, with

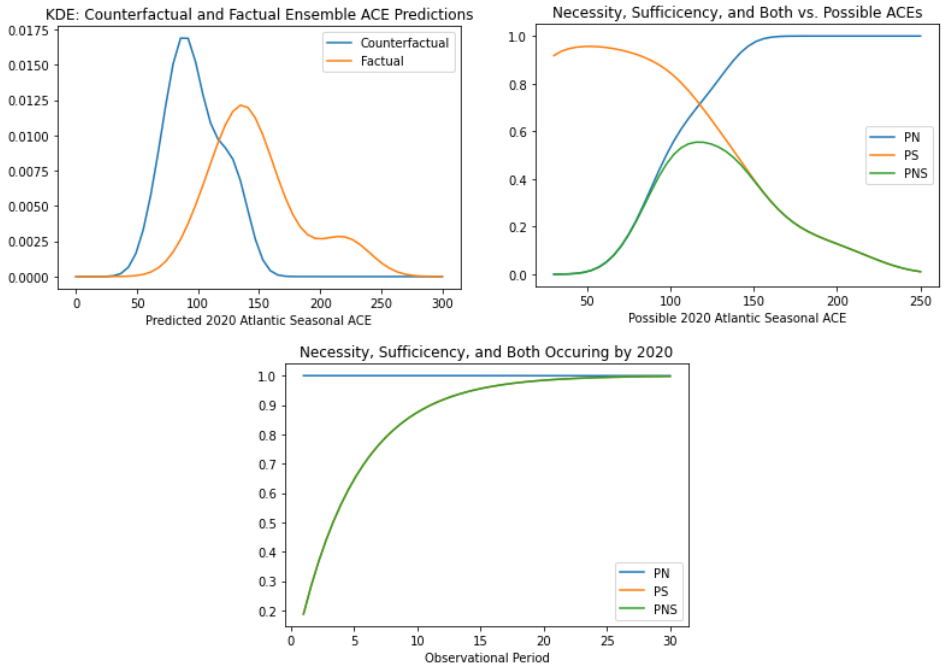

Figure 4: The Gaussian KDEs fitted to the counterfactual simulation ensembles and factual simulation ensembles (top left); PN,PS and PNS as a function of the possible 2020 ACE (top right); PN,PS,PNS as a function of time interval τ leading up to 2020 (bottom)

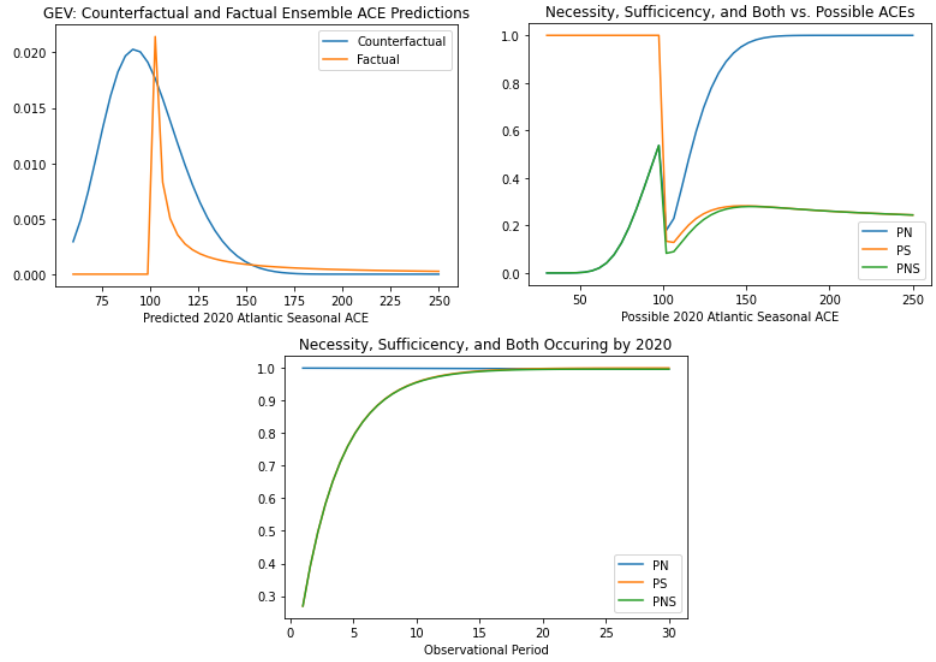

Figure 5: The GEV distribution fitted to the counterfactual simulation ensembles and factual simulation ensembles (top left); PN,PS and PNS as a function of the possible 2020 ACE (top right); PN,PS,PNS as a function of time interval τ leading up to 2020 (bottom)

endogenous variables V, exogenous and unobserved variables U, the probability distribution over $P(U)$, and a set of functions $\mathcal{F}$ that govern the values of variables $v \in V$. Multiple SCMs are compatible with one causal diagram G. Revisiting our causal diagram $\mathcal{G}$ from section 1, we could assign the domain of U_I to be discrete like A (set to 0 or 1, representing the

absence or presence of anthropogenic forcings), and N (the same, but for natural forcings). This would satisfy the definition of a discrete structural causal model (SCM) as stated in [12] and allow for the formulation of an optimization problem to get the lower/upper bounds like $min/max_{N \in \mathcal{N}(\mathcal{G})} P_N(Y_A)$ such that $P_N(Y_A) = P(Y_A)$ where N represents a possible SCM in the space of SCMs compatible with $\mathcal{G}$, P_N is the associated probability distribution under the SCM N, and Y_A is the counterfactual value of Y given $A = a$, which corresponds to the the 2020 season occurring in the present or the climate change-free world. This can be reformulated into a polynomial program according to [12] based on the distribution $P(U)$, which we could enforce to assumptions to create. The program can be solved with a polynomial solver — or approximated through Markov chain Monte Carlo (MCMC) algorithms such as Gibbs sampling as demonstrated by [12].

Of course, the best option would be to acquire actual hurricane simulation data to validate whether the predicted ACE from the simulations was actually valid or not. Continuing to search for this data would still be in order even if it would no longer render the machine learning workflow prior to the distribution fitting necessary, as it would allow us to validate the degree of accuracy of our machine learning workflow. Our workflow presumably usurps far less computing power than a proper hurricane simulation, so evidence that it may be apt at recreating simulated metrics would be a notable achievement beyond the main aim of this paper. Regardless, we were able to demonstrate a causal necessity result, and a causal sufficiency one too — provided a larger timeframe, with the data we were able to collect and constraints imposed, which is a promising result.

References

[1] Chris Mooney. Hurricane harvey was year's costliest u.s. disaster at $125 billion in damages. Washington Post, 2018.

[2] The great galveston hurricane of 1900. NOAA, 2021.

[3] A Hannart, J Pearl, F Otto, P Naveau, and M Ghil. Causal counterfactual theory for the attribution of weather and climate-related events. *Bulletin of the American Meteorological Society*, 97:99–110, 2016.

[4] Gaea. NOAA, 2020.

[5] Phil Klotzbach. Hurricane indices. Technical report, Colorado State University, 2020.

[6] B. Bell, H. Hersbach, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz Sabater, J. Nicolas, R. Radu, D. Schepers, A. Simmons, C. Soci, and J-N Thépaut. Era5 monthly averaged data on single levels from 1950 to 1978 (preliminary version). Technical report, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2020.

[7] Tanmay Asthana, Hamid Krim, Xia Sun, Siddharth Roheda, and Lian Xie. Atlantic hurricane activity prediction: A machine learning approach. *Atmosphere*, 455, 2021.

[8] Albert Kahira, Leonardo Gomez, and Rose Badia. A machine learning workflow for hurricane prediction. Technical report, 5th BSC Severo Ochoa Doctoral Symposium, 2018.

[9] C20c+ detection and attribution project. Technical report, World Climate Research Programme, 2020.

[10] Mark Risser and Michael Wehner. Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during hurricane harvey. *Geophysical Research Letters*, 44:12457–12464, 2017.

[11] Sjoukje Philip1, Sarah Kewl, and Geert Jan van Oldenborgh. A protocol for probabilistic extreme event attribution analyses. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6(2):177–203, 2020.

[12] Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. Technical report, Causal AI Lab, 1 2022.