Predicting Global Temperature Change:

An Exploratory Analysis of Using Machine Learning Techniques on Real Climate Data

Students: Kayla Casey, Joshua Carlson, Claudia Dare, Cora McAnulty

Team Name: College Dropouts

Professor: Dr. Zhengwu Zhang

Course: STOR 565 Machine Learning

Abstract

In this study, we applied machine learning techniques such as time-series analysis, ensemble methods, recurrent neural networks, and clustering on current climate data to produce accurate models of future temperature change. We used temperature change data over 58 years to train models, testing on the past five years to determine the accuracy of our models, and then used those models to predict temperature change in the future.

1 Introduction

Climate change is a persistent global issue with many confounding variables. The effects of climate change are often not felt by the ones most responsible for pollution. This study aims to predict global and regional temperature change in degrees Celsius when compared to a baseline temperature over the years 1951-1980, as provided by FAOSTAT. The main area of focus for this study was in time series forecasting, as climate data and temperature change are well represented through time series models. Clustering was done to add more depth to these models and see how climate change impacts countries based on their land use. Ensemble methods and recurrent neural networks were also created and compared to the time series models. These models operated on clusters of countries based on ecological footprint and region to simulate global temperature change in a variety of ways. Cross-validation was performed on each of the models using a training and testing set of the data, where the test set consisted of the temperature change data for the past 5-12 years and the training set contained all of the previous data. Root mean squared error (RMSE) was typically used as the measure of accuracy for this cross-validation.

1.1 Data

In this study, we utilized two main data sources. The first data source contained monthly temperature change and standard deviation data sourced from FAOSTAT (FAO). A supplementary data source used in this project contains ecological footprint data sourced from the National Footprints Accounts from the UN (NFA). The FAO dataset covers temperature change and deviation data from 1961 to 2023 in 198 countries and 39 territories. These countries are split into 28 regions which can all be separately downloaded. The values for temperature change are given per month and year. The NFA dataset contains ecological footprint data from 1961 to 2014, including information on land use in the categories of grazing land, forest land, fishing ground, and more. These categories are recorded per country for records including biocapacity per hectare, area per capita, biocapacity per capita, and ecological footprint of consumption in global hectares.

1.2 Exploratory Data Analysis

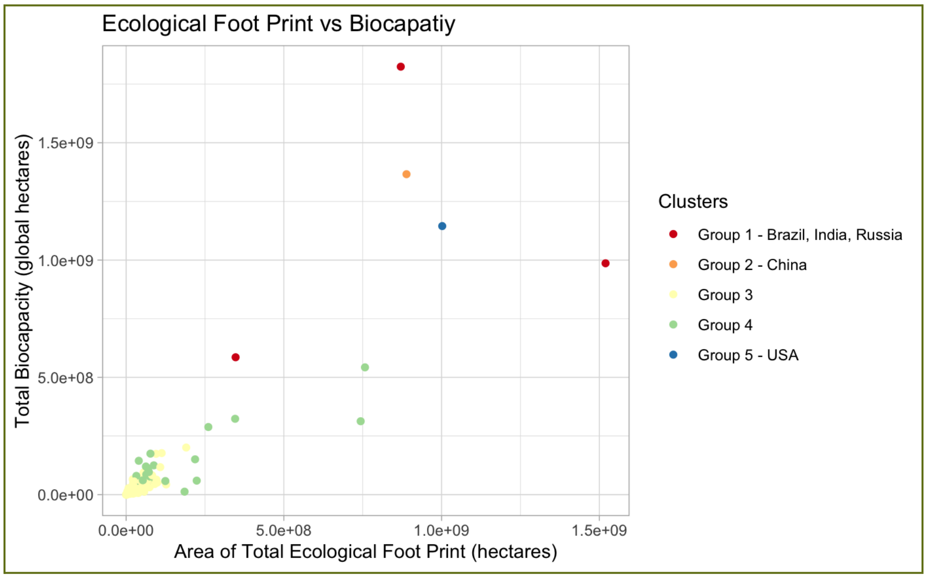
To gain insight into trends and patterns present in our data, we began the study with initial data analysis, including the creation of basic visualizations for our data. We first plotted the global temperature change per month from 1961 to 2023 and found consistent positive trends in temperature change despite some fluctuations. From this observation, we concluded that time series analysis could be used as a modeling method. Initial exploration in the NFA data showed that some countries had little or no information on certain variables for different categories of land use. From this observation, we decided that the NFA dataset would best serve as supplementary data.

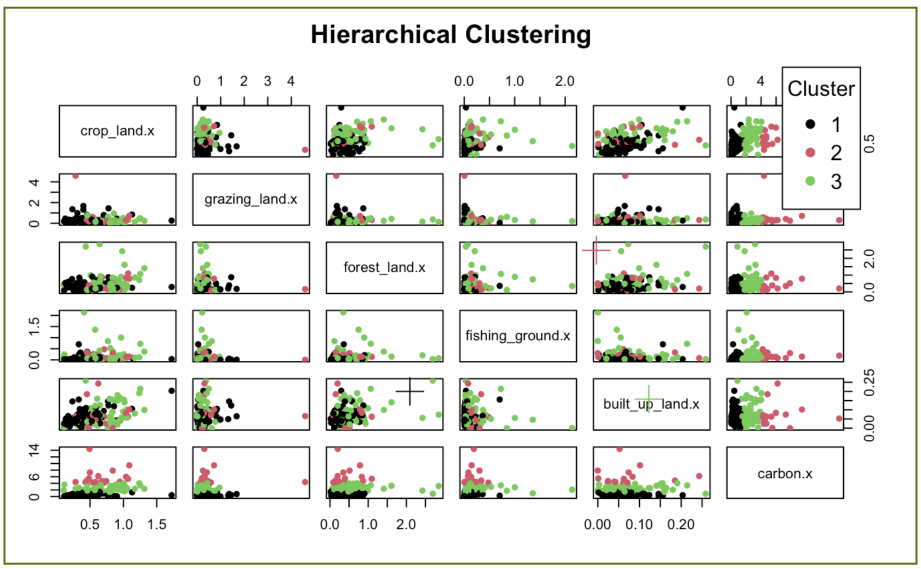
1.2.1 Clustering

From the initial exploration into the NFA dataset, we decided that a reasonable use of this data would be creating clusters on land usage. This could provide groupings for countries with similar land usage and insight into how land use directly impacts climate change. We chose to create models based on one set of clusters using the K-Means method and one set of clusters using the hierarchical method. We found that regardless of how many clusters we made, some clusters would group with only one country, some with a handful of countries, and a final cluster with all other countries. This implies that a few countries stand out in their biocapacity and total environmental footprint. We also found that hierarchical clustering created more evenly sized clusters than K-Means. Below are the groupings of K-Means and Hierarchical clustering that we created models on, as well as some plots showing ecological foots where the data is colored by cluster. The first of these plots was created using the K-means clustering, and the latter using the hierarchical clusters.

K-Means Clusters	Countries
1	Brazil, India, Russia
2	China
3	Afghanistan, Albania, Algeria, Angola, Armenia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brunei Darussalam, Bulgaria, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Chile, Colombia, Congo, Democratic Republic of Congo, Costa Rica, Cote d'Ivoire, Croatia, Cuba, Czech Republic, Denmark, Djibouti, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, French Guiana, French Polynesia, Gabon, Gambia, Georgia, Ghana, Greece, Guadeloupe, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Hungary, Iraq, Ireland, Israel, Jamaica, Jordan, Kenya, Korea, Kuwait, Kyrgyzstan, Lao People's Democratic Republic, Latvia, Lebanon, Lesotho, Liberia, Libyan Arab Jamahiriya, Lithuania, Luxembourg, Macedonia TFYR, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Moldova, Mongolia, Montenegro, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, Nicaragua, Niger, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Portugal, Qatar, Romania, Rwanda, Saint Lucia, Senegal, Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, Somalia, South Sudan, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syrian Arab Republic, Tajikistan, United Republic of Tanzania, Timor-Leste, Togo, Trinidad and Tobago, Tunisia, Turkmenistan, Uganda, United Arab Emirates, Uruguay, Uzbekistan, Bolivarian Republic of Venezuela, Yemen, Zambia, Zimbabwe
4	Argentina, Australia, Canada, France, Germany, Indonesia, Iran, Italy, Japan, Kazakhstan, Korea, Mexico, Nigeria, Poland, Saudi Arabia, South Africa, Spain, Thailand, Turkey, Ukraine, UK, Vietnam
5	USA

Hierarchical Clusters	Countries
1	Armenia, Afghanistan, Albania, Algeria, Angola, Argentina, Bangladesh, Bolivia, Brazil, Myanmar, Burundi, Cameroon, Central African Republic, Sri Lanka, Chad, Colombia, Congo, Costa Rica, Cuba, Azerbaijan, Benin, Dominican Republic, Ecuador, El Salvador, Djibouti, Georgia, Gabon, Gambia, Ghana, Guatemala, Guinea, Guyana, Haiti, India, Indonesia, Iraq, Cote d'Ivoire, Jamaica, Jordan, Kyrgyzstan, Kenya, Lao People's Democratic Republic, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mexico, Morocco, Mozambique, Moldova, Namibia, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Papua New Guinea, Paraguay, Peru. Philippines, Guinea-Bissau, Timor-Leste, Eritrea, Zimbabwe, Rwanda, Saint Lucia, Senegal, Sierra Leone, Somalia, Tajikistan, Swaziland, Syrian Arab Republic, United Republic of Tanzania, Thailand, Togo, Tunisia, Uganda, Burkina Faso, Uruguay, Uzbekistan, Viet Nam, Ethiopia, Yemen, Democratic Republic of Congo, Zambia, Sudan, South Sudan
2	Australia, Bahrain, Brunei Darussalam, Canada, Kazakhstan, Republic of Korea, Kuwait, Mongolia, Qatar, Saudi Arabia, Singapore, Turkmenistan, Trinidad and Tobago, Oman, United Arab Emirates, United States of America, Luxembourg
3	Austria, Bahamas, Barbados, Bhutan, Botswana, Belize, Bulgaria, Chile, Denmark, Belarus, Equatorial Guinea, Estonia, Fiji, Finland, France, French Guiana, French Polynesia, Germany, Bosnia and Herzegovina, Greece, Guadeloupe, Hungary, Croatia, Iran, Islamic, Ireland, Israel, Italy, Japan, Democratic People's Republic of Korea, Latvia, Lebanon, Libyan Arab Jamahiriya, Lithuania, Malaysia, Malta, Netherlands, Macedonia TFYR, Norway, Czech Republic, Poland, Portugal, Romania, Russian Federation, Slovenia, Slovakia, South Africa, Spain, Suriname, Sweden, Switzerland, United Kingdom Turkey, Ukraine, Bolivarian Republic of Venezuela, Belgium, Serbia, Montenegro, China





2 Models

2.1 Time Series Analysis

The temperature change data in the FAOSTAT data set exhibited clear variation with time. This led us to fit time series models for analysis and forecasting. We found that the Autoregressive Moving Average (ARMA) model was the best choice for fitting our time series. The autoregressive (AR) aspect of this model used lagged terms to capture a linear relationship in our model based on the past. The moving average (MA) aspect allowed captured relationships between past errors and our data, giving insight into short-term fluctuations. Thus, the ARMA model provided a comprehensive linear model that covered temporal dependencies and short-term changes in the data. There is a precedent for using ARMA/ARIMA on temperature data, which encouraged us to continue with this approach. ARMA models expect stationary and IID noise data, which we found our data exhibited after removing trends, seasonality, and other temporal dependencies.

To fit our ARMA models to the global data, we began transforming our data to exhibit stationary and IID noise behavior by removing seasonality and trends. Since we used yearly data, we did not find significant harmonic movement, meaning there was no apparent seasonality to remove. We did find quadratic trends in our data, which we removed. Once the trend was removed, we used various tests of IID noise, such as Ljung-Box Q and McLeod-Li Q, on our data and concluded with statistically significant results that it exhibited IID behavior. We then tested that the data was stationary with the Dickey-Fuller test, and were able to conclude stationary behavior. Following this, we began fitting ARMA models with various hyperparameter values to the data, training on the first 58 years of temperature change data available. We trained on a large subgroup of the data to ensure that our models contained recent data for accurate prediction of current trends. We judged the quality of our various models by comparing the fit of the models to the Auto-Correlation Function plots of the data, selecting the model with the smallest AIC value, checking that the residuals remained with IID noise behavior, and observing small, but not overfit, RMSE values. After achieving such a model, we forecasted five years into the future and compared our results to actual climate change values. Following this, we used our models to forecast temperature change into 2024 and beyond.

Following the creation of time series models on worldwide and geographical region data, we fit time series models to the clustered groups we found from the K-Means and Hierarchical clustering methods. We adhered to the model fitting process outlined above and fit individual ARMA models to each cluster. Each model used different parameter values that optimized the model's results. For example, the cluster that contained solely China was best fit to an ARMA(5, 4) model, where the first parameter is the number of lagged timesteps considered and the second is the number of lagged residuals considered. Separating the time series models for these clusters allowed us to accurately fit models to the groupings of countries and then create a larger ensemble model from them, effectively weighing each country's contribution to the model by ecological footprint and allowing us to predict worldwide temperature change from the clusters.

2.2 Ensemble Method

From our geographical region models and cluster models, we created ensemble models in order to compare their performance against the initial global model. We weighed each model evenly in these ensemble models, indicating that models featuring data from fewer countries would contribute more to the ensemble model than models featuring data from more countries. We did this intentionally to allow countries with higher ecological footprints, which were placed in clusters with fewer countries, to have a larger weight in their respective ensemble model. We found that the ensemble models performed much better than their individual parts and that the RMSEs for the ensemble models were lower than those of all of the individual clustered models.

K-Means Clusters	RMSE
1	0.222
2	0.176
3	0.167
4	0.150
5	0.176
Ensemble	**0.138**

Hierarchical Clusters	RMSE
1	0.274
2	0.328
3	0.290
Ensemble	**0.151**

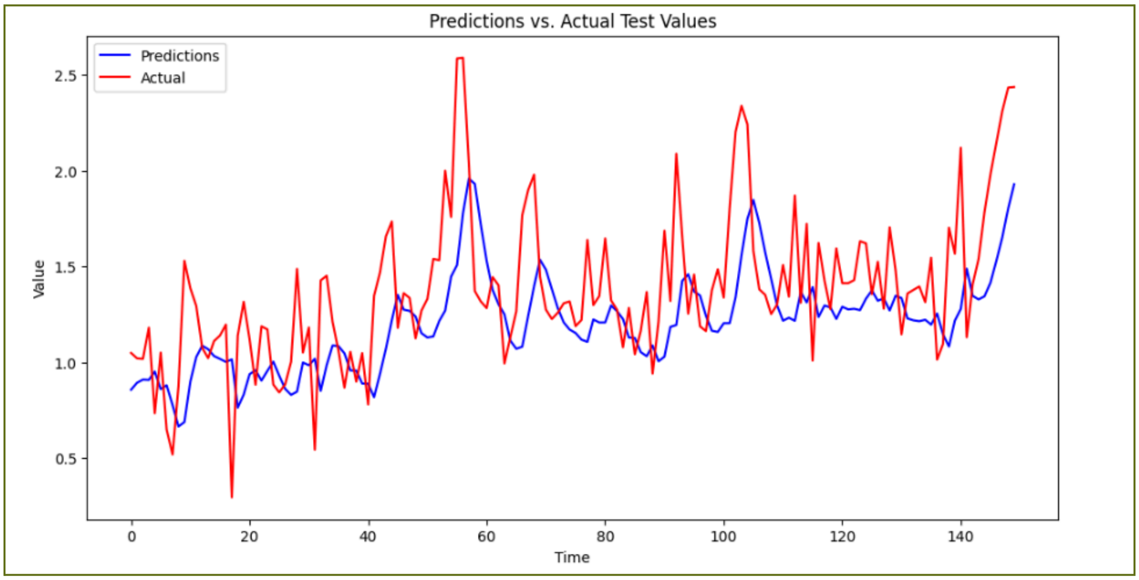
Geographical Region	RMSE
Africa	0.190
Asia	0.081
Europe	0.585
North A.	0.131
South A.	0.286
Oceania	0.368
Ensemble	**0.138**

2.3 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a broad category of neural networks that are designed for the processing of sequential data. RNNs have the ability to remember past information when processing new points, which allow them to detect trends and make future predictions for language and time series data. However, traditional RNNs can struggle to retain information over long sets of data due to a vanishing gradient problem where information begins to get lost as it is passed from one step to another. Long Short-Term Memory (LSTM) is a more recent type of RNN that solves this problem; it incorporates more complex architecture and memory that allows it to better capture long term dependencies. LSTM layers are commonly used in modern day RNNs, along with bidirectional layers, which can analyze sequential data forwards and backwards to detect trends. These layers, combined with more traditional components like deep layers and dropout layers, make RNNs a powerful tool typically utilized in time series and natural language processing applications. With this information, we determined that RNNs could be a useful instrument for the analysis of our climate change data.

We started with an extremely naive implementation of the RNN, consisting of just one LSTM layer (with a RELU activation function) and one dense layer. This initial approach also used 10 epochs, and we used it on global month-by-month average temperature change data. Although this initial approach appeared to perform modestly based on numerical indicators (cross-validation RMSE of 0.949), a glance at the plot of the predicted test data reveals that the naive RNN simply predicted a flat line. This approach completely failed to characterize the local shape of the data or the long-term trends, meaning that we required a more complex network.

Our next RNN attempt on the global temperature change data produced more encouraging results. This network contained a bidirectional layer, three LSTM layers, three dropout layers (to prevent overfitting), and two dense layers. Additionally, we increased the number of epochs to 25 and reduced the sample size at each time step. This allowed the network to better capture local spikes and dips, rather than consistently flattening out due to the consideration of too many previous time steps. Finally, to improve the long-term trend detection, we switched the loss function from mean squared error to mean absolute error, which penalizes deviations from a long-term trend more heavily.



The above figure shows the predicted vs. actual test data values for the global monthly temperature change data. This time, our predictions captured the overall upwards trend of the data as well as the locations of sharp spikes and dips (although not their amplitudes). The network also does not appear to perform significantly worse many timesteps into the future, maintaining strong performance for the 12 years taken in the test data set. The RMSE in this attempt is also less than half of that of the previous attempt at 0.393.

After tinkering with the parameters of this network, we did not find that any other changes significantly affected its performance. Thus, we applied it directly to the grouped data used in earlier discussed models. Groupings were done by continent and the use of groups from both K-means and hierarchical clustering. While results varied drastically, we were able to resolve a few key takeaways from our studies:

- Our RNN models seemed to perform far worse on groupings with a small number of countries. Particularly, in the K-means clusters, China and the U.S. each had their own group. The network performed terribly on these, as spikes in the data tended to be much higher for single countries rather than larger groups, which contained averaged data over many countries. Likewise, the continent of Europe had particularly steep spikes in its data that our model was unable to capture.
- Numerical metrics alone cannot determine a model's usefulness. In many cases, such as with the European temperature change data, the RMSE from the test set was worse than

that of the naive approach. And yet, these models still managed to predict the locations of future spikes and the long-term trends, which the naive model could not do. Although numerical characterizations of model performance certainly must play a role in its evaluation, it became clear to us over the course of our work that models which performed poorly on paper could still have their predictive uses assuming their pitfalls were well understood.

- Monthly temperature change data is simply far too volatile for our models to accurately predict temperature changes on a month-to-month basis. These networks should be used exclusively for spike location and long-term trend prediction rather than for a prediction of the temperature change at a single point in the future. For example, note the large spike located near month 60 in the figure above. Although our RNN could not come close to predicting the amplitude of the spike, it did predict that there would be a significant spike near its actual location. If we used this model to predict temperature changes in future months, for which we do not yet have true data, we might interpret predicted spikes as being within 6 months of when we should actually see an even larger amplitude spike.

In conclusion, RNNs were an extremely effective tool for this type of data analysis, but must be used with caution. Extensive parameter-tuning is necessary to ensure their accuracy and any grouped/clustered data should first be examined with EDA techniques before being fed into the network. Multiple trials should also be done in order to discover the ways in which the results of the network should be interpreted.

3 Conclusions and Predicted Results

We continue to see natural disasters enhanced by climate change all over the world which underscores the importance of predicted climate change data. In this regard, our clusters and predictive models could help identify key indicators associated with the advancement of climate change. For example, our K-means cluster models were highly influenced by each country's size, biocapacity, and total carbon footprint. This could generate many questions about how these factors contribute to a country's perceived climate responsibility. Below is a table of each model's predicted values for the averaged temperature change over the next near, along with the RMSE value of each model. Although other non-numerical factors should be considered when assessing each model's performance, it is clear that many of the time-series models performed particularly well in single-year prediction. This highlights the strength of these types of models, as well as the pitfalls of models like RNNs, which tended to have significantly higher RMSE values. However, each type of model revealed its utility during our project, which taught us that it is always important to tailor statistical models to their intended uses.

		ARMA(p,q)		Ensemble		RNN	
		Pred.	RMSE	Pred.	RMSE	Pred.	RMSE
World		1.880	0.097			1.378	0.393
Geographical Region	Africa	1.537	0.190			1.732	0.354
	Asia	1.722	0.081			1.230	0.705
	Europe	2.864	0.585			0.348	1.727
	North Amer.	1.628	0.131			0.994	0.539
	South Amer.	1.394	0.286			1.406	0.388
	Oceania	1.022	0.368			0.666	0.709
	Total			1.694	0.138	1.108	
K Means				1.725	0.138	1.016	
Hierarchical				1.781	0.151	1.237	

References

(n.d.). *Temperature change on land*. Food and Agriculture Organization of the United Nations; FAO Departments and Offices. Retrieved March 15, 2024, from https://www.fao.org/faostat/en/#data/ET

(n.d.). *NFA 2018 Edition*. Data.World. Retrieved March 15, 2024, from https://data.world/footprint/nfa-2018-edition/workspace/file?filename=NFA+2018+Edition.csv