Climate change: Surface temperature prediction and visualization

Team 33: Abhishek Pandey, Trevor McCrary, Timothy Cody, Kevin Pietruszka, Akhona Msibi

1 INTRODUCTION

Climate change is a huge problem for Earth's ecosystem. If not monitored properly, the changes can be irreversible and can lead to a devastating loss of biodiversity. While everyone knows about climate change and it's potential effects, there aren't many tools that people can use to get an accurate estimate of the same. This leads to a lack of awareness and prevents proactive changes in lifestyle and policy that can help solve the problem.

2 PROBLEM DEFINITION

There are many tools available to measure climate change related metrics like surface temperature change, rise in sea levels etc., however, they are very technical in nature and assume prior understanding of the topic. The tools not being accessible to the general population leads to a lack of awareness and cascades into exacerbating the problem of climate change.

Our team has successfully built an interface to visualize and predict surface temperatures in various cities over the course of the last 100 years and a couple of decades into the future. The interface serves as a resource for the general population to interact with, providing readily available information on climate change and its effects on surface temperature, which was previously only accessible through technical tools that assumed prior understanding of the topic. With this interface, we aim to increase awareness of the issue, and provide a reference point for areas to focus on and policy changes to make. Through this finished project, we hope to contribute to the fight against climate change by making vital information easily accessible and understandable to the general population.

3 LITERATURE SURVEY

A lot of research has been done on learning more about climate change and using mathematical models to measure and predict temperature changes. We referred to the following papers for our project:

[Sanford et al. 2022] The authors talk about improving the accuracy of climate models by incorporating novelty detection techniques. The approach could potentially be useful for this project by flagging potential outliers, which would help ensure that the model is making more accurate predictions. One potential shortcoming of the approach is that it may not be effective in cases where the input data is highly variable or noisy.

[Chen and Hwang 2000] The paper introduces fuzzy time series (FTS) to predict future temperature values based on historical temperature data. FTS forecasts by using linguistic variables and fuzzy logic to model uncertainties in time series data. By incorporating these methods, the approach can capture subtle relationships between different temperature values and help to make more accurate predictions. However, this approach is computationally expensive to implement.

[Vinnikov et al. 1990] The authors examine historical data from a variety of sources to identify trends and patterns in global temperature and precipitation levels over time. The paper could be useful for this project, as it provides a comprehensive overview of global temperature trends and patterns, but the paper focuses primarily on global trends and patterns, rather than specific local or regional variations.

[Chen et al. 2018] This paper shows a specific methodology for performing a time series forecast for temperatures. The paper showed good results in predicting forecast temperatures and the methodology could be applied to our work as the main method or as input into a more complex model. However, this methodology was tested only for one area.

[Knutti et al. 2010] The paper summarizes an IPCC meeting on best practices for building Multi-Model climate projections that focused on all of the methodologies in climate projections and how to best combine them into a coherent whole. This paper helped guide the models we built, but the meeting only described the broad-level processes.

[J. and C 2021] Temperature is affected by several different variables, including humidity, wind speed, and precipitation. The model described includes these factors for more accurate predictions, using what the paper calls an Enhanced Multivariate Prophet (EMP). The very high accuracy makes this approach appealing to our model. However, we will need to find other data sets on rainfall and humidity on a global scale.

[Xiao et al. 2019] The paper describes an ensemble approach using LSTM and Adaboost to improve the prediction accuracy of sea surface temperature. This approach combining heterogeneous models can improve the accuracy of predictions and could be useful for predicting surface-level temperatures. However, the paper's predictions were only over a small time scale (days), whereas we dealt with much larger time scales.

[Johansson et al. 2017] The web-based tool helps people with no background in science visualize the effect and impact of climate change in their future, and provides them with steps that they can take to help. Learning the design principles and technologies used for building the tool will help us improve our interface and features. Some functions and data weren't available or and were not relevant to our project.

[Lumley et al. 2022] This survey paper compares web-based climate change visualization tools by discussing their design and efficacy in meeting the needs of climate change science. The ideas from this paper could be helpful in the decision of the design of our interface and measuring its success. However, much of the discussion and the arguments are subjective.

[Kaspar et al. 2013] The paper details Germany's methods for monitoring climate change. It describes how aggregated climate measures and time series of national and regional temperatures are derived. This will be useful as it guides how to construct and derive a time series model. However, we will have to improve upon the limited scope of the model within the paper.

[Yerlikaya et al. 2020] The paper explores forecasting systems and the impact of climate change on agriculture. It provides examples of forecasting systems to predict future temperature changes and case studies to help the development of our model, but this paper only views the impact of climate change on agriculture.

[Cifuentes et al. 2020] The paper reviews machine learning strategies to forecast air temperatures and examines their advantages and disadvantages to highlight knowledge gaps. The analysis of the models, with input features, gives us ideas of which techniques may work better. However, it presents potential knowledge gaps which we will try to cover in forming a novel solution.

[Keeling and Graven 2021] This paper overviews historical records of global climate that measure atmospheric carbon dioxide. It uses time series observations of atmospheric compounds and processes that regulate global climate, and it details a correlation between atmospheric carbon and local temperature in a city in Hawaii. But it focuses on atmospheric carbon rather than carbon emissions.

[Yue et al. 2017] This paper uses statistical analysis to display the relationship between global surface temperature and observed atmospheric carbon dioxide. It focuses on increased atmospheric carbon being a lagging indicator for global average temperature using radiative forcing. It is, however, primarily focused on the statistical analysis that proves the causality of increased average global temperatures rather than the underlying systems.

[Millar et al. 2017] This paper presents a model of global temperatures with no global carbon emissions. It provides a model of a century-long decrease in temperatures followed by an increase after this period. It also cites models that claim sustained, constant atmospheric temperatures with global zero emissions. A shortcoming is that the paper focuses on multi-century timescales, which might not align with our final model's timescale.

4 PROPOSED METHOD

Our team has successfully built a web-based UI that is accessible to the general population, providing helpful features backed by machine learning. The current tools available in the market are primarily geared towards professionals, having a steep learning curve and are too niche to be useful to a larger set of people. Recognizing these shortcomings, we set out to design a user-friendly UI that could bridge this gap.

Our UI uses d3.js to create a modified choropleth map that allows people to interact and obtain predictions for different regions and cities. By providing a visually appealing and intuitive interface, we hope to encourage more people to engage with this issue and make informed decisions that contribute to a more sustainable future. With this finished project, we aim to promote widespread awareness and action towards combating climate change.

Our UI offers users the flexibility to customize their visual experience by allowing them to change the number of cities displayed on the screen. Users can easily adjust the number of cities based on their interests or needs, whether they want to focus on a specific region or gain a broader view of the data. This feature allows users to tailor the information presented to their unique circumstances and interests, making it easier to gain insights and draw conclusions that are relevant to their specific situation. By offering this level of customization, we hope to make our interface more accessible and useful to a wider audience, enabling a more informed and engaged conversation about the impact of climate change on our planet.

For every great tool, the UI is only as good as the processes and systems underlying it. Since the dataset we're working on contains data from 1980s to 2010s for multiple cities and countries, we have built a time series model to predict the changes in temperature for the next few decades. Most of the methods currently in place use classical models for time series prediction in the context of climate change. Those models work well, but more recent deep learning models like RNNs, LSTMs are shown to work much better on different time series tasks. We experimented with classical as well as deep learning models in our work. The models that we tested and considered for our work are as follows:

- Random Forest
- Decision Tree
- XGBoost
- Fuzzy time series
- S-ARIMA
- LSTM

The models are trained on 80% of the dataset (1900-1988) and evaluated by using their predictions on the remaining 20% (1989-2010) of the data. It is ensured that the training and validation sets are out of time, i.e. There are no date overlaps between the two ensuring no data leakage.

One of the main challenges we faced while building our climate change prediction model was the tradeoff between computational time and accuracy. As we experimented with different machine learning algorithms, we found that some models offered higher accuracy but required significantly longer computational time to train and make predictions. Conversely, some models could produce predictions more quickly, but with lower accuracy. To address this tradeoff, we developed a solution where we did all the necessary computations offline and stored the results in CSV files. This allowed us to provide our users with accurate and up-to-date predictions without sacrificing the performance of the web-based UI. By decoupling the computational complexity from the user interface, we were able to provide a seamless experience to our users and ensure that our predictions were both accurate and timely.

The results of the experiments are further discussed in the following section.

5 EXPERIMENTS/ EVALUATION

As previously mentioned, the primary goal for the experiments was to form an accurate time series model to predict the changes in temperature over time, reflecting the impacts of climate change. In conducting the experiments across the various models, we sought out the most accurate model for these predictions, using the mean squared error (MSE) as a core measure. Computational time was not a consideration as these computations were performed offline to be stored without real-time calculation.

Below is a table of our experimental models with their respective mean squared error when tested on our dataset.

Model	**Validation MSE**
Random Forest	0.311
XGBoost	2.67
Decision Tree	0.412
Fuzzy time series	N/A
S-ARIMA	2.93
LSTM	0.843

5.1 Random Forest

The random forest model yielded strong results in our evaluation. During the evaluation, this model presented the lowest MSE value across all tested models. This made it a promising candidate for final selection as it was fairly accurate in its predictive performance. However, the model was computationally-intensive, taking significant time to process and output the results. Computational time was not within our testbed, however, this observation was considered in the final selection process.

5.2 XGBoost

This model lacked accuracy in its predictive performance, resulting in a fairly high MSE value. This lack of accuracy limited its efficacy in our use case, and eliminated it from further consideration.

5.3 Decision Tree

The decision tree model yielded promising results from testing. It was a fairly accurate model, scoring the second lowest MSE value. As a result, it was heavily considered in the final model selection process.

5.4 Fuzzy time series

This model presented several limitations, mainly in that it could only accurately forecast 4 months into the future. As such we did not move forward with this model in our evaluation as it did not present the required predictive solution.

5.5 S-ARIMA

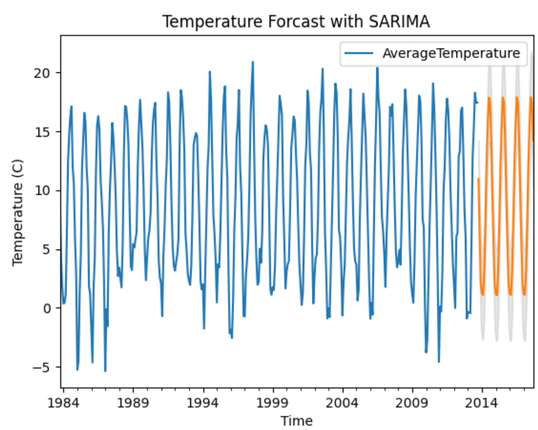

Figure 1: S-ARIMA prediction values against ground truth

The S-ARIMA model presented unique results in our testing. This model output a confidence interval for future values. Additionally, the output was slightly increasing over time, similar to historical data and models listed in our literature survey. However, the predicted output was not centered on the historical values, which led to a very high MSE value. Hence, the model did not yield accurate results for this use case.

5.6 LSTM

The Long Short term memory model presented promising results in our testing, It is especially useful for sequential data like our dataset, so this was an anticipated result. Hence, it achieved a reasonable MSE value on its prediction, and it closely followed the test data. However, as shown in figure 3, there was uncertainty in its forecasting of future values, lowering confidence in its ability to perform our predictive goals.

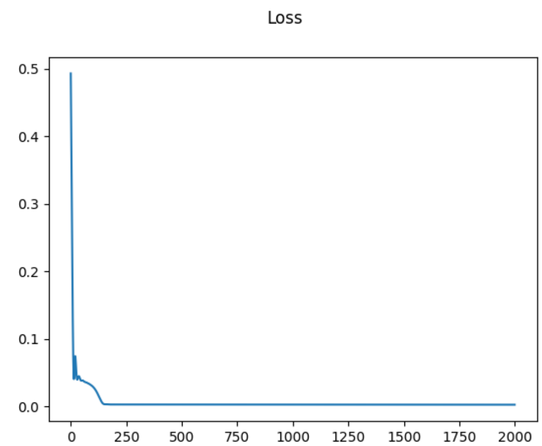

Figure 2: loss per epoch of LSTM model

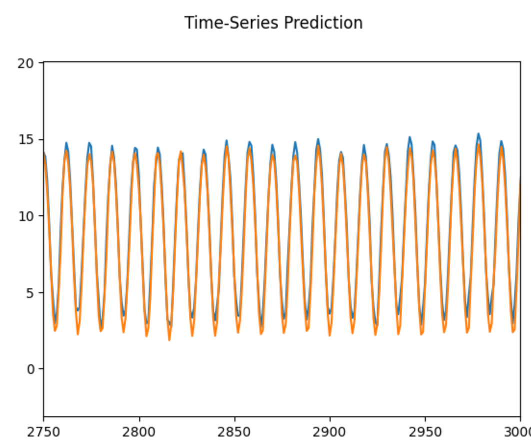

Figure 3: LSTM model prediction

5.7 Final Model

We selected the Random Forest model as our final method for predicting temperature changes due to its superior performance, as demonstrated by the lowest Mean Squared Error (MSE) among the models tested.

Despite its accuracy, however, building the Random Forest model proved to be a computationally-intensive process, requiring significant processing time to complete. To mitigate this challenge, we chose to perform all of our temperature predictions offline, before loading the results into CSV files for display in our visual interface.

By leveraging the strengths of the Random Forest model while optimizing our processing approach, we were able to create a powerful tool that provides reliable and accessible temperature predictions for different regions and cities. These were loaded into our visualization to then allow for user interaction with this data, both historical and future predictions.

5.8 Visualization

Our finished tool (Figure 4) is designed to provide users with a visually-engaging and informative experience that highlights the impacts of climate change in different regions and cities. The interface is built using d3.js, a powerful JavaScript library for data visualization that enables us to create interactive and dynamic charts and graphs.

The heart of our interface is a 3D interactive map that displays temperature changes over time for different regions and cities. Users can select the time frame they wish to view, as well as the specific cities they are interested in, and the map will display the corresponding temperature changes. The color scheme of the map is designed to be intuitive, with warmer colors indicating higher temperatures and cooler colors indicating lower temperatures.

In addition to the map, our interface also features several other interactive elements that allow users to explore temperature changes in more detail. A line chart displays the temperature changes over time for the selected city. Users can also access detailed information about specific cities, including historical temperature data and projections for the future. Overall, our interface is designed to be both accessible and informative, allowing users to gain a deeper understanding of the effects of climate change on different regions and cities.

In order to create our visualization tool, we needed a way to map the temperature changes of various cities onto a global scale. We achieved this by using longitude and latitude coordinates for each city, allowing us to accurately position them on the globe. By overlaying temperature data onto this map, users can easily visualize how climate change affects different regions and cities across the world. This approach allows for a global perspective on climate change, highlighting the interconnections between different regions.

6 CONCLUSIONS AND DISCUSSION

If the tool is successful and widely adopted, it'll have a great impact on the awareness people have about climate change. Increasing awareness will lead to quicker policy changes to mitigate the problems and will aid in decreasing the rate or arresting climate change and will potentially save a lot of lives and biodiversity.

There are multiple ways to measure the impact of a web-based tool, the most basic of which is through measuring user engagement. We can also conduct user studies with different personas to understand how people from different facets of life use the tool and if it has led to some changes in their lifestyles.

The largest risk, in this case, would be incorrect predictions, however, those are being addressed by the robust statistical and machine learning methods that we'd be utilizing for our predictions and recommendations. The rewards are great and if successful can change the course of life for many people out there.

In summary, our project aims to provide a user-friendly interface that allows individuals to explore the effects of climate change on different regions and cities. By using machine learning methods, we were able to develop a time series model that predicts changes in temperature for the next few decades. Through our web-based UI, users can interact with the data and gain insights into the potential impacts of climate change on their local environment.

Our project has several limitations that could be addressed in future work. A limitation is the computational resources required to train and run these machine learning models. While we were able to overcome this limitation by doing computations offline, there is still a need for more efficient algorithms that can provide accurate predictions with low inference latency and online training.

Despite these limitations, our project has the potential to have a significant impact on increasing awareness of climate change and its potential effects. By providing a user-friendly interface that allows individuals to explore the data, we hope to encourage more people to take action towards mitigating its effects. Our project

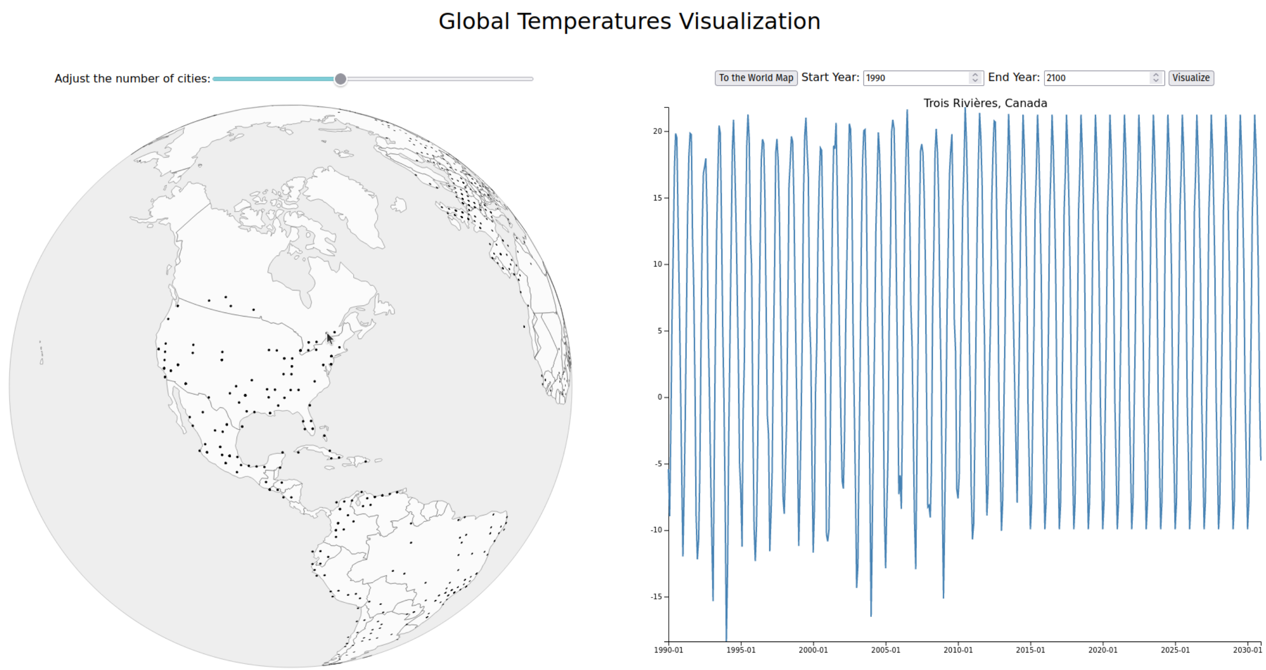

Figure 4: UI of the visualization tool

also has implications for policy changes, as increased awareness can lead to quicker policy changes that aid in decreasing climate change and potentially impacting the environment for the better.

In future work, others could extend our project by incorporating additional data sources, such as emissions data, to provide a more comprehensive picture of the effects of climate change on the environment. Overall, we believe that our project has the potential to make a significant contribution to the fight against climate change and inspire more people to take action towards building a more sustainable future.

7 CONTRIBUTIONS

Task	Weeks	People
Source and Clean Data	1	All
Develop and Train Model	2-4	All
Test Model on Data	5-6	All
Build Visualization	5-7	All
Verification of Results and Iterations	6-7	All

All team members have contributed a similar amount of effort.

REFERENCES

Peng Chen, Aichen Niu, Duanyang Liu, Wei Jiang, and Bin Ma. 2018. Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. *IOP Conference Series: Materials Science and Engineering* 394, 5 (jul 2018), 052024. https://doi.org/10.1088/1757-899X/394/5/052024

Shyi-Ming Chen and Jeng-Ren Hwang. 2000. Temperature prediction using fuzzy time series. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 30, 2 (2000), 263–275. https://doi.org/10.1109/3477.836375

Jenny Cifuentes, Geovanny Marulanda, Antonio Bello, and Javier Reneses. 2020. Air Temperature Forecasting Using Machine Learning Techniques: A Review. *Energies* 13, 16 (2020). https://doi.org/10.3390/en13164215

Jagannathan J. and Divya C. 2021. Time Series Analyzation and Prediction of Climate using Enhanced Multivariate Prophet. *International Journal of Engineering Trends and Technology* 69 (10 2021), 89–96. https://doi.org/10.14445/22315381/IJETT-V69I10P212

Jimmy Johansson, Tomasz Opach, Erik Glaas, Tina-Simone Neset, Carlo Navarra, Björn-Ola Linnér, and Jan Ketil Rød. 2017. VisAdapt: A Visualization Tool to Support Climate Change Adaptation. *IEEE Computer Graphics and Applications* 37, 2 (2017), 54–65. https://doi.org/10.1109/MCG.2016.49

F. Kaspar, G. Müller-Westermeier, E. Penda, H. Mächel, K. Zimmermann, A. Kaiser-Weiss, and T. Deutschländer. 2013. Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre. *Advances in Science and Research* 10, 1 (2013), 99–106. https://doi.org/10.5194/asr-10-99-2013

Ralph F. Keeling and Heather D. Graven. 2021. Insights from Time Series of Atmospheric Carbon Dioxide and Related Tracers. *Annual Review of Environment and Resources* 46, 1 (2021), 85–110. https://doi.org/10.1146/annurev-environ-012220-125406 arXiv:https://doi.org/10.1146/annurev-environ-012220-125406

Reto Knutti, Gabriel Abramowitz, Matthew Collins, Veronika Eyring, Peter J. Gleckler, Bruce Hewitson, and Linda Mearns. 2010. IPCC expert meeting on assessing and combining multi model climate projections. https://www.ipcc.ch/publication/ipcc-expert-meeting-on-assessing-and-combining-multi-model-climate-projections/

Sam Lumley, Renee Sieber, and Robert Roth. 2022. A framework and comparative analysis of web-based climate change visualization tools. *Computers & Graphics* 103 (2022), 19–30. https://doi.org/10.1016/j.cag.2021.12.007

Richard J Millar, Jan Fuglestvedt, Pierre Friedlingstein, Joeri Rogelj, Michael J Grubb, H Damon Matthews, Ragnhild B Skeie, Piers M Forster, David J Frame, and Myles R Allen. 2017. Continued global warming after CO2 emissions stoppage. *Nature Climate Change* 7, May (2017), May.

Clayton H Sanford, Anna Kwa, Oliver Watt-Meyer, Spencer Clark, Noah Brenowitz, Jeremy McGibbon, and Christopher Bretherton. 2022. Improving the predictions of ML-corrected climate models with novelty detection. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*. https://www.climatechange.ai/papers/neurips2022/10

K. Ya Vinnikov, P. Ya Groisman, and K. M. Lugina. 1990. Empirical Data on Contemporary Global Climate Changes (Temperature and Precipitation). *Journal of Climate* 3, 6 (1990), 662 – 677. https://doi.org/10.1175/1520-0442(1990)003<0662:EDOCGC>2.0.CO;2

Changjiang Xiao, Nengcheng Chen, Chuli Hu, Ke Wang, Jianya Gong, and Zeqiang Chen. 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing of Environment* 233 (2019), 111358. https://doi.org/10.1016/j.rse.2019.111358

Bayram Ali Yerlikaya, Seher Ömezli, and Nazlıcan Aydoğan. 2020. *Climate Change Forecasting and Modeling for the Year of 2050*. Springer International Publishing, Cham, 109–122. https://doi.org/10.1007/978-3-030-49732-3_5

Xu Yue, Shuwei Sun, Jing Cui, Jianping Zhao, Jie Liu, and Xianhong Xie. 2017. Exploring Granger causality between global average observed time series of carbon dioxide and temperature. *Journal of Atmospheric and Solar-Terrestrial Physics* 155 (2017), 11–16.