BIO Web of Conferences **130**, 06010 (2024) https://doi.org/10.1051/bioconf/202413006010
BFT-2024

A data-driven approach to assessing climate issues in coastal cities

Daniyal Kair[1] and *Amandyk* Kartbayev[1*]

[1]Kazakh-British technical university, Almaty, Kazakhstan

Abstract. This paper addresses the critical issue of climate change, a subject of considerable debate over recent decades. To provide a clear and unbiased perspective, we have developed an innovative model to forecast the future impacts of global warming. Utilizing datasets exclusively sourced from official government records ensures the integrity and neutrality of our data. Various modeling techniques were employed to optimize the fit and enhance the accuracy of our predictions. By analyzing temperature and sea-level data, we applied regression techniques to predict future values. We focused on preprocessing the data using MATLAB, and developed a function capable of forecasting both temperature and sea level changes. The findings indicate a grim future, particularly for coastal cities, which are expected to experience severe consequences.

1 Introduction

Global warming remains one of the most contentious topics in modern environmental discussions. Over the past few decades, the debate has polarized researchers and policymakers alike, with opinions varying from severe impending crises to dismissive scepticism. Some researchers, using probabilistic models, argue that the effects of climate change are devastating and imminent, posing unprecedented threats to human civilizations. In contrast, others believe these impacts are negligible and often politicize the issue, undermining the urgency of addressing it.

Despite these differing viewpoints, the implications of global warming continue to be a significant concern, not just for governmental agencies but for the public at large [1]. The undeniable reality is that no one can predict the future with absolute certainty. However, employing probabilistic models enables us to forecast potential scenarios with a degree of realism. The relevance of these predictions becomes apparent when considering that 40% of the population in the Netherlands faces a drowning risk due to rising sea levels—a direct consequence of global warming. Similarly, major urban areas like Manhattan could see significant portions of their land submerged if current trends continue [2].

Acknowledging the gravity of these projections, the research aims to conduct a thorough investigation into the effects of global warming. We hypothesize that global temperature increases serve as a proxy for broader global warming effects, notably sea-level rise [3]. Our study utilizes impartial, government-sourced datasets to maintain objectivity and employs

* Corresponding author: a.kartbayev@gmail.com

cross-validation methods to refine our models for optimal accuracy. We assess the impacts globally, predicting trends in temperature and sea-level changes across different regions, as illustrated in Figure 1.

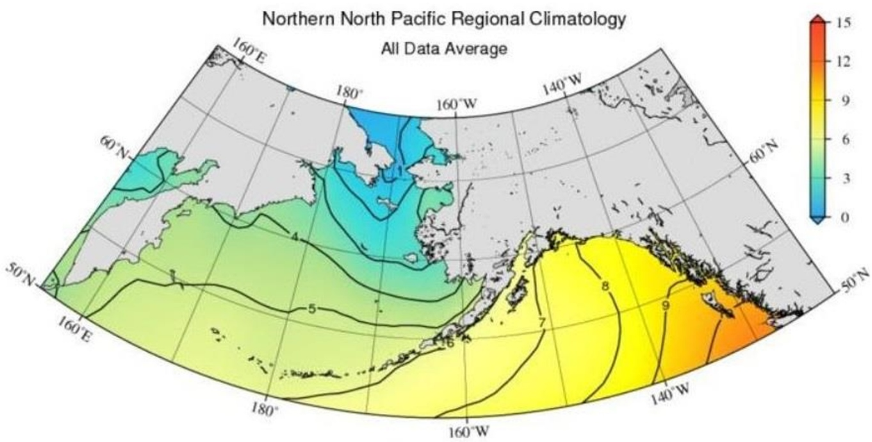

Fig. 1. Annual temperature at the surface (one degree grid).

Moreover, our research considers factors typically overlooked by standard probabilistic models, such as economics and natural resource consumption. The influence of economic activities on global warming is complex and has historically followed a non-linear trajectory, often resembling exponential growth. Similarly, the extraction and use of natural resources have intensified, contributing further to environmental degradation. By incorporating these elements into our analytical framework, we aim to present a more comprehensive model that reflects the multifaceted nature of global warming.

The impact of the study was inspired by the contentious nature of the climate debates. Intrigued by claims dismissing climate change as a fabricated issue, we were compelled to delve deeper into this topic. Our exploration began with a unique approach to data acquisition, focusing on underutilized datasets from National Oceanographic sources, which provided fresh insights away from the conventional narratives found in popular media. This approach not only enriched our understanding of the subject but also allowed us to develop original analytical models that harness this vast array of data effectively [4]. Our preliminary research activities in this field, including the choice of explored datasets and analysis of relationships through regression models, have critically reshaped the direction and basis of our study. Future sections will explore our methods and results, adding to the discourse on global warming.

2 Problem statement

The study of global warming's impact through temperature and sea level changes presents a complex data challenge, involving multiple datasets and a variety of analytical approaches [5]. In our research, we utilize three primary datasets: the Climatology dataset and the Sea Level Trends dataset from the National Oceanic and Atmospheric Administration (NOAA), along with the Global Surface Temperature dataset from NASA.

The current Climatology dataset encompasses a broad spectrum of ocean temperature measurements taken from various depths across multiple regions, including the Northern North Pacific, and the Gulf of Mexico. This expansive geographical coverage is vital for examining local temperature variations and understanding their broader environmental

https://doi.org/10.1051/bioconf/202413006010

implications. Simultaneously, the Global Surface Temperature dataset tracks temperature trends from 1880 to the present, presenting a historical perspective on climate shifts. It maps these trends across a global grid, facilitating detailed studies of enduring temperature trends, which are depicted in Figure 2. The analysis reveals a slow but steady increase in global temperatures, prompting us to divide the analysis into two geographical regions based on latitude: above and below 57 degrees north. This distinction ignites debate as initial results indicate opposing sea level reactions in these regions: an unexpected decrease in the upper region contrasts sharply with a predicted rise in the lower region.

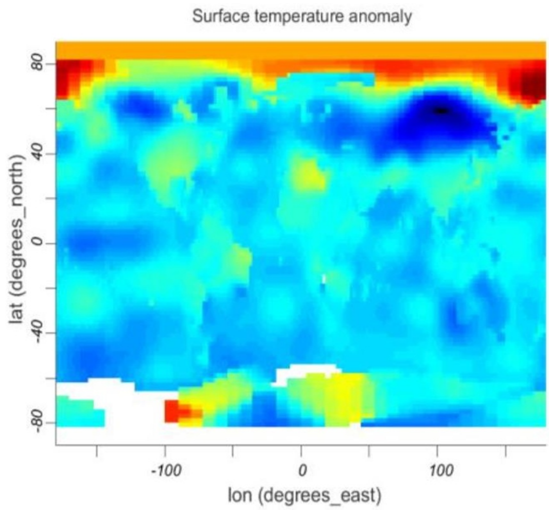

Fig. 2. Global temperature data structured in a grid format.

Handling such diverse data requires robust preprocessing, for which we employ Python and MATLAB. It was instrumental in handling the .nc files (NetCDF format), which are well-suited for array-oriented scientific data but are cumbersome for direct use in modeling due to their complex structures. Python scripts were developed to transform these .nc files into more manageable formats, such as specialized .csv files that accommodate the irregular column counts typical of real-world data [6].

A significant preprocessing task was the integration of temperature and sea level data. Given that sea level data is sparser and more localized compared to the more uniformly distributed temperature data, aligning these datasets required meticulous geographic and temporal matching. This process required sophisticated data manipulation techniques, particularly cross-joins and self-joins, to align temperature and sea level observations by longitude and latitude accurately. Such techniques are essential for handling the intricacies of the datasets, which include irregular column counts and disparate data types like sea level measurements. The integration is challenging due to the disparate nature of the data sources; temperature data is globally abundant, whereas sea level data is more sparse and localized.

This disparity necessitates a sophisticated approach to data processing. We aim to match each sea level measurement with corresponding temperature anomalies by longitude and latitude to ensure accurate modeling. The preprocessing includes managing large data sizes, which poses potential memory overflow issues [7]. Handling the large size of the datasets posed significant technical challenges, particularly the risk of memory overflow during data processing. We addressed these issues by optimizing memory management techniques in our scripts and incrementally processing data to minimize peak memory usage during data manipulation.

Furthermore, we plan to explore the critical relationships between temperature anomalies and sea level changes over time. This involves not only merging and preprocessing data from multiple sources but also establishing a methodological framework that can accommodate the nuances of each dataset [8]. The problem we face is developing a predictive model that can accurately reflect the intricate dynamics between global temperature changes and sea level fluctuations[9]. This model is designed to account for temporal fluctuations in the data, aiming to deliver insights that wide public understanding of the effects of global warming.

3 Results

We developed Python scripts to accommodate the irregular column counts typical of real-world climate data. The significant preprocessing task involved the integration of temperature and sea level data, as shown in Table 1. Given that sea level data is more localized compared to the uniformly distributed temperature data, aligning these datasets required meticulous geographic and temporal matching. This integration was crucial for accurate modeling of climate impacts, particularly in coastal areas.

Table 1. Dataset for predictive models to provide detailed risks of floods.

Category Name	Description	Unit	Data Type	Example Sources
Annual Precipitatio n	Total yearly rainfall	millimeter s	Continuou s	Meteorologic al datasets
Tidal Range	Difference between the high tide and low tide levels	meters	Continuou s	Tidal gauge measurements
Storm Surge Frequency	Number of storm surges per year	count	Discrete	Historical weather records
Storm Surge Height	Maximum height of storm surge above normal tide	meters	Continuou s	Historical weather records
Sea Level Anomalies	Deviations from the average sea level	meters	Continuou s	Satellite data
Flood Events per Year	Number of flooding events recorded annually	count	Discrete	Local disaster records
Coastline Erosion Rate	Rate at which the coastline is eroding	meters/yea r	Continuou s	Coastal surveys
Urbanizatio n Rate	Rate of urban expansion in coastal areas	percentage /year	Continuou s	Satellite imagery, census data
Flood Defense Integrity	Structural and operational condition of flood defenses	score (1-10)	Ordinal	Engineering assessments
Groundwat er Level Trends	Long-term changes in groundwater levels	meters	Continuou s	Hydrological studies

Temperature data, which is abundant and globally distributed, was merged with the sparser sea level data through complex data joining techniques. These techniques included cross-joins and self-joins, necessary to synchronize temperature and sea level observations

 https://doi.org/10.1051/bioconf/202413006010

based on precise geographic coordinates. The sophistication of these data processing steps is illustrated in Figure 3 using the Folium python library, which presents the distribution of temperature anomalies along the east coast of the US [10].

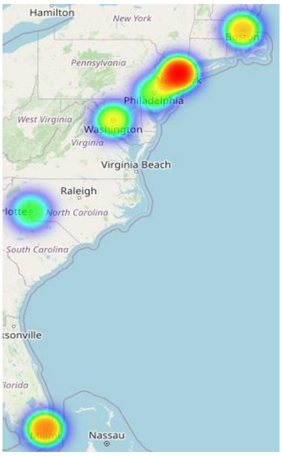

Fig. 3. Surface temperature data of east coastal cities.

The preprocessing of large datasets posed significant challenges, including potential memory overflow due to the large size of the data files. However, the Python scripts facilitated efficient data management and transformation, allowing for effective integration of temperature and sea level data. This integration is crucial for modeling and predicting climate change impacts on urban coastal areas. The significant differences between the types of data, such as global temperature data and localized sea level data, required a complex approach to achieve localized climate modeling. The integrated datasets were used to generate heatmaps to visualize climate impacts [11]. Figure 4 shows the heatmaps focused on New York City, demonstrating the areas most susceptible to increased temperatures and potential heatwaves. This visualization underscores the urban heat island effect and its amplification by global warming.

Fig. 4. New York City areas susceptible to increased temperatures.

Additionally, a specific focus was placed on flooding dangers due to rising sea levels. Using the processed data, we created detailed risk maps for coastal cities. Figure 5 depicts the flooding danger zones in New York City, highlighting areas with high vulnerability to storm surges and sea-level rise (as shown in the map). This visualization serves as a critical tool for urban planning and resilience strategies against climate-induced flooding [12].

Fig. 5. New York City areas with high vulnerability to sea-level rise.

The heatmaps and risk maps generated as part of this research not only provide insights into current conditions but also serve as predictive tools for future climate scenarios. This is crucial for developing strategies to mitigate the effects of climate change, particularly in vulnerable urban coastal areas.

4 Conclusion

This study illustrates the power of advanced data processing techniques in addressing the complexities of climate data integration. By effectively merging temperature and sea level datasets, we have enhanced our understanding of climate impacts on coastal cities. We highlight the potential pitfalls and challenges in handling large-scale climate data, such as memory overflow, and showcases the efficacy of Python scripting in overcoming these hurdles. The disparity in data distribution between temperature and sea level measurements required a sophisticated approach to data processing, which was adeptly managed through the use of advanced data joining techniques.

This research adds to the conversation about climate change by offering a data-based way to look at climate problems in coastal cities. It gives various climate researchers the tools and information they need to create effective plans to fight the harmful impacts of climate change on at risk city areas.

References

1. Hahn C, Garcia-Marti I, Sugier J, Emsley F, Beaulant A-L, Oram L, Strandberg E, Lindgren E, Sunter M, Ziska F., Climate, **10(12),** 192 (2022)
2. H. Chen, Q. Zhang, and Y. Birkelund. Energy Reports, **vol. 8,** Supplement 13, pp. 661-668 (2022)
3. L. Donadio, J. Fang, and F. Porte-Agel. Energies, **vol. 14,** no. 2, p.338 (2021)
4. A. Kartbayev. Refining Kazakh Word Alignment Using Simulation Modeling Methods for Statistical Machine Translation. Lecture Notes in Computer Science, Springer, **vol. 9362,** pp. 421-427. (2015)

5. M. A. K. Azad, A. R. M. T. Islam, M. S. Rahman, and K. Ayen. Natural Hazards, **vol. 108,** pp. 1109–1135. (2021)
6. K. Wilgan, W. Rohm, and J. Bosy. Multi-observation meteorological and GNSS data comparison with numerical weather prediction model. Atmospheric Research, **vol. 156,** pp. 29–42. (2015)
7. A. Kartbayev, U. Tukeyev, S. Sheryemetieva, A. Kalizhanova, B.K. Uuly. Journal of Theoretical and Applied Information Technology **13(96),** 4103-4113 (2018)
8. L. Naveen, H.S. Mohan. *Atmospheric Weather Prediction Using Various Machine Learning Techniques: A Survey*. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 422-428. (2019)
9. S. Madan, P. Kumar, S. Rawat and T. Choudhury. *Analysis of Weather Prediction using Machine Learning & Big Data.* 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, pp. 259-264. (2018)
10. G. Verma, P. Mittal and S. Farheen. *Real Time Weather Prediction System Using IOT and Machine Learning,* 2020 6th International Conference on Signal Processing and Communication (ICSC), Noida, India, 2020, pp. 322-324. (2020)
11. Xiaoli Ren, Xiaoyong Li, Kaijun Ren, Junqiang Song, Zichen Xu, Kefeng Deng, Xiang Wang. Deep Learning-Based Weather Prediction: A Survey. Big Data Research, vol. **23.** (2021)
12. K. M. S. A. Hennayake, R. Dinalankara and D. Y. Mudunkotuwa. *Machine Learning Based Weather Prediction Model for Short Term Weather Prediction in Sri Lanka,* 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), Negambo, pp. 299-304. (2021)