# A Machine Learning Analysis of Climate Change & Human Health Projections in Pakistan


*By*

**Zarak Jamal Khan**

**PIDE2020FMPHILETS02**

**Supervisor**

**Dr Amena Urooj**

**PIDE School of Economics**

**Pakistan Institute of Development Economics**

**Islamabad**


**2023**

Pakistan Institute of Development Economics

# Pakistan Institute of Development Economics, Islamabad

***PIDE School of Economics***

## CERTIFICATE

This is to certify that this thesis entitled **"AMachineLearning Analysis of Climate Change & Human Health Projections in Pakistan"** submitted by **Mr. Zarak Jamal Khan** is accepted in its present form by the School of Economics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in Master of Philosophy in Econometrics.

Supervisor: Dr. Amena Urooj Signature:

External Examiner: Dr. Rehana Siddique Signature: Rehana Siddique

Head,
PIDE School of Economics: Dr. Shujaat Farooq Signature:

## Author's Declaration

I **Zarak Jamal Khan** hereby state that my MPhil thesis titled **"A Machine Learning Analysis of Climate Change & Human Health Projections in Pakistan"** is my own work and has not been submitted previously by me for taking any degree from Pakistan Institute of development economics or anywhere else in the country or world at any time. If my statement is found to be incorrect even after my graduation the university has the right to withdraw my MPhil degree.

Date: 14th November, 2023

Zarak

Signature of Student

## Dedication

To my mother, in her loving memory

*To her whose heart is my heart's quiet home,*
*To my first Love, my Mother, on whose knee*
*I learnt love-lore that is not troublesome;*

- Christina Rossetti

# Acknowledgement

*I express my deepest gratitude to my supervisor, Dr. Amna Urooj, for her invaluable guidance and unwavering support throughout this thesis. Special thanks to Dr. Arif Goheer, Principal Scientist - GCICS, and Mr. Saqib Sultan Malik for their expert insights on climate change, enriching the depth of my research. I am also thankful to my friends Aqsa Gul and Azwar Muhammad Aslam for their constant encouragement and camaraderie. Their contributions have been pivotal in shaping the successful completion of this work.*

# Abstract

This thesis explores the profound challenge of global warming and climate change in Pakistan, focusing on deterioration of health. Through the utilization of machine learning techniques on climate change and causes of death datasets and is set to investigate the linkages among drivers of the climate change specifically in the context of Pakistan. The analysis reveals robust correlations between climate change and victims of respiratory diseases, while the associations with the victims of digestive problems and cardiovascular diseases are found to be comparatively less significant.

The examination of causality emerges as a potential solution to overcome the limitations of current machine-learning approaches. The interdisciplinary nature of causality, drawing from fields such as epidemiology, economics, statistics, and computer science, underscores the significance of collaboration and knowledge exchange. The research focuses on one of the fundamental task that is causal discovery. By employing causal discovery tools, the study delves into investigation and exploration of the causal linkages between climate change and human deaths, identifying both direct and indirect relationships with the drivers of climate change and leading causes of mortality in Pakistan.

While the study provides valuable insights into the intricate relationship between climate change and human health, further comprehensive analysis and extensive data are needed to obtain more precise and accurate results. The thesis emphasizes the necessity of a multidisciplinary approach to deepen our understanding of causality and climate change's health implications, leading to evidence-based policies and interventions.

In summary, this thesis underscores the urgent need to address climate change as a critical issue in Pakistan. By unraveling the correlations between climate change and the victims of respiratory diseases, it contributes to the existing body of knowledge. The research highlights the importance of causality in comprehending complex phenomena, advocates for cautious interpretation of correlations, and demonstrates the potential of causality in addressing the limitations of machine learning. By further exploring the causal pathways and gathering extensive data, the thesis aims to enhance our understanding of the relationship between climate change and human health, paving the way for effective strategies to safeguard the well-being of the population in Pakistan.

Keywords: Machine Learning, Climate Change, Forecasting, CMIP6, & NASA-GISS-E2-1-H Model

## List of Figures

# List of Tables

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| **AR6** | Assessment Report 6 |
| **CC** | Climate Change |
| **CH4** | Methan |
| **CIMP** | Coupled Model Inter-comparison Project |
| **CO2** | Carbon dioxide |
| **CVD** | Cardiovascular Disease Deaths |
| **DALYs** | Disability-adjusted life years |
| **DDD** | Digestive Diseases Deaths |
| **DFM** | Dynamic factor Model |
| **ESMs** | Earth System Models |
| **GCMs** | General Circulation Models |
| **GDP** | Gross Domestic Product |
| **GHGs** | Green House Gases |
| **GISS** | Goddard Institute for Space Studies |
| **GLOF** | Glacial lake outburst flood |
| **GPR** | Gaussian Process Regression |
| **IHME** | Institute for Health Metrics and Evaluation |
| **IPCC** | Intergovernmental Panel on Climate Change |
| **ML** | Machine Learning |
| **NASA** | National Aeronautics and Space Administration |
| **NCDs** | Non-Communicable Diseases |
| **NMBD** | Nervous Mental and Behavioral Disease Deaths |
| **NO2** | Nitrogen dioxide |
| **PCA** | Principal Component Analysis |
| **PCMCI** | Momentary conditional independence |
| **PM** | Particulate matter |
| **RCMs** | Regional Climate Models |
| **RCPs** | Representative Concentration Pathways |
| **RDD** | respiratory disease Deaths |
| **RSME** | Root Mean Squared Errors |
| **SSPs** | Shared Socioeconomic Pathways |

| | |
|---|---|
| **SVR** | Support Vector Regression |
| **Tas** | Tas Surface temperature |
| **UNFCCC** | United Nations Framework Convention on Climate Change |
| **WG-II** | Working Group – II |
| **WHO** | World Health Organization |

# TABLE OF CONTENTS

# CHAPTER 1

# INTRODUCTION

Technology is transforming human life at an extraordinary scale and pace. Artificial intelligence (AI) technology is at the core of unprecedented breakthroughs in the field of economics, medicine, engineering, and business. The advancement in machine learning technology which falls under the umbrella of AI, allows machines to outperform humans in many of the scientific domains and execute tasks that, until recently, were thought only to be carried out by humans, such as cognitive functions and language generators. It is quite safe to say that machine learning advancement has the potential to revolutionize the scientific realm at an equally fundamental level.

Traditionally Machine learning methods were utilized to perform isolated and quintessential prediction problems. The successful development of Artificial intelligence techniques in the fields of drug discovery (Protein designing), pattern matching (Visual AIs), neuro-symbolic AIs, language generators (Open AI's GPT-3 or Gopher) or IBM's Deep Mind, had unwrapped a new world of opportunities and possibilities. The automated inference technique is one of the exhilarating innovations that has gained immense value in the process of scientific discovery and creating solutions to real-world problems. Assisting researchers in choosing which hypotheses to test, which experiments to conduct, and how to extract principles and postulates detailing the range of phenomena.

The evident rise of machine learning in natural and social sciences can be attributed largely to the availability of big data. With big data availability machine learning techniques can make data speak for itself. Understanding the inherent and innate structures, relationships, and linkages within the data are made easier with machine learning techniques. For instance, in medical sciences, most traditional approaches can only function in data-driven predictive modes, which makes them incapable of realizing precision-based goals such as timely assessment and matching of the patients with the most appropriate preventative and therapeutic measures. Data interpretation and the implementation of scientific breakthroughs in health care require a deeper familiarity with machine learning and its processes (Raita et al., 2021).

Integrating machine learning algorithms with healthcare domain expertise, and further coupling these two with causal discovery is crucial to bringing about a qualitative transformation in medicine that will lead to better patient outcomes as "big data" becomes an increasingly common tool for analysing quantitative information (Raita et al., 2021).

Climate change research has evolved manifolds due to the usage of AI and ML techniques in handling Big data. One of the biggest hurdles in climate change policymaking was the availability of conclusive evidence, which wasn’t possible without innovative spatial and temporal econometric modelling, computational power, and technological advancement. Evidence and resources for learning about climate change and global health challenges are usually not readily accessible or available. AI and ML tools are extensively used to broaden the reach of the literature and improve our understanding of the effects of climate change on human health (Scheelbeek et al., 2021).

Conclusive evidence is a bedrock essential for global economic policy formulation. With the health and climate change knowledge base expanding fast, evaluations to guide public policy are becoming more labour and resource-intensive (Scheelbeek et al., 2021).To advance, implement, and make economic decisions concerning climate change, a cutting-edge knowledge base will be needed – and that is where ML and AI come into the picture, providing data-driven solutions. The econometrics knowledge base is essential for the expansion of the machine learning domain. Machine learning can play an impactful role in many broader strategies for reducing and responding to climate change and its goals (considering its impact on human health).

Machine learning is crucial for better understanding climate change and human health nexus through causal linkages. Finding a link between climate change and human health can be done using several different machine-learning tools, but the results will vary depending on the specifics of the analysis and the quantity of data at hand. This thesis on climate change, human health, and machine learning attempt to address the lack of conclusive evidence, projecting a futuristic trend of mortality and employing causal discovery. Moreover, evaluating the machine learning tools for forecasting; and application of causal discovery tools for exploration, mapping, and understanding of the causal linkages between climate change and mortality data. Utilizing the NASA climate change forecast data as input to an ML-based framework for projecting what might happen in the future using the best-performing models through the analysis of the relationship between climate change indicators and mortalities on a national scale.

## 1.1. Motivation

There is a growing belief among practitioners, researchers, and scientists that "big data" has the answers to all scientific concerns as there is a rapidly growing amount of data. The advent of data science particularly the large amount of data that gave birth to the “health data science” and “emerging climate change- machine” learning domain in recent years provides an opportunity to evaluate the data-driven perspective.

In this regard integrating machine learning algorithms with the climate change domain, and further coupling these two with causal discovery is crucial to bringing about a qualitative and quantitative transformation in solving the climate change crisis. Resulting in better outcomes as "big data" becomes an increasingly common tool for analysing quantitative information and comprehensive evidence-building. Data interpretation and the implementation of scientific breakthroughs into climate change science require a deeper familiarity with machine learning and its processes.

Most traditional approaches in climate change sciences, for instance, can only function in data-driven predictive modes, which makes them incapable of realizing precise policy matching economic goals with the most appropriate actions and measures in given time constraints. To address, the lack of evidence for efficient and effective policy-making, employing machine learning tools with the availability of big data is crucial. Given the present climate change alarming situation in Pakistan, it is urgent to explore and research the climate change situation to mitigate its adversities.

To this end, this thesis tries to attempt to explore and investigate the climate change issue in Pakistan through an ML-based framework for analysing the relationship between climate change indicators and mortality on a national scale. Addressing the lack of comprehensive and conclusive evidence of climate change on human health with the help of machine learning tools. Projecting futuristic morality trends and projecting what might happen in the future using the best-performing models. This study attempts at exploring the causal linkages through the application of causal discovery is crucial for health policy-making in Pakistan. The study comprises of application and evaluation of the machine learning tools in the emerging climate change and machine learning domain. To sum up, the manifold contribution of this study encompasses both; the methodologies employed and the innovative application of the ML framework that provide new and significant empirical insights into climate change research.

## 1.2. Research Problem

The literature on climate change in Pakistan is scant in many aspects, particularly concerning human health, adaptation, mitigation, and effective policy-making. The scarcity of the existing literature on the subject matter is a hurdle in the creation of effective policy instruments. One of the significant obstacles faced in climate change mitigation policymaking is the lack of conclusive evidence on a national scale with variability on temporal and spatial lines.

In the aftermath of devastating floods, and rising global and national climate challenges there is the utmost need for intervention policy to mitigate the socio-economic crisis as a direct result of climate change, which is central to economic growth (Sadiq & Khalil, 2022). The prime objective of this study is to predict the futuristic trend of climate change-induced mortality in Pakistan and to provide conclusive evidence on climate change's impact on human health, through the usage of novel machine learning techniques.

Another major gap is the exploration and identification of the causal pathways through which climate change affects human mortality, in particular, which has multiple hurdles in data availability and climate change literature necessary for adaption and mitigation. This study aims to fill these gaps through data-driven evidence-building. For comprehensive evidence building, utilization of big data is essential and evaluation of best machine learning tools in regard to climate change data and its projections.

*Based on the narrative of the research problem as stated in the preceding text and narrowing our research problem into "**A Machine Learning Analysis of Climate Change & Human Health Projections in Pakistan**" and have operationalized my topic into the following research questions and objectives.*

### 1.3. Research Questions

1. How mortality in relation to climate change will behave in future in case of Pakistan?
2. What are the possible avenues of application of machine learning tools in emerging climate change and machine learning domains?
3. Which interlinkages exist between climate change and mortality in Pakistan through casual discovery?

## 1.4. Objectives of the Research

- The prime objective of this study is to project the futuristic trend of mortality in relation to climate change in Pakistan by collecting conclusive evidence on climate change and human health (mortality) using machine-learning techniques.
- To apply machine learning tools in the emerging climate change and machine learning domain.
- To investigate the interlinkages between climate change & mortality for understanding and exploring the causal pathways between climate change and mortality in Pakistan, and to visualize through spatial mapping.

## 1.5. Significance of research

This study will serve as a foundation for recognizing important research and knowledge gaps and necessitating funding and resources for research climate change, health, and machine learning domains. Leveling the ground for future research in climate change and machine learning, through the application and evaluation of machine learning tools. This study reveals significant gaps in the literature for several climate-health pathways as well as temporal and spatial disparities in the data regarding health consequences, with vulnerable regions being underrepresented.

The significance of this study is three-pronged. Primarily, utilizing machine learning techniques may result in the creation of "live" evidence platforms. These could assist governments in setting priorities and supporting actions that mitigate and upcoming effects of climate change on human health. Furthermore, imploring developmental economists, environmentalists, social scientists, and policy practitioners to explore real-world issues, not through traditional approaches but rather through ML tools. Secondly, this study pioneered bridging the gap in the literature on the climate change health nexus in the context of south Asia and Pakistan. The existence of scarce literature on the subject matter restricts and narrows down the lens of climate change and policy practitioners to develop and implement policies. Scientific evidence from low-income nations, which are most affected by the health effects of climate change, is currently quite scarce. There is very little research on how climate change affects physical as well as mental health. Although the literature review found a few instances of climate change adaptation activities that claimed to improve health, were generally lacking high-quality data.

# CHAPTER 2

# Review of Literature

## 2.1. Literature on Machine Learning

Forecasting macroeconomic variable is crucial for creating effective policy solutions. The policy measure can't be developed or executed without accurate forecasting and prediction. Real Gross Domestic Product (GDP) growth is one of the most important and widely used macroeconomic indicators but predicting it is time-consuming and challenging because official data is only released with a one-quarter lag. Policymakers routinely develop policies and carry them out without having access to trustworthy data because of lags in the delivery of data. From this perspective, it would be quite advantageous to be able to precisely estimate and predict real GDP growth over time. It is challenging to forecast and foresee macroeconomic variables such as real GDP growth. Traditional economic forecasting models frequently employ d top-down, theory-driven methodologies to forecast data while taking into account the causal relationship between the variables. In these processes, the forecasters had to apply their economic judgment and intuition about the data and methodology used. The models could generate incorrect forecasts if there are any problems with the forecasters' underlying assumptions. Contrary to many traditional economic forecasting methods, machine learning models mainly concentrate on innate prediction.

Compared to traditional economic forecasting methodologies, machine learning methods are more flexible and may generate forecasts devoid of presumptions and conclusions. In fact, in line with technological advancement and the improvement in predictive capacities, machine learning models have been actively used in a range of domains, from forecasting traffic flows to predicting property prices. Machine learning techniques outperformed conventional and traditional econometrics methodologies in many studies, such as the real estate sector forecasts of American housing prices by (Plakandaras et al., 2019). Furthermore, Medeiros & Gabriel Vasconcelos (2019) show that machine learning is effective in predicting and forecasting where the data is low-frequency when it comes to inflation forecasting.

In addition to Machine learning, the dynamic factor model (DFM) framework and its extensions have been widely employed for macroeconomic time series forecasting in recent decades (Fang, 2015; Geweke, 1977; Sargent & Sims, 1975). Stock & Watson, (2002) is one of the earliest and most widely used in the field. The authors of that study demonstrated that the DFM could accurately predict the Federal Reserve Board's industrial output index using 149 months of observable macroeconomic data, outperforming the then-current gold standard models. They initially estimated the universally shared underlying traits using principle component analysis (PCA), and then they ran a linear regression of the relevant variable on the factors found.

DFM has been the go-to strategy for predicting macroeconomic time series variables for quite some time, despite the recent emergence of machine learning techniques in the domain of macroeconomic time series forecasting. For example, (Li & Chen, 2014) used 107 macroeconomic indicators from the US economy in their dataset and compared their forecasting ability to that of the DFM utilizing least absolute shrinkage and selection operator (LASSO) techniques. At all forecast horizons for all projected variables, they discovered that the LASSO-based techniques performed better than the DFM. Least angle regression was utilized by (Bai & Ng, 2008) to choose a set of predictors to forecast the variables of interest. Following the application of PCA to the selected variables, (Efron et al., 2004) used the discovered predictors in the regressions. They discovered that LARS estimates for retail sales, personal income, total employment, CPI, and industrial production were better for different subsamples of data than those from PCA using the complete data set. A substantial number of New Zealand-based predictors were utilized by (Eickmeier & Ng, 2011) discovered that one model's strategy performed better than the others.

One of the most often employed models for climate change is CMIP5. It is used by NASA and other organizations that track changes in the climate, and they have introduced new techniques like ModelE2 for atmospheric modeling (Pizzulli et al., 2021). In the context of the correlation between the factors selected i-e drivers of climate change and causes of death (Human health), (Pizzulli et al., 2021) used machine learning  approaches. The main causes of climate change and the mortality from specific diseases were analyzed in various ways utilizing neural networks and machine learning. The findings indicated a substantial association between anthropogenic climate change and human health; some diseases were found to be primarily related to risk factors, while others required a larger number of variables to establish a correlation. Developed a forecast of human adversity due to climate change. The anticipated outcome indicates that an increase in casualties is correlated with a generally growing trend in climate change components.

The ability to predict long-term global warming can be extremely useful in many different fields, including climate research, agriculture, energy, medicine, and many more. To estimate annual global warming from previously recorded values over India.(Hema et al., 2019) compared the performance of multiple Machine Learning algorithms (Linear Regression, Multi-Regression Tree, Support Vector Regression (SVR), and Lasso). They focused on building an accurate relationship between the average yearly temperature and potential components like concentrations of carbon dioxide, methane, and nitrous oxide using a reliable, efficient statistical data model on a huge data set. It was found that linear regression has the highest accuracy for predicting temperature and greenhouse gases of all the available methods. Additionally, it was discovered that CO2 is the main factor in temperature change, followed by methane ($CH_4$ and nitrogen dioxide ($N_2O$). The greenhouse gas emissions and temperature data analysis and projections revealed that global warming can be substantially decreased within a few years. Because numerous animals in addition to people are harmed by the global temperature, lowering it can benefit the entire world.

In a systematic review of the literature, (Berrang-Ford et al., 2021) methodically identified and mapped the scientific literature on climate change and health using supervised machine learning and other neural language processing (NLP) techniques such as topic modeling and geo-parsing. The review revealed that impact studies predominate in the literature on climate health, with specialized discussions of mitigation and adaptation strategies. All-cause mortality and the prevalence of infectious diseases were the most investigated health outcomes, whereas air quality and heat stress were found to be the most often examined exposures. The most often researched climate-related risks included seasonality, extreme weather events, heat, and weather variability. They also identified significant gaps in the literature about the outcomes of climate induced illnesses related to mental health, undernourishment, and maternal and child health.

A systematic review conducted by (Scheelbeek et al., 2021) on the effects of climate change adaptation efforts on public health from low and middle-income countries made use of the "Global Adaptation Mapping Initiative" database, which includes 1682 papers on climate change adaptation measures. The study found some evidence that climate change adaptation strategies may benefit human health; the overall dearth of data is alarming and represents a significant wasted learning opportunity. They proposed an urgent need to place more emphasis on the funding, design, assessment, and standardized reporting of the consequences of climate change adaptation policies to enable evidence-based policy action.

Using a NASA climate change forecast as input, an AI-assisted framework was devised to examine the relationship between climatic variability and the leading causes of fatal diseases across the globe and to forecast the future using the best correlation models ((Miller et al., 2014).

The significance and viability of automated machine learning to adequately map the big literature on climate health have been evidenced by (Berrang-Ford et al., 2021). These can serve as important inputs for analyses of the global climate and health. The lack of studies on available climate change responses is alarming and could seriously impede the development of evidence-based strategies to lessen the repercussions of climate change on human health.

Due to their increasing prevalence in daily life, the fact that artificial intelligence systems rely on uninformed associations poses serious challenges. There is a growing need for reliable Machine learning solutions as a result the scientific community invested ample time and energy into the study of causality, moving the emphasis away from philosophy and empirical trials, toward the fields of artificial intelligence and machine learning. Causation is a tool that could be used to address some of the problems that modern ML struggles with (Pearl, 2018)).

Causation is a wide-ranging notion that touches on many disciplines. Searching for possible causal correlations in observational data by combining statistics, machine learning, and data mining (Guo et al., 2016). As was previously mentioned, it is often broken down into “finding” and “inference” of causes. Analyzing the data and developing models to show the connections hidden within it is the job of causal discovery. One purpose of causal discovery is to investigate the potential outcomes of intervening in a system.

As we transition towards an era of big literature, the integration of machine learning tools with systematic evidence mapping techniques can help to retain transparency and scrutiny of scientific evaluations.

### 2.2. Climate change and Health Outcomes Nexus

Variations in the factors that cause disease or hinder treatment can cause seasonal fluctuations in death rates to shift over time. Understanding these shifts in mortality rates is instrumental in assessing the effectiveness of efforts to reduce seasonal deaths. However, the time and extent of peak-to-trough mortality rate fluctuations, as determined by the local environment, age group, gender, and medical causes are largely unexplored (Rau, 2007; Rau et al., 2018).

Building on the literature, (Parks et al., 2018) examined the seasonality in mortality rates by age group and sex in the USA and its sub-national climatic zones from 1980 to 2016 using geo-coded mortality data. They looked at the seasonal trends of general death across the board and by-cause mortality in the USA using Wavelet analytical tools, which were previously employed to examine the dynamics of meteorological phenomena and infectious diseases. To explain the timing of mortality range extremes, circular statistics and centre of gravity analysis techniques were used. They found that the percentage difference between the death rates in the months with the highest and lowest mortality rates has evolved. Death rates for both men and women in the late 40's age group peaked in December through February period and were at their lowest in the period from June through August due to injuries and disorders of the cardiorespiratory system. Between 1980 and 2016, the percentage difference in death rates between the peak and lowest months did not alter among climate zones. Additionally, it showed that after the 1990s, seasonality in all-cause mortality for children under five years dissipated considerably.

Evidence from thermal biology maintained that the spread of vector-borne diseases peaks at moderate temperatures and decreases at extremes. (Guo et al., 2016; Hansen et al., 2010; Liu et al., 2014; Mordecai EA et al., 2013; Paull et al., 2017). However, for the majority of vector-borne diseases, thermal optimums and limits are still unknown. Bridging the gap in the literature about causality between temperature variation and transmission of diseases, (Shocket et al., 2018) developed a mechanistic model for the heat response of the Ross River virus, a significant pathogen spread by mosquitoes in Australia and the Pacific Islands, they used data from lab experiments that were designed to evaluate the performance of viruses and mosquitoes over a wide temperature range. They found that at moderate temperatures (26.4 °C), transmission reaches its maximum, while at thermal bounds (17.0 °C and 31.5 °C), its zeros out. The model correctly predicted that transmission is widespread around the year in tropical regions but seasonal in temperate regions, causing the seasonal surge in human cases across the country. Most Australians reside in temperate regions, where transmission is anticipated to rise. In tropical regions, where mean temperatures are already thermally optimal, transmission is likely to decline. The model predictions have policy implications for the Australian government and mosquito control organizations to facilitate better long-term planning.

This study by (Gaythorpe et al., 2020) conducted the first analysis of the possible impact on disease burden due to climate change, which has a significant impact on Yellow Fever transmission in South America and Africa. They forecasted the intensity of transmission over the African endemic region under four different climate change scenarios by extending an existing Yellow Fever transmission model to take rainfall and a temperature appropriateness index into account. These transmission projections were used to further analyze the burden change in 2050 and 2070. Results indicated a disproportionately varying disease incidence across the zone. The likelihood that there would be an increase in annual deaths in 2050 was found to be 93.0 percent [95 per cent CI (92.7, 93.2%)] which implies future disease control initiatives will be more challenging.

The study conducted by (Rerolle et al., 2021) explains the linkage between forest fall and Malaria transmission in the Greater Mekong Sub-region (GMS) by using high-resolution forest coverage data from Hansen et al. (2010) and monthly malaria incidence data from 2013 to 2016. (Rerolle et al., 2021) demonstrated that the loss of forests in rural regions boosts malaria transmission in the early years while transmission rates dwindle afterward. Deforested areas' geographic position also mattered. Malaria rates were unaffected by deforestation within one to ten kilometers of the settlements. The transmission of malaria was found to be impacted by deforestation in a 30-kilometer radius farther away. Deforestation in heavily forested areas appears to be the primary driver of results. These results demonstrate that activities in the forest affect Malaria spread in the GMS.

Studying the Tasmanian population of Bolivia, (Jaeggi et al., 2021) tested the links between social status and health in communities where social hierarchies are not strong. They evaluated the relative wealth and income of 870 households from forty Tasmanian settlements to several outcomes, such as blood pressure, self-rated health, stress hormone levels, depressive symptoms, and multiple diseases. Results indicate that not all of the health outcomes examined were negatively impacted by poverty and inequality, contrary to what has previously been observed in industrialized cultures. However, those with lesser earnings or those who resided in more unequal neighborhoods had greater blood pressure.

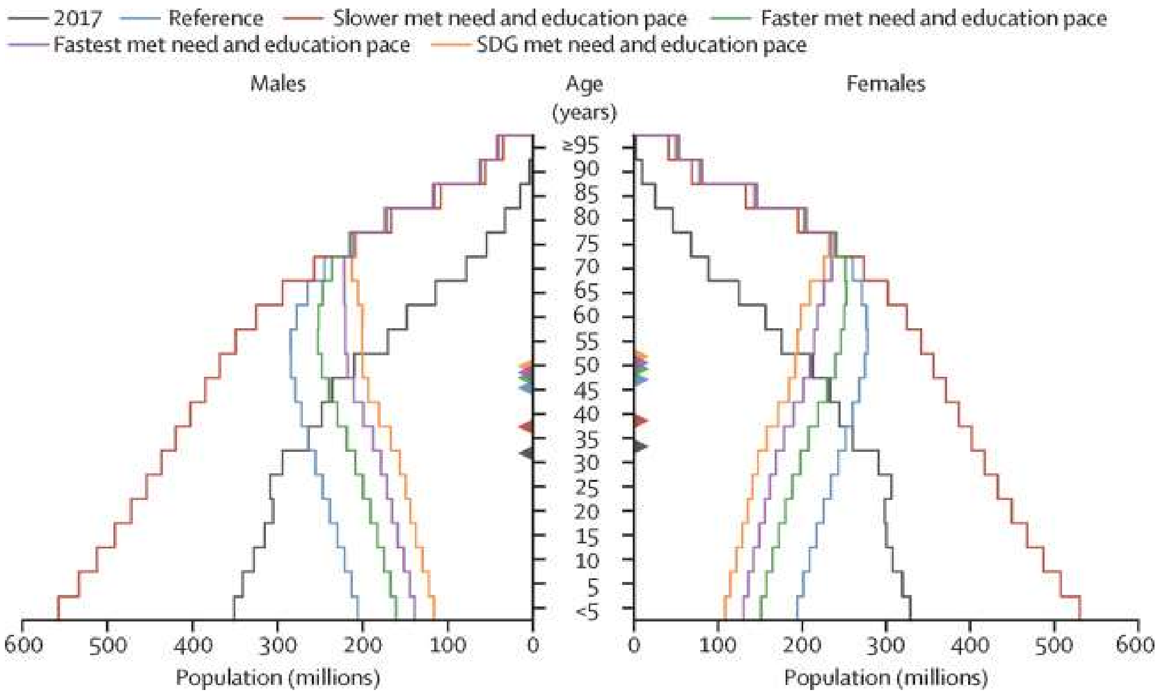
Examining the relationship between birth weight and fire-based particulate matter (PM2.5) in 54 lower-middle-income countries from 2000 to 2014. J. Li et al. (2021) in their sibling-matched case-control analysis of 227,948 neonates. The amount of fire sourced PM2.5 was used to determine each newborn's gestational exposure to landscape fire smoke (LFS). Using a fixed-effects regression model, they explored the relationships between birth weight disparities between matched siblings and within-group changes in LFS exposure. Additionally, the dichotomous outcomes of extremely low birth weight (ELB) or low birth weight (LBW) were observed. The results showed that neonatal health is compromised by prenatal exposure to LFS.

A well-designed experiment by Kunze et al. (2022), showed that increasing mean temperatures and wild fluctuation in temperature regimes both have an impact on host-pathogen interaction. The authors showed how disease dynamics would change because of climate change, and pointed out the necessity for processes that underlie species interactions in changing habitats.

**2.3. Factors explaining Climate Change & Causes of Death Linkages**

Anthropogenic activities that damage the environment have a long-term effect on economies and people's well-being on local and as well as on global scales. Anthropogenic emissions of greenhouse gases have caused an unusual surge of global warming, melting ice sheets, ice caps, and glaciers—which together produce the majority of the world's freshwater—have been exposed to degradation. Lack of clean water for irrigation and drinking is being caused by high glacial melting, catastrophic glacier advances, glacial lake outburst flood (GLOF), enormous recessions, and a negative mass balance of glaciers. In addition to the global exhaustion of freshwater supplies, it is raising the alarms of the deglaciation age. Environmental deterioration has also been putting human settlements and physical infrastructure at serious risk, demanding considerable restoration. Similar sociocultural effects of varied degrees are shown by climatic calamities, especially for people who rely completely and directly on scarce freshwater resources. Therefore, the lack of freshwater has established the foundation for the local and global social matrix, ending societal traits like relationships, trust, and networking to find a common solution to the issues. Instead of getting access to restricted freshwater, it has further enhanced communities' vulnerability to natural disasters and raised the likelihood of societal conflicts. It further accelerates the process of resource exhaustion, which widens the scope of societal issues.

***Figure 2.1:*** Change in the proportion of Pakistani older and younger population

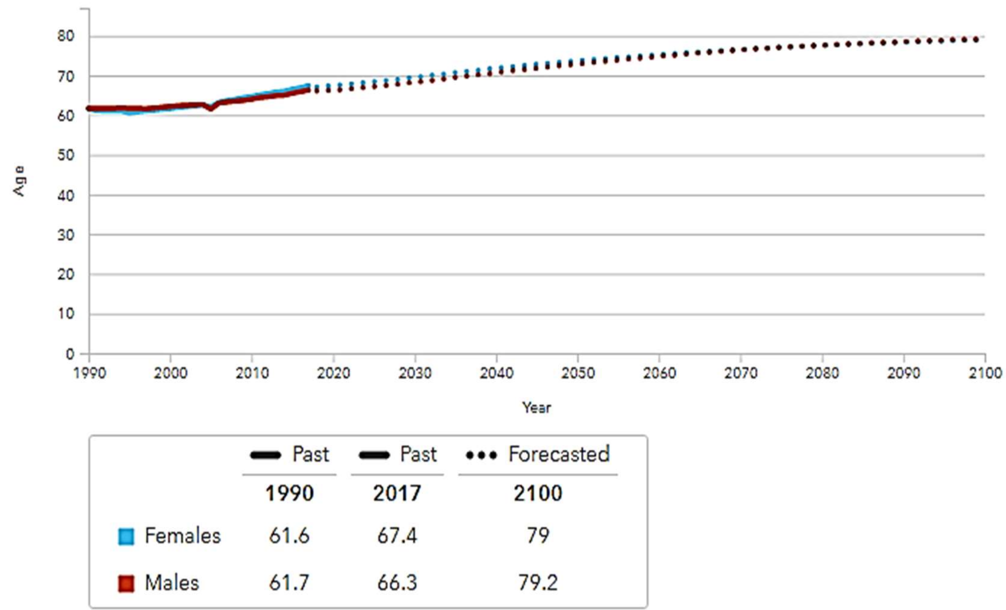



Source: Adapted from: (Vollset et al., 2020).

The figure 2.1 shows the age distribution of the population for both men and women is depicted in the graphs for 1990, 2019 (the reference scenario), and 2100 (the reference scenario). These are the outcomes from the 2017 Global Burden of Disease-based forecast statistics.

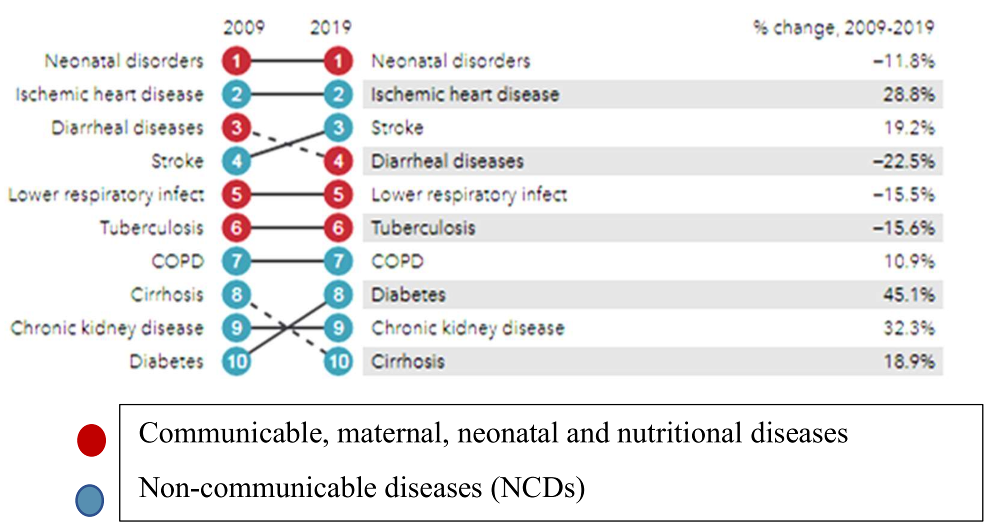
Alarming death rates are being reported in Pakistan as a result of exposure to elevated levels of air pollution, particularly particulate matter (Anjum et al., 2021). Approximately 22,000 adult premature mortalities and 163,432 DALYs (disability-adjusted life years) are lost each year in Pakistan as a result of the disease burden brought on by outdoor air pollution, according to World Bank estimates (WHO, 2020). Over the past two decades, there has also been a dramatic rise in the concentration of important air pollutants in Pakistan, including NOx, O3, and SO2. Several studies have also found that the air quality in and around Pakistan's major cities often exceeds EPA standards (Anjum et al., 2021; Hema et al., 2019; WHO, 2014). In 2019, PM2.5 concentrations in Lahore consistently exceeded national and World Health Organization guidelines. Significant effects of these increasing pollutant concentrations in Pakistan have been indicated in a few studies, but the full scope, nature of contributing variables, and consequences remain imperfectly known (Anjum et al., 2021).

**Figure 2.2:** Life Expectancy among Pakistani Males & Females




As the health and societal costs associated with breathing polluted air continue to rise, air pollution and particle matter have become a single, global problem in recent years. Particularly in developing nations like Pakistan that lack effective warning, protection, and management mechanisms, the intensity and effects of these dangers have increased. Numerous epidemiological investigations have connected poor air quality to an array of illnesses and rising death rates (Anjum et al., 2021).

**Figure 2.3:** Causes of Deaths in Pakistan for the Year 2009 & 2019



The Figure 2.3 showcase the percentage change in the deaths against the leading ten causes of deaths in Pakistan. The blue colored are the NCDs while the red color identifies the communicable, maternal, nutritional and neonatal diseases (Anjum et al., 2021).

**Figure 2.4:** Risk Factors for Death and Disability in Pakistan given by the % change in the deaths.

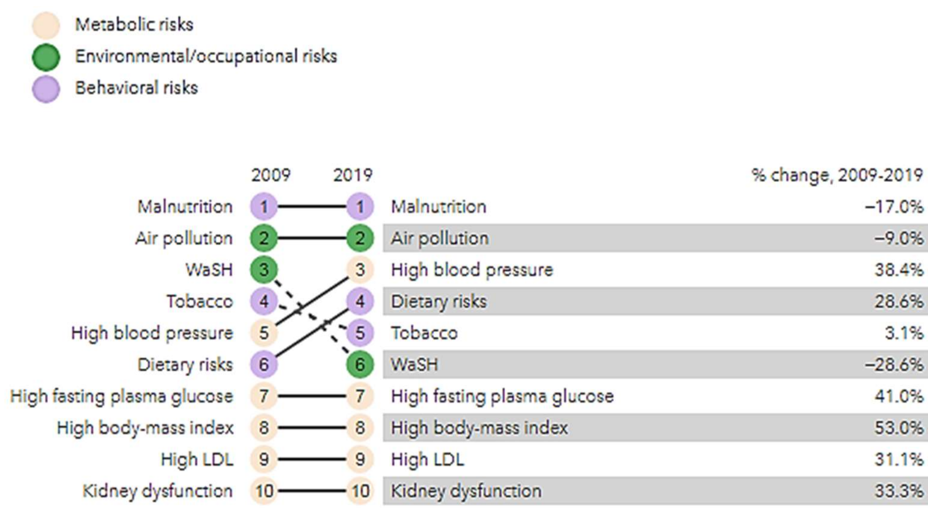
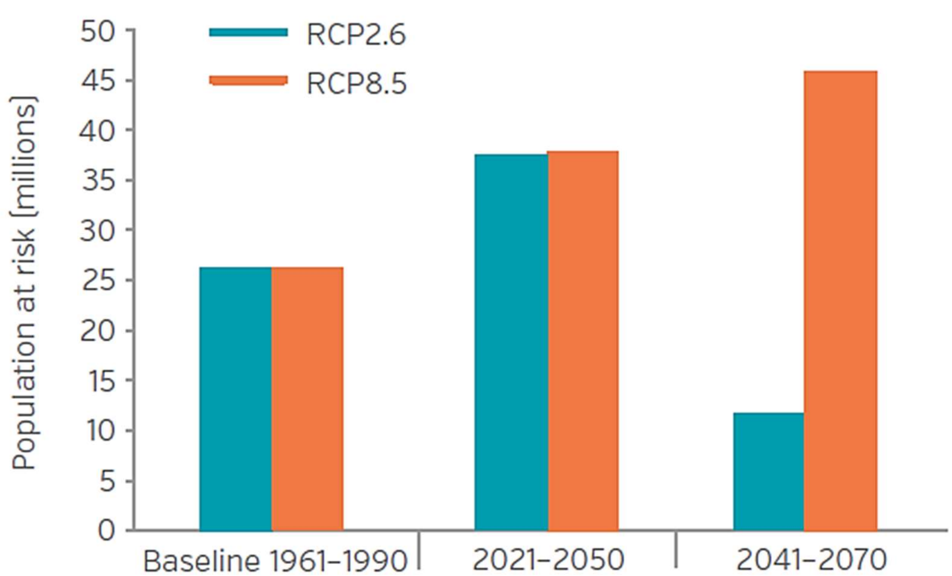


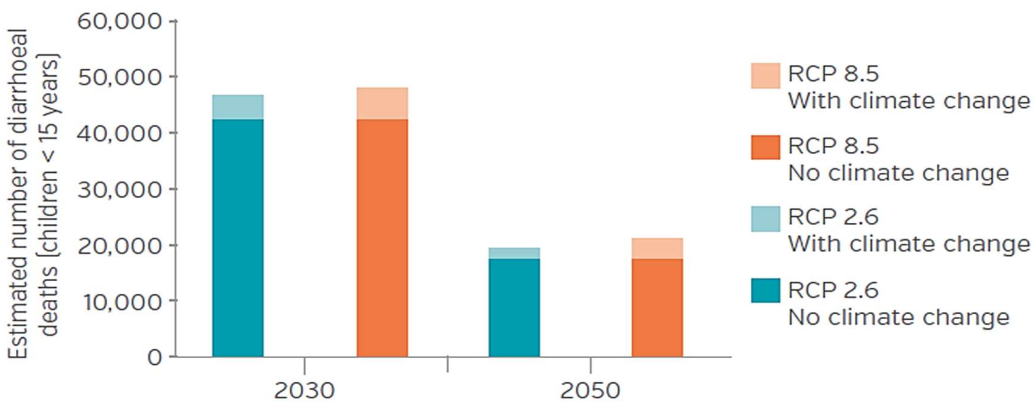

**Figure 2.5:** Infectious and Vector-Borne Diseases among the Pakistani Pollution

Malaria Incidence in Pakistan (Population in millions)



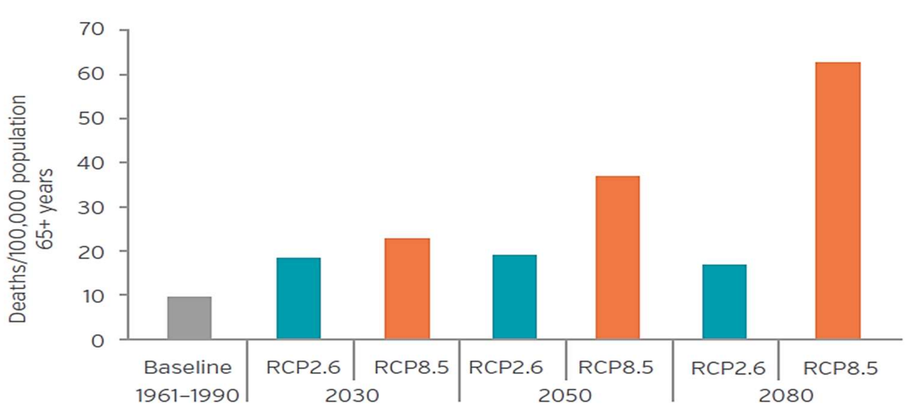
By 2070, a substantially alarming emission case predicts approximately forty-six million individuals to be vulnerable to malaria each year. However, with significant reductions in global emissions, the vulnerable lot of population can reach around twelve million annually by 2070. Additionally, the population threatened by the risk of malaria is expected to grow in regions where malaria prevalence remains constant (Anjum et al., 2021).

**Figure 2.6:** Diarrheal Disease amongst the children in Pakistan




The graph illustrates the estimated number of diarrheal deaths in children under 15 years old in the baseline year of 2008 and projections for future years under different emissions scenarios. In the high emissions scenario, it is projected that climate change will contribute to approximately 11.7% of the projected 48,200 diarrheal deaths in 2030. However, as the total number of diarrheal deaths is expected to fall by 2050 It is estimated that the percentage of mortality linked to climate change will increase to about 17.0%.

**Figure 2.7:** Heat-Related Mortality in Population Aged 65 and Over: Pakistan



The graph depicts the projection of heat-related deaths in the elderly population aged 65 years and over. Under a high emissions scenario, it shows a notable increase in the number of deaths. By 2080, the projected rate is expected to reach approximately 63 deaths per 100,000 populations, in contrast to the estimated baseline of fewer than 10 deaths per 100,000 annually during the period from 1961 to 1990. However, a swift reduction in global emissions has the potential to mitigate this impact. It could result in a significant decrease in heat-related deaths, with the projection indicating a possible limitation to around seventeen deaths in every one lac people by 2080.

The most frequently studied factors in climate health literature as categorized by (Berrang-Ford et al., 2021; Scheelbeek et al., 2021) are presented below:

Table 2.1: Indicators in climate change literature

| **Climate and Weather Events** | **Climate Change Forcing/Feedbacks** | **Health Outcome** |
|---|---|---|
| Extreme Temperatures | Temperature, $CO_2$ & $CH_4$ Concentration | All-cause mortality, Mental & Behavioral Disease |
| Frequent Heat waves | Temperature, $CO_2$ Concentration | All-cause mortality, Nervous, Mental & Behavioral Disease |
| Extreme Precipitation and Flooding | Rainfall, Temperature, $CO_2$ Concentration | Vector-Borne Diseases Digestive Diseases |
| Air Quality | $CO_2$ Concentration, $CH_4$, $NO_2$ CFCs | Respiratory Diseases, Infectious Diseases, Mortality, Nervous, Mental & Behavioral Disease |
| High Ocean Temperature & Acidification | Temperature, Anthropogenic forcing | Infectious Diseases, |
| Coastal Flooding | Seasonal Rainfall Temperature | Vector-Borne Diseases Digestive Diseases |
| Weather variability | Anthropogenic forcing, $CO_2$ Concentration | Vector-Borne Diseases Digestive Diseases, All-cause mortality, Mental & Behavioral Disease |

A review of the literature reveals that mental health, maternal & child health, and nutritional issues are the most neglected areas in climate health research. The study conducted by

(Berrang-Ford et al., 2021) deduced that in Asia and Europe, the impact of particles on air quality has received a lot of attention. In North America, where hurricanes were the most often discussed topic, extreme events were one of the top three hazards. Major hazards in Europe and Oceania at the time included heat waves. In comparison to other locations, literature from Africa and Latin America reported on rainfall and meteorological variability more frequently than other threats.

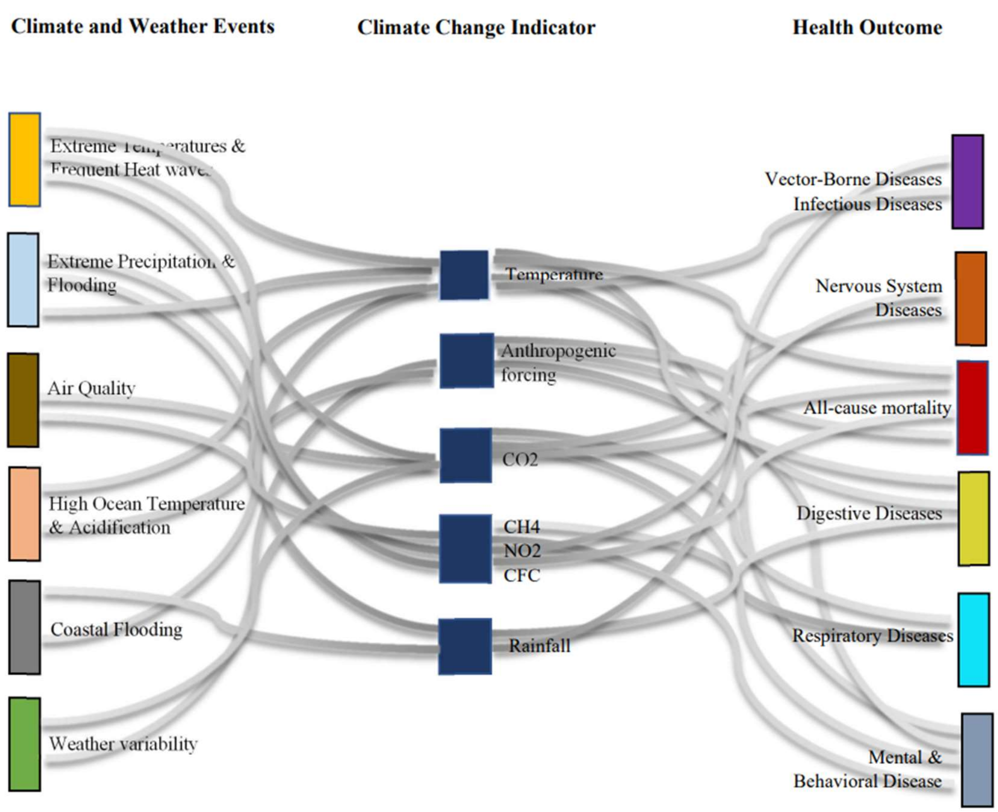
The research of (Berrang-Ford et al., 2021) also discovered that air quality, all-cause mortality, infectious diseases, and heat stress are the most common health-related issues. There are reports of a variety of health outcomes, with a strong emphasis on respiratory effects, particularly air pollution. In Asia, respiratory health was the most talked about health issue, while heat stress was among the three most talked about health issues in Europe, North America, and Oceania. The majority of the literature, especially in Asia and Europe, focuses on CCVW predictors of all-cause mortality. In addition to cholera, dengue, influenza, leptospirosis, and malaria, other common disease-specific subjects include dengue, which is the top health issue in Latin America and the second most popular topic in Asia.

Literature shows that climate health research is predominantly limited to a few geographical regions of high-income countries, whereas evidence from low-income countries which are hit hard by health issues due to climate changes, is minimal. (Berrang-Ford et al., 2021) reported that 79% of the 15,914 climate and health-related studies are focused on high- and upper-middle-income countries, mainly China. Published works on climate change and health reveal a significant income gradient. While the equivalence of several studies both for upper middle-income countries and high-income countries is misleading; a high number of publications from China explain well the under-representation of research from central Asia.

Studies in low-income areas show that infectious diseases predominate, with additional significant focuses on food and nutrition, and maternal and child health (Checkley et al., 2004; Singh et al., 2001) With lower income status, there is a clear gradient of greater emphasis placed on infectious diseases, food and nutrition, water, sanitation, and hygiene, as well as maternal and child health. This relationship is mirrored by a gradient of greater emphasis placed on chronic diseases, respiratory health, and demand on health systems. Recent research by (Pizzulli et al., 2021) has established a close correlation between climate change and human health at the global level. Some of the various diseases that are prevalent worldwide have spread more rapidly and shown deteriorating symptoms, which may be related to climate change. The World Health Organization estimates that environmental factors account for 23% of global mortality (Confalonieri, 2007).

WHO04 is the only empirical analysis of the effects of diarrhea-related global warming and is one of the finest studies on health (Kolstad & Johansson, 2011). The WHO04 analysis concluded that a 1°C increase in temperature was linked to a 5 percent increase in diarrhea and declared it to be a rough estimate based on empirical data from Fiji and Peru (Singh et al., 2001). These relationships and causal pathways are summarized in figure 2.8 below.

**Figure 2.8:** Causal linkages diagram




Globally, the growing incidence of diseases and mortality rates is triggered by climate change. Climate variability affects human life directly through changing weather patterns such as temperature extremes, precipitation, sea-level rise, and more frequent extreme events or indirectly by changes in water, air, and food quality as well as ecosystems, agriculture, industry, settlements, and the economy. Currently, the effects are minimal but are anticipated to gradually grow across the globe (Gallopín, 2006).

The myriad effects of weather variations on human health range from warnings of excessive heat and violent storms to linkages that might be less direct. Higher concentrations of Carbon Dioxide in the upper stratosphere are causing serious harm to the environment as a result of the rise in temperature (Telesca et al., 2018).

Weather and environment have an impact on the survival and behavior of disease-causing insects. Climate change also impacts the water and food quality of the affected environment which can have an impact on human health (Checkley et al., 2004). Additionally, the risks associated with human exposure to the consequences of climate change on mental health and well-being are significant (Hayes et al., 2018).

## 2.4. Conceptual Framework

WHO states that around 4 million people die every year from climate change-related or caused deaths, which roughly make up 23% of the deaths worldwide. Climate change directly causes and aggravates the diffusion and worsening of the symptoms of the diseases. Furthermore, climate change significantly contributes to the global burden of diseases and premature deaths (Pizzulli et al., 2021). Right now, at this stage in time, the effects of climate change are small but the scientific community and research project the scale of effects to drastically increase manifolds, at an unprecedented rate in the future worldwide.

Human lives and health are exposed to climatic change effects directly through variations in weather patterns that included rainfall fluctuations, rising temperatures, sea-level rise, and increase in the frequency of extreme events; and indirectly through changes in ecosystems that include changes in agriculture, industries, manufacturing, settlements; economy, water, air, and food quality.

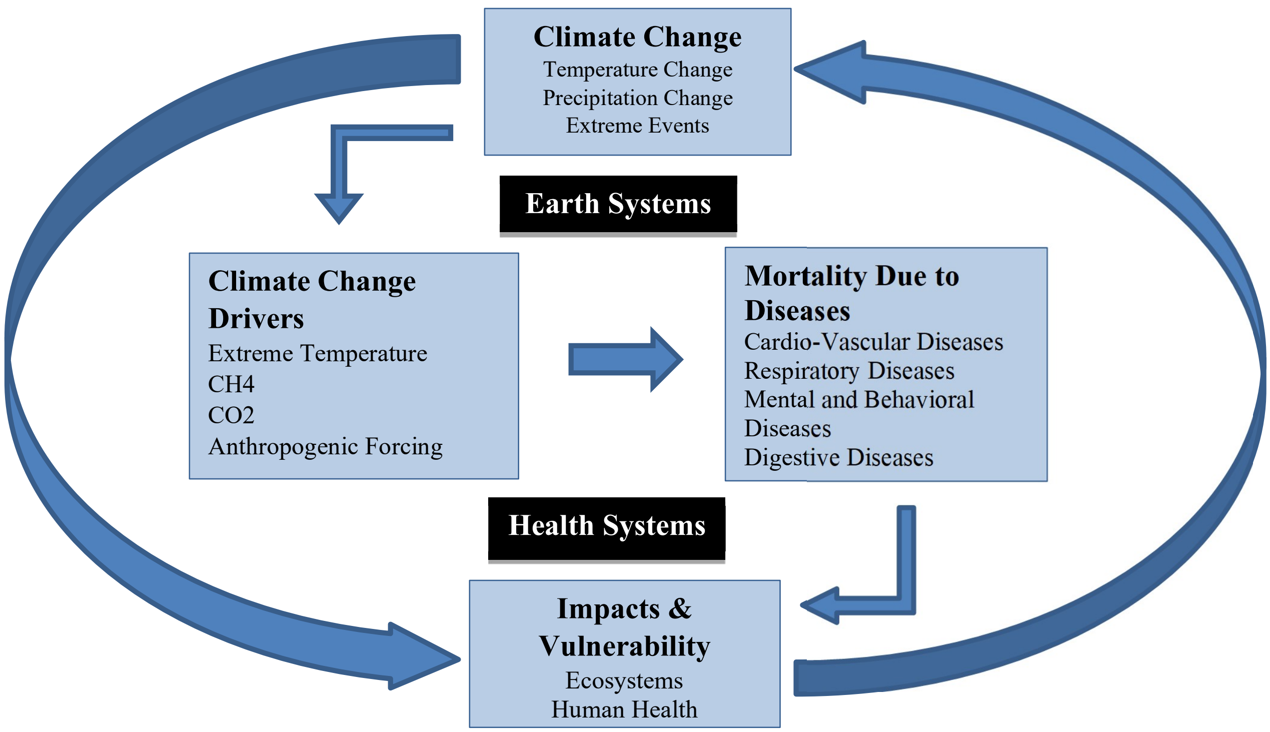
The scientific rationale used by (Pizzulli et al., 2021) to explore climate change-induced mortalities is the approach of understanding individual exposure pathways that can result in human disease is an effective method for learning how climate change affects health. Health impacts from climate change can be understood via the lens of exposure pathways, a term borrowed from the field of chemical risk assessment. Humans may react differently to various exposure routes, depending on factors including environmental context and exposure duration. Single or numerous climatic shifts, as well as geographic location, are possible risk factors for humans.

Climate change adversely impacts human health through causal channels that are either direct such as weather variability inducing human health deterioration through mutation in biological and chemical pathways resulting in diseases and indirect channels like ecosystem-induced food chain disruptions which result in a change in biodiversity in affected biomes (Kunze et al., 2022). One such example of direct causation is the recent spread of disease-transmitting vector carriers like dengue. Other indirect channels that cause human health deterioration involve socioeconomic factors such as adverse income distribution, population explosion, and regional conflicts. The conceptual framework extracted from the A-theoretical and evidence-based approach is presented below in Figure 2.2.

The climate change indicators including rising temperature, increasing concentration of methane (CH4), carbon dioxide (CO2), and manmade pollution that is anthropogenic forcing are significantly endangering human life, health, and the environment. The selection of climate change indicators in this research is based on scientific research by (Berrang-Ford et al., 2021; Pizzulli et al., 2021; Scheelbeek et al., 2021) was utilized given their significance in climate change and as well as human health domains.

The scientific rationale for using mortality as a proxy for human health is demonstrated by (Berrang-Ford et al., 2021; Pizzulli et al., 2021) in their studies. Mortality in number and rate is one traditional measure to gauge burden and to compare the impact of diseases in the domains of medical sciences, public health, health economics, and demography. Furthermore, separate studies conducted by (Melillo et al., 2014; Schwartz et al., 2015) explore the summer season which is April through September, and forecasted to see an increase in mortality as a result of global warming. Whereas in the winter season that is, October through March, studies predicted to see a drop in deaths as a result of global warming. Without any methodological modifications for potential future adaption, these findings maintain the human population at 2010 levels. Accordingly, projections for the twenty-first century indicate that the temperature-death linkages discovered for the most recent decade of evidence (1997–2006) will not change (Gosling et al., 2009; Kalkstein & Greene, 1997). These studies provide evidence that climate change-induced mortality, and further, explain the usage of mortality as the proxy for human health.

**Figure 2.9:** Framework of Climate Change Drivers and Their Impacts on Human Health




*Source: Authors own diagram, extracted from the literature review*

# CHAPTER 3

# Data and Methodology

## 3.1. Machine Learning - An Introduction

Machine learning (ML) is the combination of a set of algorithms for modelling, interpreting, and understanding complex data sets. Machine learning allows computers to learn without being explicitly programmed. Machine learning is generally understood to refer to a computer's ability to mimic human intelligence. To program computers through experience, machine learning adopts a hands-off method.

Machine learning isn't new; World War II Enigma Machine is one of the earliest forms of machine intelligence. It's only been a few years, but the concept of automating and the use of sophisticated mathematical operations on massive data is gaining a lot of traction.

ML algorithms learn without being explicitly programmed in the same way that humans do from experience, or more accurately, data. These algorithms learn, grow, adapt, and develop from fresh data. In other words, the goal of machine learning is to enable computers to make inferences without being explicitly instructed to do so. Rather, they rely on iteratively-learning algorithms to achieve this goal.

ML includes many algorithms which are further classified into two broader sets of algorithms that are supervised learning and unsupervised learning. For supervised machine learning to work, the models need to be trained on labelled data sets so that they can gradually improve their accuracy. The most widely used form of machine learning currently is supervised learning. Unsupervised machine learning involves a computer system's attempt to discern patterns in data that have not been labelled. Patterns and tendencies that aren't being actively sought out by humans can be discovered via unsupervised machine learning.

Majorly supervised learning is used when the objective of the research is to find the association or prediction in the data set while unsupervised learning is used when the objective of the research is to investigate the description of the data which could encompass dimensional reductions and clustering etc. Supervised learning leading algorithms include namely: LASSO regression, random forests, and neural networks, while unsupervised leading algorithms include namely: clustering and PCA (principal component analysis), etc.

The following table summarises different ML algorithms according to their type, usage, and advantages in health sciences (health data science).

Table 3.1: Machine Learning Algorithms

| **Algorithms** | **Type** | **Usage** | **Advantages** |
|---|---|---|---|
| Lasso | Supervised Learning | Association/ Prediction | Automatic covariate selection. Simple interpretability. |
| Neural Network | | Association/ Prediction | Many predictors & non-linear relations can be accommodated. Better prediction performance. |
| Random Forests | | Association/ Prediction | Identification of heterogeneous treatment & effects |
| Hierarchical Clustering / K-means | Unsupervised Learning | Dimensional reduction/ Clustering | Free of hypothesis. High-dimensional data could be mapped to lower dimensions. |
| Propensity Score Matching | | Counterfactual/ Casual Inference | Simple Interpretability |

*Source: Extracted from the literature review. Author's computed diagram himself.*

**Machine learning core functionality**

Machine learning begins with feeding the chosen algorithm training data. Algorithms require training data, which can be either known or unknown information that is labelled or unlabelled data. After training the algorithm, the testing phase starts in which new data is used as input and fed to the algorithm. Both forecasts and results of trained and tested phases are compared. If prediction and outcomes don't correspond, the algorithm is retrained until the objective is satisfied. This allows the machine learning system to continue to learn and produce the ideal answer, improving over time.

**Machine learning versus traditional approaches in Climate and Health domains**

Growing amounts of health data that includes clinical, pharmacological, and genetic data are handled by practitioners. There is a growing belief that "big data" could answer all the medical and scientific concerns and will convert healthcare into precision healthcare. Still, statistics by themselves don't tell much of the story or underlying causes and issues. The algorithms that encode domain (e.g., medical and biological) knowledge and causal reasoning are the ones that make a difference. The advent of data science particularly the large amount of data that gave birth to health data science in recent years provides an opportunity to re-evaluate this data-driven perspective.

Most traditional approaches used in biological sciences and medicine, for instance, can only function in data-driven predictive modes, which makes them incapable of realizing precision-based goals such as timely assessment and matching of the patients with the most appropriate preventative and therapeutic measures. Data interpretation and the implementation of breakthroughs into health care require a deeper familiarity with machine learning and its processes.

Integrating machine learning algorithms with healthcare domain expertise, and further coupling these two with causal reasoning is crucial to bringing about a qualitative transformation in medicine that will lead to better patient outcomes as "big data" becomes an increasingly common tool for analysing quantitative information.

Finding a link between climate change and human health can be done using several different traditional approaches, but the results will vary depending on the specifics of the analysis and the quantity of data at hand. This thesis utilizes a machine learning based framework for analysing the relationship between climate change indicators and mortality on a national scale, and for projecting what might happen in the future using the best-performing models.

## 3.2. Methods of Analysis

This study employs neural networks, machine learning tools, and causal discovery for the understanding, investigation, and exploration of the climate change and mortality nexus.

### 3.2.1. Machine Learning:

Machine learning is a multi-level neural network that will be employed to improve the performance and to check the validity of the results. To overcome the limitation of the neural networks machine learning techniques will be employed which include Linear regression models.

### 3.2.2. Causal discovery:

The goal of Casual Discovery is to conclude underlying causes from observational evidence. In particular, this study, casual discovery techniques will be utilized to explore and investigate the causal pathways through which climate change affects human health proxied by number of deaths due to leading diseases as measured by causes of death (frequency) in Pakistan. Causal feature selection and reconstructing interaction networks in observational multivariate time series is currently a very active area of research in many fields of science. There are two main reasons for this: increased access to extensive amounts of observational time series data in today's era of big data and research in fields where controlled experiments are impossible, unethical, or expensive such as climate, Earth systems or the human body. Correlation based studies on pairwise association networks cannot be interpreted causally. The goal of causal network reconstruction goes beyond inferring association and directionality between two time series; the objective of causal discovery is to distinguish direct from indirect dependencies and common drivers among multiple time series.

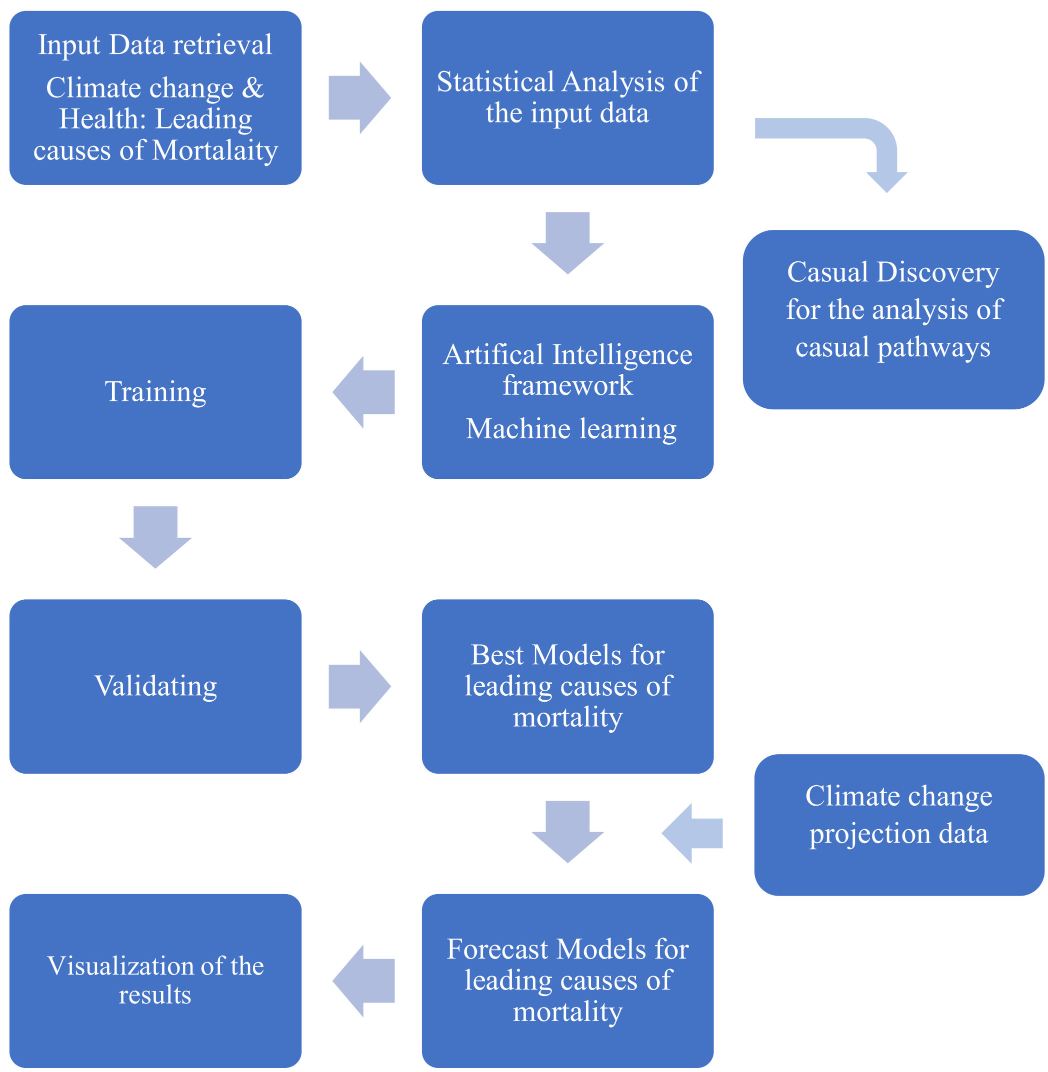
When working with both linear and nonlinear time series, the PCMCI (Runge et al., 2019) can be useful. The PC1 and MCI stages of the method are named after the two conditional independence tests they implement. The algorithm's PC1 phase uses the conditional independence method implemented in PC (skeleton phase) to identify potential dependencies between each variable at a given timestamp and all the other variables in all the previous timestamps. After accounting for auto-correlation and erroneous edge detections, the MCI (momentary conditional independence) test (Runge et al., 2019)is applied to evaluate causal links between variables across time points.

The need for reliable AI systems has prompted the development of causality techniques in machine learning (ML) studies in recent years. (Pearl, 2018) argues that to move above the constraints of present ML systems, causal reasoning is essential. The foundation of typical ML algorithms is the correlation between variables rather than sound causal structures, which can lead to inappropriate, biased, or even destructive inferences being drawn.

### 3.3. Methodological Framework

This methodological framework is adopted from the study of (Pizzulli et al., 2021). This framework employs single neural networks, machine learning tools, and causal discovery tools to explore, investigate and understand the climate change problem in the context of Pakistan. Furthermore, the mapping of results will aid in the visualization of the impactful evidence from observational data. For visualization GIS software, ArcGIS will be utilized. The usage of the causal discovery tools and ArcGIS is in addition to the existing methodological framework presented by (Pizzulli et al., 2021).

**Figure 3.1:** Methodological Framework




*Source: Authors own diagram*

## 3.4. Data Generating Process & Data Description

The following section explains the data generation process of the data sets.

### 3.4.1. Data Generating Process: CMIP6 & NASA-GISS-E2-1-H Model

Climate models play a vital role in scientists' quest to comprehend past and future climate changes. These models employ intricate simulations of the Earth's atmosphere, land, and oceans, necessitating powerful supercomputers to generate climate projections. By utilizing equations, climate models represent the underlying physics, chemistry, and biology that govern the Earth's climate system.

Given the complexity of the climate system and computational limitations, climate models divide the Earth into a grid of boxes or "grid cells." These models incorporate multiple layers spanning the atmosphere and oceans, considering factors such as temperature, air pressure, humidity, and wind speed within each cell. The spatial resolution of a model refers to the size of its grid cells, with global models typically using cells around 100km in longitude and latitude at mid-latitudes.

Climate model results yield massive amounts of data, encompassing numerous variables across space and time. These variables range from temperature and clouds to ocean salinity, generating petabytes of information. The models rely on external factors known as "forcings," which alter the amount of solar energy absorbed by the Earth or trapped by the atmosphere. Forcings include changes in solar output, long-lived greenhouse gases like CO2, CH4 (methane), N2O (nitrous oxide), halocarbons, as well as aerosols emitted through fossil fuel burning, forest fires, and volcanic eruptions. Aerosols impact incoming sunlight and influence cloud formation.

Climate models produce a comprehensive overview of the Earth's climate, providing insights into various variables across different timeframes. These outputs encompass atmospheric temperatures, humidity profiles from the surface to the upper stratosphere, as well as oceanic parameters such as temperature, salinity, and pH. Additionally, models estimate snowfall, rainfall, snow cover, glacier and ice sheet extents, along with wind speed, direction, and climate phenomena like the jet stream and ocean currents.

General Circulation Models (GCMs), also known as Global Climate Models, simulate the fundamental physics of the climate system. More recent iterations, called Earth System Models (ESMs), incorporate biogeochemical cycles and their interactions with the climate. ESMs account for the carbon and nitrogen cycles, atmospheric chemistry, ocean ecology, vegetation dynamics, and land use changes, all of which influence how the climate responds to human-induced greenhouse gas emissions. ESMs capture vegetation responses to temperature and rainfall, influencing the exchange of carbon and other greenhouse gases with the atmosphere.

#### 3.4.1.1. Coupled Model Inter-comparison Project (CMIP6)

CMIP (Coupled Model Inter-comparison Project) is a framework that enables scientists to systematically analyze, validate, and enhance General Circulation Models (GCMs) used to study the Earth's climate system. Under the CMIP umbrella, CMIP6 represents a significant global initiative that builds upon the previous CMIP5, offering improved modeling capabilities and a broader range of scenarios for comprehensive climate analysis. This note provides an overview of CMIP6, highlighting its purpose, key features, and contributions to climate science (Studies (NASA/GISS), 2018).

CMIP6 involves a substantial number of climate models contributed by various research institutions worldwide and comprises of 35 models from different modeling centers. These models encompass a diverse range of approaches, resolutions, and parameterizations, reflecting the global collaborative effort to simulate and understand the Earth's climate system.

In terms of experiments, CMIP6 encompasses a suite of coordinated simulations designed to address specific scientific questions and explore various aspects of climate variability and change. The exact number of experiments conducted in CMIP6 is not fixed and can vary depending on research goals and the interests of participating modeling groups. The experiments cover a wide range of climate-related topics, including historical simulations to reproduce past climate conditions, future projections under different emission scenarios, and specialized experiments targeting specific components of the climate system.

The primary objective of CMIP6 is to enhance our understanding of the complex interactions within the Earth's climate system and their potential impacts on regional and global climate patterns. By utilizing a diverse ensemble of climate models, CMIP6 provides robust projections that are crucial for informing policymakers, researchers, and stakeholders involved in climate change mitigation and adaptation strategies. These projections help establish future climate scenarios based on anticipated concentrations of greenhouse gases, aerosols, and other climate forcings, facilitating assessments of potential future climate conditions. Furthermore, CMIP6 serves as a vital framework for conducting coordinated climate model experiments, enabling the scientific community to gain valuable insights into the Earth's climate system. Through improved models, a range of experiments, and enhanced projections, CMIP6 contributes to our understanding of climate variability and change, supporting informed decision-making in the face of global environmental challenges.

#### 3.4.1.2.NASA-GISS-E2-1-H Model

The NASA-GISS-E2-1-H model, developed by the NASA Goddard Institute for Space Studies (GISS), is a prominent participant in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Studies (NASA/GISS), 2018). As an Earth system model, it simulates the complex interactions between the atmosphere, oceans, land surface, and cryosphere. With its high spatial resolution, advanced parameterizations, and integration of observational data, the model enhances our understanding of the Earth's climate system. It generates climate projections, analyzes emission scenarios, and incorporates observational data to improve its representation of the current climate state. The NASA-GISS-E2-1-H model contributes significantly to climate science by providing valuable insights into regional climate patterns, climate sensitivity, and the impacts of greenhouse gas emissions on global and regional climate. Its outputs inform policymakers and researchers in formulating effective strategies for climate change mitigation and adaptation.

#### 3.4.1.3 Integrated Assessment Models

Traditionally IAMs (Integrated Assessment Models) were extensively used in economic analyses to develop and evaluate climate change policies. These models combine simplified representations of climate and the economy to simulate the global economic impacts of climate change under diverse mitigation scenarios. While simpler than comprehensive global climate models (GCMs), IAMs like the Dynamic Integrated model of Climate and the Economy (DICE), The Climate Framework for Uncertainty, Negotiation and Distribution (FUND), and Policy Analysis for the Greenhouse Effect (PAGE) are frequently used by policymakers to weigh the costs and benefits of climate action. IAMs offer a simplified framework, assisting in finding optimal policies (DICE) or evaluating various parameter values in climate policies (FUND and PAGE).

These models integrate assumptions and simplifications about the climate system, and demographic, political, and economic variables. To effectively shape climate change policies, IAMS must disclose these underlying assumptions and inputs for experts to gauge their reliability. Equations within IAMs should be explicit to ensure transparency and allow users to comprehend the mechanisms behind model projections. This level of transparency, added through the incorporation of the IAMs in the IPCC working avoids the perception of IAMs as "black boxes" and empowers experts to assess the realism of model projections. Furthermore, the Coupled Model Inter-comparison Project aims to enhance the performance and reliability of global coupled ocean-atmosphere models, crucial in predicting future climate changes under various emissions scenarios.

### 3.4.2. Data Generating Process: IHME Causes of Death dataset

The Institute for Health Metrics and Evaluation (IHME) causes of death dataset is a comprehensive and significant compilation of data that offers valuable insights into the global burden of diseases and the underlying causes of mortality. The process of generating this dataset involves a series of meticulous steps aimed at ensuring the accuracy, consistency, and representativeness of the data.

Data collection for the IHME causes of death dataset entails sourcing information from a variety of reliable and diverse sources, including vital registration systems, censuses, surveys, and health records from numerous countries. These sources provide essential data on deaths, causes of death, population demographics, and other pertinent factors, forming the foundation for robust analysis.

To ensure comparability and coherence across different sources and countries, IHME undertakes a process of data harmonization. This involves employing standardized methods and definitions to reconcile discrepancies in coding systems, disease classifications, and data formats. By doing so, IHME creates a unified and consistent dataset that facilitates meaningful cross-country and cross-source comparisons.

Statistical analysis plays a crucial role in estimating cause-specific mortality rates and trends within the IHME causes of death dataset. IHME employs advanced statistical techniques and models that account for various factors, including age, sex, geography, and time. These models enable the generation of reliable and robust estimates, enhancing our understanding of cause-specific mortality patterns.

Given the inherent uncertainty associated with mortality data, IHME recognizes the need for comprehensive uncertainty assessment. Statistical methods, such as uncertainty intervals and validation techniques, are employed to quantify the precision and reliability of the estimates within the dataset. This transparent approach enables users to interpret the data while considering the associated uncertainties.

Continual updates are a fundamental aspect of the IHME causes of death dataset. IHME regularly incorporates new data sources, refines methodologies, and improves the accuracy of estimates, ensuring that the dataset remains current and reflective of the latest understanding of global mortality patterns. This iterative process allows for ongoing refinement and evolution of the dataset, enhancing its relevance and reliability over time.

Lastly, the IHME causes of death dataset is the result of a rigorous and comprehensive data generation process encompassing data collection, harmonization, statistical analysis, uncertainty assessment, data visualization, and continuous updates. This meticulous approach ensures that the dataset offers reliable and valuable information on causes of death, making it a vital resource for informing public health policies, resource allocation decisions, and the monitoring of global health trends.

### 3.5. Data Sources

Only one source of the data is chosen for climate change NASA GISS E2-1-H (Studies (NASA/GISS), 2018) and another for causes of deaths taken from IHME. The data set is chosen for its reliability and veracity, as well as the reliability of the source. Climate change data will be selected based on a literature review, identifying the main drivers of climate change which cause the most significant impact on human physical health globally and in Pakistan. The NASA GIIS database is a collection of a variety of climate data, the particular data set which is taken for this study is of a forced model of climate change worldwide.

#### 3.5.1. List of Climate change Factors

The following is the list of the climate change drivers/factors.

**i. Surface Temperature (Tas):**

Surface temperature (Tas) is a fundamental variable used in CMIP6 experiments to assess climate change. It represents the average temperature of the Earth's surface at 2 meters above ground level. Climate models simulate Tas by considering physical processes such as radiative transfer, energy balance, and heat exchange between the atmosphere and the Earth's surface. Projections of Tas under different emission scenarios provide valuable insights into global and regional temperature patterns and help evaluate the magnitude and spatial distribution of climate warming.

**ii. Atmospheric Carbon Dioxide (CO2):**

The concentration of atmospheric carbon dioxide ($CO_2$) is a crucial initial factor of climate change. In CMIP6, $CO_2$ is a key forcing variable that represents the radiative effects of increased greenhouse gas concentrations resulting from human activities. Climate models simulate historical and future scenarios of $CO_2$ emissions to understand its impact on the Earth's climate system. These models consider interactions between the atmosphere, land surface, and oceans, accounting for emissions from fossil fuel combustion, land-use changes, and natural processes. Accurate representation of $CO_2$ is essential for predicting future climate trajectories.

**iii. Atmospheric Methane ($CH_4$):**

Atmospheric methane ($CH_4$) is another significant greenhouse gas that influences climate change. CMIP6 experiments simulate $CH_4$ concentrations to understand its contribution to radiative forcing and climate variability. $CH_4$ emissions arise from natural sources such as microbial activity in wetlands, as well as human activities including agriculture and fossil fuel production. Climate models in CMIP6 incorporate complex interactions between the atmosphere, biosphere, and anthropogenic sources to capture the temporal and spatial variability of $CH_4$ concentrations. Understanding the dynamics of $CH_4$ is crucial for comprehensive climate modeling.

**iv. Anthropogenic Forcings:**

Anthropogenic forcings encompass a range of human-induced factors that impact the climate system. These forcings include greenhouse gas emissions, aerosol concentrations, land-use changes, and other human activities that alter the energy balance of the Earth's atmosphere. In CMIP6 experiments, models simulate anthropogenic forcings to assess their role in climate change. Historical and future scenarios of these forcings are considered to quantify their impacts on surface temperature, precipitation patterns, and other climate variables. Accurate representation of anthropogenic forcings is vital for understanding the drivers of climate change and developing effective mitigation strategies.

### 3.5.2 Rationale for the variable selection

The choice of climate change drivers is based on the review of the climate change existing literature (as discussed in detail in chapter 2). IPCC AR-6 highlights the climate change drivers which will largely contribute to the existing climate change and future climate change. Temperature, $CH_4$, $CO_2$ and Anthropogenic forcing (combine for all the forcing) are major drivers.

Table 3.2: A list of the variables is given below:

| **Climate Change** | **Causes of Deaths** |
|---|---|
| Temperature | Digestive Diseases |
| $CH_4$ | Nervous, Mental & Behavioral Disease |
| $CO_2$ | Cardiovascular Diseases |
| Anthropogenic Forcing | Respiratory Diseases |

### 3.6. Descriptions of Data

Statistics for the data taken as input for the period 1980–2015 was taken from the IHME Causes of Death dataset and NASA–GISS–E2-1–H (numbers based on annual fatalities for Pakistan). Table 3.3 presents descriptive statistics for the input data from NASA and WHO, covering the period from 1980 to 2015. These statistical indicators provide insights into the characteristics of various variables related to climate change and their association with global deaths.

Temperature Anomaly (°C): The mean temperature anomaly over this period was 0.034°C, indicating a slight increase in global temperatures. The standard deviation (stdev) of 0.341°C suggests considerable variability. The minimum and maximum values (-0.482°C and 1.015°C) represent the range of temperature anomalies observed.

$CO_2$ (ppm): The mean atmospheric carbon dioxide ($CO_2$) concentration was 318.58 parts per million (ppm). The stdev of 32.18 ppm reflects variations in $CO_2$ levels. The minimum and maximum values (285.20 ppm and 410.40 ppm) indicate the range of $CO_2$ concentrations recorded.

$CH_4$ (ppb): The mean atmospheric methane ($CH_4$) concentration was 1.17 parts per billion (ppb), with a stdev of 0.35 ppb. The minimum and maximum values (0.79 ppb and 1.88 ppb) represent the range of $CH_4$ concentrations.

Anthropogenic Forcing ($W\ m^{-2}$): The mean anthropogenic forcing, a measure of the impact of human activities on climate change, was 0.941 Watts per square meter ($W\ m^{-2}$). The stdev of 0.926 $W\ m^{-2}$ suggests variability in anthropogenic influences. The minimum and maximum values (0.000 $W\ m^{-2}$ and 3.421 $W\ m^{-2}$) reflect the range of anthropogenic forcing.

Victims Without External Causes: This variable represents the number of yearly deaths from various causes other than external factors. The statistics provide information on the mean (13,446,963.9), stdev (3,677,068.5), minimum (1,393,261.0), maximum (17,006,389.0), and quartiles (25%, 50%, 75%) of these deaths.

Mental and Behavioural Disorder, Respiratory Diseases, Nervous, Mental & Behavioral Disease, Digestive Diseases: These variables represent the number of yearly deaths related to specific disease categories. The statistics include the mean, stdev, minimum, maximum, and quartiles (25%, 50%, 75%) for each disease category.

Overall, these descriptive statistics provide a comprehensive summary of the input data, allowing for a better understanding of the characteristics and variability of the variables related to climate change and their association with global deaths.

**Table 3.3**: Description of the Climate change drivers and leading causes of death given below:

| | Temperature Anomaly (°C) | $CO_2$ (ppm) | $CH_4$ (ppb) | Anthropogenic Forcing ($W\ m^{-2}$) | Victims Without External Causes | Nervous, Disorder | Respiratory illnesses | Cardiovascular illnesses | Digestive illnesses |
|---|---|---|---|---|---|---|---|---|---|
| **mean** | 0.0341 | 308.58 | 1.17 | 0.936 | 44,569.9 | 102,454.5 | 357,399.7 | 281,664.2 | 1,404,489.2 |
| **stdev** | 0.331 | 31.181 | 0.352 | 0.927 | 6,706.5 | 15,76.1 | 24,668.9 | 18,370.5 | 17557.3 |
| **min** | −0.476 | 285.20 | 0.793 | 0.000 | 12,326.0 | 28,639.0 | 85,959.0 | 41,662.0 | 955,573.0 |
| **25%** | −0.201 | 295.00 | 0.86 | 0.245 | 49,049.8 | 9,366.5 | 730,557.0 | 142,181.3 | 515,161.0 |
| **50%** | −0.067 | 309.50 | 1.05 | 0.625 | 48,335.5 | 54,460.5 | 548,770.0 | 224,029.0 | 644,557.0 |
| **75%** | 0.209 | 332.20 | 1.48 | 1.417 | 13,782.0 | 22,202.0 | 276,771.5 | 411,736.8 | 787,771.5 |
| **max** | 1.015 | 410.40 | 1.88 | 3.421 | 70,638.0 | 33,349.0 | 610,627.0 | 618,126.0 | 850,382.0 |

In this chapter, we have examined the descriptive statistics for the input data from the IHME Causes of Death dataset and NASA-GISS-E2-1-H model, covering the period from 1980 to 2015(Climate change data) and from 2000 to 2015 (IHME data set). These statistics have provided valuable insights into the characteristics and variability of the variables related to climate change and their association with deaths in Pakistan.

The analysis of temperature anomaly revealed a slight increase in global temperatures, with a mean of 0.034°C and a standard deviation of 0.341°C, indicating considerable variability. The atmospheric carbon dioxide ($CO_2$) concentration showed a mean of 318.58 parts per million (ppm) and a standard deviation of 32.18 ppm, with a range from 285.20 ppm to 410.40 ppm. The mean atmospheric methane ($CH_4$) concentration was 1.17 parts per billion (ppb), with a standard deviation of 0.35 ppb and a range from 0.79 ppb to 1.88 ppb. The mean anthropogenic forcing, a measure of human activities' impact on climate change, was 0.941 Watts per square meter (W $m^{-2}$), with a standard deviation of 0.926 W $m^{-2}$ and a range from 0.000 W $m^{-2}$ to 3.421 W $m^{-2}$. The statistics for victims without external causes provided insights into the number of yearly deaths from various causes, with a mean of 44596.9, a standard deviation of 6706.5.

Furthermore, the author examined the statistics for specific disease categories, including nervous, mental and behavioral disorders, respiratory diseases, cardiovascular system diseases, and digestive diseases. These statistics encompassed the mean, standard deviation, minimum, maximum, and quartiles (25%, 50%, 75%) for each category, shedding light on the variability and characteristics of global deaths associated with these diseases.

Overall, these descriptive statistics have provided a comprehensive summary of the input data, enhancing our understanding of the variables related to climate change and their association with deaths. The variability observed in the statistics emphasizes the complexity and dynamic nature of these variables. The findings from this chapter lay the foundation for further analysis and investigation into the relationships between climate change and human health outcomes, enabling more informed decision-making and interventions to mitigate the impacts of climate change on global health. Next chapter will forecast climate change induced deaths.

# CHAPTER 4

# RESULTS

## 4.1. Overview

This chapter presents the results of machine learning by presenting the correlations between climate change and human health on a global scale. It begins by mentioning previous studies that have established connections between climate change and human health, specifically highlighting the impact on respiratory diseases and nervous system disorders. The use of machine learning is introduced as the methodology employed in this study to analyze the correlations. The results obtained through these techniques are described providing valuable insights into the influential factors in the analysis.

The topic then turns to the precise connections between changes in the climate and other diseases that have been discovered. It should be highlighted that no statistically significant association between climate change and digestive issues was found, indicating the necessity for additional variables to evaluate this issue. Since the link between climate change and mental and neurological system disorders is believed to be questionable, further research is required to confirm it. The association between rising temperatures and respiratory disorders, on the other hand, is thought to be the strongest, with heat playing a substantial role, in line with earlier findings.

The study also includes forecasts for future scenarios, indicating an anticipated increase in the number of deaths related to respiratory and mental diseases due to climate change over the next decade.

## 4.2. Analysis with Machine Learning

The Regression Learner program in Python was utilized for analysis purposes. Through this program, ML models were trained to make predictions based on the provided data. The initial analysis involved datasets containing four input values and one output value for each variable under investigation.

Each variable, namely the number of deaths attributed solely to diseases (excluding external causes), digestive disease, mental and behavioral disorders, nervous system disease, and respiratory disease, was examined independently. Their values were then correlated with the values of four drivers of the climate change: $CO_2$, $CH_4$, temperature, and anthropogenic forcing.

To identify the best solution, the model's configuration was adjusted and retrained for each scenario, systematically eliminating one cause at a time. This process continued until the optimal configuration was determined (Figure 4.2 and Figure 4.3 provide visual representations of this analysis).

To assess the model's performance, a residuals plot was utilized, as depicted in Figure 4.1. This plot visually represents the variance between the predicted and actual responses. Upon analyzing the data, it was observed that there is a strong correlation, as indicated Table 3 compares the four diseases being studied to climate change. It is crucial to remember that an upward trend in both the input and output data affects the association between climate change and digestive illnesses. Therefore, caution should be exercised when interpreting this correlation due to the substantial amount of data involved.

**Figure 4.1.** Residual plot for number of deaths predicted with two features, using Exponential Gaussian Process.

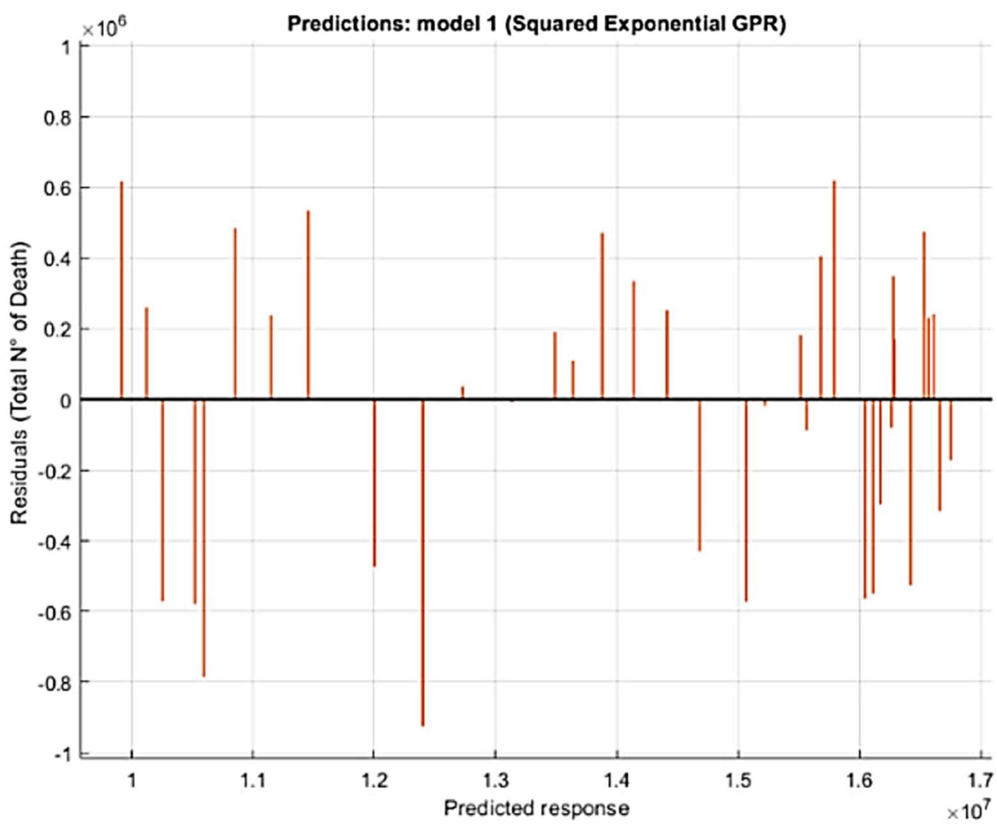


**Figure 4.2** Response plot for Cardiovascular deaths, predicted with three features, using Exponential Gaussian Process

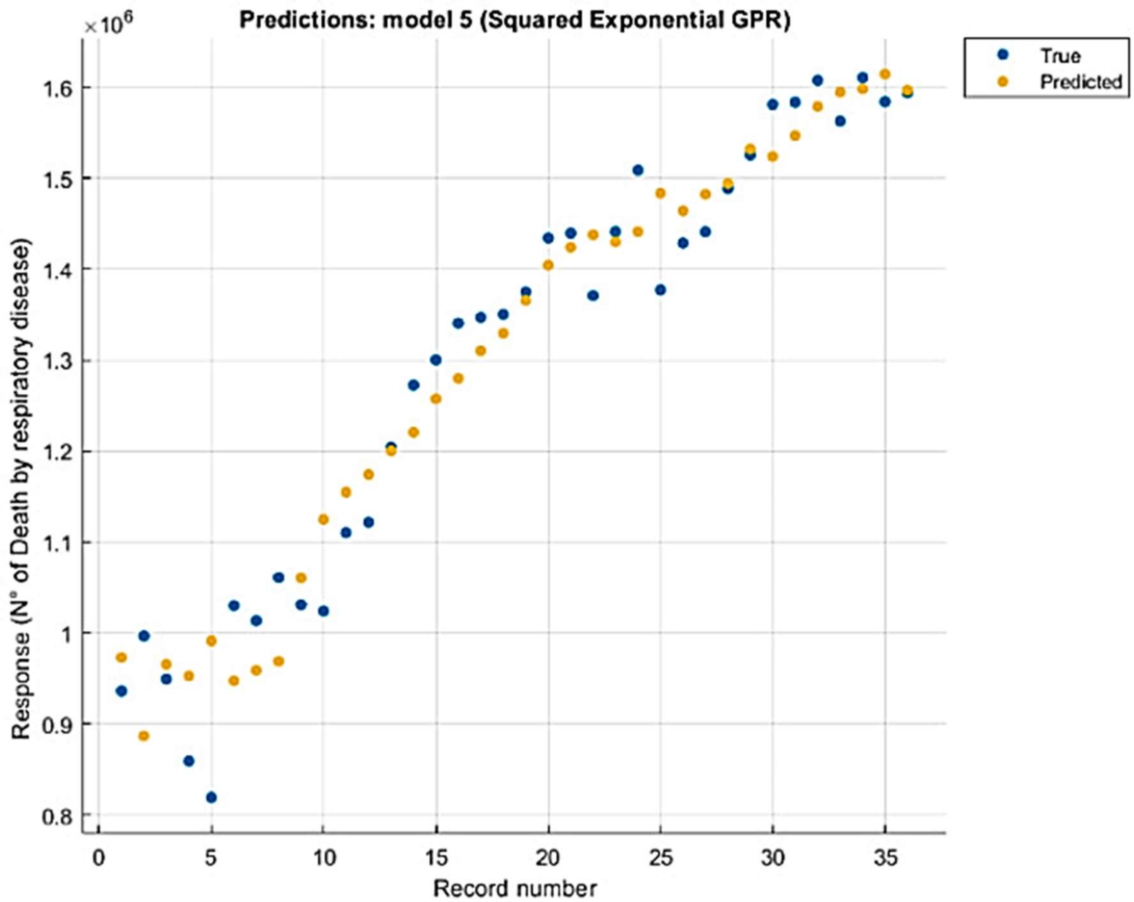


**Figure 4.3** Response plot for Mortality by respiratory diseases Forecasted with three features, using Squared Exponential Gaussian Process

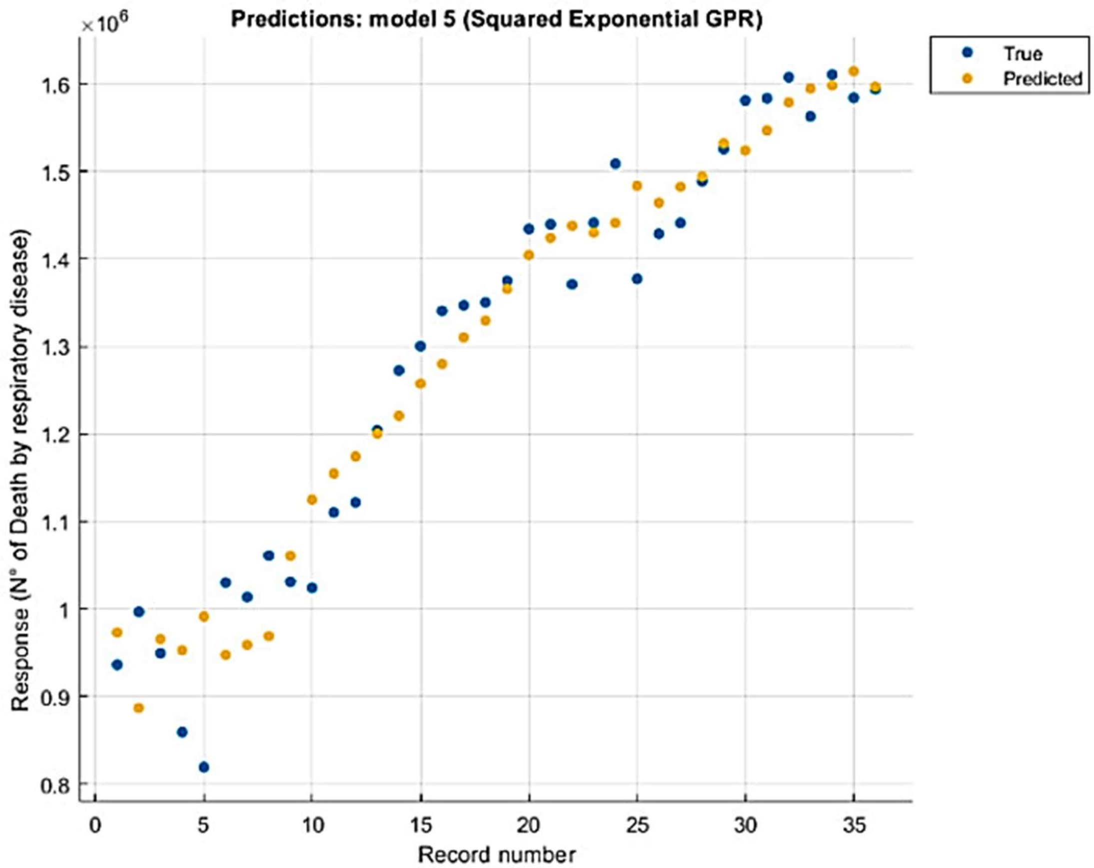



Table 4.1:Results from the analysis with the Regression Learning model

| Analyzed Variable | Regression Model | $R^2$ | Best Correlation |
|---|---|---|---|
| **Total number of Deaths without accidents** | Gaussian Squared Exponential Process | 0.96 | $CO_2$, $CH_4$ |
| **Digestive disease** | Gaussian Exponential Process | 0.98 | $CO_2$, $CH_4$, temperature |

| | | | |
|---|---|---|---|
| **Nervous system, Mental and behavioral disease** | Gaussian Exponential Process | 0.99 | $CO_2$, $CH_4$, anthropogenic forcing |
| **Cardiovascular disease** | Gaussian Exponential Process | 1.00 | $CO_2$, $CH_4$, anthropogenic forcing, temperature |
| **Respiratory disease** | Gaussian Squared Exponential Process | 0.94 | $CO_2$, $CH_4$, temperature |

### 4.3. Analysis with Causal Discovery

PCMCI is a constraint-based algorithm, Independence tests are utilized by constraint-based algorithms to establish a set of edge constraints for the graph based on observational data. The $G_2$ test, proposed by (Runge et al., 2019) is an example of such a test. Subsequently, rules are applied to determine the direction of the relationships that are discovered. However, in certain instances, the rule phase may be bypassed to generate undirected graphs. These graphs typically represent local relationships, focusing solely on the relationships of individual nodes in an undirected manner.

PCMCI is a causal discovery framework for time series datasets which are large, and it copes with both linear and non-linear timeseries. PCMCI algorithm is a powerful tool for extracting causal relationships from time-series data. It is gaining popularity when dealing with multivariate time series where variables interact dynamically. PCMCI extends the concept of conditional independence testing to time series, allowing for the identification of both direct and indirect causal connections.

The causal dependencies are presented as graphical time series. PCMCI algorithm is applied in two phases, each denotes a different conditional independence test, in the first phase it employs PC1 and in the second MCI. The PC1 relies on conditional independence strategy that is applied by PC (skeleton phase i.e., figuring out dependencies and independencies) – which is a constraint-based causal discover framework, based on the assumption that dependencies abide by the dseparation criterion - in order to reveal the dependencies that are between each variable in a specific time period and among all the variables in all the previous time periods $A_t \coprod B_{t-1} CI, A_t \coprod B_{t-2} IC$. Momentary conditional independence is applied in the next phase which by considering autocorrelation and incorrect edge detection defines dependencies among variables in different time periods.

**Table 4.2:** Evaluation Metrics

Metrics used for time series causal discovery methods are presented in the table below.

| Metric | Description |
|---|---|
| **Accuracy** | $True\ Positive + True\ Negative \Big/ True\ Positive + False\ Positive + False\ Ne$ |
| **Mean/Median error** | The disparities between the predicted values and the ground truth are assessed using various measures such as variances of mean and median, including root mean squared and squared differences |
| **Longest Common Sequence** | The length of the longest sequence of events in a time-series model is quantified |
| **Edit distance with real penalty** | The transformation of one series into another is evaluated by quantifying the number of changes, taking into account a penalty determined by the user. |
| **Euclidean Distance** | The distance between each step of the series is computed by the formula of difference between coordinate points on a plane. |
| **Dynamic time warping** | The distance between two sequences is calculated by measuring the euclidean distance between each point in the sequences, where the sequences consist of sets of time points. |

The PCMCI is initiated by assessing momentary conditional independence between pairs of variables. It then constructs a causal pathway graph as shown in figure 4.4 that captures potential causal relationships. In causal pathway graph, the relationships are represented through arrows connecting nodes, where each node corresponds to a variable. Solid arrows denote direct causal connections, and dashed arrows indicate indirect paths.

Direct relationships reflect immediate causal influences. Analysis shows the direct relationships are the links from Temperature (T) to Cardiovascular Disease Deaths (CVD) and from Temperature (T) to respiratory disease deaths (RDD). These suggest that changes in temperature can have direct effects on these health outcomes. Indirect relationships reveal more intricate connections, as the link from Cardiovascular Disease Deaths (CVD) to Nervous Mental and Behavioral Disease Deaths (NMBD) indirectly suggests that temperature changes affecting cardiovascular health could influence nervous and behavioral health. lag values to links accounts for time delays between cause and effect. These values are domain-specific and might reflect the time it takes for an impact to manifest. In our graph, lag values are provided to represent these temporal delays. A lag of -1 between Temperature Anomaly (T) and Cardiovascular Disease Deaths (CVD) indicates that temperature changes precede cardiovascular events by one time step.

1. **Temperature Anomaly (T) → Cardiovascular Disease Deaths (CVD, Lag: -1):** An increase in temperature anomalies at time t-1 is associated with an increase in cardiovascular disease deaths at time t. This indicates a one-time lag between temperature anomalies and their impact on cardiovascular health. It aligns with the understanding that extreme temperatures could trigger cardiovascular events with a delay.

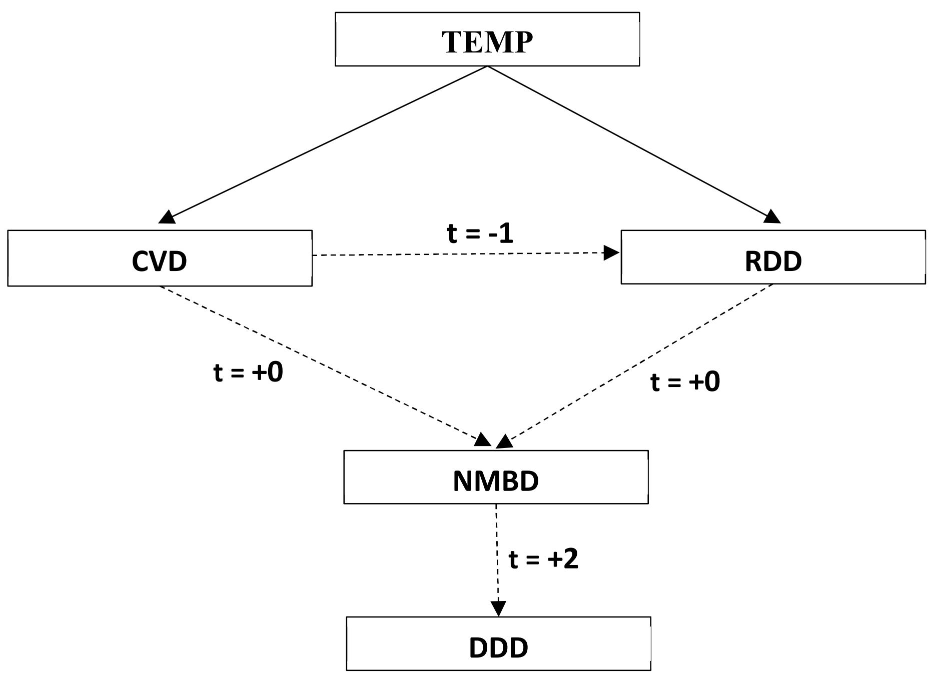
2. **Temperature Anomaly (T) → Respiratory Disease Deaths (RDD, Lag: 0):** Changes in temperature anomalies at time t have an immediate effect on respiratory disease deaths at the same time t. This implies that respiratory health is sensitive to current temperature fluctuations. Short-term variations in temperature could directly impact respiratory health outcomes.

3. **Cardiovascular Disease Deaths (CVD) → Nervous Mental and Behavioral Disease Deaths (NMBD, Lag: 0):** Changes in cardiovascular disease deaths at time t are associated with immediate changes in nervous, mental, and behavioral disease deaths at the same time t. This suggests that cardiovascular health might directly influence nervous, mental, and behavioral health outcomes, possibly due to shared physiological or psychological mechanisms.

4. **Respiratory Disease Deaths (RDD) → Nervous Mental and Behavioral Disease Deaths (NMBD, Lag: 0):** Variations in respiratory disease deaths at time t are linked with immediate variations in nervous, mental, and behavioral disease deaths at the same time. This imply that respiratory health issues might have direct psychological and behavioral effects, impacting mental and behavioral health outcomes without a delay.

5. **Nervous Mental and Behavioral Disease Deaths (NMBD) → Digestive Disease Deaths (DD, Lag: 2):** Changes in nervous, mental, and behavioral disease deaths at time t have an impact on digestive disease deaths two time steps later, at time t+2. This suggests a delayed influence of mental and behavioral health on digestive health. Mental stress or behavioral changes might gradually contribute to digestive health issues over time.

**Figure 4.4:** Causal Discovery analysis for the temporal lag connections




### 4.4. Assumptions with regard to Future Trends

Planning healthcare effectively requires an understanding of how various environmental, climatic, and developmental scenarios may affect population health. Additionally, predicting future health trends is crucial in the context of climate change since population well-being is a crucial component of adaptive capacity. A prediction function built on a previously acquired model was used to investigate these aspects. The task entailed entering fresh data for subsequent years.

The study by Miller et al. (2014) that forecasted increases in the concentration of a number of gases in the atmosphere, which are strongly linked to climate change, served as the basis for the future projection data utilised in this analysis, which were received from NASA. The previous analysis's models could only be developed with input data up to 2016. Data from 2017 onwards were used for forecasts beyond that time frame, with real data from 2017 to 2019 collected from aerial scans and analysed in the same way as previously described, yielding a global yearly average. Reference has been made to the abovementioned research article (Miller et al., 2014) for data after 2019 (Miller et al., 2014).

The CMIP6 model, which incorporates data obtained from satellites pertaining to these variables and produces a model for predicting future based on the premise that human activity patterns will not change, was used to make the prediction for the four variables of concern (CO2, CH4, heat, and anthropogenic forcing). The estimates were generated using current regulations aimed at reducing pollutants and employing more renewable energy sources.

It is plausible to anticipate a commensurate rise in future scenarios given that most future trends point to an increase in climate-related variables. The same process as before was used for integrating the latest input data, which incorporates future trends. The sequence of the data had to match that of the model analysis. A 'trainedModel' structure was used to generate predictions based on the fresh data once the algorithm was exported to the workspace of the Regression Learner. A prediction function and a model object are both present in this structure.

Out of the 5 models developed using machine learning, only two were used in simulating future trends. These models exhibited the lowest error (RMSE) and were specifically associated with victims of nervous, mental & behavioral disease. The performance of these models was nearly perfect due to the consistent upward trend observed in the input and output data.

While the forecast was extended until the year 2500, it is important to note that such distant predictions, being the result of a mathematical model, lack reliability. As a result, forecasts for the four climate variables previously studied will be created for 10-year intervals using the forecasting data as demonstrated in Figure 4.5.

An assessment of the future for the next 10 years is shown in Figure 4.6. Although extending the projection to 500 years is technically feasible, such a long prediction timescale lacks reliability and is rife with uncertainty.

**Figure 4.5:** Mortality due to Cardio-Vascular Disease CVD (actual and forecasted)

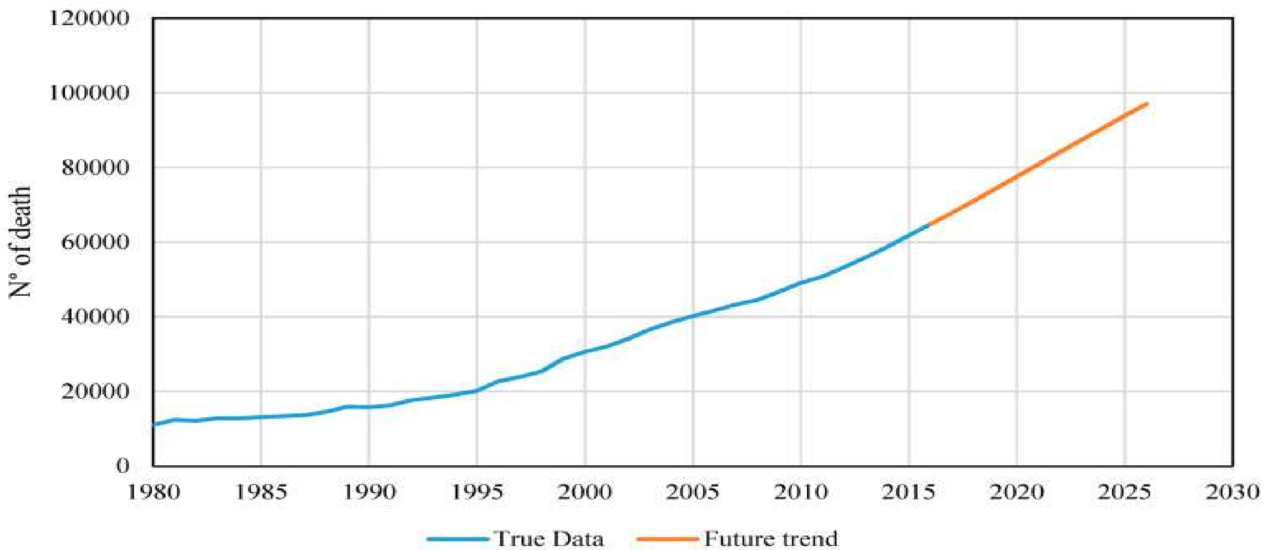
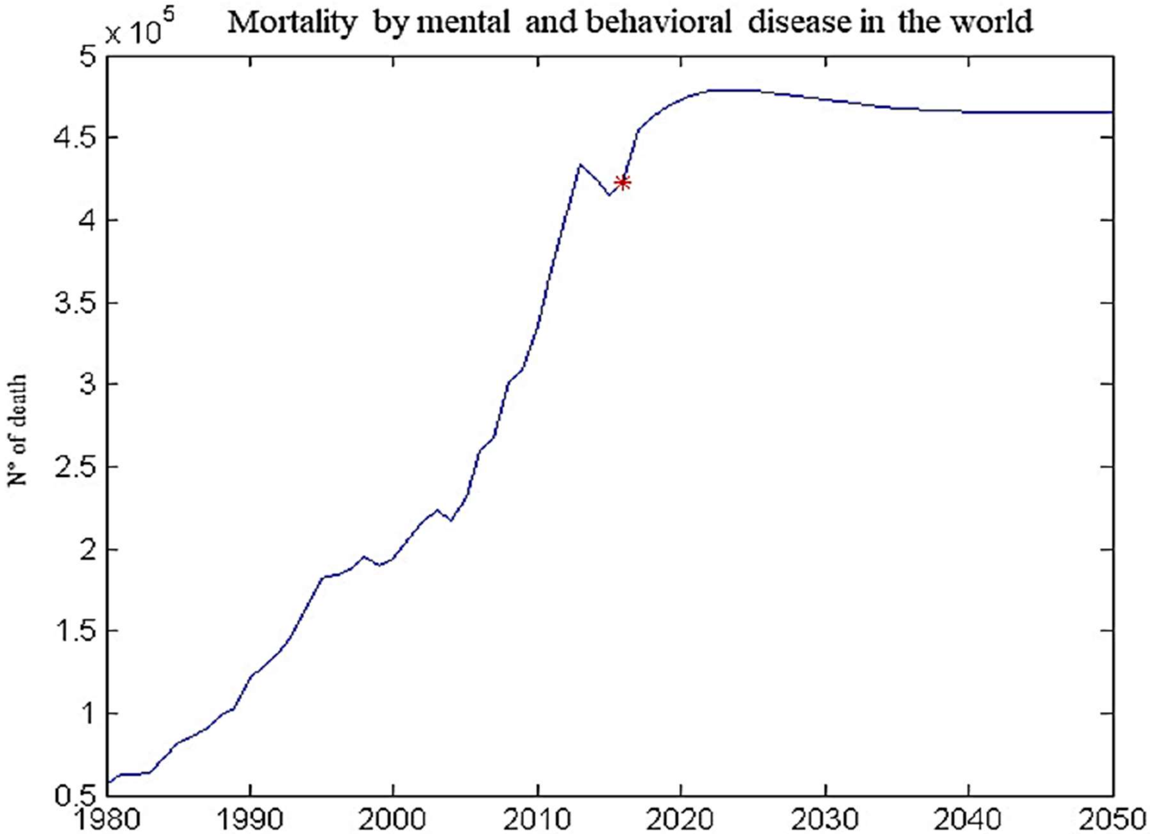


**Figure 4.6** Mortality Forecast (by Nervous, Mental & Behavioral illnesses)




The projected number of victims attributed to nervous system disease exhibits a notable increase. Over a span of 35 years, the predicted rise amounts to nearly 5000 additional deaths per year, while a 10-year period anticipates a surge of 3000 deaths. The escalating casualties resulting from climate change each year are expected to yield a more pronounced escalation in the forthcoming years compared to the past.

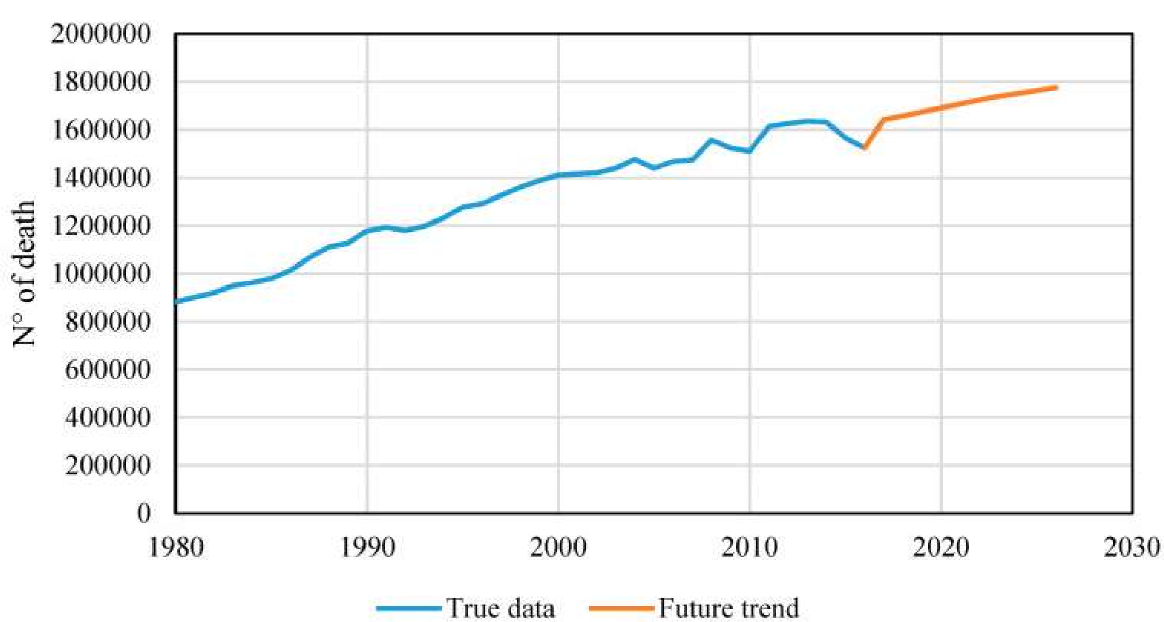
To generate the forecast, the most effective model was employed, despite the correlation being driven by the consistent upward trends observed in both the input and output curves. Consequently, the decision was made to employ neural networks for creating the forecast, aiming to address certain challenges encountered during regression using machine learning techniques. By utilizing forecast data pertaining to Temperature, $CO_2$, $CH_4$, and anthropogenic forces, simulations were conducted using the pre-trained neural network. Forecasts were successfully generated for all the variables under investigation, with Figure 4.6 specifically displaying the forecast for the nervous mental and behavioral diseases variable.

Figure 4.6 illustrates prediction based on machine learning models, revealing an upward trajectory. However, when extending the time frame under consideration, a decrease in the number of deaths becomes apparent. The fall in mortality suggests that the machine learning models have yielded somewhat better results in the case of this model.

Similar to this, a 10-year-long scenario founded on the model for patients with respiratory disorders was created. Figure 4.7 depicts a probable future scenario that is similar to the previous 10-year timeframe but differs noticeably in that the growth in casualties is less noticeable and displays an almost linear curve that is in accordance with the actual data curve. Despite not using a linear approach, the roughly linear trend indicates that the temperature variable's fluctuating trend, which is important to this model, may potentially have an impact on the future course.

As previously stated, this model holds the highest significance and aligns best with the forthcoming climate data, despite having a lower R2 index compared to other trained models. Respiratory disease-related fatalities exhibit a strong correlation with climate change, specifically with fluctuations in temperature. This includes temperature variations such as heatwaves or frost, as well as the consistent yearly rise in the Earth's average temperature.

**Figure 4.7:** Deaths due to respiratory illness (actual trend and projections)



### 4.5. Discussion:

The findings from this analysis provide strong evidence of a global correlation between climate change and human health, which aligns with previous studies  (Song et al., 2017). By applying machine learning techniques, a more comprehensive investigation yielded detailed insights into the specific environment-related threats closely associated with the diseases examined. Nervous, Mental & Behavioral Disease share a close relationship with respiratory system diseases because of same pathological grounds. Hence it becomes mandatory to treat them collectively instead of separately. Our analysis revealed a robust correlation between mental pathologies and climate change, with these diseases exhibiting comparatively lowest relative error among all the diseases we studied.

From a mathematical perspective, linkage between environmental change and respiratory illnesses is comparatively less accurate despite a 6% error. However, we observed a strong association in temperature and respiratory illnesses, independent of other climate change causes. This finding confirms a linkage between environmental change and digestive illnesses, as the trends in temperature align closely with the observed patterns in these diseases.

Our utilization of artificial intelligence has provided new insights into linkage between environmental change and health. We employed ML techniques to better understand the input variables that have the highly conspicuous influence on the research. This application of causal discovery has further reinforced the previously established correlation and enhanced the reliability of our findings for three out of the four diseases studied. The inclusion of temporal lags aligns with econometric theory, which acknowledges that causal effects might not be immediate. The lag values provide insights into the time it takes for causal impacts to propagate through the system, which is crucial for understanding policy effects or intervention planning. This interpretation helps to differentiate true causal relationships from mere correlations by considering the dynamics of cause and effect.

It is important to acknowledge that the resulting forecasts may not possess a high level of reliability. However, when compared to scenarios derived from other scientific articles, our forecasts exhibit striking similarities. The limitations of our study, including limited data availability and the multitude of factors required to explain such an intricate phenomenon, contribute to the margin of error observed. While we utilized global data to facilitate our calculations, it is worth considering that conducting analyses on specific countries and focusing on prevalent mortality types within those countries could potentially yield more accurate results. The challenges associated with obtaining comprehensive data on human health present significant obstacles to studying the correlations between health and climate change.

### 4.6. Conclusion

This study aimed to explore the association between environmental change and health at a national scale using available time series data, employing machine learning techniques for analysis. Just as research by Solomon et al. (2007) has identified associations among climate related variables and indicators of human health on a continental scale, specifically indicating the relevance of such correlations for respiratory diseases and nervous system disorders (Zhou, 2013)

The application of machine learning techniques in studying these correlations has yielded novel findings. Two distinct approaches of AI are utilized to assess and identify the strongest correlations. This study has yielded excellent results, providing deeper insights into the influential input variables in this research. The novel approaches of machine learning (ML) have affirmed the correlations that were previously identified, enhancing the viability of three out of the four studied pathologies. The analysis conducted with AI technologies has outperformed purely statistical approaches, particularly in terms of generating models with lower absolute error.

There was no connection between digestive disorders and death due to climate change as analyzed through the ML regression analysis and causal discovery analysis. Instead of solely attributing this variable to the primary drivers of changing the climate, the analysis of this variable ought to focus on trigger variables relating to the quality, availability, and other aspects of water and food. There is no reliable link between the effects of climate change and fatalities from nervous system and mental diseases. The data's almost linear and predictable pattern shows that further investigation is required to establish its authenticity.

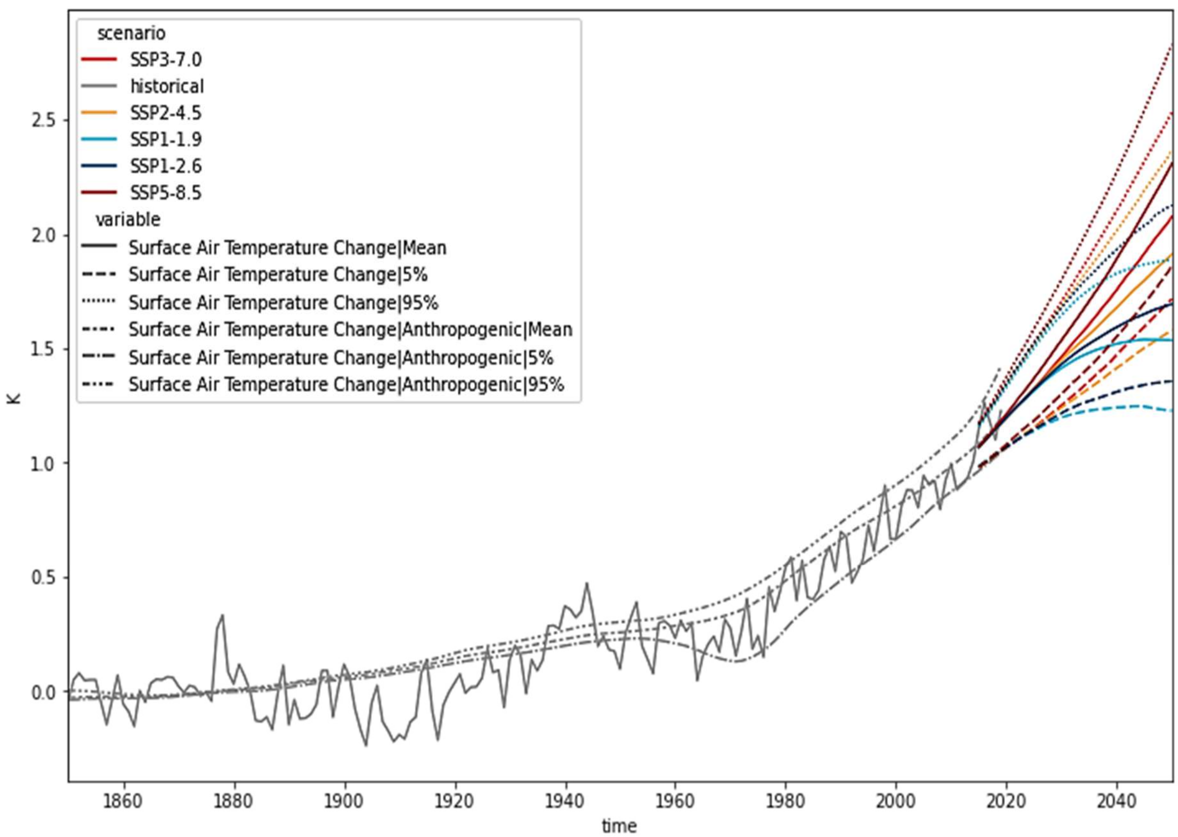
On the other hand, the association between climate change and fatalities from respiratory disorders has been shown to be the strongest (Zhou, 2013). with heat playing a crucial part in this relationship. Based on the study, a possible future scenario was predicted for both of the illnesses evaluated, indicating an increase in fatalities from respiratory and mental illnesses as a result of changing climate during the following 10 years (Guo et al., 2016). This study's overall analysis, however less comprehensive than earlier analyses, highlights the possibility of discovering a link between changing climate and human health, even when looking at a wide geographic area. In verdict, it is undeniable that a link exists between climate change and human health, particularly considering the connection between temperature changes and respiratory disease-related fatalities. Nonetheless, obtaining more extensive data and narrowing the focus of analysis to a specific geographical area would lead to more precise and accurate results.

# CHAPTER 5

# VISUALIZATIONS

This chapter presents visualization regarding the historic and future climate change trends and patterns along with the mapping of GHGs including CO2 and non-CO2 drivers, impacting the climate over time. Moreover, the chapter provides simulated visualization on the future green-house gases emissions, global surface temperature anomalies and the forecast of the global surface temperature under different scenarios. Visualizations based on CMIP-6 model and IHME causes of health datasets. Visualization were made using Python and others special visualization apps.

Global Surface Temperature Projection under different CMIP6 Shared Socioeconomic Pathways



[1]

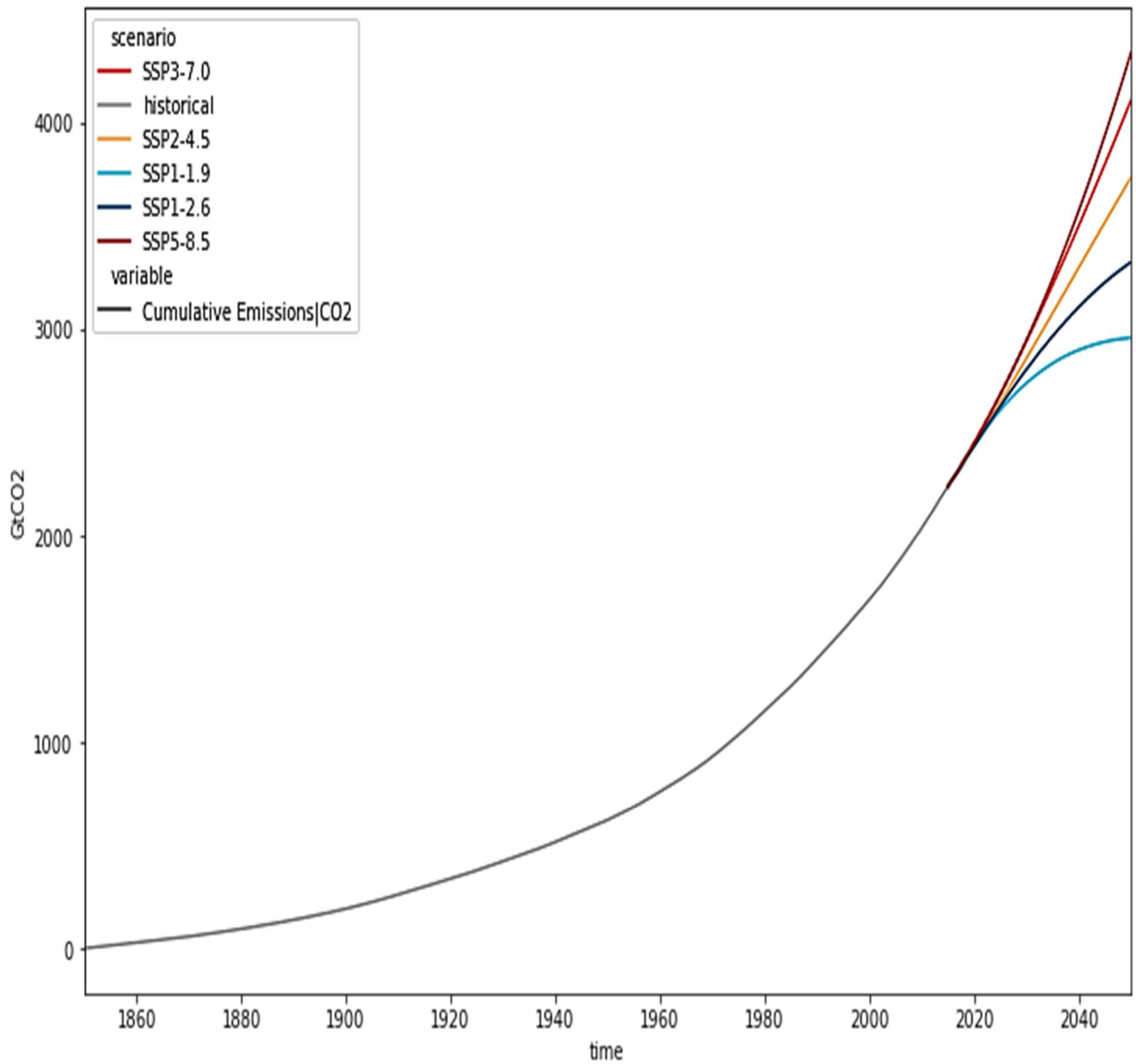
Figure 5.1: Surface temperature change projection over time under different scenarios

[1] This visualization employs python package “scmdata”, which is designed to handle simple climate model data but generalises to handle all sorts of timeseries-based data including data in the IAMC format alongside seaborn and native matplotlib as required.

Based on historical data, the increase in global surface temperature in °C between 1850 and 1900 can be correlated with cumulative CO2 emissions in GtCO2 from 1850 to 2019 which is shown by the black coloured line. The shaded range, along with its central line, provides estimates of historical human-caused surface warming. The coloured regions represent the projected range of global surface temperature with a high level of confidence, while the thick coloured lines depict the median estimate. These projections are based on cumulative CO2 emissions from 2020 to 2050 for different scenarios including SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. The expected global warming takes into account all human-induced factors and the cumulative CO2 emissions associated with each scenario. The analysis covers the period from 1850 to 2050, during which global CO2 emissions are consistently positive across all different scenarios.

Increase in global surface temperature (measured in C°) as a function of cumulative CO2 emissions given the base period of 1850-1900. Global surface temperature simulation under different shared socioeconomic pathways identified in CMIP6 model.

The red line represents the SSP 3-7.0 pathway while the yellow line represents the SSP2 -4.5 pathway. Red line represents the worst case scenario given by the SSP5-8.5 pathway. SSPs highlights the significant global warming as indicated by its various pathways such as SSP1-1.9 scenario (light blue line), targeting a 1.5°C limit, predicts an average warming of 1.4°C across multiple models. Similarly, the SSP1-2.6 (dark blue line) scenario, aligned with the RCP2.6 of AR5, indicates an average warming of 2.0°C. Conversely, the SSP5-8.5 scenario suggests a higher average warming of 5.0°C, while the new SSP3-7.0 scenario forecasts 4.1°C of global warming.




[2]

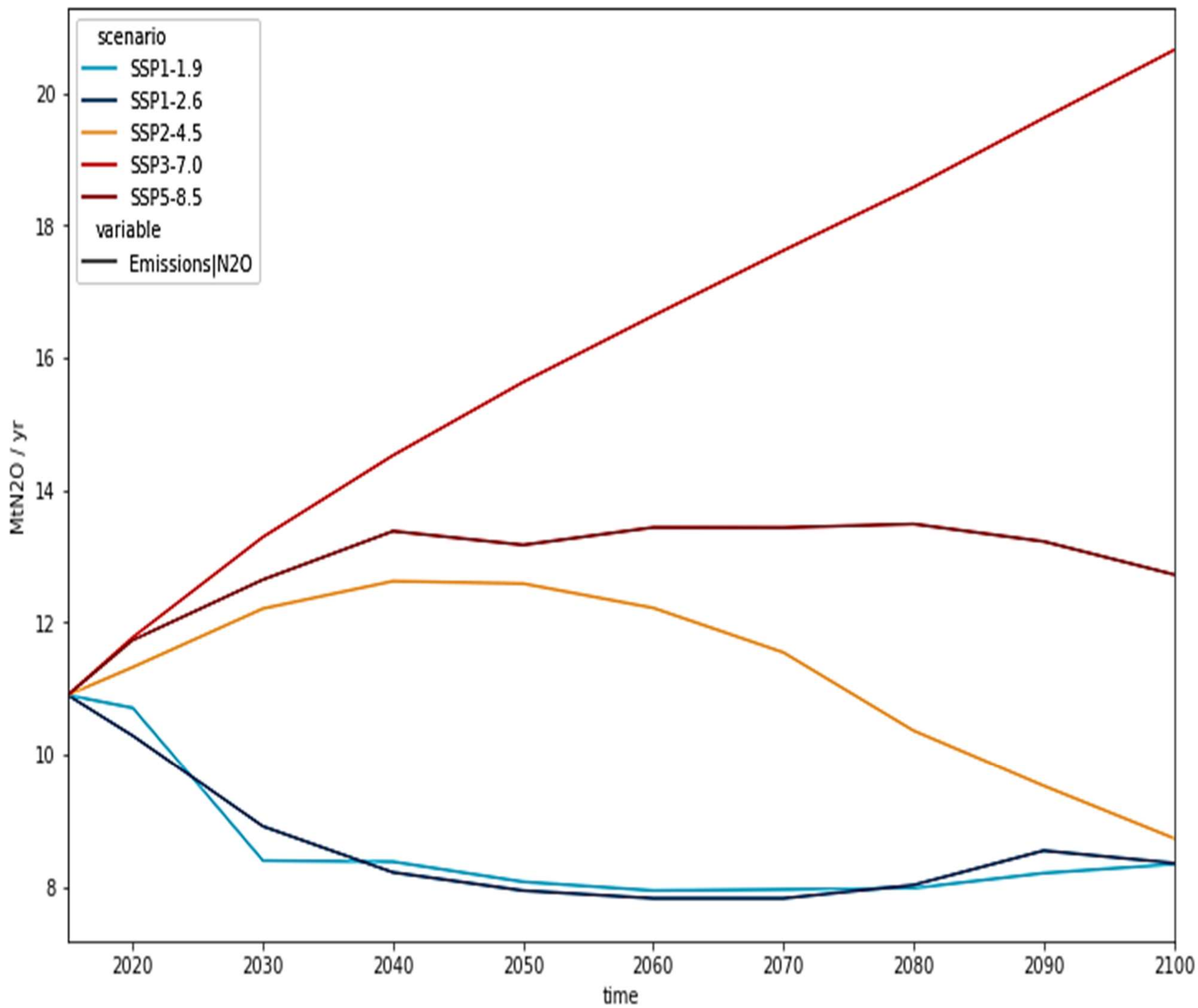
*Figure 5.2: Cumulative Carbon dioxide emission projection over time*

**5.1. Projections for GHGs:**

The following graphs shows the emissions trajectories of the anthropogenic drivers and their projections under different climate change scenarios accordingly to IPCC AR6. From the graphs

[2] This visualization employs python package "scmdata", which is designed to handle simple climate model data but generalises to handle all sorts of timeseries-based data including data in the IAMC format alongside seaborn and native matplotlib as required.

its evident that CO2 and other non-CO2 drivers, emissions will increase, resulting in the global temperature to rise leading to global warming.




3

*Figure 5.3:* N2O Emission projections

[3] This visualization uses the Python package pyam, which provides a suite of features and methods for the analysis, validation and visualization of reference data and scenario results generated by integrated assessment models, macro-energy tools and other frameworks in the domain of energy transition, climate change mitigation and sustainable development.

4

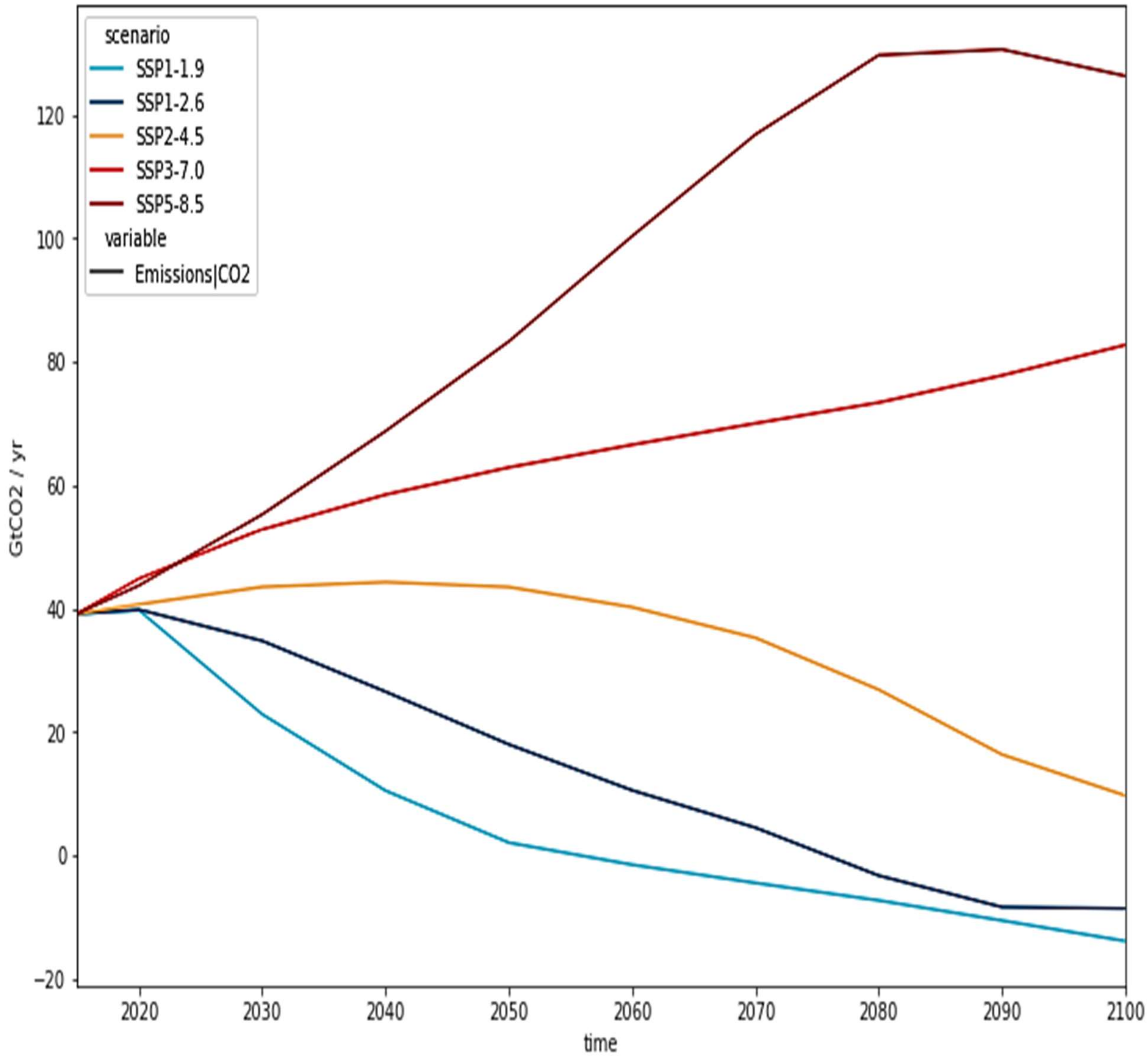
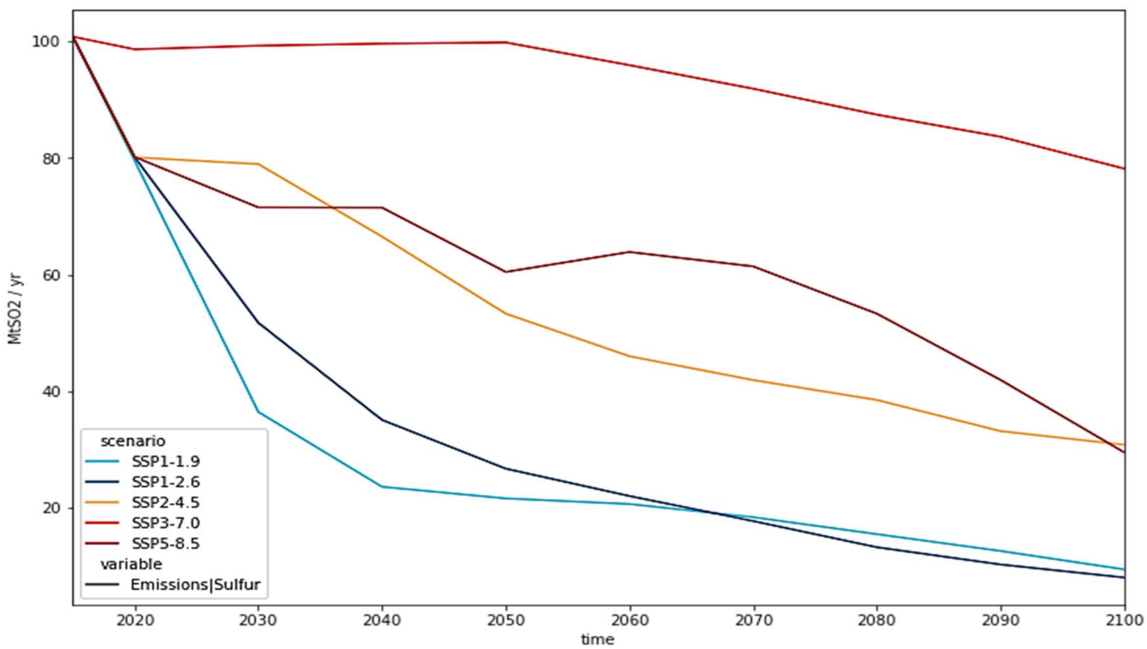



*Figure 5.4:* CO2 Emission projections

[4] This visualization uses the Python package pyam, which provides a suite of features and methods for the analysis, validation and visualization of reference data and scenario results generated by integrated assessment models, macro-energy tools and other frameworks in the domain of energy transition, climate change mitigation and sustainable development.



[5]

*Figure 5.5:* Sulfur Emissions

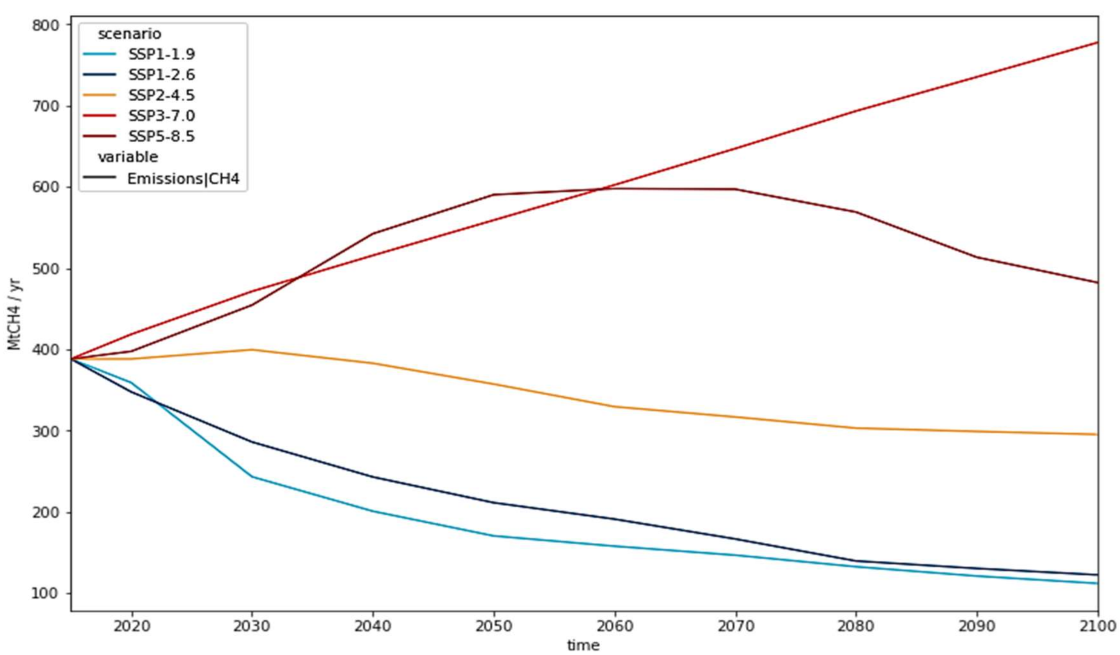



*Figure 5.6:* CH4 Emissions

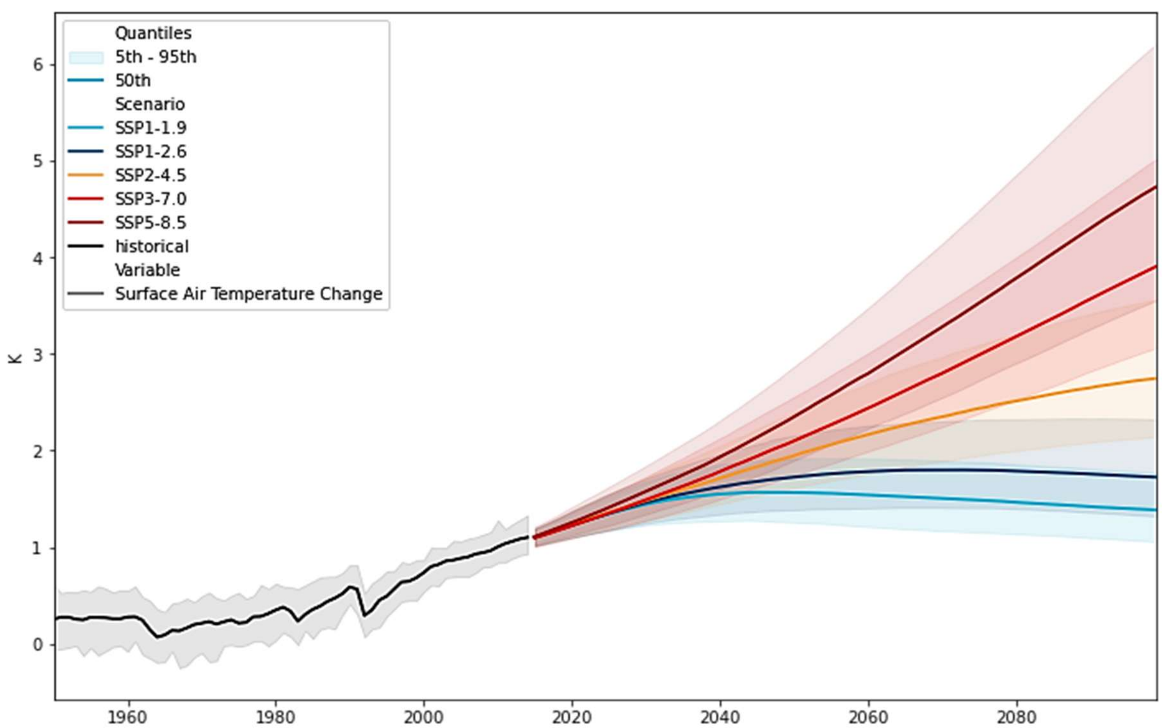
[5] This visualization uses the Python package pyam, which provides a suite of features and methods for the analysis, validation and visualization of reference data and scenario results generated by integrated assessment models, macro-energy tools and other frameworks in the domain of energy transition, climate change mitigation and sustainable development.

### 5.2. Global Temperature Anomaly:

The following maps and graphs show the global temperature anomaly for the future, for the possible future scenarios under different scenarios identified under CMIP6. Further, hindcasted maps show the global temperature anomaly for the past. Hindcasted maps validated the global temperature anomaly with the observed data. Hindcast maps are made using the data from the NASA-GISS-E2-H-1 historical stimulation experiment dataset. All the maps are generated using ArcGIS software and the graphs are made using python and other packages.

**Global surface temperature anomaly relative to 1850-1900 period under different scenarios**

Global surface temperature anomaly relative to 1850-1900 period under different scenarios under the CMIP share socioeconomic pathways. Following are the graphs for the surface air temperature change, presented with different scenarios and observed data.



[6]

*Figure 5.7:* Surface Air Temperature anomaly

[6] This visualization employs python package “scmdata”, which is designed to handle simple climate model data but generalises to handle all sorts of timeseries-based data including data in the IAMC format alongside seaborn and native matplotlib as required.

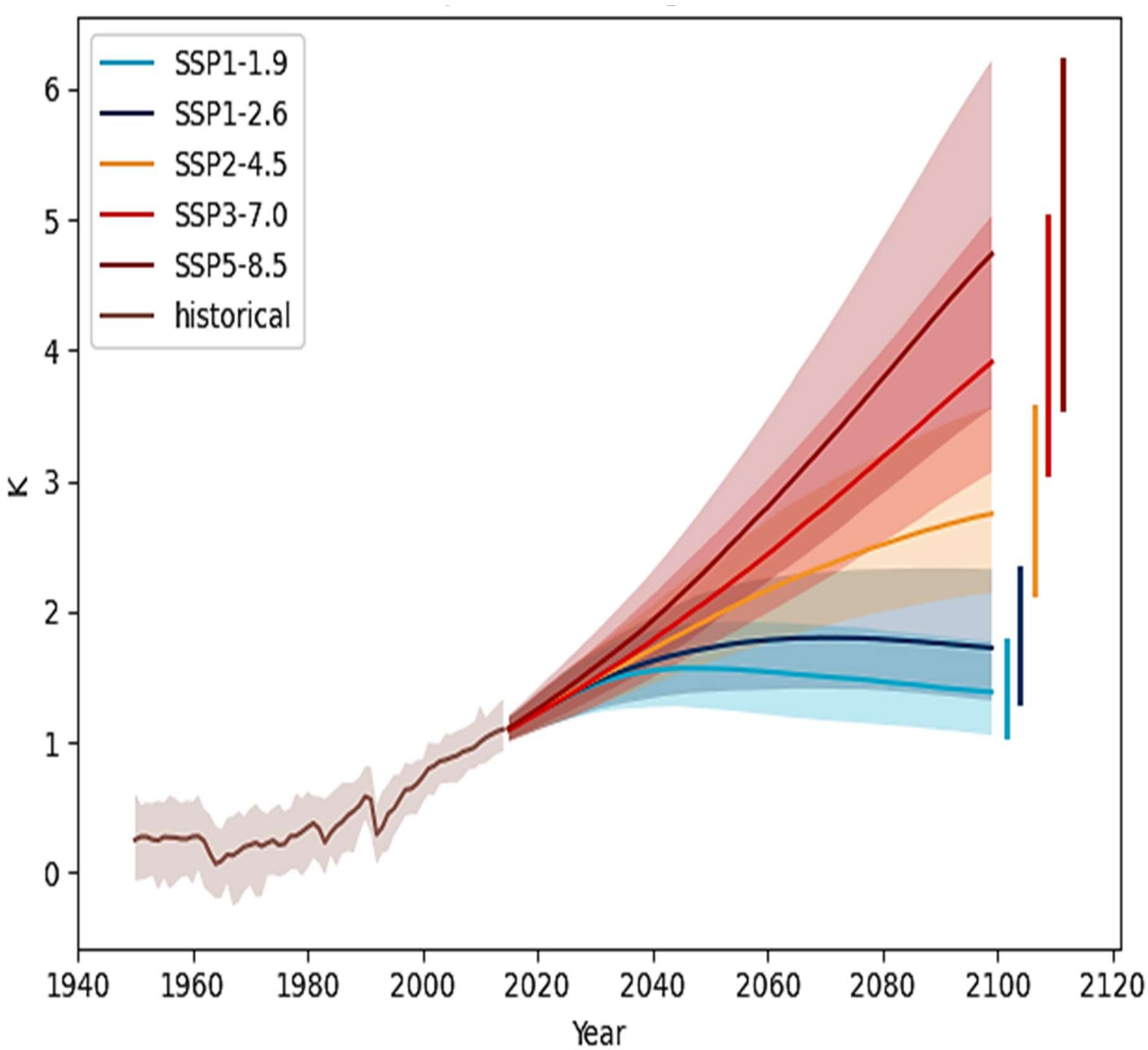
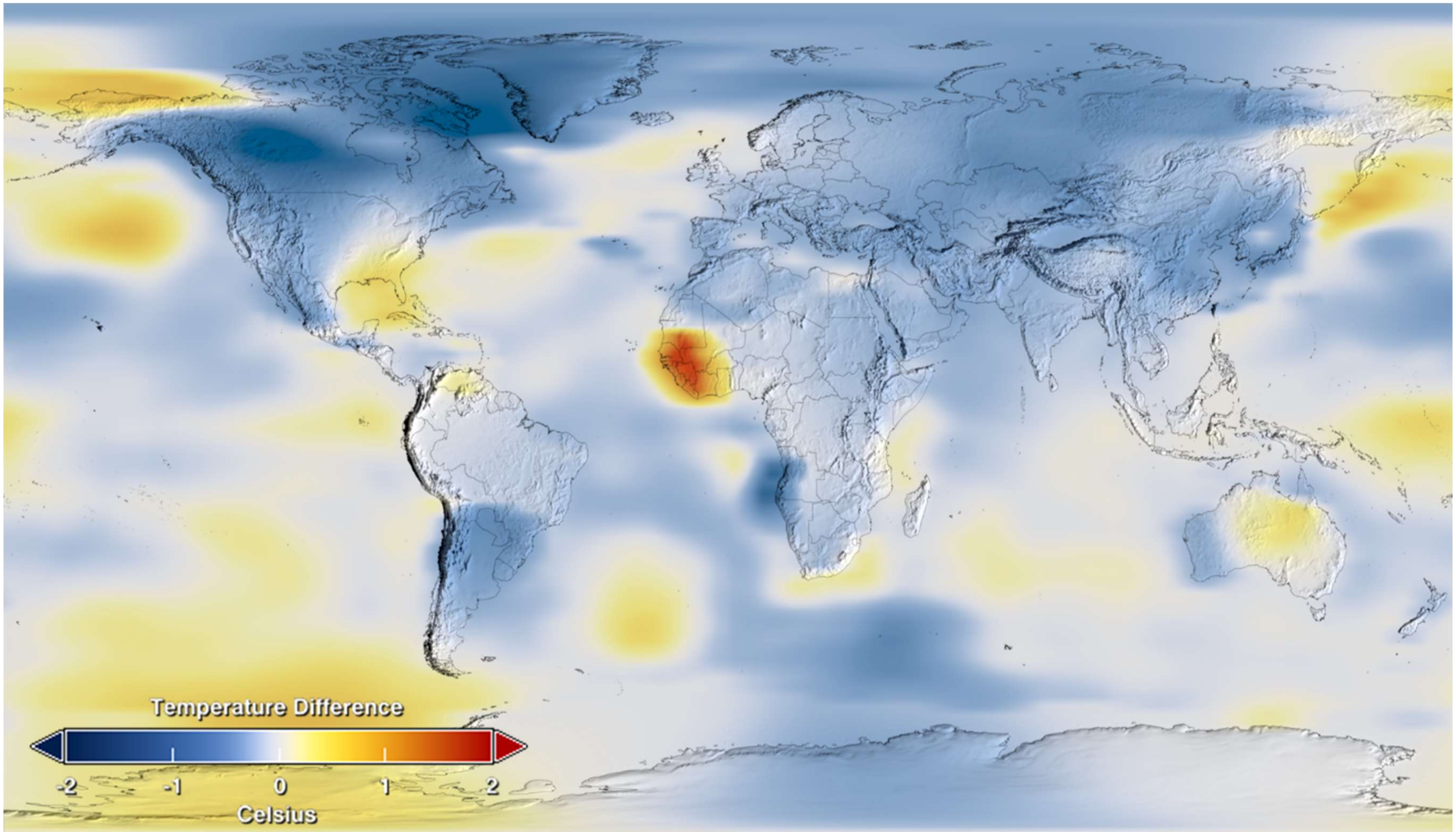



*Figure 5.8:* Surface air temperature anomaly and its projections

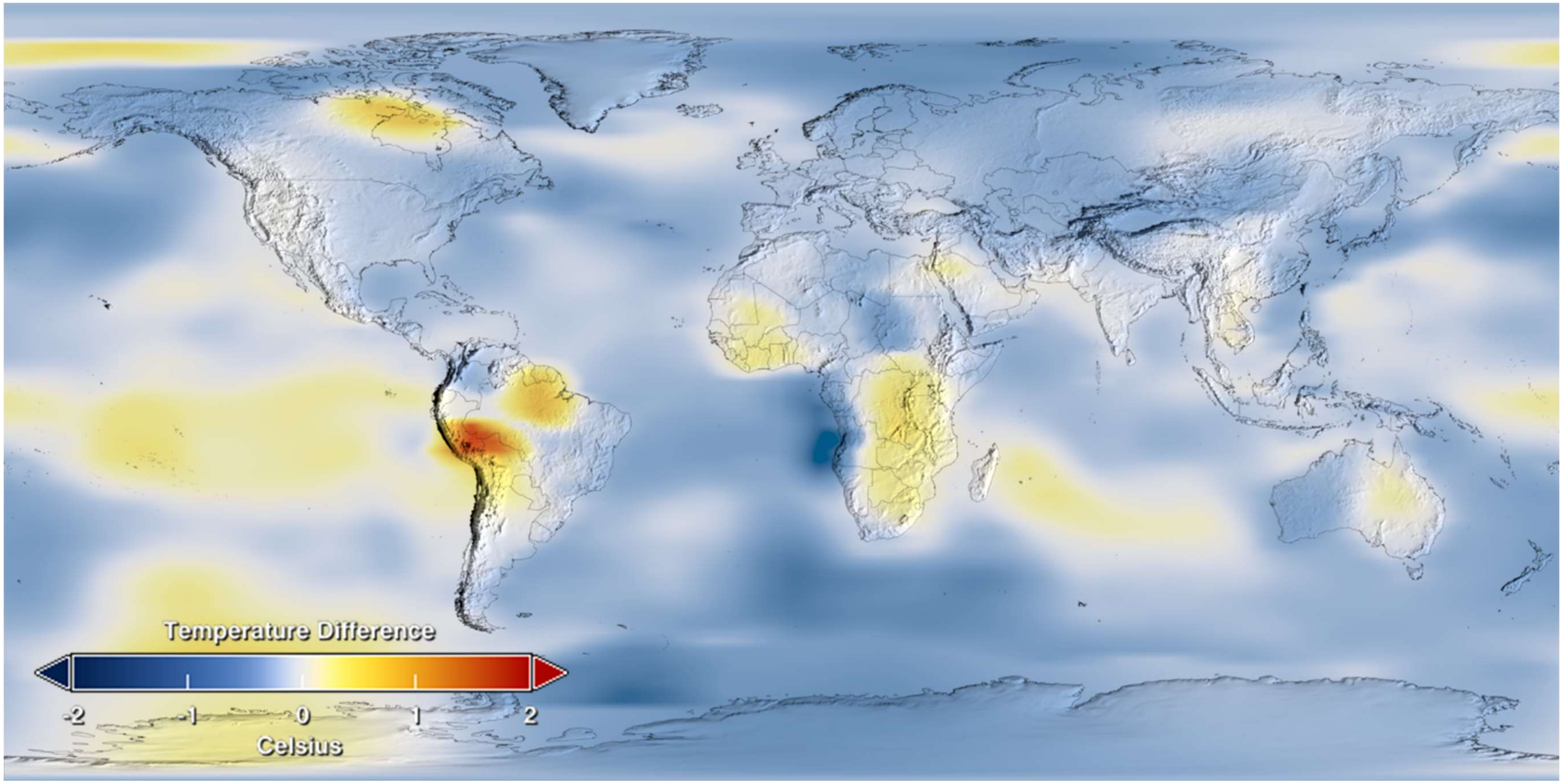
The surface air temperature changed over the years starting from 1940 to 2120. The colored lines represent the Shared socioeconomic pathways (SSP) and the confidence interval.




7

*Figure 5.9:* Global Surface temperature anomaly for year 1884 - Hind cast using NASA GISS Model dataset.

[7] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).




8

*Figure 5.10:* Global Surface temperature anomaly for year 1904 - Hind cast using NASA GISS Model dataset.

[8] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).

[9]

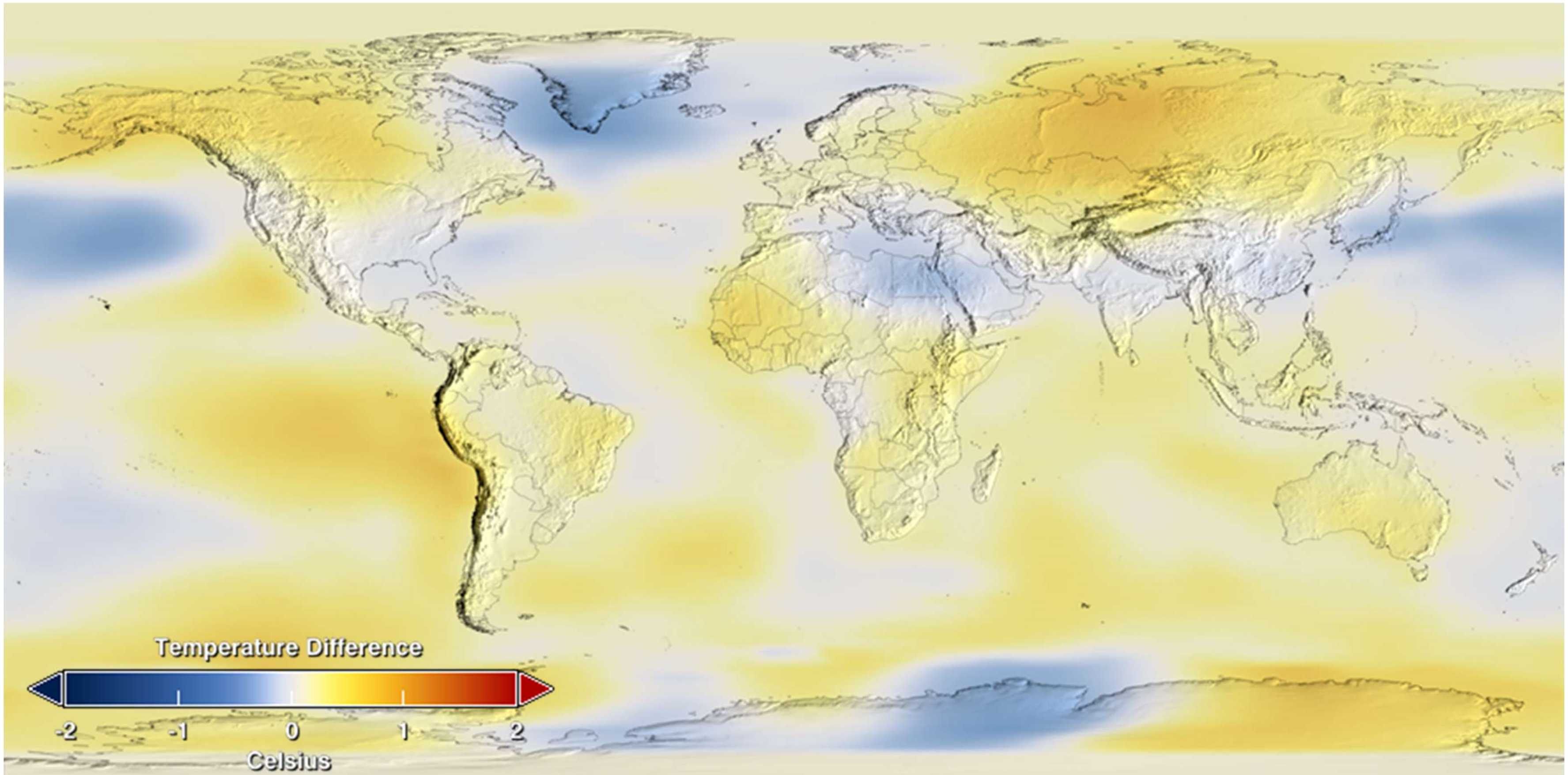
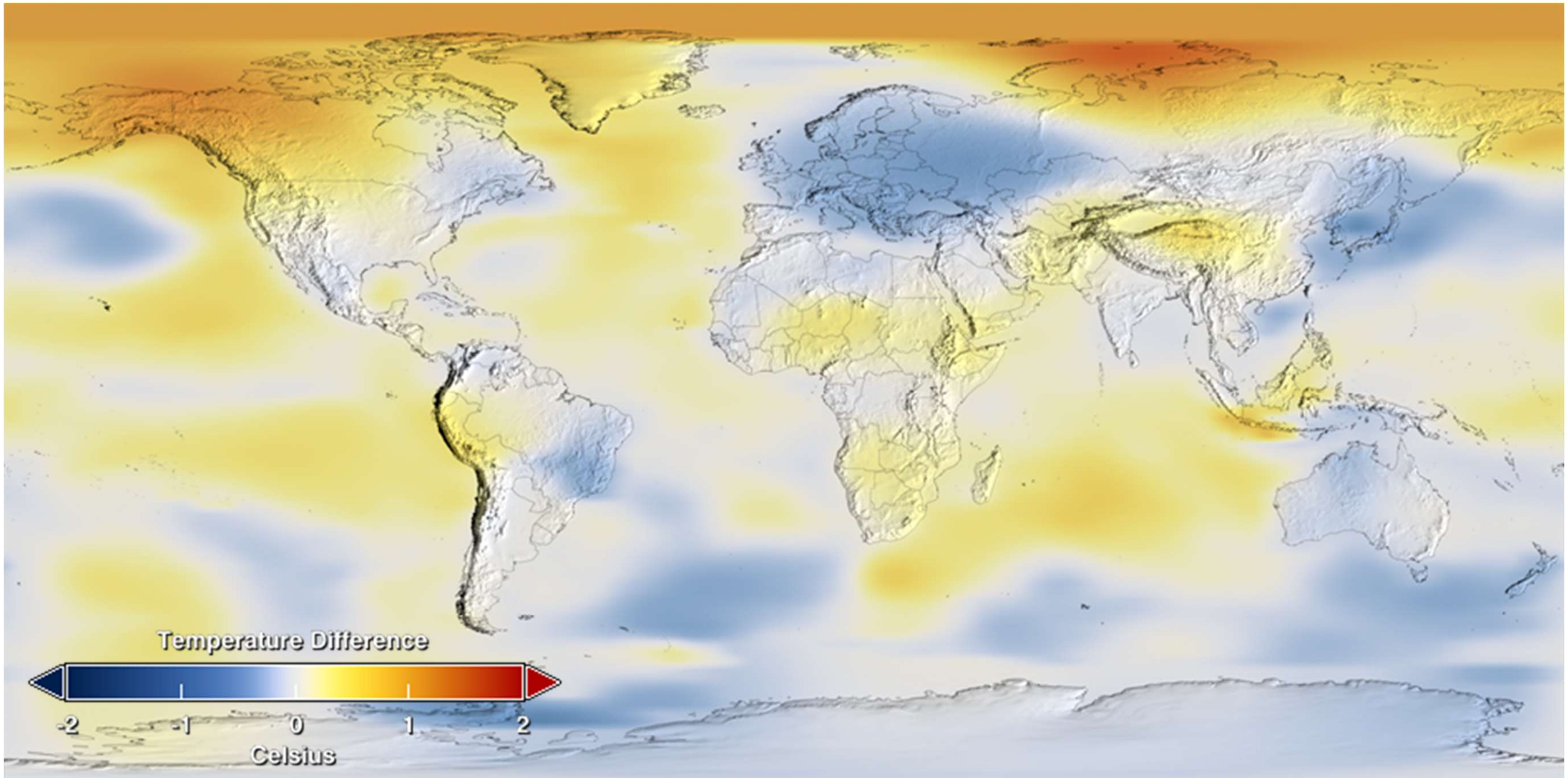



*Figure 5.11:* Global Surface temperature anomaly for year 1944 - Hind cast using NASA GISS Model dataset.

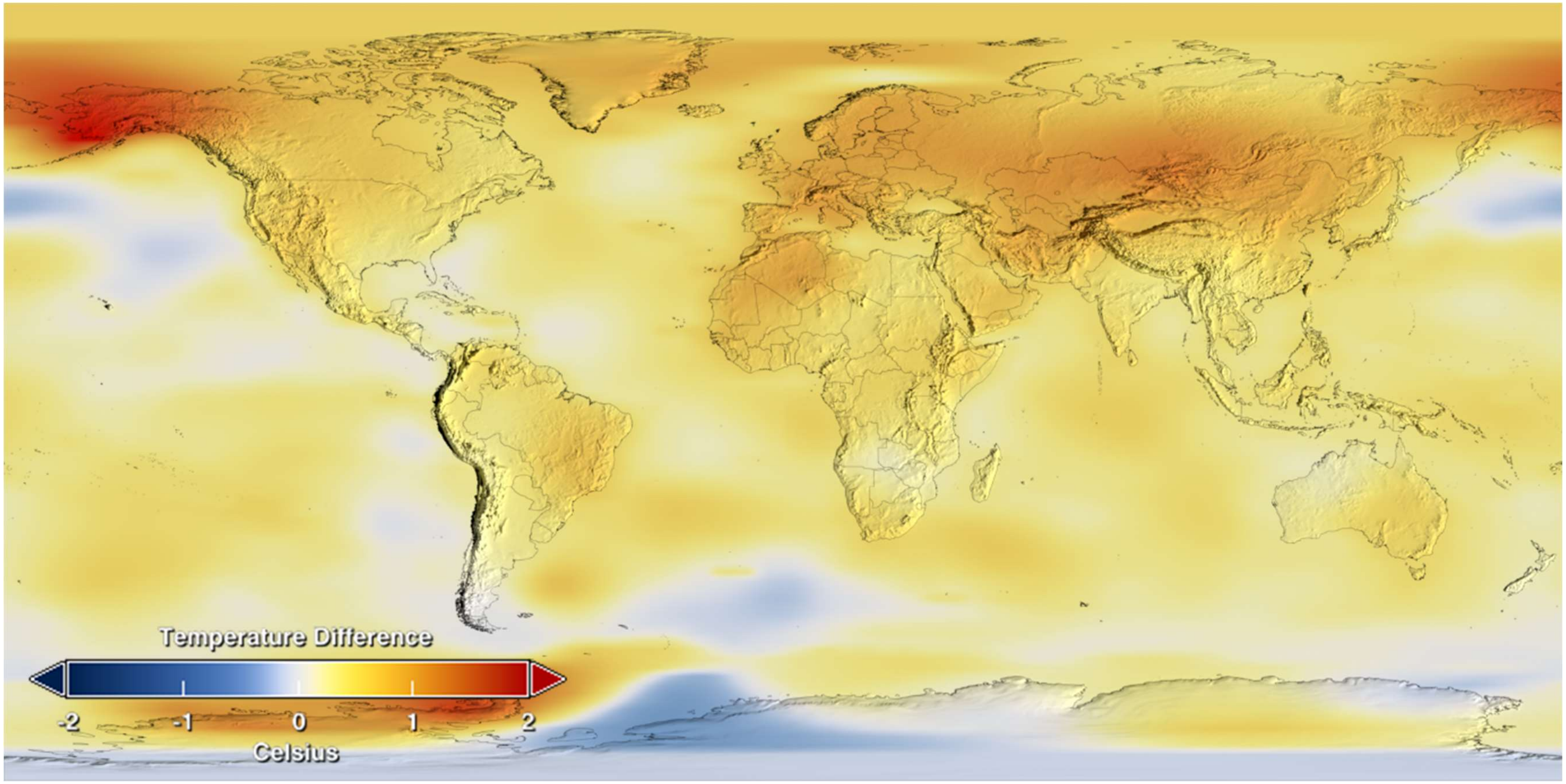
[9] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).



10

*Figure 5.12:* Global Surface temperature anomaly for year 1984 - Hind cast using NASA GISS Model dataset.

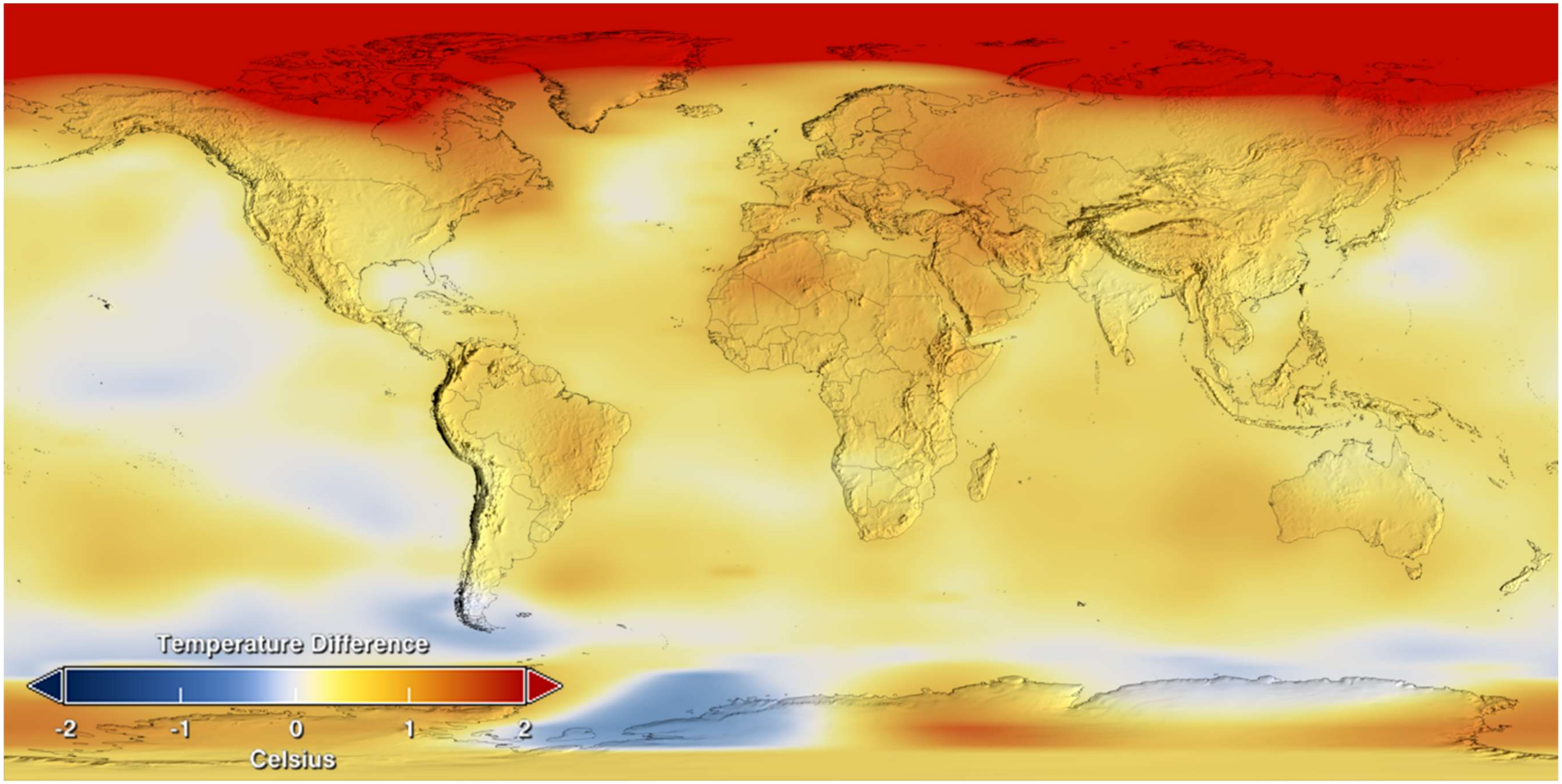
[10] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).




11

*Figure 5.13:* Global Surface temperature anomaly for year 2004 - Hind cast using NASA GISS Model dataset.

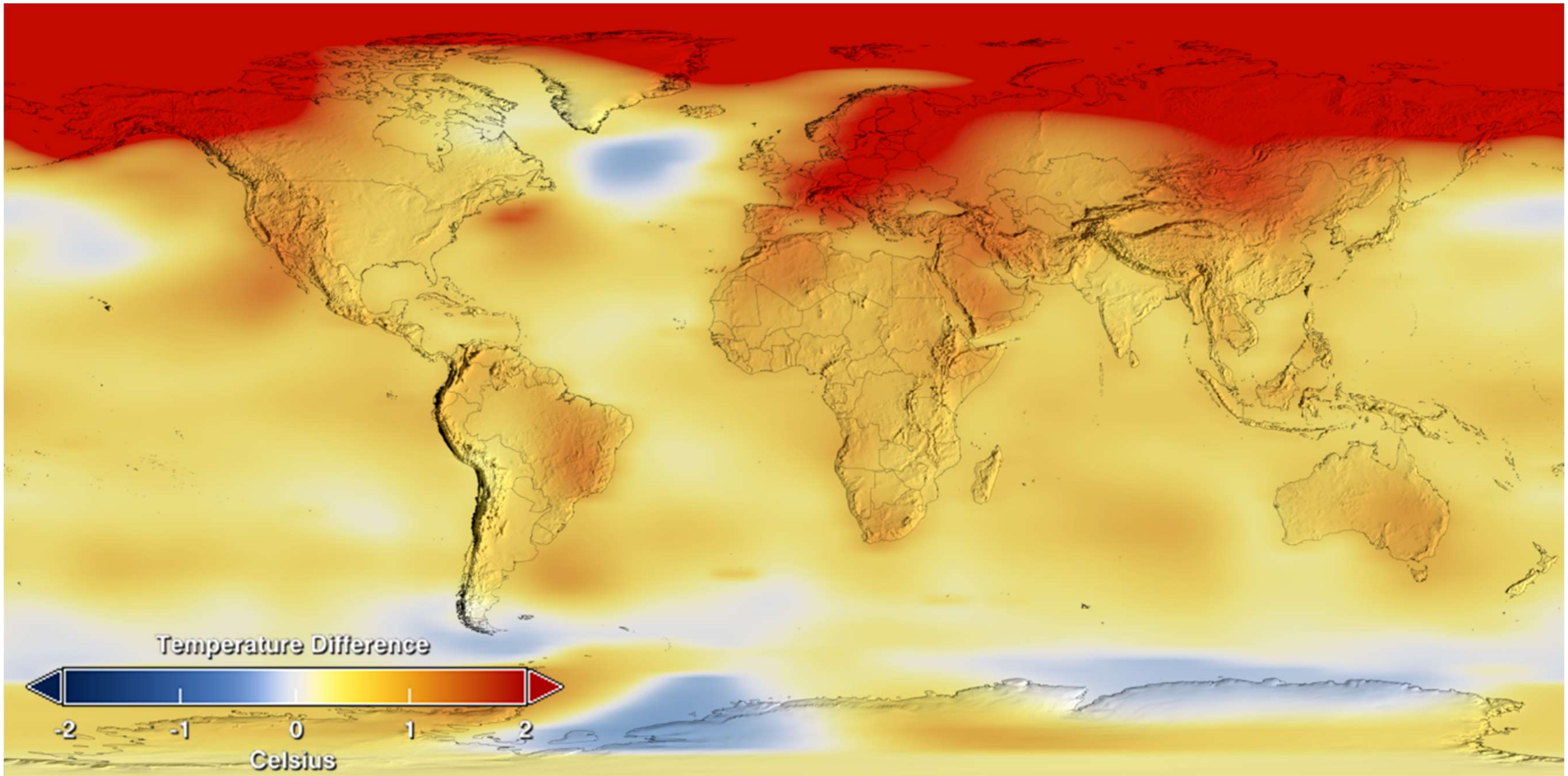
[11] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).




12

*Figure 5.14:* Global Surface temperature anomaly for year 2014 - Hind cast using NASA GISS Model dataset.

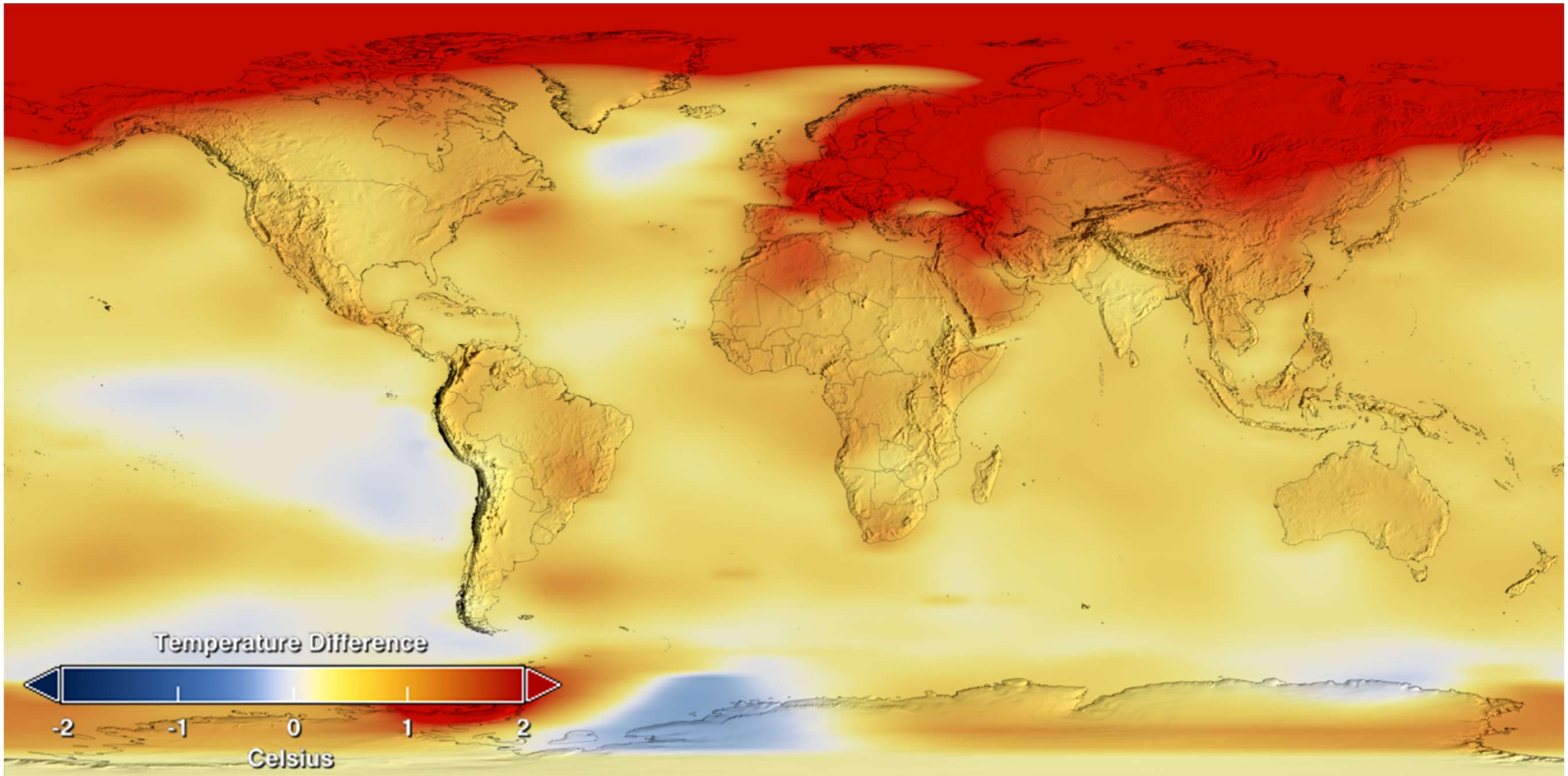
[12] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).




13

*Figure 5.15:* Global Surface temperature anomaly for year 2018 - Hind cast using NASA GISS Model dataset.

[13] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).




14

*Figure 5.16:* Global Surface temperature anomaly for year 2022 - Hind cast using NASA GISS Model dataset.

[14] This visualization (color coded map) is made using ArcGIS software employing NASA GISS Model dataset (GISTEMP surface temperature dataset).

### 5.3. Future Projections

The following data visualizations depict projected temperature changes throughout the 21st century, utilizing the climate model provided by NASA's GISS dataset (E-2-H). These visualizations present the average outcomes based on specific groups of the Shared Socioeconomic Pathways (SSPs), which were developed by the IPCC AR6. Several of these SSP scenarios have been chosen to drive climate models for CMIP6.

The SSPs encompass a wide range of potential global greenhouse gas emission and sequestration scenarios for the next century. Each pathway is assigned a numerical value based on the expected Watts per square meter, indicating the amount of heat energy retained by the climate system under each scenario. The pathways take into account the future concentrations of carbon dioxide and other greenhouse gases. Currently, the atmospheric carbon dioxide concentration stands at approximately 400 parts (previously 300 parts per million).

SSPs highlights the significant global warming as indicated by its various pathways such as SSP1-1.9 scenario, targeting a 1.5°C limit, predicts an average warming of 1.4°C across multiple models. Similarly, the SSP1-2.6 scenario, aligned with the RCP2.6 of AR5, indicates an average warming of 2.0°C. Conversely, the SSP5-8.5 scenario suggests a higher average warming of 5.0°C, while the new SSP3-7.0 scenario forecasts 4.1°C of global warming.

Each visualization represents the average output of a distinct set of models for each scenario (SSPs), including the Representative Concentration Pathways (RCPs) from the previous CMIP5. All of these models compare temperature projections from 2006-2099 against a historical baseline average from 1971-2000.

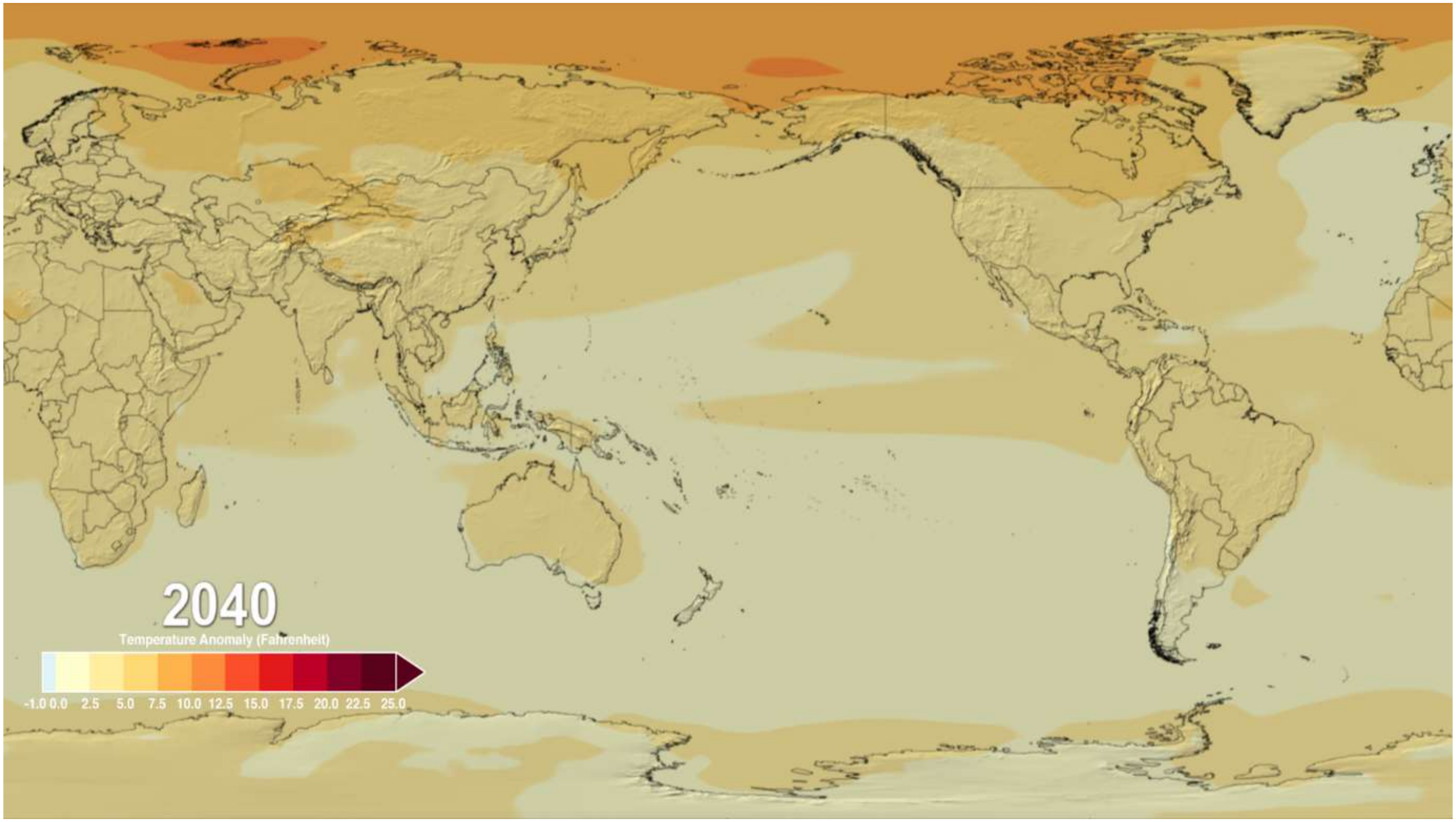
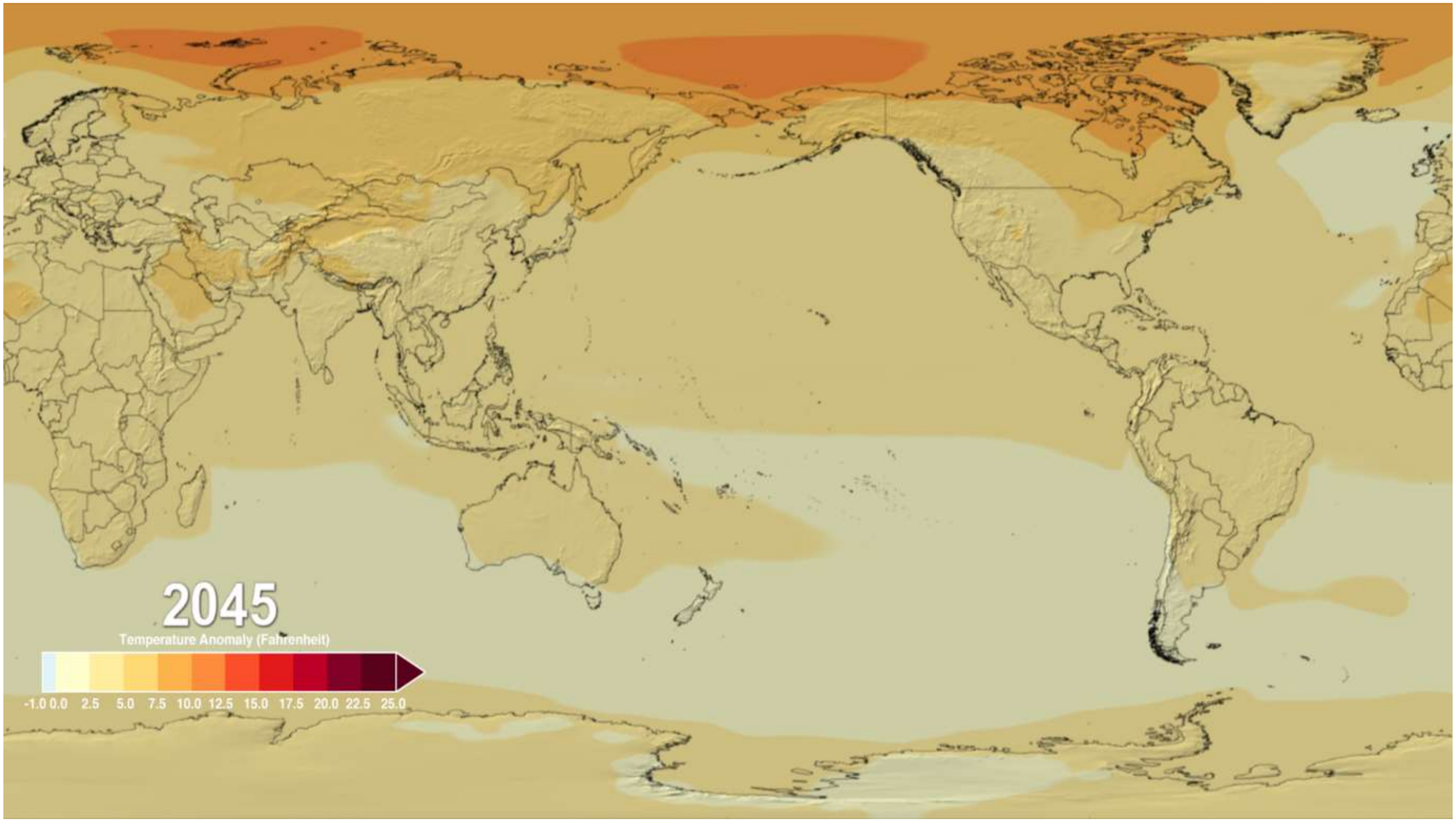


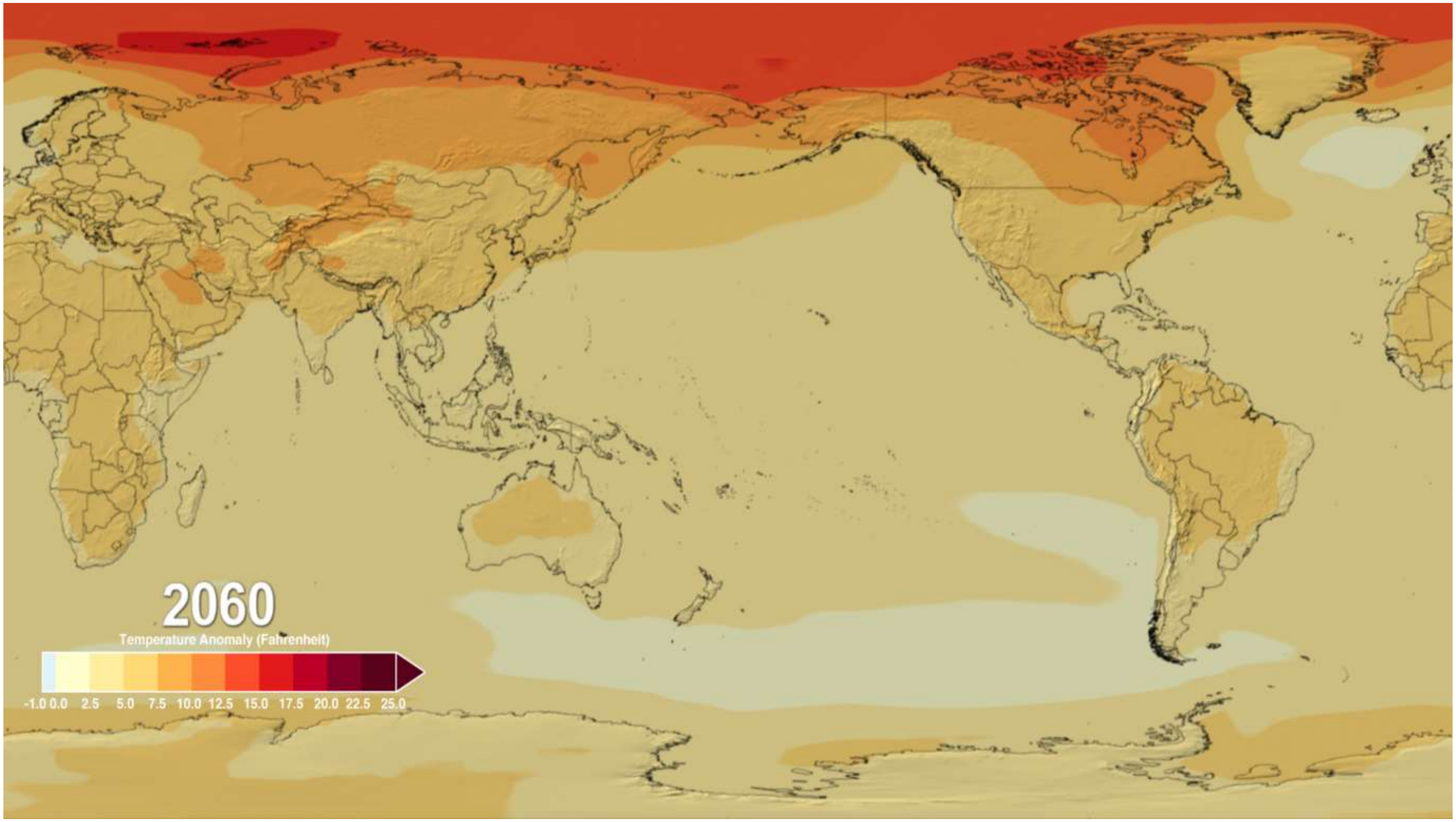

*Figure 5.17:* Global Surface temperature anomaly for year 2040 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.




16

*Figure 5.18:* Global Surface temperature anomaly for year 2045 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.

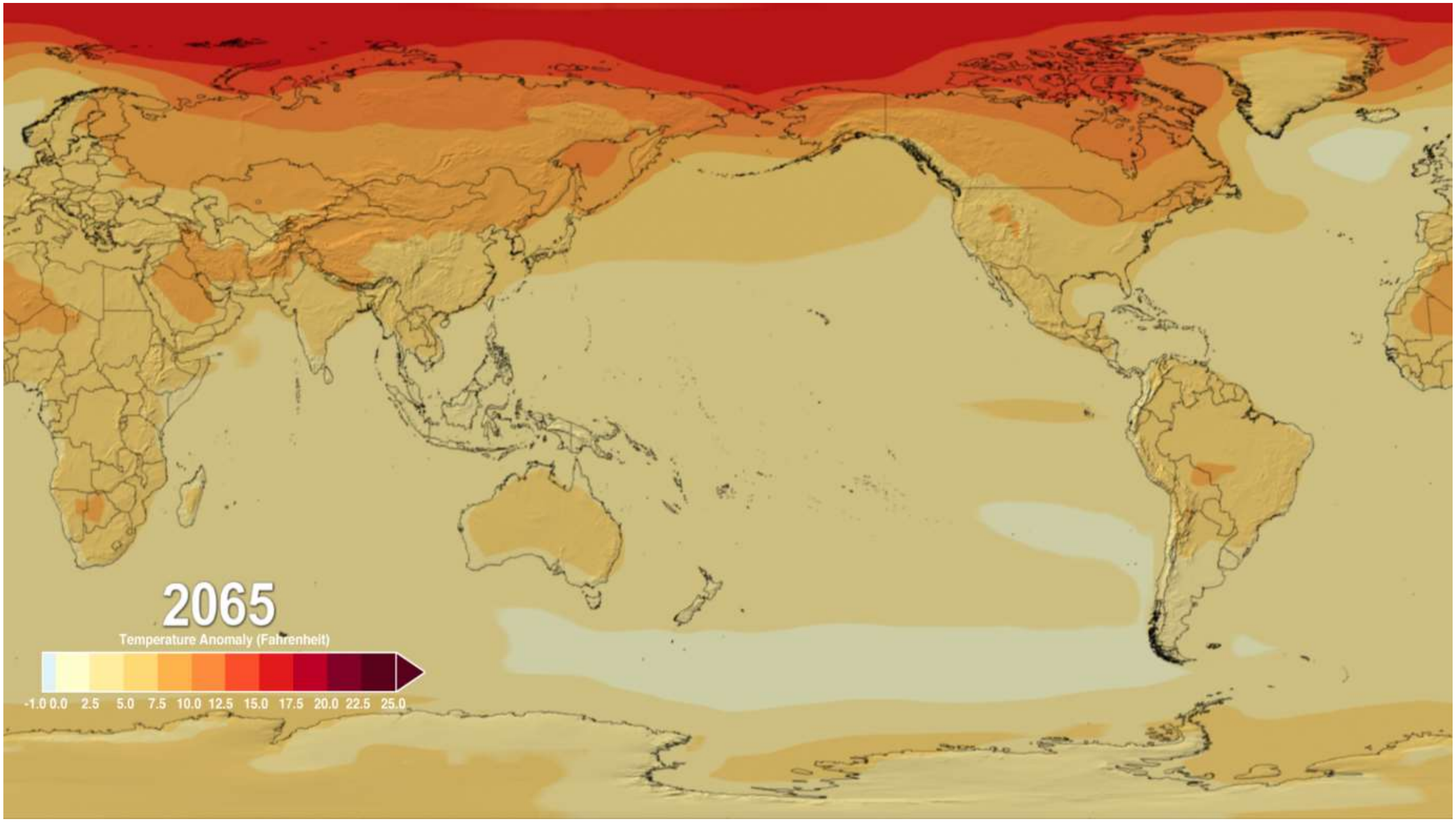
[16] This visualization (color coded map) is made using ArcGIS software employing NASA GISS E-2-H Model dataset (surface temperature dataset).




17

*Figure 5.19:* Global Surface temperature anomaly for year 2060 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.

[17] This visualization (color coded map) is made using ArcGIS software employing NASA GISS E-2-H Model dataset (surface temperature dataset).

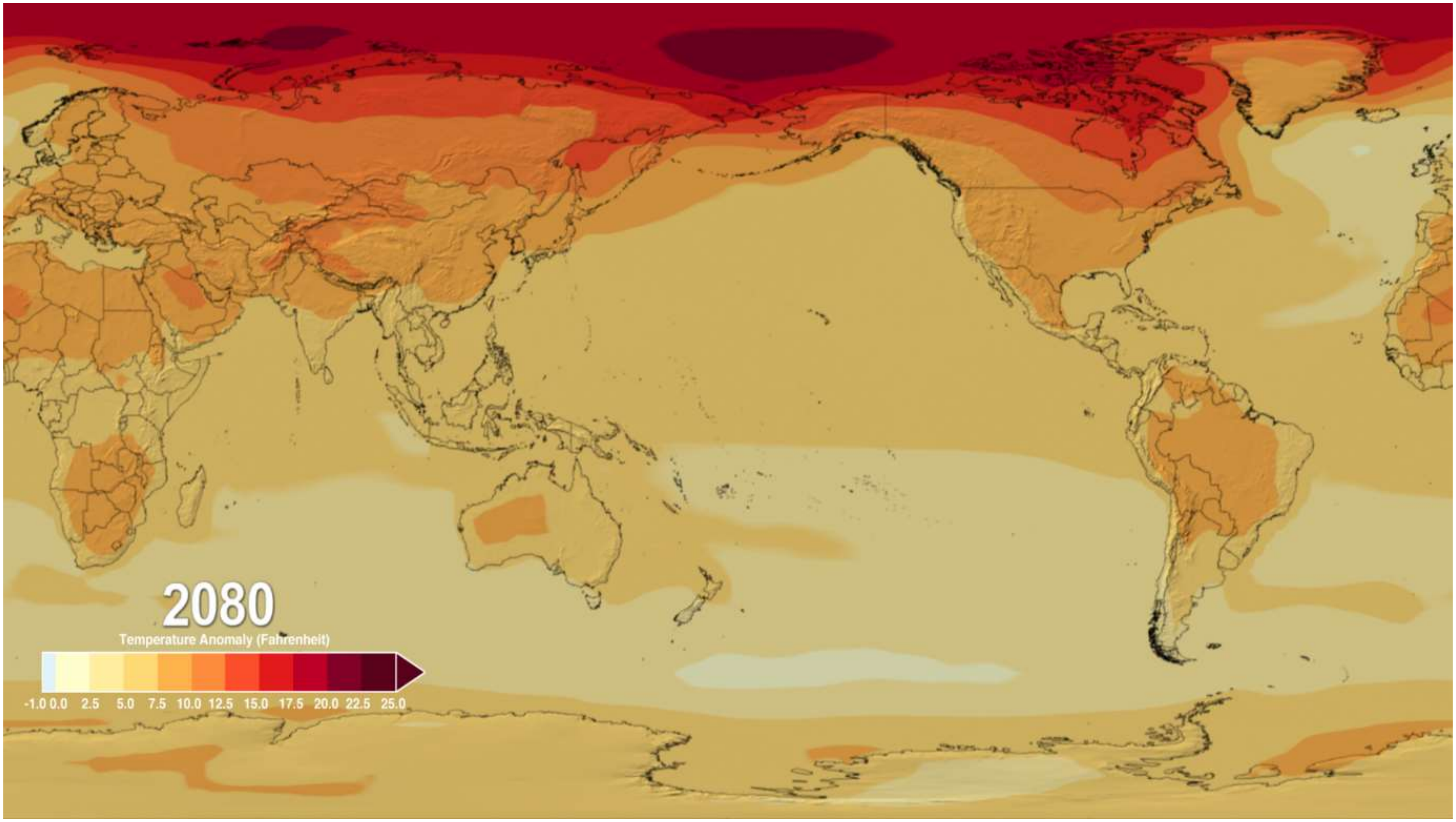



[18]

*Figure 5.20:* Global Surface temperature anomaly for year 2065 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.

---

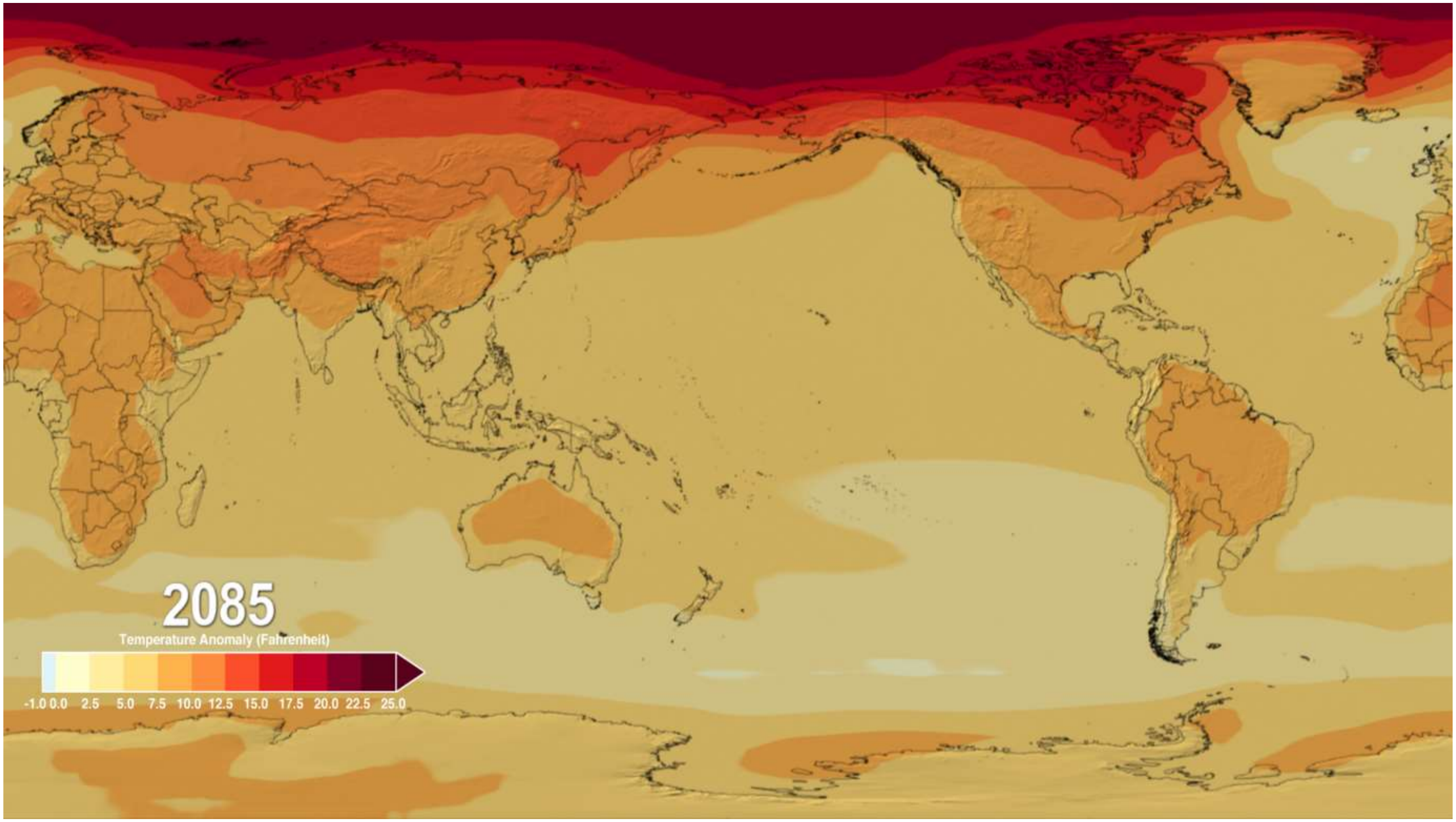
[18] This visualization (color coded map) is made using ArcGIS software employing NASA GISS E-2-H Model dataset (surface temperature dataset).




19

*Figure 5.21:*Global Surface temperature anomaly for year 2080 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.

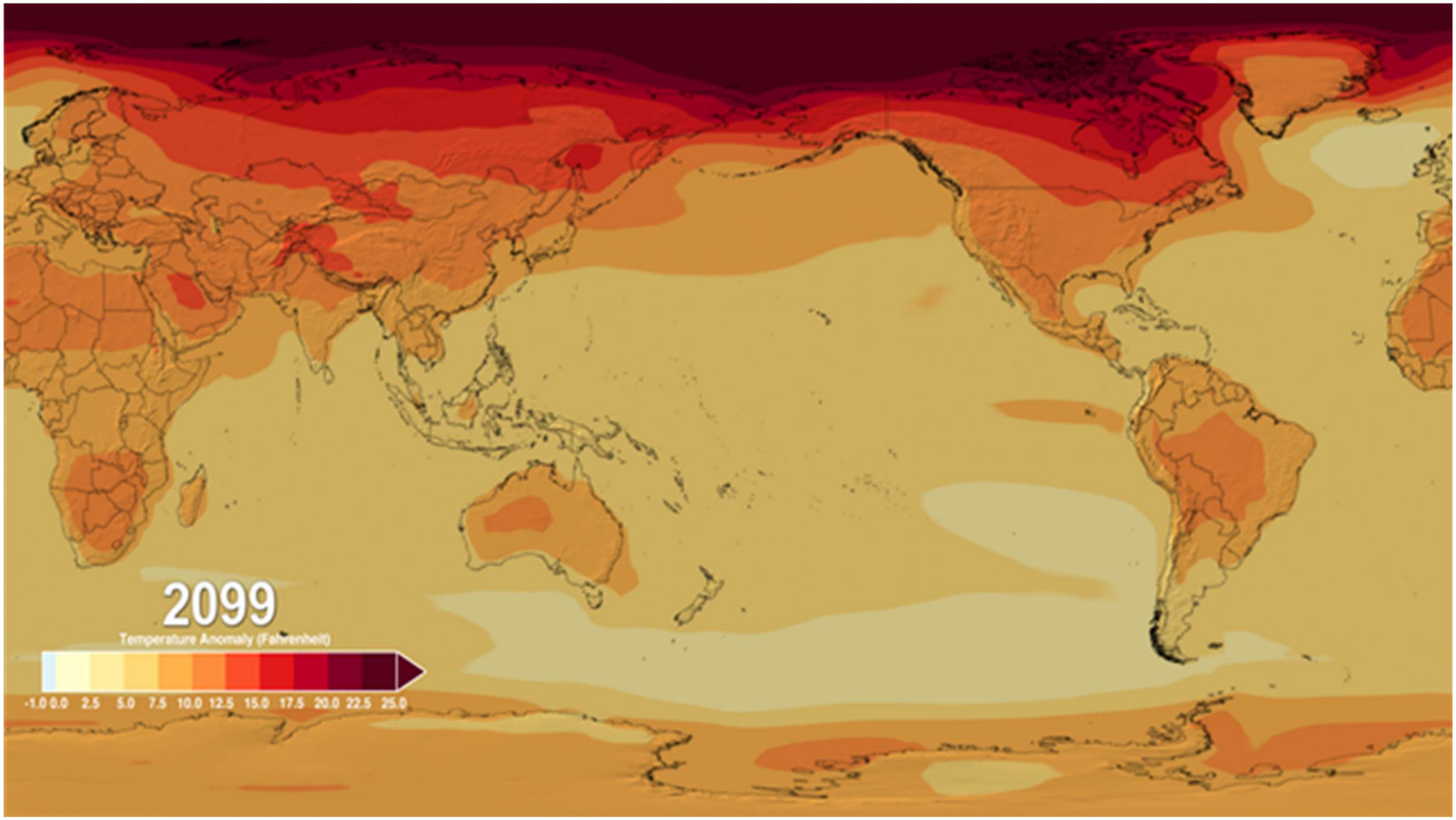
[19] This visualization (color coded map) is made using ArcGIS software employing NASA GISS E-2-H Model dataset (surface temperature dataset).




[20]

*Figure 5.22:* Global Surface temperature anomaly for year 2085 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.

[20] This visualization (color coded map) is made using ArcGIS software employing NASA GISS E-2-H Model dataset (surface temperature dataset).



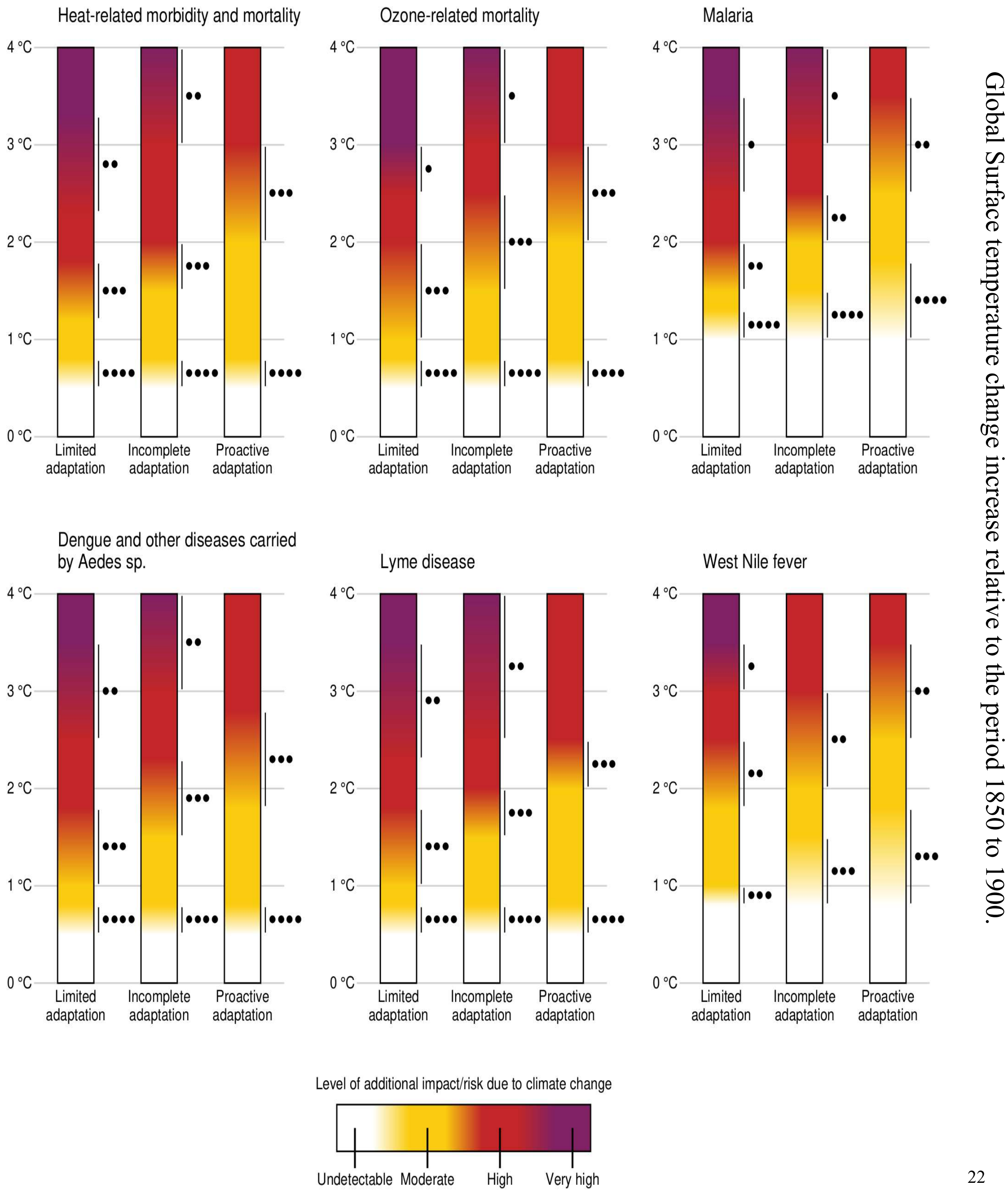

21

*Figure 5.23:* Global Surface temperature anomaly for year 2099 – Forecast Under Shared Socioeconomic Pathways (SSPs) using NASA GISS Model dataset.

[21] This visualization (color coded map) is made using ArcGIS software employing NASA GISS E-2-H Model dataset (surface temperature dataset).

### 5.4. Results for climate-sensitive health under three different adaptation scenarios:

Heath outcomes under the different adaptation scenarios given by the AR-6-WGII- Risk Assessment (IPCC) in relation to climate change are given below. With rising change in the global surface temperature relative to the base line of 1850-1900. Health outcome include the Heat related morbidity and mortality, Ozone related mortality and vector borne diseases. The risk transition outcomes are measured from Low to moderate to high.



[22]

*Figure 5.24:* Climate sensitive health outcomes under three adaptations of AR6-WGII-Risk Assessment

[22] This visualization is made using the python build app "Ember Factory" (https://climrisk.org/emberfactory). This application was made for making the burning ember diagram used in IPCC reports. Data from the AR6 -WGII- Risk Assessment for climate sensitive health outcomes was fed to this application and the burning ember diagram was visualized.

# CHAPTER 6

# CONCLUSION

Global warming as a direct result of climate change is the biggest challenge for Pakistan. In an attempt to investigate the climate change linkages to human deaths in Pakistan to identify the extent to which climate change will impact the people of Pakistan, this research employs machine learning techniques to examine the relationship between climate change and human health on a global scale. The analysis confirmed strong correlations between climate change and respiratory diseases, while the correlations with digestive problems and cardiovascular diseases were less significant. Temperature was found to play a crucial role in the relationship with respiratory diseases. More focused analysis and extensive data are needed to obtain more precise results. Causality is rooted in the historical quest to understand natural phenomena. Correlation does not imply causation, and caution must be exercised in data mining techniques that rely on correlations. The rise of artificial intelligence has highlighted the need for trustworthy machine learning tools. Causality has the potential to overcome current limitations in machine learning. The study of causality is multidisciplinary, involving fields such as epidemiology, economics, statistics, and computer science. Two main tasks in causality are causal discovery and causal inference, with different approaches and algorithms. Utilizing causal discovery tools to extrapolate and examine the causal linkages between climate change and human deaths. The results highlighted the direct and indirect relationship with the indicators of the climate change and leading causes of deaths in Pakistan.

It is worth noting that while our study provides valuable insights into the complex relationship between climate change and human health, further in-depth analysis and extensive data are necessary to obtain more precise results. A deeper understanding of causality in the context of climate change can facilitate more effective policies and interventions to mitigate the adverse health effects of climate change in Pakistan. Such measures should involve a multidisciplinary approach, incorporating expertise from various fields and focusing on the development of reliable machine learning tools.

In summary, this study underscores the urgency of addressing climate change as a formidable challenge facing Pakistan. By unraveling the correlations between climate change and respiratory diseases, we contribute to the growing body of knowledge on the subject. Our findings underscore the importance of causality in understanding complex phenomena, caution against drawing hasty causal conclusions based solely on correlations and highlight the potential of causality in overcoming limitations in machine learning. By further exploring the intricate causal pathways and gathering more extensive data, we can enhance our understanding of the relationship between climate change and human health, ultimately informing evidence-based strategies for the well-being of the people of Pakistan.

## 6.1. Recommendations

It is imperative for using correlations with caution in the realm of the big data such as climate change science, natural sciences and socioeconomic scenarios as their presence does not inherently imply causal relationships. The increasing demand for reliable machine learning tools has prompted extensive scientific and empirical research on the topic of causality, bridging interdisciplinary gaps between philosophical investigations, empirical studies, and the domains of artificial intelligence and machine learning. Causality, as a concept, holds promise in addressing the existing limitations of machine learning techniques, with causal discovery and causal inference emerging as primary tasks within this field. Causal discovery entails the extraction of causal relationships from observational data, while causal inference focuses on the testing and quantification of the impact that one variable has on another.

Within the context of Pakistan, despite its relatively minor contribution to global greenhouse gas emissions, the country remains highly susceptible to the adverse effects of climate change. Projections indicate a significant increase in mean annual temperature by approximately 6.1°C on average from 1990 to 2100 under a high emissions scenario. However, the implementation of substantial global emission reductions could limit this temperature increase to around 1.7°C. Pakistan faces numerous climate-related challenges, including the melting Himalayan glaciers, which pose a substantial threat to river flows, as well as the heightened frequency and severity of monsoons, cyclones, and saline intrusion. Furthermore, health risks such as the proliferation of vector-borne diseases and increased heat stress further compound the vulnerabilities faced by the population.

Pakistan's national development policy, "Pakistan Vision 2025," lays a strong emphasis on long-term development complementing growth in the economy and inclusion in society to overcome these complex issues. The nation is aggressively developing its institutional and technical capacities to effectively address climate-related concerns. It has launched specific measures targeted at health adaptability to climate change. In addition, Pakistan is working on creating a comprehensive national policy that considers the health effects of climate change mitigation measures. To do this, a few important proposals have been put up, one of which is the requirement for a national evaluation to gauge the health-related impacts, vulnerabilities, and adaptation options of climate change. Furthermore, efforts should be directed towards bolstering the climate resilience of health infrastructure and integrating climate information within existing disease surveillance and response systems.

Mitigation measures should involve conducting assessments to ascertain the health co-benefits of climate change mitigation policies. Furthermore, it is crucial to develop and approve a national health adaptation strategy while ensuring that health considerations are incorporated into the National Communications to the United Nations Framework Convention on Climate Change (UNFCCC).

Pakistan also faces significant risks related to inland river floods, with projections indicating that by 2030, an additional 1.5 million people may be annually exposed to such flooding due to climate change, with an additional 638,800 individuals at risk due to socioeconomic changes. The consequences of these floods extend beyond direct fatalities, encompassing secondary health effects such as impact on the production of food, availability of water, ecosystem collapse, epidemics of infectious diseases, and insect dispersal, among other things. Flooding can have long-term effects, including population displacement and post-traumatic stress disorders.

Moreover, it is worth noting that climate change has implications for the transmission of infectious diseases, as certain highly virulent infections are particularly sensitive to climate factors such as temperature, precipitation, and humidity. While socioeconomic development and health interventions have contributed to the reduction of disease burdens, climate change is expected to create more favorable conditions for disease transmission, potentially hampering progress in disease reduction efforts and placing larger populations at risk if control measures are not adequately maintained or strengthened.

Additionally, the projected rise in the average yearly temperature as well as the rising frequency and severity of heatwaves have significant effects on human health, especially for vulnerable groups like those over 65, kids, those with chronic medical conditions, those who are socially isolated, and those in vulnerable occupational groups. Under the worst-case scenario, it is anticipated that by 2080, there will be 63 heat-related deaths per 100,000 people, up from an estimated baseline of fewer than 10 deaths per 100,000 per year between 1961 and 1990. By 2080, however, rapid global emission reductions might be able to keep the number of elderly people dying from heat-related causes at roughly 17 per 100,000 people.

Moreover, climate change exerts adverse effects on agricultural production, food systems, and subsequently increases the risk of food insecurity. These consequences disproportionately affect vulnerable populations and can potentially exacerbate existing issues related to hunger and malnutrition.

Lastly, outdoor air pollution has direct and severe health consequences. Fine particulate matter, capable of penetrating the respiratory tract, has been linked to increased mortality rates from respiratory infections, lung cancer, and cardiovascular diseases.

Addressing the challenges posed by climate change requires a comprehensive and multidimensional approach. By implementing the below mentioned Policy recommendations, Pakistan can enhance its resilience, mitigate vulnerabilities, and promote sustainable development in the face of a changing climate.

1. Development & Designing of RCM: Pakistan being the most vulnerable country in the world in terms of climate change adverse impacts. With raging monsoons, flash floods and melting ice caps, harbinger for the worst and alarming situation to be faced in the future. In regard, its adamant to invest, design, plan and implement climate change mitigation policies, which are not possible without exact data and scientific evidence. Pakistan doesn't have regional climate model and relies on the global climate models (GCMs) which comes with their own drawbacks. With the development of RCM, Pakistan will be able to

accurately forecast and hindcast, historical, and futuristic simulations in context of our own weather conditions and climate. The biggest drawback of the GCMs is they don't account for the regional climatic anomalies, such as monsoon etc. With the development of RCM, Pakistan will be able to better predict the future scenarios.

Recommended Policy Action: The research underscores the limitations of using Global Climate Models (GCMs) and the necessity for regional-specific data due to Pakistan's unique climatic conditions, particularly the role of monsoons. Developing and implementing RCMs could provide more accurate climate projections and health-related assessments. Therefore, a policy direction could involve investing in and establishing RCMs to better understand local climate dynamics and their impact on health.

2. Development of Data Bank & Specializations: Development of climate change data bank is vital for the evidence building and effective policy making for the mitigation of the climate change.

3. Conduct a comprehensive national assessment: Pakistan should undertake a thorough national assessment to understand the specific impacts of climate change on health and identify vulnerabilities. This assessment should be tailored to Pakistan's unique geographic and socioeconomic context, enabling the development of targeted adaptation and mitigation strategies.

   Recommended Policy Action: Initiating a comprehensive national assessment specifically tailored to Pakistan's geographical and socioeconomic context is crucial. This could involve the establishment of a specialized data bank focused on climate change and health, ensuring data availability for policymaking and implementation.

4. Enhance resilience of health infrastructure: To effectively address the projected risks and vulnerabilities, it is essential to invest in strengthening the resilience of health infrastructure. This may involve measures such as improving the structural integrity of healthcare facilities, ensuring access to clean water and sanitation during extreme weather events, and developing robust emergency response plans.

Recommended Policy Action: Formulating policies and strategies aimed at improving the resilience of health infrastructure is essential. This might include measures to enhance the structural integrity of healthcare facilities, ensuring access to clean water and sanitation during extreme weather events, and developing robust emergency response plans to safeguard against climate-related health risks.

5. Integrate climate information into disease surveillance: Pakistan should integrate climate information into its disease surveillance and response systems. By incorporating climate data, the country can enhance its ability to detect and respond to climate-sensitive diseases, enabling the implementation of timely and targeted interventions.

6. Policy Action - Integrated Disease Surveillance and Response Systems: Integrating climate information into disease surveillance systems could enable more proactive and targeted responses to climate-sensitive diseases. This policy direction involves leveraging climate data to enhance disease surveillance, which can lead to timely interventions and better-prepared health responses. Evaluate health co-benefits of climate change mitigation: Conduct a comprehensive evaluation to determine the health co-benefits associated with climate change mitigation policies. This analysis will provide evidence of the positive impact that reducing greenhouse gas emissions can have on public health. The findings can inform policy decisions and resource allocation towards climate change mitigation measures.

   Recommended Policy Action: Policymakers could focus on conducting a thorough evaluation to identify the health co-benefits associated with climate change mitigation policies. Understanding how reducing greenhouse gas emissions positively impacts public health can inform policy decisions and resource allocation toward mitigation measures.

7. Developing a national health adaptation strategy: It is crucial for Pakistan to prioritize the development and approval of a comprehensive national health adaptation strategy. This strategy could outline specific actions and measures to address the health impacts of climate change, including capacity-building initiatives, public awareness campaigns, and community-based interventions.

Recommended Policy Action: A critical policy direction would be to prioritize the development and approval of a comprehensive national health adaptation strategy.

Incorporate health considerations in national communications: Pakistan should ensure that health considerations are appropriately integrated into its national communications submitted to the United Nations Framework Convention on Climate Change (UNFCCC). This will help raise awareness about the health implications of climate change and highlight the country's commitment to addressing these challenges.

### 6.1.2 Future Policy Direction & Measures:

Integrating the recommended policy actions previously outlined can further enrich the approach to address climate change impacts on human health in Pakistan:

1. Leveraging Government Policy for Climate Resilience: Highlighting the significance of government policies in fostering climate resilience is pivotal. Emphasize that effective policy implementation can multiply the efficiency of adaptation and mitigation measures. Acknowledge the role of policies in facilitating resilience through various sectors and spheres impacted by climate change.

2. Navigating Political Economy Concerns in Policy Alignment: Recognize that while democracies like Pakistan may prioritize citizens' long-term well-being, political economy concerns can divert government interests away from long-term adaptation. Propose strategies that align short-term political goals with long-term climate adaptation objectives.

3. One Health Concept Integration into Climate Policy: Emphasize the incorporation of the "One Health" concept into future climate change adaptation and mitigation policies in Pakistan. This integrated approach can address the interconnections between human health, animal health, and environmental health in the context of climate change.

4. Critical Aspects for Effective Policy Making: Outlining the four essential aspects that effective CC and Human Health policy could address:

    a) Highlight the significant negative impacts of CC on households, firms, and the country's economy in terms of income and mortality.

b) Stress the spatial and temporal persistence of these effects, emphasizing the need for sustained adaptation measures.

c) Acknowledge that individual adaptation efforts often fall short of completely mitigating climate impacts, thus underlining the necessity for policies that support and facilitate adaptation, leading to substantial welfare gains.

d) Considering socio-economic policies as a means to provide safety nets and reduce obstacles that hinder effective adaptation, while acknowledging the potential for political economy concerns to shift focus away from climate resilience.

5. Attraction & Optimal Allocation of International Climate Financing: Highlight the need for a strategic allocation of international climate financing towards high-impact regions and policies in Pakistan. Suggest that future work should focus on determining the most effective allocation strategies to maximize the impact of received international climate financing for adaptation.

These future policy measures accentuate the need for strong government policy support, recognition of political challenges, incorporation of the "One Health" concept, and strategic allocation of financial resources to strengthen climate adaptation and resilience in Pakistan. Integrating these considerations into the outlined policy measures can enhance the effectiveness and impact of policies addressing climate-related health issues in the country. Furthermore, by implementing these recommendations, Pakistan can strengthen its capacity to respond effectively to the health impacts of climate change, safeguard vulnerable populations, and promote sustainable development in the face of evolving climatic conditions.

**6.2 Limitations of the Study**

The following are the limitations of the present study:

1. **Exclusion of Monsoon Effect:**
   This study uses the Global Climate Models (GCMs) to forecast/hindcast climate change. GCMs estimate temperature which is in fact the climate change; through the forcings

(initial drivers of the climate) which include Solar Irradiance, GHGs emissions and Aerosols, dust, smoke, and soot. Climate feedback (processes that can either amplify or reduce the effects of climate forcings) include clouds, precipitation, forest greening and browning, ice albedo and water vapors. The GCMs estimates (hindcast/forecast) the temperature anomaly and precipitation variables as these two are central to climate change. But the GCMs excluding unique conditions to certain topographies due to its spatial resolution like the effect of monsoon which is area specific. GCMs don't take into account monsoon, which renders the precipitation variable biased for Pakistan. As monsoons play a pivotal role in the country's climatic patterns, the exclusion of this crucial factor could lead to inaccuracies in assessing the true impact of climate change on human health as the result those diseases (Vector-borne diseases & Infectious diseases) that were linked to precipitation were also excluded from the analysis.

2. **Use of Global Climate Models (GCMs) Instead of Regional Climate Models (RCMs):** This research uses Global Climate Models (GCMs) as source for the climate change data and projections, as GCMs captures broad climatic trends on a planetary scale, carries limitations due to their coarse spatial resolution, typically ranging from 100 to 300 km grid spacing. In contrast, the Regional Climate Models (RCMs) which downscale projections from GCMs to finer resolutions and account for regional climate features. RCMs offer a more localized lens with higher spatial resolution, ranging from 1 to 50 km grid spacing. Incorporating RCMs would have addressed this spatial limitation and facilitated a more comprehensive understanding of climate change's impact on human health in Pakistan. The utilization of RCMs would have corrected potential biases in GCMs and would have also account for region-specific phenomena such as monsoon patterns that are critical for understanding climate-health relationships in Pakistan.

   The downscaling of the GCMs to RCMs is time taking process and unavailability of the RCM data for CMIP6 was compelling factor to opt GCM. To address this limitation, future research should consider integrating RCMs into the methodology. This would enable a more precise examination of the unique characteristics of regional climate dynamics and human health, ultimately yielding more robust insights tailored to Pakistan's context.

3. **Limited Precision:**

The machine learning techniques employed in this research provide insights into the relationship between climate change and human health as proxied by causes of deaths, but the results are constrained by the availability of the data. More extensive and detailed data would be essential for obtaining nuanced findings.

4. **Margin of Error:**

   The availability of the health data was one of the biggest constraints for this research, therefore this is margin of error on the reliability of the data. The computational and ML biases could also affect the reliability of the projections.

5. **Contextual Variability:**

   The findings pertain to the context of Pakistan and might not be directly transferable to other regions. Local variations in climate, health infrastructure, and socio-economic conditions could result in different causal dynamics.

**Conclusion:**

In conclusion, this study sheds light on the complex relationship between climate change and human health in Pakistan. While it captures how specific climatic change interacts with human health as proxied by number of deaths due to leading diseases in Pakistan. This study provides valuable insights into correlations and potential causal pathways, it's essential to acknowledge the limitations inherent in the study's design, data, and methodologies. Further research involving rigorous causal discovery methods and collaboration across disciplines can deepen the understanding of the causal mechanisms at play and guide evidence-based policy making to address the challenges posed by climate change on human health in Pakistan.

# REFERENCES

Anjum, M. S., Ali, S. M., Imad-ud-din, M., Subhani, M. A., Anwar, M. N., Nizami, A.-S., Ashraf, U., & Khokhar, M. F. (2021). An Emerged Challenge of Air Pollution and Ever-Increasing Particulate Matter in Pakistan; A Critical Review. *Journal of Hazardous Materials*, *402*, 123943–123943. https://doi.org/10.1016/j.jhazmat.2020.123943

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317. https://doi.org/10.1016/j.jeconom.2008.08.010

Berrang-Ford, L., Sietsma, A. J., Callaghan, M., Minx, J. C., Scheelbeek, P. F. D., Haddaway, N. R., Haines, A., & Dangour, A. D. (2021). Systematic mapping of global research on climate and health: A machine learning review. *The Lancet Planetary Health*, *5*(8), e514–e525. https://doi.org/10.1016/S2542-5196(21)00179-0

Checkley, W., Gilman, R. H., Black, R. E., Epstein, L. D., Cabrera, L., Sterling, C. R., & Moulton, L. H. (2004). Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community. *Lancet*, *363*(9403), 112–118. https://doi.org/10.1016/S0140-6736(03)15261-0

Confalonieri, U. ; M., B. ;. Akhtar, R. ;. Ebi, K. L. ;. Hauengue, M. ;. Kovats, R. S. ;. Revich, B. ;. Woodward, A. (2007). *Climate Change: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Vol. 37, Issue 6, p. 431). Cambridge University Press: https://doi.org/10.2134/jeq2008.0015br

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Ishwaran, H., Knight, K., Loubes, J. M., Massart, P., Madigan, D., Ridgeway, G., Rosset, S., Zhu, J. I., Stine, R. A., Turlach, B. A., Weisberg, S., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, *32*(2), 407–499. https://doi.org/10.1214/009053604000000067

Eickmeier, S., & Ng, T. (2011). Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting*, *27*(2), 496–511. https://doi.org/10.1016/j.ijforecast.2009.10.011

Fang, Y. (2015). A Study on the Correlations between Investor Sentiment and Stock Index and Macro Economy Based on EEMD Method. *Journal of Financial Risk Management*, *04*(03), 206–215. https://doi.org/10.4236/jfrm.2015.43016

Gallopín, G. C. (2006). Linkages between vulnerability, resilience, and adaptive capacity. *Global Environmental Change*, *16*(3), 293–303. https://doi.org/10.1016/j.gloenvcha.2006.02.004

Gaythorpe, K. A. M., Hamlet, A., Cibrelus, L., Garske, T., & Ferguson, N. M. (2020). The effect of climate change on yellow fever disease burden in Africa. *ELife*, *9*, 1–27. https://doi.org/10.7554/eLife.55619

Geweke, J. (1977). The dynamic factor analysis of economic time-series models. *Latent Variable in Socioeconomic Models*, 365–383.

Gosling, S. N., Lowe, J. A., McGregor, G. R., Pelling, M., & Malamud, B. D. (2009). Associations between elevated atmospheric temperature and human mortality: A critical review of the literature. *Climatic Change*, *92*(3–4), 299–341. https://doi.org/10.1007/s10584-008-9441-x

Guo, Y., Li, S., Liu, D. L., Chen, D., Williams, G., & Tong, S. (2016). Projecting future temperature-related mortality in three largest Australian cities. *Environmental Pollution*, *208*, 66–73. https://doi.org/10.1016/j.envpol.2015.09.041

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4). https://doi.org/10.1029/2010RG000345

Hayes, K., Blashki, G., Wiseman, J., Burke, S., & Reifels, L. (2018). Climate change and mental health: Risks, impacts and priority actions. *International Journal of Mental Health Systems*, *12*(1). https://doi.org/10.1186/s13033-018-0210-6

Hema, D. D., Pal, A., Loyer, V., & Gaurav, R. (2019). Global warming prediction in India using machine learning. *International Journal of Engineering and Advanced Technology*, *9*(1), 4061–4065. https://doi.org/10.35940/ijeat.A1301.109119

Jaeggi, A. V., Blackwell, A. D., Von Rueden, C., Trumble, B. C., Stieglitz, J., Garcia, A. R., Kraft, T. S., Beheim, B. A., Hooper, P. L., Kaplan, H., & Gurven, M. (2021). Do wealth and inequality associate with health in a small-scale subsistence society? *ELife*, *10*. https://doi.org/10.7554/ELIFE.59437

Kalkstein, L. S., & Greene, J. S. (1997). An evaluation of climate/mortality relationships in large U.S. cities and the possible impacts of a climate change. *Environmental Health Perspectives*, *105*(1), 84–93. https://doi.org/10.1289/ehp.9710584

Kolstad, E. W., & Johansson, K. A. (2011). Uncertainties associated with quantifying climate change impacts on human health: A case study for diarrhea. *Environmental Health Perspectives*, *119*(3), 299–305. https://doi.org/10.1289/ehp.1002060

Kunze, C., Luijckx, P., Jackson, A. L., & Donohue, I. (2022). Alternate patterns of temperature variation bring about very different disease outcomes at different mean temperatures. *ELife*, *11*. https://doi.org/10.7554/ELIFE.72861

Li, X., & Chen, W. (2014). Facebook or Renren? A comparative study of social networking site use and social capital among Chinese international students in the United States. *Computers in Human Behavior*, *35*, 116–123. https://doi.org/10.1016/j.chb.2014.02.012

Liu, H. J., Stenlund Hans, Wilder-Smith Annelies, & Rocklöv Joacim. (2014). Vectorial Capacity of Aedes aegypti: Effects of Temperature and Implications for Global Dengue Epidemic Potential. *PLOS ONE*.

Medeiros, M. C. & Gabriel Vasconcelos. (2019). *Forecasting Inflation in a Data-Rich Environment*.

Melillo, J. M., Richmond, T. C., Yohe, G. W., & Assessment, U. N. C. (2014). Climate change impacts in the United States: The third national climate assessment. In *US Global change research program* (Vol. 841, p. 841). U.S. Global Change Research Program: Washington.

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., Delgenio, A. D., Kelley, M., Lo, K. K., Ruedy, R., Shindell, D. T., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., Faluvegi, G., Hansen, J. E., … Zhang, J. (2014). CMIP5 historical simulations (1850-2012) with GISS ModelE2. *Journal of Advances in Modeling Earth Systems*, *6*(2), 441–477. https://doi.org/10.1002/2013MS000266

Mordecai EA, Paaijmans KP, & Johnson LR. (2013). Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology Letters*.

Parks, R. M., Bennett, J. E., Foreman, K. J., Toumi, R., & Ezzati, M. (2018). National and regional seasonal dynamics of all-cause and cause-specific mortality in the USA from 1980 to 2016. *ELife*, *7*. https://doi.org/10.7554/eLife.35500

Paull, S. H., Horton, D. E., Ashfaq, M., Rastogi, D., Kramer, L. D., Diffenbaugh, N. S., & Kilpatrick, A. M. (2017). Drought and immunity determine the intensity of West Nile virus epidemics and climate change impacts. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1848), 20162078–20162078. https://doi.org/10.1098/rspb.2016.2078

Pearl, J. (2018). *Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution*. 3–3. https://doi.org/10.1145/3159652.3176182

Pizzulli, V. A., Telesca, V., & Covatariu, G. (2021). Analysis of correlation between climate change and human health based on a machine learning approach. *Healthcare (Switzerland)*, *9*(1). https://doi.org/10.3390/healthcare9010086

Plakandaras, V., Gupta, R., & Wohar, M. E. (2019). Persistence of economic uncertainty: A comprehensive analysis. *Applied Economics*, *51*(41), 4477–4498. https://doi.org/10.1080/00036846.2019.1591607

Raita, Y., Camargo, C. A., Liang, L., & Hasegawa, K. (2021). Big Data, Data Science, and Causal Inference: A Primer for Clinicians. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.678047

Rau, R. (2007). Seasonality in human mortality: A demographic approach. *Demographic Research Monographs*, 1–214.

Rau, R., Bohk-Ewald, C., Muszyńska, M. M., & Vaupel, J. W. (2018). Visualizing Mortality Dynamics in the Lexis Diagram. *Springer Series on Demographic Methods and Population Analysis*, *44*, 1–169. https://doi.org/10.1007/978-3-319-64820-0_1

Rerolle, F., Dantzer, E., Lover, A., Marshall, J. M., Hongvanthong, B., Sturrock, H. J. W., & Bennett, A. (2021). Spatio-temporal associations between deforestation and malaria incidence in Lao Pdr. *ELife*, *10*. https://doi.org/10.7554/eLife.56974

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., … Zscheischler, J. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10105-3

Sargent, T. J., & Sims, C. A. (1975). Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory. *Federal Reserve Bank of Minneapolis Working Paper*, *1*, 1–89.

Scheelbeek, P. F. D., Dangour, A. D., Jarmul, S., Turner, G., Sietsma, A. J., Minx, J. C., Callaghan, M., Ajibade, I., Austin, S. E., Biesbroek, R., Bowen, K. J., Chen, T., Davis, K., Ensor, T., Ford, J. D., Galappaththi, E. K., Joe, E. T., Musah-Surugu, I. J., Alverio, G. N., … Berrang-Ford, L. (2021). The effects on public health of climate change adaptation responses: A systematic review of evidence from low- And middle-income countries. *Environmental Research Letters*, *16*(7). https://doi.org/10.1088/1748-9326/ac092c

Schwartz, J. D., Lee, M., Kinney, P. L., Yang, S., Mills, D., Sarofim, M. C., Jones, R., Streeter, R., Juliana, A. S., Peers, J., & Horton, R. M. (2015). Projections of temperature-attributable premature deaths in 209 U.S. cities using a cluster-based Poisson approach. *Environmental Health: A Global Access Science Source*, *14*(1). https://doi.org/10.1186/s12940-015-0071-2

Shocket, M. S., Ryan, S. J., & Mordecai, E. A. (2018). Temperature explains broad patterns of Ross River virus transmission. *ELife*, *7*. https://doi.org/10.7554/eLife.37762

Singh, R. B. K., Hales, S., Wet, N. D., Raj, R., Hearnden, M., & Weinstein, P. (2001). The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environmental Health Perspectives*, *109*(2), 155–159. https://doi.org/10.1289/ehp.01109155

Solomon, S., Intergovernmental Panel on Climate Change, & Intergovernmental Panel on Climate Change (Eds.). (2007). *Climate change 2007: The physical science basis: contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Song, X., Wang, S., Hu, Y., Yue, M., Zhang, T., Liu, Y., Tian, J., & Shang, K. (2017). Impact of ambient temperature on morbidity and mortality: An overview of reviews. *Science of The Total Environment*, *586*, 241–254. https://doi.org/10.1016/j.scitotenv.2017.01.212

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. https://doi.org/10.1198/016214502388618960

Studies (NASA/GISS), N. G. I. for S. (2018). *NASA-GISS GISS-E2.1H model output prepared for CMIP6 CMIP*. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.1421

Telesca, V., Lay-Ekuakille, A., Ragosta, M., Giorgio, G. A., & Lumpungu, B. (2018). Effects on public health of heat waves to improve the urban quality of life. *Sustainability (Switzerland)*, *10*(4). https://doi.org/10.3390/su10041082

Vollset, S. E., Goren, E., Yuan, C.-W., Cao, J., Smith, A. E., Hsiao, T., Bisignano, C., Azhar, G. S., Castro, E., Chalek, J., Dolgert, A. J., Frank, T., Fukutaki, K., Hay, S. I., Lozano, R., Mokdad, A. H., Nandakumar, V., Pierce, M., Pletcher, M., … Murray, C. J. L. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease Study. *The Lancet*, *396*(10258), 1285–1306. https://doi.org/10.1016/S0140-6736(20)30677-2

WHO. (2014). Report of the Sage Working Group on. *WHO COVID-19 Global Data*, *October*, 64–64.

Zhou, L. (2013). Assessment on equivalence of control level established under the Hong Kong Convention as that established under the Basel Convention. *Maritime Safety & Environment Management*.

# Glossary

| | |
|---|---|
| **IPCC AR6** | The Intergovernmental Panel on Climate Change Sixth Assessment Report (IPCC AR6) is a comprehensive report that assesses the current state of climate science, impacts, adaptation, and mitigation efforts. It serves as a key reference for policymakers and scientists worldwide. |
| **CMIP (Coupled Model Inter-comparison Project)** | CMIP is an international effort that coordinates climate model simulations to assess and compare the performance of Earth System Models (ESMs) and General Circulation Models (GCMs) in projecting future climate changes. |
| **ESMs (Earth System Models)** | ESMs are complex computer models used to simulate the interactions and processes within the Earth's climate system, including the atmosphere, oceans, land surface, and biosphere, to project future climate changes. |
| **GCMs (General Circulation Models)** | GCMs are computer models that simulate the behavior of the Earth's climate system by dividing the planet into a grid and solving mathematical equations to project climate patterns and changes. |
| **GPR (Gaussian Process Regression)** | GPR is a statistical modeling technique that uses Gaussian processes to analyze data and make predictions by capturing underlying patterns and uncertainties in the data. |

| | |
|---|---|
| **Squared Exponential Gaussian Process Regression** | A specific form of Gaussian Process Regression that utilizes a squared exponential covariance function to model data relationships, often used for smoothing and predicting continuous functions. |
| **PCMCI (Momentary conditional independence PC Algorithm)** | PCMCI is a statistical technique used for causal discovery that assesses conditional independence relationships between variables at different time points, helping to infer causal links between variables. |
| **PM (Particulate matter)** | PM refers to tiny solid or liquid particles suspended in the air, which can be harmful to human health when inhaled. PM is often categorized based on its size, such as PM2.5 (fine particles) and PM10 (coarse particles). |
| **RCMs (Regional Climate Models)** | RCMs are specialized climate models designed to provide high-resolution climate projections for specific regions, offering more detailed information than global climate models. |
| **RCP (Representative Concentration Pathways)** | RCPs are a set of scenarios used to represent future greenhouse gas concentration trajectories. They are crucial for climate modeling and assessing potential climate impacts under different emissions scenarios. |
| **SSPs (Shared Socioeconomic Pathways)** | SSPs are a set of scenarios that describe potential future socioeconomic conditions, including population, technology, and policy choices. These pathways are used to explore how |

| | |
|---|---|
| | different socioeconomic trajectories can affect future climate change. |
| **Tas (Surface temperature)** | Tas represents the temperature of the Earth's surface, typically measured in degrees Celsius or Fahrenheit, and plays a critical role in climate assessments and studies. |
| **UNFCCC (United Nations Framework Convention on Climate Change)** | UNFCCC is an international treaty aimed at addressing climate change and its impacts. It provides the framework for annual climate conferences and the Kyoto Protocol, among other international climate agreements. |
| **IPCC-AR6-WG-II (Working Group – II):** | Working Group II of the IPCC AR6 focuses on assessing the impacts, adaptation, and vulnerabilities related to climate change, providing essential information on the consequences of climate change for society and ecosystems. |

# Appendix - A

## Climate Model

A global climate model (GCM) is a complex mathematical representation of the major climate system components (atmosphere, land surface, ocean, and sea ice), and their interactions. Earth's energy balance between the four components is the key to long-term climate prediction. The main climate system components treated in a climate model are:

- The atmospheric component, which simulates clouds and aerosols, and plays a large role in the transport of heat and water around the globe.
- The land surface component, which simulates surface characteristics such as vegetation, snow cover, soil water, rivers, and carbon storing.
- The ocean component, which simulates current movement and mixing, and biogeochemistry, since the ocean is the dominant reservoir of heat and carbon in the climate system.
- The sea ice component, which modulates solar radiation absorption and air-sea heat and water exchanges.

**Fundamentals of Climate Models**

Climate models, significant in their complexity and purpose, encapsulate nearly 18,000 pages of computer code, built and refined over years by hundreds of scientists. They range from regional to global, encompassing the dynamics of Earth's atmosphere, oceans, land, and ice-covered areas.

At their core, these models are based on fundamental physical, chemical, and biological laws governing Earth's mechanisms. These laws—such as the law of conservation of energy and the Stefan-Boltzmann Law—underpin the models' equations and principles. For instance, they abide by fundamental physical laws like the law of conservation of energy and describe phenomena such as the relationship between air temperature and water vapor pressure.

Equations governing fluid motion, like the Navier-Stokes equations, are pivotal in describing the behaviors of gases in the atmosphere and water in the ocean. These principles, translated into computer code, culminate in millions of lines forming a climate model.

- **Spatial Resolution and Modeling Process:**
    1. Climate models adopt a grid-based approach to simulate the Earth, dividing it into numerous grid cells, typically around 100km in longitude and latitude. These grids allow scientists to represent various climate processes within these defined cells. For smaller-scale processes, like convection, parameterizations—approximations—are employed to fill these gaps, simplifying and including these processes in the model.
    2. Climate models divide the globe into a three-dimensional grid of cells representing specific geographic locations and elevations. Each of the components (atmosphere, land surface, ocean, and sea ice) has equations calculated on the global grid for a set of climate variables such as temperature. In addition to model components computing how they are changing over time, the different parts exchange fluxes of heat, water, and momentum. They interact with one another as a coupled system.
    3. The model's spatial resolution determines the size and number of these grid cells. A higher resolution means more detailed regional climate information but requires significantly more computing power to execute.
- **Temporal Resolution and Model Calculations:**
    1. Models work on a time-step basis, dividing time into manageable chunks to calculate the state of the climate system. A smaller time step yields more detailed climate information but demands additional calculations, affecting the model's overall processing speed.
    2. Both spatial and temporal resolutions necessitate a compromise: higher resolutions result in more detailed outputs but require considerably more computational resources and time for modelling.

**Evolution of CMIPs**

Climate models serve as a critical tool for understanding past and future climate changes, simulating various Earth system components, and necessitating the use of some of the world's largest supercomputers. These models are constantly evolving, with different modeling groups worldwide incorporating higher spatial resolution, new physical processes, and biogeochemical cycles. These coordinated efforts are part of the Coupled Model Intercomparison Projects (CMIP).

The 2021 IPCC sixth assessment report (AR6) features the latest state-of-the-art CMIP6 models, representing a significant leap in climate modeling.

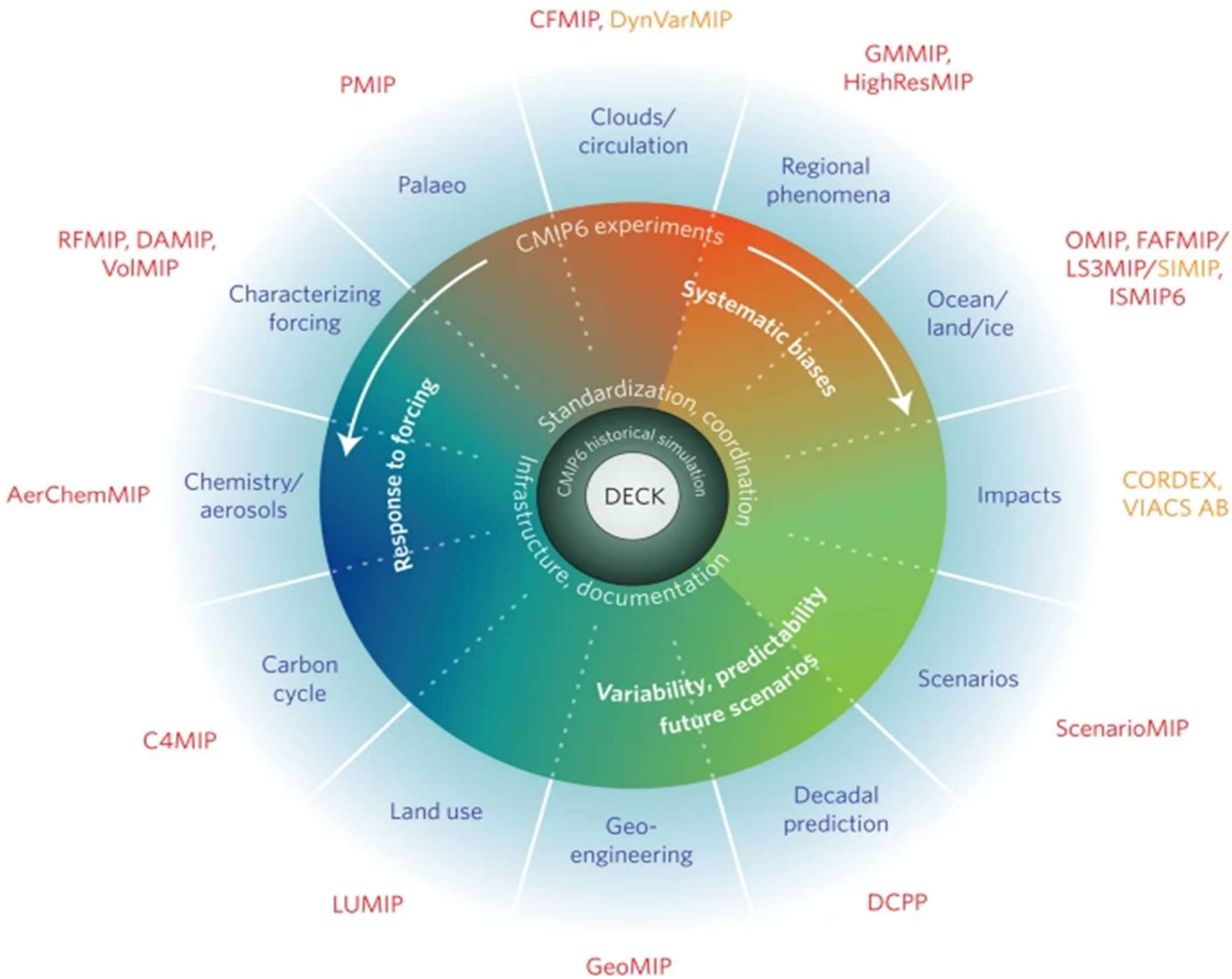
- **CMIP6 Overview:**
  CMIP6 involves "runs" from approximately 100 distinct climate models produced across 49 different modeling groups. These models exhibit higher sensitivity, leading to projections of greater warming this century, around 0.4°C warmer than similar scenarios in CMIP5. CMIP6 aims to address a growing range of scientific questions by introducing common experiments, standards, coordination, and infrastructure, facilitating the distribution of model outputs and enhancing the character of the model ensemble.
- **Integration of IAMs in CMIP6:**
  Integrated assessment models (IAMs) play a pivotal role in combining physical and economic analyses to develop and assess climate change policies. IAMs couple simplified climate and economic models to simulate the global economic impacts of climate change under various mitigation scenarios. These models inform domestic and international climate change policy, with a significant focus on three IAMs: DICE, FUND, and PAGE. IAMs provide a bridge between economists and scientists to resolve policy disagreements.
    1. ***Transparency and Assumptions in IAMs:***
       IAMs incorporate simplified representations of the climate system compared to global climate models used by climate scientists. Transparency is essential, and underlying assumptions and model inputs must be made explicit. Equations driving IAMs need to be specified to understand the mechanisms behind model projections and avoid the impression that IAMs are 'black boxes'.
    2. ***Challenges in Climate Model Robustness and Consistency:***
       Ensuring model robustness and consistency between models is a longstanding challenge for climate researchers. CMIP's major goal is to assess and improve the performance and reliability of global coupled ocean–atmosphere general circulation models used to predict future climate under various emissions scenarios.

### The DECK and CMIP Historical Simulations

Two types of processes within climate models are used: simulated and parameterized. Simulated processes are larger than grid-scale and based on bedrock scientific principles (conservation of energy, mass, and momentum). An example of a simulated process represents tropical cyclones and storm activity. Parameterized processes represent more complex processes that are smaller than grid-scale (so, they cannot be physically represented) using simpler processes. Their formulations are guided by fundamental physical principles but also make use of observational data. An example of a parameterized process represents cloud and aerosol composition.

CMIP6 features essential "diagnostic" simulations called DECK, including CO2 increases, abrupt quadrupling, and unchanged climate forcings for extended periods. These simulations also encompass historical runs driven by observed CO2 and other climate forcings, as well as future emissions scenarios. Notably, the model surface temperature "hindcast" and projections of future warming under different emission scenarios are examined.

- ***Importance of Hindcasts:***
  Hindcasts serve as valuable tools for assessing the performance of climate models. Accurate representation of past changes builds confidence in the models' ability to predict future changes.
- ***Model Tuning and Model Projections:***
  Climate models can't resolve all small-scale physics, leading to model tuning – a choice of values for processes that occur at too small a scale to simulate effectively. While most models avoid explicit tuning to match past temperature changes, discrepancies may prompt adjustments.
- ***DECK Experiments and CMIP Historical Simulations:***
  The DECK experiments provide continuity and are well-suited for quantifying and understanding important climate change response characteristics.




*Adapted from Eyring, V. et al. Geosci. Model Dev.* ***9****, 1937–1958 (2016).*

**CMIP6 MIPs**

Additionally, CMIP6 offers 22 specialized Model Intercomparison Projects (MIPs) to assess climate changes beyond basic diagnostics and historical simulations. CMIP6 incorporates various MIPs to assess climate changes in specialized areas, including aerosols and chemistry, carbon cycle, land use, and more.

The MIPs included in CMIP6 are shown in the figure below:

- The Aerosols and Chemistry Model Intercomparison Project (AerChemMIP)
- Coupled Climate Carbon Cycle Model Intercomparison Project (C4MIP)
- The Carbon Dioxide Removal Model Intercomparison Project (CDRMIP)

- Cloud Feedback Model Intercomparison Project (CFMIP)
- Detection and Attribution Model Intercomparison Project (DAMIP)
- Decadal Climate Prediction Project (DCPP)
- Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP)
- Geoengineering Model Intercomparison Project (GeoMIP)
- Global Monsoons Model Intercomparison Project (GMMIP)
- High-Resolution Model Intercomparison Project (HighResMIP)
- Ice Sheet Model Intercomparison Project for CMIP6 (ISMIP6)
- Land Surface, Snow and Soil Moisture (LS3MIP)
- Land-Use Model Intercomparison Project (LUMIP)
- Ocean Model Intercomparison Project (OMIP)
- Polar Amplification Model Intercomparison Project (PAMIP)
- Palaeoclimate Modelling Intercomparison Project (PMIP)
- Radiative Forcing Model Intercomparison Project (RFMIP)
- Scenario Model Intercomparison Project (ScenarioMIP)
- Volcanic Forcings Model Intercomparison Project (VolMIP)
- Coordinated Regional Climate Downscaling Experiment (CORDEX)
- Dynamics and Variability Model Intercomparison Project (DynVarMIP)
- Sea Ice Model Intercomparison Project (SIMIP)
- Vulnerability, Impacts, Adaptation and Climate Services Advisory Board (VIACS AB)

**Future Scenarios in CMIP6**

CMIP6 presents a significant expansion compared to CMIP5, featuring more modeling groups, a wider range of scenarios, and diverse experiments. The goal is to generate standard simulations, including future climate scenarios, which provide common greenhouse gas concentrations and climate forcings to project future changes.

- ***Shared Socioeconomic Pathways (SSPs):***

  In the lead-up to IPCC AR6, the energy modeling community introduced Shared Socioeconomic Pathways (SSPs) driven by different socioeconomic assumptions. SSPs offer distinct end-of-century climate change outcomes, allowing for a range of climate change scenarios.

- ***CMIP6 Scenarios:***

  CMIP6 introduces updated scenarios similar to CMIP5 RCPs, including SSP1-2.6, SSP2-4.5, SSP4-6.0, and SSP5-8.5. Additionally, CMIP6 adds scenarios like SSP3-7.0, SSP4-3.4, SSP5-3.4OS, and SSP1-1.9 to explore various climate change outcomes under different policy scenarios.

- ***CMIP6 Scenario Improvements:***

1. CMIP6's scenarios offer a better exploration of "no climate policy" outcomes, improving upon the limited baseline scenarios in CMIP5. These new scenarios allow climate models to explore changes and impacts at different levels of warming, including 1.5°C.
2. CMIP6 scenarios maintain end-of-century forcing similar to CMIP5 scenarios but feature different emissions pathways and CO2 and non-CO2 emissions mix.

# Appendix -2

## Machine Learning & Climate Change Science

Machine learning (ML) and artificial intelligence (AI) have seen rapid integration into numerous aspects of daily life, largely due to advancements in processor availability, connectivity, and the proliferation of big data. The impacts of these technologies are apparent across various sectors, from healthcare to transportation, internet interactions, food supply systems, and national security. As society moves closer to embracing self-driving cars, personalized medical diagnostics, speech recognition, and tailored consumer recommendations, the need to intertwine these technological advancements with the critical challenges posed by climate change becomes increasingly evident.

Climate change is a multifaceted global issue that demands holistic responses, necessitating the incorporation of ML and AI into climate science. While ML algorithms, particularly neural networks, have existed for decades, their application in the realm of climate change has been hindered by computational limitations.

The terminology surrounding these computational methods—big data, ML, and AI—encompasses diverse methodologies. Big data deals with handling complex datasets beyond the scale manageable by traditional analytical techniques. AI, a subset of computer science, focuses on training computers to execute tasks that surpass human capabilities, involving decision-making in various contexts. ML is a subset of AI that enables computers to learn from large datasets, refining or discovering linkages between different data points, such as in meteorological measurements or Earth System models.

In the context of climate and weather applications, these can be summarized as:

1. **Big Data:** Concerned with the collection and analysis of meteorological and Earth System-related measurements and high-resolution model outputs.

2. **ML:** Focuses on establishing new connections and refining existing linkages within datasets. For example, it could identify how sea surface temperature features influence weather predictions over land regions months later.

3. **AI:** Leverages the connections discovered by ML to issue automated warnings and advice to societies regarding imminent weather extremes.

The recent surge in the application of ML methods has been facilitated by enhanced computational capabilities, especially through the innovative use of graphical processing units (GPUs) that offer higher processing speeds compared to standard central processing units. Researchers also suggest the utilization of computer memory to enhance the efficiency of calculations and bring them closer to the data storage locations.

Numerical weather forecasting has seen significant advancements since the 1950s. Until recently, computational limitations necessitated solving equations on a coarse spatial grid. However, the growth in computing power has enabled ultra-fine-resolution weather forecasting models, providing grids with nearly kilometer-scale resolutions. Despite some limitations, these finer grids allow for explicit calculations of storm tracks, mesoscale cloud systems, and deep convective events.

Earth System Models (ESMs), similar to weather forecast models, predict climate change by simulating the interactions between atmospheric greenhouse gases and radiative fluxes. These models operate over centuries and encompass detailed descriptions of ocean circulations and polar ice extents. However, computational constraints prevent ESMs from operating at the ultra-high resolutions of weather forecasts. As a result, they still rely on parameterizations for crucial sub-grid processes like convection.

Around 20 research centers have developed ESMs, leading to the establishment of shared databases like the Coupled Model Intercomparison Project Phase 5 (CMIP5) and Phase 6 (CMIP6). While these models reflect two decades of ESM development, discrepancies between them persist, impacting fundamental statistics like equilibrium climate sensitivity and posing challenges for climate adaptation planning and setting target thresholds for gas concentrations.

**Dimension Reductionality**

Dimension reduction in mathematical modelling aims to explain observed phenomena by governing equations. These equations, often partial differential equations, are continuous in space and time, coupled through various terms in the climate system. However, challenges persist in reducing the dimensions to highlight dominant processes and connections within a complex system.

The three historical approaches to dimension reduction include nondimensionalization, dimensional analysis, and statistical techniques. While these methods offer valuable insights into dominant climate parameters, they face constraints such as the complexity of climate equations and the need for foreknowledge to identify related quantities.

A more recent technique, emergent constraints (ECs), aims to reduce inter-ESM discrepancies to refine projections. This involves searching for relationships between modelled climate system quantities and measurable data, which helps constrain estimates of future variables.

**Supervised ML Algorithms:**

The integration of ML and AI techniques into dimension reduction frameworks holds promise for further discoveries in climate modeling. The numerous ML methods available, both supervised and unsupervised, offer diverse applications in climate science. Supervised methods typically map inputs to outputs and are well-suited for classification and regression problems. Unsupervised methods, on the other hand, uncover patterns and connections within data without prior assumptions.

***Existing AI applications for Climate***

Many climate researchers have adopted ML methods to advance understanding of specific Earth System components in the following table. We now argue that there is enormous potential for using ML approaches also to find the more connected behaviors between multiple Earth System components, and how they aggregate to overall climate responses.

| Component of CC Research | Findings/ Development | Standard Techniques used | Reference |
|---|---|---|---|
| Climate impacts | Determine the influence of climate drivers on sand-deposition in semi-arid regions. | Artificial Neural Networks(ANN). | Buckland et al(2019) |

| | | | |
|---|---|---|---|
| Climate impacts | Estimate crop yields from satellite data. | Convolutional Neural Network (CNN). Gaussian Process(GP) Regression. | Azzariet al(2017), Burke and Lobell (2017) |
| Climate impacts | Determine the impact of water scarcity (drought) in different climatic systems. | Model Tree Ensembles(Random Forests, RF). | Yang et al(2016) |
| Climate impacts | Predict hydrological variables(evapotranspiration) from inputs of meteorological variables(precipitation, temperature) in India. | Fuzzy logic, Least Squares Support Vector Regression (LS-SVR), Artificial Neural Networks(ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS). | Goyalet al(2014) |
| Climate impacts | Assess the impact of future climate change on hydrology in India, and including for river flow. | Principal Components Analysis (PCA) and fuzzy clustering, Relevance Vector Machine (RVM). | Ghosh and Mujumdar(2008 ) |
| Climate impacts | Assess the impact of climatic change on the global hydrological cycle, with an emphasis on changes in evapotranspiration. | Model Tree Ensemble (MTE). | Jung et al(2010) |
| Climate impacts | Assess the impact of climate change on above-ground biomass. | Support Vector Machines(SVM); Artificial Neural Networks(ANN), Generalised Regression | Wu et al(2019) |

| | | | |
|---|---|---|---|
| | | Neural Network (GRNN). | |
| Climate extremes | Forecast meteorological droughts using antecedent meteorological information in Ethiopia. | Artificial Neural Network (ANN); Support Vector Regression (SVR): Wavelet Transforms | Mishra and Desai (2006), Belayneh et al (2016) |
| Climate extremes | Predict meteorological and agricultural drought conditions from satellite data. | Random Forest (RF); Gradient Boosted Regression Trees(GBRT). | Park et al(2016) |
| Climate extremes | Predict a drought index using meteorological and climate indices as inputs. | Extreme Learning Machine & Convolutional Neural Network (CNN). | Deo and Sahin (2015) |
| Climate extremes | Identify extreme weather events in the output of a global climate model. | Convolutional Neural Network (CNN); 3D Convolutional encoder-decoder. | Liu et al(2016) |
| Climate datasets | Improving estimates of min and max temperatures for incomplete timeseries. Generate better estimates of daily maximum and minimum temperatures, based on information from other nearby measurements and where accurate time of recording is undertaken. | Gaussian Process(GP) model fitted with a Markov Chain Monte Carlo (MCMC) method. | Rischard et al(2018) |

| | | | |
|---|---|---|---|
| Climate datasets | Downscale GCM precipitation fields to scales appropriate for impact assessment. | Kernel Regression (KR). | Salviet al(2017) |

**Gaussian processes**

Gaussian processes, a supervised ML method, offer a non-parametric approach to regression. Unlike linear regression methods, Gaussian processes are defined over observation functions rather than input states, allowing for explicit representation of uncertainty and prior beliefs. These methods have significant potential for climate science, especially in making out-of-sample predictions for future climate states.

1. **Gaussian Process Regression (GPR):**

   - **Methodology:** GPR is a non-parametric, Bayesian approach to regression. It's based on the assumption that the distribution of functions is Gaussian. Rather than fitting specific parameterized functions, GPR models the distribution over functions directly.
   - **Predictions:** It uses training data to make predictions about test data by assuming that the function values are jointly Gaussian.
   - **Key Features:** GPR is advantageous because it can capture complex and non-linear relationships between variables. It's also useful when there's uncertainty in the data or when few data points are available.

2. **Exponential Gaussian Process Regression (Exponential GPR):**

   - **Methodology:** Exponential GPR is a specific variant of Gaussian Process Regression that employs exponential kernels in its Gaussian process.
   - **Kernel Function:** Instead of using a standard covariance kernel function, Exponential GPR uses an exponential kernel function. This kernel function

considers the distance between pairs of points and assigns more weight to nearby points.

- **Purpose:** Exponential GPR is particularly well-suited for modelling relationships where the output tends to decrease exponentially as the input variable changes. It's effective in capturing decay patterns or situations where relationships decay over distance or time.

**Difference between GPR and Exponential GPR:** The key distinction between GPR and Exponential GPR lies in the choice of kernel function. While GPR uses a general covariance kernel function, Exponential GPR specifically uses an exponential kernel function. This choice of kernel impacts how these models capture relationships and make predictions. Exponential GPR is specialized to model scenarios where the relationships exhibit an exponential decay pattern.

In summary, both GPR and Exponential GPR are based on Gaussian processes and are used for regression tasks. GPR is a more general approach, while Exponential GPR is a specialized form of GPR that focuses on relationships that exhibit exponential decay.

n Gaussian Process Regression (GPR), the equation involves predicting the function $f(x)$ at some input point $x$, assuming a Gaussian distribution over functions. The general equation for GPR is expressed as follows:

$$f(x) \sim GP\big(m(x), K(\varkappa, x')\big)$$

Where:

- $f(x)$ represents the function to be predicted.
- $x$ is the input variable.
- $m(x)$ is the mean function, representing the prior mean assumption about the function at input $x$.
- $k(x,x')$ is the covariance (or kernel) function, specifying the similarity or correlation between any two points $x$ and $'x'$. This function provides information about the relationships or interactions between the inputs $x$ and $x'$.

This equation illustrates that the function $f(x)$ follows a Gaussian distribution with a mean function $m(x)$ and a covariance function $k(x,x')$. The GPR method allows inferring the posterior distribution over functions after observing some data, providing predictions about $f(x)$ and the associated uncertainty at new input points.

**Rationale for Using GPR & Exponential GPR in this research:**

In nonlinear regression, a first attempt might involve fitting increasingly complex polynomials,

$$Y = f(x, a_i)$$

where Y is an observation, x is a potential predictor of Y, and ai are parameters. However, in a nonlinear system such as the climate, we might not understand the precise parametric process, as this would require consideration of all possible nonlinear functions. As a supervised ML method, Gaussian processes are an alternative to such (linear) regression approaches. A Gaussian process is a collection of random variables, Y, (data observations) such that any subset of these variables has a multivariate normal distribution. Notable is the Gaussian process is defined over the observation functions, Y, rather than input state, x. The process is specified by a mean function and a covariance matrix. Combining the Gaussian process prior with a (Gaussian) likelihood based on the data, where some data is observed and some not, produces a Gaussian posterior distribution. This method enables out-of sample predictions, y*:

$$P(y^*|x^*) = \int P(y^*|\varkappa^*, f, D)P(f|D)\, df$$

Where:

- $P(y*|x,f,D)$ is the probability of the predicted value $*y*$ given the input $x$, the distribution $f$, and the data set $D$.
- $P(y*|x,f)$ is the conditional probability of $*y*$ given $x$ and $f$.
- $P(f|D)$ is the probability of the distribution $f$ given the observed data set $D$.

where f now represents a Gaussian process and D is again an observed data set. These non-parametric approaches allow explicit representation of uncertainty and prior beliefs, and are powerful ML approaches in nonlinear regression analyses.

**Conclusion:**

As the intersection of machine learning and climate science progresses, these advanced computational techniques are expected to enhance climate models, refine projections, and aid in understanding and addressing the complexities of climate change. The application of machine learning principles to climate data opens up avenues for improved understanding, predictive accuracy, and ultimately more effective climate policy planning and decision-making.

This note provides a glimpse into the intersection of machine learning and climate science, highlighting the potential and challenges posed by these advanced computational methodologies. As technological advancements continue, the integration of ML and AI into climate science is anticipated to revolutionize our understanding of climate change and drive more effective policy interventions and adaptation strategies.