VORMADA
HTTPS://VORMADA.COM/

Predicting Atmospheric CO_2 Concentrations With Machine-Learning

Author: Walid Alzohili

2024-02

Abstract

The motivation for this report stems from the need to better understand and forecast atmospheric carbon dioxide (CO_2) concentrations, driven both by the urgent issue of climate change and by scientific interest in predictive modeling. In particular, this study explores the capabilities of different machine-learning algorithms to predict CO_2 levels from the year 2022 to 2100, based on historical emission data. Given the increasing impact of CO_2 on global warming, accurate prediction models are essential for developing effective climate policies. Machine-learning offers a promising approach due to its ability to handle large datasets that contain complex nonlinear relationships.

The study's central question investigates the precision of these algorithms in estimating CO_2 levels under diverse future emission scenarios. The scope is specifically narrowed to modeling CO_2 concentrations, with a focus on four carbon emission scenarios: two that show a general increase and two that show a decrease in emissions from the year 2021. The machine-learning models developed in this research utilize two primary input features: CO_2 emissions from the burning of fossil fuels and emissions from land usage. The study aligns its emission scenarios with those presented in the sixth assessment report of the Intergovernmental Panel on Climate Change(IPCC), ensuring relevance and comparability with established climate models.

The methodical progression of this report goes from simpler to more complex machine-learning models, based on the accuracy and computational feasibility. This hierarchical approach ensures the selection of the most effective yet computationally efficient model. The performance of these models is assessed using Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results show that the hybrid deep learning model demonstrates superior performance across all scenarios, effectively capturing the complex dynamics of CO_2 levels.

Keywords: Machine-Learning, Atmospheric CO_2 Modeling, Global temperature, Greenhouse Effect, Regression, Neural Networks, Deep-Learning

Synopsis

Background

This report explores the application of machine-learning algorithms in modeling atmospheric CO_2 concentrations, focusing on their potential to predict future levels from historical emissions data. This approach addresses the limitations of traditional methods, which often struggle with the complexity and variability of environmental data, by utilizing advanced algorithms capable of handling nonlinear interactions and dynamic variables. This research aims to enhance the understanding of global temperature changes and their impacts, contributing to more effective climate policy development.

Problem

The primary challenge addressed is the accurate modeling of atmospheric CO_2 concentrations using machine-learning algorithms. This involves dealing with the complexities of climate data and the limitations of different machine-learning techniques.

Research question

The central research question: "How accurately can machine-learning algorithms model the CO_2 concentrations in the atmosphere, based on different scenarios of future CO_2 emissions?" investigates the accuracy of machine-learning algorithms in predicting atmospheric CO_2 concentrations under various emission scenarios from the year 2022 to 2100. This is measured using MSE, MAE and RMSE.

Method

The approach involves creating and evaluating multiple machine-learning models, each employing different algorithms and techniques. These models are tested against various emission scenarios, and their performance is compared to determine the most effective method.

Result

The study finds that while linear regression shows limited success, polynomial regression proves more effective in capturing nonlinear relationships in the data, whilst the hybrid deep learning model was the best performer.

Discussion

The findings offer insights into the strengths and weaknesses of different machine-learning approaches in climate modeling. The report highlights the potential of machine-learning in environmental science, especially in understanding and predicting future atmospheric CO_2 concentrations, and global temperature change.

Contents

List of Abbreviations

CH_4 - Methane
CNN - Convolutional Neural Networks
CO_2 - Carbon Dioxide
GRU - Gated Recurrent Unit
IPCC - Intergovernmental Panel on Climate Change
KNN - K-nearest Neighbors
LSTM - Long Short-Term Memory
LUC - Land-use change
MAE - Mean Absolute Error
MLP - Multi-layer Perceptron
MSE - Mean Squared Error
RMSE - Root Mean Squared Error
RNN - Recurrent Neural Networks
SVM - Support Vector Machines

Chapter 1

Introduction

1.1 Overview

This report examines the application of machine-learning algorithms in modeling atmospheric CO_2 concentrations, a critical aspect of understanding global climate dynamics. Atmospheric CO_2, as explained by (Wallace & Hobbs 2006), plays a pivotal role in climate change due to its significant greenhouse gas properties. The effective modeling of CO_2 concentrations is crucial for predicting future climate scenarios and has been traditionally challenging due to complex interactions in the climate system. This study utilizes machine-learning, an approach increasingly recognized for its ability to manage large datasets and model complex systems. Analyzing the effectiveness of machine-learning will be done by creating, and evaluating multiple machine-learning models that attempt to predict the future CO_2 concentrations in the atmosphere from the year 2022 till the year 2100 based on historical data (Ritchie & Rose 2020).

Each machine-learning model starts by predicting the future concentration of CO_2 in the atmosphere based on various hypothetical scenarios for the amount of CO_2 emissions that are released to the environment. Next the predicted concentrations will be used to predict the future temperature change for each scenario. The future hypothetical scenarios are similar to the scenarios SSP1, SSP2, SSP3, and SSP5, which are created by the IPCC (IPCC 2022*a*).

The result of this report will provide a deeper understanding, relative to previous research, of how machine-learning can be used to model the atmospheric CO_2 concentrations. Moreover it will shed light on the algorithms that can be expected to solve this problem with good accuracy. Finally, it will provide supporting arguments to clarify the factors contributing to the success or failure of these algorithms.

1.2 Problem statement

The problem that this thesis addresses is studying the effectiveness of machine-learning algorithms in modeling future atmospheric CO_2 concentrations for the years 2022-2100. The predictions will be based on various future carbon emission scenarios. Moreover, this thesis attempts to provide a performance comparison between the machine-learning models, and the more traditional method of modeling atmospheric CO_2 concentrations, which uses a carbon box model (Tomizuka 2009).

1.3 Research question

The aim of this thesis is to answer the following research question: "How accurately can machine-learning algorithms model the CO_2 concentrations in the atmosphere, based on different scenarios of future CO_2 emissions?"

1.4 Delimitations

The delimitations are the research boundaries that we as the authors have intentionally established for our study.

- This paper aims only to model the concentrations of CO_2 in the atmosphere, and not model the full carbon-cycle.
- In this paper, we will predict only the future CO_2 concentrations in the atmosphere for the years 2022 to 2100.
- For making future predictions, we will only use 4 carbon emission scenarios of which 2 are generally increasing and 2 are generally decreasing with respect to the year 2021.
- As input for the machine-learning model we will use only carbon emissions and not emissions from other greenhouse gases, however the models will be tuned to indirectly take the effects of other greenhouse gases into account.
- In this paper we will not use machine-learning to predict the environmental impact of the predicted CO_2 concentrations.

1.5 Limitations

The limitations are the external factors that impose restrictions, and boundaries on our research, and we as the authors had little to no influence or control over.

- This paper considers the predictions of CO_2 concentrations provided by the IPCC as the ground truth (IPCC 2022*a*). This is because we do not have better metrics to evaluate future predictions. Thus, we need to trust their result to be the ground truth when evaluating the machine-learning models.
- We had a very small training dataset to train the models with, i.e., from 1750 to 2022. Having a small trainingset will undermine the strengths of some machine-learning algorithms, especially deep-learning models (Alpaydin 2020).
- Unfortunately we did not have enough time nor sufficient computational power to perform more hyper-parameter tuning on the deep-learning models, which means that with more computational power it may be possible to obtain more accurate results for the MLP model and the hybrid model.

Chapter 2

Theoretical background

2.1 The carbon-cycle

The carbon cycle is a fundamental and intricate process vital to Earth's system, profoundly influencing the climate and sustaining life. It involves the transfer of carbon among Earth's major reservoirs, including the atmosphere, oceans, biosphere, and crust (Wallace & Hobbs 2006). Within this cycle, two principal domains are distinguished based on the rate of carbon turnover: 1) the fast domain, which encompasses the atmosphere, ocean surface, terrestrial vegetation, soils, and freshwater systems characterized by rapid exchanges, 2) the slow domain, which includes large carbon deposits in rocks and deep ocean sediments, where turnover occurs over much longer timescales (Ciais et al. 2014).

The carbon cycle exerts a major influence on Earth's global temperature, mainly through greenhouse gases like carbon dioxide (CO_2) and methane (CH_4). These gases trap heat within the atmosphere, maintaining conditions necessary for life (Wallace & Hobbs 2006). An increase in atmospheric CO_2 leads to higher temperatures, enhanced rainfall, and accelerated rock weathering. This, in turn, results in greater carbon deposition on the ocean floor, though the rebalancing of the slow carbon cycle through chemical weathering processes requires hundreds of thousands of years (Riebeek 2023). Human activities, especially the burning of fossil fuels and deforestation, have dramatically increased concentrations of greenhouse gases, intensifying the natural greenhouse effect and driving global warming.

The ocean serves as a relatively quicker component of the slow carbon cycle, facilitating a steady exchange of CO_2 between the ocean and the atmosphere. Prior to industrialization, this exchange remained in balance; however, the rise in atmospheric CO_2 concentrations has shifted the ocean into a net absorber of carbon. Over the course of millennia, the ocean is expected to absorb up to 85% of the excess carbon emitted by fossil fuel combustion, although the process is inherently slow, governed by the rate of ocean circulation and mixing (Riebeek 2023).

Consequently, even if carbon emissions were significantly reduced, atmospheric CO_2 levels, and hence global temperatures would not rapidly decline. The carbon already emitted will persist within Earth's slow-moving carbon reservoirs, continuing to influence the climate system. The accumulation of excess atmospheric carbon has thus initiated a feedback loop, resulting in ongoing temperature rise and further greenhouse gas emissions.

2.2 Climate sensitivity

The phenomenon of climate sensitivity is crucial in understanding how the accumulation of greenhouse gases, predominantly CO_2, influences Earth's energy balance. When more greenhouse gases are present in the atmosphere, they trap additional thermal radiation, which would otherwise escape into space. This trapping effect leads to an energy imbalance, termed "radiative forcing". This imbalance is a critical factor in assessing global warming, as it directly correlates with the rise in Earth's temperature (Smith et al. 2021).
To quantify this temperature increase, scientists use a parameter known as the climate sensitivity parameter, symbolized as λ(lambda), which is expressed in km^2/W. This parameter illustrates the temperature change per unit of radiative forcing. Without considering feedback mechanisms, λ for CO_2 stands at $0.27\,km^2/W$. However, various feedback processes, both positive (like increased water vapor leading to more warming) and negative (like aerosols from forest fires causing cooling), significantly influence this value. The exact magnitude of λ, considering these feedbacks, remains a topic of active research and is essential for accurate climate modeling (Sherwood et al. 2020).

2.3 The carbon box model

The carbon box model consists of a system of seven differential equations Eq (2.1)-Eq (2.7). Those differential equations represent rate equations for the earth compartments seen in Table 2.1 (Tomizuka 2009).

Compartment	**Corresponding equation**
Atmosphere	2.1
Surface ocean	2.2
Intermediate ocean	2.3
Deep ocean	2.4
Sediments	2.5
Biosphere	2.6
Soil	2.7

Table 2.1: Earth compartments and the corresponding differential equations

$$\frac{dN_1}{dt} = -k_{12}N_1 + k_{21}(N_2^0 + \xi(N_2 - N_2^0)) + \gamma - f + \delta + k_{51}N_5 + k_{71}N_7 \tag{2.1}$$

$$\frac{dN_2}{dt} = k_{12}N_1 - k_{21}(N_2^0 + \xi(N_2 - N_2^0)) - k_{23}N_2 + k_{32}N_3 - k_{24}N_2 \tag{2.2}$$

$$\frac{dN_3}{dt} = k_{23}N_2 - k_{32}N_3 - k_{34}N_3 + k_{43}N_4 \tag{2.3}$$

$$\frac{dN_4}{dt} = k_{34}N_3 - k_{43}N_4 + k_{24}N_2 - k_{45}N_4 \tag{2.4}$$

$$\frac{dN_5}{dt} = k_{45}N_4 - k_{51}N_5 \tag{2.5}$$

$$\frac{dN_6}{dt} = f - k_{67}N_6 - 2\delta \tag{2.6}$$

$$\frac{dN_7}{dt} = k_{67}N_6 - k_{71}N_7 + \delta \tag{2.7}$$

The term k_{ij} represents the transfer coefficient from compartment i to compartment j. The preindustrial transfer coefficients have the values in Table 2.2:

$k_{12} = 60/615$	$k_{21} = 60/842$	$k_{23} = 9/842$
$k_{24} = 43/842$	$k_{32} = 52/9744$	$k_{34} = 162/9744$
$k_{43} = 205/26280$	$k_{45} = 0.2/26280$	$k_{51} = 0.2/90000000$
$k_{67} = 62/731$	$k_{71} = 62/1238$	

Table 2.2: Preindustrial transfer coefficients (Tomizuka 2009)

The term ξ(xi) is a function that depends on atmospheric CO_2 concentration P and can be calculated according to Eq (2.8):

$$\xi(P) = 3.69 + 1.86 \times 10^{-2}P - 1.80 \times 10^{-6}P^2 \tag{2.8}$$

N_2^0 is the preindustrial content of carbon in the sea surface and has the value 842 PgC
The term γ(gamma) is the carbon emission in PgC due to fossil fuel burning.
The term δ(delta) is the carbon emission in PgC due to land usage.
The term f is the net CO_2 flux to the biosphere and is calculated according to Eq (2.9):

$$f = f_0\left(1 + \beta \ln\left(\frac{P}{P_0}\right)\right) \tag{2.9}$$

Where:

- f_0 = 62 PgC/year is the preindustrial flux
- P_0: preindustrial atmospheric concentrations of CO_2 (289 ppm)
- P: current atmospheric concentrations of CO_2
- β(beta): the fertilization factor with a default value of $\beta = 0.42$

Solving the box model gives N_1 from the first compartment, which is the content of carbon in the atmosphere for each year. To get the concentration of CO_2 we divide by 2.13 (Cornell et al. 2012).

2.4 Runge-Kutta methods

Runge-Kutta methods is a family of numerical methods that are used to solve first-order differential equations, Runge-Kutta 4 is the most used algorithm in the family and is presented (Hairer & Wanner 2015).

Given a first-order differential equation $y' = f(t, y)$ and the initial values t_0, y_0, we can use Runge-Kutta 4 to solve the equation in the range $[a, b]$ as follows:

1. Choose N which is the number of points to solve for (more points generally lead to more accurate results).
2. Generate a vector t with all the numbers to solve for. This is done by generating N evenly spaced numbers within the range $[a, b]$.
3. Calculate the step size h using the equation $h = \frac{b-a}{N}$.
4. For each number in t we can calculate the solution as follows:
   - We define k_1, k_2, k_3, k_4 as follows:
     $k_1 = h \times f(t_n, y_n)$
     $k_2 = h \times f(t_n + h/2, y_n + k_1/2)$
     $k_3 = h \times f(t_n + h/2, y_n + k_2/2)$
     $k_4 = h \times f(t_n + h, y_n + k_3)$
     Where:
     - k_1: the derivative (slope) at the beginning of the interval.
     - k_2: the derivative (slope) at the midpoint of the interval (using y and k_1).
     - k_3: the derivative (slope) at the midpoint of the interval (using y and k_2).
     - k_4: the derivative (slope) at the end of the interval.
   - We calculate the next value of the solution y at time t_{n+1} using the weighted average of k_1, k_2, k_3, k_4:
     $y_{n+1} = y_n + 1/6(k_1 + 2k_2 + 2k_3 + k_4)$
     Note: Since k_2 and k_3 have double the accuracy of k_1 and k_4, we multiply them by 2.

It is worth noting that Runge-Kutta 4 can be expanded to solve systems of differential equations, this can for example be implemented in code by treating the system of differential equations as a single function that returns the solution to each of the differential equations. Moreover, because higher-order differential equations can be converted to first-order differential equations, Runge-Kutta 4 can be used to solve higher-order differential equations and systems of higher-order differential equations.

2.5 Machine-Learning algorithms

2.5.1 Linear regression

Linear regression is a method in statistics and machine-learning for predicting a numerical outcome from numerical attributes (Géron 2022). It creates a linear function to estimate

a dependent variable from one or more independent variables. Essentially, it predicts an outcome as a combination of input features, each multiplied by a weight, plus a constant called the bias term. Training this model involves optimizing the weights to minimize the difference between the predicted and actual values, typically using the mean-squared error as the measure of this difference (Witten & Frank 2005).

2.5.2 Polynomial regression

Polynomial regression handles nonlinear relationships between variables. It expands on linear regression by adding powers of each feature as new features, allowing for modeling a wider range of data structures. This method can approximate any continuous function within a certain interval, reflecting its flexibility in modeling complex relationships in datasets (Géron 2022). Polynomial regression is powerful in predictive analytics and machine-learning and is a critical tool for capturing intricate relationships that linear models cannot, offering advanced solutions for modeling nonlinear patterns in datasets.

2.5.3 K-Nearest Neighbors

The K-Nearest Neighbors (KNN) is a simple machine-learning algorithm used for pattern recognition for both classification and regression tasks (Géron 2022). As a non-parametric, supervised-learning method, KNN distinguishes itself by depending on the proximity of data points. In regression contexts, it predicts a continuous value by averaging the values of the KNN. Due to its versatility KNN can proficiently manage various data types, making it suitable within multiple domains, such as recommendation systems and financial analysis.

2.5.4 Support Vector Machine (SVM)

The Support Vector Machine (SVM) Regressor represents an evolved form of the traditional SVM classifier, adapted for regression tasks. Unlike its classification counterpart the SVM regressor aims to fit as many instances as possible within one "street", while also minimizing margin violations. The width of this street is controlled by a hyper-parameter ϵ, where smaller values lead to more support vectors and increased model regularization. This model is noted for its ϵ-insensitivity, meaning that adding more training instances within the margin does not affect the model's predictions (Géron 2022). For nonlinear regression tasks, a kernelized SVM model, such as one using a second-degree polynomial kernel, is suitable. This approach allows SVM to perform regression on nonlinear datasets.

2.5.5 Decision Tree Regression

The Decision Tree Regressor stands out for its ability to adeptly handle complex datasets (Géron 2022). Unlike classification trees that categorize data into discrete classes, this regressor predicts continuous values, making it ideal for nuanced data modeling. Its strength lies in effectively capturing intricate, nonlinear relationships within datasets.

However, a notable challenge it faces is over-fitting, especially with diverse datasets. This necessitates the use of regularization techniques to maintain model reliability when predicting on new unseen data . The regressor utilizes the Classification and Regression Tree (CART) algorithm, which splits the training data into subsets through a recursive process, aiming to minimize the error. In regression contexts, the decision tree algorithm focuses on minimizing variance rather than impurity.

2.5.6 Multi-Layer Perceptron (MLP)

The Multi-layer Perceptron (MLP) is a fundamental structure in artificial neural networks, characterized by its layered composition. An MLP typically consists of an input layer, multiple hidden layers, and an output layer. Each layer is made up of neurons, which process inputs and pass on their output to the next layer. The neurons in the input layer receive external data, the hidden layers perform computations and transformations on this data, and the output layer produces the final outcome of the network (Géron 2022).

MLPs leverage backpropagation for training, which allows the network to adjust its weights and biases based on the errors in its output. This process involves calculating the gradient of the error with respect to each weight and bias in the network and adjusting them to minimize the error.

The architecture of an MLP, with its ability to perform nonlinear processing through multiple layers, enables it to solve complex problems that simpler models cannot. This complexity and versatility make MLPs a crucial component in the field of deep-learning and artificial intelligence (Géron 2022).

2.5.7 Long Short-Term Memory (LSTM)

The LSTM (Long Short-Term Memory) model was developed to address the vanishing and exploding gradient issues prevalent in Recurrent Neural Networks (RNNs). Its intricate architecture, featuring memory cells with nodes, sigmoidal gateways, and typically tanh activation functions, is central to its ability to effectively manage data flow and transformation. This design allows LSTMs to balance long-term and short-term memory by selectively retaining essential information while discarding irrelevant data.

Each cell in the LSTM processes inputs by incorporating the previous cell state and hidden state, along with the current input, directing this information through various gateways and activation functions. The Sigmoid nodes in these gateways either permit or restrict data flow, governed by values ranging from zero (blocking flow) to one (allowing flow), with each gateway—forget gate, input gate, and output gate serving a distinct function in regulating the information flow (Hochreiter & Schmidhuber 1997).

The cell's output includes both the regular output and new cell and hidden states, which then proceed to the next memory cell in the sequence. Thanks to this sophisticated structure, LSTMs have become preeminent in various domains like sequence prediction, natural language processing, and speech recognition, exhibiting remarkable results and an enhanced capacity for bidirectional processing in tasks such as machine translation and sentiment analysis (Brownlee 2017).

2.5.8 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU), a variant within recurrent neural network architectures, stands out for its effectiveness in processing sequential data, making it highly relevant in areas like language modeling and speech recognition. The GRU is tailored to adaptively capture dependencies across varying time scales, addressing the limitations seen in traditional RNNs, especially regarding parameter efficiency and computational speed.

At the core of the GRU's mechanism are two gates, the update gate and the reset gate. The

update gate plays a crucial role in determining how much information from previous time steps is carried forward to future states. It achieves a balance between retaining relevant past information and incorporating new data. Simultaneously, the reset gate modulates the influence of past information, allowing the model to 'forget' or discard data that is considered irrelevant. The current memory content is a synthesis of the present input and the selectively retained past information, regulated through the reset gate. This culminates in the final memory at the current time step.

The GRU's streamlined design, with fewer gates and parameters than the Long Short-Term Memory (LSTM) model, yet maintaining a robust ability to capture temporal dependencies, offers a compelling alternative to traditional RNNs. Its efficiency in handling sequence data and managing long-term dependencies, while addressing the vanishing gradient problem, makes it particularly effective in tasks where sequence data is predominant (Dey & Salem 2017).

2.5.9 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are specialized neural networks optimized for processing grid-like data, such as images and time-series. They differ from traditional networks by employing convolution—a linear operation emphasizing recent measurements for a more representative output—instead of general matrix multiplication (Goodfellow et al. 2016). CNNs are characterized by features like sparse interactions, reducing computational load, and parameter sharing, which lowers complexity and improves learning efficiency. They also employ equivariant representations, allowing for effective feature detection irrespective of precise locations.

A CNN's architecture typically consists of three stages in each layer: performing convolutions for linear activations, applying nonlinear activation functions for detection, and pooling to introduce input translation invariance. This design enables CNNs to effectively capture complex data patterns.

2.6 Research gap

While existing literature on CO_2 concentration prediction has made significant strides in applying various machine-learning models to environmental research, there remains some gaps that this study aims to address. Firstly, previous studies often limit their focus to specific machine-learning models or scenarios, neglecting a comprehensive comparison across a range of predictive algorithms under diverse emission scenarios. Additionally, there is a lack of integration between different modeling approaches, such as combining traditional carbon-cycle models with advanced machine-learning techniques.

This research aims to fill these gaps by not only comparing a wide spectrum of machine-learning models, from simple linear regressions to complex hybrid deep learning models, but also by evaluating their performance across varied future emission scenarios. This comprehensive approach is designed to enhance the predictive accuracy and reliability of CO_2 concentration forecasts, contributing to the field of environmental modeling and aiding in better informed climate policy making.

Chapter 3

Method

3.1 Choice of research method

Our choice of employing machine-learning to model the atmospheric CO_2 concentrations is inspired by the historical success that machine-learning has established in similar problems, such as weather forecasting (Hewage et al. 2020). Moreover, our motivation to use shallow machine-learning, and deep-learning approaches for this problem originates from comparatively little exploration that they have received form other researchers.

To answer the research question we have implemented 7 different machine-learning models, which can be seen in Table 3.2. This approach is good because it allows us to have a robust understanding of how machine-learning models that operating in distinct ways perform on the problem of modeling the atmospheric CO_2 concentrations.

For training the model we have used CO_2 emissions from both the burning of fossil fuels, and from land usage. Training the machine-learning models on this data is a good choice because these two emission sources have the most effect on the atmospheric CO_2 concentrations, moreover these two sources are the same ones that were used as input to the 7-compartment box model created by Akira (Tomizuka 2009). Thus, using this data will allow us to conduct a more accurate comparison between the machine-learning models and the 7-compartment carbon box model.

We have used 4 different future scenarios to make predictions on, those scenarios are similar to the scenarios presented in the IPCC report (IPCC 2022*a*). Making predictions on those scenarios is advantageous because it allows us to compare the results of our models with the results that the IPCC presents in its report.

3.2 Application of research method

3.2.1 Data preparation

The machine-learning models used in this paper take two features as input 1) CO_2 emission from the burning of fossil fuel and 2) CO_2 emissions from land usage. These data were selected from the "CO2 emissions" dataset (Ritchie & Rose 2020). The data was selected using only the rows "world" since we do not need to know how much each country or region is emitting. The historical emission data is visualized in Figure 3.1.

In the dataset, the emissions from land usage are available only from the year 1850, but since we have emission data from the burning of fossil fuel going back to 1750, we have generated emission data for land usage that are linearly increasing from the 1750 till the year 1850. Doing so will allow us to have 100 more training samples to train the models with.

As the target variables for training the machine-learning models, we have used the historic data for CO_2 concentration in the atmosphere. The historic CO_2 concentrations are visualized in Figure 3.2.

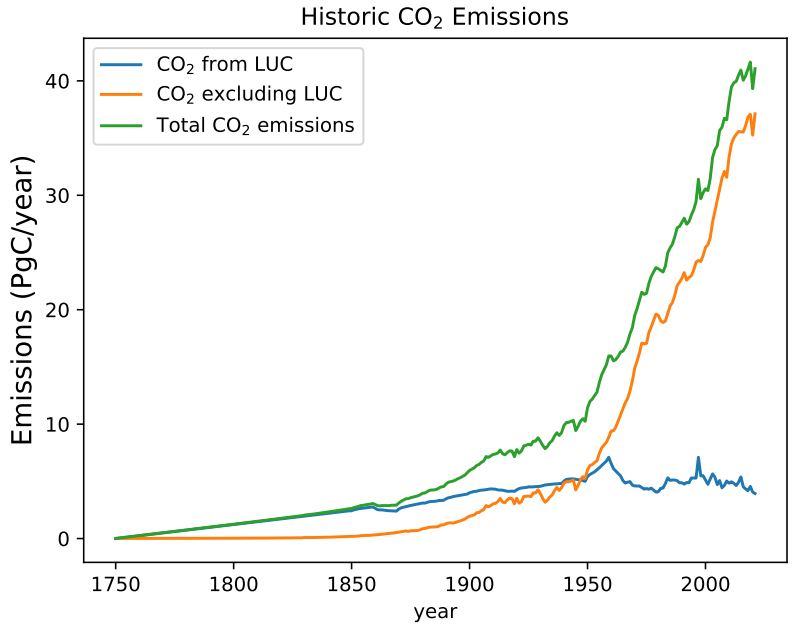

Figure 3.1: Historic CO_2 emission data for years 1750 to 2021

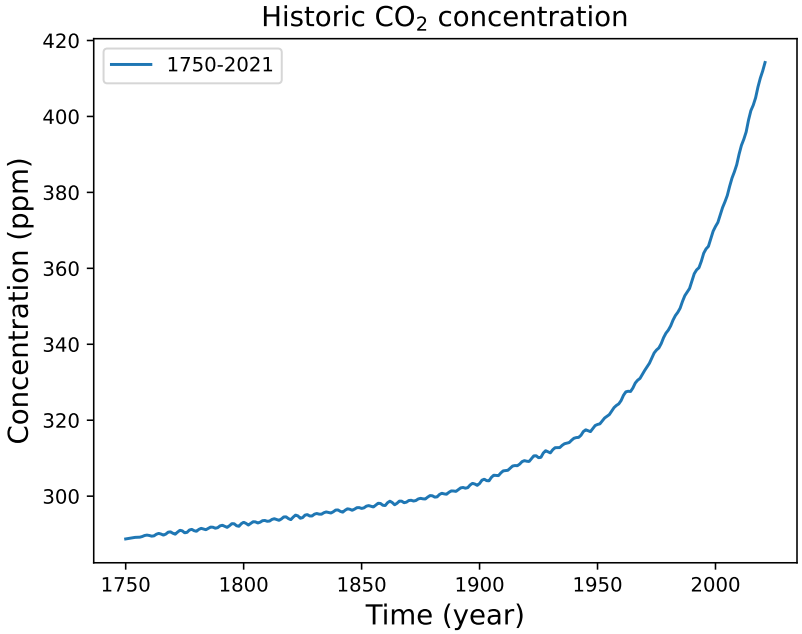

Figure 3.2: Historic CO_2 concentrations in the atmosphere for for years 1750 to 2021

The input features were normalized before using them to train the machine-learning models. This is because many machine-learning models work better when trained on normalized data. For this paper we have used Z-score normalization (Singh & Singh 2020).

Machine-learning models require splitting the data into a training-set, and a test-set. For all of the models that were made, we have used 80% of the data for the training set, and 20% for the test set. 20% of the training set was afterwards extracted as a validation set. The training set was used to train the model, and the test-set was used to evaluate the models' generalization performance (the models' performance on unseen data). Since the data is a time-series data, and our goal is to make future predictions, the splitting was done in chronological order where the test-set is chosen from the latest years in the dataset.

3.2.2 Generating future scenarios

One of the objectives of this report is to predict the future atmospheric CO_2 concentrations for different future scenarios, thus these scenarios have to be determined. We chose to generate scenarios similar to the scenarios created in the 2022 IPCC summary report (IPCC 2022*a*). Those scenarios cover the years 2021 to 2100. Table 3.1 shows the scenarios from the IPCC report and their equivalent in this report.

Figure 3.3 shows four plots, one for each future scenario. From the plots we see that for all scenarios, the future CO_2 emissions from land usage do not change from current emission levels. However for the first two scenarios we have CO_2 emissions that are mostly decreasing, though at different rates, with scenario 1 decreasing at a faster rate, to reach a total emission that is negative by the year 2100.
For the second two scenarios we have CO_2 emissions that are mostly increasing, but at different rates. Scenario 3 is increasing linearly with the same rate as historical data, and scenario 4 is increasing linearly, but at a rate that is faster than the historical data to ultimately reach 140 PgC/year by the year 2100.

IPCC scenario	**Equivalent scenario in this report**
SSP1-2.6	Scenario 1
SSP2-4.5	Scenario 2
SSP3-7.0	Scenario 3
SSP5-8.5	Scenario 4

Table 3.1: IPCC future emission scenarios and their equivalent in this report

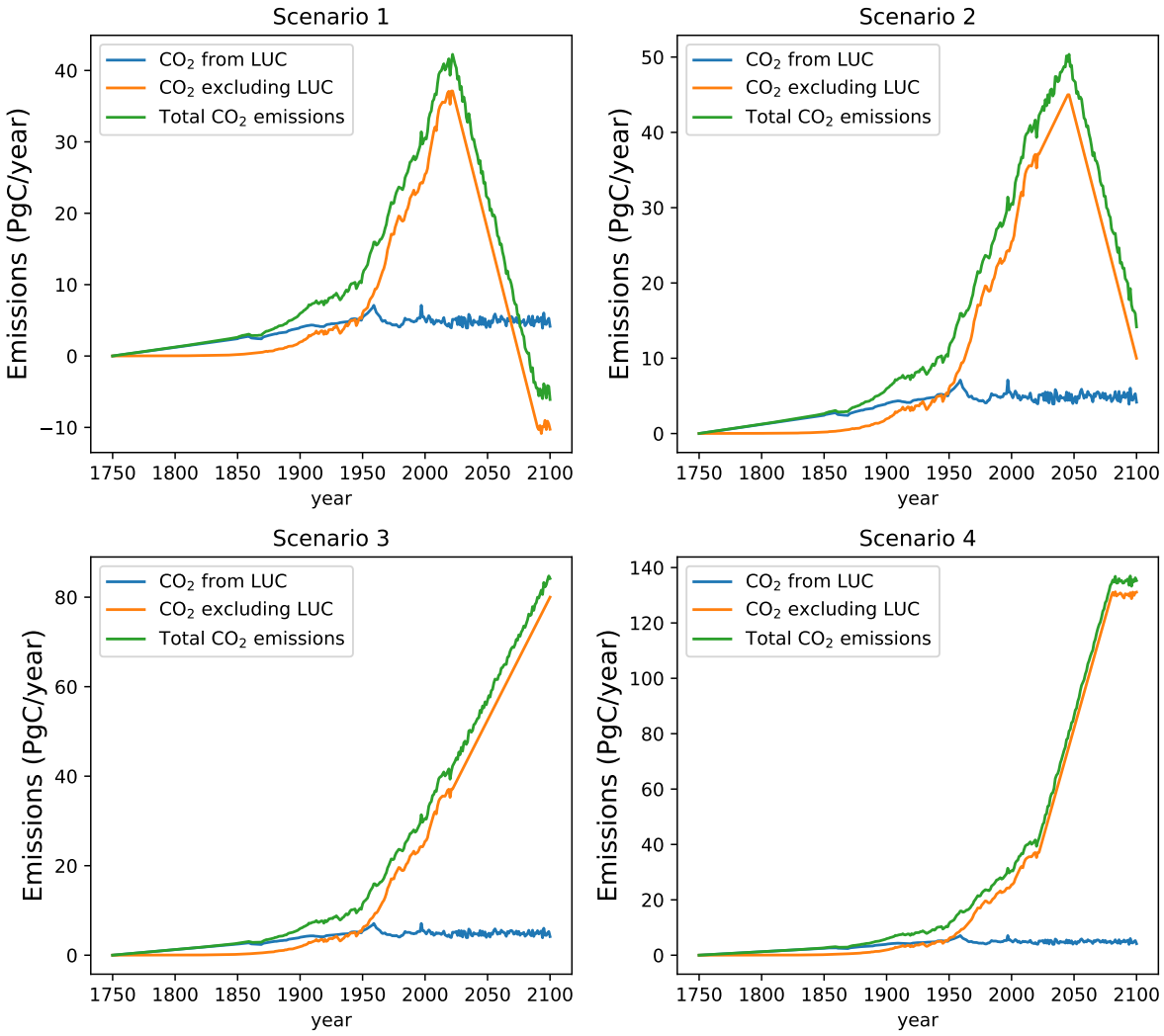

Figure 3.3: Future CO_2 scenarios

3.2.3 Modeling the CO_2 concentrations with the carbon box model

Before we can use the box model to predict future CO_2 concentration, we need to make sure that the box model works properly, moreover we may need to fine tune the free parameters of the model (in this case β). Thus we solve the system from the year 1750 to 2021 first. Note: the parameter β is a free parameter that was tuned by testing it with different values till the concentration of CO_2 given by the model matches the current concentration. The value of β has to be around 0.42 (Tomizuka 2009). The concentration of CO_2 should be 414 ppm in 2021 (Lindsey 2023), and after experimenting with different values for β we found that the value $\beta = 0.568$ gives the concentration 414 ppm. This paper uses numerical methods to solve the system of differential equations of the box model, specifically the paper uses the Runge-Kutta method.

3.2.4 Strategy

We anticipated from the beginning that our first predictive model would not be accurate enough. Thus, we planned to create multiple models starting from the simplest, and gradually switch to more complex models in case sufficiently good performance was not reached. Starting with simple models is advantageous, since simple models with good accuracy are preferred over complex models with similar accuracy. This is because simple models are

generally easier to create, easier to understand, have output that is more explainable, and are less computationally demanding to train and make predictions with (Stoffi et al. 2022). The Table 3.2 shows the models that we have created in the correct chronological order.

Some predictive models require hyper-parameter tuning to work properly. One can perform hyper-parameter tuning manually or use automatic hyper-parameter tuning, in this project we have used both techniques. For models that are not very computationally expensive such as SVMs we have performed grid search to automatically find the best hyper parameters. For the MLP model we have used random hyper-parameter search, which means that we define a range for each hyper-parameter, and then we create and train x-number of models, each time randomly choosing values of the hyper parameters from the predefined range. For the hybrid deep-learning model we have used a manual hyper-parameter tuning. This is because this model is more computational demanding, thus running automatic hyper-parameter tuning would take more time than we had available.

Since we need to choose the best predictive model from the ones we made, we have to define a cross-validation metric, in this case we have used the mean absolute error. This means that we perform the hyper-parameter tuning, and at the end we choose to keep the model that has the best cross-validation metric.

After finishing the process of hyper-parameter tuning, the models were retrained on the full dataset (training and test sets). This is because after choosing the final model, there is no longer need to have a test-set, thus we can train the model with the samples in the test-set to further improve the predictive model. Since machine-learning models generally benefit from more training data, since more data, among other things, increase the models' ability to generalize to unseen data (Alpaydin 2020). The data processing, the implementation of machine-learning models, and the data visualization, have been done using the Python programming language (Python-Software-Foundation 2023).

Order	**Predictive model**
1	Linear regression
2	Polynomial regression
3	K nearest neighbor (KNN)
4	Support vector machine (SVM)
5	Decision trees
6	Multi-layer perceptron (MLP)
7	Hybrid deep-learning model using CNN, LSTM, and GRU

Table 3.2: Predictive models in chronological order

3.2.5 Calculating the change in global temperature based on the predicted CO_2 concentrations

To predict the change in global temperature we can use theory of climate sensitivity. First we need to calculate the Radiative forcing for CO_2 using Equation 3.1 (Boucher et al. 2018).

$$\Delta F = 5.35 \times \ln\left(\frac{P}{P_0}\right) \tag{3.1}$$

Now, it is possible to calculate the increase of temperature ΔT by multiplying the Radiative forcing with the climate sensitivity parameter λ as seen in Equation 3.2.

$$\Delta T = \lambda \Delta F \tag{3.2}$$

Without feedback mechanisms $\lambda = 0.27$ km^2/W, however, feedback mechanisms are important to get accurate results, thus the value of λ was found by experiments. We know that the temperature has increased by 1.1° between the years 1850-2021 (IPCC 2022*b*). Thus, we have experimented with different values of λ till we got a temperature increase of 1.1 degree using the value $\lambda = 0.62$.

3.2.6 Performance evaluation and comparison

For this report we will consider the future predictions of global temperature, and atmospheric CO_2 concentrations provided by the 2022 IPCC summary report (IPCC 2022*a*) to be the ground truth. Therefore after getting the predictions for each scenario, the predictions will be compared with the result from the IPCC report to determine their validity. Moreover we will compare the predictions with the results obtained by using the dame data on a 7-compartment carbon box model created by (Tomizuka 2009), so that we can compare the performance of machine learning models with a more traditional model that is specifically created to model the carbon-cycle.

3.2.7 Alternative research applications

Alternative research applications could have been implemented to conduct this study. For instance, a new carbon box model could have been developed similar to the one presented by Akira (Tomizuka 2009), yet more sophisticated, with additional compartments, to assess whether increased complexity enhances performance. Another method could be to figure out a way to use reinforcement learning, and deep reinforcement learning instead of shallow machine-learning, and deep-learning algorithms. Additionally training machine-learning models on a richer emission dataset that encompasses emissions from various greenhouse gases and multiple sources, could have been explored to investigate whether superior predictive performance could be obtained.

3.3 Ethical aspects of the research

In this research we made sure to prioritize the ethical aspects, thus a number of ethical principles have been considered, and are presented bellow.

1. We have been transparent about the source of the data that we have used to train the machine learning models. Moreover this data is publicly available, so other researchers can replicate our work, and see if they can reach the similar conclusions.

2. We collected our data from trustworthy sources such as the IPCC and Our World in Data (Roser 2023).

3. We made sure to give credit to the individuals and organization whose work has been used in the making of this report.

4. We have been transparent about the limitations of our research, so when we have any shortcomings, we present them in our research so the reader can take them into consideration.

5. We have been transparent about the way we have performed the data processing and manipulation. For example we explain how we generate the different future scenarios, and how we generate the missing data for land usage emissions from for the years 1750 to 1850 (see section 3.2.1)

6. Communalism is a key ethical principle that emphasizes the responsibility of researchers to share their findings with the broader scientific community and society. Research results should be made openly available and not withheld or kept confidential. In this report, we have made a deliberate effort to present all significant aspects of the research clearly and transparently, ensuring adherence to this principle.

Chapter 4

Results

4.1 Carbon box model

Using the carbon box model for each of the scenarios gives the concentration of CO_2 as visualized in the left-side plot in figure 4.1. To get the change of temperature we have used the theory of climate sensitivity as described earlier, the results for the different scenarios can be seen in the right-side plot in figure 4.1.

From Table 4.3 we see that the carbon box model is able to produce acceptable performance for all future scenarios. The carbon box model is the numerical approach that we will use as a baseline to match(in performance) using machine learning algorithms.

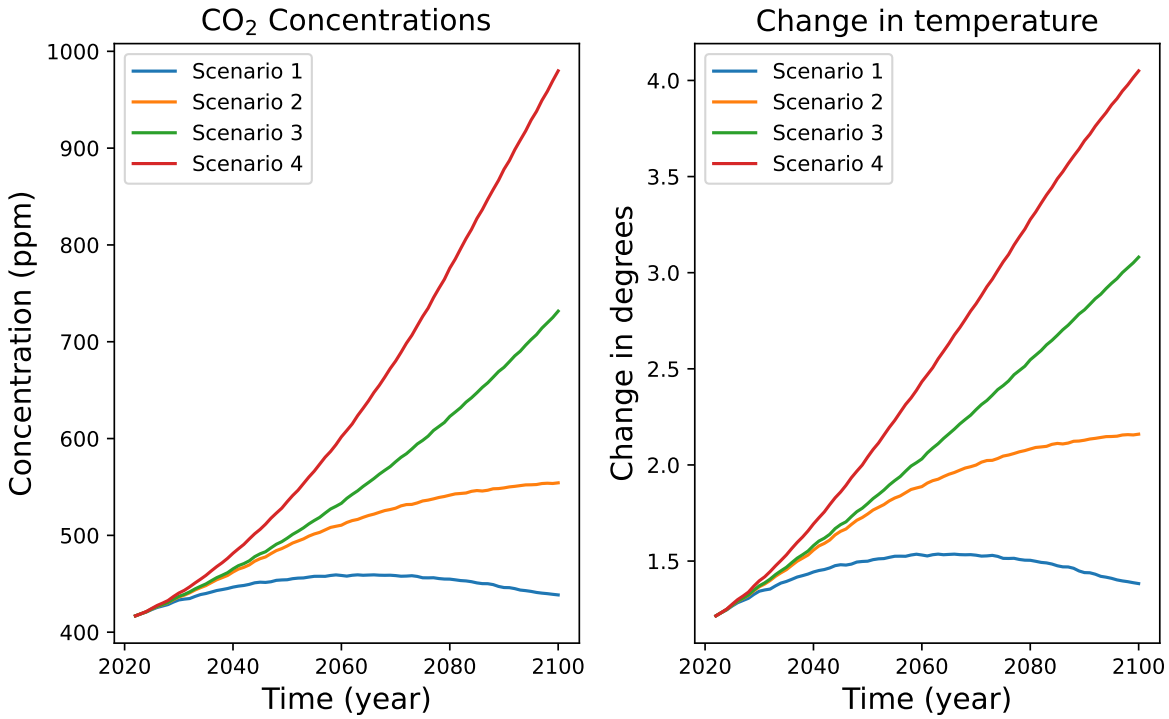

Figure 4.1: Predicted CO_2 concentrations and temperature change using the carbon box model

4.2 Linear regression

Linear regression gives the CO_2 concentrations in the left plot of Figure 4.2. Temperature changes, based on the theory of climate sensitivity, are shown in the right plot.
From Table 4.3 we see that the linear regression was able to produce acceptable performance for only 1 scenario. Linear regression is not able to perform well on this problem, because it can only fit straight lines though data, thus it can not capture complex nonlinear relationships, which are present in the case at hand.

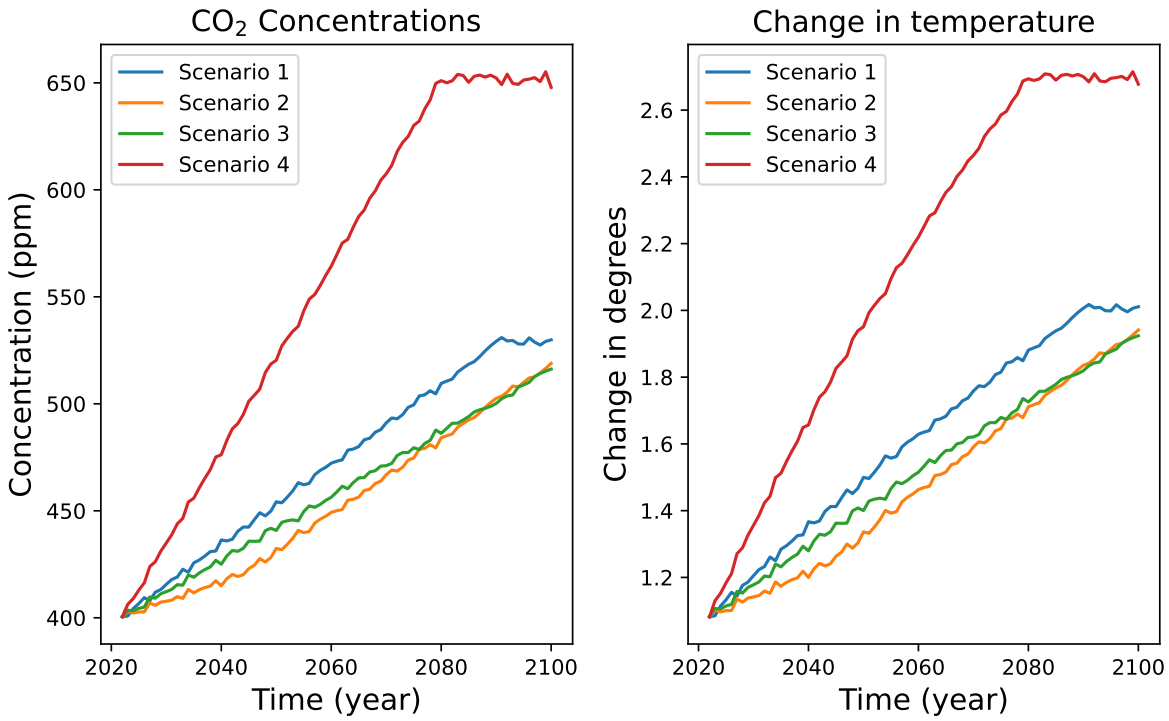

Figure 4.2: Predicted CO_2 concentrations and temperature change using linear regression

4.3 Polynomial regression

Polynomial regression gives the CO_2 concentrations in the left plot of Figure 4.3. Temperature changes, based on the theory of climate sensitivity, are shown in the right plot.
From Table 4.3 we see that the polynomial regression was able to produce acceptable performance for 3 scenarios. Polynomial regression is able to perform much better than linear regression, because unlike linear regression, it can fit a curve to the data, allowing it to capture some of the complex nonlinear relationships that are inherently present in the historical data.

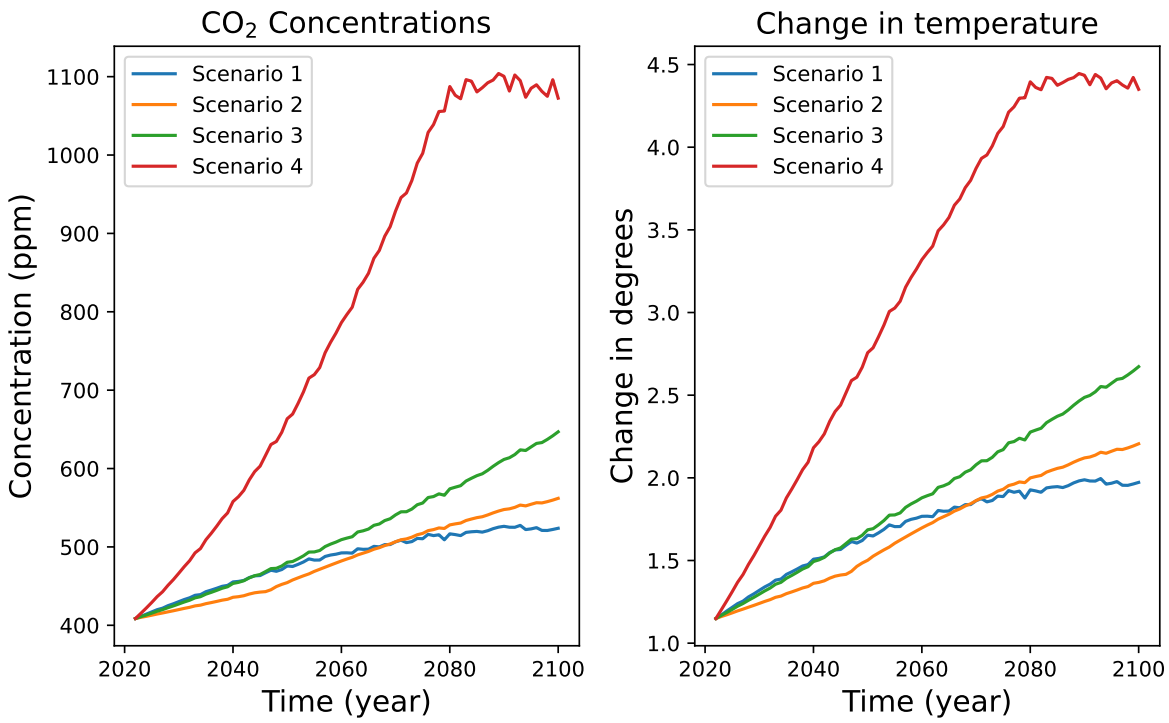

Figure 4.3: Predicted CO_2 concentrations and temperature change using polynomial regression

4.4 KNN

The KNN regressor gives the CO_2 concentrations in the left plot of Figure 4.4. Temperature changes, based on the theory of climate sensitivity, are shown in the right plot.
From Table 4.3 we see that the KNN regressor was not able to produce acceptable performance on any scenario.

The KNN regressor is not a suitable algorithm for this problem, because for any given data-sample the KNN regressor gives an output that is the average of the K-nearest neighbors to that sample (Kramer 2013). This means that the KNN regressor would not produce good results for extrapolation problems where the correct outputs are not present within the range of the training data. For example if the expected output is larger than any examples in the training data, then KNN would in the best case scenario give the largest value in the training data as output. This can clearly be seen in Figure 4.4 where scenario 1 and 2 are given almost the same value for all future years. In the same way if the expected output is smaller than any examples in the training data, then KNN would in the best case scenario give the smallest value in the training data as output.

Moreover due to the way KNN regressors work, they are not good when it comes to extrapolation on time series data like the one for the historical CO_2 concentrations that we have in this case. This is because KNN regressors can not capture the temporal relationships where past events have a lasting effect.

In scenario 1 and 2 the emissions are decreasing, but as we mentioned in the theoretical background, the carbon sinks are slow, thus the previously emitted carbon will still be present in the future, which means that in these scenarios the temporal relationships are important,

thus the KNN regressor would not perform well.

In the scenario 3 and 4, the emissions are increasing, thus we expect the concentrations to be higher than any historical concentrations, which means KNN regressor would not perform well.

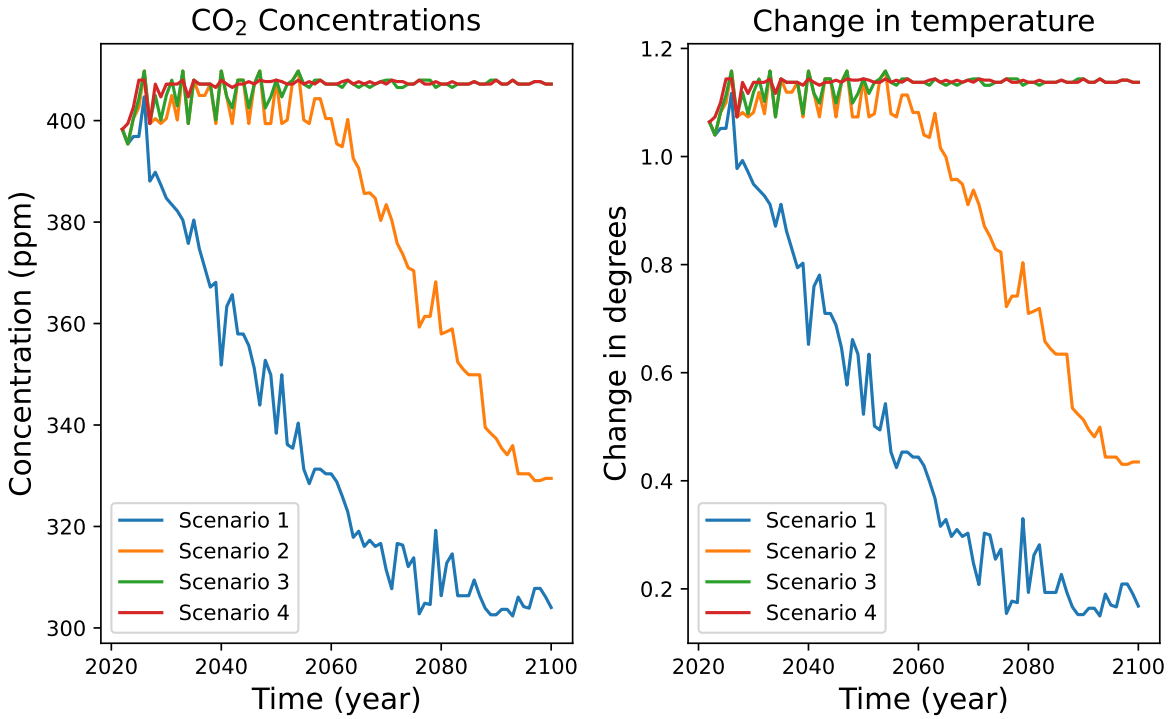

Figure 4.4: Predicted CO_2 concentrations and temperature change using the KNN regressor

4.5 SVM

The SVM regressor gives the CO_2 concentrations in the left plot of Figure 4.5. Temperature changes, based on the theory of climate sensitivity, are shown in the right plot.
From Table 4.3 we see that the SVM regressor was able to produce acceptable performance on only 1 scenario.

Unlike KNN and decision tree regressors, SVM can predict values that are not present within the range of the training data. However one factor that contributes to the poor performance of SVM regressors, is that they still suffer from the inability to learn temporal relationships in time series data.

From the Figure 4.5, we notice that the SVM regressor is predicting values that are too low in comparison with the predictions from polynomial regression, and the carbon box model, which have a better performance. This suggests that the SVM regressor is unable to adapt to the fast growing function that maps the emission values to the CO_2 concentrations in the atmosphere.

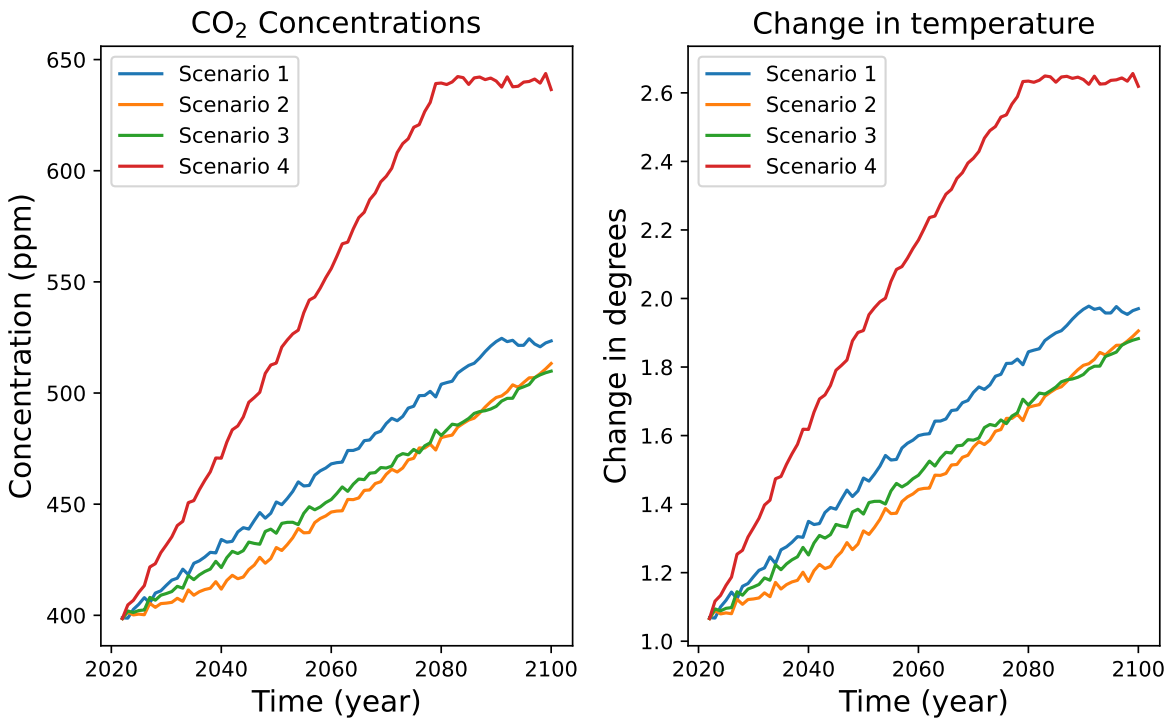

Figure 4.5: Predicted CO_2 concentrations and temperature change using the SVM regressor

4.6 Decision tree

The decision tree regression gives the CO_2 concentrations in the left plot of Figure 4.6. Temperature changes, based on climate sensitivity theory, are shown in the right plot.
From Table 4.3 we see that the Decision tree regressor was not able to produce acceptable performance on any scenario.

From Table 4.3 we notice that the predictions made by the decision tree and the KNN regressors are the same. This is because despite the fact that decision trees work in a way that is different from KNN regressors, they still suffer form the same problems.
Because of the way decision trees work (Alpaydin 2020) the predicted value is a always a value that exist in a leaf node, thus similar to KNN regressors, decision tree regressors are not suitable for extrapolation problems where the correct outputs are not present within the range of the training data. Moreover they have the same problem as KNN regressors when it comes to capturing temporal relationships where past events have a lasting effect.

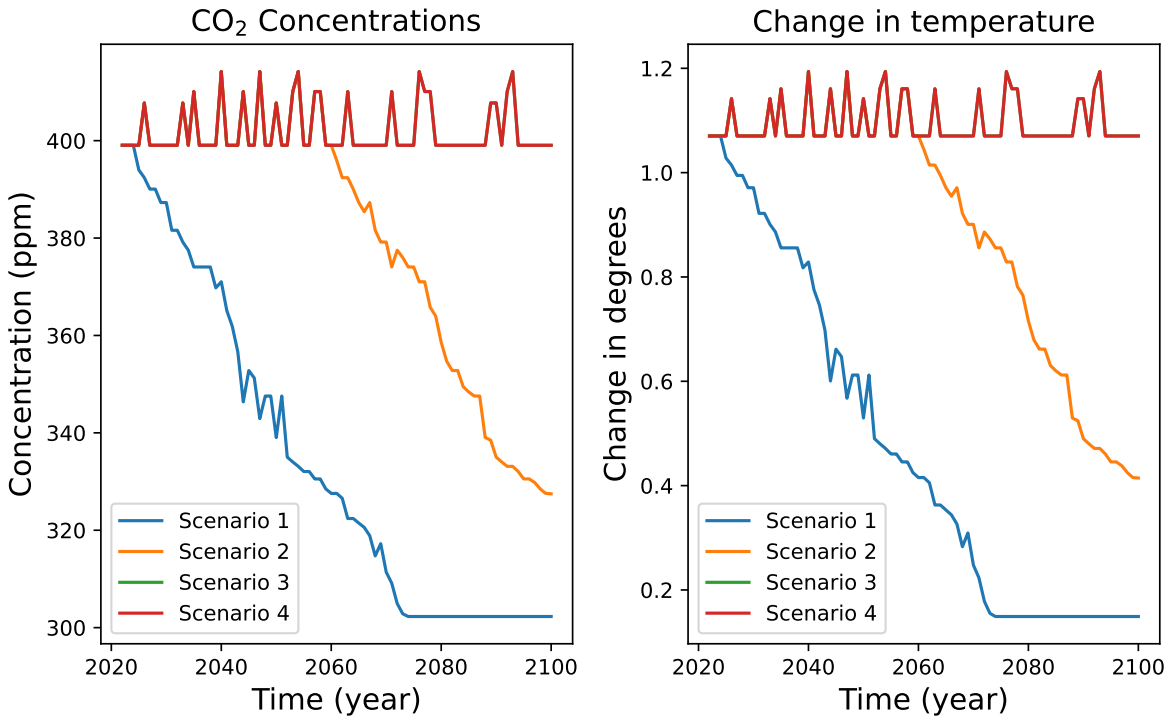

Figure 4.6: Predicted CO_2 concentrations and temperature change using the Decision Tree regressor

4.7 MLP

The MLP model gives the CO_2 concentrations in the left plot of Figure 4.7. Temperature changes, based on climate sensitivity theory, are shown in the right plot.
From Table 4.3 we see that the MLP regressor was not able to produce acceptable performance on any scenario.

The MLP model is one of the simplest deep learning models.
The poor performance of of the MLP model can be explained by the following arguments:

- While the MLP model has the theoretical ability to learn temporal relationships, it is not very good at doing so in practice, and other deep learning models perform much better in this aspect.
- The MLP is a deep learning model, and similar to other deep learning models they require large amount of training data to perform well. In the case we have at hand, there wasn't enough training examples for the MLP to learn all the relationships that are required to perform well.
- Similar to the SVM regressor discussed earlier, the MLP is predicting values that are too low, which suggests that it is unable to adapt quickly enough to the fast growing function that maps the emission values to the CO_2 concentrations in the atmosphere.

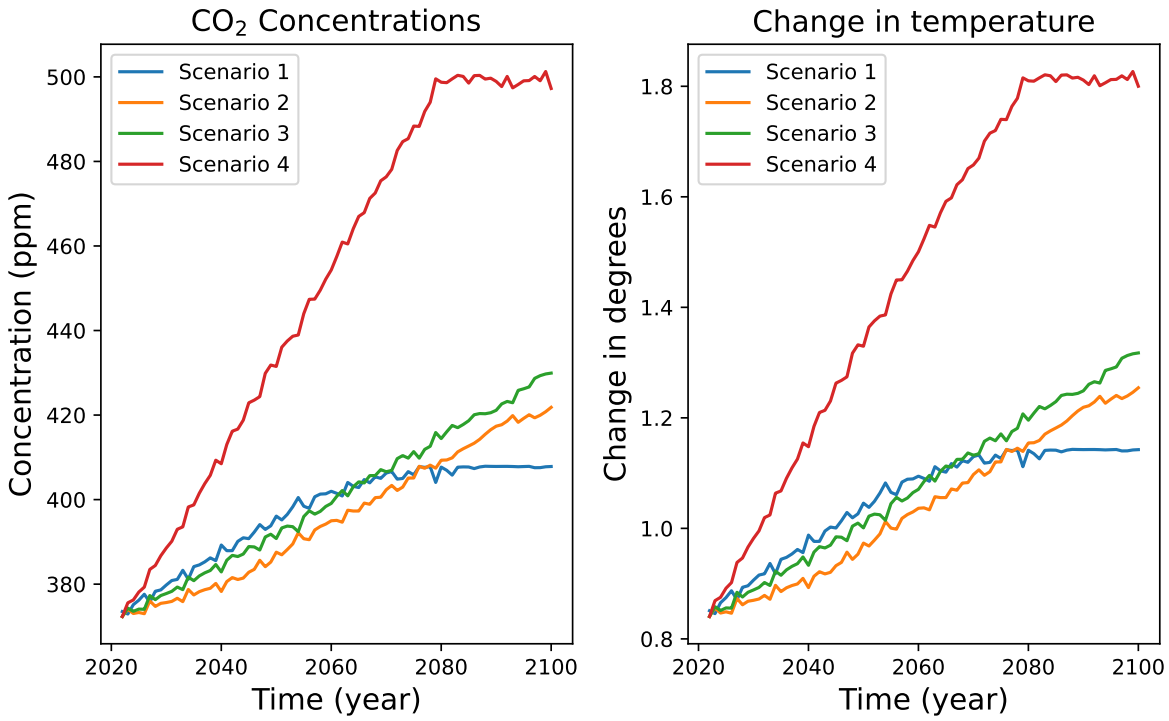

Figure 4.7: Predicted CO_2 concentrations and temperature change using the MLP model

4.8 Hybrid model

The hybrid model gives the CO_2 concentrations in the left plot of Figure 4.8. Temperature changes, based on climate sensitivity theory, are shown in the right plot.
From Table 4.3 we see that the the hybrid model is able to produce acceptable performance for all future scenarios. The performance of the hybrid model is similar to the performance of the carbon box model, but is slightly more inline with the best estimates from the IPCC.

The success of the hybrid model can be explained by the following arguments

- The LSTM layers used in this model are very well suited to learn the **long-term** temporal relationships (Zhang et al. 2021) that are present in the historical CO_2 emission and concentration data.
- The CNN layers used in this model are very well suited to learn the **short-term** temporal relationships that are present in the historical CO_2 emission and concentration data.
- Unlike the MLP model, the hybrid model is flexible enough to be able to adapt quickly to the fast growing function that maps the emission values to the CO_2 concentrations in the atmosphere.
- From Figure 4.8, we notice that the hybrid model is able to make clear distinction between scenario 1, 2 and 3, unlike linear regression, SVM and MLP, which give these models very similar predictions.

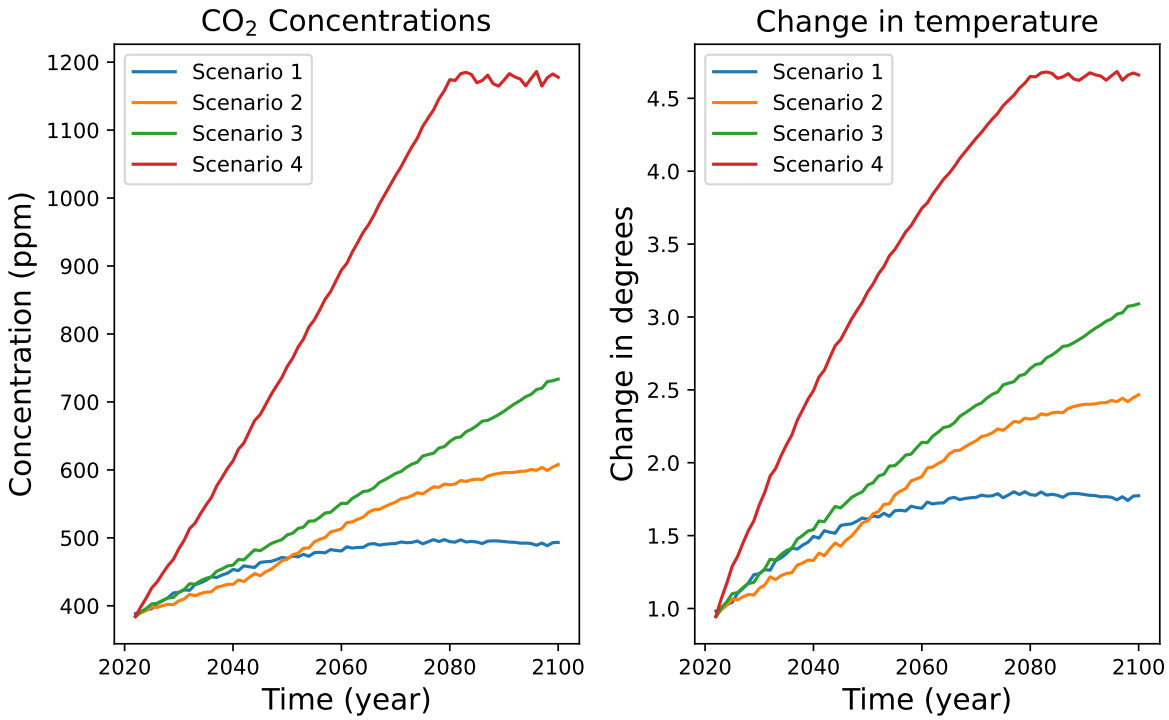

Figure 4.8: Predicted CO_2 concentrations and temperature change using the hybrid deep learning model

4.9 Result of evaluation metrics for all models

Table 4.1 shows the evaluation metrics on the test set for all models. These metrics have been used to measure the generalization of the models. For each model we ran the training and evaluation process multiple times with each set of hyper-parameters. This is to make sure that the obtained result is due to the choice of hyper-parameters, and not due to the randomness that is introduced in the training process(such as the random initialization of weights in the deep learning models.) It is worth mentioning that those validation metrics are more accurate for measuring the generalization ability on scenario 3, since this is the scenario that follows the trajectory of the historical data.

Table 4.2 shows the training errors of all models. These errors have been used for over-fitting detection.

In Table 4.3, we can see the temperature change results, for all models, moreover we can see the predictions made by the IPCC (IPCC 2022*a*)
The temperature change is measured in degrees and is calculated with respect to the year 1850, meaning that it is calculated by subtracting the historic global temperature for year 1850 from the predicted temperature for year 2100. We calculate the change in temperature with respect to the year 1850 because that is what the the IPCC report does, and we want to compare the result of this report with their predictions.

This project does not take into account emission data from greenhouse gases other than CO_2 (for example methane and nitrous oxide), however we have tuned the parameters β and λ as explained in the method section to match the observed total temperature change, thus this

project indirectly takes into account the affects of other greenhouse gases. This means that it still makes sense to compare the result of this project with the total temperature change presented in the IPCC report (IPCC 2022*a*).
To assess generalization, the validation metrics in Table 4.1 were employed. To estimate training and validation errors, the values reported in Table 4.2 are based on multiple repetitions, providing a more robust evaluation of model performance.

Model	**MSE**	**MAE**	**RMSE**
Linear regression	213.47	10.84	14.61
Polynomial regression	1172.36	28.17	34.24
KNN	1292.1	30.97	35.95
SVM	288.04	14.91	16.97
Decision tree	1887.93	35.87	43.45
MLP	2399.24	44.36	48.98
Hybrid model	842.37	20.22	29.02

Table 4.1: Evaluation metrics on test-set

Model	**MSE**	**MAE**	**RMSE**
Linear regression	9.18	1.89	3.03
Polynomial regression	3.13	1.07	1.77
KNN	1.42	0.68	1.19
SVM	10.06	1.71	3.17
Decision tree	0	0	0
MLP	1298.33	33.48	36.03
Hybrid model	31.21	4.33	5.59

Table 4.2: Training errors for all models after having trained them on the full dataset.

Model	Scenario 1	Scenario 2	Scenario 3	Scenario 4
IPCC best estimate (ground truth)	1.8	2.7	3.6	4.4
IPCC very likely range (ground truth)	1.3 to 2.4	2.1 to 3.5	2.8 to 4.6	3.3 to 5.7
Box model	1.3 ✔	2.1 ✔	3.0 ✔	4.0 ✔
Linear regression	1.9 ✔	1.9	1.8	2.6
Polynomial regression	1.9 ✔	2.1 ✔	2.6	4.3 ✔
KNN	0.1	0.3	1.0	1.0
SVM	1.9 ✔	1.8	1.8	2.5
Decision tree	0.1	0.3	1.0	1.0
MLP	1.1	1.2	1.2	1.7
Hybrid model	1.7 ✔	2.4 ✔	3.0 ✔	4.6 ✔

Table 4.3: Temperature change predictions for year 2100 in degrees Celsius with respect to the year 1850. The symbol ✔represents an acceptable result

Chapter 5

Discussion

The study's central question, assessing the efficacy of machine-learning models in predicting atmospheric CO_2 concentrations, has been addressed through comprehensive analysis. The polynomial regression model, while simpler, shows effectiveness, but lacks the precision of the more complex hybrid deep-learning model. This model, combining LSTM and CNN layers, is adept at interpreting both long-term and short-term trends in emissions data, aligning closely with IPCC predictions and the 7-compartment box model (Tomizuka 2009).

A critical limitation of this study, particularly in terms of validity, is the dependency on historical data quality and availability. The validity of our models is directly tied to the accuracy and completeness of the historical emission data. Incomplete or inaccurate data could significantly skew the models' predictions, impacting the construct validity of our findings.

The significance of our research lies in the potential application in climate policy and environmental conservation extends beyond academic contexts. Our accurate predictions of CO_2 concentrations empower policymakers and scientists to better strategize about climate change mitigation efforts. These methodologies and findings also have the potential to influence and inform public policy and heighten community awareness regarding climate change and its inherent challenges.

This research also sets a precedent for future research which should focus on enhancing the model's ability to adapt to varying emission scenarios. The integration of more input variables such as those from other greenhouse gases, and the exploration of models with better ability to extrapolate beyond current trends would be valuable. Additionally, research could explore the scalability of these models in larger, more diverse datasets and their applicability in other environmental prediction contexts.

In conclusion, this study affirms that machine-learning models, in this case the hybrid deep-learning models, are effective tools for predicting atmospheric CO_2 concentrations. As climate change remains a pressing global issue, the continued development and refinement of these models are essential for informed decision-making and policy development.

Bibliography

Alpaydin, E. (2020), *Introduction to Machine Learning*, fourth edn, The MIT Press.

Boucher, Haigh, Hauglustaine, Haywood, Myhre, Nakajima, Shi & Solomon (2018), 'Radiative forcing of climate change', *Intergovernmental Panel on Climate Change* .
URL: *https://www.ipcc.ch/site/assets/uploads/2018/03/TAR-06.pdf*

Brownlee, J. (2017), *Long Short-Term Memory Networks With Python: Develop Sequence Prediction Models With Deep Learning*, 1 edn, Jason Brownlee.

Ciais, P. et al. (2014), 'Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system', *Biogeosciences* **11**.
URL: *https://doi.org/10.5194/bg-11-3547-2014*

Cornell, S. E., Prentice, C., House, J. & Downy, C. (2012), *Understanding the Earth System*, Cambridge University Press.
URL: *https://www.cambridge.org/core/books/understanding-the-earth-system/5534588F2EE9EE776B8642EE4A952672*

Dey, R. & Salem, F. M. (2017), 'Gate-variants of gated recurrent unit (gru) neural networks'.
URL: *https://arxiv.org/abs/1701.05923*

Goodfellow, I., Bengio, Y. & Courville, A. (2016), *Deep Learning*, MIT Press.
URL: *http://www.deeplearningbook.org*

Géron, A. (2022), *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*, 2 edn, O'Reilly Media Inc.

Hairer, E. & Wanner, G. (2015), *Encyclopedia of Applied and Computational Mathematics*, first edn, Springer Berlin, Heidelberg.
URL: *https://doi.org/10.1007/978-3-540-70529-1_144*

Hewage, P., Trovati, M., Pereira, E. & Behera, A. (2020), 'Deep learning-based effective fine-grained weather forecasting model', **24**.
URL: *https://doi.org/10.1007/s10044-020-00898-1*

Hochreiter, S. & Schmidhuber, J. (1997), 'Long short-term memory', *Neural computation* **9**.
URL: *https://deeplearning.cs.cmu.edu/F23/document/readings/LSTM.pdf*

IPCC (2022*a*), 'Summary for policymakers', *Intergovernmental Panel on Climate Change* .
URL: *https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf*

IPCC (2022*b*), 'Synthesis report of the ipcc sixth assessment report (ar6)', *Intergovernmental Panel on Climate Change* .
URL: *https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_LongerReport.pdf*

Kramer, O. (2013), *Dimensionality Reduction with Unsupervised Nearest Neighbors*, first edn, Springer Berlin, Heidelberg.
URL: *https://doi.org/10.1007/978-3-642-38652-7*

Lindsey, R. (2023), 'Climate change: Atmospheric carbon dioxide'.
URL: *https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide*

Python-Software-Foundation (2023), 'Python 3.12.1 documentation'.
URL: *https://docs.python.org/3/*

Riebeek, H. (2023), 'Carbon cycle'.
URL: *https://earthobservatory.nasa.gov/features/CarbonCycle*

Ritchie, H. & Rose, M. (2020), 'Co2 emissions'. Accessed: 2023/11/14.
URL: *https://ourworldindata.org/co2-emissions*

Roser, M. (2023), 'Our world in data'.
URL: *https://repository.gheli.harvard.edu/repository/12126/*

Sherwood, S. C. et al. (2020), 'An assessment of earth's climate sensitivity using multiple lines of evidence', *Reviews of Geophysics* **58**.
URL: *https://doi.org/10.1029/2019RG000678*

Singh, D. & Singh, B. (2020), 'Investigating the impact of data normalization on classification performanc', *Applied Soft Computing* **97**.
URL: *https://www.sciencedirect.com/science/article/pii/S1568494619302947*

Smith, C. et al. (2021), 'The earth's energy budget, climate feedbacks, and climate sensitivity supplementary material'.
URL: *https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter07.pdf*

Stoffi, F., Cevolani, G. & Gnecco, G. (2022), 'Simple models in complex worlds: Occam's razor and statistical learning theory', *Minds and Machines* **32**.
URL: *https://link.springer.com/article/10.1007/s11023-022-09592-z*

Tomizuka, A. (2009), 'Is a box model effective for understanding the carbon cycle?', **77**, 156–164.

Wallace, J. M. & Hobbs, P. V. (2006), *Atmospheric Science: An Introductory Survey*, 2 edn, Academic Press.

Witten, I. H. & Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, 2 edn, Morgan Kaufmann Publishers.

Zhang, J., Zeng, Y. & Starly, B. (2021), 'Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis', **3**.
URL: *https://doi.org/10.1007/s42452-021-04427-5*