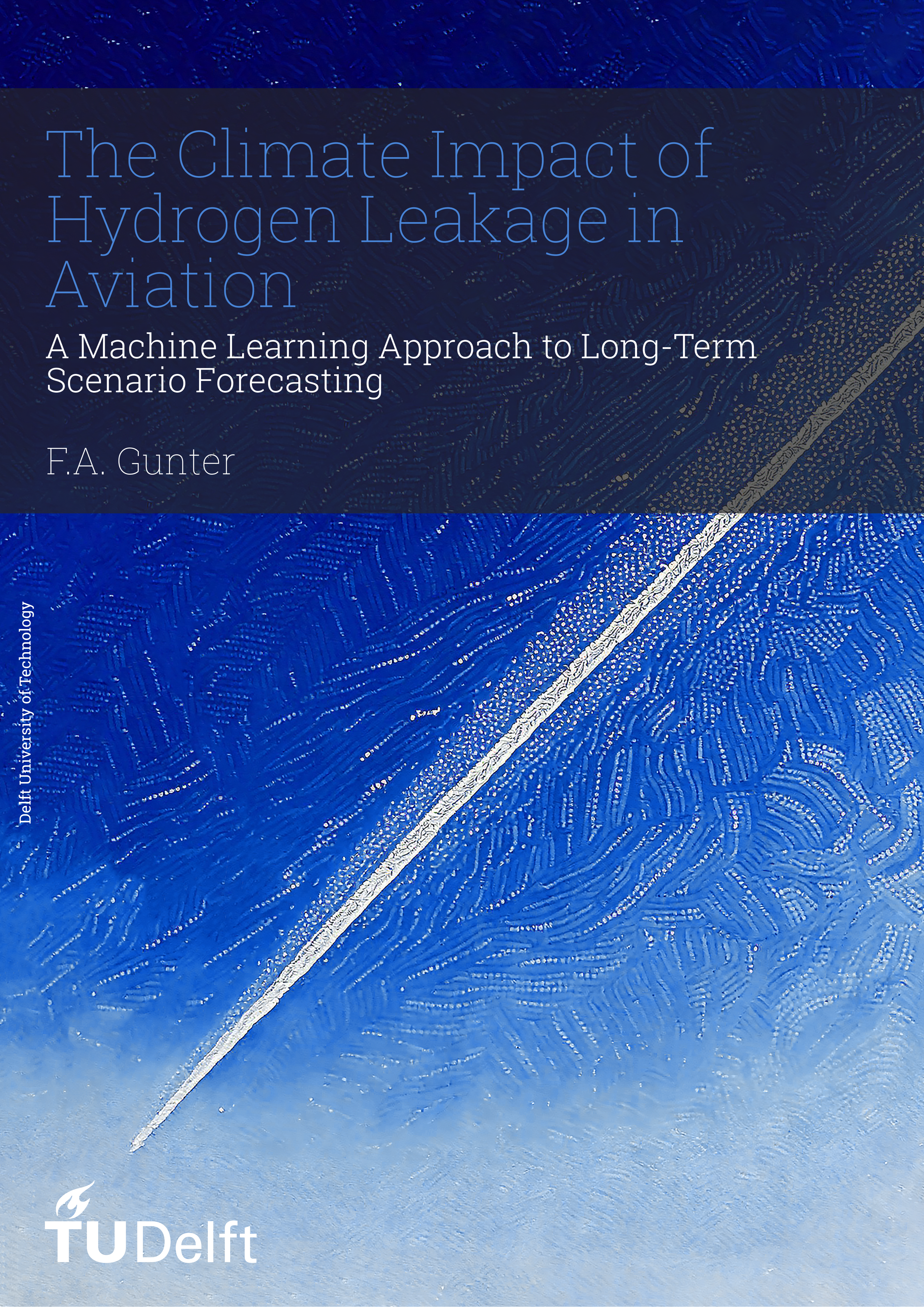


The Climate Impact of Hydrogen Leakage in Aviation

A Machine Learning Approach to Long-Term Scenario Forecasting

F.A. Gunter

Delft University of Technology

The Climate Impact of Hydrogen Leakage in Aviation

A Machine Learning Approach to Long-Term Scenario Forecasting

by

F.A. Gunter

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday, December 12th, 2024, at 15:00.

Student number:	4877764	
Project duration:	April 15, 2024 – December 12, 2024	
Thesis committee:	Dr. ir. M.F.M. Hoogreef,	TU Delft
	Prof. dr. V. Grewe,	TU Delft | DLR
	Prof. dr. M. Menenti	TU Delft
	Dr. rer. nat. S. Völk,	DLR

TU Delft

Abstract

The adoption of hydrogen as an alternative fuel in aircraft has the potential to reduce the climate effect of aviation significantly. However, hydrogen leakage during production, storage, or use can offset these benefits by altering atmospheric chemistry and composition, particularly through interactions with methane, ozone, and stratospheric water vapour. This study employs surrogate models based on recurrent and convolutional neural networks to simulate the climate effects of hydrogen leaks, achieving rapid projections 30,000 times faster than conventional climate models, with an error margin of less than 5%. This efficiency enables the quantification of uncertainties related to hydrogen leakage rates and atmospheric chemistry through Monte Carlo simulations, allowing for an assessment of their contributions to radiative forcing under various Shared Socioeconomic Pathway (SSP) climate scenarios. By 2100, the radiative forcing from aviation-induced hydrogen leaks is projected to reach 43.9 ± 21.2 mW m^{-2} ($\pm 1\sigma$) under the stringent climate change mitigation projection SSP1-2.6, accounting for 35% of aviation's total radiative forcing. Under the more conservative scenario SSP3-7.0, higher methane levels reduce the oxidative capacity of the atmosphere, lowering the projection to 17.7 ± 6.9 mW m^{-2} ($\pm 1\sigma$) which corresponds to approximately 5% of the aviation's total radiative forcing for this scenario. These findings demonstrate that hydrogen leaks have the potential to substantially contribute to aviation's total radiative forcing, with the magnitude of their impact heavily influenced by background climate conditions. It will be important to minimise hydrogen leakage to fully harness the climate benefits of transitioning to hydrogen as a fuel for aircraft.

Preface

After six and a half wonderful years as a student in Delft, it's finally time to close this chapter with the completion of this thesis. Throughout my studies, I've had the incredible opportunity to discover my true passions and to meet so many fun and inspiring people. My time with the student team AeroDelft taught me how much I enjoy tackling challenges related to climate change, a passion that has shaped a significant part of my academic journey. My decision to focus the rest of my studies on sustainable aviation led me to an unforgettable internship at Bosch in the United States and ultimately culminated in the opportunity to research the intersection of two subjects that deeply interested me: climate change and aviation.

My heartfelt thanks go to my daily supervisor, Dr. Stefan Völk, for his continuous support throughout my thesis and the interesting discussions during our weekly meetings. I am also very grateful to Professor Volker Grewe for allowing me to work on a subject I am truly passionate about and for the extensive feedback I have received over the course of this year. Next to this, I would like to thank Prof. Massimo Menenti and Dr. Maurice Hoogreef for agreeing to be part of my graduation committee.

This thesis wouldn't have been possible without the unwavering support of my friends! Whether dragging me to the UB for yet another evening of studying or convincing me to take a break and grab a drink, those moments helped me push through. Lastly, I want to give a big thank you to my family, especially my parents, for always supporting and encouraging me in everything I do. Thank you!

Floris Gunter
Delft, November 2024

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
ATR	Average Temperature Response
BAU	Business As Usual
CH_4	Methane
CMIP6	Coupled Model Intercomparison Project Phase 6
CNN	Convolutional Neural Network
CO	Carbon Monoxide
CO_2	Carbon Dioxide
ConvLSTM	Convolutional Long Short-Term Memory
CurTec	Current Technology
DU	Dobson Unit
ERF	Effective Radiative Forcing
GFDL	Geophysical Fluid Dynamics Laboratory
GHG	Greenhouse Gas
GWP	Global Warming Potential
H_2	Hydrogen
H_2O	Water Vapour
HO_2	Hydroperoxyl
INCA	Integrated Catchments (climate model)
IPCC	Intergovernmental Panel on Climate Change
IRF	Instantaneous Radiative Forcing
LF	Leakage Factor
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MCS	Monte Carlo Simulation
ML	Machine Learning
MRI-ESM2.0	Meteorological Research Institute Earth System Model 2.0
N_2O	Nitrogen Dioxide
N_x	Nitrogen Oxides
O_3	Ozone

Abbreviation	Definition
OH	Hydroxyl
OsloCTM	Oslo Chemistry Transport Model
PM	Particulate Matter
ppb	Parts Per Billion
RF	Radiative Forcing
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SD	Standard Deviation
SO_x	Sulfur Oxides
SSP	Shared Socioeconomic Pathways
STP	Standard Temperature and Pressure
SWV	Stratospheric Water Vapor
UHC	Unburned Hydrocarbons
UKESM1	UK Earth System Model 1
UV	Ultraviolet
VOC	Volatile Organic Compounds
WACCM	Whole Atmosphere Community Climate Model

Symbols

Symbol	Definition	Unit
E_s	Emission of species s	[Tg yr^{-1}]
k	Reaction rate constant	[cm^3 s^{-1}]
k_d	Deposition rate constant	[s^{-1}]
k_s	Stratospheric sink rate constant	[s^{-1}]
k_x	Auxiliary reaction rate constant	[s^{-1}]
LWH	Lower Heating Value	[MJ kg^{-1}]
M	Mass	[kg]
R_s	Radiative efficiency of species s	[W m^{-2} ppb$_s^{-1}$]
S_s	Source term of species s	[Tg]
T_s	Global mean surface temperature	[K]
α	Methane feedback yield	[-]
λ (ML)	Regularization parameter	[-]
λ (Metrics)	Climate sensitivity	[K W^{-1} m^2]
τ_s	Atmospheric lifetime of species s	[yr]

1

Introduction

Climate change remains one of humanity's most pressing challenges, with significant implications for ecosystems, economies, and societies worldwide. According to the Intergovernmental Panel on Climate Change (IPCC), anthropogenic activities, primarily burning fossil fuels, have been the principal drivers of global warming observed over the past century (IPCC, 2021a). In response to this pressing challenge, the Paris Agreement was adopted in 2015 to limit global temperature rise to well below 2°C above pre-industrial levels and to pursue efforts to restrict the increase to 1.5°C (Schleussner et al., 2016). Significant reductions in greenhouse gas (GHG) emissions are necessary across all sectors to achieve these objectives. As illustrated in Figure 1.1, the aviation sector's contribution to global CO_2 emissions is relatively minor. However, its absolute emissions increased by approximately 60% between 1990 and 2019, leading to a nearly 30% rise in its relative share. Recent studies indicate that this upward trend is expected to continue, suggesting that the aviation industry will likely fall short of the climate goals outlined in the Paris Agreement (Grewe et al., 2021). The sector's reliance on fossil fuels and the anticipated increase in air traffic contribute to its likelihood of missing the established emission reduction targets. This underscores the critical need for sustainable innovations to mitigate the aviation industry's environmental impact and align with global climate objectives.

In response to the need for sustainable aviation, hydrogen technology has emerged as a promising solution to reduce the carbon footprint of air travel (Adler and Martins, 2023). Hydrogen combustion primarily produces water vapour when used as a fuel, avoiding direct CO_2 emissions (Adler and Martins, 2023). This technological shift would be a crucial step toward achieving the decarbonization goals of international climate agreements. Adopting hydrogen-powered aircraft could revolutionize the aviation industry, significantly lowering greenhouse gas emissions and contributing to broader climate mitigation efforts (Sáez Ortuño et al., 2023). However, while the potential benefits are substantial, the total environmental impact of this transition must be carefully examined.

One of the primary concerns with the adoption of hydrogen fuel in aviation is the issue of fugitive emissions - intentional purging or unintentional leaks of hydrogen during production, storage, and use. Hydrogen, the smallest and lightest element, is a molecule that can easily escape containment. Such

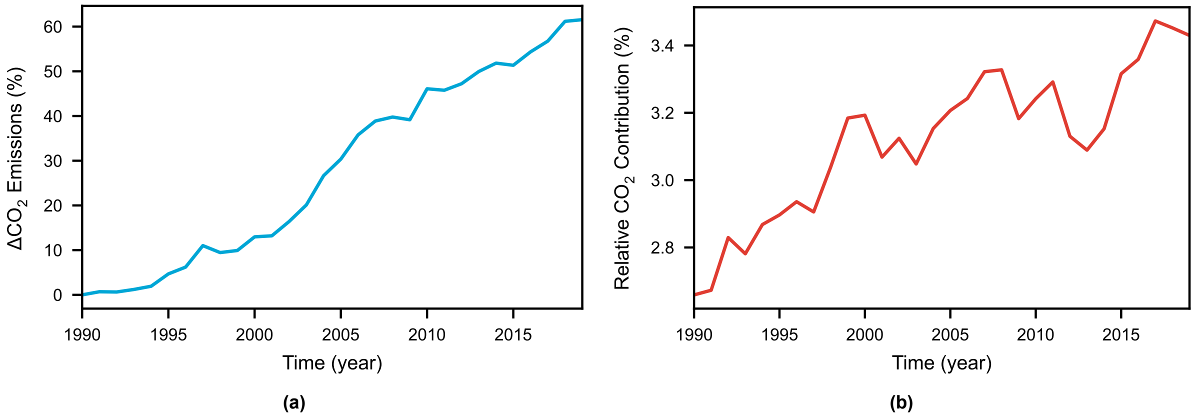


Figure 1.1: Relative increase in aviation CO_2 emissions with respect to 1990 **(a)** and the relative contribution of aviation to the total amount of CO_2 emissions **(b)** between 1990 and 2019. Data is adapted from Friedlingstein et al. (2023).

leaks can have significant atmospheric implications. Once released, hydrogen can participate in various atmospheric reactions, indirectly affecting the concentrations of other greenhouse gases, such as methane and ozone (Ehhalt and Rohrer, 2009; Derwent et al., 2006). These interactions could lead to warming, counteracting some benefits of reducing CO_2 emissions (Hauglustaine et al., 2022; Ocko and Hamburg, 2022). Understanding the extent and impact of these fugitive hydrogen emissions is crucial for accurately assessing the climate benefits of hydrogen-powered aviation.

Recent studies have employed detailed climate models to simulate the climate effect of hydrogen emissions (Sand et al., 2023; Warwick et al., 2022). However, the specific impact of hydrogen emissions from aviation sources has yet to be investigated. Furthermore, recent studies limit their analysis to the evaluation of hydrogen's climate impact under current climate conditions. However, Warwick et al. (2022) and Sand et al. (2023) indicate that the climate impact of hydrogen could be strongly influenced by background concentrations of other atmospheric species. This indicates that the climate impacts of hydrogen are intrinsically tied to the progression of future climate scenarios.

This work investigates the climatic impacts of fugitive hydrogen emissions from aviation sources on global atmospheric greenhouse gas concentrations and radiative forcings in the context of various climate scenarios. A novel model will be integrated into OpenAirClim, a climate response tool designed for simplified assessments of the chemistry-climate effects of air traffic emissions (Grewe and Stenke, 2008). In contrast to fully coupled climate models, OpenAirClim can rapidly assess a wide range of future scenarios. Similarly, this study proposed the application of Recurrent Neural Networks (RNNs) as surrogate models to substantially reduce the computational costs associated with projecting the climatic impacts of hydrogen emissions while ensuring minimal loss of accuracy. Subsequently, a Monte Carlo simulation is employed to evaluate the uncertainties related to the atmospheric chemistry of hydrogen and the various assumptions necessary for forecasting the future climate effects of fugitive emissions.

This research contributes to the scientific discourse by examining the climate impacts of hydrogen under projected future emission scenarios. More broadly, it evaluates the effectiveness of Recurrent Neural Networks (RNNs) for long-term atmospheric chemistry modelling. If successful, this approach could enable rapid assessments of hydrogen's climate impact, supporting a more thorough evaluation of hydrogen's viability as an alternative aviation fuel. The research question is formulated as follows:

> *What are the climatic impacts and associated uncertainties of fugitive hydrogen emissions from aviation on global atmospheric greenhouse gas concentrations and radiative forcings for various emission scenarios until the year 2100?*

First, an overview of related literature is provided in chapter 2. Subsequently, the methodology for developing the model capable of efficiently assessing the climate impact of hydrogen is detailed in chapter 3. A comprehensive verification, validation, and sensitivity analysis of the proposed model is then conducted in chapter 4. Afterwards, results obtained using the model are shown and discussed in chapter 5. Finally, chapter 6 concludes this thesis and provides an outlook on future research.

2

Scientific Background

This chapter offers an overview of the most relevant literature, serving as a foundation for the methodology. The chapter is divided into four sections. First, section 2.1 provides an overview of the climate impact of aviation and how hydrogen is projected to reduce this impact. Subsequently, section 2.2 explores how this climate impact can be accurately evaluated through climate metrics. The atmospheric chemistry of hydrogen is investigated in detail in section 2.3. Finally, section 2.4 examines scenario modelling approaches for evaluating the potential future impacts of hydrogen adoption in aviation.

2.1. Hydrogen for Aircraft Emission Mitigation

The aviation industry plays a significant role in global climate change, contributing to both long-term and short-term effects through the emission of various greenhouse gases and aerosols (Lee et al., 2010). While aviation accounts for approximately 2.4% of global CO_2 emissions, its total climate impact is considerably higher when non-CO_2 effects are considered (Lee et al., 2021). Figure 2.1 highlights that aviation's climate effects extend beyond CO_2 alone and that the net effect is complex and associated with significant uncertainties.

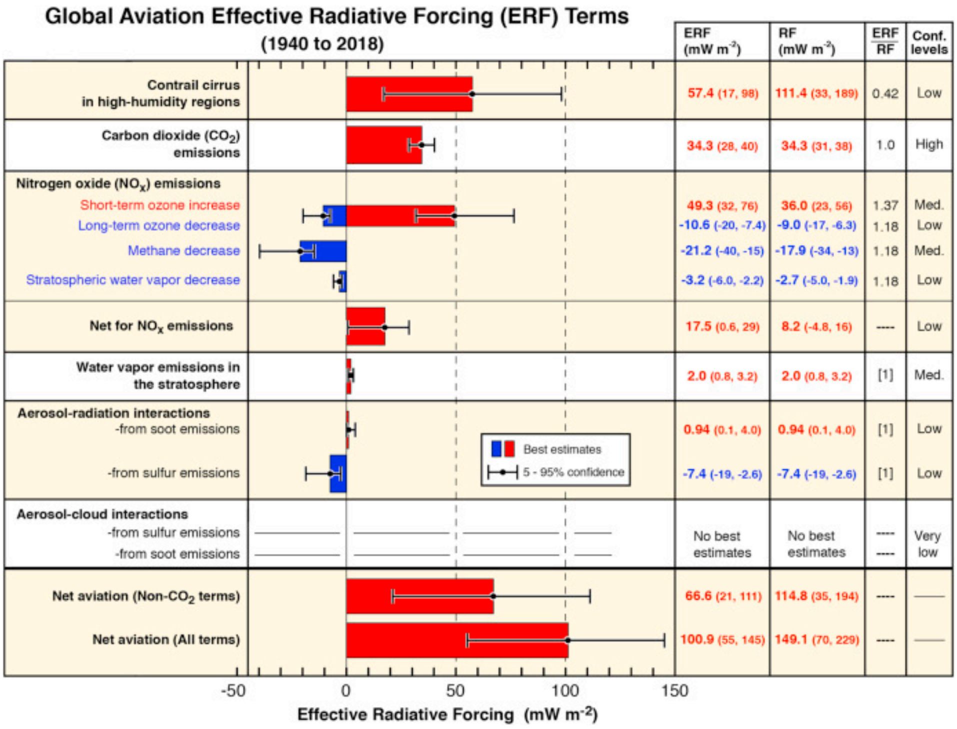


Figure 2.1: Representation of the best estimates of climate forcing components from global aviation between 1940 and 2018, taken from Lee et al. (2021). The bars depict the best estimates for effective radical forcing (ERF), and the whiskers represent the 5th to 95th percentile confidence intervals. Red bars denote warming effects, while blue bars indicate cooling effects.

In light of the significant climate impacts from aviation emissions, hydrogen is increasingly recognized as a viable alternative to conventional fossil-based fuels in the aviation sector (Adler and Martins, 2023; Gangoli Rao et al., 2020). As a clean energy carrier, hydrogen can be utilized in various propulsion technologies—including fuel cells and hydrogen combustion engines—that produce minimal carbon emissions compared to traditional jet fuels. Thus, adopting hydrogen in aviation aligns to decarbonize the industry and helps mitigate the environmental impacts of air travel.

This section will first elaborate on the effects of emissions from conventional energy sources, providing a comprehensive understanding of how pollutants such as nitrogen oxides, water vapour, contrails, sulfur oxides, and particulate matter contribute to aviation's overall climate impact. It will then explore hydrogen's role in advancing clean aviation by examining its advantages, challenges, and implications for atmospheric chemistry.

2.1.1. Aircraft Emissions

Aviation emissions are a complex mix of gases and particles that have warming and cooling effects on the climate. The primary emissions from aircraft engines include CO_2, water vapour (H_2O), nitrogen oxides (NO_x), sulfur oxides (SO_x), unburned hydrocarbons (UHC), carbon monoxide (CO), and particulate matter (PM) such as soot and sulfate aerosols (Lee et al., 2010).

Carbon Dioxide (CO_2)

CO_2 emissions are the primary contributor to anthropogenic climate change due to their significant role in enhancing the greenhouse effect. In aviation, CO_2 is produced directly from the combustion of hydrocarbon fuels such as kerosene, which powers most commercial aircraft (Lee et al., 2021). The combustion process oxidizes carbon-containing fuels, releasing CO_2 into the atmosphere, where it accumulates over time. The atmospheric lifetime of CO_2 is extensive, with a significant fraction remaining in the atmosphere for centuries to millennia (Archer et al., 2009). This long-term persistence means that CO_2 emissions have a cumulative effect on atmospheric concentrations, leading to sustained radiative forcing and global warming (IPCC, 2021b).

CO_2 emissions are intricately linked to the Earth's carbon cycle, involving exchanges among interconnected reservoirs such as the atmosphere, oceans, terrestrial biosphere, and lithosphere. These exchanges occur over various timescales—from seconds to millennia—through complex physical, chemical, and biological processes (IPCC, 2021b). Unlike many atmospheric pollutants that can be modelled using a single exponential decay function due to their relatively simple and consistent removal processes, CO_2's removal involves multiple mechanisms operating on different timescales, including rapid exchanges with the biosphere and surface oceans (years to decades), mixing into the deep ocean (centuries), and long-term geological sequestration (thousands to hundreds of thousands of years) (Shine et al., 2005). Consequently, no single timescale can accurately represent the atmospheric lifetime of CO_2, necessitating models that incorporate multiple decay modes or timescales to simulate how CO_2's concentrations change over time following emissions precisely.

Nitrogen Oxides (NO_x)

Nitrogen oxides (NO_x), consisting primarily of nitric oxide (NO) and nitrogen dioxide (NO_2), are emitted during high-temperature combustion in aircraft engines. The formation of NO_x occurs when atmospheric nitrogen and oxygen react under the extreme temperatures present in the combustion chamber (Schumann, 1997). NO_x emissions from aviation play a significant role in atmospheric chemistry, especially at cruise altitudes in the upper troposphere and lower stratosphere, as discussed by Grewe et al. (2019). At these altitudes, NO_x emissions lead to the photochemical production of tropospheric ozone (O_3), a short-lived climate forcer with a potent warming effect (IPCC, 2021b). The formation of O_3 involves a series of reactions initiated by NO_x interacting with volatile organic compounds (VOCs) and CO in the presence of sunlight (Stevenson et al., 2013). A detailed description of this chemistry is provided in section 2.3.

Simultaneously, NO_x emissions impact the atmospheric concentration of methane (CH_4), another important greenhouse gas. NO_x emissions increase the abundance of hydroxyl radicals (OH), which are the primary oxidizing agents that remove CH_4 from the atmosphere (Stevenson et al., 2004). This process reduces the atmospheric lifetime of CH_4, leading to a decrease in its concentration and a subsequent negative radiative forcing (cooling effect).

The net climate effect of NO_x emissions is the result of these opposing processes, warming from increased O_3 and cooling from decreased CH_4 (Fuglestvedt et al., 2010). Quantifying the overall impact is complex due to the non-linearities in atmospheric chemistry and the spatial variability of NO_x-induced effects (Holmes et al., 2011). Moreover, the effects are temporally distinct: O_3 changes occur rapidly and are short-lived, while CH_4 adjustments develop over longer timescales due to its atmospheric lifetime of approximately 12 years (Prather et al., 2012). Understanding the net radiative forcing from aviation NO_x emissions requires sophisticated modelling that accounts for atmospheric transport, chemical reactions, and interactions with other climate forces.

Water Vapour Emissions and Contrail Formation

Although aviation contributes a relatively small amount of H_2O compared to natural sources such as oceanic evaporation, the injection of H_2O at cruise altitudes—specifically in the upper troposphere and lower stratosphere—can have disproportionately large climatic effects. These regions are particularly sensitive to additional H_2O because it influences radiative balance and cloud formation processes, such as contrail and cirrus cloud development, which are key drivers of radiative forcing and climate warming (Gettelman et al., 2011). The persistence and radiative properties of these clouds amplify the overall climatic impact of H_2O emissions from aircraft. A substantial fraction of annual aircraft emissions occur near or within the lower stratosphere, primarily in the Northern Hemisphere, as discussed by Forster et al. (2003), where the low background humidity means that additional H_2O can perturb the radiative balance and influence atmospheric temperatures.

At cruise altitudes, the cold temperatures and low humidity conditions lead to the formation of condensation trails, or contrails, when H_2O emitted by aircraft mixes with the ambient air (Schumann, 2005). These contrails, composed of ice crystals, can persist and evolve into contrail-induced cirrus clouds under ice-supersaturated conditions, differing from natural cirrus in their microphysical properties and extent (Bock and Burkhardt, 2016). Contrail cirrus clouds generally contribute to a net warming effect on the Earth's climate, as they trap outgoing longwave radiation more effectively than they reflect incoming solar radiation, especially at night (IPCC, 2021b; Stuber et al., 2006). The overall radiative impact depends on factors like cloud optical properties, the time of day, and surface albedo (McCausland, 2024).

Sulfur Oxides, Soot and Carbon Monoxide

Sulfur oxides (SO_x), primarily sulfur dioxide (SO_2), are emitted during the combustion of sulfur-containing jet fuels (Lee et al., 2021). In the atmosphere, SO_2 oxidizes to form sulfate aerosols, which act as cloud condensation nuclei (CCN) and influence cloud properties, thereby impacting climate and air quality (Lee et al., 2021). Additionally, aircraft emissions include particulate matter such as black carbon (soot) and organic carbon particles from incomplete combustion (Schumann, 1997). These particles serve as both CCN and ice-nucleating particles (INPs), altering cloud microphysics and the Earth's radiative balance (Twomey, 1977). Sulfate aerosols generally exert a cooling effect by increasing cloud albedo, known as the Twomey effect (Twomey, 1977), whereas black carbon contributes to warming by absorbing solar radiation and reducing surface albedo when deposited on snow and ice (Bond et al., 2013). The net radiative forcing from aviation-related aerosols is complex due to these opposing effects and their interactions with clouds and atmospheric conditions (IPCC, 2021b). Furthermore, carbon monoxide (CO) emissions from aircraft also play a role in atmospheric chemistry by reacting with hydroxyl radicals (OH), thereby affecting the concentration of greenhouse gases and indirectly influencing ozone formation and overall climate forcing (Lee et al., 2021).

As the aviation industry seeks to mitigate its environmental impact, hydrogen has emerged as a promising alternative fuel due to its potential to reduce CO_2 emissions significantly. However, the adoption of hydrogen in aviation introduces new complexities in atmospheric chemistry, particularly concerning its interactions with atmospheric species such as O_3, CH_4, and hydroxyl radicals (OH) (Ehhalt and Rohrer, 2009). Understanding these interactions is crucial for accurately assessing the climate benefits and potential trade-offs associated with hydrogen-powered aviation. Consequently, it becomes imperative to analyze aircraft emissions comprehensively, considering both the reduction of traditional greenhouse gases and the introduction of hydrogen-related emissions that may influence atmospheric chemistry and climate forcing.

2.1.2. Hydrogen as an Alternative Fuel

In contrast to conventional combustion, burning hydrogen (H_2) eliminates carbon emissions such as CO_2, CO, and particulate matter (PM) (Sáez Ortuño et al., 2023). Additionally, hydrogen's main combustion byproduct is H_2O, the climate impact of which is limited when emitted in the troposphere (Lee et al., 2021). This shift supports global decarbonization goals and enhances air quality by reducing pollutant emissions. However, the transition to hydrogen-based aviation faces challenges, including the need for advanced storage solutions, the development of a robust hydrogen supply infrastructure, and ensuring the sustainability of hydrogen production.

Hydrogen Combustion

Hydrogen combustion presents a distinct emissions profile compared to kerosene, primarily due to the absence of carbon and sulfur in hydrogen fuel. Consequently, burning hydrogen does not produce CO_2, CO, unburned hydrocarbons (UHC), soot, or sulfur oxides (SO_X) (Sáez Ortuño et al., 2023). This lack of carbon-based and sulfur-based emissions significantly reduces the pollutants typically associated with kerosene combustion.

However, hydrogen combustion introduces other climate-related emissions, notably H_2O and NO_X, which contribute to contrail formation. Gauss et al. (2003) found that hydrogen-fueled aircraft emit approximately 2.55 times more H_2O per kilogram of fuel compared to kerosene aircraft, with the impact on climate being highly dependent on the altitude of emission. Higher H_2O emissions can increase contrail formation probability, especially at elevated cruise altitudes, nearly doubling atmospheric H_2O accumulation. Additionally, hydrogen combustion results in higher NO_x emissions due to its elevated flame temperature, with a reported 30% increase in NO_x emission indices compared to kerosene (Sáez Ortuño et al., 2023).

The increased H_2O emissions and the reduced presence of soot particles in hydrogen exhaust further influence contrail formation. While higher H_2O emissions enhance the likelihood of contrail formation, Grewe et al. (2017) found that the absence of soot decreases the number of ice nuclei necessary for contrail development, potentially lowering their radiative forcing. Marquart et al. (2005) suggests that hydrogen-induced contrails contribute less to global warming due to fewer ice crystals and reduced optical depth, with estimates of radiative forcing reductions ranging from 30% to 69% under various scenarios. Evidently, the climate impact of adopting hydrogen is complex, extending beyond a straightforward reduction in CO_2 emissions.

Adoption of Hydrogen Technology

In addition to the uncertainty surrounding hydrogen's precise climate impact, other challenges hinder its adoption. Implementing hydrogen as a fuel for aircraft entails significant technical challenges related to storage, energy density, and material compatibility. Hydrogen storage requires either cryogenic temperatures or high-pressure tanks to maintain its liquid or gaseous state, respectively. This introduces substantial weight and volume constraints to the aircraft design (Sáez Ortuño et al., 2023). The lower volumetric energy density of hydrogen compared to conventional jet fuels necessitates larger fuel tanks or more frequent refuelling, potentially reducing payload capacity and limiting the operational range of aircraft (Adler and Martins, 2023).

Hydrogen presents another distinct challenge: its tendency to leak. Fugitive hydrogen emissions primarily occur due to the unique physical properties of hydrogen gas, making it exceptionally challenging to contain. Hydrogen molecules are the smallest and lightest of all gases, possessing high diffusivity that allows them to permeate through materials more readily than larger molecules. This enables hydrogen to escape through microscopic cracks, pores, and even materials impermeable to other gases. Additionally, hydrogen's low viscosity and high buoyancy contribute to its propensity to leak from containment systems, as it can easily flow through small gaps and rise rapidly once released (Züttel, 2003). These characteristics necessitate specialized materials and engineering solutions to prevent leaks, which can be technically challenging and costly.

Fan et al. (2022) estimate that hydrogen leakage rates could range between 1% to 10% across different stages of hydrogen production, storage, transport, and utilization within the hydrogen economy. These estimates take into account the inherent challenges in containing hydrogen due to its small molecular size and high diffusivity, as well as the degradation of infrastructure over time. Values from this study form the basis for the hydrogen leakage estimation used in this thesis.

Assessing the climate benefits of hydrogen adoption is inherently challenging due to the interaction of complex conventional emission effects and the uncertainties associated with hydrogen's climate impacts. While hydrogen combustion mitigates CO_2 emissions, thereby reducing one of the primary drivers of anthropogenic climate change, it simultaneously introduces other climate-relevant emissions such as NO_x and H_2O. Additionally, fugitive hydrogen emissions introduce further uncertainties by potentially increasing CH_4 and O_3 levels through complex atmospheric interactions (Warwick et al., 2022). The necessity for comprehensive accounting becomes evident, as evaluating the net climate impact of hydrogen adoption requires balancing the reductions in CO_2 against the enhancements of NO_x, O_3, and H_2O. An exhaustive assessment of the climate benefit of hydrogen-powered aviation must thus incorporate both direct emissions reductions and indirect effects on various atmospheric species, necessitating modelling frameworks that can capture these interactions across different timescales and spatial scales. Furthermore, the variability in hydrogen leakage rates and the technical challenges in storage and infrastructure add layers of complexity to quantify hydrogen's overall climate benefits accurately.

Another significant challenge in evaluating aviation's impact on climate change is the complex and diverse nature of its emissions, as discussed above. This complexity hinders the direct comparison and aggregation of individual climate effects, making comprehensive assessment difficult. Fuglestvedt et al. (2010) emphasize that without a standardized approach, inconsistencies arise in quantifying and comparing these impacts, impeding the development of effective mitigation strategies. Climate metrics solve this problem by offering a systematic framework to quantify and compare the diverse effects of aviation emissions. By translating atmospheric processes into standardized measures, climate metrics enable the integration of differing timescales, radiative properties, and emissions interactions into a coherent assessment.

2.2. Climate Metrics

Climate metrics are tools used to quantify the effects of gas emissions and aerosols on the climate system or society. They serve as simplified representations of complex climate responses, enabling policymakers and researchers to assess and compare the impacts of different emissions without requiring intricate climate models, which are computationally intensive and require specialized expertise. The design of climate metrics involves a balance between simplicity and the accurate conveyance of the current scientific understanding of climate responses to emissions.

A significant challenge in developing climate metrics is their dependence on the time horizon over which they are considered (Wuebbles et al., 2010; Fuglestvedt et al., 2010; Grewe and Dahlmann, 2015; Ocko and Hamburg, 2022). This dependence becomes evident when comparing the lifetimes of two potent greenhouse gases: CO_2 and methane (CH_4). CO_2 is a long-lived climate pollutant (LLCP) with an atmospheric lifetime ranging from 300 to 1,000 years (Archer et al., 2009). In contrast, CH_4 is a short-lived climate pollutant (SLCP) with a lifetime of approximately 12 years (Prather et al., 2012). The chosen time horizon can significantly influence the emphasis placed on SLCPs versus LLCPs, potentially leading to ambiguities in results and affecting the effectiveness of climate policies (Grewe and Dahlmann, 2015).

Various climate metrics exist, each with its strengths and weaknesses, often tailored to answer specific climate-related questions. In this section, three widely used and relevant climate metrics are discussed in detail: Radiative Forcing (RF), Global Warming Potential (GWP), and Average Temperature Response (ATR).

2.2.1. Radiative Forcing

Radiative forcing is a fundamental concept used to quantify changes in the Earth's energy balance resulting from external perturbations. It is defined as "the change in the net, downward minus upward, radiative flux (expressed in $W\,m^{-2}$) at the tropopause or top of the atmosphere due to a change in an external driver of climate change" (IPCC, 2021b). Essentially, RF measures how much a given factor, such as a greenhouse gas, influences the balance between incoming solar radiation and outgoing infrared radiation, thereby affecting the planet's temperature. Positive RF leads to warming, while negative RF leads to cooling.

Three primary definitions of radiative forcing are distinguished: Instantaneous Radiative Forcing (IRF), Radiative Forcing (RF), and Effective Radiative Forcing (ERF). The differences among these definitions are illustrated in figure 2.2 and summarized below:

- **Instantaneous Radiative Forcing (IRF)**: Refers to the immediate change in radiative flux at the top of the atmosphere due to a perturbation without allowing for any changes in atmospheric temperatures or feedback processes. It provides an initial snapshot of the direct impact of a forcing agent.
- **Radiative Forcing (RF)**: Accounts for the change in net radiative flux after allowing stratospheric temperatures to adjust to radiative equilibrium while keeping surface and tropospheric temperatures fixed at their unperturbed values.
- **Effective Radiative Forcing (ERF)**: Extends RF by including rapid adjustments in the troposphere and at the surface (excluding changes in global mean surface temperature). ERF encompasses changes in clouds, H_2O, and land surface properties.

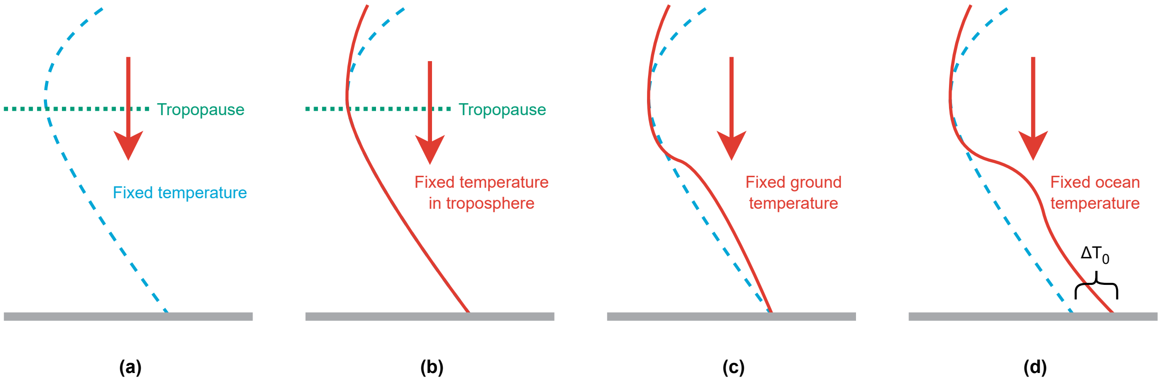


Figure 2.2: Main definitions of radiative forcing, adapted from IPCC (2021b). The definitions are **(a)** instantaneous forcing; **(b)** stratosphere-adjusted forcing; **(c)** zero-surface-temperature-change forcing; **(d)** fixed sea surface temperature forcing, allowing atmospheric and land temperatures to adjust.

Radiative forcing serves as an approximation for the change in global-mean surface temperature at equilibrium, expressed by:

$$\Delta T_s = \lambda \cdot \mathsf{RF}, \tag{2.1}$$

where ΔT_s (K) is the change in global-mean surface temperature, λ ($\mathrm{K\,W^{-1}\,m^2}$) is the climate sensitivity parameter, and RF ($\mathrm{W\,m^{-2}}$) is the radiative forcing. Due to numerous potential interactions and feedbacks, effective comparisons of radiative forcings are typically conducted globally, which may overlook regional or local effects (Wuebbles et al., 2010). Moreover, not all radiative forcings yield the exact global mean temperature change at equilibrium. The relative effectiveness of a climate agent in causing a temperature change, compared to CO_2, is referred to as its efficacy (Wuebbles et al., 2010).

Radiative forcing is the foundational concept for many other climate metrics, such as the GWP (Derwent, 1990). Its primary strength is providing a straightforward measure of changes in the Earth's energy balance due to perturbations. This enables precise and direct assessments of the impacts of various forcing agents, facilitating comparisons between diverse sources like CO_2 and CH_4.

2.2.2. Global Warming Potential

Global Warming Potential is a metric used to compare the relative abilities of different GHG's to trap heat in the atmosphere over a specific time horizon. It quantifies the cumulative radiative forcing of a GHG relative to that of CO_2, the reference gas, over a designated period, typically 20, 100, or 500 years (Derwent, 1990). The GWP of a gas is calculated by integrating the radiative forcing of a unit mass of the gas over the chosen time horizon and comparing it to the integrated radiative forcing of the same mass of CO_2:

$$\mathsf{GWP} = \frac{\int_0^H \mathsf{RF}_i(t)\,dt}{\int_0^H \mathsf{RF}_{\mathsf{CO_2}}(t)\,dt}, \tag{2.2}$$

where $\mathrm{RF}_i(t)$ and $\mathrm{RF}_{\mathrm{CO}_2}(t)$ are the time-dependent radiative forcings of the species of interest and CO_2, respectively, and H is the time horizon.

Introduced in the early 1990s by Rodhe (1990); Derwent (1990) and adopted in the First Assessment Report by the IPCC (1990), Global Warming Potential (GWP) has become the most widely used climate metric in policy-making and research, including studies on hydrogen-induced climate effects (Sand et al., 2023; Warwick et al., 2022; Derwent, 2023; Ocko and Hamburg, 2022; Hauglustaine et al., 2022). Its widespread adoption has solidified its role in assessing climate impacts, particularly in areas like aviation emissions where non-CO_2 gases are significant contributors.

However, GWP faces substantial criticism due to its dependence on the selected time horizon, which can bias assessments toward either short-lived or long-lived pollutants and potentially influence policy decisions (Megill et al., 2024; Grewe and Dahlmann, 2015). Critics argue that the commonly used 100-year horizon (GWP_{100}) may underestimate the impact of short-lived climate pollutants, suggesting the use of multiple horizons like GWP_{20} and GWP_{100} to capture both immediate and long-term effects better. Additionally, the arbitrary selection of the time horizon lacks a strong scientific foundation and may not align with specific climate goals, making GWP less suitable for analyzing the climate impact of gases with short atmospheric lifetimes, such as those associated with hydrogen (Ocko and Hamburg, 2022; Allen et al., 2018).

2.3. Atmospheric Chemistry of Hydrogen

Hydrogen (H_2) is currently present in the atmosphere with a globally averaged mixing ratio of 552.8 ppb (Pétron et al., 2024). Over the past decade, atmospheric hydrogen concentrations have increased significantly, with a rise of 20.2 ppb since 2010 (Pétron et al., 2024). This recent trend contrasts with the historically stable hydrogen levels observed prior to the last decade, indicating that recent increases may be influenced by factors such as large biomass burning events and changes in atmospheric composition (Pétron et al., 2024; Novelli et al., 1999; Ehhalt and Rohrer, 2009). This trend is likely to be enhanced by the anticipated large-scale adoption of hydrogen as an alternative energy carrier, particularly as part of efforts to decarbonize industries and transportation sectors, as discussed in section 2.1.

Atmospheric hydrogen originates from a variety of natural and anthropogenic sources. One of the primary natural sources is the oxidation of methane (CH_4) and volatile organic compounds (VOCs), which contribute around 30% and 25% to the total atmospheric hydrogen budget, respectively (Ehhalt and Rohrer, 2009; Bertagni et al., 2022; Hauglustaine et al., 2022). Direct emissions, such as industrial hydrogen production, fossil fuel combustion, and transportation activities, account for the other 45% of hydrogen in the atmosphere (Bertagni et al., 2022; Hauglustaine et al., 2022). These sources, along with potential future increases in H_2 emissions due to the widespread adoption of hydrogen as an energy source, are expected to shape the future dynamics of atmospheric hydrogen concentrations.

In contrast, the removal of atmospheric hydrogen is primarily driven by the uptake of ground bacteria through deposition. This process accounts for between 65% and 85% of the total H_2 sink and is conversely the largest uncertainty in its atmospheric budget. (Novelli et al., 1999; Ehhalt and Rohrer, 2009). The remaining hydrogen is oxidized by hydroxyl radicals (OH) through the following reaction:

$$\mathrm{H_2 + OH \longrightarrow H + H_2O} \tag{2.3}$$

Hydroxyl radicals play a crucial role in atmospheric chemistry as the primary oxidizing agent, often referred to as the ”detergent of the atmosphere” (Levy, 1971). They are highly reactive and contribute to the breakdown of a variety of atmospheric pollutants and trace gases, including methane and carbon monoxide (CO) (Prather et al., 2012).

Through its oxidation by hydroxyl radicals, excess atmospheric hydrogen indirectly contributes to the greenhouse effect in several ways: (1) by increasing methane abundance, (2) by enhancing the production of tropospheric ozone (O_3) and altering stratospheric ozone levels, and (3) by increasing stratospheric water vapour (SWV) concentrations. The oxidation of hydrogen consumes hydroxyl radicals (OH), reducing their availability for oxidizing methane. This decrease in OH concentration extends the atmospheric lifetime of methane, thereby increasing its overall abundance. Increases in CH_4 also cause an increase in atmospheric O_3. Additionally, the oxidation of hydrogen leads to higher levels of hydrogen oxides (HO_x), which promote the formation of O_3 in the troposphere through photochemical reactions. In the stratosphere, increased HO_x levels can accelerate ozone depletion processes, affecting stratospheric ozone concentrations. Moreover, the oxidation of hydrogen generates water vapor in the stratosphere, which can contribute to warming and affect ozone chemistry. The following text will provide a detailed discussion of these effects, including an overview of the underlying mechanisms and current modelling approaches for assessing hydrogen's impact on methane, ozone, and stratospheric water vapour.

2.3.1. Methane

Methane is the second most important greenhouse gas after CO_2. Despite its lower atmospheric concentration compared to CO_2, methane exhibits a global warming potential approximately 28-34 times greater over a 100-year timescale (IPCC, 2021b). This enhanced efficacy in trapping heat makes methane a potent contributor to radiative forcing, disproportionate to its abundance. Methane primarily originates from natural sources such as wetlands and anthropogenic activities like agriculture, fossil fuel extraction, and waste management (IPCC, 2021b).

Modelling of Atmospheric Methane

In 1993, Don Fisher presented a problem: chemical interactions between methane, carbon monoxide, and hydroxyl radicals could extend the atmospheric lifetime of methane perturbations. Fisher's 1995 study, using a two-dimensional global chemistry-transport model, showed that a CH_4 pulse decayed more slowly than expected, with a recovery time of about 12 years instead of the predicted 8 years. This delayed recovery was found to be constant, regardless of perturbation size, suggesting that it was not due to a nonlinear chemical response. Earlier, Isaksen and Hov (1987) had observed a CH_4-OH feedback, where a slight increase in CH4 emissions caused a disproportionate rise in CH_4 levels. Still, the connection between this feedback and Fisher's findings was not initially apparent.

In response to the problem posed by Fisher, Prather (1994) developed a general box model capable of demonstrating the effect of chemical feedback on methane perturbations. Zero-dimensional box models are suitable for simulating the mixing of species with long atmospheric lifetimes compared to the inter-hemispheric mixing time. This approach works well for methane (with a lifetime of about 9.6 years) and hydrogen (with a lifetime of around 2 years), given the typical inter-hemispheric mixing time of approximately 1 year (Warwick et al., 2022). Within Prather's model, the chemical coupling <CH4-CO-OH> is modelled using three continuity equations as shown in equation (2.4), equation (2.5) and equation (2.6). The model, governed by the reactions

$$\frac{d[CH_4]}{dt} = S_{CH_4} - k_1[OH][CH_4], \tag{2.4}$$

$$\frac{d[CO]}{dt} = S_{CO} + k_1[OH][CH_4] - k_2[OH][CO], \tag{2.5}$$

$$\frac{d[OH]}{dt} = S_{OH} - k_1[OH][CH_4] - k_2[OH][CO] - k_3[OH][X], \tag{2.6}$$

describes the time-dependent concentrations of methane, carbon monoxide, and hydroxyl radicals by incorporating the key production and loss processes governing these species. Methane is oxidized by hydroxyl radicals, leading to the formation of carbon monoxide, which in turn reacts with hydroxyl radicals to form CO_2. This creates a dynamic feedback loop in which the presence of methane and carbon monoxide influences the availability of hydroxyl radicals, which are also responsible for their removal from the atmosphere.

The model highlights how perturbations in methane concentrations can lead to changes in the concentrations of carbon monoxide and hydroxyl radicals, which in turn affect the removal rates of methane. This feedback process ultimately influences the atmospheric lifetime of methane and its impact on the Earth's radiative balance. Additionally, the model captures the nonlinear nature of these interactions, showing that increases in methane can lead to a depletion of hydroxyl radicals, which reduces the efficiency of methane removal and prolongs its atmospheric lifetime. Similarly, carbon monoxide plays a role in consuming hydroxyl radicals, further impacting methane oxidation. By representing these chemical feedbacks, Prather showed how methane perturbations propagate through the system and how changes in one component, such as methane or hydroxyl radicals, can affect the broader atmospheric oxidative capacity and greenhouse gas lifetimes. Ultimately, findings from this study led to a +40% revision of the GWP of methane (Prather, 1994).

Warwick et al. (2022) adapted Prather's box model to incorporate H_2. The extended model includes additional terms to represent the sources and sinks of hydrogen and its interactions with hydroxyl radicals. Prather's model is adapted in the following way:

$$\frac{d[OH]}{dt} = ... - k_4[OH][H_2], \tag{2.7}$$

$$\frac{d[H_2]}{dt} = S_{H_2} - k_4[OH][H_2] - \frac{1}{\tau_{dep}}[H_2], \tag{2.8}$$

where the new equation (2.8) describes the time evolution of hydrogen concentrations due to emissions, together with its atmospheric loss through OH and an additional deposition sink modelled using a lifetime. The inclusion of hydrogen introduces an additional sink for hydroxyl radicals, influencing the oxidative capacity of the atmosphere. As hydrogen reacts with OH, it reduces the availability of OH for oxidizing methane and carbon monoxide, leading to an increase in the atmospheric lifetime of methane. This interaction creates a complex feedback loop: elevated H_2 levels decrease OH concentrations, which slows down the removal of CH_4, resulting in higher methane concentrations that further consume OH. This nonlinear chemical coupling emphasizes the need to consider hydrogen when evaluating the atmospheric implications of transitioning to a hydrogen-based energy system.

Warwick et al. (2022) utilized this extended box model to calculate the boundary condition values of H_2 and CH_4 for three-dimensional chemistry-climate simulations performed with the UKESM1 Earth system model. This enables the model to capture the spatial variability of these species while ensuring their global concentrations align with the specific scenarios used in the simulation.

Bertagni et al. (2022) uses a similar approach, but with some slight adaptations to account for CH_4's feedback on H_2, and soil uptake, stratospheric loss and reactions of chlorine radicals. This results in the following additions to Prather's model:

$$\frac{d[CH_4]}{dt} = ... - k_s[CH_4], \tag{2.9}$$

$$\frac{d[OH]}{dt} = ... - k_4[OH][H_2], \tag{2.10}$$

$$\frac{d[H_2]}{dt} = S_{H_2} + \alpha[OH][CH_4] - k_4[OH][H_2] - k_d[H_2], \tag{2.11}$$

where k_s denotes the reaction rate of CH_4's additional sinks, k_d the reaction rate due H_2 deposition, and α the feedback yield of CH_4 on H_2. It should be noted that the parameter τ_{dep} used in Warwick's model, and k_d used in Bertagni's model essentially denote the same loss.

Sand et al. (2023) uses an ensemble of fully coupled climate models to assess the climate impact H_2 emissions. The use of a box model is circumvented by dividing the methane burden by its total lifetime, which includes losses from reactions with OH, the stratosphere, and soil sinks. The methane flux difference between experiments a H_2 perturbation experiment and a control experiment reflects the additional emissions needed to sustain methane levels, allowing for the calculation of feedback factors and the impact of hydrogen on methane concentrations.

The box models used in the first two studies mentioned provide a versatile, simple, and computationally efficient way of determining the principle effect of hydrogen emissions on the tropospheric chemistry of its key derivatives. Bertagni et al. (2022) investigated a pulse emission of hydrogen, resulting in a 10% increase in its concentration. This perturbation causes an instantaneous drop in OH concentrations and subsequent build-up in CH_4 that lasts several years, after which it decays similarly to a direct pulse in CH_4 emissions. The timescale of this decay is defined by the CH_4 feedback effect, modelled using α by Bertagni et al. (2022) but omitted in the model by Warwick et al. (2022). However, both studies show that the atmospheric response to hydrogen emissions is relatively insignificant compared to the oxidation of H_2 by soil bacteria. Both box models are thus sensitive to assumptions made on the tropospheric lifetime of H_2 due to soil deposition of hydrogen.

Box models are primarily used to provide the steady-state response of atmospheric species to hydrogen emissions under simple scenarios, either as boundary conditions for more detailed climate chemistry models or to determine climate metrics. However, their accuracy diminishes when more complex scenarios are considered. This is because box models assume constant reaction rates, even though these rates depend on temperature. Furthermore, they assume constant values for various sinks, while real-world interactions are much more complex (Winterstein and Jöckel, 2021). Therefore, detailed climate models, such as MRI-ESM2-0 discussed in section 2.4, model the climate chemistry and sinks of key species more accurately, albeit at the disadvantage of high computational costs.

Moreover, perturbations in CH_4 levels affect other atmospheric greenhouse gases, such as stratospheric water vapour and ozone. Due to the inherent assumption that a box model is zero-dimensional, it cannot model the transport of methane to the stratosphere, where it generates water vapour. Additionally, because ozone has a short tropospheric lifetime, box models cannot adequately capture its dynamics, preventing it from being able to capture the full climate response of H_2 emissions.

Stratospheric Water Vapour

Increased methane concentrations can lead to higher levels of stratospheric water vapour through chemical processes that occur as methane oxidizes in the atmosphere. CH_4 is gradually broken down by OH in the troposphere, producing H_2O. Due to the abundance of H_2O in the troposphere, the impact is insignificant. Some CH_4 molecules, however, are transported to the stratosphere, where they undergo further oxidation primarily via reactions with OH and excited oxygen atoms $O(^1D)$ (Winterstein and Jöckel, 2021). This causes the release of additional water vapour directly into the stratosphere. According to recent estimates, this process has significant implications for the stratospheric water budget, as methane oxidation accounts for approximately 15-20% of the total water vapour in the stratosphere (Solomon et al., 2010). The injection of water vapour into the stratosphere from methane oxidation contributes to radiative forcing, as water vapour is a potent greenhouse gas that can amplify global warming by trapping outgoing longwave radiation (IPCC, 2021b).

Nonetheless, simulations performed by Warwick et al. (2022) indicate that the total radiative forcing attributable to H_2–induced stratospheric water vapour perturbation is caused by non-methane induced chemical reactions. This finding agrees with earlier research by Paulot et al. (2021). It highlights that the climate impact of hydrogen is influenced by various atmospheric processes beyond methane interactions, underscoring the necessity for comprehensive modelling approaches to assess hydrogen's overall effect on climate change accurately.

2.3.2. Ozone

Ozone (O_3) is a triatomic molecule of oxygen that plays a critical role in Earth's atmosphere, exhibiting dual characteristics depending on its altitudinal distribution. In the stratosphere, ozone forms the ozone layer, which absorbs and scatters the majority of the Sun's harmful ultraviolet (UV) radiation, protecting living organisms from potential genetic damage and other adverse effects (Molina and Rowland, 1974). Conversely, in the troposphere, ozone is a secondary pollutant formed by photochemical reactions involving nitrogen oxides NO_X and volatile organic compounds (VOCs), contributing to urban smog and significant health risks to humans and ecosystems. The atmospheric chemistry of ozone is complex and influenced by natural processes and anthropogenic emissions, making it a critical subject of study for understanding environmental impacts and climate change and developing effective air quality management strategies (IPCC, 2021b).

Ozone undergoes continuous cycles of formation and destruction through complex photochemical reactions that play a significant role in climate regulation. In the stratosphere, ozone formation begins with the photodissociation of molecular oxygen (O_2) by ultraviolet solar radiation, producing two oxygen atoms. These highly reactive oxygen atoms subsequently collide with O_2 molecules in the presence of a third body to form ozone. The ozone molecules can absorb UV radiation, leading to their photolysis back into an oxygen molecule and an oxygen atom, thus perpetuating the ozone-oxygen cycle (Solomon et al., 2010).

Destruction mechanisms of ozone involve catalytic cycles mediated by trace gases. In the stratosphere, natural and anthropogenic gases containing chlorine and bromine, such as chlorofluorocarbons (CFCs), release halogen atoms upon photolysis, which then catalyze the breakdown of ozone into molecular oxygen (Molina and Rowland, 1974). NO_X and HO_X also participate in catalytic cycles that deplete ozone by facilitating reactions where ozone is converted back to O_2 without the consumption of the catalyst. In the troposphere, ozone formation occurs through photochemical reactions involving precursors like NO_X and VOCs under sunlight, leading to ozone accumulation at ground level, which is a major component of photochemical smog (Monks et al., 2015).

Ozone as a Greenhouse Gas

Beyond its well-known roles in shielding the Earth from ultraviolet radiation and contributing to urban smog, ozone also functions as a potent greenhouse gas (GHG) that significantly affects the Earth's radiative balance. Tropospheric ozone, generated through photochemical reactions involving pollutants like NO_X and VOCs, absorbs infrared radiation emitted from the Earth's surface, thereby trapping heat in the atmosphere and contributing to global warming (IPCC, 2021b). Although its concentration is much lower than that of CO_2, ozone has a stronger per-molecule radiative forcing effect, making it an important short-lived climate forcer (IPCC, 2021b). Stratospheric ozone also influences climate by affecting the temperature structure of the upper atmosphere, which can alter atmospheric circulation patterns and, consequently, surface climate (IPCC, 2021b).

Tropospheric ozone has a pronounced seasonal variability and varied spatial distribution that contrasts sharply with the uniform distribution of methane. This variability is largely driven by the complex interaction of photochemical production, precursor emissions, and atmospheric transport mechanisms. During the summer months in mid-latitude regions, increased solar radiation enhances photochemical reactions involving NO_X and VOCs, leading to higher concentrations of ozone (Monks et al., 2015). In contrast, during the winter months, reduced sunlight and lower temperatures diminish photochemical activity, resulting in decreased ozone production.

The spatial distribution of tropospheric ozone is highly variable due to localized sources of its precursors and regional atmospheric conditions. Industrialized and urban areas with substantial NO_X and VOC emissions often experience elevated ozone levels, particularly under stagnant atmospheric conditions that limit dispersion (Cooper et al., 2014). Additionally, long-range transport can redistribute ozone and its precursors, contributing to elevated background ozone levels in downwind regions (Cooper et al., 2014). Stratosphere-troposphere exchange processes also influence ozone distribution by injecting ozone-rich air from the stratosphere into the troposphere, particularly at mid to high latitudes during spring.

An additional significant source of NO_X emissions influencing tropospheric ozone distribution is aviation. Aircraft emit NO_X directly into the upper troposphere and lower stratosphere during cruise phases, where background NO_X concentrations are relatively low, enhancing the efficiency of ozone production in these regions (Lee et al., 2021). The physical basis behind this involves a series of photochemical reactions initiated by emitted NO_X. In the upper atmosphere, NO reacts with hydroperoxy (HO_2) and organic peroxy radicals (RO_2), produced from the oxidation of carbon monoxide (CO) and volatile organic compounds (VOCs), to form nitrogen dioxide (NO_2) (Maruhashi et al., 2024). NO_2 then photolyzes under ultraviolet sunlight to produce ozone (O_3) and regenerates NO, which can participate in further reactions, creating a catalytic cycle that efficiently generates ozone (Maruhashi et al., 2024).

The efficiency of ozone formation from aviation NO_X emissions is influenced by factors such as altitude, latitude, and atmospheric conditions. Higher altitudes receive more intense solar radiation and have lower temperatures, conditions that favour the photolysis of NO_2 and subsequent ozone production (IPCC, 2021b). Additionally, regions with low background NO_X levels and abundant ozone precursors experience a greater increase in ozone concentrations due to the nonlinear chemistry of NO_X and ozone formation (Maruhashi et al., 2024). Understanding these physical mechanisms is crucial for developing strategies to mitigate aviation's impact on climate. For instance, adjusting flight altitudes and routes to minimize NO_X emissions in sensitive regions can reduce ozone formation and its associated radiative forcing (Grewe, 2020).

Effect of Hydrogen Emissions on Tropospheric Ozone

Hydrogen emissions have a significant impact on tropospheric ozone levels through their interactions with methane, hydroxyl radicals (OH), and hydroperoxyl radicals (HO_2). Elevated methane levels, a direct consequence of H_2 emissions as described in section 2.3.1, contribute to increased ozone formation in the troposphere because methane oxidation produces ozone precursors that participate in photochemical reactions leading to ozone generation (Warwick et al., 2022). Additionally, the increase in HO_2 radicals from hydrogen oxidation plays a crucial role in ozone chemistry. HO_2 reacts with NO_X to regenerate OH and produce NO_2, which photolyzes under sunlight to release oxygen atoms that combine with molecular oxygen O_2 to form ozone. This mechanism implies that higher HO_2 concentrations can enhance ozone production in NO_X-rich environments. Sand et al. (2023) demonstrated that increased hydrogen emissions could lead to a net rise in tropospheric ozone levels due to these interactions.

Warwick et al. (2022) investigated the atmospheric implications of increased H_2 emissions, as discussed in section 2.3.1. The study found that the processes described above result in a linear increase in tropospheric ozone burden with rising H_2 surface mixing ratios, a relationship that holds across various emission scenarios and climate conditions. In contrast, the study showed that increased hydrogen emissions have a minimal impact on stratospheric ozone and the overall recovery of the ozone layer. While there are minor decreases in stratospheric ozone concentrations, particularly in the upper stratosphere and polar regions due to enhanced HO_X-mediated ozone depletion, these changes are not statistically significant when methane responses are considered. Total column ozone changes remain within the variability range of climate model projections, suggesting that even substantial increases in atmospheric H_2 are unlikely to significantly affect ozone recovery.

Estimates for hydrogen's Global Warming Potential (GWP) have become more refined over time as research has detailed hydrogen's indirect impacts. Early research by Derwent et al. (2020) underestimated the GWP of H_2 due to othe mittance of its effects on SWV. However, recent studies by Warwick et al. (2022) and Sand et al. (2023) offer more nuanced assessments. Warwick et al. (2022) calculate a total GWP_{100} for hydrogen at around 11.5, with a range of 6–18 due to various atmospheric factors. This includes tropospheric and stratospheric contributions, where the stratospheric effect of hydrogen on water vvapouraccounts for approximately a quarter to a third of the overall GWP_{100}.

Sand et al. (2023) evaluated the GWP_{100} for hydrogen using an ensemble of fully coupled climate models, providing a robust analysis of hydrogen's indirect climate impacts. Their study found a central GWP_{100} value of approximately 12, with an uncertainty range from 6 to 20. This range reflects variability across models due to different atmospheric conditions, chemical feedback, and the radiative impacts of hydrogen on methane, ozone, and stratospheric water vapour. The multi-model ensemble approach allowed for a more comprehensive assessment by capturing both tropospheric and stratospheric components, establishing a higher confidence in these GWP estimates and underscoring hydrogen's significant indirect effects on climate forcing through its interactions with other atmospheric constituents.

Both studies indicate that H_2's climate impact is expected to be dependent on background concentrations of certain species, such as CH_4. However, simulating the wide range of possible climate scenarios using the fully coupled climate models used in these studies is computationally expensive and time-consuming. As such, hydrogen's effects must be explored through surrogate or simplified models that approximate the interactions of H_2 with atmospheric species under different conditions. These models enable efficient evaluation of potential impacts across various scenarios while capturing the key feedback mechanisms associated with hydrogen's indirect effects on greenhouse gases like methane and ozone. The following section examines the use of scenario modelling in climate research.

2.4. Climate Scenario Modelling

Addressing the complexities of climate change necessitates robust tools to project future climatic conditions under varying socioeconomic trajectories. Scenario modelling serves this critical function in climate research by allowing scientists to simulate and analyze potential future states of the climate system. These models help in understanding the implications of different greenhouse gas emission pathways, guiding policymakers in formulating effective mitigation and adaptation strategies. By exploring a range of plausible scenarios, researchers can assess uncertainties and identify critical factors influencing climate outcomes (Moss et al., 2010).

2.4.1. Shared Socioeconomic Pathways (SSPs)

The Shared Socioeconomic Pathways (SSPs) are a set of five integrated scenarios that describe possible future developments in demographics, economics, technology, and environmental policies over the 21st century. They serve as a foundation for climate change research by providing narratives and quantitative data that model how global society, demographics, and economics might change, thereby influencing greenhouse gas emissions and the capacity to mitigate and adapt to climate change (O'Neill et al., 2014).

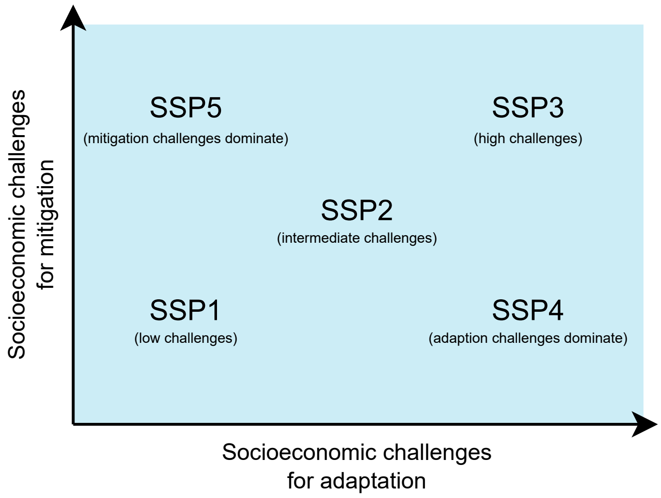


Figure 2.3: SSPs mapped in the challenges to mitigation/adaptation space, adapted from O'Neill et al. (2017).

The SSPs were developed to address the limitations of previous scenarios by separating socioeconomic projections from climate policy assumptions. This separation allows researchers to analyze the implications of different socioeconomic futures independently of climate policies, providing a more flexible framework for integrated assessments (Ebi et al., 2014).

- **SSP1 (Sustainability)**: SSP1 envisions a world shifting rapidly towards sustainable practices. Education and health investments lead to slow population growth, and there is a strong emphasis on human well-being and environmental stewardship. Technological progress is directed towards green innovations, energy efficiency improves, and reliance on fossil fuels decreases significantly. Challenges to both mitigation and adaptation are low due to proactive environmental policies and global cooperation (Riahi et al., 2017).
- **SSP2 (Middle of the Road)**: In SSP2, historical trends continue without significant shifts. Population growth is moderate, and economic development proceeds unevenly across regions. Energy systems remain diverse, with gradual improvements in technology and energy efficiency. Environmental degradation slows but does not stop, and there are modest efforts to reduce resource and energy intensity. Challenges are intermediate (O'Neill et al., 2017).
- **SSP3 (Regional Rivalry)**: SSP3 portrays a world characterized by nationalism and regional conflicts. Policies are oriented towards security and competitiveness rather than global cooperation. Economic growth is slow, especially in developing countries, and population growth is high due to limited investments in education and healthcare. Reliance on domestic energy sources leads to continued use of fossil fuels. Challenges to mitigation and adaptation are high because of the lack of international collaboration and slow technological progress (Fujimori et al., 2017).
- **SSP4 (Inequality)**: This scenario represents a world with increasing inequality, both within and between countries. A wealthy global elite drives rapid technological innovation and economic growth, while the majority of the population experiences decline. Access to education and healthcare is unequal, leading to disparities in population growth and vulnerability to climate change. Challenges to adaptation are high due to the vulnerability of disadvantaged groups, while challenges to mitigation are low to moderate as the elite adopt low-carbon technologies (Calvin et al., 2017).
- **SSP5 (Fossil-fueled Development)**: SSP5 envisions rapid economic growth powered by abundant fossil fuel resources. There is strong faith in technological solutions and minimal emphasis on environmental protection. Population growth is low due to high levels of education and urbanization. Energy consumption is high, leading to significant greenhouse gas emissions. Challenges to mitigation are high due to the dependency on fossil fuels, but adaptation challenges are low because of the wealth and technological advancements available to address climate impacts (Kriegler et al., 2017).

Each SSP outlines varying assumptions about population growth, economic development, technological progress, and environmental policies, which collectively influence greenhouse gas emissions and the capacity to address climate change.

The aviation sector plays a pivotal role in Shared Socioeconomic Pathway (SSP) modelling due to its substantial and increasing contribution to global greenhouse gas emissions. A multitude of factors, including economic growth, income levels, population dynamics, technological advancements, and policy interventions, shape aviation demand. Under SSP1 (Sustainability – Taking the Green Road), a shift towards sustainable practices fosters advancements in alternative fuels, such as biofuels and synthetic fuels, alongside significant improvements in aircraft efficiency. Policy measures encourage shifts to less carbon-intensive transportation modes wherever feasible. Consequently, even with enhanced connectivity, overall aviation emissions decline due to these technological and behavioural transformations. In contrast, SSP2 (Middle of the Road) envisions moderate growth in air travel aligned with prevailing economic trends. Technological enhancements proceed at historical rates, and policy efforts fail to markedly restrain emissions, leading to a gradual increase in aviation emissions, albeit less pronounced than in high-growth scenarios.

SSP3 (Regional Rivalry – A Rocky Road) presents a fragmented global landscape characterized by economic stagnation and regional conflicts curtailing international travel demand. However, the absence of technological advancements and continued reliance on inefficient aircraft technologies result in only minimal reductions in emissions. SSP4 (Inequality – A Road Divided) is marked by pronounced disparities, where the affluent segment of the population engages in extensive travel, often utilizing private aviation. At the same time, the majority have restricted access to air travel. Technological improvements primarily benefit the elite, yielding more efficient aircraft within this group, yet overall emissions may not decrease due to the increased frequency of travel among the wealthy (Calvin et al., 2017). Lastly, SSP5 (Fossil-fueled Development – Taking the Highway) is characterized by robust economic growth and globalization, driving a significant surge in air travel demand. Although technological innovations are realized, they are insufficient to offset the rapid increase in travel volume. In the absence of stringent policies, aviation emissions escalate substantially, thereby contributing to heightened radiative forcing (Kriegler et al., 2017).

Methane (CH_4) and ozone (O_3) are critical components in the atmospheric chemistry framework of Shared Socioeconomic Pathways (SSPs) due to their significant roles in climate forcing and atmospheric interactions. Methane, a potent greenhouse gas, possesses a global warming potential substantially higher than CO_2 over 100 years. In SSP modelling, methane emissions are projected based on various anthropogenic and natural activities, including livestock farming, rice cultivation, fossil fuel extraction, waste management, and biomass burning. Under SSP1, aggressive mitigation strategies and technological innovations result in substantial reductions in methane emissions. This scenario benefits from improved agricultural practices, dietary shifts away from methane-intensive foods, and enhanced waste management systems, collectively contributing to the decline in methane levels (Rogelj et al., 2018). Conversely, SSP3 is characterized by persistent high methane emissions due to continued reliance on traditional agricultural methods, limited technological adoption, and minimal environmental regulations. Similarly, SSP5 maintains elevated methane emissions driven by increased energy demand and agrarian production unless significant technological breakthroughs in emission control are implemented.

Tropospheric ozone, a secondary pollutant and greenhouse gas, forms from precursors such as nitrogen oxides (NO_x), carbon monoxide (CO), and volatile organic compounds (VOCs). The SSPs influence ozone levels through variations in these precursor emissions, which are affected by changes in energy production, transportation, industrial activities, and biomass burning. In the SSP1 scenario, stringent emission controls and the adoption of clean technologies lead to a reduction in precursor emissions, resulting in lower ozone concentrations and improved air quality. In contrast, both SSP3 and SSP5 scenarios experience increased emissions of ozone precursors due to high fossil fuel utilization and lax environmental policies, which elevate ozone levels and exacerbate climate warming and associated health impacts (Turnock et al., 2020).

2.4.2. Scenario Model Intercomparison Project (ScenarioMIP)

The Coupled Model Intercomparison Project Phase 6 (CMIP6) is a collaborative framework that coordinates climate modelling experiments to enhance understanding of climate variability and change (Eyring et al., 2016). Within CMIP6, the Scenario Model Intercomparison Project (ScenarioMIP) provides a suite of future climate projections based on the SSPs. ScenarioMIP aims to explore a range of plausible future climates resulting from different emissions trajectories, facilitating the assessment of climate responses and enabling cross-comparison among models.

The most extensive climate projections in ScenarioMIP have been performed using the MRI-ESM2.0 model, and results from this model form the basis of this research. This model, developed by the Meteorological Research Institute in Japan, is a comprehensive Earth system model designed to simulate the interactions between the atmosphere, ocean, land surface, and biosphere, encompassing both chemical and physical processes that influence the climate (Yukimoto et al., 2019). This model integrates the Global Spectral Model with high horizontal resolution, providing detailed representations of cloud microphysics, radiation processes, and aerosol-cloud interactions. Its oceanic processes are simulated using the MRI Community Ocean Model, which captures circulation, heat transport, and biogeochemical cycles. The land surface dynamics are accounted for by the Minimal Advanced Treatments of Surface Interaction and Runoff model, which includes vegetation dynamics, carbon and nitrogen cycles, and energy fluxes. Additionally, the model incorporates sea ice dynamics and thermodynamics to simulate polar climate processes accurately. An atmospheric chemistry module within MRI-ESM2.0 models key chemical species such as ozone, methane, and aerosols, incorporating detailed reaction mechanisms and transport processes. The MRI-ESM2.0 has been extensively utilized in ScenarioMIP simulations, contributing to a broad spectrum of experiments that evaluate future climate projections under various SSP scenarios.

3

Methodology

This chapter outlines the methodological framework employed to assess the future climate impact of fugitive hydrogen emissions from aviation. The aim is to develop a robust model that can rapidly evaluate the perturbation of methane, tropospheric ozone and stratospheric water vapour due to excess hydrogen mixing ratios in the context of future climate projections.

First, an overview of the methodology is set out in section 3.1. The approach used to model hydrogen emissions and subsequent perturbations of atmospheric hydrogen is described in section 3.2. The modelled response of methane, ozone and stratospheric water vapour to excess hydrogen emissions is explained in section 3.3, after which section 3.4 provides the method used to calculate their respective radiative forcings. Finally, section 3.5 describes the Monte Carlo simulation used to quantify the uncertainties associated with hydrogen emissions.

3.1. Overview of Method

The primary goal of this research is to evaluate the climate impact of fugitive hydrogen (H_2) emissions in the context of future emissions scenarios. As mentioned in section 2.3, H_2 is an indirect greenhouse gas, with its climate impact mainly governed through its perturbation of methane (CH_4), ozone (O_3) and stratospheric water vapour (SWV). The complexity of the interaction between H_2 and these greenhouse gases, together with the interaction of these species with other agents in the atmosphere, makes an accurate assessment of the climate impact of H_2 difficult. Moreover, increased H_2 emissions are linked to a simultaneous reduction in other greenhouse gas emissions, as H_2 is frequently advocated as a cleaner alternative to fossil fuels. This transition can result in lower emissions of carbon dioxide (CO_2), methane (CH_4), and other greenhouse gases commonly released during the combustion of conventional energy sources. Therefore, a thorough assessment of the future climate impact of H_2 emissions should encompass a detailed simulation of its atmospheric chemistry and account for the net effect of reduced emissions from conventional energy sources through exhaustive scenario analysis.

However, the computational expense associated with full-scale climate simulation limits its feasibility in scenario modelling, where multiple simulations over long timescales are required to capture the full range of possible outcomes. Therefore, models designed for rapid scenario assessments aim to approximate atmospheric and climate responses using nonlinear relationships between emissions and climate indicators. The model into which the method developed in this thesis will be integrated, OpenAirClim, simplifies the evaluation of air traffic emissions' impact by utilising pre-calculated response surfaces stored in look-up tables. Instead of relying on complex climate-chemistry simulations, it reproduces the atmospheric and climate responses through these simplified relationships. This approach is referred to as surrogate modelling. In general, the goal of surrogate modelling is to use data-driven methods to approximate complex physical or computational processes with reduced computational cost while maintaining sufficient accuracy for predictive or scenario-based analysis.

A similar approach will be employed in this study. Using data from the Coupled Model Intercomparison Project (CMIP6), a surrogate model will be developed to simulate the effects of H_2 emissions under various future climate scenarios. One of the primary activities within CMIP6, ScenarioMIP, provides simulations of a wide range of emission pathways. These simulations are leveraged to capture the diverse potential interactions of CH_4, O_3, and other atmospheric gases in future climate scenarios. Nonetheless, ScenarioMIP primarily focuses on projections based on emissions of major greenhouse gases such as CO_2, N_2O, and CH_4 as the climate impact of H_2 is indirect. While interactions involving

H_2 and other atmospheric species are considered in the ScenarioMIP model, increases in H_2 emissions are not. Due to this omission, the surrogate model must be supplemented with another method capable of modelling H_2 emissions.

As described in section 2.3, H_2 emissions lead to a direct perturbation of hydroxyl (OH), which in turn affect the atmospheric levels of CH_4, O_3, and SWV. By quantifying H_2 emissions in terms of the associated loss or production of OH and subsequently using the perturbed mixing ratios of this oxidant as input to the surrogate model, assessing the impact of H_2 using CMIP6 data becomes possible. A general framework of the methodology is shown in Figure 3.1.

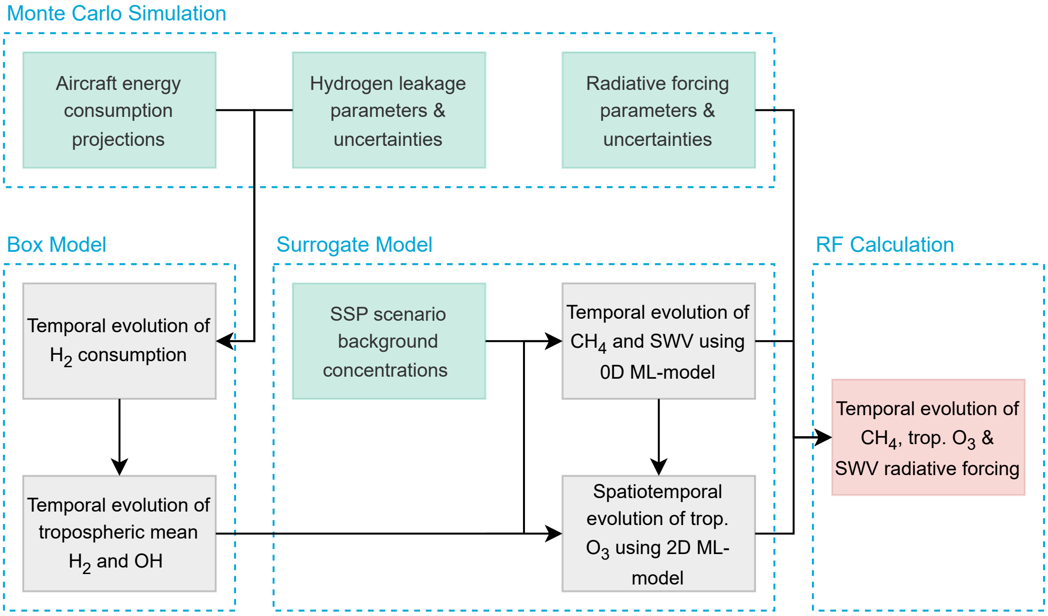


Figure 3.1: Overview of the modelling approach to derive the climatic impact of H_2 emissions. Green blocks denote inputs to the model, grey intermediate results, and red the final model result. The model is divided into four segments: the Monte Carlo Simulation, the box model, the machine learning (ML) based surrogate model, and the radiative forcing calculation.

1. **H_2 emission model**: The first step involves modelling the atmospheric perturbation of H_2 due to projected increases in emissions. Future aircraft fuel consumption scenarios are converted into projected H_2 consumption scenarios by assuming a specific temporal evolution of H_2 adoption in the aviation sector. The corresponding H_2 emissions are then estimated by applying an assumed leakage factor, representing the proportion of H_2 that escapes into the atmosphere. A box-model approach is employed to quantify the total atmospheric perturbation of H_2 due to emissions. This method is described in section 3.2.
2. **Surrogate model**: The second step uses the temporal perturbation of OH in the troposphere due to excess H_2 as input for the surrogate models. Two distinct surrogate models are developed: a zero-dimensional temporal model to estimate scenario-specific changes in mean tropospheric CH_4 and SWV, and a spatiotemporal model designed to evaluate the impact of H_2 emissions on tropospheric O_3. These models are trained using Shared Socioeconomic Pathways (SSP) data from the CMIP6 ScenarioMIP project, allowing them to predict atmospheric changes resulting from perturbed OH mixing ratios. These models' detailed development and application are discussed in section 3.3.
3. **Radiative forcing calculation**: In the third step, the temporal changes in CH_4, O_3, and SWV levels are converted into their respective effective radiative forcings. The methodology to translate these greenhouse gas mixing ratios into effective radiative forcing values is described in section 3.4.

4. **Monte Carlo simulation**: To account for the inherent uncertainties associated with the models, input parameters, and projected scenarios used in this analysis, a Monte Carlo simulation is conducted. This probabilistic approach samples a wide range of potential inputs to assess the sensitivity and range of possible outcomes, quantifying the effect of uncertainties. A detailed explanation of the Monte Carlo simulation is presented in section 3.5.

Two separate surrogate models are trained to estimate the atmospheric changes in CH_4, SWV and O_3 resulting from H_2 emissions. The effect of H_2 emissions on SWV is determined by combining the methane-induced SWV perturbation from the surrogate model with an empirical relation accounting for the non-methane-induced SWV. The following equation is used:

$$d[SWV]_{non-CH_4} = A_{SWV} \cdot d[H_2], \tag{3.1}$$

where A_{swv} is a conversion factor used to express the increase in SWV per increase in H_2 levels, equal to 0.3 $ppb_{SWV}\ ppb_{H_2}^{-1}$ (Warwick et al., 2022). Figure 3.2 provides a simplified overview of the perturbed species modelled in this work.

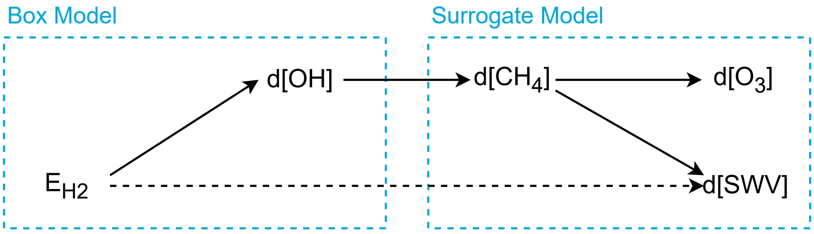
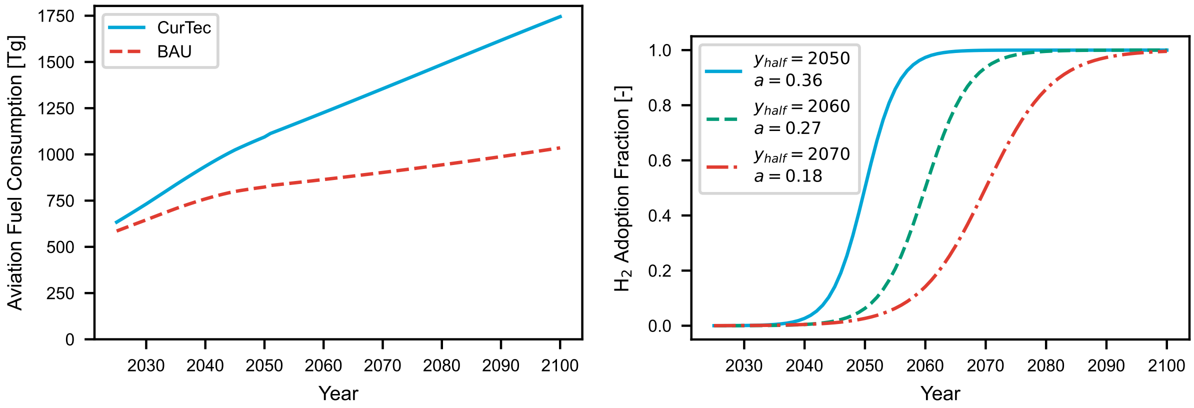


Figure 3.2: Simplified overview of how changes in CH_4, O_3 and SWV mixing ratios are modelled. Using a box model, hydrogen emissions are converted to equivalent perturbations in tropospheric OH mixing ratio. This is used as an input for the surrogate model to compute the subsequent perturbation in CH_4, O_3 and SWV. The non-methane-induced contribution of H_2 emissions to SWV is modelled using equation (3.1).

3.2. Hydrogen Emission Model

To evaluate the impact of fugitive H_2 emissions, it is necessary to project future H_2 fuel consumption. These projections are derived from conventional kerosene consumption scenarios. Grewe et al. (2021) proposed two relevant scenarios, visualised in figure 3.3.



(a) Projected fuel consumption [Tg] in the aviation sector across three different scenarios: CurTec (current technology) and BAU (business as usual), adapted from Grewe et al. (2021).

(b) Projected hydrogen adoption fraction over time for three scenarios with varying adoption rates. The parameter $y_{0.5}$ signifies the year in which 50% of the total fuel consumption is attributed to hydrogen, while a governs the slope of the curve.

Figure 3.3: Projected fuel consumption under the CurTec and BAU scenarios **(a)** and various hydrogen adoption scenarios **(b)** in the aviation sector between 2025 and 2100.

1. **CurTec** (Current Technology), which assumes aviation technology remains at current levels, leading to steady growth in emissions.
2. **BAU** (Business As Usual), which also assumes growth in aviation but factors in technological advancements to improve efficiency.

Kerosene consumption is converted to an equivalent hydrogen consumption by assuming hydrogen-powered aircraft exhibit an increase in energy use in hydrogen-powered aircraft, attributed to the increased drag from the larger wetted area needed for hydrogen storage (Sáez Ortuño et al., 2023). This allows for the straightforward calculation,

$$M_{H_2,eq} = 1.1 \cdot M_{kerosene} \frac{LWH_{kerosene}}{LWH_{H_2}}, \tag{3.2}$$

where LWH refers to the lower heating value of either kerosene or hydrogen. Due to hydrogens' higher energy content per unit mass, its equivalent energy consumption ($M_{H_2,eq}$) is lower than kerosene ($M_{kerosene}$).

The transition to hydrogen-powered aircraft is expected to take time. Airbus predicts its first deliveries of hydrogen-powered aircraft will occur in 2035[1]. From this year onwards, the adoption of hydrogen-powered aircraft will likely follow the trajectory of a technology adoption lifecycle (often referred to as an S-curve) (Savage, 1985), which is modelled as

$$f_{adoption} = \frac{1}{1 + \exp^{-m(y - y_{0.5})}}, \tag{3.3}$$

where $y_{0.5}$ represents the year when 50% of total aircraft energy consumption is from hydrogen, and m determines the rate of adoption by adjusting the slope of the S-curve. By adjusting these parameters, optimistic or pessimistic scenarios for hydrogen adoption are modelled. Figure 3.3b shows three adoption scenarios, with hydrogen-powered aircraft introduced in 2035 and reaching 50% adoption by either 2050, 2060, or 2070. Parameter m is chosen to constrain the curves to 0% adoption in 2035 and 100% by 2100 at the latest.

H_2 leakage occurs not only during application, such as in hydrogen-powered aircraft but also throughout production and delivery. Fan et al. (2022) provides estimates of leakage rates across different stages of the H_2 supply chain. By applying projected scales for each production and delivery method for 2050, the estimated leakage rates range from 1.6% to 3.0% for production and 1.1% to 2.2% for delivery, reflecting both low- and high-risk scenarios. These rates are used in conjunction with an assumed aircraft-specific leakage rate, LF_{app}.

Equations (3.4) to (3.6) describe the mass of H_2 across different stages of the supply chain. The first equation calculates the mass of H_2 required for application ($M_{H_2,app}$) by adjusting the equivalent H_2 mass ($M_{H_2,eq}$) for adoption and accounting for leakage during application (LF_{app}). The second equation adjusts this mass to account for leakage during delivery (LF_{del}), giving the mass required for delivery ($M_{H_2,del}$). Finally, the third equation calculates the mass required for production ($M_{H_2,prod}$) by adjusting for leakage during production (LF_{prod}). Each equation accounts for the losses at each stage of the H_2 supply chain.

$$M_{H_2,app} = f_{adoption} \frac{M_{H_2,eq}}{(1 - LF_{app})} \tag{3.4}$$

$$M_{H_2,del} = \frac{M_{H_2,app}}{(1 - LF_{del})} \tag{3.5}$$

$$M_{H_2,prod} = \frac{M_{H_2,del}}{(1 - LF_{prod})} \tag{3.6}$$

$$E_{H_2,fugitive} = LF_{app} \cdot M_{H_2,app} + LF_{del} \cdot M_{H_2,del} + LF_{prod} \cdot M_{H_2,prod} \tag{3.7}$$

The total fugitive H_2 emissions ($E_{H_2,fugitive}$) are determined by calculating the hydrogen masses lost at each stage—production, delivery, and application—by multiplying the required hydrogen masses by their respective leakage rates and summing the contributions from each stage.

---

[1] Airbus, Hydrogen in aviation: how close is it? accessed October 2024, `https://www.airbus.com/en/newsroom/stories/2020-10-hydrogen-in-aviation-how-close-is-it`

Since the parameters used in this estimation carry significant uncertainty—particularly because hydrogen technology is in the early stages of adoption—the uncertainties are quantified using a Monte Carlo simulation. Details of this simulation can be found in section 3.5.

Hydroxyl Perturbation due to Hydrogen Emissions

The methodology described previously provides a way to calculate the temporal evolution of H_2 emissions. Ideally, a surrogate model could be directly trained on H_2 emission data to determine its impact on GHG levels and radiative forcings. Unfortunately, simulations assessing the impact of hydrogen emissions in the context of future SSPs do not exist yet. Conversely, while the scenario simulations available through CMIP6 simulate the complex dependencies between H_2, OH, CO, CH_4, SWV, O_3 and O_3 precursors, they do not explicitly simulate the effect of hydrogen emissions. Therefore, a method must be devised to convert H_2 emissions into changes in tropospheric mean OH.

Hydrogen's climate impact, however, is governed mainly by its destruction of OH (section 2.3). Warwick et al. (2022) demonstrated that tropospheric OH response to increases in hydrogen mixing ratios is complex and dependent on the response of CH_4, as well as background levels of CO, NO_X and VOCs. The rate of change of OH with H_2, however, is linear and only dependent on the inclusion of CH_4 response to excess H_2. This means that the influence of CO, NO_X and VOCs on the rate of change of OH due to increases in H_2 mixing ratios is negligible.

Aligning with these findings, the assumption is made that the rate of change of OH per H_2 ($\frac{dOH}{dH_2}$) is only dependent on the interaction between H_2, OH and CH_4. This allows for the determination of H_2 atmospheric concentration due to emissions, together with a change in OH concentrations (dOH), using a simple box model. The input into the surrogate models is the superposition of the background OH concentration for a certain scenario (OH), along with the change in OH concentration resulting from excess hydrogen concentrations (dOH). This assumption is thoroughly validated in chapter 4.

In recent years, several researchers have proposed box models to simulate the competition between CH_4, CO and H_2 for OH, all based on the initial approach proposed by Prather (1994). Section 2.3 provides an extensive overview of these models. In this study, the model used by Bertagni et al. (2022) is used, as it accounts for additional CH_4 sinks and feedback of CH_4 on H_2 concentrations, which the model used by Warwick et al. (2022) does not. Table 3.1 presents the main constants and input parameters for the box model.

Table 3.1: Constants and input parameters used in the box-model proposed by Bertagni et al. (2022), which is used concurrently in this work.

Constant	Value	Unit	Constant	Value	Unit
k_1	$3.17 \cdot 10^{-15}$	$cm^3 s^{-1}$	k_s	0.02	$year^{-1}$
k_2	$3.80 \cdot 10^{-15}$	$cm^3 s^{-1}$	k_d	0.38	$year^{-1}$
k_3	$1.90 \cdot 10^{-13}$	$cm^3 s^{-1}$	α	0.37	-
$k_4[X]$	0.3	s^{-1}	$t_{horizon}$	100	year

In the box model described by Bertagni et al. (2022), which is described in section 2.3, the tropospheric dynamics of methane (CH_4) and its feedback with hydrogen (H_2) are governed by a set of chemical reactions and mass balance equations. The model includes reaction rates for the oxidation of CH_4, H_2, carbon monoxide (CO), and other OH-consuming species (X), with rates defined by k_1, k_2, k_3, and $k_4[X]$, respectively. Methane oxidation via OH results in a yield (α) of H_2, representing the feedback effect of methane oxidation products on atmospheric OH. The mass balance equations account for CH_4 sources (S_{CH4}) and sinks, including soil uptake (k_s) and atmospheric decay (k_d), while soil uptake of H_2 is modeled as a primary removal pathway, accounting for 75% of its loss.

Using the box model and the projected emissions from the CMIP6 database for each SSP scenario in this study, a baseline evolution of the tropospheric mean OH concentration is computed. A second simulation is then conducted, incorporating the additional hydrogen emissions estimated by the method outlined earlier in this section. The difference in OH mixing ratios between these two simulations represents the perturbation of OH levels directly caused by H_2 emissions.

For the baseline simulation, the temporal evolution of emissions for CH_4 and CO are directly taken from the CMIP6 simulations. Because CMIP6 data does not include emission or chemical production data for H_2 and OH, these values are assumed to remain constant over time. This results in values of 226 ppb $year^{-1}$ for H_2, and 1333 ppb $year^{-1}$ for OH Bertagni et al. (2022). In the perturbed scenario, the additional fugitive emissions of hydrogen are added to the baseline emissions, while all other emissions remain unchanged.

The system of ODEs is solved using an explicit fourth-order Runge-Kutta (RK4) method. This numerical technique offers a good balance between computational efficiency and accuracy for time integration. To accommodate the time-dependent nature of the emissions data, the emission time series is interpolated linearly. This interpolation allows the solver to evaluate the equations at any arbitrary time point required by the RK4 method, ensuring that the impact of emissions is accurately reflected throughout the simulation. Verification of the implementation of the box-model along with validation of its usage in this study is performed in chapter 4.

3.3. Surrogate Model

As mentioned in section 3.1, surrogate models are used to approximate complex atmospheric and climate processes with significantly reduced computational cost. OpenAirClim is a surrogate model that emulates the climate response to aircraft emissions by modelling the nonlinear relationships between emissions and climate metrics. The model developed in this research is meant to supplement OpenAirClim by rapidly assessing the climate impact of hydrogen emissions. Response modelling, however, cannot be used to determine the climate impact of hydrogen because no simulations have been performed linking localised hydrogen emissions directly to radiative forcing as of the writing of this report. Therefore, other surrogate modelling techniques capable of capturing complex, previously unseen relationships in the absence of direct simulations need to be considered.

One suitable approach is machine learning. Unlike response modelling, machine learning models can learn patterns from large datasets, even when explicit simulations are unavailable. This enables them to be trained on observed or simulated atmospheric data from other sources. Machine learning works by training algorithms on large datasets to recognise patterns and relationships within the data. Through iterative learning processes, these algorithms adjust their internal parameters to optimise predictions or classifications, enabling them to generalise and make accurate predictions on unseen data.

In particular, recurrent neural networks (RNNs) are well-suited for modelling time-dependent processes because they maintain a memory of previous inputs through their internal hidden states, allowing them to capture temporal dependencies and sequential patterns in data. RNNs are especially useful for forecasting atmospheric changes where historical data influences future outcomes, such as the delayed response of methane to hydrogen concentration perturbations. To model the spatial dependency of ozone, a recurrent neural network can be supplemented with a convolutional neural network (CNN). CNNs are particularly effective at extracting spatial features by applying convolutional filters that capture localised patterns, such as spatial gradients in ozone concentrations across different regions. Integrating CNNs for spatial dependencies with recurrent layers for temporal evolution allows the model to capture both the spatial structure and time-varying dynamics of ozone levels.

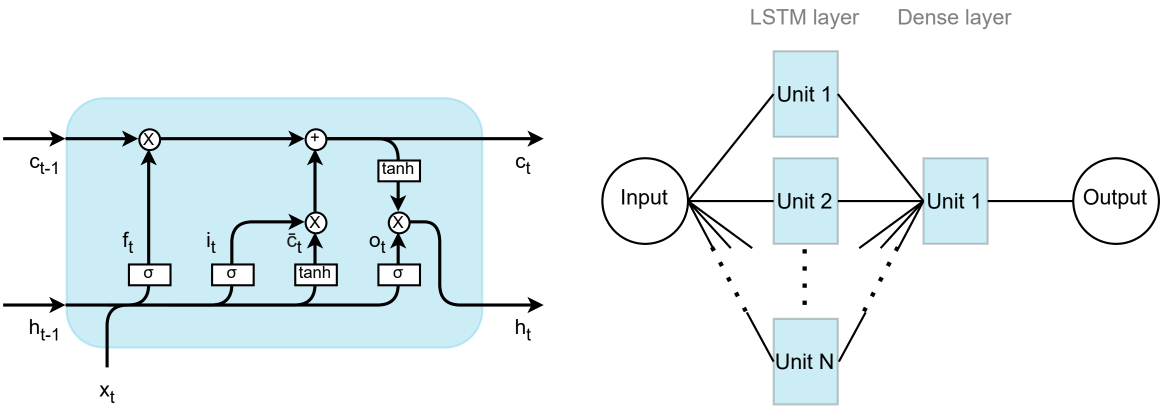
This section details how these machine-learning techniques are implemented. First, the rationale for choosing these particular neural network models will be provided with a brief overview of their functioning. Subsequently, the architecture of the models, including key hyperparameters and training strategies, will be discussed. Finally, the implementation of these models will be explained in detail, including the dataset preparation, model validation, and hyperparameter tuning.

3.3.1. Convolutional- and Recurrent Neural Networks

RNNs are a class of artificial neural networks designed to model sequential data by incorporating temporal dynamics through recurrent connections. Unlike feedforward neural networks, which process inputs independently, RNNs maintain an internal state (memory) that evolves as sequences of input data are processed over time. In a basic RNN, the network processes each time step sequentially, with hidden states carrying information from previous steps. RNNs share weights across all time steps, reducing the number of parameters compared to traditional neural networks. This weight-sharing mechanism helps in learning patterns that are consistent throughout the sequence.

Long Short-Term Memory (LSTM)

The most commonly used RNN model is the Long Short-Term Memory (LSTM) network, first introduced by Hochreiter and Schmidhuber (1997). LSTMs are designed to overcome the limitations of traditional RNNs, particularly the issue of vanishing or exploding gradients, which makes it difficult for RNNs to learn long-term dependencies in data. This is achieved by incorporating memory cells and gating mechanisms that regulate the flow of information, enabling the retention of important temporal information over longer sequences and allowing irrelevant data to be discarded. Figure 3.4 shows a typical LSTM cell architecture. Each LSTM cell contains three primary gates—the input gate, forget gate, and output gate—and an input modulation gate that creates a candidate cell state. The input gate controls how much new information from the current input is stored in the cell state, while the forget gate determines which parts of the previous cell state to retain or discard. The output gate decides how much of the cell state is exposed to the next hidden state. The cell state serves as a memory, updated via a combination of the input and forget gates, enabling the network to maintain relevant information and forget irrelevant details over time.



(a) Structure of an LSTM cell. The forget gate (f_t) modulates the retention of past information, while the input gate (i_t) regulates the integration of new data. The cell candidate $(\tilde{c}_t)$ is generated via a $\tanh$ transformation, and the updated cell state (c_t) results from combining the retained previous state with the new input. The output gate (o_t) determines the hidden state (h_t), which influences the subsequent time step.

(b) Fully connected feedforward neural network with two hidden layers. The input is fed into the first layer, consisting of N units, each connected to all units in the second hidden layer. The second layer contains one unit - used to make a single prediction - and is fully connected to the previous layer. Each connection represents a weighted input, facilitating the flow of information from input to output through multiple layers, allowing the model to learn complex patterns in the data.

Figure 3.4: Overview of an LSTM cell structure **(a)** and a LSTM model architecture with two layers **(b)**.

Individual LSTM layers can be stacked, where the output of each layer at a given time step serves as input to the next layer at the same time step. This allows the network to capture increasingly complex temporal dependencies by processing information at multiple hierarchical levels. When multiple LSTM layers are stacked, the higher layers can abstract patterns from longer time spans or more intricate combinations of features, enhancing the network's capacity to model complex sequential data. Increasing the number of LSTM layers can improve the model's ability to capture complex patterns but comes with the trade-off of longer training times and a higher risk of overfitting, especially on smaller datasets. Regularisation techniques like dropout or early stopping, along with careful tuning of layer size, are necessary to balance the depth of the model with generalisation performance.

Convolutional LSTM

While LSTM networks excel at capturing temporal dependencies in sequential data, they are limited in their ability to model spatial grids. A Convolutional Neural Network (CNN) operates by applying convolutional filters, or kernels, to input data - typically images - to automatically detect local patterns like edges, textures, or shapes. The convolution operation slides these filters across the input, producing feature maps of specific patterns. CNNs also use pooling layers, commonly max pooling, to reduce the spatial dimensions of the feature maps, retaining important information while reducing computational complexity. As the network deepens, successive layers capture increasingly abstract and complex features. Finally, fully connected layers combine these learned features for regression tasks.

Convolutional operations can be integrated into the recurrent structure of LSTMs to create the ConvLSTM architecture. This model extends traditional LSTM networks by replacing the internal matrix multiplications with convolutional operations. Specifically, the input-to-state and state-to-state transitions within the LSTM cell are performed using convolutional layers, allowing the network to capture spatial features in addition to temporal dependencies. The hidden states and cell states in ConvLSTM are treated as spatial tensors, enabling the model to process spatial grids. Figure 3.5 shows how convolution can be implemented within an LSTM layer.

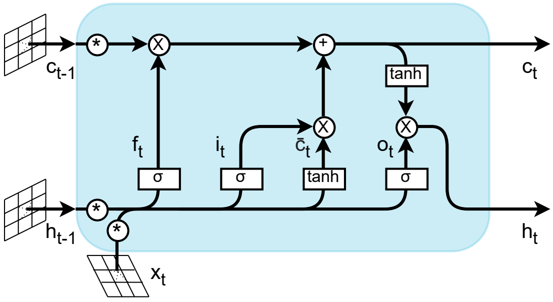


Figure 3.5: Structure of a ConvLSTM cell. The layout of the cell is similar to an LSTM cell, with the addition of spatial tensors as inputs and convolutional operations between the inputs. These convolutional operations are denoted by the * symbol.

The convolutional filters within the LSTM gates learn localised spatial features, while the recurrent connections model how these features evolve. This makes the ConvLSTM architecture specifically useful to model the spatiotemporal evolution of ozone, which has a significant spatial dependency due to its short lifetime (section 2.3). In contrast, methane is be modelled using the less complex LSTM model architecture, as its longer lifetime ensures it is well mixed in the atmosphere, negating the need for spatial resolution (Warwick et al., 2022).

3.3.2. Data Preparation

The Scenario Model Intercomparison Project (ScenarioMIP) provides a set of future climate scenarios designed to explore the range of potential outcomes of socioeconomic and policy decisions on the climate system. These datasets are be used to train a surrogate model on the effect of hydroxyl perturbations caused by hydrogen emissions on future tropospheric concentrations of methane and ozone. This allows the surrogate model to capture the impact of hydrogen emissions in the context of various emission scenarios of carbon monoxide, methane, nitrogen oxides, and other atmospheric reactants.

In this study, models are trained on data from seven distinct scenarios: SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP4-3.4, SSP4-6.0, and SSP5-8.5. Some variations exist for each of these scenarios. For example, SSP3-7.0-lowNTCF considers a scenario similar to SSP3-7.0 but with lower non-CO_2 climate forcers. However, these variations cannot be included in the training data, as large dataset similarity can influence the model's ability to generalise effectively to unseen scenarios. Including similar variations may lead to overfitting, where the model performs well on the training data but poorly on new, distinct data. To maintain model robustness and ensure diverse representation of future pathways, only the primary SSP scenarios are utilised in the training process.

Feature Engineering

Section 2.3 explains that hydrogen emissions increase the atmospheric lifetime of methane by decreasing OH concentrations. Levels of OH in the atmosphere depend on gases besides methane and hydrogen; carbon monoxide (CO), nitrogen oxides (NO_x), and volatile organic compounds (VOCs) all compete with methane for reactions with OH. However, the most significant competitors are CO and CH_4. Therefore, to predict future methane levels, an LSTM model is trained on the temporal evolution of mean tropospheric CO and OH concentrations, along with CH_4 and CO emissions. The zero-dimensional LSTM model is suitable here because both the perturbing species (H_2) and the predicted species (CH_4) exhibit negligible spatial variability due to their atmospheric lifetimes, as discussed in section 2.3.

Stratospheric water vapour (SWV) is perturbed by hydrogen through two main pathways: directly via hydrogen transport into the stratosphere and indirectly through hydrogen-induced changes in methane

(CH_4). The direct impact is computed using empirical relations that account for the transport of hydrogen and its oxidation in the stratosphere. The methane-induced perturbation is estimated using an LSTM model, which predicts CH_4 perturbations and their subsequent effect on SWV. Additionally, stratospheric temperature and the transport of H_2O from the troposphere influence SWV levels and are included as features in the model to ensure these processes are accurately represented.

Tropospheric ozone is not directly emitted; its atmospheric abundance is governed by a balance between downward transport from the stratosphere, surface deposition, and the net effects of chemical production and loss processes (Warwick et al., 2022). Photochemical ozone production occurs through the oxidation of carbon monoxide (CO), methane (CH_4), and volatile organic compounds (VOCs) in the presence of nitrogen oxides (NO_X). Consequently, variations in hydroxyl radicals (OH), NO_X, and CO levels influence the spatial distribution of tropospheric ozone in response to changes in atmospheric hydrogen (H_2) concentrations. Therefore, a ConvLSTM model is trained using data on CH_4, CO, NO_X, OH, HO_2, and stratospheric ozone concentrations to predict the spatiotemporal evolution of tropospheric ozone.

The most extensive ScenarioMIP simulations were performed by the Japanese Meteorological Research Institute using the MRI-ESM2.0 model. A brief overview of this model is provided in section 2.4. Output concentrations are provided as molar fractions in air, sampled monthly in a full 3D grid containing latitude, longitude and pressure level coordinates. To be used in the LSTM model, the data must be converted to their respective tropospheric means. This is achieved by calculating the total mass of a species in the troposphere using air mass and the dynamic location of the tropopause, which are outputs of MRI-ESM2.0. The mean tropospheric volume mixing ratios of a species can then be computed by dividing the total weight of the species by the total weight of air in the troposphere, and multiplying the result with the ratio of the molecular masses. For ozone predictions using the ConvLSTM model, the spatial distribution of species along the latitude and longitude coordinates is maintained by applying the same procedure using the relevant values for a specific cell.

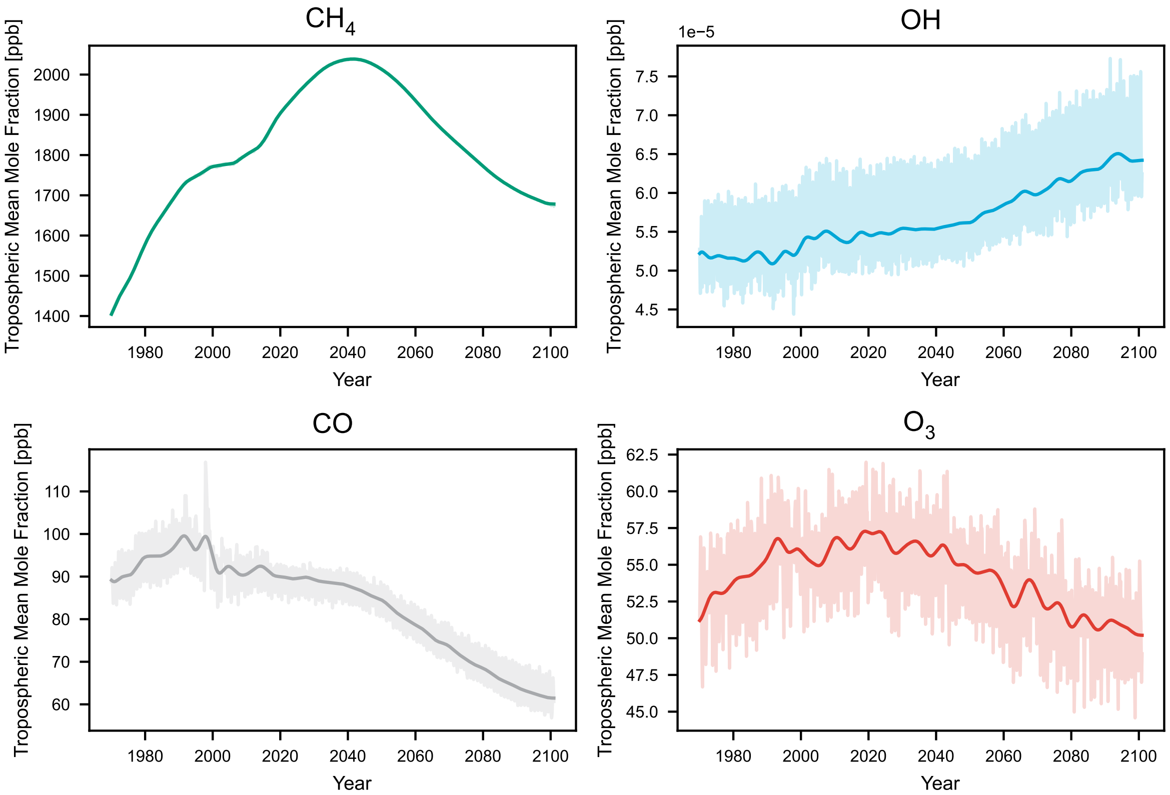


Figure 3.6: Example of applying a Butterworth filter to remove seasonal variations in four key hydrogen atmospheric chemistry species under the SSP2-4.5 scenario. Shaded lines illustrate the seasonal fluctuations of the original data, while the filtered data is represented with bold colours.

This study is focused on predicting long-term climate scenarios. Seasonal cycles that are captured by the monthly sampling of the data, tend to average out over the long timescales considered in this study. By filtering out the seasonal variation in the data, the surrogate models can be trained to capture long-term trends in the data more effectively. At the same time, the computational cost is reduced due to the reduction in data points. Therefore, the data is resampled to yearly values by applying a Butterworth filter. A Butterworth filter operates by gradually weakening fluctuations that occur faster than the specified cutoff frequency—in this case, monthly seasonal variations at 1/12 cycles per month—allowing only the slower, long-term trends in the atmospheric data to remain (S Butterworth, 1930). It is chosen for its ability to provide a smooth and flat response in the desired frequency range, effectively removing seasonal noise without distorting the underlying signal. Figure 3.6 shows the resulting timeseries after the application of a Butterworth filter.

Data Splitting

Data splitting is a fundamental process in machine learning that involves dividing the dataset into distinct subsets: training, validation, and testing. The training set is used to train the model, allowing it to learn the underlying patterns and relationships within the data. The validation set serves as a means to tune the model's hyperparameters, ensuring that the model generalises well to unseen data. Finally, the testing set is reserved for evaluating the model's final performance, providing an unbiased assessment of its predictive capabilities on new, unseen data.

To enhance the robustness of the model evaluation, K-fold cross-validation is employed, where the dataset is partitioned into K equally sized folds. The model is trained on $K - 1$ folds and validated on the remaining fold, repeating this process K times so that each fold serves as the validation set once. This technique mitigates the risk of overfitting. In this report, each scenario corresponds to a fold, resulting in the data splices shown in figure 3.7.

This approach is used to evaluate the model's performance during the verification phase, as detailed in section 4.1. Once it is determined that the model generalises effectively, with no significant discrepancies observed across different folds during validation, a final model is trained using the complete set of folds, where a random set of timesteps is denoted for validation. This final model is then assessed in section 4.2, where the "testing" dataset consists of a simplified scenario in which only OH levels are perturbed, while all other features remain fixed at present-day values. This setup enables a focused analysis to determine whether the model truly generalises well to variations in OH perturbations.

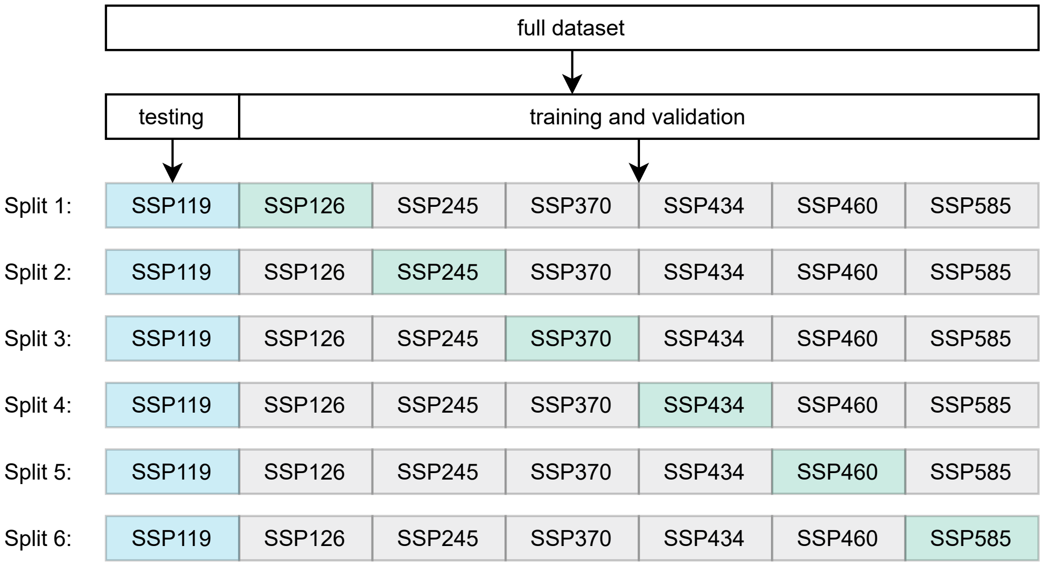


Figure 3.7: Overview of the split between training, validation and testing scenarios for SSP-1.19. Blue indicates the scenario for which the model is tested, which is the same for each split. grey indicates the scenarios used for model training, while green indicates the scenario used to validate the generalizability of the model at each epoch.

Stationarity and Normalization

Training machine learning models can often be enhanced by ensuring the data is stationary (Siami-Namini et al., 2018). Stationarity refers to a property of a time series where its statistical characteristics, such as mean, variance, and autocorrelation, remain constant over time (Siami-Namini et al., 2019). Achieving strict stationarity in the data used in this study is unfeasible due to factors like inherent trends and seasonality. However, stationarity can be approached through various preprocessing techniques.

Differencing is a common method used to remove trends and render a time series trend stationary. By computing the difference between consecutive observations, linear trends are effectively eliminated, thereby stabilising the mean of the series. Differencing alone does not address changes in variance or other higher-order moments. Nonetheless, this approach is widely utilised in time series forecasting and is therefore applied in this study (Siami-Namini et al., 2018; Haq, 2021).

Normalisation of the data is performed to scale the features to a standard range, which facilitates faster convergence during model training and improves overall model performance (Lima and Souza, 2023). It is important to normalise using only the training data statistics to prevent information leakage from the validation and testing sets. This approach ensures that the model generalises well to new data and maintains the integrity of the evaluation process.

In the time dimension, a min-max normalisation technique is applied, scaling the data to a fixed range between -1 and 1. This method preserves the temporal relationships and patterns within the data while ensuring that all features contribute equally to the model training process. An example of the resulting timeseries is shown in figure 3.8.

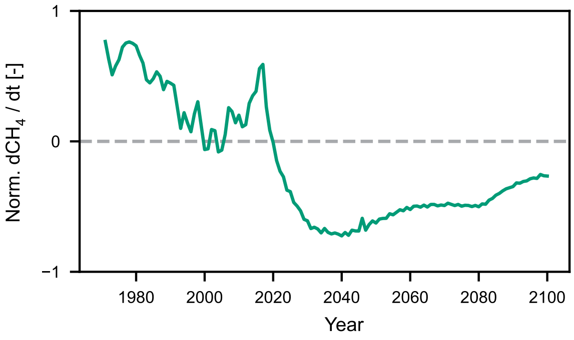


Figure 3.8: Normalized time derivative of mean tropospheric CH_4 concentrations.

For ConvLSTM, additional normalisation is required for the spatial dimensions of the data. Some atmospheric gases, such as OH and O_3, have a spatial variation of several orders of magnitude. Therefore, the data is normalised in the spatial dimension by subtracting the mean and dividing by the standard deviation across the spatial axes. Spatial normalisation ensures that each spatial feature has a consistent scale.

Window Generation and Batching

In the final step of data preparation, the training and validation datasets are divided into windows. These windows are smaller, overlapping sequences of data, each with a width consisting of input values and the prediction step size. To predict the label value at time t, the model considers known features and label values between $t - w_{\text{input}}$ and t. This approach offers two key advantages.

First, by restricting the model's input to recent data rather than the entire history, the model focuses on the most relevant information, aligning with the expected lifetimes and behaviour of the atmospheric species under study. Specifically, species with short to moderate atmospheric lifetimes do not require long historical data, reducing computational cost and minimising the risk of overfitting to irrelevant data.

Secondly, this method allows the dataset to be divided into numerous smaller training windows, enhancing the model's ability to generalise. For example, in a dataset spanning n years, $n - w_{\text{input}}$ windows can be generated, significantly increasing the number of training examples. This also prevents the model from overfitting to learned data, as the windows are randomly shuffled, which prevents the model from learning a particular sequence of time steps that belong to a scenario. The window generation process is visualised in figure 3.9.

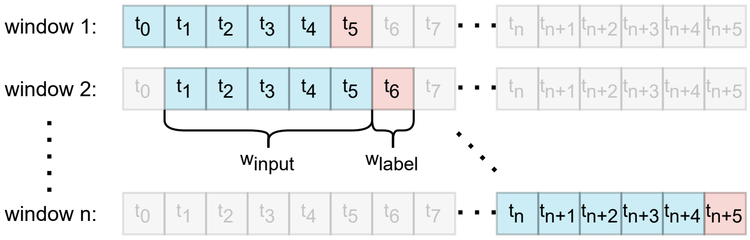


Figure 3.9: Graphic representation of the division of a single dataset containing $n + w_{input}$ timesteps. A window contains input variables for each feature with a length w_{input}, and a label (or prediction) length.

Training windows are batched together before being input into the model. The size of a batch refers to the number of windows the model will consider before its internal weights are updated. The batch size, b, is an important hyperparameter that balances the trade-off between memory usage and the ability to generalise; larger batch sizes tend to offer faster convergence at the cost of higher memory consumption, while smaller batch sizes may improve generalisation by introducing more stochasticity into the learning process (Siami-Namini et al., 2018). Batching helps stabilise gradient updates during backpropagation, leading to smoother and more consistent learning. The effects of window parameters and batch size on training are discussed in section 4.3.

3.3.3. Model Architecture

In this study, two different model architectures are used: LSTM for temporal predictions of tropospheric methane and ConvLSTM for spatiotemporal predictions of tropospheric ozone. Both models require a different architecture described below.

LSTM

The LSTM model architecture is constructed sequentially, comprising an LSTM layer, a Dense layer, and a final reshaping operation. The input tensor to the LSTM layer has the shape (n_{batch}, w_{window}, n_{features}), representing the batch size, window width, and number of features, respectively. The number of units in the LSTM layer is optimised through automatic hyperparameter tuning, as detailed in section 3.3.5. The hyperbolic tangent activation function is employed due to its output range of $[-1, 1]$, which accommodates both positive and negative values necessary for the differenced data described in section 3.3.2. This contrasts activation functions like ReLU or sigmoid, which are limited to $[0, 1]$. Additionally, the kernel and recurrent initialisers are selected using the hyperparameter tuning process.

To improve generalisation and prevent overfitting, an L2 kernel regulariser is applied. This regularisation penalises large weights, encouraging the model to remain simple and reducing the likelihood of fitting noise in the training data. Such regularisation is particularly important for time series data, which may exhibit short-term fluctuations. For more background information on activation functions, initialisers, and regularisers, refer to appendix A.

A dense layer is used after the LSTM layer to map the temporal patterns learned by the LSTM to a final output feature. The Dense layer receives the LSTM's output, applies a linear transformation using learned weights, and uses an activation function to generate a final feature or prediction, condensing the sequential information into a singular output. As the output of this final layer must be capable of being outside the range of $[-1, 1]$ (the model must predict steeper gradients than the training data if required), linear activation is used.

ConvLSTM

The ConvLSTM architecture is configured similarly. An input tensor of (n_{batch}, w_{window}, n_{rows}, $n_{columns}$, $n_{channels}$) is fed into a ConvLSTM2D layer. This layer extends the LSTM framework by incorporating convolutions into the input-to-state and state-to-state transitions, making it suitable for predicting data with spatial dependencies. In this layer, the size of the kernel that determines the receptive field for spatial feature extraction can be altered to adjust the level of spatial detail captured by the model, with larger kernels capturing broader spatial patterns and smaller kernels focusing on finer local details. Kernel size, along with the other hyperparameters, is optimised using a hyperparameter autotuning regime.

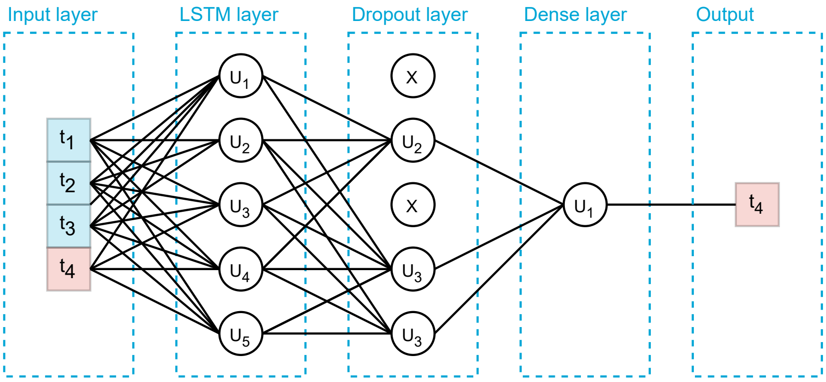


Figure 3.10: Simplified architecture of the LSTM model, showing an input layer using an input width of three timesteps to forecast one step. In this example, the LSTM layer contains five units, and the dense layer contains one unit because only one timestep is predicted. A dropout layer is configured between the LSTM and Dense layers.

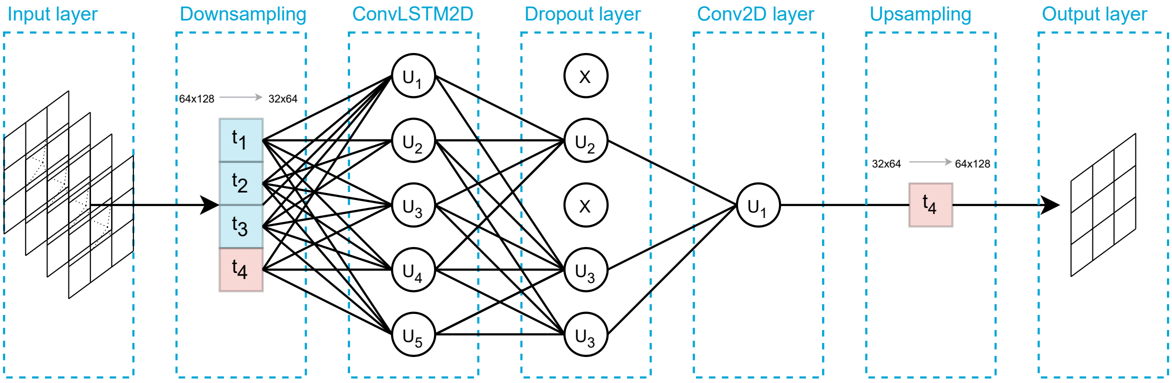


Figure 3.11: Simplified architecture of the ConvLSTM model, showing an input layer using an input width of three spatially resolved timesteps to forecast one step. In this example, the ConvLSTM2D layer contains five units, and the dense layer contains one unit because only one timestep is predicted. A dropout layer is configured between the LSTM and Dense layers. To prevent memory issues, the input layers with spatial resolutions of (64x128) are downsampled using AveragePooling.

Subsequently, a Conv2D layer is used the same way as the Dense layer in the LSTM model. Linear transformation is used to allow the model to make continuous predictions. The kernel processes the spatiotemporal features, condensing them into a single output channel.

3.3.4. Model Evaluation

The machine learning algorithm assesses its training performance after each batch is processed and its validation performance after each epoch. Performance refers to minimising the loss function, often a simple metric quantifying the error between the model prediction and the true value.

The primary metric used for evaluating the model's performance is the Root Mean Square Error (RMSE). RMSE measures the average magnitude of the errors between predicted and actual values, penalising larger errors more severely. The optimisation of the model is carried out by the Adam optimiser. The learning rate is a pivotal hyperparameter in this process, determining the size of the steps taken during optimisation. Selecting an appropriate learning rate is crucial; too high a value can lead to divergence, while too low a value may result in slow convergence.

During training, an early stopping algorithm and a learning rate scheduler are implemented to prevent the model from overfitting. Early stopping halts the training process when the validation loss increases, ensuring the model does not learn noise from the training data. Additionally, a learning rate scheduler is employed to adjust the learning rate dynamically throughout the training process. By reducing the learning rate when the improvement in validation loss plateaus, the scheduler helps the model converge more effectively to a local minimum. These concepts are explained in appendix A.

3.3.5. Hyperparameter Optimization

The hyperparameters discussed in section 3.3.3 significantly influence the performance of the model during training and inference. An optimisation scheme can be used to determine the set of parameters that yields the minimum training and validation loss. In this study, the optimal set of hyperparameters for both the LSTM and ConvLSTM models is found using a random search algorithm in favour of the more exhaustive but computationally intensive grid search. The results for LSTM and ConvLSTM can be found in table 3.2 and table 3.3, respectively.

Using the KerasTuner framework, a range of possible hyperparameter values is initialised O'Malley et al. (2019). The number of LSTM units or ConvLSTM filters is determined by sampling between 16 and 128 with a step size of 16. The rate of L2 regularisation (λ) - an additional loss used to penalise large weights - along with the learning rate of the optimiser are sampled similarly, but on a logarithmic scale between 1e^{-4} and 1e^{-2}. The kernel and recurrent initialisers are chosen from the choices visible under the sampling space column.

Table 3.2: Hyperparameter sampling techniques, sampling space and final values used in random search optimisation for the LSTM model.

Hyperparameter	Sampling technique	Sampling space	Final value
LSTM units	LinearSample(16)	16 to 128	32
LSTM kernel initializer	Choice	random normal random uniform glorot normal glorot uniform	glorot normal
LSTM recurrent initializer	Choice	orthogonal identity	orthogonal
LSTM L2 regularization λ	LogSample(-)	1e-4 to 1e-2	5.0e-3
Optimizer learning rate	LogSample(-)	1e-4 to 1e-2	0.8e-3
Dropout λ	LinearSample(0.1)	0.1 to 0.4	0.2

In the ConvLSTM model, the size of the kernel that slides over the spatial map can also be optimised. A smaller kernel grasps more local fine-grained details, while larger kernel sizes can retrieve information from a more global context. An excessively small kernel might lead to information loss, while a too-large kernel size will significantly increase the computational costs.

3.3.6. Autoregression

Both the LSTM and ConvLSTM models use a window with a certain input width to predict the value of a label one step into the future. This is done through an autoregressive algorithm, where the output of the model at each step is fed back as input for the subsequent prediction. This method allows the model to predict a time series by iterating step-by-step, leveraging the previous predictions as inputs. Autoregressive models are often preferred in time series prediction because they offer flexibility in predicting sequences of varying lengths and are more accurate for short-term forecasts, as they adjust to small variations before errors accumulate. Additionally, autoregressive models are easier to train, reduce complexity, and can better handle long-term dependencies by leveraging past predictions iteratively. This makes them more suitable for tasks requiring a fine-grained understanding of time-dependent data.

However, the autoregressive nature of this approach introduces the possibility of error propagation. Small prediction errors in early steps can accumulate over time, leading to increasingly inaccurate results in the later steps. This is especially problematic for long-term forecasts, where any initial discrepancies in the predicted values may amplify as they are used in subsequent iterations, ultimately degrading the performance of the model. Prediction errors at each time step are generally non-systematic

Table 3.3: Hyperparameter sampling techniques, sampling space and final values used in random search optimisation for the ConvLSTM model.

Hyperparameter	Sampling technique	Sampling space	Final value
ConvLSTM2D filters	LinearSample(16)	16 to 128	32
ConvLSTM2D kernel size	Choice	(2, 2) (3, 3) (4, 4)	(3, 3)
ConvLSTM2D kernel initializer	Choice	random normal random uniform glorot normal glorot uniform	glorot normal
ConvLSTM2D recurrent initializer	Choice	orthogonal identity	orthogonal
ConvLSTM2D L2 regularization λ	LogSample(-)	1e-4 to 1e-2	1.0e-3
Conv2D kernel size	Choice	(2, 2) (3, 3) (4, 4)	(2, 2)
Optimizer learning rate	LogSample(-)	1e-4 to 1e-2	0.4e-3
Dropout λ	LinearSample(0.1)	0.1 to 0.4	0.1

and, over many iterations, tend to follow a normal distribution centred around the true value of the time series. This is because the small, random deviations introduced during each step accumulate in a way that resembles a random walk, leading the final predicted value to vary symmetrically around the true target. As these errors are normally distributed, their impact can be quantified using the variance, which provides a measure of the overall spread of the prediction errors. Analysing the variance allows for a deeper understanding of the extent of the non-systematic error propagation throughout the forecast.

3.4. Radiative Forcing Calculation

The perturbed tropospheric concentrations of CH_4 and O_3 derived by the surrogate model must be converted to their respective radiative forcings to synthesise the climate impact of hydrogen emissions. The model computes the increase in radiative forcing by simulating a baseline projection and a perturbed projection. The difference between these two simulations provides the increase in mixing ratio of a certain species that can be directly attributed to increases in H_2 emissions. This allows for the conversion of mixing ratio increase of a species to its radiative forcing using radiative efficiencies - the change in radiative forcing per unit increase in the atmospheric abundance of the gas - determined by conventional climate models. Furthermore, as described in section 3.1, the contribution of SWV is directly calculated using its radiative efficiency expressed in terms of H_2.

The radiative forcings for CH_4, O_3 and SWV are calculated using the following equation:

$$\Delta RF_S = R_S \cdot \Delta C_S, \tag{3.8}$$

where ΔRF_S denoted the change in radiative forcing induced by species S, R_S the radiative efficiency and ΔC_S the change in mixing ratio of said species. For CH_4 and O_3, radiative efficiencies are directly taken from the latest report by the IPCC (2021b). The radiative efficiency of SWV with respect to increases in H_2 mixing ratios is taken from Warwick et al. (2022). Table table 3.4 provides an overview of the radiative efficiencies and their uncertainties.

The radiative forcing of CH_4 is scaled by 14% to represent its effective radiative forcing (ERF), following (IPCC, 2021b). This allows for direct comparison with results from Warwick et al. (2022) and Sand et al. (2023). For tropospheric O_3, no adjustment factor is applied, consistent with (IPCC, 2021b). Research on adjustment factors for methane-induced stratospheric water vapour (SWV) is limited; (IPCC, 2021b) suggests an adjustment factor below 1, though with low confidence. Consequently, RF values are treated as ERF, aligning with the approach of Sand et al. (2023).

3.5. Monte Carlo Simulation

Monte Carlo simulation is a statistical technique for assessing the impact of uncertainties in model parameters on the output of complex systems. In this study, a Monte Carlo simulation is used to quantify the impact of uncertainty on the future adoption of hydrogen in aviation, future leakage rates for hydrogen production, delivery, and application, constants used in the box model, and the parameters used in the radiative forcing calculations.

Table 3.4 summarises the parameters incorporated into the Monte Carlo simulation, highlighting their standard values, uncertainty distributions, range of uncertainty, and the sources they are retrieved from. If the uncertainty of a parameter is initialised as a uniform range distribution, any value between the minimum and maximum values presented has an equal likelihood of being randomly chosen by the Monte Carlo simulation. Under a normal uncertainty distribution, the parameter values are sampled from a Gaussian distribution centred around the standard value, with the spread defined by the specified standard deviation. This means that values closer to the mean (standard value) have a higher probability of being selected. In contrast, values further away are progressively less likely, following the characteristic bell curve of the normal distribution.

Table 3.4: Overview of the parameters used in the Monte Carlo simulation, along with their associated uncertainties and corresponding uncertainty distributions. Each parameter is characterised by an uncertainty range and a specific probability distribution, as outlined in the table below. During each Monte Carlo iteration, values for these parameters are sampled randomly based on their assigned distribution, ensuring variability consistent with the uncertainty ranges.

Parameter	Symbol	Unit	Standard value	Uncertainty distribution	Uncertainty range	Source
Fuel consumption scenario	-	[-]	BAU	Choice	BAU, CurTec	Grewe et al. (2021)
S-curve midpoint	t_{mid}	[year]	2060	Uniform (range)	2050 - 2070	-
S-curve steepness	m	[-]	0.27	Uniform (range)	0.36 - 0.18	-
Leakage factor application	LF_{app}	[-]	0.0300	Uniform (range)	0.0100 - 0.0500	Fan et al. (2022)
Leakage factor delivery	LF_{del}	[-]	0.0133	Uniform (range)	0.0050 - 0.0216	Fan et al. (2022)
Leakage factor production	LF_{prod}	[-]	0.0178	Uniform (range)	0.0053 - 0.0302	Fan et al. (2022)
Hydrogen deposition rate	k_d	[year^{-1}]	0.38	Uniform (range)	0.34-0.50	Warwick et al. (2022)
Methane sink rate	k_s	[year^{-1}]	0.02	Normal (% std)	10.0	Bertagni et al. (2022)
Methane feedback yield	α	[-]	0.37	Normal (% std)	10.0	Bertagni et al. (2022)
Metane RF coefficient	R_{CH_4}	[Wm^{-2}ppb^{-1}]	0.000389	Normal (% std)	20.0	IPCC (2021b)
Trop. ozone RF coefficient	R_{O_3}	[Wm^{-2}DU^{-1}]	0.0420	Normal (std)	0.005	IPCC (2021b)
Strat. water vapour RF coefficient	R_{H_2O}	[Wm^{-2}ppb^{-1}]	1e^{-4}	Normal (% std)	20.0	IPCC (2021b)

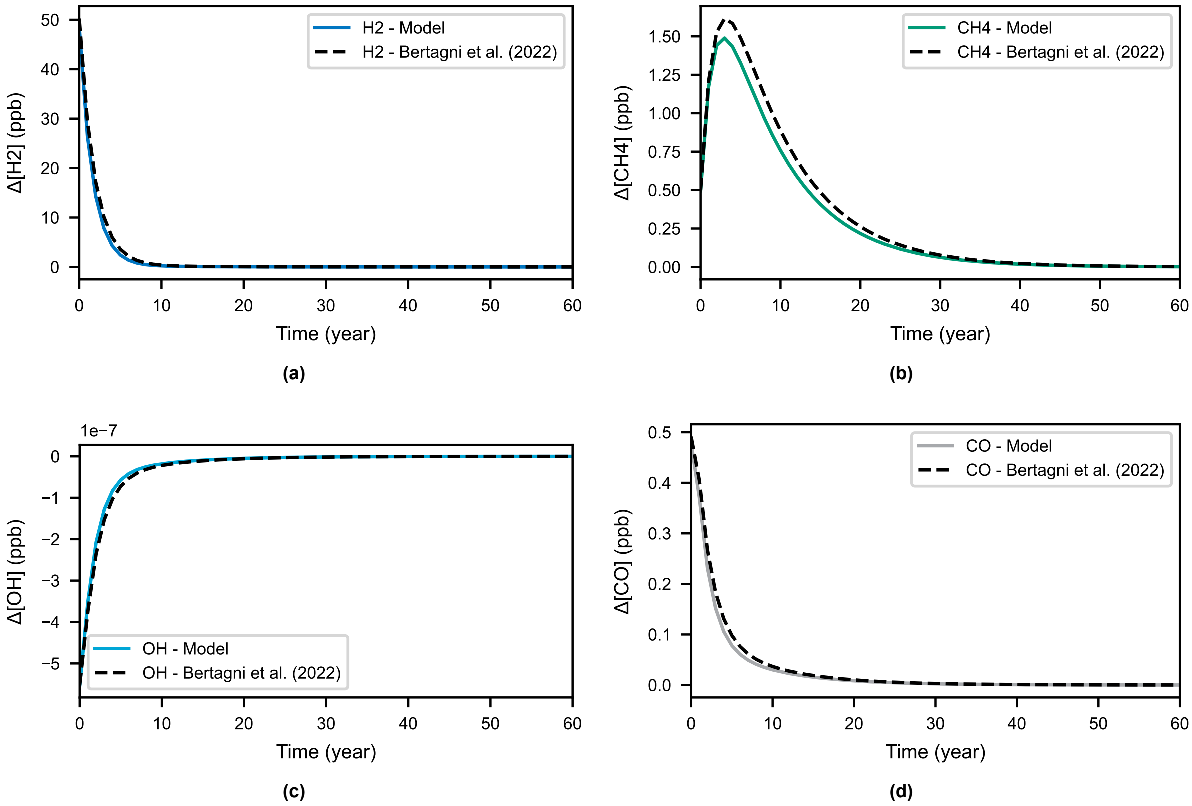
4

Model Verification & Validation

Before the model described in chapter 3 can be applied to assess the climate impact of fugitive hydrogen emissions, its accuracy and credibility must be evaluated thoroughly. Section 4.1 discusses the verification of the various models used in conjunction with this thesis. Subsequently, in section 4.2, the validity of the proposed methodology to model the climate effect of fugitive hydrogen emissions is investigated. A sensitivity analysis is performed in section 4.3 to test the robustness of the model. Finally, the findings of this chapter are summarised and critically discussed in section 4.4.

4.1. Verification

Verification refers to the process of ensuring that a computational model is correctly implemented and free from errors, confirming that the model accurately follows its intended design and specifications. The framework used in this study consists of several sub-models, each with a specific objective. The box model used in the hydrogen emission model can be verified by comparing its performance against the model described by Bertagni et al. (2022). Its implementation is correct if it performs similarly using the same input parameters.



(a) **(b)** **(c)** **(d)**

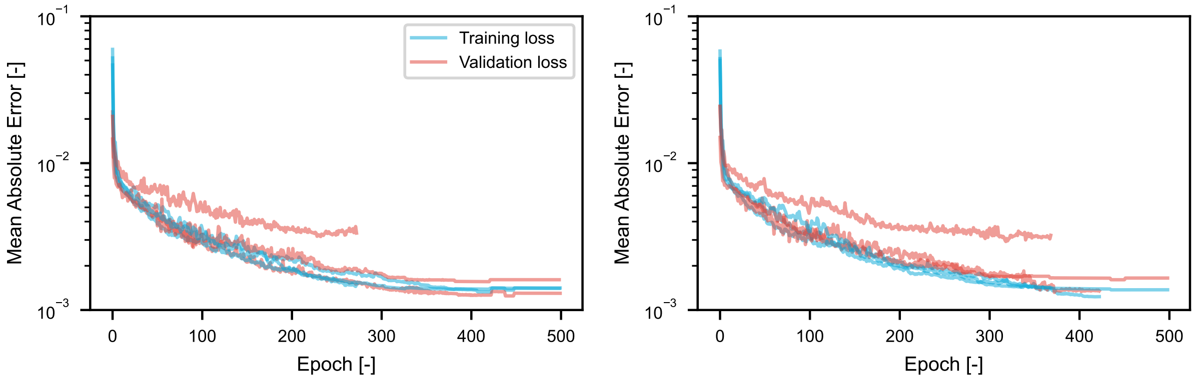
Figure 4.1: Box model response of CH_4 **(b)**, OH **(c)** and CO **(d)** to a pulse emission of H_2 **(a)** resulting in a 10% increase of its concentration plotted against results for the same simulation performed by Bertagni et al. (2022).

First, the model input parameters are verified by comparing the steady solution state of the model to the original model. The resulting concentrations of $[CH_4] = 1890$ ppb, $[H_2] = 530$ ppb, $[CO] = 80$ ppb and $[OH] = 10^6$ cm^{-3} correspond to values found by Bertagni et al. (2022) after rounding to the nearest decimal value. Subsequently, the temporal dynamics of the box model can be verified by analysing the response of the box model to a pulse emission.

At $t = 0$, hydrogen concentrations are temporarily increased by 10%, corresponding to an increase of 53 ppb in H_2 compared to the steady state solution. In figure 4.1, the tropospheric response of CH_4, CO, and OH to this perturbation of H_2 is shown. As expected, the higher levels of H_2 cause a nearly instant reduction in OH concentrations and a correlated increase in CO. The depletion of OH increases the lifetime of CH_4. This effect is visible as a delayed increase in CH_4 concentrations. The magnitude of the perturbations, along with the temporal dynamics that can be derived from figure 4.1, are all consistent with the results from Bertagni et al. (2022). This corroborates the finding that the box model used in this study was implemented correctly.

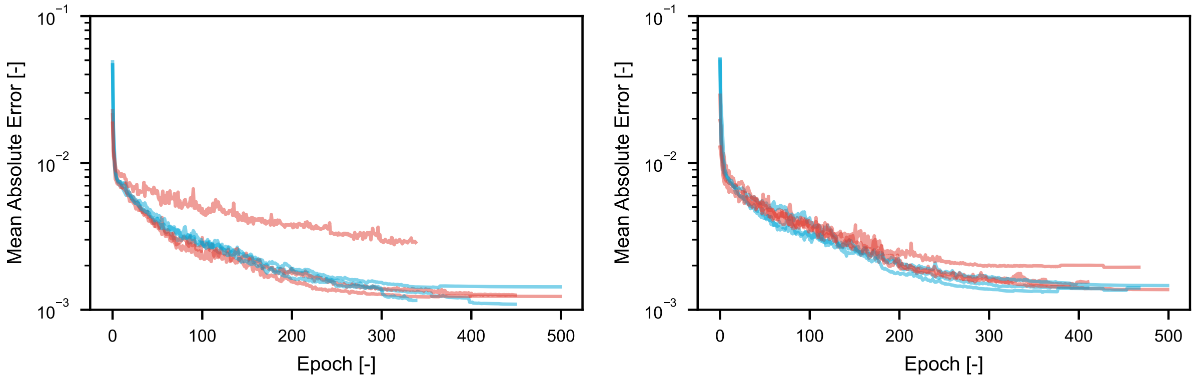
Methane

The machine learning surrogate models developed in this study require more extensive verification, as they are not derived from existing models but developed bespoke for this research. An initial method to evaluate their performance is through training and validation metric analysis. Figure 4.2 presents the training loss and validation loss for each model trained under specific scenarios, plotted against epoch number on a logarithmic scale. These loss curves are used to assess the models' convergence behaviour, identify potential overfitting, and determine the overall effectiveness of the training process. In total, all seven scenarios considered in this study are analysed. For comprehension, only the four most widely used scenarios are presented in the continuation of this thesis. Results for the other SSP scenario can be found in appendix C.



(a) SSP1-2.6

(b) SSP2-4.5



(c) SSP3-7.0

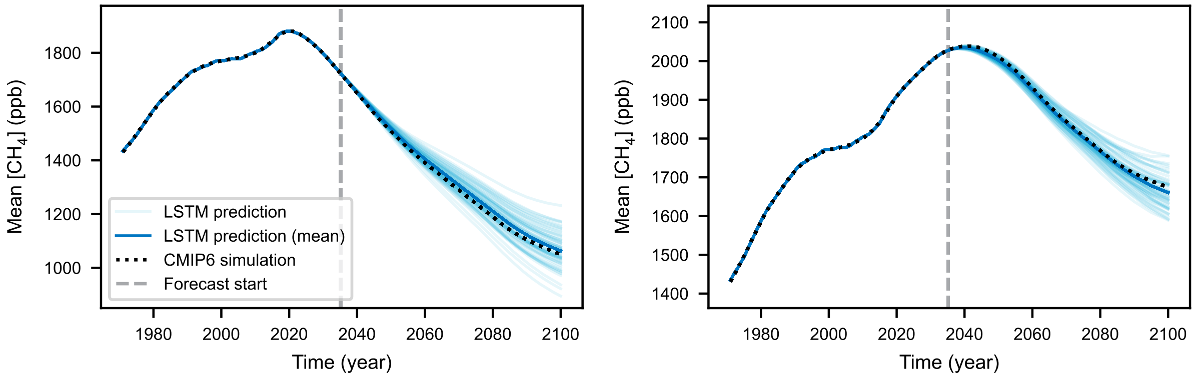
(d) SSP5-8.5

Figure 4.2: Training loss (cyan) and validation loss (red) for the CH_4 LSTM models tested on scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**, plotted as a function of epoch number on a logarithmic scale.

The plots show a decreasing training and validation loss trend for all scenarios. The initial steep decrease is typical for loss curves and shows the model is learning quickly in the first epochs. As the model converges, the slopes become less steep. Eventually, the models plateau towards the cut-off point of 500 epochs, signifying convergence. At this plateau, none of the models exhibit an increase in validation loss. The lack of divergence between the training and validation loss curves towards the latter epochs indicates that the model is not overfitting to the training data. Furthermore, all scenarios converge to similar final values for training and validation loss, suggesting the models can generalise the atmospheric chemistry of CH_4, CO and OH.

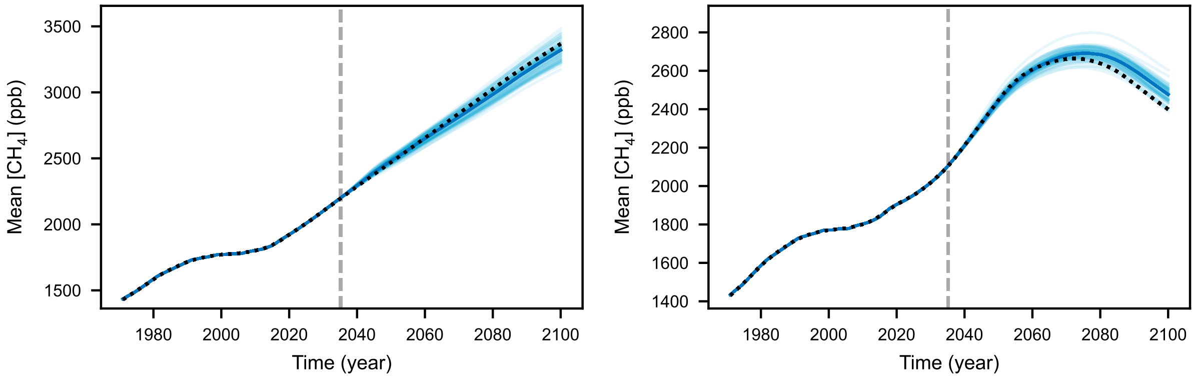
When considering the validation loss curves for the individual cross-validation folds (plotted in shaded red), it is evident that two folds fail to converge to the same extent as the other folds. This behaviour is most apparent in SSP2-4.5, SSP3-7.0 and SSP5-8.5. Here, the early stopping algorithm quits the simulation before the cut-off point of 500 epochs can be reached. Nonetheless, these curves show a larger gap between training and validation loss before early stopping is triggered. The corresponding curves can be attributed to the folds for SSP4-6.4 and SSP5-8.5 in all scenarios.

In addition to these general insights, an interesting scenario-specific observation can be noted. In scenario SSP3-7.0, the validation loss ultimately falls below the training loss. Unlike the opposite phenomenon, where validation loss increases while training loss decreases, this is not an indicator of model overfitting. This is more likely attributed to the L2 regularisation applied to the model. L2 regularisation penalises large weights by adding a term proportional to the squared magnitude of the weights to the loss function, effectively constraining the model's capacity to overfit. Therefore, the model generalises better to unseen data, which can result in the validation loss being lower than the training loss during certain stages of training.



(a) SSP1-2.6

(b) SSP2-4.5



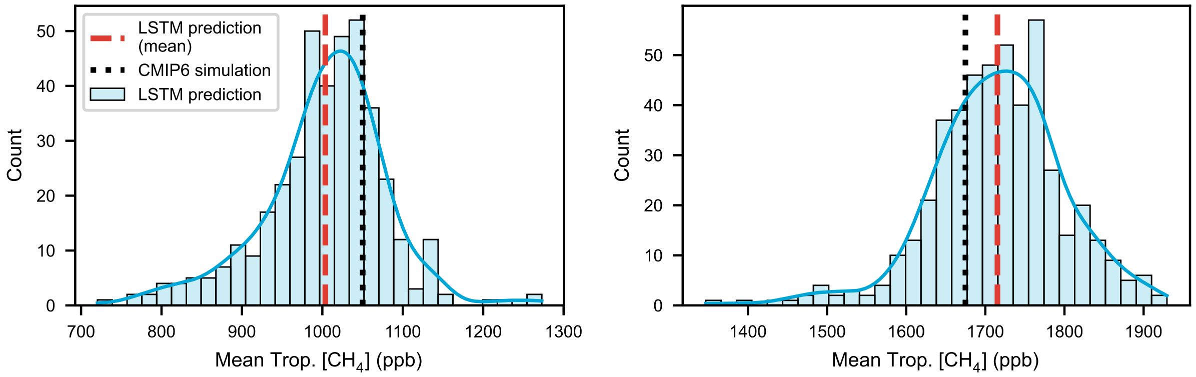
(c) SSP3-7.0

(d) SSP5-8.5

Figure 4.3: Temporal evolution of mean tropospheric CH_4 molar fractions (expressed in ppb) as simulated by the MRI-ESM2.0 model for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)** (black dots), plotted against the LSTM predictions (shaded blue) and the mean of the LSTM predictions (bold blue). The model simulates from the year 2035 onwards. Note that the y-axis is not scaled evenly between the different scenarios.

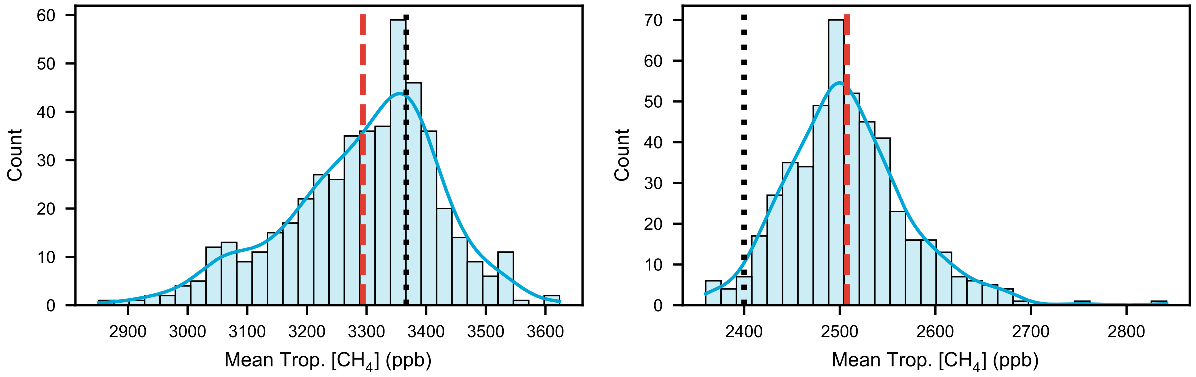
The trained models are used to make autoregressive predictions, as outlined in section 3.3. In total, models are trained on seven SSP scenarios. For every model, the test scenario is omitted. This results in six cross-validation folds for each scenario. Multiple models can be trained for each validation fold to prove that the error induced by the autoregression is purely stochastic and not systematic. In this analysis, 80 models are trained for each fold, resulting in 480 models per scenario.

Figure 4.3 shows the resulting forecast of each model, together with the mean of all forecasts, plotted against the simulated forecast from MRI-ESM2.0 on which the models are trained. After the start of the forecast in 2035, the spread in prediction steadily grows towards the end-tail of the considered period. This effect is due to error propagation and is detailed in section 3.3. This behaviour results in a distribution of predictions around the mean at the end of the simulation in 2100. The distribution at this time can be visualised using histograms, as shown in figure 4.4.



(a) SSP1-2.6

(b) SSP2-4.5



(c) SSP3-7.0

(d) SSP5-8.5

Figure 4.4: Histograms highlighting the variance of the predictions at the end-date of the simulations for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**. The spread of predictions (cyan) around the mean of predictions (dashed red) represents the variance in the results. The difference between the mean of the predictions and the simulated result from MRI-ESM2.0 is the bias of the result.

The predictions form a normal distribution around the mean, allowing for analysis of their basic statistical properties. First, the variance of the normal distribution indicates the spread of the predictions, providing insight into the model's precision. A smaller variance suggests more consistent predictions, while a larger variance indicates greater uncertainty. Conversely, bias represents a measure of systematic error in the model. It is defined as the difference between the mean of the model predictions (LSTM) and the reference values (CMIP6). A small bias suggests that the errors primarily stem from stochastic variability inherent in the autoregressive process. In contrast, a large bias indicates significant systematic errors, which suggests that the surrogate model is unfit for forecasting purposes. Table 4.1 provides an overview of the bias and standard deviation of the predictions for each scenario.

The standard deviation (SD) of the predictions is relatively consistent across scenarios, with only SSP3-7.0 showing a notable outlier in absolute terms, indicating higher uncertainty in that specific case. The other scenarios exhibit lower SDs, suggesting the model's predictions are consistent.

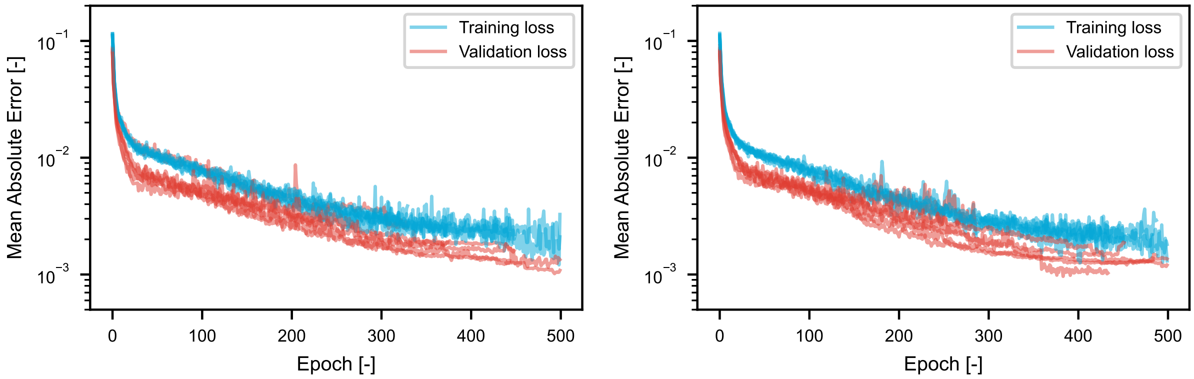
When considering the bias, the values across most scenarios are reasonably low, a positive indicator of the model's performance. For instance, SSP1-2.6 and SSP3-7.0 show small negative biases (-4.4% and -2.2%, respectively), suggesting only a slight underestimation of the predicted values compared to the CMIP6 reference. SSP2-4.5 and SSP5-8.5 also show a relatively small positive bias of 2.4% and 3.3%, indicating that the model's predictions are closely aligned with the reference values in that scenario.

Table 4.1: Overview of the bias and variance (expressed in absolute terms and relative to the predicted value) of the predictions for the considered scenarios at the end-of-simulation in 2100.

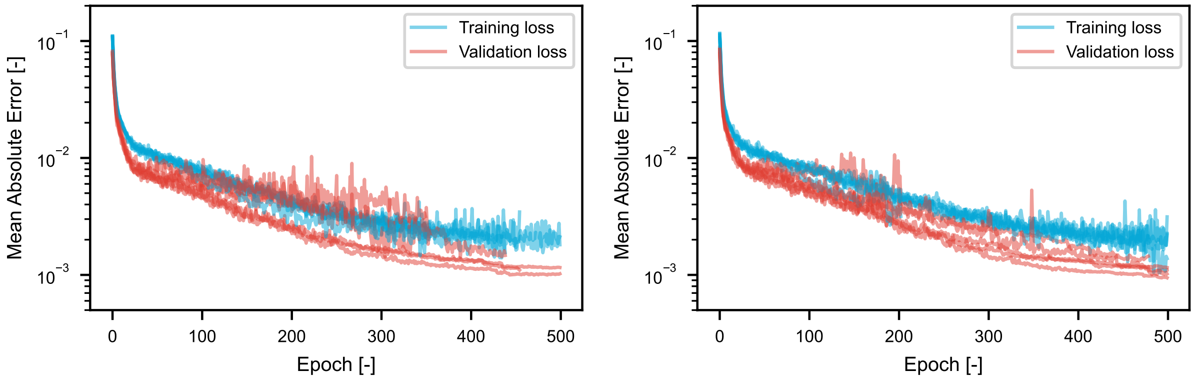
Scenario	Bias (ppb)	Bias (%)	SD (ppb)	SD (%)
SSP1-2.6	-46.4	-4.4	76.9	7.3
SSP2-4.5	40.4	2.4	82.9	4.9
SSP3-7.0	-72.8	-2.2	128.5	3.8
SSP5-8.5	110.6	3.3	63.0	2.6

Stratospheric Water Vapour

The machine learning surrogate models developed for SWV predictions undergo a similar verification process as outlined for CH_4. These models are trained to emulate the atmospheric chemistry of SWV for various SSP scenarios. Figure 4.5 illustrates the training and validation loss for each model, highlighting their convergence and generalisation capabilities. This analysis focuses on four widely utilised SSP scenarios, with supplemental results provided in appendix C.



(a) SSP1-2.6 **(b)** SSP2-4.5



(c) SSP3-7.0 **(d)** SSP5-8.5

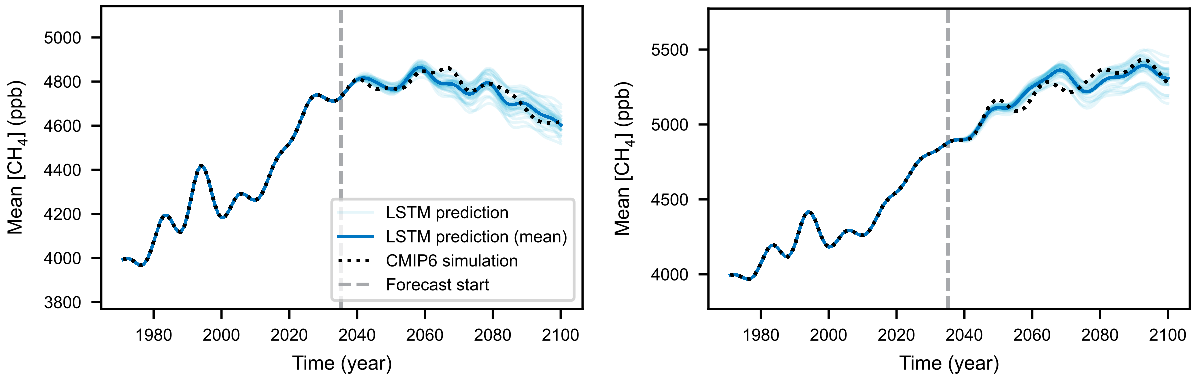
Figure 4.5: Training loss (cyan) and validation loss (red) for the SWV LSTM models tested on scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**, plotted as a function of epoch number on a logarithmic scale.

The loss curves for SWV models exhibit a decreasing trend in training and validation loss across all scenarios. The initial steep decline indicates rapid learning during the early epochs, typical of neural network training. As training progresses, the loss curves plateau, suggesting that the models are approaching convergence by the 500th epoch. Notably, no significant difference exists between training and validation loss in the later epochs, indicating that the models are not overfitting and can generalise well to unseen data.

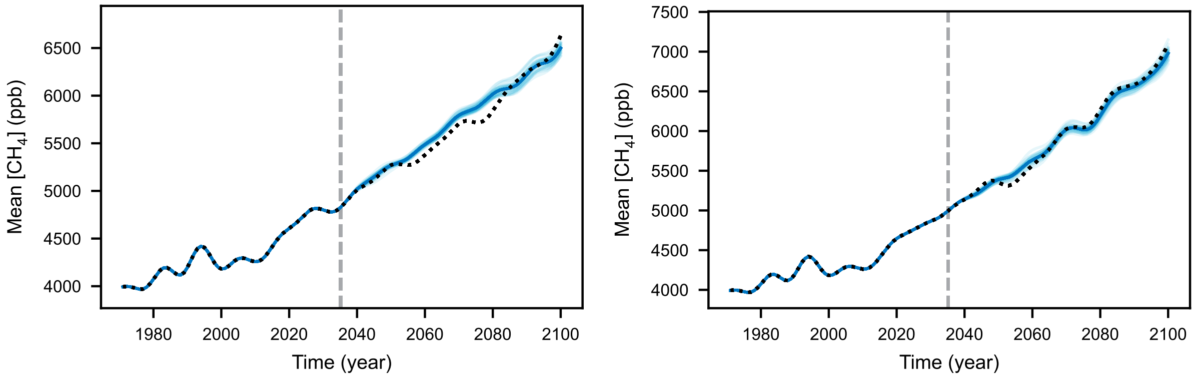
Minor variations are observed when examining the validation loss curves for individual cross-validation folds (plotted in shaded red). However, these variations are consistent across scenarios and do not indicate any significant issues with convergence or overfitting. This consistency reinforces the robustness of the training process and the models' ability to capture the underlying atmospheric chemistry of SWV.

The trained models are then used to make autoregressive predictions, as outlined in section 3.3. For each SSP scenario, models are trained using data from the other scenarios, resulting in six cross-validation folds per scenario. Multiple models are trained for each fold to account for the stochastic nature of autoregression. In this analysis, 80 models are trained per fold, totalling 480 models for each scenario.

Figure 4.7 presents histograms depicting the distribution of SWV predictions at the end of the simulation in 2100. The projections form approximately normal distributions around their means, allowing for statistical analysis of variance and bias. The spread of the predictions (cyan bars) around the mean (dashed red line) illustrates the variance. In contrast, the difference between the mean of the projections and the MRI-ESM2.0 simulation reference (black line) represents the bias.



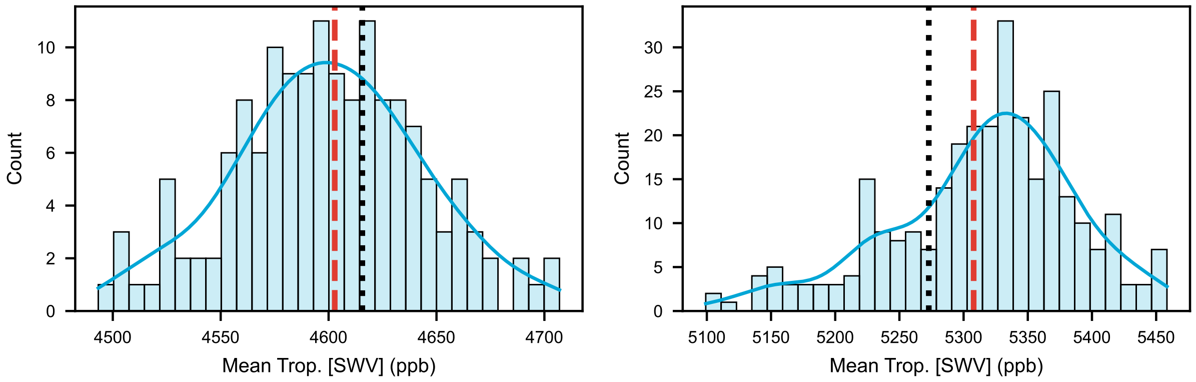
(a) SSP1-2.6 **(b)** SSP2-4.5



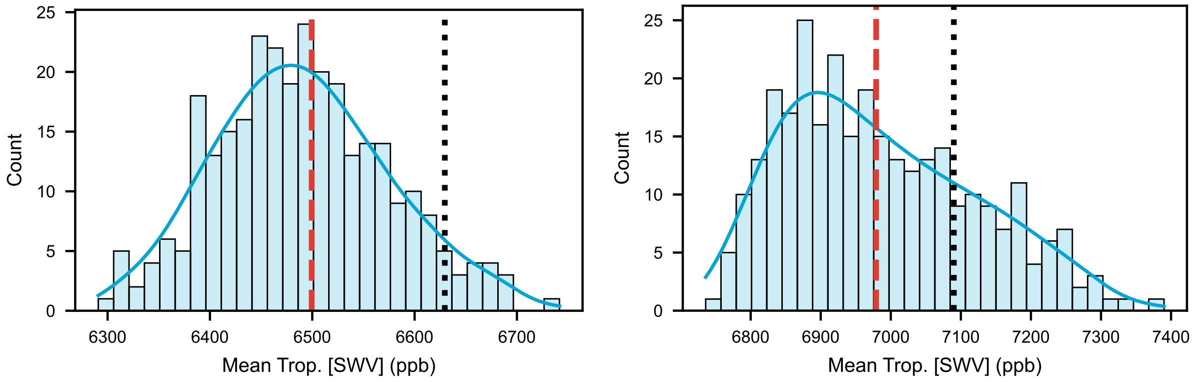
(c) SSP3-7.0 **(d)** SSP5-8.5

Figure 4.6: Temporal evolution of mean tropospheric SWV molar fractions (expressed in ppb) as simulated by the MRI-ESM2.0 model for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)** (black dots), plotted against the LSTM predictions (shaded blue) and the mean of the LSTM predictions (bold blue). The model simulates from the year 2035 onwards. Note that the y-axis is not scaled evenly between the different scenarios.

The variance and bias metrics are quantified in table 4.2. The variance, represented by the standard deviation (SD) of the predictions, provides insight into the model's precision. A lower SD indicates that the predictions are closely clustered around the mean, reflecting higher precision. Bias, calculated as the difference between the mean of the predictions and the CMIP6 reference value, indicates the model's accuracy. A small bias suggests minimal systematic error.



(a) SSP1-2.6 **(b)** SSP2-4.5



(c) SSP3-7.0 **(d)** SSP5-8.5

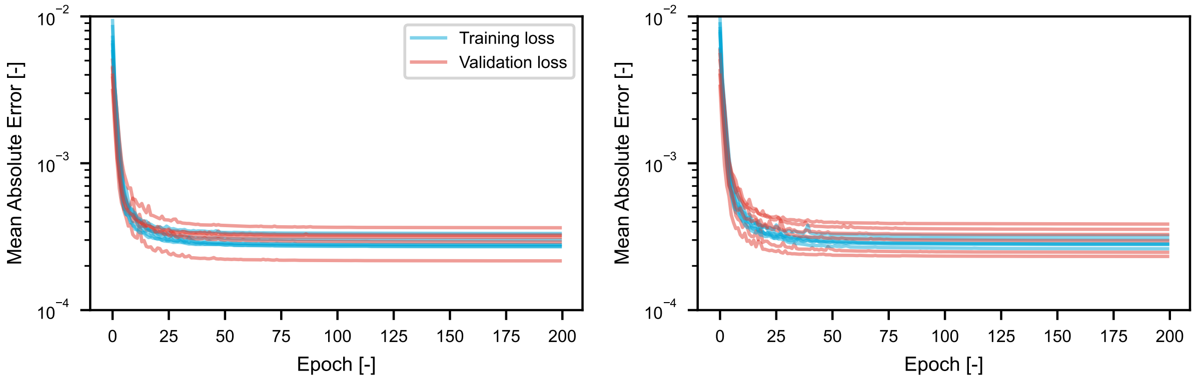
Figure 4.7: Histograms highlighting the variance of the predictions at the end-date of the simulations for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**. The spread of predictions (cyan) around the mean of predictions (dashed red) represents the variance in the results. The difference between the mean of the predictions and the simulated result from MRI-ESM2.0 is the bias of the result.

Table 4.2: Overview of the bias and variance (expressed in absolute terms and relative to the predicted value) of the SWV predictions for SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 at the end-of-simulation in 2100.

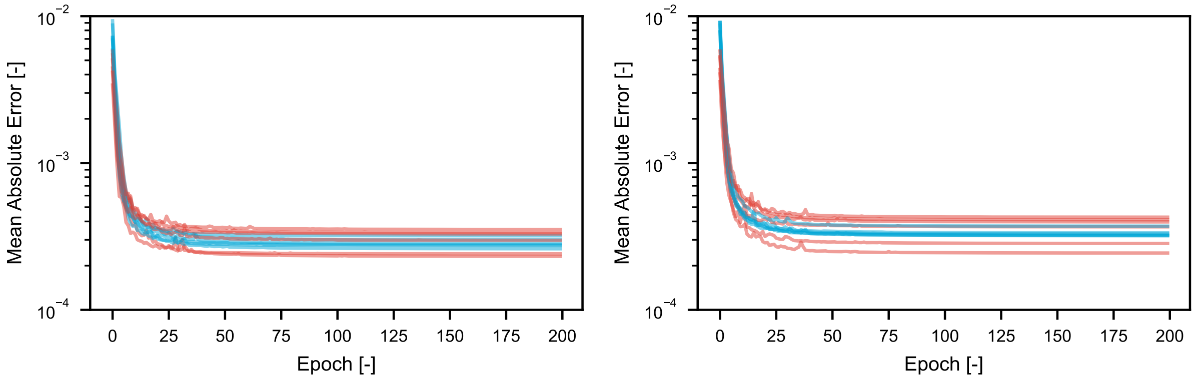
Scenario	Bias (ppb)	Bias (%)	SD (ppb)	SD (%)
SSP1-2.6	-12.8	-0.3	39.6	0.9
SSP2-4.5	34.5	0.8	68.1	1.5
SSP3-7.0	-130.1	-3.0	47.2	1.1
SSP5-8.5	-110.9	-2.5	67.4	1.5

ConvLSTM - Ozone

A similar verification approach can be applied to the ConvLSTM model, which assesses the spatiotemporal evolution of ozone concentrations due to hydrogen emissions. Figure 4.8 shows the training and validation loss for the ConvLSTM simulations. Due to the higher computational cost of resolving a spatial grid, the simulation cut-off point was set at 200 epochs. The losses for the LSTM and ConvLSTM models are not directly comparable due to differences in data preparation techniques, as discussed in section 3.3.2.



(a) SSP1-2.6 **(b)** SSP2-4.5



(c) SSP3-7.0 **(d)** SSP5-8.5

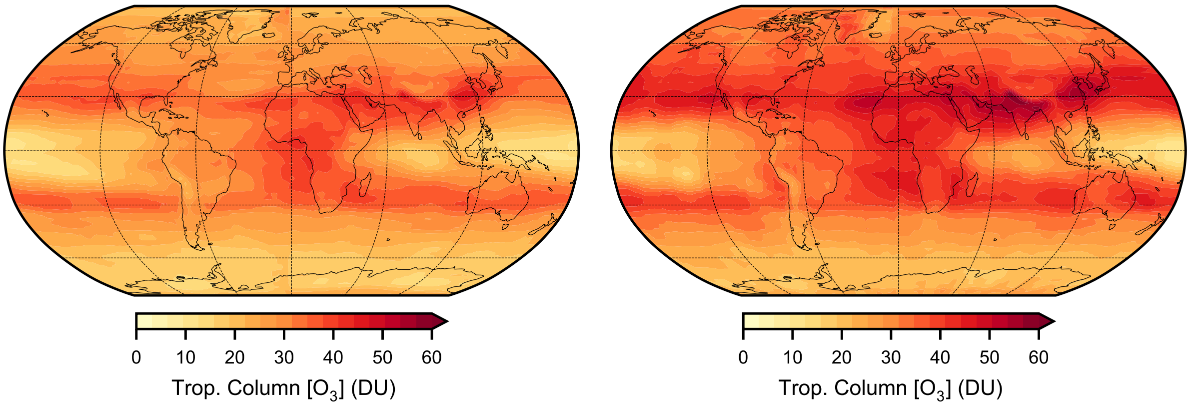
Figure 4.8: Training loss (cyan) and validation loss (red) for the O_3 ConvLSTM models tested on scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**, plotted as a function of epoch number on a logarithmic scale.

Similar to the loss history presented in figure 4.2, these plots show that the model converges as the simulation progresses through the epochs. For every scenario, the validation loss closely trails the training loss. There are no sudden increases in validation loss visible that indicate overfitting. Furthermore, apart from minor and short-lived increases in training loss, the simulations are free of erratic behaviour that would suggest problems with convergence.

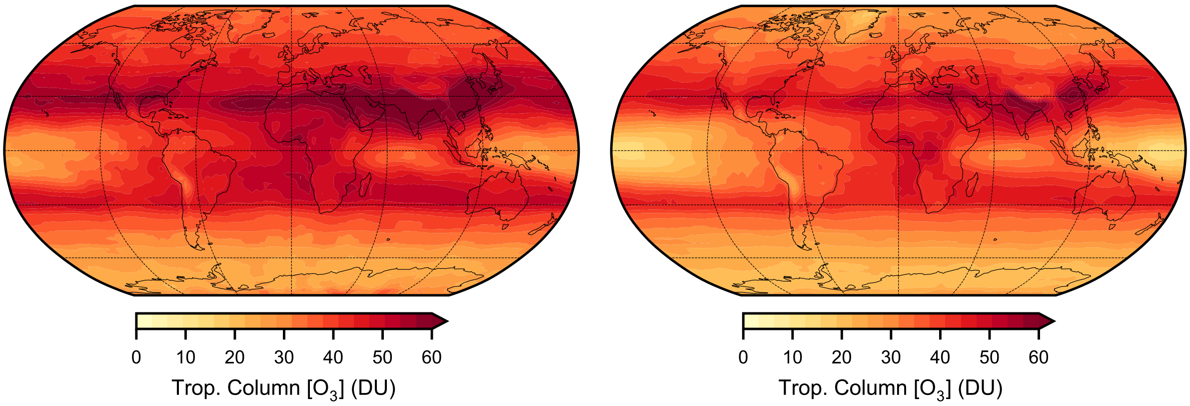
However, compared to the LSTM model, there are no cases of early stopping. This means that the maximum number of epochs may be set too low. If the model were allowed to converge for a longer period, the loss might decrease, yielding more accurate results. Nonetheless, the minor decreases achieved towards the latter epochs suggest that further loss reduction might not be worth the additional computational cost associated with longer simulation.

Given the dimensionality of the ConvLSTM predictions, a natural way to present the results is by projecting them onto a globe. Figure 4.9 shows the predicted ozone concentrations for the four SSP scenarios under study in the year 2100. One Dobson Unit (DU) corresponds to a layer of ozone that would be 0.01 millimetres thick at the surface. Only tropospheric ozone concentrations are considered, leading to lower values than visualisations that include stratospheric ozone.

Each scenario exhibits a similar latitude-dependent distribution of tropospheric ozone. The projection shows relatively low surface-level ozone concentrations in tropical regions (0°–30° latitude). This low concentration is primarily due to convective activity, which transports ozone-poor air from the boundary layer to higher altitudes. This results in a dilution effect that minimises tropospheric ozone levels in these areas. Conversely, the mid-latitudes (30°–60° latitude) exhibit significantly higher tropospheric ozone concentrations. The map indicates elevated ozone levels in regions with substantial anthro-



(a) SSP1-2.6 **(b)** SSP2-4.5



(c) SSP3-7.0 **(d)** SSP5-8.5

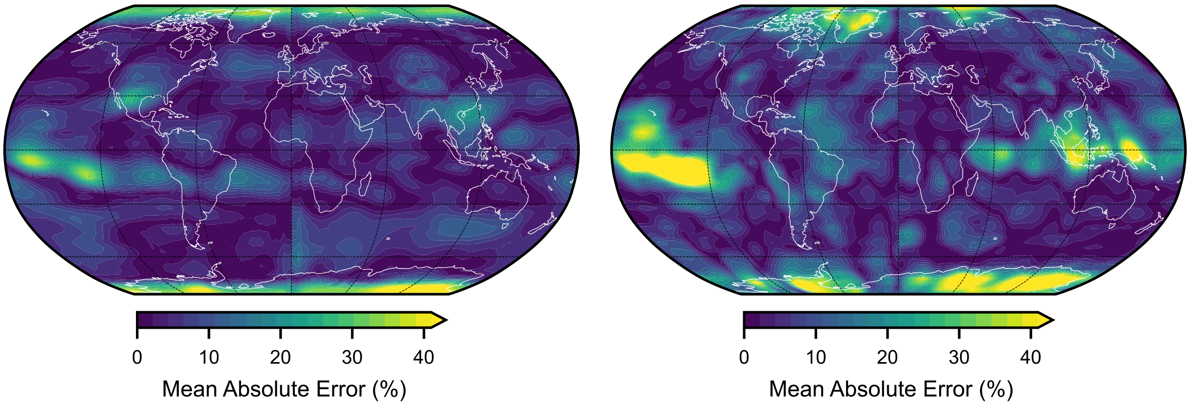
Figure 4.9: Mean tropospheric column ozone in 2100 as projected by the ConvLSTM model for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**. Ozone levels are presented in Dobson units (DU) on a scale from zero to 60, with the darkest shade of red denoting any value higher than 60 DU.

pogenic emissions of ozone precursors from industrialised and urbanised sectors. Under moderate UV radiation conditions, these precursors enhance photochemical ozone production. Additionally, atmospheric transport mechanisms, such as the subtropical jet stream, facilitate the movement of ozone-enriched air from higher latitudes and the stratosphere into mid-latitude regions, further contributing to elevated ozone concentrations observed in these areas.

At high latitudes (>60° latitude), the plots show generally lower tropospheric ozone concentrations, attributed to reduced solar radiation that limits photochemical ozone production. Seasonal variations of ozone levels are effectively filtered out by the Butterworth filter described in section 3.3.2, and thus not visible. The overall latitude-dependent distribution of tropospheric ozone shows a clear gradient: lower ozone levels in the tropics due to convective dilution, elevated concentrations in the mid-latitudes driven by anthropogenic emissions and atmospheric transport, and lower ozone levels in high latitudes influenced by decreased photochemical activity.

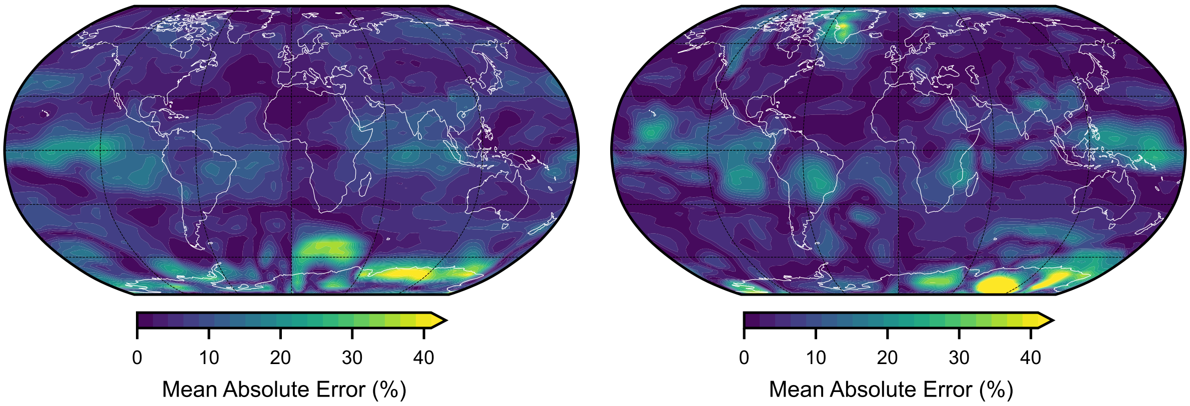
Lastly, by 2100, SSP scenarios modelling lower radiative forcing show lower overall mean tropospheric ozone concentrations compared to scenarios modelling higher radiative forcing. This increase is primarily visible near urban areas around the mid-latitudes, especially near Southeast Asia, the Middle East, and the Mediterranean. This is likely related to the increase in O_3 precursor for the projected climate scenarios in these regions.

To assess the accuracy of the spatial resolution of the model, the results from figure 4.9 can be compared to the corresponding CMIP6 projections. Figure 4.10 shows the mean absolute error between the ConvLSTM and CMIP6 projections for the considered SSP scenarios. Here, darker shades indicate a lower error, while lighter colours indicate a larger error. This figure demonstrates the model's capacity to predict ozone's spatial distribution accurately.



(a) SSP1-2.6

(b) SSP2-4.5



(c) SSP3-7.0

(d) SSP5-8.5

Figure 4.10: Spatially resolved mean absolute error (MAE) as a percentage between the ConvLSTM projections and CMIP6 data for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)**.

Each projection reveals substantial errors near the polar regions, with SSP1-2.6 and SSP2-4.5 also exhibiting pronounced inaccuracies in the tropics, particularly over Southeast Asia and the Central Pacific. When examining these patterns in conjunction with figure 4.9, it becomes evident that the errors predominantly occur in regions of low absolute ozone concentrations. Depending on the model, errors in these areas can reach up to 40%, highlighting a significant limitation in capturing ozone levels in these specific zones.

Conversely, the mid-latitudes, characterised by higher ozone concentrations, display relatively smaller errors. This pattern suggests that the model performs better in regions where precursor emissions strongly influence ozone production. The pronounced errors near the polar regions may also relate to the seasonal variability of ozone in these areas. Since the surrogate model does not explicitly account for seasonal fluctuations, discrepancies arise when compared to actual observations or high-resolution models that incorporate this variability.

The spatial distribution of bias and variance provides additional insights into the model's accuracy, as summarised in table 4.3. For instance, SSP1-2.6 and SSP2-4.5 exhibit negative biases, with SSP2-4.5 showing a particularly significant underestimation of ozone levels. Conversely, SSP3-7.0 and SSP5-8.5 show positive biases, indicating a tendency to overestimate global ozone concentrations. The variance, represented by one standard deviation, remains relatively consistent across scenarios, though it is lower for SSP3-7.0 and SSP5-8.5 compared to SSP1-1.9 and SSP2-4.5. This pattern, where bias remains consistently lower than variance, indicates that although the model's average predictions closely match the actual values, it lacks precision, resulting in variability around the expected outcomes.

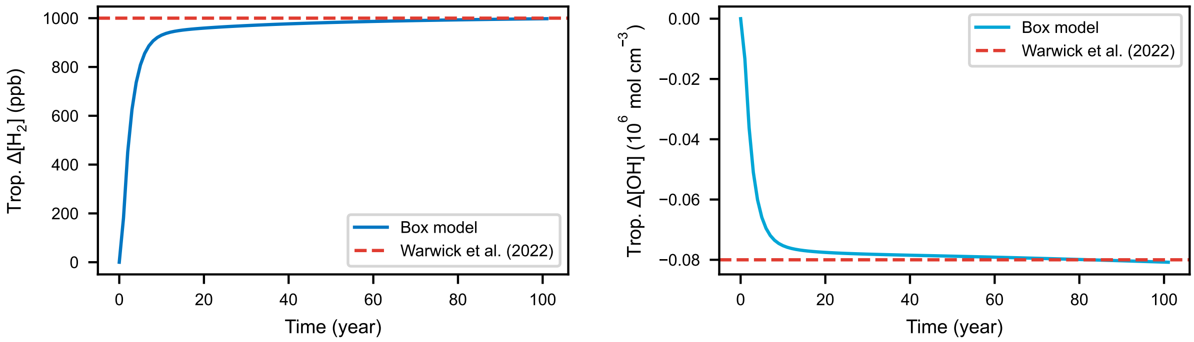
Table 4.3: Spatial bias and variance of the error between ConvLSTM O_3 predictions and CMIP6 O_3 simulations for each scenario at the end-of-simulation in 2100.

Scenario	Bias (DU)	Bias (%)	SD (DU)	SD (%)
SSP1-2.6	-1.05	-2.1	3.11	11.0
SSP2-4.5	-2.91	-9.0	3.35	10.4
SSP3-7.0	1.92	5.1	2.49	6.6
SSP5-8.5	0.19	0.6	2.95	8.9

4.2. Validation

During validation, a model is assessed for its effectiveness in solving the problem it was designed for. This includes evaluating its accuracy on training data and its ability to predict the impact of hydrogen emissions on atmospheric composition. This section compares the models' response to hydrogen emissions to similar experiments in literature, most notably those performed by Warwick et al. (2022). It is important to note that for subsequent results are obtained using a model trained on all folds, in contrast to the verification results which are derived using a cross-validation approach. The difference between these two models is explained in section 3.3.2.

The model is initialised under the same initial conditions as the baseline experiment from Warwick et al. (2022). Subsequently, two experiments are performed. The first experiment captures the model's response without hydrogen emissions. Subsequently, a second experiment is perturbed by a step in hydrogen emissions, increasing mean tropospheric hydrogen levels by 1000 ppb. The difference between these two experiments is analysed to obtain the model's specific response to hydrogen emissions. To make the results comparable to those provided by the study of Warwick et al. (2022), background concentrations of all species are fixed at their 2014 values. The response of OH simulated by the box model is shown in figure 4.11.



(a) **(b)**

Figure 4.11: Box-model response of H_2 to a step of 420 ppb/year in H_2 emissions **(a)**, resulting in a decrease in OH concentrations **(b)** over 100 years.

Using the same input parameters, and thus excluding the effect of additional CH_4 sinks and CH_4 feedback modelled by Bertagni et al. (2022), the box model used in this thesis is the same as the one employed by Warwick et al. (2022). Therefore, the response of OH to H_2 emissions approaches the steady state solution given by the same study. This result allows the direct comparison between the results from Warwick et al. (2022) and the results from the surrogate models and lays the foundation for the subsequent analysis of the model's effect on CH_4, O_3 and radiative forcing.

Figure 4.12 shows the steady state response of the surrogate models to a perturbation of OH. The error bars are generated using the averaged results of the last 25 years of simulation. The results are compared to results from the same experiment performed by Warwick et al. (2022) using the UKESM1 model and by Sand et al. (2023) using the GFDL, INCA, OsloCTM, and WACCM models.

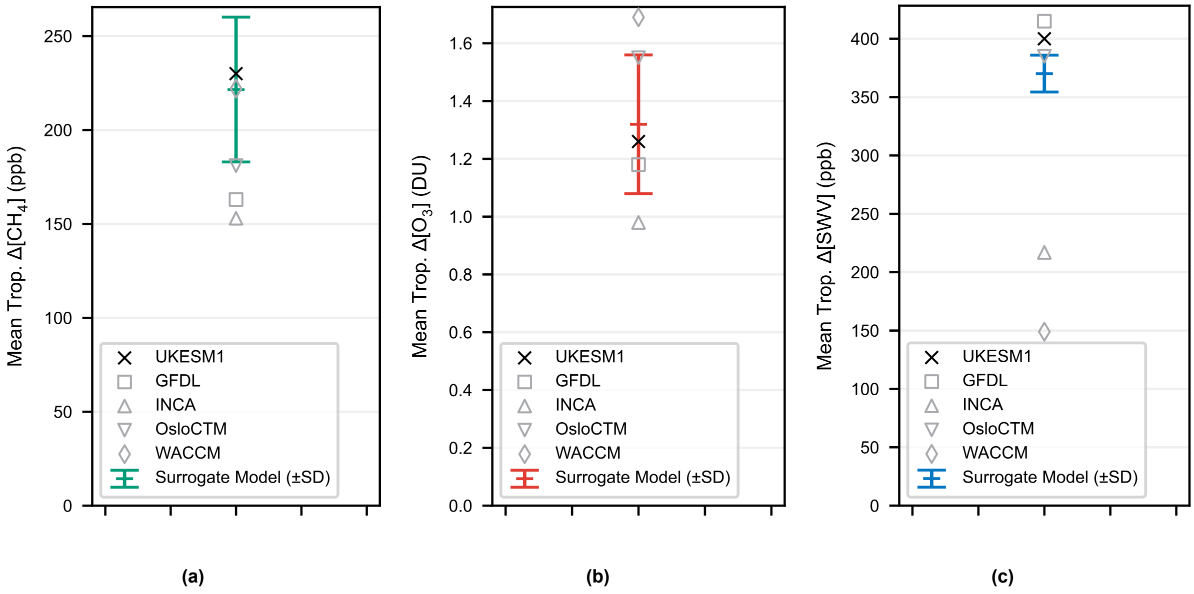


Figure 4.12: Response of the LSTM and ConvLSTM models to an OH perturbation, analysed for CH_4 (**a**), O_3 (**b**), and SWV (**c**). The results are compared with outputs from the UKESM1 model (Warwick et al., 2022) as well as the GFDL, INCA, OsloCTM, and WACCM models (Sand et al., 2023). Each result represents the mean value from the final 25 years of the simulation, with error bars reflecting one standard deviation.

The response of CH_4 shows a significant spread when compared to the spread of the results from the various models. This spread is caused by the autoregressive nature of the models, where small errors at the early stages of the projection can propagate over time. The mean of the projections is seen to be in broad agreement with the results from the UKESM1 model used in the study by Warwick et al. (2022) in particular. When compared to the results from Sand et al. (2023), the model seems to overestimate the response of CH_4 to a perturbation in OH.

Similar results are noted for the simulation of tropospheric O_3. The mean of the ensemble of models again aligns with the results from the UKESM1 model simulation, showing an increase of 1.23 DU in tropospheric O_3 in response to an increase of 1000 ppb in tropospheric H_2. The spread in the results relative to the spread in the climate model results is significantly smaller. This can be attributed to the complexity of modelling O_3 causing a larger discrepancy between different climate models.

Finally, section 4.2 shows the results of the SWV simulations. Clearly, the INCA and WACCM models differ significantly in their method of simulation SWV, resulting in significantly smaller values of SWV perturbation compared to the UKESM1, GFDL, and OsloCTM models. The results from the surrogate model show a minor spread compared to the spread in the climate models. This is a consequence of the current methodology employed to derive the effect of H_2 emissions on SWV levels. The surrogate model only computes methane-induced perturbations of SWV, with non-methane-induced perturbations calculated using simplified expressions as explained in section 3.3.3. As the non-methane-induced part accounts for about 300 ppb of SWV, the spread in the final results is relatively small.

The spatial resolution of the increase in tropospheric O_3 is plotted in figure 4.13. Only changes higher than 0.1 DU are shown. In agreement with Warwick et al. (2022), the model predicts increases in O_3 to be centred around the tropics and mid-latitudes, with the largest increases visible near Southeast Asia, the Middle East, the area spanning from Middle America to Africa and parts of the Pacific.

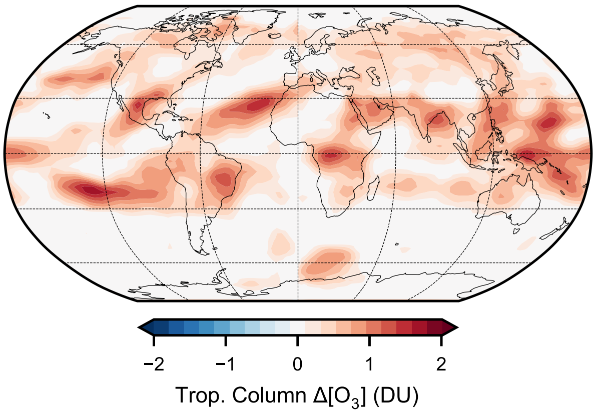


Figure 4.13: Spatially resolved response of O_3 to a perturbation of CH_4. Only changes with a magnitude higher than 0.1 DU are shown.

Radiative Forcing

The changes in tropospheric CH_4 and O_3 are converted to radiative forcings using the scaling factors shown in table 3.4. Figure 4.14 plots the contribution of CH_4, O_3 and SWV to the total radiative forcing, normalised per ppb increase in H_2.

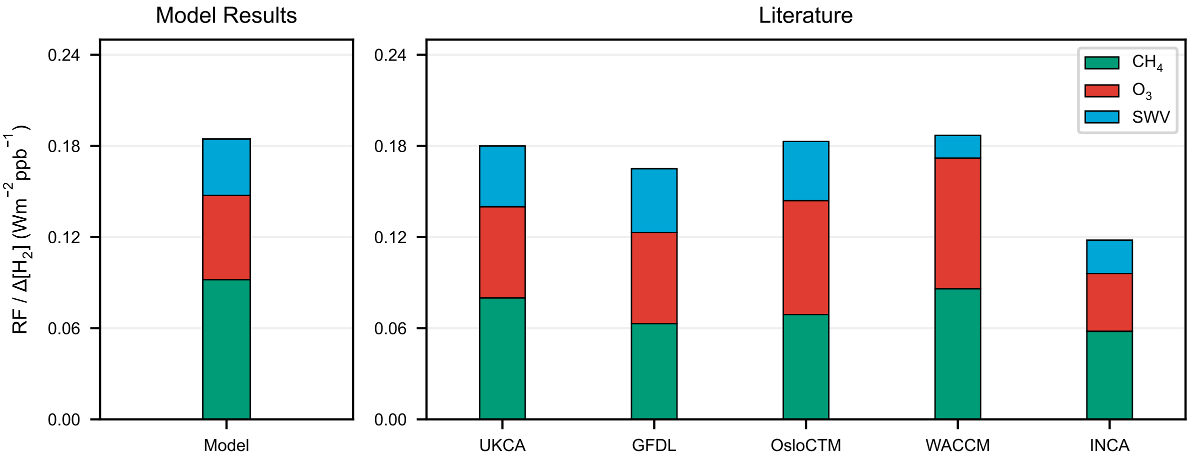


Figure 4.14: Comparison of the resulting radiative forcing normalised for unit increase in H_2 mixing ratios for the validation experiments compared to similar results from the models used in the studies by Warwick et al. (2022) and Sand et al. (2023).

The figure presents the RF results for the surrogate models developed in this study alongside results from the literature, specifically the UKESM1, GFDL, OsloCTM, WACCM, and INCA models. The vertical axis shows the RF per ppb increase in H_2 (W/m^2 per ppb), while the horizontal groupings compare the RF values from this study's models to those from previous studies.

From the results, CH_4 and O_3 emerge as the dominant contributors to RF, while SWV shows a smaller but consistent contribution. The surrogate models developed here show good agreement with the literature values for CH_4 and O_3, indicating their ability to replicate known climate impacts of H_2-driven perturbations. However, slight differences are observed for SWV, which may reflect variations in how different models represent stratospheric water vapour processes.

The consistency between the surrogate models and literature values supports the models' robustness and their applicability in projecting the climate impacts of H_2 emissions. Furthermore, the results provide confidence in using these models for further analyses, such as quantifying uncertainties through Monte Carlo simulations or exploring the effects of alternative emission scenarios.

4.3. Sensitivity Analysis

Understanding the impact of key hyperparameters is essential for achieving optimal performance in the development and optimisation of neural network models for predictive analytics. This section presents a sensitivity analysis focused on two critical hyperparameters that have not been optimised using the tuner described in section 3.3.2: batch size and window size. Furthermore, the impact of selecting the training set is analysed by presenting results from training the model exclusively on historical data.

This sensitivity analysis provides insights into the LSTM model's robustness and adaptability by systematically varying these parameters and assessing their effects on model performance. Sensitivity analysis is computationally expensive, as a full ensemble of models needs to be trained for each set of hyperparameters analysed. The ConvLSTM model is significantly more computationally expensive to train due to the larger two-dimensional datasets required for training. Full sensitivity analysis of this model, therefore, falls outside of the scope of this research. The following presents results only for the LSTM model.

Sensitivity to Window Hyperparameters

Batch size and window input width are two hyperparameters that play crucial roles in the training efficiency and performance of machine learning models. Smaller batch sizes can lead to noisy gradient estimates but may help the model escape local minima, while larger batch sizes provide more stable gradient estimates but require more memory and may converge to suboptimal solutions. Additionally, the choice of window size determines the amount of historical information the model considers, affecting its ability to capture temporal dependencies and patterns within the data. An appropriately selected window size can enhance the model's predictive accuracy by providing sufficient context, whereas an inadequate window size may lead to underfitting or overfitting. Balancing these hyperparameters is essential for optimising model performance, ensuring that the model learns effectively from the data while maintaining computational efficiency.

The effect of these parameters is analysed by computing the bias and variance in the model ensembles at end-of-simulation, similarly to section 4.1. Figure 4.15 shows the batch size and window input size heatmap evaluated on bias and variance.

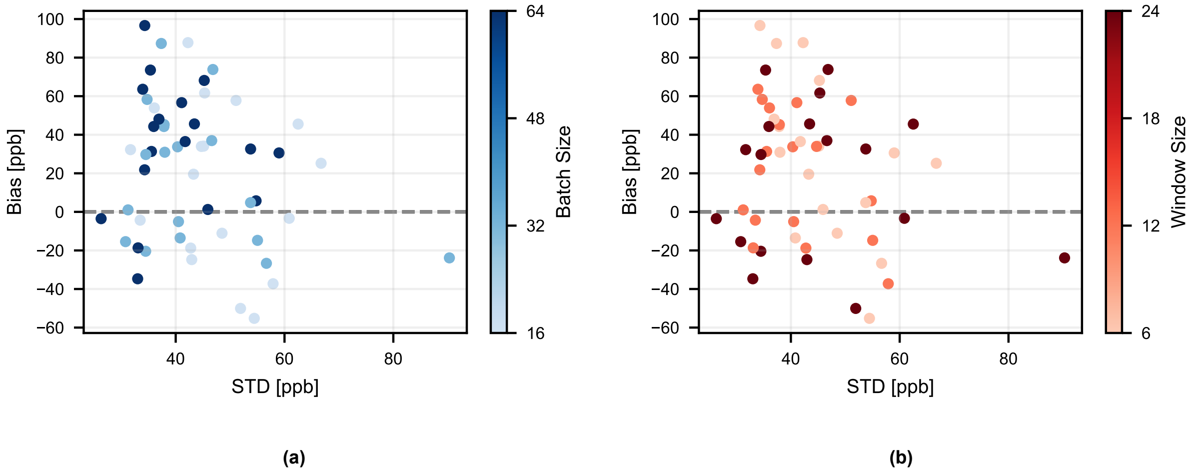


Figure 4.15: Correlation plot between the bias and variance for the nine simulations performed. An increase in batch size **(a)** or window size **(b)** is plotted as a gradient. Each experiment contains mean ensemble results from all six scenarios (SSP1-1.9 - SSP5-8.5), resulting in 54 results per graph, or 18 per hyperparameter value.

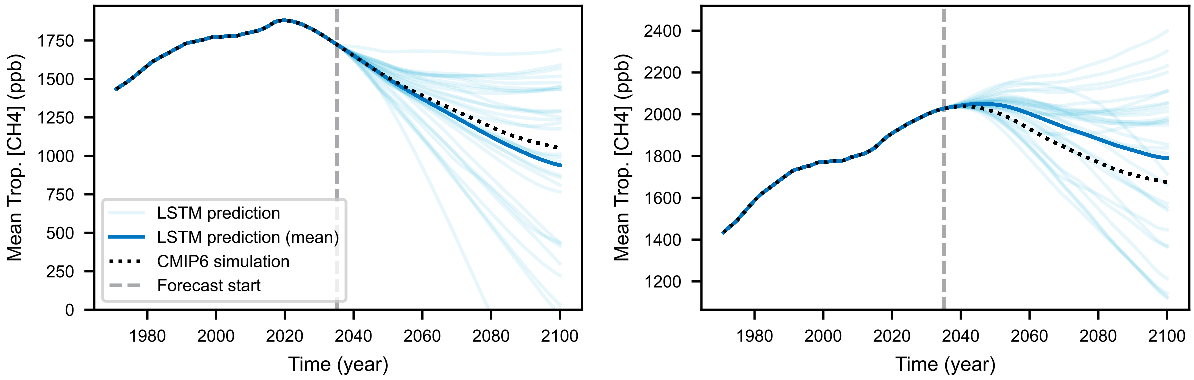
In the batch size plot, darker colours shift slightly towards the left compared to lighter shades. This indicates a slight trend where an increase in batch size decreases the variance of the model ensemble. Furthermore, a slight upward trend is visible in the bias, indicating that the model tends to overestimate more on average compared to smaller batch sizes. Neither of these trends is visible in the window size plot, where the data points are scattered rather randomly.

The absolute effect of increasing batch or window size is insignificant compared to the overall variance or bias, demonstrating that the model is robust to changes in these hyperparameters. These trends can be attributed to the effects of batch size on the training dynamics of the model ensemble. Larger batch sizes result in more accurate and stable gradient estimates, reducing forecast variance by providing a more comprehensive representation of the data distribution. While this enhanced stability improves consistency across runs, it can also constrain the model's ability to generalise, leading to higher bias as the model becomes less sensitive to data variability and converges to more rigid decision boundaries.

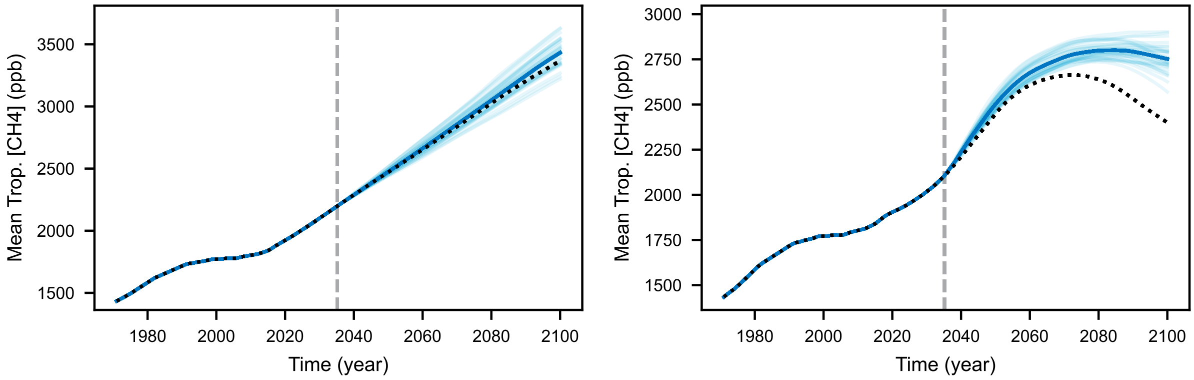
Sensitivity to Training Data

Machine learning models are data-driven; their accuracy, therefore, depends greatly on the quality of the training data. In this study, the machine learning models are trained on scenario simulations provided by the MRI-ESM2.0 model. This model has its underlying assumptions, uncertainties, and inaccuracies. Consequently, the machine-learning model might be overfitting while predicting outcomes that are not representative of real observations.

Therefore, it is useful to investigate the model's sensitivity to training data. By truncating the training data to 2014, the model will effectively be trained on simulations from the MRI model that are constrained using observations. The resulting temporal evolutions of the model ensembles for the four common SSP scenarios are provided in figure 4.16.



(a) SSP1-2.6 **(b)** SSP2-4.5



(c) SSP3-7.0 **(d)** SSP5-8.5

Figure 4.16: Temporal evolution of mean tropospheric CH_4 molar fractions (expressed in ppb) as simulated by the MRI-ESM2.0 model for SSP1-2.6 **(a)**, SSP2-4.5 **(b)**, SSP3-7.0 **(c)** and SSP5-8.5 **(d)** (black dots), plotted against the LSTM predictions (shaded red) and the mean of the LSTM predictions (bold red). The model simulates from the year 2035 onwards.

These graphs show a larger variance in the data near the tail-end of the simulations when compared to the simulations leveraging the full training dataset (figure 4.3). The variance is significantly greater for scenarios modelling future decreases in CH_4 emissions, such as SSP1-2.6 and SSP2-4.5. SSP3-7.0 remains relatively unchanged compared to the original simulations. There is a simple explanation for this phenomenon. The current historical dataset contains only data where CH_4 levels are increasing. This prevents the machine learning model from learning the relationship between the features that result in declining CH_4 emissions, causing significant divergence during inference.

4.4. Discussion

This section discusses the key findings from the model verification and validation process. It begins with an evaluation of the model's accuracy, followed by an examination of its limitations and potential applications. The section concludes with suggestions for optimizing the model architecture to enhance performance.

Model Accuracy

The loss curves for the LSTM-CH_4, LSTM-SWV and the ConvLSTM-O_3 model displayed in figure 4.2, figure 4.5 and figure 4.8 are free of erratic behaviour, show a steady convergence towards a plateau and don't exhibit any validation loss divergence throughout the simulation. This suggests that both models are training effectively without overfitting on any specific feature. Convergence between the different validation folds is also similar, further indicating that the model is generalising well between different climate projections. Simply put, this means that the model is capable of accurately predicting the response of CH_4 and O_3 at different background concentrations of their precursors. The LSTM model is seen to trigger early stopping at a lower epoch and conversely higher loss than the other folds. Because this behaviour is evident for all four testing scenarios, it can be concluded that the two scenarios (SSP4-6.4 and SSP5-8.5) less generalisable compared to the other scenarios. This phenomenon is not seen in the ConvLSTM loss curves, suggesting that the epoch cut-off is set too low. Further investigation of effect of the maximum epoch number and early stopping parameters is needed to optimise both models' learning behaviour.

The argument that the model is accurate and generalises well between the different scenarios is corroborated by the temporal evolution plots shown in figure 4.3. For all scenarios, the predictions form a Gaussian distribution, indicating that the individual error at end-of-simulation is attributable to random errors propogated during autoregression. The error of a prediction can be greatly reduced by taking the mean of an ensemble of models. The bias of the model ensamble is then indicative of the systematic error in the predictions, and can therefore provide insights into the accuracy of the machine learning model. The error for the both LSTM models is generally low, between 2% and 5%, which is deemed sufficiently accurate for the intended purpose of the model.

The spatial error in the ConvLSTM simulations shows that the model struggles to accurately forecast location-specific ozone perturbations over longer time periods. With errors >40%, the model should not be used to gain accurate insights into local effects of hydrogen emissions on tropospheric ozone. The assumptions used in this study that H_2 and CH_4 are well-mixed such that the box model can be used are not compliant with regional analysis, however, and never within the intended scope of the model. Table 4.3 shows that the model ensemble is capable of determining the mean tropospheric ozone levels within 10% of its simulated 2100 values. This outcome aligns with the model's design, which focuses on broad-scale trends over long time horizons rather than precise spatial predictions.

In section 4.3, the sensitivity of the model to key hyperparameters and training data selection is investigated. Minor improvements can be achieved by optimising for batch size, but this is a trade-off between memory usage during training and incremental decreases in model variance. Optimisation for batch size largely depends on the machine the model is run on and should be considered on a case-by-case basis.

The model is robust to changes in input width size. This indicates the model can effectively capture the necessary temporal dependencies irrespective of the specific window size. Consequently, this supports the choice of an LSTM model. These models are usually employed to model the long-term dependence of data but can also have distinct advantages in modelling atmospheric chemistry. Due to the large spread in atmospheric lifetimes and temporal dynamics of emissions and species, a 'normal' RNN might be more susceptible to overfitting to seasonal noise or incorrect dependencies between species. However, LSTM models are more computationally expensive. Future studies should, therefore, investigate the trade-off between computational cost and the suggested additional accuracy of an LSTM model.

The sensitivity analysis shows a significant dependency on the selection of training data. The substantial increase in variance for the lower SSP scenarios when only trained on historical data argues that this is a direct result of its limited scope. Because the historical data does not contain any decreases in methane, the model is incapable of learning the underlying patterns necessary to predict declining methane concentrations in these scenarios accurately. Furthermore, the reduction in size of the training dataset also increases the variance and bias of the higher SSP scenarios. This logically follows the principle that machine learning models are inherently data-driven, with their performance closely tied to the quality and quantity of the training data. Due to the computational cost associated with training the ConvLSTM model, the sensitivity analysis is only performed on the LSTM model. For a more comprehensive analysis, future studies should also replicate the sensitivity analysis for the ConvLSTM model. Nonetheless, given the similar recurrent structure of the models, it is expected that the findings will be largely consistent.

The low bias observed across all scenarios in both models, together with their robustness to hyperparameter variations as shown in the sensitivity analysis, suggests that the models generalise the atmospheric chemistry of CH_4 and O_3 accurately. The purpose of the model, however, is not necessarily accurate predictions compared to the underlying CMIP6 simulations. Rather, the aim of the model is to assess the effect of H_2 emissions on CH_4 and O_3, together with their radiative forcing. As such, validation is required to prove that the model can accurately model the consequences of H_2 perturbations on atmospheric composition. This is achieved by validating the model against similar experiments performed by Warwick et al. (2022) and Sand et al. (2023).

The surrogate model demonstrates strong agreement with results from the literature. Specifically, the validation results for CH_4 and O_3 closely align with the findings of Warwick et al. (2022). However, the largest discrepancy is observed in the SWV validation, where the model appears to underestimate SWV production. This is likely due to the simplified method used in this thesis to calculate non-methane-induced SWV perturbations. Nonetheless, conventional climate models such as INCA and WACCM exhibit even larger discrepancies, suggesting that the surrogate model provides a relatively robust and efficient alternative for assessing SWV-related impacts. Figure 4.14 shows how the surrogate models developed in this study compare with RF values from literature models, including UKESM1, GFDL, OsloCTM, WACCM, and INCA. The figure highlights the surrogate models' strong agreement with literature values for CH_4 and O_3, which dominate RF contributions, while showing slight discrepancies for SWV. Overall, these results provide great confidence that the models are capable to generalise the atmospheric chemistry of H_2, and are capable of providing accurate results for a wide range of future climate scenarios.

To account for the error resulting from the machine learning models, a standard error (SE) of 20% is appended to the Monte Carlo simulation for further analysis for both the CH_4 and O_3 models. This means the Monte Carlo simulation randomly samples an error factor from a Gaussian distribution with a mean of zero and a standard deviation of 0.2. The net increase in CH_4 or O_3 between a baseline and a perturbed simulation will then be multiplied by this factor.

Model Limitations and Use Cases

The surrogate model developed in this study employs a Long Short-Term Memory (LSTM) recurrent neural network to approximate the behavior of complex atmospheric processes. While this approach offers significant computational advantages, it is important to acknowledge its inherent limitations to ensure appropriate application and interpretation of the results.

One major limitation is the model's accuracy in capturing the full complexity of atmospheric chemistry. Despite its ability to model temporal dependencies, the LSTM neural network cannot encompass all the intricate interactions and nonlinear feedback mechanisms present in comprehensive climate models. This simplification can lead to discrepancies between the surrogate model's predictions and those of full-scale models, particularly under extreme or novel atmospheric conditions.

Furthermore, the model's performance is heavily dependent on the quality and representativeness of the training data. Since the LSTM neural network learns patterns from historical data, any biases, errors, or gaps in the underlying datasets can directly affect its predictive capabilities. The surrogate model is only as accurate as the data it was trained on and may not generalise well to conditions outside the range of the training set.

Despite these limitations, the primary strength of the LSTM-based surrogate model lies in its computational efficiency and ability to handle sequential data. Its rapid processing capabilities enable the execution of a large number of simulations in a fraction of the time required by full-scale climate models. This efficiency opens up valuable opportunities for scenario exploration and uncertainty quantification.

Use Cases

- **Scenario Analysis**: The LSTM surrogate model allows for efficient exploration of a wide array of emission scenarios and policy interventions. By rapidly simulating atmospheric responses to various inputs, it enables researchers to assess potential future states of the climate system under different assumptions. This capability is particularly useful for identifying key drivers of atmospheric composition changes and informing strategic decision-making in climate policy.
- **Uncertainty Analysis**: The computational speed of the LSTM model makes it well-suited for performing extensive uncertainty quantification. It facilitates the execution of large ensembles of simulations to evaluate how uncertainties in input parameters propagate through the model. This approach aids in quantifying confidence levels of predictions and in identifying the most influential variables affecting atmospheric chemistry and climate forcing.

Model Optimisation

To enhance the performance of the LSTM model in predicting the temporal evolution of methane concentrations, one potential improvement is the implementation of teacher forcing during training. Teacher forcing involves using the actual target outputs as inputs for the next time step during training, rather than the model's own predictions. This technique can mitigate the problem of error accumulation in autoregressive models, where small prediction errors can propagate and amplify over time, leading to significant deviations in long-term forecasts. By providing the true previous outputs during training, the LSTM can learn to correct its errors and improve its ability to handle sequences, ultimately enhancing the accuracy of long-term predictions.

The ConvLSTM model could be enhanced by increasing the depth of the convolutional layers, which could significantly improve its feature extraction capabilities. Adding more convolutional layers allows the model to learn hierarchical representations of the spatial data, capturing both low-level features like local gradients and high-level features such as broader spatial patterns and trends. Deeper architectures can enable the model to better understand complex interactions and dependencies in the atmospheric data, potentially leading to more accurate predictions of ozone concentrations across different regions and time periods.

Lastly, conducting a more extensive hyperparameter tuning process could further enhance the performance of both models. While the current approach employs random search within specified ranges for key hyperparameters, expanding the search space and utilising more sophisticated optimisation techniques like Bayesian optimisation or genetic algorithms could identify more optimal configurations. Additionally, incorporating cross-validation methods and experimenting with different architectures, activation functions, and regularisation techniques could help prevent overfitting and improve generalisation of unseen data. A thorough hyperparameter optimisation process is essential for maximising model performance and ensuring robust predictions in the context of future climate scenarios.

5
Scenario Analysis

This chapter investigates the effect of fugitive hydrogen emissions on atmospheric composition and radiative forcing using the model verified and validated in the previous chapter. First, section 5.1 presents the results from the Monte Carlo simulation. Afterwards, in section 5.2, three narratives are presented that will be explored in more detail. The results from the Monte Carlo simulation and case studies are discussed in section 5.3.

5.1. Uncertainty Analysis

The primary advantage of the model developed in this research is its ability to generate predictions rapidly, offering significant efficiency gains over conventional climate models. Traditional models, while highly accurate, require powerful supercomputers and extensive computational resources, often taking weeks or months to complete a simulation due to their complexity, grid resolution, and numerous variables. This limits their practicality for exhaustive scenario analyses or uncertainty quantification, which demands numerous iterations across a broad range of inputs.

In contrast, the model presented here approximates the behaviour of traditional models with far less computational intensity. On standard consumer hardware, it generates a 65-year climate projection in about 20 seconds, compared to a week required by supercomputers for the same task Acosta et al. (2024). This makes the surrogate model approximately 30,000 times faster, even without factoring in the hardware disparity. This efficiency enables the quantification of uncertainties in the atmospheric chemistry of H_2. The results of such uncertainty analysis, including the Monte Carlo simulation, are presented in section 5.1.

First, fugitive emissions from aviation are estimated. The resulting curves are a function of the adoption curve, aviation fuel consumption, and leakage assumptions used as inputs to the Monte Carlo simulations. Figure 5.1 presents the resulting temporal evolution of hydrogen emissions between 2025 and 2100.

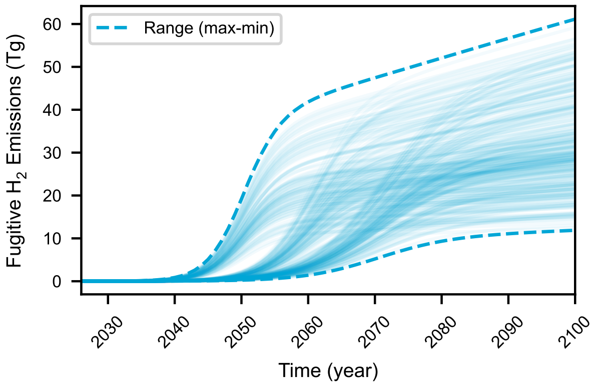
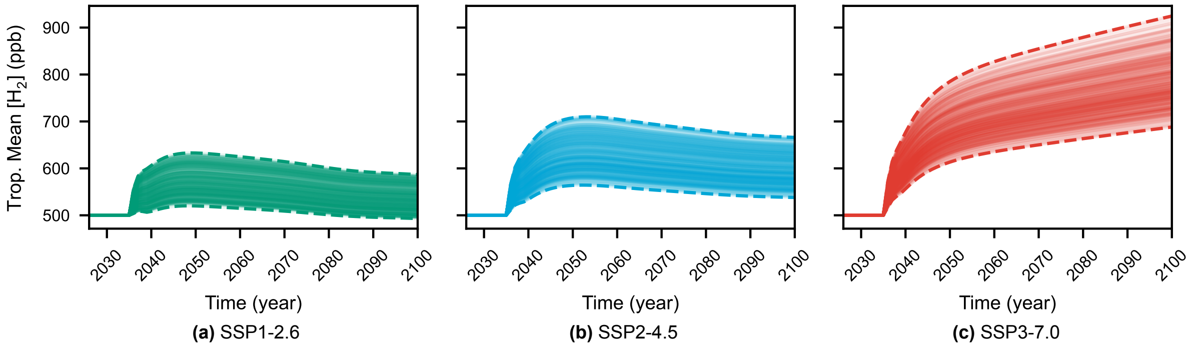


Figure 5.1: Fugitive H_2 emissions between 2025 and 2100 projected using parameters from the Monte Carlo simulation. Each shaded line indicates a single simulation, with the dashed lines indicating the maximum and minimum scenarios.

The influence of the adoption curve assumption is evident, with fugitive emissions increasing in three separate bands. Near the end of the simulation, the adoption curves have reached saturation. At this point, the spread in results is only a function of the leakage assumptions and the aircraft fuel consumption scenario. The values used in this Monte Carlo simulation, described in table 3.4, result in a significant spread of approximately 50 Tg of hydrogen in the year 2100.

The box model uses these additional hydrogen emissions as input to derive the temporal evolution of tropospheric mean H_2 and OH. At this point, background emissions for the various SSP scenarios start to influence the model, necessitating that scenario-specific results are separated. Scenario SSP5-8.5 is omitted from this analysis due to its similarity to SSP3-7.0. Results for the temporal evolution of these species are visible in figure 5.2 and figure 5.3, respectively.



(a) SSP1-2.6 **(b)** SSP2-4.5 **(c)** SSP3-7.0

Figure 5.2: Temporal evolution of tropospheric mean H_2 mixing ratios due to H_2 emissions, with scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)** or SSP3-7.0 **(c)** as background climate state. Each shaded line represents a projection from the Monte Carlo simulation, and the dashed lines represent the minimum and maximum values.

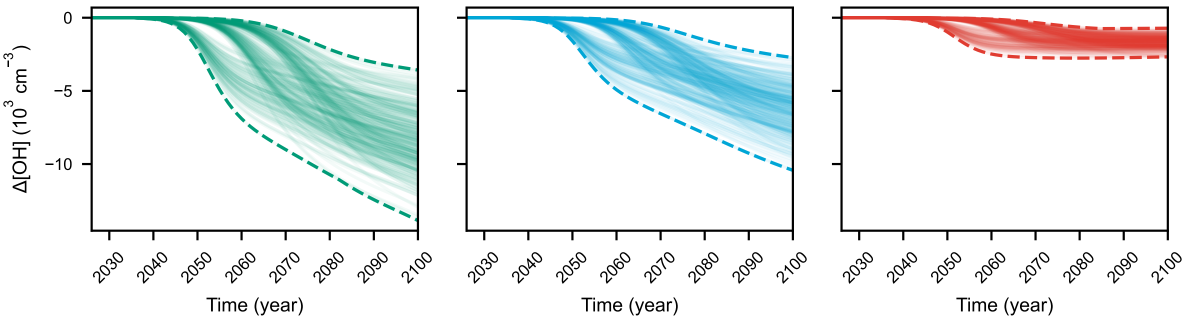
Hydrogen is not explicitly output by the CMIP6 scenarios; therefore, its emissions are initialised according to results from Bertagni et al. (2022) and constant in the years before 2035. After this year, fugitive emissions commence, showing a general increase in H_2 levels across the three scenarios. At this point, other variables parametrized by the Monte Carlo simulation are also changed. This includes the H_2 deposition lifetime k_d, the additional CH_4 sink lifetime k_s, and the CH_4 feedback yield α. This variable change (in particular a change in k_d) results in a sudden spread in projected H_2 developing between 2035 and 2040. This is not representative of reality. Nonetheless, the output of the box model is only the difference in OH between a perturbed and baseline projection. Therefore, this discrepancy cancels out, allowing for it to be ignored.

There are notable differences between the projections of the three scenarios. SSP1-2.6, for the same fugitive H_2 emissions, shows a substantially smaller increase in H_2 compared to SSP3-7.0. Furthermore, the spread of the predictions seems to increase in the same order, with SSP3-7.0 exhibiting the widest range of outcomes, while SSP1-2.6 remains relatively constrained. The differences in emissions considered in each scenario must account for these phenomena, as they are the only factors distinguishing the simulations at this stage. A full explanation of these phenomena and their implications is provided in section 5.3.

H_2 competes with CH_4 and CO for OH. Under nominal conditions, a greater increase in H_2 (SSP3-7.0) is expected to result in greater decreases in OH. However, figure 5.3 shows different behaviour. This figure plots the difference between a control simulation and a perturbation simulation. Therefore, the figures show the change in OH that can be directly attributed to increases in H_2 emissions.

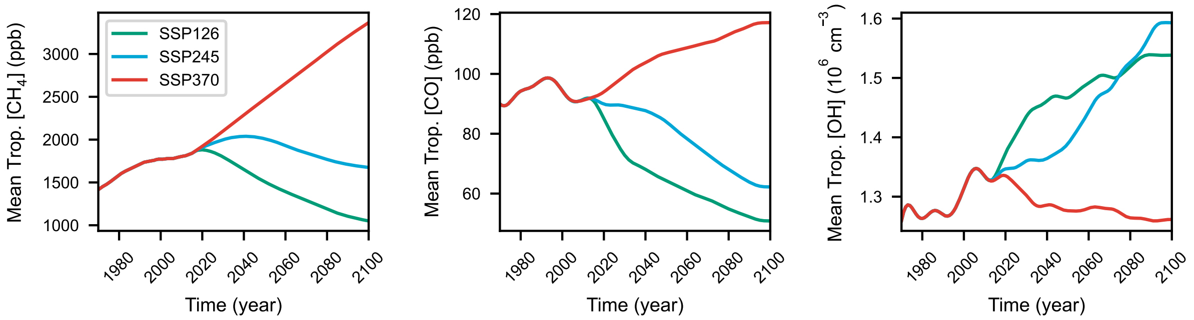
SSP1-2.6 shows a large reduction in OH, contrasting the relatively minor increase in H_2 visible in figure 5.2. Conversely, the large increase in hydrogen emissions projected for the SSP3-7.0 scenario does not lead to OH reduction of similar magnitude. This can be explained by considering the background concentrations of CH_4, CO and OH. The temporal evolutions of these species as projected by CMIP6 simulations are provided in figure 5.4, for each scenario.

This figure shows that increases in CH_4 and CO emissions reduce - or limit the increase of - OH. As a result, the oxidative capacity of the troposphere is greatly reduced under SSP3-7.0 when compared to SSP1-2.6 and SSP2-4.5. When excess hydrogen is introduced to the troposphere, less OH is available,



(a) SSP1-2.6 **(b)** SSP2-4.5 **(c)** SSP3-7.0

Figure 5.3: Temporal perturbation of tropospheric mean OH mixing ratios due to H_2 mixing ratio increases, with scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)** or SSP3-7.0 **(c)** as background climate state. Each shaded line represents a projection from the Monte Carlo simulation, and the dashed lines represent the minimum and maximum values.



(a) CH_4 **(b)** CO **(c)** OH

Figure 5.4: Temporal evolution of tropospheric mean H_2 mixing ratios due to H_2 emissions, with scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)** or SSP3-7.0 **(c)** as background climate states. Each shaded line represents a projection from the Monte Carlo simulation, and the dashed lines represent the minimum and maximum values.

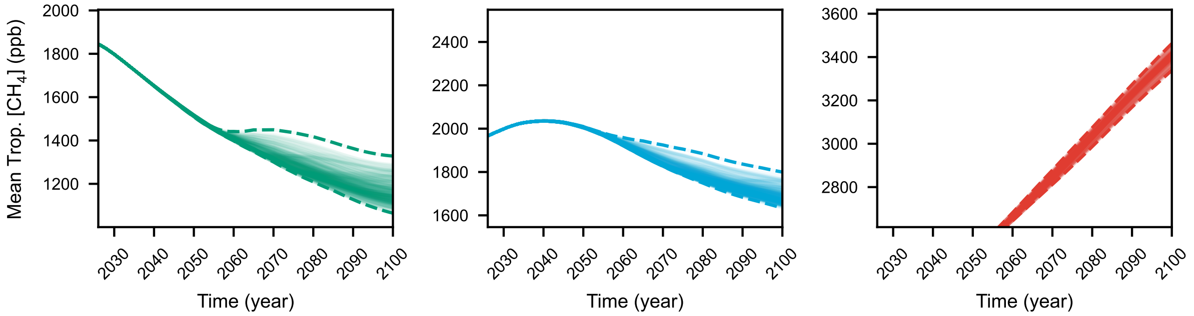
leading to higher H_2 concentrations. This can also be approached from the perspective of the effect of H_2 on the lifetime of CH_4. At elevated background CH_4 levels, additional H_2 will interact less with OH, thereby reducing its effect on CH_4 lifetime. This topic is discussed elaborately in section 5.3.

Figure 5.5 illustrates how the reduction of OH impacts CH_4. Here, the temporal evolution of mean tropospheric CH_4 mixing ratios is plotted for the various scenarios. It is important to note that while the axes in these plots are not shared, the scale of the axes is the same.

These plots demonstrate how the decreasing reduction in OH propagates into CH_4 projections. Where SSP1-2.6 shows a mean CH_4 increase of 50 ppb, this value is halved for SSP2-4.5 and almost negligible for SSP3-7.0. Notably, while SSP2-4.5 and SSP1-2.6 show similar reductions in OH, this does not seem to translate into equal increases in CH_4. This could be related to the temporal evolution of OH seen in figure 5.4, where OH levels only increase towards around 2060, in contrast to SSP1-2.6, which shows a steeper initial increase after which it levels out.

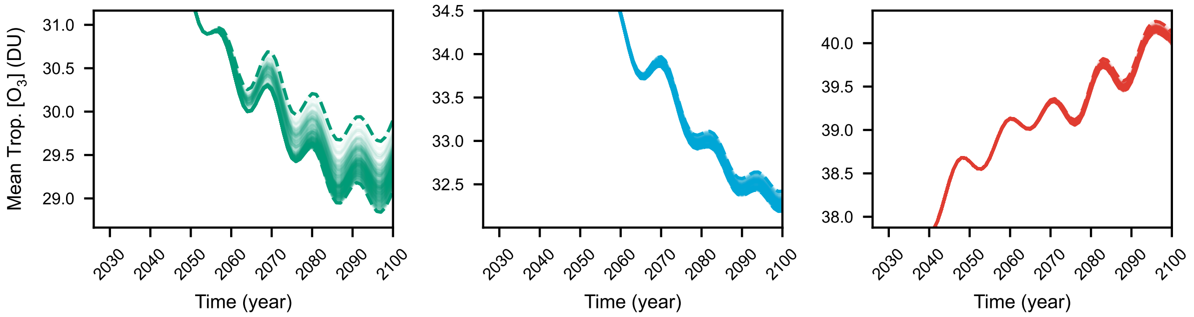
For tropospheric O_3, similar plots can be generated. These plots are presented in figure 5.6. Tropospheric O_3 is mostly formed through its interactions with CH_4. Thus, the perturbation visible in figure 5.5 is predicted to propagate similarly into the O_3 temporal evolution.

Figure 5.7 shows the resulting effective radiative forcings resulting from the Monte Carlo Simulation for the three considered SSP scenarios. The slope of the fits through these points provides information about the radiative efficiency of hydrogen during a certain background climate state—a measure of how effectively hydrogen influences radiative forcing under the given atmospheric and climatic conditions. This slope reflects the sensitivity of the radiative forcing to changes in hydrogen emissions, accounting for interactions with methane, ozone, and water vapour. By analyzing these fits, insights can be gained into the variability of hydrogen's climate impact across different scenarios and time periods. The results emphasize the role of background climate in modulating the radiative effects of hydrogen leakage, highlighting its potential implications as an alternative fuel in future climate scenarios.



(a) SSP1-2.6 **(b)** SSP2-4.5 **(c)** SSP3-7.0

Figure 5.5: Temporal evolution of tropospheric mean CH_4 mixing ratios due to OH perturbations, with scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)** or SSP3-7.0 **(c)** as background climate state. Each shaded line represents a projection from the Monte Carlo simulation, and the dashed lines represent the minimum and maximum values.



(a) SSP1-2.6 **(b)** SSP2-4.5 **(c)** SSP3-7.0

Figure 5.6: Temporal evolution of tropospheric mean O_3 mixing ratios due to CH_4 perturbations, with scenario SSP1-2.6 **(a)**, SSP2-4.5 **(b)** or SSP3-7.0 **(c)** as background climate state. Each shaded line represents a projection from the Monte Carlo simulation, and the dashed lines represent the minimum and maximum values.

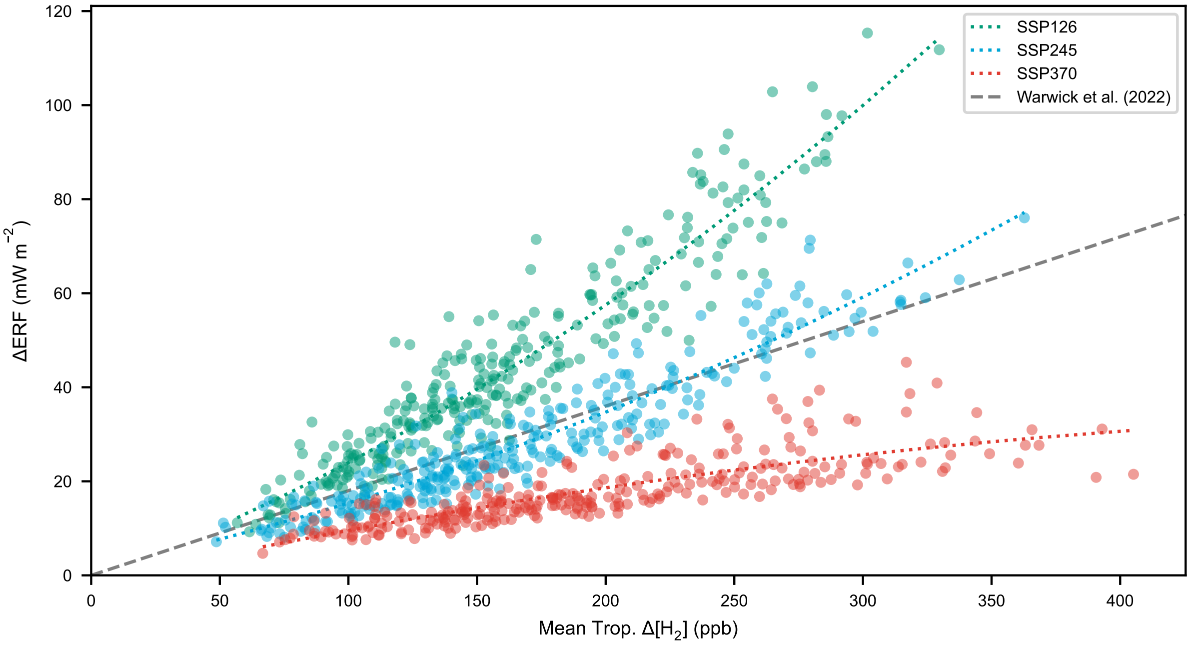


Figure 5.7: Plot of the H_2-induced change in effective radiative forcing against the increase in mean tropospheric H_2 mixing ratios - also referred to as radiative efficiency of H_2. Each dot represents a result from the Monte Carlo simulation. The results are color-coded per scenario, green represents SSP1-2.6, blue SSP2-4.5 and red SSP3-7.0. The results for each scenario are fitted using a 2nd-order polynomial. The grey dashed line represents the radiative efficiency of H_2 derived by Warwick et al. (2022).

The y-axis represents the change in superposition of CH_4, tropospheric O_3 and SWV radiative forcing evaluated in 2100. The x-axis shows the change in mean tropospheric H_2 directly attributable to fugitive H_2 emissions. It is computed by taking the difference between the perturbed and baseline scenarios and, therefore, does not account for increases in the lifetime of H_2 due to additional CH_4 emissions. This allows for a direct comparison of the radiative efficiency of H_2 with values found in the literature.

The results for each scenario reveal distinct trend separations, highlighting that the radiative efficiency of H_2 is highly sensitive to the atmospheric concentrations of other species. These trend slopes can be analyzed and compared against values reported in the literature. Reference values derived by Paulot et al. (2021) and Warwick et al. (2022) are plotted on the same graph, representing linear slopes passing through the origin. Notably, SSP1-2.6 exhibits a much steeper slope, whereas SSP2-4.5 and SSP3-7.0 demonstrate a more moderate RF increase with rising H_2 levels.

In figure 5.8, the contribution of each GHG indirectly affected by H_2 emissions to the total change in radiative forcing is plotted. Each coloured bar represents the mean of the individual contributions of a species, and the grey bar represents the combined radiative forcing. The spread of the simulations is visualised using whiskers representing one standard deviation.

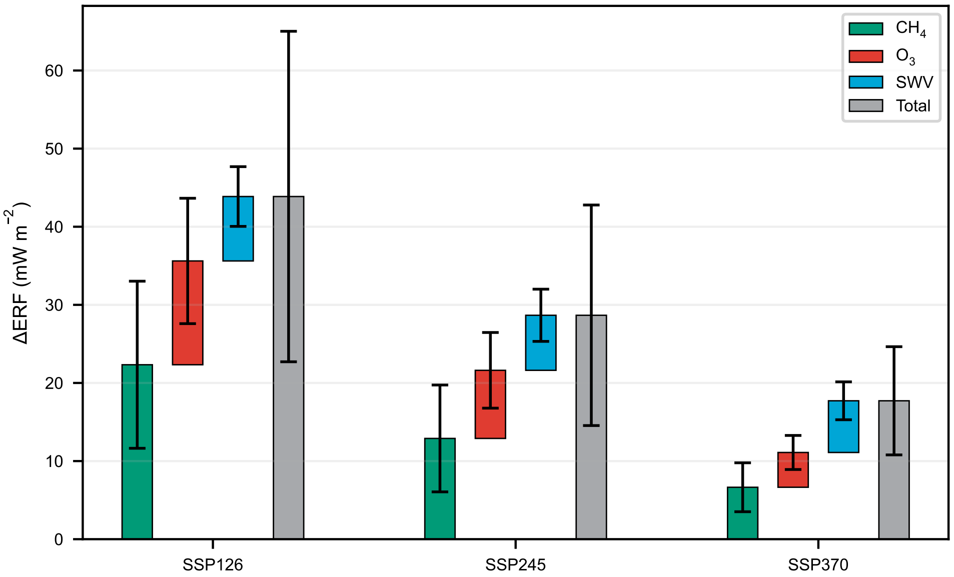


Figure 5.8: Bar plots indicating the mean contribution of CH_4, O_3 and SWV to the total radiative forcing for SSP1-2.6, SSP2-4.5, and SSP3-7.0. The black whiskers represent the uncertainty, expressed as ±1 standard deviation (SD).

Table 5.1: Summary of the mean contributions of CH_4, O_3 and SWV to the H_2-induced radiative forcing for SSP1-2.6, SSP2-4.5 and SSP3-7.0.

Scenario	**SSP1-2.6**		**SSP2-4.5**		**SSP3-7.0**	
ERF_{CH_4} (mW m^{-2}) (%)	22.3	(50.9)	12.9	(45.0)	6.7	(26.5)
ERF_{O_3} (mW m^{-2}) (%)	37.5	(30.3)	8.7	(30.4)	4.5	(25.1)
ERF_{H_2O} (mW m^{-2}) (%)	8.3	(18.8)	7.1	(24.6)	6.6	(37.3)
ERF_{tot} (mW m^{-2}) (%)	43.9	(100.0)	28.7	(100.0)	17.7	(100.0)

The relative contribution of each species to the total radiative forcing is not constant. The contribution of CH_4 and O_3 declines towards more pessimistic climate projections. Contrarily, the contribution of SWV is seen to increase relatively but also in absolute terms. This is a direct consequence of the method used to calculate SWV described in section 3.1. This calculation significantly impacts the results, necessitating an investigation of its implications. These implications are discussed in section 5.3.

The individual contribution of parameter uncertainty to the indirect radiative forcing of H_2 is determined by configuring the Monte Carlo simulation on a case-by-case basis. A set of predictions is made, allowing the Monte Carlo simulation only to sample one parameter. The resulting uncertainty in radiative forcing provides insights into the significance of a certain parameter on the radiative effect of H_2.

The results from this analysis for the SSP1-2.6 and SSP3-7.0 scenarios are plotted in figure 5.9. The x-axis represents the relative change in radiative forcing resulting from the spread in input parameters. The y-axis contains the results in terms of a box plot for each parameter. The width of the box plots, thus, represents the uncertainty in RF as a function of the underlying parameter. The parameters are sorted by their influence.

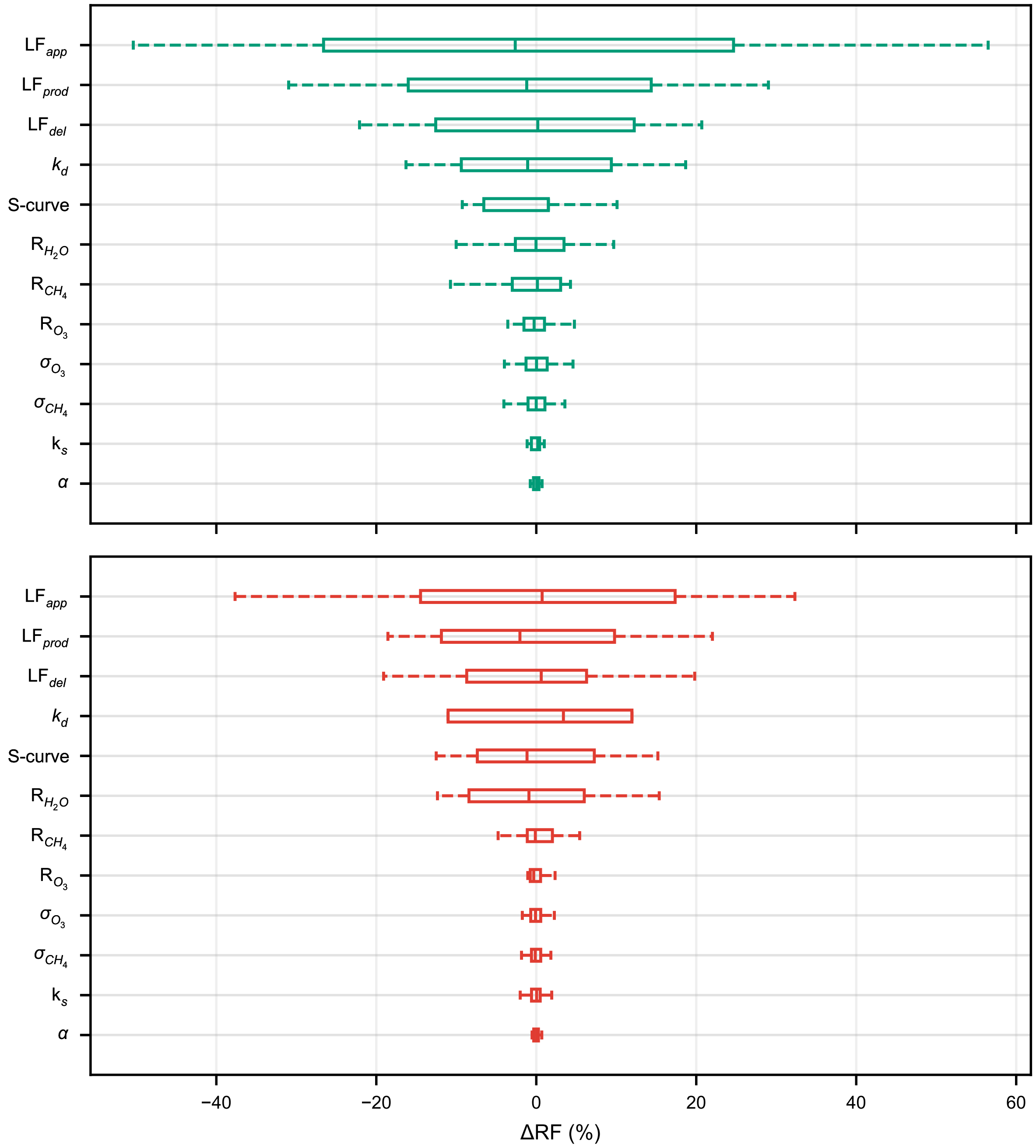


Figure 5.9: Box plot illustrating the impact of input parameter uncertainty, as detailed in table 3.4, for SSP1-2.6 (green) and SSP3-7.0 (red). The interquartile range (IQR) is divided by the median, with whiskers extending to 1.5 times the IQR.

The assumed leakage factors contribute majorly to the uncertainty in radiative forcing. Depending on the individual assumed leakage rate, the indirect radiative forcing of H_2 is seen to vary by $\pm$50%. This effect is amplified in the Monte Carlo simulation, as the total emissions calculated with these emission factors result from their multiplicative interaction. In other words, when a high application leakage factor is combined with high values for leakage during production and delivery, the resulting uncertainty will be significantly magnified, leading to a wider range of potential emissions outcomes.

The box model parameters for methane feedback and additional stratospheric sinks, α and k_s, have a minor effect on the overall uncertainty. However, the deposition lifetime of H_2 causes an uncertainty of about $\pm$ 20% in the final radiative forcing. Errors attributable to the surrogate models, σ_{CH_4}, and σ_{O_3} are also relatively minor compared to the uncertainty resulting from the box-model and hydrogen emissions scenario assumptions. When considering the uncertainties in radiative forcings, a significant increase in uncertainty for SWV between the two scenarios becomes apparent. This is a product of the increased relative contribution of SWV to the total radiative forcing shown in table 5.1.

5.2. Case Studies

While the Monte Carlo simulation offers a broad understanding of the potential impacts and uncertainties associated with fugitive hydrogen emissions, it provides limited insight into the nuanced dynamics of specific scenarios. To bridge this gap, it is beneficial to investigate case studies representing distinct narratives within the broader context of hydrogen adoption in aviation. Examining individual cases allows for a more granular analysis of how specific combinations of factors—such as adoption rates, leakage assumptions, and background atmospheric conditions—influence the outcomes.

This section presents five distinct case studies that explore different narratives for hydrogen emissions' impact on atmospheric composition and climate forcing. Each narrative embodies a unique set of assumptions and parameters, reflecting possible real-world scenarios. A specific set of input parameters relating to the hydrogen emission model will be set in each case. This includes the assumptions on the leakage factors, hydrogen adoption curve and the SSP scenario used to model background concentrations of relevant atmospheric gases. All scenarios assume that aviation will completely adopt H_2 powered aircraft by 2100. The narratives are outlined as follows:

Case 1: Sustainable Development with Minimal Hydrogen Leakage

In this scenario, the world follows the SSP1-2.6 pathway, which represents a future where sustainable development is the primary focus. Global cooperation is strong, leading to significant investments in renewable energy, technological innovation, and environmental protection. The international community successfully implements policies to limit global warming to well below 2°C above pre-industrial levels, in line with the Paris Agreement.

Hydrogen adoption in aviation is rapid and widespread. By 2050, 50% of the global aviation fleet is powered by hydrogen, reaching full adoption by 2065. The hydrogen used is primarily produced from renewable energy sources (green hydrogen), ensuring minimal carbon emissions during production. The minimal values for hydrogen leakage reflect highly efficient technologies and stringent regulations that minimize fugitive hydrogen emissions throughout the supply chain—from production and storage to transportation and refuelling. Advanced materials and monitoring systems are employed to detect and prevent leaks, ensuring that the environmental benefits of hydrogen as a clean fuel are fully realized.

Case 2: Sustainable Development with Significant Hydrogen Leakage

This scenario also operates under the SSP1-2.6 pathway of sustainable development, emphasizing global efforts to mitigate climate change and promote environmental sustainability. Hydrogen adoption in aviation is as rapid as in Case 1, reaching 50% by 2050 and 100% by 2065. The aviation industry is transitioning aggressively towards hydrogen to reduce carbon emissions.

However, in this case, the maximal values for hydrogen leakage are assumed. Despite the commitment to sustainability, the rapid scaling up of hydrogen infrastructure leads to challenges in controlling fugitive emissions. Technological limitations, insufficient regulations, or supply chain bottlenecks result in higher hydrogen leakage during production, storage, and distribution. The significant hydrogen leakage has notable implications for atmospheric chemistry. This scenario highlights the importance of not only adopting clean energy technologies but also ensuring that their implementation does not introduce new environmental challenges. It underscores the need for robust technological solutions and regulatory frameworks to manage hydrogen leakage effectively.

Case 3: Hybrid Development with Moderate Hydrogen Leakage

Under the SSP2-4.5 pathway, the world takes a ”middle of the road” approach to development. Economic growth and technological progress continue at a moderate pace, with some efforts towards sustainability but without the aggressive policies seen in SSP1. Population growth is moderate, and global disparities persist, leading to uneven progress in different regions.

Hydrogen adoption in aviation is gradual. The sector will reach 50% hydrogen-powered aircraft by 2060 and full adoption by 2085. This moderate pace reflects a balance between the desire to reduce emissions and the economic and technological challenges of overhauling existing infrastructure. The mean values for hydrogen leakage indicate that while some measures are in place to control emissions, they are not as stringent or effective as in the minimal leakage scenarios. Technological advancements and regulations lag behind the rapid adoption needed to minimize leakage fully. This scenario represents a pragmatic progression where both conventional fossil fuels and alternative energy sources coexist for an extended period.

Case 4: Fossil-Fueled Development with Minimal Hydrogen Leakage

In this scenario, emissions follow the SSP3-7.0 pathway, characterized by regional rivalry and a focus on energy security over environmental concerns. Economic growth is uneven, and international cooperation is limited. High population growth and persistent inequalities lead to increased demand for energy, predominantly met by fossil fuels.

Despite these challenges, hydrogen adoption in aviation occurs but at a slow pace, reaching 50% by 2070 and full adoption by 2100. The slow adoption reflects technological barriers, lack of investment, and policy priorities favouring fossil fuels. The limited hydrogen leakage minimizes the indirect climate impacts associated with hydrogen use, such as methane lifetime extension and ozone formation. However, the overall benefits are constrained by the slow adoption rate and the continued dominance of fossil fuels, resulting in high CO_2 emissions from the aviation sector and other industries.

Case 5: Fossil-Fueled Development with Significant Hydrogen Leakage

Also set within the SSP3-7.0 pathway, this scenario represents a world where high challenges to mitigation persist, and regional conflicts hinder global cooperation on climate action. Economic priorities focus on immediate concerns over long-term sustainability, and technological innovation is unevenly distributed.

Hydrogen adoption in aviation is slow, mirroring Case 4, with 50% adoption by 2070 and full adoption by 2100. However, this scenario assumes maximal values for hydrogen leakage. The combination of slow adoption and significant leakage reflects inadequate technological development, poor infrastructure, and weak regulatory frameworks. Significant hydrogen leakage exacerbates indirect climate impacts. Elevated atmospheric hydrogen levels reduce OH concentrations, extending the lifetime of methane and enhancing its greenhouse effect. Additionally, alterations in ozone chemistry may contribute to further radiative forcing changes.

This scenario highlights a worst-case outcome where the slow transition to cleaner energy sources is compounded by poor implementation practices, leading to both direct and indirect negative environmental impacts. It underscores the critical importance of not only adopting alternative fuels but also investing in the necessary technology and regulations to manage them effectively.

Figure 5.10 shows the resulting radiative forcing prediction for each case. In each simulation, the parameters not set according to the method described here are randomly sampled in the same way described in table 3.4.

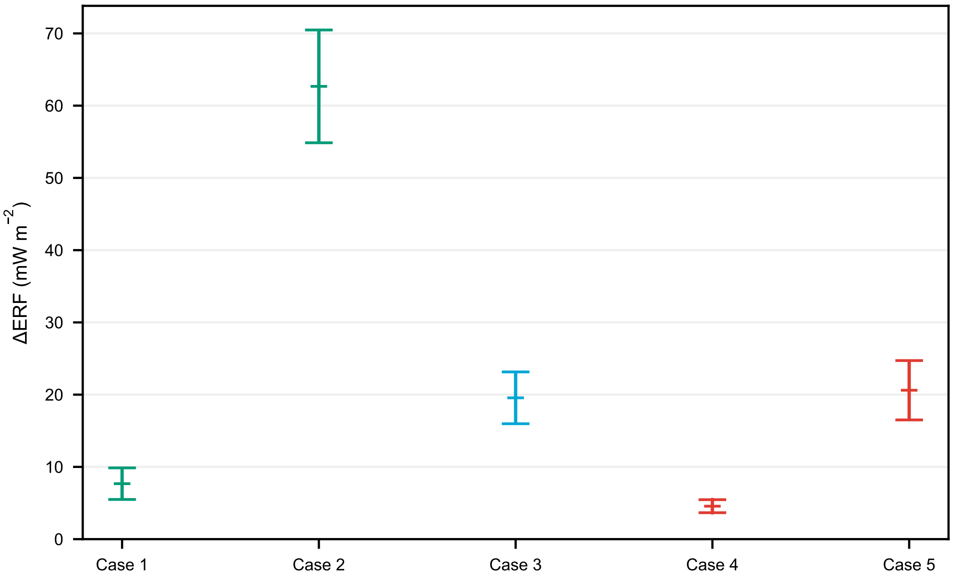


Figure 5.10: Mean change in ERF caused by fugitive H_2 emissions across five cases, each representing different leakage rates and background climate states. The results are colour-coded based on the background climate state: green for SSP1-2.6, blue for SSP2-4.5, and red for SSP3-7.0. Error bars represent ±1 SD.

Table 5.2: Summary of changes in ERF due to fugitive hydrogen emissions for various case studies describing different leakage rates and background climate states.

Scenario	ERF (mW m^{-1})	SD (mW m^{-1})
Case 1	7.7	± 2.2
Case 2	62.7	± 7.8
Case 3	19.6	± 3.6
Case 4	4.6	± 0.9
Case 5	20.6	± 4.1

The scenarios simulating low hydrogen leakage, case 1 and case 4, produce the lowest radiative forcings. At high hydrogen leakage rates, there is a significant discrepancy between cases where optimistic or pessimistic projections on climate change mitigation are considered. Case 2, which uses SSP1-2.6 for background concentrations, inhibits the largest radiative forcing. The disparity between case 2 and the other cases indicates an exponential sensitivity to hydrogen emissions under scenarios with strong climate mitigation efforts. In case 2, where SSP1-2.6 background concentrations are applied alongside high hydrogen leakage rates, the low baseline levels of greenhouse gases such as CH_4 and CO in SSP1-2.6 amplify the indirect radiative effects of hydrogen emissions.

5.3. Discussion

This section discusses the interpretation of the results from the scenario analysis, as well as the effects of the assumptions made in modelling and possible implications for future research. It will first examine the key outcomes and their significance, assess the validity and impact of the underlying assumptions, and outline potential directions for subsequent studies to build upon these findings.

Interpretation of Results

Results from the Monte Carlo analysis show how various projections for future fugitive H_2 emissions from aviation sources will impact atmospheric composition and radiative forcing. Previous studies investigating this subject - or rather fugitive H_2 emissions in general - have not considered the potential implications of changing climate states on the (Sand et al., 2023; Warwick et al., 2022; Derwent, 2023; Hauglustaine et al., 2022; Paulot et al., 2021). The surrogate model developed in this thesis allows for previously infeasible rapid assessment of the indirect climate impact of H_2, together with an exhaustive analysis of the uncertainties of the atmospheric chemistry of H_2.

Analysis of the temporal evolution of H_2 and OH illustrate how various climate states drastically influence the climate impact of H_2. Under high CH_4, the box model predicts a fiercer competition for OH in the troposphere. As a result, OH levels remain low under these conditions. The lower availability of OH results in a lower reaction rate with H_2. Consequently, scenario projections with higher methane emissions (SSP3-7.0) project higher levels of H_2 in the troposphere, while the opposite is true for scenario projections with lower CH_4 emissions. Hydrogen's indirect climate effect is mainly governed by its perturbation of OH levels and subsequent chain reaction, increasing the lifetime of CH_4 and production of O_3. Hence, the climate effect of hydrogen emissions is modelled to be greatly reduced under pessimistic climate scenarios and exacerbated under optimistic scenarios.

These findings are consistent with existing literature. Although the study by Warwick et al. (2022) did not model the effects of H_2 across various climate states as comprehensively as this thesis, it did examine the implications of reducing O_3 precursors. The study found that decreasing the CH_4 lower boundary condition by approximately 10% increases the rate at which the lifetime of CH_4 extends with each incremental increase in the surface mixing ratio of H_2. In other words, the climate impact of H_2 is amplified under lower CH_4 emissions. The alignment of this study's results with the simulations from the conventional climate model strengthens the argument that the surrogate model is properly validated.

The relative contributions of CH_4, O_3 and SWV to the indirect total radiative forcing of H_2 are given by table 5.1. The relative contribution of CH_4 and O_3 decreases when for SSP2-4.5 and SSP3-7.0. Conversely, the absolute - and thus relative - contribution of SWV increases. This absolute increase is related to the method used to model SWV. Unlike CH_4 and O_3, the radiative forcing of SWV is only partially modelled by the surrogate model. The argumentation for this decision is provided in section 3.1. This leads to the direct correlation of SWV with increases of tropospheric H_2 levels, which are found to increase for the more pessimistic climate scenarios.

The radiative efficiency used to determine the radiative forcing of H_2-induced SWV is taken directly from Warwick et al. (2022). Simulations in this study found that approximately 25% of SWV radiative forcing is through H_2's interaction with CH_4, with the remaining 75% resulting from SWV formed by the oxidation of H_2 in the troposphere transported vertically to the stratosphere, or direct oxidation of H_2 transported to the stratosphere. This results in a discrepancy in the context of this research, where the interaction between H_2 and CH_4 is drastically altered. Reduced oxidation of H_2 due to lower OH levels and subsequent smaller perturbations of CH_4 should lead to lower SWV levels than modelled in this research. As such, the model is believed to overestimate the radiative forcing of SWV and, thus, the overall radiative forcing for the pessimistic SSP3-7.0 scenarios. Following the same logic, it can be argued that the model underestimates the contribution of SWV for the more optimistic SSP scenarios. Future research should improve the accuracy of SWV modelling its alteration in the stratosphere due to H_2 emissions.

The total radiative forcing induced by H_2 is plotted against the change in the mean tropospheric H_2 mixing ratio to obtain the normalized radiative forcing. The simulation trends for various SSP scenarios, illustrated in figure 5.7, demonstrate how hydrogen's climate effect depends on the prevailing climate state. These trends can be compared to the results from experiments by Warwick et al. (2022) and Paulot et al. (2021). SSP1-2.6 shows a higher sensitivity in radiative forcing, while SSP2-4.5 and SSP3-7.0 exhibit lower sensitivities. This variation logically results from the changing sensitivity of CH_4 under these scenarios, though the outcome may initially seem counterintuitive. Simply put, this indicates that hydrogen emissions have a more detrimental effect on the climate in future emission pathways that are more sustainable. Conversely, in more polluting pathways, the impact of hydrogen emissions is significantly reduced.

The Monte Carlo simulation is employed to quantify the effect of uncertainty in the input parameters used in this model. The resulting uncertainty is extensive, with SSP1-2.6 indicating a spread in between approximately 10 and 100 mW m^{-2} in H_2-induced climate forcing. The parameter sensitivity analysis displayed in figure 5.9 indicates this is mostly a result of the leakage assumption used in the hydrogen emission model. The spread in the leakage assumption solely for aircraft alone results in an uncertainty of approximately $\pm 50\%$. In the Monte Carlo simulations, these uncertainties are compounded, which can lead to even greater uncertainty as a result of the leakage assumption. The leakage assumptions, together with the assumption dictating the adoption of hydrogen, remain predictions on technology in its early stages of development. Therefore, only time will reveal whether these assumptions prove accurate as hydrogen technology evolves and its real-world applications become more established.

The second most significant uncertainty arises from the variability in the H_2 deposition lifetime (k_d). As the primary sink of atmospheric H_2, uncertainties in k_d significantly influence the estimated climate impact of hydrogen emissions. In this study, k_d is uniformly sampled from a range of deposition lifetimes derived from the literature, following the methodology of Warwick et al. (2022). However, recent findings by Paulot et al. (2021) suggest that H_2 deposition rates may increase due to climate change, potentially acting as a damping mechanism on H_2-induced climate effects under more polluting scenarios. Consequently, further research is necessary to enhance the representation of the H_2 soil sink in box models, reducing its contribution to overall uncertainty.

The impact of other model uncertainties is relatively insignificant. Uncertainties arising from the surrogate models derived in chapter 4 play minor role in the overall uncertainty. Therefore, improvements that could be obtained by optimising the machine learning models will not likely yield useful results if uncertainties in other areas are not improved. The methane feedback yield and additional methane sink have a minor impact, indicating that the use of the box model devised by Bertagni et al. (2022) shows no major improvements over the use of the simpler model proposed by Warwick et al. (2022). The uncertainties in radiative forcing have a more significant effect, especially RF_{H_2O}. These uncertainties are well-researched and directly taken from the AR6 report by IPCC (2021b), and therefore, unlikely to be reduced significantly in the near future.

Brazzola et al. (2022) performed simulations to quantify the ERF of aviation for future climate projections. When considering zero-CO_2 fuels, this study finds a total aviation ERF of approximately 80 mW m^{-2} in 2100 for SSP1-2.6, and an ERF of approximately 300 mW m^{-2} for SSP3-7.0. Depending on the scenario, this would result in a relative contribution of fugitive H_2 emissions of around 35% for SSP1-2.6 or 5% for SSP3-7.0. This highlights the importance of limiting fugitive H_2 emissions in future aviation applications, especially as their contribution can significantly amplify the non-CO_2 effects of aviation on climate.

The case studies presented in this chapter can be used to obtain a more granular understanding of the potential climate impact of fugitive H_2 emissions. Leakage rates for the application of H_2 in aviation, delivery, and production estimated by Fan et al. (2022), several narratives are established. The most extreme narrative, using the highest estimates for H_2 leakage, the quickest adoption of H_2 in aviation, and assuming the most optimistic emission scenario, indicates an additional radiative forcing from H_2 fugitive emission of 62.7 mW m^{-2} (43%), with a standard deviation of ± 7.8 mW m^{-2}.

Given the sensitivity of the atmosphere to H_2 emissions, even low leakage rates can lead to disproportionate climate impacts by affecting the atmospheric composition of CH_4, O_3, and SWV. Therefore, mitigation strategies to minimize H_2 leakage—such as improved fuel storage and handling technologies—are critical to ensuring that the adoption of hydrogen-based fuels does not undermine their potential climate benefits.

Effect of Assumptions and Uncertainties

This model is built upon a series of assumptions and simplifications inherent to the modelling process, each of which introduces a degree of uncertainty that can influence the reliability and interpretation of the results. These assumptions pertain to various aspects, such as the representation of physical processes, the parameterization of complex interactions, and the boundary conditions or initial states of the system being modeled. Recognizing and critically assessing these underlying assumptions is essential for a comprehensive understanding of the model's outcomes and their limitations. In the following, these assumptions and their presumed or approximated effects are analysed.

1. **Aviation Fuel Consumption Projections:** This research assumes future aircraft fuel consumption according to projections by Grewe et al. (2021). These projections, however, are not harmonised with the SSP scenarios used in this study. The various socio-economic assumptions underlying the SSP scenarios could lead to differing fuel consumption trajectories, introducing additional uncertainty into the analysis. To enhance the robustness of the projections, future research should align fuel consumption estimates with the socio-economic pathways defined by the SSP frameworks.
2. **Computation of Equivalent Hydrogen Consumption:** Hydrogen emissions are derived by computing an equivalent H_2 consumption. In this calculation, H_2-powered aircraft are assumed to have a 10% increase in energy consumption due to estimated increases in wetted area (Sáez Ortuño et al., 2023). Hydrogen-powered aviation is in early stages of development, however, and its efficiency could vary significantly depending on factors such as aircraft design, hydrogen storage technology, propulsion system advancements, and the integration of hydrogen infrastructure within the aviation sector. Additionally, improvements in fuel cell or combustion technologies and the optimization of aerodynamic properties could further influence the overall efficiency of hydrogen-powered aircraft.
3. **Hydrogen Adoption:** Hydrogen adoption is modelled using an S-curve. While the effect of the rate of adoption is analysed, hydrogen adoption is always asssumed to reach 100%. This is likely unrealistic, as future aircraft may be powered by a wide range of sources, such as sustainable aviation fuels, electric sources, or conventional hydrocarbon fuels.
4. **Auxiliary Hydrogen Emissions:** To isolate the specific effect of H_2 emissions from aviation, this study does not consider fugitive emissions of H_2 from other industries or sources. Nonetheless, aviation is far from the only sector contributing to hydrogen emissions. Other industries, such as chemical manufacturing, energy production, and transportation, also release H_2, which could compound the overall climate impact. Additionally, these multiple sources of H_2 emissions may interact in complex ways, potentially influencing atmospheric chemistry and radiative forcing beyond the scope of this study. Future research should incorporate these additional sources to provide a more comprehensive assessment of hydrogen's role in climate dynamics and to better inform mitigation strategies across all relevant sectors.
5. **Box-model:** Ideally, a surrogate model can directly be trained on hydrogen emission data. Such data is not available for multiple SSP scenarios considered in this study, however, necessitating the use of a simple box model to derive the effect of H_2 on OH. Simulations by conventional climate models indicate the relation between the rate of change of OH per unit H_2 is linear and primarily dependent on background concentrations of CH_4, OH, CO and H_2 (Sand et al., 2023; Warwick et al., 2022). This allows for the use of a box model to derive mean tropospheric changes to OH attributable to H_2 emissions.

   The use of a box model, however, necessitates another assumption. Due to the extensive lifetime of H_2 and CH_4, these species are assumed to be well-mixed in the atmosphere. As such, using the tropospheric mean perturbation of OH as an input to the LSTM model which determines the effect on CH_4 is deemed valid. OH is also used as input for the ConvLSTM model. This model uses the spatial resolution of its inputs to determine the spatiotemporal evolution of ozone. The assumption that the spatial variation of an increase in OH due to H_2 emissions is negligible might give rise to errors in the ConvLSTM model. Nonetheless, validation performed in chapter 4, together with the minor impact of the model error derived in this chapter, indicate that the potential uncertainty arising from this assumption is relatively insignificant compared to the error arising from the leakage assumption or deposition lifetime of H_2.
6. **Simplification of Atmospheric Chemistry:** Similarly, the surrogate models used in this research are trained on a selection of atmospheric data, which simplifies its chemistry. The verification and validation performed in chapter 4 demonstrates that these assumptions have a minimal effect on the model's accuracy. Nonetheless, a more extensive analysis of the specific impact of including certain species in the training dataset could improve the accuracy of the model.
7. **Stratopsheric Ozone:** Aligning with findings from both Sand et al. (2023) and Warwick et al. (2022), the effect of H_2 emissions on stratospheric O_3 is taken to be negligible and therefore not considered in the model.

Implications on Future Research

This thesis demonstrates the potential of employing neural networks, specifically Long Short-Term Memory (LSTM) models, for the rapid evaluation of future climate scenarios. The LSTM model is capable of accurately assessing future climate conditions while significantly reducing the time and computational resources required by conventional climate models. Moreover, the model exhibits robust performance when trained on a general dataset, offering a distinct advantage over response models that necessitate custom simulations for each specific application. Recurrent neural networks like LSTM could be well suited for modelling non-linearities associated with atmospheric chemistry, enabling the capture of complex interactions between emissions, chemical transformations, and feedbacks within the climate system.

The scenario analysis conducted in this study reveals that the impact of H_2 emissions is highly dependent on future climate scenarios. This dependency necessitates further investigation using more detailed and accurate climate models. Results from such simulations should be employed to validate the model's accuracy. Additionally, data obtained from these comprehensive simulations can be utilized to train and iteratively refine the framework, thereby enhancing its overall precision and reliability.

More specifically, customized simulations are essential to comprehensively capture the climate effects of H_2, particularly those unique to the aviation sector. Aviation emissions are distinctive in that they occur at high altitudes, necessitating detailed simulations to accurately model the impact of H_2 emissions in the stratosphere, where their influence on radiative forcing may differ substantially. Employing bespoke simulations to train the dataset can eliminate various assumptions currently required by the model. For instance, tailored simulations can directly correlate the location of hydrogen emissions with their specific effects on radiative forcing, thereby preventing the need for multiple surrogate models or reliance on box models.

In the short term, however, the model should be improved by investigating two distinct areas. First, the most pressing challenge is accurately modelling hydrogen deposition, which is critical for a realistic simulation of its environmental impact. Addressing this requires refining the deposition mechanisms for variable surface and atmospheric conditions. Secondly, for this model specifically, implementing a stratospheric water vapour (SWV) modelling technique would enhance the accuracy of feedback processes, particularly those linked to ozone chemistry and radiative forcing.

Furthermore, the methodologies utilizing LSTM and ConvLSTM neural networks demonstrated in this thesis can be extended to other atmospheric species beyond hydrogen. These models are proficient in capturing the intricate, nonlinear interactions and feedback mechanisms inherent in atmospheric chemistry and physics. By applying these neural network architectures to other atmospheric constituents, it is possible to efficiently simulate their temporal evolution and spatial distribution under various emission scenarios. This approach not only accelerates the evaluation of their climatic impacts but also reduces the reliance on computationally intensive traditional climate models. Expanding the application of LSTM and ConvLSTM methodologies to a broader range of atmospheric species could advance climate research, providing rapid and flexible tools for assessing environmental policies and mitigation strategies.

6

Conclusions & Recommendations

This research evaluated the climate effect of fugitive hydrogen emissions from aviation sources. The findings underscore the complex interplay between hydrogen leakage, atmospheric chemistry, and future climate scenarios. Based on uncertainty analysis performed through a Monte Carlo simulation, it can be concluded that the climate effect of H_2 emissions strongly depends on future climate pathways and the associated atmospheric conditions.

The surrogate model developed in this work reduced the computational time required for forecasting atmospheric chemistry by a factor of 30000, allowing for the analysis of underlying uncertainties using Monte Carlo simulations. The developed model frameworks, leveraging Long Short-Term Memory (LSTM) and Convolutional LSTM (ConvLSTM) neural network architectures, were found to be capable of generalizing the atmospheric chemistry of methane and tropospheric ozone within an error margin of 10%. This level of accuracy demonstrates the potential of machine learning techniques in climate modelling, particularly in scenarios where traditional models are computationally prohibitive. Furthermore, the resulting uncertainty analysis indicates that the uncertainty inherent to the surrogate models is relatively minor compared to the uncertainties associated with the soil deposition of hydrogen and future predictions on leakage rates and adoption of hydrogen. This suggests that surrogate models are viable tools for preliminary assessments and can be instrumental in scenario analysis and policy development.

In particular, this study finds that besides the assumptions on aircraft leakage rates, the uncertainty in the atmospheric lifetime of hydrogen due to soil deposition has a significant effect on hydrogen's indirect impact on methane concentrations and, consequently, on radiative forcing. Variations in soil deposition rates can alter hydrogen's residence time in the atmosphere, influencing the extent to which it interacts with other greenhouse gases, such as methane and ozone. This highlights the need for further research to refine soil deposition parameters, as they are critical to accurately modelling hydrogen's overall climate impact.

Under SSP1-2.6, a climate scenario simulating effective climate change mitigation, the surrogate model developed in this work projects a H_2-induced radiative forcing of 43.9 ± 21.2 mW m^{-2} ($\pm 1\sigma$). This substantial value indicates that in scenarios where efforts are made to reduce greenhouse gas emissions, the relative impact of hydrogen leakage becomes more pronounced. Conversely, this value is lower when more pessimistic climate scenarios are considered, where the oxidative capacity of the troposphere is reduced due to higher concentrations of methane and other pollutants. For a climate scenario modelling conservative climate change mitigation and moderate leakage rates, a value of 17.7 ± 6.9 mW m^{-2} ($\pm 1\sigma$) is determined. When placed in the context of future projections of the total effective radiative forcing (ERF) from aviation, it becomes apparent that H_2 emissions represent a non-negligible contribution that could offset a substantial part of the anticipated climate benefits of transitioning to hydrogen fuel. Based on future aviation ERF estimates from Brazzola et al. (2022), fugitive hydrogen emissions are projected to contribute 5% to 35% of the ERF associated with zero-CO_2 aviation by the end of the century. This substantial share indicates that, without stringent measures to control hydrogen leakage, the anticipated reduction in the aviation sector's climate impact from adopting hydrogen fuels might not materialize as expected. In the most pessimistic scenario explored in this study—considering the highest leakage rates under SSP1-2.6—this contribution could increase further to 43%.

When lower leakage rates are assumed, the contribution of fugitive hydrogen emissions to aviation's ERF is relatively minor. For SSP1-2.6 a value of 7.7 ± 2.2 mW m^{-2} ($\pm 1\sigma$) is found, which would constitute to 8.8% of the total ERF. As expected, this value is further reduced under SSP3-7.0, resulting in a value of 4.6 ± 0.9 mW m^{-2} ($\pm 1\sigma$), which would be 1.5% of the total ERF. These results suggest that mitigating hydrogen leakage in H_2 production, transport, and storage is critical to ensuring that the adoption of hydrogen as an aviation fuel achieves its intended climate benefits.

The large discrepancy between the increment in radiative forcing between different background climate states warrants the investigation of hydrogen's climate effect using more accurate climate models. Current models may not fully capture the complex feedback mechanisms and regional variations that influence hydrogen's atmospheric behaviour. Results from advanced, high-resolution climate models should be used to validate and calibrate the performance of the surrogate models employed in this work. Furthermore, leveraging insights from these sophisticated models can help refine the surrogate models, enhancing their reliability across diverse scenarios. Such validation is crucial, as it will allow for improved predictions of radiative forcing changes due to hydrogen emissions in various Shared Socioeconomic Pathways (SSP), offering a better understanding of the potential climate impact of hydrogen as an alternative fuel. Future research should thus focus on continuous model evaluation and refinement to ensure that surrogate models remain robust tools for relevant climate assessments.

Moreover, the accuracy of the surrogate models employed in this work can be greatly improved by training them on bespoke simulations concerning aviation-specific hydrogen emissions. This would circumvent the usage of a simplistic box model and result in the direct determination of hydrogen-induced radiative forcing changes, which is especially crucial for aviation. Emissions from aircraft occur at high altitudes, and hydrogen released in the stratosphere could have a significantly different climate effect than the current assumption that hydrogen is well-mixed in the troposphere. High-altitude emissions may lead to longer atmospheric lifetimes and altered chemical interactions, potentially exacerbating their climate impact. Incorporating altitude-specific emission profiles and atmospheric dynamics into the surrogate models would enhance their precision and applicability.

Investigating the effects of reductions or increases in co-emitted species related to hydrogen use is essential for a comprehensive understanding of hydrogen's climate impact. Co-emitted species, such as nitrogen oxides (NO_x), water vapour, and other reactive compounds, can influence atmospheric chemistry to amplify or counteract the benefits of hydrogen as a low-carbon fuel. For instance, while hydrogen combustion itself does not produce CO_2, it can lead to the formation of NO_x under high-temperature conditions, which may affect both ozone and methane concentrations—two critical greenhouse gases that significantly impact radiative forcing. Additionally, increased water vapour emissions at high altitudes, a byproduct of hydrogen combustion, could contribute to cirrus cloud formation and alter atmospheric water vapour levels, further impacting radiative forcing and potentially leading to net warming effects.

Reductions or increases in these co-emitted species could have cascading effects on atmospheric processes. For example, elevated levels of NO_x may increase ozone concentrations in the troposphere, enhancing warming, whereas a reduction in NO_x could have the opposite effect. Similarly, changes in water vapour emissions at high altitudes could affect the radiation balance by influencing cloud formation and persistence. To assess these dynamics, it is crucial to model how different scenarios of co-emitted species interact with atmospheric components, particularly under varying emission rates and environmental conditions. Such modelling efforts should consider the nonlinear and region-specific chemical reactions in the atmosphere.

Finally, the neural network-based surrogate modelling approach proposed in this work could provide the foundation for future response modelling frameworks for other species, enabling rapid, computationally efficient predictions across a wide range of atmospheric and climate scenarios. By extending this methodology, researchers could explore complex feedback mechanisms, assess mitigation strategies, and improve the understanding of anthropogenic impacts on the climate system, paving the way for more informed policy decisions and proactive climate action.

Bibliography

Acosta, M. C., and Coauthors, 2024: The computational and energy cost of simulation and storage for climate science: lessons from cmip6. *Geoscientific Model Development*, **17 (8)**, 3081–3098, https://doi.org/10.5194/gmd-17-3081-2024.

Adler, E. J., and J. R. R. A. Martins, 2023: Hydrogen-powered aircraft: Fundamental concepts, key technologies, and environmental impacts. *Progress in Aerospace Sciences*, **141**, 100 922, https://doi.org/10.1016/j.paerosci.2023.100922.

Allen, M. R., K. P. Shine, J. S. Fuglestvedt, R. J. Millar, M. Cain, D. J. Frame, and A. H. Macey, 2018: A solution to the misrepresentations of CO2-equivalent emissions of short-lived climate pollutants under ambitious mitigation. *npj Climate and Atmospheric Science*, **1 (1)**, 1–8, https://doi.org/10.1038/s41612-018-0026-8.

Archer, D., and Coauthors, 2009: Atmospheric Lifetime of Fossil Fuel Carbon Dioxide. *Annual Review of Earth and Planetary Sciences*, **37 (Volume 37, 2009)**, 117–134, https://doi.org/10.1146/annurev.earth.031208.100206.

Bertagni, M. B., S. W. Pacala, F. Paulot, and A. Porporato, 2022: Risk of the hydrogen economy for atmospheric methane. *Nature Communications*, **13 (1)**, 7706, https://doi.org/10.1038/s41467-022-35419-7.

Bock, L., and U. Burkhardt, 2016: Reassessing properties and radiative forcing of contrail cirrus using a climate model. *Journal of Geophysical Research: Atmospheres*, **121 (16)**, 9717–9736, https://doi.org/10.1002/2016JD025112.

Bond, T. C., and Coauthors, 2013: Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, **118 (11)**, 5380–5552, https://doi.org/10.1002/jgrd.50171.

Brazzola, N., A. Patt, and J. Wohland, 2022: Definitions and implications of climate-neutral aviation. *Nature Climate Change*, **12 (8)**, 761–767.

Calvin, K., and Coauthors, 2017: The SSP4: A world of deepening inequality. *Global Environmental Change*, **42**, 284–296, https://doi.org/10.1016/j.gloenvcha.2016.06.010.

Cooper, O. R., and Coauthors, 2014: Global distribution and trends of tropospheric ozone: An observation-based review. *Elementa: Science of the Anthropocene*, **2**, 000 029, https://doi.org/10.12952/journal.elementa.000029.

Derwent, R., P. Simmonds, S. O'Doherty, A. Manning, W. Collins, and D. Stevenson, 2006: Global environmental impacts of the hydrogen economy. *International Journal of Nuclear Hydrogen Production and Applications*, **1 (1)**, 57–67, https://doi.org/10.1504/IJNHPA.2006.009869.

Derwent, R. G., 1990: *Trace Gases and Their Relative Contribution to the Greenhouse Effect*. Atomic Energy Research Establishment, Environmental & Medical Sciences Division, google-Books-ID: yux5AAAACAAJ.

Derwent, R. G., 2023: Global warming potential (GWP) for hydrogen: Sensitivities, uncertainties and meta-analysis. *International Journal of Hydrogen Energy*, **48 (22)**, 8328–8341, https://doi.org/10.1016/j.ijhydene.2022.11.219.

Derwent, R. G., D. S. Stevenson, S. R. Utembe, M. E. Jenkin, A. H. Khan, and D. E. Shallcross, 2020: Global modelling studies of hydrogen and its isotopomers using STOCHEM-CRI: Likely radiative forcing consequences of a future hydrogen economy. *International Journal of Hydrogen Energy*, **45 (15)**, 9211–9221, https://doi.org/10.1016/j.ijhydene.2020.01.125.

Ebi, K. L., and Coauthors, 2014: A new scenario framework for climate change research: background, process, and future directions. *Climatic Change*, **122 (3)**, 363–372, https://doi.org/10.1007/s10584-013-0912-3.

Ehhalt, D. H., and F. Rohrer, 2009: The tropospheric cycle of H2: a critical review. *Tellus B: Chemical and Physical Meteorology*, **61 (3)**, 500–535, https://doi.org/10.1111/j.1600-0889.2009.00416.x.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9 (5)**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Fan, Z., and Coauthors, 2022: Hydrogen Leakage: A Potential Risk for the Hydrogen Economy. Tech. rep., Columbia Center on Global Energy Policy, New York, United States. URL https://www.energypolicy.columbia.edu/wp-content/uploads/2022/07/HydrogenLeakageRegulations_CGEP_Commentary_063022.pdf.

Forster, C., A. Stohl, P. James, and V. Thouret, 2003: The residence times of aircraft emissions in the stratosphere using a mean emission inventory and emissions along actual flight tracks. *Journal of Geophysical Research: Atmospheres*, **108 (D12)**, https://doi.org/10.1029/2002JD002515.

Friedlingstein, P., and Coauthors, 2023: Global carbon budget 2023. *Earth System Science Data*, **15 (12)**, 5301–5369.

Fuglestvedt, J. S., and Coauthors, 2010: Transport impacts on atmosphere and climate: Metrics. *Atmospheric Environment*, **44 (37)**, 4648–4677, https://doi.org/10.1016/j.atmosenv.2009.04.044.

Fujimori, S., T. Hasegawa, T. Masui, K. Takahashi, D. S. Herran, H. Dai, Y. Hijioka, and M. Kainuma, 2017: SSP3: AIM implementation of Shared Socioeconomic Pathways. *Global Environmental Change*, **42**, 268–283, https://doi.org/10.1016/j.gloenvcha.2016.06.009.

Gangoli Rao, A., F. Yin, and H. Werij, 2020: Energy transition in aviation: the role of cryogenic fuels. *Aerospace — Open Access Aeronautics and Astronautics Journal*, **7 (12)**, 1–24, https://doi.org/10.3390/aerospace7120181.

Gauss, M., I. S. A. Isaksen, S. Wong, and W.-C. Wang, 2003: Impact of H2O emissions from cryoplanes and kerosene aircraft on the atmosphere. *Journal of Geophysical Research: Atmospheres*, **108 (D10)**, https://doi.org/10.1029/2002JD002623.

Gettelman, A., P. Hoor, L. L. Pan, W. J. Randel, M. I. Hegglin, and T. Birner, 2011: The Extratropical Upper Troposphere and Lower Stratosphere. *Reviews of Geophysics*, **49 (3)**, https://doi.org/10.1029/2011RG000355.

Grewe, V., 2020: Aviation Emissions and Climate Impacts. *Aviation and Climate Change*, F. Fichert, P. Forsyth, and H.-M. Niemeier, Eds., Roudledge, london, https://doi.org/10.4324/9781315572406.

Grewe, V., and K. Dahlmann, 2015: How ambiguous are climate metrics? And are we prepared to assess and compare the climate impact of new air traffic technologies? *Atmospheric Environment*, **106**, 373–374, https://doi.org/https://doi.org/10.1016/j.atmosenv.2015.02.039.

Grewe, V., S. Matthes, and K. Dahlmann, 2019: The contribution of aviation NOx emissions to climate change: are we ignoring methodological flaws? *Environmental Research Letters*, **14 (12)**, 121 003, https://doi.org/10.1088/1748-9326/ab5dd7.

Grewe, V., and A. Stenke, 2008: AirClim: an efficient tool for climate evaluation of aircraft technology. *Atmospheric Chemistry and Physics*, **8 (16)**, 4621–4639, https://doi.org/10.5194/acp-8-4621-2008.

Grewe, V., and Coauthors, 2017: Assessing the climate impact of the AHEAD multi-fuel blended wing body. *Meteorologische Zeitschrift*, 711–725, https://doi.org/10.1127/metz/2016/0758.

Grewe, V., and Coauthors, 2021: Evaluating the climate impact of aviation emission scenarios towards the Paris agreement including COVID-19 effects. *Nature Communications*, **12 (1)**, 3841, https://doi.org/10.1038/s41467-021-24091-y.

Haq, M. A., 2021: CDLSTM: A Novel Model for Climate Change Forecasting. *Computers, Materials & Continua*, **71**, 2363–2381, https://doi.org/10.32604/cmc.2022.023059.

Hauglustaine, D., F. Paulot, W. Collins, R. Derwent, M. Sand, and O. Boucher, 2022: Climate benefit of a future hydrogen economy. *Communications Earth & Environment*, **3 (1)**, 1–14, https://doi.org/10.1038/s43247-022-00626-z.

Hochreiter, S., and J. Schmidhuber, 1997: Long Short-Term Memory. *Neural Computation*, **9 (8)**, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Holmes, C. D., Q. Tang, and M. J. Prather, 2011: Uncertainties in climate assessment for the case of aviation NO. *Proceedings of the National Academy of Sciences*, **108 (27)**, 10 997–11 002, https://doi.org/10.1073/pnas.1101458108.

IPCC, Ed., 1990: *Climate Change 1990 - IPCC first assessment report overview and policymaker summaries and 1992 IPPC supplement*. IPCC, Geneve.

IPCC, 2021a: Climate Change 2021 – Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. Tech. rep., Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647, URL https://www.ipcc.ch/report/ar6/syr/, edition: First.

IPCC, 2021b: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Tech. rep., Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.1017/9781009157896, URL https://report.ipcc.ch/ar6/wg1/IPCC_AR6_WGI_FullReport.pdf.

Isaksen, I. S. A., and □. Hov, 1987: Calculation of trends in the tropospheric concentration of O3, OH, CO, CH4 and NOx. *Tellus B*, **39B (3)**, 271–285, https://doi.org/10.1111/j.1600-0889.1987.tb00099.x.

Kriegler, E., and Coauthors, 2017: Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century. *Global Environmental Change*, **42**, 297–315, https://doi.org/10.1016/j.gloenvcha.2016.05.015.

Lee, D. S., and Coauthors, 2010: Transport impacts on atmosphere and climate: Aviation. *Atmospheric Environment*, **44 (37)**, 4678–4734, https://doi.org/10.1016/j.atmosenv.2009.06.005.

Lee, D. S., and Coauthors, 2021: The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. *Atmospheric Environment*, **244**, 117 834, https://doi.org/10.1016/j.atmosenv.2020.117834.

Levy, H., 1971: Normal Atmosphere: Large Radical and Formaldehyde Concentrations Predicted. *Science*, **173 (3992)**, 141–143, https://doi.org/10.1126/science.173.3992.141.

Lima, F. T., and V. M. A. Souza, 2023: A Large Comparison of Normalization Methods on Time Series. *Big Data Research*, **34**, 100 407, https://doi.org/10.1016/j.bdr.2023.100407.

Marquart, S., M. Ponater, L. Strö M, and K. Gierens, 2005: An upgraded estimate of the radiative forcing of cryoplane contrails. *Meteorologische Zeitschrift*, **14 (4)**, 573–582, https://doi.org/10.1127/0941-2948/2005/0057.

Maruhashi, J., M. Mertens, V. Grewe, and I. C. Dedoussi, 2024: A multi-method assessment of the regional sensitivities between flight altitude and short-term O $_3$ climate warming from aircraft NO $_x$ emissions. *Environmental Research Letters*, **19 (5)**, 054 007, https://doi.org/10.1088/1748-9326/ad376a.

McCausland, R., 2024: Aviation contrails and their climate effect. Tech. rep., International Air Transport Association (IATA).

Megill, L., K. Deck, and V. Grewe, 2024: Alternative climate metrics to the Global Warming Potential are more suitable for assessing aviation non-CO2 effects. *Communications Earth & Environment*, **5 (1)**, 1–9, https://doi.org/10.1038/s43247-024-01423-6.

Molina, M. J., and F. S. Rowland, 1974: Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. *Nature*, **249 (5460)**, 810–812, https://doi.org/10.1038/249810a0.

Monks, P. S., and Coauthors, 2015: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics*, **15 (15)**, 8889–8973, https://doi.org/10.5194/acp-15-8889-2015.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463 (7282)**, 747–756, https://doi.org/10.1038/nature08823.

Novelli, P. C., P. M. Lang, K. A. Masarie, D. F. Hurst, R. Myers, and J. W. Elkins, 1999: Molecular hydrogen in the troposphere: Global distribution and budget. *Journal of Geophysical Research: Atmospheres*, **104 (D23)**, 30 427–30 444, https://doi.org/10.1029/1999JD900788.

Ocko, I. B., and S. P. Hamburg, 2022: Climate consequences of hydrogen emissions. *Atmospheric Chemistry and Physics*, **22 (14)**, 9349–9368, https://doi.org/10.5194/acp-22-9349-2022.

O'Malley, T., E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, and Coauthors, 2019: Kerastuner. https://github.com/keras-team/keras-tuner.

O'Neill, B. C., E. Kriegler, K. Riahi, K. L. Ebi, S. Hallegatte, T. R. Carter, R. Mathur, and D. P. van Vuuren, 2014: A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, **122 (3)**, 387–400, https://doi.org/10.1007/s10584-013-0905-2.

O'Neill, B. C., and Coauthors, 2017: The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, **42**, 169–180, https://doi.org/10.1016/j.gloenvcha.2015.01.004.

Paulot, F., D. Paynter, V. Naik, S. Malyshev, R. Menzel, and L. W. Horowitz, 2021: Global modeling of hydrogen using GFDL-AM4.1: Sensitivity of soil removal and radiative forcing. *International Journal of Hydrogen Energy*, **46 (24)**, 13 446–13 460, https://doi.org/10.1016/j.ijhydene.2021.01.088.

Prather, M. J., 1994: Lifetimes and eigenstates in atmospheric chemistry. *Geophysical Research Letters*, **21 (9)**, 801–804, https://doi.org/10.1029/94GL00840.

Prather, M. J., C. D. Holmes, and J. Hsu, 2012: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry. *Geophysical Research Letters*, **39 (9)**, https://doi.org/10.1029/2012GL051440.

Pétron, G., and Coauthors, 2024: Atmospheric H_2 observations from the NOAA Cooperative Global Air Sampling Network. *Atmospheric Measurement Techniques*, **17 (16)**, 4803–4823, https://doi.org/10.5194/amt-17-4803-2024.

Riahi, K., and Coauthors, 2017: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, **42**, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Rodhe, H., 1990: A comparison of the contribution of various gases to the greenhouse effect. *Science (New York, N.Y.)*, **248 (4960)**, 1217–1219, https://doi.org/10.1126/science.248.4960.1217.

Rogelj, J., and Coauthors, 2018: Scenarios towards limiting global mean temperature increase below 1.5 °C. *Nature Climate Change*, **8 (4)**, 325–332, https://doi.org/10.1038/s41558-018-0091-3.

S Butterworth, 1930: On the Theory of Filter Amplifiers. *Experimental Wireless and the Wireless Engineer*, **7**, 536–541.

Sand, M., and Coauthors, 2023: A multi-model assessment of the Global Warming Potential of hydrogen. *Communications Earth & Environment*, **4 (1)**, 1–12, https://doi.org/10.1038/s43247-023-00857-8.

Savage, R. L., 1985: Diffusion Research Traditions and the Spread of Policy Innovations in a Federal System. *Publius: The Journal of Federalism*, **15 (4)**, 1–28, https://doi.org/10.1093/oxfordjournals.pubjof.a037561.

Schleussner, C.-F., and Coauthors, 2016: Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change*, **6 (9)**, 827–835, https://doi.org/10.1038/nclimate3096.

Schumann, U., 1997: The impact of nitrogen oxides emissions from aircraft upon the atmosphere at flight altitudes—results from the aeronox project. *Atmospheric Environment*, **31 (12)**, 1723–1733, https://doi.org/10.1016/S1352-2310(96)00326-3.

Schumann, U., 2005: Formation, properties and climatic effects of contrails. *Comptes Rendus Physique*, **6 (4)**, 549–565, https://doi.org/10.1016/j.crhy.2005.05.002.

Shine, K., J. Fuglestvedt, K. Hailemariam, and N. Stuber, 2005: Alternatives to the Global Warming Potential for Comparing Climate Impacts of Emissions of Greenhouse Gases. *Climatic Change*, **68**, 281–302, https://doi.org/10.1007/s10584-005-1146-9.

Siami-Namini, S., N. Tavakoli, and A. S. Namin, 2019: The Performance of LSTM and BiLSTM in Forecasting Time Series. *2019 IEEE International Conference on Big Data (Big Data)*, 3285–3292, https://doi.org/10.1109/BigData47090.2019.9005997, URL https://ieeexplore.ieee.org/document/9005997/?arnumber=9005997.

Siami-Namini, S., N. Tavakoli, and A. Siami Namin, 2018: A Comparison of ARIMA and LSTM in Forecasting Time Series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401, https://doi.org/10.1109/ICMLA.2018.00227, URL https://ieeexplore.ieee.org/document/8614252/?arnumber=8614252.

Solomon, S., K. H. Rosenlof, R. W. Portmann, J. S. Daniel, S. M. Davis, T. J. Sanford, and G.-K. Plattner, 2010: Contributions of Stratospheric Water Vapor to Decadal Changes in the Rate of Global Warming. *Science*, **327 (5970)**, 1219–1223, https://doi.org/10.1126/science.1182488.

Stevenson, D. S., R. M. Doherty, M. G. Sanderson, W. J. Collins, C. E. Johnson, and R. G. Derwent, 2004: Radiative forcing from aircraft NO emissions: Mechanisms and seasonal dependence. *Journal of Geophysical Research: Atmospheres*, **109 (D17)**, https://doi.org/10.1029/2004JD004759.

Stevenson, D. S., and Coauthors, 2013: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmospheric Chemistry and Physics*, **13 (6)**, 3063–3085, https://doi.org/10.5194/acp-13-3063-2013.

Stuber, N., P. Forster, G. Rädel, and K. Shine, 2006: The importance of the diurnal and annual cycle of air traffic for contrail radiative forcing. *Nature*, **441 (7095)**, 864–867, https://doi.org/10.1038/nature04877.

Sáez Ortuño, M. n., F. Yin, A. Gangoli Rao, R. Vos, and P.-J. Proesmans, 2023: Climate Assessment of Hydrogen Combustion Aircraft: Towards a Green Aviation Sector. *AIAA SCITECH 2023 Forum*, AIAA SciTech Forum, American Institute of Aeronautics and Astronautics, https://doi.org/10.2514/6.2023-2513, URL https://arc.aiaa.org/doi/10.2514/6.2023-2513.

Turnock, S. T., and Coauthors, 2020: Historical and future changes in air pollutants from CMIP6 models. *Atmospheric Chemistry and Physics*, **20 (23)**, 14 547–14 579, https://doi.org/10.5194/acp-20-14547-2020.

Twomey, S., 1977: The Influence of Pollution on the Shortwave Albedo of Clouds. *Journal of the Atmospheric Sciences*.

Warwick, N., P. Griffiths, A. Archibald, and J. Pyle, 2022: Atmospheric implications of increased hydrogen use. *London, Crown: University of Cambridge and NCAS and Keith Shine and Unvercity of Reading*.

Winterstein, F., and P. Jöckel, 2021: Methane chemistry in a nutshell – the new submodels CH4 (v1.0) and TRSYNC (v1.0) in MESSy (v2.54.0). *Geoscientific Model Development*, **14 (2)**, 661–674, https://doi.org/10.5194/gmd-14-661-2021.

Wuebbles, D., P. Forster, H. Rogers, and R. Herman, 2010: Issues and uncertainties affecting metrics for aviation impacts on climate. *Bulletin of the American Meteorological Society*, **91 (4)**, 491–496, https://doi.org/10.1175/2009BAMS2840.1.

Yukimoto, S., and Coauthors, 2019: The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component. *Journal of the Meteorological Society of Japan*, **97**, https://doi.org/10.2151/jmsj.2019-051.

Züttel, A., 2003: Materials for hydrogen storage. *Materials Today*, **6 (9)**, 24–33, https://doi.org/10.1016/S1369-7021(03)00922-2.

A

Machine Learning Concepts

This appendix offers a concise overview of the key machine learning concepts pertinent to this thesis, serving as a foundation for understanding the methodologies employed. It highlights essential principles that underpin the analysis and modeling techniques discussed in later chapters.

Training and Inference

Training is the process of optimizing a machine learning model's parameters using a dataset. During training, the model learns to map inputs to outputs by minimizing a loss function $L(\theta)$ represents the model parameters.

The optimization is typically performed using algorithms such as stochastic gradient descent (SGD), which iteratively updates the model parameters:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t), \tag{A.1}$$

where:

- $L(\theta_t)$ is the parameter vector at iteration t,
- η is the learning rate,
- $\nabla_\theta L(\theta_t)$is the gradient of the loss function with respect to the parameters at iteration t.

Training involves the following steps:

1. **Forward Pass**: Compute the output of the model given the current parameters and input data.
2. **Loss Computation**: Calculate the loss by comparing the model's output to the true labels.
3. **Backward Pass (Backpropagation)**: Compute the gradients of the loss with respect to the model parameters.
4. **Parameter Update**: Update the model parameters using the gradients and the learning rate.

The learning rate η is a crucial hyperparameter that controls the magnitude of the parameter updates during training. It affects both the speed of convergence and the stability of the training process.

Updating parameters using gradient descent involves:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t). \tag{A.2}$$

A too large learning rate may cause the training to overshoot minima and diverge, while a too small learning rate may lead to slow convergence.

Inference refers to the phase where the trained model is used to make predictions on new, unseen data. The model applies the learned parameters θ^* to input data X to produce output $\hat{y}$:

$$\hat{\mathbf{y}} = f(\mathbf{x}; \theta^*), \tag{A.3}$$

where $f$is the function represented by the model, and θ^* are the optimized parameters obtained after training. During inference, the model parameters remain fixed.

Activation Functions

Activation functions introduce non-linearity into neural networks, enabling them to model complex patterns and relationships in data.

Sigmoid Function The sigmoid activation function is defined as:

$$\sigma(x) = \frac{1}{1+e^{-x}}. \tag{A.4}$$

Properties:

- Output range: (0,1).
- Commonly used in the output layer for binary classification tasks.
- Can cause vanishing gradient problems for large input magnitudes due to saturation.

Hyperbolic Tangent (Tanh) Function The tanh activation function is defined as:

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}. \tag{A.5}$$

Properties:

- Output range: (−1,1).
- Zero-centered output, which can be advantageous for optimization.
- Also susceptible to vanishing gradients for large input magnitudes.

Rectified Linear Unit (ReLU) The ReLU activation function is defined as:

$$\text{ReLU}(x) = \max(0, x). \tag{A.6}$$

Properties:

- Output range: [0,∞).
- Computationally efficient and mitigates the vanishing gradient problem.
- Can suffer from the ”dying ReLU” problem where neurons become inactive for all inputs.

Neural Network Architectures

Recurrent Neural Networks (RNNs) RNNs are designed to handle sequential data by maintaining a hidden state h_t that captures information from previous time steps. The hidden state is updated as follows:

$$\mathbf{h}t = \phi(\mathbf{W}hh\mathbf{h}t - 1 + \mathbf{W}xh\mathbf{x}_t + \mathbf{b}_h), \tag{A.7}$$

where:

- X_t is the input at time step t
- h_{t-1} is the hidden state from the previous time step,
- W_{hh} and W_{xh} are weight matrices,
- $b_h$is the bias vector,
- ϕ an activation function, typically tanh or ReLU.

Long Short-Term Memory (LSTM) Networks LSTMs are a type of RNN that addresses the vanishing gradient problem by incorporating memory cells and gating mechanisms to control the flow of information.

The LSTM cell comprises the following components:

$$\text{Input gate: } \mathbf{i}t = \sigma(\mathbf{W}xi\mathbf{x}t + \mathbf{W}hi\mathbf{h}t - 1 + \mathbf{W}ci\mathbf{C}t - 1 + \mathbf{b}i), \tag{A.8}$$

$$\text{Forget gate: } \mathbf{f}t = \sigma(\mathbf{W}xf\mathbf{x}t + \mathbf{W}hf\mathbf{h}t - 1 + \mathbf{W}cf\mathbf{C}_{t-1} + \mathbf{b}f), \tag{A.9}$$

$$\text{Cell candidate: } \tilde{\mathbf{C}}t = \tanh(\mathbf{W}xC\mathbf{x}t + \mathbf{W}hC\mathbf{h}t - 1 + \mathbf{b}_C), \tag{A.10}$$

$$\text{Cell state: } \mathbf{C}_t = \mathbf{f}t \odot \mathbf{C}t - 1 + \mathbf{i}t \odot \tilde{\mathbf{C}}t, \tag{A.11}$$

$$\text{Output gate: } \mathbf{o}t = \sigma(\mathbf{W}xo\mathbf{x}t + \mathbf{W}ho\mathbf{h}t - 1 + \mathbf{W}co\mathbf{C}_t + \mathbf{b}_o), \tag{A.12}$$

$$\text{Hidden state: } \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t), \tag{A.13}$$

where:

- σ is the sigmoid activation function,
- tanh is the hyperbolic tangent function,
- $\odot$ denotes element-wise multiplication,
- W matrices are weight matrices for different gates,
- b are bias vectors.

Convolutional LSTM (ConvLSTM) extends LSTM to handle spatiotemporal data by replacing matrix multiplications with convolution operations. This allows the network to capture both temporal and spatial dependencies.

The ConvLSTM equations are similar to those of LSTM but with convolution operations:

$$\mathbf{i}t = \sigma(\mathbf{W}xi * \mathbf{X}t + \mathbf{W}hi * \mathbf{H}t - 1 + \mathbf{W}ci \odot \mathbf{C}t - 1 + \mathbf{b}i), \tag{A.14}$$

$$\mathbf{f}t = \sigma(\mathbf{W}xf * \mathbf{X}t + \mathbf{W}hf * \mathbf{H}t - 1 + \mathbf{W}cf \odot \mathbf{C}_{t-1} + \mathbf{b}f), \tag{A.15}$$

$$\tilde{\mathbf{C}}t = \tanh(\mathbf{W}xC * \mathbf{X}t + \mathbf{W}hC * \mathbf{H}t - 1 + \mathbf{b}_C), \tag{A.16}$$

$$\mathbf{C}_t = \mathbf{f}t \odot \mathbf{C}t - 1 + \mathbf{i}t \odot \tilde{\mathbf{C}}t, \tag{A.17}$$

$$\mathbf{o}t = \sigma(\mathbf{W}xo * \mathbf{X}t + \mathbf{W}ho * \mathbf{H}t - 1 + \mathbf{W}co \odot \mathbf{C}_t + \mathbf{b}_o), \tag{A.18}$$

$$\mathbf{H}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t), \tag{A.19}$$

where:

- $*$ denotes the convolution operation,
- X_t is the input tensor at time t
- H_{T-1} is the hidden state tensor from the previous time step,
- Other symbols are analogous to those in the standard LSTM.

Hyperparameter Tuning

Hyperparameter tuning involves finding the optimal set of hyperparameters that yields the best model performance.

Grid Search Grid search exhaustively explores all combinations of specified hyperparameter values.

Random Search Random search samples hyperparameter combinations randomly. It has been shown to be more efficient than grid search in high-dimensional spaces.

Bayesian Optimization Bayesian optimization uses a probabilistic model (e.g., Gaussian processes) to model the performance metric as a function of hyperparameters and selects hyperparameters that are likely to improve performance.

Batching

Batching refers to the process of dividing the training dataset into smaller subsets called batches.

Batch Gradient Descent Uses the entire dataset to compute the gradients:

$$\nabla_\theta L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \nabla_\theta L^{(i)}(\theta), \tag{A.20}$$

where N is the total number of training samples.

Mini-Batch Gradient Descent Uses batches of size B to compute gradients:

$$\nabla_\theta L(\theta) = \frac{1}{B} \sum_{i=1}^{B} \nabla_\theta L^{(i)}(\theta). \tag{A.21}$$

Mini-batches provide a balance between the robustness of batch gradient descent and the computational efficiency of stochastic gradient descent.

Regularization

Regularization techniques help prevent overfitting by adding constraints to the model.

L2 Regularization Also known as weight decay, L2 regularization adds a penalty term proportional to the square of the magnitude of the weights to the loss function:

$$L_{\text{total}}(\theta) = L(\theta) + \lambda |\theta|_2^2, \tag{A.22}$$

where $\lambda$is the regularization strength.

Dropout Dropout randomly sets a fraction p of the neurons to zero during each training iteration.

Mathematically, the activation $h^{(l)}$becomes:

$$\mathbf{h}^{(l)} = \mathbf{h}^{(l)} \odot \mathbf{r}, \tag{A.23}$$

where r is a binary mask vector with elements sampled from:

$$r_i \sim \text{Bernoulli}(1 - p). \tag{A.24}$$

Early Stopping Early stopping involves monitoring the model's performance on a validation set during training and halting the training process when the validation performance begins to deteriorate.

K-Fold Cross-Validation K-fold cross-validation is used to assess the generalization capability of a model.

The dataset is partitioned into K equal-sized folds. The cross-validation process involves:

1. For each $k = 1$ to K:
   (a) Use fold k as the validation set.
   (b) Use the remaining $K - 1$ folds as the training set.
   (c) Train the model on the training set and evaluate it on the validation set.

2. Aggregate the performance metrics (e.g., accuracy, RMSE) over all K folds.

This technique provides a robust estimate of model performance by reducing the variance associated with a single train-test split.

Weight Initialisation

Proper weight initialization is essential for effective training of deep neural networks.

Random Normal Initialization Weights are initialized from a normal (Gaussian) distribution:

$$W_{ij} \sim \mathcal{N}(\mu, \sigma^2). \tag{A.25}$$

Random Uniform Initialization Weights are initialized from a uniform distribution:

$$W_{ij} \sim \mathcal{U}(a, b), \tag{A.26}$$

where a and b define the range.

Glorot Normal (Xavier Normal) The Glorot normal initializer sets the standard deviation based on the number of input and output units:

$$W_{ij} \sim \mathcal{N}\left(0, \frac{2}{n_{\text{in}} + n_{\text{out}}}\right), \tag{A.27}$$

where n_{in} and n_out are the number of input and output units of the layer.

Glorot Uniform (Xavier Uniform) Weights are initialized uniformly within a range:

$$W_{ij} \sim \mathcal{U}\left(-\sqrt{\frac{6}{n_{\text{in}} + n_{\text{out}}}}, \sqrt{\frac{6}{n_{\text{in}} + n_{\text{out}}}}\right). \tag{A.28}$$

B

ScenarioMIP Data

This appendix provides an overview of key data derived from MRI-ESM2.0 simulations conducted as part of the CMIP6 ScenarioMIP project. The data includes simulations that have been appended with historical runs from the same model, starting from 1970, to provide a seamless temporal sequence. The combined dataset offers a comprehensive view of climate projections in the context of historical trends. Before inclusion in this appendix, the data was processed according to the methods described in section 3.3.2.

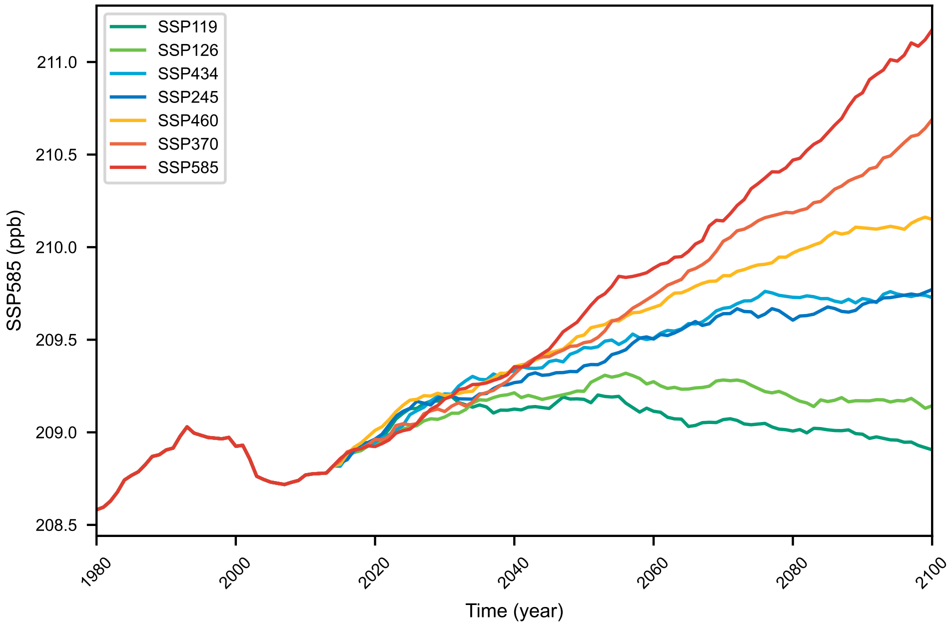


Figure B.1: Temporal evolution temperature at the tropopause for various SSP scenarios simulated with MRI-ESM2.0. Data incorporates historical runs (from 1970 until 2014) appended to the ScenarioMIP projections (from 2015 until 2100), processed as described in section 3.3.2

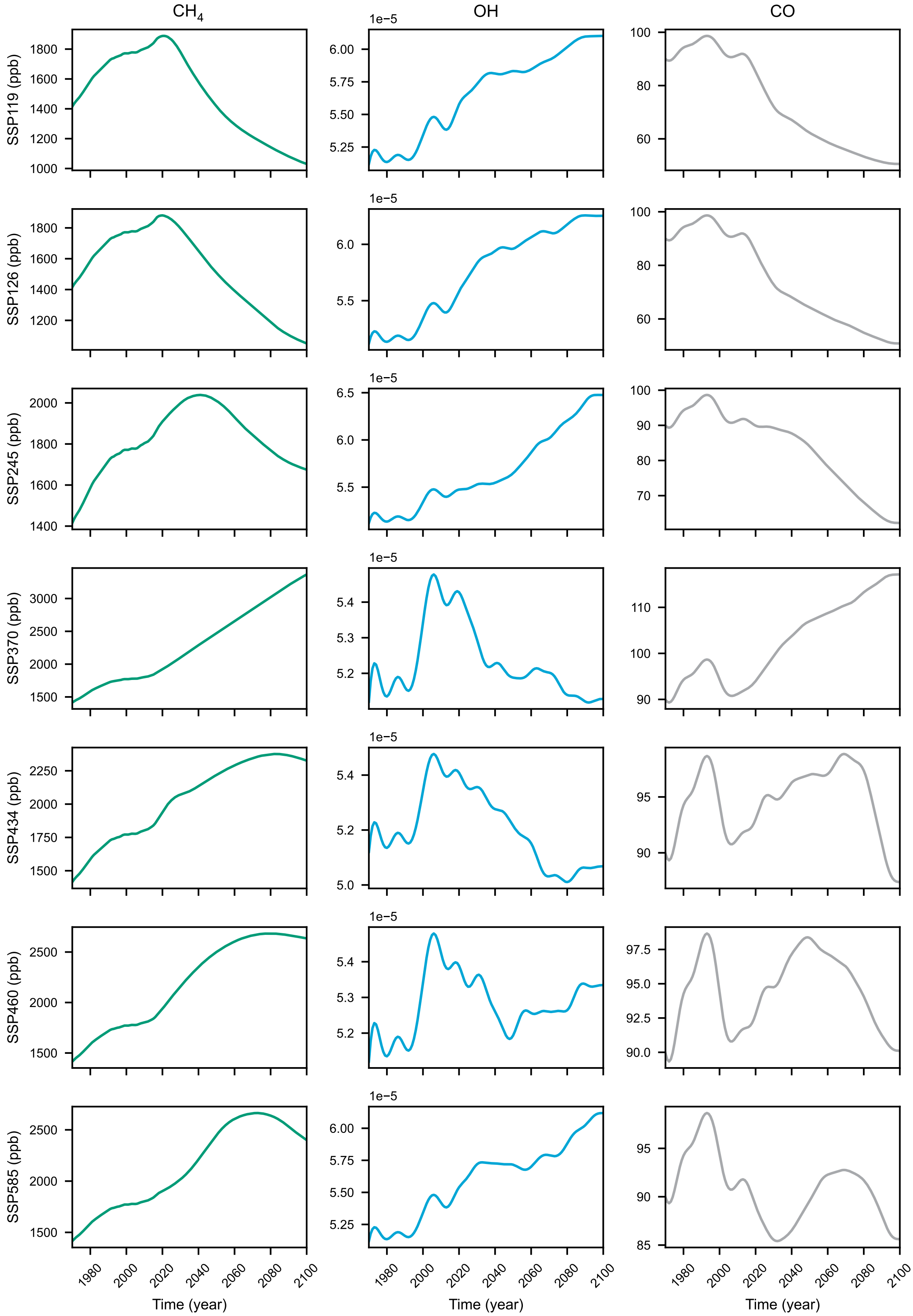


Figure B.2: Temporal evolution of CH$_4$, OH, and CO mixing ratios (ppb) for various SSP scenarios simulated with MRI-ESM2.0. Data incorporates historical runs (from 1970 until 2014) appended to the ScenarioMIP projections (from 2015 until 2100), processed as described in section 3.3.2

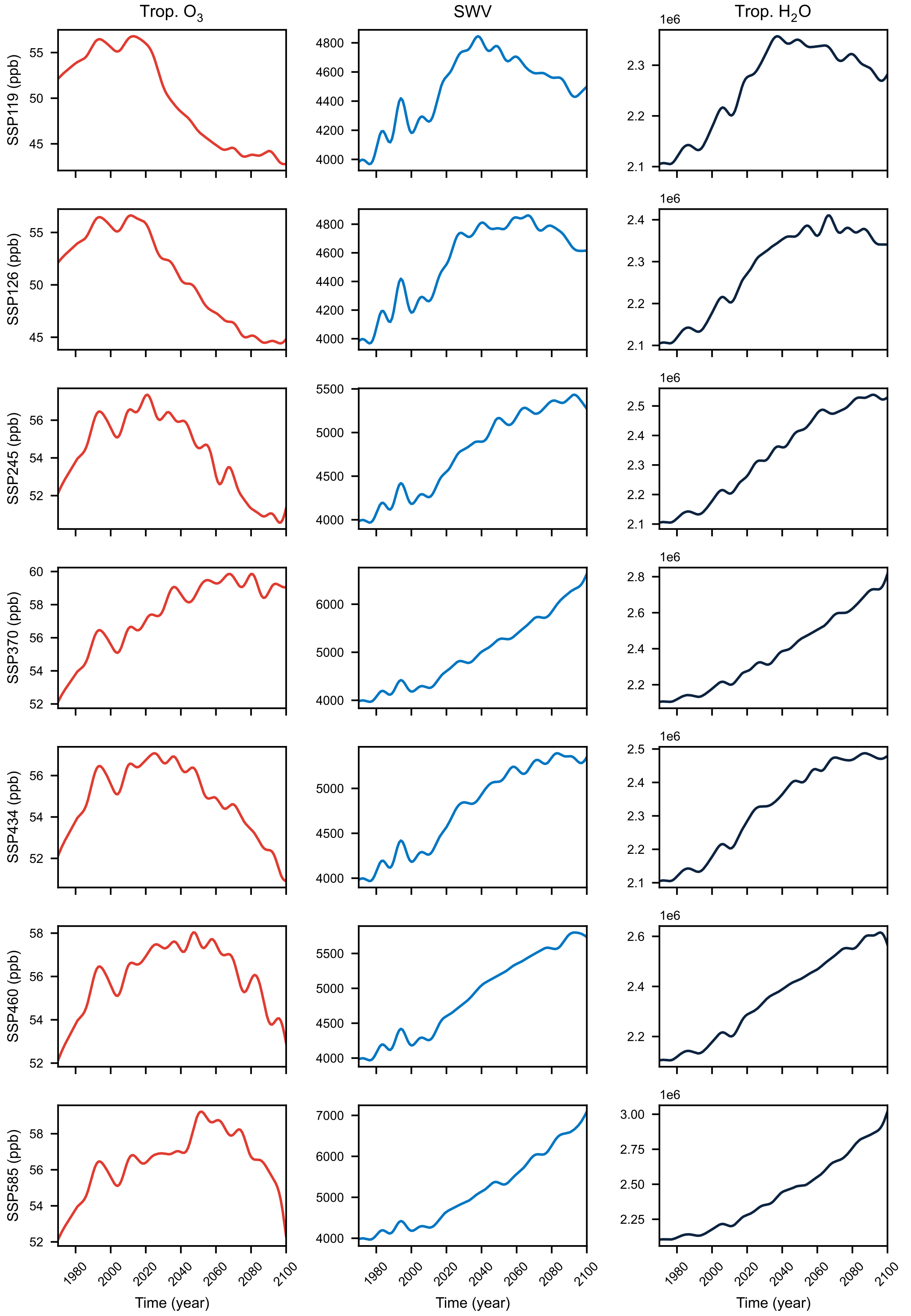


Figure B.3: Temporal evolution of Trop. O_3, SWV, and Trop. H_2O mixing ratios (ppb) for various SSP scenarios simulated with MRI-ESM2.0. Data incorporates historical runs (from 1970 until 2014) appended to the ScenarioMIP projections (from 2015 until 2100), processed as described in section 3.3.2

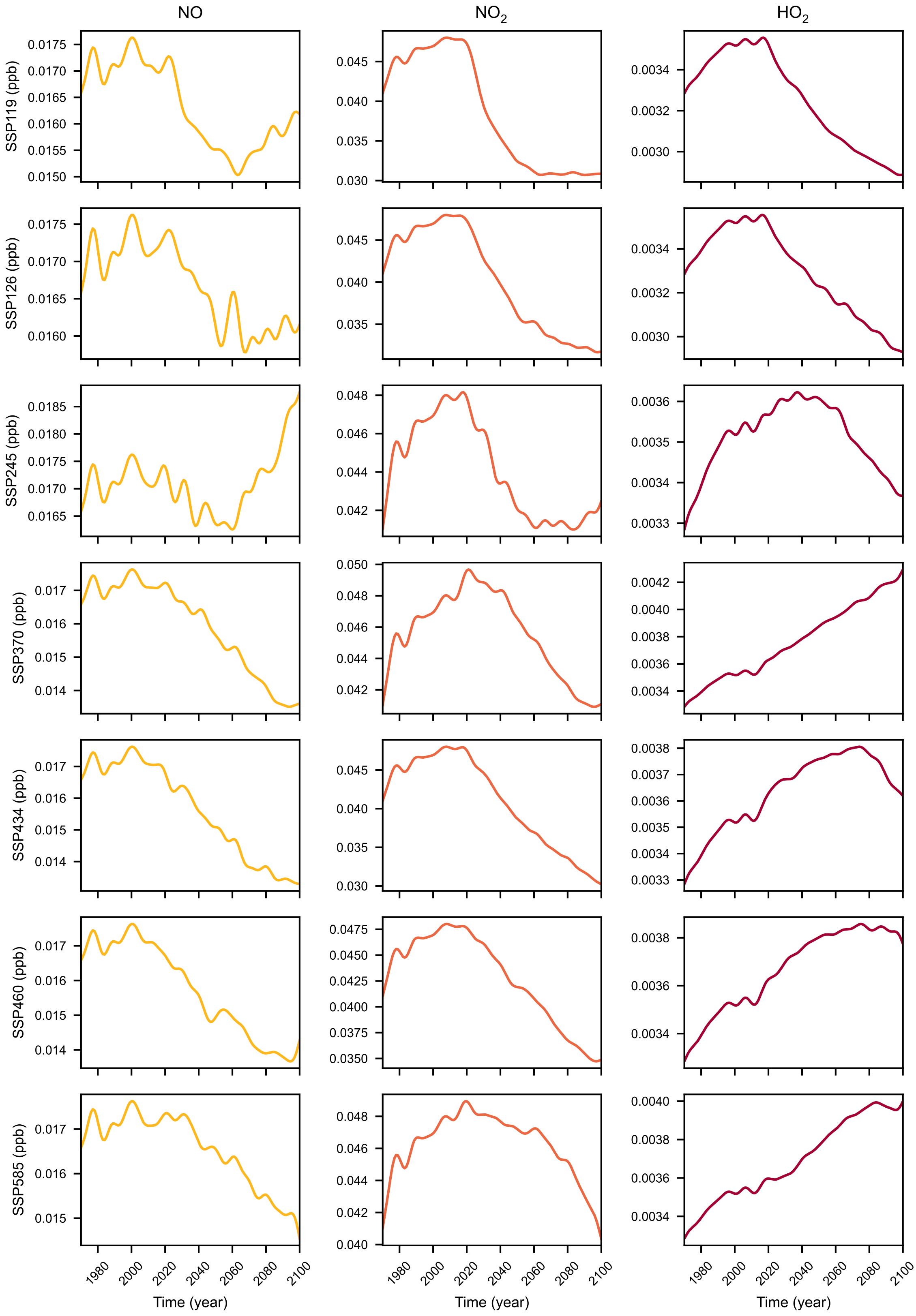


Figure B.4: Temporal evolution of NO, NO_2, and HO_2 mixing ratios (ppb) for various SSP scenarios simulated with MRI-ESM2.0. Data incorporates historical runs (from 1970 until 2014) appended to the ScenarioMIP projections (from 2015 until 2100), processed as described in section 3.3.2

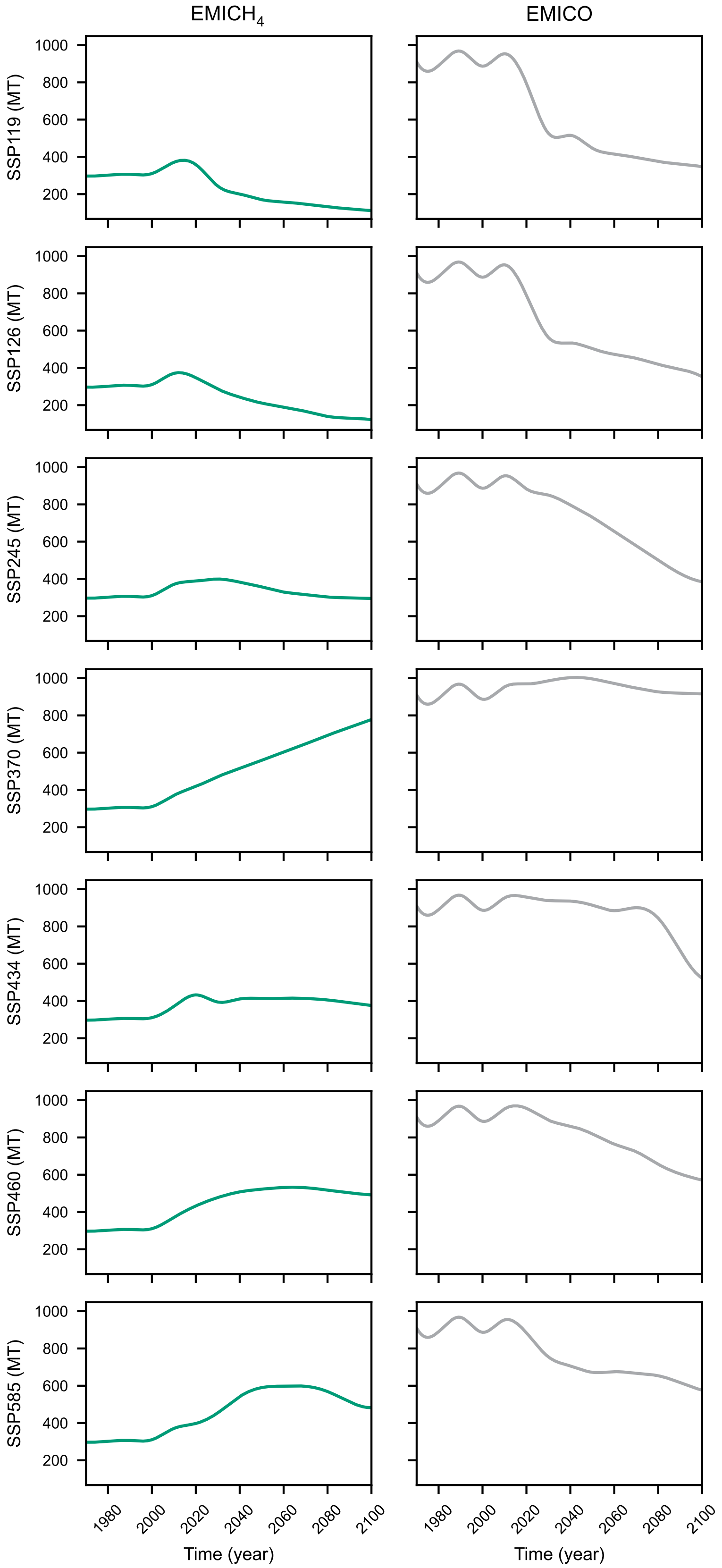


Figure B.5: Temporal evolution of CH_4 and CO emissions (MT) for various SSP scenarios simulated with MRI-ESM2.0. Data incorporates historical runs (from 1970 until 2014) appended to the ScenarioMIP projections (from 2015 until 2100), processed as described in section 3.3.2

C

Supplemental Results

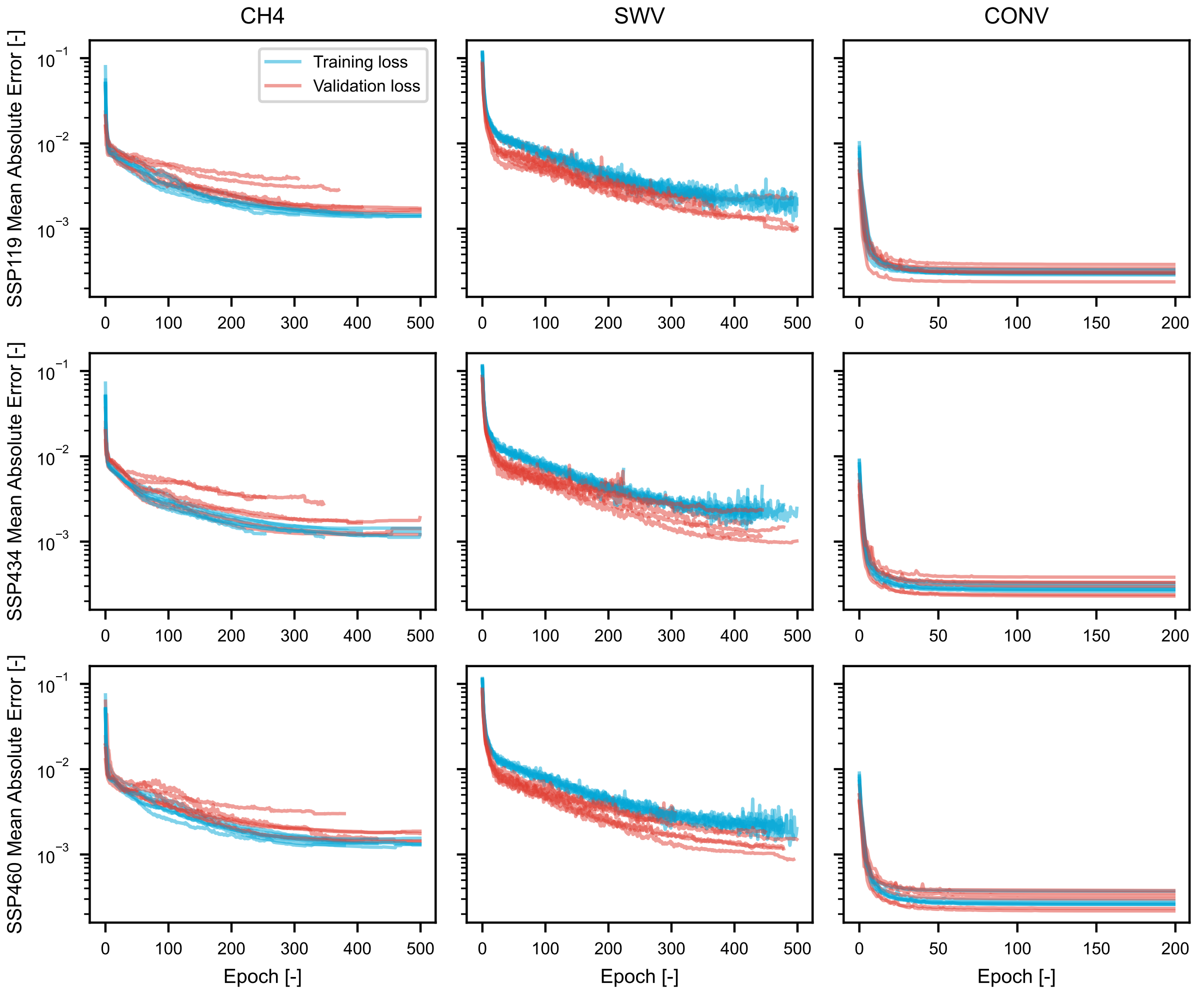


Figure C.1: Training loss (cyan) and validation loss (red) for the CH_4 and SWV LSTM models, and the O_3 ConvLSTM model, tested on scenario SSP1-1.9, SSP4-3.4 and SSP4-6.0, plotted as a function of epoch number on a logarithmic scale.

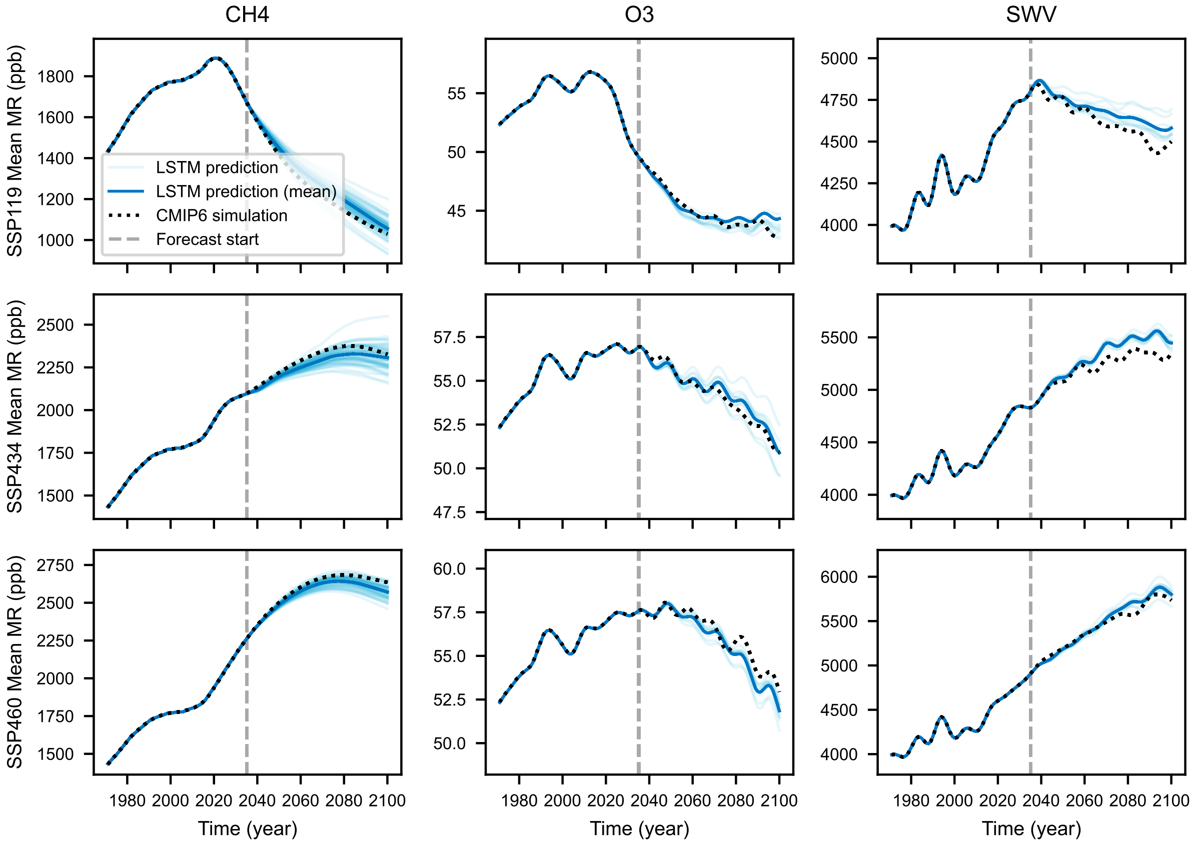


Figure C.2: Temporal evolution of mean tropospheric CH_4, O_3 and SWV molar fractions (expressed in ppb) as simulated by the MRI-ESM2.0 model for SSP1-1.9, SSP4-3.4 and SSP4-6.0 (black dots), plotted against the LSTM predictions (shaded blue) and the mean of the LSTM predictions (bold blue). The model simulates from the year 2035 onwards. Note that the y-axis is not scaled evenly between the different scenarios.