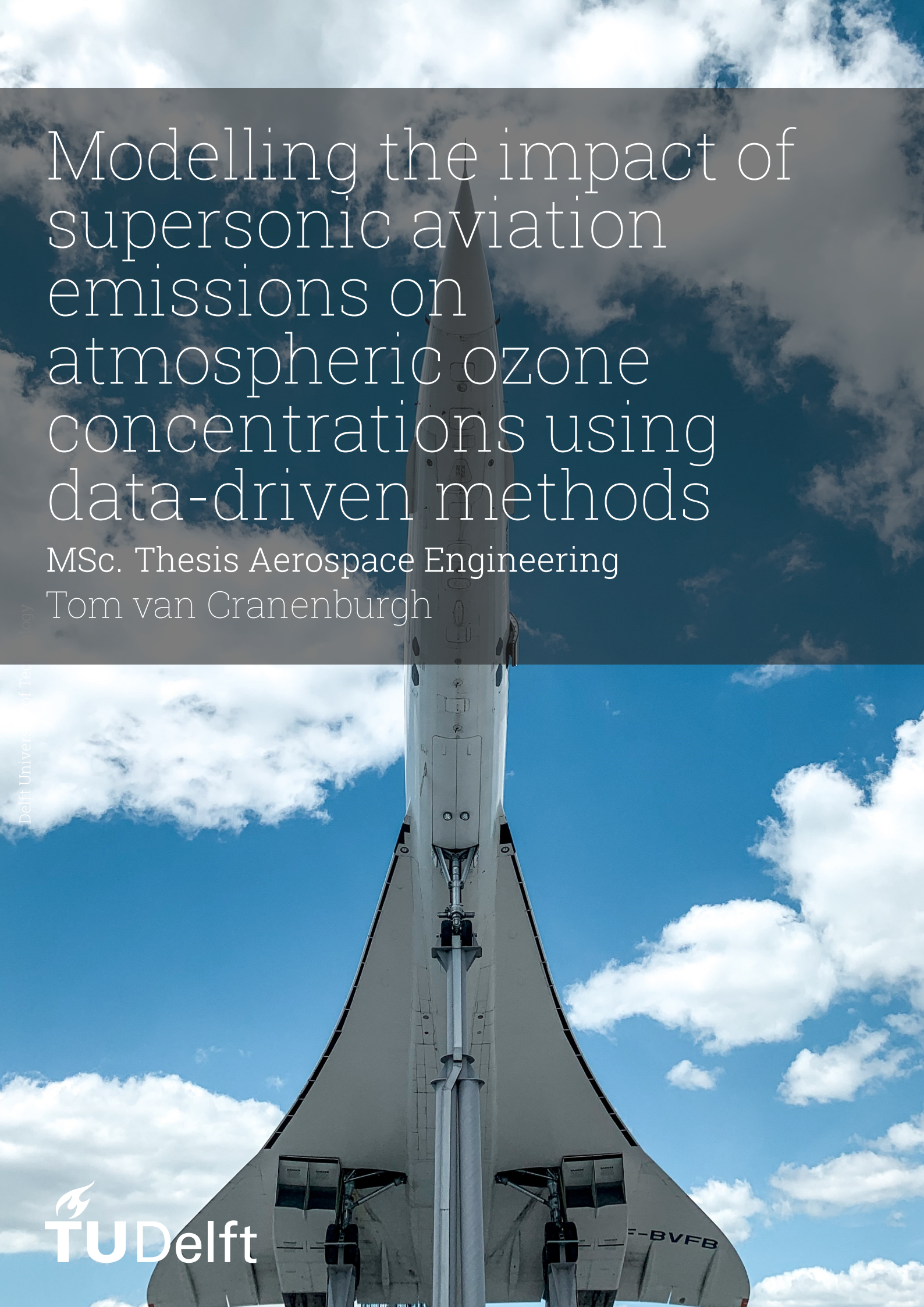
Modelling the impact of supersonic aviation emissions on atmospheric ozone concentrations using data-driven methods

MSc. Thesis Aerospace Engineering

Tom van Cranenburgh

Delft University of Technology



TU Delft

Modelling the impact of supersonic aviation emissions on atmospheric ozone concentrations using data-driven methods

by

Tom van Cranenburgh

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on March 27, 2024.

Student number:	4647645	
Project duration:	March, 2023 - March, 2024	
Thesis committee:	Prof. Dr. V. Grewe	TU Delft, Chair
	Dr. I.C. Dedoussi	TU Delft, supervisor
	ir. J.A. van 't Hoff	TU Delft, supervisor
	Dr. M. Lourenço Baptista	TU Delft, examiner
Collaborator:	Dr. U. Fasel	Imperial College London

Cover: Library, *Unsplash 2023*

TU Delft

Preface

This report is the final product of my thesis work to obtain a masters degree at the faculty of Aerospace Engineering at Delft University of Technology. Finalising this thesis marks the end of an incredible period in Delft and at the University. I am lucky to have met many amazing people and have made some unforgettable experiences. I am certain these years have fully equipped me for the future, and I am excited about the new opportunities that lie ahead!

I would like to express my gratitude to Jurriaan, Irene and Urban for their supervision during this project. First and foremost I would like to thank Jurriaan for helping me tackle this challenge, discussing together the many ways we could approach this problem and helping me think critically about all research steps and ways to communicate my work. I would also like to thank Urban, for your knowledge and your optimism, helping me understand the methods we looked at and always sharing your excitement towards this research. Last but not least I would like to thank Irene for all your thoughts and knowledge that have contributed to putting this work together. Together we managed to find an interesting approach to tackle this problem, combining all of your areas of expertise and realising very promising results!

I also want to thank my family and friends for their support. A very special thanks goes out to my girlfriend and my father, whom have helped me tremendously during this thesis and beyond. Thank you for always providing a listening ear and offering continuous support and motivation.

Enjoy the read!

Tom van Cranenburgh
Delft, March 2024

Summary

As civil supersonic aviation is likely to return in the near future, there is a need for increased understanding of the climate effects for mitigation and regulatory purposes. A large concern is the impact of emissions on ozone concentrations. Ozone formation and depletion is influenced by non-CO_2 emissions such as nitrogen oxides (NOx), water vapour (H_2O) and sulphate aerosols. The redistribution of chemical species as well as the conditions (amount of radiation influencing photolysis processes) also influence the ozone concentrations in the atmosphere. The emission location, specifically the altitude and latitude at which aircraft fly and emit, thus have a large impact the concentrations changes in ozone over time. Even though current research already addresses possible ozone depletion concerns for supersonic aviation emissions in the stratosphere, quantifying the effects for many individual emission scenarios is desired.

Changes in ozone concentrations as a result of supersonic aviation emissions are currently evaluated using computationally expensive chemistry transport models (CTMs). To evaluate atmospheric impacts for numerous emission scenarios faster methods will be required. In this research a first step is explored into building such computationally cheaper methods. A novel approach is taken by modelling the ozone response to supersonic aviation emissions as a data-driven dynamical system. We leverage dimensionality reduction techniques to capture dominant patterns in spatiotemporal atmospheric ozone data. Two objectives are thus: to enhance understanding of ozone response by analysing data-driven decompositions and to determine the extent to which data-driven dynamical systems are suitable in modelling ozone concentration changes.

Two methods are applied to monthly ozone concentration data that vary with altitude, latitude and time, obtained from GEOS-Chem simulations. Data in two flight regions, Transatlantic flight corridor (TAC) and Southern Arabian sea (SAS) region, and emission altitudes 16.2 km and 20.4 km are examined. Firstly, the proper orthogonal decomposition (POD) is applied to decompose the simulated ozone response data for a specific emission scenario into hierarchically ordered spatial modes and associated time coefficients. The hierarchical spatial modes show areas of high correlation in the data, some of which could be associated with known patterns of chemical species transport. The sparse identification of nonlinear dynamics (SINDy) algorithm is used to find a set of ordinary different equations (ODEs) describing the dynamics of the first four POD modes over time. Secondly, the dynamic mode decomposition (DMD) is applied to find a best fit linear operator that describes the evolution of the atmospheric ozone data as a dynamical system. The DMD decomposes the data into a set of spatial modes and associated frequency and growth/decay rate that describe how monthly ozone concentrations change linearly over time.

Results show the GEOS-Chem data can be reconstructed accurately with few POD and DMD modes. Analysing individual spatial modes obtained from DMD and POD analysis of different emission scenarios highlights dominant pattern differences in ozone response for emissions at different locations. Even though the spatial modes indicate areas of variability in the data, connections between observed patterns in the modes and atmospheric processes such as vertical mixing and Brewer-Dobson circulation causing poleward transport of species could be made. Reconstruction of results using DMD and POD-SINDy models reveal that especially DMD models accurately accurately reconstruct the monthly ozone concentrations over a period of several years. Furthermore, the dynamics of the DMD modes can be projected into a longer time period than the years used to fit the models, and show that DMD models accurately forecast ozone concentrations for future years. Nonlinear SINDy equations often fail to capture the exact months in which ozone peaks and lows occur, whereas DMD models accurately capture these.

We also show that the DMD models at different altitudes can be interpolated to predict monthly column ozone concentrations for emissions at an altitude that lies between these models. All in all, we

have shown that atmospheric ozone response to supersonic aviation emissions can be modelled using data-driven dynamical systems, providing a low dimensional representation. Analysis of decomposed atmospheric ozone data into spatial modes with nonlinear time coefficients, POD modes, and analysis of spatial modes that have same linear behaviour over time, provide extremely useful insights into differences between emissions scenarios and allow us to interpret the building blocks of the dynamical systems created to model the data. In the future these models can be used to get fast and accurate quantifications of aviation emissions effects.

Contents

Nomenclature

Acronyms

Acronyms	Definition
AIC	Akaike information criteria
CAEP	Committee on Aviation Environmental Protection
CTM	Chemistry transport model
DMD	Dynamic mode decomposition
DU	Dobson unit
EI	Emission index
EOF	Empirical orthogonal functions
ERF	Effective radiative forcing
GEOS-Chem	Goddard Earth Observing System chemistry model
ICAO	International Civil Aviation Organization
IPCC	Intergovernmental Panel on Climate Change
MAE	Mean absolute error
ML	Machine learning
ODE	Ordinary differential equation
PCA	Principal component analysis
POD	Proper orthogonal decomposition
RF	Radiative forcing
RMSE	Root mean square error
ROM	Reduced order modelling
RSS	Residual sum of squares
SAS	Southern Arabian sea flight region, defined as N15° to N24°
SINDy	Sparse identification of nonlinear dynamics algorithm
SST	Supersonic transport
SVD	Singular value decomposition
TAC	Transatlantic flight corridor region, defined as N48° to N57°
UT/LS	Upper troposphere/lower stratosphere
UV	Ultraviolet

Chemical symbols

Symbols	Definition
BC	Black carbon
CO	Carbon monoxide
CO_2	Carbon dioxide
CH_4	Methane
NO_x	Nitrogen oxides (NO + NO_2)
O_2	Oxygen
O_3	Ozone
SO_x	Sulfur oxides
SO_2	Sulfur dioxide

Introduction

This chapter introduces the motivation behind the thesis research, the research objective and sets the scope of this work, and includes an outline of the two documents found in this report.

Motivation

Running chemistry transport models to calculate chemical species concentrations over a large time period simulations is computationally expensive. In fact, calculating the concentrations of over three hundred chemical species over a period of 10 years using GEOS-Chem takes up to several weeks to run on a supercomputer. Even though these model provide an accurate representation and takes into account chemical interactions and transport, their computational time limits its application to evaluate atmospheric impacts for many different emission locations.

This motivates research into methods that can reduce dependency on chemistry transport models. The goal is to create models capable of predicting atmospheric impacts faster but with minimal loss in accuracy. To approach this problem many machine learning techniques could be applied, learning input and output relations from data to mimic the computations of the the GEOS-Chem model. However, in this research we look at modelling the concentration changes of ozone as a dynamical system, using dimensionality reduction and data-driven dynamical systems to model ozone concentration changes. The benefit of using such models is its interpretability. Compared to black box machine learning models that rely on deep learning techniques to learn relations, building dynamical models by decomposing data into spatial modes and using equations or frequencies and growth rates to describe the evolution of these modes, the building blocks of the model are visible and one can depict what the effect is of adding decomposed aspects of the data to the model.

Research objective and scope

The main research question that covers the scope of this research is as follows:

How can data-driven methods provide new insights into the impact of supersonic aviation emissions and be used to develop models that predict changes in atmospheric ozone?

A large benefit of these methods is that the building blocks of the model that is being constructed can be analysed, and that decomposition high dimensional spatiotemporal data into spatial modes and time coefficients has the ability to provide new insights. The research objective is thus two sided, trying to gain new insights and explore the extent to which data-driven methods are suitable for creating models that predicts atmospheric ozone changes. The two research questions are formulated as follows:

1. To what extent can data-driven methods enhance our understanding of the impact of supersonic aviation emissions on atmospheric ozone composition?
2. What data-driven techniques are effective for developing reduced-order models capable of accurately predicting changes in atmospheric ozone change along multiple dimensions?

If the methods are shown to be suitable for modelling ozone changes, attempting to create prediction models for other chemical species other chemical species of interest (H_2O, H_2O, BC) could be explored. Various species exhibit diverse atmospheric responses, encompassing differences in production and depletion rates, transport, and more. Consequently, certain methods may not be applicable to develop reduced-order models for all chemical species.

In this research ozone concentration changes and column ozone changes in Dobson Unit (DU) will be used as metrics. Examining other chemical species or the effect of mass concentrations changes on effective radiative forcing is beyond the scope of this work. The available data originates from the GEOS-Chem model. As will be discussed in later chapters, various climate transport models (CTM) exist and have been used to evaluate the effect of supersonic aviation emissions. This research does

not focus on validating the GEOS-Chem model or comparing results from various models for different scenarios, but instead focuses on using GEOS-Chem data to build a reduced order model capable of predicting effects of future scenarios.

Report outline

This report is compromised of two parts. Part I of this report contains the scientific paper that should be regarded as the final product of this master thesis. This paper includes an introduction to the problem, a description of the methodology, a discussion of the results and concluding remarks.

Part II of this report contains the Literature Study, a supplementary document previously graded under AE4020 that contains more elaborate background information. This document contains more detailed descriptions of other atmospheric impacts of supersonic aviation emissions, ozone chemistry in the stratosphere and troposphere, and a more elaborate description of the applied methods and an overview of current application of dimensionality reduction techniques and data-driven modelling.

Part I

Scientific Paper

List of Figures

Modelling the impact of supersonic aviation emissions on atmospheric ozone concentrations using data-driven methods

Tom van Cranenburgh
Supervisors: Dr. I.C. Dedoussi, J. A. van 't Hoff
Collaborator: Dr. U. Fasel (Imperial College London)

Faculty of Aerospace Engineering, Sustainable Air Transport
Delft University of Technology, The Netherlands

Abstract

With renewed interest in the development of civil supersonic aircraft, their return in the future is becoming more ever more likely. The environmental impact of emissions in the stratosphere on climate and the ozone layer therefore needs to be explored. The stratospheric ozone levels determine the amount of harmful ultraviolet radiation reaching the Earth's surface and thus the level of risk to human health and ecosystems. Ozone response is complex, varying with emission altitude and latitude and we are currently reliant on computationally expensive chemistry-transport models to calculate chemical species concentration changes resulting from supersonic aviation emissions. This paper takes a novel approach to reduce the dependency on these models, creating data-driven dynamical systems that model the global spatiotemporal atmospheric ozone response for different emission scenarios. The dynamic mode decomposition (DMD) and proper orthogonal decomposition (POD) methods are applied to atmospheric ozone data obtained from the GEOS-Chem model, and the evolution of the dominant POD spatial modes are modelled using sparse identification of nonlinear dynamics algorithm (SINDy). We show that DMD models can reconstruct monthly global column ozone changes with root mean square errors less than 0.05 Dobson unit (DU) for a period of three years. Predicting the global mean column ozone changes for the years beyond the period used to construct the models, results in errors less than 0.12 DU. Independent DMD models at two different altitudes can be interpolated to produce estimates for ozone response at an intermediate altitude. These methods can serve as a basis for low dimensional surrogate models that can be used to evaluate chemical species concentrations changes as a result of supersonic aviation emissions.

keywords: atmospheric chemistry modelling, supersonic aviation, non-CO_2 emissions, dimensionality reduction techniques, data-driven modelling, dynamic mode decomposition, proper orthogonal decomposition, sparse identification of nonlinear dynamics

1 Introduction

It is estimated that aviation emissions contribute up to 3.5 % of anthropogenic climate change [2]. While for most sectors carbon dioxide emissions (CO_2) are the most significant contributor, for aviation the CO_2 contribution amounts to only one-third of the radiative forcing and two-thirds of the radiative forcing is attributable to non-CO_2 emissions [2]. This estimate is based upon current aircraft fleets powered with kerosene fuel flying at altitudes in the range of 10 to 12 km. Amidst the challenge of making the aviation industry more sustainable and reducing the radiative forcing induced by aviation emissions, there has also been a renewed interest in the development of civil supersonic transport (SST) aircraft. These aircraft will drastically reduce travel time. Concurrently, the resulting atmospheric impact of supersonic aircraft is expected to be stronger due to their cruise altitudes in the range of 15-20 km altitude being much higher than conventional subsonic aviation [3].

Two notable SST's being developed for civil use are the Boom Overture and NASA's X-59. Boom Supersonic is developing the Boom Overture, a commercial passenger aircraft carrying up to 88 passengers, flying at a speed of Mach 1.7 and a cruise altitude of 18 km[1]. The aircraft will service a set of transoceanic routes and fly at subsonic speeds over land to minimise the nuisance of sonic booms. The Boom Overture is designed to use sustainable aviation fuel (SAF) to contribute to a net-zero carbon emission. However, even though SAF might reduce emissions of soot (black carbon) and have a net-zero CO_2 footprint, non-CO_2 emissions are still a concern [4]. NASA and Lockheed Martin are developing the X-59, a supersonic aircraft part of NASA's Quiet Supersonic Technology (QueSST)[2] mission. The aircraft is specifically designed to demonstrate and test technologies aimed at reducing the noise and impact of the sonic boom into sonic 'thump'. If successful, this can pave the way for overland civil supersonic air travel

[1] https://boomsupersonic.com/
[2] https://www.nasa.gov/mission/quesst/

in the future, something that is currently prohibited by regulations. With these two SST aircraft being in advanced stages of development, civil supersonic aviation is likely to return.

The increased cruise altitude of these aircraft increases the significance of non-CO_2 emissions compared to subsonic emissions, mainly due to the longer atmospheric lifetimes of emitted particles at higher altitudes [5]. The non-CO_2 emissions include nitrogen oxides (NO_x consisting of NO and NO_2), soot (black carbon, BC), water vapour (H_2O) and sulphurs [2]. The main concerns raised by the emissions of these chemicals at higher altitudes is the long-term change in global ozone distribution and its impact on climate [6, 7, 3, 4]. Changes in stratospheric ozone concentrations, influence the Earth's exposure to levels of ultraviolet (UV) radiation, with increased levels posing harmful threats to human health [8, 4].

The ozone concentration in the stratosphere is affected by the emissions of NO_x, H_2O and sulphate aerosols. NO_x emissions contribute to the catalytic ozone destruction in the middle- and upper stratosphere. In the lower stratosphere and upper troposphere, increased NO_x concentrations contribute to ozone production in the presence of HO_2 and sunlight [6]. H_2O acts as a catalyst in the destruction of ozone at higher altitudes [9]. Aerosols, black carbon (BC) known as soot and sulphate aerosols, have both a direct effect through absorption and scattering of radiation, and indirect effect influencing ozone concentrations [6]. Aerosols act as surfaces on which heterogeneous chemical reactions take place, facilitating the reaction of NO_x into HNO_3. The decrease of NO_x concentrations reduces O_3 depletion in the stratosphere [6]. Aerosols also scatter stratospheric UV radiation, lowering the rate of photo dissociation of O_2 and NO_2 in the lower stratosphere, which in turn are necessary for O_3 production [6]. Taking into account all of these processes is essential in achieving accurate predictions of atmospheric ozone concentrations changes.

The extent to which the supersonic non-CO_2 emissions influence ozone concentration changes is highly dependant on the residence time of the emitted species. The residence time is in turn dependant on circulation of air, the chemical interaction with other species and photolysis [9]. The redistribution of chemical species is dominated by the Brewer-Dobson circulation; an upward vertical mixing of species in the tropics and subsidence in the winter polar region that fluctuates seasonally [9]. The emission location, both altitude and latitude, impact the level of changes in ozone concentration. This has sparked interest in research into the effect of altitude and latitude, a quest to identify an ozone neutral altitude and calls for analysis into effects of individual flight scenarios [5, 10, 11, 12].

The high variability in ozone response, requires evaluation of concentration changes in ozone for many different individual emission locations. Faster methods than the current computationally expensive chemistry transport models will be required to asses these impacts for many emission scenarios. This requires surrogate models that are able to accurately and quickly evaluate atmospheric composition changes of species and climate effects for different supersonic flight scenarios (different emission altitudes and latitudes). This research is a first step in a possible alternative approach to modelling and predicting atmospheric ozone concentration changes as a result of supersonic aviation emissions. Complete understanding of the ozone concentration impact will be important for supersonic aircraft certification. Regulations regarding non-CO_2 emissions are not yet in place. Feedback from research into these effects can influence regulations and lead to re-evaluation of aircraft design and of proposed routes [5, 9, 12].

The proposed methodology to create such models is discovering dynamical systems from global spatiotemporal atmospheric chemistry data. This data is obtained from simulations using the chemistry transport model GEOS-Chem [13]. The proper orthogonal decomposition (POD) in combination with the sparse identification of nonlinear dynamics algorithm (SINDy) is applied to model the nonlinear evolution of dominant spatial modes. The second approach taken is the use of dynamic mode decomposition (DMD) to find spatially correlated structures with the same linear behaviour in time. Both methods extract low-dimensional features from data that can be used to construct reduced-order models (ROMs), which can serve for reconstruction and analysis of data, as well as a model that can predict states of the system beyond the time used to train the model. These methods currently have a large list of applications, but have been most predominantly used in fluid mechanics [14, 15, 16, 17, 18].

Similar to fluid mechanics applications, the atmospheric ozone data contains dominant patterns in both space and time. In the GEOS-Chem simulation, monthly ozone concentration changes stabilise after several years following the emission perturbation and show seasonal patterns [13]. The spatial patterns can be extracted using dimensionality reduction techniques, modelling their evolution in a linear manner (DMD) or nonlinear (POD-SINDy). These methods provide a novel approach to finding a mathematical framework that describes the ozone response to supersonic aviation emissions, learning these models purely from data.

A vast range of other methods have been applied to global spatiotemporal atmospheric chemistry data. Machine learning techniques such as auto encoders, random forest regression, neural networks have been implemented to replace chemical solver components of chemistry transport and climate models, to predict surface level species concentrations on an hourly or daily time scale [19, 20, 21, 22]. Dimensionality reduction methods applied to spatiotemporal atmospheric chemistry data, on full year simulation of surface species concentrations at latitude and longitude across time performed by Velegar et al, showed successfully extracting known major features of the data [23]. This research showed the benefits of dimensionality reduction

techniques on large spatiotemporal data sets, extracting dominant information that allows for recognising known patterns from the data and creating accurate approximations of the original data with a fraction of the memory. Yang et al. proposed surrogate modelling of atmospheric chemistry data using dimensionality reduction and SINDy to learn chemical dynamics [24]. This is an application of SINDy on known chemical dynamics interactions.

Dimensionality reduction and data-driven dynamical systems methods (POD, SINDy, and DMD) are employed in this study over other machine learning techniques for several reasons. Primarily, the interpretability offered by these methods is essential, allowing connections to be made to the physical and chemical processes inherent in atmospheric chemistry. This interpretability provides a clear advantage over 'black-box' machine learning approaches, ensuring a deeper understanding of the underlying dynamics. Additionally, the generalizability and reduced computational effort further contribute to the suitability of our chosen dimensionality reduction and data-driven dynamical systems methods for spatiotemporal atmospheric chemistry data.

Section 2 of this paper explains the methodology and the data used in more detail. Section 3 presents the results and a discussion. Section 4 includes concluding remarks and suggestions for future work.

2 Methodology

2.1 Data set

The data used is obtained from research carried out by van 't Hoff et al. [13]. This parametric study analysed the effect of emissions in two flight corridors at varying altitudes using the GEOS-Chem chemistry transport model. In this study six emission scenarios are examined, in two latitude regions and three different altitudes. The flight regions analysed are the transatlantic (TAC) flight corridor, and the southern Arabian sea (SAS) region. In the simulation the emitted species are introduced in vertical bands of GEOS-Chem's grid, with their central altitudes used as reference to the relevant emission altitude: 16.2 km, 18.3 km and 20.4 km. These acronyms and altitudes will be used as a naming convention in this research. For example, SAS162 represents the scenario of emissions in the southern Arabian sea region at 16.2 km. 8 Tg of annual fuel burn is introduced in the corresponding regions using estimated emission indices from recent works representing a possible future supersonic aircraft [13]. The resulting data contains monthly chemical species concentrations at a horizontal resolution of $4^\circ \times 5^\circ$ with 72 vertical pressure levels following a simulation in the years 2014-2023. Ozone changes across longitude are averaged in this research, as the variations with respect to latitude and altitude are more significant. Different time instances of the ozone response will be examined, as the initial years (2014-2018) show transient behaviour for ozone emissions in both regions and all altitudes. The stabilised response years (2019-2023) will be modelled as a dynamical system. Details of the exact atmospheric modelling procedure using GEOS-Chem is provided by van 't Hoff et al. [13].

2.2 Data preparation

To create a dynamical system, the temporal mean is subtracted from the original atmospheric chemistry data. In reconstruction and forecast of future atmospheric changes the mean is added to the system prediction. By subtracting the temporal mean, variations around the mean are isolated capturing the dynamic components of the data. The high dimensional atmospheric ozone data is then organised in a matrix of snapshots as follows.

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_m \\ | & | & & | \end{bmatrix} \tag{1}$$

The matrix $\mathbf{X}^{n\times m}$ represents the ozone concentration, with n representing the latitude by altitude dimension and m the number of months examined. By taking the average values across longitude and flattening latitude and altitude dimensions into a single column, the data can be organised in the format above.

2.3 Proper orthogonal decomposition

The proper orthogonal decomposition (POD) originates from applications in the field of fluid mechanics, and makes use of the singular value decomposition (SVD) in the context of spatiotemporal data [25, 26]. The SVD is a mathematical technique that can be applied to any matrix, decomposing it into singular values and orthogonal matrices [25]. The SVD applied to spatiotemporal data (known as the POD), returns dominant spatial patterns and their evolution across time. The spatial patterns are hierarchially ordered, making truncated POD approximations of the high-dimensional data using the first r dominant spatial modes straightforward. POD modes can be leveraged to create a reduced-order model (ROM) that allows for model simulations at significantly faster computational speeds. The as a basis SVD for applications in climate is often called empirical orthogonal functions (EOFs), in this paper reference will be made to the proper orthogonal decomposition as the spatiotemporal data is treated in a similar manner to POD applications on fluid mechanics. The POD seeks to find an approximation of the concentration of ozone C at spatial dimension x (both latitude and altitude) and time t:

$$\mathbf{C}(x,t) = \overline{\mathbf{C}}(x) + \sum_{k=1}^{r} \mathbf{a}_k(t)\psi_k(x) \tag{2}$$

Where $\overline{C}(x)$ represents the temporal mean. The SVD returns spatial modes $\psi_k(x)$. The time coefficients $\mathbf{a}_k(t)$ represent the evolution of the spatial modes and are found by projecting the spatial modes on the original data matrix.

2.4 Sparse identification of non-linear dynamics

The SINDy algorithm can be applied to discover governing equations that explain the dynamics of data [27, 1]. Applications of SINDy include discovering the equations of dynamical systems such as the Lorenz system. It is often assumed that the equations that govern these systems have a few active terms in right-hand side functions. As SINDy can be leveraged to find a sparse representation of the dynamics of multiple variables, we use it to model the dynamics of the POD time coefficients.

Results will show that the POD coefficients portray related behaviour, with several coefficients exhibiting oscillatory patterns at similar frequencies. As the variables are likely to be related to underlying chemical and transport process that vary seasonally, an attempt is made to model the coefficients as a coupled dynamical system. Many other methods exist to fit an equation on the evolution of the coefficients, but as the underlying dynamics exhibit some behaviour of a coupled dynamical system, the SINDy algorithm is appropriate.

The coefficients obtained from POD analysis on ozone concentrations responses, are not governed by known equations. The evolution of the coefficients are likely to vary for POD results obtained from different data sets (different time frames, different emission scenarios). SINDy attempts to find the most accurate sparse representation as an interpretable set of first order ordinary differential equations (ODEs). Even though the SINDy model might not be a perfect representation of the evolution of the coefficients, this method is a robust way to get accurate fits to describe the POD coefficients for multiple emission scenarios.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{a} = \mathbf{f}(\mathbf{a}) \tag{3}$$

The dynamics $\mathbf{f}$ of a dynamical system (POD coefficients $\mathbf{a}$) can be represented with minimum optimal number of components. SINDy seeks to find an approximation of the dynamics with a linear combination of non-zero coefficients Ξ and a library of nonlinear functions $\Theta(A)$:

$$\dot{\mathbf{A}} = \boldsymbol{\Theta}(\mathbf{A})\Xi \tag{4}$$

Two approaches are explored to robustify the SINDy algorithm and decrease its sensitivity to noise, a weak formulation and ensembling method. The weak SINDy formulation may be applied to noisy data, and was first presented by Reinbold et al. in the application of turbulent flow to discover the underlying nonlinear partial differential equations (PDEs) [27]. Ensemble-SINDy takes bootstraps of the time derivative of time coefficients matrix (with replacement), and returns the median of these coefficients as the otimal SINDy model [28]

To find a parsimonious model with an accurate fit, as few terms as possible in Ξ should be present. This solution is found by sequentially solving a least-squares problem using sequentially least thresholding (STLSQ) algorithm. This algorithm includes a sparsity-promoting threshold setting terms below the threshold to zero.

To evaluate the best SINDy fit model for a range of threshold values λ, the Akaike information criteria (AIC) is used [29]. This criteria was developed for model selection of nonlinear dynamical systems whilst promoting sparsity by adding a penalty on the number of terms included in the solution while ensuring an optimal fit. A sparse representation of the data is desired to get an interpretable description of the dynamics of the system. The minimum AIC value will return the sparsest optimal fit. The AIC is calculated using the residual sum of squares (RSS) between the actual POD time coefficients and the simulated SINDy coefficients as likelihood function.

$$\mathrm{AIC} = m\ln(\mathrm{RSS}/m) + 2k \tag{5}$$

Where m is the number of data points and k is the number of parameters used in the candidate SINDy fit model. For a finite sample size the AIC scores require the following correction.

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2(k+1)(k+2)}{(m-k-2)} \tag{6}$$

Two pre-processing steps are taken on the time-series coefficients obtained from the POD analysis. With the limited number of data points and a turbulent pattern recovered (especially for higher order modes), we found that the performance of the SINDy model could be improved by taking the following pre-processing steps. Firstly, the time series data is smoothed by taking the moving average of three consecutive data points. Secondly, the data is interpolated using quadratic interpolation to increase the number of data points. These two steps enable the creation of more data points, getting a more accurate pattern of the dynamics of the POD coefficients to improve the SINDy model fit.

2.5 POD-SINDy

Figure 1 includes a schematic of the full POD-SINDy methodology. During step 3), both Ensemble-SINDy and the weak formulation of SINDy are explored for a range of threshold values of the STLSQ optimiser algorithm. The minimum AIC of the full set of possible models obtained from ensembling and the weak formulation solutions is used. These methods are incorporated in the Python package PySINDy [30, 31]. In this manner, a robust method is created that is able to fit SINDy models on time coefficients obtained for different emissions scenarios and different subsets (years of simulation) of the data.

2.6 Dynamic mode decomposition

The dynamic mode decomposition (DMD) decomposes high dimensional spatiotemporal data into spatial structures, each with an associated frequency of oscillation and rate of growth/decay. The DMD returns

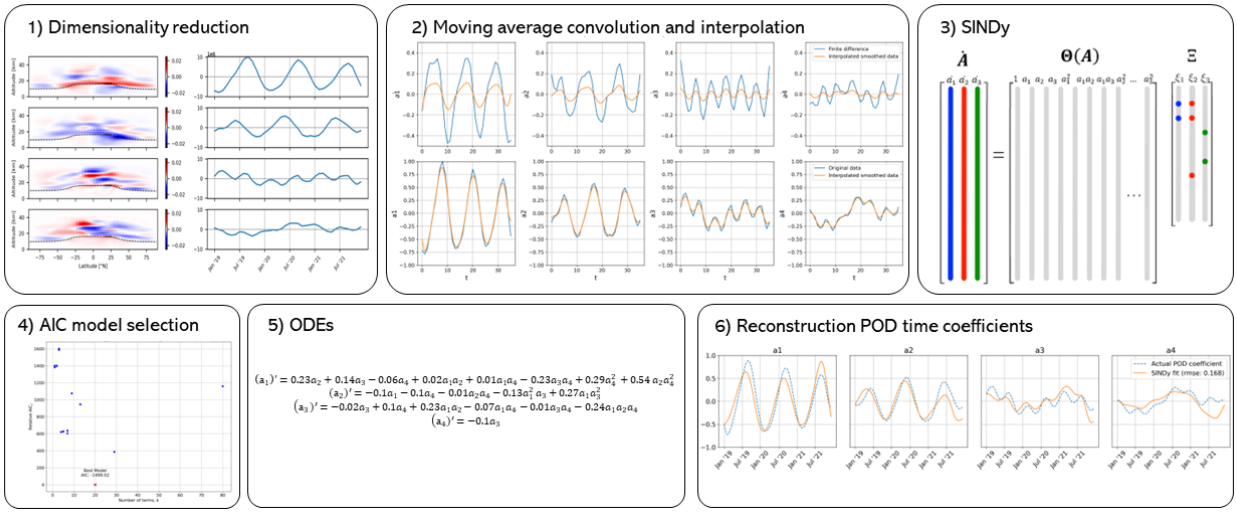


Figure 1: POD-SINDy method schematic. 1) Dimensionality reduction using POD. 2) Perform moving average convolution and interpolation on POD coefficients to reduce derivative extremes and create more data points. Lower plot shows interpolated smoothed time coefficients, upper plot shows derivatives. 3) Ensemble and weak formulation of SINDy using library of polynomial terms up to order three for a range of threshold values in optimiser, illustration adapted from Brunton et al. [1]. 4) Optimal model selection using AIC criteria. 5) Optimal model ODEs. 6) ODEs integrated to reconstruct POD time coefficients.

modes with same linear behaviour in time whereas the POD returns hierarchically ordered spatial structures based on the variance captured by a mode and an associated evolution of these modes in time.

To compute the DMD, similarly to the POD, a matrix of snapshots is arranged as matrices $\mathbf{X}$ and $\mathbf{X}'$, with matrix $\mathbf{X}'$ omitting the first snapshot and including the final snapshot of the original data.

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathrm{x}(t_1) & \mathrm{x}(t_2) & \cdots & \mathrm{x}(t_{m-1}) \\ | & | & & | \end{bmatrix} \tag{7}$$

$$\mathbf{X}' = \begin{bmatrix} | & | & & | \\ \mathrm{x}(t_2) & \mathrm{x}(t_3) & \cdots & \mathrm{x}(t_m) \\ | & | & & | \end{bmatrix} \tag{8}$$

The DMD algorithm finds the best-fit linear operator $\mathbf{A}$ that relates these snapshot matrices.

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X} \tag{9}$$

DMD steps:

1. Compute the truncated singular value decomposition of matrix. Where $\tilde{\mathbf{U}} \in \mathbb{C}^{n\times r}$ and $\mathbf{V} \in \mathbb{C}^{m\times r}$ and $\Sigma \in \mathbb{C}^{r\times r}$, dependant on the chosen rank r.
$$\mathbf{X} \approx \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{V}^* \tag{10}$$

2. Matrix $\mathbf{A}$ can be obtained by computing the pseudo inverse of $\mathbf{X}$. By projecting $\mathbf{A}$ on the POD modes $\mathbf{U}$, the full matrix $\mathbf{A}$ does not have to be computed:
$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}^*\mathbf{A}\tilde{\mathbf{U}} = \tilde{\mathbf{U}}^* X'\tilde{\mathbf{V}}\tilde{\Sigma}^{-1} \tag{11}$$

3. The eigenvalues $\mathbf{\Lambda}$ of the reduced matrix $\tilde{\mathbf{A}}$ are the same as the eigenvalues of $\mathbf{A}$. The eigendecomposition of $\tilde{\mathbf{A}}$ is performed:
$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda} \tag{12}$$

4. The DMD modes $\mathbf{\Phi}$ are reconstructed with the eigenvectors $\mathbf{W}$ and the time-shifted snapshot matrix of the original data $\mathbf{X}'$.
$$\mathbf{\Phi} = \mathbf{X}'\mathbf{V}\tilde{\Sigma}^{-1}\mathbf{W} \tag{13}$$

To robustify the DMD algorithm and decrease the influence of noisy data on recognising dynamics, an optimised version of the DMD algorithm is used that was proposed by Askham and Kutz [32]. This extension was improved upon in the bagging optimised DMD (BOP-DMD) version presented by Sashidhar and Kutz [33]. This method, incorporated into the open source Python package PyDMD, allows for constraining eigenvalues to be stable [34, 35]. In the context of atmospheric ozone changes, the concentration deviations from the mean are expected to oscillate seasonally due to photolysis and circulation of species that influence on production and depletion rates [13]. Therefore, the optimised DMD method, enforcing stable eigenvalues, can be used to fit a model that allows for stable predictions of ozone concentrations for many time steps ahead. This allows forecasting of ozone concentration values beyond the time frame of the data used to fit the DMD model.

2.7 Evaluating ozone impact

Climate metrics, such as effective radiative forcing, are not touched upon in the research but concentration

changes serve as a basis for calculating radiative forcing. To evaluate the effectiveness of the proposed methods in reconstructing the ozone concentration changes obtained from GEOS-Chem, the ozone impact across different dimensions is analysed. Ozone mass changes (ΔO_3 [kg]) at varying latitude and altitude are examined across time. We convert this to changes in column ozone in Dobson units (DU), a valueable metric for evaluating ozone impact at a spatial location, representing the total mass of ozone in a vertical column. Column ozone values at varying latitudes across time are examined, and mean global column ozone values across time are analysed. The evaluations will encompass monthly instances, capturing both transient perturbed behaviour and stabilised seasonal differences observed in the final years of simulations. The mean global ozone column change is used in other studies to quantify the overall ozone impact [4, 3]. The impacts of atmospheric ozone due to supersonic aviation are mainly influenced by circulation patterns and photolysis, and as variations with longitude is not discussed in previous research this dimension is omitted from analysis and reconstruction of data [3, 5, 6, 4].

3 Results and discussion

3.1 POD modal analysis

Drawing definitive conclusions regarding the observed patterns in POD spatial modes and its relation to atmospheric ozone response processes is challenging and somewhat arbitrary. It is important to note that these modes represent areas of high correlation in space across the dataset. Even though recognisable patterns emerge that can be associated with known atmospheric chemistry processes or species transport, drawing conclusive interpretations can be challenging. However, analysing the spatial modes and their time coefficients, and comparing these for the different emission scenarios and different time subsets can improve understanding of the actual atmospheric ozone response to the emissions. Caution should be taken in attributing specific chemical or transport phenomena solely based on spatial modes, and this analysis shows how POD modes reveal that areas of high correlation could be related to physical and chemical processes.

Applying the POD on the full 10-year data of ozone concentration changes, along latitude and altitude dimension across time, results in the spatial modes and their temporal evolution depicted in Figure 2. The first four dominant POD modes for the SAS162 scenarios are shown. The temporal mean ozone change and the results for other emission scenarios is included in Appendix A. In this analysis data from all years of simulation are included to illustrate the transient response as well as the stabilised ozone response, which is best visible by examining the time coefficient data of mode 1 in the first years.

Inspecting the dominant spatial structures in the left figures in Figure 2, mode 1 shows the general ozone response to aviation emissions in the stratosphere. In the lower stratosphere there is an increase in ozone concentrations and in the upper stratosphere a decrease. The corresponding time coefficients indicate an initial transient response, stabilising to a seasonal variation with peaks between the months July and January, and lows between February and June. The tropopause acts as a barrier which is also visible in mode 1, showing little ozone concentration changes in the tropopause region. Mode 1 also represents a clear separation of varying altitudes above which ozone depletion is dominant, and under which ozone production is dominant. The spatial modes and coefficients can be multiplied and added to the temporal mean to reconstruct the data. This explains the large negative time coefficient in the month Jan 2014, as the contribution of mode 1 will cancel a large part of the temporal mean pattern. The evolution of spatial mode 1 accurately captures the transient behaviour and a stabilised seasonal pattern in the years from 2019 onwards. The dominant pattern of ozone depletion at higher altitudes and ozone production at lower altitudes as a result of supersonic aviation emissions is described in other research [5, 6].

The spatial pattern of mode 2 varies also varies seasonally and shows two recognised transport pathways of chemical species. Firstly the blue region in the southern polar region is indicate as a region of large ozone variation from the mean ozone change. This could be associated with the Brewer-Dobson circulation that explains stratospheric transport from the tropics to the winter pole. The enhanced redistribution to the southern hemisphere for emissions in the SAS region, is captured by mode 2, and is less visible in the spatial modes of the TAC emission scenarios (see Appendix A. Secondly, vertical mixing in the mid latitudes is visible by the blue region in the region below the tropopause in mode 2. The blue and red regions, coupled with a time coefficient that changes seasonally from positive to negative shows a region of peaks and lows in ozone concentration. This could be explained by the downward transport in the mid latitudes, which is also visible in mode 3. This region of high correlation related to vertical mixing is not clearly captured by the first four dominant spatial modes of the results from emissions at higher altitudes (see spatial modes in Appendix A), indicating that this phenomena is less influential for ozone response when emitting at higher altitudes in the stratosphere.

Despite the challenge associated to relating spatial correlated structures to atmospheric processes, inspecting modes is still a valuable approach in increasing understanding of the impact on ozone of aviation emissions at different altitudes and in different regions. The first four modes from a POD analysis on scenarios SAS162, SAS204, TAC162 and TAC204 is included in Appendix A, and the first four modes from a POD analysis examining only the years 2019 to 2021 is included in Appendix B. The POD coefficients used from these years of simulation will be used to construct SINDy models.

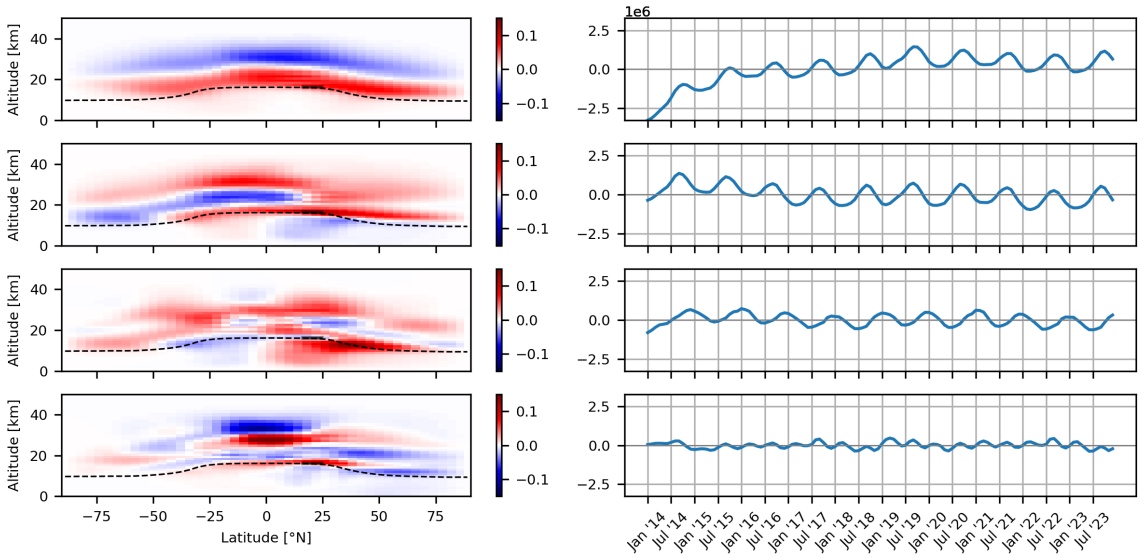


Figure 2: First four dominant POD modes SAS162 scenario. Left: spatial modes, grey rectangle indicates emission region and dotted line indicates mean tropopause height. Right: time coefficients indicating how spatial modes evolve over time.

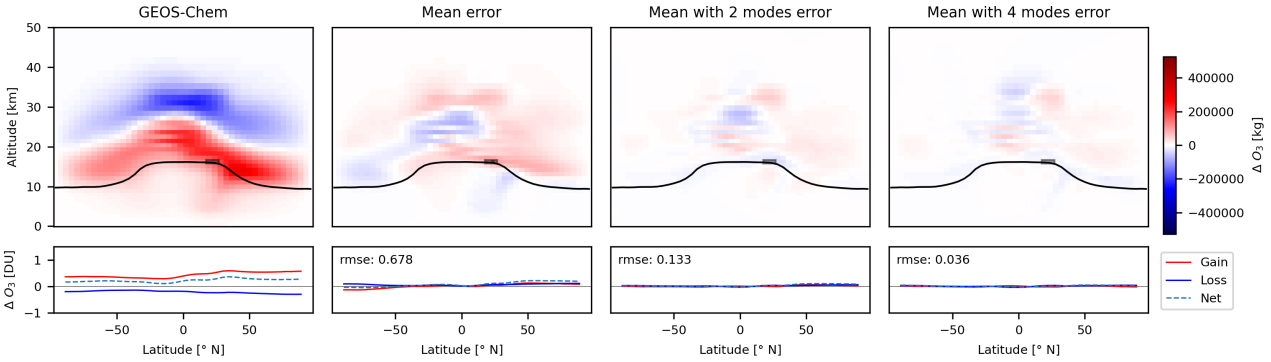


Figure 3: Ozone mass change reconstruction errors using only the mean, mean with first 2 modes and mean with first 4 modes for the SAS162 scenario March 2023. Changes in mean column ozone are shown below the zonal averages including the root mean square error (rmse).

3.2 POD reconstruction

Figure 3 shows the errors in reconstruction of the SAS162 data for March 2023. This example illustrates for a single month how the original ozone response can be reconstructed accurately with the mean response and the first four dominant modes. The rightmost figure in Figure 3 shows the error when reconstructing the data using the mean with the first four modes. The net column ozone change reconstruction across latitude (bottom right graph), results in a reconstruction with a root mean square error (RMSE) of 0.036 DU. The reconstruction with solely the mean resulted in an error of 0.678 DU, highlighting the benefit of reconstructing the data with first four modes. This creates an incentive to attempt to model the dynamics of the first 2, 3 or 4 POD modes as a coupled dynamical system using SINDy.

The first four modes together capture just above 50% of the variability of the data, as is indicated in Figure 10 in Appendix A. Even though this seems rather low, dominant patterns are covered by first dominant modes which allow for low dimensional reconstructions of the original data.

3.3 Modelling evolution of POD modes using SINDy

Finding a way to model the evolution of a set of POD modes (2, 3 or 4 modes) will aid in describing the dynamics, and creating a system that is able to predict values beyond the time length used to train the model. The time coefficients exhibit behaviour at a similar frequency, thus SINDy could be an effective manner of modelling the evolution of the system. The years 2019-2021 are used to fit a SINDy model on, leaving the years 2022-2023 as validation data.

Figure 4 shows the best ensemble fit for the first three POD coefficients. The resulting ODEs contain 20 terms to describe the system. Even though this is not a sparse system with few terms describing the dynamics, inspecting the visual reconstruction and the

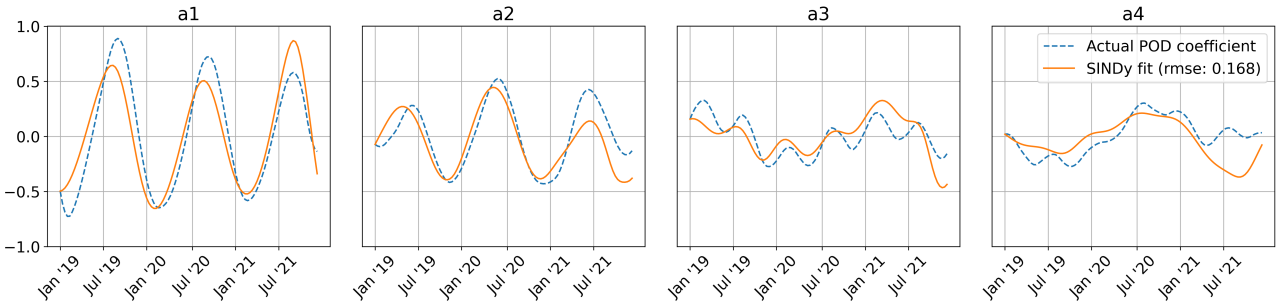


Figure 4: Best fit ensemble SINDy model on first four POD mode coefficients obtained from analysis on SAS162 emissions from year 2019-2021, solved using STLSQ with threshold 0.04. The ODEs contain 20 terms in total, results in a root mean square error (RMSE) of 0.168.

rmse of 0.168 suggest a reasonable representation of the actual POD time coefficients. As will become clear in section 3.5, when integrating these ODEs over a longer period of time, there is a mismatch in the peaks of the actual POD time coefficients and those predicted by the SINDy model. This is visible in the reconstruction of the a_2 time coefficient in Figure 4. Examining the SINDy models fits on the first four modes for the other emission scenarios, also results in representations with ODEs containing a total of more than 20 terms. These results are presented in Appendix C. Even though equations with such many terms is not a 'sparse' representation of the dynamics, inspecting the results visually shows a fit capturing the most significant patterns.

The effectiveness of this POD-SINDy method in reconstructing ozone concentrations and predicting future values are covered in section 3.5. The ODEs obtained from SINDy are integrated over a larger time period than the time used to train the mode, with the same initial condition to predict future values.

3.4 DMD modal analysis

In capturing a linear best fit model of the ozone response evolution using DMD, similar to the POD-SINDy years 2019 to 2021 is used for analysis. The optimised DMD is used enforcing stable eigenvalues.

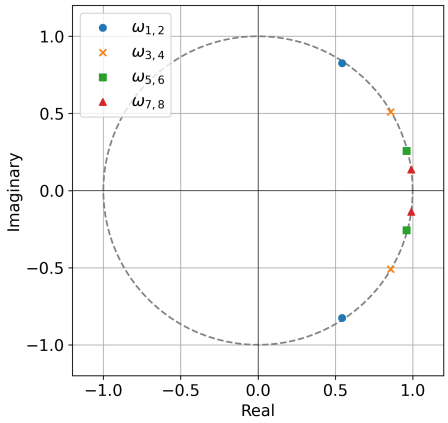


Figure 5: Discrete-time eigenvalues associated with DMD modes SAS162 scenario. As the system exhibits oscillatory behaviour, the eigendecomposition of the operator matrix contains complex conjugate pairs, resulting in paired modes.

Interpreting the spatial modes and eigenvalues in Figure 6 is more ambiguous than interpreting the spatial modes from POD analysis. The DMD finds spatial patterns that have the same frequency over time, and coupling spatial modes and their eigenvalues to chemical or transport processes is more open to different interpretations. Figure 5 shows the discrete-time eigenvalues that are projected onto the time frame on which the DMD model is trained in Figure 6. Modes 3 and 4 in Figure 6 show peaks and lows every 6 months. The DMD analysis for these years of other scenarios is included in Appendix D, and all show a half-year seasonal pattern in modes 3 and 4. Comparing these modes highlights some expected differences between ozone response to emissions in certain regions.

The DMD spatial modes and eigenvalues, lend itself to construct a prediction model as they indicate the dynamics from one time step to the next. This allows DMD model to reconstruct data for the original years on which the model was fit and also forecast for future years, these results are shown in section 3.5.

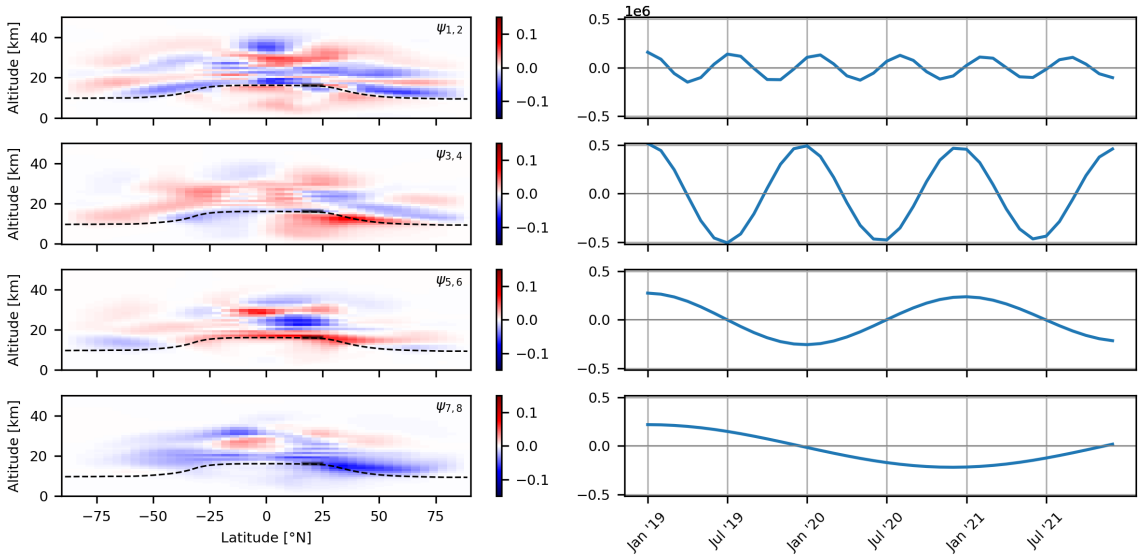


Figure 6: First eight DMD modes SAS162 scenario from analysing years 2019-2021 and their project time dynamics. Note that the modes are ordered according to the frequency of the mode, not the percentage of the variance a mode represents.

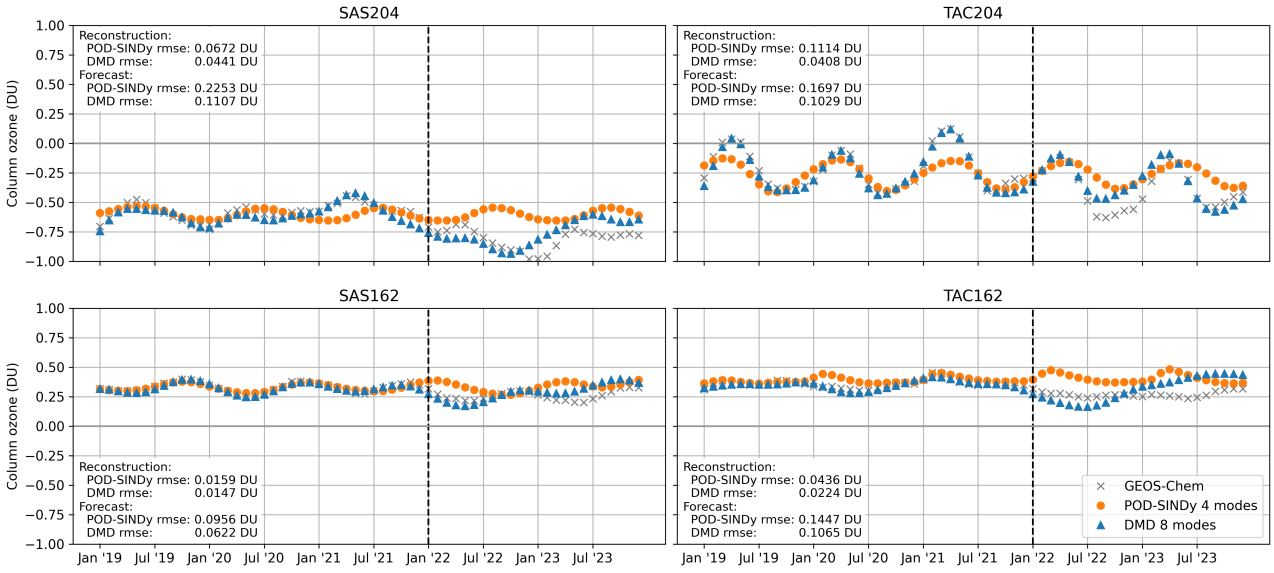


Figure 7: Global mean column ozone reconstruction and forecast for four emissions scenarios. Vertical dotted line separates reconstruction data (2019-2021) used to fit models and forecast data (2022-2023). Root mean square errors (rmse) of reconstruction and forecast period for POD-SINDy with 4 modes model and DMD with 8 modes model.

3.5 Global mean column ozone reconstruction and forecast

The ozone concentration data is reconstructed for the years on which the data is fitted (2019-2021) and forecasted for the years 2022-2023. The monthly global mean column ozone values per month in DU for all four scenarios using the DMD and POD-SINDy method are shown in Figure 7.

These results show that these methods are effective in reconstructing the high dimensional data. Both POD-SINDy, a combination of four spatial modes and four ODE's describing its evolution and the DMD, a set of eight spatial modes with an associated frequency and growth or decay rate, prove suitable methods to reconstruct atmospheric ozone data. Leveraging dimensionality reduction to capture dominant patterns and learning a dynamical system from data has proven to be effective in this application.

Examining the differences between POD-SINDy and DMD, all DMD reconstruction and forecasted monthly column ozone changes result in lower average root mean square errors than the POD-SINDy values. This is most likely due to the monthly global mean column ozone peaks and lows appearing seasonally for all emissions scenarios. For instance, examining the SAS162 results in Figure 7, the POD-SINDy forecast for years 2022-2023 shows a mismatch in the months where the actual GEOS-Chem data shows peaks and lows in global mean column ozone.THEWhere DMD modes have a specific frequency that do not change when predicting future timesteps, not influencing the position of global mean column ozone peaks and lows. The nonlinear ODE's integrated for a longer time period to form time coefficients result in a less regular pattern and not matching the actual POD time coefficients as well. This then results in a mismatch in global mean column ozone values for the POD-SINDy predicted values. A linear best fit model with a dominant mode oscillating seasonally together with other modes oscillating at higher or lower frequencies combines to create the most accurate ozone concentration reconstructions.

Reducing the dimension further to mean global column ozone change over the entire reconstruction and forecast period, a metric that illustrates overall ozone response of an emissions scenario, is represented in Table 1. Analysing the global mean column ozone values shows the accuracy for ozone changes predictions over a time period. The root mean squares errors provide a more accurate validation metric of the models, but these zero dimensional values (global mean column ozone values for several years) allows for quick interpretation of the effectiveness of models. The values in Table 1 show how well the models reconstruct values that might be of interest to policy makers or for comparison of the effects emissions at a specific region. For example, the global mean column ozone reconstructions for 2019-2021 of the SAS204 scenario, show a value of -0.596 DU for the POD-SINDy reconstruction and -0.601 for the DMD reconstruction. Despite the RMSE of the DMD reconstruction being lower, the global mean column ozone value of the POD-SINDy reconstruction is closer to the actual GEOS-Chem value. Examining the data points in Figure 7 reveals the reasons for this. The POD-SINDy column ozone values deviate less from the mean ozone value over time, and do not capture the peak ozone value in 2021 as well as the DMD model, yet the POD-SINDy mean value of these three years is closer to the GEOS-Chem simulated value than the DMD value. Analysing multiple metrics to evaluate model effectiveness is thus crucial in this case.

This also highlights the influence of the temporal mean in the reconstruction of the original data. When evaluating the effectiveness of these models we are evaluating their effectiveness in capturing deviations from the mean ozone change. Comparing the mean global column ozone values presented in Table 1, a large portion is represented by the mean ozone change, the second image shown in Figure 3. Therefore reducing down to this zero dimensional value is an interesting final step in evaluating the effectiveness of the models. But analysing the data at higher dimensions proves more effective in evaluating their accuracy in reconstruction and forecasting of the original data.

Furthermore, the models are fitted on GEOS-Chem data and evaluated against this data, which do not exhibit the behaviour of a perfect dynamical system. In fact, the GEOS-Chem data are computed based on many input variables including meteorology. Identifying the outliers in the original data biases the fitting or evaluation of the POD-SINDy and DMD models. Figure 7 shows a year with significantly more ozone depletion than previous years. This is illustrated by the GEOS-Chem data in the month Jan '23 of the SAS204 scenario. The significantly larger ozone concentration decrease in this year could be caused by different meteorological conditions, with an increased solar flux in these months that enhanced ozone depletion processes. When evaluating the errors of the model forecasts with the GEOS-Chem data this should be taken into account.

All in all, these methods show that accurate reconstruction of atmospheric ozone data can be done using these data-driven methods, and that dynamic mode decomposition models most accurately reconstruct and predict monthly global mean column ozone values. It is extremely valuable to able to represent ozone concentration changes as a dynamical system that consists of several spatial modes and an associated eigenvalue. This is a much simpler representation than the high dimensional and memomry intensive GEOS-Chem data.

3.6 Predicting ozone concentration changes at other altitude

The goal of this research is focused around finding methods that can reduce the dependancy on computationally expensive chemistry transport models like GEOS-Chem to predict chemical species concentrations changes as a result of aviation emissions. Ideally,

Table 1: Global mean column ozone changes in Dobson Units (DU) per month over the period used to construct models (2019-2021) and forecast period (2022-2023). The root mean square error of the monthly column ozone values is included.

		Global mean column ozone (DU)			Root mean square error (DU)	
		GEOS-Chem	POD-SINDy	DMD	POD-SINDy	DMD
reconstruction 2019-2021	SAS162	0.330	0.328	0.325	0.016	0.015
	TAC162	0.360	0.390	0.351	0.044	0.022
	SAS204	-0.579	-0.596	-0.601	0.067	0.044
	TAC204	-0.217	-0.265	-0.235	0.111	0.041
forecast 2022-2023	SAS162	0.262	0.339	0.288	0.096	0.062
	TAC162	0.272	0.408	0.314	0.145	0.107
	SAS204-	-0.808	-0.605	-0.761	0.225	0.111
	TAC204	-0.387	-0.265	-0.334	0.170	0.103

a surrogate model is present that is able to predict such changes for specific emission inputs in minutes or even seconds instead of weeks. In this case, we are interested in the ozone response to emissions at specific altitude and latitude. Here we attempt to predict the ozone response at an altitude between altitudes for which a surrogate DMD models has been created. Using the DMD model obtained for the 20.4 km altitude scenario and the 16.2 km altitude scenario we investigate whether these models serve as a basis for predicting another altitude for which the actual ozone response is known, emissions at 18.3 km.

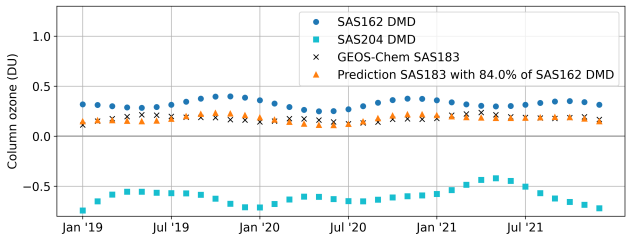


Figure 8: Interpolating between DMD solutions for altitudes 20.4 and 16.2 km to reconstruct 18.3 km scenario results in SAS region, prediction rmse is 0.035.

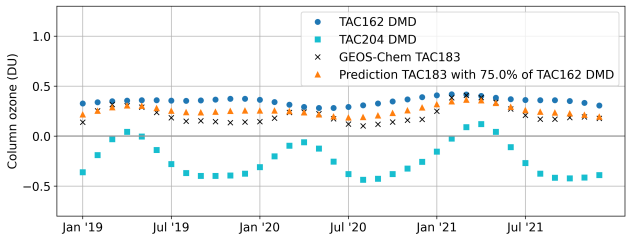


Figure 9: Interpolating between DMD solutions for altitudes 20.4 and 16.2 km to reconstruct 18.3 km scenario results in TAC region, prediction rmse is 0.037.

Figure 8 shows that to get the optimal fit for SAS183 global mean column ozone for the years 2019-2021, 84% of the SAS162 DMD model should be taken and 16% of the SAS204. A similar approach was taken for emissions in the TAC region, and reconstruction of the TAC183 results reconstruction is best achieved combining 75% of the TAC162 DMD model and 25% of the TAC204 DMD model, shown in Figure 9. This insight indicates that there is possibility to combine surrogate models trained on detailed atmospheric chemistry data from GEOS-Chem to predict the ozone response at another altitude. This demonstration is a first step into potential use of these models in the future, and how they can contribute to overcoming the computational intensity of chemistry transport models used to generate these results.

Although the interpolation ratio between the altitudes was computed by minimising the error with respect to the actual GEOS-Chem data, this result demonstrates the potential of combining sparse models to predict monthly ozone concentration changes for unknown emission locations. The DMD modes are thus a practical application for the next step in this research.

4 Concluding remarks and recommendations

We have shown that POD-SINDy and DMD can be used as tools for analysis and for surrogate modelling of the ozone response to supersonic aviation emissions. Examining the decomposition of data into spatial modes and time coefficients using two different approaches, provides insights into the differences in ozone response for different emission scenarios. Furthermore, monthly global mean column ozone values can be reconstructed with several modes with root mean square errors less than 0.05 DU.

Reducing dimensionality using the POD and modelling the evolution of these modes using SINDy provides an accurate reconstruction for emission scenarios, but SINDy is not accurate in predicting the time coefficients for all scenarios. This leads to unstable models when predicting the value of POD coefficients for future timesteps, leading to errors in the reconstruction and forecast of global mean column ozone values. The dynamic mode decomposition (DMD) is a more appropriate method to model monthly global mean column values. This returns a stable model that reconstructs and predicts the seasonal changes in ozone concentration accurately. The DMD model with stable eigenvalues serves as an accurate basis for reconstructing data and predicting future time steps, predicting monthly global mean column ozone values for years 2019-2023 (years exhibiting a stabilised ozone response to avia-

tion emissions) with root mean square errors less than 0.12 DU for all scenarios.

We have also shown that it is possible to combine the DMD models at altitudes 16.2 km and 20.4 km to reconstruct monthly global mean column ozone values for emissions at 18.3 km. This shows there is a basis for combining reduced-order models to predict the atmospheric impact of other emission scenarios, but future work would need to be done to identify interpolation ratios. By analysing the sensitivity of supersonic aviation emissions on ozone with altitude, an approximation can be made for how ozone response changes when emissions altitude changes. However, this result promises possible reduced dependency on computationally expensive chemistry transport models to predict changes atmospheric ozone concentrations at other emission locations.

The results in this study demonstrate that data-driven dynamical systems can be used to model the impact of aviation emissions on ozone concentrations. Therefore a useful next step could be to model the impact of aviation emissions on other chemical species concentrations, such as water vapour or sulphur oxides, using dynamic mode decomposition. These chemical species also exhibit seasonal concentration changes, which makes using dimensionality reduction methods to capture dominant spatial patterns with associated oscillatory behaviour over time suitable.

These methods are still only limited to capturing the stabilised ozone response. Capturing the transient perturbed response in ozone concentrations following supersonic aviation emissions (first years of data) requires a different approach. To improve the accuracy of the DMD reconstructions and forecast, an alternative approach could be to model the system using dynamic mode decomposition with control. Using a control parameter that resembles meteorological conditions at a specific time step, might improve the reconstruction of global mean column ozone values for specific months.

The application of data-driven methods on atmospheric ozone chemistry data is valuable to gain new insights into underlying dominant patterns in the data and to model the changes in zone concentrations as a simplified dynamical system.

References

[1] S. L. Brunton, J. L. Proctor, J. N. Kutz, and W. Bialek, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 15, pp. 3932–3937, 4 2016.

[2] D. S. Lee, D. W. Fahey, A. Skowron, M. R. Allen, U. Burkhardt, Q. Chen, S. J. Doherty, S. Freeman, P. M. Forster, J. Fuglestvedt, A. Gettelman, R. R. De León, L. L. Lim, M. T. Lund, R. J. Millar, B. Owen, J. E. Penner, G. Pitari, M. J. Prather, R. Sausen, and L. J. Wilcox, "The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018," *Atmospheric Environment*, vol. 244, 1 2021.

[3] R. L. Speth, S. D. Eastham, P. Prashanth, and F. Allroggen, "Global Environmental Impact of Supersonic Cruise Aircraft in the Stratosphere," Massachusetts Institute of Technology, Cambridge, Massachusetts, Tech. Rep., 2021.

[4] S. D. Eastham, T. Fritz, I. Sanz-Morère, P. Prashanth, F. Allroggen, R. G. Prinn, R. L. Speth, and S. R. Barrett, "Impacts of a near-future supersonic aircraft fleet on atmospheric composition and climate," *Environmental Science: Atmospheres*, vol. 2, no. 3, pp. 388–403, 2 2022.

[5] J. Zhang, D. Wuebbles, D. Kinnison, and S. L. Baughcum, "Potential Impacts of Supersonic Aircraft Emissions on Ozone and Resulting Forcing on Climate: An Update on Historical Analysis," *Journal of Geophysical Research: Atmospheres*, vol. 126, no. 6, 3 2021.

[6] S. Matthes, D. S. Lee, R. R. De Leon, L. Lim, B. Owen, A. Skowron, R. N. Thor, and E. Terrenoire, "Review: The Effects of Supersonic Aviation on Ozone and Climate," *Aerospace 2022, Vol. 9, Page 41*, vol. 9, no. 1, p. 41, 1 2022.

[7] V. Grewe, A. Stenke, M. Ponater, R. Sausen, G. Pitari, D. Iachetti, H. Rogers, O. Dessens, J. Pyle, I. S. A. Isaksen, L. Gulstad, O. A. Søvde, C. Marizy, and E. Pascuillo, "Climate impact of supersonic air traffic: an approach to optimize a potential future supersonic fleet-results from the EU-project SCENIC," *Atmos. Chem. Phys*, vol. 7, pp. 5129–5145, 2007.

[8] R. L. Mckenzie, P. J. Aucamp, A. F. Bais, L. O. Björn, M. Ilyas, and S. Madronich, "Ozone depletion and climate change: impacts on UV radiation," *Photochem. Photobiol. Sci*, vol. 10, pp. 182–198, 2011.

[9] J. Pletzer, D. Hauglustaine, Y. Cohen, P. Jöckel, and V. Grewe, "The climate impact of hydrogen-powered hypersonic transport," *Atmospheric Chemistry and Physics*, vol. 22, no. 21, pp. 14 323–14 354, 11 2022.

[10] J. Zhang, D. Wuebbles, J. H. Pfaender, D. Kinnison, and N. Davis, "Potential Impacts on Ozone and Climate From a Proposed Fleet of Supersonic Aircraft," *Earth's Future*, vol. 11, no. 4, p. e2022EF003409, 4 2023.

[11] T. M. Fritz, I. C. Dedoussi, S. D. Eastham, R. L. Speth, D. K. Henze, and S. R. Barrett, "Identifying the ozone-neutral aircraft cruise altitude," *Atmospheric Environment*, vol. 276, p. 119057, 5 2022.

[12] J. Pletzer and V. Grewe, "Sensitivities of atmospheric composition and climate to altitude and latitude of hypersonic aircraft emissions," *Atmos. Chem. Phys*, vol. 24, pp. 1743–1775, 2024.

[13] J. van 't Hoff, V. Grewe, and I. Dedoussi, "Sensitivities of atmospheric ozone to supersonic aircraft emissions across two flight corridors, in review," *Journal of Geophysical Research: Atmospheres*, 2023.

[14] K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, "Modal analysis of fluid flows: An overview," *AIAA Journal*, vol. 55, no. 12, pp. 4013–4041, 10 2017.

[15] K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S. T. Dawson, and C. A. Yeh, "Modal analysis of fluid flows: Applications and outlook," *AIAA Journal*, vol. 58, no. 3, pp. 998–1022, 10 2020.

[16] K. Champion, B. Lusch, J. Nathan Kutz, and S. L. Brunton, "Data-driven discovery of coordinates and governing equations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 45, pp. 22 445–22 451, 11 2019.

[17] M. A. Khodkar and P. Hassanzadeh, "A data-driven, physics-informed framework for forecasting the spatiotemporal evolution of chaotic dynamics with nonlinearities modeled as exogenous forcings," *Journal of Computational Physics*, vol. 440, p. 110412, 9 2021.

[18] J. Callaham, S. Brunton, J. N. Kutz, and D. Storti, "Multiscale model reduction for unsteady fluid flow," Ph.D. dissertation, 2022.

[19] C. A. Keller and M. J. Evans, "Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10," *Geoscientific Model Development*, vol. 12, no. 3, pp. 1209–1225, 3 2019.

[20] C. A. Keller, M. J. Evans, J. N. Kutz, and S. Pawson, "Machine learning and air quality modeling," *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, vol. 2018-January, pp. 4570–4576, 7 2017.

[21] M. M. Kelp, D. J. Jacob, J. N. Kutz, J. D. Marshall, and C. W. Tessum, "Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System," *Journal of Geophysical Research: Atmospheres*, vol. 125, no. 23, 12 2020.

[22] M. M. Kelp, D. J. Jacob, H. Lin, and M. P. Sulprizio, "An Online-Learned Neural Network Chemical Solver for Stable Long-Term Global Simulations of Atmospheric Chemistry," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 6, p. e2021MS002926, 6 2022.

[23] M. Velegar, N. Benjamin Erichson, C. A. Keller, and J. Nathan Kutz, "Scalable diagnostics for global atmospheric chemistry using Ristretto library (version 1.0)," *Geoscientific Model Development*, vol. 12, no. 4, pp. 1525–1539, 4 2019.

[24] X. Yang, L. Guo, Z. Zheng, N. Riemer, and C. W. Tessum, "Atmospheric chemistry surrogate modeling with sparse identification of nonlinear dynamics," 2024.

[25] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control*, 2021.

[26] J. Weiss, "A Tutorial on the Proper Orthogonal Decomposition," Tech. Rep., 2019.

[27] P. A. K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, "Using noisy or incomplete data to discover models of spatiotemporal dynamics," *Physical Review E*, vol. 101, no. 1, 2020.

[28] U. Fasel, J. N. Kutz, B. W. Brunton, and S. L. Brunton, "Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 478, no. 2260, 2022.

[29] N. M. Mangan, J. N. Kutz, S. L. Brunton, and J. L. Proctor, "Model selection for dynamical systems via sparse regression and information criteria," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2204, 1 2017.

[30] B. M. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Nathan Kutz, and S. L. Brunton, "PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data," *Journal of Open Source Software*, 2017.

[31] A. A. Kaptanoglu, B. M. de Silva, U. Fasel, K. Kaheman, A. J. Goldschmidt, J. L. Callaham, C. B. Delahunt, Z. G. Nicolaou, K. Champion, J.-C. Loiseau, J. N. Kutz, and S. L. Brunton, "PySINDy: A comprehensive Python package for robust sparse system identification," *Journal of Open Source Software*, vol. 7, no. 69, p. 3994, 11 2021.

[32] T. Askham and J. N. Kutz, "Variable projection methods for an optimized dynamic mode decomposition," *SIAM Journal on Applied Dynamical Systems*, vol. 17, no. 1, pp. 380–416, 2018.

[33] D. Sashidhar and J. N. Kutz, "Bagging, optimized dynamic mode decomposition for robust, stable forecasting with spatial and temporal uncertainty quantification," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2229, 2022.

[34] N. Demo, M. Tezzele, and G. Rozza, "PyDMD: Python Dynamic Mode Decomposition Software," *Journal of Open Source Software*, 2018.

[35] S. M. Ichinaga, F. Andreuzzi, N. Demo, M. Tezzele, K. Lapo, G. Rozza, S. L. Brunton, and J. N. Kutz, "PyDMD: A Python package for robust dynamic mode decomposition," 2 2024.

Appendices

A POD modal analysis

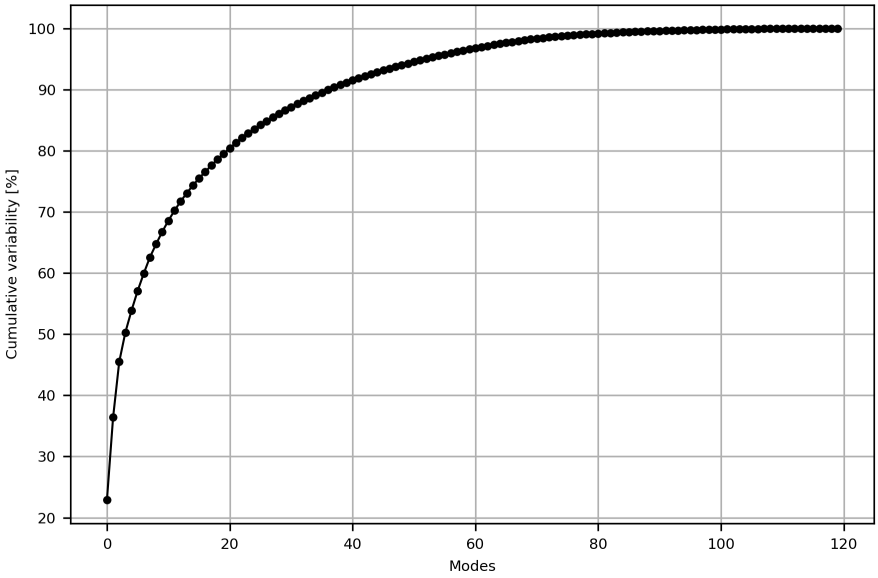


Figure 10: Cumulative variability captured by number of modes from POD analysis on TAC162 emission scenario. With 4 modes 50% of the variability of the data is represented.

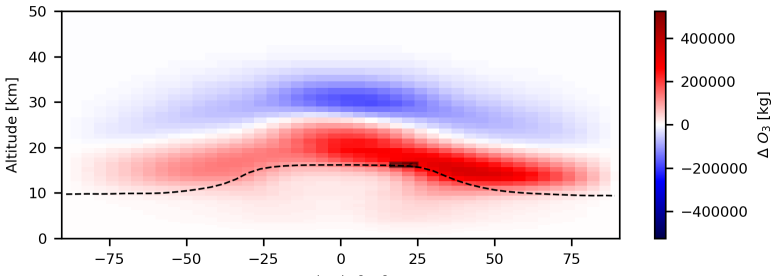


Figure 11: Mean ozone change across latitude and altitude for emission scenario SAS162 from years 2014 to 2023.

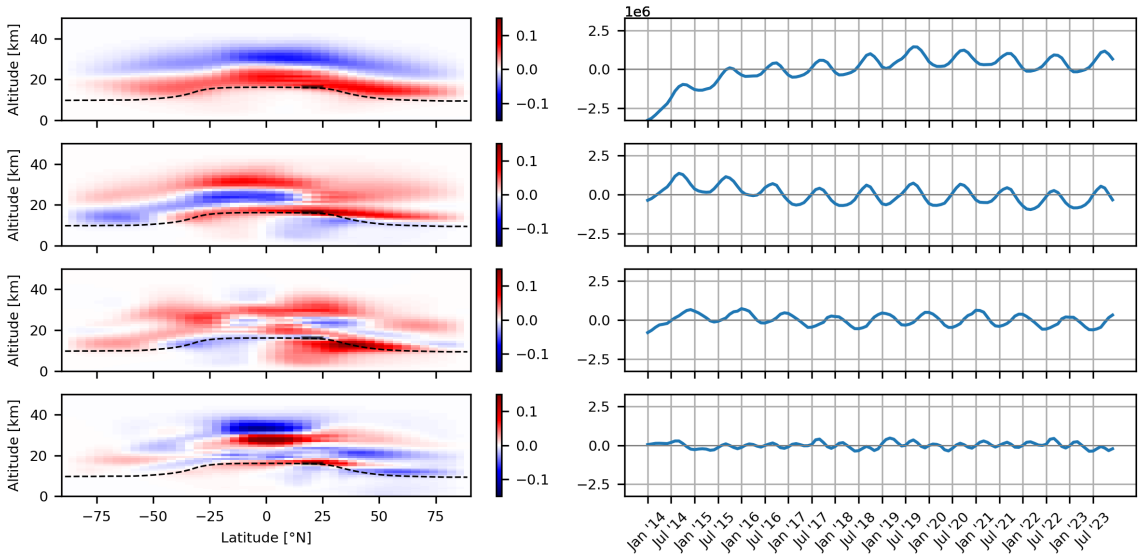


Figure 12: First four dominant POD modes for SAS162 scenario. Left: spatial modes, grey rectangle indicates emission region and dotted line indicates mean tropopause height. Right: time coefficients indicating how spatial modes evolve over time.

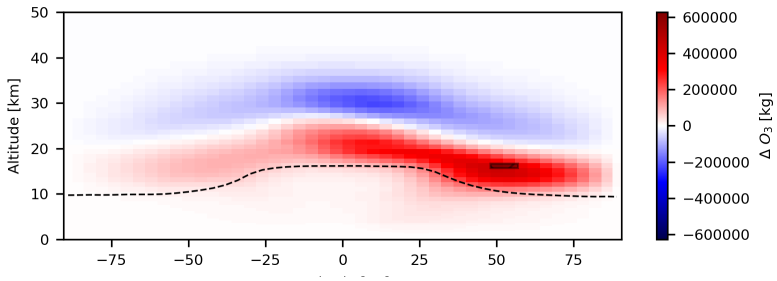


Figure 13: Mean ozone change across latitude and altitude for emission scenario TAC162 from years 2014 to 2023.

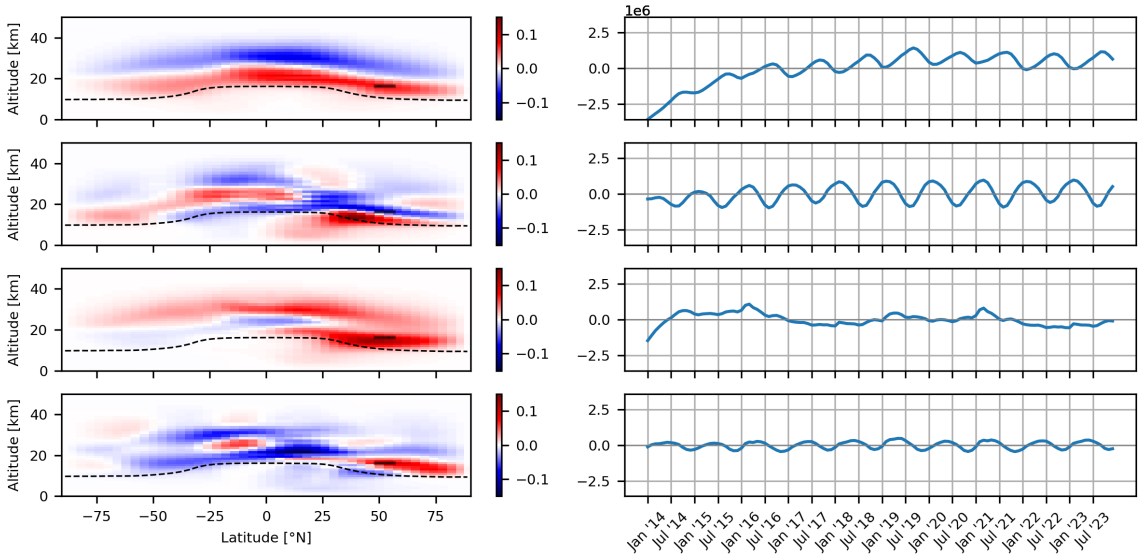


Figure 14: First four dominant POD modes for TAC162 scenario. Left: spatial modes, grey rectangle indicates emission region and dotted line indicates mean tropopause height. Right: time coefficients indicating how spatial modes evolve over time.

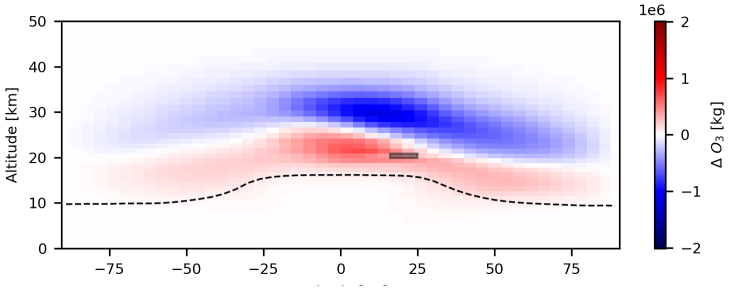


Figure 15: Mean ozone change across latitude and altitude for emission scenario SAS204 from years 2014 to 2023.

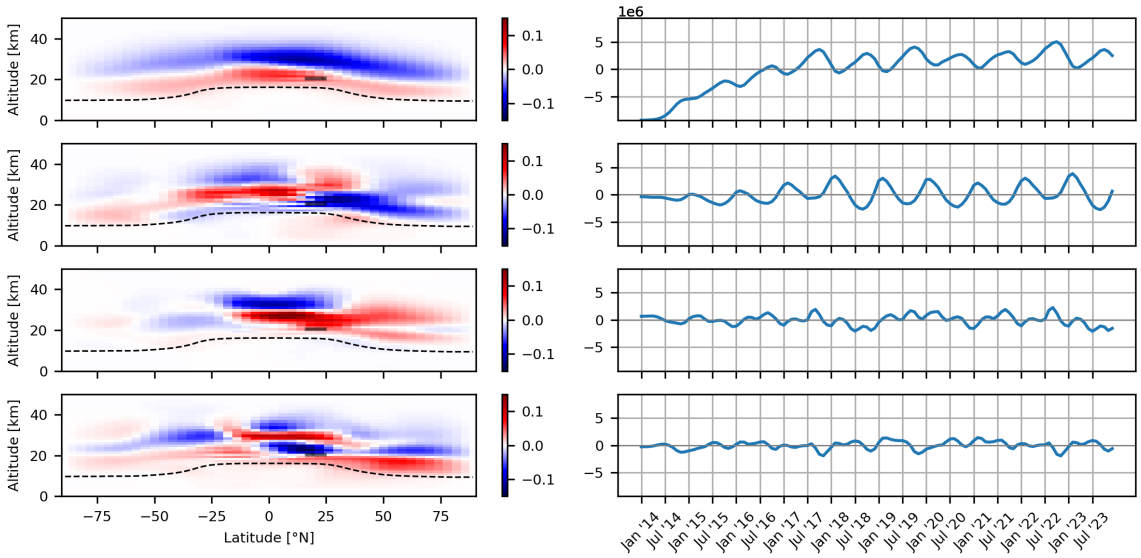


Figure 16: First four dominant POD modes for SAS204 scenario. Left: spatial modes, grey rectangle indicates emission region and dotted line indicates mean tropopause height. Right: time coefficients indicating how spatial modes evolve over time.

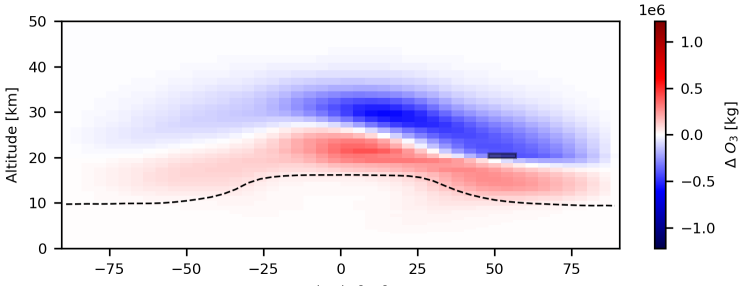


Figure 17: Mean ozone change across latitude and altitude for emission scenario TAC204 from years 2014 to 2023.

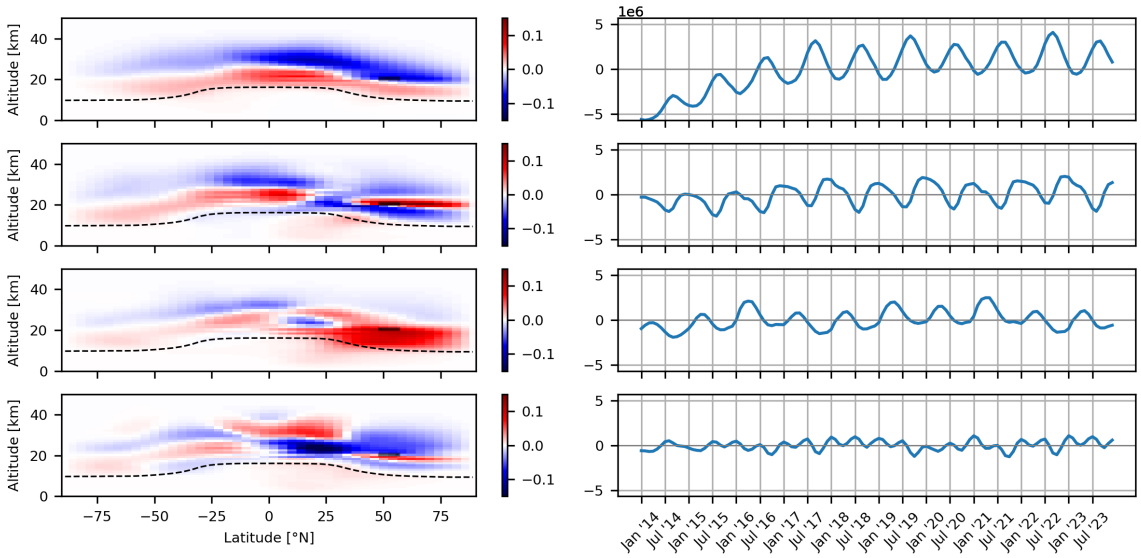


Figure 18: First four dominant POD modes for TAC204 scenario. Left: spatial modes, grey rectangle indicates emission region and dotted line indicates mean tropopause height. Right: time coefficients indicating how spatial modes evolve over time.

B POD modal analysis 2019-2021

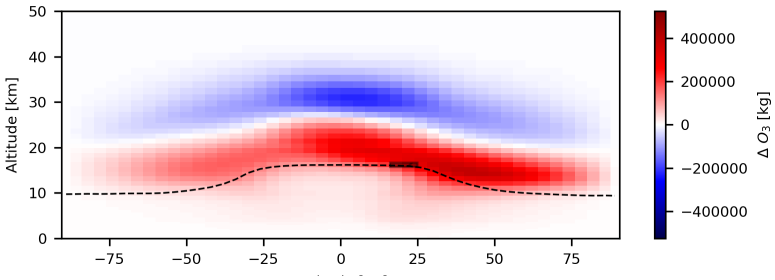


Figure 19: Mean ozone change across latitude and altitude for emission scenario SAS162 from years 2019 to 2021.

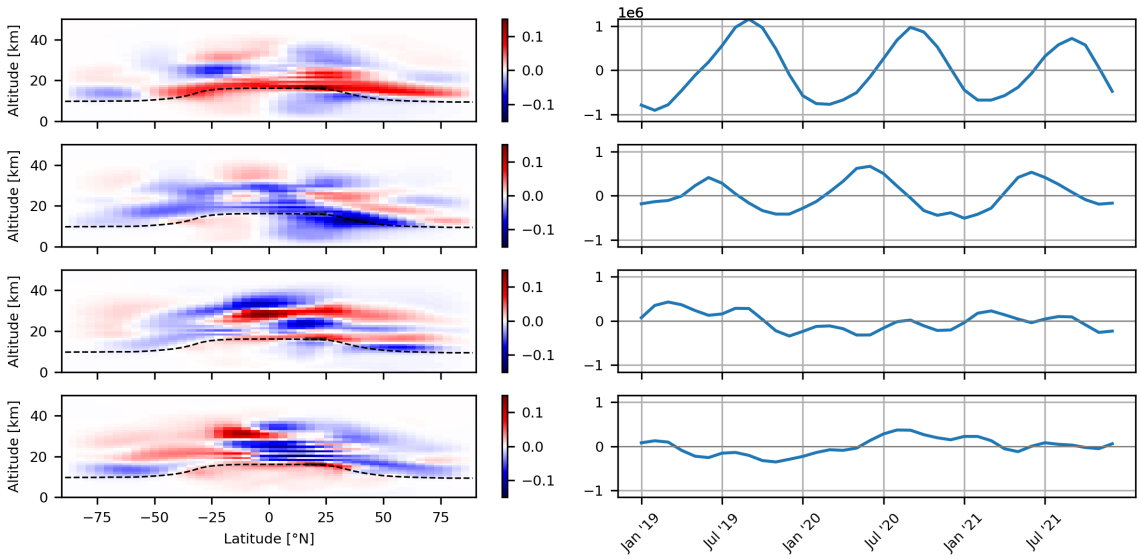


Figure 20: First four spatial modes SAS162 and associated time coefficients from POD analysis on data from 2019 to 2021.

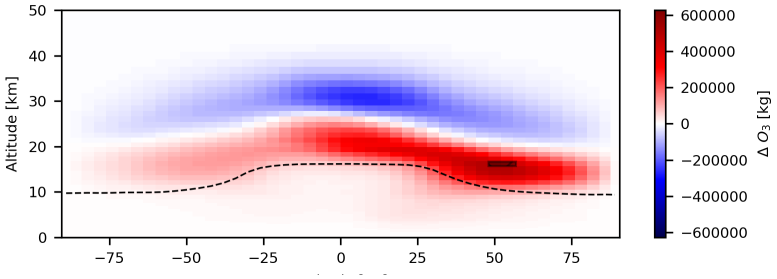


Figure 21: Mean ozone change across latitude and altitude for emission scenario TAC162 from years 2019 to 2021.

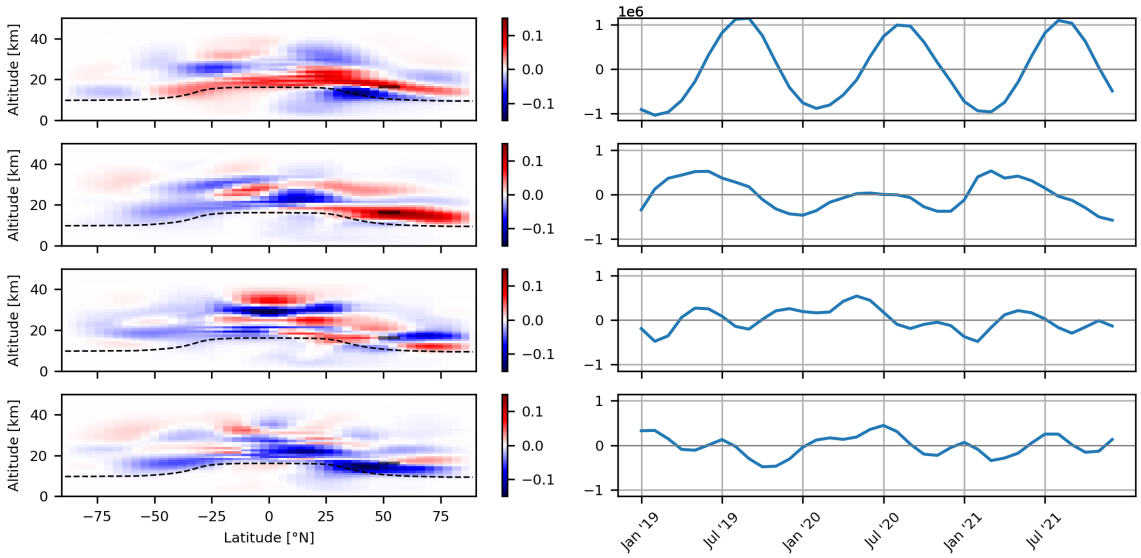


Figure 22: First four spatial modes TAC162 and associated time coefficients from POD analysis on data from 2019 to 2021.

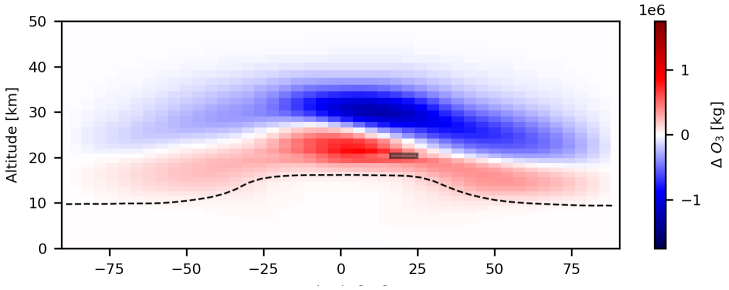


Figure 23: Mean ozone change across latitude and altitude for emission scenario SAS204 from years 2019 to 2021.

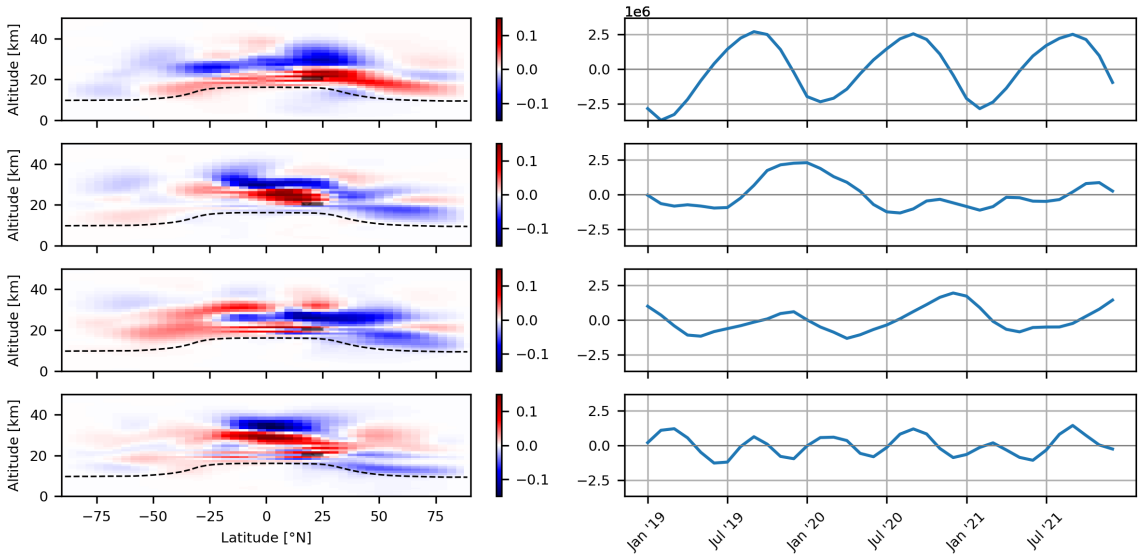


Figure 24: First four spatial modes SAS204 and associated time coefficients from POD analysis on data from 2019 to 2021.

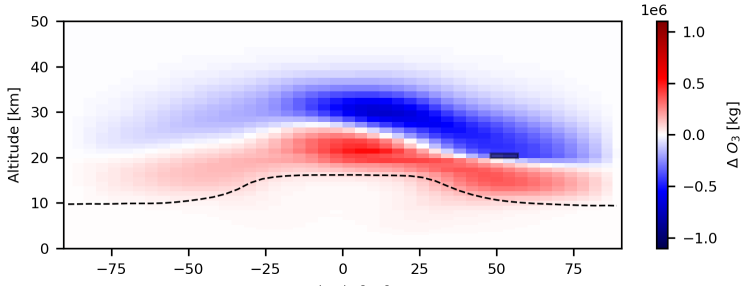


Figure 25: Mean ozone change across latitude and altitude for emission scenario TAC204 from years 2019 to 2021.

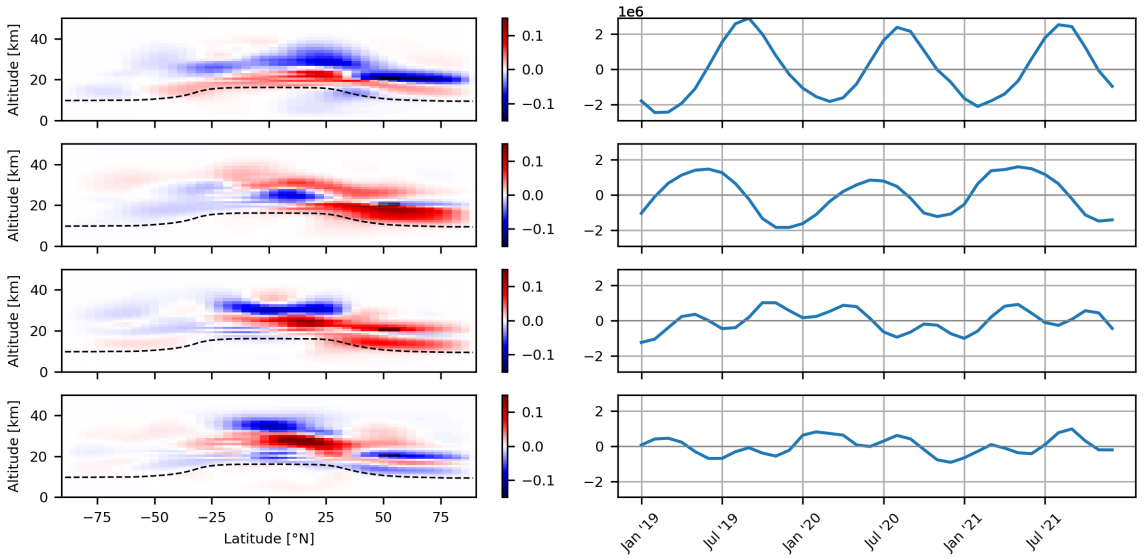


Figure 26: First four spatial modes TAC204 and associated time coefficients from POD analysis on data from 2019 to 2021.

C SINDy optimal fits

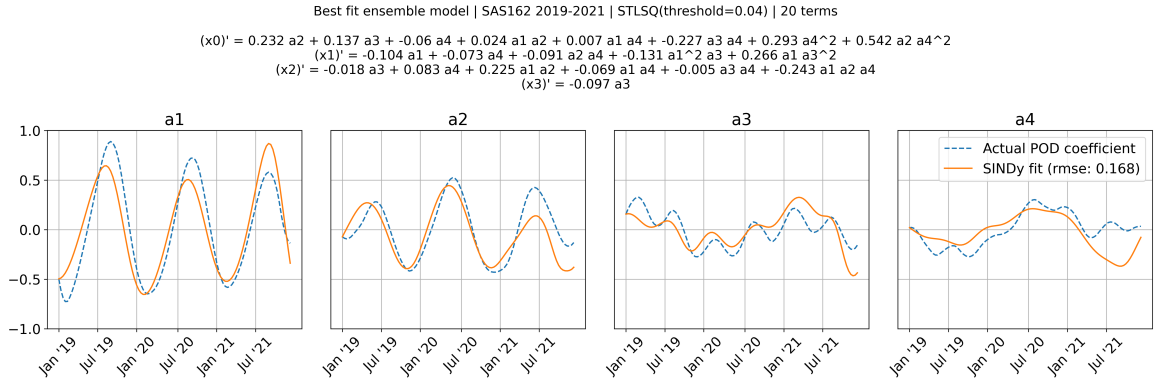


Figure 27: Best fit ensemble SINDy model on SAS162 emissions scenario on POD coefficients from year 2019-2021. Solved using threshold of 0.04. Subtitle includes resulting ODE's.

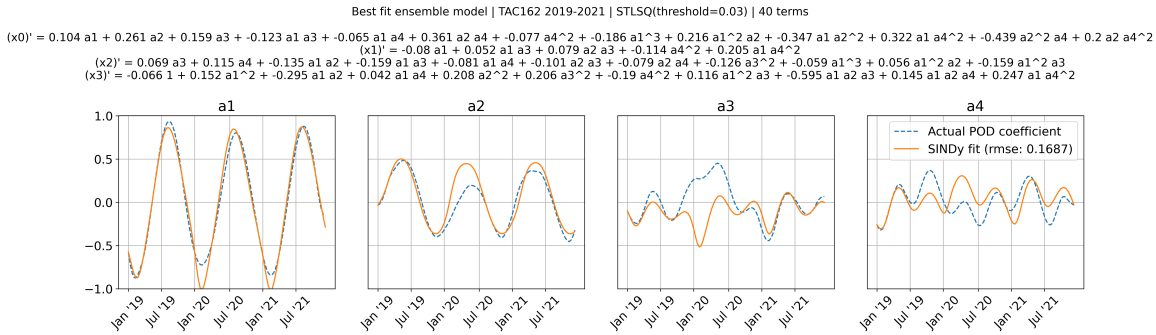


Figure 28: Best fit ensemble SINDy model on TAC162 emissions scenario on POD coefficients from year 2019-2021. Solved using threshold of 0.03. Subtitle includes resulting ODE's.

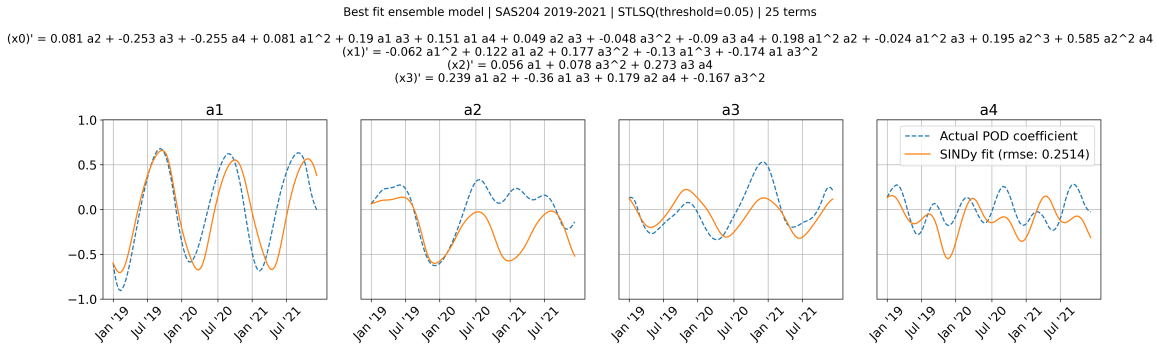


Figure 29: Best fit ensemble SINDy model on SAS204 emissions scenario on POD coefficients from year 2019-2021. Solved using threshold of 0.05. Subtitle includes resulting ODE's.

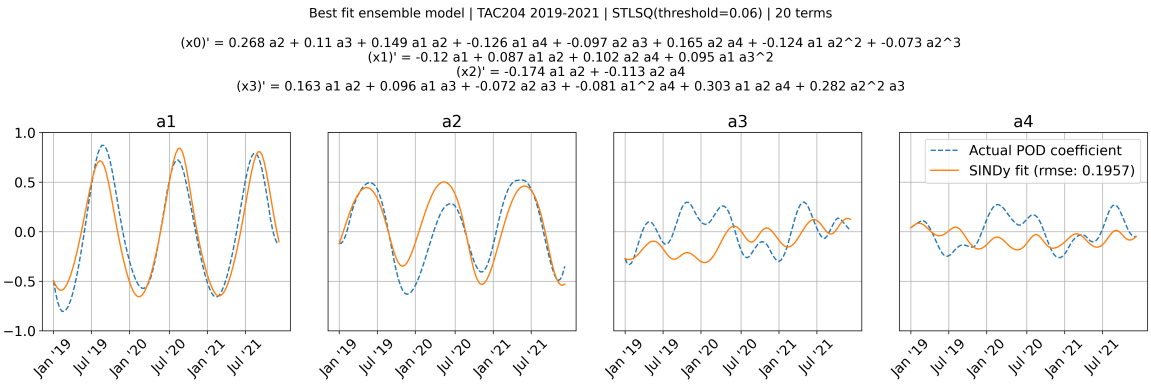


Figure 30: Best fit ensemble SINDy model on TAC204 emissions scenario on POD coefficients from year 2019-2021. Solved using threshold of 0.06. Subtitle includes resulting ODE's.

D DMD modal analysis

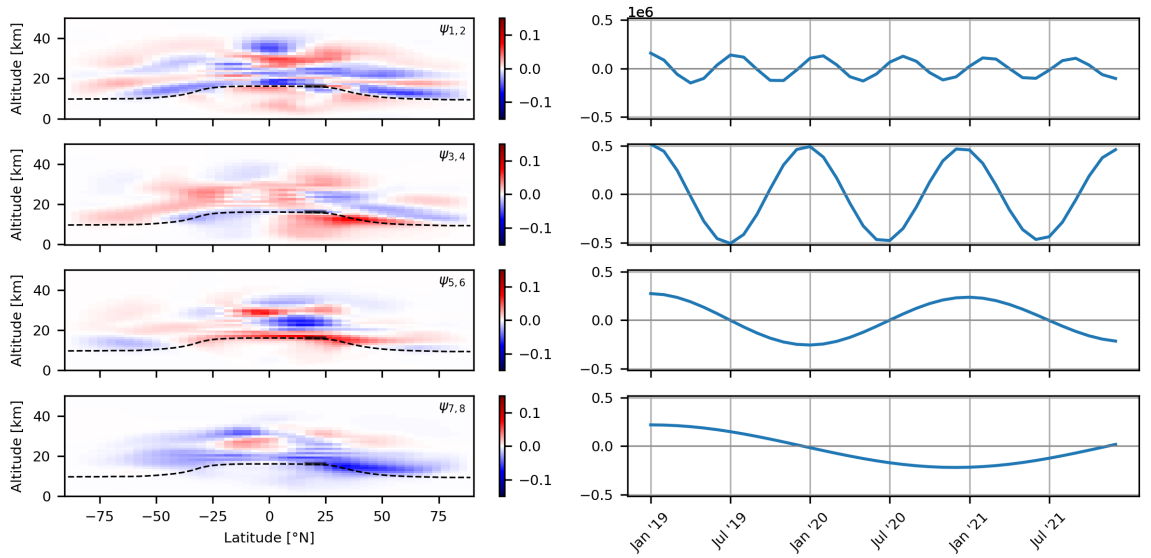


Figure 31: First eight DMD modes SAS162 and time coefficients.

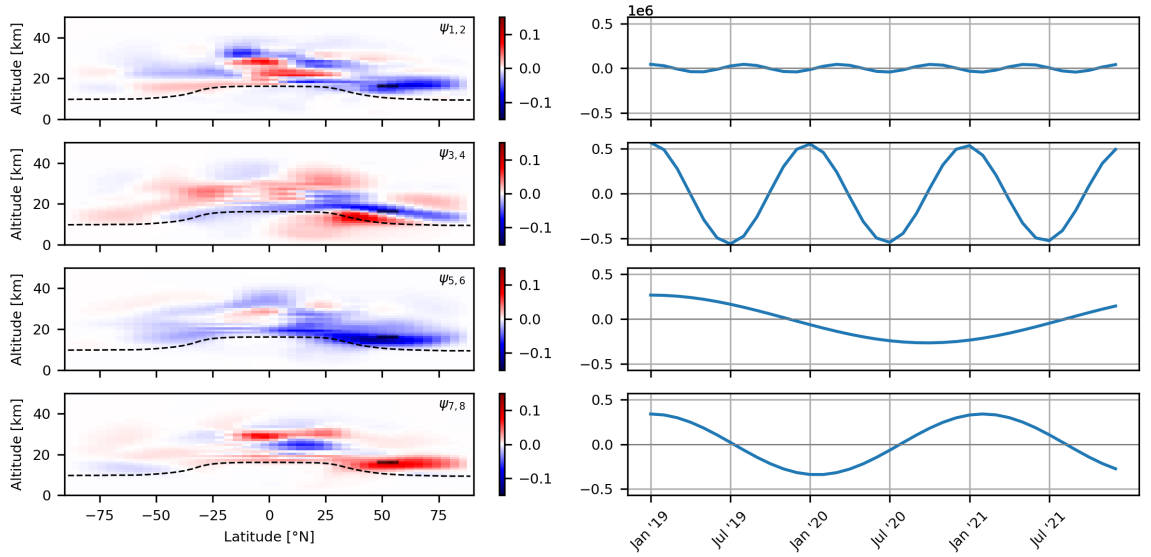


Figure 32: First eight DMD modes TAC162 and time coefficients.

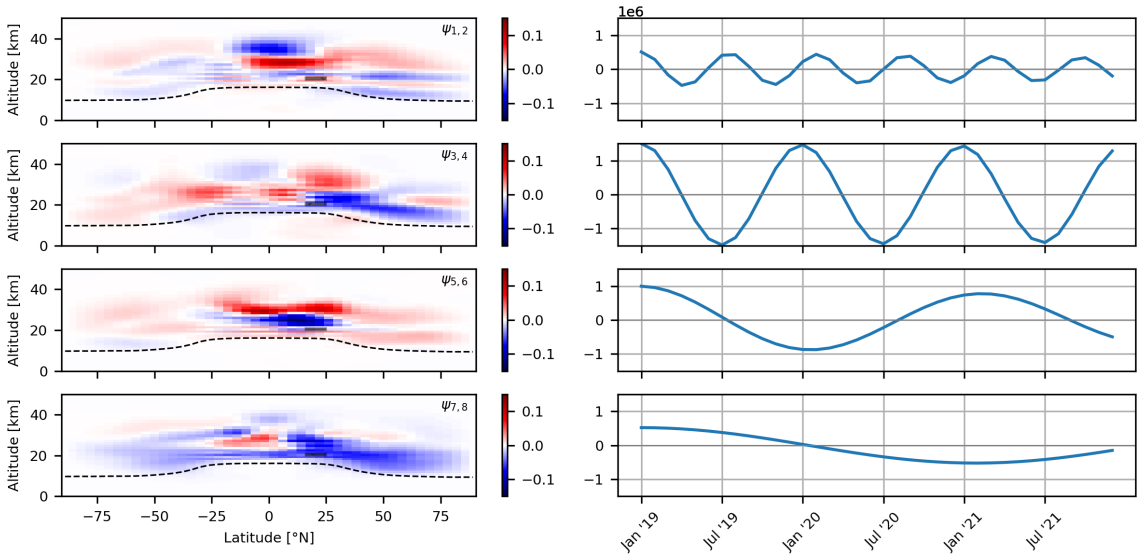


Figure 33: First eight DMD modes SAS204 and time coefficients.

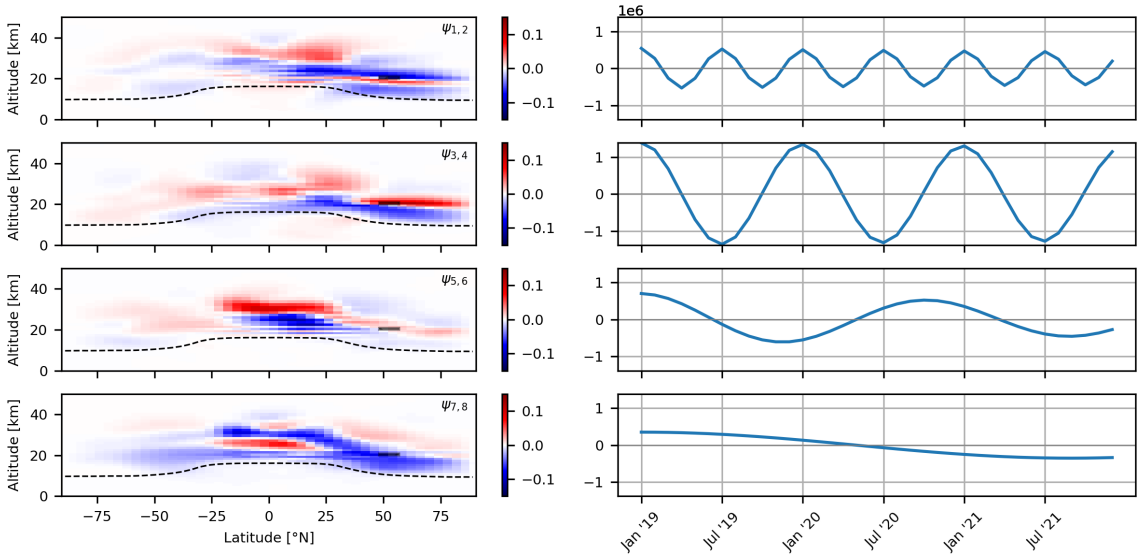


Figure 34: First eight DMD modes TAC204 and time coefficients.

Part II

Literature Study

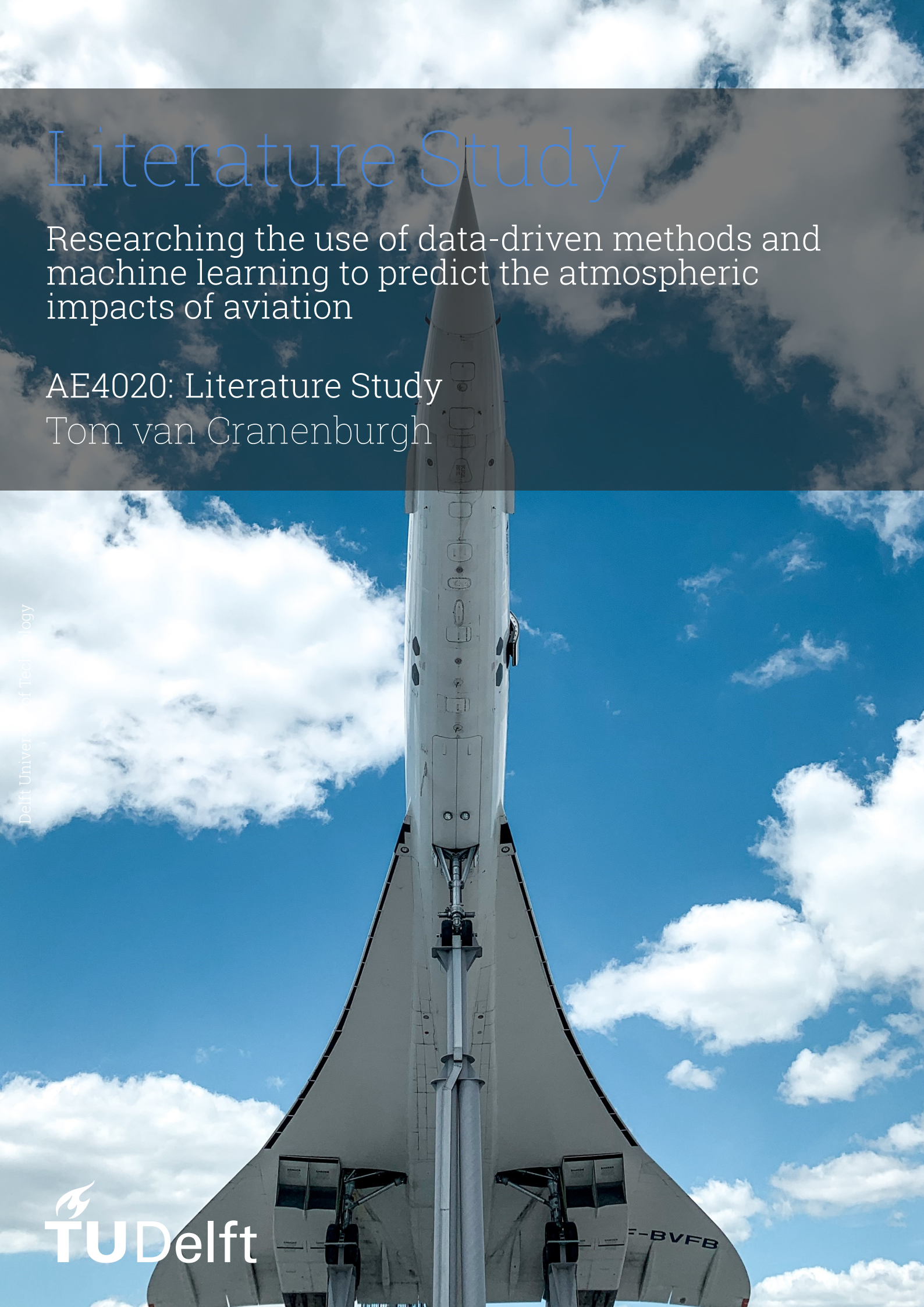
Literature Study

Researching the use of data-driven methods and machine learning to predict the atmospheric impacts of aviation

AE4020: Literature Study

Tom van Cranenburgh

Delft University of Technology



TU Delft

Literature Study

Researching the use of data-driven methods and machine learning to predict the atmospheric impacts of aviation

by

Tom van Cranenburgh
4647645

This literature study is submitted in fulfilment of the course AE4020

Supervisors:	I. (Irene) Dedoussi
	J. (Jurriaan) van 't Hoff
Institution:	Delft University of Technology
Place:	Faculty of Aerospace Engineering, Delft
Date:	Monday 4th December, 2023

Cover: Library, *Unsplash 2023*

TU Delft

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
BC	black carbon
CAEP	Committee on Aviation Environmental Protection
CO	carbon monoxide
CO_2	carbon dioxide
CTM	chemistry transport model
CH_4	methane
DMD	Dynamic mode decomposition
DU	Dobson unit
EI	emission index
EOF	empirical orthogonal functions
ERF	effective radiative forcing
GEOS-Chem	Goddard Earth Observing System chemistry model
ICAO	International Civil Aviation Organization
IPCC	Intergovernmental Panel on Climate Change
ML	machine learning
NO_x	nitrogen oxides (NO + NO_2)
O_2	oxygen
O_3	ozone
PCA	principal component analysis
POD	proper orthogonal decomposition
RF	radiative forcing
ROM	reduced order modelling
SINDy	sparse identification of nonlinear dynamics algorithm
SO_x	sulfur oxides
SO_2	sulfur dioxide
SST	supersonic transport
SVD	singular value decomposition
UV	ultraviolet
VOC	volatile organic compound

1

Introduction

Aviation emissions contribute up to 3.5% of the human related climate forcing effects. This percentage consists of both CO_2 and non-CO_2 effects [1]. The aviation industry is in constant development to meet this challenge with the prospect of radical (or revolutionary) aircraft designs taking flight in the future. In order to evaluate the impact of novel aircraft design on climate, insights into the emissions of the aircraft and their effects over different time scales are required. A common approach to gaining these insights is the use of chemistry transport models (CTM). These models capture the complexity of the chemical interactions and physical processes taking place in the atmosphere. This research focuses on gaining new insights into atmospheric impacts of aviation emissions and evaluating models able to predict future changes in atmospheric composition. In-depth knowledge on the effects of emissions is valuable for both engineers involved in designing future systems and policymakers tasked with establishing regulations for emission indices and flight conditions.

The re-emergence of commercial supersonic aircraft calls for updates on outdated regulations that specify constrains on emissions. Companies such as Spike, NASA, Overture and Boom Supersonic are working on supersonic aircraft that will dramatically reduce flying times [2]. These supersonic aircraft fly at speeds above Mach 1 and at higher cruise altitudes, in the upper troposphere and lower stratosphere, and thus have stronger impacts on the atmosphere compared to conventional subsonic aviation. The existing regulations on the emissions of these aircraft are from the time the Concorde was flying commercially (1976 to 2003). Since then, the CO_2 and non-CO_2 impacts of subsonic aircraft at specific altitudes, trajectories and year-long scenarios have developed substantially, but assessments on supersonic aviation climate effects remain superficial. To set new regulations, knowledge on the climate effects of supersonic aviation is required. The effects are evaluated using climate transport models, but such models are computationally expensive, running up to weeks on supercomputers to gain detailed results. Consequently, a faster method for evaluating the emission impact of future supersonic aircraft scenarios is desired to update policies, increase understanding and ultimately minimise environmental effects.

The climate transport models evaluate emissions on the complex atmosphere-climate system, being capable of accounting for various feedbacks in meteorology, climate and chemistry. One method of evaluating anthropogenic emissions effects is quantifying changes in the concentration of chemical species, which can be used to evaluate warming and cooling effects. The Goddard Earth Observing System Chemistry (GEOS-Chem) model, is a chemistry transport model that stands out for providing accurate representations of chemical species concentrations across spatial and temporal dimensions. However, the high computational expense hinders their use for operational and engineering decision-making.

The proposed approach to tackle the computational limitation of these models is the use of machine learning and data-driven engineering methods. The goal of this research is to create a reduced order, computationally cheaper, prediction model that is able to evaluate the effects of future scenarios, using data from GEOS-Chem models that has evaluated emission scenarios as training data. The research

question is therefore formulated as follows:

How can data-driven engineering and machine learning be used to gain new insights into the impact of supersonic aviation emissions and develop models capable of predicting the changes in atmospheric composition?

The goal of this literature review is to provide an overview on the available literature of current applications of machine learning and data-driven techniques on atmospheric chemistry data and other fields of spatio-temporal data. Additionally, this review elaborates on the climate implications of supersonic aviation and provides an explanation of the mathematical concepts behind dimensionality reduction techniques. This is done to provide more in-depth knowledge of the research topic before answering the research questions. In this report, chapter 2 includes the outline and the scope of the research, chapter 3 explains climate effects of supersonic aviation, chapter 4 contains a description of the GEOS-Chem climate model, the various data-driven reduced order modelling and machine learning methods are explained in chapter 5 and an overview of current applications of these methods in available literature is provided in chapter 6.

2

Research outline

This chapter gives an outline of the research on climate impacts of aviation emissions, reduced order modelling and machine learning methods that could be used for their evaluation. Section 2.1 describes the research problem, section 2.2 lists the research questions and sub questions. In section 2.3, the scope of the research is explained to provide context and the rationale for focusing on one specific aspect of the research problem. The planned approach to the research can be found in section 2.4.

2.1. Research Problem

With non-CO_2 aviation emissions warming the climate at twice the rate of that associated with aviation CO_2 emissions, a continued demand for accurate evaluation of anthropogenic climate change research remains [1]. As air travel demand recovers from the Covid-19 pandemic, simultaneously, novel aircraft architectures are being developed to align with the dynamic economic landscape and the growing demand for sustainable alternatives. Although supersonic flight has the great advantage of reducing flight duration its use came to an end in 2003, the year the Concorde made its last flight. However, several companies in recent years have commenced developing supersonic commercial aircraft, while also addressing environmental issues [2]. With the use of Sustainable Aviation Fuel (SAF) *Boom Supersonic*, one of the current leaders in designing a supersonic passenger plane, is promising to deliver a net carbon neutral supersonic commercial aircraft to the market, capable of carrying around 80 passengers at a cruise speed of Mach 1.7. Nevertheless, one of the largest concerns of supersonic aviation the sonic boom created at such high speeds. This extreme noise lacks social acceptance and resulted in countries banning supersonic commercial flights above land. Boom's 2022 sustainability report outlines plans to tackle this issue by avoiding supersonic speeds over land [3].

The other large concern is uncertainty in the climate impact of supersonic emissions at higher altitudes, more specifically the non-CO_2 impacts. Supersonic planes' high speed requires almost twice the fuel consumption of subsonic aircraft, increasing the chemical species concentrations emitted [4]. With the impact of contrail formation to be greatly reduced at higher cruising altitudes compared to subsonic typical altitudes, the main concerns are related to changes in concentrations of direct and indirect greenhouse gases such as ozone and water vapour [5]. Nitrogen oxide (NO_x) emissions can lead to formation of ozone but at higher altitudes also result in ozone depletion. The emitted water vapour is another concern as it is a potent greenhouse gas [6].

Quantifications on the effects of supersonic aviation is required to update the current outdated regulations. As the use of climate transport models to evaluate these is computationally expensive, there is a need for computationally cheap models that provide accurate insights into effects of supersonic aviation emissions.

2.2. Research Question

The main research question that covers the entire scope of this research is as follows:

How can data-driven engineering and machine learning be used to gain new insights into the impact of supersonic aviation emissions and develop models capable of predicting the changes in atmospheric composition?

Various methods will be explored that might suit the application of predicting climate effects in a computationally less expensive manner than current climate transport models. The research question contains two specific goals: gaining new insights and developing a model for prediction. The lower level research questions are thus:

1. To what extent can data-driven methods enhance our understanding of the impact of supersonic aviation emissions on atmospheric composition?
    (a) To what extent can data-driven algorithms capture dominant spatial atmospheric chemistry patterns and mechanisms for different chemical species and for different aviation emission scenarios?
    (b) To what extent can data-driven algorithms capture dominant temporal atmospheric chemistry patterns that occur in daily, monthly and seasonal cycles?
    (c) Which chemical species exhibit dominant spatial and temporal patterns and what mechanisms drive these patterns?
    (d) To what extent can data-driven algorithms recognise transport of chemical species?
2. What data-driven techniques are effective for developing reduced-order models capable of accurately predicting changes in atmospheric composition along multiple dimensions?
    (a) How many dimensions of the output data can accurately represent the impact of an aviation emissions on the change in concentration of a chemical species?
    (b) What are the challenges and limitations of multidimensional modelling in capturing the dynamic nature of atmospheric composition changes due to aviation emissions?
    (c) What validation methods and metrics can be used to assess and compare the accuracy of results generated by different machine learning models?

2.3. Research Objective and Scope

Section 2.1 and section 2.2 outlined the research problem and focus. As stated, there are many uncertainties in the exact effects of individual supersonic emissions scenarios. Methods to create a prediction model will be explored for various emissions scenarios (varying emission altitude, latitude and longitude) and the concentration changes of ozone (O_3). The changes in ozone concentration are the largest non-CO_2 effect in terms of radiative forcing [7]. The extent to which production of ozone in the lower stratosphere and depletion in the upper stratosphere takes place is a major concern [8].

If the methods are shown to be suitable for the application of ozone, attempting to create prediction models for other chemical species other chemical species of interest (H_2O, H_2O, BC) will be explored. Various species exhibit diverse atmospheric responses, encompassing differences in production and depletion rates, transport, and more. Consequently, certain methods may not be suitable for developing a reduced-order model for all chemical species.

This research focuses on addressing percentage and absolute mass changes of a chemical species of interests. Determining total radiative forcing is beyond the scope of this work.

Lastly, the available data originates from the GEOS-Chem model. As will be discussed in later chapters, various climate transport models (CTM) exist and have been used to evaluate the effect of supersonic aviation emissions. This research does not focus on validating the GEOS-Chem model or comparing results from various models for different scenarios, but instead focuses on using GEOS-Chem data to build a reduced order model capable of predicting effects of future scenarios.

2.4. Research Planning

The research consists of four phases. A detailed planning can be found the form of a Gantt-chart in appendix A. A description of each phase is provided below:

1. **Literature phase**: this phase consists of a comprehensive literature study and creating a research plan.
2. **Initial phase**: in the initial phase, the available data will be analysed and various dimensionality reduction methods will be examined. Methods at creating reduced order models of individual scenarios will be explored. This phase ends with a midterm meeting,
3. **Final phase**: in this phase the final prediction model will be developed, including verification and validation of the method. This includes finding a way to interpolate individual scenario models that allow for prediction of new unknown emission scenarios. This phase ends with a green light meeting in which final results are presented, giving room for final feedback that is to be included in the final version of the research report.
4. **Graduation phase**: in the last phase no large modifications or developments are made to the methodology of the project. This phase is dedicated to implementing final feedback, adjusting conclusions to complete the final report, to prepare the graduation presentation and defence.

3

Aviation emissions and climate effects

The effects of aviation on the atmosphere can be categorised into noise pollution, air pollution, nitrogen deposition, climate change and ozone layer impact. Only the latter two effects are within the scope this research, with a specific focus on the effects of non-CO_2 emissions in the troposphere and stratosphere, and their impacts on the concentration of chemical species. Noise pollution is a major concern for supersonic aircraft but is not touched upon in this research. Air pollution refers to surface level air quality effects due to aviation emissions. Section 3.1 includes an explanation of the layers of the Earth's atmosphere, and dominant transport patterns in the troposphere and stratosphere and species exchange between them. Section 3.2 describes ozone chemistry and the complex dependence of chemicals in ozone production cycles. Section 3.3 highlights all climate impacts of aviation emissions. Section 3.4 concludes with the important takeaways from this chapter. The chemistry transport model used to quantify atmospheric impacts is described in the following chapter, chapter 4.

3.1. Atmospheric dynamics

The concentration of chemical species and their interaction in the atmosphere is dependant on the dynamical properties of the atmosphere. Understanding the characteristics at varying altitude, latitude, the transport of chemical species, together with the local temperature, humidity and radiation are important to understanding chemical behaviour and the impact anthropogenic perturbations on the atmosphere. In this chapter these concepts are described.

3.1.1. Layers of the Earth's atmosphere

The layers of the atmosphere are separated by areas of varying temperature rate changes. The mesosphere, stratosphere and troposphere along with the temperature gradient in each corresponding layer and ozone concentration at a certain altitude are depicted in Figure 3.1. The thermosphere and exosphere are not shown in this figure, as they irrelevant to this research.

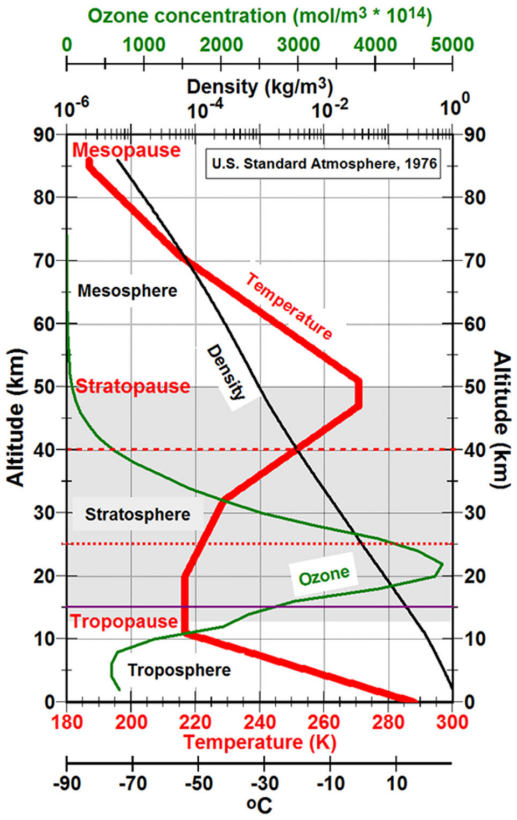


Figure 3.1: Temperature gradient and ozone concentrations across layers of the atmosphere [9]

The interaction of aviation emissions and the Earth's atmosphere takes place in the troposphere and the stratosphere. Subsonic aviation cruise altitude ranges are in the upper troposphere whereas supersonic commercial aviation is expected to fly in the lower stratosphere. The temperature in the troposphere decreases with altitude to around -50 ° C in the upper troposphere. A major difference between the troposphere and stratosphere is the portion of mass of air contained in each layer, with the troposphere containing 80% of the Earth's atmosphere's mass, resulting in higher density and pressures at these altitudes than in the stratosphere [10]. Furthermore, the warmer air present in the lower troposphere results in exchanges of energy associated with convection. The variability in the atmosphere in the lower troposphere, includes different weather patterns including those associated with high and low pressure fronts.

Figure 3.1 also shows how the concentration of ozone varies with altitude. The stratosphere contains around 90% of the atmospheric ozone, with the *ozone layer* around 20-30 km altitude [11]. The majority of ozone is produced above this region due to the intense solar radiation. Ozone in the stratosphere plays a fundamental role in keeping harmful radiation from life on earth by absorbing ultraviolet radiation. On the other hand, ozone in the troposphere (often present in smog and other forms of air pollution) is undesired, as surface level ozone is harmful to human health and crop production. The processes related to ozone production and formation are discussed in more detail in subsection 3.2.1.

The region in between the stratosphere and troposphere is known as the tropopause. This boundary is a chemically inactive, unstable region, whose height varies between 9 and 18 km above the Earth's surface. The tropopause height varies with latitude, season and weather patterns. The height of the tropopause and its thickness vary depending on the intensity of solar energy and vertical mixing, resulting in a lower tropopause height above polar regions and higher tropopause heights above the tropics.

The majority of the Earth's atmosphere contains nitrogen (78%) and oxygen (21%) and many other trace gases present in small quantities that play important roles in atmospheric chemistry and environ-

mental processes [12]. In this research, the concentration of species present in the atmosphere will be distinguished as *background concentration* and concentrations induced from aviation emissions or the results of chemical interactions of the species emitted with background concentration.

3.1.2. Transport in the troposphere

The dynamics of atmospheric gases is also influenced by weather patterns, horizontal and vertical mixing. The transport of species influences chemical reactions and total concentrations in the atmosphere and are thus relevant to understand in relation to how perturbations such as aviation emissions disrupt the system.

Convection in the troposphere is driven by temperature differences, and results in the rising of warmer, less dense air and sinking of cooler, denser air. This motion mixes the atmosphere, distributing various gases throughout the troposphere. As convection is able to rapidly transport species from lower altitudes to higher altitude, the exact transport time together with the lifetime of a species determines the impact. For instance, a short-lived species such as NO_x has a lifetime of several days to months (varying with altitude) and its transport results in the actual location of NO_x emissions being important to its impact [13]. As NO_x plays a crucial role in ozone production, examining its impact will be of interest later.

At different latitudes, the atmosphere experiences different intensities of solar radiation during different seasons due to the tilt of the axis around which the Earth spins and orbits around the sun elliptically. This results in a surplus of incoming radiation at the tropics, and a radiation deficit towards the poles. To restore equilibrium in the atmosphere, transport of energy takes place through the movement of air parcels. The prevailing winds occurring from these differences induce the redistribution of air as pictured in Figure 3.2. The motion of air parcels from the tropics towards the polar regions in the atmosphere is characterised by three circulation cells: Harley, Ferrel (or mid-latitude) and polar cells.

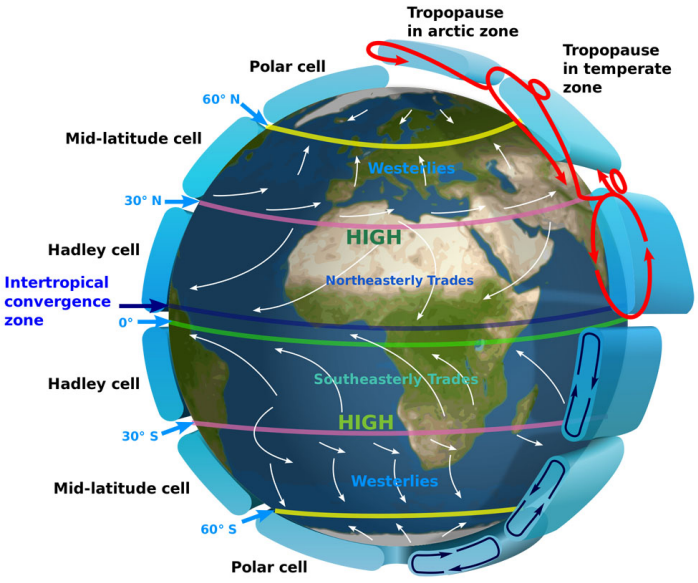


Figure 3.2: Prevailing winds and circulation cells [1]

The direction of the trade winds and westerlies can be explained by the Earth's rotation, the circulation cells and the Coriolis effect. The latter states that air parcels at lower latitudes travel faster than at higher latitudes as they are covering a larger distance in the same time. Air parcels moving from a higher latitude to a lower latitude, such as the air parcels transported by the Hadley circulation cell at low altitudes in the Northern hemisphere, will be moving slower than the ground beneath as it moves to a lower latitude. This is because the air parcel's inertial speed remains the same while the ground beneath is moving relatively quicker. This causes the air parcel to *fall behind*, resulting in trade winds travelling from East to West.

The circulation cells indicate a limited exchange of species between hemispheres. It is probable that emissions in the Northern hemisphere will not reach the Southern hemisphere, with the additional consideration that this outcome depends on the lifetime of the species. Transport of species is also highly dependent on weather patterns. Figure 3.2 depicts the common transport patterns, but it's important to note that weather patterns significantly influence both atmospheric conditions and the transport of chemicals.

3.1.3. Exchange between troposphere and stratosphere

The aformented vertical mixing and upward transport of warmer air results in transport of species between the troposphere and stratosphere. The circulation that describes the exchange is the Brewer-Dobson circulation, visualised in Figure 3.3.

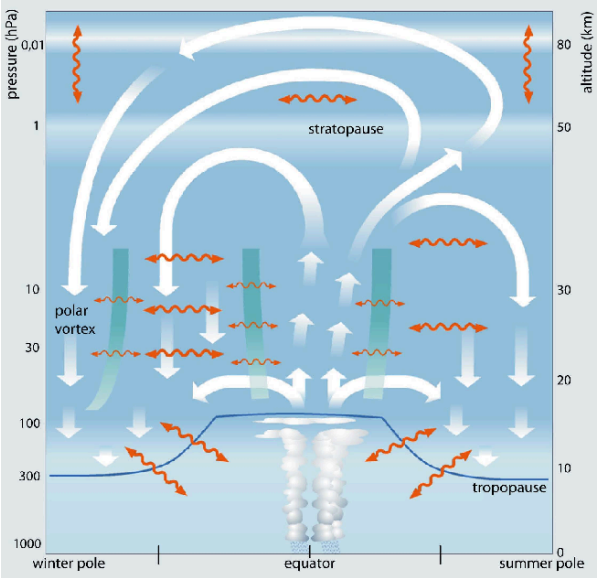


Figure 3.3: Brewer-Dobson circulation [14]

The overall results of the Brewer-Dobson circulation of air parcels is a slow updraft in the tropics, and subsidence in polar winter regions. A more significant portion of tropospheric air transport upwards travels to winter poles as it enters the stratosphere. During winter months, the winter pole experiences less radiation and thus more cooling of the stratosphere and a greater energy difference than with the summer pole, the polar vortex forms as a result. In the Northern Hemisphere the polar vortex attracting more poleward transport of air parcels to the North Pole occurs in November - March. A similar pattern occurs in the Southern Hemisphere in the months May to September. In the summer months in corresponding hemispheres, there is more horizontal transport rather than poleward transport.

Downward transport between these two layers takes place at mid latitudes, where the tropopause breaks occur due to abrupt pressure level changes [10]. The interaction between two circulation cells shown in Figure 3.2 causes jet streams, a narrow band of air with persistent direction in which wind speeds can become significant, facilitate the exchange of air between the troposphere and stratosphere. Therefore the region in which most downward transport of species from stratosphere to the troposphere takes place, is at the points of interaction of circulation cells, at 60 ◦ N/S and 30 ◦ N/S [14]. This is indicated by the red arrows on the blue line that represents the tropopause height in Figure 3.3. Understanding the distribution of gases in the stratosphere and their seasonal variations is crucial in recognising patterns from aviation emissions.

3.1.4. Atmospheric impact and climate change

Climate change refers to the disruption of the climate due to anthropogenic emissions. Generally it refers to changes over a longer period of time, mainly several years, and is often measured in temperature increase as a result of radiative forcing (RF). Radiative forcing is used to predict the expected equilibrium global mean surface temperature change that results from climate forces, measured in mWm^{-2}. The simplest definition given is the change in total irradiation at the top of the atmosphere [1]. The effective radiative forcing (ERF) is an indicator of the eventual global mean temperature response including only 'fast' atmospheric responses. The equation that relates RF and surface temperature change, where λ is the climate sensitivity parameter, is the following:

$$\Delta T_s = \lambda RF \tag{3.1}$$

The RF can be estimated by using climate models to calculate the changes in incoming short-wave solar radiation minus outgoing long-wave radiation in the atmosphere. Even though this research focuses on changes in the concentration of certain chemical species, it is important to understand that in many studies these concentration changes are used to determine RF of certain emitted species, which in turn allow researchers to make predictions on warming potential. Changes in concentration of a chemical species are commonly expressed as mass changes in certain regions of a period of time. A notable metric that will be used to assess the changes in ozone concentration is the Dobson Unit (DU), which represents the thickness of a layer of ozone from the earth's surface to the top of the atmosphere. As ozone varies with latitude but also altitude, the column ozone gives an indication of a location's UV radiation exposure. This together with ground level ozone is used to analyse smog levels and implications on health effects.

A description of a climate model is given in chapter 4 and an explanation of how gases emitted from aviation engines influence the atmosphere is provided in section 3.3. The complex chemistry of ozone, whose concentrations supersonic aircraft emissions have a significant impact on, is described in the next section.

3.2. Ozone chemistry

Surface level-ozone is harmful to human health as it affects human breathing due to increased smog concentration, thus excess oxygen atoms as a result of NO_2 emissions at lower altitudes are undesired. On the other hand, stratospheric ozone is important as it protects life on Earth from harmful levels of ultraviolet (UV) radiation from the Sun and because it is a greenhouse gas that is important to the Earth's climate [15]. Thus examining the concentrations of ozone and its dependencies at tropospheric and stratospheric altitude levels is of importance.

3.2.1. Tropospheric chemistry

NO_x emissions from aviation in the troposphere alter the composition of the atmosphere mainly through a short-term increase in ozone and a long-term decrease in methane. The long-term methane decrease leads to a feedback of less ozone production and eventually also to stratospheric water vapour [1]. The formation and depletion of ozone in the troposphere is dependant on atmospheric conditions and the presence of other particles such as NO_x.

In the troposphere, ozone is predominantly produced in the presence of NO_x following the reactions below [16]. The energy from sunlight (*hv*) initiates the photolysis of NO_2 to form NO and O. The oxygen atom (O) reacts with oxygen molecule (O_2) to form ozone (O_3) when energised with sufficient sunlight,

$$NO + HO_2 \rightarrow NO_2 + OH \tag{3.2}$$

$$NO_2 + hv \rightarrow NO + O \tag{3.3}$$

$$O_2 + O \rightarrow O_3 \tag{3.4}$$

The total ozone production is dependent on various chemicals, primarily involving the HO_x and NO_x cycles. The mechanism of ozone production in the troposphere is visualised in Figure 3.4. NO_x represents both NO_2 and NO molecules, even though engine emissions indices often specify the emission index of NO_x as g(NO_2)/kg(fuel).

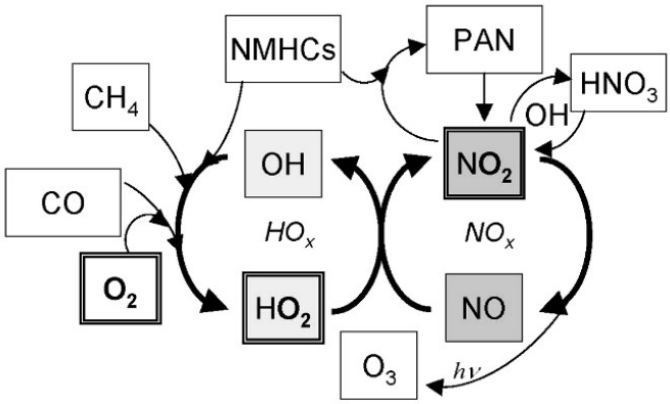


Figure 3.4: Schematic representation of Tropospheric ozone production mechanism. NMHC are non-methane hydrocarbons, PAN is peroxi-acetyl-nitrate and serves as a reservoir for nitrogen oxides, thus controlling their long term transport [17]

Figure 3.4 illustrates the complex dependencies of various chemicals in ozone production. The complete and incomplete combustion in aircraft engines emits NO_x, CO, CH_4, NMHCs and shows how all these chemicals directly and indirectly contribute to and increase in ozone. A long-term secondary effect of ozone production, is related to the methane concentration. Figure 3.4 shows that increased NO_x concentrations in the presence of hydroxil radical (OH), result in the formation of nitric acid (HNO_3). In this cycle, increased NO_x concentrations will result in higher O_3 and OH concentrations. The latter influences methane depletion according to the following process [7]:

$$CH_4 + OH \rightarrow CO_2 + H_2O \tag{3.5}$$

In the troposphere concentrations of CH_4 and CO are significantly higher than NO_x concentrations, often making NO_x the rate limiting step in tropospheric ozone production. The interaction of methane with ozone and NO_x is important to understand as both ozone and methane are greenhouse gases that influence the radiative forcing in the atmosphere.

Ozone loss in the troposphere occurs due to dry deposition and chemistry, most dominantly with the interaction of HO_2, described by the reaction below.

$$O_3 + HO2 \rightarrow OH + 2O_2 \tag{3.6}$$

In conclusion, the influence of aviation emissions, particularly NO_x, on tropospheric chemistry has wide-ranging implications for both air quality and the Earth's radiative balance. The complex interactions of chemical reactions involving NO_x, O_3, CH_3 have a significant impact on short-term production and long-term destruction of ozone.

3.2.2. Stratospheric chemistry

The stratosphere plays a crucial role in climate and ozone layer protection. Stratospheric chemistry is a complex system influenced by many chemical processes. Understanding the influence of aviation emissions such as NO_x on the disruption of the system is essential in understanding the environmental impact of supersonic aviation flying and emitting in the stratosphere.

The production of ozone in the stratosphere primarily takes place in the mid and upper stratosphere and is driven by high-energy UV radiation levels. The initial reaction involves an oxygen (O_2) molecule

photolysed into two oxygen atoms (O). Strong covalent bonds in O_2 are broken by high energy UV radiation, the free radical O reacts with oxygen molecule to form ozone.

$$O_2 + hv \rightarrow O + O \tag{3.7}$$

$$O_2 + O \rightarrow O_3 \tag{3.8}$$

The process of ozone depletion takes place in the presence of various compounds. The concentration of ozone in the stratosphere is reduced by ozone photolysis and catalysed ozone destruction. The latter is dependant on CO, NO, ClO and BrO acting as a catalyst in the reaction. The natural form of ozone depletion is the photolysis of ozone molecules, summarised by the reaction below, and take places in the upper stratosphere. The reaction is initiated by high energy UV radiation which is present at higher altitudes.

$$O_3 + hv \rightarrow O + O_2 \tag{3.9}$$

Catalysed ozone destruction takes place, the NO_x compounds NO and NO_2 act as a catalyst in ozone destruction in the following reactions. Other compounds including chlorine (Cl), bromine (Br) and hydroxyl radicals (OH) also act as catalysts in a similar manner breaking down ozone (O_3).

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{3.10}$$

$$NO_2 + O \rightarrow NO + O_2 \tag{3.11}$$

3.2.3. Influence of atmospheric conditions on reaction productivity

The efficiency of cycles that lead to increased ozone depletion increases with higher intensity UV radiation, lower air pressures decreasing likelihood of collisions with other substances, and a high concentration of ozone [18]. The upper stratosphere has these three conditions and the presence of catalytic radicals will contribute to ozone destruction reactions. The lower stratosphere also contains high ozone concentrations, but the less intense UV radiation and higher air pressures make ozone less susceptible to catalytic reactions and thus depletion. Natural ozone disassociation, the photolysis of ozone molecules, takes place mostly in the mid and upper stratosphere. The depletion as a result of interaction with catalytic radicals takes place at all altitude levels, but has a more profound impact on ozone destruction at higher altitudes because of the aforementioned favourable conditions [19].

The presence of aerosols in the stratosphere also results in a direct aerosol effect. The aerosols have the ability to absorb or scatter incoming solar radiation. Furthermore, an indirect effect occurs through heterogeneous chemistry as the aerosol particles facilitate its surface for chemical reactions to take place. One of the concerns with aviation emissions is the emissions of sulfur dioxide (SO_2) that can be converted to sulfur aerosols that enhance the direct and indirect aerosol effect. Often the sulfur content of fuel is discussed in relation to aerosol effects. The inorganic (sulfate) and carbonaceous (black carbon) aerosols are the most significant aerosol effects related to aviation emissions. The important takeaway is that increased aerosol concentrations lead to direct effects including scattering and absorbing sunlight and indirect effects through providing a surface of heterogeneous chemical reactions. Increased aerosol emissions will thus lead to more efficient destruction of ozone in the stratosphere.

3.3. Influence of aviation emissions

The extent to which aviation emissions contribute to the atmospheric impacts in the troposphere and stratosphere is complex. One of the main concerns with supersonic aviation emissions is the increased ozone depletion in the stratosphere as a result of NO_x emissions, but aerosol effects, water vapour emissions, and contrail formations are also a concern [18]. In this section the research that has been performed into both subsonic and supersonic aviation and their non-CO_2 emissions' atmospheric impacts is discussed, including the variability with respect to season, latitude, and altitude emission location.

3.3.1. Products from engine combustion

Complete or incomplete jet engine combustion results in emission of the following gasses: carbon dioxide (CO_2), nitrogen oxides (NO_x), carbon monoxide (CO), water vapour (H_2O), sulfur oxides (SO_x), particulate matter (PM), black carbon also known as soot, unburned hydrocarbons, and aerosol particles [1]. The quantity emitted by an aircraft is dependant on the thrust settings that determine the fuelburn, and the engine type that determines the emissions indices indicating mass of species emitted per unit mass of fuel burnt. Supersonic aircraft fly at higher speeds than subsonic aircraft, requiring higher temperature combustion and more fuelburn to achieve higher thrust, leading to mainly higher emission quantities than subsonic aircraft (exact quantities dependant on the emission indices of the engine type). The chemicals emitted are affected by homogeneous chemistry, microphysics, heterogeneous chemistry and aerosol dynamics and circulation. One of the effects of the emitted particles is contrail formations, which form depending on the plume composition and the atmospheric conditions at emission location.

The concentration of species emitted during landing and take-off, climb and descent are not examined in this research as they contribute significantly less to the long term climate impact [1]. The cruise altitude emissions and their effects on climate change and the impact on the ozone layer will be examined. More concretely, concentration changes in CO_2, NOx, O_3, SO_2, black carbon and H_2O will be examined to evaluate these effects. The emitted species contribute to radiative forcing of the Earth, with non-CO_2 effects responsible for two-thirds of the negative effective radiative forcing from global aviation [1].

3.3.2. Atmospheric effects of emissions

CO_2 and water vapour are direct greenhouse gases causing net positive radiative forcing (RF), warming, and ozone (O_3) and methane (CH_4) are considered indirect greenhouse gases. A significant climate forcer of supersonic aviation is water vapour (H_2O) emissions [5] [20]. Stratospheric H_2O emissions are a significant factor to climate impacts from supersonic aircraft aviation [8]. Water vapour has a longer effective lifetime at higher altitudes, causing an increase positive radiative forcing effect compared to the almost negligible radiative forcing effect of subsonic aviation emissions [18]. Emissions at higher altitudes persisting longer in the stratosphere leading to greater accumulation of aviation water vapour and greater RF. Increases in stratospheric water vapour could significant cool the stratosphere while warming the troposphere.

Ozone (O_3) and methane (CH_4) are indirect greenhouse gases as they contribute to radiative forcing. These gases are not directly emitted, but the interaction of aviation emitted gases influence the production and destruction cycles of these gases.

Aerosols absorbing shortwave radiation before it can cross the tropopause in the stratosphere produces a negative radiative forcing [18]. However, many uncertainties remain in quantifying this effect. Exploring patterns of black carbon and sulfate emissions and their concentration levels at various altitudes in this research will help quantify the effects of this.

3.3.3. Subsonic aviation emissions effect

An overview of the contribution of emissions and their radiative forcing is given in Figure 3.5.

Figure 3.5 depicts an estimate of the total global aviation contribution to effective radiative forcing from 1940 to 2018. A significant conclusion to be made is the contribution of non-CO_2 emissions on the net aviation effective radiative forcing (ERF). Furthermore, it is evident that the contrail cirrus and ozone increase due to NO_x are two major components of this non-CO_2 ERF. The emissions of NO_x is a species of interest, whose emissions at higher altitudes (above 12 km) result in short-term ozone increases, long-term ozone decrease, a methane decrease and a stratospheric water vapour decrease. The extent to which each of these effects take place is dependant on many factors, but all effects are taken into account in the net warming contribution of nitrogen oxides [1]. NO_x is an example of a complex species that is response for multiple climate responses and is thus a species to be examined on more detail.

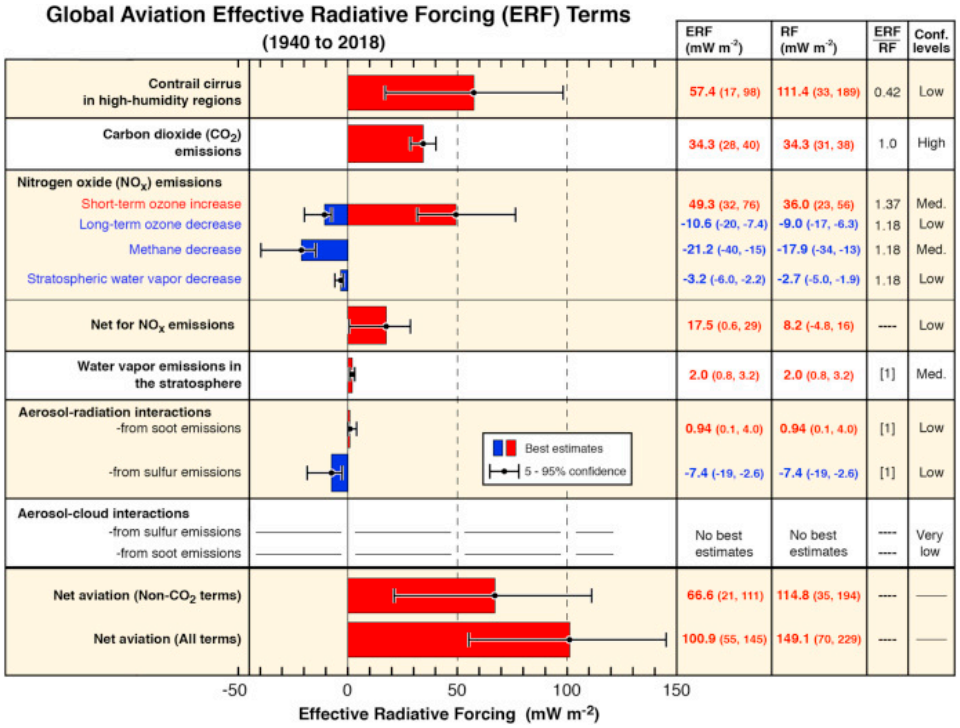


Figure 3.5: Quantifying component contributions to global aviation effective radiative forcing from 1940 to 2018 [1]

3.3.4. Supersonic aviation emissions effect

A comprehensive overview of multiple studies into the non-CO_2 effects of supersonic aviation emissions is given by Matthes et al. [6], including comparisons of individual effects and their quantified RF estimates. Several studies have investigated the effects of a fleet estimated to be flying at cruise altitudes within the range of 14-20 km. These effects include the effect of NO_x emissions on homogeneous chemistry, aerosol direct and indirect effects, stratospheric water vapour, CO_2 effects and contrail formation.

The paper stresses a highly non-linear dependence of NO_x concentrations in ozone production, and that there is high uncertainty between models on change in ozone concentration in upper troposphere and lower stratosphere as a result of NO_x emissions [6]. The primary effects of supersonic emissions in terms of RF and climate effect are changes in ozone concentration and stratospheric water vapour, thus stressing that the climate effect of a supersonic fleet is considerably larger than a subsonic fleet, with the RF being five times as large. The longer lifetimes of H_2O and other species together with a significantly higher fuel burn for supersonic aircraft causes this difference.

One of the most recent studies, performed by Eastham et al. [18] predicts a 0.046% reduction in global column ozone of a fleet at Mach 1.6, 15-17 km cruise altitude, burning 19 Tg of fuel each year. The largest radiative forcing component is change in ozone concentration, followed by water vapour. Two negative radiative forcing effects are the methane and (sulfur and black carbon) aerosol effects. Absorbing aerosols emitted in the stratosphere produce a negative RF as it absorbs shortwave radiation before it can cross the tropopause [18]. The negative RF from long-term methane response is not caused by affecting mid- and upper stratospheric catalytic ozone depletion cycles, but by reductions in lower-altitude production of ozone.

A study performed by NASA [5] finds a net non-CO_2, non-contrail radiative forcing of -0.02 $mW\ m^{-2}$ per billion seat-km, flying at 17 km cruise altitude at Mach 1.6. This negative radiative forcing is attributable to net ozone depletion, of which sulfur emissions are largely responsible followed by NO_x emissions. Furthermore, this research touches upon the formation of persistent condensation trails, known as contrails that persist for several minutes but can evolve into cirrus clouds dependant on atmospheric conditions (cold and humid). This study predicts that supersonic flights will generate fewer

persistent contrails than subsonic flights, varying per latitude and altitude due to ambient drier conditions at supersonic flight altitudes.

All studies show stratospheric water vapour and changes in ozone concentrations as the two largest radiative forcing factors of supersonic aviation emissions, but all mention sensitivities to emissions altitude and latitude, and that the predicted atmospheric changes vary with the seasons. Comparing the results of various simulations is possible as they account for fleet differences and normalise species concentration changes with fuel burn. Another factor that influences atmospheric response is the emission indices (EI) of NO_x and H_2O, with studies performing sensitivity studies showing strong ozone sensitivity to the EI of NO_x [21].

3.3.5. Sensitivity with respect to altitude, latitude and season

All studies find varying impact with altitude, and mention differences in atmospheric response with latitude and season. The ozone impact is most prominently discussed in [5–8, 18, 21].

Zhang et al. performed sensitivity experiments to evaluate responses at different altitudes (13-23 km), shows both ozone decrease and increase primarily as a result of nitrogen oxides emissions. Significant depletion takes place from 17 km up while production dominates at lower altitudes [8]. Grewe et al. showed that reducing cruise speed from Mach 2 to 1.6 (and thus its cruise altitude) could reduce both climate impact and ozone destruction by 40% [7]. At higher altitudes, the lower temperatures enhance the efficiency of catalytic ozone destruction processes involving NO_x. In general, the lower temperature slows down chemical reactions as fewer molecules have sufficient activation energy, but the catalytic destruction of ozone is less affected by this. An ozone-neutral aircraft flying altitude would be around 13-14 km according to [22]. Speth et al. [5] finds ozone-neutral cruise altitude near 14 km, dependant on fuel sulfur content and the NO_x emission index. There is also mention of an ozone self healing effect from ozone depletion at higher altitudes. The increased UV radiation at lower altitudes due to reduction in ozone at higher altitudes will lead to more oxygen photolysis at lower altitudes, and thus an increased ozone production. The direct aerosol effect and stratospheric water vapour concentrations are two other effects that have a greater impact at higher altitudes due to longer lifetimes and thus a greater radiative forcing. The study of Pletzet et al. into a more futuristic supersonic aircraft concept, a fleet of Mach 5-8 (15-35 km cruise altitude) fuelled with liquid hydrogen, in which emissions consist of water vapour and nitrogen oxides highlights the climate impact to be dependant on: atmospheric residence time of emitted species, mainly controlled by circulation of air (Brewer-Dobson), chemical interaction and photolysis [20].

Latitudinal variation of aviation emissions on climate has been the subject of research for subsonic aviation. Lower latitude emissions showed an ozone and methane RF six times larger than for higher latitude emissions [23]. Supersonic research shows that higher altitude emissions show more dominant distribution of ozone changes with respect to latitude [18]. The largest ozone destruction and production are visible towards the tropics, due to greater UV radiation at these latitudes. Furthermore, studies mention inter-hemispheric differences with larger ozone losses in the Northern hemisphere [5]. This can be explained by higher background NO_x concentrations in the Northern hemisphere. Early studies already showed the importance of taking into account stratospheric circulation of species as accumulated species accounted for significant amounts of concentration changes [24]. In this study the changes of H_2O accumulation in the tropical lower stratosphere was most important. An important difference between subsonic and supersonic emissions is that ozone concentrations are more chemically controlled in the stratosphere whereas in the troposphere they are more dynamically controlled. The turbulent motions and vertical mixing is more active in the troposphere, leading to more distribution of ozone whereas in the stratosphere, chemicals are more separated as chemical interactions and atmospheric conditions mainly control concentration change.

3.4. Conclusions

This chapter contained an overview of the literature on supersonic aviation atmospheric effects. The exact effects of emitting at various altitudes has been examined, but a need to do this for not only fleets but for individual flight regions (specific latitudes) is desired to gain more insight into the effects.

4

Atmospheric modelling: GEOS-Chem

The experimental set-up of this research includes developing a model that is able to predict the concentration of chemical species. The proposed methodology includes the use of both statistical and machine learning methods that base dominant patterns and input-output relationships on the underlying data. This section describes the set-up of the GEOS-chem model used to generate output data that will be used to identify patterns and serve as training data for the prediction model. A brief understanding of the GEOS-Chem model helps to understand how its limitations and assumptions influence this research.

The Goddard Earth Observing System Chemistry model is a 3-D chemistry transport model (CTM) that calculates atmospheric composition coupled to weather and climate. It is a Eulerian model, signifying that the air parcels are fixed in space. Species enter and leave though the fixed spaces, assessing the concentration of each species at every time step. Between each time step, the concentration changes within a grid space are determined based on the chemical interactions and meteorological conditions at that grid cell. An example of an arrangement of grid cells for a global CTM is portrayed in Figure 4.1. The model is also able to simulate horizontal and vertical exchange between columns and rows dependant on weather conditions. The entire atmosphere is simulated as a function of both space and time.

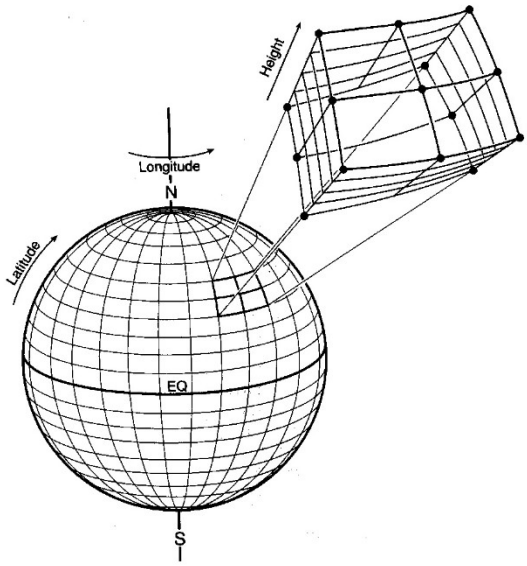


Figure 4.1: Global chemistry transport model grid schematic [25]

The advantage of the GEOS-Chem model is able to take into both account natural and anthropogenic

emissions. To calculate concentration changes, GEOS-Chem takes into account heterogeneous and homogeneous chemistry processes, aerosol process, photochemical and thermochemical reactions, deposition and transport [26]. Unlike other climate-transport models GEOS-Chem has no native components for evaluating climate impacts and changes in meteorology. Meotrological feedbacks are not incorporated in the model.

The GEOS-Chem model allows anthropogenic emissions to be added based on set latitude and altitude locations. To simulate supersonic aviation emission, likely emission locations can be based on extensive possible future supersonic routes for which there is expected to be a market. Many routes across oceans are expected to be covered by supersonic aircraft with subsonic aircraft serving shorter overland connections [5, 21].

The output *netCDF* files of the GEOS-Chem model contain chemical species concentrations at varying values latitude, longitude, altitude and time. The resolution across space and time dimensions can be configured starting from a resolution of 0.25° latitude x 0.3125° longitude at 72 altitude levels for every 10 minutes [26]. The required resolution of training data to generate a prediction model and to be able to capture dominant patterns is uncertain, therefore a range of spatial and time dimension levels of precision will be examined. Given its requirement to perform the vast amount of calculations for more than 300 chemicals at many time steps for a long simulation time, the GEOS-Chem model is run on a supercomputer. The resolution has an impact on the computational effort required to generate the desired data. For this research, data from GEOS-chem is generated in the following format.

- Chemical species concentration: 314 chemical species
- Altitude: 72 altitude levels, up to 80 km
- Latitude x Longitude: 4° x 5°
- Time: daily and monthly averages
- Time span: up to 10 years simulation after emission.

Another drawback of the GEOS-Chem model, other than the computational effort associated with running it, is the omitting of changes of atmospheric chemical species concentration feedback on meteorology. Changes in the composition of both the stratosphere and troposphere influence weather patterns, these feedbacks are not incorporated in GEOS-Chem simulations resulting in uncertainty of the simulation. Furthermore, the quantity of pollutants added to the simulation is dependant on set emission indices, which is important to take into account when comparing results with other simulations.

5

Data driven modelling

This chapter gives an overview of data-driven reduced order modelling and machine learning techniques that could be applied to spatio-temporal atmospheric chemistry data. A brief explanation of the important methods that are the basis of many reduced-order models is provided. Many of these methods have been adapted over the years for various applications, resulting in a vast range of extensions which are mentioned but not explained in detail.

Data-driven methods provide new methods of modelling, predicting and controlling complex scientific and engineering systems. Research are turning from empirical models or derivations based on first principles to data-driven approaches for a range of applications. The systems are often nonlinear, dynamic, high-dimensional, dimensions in space and time, but with dominant patterns. These dominant patterns can help to achieve the goal of predicting, representing, reducing size or controlling the system.

5.1. Singular value decomposition

The goal of dimensionality reduction is to identify coherent structures from high-dimensional data and the singular value decomposition (SVD) is a technique often used for this purpose. The SVD is a matrix decomposition technique that is used for a range of purposes. Multi-dimensional data, such as in the spatio-temporal chemical species concentrations obtained from chemistry transport models, can be reshaped into large column vectors that form a matrix. The data generated is often *low-rank*, indicating that there are a few dominant patterns that explain the high-dimensional data. The SVD is a robust method for extracting dominant patterns from data, and serves as the basis for many other methods [27].

A low-dimensional approximation of high-dimensional data can be determined by finding a dominant pattern. There is no knowledge of the underlying relations in the data, the dominant patterns are discovered purely from the data. Another benefit of the SVD is that the dominant correlations of the data are represented hierarchically and exist for any matrix. Its most important application, for which it will be used in this research, is dimensionality reduction.

The SVD can be applied to any matrix $\mathbf{X} \in \mathbb{C}^{n \times m}$ in the following form, where the columns represent *m* measurements or state of a system at different time steps from simulations. The columns ($\mathsf{x}_1 - \mathsf{x}_m$) are also referred to as snapshots.

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathsf{x}_1 & \mathsf{x}_2 & \cdots & \mathsf{x}_m \\ | & | & & | \end{bmatrix} \tag{5.1}$$

The SVD of matrix **X** is a unique decomposition of the matrix into three components.

$$\mathbf{X} = \mathbf{U}\Sigma V^* \tag{5.2}$$

Where $\mathbf{U} \in \mathbb{C}^{n\times n}$ and $\mathbf{V} \in \mathbb{C}^{n\times n}$ are unitary matrices with orthonormal columns and $\Sigma \in \mathbb{R}^{n\times m}$ is a diagonal matrix with non-negative values known as singular values.

The singular values are the square roots of the eigenvalues of covariance matrices $\mathbf{X}^T X$ and $\mathbf{X}\mathbf{X}^T$ and are arranged in descending order providing an indication on the significance of transformations described by matrix **X**. Matrices **V** and **U** contain the eigenvectors from the decompositions of $\mathbf{X}^T X$ and $\mathbf{X}\mathbf{X}^T$, respectively. The left singular vectors, columns of **U**, capture the patterns among the rows of A whereas the right singular vectors, columns of **V**, capture patterns among columns of A. In high dimensional data, often $n > m$, resulting in zero elements on the diagonal of the matrix Σ. The exact representation of **X** can done using the economy SVD [27].

One of the most useful applications of the SVD is an accurate low-rank approximation of a data matrix. As the SVD provides a hierarchy of of vectors, a rank-*r* approximation can be obtained by expressing the matrix as a sum of rank-one matrices.

$$\mathbf{X} \approx \tilde{\mathbf{X}} = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^* = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^* + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^* + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^* = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{V}^* \tag{5.3}$$

There are various theories on choosing an optimal truncation rank-*r*. Often it involves examining the rate of decay of singular values and determining a threshold point [28] or examining error sizes between reconstructed data and original data for different rank *r* values.

The terms dimensionality reduction and low-rank approximation are often used interchangeably and are similar. In the context of SVD applied on spatio-temporal data, dimensionality reduction would refer to reducing variables in a dataset by selecting the most important spatial and temporal patterns, whereas rank truncation refers to approximating the original matrix with a lower-rank matrix by choosing a rank *r* that determines the number of rank-one matrices to represent the original data.

5.2. Proper orthogonal decomposition

The SVD has been applied in a range of applications, varying from image compression to signal processing, spectral clustering or natural language processing. The application of the SVD to provide an interpretable, data-driven, hierarchical coordinate system representing high-dimensional correlated data is often called principal component analysis (PCA). PCA is the general term most commonly used for application in statistics and data analysis in different fields. The application of the SVD in meteorology and climatology to identify spatial patterns and their corresponding temporal evolution, is referred to as determining empirical orthogonal functions (EOF). The application of SVD in fluid dynamics to represent both spatial and temporal modes of a system is known as proper orthogonal decomposition (POD) [29]. PCA, EOF, and POD are all methods that have the goal to reduce the dimension and extract dominant features of data and are based on the singular value decomposition. Often, it is merely a matter of the application in which the method is applied. Throughout the remaining portion of this research, wherein the focus lies on both spatial and temporal patterns, and the atmosphere is considered a dynamical system, the term ”POD” will be used in this report to denote the method of applying the SVD.

The SVD applied on fluid flows involves removing the temporal mean flow to gain insight into dominant fluctuations from the mean. In a tutorial [29] an example is given on the analysis of flow past a cylinder. Firstly, two dimensional velocity data is concatenated into a matrix of snapshots, in which a column represents an instance in time. The average velocities at every snapshot are removed to only analyse velocity fluctuations. The SVD method is performed on the matrix of velocity fluctuations. A n-dimensional example is explained to show a hierarchy of POD modes for higher dimensional data [29]. This example shows how the SVD can be applied effectively to higher dimensional data to extract dominant spatial patterns and their temporal coefficients, and how a multiplication of the first few dominant modes with their time coefficients provides and accurate approximation of the original data.

Many extensions of the SVD/PCA method exists including method to deal with outliers and noise data (robust PCA) and promoting sparsity to force as many coefficients of principal compontents to zero (sparse PCA). To handle large datasets, the randomised SVD has been developed, which involves projecting the original data on a random subspace to end up with a smaller matrix to save computational time [27].

5.3. Dynamic mode decomposition

As the evolution of chemical species concentrations in the atmosphere can be seen as a dynamical system, data-driven approaches to modelling discrete-time dynamical systems can be explored as well. Dynamic mode decomposition (DMD) was developed in fluid dynamics, and like the POD aims to identify coherent spatio-temporal structures from high-dimensional data. The DMD algorithm identifies the best-fit linear dynamical system that described the advancement of measurements forward in time [27]. The DMD thus provides a dimensionality reduction and a model of how these modes evolve in time, as oppose to the POD that represent how the spatial modes evolve as a time series of coefficients.

Similarly to the POD, high-dimensional data can be arranged in a matrix of snapshots. These are then arranged as matrices $\mathbf{X}$ and $\mathbf{X}'$, with matrix $\mathbf{X}'$ containing not containing the first snapshot and including the final snapshot.

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathsf{x}(t_1) & \mathsf{x}(t_2) & \cdots & \mathsf{x}(t_{m-1}) \\ | & | & & | \end{bmatrix} \tag{5.4}$$

$$\mathbf{X}' = \begin{bmatrix} | & | & & | \\ \mathsf{x}(t_2) & \mathsf{x}(t_3) & \cdots & \mathsf{x}(t_m) \\ | & | & & | \end{bmatrix} \tag{5.5}$$

The DMD algorithm finds the best-fit linear operator $\mathbf{A}$ that relates these snapshot matrices.

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X} \tag{5.6}$$

The exact DMD algorithm is computed using the following steps [27]:

1. Compute the truncated singular value decomposition of matrix.

$$\mathbf{X} \approx \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{V}^* \tag{5.7}$$

2. The best fit matrix $\mathbf{A}$ can be obtained by computing the pseudo inverse of $\mathbf{X}$. As only the first $\mathbf{r}$ eigenvalues and eigenvectors are of interest, $\mathbf{A}$ can be projected on the POD modes $\mathbf{U}$:

$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}^*\mathbf{A}\tilde{\mathbf{U}} = \tilde{\mathbf{U}}^* X'\tilde{\mathbf{V}}\tilde{\Sigma}^{-1} \tag{5.8}$$

3. The eigen decomposition of $\tilde{\mathbf{A}}$ is computed to determine the DMD eigenvalues. The eigenvalues of $\mathbf{\Lambda}$ are the same as the eigenvalues of $\mathbf{A}$.

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda} \tag{5.9}$$

4. The DMD modes Φ are reconstructed with the eigenvectors $\mathbf{W}$ and the time-shifted snapshot matrix of the original data $\mathbf{X}'$.

$$\mathbf{\Phi} = \mathbf{X}'\mathbf{V}\tilde{\Sigma}^{-1}\mathbf{W} \tag{5.10}$$

In conclusion, DMD is a method that expands a system by means of data-driven spectral decomposition into spatial modes, with associated eigenvalues and amplitude that indicate the behaviour of these modes. Again, as it is data-driven, no knowledge of governing equations is required. It is important to realise that DMD is a linear regression algorithm, and the challenge of non-linear systems perhaps

requires alternative methods. One of the advantages of DMD as oppose to POD is that it is better suited for future state prediction.

As DMD has been applied in fluid dynamics but also many other applications resulting in many extensions have been developed. For instance *multi-resolution* DMD has been explored to capture systems that exhibit transient dynamics. The multi resolution DMD decomposes the dynamics into different time scales to isolate transient and irregular patterns. This was applied on global climate data to observe irregular patterns such as the El Niño phenomenon [27]. This is another reason why exploring this linear representation of high-dimensional data could be an interesting application in this research.

Other notable extensions to the methods that have been proposed include an optimised DMD that effectively handles noisy data using variable projection method for nonlinear least squares problems [30] and the bagging optimised DMD in which the algorithm is run on random subsamples of the original dataset resulting in more robust solutions [31].

5.4. Sparse identification of non-linear dynamics

Identifying nonlinear structure from data is more challenging. For many situations, the best-fit linear model obtained using DMD serves effective, but other complex systems portray more nonlinear dynamics. The sparse identification of nonlinear dynamics (SINDy) algorithm seeks to approximate the dynamics of a dynamic system in the following generalised linear model with the fewest non-zero terms ξ. The original algorithm was presented by Brunton et al. [32].

$$\mathbf{f}(\mathbf{x}) \approx \sum_{k=1}^{p} \theta_k(\mathbf{x})\xi_k = \mathbf{\Theta}(\mathbf{x})\boldsymbol{\xi} \tag{5.11}$$

The large library of candidate nonlinear functions Θ is shown in Figure 5.1, which shows an example schematic of SINDy applied to the Lorenz system. To determine a sparse model, a convex l_1-regularised regression is used, with a sparsity parameter λ that can be altered to achieve a desired accuracy and sparsity.

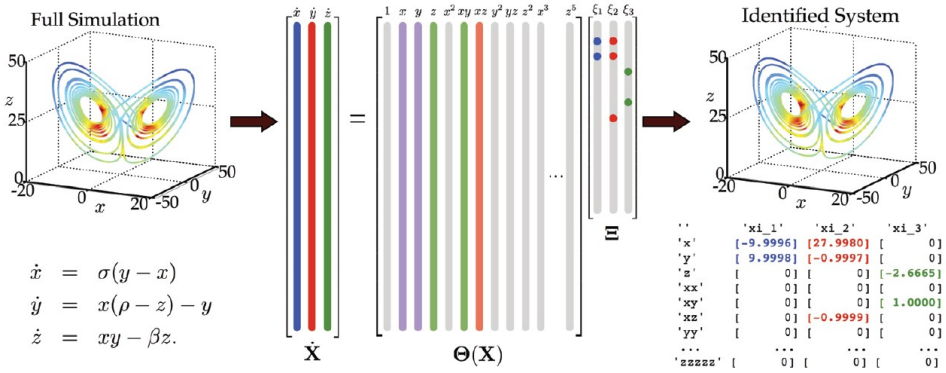


Figure 5.1: Schematic of sparse identification of nonlinear dynamics algorithm [27]

The SINDy algorithm can be used in combination with spatial modes obtained from the POD algorithm. The nonlinear temporal evolution of the POD modes could be explained with an equation found using the SINDy algorithm. In this manner SINDy based on POD coefficients has been applied to fluid flows to create reduced order models.

5.5. Other methods

Machine learning (ML) methods have been increasingly developed in the last decades, with all problems focused on estimating associations between inputs, outputs and parameters of a system using

a number of observations. In weather and climate models ML is extending from replacement of components of models to improve accuracy and reduce computational time to full replacement of models [33]. One such approach for full model replacement is the use of deep learning and neural networks, that do not require understanding of the processes, but only rely on sufficient training data. Neural networks are fundamental nonlinear function approximators, and with sufficient layer any function can be approximated [34]. However, the 'black box' idea behind many ML methods result in it not being widely applied in research, especially in weather and climate [33]. Other methods such as decision trees, random forest, clustering, might all be useful methods for other classification and regression problems.

5.6. Conclusion

Even though dimensionality reduction technique still contain may limitations such as overfitting possibility, linearity assumptions and sensitivity to outliers and missing value, by taking into account these limitations in implementation some of these can be overcome. Machine learning methods such as neural networks might prove to be efficient in representing full models at a much lower computational cost, but the 'black box' machine learning methods will limit interpretability of the system. The goal of this research is both to learn, reconstruct and predict the underlying dynamical system of chemical interactions in the atmosphere at lower computational cost. Machine learning methods might be effective at creating a computationally cheaper model capable of predicting future states of a system, but without insights into the underlying dynamics, these methods will unlikely be a starting point in this research. POD, DMD and SINDy methods will most likely provide an efficient basis for a reduced-order model that is interpretable.

6

Overview current applications

This chapter gives an overview of the applications of the dimensionality reduction techniques described in chapter 5, as well as applications of extensions of these methods and other methods. Machine learning (ML) methods, and other forms of dimensionality reduction that were not touched upon in chapter 5 are also discussed, as they often also have the goal of representing a high dimensional system as a simplified, sparse model. This chapter gives an overview of the literature on these methods in fluids flows in section 6.1 and climate modelling in section 6.2, highlighting their relevant conclusions and limitations. Section 6.3 provides clarification on why certain methods are deemed more appropriate as an initial framework for this research.

6.1. Applications on fluid flows

An overview of various machine learning methods including neural networks, autoencoders applied on fluid mechanics and their performance compared to POD modes is provided by Brunton et al. [34]. The paper highlights the benefits of using ML including reduced order modelling, control, optimisation and regression. It also stresses its benefit for dimensionality reduction and identifying low-dimensional manifolds (dominant structures) to increase understanding. In a different overview [35] data-based and operator based modal decomposition techniques are compared. The overview show the benefit of comparing and interpreting POD and DMD results to understand different aspects of the flow field, and stresses that these modes can be used to develop reduced order model. The same authors provide a more recent study in [36], providing an example of extensive comparison between POD/DMD and other methods examining various simulations: past a cylinder, airfoil wake and cavity flow. The analysis again shows clear interpretation of modes to provide physical insights.

Extracting linear and nonlinear spatial and temporal modes in fluid dynamics is also performed using autoencoders [37] or with convolutional neural networks [38]

The initial analysis of fluid flows mainly concerns interpreting the modes from POD and DMD analysis. The next step often involves predicting future states of dynamical system, for which the DMD modes with stable associated frequencies can serve as a basis [39].

Another method for dimensionality reduction and simplifying dynamics is the use of *projection onto nonlinear manifolds*. This involves finding a lower-dimensional representation of the data that captures the same dynamics of the system. Modelling the nonlinear correlations between the temporal coefficients of POD modes can be done using the SINDy algorithm or using sparse polynomial regression to learn an interpretable dynamical system model from time series of the mode coefficients, as presented by Callaham et al. [40]. The first method, models the nonlinear correlations between the modal coefficients as an *invariant manifold reduction* on a cavity flow example. The paper shows that this method could be applicable to periodic flows and not chaotic and turbulent flows, creating a low-dimensional and stable system that models underlying dynamics. Therefore, if the coefficients of POD modes of

atmospheric data demonstrate significant seasonal behaviour, and not random evolution of time coefficients, exploring this method can be interesting.

The SINDy algorithm is useful for for data driven discovery of equations in [41] and [42]. The more interesting applications of the SINDy is the use to estimate a nonlinear model of dynamical systems. Foster et al. [43] apply the SINDy to the wake of an oscillating cylinder, using 10 mode POD projection resulting in accurate reconstructions.

6.2. Applications on atmospheric chemistry data

An overview of the use of statistical and machine learning methods in aviation environmental analysis is provided by Z. Gao and D. Mavris [44].

[45] produces low-rank decompositions of 1-year simulation of global spatio-temporal atmospheric chemistry data. To reduce computational cost, the *randomised SVD* is used which makes use of projecting the original data matrix on a random matrix to obtain a smaller matrix. This research uses *non-negative matrix factorisation* to restrict the data to include only positive chemical concentrations. Sparse PCA is used to zero out very small concentrations so that modes from the SVD highlight significant contributions to dynamics. This is a great example of how extensions to known methods are successful in extracting major features from atmospheric chemistry data, and can create accurate reconstructions with few modes. It found it was possible to represent 99% of the spatio-temporal data of global ozone distribution with just 20 modes of the randomised SVD. The application of dimensionality reduction technique provided effective in reconstructing GEOS-Chem data for species of interest, and concludes it being an effective basis for a reduced-order model.

In [46] a simplified model a an atmospheric chemistry model is created that operates orders of magnitude faster faster than the original model. The number of modelled chemical species is compressed by 80% without decreasing accuracy. An encoder-operator-decoder neural network is used to reduce dimensionality and compress features, and uses *recurrent training* to recurrently integrate the chemical system to predict its evolution. It is applied to important air quality predictions (concentrations of ozone and particulate matter) over a period of several days. The results show that the simulations become numerically unstable after a longer period of time, and other features such as advection are not incorporated into the model. Therefore it remains uncertain whether the described methodology is suitable for chemistry models that require predictions over a longer period of time.

Again in the context of air quality modelling, Keller et al. [47] explores different methods to overcome the computational costs of simulating complex chemical and dynamical processes of pollutants. One method explored is the use of a *random forest regression model* as a chemical solver, which shows promising results that deviate less than 10% from the original model. The research also discusses the use of randomised matrix decomposition, sparse sampling and reduced order modelling concluding that they are suitable methods to provide a low-rank interpretable system that can be analysed much more efficiently.

Other works also explore the use of deep learning or other machine learning methods for spatio-temporal atmospheric data: such as applying random forest regression to gas-phase chemistry calculation within the GEOS-Chem model [48], or using clustering techniques to to identify patterns in transport of aviation induced NO_x and their short-term ozone radiative forcing [13], or machine learning to predict output of climate modes using random forest regression [49]. Even though examining the use of these methods is valuable, their time scales and goals vary. There are also examples of deep learning applied on empirical orthogonal functions obtained from high-dimensional atmospheric data that is then used to create a prediction model [50].

An extensive overview of applications of methods on analysis of wake behind a cylinder is provided by J. Callaham, showing a succesful use of SINDy algorithm for this application [51].

6.3. Conclusion and challenges

The literature overview shows promising results for applying statistical methods on global atmospheric chemistry data. A trend visible in the various applications, is the desired accuracy and output level on which the methods are evaluated. In the case of modelling changes in concentrations of chemical species, the dimension along which this is desired and degree of accuracy will determine whether certain methods are suitable. For instance, predicting the total global ozone change could be predicted by many methods, whereas if ozone changes at exact altitudes and latitudes are wanted other methods might be more useful.

Literature also shows that high dimensional atmospheric and fluid flow data often contain dominant low dimensional structures that are useful for both feature extraction and reconstruction of the data with few modes. However, in order to create a prediction model that allows to predict future states of a known system or predict the response of a system at different initial locations, will require the exploration of the DMD or machine learning methods.

From the following literature was concluded that the starting point of the analysis of this research, would be the application of statistical dimensionality reduction methods (those explained in more detail in chapter 5) to extract dominant features and determine their effectiveness in reconstructing the data. A future step would include examining suitable methods to use the modes from dimensionality reduction as a basis for a prediction model.

7

Conclusions

Both the renewed recent interest in supersonic aircraft development as well as the need to renewal of related environmental regulations have created the need to further explore non-CO_2 effects of supersonic aviation emission to update regulations. Although there are multiple sources of existing knowledge on the impacts of emitting various substances, the quantification of the changes (especially in column ozone) vary per study. Specifically for regulatory purposes it is desirable to gain more insight into the effects of individual flight scenarios, to understand the impact of emitting at specific latitudes and altitudes on climate.

Understanding of the current knowledge on effects of emissions of supersonic aircraft showed that the largest concerns associated with the climate effects of non-CO_2 are the stratospheric water vapour and changes in ozone concentration. The latter is a concern as the extent to which ozone depletion in the upper stratosphere and ozone production. The responses of changes in concentration of ozone are sensitive to altitude and latitude of emission, and vary seasonally indicating that monthly concentration changes along dimensions of altitude and latitude is an appropriate minimum level of accuracy that is examined for changes in ozone concentration.

An explanation of the various data driven methods, their strengths and drawbacks and an overview of applications of these methods in literature support these two methods. The dimensionality reduction techniques (POD/DMD) are suitable for high-dimensional data, and can provide insights into dominant spatial and temporal patterns underlying the atmospheric chemistry data. With few modes obtained from these methods, accurate but low dimensional reconstructions of the original data can be made. Both DMD and POD modes in combination with SINDy algorithm to describe the nonlinear evolution of modes can both be methods to explore to model future states of a system given an initial condition. Other methods, such as the use of neural networks, autoencoders and random forest regressions have also been examined, but they often require significant amounts of training data. Obtaining this training data from GEOS-Chem is a computationally expensive process, and thus does not seem a reasonable starting point in this research.

The result of this literature study is the proposed research outline described in chapter 2, which gives a step-wise approach to finding out whether data-driven reduced-order modelling methods are suitable for modelling atmospheric chemistry.

References

[1] D. S. Lee et al. “The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018”. In: *Atmospheric Environment* 244 (Jan. 2021). ISSN: 18732844. DOI: 10.1016/J.ATMOSENV.2020.117834.

[2] Ashima Sethi. *The Return of Supersonic Travel: What to Expect in the Coming Years*. June 2023.

[3] Boom Supersonic. *Sustainability Report*. Tech. rep. 2022.

[4] Simon Calder. *Twenty years after Concorde’s final touchdown, will supersonic passenger flights ever return?* 2023.

[5] Sebastian D. Eastham Raymond L. Speth and Florian Allroggen Prakash Prashanth. *Global Environmental Impact of Supersonic Cruise Aircraft in the Stratosphere*. Tech. rep. Cambridge, Massachusetts: Massachusetts Institute of Technology, 2021, p. 172.

[6] Sigrun Matthes et al. “Review: The Effects of Supersonic Aviation on Ozone and Climate”. In: *Aerospace 2022, Vol. 9, Page 41* 9.1 (Jan. 2022), p. 41. ISSN: 2226-4310. DOI: 10.3390/AEROSPACE9010041.

[7] V Grewe et al. *Climate impact of supersonic air traffic: an approach to optimize a potential future supersonic fleet-results from the EU-project SCENIC*. Tech. rep. 2007, pp. 5129–5145.

[8] Jun Zhang et al. “Potential Impacts of Supersonic Aircraft Emissions on Ozone and Resulting Forcing on Climate: An Update on Historical Analysis”. In: *Journal of Geophysical Research: Atmospheres* 126.6 (Mar. 2021). ISSN: 21698996. DOI: 10.1029/2020JD034130.

[9] Peter L. Ward. “Ozone Depletion Explains Global Warming”. In: *Current Physical Chemistry* 6.4 (May 2017), pp. 275–296. ISSN: 18779468. DOI: 10.2174/1877946806999160629080145.

[10] *Tropopause*. Oct. 2016.

[11] *Basic Ozone Layer Science | US EPA*.

[12] L L N ; Guarieiro et al. “Artigo Use and Application of Photochemical Modeling to Predict the Formation of Tropospheric Ozone”. In: ().

[13] Jin Maruhashi et al. “Transport patterns of global aviation NOxand their short-term O3radiative forcing - a machine learning approach”. In: *Atmospheric Chemistry and Physics* 22.21 (Nov. 2022), pp. 14253–14282. ISSN: 16807324. DOI: 10.5194/ACP-22-14253-2022.

[14] H. Bönisch et al. “On the structural changes in the Brewer-Dobson circulation after 2000”. In: *Atmospheric Chemistry and Physics* 11.8 (2010), pp. 3937–3948. ISSN: 16807316. DOI: 10.5194/ACP-11-3937-2011.

[15] Jun Zhang et al. “Stratospheric Ozone and Climate Forcing Sensitivity to Cruise Altitudes for Fleets of Potential Supersonic Transport Aircraft”. In: *Journal of Geophysical Research: Atmospheres* 126.16 (Aug. 2021), e2021JD034971. ISSN: 2169-8996. DOI: 10.1029/2021JD034971.

[16] Joanna D. Haigh. “The Sun and the Earth’s Climate”. In: *Living Reviews in Solar Physics* 4.1 (Dec. 2007), pp. 1–64. ISSN: 16144961. DOI: 10.12942/LRSP-2007-2/FIGURES/47.

[17] Volker Grewe. “Impact of lightning on air chemistry and climate”. In: *Lightning: Principles, Instruments and Applications: Review of Modern Lightning Research* (2009), pp. 537–549. DOI: 10.1007/978-1-4020-9079-0{_}25/COVER.

[18] Sebastian D. Eastham et al. “Impacts of a near-future supersonic aircraft fleet on atmospheric composition and climate”. In: *Environmental Science: Atmospheres* 2.3 (Feb. 2022), pp. 388–403. ISSN: 26343606. DOI: 10.1039/d1ea00081k.

[19] Guy P Brasseur, John J Orlando, and Geoffrey S Tyndall. “Atmospheric Chemistry and Global Change”. In: (1999), p. 481.

[20] Johannes Pletzer et al. “The climate impact of hydrogen-powered hypersonic transport”. In: *Atmospheric Chemistry and Physics* 22.21 (Nov. 2022), pp. 14323–14354. ISSN: 16807324. DOI: 10.5194/acp-22-14323-2022.

[21] Jun Zhang et al. “Potential Impacts on Ozone and Climate From a Proposed Fleet of Supersonic Aircraft”. In: *Earth’s Future* 11.4 (Apr. 2023), e2022EF003409. ISSN: 2328-4277. DOI: 10.1029/2022EF003409.

[22] Thibaud M. Fritz et al. “Identifying the ozone-neutral aircraft cruise altitude”. In: *Atmospheric Environment* 276 (May 2022), p. 119057. ISSN: 1352-2310. DOI: 10.1016/J.ATMOSENV.2022.119057.

[23] M. O. Köhler et al. “Latitudinal variation of the effect of aviation NOx emissions on atmospheric ozone and methane and related climate metrics”. In: *Atmospheric Environment* 64 (Jan. 2013), pp. 1–9. ISSN: 1352-2310. DOI: 10.1016/J.ATMOSENV.2012.09.013.

[24] G Pitari and E Mancini. *Climatic Impact of Future Supersonic Aircraft: Role of Water Vapour and Ozone Feedback on Circulation*. Tech. rep. 8. 2001, pp. 571–576.

[25] Guy P. Brasseur and Daniel J. Jacob. *Modeling of Atmospheric Chemistry*. Cambridge University Press, Jan. 2017, pp. 1–606. ISBN: 9781316544754. DOI: 10.1017/9781316544754.

[26] *GEOS-Chem Overview*.

[27] Steven L Brunton and J Nathan Kutz. *Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control*. Tech. rep. 2021.

[28] Matan Gavish and David L. Donoho. “The optimal hard threshold for singular values is $4/\sqrt{3}$”. In: *IEEE Transactions on Information Theory* 60.8 (2014), pp. 5040–5053. ISSN: 00189448. DOI: 10.1109/TIT.2014.2323359.

[29] Julien Weiss. “Julien: A Tutorial on the Proper Orthogonal Decomposition”. In: (2019), pp. 17–21. DOI: 10.14279/depositonce-8512.

[30] Travis Askham and J Nathan Kutz. “Variable projection methods for an optimized dynamic mode decomposition”. In: ().

[31] Diya Sashidhar and J Nathan Kutz. “Bagging, optimized dynamic mode decomposition (BOP-DMD) for robust, stable forecasting with spatial and temporal uncertainty-quantification”. In: (2021).

[32] Steven L. Brunton et al. “Discovering governing equations from data by sparse identification of nonlinear dynamical systems”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.15 (Apr. 2016), pp. 3932–3937. ISSN: 10916490. DOI: 10.1073/PNAS.1517384113/SUPPL{_}FILE/PNAS.1517384113.SAPP.PDF.

[33] Catherine O De Burgh-Day and Tennessee Leeuwenburg. “Machine learning for numerical weather and climate modelling: a review”. In: *Geosci. Model Dev* 16 (2023), pp. 6433–6477. DOI: 10.5194/gmd-16-6433-2023.

[34] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. “Machine Learning for Fluid Mechanics”. In: *Annual Review of Fluid Mechanics* 52 (May 2019), pp. 477–508. ISSN: 00664189. DOI: 10.1146/annurev-fluid-010719-060214.

[35] Kunihiko Taira et al. “Modal analysis of fluid flows: An overview”. In: *AIAA Journal* 55.12 (Oct. 2017), pp. 4013–4041. ISSN: 1533385X. DOI: 10.2514/1.J056060/ASSET/IMAGES/LARGE/FIGURE17.JPEG.

[36] Kunihiko Taira et al. “Modal analysis of fluid flows: Applications and outlook”. In: *AIAA Journal* 58.3 (Oct. 2020), pp. 998–1022. ISSN: 00011452. DOI: 10.2514/1.J058462/ASSET/IMAGES/LARGE/FIGURE19.JPEG.

[37] Kai Fukami, Taichi Nakamura, and Koji Fukagata. “Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data”. In: *Department of Mechanical Engineering, Keio University, Yokohama 223-8522,* (Aug. 2020).

[38] Takaaki Murata, Kai Fukami, and Koji Fukagata. “Nonlinear mode decomposition with convolutional neural networks for fluid dynamics”. In: *Journal of Fluid Mechanics* 882 (June 2019). DOI: 10.1017/jfm.2019.822.

[39] M. A. Khodkar and Pedram Hassanzadeh. “A data-driven, physics-informed framework for forecasting the spatiotemporal evolution of chaotic dynamics with nonlinearities modeled as exogenous forcings”. In: *Journal of Computational Physics* 440 (Sept. 2021), p. 110412. ISSN: 0021-9991. DOI: 10.1016/J.JCP.2021.110412.

[40] Jared L. Callaham, Steven L. Brunton, and Jean-Christophe Loiseau. “On the role of nonlinear correlations in reduced-order modeling”. In: *Journal of Fluid Mechanics* 938 (June 2021). DOI: 10.1017/jfm.2021.994.

[41] Kathleen Champion et al. “Data-driven discovery of coordinates and governing equations”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.45 (Nov. 2019), pp. 22445–22451. ISSN: 10916490. DOI: 10.1073/PNAS.1906995116/SUPPL{_}FILE/PNAS.1906995116.SAPP.PDF.

[42] Patrick A K Reinbold, Daniel R Gurevich, and Roman O Grigoriev. “Using noisy or incomplete data to discover models of spatiotemporal dynamics”. In: (2020). DOI: 10.1103/PhysRevE.101.010203.

[43] Joshua A. Foster et al. “Estimating a sparse nonlinear dynamical model of the flow around an oscillating cylinder in a fluid flow using SINDy”. In: *IFAC-PapersOnLine* 54.20 (Sept. 2022), tba. ISSN: 24058963. DOI: 10.1016/J.IFACOL.2021.11.149.

[44] Zhenyu Gao and Dimitri N. Mavris. “Statistics and Machine Learning in Aviation Environmental Impact Analysis: A Survey of Recent Progress”. In: *Aerospace 2022, Vol. 9, Page 750* 9.12 (Nov. 2022), p. 750. ISSN: 2226-4310. DOI: 10.3390/AEROSPACE9120750.

[45] Meghana Velegar et al. “Scalable diagnostics for global atmospheric chemistry using Ristretto library (version 1.0)”. In: *Geoscientific Model Development* 12.4 (Apr. 2019), pp. 1525–1539. ISSN: 19919603. DOI: 10.5194/GMD-12-1525-2019.

[46] Makoto M. Kelp et al. “Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System”. In: *Journal of Geophysical Research: Atmospheres* 125.23 (Dec. 2020). ISSN: 21698996. DOI: 10.1029/2020JD032759.

[47] Christoph A. Keller et al. “Machine learning and air quality modeling”. In: *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017* 2018-January (July 2017), pp. 4570–4576. DOI: 10.1109/BIGDATA.2017.8258500.

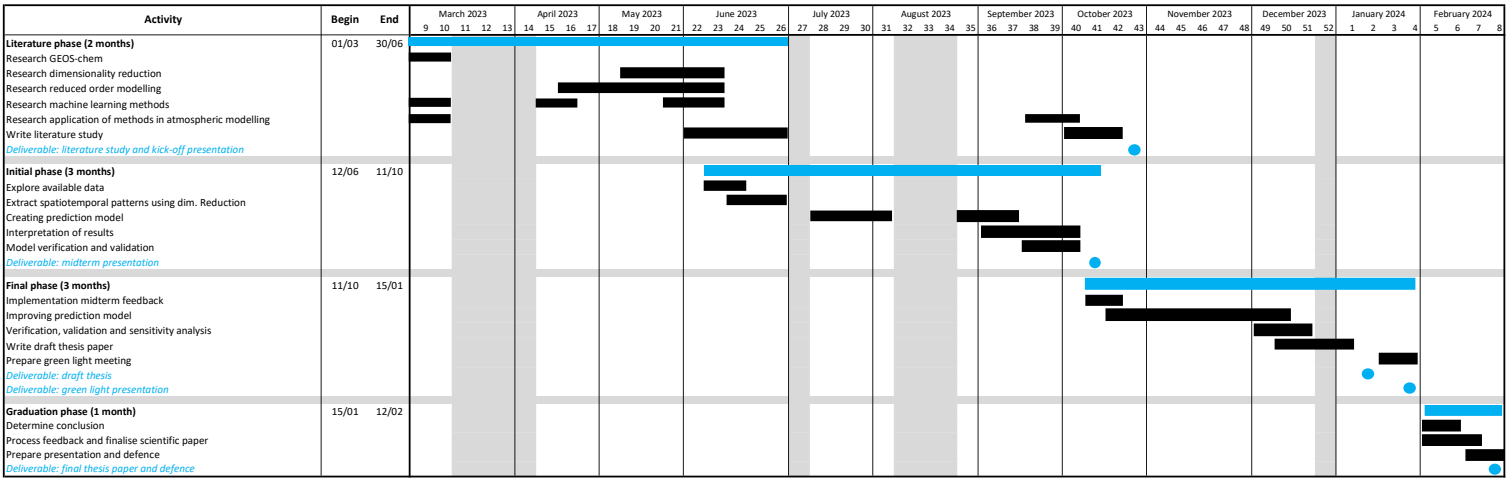
[48] A. Christoph Keller and J. Mat Evans. “Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10”. In: *Geoscientific Model Development* 12.3 (Mar. 2019), pp. 1209–1225. ISSN: 19919603. DOI: 10.5194/GMD-12-1209-2019.

[49] Gemma J. Anderson and Donald D. Lucas. “Machine Learning Predictions of a Multiresolution Climate Model Ensemble”. In: *Geophysical Research Letters* 45.9 (May 2018), pp. 4273–4280. ISSN: 19448007. DOI: 10.1029/2018GL077049.

[50] Federico Amato et al. “A novel framework for spatio-temporal prediction of environmental data using deep learning”. In: *Scientific Reports 2020 10:1* 10.1 (Dec. 2020), pp. 1–11. ISSN: 2045-2322. DOI: 10.1038/s41598-020-79148-7.

[51] Jared Callaham et al. *Multiscale model reduction for unsteady fluid flow*. Tech. rep. 2022.

A. Gantt Chart


Activity
Begin
End
March 2023
9 10 11 12 13
April 2023
14 15 16 17
May 2023
18 19 20 21
June 2023
22 23 24 25 26
July 2023
27 28 29 30
August 2023
31 32 33 34 35
September 2023
36 37 38 39
October 2023
40 41 42 43
November 2023
44 45 46 47 48
December 2023
49 50 51 52
January 2024
1 2 3 4
February 2024
5 6 7 8
Literature phase (2 months) 01/03 30/06
Research GEOS-chem
Research dimensionality reduction
Research reduced order modelling
Research machine learning methods
Research application of methods in atmospheric modelling
Write literature study
Deliverable: literature study and kick-off presentation
Initial phase (3 months) 12/06 11/10
Explore available data
Extract spatiotemporal patterns using dim. Reduction
Creating prediction model
Interpretation of results
Model verification and validation
Deliverable: midterm presentation
Final phase (3 months) 11/10 15/01
Implementation midterm feedback
Improving prediction model
Verification, validation and sensitivity analysis
Write draft thesis paper
Prepare green light meeting
Deliverable: draft thesis
Deliverable: green light presentation
Graduation phase (1 month) 15/01 12/02
Determine conclusion
Process feedback and finalise scientific paper
Prepare presentation and defence
Deliverable: final thesis paper and defence