

PHYSICAL SENSING AND PHYSICS-BASED MACHINE LEARNING  
FOR ACTIONABLE ENVIRONMENTAL INSIGHTS

by

John Waczak

APPROVED BY SUPERVISORY COMMITTEE:

---

David J. Lary, Chair

---

Christopher Simmons

---

David Lumley

---

Lindsay King

---

Joseph Izen

Copyright © 2024

John Waczak

All rights reserved

***UPDATE REQUIRED!***

PHYSICAL SENSING AND PHYSICS-BASED MACHINE LEARNING  
FOR ACTIONABLE ENVIRONMENTAL INSIGHTS

by

JOHN WACZAK, BS, PhD

DISSERTATION

Presented to the Faculty of  
The University of Texas at Dallas  
in Partial Fulfillment  
of the Requirements  
for the Degree of

DOCTOR OF PHILOSOPHY IN  
PHYSICS

THE UNIVERSITY OF TEXAS AT DALLAS  
September 2024

## ACKNOWLEDGMENTS

**UPDATE REQUIRED!** Make sure to include lab colleagues \*and\* ActivePure colleagues and a nod to Dr. Cooper, OSU teachers, and friends...

August 2024

PHYSICAL SENSING AND PHYSICS-BASED MACHINE LEARNING  
FOR ACTIONABLE ENVIRONMENTAL INSIGHTS

John Waczak, PhD  
The University of Texas at Dallas, 2024

Supervising Professor: David J. Lary, Chair

**UPDATE REQUIRED!**

## TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . . .	v
ABSTRACT . . . . .	vi
LIST OF FIGURES . . . . .	xi
LIST OF TABLES . . . . .	xiv
CHAPTER 1 INTRODUCTION . . . . .	1
1.0.1 Global Change and the Need For Improved Sensing . . . . .	2
1.0.2 Attention Towards Indoor Air Quality . . . . .	2
1.0.3 Physics Informed Machine Learning . . . . .	9
1.0.4 Dissertation Overview . . . . .	9
CHAPTER 2 PHYSICAL SENSING . . . . .	13
2.1 Coordinated Robotic Teams . . . . .	13
2.2 A Low Cost Sensor Network For Air Quality Monitoring . . . . .	13
2.3 PCO and The HEART Chamber . . . . .	14
CHAPTER 3 MACHINE LEARNING METHODS . . . . .	18
3.1 Adjoint Methods for Optimization . . . . .	19
3.1.1 Adjoint Methods for Linear Systems . . . . .	19
3.1.2 Adjoint Methods for Nonlinear Systems . . . . .	20
3.1.3 Adjoint Methods for ODEs . . . . .	22
3.2 Supervised Regression Techniques . . . . .	24
3.2.1 Neural Networks . . . . .	24
3.2.2 Gaussian Process Regression . . . . .	24
3.2.3 Decision Trees . . . . .	43
3.3 Unsupervised Classification . . . . .	43
3.3.1 Self Organizing Maps . . . . .	43
3.3.2 Generative Topographic Maps . . . . .	48
3.4 Uncertainty Quantification via Conformal Prediction . . . . .	57
3.4.1 Conformalizing Quantile Regression . . . . .	59
3.4.2 Conformalizing Scalar Uncertainty Estimates . . . . .	60

3.5	Data Assimilation . . . . .	62
3.5.1	Kalman Filter . . . . .	65
3.5.2	Extended Kalman Filter . . . . .	71
3.5.3	continuous-discrete Extended Kalman Filter . . . . .	72
3.5.4	3d-Var . . . . .	75
3.5.5	4d-Var . . . . .	77
CHAPTER 4 AN AUTONOMOUS ROBOTIC TEAM FOR THE RAPID CHARACTERIZATION OF NOVEL ENVIRONMENTS . . . . .		80
4.1	Rapid Georectification and Processing of Pushbroom Hyperspectral Imagery . . . . .	80
4.1.1	Autonomous Aerial Vehicle . . . . .	84
4.1.2	Real time processing of HSI imagery . . . . .	84
4.1.3	Georectification Procedure . . . . .	86
4.2	Results . . . . .	86
4.3	Discussion . . . . .	87
4.4	Supervised Regression with Uncertainty Quantification . . . . .	89
4.4.1	Hyperspectral Imaging and Remote Sensing . . . . .	89
4.4.2	In-situ Measurements / Robot Teams . . . . .	90
4.4.3	Julia Programming Language . . . . .	90
4.5	Materials and Methods . . . . .	90
4.5.1	Autonomous Collection and Processing of Hyperspectral Images . . . . .	90
4.5.2	Georectification . . . . .	90
4.5.3	In-situ Data Collection via Robotic Sentinels . . . . .	91
4.5.4	Exploratory Data Analysis . . . . .	92
4.5.5	Model Analysis . . . . .	93
4.6	Results . . . . .	97
4.6.1	CO . . . . .	97
4.7	Methods for Unsupervised Classification of Hyperspectral Scenes in Novel Environments . . . . .	97
CHAPTER 5 TIME SERIES METHODS FOR AIR QUALITY . . . . .		101
5.1	Time-Series Methods for Uncertainty Quantification . . . . .	101

5.1.1	Uncertainty Quantification with Temporal Variograms . . . . .	103
5.1.2	Next Steps . . . . .	106
5.2	Physics Informed modeling techniques for Air Quality Data . . . . .	107
5.2.1	A Hybrid HAVOK UDE Approach . . . . .	108
5.2.2	Hamiltonian Neural Networks . . . . .	117
CHAPTER 6	A CHEMICAL DATA ASSIMILATION FRAMEWORK FOR INDOOR AIR QUALITY . . . . .	125
6.1	Physics of Chemical Reactions: Chemical Reaction Kinetics . . . . .	125
6.1.1	Overview . . . . .	125
6.1.2	Chemical Equilibrium and the Law of Mass Action . . . . .	126
6.1.3	Reaction Rate Laws . . . . .	129
6.2	Photolysis . . . . .	134
6.3	Summary of Chemical Mechanism Kinetics . . . . .	134
6.4	Characterization of Photolysis . . . . .	136
6.4.1	Absorption Cross Sections . . . . .	136
6.4.2	Quantum Yields . . . . .	136
6.4.3	Irradiance Spectra . . . . .	136
6.4.4	Photolysis Rate Determination . . . . .	137
6.4.5	Next Steps . . . . .	138
6.5	HEART Chamber Data . . . . .	138
6.6	Chemical Data Assimilation . . . . .	139
6.7	Next Steps . . . . .	140
CHAPTER 7	LIMITATIONS AND FUTURE WORK . . . . .	149
7.1	Robot Team . . . . .	149
7.1.1	Super Resolution . . . . .	149
CHAPTER 8	CONCLUSIONS . . . . .	150
APPENDIX A	CHEMICAL REACTION MECHANISMS . . . . .	151
A.1	Simple Ion Mechanism . . . . .	151
A.1.1	Bimolecular Reactions . . . . .	151

A.1.2 Trimolecular Reactions . . . . .	206
A.1.3 Photolysis Reactions . . . . .	226
A.2 Master Chemical Mechanism . . . . .	234
REFERENCES . . . . .	235
BIOGRAPHICAL SKETCH . . . . .	241
CURRICULUM VITAE	

## LIST OF FIGURES

2.1	Details of comprehensive chemical and aerosol testing for IAQ using HEART: Holistic Environmental Aerosol and Reactive Component Auto Response Testing.	14
2.2	Visual overview of the various infrastructure components of the ActivePure lab. Note that the lab uses multiple instruments of those shown to accommodate the stated range of analytes. . . . .	15
3.1	A standard linear regression fit on some noisy data. . . . .	26
3.2	An example of polynomial regression for which we fit a quadratic polynomial to some noisy data. . . . .	31
3.3	A Gaussian Process fit to the training data illustrating the predication (means) and a $\pm 2\sigma$ uncertainty interval. We can clearly see how the uncertainty is larger for inputs far away from supplied training data. . . . .	41
3.4	Left: the original Gaussian Process obtained with our default choice of kernel parameters. Right: A much better Gaussian Process obtained after performing hyper-parameter optimization on the kernel parameters. Note how the mean function still does an excellent job fitting the data points while <i>also</i> minimizing the uncertainty of the fit. . . . .	44
3.5	A Self Organizing Map with nodes configured in a rectangular grid topology with weight vectors randomly initialized. . . . .	46
3.6	A trained Self Organizing Map of $25 \times 25$ cells in a hexagonal topology trained on a dataset consisting of the three colors red, green, and blue. . . . .	49
3.7	An example of conformal prediction for a model estimating a noisy one-dimensional target. The blue shaded region illustrates the learned confidence interval. Image taken from (Altmeyer, 2023) . . . . .	58
4.1	Components of the Autonomous Drone HSI platform. . . . .	84
4.2	Solar Irradiance spectrum captured by downwelling spectrometer. . . . .	85
4.3	Visual representation of scan-line geometry for the drone based hyperspectral imaging platform. . . . .	86
4.4	Annotated view of a hyperspectral data cube showcasing sampled spectra for a variety of constituents. . . . .	87
4.5	This is a figure. Schemes follow the same formatting. If there are multiple panels, they should be listed as: <b>(a)</b> Description of what is contained in the first panel. <b>(b)</b> Description of what is contained in the second panel. Figures should be placed in the main text near to the first time they are cited. A caption on a single line should be centered. . . . .	88

4.6	Description . . . . .	89
4.7	Correlation Matrix for Target variables measured by the boat. . . . .	93
4.8	Correlation Matrix for Radiance measurements versus Target Variables. . . . .	94
4.9	Correlation Matrix for Reflectance measurements versus Target Variables. . . . .	95
4.10	Correlation Matrix for derived measurements versus Target Variables. . . . .	96
4.11	Target distribution of CO at Scotty's Ranch. . . . .	97
4.12	Stratified train-test split for CO. . . . .	98
4.13	(a) Scatterplot for the trained random forest Crudo Oil model. (b) Quantile-Quantile plot for the same fit. . . . .	98
4.14	Ranked feature importance for Crude Oil. Importance is determined via mean absolute SHAP value. . . . .	99
4.15	Unsupervised clustering of hyper-spectral image data using the K-means algorithm.	100
5.1	Time series of particulate matter at size fractions 1.0, 2.5, and 10.0 $\mu m$ captured at a single location over one 24-hour day. . . . .	102
5.2	The empirical variogram and a variety of model fits obtained for the a PM 10.0 single-day time series . . . . .	106
5.3	Comparing the original Lorenz attractor (left) to the embedded attractor learned after performing the SVD (right). Color indicates the time along the trajectory.	112
5.4	Eigenmodes of the embedded attractor extracted from the SVD. . . . .	113
5.5	Heatmap of the linear Koopman operator together with the forcing activation. .	113
5.6	The reconstructed timeseries for $v_1$ via the learned HAVOK model. . . . .	114
5.7	A scatterplot of the resulting HAVOK fit . . . . .	115
5.8	The statistics of the learned forcing function. The sharpness of the distribution (in comparison to a Normal distribution) indicates that the forcing is <i>intermittent</i>	116
5.9	The time series for $v_1$ marked where the forcing function is above a specified threshold. . . . .	117
5.10	The embedded attractor colored by the presence of external forcing. . . . .	118
5.11	The embedded attractor for PM 2.5 time-series data. . . . .	118
5.12	Heatmap of the linear Koopman operator together with the forcing activation. .	119
5.13	The reconstructed time-series for the first three embedding coordinates using the HAVOK fit. . . . .	120

5.14	A zoomed in view of the same time-series reconstruction for the first 2.5 hours showing a very decent fit. Some kind of error appears to be accumulating over time. . . . .	121
5.15	The first embedding coordinate with forcing above a critical threshold identified. . . . .	122
5.16	The original attractor now colored by external forcing above a threshold. . . . .	123
6.1	An illustration of the broad range of reaction time scales from the long-lived nuclear decay to rapid degradation of molecules by photolysis. Figure taken from (Arnaut and Burrows, 2006) . . . . .	126
6.2	The collected absorption cross section data for Ozone, $O_3$ , across a variety of temperatures. . . . .	135
6.3	The distribution of cross section data available for $O_3$ . . . . .	135
6.4	The distribution of subsampled $O_3$ cross section data . . . . .	136
6.5	A scatter plot of the resulting GPR fit evaluated on the training set (green) and a holdout testing dataset (red). . . . .	136
6.6	Predicted $O_3$ cross section as a function of temperature. . . . .	137
6.7	Predicted $O_3$ cross section for two temperatures with the associated uncertainty visualized. At $T = 293$ K, we have many data points and therefore the associated fit uncertainty is small. For $T = 295$ K, there are many fewer data points available above 800 nm leading to a larger fit uncertainty. . . . .	137
6.8	Data for the quantum yield of $O_3$ . Sources of data are far more sparse than for the absorption cross section. . . . .	138
6.9	Distribution for the $O_3$ quantum yield data. . . . .	138
6.10	Resulting GPR fit for $O_3$ quantum yield data. Left: A scatter plot. Right: a quantile-quantile plot. . . . .	139
6.11	Predicted quantum yield as a function of temperature for $O_3$ . Despite the goodness of fit, setting the mean to nearly zero to prevent edge effects causes the GPR to approach in wavelength regions with little data. . . . .	140
6.12	Final quantum yield predictions where we have forced the value to be 1 below a reasonable threshold wavelength. . . . .	141
6.13	A plot of the spectral irradiance captured within a typical office building and averaged over 8 hours. . . . .	142
6.14	A comparison of the absorption cross section, quantum yield, and captured irradiance spectrum used to determine the photolysis rate for an $O_3$ photolysis channel. . . . .	143

## LIST OF TABLES

- |   |   |
|---|---|
| 1.1 Tabular literature review of particulate matter and health outcomes related to PM <sub>10</sub> , PM <sub>2.5</sub> , and ultrafine particulates (UFPs) exposure (based on the literature review of (Lary et al., 2015)). . . . . | 7 |
|---|---|

# CHAPTER 1

## INTRODUCTION

The rapid pace of global change poses a significant and ever-present threat to human prosperity. To facilitate the development of remediation technologies and enable effective mitigation strategies, we must make *data-driven decisions*. However, the limitations posed by the lack of highly available, highly resolved data coupled together with the computational difficulties posed by direct simulation of physics *at scale* severely constrains our ability to make the low uncertainty predictions needed to meaningfully address these challenges. The goal of this applied physics dissertation is to expand the boundary of what is possible in realm of sensing by leveraging machine learning in a *principled way* to produce actionable insights. To that end, this work invokes three key case studies to demonstrate how machine learning can be used effectively to fill to enable

To that end, the guiding question of this work can be summarized succinctly as: *How can we best utilize the measurements we have together with the physics we know to estimate the quantities we care about?*

- first we should discuss global change and the necessity for improved sensing and simulation - justify the use of machine learning in order to fill the gaps where theory provides the motivation between what we measure and what we want to predict but no easily-represented physical relationship exists. - this is a good place to bring in the - we can then discuss the advent of physics-informed machine learning, and better, scientific machine learning to fill in the gaps and not be Snake Oil Salesmen (TM). In particular the role of numerical methods has not been subsumed by machine learning. At best these techniques are complementary and any performance claims need to be carefully benchmarked so as to not oversell the abilities of machine learning. - Many of these emerging methods are promising but have seen little application to noisy, real-world data. One can only examine the Lorentz equations so much before transitioning to actual atmospheric dynamics - We can do a paragraph on the

specific areas we are trying to address: (1) using machine learning to establish the nonlinear relationship between measured spectra and constituent concentrations in water, (2) data-based techniques for evaluating the uncertainty of single sensor time series with application to low cost sensing, (3) physics based models for time series dynamics of Air Quality data (i.e. there will never be enough sensors on a low cost station to explain variability – we would need a car presence sensor, etc... but we can model most of the dynamics and treat the rest as intermittent forcing), (3) how can we fuse sensing and simulation in order to infer the abundance of chemicals below detectable limits (in particular with application to indoor air chemistry).

### **1.0.1 Global Change and the Need For Improved Sensing**

### **1.0.2 Attention Towards Indoor Air Quality**

Humans spend approximately 90% of their time indoors (Finewax et al., 2021). Consequently, the quality of indoor air has a dramatic effect on physical health, cognitive performance, and general well-being (Krebs and Luechinger, 2021; Gao et al., 2021; Carneiro et al., 2021; Ni et al., 2021; Shehab and Pope, 2019). At the beginning of the 20<sup>th</sup> century, the debate about indoor air quality standards centered around quantifying the ideal ventilation rate for a variety of indoor spaces, ultimately leading to the ventilation rate procedure of ASHRAE standard 62. Engineers using this method compute the total outdoor ventilation rate by using the predefined rate-per-person and rate-per-area values for each desired space. Adjusting the ventilation rate is clearly effective for treating stale air and reducing so-called sick-building syndrome, but when dealing with contaminants and airborne pathogens, solely adjusting the ventilation rate falls short.

HVAC systems are typically designed to circulate air throughout a building, bringing in outside air, in order to achieve comfortable thermal conditions while consuming as little energy as possible (Memarzadeh and Xu, 2012). When assigned the secondary task of

maintaining indoor air quality, ventilation operates by the same principles as a laboratory fume hood: airtightness and controlled airflow. To be effective, the ventilation system must move air along a short, unimpeded path toward the exhaust point while preventing unintended leakage (Ng et al., 2015). Despite these assumptions, most buildings are not airtight nor are they well-mixed (Emmerich et al., 2005). Further, the variable nature of indoor human interactions coupled with the plethora of possible room configurations introduces significant design challenges. For example, computational fluid dynamics simulations of the airflow in hospital rooms found that “the most important contributing factor to contaminant transmission in enclosed and mechanically ventilated environment is the path between the contaminant source and the exhaust, not the air changes per hour” (Memarzadeh and Xu, 2012). This observation is widely confirmed across the literature with multiple studies indicating that air changes per hour (ACH) can be safely lowered without increasing concentrations of chemicals of concern above tolerable levels (?Lamping and Muller, Lamping and Muller; ?). In concert with these findings, studies of the transmission of airborne pathogens suggest that “increasing the air change rate removes virus from the source room faster but also increases the rate of exposure in connected rooms. Therefore, slower air change rates reduce infectivity in connected rooms at shorter durations (Pease et al., 2021).

As airborne transmission became the clear mechanism of SARS-CoV-2 spread, multiple studies confirmed the negative consequences of high air change rates. One study observed that when ACH was increased from 1.8 to 12, the time to peak virus concentration in connected rooms decreased from 30 minutes to 11 (Pease et al., 2021). A separate contact-tracing case study found that six students were infected by a teacher in a distant room, despite the use of high-efficiency particulate air filters throughout the building (Turakhia et al., 2021). Indeed, amidst observations of increased HVAC energy consumption by upwards of 128% in one Chinese district, researchers still concluded that “HVAC systems may turn out to be the culprit of microbial contamination in enclosed spaces and deteriorate the environment due to inappropriate design and operation” (Zheng et al., 2021).

Filtration by means of in-room high-efficiency particle air (HEPA) and gas-phase filters suffer from the same limitations as discussed above. Namely, air filters only clean the air they come into contact with. This establishes a dichotomy between air quality goals and energy expenditure: pulling less outdoor air is more efficient, but pulling less air through filtration reduces efficacy (Lamping and Muller, Lamping and Muller). Often, the flow rates needed to have a positive effect are far less than the suggested ventilation rates offering opportunities for energy savings. This is the spirit of the alternative ASHRAE 62 IAQ procedure which seeks to attain acceptable indoor air quality by performance-based design criteria. One challenge of IAQP methods is one should predetermine contaminants of concern together with their acceptable levels in order to regulate building design (Zaatari et al., 2016). Buildings contain significant sources of volatile chemical products such as cleaners, disinfectants, personal care products (PCPs), and paints whose use introduces a variety of potentially harmful compounds into the indoor environment (Finewax et al., 2021). The need to reduce energy use while treating contaminants of concern and infectious microorganisms is the motivation for our response to this RFI. Improved characterization of the built environment together with *active* air treatment technologies like ActivePure provides low-cost means to optimize building energy consumption without sacrificing the quality of our air.

HVAC system operating costs account for a significant portion of the total energy consumption in buildings, which themselves account for 40% of the total energy consumption in the United States (?). In the retail sector alone, ventilation accounts for 8.4% of the total energy consumption and 27.8% of the gas energy consumption (Zaatari et al., 2016). It comes as no surprise, then, that the carbon footprint associated with those activities must decrease in the coming years in order to meet global decarbonization targets.

Furthermore, such proposals to focus on increased ventilation rates can only be realized through significant construction and renovation, and once completed, have a significant increase in running costs due to the energy required, preventing our most vulnerable communities from accessing clean air disproportionately (Lamping and Muller, Lamping and

Muller). The federal government should first recognize that **any solution that involves even a minor increase in cost or carbon footprint is a non-starter for many decision-makers.**

The authors of this response hear this market feedback on a daily basis and are constantly on the lookout for behaviors and actions by these decision-makers that align with this mindset. Changing this mindset will be incredibly challenging, especially given that the majority of Americans believe the worst of the pandemic has passed. Instead, we have the option to accept this reality and work within its constraints to achieve the greatest good. Fortunately, a ‘Clean First’ (enVerid, 2022b,a) approach would give the Federal Government the best chance of making improvements that will result in tangible results in better health and safety for all Americans.

While the passage of the Clean Air Act was a defining moment, and subsequent updates have been invaluable, the recent pandemic has unequivocally demonstrated the need for improved EPA testing and evaluation methodology that characterizes the various components of indoor air quality holistically. This is necessary for the critical evaluation and assessment of built-environment ventilation, filtration, and air-cleaning approaches. Furthermore, given the dual challenges of global pandemics and global change, the carbon footprint of various indoor air quality remediation strategies must be evaluated in order to optimize both indoor air quality and the carbon footprint. If we fail to consider both aspects, we risk inadvertently promoting an increase in health and economic disparities by failing to consider the very real fiscal cost associated with increased air changes. It is time to enable holistic optimization of both indoor air quality and energy (and carbon footprint) impacts.

The need for this dual optimization of indoor air quality and energy consumption goes beyond extreme events such as the COVID-19 pandemic. Without a doubt, the composition of the ambient environment has a significant impact on human health as well as general human physical and cognitive performance. We take this reality as a foundational premise

in areas such as Chemical, Biological, Radiological, and Nuclear (CBRN) defense, but it appears to be all too easily overlooked when it comes to everyday occupational exposures that, when aggregated across society, may be far more significant. For example, society rarely, if ever, considers the impact of chemical emissions from personal care products, bleach, air fresheners, disinfectants, and even human excretions on indoor air quality. Each of these causes our buildings to contain a slew of volatile organic compounds (Finewax et al., 2021), as well as reactive halogen species from bleaches.

Even after a significant initial indoor exposure event may have passed, the health impact associated with prior significant and persistent exposure events may well contribute to long-term health quality. For example, chemical disinfectants such as hypochlorite bleach, have been linked to respiratory ailments and have been observed to produce harmful byproducts such as chloroform, carbon tetrachloride, and N-chloroaldimines, which can take up to 3 hours to be ventilated away after the use the bleach (Finewax et al., 2021; Odabasi, 2008). EPA testing needs to consider a maximal set of components, not a minimal set of only a few criterion pollutants, and we know that COVID-19 was spread through inadequately treated indoor air. It is therefore critical that the EPA testing for indoor air quality in the built environment is rigorous, comprehensively characterizing all aspects of indoor air quality, going beyond the outdoor National Ambient Air Quality Standards (NAAQS). It is clearly time for a strategy that goes well beyond just six criteria air pollutants.

For example, some work in this area identifies a short list of species based on a metric widely used by the World Health Organization, the Disability Adjusted Life Years (DALY) indicator (means ‘loss of life corrected by disability’) (Juginović et al., 2021; Lehtomäki et al., 2018). The DALY indicator is widely used to assess the disease burden on the population and identify disease causes. According to the World Health Organization (WHO), ambient air pollution causes 4.2 million premature deaths worldwide each year. In comparison, the total number of global confirmed COVID-19 deaths in 2020 was 1.9 million, and 3.6 million in

HEALTH OUTCOMES	SHORT-TERM STUDIES			LONG-TERM STUDIES		
	PM <sub>10</sub>	PM <sub>2.5</sub>	UFP	PM <sub>10</sub>	PM <sub>2.5</sub>	UFP
<b>Mortality</b>						
All causes	XXX	XXX	X	XX	XX	X
Cardiovascular	XXX	XXX	X	XX	XX	X
Pulmonary	XXX	XXX	X	XX	XX	X
<b>Pulmonary effects</b>						
Lung function, eg, PEF	XXX	XXX	XX	XXX	XXX	
Lung function growth				XXX	XXX	
<b>Asthma and COPD exacerbation</b>						
Acute respiratory symptoms		XX	X	XXX	XXX	
Medication use			X			
Hospital admission	XX	XXX	X			
<b>Lung cancer</b>						
Cohort				XX	XX	X
Hospital admission				XX	XX	X
<b>Cardiovascular effects</b>						
Hospital admission	XXX	XXX		X	X	
<b>ECG-related endpoints</b>						
Autonomic nervous system	XXX	XXX	XX			
Myocardial substrate and vulnerability		XX	X			
<b>Vascular function</b>						
Blood pressure	XX	XXX	X			
Endothelial function	X	XX	X			
<b>Blood markers</b>						
Pro-inflammatory mediators	XX	XX	XX			
Coagulation blood markers	XX	XX	XX			
Diabetes	X	XX	X			
Endothelial function	X	X	XX			
<b>Reproduction</b>						
Premature birth	X	X				
Birth weight	XX	X				
IUR/SGA	X	X				
<b>Fetal growth</b>						
Birth defects	X					
Infant mortality	XX	X				
Sperm quality	X	X				
<b>Neurotoxic effects</b>						
Central nervous system		X	XX			

**Notes:** X, few studies. XX, many studies. XXX, large number of studies.

**Abbreviations:** UFP, ultrafine particle; PEF, peak expiratory flow; COPD, chronic obstructive pulmonary disease; IUG, intrauterine growth restriction; SGA, small for gestational age.

Table 1.1: Tabular literature review of particulate matter and health outcomes related to PM<sub>10</sub>, PM<sub>2.5</sub>, and ultrafine particulates (UFPs) exposure (based on the literature review of (Lary et al., 2015)).

2021. Even at the height of the COVID pandemic, deaths from poor air quality outnumbered those from COVID-19. The quality of outdoor ambient air, in turn, influences the quality of indoor air. In areas where outdoor air quality is already poor, increasing ventilation inevitably introduces pollutants indoors, even when filtration is employed (Srikrishna, 2022).

While death is unquestionably the most severe consequence of pollution, it is far from the only one. The biological wear and tear caused by poor air quality (allostatic load) have been shown to degrade cognitive and physical performance, as well as a variety of human health outcomes (Krebs and Luechinger, 2021; Gao et al., 2021; Carneiro et al., 2021; Ni et al., 2021; Shehab and Pope, 2019).

To illustrate, even among the indoor particulates with representatives among the criterion pollutants ( $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ ), the ultra-fine particulate matter, with a size of less than  $0.1 \mu\text{m}$ , are not included among the criterion pollutants, yet are actually some of the most dangerous to human health (Ehsanifar et al., 2021; Mojtaba et al., 2022; Ehsanifar et al., 2022; ?). The ultra-fine size fraction penetrates most deeply into the lungs, and can readily get into the bloodstream, and even into the brain across the blood-brain barrier. On this point, it is worth noting that there is a substantial increase in Alzheimer's which is associated with inflammation in the brain (Urrutia et al., 2021). Ultra-fine particulates are known to be associated with brain inflammation (Rhew et al., 2021). Further, at the other end of the size spectrum, the larger size range above  $10 \mu\text{m}$  is also currently excluded from EPA testing. This larger size range includes some airborne molds and pollens associated with some of the significant allergies prevalent across society (D'Amato et al., 2020). A similar issue is present with volatile organic compounds (VOCs), and the inappropriate lumping of very different types of organic compounds.

It is not just health outcomes, such as those listed in Table 1.1, that are associated with poor air quality. Numerous studies have demonstrated that poor air quality diminishes both human cognitive and physical performance (Krebs and Luechinger, 2021; Gao et al.,

2021; Carneiro et al., 2021; Ni et al., 2021; Shehab and Pope, 2019). Degraded cognitive performance has a significant impact on society, ranging from poor learning outcomes in schools to impaired decision-making performance and productivity at work. As a result, the ability to adequately and comprehensively quantify the performance of indoor air quality remediation systems is critical for guiding appropriate, science-based threat mitigation strategies. If HVAC control is driven by an incomplete subset of metrics/contaminants of concern, buildings will clearly fail to adapt to unexpected threats. For instance, CO<sub>2</sub> is frequently used as a control metric because it acts as a tracer gas to control the ventilation rate per person by controlling the indoor/outdoor differential. However, using CO<sub>2</sub> along with other metrics is beneficial. For example, five studies found that CO<sub>2</sub> control alone cannot control for pollutants generated outdoors or indoors independent of human activity (Zaatari et al., 2016). More recently, one COVID-19 case study discovered that the rapid spread of the disease through a nursing home ward was most likely caused by decreased ventilation by a CO<sub>2</sub> demand control ventilation system optimized solely for decreased energy-demand (Lamping and Muller, Lamping and Muller).

### **1.0.3 Physics Informed Machine Learning**

### **1.0.4 Dissertation Overview**

Much of the current machine learning landscape is dominated by highly abstract problems like image recognition, language prediction

In particular, this work focuses on three key areas

we should not throw the baby out with the bath water. Nor should we attempt to reinvent the wheel... we can use machine learning to *fill in the gaps* where our current theory can not easily provide a direct link between the data we have and the target we wish to model but a relationship is easily justified on physical grounds. Further, we can use our knowledge of physical laws to impose strong constraints on the space of available models that may

significantly improve the capability of machine learning models to extrapolate beyond the boundaries of their training data sets.

make the point that for physical theories are typically only amenable to analytic techniques when sufficiently constrained to domains with high symmetry or where nonlinear interactions can be controlled so that linear effects dominate over bounded of higher order effects (re: perturbation theory). For common real world scenarios where edges are sharp, functions aren't smooth, and approximating assumptions can not be readily applied, we typically must resort to numerical methods which seek to solve the dynamical laws on computational domains with well posed boundary and initial conditions.

In the first proposed project, we will apply machine learning techniques to provide a sensor calibration where theory suggests

To that end, this work focuses on improving sensing capabilities in three key domains: water quality, outdoor air quality, and indoor air quality.

physical sensing and physics-informed machine learning. The world we inhabit is undergoing rapid and transformative changes, with pressing environmental challenges demanding innovative solutions. At the heart of this quest lies the critical need for a deeper understanding of our complex environmental systems, coupled with the ability to derive actionable insights from the wealth of data at our disposal. This dissertation represents my contribution to addressing these challenges by bridging the gap between the physical world and advanced data-driven methodologies. Through the fusion of physical sensing technologies and machine learning, my research endeavors to unlock profound insights into environmental phenomena, ultimately facilitating informed decisions and sustainable practices in an ever-changing world.

This goal is pursued by applying physics informed approaches together with a suite of sensing and computational technologies, implementing the reusable paradigm of software defined sensors, i.e. physical sensing elements wrapped in a software layer. This software

layer can serve a variety of purposes such as calibration and the provision of enhanced or derived data products. It is part of a broader effort in the MINTS-AI laboratory at the University of Texas at Dallas. Where MINTS-AI is an acronym, Multi-Scale Multi-Use Integrated Intelligent Interactive Sensing in Service of Society for Actionable Insights.

Comprehensive environmental sensing is a timely and beneficial endeavor for a variety of reasons. The growing awareness of major environmental issues such as climate change, pollution, and habitat loss necessitates effective environmental monitoring and management. Comprehensive environmental sensing can provide real-time data on air and water quality, weather patterns, and other environmental factors, assisting in the identification and resolution of environmental issues. This assists in the development and implementation of policies and strategies aimed at reducing environmental impact and increasing sustainability. Given that, for instance, air quality can have significant effects on human health, this has particular societal value.

Comprehensive sensing of the environment can improve decision-making. The real-time and accurate data provided by environmental sensors can aid in informed decision-making regarding various aspects such as traffic management, industrial regulation, and crop planning. For instance, data on air quality can be used to inform decisions about reducing pollution levels, while data on weather patterns can help farmers to plan their crops and reduce water usage. Comprehensive sensing of the environment can be instrumental in emergency response. Real-time data on weather patterns, air quality, water levels and resources, and seismic activity can help emergency responders to prepare for and respond to natural disasters such as hurricanes, floods, and earthquakes. The quick and accurate information can enable effective and timely response, potentially saving lives and reducing the impact of the disaster.

Many advances in technology have enabled the creation of comprehensive sensing systems that can monitor and analyze data from various sensors and devices in real-time. In this

thesis we use a range of technologies including autonomous robotic teams [@Dunbabin2012; @Rubenstein2014; @Chen2017], hyperspectral imaging [@Plaza2009; @Li2018; @Zhu2017], mesh networks utilizing the Internet of Things (IoT) [@Gubbi2013; @Atzori2010; @Al-Fuqaha2015], machine learning (ML) [@Goodfellow2016; @LeCun2015; @Jordan2015], edge computing, high-performance computing, wearable sensors and modern high-performance dynamic programming languages such as Julia [@Bezanson2017] designed for numerical and scientific computing. These technologies have facilitated the collection and processing of large amounts of data from multiple sources, resulting in more accurate and comprehensive environmental monitoring.

Today, the most well known application of machine learning techniques

## CHAPTER 2

### PHYSICAL SENSING

#### 2.1 Coordinated Robotic Teams

we can discuss remote sensing, hyperspectral imaging, and the robot team here... Let's fetch some text from the 2019 paper.

- Remote Sensing: data acquisition, processing, and interpretation of images, and related data, obtained from aircraft and satellites that record the interaction between matter and electromagnetic radiation
- Source: the source of electromagnetic radiation, e.g. the sun, black-body radiation, microwave radar, etc...
- Atmospheric Radiation: The EM radiation propagating through the atmosphere. Moderated by various processes including absorption and scattering
- Earth's Surface Interactions: Amount and spectral distribution of radiation emitted/reflected by the earth's surface. This depends on
  - physical properties of the matter
  - wavelength of EM radiation that is sensed

#### 2.2 A Low Cost Sensor Network For Air Quality Monitoring

The outdoor part here is optional. Here I seek to provide specific details of the classes of low cost sensors used for PM (optical particle counters based on Mie scattering), thermistors (for temperature), humidity sensors, pressure sensors, gas sensors, etc... This can be a rough outline for those that specifically are used in my research projects.

## 2.3 PCO and The HEART Chamber

Photocatalytic Oxidation (PCO) technology aims to reproduce the outdoor process of photolysis indoors, yielding oxidative molecules that can effectively reduce viruses, bacteria, fungi, and other contaminants, including VOCs. While the technology's benefits have been well documented and replicated numerous times in real-world settings, we will focus attention, as directed by the RFI, on a highly rigorous approach to comprehensive measurements of indoor air quality and the resulting impacts that a technology, like Advanced Photocatalysis would have on any indoor atmosphere.

# HEART: Holistic Environmental Aerosol and Reactive Component Auto Response Testing

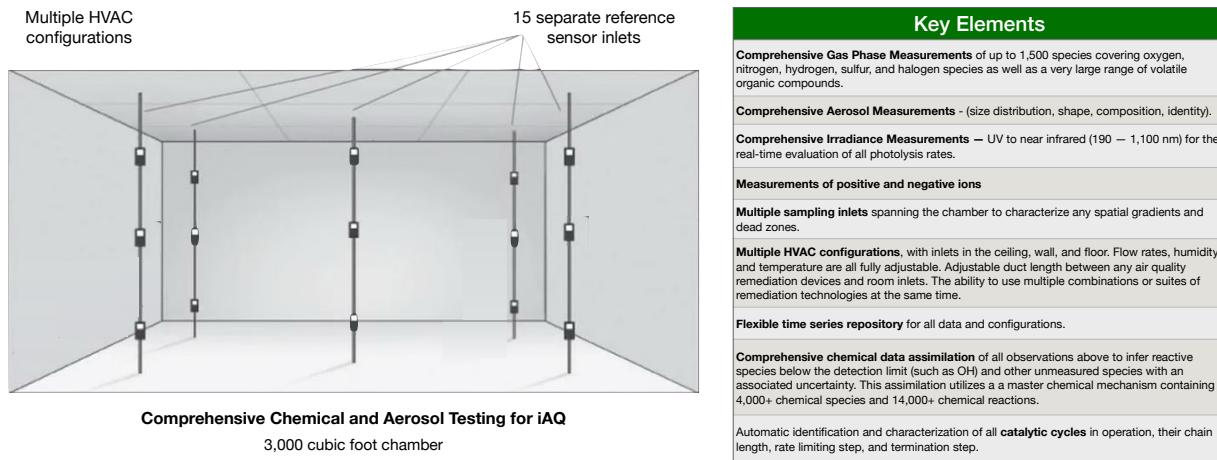


Figure 2.1: Details of comprehensive chemical and aerosol testing for IAQ using HEART: Holistic Environmental Aerosol and Reactive Component Auto Response Testing.

The HEART test chamber is designed for comprehensive characterization of air quality including the chemical composition (characterizing hundreds of chemicals), positive and negative ion density, aerosol size (from 5 nm to 100 microns), detailed morphology (shape), and composition, and individual bio-aerosol particle recognition, along with the accurately

characterized illumination environment characterized in more than 4,000 wavelengths from the UV to the near-infrared (190 — 1,100 nm) using a NIST calibrated spectrometer. The lighting environment is important, as, after all, air quality is atmospheric photochemistry in the indoor built environment, and photolysis is a key driver. Furthermore, some of the frequently advocated families of technologies make use of various parts of the ultraviolet spectrum. This high-energy region of the spectrum has important implications for indoor air quality.



Figure 2.2: Visual overview of the various infrastructure components of the ActivePure lab. Note that the lab uses multiple instruments of those shown to accommodate the stated range of analytes.

This testing capability can be used in two modalities. The first is in a static test chamber with the same volume (3,000 cubic feet) as the EPA test chamber. The chamber is equipped with a comprehensive suite of sensors just described, holistically characterizing hundreds of chemicals, the size of individual aerosols from 5 nm to 100 microns, their morphology, and their composition. This is critical because remediation technologies, such as UV lights

or various oxidation-based cleaners, may have a direct impact on the air's composition. Being able to comprehensively measure the gas phase oxygenated volatile organic compounds (OVOCs) and aerosol chemical composition is critical. Further, various cleaning compounds are often used in the indoor environment, introducing the possibility of significant levels of halogens. So as well as a suite of individual reference grade analyzers there is also a Syft 8 reagent ion mass spectrometer with a standard library of over 1,000 chemicals. The usual configuration used for the Syft is a large number of repeated sequential scans, each scan measuring a different family of species covering the nitrogen, hydrogen, sulfur, chlorine, bromine, and iodine families as well as the large range of volatile organic compounds. So for each measurement cycle, hundreds of species are measured. This is complemented by four aerosol instruments characterizing the aerosol/bio-aerosol size distribution between 5 nm and 100  $\mu\text{m}$ , the aerosol/bio-aerosol morphology, composition (using an aerosol mass spectrometer), and identity (with real-time machine learning providing a thousand member fingerprint for each individual particle, so that once trained it can recognize on the fly every single aerosol/bio-aerosol and identify what mold, pollen, pathogen, etc that it is).

Further, it is inevitable that some chemical species, for example, reactive components such as OH, will typically be below the detection limit of the measurement systems. Information on these unmeasured reactive components is naturally embedded in the shape of the time-series observations of the chemical species they are reacting with and in the partitioning between the large number of measured components that they are reacting with. This information can be effectively extracted, together with associated uncertainty, using chemical data assimilation (?Lary, 1999; Lary et al., 2003), a technique we pioneered at the University of Cambridge and NASA, and now used routinely by the EPA, NASA and NOAA in providing air quality analyses and forecasts.

Hundreds of chemicals measured in real-time in the HEART chamber, along with high-resolution irradiance spectra used to calculate real-time photolysis rates and real-time aerosol

measurements, are assimilated in a full chemical data assimilation system with a master chemical mechanism containing 4,000+ chemical species and 14,000+ chemical reactions. This enables us to accurately infer the abundance of unmeasured species, such as OH, with associated uncertainty. We use both 4D-Var chemical assimilation to initialize our system, followed by a full Kalman filter to time evolve the error covariance matrix. The chemical simulation and assimilation system won six NASA technology awards. This is the premier global approach for studying indoor air quality. Given the magnitude of the societal impacts of indoor air quality and preparing for any future pandemics, it is appropriate that the EPA uses a best-in-class methodology.

Furthermore, multiple HVAC configurations can readily be set up, with inlets located in the floor, ceiling, and walls. Further, the flow rates, humidity, and temperature can easily be adjusted, as can the length of the ductwork between any air quality remediation devices and the room inlets. In addition, various combinations or suites of remediation technologies can be used together.

In built-environment spaces, dead zones are a common issue. Fifteen inlets are strategically placed throughout the chamber to characterize the variability. The inlets are distributed across five vertical transects, with inlets at three vertical locations on each transect, 18 inches above the floor, on the centerline, and 18 inches below the ceiling.

The second modality for the test chamber is a mobile lab. This allows the performance of various remediation technologies to be tested in a wide range of ambient conditions. It also allows for the comprehensive exploration of the remediation technologies' effectiveness in a wide range of environments, from highly polluted such as in the Houston shipping channel, or downwind of a major forest fire, near a large composting facility, to unpolluted rural environments.

## CHAPTER 3

### MACHINE LEARNING METHODS

In this chapter, we discuss the relevant techniques from machine learning that we will employ throughout the rest of this dissertation.

NOTE: Good way to start is with the question: What is a model? We can discuss the differences between mechanistic models (i.e. physics equations), their free parameters (constants of nature) and other types of models such as the non-parametric, nonlinear models used in machine learning. Further, we can discuss how machine learning models often display have the desirable trait of being a **universal approximator**. This will naturally lead to a discussion of function expansions (Taylor, Fourier, other polynomial expansions, etc...) and how they scale poorly (exponential) with increasing dimension. Machine learning models essentially allow us to do the same thing: perform a function expansion given data but in a way that can scale well to arbitrary dimension (features) of data. This allows us to build predictive models without necessarily needing to prescribe \*all\* of the physics. From another perspective, this framework allows us to incorporate our physics knowledge from well-behaved or linearized domains in order to \*fit the residual\* behavior with a sophisticated data-driven approach.

discuss role in science in particular (i.e. calibration, modeling, etc...)

discuss pushback against use in science and need for methods which simultaneously provide uncertainty bounds. Also describe types of ML e.g. supervised, unsupervised, generative, etc...

this is a good place for the Chihuahua vs Muffin meme... and use this to motivate incorporating physical knowledge into the machine learning process as a key feature for scientific applications... we have more constraints!

also discuss statistical v.s. deep learning

### 3.1 Adjoint Methods for Optimization

The fundamental task of machine learning is to build data-driven models. To do this we must *fit* model parameters by taking advantage of the available training data. In many cases, this is straight forward: the sensitivities of model outputs to changes in model parameters can be backpropagated by application of the chain rule through each layer of our model. However, this naive propagation can become prohibitively expensive as we increase the size and complexity of our models. To efficiently determine these sensitivities and thereby enable the optimization of complicated models, we can instead turn the problem on it's side and consider a parallel *adjoint* system. Due to the utility of this technique for enabling the optimization of models involving implicit linear and nonlinear relationships and even differential equations, we shall begin our discussion of machine learning with a handful of these techniques which will be utilized in the remainder of this work.

#### 3.1.1 Adjoint Methods for Linear Systems

To begin, suppose a component of our model is given by the  $\theta$ -parameterized linear system

$$A(\theta)u = b(\theta). \quad (3.1)$$

If we wish to find the ideal  $\theta$  subject to some loss function  $J(u)$  we must obtain the gradients  $dJ/d\theta$  which upon expansion yields

$$\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} + \frac{\partial J}{\partial u} \frac{du}{d\theta}. \quad (3.2)$$

The Jacobian  $du/d\theta$  will be challenging to compute and depends on solving the above implicit system:

$$\begin{aligned} \frac{d}{d\theta}(Au) &= \frac{d}{d\theta}b \\ \left( \frac{dA}{d\theta} \right) u + A \left( \frac{du}{d\theta} \right) &= \frac{db}{d\theta} \end{aligned}$$

which with some slight abuse of notation in the, yields

$$\frac{du}{d\theta} = A^{-1} \left( \frac{db}{d\theta} - \frac{dA}{d\theta} u \right) \quad (3.3)$$

and therefore

$$\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} + \frac{\partial J}{\partial u} A^{-1} \left( \frac{db}{d\theta} - \frac{dA}{d\theta} u \right) \quad (3.4)$$

must require  $p$  linear solves if there are  $p$  parameters in  $\theta$ . Can we do better?

If we perform a clever *re-bracketing* so that

$$\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} + \left( \frac{\partial J}{\partial u} A^{-1} \right) \left( \frac{db}{d\theta} - \frac{dA}{d\theta} u \right) \quad (3.5)$$

then we can see that the vector  $\frac{\partial J}{\partial u} A^{-1}$  is the solution to *some* problem, say

$$\lambda^T A = \frac{\partial J}{\partial u} \quad (3.6)$$

then, we need only solve

$$A^T \lambda = \left( \frac{\partial J}{\partial u} \right)^T \quad (3.7)$$

once! This leaves us with the following procedure:

1. Solve the system  $Au = b$  for  $u$
2. Solve the adjoint system  $A^T \lambda = \left( \frac{\partial J}{\partial u} \right)^T$  for  $\lambda$
3. compute  $\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} + \lambda^T \left( \frac{db}{d\theta} - \frac{dA}{d\theta} u \right)$

### 3.1.2 Adjoint Methods for Nonlinear Systems

Having demonstrated an effective strategy for computing the gradients of cost functions with respect to underlying linear systems, a natural next step is to ask: can we do the same for nonlinear systems? Suppose now that we instead we find a component of our model to be of the form

$$f(u; \theta) = 0 \quad (3.8)$$

Then the gradient of some loss function  $J(u; \theta)$  which depends indirectly on the solution of this system will be

$$\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} + \frac{\partial J}{\partial u} \frac{du}{d\theta} \quad (3.9)$$

while the nonlinear system satisfies

$$\frac{df}{d\theta} = \frac{\partial f}{\partial u} \frac{du}{d\theta} + \frac{\partial f}{\partial \theta} = 0. \quad (3.10)$$

Therefore, the linear system that resulting from the differentiation procedure leads to

$$\frac{du}{d\theta} = - \left( \frac{\partial f}{\partial u} \right)^{-1} \frac{\partial f}{\partial \theta} \quad (3.11)$$

which we may substitute to find

$$\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} - \left( \frac{\partial J}{\partial u} \left( \frac{\partial f}{\partial u} \right)^{-1} \right) \frac{\partial f}{\partial \theta} \quad (3.12)$$

so that again, by cleverly re-bracketing the above equations we my identify that

$$\frac{\partial J}{\partial u} \left( \frac{\partial f}{\partial u} \right)^{-1}$$

is the solution to *some* linear system of the form

$$\lambda^T \left( \frac{\partial f}{\partial u} \right) = \frac{\partial J}{\partial u}. \quad (3.13)$$

Therefore, we have arrived at a procedure similar to the linear case which is as follows:

1. Solve  $f(u; \theta) = 0$  for  $u$  by your favorite algorithm.
2. Solve  $\left( \frac{\partial f}{\partial u} \right)^T \lambda = \left( \frac{\partial J}{\partial u} \right)^T$  for the adjoints  $\lambda$ .
3. Compute the gradient  $\frac{dJ}{d\theta} = \frac{\partial J}{\partial \theta} - \lambda^T \left( \frac{\partial f}{\partial \theta} \right)$ .

### 3.1.3 Adjoint Methods for ODEs

To complete the discussion, suppose that we instead seek to fit a model described by an Ordinary Differential Equations (ODE) initial-value problem (IVP) of the form

$$\begin{cases} \frac{du}{dt} = f(u, t; \theta) \\ u_0 = u(t=0) \end{cases} \quad (3.14)$$

where  $u \in \mathbb{R}^n$  denotes the time-dependent state vector, and  $f$  is a function of the state  $u$ , time  $t$ , and some number of parameters  $\theta$ .

To *fit* such a model to data, we need to be able to compute the sensitivity of some cost function  $J(u)$  to the parameters  $\theta$ , that is, we need the gradient  $dJ/d\theta$ . Obtaining these sensitivities by direct forward-mode automatic differentiation or with tape/tracer based backpropagation demands we be able to establish the derivatives for any class of ODE integrator we want to choose. This task is further complicated if we desire to use modern differential equation solver suites, like those provided by the `DifferentialEquations.jl` Julia library, which employ complicated adaptive stepping schemes. How could we possibly know *a priori* how many function evaluations will be needed to produce the output for each ODE step, and thereby how to compute the partial derivatives? One approach is to establish language-wide automatic differentiation capability for generic computer code (dubbed  $\partial P$ ) (Innes et al., 2019). An alternative approach, which we shall describe here, is to develop a set of adjoint differential equations whose solution we can use to compute the desired derivatives.

Let us suppose our loss function can be written in the form

$$J(u; \theta) = \int_0^T g(u; \theta) dt \quad (3.15)$$

where  $g$  is some function like the quadratic loss  $u^T R u$  of 4d-var. To find our desired sensitivities, we can utilize the method of Lagrange multipliers by introducing a Lagrangian of

the form

$$\begin{aligned}\mathcal{L}(u, \lambda; \theta) &:= J(u; \theta) + \int_0^T \lambda^T \left( f - \frac{du}{dt} \right) dt \\ &= \int_0^T \left[ g(u; \theta) + \lambda^T(t) \left( f - \frac{du}{dt} \right) \right] dt\end{aligned}\tag{3.16}$$

The  $\lambda$  are the so-called Lagrange multipliers, and the integrand,  $f - du/dt$ , is a clever way of adding 0 so that  $d\mathcal{L}/d\theta = dJ/d\theta$ . Evaluating this derivative, we find

$$\frac{d\mathcal{L}}{d\theta} = \int_0^T \left\{ \left( \frac{\partial g}{\partial \theta} + \frac{\partial g}{\partial u} \frac{du}{d\theta} \right) + \lambda^T(t) \left[ \frac{\partial f}{\partial \theta} + \frac{\partial f}{\partial u} \frac{du}{d\theta} - \frac{d}{dt} \frac{du}{d\theta} \right] \right\} dt.\tag{3.17}$$

The difficult term to compute is  $du/d\theta$  and so we group the terms so that

$$\frac{d\mathcal{L}}{d\theta} = \int_0^T \left[ \frac{\partial g}{\partial \theta} + \lambda^T(t) \frac{\partial f}{\partial \theta} + \left( \frac{\partial g}{\partial u} + \lambda^T(t) \frac{\partial f}{\partial u} - \lambda^T(t) \frac{d}{dt} \right) \frac{du}{d\theta} \right] dt.\tag{3.18}$$

We now observe that if we pick the  $\lambda^T(t)$  such that

$$\lambda^T(t) \frac{\partial f}{\partial \theta} + \left( \frac{\partial g}{\partial u} + \lambda^T(t) \frac{\partial f}{\partial u} - \lambda^T(t) \frac{d}{dt} \right) \frac{du}{d\theta} = 0,\tag{3.19}$$

we will not have to compute this pesky term at all! To proceed, let's apply integration by parts to move the time derivative of the final term and simplify the matter:

$$\begin{aligned}\int_0^T -\lambda^T \frac{d}{dt} \frac{du}{d\theta} dt &= -\lambda^T \frac{du}{d\theta} \Big|_0^T + \int_0^T \frac{d\lambda^T}{dt} \frac{du}{d\theta} dt \\ &= \lambda^T(0) \frac{du}{d\theta}(0) - \lambda^T(T) \frac{du}{d\theta}(T) + \int_0^T \dot{\lambda}^T \frac{du}{d\theta} dt.\end{aligned}\tag{3.20}$$

Plugging this back in to our previous expression yields

$$\frac{d\mathcal{L}}{d\theta} = \int_0^T \left[ \frac{\partial g}{\partial \theta} + \lambda^T \frac{\partial f}{\partial \theta} + \left( \frac{\partial g}{\partial u} + \lambda^T \frac{\partial f}{\partial u} + \dot{\lambda}^T \right) \frac{du}{d\theta} \right] dt + \lambda^T(0) \frac{du}{d\theta}(0) - \lambda^T(T) \frac{du}{d\theta}(T).\tag{3.21}$$

Great! In order to make the pesky term disappear, it suffices to find the  $\lambda^T(t)$  such that

$$\begin{cases} \frac{\partial g}{\partial u} + \lambda^T \frac{\partial f}{\partial u} + \frac{d\lambda^T}{dt} = 0 \\ \lambda(T) = 0 \end{cases}\tag{3.22}$$

which amounts simply solving a second differential equation from the ending time  $T$  back to  $t = 0$ . Transposing the above, we can obtain the new ODE in the standard form:

$$\begin{cases} \frac{d\lambda}{dt} = - \left( \frac{\partial f}{\partial u} \right)^T \lambda - \left( \frac{\partial g}{\partial u} \right)^T \\ \lambda(T) = 0 \end{cases} \quad (3.23)$$

Better yet, this system is linear in  $\lambda$  even though the original differential equation may be fully nonlinear with no restrictions on  $f$ ! To summarize, the procedure for determining the desired gradients is as follows:

1. Solve the ODE system given by  $du/dt = f(u, t; \theta)$  forward in time from  $t = 0$  to  $t = T$  to obtain the solution  $u(t)$ .
2. Solve the Adjoint ODE system given by  $d\lambda/dt = -(\partial f/\partial u)^T \lambda - (\partial g/\partial u)^T$  from  $t = T$  to  $t = 0$  to obtain the adjoints  $\lambda^T(t)$
3. Compute the gradient  $dJ/d\theta = d\mathcal{L}/d\theta = \int_0^T [(\partial g/\partial \theta) + \lambda^T(t)(\partial f/\partial \theta)] dt + \lambda(0) du - 0/d\theta$  by quadrature.

## 3.2 Supervised Regression Techniques

### 3.2.1 Neural Networks

### 3.2.2 Gaussian Process Regression

Based on my notes from this repo

The following is based on the book **Gaussian Processes for Machine Learning** by Carl Edward Rasmussen and Christopher K. I. Williams (Rasmussen et al., 2006). You can find the free online book [here](#).

To explain/derive the Gaussian Process model for regression, let's first consider a motivating example: **Linear Regression**. We will use this guiding example to derive GRP

from a *weight space view*. After this derivation, we will suggest a simpler, but more abstract derivation using a *function space view*. First let's set up some data we can use for training:

### The Weight Space View

We consider a dataset  $\mathcal{D}$  with  $n$  observations

$$\mathcal{D} = \left\{ (\mathbf{x}_i, y_i) \mid i = 1, \dots, n \right\} \quad (3.24)$$

- $\mathbf{x}_i$  is the  $i^{th}$  D-dimensional input (feature) vector
- $y_i$  is the  $i^{th}$  target

Linear regression is easily understood in terms of *linear algebra*. We therefore collect our dataset  $\mathcal{D}$  into a  $D \times n$  dimensional **Design Matrix**. Note that we have used a transposed definition (features are rows, records are columns) as Julia is a column-major language (like Matlab and Fortran).

$$X := \begin{pmatrix} \vdots & \vdots & \vdots \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \\ \vdots & \vdots & & \vdots \end{pmatrix} \quad (3.25)$$

and our targets into a target vector

$$\mathbf{y} := (y_1, \dots, y_n) \quad (3.26)$$

so that the full training set becomes

$$\mathcal{D} := (X, \mathbf{y}) \quad (3.27)$$

Standard linear regression is a model of the form

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{w} \quad (3.28)$$

where  $\mathbf{w}$  is the  $D$ -dimensional vector of weights. By minimizing the mean-squared-error between our model and targets, one can show that the optimal weights are given by

$$\mathbf{w} = (X X^T)^{-1} X \mathbf{y} \quad (3.29)$$

Note, this can also be easily obtained geometrically by finding the vector with the shortest distance to the hyperplane defined by the column space of  $X$ . This corresponds to solving the normal equations.

$$X X^T \mathbf{w} = X \mathbf{y} \quad (3.30)$$

The following demonstrates this procedure on a simple dataset

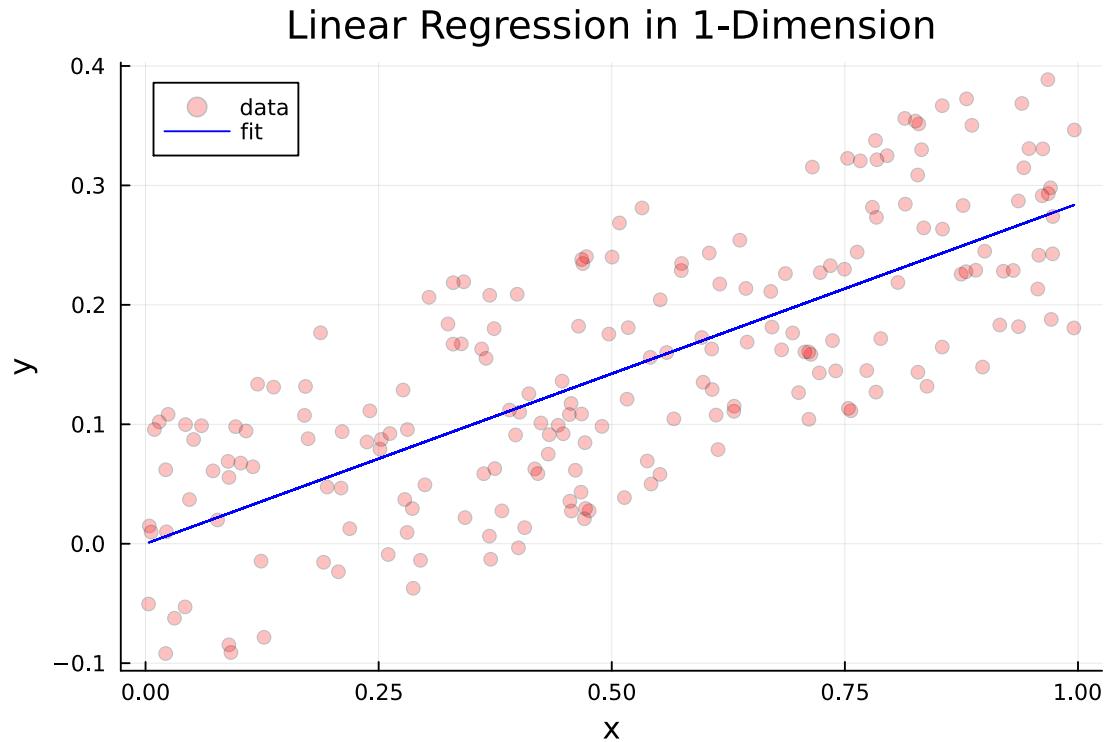


Figure 3.1: A standard linear regression fit on some noisy data.

We can also fit a y-intercept (aka *bias*) by augmenting the design matrix  $X$  to contain an extra row with all 1's, i.e.

$$X[D+1, :] = (1, \dots, 1) \quad (3.31)$$

Standard linear regression assumes that the data  $\mathcal{D}$  are perfect but we can clearly see that the above data are noisy. To account for this, we need to make our model *Bayesian* by augmenting it to consider the measurement error. We define

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{w} \quad (3.32)$$

$$\mathbf{y} = f(\mathbf{x}) + \epsilon \quad (3.33)$$

$$\epsilon \sim \mathcal{N}(0, \sigma_n^2) \quad (3.34)$$

or, in words, our observed values differ from the *truth* by identically, independently, distributed Gaussian noise with mean 0 and variance  $\sigma_n^2$ . The assumption that the noise is i.i.d. is critical because it allows us to simplify the *likelihood* function by separating out each individual contribution by our datapoints:

$$p(\mathbf{y}|X, \mathbf{w}) := \prod_i^n p(\mathbf{y}_i|\mathbf{x}_i, \mathbf{w}) \quad (3.35)$$

$$= \prod_i^n \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(\mathbf{y}_i - \mathbf{x}_i^T \mathbf{w})^2}{2\sigma_n^2}\right) \quad (3.36)$$

$$= \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_n^2} |\mathbf{y} - X^T \mathbf{w}|^2\right) \quad (3.37)$$

$$= \mathcal{N}(X^T \mathbf{w}, \sigma_n^2 I) \quad (3.38)$$

To perform inference with this updated model, we apply Baye's Rule, that is:

$$p(\mathbf{w}|\mathbf{y}, X) = \frac{p(\mathbf{y}|X, \mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|X)} \quad (3.39)$$

where

- $p(\mathbf{w}|\mathbf{y}, X)$  is the *posterior distribution*

- $p(\mathbf{y}|X, \mathbf{w})$  is the *likelihood*
- $p(\mathbf{w})$  is the *prior distribution*
- $p(\mathbf{y}|X)$  is the *marginal likelihood*, i.e. the normalization constant

It is now that the utility of choosing gaussian distributions for our likelihood and prior becomes clear. We have

$$p(\mathbf{w}|\mathbf{y}, X) \propto \exp\left(-\frac{1}{2\sigma_n^2}(\mathbf{y} - X^T\mathbf{w})^T(\mathbf{y} - X^T\mathbf{w})\right) \exp\left(-\frac{1}{2}\mathbf{w}^T\Sigma_p^{-1}\mathbf{w}\right) \quad (3.40)$$

Taking the log and expanding leads to

$$\begin{aligned} \log(p(\mathbf{w}|\mathbf{y}, X)) &= \frac{1}{2} \left[ \frac{1}{\sigma_n^2} \mathbf{y}^T \mathbf{y} - \frac{1}{\sigma_n^2} \mathbf{y}^T X^T \mathbf{w} - \frac{1}{\sigma_n^2} \mathbf{w}^T X \mathbf{y} + \frac{1}{\sigma_n^2} \mathbf{w}^T X X^T \mathbf{w} + \mathbf{w}^T \Sigma_p^{-1} \mathbf{w} \right] \\ &\quad (3.41) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2} \left[ \mathbf{w}^T \left( \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1} \right) \mathbf{w} - \left( \frac{1}{\sigma_n^2} \mathbf{y}^T X^T \right) \mathbf{w} - \mathbf{w}^T \left( \frac{1}{\sigma_n^2} X \mathbf{y} \right) + \mathbf{y}^T \frac{1}{\sigma_n^2} \mathbf{y} \right] \\ &\quad (3.42) \end{aligned}$$

$$= \mathbf{w}^T A \mathbf{w} - B^T \mathbf{w} - \mathbf{w}^T B + C \quad (3.43)$$

where we have defined

$$A := \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1} \quad (3.44)$$

$$B := \frac{1}{\sigma_n^2} X \mathbf{y} \quad (3.45)$$

$$C := \mathbf{y}^T \frac{1}{\sigma_n^2} \mathbf{y} \quad (3.46)$$

Now we can complete the square so that

$$\mathbf{w}^T A \mathbf{w} - B^T \mathbf{w} - \mathbf{w}^T B + C = (\mathbf{w} - \bar{\mathbf{w}})^T A (\mathbf{w} - \bar{\mathbf{w}}) + K \quad (3.47)$$

leading to

$$\bar{\mathbf{w}} = A^{-1} B = \frac{1}{\sigma_n^2} \left( \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1} \right)^{-1} X \mathbf{y} \quad (3.48)$$

$$K = C - \bar{\mathbf{w}}^T A \bar{\mathbf{w}} \quad (3.49)$$

Since  $K$  does not depend on  $\mathbf{w}$  directly, it may be absorbed into the normalization of  $p(\mathbf{w}|\mathbf{y}, X)$ . Thus we are left with

$$p(\mathbf{w}|\mathbf{y}, X) = \mathcal{N}\left(\bar{\mathbf{w}} = \frac{1}{\sigma_n^2} A^{-1} X \mathbf{y}, \Sigma = A^{-1}\right) \quad (3.50)$$

$$A = \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1} \quad (3.51)$$

This result gives us the gaussian distribution over the space of possible parameter vectors  $\mathbf{w}$ . To use this distribution to make predictions, consider a newly supplied testpoint  $\mathbf{x}_*$ . We want to find

$$p(y_*|\mathbf{x}_*, \mathbf{y}, X) \quad (3.52)$$

We do this by marginalizing over our weight distribution, i.e.

$$p(y_*|\mathbf{x}_*, \mathbf{y}, X) = \int_{\mathbf{w}} p(y_*|\mathbf{x}_*, \mathbf{w}) p(\mathbf{w}|\mathbf{y}, X) d\mathbf{w} \quad (3.53)$$

If we make the further assumption that testing points are i.i.d. gaussian distributed, we see that this integral is the product of two gaussians and therefore is also a gaussian. To find the mean and covariance of the predictive distribution, we check

$$\bar{y}_* = \mathbb{E}[y_*] = \mathbb{E}[\mathbf{x}_*^T \mathbf{w}] = \mathbf{x}_*^T \mathbb{E}[\mathbf{w}] = \mathbf{x}_*^T \bar{\mathbf{w}} \quad (3.54)$$

$$\text{Cov}(y_*) = \mathbb{E}[(y_* - \bar{y}_*)(y_* - \bar{y}_*)^T] \quad (3.55)$$

$$= \mathbb{E}[(\mathbf{x}_*^T \mathbf{w} - \mathbf{x}_*^T \bar{\mathbf{w}})(\mathbf{x}_*^T \mathbf{w} - \mathbf{x}_*^T \bar{\mathbf{w}})^T] \quad (3.56)$$

$$= \mathbb{E}[\mathbf{x}_*^T (\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T \mathbf{x}_*] \quad (3.57)$$

$$= \mathbf{x}_*^T \mathbb{E}[(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T] \mathbf{x}_* \quad (3.58)$$

$$= \mathbf{x}_*^T \text{Cov}(\mathbf{w}) \mathbf{x}_* \quad (3.59)$$

$$= \mathbf{x}_*^T A^{-1} \mathbf{x}_* \quad (3.60)$$

so that

$$p(y_*|\mathbf{x}_*, \mathbf{y}, X) = \mathcal{N}(\mathbf{x}_*^T \bar{\mathbf{w}}, \mathbf{x}_*^T A^{-1} \mathbf{x}_*) \quad (3.61)$$

Let's now take a break from our Bayesian regression discussion and return to the standard linear regression model for a moment. The key drawback of linear models like this is, of course, that they're *linear*!. Considering that many (most) *interesting* relationships are non-linear, how can we extend our simple linear model to enable us to perform complicated non-linear fits?

In the parlance of machine learning, the simple solution is to do feature engineering. If our initial feature vector is

$$\mathbf{x} = (x_1, \dots, x_n) \quad (3.62)$$

we can use our *expertise* to concoct new combinations of these features to produce the augmented vector

$$\tilde{\mathbf{x}} = (x_1, \dots, x_n, x_1^2, \sin(x_2), x_5 x_7 / x_4, \dots) \quad (3.63)$$

As an example, a linear classifier is unable to distinguish points inside a circle from those outside just from the  $(x, y)$  coordinates alone. Augmenting the feature vector to include the squared radius  $x^2 + y^2$  as a new feature removes this obstacle. This works because the *linear* part of linear regression only refers to the fact that our model takes *linear combinations* of feature variables to produce its output. There is no restriction that the features themselves need to be independent variables! This same idea is what makes methods like SINDy work which use libraries of polynomial combinations of base features.

Constructing new features is often more art than science. To standardize the process, let's abstract the mapping from the original feature vector  $\mathbf{x}$  to the augmented vector  $\tilde{\mathbf{x}}$ . This is accomplished via the projection map  $\phi : \mathbb{R}^D \rightarrow \mathbb{R}^N$  where

$$\mathbf{x} \mapsto \tilde{\mathbf{x}} = \phi(\mathbf{x}) \quad (3.64)$$

The result is that our linear model updates to become

$$f(\mathbf{x}) := \phi(\mathbf{x})^T \mathbf{w} \quad (3.65)$$

where the weight vector has gone from  $D$  dimensional to  $N$  dimensional. Similarly, the normal equations for  $\mathbf{w}$  update to become

$$\mathbf{w} = (\Phi\Phi^T)^{-1}\Phi\mathbf{y} \quad (3.66)$$

where  $\Phi = \phi(X)$  is the  $N \times n$  matrix resulting from applying  $\phi$  columnwise to  $X$ .

The following example shows how to use such a mapping to produce a quadratic polynomial fit.

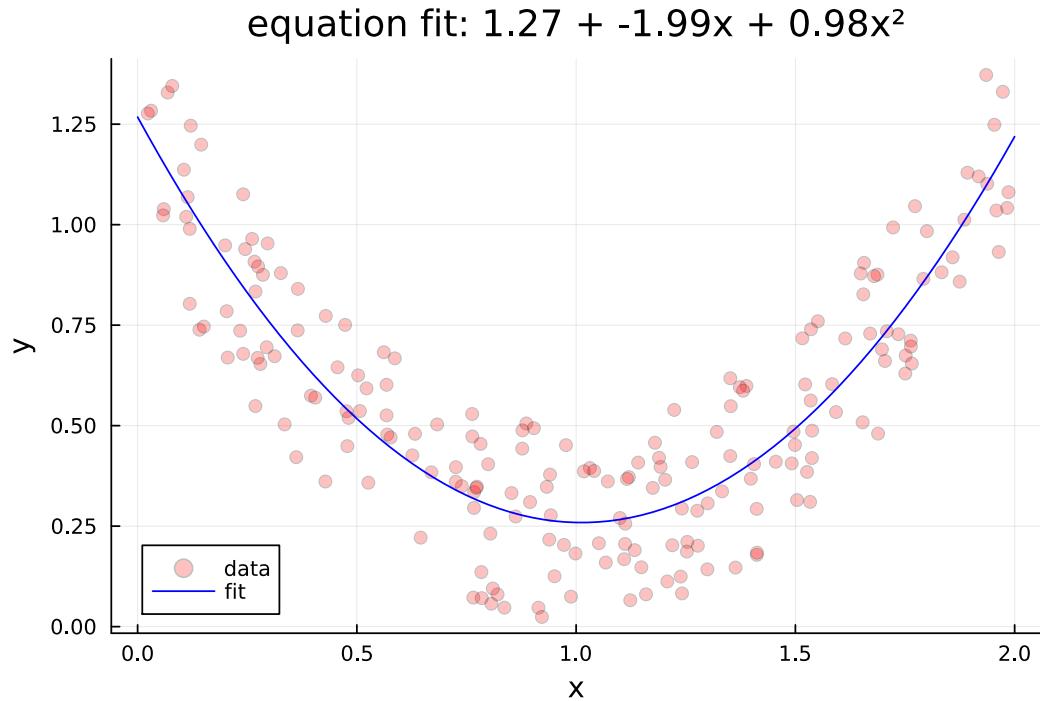


Figure 3.2: An example of polynomial regression for which we fit a quadratic polynomial to some noisy data.

We see from Figure 3.2 that our linear regression model found a great fit for a 2nd order polynomial when supplied with polynomial features. Let's update our Bayesian regression scheme to reflect the use of our feature projection map  $\phi$ . First we define

$$\Phi := \phi(X) \quad (3.67)$$

$$\phi_* := \phi(\mathbf{x}_*) \quad (3.68)$$

Our predictive distribution therefore becomes

$$p(y_* | \mathbf{x}_*, X, \mathbf{y}) = \mathcal{N} \left( \frac{1}{\sigma_n^2} \phi_*^T A^{-1} \Phi \mathbf{y}, \phi_*^T A^{-1} \phi_* \right) \quad (3.69)$$

$$A = \frac{1}{\sigma_n^2} \Phi \Phi^T + \Sigma_p^{-1} \quad (3.70)$$

Great! Now we can do our Bayesian inference with non-linear features given by  $\phi$ . There is a *massive* problem with this approach, however. Our order 2 polynomial map  $\phi$  takes us from a  $D$  dimensional feature vector to  $(D + 1)!$  many. This means that as we add more features to our feature map, the dimension of the resulting vector will quickly become prohibitively large. Looking at our current model equations, we see that the bottleneck is the matrix inversion of  $A$  which requires we invert an  $N \times N$  matrix. Our prediction (i.e. the mean) involves multiplication on the right by the  $n$  dimensional vector  $\mathbf{y}$ . With that in mind, perhaps we can reformulate the above into an equivalent form using at most an  $n \times n$  dimensional matrix.

Let  $K := \Phi^T \Sigma_p \Phi$ . Observe the following:

$$\frac{1}{\sigma_n^2} \Phi(K + \sigma_n^2 I) = \frac{1}{\sigma_n^2} \Phi (\Phi^T \Sigma_p \Phi + \sigma_n^2 I) \quad (3.71)$$

$$= \frac{1}{\sigma_n^2} \Phi \Phi^T \Sigma_p \Phi + \Phi I \quad (3.72)$$

$$= \left( \frac{1}{\sigma_n^2} \Phi \Phi^T \right) \Sigma_p \Phi + (\Phi I \Phi^{-1} \Sigma_p^{-1}) \Sigma_p \Phi \quad (3.73)$$

$$= \left( \frac{1}{\sigma_n^2} \Phi \Phi^T + \Sigma_p^{-1} \right) \Sigma_p \Phi \quad (3.74)$$

$$= A \Sigma_p \Phi \quad (3.75)$$

From there we see that

$$A^{-1} \frac{1}{\sigma_n^2} \Phi (K + \sigma_n^2 I) = \Sigma_p \Phi \quad (3.76)$$

$$\Rightarrow \frac{1}{\sigma_n^2} A^{-1} \Phi = \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \quad (3.77)$$

$$\Rightarrow \frac{1}{\sigma_n^2} \phi_*^T A^{-1} \Phi = \phi_*^T \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \quad (3.78)$$

For the covariance, we utilize the matrix inversion lemma ([ADD REFERENCE HERE](#)) which states

$$(Z + UWV^T)^{-1} = Z^{-1} - Z^{-1}U(W^{-1} + V^T Z^{-1}U)^{-1}V^T Z^{-1} \quad (3.79)$$

With the identification

$$Z^{-1} \rightarrow \Sigma_p \quad (3.80)$$

$$W^{-1} \rightarrow \sigma_n^2 I \quad (3.81)$$

$$V \rightarrow \Phi \quad (3.82)$$

$$U \rightarrow \Phi \quad (3.83)$$

we find

$$\Sigma_p - \Sigma_p \Phi (\Sigma_p + \Phi^T \Sigma_p \Phi)^{-1} \Phi^T \Sigma_p = \left( \Sigma_p^{-1} + \Phi \frac{1}{\sigma_n^2} I \Phi^T \right)^{-1} \quad (3.84)$$

$$= \left( \frac{1}{\sigma_n^2} \Phi \Phi^T + \Sigma_p^{-1} \right)^{-1} \quad (3.85)$$

$$= A^{-1} \quad (3.86)$$

Thus, we have the equivalent form for our predictive distribution:

$$p(y_* | \mathbf{x}_*, X, \mathbf{y}) = \mathcal{N}(\phi_*^T \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \mathbf{y}, \phi_*^T \Sigma_p \phi_* - \phi_*^T \Sigma_p \Phi (K + \sigma_n^2 I)^{-1} \Phi^T \Sigma_p \phi_*) \quad (3.87)$$

where the pesky  $N \times N$  term has been replaced by the  $n \times n$  matrix  $\Phi^T \Sigma_p \Phi$ .

We now make the the *key* observation that the only matrices that appear in the above expression are

$$\Phi^T \Sigma_p \Phi, \quad \phi_*^T \Sigma_p \phi_* \quad (3.88)$$

$$\phi_*^T \Sigma_p \Phi, \quad \Phi^T \Sigma_p \phi_* \quad (3.89)$$

whose matrix elements we can write abstractly as

$$\phi(\mathbf{x})^T \Sigma_p \phi(\mathbf{x}') \quad (3.90)$$

To fit our model, we must determine appropriate values for the symmetric, positive semi-definite covariance matrix  $\Sigma_p$  (and  $\sigma_n$  too, technically). Instead, we observe that this matrix product is a quadratic form which we can think of as representing an inner product on our transformed vectors:

$$K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \quad (3.91)$$

We call the function  $k(\mathbf{x}, \mathbf{x}')$  the *kernel function* or the *covariance function*.

All we need to perform the above calculations are the matrix elements of  $K$  applied to our data  $\mathcal{D}$  and any test points  $\mathbf{x}_*$  we wish to apply our model to. In effect, this means we are free to use feature vectors of any dimension, including  $\infty$ . The idea here is that the kernel function represents an inner product over *some* vector space. As it turns out, the RBF kernel corresponds to a an infinite dimensional feature vector.

There are many choices for the kernel function. One of the most popular is the RBF (radial basis function) kernel, also commonly referred to as the *squared exponential kernel*:

$$k_{\text{rbf}}(\mathbf{x}, \mathbf{x}') := \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}|\mathbf{x} - \mathbf{x}'|^2\right) \quad (3.92)$$

where  $\sigma_f^2$  is the \*signal variance\* and  $\ell$  denotes the similarity length scale.

For notational convenience, let's define

$$K := k(X, X) \quad (3.93)$$

$$K_{**} := k(X_*, X_*) \quad (3.94)$$

$$K_* := k(X, X_*) \quad (3.95)$$

then, our predictive distribution takes the final, *clean* form

$$p(\mathbf{y}_* | X_*, X, \mathbf{y}) = \mathcal{N}\left(K_*^T(K + \sigma_n^2 I)^{-1}\mathbf{y}, K_{**} - K_*^T(K + \sigma_n^2 I)^{-1}K_*\right) \quad (3.96)$$

This is the *end-result* of Gaussian Process Regression acheived via the *weight-space view*.

## The Function-space View

So far our approach has been to generalize the standard linear regression model to allow for fitting over a (possibly infinite) basis of features with consideration for measurement and model uncertainty (our Bayesian priors). In essence, the idea was to fit the distribution of all possible weights conditioned on the available training data,  $p(\mathbf{w}|X, \mathbf{y})$ . A second *equivalent* approach is to instead consider the distribution of all possible model function  $f(\mathbf{x})$ . By constructing a Bayesian prior over this space, we constrain the space of possible model functions and learn a *distribution* over all allowed model functions,  $p(f|X, \mathbf{y})$ . To do so we will need to develop the abstract machinery of distributions over function spaces. When these distributions are Gaussian, the result is called a **Gaussian process**.

By this point, we are very familiar with the Gaussian distribution, a.k.a. the Normal distribution  $\mathcal{N}(\mu, \sigma^2)$ . This distribution is defined by a mean value  $\mu$  and a variance  $\sigma^2$ . Its *big brother* is the **Multivariate Normal Distribution**,  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , described by a vector of means  $\boldsymbol{\mu}$  and a covariance matrix  $\boldsymbol{\Sigma}$ . A natural question, then, is can we generalize the concept of the Gaussian distribution from  $N$  dimensions to being defined over a continuous field? This leads us naturally to the development of a so-called **Gaussian Process**.

**Definition:** A *Gaussian Process*,  $\mathcal{GP}$ , is a collection of random variables for which any finite subset are described by a joint Gaussian distribution.

To see where this comes from, recall that in our previous derivation, we already made the assumption that all our data points  $\mathcal{D}$  are i.i.d. Gaussian distributed. A Gaussian process is the natural extension of this and makes the assumption that the continuous set from which data are sampled are \*\*so Gaussian\*\* that any finite sample will be jointly Gaussian distributed. The term \*process\* is used to distinguish between finite collections of random variables (distributions) and their continuous counterparts described here.

Because each finite subset of this continuous collection is jointly gaussian, we can completely specify a Gaussian Process with two functions: the mean function  $m(\mathbf{x})$  and the covariance function  $k(\mathbf{x}, \mathbf{x}')$ . To denote this, we typically write

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (3.97)$$

To see this in action, recall our Bayesian regression model

$$f(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{w} \quad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p) \quad (3.98)$$

where we have set the prior on  $\mathbf{w}$  to have zero mean. The mean function is given by the expectation value of our model:

$$\mathbb{E}[f(\mathbf{x})] = \phi(\mathbf{x})^T \mathbb{E}[\mathbf{w}] = 0 \quad (3.99)$$

and the covariance function is given by

$$\mathbb{E}[f(\mathbf{x})f(\mathbf{x}')]=\phi(\mathbf{x})^T\mathbb{E}[\mathbf{w}\mathbf{w}^T]\phi(\mathbf{x}')=\phi(\mathbf{x})^T\Sigma_p\phi(\mathbf{x}') \quad (3.100)$$

To repeat the point, the key feature of Gaussian processes is that finite subsets are jointly Gaussian distributed. Thus we can split our data into the testpoints  $\mathcal{D} = (X, \mathbf{y})$  and testpoints  $X_*$  and treat each collection as joint distributions with the following priors:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \quad (3.101)$$

where  $\mathbf{f} := f(X)$  and  $\mathbf{f}_* = f(X_*)$ .

To obtain our predictive distribution,  $p(\mathbf{f}_*|X_*, X, \mathbf{y})$ , we *condition the joint prior distribution* on the observations. To see how this works, consider a general joint gaussian distribution given by

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right) \quad (3.102)$$

define the centered values  $\tilde{x} := x - \mu_x$  and  $\tilde{y} := y - \mu_y$ . Define the intermediate variable

$$z := \tilde{x} - A\tilde{y} \quad (3.103)$$

Note that since we've subtracted out the mean we have  $\mathbb{E}[\tilde{x}] = \mathbb{E}[\tilde{y}] = \mathbb{E}[z] = 0$ . Let's now find  $A$ .

$$\mathbb{E}[z\tilde{y}^T] = \mathbb{E}[(\tilde{x} - A\tilde{y})\tilde{y}^T] \quad (3.104)$$

$$= \mathbb{E}[\tilde{x}\tilde{y}^T - A\tilde{y}\tilde{y}^T] \quad (3.105)$$

$$= \mathbb{E}[\tilde{x}\tilde{y}^T] - \mathbb{E}[A\tilde{y}\tilde{y}^T] \quad (3.106)$$

$$= \Sigma_{xy} - A\mathbb{E}[\tilde{y}\tilde{y}^T] \quad (3.107)$$

$$= \Sigma_{xy} - A\Sigma_{yy} \quad (3.108)$$

Therefore if we choose  $A$  so that  $z$  and  $\tilde{y}$  are independent and uncorrelated, then  $\Sigma_{zy} = \mathbb{E}[z\tilde{y}^T] = 0$ . Using this assumption, we find

$$0 = \mathbb{E}[z\tilde{y}^T] = \Sigma_{xy} - A\Sigma_{yy} \Rightarrow \boxed{A = \Sigma_{xy}\Sigma_{yy}^{-1}} \quad (3.109)$$

If we now condition  $\tilde{x}$  on  $\tilde{y}$  (i.e. look at  $\tilde{x}$  when  $\tilde{y}$  is constant), we find

$$\mathbb{E}[\tilde{x}|\tilde{y}] = A\tilde{y} + \mathbb{E}[z] \quad (3.110)$$

$$= A\tilde{y} + 0 \quad (3.111)$$

$$= \Sigma_{xy}\Sigma_{yy}^{-1} \quad (3.112)$$

$$(3.113)$$

By manipulating this expression, we can now derive  $\mathbb{E}[x|y]$  as follows:

$$\mathbb{E}[x|\tilde{y}] = \mathbb{E}[\tilde{x}|\tilde{y}] + \mu_x \quad (3.114)$$

$$= \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}\tilde{y} \quad (3.115)$$

$$(3.116)$$

$$\boxed{\mathbb{E}[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)} \quad (3.117)$$

Similarly for the covariance, we have

$$\text{Cov}(x|y) = \text{Cov}(\tilde{x} + \mu_x|\tilde{y}) \quad (3.118)$$

$$= \text{Cov}(\tilde{x} + \mu_x|\tilde{y} + \mu_y) \quad (3.119)$$

$$= \text{Cov}(\tilde{x}|(\tilde{y} + \mu_y)) \quad (3.120)$$

$$= \text{Cov}(\tilde{x}|\tilde{y}) \quad (3.121)$$

$$= \text{Cov}((z + A\tilde{y})|\tilde{y}) \quad (3.122)$$

$$= \text{Cov}(z) + AC\text{ov}(\tilde{y}) \quad (3.123)$$

$$= \text{Cov}(z) + 0 \quad (3.124)$$

$$= \mathbb{E}[zz^T] \quad (3.125)$$

$$= \mathbb{E}[(\tilde{x} - A\tilde{y})(\tilde{x} - A\tilde{y})^T] \quad (3.126)$$

$$= \mathbb{E}[\tilde{x}\tilde{x}^T - A\tilde{y}\tilde{x}^T - x(A\tilde{y})^T + A\tilde{y}\tilde{y}^TA^T] \quad (3.127)$$

$$= \Sigma_{xx} - A\Sigma_{yx} - \Sigma_{xy}A^T + A\Sigma_{yy}A^T \quad (3.128)$$

$$= \Sigma_{xx} - (\Sigma_{xy}\Sigma_{yy}^{-1})\Sigma_{yx} - \Sigma_{xy}(\Sigma_{yy}^{-1})^T\Sigma_{xy}^T + \Sigma_{xy}\Sigma_y^{-1}\Sigma_y(\Sigma_y^{-1})^T\Sigma_{xy}^T \quad (3.129)$$

$$= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T - \Sigma_{xy}(\Sigma_{yy}^{-1})^T\Sigma_{xy}^T + \Sigma_{xy}(\Sigma_{yy}^{-1})^T\Sigma_{xy}^T \quad (3.130)$$

$$= \Sigma_{xx} - \Sigma_{xy} [\Sigma_{yy}^{-1} - (\Sigma_{yy}^{-1})^T + (\Sigma_{yy}^{-1})^T] \Sigma_{xy}^T \quad (3.131)$$

$$= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \quad (3.132)$$

$$\boxed{\text{Cov}(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}} \quad (3.133)$$

Armed with this identity for joint Guassian distributions, we are ready to derive the predictive distribution for Gaussian Process Regression. We find:

$$p(\mathbf{f}_*|X_*, X, \mathbf{y} = \mathcal{N}(K_*^T K^{-1} \mathbf{f}, K_{**} - K_*^T K^{-1} K_*) \quad (3.134)$$

To account for noisy observations, we can augment our correlation function to include a noise offset. The joint distribution then becomes:

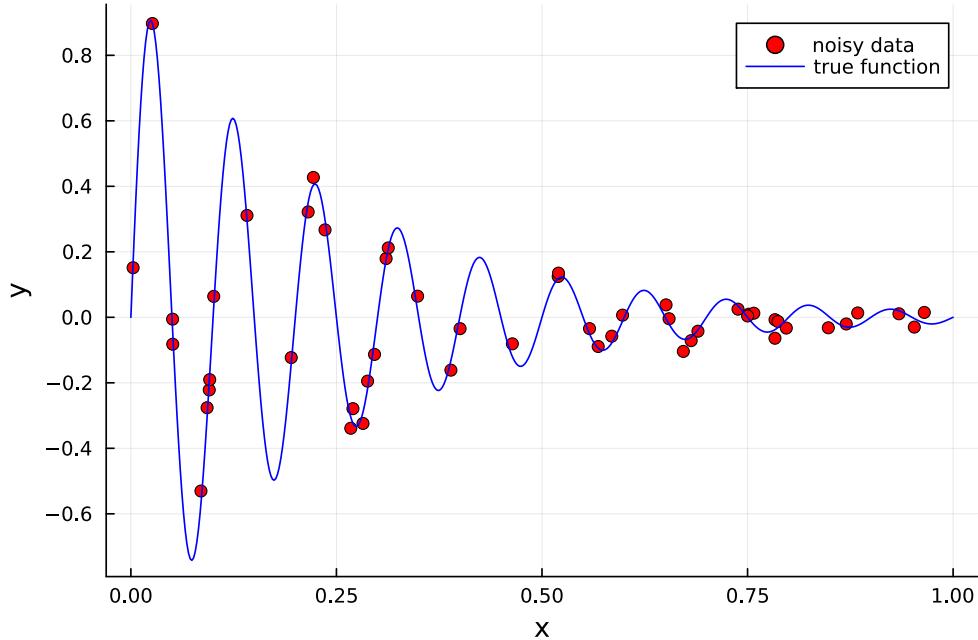
$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} K(X, X) - \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \quad (3.135)$$

which leads to the predictive distribution

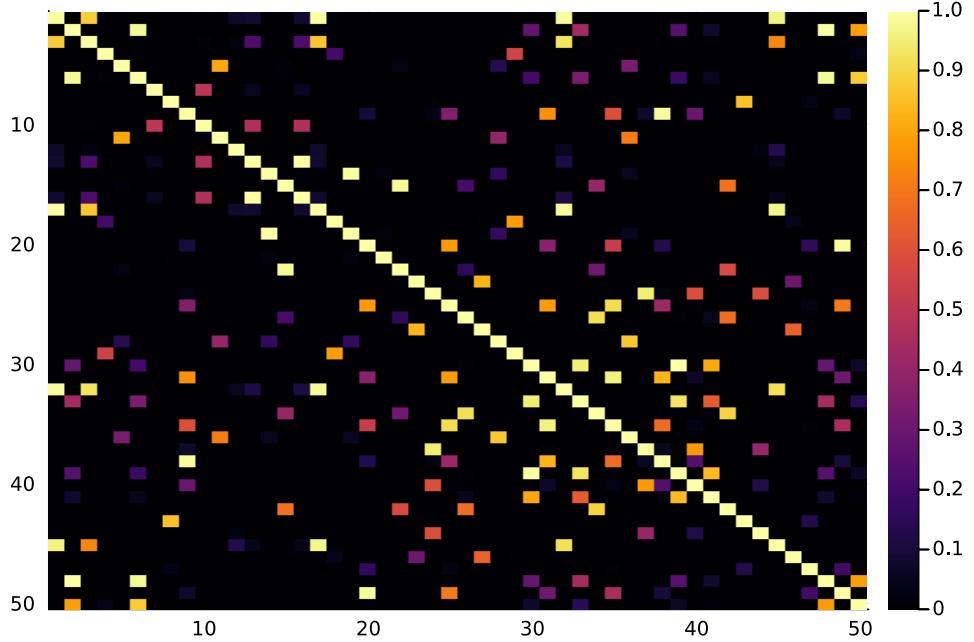
$$p(\mathbf{f}_* | X_*, X, \mathbf{y}) = \mathcal{N} \left( K_*^T [K + \sigma_n^2 I]^{-1} \mathbf{f}, K_{**} - K_*^T [K + \sigma_n^2 I]^{-1} K_* \right) \quad (3.136)$$

## Doing it in Julia

To see how we can use this in practice, let's first construct some sample data which we seek to model as a Gaussian Process



The excellent package KernelFunctions.jl provides a clean interface to create various kernel functions and apply them to data to create our  $K$ -matrices. Due to the fact that kernel functions obey composition laws, we can easily build up complicated Kernels from basic pieces via function composition with  $\circ$



Unsurprisingly, there is a lot of activation on the diagonal as for a single datapoint  $\mathbf{x}$ , we have

$$k(\mathbf{x}, \mathbf{x}) = \exp\left(-\frac{0}{2\ell^2}\right) = 1.0 \quad (3.137)$$

The package `AbstractGPs.jl` provides an excellent way to define Gaussian Processes by supplying mean and kernel functions. We can then sample from our GPs with a simple interface designed to extend the basic functions from `Statistics.jl`. From an `AbstractGP` we can construct a `FiniteGP` by *indexing* into our datasets. The procedure can be summarized as follows

1. Build a kernel function  $k(\cdot, \cdot)$  via composition using `KernelFunctions.jl`
2. Construct an a Gaussian Process  $f \sim \mathcal{GP}$  abstractly using `AbstractGPs.jl`
3. Construct a finite representation of our GP,  $f_x$ , over our training data
4. Construct a posterior Gaussian Process from  $f_x$  and our training targets  $\mathbf{y}$ .
5. Construct a finite representation of the posterior GP applied to our prediction data

6. Sample this final distribution to obtain a prediction via `mean()` and variances via `var()`. Alternatively, we can obtain a multivariate normal distribution for each point by calling `marginals()`.

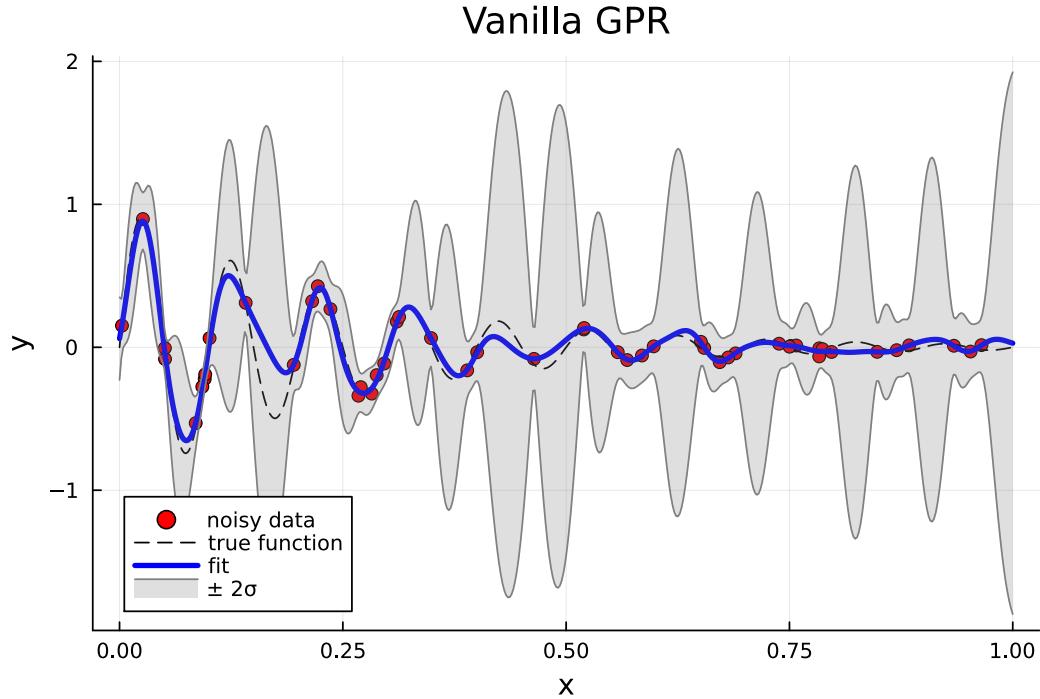


Figure 3.3: A Gaussian Process fit to the training data illustrating the prediction (means) and a  $\pm 2\sigma$  uncertainty interval. We can clearly see how the uncertainty is larger for inputs far away from supplied training data.

## Fitting the Kernel Hyperparameters

At this point, it is easy to think we are finished; we *already* fit the Gaussian process. However, we were forced to choose values for both  $\ell$  and  $\sigma^2$  kernel parameters. How can we optimally select the ideal hyperparameters for our Gaussian Process? This question leads us into the realm of Bayesian Model Selection. Rather than focusing specifically on our Gaussian Process model, let's take a step back and think about the process of model selection from a Bayesian perspective.

There are several levels of parameters in machine learning. At the lowest level, we have the model weights  $\mathbf{w}$ . Above that, we have model hyperparameters,  $\theta$ . At the top we have model structure  $\mathcal{H}$ . In our Bayesian framework, we can consider prior distributions defined at each of these levels which codify credences for how much we trust a particular model, hyperparameter, etc. At the bottom, we have

$$p(\mathbf{w}|X, \mathbf{y}, \theta, \mathcal{H}_i) = \frac{p(\mathbf{y}|X, \mathbf{w}, \theta, \mathcal{H}_i)p(\mathbf{w}|\theta, \mathcal{H}_i)}{p(\mathbf{y}|X, \theta, \mathcal{H}_i)} \quad (3.138)$$

If this looks confusing, consider Bayes rule for 3 events  $R, H, S$ . We have:

$$P(R|H, S) = \frac{P(R, H, S)}{P(H, S)} \quad (3.139)$$

$$= \frac{P(H|R, S)P(R, S)}{P(H, S)} \quad (3.140)$$

$$= \frac{P(H|R, S)P(R|S)P(S)}{P(H|S)P(S)} \quad (3.141)$$

$$= \frac{P(H|R, S)P(R|S)}{P(H|S)} \quad (3.142)$$

To get the result, just think of  $\theta$  and  $\mathcal{H}_i$  as a single *event* and translate the above to distribution functions. See stack exchange link.

The prior  $p(\mathbf{w}|\theta, \mathcal{H}_i)$  encodes any knowledge we have about the parameters prior to seeing the data. The denominator is the *marginal likelihood* and is given by

$$p(\mathbf{y}|X, \theta, \mathcal{H}_i) = \int d\mathbf{w} p(\mathbf{y}|X, \mathbf{w}, \theta, \mathcal{H}_i)p(\mathbf{w}|\theta, \mathcal{H}_i) \quad (3.143)$$

The next level up is to express the distribution of hyper-parameters  $\theta$ :

$$p(\theta|X, \mathbf{y}, \mathcal{H}_i) = \frac{p(\mathbf{y}|X, \theta, \mathcal{H}_i)p(\theta|\mathcal{H}_i)}{p(\mathbf{y}|X, \mathcal{H}_i)}. \quad (3.144)$$

Here  $p(\theta|\mathcal{H}_i)$  is called the *hyper-prior*. Similarly, the normalization constant is given by

$$p(\mathbf{y}|X, \mathcal{H}_i) = \int d\theta p(\mathbf{y}|X, \theta, \mathcal{H}_i)p(\theta|\mathcal{H}_i). \quad (3.145)$$

Finally, at the top level we have the set of possible model structures  $\{\mathcal{H}_i\}$ . This leads to

$$p(\mathcal{H}_i|X, \mathbf{y}) = \frac{p(\mathbf{y}|X, \mathcal{H}_i)p(\mathcal{H}_i)}{p(\mathbf{y}|X)} \quad (3.146)$$

with normalization constant

$$p(\mathbf{y}|X) = \sum_i p(\mathbf{y}|X, \mathcal{H}_i)p(\mathcal{H}_i). \quad (3.147)$$

Depending on the model details, these integrals may be intractable without approximations or Monte Carlo methods. Since we rarely have sufficient knowledge to form a hyperparameter prior, one often attempts to maximize the marginal likelihood  $p(\mathbf{y}|X, \theta, \mathcal{H}_i)$  with respect to the hyperparameters  $\theta$  instead. This is known as Type II Maximum Likelihood Estimation. (Refer to prior section on this in the Theoretical Techniques chapter). In the case of Gaussian Process Regression, we are once again saved by the fact that every piece has a convenient functional from resulting in analytically tractible integrals for the marginal likelihood function. We find

$$\ln p(\mathbf{y}|X, \theta) = -\frac{1}{2}\mathbf{y}^T(K_f + \sigma_n^2 I)^{-1}\mathbf{y} - \frac{1}{2}\ln|K_f + \sigma_n^2 I| - \frac{n}{2}\ln(2\pi) \quad (3.148)$$

where we have employed the natural logarithm remove the pesky exponentials and arrive at a very convenient form. Note that because the lograithm is monotonically increasing, the maximum of the log-marginal-likelihood will be the same as the marginal-likelihood itself. Provided this functional form, we can use our favorite optimization routine to obtain the kernel hyperparameters which maximize the likelihood of obtaining our data as illustrated in the following comparison figure.

### 3.2.3 Decision Trees

## 3.3 Unsupervised Classification

### 3.3.1 Self Organizing Maps

Self organizing maps (SOMs) are an unsupervised machine learning technique developed by Kohonen (Kohonen, 1982) based on the simple biological principle that *neurons near each*

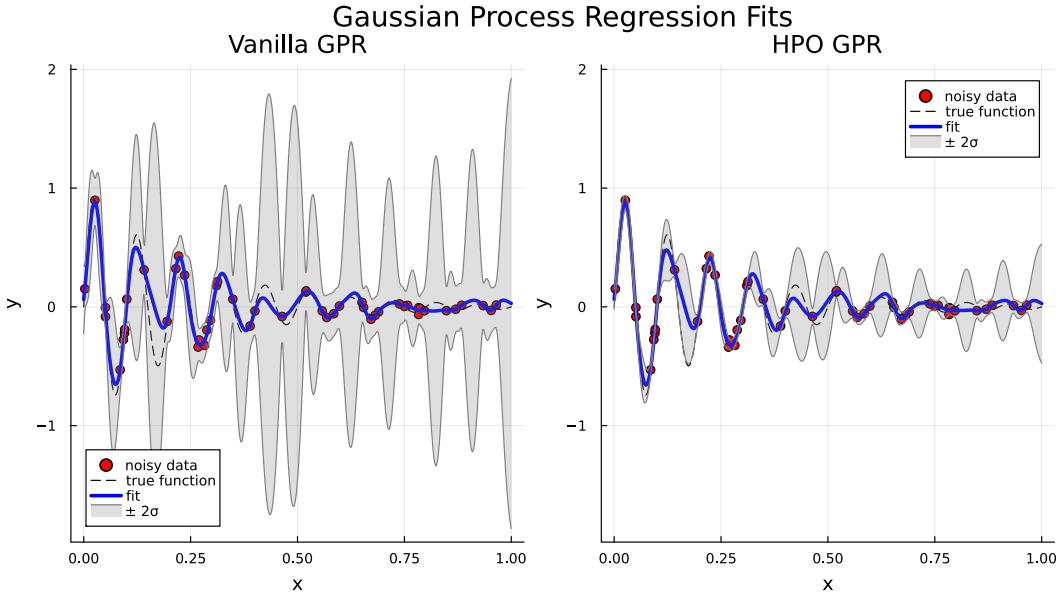


Figure 3.4: Left: the original Gaussian Process obtained with our default choice of kernel parameters. Right: A much better Gaussian Process obtained after performing hyperparameter optimization on the kernel parameters. Note how the mean function still does an excellent job fitting the data points while *also* minimizing the uncertainty of the fit.

*other fire together.* This observation that the topological *closeness* of similar computational units is a critical feature of intelligent systems leads to a natural reinterpretation of the familiar perceptron model into a new form amenable for a variety of clustering and dimensionality reduction tasks. In particular, the SOM enables a rapid unsupervised classification of multidimensional data into a (typically) one or two dimensional *simplcial complex*, the discrete realization of a topological manifold, whose vertices correspond to representative points in the original data space  $\mathcal{D}$ . While a tad esoteric compared to other popular unsupervised methods like KMeans clustering or DBSCAN, the SOM distinguishes itself with the added benefit that it's training procedure guarantees nodes (i.e. classes) close to each other in the feature manifold share similar weights. This additional structure makes the SOM particularly attractive when an interpretation of the discovered clusters as well as the relationships between them is desired.

The original treatment of the SOM by Kohonen was made in terms of processing units, sensory signals, and relaying networks (Kohonen, 1982), however, in the modern era of deep learning, a more easily digestible derivation can be obtained by re-interpreting the weights of a simple perceptron model to provide the foundation for a clustering approach. As described in previously, a perceptron is a function of the form

$$\mathbf{y} = \sigma. (W\mathbf{x}) \quad (3.149)$$

where  $W \in \mathbb{R}^{n \times m}$  is a matrix of weights which transform the input  $\mathbf{x} \in \mathbb{R}^m$  into  $\mathbb{R}^n$  and  $\sigma$  is a nonlinear *activation function* applied element-wise to the outputs of the matrix multiplication (indicated by the . syntax). If we instead think of the weight matrix as an ordered collection of vectors  $\{\mathbf{w}_i\}_{i=1}^n$ , then this formula can be further decomposed into

$$(\mathbf{y})_i = \sigma(\mathbf{w}_i^T \mathbf{x}) = \sigma(\langle \mathbf{w}_i, \mathbf{x} \rangle) \quad (3.150)$$

The function of the perceptron is now clear: given an input vector  $\mathbf{x}$  and a collection of  $n$ -many weight vectors  $\mathbf{w}_i$ , compute the  $n$ -many inner products of  $\mathbf{x}$  with each weight vector  $\mathbf{w}_i$ , apply the nonlinear activation function  $\sigma$ , and concatenate the results.

If we now allow ourselves to imagine the weight vectors  $\mathbf{w}_i$  as members of the same vector space as the input  $\mathbf{x}$ , a reasonable question to ask is: *how similar is the input  $\mathbf{x}$  to each  $\mathbf{w}_i$* . Further, the application of the inner product  $\langle \cdot, \cdot \rangle$  suggests we may answer this question in terms of the distance

$$\langle \mathbf{w}_i - \mathbf{x}, \mathbf{w}_i - \mathbf{x} \rangle = d(\mathbf{w}_i, \mathbf{x})^2. \quad (3.151)$$

In other words, given a set of weight vectors  $\mathbf{w}_i$  which we may now think of as the cluster centers for our unsupervised model, we can measure the similarity between a given datum  $\mathbf{x}_j$  and each cluster by computing the distance

$$d_{ij} = d(\mathbf{w}_i, \mathbf{x}_j). \quad (3.152)$$

To make this mapping useful, we should prescribe a *training procedure* which updates the vectors  $\mathbf{w}_i$  based on all of the available data points. Further, our goal is to establish some kind of interpretable relationship between the various weights  $\mathbf{w}_i$  so that they *more than* disjoint classes. The SOM achieves this by providing a lower-dimensional grid (typically 2-dimensional) whose vertices are understood to be the location of each of the weight vectors. The procedure is as follows:

First, initialize a grid of weight vectors  $\mathbf{w}_i \in \mathbb{R}^m$  which we collect into a matrix  $W \in \mathbb{R}^{m \times n}$ . Assign coordinates to each weight vector by mapping each  $\mathbf{w}_i$  to a vertex in the grid. The *topology* of the grid is a hyperparameter which dictates the spatial relationship between neighboring nodes. Popular choices are: flat rectangular, rectangular on a cylinder (one periodic boundary condition), rectangular on a torus (two periodic boundary conditions), hexagonal (more equidistant neighbors per node than rectangular), and spherical. One can

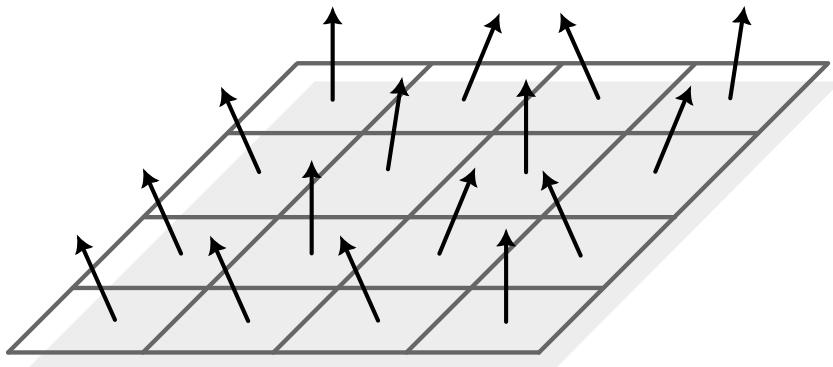


Figure 3.5: A Self Organizing Map with nodes configured in a rectangular grid topology with weight vectors randomly initialized.

initialize the weights in a variety of ways. Some common choices are

- Randomly initialize them as in figure 3.5.

- Initialize them to the value of  $n$ -many randomly selected data samples  $\mathbf{x}_i$ .
- Initialize them to the first  $n$ -many principal components of the dataset.

Next, we proceed to update the weights according to the following steps for each  $\mathbf{x}_k$  datum:

1. Compute the *best match unit* (BMU) as

$$\mathbf{w}_i^{\text{best}} = \arg \min_{\mathbf{w}_j} (d(\mathbf{x}_k, \mathbf{w}_j)^2) \quad (3.153)$$

where  $d(\cdot, \cdot)$  is some suitably chosen distance function. One often defaults to the Euclidean metric.

2. Choose a radius  $\sigma_t$  which we will use to identify the correct update for to apply to all nodes relative to  $\mathbf{w}_i^{\text{best}}$ .
3. Update *all* the weights according to the equation

$$w_m^{t+1} = w_m^t + \eta(t) f(x_m, y_m, \sigma_t) (\mathbf{x}_k - \mathbf{w}_n^t) \quad (3.154)$$

where  $\eta(t)$  is the learning rate, and  $f(x_m, y_m, \sigma_t)$  is the *neighborhood* function which defines the relative amount each node  $\mathbf{w}_m$  is updated according to its coordinate distance from the best match unit (i.e. the distance between the points  $(x_m, y_m)$  and  $(x_{\text{best}}, y_{\text{best}})$  in the two dimensional case). Typically one chooses a the learning rate to evolve according to the schedule

$$\eta(t) = \eta_0 \exp(-\lambda t) \quad (3.155)$$

with the radius  $\sigma_t$  evolving similarly according to

$$\sigma_t = \sigma_0 \exp(-\beta t). \quad (3.156)$$

Popular choices for the neighborhood function are the Gaussian function, Mexican-hat function, a cone function, and a cylinder function.

## A simple example: partitioning color spaces

As an illustrative example, consider the minimal dataset formed by the colors red, green, and blue which as vectors may be written as

$$\text{Red} = (1, 0, 0) \quad (3.157)$$

$$\text{Green} = (0, 1, 0) \quad (3.158)$$

$$\text{Blue} = (0, 0, 1). \quad (3.159)$$

Now let us suppose we want to train an SOM to generate a classification for colors using only these three as the supplied training data. Figure 3.6 shows exactly this. Here we have trained an SOM of size  $25 \times 25$  cells in a flat hexagonal topology. As the weight vectors are themselves vectors of length 3, we can then visualize the trained SOM by coloring each corresponding node by the color learned in its weight vector. What we find is that the resulting map has clearly separated red, green, and blue into distinct regions in the grid, and more importantly, there is a smooth gradient between neighboring cells reflecting the fact that neighboring classes are more *similar* than cells with a large separation distance. This is not the case for many other common clustering techniques like k-nearest neighbors or DBSCAN for which relationship between learned classes is challenging to interpret.

To enable this demonstration and its subsequent use throughout the rest of this dissertation, I have created a new open-source implementation of the SOM algorithm for the Julia programming language called `SelfOrganizingMaps.jl` which has been added to the general registry and can be freely downloaded for use by anyone. The repository for the code as well as the associated documentation can be found at this site.

### 3.3.2 Generative Topographic Maps

Now that we've developed the SOM algorithm, we see upon reflection that there are a few drawbacks to the methods, namely

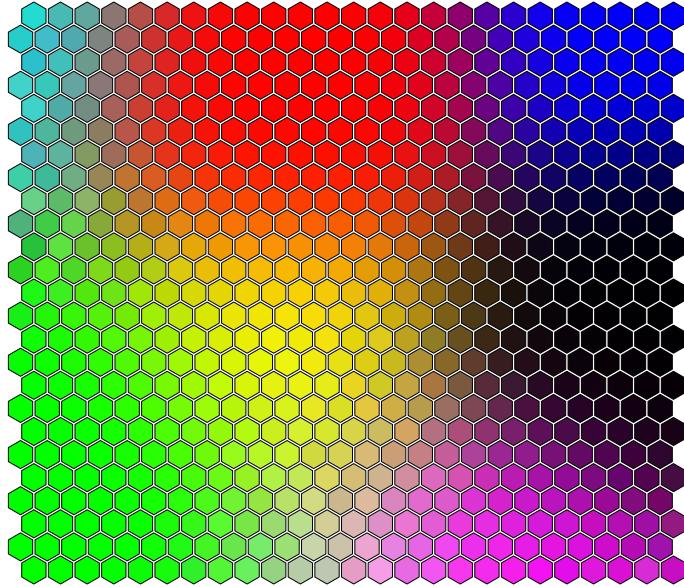


Figure 3.6: A trained Self Organizing Map of  $25 \times 25$  cells in a hexagonal topology trained on a dataset consisting of the three colors red, green, and blue.

- There is no probabilistic interpretation of the fitted SOM? Are we to trust the fit results with 100% certainty?
- There is no clearly stopping criteria for the learning process? How many epochs should be performed? Do you simply stop the training once the neighbor radius  $\sigma_t$  is less than the minimum neighbor distance?

With this in mind, we now seek to introduce a new method based on the SOM which will allow us to address the shortcomings by taking a principled, probabilistic approach, which will be called the *Generative Topographic Mapping* (GTM). This technique was first introduced by Bishop et al. in (Bishop et al., 1998) and our derivation of the method will follow their presentation closely.

Both the SOM and GTM rely on the assumption that our high dimensional dataset  $\mathcal{D} \subseteq \mathbb{R}^D$  can be accurately represented as being constrained to a lower dimensional submanifold. In the SOM this is achieved by assigning coordinates to each of the weight vectors

according to some predetermined grid. In the GTM, however, we make a slight change in our perspective and say that the high dimensional data we have are *generated* from some lower dimensional set of *latent vectors* living in  $\mathbb{R}^L$  with  $L < D$ . For ease of visualization, it is standard to chose  $L = 2$ . Our goal, then, is to learn a mapping from the latent space to the data space which *maximizes the likelihood* of obtaining our dataset given the set of latent variables.

In the coming derivation, we shall assume the following notations:

$$\begin{aligned} \mathbf{x} \in \mathbb{R}^L &\quad \text{A latent vector} \\ \mathbf{t} \in \mathbb{R}^D &\quad \text{A data vector} \\ \mathbf{y} = \mathbf{y}(\mathbf{x}; W) &\quad \text{The transformation } \mathbf{y} : \mathbb{R}^L \rightarrow \mathbb{R}^D \\ W &\quad \text{The model weights} \end{aligned} \tag{3.160}$$

To get started, let us assume that our data can be reasonably described by a multivariate normal distribution such that

$$p(\mathbf{t}|\mathbf{x}; W, \beta) = \mathcal{N}\left(\mathbf{y}(\mathbf{x}; W), \frac{1}{\beta}\right), \tag{3.161}$$

that is, our data follow a normal distribution with mean given by the transformation function  $\mathbf{y}$  applied to the latent variables  $\mathbf{x}$  with covariance  $\mathbf{1}\beta$ . Suppose that we have some model for the prior distribution of our latent variables,  $p(\mathbf{x})$ , then by integration would find

$$p(\mathbf{t}|W, \beta) = \int d\mathbf{x} p(\mathbf{t}|\mathbf{x}; W, \beta)p(\mathbf{x}) \tag{3.162}$$

which is the distribution we really care about, namely, the distribution of data  $\mathbf{t} \in \mathcal{D}$  given our choice of model weights  $W$  and covariance  $\beta^{-1}$ . To make the problem tractible, we need an integrable distribution for  $p(\mathbf{x})$ . We therefore take inspiration from the SOM and choose  $\mathbf{x}$  to be precisely distributed over a regular grid of points with equal weights. That is

$$p(\mathbf{x}) = \frac{1}{K} \sum_k^K \delta(\mathbf{x} - \mathbf{x}_k) \tag{3.163}$$

where  $\mathbf{x}_k$  are the  $K$ -many grid points such that for any  $\mathbf{x}$ , there is a probability of precisely  $1K$  that it *came from* one of the nodes  $\mathbf{x}_k$ .

**NOTE:** Add figure of GTM grid here. (See pg 27 of notebook)

As mathematicians (or physicists, or data scientists, etc...) we rejoice at the wise decision to incorporate the integral-zapping Dirac-delta distributions which enables us to simplify the data distribution to become

$$\begin{aligned} p(\mathbf{t}|W, \beta) &= \int d\mathbf{x} p(\mathbf{t}|\mathbf{x}; W, \beta) \frac{1}{K} \sum_k^K \delta(\mathbf{x} - \mathbf{x}_k) \\ &= \frac{1}{K} \sum_k^K p(\mathbf{t}|\mathbf{x}_k; W, \beta). \end{aligned} \quad (3.164)$$

Now provided we have in supply a dataset with  $N$ -many records, i.e.  $\mathcal{D}^N \subseteq \mathcal{D} = \{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N\}$ , how do we chose the *optimal* parameters  $W$  and  $\beta$ ? We need an objective to optimize!

If we assume the  $\mathbf{t}_i$  are independently, identically distributed, then the *likelihood* of obtaining the data given a choice of parameters (i.e.  $p(\mathcal{D}^N|W, \beta)$ ) is

$$\begin{aligned} \mathcal{L}(W, \beta) &= \prod_n^N p(\mathbf{t}_n|W, \beta) \\ &= \prod_n^N \left( \frac{1}{K} \sum_k^K p(\mathbf{t}_n|\mathbf{x}_k; W, \beta) \right). \end{aligned} \quad (3.165)$$

Taking the logarithm (which must share the same optimum as it is a monotonically increasing function) yields the simpler expression

$$\ell(W, \beta) = \log(\mathcal{L}(W, \beta)) = \sum_n^N \log \left( \frac{1}{K} \sum_k^K p(\mathbf{t}_n|\mathbf{x}_k; W, \beta) \right) \quad (3.166)$$

Maximizing this log-likelihood function is known as *maximum likelihood estimation* (MLE). But how should we optimize this function? We could use techniques like gradient descent or ADAM, however the probabilistic nature of the GTM suggests we may be able to obtain an *expectation-maximization* (EM) routine for a suitable choice of transformation  $\mathbf{y}(\mathbf{x}, W)$ .

Suppose we already have some guesses  $W_o$  and  $\beta_o$  for the parameters ( $o$  for *old*). To simplify the derivation, we define  $\theta_o = (W_o, \dots, \beta_o)$ . Then we can compute the responsibilities  $r_{kn}$  as

$$\begin{aligned}
r_{kn} &:= p(\mathbf{x}_k | \mathbf{t}_n, \theta_o) \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) p(\mathbf{x}_k | \theta_o)}{p(\mathbf{t}_n | \theta_o)} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) p(\mathbf{x}_k | \theta_o)}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o) p(\mathbf{x}_{k'} | \theta_o)} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) \frac{p(\mathbf{x}_k, \theta_o)}{p(\theta_o)}}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o) \frac{p(\mathbf{x}_{k'}, \theta_o)}{p(\theta_o)}} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) p(\mathbf{x}_k, \theta_o)}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o) p(\mathbf{x}_{k'}, \theta_o)} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) p(\mathbf{x}_k) p(\theta_o)}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o) p(\mathbf{x}_{k'}) p(\theta_o)} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) p(\mathbf{x}_k)}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o) p(\mathbf{x}_{k'})} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o) \frac{1}{K}}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o) \frac{1}{K}} \\
&= \frac{p(\mathbf{t}_n | \mathbf{x}_k, \theta_o)}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}, \theta_o)}
\end{aligned} \tag{3.167}$$

which we will soon utilize in our EM procedure.

Now, we must chose a form for the transformation  $\mathbf{y}(\mathbf{x}, W)$  after which we will have all of the information we need to perform the computation. A convenient choice is to use a kernelized regression strategy so that

$$\mathbf{y} := \phi^T(\mathbf{x})W \tag{3.168}$$

where  $W \in \mathbb{R}^{M \times D}$  are some constant parameters and  $\phi : \mathbb{R}^L \rightarrow \mathbb{R}^M$  is a transformation employing  $M$ -many basis functions  $\phi_m$ . To capture linear and nonlinear effects, we chose

$$\phi_m(\mathbf{x}) = \begin{cases} \exp(-\frac{|\mathbf{x} - \mu_m|^2}{2\sigma}), & m < M \\ 1, & m = M \end{cases} \tag{3.169}$$

that is,  $M - 1$  Gaussians and a term for a bias offset. If we apply the transformation to each of the  $K$ -many latent nodes, then we may collect the resulting vectors into a matrix,

$$Y = \Phi W, \quad (3.170)$$

where  $\Phi \in \mathbb{R}^{K \times M}$  with  $\Phi_{km} = \phi_m(\mathbf{x}_k)$ .

Now that we are able to compute the responsibilities of each latent node to the observed data (the expectation step), we need to compute the relevant derivatives which we will use in the maximization step. For this, we treat the responsibilities  $r_{kn}$  as fixed, and solve for the relevant updates to  $W$  and  $\beta$  which make these derivatives 0 and therefore optimize  $\ell$ . Let's begin by differentiating  $\ell$  with respect to the weight matrix  $W$ . We have

$$0 = \frac{\partial}{\partial W_{md}} \ell(W, \beta) \quad (3.171)$$

$$= \frac{\partial}{\partial W_{md}} \sum_n^N \log \left( \frac{1}{K} \sum_k^K p(\mathbf{t}_n | \mathbf{x}_k; W, \beta) \right) \quad (3.172)$$

$$= \sum_n^N \frac{1}{\frac{1}{K} \sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}; W, \beta)} \frac{1}{K} \frac{\partial}{\partial W_{md}} \sum_k^K p(\mathbf{t}_n | \mathbf{x}_k; W, \beta) \quad (3.173)$$

Noting that  $p(\mathbf{t}_n | \mathbf{x}_k; W, \beta) = \mathcal{N}(\mathbf{y}_k, \beta^{-1})$ , then the derivative of the exponential yields

$$0 = \sum_n^N \sum_k^K \frac{p(\mathbf{t}_n | \mathbf{x}_k; W, \beta)}{\sum_{k'}^K p(\mathbf{t}_n | \mathbf{x}_{k'}; W, \beta)} \frac{\partial}{\partial W_{md}} \left( -\frac{\beta}{2} |\mathbf{t}_n - \mathbf{y}_k|^2 \right) \quad (3.174)$$

$$= \sum_n^N \sum_k^K r_{kn} \frac{\partial}{\partial W_{md}} \left( -\frac{\beta}{2} |\mathbf{t}_n - \mathbf{y}_k|^2 \right) \quad (3.175)$$

$$= \sum_n^N \sum_k^K (-\beta) r_{kn} \sum_q^D (y_k^q - t_n^q) \frac{\partial}{\partial W_{md}} \sum_s^M \phi_s(\mathbf{x}_k) W_{sq} \quad (3.176)$$

$$= \sum_n^N \sum_k^K \sum_q^D (-\beta) r_{kn} (y_k^q - t_n^q) \sum_s^M \phi_s(\mathbf{x}_k) \delta_{sm} \delta_{qd} \quad (3.177)$$

$$= \sum_n^N \sum_k^K (-\beta) r_{kn} (y_k^d - t_n^d) \phi_m(\mathbf{x}_k) \quad (3.178)$$

Looking closely at the above expression, we see that there are two free indices suggesting we may write the above as a matrix expression. To do so, let's introduce a diagonal matrix  $G$  such that  $G_{kk} = \sum_n r_{kn}$ . Then upon rearrangement, we find

$$\sum_n^N \sum_k^K r_{kn} y_k^d \phi_m(\mathbf{x}_k) = \sum_n^N \sum_k^K r_{kn} t_n^d \phi_m(\mathbf{x}_k) \quad (3.179)$$

$$\sum_n \sum_k r_{kn} \left( \sum_s W_{sd} \phi_s(\mathbf{x}_k) \right) \phi_m(\mathbf{x}_k) = \sum_n \sum_k r_{kn} t_n^d \phi_m(\mathbf{x}_k) \quad (3.180)$$

$$\sum_n \sum_k \sum_s r_{kn} \Phi_{ks} W_{sd} \Phi_{km} = \sum_n \sum_k r_{kn} t_n^d \Phi_{km} \quad (3.181)$$

$$\sum_k \sum_s \left( \sum_n r_{kn} \right) \Phi_{ks} W_{sd} \Phi_{km} = \sum_n \sum_k r_{kn} t_n^d \Phi_{km} \quad (3.182)$$

$$\sum_k \sum_s G_{kk} \Phi_{ks} W_{sd} \Phi_{km} = \sum_n \sum_k r_{kn} t_n^d \Phi_{km} \quad (3.183)$$

$$\sum_k \sum_s (\Phi_{km})^T G_{kk} \Phi_{ks} W_{sd} = \sum_n \sum_k (\Phi_{km})^T r_{kn} t_n^d \quad (3.184)$$

$$(\Phi^T G \Phi W)_{md} = (\Phi^T R T)_{md} \quad (3.185)$$

where we have defined  $T_{nd} = t_n^d$  and  $R_{kn} = r_{kn}$ .

By the same procedure, we find that differentiation with respect to  $\beta$  leads to

$$\frac{1}{\beta} = \frac{1}{ND} \sum_n^N \sum_k^K r_{kn} |\mathbf{y}_k - \mathbf{t}_n|^2 \quad (3.186)$$

so that together the maximization step amounts to solving

$$\begin{cases} \Phi^T G_{\text{old}} \Phi W_{\text{new}} = \Phi^T R_{\text{old}} T \\ \frac{1}{\beta_{\text{new}}} = \frac{1}{ND} \sum_n^N \sum_k^K R_{kn}^{(\text{old})} |\mathbf{y}_k - \mathbf{t}_n|^2 \end{cases} \quad (3.187)$$

The last piece we need for the training algorithm is a method to initialize the parameters  $W$  and  $\beta$ . As before with the SOM, there are a few different options. In particular we might

- Randomly initialize the weight matrix  $W$

- Use PCA to initialize the weights to a linear model. To do this, one computes the data covariance matrix  $U$  keeping only the first two columns. We then initialize  $W$  so that

$$W\Phi^T \approx UX^T \quad (3.188)$$

and then set  $\beta^{-1}$  to the third principal component variance.

In summary, the GTM training procedure is as follows

1. Generate a grid of latent points  $\{\mathbf{x}_k\}_{k=1}^K$ .
2. Generate basis function centers  $\{\mu_m\}_{m=1}^M$ .
3. Select an basis function width  $\sigma$ .
4. Compute the matrix of activations  $\Phi$  where

$$\Phi_{mk} = \Phi_m(\mathbf{x}_k) \quad (3.189)$$

5. Initialize the weight matrix  $W$  randomly or with PCA
6. Initialize  $\beta$  randomly or with PCA
7. (Optional) select a value for  $\alpha$  to enable weight regularization, i.e.  $(\Phi^T G \Phi + \frac{\alpha}{\beta} I)W = \Phi^T RT$ . This corresponds to specifying a prior distribution on the weights  $W$  such that  $p(W) \propto \exp(-\alpha||W||^2/2)$ .
8. Compute the difference matrix  $\Delta := |T - \Phi W|^2$
9. Repeat the following until convergence:
  - (Expectation step) Compute the responsibility matrix  $R$  using  $\Delta$  and  $\beta$ .
  - (b) Compute  $G$  from  $R$

- (c) (Maximization step) Compute the update to the weight matrix  $W$  with  $W_{\text{new}} = (\Phi^T G \Phi)^{-1} \Phi^T R T$
- (d) Compute the updated difference matrix  $\Delta$
- (e) Compute the updated  $\beta$

Excellent! We now have a robust procedure to fit the GTM model. With a fitted GTM in hand, how do we utilize the results? The main idea is that the fitted GTM provides us all of the relevant probability distributions to explain how we obtained our current data  $\mathbf{t}_n$  from the latent vectors  $\mathbf{x}_k$ . Using Baye's rule, we can invert this relationship using our responsibility matrix  $R$  to understand the contributions of each latent vector  $\mathbf{x}_k$  to each data. Since this would lead to a different  $R$  matrix for each datapoint, one often computes a handful of summary statistics using  $R$  rather than using the entire result. For example, we find the mean of the distribution to be

$$\begin{aligned}
\langle \mathbf{x} | \mathbf{t}_n; W, \beta \rangle &= \int d\mathbf{x} p(\mathbf{x} | \mathbf{t}_n; W, \beta) \mathbf{x} \\
&= \int d\mathbf{x} \frac{p(\mathbf{t}_n | \mathbf{x}; W, \beta) p(\mathbf{x})}{\int d\xi p(\mathbf{t}_n | \xi; W, \beta) p(\xi)} \mathbf{x} \\
&= \int \sum_k \frac{p(\mathbf{t}_n | \mathbf{x}, W, \beta) \mathbf{x} \delta(\mathbf{x} - \mathbf{x}_k)}{\sum_{k'} p(\mathbf{t}_n | \mathbf{x}_{k'}; W, \beta)} d\mathbf{x} \\
&= \sum_k \frac{p(\mathbf{t}_n | \mathbf{x}_k; W, \beta) \mathbf{x}_k}{\sum_{k'} p(\mathbf{t}_n | \mathbf{x}_{k'}; W, \beta)} \\
&= \sum_k^K R_{kn} \mathbf{x}_k
\end{aligned} \tag{3.190}$$

However, in cases where the distribution is multi-modal, the mean can be misleading. It can therefore also be useful to compute the mode of the distribution which is given by

$$\mathbf{x}_{\text{mode}} = \mathbf{x}_{k_{\max}} \tag{3.191}$$

where  $k_{\max} = \arg \max_k (R_{kn})$ .

To provide an easy to use open-source implementation of the GTM algorithm, I have created a publicly accessible github repo with a Julia implementation that comports with the `MLJ.jl` machine learning framework. The repo can be found [here](#) and makes it easy to incorporate GTMs into complicated machine learning pipelines.

UPDATE REQUIRED: Add demo use for GTM like we did for the SOM.

### 3.4 Uncertainty Quantification via Conformal Prediction

Having established a variety of machine learning methods which we will use throughout the rest of this dissertation, a natural next question is: How can we evaluate our confidence in the predictions of machine learning models? For some methods like Gaussian Process Regression, our models naturally output distributions which we can evaluate to provide predictions via the mean, and uncertainty estimates via the standard deviation or some similar statistic. For models like Neural Networks and Decision Trees which are not inherently probabilistic, there is no obvious way to extract uncertainty estimates purely from the model's predictions. One standard approach is to attempt to evaluate our confidence in a particular model by comparing the predictions of many copies of the same model trained on complementary cross validation folds of the original training set. By examining the prediction variance due to variation in the supplied training data, we can establish some expectation for uncertainty in our model's output.

Clearly this approach is far from perfect. Therefore, in order to prevent ourselves from biasing towards using only those methods like Gaussian Process Regression which output distributions by default, we would like to develop a robust procedure for augmenting *any* regression model with the capability to simultaneously estimate target values and *confidence intervals* by taking advantage of our (often) large datasets to establish sufficient statistics. Further, we do not wish to make any assumptions about the particular target distributions; without sufficient reason to expect a variable to be normally distributed why should we make

this *strong* assumption? This is the approach taken by *Conformal Prediction* which we will develop in this section and utilize throughout the rest of this dissertation.

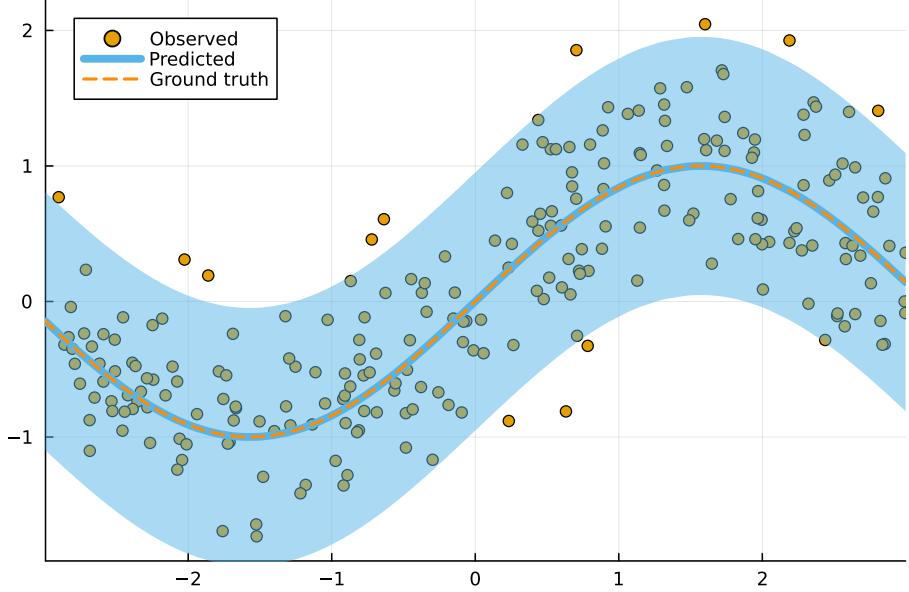


Figure 3.7: An example of conformal prediction for a model estimating a noisy one-dimensional target. The blue shaded region illustrates the learned confidence interval. Image taken from (Altmeyer, 2023)

As discussed in the previous section, the typical machine learning procedure involves partitioning our dataset into independent training and testing batches (or  $k$ -many cross-validation folds) so that we can evaluate a trained model's performance and ensure we have not run afoul of over/under-fitting. In conformal prediction, we introduce an additional split so that our original dataset  $\mathcal{D}$  can be decomposed into  $\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{cal}} \cup \mathcal{D}_{\text{test}}$ . We then start with *any* heuristic notion of model uncertainty, and make the estimated confidence interval rigorous by adjusting the output to *guarantee* the desired coverage is satisfied on the calibration set  $\mathcal{D}_{\text{cal}}$ .

A high level outline of the procedure is as follows

1. Define a score function  $s(x, y) \in \mathbb{R}$  using some heuristic notion of uncertainty so that large  $s$  means *worse* agreement between  $\hat{f}(x)$  and the target  $y$ .

2. Compute  $\hat{q}$  as the  $\frac{\lceil(n+1)(1-\alpha)\rceil}{n}$  quantile of the scores  $s_i = s(x_i, y_i)$  generated from the calibration set  $\mathcal{D}_{\text{cal}}$ .
3. Use  $\hat{q}$  to form valid prediction intervals.

The question then becomes: How can we obtain accurate uncertainty estimates if the score function  $s$  need not be accurate? So long as the scores  $s_i$  accurately rank estimates from best to worst model error, then all we need to do is *calibrate* the scores to achieve a confidence interval with the desired coverage, i.e. so that the interval is always large enough to guarantee the coverage we require. It may be that our predicted interval is *too large* for a poor scoring function. As an example, let's demonstrate how we can form correct estimates for the specific problem of quantile regression.

### 3.4.1 Conformalizing Quantile Regression

In quantile regression the goal is to predict the  $\gamma^{th}$  quantile of the targets  $y$  given the input features  $x$ . Let us denote  $t_\gamma(x)$  as the true value for the  $\gamma^{th}$  quantile and  $\hat{t}_\gamma(x)$  as our predicted value. In this framework, we could then expect the interval  $[t_{0.05}, t_{0.95}]$  to constitute a 90% coverage confidence interval. However, our fitted values  $\hat{t}_\gamma$  will certainly be imperfect so that the actual interval obtained will not conform to the desired 90%. Suppose we have a holdout dataset  $\mathcal{D}_{\text{cal}}$  as previously mentioned, and for notational consistency, let  $\alpha$  be the coverage so that we may write our predicted interval as  $[\hat{t}_{\alpha/2}, \hat{t}_{1-\alpha/2}]$ . A reasonable score function (i.e. uncertainty heuristic) would be

$$s(x, y) = \max \{ \hat{t}_{\alpha/2}(x) - y, y - \hat{t}_{1-\alpha/2}(x) \} \quad (3.192)$$

so that the score is the distance to the nearest of our predicted quantiles. From this score, we may then form

$$\hat{q} = \text{Quantile} \left( \{s(x_i, y_i) | (x_i, y_i) \in \mathcal{D}_{\text{cal}}\}, \frac{\lceil(n+1)(1-\alpha)\rceil}{n} \right), \quad (3.193)$$

that is,  $\hat{q}$  is the  $\frac{[(n+1)(1-\alpha)]}{n}$ -th quantile of the scores from our calibration set. Using these corrections, we then output the *conformalized* interval

$$C(x) = [\hat{t}_{\alpha/2}(x) - \hat{q}, \hat{t}_{1-\alpha/2}(x) + \hat{q}] \quad (3.194)$$

which we have guaranteed achieves the desired coverage on the holdout set. If  $\mathcal{D}_{\text{cal}}$  is sufficiently representative, then we will have achieved a rigorous interval for a coverage of  $\alpha$ . The effect is that  $\hat{q}$  adjusts our previously estimated interval so that a positive  $\hat{q}$  widens the interval and a negative  $\hat{q}$  shrinks it in order to obtain the appropriate coverage.

How do we implement this in practice? Any model that we fit by minimizing the mean squared error (e.g. a Neural Network) can be augmented to instead fit the so-called *pinball loss*

$$L_\gamma(\hat{t}_\gamma, y) = \begin{cases} (y - \hat{t}_\gamma)\gamma; & \hat{t}_\gamma < y \\ (\hat{t}_\gamma - y)(1 - \gamma); & \hat{t}_\gamma \geq y \end{cases} \quad (3.195)$$

For  $\gamma = 0.5$ , this reduces to the standard  $\ell_1$  loss  $\ell_1(\hat{t}, y) = |\hat{t} - y|/2$  which encourages models to learn the median (i.e. the 0.5 quantile).

**NOTE:** Add plot of pinball loss for reference

### 3.4.2 Conformalizing Scalar Uncertainty Estimates

Another common method for producing uncertainty estimates is to simultaneously predict the mean and standard deviation of the target distribution by assuming a Gaussian shape. More generally, we may seek to estimate a model uncertainty directly by first training a model,  $\hat{f}(x)$ , to predict the target  $y$ , and then train a second model  $\hat{r}(x)$  to predict the residual error, that is

$$\hat{r} = |y - \hat{f}(x)|. \quad (3.196)$$

In either case, one could then form a confidence interval as

$$[\hat{f}(x) - \hat{r}(x), \hat{f}(x) + \hat{r}(x)]. \quad (3.197)$$

Other reasonable approaches to extract a heuristic uncertainty from a trained model might include

- Compute the output variance of an ensemble model  $\hat{f}$
- Compute the prediction variance obtained when randomly dropping out nodes from a neural network.
- Estimate the variance of  $\hat{f}$  resulting from small perturbations of the input  $x$ .

For all of these cases, we may call the uncertainty estimate  $u(x)$  and form the score function

$$s(x, y) = \frac{|y - \hat{f}(x)|}{u(x)} \quad (3.198)$$

so that the score function  $s$  measures how well our uncertainty estimate  $u(x)$  matches the actual error  $|y - \hat{f}(x)|$ . If the error is large our predicted uncertainty is small, then  $s$  will be large. Consequently, the score function tells us the correction we must apply to our estimate  $u(x)$  in order to predict the actual model error, that is,

$$|y - \hat{f}(x)| = s(x, y)u(x). \quad (3.199)$$

If we now take  $\hat{q}$  to be the  $\frac{[(n+1)(1-\alpha)]}{n}$ -th quantile as before, we are guaranteed that

$$p(s(x_{\text{cal}}, y_{\text{call}}) \leq \hat{q}) \geq 1 - \alpha \quad (3.200)$$

An easy-to-use implementation of these methods is provided by the open-source Julia package `ConformalPrediction.jl` which is designed to integrate into the larger `MLJ.jl` machine learning framework. We will utilize these techniques throughout the remainder of this dissertation in order to establish uncertainty estimates for *all* of our regression models.

**UPDATE REQUIRED:** we should add a couple more paragraphs referencing the key papers that develop this technique as well as a few example use cases it has been applied to.

### 3.5 Data Assimilation

The proper application of scientific models to make real-world predictions requires that we commit ourselves to a full accounting of all possible sources of uncertainty when reporting results. Further, the explosion of *big data* across scientific fields provides a plethora observational data that our models are typically unequipped to incorporate when making predictions. The field of *Data Assimilation* addresses this problem by providing a family of techniques engineered to combine model output together with observational data whilst enabling a complete accounting the sources of uncertainty. For chaotic systems in particular, data assimilation enables integration on long time scales that would be impossible via models alone. In this overview, we will follow the examples from (Ahmed et al., 2020).

Data assimilation can be cleanly developed in the framework of discrete dynamical systems. Since, at the end of the day, all of our scientific models must be discretized to be evaluated numerically, this is a reasonable course of action. Our goal is to find the best prediction for the system state vector  $u$  that combines our model predictions, also known as *forecasts*, with observational data. Model predictions are described via the discrete update equation:

$$u_{k+1} = \mathcal{M}(u_k; \theta) \quad (3.201)$$

For ODE systems,  $\mathcal{M}$  represents the time integration scheme for a models like

$$\frac{du}{dt} = f(u, t; \theta), \quad (3.202)$$

in other words, a particular choice of ODE integration scheme (Runge Kutta for example) used to evolve the state vector  $u$  from time  $u_k$  to  $u_{k+1}$ .

To measure the performance of our assimilation scheme, we denote the *true* value of the state vector as  $u^{(t)}$ . The output of our model is denoted  $u^{(b)}$  ( $b$  superscript for *background*). The discrepancy between the true value and our forecast is denoted  $\xi^{(b)} = u^{(t)} - u^{(b)}$  characterizing the extent to which our model prediction is imperfect. The observations of our system

are denoted by  $w_k = w(t_k)$ . These observations do not necessarily need to be components of the state vector  $u$ , but rather, are related to it via the *observation function*,  $h : u_k \mapsto w_k$ . For example, one may attempt to predict surface sea temperatures using data assimilation with data from satellite observations. The function  $h$  would then be the Stefan-Boltzmann relating the measured spectra to temperature. However, real world data is *also* imperfect, which we must take into account let we over constrain our model with poor quality data.

We write

$$w_k = h(u_k) + \xi_k^{(m)} \quad (3.203)$$

where  $\xi_k^{(m)}$  denotes this measurement error.

Given our model predictions  $u_k^{(b)}$  and observations  $w_k$ , we seek to obtain the *optimal* or best-possible prediction called the **analysis**,  $u^{(a)}$ . Despite our care, this analysis will still be imperfect, so we further define the analysis error as

$$\xi^{(a)} = u^{(t)} - u^{(a)} \quad (3.204)$$

In summary, we have defined the following relevant quantities (at time  $t_k$ ):

$$u_k^{(t)} \in \mathbb{R}^n \quad \text{the true state vector} \quad (3.205)$$

$$u_k^{(b)} \in \mathbb{R}^n \quad \text{the model forecast} \quad (3.206)$$

$$u_k^{(a)} \in \mathbb{R}^n \quad \text{the analysis} \quad (3.207)$$

$$w_k \in \mathbb{R}^m \quad \text{the observation vector} \quad (3.208)$$

$$\xi^{(b)} \in \mathbb{R}^n \quad \text{the model forecast error} \quad (3.209)$$

$$\xi^{(m)} \in \mathbb{R}^m \quad \text{the observation noise vector} \quad (3.210)$$

$$\xi^{(a)} \in \mathbb{R}^n \quad \text{the analysis error} \quad (3.211)$$

$$\xi^{(p)} \in \mathbb{R}^n \quad \text{the process noise if we used our model on the true state} \quad (3.212)$$

$$\mathcal{M} : \mathbb{R}^n \rightarrow \mathbb{R}^n \quad \text{the discrete model evolution function} \quad (3.213)$$

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^n \quad \text{differential equation model (the right hand side)} \quad (3.214)$$

$$h : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad \text{observation function} \quad (3.215)$$

To move forward, we now establish the following (prior) assumptions about the distribution of errors in each component in order to make it possible to derive a closed form for the final analysis. We require

$$\mathbb{E}[\xi_k^{(b)}] = 0 \quad \mathbb{E}[\xi_k^{(b)}(\xi_j^{(b)})^T] = 0 \text{ for } k \neq j \quad (3.216)$$

$$\mathbb{E}[\xi_k^{(m)}] = 0 \quad \mathbb{E}[\xi_k^{(m)}(\xi_j^{(m)})^T] = 0 \text{ for } k \neq j \quad (3.217)$$

$$\mathbb{E}[\xi_k^{(b)}(u_0)^T] = 0 \quad \mathbb{E}[\xi_k^{(m)}(u_0)^T] = 0 \quad (3.218)$$

$$\mathbb{E}[\xi_k^{(b)}\xi_j^{(m)}] = 0 \quad (3.219)$$

or in words, the errors in our model and observation are unbiased (e.g. mean zero) and uncorrelated.

We also define the error covariance matrices

$$Q_k := \mathbb{E}[\xi_k^{(p)}(\xi_k^{(p)})^T] \quad (3.220)$$

$$R_k := \mathbb{E}[\xi_k^{(m)}(\xi_k^{(m)})^T] \quad (3.221)$$

$$B_k := \mathbb{E}[\xi_k^{(b)}(\xi_k^{(b)})^T] \quad (3.222)$$

which we will use in our consideration of the final error of our analysis.

### 3.5.1 Kalman Filter

Given some model for the error covariance matrices  $Q_k$  and  $R_k$ , we would like a method that propagates *both* our model *and* the errors forward. This way we may guarantee that the accuracy of our analysis doesn't come at the cost of higher uncertainty.

The original implementation of the Kalman filter was for strictly linear systems, in particular as a filter for signal processing. We will first develop the analysis for this simplified case and then will generalize to the *Extended Kalman Filter* (EKF) that can handle fully nonlinear situations.

In the linear case, our system may be written as

$$u_{k+1}^{(t)} = M_k u_k^{(t)} + \xi_{k+1}^{(p)} \quad (3.223)$$

$$w_k = H_k u_k^{(t)} + \xi_k^{(m)} \quad (3.224)$$

where  $M_k$  and  $H_k$  are now matrices defining the linear problem.

The goal of the Kalman filter is to derive the analysis  $u^{(a)}$  which minimizes the trace of the analysis error covariance matrix (i.e. sum of squared errors):

$$\text{Tr}(P_k) := \mathbb{E}[(u_k^{(t)} - u_k^{(a)})^T(u_k^{(t)} - u_k^{(a)})] \quad (3.225)$$

Finding the analysis consists of two steps: the forecast step and the assimilation step.

Beginning with the forecasting step, assume we have the analysis at time  $t_k$  denoted  $u_k^{(a)}$ .

Then the forecast for time  $t_{k+1}$  is

$$u_{k+1}^{(b)} = M_k u_k^{(a)}, \quad (3.226)$$

or in words, we obtain the next prediction by evolving the previous analysis forward. The background error is therefore

$$\xi_{k+1}^{(b)} = u_{k+1}^{(t)} - u_{k+1}^{(b)} \quad (3.227)$$

$$= M_k u_k^{(t)} + \xi_{k+1}^{(p)} - M_k u_k^{(a)} \quad (3.228)$$

$$= M_k \left( u_k^{(t)} - u_k^{(a)} \right) + \xi_{k+1}^{(p)} \quad (3.229)$$

$$= M_k \xi_k^{(a)} + \xi_{k+1}^{(p)} \quad (3.230)$$

We may now evaluate the covariance matrix of our background estimate as:

$$B_{k+1} = \mathbb{E}[\xi_{k+1}^{(b)} (\xi_{k+1}^{(b)})^T] \quad (3.231)$$

$$= \mathbb{E} \left[ \left( M_k \xi_k^{(a)} + \xi_{k+1}^{(p)} \right) \left( M_k \xi_k^{(a)} + \xi_{k+1}^{(p)} \right)^T \right] \quad (3.232)$$

$$(3.233)$$

If we presume that  $\mathbb{E}[\xi_k^{(b)} (\xi_{k+1}^{(p)})^T] = 0$ , then the cross terms vanish and we are left with

$B_{k+1} = M_k P_k M_k^T + Q_{k+1}$

$$(3.234)$$

Thus we now have the background (i.e forecast) estimate of the state at  $t_{k+1}$  and its covariance matrix. Given a measurement  $w_{k+1}$  at the same time with covariance matrix  $R_{k+1}$ , we may now perform the assimilation step where we fuse the two sources of information to obtain  $u_{k+1}^{(a)}$  and  $P_{k+1}$ .

Let's suppose that the analysis has the form

$$u_{k+1}^{(a)} = \nu + K_{k+1} w_{k+1} \quad (3.235)$$

for some vector  $\nu \in \mathbb{R}^n$  and matrix  $K_{k+1} \in \mathbb{R}^{m \times n}$ . In a perfect world, we would have

$$\mathbb{E}[u_k^{(t)} - u_k^{(a)}] = 0. \text{ Therefore,}$$

$$0 = \mathbb{E}[u_k^{(t)} - u_k^{(a)}] \quad (3.236)$$

$$= \mathbb{E}[(u_k^{(b)} + \xi_k^{(b)}) - (\nu + K_k w_k)] \quad (3.237)$$

$$= \mathbb{E}[(u_k^{(b)} + \xi_k^{(b)}) - (\nu + K_k H_k u_k^{(t)} + K_k \xi_k^{(m)})] \quad (3.238)$$

$$= \mathbb{E}[u_k^{(b)}] + \mathbb{E}[\xi_k^{(b)}] - \mathbb{E}[\nu] - K_k H_k \mathbb{E}[u_k^{(t)}] - K_k \mathbb{E}[\xi_k^{(m)}] \quad (3.239)$$

$$= u_k^{(b)} + 0 - \nu - K_k H_k u_k^{(b)} - 0 \quad (3.240)$$

$$= u_k^{(b)} - \nu - K_k H_k u_k^{(b)} \quad (3.241)$$

$$\Rightarrow \nu = u_k^{(b)} - K_k H_k u_k^{(b)} \quad (3.242)$$

which we now substitute to obtain

$$\boxed{u_k^{(a)} = u_k^{(b)} + K_k(w_k - H_k u_k^{(b)})} \quad (3.243)$$

Now that we know the form for the analysis we may derive the optimal matrix  $K_k$  by optimization of  $P_k$ . We have

$$\xi_k^{(a)} = u_k^{(t)} - u_k^{(a)} \quad (3.244)$$

$$= M_{k-1}u_{k-1}^{(t)} + \xi_k^{(p)} - u_k^{(b)} - K_k \left( w_k - H_k u_k^{(b)} \right) \quad (3.245)$$

$$= M_{k-1}u_{k-1}^{(t)} + \xi_k^{(p)} - M_{k-1}u_{k-1}^{(a)} - K_k \left( H_k u_k^{(t)} + \xi_k^{(m)} - H_k u_k^{(b)} \right) \quad (3.246)$$

$$= M_{k-1}u_{k-1}^{(t)} + \xi_k^{(p)} - M_{k-1}u_{k-1}^{(a)} - K_k H_k u_k^{(t)} - K_k \xi_k^{(m)} + K_k H_k u_k^{(b)} \quad (3.247)$$

$$= M_{k-1}u_{k-1}^{(t)} + \xi_k^{(p)} - M_{k-1}u_{k-1}^{(a)} - K_k H_k u_k^{(t)} - K_k \xi_k^{(m)} + K_k H_k u_k^{(b)} \quad (3.248)$$

$$= \left\{ M_{k-1}(\xi_{k-1}^{(a)} + u_{k-1}^{(a)}) + \xi_k^{(p)} - M_{k-1}u_{k-1}^{(a)} \right\} - K_k H_k u_k^{(t)} - K_k \xi_k^{(m)} + K_k H_k u_k^{(b)} \quad (3.249)$$

$$= \left\{ M_{k-1}\xi_{k-1}^{(a)} + \xi_k^{(p)} \right\} - K_k H_k u_k^{(t)} + K_k H_k u_k^{(b)} - K_k \xi_k^{(m)} \quad (3.250)$$

$$= M_{k-1}\xi_{k-1}^{(a)} + \xi_k^{(p)} - K_k H_k(M_{k-1}u_{k-1}^{(t)} + \xi_k^{(b)}) + K_k H_k u_k^{(b)} - K_k \xi_k^{(m)} \quad (3.251)$$

$$= M_{k-1}\xi_{k-1}^{(a)} + \xi_k^{(p)} - K_k H_k M_{k-1}(\xi_{k-1}^{(a)} + u_{k-1}^a) - K_k H_k \xi_k^{(b)} + K_k H_k u_k^{(b)} - K_k \xi_k^{(m)} \quad (3.252)$$

$$= M_{k-1}\xi_{k-1}^{(a)} + \xi_k^{(p)} - K_k H_k M_{k-1}(\xi_{k-1}^{(a)} + u_{k-1}^a) - K_k H_k \xi_k^{(b)} + K_k H_k M_{k-1}u_{k-1}^{(a)} - K_k \xi_k^{(m)} \quad (3.253)$$

$$= M_{k-1}\xi_{k-1}^{(a)} + \xi_k^{(p)} - K_k H_k M_{k-1}\xi_{k-1}^{(a)} - K_k H_k \xi_k^{(b)} - K_k \xi_k^{(m)} \quad (3.254)$$

$$= (I - K_k H_k)(M_{k-1}\xi_{k-1}^{(a)} - \xi_k^p) - K_k \xi_k^{(m)} \quad (3.255)$$

$$(3.256)$$

and therefore the covariance matrix is

$$P_k = \mathbb{E}[\xi_k^{(a)}(\xi_k^{(a)})^T] \quad (3.257)$$

$$\begin{aligned} &= \mathbb{E}\left[\left((I - K_k H_k)(M_{k-1}\xi_{k-1}^{(a)} - \xi_k^p) - K_k \xi_k^{(m)}\right)\left((I - K_k H_k)(M_{k-1}\xi_{k-1}^{(a)} - \xi_k^p) - K_k \xi_k^{(m)}\right)^T\right] \\ &\quad (3.258) \end{aligned}$$

$$= (I - K_k H_k) M_{k-1} \mathbb{E}[\xi_{k-1}^{(a)}(\xi_{k-1}^{(a)})^T] M_{k-1}^T (I - K_k H_k)^T \quad (3.259)$$

$$\begin{aligned} &\quad + (I - K_k H_k) \mathbb{E}[\xi_k^{(p)}(\xi_k^{(p)})^T] (I - K_k H_k)^T \\ &\quad - K_k \mathbb{E}[\xi_k^{(m)}(\xi_k^{(m)})^T] K_k^T \\ &= (I - K_k H_k) B_k (I - K_k H_k)^T - K_k R_k K_k^T \quad (3.260) \end{aligned}$$

Now all that remains is to derive the specific form for  $K_k$ . The Kalman filter is defined as that  $K_k$  which minimizes the sum of squared analysis errors, i.e. the trace of the analysis error covariance matrix. Therefore, the following identities from matrix calculus will be useful:

$$\nabla_A \text{tr}(AB) = B^T \quad (3.261)$$

$$\nabla_A \text{tr}(BA^T) = B \quad (3.262)$$

$$\nabla_A \text{tr}(ABA^T) = AB^T + AB \quad (3.263)$$

$$(3.264)$$

from which we obtain

$$0 = \nabla_{K_k} \text{tr}(P_k) \quad (3.265)$$

$$= \nabla_{K_k} \left\{ B_k - B_k H_k^T K_k^T - K_k H_k B_k + K_k H_k B_k H_k^T B_k^T - K_k R_k K_k^T \right\} \quad (3.266)$$

$$= -B_k H_k^T - (H_k B_k)^T + K_k [H_k B_k H_k^T + (H_k B_k H_k^T)^T - R_k + R_k^T] \quad (3.267)$$

$$= -2B_k H_k^T + 2K_k (H_k B_k H_k^T - R_k) \quad (3.268)$$

$$\Rightarrow K_k = B_k H_k^T [H_k B_k H_k^T - R_k]^{-1} \quad (3.269)$$

we now substitute this result to obtain a simplified form for  $P_k$ :

$$P_k = (I - K_k H_k) B_k (I - K_k H_k)^T + K_k R_k K_k^T \quad (3.270)$$

$$= (I - K_k H_k) B_k - (I - K_k H_k) B_k (K_k H_k)^T + K_k R_k K_k^T \quad (3.271)$$

$$= (I - K_k H_k) B_k - \left\{ (I - K_k H_k) B_k (K_k H_k)^T + K_k R_k K_k^T \right\} \quad (3.272)$$

$$= (I - K_k H_k) B_k - \left\{ (I - K_k H_k) B_k H_k^T K_k^T + K_k R_k K_k^T \right\} \quad (3.273)$$

$$= (I - K_k H_k) B_k - \left\{ (I - K_k H_k) B_k H_k^T + K_k R_k \right\} K_k^T \quad (3.274)$$

$$= (I - K_k H_k) B_k - \left\{ B_k H_k^T - K_k (H_k B_k H_k^T + R_k) \right\} K_k^T \quad (3.275)$$

$$= (I - K_k H_k) B_k - \left\{ B_k H_k^T - B_k H_k^T \right\} K_k^T \quad (3.276)$$

$$= (I - K_k H_k) B_k \quad (3.277)$$

where we have used the fact that covariance matrices are symmetric.

Now that we have all of the pieces, let's summarize the procedure:

1. **Initialization:** We must set the system to some initial condition. This means we must define  $u_0^{(a)}$  and  $P_0$ . We must also come up with a model for the process noise covariance  $Q_k$  and measurement error covariance  $R_k$ .
2. **Forecast Step:** Starting with the analysis state vector  $u_{k-1}^{(a)}$  and covariance  $P_{k-1}$ , we apply the model time evolution operator  $M_{k-1}$  to obtain the predicted state  $u_k^{(b)}$  and covariance  $B_k$  at the next time step.

$$u_k^{(b)} = M_{k-1} u_{k-1}^{(a)} \quad (3.278)$$

$$B_k = M_{k-1} P_{k-1} M_{k-1}^T + Q_k \quad (3.279)$$

3. **Assimilation Step:** Combine our observations  $w_k$  and their associated uncertainties contained in  $R_k$  together with the background prediction  $u_k^{(b)}$  and its error covariance

matrix  $B_k$  to obtain the analysis state  $u_k^{(a)}$  and error covariance  $P_k$ .

$$K_k = B_k H_k^T \left[ H_k B_k H_k^T - R_k \right]^{-1} \quad (3.280)$$

$$u_k^{(a)} = u_k^{(b)} + K_k(w_k - H_k u_k^{(b)}) \quad (3.281)$$

$$P_k = (I - K_k H_k) B_k \quad (3.282)$$

### 3.5.2 Extended Kalman Filter

Given the nonlinear nature of many scientific models, it is desirable to extend our notion of the *Kalman Filter* to be able to handle nonlinear models  $f(\cdot)$  (and by extension, their update function  $\mathcal{M}(\cdot)$ ), and nonlinear observation functions  $h(\cdot)$ . This can be accomplished so long as these functions are sufficiently smooth ( $C^1$  to be precise) so as to admit valid Taylor approximations to first order. That is,

$$\mathcal{M}(u_k) \approx \mathcal{M}(u_k^{(a)}) + D_M(u_k^{(a)})\xi_k^{(a)} \quad h(u_k) \approx h(u_k^{(b)}) + D_h(u_k^{(b)})\xi_k^{(b)} \quad (3.283)$$

$$D_M := \left[ \frac{\partial \mathcal{M}_i}{\partial u_j} \right] \quad D_h := \left[ \frac{\partial h_i}{\partial u_j} \right] \quad (3.284)$$

where  $\mathcal{M}_i$  and  $h_i$  denote the  $i$ th component functions of  $\mathcal{M}$  and  $h$  respectively. For the so-called *Extended Kalman Filter* (EKF), approach. We assume that the analysis errors can be reasonably approximated to evolve linearly such that we can use the above *linearized* approximations to our model update and observation functions to reasonably model our uncertainties.

We therefore replace the (previously) linear functions  $M_k$  and  $H_k$  with their nonlinear Jacobian derived counterparts to obtain the following procedure

1. **Initialization:** To begin we must choose values for  $u_0^{(a)}$  and  $P_0$ . We must also provide models for  $Q_k$  and  $R_k$ .

2. **Forecast Step:**

$$u_k^{(b)} = \mathcal{M}(u_{k-1}^{(a)}) \quad (3.285)$$

$$B_k = D_M(u_{k-1}^{(a)}) P_{k-1} D_M^T(u_{k-1}^{(a)}) + Q_k \quad (3.286)$$

### 3. Assimilation Step:

$$K_k = B_k D_h^T(u_k^{(b)}) \left[ D_h(u_k^{(b)}) B_k D_h^T(u_k^{(b)}) + R_k \right]^{-1} \quad (3.287)$$

$$u_k^{(a)} = u_k^{(b)} + K_k(w_k - h(u_k^{(b)})) \quad (3.288)$$

$$P_k = \left( I - K_k D_h(u_k^{(b)}) \right) B_k \quad (3.289)$$

Note that the most challenging piece in this implementation is to stably compute the model update Jacobian  $D_M$ , as this is not just the Jacobian of our continuous model  $f(\cdot)$ . For example, if we are modeling the state of the system via a set of ODEs, then  $D_M$  is the Jacobian of the output of a single integrator step with respect to the initial starting state. This can pose significant computational challenges where adaptive solvers are desired as we must be able to back-propagate out derivative information for all evaluations of the function  $f(\cdot)$  taken during the step.

#### 3.5.3 continuous-discrete Extended Kalman Filter

So far in our discussion of data assimilation techniques we have limited ourselves to discretized realizations of continuous models and measurements. This allowed us to derive reasonable update equations for the error covariance matrix and analysis by considering the changes that occur by application of our model evolution operator  $\mathcal{M}$ . In practice though the case tends to be slightly different: the underlying dynamics are understood to evolve continuously despite our access to finite time measurements at some frequency. For this reason, it would be preferable to utilize our knowledge of the model covariance  $Q$  together with the continuous dynamics of the model itself to allow the background covariance matrix to evolve continuously between measurements. Then, only when we have access to new data do we need to perform a discrete update by computing the Kalman gain and generating a new analysis as before. As we shall see, this strategy yields an augmented dynamical system

for which

$$\begin{cases} \frac{du}{dt} = f(u, t; \theta) + \xi^{(b)}(t) \\ \frac{P}{dt} = J_u(t)P(t) + P(t)J_u^T(t) + Q(t) \end{cases} \quad (3.290)$$

where  $J_u = \frac{\partial f}{\partial u}$  is the Jacobian of the system with respect to the state vector  $u$ . To derive this new form for the Extended Kalman filter, let us first begin with a linear system for which *both* the underlying dynamics continuously, that is

$$\begin{cases} \dot{u} = Mu + \xi^{(b)} \\ w_k = Hu_k + \xi_k^{(m)} \end{cases} \quad (3.291)$$

where  $\xi^{(b)}$  now represents a Weiner process (e.g. Gaussian white noise process) of covariance  $Q(t)$ . Presuming the model covariance is still uncorrelated with the measurement means that we now have:

$$\mathbb{E}[\xi^{(b)}(t)(\xi^{(b)}(\tau))^T] = Q(t)\delta(t - \tau) \quad (3.292)$$

$$\mathbb{E}[\xi^{(m)}(\xi^{(b)}(t))^T] = 0 \quad (3.293)$$

To move forward, we now need to establish the relationship between our previous  $Q_k$  and  $Q$  so that in the limit as  $\Delta t \rightarrow 0$ , we may obtain the new dynamical equations. If we understand  $Q_k$  to be the model error *accumulated* over the integration window, then it follows that

$$\begin{aligned} Q_k &\approx \int \int d\zeta d\eta \mathbb{E}[\xi^{(b)}(\zeta)(\xi^{(b)}(\eta))^T] \\ &= \int \int d\zeta d\eta Q(\zeta)\delta(\zeta - \eta) \\ &= \int_{t_{k-1}}^{t_k} Q(\zeta)d\zeta \\ &\approx Q(t_k)\Delta t \end{aligned} \quad (3.294)$$

This tells us all that we need! For small enough  $\Delta t$ , we may approximate our continuous dynamics by

$$\frac{u_{k+1} - u_k}{\Delta t} = Mu_k + \xi_k^{(b)} \quad (3.295)$$

and therefore

$$u_{k+1} = (I - \Delta t M) u_k + \xi_k^{(b)} \quad (3.296)$$

from which we can identify our linear  $M_k$  matrix from the original discrete Kalman filter as  $(I + \Delta t M)$  in the discretized version of this new continuous model. Substituting this in to our expression for the forecast step yields

$$B_{k+1} = (I + \Delta t M) P_k (I + \Delta t M)^T + Q \Delta t. \quad (3.297)$$

Expanding this expression and keeping only terms of  $\mathcal{O}(\Delta t)$ , we find

$$\begin{aligned} B_{k+1} &= P_k + \Delta t M P_k + P_k M^T \Delta t + M P_k M^T \Delta t^2 + Q \Delta t \\ &\approx P_k + \Delta t M P_k + \Delta t P_k M^T + Q \Delta t \end{aligned} \quad (3.298)$$

As we shrink  $\Delta t$  to zero between observations, there is no longer a distinction between the background covariance  $B_k$  and the analysis covariance  $P_k$  so that we may write rearrange and take the limit

$$\lim_{\Delta t \rightarrow 0} \frac{P_{k+1} - P_k}{\Delta t} = \lim_{\Delta t \rightarrow 0} M P_k + P_k M^T + Q \quad (3.299)$$

which evaluates to

$$\frac{d}{dt} P = M P(t) + P(t) M^T + Q(t)$$

(3.300)

To extend this linear version to the case of nonlinear model and observation functions, we linearize our nonlinear model so that  $M \mapsto J_u = \partial f / \partial u$ . This then leads to the continuous-discrete Extended Kalman Filter for which the procedure is as follows

1. **Initialization:** To begin we choose initial values for  $u_0^{(a)}$  and  $P_0$ . We must also provide initial covariance values  $R_0$  and  $Q(0)$ .
2. **Forecast Step:** Integrate the dynamical system

$$\begin{cases} \frac{du}{dt} = f(u, t; \theta) \\ \frac{dP}{dt} = J_u P(t) + P(t) J_u^T + Q(t) \end{cases} \quad (3.301)$$

from  $t_{k-1}$  to  $t_k$  to obtain  $u_k$  and  $P_k$

**3. Assimilation Step:** Compute the Kalman gain and update the state vector and associated covariance matrix according to

$$K_k = P_k D_h^T(u_k) [D_h(u_k) P_k D_h^T(u_k) + R_k]^{-1} \quad (3.302)$$

$$u_k \mapsto u_k + K_k(w_k - h(u_k)) \quad (3.303)$$

$$P_k \mapsto (I - K_k D_h(u_k)) P_k \quad (3.304)$$

### 3.5.4 3d-Var

For the *Kalman Filter* and the *EKF*, we derived an optimal way to combine observation with simulation so as to minimize the trace of the analysis error covariance matrix,  $P_k$ . An alternative approach is to recast the problem as a pure optimization problem where rather than finding a filter  $K_k$  that will add an innovation to  $u_k^{(b)}$  to obtain the analysis  $u_k^{(a)}$ , we obtain the analysis by directly optimizing a cost function

$$J(u) = \frac{1}{2} (w - h(u))^T R^{-1} (w - h(u)) + \frac{1}{2} (u - u^{(b)})^T B^{-1} \frac{1}{2} (u - u^{(b)}) \quad (3.305)$$

which we can justify as coming from the joint probability distribution (assuming Gaussian errors)

$$\mathcal{P}(u|w) = C \exp \left( -\frac{1}{2} (u - u^{(b)})^T B^{-1} \frac{1}{2} (u - u^{(b)}) \right) \cdot \exp \left( -\frac{1}{2} (w - h(u))^T R^{-1} (w - h(u)) \right) \quad (3.306)$$

with model error covariance  $B$  and measurement error covariance  $R$  as before. This is clearly a *very strong assumption*.

To optimize  $J(u)$ , we begin by taking it's gradient.

$$\nabla_u J(u) = -D_h^T R^{-1} (w - h(u)) + B^{-1} (u - u^{(b)}) \quad (3.307)$$

Thus, finding the analysis  $u^{(a)}$  ammounts to solving the system

$$a - D_h^T R^{-1} (w - h(u^{(a)})) + B^{-1} (u^{(a)} - u^{(b)}) = 0 \quad (3.308)$$

As for Kalman filtering, let's begin with the assumption that our model and observation function are linear. Suppose that we have  $h(u) = Hu$  so that  $D_h(u) = H$ . It follows that

$$D_h^T R^{-1} (w - Hu^{(a)}) = B^{-1} (u^{(a)} - u^{(b)}) \quad (3.309)$$

$$D_h^T R^{-1} w - D_h^T R^{-1} Hu^{(a)} = B^{-1} u^{(a)} - B^{-1} u^{(b)} \quad (3.310)$$

$$(D_h^T R^{-1} H + B^{-1}) u^{(a)} = D_h^T R^{-1} + B^{-1} u^{(b)} \quad (3.311)$$

$$(H^T R^{-1} H + B^{-1}) u^{(a)} = H^T R^{-1} + B^{-1} u^{(b)} \quad (3.312)$$

Thus we see that the analysis is given by

$$u^{(a)} = u^{(b)} + B H^T (R + H B^T H)^{-1} (w - H u^{(b)}) \quad (3.313)$$

which agrees with what we found for the Linear Kalman Filter.

Moving on to the non-linear case, we can expand  $h$  about an initial guess  $u^{(c)}$  which we will later choose to be  $u^{(b)}$  for convenience.

$$h(u^{(a)}) \approx h(u^{(c)}) + D_h(u^{(c)}) \Delta u \quad (3.314)$$

Using this, we have

$$D_h^T(u^{(a)}) R^{-1} (w - h(u^{(a)})) = B^{-1} (u^{(a)} - u^{(b)}) \quad (3.315)$$

$$D_h^T(u^{(a)}) R^{-1} (w - h(u^{(c)}) - D_h(u^{(c)}) \Delta u) \approx B^{-1} (u^{(c)} + \Delta u - u^{(b)}) \quad (3.316)$$

$$D_h^T(u^{(c)}) R^{-1} (w - h(u^{(c)}) - D_h(u^{(c)}) \Delta u) \approx B^{-1} (u^{(c)} + \Delta u - u^{(b)}) \quad (3.317)$$

which we now solve for the update  $\Delta u$  to obtain the linear system

$$(B^{-1} + D_h^T(u^{(c)}) R^{-1} D_h(u^{(c)})) \Delta u = B^{-1} (u^{(b)} - u^{(c)}) + D_h^T(u^{(c)}) R^{-1} (w - h(u^{(c)})) \quad (3.318)$$

Thus we have the following prescription

1. To begin, take  $u^{(c)} == u^{(b)}$ .

2. Solve the system

$$(B^{-1} + D_h^T(u^{(c)})R^{-1}D_h(u^{(c)})) \Delta u = B^{-1}(u^{(b)} - u^{(c)}) + D_h^T(u^{(c)})R^{-1}(w - h(u^{(c)})) \quad (3.319)$$

to obtain  $\Delta u$

3. Update your guess using your favorite optimization algorithm. For example, in steppest descent, choose a learning rate  $\eta$  and set

$$u_{\text{new}}^{(c)} = u_{\text{prev}}^{(c)} - \eta \Delta u \quad (3.320)$$

4. Repeat the procedure until  $|u_{\text{new}}^{(c)} - u_{\text{prev}}^{(c)}|$  converges to a desired tolerance.

In both the linear and nonlinear case, it should be noted that we have not added time indices to our state vectors. This is an indication that the 3d-var procedure is performed *at every time where you have observation data*.

### 3.5.5 4d-Var

The 3d-Var algorithm attempts to optimize a cost function point-by-point to obtain the ideal analysis at each time where we have observation data. This can become computationally expensive as we require model evaluations *and* an optimization routine for every observation point. An alternative approach is to simultaneously optimize across all observations in order to obtain the ideal *initial condition* which achieves the best model fit. In other words, this amounts to parameter fitting for our model where we think of the initial conditions as the parameters. The difference is two-fold: We extend the usual  $\ell_2$  function into a quadratic form utilizing the observation error covariance matrix  $R_k$ . We also typically only have a small number of observables related to the state vector (potentially non-linearly) and therefore

must transform the state vector  $u_k$  into the observation space via  $h$  before computing the loss against our observations  $w_k$ . This results in a loss function of the form

$$\begin{aligned} J(u_0) &= \frac{1}{2} \left( u_0 - u_0^{(b)} \right)^T B^{-1} \left( u_0 - u_0^{(b)} \right) + \frac{1}{2} \sum_k (w_k - h(u_k))^T R_k^{-1} (w_k - h(u_k)) \quad (3.321) \\ &= J_b(u_0) + J_m(u_0). \end{aligned}$$

Note that the first term is useful if we already have an initial guess  $u_0^{(b)}$  for the initial condition in mind. If we do not have one, we may omit this term.

As before, we now want to optimize this cost function. To do so, we first observe that

$$u_k = \mathcal{M}^{(k)}(u_0; \theta) \quad (3.322)$$

It is easy to obtain the gradient of  $J_b$  so we shall focus on the second term. We find that

$$\nabla_{u_0} J_m = \nabla_{u_0} \left\{ \sum_k \frac{1}{2} (w_k - h(u_k))^T R_k^{-1} (w_k - h(u_k)) \right\} \quad (3.323)$$

$$= - \sum_k \left[ \frac{\partial}{\partial u_0} h(\mathcal{M}^{(k-1)}(u_0)) \right]^T R_k^{-1} (w_k - h(u_k)) \quad (3.324)$$

$$= - \sum_k [D_h(u_k) D_M(u_{k-1}) D_M(u_{k-2}) \cdots D_M(u_0)]^T R_k^{-1} (w_k - h(u_k)) \quad (3.325)$$

$$= - \sum_k [D_M^T(u_0) D_M^T(u_1) \cdots D_M^T(u_{k-1}) D_h^T(u_k)] R_k^{-1} (w_k - h(u_k)) \quad (3.326)$$

With this in hand, the procedure is nearly identical to 3d-var:

1. Integrate your model forward to obtain  $\{u_k\}$
2. Evaluate each of the  $D_M^T(u_{k-1:0})$  and  $D_h(u_k)$ .
3. Using these values, compute  $\nabla J_m(u)$
4. Set  $u_0^{(new)} = u_0^{(prev)} - \eta \nabla J(u_0^{(prev)})$
5. Stop when  $|u_0^{(new)} - u_0^{(prev)}|$  converges to your desired tolerance.

You can of course substitute another optimization scheme after step 3.

There are a few further observations that we can make about this procedure. The first, is that should we feel our model  $f(\cdot)$  is not sufficient to fit the entire time series without the periodic corrections of the Kalman filter (perhaps there is some missing physics we haven't captured in our model), then it may be ideal to restrict the optimization to a subset of the full time series of observations. The second observation is that the above formulation has assumed the adjoint  $D_M^T$  can be computed easily. We can take advantage of the methods from the *Adjoint Sensitivity Analysis* of ODEs to take advantage of modern methods for automatic differentiation of ODE solutions. Finally, it should be noted that the *4d-var* algorithm yields the optimal starting conditions assuming our model is perfect. At the sacrifice of smoothness, it is often desirable to first obtain the ideal initial condition with *4d-var* and *then* employ a filtering method like the EKF to obtain the final analysis.

# CHAPTER 4

## AN AUTONOMOUS ROBOTIC TEAM FOR THE RAPID CHARACTERIZATION OF NOVEL ENVIRONMENTS

### 4.1 Rapid Georectification and Processing of Pushbroom Hyperspectral Imagery

Recent developments in hyperspectral imaging technology have led to dramatic reductions in both size and weight of imaging platforms. Due to these improvements, it is now possible to incorporate the technology as the payload of highly mobile autonomous aerial vehicles such as drones. However, the massive volume of hyperperspectral datacubes poses significant computational challenges to their adoption in real-time applications. In this paper, we demonstrate a procedure for the rapid georeferencing of pushbroom imagery and demonstrate its application for real-time remote sensing... Results indicate that a typical image can be processed in 10 seconds while it takes  $x$  seconds to capture.

For decades, multi-spectral imagers\* have seen wide spread adoption in the remote sensing community as a means to take advantage of the wealth of information contained in the reflectance spectra of materials. In addition to the three color filters of traditional cameras, \*multi-spectral\* imagers, like those deployed on MODIS, Sentinel 2, and other satellite missions, capture many additional features by utilizing wavelength bands ranging from the near-UV, through the visible spectrum, and into the Infrared. With this additional information, multi-spectral remote sensing platforms are able to aid in a variety of domains from tracking land change, characterizing deforestation, monitoring erosion, evaluating crop health, and many others (add references here).

(\*\*add a figure showing the different wavelength bands of publically available satellites\*\*)

These uses are justified by the reflectance features of materials across the electromagnetic spectrum; water has vibrational modes in the IR, pigments have absorption peaks in the

visible, etc... (discuss the particular features present in different regions of the reflectance spectrum). Many currently used spectral indices like the \*normalized difference vegetation index\* (NDVI) take advantage of these spectral regions by comparing ratios of pigment sensitive passbands to the constant signals infrared to infer the abundance of chlorophyll, and consequently, the health of plants, to identify forest canopy, etc.. (\*\*add citations and expand\*\*)

However, despite the plethora of successful applications of multi-spectral imaging, more can be accomplished with the additional information provided by fully resolved spectra. For example, in the laboratory, spectrophotometry allows the direct determination of the concentrations of chemicals constituents in solution (\*\*add more detail and references\*\*) by deconvolution of a sample spectrum against libraries carefully collected of reference spectra: individual chemicals be uniquely identified by the characteristic location and shape of their absorption features. We should mention something about algal blooms—harmful species display shifted reflectance peaks that can be used to identify toxic species. This information is effectively filtered away by the broad passbands of existing (find the spectra used in that book that show the different species of algae). To that end, \*hyperspectral imagers\* (HSI), which sample hundreds of wavelength bands at each pixel, have become the natural next step for remote sensing platforms with many planned to be deployed in the coming years (\*\*add references about soon-to-be-deployed HSI satellites\*\*).

At the terrestrial level, drones (commonly, quadcopters, octocopters, and other similar multi-rotor craft) equipped with cameras are able to utilize techniques of photogrammetry together with continuously sampled imagery to produce high quality digital elevation maps, high quality mosaics, 3-dimensional reconstructions, etc... These capabilities provide significant aid for structural analysis, smart agriculture, etc... Today, kilogram-scale HSI can be comfortably mounted to the payload of film-scale drones such as the AltaX and advancements in spectral sensing such as (\*\*reference Ethan Minot's recent paper\*\*) suggest that sizes of HSI will continue to shrink further expanding their application in this domain.

The increased spectral resolution of HSI systems poses unique challenges to their adoption for real time applications primarily stemming from the considerable size of generated data files. Current data collection workflows see researchers first perform the aerial survey (data collection) and then transfer data to ground based computers for post processing. This workflow is well established in the remote sensing community where, as an example, compressed raw imagery from Sentinel-2 are transferred to the ground and then subsequently post processed into their final L1C (top of atmosphere) and L2A (bottom of atmosphere) data products (\*\*add citation here\*\*). Drone based applications often operate in a similar manner: images or video are collected by a survey and then post-processed and analyzed with software such as Open Drone Map to produce the desired data products (tile mosaics, 3d reconstructions, etc.) (\*\*add citation\*\*). For an HSI platform to function in real time, three key computational tasks are critical: 1. **FileIO**: captured imagery need to be quickly read by the on-board processing computer 2. **Post-processing**: Raw imagery need to be rapidly converted to the chosen data product (typically, Reflectance), and importantly, must be georeferenced so that each image pixel can be located on the ground. 3. **Ground Transfer**: Sufficient wireless communication must be available to transmit the final

The first can be readily accomplished by means of light weight, high volume solid state drives. To address the second, we need both sufficient compute and optimized processing software. Finally, ground transfer of final post-processed data products can be accomplished in a variety of ways. As we rarely need the full hyper-spectral datacube right away, once can generate the desired data products on board (NDVI for example) and transfer only the relevant information to a ground station

We should now make the argument about the need for improved methods for processing of HSI due the their dramatically larger size, i.e. the standard workflow of collecting imagery and then post processing (re: Open Drone Map) is a fantastic and well tested solution

but prevents the use in time-critical applications. We can also reference the continued improvements to single-board computers such as the raspberry pi 4b (8 Gb of Ram), Jetson family (GPU equipped), and intel nuc (powerful work horses with minuscule form factors). To make real-time possible we need two things: 1. Conversion from raw data (digital counts) to physical units like Reflectance 2. Georeferencing of captured imagery so features can be quickly geolocated

We can always re-process the imagery later for an in-depth reanalysis, but for time-critical applications, we need all of the processing to happen on board the drone.

Add a paragraph about georeferencing in particular: - Georectification = georeference + orthorectification - outline other papers that have discussed georectification techniques: - ground control points (not reasonable for dangerous or water-based environments) - IMU / GPS - different type of sensors: square, pushbroom, whiskbroom

Discuss configurations of imagers and the development of georeferencing strategies, e.g. starting with the Muller paper and going to today. Mention potential for the incorporation of digital elevation maps together enabled by GPUs and video game engines (\*\*reference that one paper that suggests using Unity or something similar for the projective geometry optimized for Cuda on NVIDIA GPUs\*\*).

In our previous work, we demonstrated a prototype autonomous robot team employing a drone based hyper-spectral imager which can learn the mapping from reflectance spectra to concentrations of variety of chemicals-of-concern by utilizing the information contained in the reflectance spectra captured with a HSI together with a (Lary et al., 2021). In this paper, we present a procedure based on the method of (Muller et al., 2002) for the rapid processing and georeferencing of imagery captured by a pushbroom HSI mounted on an autonomous. Associated code can be freely accessed and downloaded in (\*\*add reference to our repository\*\*)

(\*\*This point can be saved for the supervised learning paper\*\*) what we really want (usually) isn't the full spectra at each pixel

IMU - inertial measurement unit

#### 4.1.1 Autonomous Aerial Vehicle

subsection detailing the robot team, and specifically, the drone setup (HSI, NUC, Flir, etc...)

- HSI takes images - IMU/GPS onboard position/orientation determination - Intel NUC 1: manages HSI - SSD - Intel NUC 2: Processing - Ubiquity long-range wifi antenna (for streaming data products)

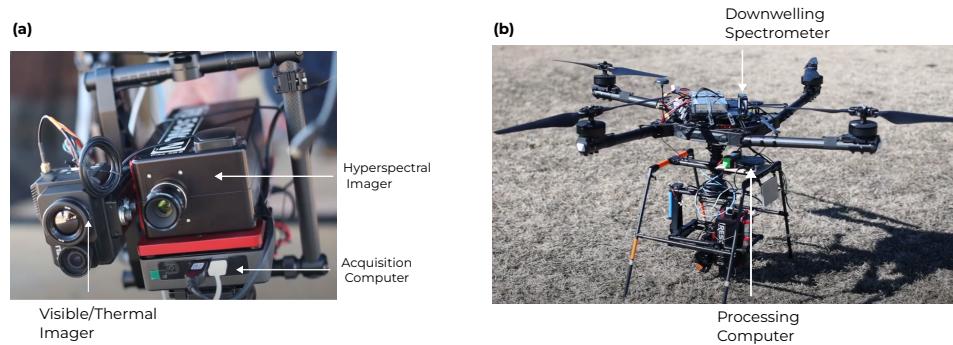


Figure 4.1: Components of the Autonomous Drone HSI platform.

#### 4.1.2 Real time processing of HSI imagery

- overview of processing pipeline

- Collection of raw HSI by imager - Reading stored binary (ENVI) radiance data into tensor - Reading Flight Data - Interpolate flight data from IMU to match HSI times - Reading stored downwelling irradiance spectrum - Interpolate irradiance spectrum to match HSI  $\lambda$ s - Conversion to Reflectance under the assumption of a Lambertian surface (perfectly diffuse) - Georectify to obtain new coordinates - resample to regular grid - compute derived  $\lambda$ -metrics (NDVI, etc...) - serve result (save to HDF5 or serve via web map service over wifi)

Georectification Procedure.

- collect datacube

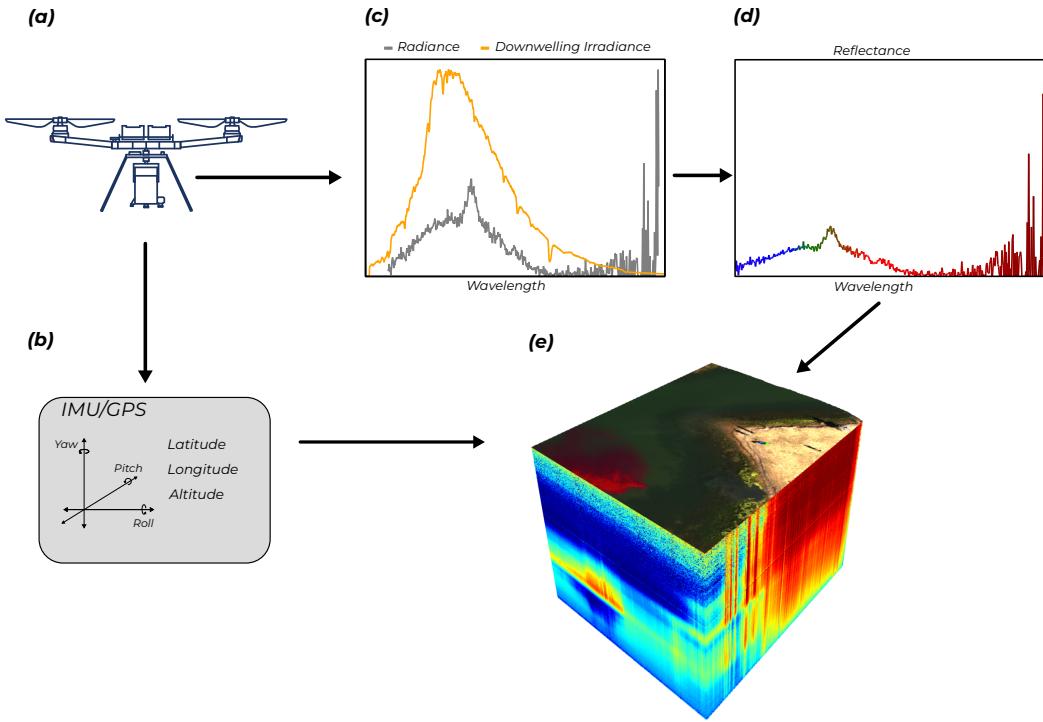


Figure 4.2: Solar Irradiance spectrum captured by downwelling spectrometer.

- read HSI in ENVI format
- read Flight Data
- interpolate flight data to HSI times
- read Downwelling irradiance spectrum
- convert radiance to reflectance
- georectify to obtain new coordinates
- resample to regular grid
- save results to HDF5

Code and data availability via OSN and Github

#### 4.1.3 Georectification Procedure

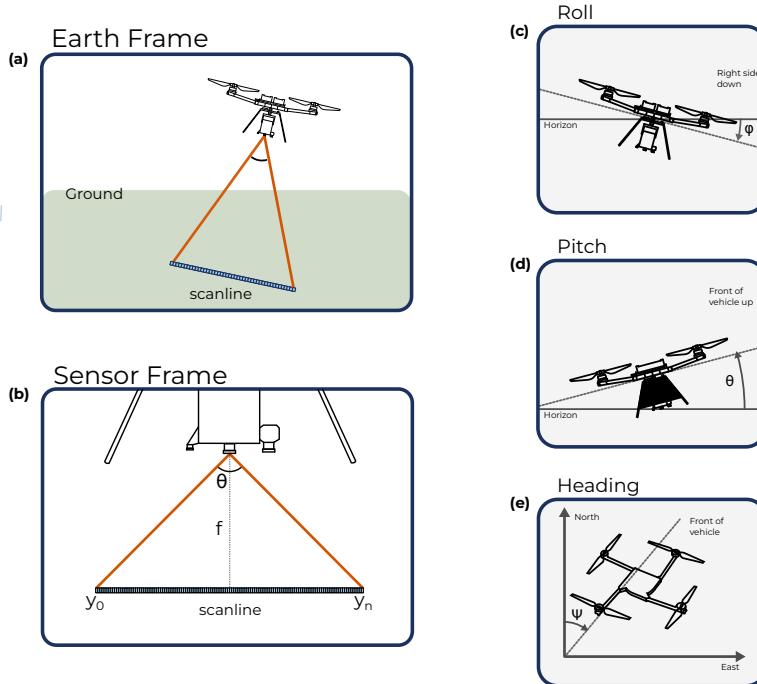


Figure 4.3: Visual representation of scan-line geometry for the drone based hyperspectral imaging platform.

## 4.2 Results

NOTE: we need to update these timings by first copying the HSI to the local drive (off of the external hard drive). This will probably have some effect on the read/write speeds for the test.

**NOTE** Stuff to do in this section:

- Create table with processing timings for each step on a typical HSI
- Create second table comparing time for full processing including resampling to a variety of scales (e.g. 5 cm, 10 cm, 25 cm) We will want a version of the processing function that does not include the uncertainties and the spectral indices

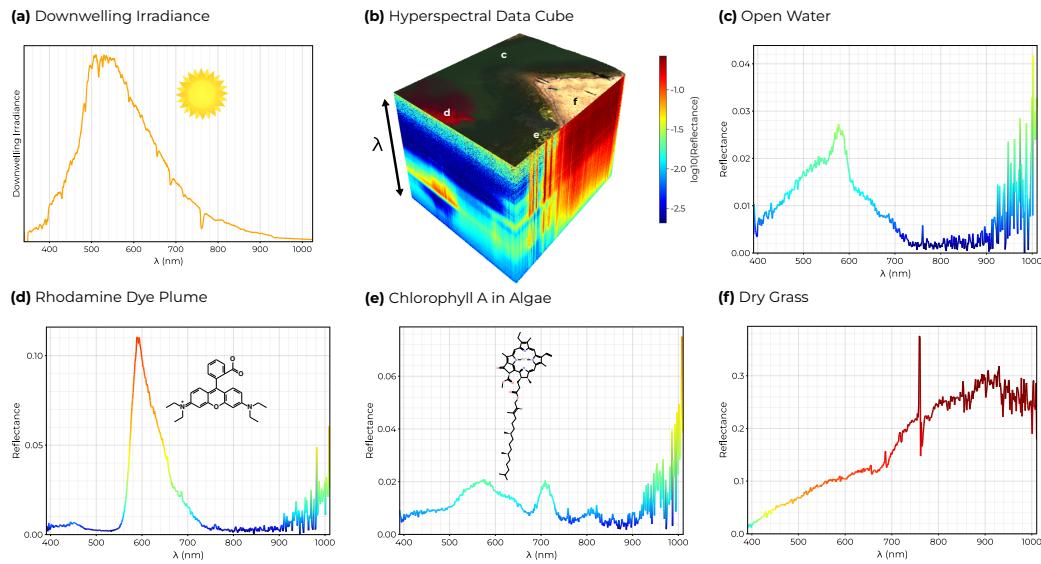


Figure 4.4: Annotated view of a hyperspectral data cube showcasing sampled spectra for a variety of constituents.

- create pipeline visualization showing size of each of the data products and their step in the pipeline (ending with a dataproduct like the NDVI)
- Create a figure showing the georeferenced image on top of the satellite background
- Create another figure of the entire map with a spectral index applied to one of the dye-release datasets (to demonstrate use for real-time application like oil-spill)
- create a figure of a sample HSI datacube in 3-d like from the previous paper with the rgb image on the top.

### 4.3 Discussion

- Incorporation of drone into autonomous robotic team - Long range wireless network (ubiquiti) for data downlink - Incorporation of edge-ML for direct determination of chemical concentration - Utility for navigation, waste removal, etc...

Discuss drawbacks (striping in images)



Figure 4.5: This is a figure. Schemes follow the same formatting. If there are multiple panels, they should be listed as: (a) Description of what is contained in the first panel. (b) Description of what is contained in the second panel. Figures should be placed in the main text near to the first time they are cited. A caption on a single line should be centered.

Authors should discuss the results and how they can be interpreted from the perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

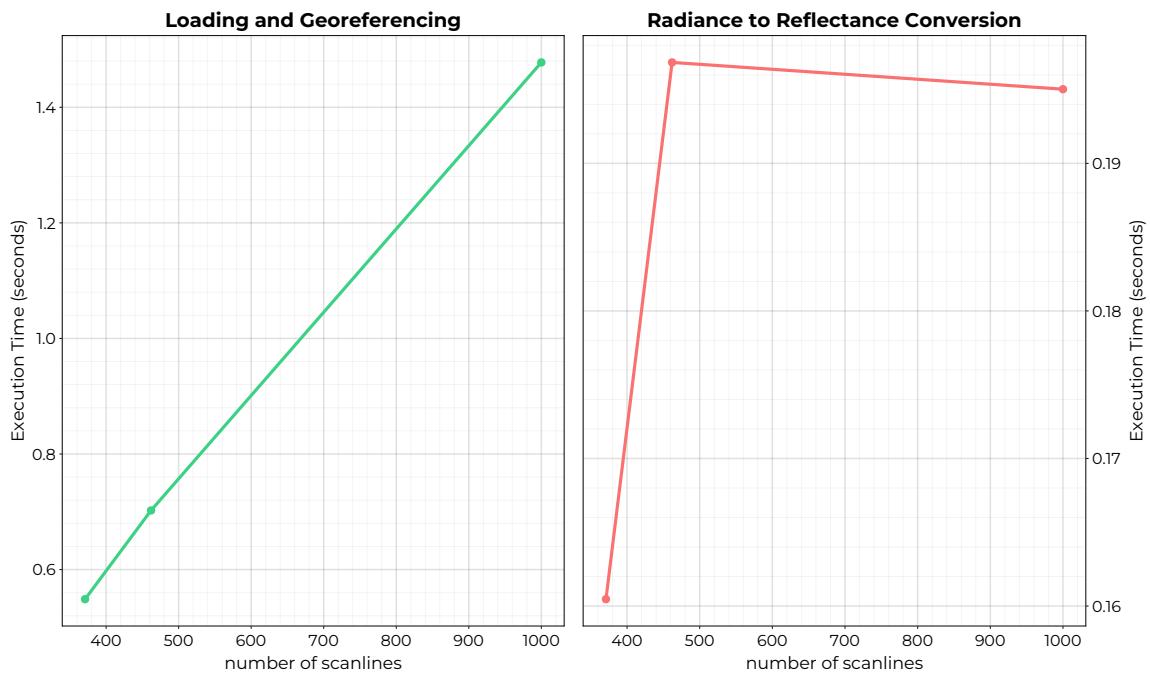


Figure 4.6: Description

## 4.4 Supervised Regression with Uncertainty Quantification

### 4.4.1 Hyperspectral Imaging and Remote Sensing

paper creating new spectral index for wheat rust (Zheng et al., 2018)

overview of NDVI (Huang et al., 2020)

paper using spectral indices for desertification (Lamqadem et al., 2018)

**HSI sattelite systems**

**HSI drone systems**

**Machine Learning using HSI**

add Xiao He reference here (Yu et al., 2021)

#### **4.4.2 In-situ Measurements / Robot Teams**

**Meteorology**

**Atmospheric Sensing**

add in reference to our AQ sensor network

#### **4.4.3 Julia Programming Language**

**scientific computing**

**differentiable programming ( $\partial P$ )**

**MLJ**

discuss general *learning network* framework and utilization of Directed Acyclic Graphs for pipeline construction

basic overview of MLJ as a Julia package (Blaom et al., 2020)

detailed description of *learning networks* (Blaom and Vollmer, 2020)

### **4.5 Materials and Methods**

#### **4.5.1 Autonomous Collection and Processing of Hyperspectral Images**

website for alta x (fre, 2021)

$$R = \pi \frac{L}{E} \quad (4.1)$$

#### **4.5.2 Georectification**

paper by Muller (Muller et al., 2002)

paper by Baumker (Bäumker and Heimes, 2001)

paper by Mostafa (Mostafa and Schwarz, 2000)

NOTE: need to double check (Muller et al., 2002) for convention used ( $\Phi$  vs  $\phi$ )

Heading correction due to longitude deviation in UTM system:

$$\kappa_{\text{geodetic}} = \kappa_{\text{geographic}} - \tan^{-1}(\sin \Phi \tan(\lambda - \lambda_{cm})) \quad (4.2)$$

scale factor:

$$s = \frac{z_{\text{sensor}} - z_{\text{ground}}}{f} \quad (4.3)$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_{\text{Object}}^{\text{UTM}} = f_{\text{geo}}^{\text{UTM}} \begin{pmatrix} \lambda \\ \Phi \\ z \end{pmatrix}_{\text{GPS}}^{\text{geo}} + s T_{\text{IMU}}^{\text{UTM}} R_{\text{sensor}}^{\text{IMU}}(\kappa, \phi, \omega) \begin{pmatrix} 0 \\ y_i \\ f \end{pmatrix}_{\text{object}}^{\text{sensor}} \quad (4.4)$$

#### 4.5.3 In-situ Data Collection via Robotic Sentinels

autonomous boat website (??, Ott)

**data collection process**

## **Data Collocation**

**CO**

**CDOM**

**HDO**

$Na^+$

$Ca^{2+}$

## **Other Targets**

### **Feature Engineering with Remote Sensing Indices**

include discussion of PCA as well as Standardization

#### **4.5.4 Exploratory Data Analysis**

##### **Variable Correlations**

##### **Machine Learning Models**

include Linear Models, Neural Networks, Decision Trees, Bagging, Boosting, and Stacking

The field of Machine Learning has seen the invention of a plethora of models able to perform inference for regression tasks. Naturally, this leads to the challenge of evaluating which model is best for a given problem. In 1992 David Wolpert developed the notion of Stacked Generalization to divorce arbitrary human choice from the model selection process. [1] The key idea is as follows: fit an instance of each model compatible with your dataset. Each model learns different features of the data so that an adjudicating model (called the

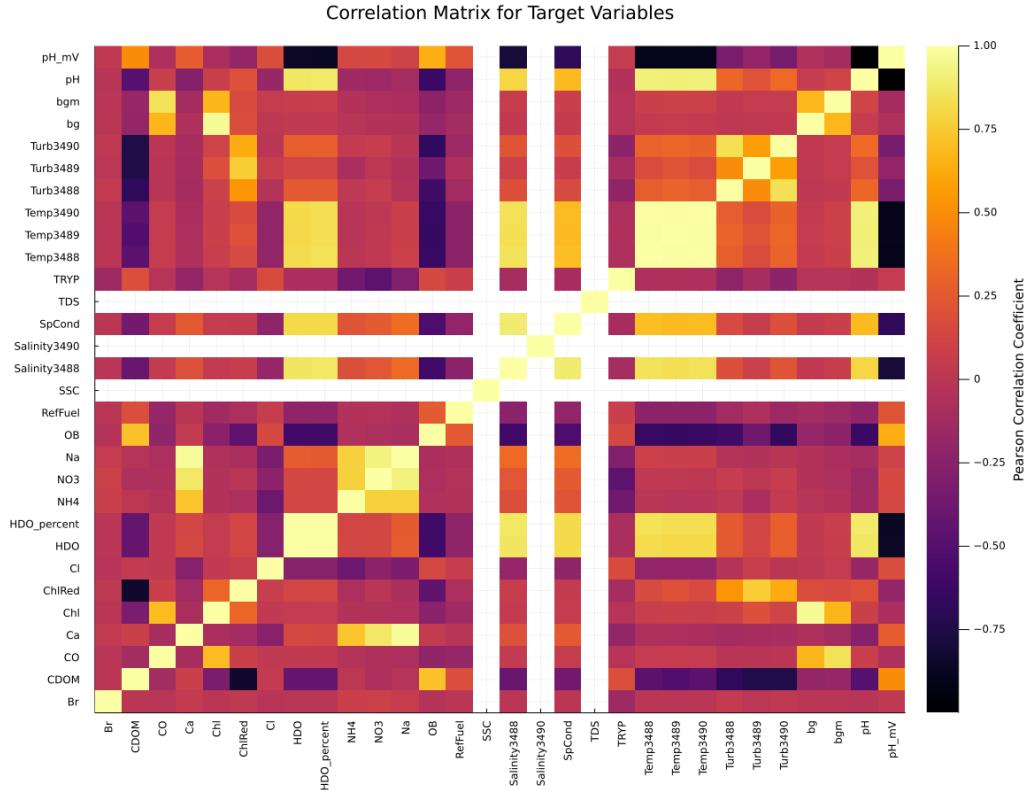


Figure 4.7: Correlation Matrix for Target variables measured by the boat.

super-learner) can then be trained to map the predictions from each individual learner to a final output prediction. This is incredibly easy to accomplish in the Julia programming language via the MLJ machine learning framework [2]. We train a variety of learners including artificial neural networks, bagged decision trees, boosted decision trees, polynomial regressors, quantile regressors, LASSO regressors, and K-nearest-neighbors regressors. To ensure proper data hygiene, individual learners are first trained on separate cross validation folds so as not to bias the super-learner. After training the super-learner, each individual model is then trained on the full dataset to produce the best possible model.

paper on stacking (Wolpert, 1992)

#### 4.5.5 Model Analysis

aleatoric and epistemic uncertainty

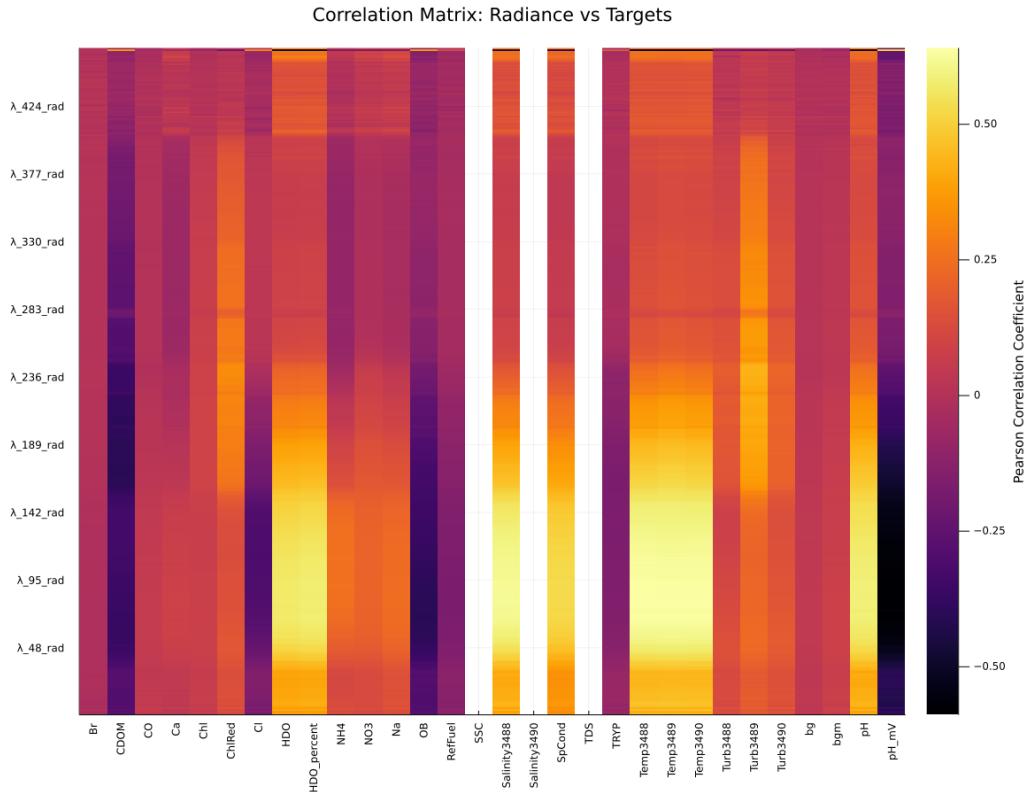


Figure 4.8: Correlation Matrix for Radiance measurements versus Target Variables.

The black-box nature most machine learning methods makes it difficult to consider how errors propagate through a trained model during the inference process. For many problems such as image classification this is not an issue, however, in physical sensing error considerations are vital to establish detection limits and prediction confidence. Typically, there are three considerations we must make: measurement uncertainty— how much we can trust our model’s inputs, measurement representativeness— how well each data point represents data within a local neighborhood, and model error— how far the prediction is from the true value. Machine learning models are typically trained to minimize the model error. The Julia programming language makes it possible to handle all three.

To address measurement uncertainty, we take advantage of Julia’s advanced automatic differentiation capabilities to enable differentiable programming ( P). This means that we can take derivatives of our trained ML model with respect to each input feature to perform

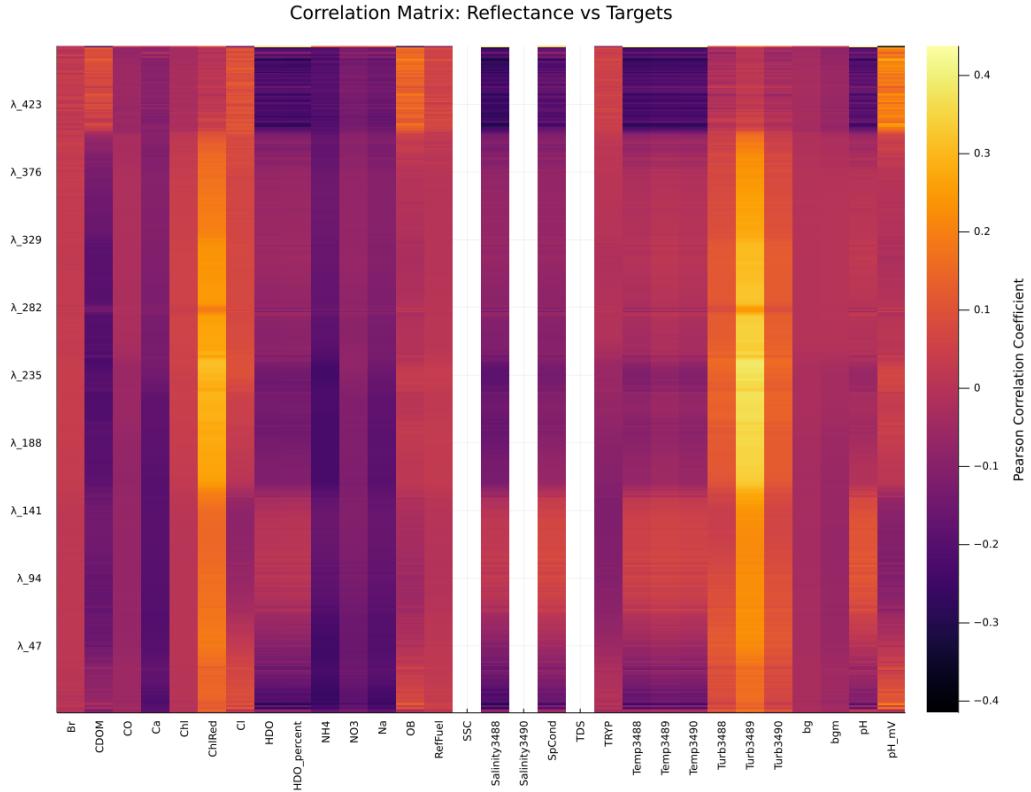


Figure 4.9: Correlation Matrix for Reflectance measurements versus Target Variables.

a sensitivity analysis. The final output uncertainty is then given via linear error propagation theory. [5]

Measurement representativeness can be trickier to handle. For example, in fluids it is common for two species to form a mixing layer that presents as a sharp discontinuity in measurement values. To quantify the degree to which neighboring points diverge we compute the average deviation across input measurements in a neighborhood around each data sample. This information can then be presented with model outputs to identify the regions where our model is valid. Similarly for ensemble ML models we compute the prediction deviation across all base learners to identify the representativeness of a data sample. Inputs that cause a large spread in base learner predictions indicate low representativeness in the training set. This analysis can then be used to plan further data collection missions that minimize time spent collecting data we have already present in our models.

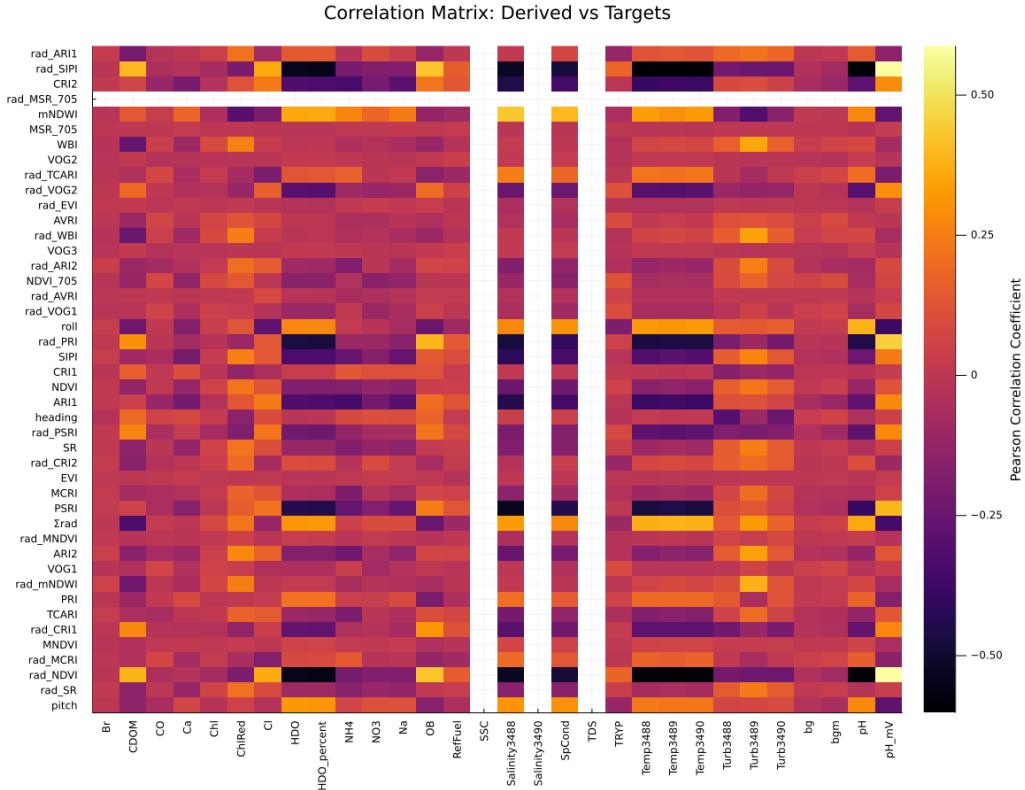


Figure 4.10: Correlation Matrix for derived measurements versus Target Variables.

## Uncertainty Propagation via $(\partial P)$

### Feature Ranking via SHAP values

shape value paper used in ShapML.jl (Strumbelj and Kononenko, 2013)

### Data Representativeness

i.e. the average deviation across pixels used average during collocation process. Also, we can look at relative disagreement between model predictions in bagged ensemble.

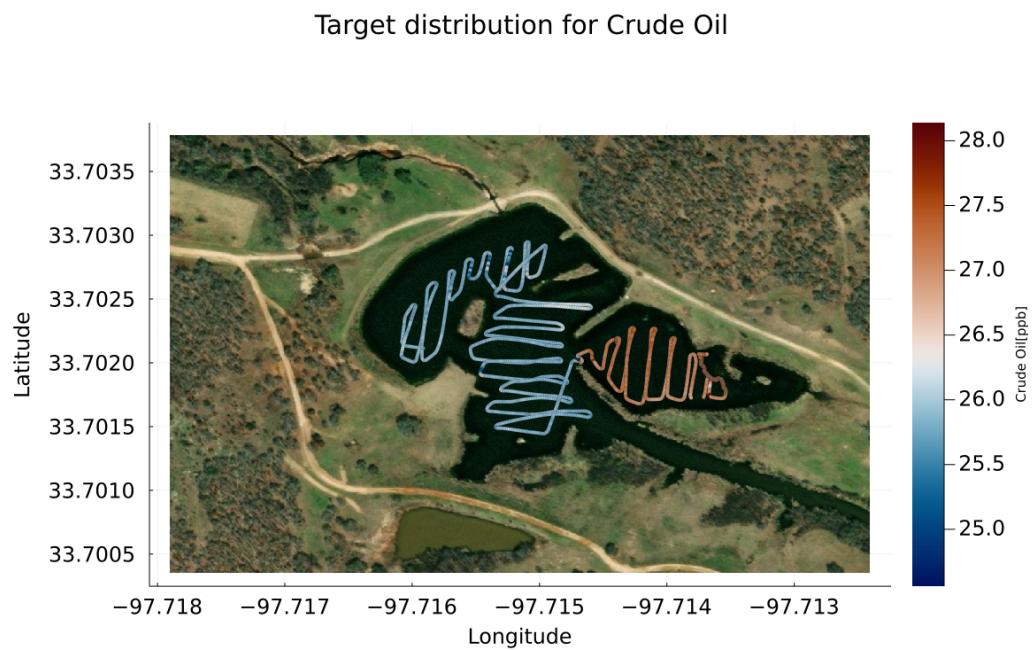


Figure 4.11: Target distribution of CO at Scotty's Ranch.

## 4.6 Results

### 4.6.1 CO

## 4.7 Methods for Unsupervised Classification of Hyperspectral Scenes in Novel Environments

### Unsupervised Classification

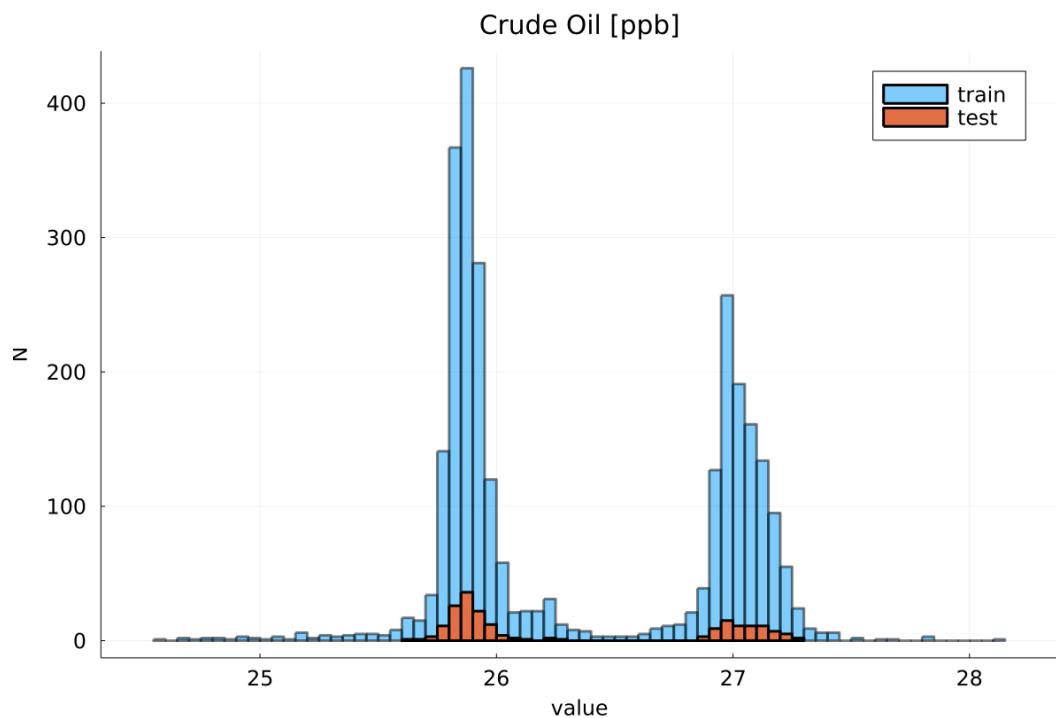


Figure 4.12: Stratified train-test split for CO.

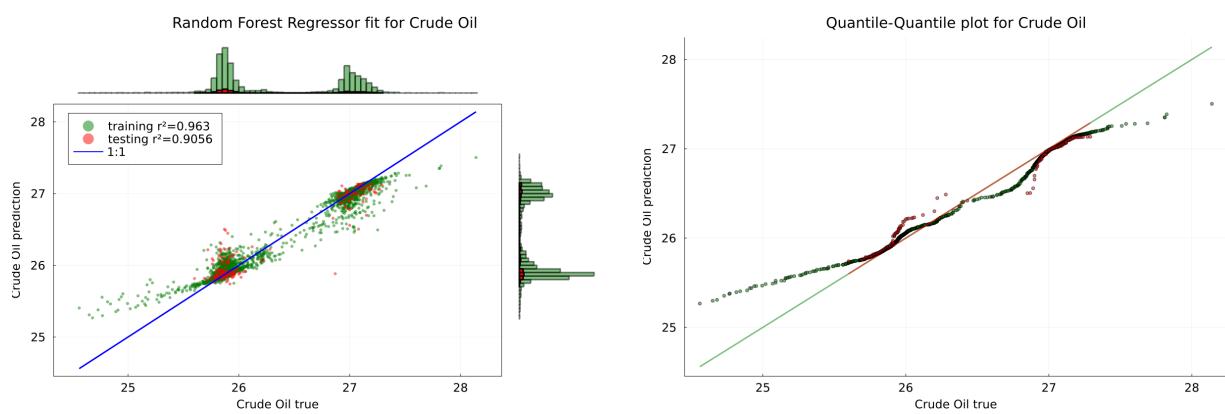


Figure 4.13: (a) Scatterplot for the trained random forest Crudo Oil model. (b) Quantile-Quantile plot for the same fit.

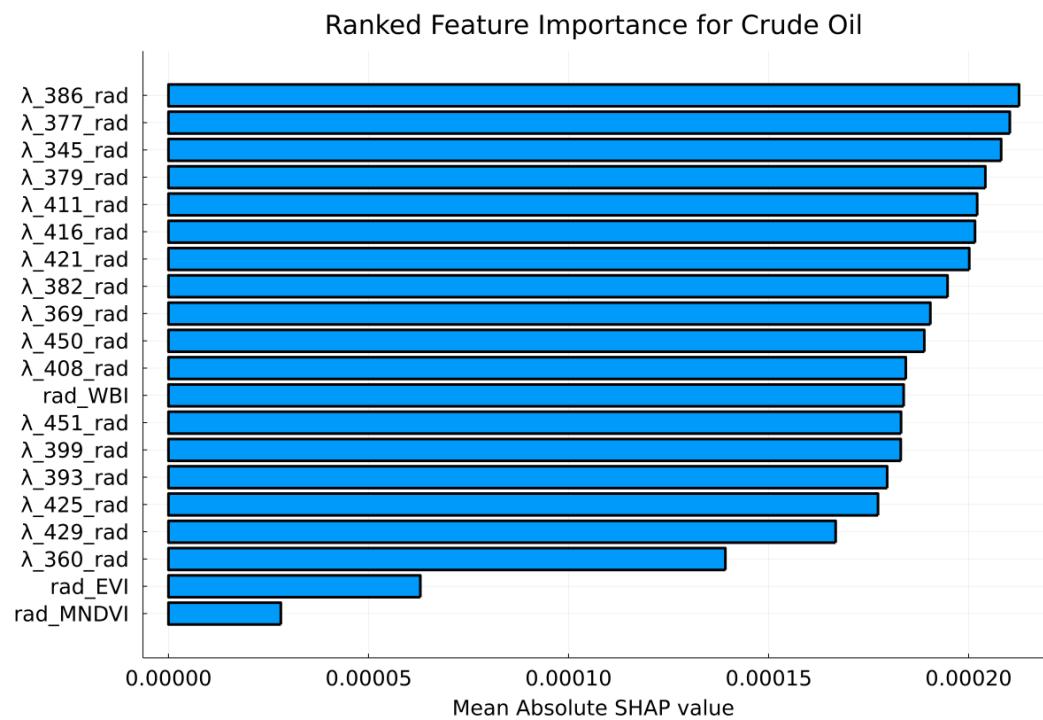


Figure 4.14: Ranked feature importance for Crude Oil. Importance is determined via mean absolute SHAP value.

### K-means Clustering of HSI Imagery (k=10)



Figure 4.15: Unsupervised clustering of hyper-spectral image data using the K-means algorithm.

## CHAPTER 5

### TIME SERIES METHODS FOR AIR QUALITY

With the proliferation of low-cost sensors for a wide range of IoT and wearable biometric applications, a systematic approach to both quality control and performing comprehensive time series analysis has significant value. It is highly desirable to be able to answer some basic questions using *just* the time series alone. For example: What is the likely sensor uncertainty given a time series of observations? How frequently should observations be made to adequately resolve typical temporal variability? How representative is a single observation of what one expects to see over a temporal (and spatial) window? Can we construct predictive models for the time series of a single sensor given a sufficient volume of sample data? In this chapter we seek to address these questions with direct application to the data collected by our low cost air quality sensor network. First, we will demonstrate how the *temporal variogram* provides a way to assess the intrinsic sensor uncertainty from a time series, and in the process, develop an open-source Julia implementation of a variety of variogram methods that can be used for any generic time series. Next, we present two techniques for physics-based time series modeling: the Hankel Alternative View Of Koopman (HAVOK) method, and the Hamiltonian Neural Network. It should be noted that despite the rapid pace of development in the fields of data-driven and scientific machine learning, many recently developed techniques like the Universal Differential Equations (UDEs), Hamiltonian Neural Networks, and others have yet to see wide spread application on noisy, real-world datasets. Our secondary goal for this chapter is therefore to demonstrate how, with some slight modifications, these techniques can be applied to real-world problems.

#### **5.1 Time-Series Methods for Uncertainty Quantification**

As we've already demonstrated, the shrinking cost of sensing technologies has improved our ability to create dense sensing networks. In order to make effective use of these sensors, and to

provide high quality data that can be used for critical decision making, it is vital we establish both the relevant sampling time scale and reasonable uncertainty estimates for measurements obtained by these sensors. For low cost sensors in particular, the manufacturer supplied uncertainty estimates tend to be highly conservative so as to minimize their responsibility for variation in device performance, and in general, to prevent unnecessary returns. Consider for example, the sample time series for particulate matter concentrations at a variety of size fractions collected by one of our low-cost sensing nodes for a single 24-hour period (figure 5.1: These sample data illustrate many typical features of air quality time series: there is

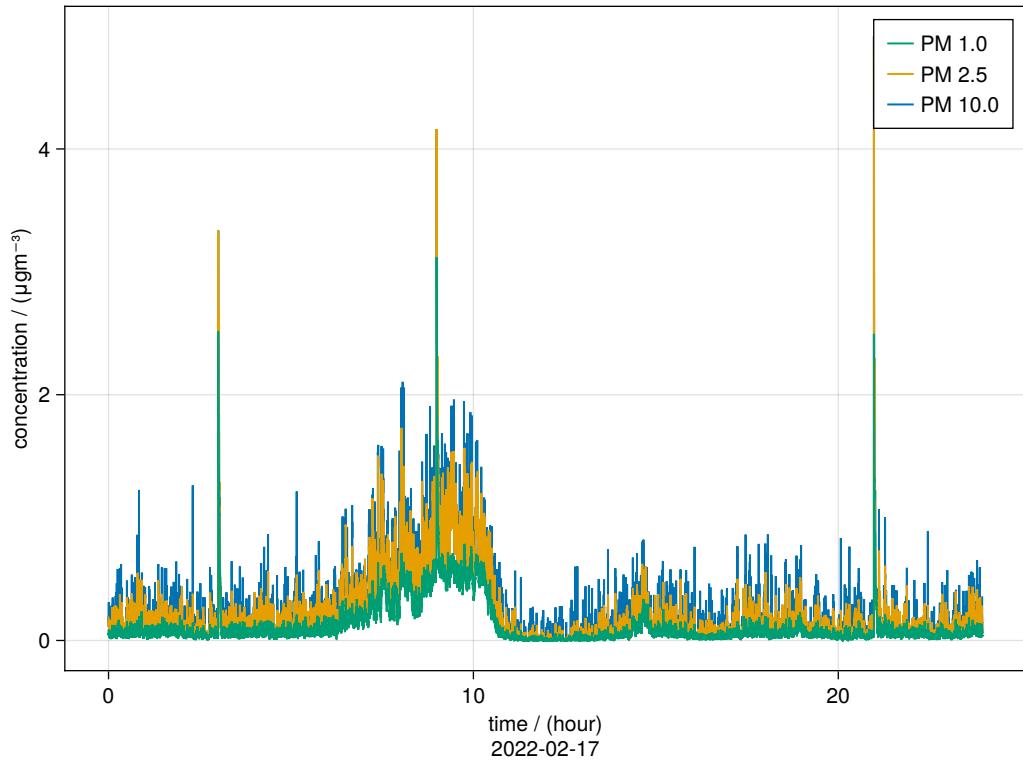


Figure 5.1: Time series of particulate matter at size fractions 1.0, 2.5, and  $10.0 \mu\text{m}$  captured at a single location over one 24-hour day.

a baseline level of noise, there are general periodic trends (an increase in concentration at all size fractions near morning traffic around hour 10), and there are intermittent transient events. If we suppose that we *already* know what the shortest time resolution needed to

resolve the important transients is, say  $\Delta t$ , then a straight forward approach to develop a baseline notion of uncertainty for a time series,  $z(t)$ , is to convolve a rolling window of width  $\Delta t$  across the signal so that at any time  $t$  we may define the associated *pseudo-observation* to be

$$\bar{z}(t) = \text{mean}(\{z_j\}) \pm \text{std}(\{z_j\}); \quad \text{where } j \in \left\{ j \mid t - \frac{\Delta t}{2} \leq t_j \leq t + \frac{\Delta t}{2} \right\} \quad (5.1)$$

where  $\text{std}(\{z_j\})$  is called the *representativeness uncertainty* as it measures how representative a single measurement is for the entire sampling window.

Unfortunately it may be difficult to know the relevant times scales of key transient events in advance. Given the fact that it is increasingly common for modern sensing systems (like our low-cost PM monitors) to provide sampling rates north of 1 Hz, we do not want to *shoot ourselves in the foot* by arbitrarily choosing an averaging window that will smooth away all the relevant features. Therefore, we seek to develop an alternative technique that can simultaneously establish the relevant temporal resolution for sampling whilst also enabling us to estimate the aleatoric, that is, intrinsic uncertainty of the sensing system.

### 5.1.1 Uncertainty Quantification with Temporal Variograms

The key feature that differentiates time series from other data sources is that at short time scales, neighboring samples are not independent; measurements taken close in time tend to be more correlated than measurements taken far apart. One way to measure this idea of self-similarity is by the auto-correlation function, which for a continuous signal  $z(t)$  is

$$R_{zz}(\Delta t) = \int_{-\infty}^{\infty} z(t + \Delta t) z^*(\Delta t) dt. \quad (5.2)$$

This can be efficiently computed for real-valued time series in  $\mathcal{O}(n \log n)$  using the *Fast Fourier Transform* via

$$R_{zz}(\Delta t) = \mathcal{F}^{-1} [\mathcal{F}(z) \cdot (\mathcal{F}(z))^*] \quad (5.3)$$

The first peak of this autocorrelation function is an excellent method to identify the relevant time-scale: high autocorrelation means we gain little extra information between points offset by a time lag of  $\Delta t$ . This puts us on the right track but does not provide a straight forward interpretation for an intrinsic uncertainty. Taking motivation from the representativeness uncertainty, let us instead consider the expected variance between neighboring samples as a function of the time lag  $\Delta t$ . That is,

$$2\gamma(\Delta t) = \text{Var}(z(t + \Delta t) - z(t)) \quad (5.4)$$

where  $\gamma$  is called the *semi-variogram* for  $z(t)$ . Expanding this definition yields

$$2\gamma(\Delta t) = \mathbb{E}[(z(t + \delta t) - z(t) - \mathbb{E}(z(t + \Delta t) - z(t)))^2] \quad (5.5)$$

which for timescales where  $z(t)$  is sufficiently stationary simplifies to yield

$$\gamma(\Delta t) = \frac{1}{2}\mathbb{E}[(z(t + \Delta t) - z(t))^2]. \quad (5.6)$$

For a perfect measurement system we would expect that taking  $\lim_{\Delta t \rightarrow 0} \gamma(\Delta t) = 0$ , since measurements taken at the same time by the same instrument *should* be identical. Intrinsic uncertainty will therefore manifest as a non-zero limit and taking the square-root of the variogram will yield units that match our original time series and provide a reasonable uncertainty estimate. For a given set of samples, we construct the *empirical variogram* as

$$\hat{\gamma}(\Delta t) = \frac{1}{2N(\Delta t)} \sum_i^{N(\Delta t)} (z(t_i + \Delta t) - z(t_i))^2 \quad (5.7)$$

where  $N(\Delta t)$  are the number of available observation pairs  $(z_i, z(t_i + \Delta t))$  for a lag of  $\Delta t$ .

There are three key parameters of the variogram  $\gamma(\Delta t)$  which we want to estimate from  $\hat{\gamma}(\Delta t)$ :

- **Nugget:** The  $y$ -intercept of  $\gamma(\Delta t)$  representing the limit as  $\Delta t \rightarrow 0$ .

- **Sill:** The value of  $\gamma(\Delta t)$  as  $\Delta t \rightarrow \infty$ . In other words, the expected variance for uncorrelated samples of our time series. Note the **partial sill** is defined to be the sill minus the nugget.
- **Range:** The  $\Delta t$  at which  $\gamma$  realizes its asymptote. Effectively this is the time lag beyond which samples are uncorrelated and can be used as a proxy for the relevant sampling time scale.

To do this, we first compute  $\hat{\gamma}(\Delta t)$  as above and then fit a model to  $\gamma(\Delta t)$ . There are a number of popular choices depending on the structure of a particular time series. In my open source Julia implementation, `TimeSeriesTools.jl`, the following models are made available:

- **Spherical:**

$$\gamma(\Delta t) = b + C_0 \left( 1.5 \frac{\Delta t}{r} - 0.5 \frac{(\Delta t)^2}{r^2} \right) \quad (5.8)$$

- **Exponential:**

$$\gamma(\Delta t) = b + C_0 \left( 1 - \exp \left( -\frac{\Delta t}{r} \right) \right) \quad (5.9)$$

- **Gaussian:**

$$\gamma(\Delta t) = b + C_0 \left( 1 - \exp \left( -\frac{(\Delta t)^2}{r^2} \right) \right) \quad (5.10)$$

- **Circular:**

$$\gamma(\Delta t) = b + C_0 \left( 1 - (2\pi) \cos^{-1} \left( \frac{\Delta t}{r} \right) + \frac{2\Delta t}{\pi r} \sqrt{1 - \frac{(\Delta t)^2}{r^2}} \right) \quad (5.11)$$

- **Cubic:**

$$\gamma(\Delta t) = b + C_0 \left( 7 \left( \frac{\Delta t}{r} \right)^2 - 8.75 \left( \frac{\Delta t}{r} \right)^3 + 3.5 \left( \frac{\Delta t}{r} \right)^5 - 0.75 \left( \frac{\Delta t}{r} \right)^7 \right) \quad (5.12)$$

- **Pentaspherical:**

$$\gamma(\Delta t) = b + C_0 \left( \frac{15}{8} \left( \frac{\Delta t}{r} \right) - \frac{5}{4} \left( \frac{\Delta t}{r} \right)^3 + \frac{3}{8} \left( \frac{\Delta t}{r} \right)^5 \right) \quad (5.13)$$

- **Sine Hole**

$$\gamma(\Delta t) = b + C_0 \left( 1 - \frac{\sin(\pi\Delta t/r)}{\pi\Delta t/r} \right) \quad (5.14)$$

where  $C_0$  is the partial sill,  $b$  is the nugget, and  $r$  is the range. Figure 5.2 demonstrates these fits for the PM<sub>10</sub> time series from above.

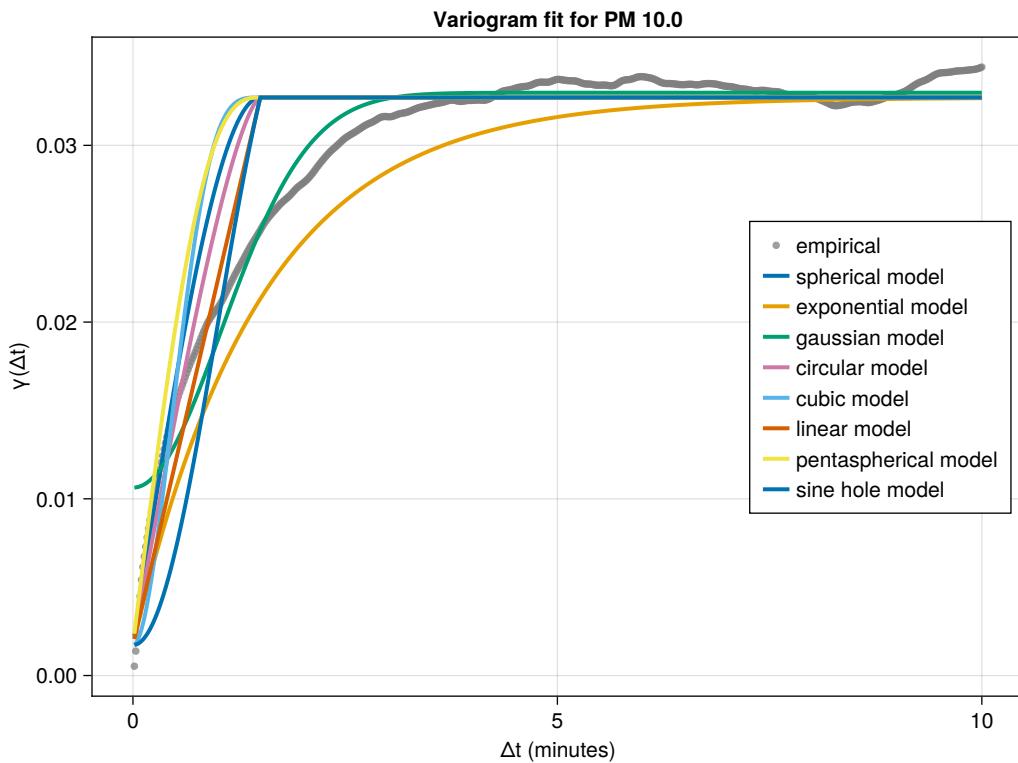


Figure 5.2: The empirical variogram and a variety of model fits obtained for the a PM 10.0 single-day time series

### 5.1.2 Next Steps

As of right now, I have developed an open source library to compute the temporal variogram and extract the estimated uncertainty for generic time series data. I have also developed a

pipeline for the acquisition and storage of time series data in publicly available S3 buckets on the Open Storage Network. My plan now is to evaluate the uncertainty for a variety of sensors from our network for data collected across the previous 1-3 years. We will then be able to investigate how stable these uncertainty estimates are as a function of time. For example, do we see any seasonal trends during the coldest winter months and warmest summer months? Does the uncertainty appear to increase over time so that we might be able to automatically identify when a sensor should be serviced/replaced?

## 5.2 Physics Informed modeling techniques for Air Quality Data

In order to provide actionable insights we must be able to effectively model the dynamics of our collected time series. In a perfect world, we would measure all relevant physical quantities such that the time evolution of local air quality at each sensor could be obtained by simulating the relevant micro-physics. However, low cost sensor networks are not equipped with all the necessary reference grade instruments needed to perform such simulations; accurate winds speed and direction sensors alone can cost hundreds to thousands of dollars and remote sensing data products are often unreliable at the ground level (i.e. in the human head space). We therefore are motivated to develop techniques to model our collected time series using only the data provided at a single node. There are many approaches to this task in the statistics and machine learning literature including statistical models like ARIMA and deep learning methods like Recurrent Neural Networks (Adhikari and Agrawal, 2013; Petneházi, 2019). While these methods can often lead to robust short term predictions, they do not incorporate prior physics knowledge and therefore are not primed to take advantage of underlying dynamical laws. Recently two interesting physics-informed, data driven techniques have been developed for just this type of scenario. The first we shall examine is the so-called *Hankel Alternative View Of Koopman* (HAVOK) framework which extends the principle of dynamic mode decomposition to nonlinear systems (Brunton et al., 2017). The second is

an technique dubbed the *Hamiltonian Neural Network* which extends the notion of a Neural Ordinary Differential Equation to allow a neural network to learn coordinate transformations or the original time series data which satisfy *Hamiltons equations* (Greydanus et al., 2019).

### 5.2.1 A Hybrid HAVOK UDE Approach

If we are to take a physics informed approach to modeling our time series, a good place to start is with the question: *How much information about the physical state of a system can be extracted from a single time series*. The answer, surprisingly, is *more than you might expect*. For simplicity let us consider a discrete, deterministic dynamical system described by a smooth function  $f : M \rightarrow M$  that maps points in the state-space manifold  $M$  to new points in the same manifold. We can define an *observable*  $y : M \rightarrow \mathbb{R}$  as a smooth function which maps a state  $x_t \in M$  to a number. The Takens Embedding Theorem then suggests that if the dynamical system described by  $f$  has a strange attractor of box counting dimension  $d$ , then there exists a  $k \geq 2d$  such that the embedding  $\mathbf{y} : M \rightarrow \mathbb{R}^k$  given by

$$\mathbf{y}(x) = (y(x), y(f(x)), y(f^2(x)), \dots, y(f^{k-1}(x))) \quad (5.15)$$

is a diffeomorphism. In other words, the time series  $(y(x_t), y(x_{t+1}), y(x_{t+2}), \dots)$  formed by tracking the evolution of a single point  $x_t$  in the state manifold  $M$  under the dynamical law  $x_k \mapsto x_{k+1} = f(x_k)$  captures the entire structure of the original dynamics in the state manifold (Takens, 2006). This remarkable result provides justification for our modeling and suggests that for practical purposes, we will want to construct our models by considering dynamics on vectors formed by so called *time delay embeddings* like that of the previous equation.

With this justification in hand, the next logical step is to propose a model for the dynamics which describe the evolution of such delay-embeddings. To do this, we will utilize techniques from Koopman Operator theory. Koopman operator theory provides a framework for understanding the time evolution of a dynamical system at the level of observable

functions rather than directly from the state itself. This is highly analogous to the equivalence between the Schrodinger and Heisenberg pictures in Quantum Mechanics where the former provides time evolution via dynamics of the state vector,  $|\alpha(t)\rangle$ , and the later provides dynamics on the observable operators,  $X(t)$ , of a system (Sakurai and Commins, 1995, p 80-84). In this same spirit, we seek to identify an operator  $\mathcal{K}$ , called the *Koopman Operator* for our dynamical system which satisfies

$$\mathcal{K}[y(x_k)] = y(f(x_k)) = y(x_{k+1}), \quad (5.16)$$

that is,  $\mathcal{K}$  takes an observable  $y$  and maps it to a new value which is just the observable evaluated at the *next* state after an update from  $f$ . This is highly appealing because such an operator (if it exists) is *linear* and results in a *linear* dynamical system for the time series of an observable, namely

$$y_t = \mathcal{K}^t(y_0) = y(f^t(x_0)). \quad (5.17)$$

The goal should now be clear: if we are able to utilize the data from our time series (interpreted as an *observable* of the underlying system's state space) to estimate the operator  $\mathcal{K}$ , then we will be able to model the dynamics of our time series for which a time delay embedding of sufficient length should produce an attractor that is diffeomorphic to the underlying dynamics of the system's true state. Unfortunately, the price paid for the linearity of  $\mathcal{K}$  is that its representation may demand an infinite dimensional basis (Brunton et al., 2021). Therefore, in practice, we hope that sufficient data will enable us to identify a suitable subspace of  $M$  for which a finite approximation of  $\mathcal{K}$ , say  $K$ , keeps our observables bounded. One approach (add citation for Mezic et al here) is to attempt an eigen decomposition of  $K$  using our time series data which relies on carefully crafting the observable functions  $y$ . In our case, we do not have direct control over the available observables; we are limited to the sensors provided by each sensing node. As an alternative approach, we take inspiration from

the popular method of Dynamic Mode Decomposition (DMD) to enable approximations of the Koopman operator from time series observations.

Originally, DMD was introduced to allow for the decomposition of expensive fluid dynamics simulations into a basis of *modes* (in the spirit of normal modes) such that the time evolution across the entire domain can be obtained by a linear combination of such modes with some simple time-dependent coefficients. With this as our inspiration, we can take a similar approach by forming a so called *Hankel matrix*,  $H$ , from our time series as

$$H = \begin{bmatrix} y_1 & y_2 & \dots & y_p \\ y_2 & y_3 & \dots & y_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_q & y_{q+1} & \dots & y_{p+q} \end{bmatrix} \quad (5.18)$$

(NOTE: we should update this to match our definition i.e. starting at  $y_1$  or  $y_q$  as the first element) where each column is a shifted time delay embedding of length  $q$  and each row corresponds to a subsample of our observable time series of length  $p$ . We can decompose this matrix using the SVD so that

$$H = U\Sigma V^T, \quad (5.19)$$

where the singular values  $\sigma_k = \Sigma_{kk}$  are ranked hierarchically by the relative contribution to  $H$ . If the singular values drop off fast enough, a reasonable approximation to  $H$  can be obtained by keeping only  $r$ -many singular values and singular vectors from the decomposition. This provides us a data-driven approach to approximating the Koopman operator for a subspace (that is a subset of singular vectors) that is approximately invariant. To see this, we can rewrite our Hankel matrix in terms of  $\mathcal{K}$  so that

$$H = \begin{bmatrix} y_1 & \mathcal{K}(y_1) & \dots & \mathcal{K}^{p-1}(y_1) \\ \mathcal{K}(y_1) & \mathcal{K}^2(y_1) & \dots & \mathcal{K}^p(y_1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{K}^{q-1}(x_1) & \mathcal{K}^q(x_1) & \dots & \mathcal{K}^{p+q-1}(x_1) \end{bmatrix} \quad (5.20)$$

this direct connection to the propagation by  $\mathcal{K}$  suggests we *should* be able to successfully obtain a linear model describing the dynamics of the singular vectors  $v_i(t)$ . The potential for multiple attractors as well as the challenges of noise in real datasets suggests that a linear model alone may not be good enough to recover the full dynamics and therefore the HAVOK model proposed by Brunton seeks to fit a model on the first  $r - 1$  singular vectors with *external forcing* provided by the  $r^{\text{th}}$  component:

$$\frac{d}{dt}\mathbf{v}(t) = A\mathbf{v}(t) + Bv_r(t) \quad (5.21)$$

where  $\mathbf{v} = (v_1, v_2, \dots, v_{r-1})^T$ . To demonstrate this approach, consider the famous Lorenz dynamical system given by

$$\dot{x} = \sigma(y - x) \quad (5.22)$$

$$\dot{y} = x(\rho - z) - y \quad (5.23)$$

$$\dot{z} = xy - \beta z \quad (5.24)$$

with parameters chosen so that  $\sigma = 10$ ,  $\rho = 20$ , and  $\beta = 8/3$ . Because the SVD provides a hierarchical ranking of the singular vectors by the relative scale of the singular values, we imagine this  $v_r(t)$  to be an intermittent external forcing due to the fact that it represents a low energy mode. A visualization of a sample trajectory for this system as well as the *embedded* attractor formed from the first three principal vectors  $v_1, v_2, v_3$  of the singular value decomposition of the associated Hankel matrix for the “observable”  $x(t)$  are illustrated in Figure 5.3.

This is a clear demonstration of Takens embedding theorem in action: the attractor formed from the first three singular vectors of the Hankel matrix for the single observable  $x(t)$  appear to capture key features of the original attractor formed from the state vector  $(x(t), y(t), z(t))^T$ . The rows of the matrix  $U$  correspond to the *modes* of the singular value decomposition and allow us to decompose the original delay embedding of the observable

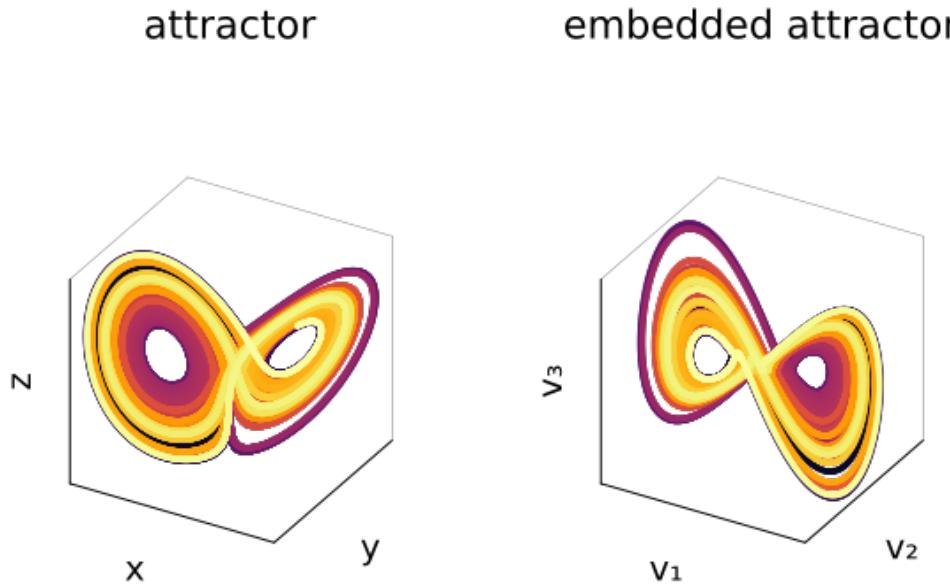


Figure 5.3: Comparing the original Lorenz attractor (left) to the embedded attractor learned after performing the SVD (right). Color indicates the time along the trajectory.

time series into this new representation. In this example, we have chosen  $r = 15$  as is done in (Brunton et al., 2017). A more programmatic approach is to evaluate  $r$  based on some information criterion. Alternatively, this can be left as a hyperparameter and adjusted to achieve a better fit. These modes are visualized in Figure 5.4 below.

The matrices  $A$  and  $B$  for our HAVOK model are obtained by standard linear regression. (We should add further description of how to form the linear system here, .e.g. we do linear regression on a non-square matrix of weights to let us fit  $A$  and  $B$  simultaneously.) The coefficients of these matrices for the Lorenz system described above are illustrated in Figure 5.5. We note that the form of the  $A$  matrix appears to closely follow the Toeplitz structure one would expect from the matrix of a finite difference stencil. We can then examine the time series obtained by integrating the HAVOK model forward to see how well we model the original dynamics of the singular vectors  $v_i$ . The first of these is shown in Figure 5.6 To further evaluate the quality of the resulting model, we can visualize the scatter diagram

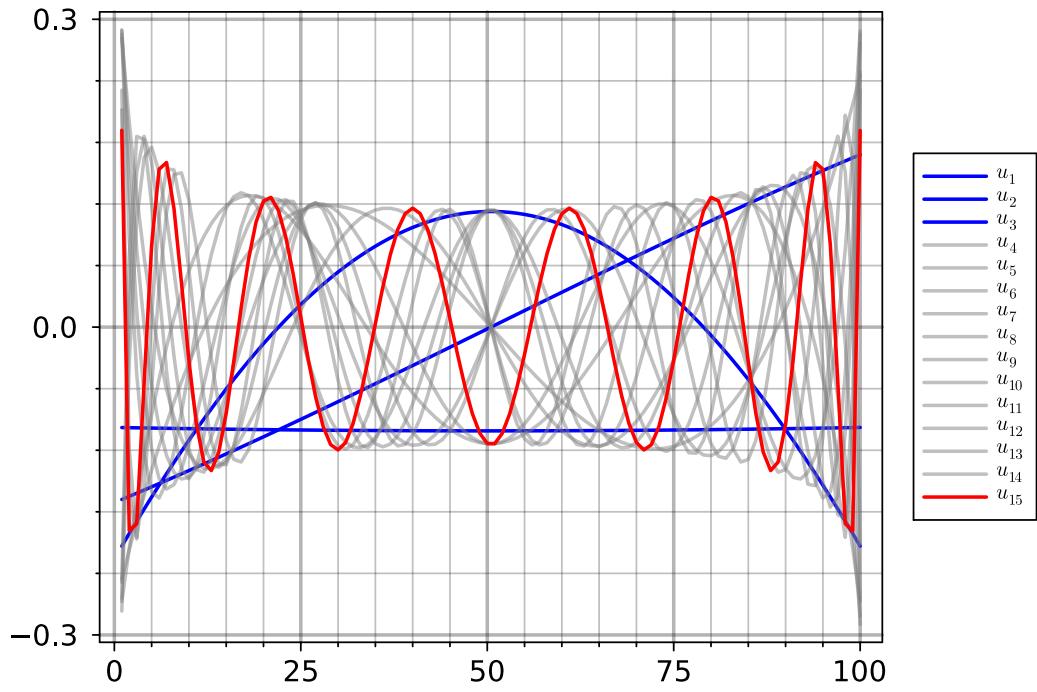


Figure 5.4: Eigenmodes of the embedded attractor extracted from the SVD.

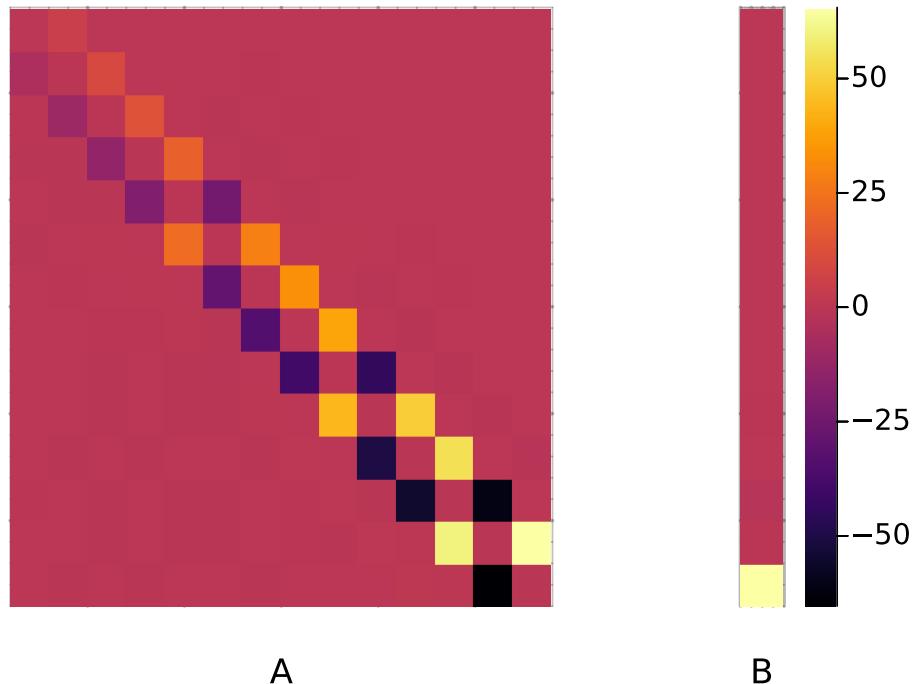


Figure 5.5: Heatmap of the linear Koopman operator together with the forcing activation.

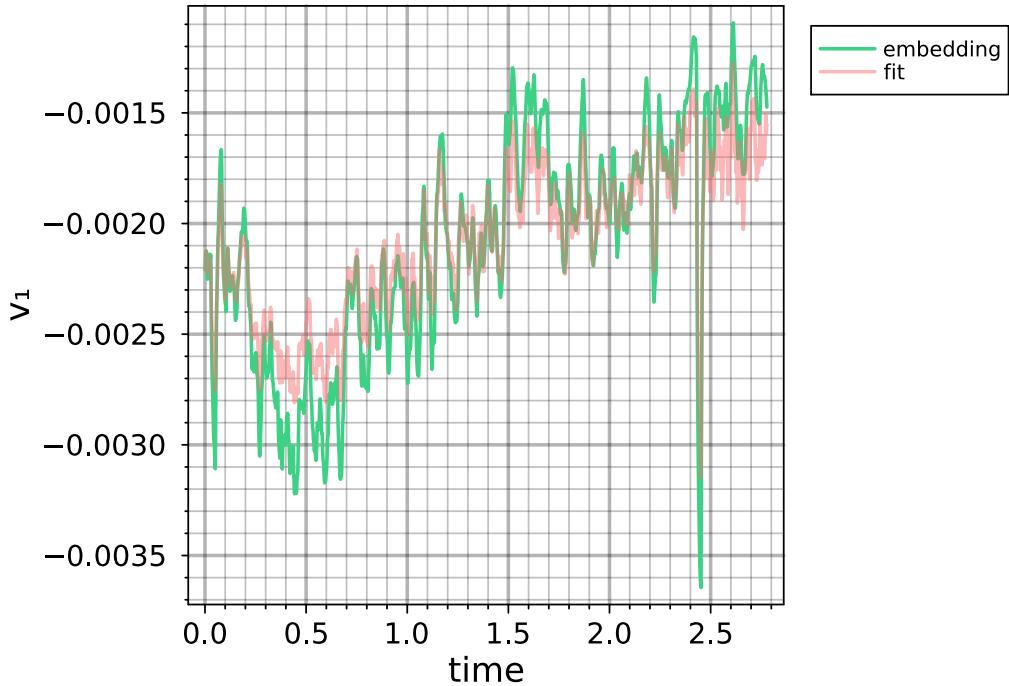


Figure 5.6: The reconstructed timeseries for  $v_1$  via the learned HAVOK model.

of the predicted values for  $v_1(t)$  on training data (the points that were used in the Hankel Matrix) as well as some holdout data. The results are shown in Figure 5.7.

Finally, we can visualize the original time series series and the embedded attractor by differentiating regions with high external forcing (salmon colored) from the regions where the observable is approximately linear (green colored), that is where the dynamics of the observable are invariant to our learned Koopman operator. This is illustrated in Figures 5.9 and 5.10.

## Results so far

What happens if we try and apply this HAVOK approach to a sample of our PM time series? This section details preliminary results with the method. Figure 5.11 illustrated the embedded attractor obtained by visualizing the first 3 singular vectors of the SVD for a  $\text{PM}_{2.5}$  time series. After attempting a variety of cut of values  $r$  as well as a variable number

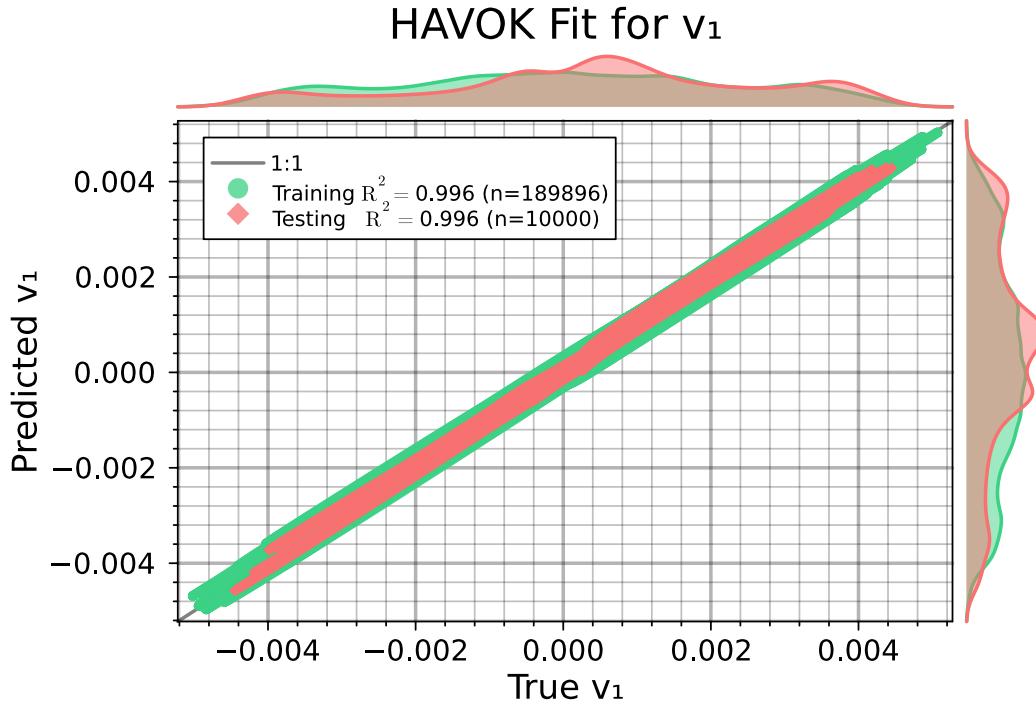


Figure 5.7: A scatterplot of the resulting HAVOK fit

of forcing terms, the *best fit* values for the  $A$  and  $B$  matrices are visualized in 5.12. As in the case of the Lorenz system, the  $A$  matrix has an interesting structure of alternating positive and negative coefficients in an anti-symmetric pattern along the diagonals. The learned coefficients of the external forcing activation matrix  $B$  appear to be biased towards negative values. Time series obtained for the first three components are visualized in Figure ???. In figure 5.14 we show the same data limited to the first 2.5 hours. This appears to suggest that the HAVOK models has learned a reasonable representation of the dynamics, however errors appear to accumulate over past 2.5 hours such that the model fails for long time integration.

Finally, if we examine the original time series and embedded attractor colored by the magnitude of the first forcing vector, we obtain Figures 5.15, and 5.16.

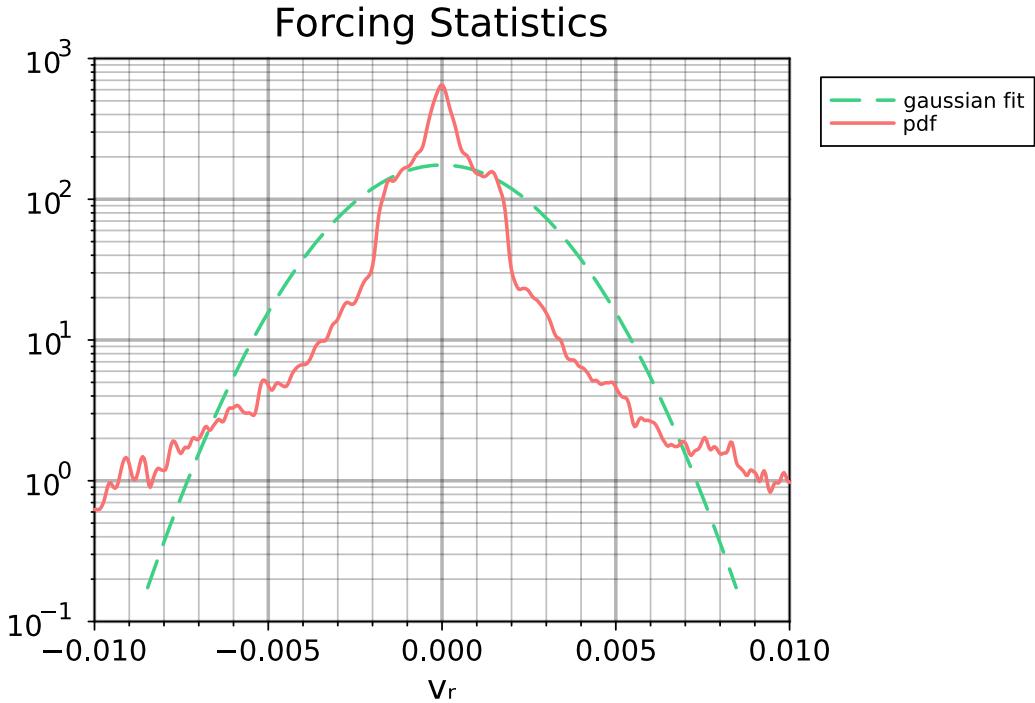


Figure 5.8: The statistics of the learned forcing function. The sharpness of the distribution (in comparison to a Normal distribution) indicates that the forcing is *intermittent*

## Next Steps

The preliminary results of our application of HAVOK to model PM data are very encouraging. The time series obtained in Figure 5.14 suggest we have achieved a reasonable first order model, however as the integration period extends, the agreement is quickly lost. This is likely due to our observations not being entirely invariant under the approximated Koopman operator and therefore it is unreasonable to expect that our model will be successful with linear and external forcing terms alone. To address this, I plan to combine the HAVOK approach with the Universal Differential Equations (UDEs) framework established by Rackauckas et al in (Rackauckas et al., 2020). That is, we can augment our learned HAVOK model to become

$$\frac{d\mathbf{v}}{dt} = A\mathbf{v}(t) + Bv_r(t) + \text{NN}(\mathbf{v}, t) \quad (5.25)$$

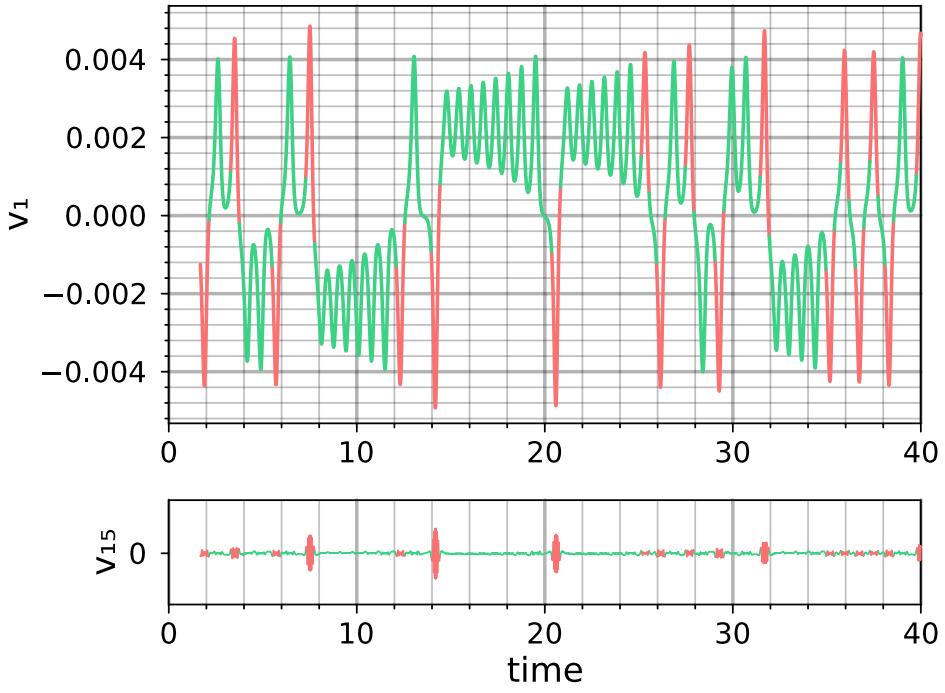


Figure 5.9: The time series for  $v_1$  marked where the forcing function is above a specified threshold.

where  $\text{NN}(\mathbf{v}, t)$  is a neural network. Since neural networks are universal function approximators, we can train the NN to learn the missing dynamics not present in our simple HAVOK model. Once we have accomplished this, we can then use sparse regression techniques such as SINDy as in (Brunton et al., 2016) to obtain a functional form for the missing nonlinear terms which complete the model.

### 5.2.2 Hamiltonian Neural Networks

The HAVOK modeling approach outlined in the previous section can largely be understood as a *data-driven* method. By carefully curating the Hankel matrix from our time series observations, we open the door to techniques from linear algebra that enable the generation of a straight-forward dynamical model. As an alternative method, let us now consider how we can use neural networks to directly model our time series data by forcing a neural network to learn useful transformations that simplify the modeling task.

## Attractor with Intermittent Forcing

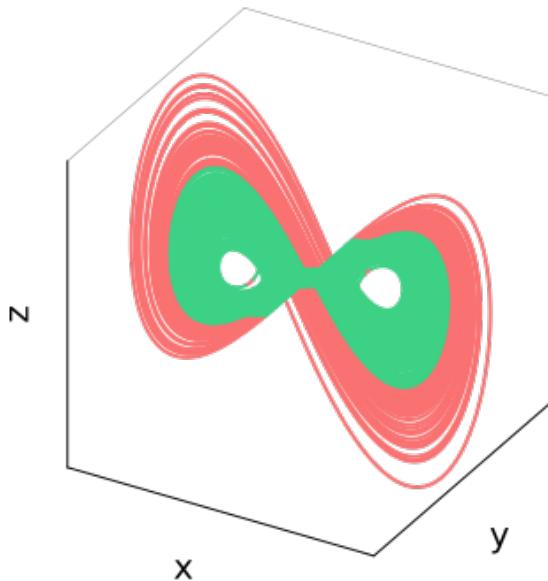


Figure 5.10: The embedded attractor colored by the presence of external forcing.

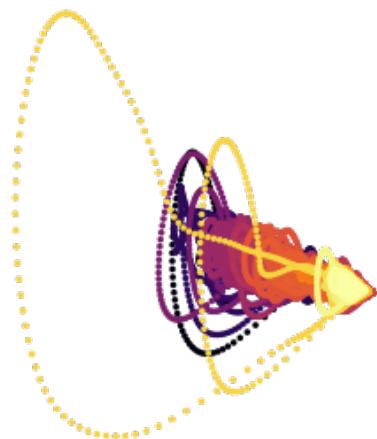


Figure 5.11: The embedded attractor for PM 2.5 time-series data.

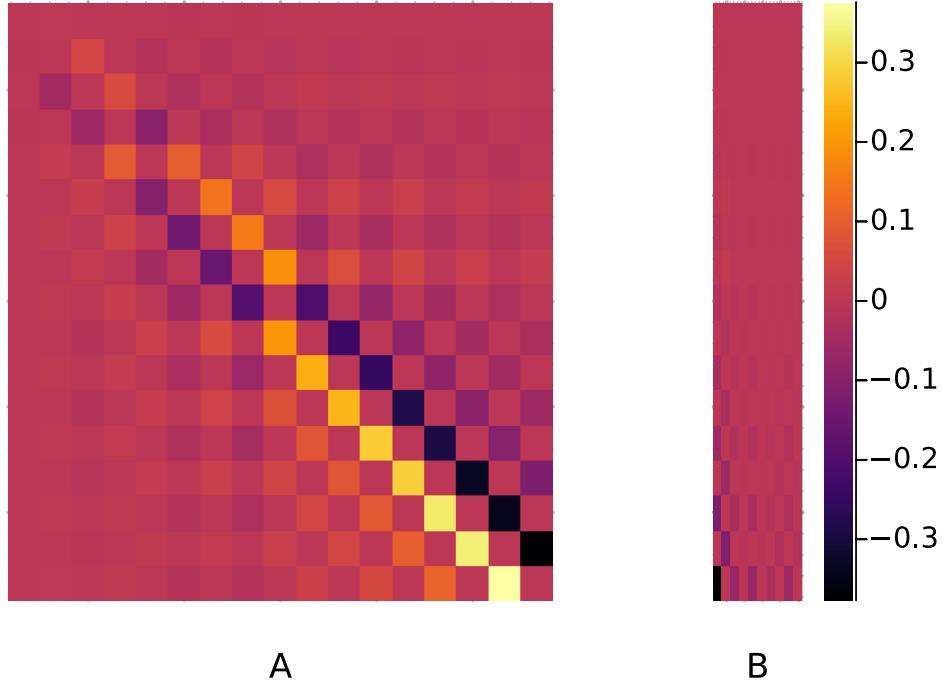


Figure 5.12: Heatmap of the linear Koopman operator together with the forcing activation.

For a number of years one of the most popular deep learning approaches for time series forecasting was the family of so-called *recurrent neural networks* (Petneházi, 2019). These models take the form

$$u_k = u_{k-1} + \text{NN}(u_{k-1}; \theta) \quad (5.26)$$

so that the output of the  $k^{\text{th}}$  layer is formed by starting with the previous output and correcting by the output of a neural network  $\text{NN}(u_{k-1}, \theta)$ . Recently Chen et al observed that this equation can be understood to be the discrete Euler scheme implementation for a continuous ODE of the form

$$\dot{u} = \text{NN}(u, \theta). \quad (5.27)$$

Therefore, by utilizing our adjoint method for determining the sensitivity of an ODE model to the parameters  $\theta$ , once can construct ODE layers whose output is the solution to a differential equations (Chen et al., 2018). By extension, a reasonable time series forecasting

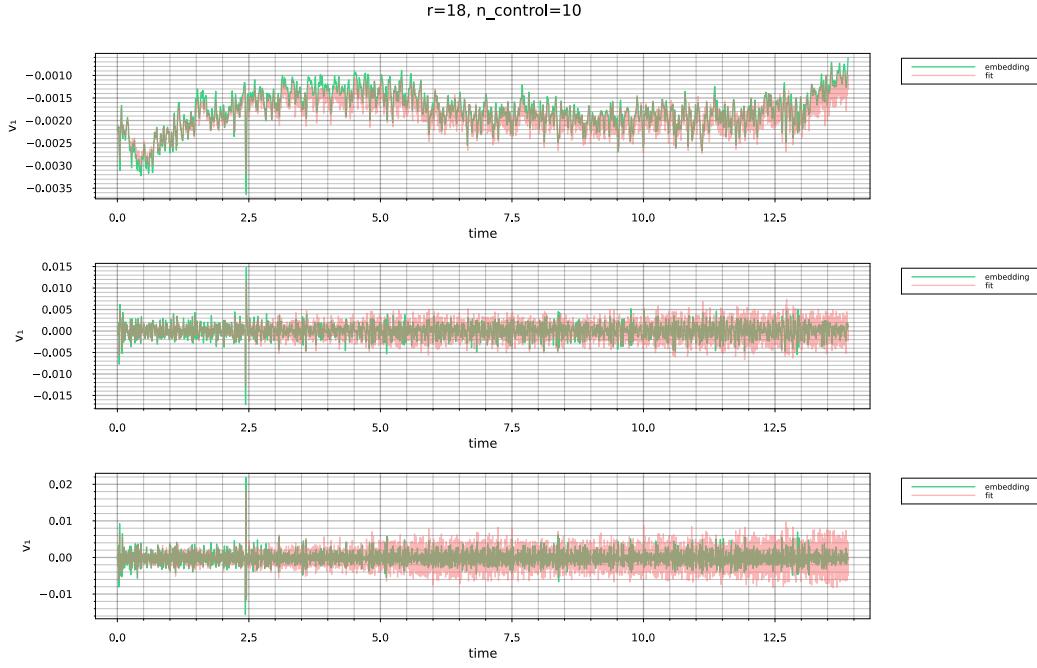


Figure 5.13: The reconstructed time-series for the first three embedding coordinates using the HAVOK fit.

approach would be to learn a model such that our time series  $z(t)$  satisfies

$$\dot{z}(t) = \text{NN}(z, t; \theta). \quad (5.28)$$

This technique is very general however it does not take advantage of any physical constraints imposed for time series sampling real, physical processes. Much effort is now being spent on how to regularize this approach to produce physics-informed Neural ODEs such that obey physical constraints like energy conservation (?). Still, these physics-informed constraints are not themselves built into the model (i.e. the neural network) but rather the weights are encouraged during training to conform to physics priors. An interesting alternative development came from Greydanus et al who took a different approach by incorporating physical laws directly into the neural network structure (Greydanus et al., 2019). Their method, the Hamiltonian Neural Network, seeks to train a neural network so that it

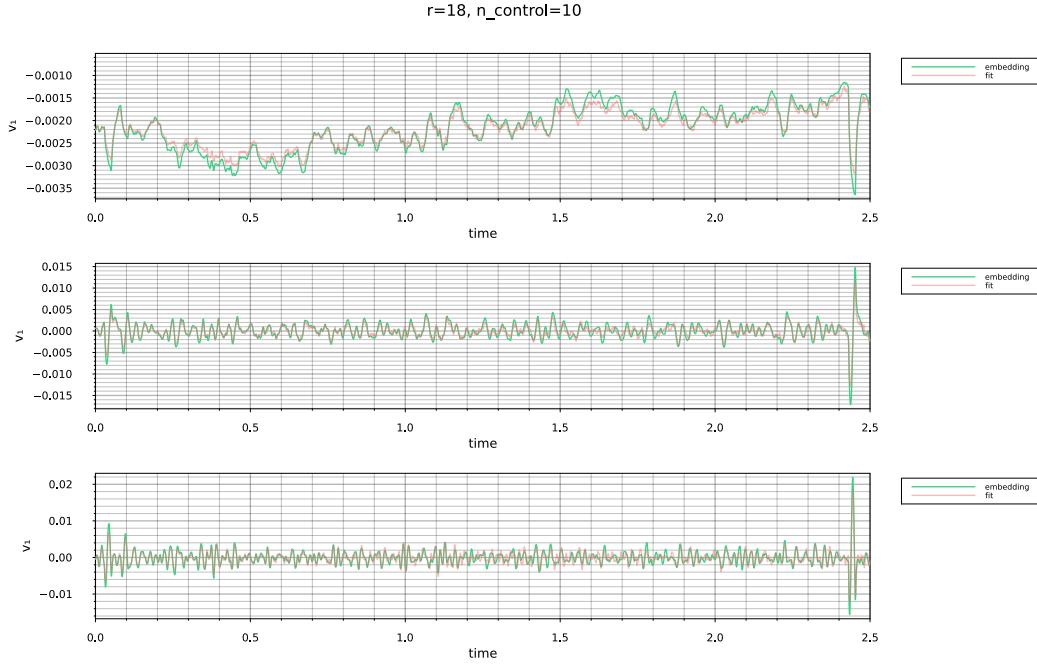


Figure 5.14: A zoomed in view of the same time-series reconstruction for the first 2.5 hours showing a very decent fit. Some kind of error appears to be accumulating over time.

satisfies Hamilton's equations, namely

$$\begin{aligned} H &= \text{NN}(q, p, \theta) \\ \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q} \end{aligned} \tag{5.29}$$

where  $q$  and  $p$  are the generalized coordinates and momenta for a system. By constructing a loss function of the form

$$\mathcal{L}_{\text{HNN}} = \left| \frac{\partial H}{\partial p} - \dot{q} \right|_2 + \left| \frac{\partial H}{\partial q} + \dot{p} \right|_2 \tag{5.30}$$

they found that the HNN is able to learn a valid representation of the Hamiltonian for a variety of physical systems (spring, double pendulum, etc.). Further, since Hamilton's equations must be integrated to obtain trajectories for a system, the learned Hamiltonian can easily explore counterfactual configurations not present in the supplied training data by

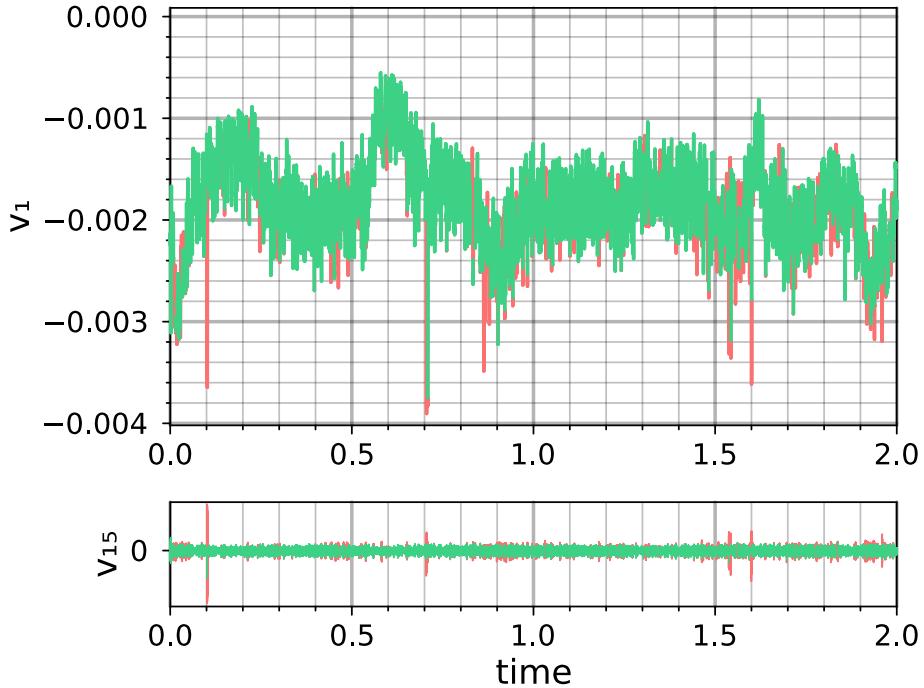


Figure 5.15: The first embedding coordinate with forcing above a critical threshold identified.

simply adjusting the Hamiltonian by addition of a constant (i.e. shifting the total energy). Because the total energy is a constant of motion, learning a representation of the Hamiltonian also improved the long term energy conservation of integrated trajectories as compared to directly attempting to predict  $\dot{q}, \dot{p}$  with a neural network alone.

To make approach feasible for real data, one must supply both position and momenta time series which may not be readily available. In their final example, the authors demonstrate how to accomplish this with an autoencoder network. By using additional encoder/decoder models, they were able to take the raw images of a *video* of a swinging pendulum and *learn* an effective transformation into a  $q, p$  pair that they then use to learn the Hamiltonian. Specifically, the HNN loss function is augmented to include two additional terms. The first is the standard autoencoder loss which demands an input that is encoded and then decoded to be similar to the original input. The second is an additional loss term used to encourage the moment portion of the encoder to be some function of the velocity of the position component,

## Attractor with Intermittent Forcing

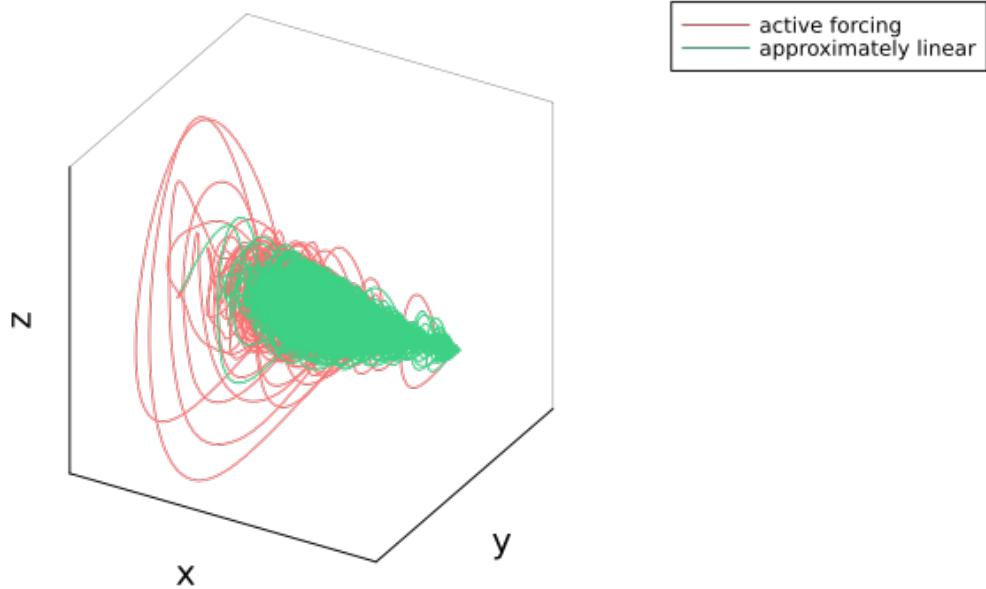


Figure 5.16: The original attractor now colored by external forcing above a threshold.

that is,

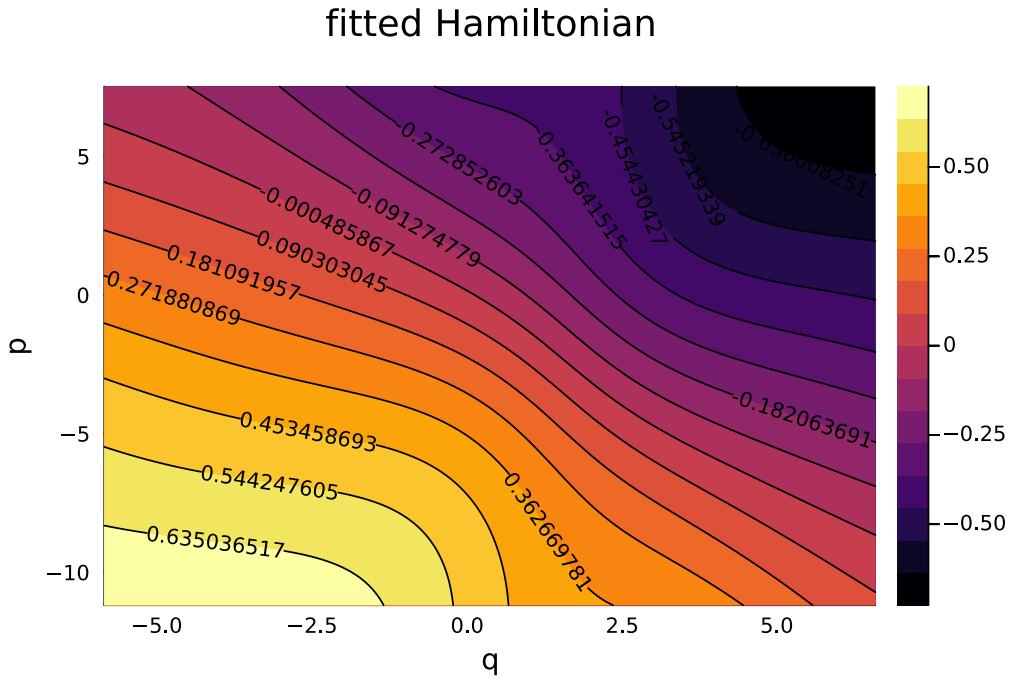
$$\mathcal{L}_{\text{momentum}} = |p^t - (q^{t+1} - q^t)|_2 \quad (5.31)$$

## Results

I have implemented the HNN model for time series data collected by one of our central nodes. For this implementation I have joined time smoothed time series for PM<sub>2.5</sub> together with temperature, pressure, and humidity time series from the same node. Using this dataset with the HNN training procedure outlined above, I have been able to generate preliminary mappings of the learned Hamiltonian such as those in Figure ??

## Next Steps

Now that I've developed the code to train an HNN, my plan is to attempt to train one for a data set consisting of roughly 1 week of observations from a Central Node to make sure



we include a few iterations of the diurnal cycle. With the trained HNN we can visualize the trajectory of the system through phase space, in particular, as it relates to the level surfaces of the learned Hamiltonian. I expect that for relatively calm conditions and short time scales, the trajectories will conform to the level sets of  $H$  so that the total energy is conserved. I am particularly interested in whether or not the magnitude of the gradient of  $H$  can be used as an indicator of potential external pollution sources. Similarly, I am curious about how long a trained HNN will be valid for a single sensor, as well as whether or not an HNN trained for a single node will be transferable to multiple nodes.

## CHAPTER 6

# A CHEMICAL DATA ASSIMILATION FRAMEWORK FOR INDOOR AIR QUALITY

## 6.1 Physics of Chemical Reactions: Chemical Reaction Kinetics

### 6.1.1 Overview

Since the early successes of Newton's descriptions of mechanics by means of simple forces acting on masses, scientists have sought to understand the dynamics of chemical reactions in terms of the detailed microphysics of molecular collisions. As we shall see, this approach can be utilized productively to justify the complicated temperature and pressure dependence of the reaction rate coefficients of many elementary reaction. However, even when considering the asymmetric structure of many molecules, and therefore, the dependence on orientation at the collision site, kinetic theory alone is unable to model reaction rates in all relevant temperature and pressure regimes. To do this, one can utilize the modern treatment of *Potential Energy Surface* (PES) theory together with the notion of short-lived, unstable intermediate reaction states to calculate reaction rate coefficient functions for specific reactants. *Ab initio* solution of the Schrodinger equation for the relevant nuclear geometries ( $3N$  reaction coordinates for  $N$  atoms) together with scattering and spectroscopic methods as suggested by (Neumark, 1992) have led to significant improvements in our understanding of reaction dynamics.

However the computational complexity of this task makes it prohibitively expensive to perform at the scale required for our desired chemical mechanism which consists of many hundreds to thousands of reactants together with as many as 16,000 unique reactions. Therefore in the following discussion, we shall primarily utilize the kinetic theory to justify the functional form for *most* rate coefficients with some reference to the extensions made by PES theory. We note that in practice, kinetic evaluations such as the periodic reports from the

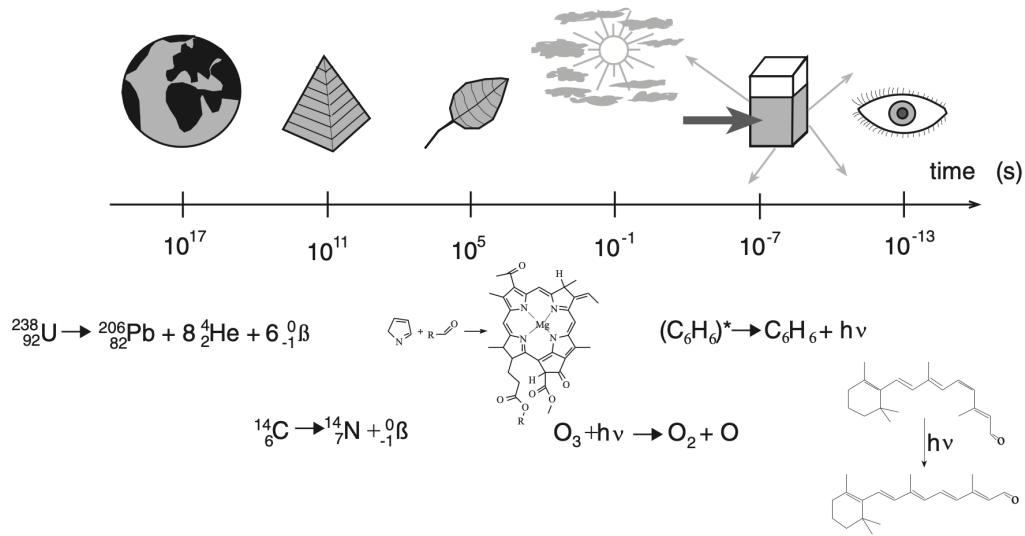


Figure 6.1: An illustration of the broad range of reaction time scales from the long-lived nuclear decay to rapid degradation of molecules by photolysis. Figure taken from (Arnaut and Burrows, 2006)

NASA Jet Propulsion Laboratory (Burkholder et al., 2020) utilize (justified) empirical fits to provide suggested functional forms for reaction rate coefficients.

### 6.1.2 Chemical Equilibrium and the Law of Mass Action

Before outlining the dynamics involved during complex chains of chemical reactions, it is worth spending some time to consider how we should treat chemical equilibrium. Usually in these scenarios we can not hold the internal energy fixed due to interactions with the environment, but rather, the temperature and pressure can often be treated as so. For example, we may be interested in gas-phase reactions occurring in ambient indoor air at or near standard temperature and pressure. In such scenarios, one finds the relevant potential energy to be the Gibbs, given by

$$G = U - TS + PV, \quad (6.1)$$

which is minimized in equilibrium under constant temperature and pressure.

This equation leads to the convenient thermodynamic identity

$$dG = -SdT + VdP + \sum_i \mu_i dN_i \quad (6.2)$$

from which we may identify the *chemical potential* of the  $i^{th}$  species as

$$\mu_i = \left( \frac{\partial G}{\partial N_i} \right)_{T,P,N_j \neq i}. \quad (6.3)$$

The fact that each  $\mu_i$  depends only on intensive state variables allows us to further simplify the relationship by considering what would happen were we to gradually increase the size of the system while maintaining the values of intensive parameters  $T$ ,  $P$ ,  $\mu_i$ . The result is  $G$  must increase in direct proportion to the increase in each  $N_i$ , that is:

$$G = \sum_i \mu_i N_i. \quad (6.4)$$

From equation 6.4 it is clear that the  $\mu_i$  can be understood as molecular *potentials* (i.e. chemical energy per molecule) in analogy to the notion of electric potential as a energy per unit charge.

For an ideal gas consisting of a single component we can combine equation 6.4 together with the identity

$$V = \left( \frac{\partial G}{\partial P} \right)_{T,N} \quad (6.5)$$

to obtain

$$\left( \frac{\partial \mu}{\partial P} \right)_{T,N} = \left( \frac{\partial}{\partial P} \frac{G}{N} \right)_{T,N} = \frac{1}{N} \left( \frac{\partial G}{\partial P} \right)_{T,N} = \frac{V}{N} = \frac{kT}{P} \quad (6.6)$$

so that by integration from a reference pressure, say  $P_0 = 1$  atm, we obtain the handy expression for the chemical potential

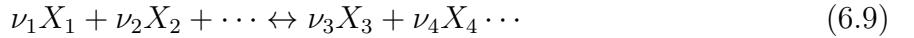
$$\mu(T, P) = \mu(T, P_0) + kT \ln(P/P_0) \quad (6.7)$$

which we shall use again momentarily.

Returning now to the notion of chemical equilibrium, recall that we must have  $dG = 0$  so that  $G$  is minimized. At constant temperature and pressure, this means

$$0 = dG = -SdT + VdP + \sum_i \mu_i dN_i = \sum_i \mu_i dN_i \quad (6.8)$$

and therefore, a generic reaction of the form



with chemical species  $X_i$  and stoichiometric coefficients  $\nu_i$  satisfies the condition that

$$\nu_1 \mu_1 + \nu_2 \mu_2 + \cdots = \nu_3 \mu_3 + \nu_4 \mu_4 + \cdots. \quad (6.10)$$

If we now make use of equation 6.7 with the identification of  $\mu^0 = \mu(T, P_0)$ , then we obtain

$$\sum_i^{\text{products}} \nu_i \mu_i^0 + \nu_i kT \ln(P_i/P_0) = \sum_j^{\text{reactants}} \nu_j \mu_j^0 + \nu_j kT \ln(P_j/P_0). \quad (6.11)$$

collecting terms involving the  $\mu_k^0$  to one side and multiplying through by Avogadro's constant, we obtain

$$RT \ln \left( \frac{\prod_j^{\text{reactants}} \left(\frac{P_j}{P_0}\right)^{\nu_j}}{\prod_j^{\text{products}} \left(\frac{P_j}{P_0}\right)^{\nu_j}} \right) = R \left( \sum_j^{\text{reactants}} \nu_j \mu_j^0 - \sum_i^{\text{products}} \nu_i \mu_i^0 \right) = \Delta G^0 \quad (6.12)$$

so that by exponentiation, we arrive at the simple expression:

$$\frac{\prod_j^{\text{products}} \left(\frac{P_j}{P_0}\right)^{\nu_j}}{\prod_i^{\text{reactants}} \left(\frac{P_i}{P_0}\right)^{\nu_i}} = \exp(-\Delta G^0/RT) \quad (6.13)$$

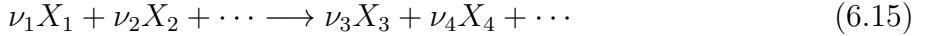
which through further application of the ideal gas law yields

$$\frac{\prod_j^{\text{products}} (X_j)^{\nu_j}}{\prod_i^{\text{reactants}} (X_i)^{\nu_i}} = K_{\text{eq}} \quad (6.14)$$

Here  $K_{\text{eq}}$  is a temperature dependent constant called the *equilibrium constant* for the reaction, and equation 6.14 is called the *law of mass action*. This expression indicates what we can expect to find if we allow our reactive system to proceed far enough to reach equilibrium. We shall later utilize this expression to perform a thermodynamic *sanity check* as is described in (Boldi, 1994).

### 6.1.3 Reaction Rate Laws

Having established the expected behavior at equilibrium, our task now is to establish the correct dynamical laws describing the variety of reactions which take place. To begin, let us consider again a generic chemical reaction of the form



To describe the dynamical process of a reaction, we can introduce a parameter  $\xi$  called the *reaction extent* such that at any time we have

$$\xi(t) = \frac{|N_i(t) - N_i(0)|}{\nu_i} \quad (6.16)$$

where  $N_i(t)$  is the number of the  $i^{\text{th}}$  species and  $\nu_i$  is the stoichiometric coefficient. The reaction rate is then easily understood as the rate of change of the reaction extent,

$$r := \frac{d\xi}{dt} = \frac{1}{\nu_i} \left| \frac{dN_i(t)}{dt} \right|. \quad (6.17)$$

Manipulating this expression to introduce the volume then leads us to

$$r(t) = \frac{1}{\nu_i} \left| \frac{dN_i}{dt} \frac{V}{V} \right| = \frac{V}{\nu_i} \left| \frac{d[X_i]}{dt} \right| \quad (6.18)$$

where  $[X_i]$  denotes the concentration (number density) of the  $i^{\text{th}}$  species participating in the reaction. All this is to say that upon rearranging the expression, we find

$$\left| \frac{d[X_i]}{dt} \right| = \nu_i \frac{r(t)}{V} = \nu_i v, \quad (6.19)$$

or in words, the absolute change in concentration of the  $i^{th}$  species as a function of time is proportional the quantity  $v = r(t)/V$  (called the reaction *velocity*) scaled by the stoichiometric coefficient  $\nu_i$  of the  $i^{th}$  species. For the purposes of our modeling tasks, we must now establish appropriate forms for the reaction velocity in terms of the relevant thermodynamic variables (i.e. temperature, and pressure) and constituent concentrations.

To begin, let us consider a bimolecular reaction



The simplest approach to modeling the reaction velocity for reactions of this type is to make the assumption that *each molecular collision leads to a reaction*. From this perspective, should then be able to derive the reaction velocity using the established statistics of molecular velocities together with appropriate data for the size of each reactant.

Let us treat reacting species as *hard spheres* of radii  $r_A$  and  $r_B$  respectively, or in other words, the molecules are spheres which only interact by contact (no long range interactions). Then a collision will occurs if the separation  $d_{AB}$  satisfies

$$d_{AB} = r_A + r_B \quad (6.21)$$

**NOTE:** Add image of hard spheres here

To start, let's consider collisions where  $B$  are stationary and  $A$  is seen to move with velocity  $\mathbf{v}_A$ .

**NOTE:** Add image of hard spheres forming cylinder here

Then during a time  $dt$ , the molecule  $A$  sweeps out a cylindrical volume

$$dV = \pi d_{AB}^2 v_A dt \quad (6.22)$$

in which collisions may occur. Supposing we have a density  $N_B/V$  of species  $B$ , then the collision rate (collisions per second) for a single particle of  $A$  will be

$$\frac{dV}{dt} \frac{N_B}{V} = \frac{\pi d_{AB}^2 v_A N_B}{V} \quad (6.23)$$

and therefore, the reaction velocity for  $N_A$  particles with average velocity  $\bar{v}_A$  is

$$v = \frac{\pi d_{AB}^2 \bar{v}_A N_A N_B}{V^2} \quad (6.24)$$

if instead particles of both  $A$  and  $B$  move relative to each other, then their *relative* velocity is given by the law of cosines:

$$v_{AB}^2 = v_A^2 + v_B^2 - 2v_A v_B \cos(\theta) \quad (6.25)$$

which, since all directions are equally probable, yields an average relative velocity of

$$\overline{v_{AB}}^2 = \sqrt{\overline{v_A}^2 + \overline{v_B}^2}. \quad (6.26)$$

The relative velocity describes the motion of the reduced mass  $\mu = m_A m_B / (m_A + m_B)$  and therefore under the standard Maxwellian velocity distribution, leads to

$$\overline{v_{AB}} = \sqrt{\frac{8kT}{\pi\mu}} \quad (6.27)$$

Combining everything together yields the reaction velocity

$$v = \pi d_{AB}^2 \frac{N_A}{V} \frac{N_B}{V} \sqrt{\frac{8kT}{\pi\mu}} = \pi d_{AB}^2 [A][B] \sqrt{\frac{8kT}{\pi\mu}} \quad (6.28)$$

which we might further simplify as

$$v = k[A][B] \quad (6.29)$$

$$k = \pi d_{AB}^2 \sqrt{\frac{8kT}{\pi\mu}} = \sigma \sqrt{\frac{8k_B T}{\pi\mu}} \quad (6.30)$$

where  $k$  is called the *reaction rate coefficient* and  $\sigma$  is the *reaction cross section*.

Excellent! We have discovered a couple of key features for the reaction velocity, namely, that it depends on a polynomial combination of reactant concentrations, *and* that there is clear temperature dependence due to the relationship between molecular velocities and temperature.

There are, however, some obvious limitations.

1. Not all collisions occur with orientations favorable for reaction (i.e. the hard sphere model isn't realistic for species).
2. Not all collisions will have enough energy for the reaction to proceed.

These limitations were well known, and in particular, Arrhenius suggested a competing function based on empirical studies of with

$$k = A \exp(-\alpha/T) \quad (6.31)$$

where  $\alpha$  is some constant that depends on the reaction taking place. We can address the first point in a *hand-wavy* manner by simply including a geometric correction factor,  $g \leq 1$ , to account for the distribution of favorable orientations. To address the second point, it is worth establishing some minimal energy required for the reaction,  $E_a$ , by examining our Maxwellian distribution in closer detail. As we shall see, this will allow us to recover the exponential dependence suggested by Arrhenius.

The velocities  $\mathbf{v}_A$  and  $\mathbf{v}_B$  of each colliding pair of reactants define a plane. Therefore, for ease of calculation, we approximate the distribution of velocities near the collision site by the 2-dimensional Maxwellian speed distribution:

$$f(v) = \frac{\mu}{k_B T} v \exp\left(-\frac{\mu v^2}{k_B T}\right) \quad (6.32)$$

so that the fraction of particles with speed in the range  $[v, v + dv]$  is

$$\frac{dN(v)}{N_{tot}} = f(v)dv = \frac{\mu}{k_B T} v \exp\left(-\frac{\mu v^2}{k_B T}\right) dv. \quad (6.33)$$

For an ideal gas under no external forces, we may identify the energy  $\epsilon = \mu v^2/2$  so that  $d\epsilon = \mu v dv$ . Therefore, in energy space, this ratio becomes

$$\frac{dN(\epsilon)}{N_{tot}} = \frac{1}{k_B T} \exp(-\epsilon/k_B T) d\epsilon \quad (6.34)$$

If  $E_a$  is the minimum (*activation*) Energy required to engage the reactants, then the ratio of reactants with sufficient energy for the reaction to proceed is determined by integration to be

$$\frac{N(\epsilon)}{N_{tot}} \Big|_{\epsilon > E_a} = \int_{E_a}^{\infty} \frac{1}{k_B T} \exp(-\epsilon/k_B T) d\epsilon = \exp(-E_a/k_B T) \quad (6.35)$$

so that we may justify an additional correction to our reaction rate coefficient

$$k = g\pi d_{AB}^2 \sqrt{\frac{8k_B T}{\pi\mu}} \exp(-E_a/k_B T) \quad (6.36)$$

The final augmentation we can perform without before leaving collision theory behind is to observe that the cross-section  $\sigma$  was treated independently from the requirement of a minimum activation energy. With that in mind, suppose instead that a reaction cross section *only makes sense* if the reactants have the required minimum energy, that is

$$\sigma(\epsilon) = \begin{cases} \pi d_{AB}^2 & \epsilon > E_A \\ 0 & \text{otherwise} \end{cases} \quad (6.37)$$

If this is true, we can not keep the cross section outside of the ensemble average.

Allowing ourselves to utilize the full, three-dimensional Maxwellian speed distribution, we have

$$\begin{aligned} k &= \int_0^{\infty} \sigma(v) v f(v) dv \\ &= 4 \left( \frac{\mu}{2\pi k_B T} \right)^{3/2} \int_0^{\infty} \sigma(v) v \cdot v^2 \exp\left(-\frac{\mu v^2}{2k_B T}\right) dv \end{aligned} \quad (6.38)$$

which in terms of energy yields

$$\begin{aligned} k(T) &= \left( \frac{1}{\pi\mu} \right)^{1/2} \left( \frac{2}{k_B T} \right)^{3/2} \int_0^{\infty} \epsilon \sigma(\epsilon) \exp(-\epsilon/k_B T) d\epsilon \\ &= \left( \frac{1}{\pi\mu} \right)^{1/2} \left( \frac{2}{k_B T} \right)^{3/2} \int_{E_a}^{\infty} \epsilon \pi d_{AB}^2 \exp(-\epsilon/k_B T) d\epsilon \\ &= \pi d_{AB}^2 \sqrt{\frac{8k_B T}{\pi\mu}} \left( 1 + \frac{E_a}{k_B T} \right) \exp(-E_a/k_B T) \end{aligned} \quad (6.39)$$

In summary, we have established that for bimolecular reactions, the (simple) theory of molecular collisions allows us to model the reaction velocity by

$$v = k(T)[A][B] \quad (6.40)$$

$$k(T) \approx \pi d_{AB}^2 \sqrt{\frac{8k_B T}{\pi \mu}} \left( 1 + \frac{E_a}{k_B T} \right) \exp(-E_a/k_B T) \quad (6.41)$$

To derive a more accurate form for the rate coefficient  $k$ , we can result to PES, simulation, and statistical sampling techniques (Partridge et al., 1993; Wu et al., 2006, for example). For our purposes, it is enough to have justified the kinds of functional dependence on temperature we can expect to encounter when collating available data on rate coefficients in the present literature.

## 6.2 Photolysis

Give an overview for photolysis (perhaps borrow from Dr. Lary's Textbook)

For experimental reasons, the wavelengths ( $\lambda$ ) most commonly used for initiating photochemical processes vary between the ultraviolet (200 – 250 nm) and the near infrared (750 – 800 nm). Light at these wavelengths has an energy which corresponds roughly to 600 – 50 kJ mol<sup>-1</sup>, which is very close to the energies of many chemical bonds.

## 6.3 Summary of Chemical Mechanism Kinetics

Give a summary of kinetics for each of the 3 reaction types. Then finish by writing the form of the combined differential equations for the entire system (and it's Jacobian)

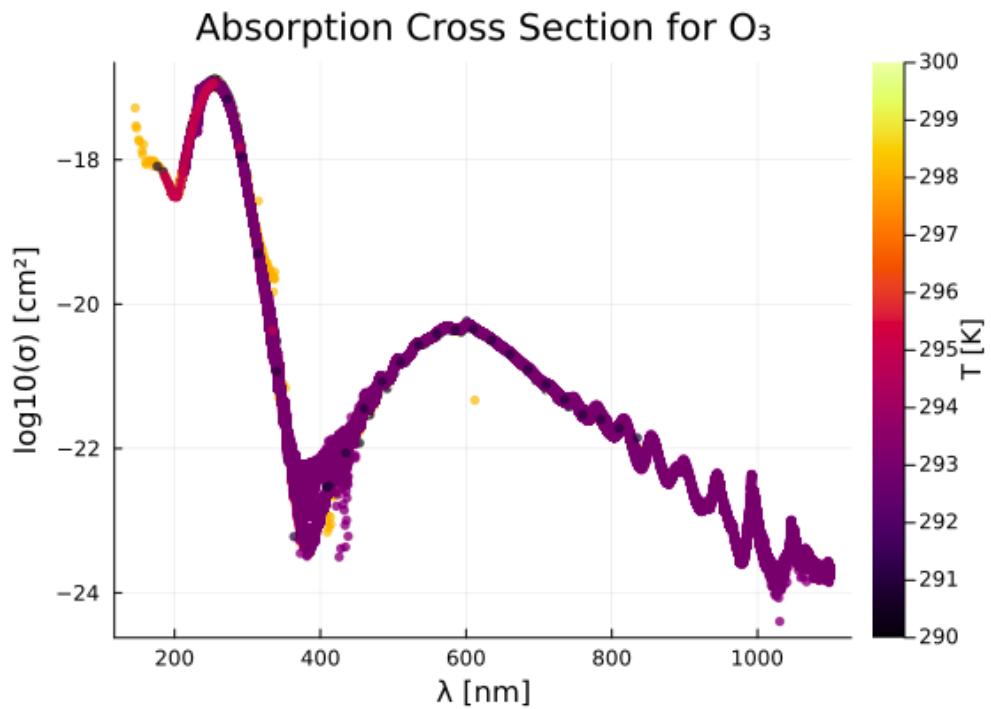


Figure 6.2: The collected absorption cross section data for Ozone, O<sub>3</sub>, across a variety of temperatures.

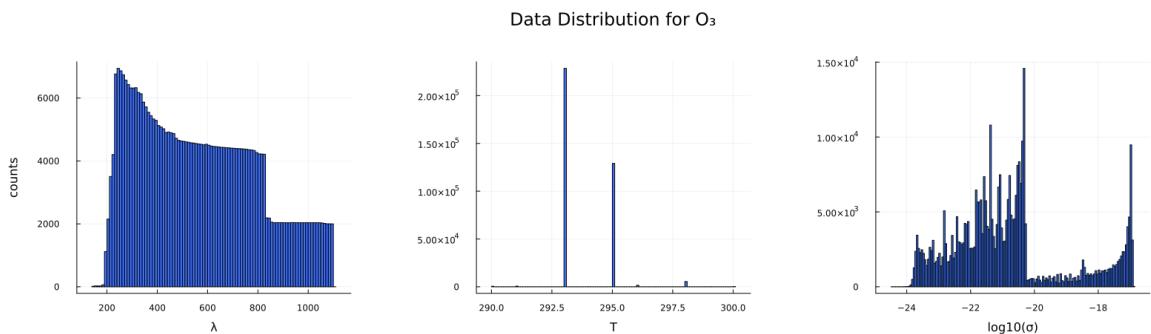


Figure 6.3: The distribution of cross section data available for O<sub>3</sub>.

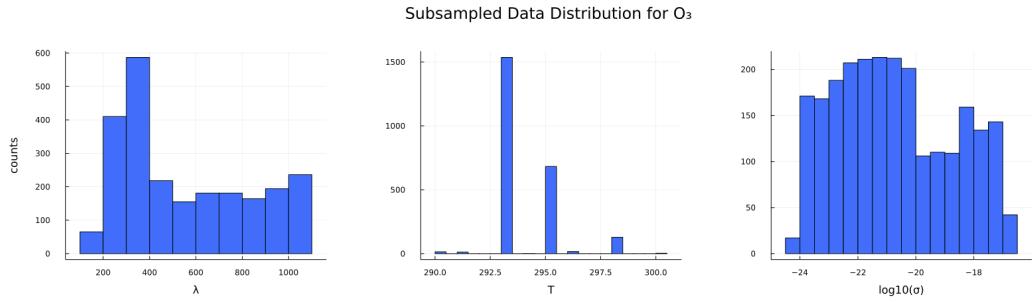


Figure 6.4: The distribution of subsampled O<sub>3</sub> cross section data

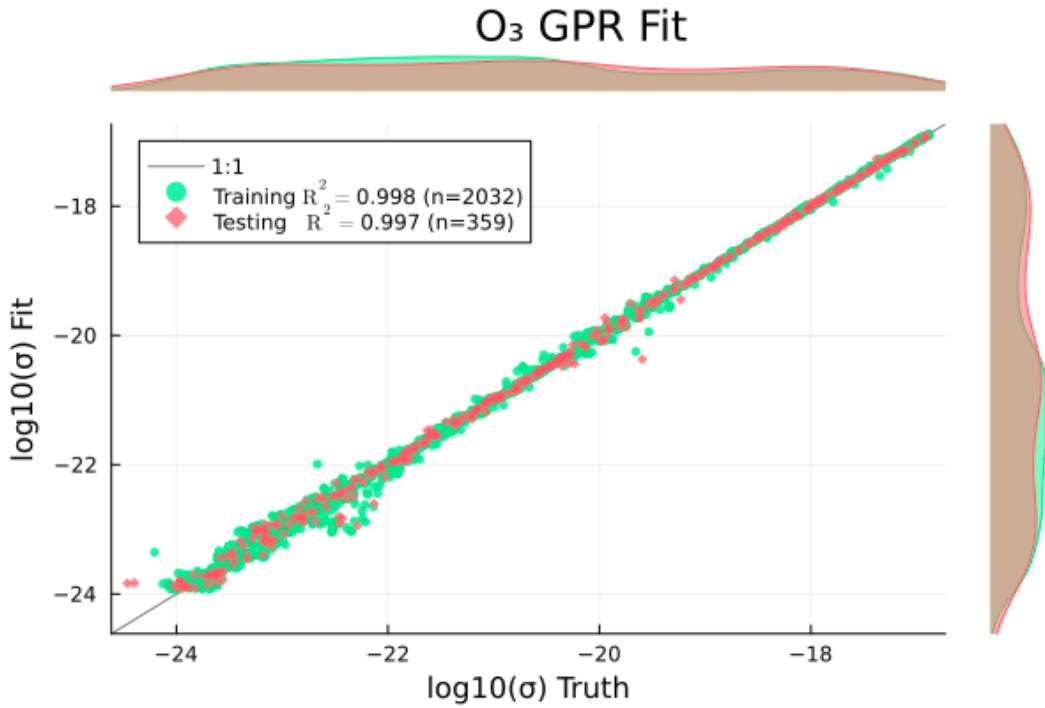


Figure 6.5: A scatter plot of the resulting GPR fit evaluated on the training set (green) and a holdout testing dataset (red).

## 6.4 Characterization of Photolysis

### 6.4.1 Absorption Cross Sections

### 6.4.2 Quantum Yields

### 6.4.3 Irradiance Spectra

Here we can discuss the collection of irradiance spectra using our UV-VIS Ocean Optics spectrometer. Importantly, we can discuss the correct unit conversion so as to obtain *photon fluxes* rather than just  $\mu$ Ws energy units.

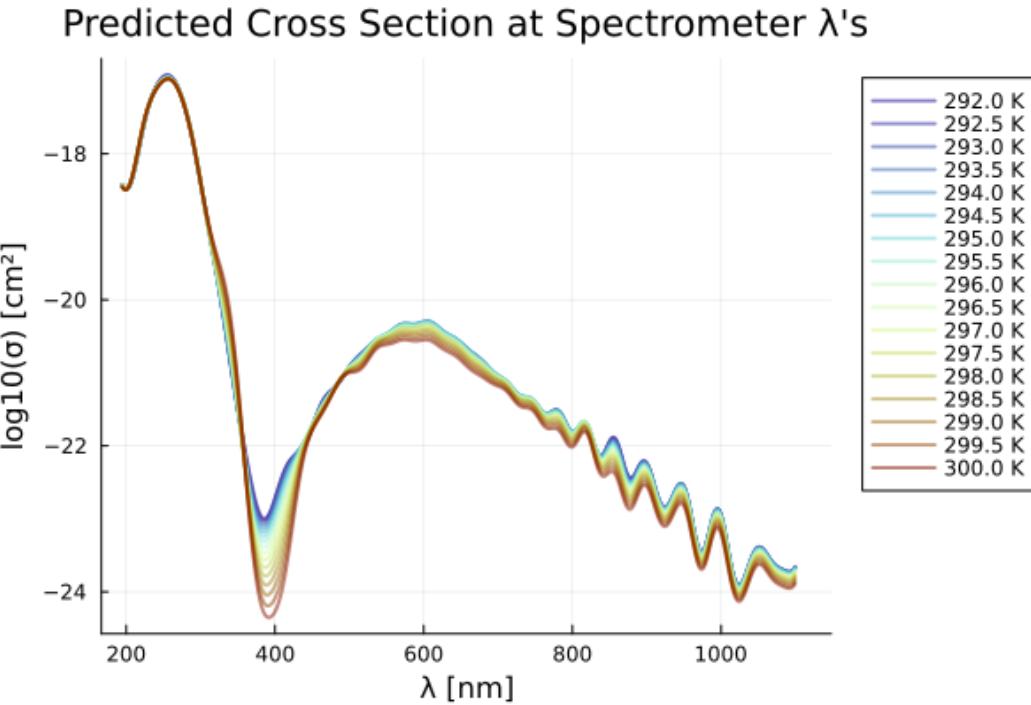


Figure 6.6: Predicted O<sub>3</sub> cross section as a function of temperature.

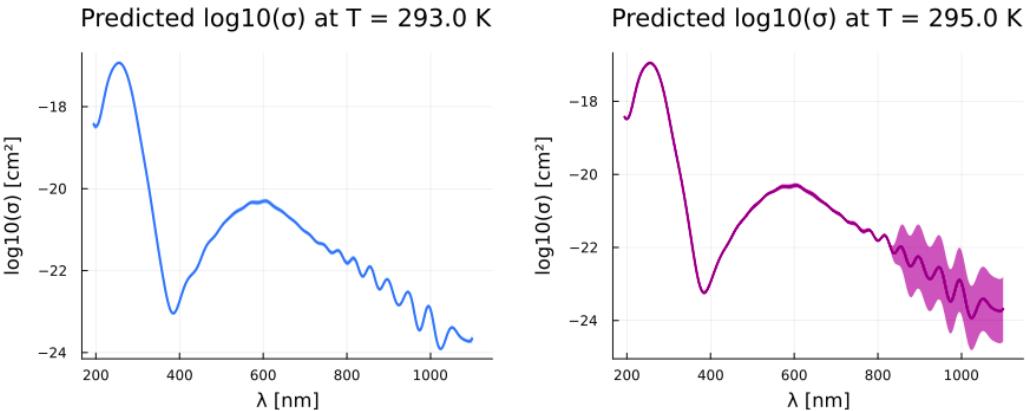


Figure 6.7: Predicted O<sub>3</sub> cross section for two temperatures with the associated uncertainty visualized. At  $T = 293$  K, we have many data points and therefore the associated fit uncertainty is small. For  $T = 295$  K, there are many fewer data points available above 800 nm leading to a larger fit uncertainty.

#### 6.4.4 Photolysis Rate Determination

Here we can discuss how we perform the numerical integration together with the captured intensity spectra

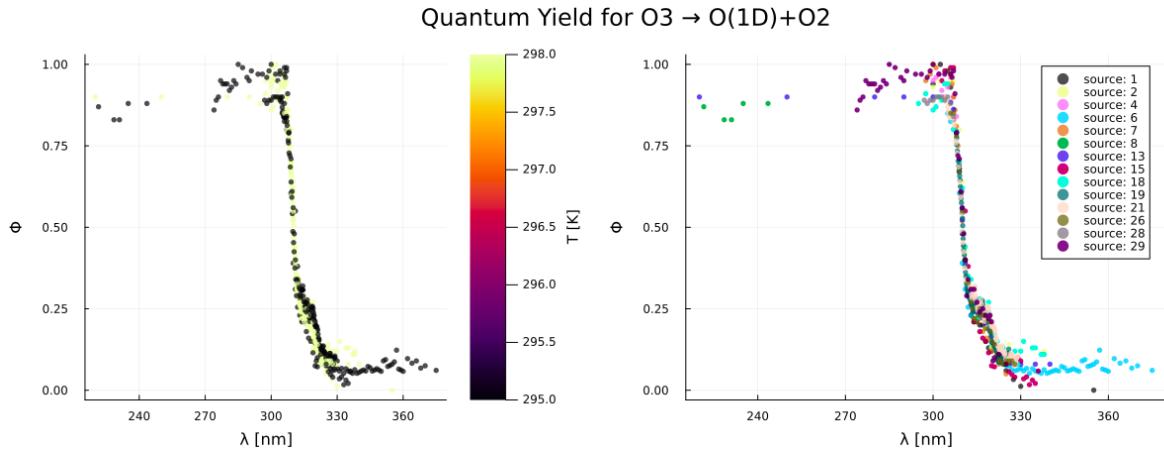


Figure 6.8: Data for the quantum yield of  $O_3$ . Sources of data are far more sparse than for the absorption cross section.

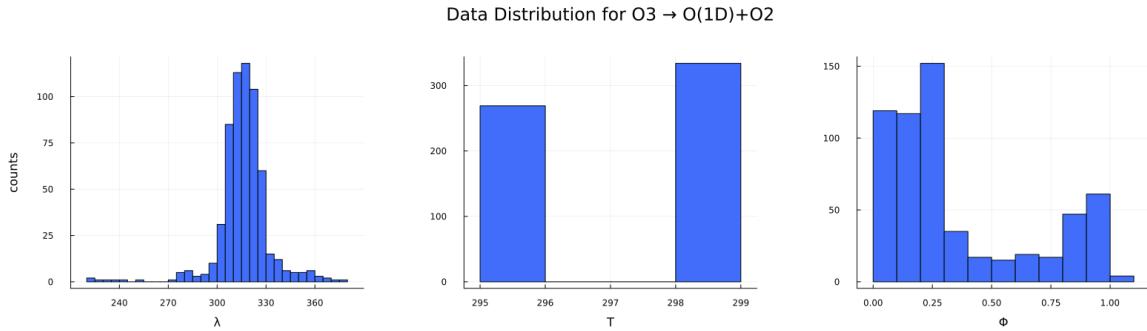


Figure 6.9: Distribution for the  $O_3$  quantum yield data.

#### 6.4.5 Next Steps

Discuss use of Conformal Prediction as a way to enable fits using different regressors (ideally XGBoost or some other boosted/bagged tree method) without sacrificing the necessary uncertainty estimates. Discuss the potential propagation of uncertainty

### 6.5 HEART Chamber Data

Initial data collected for evaluation of different PCO configurations.

The  $H_2O_2$  is a good demonstrate of a high uncertainty instrument close to the detection limits.

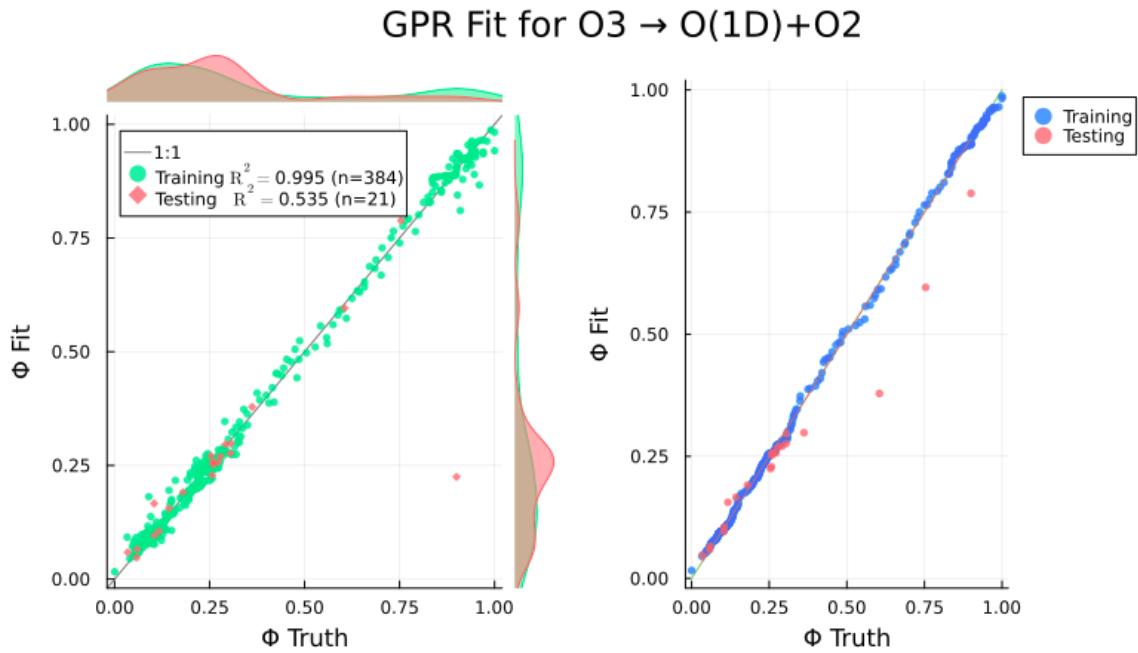


Figure 6.10: Resulting GPR fit for O<sub>3</sub> quantum yield data. Left: A scatter plot. Right: a quantile-quantile plot.

## 6.6 Chemical Data Assimilation

Discuss the mechanisms used. Also discuss the generation of pseudo observations with uncertainties determined by considering both the temporal variogram and the representativeness for the pseudo-observation period. We should probably also come up with some number of observations performed...

Here we can discuss the 2 stage procedure where we first use ADAM and then we finish with BFGS to prevent getting stuck in a local minimum.

EKF results

Add in numbers/plots for NO<sub>3</sub><sup>-</sup> ions from latest implementation of the mechanism.

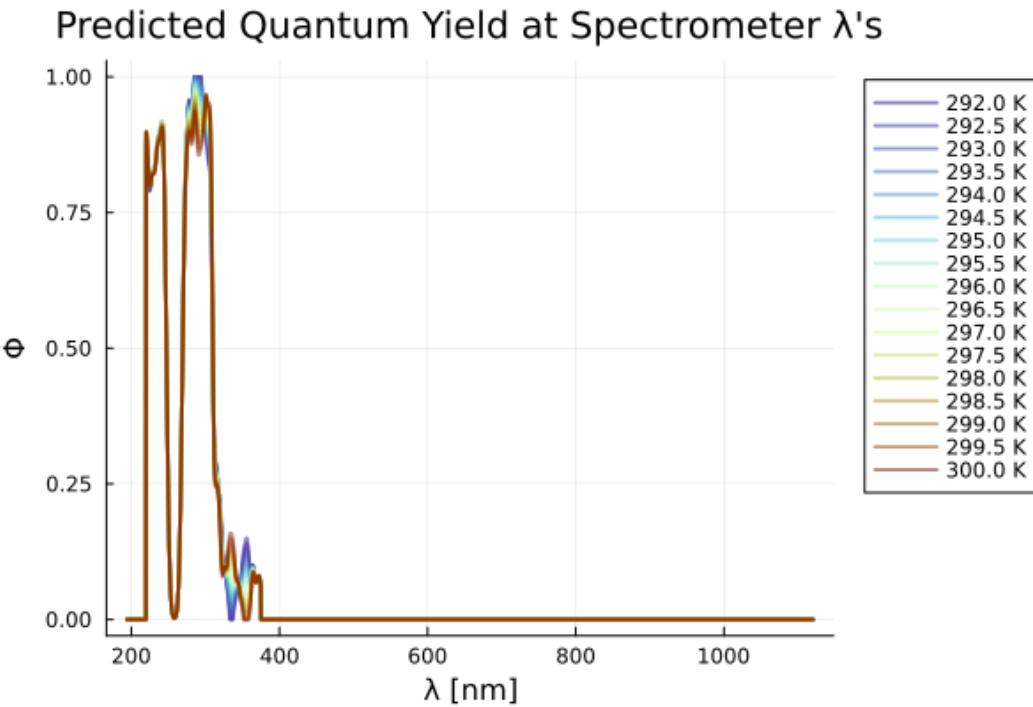


Figure 6.11: Predicted quantum yield as a function of temperature for  $O_3$ . Despite the goodness of fit, setting the mean to nearly zero to prevent edge effects causes the GPR to approach in wavelength regions with little data.

## 6.7 Next Steps

- Extend mechanism to include heterogeneous (mixed phase) reaction types
- Extend mechanism to include fixed rate surface deposition
- Extend mechanism to include direct generation of  $O_2^-$  at some rate (learned via 4d-var) for the PCO on portions
- Incorporate the cd-EKF and continue to refine integration techniques
- Perform fixed temperature/pressure long time integration tests per the Boldi thesis.
- Update reaction rates further with recent JPL Kinetic evaluations.

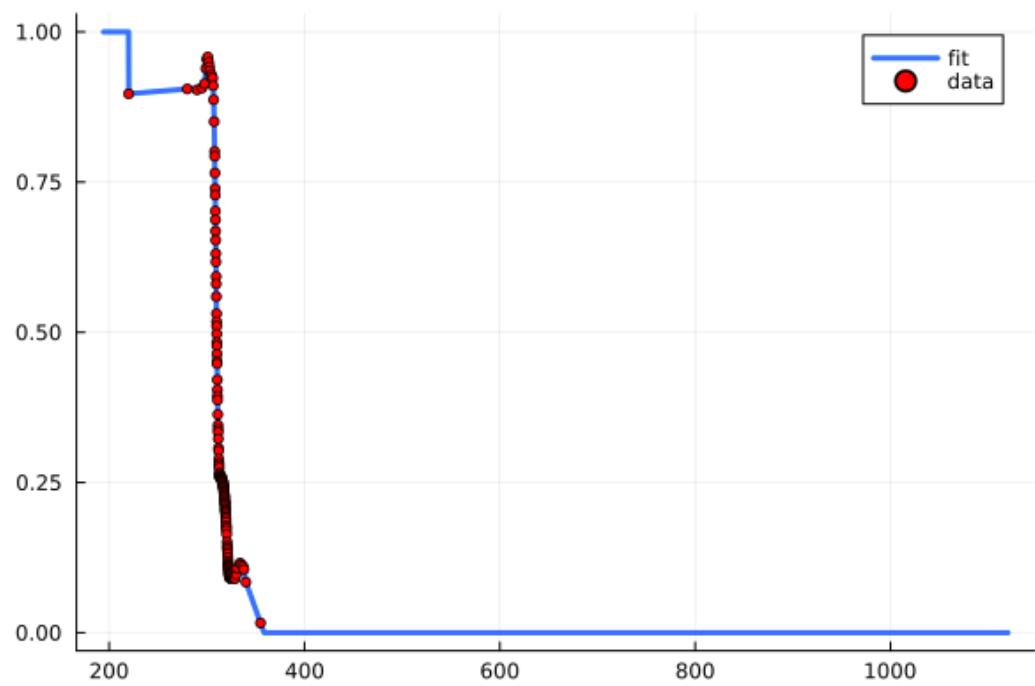


Figure 6.12: Final quantum yield predictions where we have forced the value to be 1 below a reasonable threshold wavelength.

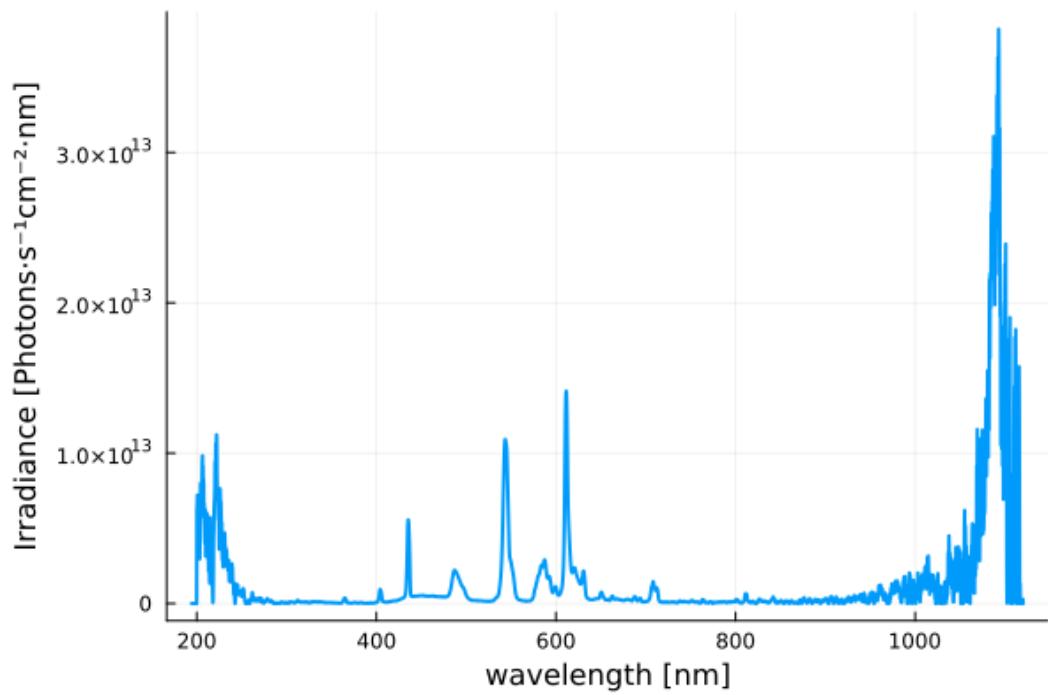


Figure 6.13: A plot of the spectral irradiance captured within a typical office building and averaged over 8 hours.

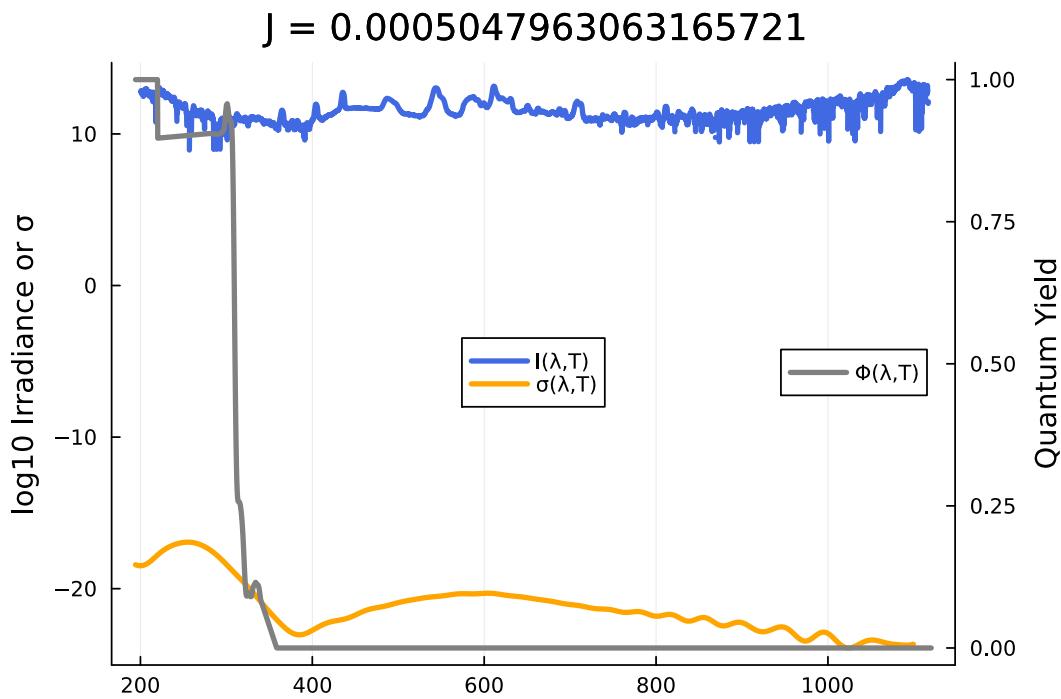
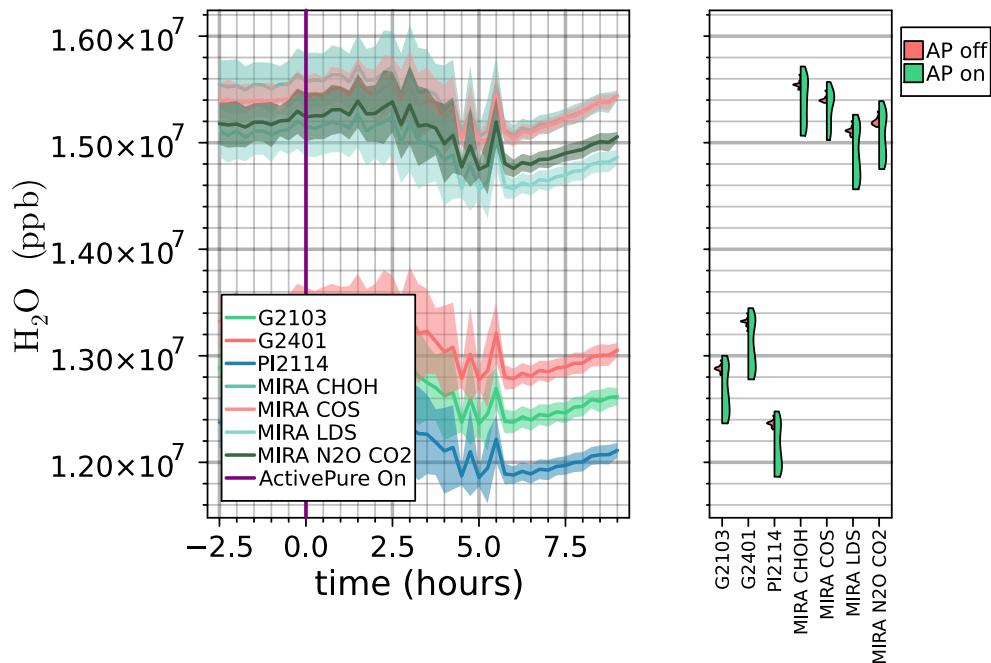
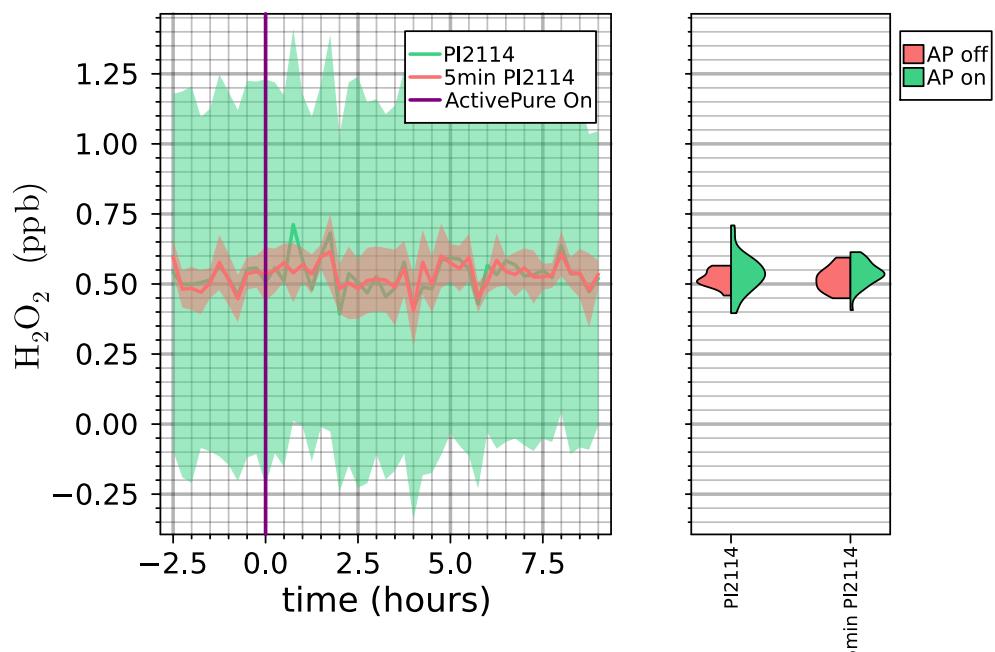
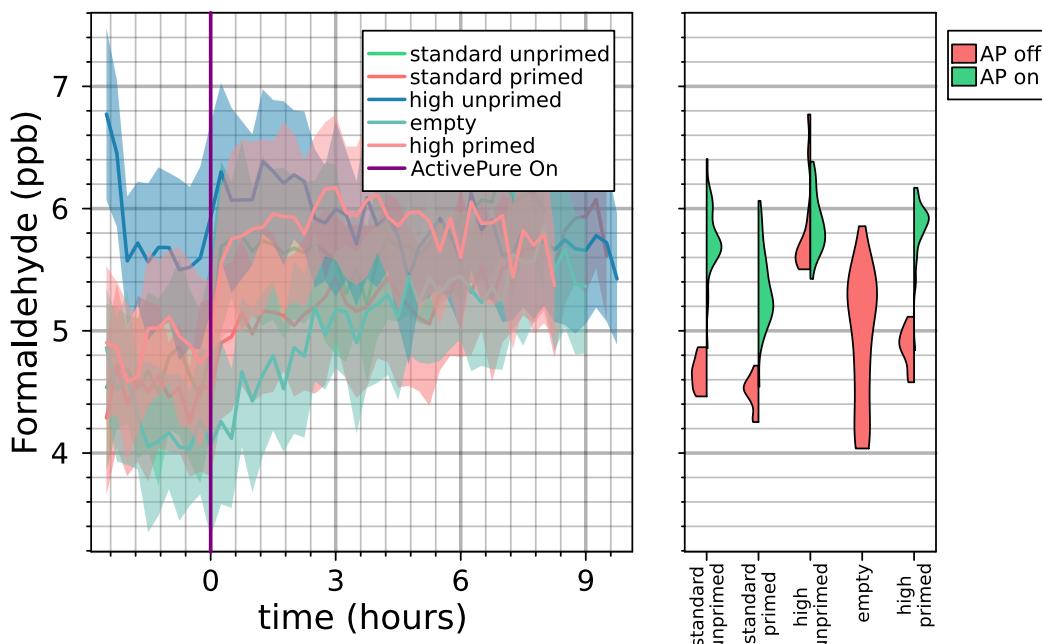
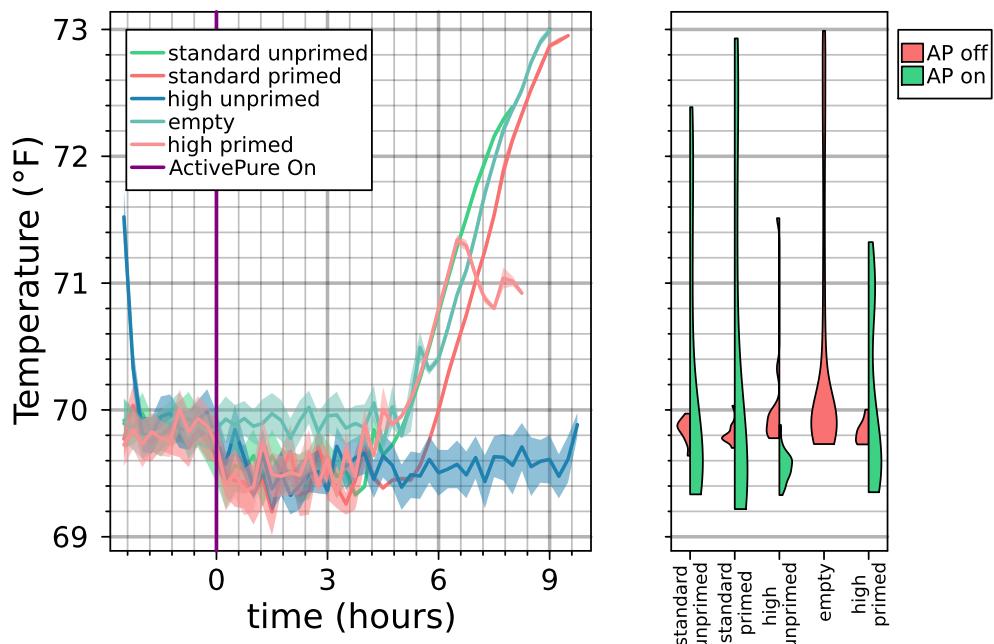
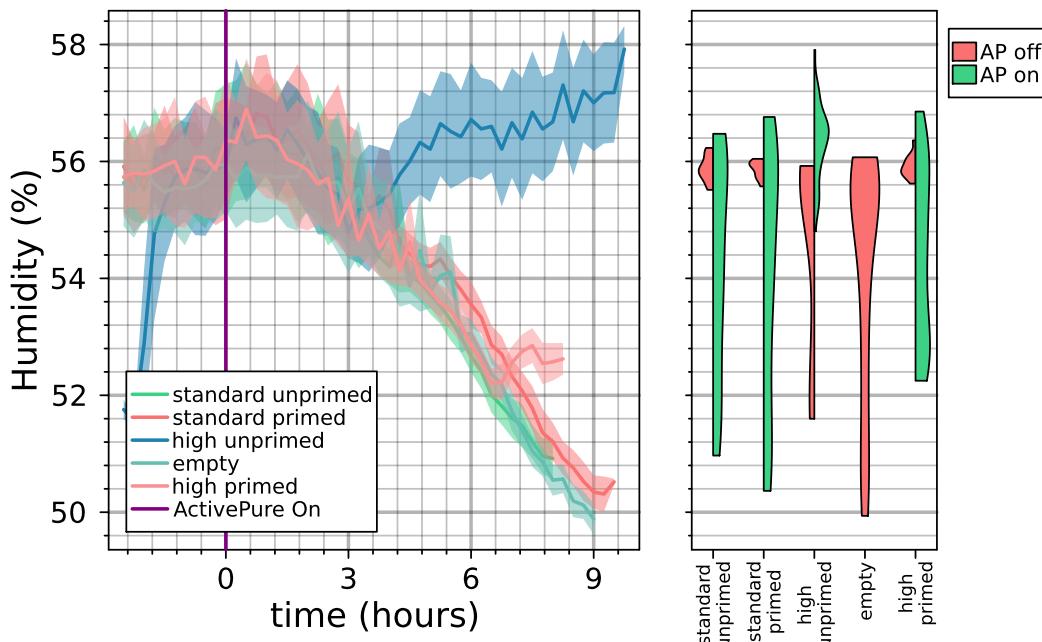
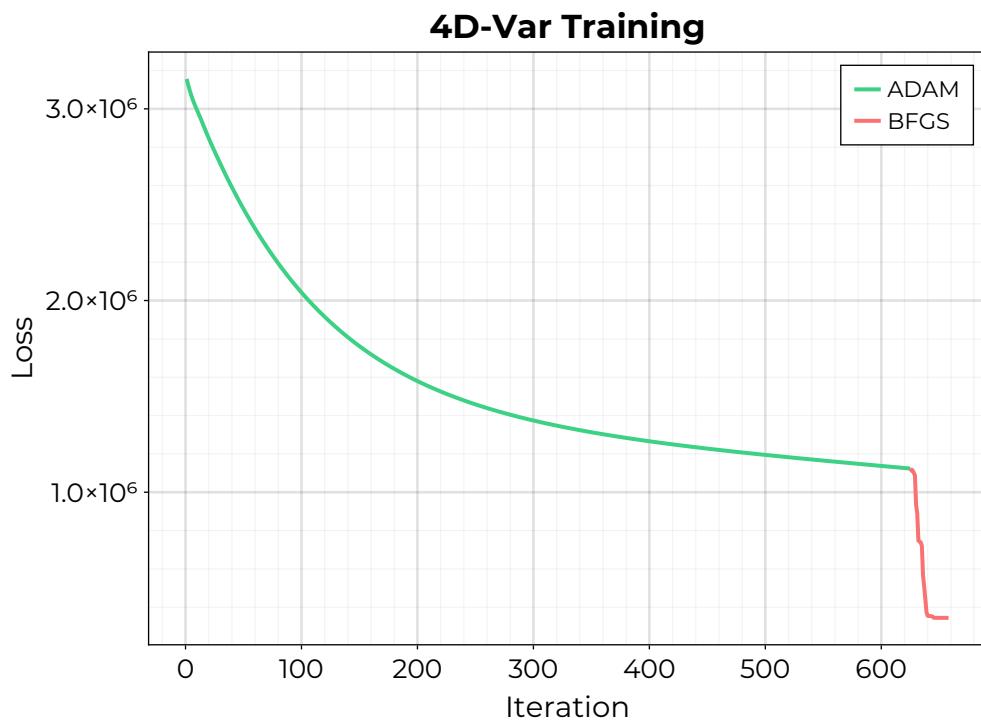
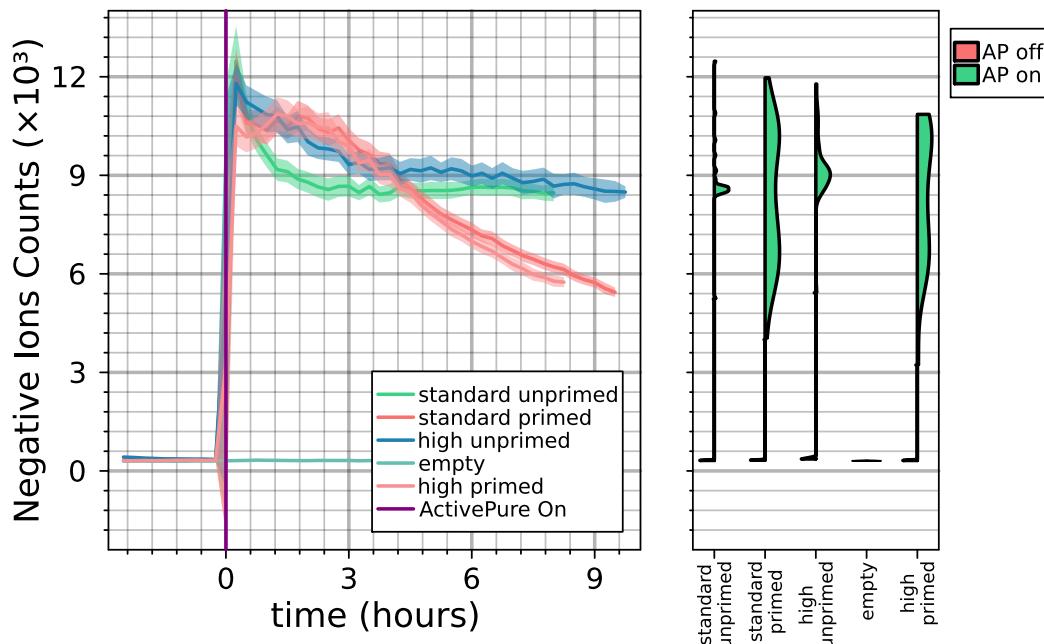


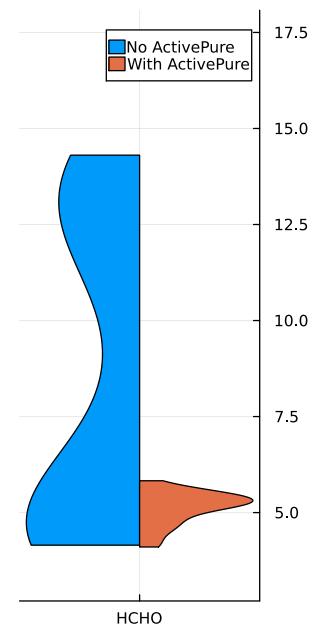
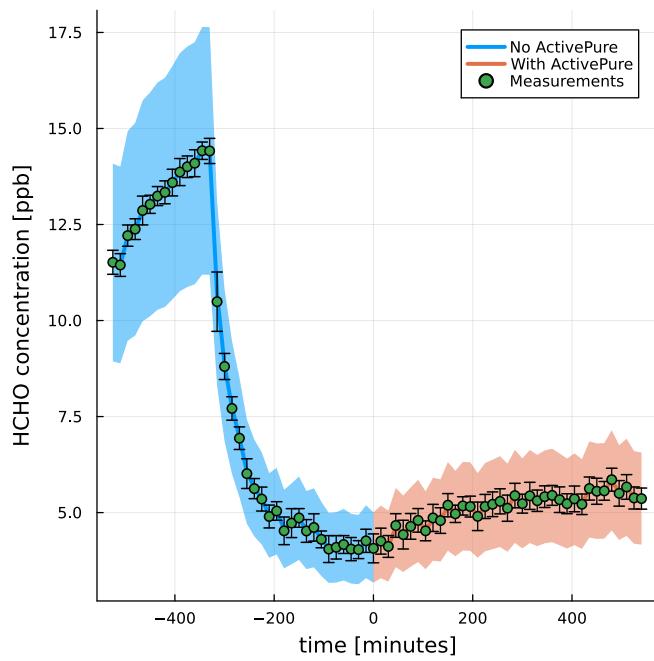
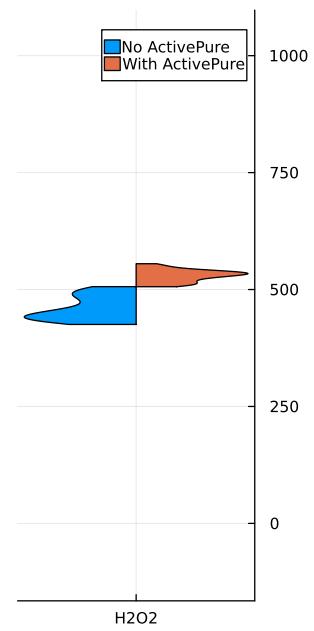
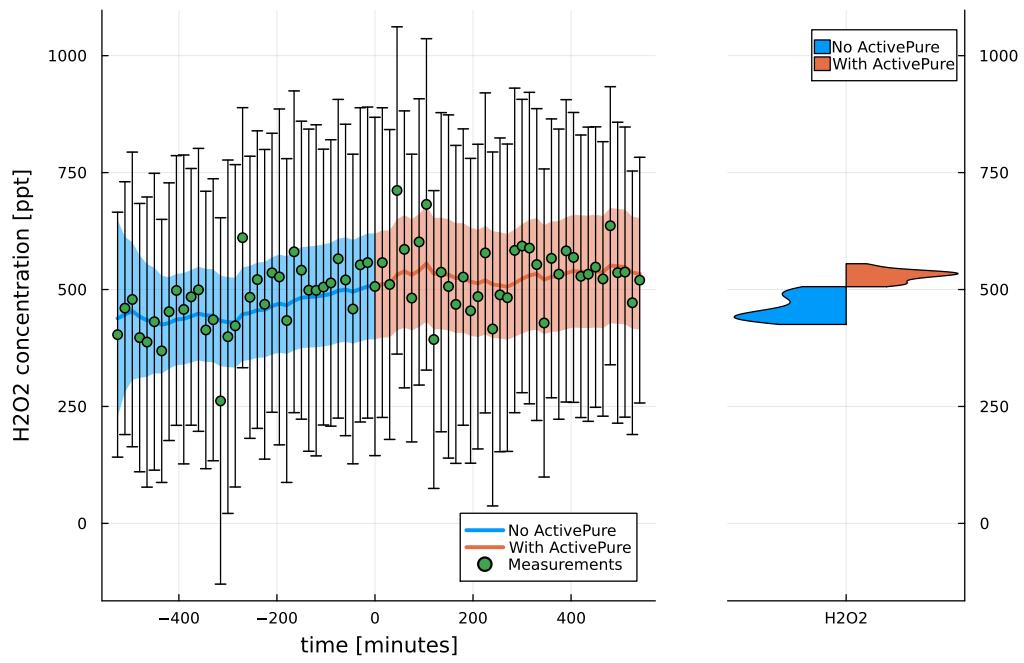
Figure 6.14: A comparison of the absorption cross section, quantum yield, and captured irradiance spectrum used to determine the photolysis rate for an O<sub>3</sub> photolysis channel.

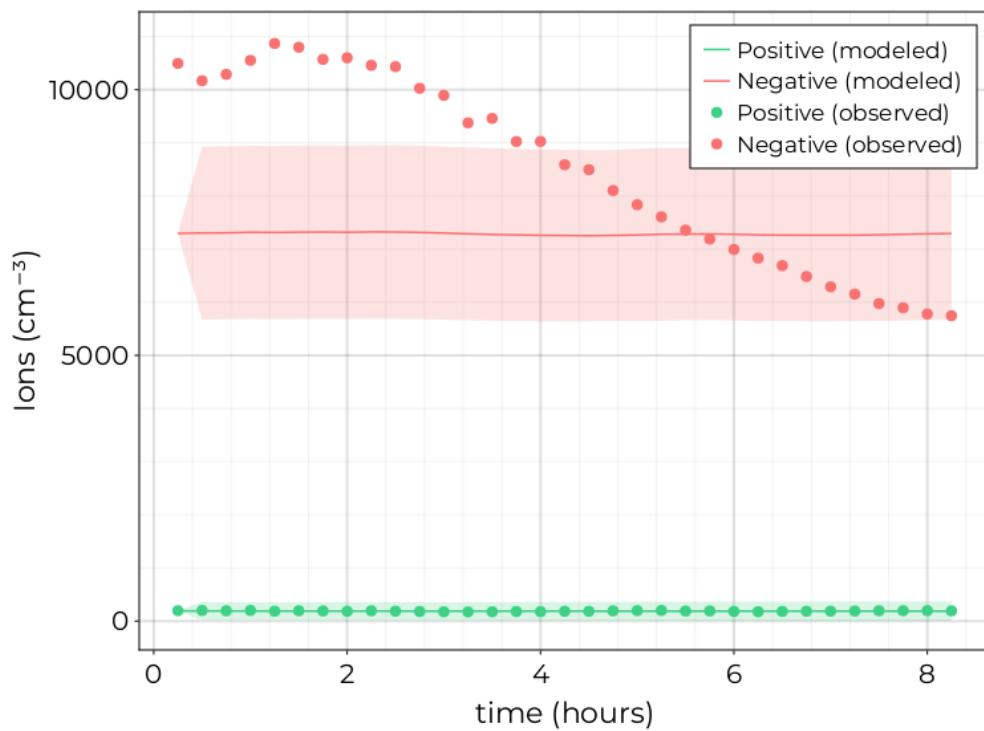
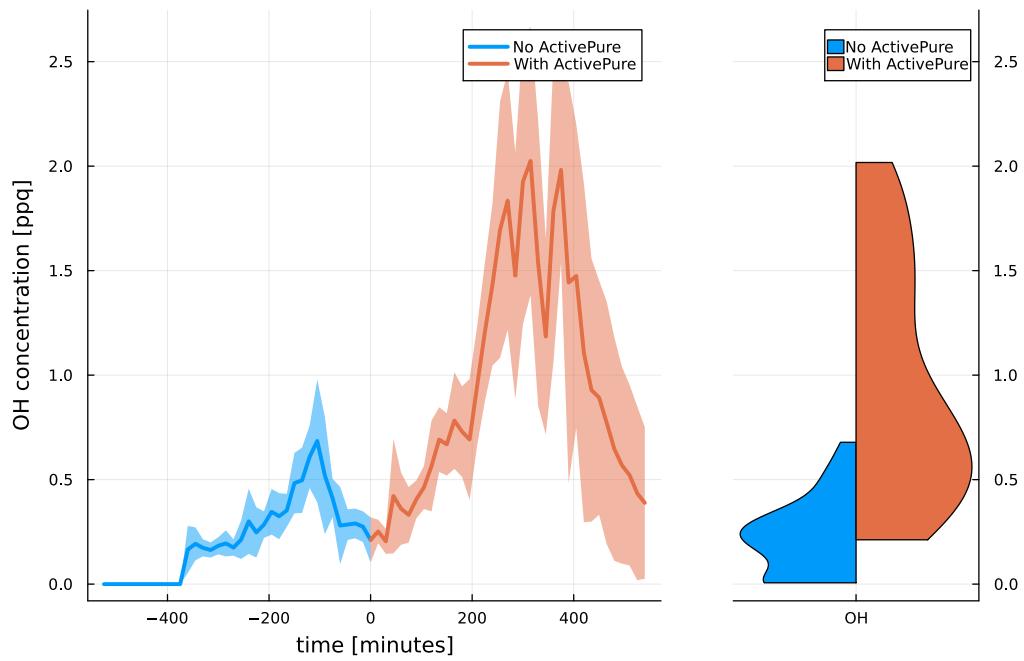












## CHAPTER 7

### LIMITATIONS AND FUTURE WORK

#### **7.1 Robot Team**

##### **7.1.1 Super Resolution**

Propose methods for spatial/temporal/spectral super-resolution. In particular, comment on upcoming satellite deployments like EnMAP and others which will provide hyperspectral imagery.

## CHAPTER 8

## CONCLUSIONS

**UPDATE REQUIRED!!!**

## APPENDIX A

### CHEMICAL REACTION MECHANISMS

UPDATE REQUIRED!

#### A.1 Simple Ion Mechanism

Here we can include the automatically generated documentation for the Ion Mechanism and any others.

##### A.1.1 Bimolecular Reactions

#	Bimolecular Reaction	Reaction Rate Coeff
1	$\text{Cl} + \text{HNO}_3 \longrightarrow \text{HCl} + \text{NO}_3$	$k = (2.0\text{e}-16)$
2	$\text{Br} + \text{H}_2 \longrightarrow \text{H} + \text{HBr}$	$k = (8.0\text{e}-11) \exp(-8970.0/T)(T/298.0)^{0.43}$
3	$\text{Br} + \text{O}_3 \longrightarrow \text{O}_2 + \text{BrO}$	$k = (1.7\text{e}-11) \exp(-800.0/T)$
4	$\text{Br} + \text{OH} \longrightarrow \text{O}({}^3\text{P}) + \text{HBr}$	$k = (4.56\text{e}-12) \exp(-8715.0/T)$

5	$\text{Br} + \text{CH}_4 \longrightarrow \text{HBr} + \text{CH}_3$	$k = (7.8\text{e-}11) \exp(-76330.0/T)$
6	$\text{Br} + \text{H}_2\text{O} \longrightarrow \text{OH} + \text{HBr}$	$k = 0$
7	$\text{Br} + \text{HCO} \longrightarrow \text{CO} + \text{HBr}$	$k = (2.8\text{e-}10) \exp(-800.0/T)$
8	$\text{Br} + \text{HO}_2 \longrightarrow \text{O}_2 + \text{HBr}$	$k = (1.5\text{e-}11) \exp(-600.0/T)$
9	$\text{Br} + \text{N}_2\text{O} \longrightarrow \text{N}_2 + \text{BrO}$	$k = (3.32\text{e-}10) \exp(-154800.0/T)$
10	$\text{Br} + \text{NO}_3 \longrightarrow \text{BrO} + \text{NO}_2$	$k = (1.6\text{e-}11)$
11	$\text{Br} + \text{H}_2\text{O}_2 \longrightarrow \text{HBr} + \text{HO}_2$	$k = (1.0\text{e-}11) \exp(-3000.0/T)$

12	$\text{Br} + \text{HCHO} \longrightarrow \text{HBr} + \text{HCO}$	$k = (1.7\text{e-}11) \exp(-800.0/T)$
13	$\text{Br} + \text{CH}_3\text{OH} \longrightarrow$ $\text{HBr} + \text{HO}_2 + \text{HCHO}$	$k = (7.59\text{e-}13) \exp(-3119.0/T)$
14	$\text{Br} + \text{OClO} \longrightarrow \text{BrO} + \text{ClO}$	$k = (2.6\text{e-}11) \exp(-1300.0/T)$
15	$\text{Br} + \text{Cl}_2\text{O}_2 \longrightarrow \text{BrCl} + \text{ClOO}$	$k = (3.0\text{e-}12)$
16	$\text{Br} + \text{CH}_3\text{OOH} \longrightarrow \text{HBr} + \text{CH}_3\text{O}_2$	$k = (2.63\text{e-}12) \exp(-1610.0/T)$
17	$\text{Br} + \text{BrONO}_2 \longrightarrow \text{Br}_2 + \text{NO}_3$	$k = (6.3\text{e-}11) \exp(213.0/T)$
18	$\text{BrO} + \text{BrO} \longrightarrow$ $\text{Br} + \text{O}_2 + \text{Br}$	$k = (9.23\text{e-}13) \exp(250.0/T)$

19	$\text{BrO} + \text{BrO} \longrightarrow \text{O}_2 + \text{Br}_2$	$k = (1.8\text{e-}13) \exp(250.0/T)$
20	$\text{BrO} + \text{ClO} \longrightarrow \text{Br} + \text{OClO}$	$k = (1.6\text{e-}12) \exp(430.0/T)$
21	$\text{BrO} + \text{ClO} \longrightarrow \text{Br} + \text{ClOO}$	$k = (2.9\text{e-}12) \exp(220.0/T)$
22	$\text{BrO} + \text{ClO} \longrightarrow \text{O}_2 + \text{BrCl}$	$k = (5.8\text{e-}13) \exp(170.0/T)$
23	$\text{BrO} + \text{HO}_2 \longrightarrow \text{O}_3 + \text{HBr}$	$k = \exp(500.0/T)$
24	$\text{BrO} + \text{HO}_2 \longrightarrow \text{O}_2 + \text{HOBr}$	$k = (3.4\text{e-}12) \exp(540.0/T)$
25	$\text{BrO} + \text{NO}_3 \longrightarrow$ $\text{Br} + \text{O}_2 + \text{NO}_2$	$k = (1.0\text{e-}12)$

26	$\text{BrO} + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{Br} + \text{CH}_3\text{O}$	$k = (3.23\text{e-}11) \exp(-332.0/T)$
27	$\text{CH}_3 + \text{H}_2\text{O} \longrightarrow \text{OH} + \text{CH}_4$	$k = (1.2\text{e-}14) \exp(-62190.0/T)(T/298.0)^{2.9}$
28	$\text{CH}_3 + \text{HCl} \longrightarrow \text{Cl} + \text{CH}_4$	$k = (3.88\text{e-}13) \exp(-9640.0/T)$
29	$\text{CH}_4 + \text{H}_2\text{O} \longrightarrow \text{CH}_3 + \text{H}_2\text{O}_2$	$k = (3.0\text{e-}13) \exp(-77740.0/T)$
30	$\text{CH}_4 + \text{HCO} \longrightarrow \text{CH}_3 + \text{HCHO}$	$k = (1.36\text{e-}13) \exp(-94200.0/T)(T/298.0)^{2.85}$
31	$\text{CH}_4 + \text{CH}_3\text{O} \longrightarrow \text{CH}_3 + \text{CH}_3\text{OH}$	$k = (2.6\text{e-}13) \exp(-37000.0/T)$

32	$\text{CH}_4 + \text{CH}_3\text{O}_2 \longrightarrow \text{CH}_3 + \text{CH}_3\text{OOH}$	$k = (3.0\text{e-}13) \exp(-77320.0/T)$
33	$\text{CO} + \text{HO}_2 \longrightarrow \text{OH} + \text{CO}_2$	$k = (2.5\text{e-}10) \exp(-98940.0/T)$
34	$\text{CO} + \text{CH}_3\text{O} \longrightarrow \text{CH}_3 + \text{CO}_2$	$k = (2.6\text{e-}11) \exp(-49390.0/T)$
35	$\text{CO} + \text{NO}_2 \longrightarrow \text{NO} + \text{CO}_2$	$k = (1.48\text{e-}10) \exp(-141420.0/T)$
36	$\text{CO} + \text{NO}_3 \longrightarrow \text{NO}_2 + \text{CO}_2$	$k = (4.0\text{e-}19)$
37	$\text{Cl} + \text{H}_2 \longrightarrow \text{H} + \text{HCl}$	$k = (3.7\text{e-}11) \exp(-2300.0/T)$
38	$\text{Cl} + \text{O}_3 \longrightarrow \text{O}_2 + \text{ClO}$	$k = (2.9\text{e-}11) \exp(-260.0/T)$

39	$\text{Cl} + \text{CH}_4 \longrightarrow \text{HCl} + \text{CH}_3$	$k = (1.1\text{e-}11) \exp(-1400.0/T)$
40	$\text{Cl} + \text{H}_2\text{O} \longrightarrow \text{OH} + \text{HCl}$	$k = (2.79\text{e-}11) \exp(-72080.0/T)$
41	$\text{Cl} + \text{HCl} \longrightarrow \text{H} + \text{Cl}_2$	$k = (1.66\text{e-}7) \exp(-198590.0/T)$
42	$\text{Cl} + \text{HO}_2 \longrightarrow \text{O}_2 + \text{HCl}$	$k = (1.8\text{e-}11) \exp(170.0/T)$
43	$\text{Cl} + \text{HO}_2 \longrightarrow \text{OH} + \text{ClO}$	$k = (4.1\text{e-}11) \exp(-450.0/T)$
44	$\text{Cl} + \text{N}_2\text{O} \longrightarrow \text{N}_2 + \text{ClO}$	$k = (2.16\text{e-}10) \exp(-140150.0/T)$
45	$\text{Cl} + \text{NO}_3 \longrightarrow \text{ClO} + \text{NO}_2$	$k = (2.4\text{e-}11)$

46	$\text{Cl} + \text{ClOO} \longrightarrow \text{ClO} + \text{ClO}$	$k = (1.2\text{e-}11)$
47	$\text{Cl} + \text{ClOO} \longrightarrow \text{O}_2 + \text{Cl}_2$	$k = (2.3\text{e-}10)$
48	$\text{Cl} + \text{H}_2\text{O}_2 \longrightarrow \text{HCl} + \text{HO}_2$	$k = (1.1\text{e-}11) \exp(-980.0/T)$
49	$\text{Cl} + \text{HCHO} \longrightarrow \text{HCl} + \text{HCO}$	$k = (8.2\text{e-}11) \exp(-34.0/T)$
50	$\text{Cl} + \text{HOCl} \longrightarrow \text{OH} + \text{Cl}_2$	$k = (2.5\text{e-}12) \exp(-130.0/T)$
51	$\text{Cl} + \text{CH}_3\text{O}_2 \longrightarrow \text{CH}_3\text{O} + \text{ClO}$	$k = (7.7\text{e-}11)$
52	$\text{Cl} + \text{OClO} \longrightarrow \text{ClO} + \text{ClO}$	$k = (3.4\text{e-}11) \exp(160.0/T)$

53	$\text{Cl} + \text{Cl}_2\text{O}_2 \longrightarrow \text{Cl}_2 + \text{ClOO}$	$k = (1.0\text{e}-10)$
54	$\text{Cl} + \text{CH}_3\text{OCl} \longrightarrow$ $\text{Cl}_2 + \text{HO}_2 + \text{HCHO}$	$k = (4.9\text{e}-11)$
55	$\text{Cl} + \text{CH}_3\text{OOH} \longrightarrow \text{HCl} + \text{CH}_3\text{O}_2$	$k = (5.9\text{e}-11)$
56	$\text{Cl} + \text{BrONO}_2 \longrightarrow \text{NO}_3 + \text{BrCl}$	$k = (2.0\text{e}-11) \exp(329.0/T)$
57	$\text{Cl} + \text{ClONO}_2 \longrightarrow \text{Cl}_2 + \text{NO}_3$	$k = (6.5\text{e}-12) \exp(135.0/T)$
58	$\text{Cl} + \text{CH}_3\text{ONO}_2 \longrightarrow$ $\text{NO}_2 + \text{HCl} + \text{HCHO}$	$k = (1.3\text{e}-11) \exp(-1200.0/T)$
59	$\text{Cl} + \text{CH}_3\text{O}_2\text{NO}_2 \longrightarrow$ $\text{NO}_3 + \text{HCl} + \text{HCHO}$	$k = (4.0\text{e}-13) \exp(-640.0/T)$

60	$\text{ClO} + \text{ClO} \longrightarrow \text{Cl} + \text{ClOO}$	$k = (3.0\text{e-}11) \exp(-2450.0/T)$
61	$\text{ClO} + \text{ClO} \longrightarrow \text{O}_2 + \text{Cl}_2$	$k = (1.0\text{e-}12) \exp(-1590.0/T)$
62	$\text{ClO} + \text{ClO} \longrightarrow \text{Cl} + \text{OClO}$	$k = (3.5\text{e-}13) \exp(-1370.0/T)$
63	$\text{ClO} + \text{HO}_2 \longrightarrow \text{O}_2 + \text{HOCl}$	$k = (4.8\text{e-}13) \exp(700.0/T)$
64	$\text{ClO} + \text{HO}_2 \longrightarrow \text{O}_3 + \text{HCl}$	$k = \exp(710.0/T)$
65	$\text{ClO} + \text{NO}_3 \longrightarrow \text{NO}_2 + \text{ClOO}$	$k = (4.7\text{e-}13)$
66	$\text{ClO} + \text{NO}_3 \longrightarrow \text{NO}_2 + \text{OClO}$	$k = 0$

67	$\text{ClO} + \text{HCHO} \longrightarrow \text{HCO} + \text{HOCl}$	$k = (1.0\text{e}-15)$
68	$\text{ClO} + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{CH}_3\text{OCl}$	$k = (2.6\text{e}-13) \exp(263.0/T)$
69	$\text{ClO} + \text{CH}_3\text{O}_2 \longrightarrow \text{CH}_3\text{O} + \text{ClOO}$	$k = (4.9\text{e}-12) \exp(-330.0/T)$
70	$\text{ClO} + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{Cl} + \text{CH}_3\text{O}$	$k = (4.91\text{e}-12) \exp(-332.0/T)$
71	$\text{H} + \text{O}_3 \longrightarrow \text{OH} + \text{O}_2$	$k = (1.4\text{e}-10) \exp(-480.0/T)$
72	$\text{H} + \text{Br}_2 \longrightarrow \text{Br} + \text{HBr}$	$k = (1.84\text{e}-10) \exp(-558.0/T)(T/298.0)^{0.5}$

73	$H + CH_4 \longrightarrow H_2 + CH_3$	$k = (5.82e-13) \exp(-33630.0/T)(T/298.0)^{3.0}$
74	$H + H_2O \longrightarrow OH + H_2$	$k = (6.83e-12) \exp(-80810.0/T)(T/298.0)^{1.6}$
75	$H + HCl \longrightarrow H_2 + Cl$	$k = (1.32e-11) \exp(-14220.0/T)$
76	$H + HO_2 \longrightarrow O(^3P) + H_2O$	$k = (2.4e-12)$
77	$H + HO_2 \longrightarrow OH + OH$	$k = (7.2e-11)$
78	$H + HO_2 \longrightarrow H_2 + O_2$	$k = (5.6e-12)$

79	$H + N_2O \longrightarrow N_2 + OH$	$k = (9.793e-11) \exp(-56640.0/T)$
80	$H + NO_2 \longrightarrow OH + NO$	$k = (4.0e-10) \exp(-340.0/T)$
81	$H + HONO \longrightarrow H_2 + NO_2$	$k = (2.0e-11) \exp(-3700.0/T)$
82	$H + CH_3Br \longrightarrow CH_3 + HBr$	$k = (4.34e-11) \exp(-2646.0/T)$
83	$H_2 + Br_2 \longrightarrow HBr + HBr$	$k = (6.81e-9) \exp(-20430.0/T)$
84	$H_2O + N_2O_5 \longrightarrow HNO_3 + HNO_3$	$k = (2.0e-21)$

85	$\text{H}_2\text{O} + \text{ClONO}_2 \longrightarrow \text{HOCl} + \text{HNO}_3$	$k = (2.0\text{e-}21)$
86	$\text{HCO} + \text{H}_2\text{O} \longrightarrow \text{OH} + \text{HCHO}$	$k = (8.54\text{e-}13) \exp(-109300.0/T)(T/298.0)^{1.35}$
87	$\text{HCO} + \text{HONO} \longrightarrow \text{NO}_2 + \text{HCHO}$	$k = (2.0\text{e-}21)T^{2.37} \exp(-1940.0/T)$
88	$\text{HCl} + \text{NO}_3 \longrightarrow \text{Cl} + \text{HNO}_3$	$k = (5.0\text{e-}17)$
89	$\text{HCl} + \text{N}_2\text{O}_5 \longrightarrow \text{HNO}_3 + \text{ClNO}_2$	$k = (6.97\text{e-}21)$
90	$\text{HCl} + \text{ClONO}_2 \longrightarrow \text{Cl}_2 + \text{HNO}_3$	$k = (1.0\text{e-}20)$

91	$\text{HCl} + \text{HO}_2\text{NO}_2 \longrightarrow \text{HOCl} + \text{HNO}_3$	$k = (1.0\text{e-}21)$
92	$\text{HO}_2 + \text{HO}_2 \longrightarrow \text{O}_2 + \text{H}_2\text{O}_2$	$k = (2.2\text{e-}13) \exp(600.0/T)(1.0 + (0.6*P/1013.25))$
93	$\text{HO}_2 + \text{NO}_3 \longrightarrow$ $\text{O}_2 + \text{OH} + \text{NO}_2$	$k = (2.15\text{e-}12)$
94	$\text{HO}_2 + \text{NO}_3 \longrightarrow \text{O}_2 + \text{HNO}_3$	$k = (2.15\text{e-}12)$
95	$\text{HO}_2 + \text{HCHO} \longrightarrow \text{HCO} + \text{H}_2\text{O}_2$	$k = (3.3\text{e-}12) \exp(-48800.0/T)$
96	$\text{HO}_2 + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{CH}_3\text{OOH}$	$k = (3.42\text{e-}13) \exp(780.0/T)$

97	$\text{HO}_2 + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{H}_2\text{O} + \text{HCHO}$	$k = (3.42\text{e-}14) \exp(780.0/T)$
98	$\text{CH}_3\text{O}_2 + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{CH}_3\text{O} + \text{CH}_3\text{O}$	$k = (3.67\text{e-}14) \exp(365.0/T)$
99	$\text{CH}_3\text{O}_2 + \text{CH}_3\text{O}_2 \longrightarrow \text{O}_2 + \text{CH}_3\text{OH} + \text{HCHO}$	$k = (6.6\text{e-}14) \exp(365.0/T)$
100	$\text{N} + \text{NO} \longrightarrow \text{O}({}^3\text{P}) + \text{N}_2$	$k = (3.1\text{e-}11)$
101	$\text{N} + \text{O}_2 \longrightarrow \text{O}({}^3\text{P}) + \text{NO}$	$k = (4.4\text{e-}12) \exp(-3220.0/T)$
102	$\text{N} + \text{O}_3 \longrightarrow \text{NO} + \text{O}_2$	$k = (1.0\text{e-}16)$
103	$\text{N} + \text{OH} \longrightarrow \text{H} + \text{NO}$	$k = (3.8\text{e-}11) \exp(85.0/T)$

104	$\text{N} + \text{NO}_2 \longrightarrow \text{O}({}^3\text{P}) + \text{N}_2\text{O}$	$k = (3.0\text{e-}12)$
105	$\text{NO} + \text{O}_3 \longrightarrow \text{O}_2 + \text{NO}_2$	$k = (1.8\text{e-}12) \exp(-1370.0/T)$
106	$\text{NO} + \text{BrO} \longrightarrow \text{Br} + \text{NO}_2$	$k = (8.8\text{e-}12) \exp(260.0/T)$
107	$\text{NO} + \text{ClO} \longrightarrow \text{Cl} + \text{NO}_2$	$k = (6.4\text{e-}12) \exp(290.0/T)$
108	$\text{NO} + \text{HO}_2 \longrightarrow \text{OH} + \text{NO}_2$	$k = (3.5\text{e-}12) \exp(250.0/T)$
109	$\text{NO} + \text{N}_2\text{O} \longrightarrow \text{N}_2 + \text{NO}_2$	$k = (2.51\text{e-}13) \exp(-206270.0/T)$
110	$\text{NO} + \text{NO}_3 \longrightarrow \text{NO}_2 + \text{NO}_2$	$k = (1.5\text{e-}11) \exp(170.0/T)$

111	$\text{NO} + \text{CH}_3\text{O}_2 \longrightarrow \text{CH}_3\text{O} + \text{NO}_2$	$k = (4.1\text{e-}12) \exp(180.0/T)$
112	$\text{NO} + \text{CH}_3\text{O}_2 \longrightarrow \text{CH}_3\text{ONO}_2$	$k = (4.1\text{e-}15) \exp(180.0/T)$
113	$\text{NO} + \text{OCLO} \longrightarrow \text{NO}_2 + \text{ClO}$	$k = (3.4\text{e-}13)$
114	$\text{NO} + \text{HNO}_3 \longrightarrow \text{NO}_2 + \text{HONO}$	$k = (7.43\text{e-}21)$
115	$\begin{aligned} \text{NO}_2 + \text{NO}_3 &\longrightarrow \\ \text{O}_2 + \text{NO} + \text{NO}_2 & \end{aligned}$	$k = (4.5\text{e-}14) \exp(-1260.0/T)$
116	$\text{NO}_3 + \text{HBr} \longrightarrow \text{Br} + \text{HNO}_3$	$k = (1.0\text{e-}16)$
117	$\text{NO}_3 + \text{HCHO} \longrightarrow \text{HCO} + \text{HNO}_3$	$k = (5.8\text{e-}16)$

118	$\text{NO}_3 + \text{CH}_3\text{OH} \longrightarrow \text{CH}_3\text{O} + \text{HNO}_3$	$k = (1.3\text{e-}12) \exp(-2560.0/T)$
119	$\text{NO}_3 + \text{CH}_3\text{O}_2 \longrightarrow$ $\text{O}_2 + \text{CH}_3\text{O} + \text{NO}_2$	$k = (1.0\text{e-}12)$
120	$\text{NO}_3 + \text{OClO} \longrightarrow$ $\text{O}_2 + \text{ClO} + \text{NO}_2$	$k = (2.0\text{e-}15)$
121	$\text{O(^1D)} + \text{CO} \longrightarrow \text{CO}_2$	$k = (7.3\text{e-}11)$
122	$\text{O(^1D)} + \text{H}_2 \longrightarrow \text{H} + \text{OH}$	$k = (1.1\text{e-}10)$
123	$\text{O(^1D)} + \text{N}_2 \longrightarrow \text{O(^3P)} + \text{N}_2$	$k = (1.8\text{e-}11) \exp(107.0/T)$
124	$\text{O(^1D)} + \text{NO} \longrightarrow \text{N} + \text{O}_2$	$k = (5.0\text{e-}11)$

125	$O(^1D) + O_2 \longrightarrow O(^3P) + O_2$	$k = (3.2\text{e-}11) \exp(67.0/T)$
126	$O(^1D) + O_3 \longrightarrow$ $O(^3P) + O(^3P) + O_2$	$k = (1.2\text{e-}10)$
127	$O(^1D) + O_3 \longrightarrow O_2 + O_2$	$k = (1.2\text{e-}10)$
128	$O(^1D) + CH_4 \longrightarrow OH + CH_3$	$k = (1.13\text{e-}10)$
129	$O(^1D) + CH_4 \longrightarrow H_2 + HCHO$	$k = (7.5\text{e-}12)$
130	$O(^1D) + CH_4 \longrightarrow$ $H + H + HCHO$	$k = (3.0\text{e-}11)$
131	$O(^1D) + CO_2 \longrightarrow O(^3P) + CO_2$	$k = (7.4\text{e-}11) \exp(120.0/T)$

132	$O(^1D) + Cl_2 \longrightarrow O(^3P) + Cl_2$	$k = (7.0e-11)$
133	$O(^1D) + Cl_2 \longrightarrow Cl + ClO$	$k = (2.1e-10)$
134	$O(^1D) + H_2O \longrightarrow OH + OH$	$k = (2.2e-10)$
135	$O(^1D) + HBr \longrightarrow O(^3P) + HBr$	$k = (3.0e-11)$
136	$O(^1D) + HBr \longrightarrow OH + Br$	$k = (1.5e-10)$
137	$O(^1D) + HCl \longrightarrow O(^3P) + HCl$	$k = (1.35e-11)$
138	$O(^1D) + HCl \longrightarrow H + ClO$	$k = (3.6e-11)$

139	$O(^1D) + HCl \longrightarrow OH + Cl$	$k = (1.0e-10)$
140	$O(^1D) + N_2O \longrightarrow N + NO_2$	$k = (2.45e-13)$
141	$O(^1D) + N_2O \longrightarrow NO + NO$	$k = (7.2e-11)$
142	$O(^1D) + N_2O \longrightarrow N_2 + O_2$	$k = (4.4e-11)$
143	$O(^1D) + CF_2Cl_2 \longrightarrow Cl + Cl$	$k = (1.4e-10)$
144	$O(^3P) + H_2 \longrightarrow H + OH$	$k = (9.0e-18)$
145	$O(^3P) + O_3 \longrightarrow O_2 + O_2$	$k = (8.0e-12) \exp(-2060.0/T)$

146	$O(^3P) + OH \longrightarrow H + O_2$	$k = (2.3e-11) \exp(110.0/T)$
147	$O(^3P) + Br_2 \longrightarrow Br + BrO$	$k = (1.4e-11)$
148	$O(^3P) + BrO \longrightarrow O_2 + Br$	$k = (1.9e-11) \exp(230.0/T)$
149	$O(^3P) + CH_3 \longrightarrow H + HCHO$	$k = (1.4e-10)$
150	$O(^3P) + CH_4 \longrightarrow OH + CH_3$	$k = (8.32e-12) \exp(-35500.0/T)(T/298.0)^{1.56}$
151	$O(^3P) + ClO \longrightarrow Cl + O_2$	$k = (3.0e-11) \exp(70.0/T)$
152	$O(^3P) + H_2O \longrightarrow OH + OH$	$k = (1.85e-11) \exp(-71260.0/T)(T/298.0)^{0.946}$

153	$O(^3P) + HBr \longrightarrow OH + Br$	$k = (6.6\text{e-}12) \exp(-1540.0/T)$
154	$O(^3P) + HCl \longrightarrow OH + Cl$	$k = (1.0\text{e-}11) \exp(-3340.0/T)$
155	$O(^3P) + HO_2 \longrightarrow OH + O_2$	$k = (2.7\text{e-}11) \exp(224.0/T)$
156	$O(^3P) + NO_2 \longrightarrow O_2 + NO$	$k = (6.5\text{e-}12) \exp(120.0/T)$
157	$O(^3P) + NO_3 \longrightarrow O_2 + NO_2$	$k = (1.7\text{e-}11)$
158	$O(^3P) + BrCl \longrightarrow Cl + BrO$	$k = (2.2\text{e-}11)$
159	$O(^3P) + H_2O_2 \longrightarrow OH + HO_2$	$k = (1.4\text{e-}12) \exp(-2000.0/T)$

160	$O(^3P) + HCHO \longrightarrow OH + HCO$	$k = (3.4e-11) \exp(-1600.0/T)$
161	$O(^3P) + HOBr \longrightarrow OH + BrO$	$k = (1.4e-10) \exp(-430.0/T)$
162	$O(^3P) + HOCl \longrightarrow OH + ClO$	$k = (1.7e-13)$
163	$O(^3P) + HONO \longrightarrow OH + NO_2$	$k = (2.0e-11) \exp(-3000.0/T)$
164	$O(^3P) + OClO \longrightarrow O_2 + ClO$	$k = (2.5e-12) \exp(-950.0/T)$
165	$O(^3P) + CH_3Br \longrightarrow OH + Br$	$k = (7.6e-13) \exp(-890.0/T)$
166	$O(^3P) + HNO_3 \longrightarrow OH + NO_3$	$k = (3.0e-17)$

167	$O(^3P) + ClONO_2 \longrightarrow NO_2 + OClO$	$k = (1.5e-12) \exp(-800.0/T)$
168	$O(^3P) + ClONO_2 \longrightarrow NO_3 + ClO$	$k = (1.5e-12) \exp(-800.0/T)$
169	$O(^3P) + HO_2NO_2 \longrightarrow OH + O_2 + NO_2$	$k = (7.8e-11) \exp(-3400.0/T)$
170	$O_2 + CH_3 \longrightarrow OH + HCHO$	$k = (3.0e-16)$
171	$O_2 + CH_4 \longrightarrow CH_3 + HO_2$	$k = (6.7e-11) \exp(-238120.0/T)(T/298.0)^{1.56}$
172	$O_2 + HCO \longrightarrow CO + HO_2$	$k = (5.5e-12)$

173	$O_2 + CH_3O \longrightarrow HO_2 + HCHO$	$k = (7.2e-14) \exp(-1080.0/T)$
174	$O_3 + BrO \longrightarrow$ $Br + O_2 + O_2$	$k = (5.0e-15)$
175	$O_3 + CH_3 \longrightarrow O_2 + CH_3O$	$k = (5.1e-12) \exp(-210.0/T)$
176	$O_3 + CH_3 \longrightarrow$ $H + O_2 + HCHO$	$k = (5.1e-12) \exp(-210.0/T)$
177	$O_3 + ClO \longrightarrow O_2 + ClOO$	$k = (1.5e-17)$
178	$O_3 + ClO \longrightarrow O_2 + OCLO$	$k = (1.0e-18) \exp(-4000.0/T)$
179	$O_3 + HO_2 \longrightarrow$ $OH + O_2 + O_2$	$k = (1.4e-14) \exp(-600.0/T)$

180	$O_3 + NO_2 \longrightarrow O_2 + NO_3$	$k = (1.2e-13) \exp(-2450.0/T)$
181	$O_3 + HONO \longrightarrow O_2 + HNO_3$	$k = (5.0e-19)$
182	$O_3 + Cl_2O_2 \longrightarrow$ $O_2 + ClO + ClOO$	$k = (1.0e-19)$
183	$OH + CO \longrightarrow H + CO_2$	$k = (1.5e-13)(1.0 + (0.6 * P/1013.25))$
184	$OH + H_2 \longrightarrow H + H_2O$	$k = (5.0e-12) \exp(-2000.0/T)$
185	$OH + O_3 \longrightarrow O_2 + HO_2$	$k = (1.9e-12) \exp(-1000.0/T)$
186	$OH + OH \longrightarrow O(^3P) + H_2O$	$k = (4.2e-12) \exp(-240.0/T)$

187	$\text{OH} + \text{Br}_2 \longrightarrow \text{Br} + \text{HOBr}$	$k = (1.2\text{e-}11) \exp(400.0/T)$
188	$\text{OH} + \text{BrO} \longrightarrow \text{Br} + \text{HO}_2$	$k = (7.425\text{e-}11)$
189	$\text{OH} + \text{BrO} \longrightarrow \text{O}_2 + \text{HBr}$	$k = (7.5\text{e-}13)$
190	$\text{OH} + \text{CH}_4 \longrightarrow \text{H}_2\text{O} + \text{CH}_3$	$k = (2.8\text{e-}14)T^{0.667} \exp(-1575.0/T)$
191	$\text{OH} + \text{Cl}_2 \longrightarrow \text{Cl} + \text{HOCl}$	$k = (1.4\text{e-}12) \exp(-900.0/T)$
192	$\text{OH} + \text{ClO} \longrightarrow \text{Cl} + \text{HO}_2$	$k = (9.35\text{e-}12) \exp(120.0/T)$
193	$\text{OH} + \text{ClO} \longrightarrow \text{O}_2 + \text{HCl}$	$k = (1.65\text{e-}12) \exp(120.0/T)$

194	$\text{OH} + \text{HBr} \longrightarrow \text{Br} + \text{H}_2\text{O}$	$k = (1.1\text{e-}11)$
195	$\text{OH} + \text{HCl} \longrightarrow \text{Cl} + \text{H}_2\text{O}$	$k = (2.6\text{e-}12) \exp(-350.0/T)$
196	$\text{OH} + \text{HO}_2 \longrightarrow \text{O}_2 + \text{H}_2\text{O}$	$k = (4.8\text{e-}11) \exp(250.0/T)$
197	$\text{OH} + \text{N}_2\text{O} \longrightarrow \text{N}_2 + \text{HO}_2$	$k = (3.8\text{e-}17)$
198	$\text{OH} + \text{NO}_3 \longrightarrow \text{HO}_2 + \text{NO}_2$	$k = (2.3\text{e-}11)$
199	$\text{OH} + \text{H}_2\text{O}_2 \longrightarrow \text{H}_2\text{O} + \text{HO}_2$	$k = (2.9\text{e-}12) \exp(-160.0/T)$
200	$\text{OH} + \text{HCHO} \longrightarrow \text{H}_2\text{O} + \text{HCO}$	$k = (8.8\text{e-}12) \exp(25.0/T)$

201	$\text{OH} + \text{HOCl} \longrightarrow \text{ClO} + \text{H}_2\text{O}$	$k = (3.0\text{e-}12) \exp(-500.0/T)$
202	$\text{OH} + \text{HONO} \longrightarrow \text{H}_2\text{O} + \text{NO}_2$	$k = (1.8\text{e-}11) \exp(-390.0/T)$
203	$\text{OH} + \text{CH}_3\text{OH} \longrightarrow \text{H}_2\text{O} + \text{CH}_3\text{O}$	$k = (5.0\text{e-}13) \exp(-380.0/T)$
204	$\text{OH} + \text{OCLO} \longrightarrow \text{O}_2 + \text{HOCl}$	$k = (4.5\text{e-}13) \exp(800.0/T)$
205	$\text{OH} + \text{ClNO}_2 \longrightarrow \text{NO}_2 + \text{HOCl}$	$k = (3.5\text{e-}14)$
206	$\text{OH} + \text{HNO}_3 \longrightarrow \text{H}_2\text{O} + \text{NO}_3$	$k = 2.4\text{e-}14 \cdot \exp(460.0/T) + \frac{6.50\text{e-}34 \cdot M \cdot \exp(1335.0/T)}{1 + \frac{6.50\text{e-}34 \cdot M \cdot \exp(1335.0/T)}{2.7\text{e-}17 \cdot \exp(2199.0/T)}}$
207	$\text{OH} + \text{CH}_3\text{OCl} \longrightarrow \text{Cl} + \text{H}_2\text{O} + \text{HCHO}$	$k = (2.4\text{e-}12) \exp(-360.0/T)$

208	$\text{OH} + \text{CH}_3\text{OOH} \longrightarrow \text{H}_2\text{O} + \text{CH}_3\text{O}_2$	$k = (1.9\text{e-}12) \exp(190.0/T)$
209	$\text{OH} + \text{CH}_3\text{OOH} \longrightarrow$ $\text{OH} + \text{H}_2\text{O} + \text{HCHO}$	$k = (1.0\text{e-}12) \exp(190.0/T)$
210	$\text{OH} + \text{ClONO}_2 \longrightarrow \text{NO}_3 + \text{HOCl}$	$k = (1.2\text{e-}12) \exp(-330.0/T)$
211	$\text{OH} + \text{HO}_2\text{NO}_2 \longrightarrow$ $\text{O}_2 + \text{H}_2\text{O} + \text{NO}_2$	$k = (1.5\text{e-}12) \exp(360.0/T)$
212	$\text{OH} + \text{CH}_3\text{ONO}_2 \longrightarrow$ $\text{NO}_2 + \text{H}_2\text{O} + \text{HCHO}$	$k = (4.0\text{e-}13) \exp(-845.0/T)$
213	$\text{OH} + \text{CH}_3\text{O}_2\text{NO}_2 \longrightarrow$ $\text{NO}_3 + \text{H}_2\text{O} + \text{HCHO}$	$k = (1.0\text{e-}13) \exp(-640.0/T)$

214	$\text{N}_2^+ + \text{O}({}^3\text{P}) \longrightarrow \text{NO}^+ + \text{N}$	$k = (1.4\text{e-}10)$
215	$\text{N}_2^+ + \text{O}_2 \longrightarrow \text{O}_2^+ + \text{N}_2$	$k = (6.0\text{e-}11)$
216	$\text{N}_2^+ + \text{NO} \longrightarrow \text{NO}^+ + \text{N}_2$	$k = (5.0\text{e-}10)$
217	$\text{N}_2^+ + \text{CO}_2 \longrightarrow \text{CO}_2^+ + \text{N}_2$	$k = (7.7\text{e-}10)$
218	$\text{N}_2^+ + \text{CO} \longrightarrow \text{CO}^+ + \text{N}_2$	$k = (7.4\text{e-}11)$
219	$\text{N}^+ + \text{O}_2 \longrightarrow \text{O}_2^+ + \text{N}$	$k = (2.8\text{e-}10)$
220	$\text{N}^+ + \text{NO} \longrightarrow \text{NO}^+ + \text{N}$	$k = (8.0\text{e-}10)$

221	$N^+ + CO_2 \longrightarrow CO_2^+ + N$	$k = (7.5e-10)$
222	$N^+ + CO_2 \longrightarrow CO^+ + NO$	$k = (2.5e-10)$
223	$O^+ + N_2 \longrightarrow NO^+ + N$	$k = (1.2e-12)$
224	$O^+ + O_2 \longrightarrow O_2^+ + O(^3P)$	$k = (1.2e-12)$
225	$O^+ + NO \longrightarrow NO^+ + O(^3P)$	$k = (1.1e-11)$
226	$O^+ + CO_2 \longrightarrow CO_2^+ + O(^3P)$	$k = (1.1e-9)$
227	$O^+ + CO_2 \longrightarrow O_2^+ + CO$	$k = (1.1e-9)$

228	$O_2^+ + NO \longrightarrow NO^+ + O_2$	$k = (4.4\text{e-}10)$
229	$Ar^+ + N_2 \longrightarrow N_2^+ + Ar$	$k = (2.5\text{e-}12)$
230	$Ar^+ + CO_2 \longrightarrow CO_2^+ + Ar$	$k = (4.2\text{e-}10)$
231	$Ar^+ + O_2 \longrightarrow O_2^+ + Ar$	$k = (5.5\text{e-}11)$
232	$Ar^+ + NO \longrightarrow NO^+ + Ar$	$k = (2.0\text{e-}10)$
233	$CO_2^+ + O_2 \longrightarrow O_2^+(CO_2)$	$k = (1.0\text{e-}10)$
234	$CO_2^+ + NO \longrightarrow NO^+ + CO_2$	$k = (2.0\text{e-}11)$

235	$\text{CO}_2^+ + \text{CO} \longrightarrow \text{CO}^+ + \text{CO}_2$	$k = (1.9\text{e-}12)$
236	$\text{CO}^+ + \text{CO}_2 \longrightarrow \text{CO}_2^+ + \text{CO}$	$k = (1.42\text{e-}9)$
237	$\text{CO}_2^+ + \text{O}_2 \longrightarrow \text{O}_2^+ + 2\text{CO}_2$	$k = (1.0\text{e-}10)$
238	$\text{CO}_2^+ + \text{O}_2 \longrightarrow \text{O}_2^+(\text{CO}_2) + \text{CO}_2$	$k = (2.7\text{e-}11)$
239	$\text{O}_2^+(\text{CO}_2) + \text{CO}_2 \longrightarrow \text{O}_2^+ + 2\text{CO}_2$	$k = (2.4\text{e-}13)$
240	$\text{O}_2^+(\text{CO}_2) + \text{H}_2\text{O} \longrightarrow \text{O}_2^+(\text{H}_2\text{O}) + \text{CO}_2$	$k = (1.1\text{e-}9)$
241	$\text{O}_2^+(\text{CO}_2)_2 + \text{H}_2\text{O} \longrightarrow \text{O}_2^+(\text{H}_2\text{O}) + 2\text{CO}_2$	$k = (2.3\text{e-}9)$

242	$O_2^+(H_2O) + H_2O \longrightarrow H_3O^+ + OH + O_2$	$k = (2.04e-10)$
243	$O_2^+(H_2O) + H_2O \longrightarrow H_3O^+OH + O_2$	$k = (9.96e-10)$
244	$O_2^+(H_2O)_2 + H_2O \longrightarrow O_2^+(H_2O) + 2H_2O$	$k = (3.24e-10)$
245	$O_2^+O_2 + CO_2 \longrightarrow O_2^+(CO_2) + O_2$	$k = (4.0e-13)$
246	$O_2^+O_2 + H_2O \longrightarrow O_2^+(H_2O) + O_2$	$k = (1.7e-9)$
247	$H_3O^+OH + H_2O \longrightarrow H_3O^+(H_2O) + OH$	$k = (1.4e-9)$

248	$O(^3P) + e^- \longrightarrow O^-$	$k = (1.3e-15)$
249	$O_3 + e^- \longrightarrow O^- + O_2$	$k = (9.1e-12) \exp(-46.0/T)(T/300.0)^{-1.0}$
250	$O_3^- + CO_2 \longrightarrow CO_3^- + O_2$	$k = (5.5e-10)$
251	$CO_3^- + NO \longrightarrow NO_2^- + CO_2$	$k = (1.0e-11)$
252	$CO_3^- + NO_2 \longrightarrow NO_3^- + CO_2$	$k = (2.0e-10)$
253	$CO_4^- + NO \longrightarrow NO_3^- + CO_2$	$k = (4.8e-11)$
254	$CO_4^- + O(^3P) \longrightarrow CO_3^- + O_2$	$k = (1.4e-10)$

255	$\text{CO}_4^- + \text{O}_3 \longrightarrow \text{O}_3^- + \text{CO}_2 + \text{O}_2$	$k = (1.3\text{e-}10)$
256	$\text{NO}_2^- + \text{O}_3 \longrightarrow \text{NO}_3^- + \text{O}_2$	$k = (1.2\text{e-}10)$
257	$\text{NO}_3^- + \text{CO}_2 \longrightarrow \text{CO}_3^- + \text{NO}_2$	$k = (1.0\text{e-}11)$
258	$\text{O}_2^+\text{O}_2 + h\nu \longrightarrow \text{O}_2^+ + \text{O}_2$	$k = (0.13)$
259	$\text{O}_2^+(\text{H}_2\text{O}) + h\nu \longrightarrow \text{O}_2^+ + \text{H}_2\text{O}$	$k = (0.18)$
260	$\text{O}_2^+(\text{H}_2\text{O})_2 + h\nu \longrightarrow \text{O}_2^+(\text{H}_2\text{O}) + \text{H}_2\text{O}$	$k = (0.27)$
261	$\text{CO}_3^- + h\nu \longrightarrow \text{e}^- + \text{CO}_2 + \text{O}(^3\text{P})$	$k = (0.01)$

262	$\text{CO}_3^- + h\nu \longrightarrow \text{e}^- + \text{CO}_2$	$k = (0.076)$
263	$\text{CO}_4^- + h\nu \longrightarrow \text{O}_2^- + \text{CO}_2$	$k = (0.0028)$
264	$\text{NO}_2^- + h\nu \longrightarrow \text{e}^- + \text{NO}_2$	$k = (0.018)$
265	$\text{NO}_3^- + h\nu \longrightarrow \text{O}(^3\text{P}) + \text{NO}_2^-$	$k = (0.026)$
266	$\text{NO}_3^- + h\nu \longrightarrow \text{e}^- + \text{NO} + \text{O}_2$	$k = (0.002)$
267	$\text{O}^- + h\nu \longrightarrow \text{e}^- + \text{O}(^3\text{P})$	$k = (0.62)$
268	$\text{O}_2^- + h\nu \longrightarrow \text{e}^- + \text{O}_2$	$k = (0.17)$

269	$O_3^- + h\nu \longrightarrow e^- + O_3$	$k = (0.021)$
270	$O_3^- + h\nu \longrightarrow O^- + O_2$	$k = (0.21)$
271	$N_2^+ + e^- \longrightarrow 2N$	$k = (2.7e-7)$
272	$N_2^+ + O^- \longrightarrow N_2 + O(^3P)$	$k = (1.5e-7)$
273	$N_2^+ + O_2^- \longrightarrow N_2 + O_2$	$k = (1.5e-7)$
274	$N_2^+ + O_3^- \longrightarrow N_2 + O_3$	$k = (1.5e-7)$
275	$N_2^+ + NO_2^- \longrightarrow N_2 + NO_2$	$k = (1.5e-7)$

276	$\text{N}_2^+ + \text{NO}_3^- \longrightarrow \text{N}_2 + \text{NO}_3$	$k = (1.7\text{e-}7)$
277	$\text{N}_2^+ + \text{CO}_3^- \longrightarrow$ $\text{N}_2 + \text{CO}_2 + \text{O}(^3\text{P})$	$k = (1.5\text{e-}7)$
278	$\text{N}_2^+ + \text{NO}_2^- (\text{H}_2\text{O})_2 \longrightarrow$ $\text{N}_2 + \text{NO}_2 + 2\text{H}_2\text{O}$	$k = (1.5\text{e-}7)$
279	$\text{N}_2^+ + \text{CO}_4^- \longrightarrow$ $\text{N}_2 + \text{CO}_2 + \text{O}_2$	$k = (1.5\text{e-}7)$
280	$\text{O}_2^+ + \text{e}^- \longrightarrow 2\text{O}(^3\text{P})$	$k = (2.8\text{e-}7)(T/300.0)^{-0.63}$
281	$\text{O}_2^+ + \text{O}^- \longrightarrow \text{O}_2 + \text{O}(^3\text{P})$	$k = (1.0\text{e-}7)(T/300.0)^{-0.5}$
282	$\text{O}_2^+ + \text{O}_2^- \longrightarrow 2\text{O}_2$	$k = (4.2\text{e-}7)(T/300.0)^{-0.5}$

283	$O_2^+ + O_3^- \longrightarrow O_2 + O_3$	$k = (4.0e-7)(T/300.0)^{-0.5}$
284	$O_2^+ + NO_2^- \longrightarrow O_2 + NO_2$	$k = (4.1e-7)(T/300.0)^{-0.5}$
285	$O_2^+ + NO_3^- \longrightarrow O_2 + NO_3$	$k = (1.3e-7)(T/300.0)^{-0.5}$
286	$O_2^+ + CO_3^- \longrightarrow \\ O_2 + CO_2 + O(^3P)$	$k = (2.0e-7)(T/300.0)^{-0.5}$
287	$O_2^+ + NO_2^-(H_2O)_2 \longrightarrow \\ O_2 + NO_2 + 2H_2O$	$k = (2.0e-7)(T/300.0)^{-0.5}$
288	$O_2^+ + CO_4^- \longrightarrow 2O_2 + 2CO_2$	$k = (2.0e-7)$
289	$NO^+ + e^- \longrightarrow N + O(^3P)$	$k = (4.0e-7)(T/300.0)^{-1.0}$

290	$\text{NO}^+ + \text{O}^- \longrightarrow \text{NO} + \text{O}({}^3\text{P})$	$k = (4.9\text{e-}7)(T/300.0)^{-0.5}$
291	$\text{NO}^+ + \text{O}_2^- \longrightarrow \text{NO} + \text{O}_2$	$k = (6.0\text{e-}7)(T/300.0)^{-0.5}$
292	$\text{NO}^+ + \text{O}_3^- \longrightarrow \text{NO} + \text{O}_3$	$k = (1.0\text{e-}7)(T/300.0)^{-0.5}$
293	$\text{NO}^+ + \text{NO}_2^- \longrightarrow \text{NO} + \text{NO}_2$	$k = (1.0\text{e-}7)(T/300.0)^{0.5}$
294	$\text{NO}^+ + \text{NO}_3^- \longrightarrow \text{NO}_2 + \text{NO}_3$	$k = (1.3\text{e-}7)(T/300.0)^{-0.5}$
295	$\begin{aligned} \text{NO}^+ + \text{CO}_3^- &\longrightarrow \\ \text{NO} + \text{CO}_2 + \text{O}({}^3\text{P}) & \end{aligned}$	$k = (1.2\text{e-}7)(T/300.0)^{-0.5}$
296	$\begin{aligned} \text{NO}^+ + \text{NO}_2^-(\text{H}_2\text{O})_2 &\longrightarrow \\ \text{NO} + \text{NO}_2 + 2\text{H}_2\text{O} & \end{aligned}$	$k = (1.0\text{e-}7)(T/300.0)^{-0.5}$

297	$\text{NO}^+ + \text{CO}_4^- \longrightarrow \text{NO} + \text{CO}_2 + \text{O}_2$	$k = (1.0\text{e-}7)(T/300.0)^{-0.5}$
298	$\text{CO}_2^+ + \text{e}^- \longrightarrow \text{CO} + \text{O}(^3\text{P})$	$k = (2.7\text{e-}7)$
299	$\text{CO}_2^+ + \text{O}^- \longrightarrow \text{CO}_2 + \text{O}(^3\text{P})$	$k = (1.0\text{e-}7)(T/300.0)^{-0.5}$
300	$\text{CO}_2^+ + \text{O}_2^- \longrightarrow \text{CO}_2 + \text{O}_2$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
301	$\text{CO}_2^+ + \text{O}_3^- \longrightarrow \text{CO}_2 + \text{O}_3$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
302	$\text{CO}_2^+ + \text{NO}_2^- \longrightarrow \text{CO}_2 + \text{NO}_2$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
303	$\text{CO}_2^+ + \text{NO}_3^- \longrightarrow \text{CO}_2 + \text{NO}_3$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$

304	$\text{CO}_2^+ + \text{CO}_3^- \longrightarrow 2\text{CO}_2 + \text{O}({}^3\text{P})$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
305	$\text{CO}_2^+ + \text{NO}_2^-(\text{H}_2\text{O})_2 \longrightarrow \text{CO}_2 + \text{NO}_2 + 2\text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
306	$\text{CO}_2^+ + \text{CO}_4^- \longrightarrow 2\text{CO}_2 + \text{O}_2$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
307	$\text{H}_3\text{O}^+ + \text{e}^- \longrightarrow \text{H}_2 + \text{OH}$	$k = (1.3\text{e-}6)(T/300.0)^{-0.7}$
308	$\text{H}_3\text{O}^+ + \text{O}^- \longrightarrow \text{H}_2 + \text{O}({}^3\text{P}) + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
309	$\text{H}_3\text{O}^+ + \text{O}_2^- \longrightarrow \text{H}_2 + \text{O}_2 + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
310	$\text{H}_3\text{O}^+ + \text{O}_3^- \longrightarrow \text{H}_2 + \text{O}_3 + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$

311	$\text{H}_3\text{O}^+ + \text{NO}_2^- \longrightarrow \text{H}_2 + \text{NO}_2 + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
312	$\text{H}_3\text{O}^+ + \text{NO}_3^- \longrightarrow \text{H}_2 + \text{NO}_3 + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
313	$\text{H}_3\text{O}^+ + \text{CO}_3^- \longrightarrow \text{H}_2 + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
314	$\text{H}_3\text{O}^+ + \text{NO}_2^-(\text{H}_2\text{O})_2 \longrightarrow \text{H}_2 + \text{NO}_2 + 2\text{H}_2\text{O} + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
315	$\text{H}_3\text{O}^+ + \text{CO}_4^- \longrightarrow \text{H}_2 + \text{CO}_2 + \text{O}_2 + \text{OH}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
316	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{e}^- \longrightarrow \text{H}_2 + \text{OH} + \text{H}_2\text{O}$	$k = (2.8\text{e-}6)(T/300.0)^{-0.15}$
317	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{O}_2^- \longrightarrow \text{H}_2 + \text{O}_2 + \text{OH} + \text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$

318	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{O}_3^- \longrightarrow \text{H}_2 + \text{O}_3 + \text{OH} + \text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
319	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{NO}_2^- \longrightarrow \text{H}_2 + \text{NO}_2 + \text{OH} + \text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
320	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{NO}_3^- \longrightarrow \text{H}_2 + \text{NO}_3 + \text{OH} + \text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
321	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{CO}_3^- \longrightarrow \text{H}_2 + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{OH} + \text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
322	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{NO}_2^-(\text{H}_2\text{O})_2 \longrightarrow \text{H}_2 + \text{NO}_2 + \text{OH} + 3\text{H}_2\text{O}$	$k = (2.0\text{e-}7)(T/300.0)^{-0.5}$
323	$\text{CO}_2^+\text{CO}_2 + \text{e}^- \longrightarrow 2\text{CO}_2$	$k = (2.0\text{e-}6)(T/300.0)^{-1.0}$
324	$\text{O}_2^+\text{O}_2 + \text{e}^- \longrightarrow 2\text{O}_2$	$k = (2.0\text{e-}6)(T/300.0)^{-1.0}$

325	$O_2^+O_2 + O^- \longrightarrow 2O_2 + O(^3P)$	$k = (2.0e-7)(T/300.0)^{-0.5}$
326	$O_2^+O_2 + O_2^- \longrightarrow 3O_2$	$k = (2.0e-7)(T/300.0)^{-0.5}$
327	$O_2^+O_2 + O_3^- \longrightarrow 2O_2 + O_3$	$k = (2.0e-7)(T/300.0)^{-0.5}$
328	$O_2^+O_2 + NO_3^- \longrightarrow 2O_2 + NO_3$	$k = (2.0e-7)(T/300.0)^{-0.5}$
329	$O_2^+O_2 + CO_3^- \longrightarrow$ $2O_2 + CO_2 + O(^3P)$	$k = (2.0e-7)(T/300.0)^{-0.5}$
330	$O_2^+O_2 + NO_2^- (H_2O)_2 \longrightarrow$ $2O_2 + NO_2 + 2H_2O$	$k = (2.0e-7)(T/300.0)^{-0.5}$
331	$O_2^+O_2 + CO_4^- \longrightarrow 3O_2 + CO_2$	$k = (2.0e-7)(T/300.0)^{-0.5}$

332	$O_2^- + HNO_3 \longrightarrow NO_3^- + HO_2$	$k = (2.8e-9)$
333	$O_2^- + HCl \longrightarrow Cl^- + HO_2$	$k = (1.6e-9)$
334	$O^- + O_3 \longrightarrow O_3^- + O(^3P)$	$k = (8.0e-10)$
335	$O^- + NO_2 \longrightarrow NO_2^- + O(^3P)$	$k = (1.0e-9)$
336	$O^- + H_2 \longrightarrow OH^- + H$	$k = (3.2e-11)$
337	$O^- + CH_4 \longrightarrow OH^- + CH_3$	$k = (1.0e-10)$
338	$O^- + H_2 \longrightarrow H_2O + e^-$	$k = (6.0e-10)$

339	$O^- + O(^3P) \longrightarrow O_2 + e^-$	$k = (1.9e-10)$
340	$O^- + NO \longrightarrow NO_2 + e^-$	$k = (2.8e-10)$
341	$O^- + HCl \longrightarrow Cl^- + OH$	$k = (2.0e-9)$
342	$O^- + HNO_3 \longrightarrow NO_3^- + OH$	$k = (3.0e-9)$
343	$O_2^- + O(^3P) \longrightarrow O_2 + O^-$	$k = (1.5e-10)$
344	$O_2^- + O_3 \longrightarrow O_2 + O_3^-$	$k = (7.8e-10)$
345	$O_2^- + NO_2 \longrightarrow NO_2^- + O_2$	$k = (7.8e-10)$

346	$O_2^- + O(^3P) \longrightarrow e^- + O_3$	$k = (1.5e-10)$
347	$O_2^- + H \longrightarrow e^- + HO_2$	$k = (1.5e-10)$
348	$O_2^-(H_2O) + CO_2 \longrightarrow CO_4^- + HO_2$	$k = (5.8e-10)$
349	$O_3^- + O(^3P) \longrightarrow O_2^- + O_2$	$k = (2.5e-10)$
350	$O_3^- + H \longrightarrow OH^- + O_2$	$k = (8.4e-10)$
351	$O_3^- + NO_2 \longrightarrow NO_3^- + O_2$	$k = (2.8e-10)$
352	$O_3^- + NO \longrightarrow NO_3^- + O(^3P)$	$k = (4.5e-12)$

353	$\text{CO}_3^- + \text{N}_2\text{O}_5 \longrightarrow \text{NO}_3^- + \text{CO}_2 + \text{NO}_3$	$k = (2.8\text{e-}10)$
354	$\text{CO}_3^- + \text{O}({}^3\text{P}) \longrightarrow \text{O}_2^- + \text{CO}_2$	$k = (1.1\text{e-}10)$
355	$\text{CO}_3^- + \text{H} \longrightarrow \text{OH}^- + \text{CO}_2$	$k = (1.7\text{e-}10)$
356	$\text{NO}_2^- + \text{NO}_2 \longrightarrow \text{NO}_3^- + \text{NO}$	$k = (2.0\text{e-}13)$
357	$\text{NO}_2^- + \text{H} \longrightarrow \text{OH}^- + \text{NO}$	$k = (3.0\text{e-}10)$
358	$\text{NO}_2^- + \text{HCl} \longrightarrow \text{Cl}^- + \text{HONO}$	$k = (1.4\text{e-}9)$
359	$\text{NO}_2^- + \text{HNO}_3 \longrightarrow \text{NO}_3^- + \text{HONO}$	$k = (1.6\text{e-}9)$

360	$\text{NO}_2^- + \text{N}_2\text{O}_5 \longrightarrow \text{NO}_3^- + 2\text{NO}_2$	$k = (7.0\text{e-}10)$
361	$\text{OH}^- + \text{NO}_2 \longrightarrow \text{NO}_2^- + 2\text{OH}$	$k = (1.1\text{e-}9)$
362	$\text{OH}^- + \text{O}_3 \longrightarrow \text{O}_3^- + 2\text{OH}$	$k = (9.0\text{e-}10)$
363	$\text{OH}^- + \text{H} \longrightarrow \text{e}^- + 2\text{H}_2\text{O}$	$k = (1.4\text{e-}9)$
364	$\text{OH}^- + \text{O}({}^3\text{P}) \longrightarrow \text{e}^- + 2\text{HO}_2$	$k = (2.0\text{e-}10)$
365	$\text{O}_4^- + \text{O}({}^3\text{P}) \longrightarrow \text{O}_3^- + 2\text{O}_2$	$k = (4.0\text{e-}10)$
366	$\text{O}_4^- + \text{CO}_2 \longrightarrow \text{CO}_4^- + 2\text{O}_2$	$k = (4.3\text{e-}10)$

367	$O_4^- + NO \longrightarrow NO_3^- + 2O_2$	$k = (2.5\text{e-}10)$
368	$O_4^- + H_2O \longrightarrow O_2^-(H_2O) + 2O_2$	$k = (1.0\text{e-}10)$
369	$CO_4^- + H \longrightarrow CO_3^- + 2OH$	$k = (2.2\text{e-}10)$
370	$CO_4^- + H_2O \longrightarrow O_2^-(H_2O) + 2CO_2$	$k = (2.5\text{e-}10)$
371	$Cl^- + H \longrightarrow HCl + 2e^-$	$k = (9.3\text{e-}10)$
372	$Cl^- + HNO_3 \longrightarrow HCl + 2NO_3^-$	$k = (1.6\text{e-}9)$
373	$Cl^- + N_2O_5 \longrightarrow NO_3^- + 2ClNO_2$	$k = (9.4\text{e-}10)$

### A.1.2 Trimolecular Reactions

#	Trimolecular Reaction	Reaction Rate Coeff
1	$\text{Br} + \text{NO}_2 \longrightarrow \text{BrONO} + \text{M}$	$k_0 = M(4.2\text{e-}31)(T/300.0)^{-2.4}$ $k_\infty = (2.7\text{e-}11)(T/300.0)^{0.0}$ $f_c = (0.55)$
2	$\text{BrO} + \text{NO}_2 \longrightarrow \text{BrONO}_2 + \text{M}$	$k_0 = M(5.2\text{e-}31)(T/300.0)^{-3.2}$ $k_\infty = (6.9\text{e-}12)(T/300.0)^{-2.9}$ $f_c = (0.6)$
3	$\text{CH}_3 + \text{O}_2 \longrightarrow \text{CH}_3\text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}30)(T/300.0)^{-3.3}$ $k_\infty = (2.2\text{e-}12)(T/300.0)^{1.0}$ $f_c = (0.27)$
4	$\text{CO} + \text{H} \longrightarrow \text{HCO} + \text{M}$	$k_0 = M(1.9\text{e-}33) \exp(-7000.0/T)(T/300.0)^{0.0}$ $k_\infty = (T/298.0)^{0.0}$ $f_c = (0.0)$

5	$\text{CO} + \text{O}({}^3\text{P}) \longrightarrow \text{CO}_2 + \text{M}$	$k_0 = M(1.7\text{e-}33) \exp(-12550.0/T)(T/300.0)^{0.0}$ $k_\infty = (4.2\text{e-}12) \exp(-199540.0/T)(T/298.0)^{0.0}$ $f_c = (0.6)$
6	$\text{Cl}_2\text{O}_2 + \text{M} \longrightarrow \text{ClO} + \text{ClO}$	$k_0 = M(1.35\text{e-}5) \exp(-8720.0/T)(T/300.0)^{1.0}$ $k_\infty = (1.8\text{e}15) \exp(-8450.0/T)(T/300.0)^{0.0}$ $f_c = (0.6)$
7	$\text{Cl} + \text{NO}_2 \longrightarrow \text{ClNO}_2 + \text{M}$	$k_0 = M(1.8\text{e-}31)(T/300.0)^{-2.0}$ $k_\infty = (1.0\text{e-}10)(T/300.0)^{-1.0}$ $f_c = (0.6)$
8	$\text{Cl} + \text{O}_2 \longrightarrow \text{ClOO} + \text{M}$	$k_0 = M(2.7\text{e-}33)(T/300.0)^{-1.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
9	$\text{ClO} + \text{ClO} \longrightarrow \text{Cl}_2\text{O}_2 + \text{M}$	$k_0 = M(2.2\text{e-}32)(T/300.0)^{-3.1}$ $k_\infty = (3.5\text{e-}12)(T/300.0)^{-1.0}$ $f_c = (0.6)$

10	$\text{ClO} + \text{NO}_2 \longrightarrow \text{ClONO}_2 + \text{M}$	$k_0 = M(1.6\text{e-}31)(T/300.0)^{-3.4}$ $k_\infty = (2.0\text{e-}11)(T/300.0)^{0.0}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/430.0)$
11	$\text{ClONO}_2 + \text{M} \longrightarrow \text{ClO} + \text{NO}_2$	$k_0 = M(T/300.0)^{0.0}$ $k_\infty = (2.75\text{e-}6) \exp(-95100.0/T)(T/300.0)^{0.0}$ $f_c = (0.0)$
12	$\text{ClOO} + \text{M} \longrightarrow \text{Cl} + \text{O}_2$	$k_0 = M(2.8\text{e-}10) \exp(-1820.0/T)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
13	$\text{HO}_2 + \text{HO}_2 \longrightarrow \text{H}_2\text{O}_2 + \text{O}_2$	$k_0 = M(1.9\text{e-}33) \exp(980.0/T)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

14	$H + O_2 \longrightarrow HO_2 + M$	$k_0 = M(6.2e-32)(T/300.0)^{-1.6}$ $k_\infty = (7.5e-11)(T/300.0)^{0.0}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/498.0)$
15	$HO_2 + NO_2 \longrightarrow HO_2NO_2 + M$	$k_0 = M(1.8e-31)(T/300.0)^{-3.2}$ $k_\infty = (4.7e-12)(T/300.0)^{-1.4}$ $f_c = (0.6)$
16	$HO_2NO_2 + M \longrightarrow HO_2 + NO_2$	$k_0 = M(5.0e-6) \exp(-10000.0/T)(T/300.0)^{0.0}$ $k_\infty = (2.6e15) \exp(-10900.0/T)(T/300.0)^{0.0}$ $f_c = (0.6)$
17	$HONO + M \longrightarrow NO + OH$	$k_0 = M(6.4e6)T^{-3.8} \exp(-25340.0/T)$ $k_\infty = (1.2e19)T^{-1.23} \exp(-25010.0/T)$ $f_c = (0.62)$

18	$\text{CH}_3\text{O}_2\text{NO}_2 + \text{M} \longrightarrow \text{CH}_3\text{O}_2 + \text{NO}_2$	$k_0 = M(9.0\text{e-}5) \exp(-9690.0/T)(T/300.0)^{0.0}$ $k_\infty = (1.1\text{e}16) \exp(-10560.0/T)(T/300.0)^{0.0}$ $f_c = (0.4)$
19	$\text{CH}_3\text{O} + \text{NO}_2 \longrightarrow \text{CH}_3\text{ONO}_2 + \text{M}$	$k_0 = M(2.8\text{e-}29)(T/300.0)^{-4.5}$ $k_\infty = (2.0\text{e-}11)(T/300.0)^{0.0}$ $f_c = (0.44)$
20	$\text{CH}_3\text{O}_2 + \text{NO}_2 \longrightarrow \text{CH}_3\text{O}_2\text{NO}_2 + \text{M}$	$k_0 = M(2.5\text{e-}30)(T/300.0)^{-5.5}$ $k_\infty = (7.5\text{e-}12)(T/300.0)^{0.0}$ $f_c = (0.4)$
21	$\text{N}_2 + \text{O}({}^1\text{D}) \longrightarrow \text{N}_2\text{O} + \text{M}$	$k_0 = M(3.5\text{e-}37)(T/300.0)^{-0.6}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
22	$\text{N}_2\text{O}_5 + \text{M} \longrightarrow \text{NO}_2 + \text{NO}_3$	$k_0 = M(0.0022) \exp(-11080.0/T)(T/300.0)^{-4.4}$ $k_\infty = (9.7\text{e}14) \exp(-11080.0/T)(T/300.0)^{0.1}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/250.0) + \exp(ext rm - 1)$

23	$\text{NO}_2 + \text{NO}_3 \longrightarrow \text{N}_2\text{O}_5 + \text{M}$	$k_0 = M(2.7\text{e-}30)(T/300.0)^{-3.4}$ $k_\infty = (2.0\text{e-}12)(T/300.0)^{0.2}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/250.0) + \exp(extrm-1)$
24	$\text{NO}_2 + \text{O}({}^3\text{P}) \longrightarrow \text{NO}_3 + \text{M}$	$k_0 = M(9.0\text{e-}32)(T/300.0)^{-2.0}$ $k_\infty = (2.2\text{e-}11)(T/300.0)^{0.0}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/1300.0)$
25	$\text{NO}_2 + \text{OH} \longrightarrow \text{HNO}_3 + \text{M}$	$k_0 = M(2.47\text{e-}30)(T/300.0)^{-2.97}$ $k_\infty = (1.45\text{e-}11)(T/300.0)^{-2.77}$ $f_c = (0.6)$
26	$\text{NO} + \text{NO} \longrightarrow \text{NO}_2 + \text{NO}_2$	$k_0 = M \exp(530.0/T)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

27	$\text{NO} + \text{O}({}^3\text{P}) \longrightarrow \text{NO}_2 + \text{M}$	$k_0 = M(1.0\text{e-}31)(T/300.0)^{-1.6}$ $k_\infty = (3.0\text{e-}11)(T/300.0)^{0.3}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/1850.0)$
28	$\text{NO} + \text{OH} \longrightarrow \text{HONO} + \text{M}$	$k_0 = M(7.4\text{e-}31)(T/300.0)^{-2.4}$ $k_\infty = (3.2\text{e-}11)(T/300.0)^{0.0}$ $f_c = (0.0) + (0.0) \cdot \exp(-T/1300.0)$
29	$\text{O}({}^3\text{P}) + \text{O}_2 \longrightarrow \text{O}_3 + \text{M}$	$k_0 = M(5.64\text{e-}34)(T/300.0)^{-2.8}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
30	$\text{OH} + \text{OH} \longrightarrow \text{H}_2\text{O}_2 + \text{M}$	$k_0 = M(8.0\text{e-}31)(T/300.0)^{-0.8}$ $k_\infty = (3.0\text{e-}11)(T/300.0)^{0.0}$ $f_c = (0.5)$

31	$\text{CH}_3\text{ONO}_2 + \text{M} \longrightarrow \text{CH}_3\text{O} + \text{NO}_2$	$k_0 = M(T/300.0)^{0.0}$ $k_\infty = (1.0\text{e}13) \exp(-140.0/T)(T/300.0)^{0.0}$ $f_c = (0.0)$
32	$\text{O}_2^+ + \text{O}_2 \longrightarrow \text{O}_2^+\text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e}-30)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
33	$\text{O}_2^+ + \text{H}_2\text{O} \longrightarrow \text{O}_2^+(\text{H}_2\text{O}) + \text{M}$	$k_0 = M(2.8\text{e}-28)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
34	$\text{O}_2^+ + \text{CO}_2 \longrightarrow \text{O}_2^+(\text{CO}_2) + \text{M}$	$k_0 = M(1.7\text{e}-29)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
35	$\text{CO}_2^+ + \text{CO}_2 \longrightarrow \text{CO}_2^+\text{CO}_2 + \text{M}$	$k_0 = M(2.5\text{e}-28)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

36	$O_2^+(CO_2) + CO_2 \longrightarrow$ $O_2^+(CO_2)_2 + M$	$k_0 = M(5.0e-30)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
37	$O_2^+(H_2O) + H_2O \longrightarrow$ $O_2^+(H_2O)_2 + M$	$k_0 = M(1.3e-27)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
38	$H_3O^+ + H_2O \longrightarrow H_3O^+(H_2O) + M$	$k_0 = M(3.4e-27)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
39	$H_3O^+(H_2O) + H_2O \longrightarrow$ $H_3O^+(H_2O)_2 + M$	$k_0 = M(2.3e-27)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
40	$H_3O^+(H_2O) + M \longrightarrow$ $H_3O^+ + H_2O + M$	$k_0 = M(7.0e-26)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

41	$\begin{aligned} \text{H}_3\text{O}^+(\text{H}_2\text{O})_2 + \text{H}_2\text{O} &\longrightarrow \\ \text{H}_3\text{O}^+(\text{H}_2\text{O})_3 + \text{M} \end{aligned}$	$k_0 = M(2.4\text{e-}27)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
42	$\begin{aligned} \text{H}_3\text{O}^+(\text{H}_2\text{O})_2 + \text{M} &\longrightarrow \\ \text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{H}_2\text{O} + \text{M} \end{aligned}$	$k_0 = M(7.0\text{e-}18)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
43	$\begin{aligned} \text{H}_3\text{O}^+(\text{H}_2\text{O})_3 + \text{H}_2\text{O} &\longrightarrow \\ \text{H}_3\text{O}^+(\text{H}_2\text{O})_4 + \text{M} \end{aligned}$	$k_0 = M(9.0\text{e-}28)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
44	$\begin{aligned} \text{H}_3\text{O}^+(\text{H}_2\text{O})_3 + \text{M} &\longrightarrow \\ \text{H}_3\text{O}^+(\text{H}_2\text{O})_2 + \text{H}_2\text{O} + \text{M} \end{aligned}$	$k_0 = M(4.0\text{e-}14)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
45	$\begin{aligned} \text{H}_3\text{O}^+(\text{H}_2\text{O})_4 + \text{M} &\longrightarrow \\ \text{H}_3\text{O}^+(\text{H}_2\text{O})_3 + \text{H}_2\text{O} + \text{M} \end{aligned}$	$k_0 = M(6.0\text{e-}12)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

46	$O_2 + e^- \longrightarrow O_2^- + M$	$k_0 = M(2.0e-31) \exp(-600.0/T)(T/300.0)^{1.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
47	$O^- + CO_2 \longrightarrow$ $CO_3^- + CO_2 + M$	$k_0 = M(1.1e-27)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
48	$O_2^- + CO_2 \longrightarrow CO_4^- + M$	$k_0 = M(1.3e-29)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
49	$O_2^+ + O^- \longrightarrow$ $O_2 + O(^3P) + M$	$k_0 = M(3.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
50	$O_2^+ + O_2^- \longrightarrow 2O_2 + M$	$k_0 = M(3.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

51	$O_2^+ + O_3^- \longrightarrow$ $O_2 + O_3 + M$	$k_0 = M(3.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
52	$O_2^+ + NO_2^- \longrightarrow$ $O_2 + NO_2 + M$	$k_0 = M(3.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
53	$O_2^+ + NO_3^- \longrightarrow$ $O_2 + NO_3 + M$	$k_0 = M(3.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
54	$O_2^+ + CO_3^- \longrightarrow$ $O_2 + CO_2 + O(^3P) + M$	$k_0 = M(3.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
55	$O_2^+ + NO_2^- (H_2O)_2 \longrightarrow$ $O_2 + NO_2 + 2H_2O + M$	$k_0 = M(1.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

56	$\text{NO}^+ + \text{O}^- \longrightarrow$ $\text{NO} + \text{O}({}^3\text{P}) + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
57	$\text{NO}^+ + \text{O}_2^- \longrightarrow$ $\text{NO} + \text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
58	$\text{NO}^+ + \text{O}_3^- \longrightarrow$ $\text{NO} + \text{O}_3 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
59	$\text{NO}^+ + \text{NO}_2^- \longrightarrow$ $\text{NO} + \text{NO}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
60	$\text{NO}^+ + \text{NO}_3^- \longrightarrow$ $\text{NO} + \text{NO}_3 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

61	$\text{NO}^+ + \text{CO}_3^- \longrightarrow$ $\text{NO} + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
62	$\text{NO}^+ + \text{NO}_2^- (\text{H}_2\text{O})_2 \longrightarrow$ $\text{NO} + \text{NO}_2 + 2\text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
63	$\text{NO}^+ + \text{CO}_4^- \longrightarrow$ $\text{NO} + \text{CO}_2 + \text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
64	$\text{CO}_2^+ + \text{O}_2^- \longrightarrow$ $\text{CO}_2 + \text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
65	$\text{CO}_2^+ + \text{O}_3^- \longrightarrow$ $\text{CO}_2 + \text{O}_3 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

66	$\text{CO}_2^+ + \text{NO}_2^- \longrightarrow$ $\text{CO}_2 + \text{NO}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
67	$\text{CO}_2^+ + \text{NO}_3^- \longrightarrow$ $\text{CO}_2 + \text{NO}_3 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
68	$\text{CO}_2^+ + \text{CO}_3^- \longrightarrow$ $\text{CO}_2 + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
69	$\text{CO}_2^+ + \text{NO}_2^-(\text{H}_2\text{O})_2 \longrightarrow$ $\text{CO}_2 + \text{NO}_2 + 2\text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
70	$\text{CO}_2^+ + \text{CO}_4^- \longrightarrow$ $2\text{CO}_2 + \text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

71	$\text{H}_3\text{O}^+ + \text{O}^- \longrightarrow$ $\text{H}_2 + \text{O}(^3\text{P}) + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
72	$\text{H}_3\text{O}^+ + \text{O}_2^- \longrightarrow$ $\text{H}_2 + \text{O}_2 + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
73	$\text{H}_3\text{O}^+ + \text{O}_3^- \longrightarrow$ $\text{H}_2 + \text{O}_3 + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
74	$\text{H}_3\text{O}^+ + \text{NO}_2^- \longrightarrow$ $\text{H}_2 + \text{NO}_2 + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
75	$\text{H}_3\text{O}^+ + \text{NO}_3^- \longrightarrow$ $\text{H}_2 + \text{NO}_3 + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

76	$\text{H}_3\text{O}^+ + \text{CO}_3^- \longrightarrow$ $\text{H}_2 + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
77	$\text{H}_3\text{O}^+ + \text{NO}_2^-(\text{H}_2\text{O})_2 \longrightarrow$ $\text{H}_2 + \text{NO}_2 + 2\text{H}_2\text{O} + \text{OH} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
78	$\text{H}_3\text{O}^+ + \text{CO}_4^- \longrightarrow$ $\text{H}_2 + \text{CO}_2 + \text{O}_2 + \text{OH} + \text{M}$	$k_0 = M(3.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
79	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{O}^- \longrightarrow$ $\text{H}_2 + \text{O}({}^3\text{P}) + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
80	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{O}_2^- \longrightarrow$ $\text{H}_2 + \text{O}_2 + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

81	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{O}_3^- \longrightarrow$ $\text{H}_2 + \text{O}_3 + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
82	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{NO}_2^- \longrightarrow$ $\text{H}_2 + \text{NO}_2 + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
83	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{NO}_3^- \longrightarrow$ $\text{H}_2 + \text{NO}_3 + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
84	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{CO}_3^- \longrightarrow$ $\text{H}_2 + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
85	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{NO}_2^-(\text{H}_2\text{O})_2 \longrightarrow$ $\text{H}_2 + \text{NO}_2 + \text{OH} + 3\text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

86	$\text{H}_3\text{O}^+(\text{H}_2\text{O}) + \text{CO}_4^- \longrightarrow$ $\text{H}_2 + \text{CO}_2 + \text{O}_2 + \text{OH} + \text{H}_2\text{O} + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
87	$\text{O}_2^+\text{O}_2 + \text{O}^- \longrightarrow$ $\text{O}_2 + \text{O}({}^3\text{P}) + \text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
88	$\text{O}_2^+\text{O}_2 + \text{O}_2^- \longrightarrow 3\text{O}_2 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
89	$\text{O}_2^+\text{O}_2 + \text{NO}_3^- \longrightarrow$ $2\text{O}_2 + \text{NO}_3 + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
90	$\text{O}_2^+\text{O}_2 + \text{CO}_3^- \longrightarrow$ $2\text{O}_2 + \text{CO}_2 + \text{O}({}^3\text{P}) + \text{M}$	$k_0 = M(1.0\text{e-}25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$

91	$O_2^+ O_2 + NO_2^- (H_2O)_2 \longrightarrow$ $2O_2 + NO_2 + 2H_2O + M$	$k_0 = M(1.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
92	$O_2^+ O_2 + CO_4^- \longrightarrow$ $3O_2 + CO_2 + M$	$k_0 = M(1.0e-25)(T/300.0)^{-2.5}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
93	$O_2^- + O_2 \longrightarrow O_4^- + M$	$k_0 = M(3.4e-31)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
94	$O_2^- + H_2O \longrightarrow O_2^- (H_2O) + M$	$k_0 = M(2.2e-28)(T/300.0)^{0.0}$ $k_\infty = (T/300.0)^{0.0}$ $f_c = (0.0)$
95	$O_2^- (H_2O) + NO \longrightarrow$ $NO_3^- + H_2O + M$	$k_0 = M(T/300.0)^{0.0}$ $k_\infty = (2.0e-10)(T/300.0)^{0.0}$ $f_c = (0.0)$

### A.1.3 Photolysis Reactions

#	Photolysis Reaction
1	$\text{Br}_2 + h\nu \longrightarrow \text{Br} + \text{Br}$
2	$\text{BrCl} + h\nu \longrightarrow \text{Cl} + \text{Br}$
3	$\text{BrO} + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{Br}$
4	$\text{BrO} + h\nu \longrightarrow \text{Br} + \text{O}({}^1\text{D})$
5	$\text{BrONO} + h\nu \longrightarrow \text{Br} + \text{NO}_2$
6	$\text{BrONO}_2 + h\nu \longrightarrow \text{NO}_2 + \text{BrO}$
7	$\text{BrONO}_2 + h\nu \longrightarrow \text{Br} + \text{NO}_3$

8	$\text{CF}_2\text{Cl}_2 + h\nu \longrightarrow \text{Cl} + \text{Cl}$
9	$\text{CH}_3\text{Br} + h\nu \longrightarrow \text{Br} + \text{CH}_3$
10	$\text{CH}_4 + h\nu \longrightarrow \text{H} + \text{CH}_3$
11	$\text{CO}_2 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{CO}$
12	$\text{CO}_2 + h\nu \longrightarrow \text{CO} + \text{O}({}^1\text{D})$
13	$\text{Cl}_2 + h\nu \longrightarrow \text{Cl} + \text{Cl}$
14	$\text{Cl}_2\text{O}_2 + h\nu \longrightarrow \text{Cl} + \text{ClOO}$

15	$\text{ClNO}_2 + h\nu \longrightarrow \text{Cl} + \text{NO}_2$
16	$\text{ClO} + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{Cl}$
17	$\text{ClONO}_2 + h\nu \longrightarrow \text{Cl} + \text{NO}_3$
18	$\text{ClOO} + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{ClO}$
19	$\text{H}_2\text{O} + h\nu \longrightarrow$ $\text{H} + \text{H} + \text{O}({}^3\text{P})$
20	$\text{H}_2\text{O} + h\nu \longrightarrow \text{H}_2 + \text{O}({}^1\text{D})$
21	$\text{H}_2\text{O} + h\nu \longrightarrow \text{H} + \text{OH}$

22	$\text{H}_2\text{O} + h\nu \longrightarrow \text{H}_2 + \text{O}^+$
23	$\text{H}_2\text{O}_2 + h\nu \longrightarrow \text{OH} + \text{OH}$
24	$\text{HCHO} + h\nu \longrightarrow \text{H}_2 + \text{CO}$
25	$\text{HCHO} + h\nu \longrightarrow \text{H} + \text{HCO}$
26	$\text{HCHO} + h\nu \longrightarrow$ $\text{H} + \text{H} + \text{CO}$
27	$\text{HCl} + h\nu \longrightarrow \text{H} + \text{Cl}$
28	$\text{HO}_2 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{OH}$

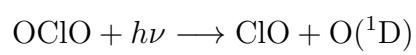
29	$\text{HO}_2\text{NO}_2 + h\nu \longrightarrow \text{NO}_2 + \text{HO}_2$
30	$\text{HO}_2\text{NO}_2 + h\nu \longrightarrow \text{OH} + \text{NO}_3$
31	$\text{HOBr} + h\nu \longrightarrow \text{OH} + \text{Br}$
32	$\text{HOCl} + h\nu \longrightarrow \text{Cl} + \text{OH}$
33	$\text{HONO} + h\nu \longrightarrow \text{NO} + \text{OH}$
34	$\text{HNO}_3 + h\nu \longrightarrow \text{OH} + \text{NO}_2$
35	$\text{HNO}_3 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{HONO}$

36	$\text{CH}_3\text{O}_2\text{NO}_2 + h\nu \longrightarrow \text{NO}_2 + \text{CH}_3\text{O}_2$
37	$\text{CH}_3\text{O}_2\text{NO}_2 + h\nu \longrightarrow \text{NO}_3 + \text{CH}_3\text{O}$
38	$\text{CH}_3\text{OCl} + h\nu \longrightarrow \text{Cl} + \text{CH}_3\text{O}$
39	$\text{CH}_3\text{OH} + h\nu \longrightarrow \text{H}_2 + \text{HCHO}$
40	$\text{CH}_3\text{OH} + h\nu \longrightarrow \text{OH} + \text{CH}_3$
41	$\text{CH}_3\text{ONO}_2 + h\nu \longrightarrow \text{NO}_2 + \text{CH}_3\text{O}$
42	$\text{CH}_3\text{OOH} + h\nu \longrightarrow \text{OH} + \text{CH}_3\text{O}$

43	$\text{CH}_3\text{OOH} + h\nu \longrightarrow$ $\text{O}({}^3\text{P}) + \text{H} + \text{CH}_3\text{O}$
44	$\text{N}_2\text{O} + h\nu \longrightarrow \text{N}_2 + \text{O}({}^1\text{D})$
45	$\text{N}_2\text{O}_5 + h\nu \longrightarrow$ $\text{O}({}^3\text{P}) + \text{NO} + \text{NO}_3$
46	$\text{N}_2\text{O}_5 + h\nu \longrightarrow \text{NO}_2 + \text{NO}_3$
47	$\text{NO} + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{N}$
48	$\text{NO}_2 + h\nu \longrightarrow \text{NO} + \text{O}({}^1\text{D})$
49	$\text{NO}_2 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{NO}$

50	$\text{NO}_3 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{NO}_2$
51	$\text{NO}_3 + h\nu \longrightarrow \text{O}_2 + \text{NO}$
52	$\text{O}_2 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{O}({}^1\text{D})$
53	$\text{O}_2 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{O}({}^3\text{P})$
54	$\text{O}_3 + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{O}_2$
55	$\text{O}_3 + h\nu \longrightarrow \text{O}_2 + \text{O}({}^1\text{D})$
56	$\text{OCIO} + h\nu \longrightarrow \text{O}({}^3\text{P}) + \text{ClO}$

57



## A.2 Master Chemical Mechanism

## REFERENCES

- Otter. <https://www.maritimerobotics.com/otter>.
- (2021, May). Introducing alta x. <https://freeflysystems.com/alta-x>.
- Adhikari, R. and R. K. Agrawal (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.
- Ahmed, S. E., S. Pawar, and O. San (2020). Pyda: A hands-on introduction to dynamical data assimilation with python. *Fluids* 5(4), 225.
- Altmeyer, P. (2023). Conformal prediction.
- Arnaut, L. and H. Burrows (2006). *Chemical kinetics: from molecular structure to chemical reactivity*. Elsevier.
- Bäumker, M. and F. Heimes (2001). New calibration and computing method for direct georeferencing of image and scanner data using the position and angular data of an hybrid inertial navigation system. In *OEEPE Workshop, Integrated Sensor Orientation*, pp. 1–17.
- Bishop, C. M., M. Svensén, and C. K. Williams (1998). Gtm: The generative topographic mapping. *Neural computation* 10(1), 215–234.
- Blaom, A. D., F. Kiraly, T. Lienart, Y. Simillides, D. Arenas, and S. J. Vollmer (2020). MLJ: A julia package for composable machine learning. *Journal of Open Source Software* 5(55), 2704.
- Blaom, A. D. and S. J. Vollmer (2020). Flexible model composition in machine learning and its implementation in MLJ.
- Boldi, R. A. (1994). A model of the ion chemistry of electrified convection.
- Brunton, S. L., B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz (2017). Chaos as an intermittently forced linear system. *Nature communications* 8(1), 19.
- Brunton, S. L., M. Budisic, E. Kaiser, and J. N. Kutz (2021). Modern koopman theory for dynamical systems. *arXiv preprint arXiv:2102.12086*.
- Brunton, S. L., J. L. Proctor, and J. N. Kutz (2016). Sparse identification of nonlinear dynamics with control (sindyc). *IFAC-PapersOnLine* 49(18), 710–715.
- Burkholder, J., S. Sander, J. Abbatt, J. Barker, C. Cappa, J. Crounse, T. Dibble, R. Huie, C. Kolb, M. Kurylo, et al. (2020). Chemical kinetics and photochemical data for use in atmospheric studies; evaluation number 19. Technical report, Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space ....

- Carneiro, J., M. A. Cole, and E. Strobl (2021). The effects of air pollution on students' cognitive performance: Evidence from brazilian university entrance tests. *Journal of the Association of Environmental and Resource Economists* 8, 1051 – 1077.
- Chen, R. T., Y. Rubanova, J. Bettencourt, and D. K. Duvenaud (2018). Neural ordinary differential equations. *Advances in neural information processing systems* 31.
- D'Amato, G., H. J. Chong-Neto, O. P. M. Ortega, C. Vitale, I. J. Ansotegui, N. A. Rosário, T. Haathela, C. Galán, R. Pawankar, M. Murrieta, L. Cecchi, C. Bergmann, E. Ridolo, G. D. Ramon, S. N. G. Díaz, M. D'Amato, and I. Annesi-Maesano (2020). The effects of climate change on respiratory allergy and asthma induced by pollen and mold allergens. *Allergy* 75, 2219 – 2228.
- Ehsanifar, M., Z. Montazeri, and M. Rafati (2022). Exposure to urban air pollution nanoparticles: Oxidative stress and cardiovascular disease. *Journal of Biomedical Research & Environmental Sciences*.
- Ehsanifar, M., B. Ss, and F. Farokhmanesh (2021). Exposure to ambient ultra-fine particles and stroke. *Journal of Biomedical Research & Environmental Sciences*.
- Emmerich, S. J., T. McDowell, and W. Anis (2005). Investigation of the impact of commercial building envelope airtightness on HVAC energy use. Technical Report NIST IR 7238, National Institute of Standards and Technology, Gaithersburg, MD.
- enVerid (2022a). How to achieve sustainable indoor air quality: A roadmap to simultaneously improving indoor air quality & meeting building decarbonization and climate resiliency goals. <https://enverid.com/resources/learning/how-to-achieve-sustainable-indoor-air-quality-a-roadmap-to-simultaneously-improving>. Last accessed on 2022-12-01.
- enVerid (2022b). Leaders in indoor air quality and energy efficiency share framework for achieving healthy indoor air while decarbonizing buildings. <https://enverid.com/awards/leaders-in-indoor-air-quality-and-energy-efficiency-share-framework-for-achieving>. Last accessed on 2022-12-01.
- Finewax, Z., D. Pagonis, M. S. Clafflin, A. V. Handschy, W. L. Brown, O. Jenks, B. A. Nault, D. A. Day, B. M. Lerner, J. L. Jimenez, P. J. Ziemann, and J. A. Gouw (2021, September). Quantification and source characterization of volatile organic compounds from exercising and application of chlorine-based cleaning products in a university athletic center. *Indoor Air* 31(5), 1323–1339.
- Gao, X., B. Coull, X. Lin, P. Vokonas, A. Spiro, L. Hou, J. Schwartz, and A. A. Bacarelli (2021). Short-term air pollution, cognitive performance, and nonsteroidal anti-inflammatory drug use in the veterans affairs normative aging study. *Nature aging* 1 5, 430.

- Greydanus, S., M. Dzamba, and J. Yosinski (2019). Hamiltonian neural networks. *Advances in neural information processing systems* 32.
- Huang, S., L. Tang, J. P. Hupy, Y. Wang, and G. Shao (2020). A commentary review on the use of normalized difference vegetation index (ndvi) in the era of popular remote sensing. *Journal of Forestry Research* 32, 1–6.
- Innes, M., A. Edelman, K. Fischer, C. Rackauckas, E. Saba, V. B. Shah, and W. Tebbutt (2019). A differentiable programming system to bridge machine learning and scientific computing. *arXiv preprint arXiv:1907.07587*.
- Juginović, A., M. Vuković, I. Aranza, and V. Biloš (2021). Health impacts of air pollution exposure from 1990 to 2019 in 43 european countries. *Scientific Reports* 11(1), 22516.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics* 43(1), 59–69.
- Krebs, B. and S. Luechinger (2021). Air pollution, cognitive performance, and the role of task proficiency. *Planetary Health eJournal*.
- Lamping, G. A. and C. O. Muller. Air Cleaning in Practice – School Sustainability and Commercial Building Field Study Results. pp. 13.
- Lamqadem, A. A., H. Saber, and B. Pradhan (2018). Quantitative assessment of desertification in an arid oasis using remote sensing data and spectral index techniques. *Remote. Sens.* 10, 1862.
- Lary, D. J. (1999). Data assimilation: a powerful tool for atmospheric chemistry. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 357(1763), 3445–3457.
- Lary, D. J., B. V. Khattatov, and H. Y. Mussa (2003). Chemical data assimilation: A case study of solar occultation data from the ATLAS 1 mission of the Atmospheric Trace Molecule Spectroscopy Experiment (ATMOS). *Journal of Geophysical Research-Atmospheres* 108(D15).
- Lary, D. J., T. Lary, and B. Sattler (2015). Using machine learning to estimate global pm2.5 for environmental health studies. *Environmental Health Insights* 9s1, EHI.S15664. PMID: 26005352.
- Lary, D. J., D. Schaefer, J. Waczak, A. Aker, A. Barbosa, L. O. Wijeratne, S. Talebi, B. Fernando, J. Sadler, T. Lary, et al. (2021). Autonomous learning of new environments with a robotic team employing hyper-spectral remote sensing, comprehensive in-situ sensing and machine learning. *Sensors* 21(6), 2240.

- Lehtomäki, H., A. Korhonen, A. Asikainen, N. Karvosenoja, K. J. Kupiainen, V.-V. Paunu, M. Savolahti, M. Sofiev, Y. Palamarchuk, A. Karppinen, J. Kukkonen, and O. Hänninen (2018). Health impacts of ambient air pollution in finland. *International Journal of Environmental Research and Public Health* 15.
- Memarzadeh, F. and W. Xu (2012, March). Role of air changes per hour (ACH) in possible transmission of airborne infections. *Building Simulation* 5(1), 15–28.
- Mojtaba, E., M. Zeinab, and R. Mehravar (2022). Neurotoxicity related exposure to ambient nanoparticles. *Journal of Neuroscience and Neurological Disorders*.
- Mostafa, M. M. R. and K. P. Schwarz (2000). A multi-sensor system for airborne image capture and georeferencing. *Photogrammetric Engineering and Remote Sensing* 66, 1417–1424.
- Muller, R., M. Lehner, R. Muller, P. Reinartz, M. Schroeder, and B. Vollmer (2002). A program for direct georeferencing of airborne and spaceborne line scanner images. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* 34(1), 148–153.
- Neumark, D. M. (1992). Transition state spectroscopy of bimolecular chemical reactions. *Annual Review of Physical Chemistry* 43(1), 153–176.
- Ng, L. C., A. K. Persily, and S. J. Emmerich (2015, October). IAQ and energy impacts of ventilation strategies and building envelope airtightness in a big box retail building. *Building and Environment* 92, 627–634.
- Ni, Y., C. T. Loftus, A. Szapiro, M. T. Young, M. F. Hazlehurst, L. E. Murphy, F. A. Tylavsky, A. Mason, K. Z. Lewinn, S. Sathyaranayana, E. S. Barrett, N. R. Bush, and C. J. Karr (2021). Associations of pre- and postnatal air pollution exposures with child cognitive performance and behavior: A multi-cohort study. *ISSE Conference Abstracts*.
- Odabasi, M. (2008, March). Halogenated Volatile Organic Compounds from the Use of Chlorine-Bleach-Containing Household Products. *Environmental Science & Technology* 42(5), 1445–1451. Publisher: American Chemical Society.
- Partridge, H., C. W. Bauschlicher Jr, J. R. Stallcop, and E. Levin (1993). Ab initio potential energy surface for h–h<sub>2</sub>. *The Journal of chemical physics* 99(8), 5951–5960.
- Pease, L. F., N. Wang, T. I. Salsbury, R. M. Underhill, J. E. Flaherty, A. Vlachokostas, G. Kulkarni, and D. P. James (2021, June). Investigation of potential aerosol transmission and infectivity of SARS-CoV-2 through central ventilation systems. *Building and Environment* 197, 107633.

- Petneházi, G. (2019). Recurrent neural networks for time series forecasting. *arXiv preprint arXiv:1901.00069*.
- Rackauckas, C., D. Kelly, Q. Nie, J. Li, C. Warner, M. Dhairyा, J. Fang, E. Zhou, R. Supekar, S. Sandhu, et al. (2020). Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2012.09345*.
- Rasmussen, C. E., C. K. Williams, et al. (2006). *Gaussian processes for machine learning*, Volume 1. Springer.
- Rhew, S. H., J. Kravchenko, and H. K. Lyerly (2021). Exposure to low-dose ambient fine particulate matter pm2.5 and alzheimer's disease, non-alzheimer's dementia, and parkinson's disease in north carolina. *PLoS ONE* 16.
- Sakurai, J. J. and E. D. Commins (1995). Modern quantum mechanics, revised edition.
- Shehab, M. and F. D. Pope (2019). Effects of short-term exposure to particulate matter air pollution on cognitive performance. *Scientific Reports* 9.
- Srikrishna, D. (2022, November). Long-term experience with rapid air filtration (6 to 15 air changes per hour) in a K-5 elementary school using HEPA and Do-It-Yourself (DIY) air purifiers during the COVID-19 pandemic. preprint, Infectious Diseases (except HIV/AIDS).
- Strumbelj, E. and I. Kononenko (2013). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems* 41, 647–665.
- Takens, F. (2006). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pp. 366–381. Springer.
- Turakhia, Y., B. Thornlow, A. S. Hinrichs, N. De Maio, L. Gozashti, R. Lanfear, D. Haussler, and R. Corbett-Detig (2021, June). Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nature Genetics* 53(6), 809–816.
- Urrutia, P. J., D. A. Bórquez, and M. T. Núñez (2021). Inflaming the brain with iron. *Antioxidants* 10.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks* 5, 241–259.
- Wu, T., H.-J. Werner, and U. Manthe (2006). Accurate potential energy surface and quantum reaction rate calculations for the h+ ch4 → h2+ ch3 reaction. *The Journal of chemical physics* 124(16).
- Yu, X., D. J. Lary, and C. S. Simmons (2021). Pm2. 5 modeling and historical reconstruction over the continental usa utilizing goes-16 aod. *Remote Sensing* 13(23), 4788.

Zaatari, M., A. Novoselac, and J. Siegel (2016, May). Impact of ventilation and filtration strategies on energy consumption and exposures in retail stores. *Building and Environment* 100, 186–196.

Zheng, Q., W. Huang, X. Cui, Y. Shi, and L. Liu (2018). New spectral index for detecting wheat yellow rust using sentinel-2 multispectral imagery. *Sensors (Basel, Switzerland)* 18.

Zheng, W., J. Hu, Z. Wang, J. Li, Z. Fu, H. Li, J. Jurasz, S. Chou, and J. Yan (2021, August). COVID-19 Impact on Operation and Energy Consumption of Heating, Ventilation and Air-Conditioning (HVAC) Systems. *Advances in Applied Energy* 3, 100040.

## BIOGRAPHICAL SKETCH

**UPDATE REQUIRED!!!**

# **CURRICULUM VITAE**

**UPDATE REQUIRED!**