



Inference and Machine Learning across the Spatial Scales in Geophysics

Alexandre Szenicer

Linacre College

University of Oxford

A thesis presented for the degree of

Doctor of Philosophy

Trinity 2021

Copyright © 2021 by Alexandre Szenicer
All Rights Reserved

When we try to pick out anything by itself, we find
it hitched to everything else in the universe.

— John Muir

Acknowledgements

My time in Oxford has been one of growth, both intellectually and personally, for which I am indebted to many people.

First of all, I would like to thank my supervisor Tarje, for giving me advice and support when I needed it, but also great freedom to find my interests and pursue them.

I would also like to say a huge thank you to my friends at Linacre College, whose company has been the most wonderful thing about my PhD. To my 2016 crew of fellow PhDs: Dan, Abbie, Derek, Sapna, Kento, and co, for always sticking together through the years. To the residents of Griffiths ground floor 2017, the best floormates one could ask for: Omar, Peter, Joseph, Mehran, Alex and Farhana, for the long discussions in the kitchen running into the night, the meals, the arabic lessons, and all the great times. To many other Linacre friends, for simply hanging out and chatting in the CR, rowing, hiking, partying, and making Oxford such a great place to live in: Ashley, Angad, Oli, Emily, Cami, Tanvi, Chido, Gerardo, Salam, Nehir, Anna, Maya, ...

My deepest gratitude goes to my family, for always being there for me, for their encouragement and unwavering support, neither for the first nor the last time. And to Hailey, for the russian dumplings and TV shows, for all the nicknames, and for being such a big part of my life.

A heartfelt salute to the KAIP team, for the great times in the Sun, full of mango salad and dad jokes, especially during the bleakest winter months in Oxford.

A big thank you to the friends and colleagues from the Seismology group for the friendly and supportive work environment and the stimulating discussions; and in

particular office 30.13: Marta, Harriet and Eoghan, for going through the highs and lows together.

I would also like to gratefully acknowledge the support of the anonymous donor of my PhD fund, who has allowed me to spend such a wonderful time in Oxford.

Abstract

The Anthropocene and the concomitant Great Acceleration – defined by super-exponential growth rates in a collection of measures of human activity – are having a tremendous impact on Earth and its ecosystems. To better understand the consequences of our activity, we must increase our knowledge of all the elements in the Geosystem, loosely defined as Earth, the oceans, atmosphere, and the Sun, and in constant evolution as human activity ties in new elements, e.g. the Moon or Mars. In this thesis, I am focusing on improving our knowledge of elements pertaining to geophysics, while also showcasing the diversity of thematics in the Geosystem. I also emphasize the common grounds and cross-cutting themes in our sundry of projects, in particular the importance of a data-centric approach, including the creation of new sensors, and the necessity of both methodological improvements to our numerical tools, and the adoption of novel methods from machine learning to help us exploit the newly available vast amount of data. I begin with wildlife conservation, where, with collaborators, we prototype an elephant monitoring system based on the classification of animal generated seismic signals with deep learning. We achieve promising results, with high accuracies in classification of elephants and their behaviour, and with further research I believe our method will bring about a seismic shift to wildlife monitoring. Next, I present a novel method for the computation of Frechet sensitivity kernels for full waveform inversion based on the discrete adjoint approach, resulting in a more advantageous frequency scaling of the computational cost than traditional methods. This discount is crucial to enable us to image the interior of the Earth at higher resolutions, which is necessary to unlock the answers to some of the biggest questions in Earth science, e.g. related to the mechanisms of mantle convection and plate tectonics. Finally, I introduce how, with my teammates, we developed a deep learning replacement to a now defunct Sun-monitoring spectrograph

onboard a NASA satellite. The best performing model, a combined linear-CNN architecture, achieves 1.6% median error, virtually reviving the spectrograph and restoring a vital data source, for instance for satellite operators to model atmospheric dynamics and avoid catastrophic collisions.

Table of Contents

1	Introduction	1
1.1	The clock is ticking	1
1.2	An evolving field	3
1.3	Thesis Outline	6
1.3.1	Seismic savanna: Machine learning for classifying wildlife and behaviours using ground-based vibration field record- ings	7
1.3.2	A complexity-driven framework for waveform tomogra- phy with discrete adjoints	8
1.3.3	A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance	10
2	Seismic savanna: Machine learning for classifying wildlife and be- haviours using ground-based vibration field recordings	12
	Preface	12
	Significance statement	17
2.1	Introduction	18
2.2	Results	23
2.2.1	SeisSavanna: An out-of-the-box dataset	23
2.2.2	Classifying elephants versus other seismic sources	26
2.2.2.1	Generalising in time	26
2.2.2.2	Generalising to new stations	30
2.2.3	Classifying behaviour	31
2.2.3.1	Locomotion versus vocalisation	31
2.2.3.2	Walking versus running	32

2.3	Discussion	34
2.4	Methods	40
2.4.1	Fieldwork	40
2.4.2	Data processing and dataset creation	42
2.4.2.1	Camera trap data	42
2.4.2.2	Seismic data	43
2.4.2.3	Spectrograms	43
2.4.3	Modelling	43
2.4.4	Metrics	45
2.4.5	Data augmentation	45
2.5	Supplementary Material	46
2.5.1	Models details	46
2.5.2	Data generation algorithm	48
	Acknowledgments	50
	Author contributions	50
	Competing interests	51
	Data and materials availability	51
3	A complexity-driven framework for waveform tomography with discrete adjoints	52
3.1	Introduction	54
3.2	Theory	59
3.2.1	Forward problem: AxiSEM3D	60
3.2.2	Inverse problem	63
3.2.3	Continuous adjoint	66
3.2.4	Discrete adjoint	67
3.3	Continuous adjoint kernels	75
3.3.1	Physical adjoint wavefields	75

3.3.2	Continuous kernels benchmark	79
3.4	Discrete adjoint kernels	83
3.4.1	Computational cost	84
3.4.2	Event kernels	86
3.5	Discussion and conclusions	92
	Acknowledgments	94
	Postface	96
	Generating sensitivity kernels with deep learning	96
	Encoder-Decoder networks for seismogram super-resolution	99
4	A Deep Learning Virtual Instrument for Monitoring Extreme UV So-	
	lar Spectral Irradiance	104
	Preface	104
4.1	Introduction	108
4.2	Results	112
4.2.1	Data Setup and Evaluation Criteria	112
4.2.2	Baseline Models	114
4.2.3	Convolutional Neural Networks	118
4.2.4	Opening the black-box: Analysis of Models	123
4.3	Discussion	126
4.4	Materials and Methods	128
4.4.1	Models	129
	Acknowledgments	132
	Funding	132
	Author Contributions	132
	Competing interests	133
	Data and materials availability	133
	Supplementary material	133

5 Discussion and conclusions	135
-------------------------------------	------------

Appendix A: A Machine Learning Dataset Prepared From the NASA So-	
lar Dynamics Observatory Mission	142
A.1 Introduction	144
A.2 Examination of Raw Data Products	148
A.3 Processed Data Preparation	151
A.3.1 AIA	151
A.3.2 HMI	153
A.3.3 EVE	155
A.3.4 Temporal Downsampling and Synchronization	156
A.3.5 Data	157
A.4 Protocols and Metrics	157
A.5 Results	160
A.5.1 EVE-to-EVE Prediction	161
A.5.2 HMI-to-AIA Prediction	164
A.6 Conclusion	167
Acknowledgments	169
Data availability	169

List of Figures

2.1	A deployed seismometer in Mpala.	13
2.2	The author hammering away.	14
2.3	Three samples from animal generated seismograms, and the associated camera trap picture.	22
2.4	Some statistics of SeisSavanna.	25
2.5	Test set results of the finetuned SqueezeNet for datasets with increasing maximal distances, and for different test stations.	28
2.6	Test set results of the finetuned Squeezenet as a function of distance for the 150m dataset.	29
2.7	Spectrograms of elephants walking at two different distances.	29
2.8	Spectrograms of an elephant rumbling and an elephant walking.	31
2.9	Spectrograms of an elephant running and an elephant walking.	33
2.10	Fieldwork site.	41
3.1	Coordinate system and Fourier expansion in AxiSEM3D.	60
3.2	Constant and learned Fourier expansions.	76
3.3	Converging seismograms for increasing values of the n_u^{const} field in s40rts.	78
3.4	Kernels computed with AxiSEM3D in the continuous adjoint framework with the Fourier expansion n_u^{sup}	80
3.5	a) PREM kernel used in the benchmark. b) Azimuthal cut through 500km radius bubble at 5% strength used in benchmarks.	81
3.6	Traveltime perturbations predicted by kernels computed in PREM and S40RTS, compared with forward modelling prediction, for different perturbation geometries and strengths.	82

3.7	Seismograms obtained from an axial source, and the same seismograms with an off-axis source computed in the Fourier expansion learned for the 3D bubble.	85
3.8	Slices through continuous and discrete adjoint kernels.	86
3.9	Traveltime perturbations predicted by discrete adjoint kernels and continuous adjoint kernels, compared to cross correlations predictions.	87
3.10	Computational cost of the forward and the adjoint simulation for different seismic periods, in the continuous and the discrete adjoint approach.	88
3.11	Event kernel at 15s with a magnitude 5.8 earthquake in Virginia, and 5 receivers, in S40RTS + crust1.0.	90
3.12	Event kernel at 5s with a magnitude 5.8 earthquake in Virginia, and 3 receivers, in S40RTS, at epicentral distance 50°.	91
3.13	Ground truth and results from the trained CNN on validation data.	97
3.14	Models used to generate data for the super-resolution seismogram problem.	101
3.15	Results from the trained encoder-decoder CNN on the test set, generated from model 4.	102
4.1	Spectral and temperature response of AIA and EVE.	110
4.2	Proposed Linear-CNN Neural Network Architecture.	120
4.3	Emission prediction of FeXX line for several models.	123
4.4	Linear weights for the Huber Mean+Std model per line.	125
4.5	Spatially rearranged predictions.	126
4.6	Emission prediction of Fe XVIII line for several models.	133
4.7	Emission prediction of He II line for several models.	134
4.8	Emission prediction of Fe XI line for several models.	134

A.1	The average AIA wavelength band data counts for level 1 data products over time.	147
A.2	Example of a corrupt observation from the AIA instrument.	148
A.3	Mean intensity variation after correcting for exposure time, degradation, and orbital variation for AIA and EVE.	152
A.4	Co-aligned AIA and HMI dataset around a solar maximum and a solar minimum.	155
A.5	Results from the 2 hour EVE Prediction experiment.	159
A.6	Results for HMI to AIA translation.	164

List of Tables

2.1	Metrics for additional models.	48
3.1	Normalised mean absolute error between offaxis seismogram and axial seismogram.	77
4.1	Per-line average relative error for the methods evaluated in this paper.	116
A.1	EVE Emission lines, their wavelength, and temperature of the emission plasma.	154
A.2	Results for HMI to AIA prediction.	163
A.3	Dataset urls and Digital Object Identifiers	170

1 | Introduction

1.1 The clock is ticking

Understanding the Earth system has never been as important as it is today. Humanity is on an accelerating curve of open-ended growth across several measures of human activity (population, real GDP, energy use, telecommunications, ...) ([Steffen et al., 2015](#)), and this is increasingly jeopardising both our planet and ourselves. By gaining an unprecedented level of knowledge of the Geosystem composed of the Earth, oceans, atmosphere, and Sun, we can design a sustainable approach to the future.

Humanity's existence on Earth has been marked by paradigm-changing innovations happening more and more frequently, starting with Agriculture around 10000 years BCE ([Zeder, 2011](#)), to the 20th century and the atomic era, the invention of computers, transistors, the internet, etc. Concomitant to these revolutions, economic and lifestyle indicators such as GDP and life expectancy have followed a similar trajectory of super-exponential growth ([Riley, 2001](#); [Roser, 2013](#)). For instance, it took millions of years for the human population to reach one billion individuals at the dawn of the 19th century, then 123 years to reach 2 billion in 1927, 33 years to add another billion, and a mere 39 more years to double that to 6 billion in 1999 ([Bacci, 2017](#)). The present situation is that we are roughly 7.8 billion humans on the planet, and rapidly increasing, with an estimated population of about 10 billion by 2050 ([Roser, 2013](#); [Gerland et al., 2014](#)). What is more, we are a rapidly urbanising population. We recently passed the milestone of half of the world population being urbanised. By 2050, the figure is projected to be two thirds of the world population living in cities ([United Nations, 2019](#)). This is a frenetic pace of an average of three million people moving to cities every month!

Our increasing population coupled with fast urbanisation and with the increasing quality of life as proxied by life expectancy, health, wealth, etc; are putting an ever-greater strain on the planet. All these people need to be fed, they need houses, phones, entertainment, and a plethora of other things. This means that we must find land to build cities, land to plant crops and raise cattle, provide always larger quantities of food, mine ores and rare metals, drill for oil, and procure a multitude of other resources. There are myriads of deleterious consequences to this ever expanding lifestyle, both for the Geosystem and humanity. For instance, the increase in population and urbanisation has forced the creation of cities in zones exposed to natural disasters such as active faults, flood plains, or next to active volcanoes ([Wamsler, 2014](#)). To safeguard the lives of the inhabitants, it is vital to better our understanding of these systems, e.g. the likelihood of the fault rupturing and the resulting shaking map and infrastructural damage, or the probability of floods of a given severity and an accurate flooding model for the region at risk. There are also completely new and unexpected challenges that crop up at the boundaries of the Geosystem as we continue our urban and technological expansion. A good example is our innumerable satellites orbiting the Earth, providing us with internet, telecommunications, GPS, information on our planet, etc. Our modern lifestyle is entirely dependent on these satellites, and yet this system is surprisingly fragile. With the increasing number of spacecrafts in orbit, and the increasing number of them being decommissioned and turning into space junk, collisions that would destroy vital satellites are more and more likely ([Kessler and Cour-Palais, 1978](#); [Liou and Johnson, 2006](#)). Avoiding this scenario requires having a good understanding of the dynamics of the atmospheric layers that host these machines, because any changes in temperature and density are reflected in the orbits and thus must be compensated for. Taking this line of reasoning one step further, we can see how very soon the Geosystem will also absorb

extraterrestrial bodies, such as the Moon and Mars, that some institutions plan to colonize in the coming decades, and there are many competing theories for the best way to utilize these otherworldly resources. We could envisage using the Moon to outsource all our heavy industry and let Earth be a heaven of greenery, or maybe terraform Mars and create large human settlements that would lighten the population load on Earth. Whatever the way forward, the integration of these elements into the Geosystem will be paradigm shifting, and we need to acquire as much information and knowledge on these systems as possible. To this end, there are increasingly many missions that are sent into space to deploy instruments on other planets to understand their structure and their history. For instance, the In-Sight mission ([Panning et al., 2017](#)) has been designed to study the deep interior of Mars, in particular its 3D structure by deploying a seismometer, and its internal heat flow with a heat probe (during my PhD, I co-authored a paper exploring methods to detect and characterize marsquakes using simulated data, see [van Driel et al. \(2019\)](#)).

If left unaddressed, all these issues will only grow worse and eventually result in profound and disastrous consequences for all the actors involved. We must therefore understand the Geosystem better in order to address these questions and develop policies and technologies to avoid the worst case scenario developments.

1.2 An evolving field

The Geosystem is a complex dynamical system ([Bar-Yam, 2002](#)) with many moving parts that interact with one another in highly non-linear ways. We can view it as a giant causal graph ([Pearl, 2009](#)), whose nodes are elements of the Geosystem, e.g. atmospheric cells, volcanoes, oceans, faults, ecosystems, etc. It is particularly important to acknowledge the causal links between these nodes, because

modelling the subsystems in isolation can lead to drastic errors. Take for instance the previous example of flooding. Climate change is likely to increase the frequency and intensity of extreme events such as storms and hurricanes ([Knutson et al., 2010, 2020](#)), and not taking this into account in our modelling of floods in a given area could lead to dire consequences. Taken together with human society, the Geosystem becomes a complex adaptive system ([Lansing, 2003](#)), which not only has intricate dynamics, but whose nodes also evolve and adapt through time. For example, in response to climate change, a country may lean towards sustainable policies, or may blatantly disregard the problem, based on the will of the individual agents constituting that system (in a democracy, the people). What is required to address the issues that arise from humanity's growing interconnect- edness with the Geosystem is to have as good an understanding as possible of the different constitutive parts of this giant causal graph. In this way, we can assess with eyes unclouded the causal consequences of our actions on the en- vironment, but also understand the risk to our society and species. The study of these questions is fundamentally interdisciplinary, and involves fields such as physics, chemistry, social sciences, and biology. What is more, the processes at play range, on the spatial scale, from individual atoms (e.g. chemical reactions in the ocean/atmosphere/lithosphere) to the solar system (e.g. impact of the Sun on the atmospheric layers of the Earth); and on the temporal scale from fractions of seconds (e.g. lightnings) to billion of years (e.g. mantle convection). We will collectively refer to this melting pot of disciplines as the geosciences.

Because of the sheer size and complexity of many of the systems in the geo- sciences, this field has always been reliant on data and simulations to answer its questions. In certain subfields of the geosciences such as seismology, the ques- tions are additionally complicated by our inability to directly observe the interior of the Earth, and instead we must solve an inverse problem connecting physical

properties inside the planet, to observables on its surface ([Tarantola, 1984](#)). A similar situation arises in other subfields such as oceanography or atmospheric physics where the data fit of numerical methods is increased with variational data assimilation (VAR), in which initial conditions, boundary conditions, and various parameters are inverted for ([Huang et al., 2005](#)).

Since data and computational models are the main buttresses of geosciences, we must therefore focus on improving and expanding our sensor capabilities, as well as refining and accelerating our simulations, the latter both through methodological and hardware improvements. Progress on these fronts has been tremendous in recent years. Indeed, together with many other fields, the geosciences are undergoing a revolution because of the explosion in computational power, hardware capabilities, sensors and connected devices. For an example of new observations and sensors, there are now numerous satellites dedicated to Earth observation which are monitoring various aspects of the Geosystem. These instruments have data starting as early as the 1970s ([USGS, 2016](#)) on a daily basis and on spatial scales ranging from kilometres to, for the most recent instruments, as high a resolution as 50cm. Earth observing satellites have traditionally been deployed by large public institutions (e.g. NASA, ESA, JAXA) that make their data available for free, but there are many new private players entering the field (e.g. Planet Labs). Another example is the always expanding networks of seismic stations around the globe, for which data is mostly open source and even contains travel time picks for various seismic waves ([Willemann and Storchak, 2001](#)). These stations are now steadily supplemented with innovative instruments that allow to record waves in difficult environments, such as ocean bottom seismometers ([Bialas and Flueh, 1999](#)) and MERMAIDs ([Sukhovich et al., 2015](#)).

As for simulation software and the resulting synthetic data, many communities in the geosciences build and maintain highly performant open-source models. For

instance, the Community Earth System Model (CESM) ([Kay et al., 2015](#)) is a fully-coupled software simulating the Geosystem, as composed of atmospheric, ocean, ice, carbon cycle, and other components. In seismology, AxiSEM3D ([Leng et al., 2019](#)) is a software that allows to solve the wave equation and thus simulate earthquakes (or otherworldly quakes!).

Taken together, the progress in observations and simulations has led to a deluge of data, as a response to which the field must adapt and create new methods to incorporate and exploit all this information to its fullest. One of the most promising avenues to tackle the challenges of increasing datasets is machine learning (ML), which has already revolutionised many aspects of our lives, and could significantly contribute to our understanding of the Geosystem. While ML has been used in the geosciences since the 1990s ([Dowla et al., 1990](#); [Dysart and Pulli, 1990](#); [Dai and MacBeth, 1995](#)), new developments in methodology and GPU hardware, together with the vast new amounts of data, have propelled ML to the top contenders for solving new problems. ML is now pervading many aspects in geosciences, from automating tasks (e.g. earthquake detection ([Mousavi et al., 2020](#))), to solving inverse problems directly ([McCann et al., 2017](#)), or using ML to act as a replacement for expensive simulations ([DeVries et al., 2017](#); [Smith et al., 2020](#)).

1.3 Thesis Outline

In this PhD, I have been pursuing the goal of improving our understanding of different elements of the Geosystem through some of the avenues mentioned above. In particular, I have been focusing on projects in which we can improve and expand our sensor capabilities, create methodological improvements in numerical modelling and inverse problems, or use machine learning to exploit vast amount of data. What is more, the projects tackled in this PhD reflect the richness and di-

versity of the Geosystem, and its interconnectedness with various disciplines such as biology, conservation policy, astrophysics, engineering, etc. They span orders of magnitude in spatial scales, from a few tens of meters, to the whole Earth, and finally the Sun-Earth system. And yet they are all coupled together and form an integral part of the Geosystem, and each one of them is every bit as crucial as the next in order for us to have a holistic, sustainable and informed approach to our future.

1.3.1 Seismic savanna: Machine learning for classifying wildlife and behaviours using ground-based vibration field recordings

In chapter 1, we will describe how we used seismometers and deep learning in order to design an elephant monitoring system, which exemplifies the crunch between humanity and other ecosystems on Earth.

Our planet is facing its sixth mass extinction, with hundreds of species disappearing largely because of human activity ([Barnosky et al., 2004](#); [Bellwood et al., 2004](#); [Scheele et al., 2019](#); [Ceballos et al., 2020](#)). In order to mitigate this existential threat of biodiversity loss, it is vital to monitor and thereby understand and protect endangered species, from the small golden frog of Panama, to elephants and other megafauna in their habitat. Large endangered land species are usually monitored using GPS collars ([Galanti et al., 2000](#); [Blake et al., 2001](#); [Ngene et al., 2010](#)) or camera traps ([Kays et al., 2009](#); [Smit et al., 2019](#)), and, while vital, these methods all have drawbacks. We propose to use seismic data for wildlife monitoring, to detect footsteps, ground-coupled vocalisations, and other seismogenic animal behaviours. Seismic data boast several crucial benefits compared to other methods: seismometers record with 360° azimuthal range, do not

encounter line of sight issues, and can detect wildlife signals up to hundreds of meters away. Seismic instruments do not interfere with animals in any manner as they are buried and can be deployed in concealed areas without affecting sensitivity, which also reduces the chances of sensor damage caused by the environment or fauna. Finally, low data volumes delivering information-rich, spatio-temporal and multiscale signals, combined with modest power requirements make them ideal for prolonged autonomous deployments. We present a novel approach using convolutional neural networks to detect elephants and classify their behaviours with high accuracy by using seismic field data acquired from a multisensor deployment in Kenya. Crucially we open-source SeisSavanna: a unique dataset of seismic signals from wildlife, curated specifically for machine learning applications.

1.3.2 A complexity-driven framework for waveform tomography with discrete adjoints

In chapter 2 a novel approach to seismic tomography is presented, which utilises the smoothness of wavefields, and discrete adjoints, to speed up seismic inversion. Seismic tomography produces 3D models of the interior of our planet, whose structures are a cornerstone input for geological and geophysical studies that aim to explain the creation, history, and dynamics of our planet, such as mantle convection ([Van der Hilst et al., 1997](#); [Ritsema and Van Heijst, 2000](#)), uprising mantle plumes ([Bijwaard and Spakman, 1999](#); [Montelli et al., 2004a](#); [Zhao, 2004](#); [French and Romanowicz, 2015](#)), continental evolution ([Ritsema et al., 1999](#)), reconstruction of tectonic plate configurations ([Van der Voo et al., 1999](#); [Van Der Meer et al., 2010](#); [Sigloch and Mihalynuk, 2013](#)) and thermal/compositional components of structural heterogeneities ([Schuberth et al., 2009b,a](#)).

Full waveform inversion (FWI) ([Tarantola et al., 1982](#); [Tromp et al., 2005](#); [Fichtner, 2010](#)) is the most physically accurate of the currently existing deterministic approaches to seismic tomography, but also the most computationally expensive. FWI starts out with an initial Earth model, computes synthetic seismograms by solving the wave equation, and compares these synthetics to real seismograms recorded on seismic stations around the globe. From the misfit of synthetic and real seismograms, we can compute gradients (called Frechet sensitivity kernels in seismology) to iteratively update the Earth model. A global scale FWI at the resolutions required to help answer the most pressing questions in the field, such as the precise nature of mantle convection or the origins of plate tectonics ([Council et al., 2008](#)), etc, is currently computationally out of reach, with costs in the order of billion of CPU hours. Our novel approach exploits the azimuthal smoothness of seismic wavefields to compute Frechet kernels with a much more advantageous frequency scaling than traditional methods, and paves the way to answering new questions in seismology and downstream fields, in order to better our understanding of the solid Earth.

At the end of the chapter, in a postface, we also discuss two early stage projects that have been shelved for now, which explore applying deep learning methods to speed up the forward and inverse problems. These works are important parallel investigations to more traditional approaches to speed up computations such as the discrete adjoint method we introduce in this chapter. In the first project we try to generate sensitivity kernels using convolutional neural networks (CNN) from only a few quantities such as source and receiver position, moment tensor, and measurement. The second project deals with frequency upscaling of seismograms, whereby we use encoder-decoder CNNs to map a low frequency seismogram to a high frequency one.

1.3.3 A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance

Finally, in chapter 3 we present a deep learning based virtual instrument, which uses images of the Sun to produce solar irradiance power spectra which are very important, for instance for satellite operators to perform orbital adjustments.

The extreme ultraviolet (EUV) radiation from the Sun is the dominant driver of the Earth's thermosphere/ionosphere system. During periods of elevated solar activity, enhanced solar EUV driving causes adverse space weather effects such as radio communication blackouts, increased aerodynamic drag on satellites in low-earth orbit, and scintillation of Global Navigation Satellite Systems (GNSS) signals ([Lean, 1987](#); [Tobiska et al., 2000](#); [Solomon et al., 2010](#)). NASA's Solar Dynamics Observatory ([Pesnell et al., 2011](#)) was launched in 2010 with the goal to help us understand solar activity and how their variations affect life on Earth. SDO carries three instruments. The EUV Variability Experiment (EVE) instrument suite ([Woods et al., 2010](#)) provides EUV spectral solar irradiance (SSI) measurements integrated over the full Sun. The Atmospheric Imaging Assembly (AIA) instrument ([Lemen et al., 2011](#)) captures full-disk images of the Sun at 4096 x 4096 pixels in seven EUV channels, two UV and one visible wavelength channels. The Helioseismic and Magnetic Imager (HMI) ([Scherrer et al., 2012](#)) delivers photospheric vector magnetograms and Dopplergrams of the full Sun at 4096 x 4096 pixels. All three instruments continue to co-observe the Sun. However, an electrical malfunction of the EVE MEGS-A channel compromised our ability to monitor EUV SSI. The goal of this work is to fill this gap in measurement capabilities with a virtual replacement for MEGS-A. Toward this goal, we examine how well physics-based models and learning-based models can use narrowband AIA images to reproduce MEGS-A EUV SSI measurements. Our best

performing ML model based on Convolutional Neural Networks (CNNs) outperforms other ML and physics-based models, yielding average relative errors of under 4.6% (maximum error over all MEGS-A emission lines) and more typically 1.6% (median over all lines). Additionally, we provide evidence that the proposed method is solving this mapping in a way that makes physical sense and by paying attention to the magnetic structures known to drive EUV SSI variability.

2 | Seismic savanna: Machine learning for classifying wildlife and behaviours using ground-based vibration field recordings

Preface

Before I include the paper for this chapter, I would like to preface this work by providing some wider context about the project. The venture undertaken in Kenya is an ambitious multidisciplinary attempt to better understand elephant behaviour, as well as to develop novel methods for monitoring that will contribute to elephant conservation. While the material in this paper is mostly focused on prototyping the monitoring system, the project as a whole has numerous branches of investigation which can provide a number of tantalising insights into elephants. Central to these investigations is the dataset acquired in Kenya over five weeks of fieldwork, which I had the incredible opportunity to take part in (and co-lead for several weeks). In this fieldwork, we deployed seismometers, camera traps, as well as microphones, in the Mpala Research Center in Kenya (see figure 2.1 for a picture of a deployed seismometer).



Figure 2.1: A deployed seismometer. The instrument is buried inside a cage to protect it from animals.

While this chapter focuses on utilising only a subset of the whole data, i.e. chunks of seismograms and camera trap labels to design a monitoring system; there are other very interesting research directions enabled by this dataset. For example, another study has focused on designing an elephant localisation system by triangulating elephant rumbles recorded on microphones and seismometers ([Reinwald et al., 2021](#)). An interesting segue from these signal propagation papers, is to explore their implications for animal communication. Since we can record footsteps and rumbles propagating through the ground over large distances, do elephants perhaps use these signals to communicate? Another direction of research is to try and utilize the seismometers to image the local subsurface, either by using controlled sources (e.g. we have been hammering the ground in many locations,

providing reference sources, as can be seen in figure 2.2), noise tomography, or perhaps even using animal generated signals. Determining the local velocity structure could in turn greatly improve the capabilities of our monitoring systems, for instance by allowing us to triangulate animals detected on several seismometers.



Figure 2.2: The author hammering away. We hammered in many different locations to provide some relatively controlled sources which could potentially be used for imaging.

Submitted to PNAS.

Note: We have not included links to codes and dataset, as we are awaiting publication to release them.

Alexandre Szenicer¹, Michael Reinwald², Ben Moseley³, Tarje Nissen-Meyer^{1,1}, Zacharia Mutinda Muteti⁵, Sandy Oduor⁵, Alex McDermott-Roberts², Atilim Gunes Baydin^{3,4}, Beth Mortimer²

Departments of: ¹ Earth Sciences, ²Zoology, ³Computer Science, ⁴ Engineering Science, University of Oxford, UK.

⁵ Mpala Research Centre, Kenya.

¹corresponding author: tarje.nissen-meyer@earth.ox.ac.uk

Abstract

We develop a machine learning approach to detect and discriminate elephants from other species, and to recognise important behaviours such as running and rumbling, based only on data from recordings of seismic energy generated by the animals. We demonstrate our approach using data acquired in the Kenyan savanna, consisting of 8000 hours seismic recordings and 250k camera trap pictures. Our classifiers, different convolutional neural networks trained on seismograms and spectrograms, achieved 80–90% balanced accuracy in detecting elephants up to 100 meters away, and over 90% balanced accuracy in recognising running and rumbling behaviours from the seismic data. We release the dataset used in this study: SeisSavanna represents a unique collection of seismic signals with the associated wildlife species and behaviour. Our results suggest that seismic data offer substantial benefits for monitoring wildlife, and we propose to further develop our methods using dense arrays that could result in a seismic shift for wildlife monitoring.

Significance statement

Our planet is facing its sixth mass extinction, with hundreds of species disappearing largely because of human activity. To mitigate this existential threat of biodiversity loss, we must monitor and protect endangered species, from the small golden frog of Panama, to elephants and other megafauna. We present a novel approach using deep learning to detect elephants and classify their behaviours with high accuracy by using seismic field data acquired from a multisensor deployment in Kenya. Our results showcase the promising potential of seismic data to be used in conservation to monitor animals in near real-time, and alert to behaviours of concern. Crucially, we open-source SeisSavanna: a unique dataset of seismic signals from wildlife, curated specifically for machine learning applications.

2.1 Introduction

The ecosystems on our planet are facing an existential crisis, which authors increasingly call the sixth extinction ([Barnosky et al., 2004](#); [Ceballos et al., 2020](#)). In mere decades, hundreds of species have become either endangered or extinct because of human activity, either directly through hunting, or indirectly through side effects such as ocean acidification and spread of invasive species ([Ceballos et al., 2020](#)). For instance, about 90 amphibian species have disappeared in the last few decades, with a total of about 500 species on the decline ([Scheele et al., 2019](#)), or worldwide declines of coral reefs that support some of the most diverse ecosystems in otherwise barren tropical waters ([Bellwood et al., 2004](#)). Mammals, too, are adversely impacted and increasingly jeopardised, and many hold the unenviable position of being among the most famous endangered species, such as giant pandas, tigers, or blue whales.

Of particular interest to us are African elephants (*Loxodonta Africana*), who are poached for their ivory that is falsely believed to have healing powers and reaches high prices on the black market ([Wasser et al., 2008](#)), but are also killed because of increasing human-elephant conflicts. Populations of African elephants have plummeted since the 1980s, with over a million of individuals estimated in the 1970s, down to about 250 000 in the late 80s ([Douglas-Hamilton, 1987](#)). After a recovery until the mid 2000s thanks to new laws and conservation efforts, the numbers are now once again on the decline and the species is classified as vulnerable, with about 350 000 individuals on the African continent ([Chase et al., 2016](#)).

A crucial aspect of elephant conservation is real-time monitoring of the animals and their behaviour, for instance to understand if a herd is getting dangerously

close to a human settlement, or if the animals start running, which could signify an attack by poachers (Sukumar, 2003). There are several approaches currently employed to monitor elephants that each have strengths and weaknesses. One possibility is to fit individual GPS collars to animals (Galanti et al., 2000; Blake et al., 2001; Ngene et al., 2010), which allows to track them in real time, but unfortunately this is very expensive as it requires deploying a specialised team to tranquilise the elephant and fit the collar. A cheaper and commonly used option is to deploy camera traps, which trigger through an infrared motion sensor (Kays et al., 2009; Smit et al., 2019). A drawback of that method is the limited detection range of the infrared sensor, typically an angle of roughly 60° and a range of about 30m (Randler and Kalb, 2018) in open terrain. The sensor is further influenced by environmental conditions such as temperature, and limited visibility due to vegetation and topography. One could perhaps deploy continuously recording cameras that do not have triggering mechanisms, but then again one runs into difficulties with obstacles to the field of view, massive data volumes, and expensive computations, which are all drawbacks for real-time monitoring. There are also promising new airborne methods emerging for real time monitoring by using drones (Mangewa et al., 2019), but they also have limitations such as field of view obstruction and environmental disruption due to noise.

In this study, we investigate the possibility of using seismometers and machine learning to distinguish elephants from other species, as well as classify their behaviour, from the ground motion resulting from their footsteps. Traditionally, seismic recordings belong to the realm of geophysics and seismology, where they are used to determine earthquake properties and the structure of the planet. They have led to discoveries such as the layered structure of the Earth (Oldham, 1906), the outer and inner core (Jeffreys, 1926; Lehmann, 1936), and the Moho (Mohorovicic, 1909), either by direct inspection of recorded seismograms after earthquakes,

or deriving 3D images of Earth’s interior through inversion of data in a process called seismic tomography for seismic traveltimes ([Dziewonski et al., 1977](#)) or full waveforms ([Hosseini et al., 2020](#); [Szenicer et al., 2020](#)). Additionally, seismic recordings have also been employed for footfall detection in humans ([Clemente et al., 2019](#)) and they seem a natural candidate for detecting much larger animals such as elephants. [Mortimer et al. \(2018\)](#) have performed modelling with limited recorded data that suggests that, depending on behaviour and geology, seismic signals from elephants can be detected at distances up to hundreds of meters or even kilometers, and thus potentially have much larger ranges than camera traps. What is more, seismometers are not limited by obstacles in the line of sight, nor do they have an azimuthal range limit. Finally, because of their low data volume and power requirements, seismic stations are routinely deployed for months at a time without the need for maintenance or inspection, and therefore can be good candidates for monitoring in remote locations. All of the above suggests that using seismic data can be a promising approach to real-time monitoring, with potentially significant benefits over existing methods. Some work has been done in this direction, for example [Wood et al. \(2005\)](#) have employed geophones and seismic data to recognise elephants, but on a limited scale, with only a few dozen datapoints, and handcrafted features extracted from the seismic data. Crucially, they only had one seismic station, and therefore could not investigate the generality of their method. [Sugumar and Jayaparvathy \(2013\)](#) have also investigated using seismic data for elephant classification, but have only tested their method on a few dozen simulated datapoints. [Lamb et al. \(2021\)](#) have recently deployed Raspberry Shake sensors in South Africa to assess their viability for monitoring seismic vocalizations and locomotion of elephants, but have not designed algorithms to detect and classify the animals. All these results are encouraging, and in this study we take this idea to the next step, with large amounts of field data con-

sisting of seismic and camera-trap recordings, and a fully data-driven automated deep learning approach.

Deep learning has pervaded many aspects of our lives in recent years, and has achieved great successes in fields as varied as natural language processing ([Brown et al., 2020](#)) and medical diagnostics ([Shamout et al., 2019](#)). It is also increasingly used in the sciences, with results such as solving protein folding ([Senior et al., 2020](#)), virtually fixing satellite instruments ([Szenicer et al., 2019](#)), or learning to solve the seismic wave equation ([Moseley et al., 2020](#)). Deep learning is a natural candidate to tackle complex unstructured data such as seismograms, and has already been employed on seismic data to detect earthquakes ([Mousavi et al., 2020](#)), estimate their magnitude ([Mousavi and Beroza, 2020](#)), and classify seismic phases ([Mousavi et al., 2020](#)). It is steadily more being adopted in conservation, for example for counting elephant populations from high resolution satellite imagery ([Duporge et al., 2020](#)), or for detection of whales from their vocalisations ([Shiu et al., 2020](#)), and we believe it is vital to further incorporate this technology into conservation efforts.

We conducted field work at the Mpala research centre in Kenya, where we deployed 22 seismometers and 30 camera traps running continuously for three weeks, in February-March 2019. The cameras are deployed in the vicinity of some of the seismometers, and constitute the basis for defining labels for the seismic signals. We acquired a rich dataset of more than 8,000 hours of seismic recordings and over 250,000 images, which were later annotated for species and animal behaviour, so that for each image containing an animal we can extract the corresponding seismogram (see figure 2.3). In figure 2.3, the seismograms exemplify the diversity and complexity of the signals present in our data. We can see a clear difference between the giraffe and the elephant seismograms, the hooved footsteps producing a clearer, more impulsive signal. And yet the relatively small and

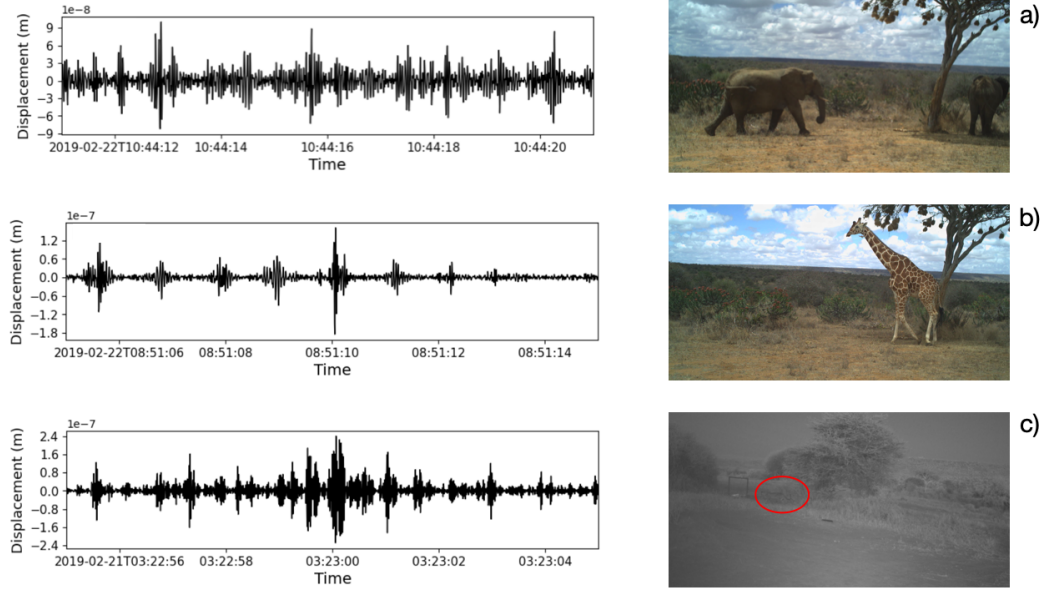


Figure 2.3: Camera trap pictures providing animal species and behaviour labels, and the corresponding vertical component seismograms extracted from nearby seismometers. The amplitude scales are purposefully not equalized between the plots, so that details of each seismogram can be distinguishable. In panel a), we can see footsteps generated by two elephants on their way to a tree, whose shade provides relief from the heat, and its bark from itches. In b) in the same location, a giraffe walking. In panel c), a leopard sneaking past a seismometer at night.

lightfooted leopard, simply by its happenstance close proximity to a seismometer, could easily be mistaken for a five ton elephant. One must therefore bear in mind the central roles played by several variables such as distance, soil and geology, number of animals, behavior etc.

Using this dataset, we develop deep learning classifiers which use convolutional neural networks (CNN) to recognise elephants and their behaviour from the spectrograms of the seismic signals, which represent the temporal evolution of the frequency content of the signal. In particular, we train a classifier to recognise elephants with good accuracy up to 150 meters when trained and tested on the same set of seismometers. We also demonstrate that the network can recognise

elephants on previously unseen seismic stations, albeit with lower accuracy when environmental conditions are very different from the training set, e.g. on geographically distant stations with different terrains. Finally, we train a CNN to distinguish running elephants from walking ones, and a CNN to detect elephant rumbles (vocal sounds) which couple to the ground. Both signal types represent important behaviours to monitor from the perspective of conservation and biology. Our dataset and code to replicate the results are fully open source and available: <https://github.com/AlexSzen/SeisSavanna>

2.2 Results

2.2.1 SeisSavanna: An out-of-the-box dataset

One of the principal contributions of this paper is to open source the dataset used in this work, which we name SeisSavanna. The data are in netcdf format (Rew and Davis, 1990), and are separated into three main files. The first file is focused on the task of species classification. It is composed of seismograms, the corresponding picture that generated the sighting, the species (i.e. the label), and other information such as seismometer name and distance of the animals from the seismometer. We provide a “master” dataset that includes animal sightings up to 150 meters from any given seismometer. We also provide scripts that easily reduce this dataset to any desired maximal distance, such as 40 or 60 meters. Likewise, we include sightings from all available seismometers, as well as tools to select desired subsets of stations. In this way, we want to provide maximal flexibility to the user and enable creative applications. The second netcdf file is specific to elephant behaviour, and only contains elephant sightings up to 150 meters away, with the label being either "running" or "walking". Finally, the last netcdf file is focused on the task of detecting elephant rumbles recorded on seismometers, and

as such the class labels are "rumble" or "not rumble". The rumble signals do not have distance information, while the non-rumble signals come from sightings of animals up to 150 meters away. To the best of our knowledge, SeisSavanna is the first dataset of its kind, which contains large amounts of seismic data and images from wild animals, made open-source, and curated for ML applications.

In figure 2.4 we provide some statistics about the species classification file in SeisSavanna. In particular, we plot the data counts with respect to station, distance from seismometer, and species. We can see that most of the events come from the stations within a 200 meters radius from the waterhole, since this location is both the most instrumented with camera traps, but also sees the heaviest animal traffic being one of only two locations within the area bearing water. The two more remote stations, NWP05 and>NNL62, have relatively low amounts of data. There is also a class imbalance, where the majority of datapoints come from elephants (about 60,000), with other species having from a few dozen to a few thousand samples. We also plot several examples of spectrograms coming from different species in panel d).

It is interesting to observe the differences in signal shape, and relate them to biological features. For example, we can see a clear difference in the frequency of footsteps between a relatively small animal (e.g., human, warthog, hyena) and a giraffe or a hippo. Similarly, there are noteworthy differences in the frequency content of the signal. For example, some signals demonstrate more impulsive sources with a relatively white spectrum (e.g., giraffe or human), whereas others exhibit a softer original impact (e.g., elephants), with a more concentrated spectrum. Similar patterns in the spectra of footfalls have been observed by (Wood et al., 2005). However, when making this observation, it is important to bear in mind the impact of signal propagation (and thus distance) on seismic data, which progressively removes high frequencies. A detailed description of our data col-

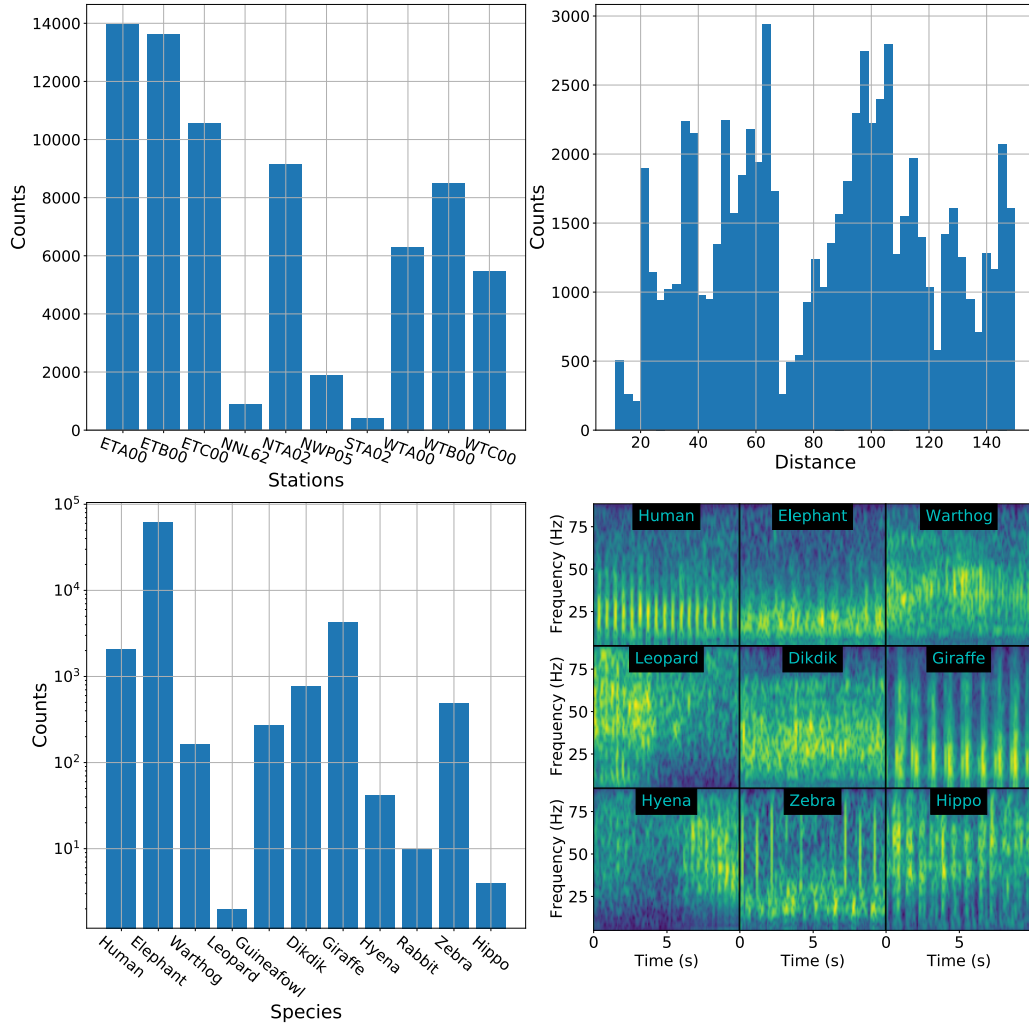


Figure 2.4: We provide a dataset of 10s long seismic chunks that contain labels of animals located up to 150 meters distance from any given seismometer. For each chunk, we provide the class (species), the station it is taken from (see Fig. 2.10 for a map), the distance from the station, and the image that generated the label. Panel a) shows the data count for each station. We can see that the majority of the data comes from around the waterhole, particularly the ET and WT stations. Panel b) shows the distribution of data with respect to distance. Panel c) displays the class counts, where the y axis is on a log scale, given the overwhelming majority of elephant signals. Finally, panel d) shows the spectrograms of several events from different classes in the dataset.

lection process is given in the Methods section.

2.2.2 Classifying elephants versus other seismic sources

2.2.2.1 Generalising in time

First we focus on the task of discriminating elephants from other species, while keeping environmental conditions mostly constant, by training and testing on the same set of seismic stations. We select a subset of stations (ETA00, WTA00, STA02, NTA02, NWP05, and>NNL62) for which the data are sorted in time. We then select the first 15% of the data for the test set, the following 15% for validation, and the remaining data for training. This split in time allows us to have independent training/validation/test sets. Indeed, we cannot simply randomise our samples before splitting, as is common in many machine learning applications, since the samples are not independent and identically distributed (i.i.d). We report the results for several maximal distances of animals in the dataset, viz. 40, 60, 80, 100, and 150 meters, which allows us to probe the limits of our detection range.

We trained different machine learning models on this task for the 60 meters dataset: several baselines including a logistic regression, gradient boosted trees, and a multi-layer perceptron, all on flattened seismograms/spectrograms. Then we trained an all-convolutional 1D-CNN inspired by [Springenberg et al. \(2014\)](#) on the seismogram data, an all-convolutional 2D-CNN on the spectrogram data, and finally a Squeezenet ([Iandola et al., 2016](#)) either trained from scratch, or pretrained on ImageNet and finetuned on spectrograms. The last approach is motivated by research in environmental sound classification, in which ImageNet pretrained models deliver state of the art results ([Palanisamy et al., 2020](#)). Details of the models' architectures and hyperparameters used during training can be found in the supplementary materials.

We report three metrics for the classification tasks in this paper: true positive rate (elephant accuracy in this section), true negative rate (non-elephant accuracy in this section), and balanced accuracy (average of true positive rate and true negative rate), which provides an interpretable metric for imbalanced classification tasks. See Methods section for more details.

In this paper, we only report results from the best performing model on the task, but results from additional models are provided in the supplementary material. For the elephant/non-elephant classification with generalisation in time on the 60m dataset, the best performing model is the finetuned Squeezenet trained on spectrogram data, which we therefore use on the datasets for other maximal distances. In figure 2.5 a), we plot the balanced accuracy achieved by the network on the test set of datasets with increasing maximal distances. We can see that the model achieves close to 90% balanced accuracy up to 80 meters distance, which then decreases to 82% at 100 meters, and decreases further to 73% for 150 meters. The drop in accuracy is further confirmed by inspecting classification accuracy as a function of distance, as can be seen in figure 2.6. The plot suggests that past around 100 meters, the accuracy drops off, likely because increasingly more seismograms have lower signal to noise ratios. This drop in accuracy is due to the labels becoming corrupted, i.e. a signal labelled elephant while it is only noise on the seismometer, as can be seen in figure 2.7. The extent of signal propagation depends on environmental conditions such as force of the impact, number of animals, local geology, etc. As such, label corruption can also happen at closer ranges, but on average increases with distance from the seismometer.

We believe that these results are very promising, and they showcase the great potential of using seismic data to accurately detect elephants up to relatively large distances: around 100 meters with our current dataset.

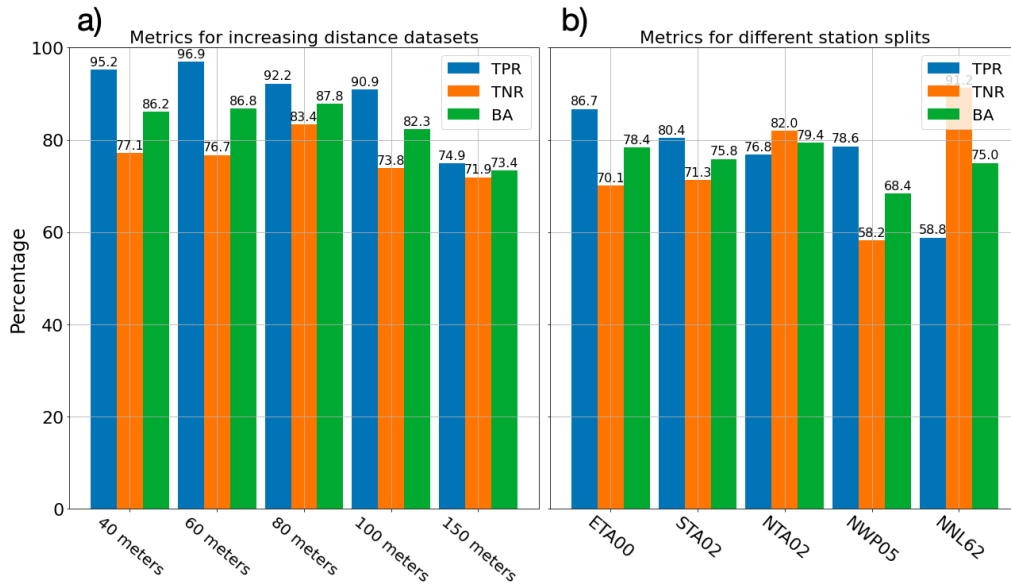


Figure 2.5: a) true positive rate (TPR, elephant accuracy), true negative rate (TNR, non-elephant accuracy), and balanced accuracy (BA, average of elephant and non-elephant accuracies) achieved by the finetuned Squeezenet on the test set for datasets with increasing maximal distance of animals. b) true positive rate (TPR, elephant accuracy), true negative rate (TNR, non-elephant accuracy), and balanced accuracy (BA, average of elephant and non-elephant accuracies) achieved by the finetuned Squeezenet on the test set for different test stations, which probes the ability of the network to generalise to new environmental conditions. We annotate the histograms with the value of the metric for easier interpretation.

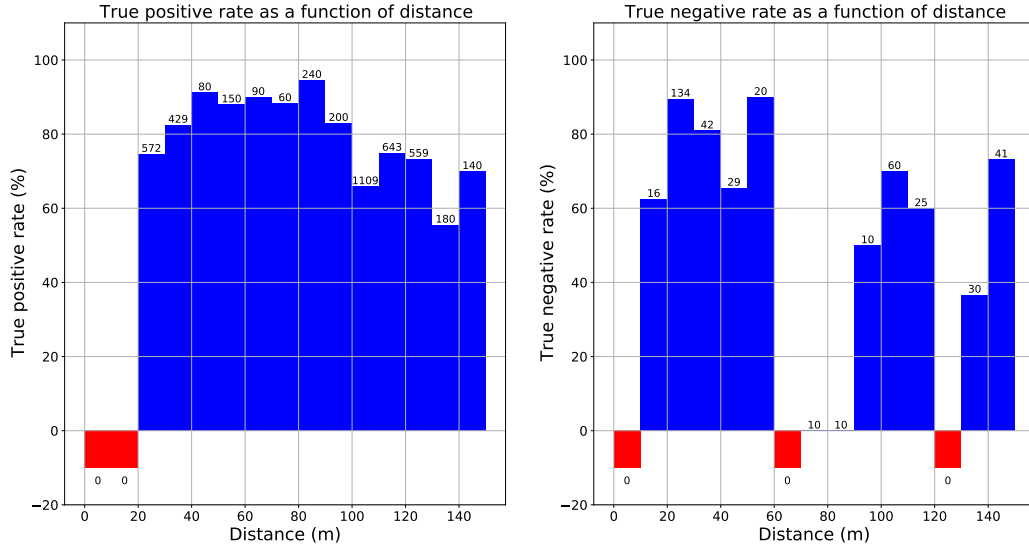


Figure 2.6: True positive rate (elephant accuracy) and true negative rate (non-elephant accuracy) as a function of distance on the test set of the 150 meters dataset, for a finetuned Squeezenet. We annotate the histogram with numbers of samples to clarify the absence of data for certain distances.

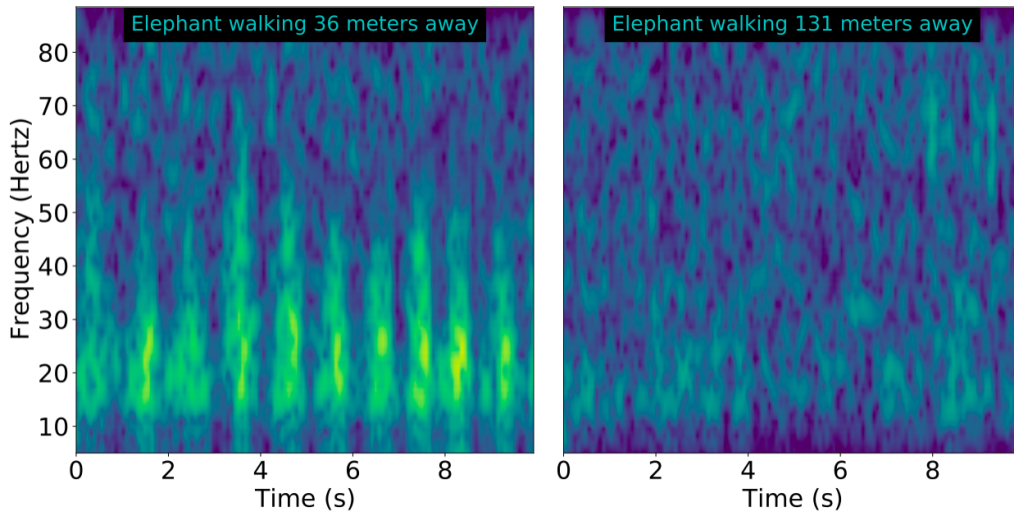


Figure 2.7: Spectrograms of two elephant samples, recorded on the station NWP05 at different distances. We can see that for the elephant walking 131 meters away, the camera trap provides an elephant label, but the seismic signal becomes so faint after propagation that the datapoint looks like noise.

2.2.2.2 Generalising to new stations

In this section, we still classify signals into elephant/non-elephant, but in a more ambitious training scenario. Indeed, we train on a subset of stations, and test on a different subset. Achieving across-station generalisation would in theory allow us to only conduct data-gathering fieldwork once, and then add new stations to our existing network, or even perform the analysis with this same pretrained network in new field deployments and locations. Each subset uses a new seismometer as a test set, in order to investigate generalisation ability to different environments. It is noteworthy to appreciate that the terrain across our 22 stations varied quite significantly, including topographic and vegetation changes, soil conditions from wet mud to unconsolidated sand and rocky terrain. All the splits are for a maximal distance of 60 meters, because we want to decouple the effect of the environment from effects such as label degradation with distance mentioned in the previous section.

The best performing model for the across-station splits is the finetuned Squeezenet. In figure 2.5b), we plot the balanced accuracy, true positive rate (elephant accuracy) and true negative rate (non-elephant accuracy), for the test set of each station split. We still achieve good balanced accuracy, albeit lower than when generalising in time. As expected, the network achieves lower balanced accuracy when the test set is NWP05 or>NNL62, compared to the other splits. This is noteworthy because these two stations are the furthest away from the waterhole where most of the data comes from, and have therefore the most different geological and environmental conditions, for instance hard rocky ground at>NNL62, as opposed to wet muddy conditions at the waterhole. This suggests that, as expected, we need to acquire data from a varied collection of environmental conditions to improve our generalisation capabilities.

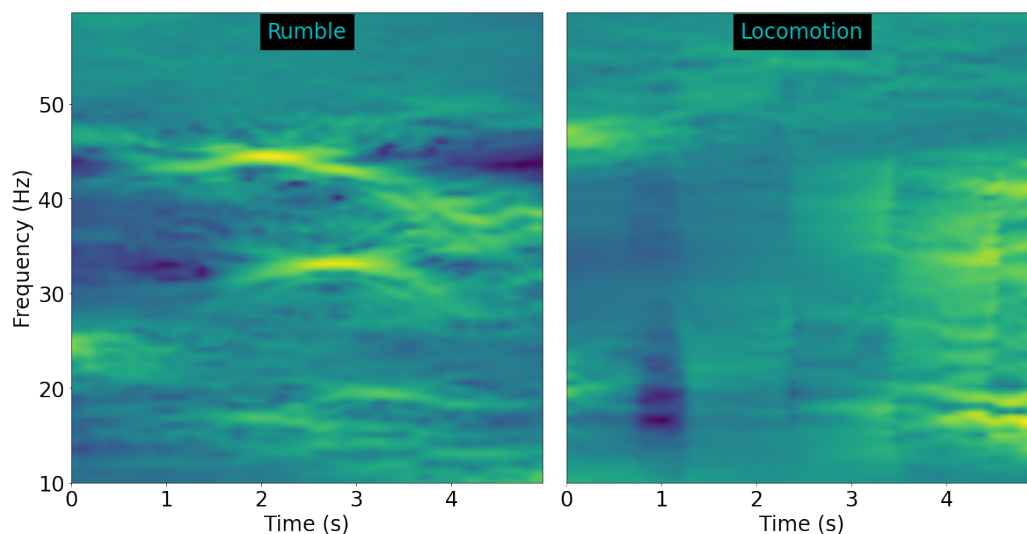


Figure 2.8: Spectrograms of an elephant rumble (left) versus elephant locomotion (right). The rumble has a very distinctive spectral signature, which is characteristic of a vocalisation signal.

2.2.3 Classifying behaviour

2.2.3.1 Locomotion versus vocalisation

Besides recognising and tracking elephants, a central objective of a monitoring system is to be able to alert authorities to potentially perilous situations for the animals. A behaviour of interest is elephant vocalisation, in particular in the form of rumbling, because they couple to the ground and can be observed in our seismic signals. Elephants use these vocalisations as communication, e.g., for greetings, warnings about imminent threats, or communicating movement (Poole, 2011). While there is research looking at automatically detecting and classifying elephant vocalisations (Leonid and Jayaparvathy, 2020), it is performed with small sample sizes, and crucially on acoustic data from microphones. To the best of our knowledge it is the first time this is done from seismic signals, with large amounts of data, and using deep learning.

Because rumbles cannot be identified from images, we manually went through a portion of the data to look for vocalisations. The rumbles have a characteristic shape that is easily recognisable, and in stark contrast to an impulsive locomotion signal, as we can see in figure 2.8. We therefore manually label 1,500 rumbles on two seismic stations (ETA00 and EEL11), and combine these into a dataset with locomotion signals from all other species together (elephants, giraffes, etc). The spectrograms are processed differently than in the previous section, and we apply the structure tensor (Harris et al., 1988) of the spectrogram as a filter. This approach enhances the rumbles which contain sharp contours along the frequency axis, and diminishes the broadband locomotion signals.

The finetuned squeezenet achieves the best performance on the rumble classification task, with 96.1% balanced accuracy, of which the true positive rate (rumble accuracy) is 97.3%, and true negative (non-rumble accuracy) is 94.8%. This provides us with a highly accurate automated seismic rumble detector, which is yet another reason to push further with using seismic data for elephant conservation.

2.2.3.2 Walking versus running

Another common behaviour exhibited in times of danger is running (Sukumar, 2003), either as an attack or defense mechanism from poaching for instance. Therefore, once we have identified an elephant signal, we additionally attempt to train a classifier to determine whether the elephants are running or not.

This task is particularly difficult for elephants, because their gait transition is not as pronounced and clear as other animals such as e.g. in humans or zebras (Ren and Hutchinson, 2008). Therefore, the main differences we can expect in the signal comes either from the frequency of the footfalls (as can be seen in figure 2.9) or their intensity. However, these two effects can also be confounded by the

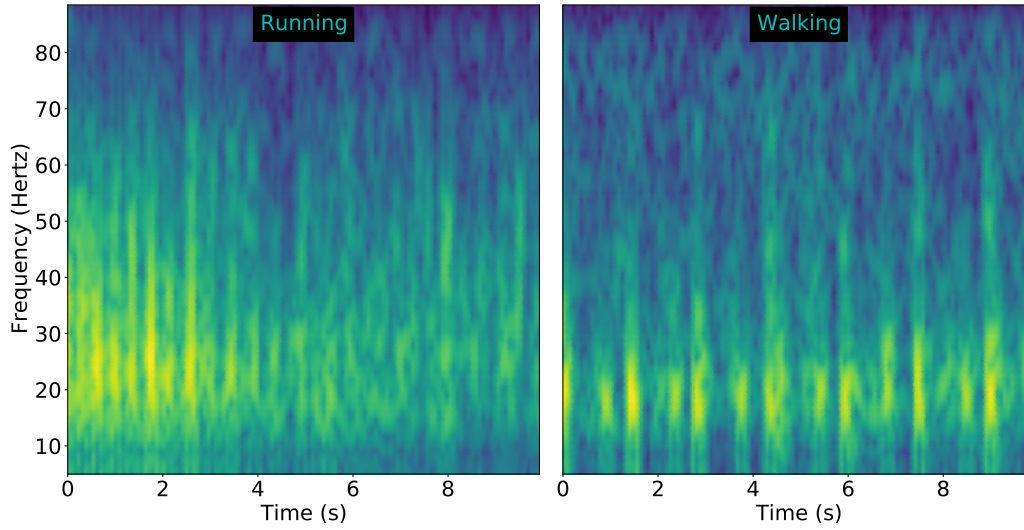


Figure 2.9: Spectrograms of an elephant running (left) versus elephant walking (right). In this instance we can see a clear difference in footstep frequency between the two signals.

number of animals or the distance of the animals, e.g. several elephants walking close to a seismometer could look the same as elephants running further away. Nevertheless, we include all sightings with up to 100 animals in the image, in order to be as realistic as possible in the type of signals we cover, and also increase our sample size.

The behaviour dataset is generated similarly to the species dataset, except that we only select sightings of elephants, and use behavioural notes for class attribution to "running" or "walking". The 60 meters dataset contains 309 samples of the running class, and 18726 of non running. We take particular care to only keep running data for which the signal is very clear, since having erroneous labels with such low sample size drastically degrades performance. To this end, we visually inspect the running data and refine the labels by discarding samples for which it is clear that the signal has not been recorded at the seismometer. While we only refine labels for the samples from the "running" class (the large amount of

samples in the "walking" class makes visual inspection too time consuming), and it is inevitable that the "walking" class will contain some unnoticed instances of running, the majority of the labels should be correct and allow the network to successfully discriminate between the signals.

The best performing model on this task is the 2D-CNN trained from scratch, which on the test set achieves a balanced accuracy of 93.9%, with a perfect true positive rate (running accuracy) of 100%, and a true negative rate (walking accuracy) of 87.8% . Interestingly, this is the only task for which the 2D-CNN trained from scratch outperformed the pretrained Squeezenet, which achieved 75% balanced accuracy.

Once again, these results show great promise for the use of seismic data to not only detect elephants, but also understand whether they display behaviour that potentially indicates a perilous situation.

2.3 Discussion

In this work we report a novel approach to monitoring wildlife, including African elephants, which utilises deep learning to classify seismic signals recorded on seismometers. We open-source all the code required to reproduce our results.

Crucially, we open-source SeisSavanna: the seismic wildlife dataset used in this work, with the hope that it will foster collaboration between the fields of machine learning, geophysics, and biology. The species file in SeisSavanna contains 70k sightings from 11 different species, for distances up to 150 meters. We also provide two elephant behaviour files, one containing sightings of elephants running and walking, with respectively 1.5k and 61k samples up to 150 meters; and a rumbles file containing 1.5k rumbles and 25k locomotion signals. While this

dataset is intended to benefit the investigation of new methods to use seismic data for conservation, it is worth noting that it is also very useful for computer vision tasks, since for each seismogram in the species and running files, we provide the corresponding picture that generated the sighting.

By applying deep learning methods to these datasets, we are able to distinguish elephants from other species with 80–90% balanced accuracy for distances of up to 100 meters, recognise elephant vocalisations in the form of rumbles with 96% balanced accuracy, and distinguish walking from running in elephants with 94% balanced accuracy for distances up to 60 meters. Our best performing approaches are CNNs applied on spectrograms, in particular finetuning a Squeezenet pre-trained on ImageNet, or training an all convolutional 2D-CNN from scratch.

The aforementioned results are very encouraging, as deep learning methods performed very well in the tasks we set them, with good accuracy despite the challenging natural context in which data was collected. In particular, the variable field terrain (e.g., wet and muddy conditions, rocky ridges, hard stone and sand, gorges) and the noisy conditions of the sites (e.g., many different animal species, large variation in number of animals, cars, etc) are usually seen as compromising factors for generalisation of such an approach, but our results indicate that these can be handled well. Our results also highlight some of the limitations and key variables of the dataset which we discuss in more detail below, namely: 1. Uncertainty in the labels due to partial camera trap coverage; 2. Decrease of signal to noise ratios (SNR) with distance, which causes corruption in the labels; 3. Effect of local environmental conditions and their impact on the ability of the network to generalise to new locations.

We now address each of these points in more detail, before outlining how these limitations can be mitigated.

To begin with, one crucial consideration about the data, which is relevant to all the tasks we are tackling, is the relatively high level of uncertainty present in the signal. Indeed, the seismometers record data with a 360° azimuthal range, whereas the animal species labels are provided by camera traps, which have a range of about 60° and about 30 meters. This means that there is a sizeable blind spot in our labelling procedure, and to give an extreme example, there could be an elephant walking a few meters from a seismometer, while the label provided by the camera is that of a zebra 50 meters away. Naturally, the fieldwork was designed to try and minimize these scenarios, notably by placing camera traps covering all directions in areas most propitious to animal traffic, as recommended by our local field guide. Therefore, the corruption of labels should not be too prevalent, but must nevertheless be kept in mind. Considering this uncertainty, the accuracies achieved in this study are all the more encouraging.

Another important limitation of the data to bear in mind, as highlighted by the decrease in accuracy in classifying elephants versus non elephants for the 150 meters dataset, is the increasing corruption of the labels in the dataset and decreasing signal to noise ratios (SNR). Indeed, as we include sightings of animals at increasingly large distances, seismic signals are less and less likely to propagate all the way to a seismometer. Therefore, if we have a label saying "elephant at 500 meters", the extracted chunk will most likely contain only noise. This causes problems both during training (showing very similar noise signals both for the "elephant" and "non elephant" classes) and during testing (labels are corrupted and therefore metrics are biased).

Finally, attempting to classify elephants from non elephants and generalise to new stations highlights the importance of the local environment of the station, and the need for more diverse datasets to improve generalisation capabilities to new environments. Indeed, it is noteworthy that the balanced accuracy is higher when

the test set station is similar to the stations in the training set, and lower when the test station is remote or with a different environment, which is physically coherent, because factors such as terrain and geology have important impacts on signal propagation (Mortimer et al., 2018). It is also in line with general results in deep learning, whereby training and test data have to be generated from the same distribution for the network to produce good results, and is the reason why the machine learning community strives to produce datasets containing enormous amounts of labeled data (e.g. ImageNet (Deng et al., 2009) for image classification with 14,197,122 pictures in 1000 classes, or MS COCO (Lin et al., 2014) for object detection with 328k pictures annotated with bounding boxes, segmentation masks, natural language captions, etc).

How might we address or mitigate the limitations outlined above? An obvious solution is to acquire more data, from more diverse environments. This will allow us to improve accuracy, generalise better to new stations, but also test refined classification tasks, such as multispecies classification. A straightforward way to get more labeled data is to label the remaining camera trap pictures. Another possibility worth exploring is to deploy dense arrays of devices such as geophones or accelerometers. These devices are inexpensive and easy to deploy, and can therefore be used to sample many different locations, which is a boon both for the monitoring range since we can cover more area, but also for the performance of the method, as we will sample more varied environmental conditions and generalise better. Dense arrays of instruments can also be beneficially exploited, e.g. to enhance signal to noise ratios with stacking, but also with more advanced array methods (Rost and Thomas, 2002). However, more data is not the only solution to overcome the outlined limitations of our approach.

A very successful way to improve generalisation ability in machine learning is the use of data augmentation methods, whereby we create fake training examples by

modifying existing data in realistic ways (e.g., rotating or distorting pictures in image classification ([Shorten and Khoshgoftaar, 2019](#))). We attempted to use data augmentation methods designed for seismic signals ([Zhu et al., 2020](#)) (see Methods section for more details), but have not seen a consistent improvement in accuracy. This likely means that for our task, these methods do not produce realistic or useful datapoints, and in future work we will address the design of specialised data augmentation techniques that will help us generalise to new environments and therefore across stations. For instance, we are planning to investigate synthetic signal based augmentation, by exploiting generative adversarial networks (GAN), which can be used to produce realistic new training examples ([Frid-Adar et al., 2018](#); [Li et al., 2018](#)). In particular, we would like to be able to create data that appears to come from new environments, which is key to across-station generalisation.

To tackle the uncertainty in the labels, future fieldwork should surround each seismometer with several camera traps in order to provide full azimuthal coverage and remove label corruption altogether. To address the issue of label corruption due to increasing distances, one can improve the existing dataset by sifting through all the datapoints and manually refining the labels by removing samples that contain no signal. It is a slow and time consuming process, which unfortunately is hard to automate by simply removing low SNR samples, because doing so also removes many valuable low amplitude signals that have comparable SNRs.

Finally, we outline a few additional avenues for future research which we believe could improve our results. To begin with, it would be straightforward and interesting to attempt to directly regress on the distance and the angle of an animal, in addition to simply classifying the species. To put it another way, we would be learning to localise, rather than simply detect the animal. However, this could be challenging due to the noisy nature of the data, such as fairly large uncertainty in

the location measurement, coupled with strong effects of the local environment. Another interesting direction of investigation which can be pursued with the current dataset is to utilise array methods to increase the signal to noise ratio in the data. Indeed, there are numerous methods derived from seismic exploration and seismology which could be tried out, such as weighted stacking, double beam, N-th root process, etc.

Overall, our results show that information-bearing signals propagate over large distances (over 100 meters), and that seismic data are a very promising avenue for monitoring wildlife and their behaviour with automated techniques. Together with the release of seismic wildlife datasets, this opens up the door for many applications. First and foremost, this approach paves the way for the development of elephant seismic monitoring systems for conservation, with the potential for near real-time capability. We have shown that seismic data offers great benefits for detecting elephants, in terms of sensitive detection up to tens of meters range and accurate classification of both the species and behaviours. The detection of elephant running is of particular interest since this potentially indicates a situation of welfare concern ([Sukumar, 2003](#)). In the future, in combination with the deployment of dense arrays of geophones or accelerometers, to improve signal to noise ratios and sample more varied environments, we envisage autonomous sensor systems that can record, collate and analyse seismic data streams in near real-time. This has broad applications for elephant monitoring, whether for the study of their behaviour and communication in the wild, or as information for rangers to respond to behaviours of concern, such as elephants running. With further research, this approach could have applications beyond elephants to detect, classify and monitor a range of animals within their remote habitats.

2.4 Methods

2.4.1 Fieldwork

We conducted fieldwork in the Mpala research centre, located on the Laikipia Plateau in central Kenya, between February and March 2019. The research centre is geographically diverse, defined by a semi-arid savanna landscape and the catchments of two rivers, supporting a wide range of wildlife. In particular, Mpala is the temporary home of over 6,000 migrating elephants, making it a great location for our experiment.

We deployed 22 6TD Gralp seismometers on a loan from SEIS-UK, and 30 Reconyx Hyperfire 2 camera traps, for a duration of three weeks. The seismometers record ground motion data in three orthogonal components at a 200Hz sampling rate and were buried around 70cm beneath the surface. The camera traps have an infrared motion detector that triggers the recording of 10 pictures at 1s intervals, and were mostly attached to trees and poles at human-eye height. The theoretical range of the sensor is up to 30m in a 60 degrees cone, although this is often lower in practice due to environmental conditions.

All the instruments were deployed in an area centered around a watering hole where many species congregated every day. As can be seen in figure 2.10, most of the instruments were deployed within a few hundred meters of the hole in order to maximize the amount of data; but others were also deployed more remote locations, so as to add diversity in terms of geology and terrain. For example, the terrain around the waterhole is wet, muddy, and with trees and bushes on the outskirts, the NWP05 station is more arid with rocky and sandy terrains, while the remote>NNL62 has a rocky geology as well as a river gorge that is often crossed by animals.

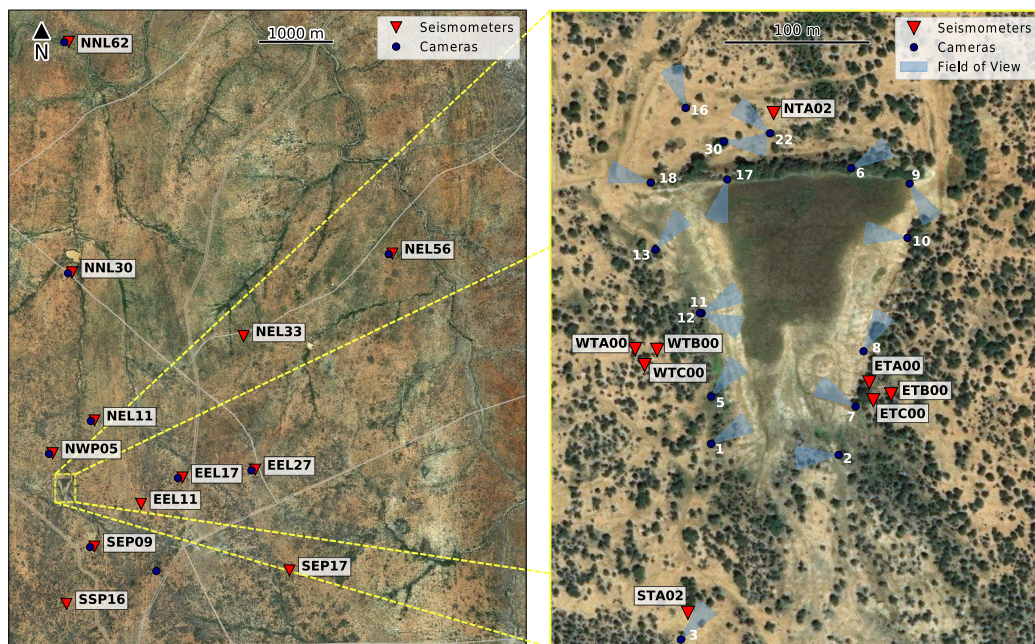


Figure 2.10: Site of fieldwork, with annotations for the positions of camera traps and seismometers. Note that the 'field of view' cone is that of the infrared sensor, and when the sensor is triggered, the camera records a picture with a much larger field of view. The name of seismometers is coded in reference to the watering hole, e.g. WTA00 is due West of the hole at a "distance" of 0 meters, STA02 is due South at distance of 150m, NWP05 is North-West at a distance of 500 meters, etc.

2.4.2 Data processing and dataset creation

We recorded over 250,000 pictures on the camera traps and 8,000 hours of seismograms. One of the contributions of this paper is a unique dataset of seismic recordings with the corresponding label of the species generating the signal. In this section we focus on the steps for the processing and creation of the dataset.

2.4.2.1 Camera trap data

The pictures were labeled, resulting in a total of around 190,000 annotated pictures, of which around 82,000 contained animals, and the rest were caused by misfiring of the infrared sensor, e.g. due to wind. The annotation process consisted of not only labeling the species and number of animals, but also the behaviour exhibited in the picture, as was determined by the sequence of 10 pictures recorded per trigger. What is more, each picture contains an annotation of the position of groups of animals in reference to the camera trap. In order to do this, for each camera trap some reference point in the visible range was measured during field work, and subsequently used to interpolate for positions. The resulting distances are not very precise (we estimate that there is an error radius of about 10m), but still informative and useful.

Once we have an annotated csv file with all the pictures, we can start constructing a dataset. To this end, we define a series of filters to select the sightings in the csv file. These include: maximal distance of an animal from a seismometer, maximum number of animals in the pictures, list of species of interest, list of behaviours of interest, etc. From these filtered sightings we extract the corresponding seismograms, with a length of 10 seconds. More detail on the generation procedure can be found in algorithm 1, in the supplementary material. Each seismogram has a class corresponding to "Elephant" or "Not Elephant". The "Not Elephant" class

contains all other types of signal detected by the camera trap: zebras, camels, giraffes, humans, cars, etc.

2.4.2.2 Seismic data

The seismic data were preprocessed by removing the instrument response from each seismometer, and we applied a bandpass filter on the extracted seismograms between 5 and 95Hz. No other preprocessing is applied, so that additional processing steps can be applied at runtime as a hyperparameter.

2.4.2.3 Spectrograms

We also generate spectrograms from the extracted seismic chunks, in order to experiment with 2D CNNs. The spectrograms are obtained by applying a short time Fourier transform (STFT) to each of the extracted seismograms. The signal is split into overlapping time windows, for each of which a fast Fourier transform (FFT) is computed. We choose a time window of 40 samples (0.2 seconds) in order to capture the spectral information in a step. The overlap of time windows is 39 samples (0.195 seconds).

Spectrograms represent the evolution of the frequency spectrum of a signal through time, and have proven successful in environmental sound recognition applications (Salamon and Bello, 2017; Palanisamy et al., 2020). Following (Salamon and Bello, 2017), we also convert the spectrograms to the mel scale with 128 bins and convert to a logarithmic scale.

2.4.3 Modelling

The classification problem is framed as supervised learning, for which we introduce a few key concepts. Supervised learning aims to perform function approx-

imation given a set of N training examples of inputs $X_i : X_i \in \mathbf{X}$ and desired outputs $y_i : y_i \in \mathbf{Y}$. In our case, each X_i may be the 2048×3 seismogram input $\in \mathbb{R}^{2048 \times 3}$, or the $128 \times 128 \times 3$ spectrogram input $\in \mathbb{R}^{128 \times 128 \times 3}$; and each y_i is a class label corresponding to the species. Each of our models is a function parametrized by θ , or $f(X_i, \theta) : \mathbf{X} \rightarrow \mathbf{Y}$. We fit these functions by optimizing parameters to minimize the empirical risk for a given loss function (which measures how close $f(X_i, \theta)$ is to y_i), $\theta^* = \operatorname{argmin}_{\theta} \frac{1}{N} L(f(X_i, \theta), y_i)$ or, in other words, θ^* is the set of parameters that minimizes the average loss occurred on the validation set. How this is optimised depends on the form of X_i , f , and L ; but generally a standard optimisation technique such as gradient descent is used. For our task of binary classification, we use the binary cross entropy loss.

We investigate multiple models on this classification task, such as logistic regression, gradient boosted trees, and multi-layer perceptrons (see [Bishop \(2006\)](#) for a thorough introduction to these models), on either seismograms or spectrograms. Our best performing models are different architectures of convolutional neural networks (CNN), a complete introduction to which can be found in [Goodfellow et al. \(2016\)](#). A CNN represents the output of $f(X_i, \theta)$ as a transformation of the input involving multiple layers of convolution or linear filtering, interleaved with non-linearities, resulting in a complex non-linear function built of smaller simple components. The parameter vector θ is then all of the parameters of the convolutions. While the resulting empirical risk minimization problem of minimizing the loss with respect to θ is highly non-convex, the local minima of the networks tend to perform well in practice across most problem domains.

In this study, we experiment with different types of CNNs: 1D CNNs trained from scratch on the seismograms, 2D CNNs trained from scratch on spectrograms, and a Squeezenet pretrained on ImageNet ([Deng et al., 2009](#)) and then finetuned on spectrograms. The latter approach is very successful in environmental sound

classification ([Palanisamy et al., 2020](#)), and is therefore a natural candidate to apply to our dataset.

2.4.4 Metrics

We have a relatively imbalanced dataset with more sightings of elephants than non elephants, and class imbalances for the behavioural datasets as well, therefore we have to use appropriate metrics to account for that fact, as usual metrics (e.g. accuracy) are overly optimistic for imbalanced datasets. With TP denoting true positives, TN true negatives, FP false positives, and FN false negative, we report true positive rate $TPR = TP/(TP + FN)$, true negative rate $TNR = TN/(TN + FP)$, and balanced accuracy $BA = (TPR + TNR)/2$. By averaging the true positive and negative rates, BA offers a balanced and interpretable measure of the quality of binary classification, even in the presence of class imbalance.

2.4.5 Data augmentation

Data augmentation is a standard machine learning method to improve the generalisation capabilities of a network. It consists in modifying datapoints from the training set in realistic ways in order to expose the network to a more diverse dataset, which is more likely to be representative of the data in the test set.

A crucial requirement of data augmentation methods is that they must create realistic datapoints. While in traditional image classification of natural images such as cats and dogs, common augmentation techniques include image rotations and distortions ([Shorten and Khoshgoftaar, 2019](#)), it would be unsuitable for seismic data, since e.g. rotating a spectrogram would not result in a realistic new datapoint.

There have been several methods proposed for earthquake seismic data such as channel dropout, event stacking, false positive noise, etc (Zhu et al., 2020). We select two methods that appear most suitable for our task, which are event stacking and addition of noise. The former corresponds to adding together different seismograms from the same class. The intuition is that if you add several different signals of elephants, the result should look like several elephants walking at the same time. Noise addition simply consists in adding seismic noise to the signal, to artificially create more diverse signal-to-noise ratios in the data. To add seismic noise, we randomly extracted seismic chunks from the data, which we visually inspected to confirm absence of prominent signals, and subsequently randomly added them to the samples during model training.

2.5 Supplementary Material

2.5.1 Models details

In table 2.1 we report results on the 60 meters dataset split in time, for the elephant/non elephant classification task, for several more models that we have explored in this work. Below are details of their architectures and hyperparameters. Let $C2d(k,n,s,p)$ denote a 2D convolutional layer with n filters of size $k \times k$, applied with stride s and padding p ; $C1d(k,n,s,p)$ denote a 1D convolutional layer with n filters of size k , applied with stride s and padding p ; BN denote batchnorm; R denote a LeakyReLU, and $FC(i,o)$ denote a fully connected layer with input dimension i and output dimension o . We select the model weights that achieved best the performance on the validation set.

1. Logistic regression: linear layer in Pytorch, taking the sum of flattened spectrograms as input, i.e. with architecture $FC(128*128, 2)$. Trained for 10 epochs with a batch size of 32, Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate

of $10e^{-5}$

2. 1D-MLP: Multi layer perceptron trained on the vertical component of seismograms. Architecture: FC(2048, 512), Dropout(p=0.5), R, FC(512, 2). Trained 10 epochs with a batch size of 32, Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of $10e^{-4}$.

3. 1D-CNN: All convolutional CNN trained on seismograms. Architecture: C1d(11,32,1,1), BN, R, C1d(4,64,2,1), C1d(11,64,1,1), BN, R, C1d(4,128,2,1), C1d(11,128,1,1), BN, R, C1d(4,256,2,1), C1d(3,256,1,1), BN, R, C1d(4,256,2,1), C1d(3,256,1,1), BN, R, C1d(4,256,2,1), C1d(1,256,1,1), BN, R, C1d(1,2,2,1), BN, R. The result is then averaged over the spatial dimension to produce class predictions. Trained for 100 epochs with a batch size of 256, Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of $10e^{-4}$.

4. 2D-CNN: All convolutional CNN trained on spectrograms. Architecture: C2d(3,32,1,1), BN, R, C2d(4,64,2,1), C1d(3,64,1,1), BN, R, C2d(4,128,2,1), C2d(3,128,1,1), BN, R, C2d(4,256,2,1), C1d(3,256,1,1), BN, R, C2d(4,256,2,1), C2d(3,256,1,1), BN, R, C2d(4,256,2,1), C1d(1,256,1,1), BN, R, C2d(1,2,2,1), BN, R. The result is then averaged over the spatial dimensions to produce class predictions. Trained for 10 epochs with a batch size of 32, Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of $10e^{-6}$.

Metric \ Max distance	TPR	TNR	BA
2D-LR	92.9	38.1	65.5
1D-MLP	82.4	25.3	53.9
1D-CNN	78.6	54.3	66.5
2D-CNN	89.0	88.8	88.9
Squeezenet	92.3	84.6	88.9

Table 2.1: True positive rate (elephant accuracy), true negative rate (non-elephant accuracy), and balanced accuracy $((\text{TPR} + \text{TNR})/2)$ on test set for the 60 meters maximal distance dataset, for a collection of models we experimented with. We only report models we successfully trained, while for others we had issues. The models are: a logistic regression trained on flattened spectrograms (2D-LR), a multi layer perceptron trained on seismograms (1D-MLP), an all-convolutional CNN trained on seismograms (1D-CNN), an all-convolutional CNN trained on spectrograms (2D-CNN), and a Squeezenet pretrained on ImageNet and finetuned on spectrograms.

2.5.2 Data generation algorithm

In algorithm 1, we provide more details on how the chunks are generated from the raw csv file, in order to satisfy filtering options, and reduce the chance of label corruption.

Algorithm 1 Based on a spreadsheet of sightings, i.e. annotations from camera trap pictures containing species, distance to camera, number of animals, and behavioural notes; we filter out sightings based on several criteria such as number of animals and distance to seismometer, to generate a seismic wildlife dataset.

$N_{xl} \in \mathbb{N}$: Total number of sightings in spreadsheet.

$N_{sm} \in \mathbb{N}$: Total number of seismometers

$N_{ma} \in \mathbb{N}$: Maximum number of animals in a sighting.

$N_{md} \in \mathbb{R}^+$: Maximum distance of animal from a seismometer.

L_{sp} : List of species of interest.

L_{bh} : List of behaviours of interest (e.g. walking, running, ...).

$S \in \mathbb{B}$: Whether to keep only sightings with a single species.

$\rho : \{\rho_i\}_{i=1}^{N_{xl}}$: Annotated sightings.

$\alpha : \{\alpha_i \in \mathbb{N}\}_{i=1}^{N_{xl}}$: Number of animals in sighting.

$\beta : \{\beta_i \in \mathbb{R}^+\}_{i=1}^{N_{xl}}$: GPS coordinates of animals in sighting.

$\gamma : \{\gamma_i \in \mathbb{N}\}_{i=1}^{N_{xl}}$: Number of species in sighting.

$\delta : \{\delta_i\}_{i=1}^{N_{xl}}$: Species in the sighting.

$\epsilon : \{\epsilon_i\}_{i=1}^{N_{xl}}$: Behaviours of animals in sighting.

$\tau : \{\tau_i\}_{i=1}^{N_{xl}}$: Time of sighting

$\sigma : \{\sigma_i\}_{i=1}^{N_{sm}}$: Seismometers

$\zeta : \{\zeta_i\}_{i=1}^{N_{sm}}$: GPS coordinates of seismometers

$\eta : \{\eta_{ij}\}_{i,j=1,1}^{N_{xl}, N_{sm}}$: Distance of sighting i from seismometer j.

```

1: function FILTERSIGHTINGS( $N_{ma}, N_{md}, L_{sp}, L_{bh}, S$ )
2:   Set L to empty list.
3:   for  $i = 1, \dots, N_{xl}$  do
4:     if  $\alpha_i \leq N_{ma} \wedge ((S \wedge (\gamma_i == 1)) \vee !S) \wedge (\delta_i \in L_{sp}) \wedge (\epsilon_i \in L_{bh})$  then
5:       Set I to empty list.
6:       Set  $N_k = 1$   $\triangleright$  Nearby sightings can interfere with correct labelling
7:       while  $(\tau_{i+N_k} - \tau_i) \leq 10$  seconds do
8:         Append  $\rho_{i+N_k}$  to I.
9:          $N_k + = 1$ 
10:      for  $j = 1, \dots, N_{sm}$  do
11:         $\eta_{ij} = \text{gps2m}(\beta_i, \delta_j)$ 
12:        if  $\eta_{ij} \leq N_{md}$  then
13:          for  $k = 1, \dots, N_k$  do
14:             $\eta_{(i+k)j} = \text{gps2m}(\beta_{i+k}, \delta_j)$ 
15:            Set add=True
16:            if  $\eta_{(i+k)j} \leq \eta_{ij}$  then
17:              Set add=False
18:              Continue
19:          if add then
20:            Append  $[\tau_i, \sigma_j, \delta_i, \eta_{ij}, \gamma_i]$  to L.
21:   return L

```

Acknowledgements

We would like to express our gratitude to all the staff at the Mpala Research Centre for assisting with fieldwork operations and for creating a warm and welcoming environment, in particular Dino Martins and Cosmas Nzomo. We also thank Fritz Vollrath for discussion throughout and help with project logistics. We thank Frank Pope and staff at Save the Elephants for help getting research permits and Kenyan Wildlife Service affiliations, and Paula Koelemeijer for help with sensor deployment.

A. Szenicer thanks the anonymous donor of his PhD grant. We thank National Geographic (NGS-50019R-18), Royal Society (URF R1 191033), John Fell Oxford University Press Research Fund, Royal Commission for the Exhibition of 1851 for funding. This research has been supported by the Centre for Doctoral Training in Autonomous Intelligent Machines and Systems at the University of Oxford, Oxford, UK, and the UK Engineering and Physical Sciences Research Council. A. G. Baydin is supported by EPSRC/MURI grant EP/N019474/1 and by Lawrence Berkeley National Lab.

Author contributions

T. Nissen-Meyer and B. Mortimer conceived the project, undertook project administration and obtained funding for the project.

A. Szenicer, B. Moseley, T. Nissen-Meyer, B. Mortimer, and S. Oduor conducted fieldwork and collected data.

A. Szenicer processed the data and worked on the models.

M. Reinwald labeled the rumbles from the seismic data.

Z. M. Muteti annotated the pictures, A. McDermott-Roberts and A. Szenicer re-

fined labels.

T. Nissen-Meyer and A. G. Baydin supervised A. Szenicer, T. Nissen-Meyer supervised B. Moseley, B. Mortimer supervised A. McDermott-Roberts and M. Reinwald, B. Mortimer and S. Oduor supervised Z. M. Muteti.

A. Szenicer wrote the manuscript, which was reviewed, corrected, and approved by all authors.

Competing interests

The authors declare no competing interests.

Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

The full SeisSavanna dataset and codes to replicate the main results of the paper can be found at: <https://github.com/AlexSzen/SeisSavanna>

3 | A complexity-driven framework for wave-form tomography with discrete adjoints

Published in Geophysics Journal International, <https://doi.org/10.1093/gji/ggaa349>

Alexandre Szenicer¹, Kuangdai Leng¹, Tarje Nissen-Meyer¹

¹ Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, United Kingdom.

Abstract

We develop a new approach for computing Frechet sensitivity kernels in full waveform inversion by using the discrete adjoint approach in addition to the widely used continuous adjoint approach for seismic waveform inversion. This method is particularly well suited for the forward solver AxiSEM3D, a combination of the spectral-element method (SEM) and a Fourier pseudo-spectral method, which allows for a sparse azimuthal wavefield parametrization adaptive to wavefield complexity, leading to lower computational costs and better frequency scaling than conventional 3D solvers. We implement the continuous adjoint method to serve as a benchmark, additionally allowing for simulating off-axis sources in axisymmetric or 3D models. The kernels generated by both methods are compared to each other, and benchmarked against theoretical predictions based on linearized Born theory, providing an excellent fit to this independent reference solution. Our verification benchmarks show that the discrete adjoint method can produce exact kernels, largely identical to continuous kernels. While using the continuous adjoint method we lose the computational advantage and fall back on a full-3D frequency scaling, using the discrete adjoint retains the speedup offered by AxiSEM3D. We also discuss the creation of a data-coverage based mesh to run the simulations on during the inversion process, which would allow to exploit the flexibility of the Fourier parametrization and thus the speedup offered by our method.

Keywords: Inverse theory, Computational seismology, Seismic tomography, Theoretical seismology.

3.1 Introduction

In the beginning of the twentieth century, observations of seismic signals from teleseismic and regional earthquakes have provided a first glimpse into the structure of our planet (Oldham, 1906), with fundamental discoveries such as the Moho (Mohorovicic, 1909), and the outer and inner cores (Jeffreys, 1926; Lehmann, 1936). These early models are spherically symmetric with only radial variations in the structure of the planet, and yet account for the bulk of the observed seismic phases. Explaining the remaining discrepancies in seismic data by introducing 3D heterogeneities is the purpose of global seismic tomography. It might at first seem inconsequential to try and model the details of the internal structure of the planet in order to explain some relatively minor discrepancies in arrival times of seismic phases, and yet these 3D structures are a cornerstone input for geological and geophysical studies that aim to explain the creation, history, and dynamics of our planet, such as mantle convection (Van der Hilst et al., 1997; Ritsema and Van Heijst, 2000), uprising mantle plumes (Bijwaard and Spakman, 1999; Montelli et al., 2004a; Zhao, 2004; French and Romanowicz, 2015), continental evolution (Ritsema et al., 1999), reconstruction of tectonic plate configurations (Van der Voo et al., 1999; Sigloch and Mihalynuk, 2013) and thermal/compositional components of structural heterogeneities (Schuberth et al., 2009b,a).

The development of global seismic tomography started in the 1970s with methods based on arrival times of different seismic phases modelled with ray theory (Dziewonski et al., 1977), and opened doors for a plethora of global P and S wave velocity models of the Earth (Van der Hilst et al., 1997; Grand et al., 1997; Kustowski et al., 2008; Ritsema et al., 2011; Schaeffer and Lebedev, 2013; Auer et al., 2014; Hosseini et al., 2020). The theoretical framework used for these models assumes that the traveltimes of seismic waves only depend on the structure seen by

an infinitely thin geometric ray path connecting the source and receiver. There are considerable benefits to using these methods, such as availability of large datasets of hand-picked arrival times and earthquake locations from organisations such as the International Seismological Centre ([Willemann and Storchak, 2001](#)), as well as computational efficiency. However, the main limitation is that they are only valid when the wavelength of waves used for the measurements is very small compared to the size of structural heterogeneities, and for those wave types which can be represented by ray theory, for instance neglecting diffracted ([Hosseini and Sigloch, 2015](#)) and triplicated waves ([Stähler et al., 2012](#)).

Global seismic tomography transitioned to more physically accurate, yet computationally demanding methods such as finite frequency theory ([Dahlen et al., 2000](#); [Sigloch and Nolet, 2006](#); [Nissen-Meyer et al., 2007](#)), leading to tomography efforts whereby kernels are calculated either by ray theory ([Montelli et al., 2004b](#); [Hosseini et al., 2020](#)), or by wavefields in spherically symmetric structures ([Colombi et al., 2014](#)). Finally, full waveform inversion (FWI) based on adjoint methods ([Fichtner et al., 2009](#); [Bozdağ et al., 2016](#)) provides a framework for iteratively improving 3D models with gradient methods, even for mildly non-linear regimes. All these methods aim to resolve several limitations from ray-theoretical approaches by incorporating information from the whole frequency-dependent waveform. This provides advantages such as quantifying the quality of observations, for instance extracting new observables from the seismograms such as their frequency-dependent amplitudes ([Sigloch et al., 2008](#)), and correctly modelling multiscale effects from structural heterogeneities that are unaccounted for in ray-based methods, such as diffraction and wavefront healing. Finite frequency tomography assumes that the perturbation of the measurement is a linear function of a perturbation in the model parameters, and these two perturbations are connected by a so-called sensitivity kernel. Though such linearization brings

many benefits such as a drastically lower computational cost and the ability to include large waveform datasets, it inevitably limits the capability of resolving structures with a higher perturbation strength or a finer scale (Tromp et al., 2008). Similarly to finite-frequency tomography, FWI compares synthetic seismograms to observed ones through a misfit function, and the objective is to minimize that function. One of the main differences from finite-frequency tomography is that in FWI the objective function is minimized with a gradient-based iterative procedure. In this way, the non-linear relation between observables and model parameters is at least partially accounted for as long as the nonlinearity does not cause cycle skipping leading to local minima in the misfit function. FWI is based on full 3D simulations, and allows for 3D starting models to be then improved iteratively. In certain settings such as exploration geophysics or basin scales, this may be inevitable (Virieux and Operto, 2009; Tape et al., 2009).

In the mathematical literature, FWI is called a PDE-constrained optimization problem, where we aim to minimize a function χ (the misfit), subject to constraints imposed by a partial differential equation $\mathcal{L}$ (the wave equation in our case) and associated initial and boundary conditions. To solve this minimization problem, a standard way is to compute $d_{\mathbf{m}}\chi$, the gradient of χ in the model space $\mathbf{m}$. To do this, one possibility is to evaluate $\mathcal{L}$ for each model parameter and compute the gradient by finite differences, which would require solving $\mathcal{L}$ for $O(n_m)$ times, where n_m is the dimension of the parameter vector $\mathbf{m}$. In global scale tomography n_m can be millions. A more elegant (but more importantly, computationally efficient) way is using the continuous adjoint method, through which the gradient $d_{\mathbf{m}}\chi$ (called Frechet sensitivity kernels in seismic tomography) can be obtained by just two solutions of $\mathcal{L}$. Continuous adjoint methods have been used for quite some time in optimal control theory (Lions, 1971) and in computational fluid dynamics (CFD) (Pironneau, 1974), but they were first introduced in seismology

by (Tarantola, 1984) for seismic reflection data. While FWI is now well developed (Bozdağ et al., 2016), we still lack the computational power to run it at short simulation periods on a global scale (Nissen-Meyer et al., 2014), which would be desirable to obtain a finer picture of the interior of the planet. As a crude example of the challenging computational cost of using 3D solvers, a global scale FWI tomography at 7s ($\approx 10^4$ CPU hours to produce an hour of seismograms (Leng et al., 2016)) with 3000 earthquakes and 30 iterations of gradient descent would have a computational cost on the order of one billion CPU hours. While this could be in reach with the fastest supercomputers on Earth and moving towards exascale (Bozdağ et al., 2016), keeping a supercomputer with 10^4 cores busy for more than ten years is still a formidable task and any means of significantly reducing this cost are welcome, especially if such resources are out of reach.

In summary, global seismic tomography can, for the purpose of the argument in this paper, be crudely divided into methods based on (1) ray theory but using large amounts of traveltimes data for great data coverage, (2) finite frequency theory acknowledging the finite sensitivity of a wave due to its frequency-dependent character leading to physically more accurate representation of the inverse problem and a wider class of data measurements, and (3) full-waveform inversion allowing to iteratively update 3D models, thereby accounting for mildly nonlinear relations between seismic measurements and underlying structure. Each method has advantages and disadvantages, but the continually increasing computational efficiency in hardware adhering to Moore’s Law increases the interest in computationally more intensive methods.

In this paper, we propose a framework for the computation of 3D Frechet kernels in any background model based on the discrete adjoint approach, leading to a significant speedup compared to conventional waveform inversion. The discrete adjoint approach has been developed and used extensively in computational fluid

dynamics for complex shape optimization problems (Elliott and Peraire, 1997) because continuous methods had poor convergence rates (Giles et al., 2003), but until now it has not become sufficiently relevant in seismology compared to the continuous adjoint approach. This however changed when the azimuthal smoothness of a global wavefield was discovered by (Leng et al., 2016), which naturally gave rise to a spectral-element pseudo-spectral method AxiSEM3D (Leng et al., 2016, 2019) adapted to wavefield complexity: As the wavefield is sparsely represented on a Fourier basis, it leads to a drastic speedup in computational cost. Building on Leng et al.’s idea on azimuthal smoothness, this has also been identified by van Driel et al. (2020), who showed that the discrete adjoint approach can obtain exact 2D Frechet kernels at a much lower computational cost. They achieved such adaptability by an alternative implementation of exploiting azimuthal smoothness by azimuthally adaptive meshing instead of Fourier series, thereby retaining the conventional 3D SEM solver strategy at the expense of having to mesh in 3D. Thrastarson et al. (2020) present an FWI workflow compatible with the discrete adjoint method for 3D global applications, but thus far using 50s discrete adjoint kernels in spherically symmetric models.

Our new approach computes 3D kernels in 3D Earth models and at high frequencies for global scale applications using AxiSEM3D, in which the 3D globe is discretized by a 2D SEM mesh coupled with a Fourier expansion over the azimuthal direction. The azimuthal adaptivity is achieved in AxiSEM3D by varying the number of terms in the Fourier expansion. This discretization is irregular and for the full inversion process can be adapted to the resolvable complexity of the structures based on data coverage, which is what makes the method computationally cheaper. We thus require the resolution of the mesh (and thus model parameters) to be fixed prior to the inversion. Such a priori inversion meshes with spatially varying element sizes based on data coverage are standardly used to reflect that

not all regions have the same resolution (Nolet and Montelli, 2005; Auer et al., 2014; Zaroli et al., 2015; Hosseini et al., 2020)

3.2 Theory

We set ourselves the task of imaging the interior of the Earth, based on a dataset of seismograms recorded after earthquakes on receivers across the globe, $\mathbf{d}(\mathbf{x}_r, t)$, where $\mathbf{x}_r$ are the receiver positions. In full waveform inversion (FWI), we simulate an earthquake by solving the elastodynamic momentum equation in a given Earth model parametrised by a vector $\mathbf{m}$ containing various physical properties (P and S wave velocities, density, etc), and record the resulting synthetic seismograms $\mathbf{u}(\mathbf{x}_r, \mathbf{m}, t)$. We then define a misfit $\chi(\mathbf{m})$ that compares synthetic to observed seismograms ($\mathbf{u}$ and $\mathbf{d}$ respectively), e.g. using the Euclidian norm $\chi(\mathbf{m}) = \int_T dt \|\mathbf{u}(\mathbf{m}, t) - \mathbf{d}(t)\|_2^2$ for waveform tomography, but there is a plethora of other norms one could use. We want the synthetic seismograms to match the data as closely as possible, and hence we want to minimize our misfit by updating the model parameters $\mathbf{m}$ by using gradients of the misfit obtained through the (continuous or discrete) adjoint method. Comprehensive treatments of the adjoint approach in seismology can be found in (Tromp et al., 2005; Fichtner, 2010). The computational effort in FWI is largely determined by the efficiency of its underlying wave solver which computes forward and adjoint wavefields.

The crux of adjoint tomography is thus to have a powerful PDE-solver, which one can leverage to efficiently compute gradients of the misfit function. This is particularly true in our case, since we exploit the unique combination of spectral-element/pseudo spectral methods in AxiSEM3D to compute gradients differently than in other seismic tomography software, and thus cut computational cost. Therefore, to better understand our implementation of the adjoint method, we will first

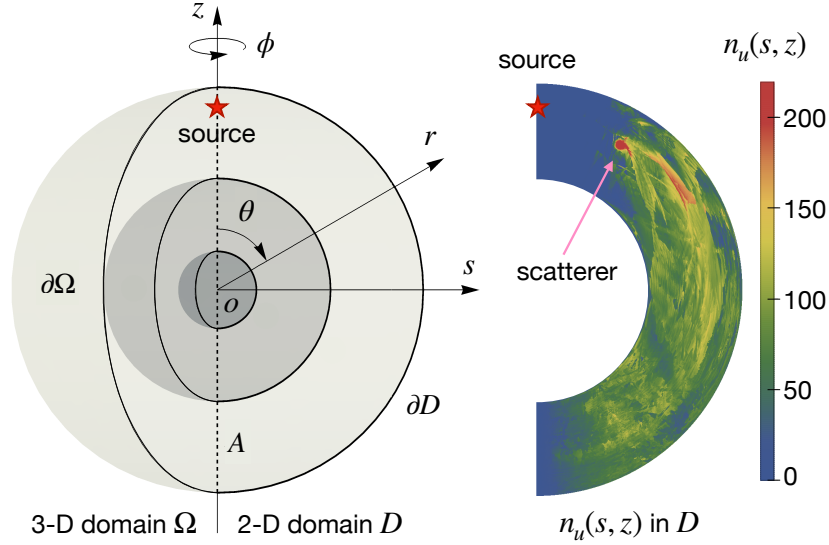


Figure 3.1: Figure from (Leng et al., 2020) illustrating the coordinate system in AxiSEM3D, and the Fourier expansion. The mesh is source-centered so that the source lies on the axis, the slice plane contains an SEM mesh, and the azimuthal direction is expanded in Fourier series. In the source-centred coordinate system (s, ϕ, z) , we reduce the 3-D computational domain Ω to a 2-D meridian domain D through a Fourier series characterisation of the ϕ -dimension. At different locations in D , the Fourier series may contain different number of terms, i.e., $n_u = n_u(s, z)$, determined by the azimuthal complexity of the wavefield. In this example, we show $n_u(s, z)$ for a PREM model including a spherical heterogeneity (with a radius of 160 km and a 50% velocity reduction, located at 30° in distance and 1200 km in depth), obtained by wavefield scanning at a 5 s period and with a 1-hr record length.

give an overview of the functionality of AxiSEM3D.

3.2.1 Forward problem: AxiSEM3D

We rely on the computational speedup offered by AxiSEM3D for solving the wave equation. This speedup is achieved by a combination of spectral element/pseudo spectral methods. The wave equation is solved on a D-shaped spectral elements (SEM) mesh, and the azimuthal direction is solved in Fourier space, see figure 3.1.

Consider the wave equation written in condensed form

$$L(\mathbf{u}, \mathbf{m}) = \mathbf{f} \quad (3.1)$$

where $L()$ is the wave operator, $\mathbf{u}$ is the displacement, $\mathbf{m}$ the model parameters and $\mathbf{f}$ the source term.

AxiSEM3D takes advantage of the observation that in the azimuthal direction, the wavefields propagating through the Earth are relatively smooth, and that complexity is localized. Therefore, the quantities $\mathbf{u}$, $\mathbf{m}$, $\mathbf{f}$ are expanded in Fourier series in the azimuthal direction. Furthermore, the earthquake source is positioned on the axis of the mesh, rendering the expansion of the force term analytical.

$$\mathbf{u}(s, z, \phi) = \sum_{|\alpha| \leq n_u} \mathbf{u}^\alpha(s, z) e^{i\alpha\phi} \quad (3.2)$$

$$\mathbf{m}(s, z, \phi) = \sum_{|\beta| \leq n_m} \mathbf{m}^\beta(s, z) e^{i\beta\phi} \quad (3.3)$$

$$\mathbf{f}(s, z, \phi) = \sum_{|\gamma| \leq n_f} \mathbf{f}^\gamma(s, z) e^{i\gamma\phi} \quad (3.4)$$

One can think of this Fourier expansion as corresponding to the number of grid points in the azimuthal direction at a given point (s, z) . If at a point (s, z) we have n_u Fourier coefficients, it is equivalent to $2 \times n_u + 1$ physical grid points in the azimuthal direction. The power of this approach is that usually n_u is much smaller than the number of azimuthal grid points in a standard 3D SEM mesh, and that one can locally choose n_u depending on the complexity of the model. What is more, AxiSEM3D contains a feature to derive a Fourier expansion that is suf-

ficient to accurately model the wavefield in a given Earth model, called wavefield learning (Leng et al., 2019): Starting from a high Fourier expansion, the solver runs a simulation and progressively reduces the Fourier expansion while conserving the accuracy of the wavefield. The resulting Fourier field is called the wisdom Fourier expansion $n_u^{\text{wisdom}}(s, z)$. These aspects deliver a tremendous speedup when compared to classic 3D solvers (Leng et al., 2016).

At a given location, n_u can be bounded from above such that the wavelength resolution in the ϕ -direction matches that in D , namely (Leng et al., 2020)

$$n_u^{\text{sup}}(s, z) = \frac{2\pi s}{\lambda(s, z)} \times N_D \times \frac{1}{2} = \frac{\pi N_D s}{\lambda(s, z)} \quad (3.5)$$

where $\lambda(s, z)$ denotes the spatial resolution taken as the local wavelength (S-wavelength in solid or P-wavelength in fluid), and thus $2\pi s/\lambda$ the number of wavelengths along the circle generated by revolving the point (s, z) , and N_D the number of grid points in D to resolve one wavelength in the s - or z -direction, a constant normally ranging from 6 to 8 in SEM, and the rearward 1/2 stems from the equivalence of one complex term in the Fourier series and two grid points along the circle. Given the Earth model or $\lambda(s, z)$, n_u^{sup} scales with the distance s to the axis.

As mentioned above, a requirement of this approach is that a source necessarily has to be placed on the axis of the D-shaped mesh. Indeed, because in the azimuthal direction the source term is a Dirac delta, it would theoretically require an infinite Fourier expansion to model accurately. However, in the next section we show how in practice we can accurately simulate off-axis sources by setting the Fourier expansion close to n_u^{sup} , and how we can exploit that to compute exact sensitivity kernels using the discrete adjoint approach. In summary, AxiSEM3D efficiently delivers accurate 3D wavefields for 3D Earth models with internal and

external boundary undulations, including arbitrarily anisotropic ([van Driel and Nissen-Meyer, 2014](#); [Tesoniero et al., 2020](#)) viscoelastic rheologies.

Now that we understand in more detail the framework of AxiSEM3D, as well as its strengths and limitations, we can look at the adjoint method for computing sensitivity kernels, and how we can leverage AxiSEM3D to exploit smoothness and speed up the computations.

3.2.2 Inverse problem

We provide a summary of the theory behind seismic waveform inversion, which consists of determining the 3D model parameters based on measured seismograms. As mentioned in the introduction, we will focus on full waveform inversion (FWI), which exploits the entire waveform (or a time window within it), and allows to obtain Frechet kernels using the adjoint approach. Adjoint methods have been used for a long time in optimal control theory and in fluid dynamics. They were first introduced for FWI by ([Tarantola, 1984](#)) and thorough theoretical derivations can be found in ([Tromp et al., 2005](#)) and ([Fichtner, 2010](#)).

The concept of adjoint methods is very general and applies to any kind of partial differential equation. Moreover, the key idea applies already to a finite-dimensional system of linear equations. In fact, by definition, the "adjoint" of a linear operator is just the generalization of the transpose of a matrix (which is a finite-dimensional linear operator).

Generally speaking, the adjoint method is a numerical method for efficiently computing the gradient of a function or operator in a numerical optimization problem. Let us illustrate the idea behind the adjoint method through a simple example. Consider a function $f(x) : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ for which we would like to determine the gradient $d_p f$ (e.g. to minimize f), and the relationship $g(x, p) = 0$ for a function

$g(x, p) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_x}$. Additionally, suppose f is expensive to evaluate and p is high dimensional, such that computing $d_p f$ by finite differences is prohibitively costly. Let us see how, by using the adjoint method we can obtain the gradient by solving $g(x, p) = 0$ once.

We have $d_p f = d_p f(x(p)) = \partial_x f d_p x$. What is more, $g(x, p) = 0$ everywhere means that $d_p g = 0$. Expanding the total derivative, $\partial_x g \partial_p x + \partial_p g = 0$ or $\partial_p x = -\partial_x g^{-1} \partial_p g$. Reinjecting into the expression for $d_p f$, we get $d_p f = -\partial_x f \partial_x g^{-1} \partial_p g$. We can notice that $-\partial_x f \partial_x g^{-1}$ is a row vector times an $n_x \times n_x$ matrix, which can be understood as the solution to the linear equation $\partial_x g^T \lambda = \partial_x f^T$, and finally $d_p f = \lambda^T \partial_p g$. As mentioned in the previous paragraph, the matrix conjugate transpose (here simply transpose because we are working with real numbers) is called the adjoint, and hence λ is called the adjoint variable, and $d_p f = \lambda^T \partial_p g$ the adjoint equation.

We will detail two ways of looking at the calculation of the gradient, namely the continuous adjoint and the discrete adjoint approaches. Note that the approach used will not affect the outcome of the inverse solution as the kernels (gradients) used to update each iteration are identical. However, and this is the crucial point, the discrete adjoint in our AxiSEM3D context provides a much more efficient time to solution as shown further below, especially at higher frequencies and thus opens doors to new parameter regimes such as using more data or higher frequencies.

Let us define a misfit $\chi(\mathbf{u})$ (this is our function f in our previous example) that compares synthetic to observed seismograms ($\mathbf{u}$ and $\mathbf{d}$ respectively) at a number of receiver locations $\mathbf{x}_r$, e.g. using the Euclidian norm $\chi(\mathbf{u}) = \frac{1}{2} \sum_{r=1}^N \int_T dt ||\mathbf{u}(\mathbf{x}_r, \mathbf{m}, t) - \mathbf{d}(\mathbf{x}_r, t)||^2$, which we want to minimize. The perturbation of the misfit induced by a perturbation in the model parameters is the total derivative:

$$\delta\chi = \frac{d\chi}{d\mathbf{m}} \cdot \delta\mathbf{m} \quad (3.6)$$

As we have seen, adjoint methods allow us to obtain gradients $d\chi/d\mathbf{m}$ by solving a second equation

$$L^\dagger(\mathbf{u}^\dagger, \mathbf{m}) = \mathbf{f}^\dagger \quad (3.7)$$

where $L^\dagger$ is the adjoint wave operator, $\mathbf{u}^\dagger$ the adjoint wavefield and $\mathbf{f}^\dagger = -\partial\chi/\partial\mathbf{u}$ is the adjoint source term (Tromp et al., 2005; Fichtner, 2010). The continuous wave operator is self-adjoint, meaning we can solve the adjoint equations with the same solver used in the forward simulation, simply with different initial conditions.

In conventional derivations of kernels in seismology (Tromp et al., 2005; Fichtner, 2010), running the adjoint simulation delivers the adjoint wavefield $\mathbf{u}^\dagger$. From there, we can compute the directional derivative (called sensitivity kernel) for various model parameters. For instance with respect to P wave velocity v_P , the sensitivity kernels are (Tromp et al., 2005; Fichtner, 2010) :

$$K_{v_P} = 2\rho v_P \int_T \nabla \cdot \mathbf{u}^\dagger(t) \nabla \cdot \mathbf{u}(T-t) dt \quad (3.8)$$

and for a given model parameter $\mathbf{m}$, we have the misfit variation

$$\frac{d\chi}{d\mathbf{m}} \cdot \delta\mathbf{m} = \int_\Omega K_m(\mathbf{r}) \delta\mathbf{m}(\mathbf{r}) d\mathbf{r}. \quad (3.9)$$

This equality provides the essential, locally linearized relation between measurement and model for FWI. The solution to the inverse problem is the final model

$\mathbf{m}$ for the case where its corresponding residual misfit $\delta\chi$ is deemed acceptably small. Note that we deliberately keep $\mathbf{m}$ generic; the theory below and our corresponding methodology allow for computing sensitivity kernel for any anisotropic element of the elasticity tensor, as well as attenuation, irrespective of whether the background model used in the forward problem has the same model complexity.

3.2.3 Continuous adjoint

In the equations presented above, which is the common approach in seismology, we can see that we haven't specified any simulation grid: the derivation of the adjoint system is performed on the continuous equations, and we can then discretize and solve our system.

From the previous section, the adjoint equation as derived in the continuous adjoint framework (Tromp et al., 2005; Fichtner, 2010) is given by eq. (3.7). Let us now discretise this equation in space, to obtain the semi-discrete representation of the continuous adjoint equation

$$\ddot{\mathbf{u}}^\dagger = \mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m})\mathbf{u}^\dagger + \mathbf{M}^{-1}\mathbf{f}^\dagger \quad (3.10)$$

with initial conditions $\mathbf{u}^\dagger(T) = \mathbf{0}$ and $\dot{\mathbf{u}}^\dagger(T) = \mathbf{0}$, and **sans-serif font** always denoting discretised quantities. The dots denote derivative in time, $\mathbf{u}^\dagger$ is the discretised adjoint displacement wavefield, $\mathbf{m}$ are the discretised model parameters, $\mathbf{M}$ is the discretised mass matrix (which is diagonal in AxiSEM3D), $\mathbf{K}$ is the discretised stiffness matrix, and $\mathbf{f}^\dagger$ the discretised adjoint source term (which is a derivative of the misfit function). Note that in this approach, we have to choose the discretisation such that the adjoint wavefield $\mathbf{u}^\dagger$ is physical. In the case of AxiSEM3D this means a very high and expensive Fourier expansion to accurately express an off-axis source. This approach of optimize-then-discretize is called the continu-

ous adjoint approach. There is one bottleneck to this approach with AxiSEM3D. Indeed, if we want to compute the adjoint wavefield originating from a number of globally distributed receivers simultaneously as is customary in adjoint-based tomography, we need to place sources off-axis. As we have discussed above, in AxiSEM3D's context, this requires modeling a point force source in Fourier space, which has an infinite Fourier expansion. In practice, this means that we need to use a rather large Fourier field, sufficient to accurately model in azimuth the wavelength of the dominant period of the simulation. We investigate this in the next section, by plotting seismograms for increasingly high Fourier fields, and we show that they converge to a physically correct seismogram.

3.2.4 Discrete adjoint

An alternative approach to optimize-then-discretize is the discretize-then-optimize paradigm. In the former, we first optimise the continuous system and then discretise the resulting equations, and in the latter we first discretise the wave equation on the forward grid, and then optimise with respect to the discrete model parameters. In this section, we will focus on showing that the two approaches are equivalent, in that they give rise to the same adjoint equations. However, we'll see that the discrete adjoint approach imposes fewer constraints on the discretisation of the adjoint equation, which is very beneficial in AxiSEM3D.

Starting directly from the discrete wave equation with the forward model discretisation and optimising with respect to those discrete model parameters, we can arrive at eq. (3.10) but without the expensive Fourier expansion for the source term. We demonstrate that we can re-use the same expansion as the forward simulation, which in AxiSEM3D is very advantageous because it exploits the azimuthal smoothness of the wavefield. The adjoint wavefield $\mathbf{u}^\dagger$ will not be physical anymore, but the kernels will still be exact with respect to the model parameters of

the forward simulation.

We derive the adjoint equations in the discrete adjoint framework following the Lagrange multipliers approach from (Liu and Tromp, 2008), except that they start with the continuous wave equation, and we start with the discretised one. The equivalence of the two approaches may seem evident, but for clarity it shall be useful to re-derive it explicitly. As we will see, we cast the computation of the kernels as a constrained optimisation problem to minimise the misfit function. The adjoint wave equation then arises as a convenient means to remove terms that are intractable to compute, thus rendering the calculation of the kernels much simpler, by two solutions of the wave equation.

Let us begin by given a brief introduction to the method of Lagrange multipliers. Suppose we wish to find the minimum of a function $f(x)$, subject to the constraint $g(x) = 0$. We can construct the Lagrangian function $\mathcal{L} = f(x) - \lambda g(x)$, where λ is called a Lagrange multiplier, and we can solve the original optimization problem by finding the stationary points of the Lagrangian. As such, the method of Lagrange multipliers is just a reformulation of a constrained optimisation problem into an unconstrained one, such that we can simply perform derivative tests to find the extrema. In the context of the adjoint method, we will not be solving the optimisation problem analytically, but as we will see in the next paragraph, deriving the Lagrangian will allow us to find computationally efficient expressions for the gradients of the function we wish to optimize.

Consider again the wave operator $L(\mathbf{u}, \mathbf{m}) = \mathbf{f}$. We can discretise this operator in space on the grid of the forward simulation similar to eq (3.10), to obtain the semi-discrete ordinary differential equation

$$\ddot{\mathbf{u}} = \mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m})\mathbf{u} + \mathbf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m}) \quad (3.11)$$

with some initial conditions $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$ and $\mathbf{h}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$. The dots denote derivative in time, $\mathbf{u}$ is the discretised displacement wavefield, $\mathbf{m}$ are the discretised model parameters for the forward problem, $\mathbf{M}$ is the discretised mass matrix, $\mathbf{K}$ is the discretised stiffness matrix, and $\mathbf{f}$ the discretised source term. Note that $\mathbf{K}$ is time-independent since attenuation can be incorporated in the Earth model parameters without time dependence, as shown by (Fichtner and van Driel, 2014). Note also that unlike eq. (3.10), the discretisation is that of the forward problem, so a relatively low Fourier expansion in AxiSEM3D. For convenience of manipulation we will rewrite the semi-discretised wave equation as

$$\ddot{\mathbf{u}} = \mathbf{l}(\mathbf{u}, \mathbf{m}, t) \quad (3.12)$$

where $\mathbf{l}(\mathbf{u}, \mathbf{m}, t) = \mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m})\mathbf{u}(t) + \mathbf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m})$.

As previously stated, the goal of seismic tomography is to minimize a misfit function of some measurement between synthetic and observed seismograms with respect to the model parameters. Such a measurement choice could be the Euclidian norm $\chi(\mathbf{u}) = \frac{1}{2} \sum_{r=1}^N \int_0^T \|\mathbf{u}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)\|^2 dt$ which we write as $\chi(\mathbf{u}) = \int_0^T \Xi(\mathbf{u}, t) dt$, where Ξ is a concatenated, generic notation for any chosen measurement. This minimization is constrained by the semi-discrete wave equation $\ddot{\mathbf{u}} = \mathbf{l}(\mathbf{u}, \mathbf{m}, t)$, and initial conditions $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$ and $\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}) = \mathbf{0}$. Now we will solve this constrained optimisation problem to re-derive eq. (3.10) but starting from the semi-discrete wave equation and thus obeying the forward discretisation, by using Lagrange multipliers. To do so, we start with the Lagrangian to describe the dynamics of this system as

$$\mathcal{L} = \int_0^T [\Xi(\mathbf{u}, t) + \boldsymbol{\lambda}^T(t) \cdot (\ddot{\mathbf{u}} - \mathbf{l}(\mathbf{u}, \mathbf{m}, t))] dt + \boldsymbol{\mu}^T \cdot \mathbf{g}(\mathbf{u}(0), \mathbf{m}) + \boldsymbol{\eta}^T \cdot \mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}), \quad (3.13)$$

where λ , μ and η are the vector Lagrange multipliers. Note that because the constraints $\ddot{\mathbf{u}} = \mathbf{l}(\mathbf{u}, \mathbf{m}, t)$, $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$, $\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}) = \mathbf{0}$ are (by construction) always satisfied, we have $d_{\mathbf{m}}\mathcal{L} = d_{\mathbf{m}}\chi$, and we are also free to set the multipliers λ , μ and η to any value. This is crucial, as it will allow us to set the multipliers to a value that conveniently makes some incomputable terms vanish in the expression of the kernel, and it is the beauty of the adjoint method. We can now derive the gradient $d_{\mathbf{m}}\mathcal{L}$. As stated above, because the constraints are satisfied by construction and $d_{\mathbf{m}}\mathcal{L} = d_{\mathbf{m}}\chi$, we are actually computing the gradient $d_{\mathbf{m}}\chi$. Note again that in this section, all gradients and partial derivatives are discrete operators, with respect to the forward discretisation.

The Lagrangian variation with respect to discrete model parameters becomes

$$\begin{aligned}
d_{\mathbf{m}}\mathcal{L} = & \int_0^T [\partial_{\mathbf{u}}\Xi(\mathbf{u}, t)d_{\mathbf{m}}\mathbf{u}(t) + \partial_{\mathbf{m}}\Xi(\mathbf{u}, t) \\
& + \lambda^T(t) \cdot (d_{\mathbf{m}}\ddot{\mathbf{u}} - \partial_{\mathbf{u}}\mathbf{l}(\mathbf{u}, \mathbf{m}, t)d_{\mathbf{m}}\mathbf{u}(t) - \partial_{\mathbf{m}}\mathbf{l}(\mathbf{u}, \mathbf{m}, t))]dt \\
& + \mu^T \cdot (\partial_{\mathbf{u}(0)}\mathbf{g}(\mathbf{u}(0), \mathbf{m})d_{\mathbf{m}}\mathbf{u}(0) + \partial_{\mathbf{m}}\mathbf{g}(\mathbf{u}(0), \mathbf{m}) \\
& + \eta^T \cdot (\partial_{\dot{\mathbf{u}}(0)}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m})d_{\mathbf{m}}\dot{\mathbf{u}}(0) + \partial_{\mathbf{m}}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m})).
\end{aligned} \tag{3.14}$$

Integrating the term $\int_0^T \lambda^T(t)d_{\mathbf{m}}\ddot{\mathbf{u}}(t)dt$ twice by parts yields

$$\int_0^T \lambda^T(t)d_{\mathbf{m}}\ddot{\mathbf{u}}(t)dt = [\lambda^T(t)d_{\mathbf{m}}\dot{\mathbf{u}}(t)]_0^T - [\dot{\lambda}^T(t)d_{\mathbf{m}}\mathbf{u}(t)]_0^T + \int_0^T \ddot{\lambda}^T(t)d_{\mathbf{m}}\mathbf{u}(t)dt$$

Reinjecting this into eq. (3.14) and grouping terms by wavefield derivatives with

respect to model parameters, we obtain

$$\begin{aligned}
d_{\mathbf{m}}\mathcal{L} = & \int_0^T [(\partial_{\mathbf{u}}\Xi(\mathbf{u}, t) - \boldsymbol{\lambda}^T(t)\partial_{\mathbf{u}}\mathbf{l}(\mathbf{u}, \mathbf{m}, t) + \ddot{\boldsymbol{\lambda}}^T(t))d_{\mathbf{m}}\mathbf{u}(t) + \partial_{\mathbf{m}}\Xi(\mathbf{u}, t) - \boldsymbol{\lambda}^T(t)\partial_{\mathbf{m}}\mathbf{l}(\mathbf{u}, \mathbf{m}, t)]dt \\
& + (\boldsymbol{\mu}^T\partial_{\mathbf{u}(0)}\mathbf{g}(\mathbf{u}(0), \mathbf{m}) + \dot{\boldsymbol{\lambda}}^T(0))d_{\mathbf{m}}\mathbf{u}(0) \\
& + (\boldsymbol{\eta}^T\partial_{\mathbf{u}(0)}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}) - \boldsymbol{\lambda}^T(0))d_{\mathbf{m}}\dot{\mathbf{u}}(0) \\
& - \dot{\boldsymbol{\lambda}}^T(T)d_{\mathbf{m}}\mathbf{u}(T) \\
& + \boldsymbol{\lambda}^T(T)d_{\mathbf{m}}\dot{\mathbf{u}}(T) \\
& + \boldsymbol{\mu}^T\partial_{\mathbf{m}}\mathbf{g}(\mathbf{u}(0), \mathbf{m}) + \boldsymbol{\eta}^T\partial_{\mathbf{m}}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m}).
\end{aligned} \tag{3.15}$$

Now we can freely set the Lagrange multiplier values to what we deem convenient because the constraints are satisfied by construction. What is more, all the terms that are derivatives of the wavefield with respect to model parameters such as $\partial_{\mathbf{m}}\mathbf{u}$ are not straightforward to compute because they would require many runs to do finite-differences, which is intractable with many model parameters. Therefore, we can set $\boldsymbol{\lambda}^T(T) = \mathbf{0}$, and $\dot{\boldsymbol{\lambda}}^T(T) = \mathbf{0}$, and $\boldsymbol{\eta}^T = \boldsymbol{\lambda}^T(0)\partial_{\mathbf{u}(0)}\mathbf{h}(\dot{\mathbf{u}}(0), \mathbf{m})^{-1}$, and $\boldsymbol{\mu}^T = -\dot{\boldsymbol{\lambda}}^T(0)\partial_{\mathbf{u}(0)}\mathbf{g}(\mathbf{u}(0), \mathbf{m})^{-1}$, and finally $\partial_{\mathbf{u}}\Xi(\mathbf{u}, t) - \boldsymbol{\lambda}^T(t)\partial_{\mathbf{u}}\mathbf{l}(\mathbf{u}, \mathbf{m}, t) + \ddot{\boldsymbol{\lambda}}^T(t) = \mathbf{0}$, and this removes all the intractable terms. The last identity is nothing but the adjoint equation

$$\ddot{\boldsymbol{\lambda}}(t) = \partial_{\mathbf{u}}\mathbf{l}(\mathbf{u}, \mathbf{m}, t)^T\boldsymbol{\lambda}(t) - \partial_{\mathbf{u}}\Xi(\mathbf{u}, t)^T \tag{3.16}$$

with initial conditions $\boldsymbol{\lambda}(T) = \mathbf{0}$, and $\dot{\boldsymbol{\lambda}}(T) = \mathbf{0}$. We can further simplify it by expanding $\mathbf{l}(\mathbf{u}, \mathbf{m}, t)$. Because the adjoint equation is solved backwards in time due to initial conditions, we rewrite it in terms of $T - t$. Finally, we rename $\boldsymbol{\lambda}$ as

$\mathbf{u}^\dagger$ and apply Leibniz's rule to replace $\partial_{\mathbf{u}}\Xi$ with $\partial_{\mathbf{u}}\chi$, to obtain

$$\ddot{\mathbf{u}}^\dagger(T-t) = (\mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m}))^T \mathbf{u}^\dagger(T-t) - \partial_{\mathbf{u}}\chi(\mathbf{u}, T-t)^T \quad (3.17)$$

with initial conditions $\mathbf{u}^\dagger(T) = \mathbf{0}$, and $\dot{\mathbf{u}}^\dagger(T) = \mathbf{0}$. We see that eq. (3.17) only differs from the initial semi-discrete wave equation eq. (3.12) by the adjoint operator, and the source term which is a partial derivative of the misfit function of choice. It also only differs from eq. (3.10) by the adjoint operator. A last step is then to verify whether we can use the same solver to solve the adjoint wave equation, i.e. whether the discrete operator $(\mathbf{M}^{-1}\mathbf{K})$ is self-adjoint. First, from (Leng et al., 2016) we know that $\mathbf{M}$ is diagonal in AxiSEM3D, so we only need to determine whether $\mathbf{K}$ is self adjoint. Furthermore, from (Leng et al., 2016), the computation of the stiffness matrix $\mathbf{K}$ in AxiSEM3D involves five operations: applying a differential operator $\mathbf{D}_1$, a Fourier transform operator $\mathbf{P}_1$ which is unitary and linear, multiplying with symmetric 3D elasticity matrix $\mathbf{C}$, then applying the inverse Fourier transform $\mathbf{P}_2 = \mathbf{P}_1^{-1} = \mathbf{P}_1^T$, and then applying the same differential operator $\mathbf{D}_2 = \mathbf{D}_1^T$. Hence $\mathbf{K} = \mathbf{D}_2\mathbf{P}_2\mathbf{C}\mathbf{P}_1\mathbf{D}_1$ is self adjoint.

We now see that eq. (3.17) is the same as eq. (3.10), except that it is discretised identically to the forward simulation (avoiding a high Fourier expansion for off-axis sources), thus rendering it much cheaper to compute with AxiSEM3D than the continuous adjoint. As mentioned previously, this means that the wavefield $\mathbf{u}^\dagger$ does not need to be physical in the discrete adjoint approach, but the gradients are exact with respect to the discretised model parameters. We can illustrate the difference in the case of AxiSEM3D: the wave equation is discretized on a 2D SEM mesh and in Fourier series in the azimuth direction. It is important to notice that the Fourier expansion is finite and dictated by the forward simulation with wavefield learning. We thus get the following accurate representation of the

wavefield:

$$\mathbf{u}(s(\xi, \eta), z(\xi, \eta), \phi, t) = \sum_{|\alpha| \leq n_u} \sum_{p,q} \mathbf{u}_{pq}^\alpha(t) l_p^N(\xi, \eta) l_q^N(\xi, \eta) e^{i\alpha\phi}, \quad (3.18)$$

where n_u is the local Fourier expansion order, $\mathbf{u}_{pq}^\alpha$ contains the Fourier coefficient α at the node p, q , and $l_p^N(\xi, \eta)$ is the p th Lagrange interpolating function of polynomial order N . We can then start from the discrete adjoint equation source term in eq. (3.17), and use our AxiSEM3D notation:

$$-\frac{\partial}{\partial \mathbf{u}_{pq}^\alpha} \chi(\mathbf{u}_{pq}^\alpha) = -\frac{\partial}{\partial \mathbf{u}} \chi(\mathbf{u}) \frac{\partial \mathbf{u}}{\partial \mathbf{u}_{pq}^\alpha} = -\frac{\partial}{\partial \mathbf{u}} \chi(\mathbf{u}) l_p^N l_q^N e^{i\alpha\phi} \quad (3.19)$$

where the result is the continuous adjoint source term in eq. (3.7), interpolated onto the forward simulation mesh.

The implementation of the discrete adjoint source is therefore equivalent to that of the discretised continuous adjoint source, with the difference that the Fourier expansion is imposed by the discretization of the forward simulation, and thus does not need to model a physically accurate point force source by using a very high and expensive Fourier expansion.

Finally, by removing all the vanishing terms due to the appropriately chosen Lagrange multipliers, as well as the vanishing partial derivatives of the initial conditions (usually just the earthquake, so independent of $\mathbf{m}$), we obtain the expression for the gradients with respect to model parameters

$$d_{\mathbf{m}} \chi = d_{\mathbf{m}} \mathcal{L} = - \int_0^T \boldsymbol{\lambda}^T (T-t) \partial_{\mathbf{m}} (\mathbf{M}^{-1}(\mathbf{m}) \mathbf{K}(\mathbf{m}) \mathbf{u}(t) + \mathbf{M}^{-1}(\mathbf{m}) \mathbf{f}(\mathbf{m})) dt \quad (3.20)$$

We now have the well known recipe to compute the gradients using the adjoint

method:

1. Integrate $\ddot{\mathbf{u}} = \mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m})\mathbf{u} + \mathbf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m})$ from 0 to T, with initial conditions $\mathbf{g}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$ and $\mathbf{h}(\mathbf{u}(0), \mathbf{m}) = \mathbf{0}$.
2. Integrate $\ddot{\mathbf{u}}^\dagger(T-t) = \mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m})\mathbf{u}^\dagger(T-t) - \partial_{\mathbf{u}}\chi(\mathbf{u}, T-t)$ from T to 0 with initial conditions $\mathbf{u}^\dagger(T) = \mathbf{0}$, and $\dot{\mathbf{u}}^\dagger(T) = \mathbf{0}$.
3. Compute the gradients with $d_{\mathbf{m}}\chi = -\int_0^T \lambda^T(T-t)\partial_{\mathbf{m}}(\mathbf{M}^{-1}(\mathbf{m})\mathbf{K}(\mathbf{m})\mathbf{u}(t) + \mathbf{M}^{-1}(\mathbf{m})\mathbf{f}(\mathbf{m}))dt$

As we have seen for the adjoint source in eq. (3.19), we do not need to bother re-deriving the expressions of the gradient for any specific parameter, which involves a lot of tedious algebra, as this will be identical to previous derivations obtained in the continuous adjoint framework. More importantly, we have established the equivalence of the continuous and discrete adjoint framework, i.e. the operations of adjoint and discretisation commute in our system. Therefore, we can appropriately discretise the results obtained in the continuous adjoint framework of (Tromp et al., 2005; Fichtner, 2010).

Note again that in the discrete adjoint case all the quantities are discretised on the grid of the forward simulation. For methods where the mesh is source- and complexity-independent, whether to use the discrete or continuous adjoint formulation does not matter. However in AxiSEM3D this is a crucial point, as it allows us to drastically cut down the Fourier expansion of the adjoint simulation, and thus the computational time. Very importantly, the discrete adjoint formulation highlights the fact that one can only update the discrete model parameters used in the mesh of the forward simulation. This is of course no different in the continuous adjoint case, but provides a direct and natural connection between the discretisation of the forward, adjoint and inverse problems: For instance, in the azimuthal direction we can design meshes that are much coarser than a standard 3D SEM

mesh, reflecting our limited data coverage and resolution. Such meshes of spatially varying element size are already used in tomographic inversions (Nolet and Montelli, 2005; Hosseini et al., 2020).

3.3 Continuous adjoint kernels

In this section we demonstrate how we obtain exact kernels in the continuous adjoint framework. As mentioned in the previous section, we need to model a point force source off-axis (at the receiver position), which in AxiSEM3D theoretically means an infinite Fourier expansion. First, we show that in practice, we just need a Fourier expansion that is high enough to correctly approximate the dominant wavelength in the simulation.

3.3.1 Physical adjoint wavefields

We model our source to be a nascent Dirac in azimuth, approximated by a Gaussian.

$$\delta_{n_u}(\phi) = \frac{2n_u}{\sqrt{\pi}} e^{-(2n_u\phi)^2} \quad (3.21)$$

Where n_u is the local order of Fourier expansion, and ϕ is the azimuth. We postulate that given a high enough Fourier expansion, this representation accurately models a point force. We use a constant Fourier expansion n_u^{const} pictured in figure 3.2 a). As we increase n_u^{const} we get to n_u^{sup} , which corresponds to the case where the azimuthal direction is sampled as densely as the D-shaped SEM mesh, and we do not exploit any structural smoothness, similarly to a traditional 3D wave equation solver.

We then put the off axis source on the equator. This is the worst case scenario

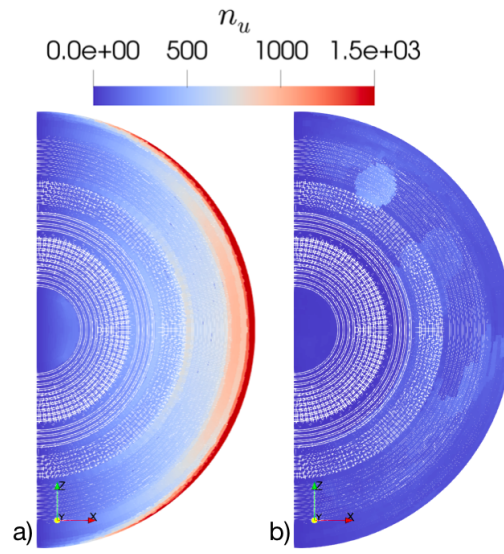


Figure 3.2: a): Constant Fourier expansion n_u^{const} . One can increase this constant Fourier expansion until arriving to the upper limit n_u^{sup} . b): Fourier expansion learned n_u^{wisdom} from a forward simulation on PREM (Dziewonski and Anderson, 1981) containing a 3D bubble. We can see that the expansion needs to be higher where the 3D bubble is located to capture the wavefield complexity arising from effects such as scattering. However, overall the wisdom Fourier expansion is much lower than the constant Fourier expansion.

n_u field \ (Lat, Lon)	(20°, −60°)	(−40°, 60°)	(20°, 90°)
$n_u^{\text{const}} = 100$	59.69	43.39	38.97
$n_u^{\text{const}} = 250$	30.65	18.02	14.05
$n_u^{\text{const}} = 500$	12.27	5.98	4.41
$n_u^{\text{const}} = 1500$	1.68	0.74	0.54
$n_u^{\text{sup}} = 2800$	0.51	0.23	0.17

Table 3.1: Normalised mean absolute error between offaxis seismogram and axial seismogram (in %) $\int_t (|s^{\text{offaxis}} - s^{\text{axial}}|) / \int_t |s^{\text{axial}}|$ with increasing n_u field for the stations displayed in figure 3.3.

where the azimuthal extent of the wavelength is the largest, thus requiring the largest Fourier expansion to model accurately. Finally, we place receivers on several locations around the Earth. We record seismograms for increasing spatially constant Fourier expansions n_u^{const} , shown in figure 3.3. Simulations are run at 30s dominant period, both in a spherically symmetric Earth model (PREM (Dziewon-ski and Anderson, 1981)), and in a 3D model (S40RTS (Ritsema et al., 2011)), to underline the fact that the computations work in fully 3D models.

Looking at the resulting seismograms and at the mean absolute errors errors in 3.1, we can see that they quickly converge to the reference seismograms obtained with a source positioned on the axis, as n_u^{const} gets closer to n_u^{sup} , showing that we accurately compute wavefields from an off-axis source by simply having a high enough Fourier expansion. This result intuitively makes sense because as we increase the Fourier expansion, we get closer and closer to a fully 3D mesh as used in other SEM solvers such as SPECFEM, except that our azimuthal dimension is not discretized with spectral elements.

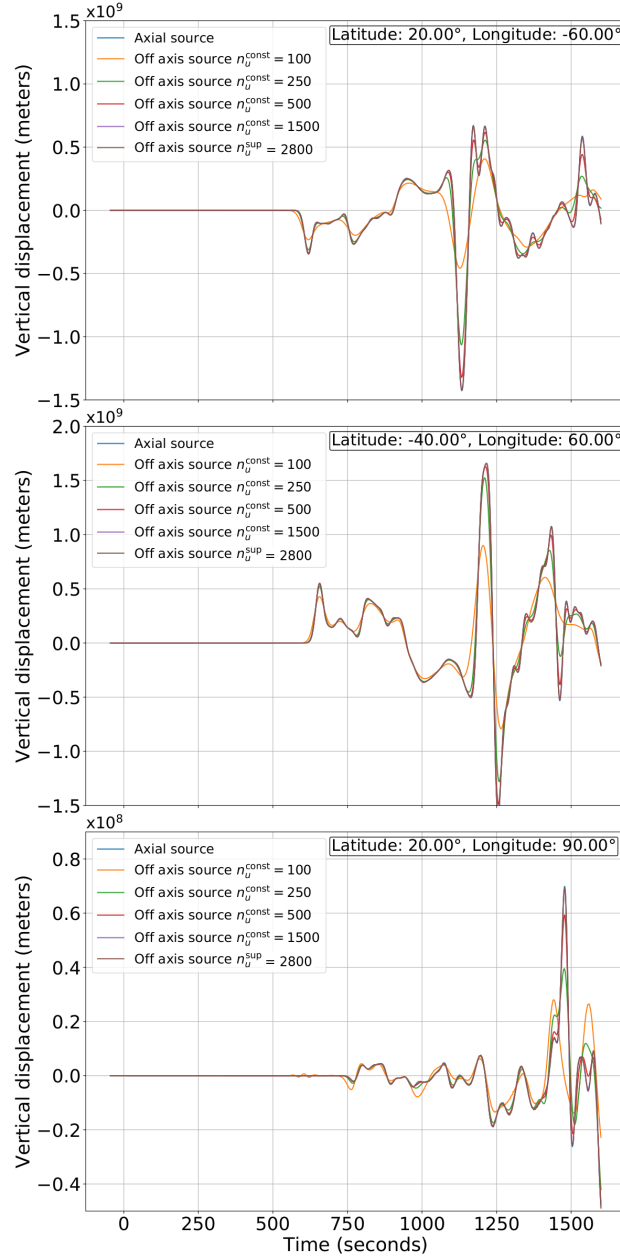


Figure 3.3: Converging seismograms for increasing values of the n_u^{const} field in s40rts, with a point force at the equator at 0 degrees azimuth. When the source is positioned on the axis of the SEM mesh, there is an analytical solution and the seismograms are exact, it is thus our reference. When the source is positioned off-axis, the seismograms converge to the exact ones as we increase n_u^{const} and get closer to n_u^{sup} . Amplitudes are scaled up to avoid floating point errors.

3.3.2 Continuous kernels benchmark

Now that we have established the ability of AxiSEM3D to accurately model off axis sources, we can compute kernels in the continuous adjoint framework. In the following figures, our source is located at the surface on the geographic North pole. We inject an M_{rr} only moment tensor, and our simulations are at a dominant period of 30s. The receiver is placed at an epicentral distance of 60 degrees. The computation is performed at n_u^{sup} , ensuring we propagate physical adjoint wavefields. We run tests in PREM and in S40RTS. For the benchmarks we do not add a 3D crust because it makes simulations quite a bit more expensive, but in section 3.4.2 we run a 15s simulation to produce a collection of kernels for a given earthquake in S40RTS + crust1.0 (Laske et al., 2013).

In figure 3.4 a) we can see a cut through the kernel obtained using the continuous adjoint approach for PREM, and in 3.4 b) for S40RTS. The fact that both kernels are essentially identical (as previously noted by (Liu and Tromp, 2008)) raises interesting questions about how sensitive to 3D structure the wavefield really is.

In traveltime tomography, we can examine the cross-correlation traveltime update δT_{xcorr} caused by a small perturbation to the model parameters (e.g. for a P wave velocity perturbation δv_P), computed by cross correlation between synthetic seismograms from the initial model, and perturbed seismograms from the model + perturbation (Dahlen et al., 2000). To benchmark our kernels, we can compare δT_{xcorr} to the traveltime perturbation predicted by the kernels

$$\delta T_{\text{ker}} = \int_{\Omega} K_{v_P}(\mathbf{r}) \delta v_P(\mathbf{r}) d\mathbf{r} = \langle K_{v_P}, \delta v_P \rangle_{\Omega}. \quad (3.22)$$

We want to emphasize that our framework is capable of computing kernels for other misfit measurements, for instance based on amplitude misfit, L2 misfit,

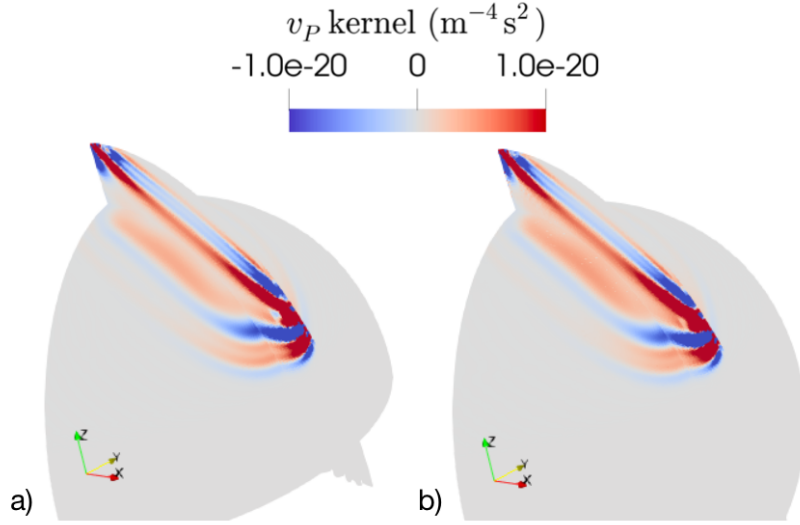


Figure 3.4: Kernel computed with AxiSEM3D in the continuous adjoint framework with the Fourier expansion n_u^{sup} , ensuring the adjoint wavefields are physical. a): result in PREM. b): result in s40rts. Note that the kernels are very similar.

phase and envelope etc. Indeed, this simply corresponds to a different adjoint source time function, as shown in the theoretical section. However, we choose cross-correlation traveltimes because of their linear stability with the strength of velocity perturbations, as shown in (Mercerat and Nolet, 2013). Indeed, we know that for the perturbations we choose, the linear traveltimes update is correct, and hence we expect a correct kernel to reproduce it. If we were to choose a more non-linear misfit such as amplitude, any error in the benchmark would inevitably be attributable either to the kernel accuracy, or to the non-linearity of the measurement (i.e. the accuracy of the Born approximation for that measurement), which would then not allow us to quantify the accuracy of the adjoint implementation.

Because our mesh is not a standard 3D mesh, we need to perform some manipulations to be able to compute δT_{ker} via this volume integral. From (Leng et al., 2016), we know we can rewrite the integral over the 3D domain Ω in the following

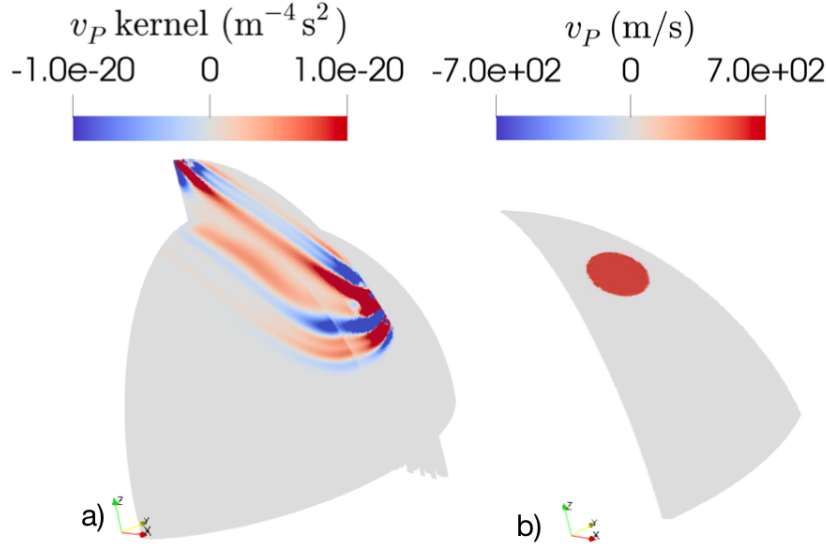


Figure 3.5: a) PREM kernel used in the benchmark. b) Azimuthal cut through 500km radius bubble at 5% strength used in benchmarks.

way:

$$\langle K_{v_P}, \delta \mathbf{v}_P \rangle_\Omega = 2\pi \sum_{|\beta| \leq n_u} \langle K_{v_P}^{-\beta}, \delta \mathbf{v}_P^\beta \rangle_D \quad (3.23)$$

where the right hand side is now an integral over the D shaped SEM mesh and n_u is the local Fourier expansion. What is more, the sum over complex Fourier coefficients simplifies: because the azimuthal quantities are purely real, the Fourier space quantities are hermitian symmetric.

$$2\pi \sum_{|\beta| \leq n_u} \langle K_{v_P}^{-\beta}, \delta \mathbf{v}_P^\beta \rangle_D = 4\pi \sum_{\beta \leq n_u} \Re \langle K_{v_P}^\beta, \delta \mathbf{v}_P^\beta \rangle_D \quad (3.24)$$

where $\Re$ denotes taking the real part.

The model perturbations we use are spherical (or "bubbles"), such as the one depicted in Figure 3.5 b). The center of the bubble is fixed, but we vary its radius and its strength relative to the background model.

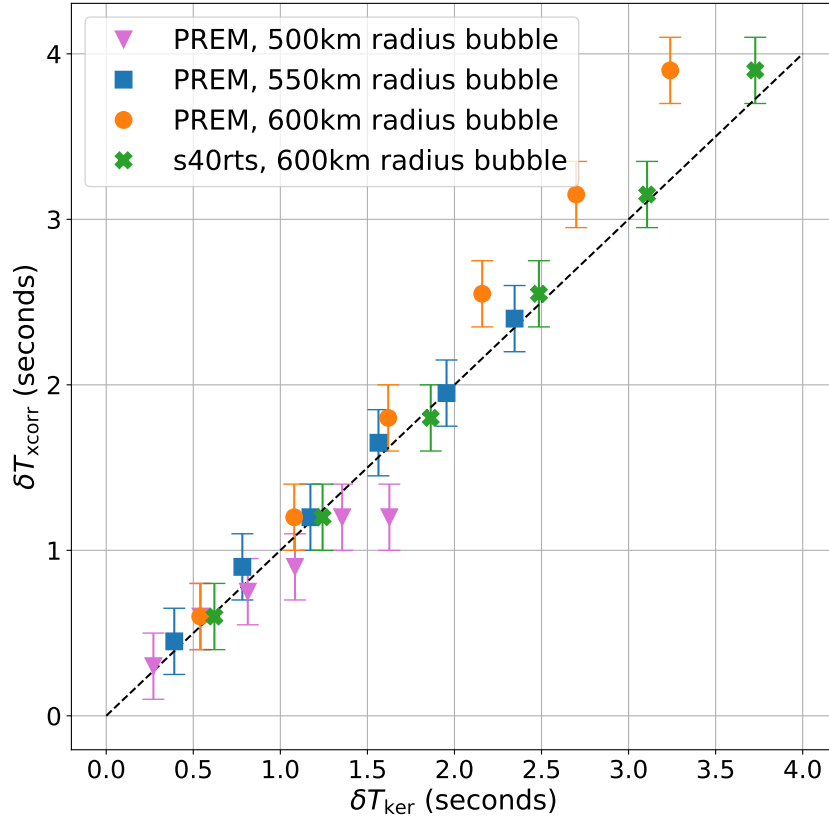


Figure 3.6: Traveltime perturbations predicted by kernels computed in PREM and S40RTS, compared with forward modelling prediction, for different perturbation geometries and strengths.

To start with, we compute δT_{ker} in the continuous adjoint paradigm, i.e. by using a high Fourier expansion that ensures that our adjoint wavefields are physical, and thus reproduce kernels such as in (Liu and Tromp, 2008). In figure 3.6 we plot the traveltime updates for 3D bubbles of varying radius and strength.

We can see that the traveltime perturbations computed from the kernels match the cross correlation values while the size and strength of the model perturbation allows to use the Born approximation. It is noteworthy to recognize that the domain of validity of the Born approximation seems to be different for PREM and s40rts.

We have shown that we can compute accurate sensitivity kernels in the continuous

adjoint framework. In the paradigm of AxiSEM3D this is not the most interesting approach because we lose most of the computational speedup offered by the flexible Fourier expansion, preserving the advantage mostly for the forward simulation.

3.4 Discrete adjoint kernels

Now that we have shown that the continuous adjoint kernels produce correct predictions, we can turn our attention to discrete adjoint kernels. In the discrete adjoint case, we do not enforce the adjoint simulation to produce physical wavefields. Therefore, we do not need a high and dense Fourier expansion, and we can simply reuse the Fourier expansion from the forward simulation. For our benchmarks, we will use a wisdom Fourier expansion that is sufficient to accurately model wavefields in our perturbed velocity model which includes the bubble. AxiSEM3D contains a feature that allows to obtain this sufficient expansion, called wavefield learning (Leng et al., 2019): Starting from a high Fourier expansion, the solver runs a simulation and progressively reduce the Fourier expansion while conserving the accuracy of the wavefield. We use the "wisdom" (i.e. learned) Fourier expansion shown in figure 3.2, which is learned for the 500km radius bubble. The learned expansion is significantly lower than the initial dense Fourier expansion.

We emphasize the fact that the wavefield produced by an off-axis source with this reduced Fourier expansion is not physical, as we can see on figure 3.7 where a seismogram from the off-axis source with the new learned Fourier expansion is compared to the exact seismogram from the same station recorded with the source positioned on the axis of the SEM mesh. However, as we saw in theory and will now see in practice, this is not a requirement to compute accurate gradients for the

discretized model parameters. This is the key idea of the discrete adjoint approach.

In figure 3.8 b) we look at the same slice from the kernel as previously, and it appears to be very similar to the one obtained from the continuous adjoint (3.8 a)), barring artefacts localized around the adjoint source (3.8 c)). Kernels are routinely smoothed and filtered around the source and receiver areas because of near-fields effects, therefore we do not concern ourselves with a discrepancy at these specific locations.

We can now repeat the experiment from the previous section, in which we computed the traveltimes updates obtained from the kernels for a given bubble, but this time using the discrete adjoint kernels computed in the wisdom Fourier expansion. The results are shown in figure 3.9, and as expected the kernels computed in the discrete adjoint framework produce the same traveltimes perturbation as the kernels computed with the continuous adjoint.

3.4.1 Computational cost

The appealing aspect of our implementation of the discrete adjoint approach is that it decreases computational cost for kernels compared to other 3D SEM based software. Indeed, the learned wisdom expansion for a given inversion is mostly frequency independent, but instead driven by structural complexity. The computation of the kernels, besides the I/O overhead for loading and dumping wavefields, only consist of a few extra FFTs to convolve the forward and adjoint wavefields, and is thus of the same time complexity as a standard simulation, i.e. $O(\omega^3)$ as opposed to $O(\omega^4)$ for methods such as SPECFEM (Leng et al., 2016). If we want to use the continuous adjoint framework however, the complexity will rise to $O(\omega^4)$, because the n_u^{sup} Fourier expansion is frequency dependent, just like in a standard 3D SEM method.

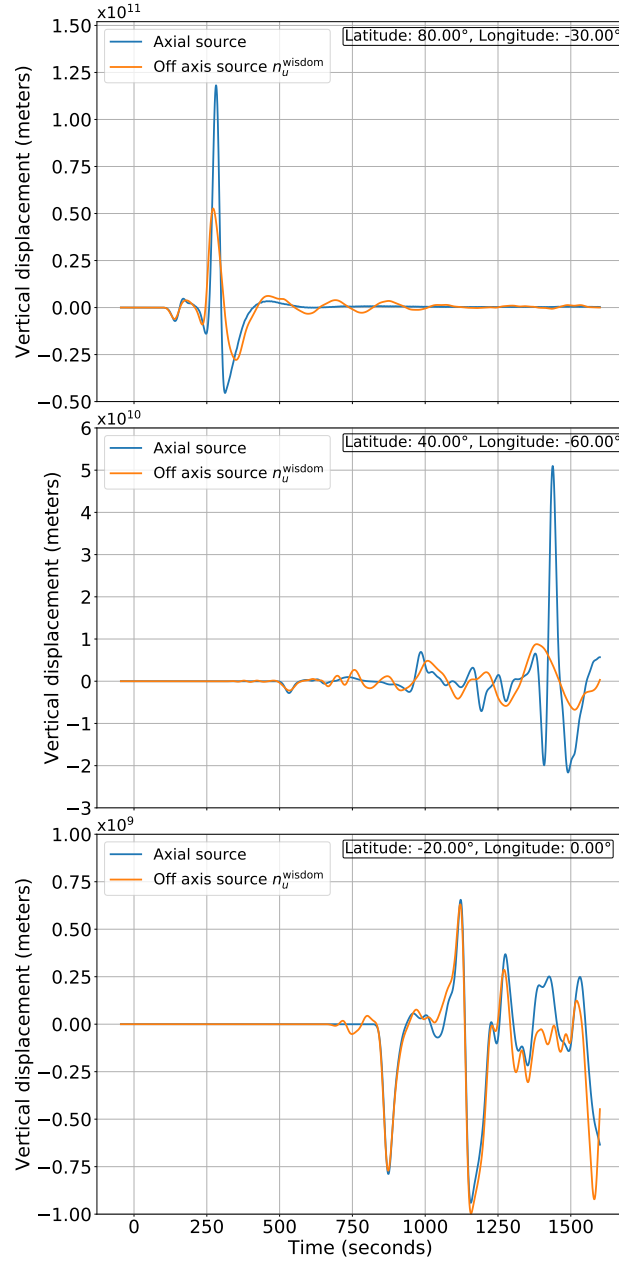


Figure 3.7: Seismograms obtained from an axial source, and the same seismograms with an off-axis source computed in the Fourier expansion learned for the 3D bubble, see Figure 3.2 b). As expected, the off-axis source with the n_u^{wisdom} expansion does not provide physically correct wavefields. This is the essence of the discrete adjoint method and the crucial point is that, as shown in figure 3.8 and figure 3.9, the kernels produced by these wavefields are exact. Amplitudes are scaled up to avoid floating point errors.

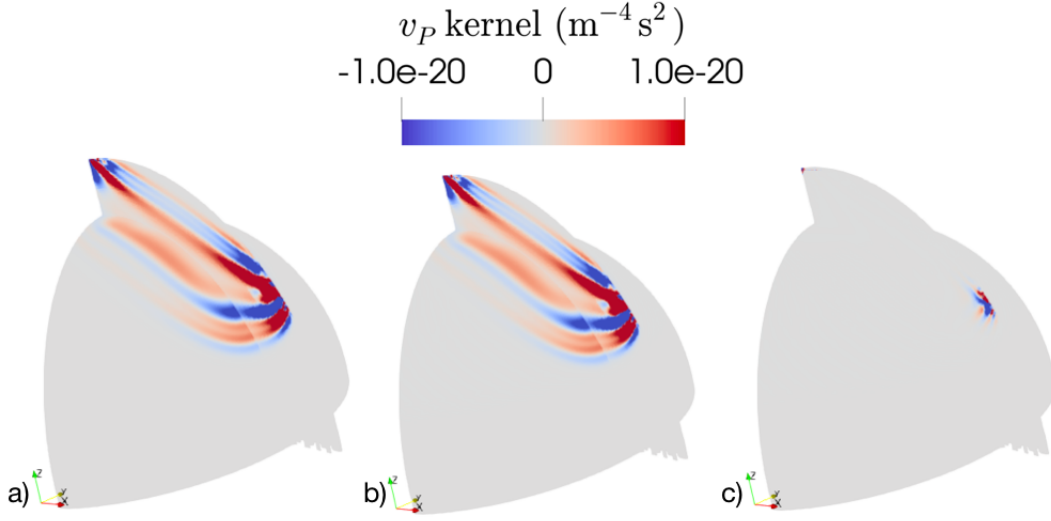


Figure 3.8: a): Cut from the continuous adjoint v_P kernel in PREM. b): Cut from the discrete adjoint v_P kernel in PREM. c): Difference between a) and b) kernels. The kernels are identical, except for some artefacts at the location of the adjoint source.

In figure 3.10 we compare the computational cost of a standard forward simulation with AxiSEM3D and a kernel computation, both in a discrete and continuous adjoint settings. The simulations are run in S40RTS and they produce 700s of seismograms. The figure reveals two main features: First, we see that for both the continuous and discrete adjoint approaches, the computational cost of the kernels scales similarly to a simple forward simulation. Furthermore, the discrete adjoint approach to kernel computation indeed scales much more advantageously with simulation period than the continuous adjoint approach.

3.4.2 Event kernels

In the above, we computed kernels from an individual source/receiver pair so that we can perform benchmarks. However, in FWI one typically computes kernels for all n_r receivers at once for any given earthquake owing to the efficiency of

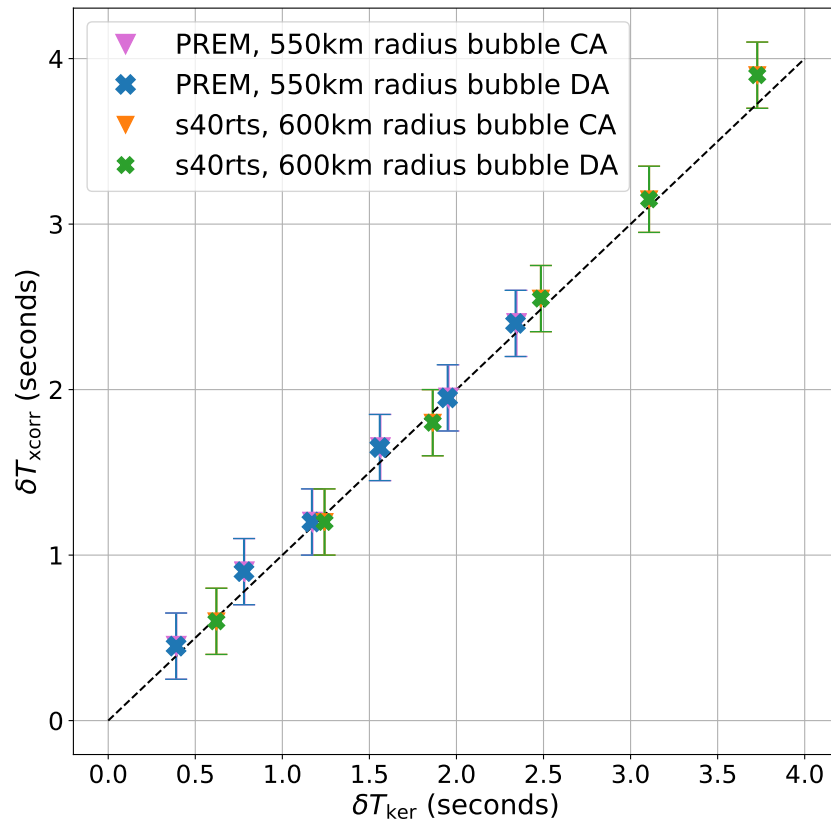


Figure 3.9: Traveltime perturbations predicted by discrete adjoint (DA) kernels and continuous adjoint (CA) kernels, compared to cross correlations predictions. The results from continuous and discrete adjoint kernels are identical.

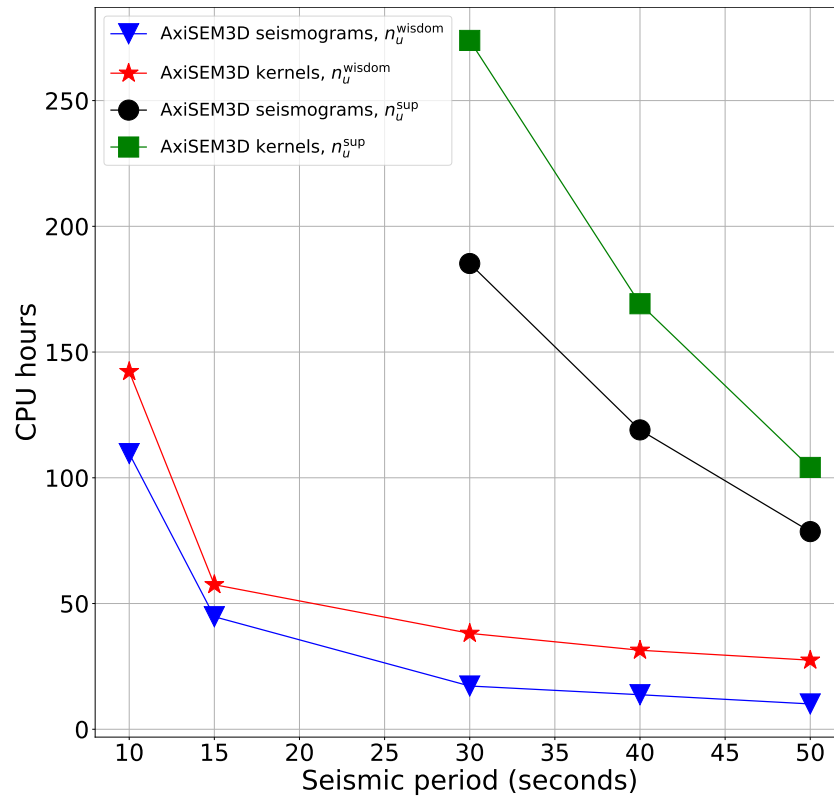


Figure 3.10: Computational cost of the forward ("seismograms") and the adjoint ("kernels") simulation for different seismic periods, in the continuous ("sup") and the discrete ("wisdom") adjoint approach.

the adjoint approach. Our discrete-adjoint framework is naturally capable of this feature. We compute an event kernel at 15s and one at 5s from the 2011 Virginia earthquake at 12km depth with magnitude 5.8. For an event kernel, the individual source time functions are multiplied by a measured cross correlation time delay between synthetics and real data. Since our example is purely synthetic, we introduce time delays for each station reminiscent of having actual observations at hand, by simply perturbing the computed seismograms by a given time shift. While in the benchmarks we did not include a 3D crust, we now run the simulation at 15s in S40RTS + crust1.0 (Ritsema et al., 2011; Laske et al., 2013), and we perform the run in a Fourier expansion n_u^{wisdom} learned during the forward simulation. We only place 5 receivers to render details discernible, but the same simulation principle and cost applies to any number of receivers. We compute latitudinal slices, as well as constant radius shells, shown in figure 3.11. The receivers are placed at various latitudes to illustrate different kernel shapes, from a close 20° epicentral distance kernel, to a much deeper penetrating kernel at 110°. We conveniently choose artificial travel time delays so that kernels at different latitudes are clearly visible on the same color scale. We can clearly see the traditional banana-doughnut shaped sensitivity of body waves.

For the 5s run in figure 3.12, we place three receivers at equal epicentral distance of 50°, so that we can choose artificial time delays that highlight changes due to measurements in similar kernels. The figure illustrates how emphasis is placed on regions where there is more misfit between synthetics and observations. We perform the simulation in S40RTS.

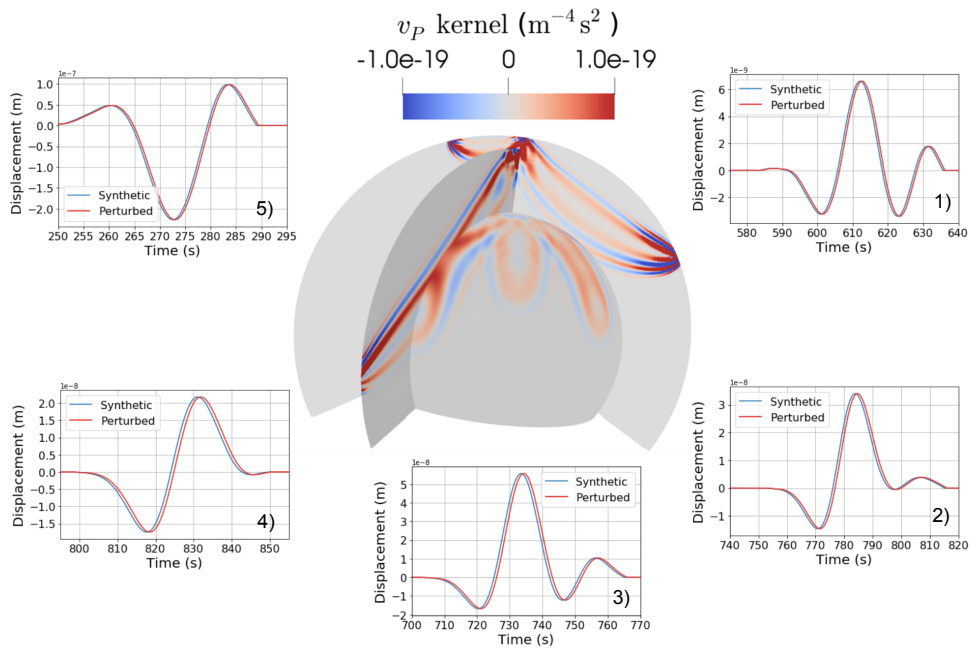


Figure 3.11: Event kernel at 15s with a magnitude 5.8 earthquake in Virginia, and 5 receivers, in S40RTS + crust1.0. We display three slices through the kernels, as well as a shell-like cut at 4000km radius. We also display the measurement used to compute the source time function of each kernel. "Synthetic" refers to the seismogram computed in S40RTS+crust1.0, and "Perturbed" refers to that same seismogram shifted by a small time offset, which artificially creates an observed seismogram. The time shifts used are (from seismograms 1) to 5)) 0.7s, 1s, 1s, 1s and 0.6s

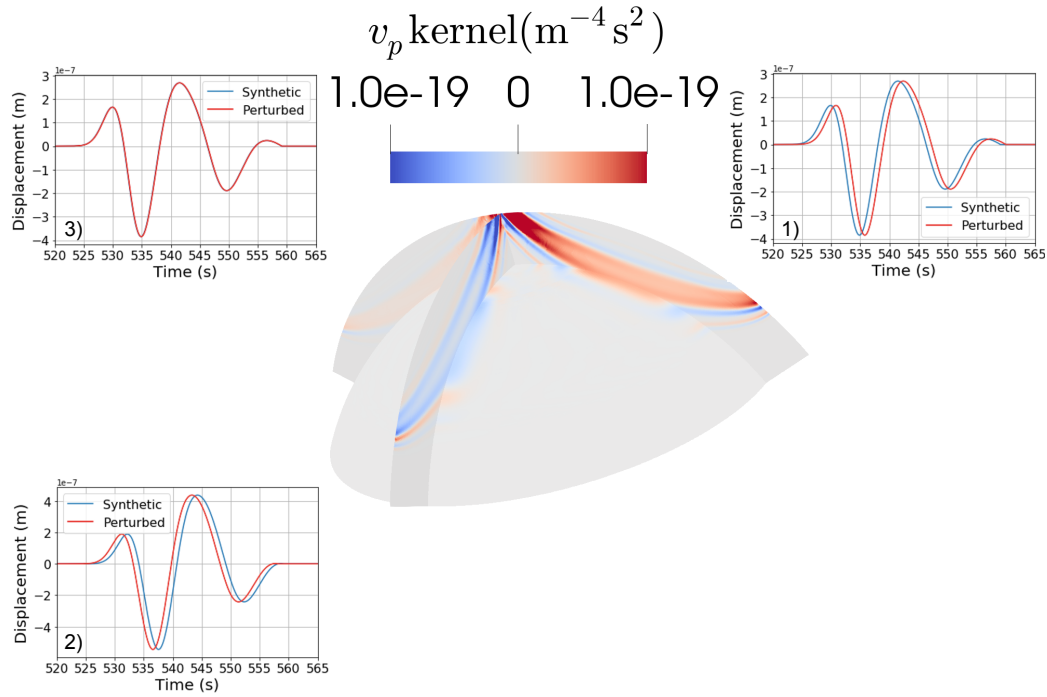


Figure 3.12: Event kernel at 5s with a magnitude 5.8 earthquake in Virginia, and 3 receivers, in S40RTS, at epicentral distance 50° . We display three slices through the kernels, as well as a shell-like cut at 5000km radius. We also display the measurement used to compute the source time function of each kernel. "Synthetic" refers to the seismogram computed in S40RTS, and "Perturbed" refers to that same seismogram shifted by a small time offset, which artificially creates an observed seismogram. The time shifts used are (from seismograms 1) to 3)) 1s, -1s, and 0.1s

3.5 Discussion and conclusions

We are introducing a new framework for computing the sensitivity kernels using adjoint methods in the context of an accelerated approach relying on the forward solver AxiSEM3D. Our approach exploits the inherent azimuthal smoothness of a global wavefield by using AxiSEM3D for both forward and adjoint simulations, which allows to flexibly adjust the mesh according to the desired complexity and thus lends itself very well to the discrete adjoint approach. We compute kernels in the continuous adjoint and discrete adjoint frameworks in global scale tomographic models, and we show that our kernels are exact by comparing them with cross correlation traveltime perturbations predicted from forward simulations with AxiSEM3D. The continuous adjoint approach is built upon the novel implementation of off-axis sources which are required to leverage the benefits of the adjoint method by igniting them simultaneously. Of course, our off-axis source implementation can also be used for other problems such as finite kinematic earthquake sources and source encoding ([Tromp and Bachmann, 2019](#)).

By using the discrete adjoint method, we are able to obtain kernels at a reduced computational cost and with reduced disk storage compared to traditional spectral element methods. Indeed, our method retains the $O(\omega^3)$ scaling with simulation frequency from AxiSEM3D for tomographic mantle models, as opposed to $O(\omega^4)$ for other software. Even though disk storage in AxiSEM3D is much lower than in other methods because of the flexible mesh parametrisation, we will in future work implement well documented checkpointing algorithms ([Komatitsch et al., 2016](#)), which allow to only store a subset of the time steps of the forward wavefield. These methods are more computationally expensive because they require recomputing part of the forward wavefield during the adjoint simulation. However, this does not change the frequency scaling and our method will remain $O(\omega^3)$, which

is the key benefit of our implementation.

The main idea of the discrete adjoint method, thanks to the discrete wave operator being self adjoint and therefore hermitian, is that it is not necessary for the adjoint simulation to produce physically accurate wavefields, instead the only requirement is that it is run on the same discretization as the forward simulation. This allows for the obtained kernels to be exact for the given discretized model parameters. In our case, this means we can keep the flexible Fourier expansion enabled by AxiSEM3D, and thus apply the complexity-adapted efficiency and speedup from the forward problem in the same manner for the inverse problem. However, we need to be cautious when choosing the Fourier expansion we work in. In our benchmarks we assumed a known expansion corresponding to the perturbed Earth model. In a tomographic inversion however, we have no knowledge of what is inside the Earth. If choosing an arbitrary Fourier expansion to work in, one may run the risk that the expansion is not high enough to accurately model the actual velocity model resolved by the data. This, of course, is always the case for any inversion, but the discrete adjoint approach illuminates this pre-defined fixed resolution problem clearly. This situation is identical to the problem of designing a tomographic inversion mesh, where one risks missing structure but also possibly overfitting the data. In that case, the mesh structure is determined by data coverage and resolution, with regions of dense ray paths having a fine meshing, and a coarser one for regions with less ray paths. Essentially, this means fixing the model resolution prior to running the inversion, which is exactly what we are interested in, because it will have very dense regions for which we'll have a high Fourier expansion, and very sparse regions for which the expansion will be very low. In subsequent work we will be designing such a mesh adapted to our framework, and demonstrating an example of global seismic tomography using this novel framework.

Acknowledgments

The authors thank Christian Böhm and Martin van Driel for helpful discussions related to the discrete adjoint method during and since the TIDES summer school in Oxford (2017). Comments by Pierre Jacquet, editor Sidao Ni and an anonymous reviewer significantly improved the quality of the manuscript. AS thanks the anonymous donor of his PhD grant which allowed this work to be carried out. KL and TNM were supported by NERC grant NE/R012199/1. The numerical simulations for this work were performed on TNM's student and NERC project grants on the ARCHER supercomputer, whose support is kindly acknowledged.

Statement of Authorship for joint/multi-authored papers for PGR thesis
To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (**only required where there isn't already a statement of contribution within the paper itself**).

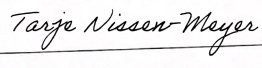
Title of Paper	A complexity-driven framework for waveform tomography with discrete adjoints
Publication Status	Published
Publication Details	Szenicer, A., Leng, K., and Nissen-Meyer, T (2019). A complexity-driven framework for waveform tomography with discrete adjoint. Geophysical Journal International 223 (2), 1247-1264

Student Confirmation

Student Name:	Alexandre Szenicer		
Contribution to the Paper	Wrote the code, performed analyses, wrote the manuscript.		
Signature		Date	30/04/2021

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Dr Tarje Nissen-Meyer		
Supervisor comments	Alex was the core author behind the underlying research and all aspects of this publication.		
Signature		Date	May 5, 2021

This completed form should be included in the thesis, at the end of the relevant chapter.

Postface

In this chapter, we have seen the central role simulations play in inference problems, and in particular the need to make these simulations cheaper to be able to explore problems at higher resolution while maintaining reasonable computational costs. In this postface I want to succinctly discuss two projects I worked on which aim to cut computational cost in the forward and inverse problem by leveraging deep learning. Unfortunately, I had to shelve these ideas for the time being in order to prioritize the other projects presented in this PhD. Nevertheless, I think they are worth including in this thesis, because they showed encouraging first results, and could be very interesting to explore further. All the data and codes are available upon reasonable request for those interested in continuing these projects.

Generating sensitivity kernels with deep learning

Frechet sensitivity kernels are central to seismic tomography, since they provide us with the parameter updates to produce our Earth model. The most physically accurate Frechet kernels, which incorporate the full physics of wave propagation, are obtained using the adjoint method, by convolving a forward and an adjoint wavefield. We set out to learn to generate kernels using a deep learning model, which could drastically cut computational and storage requirements. In particular, we'd like to be able to generate a kernel simply by providing source and receiver positions, moment tensor, and a measurement. To this end, we generate a database of 40s P wave kernels in PREM, with source depths ranging from 0 to 800km spaced every 10km, and receiver angles from 10 to 160 degrees, spaced every 0.5 degree. The kernels are computed using MC Kernel (unpublished yet, and in the last stages of benchmarking, available on [github](#)) (Stähler et al., 2016), which is based on AxiSEM wavefields (Nissen-Meyer et al., 2007) and allows for fast

computation of kernels in a 1D background model. The resulting kernels live on an inversion mesh, so we preprocess them by projecting them onto a set of 2D azimuthal slices on a square 128x128 grid, to match a traditional computer vision image format.

The network we use is an expanding net inspired by [Dosovitskiy et al. \(2016\)](#), where we feed each of the inputs through a fully connected layer, concatenate the resulting representations and reshape them into a (height, width, channel) format, then expand that representation using transpose convolutions until we arrive at the shape of our kernel. We use a per pixel mean squared error loss, and randomly split our data into a training and validation sets.

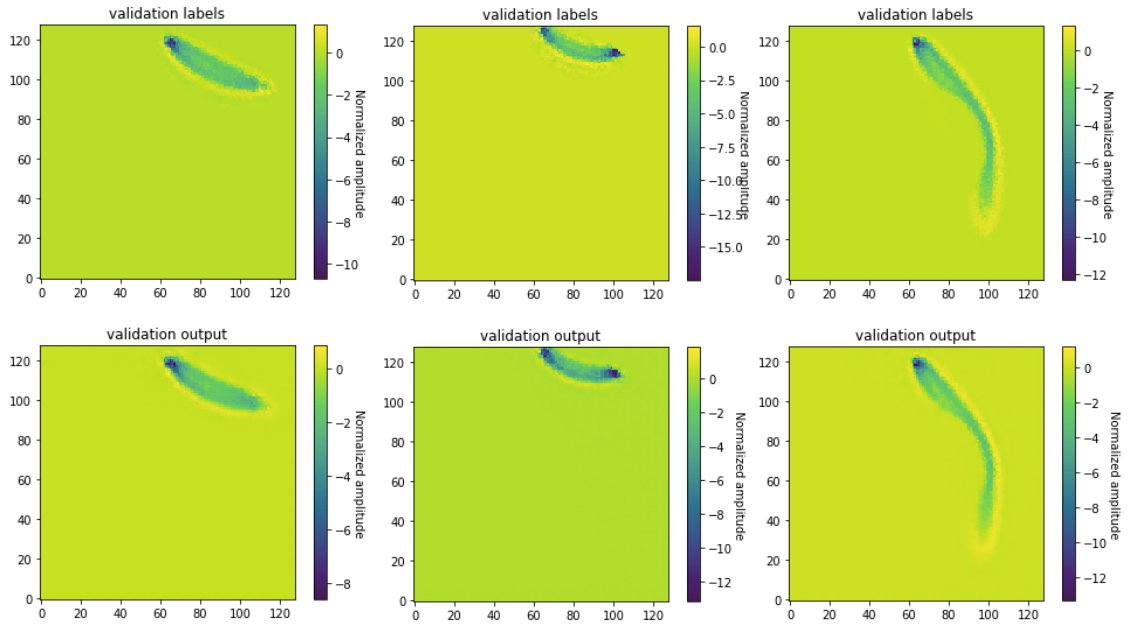


Figure 3.13: Ground truth and results from the trained CNN on validation data.

In figure 3.13 we can see some ground truth and predictions from our network. From visual inspection, the network seems to be able to learn to generate ker-

nels, although this is a very early result and warrants several important discussion points. First of all, since the training and validation splits are random, the validation data is very similar to the training data, and it is therefore not surprising we can produce a correct looking kernel. The main point for future work is to investigate how much can we generalise to unseen situations, i.e. if we only train on kernels with angles 10-30 degrees and 70-120 degrees, can we produce an accurate 50 degree kernel. Similarly, we want to be able to generalise to unseen moment tensors, source depths, etc. Indeed, neural networks are universal function approximators, and what we want to investigate in this project is, how much (or rather how little) datapoints do we need to provide, for the network to be able to accurately interpolate to the intermediate cases. Another important point is that we haven't so far investigated the accuracy of the kernel quantitatively, in particular if the generated kernel slices sum up to the correct travel time of the measured wave. In the future, we will regularize our network by adding a loss term on the sum of the kernel to enforce this condition. Finally, our current approach produces 2D azimuthal kernel slices, but we haven't yet designed the mapping back to the inversion mesh, where the parameter updates happen. In the future, we plan to transition to more modern approaches of geometric deep learning, which allow to perform learning directly on irregular meshes.

In conclusion, we believe that learning to generate kernels with deep learning is a promising approach to speed up the inversion process, and we hope that the databases and early results presented in this section can be picked up and greatly expanded by future students.

Encoder-Decoder networks for frequency upscaling of seismograms

In the previous paper, we have seen that computational costs of forward modelling methods have power law scalings with increasing frequency: $O(\omega^4)$ for traditional solvers, and $O(\omega^3)$ for AxiSEM3D. This means that if we want to double the frequency of our simulation, the computational time increases by 2^3 , which can very quickly become intractable as we go to high frequencies. And yet, higher frequency is a crucial direction for simulations. For instance, on a global scale, they could help us understand the nature of mantle dynamics and the origin of plate tectonics (Lei et al., 2020), by imaging new features invisible at larger wavelengths, such as Ultra Low Velocity Zones (ULVZ) (Thorne et al., 2021). On a local scale, high frequencies are important for instance in seismic hazard assessment, to reach the resonance regime of engineering structures, which can cause devastating damage. With this in mind, we attempt to design a deep learning approach for upscaling seismograms in frequency. There have recently been many approaches to produce seismic wavefields (or solve PDEs in general) using deep learning (Raissi et al., 2019; Smith et al., 2020; Moseley et al., 2020). Our approach is different in that we provide a low frequency seismogram as input, which contains a lot of information about the propagation medium, the source mechanism, etc. As such, we hypothesize that it is an easier problem to solve than fully approximating wave propagation using neural networks, and could be slower (we need to run a low frequency forward problem first) but more accurate, and could still be an incredibly useful tool for many applications. For instance, in the domain of earthquake hazard monitoring, one may want to produce shaking maps of an area when an earthquake occurs, in order to understand the most dangerous areas where rescuers need to be sent. In that case, one needs to run simulations

very quickly and accurately, because time is of the essence. A potential solution could be to run an accurate physics-based solver at relatively low frequencies, and upscale the resulting seismic fields using deep learning, in order to drastically cut computational time while attaining good accuracy of the shaking maps.

To investigate this approach, we generate low and high frequency seismograms, and attempt to map one to the other, while trying to generalise to novel source locations and velocity models. To begin with, we use a finite difference 2D acoustic wave equation solver, because it allows us to rapidly produce training data, and is an easier problem than mapping 3D viscoelastic seismograms.

We randomly generate 5 sources from model 1, one source from model 2, and one source from model 3, at a low frequency $f_1 = 10\text{Hz}$ and a high frequency $f_2 = 20\text{Hz}$. This generates 7 (number of sources) x 256 x 256 (size of the grid) data points, and we use the data from model 1 for training, and from model 2 and 3 as validation. All the sources are point forces with a gaussian source time function. We then train an encoder-decoder CNN to reconstruct high frequency seismograms from low frequency ones. A noteworthy fact is the low amount of simulations needed for our approach. Indeed, with one source, every grid point becomes a datapoint. While in practice we downsample the computational grid because nearby points are very correlated, each simulation still provides a large number of training data, which is in contrast to approaches such as ([Moseley et al., 2020](#)) which need each snapshot of the grid as a datapoint.

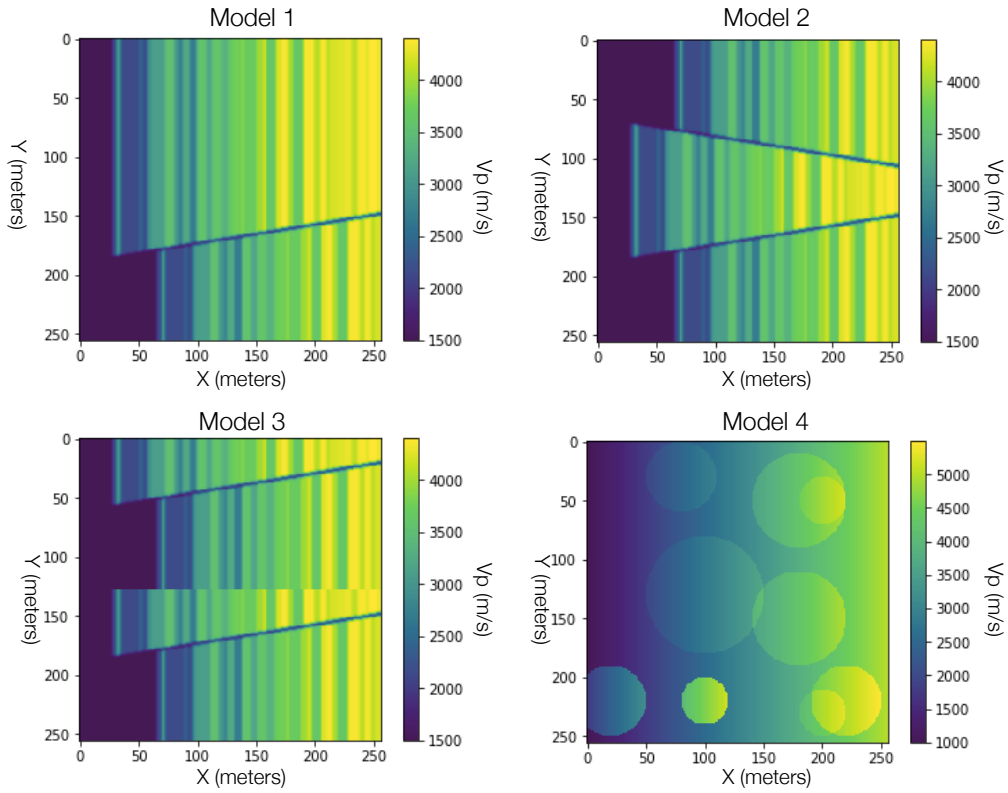


Figure 3.14: Models used to generate data for the super-resolution seismogram problem. Model 1 is used for training with 5 sources, models 2 and 3 are used for validation with one source each. Finally, model 4, which is very different from the previous three models, is used for testing.

The interesting part is the generalization ability of our network to new, unseen models. For instance if we randomly ignite a source in model 4 and run our CNN on a seismogram, we match the ground truth really well, even though the velocity structure is very different from the previous models (see 3.15).

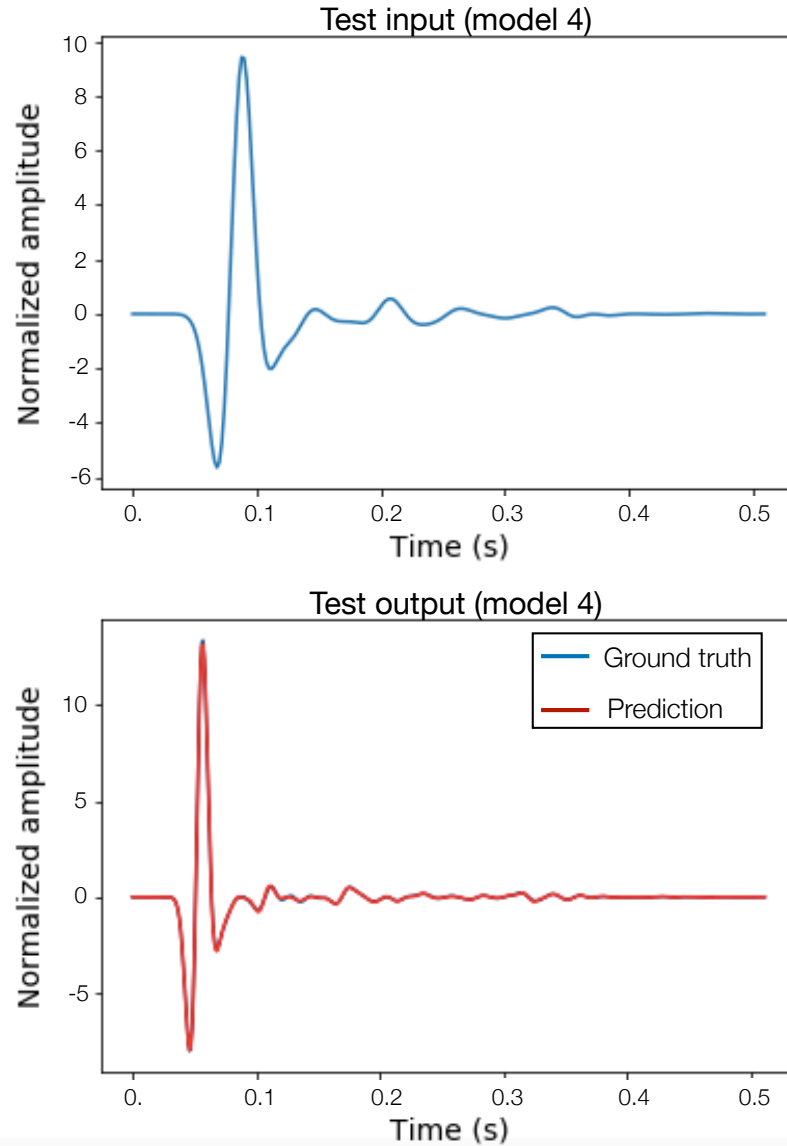


Figure 3.15: Results from the trained encoder-decoder CNN on the test set, generated from model 4 (see figure 3.14). Despite strong differences in the velocity model from those of the training and validation sets, we match the ground truth very well.

We believe these results are promising as an alternative approach to using neural

networks to directly solve PDEs. While the experiments so far are only run on 2D acoustic seismograms, the generalisation abilities of the network are encouraging. Our approach is in its very early stages, and there are several interesting steps to be taken next. First, we can test the generalisation abilities of our network further by running inference on increasingly different velocity models. Next, a very interesting investigation would be to be able to choose the frequency of the input and output seismograms. Indeed, currently we just double the frequency from 10 to 20Hz, but ideally we would like to be able to input a seismogram of any frequency and specify a frequency scaling. We believe simple methods such as passing a frequency scaling value to a fully connected layer, then concatenating the output with the latent space of the encoder-decoder, could be a good first approach. Finally, one crucial direction for realistic and useful applications would be to transition to 3D, first acoustic then fully viscoelastic.

4 | A Deep Learning Virtual Instrument for Monitoring Extreme UV Solar Spectral Irradiance

Preface

As we enter the final chapter of this thesis, I think it is worth prefacing the paper with a few words on my intellectual state of affairs at the time, as this project played a pivotal role in my development as a researcher. To be sure, this piece of work is in some sense the crystallisation of a desire to learn more about machine learning and apply it to my research in geophysics, which started quite early in my PhD. Before arriving to Oxford, I was completely oblivious to ML, and I discovered it by chance, with two of my College floormates undertaking PhDs in this field. Fairly quickly, I realised not only the huge potential of ML for geophysics, but also the incredible compatibility it has with my pathological curiosity, which evermore draws me to new fields of science and novel applications. Indeed, I found out that ML is a very versatile tool that I could successfully apply to many a problem in geophysics. I dare say that this endeavour has been fruitful, since, throughout my PhD, through various collaborations on either the main projects included in this thesis or on side projects in my free time, I have had the chance to work on applications of ML to many problems in geophysics related to satellites, elephant conservation, wildfire mitigation, earthquake hazard, etc; but also expand my intellectual horizon and explore fields outside of geophysics, such as shark conservation or real estate. Circling back to the personal importance of the work in this chapter, it is being accepted into the Frontier Development Lab with little real world practice in ML, gaining heaps of experience and learning from

incredible mentors, and subsequently having the freedom and trust from my supervisor to keep working on this project for my PhD to eventually achieve exciting results and a prestigious publication, which I credit for truly launching me onto the path I now walk on.

Published in Sciences Advances, DOI: 10.1126/sciadv.aaw6548

Note: While this project was started at the FDL program in summer 2018, a lot of the work has been carried out after the program ended, and until end of 2019, when I was again holding the DPhil status.

Note 2: More details about the dataset used in the work below can be found in the paper in appendix A.

Alexandre Szenicer^{1,2*†}, David F. Fouhey^{3,2*†}, Andres Munoz-Jaramillo^{4,2}, Paul J. Wright^{5,6,2}, Rajat Thomas^{7,2}, Richard Galvez^{8,2}, Meng Jin^{9,10,2} and Mark C. M. Cheung^{9,5,2}

¹ Department of Earth Sciences, University of Oxford; ² Frontier Development Lab; ³ Electrical Engineering and Computer Science Department, University of Michigan; ⁴ Southwest Research Institute; ⁵ Hansen Experimental Physics Laboratory, Stanford University; ⁶ SUPA School of Physics and Astronomy, University of Glasgow; ⁷ Department of Psychiatry, Academic Medical Centre, University of Amsterdam; ⁸ New York University, Center for Data Science, Center for Cosmology and Particle Physics; ⁹ Lockheed Martin Solar and Astrophysics Laboratory; ¹⁰ SETI Institute

* Corresponding authors, alexandre.szenicer@linacre.ox.ac.uk, fouhey@umich.edu

† These authors contributed equally to this work.

Abstract

Measurements of the extreme ultraviolet (EUV) solar spectral irradiance (SSI) are essential for understanding and estimating the drivers of space weather effects, such as radio blackouts, and increased aerodynamic drag on satellites during periods of enhanced solar activity. Our ability to monitor EUV SSI was compromised in 2014 due to an electrical malfunction of the Multiple EUV Grating Spectrograph A (MEGS-A) channel onboard NASA's Solar Dynamics Observatory (SDO). We propose a virtual replacement for MEGS-A by applying machine learning (ML) techniques to narrowband EUV images taken by the Atmospheric Imaging Assembly (AIA) onboard SDO. In particular, we show how to learn a mapping from AIA images to MEGS-A irradiance measurements using data from 2010 to 2014, when MEGS-A and AIA co-observed the Sun. We describe a protocol and baselines for measuring the performance of models. Our best performing ML model based on Convolutional Neural Networks (CNNs) outperforms other ML models, and a Differential Emission Measure (DEM) based approach, yielding average relative errors of under 4.6% (maximum error over all MEGS-A emission lines) and more typically 1.6% (median over all lines). Additionally, we provide evidence that the proposed method is solving this mapping in a way that makes physical sense and by paying attention to the magnetic structures known to drive EUV SSI variability.

4.1 Introduction

The extreme ultraviolet (EUV) radiation from the Sun is the dominant driver of the Earth's thermosphere/ionosphere system. During periods of elevated solar activity, enhanced solar EUV driving causes adverse space weather effects such as radio communication blackouts, increased aerodynamic drag on satellites in low-earth orbit, and scintillation of Global Navigation Satellite Systems (GNSS) signals ([Lean, 1987](#); [Tobiska et al., 2000](#); [Solomon et al., 2010](#)). NASA's Solar Dynamics Observatory ([Pesnell et al., 2011](#)) was launched in 2010 with the goal to help us understand solar activity and how their variations affect life on Earth. SDO carries three instruments. The EUV Variability Experiment (EVE) instrument suite ([Woods et al., 2010](#)) provides EUV spectral solar irradiance (SSI) measurements integrated over the full Sun. The spectrum measured by the MEGS-A and B modules of EVE contains emission lines from a variety of ions that exist at temperatures ranging from 7000 K to 9.3 MK. Of these, 39 lines can be extracted with relative ease and these define the Level 2 Lines data product. The Atmospheric Imaging Assembly (AIA) instrument ([Lemen et al., 2011](#)) captures full-disk images of the Sun at 4096 x 4096 pixels in seven EUV channels, two UV and one visible wavelength channels. The Helioseismic and Magnetic Imager (HMI) ([Scherrer et al., 2012](#)) delivers photospheric vector magnetograms and Dopplergrams of the full Sun at 4096 x 4096 pixels. The AIA and EVE instruments are two separate, although complementary, entities, and they are both crucial in their own right. EVE is meant to give us detailed information about the radiative budget of the Sun and how it affects the Earth's atmosphere, but without any spatial resolution. On the other hand, AIA is meant to give us information about the structure of the Corona, its density, temperature, and evolution, but with lower spectral resolution than EVE. Presently, AIA, HMI, and the MEGS-B mod-

ule of EVE continue to co-observe the Sun. However, an electrical malfunction of the EVE MEGS-A channel compromised our ability to monitor EUV SSI in the 5-37 nm wavelength range, which, as shown in figure 4.1, contains roughly 60% of the solar irradiance in the EUV. The goal of this work is to fill this gap in measurement capabilities with a virtual replacement for MEGS-A. Toward this goal, we examine how well a differential emission measure (DEM) based model and learning-based models can use narrowband AIA images to reproduce MEGS-A EUV SSI measurements.

This approach is feasible due to both a wealth of data from when AIA and EVE instruments co-observed the Sun and the common underlying solar dynamics that generate both sets of observations. Before it was taken offline, both the MEGS-A channel and the AIA instrument observed the Sun for roughly 4 years together, producing a set of paired AIA and MEGS-A observations from which we can learn a data-driven mapping. This mapping ought to exist because of three reasons:

1. The same population of solar plasma is expected to be responsible for emitting the EUV radiation observed by the two instruments.
2. As shown in figure 4.1, images taken by the AIA instrument have a significant overlap with the EUV spectra measured by EVE's MEGS-A module.
3. As shown in figure 4.1, each of the AIA filters has a different and overlapping response to radiation emitted by plasmas of different temperature, enabling the use of AIA filters as basis of a mapping between plasma temperature and irradiance.

There are currently two main approaches for using information contained in AIA images to reconstruct solar EUV irradiance. The first one uses AIA images to estimate the DEM of the solar corona (a combined estimate of the density and temperature of the coronal plasma along the line of sight) (Cheung et al., 2015), in combination with the CHIANTI Atomic Database (Dere et al., 1997; Landi and

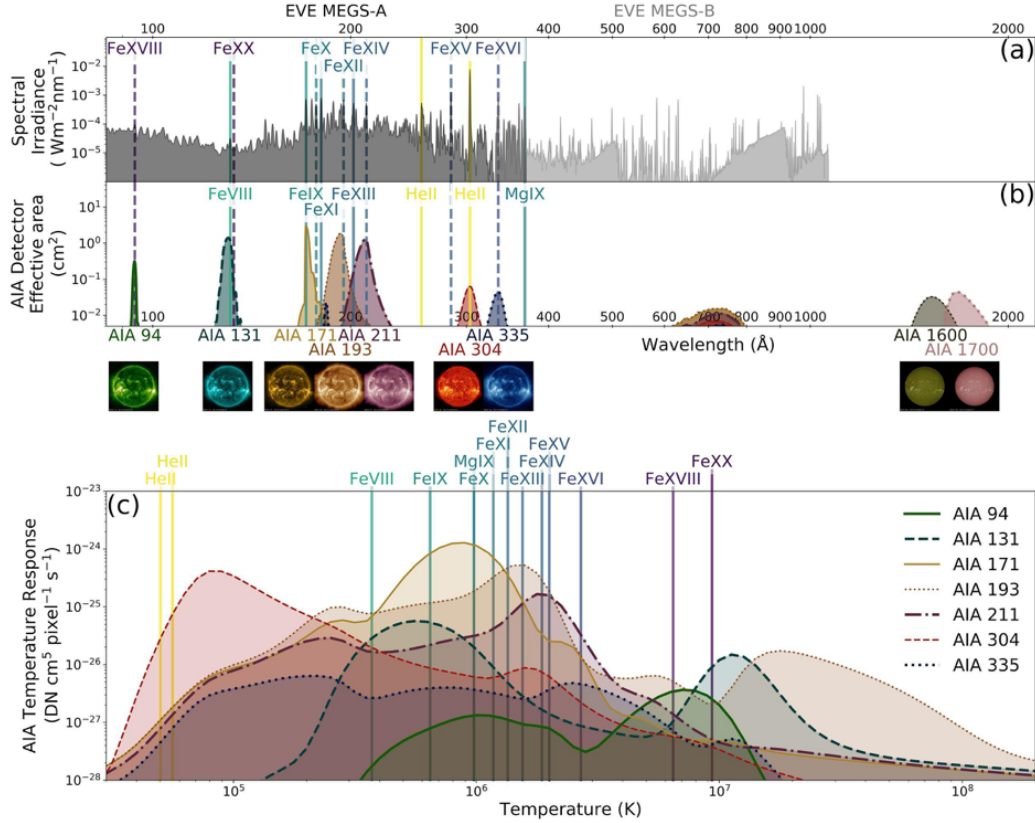


Figure 4.1: Spectral and temperature response of AIA and EVE. The AIA instrument on board of SDO makes narrow band images whose spectral response (b) overlaps with the spectra measured by the EVE instrument (a). Each of the 14 emission lines we recover in this project (shown in all panels as vertical lines) have a characteristic temperature associated with them (denoted using shades of color). Given that each of the AIA filters assembles light emitted by plasma of a wide range of temperatures (c), it is possible to combine their information to recover the part of the spectrum that used to be measured by the MECS-A instrument, which is no longer operational. For this EVE spectra (a), the MECS-A spectral region contains 60% of the total solar EUV irradiance. AIA images and EVE data from Jan-05-2014.

[Phillips, 2006](#)) to produce a full integrated spectrum. The second uses an empirical segmentation and classification of pixels of AIA images into different types of solar structures, each with a different radiative output ([Fontenla et al., 2016](#)), which is also integrated to produce a complete spectrum. However, due to the relatively sparse wavelength coverage of AIA, simplifying physical assumptions used in both methods (i.e. the assumption that emission is formed under coronal equilibrium conditions and that elemental abundances are uniform in the solar corona), and the empirical segmentation of images, inaccuracies in the mapping from AIA to MEGS-A are expected. In this work we examine how a data-driven approach can be used to provide an improved mapping between these two instruments.

We propose a data-driven approach that fuses two complementary models and takes advantage of recent developments in machine learning (ML). The first component is a linear model that transforms global summary statistics about the AIA input to produce the EVE MEGS-A output; this handles the bulk of the mapping, especially during quiet periods in solar activity. The second component is a convolutional neural network (CNN) model that uses AIA images to additively correct the output of this linear model, often during flaring periods. Both components are learned from data.

Given that this is the first application of CNNs to the problem of EUV spectral irradiance reconstruction, one of the contributions of this paper is a proper protocol and baselines for such an application. These include training and testing splits that can evaluate whether the method is generalizing as opposed to memorizing, and proper baselines to examine and better understand the contributions of more advanced methods. We also report results from a DEM-based model that aims to directly reconstruct the DEM distribution, as well as simple statistical models for the mapping problem, which can serve as benchmarks that future, more complex

methods of inferring EUV irradiance need to outperform.

Our proposed method is able to achieve strong performance with a median (across MEGS-A lines) average relative error of 1.6% and maximum error (on Fe XVI) of 4.6%, outperforming both our baseline models: a linear model on engineered features and a physics method based on DEM inversions. Furthermore, while our proposed method improves the overall results, it is significantly better on channels that are spectrally far away from the AIA observations than our baselines, and in circumstances that violate the underlying assumptions of the DEM-based model. Additionally, although ML models have a reputation for being indiscernible black-boxes, we show evidence that our learned model is solving the task in a sensible way: a careful specification of our neural network architecture provides spatially resolved maps that suggest how the CNN is using information to infer irradiance, and these maps match our physical intuition.

4.2 Results

4.2.1 Data Setup and Evaluation Criteria

We train and evaluate our learned mappings using data from the period between 1 January 2011 and 26 May 2014, when the SDO/AIA imager and the SDO/EVE spectrograph co-observed the Sun. Our AIA data derives from (Galvez et al., 2019), consisting of 9-channel AIA images at 6m cadence (in total: 264K images) that have been corrected for angular resolution variation and instrument degradation. We further downsample this data spatially to 256 x 256. The EVE MEGS-A data is also obtained from (Galvez et al., 2019), consisting of a 15-channel signal at 10s cadence. The Fe XVI 361 Å emission line has too few observations (available less than 1% of the time compared to other lines) to learn a meaningful

model and thus we work with the remaining 14 lines. We assign to each AIA observation the MEGS-A observation that is closest to it in time (mean/max time distance 8.5s/12s). Thus, the problem becomes the mapping of a $256 \times 256 \times 9$ measurement to a $1 \times 1 \times 14$ measurement.

Since the Sun’s dynamics at the level of total EUV irradiance move at a far slower rate than our observations, it is crucial to properly split the data into training (used for fitting model parameters), validation (for choosing model hyperparameters), and test (for evaluating prediction) sets. In particular, randomly splitting the data, commonly done in other settings, will yield overly optimistic performance estimates because access to one observation makes it easy to predict the subsequent observation. We therefore split our data year-wise. We set aside years 2012 and 2013 as testing data, and use the remaining data, covering 2011 and 2014, for training and validation. We split this sequentially into 70% training (1 January 2011 â 18 December 2011) and 30% validation (18 December 2011 - 31 December 2011 and 1 January 2014 - 27 May 2014).

We evaluate predictions using the relative absolute error (i.e., if y is the ground-truth observed irradiance and y_p is the predicted, we report $|y - y_p|/|y|$). This produces an error for every single observation and line. We summarize results into tables by taking the average over the observations, i.e., producing an average relative absolute error. We additionally report results only during flaring conditions. Rather than manually specifying flaring times, we use ground-truth EVE Fe XX observations as a proxy variable for flare conditions, and define all observations in the top 5% of Fe XX emissions across the whole dataset as flares.

4.2.2 Baseline Models

To put our results into proper context, we first report the results of a number of simpler methods. We begin with an approach that models the physics of the problem, and then report the performance of a set of models that are data-driven, but use only non-spatially resolved summary statistics about AIA measurements. These results additionally provide future experiments about MEGS-A nowcasting with proper baselines.

Differential Emission Measure Model. To illustrate the need for a data-driven solution, we first analyze the results of a model that uses only a physical model. In particular, we transform each set of 6 EUV images (i.e. all AIA EUV channels except the 30.4 nm) into DEM maps (Cheung et al., 2015) spanning 18 temperature bins from $\log T = 5.5 - 7.2$ K, at $\Delta \log(T) = 0.1$. This is done with the assumption that the emitting plasma is optically thin, that the emission is formed under coronal equilibrium conditions and that elemental abundances are uniform in the solar corona. With the same assumptions, the DEM solutions are then used to compute the emission of EUV lines observed by EVE MEGS-A. Once the pixel-by-pixel EUV line emissions have been obtained, we integrate them spatially to obtain the line irradiances (as reported by MEGS-A). The specific inversion method used here is the neural DEM method (DeepEM) outlined in Wright et al.. This implementation is able to provide solutions where the underlying Linear Programming (LP) solver of standard basis pursuit DEM inversion technique (Cheung et al., 2015) fails to converge on a satisfactory solution. The DeepEM method was trained on data from 2011, mirroring our models.

Channel Summary Statistics Models. We expect that many of the MEGS-A observations can be explained empirically by the total intensity of the AIA images since AIA observes data at nearby wavelengths to the MEGS-A spectrum (see

figure 4.1). To demonstrate this, we use features that are the result of averaging spatially, i.e., producing a single feature per-AIA channel and then fit a series of models to these summary statistics features. As features, we use the average AIA data-count and standard deviation of each AIA image (AIA_μ and AIA_σ , both 9-dimensional feature vectors), i.e. $AIA_{(\mu,k)} = \frac{1}{256^2} \sum_{i=1, j=1}^{256, 256} AIA_{(i,j,k)}$ and similarly for AIA_σ . The first feature captures the total irradiance and the second reveals the extent to which that irradiance is constant across the Sun. While AIA_σ is non-linear with respect to the original AIA image, AIA_μ is linear and any linear model fit to it demonstrates to what extent a properly parameterized linear model explains the observed MEGS-A data.

We report results for a number of models. All model parameters were fit with the training set and relevant hyperparameters were fit on the validation set. See materials and methods section for more precise details. (i) We begin with the most basic model, a least-squares fit to the average AIA image, or L2-mean, which represents the most basic data-driven approach to this problem. (ii) Our most complex and effective method is Huber-mean-std, or finding a linear model on AIA_μ and AIA_σ features that minimizes the robust Huber loss; regularization and Huber parameters were optimized on the validation set. While it has few parameters, this model already involves both feature and model engineering. (iii) To evaluate the importance of the loss function compared to standard least-squares, we also report results for L2-mean-std, or the same model, but fit to minimize a squared Euclidean distance; (iv) Similarly, to evaluate whether the non-linear standard-deviation features help, we report Huber-mean or fitting the parameters only to AIA_μ . Note that since the mean is a linear operator, this evaluates how well a properly fit linear model can explain performance.

We report absolute relative error for the above methods in table 4.1. Before discussing the results in detail, we point out that some of the MEGS-A channels

Line and log T	He II 4.7 303Å	He II 4.8 256Å	Fe VIII 5.6 131Å	Fe IX 5.8 171Å	Fe X 6.0 177Å	Mg IX 6.0 368Å	Fe XI 6.1 180Å	Fe XII 6.1 195Å	Fe XIII 6.2 202Å	Fe XIV 6.3 211Å	Fe XV 6.3 284Å	Fe XVI 6.4 335Å	Fe XVIII 6.8 93Å	Fe XX 7.0 132Å	Min	Mean	Median	Max
Overall																		
Physics	8.66	6.09	3.53	2.76	2.86	3.58	2.66	2.62	2.92	2.78	4.90	5.69	13.07	13.39	2.62	5.39	3.55	13.39
L2 Mean	2.70	2.48	1.95	1.53	1.95	3.80	1.76	1.63	2.05	1.97	4.28	10.03	1.78	11.71	1.53	3.54	2.01	11.71
L2 Mean-Std	2.48	2.40	2.72	2.15	2.77	4.56	2.30	2.30	2.63	3.34	3.81	7.48	2.53	10.85	2.15	3.74	2.67	10.85
Huber Mean	1.15	1.36	1.70	0.68	2.00	4.90	2.36	1.83	1.80	2.00	5.63	10.51	1.74	3.55	0.68	2.94	1.92	10.51
Huber Mean-Std	1.27	1.74	1.21	0.79	2.04	4.71	1.92	1.39	1.33	2.44	4.92	7.12	1.44	2.67	0.79	2.50	1.83	7.12
Linear+CNN	1.31	1.24	1.67	1.95	1.19	3.21	1.15	1.00	1.50	2.40	3.72	4.61	1.28	2.07	1.00	2.02	1.59	4.61
Flare Conditions (FeXX 95 percentile)																		
Physics	18.18	15.48	16.57	3.87	1.91	3.23	2.33	2.60	4.80	2.91	4.41	4.83	66.20	64.00	1.91	15.09	4.60	66.20
L2 Mean	2.16	2.25	2.99	2.07	3.02	5.06	2.84	2.57	2.94	2.93	3.20	5.98	4.36	11.52	2.07	3.85	2.97	11.52
L2 Mean-Std	4.18	4.64	5.35	3.71	4.79	7.02	3.96	4.13	4.18	5.06	5.62	5.70	5.20	11.09	3.71	5.33	4.93	11.09
Huber Mean	1.58	1.47	2.15	0.98	1.96	4.59	1.72	1.27	1.25	1.76	4.31	8.45	2.64	17.91	0.98	3.72	1.86	17.91
Huber Mean-Std	2.05	2.39	1.81	1.24	3.07	4.82	2.36	1.94	2.82	2.71	4.50	5.23	4.04	10.67	1.24	3.55	2.77	10.67
Linear+CNN	1.89	1.55	1.69	1.79	1.48	3.32	1.30	1.41	2.30	2.15	3.27	3.85	2.03	7.55	1.30	2.54	1.96	7.55

Table 4.1: Per-line average relative error for the methods evaluated in this paper. We provide summary statistics for different models for all lines, where the values are percentages of the average relative error defined in the Results - Data setup and evaluation criteria section. The top part of the table indicates results averaged over the entire test set, whereas for the bottom part only data points where the Fe XX line is in the 95th percentile are kept, to focus on flare conditions. Numbers in bold indicate the best result within the given column.

should be readily predicted from AIA observations while others should not: some MECS-A channels are spectrally close to AIA observations.

The DEM-based model obtains strong performance on a number of lines with 6 / 14 lines predicted with under 3% error. Its performance on the He II 304 Å is far worse, since the plasma is optically thick for this line and since it forms in non-equilibrium conditions (14). For these reasons, AIA DEM inversions do not even use data from the He II 304 Å channel. With this in mind, it is not a surprise that the pure DEM-based model gives predicted He II 304 Å irradiances that are a factor 37 smaller than the corresponding MECS-A measurement. The metrics for this line quoted in table 4.1 have been adjusted to account for this huge discrepancy.

Some other lines, such as Fe XX (which is sensitive to plasma at about 10 million K) have discrepancies of 10% during quiet conditions but the average error can increase to 60% during flares. This may be due to a number of reasons. Firstly,

saturation of the AIA CCD detectors during flares can result in bleeding of signal along columns of pixels that overlap with the flaring site. Secondly, automatic exposure control during flaring times (which aim to mitigate CCD saturation) can degrade the signal-to-noise of AIA images. Thirdly, there is evidence that coronal plasma in the quiescent state may have different elemental abundances than plasma that have been impulsively evaporated to fill flare loops ([Warren, 2014](#)).

A properly fit linear model (Huber-Mean) results in better performance on most lines compared to the physics model and strong performance on many lines overall. Many emission lines (Fe VIII, Fe IX, Fe XII, Fe XIII, He II) can be explained well by a linear transformation of the average of the AIA observation, with average relative errors close to 1% and as low as 0.7%. While the linear transformation is simple, choosing the right loss function is important: the model fit with Huber regression consistently improves over L2.

While a linear model applied to the average of the AIA observation does well on many lines, it performs poorly on many others, including Fe XX, Fe XV, Fe XVI, and Mg IX. Adding standard deviations as a feature gives a proxy variable for the presence of flares, and thus Huber-Mean-Std method improves performance on these flare sensitive lines, reducing e.g., Fe XVI prediction from 10.5% relative error to 7.1%). This performance gain on flare-sensitive lines comes at the cost of modest increases in error for the lines that are modeled well by a linear transformation of the AIA data: the larger capacity models start fitting to noise in order to better explain the training data. Nonetheless, during flaring conditions, performance degrades: the Fe XX emission prediction error is 440% of the error seen over all conditions.

The best performing linear model (Huber-Mean-Std) is outperformed on some lines (Fe XV, Fe XVI, Mg IX) by the physics model and one might naturally won-

der whether the linear model is simply underpowered. We experimented with two standard ways of adding expressiveness to the model: adding pairwise interaction features between all the terms (Hastie et al., 2009) and applying gradient boosting regression trees (Friedman, 2001), a standard non-linear model. Both failed to substantially improve upon performance of the Huber loss on mean+std model, with average errors of 2.5% and 2.8% respectively, as compared to 2.5%. Additional information, it seems, should come from either additional physics knowledge to provide a better fitting model (as used in the DEM-based model) or via examining the spatial information in the image (e.g., actually looking at the flare rather than inferring their presence via features).

4.2.3 Convolutional Neural Networks

Having demonstrated what appear to be the limits of models that consider only first and second moment summary statistics of AIA data, we now present results from models that examine them spatially, in particular convolutional neural networks. These models build a single function that is optimized end-to-end, consisting of interleaved convolutions (that aggregate spatial features) and non-linearities (that enable the composition of these convolutions to learn non-linear functions). In the process, the model discovers features that, when extracted from the image, help make good inferences.

These models face large learning challenges when applied to the AIA to MEGS-A nowcasting problem. First, there is very little supervision available for learning the mapping: only 80K data points are available for training and these data points are densely sampled in time, causing correlations and decreasing the effective number of data points. Second, a great deal, but not all, of the mapping can be explained by a linear transformation of basic summary statistics of the AIA irradiance, and the CNN must capture this linear mapping while also incorporating

other features visible in the images. While the CNN could, in principle, compute the average AIA intensity via a set of properly designed filters, the CNN’s parameterization and random initialization in practice might encourage it to converge to a more complex solution (e.g., consisting of a series of basis expansions followed by a linear model).

We avoid this problem by training a model that explicitly corrects the predictions of the linear model as shown in the bold components of figure 4.2. We fit a linear model to the training set, using mean and standard deviation features to represent each AIA observation and minimizing huber loss, and then compute the residual, $y_{diff} = y - y_{linear}$, between ground truth EVE irradiance and the linear model’s prediction on the training set. This y_{diff} serves as the learning target for the CNN. We adjust the input of the CNN to compensate for this: the CNN is predicting a variable that has had, in some sense, the information about its average image removed and must then focus on the spatial details. We accordingly compute the per-channel mean of the image, and subtract it, yielding an image that is, per-channel, each pixel’s deviation from the mean.

At inference time, the network produces an estimate of y_{diff} , which is added to the prediction from the linear model. The inference-time procedure is thus computing per-channel statistics over the input and making a prediction using the linear model, using these per-channel mean of the AIA image to convert the AIA image to a deviation-from-the-mean image, and making a correction prediction via a CNN.

As a side-benefit, this method also uses the linear model to explain a great deal of the variability in the output, which we will show later has a number of advantages. One such advantage is that the linear model can be turned into a spatial model as well (i.e., produce a per-pixel prediction), which we demonstrate in a subsequent

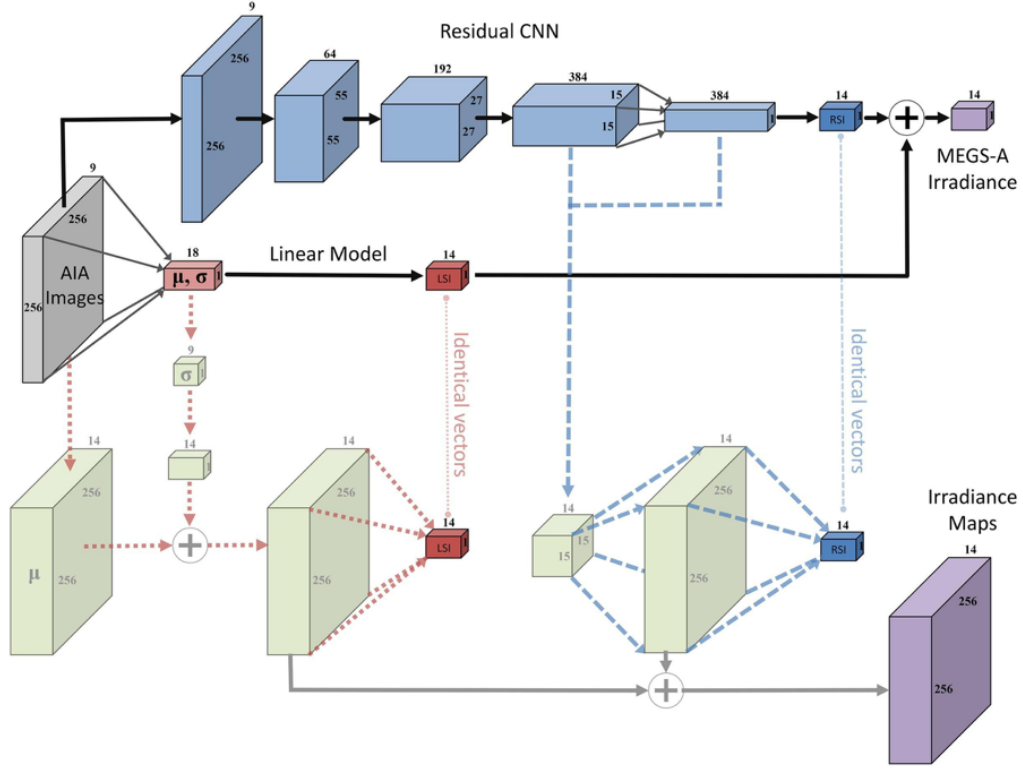


Figure 4.2: Proposed Neural Network Architecture. After computing summary statistics of the input AIA images and making a prediction via a linear model, a convolutional neural network makes a prediction that corrects this linear model. The combined linear+CNN model is shown in bold colors and arrows. The numbers attached to the boxes denote the sizes of the representations of the data as it goes through the network, e.g. the first block annotated with 256, 256 and 9 represents an input of 256x256 pixels and 9 channels. We produce spatially resolved maps in units of irradiance to validate how the CNN is operating (see figure 4.5) by rearranging commutative operations in the last two layers of the CNN (blue dashed path) and the linear model (red dashed path). These operations yield identical outputs as the original (bold) model (illustrated with faint vertical lines), but recasting them this way enable the diagnosis of the model’s operation.

section.

CNN Model configurations: Due to the size of the data (recall the data is both <80K observations and samples are highly non-iid), we experimented primarily with smaller networks. We obtained strong performance with a shallow network derived from AlexNet (Krizhevsky, 2014) with approximately 1 million parameters. We could obtain similar performance from more standard networks, including an 18-layer Resnet (He et al., 2016b) (11.7 million parameters), but found that the high capacity of the model coupled with limited variability of the data led to severe overfitting even with regularization. In contrast, the shallower approach worked more consistently. All networks were trained from random initialization (see implementation details in Materials and Method) to minimize the Huber loss. While transfer learning typically improves results, early experiments yielded negative results when initializing with ILSVRC-pretrained models (Russakovsky et al., 2015) (by replicating the RGB filters 3x in the channel dimension to create 9-channel filters similar to (Carreira and Zisserman, 2017)).

Our proposed method, which we denote ANet3 consists of three convolutional layers derived from Alexnet, followed by average pooling, and a linear map. In particular, we use the convolutional blocks from (Krizhevsky, 2014) (i.e. convolutions, ReLUs, and max-pools). We adapt the input channel count to match AIA and append a batch normalization layer after each convolution since batch-norm has been shown to be empirically effective. After 3 convolutions, the feature map is averaged just as in a Resnet (He et al., 2016b), passed through dropout for regularization, and then linearly transformed into the EVE prediction.

Results: We report results in 4.1. The method obtains the lowest overall mean error. This error, however, conflates lines for which a linear transformation of the total AIA-observed irradiance suffices and lines for which a linear transformation

is a poor model. The proposed model improves on the best of either Huber-Mean and Huber-Mean-Std on 9/14 lines, with especially strong performance gains on modeling Fe XX, Fe XV, Fe XVI, and Mg IX, achieving a 26% error reduction on Fe XVI. This trend continues under flaring conditions, where the CNN reduces the error rate on Fe XX emission prediction from 10.7% (Huber Mean-Std) to 7.6%. At the same time, as was the case with adding standard deviation features, the increased model capacity leads to performance losses on lines that are well-described by a linear model: in particular lines like Fe IX, Fe XIV, He II are slightly degraded by the use of a CNN.

The methods used are far shallower than typical approaches used in learning-based computer vision and one might reasonably wonder whether a more off-the-shelf approach might work. Here, we report a number of additional experiments that help justify our approach. Our early efforts tried a standard Resnet18 (He et al., 2016b) trained to directly map from AIA to MEGS-A. We trained these using standard parameters and ones in adjacent orders of magnitude, but the networks overfit quickly, even with regularization, obtaining minimum validation loss in very early epochs and producing poor results. We cannot rule out the possibility that different parameter settings are needed on this problem, but the number of parameters vastly outnumbers the number of independent data points which would point to the problem being that the network is simply far too parameterized.

We additionally show the distribution of predicted-vs-actual irradiances in figure 4.3 for Fe XX emissions. The DEM inversion approach systematically overestimates irradiance. The least-squares fit to mean AIA data-count produces a substantially better fit, but still cannot accurately model emissions. Changing the loss function for fitting the model and adding additional features produces a much tighter fit and as seen in figure 4.3 (and the 20% relative decrease in error), using a CNN even further reduces errors.

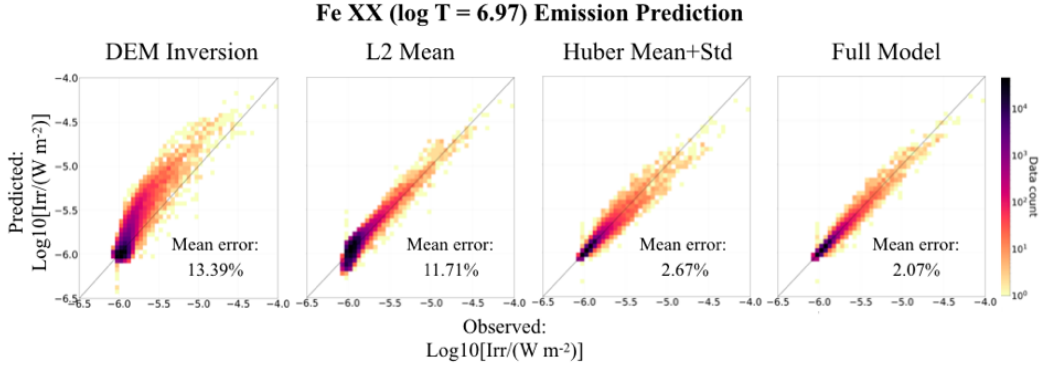


Figure 4.3: Emission prediction of FeXX line for several models. We plot a histogram of the results of the prediction on the test set for different model implementations, with a log scale colorbar. The closer the points are scattered around the line, the better the predictions are when compared to observations. We can see that the data driven models greatly improve upon the DEM inversion, and the full model which includes the CNN further increases accuracy, especially on the less frequent, higher amplitude flares.

While the goal of this study is to reproduce the 14 lines of EUV SSI, we performed an additional experiment with the same method on total irradiance to test the generality of the method. Specifically, we trained the same linear+CNN model to predict the SSI over the entire MEGS-A range as an additional output. This model performs similarly well, obtaining a 1.22% relative error on the total irradiance channel.

4.2.4 Opening the black-box: Analysis of Models

In addition to looking at the predictive abilities of the model, we can analyze the models to evaluate how they are solving the problem. We report analysis of both a channel summary statistic model and the full model. Both suggest that the learning models are, generally, solving the problem in the correct way.

One of the simplest ways to visualize a model is to examine its weights. In figure 4.4, we show visualizations of the linear weights for the model using Huber loss

on Mean+Std AIA features. As expected, the largest contributor for each emission line tends to be the closest AIA filter. This takes advantage of the direct overlap between most of the MEGS-A emission lines and at least one AIA filter wavelength response (the only exceptions being He II, Fe XV, and Mg IX). However, many of the emission lines use information from more than one AIA filter, as well as information about the variability of the irradiance to adjust their predictions. A careful comparison between each of the emission line weights in figure 4.3 with the AIA filters temperature response of figure 4.1C shows that our model is indeed using the AIA filters' temperature response as a basis for the reconstruction of EUV SSI. Fe XVIII is a nice example of this behavior: figure 4.1B shows that this line is centered on the AIA 94 filter, but figure 4.4 shows clear contributions from the 131, 193, 211, and 335 channels. All these channels can be seen to have similar values of response at the temperature of the Fe XVIII line in figure 4.1C. This means that our model is essentially working the way the DEM-based models do (i.e. performing an internal assessment of the density and temperatures of the solar plasma in order to estimate irradiance), while using wavelength proximity to enhance the recovery of the MEGS-A emission lines using the AIA image filters.

In the case of the combined linear+CNN model, as done in (Barron et al., 2013) and similar to (Zhou et al., 2016), it is possible to change the order of the averaging and linear combination operations in the later stages of the CNN and linear models to produce spatially resolved irradiance maps that can be used as a diagnostic of the operation of the model. For example, the CNN concludes by taking a 15x15 384-channel image, averaging it spatially to a 384-dimensional feature vector, and applying a linear transformation to produce the final 14-dimensional EVE predictions. This is equivalent to a model in which the linear transformation is applied per-pixel in the 15x15 384-channel image (producing a 15x15 14-channel image) followed by spatial averaging to a 14-dimensional result. To get a sense of how

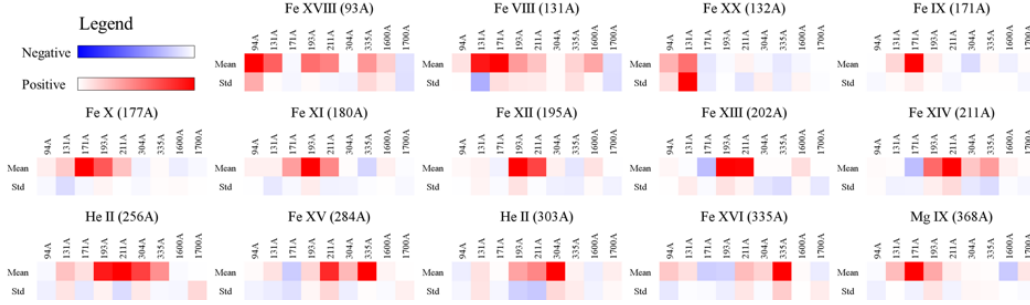


Figure 4.4: Linear weights for the Huber Mean+Std model per line. We visualize heatmaps, where the red indicates positive weights and blue negative weights with intensity proportional to weight. The first row is features that are the average (i.e., total irradiance); the second is features that are the standard deviation (i.e., variance of irradiance). Most MEGS-A lines are primarily a function of nearby AIA observations (e.g., Fe IX is overwhelmingly just a rescaling of the AIA data). Other lines make use of the standard deviation features (e.g., Fe XX is primarily driven by variance in 131 Å).

the CNN is solving the task, we can examine this pre-averaged 14-dimensional prediction.

As shown in the faint components of figure 4.2, we apply this technique to our best model by splitting the final prediction into the sum of three linear models applied to: standard deviation AIA features (linear model); average AIA features (linear model); and the average of the last convolution layer (CNN). The first computation remains a per-channel scalar, and we rearrange the latter two, yielding spatially resolved feature maps at 15x15 and 256x256 respectively. We produce a final spatially resolved map by bilinearly upsampling all components of the 15x15 images to 256x256 and summing all the feature activation maps in the last convolutional layer. When averaged, this spatial map produces the same results as the original computation graph, but spatially resolved, as can be seen in figure 4.5. These maps look remarkably like one would expect from a spectroscopic imager with bright patches that match the location of active magnetic regions, surrounded by a darker quiet Sun, and even brighter flaring regions. This is evidence that our

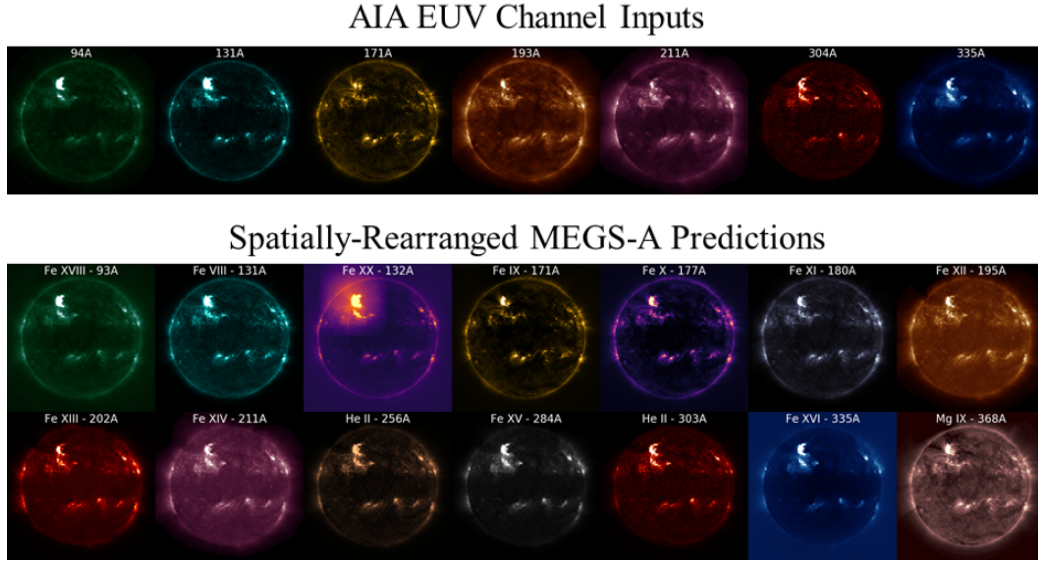


Figure 4.5: Spatially rearranged predictions. Results from the model after it has been re-arranged to produce spatial results. Global (i.e., image-sized 1×1) features are due to linear model applied to standard deviation features; low-frequency (i.e., blocky 15×15) features are due to the CNN, which has low spatial resolution; high frequency details are due to the linear model on average AIA features (i.e., 256×256). The linear model has learned a largely correct model mapping AIA to MEGS-A which is corrected, especially during flaring events, by the CNN.

combined model is paying attention to the right sources of EUV irradiance in a way that makes physical sense. This opens tantalizing opportunities to further constrain them, validate them, and exploit their physicality in future work; showing the potential that deep learning has to enhance the scientific output of solar data.

4.3 Discussion

We have implemented an approach for nowcasting MEGS-A spectral irradiance measurements from AIA images, achieving a median error of 1.6% across MEGS-A lines and outperforming a variety of alternate machine learning approaches and

a DEM-based approach. In particular, our approach produces substantially better performance on a variety of lines that are not well reconstructed by our DEM model (e.g., the flare-sensitive Fe XX line). In addition to quantitatively characterizing the performance of the method, our analysis has produced evidence that the model is solving the AIA to MEGS-A mapping in a physically sensible way. Although it is too early to claim that our approach can produce spatially resolved maps of EUV irradiance based on AIA and EVE data, the results are promising enough that they are worth pursuing. In future work we plan to integrate the data-driven and physically-based techniques, to produce models that by design solve the problem in a physically valid way.

In this work we have demonstrated that we can leverage 4 years worth of AIA and MEGS-A to create a virtual MEGS-A that will be available as long as AIA is functioning. This showcases the incredible potential that CNNs, and deep learning in general, have for heralding a new age of virtual instruments. Other important examples of this potential include the estimation of photospheric velocities based on continuum images ([Ramos et al., 2017](#)), and the assembly of super-resolution magnetograms based on magneto-hydrodynamic simulations and photospheric magnetograms ([Baso and Ramos, 2018](#)). It is important to stress that these virtual instruments will not be substitutes for their hardware counterparts, as it is clear that this work would have been impossible without the data taken by AIA and the MEGS-A/EVE instruments. However, these virtual instruments will leverage existing and historic scientific instruments to yield similar levels of scientific data products as hardware missions currently do.

4.4 Materials and Methods

Throughout we use supervised machine learning methods. A complete introduction is beyond the scope of this paper and can be found in (Hastie et al., 2009) but to provide context, we will briefly introduce a few key concepts. Supervised learning aims to perform function approximation given a set of N training examples of inputs $X_i : X_i \in \mathbf{X}$ and desired outputs $y_i : y_i \in \mathbf{Y}$. In our case, each X_i may be the $256 \times 256 \times 9$ AIA input $\in \mathbb{R}^{256 \times 256 \times 9}$ or a 19-D summary statistic $\in \mathbb{R}^{19}$; and each y_i is a MEGS-A line. Each of our models is a function parametrized by θ , or $f(X_i, \theta) : \mathbf{X} \rightarrow \mathbf{Y}$. We fit these functions by optimizing parameters to minimize the empirical risk for a given loss function (which measures how close $f(X_i, \theta)$ is to y_i), $\theta^* = \operatorname{argmin}_{\theta} \frac{1}{N} L(f(X_i, \theta), y_i)$ or, in other words, θ^* is the set of parameters that minimizes the average loss occurred on the validation set. How this is optimised depends on the form of X_i , f , and L ; but generally a standard optimisation technique such as gradient descent is used.

Some of our initial models are linear regression. For these, we model $f(X_i; \theta)$ as $f(X_i; \theta) = \sum_{j=1}^D \theta_j X_{ij} = \theta^T X_i$ where D is the dimensionality of the input (typically including a constant term to act as a bias). If we set the loss function to $L(f(X_i; \theta), y) = (f(X_i; \theta) - y)^2$, the resulting minimization problem is ordinary least-squares and can be solved analytically both with and without Tikhonov/ridge-regression regularization. For more complex loss functions (e.g., Huber), we can optimize the loss function with gradient descent: starting with an initial point, we take the gradient of the loss (plus any regularization) with respect to θ and iteratively update θ to follow the negative gradient. Our full model incorporates a convolutional neural network (CNN), a complete introduction to which can be found in (Goodfellow et al., 2016). We have seen that the linear model represents the output of $f(X_i; \theta)$ as a linear or affine transformation of the

input. Similarly, a CNN represents the output of $f(X_i; \theta)$ as a transformation of the input, this time involving multiple layers of convolution or linear filtering, interleaved with non-linearities, resulting in a complex non-linear function built of smaller simple components. The parameter vector θ is then all of the parameters of the convolutions. While the resulting empirical risk minimization problem of minimizing the loss with respect to θ is highly non-convex, the local minima of the networks tend to perform well in practice across most problem domains.

The advantage of such a model is that the desired output variables are often linear transformations of the input variables and the network can discover a transformation from inputs to a feature space which is related by a linear transformation to the outputs. Without this sort of model, it is common to need a heuristic non-linear preprocessing step (e.g., in our work for the baseline models, taking the standard deviation of the input). In the case of CNNs, this transformation is discovered via the data to minimize the final objective.

4.4.1 Models

Here, we report a number of implementation details about the models:

Channel Summary Statistics Model

As is common practice in machine learning, we append 1 as a feature before fitting all linear models to add a bias. This results in a 10-dimensional feature vector for AIA mean models and a 19-dimensional one for mean and standard deviation models.

L2 Regression: This can be solved geometrically and has no hyperparameters.

Huber Regression: We fit a model using Scikit Learn's SGD Regressor per EVE channel, using a huber loss and L2 regularization. We standardize the variables (subtracting mean and dividing by standard deviation from the train set), and use

the validation set to pick the regularization strength and huber delta parameter.

Gradient Boosting Regression Trees: We fit a model with Scikit Learn’s Gradient Boosting Regressor, minimizing huber loss to set weights and fitting 100 trees. We use the validation set to pick the huber delta parameter, tree depth, and minimum number of samples per leaf. We found the default optimization parameters (e.g., convergence criteria) to work poorly with EVE’s small magnitude and multiplied the target variables by a factor of 10^5 .

Convolutional Neural Net: We use a modified Alexnet ([Krizhevsky, 2014](#)) containing 3 convolutional layers, followed by average pooling and an affine (i.e., linear + bias) transformation. Let $C(k,n,s,p)$ denote a convolutional layer with n filters of size $k \times k$, applied with stride s and padding p ; BN denote batchnorm; R denote a ReLU; and $MP(k,s)$ denote max-pooling over a $k \times k$ grid with stride s . The convolutional block is then: $C(11,64,4,2)$, BN, R, $MP(3,2)$, $C(5,192,1,2)$, BN, R, $MP(3,2)$, $C(3,384,1,1)$, BN, R. This is then followed by average pooling, dropout ($p=0.5$) and a fully connected layer with 14 outputs. We use dropout due to the small amount of data.

The network is trained following standard practices from a random initialization with stochastic gradient descent. We use pytorch defaults for initialization except for convolutional layers, which are initialized with the Kaiming normal method. After initialization, we train the network with stochastic gradient descent with Nesterov momentum ([Sutskever et al., 2013b](#)) and weight decay of 10^{-9} for 24 epochs. We start with a learning rate of 0.1 and multiply it by 0.1 every 8 epochs. We applied gradient clipping (to have maximum norm 0.5), which we found anecdotally to improve performance.

Properly scaling the inputs is important. We divide each channel of an input AIA image with an average computed over the training set (and divide by this same

value at test time). We treat the residual EVE similarly, dividing by the average residual over the training set, but we also multiply it by 10^2 to make the order of magnitude of the target values match our network’s initial outputs.

Acknowledgments

CHIANTI is a collaborative project involving George Mason University, the University of Michigan (United States), and the University of Cambridge (United Kingdom). The AIA, EVE, and HMI instruments are instruments onboard the SDO, a mission for NASA's Living With a Star Program. The authors thank James P. Mason for help with the EVE level 2 data. The authors would like to acknowledge the developers of PyTorch and scikit-learn.

Funding

This project was conducted during the 2018 NASA Frontier Development Lab (FDL) program, a public-private partnership between NASA, the SETI Institute and commercial partners. The authors wish to thank, in particular NASA, IBM and Lockheed Martin for supporting this project. MCMC and MJ acknowledge support from NASA's SDO/AIA (NNG04EA00C) contract to the LMSAL.

Author Contributions

M.C.M. Cheung conceived the project.

M. Jin and D. Fouhey worked on data preprocessing.

P. J. Wright worked on the neural DEM-based model, with contributions from M. C. M. Cheung, R. Thomas, R. Galvez, and M. Jin.

D. Fouhey and A. Szenicer worked on the data-driven models. A. Munoz-Jaramillo worked on the interpretation of the model weights and irradiance maps.

A. Munoz-Jaramillo worked on data visualization with contributions from M. C. M. Cheung, D. Fouhey and A. Szenicer.

D. Fouhey, A. Szenicer, and A. Munoz-Jaramillo took the lead on writing the manuscript, which was reviewed, corrected, and approved by all authors.

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The full dataset of AIA images and EVE spectra used to obtain the results in the paper can be found at <https://purl.stanford.edu/nk828sc2920>. All codes used in the project can be found at <https://github.com/AlexSzen/sw-irradiance>

Supplementary material

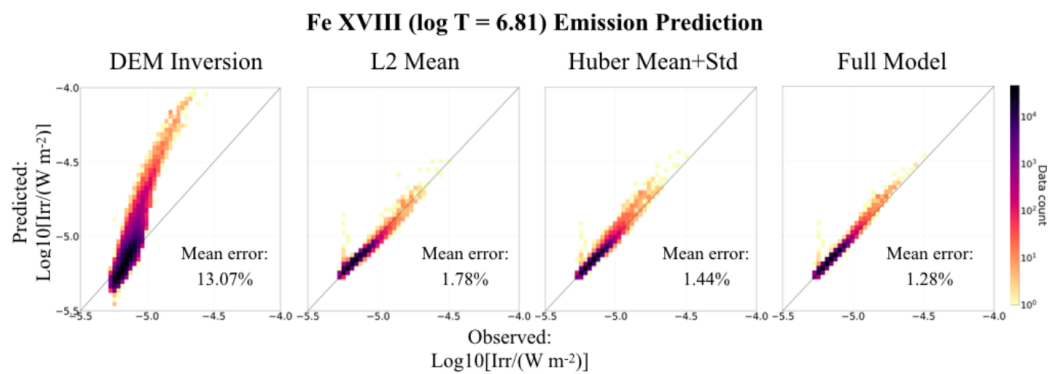


Figure 4.6: Emission prediction of Fe XVIII line for several models. We plot a histogram of the results of the prediction on the test set for different model implementations, with a log scale colorbar. The closer the points are scattered around the line, the better the predictions are when compared to observations.

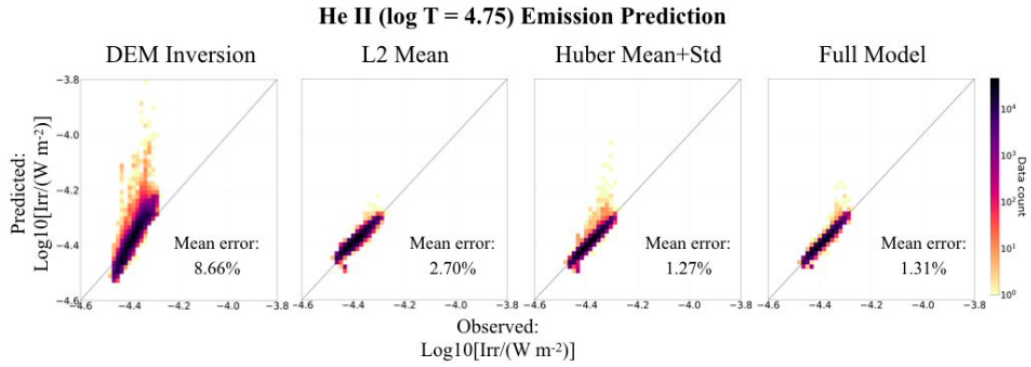


Figure 4.7: Emission prediction of He II line for several models. We plot a histogram of the results of the prediction on the test set for different model implementations, with a log scale colorbar. The closer the points are scattered around the line, the better the predictions are when compared to observations.

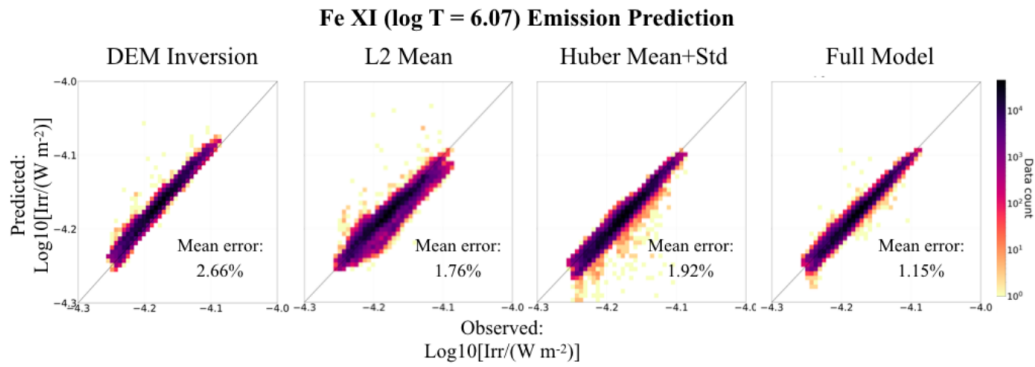


Figure 4.8: Emission prediction of Fe XI line for several models. We plot a histogram of the results of the prediction on the test set for different model implementations, with a log scale colorbar. The closer the points are scattered around the line, the better the predictions are when compared to observations.

5 | Discussion and conclusions

The Geosystem is composed of many interacting subsystems that each hinge on one another, and are important to understand in their own right. Only through bettering our knowledge of all these subsystems will we arrive at a global approach to sustainability, i.e. understanding the impact of our actions on a global and holistic level, which will allow us to thrive as a species in the long term. Understanding the Geosystem requires developing new theories, new sensors (physical or virtual!), creating new data sources, and designing new modelling methods. Enormous progress has been made in recent years on these fronts, with computers following Moore's law, sensors and instruments being developed and deployed at a rapid pace, and clever advances in numerical methods, whether physics or data-driven. In this PhD, I have chipped away at the mountain of work that lies ahead of the new generation of scientists, and tried to contribute to some projects in geophysics that we found particularly interesting and which are well suited to a computational and machine learning approach. The scientific triptych presented in this PhD is a good illustration of the diversity and richness of the Geosystem, the disciplines it is composed of, and the methods needed to improve our knowledge of it.

We have started our journey with a problem of ecology and conservation: monitoring elephants to mitigate poaching and human-elephant conflicts. While the consequences of an ecological disaster such as the extinction of elephants are far reaching, the problem of monitoring itself is defined on a spatial scale of a few hundred meters. We have tackled it with a composite approach of geophysics and deep learning, and have delivered promising results that will hopefully lead to more in-depth studies of this method. In particular, we have constructed a deep learning based classifier trained on spectrograms produced from seismic data, to

either identify elephants from other species, or recognise important behaviour for welfare such as running or rumbling. Our classifier achieved between 80–90% balanced accuracy in detecting elephants for distances up to 100 meters, as well as over 90% balanced accuracy in recognising elephant rumbles from locomotion, and elephants running from walking. What is more, our methods were developed using a unique dataset acquired on fieldwork in Kenya, which consists of over 8000 hours of seismograms and 250k camera trap pictures, and represents a diverse ecosystem with varied animal species (elephants, giraffe, zebras, humans, etc) and terrains (wet and muddy, rocky, sandy, ...). We release a curated version of this dataset tailored for machine learning applications, dubbed SeisSavanna, to promote collaboration between different scientific fields and enable creative solutions that will benefit elephant conservation. These contributions provide a prototype and stepping stone for further developments to our system, which we hope will become a realistic and usable tool, and will be deployed to monitor wildlife. To achieve this, we propose several avenues for future research, from both the software and hardware side. First of all, we believe that there can be significant benefits to further processing and refinement of the dataset, in order to increase the amount of labelled data, and the diversity of environmental conditions in the data. Another important direction to explore is that of data augmentation. While we have investigated the use of two augmentation methods that were designed for earthquake signals, we found no consistent increase in accuracy from using them, which means that we need to design specialized approaches for wildlife seismic signals, e.g. by exploiting generative adversarial networks (GANs) ([Goodfellow et al., 2014](#)). A third interesting path to explore from a methodological perspective is that of specialised deep learning architectures. While in the present work we have used several different CNN designs working on both 1D (seismogram) and 2D (spectrogram) signals, we believe that investigating specialised architectures

for the seismogram format could be beneficial. Indeed, seismogram chunks are relatively long vectors (2048 samples in our case), with both short term (individual footsteps) and long term (gait characteristics) structure. Exploring dedicated architectures that can exploit this structure, such as dilated convolutions ([Chen et al., 2017](#)) that allow an exponential increase in the CNN’s field of view. [Moseley et al. \(2018\)](#) used such an approach to simulate the response of acoustic seismic waves in horizontally layered media with a WaveNet ([Oord et al., 2016](#)), or attention mechanisms ([Vaswani et al., 2017](#)) that can focus on specific parts of a signal, could be a very promising approach. Finally, from a hardware perspective, we envision longer deployments with dense arrays of lightweight seismic sensors, and full azimuthal coverage with camera traps. This could allow to further investigate the generalisation capabilities of our method to new environments, as well as getting rid of label uncertainty and corruption.

Our next stop lives on an entirely different spatial scale: the whole planet. Global seismic tomography is a crucial tool that allows us to create structural models of the interior of the Earth, which can in turn answer questions about the creation, history, and evolution of our planet. However, the most powerful and physically accurate deterministic methods for seismic tomography are still computationally prohibitive if we want to run them at high resolution, which is where things get really interesting. Indeed, short wavelengths details are important for understanding mantle dynamics and related plate tectonic processes(e.g. the origin of hotspots and the driving mechanisms of plate motion and earthquakes), because clues to the answers can be found in Earth’s interior (subducted oceanic crust, geochemical blobs with primordial noble gas isotopes, partial melt, ...) with high frequencies. From a more applied perspective, high resolution images of the subsurface are important to locate earthquakes and nuclear explosions, or to understand earthquake hazard in populated regions.

We design a new framework for FWI based on discrete adjoints to compute sensitivity kernels. Together with the solver AxiSEM3D, this allows us to speed up the computation of the kernels by exploiting the azimuthal smoothness of seismic wavefields. Indeed, in AxiSEM3D we adopt a flexible Fourier expansion that allows us to adapt to local complexity, and thus have $O(\omega^3)$ frequency scaling, rather than $O(\omega^4)$ for traditional software. While we lose this computational advantage in the adjoint simulation when using the continuous adjoint approach, switching to discrete adjoints allows us to retain the $O(\omega^3)$ scaling. This comes at the expense of allowing more complex Earth models, as parametrised by the dense mesh in traditional software. Indeed, in the discrete adjoint approach, the trick is to work in the discretisation imposed by the forward simulation, and only update those discrete model parameters, which in turn lifts the requirement of having physically accurate adjoint wavefields. Naturally, we need to be cautious when adopting this discretisation. Indeed, because we do not know a priori what is the structural complexity of the Earth, we cannot adopt an overly simplistic parametrisation, since we would essentially underfit the data. One promising approach which we plan to explore in future work is to adopt a model parametrisation defined by our data coverage, as is done in tomographic inversions ([Hosseini et al., 2020](#)): zones with dense ray coverages (e.g. USA) will have a fine parametrisation because we can resolve Earth’s structure in a lot of detail, whereas zones with sparse coverage (e.g. oceans) will have much larger mesh cells, because the low amount of seismic rays does not allow us to resolve small structures. We will therefore be fixing our mesh resolution (and thus our Fourier expansion and computational complexity) prior to the inversion, which will allow us to take advantage of the discrete adjoint kernels and cut computational costs. In subsequent work we will be designing such a mesh adapted to our framework, and demonstrating an example of global seismic tomography using this novel frame-

work. Designing meshes based on data coverage is a powerful approach which can speed up FWI in the mid-term future (at least several decades), but as we are increasingly instrumenting the planet, including oceans with ocean bottom seismometers (OBS) (Bialas and Flueh, 1999) and MERMAIDs (Sukhovich et al., 2015), this approach will eventually become less advantageous given that our data coverage and thus mesh resolution will get finer and finer. Looking into the long term, we therefore plan to explore methods which would allow us to determine which sections of the mesh need an increase in their Fourier expansion, and update them along with the model parameters during the inversion process. Potential approaches could be inspired by techniques such as adaptive mesh refinement (Berger and Oliger, 1984), or transdimensional inversion (Sambridge et al., 2013). We also present early stage research in which we aim to accelerate forward and inverse problems by using deep learning. For the deep learning inversion, we generate sensitivity kernels from source and receiver position, source characteristics, and a waveform measurement; by using an expanding CNN, trained on a database of kernels produced using MC Kernel. We qualitatively show that the network is able to learn the mapping, and we outline next steps for this project. To accelerate forward simulations, we propose to map low frequency seismograms to a higher frequency. We showcase some promising results by training an encoder-decoder CNN on 2D acoustic synthetics, where we show the ability of our network to generalise to unseen velocity models while being trained on just a handful of simulations.

Finally, the last chapter in the PhD takes another step up spatially, as we are now interested in monitoring the Sun and understanding its impact on the thermosphere and ionosphere of the Earth. The extreme ultraviolet radiation from the Sun is the major driver of these atmospheric layers, and enhanced activity can cause various deleterious effects such as increased drag on low orbit satellite. High EUV can

also cause other noxious effects such as radio blackouts or GNSS scintillation. Our ability to monitor EUV SSI was compromised in 2014 due to an electrical malfunction of the Multiple EUV Grating Spectrograph A (MEGS-A) channel onboard NASA’s Solar Dynamics Observatory (SDO).

We develop a deep learning model for nowcasting EVE MEGS-A spectra from AIA images, which achieves 1.6% median error. We show that our model performs better than a variety of baselines which include a physics approach based on the DEM model, as well as different classical machine learning methods such as linear regression and random forests. Our best performing model is a combination of a linear module which explains the bulk of the MEGS-A spectrum, and a CNN module trained on residuals from the linear module. Importantly, we show that both the linear and CNN modules perform the mapping in a physically sensible way. Indeed, inspection of the linear weights shows that MEGS-A lines use as features spectrally close AIA observations. What is more, upsampling the data transformation in the penultimate layer of the CNN, we can see that the network preserves the shape of spatial structures of the Sun. It is important to highlight that the physics based DEM model does not perform well because of our incomplete knowledge of the physics at play. Indeed, several assumptions are made, such as coronal equilibrium conditions, optically thin plasma, and uniform elemental abundances in the corona. However, it turns out we can bypass these approximations and obtain very good results with a deep learning model, which essentially learns the physics. Could we perhaps design models which we could interrogate and use to guide us to a better knowledge of the physics? To this end, we could explore deep learning models that are explicitly constrained to respect physical laws, such as physically inspired neural networks (PINNs) (Raissi et al., 2019), which allow to find solutions to PDEs or discover the PDEs themselves in a data-driven way. We could apply these methods to our problem setting in order to, for

instance, get a better understanding of elemental abundances in the solar corona. It is also important to emphasise that this result is not meant to show that machine learning by itself voids the need for proper scientific analysis. Rather it highlights how scientific expertise, when accelerated by high-performance computing and machine learning, can help us more effectively exploit existing data sets. What is more, our results presage the advent of virtual instruments to supplement existing hardware ones, by exploiting large datasets of co-observing sensors. Virtual instruments are naturally not limited to observations of the Sun, and we can envisage applying similar procedures to terrestrial instruments and Earth observing satellites. For instance, there are currently several different Earth observation satellites in orbit, such as Landsat 8 ([USGS, 2016](#)), Aqua and Terra ([Friedl et al., 2002](#)), and Sentinel 2 ([Drusch et al., 2012](#)), among others. These satellites carry various instruments that observe the planet in different spectral bands, with different spatial resolutions, and different revisit times. We can contemplate constructing mappings between these different observations to virtually augment the suite of instruments of each satellite. Another example is directly linked to Chapter 2 and elephant monitoring. As discussed earlier, a promising avenue of further research for our monitoring system is to use denser deployments of small, low cost instruments. These geophones will have different characteristics from the high-end seismometers we have deployed (e.g. different frequency range), but perhaps by deploying a few seismometers to co-observe with the cheaper geophones, we could map the lower quality data to the higher quality one, essentially creating super-resolution seismograms.

Appendix A: A Machine Learning Dataset Prepared From the NASA Solar Dynamics Observatory Mission

Published in Astrophysical Journal Supplement Series, DOI: 10.3847/1538-4365/ab1005

Richard Galvez^{1†}, David F. Fouhey^{2†}, Meng Jin^{3,4}, Alexandre Szenicer⁵, Andres Munoz-Jaramillo⁶, Mark C. M. Cheung^{3,7}, Paul J. Wright⁸, Monica G. Bobra⁷, Yang Liu⁷, James Mason⁹ and Rajat Thomas¹⁰.

¹ Center for Data Science, New York University; ² Electrical Engineering and Computer Science Department, University of Michigan ; ³ Lockheed Martin Solar and Astrophysics Laboratory; ⁴ SETI Institute; ⁵ Department of Earth Sciences, University of Oxford; ⁶ Southwest Research Institute; ⁷ Hansen Experimental Physics Laboratory, Stanford University; ⁸ SUPA School of Physics and Astronomy, University of Glasgow; ⁹ NASA Goddard Space Flight Center; ¹⁰ Department of Psychiatry, Academic Medical Centre, University of Amsterdam

* Corresponding author: Richard Galvez, richard.galvez@nyu.edu

† These authors contributed equally to this work.

Abstract

In this paper we present a curated dataset from the NASA Solar Dynamics Observatory (SDO) mission in a format suitable for machine learning research. Beginning from level 1 scientific products we have processed various instrumental corrections, downsampled to manageable spatial and temporal resolutions, and synchronized observations spatially and temporally. We illustrate the use of this dataset with two example applications: forecasting future EVE irradiance from present EVE irradiance and translating HMI observations into AIA observations. For each application we provide metrics and baselines for future model comparison. We anticipate this curated dataset will facilitate machine learning research in heliophysics and the physical sciences generally, increasing the scientific return of the SDO mission. This work is a direct result of the 2018 NASA Frontier Development Laboratory Program. Please see the appendix for access to the dataset, totaling 6.5TBs.

1. Introduction

Launched in 2010, NASA's Solar Dynamics Observatory (SDO; [Pesnell et al., 2012](#)) has been continuously monitoring the Sun's activity and delivering valuable scientific data for heliophysics researchers with the use of three instruments:

- The Atmospheric Imaging Assembly (AIA [Lemen et al., 2012](#)) which captures 4096×4096 resolution images (with 0.6 arcsec pixel size) of the full Sun in two ultraviolet (UV; centered at 1600 & 1700 Å) wavelength bands, seven Extreme Ultraviolet (EUV) wavelength bands (centered at 94, 131, 171, 193, 211, 304 and 335 Å) and one visible wavelength (centered at 4500 Å).
- The Helioseismic and Magnetic Imager (HMI [Schou et al., 2012](#)) captures visible wavelength filtergrams of the full Sun at 4096×4096 resolution (pixel size of 0.5 arcsec), which are then processed into a number of products, including photospheric dopplergrams, line-of-sight magnetograms and vector magnetograms ([Hoeksema et al., 2014](#)).
- The EUV Variability Experiment (EVE; [Woods et al., 2012](#)) monitors the solar EUV spectral irradiance from 1 to 1050 Å. This is done by utilizing multiple EUV Grating Spectrographs (MEGS) which disperse EUV light from the full disk of the Sun and its corona onto a 1024 x 2048 CCD.

Calibrated level 1 scientific data from the AIA and HMI instruments are accessible from the Joint Science Operations Center¹ (JSOC) at Stanford University, Lockheed Martin Solar & Astrophysics Laboratory, and affiliate science data centers; while science data from the EVE instrument are accessible from the EVE

¹<http://jsoc.stanford.edu>

Science Operations Center² at the Laboratory for Atmospheric and Space Physics (LASP) at the University of Colorado, Boulder.

The SDO mission has been scientifically prolific. In the eight years after launch, over 3000 refereed scientific publications³ have made use of SDO data. This success can be attributed to the reliability of the spacecraft and its instruments, the consistency and quality of the observations, the mission's open data policy, and the ease of online data access from the affiliate science data centers. The large volume of structured, calibrated scientific data (over 12 Petabytes and counting) is poised for exploratory analysis from machine learning methods as well as more traditional approaches. In early pioneering works, supervised learning techniques have been applied to the prediction of solar flares using HMI vector magnetograms (e.g., [Bobra and Couvidat, 2015](#)), as well as HMI and AIA imagery in [Jonas et al. \(2018\)](#). Deep learning applications have begun to emerge from the heliophysics community as well, with exemplary cases illustrated in [Colak and Qahwaji \(2013\)](#) and [Huang et al. \(2018\)](#), and [Wright et al. \(2019\)](#) presenting a more recent treatment using the dataset presented here.

While level 1 data are easily accessible, pre-processing these data for scientific analysis often requires specialized heliophysics knowledge. The necessity for such pre-processing may act as an unnecessary hurdle for non-heliophysics machine learning researchers whom may wish to experiment with datasets from the physical sciences, but are unaware of domain-specific nuances (e.g., that images must be spatially and temporally adjusted).

The first contribution of this paper is a curated SDO dataset that is mission-ready for machine learning applications. Our aim is to supply this standardized dataset for heliophysicists who wish to use machine learning in their own research, as well

²<http://lasp.colorado.edu/home/eve/data>

³<https://sdo.gsfc.nasa.gov/mission/publications.php>

as machine learning researchers who wish to develop models specialized for the physical sciences. In Section 5, we examine current available data products, the pitfalls for their direct use in machine learning tasks, as well as what corrections and adjustments they warrant. These corrections are incorporated into our data preparation procedures that are discussed in Section 5.

The second contribution of this paper are protocols, metrics, and baseline models. We introduce evaluation protocols and metrics in Section 5, and baseline models in Section 5 where we tackle the tasks of predicting irradiance using present and future EVE data, as well as translating 3 HMI channels into 9 AIA channels. We believe these models contain generic enough components with providing useful benchmarks, and highlighting the most dangerous pitfalls, for most subsequent SDO machine learning applications.

By providing these standardized data products along with accompanying protocols, metrics, and baselines, our aim is to remove unnecessary hurdles for future machine learning research in heliophysics, and the physical sciences more broadly.

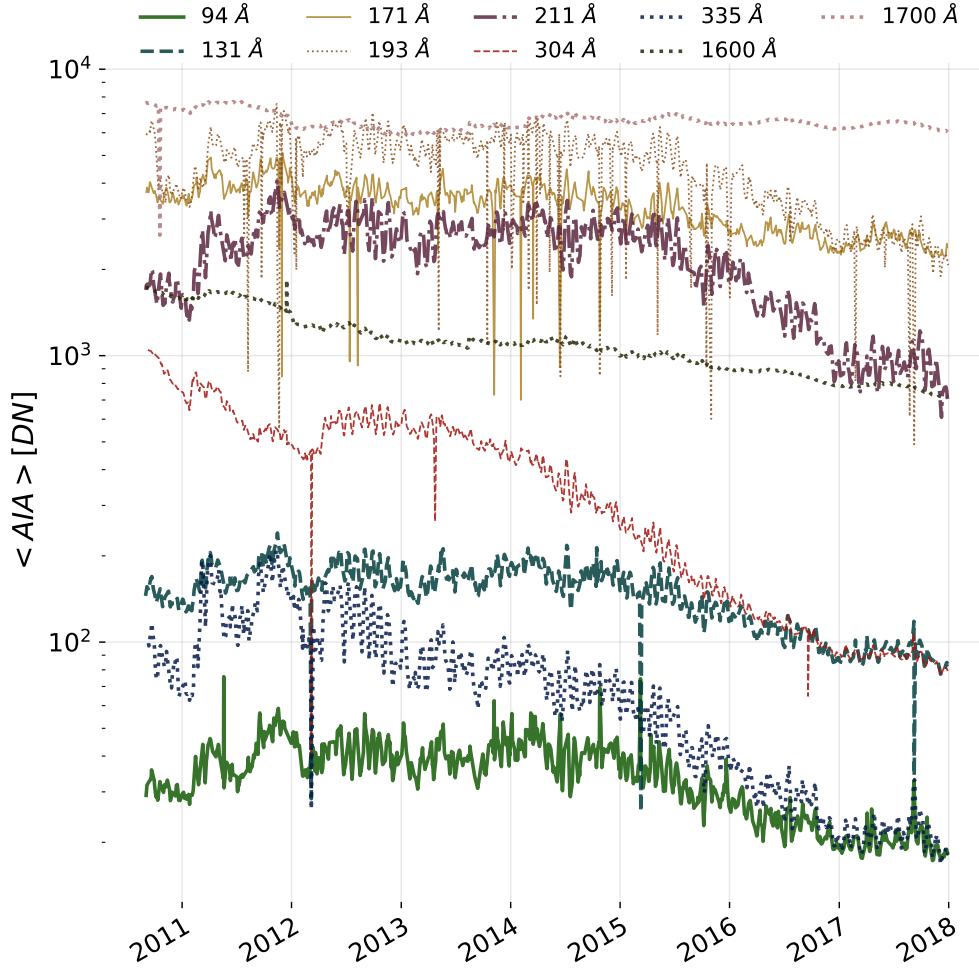


Figure A.1: The average AIA wavelength band data counts for level 1 data products over time, presented here randomly downsampled to 1k observations. The secular downward trend is caused by instrument degradation over time, while the spurious per-channel drops are caused by the instrument's automatic exposure control mechanism.

2. Examination of Raw Data Products

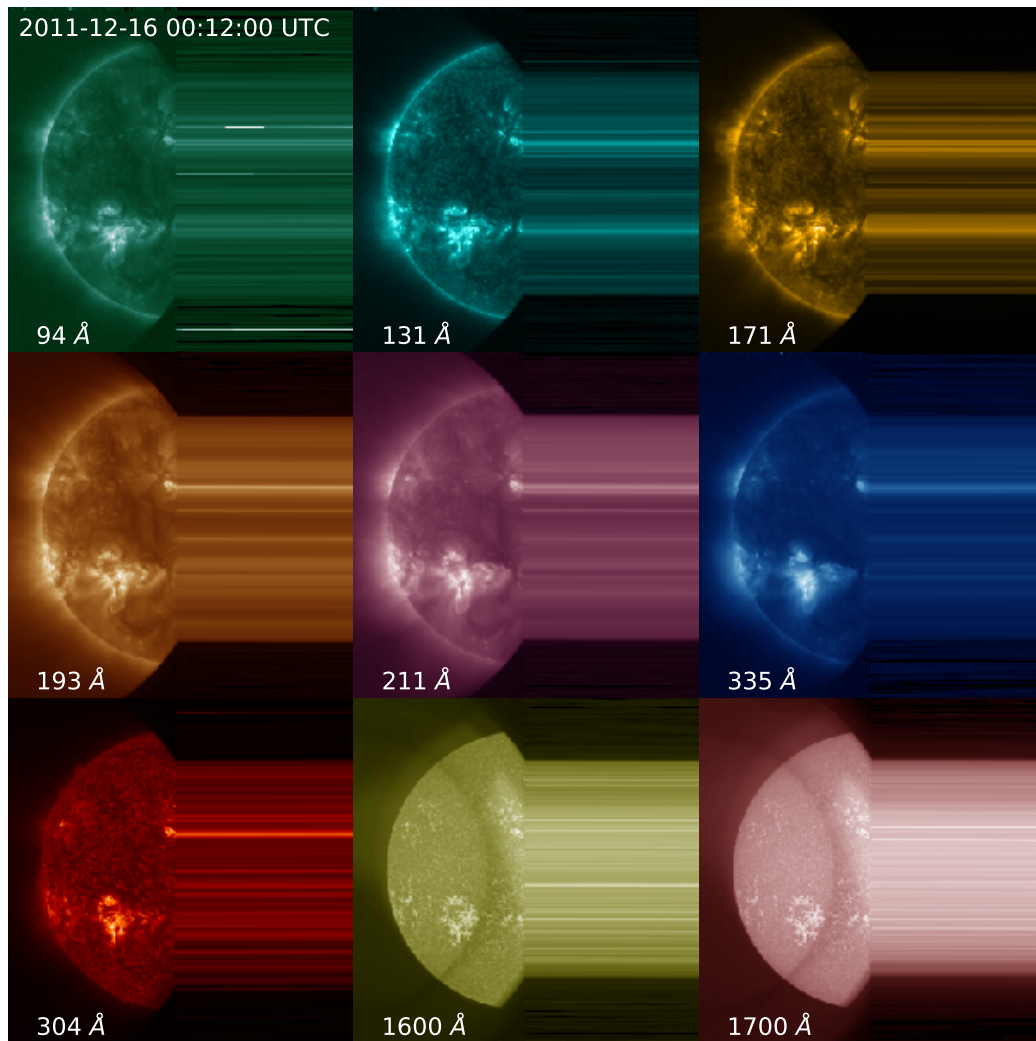


Figure A.2: Example of a corrupt observation from the AIA instrument. Utilizing such observations during the training phase of a machine learning model may compromise its predictive capability.

We first examine existing raw data products available from SDO for each of the three instruments (level 1 science data products for AIA; `hmi.B_720s` for HMI; and the EVE version 5 IDL saveset).

While heliophysics researchers are likely aware of corrections that must be applied to this data, and the fact that AIA measurements have heterogeneous exposure times, it is unrealistic to expect the same from researchers in other fields (e.g. the data set of [Kucuk et al., 2017](#), were compiled from quicklook JPEG2000 images that have compressed dynamic range and do not account for instrumental degradation). We therefore process these corrections by identifying and removing corrupt observations (e.g., images taken during instrument anomalies), adjust detected intensities for heterogeneous exposure times, and fix instrument artifacts that introduce spurious trends.

If such corrupt observations or various sources of heterogeneity are not removed, any subsequent machine learning model will likely learn to emulate these incorrect observations as well as any spurious trends, and will not be able to isolate the physical dynamics. Exposing such corrupt data during model training may also compromise predictive quality; or worse, the model may even learn to emulate non-physical aberrations and instrumental noise. See [Figure A.2](#) for an example of one such unwanted AIA observation.

To identify each instrument’s possible issues, we visualize each instrument’s data by taking the average channel values (i.e., AIA wavelength band data counts, HMI vector field components, and EVE irradiance values) and plot them over time. We then identify the underlying causes of nonphysical aberrations and what necessary corrections are needed to standardize the data. Below we report our analysis for AIA and outline where HMI required similar adjustments; EVE level-3 data products already address all instrumental issues so we only adjusted for time synchronicity. We describe these corrections in [Section 5](#).

The average channel values for the AIA level 1 data products, as plotted in [Figure A.1](#), shows the data heterogeneity as well as the presence of corrupt obser-

uations. These are visible in this figure as isolated downward spikes, while the secular downward trend is indicative of degradation over the lifetime of the instrument.

Corrupt observations arise due to a variety of reasons such as data reported during calibration maneuvers, eclipse periods, or the occasional instrument anomaly. Such data, flagged by a non-zero value of the `QUALITY` keyword for both the AIA and HMI instruments, are not intended for scientific analysis and are removed from our dataset. One of the main sources of heterogeneity in AIA data responds to its instrument design: AIA instrumentation is not designed to directly measure irradiance, but rather data numbers (DNs) as tabulated by the activation of the CCD instrument. While it is intended that DN values are proportional to the flux of photons at a specific wavelength (Boerner et al., 2012), the factor of proportionality is not constant in time. For instance, the camera exposure time t_{exp} is not constant due to instrumental automatic exposure control (AEC); e.g. in times of flares when certain regions on the Sun become especially bright, AEC reduces the nominal exposure time from a few seconds to tens of milliseconds. Due to these factors, when the AEC is activated, the mean registered DNs are drastically reduced; an effect easily compensated for by adjusting for the exposure time.

The visible downward trend in Figure A.1 is caused by the gradual in-orbit degradation of the AIA instrument. This degradation is purely due to CCD corrosion over time. Because AIA is calibrated against EVE, which is itself bootstrapped to a program of regular EUV spectral irradiances as measured from sounding rocket launches (Boerner et al., 2014), the time-dependent profile of the AIA instrument is well understood independently of the solar-cycle. This instrumental understanding allows us to correct for AIA instrument degradation by simply applying the `aia_get_response` routine in the SolarSoft software package (Freeland and Handy, 1998).

Lastly, there is a more subtle non-monotonic heterogeneity caused by SDO’s orbit around the Sun. SDO is in a geosynchronous orbit around the Earth, which itself is in a slightly elliptical orbit around the solar system barycenter. The elliptical orbit causes the size of the Sun (in DNs registered on the CCD) to vary by about 10% over the course of a year. This is not an intrinsic feature of solar evolution. We compensate for this effect by resizing AIA and HMI images such that the size of the solar disk is fixed to some size R_s . In particular, we can scale the AIA and HMI images by a factor of R_s/R_{obs} , where R_{obs} can be obtained from `RSUN_OBS` keyword in the level 1 FITS header.

3. Processed Data Preparation

We now describe how our processed dataset is produced in detail. First, we describe how the needed corrections outlined in Section 5 are applied to each instrument, and how temporal synchronization is computed. We first begin by removing the non-zero `QUALITY` observations from both AIA and HMI. We then spatially downsample to produce a more manageable dataset, while being careful to emulate what a lower resolution instruments would observe.

3.1. AIA

We begin with the 4096×4096 pixel level 1 ([Lemen et al., 2012](#), dark subtracted, flat fielded and despiked) data products and process them as described below:

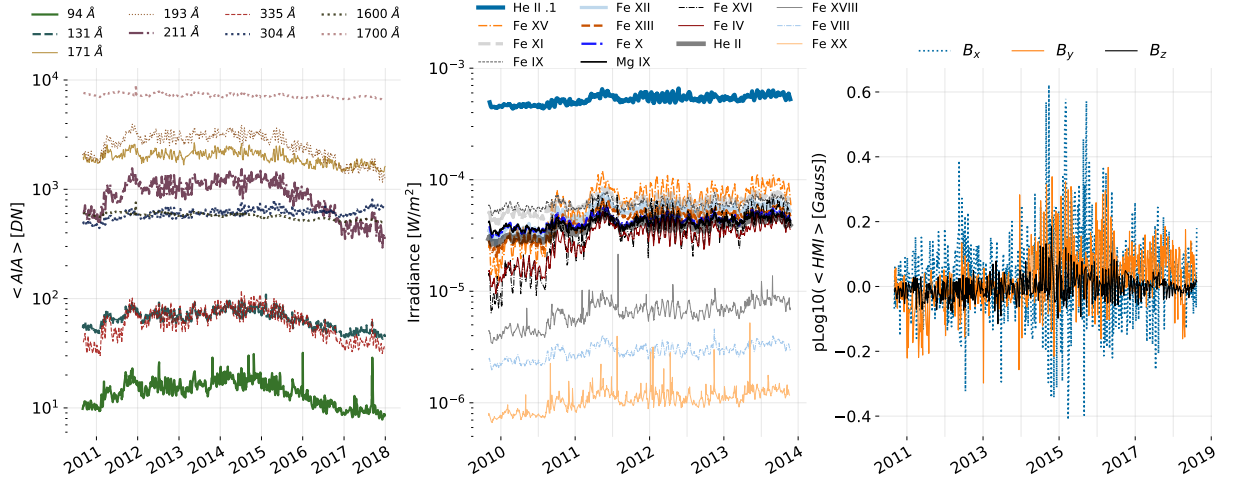


Figure A.3: Mean intensity variation after correcting for exposure time, degradation, and orbital variation for AIA (Left Panel) and EVE (Middle Panel). Middle Panel: We display only the 14 emission lines covered by the MEGS-A instrument, for illustration purposes, out of the 37 in our data product (see Table A.1). Right Panel: The signed pseudo logarithm of the mean field values for B_x , B_y , and B_z from HMI after correcting for orbital variation.

1. The raw images are rotated and resized onto a common grid (still 4096×4096 pixels) such that the pixel size is 0.600 arcsec, and $\hat{x}$ and $\hat{y}$ (the first and second image dimensions) are aligned with the solar west and north directions, respectively.
2. Images are re-binned by averaging neighboring 4×4 pixel blocks such that the resultant image has size 1024×1024 pixels (with a final pixel size of 2.400 arcsec). Resultant images are processed at a 2 min cadence producing the so-called Synoptic series⁴.
3. The AIA images are then normalized by exposure time and corrected for instrument degradation, while corrections for elliptical orbital variation are applied with a fixed disk size of R_s of 976 arcsec.

⁴<http://jsoc.stanford.edu/data/aia/synoptic/>

4. Finally, the images are downsampled again by *summing* in local blocks, which emulates the expected observation of a lower resolution instrument. The final interpolated images have 512×512 pixels with pixel size of ~ 4.8 arcsec.

3.2. HMI

We start with the original HMI JSOC data series `him.B_720s`, which provides the magnetic vector field strength, inclination angle, and azimuth (Hoeksema et al., 2014). We process this to calculate full-disk vector field observations in B_x , B_y , and B_z components with 12 minute cadence. The $+x$ direction points to the solar west, $+y$ to the north, and $+z$ out of the image plane (i.e., line-of-sight). Additionally, as with the AIA instrument, although the original image size is 4096×4096 , the pixel resolution is different. We therefore further co-aligned HMI and AIA data so that they have the same spatial sampling. The major processing steps for the HMI observations are as follows:

1. We begin by converting the original HMI JSOC data series `hmi.B_720s` vector field data with the disambiguation solution of `disambig.fits` to B_x , B_y , and B_z components, spatially co-aligning with AIA observations using the FITS header information.
2. The HMI images were also corrected for orbital variation with a fixed disk size R_s of 976 arcseconds throughout.
3. Finally, we downsampled the data by *averaging* in local blocks, which emulates the expected observation at the target lower resolution. The final interpolated images have 512×512 pixels with pixel size of ~ 4.8 arcsec.

Table A.1: EVE Emission lines, their wavelength, and temperature of the emission plasma. The MEGS-A emission lines used in Section 5 are highlighted in purple.

Emission Line	Wavelength	Temperature
Fe18	93.9 Å	6.46×10^6 K
Ne7	127.7 Å	5.01×10^5 K
Fe8	131.2 Å	3.71×10^5 K
Fe20	132.8 Å	9.33×10^6 K
Ne5	148.7 Å	3.16×10^5 K
Fe9	171.0 Å	6.46×10^5 K
Fe10	177.2 Å	9.77×10^5 K
Fe11	180.4 Å	1.17×10^6 K
Fe12	195.1 Å	1.35×10^6 K
Fe13	202.0 Å	1.55×10^6 K
Fe14	211.3 Å	1.86×10^6 K
He2	256.3 Å	5.62×10^4 K
Fe15	284.2 Å	5.01×10^4 K
He2	303.8 Å	1.99×10^6 K
Fe16	335.4 Å	2.69×10^6 K
Fe16	360.8 Å	2.69×10^6 K
Mg9	368.1 Å	9.77×10^5 K
Mg9	443.7 Å	1.00×10^6 K
Ne7	465.2 Å	3.98×10^5 K
Si12	499.4 Å	1.99×10^6 K
O3	525.8 Å	7.94×10^4 K
O4	554.4 Å	1.99×10^5 K
He1	584.0 Å	1.99×10^4 K
O3	599.6 Å	7.94×10^4 K
Mg10	624.9 Å	1.26×10^6 K
O5	629.7 Å	2.51×10^5 K
O2	718.5 Å	6.31×10^4 K
N4	765.1 Å	1.58×10^5 K
Ne8	770.4 Å	6.31×10^5 K
O4	790.2 Å	1.99×10^5 K
H1	972.5 Å	5.01×10^4 K
C3	977.0 Å	5.01×10^4 K
H1	1025.7 Å	5.01×10^4 K
O6	1031.9 Å	2.51×10^5 K

3.3. EVE

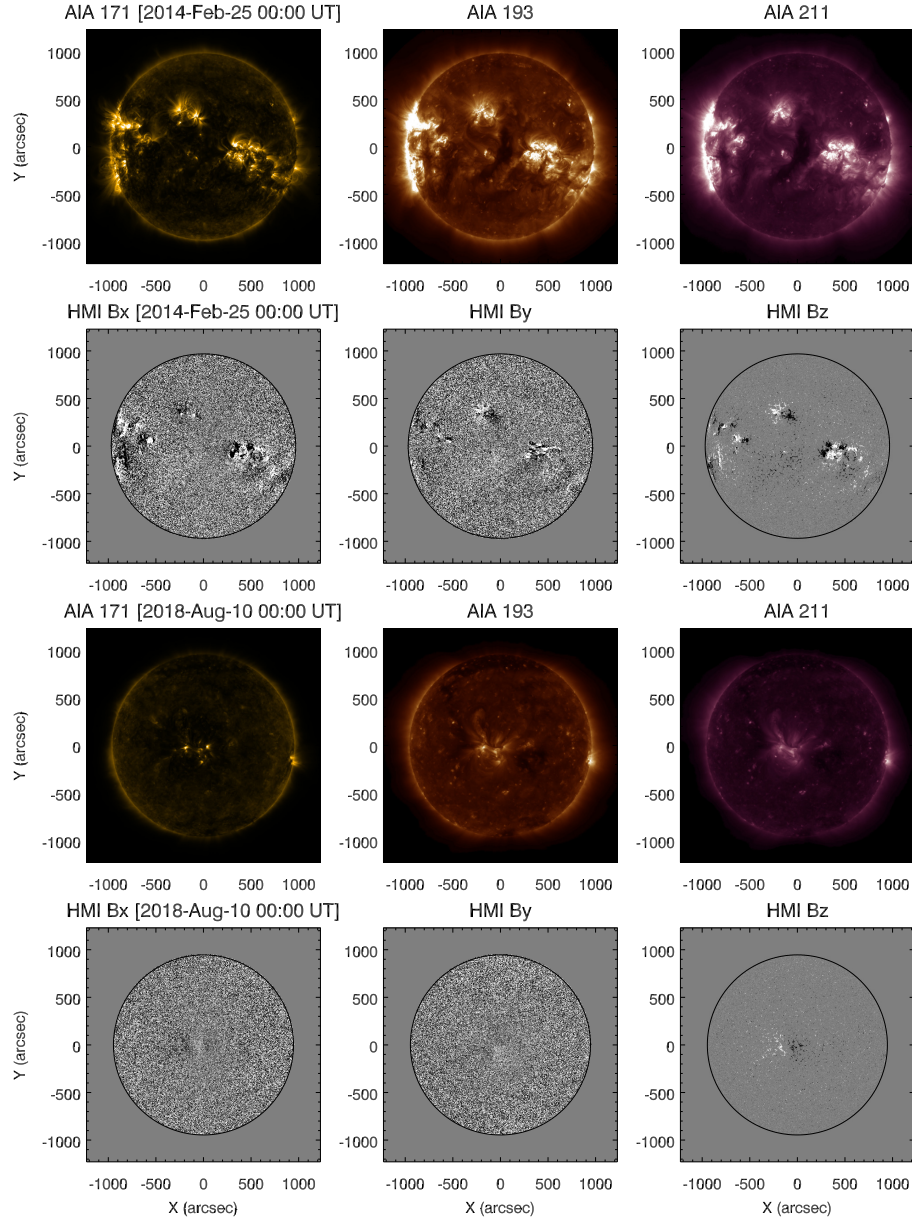


Figure A.4: Upper Panel: Co-aligned AIA and HMI dataset around solar maximum (2014-February-25 00:00 UT) of solar cycle 24. Three selected AIA wavebands (171, 193, 211 Å) are shown; Bottom Panel: Co-aligned AIA and HMI dataset around solar minimum (2018-August-10 00:00 UT) of solar cycle 24. The black circle in the HMI magnetogram shows the location of solar limb.

EVE spectra are assembled from a battery of instruments including the Multiple EUV Grating Spectrographs (MEGS-A, B, P), Solar Aspect Monitor (SAM), and the Extreme ultraviolet SpectroPhotometer (ESP). Each of these instruments covers a different wavelength range in the EUV spectra and are cross calibrated to produce EVE's data products.

The EVE data released in this dataset is extracted from a specially prepared EVE version 5 IDL saveset, including 39 emission lines (see Table A.1) during the time window between 2010-05-01 and 2014-05-26. The end date of this dataset corresponds to the failure of the MEGS-A instrument, which covered the range between 30 and 370. The EVE data have already been calibrated with physical units of Wm^{-2} , scaled to 1 AU, and corrected for degradation, requiring no subsequent calibration. The only processing we perform is to convert from IDL to NumPy Arrays and temporally synchronize with the AIA and HMI observations.

3.4. Temporal Downsampling and Synchronization

One of the goals of this paper is to produce a dataset that is temporally and spatially synchronized for the three SDO data products at manageable resolutions. While our scaling to a fixed solar disk size automatically ensures the spatial synchronization of AIA and HMI, all SDO data instruments observe at different cadences (AIA: 2 minutes, HMI: 12 minutes, EVE: 10 seconds) and are not necessarily aligned in time.

In order to perform the temporal synchronization, we downsample AIA to a 6 minute cadence and match the nearest EVE observation within a mean/max time window of 8.5s/12s. This yields a final dataset consisting of AIA and EVE observations each at 6 minute cadence, with accompanying HMI observations occurring

at every other time step⁵ with a 12 minute cadence.

3.5. Data

This produces the final dataset for the paper totalling ~ 6.5 TB, made available through the Stanford Digital Repository⁶ (please see appendix for a list of URLs). The data are individually packed monthly and for each waveband/component of AIA/HMI, all in the NumPy format. The EVE data are packed in a single TAR file. We show the average value as a function of time for the three products in Figure A.3, which demonstrates the removal of spurious trends and artifacts. We also show in Figure A.4 an example of co-aligned AIA and HMI observations. The upper panel shows the observation near the solar maximum of cycle 24 (2014 February 25 00:00 UT), exhibiting several active regions with strong magnetic field magnitudes and associated EUV emission. The bottom panel shows the observation near the solar minimum of cycle 24 (2018 August 10 00:00 UT), displaying only one active region with a comparatively weak magnetic field and EUV brightness.

4. Protocols and Metrics

We expect that this data will be of interest for machine learning applications in heliophysics and a simple-access dataset for the testing of machine learning models in the physical sciences. To facilitate this, we have defined standard protocols and metrics to aid future work with this data.

Data splits: There is large temporal coherence in the data since large-scale structures on the Sun evolve at timescales beyond days and months. This leads to issues

⁵With the exception of non-zero QUALITY observations for either AIA or HMI.

⁶<https://purl.stanford.edu/nk828sc2920>

with randomly sampled splits of the data, often done in machine learning settings with uncorrelated data. In particular, randomly sampled splits will lead to training and testing observations that are separated by days or even minutes. While these observations are indeed distinct points in time, they are generated by virtually the same large-scale structures.

In practice, this means that experiments on randomly split data will be unable to identify overfitting and likely lead to overly optimistic estimates of generalization performance. The specific issue is that when deploying a model, one tests it on large-scale structures and conditions that are different than the training data. However, if the data is split randomly, the model is never actually evaluated on unseen large-scale structures due to temporal coherence. Therefore, there is no indication of whether the model's performance is due to generalizing well or if it is simply explained by the model overfitting to the particularities of the limited large-scale structures observed at training time.

To preclude this, we have split our data in temporal blocks that break this correlation, consisting of (i) a training set used to fit model parameters (e.g., the filter weights of a convolutional neural network); (ii) a validation set used to set model hyperparameters (e.g., the learning rate for training a network); and (iii) a test set used to evaluate out-of-sample model performance.

All of our data splits are performed over the years (2011-2014), the time period for which all three SDO instruments (AIA, HMI, EVE) were active. This time period provides a dataset large enough to support the training of modern models that require copious amounts of data. We set aside years 2012 and 2013 for testing purposes, supplying a wide variety of solar conditions. Years 2011 and 2014 are split into training and validation such that 70% of available EVE observations are used for training (until mid-December 2011) and 30% are used for validation.

Of course, other splits are possible, especially for problems not relying on EVE observations. We therefore encourage the community to experiment with various data splits, with the cautionary advice that splits should be constructed in temporal blocks as opposed to random sampling.

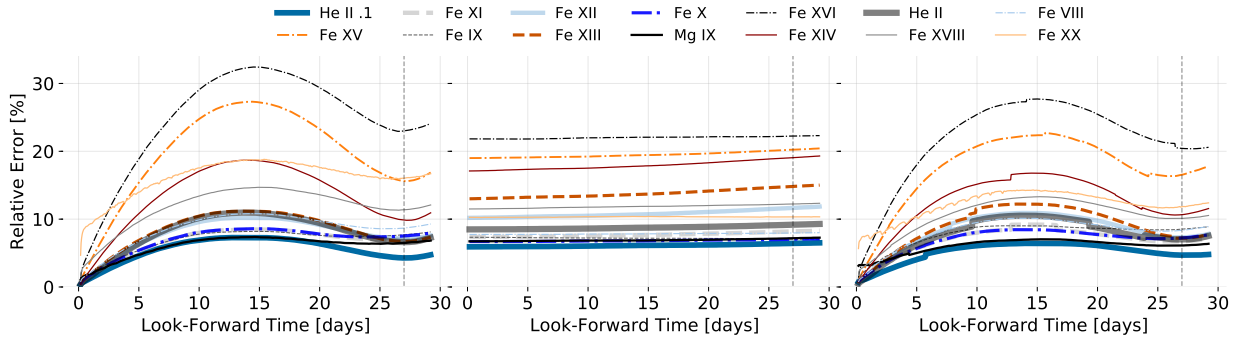


Figure A.5: Results from the 2 hour EVE Prediction experiment. Left Panel: Persistence model, Middle Panel: Average assumption model, Right Panel: Ridge regression model. This prediction exercise is performed for illustration purposes on 14 MEGS-A emission lines (see Table A.1). The forecast errors of all intervals, N hours apart, contained in the years 2012 and 2013 are averaged to produce the average error plotted for an N hour look-forward time in these figures. The average model surprisingly predicts 7 of the 14 EVE lines within 10% error not showing much overall variation, while the persistence model achieves this same performance for 10 lines. The ridge regression model often outperforms the persistence model overall, but not in all conditions and not by a substantial margin. The linear and persistence models both show periodic trends consistent with one solar rotation. See Section 5 for discussion.

Metrics: All of the metrics reported in Section 5 are derived from the normalized absolute error, or $|y_i - \hat{y}_i|/y_i$ where $\hat{y}$ is the model prediction and y the measurement for data point i . For scalar quantities like EVE prediction, that are intrinsically already averaged over the Sun, we report the average normalized absolute error over all samples in the test set.

For images (e.g., AIA prediction) that are not already spatially integrated, we report a number of metrics. First, we report the average normalized absolute error

averaging first over each predicted image’s valid pixels, and then over the images. In computer vision research, this average has been noted to often poorly characterize the performance on most pixels [Scharstein and Szeliski \(2002\)](#) since it can be arbitrarily changed by a small number of large errors. Thus, we also report the percentage of good pixels metric, or the fraction of image pixels with normalized absolute error less than a fixed percentage t for $t = 10\%, 20\%, 50\%$.

5. Results

In this section, we provide baseline metrics for simple machine learning applications utilizing the proposed dataset, all implemented in the `PyTorch` library [Paszke et al. \(2017\)](#). These examples were chosen to illustrate what performance metrics should be expected from future models, as well as supplying simple examples for typical use cases. To this end, we have selected and evaluated two problems that demonstrate the temporal nature of the data as well as the alignment between the two spatially-resolved sensors: (i) Predicting future EVE from present EVE, and (ii) Transforming HMI observations to AIA observations. These generic models may be re-applied to a wide variety of other problems not discussed in this section, for instance predicting future AIA from current AIA, or predicting EVE from AIA.

We stress that our baselines are not intended to be the top-performing solutions, but rather a rubric that shows how well a simple data-driven approach would perform. This serves two functions: First, future model implementations that are more complex should out-perform these baselines in the metrics we propose or other such metrics (e.g., focusing only on flaring events); and secondly, the baselines provide context necessary to properly evaluate a future model’s performance. For instance, while a more complex model may achieve a low error rate such as

5%, if our baseline already achieved a similar score, the complexity of the new model may not be warranted.

5.1. EVE-to-EVE Prediction

The goal of this task is to predict future EVE observations given current EVE observations at a future time ranging from a few hours up to a full solar rotation. In order to provide statistically sound benchmarks in light of strong solar variability, we calculate the average relative error over all predictions for look-forward times of various fixed sizes. This statistical approach informs to what extent we can predict overall future phenomena for a given look-forward time, as well as account for strong solar variability.

There are two main sources of solar variability for this timescale. In shorter time-scales, the main source of variability are flares, which increase the EUV radiative output of the Sun by several orders of magnitude in time-scales of minutes and hours. The second is solar rotation itself (27 days at the synodic Carrington rotation rate). Rotation modulates EUV irradiance because active regions (bright in the EUV) have lifespans of 14 to 55 days and can come in and out of view as the Sun rotates. This active region permanence induces strong temporal correlations at look-forward times greater than 27 days as illustrated by the periodicity in Figure A.5, as the Sun’s “same face” rotates into view. For model evaluation, we choose a total look-forward time of 29 days, a duration long enough to expose the irradiance periodicity.

For our input data we use the MEGS-A lines listed in Table A.1 with the exception of Fe XVI 361 Å which is the most sparsely measured line with only $\sim 1\%$ of the average number of line measurements.

Baselines: For this problem we report 3 simple baselines:

1. *Persistence.* This model assumes that all future observations of the Sun will be identical to its current state. Thus, for any time jump, it predicts that the future EVE observation will be the same as the current EVE observation.
2. *Constant.* This model assumes that the Sun produces a constant EUV irradiance and therefore gives a constant prediction irrespective of the current EVE observation. We set this constant to the training set average per line.
3. *Linear.* This model assumes that future observations are a linear transformation of the current observations plus a constant bias. We fit this model using ridge regression, or a linear model with Tikhonov/L2 regularization. In particular, for a given spectral line and look-forward time, if x_i is the current measurement and y_i the corresponding future average observation, we solve for w such that $\lambda ||w||_2^2 + \sum_i^n ||x_i^T - y_i||$ is minimized for all instances i . We set λ per model by grid search to minimize the average normalized absolute error, doing 2-fold cross validation on the training set.

Results: We evaluate the average normalized absolute error for these models for look-forward times ranging from 2 hours to 29 days in steps of 2 hours and report our results in Figure A.5. The linear and persistence both show trends corresponding to the solar rotation: their errors peak at approximately half a solar rotation and reduce steadily until a full rotation occurs, thus confirming the strong correlation between observations separated by exact rotations. The average model’s error on the other hand is effectively constant; small variations occur because pairs of 1 day jump observations exist from January 1, 2012 up to December 30, 2013; while pairs of 29 day jumps can only be tested up to December 2, 2013.

Collectively, the results underscore the importance of having good baselines via

the surprising effectiveness of even trivial models such as the persistence or average models. For instance, although the average model entirely ignores the current EVE observation, it is able to predict 7 of the 14 EVE lines with less than 10% average normalized absolute error; and similarly, at a look-forward time of 27 days, 10 lines can still be predicted within 10% error by the persistence model.

It is true that the linear model frequently improves on the persistence model, especially for high-error lines like Fe XVI and Fe XV, and look-forward times much less than a full rotation. However, for many look-forward times and lines, the trivial persistence model actually outperforms the relatively complex linear model, demonstrating how simple baselines may assist in properly assessing the effectiveness of a machine learning model.

	Mean (Lower Better)										% Pixels < 10% Error (Higher Better)									
	Avg	94	131	171	193	211	304	335	1600	1700	Avg	94	131	171	193	211	304	335	1600	1700
3 Layer	2.08	0.80	0.98	4.63	5.39	3.70	0.85	0.72	0.73	0.90	15.4	12.0	13.8	12.1	14.7	12.5	15.8	11.1	22.9	23.5
7 Layer	0.83	<u>0.35</u>	<u>0.38</u>	<u>1.55</u>	1.94	1.27	0.54	0.45	0.46	0.52	18.1	19.5	17.7	14.3	15.6	14.2	15.2	14.2	24.3	27.7
11 Layer	<u>0.75</u>	0.37	0.40	1.66	<u>1.55</u>	<u>1.06</u>	<u>0.47</u>	<u>0.37</u>	<u>0.42</u>	<u>0.47</u>	<u>20.8</u>	<u>20.3</u>	<u>20.6</u>	<u>16.4</u>	<u>17.6</u>	<u>16.3</u>	<u>18.2</u>	<u>17.4</u>	<u>28.9</u>	<u>31.4</u>
	% Pixels < 20% Error										% Pixels < 50% Error (Higher Better)									
	Avg	94	131	171	193	211	304	335	1600	1700	Avg	94	131	171	193	211	304	335	1600	1700
3 Layer	29.2	23.7	27.0	23.6	28.8	24.9	31.1	22.2	40.5	40.7	58.0	53.9	56.3	49.8	61.6	57.4	65.3	54.4	62.5	60.6
7 Layer	34.5	38.3	34.7	27.7	30.4	27.9	30.0	28.2	44.1	49.5	68.9	<u>80.0</u>	74.5	58.0	64.0	61.5	65.1	64.2	74.5	78.7
11 Layer	<u>39.5</u>	<u>39.2</u>	<u>39.6</u>	<u>31.9</u>	<u>34.3</u>	<u>31.9</u>	<u>35.5</u>	<u>34.6</u>	<u>52.2</u>	<u>56.6</u>	<u>75.0</u>	77.5	<u>76.4</u>	<u>65.1</u>	<u>70.6</u>	<u>68.2</u>	<u>73.8</u>	<u>77.2</u>	<u>81.6</u>	<u>84.3</u>

Table A.2: Results for HMI to AIA prediction. The top performing method is underlined.

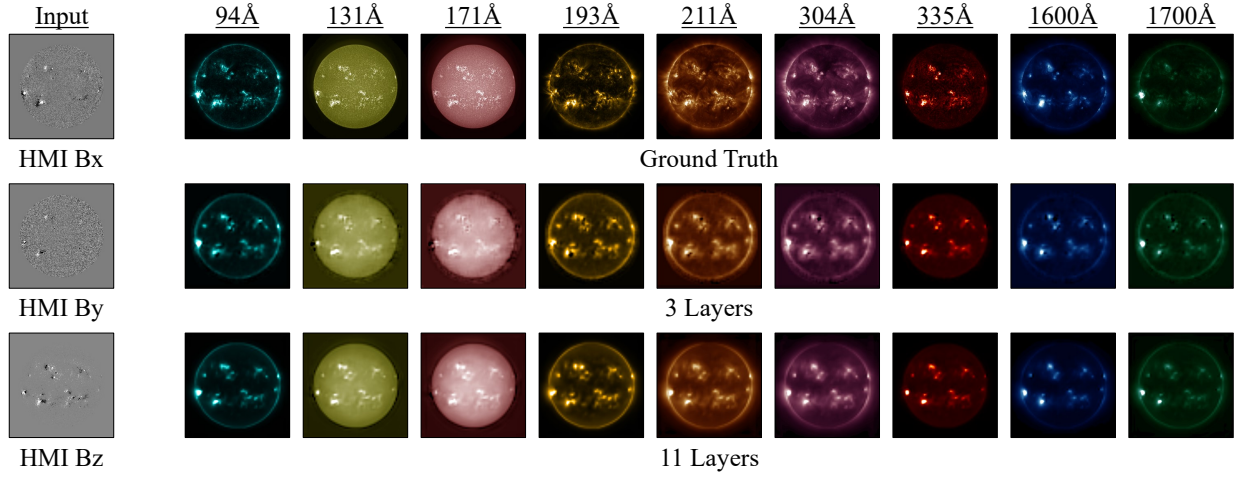


Figure A.6: Results for HMI to AIA translation. The Left Panel shows the HMI inputs, while the Right Panel shows the ground-truth AIA (Top Panel) as well as the predicted AIA from a 3-layer network (Middle Panel) and 11-layer network (Bottom). While the 3-layer network performs well, additional layers (i) reduce artifacts (especially in 131 Å and 171 Å) presumably due to the depth, and (ii) better resolve coronal predictions (especially in 211 Å and 304 Å), presumably due to the larger receptive field caused by the additional layers.

5.2. HMI-to-AIA Prediction

We now move on to an example which demonstrates how a convolutional deep learning model may exploit the spatial richness of our dataset. In this application we show how a mapping between the HMI and AIA instruments is learned by treating it as an image-to-image translation problem. Such an approach is common in computer vision research, with applications as diverse as labeling each pixel in the scene with a category label (e.g., building) (Shelhamer et al., 2016), generating images from sketches (Isola et al., 2017), inferring 3D properties of scenes (Wang et al., 2015), or detecting the pose of humans (Cao et al., 2017).

We have physical reason to expect that there exists a mapping between the HMI and AIA instruments. While the HMI instrument infers information about the solar magnetic field from the solar photosphere, the AIA instrument measures

UV/EUV emission from the solar chromosphere and corona. Since the chromosphere and corona are spatially structured by the presence of strong magnetic fields, UV/EUV emission will typically reflect information about the magnetic field through its spatial distribution, and *vice versa*. Here we show how a simple convolutional model can realize the mapping from HMI to AIA.

Baseline: Our baseline is a deep convolutional neural network. This is a function composed of alternating convolutions and non-linearities that maps one multi-channel image (e.g., a 3-channel 256x256 image) to another (e.g., a 9-channel 256x256 image). This function can be fit to a dataset of inputs (i.e., HMI) and desired outputs (i.e., AIA) via standard optimization procedures. Throughout we work with 256x256 images.

We adopt a basic approach for our network consisting of a three parts: (i) an initial feature extraction following ResNet (He et al., 2016a) consisting of a 7x7 convolution with stride 2 followed by 3x3 max-pooling with stride 2, which expands the receptive fields of subsequent feature maps; (ii) a variable number of 3x3 convolutions with stride 1; (iii) 3x3 convolution yielding a 9-channel prediction followed by 4x bilinear upsampling. All intermediate convolutions have 128 filters and are followed by a Rectified Linear Unit (Nair and Hinton, 2010) and Batch normalization (Ioffe and Szegedy, 2015). By varying the number of intermediate convolutions blocks in part ii, we can control both parameter count and effective receptive field of the network. We report results with 3, 7, and 11 layers (i.e., with 2, 6, and 10 hidden layers, including the initial convolution in part i).

We train the parameters of the network (e.g., filter weights and biases) via back-propagation and mini-batch stochastic gradient descent (SGD) to minimize the mean-squared error of the prediction. Specifically, we use SGD with Nesterov

momentum (Sutskever et al., 2013a) with momentum 0.99, weight decay 10^{-8} , and batch size 32. We start with a learning rate of 10^{-3} , which we multiply by 0.1 every 5 epochs, and train for 15 epochs. We checkpoint the network at the end of every epoch and take the network with lowest validation loss. To help learning, we divide inputs and outputs per-channel by their average over the training set (i.e., network is trained to predict the AIA 94 Å image divided by the empirical mean of AIA 94 Å over the training set).

Results: We show sample qualitative results in Figure A.6 for 3 and 11 layers. Even with a small number of hidden layers, a simple data-driven approach does a good job of getting the general shape and features of the Sun, which suggests that results that get general features can be explained with relatively simple models and that more complex models must provide additional results compared to this. Adding more layers helps reduce artifacts at the edge of the disk in the 131 Å and 171 Å channels and more accurately resolve the corona in the 211 Å and 304 Å channels. The shallower network has difficulty accurately resolving the corona because each prediction is made from a small portion of the Sun; thus it produces a halo-like effect around the entire Sun, as opposed to at specific locations on the disk.

Quantitatively, increasing depth produces strong improvements, as seen in Table A.2. With a relatively unsophisticated deep network, 75% of the pixels of AIA across all channels can be predicted within 50% relative error from HMI observations. As seen by the percentage good pixels metrics, 1600Å and 1700Å observations appear to be among the easier to predict, and are almost always a few percentage points higher across both network depths and good pixel thresholds. This serves as a good sanity check on the results, since the photospheric and chromospheric brightness features in these two channels are known to be highly correlated to the photospheric distribution of magnetic fields.

6. Conclusion

In this paper we present a curated, high quality dataset from all three SDO instruments primed for machine learning research. We have preprocessed this data by downsampling AIA and HMI images from 4096×4096 to 512×512 pixels; removed `QUALITY` $\neq 0$ observations, corrected for instrumental degradation over time, and applied exposure corrections that account for Earth’s elliptical orbit as well as AIA’s automatic exposure control. We also have ensured that both AIA and HMI data are spatially co-located, have identical angular resolutions, and that all instruments are chrono-synchronous.

We also highlight some of the potential pitfalls of blindly applying machine learning techniques to solar data, or even more broadly:

1. To maximize its versatility, SDO data products are nuanced and assume an expert-level understanding of its instrumental design and limitations. Using them without this knowledge may lead to incorrect results and invalid conclusions.
2. Most of the physical processes that drive solar variability occur at a much slower cadence than that of SDO’s instruments (hours and days vs. minutes and seconds, respectively), requiring special care with the splitting of training, validation, and test sets. Splits must be performed along temporal blocks and not by random sampling, as is done in other settings with uncorrelated data samples. Random sampling in this case will lead to an overly optimistic estimate of validation error, leading to an inability to identify whether a model will generalize properly to future observations or has instead overfit to its data.
3. Due to the relatively short timescales of solar variability, the simple fore-

casting models of permanence and climatological averages perform exceptionally well in hourly and daily timescales. Due to this, error estimates of more advanced models are not meaningful in an absolute sense, but rather only when compared to these simple baselines.

Finally, we provide a series of baselines that take advantage of this dataset to produce EVE time-forecasts and HMI→AIA reconstructions. These examples are meant to illustrate some of the applications made possible by combining these data with ML techniques, as well as what heuristic performance measures one should expect to compare their own model implementations with.

As with many fields, Heliophysics has entered a data-rich age in which the human intellect alone is incapable of processing the copious amounts of data gathered by NASA's ever-growing spacecraft fleet. Fortunately, the ongoing revolution in machine learning research will power a new age of data inference and physical insight that maximizes the scientific output of these data-rich missions. It is important however for heliophysicists and computer scientists to work together to understand the properties and limitations of both the raw data and the ML techniques. If special care is not taken in understanding such limitations, we may unfortunately see a large amount of incorrect, overly optimistic, or worse, misleading research. Interdisciplinary programs such as NASA's Frontier Development Laboratory can provide a vital common ground to facilitate this skill transfer and will be highly critical for the successful and fruitful development of ML techniques in the astrophysical sciences.

Acknowledgements

This project was conducted during the 2018 NASA Frontier Development Lab (FDL) program, a public/private partnership between NASA and SETI and industry partners including NVIDIA Corporation, Lockheed Martin and IBM. The authors thank IBM (especially Naeem Altaf) for generously providing computing resources on the IBM Cloud. We gratefully thank our mentors for guidance and useful discussion, as well as the SETI Institute for their hospitality. RG acknowledges support from the Moore-Sloan Data Science Environment at New York University and thanks Rob Fergus for useful discussion. The authors acknowledge support from NASA’s SDO/AIA contract (NNG04EA00C) to LMSAL. AIA is an instrument onboard the Solar Dynamics Observatory, a mission for NASA’s Living With a Star program. MCMC acknowledges support from NASA’s Heliophysics Grand Challenges Research grant (NNX14AI14G). PyTorch ([Paszke et al., 2017](#)), SunPy ([SunPy Community et al., 2015](#)), SolarSoft ([Freeland and Handy, 1998](#)).

Data availability

The dataset is made available through the Stanford Digital Repository, partitioned by year. The data may be obtained via the links provided in Table 2, totalling 6.5TBs.

Table A.3: Dataset urls and Digital Object Identifiers

Stanford Digital Repository url	DOI
https://purl.stanford.edu/vk217bh4910	[10.25740/ppax-bf07] https://doi.org/10.25740/ppax-bf07
https://purl.stanford.edu/jc488jb7715	[10.25740/sb4q-wj06] https://doi.org/10.25740/sb4q-wj06
https://purl.stanford.edu/dc156hp0190	[10.25740/1vyz-b592] https://doi.org/10.25740/1vyz-b592
https://purl.stanford.edu/km388vz4371	[10.25740/2zme-3q44] https://doi.org/10.25740/2zme-3q44
https://purl.stanford.edu/sr325xz9271	[10.25740/3jhw-x180] https://doi.org/10.25740/3jhw-x180
https://purl.stanford.edu/qw012qy2533	[10.25740/0fbp-re41] https://doi.org/10.25740/0fbp-re41
https://purl.stanford.edu/vf806tr8954	[10.25740/64cr-bc95] https://doi.org/10.25740/64cr-bc95
https://purl.stanford.edu/kp222tm1554	[10.25740/c8bw-ar96] https://doi.org/10.25740/c8bw-ar96
https://purl.stanford.edu/nk828sc2920	[10.25740/pknx-5s37] https://doi.org/10.25740/pknx-5s37

Bibliography

- L. Auer, L. Boschi, T. Becker, T. Nissen-Meyer, and D. Giardini. Savani: A variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets. *Journal of Geophysical Research: Solid Earth*, 119(4):3006–3034, 2014.
- M. L. Bacci. *A concise history of world population*. John Wiley & Sons, 2017.
- Y. Bar-Yam. General features of complex systems. *Encyclopedia of Life Support Systems (EOLSS)*, UNESCO, EOLSS Publishers, Oxford, UK, 1, 2002.
- A. D. Barnosky, P. L. Koch, R. S. Feranec, S. L. Wing, and A. B. Shabel. Assessing the causes of late pleistocene extinctions on the continents. *science*, 306(5693):70–75, 2004.
- J. T. Barron, M. D. Biggin, P. Arbelaez, D. W. Knowles, S. V. Keranen, and J. Malik. Volumetric semantic segmentation using pyramid context features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3448–3455, 2013.
- C. D. Baso and A. A. Ramos. Enhancing sdo/hmi images using deep learning. *Astronomy & Astrophysics*, 614:A5, 2018.
- D. R. Bellwood, T. P. Hughes, C. Folke, and M. Nyström. Confronting the coral reef crisis. *Nature*, 429(6994):827–833, 2004.
- M. J. Berger and J. Oliger. Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of computational Physics*, 53(3):484–512, 1984.
- J. Bialas and E. R. Flueh. Ocean bottom seismometers. *Sea Technol*, 40(4):41–46, 1999.

- H. Bijwaard and W. Spakman. Tomographic evidence for a narrow whole mantle plume below iceland. *Earth and Planetary Science Letters*, 166(3-4):121–126, 1999.
- C. M. Bishop. Pattern recognition. *Machine learning*, 128(9), 2006.
- S. Blake, I. Douglas-Hamilton, and W. Karesh. Gps telemetry of forest elephants in central africa: results of a preliminary study. *African journal of ecology*, 39(2):178–186, 2001.
- M. G. Bobra and S. Couvidat. Solar Flare Prediction Using SDO/HMI Vector Magnetic Field Data with a Machine-learning Algorithm. , 798:135, Jan. 2015. doi: 10.1088/0004-637X/798/2/135.
- P. Boerner, C. Edwards, J. Lemen, A. Rausch, C. Schrijver, R. Shine, L. Shing, R. Stern, T. Tarbell, A. Title, C. J. Wolfson, R. Soufli, E. Spiller, E. Gullikson, D. McKenzie, D. Windt, L. Golub, W. Podgorski, P. Testa, and M. Weber. Initial Calibration of the Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). , 275:41–66, Jan. 2012. doi: 10.1007/s11207-011-9804-8.
- P. F. Boerner, P. Testa, H. Warren, M. A. Weber, and C. J. Schrijver. Photometric and Thermal Cross-calibration of Solar EUV Instruments. , 289:2377–2397, June 2014. doi: 10.1007/s11207-013-0452-z.
- E. Bozdağ, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire. Global adjoint tomography: first-generation model. *Geophysical Journal International*, 207(3):1739–1766, 2016.
- T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Nee-lakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

- Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Computer Vision and Pattern Recognition*, 2017.
- J. Carreira and A. Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017.
- G. Ceballos, P. R. Ehrlich, and P. H. Raven. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences*, 117(24):13596–13602, 2020.
- M. J. Chase, S. Schlossberg, C. R. Griffin, P. J. Bouché, S. W. Djene, P. W. Elkan, S. Ferreira, F. Grossman, E. M. Kohi, K. Landen, et al. Continent-wide survey reveals massive decline in african savannah elephants. *PeerJ*, 4:e2354, 2016.
- L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- M. C. Cheung, P. Boerner, C. Schrijver, P. Testa, F. Chen, H. Peter, and A. Malanushenko. Thermal diagnostics with the atmospheric imaging assembly on board the solar dynamics observatory: a validated method for differential emission measure inversions. *The Astrophysical Journal*, 807(2):143, 2015.
- J. Clemente, W. Song, M. Valero, F. Li, and X. Liy. Indoor person identification and fall detection through non-intrusive floor seismic sensing. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 417–424. IEEE, 2019.
- T. Colak and R. Qahwaji. Prediction of extreme ultraviolet variability experiment (eve)/extreme ultraviolet spectro-photometer (esp) irradiance from solar

- dynamics observatory (sdo)/atmospheric imaging assembly (aia) images using fuzzy image processing and machine learning. *Solar Physics*, 283(1):143–156, Mar 2013. ISSN 1573-093X. doi: 10.1007/s11207-011-9880-9. URL <https://doi.org/10.1007/s11207-011-9880-9>.
- A. Colombi, T. Nissen-Meyer, L. Boschi, and D. Giardini. Seismic waveform inversion for core–mantle boundary topography. *Geophysical Journal International*, 198(1):55–71, 2014.
- N. R. Council et al. *Origin and evolution of Earth: Research questions for a changing planet*. National Academies Press, 2008.
- F. Dahlen, S.-H. Hung, and G. Nolet. Fréchet kernels for finite-frequency travel-timesâi. theory. *Geophysical Journal International*, 141(1):157–174, 2000.
- H. Dai and C. MacBeth. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical journal international*, 120(3):758–774, 1995.
- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- K. Dere, E. Landi, H. Mason, B. M. Fossi, and P. Young. Chianti-an atomic database for emission lines-i. wavelengths greater than 50 Å. *Astronomy and Astrophysics Supplement Series*, 125(1):149–173, 1997.
- P. M. DeVries, T. B. Thompson, and B. J. Meade. Enabling large-scale viscoelastic calculations via neural network acceleration. *Geophysical Research Letters*, 44(6):2662–2669, 2017.
- A. Dosovitskiy, J. T. Springenberg, M. Tatarchenko, and T. Brox. Learning to

- generate chairs, tables and cars with convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):692–705, 2016.
- I. Douglas-Hamilton. African elephants: population trends and their causes. *Oryx*, 21(1):11–24, 1987.
- F. U. Dowla, S. R. Taylor, and R. W. Anderson. Seismic discrimination with artificial neural networks: preliminary results with regional spectral data. *Bulletin of the Seismological Society of America*, 80(5):1346–1373, 1990.
- M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, et al. Sentinel-2: Esa’s optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.
- I. Duporge, O. Isupova, S. Reece, D. W. Macdonald, and T. Wang. Using very-high-resolution satellite imagery and deep learning to detect and count african elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation*, 2020.
- P. S. Dysart and J. J. Pulli. Regional seismic event classification at the noress array: seismological measurements and the use of trained neural networks. *Bulletin of the Seismological Society of America*, 80(6B):1910–1933, 1990.
- A. M. Dziewonski and D. L. Anderson. Preliminary reference earth model. *Physics of the earth and planetary interiors*, 25(4):297–356, 1981.
- A. M. Dziewonski, B. H. Hager, and R. J. O’Connell. Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research*, 82(2):239–255, 1977.
- J. Elliott and J. Peraire. Practical three-dimensional aerodynamic design and optimization using unstructured meshes. *AIAA journal*, 35(9):1479–1485, 1997.

- A. Fichtner. *Full seismic waveform modelling and inversion*. Springer Science & Business Media, 2010.
- A. Fichtner and M. van Driel. Models and fréchet kernels for frequency-(in) dependent q . *Geophysical Journal International*, 198(3):1878–1889, 2014.
- A. Fichtner, B. L. Kennett, H. Igel, and H.-P. Bunge. Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725, 2009.
- J. Fontenla, M. Codrescu, M. Fedrizzi, T. Fuller-Rowell, F. Hill, E. Landi, and T. Woods. Five years of synthesis of solar spectral irradiance from sdiid/sisa and sdo/aia images. *The Astrophysical Journal*, 834(1):54, 2016.
- S. Freeland and B. Handy. Data analysis with the solarsoft system. *Solar Physics*, 182(2):497–500, Oct 1998. ISSN 1573-093X. doi: 10.1023/A:1005038224881. URL <https://doi.org/10.1023/A:1005038224881>.
- S. W. French and B. Romanowicz. Broad plumes rooted at the base of the earth’s mantle beneath major hotspots. *Nature*, 525(7567):95–99, 2015.
- M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321:321–331, 2018.
- M. A. Friedl, D. K. McIver, J. C. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, et al. Global land cover mapping from modis: algorithms and early results. *Remote sensing of Environment*, 83(1-2):287–302, 2002.
- J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

- V. Galanti, G. Tosi, R. Rossi, and C. Foley. The use of gps radio-collars to track elephants (*loxodonta africana*) in the tarangire national park (tanzania). *Hystrix-the Italian Journal of Mammalogy*, 11(2), 2000.
- R. Galvez, D. F. Fouhey, M. Jin, A. Szenicer, A. Muñoz-Jaramillo, M. C. Cheung, P. J. Wright, M. G. Bobra, Y. Liu, J. Mason, et al. A machine-learning data set prepared from the nasa solar dynamics observatory mission. *The Astrophysical Journal Supplement Series*, 242(1):7, 2019.
- P. Gerland, A. E. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema, B. K. Fosdick, J. Chunn, N. Lalic, et al. World population stabilization unlikely this century. *Science*, 346(6206):234–237, 2014.
- M. B. Giles, M. C. Duta, J.-D. Muller, and N. A. Pierce. Algorithm developments for discrete adjoint methods. *AIAA journal*, 41(2):198–205, 2003.
- I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.
- S. P. Grand, R. D. Van der Hilst, and S. Widiyantoro. High resolution global tomography: a snapshot of convection in the earth. *Geological Society of America Today*, 7(4), 1997.
- C. G. Harris, M. Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.
- T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*, 2016a.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016b.
- J. T. Hoeksema, Y. Liu, K. Hayashi, X. Sun, J. Schou, S. Couvidat, A. Norton, M. Bobra, R. Centeno, K. D. Leka, G. Barnes, and M. Turmon. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: Overview and Performance. , 289:3483–3530, Sept. 2014. doi: 10.1007/s11207-014-0516-8.
- K. Hosseini and K. Sigloch. Multifrequency measurements of core-diffracted p waves (pdiff) for global waveform tomography. *Geophysical Journal International*, 203(1):506–521, 2015.
- K. Hosseini, K. Sigloch, M. Tsekhmistrenko, A. Zaheri, T. Nissen-Meyer, and H. Igel. Global mantle structure from multifrequency tomography using p, pp and p-diffracted waves. *Geophysical Journal International*, 220(1):96–141, 2020.
- S. Huang, J. Xiang, H. Du, and X. Cao. Inverse problems in atmospheric science and their application. In *Journal of Physics: Conference Series*, volume 12, page 005. IOP Publishing, 2005.
- X. Huang, H. Wang, L. Xu, J. Liu, R. Li, and X. Dai. Deep learning based solar flare forecasting model. i. results for line-of-sight magnetograms. *The Astrophysical Journal*, 856(1):7, mar 2018. doi: 10.3847/1538-4357/aaae00. URL <https://doi.org/10.3847%2F1538-4357%2Faaae00>.
- F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer.

- Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, 2015.
- P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Computer Vision and Pattern Recognition*, 2017.
- H. Jeffreys. The rigidity of the earth’s central core. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 1(7):371–383, 1926.
- E. Jonas, M. Bobra, V. Shankar, J. Todd Hoeksema, and B. Recht. Flare Prediction Using Photospheric and Coronal Image Data. , 293:48, Mar. 2018. doi: 10.1007/s11207-018-1258-9.
- J. E. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. M. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, et al. The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349, 2015.
- R. Kays, B. Kranstauber, P. Jansen, C. Carbone, M. Rowcliffe, T. Fountain, and S. Tilak. Camera traps as sensor networks for monitoring animal communities. In *2009 IEEE 34th Conference on Local Computer Networks*, pages 811–818. IEEE, 2009.
- D. J. Kessler and B. G. Cour-Palais. Collision frequency of artificial satellites: The creation of a debris belt. *Journal of Geophysical Research: Space Physics*, 83(A6):2637–2646, 1978.

- T. Knutson, S. J. Camargo, J. C. Chan, K. Emanuel, C.-H. Ho, J. Kossin, M. Mo-hapatra, M. Satoh, M. Sugi, K. Walsh, et al. Tropical cyclones and climate change assessment: Part ii: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, 101(3):E303–E322, 2020.
- T. R. Knutson, J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. Srivastava, and M. Sugi. Tropical cyclones and climate change. *Nature geoscience*, 3(3):157–163, 2010.
- D. Komatitsch, Z. Xie, E. Bozdağ, E. Sales de Andrade, D. Peter, Q. Liu, and J. Tromp. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion. *Geophysical Journal International*, 206(3):1467–1478, 2016.
- A. Krizhevsky. One weird trick for parallelizing convolutional neural networks. *arXiv preprint arXiv:1404.5997*, 2014.
- A. Kucuk, J. M. Banda, and R. A. Angryk. A large-scale solar dynamics observatory image dataset for computer vision applications. *Scientific Data, Volume 4, id. 170096 (2017)*, 4:170096, July 2017. doi: 10.1038/sdata.2017.96.
- B. Kustowski, G. Ekström, and A. Dziewoński. Anisotropic shear-wave velocity structure of the earth’s mantle: A global model. *Journal of Geophysical Research: Solid Earth*, 113(B6), 2008.
- O. D. Lamb, M. J. Shore, J. M. Lees, S. J. Lee, and S. M. Hensman. Assessing raspberry shake and boom sensors for recording african elephant acoustic vocalizations. *Frontiers in Conservation Science*, 1:10, 2021.
- E. Landi and K. Phillips. Chiantiâan atomic database for emission lines. viii. comparison with solar flare spectra from the solar maximum mission flat crystal spectrometer. *The Astrophysical Journal Supplement Series*, 166(1):421, 2006.

- J. S. Lansing. Complex adaptive systems. *Annual review of anthropology*, 32(1): 183–204, 2003.
- G. Laske, G. Masters, Z. Ma, and M. Pasyanos. Update on crust1.0: a 1-degree global model of earth's crust. In *Geophys. res. abstr*, volume 15, page 2658, 2013.
- J. Lean. Solar ultraviolet irradiance variations: A review. *Journal of Geophysical Research: Atmospheres*, 92(D1):839–868, 1987.
- I. Lehmann. P'. *Bureau Central Sismologique International Strasbourg: Publications du Bureau Central Scientifiques*, 14:87–115, 1936.
- W. Lei, Y. Ruan, E. Bozdağ, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire. Global adjoint tomography model glad-m25. *Geophysical Journal International*, 223(1):1–21, 2020.
- J. R. Lemen, D. J. Akin, P. F. Boerner, C. Chou, J. F. Drake, D. W. Duncan, C. G. Edwards, F. M. Friedlaender, G. F. Heyman, N. E. Hurlburt, et al. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). In *The solar dynamics observatory*, pages 17–40. Springer, 2011.
- J. R. Lemen, A. M. Title, D. J. Akin, P. F. Boerner, C. Chou, J. F. Drake, D. W. Duncan, C. G. Edwards, F. M. Friedlaender, G. F. Heyman, N. E. Hurlburt, N. L. Katz, G. D. Kushner, M. Levay, R. W. Lindgren, D. P. Mathur, E. L. McFeaters, S. Mitchell, R. A. Rehse, C. J. Schrijver, L. A. Springer, R. A. Stern, T. D. Tarbell, J.-P. Wuelser, C. J. Wolfson, C. Yanari, J. A. Bookbinder, P. N. Cheimets, D. Caldwell, E. E. Deluca, R. Gates, L. Golub, S. Park, W. A. Podgorski, R. I. Bush, P. H. Scherrer, M. A. Gummin, P. Smith, G. Auken, P. Jerram, P. Pool, R. Soufli, D. L. Windt, S. Beardsley, M. Clapp, J. Lang, and N. Waltham.

- The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). , 275:17–40, Jan. 2012. doi: 10.1007/s11207-011-9776-8.
- K. Leng, T. Nissen-Meyer, and M. van Driel. Efficient global wave propagation adapted to 3-d structural complexity: a pseudospectral/spectral-element approach. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 207(3):1700–1721, 2016.
- K. Leng, T. Nissen-Meyer, M. Van Driel, K. Hosseini, and D. Al-Attar. Axisem3d: broad-band seismic wavefields in 3-d global earth models with undulating discontinuities. *Geophysical Journal International*, 217(3):2125–2146, 2019.
- K. Leng, J. Korenaga, and T. Nissen-Meyer. 3-d scattering of elastic waves by small-scale heterogeneities in the earth’s mantle. *Geophysical Journal International*, 223(1):502–525, 2020.
- T. T. Leonid and R. Jayaparvathy. Statistical-model based voice activity identification for human-elephant conflict mitigation. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–7, 2020.
- Z. Li, M.-A. Meier, E. Hauksson, Z. Zhan, and J. Andrews. Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, 45(10):4773–4779, 2018.
- T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- J. L. Lions. Optimal control of systems governed by partial differential equations problèmes aux limites. 1971.
- J. Liou and N. L. Johnson. Risks in space from orbiting debris. *SCIENCE-NEW YORK THEN WASHINGTON-*, 311(5759):340, 2006.

- Q. Liu and J. Tromp. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods. *Geophysical Journal International*, 174(1):265–286, 2008.
- L. J. Mangewa, P. A. Ndakidemi, and L. K. Munishi. Integrating uav technology in an ecological monitoring system for community wildlife management areas in tanzania. *Sustainability*, 11(21):6116, 2019.
- M. T. McCann, K. H. Jin, and M. Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017.
- E. D. Mercerat and G. Nolet. On the linearity of cross-correlation delay times in finite-frequency tomography. *Geophysical Journal International*, 192(2):681–687, 2013.
- A. Mohorovicic. Earthquake of 8 october 1909 (potres od 8), 1909.
- R. Montelli, G. Nolet, F. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung. Finite-frequency tomography reveals a variety of plumes in the mantle. *Science*, 303(5656):338–343, 2004a.
- R. Montelli, G. Nolet, G. Masters, F. Dahlen, and S.-H. Hung. Global p and pp traveltimes tomography: rays versus waves. *Geophysical Journal International*, 158(2):637–654, 2004b.
- B. Mortimer, W. L. Rees, P. Koelemeijer, and T. Nissen-Meyer. Classifying elephant behaviour through seismic vibrations. *Current Biology*, 28(9):R547–R548, 2018.
- B. Moseley, A. Markham, and T. Nissen-Meyer. Fast approximate simulation of seismic waves with deep learning. *arXiv preprint arXiv:1807.06873*, 2018.

- B. Moseley, T. Nissen-Meyer, and A. Markham. Deep learning for fast simulation of seismic waves in complex media. *Solid Earth*, 11(4):1527–1549, 2020.
- S. M. Mousavi and G. C. Beroza. A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, 47(1):e2019GL085976, 2020.
- S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1):1–12, 2020.
- V. Nair and G. Hinton. Rectified linear units improve restricted boltzmann machines. In *International Conference on Machine Learning*, 2010.
- S. M. Ngene, H. Van Gils, S. E. Van Wieren, H. Rasmussen, A. K. Skidmore, H. H. Prins, A. G. Toxopeus, P. Omondi, and I. Douglas-Hamilton. The ranging patterns of elephants in marsabit protected area, kenya: The use of satellite-linked gps collars. *African journal of ecology*, 48(2):386–400, 2010.
- T. Nissen-Meyer, F. Dahlen, and A. Fournier. Spherical-earth frechet sensitivity kernels. *Geophysical Journal International*, 168(3):1051–1066, 2007.
- T. Nissen-Meyer, M. van Driel, S. C. Stähler, K. Hosseini, S. Hempel, L. Auer, A. Colombi, and A. Fournier. Axisem: broadband 3-d seismic wavefields in axisymmetric media. *Solid Earth*, 5(1):425–445, 2014.
- G. Nolet and R. Montelli. Optimal parametrization of tomographic models. *Geophysical Journal International*, 161(2):365–372, 2005.
- R. D. Oldham. The constitution of the interior of the earth, as revealed by earthquakes. *Quarterly Journal of the Geological Society*, 62(1-4):456–475, 1906.
- A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves,

- N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
- K. Palanisamy, D. Singhania, and A. Yao. Rethinking cnn models for audio classification. *arXiv preprint arXiv:2007.11154*, 2020.
- M. P. Panning, P. Lognonné, W. B. Banerdt, R. Garcia, M. Golombek, S. Kedar, B. Knapmeyer-Endrun, A. Mocquet, N. A. Teanby, J. Tromp, et al. Planned products of the mars structure service for the insight mission to mars. *Space Science Reviews*, 211(1):611–650, 2017.
- A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.
- J. Pearl. *Causality*. Cambridge university press, 2009.
- W. D. Pesnell, B. J. Thompson, and P. Chamberlin. The solar dynamics observatory (sdo). In *The Solar Dynamics Observatory*, pages 3–15. Springer, 2011.
- W. D. Pesnell, B. J. Thompson, and P. C. Chamberlin. The Solar Dynamics Observatory (SDO). , 275:3–15, Jan. 2012. doi: 10.1007/s11207-011-9841-3.
- O. Pironneau. On optimum design in fluid mechanics. *Journal of Fluid Mechanics*, 64(1):97–110, 1974.
- J. H. Poole. Behavioral contexts of elephant acoustic communication. *The Amboseli elephants: a long-term perspective on a long-lived mammal. Chicago: The University of Chicago*, pages 125–161, 2011.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems

- involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- A. A. Ramos, I. Requerey, and N. Vitas. Deepvel: Deep learning for the estimation of horizontal velocities at the solar surface. *Astronomy & Astrophysics*, 604: A11, 2017.
- C. Randler and N. Kalb. Distance and size matters: A comparison of six wildlife camera traps and their usefulness for wild birds. *Ecology and evolution*, 8(14): 7151–7163, 2018.
- M. Reinwald, B. Moseley, A. Szenicer, T. Nissen-Meyer, S. Oduor, F. Vollrath, A. Markham, and B. Mortimer. Seismic localization of elephant rumbles as a monitoring approach. *Journal of the Royal Society Interface*, 18(180): 20210264, 2021.
- L. Ren and J. R. Hutchinson. The three-dimensional locomotor dynamics of african (*loxodonta africana*) and asian (*elephas maximus*) elephants reveal a smooth gait transition at moderate speed. *Journal of the Royal Society Interface*, 5(19):195–211, 2008.
- R. Rew and G. Davis. Netcdf: an interface for scientific data access. *IEEE computer graphics and applications*, 10(4):76–82, 1990.
- J. C. Riley. *Rising life expectancy: a global history*. Cambridge University Press, 2001.
- J. Ritsema and H.-J. Van Heijst. Seismic imaging of structural heterogeneity in earth’s mantle: evidence for large-scale mantle flow. *Science Progress (1933-)*, pages 243–259, 2000.
- J. Ritsema, H. J. van Heijst, and J. H. Woodhouse. Complex shear wave velocity

- structure imaged beneath africa and iceland. *Science*, 286(5446):1925–1928, 1999.
- J. Ritsema, a. A. Deuss, H. Van Heijst, and J. Woodhouse. S40rts: a degree-40 shear-velocity model for the mantle from new rayleigh wave dispersion, teleseismic traveltimes and normal-mode splitting function measurements. *Geophysical Journal International*, 184(3):1223–1236, 2011.
- M. Roser. Economic growth. *Our World in Data*, 2013. <https://ourworldindata.org/economic-growth>.
- S. Rost and C. Thomas. Array seismology: Methods and applications. *Reviews of geophysics*, 40(3):2–1, 2002.
- O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- J. Salamon and J. P. Bello. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3):279–283, 2017.
- M. Sambridge, T. Bodin, K. Gallagher, and H. Tkalčić. Transdimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110547, 2013.
- A. Schaeffer and S. Lebedev. Global shear speed structure of the upper mantle and transition zone. *Geophysical Journal International*, 194(1):417–449, 2013.
- D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.

- B. C. Scheele, F. Pasmans, L. F. Skerratt, L. Berger, A. Martel, W. Beukema, A. A. Acevedo, P. A. Burrowes, T. Carvalho, A. Catenazzi, et al. Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science*, 363 (6434):1459–1463, 2019.
- P. H. Scherrer, J. Schou, R. Bush, A. Kosovichev, R. Bogart, J. Hoeksema, Y. Liu, T. Duvall, J. Zhao, C. Schrijver, et al. The helioseismic and magnetic imager (hmi) investigation for the solar dynamics observatory (sdo). *Solar Physics*, 275(1):207–227, 2012.
- J. Schou, P. H. Scherrer, R. I. Bush, R. Wachter, S. Couvidat, M. C. Rabello-Soares, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, D. J. Akin, B. A. Allard, J. W. Miles, R. Rairden, R. A. Shine, T. D. Tarbell, A. M. Title, C. J. Wolfson, D. F. Elmore, A. A. Norton, and S. Tomczyk. Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO). , 275:229–259, Jan. 2012. doi: 10.1007/s11207-011-9842-2.
- B. S. Schuberth, H.-P. Bunge, and J. Ritsema. Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone? *Geochemistry, Geophysics, Geosystems*, 10(5), 2009a.
- B. S. Schuberth, H.-P. Bunge, G. Steinle-Neumann, C. Moder, and J. Oeser. Thermal versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: High plume excess temperatures in the lowermost mantle. *Geochemistry, Geophysics, Geosystems*, 10(1), 2009b.
- A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.

- F. E. Shamout, T. Zhu, P. Sharma, P. J. Watkinson, and D. A. Clifton. Deep interpretable early warning system for the detection of clinical deterioration. *IEEE journal of biomedical and health informatics*, 24(2):437–446, 2019.
- E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 2016. URL <http://arxiv.org/abs/1605.06211>.
- Y. Shiu, K. Palmer, M. A. Roch, E. Fleishman, X. Liu, E.-M. Nosal, T. Helble, D. Cholewiak, D. Gillespie, and H. Klinck. Deep neural networks for automated detection of marine mammal species. *Scientific reports*, 10(1):1–12, 2020.
- C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019.
- K. Sigloch and M. G. Mihalynuk. Intra-oceanic subduction shaped the assembly of cordilleran north america. *Nature*, 496(7443):50–56, 2013.
- K. Sigloch and G. Nolet. Measuring finite-frequency body-wave amplitudes and traveltimes. *Geophysical Journal International*, 167(1):271–287, 2006.
- K. Sigloch, N. McQuarrie, and G. Nolet. Two-stage subduction history under north america inferred from multiple-frequency tomography. *Nature Geoscience*, 1(7):458–462, 2008.
- J. Smit, R. A. Pozo, J. J. Cusack, K. Nowak, and T. Jones. Using camera traps to study the age–sex structure and behaviour of crop-using elephants *loxodonta africana* in udzungwa mountains national park, tanzania. *Oryx*, 53(2):368–376, 2019.
- J. D. Smith, K. Azizzadenesheli, and Z. E. Ross. Eikonet: Solving the eikonal equation with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

- S. C. Solomon, T. N. Woods, L. V. Didkovsky, J. T. Emmert, and L. Qian. Anomalous low solar extreme-ultraviolet irradiance and thermospheric density during solar minimum. *Geophysical Research Letters*, 37(16), 2010.
- J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.
- S. C. Stähler, K. Sigloch, and T. Nissen-Meyer. Triplicated p-wave measurements for waveform tomography of the mantle transition zone. *Solid Earth*, 3(2): 339–354, 2012.
- S. C. Stähler, M. van Driel, L. Auer, K. Hosseini, K. Sigloch, and T. Nissen-Meyer. Mc kernel: Broadband waveform sensitivity kernels for seismic tomography. In *EGU General Assembly Conference Abstracts*, pages EPSC2016–7020, 2016.
- W. Steffen, W. Broadgate, L. Deutsch, O. Gaffney, and C. Ludwig. The trajectory of the anthropocene: the great acceleration. *The Anthropocene Review*, 2(1): 81–98, 2015.
- S. Sugumar and R. Jayaparvathy. An early warning system for elephant intrusion along the forest border areas. *current science*, pages 1515–1526, 2013.
- A. Sukhovich, S. Bonnieux, Y. Hello, J.-O. Irisson, F. J. Simons, and G. Nolet. Seismic monitoring in the oceans by autonomous floats. *Nature communications*, 6(1):1–6, 2015.
- R. Sukumar. *The living elephants: evolutionary ecology, behaviour, and conservation*. Oxford University Press, 2003.
- T. SunPy Community, S. J. Mumford, S. Christe, D. Pérez-Suárez, J. Ireland, A. Y. Shih, A. R. Inglis, S. Liedtke, R. J. Hewett, F. Mayer, K. Hughitt, N. Freij, T. Meszaros, S. M. Bennett, M. Malocha, J. Evans, A. Agrawal, A. J. Leonard, T. P. Robitaille, B. Mampaey, J. Iván Campos-Rozo, and M. S. Kirk.

- SunPyPython for solar physics. *Computational Science and Discovery*, 8(1): 014009, Jan. 2015. doi: 10.1088/1749-4699/8/1/014009.
- I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning*, 2013a.
- I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013b.
- A. Szenicer, D. F. Fouhey, A. Munoz-Jaramillo, P. J. Wright, R. Thomas, R. Galvez, M. Jin, and M. C. Cheung. A deep learning virtual instrument for monitoring extreme uv solar spectral irradiance. *Science advances*, 5(10): eaaw6548, 2019.
- A. Szenicer, K. Leng, and T. Nissen-Meyer. A complexity-driven framework for waveform tomography with discrete adjoints. *Geophysical Journal International*, 223(2):1247–1264, 2020.
- C. Tape, Q. Liu, A. Maggi, and J. Tromp. Adjoint tomography of the southern california crust. *Science*, 325(5943):988–992, 2009.
- A. Tarantola. Linearized inversion of seismic reflection data. *Geophysical prospecting*, 32(6):998–1015, 1984.
- A. Tarantola, B. Valette, et al. Inverse problems= quest for information. *Journal of geophysics*, 50(1):159–170, 1982.
- A. Tesoniero, K. Leng, M. D. Long, and T. Nissen-Meyer. Full wave sensitivity of sk (k) s phases to arbitrary anisotropy in the upper and lower mantle. *Geophysical Journal International*, 222(1):412–435, 2020.

- M. S. Thorne, K. Leng, S. Pachhai, S. Rost, J. Wicks, and T. Nissen-Meyer. The most parsimonious ultralow-velocity zone distribution from highly anomalous spdk waveforms. *Geochemistry, Geophysics, Geosystems*, 22(1): e2020GC009467, 2021.
- S. Thrastarson, M. van Driel, L. Krischer, C. Boehm, M. Afanasiev, D.-P. Van Herwaarden, and A. Fichtner. Accelerating numerical wave propagation by wave-field adapted meshes. part ii: full-waveform inversion. *Geophysical Journal International*, 221(3):1591–1604, 2020.
- W. K. Tobiska, T. Woods, F. Eparvier, R. Viereck, L. Floyd, D. Bouwer, G. Rottman, and O. White. The solar2000 empirical solar irradiance model and forecast tool. *Journal of Atmospheric and Solar-Terrestrial Physics*, 62(14):1233–1250, 2000.
- J. Tromp and E. Bachmann. Source encoding for adjoint tomography. *Geophysical Journal International*, 218(3):2019–2044, 2019.
- J. Tromp, C. Tape, and Q. Liu. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1): 195–216, 2005.
- J. Tromp, D. Komatitsch, and Q. Liu. Spectral-element and adjoint methods in seismology. *Communications in Computational Physics*, 3(1):1–32, 2008.
- United Nations. World urbanization prospects 2018: Highlights, 2019.
- USGS. Landsat – earth observation satellites, 2016.
- R. D. Van der Hilst, S. Widiyantoro, and E. Engdahl. Evidence for deep mantle circulation from global tomography. *Nature*, 386(6625):578–584, 1997.
- D. G. Van Der Meer, W. Spakman, D. J. Van Hinsbergen, M. L. Amaru, and

- T. H. Torsvik. Towards absolute plate motions constrained by lower-mantle slab remnants. *Nature Geoscience*, 3(1):36–40, 2010.
- R. Van der Voo, W. Spakman, and H. Bijwaard. Mesozoic subducted slabs under siberia. *Nature*, 397(6716):246–249, 1999.
- M. van Driel and T. Nissen-Meyer. Seismic wave propagation in fully anisotropic axisymmetric media. *Geophysical Journal International*, 199(2):880–893, 2014.
- M. van Driel, S. Ceylan, J. F. Clinton, D. Giardini, H. Alemany, A. Allam, D. Ambrois, J. Balestra, B. Banerdt, D. Becker, et al. Preparing for insight: Evaluation of the blind test for martian seismicity. *Seismological Research Letters*, 90(4):1518–1534, 2019.
- M. van Driel, C. Boehm, L. Krischer, and M. Afanasiev. Accelerating numerical wave propagation using wavefield adapted meshes. part i: forward and adjoint modelling. *Geophysical Journal International*, 221(3):1580–1590, 2020.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.
- J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.
- C. Wamsler. *Cities, disaster risk and adaptation*. Routledge, 2014.
- X. Wang, D. F. Fouhey, and A. Gupta. Designing deep networks for surface normal estimation. In *CVPR*, 2015.
- H. P. Warren. Measurements of absolute abundances in solar flares. *The Astrophysical Journal Letters*, 786(1):L2, 2014.

- S. K. Wasser, W. Joseph Clark, O. Drori, E. Stephen Kisamo, C. Mailand, B. Mutayoba, and M. Stephens. Combating the illegal trade in african elephant ivory with dna forensics. *Conservation Biology*, 22(4):1065–1071, 2008.
- R. J. Willemann and D. A. Storchak. Data collection at the international seismological centre. *Seismological Research Letters*, 72(4):440–453, 2001.
- J. D. Wood, C. E. O’Connell-Rodwell, and S. L. Klemperer. Using seismic sensors to detect elephants and other large mammals: a potential census technique. *Journal of Applied Ecology*, pages 587–594, 2005.
- T. Woods, F. Eparvier, R. Hock, A. Jones, D. Woodraska, D. Judge, L. Didkovsky, J. Lean, J. Mariska, H. Warren, et al. Extreme ultraviolet variability experiment (eve) on the solar dynamics observatory (sdo): Overview of science objectives, instrument design, data products, and model developments. *The solar dynamics observatory*, pages 115–143, 2010.
- T. N. Woods, F. G. Eparvier, R. Hock, A. R. Jones, D. Woodraska, D. Judge, L. Didkovsky, J. Lean, J. Mariska, H. Warren, D. McMullin, P. Chamberlin, G. Berthiaume, S. Bailey, T. Fuller-Rowell, J. Sojka, W. K. Tobiska, and R. Viereck. Extreme Ultraviolet Variability Experiment (EVE) on the Solar Dynamics Observatory (SDO): Overview of Science Objectives, Instrument Design, Data Products, and Model Developments. , 275:115–143, Jan. 2012. doi: 10.1007/s11207-009-9487-6.
- P. J. Wright, M. Cheung, R. Thomas, R. Galvez, A. Szenicer, M. Jin, A. Muñoz-Jaramillo, and D. Fouhey. Deepem: Demonstrating a deep learning approach to dem inversion. *Zenodo*.
- P. J. Wright, M. C. M. Cheung, R. Thomas, R. Galvez, A. Szenicer, M. Jin, A. Muñoz-Jaramillo, and D. Fouhey. Deepem : A deep learning approach for dem inversion. In M. G. Bobra and

. URL <https://helioml.github.io/HelioML>.

C. Zaroli, S. Lambotte, and J.-J. L  v  que. Joint inversion of normal-mode and finite-frequency s-wave data using an irregular tomographic grid. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 203(3): 1665–1681, 2015.

M. A. Zeder. The origins of agriculture in the near east. *Current Anthropology*, 52(S4):S221–S235, 2011.

D. Zhao. Global tomographic images of mantle plumes and subducting slabs: insight into deep earth dynamics. *Physics of the Earth and Planetary Interiors*, 146(1-2):3–34, 2004.

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.

W. Zhu, S. M. Mousavi, and G. Berosa. Seismic signal augmentation to improve generalization of deep neural networks. *Advances in Geophysics*, 61:151–177, 2020.