

Deep Learning, Shallow Dips

Transit light curves have never been so trendy

Mario Morvan

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Physics and Astronomy
University College London

September 2022

I, Mario Morvan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

At the crossroad between photometry and time-domain astronomy, light curves are invaluable data objects to study distant events and sources of light even when they can not be spatially resolved. In particular, the field of exoplanet sciences has tremendously benefited from acquired stellar light curves to detect and characterise a majority of the outer worlds that we know today. Yet, their analysis is challenged by the astrophysical and instrumental noise often diluting the signals of interest. For instance, the detection of *shallow dips* caused by transiting exoplanets in stellar light curves typically require a precision of the order of 1 ppm to 100 ppm in units of stellar flux, and their very study directly depends upon our capacity to correct for instrumental and stellar *trends*.

The increasing number of light curves acquired from space and ground-based telescopes—of the order of billions—opens up the possibility for global, efficient, automated processing algorithms to replace individual, parametric and hard-coded ones. Luckily, the field of deep learning is also progressing fast, revolutionising time series problems and applications. This reinforces the incentive to develop data-driven approaches hand-in-hand with existing scientific models and expertise.

With the study of exoplanetary transits in focus, I developed automated approaches to learn and correct for the time-correlated noise in and across light curves. In particular, I present (i) a deep recurrent model trained via a forecasting objective to detrend individual transit light curves (e.g. from the Spitzer space telescope); (ii) the power of a Transformer-based model leveraging whole datasets of light curves (e.g. from large transit surveys) to learn the trend via a masked objective; (iii) a hybrid and flexible framework to combine neural networks with transit physics.

Impact Statement

Broadly, this work contributes to the fields of time series analysis and exoplanet sciences. It contains a variety of novel ideas and techniques, all relating to the underlying idea of *automatically learning the temporal noise structure in order to better identify the (planetary) signal(s)*. While the technical frame of this thesis might be defined as "self-supervised sequential deep learning applied to astronomical time series", the prime task at hand is the precise measurement of exoplanetary transits in stellar light curves.

Chapters 3 and 4 present general and novel ideas to detrend exoplanetary data with examples of applications to Spitzer and TESS light curves. As the problem of correcting time-correlated noise in transit light curves exists for any instrument and star, such techniques might be useful to analyse data from other current and future exoplanet space missions such as TESS, JWST, PLATO, or Ariel. Chapter 5 presents a pioneering framework paving the way for more efficient optimisation methods for transit physics, methods which will in turn play a crucial role for large-scale studies of transiting planets as we are discovering ever more of them.

Beyond research in exoplanet sciences, this work has been very well received by the machine learning community, as these excerpts from different reviewers witness:

- "The study has made clear advancements in the study of machine learning on stellar light curves on multiple fronts."
- "A method which is creative, simple, elegant, and widely applicable."
- "Potential applications of this work extend to time series data in other fields, including planetary science, the geosciences, geophysics, etc."

Time series data is indeed ubiquitous, and other applications such as health, energy, economy, industry... also have large datasets of noisy, complex, correlated time series. The problems of detrending, denoising, outlier detection, signal detection, decomposing and classifying light curves directly relate to time series tasks from other fields as well. It is also worth noting that the successes of machine learning on time series appear lagging behind the progress on images or natural language. Consequently, it is understandable that successful applications of innovative self-supervised models to time series are welcome at this stage by the machine learning community as potential drivers of change for time series analysis and its applications. As another example of machine learning applied to time series, [Appendix C](#) features an impactful application in the field of wastewater-based epidemiology produced during my 6-month placement at the Department of Health and Social Care.

Substantial progress is expected in the field machine learning applied to astronomical time series. If this work can contribute—even by a small step—towards the next generation of data-based techniques to help boost our physical understanding of exoplanets and stars, my objective will be considered met.

Finally, I strongly believe in the open science and free software movements, as key ingredients for a more accessible, democratic, collaborative and arguably effective science. All the data used for this work, as well as all the code and publications produced are publicly available.

Acknowledgements

It takes more than a PhD student to complete a PhD degree and write up a PhD thesis. I would like to thank everyone who has helped me get there directly or indirectly.

My eternal gratitude goes to Dr. Ingo Waldmann, my main supervisor, for his incredibly positive and instructive guidance continuously during this journey.

I would hardly have got anywhere near where I am now without Dr. Angelos Tsiaras, Dr. Nikolaos Nikolaou, Prof. Jan-Peter Muller and Prof. Giovanna Tinetti from whom I have learned and been inspired so much.

A million thanks to all UCL colleagues who have contributed to making these four years a fun and enlightening journey: Dr. Gordon Yip, Dr. Quentin Changeat, Dr. Billy Edwards... and the whole adorable exoplanet team.

If my placement at the EMHP/DHSC¹ has been so interesting and inspiring, that was undoubtedly thanks to my placement supervisor Jasmine Grimsley. A huge thanks to her and the whole team there.

I have been immensely fortunate to work with all of these people throughout these four years.

The final word of gratitude goes to my family, especially my sister, mum and dad, who have been most encouraging and inspiring in all of my life choices.

This being said, I realise that I should probably also acknowledge the artificial intelligence that wrote this thesis for me. It has been most helpful, surprisingly effective at reading my thoughts and explaining my work. It even picked up my own writing style—when reading this thesis, it really feels like I did write these words²!

¹Environmental Monitoring For Health Protection in the Department of Health and Social Care

²Note: this thesis was submitted just two months before the public release of ChatGPT in November 2022, making the use of AI chatbots widely available and this joke closer to believable.

Contents

List of Figures	11
List of Tables	14
1 Introduction	15
1.1 Astronomical Photometry	16
1.2 The Exoplanet Revolution	18
1.2.1 From planets to exoplanets	18
1.2.2 The transit method	20
1.2.3 From detection to characterisation	23
1.2.4 The future of exoplanet discoveries	26
1.3 The Machine Learning Revolution	31
1.3.1 Genesis	31
1.3.2 Tasks and paradigms	33
1.3.3 Machine learning in astronomy and exoplanet sciences	34
1.3.4 Challenges and opportunities	36
1.4 Thesis Outline	37
2 Light Curves	39
2.1 Formalism	39
2.2 Transit Light Curves	44
2.2.1 Planetary dynamics	44
2.2.2 Shape calculation	46
2.2.3 Astrophysical assumptions	47

2.3	Measurements	49
2.3.1	Large transit surveys	49
2.3.2	Characterisation missions	50
2.3.3	Instrumental systematics	53
2.4	Stellar Variability	55
2.4.1	Types of variables	55
2.4.2	Synergies with exoplanets	60
2.5	Detrending	61
2.5.1	A milestone in light curve processing	61
2.5.2	Detrending exoplanetary light curves	62
2.5.3	Let machines learn the trend	63
3	Autoregressive Light Curve Modelling with Recurrent Neural Networks	65
3.1	Introduction	66
3.2	Recurrent Neural Networks	68
3.3	The Transit Light Curve Detrending LSTM Model	71
3.3.1	Model description	71
3.3.2	Application to transit light curves	74
3.4	Applications	75
3.4.1	Preprocessing	75
3.4.2	Tuning & testing	76
3.4.3	HD 189733 b	79
3.4.4	WASP-96 b	83
3.4.5	WASP-17 b	85
3.5	Summary and Discussion	88
4	Self-supervised Representations of Light Curves with Transformers	90
4.1	Introduction	91
4.2	Background: the Transformer Model	93
4.2.1	Attention mechanisms	93
4.2.2	The Transformer model	95

4.3	Methodology	97
4.3.1	Problem formulation	97
4.3.2	Denoising Time Series Transformer	97
4.3.3	Masking strategy	99
4.4	Experiments	100
4.4.1	Dataset	100
4.4.2	Training and evaluation	101
4.4.3	Results	102
4.4.4	1D attention maps	103
4.5	Summary and Discussion	105
5	A Hybrid Physics & Deep Learning Framework for Transit Inference	106
5.1	Introduction	107
5.2	PyLightcurve-torch	109
5.2.1	Code design	109
5.2.2	Performance	110
5.3	Applications	112
5.3.1	Generation	112
5.3.2	Gradient-based optimisation	112
5.3.3	Combination with neural networks	113
5.3.4	Regression experiment	114
5.4	Summary and Discussion	119
6	Conclusions	120
6.1	Summary	120
6.2	Future Work	121
6.3	Using Artificial Intelligence Humanly	122
	Appendices	124
A	Acronyms & Notations	124

B	Ariel Machine Learning Challenges 2019 and 2021	128
B.1	Introduction: Why Organising Machine Learning Challenges? . . .	128
B.2	Ariel Machine Learning Challenge 2019	129
B.2.1	Introduction	129
B.2.2	The Challenge	132
B.2.3	Top-5 Solutions	142
B.2.4	What the winning models teach us	155
B.2.5	Conclusions	157
B.3	Ariel Machine Learning Challenge 2021	158
B.3.1	Data	159
B.3.2	Tasks and application scenarios	162
B.3.3	Metrics	163
B.3.4	Baseline and code available	164
B.3.5	Results	165
C	Quantifying Prevalence of COVID-19 from Wastewater Data	166
C.1	Overview	167
C.2	Introduction	167
C.3	Results	170
C.4	Discussion	176
C.5	Methods	180
C.5.1	Data	180
C.5.2	Models	184
	Bibliography	188

List of Figures

1.1	Ptolemy's star catalogue	17
1.2	Transit of Venus 1874	21
1.3	Schematic transit light curve	22
1.4	Number of detected planets over time	24
1.5	Distribution of stellar masses and exoplanetary radii	25
1.6	Joint analysis of 30 exoplanet atmospheres with HST data	27
1.7	Joint analysis of 25 exoplanet atmospheres with HST and Spitzer data	28
1.8	Diagram of a single-layer artificial neural network	32
1.9	Use of machine learning in astronomical publications over time	35
2.1	Simulated time series with trend, white noise and transient events	41
2.2	Diagram of a Keplerian orbit in 3D	45
2.3	Schematic view of the star surface during transit	46
2.4	TESS detector housing and assembly	51
2.5	Field-of-views of Spitzer detectors projected onto the sky	52
2.6	Examples of systematics from Kepler and TESS	56
2.7	Examples of systematics from Spitzer/IRAC	57
2.8	Light curve photometry of 7 variable stars	59
2.9	Diagram of common problems on transit light curves	62
3.1	Diagram of a recurrent neural network	69
3.2	Diagram of an LSTM cell	70
3.3	TLCD-LSTM model overview	72
3.4	TLCD-LSTM test predictions on out-of-transit ranges	77

3.5	Summary of HD 189733 b data used as model input	80
3.6	Plot of the best systematics predictions for HD 189733 b light curves	80
3.7	Transit fits, residuals and ACF for six HD 189733 b light curves	81
3.8	Comparison of transit depths between this work and Agol et al. (2010)	82
3.9	Fits for additional parameters for six transits of HD 189733 b	82
3.10	Fitted Spitzer transits of WASP-96 b for IRAC channels 1 and 2	84
3.11	Detrended and residual IRAC light curves of WASP-17 b	87
3.12	Transmission spectrum of WASP-17 b	87
4.1	Multi-head scaled dot-product attention mechanism	94
4.2	Overview of the Transformer model architecture	96
4.3	Schematic overview of the DTST	98
4.4	Prediction summary for four different test stars	104
4.5	Visualisation of 1D attention maps on 16 input light curves	105
5.1	Summary of computational efficiency of PyLightcurve-torch	111
5.2	Schematic views of hybrid deep learning architectures	114
5.3	Learning and validation curves for the regression experiment	117
5.4	Simulated light curves and predictions	118
B.1	Examples of simulations for two wavelength channels	136
B.2	The DL model architecture proposed by the Major Tom team	147
B.3	Preprocessing of the light curves by the BV Labs team (ranked 3rd)	148
B.4	Regression pipeline of IWF-KNOW (ranked 4th)	150
B.5	Data preprocessing performed by the TU Dortmund University team	153
B.6	The first 3 layers of TU Dortmund University team's model	154
B.7	Stellar noise simulations	159
B.8	Example of typical instrumental noise in HST/WFC3 data	160
B.9	Example of simulated Ariel atmospheric transmission spectra	161
B.10	Scaling factor as a function of wavelength	164
C.1	Geographical summary of the wastewater data	171

C.2	Comparison of the outputs from the phenomenological model to CIS	173
C.3	GBRT model performance across regions of England	173
C.4	Predictions versus targets of SARS-CoV-2 CIS prevalence	175
C.5	Lead and lag analysis of WW compared to clinical testing	175

List of Tables

2.1	Summary of datasets from wide-field photometric surveys	49
3.1	Summary of training and prediction input/output modes	72
3.2	TLCD-LSTM model's hyperparameters	78
3.3	Performance summary on out-of-transit ranges	79
3.4	Fitted physical parameters for the 6 HD 189733 b transits	81
3.5	Details of the WASP-96 system used in this study	83
3.6	Orbital parameters of WASP-96 b across different datasets	84
3.7	Stellar and planetary parameters for WASP-17 b used in this study .	86
4.1	Fixed hyperparameters of the DTST	99
4.2	Denoising performance on 3168 TESS light curves from Sector 1 . .	102
5.1	Summary of parameters used in the regression experiment	115
A.1	List of acronyms	124
A.2	List of notations	126
B.1	Final leaderboard of the 2019 competition	143
B.2	CNN architecture used in the winning solution	145

Chapter 1

Introduction

"I know that I am mortal by nature and ephemeral, but when I trace at my pleasure the windings to and fro of the heavenly bodies, I no longer touch Earth with my feet. I stand in the presence of Zeus himself and take my fill of ambrosia."

Claudius Ptolemy, *Almagest*

Exoplanet sciences and deep learning—the two main pillars of this thesis—are two very recent and fashionable fields. As news media are well aware, exoplanets are among the best places to search for extra-terrestrial intelligence (SETI), whereas deep learning is currently viewed as one of the most promising ingredients for artificial general intelligence (AGI). And to add an extra dose of fashion, progress in SETI is likely to be supported by techniques of artificial intelligence. However, to the great disappointment of the reader, the present work is not directly concerned with neither SETI nor AGI endeavours¹. There are nonetheless many other reasons why time-domain astronomical photometry, exoplanet sciences and deep learning are worthy of interest. The following introduction aims at discussing the progress, challenges and relations of and between these fields, thus serving as a motivational background for the work presented in the subsequent chapters.

¹These are left for future work.

1.1 Astronomical Photometry

Photometry—the science of the measurement of light—is one of the most ancient and powerful tool to study distant astronomical objects. In fact, it is only fairly recent in the history of astronomy that other types of space messengers than light have been used, with the first detection of cosmic rays in 1912 ([Hess, 1912](#), resulting in a Nobel Prize in Physics awarded to Hess in 1936), cosmic neutrinos ([Davis et al., 1968](#), resulting in half a Nobel Prize in Physics awarded to Davis and Koshiba in 2002) and the first detection of gravitational waves emitted by two merging black holes ([Abbott et al., 2016](#), resulting in a Nobel Prize in Physics awarded to Weiss, Barish and Thorne in 2017). If neutrino and gravitational waves astronomies can peek at regions and events otherwise hidden from our telescopes’ view, electromagnetic waves still remain our prime messenger to monitor distant objects and try to make sense of our Universe.

First records of attempts to measure the apparent brightness of stars date back to Greek astronomer Hipparchus, during the second century BC. Hipparchus’ work was then used as a basis by Claudius Ptolemy to produce a catalogue of 1022 stars classified according to what he defined as their *magnitude* (see a manuscript extract from Ptolemy’s star catalogue on [Figure 1.1](#)). Although the precise definition of magnitude has evolved since then, we are still using a similar scale to quantify the apparent brightness of stars, whereby brightest stars are of the first magnitude while the faintest stars one can see with unaided eyes are of sixth magnitude.

This was the birth of visual photometry, i.e. the direct measurement of the light radiated by astronomical objects with the human eye. Visual photometry prevailed until photometric instruments were invented nearly two millennia later, thus providing much more accurate measurement of brightness, and not any more in relative terms only. Photographic plates have been the best way to record light and map the sky from the end of the 19th century, and it is only a few decades ago that charge-coupled devices (CCDs) replaced photographic plates, thus revolutionising photometry by enabling digitally recorded photometric measurements of unprecedented accuracy.

οενττωαριστερεα ιονισοομνηρω	ΥΑΡΟ	ΑΙΣ	ΝΟ	ΟΙΣ	Ε
τωναενττωαριστερεακνημνη βονοτιωτερος	ΥΑΡΟ	ΗΓ	ΝΟ	Ι	Ε
οβορειωτεροςαυτωνυποτογονυ	ΥΑΡΟ	ΖΔΓ	ΝΟ	Ο	Ε
τ'ενττωαριστερεα του υδατος αποτχειροςονρ	ΥΑΡΟ	ΙΕ	ΒΟ	Β	Α
οεχομενοςεικνοτου του προειρημενου	ΥΑΡΟ	ΙΑΛΓ	ΒΟ	ΟΣ	Α
οτουτ'εχομενοςμετατηνκαμνη	ΥΑΡΟ	ΙΖΕ	ΠΟ	ΑΣ	Α
οεπ'τουτ'ωεπομενος	ΥΑΡΟ	Κ	ΝΟ	Ι	Α
οτουτ'ενκαμνηπομεσνμβριας	ΥΑΡΟ	ΚΛ	ΝΟ	ΑΙΣ	Α
τ'απομεσνμβριας αυτου βοβορειωτερος	ΥΑΡΟ	ΙΘ	ΝΟ	ΕΛ	Α
ονοτιωτεροςαυτωνυπο	ΥΑΡΟ	ΙΘΛΓ	ΝΟ	ΑΣ	Α
οαεστωαυτωνπροσμεσνμβριας	ΥΑΡΟ	ΚΛΓ	ΝΟ	ΗΔ	Ε
τωνμεταυτωνεσυνεχωνοπροηγουμενος	ΥΑΡΟ	ΚΒΓ	ΝΟ	ΙΑ	Ε
οεπομενοςαυτων	ΥΑΡΟ	ΚΓΣ	ΝΟ	ΙΛΓ	Ε
τ'ενττωαριστερεα του υδατος	ΥΑΡΟ	ΚΑΙΣ	ΝΟ	ΙΑ	Ε
οεπομενοςαυτων	ΥΑΡΟ	ΚΒΣ	ΝΟ	ΙΑΛΑ	Ε
ομοιωστωνεφεζηςγοβορειος	ΥΑΡΟ	ΚΓΣ	ΝΟ	ΙΕΙΣ	Ε
ονοτιωτεροςτ'τριων	ΥΑΡΟ	ΙΖ	ΝΟ	ΙΑΣ	Α
οεπομενοςαυτων	ΥΑΡΟ	ΙΗΓ	ΝΟ	ΙΕΛΑ	Α
τ'ενττωαριστερεα του υδατος	ΥΑΡΟ	ΙΖΛ	ΝΟ	ΙΕ	Α
τωνλοιπωνβονοτιωτερος	ΥΑΡΟ	ΙΑΛΓ	ΝΟ	ΙΑΛΔ	Α
οβορειωτεροςαυτων	ΥΑΡΟ	ΙΕΓ	ΝΟ	ΙΕΓ	Α
οεσχατοςυδατοςκαιεπιτ'στοματοςτ'ηιχθους	ΥΑΡΟ	ΙΙΣ	ΝΟ	ΙΔ	Α
αστερ'εσμ'ων αμεγεδα γα ιηε'ΙΣ Α	ΥΑΡΟ	Ζ	ΝΟ	ΚΓ	Α
οιπεριτονυδαροχοοημορφωτοι					
τ'επομ'ηνκαμνητου υδατοςγονηγουμεν	ΥΑΡΟ	ΚΣΙ	ΝΟ	ΙΕΛ	Α
τ'λοιπωνβοβορειωτερος	ΥΑΡΟ	ΚΒΙ	ΝΟ	ΙΑΙΣ	Α
ονοτιωτεροςαυτων	ΥΑΡΟ	ΚΒ	ΝΟ	ΙΗΔ	Α
αστερε'εφεζης					
ιχθουναστεριςμος	ΥΑΡΟ	ΚΑΙΣ	ΒΟ	ΟΑ	Α
οενττωαριστερεα του προηγουμενιχθους	ΥΑΡΟ	ΚΑΣ	ΒΟ	ΖΛ	Α
τ'ενττωαριστερεα του βονοτιωτερος	ΥΑΡΟ	ΚΣ	ΒΟ	ΟΓ	Α
οβορειωτεροςαυτων	ΥΑΡΟ	ΚΗΣ	ΒΟ	ΟΛ	Α
τωνενττωνωτωνβονοτιω	ΙΧΘΥ	ΟΓ	ΒΟ	ΖΛ	Α
οεπομενοςαυτων	ΥΑΡΟ	ΚΣ	ΒΟ	ΑΛ	Α
τ'ενττωαριστερεα του βονοτιω	ΥΑΡΟ	ΚΘΙ	ΒΟ	Π	Α
οεπομενοςαυτων	ΙΧΘΥ	Σ	ΒΟ	ΣΓ	Α
οενττωαριστερεα του υδατος	ΙΧΘΥ	ΙΑ	ΒΟ	ΕΛΑ	Α
τ'κατατολινοναυτουονρωτοσαποτ'δυνας	ΙΧΘΥ	Π	ΒΟ	ΠΛΑ	Α
οεπομενοςαυτων	ΙΧΘΥ	ΙΖΣ	ΒΟ	ΒΑ	Α
τ'εφεζηςαμνητωνγονηγουμεν	ΙΧΘΥ	ΚΣ	ΒΟ	ΑΣ	Α
οεπομενοςαυτων	ΙΧΘΥ	ΚΓ	ΝΟ	Σ	Α
τ'υπαυτ'ενκαμνημνημ'ωνβονοτιωτερος	ΙΧΘΥ	ΚΒΙ	ΝΟ	Β	Α
ονοτιωτεροςαυτων	ΙΧΘΥ	ΚΓ	ΝΟ	Ε	Α

Figure 1.1: The oldest manuscript of Ptolemy's Almagest dating from the ninth century, written in ancient Greek. It contains named stars with positions and magnitudes grouped by constellations, marking the onset of astronomical photometry (public domain).

Photometry is also one of the most useful tool to study astronomical objects whose luminosity vary in human timescales, i.e. from seconds to decades. Indeed, by measuring an object's brightness at different points in time, one effectively builds a *light curve* whose variability patterns often contain invaluable information to understand the underlying processes responsible for the emission, reflection, obstruction or deflection² of light. Light curves measurements thus enable studies of our Solar System through asteroids, comets, planets and minor planets; but also of our galactic and extragalactic surroundings through e.g. variable stars, supernovae, active galactic nuclei, etc.

Signals sought in light curves are greatly varied in nature, sometimes regular and periodic (e.g. some types of variable stars) and some other times transient and abrupt (e.g. supernovae). Although various sources of astrophysical variability are entangled, the main hurdle to study and interpret light curves is often the quality of the measurement itself. Indeed, there are many causes of artificial variability in measured light curves, and the breadth of signals we are able to study ineluctably relies on our ability to identify and correct for these instrumental artefacts.

With the multiplication of increasingly ambitious photometric missions and surveys over the past few decades, the volumes of light curves currently available have clearly become unmanageable for individual treatments. The number of light curves openly available is of the order of billions, with orders of petabytes of associated data. While individual surveys are using automated pipelines to process light curves prior to scientific use, the scientific potential in designing models trained on whole datasets of light curves using artificial intelligence is already considerable today and is only expected to grow in the future.

1.2 The Exoplanet Revolution

1.2.1 From planets to exoplanets

For a long time the six innermost planets of the Solar System were the only planets we knew of. Indeed, despite various early reports of their existence (Uranus was

²Typically via gravitational lensing, see mention of *microlensing* in Section 1.2.1

referenced as a star in Ptolemy's *Almagest*), Uranus and Neptune's apparent stillness in the sky caused them to be considered as stars up until 1781 ([Herschel and Watson, 1781](#)) and 1846 ([Kollerstrom, 2006](#)) respectively. Interestingly enough, Neptune's first claim of existence as a planet was motivated indirectly by the observation of perturbations in the orbit of Uranus in 1821 ([Jarrell, 2007](#)).

The existence of planets beyond our Solar System was suggested by Giordanni Bruno in his work "De l'Infinito, Universo e Mondi" in 1584 and later by others (e.g. [Newton et al., 1846](#); [Struve, 1952](#)), but it remained speculation until the end of the 20th century when an exoplanet orbiting the pulsar PSR B1257+12 was detected in 1992 ([Wolszczan and Frail, 1992](#)) and the first exoplanet 51 Pegasi b orbiting a main-sequence star was detected in 1995 ([Mayor and Queloz, 1995](#)). Pulsar planets are now known to be very rare, unlike planets around main-sequence stars now thought to be very common. The discovery of 51 Pegasi b is thus considered as the emblematic starting point of the exoplanet revolution now under way and its main contributors Mayor and Queloz were consequently awarded with a Nobel Prize in physics in 2019.

As for Neptune, the claim of existence of these exoplanets was made from observed perturbations to an expected signal, in one case the pulsar's constant period and in the other the host star's constant radial velocity. As well known from Newton's laws of dynamics, planets cause their host star to orbit around the system's barycentre, and this is precisely this movement which was detected via Doppler effect and then attributed to the presence of orbiting planets. In addition to these two techniques called *pulsar timing* and *radial velocities*, there are several other proposed methods to detect exoplanets, the main ones being *direct imaging*, *transits*, *microlensing* and *astrometry*. Here is a brief summary of these methods in the context of exoplanet detection:

- pulsar timing: measurement of the periodic variations in pulsar period attributed to gravitational influence from a companion planet (see e.g. [Bailes et al., 1991](#));
- radial velocities: measurement of the periodic variations of a star's radial

velocity—through the blueshift and redshift of its spectral lines—attributed to the gravitational influence from a companion planet (see e.g. [Mayor and Queloz, 1995](#));

- astrometry: measurement of the periodic angular variations of a star position attributed to the gravitational influence from a companion planet (see e.g. [Gliese, 1982](#));
- direct imaging: measurement of a point source of light—emitted or reflected—attributed to a planet (see e.g. [Claudi, 2016](#));
- microlensing: measurement of the transient amplification of a background light source attributed to a foreground planet, according to general relativity ([Paczynski, 1986](#));
- transit: partial obscuration of a star attributed to the passage of a companion planet in front of it;

The most effective method so far at detecting and studying exoplanets has been the transit method, and it is indeed the focus of this work.

1.2.2 The transit method

The first planet ever observed transiting across the Sun was Mercury in 1631 ([Gassendi, 1632](#)), following its prediction by Johannes Kepler published the year before. Not only did this observation serve as a confirmation of Kepler's laws of motion, but it also allowed correcting orbital elements of Mercury and its apparent diameter, thereby "supplying for the first time an indisputable measurement of the apparent magnitude of a planetary disc" ([Van Helden, 1976](#)). The only other Solar System planet that can be seen transiting from the Earth is Venus. The transit of Venus was observed for the first time in 1639, and Figure 1.2 shows a picture from a later event taken in 1874 from Japan by Pierre Janssen.

More than two centuries later, it was suggested that the same phenomenon could in principle be observed when exoplanets eclipse their host star. Lardner indeed claimed that "periodical disappearance or total obscuration of stars may arise

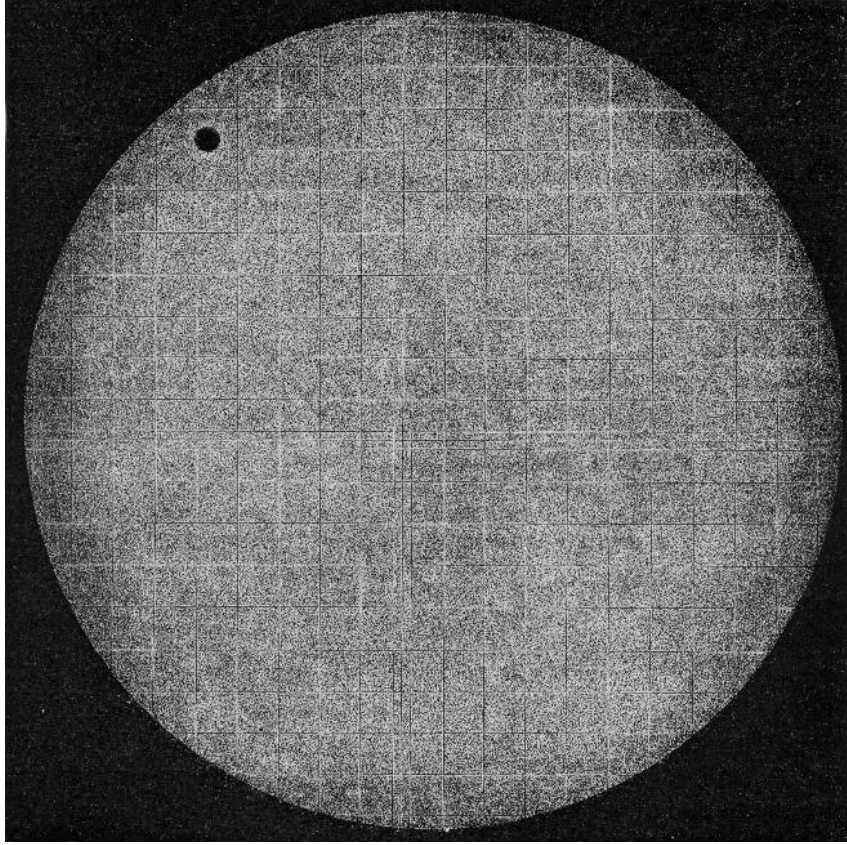


Figure 1.2: Picture of Venus transiting across the Sun, observed from Japan in 1874.
(public domain)

from transits of the star by its attending planets" (Lardner, 1858), with the idea later developed by Struve (1952); Rosenblatt (1971); Borucki and Summers (1984); Borucki et al. (1985).

However, it is only in 1999 that two teams of scientists independently detected the transit of an exoplanet for the first time (Henry et al., 1999, 2000; Charbonneau et al., 2000). These milestone detections made with ground-based 2 m-telescopes constrained the parameters of the planet HD 209458 b previously known from radial velocity measurements. Indeed, the depth of the relative decrease in flux ΔF measured during transit tells us about the relative ratio of stellar and planetary radii: $\delta = \frac{\Delta F}{F} \approx \frac{R_P^2}{R_*^2}$, giving access to the planetary radius $R_P = 1.42 \pm 0.10 R_J$ (in Jupiter radii) since the stellar radius R_* was known from other methods. In addition, the determination of the inclination angle i (angle between system's orbit and the plane perpendicular to the line of sight) enabled by transit observations constrained the

mass of the planet to be $M_P = 0.62 \pm 0.05 M_J$ (in Jupiter masses). Because of its Jupiter-like size and proximity to its star, HD 209458 b would later be named a "hot Jupiter", which are among the most probable and easiest planets to find using either transit or radial velocities methods.

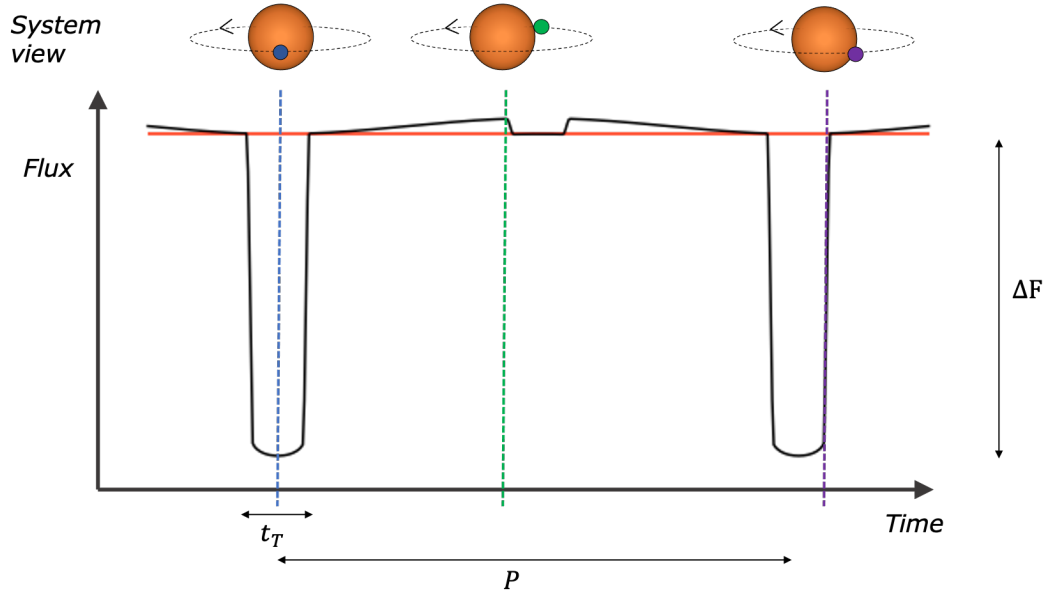


Figure 1.3: Schematic view of a light curve for a transiting exoplanetary system, showing two transits and one eclipse. The horizontal red line denotes the stellar flux in absence of a planet, the blue, green and purple lines point at specific times at mid-transit, eclipse first contact, and transit egress respectively. The top three diagrams show views of the system corresponding to these three different times.

Given the far greater distance of all stars from us compared to the Sun's, it is impossible with today's technology to directly image exoplanets during transit (see [van Belle et al. 2012](#) for a discussion on the prospect of directly imaging exoplanet transits), and as of today Solar System planets remain the only ones whose transits have been directly imaged. Wherever the *transit method* is mentioned throughout this work, it should be understood that we are referring to the study of *photometric transits* rather than directly imaged transits. The transit method relies on the precise analysis of stellar light curves to detect transits and solve the inverse problem of inferring planetary and orbital parameters from the data given some assumptions about the geometry and dynamics of the system. This is generally done by fitting a transit model to an observed light curve containing one or multiple transit events

(see Figure 1.3), resulting in the determination of the following key parameters: the ratio of planetary and stellar radii R_P/R_* ; the ratio of orbital semi-major axis over stellar radius a/R_* ; the impact parameter b and the orbital period P .

From there, it should already be apparent that our understanding of exoplanets directly depends upon our knowledge of their host stars. Indeed, given the above output parameters of the transit fit, estimates for the planetary radius R_P and orbital parameters a and i will only be as precise as the stellar radius R_* . Without knowledge of R_* and therefore of R_P , an exoplanet's transit shape might even well be mistaken for the transit of a stellar companion. Furthermore, the transit fit itself would be affected by any unaccounted blended sources or stellar variability detectable on timescales of the event.

In addition to the observation of transits, it is in some cases also possible to detect the occultation of planets when they pass behind their star with respect to us, an event named *secondary eclipses*, and to be distinguished from the *primary eclipses* (i.e. transits). In that case no stellar light is blocked as in the case of transit, but as the planet is occulted by its star, its emitted or reflected light will be hidden from us during the eclipse, thus producing a decrease in brightness also measurable in the light curve. The first such detections of planetary emission with a secondary eclipse were made by [Deming et al. \(2005\)](#) and [Charbonneau et al. \(2005\)](#).

1.2.3 From detection to characterisation

To date, more than 5000 planets have been confirmed. Most of those have only been detected through indirect methods, with only a few dozens of directly imaged planets which have been confirmed. By far, the transit method has been the most effective to detect planets to date, in particular owing to the transit surveys carried out through the CoRoT (Convection, Rotation and planetary Transits, [Baglin, 2003](#)), Kepler ([Borucki et al., 2010](#)) and TESS (Transiting Exoplanet Survey Satellite [Ricker et al., 2015](#)) missions. The Kepler survey alone has yielded nearly 3000 planets, while the TESS mission is currently finding thousands of new candidates to be confirmed.

This has enabled to perform statistical studies of exoplanet demographics, and has already yielded several important findings. First, we now know that forming

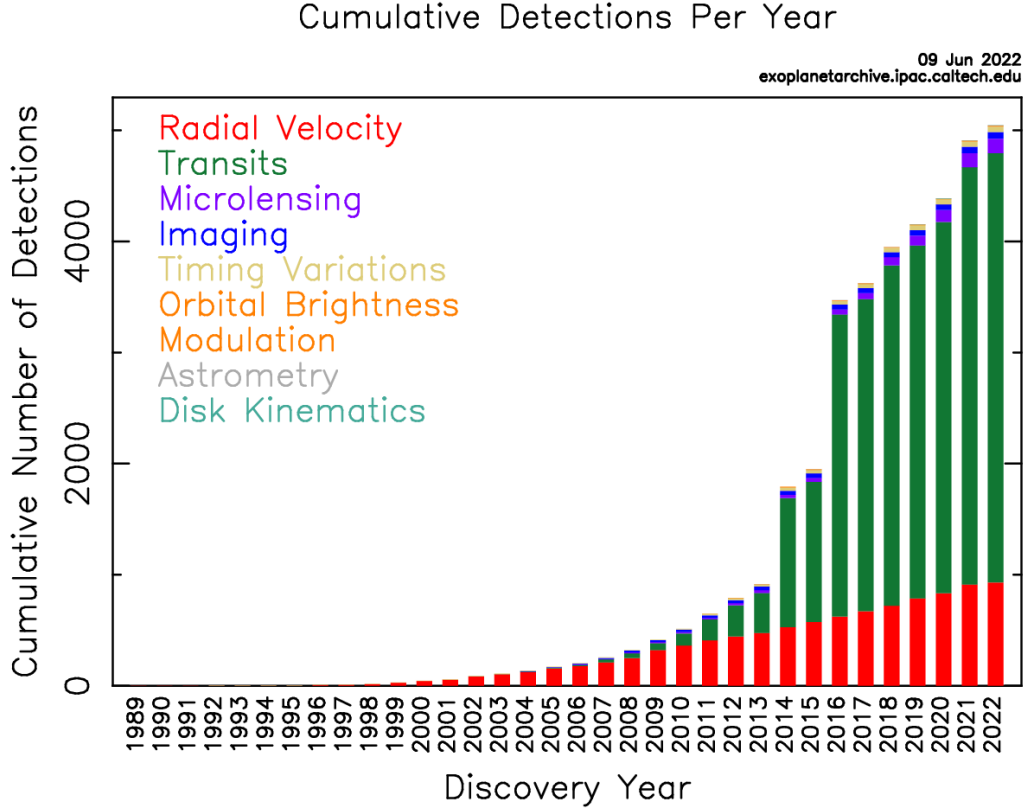


Figure 1.4: Timeline of cumulative number of detected exoplanets by method. Figure from [the Nasa Exoplanet Archive](https://exoplanetarchive.ipac.caltech.edu).

and retaining planets is common among stars, with an estimated total number of planets larger than the population of stars in our galaxy (Cassan et al., 2012). A rich diversity of orbital configurations and planetary parameters was observed, with distant systems very different to the Solar System, starting with the first planet 51 Pegasi b being a hot Jupiter. Small planets ($< 3.5 R_{\oplus}$), albeit harder to detect, appear to be much more common than giant planets ($> 3.5 R_{\oplus}$) (Howard et al., 2012). Combined with stellar data from the Gaia mission (Collaboration, 2016), it was later discovered that the distribution of planetary radii is actually bimodal (Fulton et al., 2017; Fulton and Petigura, 2018; Berger et al., 2020), separating the small planets into two sub-categories called *sub-Neptunes* ($1.8 R_{\oplus} < R_p < 3.5 R_{\oplus}$) and *super-Earths* ($R_p < 1.8 R_{\oplus}$), although no example from these statistically predominant categories exist in the Solar System. The distribution of radii for Kepler planets is shown on Figure 1.5, as a function of the stellar mass computed from Gaia data, showing the *radius valley* somewhere between 1 and $2 R_{\oplus}$. The main hypotheses

to explain the radius valley are the photoevaporation effect (Vidal-Madjar et al., 2004; Linsky et al., 2010) and core-powered mass loss (e.g. Gupta and Schlichting, 2020). Photoevaporation refers to the evaporation of the outer gassy layers of a planet caused by X-ray or extreme ultraviolet radiation, while core-powered mass loss refers to the cooling luminosity of a planet's core.

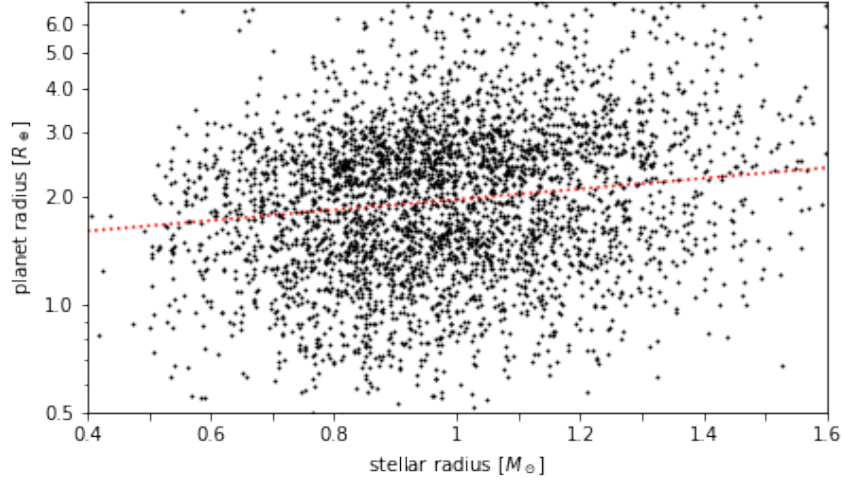


Figure 1.5: Distribution of stellar masses and exoplanetary radii with data from Berger et al. (2020). A linear boundary (red dotted line) can be traced between the two populations (upper and lower) of sub-Neptunes and super-Earths separated by the so-called radius valley.

Characterisation of exoplanets through transit observations does not stop at radius or density estimates, and it is indeed possible to obtain information about their atmospheres by measuring the spectra transmitted or emitted by exoplanets. This is done by using spectroscopic or multi-band photometric transit observations, following a principle laid out by Angel et al. (1986). In short, the mix of atoms and molecules present in exoplanets' atmospheres absorbs, emits or scatters light differently at different wavelengths, creating an apparent variation of the radius depending on the observed wavelength. The first detection of the atmospheric signature of an exoplanet was achieved by Charbonneau et al. (2002). Analysing the content their of atmospheres can provide precious information about planet history and interior. For instance, the metallicity or carbon-to-oxygen ratio can be leveraged to constrain the formation and migration (Venturini et al., 2016; Madhusudhan et al.,

2017; Turrini et al., 2021), and planets such as hot rocky planets with atmospheres are thought to have interior and magma composition directly linked to their atmosphere (Ito et al., 2022).

Exoplanetary atmosphere studies require ultra-high precision observations of the order of a few 10 parts-per-million (ppm), only really observable through our best telescopes and preferably from space. Since transit signals are larger for larger planets, most of the observed atmospheres have been detected around Jupiter-size planets so far, with only a minority of atmospheres around Earth-sized or sub-Neptune planets (e.g. McArthur et al., 2004; Léger et al., 2009; Barragán et al., 2018; Espinoza et al., 2020; Tsiaras et al., 2016, 2019; Benneke et al., 2019).

In recent years there have been population studies of atmospheres, with for example Sing et al. (2016); Tsiaras et al. (2018); Pinhas et al. (2019); Changeat et al. (2022) making use of the whole legacy of exoplanet observations with NASA's Hubble Space Telescope (HST) and Spitzer space telescopes (Werner et al., 2004). An example of 30 exoplanetary atmospheres' spectra and their associated retrieved atmospheric content from Tsiaras et al.'s article is shown on Figure 1.6. In addition, the future is bright for the study of exoplanets and their atmospheres.

Despite this fulgurant progress in the last 25 years since the birth of exoplanet sciences, there are important questions which remain unanswered yet. For example: how are planet parameters distributed? How do planets form and evolve? What are their interiors made of? What are the possibilities, conditions and markers for life on exoplanets? The ubiquity and striking diversity of observed exoplanets motivates each of these questions, which will in turn help us shed light on our own Solar System and planet.

1.2.4 The future of exoplanet discoveries

The future is bright for exoplanet discoveries, with new dedicated instruments and detection methods expected to pursue the initiated revolution. Indeed, most national and international space agencies worldwide are currently planning several exoplanet-related missions, integrating exoplanet-related questions among their core scientific objectives and therefore allocating a significant part of their budget to exoplanet

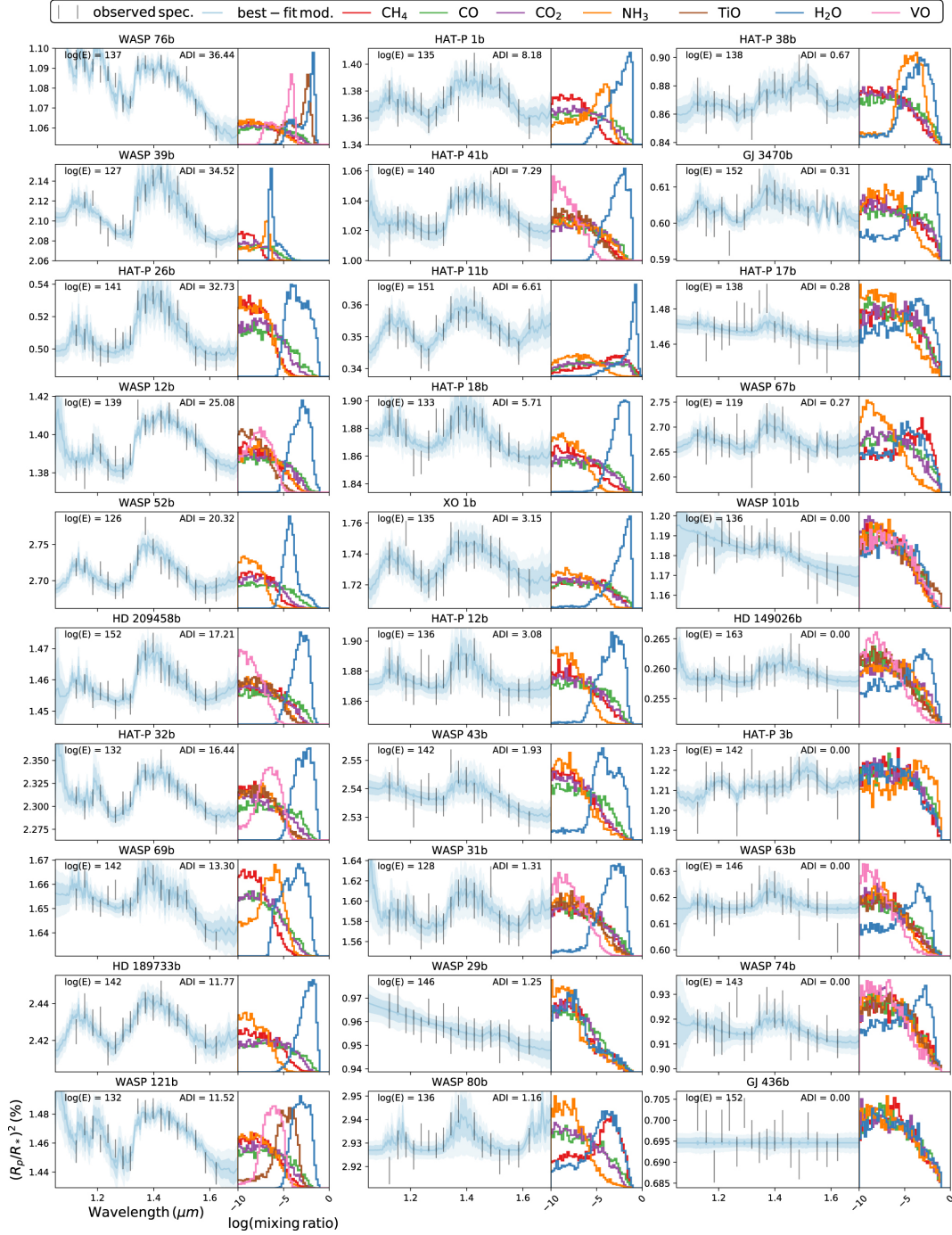


Figure 1.6: Homogeneous analysis of 30 exoplanet atmospheres from [Tsiaras et al. \(2018\)](#).

Every left-hand side plot shows the processed spectra (in units of transit depth variations, with error bars) measured from HST's WFC3 instrument and the associated best-fit atmospheric model in blue with its confidence interval (shaded blue). On each right-hand side plot are the 1D histograms for the posterior distributions of molecular abundances allowed for each model.

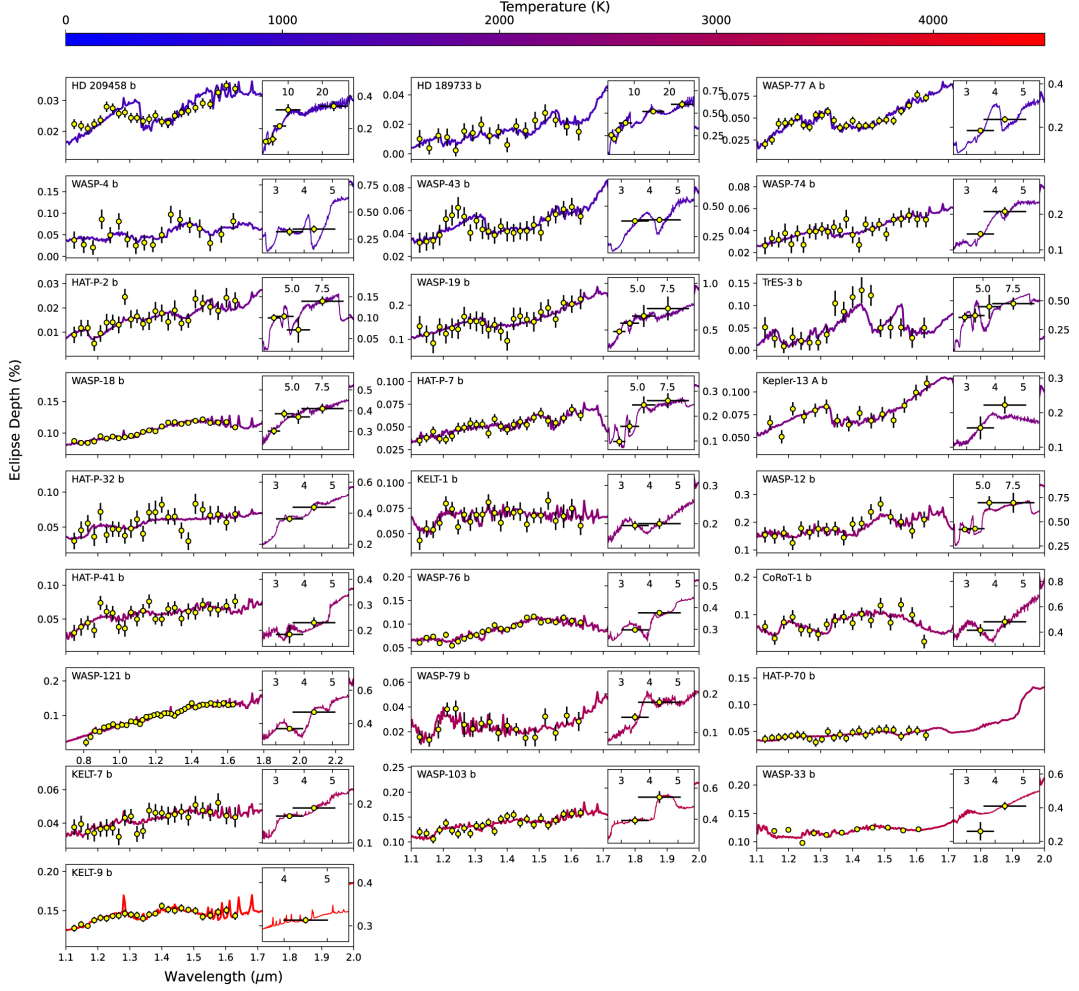


Figure 1.7: Homogeneous analysis of 25 hot Jupiter atmospheres from [Changeat et al. \(2022\)](#) through their emission spectra measured by HST and Spitzer between 1.1 and 2.2 μm . Best fit models are showed in continuous lines, coloured from the coolest retrieved temperatures in blue to the hottest temperatures in red.

sciences—see for instance [NASA’s Astro2020 decadal survey](#), [ESA’s 2015–2025 Cosmic Vision](#), the Chinese Academy of Sciences ([Ye, 2022](#)), [the Canadian Space Agency](#). Some aspects of the future evolution of exoplanets sciences are discussed in the following paragraphs, keeping the transit method in focus as this is the most relevant to this work and is to remain one of the most important techniques in the coming decades.

Detection methods

The transit method is expected to continue to play a crucial role when it comes to detecting and studying exoplanets in the coming decades (e.g. continuing Kepler

and TESS legacy with PLATO³, [Catala and The PLATO Consortium, 2009](#)), with synergies with other detection methods. Radial velocity surveys will continue to lead the follow-up confirmation and mass determination of transit candidates. The Nancy Grace Roman Space Telescope ([Green et al., 2012](#)) is expected to detect thousands of exoplanets through microlensing events. The upcoming high precision astrometric mission Gaia expected to discover of the order of $\sim 10^4$ new planets using astrometry ([Lattanzi et al., 2000](#); [Sozzetti et al., 2001](#); [Casertano et al., 2001, 2008](#); [Sozzetti et al., 2014](#); [Perryman et al., 2014](#)). Gaia is also expected to detect number of planets with transit photometry. A small fraction of its discovered planets will be observed both using astrometry and photometry, providing the full orbital solution, radius and mass for those estimated ~ 100 planets. Going forward, a NIR astrometric mission such as Gaia NIR ([Hobbs and Høg, 2017](#)) would allow probing for hidden regions in the Milky Way, facilitating to detect planets in young stellar environments. Furthermore, other innovative methods will also emerge or develop, such as gravitational waves detections ([Tamanini and Danielski, 2019](#); [Danielski and Tamanini, 2020](#)), interferometry ([Lacour et al., 2019](#)), transit imaging ([van Belle et al., 2012](#)).

Cutting-edge detections

While the faintness of exoplanets signals makes it hard to detect anything but the most obvious planets, detection limits will be pushed back as our instruments and data processing capabilities improve. If there has been one candidate of extragalactic exoplanet detection ([Di Stefano et al., 2020](#)), all the planets discovered so far have been in the Milky Way. In addition, if some exorings ([Kenworthy and Mamajek, 2015](#)) and exocomets ([Ferlet et al., 1987](#); [Beust et al., 1990](#); [Boyajian et al., 2016](#); [Welsh and Montgomery, 2018](#); [Rappaport et al., 2018](#); [Welsh and Montgomery, 2019](#); [Kennedy et al., 2019](#)) have already been detected, exomoons ([Teachey and Kipping, 2018](#); [Kipping, 2020](#)), planets around black holes still remaining to be officially discovered, as we expect them to be very common but requiring better photometric precision and/or innovative processing techniques. Open clusters and

³PLANetary Transits and Oscillations of stars

certain types of variable stars provide us challenging cases, pushing our detection abilities to their limits, and we expect more of those in the future.

Improving planetary characterisation

Building on the serendipitous successes from the first generation of space-based exoplanet characterisation missions HST and Spitzer, several current and planned space missions now integrate in their design the study of exoplanets', improving considerably the precision of measurements and wavelength coverage, with e.g. Cheops ([Benz et al., 2021](#)), JWST (James Webb Space Telescope, [Gardner et al., 2006](#)), PLATO ([Catala and The PLATO Consortium, 2009](#)), Ariel (Atmospheric Remote-sensing Infrared Exoplanet Large-survey, [Tinetti et al., 2021](#)) and Tinkle ([Edwards et al., 2019](#)). Beyond just transit spectroscopy from space, the JWST has already proven to exceed its predicted precision when it comes to directly imaging exoplanets ([Carter et al., 2022](#)).

From the ground, it will soon be possible to leverage telescope facilities with much larger aperture than in space with the improvement of adaptive optics to circumvent atmospheric perturbations, ELT (Extremely Large Telescope, [Gilmozzi and Spyromilio, 2007](#)), GMT (Giant Magellan Telescope, [Johns et al., 2012](#)), TMT (The Thirty Meter Telescope, [Sanders, 2013](#)).

These technological developments will enable more detailed inferences about the surface (e.g. [Cowan et al., 2009](#); [Cowan and Fujii, 2018](#)), interiors ([Helled et al., 2022](#)), climates (e.g. [Cowan et al., 2015](#)), therefore progressing considerably our understanding of these outer worlds.

Improved stellar characterisation

As the adage "know thy star, know thy planet" suggests, understanding host stars is paramount for exoplanet characterisation and detection. Indeed, all fundamental properties of stars are important to derive precise planetary parameters: effective temperature T_{eff} , radius R_* , luminosity L , surface gravity g , chemical composition, mass M_* , density ρ_* and age. The Gaia survey is providing a wealth of stellar information for millions to billions of nearby stars, thus enabling to reduce considerably the uncertainties on planetary parameters. Furthermore, understanding better the causes

of stellar activity (such as stellar spots, granulation, flares, see Section 2.4) will help mitigate its effects on observations and pinpoint its effects on exoplanets (e.g. habitability). This can be through solar and stellar interferometry and spectroscopy, but it has also been predicted that the next generation of space-based telescopes for transit spectroscopy will be able to constrain star spot temperatures directly from observations (Bruno et al., 2022).

Improvement of data processing and analysis

All of these advances, initiated by ever-increasing technological capabilities, will only increase the requirements for automated processing of stellar and exoplanets data. Large datasets of light curves—among others data products—will in turn benefit from using machine learning techniques to help modelling what can not be modelled analytically, help detect what can not be detected manually and speed up processing times.

1.3 The Machine Learning Revolution

1.3.1 Genesis

According to one of its pioneers Tom M. Mitchell, "machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience", i.e. without explicitly being programmed. As such, machine learning (ML) is considered a branch of artificial intelligence (AI), a broader field of research concerned with mimicking human intelligence with computers. Machine learning algorithms are generally called *models*, and the process of improvement through experience is generally called *training*. Whereas classical algorithms typically require explicit programming of the rules and functions to perform a specific task, machine learning models learn from the available data to provide us with a function, and can therefore be repurposed for a variety of tasks or subtasks by simply changing the data they are trained with.

One key milestone in deep learning was when the *backpropagation* algorithm was first used to train artificial neural networks (ANNs, see an example on Figure 1.8) in 1986 (Rumelhart et al.). Although ANNs existed before, backpropagation was

the missing piece in the overall procedure to train them. It finally provided a simple and efficient way to compute gradients of functions represented by arbitrarily large a deep ANNs, based on the chain rules for partial derivatives of functions with multiple variables. Today backpropagation is at the core of the definition of *deep learning*⁴, whose models are made up of differentiable modules—inspired from the biological brain or not—trained end-to-end using this simple but effective mechanism.

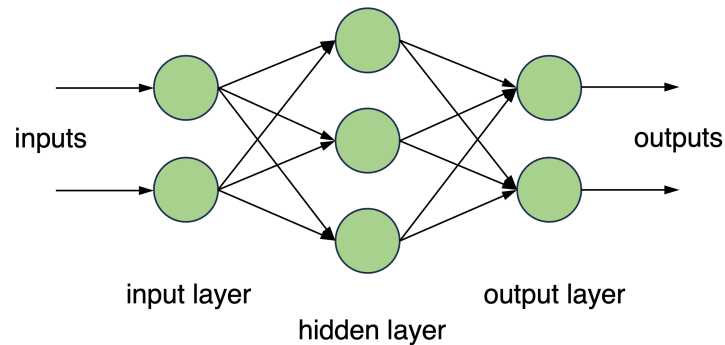


Figure 1.8: Diagram of a fully connected neural network with 2 input *neurons*, 3 hidden neurons (composing the hidden layer), and 2 output neurons. Fully connected neural networks are the simplest form of ANN architecture where all connections exist between pairs of neurons from two consecutive layers. Each neuron acts as a multiplier to build an arbitrarily complex function from multiple simple differentiable units, with information flowing strictly from left to right during a *forward pass*. Backpropagation enables information to flow in the opposite direction to the arrows, providing an efficient way to compute the output gradients with respect to the input parameters, and hence perform gradient-based minimisation⁵.

In the 1990s, several machine learning algorithms such as support vector machines (Boser et al., 1992) or kernel machines (Hofmann et al., 2008) were developed, as well recurrent neural networks (RNNs, Siegelmann and Sontag 1995) were popularised and successfully applied to real problems. This contributed to the establishment of machine learning as a recognised field and the advent of a data-driven paradigm opposed to the knowledge-driven paradigm.

Deep learning soared in the years 2000s with e.g. Hinton and Salakhutdinov (2006), when increased computational power unleashed the power of multi-layer

⁴See Goodfellow et al. (2016) for a comprehensive book about deep learning history and theory.

⁵Neural networks are trained via gradient-based minimisation of a loss function. In each iteration of training, each parameter of the model (weight) receives an update proportional to the partial derivative of the loss w.r.t the current weight.

networks. However, deep learning is often considered to have reached a hegemonic turning point with [Krizhevsky et al. \(2012\)](#), when the error rate of the top-5 classifications in the LSVRC (Large Scale Visual Recognition Challenge) was brought from 26.1 % to 15.3 % using a particular type of ANN called *convolutional neural network* (CNN). This particular type of *architecture* was particularly well suited and efficient for processing images. This brought a lot of attention to the field, and from there followed a series of innovative models adapted to various imaging tasks, followed by significant progress on other types of data such as video, language, speech, audio (e.g. [LeCun et al., 2015](#)), but also symbolic equations (e.g. [Lample and Charton, 2019](#)), spherical images (e.g. [Cohen et al., 2018](#); [Cobb et al., 2021](#); [McEwen et al., 2021](#)), graphs (e.g. [Scarselli et al., 2009](#)) etc.

1.3.2 Tasks and paradigms

The goal of ML models is to solve *tasks*, i.e. generic problems to be solved independently of the specifics of the data or application itself. Broadly, a task is specified by a data structure (inputs and outputs), and an evaluation criterion or *metrics*.

Depending on the task and the data available, models are trained using one or several learning strategies, which often fall in one of the three main learning paradigms below⁶.

Supervised learning is the archetype of a machine learning paradigm, whereby models' outputs are directly evaluated with respect to the ground truth. Classification and regression are two broad tasks typically solved using supervised learning. For classification, the targets to be predicted are discrete labels or categories, whereas for regression the targets are scalar or vectors of arbitrary dimensions.

Reinforcement learning relates to the idea of solving a task via "trial and error". A reinforcement learning task is specified by an environment, which an agent can interact with (reason why the word "agent" is substituted to "model" in that case) by its possible actions. From a given environment (Markovian) state, the agent is programmed so as to produce actions (model's outputs) which will result in a change

⁶See [Hastie et al. \(2013\)](#) for an in-depth definition and discussion of supervised and unsupervised learning.

of the environment. A reward function needs to be specified to inform the agent on how to adapt. See e.g. [Sutton and Barto \(2015\)](#) for a comprehensive introduction to reinforcement learning.

Unsupervised learning consists in learning a task without any available ground truth. For instance, clustering is an unsupervised task where the model learns to group together similar inputs, but without being shown any example of those clusters a priori.

Although those are the main ones, there are several other learning strategies. *Self-supervised learning*, for instance, consists in leveraging unlabelled data—either from the training set or from an external dataset—by using a carefully engineered proxy task whose targets are created from the data itself. Typically, self-supervised models can help to leverage large amounts of data via pre-training on large datasets and fine-tuning on the task(s) and dataset of interest. Chapter 4 presents a way to train self-supervised model on datasets of light curves.

1.3.3 Machine learning in astronomy and exoplanet sciences

A vast majority of physics and astronomy consists in trying to make sense of carefully acquired data. This qualitative interpretation of the physical world often involves solving subtasks such as classification, regression, clustering, forecasting, simulating, etc. Since ML models excel at these tasks in a variety of contexts, it should not come as a surprise to find an increasing use of ML models to help solve physics problems. Still, it is particularly striking to see the recent surge in the three last years in the fraction of publications in physics and astronomy which mention the terms "machine learning", "deep learning" or "neural networks" (see Figure 1.9).

This trend is a consequence of the progress in machine learning, but is also fostered by the apparition and multiplication of large datasets in science. More than just a fashion effect, there are reasons to believe that this is a real shift in the way scientific research is conducted, from "tightly controlled, theory-guided experiments towards an approach based on data-driven searches" ([Ourmazd, 2020](#)). It affects nearly all fields of observational science, especially those with significant amounts of data, e.g. in astronomy (e.g. [Borne et al., 2009](#); [Kremer et al., 2017](#); [Baron, 2019](#);

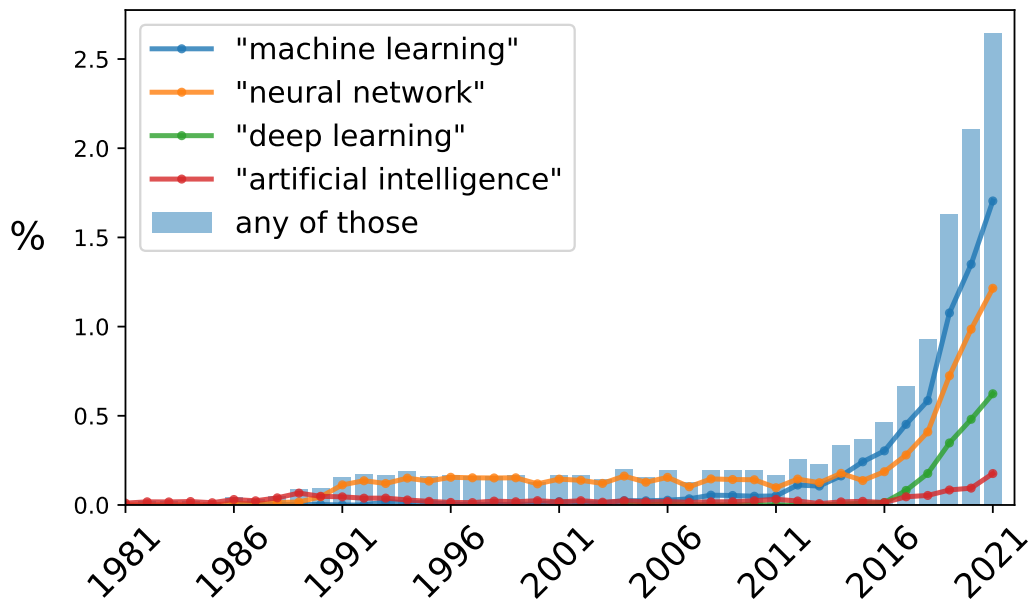


Figure 1.9: Fraction of astronomy articles on ADS⁷ mentioning the terms "machine learning", "deep learning", "neural network" or "artificial intelligence" over time. Note that the same trend is visible when restricting the corpus to refereed publications only, or extending to the whole *physics* collection.

Sen et al., 2019), but also in particle physics (e.g. Radovic et al., 2018), high-energy-density physics (e.g. Hatfield et al., 2021) or genomics (e.g. Libbrecht and Noble, 2015; Eraslan et al., 2019), to cite only a few.

The field of exoplanet sciences is no exception, with a booming number of publications making use of ML. This is apparent in various subfields including transit vetting (see Section 2.5 for more details), for which a few early works have used random forest algorithms (McCauliff et al., 2015; Thompson et al., 2015) before the use of neural networks was popularised in the field (e.g. Kipf and Welling, 2017; Pearson et al., 2018; Zucker and Giryes, 2018; Shallue and Vanderburg, 2018; Ansdell et al., 2018; Osborn et al., 2019; Dattilo et al., 2019; Chaushev et al., 2019; Schanche et al., 2019; Yu et al., 2019; Yeh and Jiang, 2020; Rao et al., 2021); regression of planetary parameters (Alibert and Venturini, 2019; Baumeister et al., 2020; Tasker et al., 2020; Auddy and Lin, 2020; Nikolaou et al., 2020); planetary dynamics modelling (Tamayo et al., 2016; Lam and Kipping, 2018; Kong et al., 2021; Valencia et al., 2019; Cambioni et al., 2019; Hinkel et al., 2019; Timpe et al.,

⁷SAO/NASA's Astrophysics Data System

2020; Pearson, 2019); atmospheric studies (Waldmann, 2016; Márquez-Neila et al., 2018; Zingales and Waldmann, 2018; Passegger et al., 2020; Johnsen et al., 2020; Fisher et al., 2020; Yip et al., 2021; Pham and Kaltenegger, 2021; Hayes et al., 2020; Nixon and Madhusudhan, 2020; Soboczenski et al., 2018; Cobb et al., 2019; Himes et al., 2022; Guzmán-Mesa et al., 2020); microlensing detections (e.g. Wyrzykowski et al., 2015, 2016; Chu et al., 2019; Godines et al., 2019; Mróz, 2020; Zhang et al., 2021a,b); direct imaging (Gomez Gonzalez et al., 2018; Yip et al., 2020b); and instrument data processing which will be discussed at length in Chapter 2.

More widely in the field of time-domain high-precision photometry, there have been numerous works developing supervised algorithms for classification of variable stars (Debosscher et al., 2007, 2009; Blomme et al., 2010; Aguirre et al., 2019; Martínez-Palomera et al., 2020; Becker et al., 2020; Hosenie et al., 2020; Barbara et al., 2022).

ML has also started to show promising results on more abstract problems such as solving intractable ordinary differential equations with physics-informed ML (Karniadakis et al., 2021), rediscovering the laws of physics through symbolic regression (Cranmer et al., 2020) or even searching for new fundamental laws of physics (Karagiorgi et al., 2022).

Despite its tremendous progress in artificial intelligence, its recent gain of popularity in physics, and its bright years to come, there are a number of challenges and pitfalls inherent to machine learning that we should be aware of when using it for scientific applications.

1.3.4 Challenges and opportunities

In astrophysics, the ground truth is scarcely available. Labels and categories are generally harder to collect or even define, and quantities are but mere estimates of the truth with associated uncertainties—estimated, too. The use of supervised learning is therefore limited in scope regarding its applicability to real physics data. And whereas recent developments of probabilistic deep learning have suggested ways to help integrate the notion of uncertainty in deep learning models (Abdar et al., 2021), this is still fairly experimental and most of the ML models used so far in physics

have been lacking this ability.

Deep learning remains elusive in that there is not one satisfying theory behind it. This, added to the lack of familiarity of scientists with ML, contribute to a reluctance to use ML models in physics. Of interest to the physicists are the growing interest in deep learning *interpretability* and *explainability* (e.g. [Ras et al., 2021](#)), which will help facilitate the dialogue between the two fields. Interestingly, physics could also have its role to play in understanding deep learning ([Zdeborová, 2020](#)).

As the bonds between applied science and ML continue to strengthen, physics is going to motivate the development of innovative ideas in unsupervised and self-supervised ML. There is thus a double incentive in getting the ML community interested in solving physics problems: on the one hand it lays the potential for scientific breakthroughs, and on the other hand it lays the potential for ML to find new challenges, extending the range of its problems, constraints and models. The day when ML systems do science better than humans will mark a milestone in the development of artificial intelligence.

1.4 Thesis Outline

Detecting exoplanets can be compared to finding needles in a haystack of noise. Moreover, we would like to precisely study the found needles while they remain diluted in an endless ocean of haystack. This thesis presents a set of tools for automated and efficient processing of noisy light curves. It addresses the problem of correcting the various undesired sources of variability (composing the haystack), by bridging the gap between deep learning (our machinery) and the physics of exoplanets (needles).

Chapter 2 is dedicated to light curves, the data objects at the very core of this work. It brings together the classical formalism of time series analysis useful for exoplanet light curves modelling with the physical causes for their variability. It also formalises the central problem of detrending light curves and links it with existing works of interest. This chapter will thus equip us with the minimal definitions, tools and background upon which this work is constructed.

In Chapter 3, a method to detrend individual observations is presented. It uses

recurrent neural networks to learn the trend in transit light curves with a forecasting objective. Its effectiveness is demonstrated on data from the Spitzer space telescope.

In Chapter 4, another method is introduced, aimed at global detrending of large datasets of light curves. It makes use of a Transformer (Vaswani et al., 2017) encoder trained with a masked objective to learn the trend in multiple time series. An application is shown on a whole sector of TESS light curves.

Chapter 5 introduces a hybrid framework to combine transit physics with deep learning. This relies upon the first automatically differentiable transit code PyLightcurve-torch (in PyTorch). An application is presented on simulated data making use of a hybrid regression loss, and several further possibilities opened by this code are discussed.

The techniques introduced in Chapters 3 and 4 both aim at precisely modelling the time-correlated noise or trend to improve the accuracy of transit studies: "Know your hay...". Chapter 5 proposes a way to integrate the introduced techniques—or in fact any deep learning algorithm—along with an exoplanet transit model.

Chapter 2

Light Curves

"How far that little candle throws his
beams! So shines a good deed in a
weary world."

William Shakespeare,
The Merchant of Venice

Light curves are the fundamental data bricks of this thesis. They represent at the same time: (i) the product of a photometric pipeline processing raw images from telescopes; (ii) entanglements of astrophysical and instrumental variability; (iii) the input data for the methods introduced in this work. Given the centrality and complexity of these objects, this chapter aims at setting some foundations by formalising various definitions, models and problems related to light curves.

2.1 Formalism

Light curves are time series of brightness, i.e. successive flux measurements (F_1, F_2, \dots, F_n) of an object at different times (t_1, t_2, \dots, t_n) , where $n \in \mathbb{N}$ is the total number of measurements.

In practice, every measured light curve is integrated spatially over a *field of view* which contains not only the object of interest, but also potential other sources of light and noise as well. Furthermore, every measurement is also integrated over an *exposure time* and a flux value at a given time is thus actually estimated from all the light detected during this exposure time. Finally, every observation is also integrated

over a spectral bandpass, and one therefore needs to remember that any observed variation of light will be specific to this particular bandpass¹. In the following we will consider that the light is measured and converted to electric current through a detector (e.g. CCD), measured in e^-/s .

Light curves being time series, it is useful to introduce several definitions inherited from time series analysis. In the following we will call (x_1, \dots, x_n) a general time series to be distinguished from the notation (F_1, \dots, F_n) reserved for light curves.

Time sampling

Time series measured at constant time intervals ($\Delta t := t_i - t_{i-1}, \forall i$) are said to be *regularly sampled*. Since many tools and algorithms for time series analysis only apply in regularly sampled scenarios, the non fulfilment of this property calls for extra-care in the choice of processing techniques. When regularly sampled, the constant interval Δt between successive times is called *cadence*.

Throughout this work, time series and light curves will be assumed regularly sampled unless stated otherwise. This refers both to the integration time and to the time in between measurements. With this assumption we can then use directly $t = 1, \dots, n$ as time index for generic values.

Dimensionality

If one scalar value only is available at each time step, a time series is said *univariate*. Conversely, if multiple values $x_t^1, x_t^2, \dots, x_t^d$ are available at each time step, the corresponding time series is said to be *multivariate*, and the different values across its d dimensions are also called *features*. When additional time-dependent variables are measured along with a time series of interest, these might have an explanatory power and are called *covariates* or *covariate* time series.

Common cases of multivariate light curves include e.g. flux values for individual pixels p^1, p^2, \dots, p^d composing a given field of view in detectors—the individual light curves are often called *pixel light curves*—, or flux values measured at various wavelengths $\lambda^1, \lambda^2, \dots, \lambda^d$ —often called *spectral light curves*. Common cases of covariate time series associated with light curves are additional instrumental

¹some sources may exhibit variability in colour which will impact each spectral band variability such as doppler-beamed sources, with e.g. [Herrero et al. \(2014\)](#)

measurements such as temperature or source position on the detector are available.

Stochasticity

Time series can be considered random processes, i.e. that each sample x_t is drawn from an underlying probability distribution p_t with associated summary statistics such as mean μ_t or standard deviation σ_t .

When the standard deviation exists and is constant, time series are said *homoscedastic*. Conversely, *heteroscedastic* time series have changes in variance.

Trend

Broadly speaking, trends in time series are associated with the change in mean value μ_t over time, i.e. its deterministic component. In practice, this definition can vary depending on the context or method to decompose the time series. In this work, we define the *trend* as the time series of mean values μ_1, \dots, μ_t once removed the potential signal of interest. Note that this definition can include period patterns, or even variability patterns on a shorter timescale than the signal.

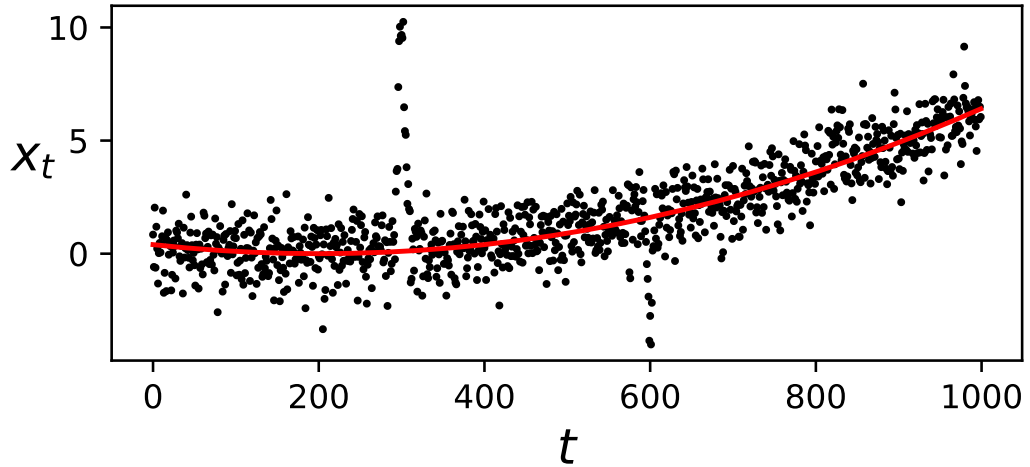


Figure 2.1: Example of a synthetic non-stationary time series (black dots) with second-order polynomial trend (red line), Gaussian white noise of variance unity, and two transient events around time step 300 and 600.

Signal and noise

We distinguish between two broad categories of variability in a time series: *signal* and *noise*. While the signal is associated with the studied object, all of the other variability components are part of the *noise*. Therefore, the same source of variability

can either be considered as signal or as noise, depending on the science objective.

Stationarity

A random process $(\epsilon_1, \dots, \epsilon_n)$ is said to be stationary when its covariance between two elements only depends on the time difference s between them:

$$\text{cov}(\epsilon_t, \epsilon_{t-s}) = \sigma_s^2, \forall s,$$

, where σ_s is constant for each time difference s .

Types of non-stationarity include presence of a trend, heteroscedasticity and change points (discontinuities in mean or variance). Many algorithms in time series analysis are assuming stationarity in the data, and therefore require correction of any trend or change point to be applicable.

Frequency domain

Time series can be analysed in the frequency domain using a frequency-based method such as Fourier analysis, wavelet analysis, and others. In this context, time series admit a power spectrum which represents the frequency distribution of the squared values x_t^2 (often called power). The power spectrum can be conveniently defined as the Fourier transform of the autocorrelation function (e.g. [Ricker, 2003](#)). Frequency methods can be particularly useful to study periodic signals, but can sometimes be impractical and often require stationarity, a condition rarely met with the raw light curves studied in this work.

White noise

A *white noise* process is a random process $(\epsilon_1, \dots, \epsilon_n)$ whose elements (i) have zero mean: $E(\epsilon_t) = 0, \forall t$; (ii) are independent: $\text{cov}(\epsilon_t, \epsilon_{t-s}) = 0, \forall s \neq 0$; (iii) and have constant variance: $\text{var}(\epsilon_t) = \sigma_w^2, \forall t$. Equivalently, a white noise process can be described as a process with flat power spectrum.

Common sources of photometric noise commonly modelled as white noise processes include the shot-noise or the readout noise. The shot-noise (also called Poisson noise or photon noise) is associated with the discrete nature of light emitted by radiative sources, i.e. the fluctuations in the count of photons received by our

detector during a given exposure time is significant and generates scatter in a light curve of standard deviation $\sigma_{shot} = \sqrt{n_{photons}}$ where $n_{photons}$ is the number of photons collected during the exposure time. Note that for large photon counts, the Poisson distribution is very well approximated by a normal distribution with same mean $n_{photons}$ and standard deviation $\sqrt{n_{photons}}$. The *readout noise* is a measurement noise resulting of the imperfect conversion of photons into electric current in a detector. In addition, electronic detectors are also subject to *dark current* noise due to their crystallographic defects, which also has a shot noise component in addition to a fixed mean component.

Red noise, pink noise

Red noise denotes a common case of non-stationary process with (i) zero mean: $E(x_t) = 0, \forall t$; (ii) constant variance: $var(x_t) = \sigma_r^2, \forall t$; (iii) serial correlation between successive values: $cov(x_t, x_{t-1}) \neq 0$.

Pink noise or $1/f$ noise refers to a process whose power spectrum is inversely proportional to the frequency of the signal.

Signal-to-noise ratio

Under the assumption of white noise as the only source of the noise, the signal-to-noise ratio (SNR) is defined as:

$$SNR = \frac{\delta}{\sigma_w} \sqrt{n_{signals}}.$$

Where δ is the signal power, equal to the transit depth for transit observations, and $n_{signals}$ is the number of observed signals (transits).

For a CCD detector,

$$\sigma_w = \sqrt{\sigma_{source}^2 + \sigma_{background}^2 + \sigma_{dark}^2 + \sigma_{readout}^2},$$

where $\sigma_{background}$, σ_{dark} , and $\sigma_{readout}$ are the standard deviations associated with the background noise², dark current and readout noise.

²The background noise consists in the resulting light from all background sources present the target in the aperture considered for photometric analysis. It can be either stationary or non-stationary.

In the presence of red noise and white noise, Pont et al. (2006) proposed the formula:

$$SNR = \delta \frac{n}{\sqrt{\sum_{k=1}^{N_{tr}} n_k^2 \left(\frac{\sigma_w^2}{n_k} + \sigma_r^2 \right)}}.$$

Multiple time series

Many classical methods in time series analysis are based on the individual processing of every single time series, regardless of there being other time series or not.

When multiple time series are available for a given problem, they form a *dataset* of multiple time series—simultaneous or not. This opens the possibility for global models to be built, leveraging the variety and similarities between light curves to solve the problem.

The large number of light curves available as well as their disparities of frequencies, samplings and underlying processes and associated noise distributions strongly motivates the development of such data-driven, global methods.

2.2 Transit Light Curves

Hereafter the analytical model used for exoplanet transits throughout this work is presented.

2.2.1 Planetary dynamics

The movement of planets can be described to a sufficient degree of accuracy using Kepler laws of motion, which can be derived from Newtonian's two-body dynamics under the assumption that $M_p \ll M_*$. In that context, a planet's orbit around its star on an elliptical orbit expressed as:

$$r = \frac{a(1 - e^2)}{1 + e \cos f},$$

where r is the distance between the stellar and planetary centres, a is the semi-major axis, e the eccentricity and v the true anomaly of the system represented on Figure 2.2.

Kepler's third law relates the semi-major axis and period of the system to its

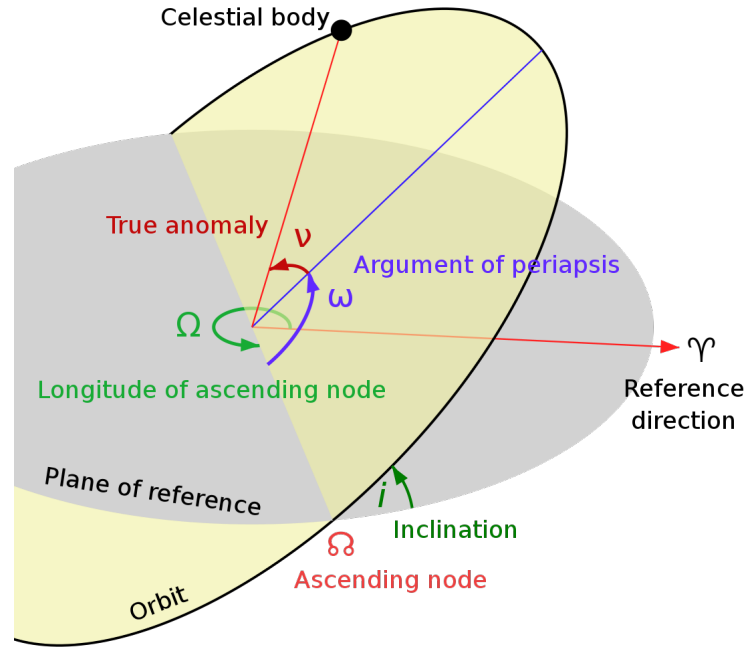


Figure 2.2: Diagram of a Keplerian orbit in three dimensions. The celestial body is moving along an elliptical orbit inclined with respect to the plane of reference—plane perpendicular to the line of sight—by an angle i (image by Lasunncty under CC BY-SA 3.0)

total mass, which can often be approximated into M_* assuming $M_P \ll M_*$:

$$a^3 = \frac{G(M_* + M_P)}{4\pi^2} P^2 \approx \frac{G(M_*)}{4\pi^2} P^2.$$

From Kepler's laws one can derive the total transit duration (see e.g. [Tingley and Sackett, 2005](#)):

$$t_T = 2(R_* + R_P) \left(\frac{P}{2\pi G(M_* + M_P)} \right)^{\frac{1}{3}} \frac{\sqrt{1-e^2}}{1 + e \cos v_0} \sqrt{1 - \frac{r_0^2 \cos^2 i}{(R_* + R_*)^2}},$$

where $v_0 = \frac{\pi}{2} - \omega$ is the true anomaly at mid-transit, and $r_0 = a(1 - e^2)/(1 + e \cos v_0)$ is the separation at mid-transit. Note that the effect of orbital inclination is concentrated in the last term, and the effect of eccentricity is concentrated in the penultimate term.

2.2.2 Shape calculation

A schematic view of the system's geometry and notations is shown on Figure 2.3. Spherical symmetry is assumed.

Following the formalism from [Mandel and Agol \(2002\)](#), the stellar area $\lambda(p, z)$ obstructed by the planet can conveniently be expressed as a function of the ratio of radii $p = R_p/R_*$ and separation between centres in units of stellar radii $z = d/R_*$. Outside transit ($1 + p \leq z$) the stellar disk is not obstructed $\lambda(p, z) = 0$. When the planetary and stellar disks are fully overlapping ($z \leq 1 - p$), then the obstructed area is simply $\lambda(p, z) = p^2$. Finally, in the case when only one part of the disks are overlapping ($1 - p < z < p + 1$):

$$\lambda(p, z) = \frac{1}{\pi} \left[p^2 \kappa_0 + \kappa_1 - \sqrt{\frac{4z^2 - (1 + z^2 - p^2)^2}{4}} \right],$$

where $\kappa_0 = \cos^{-1} [(p^2 + z^2 - 1) / 2pz]$, and $\kappa_1 = \cos^{-1} [(1 - p^2 + z^2) / 2z]$.

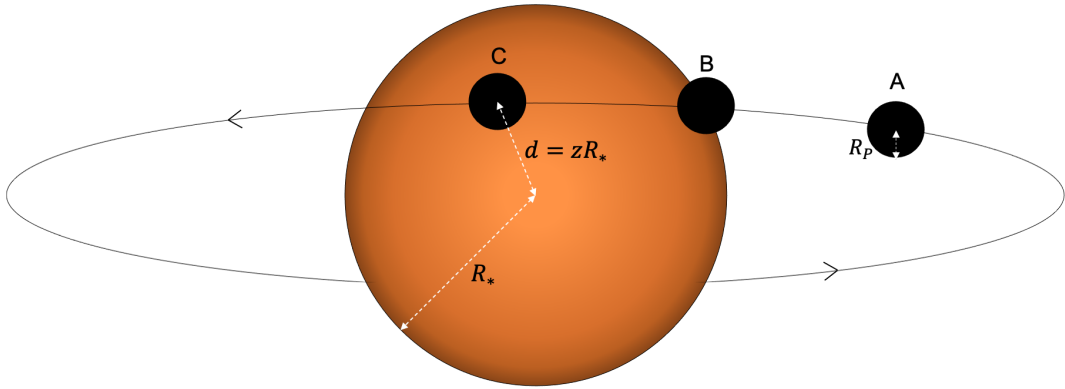


Figure 2.3: Schematic view of the host star surface during transit, as would be seen from the Solar System if we had a telescope with a good enough resolving power. A, B and C show different times along the orbits, with C out-of-transit, B during ingress and A during transit.

The flux emitted by the star is not equally distributed on the stellar surface, with typically fainter emission towards the limbs of the star. This is commonly accounted for through a limb-darkening function $I(r)$ which gives the ratio of the flux emitted at a radius r by the central flux ($r = 0$), under assumption of spherical symmetry of the star.

Under the presence of limb-darkening $I(r)$:

$$\Delta F(p, z) = \left[\int_0^1 dr 2r I(r) \right]^{-1} \int_0^1 dr I(r) \frac{d[\lambda(p/r, z/r) r^2]}{dr}.$$

Linear laws specify a functional form for $I(r)$. Common limb-darkening laws are taken as a polynomial function of r or its square root:

- linear limb-darkening: $I(r) = 1 - u(1 - r)$ (Schwarzschild, 1906)
- quadratic limb-darkening: $I(r) = 1 - a(1 - r) - b(1 - r)^2$ (Kopal, 1950)
- square-root limb-darkening: $I(r) = 1 - c(1 - r) - d(1 - \sqrt{r})$ (Diaz-Cordoves and Gimenez, 1992)
- four-terms limb-darkening: $I(r) = 1 - \sum_{k=1}^4 a_k(a - r^{k/2})$ (Claret, 2000)

where u, a, b, c, d, e, f and a_k are the limb-darkening coefficients.

Under the assumption of small planets $R_p \ll R_*$ and a common limb-darkening law, the above integral can be computed analytically with a loss in flux accuracy of about 2 % for $p \lesssim 0.1$ (Mandel and Agol, 2002). However, a careful split of the integral in three different regions as suggested by Tsiaras (2017) still allows efficient computation for various common limb-darkening laws without any loss of precision. The code PyLightcurve-torch implemented for this work and presented in Chapter 5 directly relies on this formulation. The effects of limb darkening and associated model biases have been studied in different contexts, and still remain an active and integral part of the effort to improve transit modelling (e.g. Espinoza and Jordán, 2016; Maxted, 2018; Morello et al., 2021).

2.2.3 Astrophysical assumptions

The above formulation of the transit light curve relies on multiple assumptions listed thereafter.

Constant stellar flux

The above transit light curve model supposes that star flux does not vary in time. As discussed in Section 2.4, most stars are actually variable, and this assumption

will be reasonably valid as long the stellar variability during and around transit is negligible with respect to the transit signal: $\sigma_{*,var} \ll p^2$. One of the objectives behind developing general detrending methods is to mitigate the impact that stellar variability can have on the detection and precise fit of transits.

Sphericity

If the host star is not spherical, the transit shape is expected to be asymmetrical. However, this effect was suggested to be negligible (of the ppm level) by [Hui and Seager \(2002\)](#). Planet oblateness and associated rotation is also expected to lead to a small and degenerate effect in the transit light curve (see e.g. [Carter and Winn, 2010](#)).

Star flux is spatially well-behaved following limb darkening laws

The above model supposes that the stellar surface is perfectly described by one of the limb-darkening laws, which supposes the absence of activity such as stellar spots, and flares. In practice, these two types of inhomogeneities can bias the transit parameters. Furthermore, limb darkening laws are approximate Taylor developments, and can be unadapted to describe some particular stars such as fast rotators.

Newton's laws describe the movement

Newton's laws are not the most accurate account of gravity, and for precise orbital solving it would be more accurate to use general relativity instead. The effect of apparent apsidal precession due to GR on the timing of transiting exoplanets was quantified in e.g. [Rafikov \(2009\)](#).

Host star stillness with respect to Earth

The relative motion of the system with respect to Earth (assuming our telescope is fixed with respect to Earth) causes slight variations of its distance and viewing angles. These can cause biases and drifts of several quantities, with secular or periodic effect due to proper motion or Earth rotation, as first suggested in [Scharf \(2007\)](#).

No moon, rings or other features

Exorings and exomoons are expected to confound the transit shape and apparent radius (e.g. [Kenworthy and Mamajek, 2015](#)).

2.3 Measurements

This section introduces the various exoplanet space missions and associated datasets of photometric light curves. It also specifies the most important systematic trends present in those datasets because of the measurements.

2.3.1 Large transit surveys

Following the first exoplanet transit detections, several telescopes were built to monitor stellar variability and search for transits with long and high-precision photometric observations, thus surveying large regions in the sky for long periods of time.

Of those, there have been multiple ground-based telescopes with e.g. KELT, HATNet, SuperWASP, NGTS (Wheatley et al., 2018); and several space missions: MOST, CoRoT, Kepler, K2 (Howell et al., 2014), TESS, and PLATO with a launch date planned for 2026. These surveys vary in many aspects concerning namely their instrument associated precision, targeted stars (type and location), duration of observations, etc. Some of their characteristics are summarised in Table 2.1.

Name	# targets	cadence	max duration	bandpass
SuperWASP	18×10^6	1.5 min	4 y	400 nm–700 nm
CoRoT LC	1.2×10^5	512 s	180 d	370 nm–1000 nm
NGTS DR2	6×10^5	13 s	3 months	520 nm–890 nm
Kepler SC	4000	1 min	4 y	423 nm–897 nm
Kepler LC	197 096	29.4 min	4 y	423 nm–897 nm
Kepler IC	6.5×10^6	29.4 min	4 y	423 nm–897 nm
K2	40 000	1 min - 30 min	75 d	423 nm–897 nm
TESS SC	200 000	2 min	30 d	600 nm–1000 nm
TESS LC	10^7	30 min	30 d	600 nm–1000 nm
PLATO	144 000	30 s	1825 d	500 nm–1000 nm

Table 2.1: Summary of the main wide-field photometric surveys and their characteristics. Here "LC" refers to Long Cadence and SC to short cadence. IC refer to the input targets actually visible.

Below we present Kepler and TESS, two of the missions whose surveys have proved particularly important and successful in the search for transiting exoplanets.

The Kepler Space Telescope was launched in 2009 with the objective of detecting small planets and studying the properties of exoplanetary systems - sizes, multiplicity,

host stars. It observed nearly 200 000 stars focusing on one region in the sky in Cygnus and Lyra constellations during its 4-year primary mission, yielding more than 2700 discoveries of transiting planets. Key to Kepler's success has been its long-term monitoring and near shot-noise photometric accuracy for stars ranging from magnitudes $K_p = 7$ to $K_p = 13.6$ (Koch et al., 2010), and a measurement noise still lower than the shot-noise for K_p between 13.6 and 17

After the loss of two reaction wheels in 2013, it was then ingeniously repurposed and started its extended K2 mission, observing 150 000 more stars in 80 days campaigns and in different fields around the ecliptic plane, while still retaining a 6-hr photometric precision of 80 ppm at $V = 12$ (Howell et al., 2014). More than 500 transiting exoplanets were discovered thanks to the K2 mission.

TESS was launched in 2018 with the objective of detecting planets around nearby and bright stars, with several thousand detections expected. It is located on an 13.7-d elliptical orbit with a 2:1 resonance with the Moon's orbit, inclined with respect to the ecliptic to eliminate Earth and Moon's eclipses. It has already surveyed more than ten millions stars in nearly all directions, among which 200 000 main targets with *short cadence* (2 min) during its primary mission. It is now undergoing extended missions, revisiting some regions and extending its collection modes to include an *ultra short cadence* of 20 sec.

TESS wide field-of-view ($24 \times 96^\circ$) is covered through four identical refractive cameras. Each camera contains a detector assembly with four 2048×2048 CCD arrays in their focal plane, with each $15 \mu\text{m}$ pixel covering $21''$. An example of detector is shown on Figure 2.4.

2.3.2 Characterisation missions

In parallel, targeted characterisation observations were performed from space with Hubble Space Telescope and Spitzer space telescope. Both telescopes have a much narrower field of view, increasing the SNR, but restricting their observations to single targets. It is striking that neither of these space telescopes was conceived with the objective to observe exoplanets, and yet with some ingenious of their operation modes and processing, they have proven sensitive enough and have unveiled some of

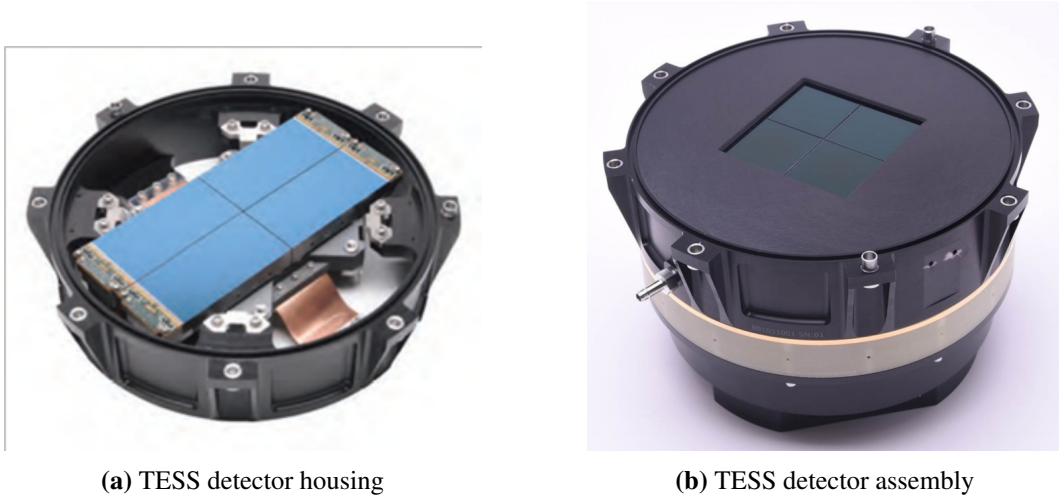


Figure 2.4: Simulated view of one TESS detector housing and assembly, from the TESS instrument handbook ([Vanderspek et al., 2018](#)).

the most dramatic discoveries about outer-worlds (e.g. [Deming and Knutson, 2020](#)). Present and future missions JWST, CHEOPS, Twinkle and Ariel are all expected to build on HST and Spitzer's legacy and help pushing exoplanet characterisation to the next step. Below we introduce some characteristics about Spitzer detectors as these are relevant to the analysis presented in Chapter 3.

The Spitzer Space Telescope was launched in 2003, onboarding three instruments: the Infrared Array camera (IRAC, [Fazio et al., 2004](#)), the InfraRed Spectrometer (IRS, [Houck et al., 2004](#)) and the Multiband Imaging Photometer for Spitzer (MIPS, [Rieke et al., 2004](#)). It was several orders of magnitude more precise than the previous space-borne infrared detectors, revealing the infrared universe as we had never seen it before. Compared to Hubble in geocentric orbit, Spitzer's innovative heliocentric Earth-trailing orbit enabled efficient cooling and no contamination from the Earth and Moon. After the helium cryogen used to cool the instruments was exhausted in 2009, Spitzer entered its "warm mission" (28 K) following its "cool mission" phase, which lasted until being decommissioned in January 2020.

The three Spitzer instruments covered a wide range of infrared wavelengths from the near to the far-infrared $3\text{ }\mu\text{m}$ to $160\text{ }\mu\text{m}$. IRAC (Figure 2.5) was equipped with four different 256×256 -pixels detectors taking separate images at $3.6\text{ }\mu\text{m}$, $4.5\text{ }\mu\text{m}$, $5.8\text{ }\mu\text{m}$ and $8.0\text{ }\mu\text{m}$. The two channels of shorter wavelength had detectors

made of indium antimonide (InSb), and the two higher wavelength channels 3 and 4 had silicon detectors treated with arsenic (Si:As). IRAC channels 1 and 2 were the sole among Spitzer's detectors which remained at use after depletion of helium cryogen in 2009. Observation modes were enhanced before the start of the warm mission to improve various types of observation such as exoplanets. The whole of IRAC data can now be accessed on the Spitzer Heritage Archive ([Spitzer Heritage Archive](#)).

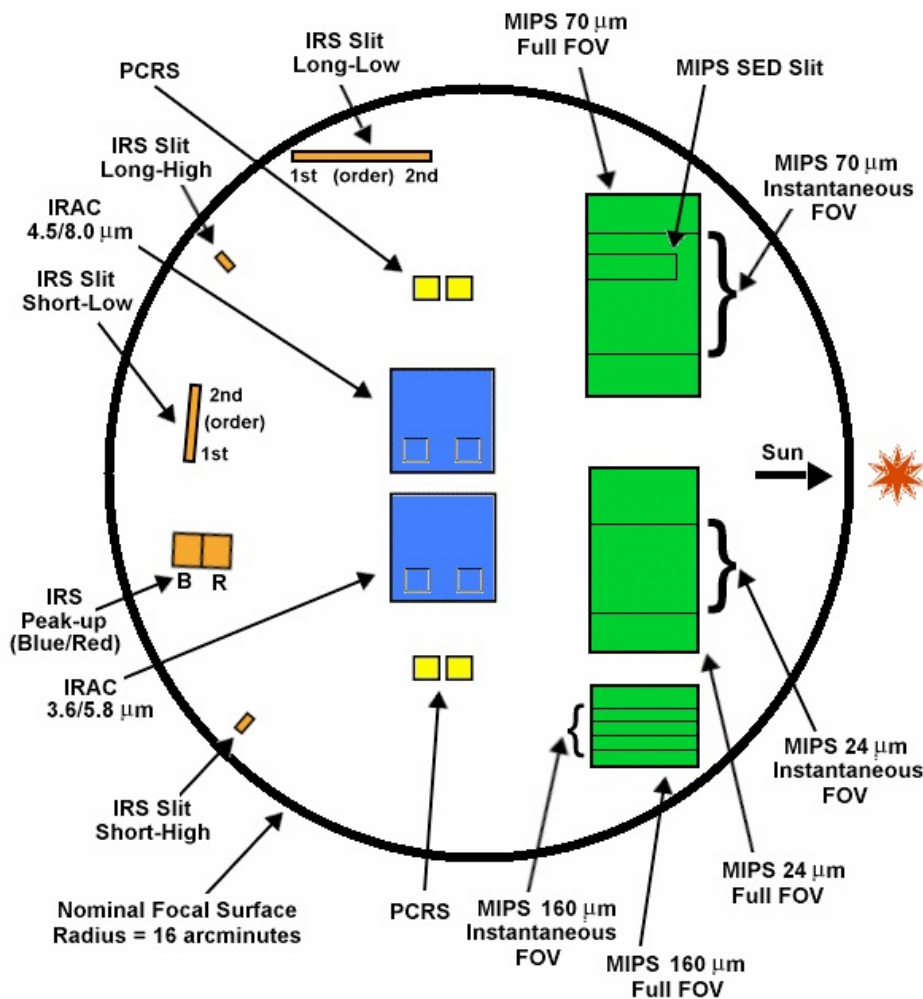


Figure 2.5: Field-of-views of Spitzer detectors projected onto the sky. The IRAC array fields are located in the central blue squares, and their respective sub-array fields are located in each square's lower corners. IRS slits are located on the left of the diagram and the various MIPS detectors in the larger green rectangles on the right. Si PIN photodiode detectors are used as pointing control reference sensors (PCRS) to recalibrate the pointing of the telescope every twelve hours. Figure from the Spitzer observation manual ([Spitzer Science Center, 2007](#))

The tale that Hubble and Spitzer have told us have only been possible through meticulous observing choices and data post-processing.

2.3.3 Instrumental systematics

There are a number of artificial effects linked to the acquisition and processing of photometric light curves, commonly referred to as *instrumental systematics* or merely *systematics*—as opposed to the purely (theoretical) astrophysical light curve. As part of the standard processing of raw photometric data, the main corrections which need applying are dark current subtraction, flat-field calibration³ and background subtraction. In addition, there can also be time-correlated artefacts left after standard post-processing.

The type of systematics observed largely depends on the detector material⁴, pixel size, substrate, and well depth, as well as on the overall spacecraft pointing and thermal stability. In order to achieve high photometric precision, one has to correct for the instrumental systematics to obtain an astrophysical light curve ready for scientific analysis. Below is a non-exhaustive list of some of the main causes of systematics and their effects on light curves.

Pointing variations are induced by various types of spacecraft motion including drifts, wobble and jitter. Each telescope uses a set of guide stars to control their pointing orientation. Any spacecraft motion therefore affects the quality of the alignment between the star tracker and the boresight of the telescope. In long-term observations (e.g. Kepler, TESS), regular readjustment of the spacecraft pointing is carried out through *momentum dumps*. When coupled with *intra-pixel variations*, pointing variations can lead to significant variations of up to 8 % in IRAC observations (e.g. Charbonneau et al., 2005; Grillmair et al., 2012; Ingalls et al., 2012, see an example on Figure 2.7a). The extent of the movement can be monitored through the *centroid positions* of a source (X, Y positions of the centre of light), correlated with the observed flux variations.

³Flat-field correction aims at accounting for the differences between pixel sensitivities (inter-pixel variations) on the detector.

⁴TESS, CoRoT, CHEOPS and Kepler are using CCDs while HST/WFC3 & NICMOS are MCT (Mercury Cadmium Telluride) and Spitzer/IRAC channels are InSb & Si:As detectors.

Scattered light from the Earth, the Moon or Mars can affect observations if not properly shielded. On TESS detector, it typically affects 10–15% of the full field-of-view at a level about 6 times the background level. IRAC detector saw patterns of scattered light from bright stars outside but close to the field-of-view (Hora et al., 2004).

Cosmic rays and solar energetic particles strike pixels on the detector unpredictably, often causing permanent local changes in pixel sensitivity. These are called *sudden pixel sensitivity drop-out* (SPSD) in Kepler photometric pipeline. The unpredictable jumps they induce are typically modelled by step functions (Stumpe et al., 2012). The cosmic ray mitigation is done onboard of TESS (except for the latest ultra-short cadence of the extended mission) in real-time by the data processing unit, therefore alleviating the need for PSD correction and reducing the number of outliers due to cosmic rays or solar energetic particles. IRAC detector was receiving cosmic rays at a rate of 3 pixels per second in channels 1 and 2 and 5 pixels per second in channels 3 and 4 (Fazio et al., 2004), which need to be discarded at post-processing time.

Anomalies are classified and flagged for each telescope. Most of them are flagged and discarded through the standard processing pipelines, and originate from the causes mentioned before (telemetry, pointing, energetic particles, etc.). They can affect individual targets (e.g. undershoot columns) up to all targets on a detector (e.g. fine-pointing anomalies). It is key to consider those (i.e., often by discarding) as they can have a drastic effect in the subsequent light curve analysis. Since discarding these anomalies may induce gaps in light curves, this might in turn cause a missing value problem whenever continuity is required. The methods we present in Chapters 3 and 4 are made to impute gaps as well as possible, and the latter one is also designed to be robust to persisting outliers.

Charges accumulation in detector traps (also called *persistence*) is the best explanation (Agol et al., 2010) to date for ramp-like artefacts (see Figure 2.7b) witnessed in IRAC (e.g. Deming et al., 2006; Knutson et al., 2007; Désert et al., 2009), WFC3 (e.g. Berta et al., 2012; Zhou et al., 2017) and JWST detectors. This effect is discussed further in Chapter 3, which also presents a method to correct for it.

Cotrends refer to the shared systematic trends by all stars on a detector. There can be various causes of shared systematics, e.g. temperature variations, differential velocity aberration, residual spacecraft pointing errors, mechanical vibrations, electrical interference with other devices onboard, or injected artificial variability during post-processing. Cotrends depend on the position on the detector, but also on target-specific factors such as source’s brightness. The variability associated with cotrends is typically of the order of $\gtrsim 10$ days, and therefore primarily affects long-term observations than single transits. Identifying cotrends from a set of quiet targets and fitting them to every light curve on a detector is a common processing step in Kepler and TESS processing pipelines to reduce instrument noise and produce astrophysical light curves—e.g. through the PDC module (Stumpe et al., 2012; Smith et al., 2012, see e.g. Figure 2.6b).

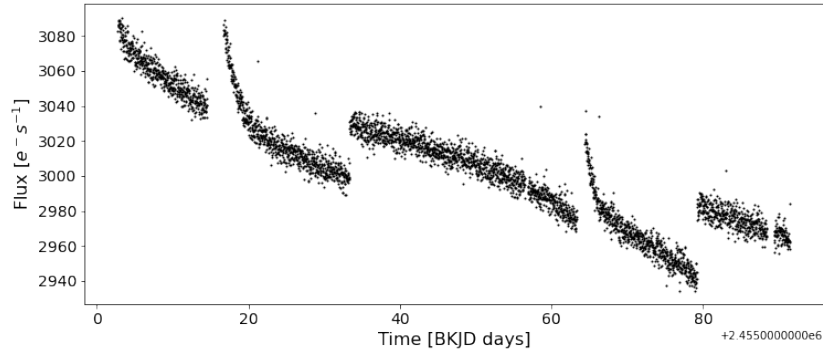
2.4 Stellar Variability

There are numerous reasons why a star’s brightness might vary. Some types of stellar variability affect the detectability and characterisation of exoplanets, whereas some active stars also dramatically affect their environment, therefore reducing the chances of life to develop. After all, even the transits or phase curves from small companions such as brown dwarfs and exoplanets can be considered as sources of extrinsic stellar variability.

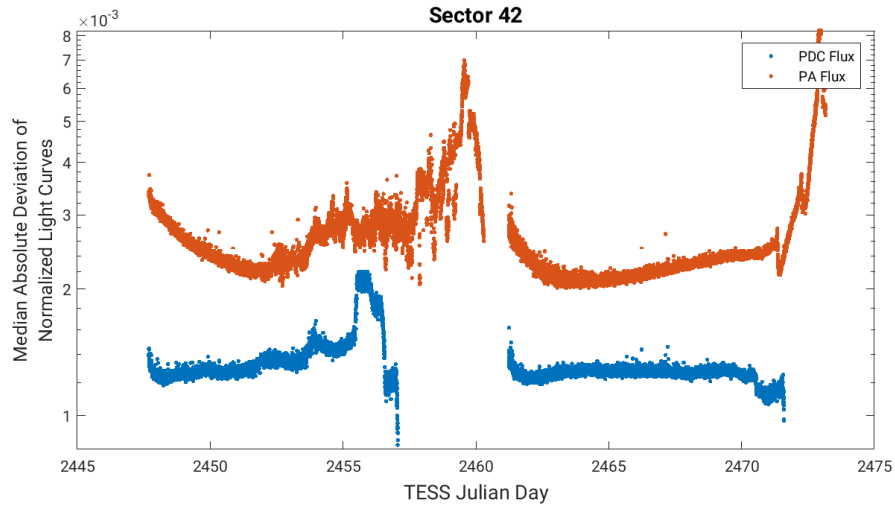
Monitoring nearby bright stars is a time-consuming and long term effort (to which many amateur astronomers have contributed). In particular, the Sun’s activity has been monitored for centuries through Sunspot counts, and more recently through dedicated space missions, thus providing a well studied example of its kind. Still, space-based stellar surveys have revolutionised variable stars astronomy, by providing high-precision and short cadence continuous monitoring of stars for long periods of time, without being only limited to the brightest stars and reducing human errors.

2.4.1 Types of variables

Variable stars are classified in a system with dozens of types and subtypes. At the top of the hierarchy there are two broad categories of *intrinsic* and *extrinsic* variables,

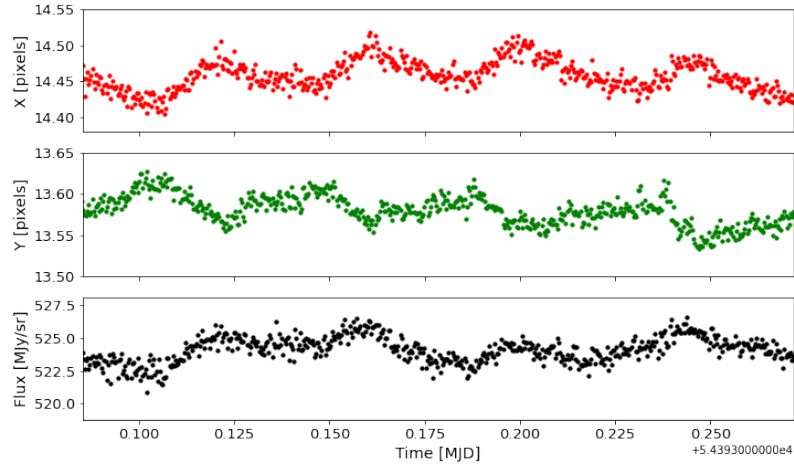


(a) SAP light curve of Kepler target 8703536 showing several common systematic effects present in Kepler data. The four main discontinuities are caused by various events, with from left to right: safe mode, pointing offset, Earth point and pointing offset; with safe mode and Earth point followed by thermal stabilisation ramps. Figure originally inspired from [Kinemuchi et al. \(2012\)](#) and reproduced using open Kepler data from the Mikulski Archive for Space Telescopes ([MAST](#)).

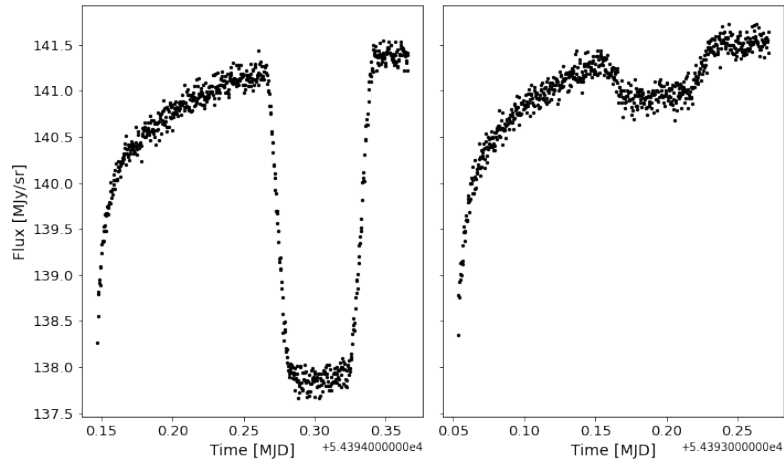


(b) Effect of cotrending on TESS data. The orange curve shows the mean absolute deviation (MAD, measure of scatter) of normalised 2-min light curves in Sector 42 after simple photometric analysis, while the blue curve shows a smaller MAD after cotrending the same light curves with the PDC pipeline. Plot from the TESS release notes for Sector 42 ([Fausnaugh et al., 2021](#)).

Figure 2.6: Examples of instrumental systematics from Kepler and TESS.



(a) Spitzer's pointing wobble originates from the pointing inaccuracy in X (top) and Y (centre) positions, and results in the raw light curve (bottom) correlated with the centroid positions. Photometry and centroid movements on this figure were obtained from observation ID 22807808 with IRAC channel 4.



(b) One transit (ID 22807808) and one eclipse (ID 22809344) of HD 189733 b observed at $8\mu\text{m}$ with Spitzer. The artificial ramps seen on the two light curves are thought to originate from charges trapped in the detector, and are commonly found in IRAC channels 3 & 4.

Figure 2.7: Examples of major types of instrumental systematics found in Spitzer/IRAC data, publicly available on the [SHA](#).

depending on whether the mechanisms of variability are internal to the star itself or are induced by dynamical effects. The intrinsic category is composed of the pulsating and the eruptive variables, and the extrinsic category is composed of the rotating stars and multiple systems. Each of this variability class is then further decomposed into many types and subtypes, with an imperfect and evolving nomenclature reflecting our limited understanding about them.

Pulsating variables expand and contract periodically with periods ranging from tens of seconds (pulsating white dwarves and subdwarf O-type stars) to thousands of days (long-period variables). More generally, the study of stellar seismic waves was given the name *asteroseismology*. The Sun itself pulsates at the level of $\sim 3 \times 10^{-6}$ mag (Barban et al., 2004). However, it is when they evolve through the instability strip that stellar pulsations tend to generate stronger variability ranging from milimagnitude to several magnitudes. One of the astounding achievements of the Kepler and K2 missions has been to provide asteroseismic information for hundreds of bright main-sequence stars and tens of thousands of red giants (Molnár et al., 2016).

Eruptive variables exhibit dramatic outbursts (e.g. novae and supernovae types) or fading events (e.g. R Coronae Borealis type) lasting a few days to a few hundreds of days. They originate in binary systems made of a main-sequence or giant star and a white dwarf star surrounded by an accretion disk.

Rotating variables feature changes in brightness due to their inhomogeneous rotating. Single hot stars of types A, B, and F, can have long-lived surface spots of various chemical compositions. These spots are long-lived (typically years to decades, Mathys et al., 2020), and should not be confused with surface spots of various temperatures seen in cooler active stars. These latter spots are more complex temporally as they can be relatively shorter (from one day to months, García et al., 2014; McQuillan et al., 2014) and cover a significant fraction of the star surface.

Eclipsing binaries (EB) are systems of binary stars eclipsing each other. Similarly to planets obstructing the light from or being eclipsed by their host stars, primary and secondary stars in binary systems can also eclipse each other when sufficiently aligned with respect to us. Analysis of the EB light curves can reveal information

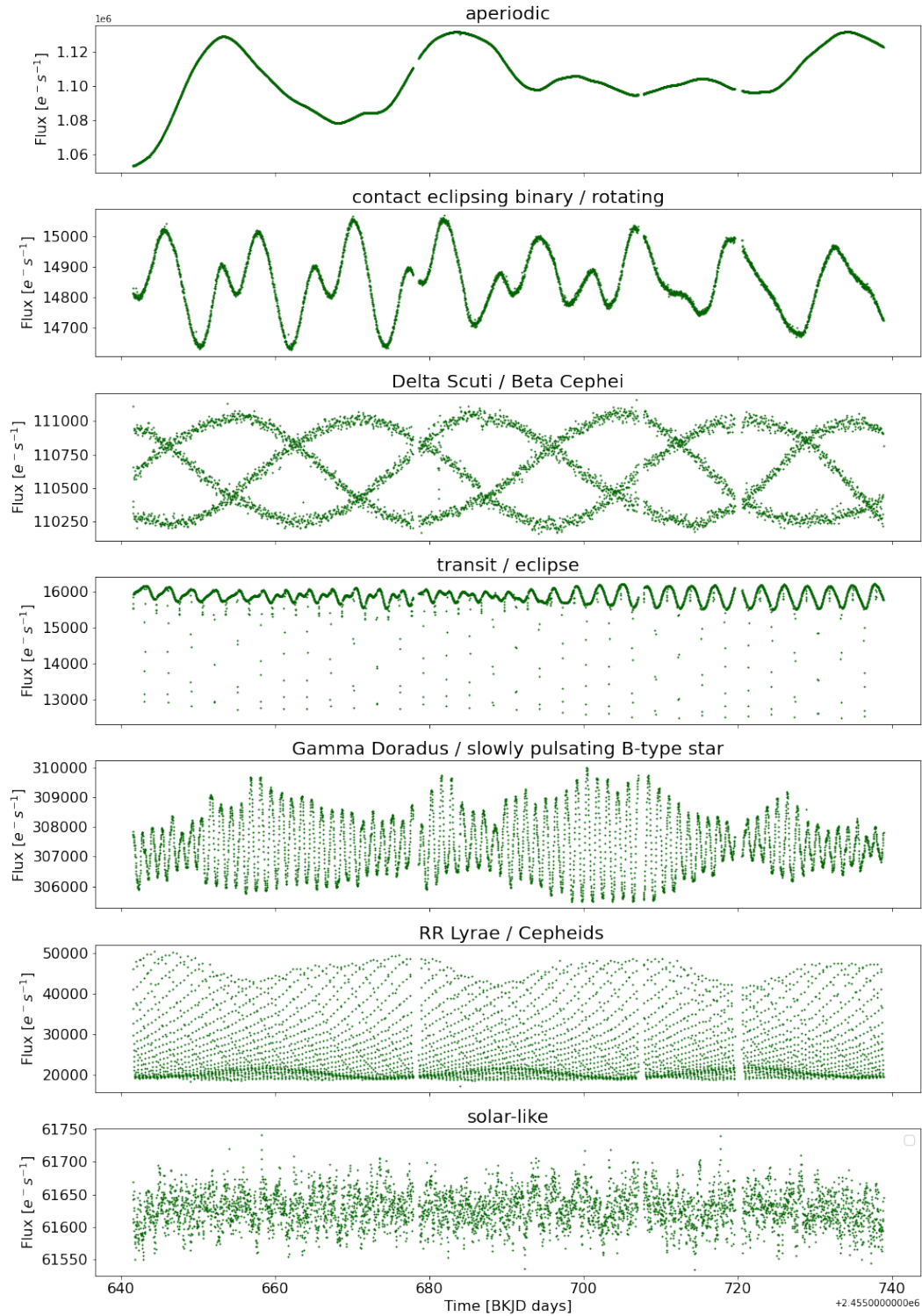


Figure 2.8: Kepler light curves for 7 different types of variable stars. From top to bottom: (i) aperiodic variable, (ii) rotational variable, (iii) δ Scuti variable, (iv) eclipsing binary, (v) a g-mode pulsator, (vi) Cepheid, (vii) solar-like pulsator. Figure produced using open Kepler Q9 data from MAST and variability classification from Audenaert et al. (2021).

about the orbital period, masses, star and orbit shapes of the system.

2.4.2 Synergies with exoplanets

The advent of space missions and surveys to search for exoplanets has resulted in millions of stellar light curves being analysed to look for transits. Incidentally, variable stars astronomy has greatly benefited from the unprecedented dataset of high-quality light curves provided by CoRoT, Kepler and K2. Conversely, the progress made in the understanding of variable stars benefits the search and study of exoplanets in various ways. Stellar variability can generate variability in time and wavelength. As such, the study of exoplanets can be affected in various ways: (i) detection is biased towards exoplanets around single quiet stars (e.g. [Gaidos and Mann, 2013](#)); (ii) transit parameters are likely affected by unaccounted or poorly understood variability. We discuss some of the synergies between the study of variable stars and the study of exoplanets.

Photometric variability thus needs to be corrected as well as possible, but without affecting the transit signal. This is one of the problems addressed in the present work (see Section 2.5 for an introduction to various detrending methods and Chapters 3 and 4 for two works aimed at addressing this).

Transiting planets are harder to detect around multiple star systems because their signal is diluted in the flux resulting from the multiple stars. Even if they are detected, the lack of knowledge on the eventual multiplicity of the host star will result in underestimating the planetary radius. In addition, eclipsing binaries remain the main source of confusion in vetting exoplanet candidates, as the eclipse of a stellar companion can easily mimic the primary transit of an exoplanet.

More than 400 years after Johannes Kepler mistook a sunspot for the transit of Mercury ([Kepler, 1871](#)), stellar spots are still affecting the search and characterisation of transiting planets—ironically in the data acquired by the Kepler Space Telescope. Short-lived stellar spots and faculae can create gaps or bumps in light curves, thus affecting the shape of exoplanet transits (e.g. [García et al., 2014](#)) and potential transmission spectroscopy (e.g. [Rackham et al., 2018a](#)). Even when unocculted, spots do not affect the shape of the light curve itself, but their inhomogeneous

distribution across the stellar surface will still generate bias on the transit depth when using the transit model defined in Section 2.2. This effect is less apparent and harder to directly account for from photometry only.

Asteroseismology can provide stellar mass, radius and age for stars too far away to be resolved with high-resolution spectroscopy. When done for the host stars of transiting planets, this reduction in stellar radius uncertainty is directly mirrored in the planetary radius uncertainty too ([Christensen-Dalsgaard et al., 2010](#)).

2.5 Detrending

In astronomical photometry, the sources of noise are various, instrument and star-dependent as seen in the previous sections. Furthermore, the associated photometric variability can hardly be modelled analytically. For these reasons the overall denoising and detrending of light curves proves challenging, and new instruments and datasets often require for new detrending methods.

2.5.1 A milestone in light curve processing

No matter what is the scientific objective, detrending is a cornerstone step in the analysis of photometric light curves. The task itself depends on the nature of the study carried out. In the context of stellar studies, detrending might indeed consist on the sole correction of instrumental systematics while preserving the stellar variability is paramount⁵. In the context of exoplanets study, however, all but the planetary signals have to be corrected. The most common problems on light curves—with a focus on the study of exoplanetary transits—are represented on Figure 2.9, organised hierarchically from general problems involving any stellar light curve on the top, transits in the middle, and specifically exoplanets on the bottom. These tasks can further be separated in two the broad ML tasks: classification and regression. Of particular importance is the fact that light curve denoising affects the quality of all the subsequent problems.

⁵Although even in that case, some noise sources of astrophysical origin might still need to be corrected, e.g. cosmic rays and stellar background; or disentangled, e.g. stellar companions.

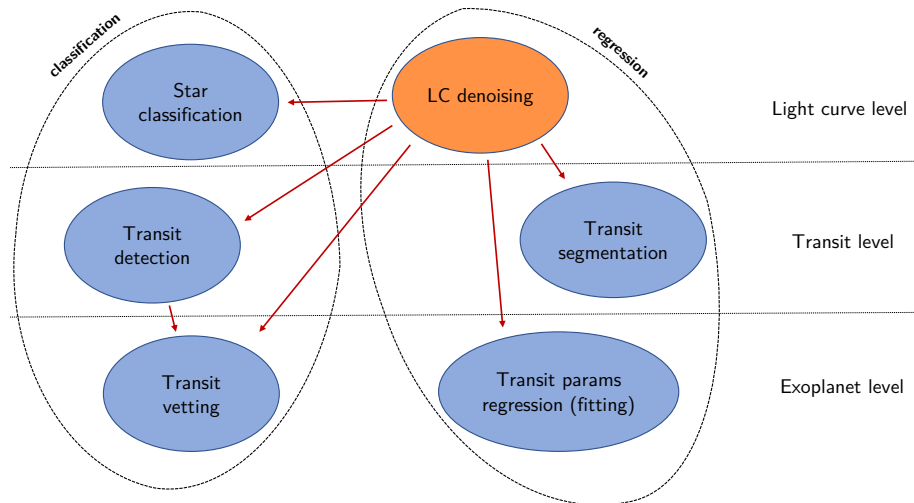


Figure 2.9: Diagram of common problems on transit light curves. Light curve (LC) denoising, which includes instrumental or stellar detrending and outlier removal, acts as a bottleneck step before all the other problems.

2.5.2 Detrending exoplanetary light curves

There are several classical approaches used to light curve detrending for exoplanet transit detection or fitting. We list the techniques relevant to this work below, grouping them by technique and intent⁶.

Early detrending methods on IRAC data were performing a polynomial fit on the photometric light curves after decorrelation using the centroid positions (e.g. Charbonneau et al., 2005; Knutson et al., 2008; Stevenson et al., 2012). These ideas were later adapted for Kepler and K2 detrending (Vanderburg and Johnson, 2014; Armstrong et al., 2015; Lund et al., 2015).

An alternative to using the centroid positions is to use pixel-level time series before photometric integration to decorrelate the signal from systematics (mainly pointing variations). Several methods were developed based on this idea, using independent component analysis (e.g. Waldmann et al., 2013; Morello, 2015) or pixel-level decorrelation (Deming et al., 2015; Zhang et al., 2018). The main techniques were compared by Ingalls et al. (2016) in an attempt to quantify their

⁶Various detrending algorithms may address different sorts of denoising depending on the data and the science objective.

absolute and relative performance on simulated data.

Gaussian Processes (GPs) offer a flexible, non-parametric, probabilistic method to simultaneously fit for systematics and transits, and have been used successfully on multiple instruments such as HST’s NICMOS⁷ (e.g. [Gibson et al., 2012](#)), IRAC (e.g. [Evans et al., 2015](#)), K2 ([Aigrain et al., 2015, 2016](#)). However, GPs also come with potential drawbacks, such as the difficulty to justify the choice of kernel, a $O(N^3)$ complexity with number of data points preventing large-scale application, and their tendency to overfit the systematics.

[Carter and Winn \(2009\)](#) proposed a method to model the systematics in the wavelet space and show how this can help to fit transits in individual observations. The idea of using frequency-based transforms to detrend light curves would later be employed for several other methods (e.g. [Morello et al., 2016](#)).

For wide-field surveys in particular, several methods are relying on linear regression from a set of quiet targets present in the same field of view. This is the case of the trend filtering algorithm ([Kovács et al., 2005](#); [Politsch et al., 2020b,a](#)) using a carefully chosen set of quiet stars on the detector, of Kepler PDC pipeline ([Smith et al., 2012](#); [Stumpe et al., 2012](#)), and causal decorrelation pipelines ([Wang et al., 2016](#); [Hattori et al., 2022](#)).

For blind transit search, a full detrending of instrumental and stellar variability is needed, and several classical methods have been used, often relying on sliding window-based filters, splines or ARIMA⁸ models (e.g. [Caceres et al., 2019](#)). [Hippke et al. \(2019\)](#) carried out an extensive comparison of most classical detrending techniques for transit detection in Kepler, K2, and TESS light curves.

2.5.3 Let machines learn the trend

Here we set out the grounds for a detrending approach where the trend would be generally learned with a neural network with minimal assumptions. In this respect this is a slightly different approach than the traditional fits mentioned in the techniques before.

⁷Near Infrared Camera and Multi-Object Spectrometer

⁸Autoregressive integrated moving average.

Although there is no clear formulation and benchmark in the machine learning world for the problem of time series detrending, there are a number of works which address some aspects of detrending indirectly to solve their tasks. These are not applied to astronomical light curves but can be relevant from a technical point of view. For instance, [Dai et al. \(2017\)](#) use a fully connected network to refine the smoothing module used in input, predict the time-varying trend, and show that this improves their forecasting performance; [Wang et al. \(2018\)](#) perform a trainable wavelet decomposition for interpretable time series forecasting; [Li et al. \(2021\)](#) use a variational autoencoder to smooth the data, learn the trend and perform anomaly detection; [Sun et al. \(2021\)](#) propose to model the autocorrelation parameter ρ to improve regression and forecasting on time series. [Woo et al. \(2022\)](#) design a specific attention mechanism⁹ which performs a time series decomposition inside the network and help forecast.

Although the importance of light curve detrending is widely recognised, there is arguably room for more coordinated effort to design optimal detrending solutions. This might involve but is not limited to setting up benchmark datasets and data challenges (see Appendix B), making simulators, data and detrending methods—articles and code—openly accessible, and carrying out comparisons between existing tools.

⁹See Section 4.2.1

Chapter 3

Autoregressive Light Curve Modelling with Recurrent Neural Networks

"We can only see a short distance
ahead, but we can see plenty there
that needs to be done."

Alan Turing,

Computing Machinery and Intelligence

This Chapter is chiefly based on [Morvan et al. \(2020\)](#): "Detrending Exoplanetary Transit Light Curves with Long Short-Term Memory Networks". The Astronomical Journal 159 (3): 109. It also features two applications of the TLCD-LSTM model used in [Yip et al. \(2020a\)](#) and [Saba et al. \(2022\)](#) in sections 3.4.4 and 3.4.5 respectively.

In this Chapter we propose a new detrending method to reconstruct the stellar flux with trends during transit. We train a probabilistic Long Short-Term Memory (LSTM) network to predict the next data point of the light curve during the out-of-transit, and use this model to reconstruct a transit-free light curve—i.e. including only the systematics—during transit. By making no specific assumption about the instrument, this provides a general way to correct for the systematics and perform a subsequent transit fit. The name of the proposed model is TLCD-LSTM, standing for Transit Light Curve Detrending LSTM. Here we present the results on data from six transits of HD 189733 b, two transits of WASP-96 b and two transits of WASP-17 b measured with the IRAC camera on board the Spitzer Space Telescope, and discuss some of its possible further applications.

3.1 Introduction

The total flux $F(t)$ received by a detector at time t can be broken down as follows:

1. Star flux: $F_*(t)$
2. Planetary signal: $\delta(t)$
3. Background stars and transient events: $F_b(t)$
4. Noise and instrumental systematics: $G(\cdot)$

The total flux received by each pixel of the detector can then be written as $F(t) = G((1 - \delta(t))F_*(t) + F_b(t))$, where F_* and F_b may vary depending on the position on the detector and are then subject to instrumental systematics. We will refer to individual pixel time series as *pixel light curves*, and to the summed contribution of pixels over time as a *raw light curve*.

One of the main instrumental systematic trend observed both with the Hubble WFC3 and the Spitzer IRAC cameras are the so-called *ramp* effect (Knutson et al., 2007), hypothesised to be due to the charge trapping in the detector (Agol et al., 2010), and intra-pixel and inter-pixel variations which are correlated with the position

of the source on the detector which shows variations in quantum efficiency across different pixels¹.

Footprints of these entangled variability sources can be found in additional instrumental data collected besides the detector raw flux. In particular, the centre and scale of the stellar point spread function (PSF) can be processed to give valuable information on the systematics while being mostly uncorrelated with the planetary signal itself.

Considering the analysis of time-correlated light curves with the end goal of detrending transit light curves and extracting the transit parameters as precisely as possible, one can approach the problem in several ways. Indeed, the disentanglement of various independent signals might naturally guide one toward *blind source separation* techniques, which have been applied on this problem (Waldmann, 2012; Morello et al., 2014, 2016) using the pixel light curves as correlated components. In a complementary way, signal processing analysis techniques have also been used to denoise the raw or pixel light curves, with *Gaussian processes* (e.g. Gibson et al., 2012), *pixel level decorrelation* (e.g. Deming et al., 2015) or *wavelet analysis* (e.g. Carter and Winn, 2009; Thatte et al., 2010; Morello et al., 2016). Here we choose the angle of *interpolation*, i.e. we want to provide predictions for the raw light curve during the transit time provided the out-of transit parts of the light curves. The interpolation method presented here is non-linear and thus capable of capturing complex long term dependencies in the light curve.

Here we make use of a *long short-term memory (LSTM) neural network* (Hochreiter and Schmidhuber, 1997) to interpolate the flux of a raw light curve during the transit, given additional time series data coming from the PSF centroid. The LSTM network learns to predict the next value of the light curve at each time step. The predictions of future time steps are then performed in a probabilistic manner using *ancestral sampling*, i.e. by injecting the current prediction as input to the subsequent prediction and so on. We thus assume that the pre-transit and post-transit information, along with additional data such as centroid time series, are sufficient to predict the

¹This effect is described in the IRAC instrument handbook: <http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac>

flux that the detector would have received in the absence of a planetary transit.

This paper is organised as follows: Section 3.2 contains background information about neural networks, Section 3.3 presents the interpolating model and how it can be used for transit light curve fitting, and finally Section 3.4 is dedicated to an application on Spitzer data.

3.2 Recurrent Neural Networks

In a typical *supervised statistical learning* task, the goal is to learn a model $h(x) \simeq y$ that maps an input x to an output y ² given examples of pairs (x, y) in such a way that the expected error of future predictions is minimised.

Fully connected neural networks or *multi-layer perceptrons (MLPs)* represent the simplest architecture of artificial neural networks. No feedback connections exist in these models. Every layer consists of a set of neurons and the neurons of the input layer represent each of the original input variables x . The output of each neuron is a scalar value and is used as input for the neurons of the next layer. Each subsequent layer transforms a linear combination of the outputs of the neurons of the previous layer using an *activation function* σ : $h_{l+1} = \sigma(W_l h_l + b_l)$ where W_l is a matrix of multiplicative weights, b_l the bias vector, h_l the vector of units and σ_l the *activation function*, all at layer l . If we interchangeably write h_l for the function represented at layer l as well as its output, the full function represented by a fully connected network can then be written: $y = h_D(h_{D-1}(\dots h_1(X)))$ where D is the *depth* of the network. Note that the use of non-linear activation functions is key to obtain a non-linear predictor.

The main characteristic of RNNs is that they allow for *recurrent connections*³. If we consider an input sequence $\{x_1, x_2, \dots\}$ of vectors, a recurrent hidden layer will thus process it sequentially, receiving at step t both the input x_t and other previous *hidden state(s)* in order to compute the current state h_t . A typical example is shown on Figure 3.1, where the recurrence occurs between the hidden units of the same

²Note that x and y can be scalars, or more generally n -dimensional vectors.

³This means that—unlike in fully connected neural networks—in RNNs the output of neurons from one layer can be used as input for neurons of the same or a previous layer.

layer: $h_t = h_t(x_t, h_{t-1})$. Compared to MLPs, RNNs allow us to reduce the number

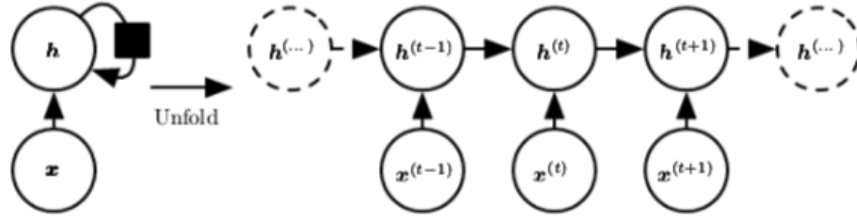


Figure 3.1: An example of a single-layer recurrent neural network with no output from Goodfellow et al. (2016). On the left is a folded recurrent cell (or layer) and on the right is the unfolded representation of the cell with recurrent connections flowing from left to right. In the presented model $x(t)$ represents an input vector containing the normalised flux value at t and covariate values—centroid positions at t in our case.

of parameters of the network by sharing weights between time-steps while seeking temporal patterns in the data. A variant of the standard unidirectional RNN described previously is called bidirectional RNN (Schuster and Paliwal, 1997) and consists in keeping two cells (non interacting) to process an input sequence both in natural and reverse directions.

In practice, several more sophisticated recurrent architectures are often more effective than vanilla RNNs, with most being variants of the long short-term memory (LSTM, Hochreiter and Schmidhuber, 1997) architecture whose cell is shown in Figure 3.2. LSTM networks have proven successful in a large range of applications including unconstrained handwriting recognition (Graves et al., 2009), speech recognition (Graves et al., 2013), machine translation (Brown et al., 2020), to cite only a few. An LSTM cell contains four different *gates* (see Figure 3.2), allowing the network to either retain or forget information from the past of the input sequence. This enables the relevant long-term time dependencies to be picked up more easily. The main addition in LSTMs compared to the basic RNNs has been to introduce self-loops, which are conditioned on the context and controlled by the gates. Below we state the detailed update formulae for the gates and states composing each LSTM unit:

- The *input gate*: $i_t = W_{ix}x_t + W_{ih}h_{t-1} + b_i$

- The *forget gate*: $f_t = W_{fx}x_t + W_{fh}h_{t-1} + b_f$
- The *output gate*: $o_t = W_{ox}x_t + W_{oh}h_{t-1} + b_o$
- The *cell state*: $c_t = \sigma(f_t) \odot c_{t-1} + \sigma(i_t) \odot \tanh(j_t)$
- The output vector: $h_t = \sigma(o_t) \odot \tanh(c_t)$

Where t denotes the time step, W_{ab} the matrix of weights relative to the vectors a and b , b_a the bias vector relative to a , \odot the Hadamart (i.e. entry-wise) product and σ is the activation function, typically a logistic, sigmoid or tanh function.

Incidentally, these types of gated RNNs also have the advantage of being easier to train than basic RNNs, by alleviating the well known vanishing or exploding gradient issue (Kolen and Kremer, 2001).

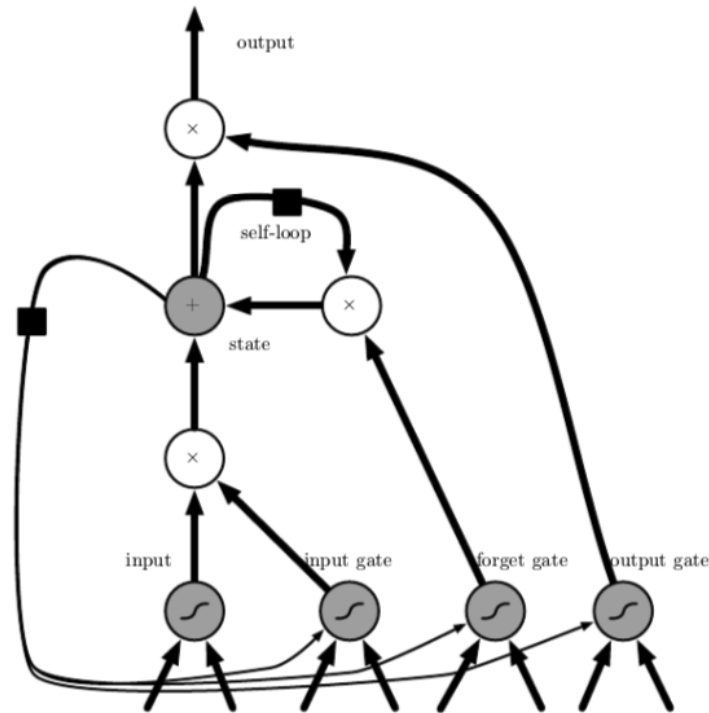


Figure 3.2: An LSTM cell from Goodfellow et al. (2016), which replaces a usual hidden unit (i.e. neuron) in a fully connected neural network. The input, forget and output gating units enable the cell to accumulate or shut off respectively the current input, long-term dependencies and output through a sigmoid activation function. The square indicates a delay of one time step, and circles indicate simple mathematical operators such as additions, multiplications of sigmoid functions.

3.3 The Transit Light Curve Detrending LSTM

Model

Here we describe the proposed model to interpolate a time series on a pre-defined prediction range. As the final goal of this paper is to study the transit signal contained in the interpolation range after correction of the systematic errors, we name the method *Transit Light Curve Detrending LSTM (TLCD-LSTM)*.

The model is inspired from the deep autoregressive neural network model described in [Salinas et al. \(2017\)](#). It assumes that temporal relations exist in the time series and learns to predict the next step in the training range of the input time series. It can also make use of additional data available for prediction contained in the so-called *covariate time series*, which is to be distinguished from the main time series to be predicted. In general, one can consider both the main and covariate time series to be multi-variate, i.e. to be composed of several time series each.

The TLCD-LSTM is specifically adapted for interpolation within a given range, and therefore differs from [Salinas et al. \(2017\)](#) mainly in that the values it predicts are not in the future (i.e. the end of the time series) but in time steps somewhere within the time series.

3.3.1 Model description

Let us denote with $\{x_1, x_2, \dots, x_T\}$ (abbreviated $\{x_t\}$) the *main time series* of length T we ought to interpolate on the prediction range $[t_1..t_2]$ with t_1 and t_2 integers in $[1..T]$, and $\{z_1, z_2, \dots, z_T\}$ (abbreviated $\{z_t\}$) the time series of covariates, which constitute additional data available for prediction on the whole time range. Finally, let us also denote with $\{y_1, y_2, \dots, y_T\}$ (abbreviated $\{y_t\}$) the *target time series*, identical to the main time series in the training range but which may differ in the prediction range. In the case of $\{x_t\}$ being a transit light curve, $\{y_t\}$ is the hypothetical light curve without any transit signal.

As sketched in Figure 3.3, each value of the input time series passes through a stack of LSTM layers, the output of which branches into two distinct fully connected layers outputting two parameters $\hat{\mu}_t$ and $\hat{\sigma}_t$ at each time-step, which are the predicted

mean and standard deviation for the distribution of the current value x_t , respectively. The same network is used both for the training and prediction ranges with only the

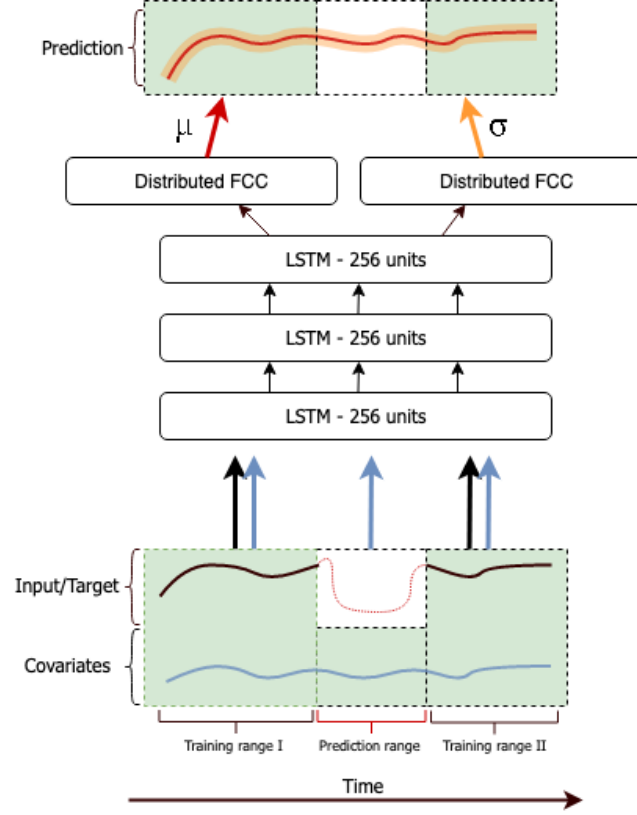


Figure 3.3: Overview of the interpolating probabilistic TLCD-LSTM network. The main and covariate time series are processed through three LSTM layers consisting of 256 units each, and then decoded into two outputs for each of the interpolated points: the mean and the standard deviation.

inputs differing in each case.

Mode	Range	Inputs at t	Output at t
Training	$[1..t_1 - 1] \cup [t_2 + 1..T]$	x_{t-1}, z_t	$(\hat{\mu}_t, \hat{\sigma}_t)$
Prediction	$[t_1..t_2]$	$\widehat{\mu}_{t-1}, z_t$	$(\hat{\mu}_t, \hat{\sigma}_t)$

Table 3.1: Summary of training and prediction input/output modes with respect to the time t .

At each time step t , the network predicts the current value x_t from all past time steps x_1, \dots, x_{t-1} as well as from the current covariate z_t . While the actual previous time series value x_{t-1} is used as input in the training ranges, in the prediction range the previous prediction μ_{t-1} is injected as an input instead of it (see Table 3.1).

3.3.1.1 Training the model

We assume each value z_t is sampled from a normal distribution:

$$z_t \sim \mathcal{N}(\mu, \sigma^2).$$

The loss function is then computed as the product of individual likelihoods outside the prediction range:

$$\mathcal{L}(\mu_{t_1..t_2}) = \prod_{t \in [1..t_1-1] \cup [t_2+1..T]} \frac{e^{(x_t - \hat{\mu}_t)^2 / 2\hat{\sigma}_t}}{\sqrt{2\pi\hat{\sigma}_t}}.$$

Note that the log-loss is only computed in the training ranges. However, the last output of the prediction range is taken as the first input of the second training range, thus providing a way to link together the outputs in the different ranges.

3.3.1.2 Generating predictions

There are several ways one can generate predictions in the interval $[t_1..t_2]$, once a model is trained. Since the network's outputs consist of parameters of a probability distribution, the simplest one is to directly take the vector of predicted means $\hat{y}_t = \mu_t$. However, one can also generate a *trace* by drawing every value from the Gaussian distribution at every time step in the prediction range: $\hat{y}_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$, and injecting each of these predictions as input for the next time step. Multiple traces obtained with this process then represent the joint predicted distribution (of which they are samples) in a more general way than merely using the vector of mean values. To generate a single vector of predictions from multiple traces, one can select the median or mean value at every time step to construct the *median trace* or *mean trace* on the prediction range. In Section 3.4, we focus on the simplest approach, i.e. selecting the output means and standard deviations.

3.3.1.3 Covariate features

The covariates time series $\{z_t\}$ can consist of single-dimensional or multidimensional data available both in the training and prediction ranges. It is used by the network as additional information besides the target time series. This works merely by concate-

nating x_t (conversely \hat{x}_t in prediction mode) to the covariate data z_t to construct the new input to the network at every time step. Ideally, one wants $\{z_t\}$ to be correlated with the target time series. Several time series might be related to the time-correlated noise we intend to correct, and therefore can be used as covariate data in the model. In the application presented in Section 3.4 we suggest the use of PSF-related time series, namely the instrument’s PSF centres and widths of a 2D Gaussian fit on the images at every time step. One could also think of other potentially relevant information such as simultaneous host star activity, calibration data relative to the detector and estimations of background flux. For ground-based applications, information about airmass, seeing and weather patterns could be included.

3.3.2 Application to transit light curves

Here we discuss the use of the interpolating model specifically to transit light curves.

The transit signal must be contained within the prediction range. This requires either to know beforehand when the transit occurs, or to adapt the prediction range during the first phase of the training. Pre-transit and post-transit data are used for training the network, and are assumed to not be contaminated by any transit event. They can however contain any sort of variability coming from the star, the background or the instrument. In fact, the model aims at picking up variations due to all sources other than a transit event in order to predict the flux due to these sources alone during the transit time.

We perform a transit fit at each evaluation step even though our model does not strictly require it for the training. This is done for two main reasons:

1) The transit fit can be used as a proxy to evaluate the quality of the prediction and provides us with a criterion for early-stopping the training of our model. The transit fit is performed on the detrended light curve normalised with respect to the star ($1 - \delta_t$). To obtain a light curve normalised with respect to the star, three steps are required: transformation to the original units $y_t \rightarrow y'_t$, subtraction of background flux $F_b(t)$, and division of the background subtracted raw light curve by the predicted star flux:

$$1 - \delta(t) = \frac{F_{received}(t) - F_b(t)}{F_*(t) - F_b(t)}.$$

With the time series notations where x'_t and \hat{y}'_t are the input and mean prediction of the neural network in the original units:

$$1 - \delta_t = \frac{x'_t - F_{b,t}}{\hat{y}'_t - F_{b,t}}.$$

2) We can use the transit fit to adapt the prediction range $[t_1..t_2]$ during training so that it matches better the actual transit range of the data. This can be done by extracting the fitted mid-transit time and transit duration to compute the times for the beginning and end of transit.

3.4 Applications

Here I describe the different preprocessing and hyperparameter optimisation applied to the raw subarray data for HD 189733 b, before presenting applications for HD 189733 b, WASP-17 b and WASP-96 b.

3.4.1 Preprocessing

Outlier removal: Due to a number of causes, such as remaining cosmic rays or bad pixels, the flux on individual pixels can exhibit large fluctuations within short timescales ($\approx 1\text{sec}$). These abnormal values are identified by computing the absolute difference of the pixels' flux with their corresponding median within a time window of width 5 (2 sec) exposures. The values of the median-subtracted time series greater than 4σ away are then replaced by the median values, where σ is the standard deviation of the time series.

Raw light curve extraction: In order to limit the influence of background light and focus on the brightest pixels of the stellar PSF, 3×3 , 5×5 and 7×7 pixel regions are extracted around the brightest pixel. The raw light curve is then obtained by summing all the individual pixel light curves.

Centroid fitting: As mentioned earlier the centroid position time series are highly correlated with the flux received by the detector. In order to compute the centroids, we perform a two-dimensional Gaussian fit with offset to the data at every time step, and hence extract four useful time series, two of which are monitoring the position

of the centre on the detector and two for the width of the Gaussian. As discussed in Agol et al. (2010), this method provides by far a better estimate of the centroids over other methods such as the flux-weighted ratio extraction.

Background extraction: The background flux contribution to the total flux, although minor, increases with the aperture size used for the light curve extraction. We estimate it here by taking the median flux value of the pixels located in the four corners of each frame, corners delimited by the complement of a circular aperture of radius 16. It accounts for 0.67% to 1.2% in our analysis of HD 189733 b, and should therefore be taken into account. However, as the background estimation is necessarily approximate, we advocate to still interpolate on the raw light curve directly, and only correct for it just before the transit fit.

Normalisation: The raw light curve and centroid time series are all locally standardised, i.e. individually centred around a mean value of zero and rescaled to have their standard deviation equal to one.

The preprocessed raw light curves and centroid X/Y positions are shown on Figure 3.5. Note the diversity of effects among them, showing more or less stochastic noise, ramps or jitter.

3.4.2 Tuning & testing

Since the ground truth—i.e. the stellar and instrumental flux without transit—is not available during transit, we first train and validate our model first on the pre-transit part of HD 189733 b’s light curves instead, containing the first 250 time steps as shown on Figure 3.4. We mask a prediction range (as shown inside the two red lines) that is used for validation while only the regions before and after this range are used as training data. Predictions (in orange on Figure 3.4) are obtained by averaging 50 traces sampled from the model output distribution.

Hyperparameter optimisation

We perform a grid search over different types of inputs and hyperparameters using the train/validation split presented just above. More specifically, we vary the aperture width of the sub-array used for computing the raw light curve between 5 and 7 pixels; we experiment with including and excluding covariate features, namely: 1) excluding

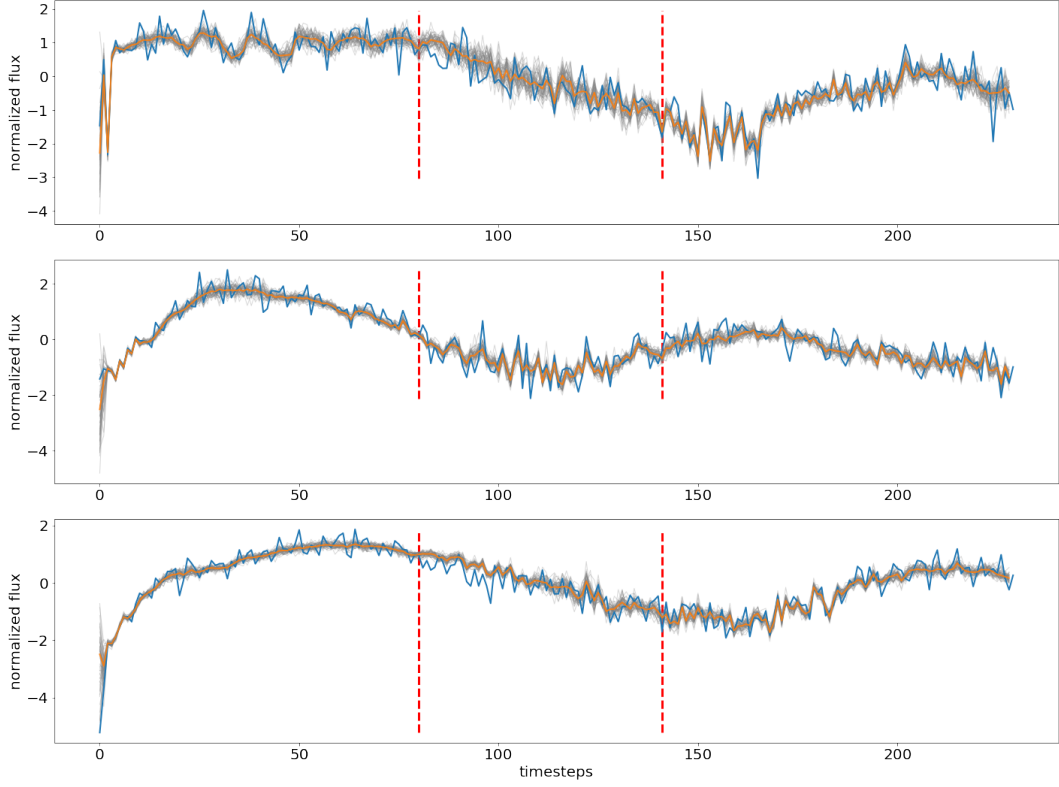


Figure 3.4: Example of interpolations on 3 light curves containing no transit for validation and hyperparameter tuning. The prediction range is located inside the vertical dashed lines. The raw light curve is displayed in blue, the predicted traces in grey and the median prediction in orange.

covariate features altogether; 2) including centroid time series, and 3) including centroid and PSF width time series. Furthermore, we vary the number of layers (between 1 and 4), units per layer (powers of 2 up to 1024 and dropout rate (between 0% and 50% in steps of 1%)⁴ values for the LSTM block; and a unidirectional or bidirectional network⁵. We train each different model on the 6 light curves and 3 different prediction ranges, monitoring the average MSE for these 18 predictions and using it as a criterion for early stopping and comparison between the different models.

From these tests we observe the following:

- Including the centroid information improves the quality of the prediction by a

⁴Dropout is a common regularisation technique in deep learning consisting in randomly reinitialising a fraction of the neurons of a given layer. The dropout rate refers to this fraction.

⁵In this ‘bidirectional’ implementation we simply average the outputs of two models processing the input sequence in natural and reverse orders.

factor of ~ 2 , and including the PSF widths time series besides the centroids brings a further increase in MSE.

- Dropping 3% of the recurrent units improves slightly the predictions, especially when the number of parameters of the network increases.
- Using a bidirectional network slightly decreases the quality of predictions.

Training was performed using the ADAM optimiser (Kingma and Ba, 2015) with parameter values $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\varepsilon = 10^{-8}$. The learning rate was decreased from 0.01 to 0.0001 using a polynomial decay law with exponent 20. We train the model using a batch size of 6 (all the light curves) for faster training.

Parameter	Value
Number of LSTM layers	3
Number of units per layer	256
Recurrent drop-out rate	3 %
Initial bias values	0.0
Batch size	6

Table 3.2: Model’s original hyperparameters tuned for HD 189733 b light curves.

3.4.2.1 Performance

We present in Table 3.3 the results of the best tested model in the explored grid. As a reference for the performance of the interpolation, we include a baseline model, which is a linear composition of the centroid X/Y time series $\{z_t^X\}$ and $\{z_t^Y\}$:

$$\hat{y}_t = a + bz_t^X + cz_t^Y,$$

where $a, b, c \in \mathbb{R}$. The model is trained on the training ranges⁶ and evaluated in the prediction ranges. The metrics computed for both models include the MSE and the mean-absolute error (MAE), defined as:

$$MSE = \frac{1}{N} \sum_{t_1}^{t_2} (\hat{y}_t - y_t)^2.,$$

⁶The model was fitted using scikit-learn’s linear regression module: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

$$MAE = \frac{1}{N} \sum_{t_1}^{t_2} |\hat{y}_t - y_t|,$$

where N is the number of observations, and σ_{noise} an estimate of the noise level computed by taking the mean value of the running standard deviation of width 15 over each input light curve.

LC instance	Baseline	This model
MSE #1	0.192	0.156
MSE #2	0.782	0.196
MSE #3	0.606	0.138
MSE #4	0.0921	0.0529
MSE #5	0.275	0.249
MSE #6	0.688	0.0837
Mean MSE	0.439	0.124
Mean MAE	0.503	0.367

Table 3.3: Comparison of performance on the 6 light curves for each model. Every value is averaged between prediction and actual value over 3 different ranges of length 60 and starting respectively at time steps 80, 100 and 120. The three last lines show the mean performance over all light curves and ranges in terms of MSE, MAE and SNR.

Despite its simplicity, this baseline model provides a satisfactory interpolation $\{x_t\}$, and this is why it was chosen here as a reference for the MSE. Furthermore, since the TLCD-LSTM also uses the centroid time series, the increase in performance seen on Table 3.3 can directly be interpreted as the improvement brought by the LSTM’s ability to identify temporal dependencies in the raw light curve.

3.4.3 HD 189733 b

We present an application to 6 transit observations of planet HD 189733 b from the Spitzer/IRAC detector at $8 \mu\text{m}$, collected in 2007 and 2008 (PI: E. Agol, program: 40238). This hot-Jupiter planet has been extensively studied and makes a good candidate for bench-marking our method. In this wavelength channel, the ramp effect can be heavily pronounced (Agol et al., 2010), while the intra-pixel variations due to pointing jitter are less important than at shorter wavelengths.

Results Using the optimised hyperparameters listed in 3.4.2 and after training the model for 3000 epochs we extract the output of the network for the whole time

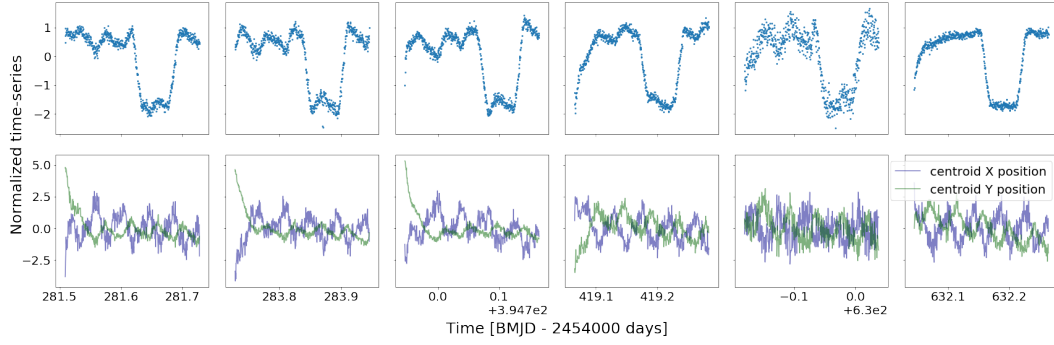


Figure 3.5: (Top) 6 Spitzer/IRAC $8\mu\text{m}$ raw transit light curves of HD 189733 b after preprocessing. (Bottom) X/Y centroid positions of the PSF.

ranges, shown in red on Figure 3.6. Note that the decreasing learning rate used guarantees the convergence of the network towards a stable solution. Visually, the model seems to be able to pick up the trends and variability of each time series, while joining smoothly the pre-and-post-transit ranges where the ground truth is known.

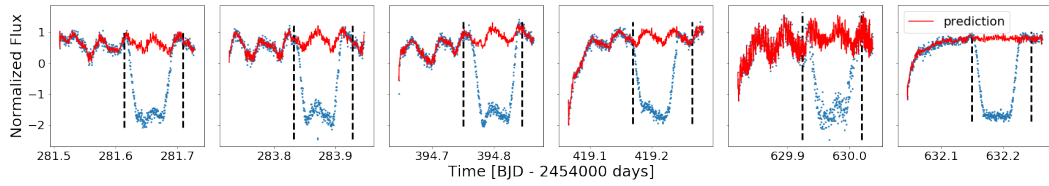


Figure 3.6: Raw data (blue) and model output, i.e. interpolated light curve in the absence of transit (red) for the light curves. Dashed vertical lines indicate the initial prediction ranges.

The last step is to perform the transit fit on the detrended light curve $\{1 - \delta_t\}$ normalised with respect to the stellar flux $F_*(t)$. Since the limb darkening effect is minor at $8\mu\text{m}$, we chose a transit model with linear limb darkening (bound between 0.05 and 0.25), and compute the best fit using a Markov Chain Monte Carlo optimisation procedure⁷ (e.g. Tsiraras et al., 2016). The fitted parameters are R_p/R_s , the mid-time transit time t_c , a linear limb darkening coefficient u , the orbit inclination i and orbital semi-major axis relatively to the stellar radius a/R_s . The fitted model, residuals and auto-correlated functions (ACF) are shown in Figure 3.7 and the fitted parameters are presented in Table 3.4. The higher variance present in the residuals of the 5th light curve is due to a higher noise level in the input data for this light curve.

⁷The transit model was fitted using PylightCurve package:

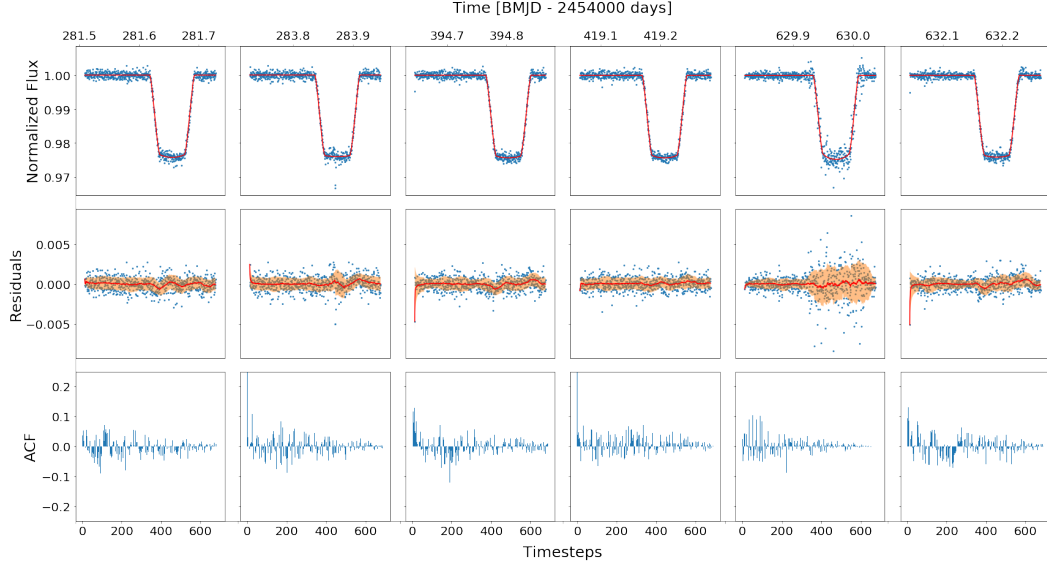


Figure 3.7: (*Top*) Best transit fit (red curve) to the detrended light curve (blue points) normalised with respect to the stellar flux. (*Centre*) Fit residuals (blue points) along with the moving average (red curve) and standard deviation (orange) of the residuals. (*Bottom*) Auto-Correlated Function of the residuals

t_c (BJD-2454000)	R_p/R_*	i (deg)	a/R_*	u
281.655329 ± 0.000046	0.15489 ± 0.00018	85.7682 ± 0.0502	8.971 ± 0.045	0.141 ± 0.020
283.873934 ± 0.000049	0.15477 ± 0.00024	85.7277 ± 0.0698	8.901 ± 0.069	0.051 ± 0.018
394.802829 ± 0.000045	0.15564 ± 0.00020	85.5926 ± 0.0771	8.799 ± 0.065	0.093 ± 0.037
419.206955 ± 0.000070	0.15520 ± 0.00015	85.8120 ± 0.0881	8.992 ± 0.075	0.129 ± 0.026
629.971770 ± 0.000097	0.15523 ± 0.00042	85.9760 ± 0.1263	8.999 ± 0.106	0.248 ± 0.028
632.190498 ± 0.000046	0.15488 ± 0.00019	85.5862 ± 0.0747	8.782 ± 0.057	0.097 ± 0.028

Table 3.4: Fitted physical parameters for each of the 6 HD 189733 b transits.

We compare the retrieved transit depths with the results published in [Agol et al. \(2010\)](#) for the same data set and preprocessing steps (Figure 3.8). Although slightly smaller, predictions are still scattered with a standard deviation of 91.7ppm instead of 144ppm. The mean weighted by the standard deviations of the 6 transit depths is also found to be slightly smaller in our case by 94ppm $\approx 4\sigma$. Since variations in planetary flux are not expected to account for such a large transit depth variation (see e.g. §6.6 in [Agol et al., 2010](#)), this scatter is likely to be either of stellar origin or artificially introduced by the detrending method. In that respect, the TLCD-LSTM seems to provide more consistent data processing across the multiple epochs, but further high-precision observations will be needed at this wavelength to resolve the

remaining discrepancy and reduce the uncertainty on the radius⁸.

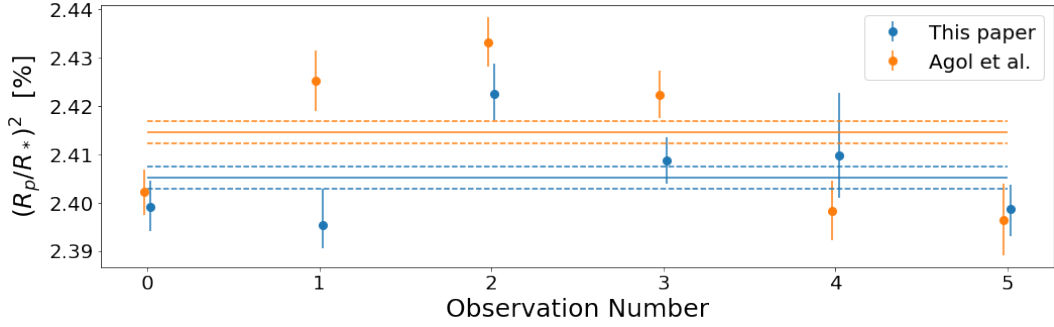


Figure 3.8: Comparison of fitted transit depths between this work and Agol et al. (2010), for the six transit observations of HD 189733 b. The horizontal lines show the means of the observations from both papers weighted by their respective standard deviations. The dotted lines show the standard deviations of these weighted means.

Finally, we present the results for additional transit parameters on Figure 3.9.

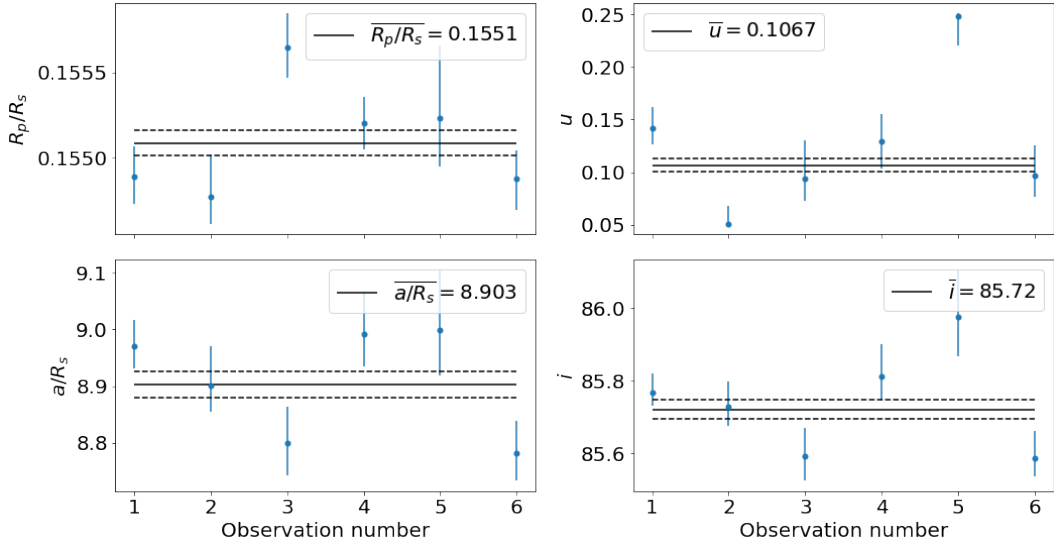


Figure 3.9: Additional plots showing the fitted parameters R_p/R_s , u , a/R_s and i for each of the 6 light curves, as well as their weighted mean and associated standard deviation.

The performance of the TLCD-LSTM method on this dataset is encouraging as it reduces some of the unanswered transit depth variability after detrending, and the open-source access of the code makes it possible for anyone to reproduce these results or use the method on different datasets. The chosen hyperparameters—optimised for

⁸HD 189733 b is featured in the first cycle of JWST’s General Observers Program.

this dataset—are expected to provide good results for similar datasets, but it might be worth fine-tuning some parameters if the data varies largely. More generally, the TLCD-LSTM is expected to work well whenever there is sufficient out-of-transit data (a rule of thumb would be similar or longer out-of-transit baseline compared to the transit duration). However, more tests between multiple detrending methods will be needed on simulated data to determine which methods are most effective in different scenarios. Infrared observations with JWST will also help confirm which results are more likely to be correct and refine our detrending methods.

3.4.4 WASP-96 b

WASP-96 b is a transiting Jupiter-sized and Saturn-mass planet discovered by [Hellier et al. \(2014\)](#) as part of the WASP-South survey. It orbits a Sun-like G8 star with a semi-major axis of 0.0453 ± 0.0013 au, making it a hot-Jupiter (or hot-Saturn, see e.g. [Nikolov et al., 2018](#)) planet. Various stellar and orbital parameters used throughout this study are shown on Table 3.5.

Parameters	Value
$R_s [R_\oplus]$	$1.05 \pm 0.05^*$
$M_s [M_\oplus]$	$1.06 \pm 0.09^*$
$T_s [K]$	5540 ± 140
$M_p [M_{Jup}]$	$0.48 \pm 0.03^*$
$R_p [R_{Jup}]$	$1.20 \pm 0.06^*$
$T_{Eff} (K)$	$1285 \pm 40^*$
a/R_s	$8.84 \pm 0.1^\dagger$
$i [deg]$	$85.14 \pm 0.2^\dagger$
$P_{orb} [days]$	$3.4252602 \pm 0.0000027^*$
$T_{mid} [BJD_{TDB}]$	$2456258.062876 \pm 0.0002^*$
[*] Hellier et al. (2014)	[†] Nikolov et al. (2018)

Table 3.5: Details of the WASP-96 system used in this study. Table adapted from [Yip et al. \(2020a\)](#).

[Yip et al. \(2020a\)](#) presented a case study on the compatibility of observations from multiple instruments. In particular, we reanalysed observations of WASP-96 b from Hubble WFC3, VLT’s FORS2 spectrograph, Spitzer IRAC channels 1 and 2 and TESS to perform an experiment on their joint contribution to the atmospheric modelling of WASP-96 b.

My main contribution to this study was to detrend the two IRAC observations of WASP-96 b obtained during the program ID 14255 and using the TLCD-LSTM. The architecture is the same as for the case of HD 189733 b study except for the number of hidden units reduced to 64 and dropout rate set to 0.1 in order to prevent over fitting on the out-of-transit. Figure 3.10 shows the detrended light curves and the best-fit model to the data. For both data sets, the only free transit parameters were the planet-to-star radius ratio and the transit mid-time, with the other model parameters fixed to those in Table 3.5. We also fitted the detrended light curve while allowing the planet semi-major axis to star radius ratio (a/R_s) and inclination to vary. For both channels, the retrieved values for the transit depth and epoch remain very close to the ones found with the two orbital parameters fixed to the values from Nikolov et al. (2018). Furthermore, values retrieved for the semi-major axis and inclination shown in 3.6 are compatible with those from Nikolov et al. (2018) to 1σ . No autocorrelation appear in the residuals of both channels, but note that the root-mean-squared error RMSE of the residuals is smaller for channel 2 by 1.4 mag.

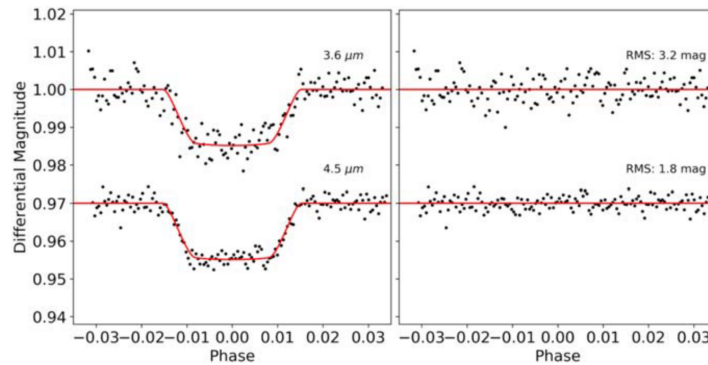


Figure 3.10: Fitted Spitzer transits of WASP-96 b for IRAC channels 1 (top) and 2 (bottom). Left: detrended light curves and best-fit model. Right: residuals from the fit. Figure reproduced from Yip et al. (2020a).

Parameter	FORS2 Nikolov et al. (2018)	TESS Yip et al. (2020a)	IRAC C1	IRAC C2
a/R_s	8.84 ± 0.1	$8.85^{+0.62}_{-0.10}$	$8.66^{+0.14}_{-0.12}$	$8.78^{+0.06}_{-0.06}$
i [deg]	85.14 ± 0.2	$85.55^{+0.39}_{-0.34}$	$85.36^{+0.14}_{-0.13}$	$85.21^{+0.07}_{-0.06}$

Table 3.6: Comparison of WASP-96 b orbital parameters between different datasets from Yip et al. (2020a).

The study concludes by warning over combining observations from different

instruments, especially when their wavelength coverage do not overlap. In addition, since Spitzer IRAC bands cover spectral regions where carbon-bearing molecules such as CH_4 , CO , and CO_2 absorb, a miscalibration of the instrument or combination with other observations could lead to wrong estimates of the C/O ratio.

3.4.5 WASP-17 b

WASP-17 b is one of the least dense hot Jupiters known to date (Anderson et al., 2010). In the study conducted by Saba et al. (2022), we downloaded the WASP-17 b transit observations taken with IRAC channel 1 and 2 in 2013 (P.I. Jean-Michel Desert, programme 90092). The two observations were processed separately using the TLCD-LSTM⁹. First, photometric light curves are extracted by using a circular aperture around the centre of light (with a radius of 3.25 pixels) and subtracting the background light computed at each time step using the median flux in an annulus of inner radius equal to 15 pixels. The neural network is then trained only on the pre-ingress and post-egress parts of the light curves (training set) using 2-D Gaussian centroid time series in order to learn only the systematics while excluding the signal. However, intermediate transit fits are performed during training on the temporary detrended light curve in order to keep track of the model’s progress and provide an additional stopping criterion. A first run is done with large margins around the expected ingress and egress times to ensure that the transits do not overlap with the training ranges. The margins are then refined after this first run to include 105 % of the transit duration centred on the mid-transit time. In practice, the results are very stable and unaffected by the chosen margins provided these remain low (under $\approx 10\%$ of the transit duration). Several architecture and learning parameters are then compared, starting with the hyperparameters optimised for HD 189733 b presented in Section 3.4.3. The data used for validation are the 5% of time steps around the transit, and the test set corresponds to the time steps during transits. For both channels we find that 2 LSTM layers of 512 units and 10% dropout trained with an Adam optimiser and decay rate of $\beta = 0.95$ provide an optimal residual noise on the validation data while avoiding overfitting. After the model was trained for 50

⁹<https://github.com/ucl-exoplanets/deepARTransit>

epochs, the best network—i.e. with the lowest residual noise—is saved along with the corresponding detrended light curve, obtained by dividing the raw light curve by the network’s prediction. A final transit fit is performed using a Markov Chain Monte Carlo procedure embedded in the PyLightcurve package. Fixed planetary parameters are those listed in Table 3.7 or computed using the ExoTETHyS open package (Morello et al., 2020) while the mid-transit time and R_p/R_* were let free.

Stellar & Planetary Parameters	
T_* [K]	6550 ± 100 *
$\log_{10}(g)$ [cm/s^2]	4.24 ± 0.13 *
[Fe/H]	-0.25 ± 0.09 *
a/R_*	$7.27^{+0.21}_{-0.44}$
e	0 (fixed)
i [$^\circ$]	$87.96^{+1.34}_{-1.56}$
ω	0 (fixed)
P [days]	$3.7354845 \pm 1.9 \times 10^{-6}$ †
*Anderson et al. (2011)	
†Southworth et al. (2012)	

Table 3.7: Stellar and planetary parameters for WASP-17 b used for the data reduction with Iracis and TLCD-LSTM, and the atmospheric modelling with TauREx3. Table adapted from Saba et al. (2022).

Raw, detrended, fitted, and residual light curves are shown on Figure 3.11. Most of the time-correlated noise is accounted for by the TLCD-LSTM detrending, especially the jitter noise present in Channel 1’s observation.

Saba et al. (2022) then used the fitted transit depths from detrended IRAC observations as well as data from Hubble WFC3 and STIS to perform a point atmospheric retrieval using TauREx retrieval framework (Al-Refaie et al., 2021). We performed multiple careful experiments to evaluate the influence of each instrument and stellar activity assumption on the results. Regardless of the stellar activity being modelled or not and which instrumental dataset is included, the analysis finds presence of water in the solar/super-solar regime and possible traces of AlO and TiH in the atmosphere of WASP-17 b. The atmosphere of the inflated hot Jupiter appears to be free from both an optically thick layer of grey clouds and hazes.

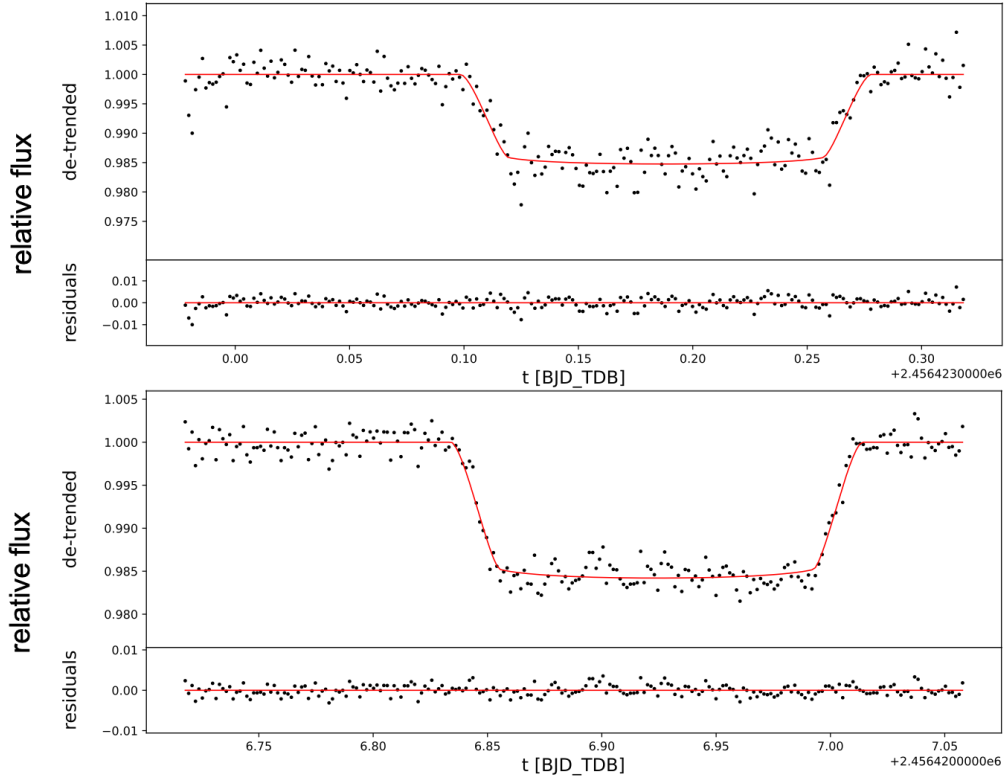


Figure 3.11: Detrended light curves, fits (red-line) and corresponding residuals for IRAC observations of WASP-17 b with channel 1 (top) and channel 2 (bottom). Figure adapted from [Saba et al. \(2022\)](#).

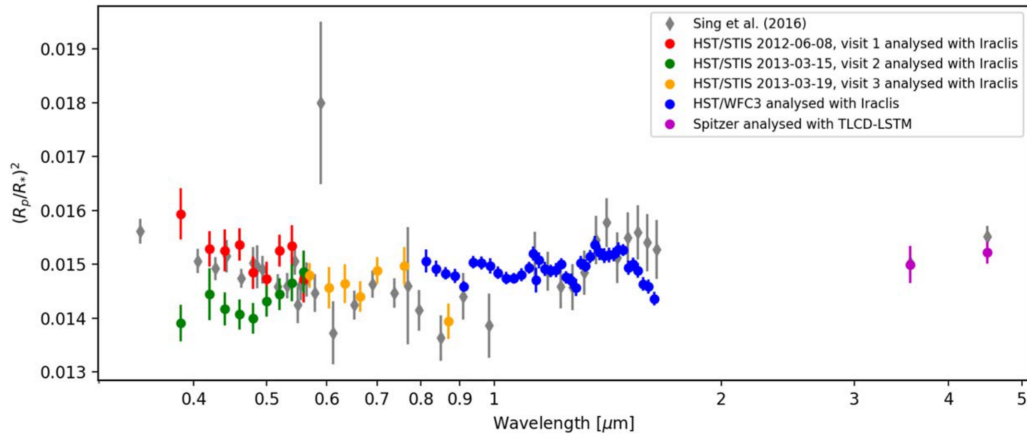


Figure 3.12: Transmission spectrum of WASP-17 b, constituted by HST/STIS, HST/WFC3, and Spitzer/IRAC channel 1 and 2 data analysed in this study compared to the data from [Sing et al. \(2016\)](#). Figure from [Saba et al. \(2022\)](#).

3.5 Summary and Discussion

We presented a deep learning model suitable for interpolating noisy light curves with covariates, and showed how it can be used to predict the variability of stellar light curves for subsequent transit fit. This approach has the advantage of making only little assumption on the types of noise, systematics or transit shape: it merely requires to know approximately when the transit takes place and assumes that centroid time series are uncorrelated with the transit.

The presented method is similar to the Gaussian Process (GP) approach ([Gibson et al., 2012](#); [Rasmussen and Williams, 2005](#)) in that they both construct highly non-linear models, avoid explicit physical modelling of the systematics and provide probabilistic predictions. However, they differ in various aspects:

1) The neural network light curve interpolation approach we propose does not need any transit model whereas it is included in the kernel of the GP. This makes the TLCD-LSTM approach more generally applicable as it does not depend on a pre-chosen mean or kernel functions.

2) The GP approach requires fewer parameters to train and provide fully Bayesian predictions compared to our LSTM-based approach. The smaller number of free parameters may make GPs the preferred choice for short time series. However, GPs computation scales more poorly with the number of data points, preventing them to be applicable to datasets of more than ≈ 1000 time steps without binning of the time series¹⁰. The proposed interpolating LSTM can on the other hand be applied to longer or multiple light curves as commonly found in Kepler and TESS time series allowing for even very long period variability to be captured in the predictive LSTM model. This is because the computational complexity in the case of GPs mainly depends on the number of data points, while in the case of the artificial neural networks in the architecture chosen (i.e. the number of layers, number of nodes per layer & type of layers in our case).

While the current implementation still relies on a few preprocessing steps such

¹⁰Although more scalable GPs implementations have recently been proposed, hence allowing for processing of increasingly longer time series (e.g. [Foreman-Mackey et al., 2017](#); [Farr et al., 2018](#), for fast GPs implementation in astronomical time series)

as computing the raw light curve or centroids fits, it constitutes a first step towards the ultimate goal of developing an end-to-end detrending pipeline where the input would be the raw pixel light curves or focal plane images. Furthermore, while we trained our network on data from six real light curves only, taking advantage of numerous light curves, real or simulated, would allow developing a more general detrending approach for each instrument. LSTMs allow for efficient *transfer learning* between data sets and instruments (e.g. Kepler to TESS). Transfer learning generally consists of training a model on a first dataset or task (also called pre-training), and *fine-tuning*—i.e. continue the training of—some or all of the model’s weights on another dataset or task. Transfer learning is often useful in low data regimes with multiple similar datasets, and may therefore be particularly relevant for modelling multiple light curve datasets with similar instrumental systematics or stellar noise, from the same or various instruments.

As we have firmly entered the era of ‘big data’ in planet detection (e.g. Kepler, TESS and ground based surveys) and with upcoming characterisation missions and instruments (e.g. JWST, Ariel, CHEOPS, the ELT, TMT and GMT), the opportunities for data detrending and modelling with scalable deep learning methods, capable of processing large numbers of high dimensional data will become increasingly prevalent in the future.

Software

The data and code used in this paper are available on GitHub under a Creative Commons Attribution 4.0 International Licence (<https://github.com/ucl-exoplanets/deepARTransit>, archived on Zenodo) and a MIT Licence (<https://github.com/ucl-exoplanets/pylightcurve>).

Chapter 4

Self-supervised Representations of Light Curves with Transformers

"There is no education but
self-education."

Charlotte Mason,
Towards a Psychology of Education

This Chapter is based on [Morvan et al. \(2022b\)](#): "Don't Pay Attention to the Noise: Learning Self-Supervised Representations of Light Curves with a Denoising Time Series Transformer", Machine Learning for Astrophysics Workshop, 39th International Conference on Machine Learning (ICML).

Astrophysical light curves are particularly challenging data objects due to the intensity and variety of noise contaminating them. Yet, despite the astronomical volumes of light curves available, the majority of algorithms used to process them are still operating on a per-sample basis, preventing information to be shared when processing multiple light curves. To remedy this, we propose a conceptually simple model—called Denoising Time Series Transformer (DTST)—and show that it excels at removing the noise in datasets of time series when trained with a masked objective, even when no clean targets are available. The Transformer (Vaswani et al., 2017) is a recent sequential model which relies mainly on a mechanism called *attention* broadly mimicking human attention, and which has reached state-of-the-art performance in a wide variety of tasks on language, vision and time series. In this work we adapt and use a Transformer model to reconstruct light curves with randomly masked chunks in order to learn the self and cross-correlations in and between light curves, and eventually improve imputation and detrending. The use of attention enables rich and illustrative queries into the light curves representations learned by the model. We present experiments on real stellar light curves from the Transiting Exoplanet Space Satellite (TESS), showing advantages of our approach compared to traditional denoising techniques. Automatic and efficient methods such as the DTST might play a crucial role in processing light curves in the future, for their potential added value in filling gaps, upsampling, identifying and correcting for anomalies, precise detrending for fitting or detecting challenging transit events such as single events or for exoplanets orbiting very active stars.

4.1 Introduction

Although the use of deep learning has started to emerge to successfully address some problems related to light curves (see Chapter 1) these often address only the later stages of data analysis and are limited to building supervised learning models. These models are indeed generally trained on scarcely labelled or simulated data and thus suffer from biases or small training sizes when applied to new or full datasets. On the other hand, there already exist large datasets consisting of thousands to billions of

light curves (see Chapter 2) with many more being generated by existing and future telescopes. We believe that tailored deep learning models will be able to leverage these large datasets to improve the efficacy and efficiency of light curve processing in a self-supervised, semi-supervised or unsupervised manner.

The self-attention mechanism (Parikh et al., 2016) and the Transformer architecture (Vaswani et al., 2017) have initiated a revolution in the field of natural language processing¹ (e.g. Devlin et al., 2019; Brown et al., 2020) and later computer vision (Khan et al., 2021). Transformers exhibit good generalisation, and offer easier training and better scalability compared to Long Short-Term Memory Networks (Hochreiter and Schmidhuber, 1997). Work is under way to adapt the Transformer architecture for time series tasks such as forecasting (e.g. Li et al., 2020; Zhou et al., 2021; Woo et al., 2022) or classification of astronomical time series (e.g. Allam Jr. and McEwen, 2021). In their study Zerveas et al. (2021) successfully pre-trained a Time Series Transformer via a masked objective before fine-tuning it for classification and regression. Even though the use of masked objectives is common in the aforementioned works, here our main objective is to denoise the time series. The masked objective allows us to solve the problem by means of a proxy imputation task without requiring any fine-tuning.

Our main contributions consist in: (i) introducing a simple self-supervised framework to perform time series denoising without access to clean targets; (ii) demonstrating how a Transformer encoder with minimal modification can perform light curve denoising effectively, leveraging the number and diversity of available inputs, (iii) producing flexible² and interpretable predictions by visualising attention scores associated with imputation and denoising of sequences.

We present experiments on real light curves from the Transiting Exoplanet Survey Satellite (TESS, Ricker et al., 2015). This is the first time a deep learning model is proposed to try to address both imputation and denoising on a dataset of light curves.

¹Natural language processing is often abbreviated "NLP".

²The flexibility of the model lies in its capability to handle inputs with missing values, variable sizes and generating processes characterised by different variances.

4.2 Background: the Transformer Model

If early uses of attention in the context of deep learning (e.g. [Bahdanau et al., 2016](#); [Kim et al., 2017](#)) were using a hybrid form of convolutional or recurrent modules along with attention modules, the main advances of the Transformer model is that it solely relied on attention and makes no use of convolution or recurrent operations. Attention overcomes two limitations of recurrent cells: it allows for long-term connections across the input sequence and scale easily to large batches since their computation essentially only involves matrix multiplications. It thus provided a conceptually simple sequential model which took over as the go-to baseline for a number of tasks in various fields, extending far beyond its original NLP applications. Here we are interested in the Transformer for its ability to model complex relationships in and between sequences (such as light curves), its ability to generalise to multiple datasets and tasks, as well as the interpretability provided by the attention mechanism which enables to interpret a model by visualising the weights assigned to each input³.

4.2.1 Attention mechanisms

The attention mechanism in deep learning takes its roots from the cognitive function of attention, recognised indispensable for human beings to selectively focus on some parts of their sensorial and memory inputs depending on the context. Intuitively, an attention module dynamically attends preferentially to certain parts of an input sequence by computing the relations between all the elements of its *key*. Its computation can generally be split in two steps: computing the attention distribution over the input, and computing the context vector according to the attention distribution.

There have historically been two main attention mechanisms, additive attention ([Bahdanau et al., 2016](#)), and dot-product attention ([Luong et al., 2015](#)). The multi-head scaled dot-product attention from [Vaswani et al. \(2017\)](#) introduces two changes from the classical dot-product attention: (i) a scaling factor $\frac{1}{d_k}$ and a split of the input along the feature dimension into several heads, as shown on Figure 4.1.

The attention score is computed from a set of queries Q , keys K and values V

³See an example of attention visualisation in Section 4.4.4.

as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

where the component i of the *softmax* of a vector x is defined as:

$$\text{softmax}(x)_i = \frac{e^{-x_i}}{\sum_k e^{x_k}}.$$

The Transformer model uses self-attention, i.e. that the queries, keys and values all relate to the same sequence by way of linear projections.

Dot-product attention is allegedly faster than additive attention, but its relative performance degrades for large d_k because of gradients becoming insignificant, hence the use of a scaling factor to mitigate this effect.

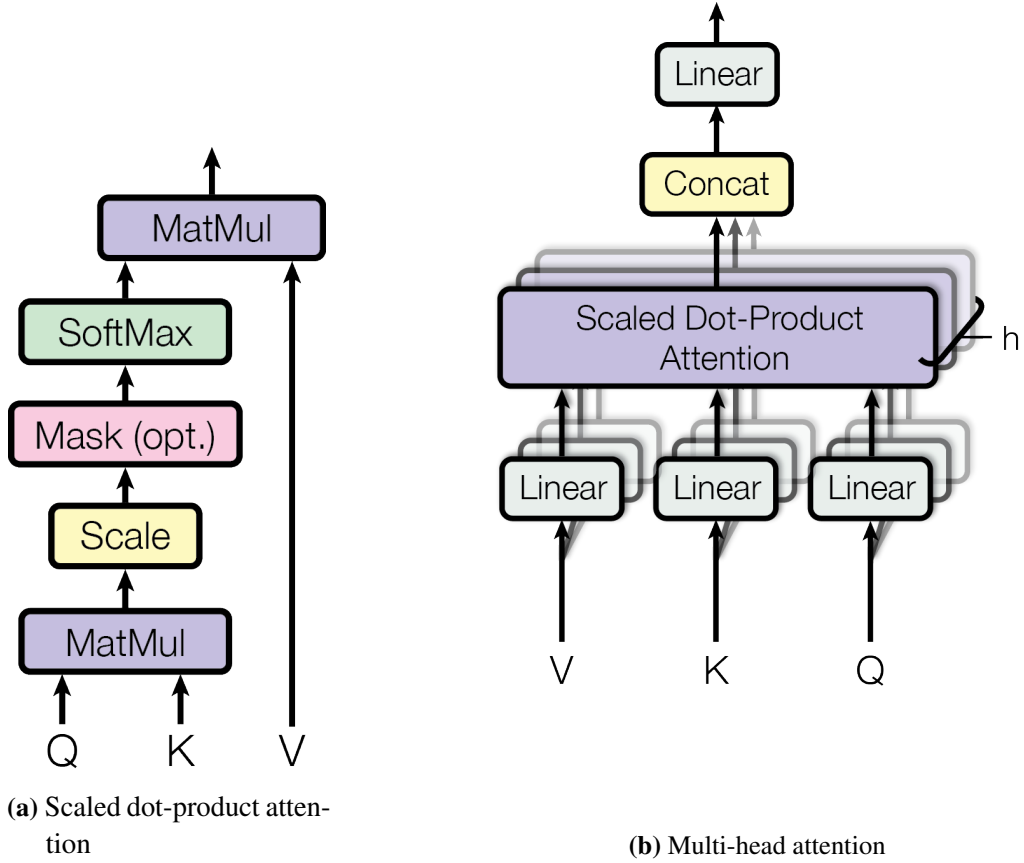


Figure 4.1: Multi-head scaled dot-product attention mechanism from Vaswani et al. (2017).

The vanilla attention mechanisms (both additive and dot-product) have a complexity of $O(T^2)$ both in time and memory. This can prevent the use of vanilla

attention for long sequences ($\gtrsim 1000$ cadences or tokens). Recently several alternative attention mechanisms have been proposed to mitigate this issue, often relying on some sparsity of the computed vectors (see e.g. Informer (Zhou et al., 2021) in $O(T \log T)$, Linformer (Wang et al., 2020) and FEDformer (Zhou et al., 2022) in $O(T)$).

4.2.2 The Transformer model

Figure 4.2 shows an overview of the architecture of the original Transformer model from Vaswani et al. (2017).

Input embedding

Input and output embeddings are typically learned through a distributed linear layer projecting the input into the transformer dimension d . Positional encodings are taken as shifted sines of various frequencies, and are aimed at providing sequential information about the input:

$$PE_{(pos, 2i)} = \sin\left(pos/10000^{2i/d_{\text{model}}}\right)$$

$$PE_{(pos, 2i+1)} = \cos\left(pos/10000^{2i/d_{\text{model}}}\right).$$

There are other types of positional encodings, fixed or learned (Gehring et al., 2017), and also some proposed types more specific to time series with explicit time information (e.g. Kazemi et al., 2019).

Transformer module

Each input is processed through one or several layers of the Transformer encoder composed with the following steps:

- a multi-head scaled dot-product attention module (orange blocks on Figure 4.2);
- a residual connection from the input added to the attention output: $x \rightarrow x + \text{Attention}(x)$ (yellow blocks on Figure 4.2);
- a layer-normalisation operation (Ba et al., 2016, yellow blocks on Figure 4.2);

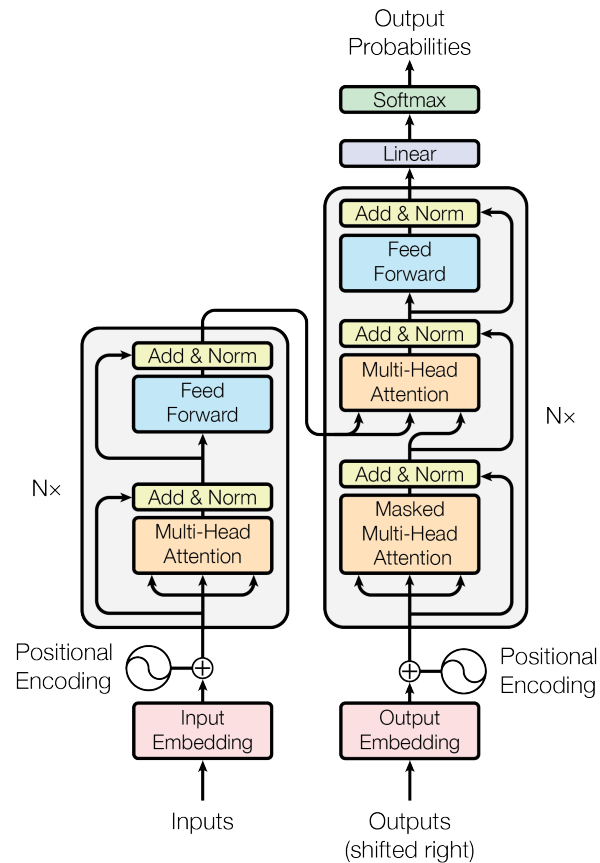


Figure 4.2: Overview of the Transformer model architecture from Vaswani et al. (2017). The left part shows one encoder layer and the right part shows one decoder layer. In practice, a number of such layers are stacked onto one another in the encoder and decoder.

- a fully connected 2-layer module distributed (i.e. applied to each input identically) and with a ReLU activation function (Agarap, 2019, (blue blocks on Figure 4.2));
- another residual connection connected to the fully connected layer output;
- another layer-normalisation operation;

The Transformer decoder module contains the same modules as the encoder, as well as an additional multi-head self-attention & residual connection & layer-normalisation block to process the target embedding. The second attention block then receives as input the concatenation of the results of the target attention block and the output of the encoder.

While the full Transformer model consists of an encoder and a decoder (see Figure 4.2), there are cases when the encoder or the decoder only can be sufficient to solve a problem. The model presented in the Section 4.3 only uses a transformer encoder with a linear decoder instead.

4.3 Methodology

4.3.1 Problem formulation

Given a univariate time series $x = \{x_1, \dots, x_t, \dots, x_T\} \in R^T$ we seek to predict its *trend*⁴ $y \in R^T$ which has been corrupted by a noise process ε such as $x_t = y_t + \varepsilon_t$ for each time step t . No assumption is made about the corruption process except its independence from the trend. In particular, ε can be heteroscedastic and non-Gaussian.

Let us consider a generic model f solely fed with corrupted time series, i.e. trained without clean targets in a *Noise2Self* setting (Batson and Royer, 2019). After masking a fraction of each input x with randomly generated masks m , f produces predictions $\hat{y} = f(x)$ of the same length as the input but is trained with a regression loss computed solely on the masked values: $\mathcal{L}(f(x), x, m)$. This masked objective guarantees the independence of the predictions with respect to the local values and their associated noise. If missing values are present in the dataset, they are treated in the same way as randomly masked values, making the method robust to missing values. The only difference is that predictions for truly missing values are not included in the calculation of the training loss.

4.3.2 Denoising Time Series Transformer

An overview of the DTST is shown on Figure 4.3. For each input time series an input mask is generated combining missing values and artificially masked values (at training only). Masked and standardised inputs are linearly projected into input embeddings $z \in R^{L \times D}$ of the model's dimension D . Input embeddings corresponding to masked positions are replaced by a learnable vector of dimension D , inspired by the mask token used in Devlin et al. (2019). This is a robust way of informing

⁴The ‘trend’ here can contain low frequency variability, aperiodic or periodic patterns.

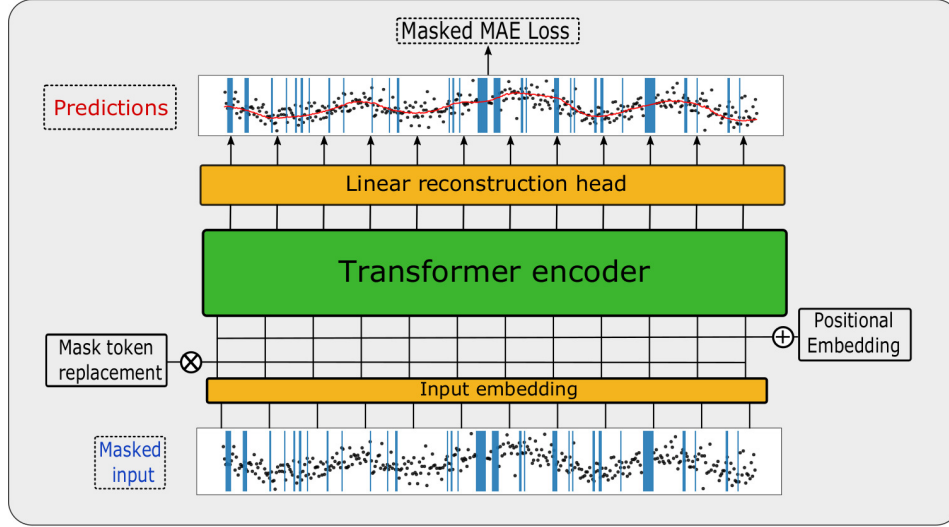


Figure 4.3: Schematic overview of the model learning trend representations of inputs with a masked objective. Masks represented in shaded blue areas include both missing and randomly masked time steps during training. At test time, only truly missing values are masked. Yellow modules represent the time-distributed linear embedding and the prediction head respectively.

the model of the masked input positions. Additionally, we have found a learnable vector to perform better than replacing masked values by zero as is often done for time series imputation models (e.g. [Cao et al., 2018](#); [Zerveas et al., 2021](#); [Yi et al., 2020](#)). However, using this scheme on its own affects the quality of the predictions outside the masks and for this reason we replace 10% of masked values in input with uniformly sampled values between -2 and 2 and do not replace their projected input embedding by the mask embedding. This setting was the most effective we have tried among several ones listed in Section 4.3.3.

Positional encodings⁵ are then added to the input embeddings to provide positional information. Since they produced better results than trainable positional encodings in our experiments we used the same fixed positional encodings as in [Vaswani et al. \(2017\)](#). We use a light version of the original Transformer encoder with the hyperparameters fixed to the values shown in Table 4.1. Each encoder's output is finally projected back into the input dimension using a distributed $D \times 1$ linear layer.

⁵In the context of Transformer models, "positions" refer to the order of the input sequence, which is here taken to be the cadence number.

PARAMETER	VALUE
Learning rate	0.001
Batch size	64
Dim. model	64
Dim. feedforward	128
Num layers	3
Num. heads	8
Masking ratio	30%
Average masking length	10

Table 4.1: Fixed hyperparameters for all presented experiments after hyperparameter optimisation. For training we used Adam (Kingma and Ba, 2015) optimiser with learning rate 0.001 and $\beta = (0.9, 0.999)$.

4.3.3 Masking strategy

Masking Patterns

We explored various generating mechanisms for the distributions of masked values in each input. Given a fixed ratio of values to mask, we tested using a Bernoulli distribution for time-independent masking and a geometric distribution over masked blocks lengths. The geometric distribution was used to impose longer masks and thus a more challenging imputation task to the model. Mean block lengths of 5, 10 and 20 were tested and final results were presented for a window of 10 as it offered the best compromise between the length of signals to impute and denoising performance. Intuitively, the length of masked blocks will control the degree of temporal locality of the noise processes to remove, and using wider masked regions will indeed force the model to make use of longer-term dependencies to make accurate predictions.

Masking Ratios

We set the masking ratio to 30% after experimenting with 10%, 20%, 30%, 40% and 50%. Heuristically, increasing the masking ratio speeds up training but also affects performance as fewer inputs are available for prediction. When values are missing in the inputs, the masking ratio is considered with respect to the number of non-missing time steps. This maintains the ratio of data used for training constant while avoiding degenerate cases where the random mask would be empty or would cover the entirety of the non-missing input.

Replacement Strategy

We considered various replacement strategies for masked input values: (i) replacing by zero, (ii) by a uniformly random value centred on zero, (iii) by a special learnable vector in the model’s space, and (iv) keeping the original values. Case (iv) was quickly discarded as it led to overfitting the noise. Case (iii) is directly inspired from [Devlin et al. \(2019\)](#), where the authors replace masked positions in the input embedding by one single vector of dimension D (i.e. at every time step) which is learned automatically, with the aim of developing an internal representation of what should be a masked value. While option (iii) offered the best imputation performance, we observed that it performed poorly on its own for denoising the full inputs, and that this issue was mitigated by using case (ii) for a random fraction of input time steps, even as small as 10%. This can be understood as an extra corruption operation on the input, thus forcing the model to provide coherent predictions even outside the regions whose embeddings are more explicitly masked with a dedicated vector.

4.4 Experiments

4.4.1 Dataset

We present experiments on a dataset of light curves from the TESS satellite, acquired during the first visit of its first sector in 2018. TESS light curves are challenging because of their length (20,076 time steps for short cadence data spread over 30 days), their noise level, residual instrumental systematics and missing blocks.

We select 2 minutes cadence light curves at the Presearch Data Conditioning (PDC) stage, i.e. after removal of the main instrument systematics, cosmic rays and background noise with the standard TESS pipeline ([Jenkins et al., 2016](#)). After rejection of 50 light curves with negative flux, the dataset contains 15839 light curves. We selected 20% of all light curves uniformly at random for testing and the remaining 80% for training and validation (with 10% of the training light curves kept for validation).

4.4.2 Training and evaluation

Because of their length we randomly crop each light curve to select 400 consecutive time steps. A random mask is then generated before subtracting the mean and dividing by the standard deviation of the non-masked values for each input segment. This procedure can be seen as a data augmentation step, as the combination of cropping and masking operations will produce different inputs at each epoch.

For training the DTST we use the noise-scaled masked mean absolute error (NMMAE) defined as: $\text{NMMAE}(\hat{y}, x, m) = \frac{1}{M * n(x)} \sum_{t=1}^T m_t |x_t - \hat{y}_t|$, where m is a binary mask equal to 1 for masked time steps and 0 otherwise, \hat{y} is the model's output prediction, $M = \sum_{t=1}^T m_t$ is the total number of masked steps in x and $n(x)$ is an estimate of the local noise by computing the average moving standard deviation with a window of size 10 and a step of 5. Compared to the mean-squared error, the mean absolute error (MAE) is more robust to outliers while rescaling using $n(x)$ helps to account for different variabilities in the training data. Predictions for the full light curves are then obtained by stitching together the predictions for segments of 400 time steps. In practice, evaluation segments are designed so as to allow overlaps of 50 steps and remove the outer 25 steps for each prediction.

As evaluation metrics we use the MAE and the inter-quartile range (IQR) of the detrended light curve as a measure of the residual noise, both expressed as percentages of the stellar flux. For both measures, lower values are desirable.

We compare the DTST to the median filter and Tukey's biweight algorithms with implementations from [Hippke et al. \(2019\)](#) as baselines. These have shown the best performance in removing the noise prior to detecting exoplanets in Kepler and TESS data compared to other methods presented in that study. Both methods require to set the window length—in time units for Tukey's algorithm and in number of cadences for the median filter. For comparison we select two window lengths: a long window of 6 hours (~ 300 time steps) which is adapted for exoplanet transit detection and a short window of ~ 2 hours which provides comparable denoising scores to the DTST but overfits some of the high frequency variability.

Computational Efficiency

Even though training the DTST on thousands of time series can take up to several hours on a single V100 GPU, its inference cost remains very low with around $10 \mu s$ for a full TESS light curve unfolded in windows of length 400 passed as a batch of size ~ 60 . This is to be compared with $\sim 50 \mu s$ and $\sim 173 \mu s$ per TESS light curve for the efficient Wotan implementations of biweight and median filter respectively.

4.4.3 Results

After experimenting with various architectures and masking scenarios (see section 4.3.3) on the training and validation sets, we evaluated the DTST and the baselines on the test set. Results are presented in Table 4.2. On average, the DTST provides the smallest residual noise and auto-correlation out of the several baselines evaluated here. The difficulty for traditional techniques here lies in reconciling the diversity of the stellar processes composing the dataset, and it is therefore understandable that a single cadence-based or window-based filter with a fixed window size will either fail to denoise targets with high variability or overfit the noise on those with low variability.

METRIC	MEDIAN FILTER		TUKEY's BIWEIGHT		DTST
	65 steps	181 steps	2 hours	6 hours	400 steps
MAE	0.244%	0.286%	0.245%	0.286%	0.235%
IQR	0.393%	0.465%	0.398%	0.469%	0.385%

Table 4.2: Denoising performance on 3168 TESS light curves from Sector 1. Averaged errors are given in percentage of the stellar flux. Window sizes considered by the three algorithms to make predictions are shown on the second line.

We show examples of predictions in Figure 4.4 for different test samples showing a range of variability patterns. We corrupted the two inputs on the left (Figures 4.4a and 4.4b) with random masks similar to those used during training. The predicted time series shown in red on each upper sub-plot plot shows very good agreement with the expected trend for both masked and unmasked input time steps. In dashed green line is shown a the result of a median filter on each light curve with a window of 65 cadences (equivalent to 2 hours). While it provides good results for slowly varying stellar processes (Figure 4.4b), this setting fails to account for faster

processes (Figures 4.4a and 4.4c) or inputs with many missing values (Figure 4.4d).

On each third sub-plot we show the residual light curve in units of stellar flux. The associated autocorrelation function (ACF) of the model-fit residual is plotted on the last subplot of each figure. The ACF is a useful tool to analyse the significance of residual time correlations as in Figure 4.4c. Target 140045538 indeed shows bursts of flaring activity which are not predicted by the DTST and therefore leave a significant signature in the ACF. As short transients or planetary transits may lie at the border between noise, outliers and signals, further fine-tuning of the model may be needed for either predicting or ignoring them consistently.

4.4.4 1D attention maps

Attention from the model’s output to the input time series is computed using *Attention Rollout* (Abnar and Zuidema, 2020). This procedure consists in recursively multiplying matrices of attention weights through the transformer layers, thus accounting for mixing of attention in the network. Figure 4.5 shows more examples of predictions overlayed with their corresponding input with Attention Rollout used to highlight time steps with greater attention.

This enables us to visualise which parts of the inputs received more attention for producing the outputs, both during training and validation. Thus we are using the generated attention maps both for orienting the model’s development and interpreting its predictions.

Our first observation is that both input tails often receive high Attention Rollout scores. This is understandable as these lack context on either their left or right and therefore prove more challenging to predict. We also observed that large masked regions receive generally less attention than non masked regions. This is in fact a useful check during the model’s development to verify if the model manages to distinguish between the mask representation and the real values. Furthermore, values surrounding the identified gaps often show greater attention than average, probably as they are particularly relevant for the prediction of masked values. Finally it is often interesting to look at the attention patterns for time steps corresponding to unexpected flux values. Those are sometimes ignored such as the rightmost flaring

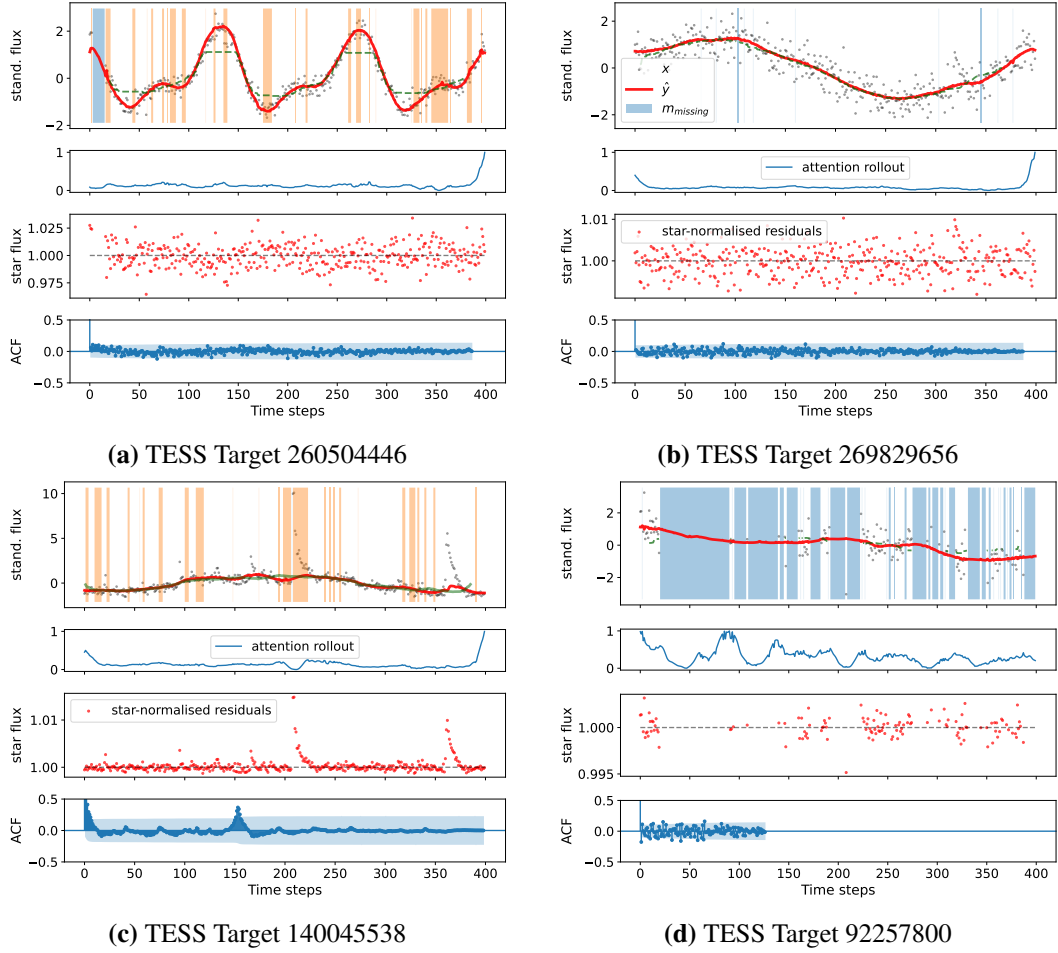


Figure 4.4: Prediction summary for four different test stars. On the left are two examples with random artificial masks (in orange) mimicking the training process. On the right are two untouched inputs, where the blue shaded masks indicate truly missing data in input. Each sub-figure contains from top to bottom: (i) inputs as black dots, the DTST’s predictions as red line, and median filter with window of 65 cadences as green dashed line, (ii) rolling attention time series scaled between 0 and 1, (iii) the star-normalised residual errors and (iv) the auto-correlation function with missing data ignored.

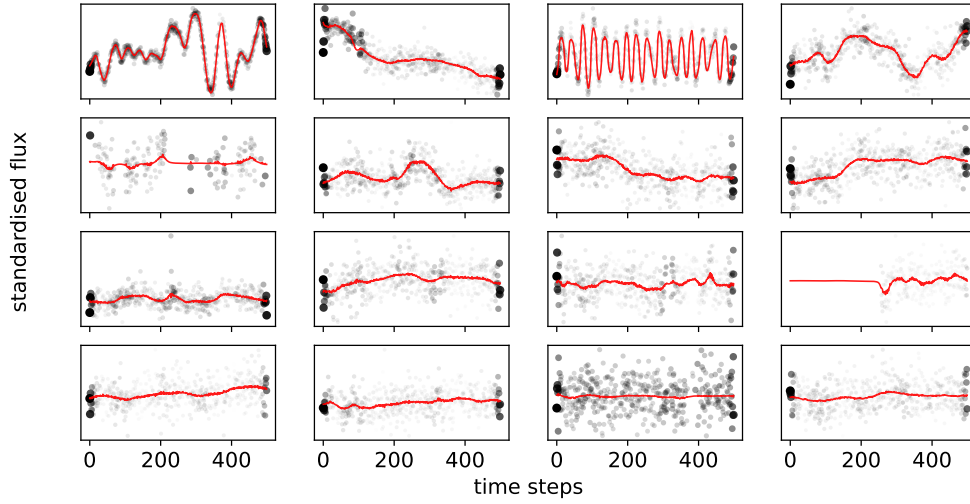


Figure 4.5: 1D attention maps computed with Attention Rollout and overlaid with predictions (red lines) for 16 random light curves from TESS dataset. The size and opacity of individual point inputs is directly proportional to the Attention Rollout values.

event on Figure 4.4c or conversely receive more attention than average when they can inform predictions.

4.5 Summary and Discussion

In this work we presented a conceptually simple framework to denoise time series via a proxy imputation task. We performed experiments and showed how such an approach based on a Transformer encoder architecture is effective at removing the noise in light curves from the TESS satellite. Compared to traditional techniques, this model can offer flexibility and increased performance when preprocessing large datasets of light curves. Further works will extend these experiments to other real and simulated datasets while assessing the generalisation power and possible gain from using a pre-trained model. Finally we would like to use this approach as a basis for downstream tasks such as event detection, imputation and upsampling.

Chapter 5

A Hybrid Physics & Deep Learning Framework for Transit Inference

"Artificial intelligence is extending what we can do with our abilities. In this way, it's letting us become more human."

Yann LeCun

This Chapter is based on [Morvan et al. \(2021\)](#): "PyLightcurve-Torch: A Transit Modelling Package for Deep Learning Applications in PyTorch". Publications of the Astronomical Society of the Pacific 133 (1021): 034505.

We present a new open-source python package¹, based on PyLightcurve (e.g. Tsiaras et al., 2016) and PyTorch (Paszke et al., 2019), tailored for efficient computation and automatic differentiation of exoplanetary transits. The classes and functions implemented are fully vectorised, natively GPU-compatible and differentiable with respect to the stellar and planetary parameters. This makes PyLightcurve-torch suitable for traditional forward computation of transits, but also extends the range of possible applications with inference and optimisation algorithms requiring access to the gradients of the physical model. Such hybrid physics/DL models allow for a higher level of robustness, flexibility and explainability, and foster the use of deep learning in exoplanets research.

5.1 Introduction

Exoplanets science discoveries have relied largely on our ability to extract precise information from stellar light curves. In the case of transiting exoplanets, this requires high precision photometric or spectroscopic measurements, and involves a transit model with one or several parameters to be determined. However, transit light curves often contain other sources of temporal variability caused by the instrument or the host star, which need to be accounted for in a preliminary or joint processing step. Consequently, forward modelling of transits needs to be considered as part of the data processing pipeline yielding the planetary parameters.

The complexity and growing amount of exoplanet light curves hint at the use of deep learning to help alleviate the issues encountered with traditional modelling techniques. Indeed, tremendous progress has been made recently in time series analysis owing to the recent development of deep learning, producing several successful applications and promising solutions for problems ranging from time series classification (Fawaz et al., 2019) and forecasting (Hewamalage et al., 2020) to denoising and anomaly detection (Chalapathy and Chawla, 2019). Amongst the various fields benefiting from such technical progress, astronomy has not been an

¹PyLightcurve-torch can be found on the following GitHub repository: <https://github.com/ucl-exoplanets/pylightcurve-torch>

exception. Several recent contributions to exoplanetary sciences (see 1.3.3) involve the use of deep learning methods.

The recent successes of deep learning can be traced back to the introduction of backpropagation as a technique which has allowed for the efficient optimisation of neural networks (Rumelhart et al., 1986). Today all major deep learning frameworks including TensorFlow (Abadi et al., 2016) and PyTorch implement a way to automatically compute gradients of scalar outputs of functions, with respect to their inputs and parameters. To use and include a function in a data flow graph created by one of the deep frameworks above, the function must first be implemented using the building blocks of the framework: the functions and the numerical objects (multidimensional arrays, often called *tensors*) specific to the framework’s language. So far, to the best of the authors’ knowledge, none of the existing transit modelling codes has been designed to allow automatic differentiation and thereby joint end-to-end training with artificial neural networks. This is precisely the purpose of PyLightcurve-torch, which provides a user-friendly transit modelling tool to facilitate the generation, inference and optimisation of transit models in a framework compatible with deep learning modules. We chose the language of PyTorch due to its user-friendliness, flexibility, efficiency and growing popularity. As the PyTorch syntax is very close to that of NumPy, it facilitates the easy conversion of NumPy codes to PyTorch, and reduces the learning curve for research communities who are used to scientific programming in NumPy.

PyLightcurve-torch is adapted from PyLightcurve², which is one of the most efficient open-source transit modelling packages available. PyLightcurve performs some numerical approximations rather than solving the fully analytical transit model to enable vectorisation of computations with NumPy (Harris et al., 2020) and gain in efficiency, thus providing a good template for designing differentiable and scalable code. Four different limb-darkening laws are natively available in PyLightcurve, as well as several utilities for database access and fitting that we are not considering here. For more details about PyLightcurve’s physics models, implementation and

²<https://github.com/ucl-exoplanets/pylightcurve>

performance, see [Tsiaras et al. \(2016\)](#).

The remainder of this chapter discusses the implementation (Section 5.2.1), performance (Section 5.2.2) and applicability (Section 5.3) of *PyLightcurve-torch*.

5.2 *PyLightcurve-torch*

5.2.1 Code design

The numerical programming code was adapted from *PyLightcurve* transit modelling library (e.g. [Tsiaras et al., 2016](#)). Indeed, the main functions `exoplanet_orbit`, `transit_duration`, `transit_flux_drop`, `transit` and `eclipse` have been translated to PyTorch while preserving their names, structure, and parameters. However, several major changes have been introduced along with the conversion to PyTorch. These have been summarised in the list below.

From NumPy to PyTorch

NumPy arrays and operations have been converted respectively to PyTorch tensors and their corresponding operations. This means that the input parameters of corresponding main functions must now be of type `torch.tensor` and of shape broadcastable to `(batch_size,)` where `batch_size` is the number of instances of each transit parameter. While PyTorch tensors share many similarities with NumPy arrays, they further allow for **GPU acceleration** and **automatic differentiation**.

Vectorisation

The main functions have been further vectorised to allow inputting 1D-arrays for each transit parameter in addition to scalars. Expressed in another way, this allows the user to provide batches of inputs to the main functions instead of individual sets of parameters. In the PyTorch version of *PyLightcurve*, this not only enables batch learning - i.e. optimisation based on groups of observations considered jointly - but also fully leverages the GPU acceleration advantages on multidimensional tensors.

Flexibility of input shapes

By allowing inputs of parameters of broadcastable shapes, the use of the main functions remains intuitively flexible, while saving memory and time in some specific cases. Indeed, when intermediate computations can be shared across a batch, such as

the planetary positions vector for scalar orbital parameters, only a vector of batch dimension 1 is computed and used for later computations of transit flux drops, even if the latter are multidimensional.

A `TransitModule` class

has been implemented to facilitate the use of `transit` and `eclipse` functions, their optimisation and embedding in deep learning pipelines. Indeed, `TransitModule` first manages transit parameters and intermediate computations in an object-oriented fashion (see Listing 5.1 for a basic example). For convenience, the transit parameters passed as attributes of a module undergo checks, type, shape and device casting to make sure correct inputs are passed to the PyTorch functions performing the actual transit computations. Secondly, `TransitModule` inherits the `torch.nn.Module` along with its methods and parameters internal management. Furthermore, as the main parent class of all neural networks implemented in PyTorch, instances of `torch.nn.Module` can easily be combined, facilitating the embedding or combination of our transit models with neural networks.

Listing 5.1: Basic use of wrapper class `TransitModule` computing transit and/or eclipse flux while inheriting `torch.Module` class.

```
from pylightcurve_torch import TransitModule

# Model definition
tm = TransitModule(time, **params) # time array-like object and dict of params
tm.activate_gradient('rp')         # gradient activation for 'rp'

# Forward transit computation
flux = tm()                         # with module's defined parameters
flux = tm(e=0., t0=3.4)             # with substituted external parameters

# Loss and backward pass
err = loss(flux, **data)             # loss function evaluation
err.backward()                      # backward propagation of gradients
tm.rp.grad                          # gradient access for parameter 'rp'
```

5.2.2 Performance

Several tests were conducted to assess the accuracy and efficiency of `PyLightcurve-torch`, taking as reference `PyLightcurve` since the former was based on the latter.

Moreover, PyLightcurve passes an wide range of unit-tests, has been cross-checked against other available transit codes and has been in use in various projects for several years. First, a sanity check was carried out to ensure both codes provide the same outputs when provided with the same inputs, up to a precision level below 0.1 ppm, on average, for default precision settings.

Two experiments were then performed with the aim of comparing the computational efficiency of both codes on CPU and GPU machines. In the first case the time array/tensor length was varied while keeping fixed the number of transit parameters, and in the second case the time array/tensor was fixed to 1000 while the number of parameters varied. The results are presented in Figure 5.1, suggesting very similar performances for the NumPy and the torch-cpu versions of PyLightcurve transit function. However, the GPU runs show a significant reduction in computation time, which is no longer increasing linearly with the input sizes but rather plateauing under $\sim 10ms$ for time tensor sizes smaller than 10^6 and batch size smaller than ~ 256 . Although these specific thresholds depend on the architecture used - which in our case consisted of 10 CPU cores, 70 GB memory and 1 Tesla-V100 GPU core - we expect GPUs to bring significant improvements in most configurations and use cases.

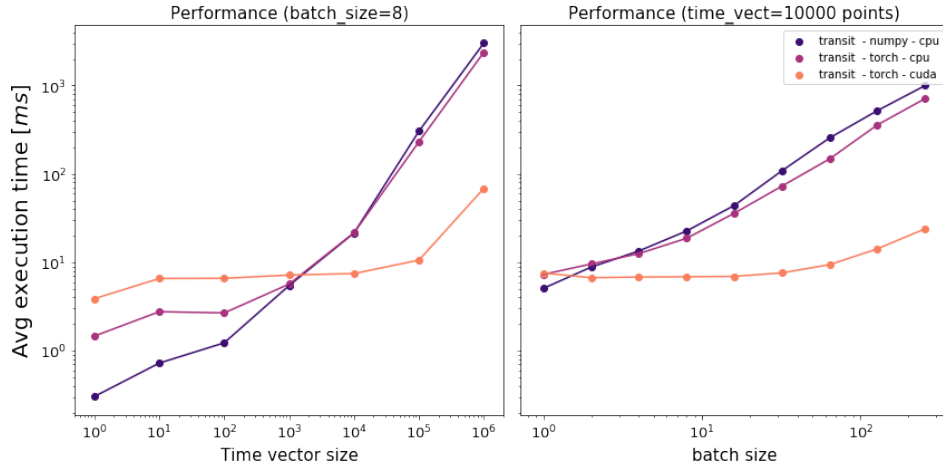


Figure 5.1: Execution times of PyLightcurve-torch transit functions in NumPy and PyTorch, averaged over 50 executions. PyTorch functions have been run onto a CPU cluster (purple curve) and a GPU core (orange curve). **Left:** batch size fixed to 8 while increasing the length of the input time series. **Right:** the input time series size was fixed to 10000 points while increasing the batch size.

The ability to handle large datasets and maintain efficiency while computing

a transit function or its gradient opens up new possibilities for processing exoplanet light curve datasets. Furthermore, automatic differentiability now allows the transit function or `TransitModule` to be included dynamically in deep learning pipelines and involved in their end-to-end optimisation. The next section discusses the potential for new applications brought by `PyLightcurve-torch`.

5.3 Applications

This section, rather than being an exhaustive list of applications, aims to provide a quick overview of the types of novel applications afforded by the use of `PyLightcurve-torch` for designing both generative and discriminative models.

5.3.1 Generation

`PyLightcurve-torch` can be used in a static generative mode, to efficiently simulate primary and secondary transit light curves. In this case the gradients may not need to be activated (static mode: `requires_grad=False`), depending on the problem considered. Indeed, one can, for instance, create artificial datasets statically and use them to train models for various problems such as transit depth regression or event classification. While the simulated transits, created as PyTorch tensors, enable GPU acceleration, the conversion to NumPy arrays is still possible and very cheap computationally, simply by means of calling the `.numpy()` method available for `torch.Tensor` objects.

5.3.2 Gradient-based optimisation

Having efficient access to the transit model's output gradient with respect to the transit parameters enables gradient-based optimisation without having to use approximate methods to compute gradients. Gradients can indeed be efficiently computed using *reverse-mode automatic differentiation* (i.e. backpropagation) in the case of a single or few outputs, which is the case in the context of loss functions to be optimised in machine learning. As long as a the transit model output is used by differentiable functions to compute a scalar output, `PyLightcurve-torch` will therefore allow efficient backward gradient-optimisation through different parts of the pipeline (see next Subsection for a discussion on combining neural networks with a differen-

table transit model).³ PyTorch conveniently implements a number of off-the-shelf optimisers which make use of the first or higher order derivatives of parameters. Moreover, several popular MCMC sampling algorithms such as Hamiltonian Monte Carlo and NUTS (Hoffman and Gelman, 2011) also require the availability of gradients, hence opening up the possibility for this class of MCMC algorithms to be used to derive posterior distributions for our transit models. Implementations of HMC and NUTS are available in the probabilistic programming language Pyro (Bingham et al., 2019), which is also based on Python and PyTorch. Besides exact Bayesian inference with MCMC, it is worth noting that Pyro also provides a convenient framework for probabilistic, flexible and deep inference of parameters, and has been designed to be fully compatible with PyTorch tensors, functions and modules.

5.3.3 Combination with neural networks

The flexibility afforded by the autograd package and torch modules makes it particularly easy to connect the input and the output of our transit model with other differentiable functions and modules. All parameters can then be optimised in an end-to-end mode through the computational graph automatically built when defining and operating on the tensors. In practice, this means that any other differentiable module or function can:

- provide the transit parameters as **input** of the transit model. A schematic example of this setup is shown in Figure 5.2a.
- be used in **parallel** with the transit model to provide other scalars, vectors or time series to be combined subsequently with the transit output. This setup is, for example, suitable for time series decomposition of transit light curves by means of generative models. A schematic example of this setup is shown in Figure 5.2b.
- be applied to the **output** time series of the transit model, transforming it to another time series, a vector or a scalar. Note that to perform gradient-based

³In the opposite situation in which a single input is used to compute multiple final outputs, forward-mode automatic differentiation will be the most efficient way to compute gradients.

optimisation, a loss function outputting a scalar value will need to be evaluated at the end of any functional flow. A schematic example of this setup is shown on Figure 5.2c.

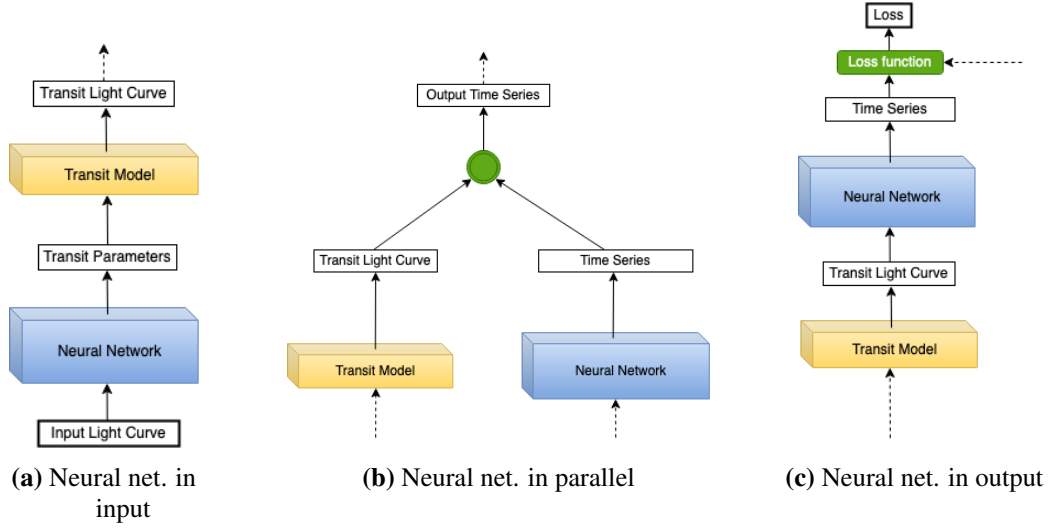


Figure 5.2: Schematic building blocks showing examples of combination between a neural network and differentiable transit model inputs (a) and/or outputs (b and c). Dotted arrows indicate possible additional inputs and outputs. The circular green node represents here a non-parametric operation such as item-wise multiplication or addition. The loss function (in green) computes a scalar value from a time series in this case, optionally using some additional scalar or tensor values.

5.3.4 Regression experiment

Let us consider a simple example of regression problem where we use the transit module as an extra term in the objective loss function of a neural network. As such, this experiment falls into the case (a) discussed above, where a neural network output is used as input to a transit model. In the present experiment, the main task is to build a predictor for $y = R_P/R_s$, assuming the other transit parameters, θ , to be known. Note that θ parameters are only required during training, and the regression models presented below do not require to know θ at inference time. It is worth highlighting that this is therefore a simulation setting in which we still need to know some parameters during training, but even in a case in which we would be training models on real observations, it is not unrealistic to assume that these planetary and stellar parameters are known from other methods or previous studies and focus on

the sole determination of R_p/R_* .

The dataset is composed of 2000 light curves generated using PyLightcurve-torch with added Gaussian noise. Additional details about the experimental setup are provided in Table 5.1. All the light curves are univariate time series with $T = 1000$ uniform time steps and parameters fixed except for the target R_p/R_* and the inclination i which are sampled from uniform distributions⁴. Note that in this experiment i is not included in the output targets, and that its variability is solely aimed at increasing the complexity of R_p/R_* prediction. We denote by Y the set of ground-truth targets and by \mathcal{M} the transit model which generated the light curves.

Parameters	Distribution	Values
Ratio of radii R_p/R_*	uniform	0.01 - 0.1
Inclination i (deg)	uniform	87.5 - 92.5
Limb-darkening method	constant	linear
Limb darkening coefficient	constant	0.2
Normalised semi-major axis a/R_*	constant	15
Argument of periastron (deg)	constant	50
Eccentricity e	constant	0
Gaussian noise standard deviation	constant	10^{-5}
Light curve lengths T (points)	constant	1000
Light curves time span (days)	constant	10
Mid-transit time from start (days)	constant	5
Orbital period P (days)	constant	100
Training size	-	1200
Validation size	-	400
Test size	-	400
Optimiser	-	Adam
Initial Learning Rate	steps of 10^{-6}	10^{-6} - 10^{-5}
Max epochs	constant	20,000

Table 5.1: Physical and training parameters used for the regression experiment. Training, validation and test sizes indicate the number of light curves randomly generated to build the training, validation and test sets respectively. The distributions the parameters are sampled from are identical for the three datasets. Whenever the distribution is not constant, the values indicate the extremes of the corresponding sampling ranges.

Two identical neural network models m_1 and m_2 are trained on this regression

⁴Note that it is in principle the distribution of $\cos i$ which should be uniform assuming a uniform orientation of planetary orbits. In practice, this does not play an important role here as $\cos i$ is nearly linear ($\cos \pi/2 - x \sim x$, for $x \sim 0$) on the considered range of inclinations 87.5 - 92.5 deg.

problem using *different loss functions*. Indeed, m_1 is only trained with a mean-squared error loss between the target parameters y and their associated predicted value \hat{y} :

$$\mathcal{L}_1 = \frac{1}{|Y|} \sum_{y \in Y} (y - \hat{y})^2,$$

whereas m_2 is also trained to reconstruct the transit flux by embedding the transit model as an additional term in the loss function:

$$\mathcal{L}_2 = \lambda \mathcal{L}_{\text{regression}} + (1 - \lambda) \mathcal{L}_{\text{reconstruction}}.$$

where :

- $\mathcal{L}_{\text{regression}}$ is identical to \mathcal{L}_1 but with predicted values \hat{y} from m_2 instead.
- The second term is a **transit reconstruction loss**, measured as the mean-squared error between the input light curve and the reconstructed transit light curve:

$$\mathcal{L}_{\text{reconstruction}} = \frac{1}{|Y|T} \sum_{y \in Y} \sum_{t \in 1..T} (\mathcal{M}(\hat{y}, \theta)_t - x_t)^2,$$

- λ is a scalar hyperparameter balancing the relative importance between the two loss terms.

The neural network architecture chosen consists of 4 convolutional blocks followed by 2 linear layers. Both models are trained using the Adam optimiser for 20000 steps. Even though the networks and λ parameters would be suited for hyperparameter optimisation, finding the optimal solution to this problem would go beyond the scope of this study, and here we simply present the results for a comparison of cases $\lambda = 0$ ($\mathcal{L}_2 = \mathcal{L}_{\text{regression}} = \mathcal{L}_1$) and $\lambda = 0.5$ (equal contributions from $\mathcal{L}_{\text{regression}}$ and $\mathcal{L}_{\text{reconstruction}}$).

The performance of both models is measured with the mean-squared-error of the predicted transit depths. The evolution of this metric evaluated on a validation subset of 400 light curves during learning is presented in Figure 5.3, which shows a significantly lower validation error for the model m_2 trained with the transit

reconstruction term in the loss function. Indeed, this model reaches m_1 's final performance far before m_1 (in only about 2000 epochs) and continues to further improve its performance until the maximum number of epochs (20000) is reached. This indicates a clear advantage given to the model informed from the transit function and jointly trained with the proxy task of reconstructing the transit shape.

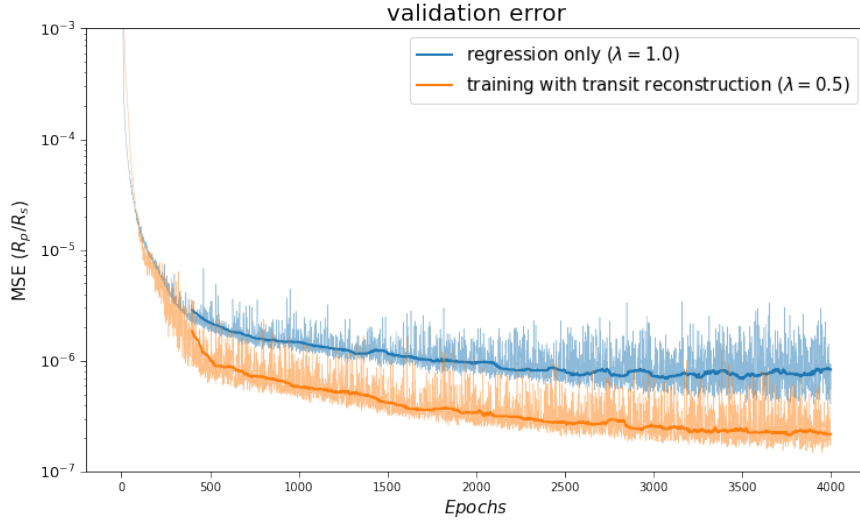


Figure 5.3: Validation mean-squared errors for R_P/R_* prediction with (orange curve) and without (blue curve) including a transit reconstruction term in the training loss function. The thicker orange and blue curves merely show the moving median of the respective validation errors with a window size of 100 steps.

For information, we plot some light curves randomly selected from the training and test sets on Figure 5.4, as well as reconstructed transit light curves from models' predictions and least-square fits of the R_P/R_S parameter (in red). Both m_1 and m_2 seem to provide good predictions on training samples (top plots, where the reconstructions from the models' predictions agree well with the least-square fit). However, m_2 —orange line, which was trained using an additional transit reconstruction term in the loss—provides better R_P/R_* predictions than m_1 (blue line) on the test set (bottom plots).

Finally, while the least-square fits are given for comparison, this optimisation process is different by nature from the trained neural networks, because it requires a transit model with known stellar/planetary parameters in order to fit R_P/R_S whereas

neural networks do not at inference time. In terms of computational efficiency, neural networks will take additional time to be trained (typically a few minutes on a V100 GPU for the dataset presented above), but will then be faster than a least-square fit when it comes to making new predictions.

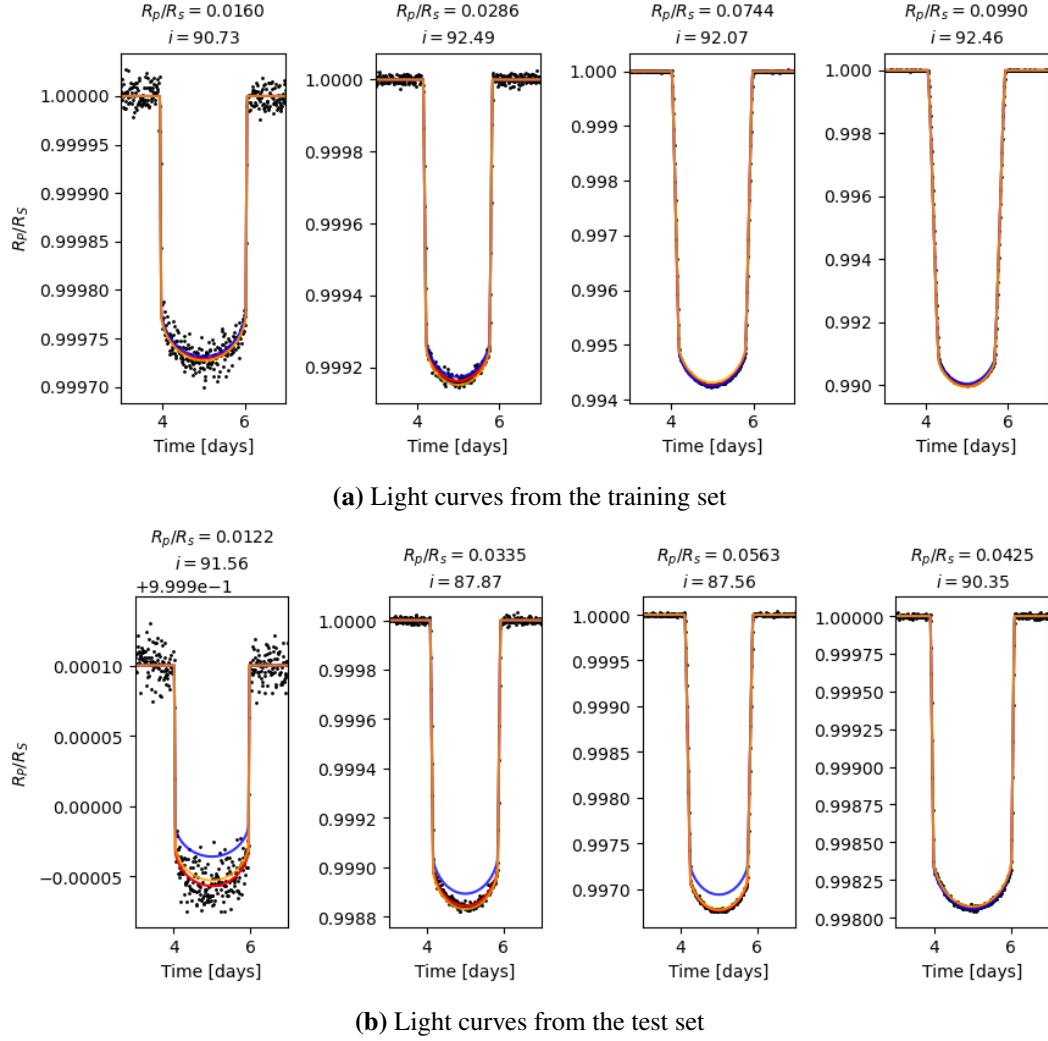


Figure 5.4: Examples of light curves simulated using PyLightcurve-torch for a regression experiment. Solid lines show the predictions from the two networks m_1 (in blue) and m_2 (in orange, physics-aided during training) and least-square fit (in red) for comparison. (a) shows light curves randomly selected from the training set and (b) shows light curves randomly selected from the test set, i.e. which have not been used for training nor validation. Light curves were cropped to show only four days centred on the transit. Ground truth targets R_p/R_* and varying inclinations i (in degrees) are shown above each plot.

5.4 Summary and Discussion

In this chapter we outlined several possible uses of PyLightcurve-torch for generating and optimising transits as well as embedding them in machine learning flows optimised end-to-end by gradient descent. While these examples are aimed at suggesting basic use cases, the list is not exhaustive and various other uses might be designed in the future. Furthermore, the rich ecosystem of open-source libraries building up in PyTorch should provide ideas and help in the development of more complex applications, making use of existing support - e.g. for deep probabilistic programming, deep Gaussian processes, Bayesian hyperparameter optimisation, gradient boosting and various implemented neural network architectures⁵.

By providing a differentiable and GPU-accelerated transit code, PyLightcurve-torch aims to facilitate and widen the use of deep learning in exoplanets research. It bridges the gap between the precision of physical transit models and the scalability of neural networks, allowing for the efficient modelling of thousands of transit light curves now commonly available with exoplanets transit surveys such as Kepler and TESS from space or HATNet, SuperWASP and NGTS from the ground. Conversely, we hope that this code will also make exoplanets science more accessible to the machine learning community and more generally inspire the development of physics-based deep learning applications.

⁵Some basic applications and related tutorials of PyLightcurve-torch are available on this repository: <https://github.com/mariomorvan/pylightcurve-torch-tutorials>.

Chapter 6

Conclusions

"Computers are useless. They can only give you answers."

Pablo Picasso

6.1 Summary

The study of exoplanets is currently one of the most challenging fields in observational astronomy, primarily because of their very faint signals and the existence of multiple observation biases. While our instruments are getting more and more precise, our need for improved precision in the post-processing of astronomical and exoplanet data remains present. Increasing volumes of astronomical data, on the other hand, open the door to data-driven techniques to replace or complement existing processing techniques where appropriate. This work was mainly concerned with the problem of light curve detrending with the study of exoplanetary transits in focus. I set out to automatically learn the instrumental and stellar trends in order to correct for it, and introduced general ideas and methods to do so. I detailed the choices behind the introduced models adapted from sequential deep learning, and presented applications to Spitzer/IRAC, TESS and simulated light curves.

These methods are innovative in that these are the first deep learning methods to process light curves affected by noise and trends. In that respect, this work is opening new questions and will hopefully be pursued and useful to the community.

6.2 Future Work

Here I suggest possible extensions and improvements to this work.

More application datasets

First, developed models could be trained on more real data from the same instruments—such as the all IRAC targets or TESS light curves from full-frame images—or data from different telescopes—e.g. NGTS, CoRoT, ground-based surveys, etc. Larger training volumes are expected to improve the performance for certain applications but also require more care and data preparation. In particular, it would be interesting to explore the possibility for weights transfer between representations learned from different instruments.

Spectroscopic light curves

Since all the models introduced here (deep-learning-based and differentiable transit models) support multiple input dimensions, they could directly be applied to spectral light curves instead of photometric light curves. Another motivation for this is that LSTMs and Transformers are known to perform well in multivariate settings by capturing complex relationships along the feature dimension. We could therefore envisage applications of these models or alike to spectroscopic data, e.g. from HST, JWST, Ariel, or ground-based spectrometers.

Pixel-level

It may be beneficial for an optimal detrending pipeline to operate at pixel level instead of the integrated light curve level (i.e. post-photometry). Again, the various pixels contained in one target’s PSF could perfectly be considered as another feature dimension of the same time series and be fed into our models as a multivariate time series. However, the varying aperture masks between targets (and along the time axis) may complicate the direct application of global models to varying input formats. Nevertheless, I believe that there is potential in linking the steps of photometric reduction and detrending.

Simulation-based validation and pre-training

The use of existing and future instrument simulators could help validate and improve current techniques. Indeed, working with simulated data more systematically will

help monitor the true performance (detrending, detection, fitting...) in various controlled scenarios and experiments. The downside is that a simulation may not perfectly match the observations, and any inference on simulated data may not directly translate into real data. Nevertheless, it might still be possible to leverage large amounts of simulated data through self-supervised pre-training—known to produce highly generalisable representations—for subsequent transfer learning and fine-tuning on a smaller dataset of real data.

6.3 Using Artificial Intelligence Humanly

Will machines replace scientists? The role of AI in exoplanetary sciences and more generally in astronomy appears to be soaring. Real opportunities may be diluted in an avalanche of fashion, and predicting the actual spread of AI techniques in astronomy in a few years time would be no simple task. There are several challenges to solve along the way to make the most of AI techniques intelligently (e.g. interpretability, probabilities, label scarcity, etc.). Currently, computers can not truly mimic our thinking or replace the full process of scientific inference—and even if they did, we would still need to make sure we understand their output. Nevertheless, DL techniques are the best candidates to date to perform efficient pattern-recognition on large astronomical datasets. Furthermore, there is a true potential for AI to complement and support researchers in their daily research. Whether "machines will replace human scientific research" is a rather ill-posed question. Machines can certainly automate some of our most tedious tasks, but this is only going to give us more time to think of more complex questions. Making machines working alongside scientists rather than in their place, here is a more realistic and interesting challenge.

As a final note of caution, there is no reason a priori why technological progress should necessarily translate into an increase in human intelligence. The invention of new technologies and algorithms as well as the exploration of our Universe are intellectually thrilling, but let them not blind our basic faculties of rational and ethical judgement as a species. One of the most obvious and practical learnings from planetary and exoplanetary studies is that the Earth is the only place where

we can reasonably envisage a sustainable future for our species in the centuries to come. From there, the active deterioration and threatening of the conditions of life on Earth as currently carried out by humans can justly raise serious doubts concerning our intelligence as a global species. Finally, the development and use of artificial intelligence techniques for applications in the real world should be subject to systematic ethical challenging, testing and legal ruling to ensure damages are avoided and risks minimised.

Appendix A

Acronyms & Notations

Table A.1: List of acronyms, either astronomy-related (top) or ML-related (bottom).

Acronym	Definition
Ariel	Atmospheric Remote-sensing Infrared Exoplanet Large-survey
CoRoT	Convection, Rotation and planetary Transits
ELT	Extremely Large Telescope
EB	Eclipsing Binary
EPSC	EuroPlanet Science Congress
GMT	Giant Magellan Telescope
HATNet	Hungarian-made Automated Telescope Network
HST	Hubble Space Telescope
IRAC	InfraRed Array Camera
JWST	James Webb Space Telescope
KELT	Kilodegree Extremely Little Telescope
MAST	Mikulski Archive for Space Telescopes
MOST	Microvariability and Oscillations of STars
NASA	National Aeronautics and Space Administration
NGTS	Next-Generation Transit Survey
NICMOS	Near Infrared Camera and Multi-Object Spectrometer
PDC	Pre-search Data Conditioning
PLATO	PLAnetary Transits and Oscillations of stars

PSF	Point Spread Function
SNR	Signal-to-Noise Ratio
SETI	Search for Extra-Terrestrial Intelligence
SAO	Smithsonian Astrophysical Observatory
SPSD	Sudden Pixel Sensitivity Drop-out
STIS	Space Telescope Imaging Spectrograph
TESS	Transiting Exoplanet Survey Satellite
TiH	Titanium Hydride
TMT	Thirty Meter Telescope
WASP	Wide Angle Search for Planets
WFC3	Wide Field Camera 3
<hr/>	
ACF	Auto-Correlation Function
Arima	Autoregressive Integrated Moving Average
AI	Artificial Intelligence
ANN	Artificial Neural Network
AGI	Artificial General Intelligence
BN	Batch Normalisation
CNN	Convolutional Neural Network
CCD	Charge-Coupled Device
CPU	Central Processing Unit
DTST	Denoising Time Series Transformer
ECML	European Conference on Machine Learning
GP	Gaussian Process
GPU	Graphics Processing Unit
ICML	International Conference on Machine Learning
IQR	InterQuartile Range
LSTM	Long Short-Term Memory network
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MCT	Mercury Cadmium Telluride

ML	Machine Learning
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
NeurIPS	Neural Information Processing Systems
NLP	Natural Language Processing
NMMAE	Noise-scaled Masked Mean Absolute Error
PE	Positional Encoding
PKDD	Principles and Practice of Knowledge Discovery in Databases
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
TLCD-LSTM	Transit Light Curve Detrending LSTM

Table A.2: List of notations.

Notation	Definition
a	orbital semi-major axis
AlO	aluminium oxide
b	impact parameter
CH ₄	methane
CO	carbon oxyde
CO ₂	carbon dioxyde
C/O	carbon-to-oxygen ratio
δ	transit depth
e	orbital eccentricity
f	true anomaly
F	flux
g	surface gravity
i	orbital inclination
$I(r)$	limb-darkening law

$InSb$	indium antimonide
K_P	Kepler magnitude
μ	mean
v_0	true anomaly at mid-transit
r	star–planet separation
R_P	planet radius
R_*	star radius
R_\oplus	Earth radius
R_J	Jupiter radius
σ	standard deviation
$Si : As$	arsenic-doped silicon
t	time
T_{eff}	effective temperature
V	visual magnitude

Appendix B

Ariel Machine Learning Challenges 2019 and 2021

This chapter is mainly based on [Nikolaou et al. \(2020\)](#), preprint associated with a submission in review at the Astrophysical Journal led by Nikolaos Nikolaou. For both 2019's and 2021's editions of the Ariel machine learning challenge, my main contribution was to beta-test the challenge's data, and design a baseline solution for participants to get started and compare to.

B.1 Introduction: Why Organising Machine Learning Challenges?

There are various outstanding challenges in exoplanet sciences, and some of them are arguably suited for data-driven machine-learning solutions. Developing and testing these solutions requires both the expertise of exoplanet astronomers on the one hand, and that of computer scientists and machine learning scientists on the other hand. To this end, fostering collaboration between both communities is likely to help solve some of our data problems better and sooner.

Data challenges directly promote the interaction between the astrophysics and machine learning communities. The Ariel ML challenges organised targeted both audiences by being officially organised in the context of the ECML-PKDD¹ 2019,

¹ECML-PKDD, the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, is one of the leading academic conferences on machine learning

2021 conferences while being widely advertised the astronomy community as well.

The two first editions of the challenge have been very successful with more than 100 teams participating from all over the world each year, attracting the interest of researchers from both communities—as evidenced from the top-5 ranked teams and the solutions they submitted. Another advantage of data challenges is to set up benchmark datasets which can still be used after the challenge for researchers to come back to and compare to.

Organising a data challenge is a challenge in itself. Indeed, it takes a lot of work from a whole team to put together a dataset, a problem, a website, etc. within the allocated timeline (for a conference’s track). The two first editions of the Ariel ML challenge went far beyond our expectations, with many participating teams, innovative solutions and new collaborations initiated. This confirmed that the gains were worth the effort, and encouraged us to continue hosting challenges. The current edition of the Ariel Machine Learning challenge is led by [Yip et al. \(2022\)](#) with a dataset generated by [Changeat and Yip \(2022\)](#), and it is hosted at the NeurIPS² conference. This year the goal is to identify a reliable and scalable method to perform planetary characterisation from simulated atmospheric spectra.

In [B.2](#) we describe 2019’s challenge and discuss its top-5 solutions. In [B.3](#) we present the 2021’s update on 2019’s edition.

B.2 Ariel Machine Learning Challenge 2019

B.2.1 Introduction

One of the outstanding challenges to high-precision spectrophotometry of exoplanets is the presence of stellar noise. Here we will address in particular the presence of occulted star spots in the spectrophotometric light curves of the Ariel space mission. The chromatic dependence of spots and faculae can adversely affect the measured exoplanetary transmission spectrum (through a biasing of the derived transit depth) as well as affect other light curve parameters, such as limb-darkening, mid-transit times. This is discussed in detail in (e.g. [Sing et al., 2015](#); [Nikolov et al., 2013](#); [Rabus et al.,](#)

and knowledge discovery, held in Europe every year.

²Neural Information Processing Systems

2009; McCullough et al., 2014; Rackham et al., 2018b, 2019; Zellem et al., 2017; Iyer and Line, 2020, and references therein). There exists a large body of literature on modelling star spot signatures in photometric and radial velocity data (e.g. Boisse et al., 2012; Dumusque et al., 2011, 2014; Lanza et al., 2011; Aigrain et al., 2012; Herrero et al., 2016; Zhao and Tinney, 2020; Gilbertson et al., 2020; Lisogorskyi et al., 2020). Recently, Rosich et al. (2020) proposed a correction of the chromatic effects using Bayesian inverse modelling of long duration spectrophotometric time series data with promising results.

In this publication, we explore the use of machine learning techniques to detect and correct for spot crossings in simulated data of the Ariel space mission. In particular, we report on the top five results of the 1st Ariel Mission Machine Learning Challenge (henceforth: the Challenge), which was concerned with the task of *correcting transiting exoplanet light curves for the presence of stellar spots*. The primary goal of the Challenge was thus to investigate if machine learning approaches are in principle suited to correcting star spot crossings in spectrophotometric light curves across a large range of stellar and planetary parameters as well as observational signal-to-noise regimes.

As many problems in the field of exoplanetary science, the issue of star spot crossings is characterised by a combination of challenges: (i) a large amount of data to process³, (ii) low signal-to-noise ratio, (iii) an underlying pattern which is non-linear and whose parametric form is a priori unknown, (iv) the available information comes in multiple forms (time dependent and independent), and finally (v) a high degree of degeneracy. These issues are commonly addressed by machine learning approaches.

This takes us to the second objective of the Challenge: promoting the interaction between the astrophysics and the machine learning communities. To this end, the Challenge targeted both audiences by being officially organised in the context of the ECML-PKDD 2019 conference⁴ and also having a strong presence in the joint

³As Ariel is an upcoming space mission, the data in our case are obtained via simulations.

⁴ECML-PKDD, the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, is one of the leading academic conferences on machine learning and knowledge discovery, held in Europe every year.

EPSC-DPS 2019⁵ conference via a dedicated session. The Challenge ran from April to August 2019. More than 120 teams participated, and it attracted the interest of researchers from both communities—as evidenced from the top-5 ranked teams and the solutions they submitted. As such, we consider the secondary objective of the Challenge has been met successfully.

But what of the main goal of the Challenge, i.e. automating the extraction of useful parameters from transiting exoplanet light curves in the presence of stellar spots? Many solutions achieved the desired precision of 10 ppm in photometric flux for correctly predicting the relative transit depth per each wavelength from the noisy light curves.

The solutions of the top-5 ranking teams that participated in the Challenge are presented in detail in this paper. Most solutions amount to constructing highly non-linear (w.r.t. the raw data) models with minimal preprocessing using artificial neural networks and/or ensemble learning methods⁶. As we will see however, there exist comparably good—in terms of the precision of the obtained predictions—approaches that involve obtaining meaningful (i.e. informed by physics) statistics from the light curves and then training models that are linear w.r.t. them.

Just like the Challenge itself, this paper also intends to serve a dual purpose. Its primary goal is to describe the research problem of obtaining good predictions of the relative transit depth per each wavelength from simulated Ariel-like light curves distorted by photon noise and stellar spot noise, along with the solutions provided by the Challenge’s winners and their implications. Its secondary aim is to promote interaction between exoplanetary scientists and machine learning researchers. As such it is written in a language accessible to both audiences and—we hope—it contains useful information for exoplanetary scientists wishing to organise their own machine learning challenge.

⁵The European Planetary Science Congress (EPSC) and the American Astronomical Society’s Division of Planetary Science (DPS) held a Joint Meeting at 2019.

⁶Ensemble methods are machine learning algorithms that construct powerful predictive models by combining multiple weaker predictors (Polikar, 2006).

B.2.2 The Challenge

B.2.2.1 Data Generation

For the purposes of the Challenge, we used the Ariel target-list produced by [Edwards et al. \(2019\)](#) to generate simulated light curves for all the 2097 planets in the list. For every planet we produced 55 light curves, one for each wavelength channel corresponding to Ariel Tier 2 resolution (between 0.5 and 8.0 μm). In addition, all the light curves covered observations of 5 hours, centred around the transit, with a time step of one minute. The uniformity in wavelength and time resolution is not realistic, but only used to make the dataset accessible to all the participants without the need of renormalisation (which would require knowledge on the transit modelling).

The simulated light curves were computed as follows:

1. As a first step we calculated the limb-darkening coefficients (using the quadratic law) for every host star in the target list and for every wavelength channel. We used the ExoTETHyS package ([Morello et al., 2020](#)) and the stellar parameters for temperature and gravity provided in the target list, assuming zero metallicity for all the stars (the effect of metallicity is not strong). Also, we did not use the Ariel throughput as in this study we were only interested in the narrow wavelength channels, and any intra-channel variations due to the Ariel throughput are minimal.
2. We then calculated the planet-to-star radius ratio, R_p/R_* , for every planet in the target list and for every wavelength channel. This calculation was made using the TauRex atmospheric retrieval framework for exoplanets [Waldmann et al. \(2015\)](#) and the planet parameters for temperature, mass and radius (all provided in the target list), assuming the presence of water vapour and methane in the atmosphere with abundances that varied uniformly at random from planet to planet between 0.001% and 0.1%. The values for the abundances were an arbitrary choice, as the scope of using a spectrum was only to include some variability, of any kind, in the R_p/R_* parameter from one wavelength channel

to another.

3. The next step was to define the spot model parameters for every host star in the target list. These parameters were:

- Spot coverage: This parameter corresponds to the percentage of the stellar surface that is covered by spots. We set this parameter to 10% for every host star in the target list. In reality this parameter decreases with stellar temperature, and initially we incorporated this in the model. However, it became clear that in the more realistic case, the number of spots that influence the light curves is very small, leading to almost noise-free data. For this reason, we chose to use the fixed value of 10% in order to have a stronger spot effect on our data. This choice resulted in a simulated dataset with many more spot-crossing events than in a real dataset, suitable for the purposes of the challenge.
- Spot temperature: This parameter corresponds to the effective temperature of the spots, which is naturally lower than the effective temperature of the star. We calculated this parameter for every host star in the target list as a function of its temperature (T_* , provided in the target list), as described in [Sarkar \(2017\)](#), adjusted from [Andersen and Korhonen \(2015\)](#):

$$T_{\text{spot}} = T_* - (0.0001343 \times T_*^2 - 0.6849 \times T_* + 1180.0) \quad (\text{B.1})$$

- Spot contrast: This corresponds to the contrast between the brightness of the stellar surface and the brightness of the spots. We calculated this parameter for every host star in the target list and for every wavelength channel by integrating the respective PHOENIX stellar models ([Husser et al., 2013](#)) within each wavelength channel and dividing them.
4. Following the definition of the spot model parameters we created a set of spots for every host star in the target list. The spots were generated one by one, until

the 10% surface coverage was reached, and it was given three parameters:

- Latitude - uniformly at random generated number between -85 and 85 degrees
- Longitude - uniformly at random generated number between 0 and 360 degrees
- Angular diameter - randomly generated using a log-normal distribution, as described in [Sarkar \(2017\)](#), based on [Bogdan et al. \(1988\)](#):

$$\frac{dN}{dA} = M_A \exp \left[-\frac{(\ln A - \ln \langle A \rangle)^2}{2 \ln \sigma_A} \right] \quad (\text{B.2})$$

where N is the number of spots, A is the area of the spots, M_A is the maximum of the distribution (adjusted to result in 10% of total coverage), $\langle A \rangle = 0.62 \times 10^{-6} A_{1/2\odot}$ is the mean of the distribution, and $\sigma_A = 3.8 \times 10^{-6} A_{1/2\odot}$ is the standard deviation of the distribution.

5. With the set of spots generated for each star in the target list, we used the KSint package ([Montalto et al., 2014](#)) to generate the spot-distorted light curves for every planet in the target list and for every wavelength channel. The input parameters for each light curve were: the set of spots, (number, position and dimensions of all the spots), the spot contrast parameter, the limb-darkening coefficients, the planet-to-star radius ratio, the stellar density (calculated from the stellar mass and radius provided in the target list), and the planet orbital parameters (period and inclination, provided in the target list) and a viewing angle to make sure that the transit happens at the middle of the observation.
6. The final step was to add Gaussian noise to the light curve. No additional instrument systematics were assumed, as we aimed for the challenge to focus on correcting for the noise resulted from the stellar spots. The standard deviation of the Gaussian noise added was calculated from the overall noise on the transit depth estimation provided in the target list. This noise value depends on the stellar magnitude, the stellar temperature, the wavelength channel and

the characteristics of the Ariel instrument. It is beyond the scope of this work to describe exactly how this level of noise is estimated. We refer the interested reader to [Edwards et al. \(2019\)](#) for a detailed description.

This process resulted in generating data for 2097 simulated observations, consisting of 55 light curves each (one per wavelength). We repeated the process 10 times with different instances of the spot set (step 4). This resulted in 20970 simulated observations consisting of 55 light curves each, distorted by stellar spots. Finally, for each instance of the spot set, 10 different instances of additive Gaussian photon noise were introduced (step 5). This resulted in 209700 simulated observations consisting of 55 light curves each, distorted by both stellar spot and photon noise. These 209700 simulated observations formed the final dataset of the Challenge. The different instances of the spot set were included to mimic multi-epoch observations, where the spot pattern is expected to change, while the different instances of additive Gaussian photon noise were included to mimic continuous observation where the spot pattern is not expected to change. Note that the two sources of noise (spots and Gaussian) are treated as independent. Most of the generated light curves only contained a single transit event, however a few of them included planets with small enough orbital periods to allow for multiple transits⁷.

Naturally, these details were unknown to the participants and neither were they used in the baseline solution. The aim of the Challenge was to infer the relative radii, either by explicitly modelling and subtracting, or by learning to ignore the photon and/or the stellar noise (or both).

B.2.2.2 Dataset Description & Problem Statement

Each *data point* (a.k.a. an *observation* or an *example* in machine learning terminology) consists of a set of 55 noisy light curves (one per wavelength, corresponding to Ariel Tier 3 target resolution). Each light curve is a time series of 300 time steps corresponding to 5 hours of observation by Ariel. We shall denote with $x_{ij}^{(t)}$ the relative flux at time step $t \in [1, 2, \dots, 300]$ of the light curve at wavelength $j \in [1, 2, \dots, 55]$ of the i -th example. By $\mathbf{x}_{ij} = [x_{ij}^{(1)}, x_{ij}^{(2)}, \dots, x_{ij}^{(300)}]^\top$ we denote the entire light

⁷In case the light curve contained multiple transits, one of them was centred.

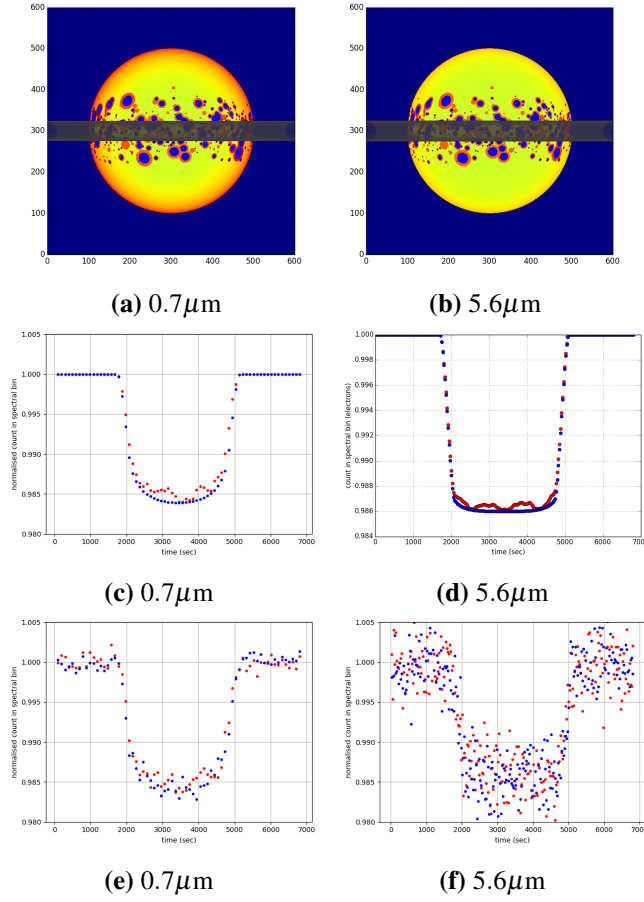


Figure B.1: Examples of simulations for two of the 55 wavelength channels, $0.7\mu\text{m}$ and $5.6\mu\text{m}$. (a) & (b), stellar surface simulations of a spotty star. Grey line shows the planetary transit trajectory. The stellar surface limb brightness varies with wavelength. (c) & (d) Normalised observed flux as the planetary transits across the star without stellar photon noise. Blue shows the perfect transit across a spotless star; red shows the transit across a spotty star. (e) & (f) same as (c) & (d) but with stellar photon noise added.

curve at wavelength j of the i -th example. Finally, with $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i55}]$ we denote all 55 light curves of the i -th example.

Along with the light curves, 6 additional stellar and planetary parameters (all knowable in advance) were provided: the orbital period, stellar temperature, stellar surface gravity, stellar radius, stellar mass & stellar K magnitude. We shall denote these as $z_{i1}, z_{i2}, \dots, z_{i6}$, respectively, for the i -th example. Finally, with $\mathbf{z}_i = [z_{i1}, z_{i2}, \dots, z_{i6}]$ we shall refer to all the additional parameters of the i -th example collectively.

The noisy light curves and the 6 additional stellar and planetary parameters all

constitute the quantities known in advance that we can use to alleviate the problem of stellar spots. In machine learning terminology they are the *features (independent variables)* in our prediction task.

The goal is to construct a model that uses these to predict a set of 55 real values, the relative radii $[R_p/R_*]_{ij}$ (one per wavelength j , for any given data point i). In machine learning terminology this is a *multi-target regression task*. The relative radii to be predicted are the *targets (dependent variables)* of the multi-target regression problem. For convenience, we shall henceforth denote the relative radius at wavelength j of the i -th example, $[R_p/R_*]_{ij}$, with y_{ij} . Finally, with $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{i55}]$ we shall refer to all the relative radii of the i -th example collectively. Note the planet to host star relative radius R_p/R_* is directly connected to the transit depth of the light curve, as the latter is equal to $\left(\frac{R_p}{R_*}\right)^2$.

The value of the 55 targets is known only for the *training examples* (the *statistical sample*). The goal of the learning task is—ideally—to construct a model $f(\mathbf{X}, \mathbf{z}) = \hat{\mathbf{y}}$ such that $\mathbb{E}[L(\mathbf{y}, \hat{\mathbf{y}})]$ is minimised, where $L(\mathbf{y}, \hat{\mathbf{y}})$ denotes some measure of difference between the predictions $\hat{\mathbf{y}}$ and their corresponding true values \mathbf{y} and \mathbb{E} denotes expectation over the joint distribution of $\mathbf{X}, \mathbf{z}, \mathbf{y}$, i.e.—in statistical terminology—the underlying *population* from which the sample is drawn.

Once models are trained, they are evaluated on a separate *test set*. The predictive performance of a model on a previously unseen test set (drawn from the same distribution as the training set), serves as a proxy for its performance in the population, the latter being intractable. The features of the test set examples $\{(\mathbf{X}_i, \mathbf{z}_i) | i \in \text{Test}\}$ were provided to the participants and they had to upload their model’s predictions $\{\hat{\mathbf{y}}_i | i \in \text{Test}\}$ on them. The ground truth $\{\mathbf{y}_i | i \in \text{Test}\}$ for the test set examples was unknown to the participants in the duration of the Challenge. It was only used to produce a ranking score for their submitted solution, which we describe in the next subsection.

B.2.2.3 Evaluation

All data points generated for a uniformly random set of 1677 out of the 2097 of the total planets (i.e. about 80% of the generated data points) were used as training data.

All data points generated for the remaining 420 planets were used to form the test set (i.e. were only used for evaluation). That is, the training and test sets not only contained no data point in common, but they also contained no data point from the same planet in common.

After producing a model (i.e. a solution to the problem), the participants could upload the predictions of the model on the Challenge’s website. Subsequently, this would assign a score on the model based on the quality of the predictions. The participants were ranked on a leaderboard on the basis of their best solution and the progress of each participant’s solutions in time was tracked to inform them of the impact of each change they made on the resulting model’s predictive performance. The leaderboard ranking determined the winners of the Challenge that would receive prizes (top 2 participants) and the top-5 participants whose solutions we will present in Section B.2.3.

The score assigned to each solution was a weighted average of the absolute error per target (i.e. on the relative radii) across all test set examples i and all wavelengths j :

$$Score = 10^4 - \frac{\sum_{i \in Test} \sum_{j=1}^{55} w_{ij} 2y_{ij} |\hat{y}_{ij} - y_{ij}|}{\sum_{i \in Test} \sum_{j=1}^{55} w_{ij}} 10^6, \quad (B.3)$$

where y_{ij} is the true relative radius and \hat{y}_{ij} the predicted relative radius of the j -th wavelength of the i -th test set example and the corresponding weight w_{ij} is given by:

$$w_{ij} = \frac{1}{\sigma_{ij}^2 \delta_{F_{ij}}^2}, \quad (B.4)$$

with σ_{ij}^2 being the variance of the relative stellar flux caused by the observing instrument at the j -th wavelength of the i -th example and $\delta_{F_{ij}}^2$ the variation of the relative stellar flux caused by stellar spots in the j -th wavelength of the i -th example. The value of σ_{ij} is an estimation based on an Ariel-like instrument, given its current

design, while $\delta_{F_{ij}}$ is calculated based on stellar flux F_{ij}^{star} and the spot flux F_{ij}^{spot} in the j -th wavelength of the i -th example:

$$\delta_{F_{ij}} = 0.1 \left(1 - \frac{F_{ij}^{spot}}{F_{ij}^{star}} \right). \quad (\text{B.5})$$

As we see, both sources of noise (photon & stellar spot) are wavelength-dependent and target-dependent (they depend on the star, therefore are different for each data point).

The higher the score, the better the solution's ranking. The maximum achievable score is 10000 (if $\hat{y}_{ij} = y_{ij}, \forall i, j$). The score is not lower-bounded (i.e. can be negative), but even naive 'reasonable' models (e.g. predicting the average target value for all test data points) would not produce scores below 4000.

The weights w_{ij} of each target were unknown to the participants⁸. A sensible strategy would thus be to try to predict all of them reasonably well. In other words, to train a model to minimise an unweighted loss $L(\mathbf{y}_i, \hat{\mathbf{y}}_i)$ like the Mean Squared Error (MSE), $L(\mathbf{y}_i, \hat{\mathbf{y}}_i) = (\hat{\mathbf{y}}_i - \mathbf{y}_i)^2$, the Mean Absolute Error (MAE), $L(\mathbf{y}_i, \hat{\mathbf{y}}_i) = |\hat{\mathbf{y}}_i - \mathbf{y}_i|$, or their relative error counterparts: $L(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \left(\frac{\hat{\mathbf{y}}_i - \mathbf{y}_i}{\hat{\mathbf{y}}_i} \right)^2$ or $L(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \frac{|\hat{\mathbf{y}}_i - \mathbf{y}_i|}{\hat{\mathbf{y}}_i}$, respectively. Indeed, this is the approach taken by the top-5 participants and in training the baseline model.

B.2.2.4 Rules, Logistics & Organisation

To allow for the broadest possible participation, the set of rules of the Challenge was the minimal possible. There was no restriction on the models, algorithms or data preprocessing techniques, neither on the programming languages, environments nor tools used for their implementation. The participants were also free to use data augmentation techniques, pre-trained models or any prior domain knowledge not included in the provided dataset. Finally, they were free to choose their own way of splitting the training data between training and validation sets.

⁸For transparency of the evaluation process, the w_{ij} coefficients of the test set examples, along with the ground truth (target values y_{ij}) became available after the end of the Challenge.

The participants were limited to 1 submission every 24 hours. This was a measure taken to limit traffic on our website and—most crucially—to prevent the extent to which the solutions would be overfitting to the test set. Indeed, although the test set contains previously unseen examples by the model and the participants could not have access to the ground truth itself, the presence of a leaderboard is effectively causing some *information leakage from the test set*. Simply put, just adapting the strategies to the ranking score signal, participants could increase their scores by effectively overfitting on the particular test set. Limiting the number of daily submissions alleviated this effect. In retrospect, an even stronger strategy to prevent this would have been to only use part of the test set to produce the leaderboard ranking score during the Challenge and only use the full test set to produce the final ranking after the Challenge closes. In future machine learning challenges we will adopt this evaluation scheme. For now, we should keep in mind that small differences in the ranking scores of solutions presented in Section B.2.3 are not necessarily indicative of true generalisation (i.e. ability to predict well on new examples).

The participants were allowed to form teams, provided they participated in only one entry. The remaining rules handled how prizes would be split among teams, how ties would be handled and ensuring that any winning entry would have to beat the baseline model.

B.2.2.5 Description of Solutions

To facilitate comparisons among the solutions discussed in the paper and to demonstrate the typical steps of training and evaluating models using machine learning methodology, we split the description of the solutions into 3 parts: (i) preprocessing, (ii) model / architecture, (iii) training / optimisation.

The ‘preprocessing’ part will describe any transformation of the raw data (either in terms of features or of observations) before giving it as input to a learning algorithm. The ‘model’ part is concerned with the general *class of models* (i.e. their *parametric form*) which the learning algorithm is exploring (e.g. artificial neural

networks of a given *architecture*⁹, random forests of 10 trees of maximal depth 5, linear models of the form $y = ax_1 + bx_2 + c$). Finally, the ‘training’ part is concerned with the specifics of the optimisation of the *parameters* of the model (i.e. the weights of the neural network, the derivation of the decision trees or the inference of the linear coefficients a, b, c in the examples below). It covers the *hyperparameters* used in the learning/optimisation algorithm, along with the loss function it minimises and the final evaluation method.

Wherever necessary, we will clarify the purposes behind modelling choices or training methodologies in all solutions described. However, a detailed treatment of models like artificial neural networks (ANNs) is beyond the scope of this paper. We direct the interested reader to [Goodfellow et al. \(2016\)](#) and [Chollet \(2017\)](#).

B.2.2.6 Baseline Solution

As a baseline solution, we trained a fully connected ANN¹⁰ on a sample of 5000 training examples selected uniformly at random. The neural network uses all 55 noisy light curves, \mathbf{X}_i to predict the 55 relative radii directly. It does not make use of the additional stellar & planetary parameters \mathbf{z}_i .

Preprocessing

The noisy light curves have undergone the following preprocessing steps:

i) Each light curve was smoothed using a moving median of window 3 (i.e. each value replaced by the median of itself and its two adjacent values). This was done to remove flux values that are obvious outliers.

ii) In any light curve, any value (relative flux) that was above 1 was clipped to 1. This was done because the maximal relative flux during transit is 1.

iii) All values were normalised for the transit depths to lie roughly within the range $[0, 1]$. Doing so allows for faster and more stable training of models like ANNs. The normalisation was carried out per wavelength and was performed as follows:

First, we computed the average transit depths per wavelength from the target

⁹By the term "architecture" we collectively refer to the number, type and connectivity of the neurons comprising a neural network.

¹⁰Fully connected ANNs are the earliest and most popular type of ANN architecture. They are also known as multi-layer perceptrons (MLPs) or ‘dense’ neural networks.

values \bar{y}_j on a sample of 10000 random training examples. For every wavelength j , we then applied the transformation:

$$x_{ij}^{(t)} \leftarrow (x_{ij}^{(t)} - (1 - 2\bar{y}_j^2))/2\bar{y}_j^2.$$

This was done to have the maximal relative flux values at exactly 1 and the transit depths around 0, leveraging the fact that the transit depths of the light curves are the squares of the relative radii (targets).

Model/Architecture

We used a fully connected ANN with 5 2D-hidden layers, all of which consisted of $1024 \text{ units} \times 55 \text{ channels}$, the j -th channel receiving as input the light curve \mathbf{x}_{ij} for each example. After these, we added a flattening layer followed by a linear layer of 55 outputs, the j -th output corresponding to the predicted relative radius \hat{y}_{ij} of each example. All other activation functions were rectified linear units (ReLU).

Training/Optimisation

No batch normalisation, regularisation or dropout was applied in the training of the baseline model. The 5000 observations used were split into 4020 training and 980 validation examples (i.e. approximately 80% training & 20% validation split) in such a way that the two sets contained no planets in common. The model was trained by minimising the average MSE across all wavelengths using the Adam optimiser (Kingma and Ba, 2015) with a learning rate of 10^{-4} decaying with a rate of 0.01 and a batch size of 128. All remaining hyperparameters were set to default Keras¹¹ values. The model was trained for a maximum number of 5 epochs without early stopping.

B.2.3 Top-5 Solutions

By the end of the Challenge, 13 teams had beaten the score attained by the baseline solution we just presented. In this section, we will present the top-5 ranked solutions. Their relative ranking in the final leaderboard and scores they achieved under Eq.(B.3) are shown in Table B.1.

¹¹<https://keras.io>

Rank	Team	Score
1	SpaceMeerkat	9813
2	Major Tom	9812
3	BV Labs	9808
4	IWF-KNOW	9805
5	TU Dortmund University	9795
14	Baseline	8726

Table B.1: Final leaderboard showing rank & score under Eq.(B.3) achieved by each of the top-5 entries and the baseline.

B.2.3.1 SpaceMeerkat’s solution

The *SpaceMeerkat* team is composed of James M. Dawson, an Astrophysics PhD student at Cardiff University. *SpaceMeerkat*’s solution is a 1D-CNN, designed to retain architectural simplicity, while exploiting the power of GPU accelerated machine learning. The largest gain in the model’s predictive power came from the extensive testing of different preprocessing operations.

Preprocessing

The data was split into 80% training and 20% test sets. In order to remove outlier flux values in the raw light curves, an initial smoothing was conducted on each time series \mathbf{x}_{ij} . The mean flux value in each non-overlapping bin of width 5 was calculated in-place along each time series leaving each observation \mathbf{X}_i as a smoothed multichannel array of dimensions 60×55 . A normalisation operation was performed on the training set prior to its use for training machine learning models. For each of the 55 wavelengths, the medians across all data points of the lowest 1% of flux values in each light curve for a given wavelength were calculated. These 55 percentile medians (henceforth ‘median offsets’) are therefore equal to

$$\kappa_j = \text{med}\{\mathbf{P}_{1\%}(\mathbf{x}_{ij}^{(t')})\}, \quad (\text{B.6})$$

where $\mathbf{P}_{1\%}(\mathbf{x}_{ij}^{(t')})$ denotes the 1st percentile of the set of all flux values $x_{ij}^{(t')}$, $t' \in \{1, 2, \dots, 60\}$ for a given data point i and wavelength j , and $\text{med}\{\cdot\}$ denotes median across all data points i . The light curves were then divided by 1 minus the median

offsets and the resulting flux values were thus

$$x_{ij}^{(t')} \leftarrow x_{ij}^{(t')} / (1 - \kappa_j).$$

This normalisation allowed the data to lie roughly within the range $[0, 1]$ but with leniency for allowing the existence of extremely shallow or deep transits. Any remaining flux values above the normalisation range were clipped to 1. This was done to encourage the model to focus on the lower flux valued regions where most of the transit-depth information lies. The preprocessing of light curves makes use of Astropy¹², a community-developed Python package for Astronomy (Astropy Collaboration et al., 2013, 2018).

Model/Architecture

The model used in this solution is a convolutional neural network (CNN) (LeCun et al., 1995)¹³. The data is presented to the CNN as a 1D vector and 1D convolutions & pooling operations are applied in order to maintain a principled simplicity to the final solution. The architecture of the CNN is shown in Table B.2. The model was built using PyTorch 0.4.1¹⁴, an open-source machine learning framework for Python users. The output of layer ‘Lc5’ in Table B.2 is concatenated with the additional stellar & planetary parameters: the orbital period, stellar surface gravity, stellar radius, stellar mass & stellar K magnitude, i.e. $[z_{i1}, z_{i3}, \dots, z_{i6}]$ for each example, to form the 1D linear input for layer ‘Lc6’. The additional parameters did not undergo any normalisation and were presented to the network in their raw form.

Training/Optimisation

The CNN was trained for 75 epochs (i.e. was presented with the entire training set 75 times), on a single NVIDIA TITAN Xp GPU. The model was trained using batches

¹²<http://www.astropy.org>

¹³CNNs are designed to excel in tasks in which translational invariance is important, i.e. we are looking for particular patterns anywhere in the input data. As such, they are especially popular in image-based tasks. However, they are very successful even outside this setting, as they effectively reduce the number of trainable parameters of a neural network (compared to a feedforward ANN of the same depth). This means they are more computationally efficient to train and more resistant to overfitting.

¹⁴<http://pytorch.org/>

Name	Layer/Operation	Dimensions	Filter
Input	None	(256,1,1,3300)	None
Conv1	1D convolution	(256,32,1,3300)	(1,3)
ReLU	ReLU	None	None
AP1	1D average pool	(256,32,1,1650)	(1,2)
Conv2	1D convolution	(256,64,1,1650)	(1,3)
ReLU	ReLU	None	None
AP1	1D average pool	(256,64,1,825)	(1,2)
Conv3	1D convolution	(256,128,1,825)	(1,3)
ReLU	ReLU	None	None
AP1	1D average pool	(256,128,1,275)	(1,2)
Lc1	Linear	(256,1,1, 35200)	None
ReLU	ReLU	None	None
Lc2	Linear	(256,1,1, 2048)	None
ReLU	ReLU	None	None
Lc3	Linear	(256,1,1, 1024)	None
ReLU	ReLU	None	None
Lc4	Linear	(256,1,1, 512)	None
ReLU	ReLU	None	None
Lc5	Linear	(256,1,1, 256)	None
ReLU	ReLU	None	None
Lc6	Linear	(256,1,1, 60)	None
Output	None	(256,1,1,55)	None

Table B.2: The CNN architecture used in the solution by the SpaceMeerkat team (Ranked 1st). The table follows the standard PyTorch format. The 1st column lists the name of each layer/operation, the 2nd column its type, the 3rd the dimensions of its output tensors (hence inputs to the next layer). These follow the convention (batch size, number of channels, height, width). The filter column shows the dimensions (height, width) of kernels used to perform the convolution and pooling operations. Layer ‘Lc6’ is notable as this is where the additional planetary parameters \mathbf{z} are introduced into the network.

of 256 examples. Rather than presenting the CNN with examples of dimensions 60×55 (as generated by the preprocessing step), each example was flattened into a single vector of length 3300. Initial investigation showed that 1D convolutions over the flattened inputs produced significantly better results than 2D convolutions over the 2D preprocessed inputs. The model was trained by minimising the MSE loss (see §B.2.2.3) using the standard Adam optimiser and an initial learning rate of 1×10^{-3} decaying by 10% the existing rate, every epoch. No early stopping was used, as we observed no increase of the validation error during training to indicate the presence of overfitting. No additional form of regularisation (e.g. batch

normalisation, dropout, or explicit regularisation) was used in the training procedure. All remaining hyperparameters were set to default PyTorch values. The code for this solution is publicly available on GitHub¹⁵.

B.2.3.2 Major Tom's solution

Major Tom took second place on the ARIEL ML challenge scoreboard. The team composed of machine learning researchers from the Data Science Research and Analytics (DSAR) group at the University of Tuebingen (Germany). The goal of the team's solution is to provide an easy-to-use ML tool, with minimal data preprocessing effort and a fast inference step. The result is a fully-integrated deep learning solution whose final code is publicly available online¹⁶.

Preprocessing

The main motivation behind this solution was to create a robust statistical model that can handle outliers and noisy data. Therefore, we deliberately do not apply any heavy preprocessing to the data beyond the rescaling of the features and the targets. Since all measurements in time series \mathbf{x}_{ij} are mostly distributed around 1 (see, for example Figure B.3), we used the following rescaling of the data, in order to emphasise the differences between measurements:

$$x_{ij}^{(t)} \leftarrow (x_{ij}^{(t)} - 1) \times 1000.$$

We apply a similar transformation technique to the target variable y :

$$y_j \leftarrow y_j \times 1000.$$

Model/Architecture

We used a multiple-input and multiple-output ANN model with fully-connected (FC), batch normalisation (BN) (Ioffe and Szegedy, 2015), and Dropout (Srivastava et al., 2014) layers¹⁷. The final architecture is presented in Figure B.2. It consists

¹⁵Solution by *SpaceMeerkat* (Ranked 1st): <https://github.com/SpaceMeerkat/ARIEL-ML-Challenge>

¹⁶Solution by *Major Tom* (Ranked 2nd): <https://github.com/unnir/Ariel-Space-Mission-Machine-Learning-Challenge>

¹⁷Both batch normalisation and dropout are commonly used techniques to prevent overfitting in neural networks.

of two separate branches. The first branch uses as input the light curves \mathbf{X}_i , and the second, the additional stellar & planetary parameters \mathbf{z}_i . After several non-linear transformations, the outputs of the two branches are concatenated into one and higher level non-linear features combining information from both are extracted. The output layer has 55 neurons, the j -th neuron mapping to the (rescaled) predicted relative radius y_{ij} of a given example. We utilised exponential linear unit (ELU) activations in all but the last two layers, where ReLUs and linear activation functions are used, respectively.

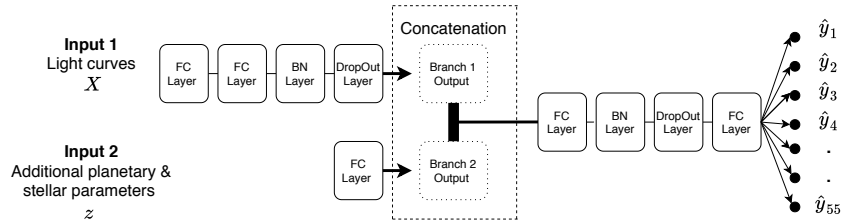


Figure B.2: The deep learning model architecture proposed by the Major Tom team (Ranked 2nd). The model has two separate inputs: one for the measurements \mathbf{X}_i , the second for the additional stellar & planetary parameters \mathbf{z}_i . The two branches are subsequently concatenated and higher level non-linear features combining information from both are extracted.

Training/Inference

We train the ANN using the *NAdam* optimisation algorithm (Dozat, 2016) and a *cyclic learning rate* as described in Smith (2017). The number of epochs was set to 1000, and the batch size to 3048. We selected the MSE as the loss function. We train the proposed model using 10-fold Cross-Validation with early stopping based on the validation loss with the patience equals to 20. The neural network was implemented using the Keras/Tensorflow deep learning framework (Abadi et al. (2015)). The entire training step took ≈ 30 hours using a single NVIDIA P100 GPU.

For the inference step, we used an ensemble consisting of all 10 models produced in the cross-validation steps; the final prediction is the average of all estimates from the 10 models.

B.2.3.3 BVLabs' solution

The team BVLabs took third place in the challenge. It is composed of researchers and data scientists from the Jožef Stefan Institute and Bias Variance Labs. The

team’s solution relied on denoising the input data, the use of tree ensembles and fully-connected neural networks.

Preprocessing

For each star-planet pair, we have 10 stellar spot noise instances and for each stellar spot noise instance we have 10 Gaussian noise instances. The data for each star-planet pair can therefore be represented as a tensor with dimensions $(10, 10, 55, 300)$. For a fixed stellar spot noise instance, we computed the element-wise mean flux matrix over the 10 Gaussian noise instances which decreases the noise in the data. This can be seen as aggregating multiple measurements of the same target to decrease the variance of the observation. We are left with tensors with dimensions $(10, 55, 300)$. Next, we compute element-wise medians over the 10 stellar spot noise instances, leaving us with tensors with dimensions $(55, 300)$. An example of the result of this denoising process is presented in Figure B.3a.

The maximum flux (without noise) is always 1, whereas the minimal flux gives information about the planet radius. To further compensate for the noise, we do not use the minimal flux directly. Instead, we calculate two values: the minimum of the average of 3 consecutive flux values, and the median of the 10 lowest flux values. An example of the extracted values is shown in Figure B.3b.

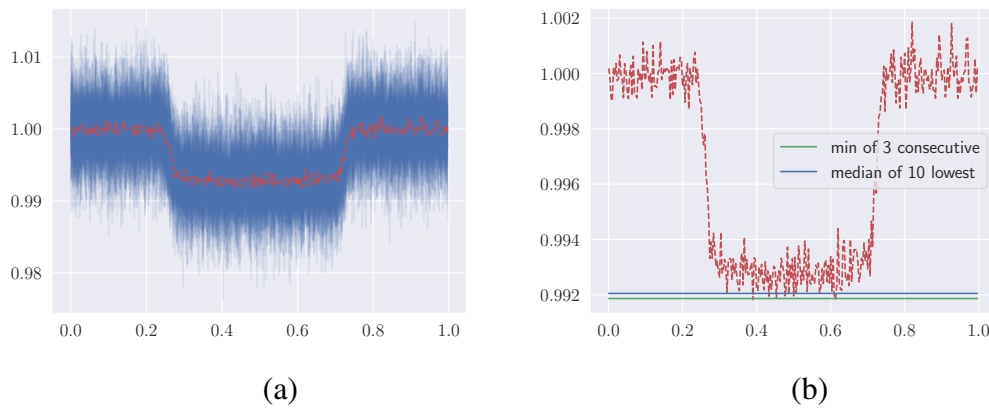


Figure B.3: Preprocessing of the light curves by the BV Labs team (ranked 3rd). Image (a) shows the light curve before (blue) and after (red) noise instance aggregation. Image (b) shows the features extracted from the denoised data. Both images show the data for star-planet pair 113, channel 25.

We also estimated the amount of energy that stars emit at operating wavelengths

of the ARIEL spacecraft. Tinetti et al. (2016) list the 5 operating ranges of ARIEL. We divided each range into 11 bins of equal length, to get the estimates of the 55 wavelengths. To calculate the energy at a given wavelength, we used Planck's law

$$B(\lambda, T) \propto \frac{1}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1},$$

where λ is the wave-length, h is the Planck's constant, k_B is the Boltzmann's constant and c is the speed of light. The star temperature T was one of the 6 stellar and planetary parameters (see Section B.2.2.1). In total, we used 171 ($3 \cdot 55 + 6$) features: 3 features for each of the 55 channels (the 2 extracted from the flux values and the energy emitted) and the 6 stellar and planetary parameters.

Model & Training

Our best performing model was a heterogeneous ensemble consisting of three models. The first model was a random forest of 500 trees (Breiman, 2001), as implemented in `scikit learn`¹⁸. The second model was an extreme gradient boosting (Friedman, 2001) ensemble of 150 trees, as implemented in the `xgboost` library¹⁹. For both methods the parameters were optimised with cross-validation, and a separate model was learned for each channel. The third model was a multi-target (one model for all 55 channels) fully connected neural network with one hidden layer of 100 neurons. We used batch normalisation, dropout (with a rate of 0.2) and ReLU activations. The network was optimised with the Adam optimiser for 1000 epochs, with a constant learning rate 10^{-3} . As the loss function, average MSE across all targets was used. The network was implemented in PyTorch.

The weights of these 3 models in the final heterogeneous ensemble were optimised manually, with the best results obtained with a weight 0.15 assigned to the random forest, 0.25 to XGBoost and 0.6 to the neural network. The code is available online²⁰.

¹⁸<https://scikit-learn.org/>

¹⁹<https://github.com/dmlc/xgboost>

²⁰Solution by *BV Labs* (Ranked 3rd): <https://github.com/bvl-ariel/bvl-ariel>.

B.2.3.4 IWF-KNOW's solution

IWF-KNOW took the fourth place on the ARIEL ML challenge scoreboard, and comprised of researchers and data scientists from the *Space Research Institute* (Austria), *Know-Center* (Austria) and the *University of Passau* (Germany). In contrast to the other top scorers who relied on deep learning approaches, their solution is based on a set of linear regressors, each of which is fast to train and easy to interpret (see Figure B.4). The corresponding scripts can be found on Zenodo²¹.

Preprocessing

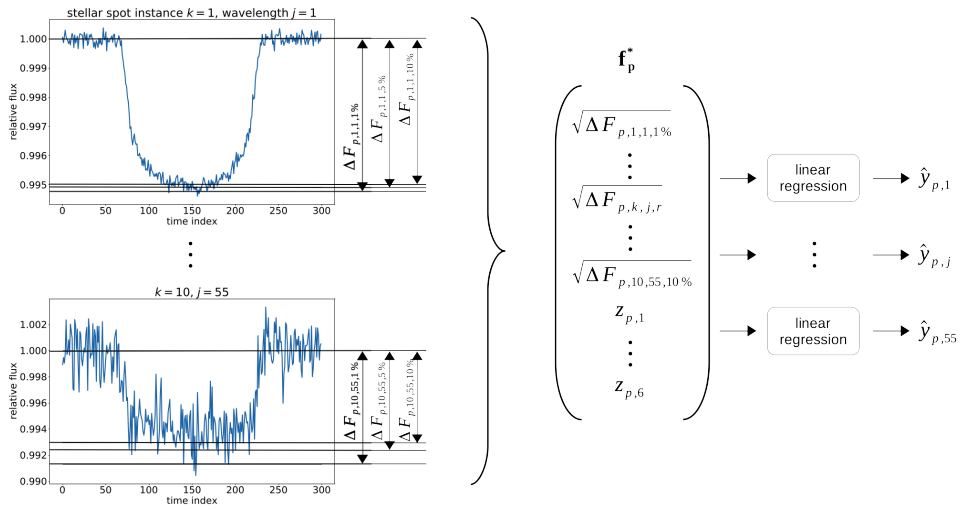


Figure B.4: Regression pipeline of IWF-KNOW (ranked 4th). The light curves on the left are two examples in $\tilde{\mathbf{X}}_{p,k}$. The minima of the light curves were estimated using the 1st, 5th, and 10th percentiles. Subsequently, the minima were used to calculate the dips of the light curves $\Delta F_{p,k,j,r}$. The square root of all light curve dips $\Delta F_{p,k,j,r}$ belonging to the same planet p (i.e. including all wavelengths j and all stellar spot instances k), and additionally the stellar and planetary parameters $z_{p,1}, \dots, z_{p,6}$, were then gathered in the feature vector \mathbf{f}_p^* . The feature vector was z-score normalised (not shown in the graphic). Eventually, linear regressions were used to calculate the relative planet radius for each wavelength j .

We re-indexed the examples \mathbf{X}_i , each of size 300×55 , in a new matrix $\mathbf{X}_{p,k,l}$, where $p \in \{1, 2, \dots, 2097\}$ indexes the planet, $k \in \{1, 2, \dots, 10\}$ the stellar spot instance, and $l \in \{1, 2, \dots, 10\}$ the photon noise instance. To reduce the photon

²¹Solution by *KNOW-IWF* (Ranked 4th) available under the DOI 10.5281/zenodo.3981141: <https://doi.org/10.5281/zenodo.3981141>.

noise, we averaged the examples $\mathbf{X}_{p,k,l}$ over the photon noise instances l belonging to the same planet p and stellar spot noise instance k , yielding the noise-reduced example matrix $\tilde{\mathbf{X}}_{p,k} = \frac{1}{10} \sum_{l=1}^{10} \mathbf{X}_{p,k,l}$. $\tilde{\mathbf{X}}_{p,k}$ was of size 300×55 and comprised of the light curves for each wavelength. Subsequently, we calculated the differences between the maxima and minima of each light curve in $\tilde{\mathbf{X}}_{p,k}$. The maxima were assumed to be 1 as the light curves were already normalised, and the minima were estimated as the 1st, 5th, and 10th percentiles. This yielded estimates $\Delta F_{p,k,j,r}$ of the dip of the relative light curve caused by a transit of planet p for stellar spot noise instance k , wavelength j , and $r \in \{1\%, 5\%, 10\%\}$ corresponding to the 1st, 5th, and 10th percentiles. As the true dip $\Delta F_{p,j}$ of the relative light curve is approximately equal to the quadratic relative planet radius $\left(\frac{R_{p,j}}{R_{*,j}}\right)^2$, we took the square root of $\Delta F_{p,k,j,r}$ to obtain estimates of the relative planet radii:

$$\frac{R_{p,j}}{R_{*,j}} \approx \sqrt{\Delta F_{p,k,j,r}}.$$

We then built a feature vector \mathbf{f}_p comprised of the estimated relative planet radii belonging to planet p :

$$\mathbf{f}_p = \left[\sqrt{\Delta F_{p,1,1,1\%}}, \dots, \sqrt{\Delta F_{p,k,j,r}}, \dots, \sqrt{\Delta F_{p,10,55,10\%}} \right].$$

The feature vectors \mathbf{f}_p were augmented by the stellar and planetary parameters provided. For that, we averaged the 6 stellar and planetary parameters $z_{i1}, z_{i2}, \dots, z_{i6}$ over all photon noise and stellar spot noise instances belonging to the same planet yielding $\mathbf{z}_p = [z_{p,1}, z_{p,2}, \dots, z_{p,6}]$. The averaged stellar and planetary parameters \mathbf{z}_p were then appended to the feature vectors \mathbf{f}_p yielding the augmented feature vectors \mathbf{f}_p^* . The length of \mathbf{f}_p^* was 1656, which resulted from 55 wavelengths, 3 percentile-based dip estimations, 10 spot noise instances, and 6 stellar and planetary features ($55 \times 3 \times 10 + 6$). Strictly speaking, the averaging was not necessary as the stellar and planetary parameters were the same for all instances of a planet (i.e. no noise was added to the stellar and planetary parameters). Finally, the extended feature

vectors \mathbf{f}_p^* were z-score normalised²², separately for the training and test set, thus avoiding information leakage from the test set into the training set.

We also re-indexed the scalar targets $y_{i,j}$ in the training set as $y_{p,k,l,j}$. Subsequently, we aggregated targets by averaging over all stellar spot noise instances k and photon noise instances l belonging to the same planet p , yielding the targets $y_{p,j}$. However, averaging was again not strictly necessary as all photon noise and stellar spot noise instances of a planet had the same relative radius in the provided dataset.

Model & Training

We set up a multiple linear regression model per wavelength j , resulting in 55 regression models:

$$y_{p,j} = \beta_{0,j} + \mathbf{f}_p^{*\top} \beta_j + \varepsilon_j,$$

with β_j being the parameter vector of the model for wavelength j , $\beta_{0,j}$ the intercept term, and ε_j the error term.

The parameters $\beta_{0,j}$ and β_j of the regression model were determined using least-squares estimation, which requires the estimation of the covariance matrix of \mathbf{f}_p^* . Because of the relatively large size of \mathbf{f}_p^* , we estimated the covariance matrix with the shrinkage method from [Ledoit and Wolf \(2004\)](#), which computes the shrinkage coefficient explicitly. The parameters were found using all examples from the training set. Following this, we used the regression models to predict all relative radii of the planets p in the test set with wavelength j :

$$\hat{y}_{p,j} = \beta_{0,j} + \mathbf{f}_p^{*\top} \beta_j.$$

The predicted relative radii $\hat{y}_{p,j}$ were re-indexed to the original indices $\hat{y}_{i,j}$ by copying $\hat{y}_{p,j}$ to all corresponding stellar spot noise instances and photon noise instances.

The only hyperparameters in our model were the percentiles used for estimating the minima of the light curve dips. We found these parameters by trial and error and refrained from fine-tuning them further.

²²This type of normalisation, also known as ‘standardisation’ is performed by subtracting for each feature of a given example the mean value of that feature across all examples and dividing by its standard deviation.

B.2.3.5 TU Dortmund University

The team from TU Dortmund University, consisting of researchers working on applying machine learning algorithms in astroparticle physics, landed the 5th place on the leaderboard, going under the alias ‘Basel321’ during the Challenge. Their implementation is publicly available²³. It embraces three central ideas: i) the preprocessing simplifies the input time series, yet retains much of their information in auxiliary features; ii) the baseline architecture is largely retained, but consists of 2 input branches: one using as input these auxiliary features and the other using as inputs the stellar and planetary parameters; and iii) a bagging ensemble is created, in which each member is trained on data that have undergone slightly altered preprocessing.

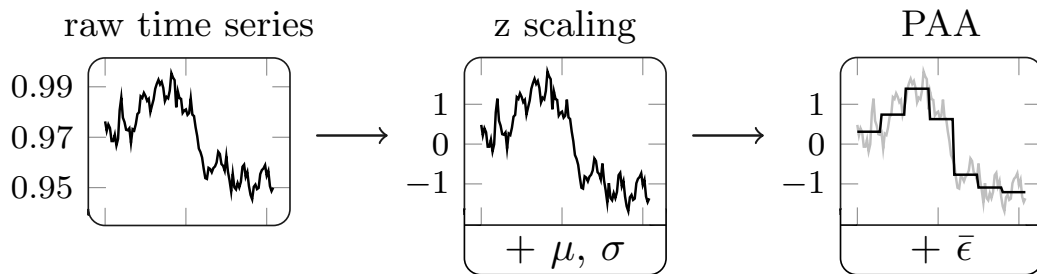


Figure B.5: The TU Dortmund University team (Ranked 5th) simplifies the raw data with piecewise aggregate approximations (PAA) of z-scaled time series. The information lost during these transformations is retained in auxiliary features. Namely, the z-scaling produces time series with zero mean and unit variance, but the original means $\mu \in \mathbb{R}$ and variances $\sigma \in \mathbb{R}$ of each channel and observation are kept. The PAA consists of only one average value in each equal segment, but the overall reconstruction errors $\bar{\epsilon} \in \mathbb{R}$ are maintained.

Preprocessing

Figure B.5 shows how the input data are simplified by the use of z-scaled piecewise aggregate approximations (PAA) Keogh and Pazzani (2000), of which the lost information is retained in the auxiliary features μ , σ and $\bar{\epsilon}$. These features describe each time series on a global level, while the PAA output captures the local shape. Namely, the PAA output is simply the average flux value in each of n_{paa} equal-sized segments. The z-scaled PAA representation facilitates learning due to the decreased number of dimensions and due to the uniform scale in each dimension.

²³Solution by TU Dortmund University (Ranked 5th): <https://bitbucket.org/zagazao/ecml-discovery-challenge>

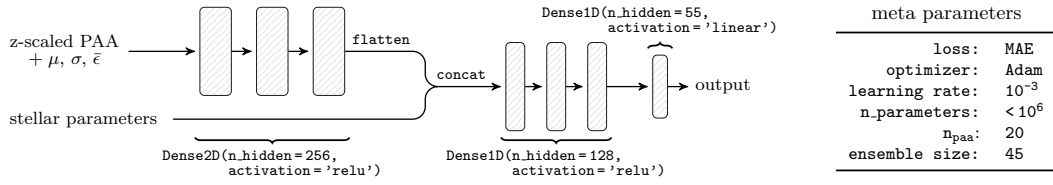


Figure B.6: The first three layers of model used by the TU Dortmund University team (Ranked 5th) derive abstract features from each time series that is represented by a PAA. The auxiliary features μ , σ and $\bar{\epsilon}$ and the stellar parameters are also fed into the network. The last four layers combine these different kinds of inputs. A randomized parameter search has been employed to tune the number of layers and their size.

These properties are particularly relevant in dense neural networks like the baseline solution, which can suffer from numerous model parameters if the input dimension is large.

Model architecture and training

A fully connected ANN is trained on the extracted features and the planetary and stellar parameters. The architecture used, shown in Figure B.6, is similar to the baseline, but it includes one branch for the auxiliary features and one for the planetary and stellar parameters. Figure B.6 also lists the associated hyperparameters.

Multiple instances of the above architecture, were then combined in a bagging ensemble. To increase the diversity, each ensemble member shifted its input by a different number $n \in [0, n_{paa})$ of time steps. This alteration is performed already before the preprocessing, so that each ensemble member uses different PAA segments. The final prediction was the median among all ensemble members' predictions.

Observations

Regarding the feature representation extracted in the preprocessing step, we observed the following: i) a linear regression on the z-scaled PAA representation is already able to outperform the baseline solution; ii) it is critical to maintain the information lost during this type of preprocessing—this is achieved by the auxiliary features; and iii) the use of shifting segments has remedied the fact that one set of PAA segments may not be optimal for all observations.

B.2.4 What the winning models teach us

We should stress again that the final score differences among the top-5 ranked solutions, as shown on Table B.1, are statistically negligible and should thus be regarded as equivalent in terms of predictive power in our simulated data. Having clarified this, these solutions provide us with some interesting insights regarding the problem.

First, we observe that all 5 solutions make use of the additional stellar and planetary parameters (orbital period, stellar temperature, stellar surface gravity, stellar radius, stellar mass & stellar K magnitude). This shows that these features indeed contain relevant information for uncovering the transit depths in light curves contaminated by the presence of stellar spots. Moreover, this information is not redundant given the noisy light curves.

Another interesting observation is that most solutions involve the use of highly non-linear non-parametric or over-parameterised²⁴ models w.r.t. the original features, like ANNs and/or ensembles of learners. More specifically, 4 out of 5 teams use deep learning approaches (*SpaceMeerkat*, *Major Tom*, *BV Labs* & *TU Dortmund University*) and 3 out of 5 (*Major Tom*, *BV Labs* & *TU Dortmund University* teams) use ensemble learning methods. The *Major Tom* team does not apply any preprocessing of the data provided beyond feature normalisation, leaving all feature extraction to be implicitly performed by the ANN, using appropriate regularisation techniques (batch normalisation & dropout) to prevent overfitting.

In contrast to this, the *IWF-KNOW* team relied on the extraction of non-linear features from the original inputs informed by domain knowledge. They then trained simple linear models in this new feature space.

The above are indicative of the non-linear nature of the problem. They also showcase the flexibility of machine learning and computational statistics methods in

²⁴The term ‘non-parametric’ applies to models that are not restricted to a predetermined number of parameters. They can therefore adjust their complexity to the data at hand. Ensemble models can fall in this class. The term ‘over-parameterised’ refers to parametric models having a number of learnable parameters that exceeds the number of data points. ANNs can fall in this class. Through appropriate use of regularisation methods it is possible to avoid overfitting even when fitting models of such high complexity.

building models that capture this nonlinearity. One can extract informative features given domain knowledge to capture it and then use simple and explainable models like linear regression trained on them. Alternatively, one can simply use powerful overparameterised models, like ANNs and ensemble methods to implicitly learn transformations of the original feature space that are useful for the purposes of predicting the transit depth.

Extracting a few meaningful features informed by domain knowledge (*IWF-KNOW & BV Labs*) or appropriately summarising the light curve information using signal processing techniques (*SpaceMeerkat & TU Dortmund University*) allows for simpler models to be trained in the lower-dimensional extracted feature space. This allows for faster training and can also ultimately reduce overfitting.

A more detailed look into how the 5 solutions control for overfitting also reveals they follow quite different approaches. *SpaceMeerkat* uses a CNN rather than a fully connected ANN to reduce the number of effective learnable parameters. *Major Tom* uses a fully connected ANN but controls for its complexity via batch normalisation, dropout and the use of an ensemble of trained ANNs, rather than a single model. *BV Labs* also make extensive use of ensembling and their neural network learner also uses batch normalisation and dropout. The fact that they operate on a much lower dimensional feature space (only 171 features per data point) also aids in reducing overfitting. *IWF-KNOW* use linear regression models, which are characterised with high bias (i.e. more prone to underfitting than overfitting). They also operate on a lower-dimensional space (1656 features per data point) and apply shrinkage. Last but not least, *TU Dortmund University* makes use of an ensemble which is interestingly built on data having undergone slightly different preprocessing. Training on perturbed inputs results in making them more robust to overfitting.

Two of the top-5 teams (*BV Labs & IWF-KNOW*) made use of the fact that the training data contained multiple data points corresponding to the same planet under (10 different photon noise and 10 different stellar spot noise instances). They treated the two noise sources as independent and averaged these out or took the median to obtain less noisy light curves. This was a sensible thing to do and such a scenario

would indeed occur if multiple observations of the same target were to be obtained.

Finally, ignoring outlier flux values via smoothing/downsampling the light curves (*SpaceMeerkat*), clipping values above 1 (*SpaceMeerkat* & *BVLabs*) or by extracting summary statistics from the light curve and using them as features (*SpaceMeerkat*, *BVLabs*, *IWF-KNOW* & *TU Dortmund University*) proved a useful strategy in building more robust models.

B.2.5 Conclusions

Correcting transit light curves for the effects of stellar spots is a challenging problem, progress in which can have a high impact on exoplanetary science and exoplanet atmosphere characterisation in particular.

The primary goal of the *Ariel Mission's 1st Machine Learning Challenge* was to investigate the existence of fully automated solutions to this task that predict the transit depth with a precision of the order of 10^{-5} with the use of machine learning and computational statistics methodologies. The secondary goal was to bridge the machine learning and exoplanetary science communities. As we saw, both of these goals were met with success.

The aim of this work is to serve as a starting point for further interaction between the two communities. We described the data generation, the problem outline and the organisational aspects of the Challenge. We intend this to serve as a reference for the organisation of future challenges in data analysis for exoplanetary science. In the interests of communicating the modelling outcomes of the Challenge, we also presented, analysed and compared the top-5 ranked solutions submitted by the participants.

As evidenced by the top-5 entries, the Challenge indeed attracted the interest of both exoplanetary scientists and machine learning experts. The participants cover an impressive breadth of academic backgrounds and the submitted solutions an equally impressive range of approaches, from linear regression to convolutional neural networks.

The solutions obtained demonstrate that it is indeed feasible to fully automate the process of efficiently correcting light curves for the effect of stellar spots to the

desired precision. One key insight obtained is that additional stellar and planetary parameters (orbital period, stellar temperature, stellar surface gravity, stellar radius, stellar mass & stellar K magnitude) can greatly improve the derivation of correct transit depths in the face of stellar spots.

Good solutions can be obtained by a wide range of modelling methodologies. They include simple, easily interpretable models, like linear regression, built on features derived from clever feature engineering, informed by exoplanet science theory. Other solutions amount to training complex machine learning models using deep learning or ensemble learning, which automate the extraction of useful features from minimally preprocessed—even raw—data. In the latter case, especially for ANN models it is crucial to take measures to prevent overfitting. These can include dimensionality reduction, ensembling, use of convolutional filters, batch normalisation, dropout, training using perturbed data and combinations thereof.

The next steps of this work include refinement of the proposed solutions to handle more realistic simulated data, possibly involving both stellar spots and faculae (areas of the host star characterised by increased temperature). Upon successful performance on these, the provided solutions can then be used in the analysis pipeline of Ariel data or adapted to other instruments.

B.3 Ariel Machine Learning Challenge 2021

For the 2021 edition of the Ariel ML challenge, we built on the first edition, included realistic end-to-end simulations of exoplanet atmosphere as well as stellar noise and instrument systematics. This is a significantly more challenging task as the non-linear science signal and stellar noise are convolved with the non-linear and time-varying responses of the instrument.

The competition presented used data generated by ArielSim, the simulator of the European Space Agency’s upcoming Ariel mission, whose objective is to characterise the atmosphere of 1000 exoplanets. The data consist of pairs of light curves corrupted by stellar spots and instrument systematics and the corresponding clean ones, along with auxiliary observation information. The goal is to correct the

light curve for the presence of these distortions (signal denoising) and predict the correct transit depths. This is a yet unsolved problem in the community and the main hurdle to analysing smaller and more Earth-like planets (i.e. fainter signals). Solving helps improve our understanding of the characteristics of currently confirmed exoplanets, and pushing our ability to characterise the atmospheres of extrasolar planets to unprecedented precision.

B.3.1 Data

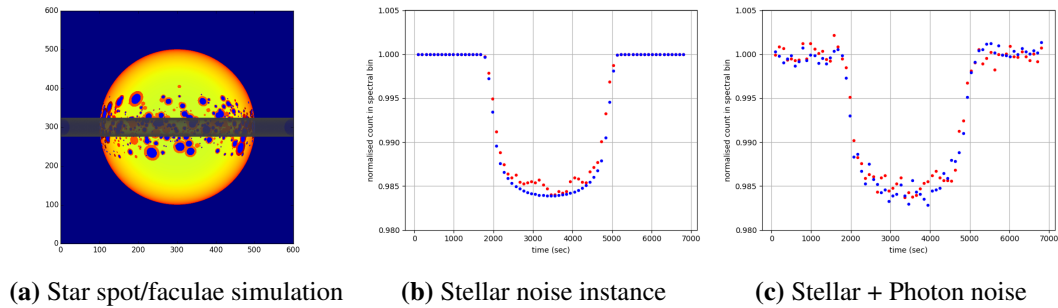


Figure B.7: Stellar noise simulations. A) 3D simulation of a with star spot and faculae activity. The grey strip indicates the transit cord of the exoplanet. B) Light curve with spot crossing noise only (red) compared to a noise-free model (blue). C) same as B) but, adding the expected photon (Poisson) noise. See figure B.8 for an instrumental noise instance example.

The Ariel space mission will observe the planet-star system with 106 spectral channels—i.e. producing 106 observed time series, also known as light curves per observation. The spectroscopic wavelengths range from 0.5 to 7.8 μm . Figure B.7 shows typically observed time series for one of the 106 channels assuming photon noise only. Figure B.8 shows an example of instrument distortions typically observed with space-based infra-red instruments. We will simulate stellar noise effects and convolve them with realistic instrument simulations to produce highly realistic simulations of the expected Ariel space mission data.

The provided **training data** will consist of the following information:

- **Corrupted light curves** (noisy signals, Figure B.8, top plot as an example): light curves at different wavelengths of planets transiting stars with stellar spots / faculae, under various spot / faculae distributions and levels of additional stellar photon noise added to the signal. We will simulate full instrument

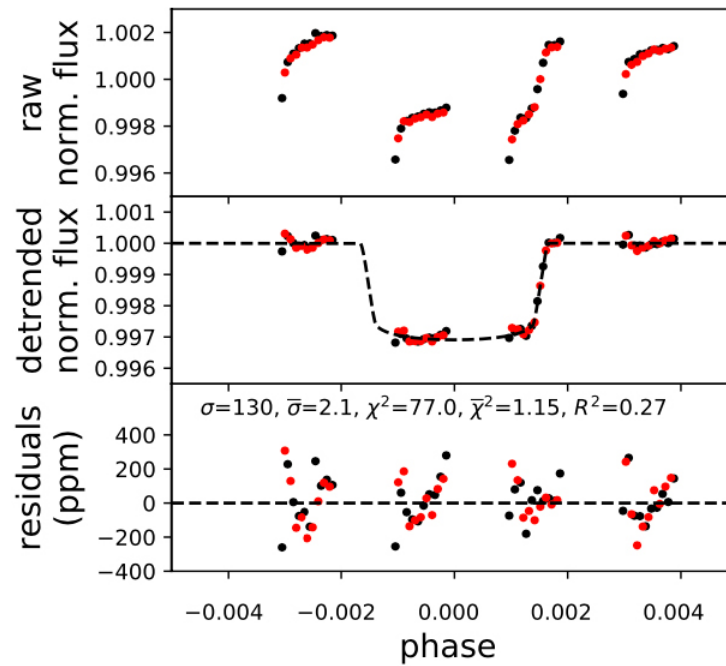


Figure B.8: Example of typical instrumental noise in exoplanet observations from [Tsiaras et al. \(2019\)](#). Top: Observations obtained with the Hubble Space Telescope. Hook-like distortions are typical artefacts seen in infra-red detectors. We will simulate these instrument responses for the Ariel data in this competition. Middle: Parametrically detrended data following an idealised transit model (dotted line). Bottom: Residuals. We can still see very significant auto-correlative noise in the residuals, setting a limit to what can be achieved with current methods. This Discovery Challenge addresses this issue to (hopefully) significantly decrease residual non-Gaussian noise.

systematics (e.g. time-dependent trends, inhomogeneous pixel quantum efficiency, optical distortion effects) and convolve the planetary and stellar signals with the instrument response.

- **Clean light curves** (denoised signals, Figure B.7, middle plots, blue curve): corresponding signals without stellar, instrument or photon noise.
- **Auxiliary observation parameters:** These include astrophysical information: the planet radius, the star radius, the star temperature, the orbital semi-major axis and the orbital inclination; and instrumental information: position of spectrum on detector, space-craft jitter, detector dark-current, temperature fluctuations.

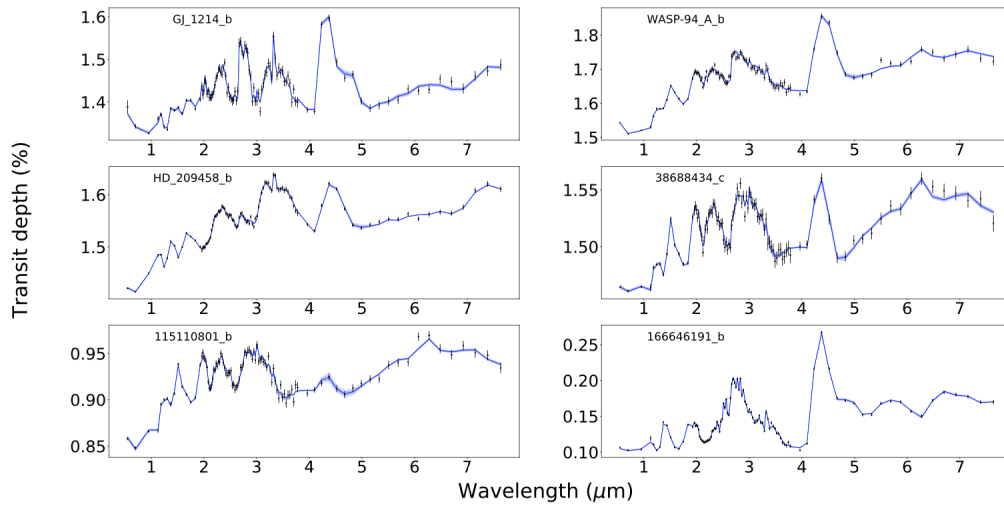


Figure B.9: Example of atmospheric transmission spectra of simulated exoplanets for the Ariel Space Mission target reference sample (Changeat et al., 2020). Each planet has a unique chemistry which is reflected in characteristic ‘bumps and troughs’ in the observed spectrum. Here, each spectral point corresponds to an individual time series measurement (e.g. Figure B.7). Distortions in the observed time series lead to uncertainties and biases in the final atmospheric spectra and hinder our understanding of these foreign worlds.

The **test data** will contain the same information except for the clean light curves (denoised signals) which will be used as the ground truth.

More specifically, we aim to provide a dataset comprising **around 10^6 data points (70% training, 30% test)**. The **training examples will consist of two ascii files, one for the corrupted light curves and one for the corresponding clean ones**. Both files have the following format:

- **Header line:**

The header line contains the auxiliary observation parameters and its format is:

#planet radius, star radius, star temperature, orbital semi-major axis, orbital inclination, normalised noise variance per wavelength

- **light curves::**

A 1200×106 table of white space separated real values. The 106 **columns** of the table correspond to the different wavelengths (i.e. detector channels),

the 1200 **rows** to time-steps (orbital phase from mid-transit) and the values themselves to *relative flux*²⁵ (intensity of incoming light).

Each column is therefore a separate light curve (at wavelength λ), i.e. a plot of the relative flux vs. time. As for the units of measurement, planet radius is measured in Jupiter radii, star radius in Solar radii, star temperature in Kelvin, orbital semi-major axis in astronomical units (au), orbital inclination in degrees, wavelengths in micrometers and time in seconds.

Test examples will consist of **just the corrupted light curve ascii file**.

All the data used in the competition was generated by the organisers, who are leading the Ariel mission’s data analysis efforts, using open data and open or custom-made tools, so no permissions are needed to generate the dataset. We were therefore able to easily run many simulations to generate a large, diverse dataset to make the competition interesting and draw conclusive results. With the complexity of the problem, both positive results and negative results are of interest to us.

On launch date (1st of April 2021) all the generated training data (corrupted light curves, clean light curves & auxiliary observation parameters) and all test data without the ground truth (i.e. without the clean light curves) were made freely available on the competition’s website. We set up a web portal based on the website of the first edition with an improved back-end performance to accelerate test-data upload and back-end validation. After the competition ends (1st of July 2021) the ground truth (clean light curves for test data) were also made openly available on the competition’s website. This helped keep the ground truth confidential until the end of the competition, while ensuring transparency of the final results.

B.3.2 Tasks and application scenarios

The learning task proposed is effectively one of multiple signals denoising. Given a set of noisy signals (light curves at different wavelengths of an exoplanet transiting, corrupted by instrument and stellar noise) along with some auxiliary information, the goal is to generate the corresponding set of noise-free signals (light curves at

²⁵This normalised flux ranges within $[0, 1]$ in all wavelengths. Typical values range from about 0.98 to 1 as the relative decrease of the incoming stellar light due to a transiting exoplanet is very small, i.e. around 0.1–2 %.

these wavelengths without the effect of instrument distortions and stellar noise). The participants were therefore expected to produce predicted clean light curves of the same format as that described in Section B.3.1 for all the test examples provided.

Current approaches have not been successfully used on multi-wavelength spectroscopic data. This will be the first attempt of detrending data of this complexity in the field of exoplanets and, as such, it will be challenging. However, in theory, a pattern for removing instrument distortions effects should be retrievable by combining information across different wavelengths.

B.3.3 Metrics

Submissions were evaluated quantitatively: for each test example $i \in \{1, N_{test}\}$ (set of light curves at different wavelengths) we need a measure that captures the difference between the predicted (denoised) light curves and the ground truth.

On each wavelength λ , we will use the normalised Root Mean Squared Error between the predicted output time series $\hat{Y}_i(\lambda)$ and the ground truth value $Y_i(\lambda)$ to measure this difference.

$$RMS_i(\lambda) = \frac{1}{\sigma_*(\lambda)} \sqrt{(Y_i(\lambda) - \hat{Y}_i(\lambda))^2} \quad (\text{B.7})$$

where $\sigma_*(\lambda)$ is the normalised stellar flux standard deviation, accounting for the varying signal-to-noise of the observation as a function of wavelength channel.

Next, we will compute the weighted average over all wavelengths to derive the overall normalised score on the i -th test example.

$$s(i) = \frac{1}{N_\lambda} \sum_{\lambda} \frac{RMS_i(\lambda)}{\delta_f(\lambda)} \quad (\text{B.8})$$

The weighting factor, $\delta_f(\lambda)$, is derived by taking into account the flux contribution of spot+faculae as function of the total stellar flux. This takes into account the decreasing spot+faculae strength as function of increasing wavelength

$$\delta_f(\lambda) = \frac{F_{spot}(\lambda) \times f_{spot} + (1 - f_{spot}) \times F_*(\lambda)}{F_*(\lambda)} \quad (\text{B.9})$$

where F_{spot} and F_* are the spot and stellar fluxes respectively (here modelled with a Planck function) and f_{spot} the spot filling factor as fraction of total stellar surface observed. Figure B.10 shows the weighting function for a sun like star. Added instrument distortions are partially wavelength dependent but are case dependent and not necessarily a homogeneously varying function of wavelength.

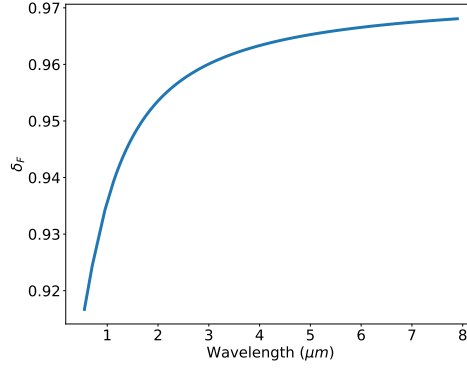


Figure B.10: Scaling factor, δ_f , as function of wavelength for a solar type star (5500 K), and 10% spot coverage.

The final performance is measured by the average normalised score across the entire test set minus the variance in normalisation scores across the test set

$$S = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} S(i) - \text{Var}(s). \quad (\text{B.10})$$

The score encourages consistently low residual systematic noise across the test set. We do not normalise this score but note that its upper limit is given by the degree of Normally distributed noise present in the test data set.

B.3.4 Baseline and code available

There are currently no solutions to the problem of automatically removing the effects of instrument and stellar distortions from the light curves of spectroscopically observed transiting exoplanets on realistic data. As a baseline, we trained a feedforward neural network on a subset of training examples selected randomly, while monitoring a validation score on some other random light curves. The neural network uses all 55 noisy light curves to predict the 55 relative radii directly. However, it does not

use the stellar parameters, nor does it use any additional physics knowledge or ML technique to do so. Before being passed into the network, the noisy light curves undergo the following preprocessing steps: i) 1 is removed to all light curve to have the asymptotic stellar flux centred around 0; ii) values are rescaled by dividing by 0.04 for the standard deviation to be closer to unity.

The PyTorch code for the baseline as well as more details and advice was made available for participants²⁶, and its evaluation set up a baseline score of 9617 for the problem.

B.3.5 Results

Results of the 2021 competition exceeded those of the same edition (given the same metrics), even though the noise was more realistic and diverse. The top-5 solutions reused ideas and models from 2019's challenge (e.g. CNNs, random forest, ensembling), building on the learnings of the first edition. However, the winning solution from ML Analytics differed in that it used extreme learning machines (ELMs) rather than more traditional deep learning approaches. ELMs are single-layer networks with random fixed weights. They are weak predictors, and can suitably be ensembled in a gradient boosting algorithm. The fast training of ELMs ($\sim ms$) enabled ML Analytics to try many ideas and perform a comprehensive hyperparameter optimisation, in what proved to be a consistent and incremental building of the final model whose occasional submissions kept on improving steadily.

²⁶publicly available on GitHub at: <https://github.com/ucl-exoplanets/ML-challenge-baseline/>

Appendix C

Quantifying Prevalence of COVID-19 from Wastewater Data

This Chapter is based on [Morvan et al. \(2022a\)](#): An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence published in Nature Communications, for which I contributed in equal proportions with Anna Lo Jacomo, Kathleen M. O'Reilly and Leon Danon. Anna Lo Jacomo was the main force behind the forward model, whereas I led the work on the regression model, Leon Danon provided high-level supervision and Kathleen O'Reilly coordinated the various collaborations and finalised the article write-up. This article is a product from the 6-months placement secondment at the UK Environmental Monitoring for Health Protection, undergone as part of UCL's CDT in DIS placement scheme requiring 6 months in a non-academic set-up.

Despite the sheer disconnect between the objects of study during my PhD work and my placement work, there are nonetheless a few similarities which might be worth noting. Both works are indeed concerned with noisy time series: in one case stellar light curves and in the other time series of wastewater measurements. These time series have various confounders, whose relation with the signal must be understood and disentangled as well as possible. Then of course, the signals and science objectives are very different in nature, partly also for the operational reason that parts of the placement work needed to be delivered quickly, communicated widely and thus naturally favoured simple models over complex ones—even if that

meant losing some performance. The particular work presented below deals with temporal data without explicitly modelling it as time series. It does so simply by considering covariate data whose correlations with the main variable of interest (concentration of SARS-CoV-2 RNA in wastewater) enable to get rid of the time variable and consider the whole dataset as tabular. It is interesting to see how well this approach has worked in this case, but it should also be caveated that for some specific tasks (e.g. forecasting) modelling the data as actual time series might be more effective.

C.1 Overview

Accurate surveillance of the COVID-19 pandemic can be weakened by under-reporting of cases, particularly due to asymptomatic or pre-symptomatic infections, resulting in bias. Quantification of SARS-CoV-2 RNA in wastewater can be used to infer infection prevalence, but uncertainty in sensitivity and considerable variability has meant that accurate measurement remains elusive. Here, we use data from 45 sewage sites in England, covering 31% of the population, and estimate SARS-CoV-2 prevalence to within 1.1% of estimates from representative prevalence surveys (with 95% confidence). Using machine learning and phenomenological models, we show that differences between sampled sites, particularly the wastewater flow rate, influence prevalence estimation and require careful interpretation. We find that SARS-CoV-2 signals in wastewater appear 4–5 days earlier in comparison to clinical testing data but are coincident with prevalence surveys suggesting that wastewater surveillance can be a leading indicator for symptomatic viral infections. Surveillance for viruses in wastewater complements and strengthens clinical surveillance, with significant implications for public health.

C.2 Introduction

Estimates of SARS-CoV-2 infection prevalence are essential to understand COVID-19 disease burden and the impact of public health interventions ([The Lancet, 2020](#); [Franceschi et al., 2021](#); [Pouwels et al., 2021](#)). The sensitivity of surveillance for COVID-19 varies for several epidemiological, administrative, political and financial

reasons, meaning that reported cases are likely to be an underestimate of actual cases (Noushad and Al-Saqqaf, 2021; Richterich, 2020; Wu et al., 2020b,a). Further, only a proportion of infections result in symptomatic disease; estimates range considerably across studies but in a recent meta-analysis that aimed to account for potential biases in reporting an average of 64.9% (95% CI 60.1–69.3%) was estimated (Sah et al., 2021), so there is a disconnect between infections that result in transmission and reported cases via clinical surveillance. To monitor the trajectory of the COVID-19 pandemic and reduce the impact of bias due to any single source of information, it is essential to have multiple measures of prevalence with well-understood sources of bias and uncertainty.

Based on its proven success applied to other infectious diseases and markers of human health and behaviour (Feng et al., 2018; Choi et al., 2018), wastewater (WW) surveillance for SARS-CoV-2 has been established in many countries, including England, since the start of the pandemic (Baraniuk, 2020; Hillary et al., 2021). Early studies indicated that fragments of RNA corresponding to the SARS-CoV-2 viral genome were detectable in WW (Medema et al., 2020), and in quantities that some quantitative measure could be established. However, laboratory protocols required refinement to establish a method that provides consistent measures of RNA with sufficient sensitivity and to be used at scale. Quantitative reverse-transcriptase PCR (RT-qPCR) of target genes of the virus genome is now routinely performed on concentrated samples from wastewater.

A challenge inherent to WW surveillance is the potential impact of environmental and biochemical attributes on the detection and quantification of the virus concentration. In England and other countries that utilise combined sewer networks to transport sewage, the wastewater inflowing at the sewage treatment works (STWs) typically comprises a combination of raw sewage, household effluent (e.g. from washing and cleaning), agricultural run-off, rainwater/snow melt, and trade waste from industry (Boogaerts et al., 2021). The percentage volume of human-derived excreta likely to contain virus RNA (i.e., urine, faeces, nasal discharge, sputum, blood) (Tran et al., 2021) in the collected wastewater sample is likely to vary because

of additional inflow detailed above, which will dilute and may degrade the target analyte concentration, reducing the sensitivity of lab assays (Wilder et al., 2021). In order to overcome these challenges and infer an estimate of prevalence from WW samples, additional data and statistical models can be used, and validation of model outputs using reliable estimates of prevalence is critical. However, the associated biases in clinical surveillance and the impact of uncertainties associated with environmental monitoring of viruses in sewer networks present significant challenges when considering prevalence estimation using WW measurements.

Back-calculating or estimating the quantity of chemical compounds (e.g., licit and illicit pharmaceuticals) or stressors (e.g., pathogens) by targeting indicative analytes present in WW is a common feature of wastewater-based epidemiology (WBE). For example, WBE has been applied successfully in estimating illegal drug consumption (Feng et al., 2018), the degree of antibiotic resistance in a population (Hutinel et al., 2019), among other applications. While most studies have used WW to track disease trends (i.e. increase/decrease), a number of studies have attempted to directly quantify prevalence from SARS-CoV-2 measurements, along with biological and hydrological parameters (Wu et al., 2020a; Arora et al., 2020; Ahmed et al., 2020). Broadly, studies using back-calculation for SARS-CoV-2 generally consider that disease prevalence is equal to the load of RNA in the sample, divided by the load of RNA produced by one infected person (Ahmed et al., 2020). The underlying assumptions are that viral RNA is released proportionally to wastewater and perfectly mixed in the sewers, and that there are no significant losses of virus RNA in the network that lead to a decrease in measurement representativeness. Variations of this hypothesis have been suggested to account for additional "signal loss" factors, for example decay, flow dilution, and temporal shedding patterns in the population (Ahmed et al., 2020).

The Office for National Statistics (ONS) Covid-19 Infection Survey (CIS) was established in the UK early in the pandemic to assess the prevalence of individuals in the community testing positive for SARS-CoV-2 (otherwise known as "positivity") through nasopharyngeal sampling of individuals living in randomly selected private

households from the UK3. This survey has been essential to understand the dynamics of SARS-CoV-2 by estimating community positivity rates, and further to estimate these rates at regional and subregional scales. The wide availability of WW samples from July 2020 in England and subregional positivity estimates from the CIS provides a unique dataset to investigate and validate WW as a reliable estimate of prevalence to support public health actions. In this study, we estimate the prevalence of SARS-CoV-2 infection in the community and establish what additional data and analyses are required to have accurate and robust estimates of prevalence from WW.

C.3 Results

We analysed data collected between July 2020 and March 2021 from 45 sewage treatment works (STWs) across England (Fig. 1) covering an estimated 31% of the population. For each site, an average of four samples were collected per week, by either grab (46%) or composite (54%) sampling. Additional metadata were collected on inorganics and other wastewater characteristics.

A Map of Coronavirus Infection Survey (CIS) regions (outlined in blue) and the locations of wastewater (WW) catchments used in this study (in red). B Regional 7-day rolling averages (median) of CIS prevalence estimates (black) with 95% credible intervals using Bayesian modelling (grey regions), with corresponding predictions of prevalence from WW data only (blue) with 95% confidence interval from bootstrapping (blue vertical lines), and raw SARS-CoV-2 concentrations (yellow, right axis). The WW prevalence estimates are provided at a subregional level and combined to produce regional estimates for comparison.

Translation of raw WW data to prevalence estimates are illustrated using a phenomenological model that considers infection prevalence, shedding and stool generation, and the volume of water in the sewage column (see the “Methods” section). The assumptions of the model results in a linear relationship between prevalence and RNA concentrations. Sensitivity analysis illustrated that viral concentration in stool is the largest source of uncertainty in this approach (Fig. S2). Using average values of the shedding rate from clinical studies ([Hoffmann and Alsing](#),

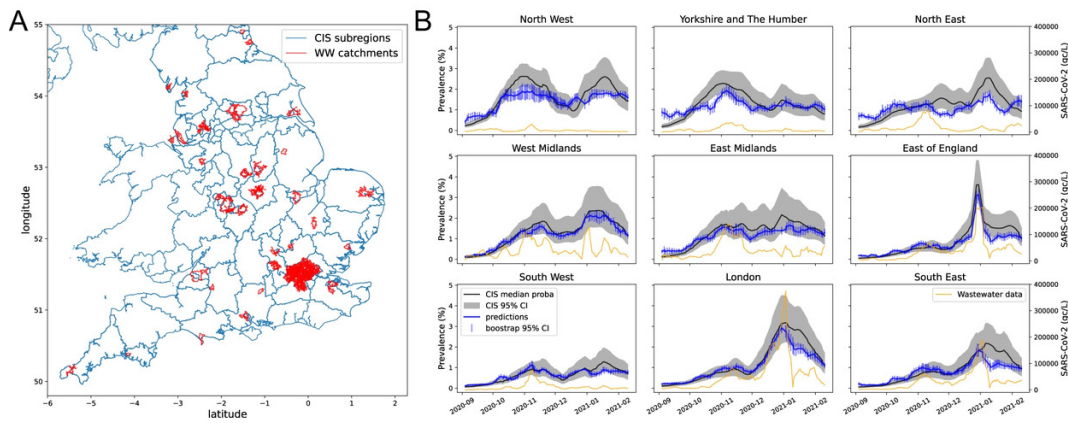


Figure C.1: Geographical summary of the data used to estimate SARS-CoV-2 from wastewater.

2021) gives a relatively good fit with observations from wastewater and CIS data in terms of average magnitude, but with high variability across individual samples (Fig. 2A). However, that variability is commensurate with the uncertainty in the appropriate hydrological and biological values used for the calculation. Comparing these model estimates to data indicates that for more than half of the CIS subregions in the study (60%) the model assumptions illustrate a valid relationship with the data (more than 60% of sample points fall within the 50% confidence interval of the model). Sites showing a poorer fit, have either relatively low (28%) or relatively high (12%) concentrations per positivity rate (Fig. 2B and C). Lower than expected concentrations could be caused by unusually high per capita flow rates (such as groundwater infiltration), or degradation of RNA during transit due to physical or chemical characteristics of the network (such as numerous pumping stations, or consistently atypical pH). The method of sample collection, together with limited homogenisation of the "sewage parcel", could also lead to unrepresentative (either low or high) concentrations, or indeed unaccounted sewage discharge could also affect measurement (Ort et al., 2010). Including an additional factor to account for degradation might provide a better model assumption for sites showing relatively low concentrations (e.g. subregion B in Fig. 2A). In some subregions, the relationship between concentrations and prevalence is not well explained by the (static) linear

model. A possible reason may be interactions between disease dynamics and shedding where infection in a growing epidemic appears to have increased viral load (Hay et al., 2021); inclusion of this in the model would require a non-linear model.

An example fit between the phenomenological model estimates (green region) and the wastewater and prevalence data from three CIS subregions (blue dots), selected to illustrate three cases: subregion (A) (good correspondence), subregion (B) (concentrations tend to be low), and subregion (C) (concentrations tend to be high). Model estimates of prevalence from WW data are in the same order of magnitude and follow the shape of the relationship between concentrations and prevalence using distributions of likely parameter values, but confidence intervals are wide. The combined uncertainty in parameter values exceeds the variability seen in the data. B The percentage of data points within each subregion that fall within the 50% credible interval of the phenomenological model. C The median concentrations per positivity rate. Only CIS subregions that overlap with the original 44 wastewater catchment sites are shown. Sites with a poor fit to the model (yellow in sub-plot B) show either relatively low (dark blue) or relatively high (dark red) concentrations in sub-plot (C). Sites with a good fit to the model (dark green) tend to show intermediate concentrations (white).

Combining WW data, site-level and sample-level variables within a statistical modelling framework (exemplified using a gradient boosted regression tree (GBRT) model, see the “Methods” section) to estimate prevalence of SARS-CoV-2 provides reliable metrics across regions and throughout the evolving epidemiology of the COVID-19 pandemic in England. Using this model, SARS-CoV-2 prevalence was tracked within 1.1% (with 95% confidence) from the CIS (Fig. 1B). When the GBRT model with covariates was aggregated to regional level, an average mean absolute error (MAE) was obtained, with the West Midlands performing above average and the North East performing below average (MAE of 0.12 and 0.19, respectively) (Fig. 3B). We focus the results of the modelling to a regional level here, but have carried out the analysis at a subregional level to inform the public health response at these smaller levels of aggregation (Fig. S6).

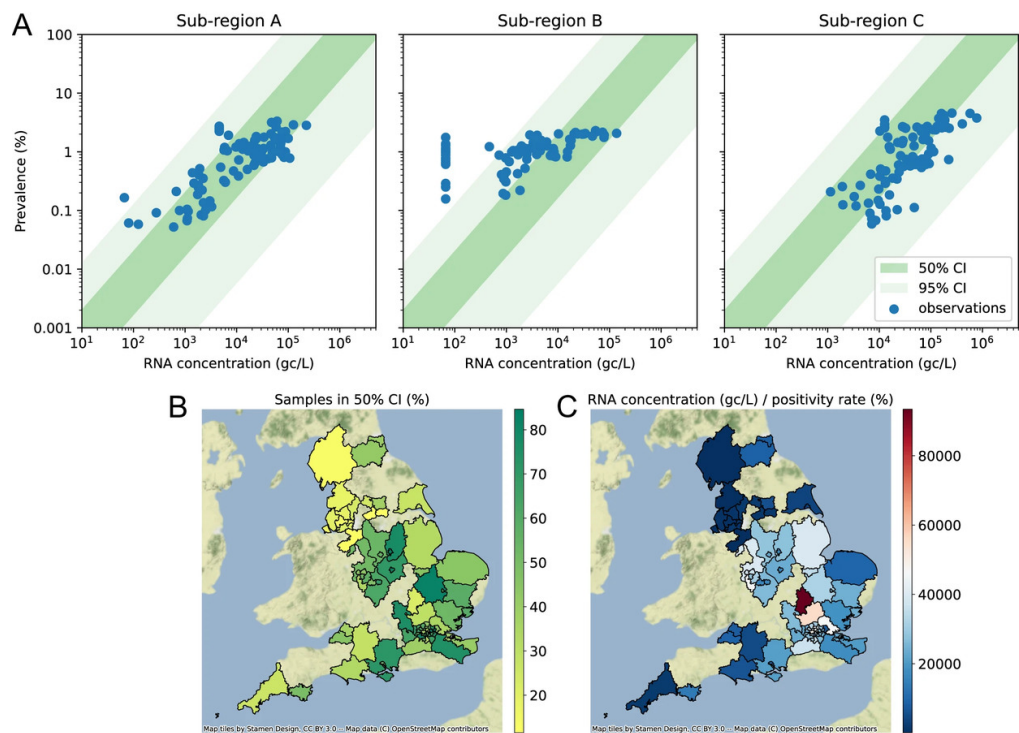


Figure C.2: Comparison of the outputs from the phenomenological model to CIS prevalence estimates.

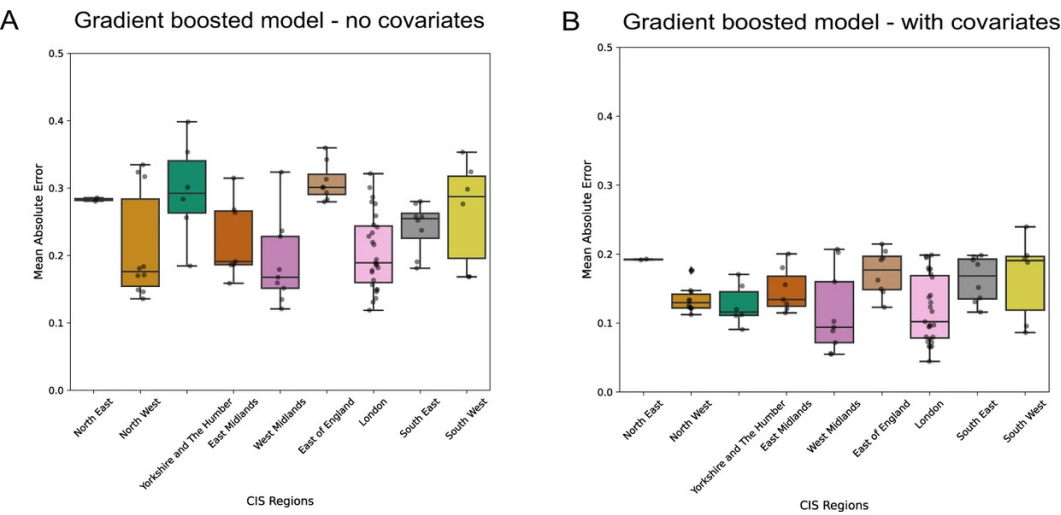


Figure C.3: Gradient boosted regression tree (GBRT) model performance across England regions.

The GBRT model was found to be the best of the candidate models that were developed to interrogate the data and identify what variables, in addition to the raw RNA concentrations, would provide accurate estimates of prevalence. Different candidate models (linear, linear with random effects and GBRT), were evaluated and compared using the MAE between predictions and median positivity rates estimates from CIS. The addition of temporally varying data such as ammonia concentration, the fraction of samples below the limit of detection and quantification and site-specific data such as population coverage greatly improved the overall fit where GBRT the average MAE per CIS subregion reduced in value when compared with a model trained on SARS-COV-2 concentrations alone (Fig. 3A and B). Further collation of additional site-level characteristics through consultation with water companies and characterisation of the catchment area showed an additional reduction in bias in the model's residuals distribution against these characteristics (Fig. S7) highlighting the robustness of our final model to wastewater network differences. While the GBRT model will be applied to estimate prevalence in England, the relative contributions of each variable and partial dependency plots (Fig. S8) are used to illustrate the direction of their effects and provide guidance for use outside this application. However, exploration of the site-specific random effects (within the random effects model) illustrated that there was considerable variability in MAE within sites that had yet to be fully accounted for. These WW data were collected in England across a time period where the prevalence of infection has varied considerably as a result of epidemic emergence and suppression through non-pharmaceutical interventions. The statistical modelling presented here illustrates that prevalence estimates are accurate and precise across a wide range of prevalence values (Fig. 4). The prevalence is tracked within 1.1% (with 95% confidence) for the GBRT model and is more precise at higher values of prevalence. Comparison between the random effects and GBRT model illustrates reduced precision and over estimation of prevalence at lower values of prevalence for the random effects model.

A lead and lag analysis was performed using the regression models on CIS estimates. Sampling dates for WW were shifted between -10 and 20 days with daily

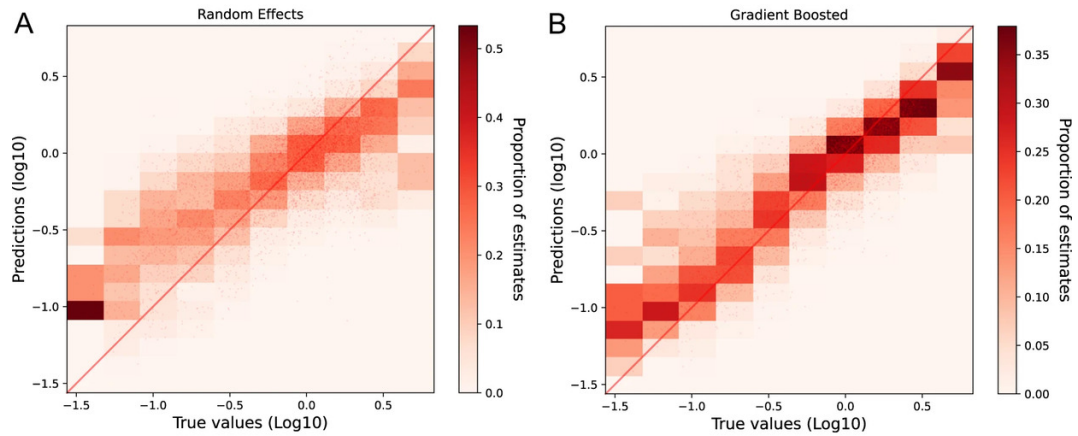


Figure C.4: Conditional predictions of SARS-CoV-2 prevalence from WW compared to the CIS positivity estimates (“True value”) in the log10 space.

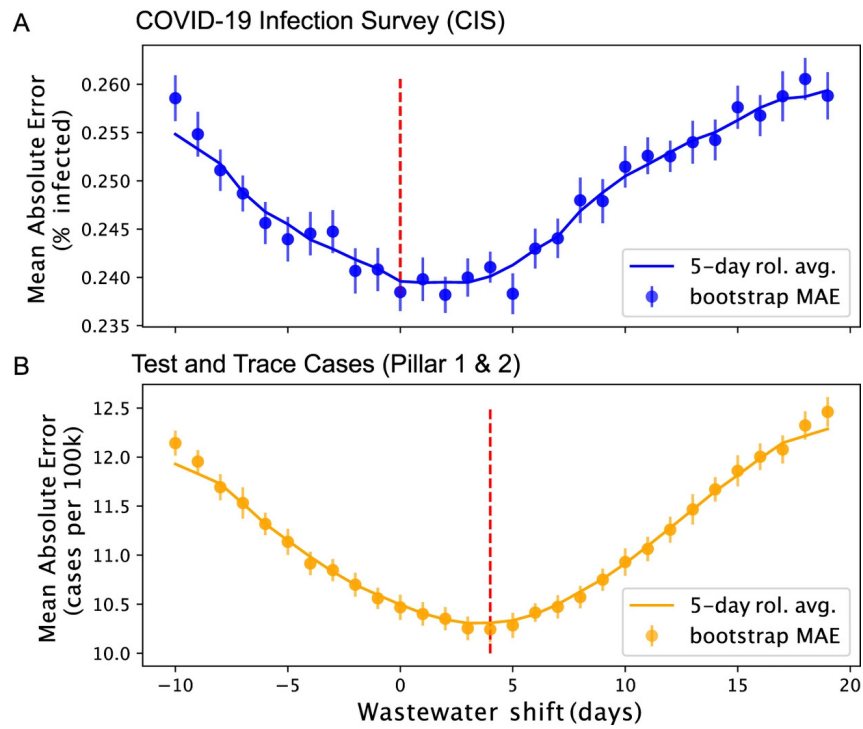


Figure C.5: Lead and lag analysis of the WW data when compared to (A) CIS and (B) Test and Trace cases.

increments, while training a model at each step to predict the CIS positivity rates in outputs, whose dates had been fixed. At each step, the models were evaluated using the bootstrapped MAE, producing a curve of prediction errors as a function of the wastewater lag (Fig. 5). Results show a minimum value of the smoothed error curve between 0 and 2 shifted days, indicating no clear advantage to predict CIS backward or forward in time from WW data. For comparison, this analysis was replicated on Pillar 1&2 data from Test and Trace. In this case the regression outputs were the case rates reported by Test and Trace until May 17, 2021, smoothed with a 7-day centred window to remove weekly periodicity and preserve consistency of reporting dates. In addition, the WW dataset was stripped to contain samples only up to 20 days before May 17, 2021 to ensure the stability of dataset sizes during the analysis. In this case the MAE is minimal between +3 and +5 days shift, suggesting an approximate 4-day lead of WW surveillance date over reported Test and Trace cases (Fig. 5B).

The shift (in days) associated with the minimal error is indicated by the red dotted line. A minimal error reached for a positive number of WW shifted days can be interpreted as a lead from wastewater by as many days. No clear lead of WW over CIS has been observed in this analysis, but an approximate 4 days lead over T&T has been observed. Vertical bars indicate the 95% confidence intervals of the mean absolute error across regions.

C.4 Discussion

We have shown that concentrations of SARS-CoV-2 RNA collected from wastewater in 45 sites in England, combined with essential related variables can provide reliable estimates of prevalence of SARS-CoV-2 infection within a population. Site-specific characteristics including limit of detection, dilution, network characteristics and other unexplained data, mean that a mechanistic model alone fails to capture the full variability in the data. However, by using locally explicit information and hierarchical models, we can understand these differences and account for them in a data-driven manner. We used our best fitting model to observe a 4-day lead in WW concentration over mainly symptomatic testing through routine surveillance. This

lead in can be explained by transmission often occurring prior to symptomatic illness (He et al., 2020), which, it would be reasonable to assume, would be reflected in shedding of SARS-CoV-2 RNA in stool, and subsequently detected in WW. This 4-day lead illustrates the potential of WW to be an early warning tool, even in a setting such as England in 2020–2021 which had comprehensive clinical surveillance at the time. In circumstances with limited clinical surveillance, WW can provide accurate and timely estimates of regional prevalence with just a few samples.

A strength of our analysis comes from using data on prospective surveys of infection prevalence as well as reported cases of COVID-19. The ONS CIS survey has been designed to minimise bias in prevalence estimates by incentivising participation, and accounting for under-representation of specific groups of individuals. SARS-CoV-2 is shed in faeces of both symptomatic and asymptomatic individuals and so even with perfect surveillance for cases of COVID-19 the correlation between infection and case reporting will be imperfect due to differences in the probability of being symptomatic across regions. Furthermore, variability in reporting of symptomatic disease to the Test and Trace programme (and alternative surveillance systems specific to other countries) across socio-demographic strata is well documented (Byambasuren et al., 2021; Russell et al., 2021; Kendall et al., 2020), raising questions on whether clinical surveillance is an accurate and reliable estimate of disease burden. Indeed, previous comparisons of WW to disease incidence have provided variable results where under-reporting of cases has been hypothesised (Wu et al., 2020a; Hillary et al., 2021; Xiao et al., 2022; Peccia et al., 2020; Randazzo et al., 2020). For these reasons among others, we argue that inferred estimates of prevalence from wastewater RNA concentrations, a measure agnostic to the source of virus, are an informative addition to clinical disease surveillance. Furthermore, WW surveillance in England has since been expanded to include additional smaller sites within networks which may enable estimates of prevalence at a smaller geographical scale than described here.

The validation of prevalence estimates derived from wastewater illustrates the critical value of collecting metadata in addition to raw RNA concentrations alone,

and confirms findings from more established WW surveillance systems (O'Reilly et al., 2020). Several of the variables (suspended solids, ammonia concentration, phosphate) indicate organic and inorganic substances and likely approximate dilution of raw sewage with other sources of flow. The relationship is not likely to be linear as these variables also indicate the presence of agricultural runoff and so can also act as confounders. Inclusion of sample pH (with some degree of imputation in this dataset due to missingness) appears to further improve model prediction, and may reflect degradation of viral fragments at lower pH values that may reduce the sensitivity of RT-PCR, and has been previously observed in WW surveillance for poliovirus (Hamisu et al., 2020). While the composite fraction, replicate samples below limit of detection/quantification, reception delay, percentage of the catchment population, the catchment area and population fraction indicate a linear relationship, their inclusion still improves model prediction. It is interesting to note that the percentage of samples from each WW site that were composite (as opposed to "grab") had only a moderate effect on prevalence estimation. The use of composite samplers was dependent on their availability during the pandemic, with grab samples used for convenience as opposed to strategic intent. Ideally, this finding should be investigated using a comparative study design to investigate the possible added benefit of estimating SARS-CoV-2 from either approach. The probability of samples being below the limit of detection reduces as CIS positivity increases, which is to be expected, and its inclusion on the GBRT model improves the prediction of prevalence, perhaps providing further information when the estimated SARS-CoV-2 is less reliable at low prevalence. Further work will establish how these indicators should be used in settings with no measure of infection prevalence to improve inference of WW data.

This analysis has illustrated the predictive ability of WW at a time when comparatively few individuals were vaccinated against COVID-19 in England (by 1 March 2021 30% had received at least one dose). As vaccination increases the relationship between infection and faecal shedding may change and the predictive ability of the model will need to be monitored and potentially adapted to account for this. Studies of viral load in vaccinated but infected individuals have illustrated that

less virus is shed in the nasopharynx by vaccinated individuals ([Levine-Tiefenbrun et al., 2021](#); [McEllistrem et al., 2021](#); [Pritchard et al., 2021](#)), but there are currently no data on shedding of SARS-CoV-2 viral fragments in stool from infected but vaccinated individuals. Moreover, a study of healthcare workers where the alpha variant was dominant reported no difference in viral load by vaccination status ([Ioannou et al., 2021](#)), and preliminary analysis from England has not identified noticeable differences in WW through to March 2021 ([Hoffmann et al., 2021](#)).

The use of WW to infer prevalence is reliant on a converging sewer network that samples a sufficient proportion of the population of interest. Remote populations, however, such as islands or rural communities, may be served by septic tank systems, leading to blind-spots in observations, especially if there is a relationship between income and centralised waste removal provision. Consequently, the benefits of wastewater-based estimates are less obvious for low density settings. Additionally, the impact of sewage effluent from hospitals has not been accounted for in this analysis which could result in an over-estimation of prevalence within a population when compared to the CIS (as hospitalised individuals are not included in the sampling). Further work will investigate the impact of hospitals and other potential sources of bias. Our analysis illustrated the added benefit of including additional metadata within a statistical model to infer prevalence highlighting the importance of site-specific characteristics. Therefore, the use of our inference model outside the setting presented here should be avoided in the absence of further external validation and local information.

Nonetheless, close monitoring of emerging SARS-CoV-2 variants with changed phenotypic properties ([Davies et al., 2021a,b](#)) will continue to be needed; meta-genomic analysis of wastewater samples provides insight into the genomic diversity of virus in the community. Also, sampling of sites that cover small catchment areas will remain capable of revealing localised spikes in incidence used to detect hotspots and inform local public health authorities. Finally, the investment that has already been made in WW surveillance systems across England, combined with insights from analysis pipelines such as ours, can be leveraged to other communicable and non-

communicable diseases and human behaviours impacting personal and community health (Hassard et al., 2021; Polo et al., 2020). As the pandemic continues to evolve, and the threat to society becomes less acute, it is likely that surveillance of SARS-CoV-2 will need to become more sustainable, making the most of those investments. These data streams will remain an important feature for public health surveillance, complementing clinical surveillance as the country emerges out of the acute phase of the COVID-19 pandemic.

C.5 Methods

C.5.1 Data

Wastewater data from 45 sites in England

The WW and associated metadata used for the analysis are summarised in Table S1 (a correlation matrix between variables is also provided Fig. S1). Untreated influent WW were collected from each sewage treatment works located across England. Samples were either collected as "grab" samples (46%) or from a composite (24 h) sampler, and were transported to the laboratory and stored at 4 °C until processed (within 24 h). Physio-chemical analyses were carried out prior to further analysis. The physio-chemical analyses include quantification of pH, ammonia, orthophosphates, and suspended solids. For quantification of SARS-CoV-2 RNA from WW up to 150 µm of each sample was subjected to concentration and RNA extraction. The full details of the protocols are described in Farkas et al. (2021) and Walker et al. (2022), adhere to the MIQE guidelines (Bustin et al., 2009). The use of WW as a public health tool was rapidly expanded in scope in early 2020 using the protocol in Farkas et al. where the secondary concentration step (PEG precipitation) required an 18 h incubation step. An alternative procedure was later identified with a shorter incubation step (using ammonium sulfate precipitation) (Walker et al., 2022), with equivalent results, and was adopted on the 1 January 2021. The WW quantification described by Walker includes the phage Phi6 as a process positive control instead of PRRSV that was used in Farkas. Further details of the protocol are provided in the SI. Both procedures use the same extraction and RT-qPCR steps.

When the impact on LoD was investigated between protocol no difference in LoD was identified ($p = 0.356$ in an anova). The RT-qPCR assays focus on detection of the N1 and E gene, and here our analyses is on quantification of the N1 gene (see the SI for details of the primer used). A 10-fold dilution series of RNA standards within the range of 104–101 gene copies per microlitre (gc/ μ L) was included on each RT-qPCR plate to generate a standard curve. Standard curves were accepted if the slope of the log10 RNA standard concentration versus Cq was between -3.1 and -3.6 and if the r^2 for the curve was > 0.98 , a summary table of these data are presented in the SI. For each sample two replicate Cq values were used to calculate the gc/l in the original sample, based on the standard curves.

The limit of detection (LoD—the lowest concentration where all replicates were positive) and limit of quantification (LoQ—the lowest concentration where the coefficient of variance was below 0.25) were determined by running WW extracts (devoid of RNA) spiked with nominal concentrations of SARS-CoV-2 ranging from 100 to 2 gc/ μ L in replicates of 10. For the N1 gene the LoD was 1.7 gc/ μ L and the LoQ was 11.8 gc/ μ L in the protocol described by [Farkas et al. \(2021\)](#), and the protocol described by [Walker et al. \(2022\)](#) the LoD was 0.4 gc/ μ L and the LoQ was 4 gc/ μ L. Note that these estimates of LoD and LoQ should be regarded as theoretical limits, and in practice the limits are likely to be higher and vary. Figure C.4 illustrates that as CIS positivity increases replicate samples below LoD are less likely, and there is some evidence of site-specific variability, which is the subject of further investigation.

Of the total 6228 samples supplied from the lab in the time-frame of this analysis, 1365 samples returned a value of "NA" for SARS-CoV-2 (meaning that the submitted sample did not provide meaningful results for further use), and were removed from the analysis, leaving 4863 observations that were taken forward. Of these observations, 24.5% of replicate samples were below the LoD and 33.7% were below the LoQ; these values were retained in the analysis. Finally, a log10 transformation was applied to all the concentration variables and the target variables to reduce the heavy skewness of the distribution.

The ONS Coronavirus infection survey

The ONS COVID-19 infection survey data are used to infer subregional estimates of positivity (Pouwels et al., 2021). CIS is a large household survey with longitudinal follow-up (ISRCTN21086382). The study received ethical approval from the South Central Berkshire B Research Ethics Committee (20/SC/0195). Private households are randomly selected on a continuous basis from address lists and previous surveys to provide a representative sample across the UK. For the current study, only data from England was used. At the first visit, participants were asked for (optional) consent for follow-up visits every week for the next month, then monthly for 12 months from enrolment. At each visit, enrolled household members provided a nose and throat self-swab following instructions from the study worker. The CIS was designed to test 150,000 people every 2 weeks across England in October 2020, and this sample size was designed to correspond with 15,000–20,000 individuals in each of the nine governmental office regions (North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, East of England, London, South East, South West), providing an approximate 0.1%, 0.2%, and 0.5% margin of error on 0.1%, 0.5%, and 2%, respectively. In September 2020 the study design was adapted to have sufficient power to estimate prevalence at a subregional level, resulting in further increase in sample size of approximately 4-fold.

For the time periods relevant to this study (July 2020–March 2021), the number of participants per two-week period varied from 31,294 to 183,167 with an average of 126,655 participants, and typically up to 90% of participants had at least 5 visits. These participants were recruited from approximately 64,586 (range 14,965–93,940) households within any 2-week period. Further details are provided in the statistical bulletins provided by the Office for National Statistics ¹.

Linkage of WW data to CIS data

Individual level data from the CIS are not available due to confidentiality agreements, and so subregional estimates of positivity are the most geographically precise esti-

¹<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases>

mates of SARS-CoV-2 prevalence available. Within England the nine regions are divided into 119 subregions. The subregional estimates of positivity were computed weekly, and made publicly available (for example, see [“Sub-regional analysis for the UK” on the ONS website](#)).

To link the CIS subregions to wastewater data, we need to make the statistical assumption that wastewater data are uniformly distributed within subregions. Thus, prevalence estimates from wastewater should be an unbiased estimate from the general population within each subregion. In order to implement this the wastewater catchment areas were mapped onto CIS subregions. Typically, the catchment area for each treatment plant is smaller in size than a CIS subregion, so the catchment collects wastewater from just one subregion (this is the case for 56 of the 83 subregions), but there can be many catchments per subregion (a schematic example is given in Fig. S3, see the example of subregion C and the numerous catchments within this region). However, for some regions of England (especially in the greater London area) catchment areas cover multiple subregions. The mapping of the catchment areas to geographical areas were made available by the water companies. The result of the mapping is that each site-level dataset has a corresponding CIS subregion assigned to it. Where sites covered multiple subregions, the data were duplicated, and each duplicate assigned a subregion with a corresponding portion of the subregion covered. These proportions are used later when estimating prevalence at a subregion (and region) level, where the calculated proportions are used as weights.

A total of 4863 wastewater samples are available for the 45 sites within the dataset (over 214 days), corresponding to approximately 3 samples per site and week. These values were linearly interpolated to daily estimates for each site. These estimates were then merged onto the CIS dataset with an average gc/l for each daily estimate of CIS positivity, where multiple sites were combined using the calculated weights. With the details of the number of unique properties included or not in each LSOA/catchment intersection, the subregional coverage (population covered per WW site) is then inferred. The Test & Trace Pillar 1&2 case rates at Layer Super Output Areas (LSOA) level

Case rates of the number of new people infected per 100,000 individuals from Pillar 1&2 data were available from Public Health England. These data were aggregated to wastewater catchment level using a mapping from LSOA to catchment areas. Furthermore, the case rates were smoothed using a 7-day moving average to remove any artefactual weekly periodicity and therefore provide a better estimate of incidence. The same mapping procedure described for the CIS data is used here.

C.5.2 Models

Phenomenological model

The phenomenological model considers that prevalence (P , or the proportion of population infected) is related to C (the measured viral concentration in the sample, in units of gene copies per litre), by (Eq. (1)): $P = \frac{C \times Q_P}{S \times V}$ (1) where Q_P is the wastewater generated by one person per day (in L/day), S is the mean shedding rate, or the concentration of virus in stool of infected people (in gc/ml) and V is the mean volume of stool per person per day (in ml/day). A similar equation is used in [Ahmed et al. \(2020\)](#) using total flow rather than per person flow. We use the same model for all sites. Mean flow is set to 400 L/person/day, based data from 15 sites where flow data is routinely collected. Mean shedding rate is assumed to be 1.9×10^6 gc/mL from ref. 20. A mean stool volume of 128 g/person/day is assumed based on a review of 95 clinical studies—the majority being UK-based ([Rose et al., 2015](#)), and factor of 1.06 ml/g is used to convert from g to ml ([Penn et al., 2018](#)).

The model is used to identify which variables result in a considerable variability in the output (P). Two methods were utilised in the sensitivity analysis; variance based sensitivity analysis (VBSA) and PAWN. The VBSA is a global sensitivity analysis where the variance of the output is decomposed into fractions attributed to the inputs. The PAWN approach considers the entire distribution of the outcome (using the cumulative distribution function), which can be useful in cases where variance is not an adequate proxy of uncertainty.

Spatiotemporal analysis to obtain CIS prevalence estimates

Bayesian multilevel regression and post-stratification (MRP) is an increasingly used statistical technique to obtain representative estimates of prevalence or preferences at

the national and smaller regional levels³. By using random effects in the multilevel model stable estimates can be obtained for subnational levels from relative small samples or relatively rare outcomes. However, if there is an underlying spatial structure this needs to be captured by the MRP methodology to avoid biased estimates based on a model that assumes independent group-level errors. [Gao et al. \(2021\)](#) recently proposed a spatial MRP using a Besag-York-Mollié specification for the regional effect.

Here we extended the spatial MRP approach proposed by Gao et al. to a spatio-temporal context by adding a temporal component to the model. For the temporal components we use autoregressive or random walk processes with discrete time indices (weeks) to capture likely temporal effects in the MRP model. The choice of the type of directed conditional distribution for the time effect (random walk or autoregressive) type of space-time interaction (type I-IV ([Blangiardo et al., 2013](#))), and inclusion of additional covariates was guided by comparing the Watanabe-Akaike information criterion (WAIC) of the models. A type I space-time interaction, which assumes no spatial and/or temporal structure on the interaction, with first-order autoregressive terms were selected based on the WAIC.

The following covariates and interactions were considered for the MRP: age (2–11, 12–16, 17–24, 25–34, 35–49, 50–69, 70+); sex; ethnicity (white/non-white), CIS area; region (9 regions in England); time and two-way interactions of age and time, ethnicity and time, area and time, region and ethnicity, and region and age. After running the spatio-temporal regression model, post-stratification was used to obtain representative estimates of the outcome prevalence in the target population. Post-stratification tables were based on the conditional distribution of age and sex by area from ONS. The conditional distribution of ethnicity by these categories were obtained from the ETHPOP database ([Wohland et al., 2011](#)). Using the population sizes of each post-stratification cell of the target population, MRP adjusts for residual non-representative by post-stratifying by the percentage of each type in the actual overall population ([Gao et al., 2021](#)). The outputs of these analyses consist of median estimates of percent positivity rates and associated 95% credible intervals available

weekly at CIS subregional aggregation level between August 31, 2020 and February 14, 2021. For application in this study the estimates were up-sampled daily by linear interpolation between weekly estimates for each subregion and joined to the wastewater dataset using mappings from Lower Tier Local Authority (LTLA) to CIS subregions and from LTLA to catchment areas. Note that the resulting joined dataset was at CIS subregional level, where STW's contributions were weighted by their population covered in each overlapping subregion.

Modelling the population prevalence from WW and associated metadata

The prevalence of SARS-CoV-2 at a subregional level was estimated using WW and associated metadata. A set of candidate statistical models were used to examine the relationship of RNA concentration to the median of the posterior estimate of the CIS positivity rate. These were: (1) linear regression, (2) linear regression with random effects intercept, (3) linear regression with random effects intercept and slope, (4) gradient boosted regression tree (machine learning) models. Their out-of-sample predictive ability used to determine which model most effectively estimates prevalence. Random effects models

This model is a linear model where WW and the metadata have random effects on the slope and the intercept. The Python statsmodels package was used to implement these models. For a Bayesian description of the model with another application case in WBE (see ref. 36). The linear regression and linear regression with random effects intercept had the poorest performance and are not described further. Gradient Boosted Regression Trees

This model consists of a linear combination of non-linear predictors (also known as decision trees) trained by gradient descent ([Friedman, 2001](#)). Its performance has been shown on many regression examples ([Touzani et al., 2018](#); [Zhang and Haghani, 2015](#); [Chang et al., 2018](#)), and it is especially good at combining a large number or variety of input variables in a non-linear way. The implementation chosen here is an “Extreme Gradient Boosting” from python XGBoost package, which simply refers to an efficiently optimised Gradient Boosting Trees regressor using second-order optimisers.

Model evaluation

The models were compared using the mean absolute error (MAE) from out-of-sample prediction (to limit over-fitting of the models). The out-of-sample prediction was carried out using repeated random sampling: 50 random splits were generated in the available dataset, in each sample 80% of the data were retained for training and the remaining 20% for testing. The MAE was generated for each dataset and the combined MAE were obtained by averaging the test results from all samples. If $y_i, (i = 1, \dots, n)$ are the testing data for sample i , and \hat{y}_i the predictions from the model, the mean absolute error associated with the i th test set of size n is

$$MAE_i = \frac{1}{n} \sum_{i=1}^n |10_i^y - 10_i^{\hat{y}}| \quad (2)$$

The estimated MAE, and associated standard deviation and 95% confidence intervals are then computed as $MAE = \frac{1}{N_{samples}} \sum_{i \in samples} MAE_i$ (3) $\sigma_{MAE} = \sqrt{\frac{\sum_{i \in samples} (MAE_i - MAE)^2}{N_{samples}}}$ (4) $CI_{95\%} \approx \left[\frac{-2\sigma_{MAE}}{\sqrt{N_{samples}}}, \frac{2\sigma_{MAE}}{\sqrt{N_{samples}}} \right]$ (5)

Model performance across STWs' characteristics

In October and November 2020, interviews were conducted with nine water utilities in England to document information on sewer network characteristics that could impact model performance (average daily flow, proportion of pumping in catchment, combined vs foul sewers etc.) The knowledge gathered from the interviews and a related questionnaire provided both qualitative and quantitative data, and transformed into variables used to assess model performance. An average MAE was obtained for each wastewater treatment plant by averaging the scores of CIS subregions included in the catchment weighted by the population covered. These resulting WTP MAE were then plotted across sites characteristics (Fig. S7). Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked.

Data availability

The data used in this study are available online at [this url](#). The full data that support the findings of this study are available alongside the code, within the repository provided in the code availability statement.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *arXiv:1605.08695 [cs]*.

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B., Affeldt, C., Agathos, M., Agatsuma, K., Aggarwal, N., Aguiar, O. D., Aiello, L., Ain, A., Ajith, P., Allen, B., Allocca, A., Altin, P. A., Anderson, S. B., Anderson, W. G., Arai, K., Arain, M. A., Araya, M. C., Arceneaux, C. C., Areeda, J. S., Arnaud, N., Arun, K. G., Ascenzi, S., Ashton, G., Ast, M., Aston, S. M., Astone, P., Aufmuth, P., Aulbert, C., Babak, S., Bacon, P., Bader, M. K. M., Baker, P. T., Baldaccini, F., Ballardin, G., Ballmer, S. W., Barayoga, J. C., Barclay, S. E., Barish, B. C., Barker, D., Barone, F., Barr, B., Barsotti, L., Barsuglia, M., Barta, D., Bartlett, J., Barton, M. A., Bartos, I., Bassiri, R., Basti, A., Batch, J. C., Baune, C., Bavigadda, V., Bazzan, M., Behnke, B., Bejger, M., Belczynski, C., Bell, A. S., Bell, C. J., Berger, B. K., Bergman, J., Bergmann, G., Berry, C. P. L., Bersanetti, D., Bertolini, A., Betzwieser, J., Bhagwat, S., Bhandare, R., Bilenko, I. A., Billingsley, G., Birch, J., Birney, R., Birnholtz, O., Biscans, S., Bisht, A., Bitossi, M., Biwer, C.,

Bizouard, M. A., Blackburn, J. K., Blair, C. D., Blair, D. G., Blair, R. M., Bloemen, S., Bock, O., Bodiya, T. P., Boer, M., Bogaert, G., Bogan, C., Bohe, A., Bojtos, P., Bond, C., Bondu, F., Bonnand, R., Boom, B. A., Bork, R., Boschi, V., Bose, S., Bouffanais, Y., Bozzi, A., Bradaschia, C., Brady, P. R., Braginsky, V. B., Branchesi, M., Brau, J. E., Briant, T., Brillet, A., Brinkmann, M., Brisson, V., Brockill, P., Brooks, A. F., Brown, D. A., Brown, D. D., Brown, N. M., Buchanan, C. C., Buikema, A., Bulik, T., Bulten, H. J., Buonanno, A., Buskulic, D., Buy, C., Byer, R. L., Cabero, M., Cadonati, L., Cagnoli, G., Cahillane, C., Bustillo, J. C., Callister, T., Calloni, E., Camp, J. B., Cannon, K. C., Cao, J., Capano, C. D., Capocasa, E., Carbognani, F., Caride, S., Diaz, J. C., Casentini, C., Caudill, S., Cavaglià, M., Cavalier, F., Cavalieri, R., Cella, G., Cepeda, C. B., Baiardi, L. C., Cerretani, G., Cesarini, E., Chakraborty, R., Chalermongsak, T., Chamberlin, S. J., Chan, M., Chao, S., Charlton, P., Chassande-Mottin, E., Chen, H. Y., Chen, Y., Cheng, C., Chincarini, A., Chiummo, A., Cho, H. S., Cho, M., Chow, J. H., Christensen, N., Chu, Q., Chua, S., Chung, S., Ciani, G., Clara, F., Clark, J. A., Cleva, F., Coccia, E., Cohadon, P.-F., Colla, A., Collette, C. G., Cominsky, L., Constancio, M., Conte, A., Conti, L., Cook, D., Corbitt, T. R., Cornish, N., Corsi, A., Cortese, S., Costa, C. A., Coughlin, M. W., Coughlin, S. B., Coulon, J.-P., Countryman, S. T., Couvares, P., Cowan, E. E., Coward, D. M., Cowart, M. J., Coyne, D. C., Coyne, R., Craig, K., Creighton, J. D. E., Creighton, T. D., Cripe, J., Crowder, S. G., Cruise, A. M., Cumming, A., Cunningham, L., Cuoco, E., Canton, T. D., Danilishin, S. L., D'Antonio, S., Danzmann, K., Darman, N. S., Da Silva Costa, C. F., Dattilo, V., Dave, I., Daveloza, H. P., Davier, M., Davies, G. S., Daw, E. J., Day, R., De, S., DeBra, D., Debreczeni, G., Degallaix, J., De Laurentis, M., Deléglise, S., Del Pozzo, W., Denker, T., Dent, T., Dereli, H., Dergachev, V., DeRosa, R. T., De Rosa, R., DeSalvo, R., Dhurandhar, S., Díaz, M. C., Di Fiore, L., Di Giovanni, M., Di Lieto, A., Di Pace, S., Di Palma, I., Di Virgilio, A., Dojcinoski, G., Dolique, V., Donovan, F., Dooley, K. L., Doravari, S., Douglas, R., Downes, T. P., Drago, M., Drever, R. W. P., Driggers, J. C., Du, Z., Ducrot, M., Dwyer, S. E., Edo, T. B., Edwards, M. C., Effler, A., Eggenstein, H.-B., Ehrens, P., Eichholz, J., Eikenberry, S. S., Engels, W., Essick, R. C., Etzel, T., Evans, M., Evans, T. M., Everett, R., Factourovich, M., Fafone, V., Fair, H., Fairhurst, S., Fan, X., Fang, Q., Farinon, S., Farr, B., Farr, W. M., Favata, M., Fays, M., Fehrmann, H., Fejer, M. M., Feldbaum, D., Ferrante, I., Ferreira, E. C., Ferrini, F., Fidecaro, F., Finn, L. S., Fiori, I., Fiorucci, D., Fisher, R. P., Flaminio, R.,

- Fletcher, M., Fong, H., Fournier, J.-D., Franco, S., Frasca, S., Frasconi, F., Frede, M., Frei, Z., Freise, A., Frey, R., Frey, V., Fricke, T. T., Fritschel, P., Frolov, V. V., Fulda, P., Fyffe, M., Gabbard, H. A. G., Gair, J. R., Gammaitoni, L., Gaonkar, S. G., Garufi, F., Gatto, A., Gaur, G., Gehrels, N., Gemme, G., Gendre, B., Genin, E., Gennai, A., George, J., Gergely, L., Germain, V., Ghosh, A., Ghosh, A., Ghosh, S., Giaime, J. A., Giardina, K. D., Giazotto, A., Gill, K., Glaefke, A., Gleason, J. R., Goetz, E., Goetz, R., Gondan, L., González, G., Castro, J. M. G., Gopakumar, A., Gordon, N. A., Gorodetsky, M. L., Gossan, S. E., Gosselin, M., Gouaty, R., Graef, C., Graff, P. B., Granata, M., Grant, A., Gras, S., Gray, C., Greco, G., Green, A. C., Greenhalgh, R. J. S., Groot, P., Grote, H., Grunewald, S., Guidi, G. M., Guo, X., Gupta, A., Gupta, M. K., Gushwa, K. E., Gustafson, E. K., Gustafson, R., Hacker, J. J., Hall, B. R., Hall, E. D., Hammond, G., Haney, M., Hanke, M. M., Hanks, J., Hanna, C., Hannam, M. D., Hanson, J., Hardwick, T., Harms, J., Harry, G. M., Harry, I. W., Hart, M. J., Hartman, M. T., Haster, C.-J., Haughian, K., Healy, J., Heefner, J., Heidmann, A., Heintze, M. C., Heinzl, G., Heitmann, H., Hello, P., Hemming, G., Hendry, M., Heng, I. S., Hennig, J., Heptonstall, A. W., Heurs, M., Hild, S., Hoak, D., Hodge, K. A., Hofman, D., Hollitt, S. E., Holt, K., Holz, D. E., Hopkins, P., Hosken, D. J., Hough, J., Houston, E. A., Howell, E. J., Hu, Y. M., Huang, S., Huerta, E. A., Huet, D., Hughey, B., Husa, S., Huttner, S. H., Huynh-Dinh, T., Idrisy, A., Indik, N., Ingram, D. R., Inta, R., Isa, H. N., Isac, J.-M., Isi, M., Islas, G., Isogai, T., Iyer, B. R., Izumi, K., Jacobson, M. B., Jacqmin, T., Jang, H., Jani, K., Jaranowski, P., Jawahar, S., Jiménez-Forteza, F., Johnson, W. W., Johnson-McDaniel, N. K., Jones, D. I., Jones, R., Jonker, R. J. G., Ju, L., Haris, K., Kalaghatgi, C. V., Kalogera, V., Kandhasamy, S., Kang, G., Kanner, J. B., Karki, S., Kasprzack, M., Katsavounidis, E., Katzman, W., Kaufer, S., Kaur, T., Kawabe, K., Kawazoe, F., Kéfélian, F., Kehl, M. S., Keitel, D., Kelley, D. B., Kells, W., Kennedy, R., Keppel, D. G., Key, J. S., Khalaidovski, A., Khalili, F. Y., Khan, I., Khan, S., Khan, Z., Khazanov, E. A., Kijbunchoo, N., Kim, C., Kim, J., Kim, K., Kim, N.-G., Kim, N., Kim, Y.-M., King, E. J., King, P. J., Kinzel, D. L., Kissel, J. S., Kleybolte, L., Klimenko, S., Koehlenbeck, S. M., Kokeyama, K., Koley, S., Kondrashov, V., Kontos, A., Koranda, S., Korobko, M., Korth, W. Z., Kowalska, I., Kozak, D. B., Kringel, V., Krishnan, B., Królak, A., Krueger, C., Kuehn, G., Kumar, P., Kumar, R., Kuo, L., Kutynia, A., Kwee, P., Lackey, B. D., Landry, M., Lange, J., Lantz, B., Lasky, P. D., Lazzarini, A., Lazzaro, C., Leaci, P., Leavey, S., Lebigot, E. O., Lee, C. H., Lee, H. K.,

Lee, H. M., Lee, K., Lenon, A., Leonardi, M., Leong, J. R., Leroy, N., Letendre, N., Levin, Y., Levine, B. M., Li, T. G. F., Libson, A., Littenberg, T. B., Lockerbie, N. A., Logue, J., Lombardi, A. L., London, L. T., Lord, J. E., Lorenzini, M., Lorient, V., Lormand, M., Losurdo, G., Lough, J. D., Lousto, C. O., Lovelace, G., Lück, H., Lundgren, A. P., Luo, J., Lynch, R., Ma, Y., MacDonald, T., Machenschalk, B., MacInnis, M., Macleod, D. M., Magaña-Sandoval, F., Magee, R. M., Mageswaran, M., Majorana, E., Maksimovic, I., Malvezzi, V., Man, N., Mandel, I., Mandic, V., Mangano, V., Mansell, G. L., Manske, M., Mantovani, M., Marchesoni, F., Marion, F., Márka, S., Márka, Z., Markosyan, A. S., Maros, E., Martelli, F., Martellini, L., Martin, I. W., Martin, R. M., Martynov, D. V., Marx, J. N., Mason, K., Masserot, A., Massinger, T. J., Masso-Reid, M., Matichard, F., Matone, L., Mavalvala, N., Mazumder, N., Mazzolo, G., McCarthy, R., McClelland, D. E., McCormick, S., McGuire, S. C., McIntyre, G., McIver, J., McManus, D. J., McWilliams, S. T., Meacher, D., Meadors, G. D., Meidam, J., Melatos, A., Mendell, G., Mendoza-Gandara, D., Mercer, R. A., Merilh, E., Merzougui, M., Meshkov, S., Messenger, C., Messick, C., Meyers, P. M., Mezzani, F., Miao, H., Michel, C., Middleton, H., Mikhailov, E. E., Milano, L., Miller, J., Millhouse, M., Minenkov, Y., Ming, J., Mirshekari, S., Mishra, C., Mitra, S., Mitrofanov, V. P., Mitselmakher, G., Mittleman, R., Moggi, A., Mohan, M., Mohapatra, S. R. P., Montani, M., Moore, B. C., Moore, C. J., Moraru, D., Moreno, G., Morriss, S. R., Mossavi, K., Mours, B., Mow-Lowry, C. M., Mueller, C. L., Mueller, G., Muir, A. W., Mukherjee, A., Mukherjee, D., Mukherjee, S., Mukund, N., Mullavey, A., Munch, J., Murphy, D. J., Murray, P. G., Mytidis, A., Nardecchia, I., Naticchioni, L., Nayak, R. K., Necula, V., Nedkova, K., Nelemans, G., Neri, M., Neunzert, A., Newton, G., Nguyen, T. T., Nielsen, A. B., Nissanke, S., Nitz, A., Nocera, F., Nolting, D., Normandin, M. E. N., Nuttall, L. K., Oberling, J., Ochsner, E., O'Dell, J., Oelker, E., Ogini, G. H., Oh, J. J., Oh, S. H., Ohme, F., Oliver, M., Oppermann, P., Oram, R. J., O'Reilly, B., O'Shaughnessy, R., Ott, C. D., Ottaway, D. J., Ottens, R. S., Overmire, H., Owen, B. J., Pai, A., Pai, S. A., Palamos, J. R., Palashov, O., Palomba, C., Pal-Singh, A., Pan, H., Pan, Y., Pankow, C., Pannarale, F., Pant, B. C., Paoletti, F., Paoli, A., Papa, M. A., Paris, H. R., Parker, W., Pascucci, D., Pasqualetti, A., Passaquieti, R., Passuello, D., Patricelli, B., Patrick, Z., Pearlstone, B. L., Pedraza, M., Pedurand, R., Pekowsky, L., Pele, A., Penn, S., Perreca, A., Pfeiffer, H. P., Phelps, M., Piccinni, O., Pichot, M., Pickenpack, M., Piergiovanni, F., Pierro, V., Pillant, G., Pinard, L., Pinto, I. M., Pitkin,

M., Poeld, J. H., Poggiani, R., Popolizio, P., Post, A., Powell, J., Prasad, J., Predoi, V., Premachandra, S. S., Prestegard, T., Price, L. R., Prijatelj, M., Principe, M., Privitera, S., Prix, R., Prodi, G. A., Prokhorov, L., Puncken, O., Punturo, M., Puppo, P., Pürner, M., Qi, H., Qin, J., Quetschke, V., Quintero, E. A., Quitzow-James, R., Raab, F. J., Rabeling, D. S., Radkins, H., Raffai, P., Raja, S., Rakhmanov, M., Ramet, C. R., Rapagnani, P., Raymond, V., Razzano, M., Re, V., Read, J., Reed, C. M., Regimbau, T., Rei, L., Reid, S., Reitze, D. H., Rew, H., Reyes, S. D., Ricci, F., Riles, K., Robertson, N. A., Robie, R., Robinet, F., Rocchi, A., Rolland, L., Rollins, J. G., Roma, V. J., Romano, J. D., Romano, R., Romanov, G., Romie, J. H., Rosińska, D., Rowan, S., Rüdiger, A., Ruggi, P., Ryan, K., Sachdev, S., Sadecki, T., Sadeghian, L., Salconi, L., Saleem, M., Salemi, F., Samajdar, A., Sammut, L., Sampson, L. M., Sanchez, E. J., Sandberg, V., Sandeen, B., Sanders, G. H., Sanders, J. R., Sassolas, B., Sathyaprakash, B. S., Saulson, P. R., Sauter, O., Savage, R. L., Sawadsky, A., Schale, P., Schilling, R., Schmidt, J., Schmidt, P., Schnabel, R., Schofield, R. M. S., Schönbeck, A., Schreiber, E., Schuette, D., Schutz, B. F., Scott, J., Scott, S. M., Sellers, D., Sengupta, A. S., Sentenac, D., Sequino, V., Sergeev, A., Serna, G., Setyawati, Y., Seigny, A., Shaddock, D. A., Shaffer, T., Shah, S., Shahriar, M. S., Shaltev, M., Shao, Z., Shapiro, B., Shawhan, P., Sheperd, A., Shoemaker, D. H., Shoemaker, D. M., Siellez, K., Siemens, X., Sigg, D., Silva, A. D., Simakov, D., Singer, A., Singer, L. P., Singh, A., Singh, R., Singhal, A., Sintès, A. M., Slagmolen, B. J. J., Smith, J. R., Smith, M. R., Smith, N. D., Smith, R. J. E., Son, E. J., Sorazu, B., Sorrentino, F., Souradeep, T., Srivastava, A. K., Staley, A., Steinke, M., Steinlechner, J., Steinlechner, S., Steinmeyer, D., Stephens, B. C., Stevenson, S. P., Stone, R., Strain, K. A., Straniero, N., Stratta, G., Strauss, N. A., Strigin, S., Sturani, R., Stuver, A. L., Summerscales, T. Z., Sun, L., Sutton, P. J., Swinkels, B. L., Szczepańczyk, M. J., Tacca, M., Talukder, D., Tanner, D. B., Tápai, M., Tarabrin, S. P., Taracchini, A., Taylor, R., Theeg, T., Thirugnanasambandam, M. P., Thomas, E. G., Thomas, M., Thomas, P., Thorne, K. A., Thorne, K. S., Thrane, E., Tiwari, S., Tiwari, V., Tokmakov, K. V., Tomlinson, C., Tonelli, M., Torres, C. V., Torrie, C. I., Töyrä, D., Travasso, F., Traylor, G., Trifirò, D., Tringali, M. C., Trozzo, L., Tse, M., Turconi, M., Tuyenbayev, D., Ugolini, D., Unnikrishnan, C. S., Urban, A. L., Usman, S. A., Vahlbruch, H., Vajente, G., Valdes, G., Vallisneri, M., van Bakel, N., van Beuzekom, M., van den Brand, J. F. J., Van Den Broeck, C., Vander-Hyde, D. C., van der Schaaf, L., van Heijningen, J. V., van Veggel, A. A., Vardaro, M., Vass, S., Vasúth, M.,

- Vaulin, R., Vecchio, A., Vedovato, G., Veitch, J., Veitch, P. J., Venkateswara, K., Verkindt, D., Vetrano, F., Viceré, A., Vinciguerra, S., Vine, D. J., Vinet, J.-Y., Vitale, S., Vo, T., Vocca, H., Vorvick, C., Voss, D., Vousden, W. D., Vyatchanin, S. P., Wade, A. R., Wade, L. E., Wade, M., Waldman, S. J., Walker, M., Wallace, L., Walsh, S., Wang, G., Wang, H., Wang, M., Wang, X., Wang, Y., Ward, H., Ward, R. L., Warner, J., Was, M., Weaver, B., Wei, L.-W., Weinert, M., Weinstein, A. J., Weiss, R., Welborn, T., Wen, L., Weßels, P., Westphal, T., Wette, K., Whelan, J. T., Whitcomb, S. E., White, D. J., Whiting, B. F., Wiesner, K., Wilkinson, C., Willems, P. A., Williams, L., Williams, R. D., Williamson, A. R., Willis, J. L., Willke, B., Wimmer, M. H., Winkelmann, L., Winkler, W., Wipf, C. C., Wiseman, A. G., Wittel, H., Woan, G., Worden, J., Wright, J. L., Wu, G., Yablon, J., Yakushin, I., Yam, W., Yamamoto, H., Yancey, C. C., Yap, M. J., Yu, H., Yvert, M., Zadrożny, A., Zangrando, L., Zanolin, M., Zendri, J.-P., Zevin, M., Zhang, F., Zhang, L., Zhang, M., Zhang, Y., Zhao, C., Zhou, M., Zhou, Z., Zhu, X. J., Zucker, M. E., Zuraw, S. E., Zweizig, J., and LIGO Scientific Collaboration and Virgo Collaboration (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6).
- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S. (2021). A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Information Fusion*, 76:243–297.
- Abnar, S. and Zuidema, W. (2020). Quantifying Attention Flow in Transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4190–4197, Online. Association for Computational Linguistics.
- Agarap, A. F. (2019). Deep Learning using Rectified Linear Units (ReLU).
- Agol, E., Cowan, N. B., Knutson, H. A., Deming, D., Steffen, J. H., Henry, G. W., and Charbonneau, D. (2010). The Climate of HD 189733b from Fourteen Transits and Eclipses Measured by Spitzer. *The Astrophysical Journal*, 721:1861–1877.
- Aguirre, C., Pichara, K., and Becker, I. (2019). Deep multi-survey classification of variable stars. *Monthly Notices of the Royal Astronomical Society*, 482:5078–5092.

- Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tschärke, B., Verhagen, R., Smith, W. J. M., Zaug, J., Dierens, L., Hugenholz, P., Thomas, K. V., and Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *The Science of the Total Environment*, 728:138764.
- Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., and Roberts, S. J. (2015). Precise time series photometry for the Kepler-2.0 mission. *Monthly Notices of the Royal Astronomical Society*, 447:2880–2893.
- Aigrain, S., Parviainen, H., and Pope, B. J. S. (2016). K2SC: Flexible systematics correction and detrending of K2 light curves using Gaussian process regression. *Monthly Notices of the Royal Astronomical Society*, 459:2408–2419.
- Aigrain, S., Pont, F., and Zucker, S. (2012). A simple method to estimate radial velocity variations due to stellar activity using photometry: Activity-induced RV variations from photometry. *Monthly Notices of the Royal Astronomical Society*, 419(4):3147–3158.
- Al-Refaie, A. F., Changeat, Q., Waldmann, I. P., and Tinetti, G. (2021). TauREx 3: A Fast, Dynamic, and Extendable Framework for Retrievals. *The Astrophysical Journal*, 917(1):37.
- Alibert, Y. and Venturini, J. (2019). Using deep neural networks to compute the mass of forming planets. *Astronomy & Astrophysics*, 626:A21.
- Allam Jr., T. and McEwen, J. D. (2021). Paying Attention to Astronomical Transients: Photometric Classification with the Time-Series Transformer. *arXiv:2105.06178 [astro-ph]*.
- Andersen, J. M. and Korhonen, H. (2015). Stellar activity as noise in exoplanet detection - II. Application to M dwarfs. *Monthly Notices of the RAS*, 448(4):3053–3069.
- Anderson, D. R., Hellier, C., Gillon, M., Triaud, A. H. M. J., Smalley, B., Hebb, L., Collier Cameron, A., Maxted, P. F. L., Queloz, D., West, R. G., Bentley, S. J., Enoch, B., Horne, K., Lister, T. A., Mayor, M., Parley, N. R., Pepe, F., Pollacco, D., Ségransan,

- D., Udry, S., and Wilson, D. M. (2010). WASP-17b: An Ultra-Low Density Planet in a Probable Retrograde Orbit. *The Astrophysical Journal*, 709:159–167.
- Anderson, D. R., Smith, A. M. S., Lanotte, A. A., Barman, T. S., Collier Cameron, A., Campo, C. J., Gillon, M., Harrington, J., Hellier, C., Maxted, P. F. L., Queloz, D., Triaud, A. H. M. J., and Wheatley, P. J. (2011). Thermal emission at 4.5 and 8 μm of WASP-17b, an extremely large planet in a slightly eccentric orbit. *Monthly Notices of the Royal Astronomical Society*, 416:2108–2122.
- Angel, J. R. P., Cheng, A. Y. S., and Woolf, N. J. (1986). A space telescope for infrared spectroscopy of Earth-like planets. *Nature*, 322(6077):341–343.
- Ansdell, M., Ioannou, Y., Osborn, H. P., Sasdelli, M., Smith, J. C., Jenkins, J. M., Raissi, C., and Angerhausen, D. (2018). Scientific Domain Knowledge Improves Exoplanet Transit Classification with Deep Learning. *The Astrophysical Journal*, 869(1):L7.
- Armstrong, D. J., Kirk, J., Lam, K. W. F., McCormac, J., Walker, S. R., Brown, D. J. A., Osborn, H. P., Pollacco, D. L., and Spake, J. (2015). K2 Variable Catalogue: Variable stars and eclipsing binaries in K2 campaigns 1 and 0. *Astronomy & Astrophysics*, 579:A19.
- Arora, S., Nag, A., Sethi, J., Rajvanshi, J., Saxena, S., Shrivastava, S. K., and Gupta, A. B. (2020). Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India. *Water Science and Technology: A Journal of the International Association on Water Pollution Research*, 82(12):2823–2836.
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., Babej, T., Bach, Y. P., Bachetti, M., Bakanov, A. V., Bamford, S. P., Barentsen, G., Barmby, P., Baumbach, A., Berry, K. L., Biscani, F., Boquien, M., Bostroem, K. A., Bouma, L. G., Brammer, G. B., Bray, E. M., Breytenbach, H., Buddelmeijer, H., Burke, D. J., Calderone, G., Cano Rodríguez, J. L., Cara, M., Cardoso, J. V. M., Cheedella, S., Copin, Y., Corrales, L., Crichton, D., D’Avella, D., Deil, C., Depagne, E., Dietrich, J. P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L. A., Finethy, T., Fox, R. T., Garrison, L. H., Gibbons, S. L. J., Goldstein, D. A., Gommers, R., Greco, J. P., Greenfield, P., Groener, A. M.,

- Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A. J., Hosseinzadeh, G., Hu, L., Hunkeler, J. S., Ivezić, Z., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N. S., Kerzendorf, W. E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A. M., Kumar, A., Lee, A., Lenz, D., Littlefair, S. P., Ma, Z., Macleod, D. M., Mastropietro, M., McCully, C., Montagnac, S., Morris, B. M., Mueller, M., Mumford, S. J., Muna, D., Murphy, N. A., Nelson, S., Nguyen, G. H., Ninan, J. P., Nöthe, M., Ogaz, S., Oh, S., Parejko, J. K., Parley, N., Pascual, S., Patil, R., Patil, A. A., Plunkett, A. L., Prochaska, J. X., Rastogi, T., Reddy Janga, V., Sabater, J., Sakurikar, P., Seifert, M., Sherbert, L. E., Sherwood-Taylor, H., Shih, A. Y., Sick, J., Silbiger, M. T., Singanamalla, S., Singer, L. P., Sladen, P. H., Sooley, K. A., Sornarajah, S., Streicher, O., Teuben, P., Thomas, S. W., Tremblay, G. R., Turner, J. E. H., Terrón, V., van Kerkwijk, M. H., de la Vega, A., Watkins, L. L., Weaver, B. A., Whitmore, J. B., Woillez, J., Zabalza, V., and Astropy Contributors (2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astronomical Journal*, 156:123.
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., and Streicher, O. (2013). Astropy: A community Python package for astronomy. *Astronomy and Astrophysics*, 558:A33.
- Auddy, S. and Lin, M.-K. (2020). A Machine Learning Model to Infer Planet Masses from Gaps Observed in Protoplanetary Disks. *The Astrophysical Journal*, 900(1):62.
- Audenaert, J., Kuszlewicz, J. S., Handberg, R., Tkachenko, A., Armstrong, D. J., Hon, M., Kgoadi, R., Lund, M. N., Bell, K. J., Bugnet, L., Bowman, D. M., Johnston, C., García, R. A., Stello, D., Molnár, L., Plachy, E., Buzasi, D., Aerts, C., and T'DA Collaboration (2021). TESS Data for Asteroseismology (T'DA) Stellar Variability Classification Pipeline: Setup and Application to the Kepler Q9 Data. *The Astronomical Journal*, 162:209.
- Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer Normalization.

- Baglin, A. (2003). COROT: A minisat for pionnier science, asteroseismology and planets finding. *Advances in Space Research*, 31(2):345–349.
- Bahdanau, D., Cho, K., and Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate.
- Bailes, M., Lyne, A. G., and Shemar, S. L. (1991). A planet orbiting the neutron star PSR1829–10. *Nature*, 352(6333):311–313.
- Baraniuk, C. (2020). Sewage monitoring is the UK’s next defence against covid-19. *BMJ (Clinical research ed.)*, 370:m2599.
- Barban, C., Hill, F., and Kras, S. (2004). Simultaneous Velocity-Intensity Spectral and Cross-Spectral Fitting of Helioseismic Data. *The Astrophysical Journal*, 602(1):516–527.
- Barbara, N. H., Bedding, T. R., Fulcher, B. D., Murphy, S. J., and Van Reeth, T. (2022). Classifying Kepler light curves for 12,000 A and F stars using supervised feature-based machine learning. *Monthly Notices of the Royal Astronomical Society*, 514(2):2793–2804.
- Baron, D. (2019). Machine Learning in Astronomy: A practical overview.
- Barragán, O., Gandolfi, D., Dai, F., Livingston, J., Persson, C. M., Hirano, T., Narita, N., Csizmadia, S., Winn, J. N., Nespral, D., Prieto-Arranz, J., Smith, A. M. S., Nowak, G., Albrecht, S., Antoniciello, G., Justesen, A. B., Cabrera, J., Cochran, W. D., Deeg, H., Eigtmüller, P., Endl, M., Erikson, A., Fridlund, M., Fukui, A., Grziwa, S., Guenther, E., Hatzes, A. P., Hidalgo, D., Johnson, M. C., Korth, J., Palte, E., Patzold, M., Rauer, H., Tanaka, Y., and Eylen, V. V. (2018). K2-141 b - A 5-M \oplus super-Earth transiting a K7 V star every 6.7 h. *Astronomy & Astrophysics*, 612:A95.
- Batson, J. and Royer, L. (2019). Noise2Self: Blind Denoising by Self-Supervision. In *Proceedings of the 36th International Conference on Machine Learning*, pages 524–533. PMLR.
- Baumeister, P., Padovan, S., Tosi, N., Montavon, G., Nettelmann, N., MacKenzie, J., and Godolt, M. (2020). Machine-learning Inference of the Interior Structure of Low-mass Exoplanets. *The Astrophysical Journal*, 889(1):42.

- Becker, I., Pichara, K., Catelan, M., Protopapas, P., Aguirre, C., and Nikzat, F. (2020). Scalable end-to-end recurrent neural network for variable star classification. *Monthly Notices of the Royal Astronomical Society*, 493:2981–2995.
- Benneke, B., Wong, I., Piaulet, C., Knutson, H. A., Lothringer, J., Morley, C. V., Crossfield, I. J. M., Gao, P., Greene, T. P., Dressing, C., Dragomir, D., Howard, A. W., McCullough, P. R., Kempton, E. M.-R., Fortney, J. J., and Fraine, J. (2019). Water Vapor and Clouds on the Habitable-Zone Sub-Neptune Exoplanet K2-18b. *The Astrophysical Journal Letters*, 887(1):L14.
- Benz, W., Broeg, C., Fortier, A., Rando, N., Beck, T., Beck, M., Queloz, D., Ehrenreich, D., Maxted, P. F. L., Isaak, K. G., Billot, N., Alibert, Y., Alonso, R., António, C., Asquier, J., Bandy, T., Bárczy, T., Barrado, D., Barros, S. C. C., Baumjohann, W., Bekkelien, A., Bergomi, M., Biondi, F., Bonfils, X., Borsato, L., Brandeker, A., Busch, M. D., Cabrera, J., Cessa, V., Charnoz, S., Chazelas, B., Collier Cameron, A., Corral Van Damme, C., Cortes, D., Davies, M. B., Deleuil, M., Deline, A., Delrez, L., Demangeon, O., Demory, B. O., Erikson, A., Farinato, J., Fossati, L., Fridlund, M., Futyan, D., Gandolfi, D., Garcia Munoz, A., Gillon, M., Guterman, P., Gutierrez, A., Hasiba, J., Heng, K., Hernandez, E., Hoyer, S., Kiss, L. L., Kovacs, Z., Kuntzer, T., Laskar, J., Lecavelier des Etangs, A., Lendl, M., López, A., Lora, I., Lovis, C., Lüftinger, T., Magrin, D., Malvasio, L., Marafatto, L., Michaelis, H., de Miguel, D., Modrego, D., Munari, M., Nascimbeni, V., Olofsson, G., Ottacher, H., Ottensamer, R., Pagano, I., Palacios, R., Pallé, E., Peter, G., Piazza, D., Piotto, G., Pizarro, A., Pollaco, D., Ragazzoni, R., Ratti, F., Rauer, H., Ribas, I., Rieder, M., Rohlfs, R., Safa, F., Salatti, M., Santos, N. C., Scandariato, G., Ségransan, D., Simon, A. E., Smith, A. M. S., Sordet, M., Sousa, S. G., Steller, M., Szabó, G. M., Szoke, J., Thomas, N., Tschentscher, M., Udry, S., Van Grootel, V., Viotto, V., Walter, I., Walton, N. A., Wildi, F., and Wolter, D. (2021). The CHEOPS mission. *Experimental Astronomy*, 51:109–151.
- Berger, T. A., Huber, D., Gaidos, E., van Saders, J. L., and Weiss, L. M. (2020). The Gaia-Kepler Stellar Properties Catalog. II. Planet Radius Demographics as a Function of Stellar Mass and Age. *The Astronomical Journal*, 160:108.
- Berta, Z. K., Charbonneau, D., Désert, J.-M., Miller-Ricci Kempton, E., McCullough, P. R., Burke, C. J., Fortney, J. J., Irwin, J., Nutzman, P., and Homeier, D. (2012). The Flat

- Transmission Spectrum of the Super-Earth GJ1214b from Wide Field Camera 3 on the Hubble Space Telescope. *The Astrophysical Journal*, 747:35.
- Beust, H., Lagrange-Henri, A. M., Vidal-Madjar, A., and Ferlet, R. (1990). The beta Pictoris circumstellar disk. X. Numerical simulations of infalling evaporating bodies. *Astronomy and Astrophysics*, 236:202.
- Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D. (2019). Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 20(28):1–6.
- Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 7:39–55.
- Blomme, J., Debosscher, J., De Ridder, J., Aerts, C., Gilliland, R. L., Christensen-Dalsgaard, J., Kjeldsen, H., Brown, T. M., Borucki, W. J., Koch, D., Jenkins, J. M., Kurtz, D. W., Stello, D., Stevens, I. R., Suran, M. D., and Derekas, A. (2010). Automated Classification of Variable Stars in the Asteroseismology Program of the Kepler Space Mission. *The Astrophysical Journal*, 713:L204–L207.
- Bogdan, T. J., Gilman, P. A., Lerche, I., and Howard, R. (1988). Distribution of Sunspot Umbral Areas: 1917–1982. *Astrophysical Journal*, 327:451.
- Boisse, I., Bonfils, X., and Santos, N. C. (2012). SOAP: A tool for the fast computation of photometry and radial velocity induced by stellar spots. *Astronomy & Astrophysics*, 545:A109.
- Boogaerts, T., Ahmed, F., Choi, P. M., Tschärke, B., O’Brien, J., De Loof, H., Gao, J., Thai, P., Thomas, K., Mueller, J. F., Hall, W., Covaci, A., and van Nuijs, A. L. N. (2021). Current and future perspectives for wastewater-based epidemiology as a monitoring tool for pharmaceutical use. *The Science of the Total Environment*, 789:148047.
- Borne, K., Accomazzi, A., Bloom, J., Brunner, R., Burke, D., Butler, N., Chernoff, D. F., Connolly, B., Connolly, A., Connors, A., Cutler, C., Desai, S., Djorgovski, G., Feigelson, E., Finn, L. S., Freeman, P., Graham, M., Gray, N., Graziani, C., Guinan, E. F., Hakkila, J., Jacoby, S., Jefferys, W., Kashyap, B., Kelly, B., Knuth, K., Lamb, D. Q., Lee, H., Lored, T., Mahabal, A., Mateo, M., McCollum, B., Muench, A., Pesenson, M., Petrosian, V.,

- Primini, F., Protopapas, P., Ptak, A., Quashnock, J., Raddick, M. J., Rocha, G., Ross, N., Rottler, L., Scargle, J., Siemiginowska, A., Song, I., Szalay, A., Tyson, J. A., Vestrand, T., Wallin, J., Wandelt, B., Wasserman, I. M., Way, M., Weinberg, M., Zezas, A., Anderes, E., Babu, J., Becla, J., Berger, J., Bickel, P. J., Clyde, M., Davidson, I., van Dyk, D., Eastman, T., Efron, B., Genovese, C., Gray, A., Jang, W., Kolaczyk, E. D., Kubica, J., Loh, J. M., Meng, X.-L., Moore, A., Morris, R., Park, T., Pike, R., Rice, J., Richards, J., Ruppert, D., Saito, N., Schafer, C., Stark, P. B., Stein, M., Sun, J., Wang, D., Wang, Z., Wasserman, L., Wegman, E. J., Willett, R., Wolpert, R., and Woodroffe, M. (2009). Astrominformatics: A 21st Century Approach to Astronomy. In *Astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010, page P6, eprint: arXiv:0909.3892.
- Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Dupree, A. K., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Kondo, Y., Latham, D. W., Marcy, G. W., Meibom, S., Kjeldsen, H., Lissauer, J. J., Monet, D. G., Morrison, D., Sasselov, D., Tarter, J., Boss, A., Brownlee, D., Owen, T., Buzasi, D., Charbonneau, D., Doyle, L., Fortney, J., Ford, E. B., Holman, M. J., Seager, S., Steffen, J. H., Welsh, W. F., Rowe, J., Anderson, H., Buchhave, L., Ciardi, D., Walkowicz, L., Sherry, W., Horch, E., Isaacson, H., Everett, M. E., Fischer, D., Torres, G., Johnson, J. A., Endl, M., MacQueen, P., Bryson, S. T., Dotson, J., Haas, M., Kolodziejczak, J., Van Cleve, J., Chandrasekaran, H., Twicken, J. D., Quintana, E. V., Clarke, B. D., Allen, C., Li, J., Wu, H., Tenenbaum, P., Verner, E., Bruhweiler, F., Barnes, J., and Prsa, A. (2010). Kepler Planet-Detection Mission: Introduction and First Results. *Science*, 327:977.
- Borucki, W. J., Scargle, J. D., and Hudson, H. S. (1985). Detectability of extrasolar planetary transits. *The Astrophysical Journal*, 291:852.
- Borucki, W. J. and Summers, A. L. (1984). The photometric method of detecting other planetary systems. *Icarus*, 58(1):121–134.
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92*, pages 144–152, New York, NY, USA. Association for Computing Machinery.

- Boyajian, T. S., LaCourse, D. M., Rappaport, S. A., Fabrycky, D., Fischer, D. A., Gandolfi, D., Kennedy, G. M., Korhonen, H., Liu, M. C., Moor, A., Olah, K., Vida, K., Wyatt, M. C., Best, W. M. J., Brewer, J., Ciesla, F., Csák, B., Deeg, H. J., Dupuy, T. J., Handler, G., Heng, K., Howell, S. B., Ishikawa, S. T., Kovács, J., Kozakis, T., Kriskovics, L., Lehtinen, J., Lintott, C., Lynn, S., Nespral, D., Nikbakhsh, S., Schawinski, K., Schmitt, J. R., Smith, A. M., Szabo, G., Szabo, R., Viuhö, J., Wang, J., Weiksnar, A., Bosch, M., Connors, J. L., Goodman, S., Green, G., Hoekstra, A. J., Jebson, T., Jek, K. J., Omohundro, M. R., Schwengeler, H. M., and Szewczyk, A. (2016). Planet Hunters IX. KIC 8462852 – where’s the flux?★†. *Monthly Notices of the Royal Astronomical Society*, 457(4):3988–4004.
- Breiman, L. (2001). Random forests. In *Machine Learning*, pages 5–32.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Bruno, G., Lewis, N. K., Valenti, J. A., Pagano, I., Wilson, T. J., Schlawin, E., Lothringer, J., Lanza, A. F., Fraine, J., Scandariato, G., Micela, G., and Cracchiolo, G. (2022). Hiding in plain sight: Observing planet-starspot crossings with the James Webb Space Telescope. *Monthly Notices of the Royal Astronomical Society*, 509:5030–5045.
- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., and Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4):611–622.
- Byambasuren, O., Dobler, C. C., Bell, K., Rojas, D. P., Clark, J., McLaws, M.-L., and Glasziou, P. (2021). Comparison of seroprevalence of SARS-CoV-2 infections with cumulative and imputed COVID-19 cases: Systematic review. *PloS One*, 16(4):e0248946.

- Caceres, G. A., Feigelson, E. D., Babu, G. J., Bahamonde, N., Christen, A., Bertin, K., Meza, C., and Curé, M. (2019). AutoRegressive Planet Search: Methodology. *The Astronomical Journal*, 158(2):57.
- Cambioni, S., Asphaug, E., Emsenhuber, A., Gabriel, T. S. J., Furfaro, R., and Schwartz, S. R. (2019). Realistic On-the-fly Outcomes of Planetary Collisions: Machine Learning Applied to Simulations of Giant Impacts. *The Astrophysical Journal*, 875(1):40.
- Cao, W., Wang, D., Li, J., Zhou, H., Li, L., and Li, Y. (2018). BRITS: Bidirectional Recurrent Imputation for Time Series. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 6775–6785. Curran Associates, Inc.
- Carter, A. L., Hinkley, S., Kammerer, J., Skemer, A., Biller, B. A., Leisenring, J. M., Millar-Blanchaer, M. A., Petrus, S., Stone, J. M., Ward-Duong, K., Wang, J. J., Girard, J. H., Hines, D. C., Perrin, M. D., Pueyo, L., Balmer, W. O., Bonavita, M., Bonnefoy, M., Chauvin, G., Choquet, E., Christiaens, V., Danielski, C., Kennedy, G. M., Matthews, E. C., Miles, B. E., Patapis, P., Ray, S., Rickman, E., Sallum, S., Stapelfeldt, K. R., Whiteford, N., Zhou, Y., Absil, O., Boccaletti, A., Booth, M., Bowler, B. P., Chen, C. H., Currie, T., Fortney, J. J., Grady, C. A., Greenbaum, A. Z., Henning, T., Hoch, K. K. W., Janson, M., Kalas, P., Kenworthy, M. A., Kervella, P., Kraus, A. L., Lagage, P.-O., Liu, M. C., Macintosh, B., Marino, S., Marley, M. S., Marois, C., Matthews, B. C., Mawet, D., McElwain, M. W., Metchev, S., Meyer, M. R., Molliere, P., Moran, S. E., Morley, C. V., Mukherjee, S., Pantin, E., Quirrenbach, A., Rebollido, I., Ren, B. B., Schneider, G., Vasist, M., Worthen, K., Wyatt, M. C., Briesemeister, Z. W., Bryan, M. L., Calissendorff, P., Cantalloube, F., Cugno, G., De Furio, M., Dupuy, T. J., Factor, S. M., Faherty, J. K., Fitzgerald, M. P., Franson, K., Gonzales, E. C., Hood, C. E., Howe, A. R., Kuzuhara, M., Lagrange, A.-M., Lawson, K., Lazzoni, C., Lew, B. W. P., Liu, P., Lloyd, J. L.-S. J. P., Martinez, R. A., Mazoyer, J., Quanz, S. P., Redai, J. A., Samland, M., Schlieder, J. E., Tamura, M., Tan, X., Uyama, T., Vigan, A., Vos, J. M., Wagner, K., Wolff, S. G., Ygouf, M., Zhang, X., Zhang, K., and Zhang, Z. (2022). The JWST Early Release Science Program for Direct Observations of Exoplanetary Systems I: High Contrast Imaging of the Exoplanet HIP 65426 b from 2-16 μm .
- Carter, J. A. and Winn, J. N. (2009). Parameter Estimation from Time-series Data with

- Correlated Errors: A Wavelet-based Method and its Application to Transit Light Curves. *The Astrophysical Journal*, 704:51–67.
- Carter, J. A. and Winn, J. N. (2010). The Detectability of Transit Depth Variations Due to Exoplanetary Oblateness and Spin Precession. *The Astrophysical Journal*, 716:850–856.
- Casertano, S., Lattanzi, M. G., Sozzetti, A., Spagna, A., Jancart, S., Morbidelli, R., Pannunzio, R., Pourbaix, D., and Queloz, D. (2001). Detection and measurement of planetary systems with GAIA. *Astronomy and Astrophysics*, 373(2):L21–L24.
- Casertano, S., Lattanzi, M. G., Sozzetti, A., Spagna, A., Jancart, S., Morbidelli, R., Pannunzio, R., Pourbaix, D., and Queloz, D. (2008). Double-blind test program for astrometric planet detection with Gaia. *Astronomy and Astrophysics*, 482(2):699–729.
- Cassan, A., Kubas, D., Beaulieu, J.-P., Dominik, M., Horne, K., Greenhill, J., Wambsganss, J., Menzies, J., Williams, A., Jorgensen, U. G., Udalski, A., Bennett, D. P., Albrow, M. D., Batista, V., Brilliant, S., Caldwell, J. A. R., Cole, A., Coutures, C., Cook, K. H., Dieters, S., Prester, D. D., Donatowicz, J., Fouque, P., Hill, K., Kains, N., Kane, S., Marquette, J.-B., Martin, R., Pollard, K. R., Sahu, K. C., Vinter, C., Warren, D., Watson, B., Zub, M., Sumi, T., Szymanski, M. K., Kubiak, M., Poleski, R., Soszynski, I., Ulaczyk, K., Pietrzynski, G., and Wyrzykowski, L. (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature*, 481(7380):167–169.
- Catala, C. and The PLATO Consortium (2009). PLATO: PLANetary Transits and Oscillations of stars. *Experimental Astronomy*, 23(1):329–356.
- Chalapathy, R. and Chawla, S. (2019). Deep Learning for Anomaly Detection: A Survey. *arXiv e-prints*, 1901:arXiv:1901.03407.
- Chang, Y.-C., Chang, K.-H., and Wu, G.-J. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing*, 73:914–920.
- Changeat, Q., Al-Refaie, A., Mugnai, L. V., Edwards, B., Waldmann, I. P., Pascale, E., and Tinetti, G. (2020). Alfnoor: A Retrieval Simulation of the Ariel Target List. *The Astronomical Journal*, 160:80.

- Changeat, Q., Edwards, B., Al-Refaie, A. F., Tsiaras, A., Skinner, J. W., Cho, J. Y.-K., Yip, K. H., Anisman, L., Ikoma, M., Bieger, M. F., Venot, O., Shibata, S., Waldmann, I. P., and Tinetti, G. (2022). Five key exoplanet questions answered via the analysis of 25 hot Jupiter atmospheres in eclipse. *The Astrophysical Journal Supplement Series*, 260(1):3.
- Changeat, Q. and Yip, K. H. (2022). ESA-Ariel Data Challenge NeurIPS 2022: Introduction to exo-atmospheric studies and presentation of the Ariel Big Challenge (ABC) Database.
- Charbonneau, D., Allen, L. E., Megeath, S. T., Torres, G., Alonso, R., Brown, T. M., Gilliland, R. L., Latham, D. W., Mandushev, G., O'Donovan, F. T., and Sozzetti, A. (2005). Detection of Thermal Emission from an Extrasolar Planet. *The Astrophysical Journal*, 626:523–529.
- Charbonneau, D., Brown, T. M., Latham, D. W., and Mayor, M. (2000). Detection of Planetary Transits Across a Sun-like Star. *The Astrophysical Journal*, 529(1):L45–L48.
- Charbonneau, D., Brown, T. M., Noyes, R. W., and Gilliland, R. L. (2002). Detection of an Extrasolar Planet Atmosphere*. *The Astrophysical Journal*, 568(1):377.
- Chaushev, A., Raynard, L., Goad, M. R., Eig Müller, P., Armstrong, D. J., Briegal, J. T., Burleigh, M. R., Casewell, S. L., Gill, S., Jenkins, J. S., Nielsen, L. D., Watson, C. A., West, R. G., Wheatley, P. J., Udry, S., and Vines, J. I. (2019). Classifying exoplanet candidates with convolutional neural networks: Application to the Next Generation Transit Survey. *Monthly Notices of the Royal Astronomical Society*, 488(4):5232.
- Choi, P. M., Tschärke, B. J., Donner, E., O'Brien, J. W., Grant, S. C., Kaserzon, S. L., Mackie, R., O'Malley, E., Crosbie, N. D., Thomas, K. V., and Mueller, J. F. (2018). Wastewater-based epidemiology biomarkers: Past, present and future. *TrAC Trends in Analytical Chemistry*, 105:453–469.
- Chollet, F. (2017). *Deep Learning with Python*. Manning Publications Co., USA, 1st edition.
- Christensen-Dalsgaard, J., Kjeldsen, H., Brown, T. M., Gilliland, R. L., Arentoft, T., Frandsen, S., Quirion, P. O., Borucki, W. J., Koch, D., and Jenkins, J. M. (2010). Asteroseismic investigation of known planet hosts in the kepler field. *Astrophysical Journal Letters*, 713(2 PART 2):L164–L168.

- Chu, S., Wagstaff, K., Bryden, G., and Shvartzvald, Y. (2019). Automatic Detection of Microlensing Events in the Galactic Bulge using Machine Learning Techniques. *Publications of the Astronomical Society of the Pacific*, 523:127.
- Claret, A. (2000). A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for $-5.0 \leq \log[M/H] \leq +1$, $2000 \text{ K} \leq T_{\text{eff}} \leq 50000 \text{ K}$ at several surface gravities. *Astronomy and Astrophysics*, 363:1081–1190.
- Claudi, R. (2016). Direct Imaging of Faint Companions. *Astrophysics and Space Science Library*, 428:183.
- Cobb, A. D., Himes, M. D., Soboczenski, F., Zorzan, S., O’Beirne, M. D., Güneş Baydin, A., Gal, Y., Domagal-Goldman, S. D., Arney, G. N., Angerhausen, D., and 2018 NASA FDL Astrobiology Team, I. (2019). An Ensemble of Bayesian Neural Networks for Exoplanetary Atmospheric Retrieval. *The Astronomical Journal*, 158:33.
- Cobb, O. J., Wallis, C. G. R., Mavor-Parker, A. N., Marignier, A., Price, M. A., d’Avezac, M., and McEwen, J. D. (2021). Efficient Generalized Spherical CNNs.
- Cohen, T. S., Geiger, M., Koehler, J., and Welling, M. (2018). Spherical CNNs.
- Collaboration, G. (2016). The Gaia mission. *Astronomy & Astrophysics*, 595:A1.
- Cowan, N. B., Agol, E., Meadows, V. S., Robinson, T., Livengood, T. A., Deming, D., Lisse, C. M., A’Hearn, M. F., Wellnitz, D. D., Seager, S., Charbonneau, D., and EPOXI Team (2009). Alien Maps of an Ocean-bearing World. *The Astrophysical Journal*, 700:915–923.
- Cowan, N. B. and Fujii, Y. (2018). Mapping Exoplanets. In Deeg, H. J. and Belmonte, J. A., editors, *Handbook of Exoplanets*, pages 1469–1484. Springer International Publishing, Cham.
- Cowan, N. B., Greene, T., Angerhausen, D., Batalha, N. E., Clampin, M., Colón, K., Crossfield, I. J. M., Fortney, J. J., Gaudi, B. S., Harrington, J., Iro, N., Lillie, C. F., Linsky, J. L., Lopez-Morales, M., Mandell, A. M., and Stevenson, K. B. (2015). Characterizing Transiting Planet Atmospheres through 2025. *Publications of the Astronomical Society of the Pacific*, 127:311.

- Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S. (2020). Discovering Symbolic Models from Deep Learning with Inductive Biases. *arXiv:2006.11287 [astro-ph, physics:physics, stat]*.
- Dai, X., Fu, R., Lin, Y., Li, L., and Wang, F.-Y. (2017). DeepTrend: A Deep Hierarchical Neural Network for Traffic Flow Prediction. *arXiv:1707.03213 [cs.LG]*.
- Danielski, C. and Tamanini, N. (2020). Will gravitational waves discover the first extragalactic planetary system? *International Journal of Modern Physics D*, 29:2043007.
- Dattilo, A., Vanderburg, A., Shallue, C. J., Mayo, A. W., Berlind, P., Bieryla, A., Calkins, M. L., Esquerdo, G. A., Everett, M. E., Howell, S. B., Latham, D. W., Scott, N. J., and Yu, L. (2019). Identifying Exoplanets with Deep Learning II: Two New Super-Earths Uncovered by a Neural Network in K2 Data. *The Astronomical Journal*, 157(5):169.
- Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., Pearson, C. A. B., Russell, T. W., Tully, D. C., Washburne, A. D., Wenseleers, T., Gimma, A., Waites, W., Wong, K. L. M., van Zandvoort, K., Silverman, J. D., CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz, K., Keogh, R., Eggo, R. M., Funk, S., Jit, M., Atkins, K. E., and Edmunds, W. J. (2021a). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science (New York, N.Y.)*, 372(6538):eabg3055.
- Davies, N. G., Jarvis, C. I., CMMID COVID-19 Working Group, Edmunds, W. J., Jewell, N. P., Diaz-Ordaz, K., and Keogh, R. H. (2021b). Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*, 593(7858):270–274.
- Davis, R., Harmer, D. S., and Hoffman, K. C. (1968). Search for Neutrinos from the Sun. *Physical Review Letters*, 20(21):1205–1209.
- Debosscher, J., Sarro, L. M., Aerts, C., Cuypers, J., Vandenbussche, B., Garrido, R., and Solano, E. (2007). Automated supervised classification of variable stars - I. Methodology. *Astronomy & Astrophysics*, 475(3):1159–1183.
- Debosscher, J., Sarro, L. M., López, M., Deleuil, M., Aerts, C., Auvergne, M., Baglin, A., Baudin, F., Chadid, M., Charpinet, S., Cuypers, J., Ridder, J. D., Garrido, R., Hubert, A. M., Janot-Pacheco, E., Jorda, L., Kaiser, A., Kallinger, T., Kollath, Z., Maceroni,

- C., Mathias, P., Michel, E., Moutou, C., Neiner, C., Ollivier, M., Samadi, R., Solano, E., Surace, C., Vandenbussche, B., and Weiss, W. W. (2009). Automated supervised classification of variable stars in the CoRoT programme - Method and application to the first four exoplanet fields. *Astronomy & Astrophysics*, 506(1):519–534.
- Deming, D., Harrington, J., Seager, S., and Richardson, L. J. (2006). Strong Infrared Emission from the Extrasolar Planet HD 189733b. *The Astrophysical Journal*, 644:560–564.
- Deming, D., Knutson, H., Kammer, J., Fulton, B. J., Ingalls, J., Carey, S., Burrows, A., Fortney, J. J., Todorov, K., Agol, E., Cowan, N., Desert, J.-M., Fraine, J., Langton, J., Morley, C., and Showman, A. P. (2015). Spitzer Secondary Eclipses of the Dense, Modestly-irradiated, Giant Exoplanet HAT-P-20b Using Pixel-Level Decorrelation. *The Astrophysical Journal*, 805(2):132.
- Deming, D. and Knutson, H. A. (2020). Highlights of exoplanetary science from Spitzer. *Nature Astronomy*, 4:453–466.
- Deming, D., Seager, S., Richardson, L. J., and Harrington, J. (2005). Infrared radiation from an extrasolar planet. *Nature*, 434(7034):740–743.
- Désert, J.-M., Lecavelier des Etangs, A., Hébrard, G., Sing, D. K., Ehrenreich, D., Ferlet, R., and Vidal-Madjar, A. (2009). Search for Carbon Monoxide in the Atmosphere of the Transiting Exoplanet HD 189733b. *The Astrophysical Journal*, 699:478–485.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Di Stefano, R., Berndtsson, J., Urquhart, R., Soria, R., Kashyap, V. L., Carmichael, T. W., and Imara, N. (2020). M51-ULS-1b: The First Candidate for a Planet in an External Galaxy. *arXiv:2009.08987 [astro-ph]*.
- Diaz-Cordoves, J. and Gimenez, A. (1992). A new nonlinear approximation to the limb-darkening of hot stars. *Astronomy and Astrophysics*, 259(1):227–231.

- Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. In *Proceedings of the 10th International Conference on Learning Representations*.
- Dumusque, X., Boisse, I., and Santos, N. C. (2014). SOAP 2.0: A Tool to Estimate the Photometric and Radial Velocity Variations Induced by Stellar Spots and Plages. *The Astrophysical Journal*, 796:132.
- Dumusque, X., Santos, N. C., Udry, S., Lovis, C., and Bonfils, X. (2011). Planetary detection limits taking into account stellar noise. II. Effect of stellar spot groups on radial-velocities. *Astronomy and Astrophysics*, 527:A82.
- Edwards, B., Mugnai, L., Tinetti, G., Pascale, E., and Sarkar, S. (2019). An Updated Study of Potential Targets for Ariel. *Astronomical Journal*, 157(6):242.
- Edwards, B., Rice, M., Zingales, T., Tessenyi, M., Waldmann, I., Tinetti, G., Pascale, E., Savini, G., and Sarkar, S. (2019). Exoplanet spectroscopy and photometry with the Twinkle space telescope. *Experimental Astronomy*, 47:29–63.
- Eraslan, G., Avsec, Ž., Gagneur, J., and Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. *Nature Reviews Genetics*, 20(7):389–403.
- Espinoza, N., Brahm, R., Henning, T., Jordán, A., Dorn, C., Rojas, F., Sarkis, P., Kossakowski, D., Schlecker, M., Díaz, M. R., Jenkins, J. S., Aguilera-Gomez, C., Jenkins, J. M., Twicken, J. D., Collins, K. A., Lissauer, J., Armstrong, D. J., Adibekyan, V., Barrado, D., Barros, S. C. C., Battley, M., Bayliss, D., Bouchy, F., Bryant, E. M., Cooke, B. F., Demangeon, O. D. S., Dumusque, X., Figueira, P., Giles, H., Lillo-Box, J., Lovis, C., Nielsen, L. D., Pepe, F., Pollacco, D., Santos, N. C., Sousa, S. G., Udry, S., Wheatley, P. J., Turner, O., Marmier, M., Ségransan, D., Ricker, G., Latham, D., Seager, S., Winn, J. N., Kielkopf, J. F., Hart, R., Wingham, G., Jensen, E. L. N., Hełminiak, K. G., Tokovinin, A., Briceño, C., Ziegler, C., Law, N. M., Mann, A. W., Daylan, T., Doty, J. P., Guerrero, N., Boyd, P., Crossfield, I., Morris, R. L., Henze, C. E., and Chacon, A. D. (2020). HD 213885b: A transiting 1-d-period super-Earth with an Earth-like composition around a bright ($V = 7.9$) star unveiled by TESS. *Monthly Notices of the Royal Astronomical Society*, 491:2982–2999.
- Espinoza, N. and Jordán, A. (2016). Limb darkening and exoplanets – II. Choosing the best law for optimal retrieval of transit parameters. *Monthly Notices of the Royal Astronomical Society*, 457(4):3573–3581.

- Evans, T. M., Aigrain, S., Gibson, N., Barstow, J. K., Amundsen, D. S., Tremblin, P., and Mourier, P. (2015). A uniform analysis of HD209458b Spitzer/IRAC lightcurves with Gaussian process models. *Monthly Notices of the Royal Astronomical Society*, 451(1):680–694.
- Farkas, K., Hillary, L. S., Thorpe, J., Walker, D. I., Lowther, J. A., McDonald, J. E., Malham, S. K., and Jones, D. L. (2021). Correction: Farkas et al. Concentration and Quantification of SARS-CoV-2 RNA in Wastewater Using Polyethylene Glycol-Based Concentration and qRT-PCR. *Methods Protoc.* 2021, 4, 17. *Methods and Protocols*, 4(4):82.
- Farr, W. M., Pope, B. J. S., Davies, G. R., North, T. S. H., White, T. R., Barrett, J. W., Miglio, A., Lund, M. N., Antoci, V., Fredslund Andersen, M., Grundahl, F., and Huber, D. (2018). Aldebaran b’s Temperate Past Uncovered in Planet Search Data. *The Astrophysical Journal Letters*, 865:L20.
- Fausnaugh, M. M., Burke, C. J., Caldwell, D. A., Jenkins, J. M., Smith, J. C., Tickwen, J. D., Vanderspek, R., Doty, J. P., Ting, E. B., and Villasenor, J. S. (2021). TESS Data Release Notes: Sector 42, DR60. Technical Report 20210023885, NASA.
- Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, 33(4):917–963.
- Fazio, G. G., Hora, J. L., Allen, L. E., Ashby, M. L. N., Barmby, P., Deutsch, L. K., Huang, J. S., Kleiner, S., Marengo, M., Megeath, S. T., Melnick, G. J., Pahre, M. A., Patten, B. M., Polizotti, J., Smith, H. A., Taylor, R. S., Wang, Z., Willner, S. P., Hoffmann, W. F., Pipher, J. L., Forrest, W. J., McMurty, C. W., McCreight, C. R., McKelvey, M. E., McMurray, R. E., Koch, D. G., Moseley, S. H., Arendt, R. G., Mentzell, J. E., Marx, C. T., Losch, P., Mayman, P., Eichhorn, W., Krebs, D., Jhabvala, M., Gezari, D. Y., Fixsen, D. J., Flores, J., Shakoorzadeh, K., Jungo, R., Hakun, C., Workman, L., Karpati, G., Kichak, R., Whitley, R., Mann, S., Tollestrup, E. V., Eisenhardt, P., Stern, D., Gorjian, V., Bhattacharya, B., Carey, S., Nelson, B. O., Glaccum, W. J., Lacy, M., Lowrance, P. J., Laine, S., Reach, W. T., Stauffer, J. A., Surace, J. A., Wilson, G., Wright, E. L., Hoffman, A., Domingo, G., and Cohen, M. (2004). The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *The Astrophysical Journal Supplement Series*, 154:10–17.

- Feng, L., Zhang, W., and Li, X. (2018). Monitoring of regional drug abuse through wastewater-based epidemiology—A critical review. *Science China Earth Sciences*, 61(3):239–255.
- Ferlet, R., Hobbs, L. M., and Vidal-Madjar, A. (1987). The beta Pictoris circumstellar disk. V. Time variations of the Ca II-K line. *Astronomy and Astrophysics*, 185:267–270.
- Fisher, C., Hoeijmakers, H. J., Kitzmann, D., Márquez-Neila, P., Grimm, S. L., Sznitman, R., and Heng, K. (2020). Interpreting High-resolution Spectroscopy of Exoplanets using Cross-correlations and Supervised Machine Learning. *The Astronomical Journal*, 159(5):192.
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., and Angus, R. (2017). Fast and Scalable Gaussian Process Modeling with Applications to Astronomical Time Series. *The Astronomical Journal*, 154:220.
- Franceschi, V. B., Santos, A. S., Glaeser, A. B., Paiz, J. C., Caldana, G. D., Machado Lessa, C. L., de Menezes Mayer, A., Küchle, J. G., Gazzola Zen, P. R., Vigo, A., Winck, A. T., Rotta, L. N., and Thompson, C. E. (2021). Population-based prevalence surveys during the Covid-19 pandemic: A systematic review. *Reviews in Medical Virology*, 31(4):e2200.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.
- Fulton, B. J. and Petigura, E. A. (2018). The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap. *The Astronomical Journal*, 156:264.
- Fulton, B. J., Petigura, E. A., Howard, A. W., Isaacson, H., Marcy, G. W., Cargile, P. A., Hebb, L., Weiss, L. M., Johnson, J. A., Morton, T. D., Sinukoff, E., Crossfield, I. J. M., and Hirsch, L. A. (2017). The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets. *The Astronomical Journal*, 154:109.
- Gaidos, E. and Mann, A. W. (2013). Objects in Kepler’s Mirror May be Larger Than They Appear: Bias and Selection Effects in Transiting Planet Surveys. *The Astrophysical Journal*, 762:41.

- Gao, Y., Kennedy, L., Simpson, D., and Gelman, A. (2021). Improving multilevel regression and poststratification with structured priors. *Bayesian Analysis*, 16(3):719–744.
- García, R. A., Ceillier, T., Salabert, D., Mathur, S., van Saders, J. L., Pinsonneault, M., Ballot, J., Beck, P. G., Bloemen, S., Campante, T. L., Davies, G. R., do Nascimento, J.-D., Mathis, S., Metcalfe, T. S., Nielsen, M. B., Suárez, J. C., Chaplin, W. J., Jiménez, A., and Karoff, C. (2014). Rotation and magnetism of Kepler pulsating solar-like stars. Towards asteroseismically calibrated age-rotation relations. *Astronomy and Astrophysics*, 572:A34.
- Gardner, J. P., Mather, J. C., Clampin, M., Doyon, R., Greenhouse, M. A., Hammel, H. B., Hutchings, J. B., Jakobsen, P., Lilly, S. J., Long, K. S., Lunine, J. I., McCaughrean, M. J., Mountain, M., Nella, J., Rieke, G. H., Rieke, M. J., Rix, H.-W., Smith, E. P., Sonneborn, G., Stiavelli, M., Stockman, H. S., Windhorst, R. A., and Wright, G. S. (2006). The James Webb Space Telescope. *Space Science Reviews*, 123:485–606.
- Gassendi, P. (1632). *Mercurius in Sole visus, et Venus inuisa Parisiis, anno 1631. Pro voto, & admonitione Keppleri. Per Petrum Gassendum, .. sumptibus Sebastiani Cramoisy, via Iacobaea, sub Ciconiis.*
- Gehring, J., Auli, M., Grangier, D., Yarats, D., and Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML’17*, pages 1243–1252, Sydney, NSW, Australia. JMLR.org.
- Gibson, N. P., Aigrain, S., Roberts, S., Evans, T. M., Osborne, M., and Pont, F. (2012). A Gaussian process framework for modelling instrumental systematics: Application to transmission spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 419(3):2683–2694.
- Gilbertson, C., Ford, E. B., Jones, D. E., and Stenning, D. C. (2020). Towards Extremely Precise Radial Velocities: II. A Tool For Using Multivariate Gaussian Processes to Model Stellar Activity. *arXiv:2009.01085 [astro-ph]*. arXiv: 2009.01085.
- Gilmozzi, R. and Spyromilio, J. (2007). The European Extremely Large Telescope (E-ELT). *The Messenger*, 127:11.

- Gliese, W. (1982). Detectable Perturbations in the Proper Motions of the Nearest Stars caused by Jupiter Like Companions? *Scientific Aspects of the Hipparcos Space Astrometry Mission.*, 177:193–194.
- Godines, D., Bachelet, E., Narayan, G., and Street, R. A. (2019). A Machine Learning Classifier for Microlensing in Wide-Field Surveys. *Astronomy and Computing*, 28:100298.
- Gomez Gonzalez, C. A., Absil, O., and Van Droogenbroeck, M. (2018). Supervised detection of exoplanets in high-contrast imaging sequences. *Astronomy & Astrophysics*, 613:A71.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., and Schmidhuber, J. (2009). A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868.
- Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech Recognition with Deep Recurrent Neural Networks. *arXiv:1303.5778 [cs]*.
- Green, J., Schechter, P., Baltay, C., Bean, R., Bennett, D., Brown, R., Conselice, C., Donahue, M., Fan, X., Gaudi, B. S., Hirata, C., Kalirai, J., Lauer, T., Nichol, B., Padmanabhan, N., Perlmutter, S., Rauscher, B., Rhodes, J., Roellig, T., Stern, D., Sumi, T., Tanner, A., Wang, Y., Weinberg, D., Wright, E., Gehrels, N., Sambruna, R., Traub, W., Anderson, J., Cook, K., Garnavich, P., Hillenbrand, L., Ivezić, Z., Kerins, E., Lunine, J., McDonald, P., Penny, M., Phillips, M., Rieke, G., Riess, A., van der Marel, R., Barry, R. K., Cheng, E., Content, D., Cutri, R., Goullioud, R., Grady, K., Helou, G., Jackson, C., Kruk, J., Melton, M., Peddie, C., Rioux, N., and Seiffert, M. (2012). Wide-Field InfraRed Survey Telescope (WFIRST) Final Report.
- Grillmair, C. J., Carey, S. J., Stauffer, J. R., Fisher, M. E., Olds, R., Ingalls, J. G., Krick, J. E., Glaccum, W. J., Laine, S., Lowrance, P. J., and Surace, J. A. (2012). Pointing effects and their consequences for Spitzer IRAC exoplanet observations. In *Observatory Operations: Strategies, Processes, and Systems IV*, volume 8448, pages 450–458. SPIE.
- Gupta, A. and Schlichting, H. E. (2020). Signatures of the Core-Powered Mass-Loss Mechanism in the Exoplanet Population: Dependence on Stellar Properties and Observational Predictions. *Monthly Notices of the Royal Astronomical Society*, 493(1):792–806.

- Guzmán-Mesa, A., Kitzmann, D., Fisher, C., Burgasser, A. J., Hoeijmakers, H. J., Márquez-Neila, P., Grimm, S. L., Mandell, A. M., Sznitman, R., and Heng, K. (2020). Information Content of JWST NIRSpec Transmission Spectra of Warm Neptunes. *The Astronomical Journal*, 160(1):15.
- Hamisu, A. W., Blake, I. M., Sume, G., Braka, F., Jimoh, A., Dahiru, H., Bonos, M., Dankoli, R., Mamuda Bello, A., Yusuf, K. M., Lawal, N. M., Ahmed, F., Aliyu, Z., John, D., Nwachukwu, T. E., Ayeni, M. F., Gumede-Moeletsi, N., Veltsos, P., Giri, S., Praharaj, I., Metilda, A., Bandyopadhyay, A., Diop, O. M., and Grassly, N. C. (2020). Characterizing Environmental Surveillance Sites in Nigeria and Their Sensitivity to Detect Poliovirus and Other Enteroviruses. *The Journal of Infectious Diseases*, page jiaa175.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- Hassard, F., Lundy, L., Singer, A. C., Grimsley, J., and Di Cesare, M. (2021). Innovation in wastewater near-source tracking for rapid identification of COVID-19 in schools. *The Lancet. Microbe*, 2(1):e4–e5.
- Hastie, T., Tibshirani, R., and Friedman, J. (2013). *The Elements of Statistical Learning – Data Mining, Inference and Prediction*. Springer Series in Statistics. Springer.
- Hatfield, P. W., Gaffney, J. A., Anderson, G. J., Ali, S., Antonelli, L., Başgeçmez du Pree, S., Citrin, J., Fajardo, M., Knapp, P., Kettle, B., Kustowski, B., MacDonald, M. J., Mariscal, D., Martin, M. E., Nagayama, T., Palmer, C. A. J., Peterson, J. L., Rose, S., Ruby, J. J., Shneider, C., Streeter, M. J. V., Trickey, W., and Williams, B. (2021). The data-driven future of high-energy-density physics. *Nature*, 593(7859):351–361.
- Hattori, S., Foreman-Mackey, D., Hogg, D. W., Montet, B. T., Angus, R., Pritchard, T. A., Curtis, J. L., and Schölkopf, B. (2022). The unpopular Package: A Data-driven Approach to De-trend TESS Full Frame Image Light Curves. *The Astronomical Journal*, 163(6):284.

- Hay, J. A., Kennedy-Shaffer, L., Kanjilal, S., Lennon, N. J., Gabriel, S. B., Lipsitch, M., and Mina, M. J. (2021). Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science (New York, N.Y.)*, 373(6552):eabh0635.
- Hayes, J. J. C., Kerins, E., Awiphan, S., McDonald, I., Morgan, J. S., Chuanraksasat, P., Komonjinda, S., Sanguansak, N., Kittara, P., and (SPEARNET) (2020). Optimizing exoplanet atmosphere retrieval using unsupervised machine-learning classification. *Monthly Notices of the Royal Astronomical Society*, 494(3):4492–4508.
- He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5):672–675.
- Helled, R., Werner, S., Dorn, C., Guillot, T., Ikoma, M., Ito, Y., Kama, M., Lichtenberg, T., Miguel, Y., Shorttle, O., Tackley, P. J., Valencia, D., and Vazan, A. (2022). Ariel planetary interiors White Paper. *Experimental Astronomy*, 53:323–356.
- Hellier, C., Anderson, D. R., Cameron, A. C., Delrez, L., Gillon, M., Jehin, E., Lendl, M., Maxted, P. F. L., Pepe, F., Pollacco, D., Queloz, D., Ségransan, D., Smalley, B., Smith, A. M. S., Southworth, J., Triaud, A. H. M. J., Udry, S., and West, R. G. (2014). Transiting hot Jupiters from WASP-South, Euler and TRAPPIST: WASP-95b to WASP-101b. *Monthly Notices of the Royal Astronomical Society*, 440(3):1982–1992.
- Henry, G. W., Marcy, G., Butler, R. P., and Vogt, S. S. (1999). HD 209458. *International Astronomical Union Circular*, 7307:1.
- Henry, G. W., Marcy, G. W., Butler, R. P., and Vogt, S. S. (2000). A Transiting “51 Peg-like” Planet. *The Astrophysical Journal*, 529(1):L41–L44.
- Herrero, E., Lanza, A. F., Ribas, I., Jordi, C., Cameron, A. C., and Morales, J. C. (2014). Doppler-beaming in the Kepler light curve of LHS 6343 A. *Astronomy & Astrophysics*, 563:A104.
- Herrero, E., Ribas, I., Jordi, C., Morales, J. C., Perger, M., and Rosich, A. (2016). Modelling the photosphere of active stars for planet detection and characterization. *Astronomy & Astrophysics*, 586:A131.

- Herschel, W. and Watson, W. (1781). XXXII. Account of a comet. *Philosophical Transactions of the Royal Society of London*, 71:492–501.
- Hess, V. F. (1912). über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten. *Phys. Z.*, 13:1084–1091.
- Hewamalage, H., Bergmeir, C., and Bandara, K. (2020). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427.
- Hillary, L. S., Farkas, K., Maher, K. H., Lucaci, A., Thorpe, J., Distaso, M. A., Gaze, W. H., Paterson, S., Burke, T., Connor, T. R., McDonald, J. E., Malham, S. K., and Jones, D. L. (2021). Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Research*, 200:117214.
- Himes, M. D., Harrington, J., Cobb, A. D., Baydin, A. G., Soboczenski, F., O’Beirne, M. D., Zorzan, S., Wright, D. C., Scheffer, Z., Domagal-Goldman, S. D., and Arney, G. N. (2022). Accurate Machine Learning Atmospheric Retrieval via a Neural Network Surrogate Model for Radiative Transfer. *The Planetary Science Journal*, 3(4):91.
- Hinkel, N. R., Unterborn, C., Kane, S. R., Somers, G., and Galvez, R. (2019). A Recommendation Algorithm to Predict Giant Exoplanet Host Stars Using Stellar Elemental Abundances. *The Astrophysical Journal*, 880(1):49.
- Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507.
- Hippke, M., David, T. J., Mulders, G. D., and Heller, R. (2019). Wotan: Comprehensive time-series de-trending in Python. *The Astronomical Journal*, 158(4):143.
- Hobbs, D. and Høg, E. (2017). GaiaNIR – A future all-sky astrometry mission. *Proceedings of the International Astronomical Union*, 12(S330):67–70.
- Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.
- Hoffman, M. D. and Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv:1111.4246 [cs, stat]*.

- Hoffmann, T. and Alsing, J. (2021). Faecal shedding models for SARS-CoV-2 RNA amongst hospitalised patients and implications for wastewater-based epidemiology.
- Hoffmann, T., McIntyre-Nolan, S., Bunce, J. T., Grimsley, J., Hart, A., Jacomo, L., Morvan, M., Robins, K., Wade, M. J., Watts, G., Engeli, A., and Henderson, G. (2021). Current environmental monitoring cannot constrain the effect of vaccines on SARS-CoV-2 transmission: Report for SAGE 08/04/202. Technical report, Strategic Advisory Group of Experts.
- Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel Methods in Machine Learning. *The Annals of Statistics*, 36(3):1171–1220.
- Hora, J. L., Fazio, G. G., Allen, L. E., Ashby, M. L. N., Barmby, P., Deutsch, L. K., Huang, J. S., Marengo, M., Megeath, S. T., Melnick, G. J., Pahre, M. A., Patten, B. M., Smith, H. A., Wang, Z., Willner, S. P., Hoffmann, W. F., Pipher, J. L., Forrest, W. J., McMurtry, C. W., McCreight, C. R., McKelvey, M. E., McMurray, Jr., R. E., Moseley, S. H., Arendt, R. G., Mentzell, J. E., Marx, C. T., Fixsen, D. J., Tollestrup, E. V., Eisenhardt, P. R., Stern, D., Gorjian, V., Bhattacharya, B., Carey, S. J., Glaccum, W. J., Lacy, M. D., Lowrance, P. J., Laine, S. J., Nelson, B. O., Reach, W. T., Stauffer, J. R., Surace, J. A., Wilson, G., and Wright, E. L. (2004). In-flight performance and calibration of the Infrared Array Camera (IRAC) for the Spitzer Space Telescope. In Mather, J. C., editor, *SPIE Astronomical Telescopes + Instrumentation*, page 77, USA.
- Hosenie, Z., Lyon, R., Stappers, B., Mootoovaloo, A., and McBride, V. (2020). Imbalance Learning for Variable Star Classification. *Monthly Notices of the Royal Astronomical Society*, 493(4):6050–6059.
- Houck, J. R., Roellig, T. L., van Cleve, J., Forrest, W. J., Herter, T., Lawrence, C. R., Matthews, K., Reitsema, H. J., Soifer, B. T., Watson, D. M., Weedman, D., Huisjen, M., Troeltzsch, J., Barry, D. J., Bernard-Salas, J., Blacken, C. E., Brandl, B. R., Charmandaris, V., Devost, D., Gull, G. E., Hall, P., Henderson, C. P., Higdon, S. J. U., Pirger, B. E., Schoenwald, J., Sloan, G. C., Uchida, K. I., Appleton, P. N., Armus, L., Burgdorf, M. J., Fajardo-Acosta, S. B., Grillmair, C. J., Ingalls, J. G., Morris, P. W., and Teplitz, H. I. (2004). The Infrared Spectrograph (IRS) on the Spitzer Space Telescope. *The Astrophysical Journal Supplement Series*, 154:18–24.

- Howard, A. W., Marcy, G. W., Bryson, S. T., Jenkins, J. M., Rowe, J. F., Batalha, N. M., Borucki, W. J., Koch, D. G., Dunham, E. W., Gautier, III, T. N., Van Cleve, J., Cochran, W. D., Latham, D. W., Lissauer, J. J., Torres, G., Brown, T. M., Gilliland, R. L., Buchhave, L. A., Caldwell, D. A., Christensen-Dalsgaard, J., Ciardi, D., Fressin, F., Haas, M. R., Howell, S. B., Kjeldsen, H., Seager, S., Rogers, L., Sasselov, D. D., Steffen, J. H., Basri, G. S., Charbonneau, D., Christiansen, J., Clarke, B., Dupree, A., Fabrycky, D. C., Fischer, D. A., Ford, E. B., Fortney, J. J., Tarter, J., Girouard, F. R., Holman, M. J., Johnson, J. A., Klaus, T. C., Machalek, P., Moorhead, A. V., Morehead, R. C., Ragozzine, D., Tenenbaum, P., Twicken, J. D., Quinn, S. N., Isaacson, H., Shporer, A., Lucas, P. W., Walkowicz, L. M., Welsh, W. F., Boss, A., Devore, E., Gould, A., Smith, J. C., Morris, R. L., Prsa, A., Morton, T. D., Still, M., Thompson, S. E., Mullally, F., Endl, M., and MacQueen, P. J. (2012). Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler. *The Astrophysical Journal Supplement Series*, 201:15.
- Howell, S. B., Sobeck, C., Haas, M., Still, M., Barclay, T., Mullally, F., Troeltzsch, J., Aigrain, S., Bryson, S. T., Caldwell, D., Chaplin, W. J., Cochran, W. D., Huber, D., Marcy, G. W., Miglio, A., Najita, J. R., Smith, M., Twicken, J. D., and Fortney, J. J. (2014). The K2 Mission: Characterization and Early Results. *Publications of the Astronomical Society of the Pacific*, 126(938):398.
- Hui, L. and Seager, S. (2002). Atmospheric Lensing and Oblateness Effects during an Extrasolar Planetary Transit. *The Astrophysical Journal*, 572:540–555.
- Husser, T. O., Wende-von Berg, S., Dreizler, S., Homeier, D., Reiners, A., Barman, T., and Hauschildt, P. H. (2013). A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. *Astronomy and Astrophysics*, 553:A6.
- Hutinel, M., Huijbers, P. M. C., Fick, J., Åhrén, C., Larsson, D. G. J., and Flach, C.-F. (2019). Population-level surveillance of antibiotic resistance in *Escherichia coli* through sewage analysis. *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 24(37).
- Ingalls, J. G., Krick, J. E., Carey, S. J., Laine, S., Surace, J. A., Glaccum, W. J., Grillmair, C. C., and Lowrance, P. J. (2012). Intra-pixel gain variations and high-precision photometry with the Infrared Array Camera (IRAC). In *Space Telescopes and Instrumentation*

- 2012: *Optical, Infrared, and Millimeter Wave*, volume 8442, page 84421Y. International Society for Optics and Photonics.
- Ingalls, J. G., Krick, J. E., Carey, S. J., Stauffer, J. R., Lowrance, P. J., Grillmair, C. J., Buzasi, D., Deming, D., Diamond-Lowe, H., Evans, T. M., Morello, G., Stevenson, K. B., Wong, I., Capak, P., Glaccum, W., Laine, S., Surace, J., and Storrie-Lombardi, L. (2016). Repeatability and Accuracy of Exoplanet Eclipse Depths Measured with Post-cryogenic Spitzer. *The Astronomical Journal*, 152:44.
- Ioannou, P., Karakostas, S., Astrinaki, E., Saplamidou, S., Vitsaxaki, E., Hamilos, G., Sourvinos, G., and Kofteridis, D. P. (2021). Transmission of SARS-CoV-2 variant B.1.1.7 among vaccinated health care workers. *Infectious Diseases (London, England)*, 53(11):876–879.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Ito, Y., Changeat, Q., Edwards, B., Al-Refaie, A., Tinetti, G., and Ikoma, M. (2022). Detectability of Rocky-Vapour atmospheres on super-Earths with Ariel. *Experimental Astronomy*, 53(2):357–374.
- Iyer, A. R. and Line, M. R. (2020). The Influence of Stellar Contamination on the Interpretation of Near-infrared Transmission Spectra of Sub-Neptune Worlds around M-dwarfs. *The Astrophysical Journal*, 889(2):78.
- Jarrell, R. A. (2007). Bouvard, Alexis. In Hockey, T., Trimble, V., Williams, T. R., Bracher, K., Jarrell, R. A., Marché, J. D., Ragep, F. J., Palmeri, J., and Bolt, M., editors, *The Biographical Encyclopedia of Astronomers*, pages 157–158. Springer, New York, NY.
- Jenkins, J. M., Twicken, J. D., McCauliff, S., Campbell, J., Sanderfer, D., Lung, D., Mansouri-Samani, M., Girouard, F., Tenenbaum, P., Klaus, T., Smith, J. C., Caldwell, D. A., Chacon, A. D., Henze, C., Heiges, C., Latham, D. W., Morgan, E., Swade, D., Rinehart, S., and Vanderspek, R. (2016). The TESS science processing operations center. In *Software and Cyberinfrastructure for Astronomy IV*, volume 9913, page 99133E. International Society for Optics and Photonics.

- Johns, M., McCarthy, P., Raybould, K., Bouchez, A., Farahani, A., Filgueira, J., Jacoby, G., Shectman, S., and Sheehan, M. (2012). Giant Magellan Telescope: Overview. In *Ground-Based and Airborne Telescopes IV*, volume 8444, pages 526–541. SPIE.
- Johnsen, T. K., Marley, M. S., and Gulick, V. C. (2020). A Multilayer Perceptron for Obtaining Quick Parameter Estimations of Cool Exoplanets from Geometric Albedo Spectra. *Publications of the Astronomical Society of the Pacific*, 132(1010):044502.
- Karagiorgi, G., Kasieczka, G., Kravitz, S., Nachman, B., and Shih, D. (2022). Machine learning in the search for new fundamental physics. *Nature Reviews Physics*, 4(6):399–412.
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440.
- Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., Wu, S., Smyth, C., Poupart, P., and Brubaker, M. (2019). Time2Vec: Learning a Vector Representation of Time. *arXiv:1907.05321 [cs]*.
- Kendall, M., Milsom, L., Abeler-Dörner, L., Wymant, C., Ferretti, L., Briers, M., Holmes, C., Bonsall, D., Abeler, J., and Fraser, C. (2020). Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: A preliminary analysis. *The Lancet. Digital Health*, 2(12):e658–e666.
- Kennedy, G. M., Hope, G., Hodgkin, S. T., and Wyatt, M. C. (2019). An automated search for transiting exocomets. *Monthly Notices of the Royal Astronomical Society*, 482:5587–5596.
- Kenworthy, M. A. and Mamajek, E. E. (2015). A Transiting Extrasolar Ring System: Indirect Evidence for Exosatellite Formation? In *European Planetary Science Congress 2015*, page 2.
- Keogh, E. J. and Pazzani, M. J. (2000). Scaling up dynamic time warping for datamining applications. In Ramakrishnan, R., Stolfo, S. J., Bayardo, R. J., and Parsa, I., editors, *Proc. of the 6th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 285–289. ACM.
- Kepler, J. (1871). *Joannis Kepleri Astronomi Opera Omnia*. Heyder & Zimmer.

- Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2021). Transformers in Vision: A Survey. *ACM Computing Surveys*.
- Kim, Y., Denton, C., Hoang, L., and Rush, A. M. (2017). Structured Attention Networks.
- Kinemuchi, K., Barclay, T., Fanelli, M., Pepper, J., Still, M., and Howell, S. B. (2012). Demystifying Kepler Data: A Primer for Systematic Artifact Mitigation. *Publications of the Astronomical Society of the Pacific*, 124(919):963–984.
- Kingma, D. P. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. *ICLR*.
- Kipf, T. N. and Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907 [cs, stat]*.
- Kipping, D. (2020). An Independent Analysis of the Six Recently Claimed Exomoon Candidates. *The Astrophysical Journal*, 900(2):L44.
- Knutson, H., Agol, E., Charbonneau, D., Cowan, N., Fortney, J., and Showman, A. (2008). Mapping the Atmospheres of the Smallest Transiting Exoplanets. *Spitzer Proposal*, page 50056.
- Knutson, H. A., Charbonneau, D., Allen, L. E., Fortney, J. J., Agol, E., Cowan, N. B., Showman, A. P., Cooper, C. S., and Megeath, S. T. (2007). A map of the day-night contrast of the extrasolar planet HD 189733b. *Nature*, 447:183–186.
- Koch, D. G., Borucki, W. J., Basri, G., Batalha, N. M., Brown, T. M., Caldwell, D., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Gautier III, T. N., Geary, J. C., Gilliland, R. L., Gould, A., Jenkins, J., Kondo, Y., Latham, D. W., Lissauer, J. J., Marcy, G., Monet, D., Sasselov, D., Boss, A., Brownlee, D., Caldwell, J., Dupree, A. K., Howell, S. B., Kjeldsen, H., Meibom, S., Morrison, D., Owen, T., Reitsema, H., Tarter, J., Bryson, S. T., Dotson, J. L., Gazis, P., Haas, M. R., Kolodziejczak, J., Rowe, J. F., Van Cleve, J. E., Allen, C., Chandrasekaran, H., Clarke, B. D., Li, J., Quintana, E. V., Tenenbaum, P., Twicken, J. D., and Wu, H. (2010). Kepler Mission Design, Realized Photometric Performance, and Early Science. *The Astrophysical Journal*, 713(2):L79–L86.

- Kolen, J. F. and Kremer, S. C. (2001). Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies. In *A Field Guide to Dynamical Recurrent Networks*. IEEE.
- Kollerstrom, N. (2006). An Hiatus in History: The British Claim for Neptune’s Co-Prediction, 1845–1846: Part 1. *History of Science*, 44(1):1–28.
- Kong, Z., Jiang, J. H., Zhu, Z.-H., Fahy, K. A., and Burn, R. (2021). Analyzing the Stability of Non-coplanar Circumbinary Planets using Machine Learning.
- Kopal, Z. (1950). Detailed effects of limb darkening upon light and velocity curves of close binary systems. *Harvard College Observatory Circular*, 454:1–12.
- Kovács, G., Bakos, G., and Noyes, R. W. (2005). A trend filtering algorithm for wide-field variability surveys. *Monthly Notices of the Royal Astronomical Society*, 356(2):557–567.
- Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S., and Igel, C. (2017). Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy. *IEEE Intelligent Systems*, 32(2):16–22.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.
- Lacour, S., Nowak, M., Wang, J., Pfuhl, O., Eisenhauer, F., Abuter, R., Amorim, A., Anugu, N., Benisty, M., Berger, J. P., Beust, H., Blind, N., Bonnefoy, M., Bonnet, H., Bourget, P., Brandner, W., Buron, A., Collin, C., Charnay, B., Chapron, F., Clénet, Y., Coudé Du Foresto, V., de Zeeuw, P. T., Deen, C., Dembet, R., Dexter, J., Duvert, G., Eckart, A., Förster Schreiber, N. M., Fédou, P., Garcia, P., Garcia Lopez, R., Gao, F., Gendron, E., Genzel, R., Gillessen, S., Gordo, P., Greenbaum, A., Habibi, M., Haubois, X., Haußmann, F., Henning, T., Hippler, S., Horrobin, M., Hubert, Z., Jimenez Rosales, A., Jocu, L., Kendrew, S., Kervella, P., Kolb, J., Lagrange, A.-M., Lapeyrère, V., Le Bouquin, J.-B., Léna, P., Lippa, M., Lenzen, R., Maire, A.-L., Mollière, P., Ott, T., Paumard, T., Perraut, K., Perrin, G., Pueyo, L., Rabien, S., Ramírez, A., Rau, C., Rodríguez-Coira, G., Rousset, G., Sanchez-Bermudez, J., Scheithauer, S., Schuhler, N., Straub, O., Straubmeier, C., Sturm, E., Tacconi, L. J., Vincent, F., van Dishoeck, E. F., von Fellenberg, S., Wank, I.,

- Waisberg, I., Widmann, F., Wieprecht, E., Wiest, M., Wiezorrek, E., Woillez, J., Yazici, S., Ziegler, D., and Zins, G. (2019). First direct detection of an exoplanet by optical interferometry. Astrometry and K-band spectroscopy of HR 8799 e. *Astronomy and Astrophysics*, 623:L11.
- Lam, C. and Kipping, D. (2018). A machine learns to predict the stability of circumbinary planets. *Monthly Notices of the Royal Astronomical Society*, 476(4):5692–5697.
- Lample, G. and Charton, F. (2019). Deep Learning for Symbolic Mathematics. *arXiv:1912.01412 [cs]*.
- Lanza, A. F., Boisse, I., Bouchy, F., Bonomo, A. S., and Moutou, C. (2011). Deriving the radial-velocity variations induced by stellar activity from high-precision photometry: Test on HD 189733 with simultaneous MOST/SOPHIE data. *Astronomy & Astrophysics*, 533:A44.
- Lardner, D. (1858). *Hand-Books of Natural Philosophy and Astronomy*. Blanchard and Lea.
- Lattanzi, M. G., Spagna, A., Sozzetti, A., and Casertano, S. (2000). Space-borne global astrometric surveys: The hunt for extrasolar planets. *Monthly Notices of the Royal Astronomical Society*, 317:211–224.
- LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365 – 411.
- Léger, A., Rouan, D., Schneider, J., Barge, P., Fridlund, M., Samuel, B., Ollivier, M., Guenther, E., Deleuil, M., Deeg, H. J., Auvergne, M., Alonso, R., Aigrain, S., Alapini, A., Almenara, J. M., Baglin, A., Barbieri, M., Bruntt, H., Bordé, P., Bouchy, F., Cabrera, J., Catala, C., Carone, L., Carpano, S., Csizmadia, S., Dvorak, R., Erikson, A., Ferraz-Mello, S., Foing, B., Fressin, F., Gandolfi, D., Gillon, M., Gondoin, P., Grasset, O., Guillot, T., Hatzes, A., Hébrard, G., Jorda, L., Lammer, H., Llebaria, A., Loeillet, B., Mayor, M., Mazeh, T., Moutou, C., Pätzold, M., Pont, F., Queloz, D., Rauer, H., Renner, S., Samadi,

- R., Shporer, A., Sotin, C., Tingley, B., Wuchterl, G., Adda, M., Agogu, P., Appourchaux, T., Ballans, H., Baron, P., Beaufort, T., Bellenger, R., Berlin, R., Bernardi, P., Blouin, D., Baudin, F., Bodin, P., Boisdard, L., Boit, L., Bonneau, F., Borzeix, S., Briet, R., Buey, J.-T., Butler, B., Cailleau, D., Cautain, R., Chabaud, P.-Y., Chaintreuil, S., Chiavassa, F., Costes, V., Cuna Parrho, V., de Oliveira Fialho, F., Decaudin, M., Defise, J.-M., Djalal, S., Epstein, G., Exil, G.-E., Fauré, C., Fenouillet, T., Gaboriaud, A., Gallic, A., Gamet, P., Gavalda, P., Grolleau, E., Gruneisen, R., Gueguen, L., Guis, V., Guivarc'h, V., Guterman, P., Hallouard, D., Hasiba, J., Heuripeau, F., Huntzinger, G., Hustaix, H., Imad, C., Imbert, C., Johlander, B., Jouret, M., Journoud, P., Karioty, F., Kerjean, L., Lafaille, V., Lafond, L., Lam-Trong, T., Landiech, P., Lapeyrere, V., Larqué, T., Laudet, P., Lautier, N., Lecann, H., Lefevre, L., Leruyet, B., Levacher, P., Magnan, A., Mazy, E., Mertens, F., Mesnager, J.-M., Meunier, J.-C., Michel, J.-P., Monjoin, W., Naudet, D., Nguyen-Kim, K., Orcesi, J.-L., Ottacher, H., Perez, R., Peter, G., Plasson, P., Plessier, J.-Y., Pontet, B., Pradines, A., Quentin, C., Reynaud, J.-L., Rolland, G., Rollenhagen, F., Romagnan, R., Russ, N., Schmidt, R., Schwartz, N., Sebbag, I., Sedes, G., Smit, H., Steller, M. B., Sunter, W., Surace, C., Tello, M., Tiphène, D., Toulouse, P., Ulmer, B., Vandermarcq, O., Vergnault, E., Vuillemin, A., and Zanatta, P. (2009). Transiting exoplanets from the CoRoT space mission. VIII. CoRoT-7b: The first super-Earth with measured radius. *Astronomy and Astrophysics*, 506(1):287–302.
- Levine-Tiefenbrun, M., Yelin, I., Katz, R., Herzel, E., Golan, Z., Schreiber, L., Wolf, T., Nadler, V., Ben-Tov, A., Kuint, J., Gazit, S., Patalon, T., Chodick, G., and Kishony, R. (2021). Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nature Medicine*, 27(5):790–792.
- Li, L., Yan, J., Wang, H., and Jin, Y. (2021). Anomaly Detection of Time Series With Smoothness-Inducing Sequential Variational Auto-Encoder. *IEEE transactions on neural networks and learning systems*, 32(3):1177–1191.
- Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., and Yan, X. (2020). Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. *arXiv:1907.00235 [cs, stat]*.
- Libbrecht, M. W. and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6):321–332.

- Linsky, J. L., Yang, H., France, K., Froning, C. S., Green, J. C., Stocke, J. T., and Osterman, S. N. (2010). OBSERVATIONS OF MASS LOSS FROM THE TRANSITING EXOPLANET HD 209458b. *The Astrophysical Journal*, 717(2):1291–1299.
- Lisogorskyi, M., Boro Saikia, S., Jeffers, S. V., Jones, H. R. A., Morin, J., Mengel, M., Reiners, A., Vidotto, A. A., and Petit, P. (2020). The impact of unresolved magnetic spots on high-precision radial velocity measurements. *Monthly Notices of the Royal Astronomical Society*, 497:4009–4021.
- Lund, M. N., Handberg, R., Davies, G. R., Chaplin, W. J., and Jones, C. D. (2015). K2P2—A Photometry Pipeline for the K2 Mission. *The Astrophysical Journal*, 806:30.
- Luong, M.-T., Pham, H., and Manning, C. D. (2015). Effective Approaches to Attention-based Neural Machine Translation.
- Madhusudhan, N., Bitsch, B., Johansen, A., and Eriksson, L. (2017). Atmospheric signatures of giant exoplanet formation by pebble accretion. *Monthly Notices of the Royal Astronomical Society*, 469(4):4102–4115.
- Mandel, K. and Agol, E. (2002). Analytic Light Curves for Planetary Transit Searches. *The Astrophysical Journal Letters*, 580:L171–L175.
- Márquez-Neila, P., Fisher, C., Sznitman, R., and Heng, K. (2018). Supervised machine learning for analysing spectra of exoplanetary atmospheres. *Nature Astronomy*, 2(9):719–724.
- Martínez-Palomera, J., Bloom, J. S., and Abrahams, E. S. (2020). Deep Generative Modeling of Periodic Variable Stars Using Physical Parameters. *arXiv:2005.07773 [astro-ph]*.
- Mathys, G., Kurtz, D. W., and Holdsworth, D. L. (2020). Long-period Ap stars discovered with TESS data. *Astronomy and Astrophysics*, 639:A31.
- Maxted, P. F. L. (2018). Comparison of the power-2 limb-darkening law from the STAGGER-grid to *Kepler* light curves of transiting exoplanets. *Astronomy & Astrophysics*, 616:A39.
- Mayor, M. and Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. *Nature*, 378(6555):355–359.

- McArthur, B. E., Endl, M., Cochran, W. D., Benedict, G. F., Fischer, D. A., Marcy, G. W., Butler, R. P., Naef, D., Mayor, M., Queloz, D., Udry, S., and Harrison, T. E. (2004). Detection of a Neptune-Mass Planet in the *P1* Cancri System Using the Hobby-Eberly Telescope. *The Astrophysical Journal*, 614(1):L81.
- McCauliff, S. D., Jenkins, J. M., Catanzarite, J., Burke, C. J., Coughlin, J. L., Twicken, J. D., Tenenbaum, P., Seader, S., Li, J., and Cote, M. (2015). Automatic Classification of Kepler Planetary Transit Candidates. *The Astrophysical Journal*, 806:6.
- McCullough, P. R., Crouzet, N., Deming, D., and Madhusudhan, N. (2014). Water Vapor in the Spectrum of the Extrasolar Planet HD 189733b: 1. the Transit. *The Astrophysical Journal*, 791(1):55.
- McEllistrem, M. C., Clancy, C. J., Buehrle, D. J., Lucas, A., and Decker, B. K. (2021). Single Dose of an mRNA Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) Vaccine Is Associated With Lower Nasopharyngeal Viral Load Among Nursing Home Residents With Asymptomatic Coronavirus Disease 2019 (COVID-19). *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 73(6):e1365–e1367.
- McEwen, J. D., Wallis, C. G. R., and Mavor-Parker, A. N. (2021). Scattering Networks on the Sphere for Scalable and Rotationally Equivariant Spherical CNNs.
- McQuillan, A., Mazeh, T., and Aigrain, S. (2014). Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample. *The Astrophysical Journal Supplement Series*, 211:24.
- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., and Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*, 7(7):511–516.
- Molnár, L., Szabó, R., and Plachy, E. (2016). Variable Stars with the Kepler Space Telescope. *Journal of the American Association of Variable Star Observers (JAAVSO)*, 44:168.

- Montalto, M., Boué, G., Oshagh, M., Boisse, I., Bruno, G., and Santos, N. C. (2014). Improvements on analytic modelling of stellar spots. *Monthly Notices of the RAS*, 444(2):1721–1728.
- Morello, G. (2015). A BLIND METHOD TO DETREND INSTRUMENTAL SYSTEMATICS IN EXOPLANETARY LIGHT CURVES. *The Astrophysical Journal*, 808(1):56.
- Morello, G., Claret, A., Martin-Lagarde, M., Cossou, C., Tsiaras, A., and Lagage, P.-O. (2020). The ExoTETHyS Package: Tools for Exoplanetary Transits around Host Stars. *The Astronomical Journal*, 159(2):75.
- Morello, G., Claret, A., Martin-Lagarde, M., Cossou, C., Tsiaras, A., and Lagage, P. O. (2020). The ExoTETHyS Package: Tools for Exoplanetary Transits around Host Stars. *Astronomical Journal*, 159(2):75.
- Morello, G., Waldmann, I. P., and Tinetti, G. (2016). Repeatability of Spitzer/IRAC exoplanetary eclipses with Independent Component Analysis. *The Astrophysical Journal*, 820(2):86.
- Morello, G., Waldmann, I. P., Tinetti, G., Peres, G., Micela, G., and Howarth, I. D. (2014). A New Look at Spitzer Primary Transit Observations of the Exoplanet HD 189733b. *The Astrophysical Journal*, 786:22.
- Morello, G., Zingales, T., Martin-Lagarde, M., Gastaud, R., and Lagage, P.-O. (2021). Phase-curve Pollution of Exoplanet Transmission Spectra. *The Astronomical Journal*, 161:174.
- Morvan, M., Jacomo, A. L., Souque, C., Wade, M. J., Hoffmann, T., Pouwels, K., Lilley, C., Singer, A. C., Porter, J., Evens, N. P., Walker, D. I., Bunce, J. T., Engeli, A., Grimsley, J., O'Reilly, K. M., and Danon, L. (2022a). An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nature Communications*, 13(1):4313.
- Morvan, M., Nikolaou, N., Tsiaras, A., and Waldmann, I. P. (2020). Detrending Exoplanetary Transit Light Curves with Long Short-term Memory Networks. *The Astronomical Journal*, 159(3):109.

- Morvan, M., Nikolaou, N., Yip, K. H., and Waldmann, I. (2022b). Don't Pay Attention to the Noise: Learning Self-supervised Representations of Light Curves with a Denoising Time Series Transformer.
- Morvan, M., Tsiaras, A., Nikolaou, N., and Waldmann, I. P. (2021). PyLightcurve-torch: A transit modeling package for deep learning applications in PyTorch. *Publications of the Astronomical Society of the Pacific*, 133(1021):034505.
- Mróz, P. (2020). Identifying Microlensing Events Using Neural Networks. *Acta Astronomica*, 70(3):169–180.
- Newton, I., Chittenden, N. W. L. o. S. I. N., Adee, D., Motte, A., and Hill, T. P. E. A. m. b. C.-B. (c1846). *Newton's Principia : The Mathematical Principles of Natural Philosophy*. New-York : Published by Daniel Adee.
- Nikolaou, N., Waldmann, I. P., Tsiaras, A., Morvan, M., Edwards, B., Hou Yip, K., Tinetti, G., Sarkar, S., Dawson, J. M., Borisov, V., Kasneci, G., Petkovic, M., Stepisnik, T., Al-Ubaidi, T., Bailey, R. L., Granitzer, M., Julka, S., Kern, R., Ofner, P., Wagner, S., Heppe, L., Bunse, M., and Morik, K. (2020). Lessons Learned from the 1st ARIEL Machine Learning Challenge: Correcting Transiting Exoplanet Light Curves for Stellar Spots. *arXiv e-prints*, 2010:arXiv:2010.15996.
- Nikolov, N., Sing, D. K., Fortney, J. J., Goyal, J. M., Drummond, B., Evans, T. M., Gibson, N. P., De Mooij, E. J. W., Rustamkulov, Z., Wakeford, H. R., Smalley, B., Burgasser, A. J., Hellier, C., Helling, C., Mayne, N. J., Madhusudhan, N., Kataria, T., Baines, J., Carter, A. L., Ballester, G. E., Barstow, J. K., McCleery, J., and Spake, J. J. (2018). An absolute sodium abundance for a cloud-free 'hot Saturn' exoplanet. *Nature*, 557(7706):526–529.
- Nikolov, N., Sing, D. K., Pont, F., Burrows, A. S., Fortney, J. J., Ballester, G. E., Evans, T. M., Huitson, C. M., Wakeford, H. R., Wilson, P. A., Aigrain, S., Deming, D., Gibson, N. P., Henry, G. W., Knutson, H., Lecavelier des Etangs, A., Showman, A. P., Vidal-Madjar, A., and Zahnle, K. (2013). Hubble Space Telescope hot Jupiter transmission spectral survey: a detection of Na and strong optical absorption in HAT-P-1b. *Monthly Notices of the Royal Astronomical Society*, 437(1):46–66. _eprint: <https://academic.oup.com/mnras/article-pdf/437/1/46/18466285/stt1859.pdf>.

- Nixon, M. C. and Madhusudhan, N. (2020). Assessment of supervised machine learning for atmospheric retrieval of exoplanets. *Monthly Notices of the Royal Astronomical Society*, 496(1):269–281.
- Noushad, M. and Al-Saqqaf, I. S. (2021). COVID-19 case fatality rates can be highly misleading in resource-poor and fragile nations: The case of Yemen. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 27(4):509–510.
- O'Reilly, K. M., Allen, D. J., Fine, P., and Asghar, H. (2020). The challenges of informative wastewater sampling for SARS-CoV-2 must be met: Lessons from polio eradication. *The Lancet. Microbe*, 1(5):e189–e190.
- Ort, C., Lawrence, M. G., Reungoat, J., and Mueller, J. F. (2010). Sampling for PPCPs in wastewater systems: Comparison of different sampling modes and optimization strategies. *Environmental Science & Technology*, 44(16):6289–6296.
- Osborn, H. P., Ansdell, M., Ioannou, Y., Sasdelli, M., Angerhausen, D., Caldwell, D. A., Jenkins, J. M., Räissi, C., and Smith, J. C. (2019). Rapid Classification of TESS Planet Candidates with Convolutional Neural Networks. *arXiv:1902.08544 [astro-ph]*.
- Ourmazd, A. (2020). Science in the age of machine learning. *Nature Reviews Physics*, 2(7):342–343.
- Paczynski, B. (1986). Gravitational Microlensing at Large Optical Depth. *The Astrophysical Journal*, 301:503.
- Parikh, A., Täckström, O., Das, D., and Uszkoreit, J. (2016). A Decomposable Attention Model for Natural Language Inference. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2249–2255, Austin, Texas. Association for Computational Linguistics.
- Passegger, V. M., Bello-García, A., Ordieres-Meré, J., Caballero, J. A., Schweitzer, A., González-Marcos, A., Ribas, I., Reiners, A., Quirrenbach, A., Amado, P. J., Azzaro, M., Bauer, F. F., Béjar, V. J. S., Cortés-Contreras, M., Dreizler, S., Hatzes, A. P., Henning, T., Jeffers, S. V., Kaminski, A., Kürster, M., Lafarga, M., Marfil, E., Montes, D., Morales, J. C., Nagel, E., Sarro, L. M., Solano, E., Tabernero, H. M., and Zechmeister,

- M. (2020). The CARMENES search for exoplanets around M dwarfs – A deep learning approach to determine fundamental parameters of target stars. *Astronomy & Astrophysics*, 642(A22):16.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv:1912.01703 [cs, stat]*.
- Pearson, K. A. (2019). A search for multi-planet systems with TESS using a Bayesian N-body retrieval and machine learning. *The Astronomical Journal*, 158(6):243.
- Pearson, K. A., Palafox, L., and Griffith, C. A. (2018). Searching for exoplanets using artificial intelligence. *Monthly Notices of the Royal Astronomical Society*, 474(1):478–491.
- Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., Weinberger, D. M., Arnold, W., and Omer, S. B. (2020). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology*, 38(10):1164–1167.
- Penn, R., Ward, B. J., Strande, L., and Maurer, M. (2018). Review of synthetic human faeces and faecal sludge for sanitation and wastewater research. *Water Research*, 132:222–240.
- Perryman, M., Hartman, J., Bakos, G. Á., and Lindegren, L. (2014). Astrometric Exoplanet Detection with Gaia. *The Astrophysical Journal*, 797:14.
- Pham, D. and Kaltenegger, L. (2021). Color classification of Earth-like planets with machine learning. *Monthly Notices of the Royal Astronomical Society*, 504(4):6106–6116.
- Pinhas, A., Madhusudhan, N., Gandhi, S., and MacDonald, R. J. (2019). H₂O abundances and cloud properties in ten hot giant exoplanets. *Monthly Notices of the Royal Astronomical Society*, 482(2):1485–1498.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3):21–45.

- Politsch, C. A., Cisewski-Kehe, J., Croft, R. A. C., and Wasserman, L. (2020a). Trend Filtering – II. Denoising Astronomical Signals with Varying Degrees of Smoothness. *Monthly Notices of the Royal Astronomical Society*, 492(3):4019–4032.
- Politsch, C. A., Cisewski-Kehe, J., Croft, R. A. C., and Wasserman, L. (2020b). Trend filtering – I. A modern statistical tool for time-domain astronomy and astronomical spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 492(3):4005–4018.
- Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D. L., Singer, A. C., Graham, D. W., and Romalde, J. L. (2020). Making waves: Wastewater-based epidemiology for COVID-19 - approaches and challenges for surveillance and prediction. *Water Research*, 186:116404.
- Pont, F., Zucker, S., and Queloz, D. (2006). The effect of red noise on planetary transit detection. *Monthly Notices of the Royal Astronomical Society*, 373(1):231–242.
- Pouwels, K. B., House, T., Pritchard, E., Robotham, J. V., Birrell, P. J., Gelman, A., Vihta, K.-D., Bowers, N., Boreham, I., Thomas, H., Lewis, J., Bell, I., Bell, J. I., Newton, J. N., Farrar, J., Diamond, I., Benton, P., Walker, A. S., and COVID-19 Infection Survey Team (2021). Community prevalence of SARS-CoV-2 in England from April to November, 2020: Results from the ONS Coronavirus Infection Survey. *The Lancet. Public Health*, 6(1):e30–e38.
- Pritchard, E., Matthews, P. C., Stoesser, N., Eyre, D. W., Gethings, O., Vihta, K.-D., Jones, J., House, T., VanSteenHouse, H., Bell, I., Bell, J. I., Newton, J. N., Farrar, J., Diamond, I., Rourke, E., Studley, R., Crook, D., Peto, T. E. A., Walker, A. S., and Pouwels, K. B. (2021). Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nature Medicine*, 27(8):1370–1378.
- Rabus, M., Alonso, R., Belmonte, J. A., Deeg, H. J., Gilliland, R. L., Almenara, J. M., Brown, T. M., Charbonneau, D., and Mandushev, G. (2009). A cool starspot or a second transiting planet in the TrES-1 system? *Astronomy & Astrophysics*, 494(1):391–397.
- Rackham, B. V., Apai, D., and Giampapa, M. S. (2018a). The Transit Light Source Effect: False Spectral Features and Incorrect Densities for M-dwarf Transiting Planets. *The Astrophysical Journal*, 853:122.

- Rackham, B. V., Apai, D., and Giampapa, M. S. (2018b). The Transit Light Source Effect: False Spectral Features and Incorrect Densities for M-dwarf Transiting Planets. *The Astrophysical Journal*, 853(2):122.
- Rackham, B. V., Apai, D., and Giampapa, M. S. (2019). The Transit Light Source Effect. II. The Impact of Stellar Heterogeneity on Transmission Spectra of Planets Orbiting Broadly Sun-like Stars. *The Astronomical Journal*, 157(3):96.
- Radovic, A., Williams, M., Rousseau, D., Kagan, M., Bonacorsi, D., Himmel, A., Aurisano, A., Terao, K., and Wongjirad, T. (2018). Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716):41–48.
- Rafikov, R. R. (2009). Stellar proper motion and the timing of planetary transits. *The Astrophysical Journal*, 700(2):965–970.
- Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., and Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Research*, 181:115942.
- Rao, S., Mahabal, A., Rao, N., and Raghavendra, C. (2021). Nigraha: Machine-learning-based pipeline to identify and evaluate planet candidates from TESS. *Monthly Notices of the Royal Astronomical Society*, 502:2845–2858.
- Rappaport, S., Vanderburg, A., Jacobs, T., LaCourse, D., Jenkins, J., Kraus, A., Rizzuto, A., Latham, D. W., Bieryla, A., Lazarevic, M., and Schmitt, A. (2018). Likely transiting exocomets detected by Kepler. *Monthly Notices of the Royal Astronomical Society*, 474:1453–1468.
- Ras, G., Xie, N., van Gerven, M., and Doran, D. (2021). Explainable Deep Learning: A Field Guide for the Uninitiated.
- Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Richterich, P. (2020). Severe underestimation of COVID-19 case numbers: Effect of epidemic growth rate and test restrictions.
- Ricker, D. W. (2003). *Echo Signal Processing*. Kluwer Academic Publishers, USA.

- Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. A., Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J. M., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., and Villaseñor, J. (2015). Transiting Exoplanet Survey Satellite (TESS). *Journal of Astronomical Telescopes, Instruments, and Systems*, 1:014003.
- Rieke, G. H., Young, E. T., Cadien, J., Engelbracht, C. W., Gordon, K. D., Kelly, D. M., Low, F. J., Misselt, K. A., Morrison, J. E., Muzerolle, J., Rivlis, G., Stansberry, J. A., Beeman, J. W., Haller, E. E., Frayer, D. T., Latter, W. B., Noriega-Crespo, A., Padgett, D. L., Hines, D. C., Bean, J. D., Burmester, W., Heim, G. B., Glenn, T., Ordonez, R., Schwenker, J. P., Siewert, S., Strecker, D. W., Tennant, S., Troeltzsch, J. R., Unruh, B., Warden, R. M., Ade, P. A. R., Alonso-Herrero, A., Blaylock, M., Dole, H., Egami, E., Hinz, J. L., LeFloch, E., Papovich, C., Perez-Gonzalez, P. G., Rieke, M. J., Smith, P. S., Su, K. Y. L., Bennett, L., Henderson, D., Lu, N., Masci, F. J., Pesenson, M., Rebull, L., Rho, J., Keene, J., Stolovy, S., Wachter, S., Wheaton, W., Richards, P. L., Garner, H. W., Hegge, M., Henderson, M. L., MacFeely, K. I., Michika, D., Miller, C. D., Neitenbach, M., Winghart, J., Woodruff, R., Arens, E., Beichman, C. A., Gaalema, S. D., Iii, T. N. G., Lada, C. J., Mould, J., Neugebauer, G. X., and Stapelfeldt, K. R. (2004). On-orbit performance of the MIPS instrument. In *Optical, Infrared, and Millimeter Space Telescopes*, volume 5487, pages 50–61. SPIE.
- Rose, C., Parker, A., Jefferson, B., and Cartmell, E. (2015). The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology. *Critical Reviews in Environmental Science and Technology*, 45(17):1827–1879.
- Rosenblatt, F. (1971). A two-color photometric method for detection of extra-solar planetary systems. *Icarus*, 14(1):71–93.
- Rosich, A., Herrero, E., Mallonn, M., Ribas, I., Morales, J. C., Perger, M., Anglada-Escudé,

- G., and Granzer, T. (2020). Correcting for chromatic stellar activity effects in transits with multiband photometric monitoring: Application to WASP-52. *Astronomy and Astrophysics*, 641:A82.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533.
- Russell, T. W., Wu, J. T., Clifford, S., Edmunds, W. J., Kucharski, A. J., Jit, M., and Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group (2021). Effect of internationally imported cases on internal spread of COVID-19: A mathematical modelling study. *The Lancet. Public Health*, 6(1):e12–e20.
- Saba, A., Tsiaras, A., Morvan, M., Thompson, A., Changeat, Q., Edwards, B., Jolly, A., Waldmann, I., and Tinetti, G. (2022). The Transmission Spectrum of WASP-17 b From the Optical to the Near-infrared Wavelengths: Combining STIS, WFC3, and IRAC Data Sets. *The Astronomical Journal*, 164(1):2.
- Sah, P., Fitzpatrick, M. C., Zimmer, C. F., Abdollahi, E., Juden-Kelly, L., Moghadas, S. M., Singer, B. H., and Galvani, A. P. (2021). Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 118(34):e2109229118.
- Salinas, D., Flunkert, V., and Gasthaus, J. (2017). DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *arXiv:1704.04110 [cs, stat]*.
- Sanders, G. H. (2013). The Thirty Meter Telescope (TMT): An International Observatory. *Journal of Astrophysics and Astronomy*, 34:81–86.
- Sarkar, S. (2017). *Exoplanet transit spectroscopy: development and application of a generic time domain simulator [PhD Thesis]*. Cardiff University.
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2009). The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1):61–80.
- Schanche, N., Cameron, A. C., Hébrard, G., Nielsen, L., Triaud, A. H. M. J., Almenara, J. M., Alsubai, K. A., Anderson, D. R., Armstrong, D. J., Barros, S. C. C., Bouchy, F., Boumis, P., Brown, D. J. A., Faedi, F., Hay, K., Hebb, L., Kiefer, F., Mancini, L., Maxted, P. F. L.,

- Palle, E., Pollacco, D. L., Queloz, D., Smalley, B., Udry, S., West, R., and Wheatley, P. J. (2019). Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys. *Monthly Notices of the Royal Astronomical Society*, 483(4):5534–5547.
- Scharf, C. A. (2007). Exoplanet Transit Parallax. *The Astrophysical Journal*, 661(2):1218–1221.
- Schuster, M. and Paliwal, K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.
- Schwarzschild, K. (1906). On the equilibrium of the Sun’s atmosphere. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Math.-phys. Klasse*, 195:41–53.
- Sen, R., Yu, H.-F., and Dhillon, I. (2019). Think Globally, Act Locally: A Deep Neural Network Approach to High-Dimensional Time Series Forecasting. *arXiv:1905.03806 [cs, stat]*.
- Shallue, C. J. and Vanderburg, A. (2018). Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. *The Astronomical Journal*, 155(2):94.
- Siegelmann, H. T. and Sontag, E. D. (1995). On the Computational Power of Neural Nets. *Journal of Computer and System Sciences*, 50(1):132–150.
- Sing, D. K., Fortney, J. J., Nikolov, N., Wakeford, H. R., Kataria, T., Evans, T. M., Aigrain, S., Ballester, G. E., Burrows, A. S., Deming, D., and al, e. (2015). A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature*, 529(7584):59–62. Publisher: Springer Science and Business Media LLC.
- Sing, D. K., Fortney, J. J., Nikolov, N., Wakeford, H. R., Kataria, T., Evans, T. M., Aigrain, S., Ballester, G. E., Burrows, A. S., Deming, D., Désert, J.-M., Gibson, N. P., Henry, G. W., Huitson, C. M., Knutson, H. A., des Etangs, A. L., Pont, F., Showman, A. P., Vidal-Madjar, A., Williamson, M. H., and Wilson, P. A. (2016). A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature*, 529(7584):59–62.

- Smith, J. C., Stumpe, M. C., Van Cleve, J. E., Jenkins, J. M., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Kolodziejczak, J. J., McCauliff, S. D., Morris, R. L., and Twicken, J. D. (2012). Kepler Presearch Data Conditioning II - A Bayesian Approach to Systematic Error Correction. *Publications of the Astronomical Society of the Pacific*, 124(919):1000–1014.
- Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 464–472. IEEE.
- Soboczenski, F., Himes, M. D., O’Beirne, M. D., Zorzan, S., Baydin, A. G., Cobb, A. D., Gal, Y., Angerhausen, D., Mascaro, M., Arney, G. N., and Domagal-Goldman, S. D. (2018). Bayesian Deep Learning for Exoplanet Atmospheric Retrieval. *arXiv:1811.03390 [astro-ph]*.
- Southworth, J., Hinse, T. C., Dominik, M., Fang, X.-S., Harpsøe, K., Jørgensen, U. G., Kerins, E., Liebig, C., Mancini, L., Skottfelt, J., Anderson, D. R., Smalley, B., Tregloan-Reed, J., Wertz, O., Alsubai, K. A., Bozza, V., Calchi Novati, S., Dreizler, S., Gu, S.-H., Hundertmark, M., Jessen-Hansen, J., Kains, N., Kjeldsen, H., Lund, M. N., Lundkvist, M., Mathiasen, M., Penny, M. T., Rahvar, S., Ricci, D., Scarpetta, G., Snodgrass, C., and Surdej, J. (2012). High-precision photometry by telescope defocusing – IV. Confirmation of the huge radius of WASP-17 b. *Monthly Notices of the Royal Astronomical Society*, 426(2):1338–1348.
- Sozzetti, A., Casertano, S., Lattanzi, M. G., and Spagna, A. (2001). Detection and measurement of planetary systems with GAIA. *Astronomy & Astrophysics*, 373(3):L21–L24.
- Sozzetti, A., Giacobbe, P., Lattanzi, M. G., Micela, G., Morbidelli, R., and Tinetti, G. (2014). Astrometric detection of giant planets around nearby M dwarfs: The Gaia potential. *Monthly Notices of the Royal Astronomical Society*, 437(1):497–509.
- Spitzer Science Center (2007). *Spitzer Space Telescope Observer’s Manual*.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.
- Stevenson, K. B., Harrington, J., Fortney, J. J., Lored, T. J., Hardy, R. A., Nymeyer, S., Bowman, W. C., Cubillos, P., Bowman, M. O., and Hardin, M. (2012). Transit and Eclipse

- Analyses of the Exoplanet HD 149026b Using BLISS Mapping. *The Astrophysical Journal*, 754:136.
- Struve, O. (1952). Proposal for a project of high-precision stellar radial velocity work. *The Observatory*, 72:199–200.
- Stumpe, M. C., Smith, J. C., Van Cleve, J. E., Twicken, J. D., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Jenkins, J. M., Kolodziejczak, J. J., McCauliff, S. D., and Morris, R. L. (2012). Kepler Presearch Data Conditioning I - Architecture and Algorithms for Error Correction in Kepler Light Curves. *Publications of the Astronomical Society of the Pacific*, 124(919):985–999.
- Sun, F.-K., Lang, C. I., and Boning, D. S. (2021). Adjusting for Autocorrelated Errors in Neural Networks for Time Series Regression and Forecasting. *arXiv:2101.12578 [cs, stat]*.
- Sutton, R. S. and Barto, A. G. (2015). *Reinforcement Learning: An Introduction*. The MIT Press.
- Tamanini, N. and Danielski, C. (2019). The gravitational-wave detection of exoplanets orbiting white dwarf binaries using LISA. *Nature Astronomy*, 3(9):858–866.
- Tamayo, D., Silburt, A., Valencia, D., Menou, K., Ali-Dib, M., Petrovich, C., Huang, C. X., Rein, H., van Laerhoven, C., Paradise, A., Obertas, A., and Murray, N. (2016). A Machine Learns to Predict the Stability of Tightly Packed Planetary Systems. *The Astrophysical Journal*, 832:L22.
- Tasker, E. J., Laneuville, M., and Guttenberg, N. (2020). Estimating Planetary Mass with Deep Learning. *The Astronomical Journal*, 159(2):41.
- Teachey, A. and Kipping, D. M. (2018). Evidence for a large exomoon orbiting Kepler-1625b. *Science Advances*, 4(10):eaav1784.
- Thatte, A., Deroo, P., and Swain, M. R. (2010). Selective principal component extraction and reconstruction: A novel method for ground based exoplanet spectroscopy. *Astronomy & Astrophysics*, 523:A35.

- The Lancet, n. (2020). COVID-19: Fighting panic with information. *Lancet (London, England)*, 395(10224):537.
- Thompson, S. E., Mullally, F., Coughlin, J., Christiansen, J. L., Henze, C. E., Haas, M. R., and Burke, C. J. (2015). A MACHINE LEARNING TECHNIQUE TO IDENTIFY TRANSIT SHAPED SIGNALS. *The Astrophysical Journal*, 812(1):46.
- Timpe, M. L., Han Veiga, M., Knabenhans, M., Stadel, J., and Marelli, S. (2020). Machine learning applied to simulations of collisions between rotating, differentiated planets. *Computational Astrophysics and Cosmology*, 7(1):2.
- Tinetti, G., Drossart, P., Eccleston, P., Hartogh, P., Heske, A., Leconte, J., Micela, G., Ollivier, M., Pilbratt, G., Puig, L., Turrini, D., Vandenbussche, B., Wolkenberg, P., Pascale, E., Beaulieu, J.-P., Güdel, M., Min, M., Rataj, M., Ray, T., Ribas, I., Barstow, J., Bowles, N., Coustenis, A., du Foresto, V. C., Decin, L., Encrenaz, T., Forget, F., Friswell, M., Griffin, M., Lagage, P. O., Malaguti, P., Moneti, A., Morales, J. C., Pace, E., Rocchetto, M., Sarkar, S., Selsis, F., Taylor, W., Tennyson, J., Venot, O., Waldmann, I. P., Wright, G., Zingales, T., and Zapatero-Orsio, M. R. (2016). The science of ARIEL (Atmospheric Remote-sensing Infrared Exoplanet Large-survey). In MacEwen, H. A., Fazio, G. G., Lystrup, M., Batalha, N., Siegler, N., and Tong, E. C., editors, *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, volume 9904, pages 658 – 667. International Society for Optics and Photonics, SPIE.
- Tinetti, G., Eccleston, P., Haswell, C., Lagage, P.-O., Leconte, J., Lüftinger, T., Micela, G., Min, M., Pilbratt, G., Puig, L., Swain, M., Testi, L., Turrini, D., Vandenbussche, B., Orsio, M. R. Z., Aret, A., Beaulieu, J.-P., Buchhave, L., Ferus, M., Griffin, M., Guedel, M., Hartogh, P., Machado, P., Malaguti, G., Pallé, E., Rataj, M., Ray, T., Ribas, I., Szabó, R., Tan, J., Werner, S., Ratti, F., Scharnberg, C., Salvignol, J.-C., Boudin, N., Halain, J.-P., Haag, M., Crouzet, P.-E., Kohley, R., Symonds, K., Renk, F., Caldwell, A., Abreu, M., Alonso, G., Amiaux, J., Berthé, M., Bishop, G., Bowles, N., Carmona, M., Coffey, D., Colomé, J., Crook, M., Désjonqueres, L., Díaz, J. J., Drummond, R., Focardi, M., Gómez, J. M., Holmes, W., Krijger, M., Kovacs, Z., Hunt, T., Machado, R., Morgante, G., Ollivier, M., Ottensamer, R., Pace, E., Pagano, T., Pascale, E., Pearson, C., Pedersen, S. M., Pniel, M., Roose, S., Savini, G., Stamper, R., Szivovicza, P., Szoke, J., Tosh, I., Vilardell, F., Barstow, J., Borsato, L., Casewell, S., Changeat, Q., Charnay, B., Civiš, S.,

- du Foresto, V. C., Coustenis, A., Cowan, N., Danielski, C., Demangeon, O., Drossart, P., Edwards, B. N., Gilli, G., Encrenaz, T., Kiss, C., Kokori, A., Ikoma, M., Morales, J. C., Mendonça, J., Moneti, A., Mugnai, L., Muñoz, A. G., Helled, R., Kama, M., Miguel, Y., Nikolaou, N., Pagano, I., Panic, O., Rengel, M., Rickman, H., Rocchetto, M., Sarkar, S., Selsis, F., Tennyson, J., Tsiaras, A., Venot, O., Vida, K., Waldmann, I. P., Yurchenko, S., Szabó, G., Zellem, R., Al-Refaie, A., Alvarez, J. P., Anisman, L., Arhancet, A., Ateca, J., Baeyens, R., Barnes, J. R., Bell, T., Benatti, S., Biazzo, K., Błęcka, M., Bonomo, A. S., Bosch, J., Bossini, D., Bourgalais, J., Brienza, D., Brucalassi, A., Bruno, G., Caines, H., Calcutt, S., Campante, T., Canestrari, R., Cann, N., Casali, G., Casas, A., Cassone, G., Cara, C., Carmona, M., Carone, L., Carrasco, N., Changeat, Q., Chioetto, P., Cortecchia, F., Czupalla, M., Chubb, K. L., Ciaravella, A., Claret, A., Claudi, R., Codella, C., Comas, M. G., Cracchiolo, G., Cubillos, P., Da Peppo, V., Decin, L., Dejabrun, C., Delgado-Mena, E., Di Giorgio, A., Diolaiti, E., Dorn, C., Doublier, V., Doumayrou, E., Dransfield, G., Dumaye, L., Dunford, E., Escobar, A. J., Van Eylen, V., Farina, M., Fedele, D., Fernández, A., Fleury, B., Fonte, S., Fontignie, J., Fossati, L., Funke, B., Galy, C., Garai, Z., García, A., García-Rigo, A., Garufi, A., Sacco, G. G., Giacobbe, P., Gómez, A., Gonzalez, A., Gonzalez-Galindo, F., Grassi, D., Griffith, C., Guarcello, M. G., Goujon, A., Gressier, A., Grzegorzczak, A., Guillot, T., Guilluy, G., Hargrave, P., Hellin, M.-L., Herrero, E., Hills, M., Horeau, B., Ito, Y., Jessen, N. C., Kabath, P., Kálmán, S., Kawashima, Y., Kimura, T., Knížek, A., Kreidberg, L., Kruid, R., Kruijssen, D. J. M., Kubelík, P., Lara, L., Lebonnois, S., Lee, D., Lefevre, M., Lichtenberg, T., Locci, D., Lombini, M., Lopez, A. S., Lorenzani, A., MacDonald, R., Magrini, L., Maldonado, J., Marcq, E., Migliorini, A., Modirrousta-Galian, D., Molaverdikhani, K., Molinari, S., Mollière, P., Moreau, V., Morello, G., Morinaud, G., Morvan, M., Moses, J. I., Mouzali, S., Nakhjiri, N., Naponiello, L., Narita, N., Nascimbeni, V., Nikolaou, A., Noce, V., Oliva, F., Palladino, P., Papageorgiou, A., Parmentier, V., Peres, G., Pérez, J., Perez-Hoyos, S., Perger, M., Pestellini, C. C., Petralia, A., Philippon, A., Piccialli, A., Pignatari, M., Piotto, G., Podio, L., Polenta, G., Preti, G., Pribulla, T., Puertas, M. L., Rainer, M., Reess, J.-M., Rimmer, P., Robert, S., Rosich, A., Rossi, L., Rust, D., Saleh, A., Sanna, N., Schisano, E., Schreiber, L., Schwartz, V., Scippa, A., Seli, B., Shibata, S., Simpson, C., Shorttle, O., Skaf, N., Skup, K., Sobiecki, M., Sousa, S., Sozzetti, A., Šponer, J., Steiger, L., Tanga, P., Tackley, P., Taylor, J., Tecza, M., Terenzi, L., Tremblin, P., Tozzi, A., Triaud, A., Trompet, L., Tsai, S.-M., Tsantaki,

- M., Valencia, D., Vandaele, A. C., Van der Swaelmen, M., Vardan, A., Vasisht, G., Vazan, A., Del Vecchio, C., Waltham, D., Wawer, P., Widemann, T., Wolkenberg, P., Yip, G. H., Yung, Y., Zilinskas, M., Zingales, T., and Zuppella, P. (2021). Ariel: Enabling planetary science across light-years. *arXiv:2104.04824 [astro-ph]*.
- Tingley, B. and Sackett, P. D. (2005). A Photometric Diagnostic to Aid in the Identification of Transiting Extra-Solar Planets. *The Astrophysical Journal*, 627(2):1011–1018.
- Touzani, S., Granderson, J., and Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158:1533–1543.
- Tran, H. N., Le, G. T., Nguyen, D. T., Juang, R.-S., Rinklebe, J., Bhatnagar, A., Lima, E. C., Iqbal, H. M. N., Sarmah, A. K., and Chao, H.-P. (2021). SARS-CoV-2 coronavirus in water and wastewater: A critical review about presence and concern. *Environmental Research*, 193:110265.
- Tsiaras, A. (2017). *Towards a Population of Exoplanetary Atmospheres*. Doctoral, UCL (University College London).
- Tsiaras, A., Rocchetto, M., Waldmann, I. P., Venot, O., Varley, R., Morello, G., Damiano, M., Tinetti, G., Barton, E. J., Yurchenko, S. N., and Tennyson, J. (2016). Detection of an Atmosphere Around the Super-Earth 55 Cancri e. *The Astrophysical Journal*, 820(2):99.
- Tsiaras, A., Waldmann, I. P., Tinetti, G., Tennyson, J., and Yurchenko, S. N. (2019). Water vapour in the atmosphere of the habitable-zone eight-Earth-mass planet K2-18 b. *Nature Astronomy*, 3(12):1086–1091.
- Tsiaras, A., Waldmann, I. P., Zingales, T., Rocchetto, M., Morello, G., Damiano, M., Karpouzias, K., Tinetti, G., McKemmish, L. K., Tennyson, J., and Yurchenko, S. N. (2018). A Population Study of Gaseous Exoplanets. *The Astronomical Journal*, 155(4):156.
- Turrini, D., Schisano, E., Fonte, S., Molinari, S., Politi, R., Fedele, D., Panic, O., Kama, M., Changeat, Q., and Tinetti, G. (2021). Tracing the formation history of giant planets in protoplanetary disks with Carbon, Oxygen, Nitrogen and Sulphur. *The Astrophysical Journal*, 909(1):40.

- Valencia, D., Paracha, E., and Jackson, A. P. (2019). Can a Machine Learn the Outcome of Planetary Collisions? *The Astrophysical Journal*, 882(1):35.
- van Belle, G. T., von Braun, K., Boyajian, T., and Schaefer, G. (2012). Direct Imaging of Planet Transit Events. *Proceedings of the International Astronomical Union*, 8(S293):378–381.
- Van Helden, A. (1976). The Importance of the Transit of Mercury of 1631. *Journal for the History of Astronomy*, 7(1):1–10.
- Vanderburg, A. and Johnson, J. A. (2014). A Technique for Extracting Highly Precise Photometry for the Two-Wheeled Kepler Mission. *Publications of the Astronomical Society of the Pacific*, 126(944):948–958.
- Vanderspek, R., Doty, J. P., Fausnaugh, M., Villasenor, J. N., Jenkins, J. M., Berta-Thompson, Z. K., Burke, C. J., and Ricker, G. R. (2018). TESS Instrument Handbook. Technical Report Version 0.1, TESS Mission.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Venturini, J., Alibert, Y., and Benz, W. (2016). Planet formation with envelope enrichment: New insights on planetary diversity. *Astronomy & Astrophysics*, 596:A90.
- Vidal-Madjar, A., Désert, J. M., Lecavelier des Etangs, A., Hébrard, G., Ballester, G. E., Ehrenreich, D., Ferlet, R., McConnell, J. C., Mayor, M., and Parkinson, C. D. (2004). Detection of Oxygen and Carbon in the Hydrodynamically Escaping Atmosphere of the Extrasolar Planet HD 209458b. *The Astrophysical Journal*, 604:L69–L72.
- Waldmann, I. P. (2012). Of "Cocktail Parties" and Exoplanets. *The Astrophysical Journal*, 747:12.
- Waldmann, I. P. (2016). Dreaming of atmospheres. *The Astrophysical Journal*, 820(2):107.
- Waldmann, I. P., Tinetti, G., Deroo, P., Hollis, M. D. J., Yurchenko, S. N., and Tennyson, J. (2013). Blind Extraction of an Exoplanetary Spectrum through Independent Component Analysis. *The Astrophysical Journal*, 766(1):7.

- Waldmann, I. P., Tinetti, G., Rocchetto, M., Barton, E. J., Yurchenko, S. N., and Tennyson, J. (2015). Tau-REx I: A Next Generation Retrieval Code for Exoplanetary Atmospheres. *Astrophysical Journal*, 802(2):107.
- Walker, D. I., Lowther, J., Evens, N., Warren, J., Porter, J., and Farkas, K. (2022). Generic Protocol version 1.0. Technical report, Centre for Environment, Fisheries and Aquaculture Science.
- Wang, D., Hogg, D. W., Foreman-Mackey, D., and Schölkopf, B. (2016). A Causal, Data-Driven Approach to Modeling the Kepler Data. *Publications of the Astronomical Society of the Pacific*, 128(967):094503.
- Wang, J., Wang, Z., Li, J., and Wu, J. (2018). Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis. *arXiv:1806.08946 [cs, eess, stat]*.
- Wang, S., Li, B. Z., Khabsa, M., Fang, H., and Ma, H. (2020). Linformer: Self-Attention with Linear Complexity. *arXiv:2006.04768 [cs, stat]*.
- Welsh, B. Y. and Montgomery, S. L. (2018). Further detections of exocomet absorbing gas around Southern hemisphere A-type stars with known debris discs. *Monthly Notices of the Royal Astronomical Society*, 474:1515–1525.
- Welsh, B. Y. and Montgomery, S. L. (2019). Warm Circumstellar Disk Gas Surrounding the 1.4hspace0.167emGyr Old F2V Star HD 109085? *Research Notes of the AAS*, 3(1):25.
- Werner, M. W., Roellig, T. L., Low, F. J., Rieke, G. H., Rieke, M., Hoffmann, W. F., Young, E., Houck, J. R., Brandl, B., Fazio, G. G., Hora, J. L., Gehrz, R. D., Helou, G., Soifer, B. T., Stauffer, J., Keene, J., Eisenhardt, P., Gallagher, D., Gautier, T. N., Irace, W., Lawrence, C. R., Simmons, L., Van Cleve, J. E., Jura, M., Wright, E. L., and Cruikshank, D. P. (2004). The Spitzer Space Telescope Mission. *The Astrophysical Journal Supplement Series*, 154:1–9.
- Wheatley, P. J., West, R. G., Goad, M. R., Jenkins, J. S., Pollacco, D. L., Queloz, D., Rauer, H., Udry, S., Watson, C. A., Chazelas, B., Eigmüller, P., Lambert, G., Genolet, L., McCormac, J., Walker, S., Armstrong, D. J., Bayliss, D., Bento, J., Bouchy, F., Burleigh, M. R., Cabrera, J., Casewell, S. L., Chaushev, A., Chote, P., Csizmadia, S., Erikson, A., Faedi, F., Foxell, E., Gänsicke, B. T., Gillen, E., Grange, A., Günther, M. N., Hodgkin,

- S. T., Jackman, J., Jordán, A., Loudén, T., Metrailler, L., Moyano, M., Nielsen, L. D., Osborn, H. P., Poppenhaeger, K., Raddi, R., Raynard, L., Smith, A. M. S., Soto, M., and Titz-Weider, R. (2018). The Next Generation Transit Survey (NGTS). *Monthly Notices of the Royal Astronomical Society*, 475:4476–4493.
- Wilder, M. L., Middleton, F., Larsen, D. A., Du, Q., Fenty, A., Zeng, T., Insaf, T., Kilaru, P., Collins, M., Kmush, B., and Green, H. C. (2021). Co-quantification of crAssphage increases confidence in wastewater-based epidemiology for SARS-CoV-2 in low prevalence areas. *Water Research X*, 11:100100.
- Wohland, P., Burkitt, M., Norman, P., Rees, P., Boden, P., and Durham, H. (2011). ETHPOP Database : Ethnic group population trends and projections for UK local areas: Dissemination of innovative data inputs, model outputs, documentation and skills. <https://eprints.ncl.ac.uk>.
- Wolszczan, A. and Frail, D. A. (1992). A planetary system around the millisecond pulsar PSR1257 + 12. *Nature*, 355(6356):145–147.
- Woo, G., Liu, C., Sahoo, D., Kumar, A., and Hoi, S. (2022). ETSformer: Exponential Smoothing Transformers for Time-series Forecasting. *ArXiv*.
- Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W. L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A. D., Erickson, T. B., Chai, P. R., Thompson, J., and Alm, E. J. (2020a). SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems*, 5(4):e00614–20.
- Wu, S. L., Mertens, A. N., Crider, Y. S., Nguyen, A., Pokpongkiat, N. N., Djajadi, S., Seth, A., Hsiang, M. S., Colford, J. M., Reingold, A., Arnold, B. F., Hubbard, A., and Benjamin-Chung, J. (2020b). Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications*, 11(1):4507.
- Wyrzykowski, \., Rynkiewicz, A. E., Skowron, J., Kozłowski, S., Udalski, A., Szymański, M. K., Kubiak, M., Soszyński, I., Pietrzyński, G., Poleski, R., Pietrukowicz, P., and Pawlak, M. (2015). OGLE-III MICROLENSING EVENTS AND THE STRUCTURE OF THE GALACTIC BULGE. *The Astrophysical Journal Supplement Series*, 216(1):12.

- Wyrzykowski, Ł., Kostrzewa-Rutkowska, Z., Skowron, J., Rybicki, K. A., Mróz, P., Kozłowski, S., Udalski, A., Szymański, M. K., Pietrzyński, G., Soszyński, I., Ulaczyk, K., Pietrukowicz, P., Poleski, R., Pawlak, M., Iłkiewicz, K., and Rattenbury, N. J. (2016). Black hole, neutron star and white dwarf candidates from microlensing with OGLE-III★. *Monthly Notices of the Royal Astronomical Society*, 458(3):3012–3026.
- Xiao, A., Wu, F., Bushman, M., Zhang, J., Imakaev, M., Chai, P. R., Duvallet, C., Endo, N., Erickson, T. B., Armas, F., Arnold, B., Chen, H., Chandra, F., Ghaeli, N., Gu, X., Hanage, W. P., Lee, W. L., Matus, M., McElroy, K. A., Moniz, K., Rhode, S. F., Thompson, J., and Alm, E. J. (2022). Metrics to relate COVID-19 wastewater data to clinical testing dynamics. *Water Research*, 212:118070.
- Ye, Y. (2022). China is hatching a plan to find Earth 2.0. *Nature*, 604(7906):415–415.
- Yeh, L.-C. and Jiang, I.-G. (2020). Searching for Possible Exoplanet Transits from BRITe Data through a Machine Learning Technique. *Publications of the Astronomical Society of the Pacific*, 133(1019):014401.
- Yi, J., Lee, J., Kim, K. J., Hwang, S. J., and Yang, E. (2020). Why Not to Use Zero Imputation? Correcting Sparsity Bias in Training Neural Networks. In *ICLR*.
- Yip, K. H., Changeat, Q., Edwards, B., Morvan, M., Chubb, K. L., Tsiaras, A., Waldmann, I. P., and Tinetti, G. (2020a). On the Compatibility of Ground-based and Space-based Data: WASP-96 b, an Example. *The Astronomical Journal*, 161(1):4.
- Yip, K. H., Changeat, Q., Nikolaou, N., Morvan, M., Edwards, B., Waldmann, I. P., and Tinetti, G. (2021). Peeking inside the Black Box: Interpreting Deep-learning Models for Exoplanet Atmospheric Retrievals. *The Astronomical Journal*, 162:195.
- Yip, K. H., Nikolaou, N., Coronica, P., Tsiaras, A., Edwards, B., Changeat, Q., Morvan, M., Biller, B., Hinkley, S., Salmond, J., Archer, M., Sumption, P., Choquet, E., Soummer, R., Pueyo, L., and Waldmann, I. P. (2020b). Pushing the Limits of Exoplanet Discovery via Direct Imaging with Deep Learning. In Brefeld, U., Fromont, E., Hotho, A., Knobbe, A., Maathuis, M., and Robardet, C., editors, *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science, pages 322–338, Cham. Springer International Publishing.

- Yip, K. H., Waldmann, I. P., Changeat, Q., Morvan, M., Al-Refaie, A. F., Edwards, B., Nikolaou, N., Tsiaras, A., de Oliveira, C. A., Lagage, P.-O., Jenner, C., Cho, J. Y.-K., Thiyaalingam, J., and Tinetti, G. (2022). ESA-Ariel Data Challenge NeurIPS 2022: Inferring Physical Properties of Exoplanets From Next-Generation Telescopes.
- Yu, L., Vanderburg, A., Huang, C., Shallue, C. J., Crossfield, I. J. M., Gaudi, B. S., Daylan, T., Dattilo, A., Armstrong, D. J., Ricker, G. R., Vanderspek, R. K., Latham, D. W., Seager, S., Dittmann, J., Doty, J. P., Glidden, A., and Quinn, S. N. (2019). Identifying Exoplanets with Deep Learning. III. Automated Triage and Vetting of TESS Candidates. *The Astronomical Journal*, 158:25.
- Zdeborová, L. (2020). Understanding deep learning is also a job for physicists. *Nature Physics*, 16(6):602–604.
- Zellem, R. T., Swain, M. R., Roudier, G., Shkolnik, E. L., Creech-Eakman, M. J., Ciardi, D. R., Line, M. R., Iyer, A. R., Bryden, G., Llama, J., and Fahy, K. A. (2017). Forecasting the Impact of Stellar Activity on Transiting Exoplanet Spectra. *The Astrophysical Journal*, 844(1):27.
- Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., and Eickhoff, C. (2021). A Transformer-based Framework for Multivariate Time Series Representation Learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, pages 2114–2124, New York, NY, USA. Association for Computing Machinery.
- Zhang, K., Bloom, J. S., Gaudi, B. S., Lanusse, F., Lam, C., and Lu, J. (2021a). Automating Inference of Binary Microlensing Events with Neural Density Estimation.
- Zhang, K., Bloom, J. S., Gaudi, B. S., Lanusse, F., Lam, C., and Lu, J. R. (2021b). Real-Time Likelihood-Free Inference of Roman Binary Microlensing Events with Amortized Neural Posterior Estimation. *The Astronomical Journal*, 161(6):262.
- Zhang, M., Knutson, H. A., Kataria, T., Schwartz, J. C., Cowan, N. B., Showman, A. P., Burrows, A., Fortney, J. J., Todorov, K., Desert, J.-M., Agol, E., and Deming, D. (2018). Phase Curves of WASP-33b and HD 149026b and a New Correlation between Phase Curve Offset and Irradiation Temperature. *The Astronomical Journal*, 155(2):83.

- Zhang, Y. and Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58:308–324.
- Zhao, J. and Tinney, C. G. (2020). FIESTA – disentangling stellar variability from exoplanets in the Fourier domain. *Monthly Notices of the Royal Astronomical Society*, 491(3):4131–4146.
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *The Thirty-Fifth {AAAI} Conference on Artificial Intelligence, {AAAI} 2021, Virtual Conference*, volume 35, pages 11106–11115. {AAAI} Press.
- Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R. (2022). FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, pages 27268–27286. PMLR.
- Zhou, Y., Apai, D., Lew, B. W. P., and Schneider, G. (2017). A Physical Model-based Correction for Charge Traps in the *Hubble Space Telescope* ’s Wide Field Camera 3 Near-IR Detector and Its Applications to Transiting Exoplanets and Brown Dwarfs. *The Astronomical Journal*, 153(6):243.
- Zingales, T. and Waldmann, I. P. (2018). ExoGAN: Retrieving Exoplanetary Atmospheres Using Deep Convolutional Generative Adversarial Networks. *The Astronomical Journal*, 156(6):268.
- Zucker, S. and Giryes, R. (2018). Shallow Transits - Deep Learning I: Feasibility Study of Deep Learning to Detect Periodic Transits of Exoplanets. *The Astronomical Journal*, 155(4):147.