UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

CORSO DI DOTTORATO DI RICERCA IN ASTRONOMIA

CICLO XXXII

Habitability Studies of Super Earths Atmospheres

Tesi redatta con il contributo finanziario
dell’Istituto Nazionale di Astrofisica

Direttore della scuola: Ch.mo Prof. Giampaolo Piotto
Supervisore: Dr. Riccardo Claudi

Valutatori Esterni:

Dr. Luca Fossati
Dr. David Pawlowski

Dottoranda:

Eleonora Alei

Per Andrea, Elisa,
Maria e Cesare.

We stand on a great threshold in the human history of space exploration. If life is prevalent in our neighborhood of the galaxy, it is within our resources and technological reach to be the first generation in human history to finally cross this threshold, and to learn if there is life of any kind beyond Earth.

— **Sara Seager (1971-present)**
(Canadian-American astronomer and planetary scientist)

Contents

Acknowledgements

Throughout the whole Ph.D. project I received great support, funding and help by many INAF members.

I would first like to thank my supervisor, Dr. Riccardo Claudi (INAF - Astronomical Observatory of Padova), for the support during the Ph.D., as well as for allowing me to connect with the astronomical community through collaborations and meetings all over the world.

I thank also the internal and external referees of this work, Dr. Silvano Desidera, Dr. Valerio Nascimbeni, Dr. Luca Fossati, and Dr. David Pawlowski, for their insightful comments.

I appreciate the support of the other members of the exoplanetary group in Padua, as well as the Italian community GAPS, on this project.

I am very grateful to Prof. Cesare Cecchi-Pestellini, Dr. Giambattista Aresu, Dr. Antonino Petralia, Dr. Giusi Micela, Dr. Daniele Locci, and Dr. Angela Ciaravella from the INAF - Astronomical Observatory of Palermo and Cagliari for the support and collaboration in developing new features in `MAGRATHEA`.

Similarly, Dr. Marco Molinaro and Dr. Andrea Bigamini from the INAF - Astronomical Observatory of Trieste helped a great deal in the development of `Exo-MerCat`, as well as in connecting me with the Virtual Observatory community, and for this, I thank them as well.

I would like to acknowledge Prof. Nicoletta La Rocca, Dr. Luca Poletto, Dr. Lorenzo Cocola, Dr. Mariano Battistuzzi, Dr. Bernardo Salasnich, Dr. Emanuele Pace, and Dr. Marco Sergio Erculiani and all the people involved the “Atmosphere in a Test Tube” project.

I’d thank also my colleagues Serena, Ilaria, Cecilia, Mariangela, Elisa, Federico, Daniele, and Davide for their friendship and support in all aspects of my life, in and out of the office.

I would finally like to thank my family and friends, for immutable encouragement and support.

Abstract

Over the centuries, mankind always wondered whether other worlds existed, as well as other life forms upon their surfaces. This topic was considered “science-fiction” a few decades ago, but now it’s becoming more and more realistic: actually, different planets do exist, and some of them could bear life. Up to date, a database of more than 7000 planets or candidates is continuously updated, as current facilities keep on discovering new ones – a simple estimate suggests that in our galaxy tens of billions of new planets await to be revealed.

In particular, a class of planets seems to be predominant in the Milky Way, while being absent in our Solar System: these are the Super Earths, terrestrial planets more massive than the Earth, but still interesting targets for the study of the formation of planetary systems and potential homes of other life-forms. These are the ideal targets for the search of life in the universe, and it should be no wonder that many future ground- and space-based missions aim to detect more of those.

Being extrasolar planetology a relatively new field of astrophysics, many things need to be set and done. In this case, a full characterization of the different atmospheres Super Earths are likely to have is needed to correctly understand observational data; also, one should retrieve information about their formation and be able to presume that some form of life exists on the surface. To do that, theoretical modeling is needed: by simulating a simple imaginary exoplanet, one could have a better understanding of how all active processes interact within the system and what observable features they express so that they could be recognized when observing a real exoplanet.

In this Ph.D. project, I explored the topic in various ways, starting with an introduction to the aims and methodologies in Chapter 1. Then, it proceeds with an overview about the detection techniques in Chapter 2, the current knowledge about Super Earths in Chapter 3, and the concept of habitability in Chapter 4.

At first, the sample of currently known exoplanets (and Super Earths in particular) needed to be retrieved. This step was challenging, since the ever-evolving vetting of candidates required a quick response by the various archives which often lagged behind schedule. Furthermore, the lack of uniformity for the notation of planetary objects makes the job of the maintainers extremely complicated and subject to errors.

In the framework of a standardized, Virtual Observatory (VO) aware treatment of the exoplanets, I developed `Exo-MerCat`, the `Exoplanet Merged Catalog`, which collects data from the most important online archives and merges the information while correcting nomenclature, status, and coordinate issues. This catalog is now a VO resource and has been positively accepted by the International Virtual Observatory Alliance (IVOA), as well as being used for PLATO and ARIEL space missions. It is described in Chapter 5.

`Exo-MerCat` allowed retrieving the sample of known Super Earths, which is then used as input to create a grid of atmospheric models to be run with the 1D radiative-convective model `MAGRATHEA` (in Chapter 6), which I contributed to develop.

`MAGRATHEA` can model Earth and Mars-like atmospheres, covering a wide range of the physical and chemical parameter space. It is able to calculate the radiative-convective equilibrium solution of an atmosphere in a very short time (a few hours of computational time), allowing us to fulfill a grid of 18000 atmospheric models of theoretical planets and 2400 ones of observed planets, as retrieved by the `Exo-MerCat` catalog. These are discussed in Chapter 7.

These models can be useful to study under which physical and atmospheric conditions it is possible to find liquid water on the surface of the planet, an essential requirement for the habitability of exoplanets.

The atmospheric pressure-temperature profiles for the observed Super Earths grid, as retrieved by `MAGRATHEA`, were then used as an input for the `Exoplanet Ozone Model` or `EOM`, to produce the corresponding ozone profile. The study of the ozone abundance on terrestrial exoplanets is important for both observational and habitability purposes: O_3 is a biosignature that could in principle be detected on life-bearing exoplanets, and its shielding effect is crucial for life to prosper. The `EOM` is, at present, still at its early stages, but can efficiently produce the ozone profile of an atmosphere by solving the (photo-induced and thermal) chemistry of the oxygen-related species. Both description of the code and results can be found in Chapter 8.

The laboratory experiments performed at the Biology Department of the University of Padua, described shortly in Chapter 9, can benefit from the theoretical results from `MAGRATHEA`. The physical and chemical conditions at the surface are reproduced in the laboratory, forming exotic environments at which cyanobacteria are exposed. The study of the survival rate and the variation of the chemical composition caused by the presence of biological activity can be thus performed: this is indispensable in order to understand if, and when, a habitable planet can be actually inhabited.

Sommario

Nei secoli, ci siamo sempre chiesti se esistessero altri mondi e altre forme di vita sulla superficie di questi. Questo argomento è stato considerato spesso “fantascienza” fino a pochi decenni fa, ma ora sta diventando sempre più realistico: in realtà, pianeti diversi esistono ed alcuni di essi possono ospitare la vita. Ad oggi, un archivio di più di 7000 pianeti confermati o candidati è costantemente aggiornato, al passo con gli strumenti che ne scoprono sempre di più - una semplice stima suggerirebbe che, solo nella nostra Galassia, decine di miliardi di nuovi pianeti aspettano di essere scoperti.

In particolare, una classe di pianeti sembra essere predominante nella Via Lattea, mentre è assente nel nostro Sistema Solare: le Super Terre, pianeti terrestri più massicci della Terra, ma candidati interessantissimi per lo studio della formazione dei sistemi planetari, oltre ad essere potenziali luoghi ospitali per altre forme di vita. Questi pianeti sono i soggetti ideali per la ricerca della vita nell’universo e non dovrebbe sorprendere che molte future missioni da terra e da spazio aspirino a trovarne sempre di più.

Essendo la planetologia extrasolare un campo relativamente nuovo dell’astrofisica, molte cose devono essere ancora studiate. In questo caso, una caratterizzazione più dettagliata delle possibili atmosfere di Super Terre è necessaria per comprendere meglio le osservazioni; inoltre, bisognerebbe ricavare informazioni sulla loro formazione e poter presumere se, ed in quali casi, forme di vita potrebbero esistere su quei pianeti. Per fare ciò, un approccio teorico è necessario: simulando un pianeta in maniera semplificata, si potrebbe avere una migliore comprensione di come tutti i processi attivi interagiscono tra loro e quali osservabili producono, affinché possano essere identificate quando si osserva un vero esopianeta.

In questo progetto di Dottorato, ho esplorato l’argomento da diversi punti di vista, iniziando con una introduzione (Capitolo 1) allo scopo e alle metodologie. In seguito, è riportata una sintesi dei metodi di scoperta (Capitolo 2), di ciò che è noto ad oggi sulle Super Terre (Capitolo 3), e del concetto di abitabilità (Capitolo 4).

Innanzitutto, un insieme di tutti gli esopianeti noti (e Super Terre in particolare), doveva essere ricavato. Questo processo è stato difficoltoso, poiché la continua conferma (o smentita) dei candidati pianeti richiedeva una risposta rapida da parte dei mantenitori dei principali archivi, la quale spesso era soggetta a ritardi. Inoltre, la notazione degli oggetti planetari non segue – ad oggi – una procedura standard. Ciò rende il lavoro dei mantenitori estremamente complicato e soggetto ad errori. In una prospettiva di un trattamento standardizzato dei dati, il quale possa

rientrare nei canoni del Virtual Observatory (VO), ho sviluppato `Exo-MerCat`, al fine di collezionare dati dai più importanti archivi online, incrociando le informazioni e correggendo problemi di nomenclatura, status e coordinate. Questo catalogo, descritto nel Capitolo 5, è ora una risorsa VO ed è stato accettato positivamente dall'International Virtual Observatory Alliance (IVOA), oltre ad essere usato per le missioni spaziali PLATO e ARIEL.

`Exo-MerCat` ha permesso di ricavare l'insieme di Super Terre note, usato poi per creare una griglia di modelli atmosferici utilizzata dal modello 1D radiativo-convettivo `MAGRATHEA` (descritto nel Capitolo 6), che ho contribuito a sviluppare.

`MAGRATHEA` riesce a riprodurre atmosfere di tipo terrestre e marziano, coprendo un largo intervallo di parametri fisici e chimici. Il codice calcola il profilo di equilibrio radiativo-convettivo di una atmosfera in poche ore di tempo computazionale, consentendoci di riempire una griglia di 18000 modelli di pianeti teorici e una di 2400 modelli di pianeti osservati, ricavati dall'insieme prodotto da `Exo-MerCat`. I risultati prodotti sono descritti nel Capitolo 7.

Questi modelli possono essere utili per studiare sotto quali condizioni fisiche e atmosferiche è possibile trovare acqua liquida sulla superficie di un pianeta, requisito essenziale per l'abitabilità degli esopianeti.

I modelli atmosferici delle Super Terre osservate ricavati da `MAGRATHEA` sono stati quindi usati come input per l'`Exoplanet Ozone Model` al fine di produrre la concentrazione di ozono corrispondente ai profili stessi. Lo studio dell'ozono sulle atmosfere di esopianeti di tipo terrestre è importante per scopi osservativi e considerazioni sull'abitabilità: l'ozono è una "biosignature" che potrebbe essere osservata in spettri di esopianeti abitati, e il suo effetto schermante è essenziale per la proliferazione di forme di vita. Questo codice è, al momento, ancora preliminare, ma può riprodurre il profilo dell'abbondanza di ozono di una atmosfera risolvendo la chimica foto-indotta e termica delle specie legate all'ossigeno. La descrizione del codice ed i risultati ottenuti sono descritti nel Capitolo 8.

I risultati teorici ottenuti dai vari codici sono utili agli esperimenti di laboratorio effettuati al Dipartimento di Biologia dell'Università di Padova. Considerando alcune atmosfere calcolate da `MAGRATHEA`, si possono riprodurre le condizioni fisiche e chimiche alla superficie in laboratorio, formando atmosfere esotiche ed esponendo cianobatteri a queste. Lo studio della sopravvivenza e dell'adattamento dei batteri, così come della variazione della composizione chimica causata dall'attività biologica, può essere eseguito. Ciò è indispensabile per comprendere se, e sotto quali condizioni, un pianeta abitabile può essere effettivamente abitato.

List of Figures

List of Tables

List of Listings

1 Introduction

Exoplanetology is, as of now, one of the most flourishing fields of astrophysics. At the time of writing, more than 7000 confirmed or candidate exoplanets have been discovered, and many are to be revealed with current and future facilities. One of the most interesting revelations was the detection of an ever-growing population of terrestrial planets, up to ten times more massive than the Earth. These are the Super Earths; surprisingly, the Solar System is lacking planets belonging to this class, but they appear to be common around other main-sequence stars.

With the current technologies, the possibility of sending a probe even to the closest terrestrial exoplanet, Proxima Cen b at 4.24 light-years [Anglada-Escudé et al., 2016], to detect molecules in its atmosphere and to define its internal composition is definitely faint. The only way for an astronomer to characterize an exoplanet is by means of indirect techniques, such as the spectroscopic analysis of its atmosphere.

In particular, spectroscopic studies of the atmospheres of Super Earths are of paramount importance to probe their composition, which could provide hints about the formation of such planets, their evolution and the conditions under which life could form on their surfaces.

However, this is not a simple task with the currently available facilities, since the emitted or reflected radiation coming from a planetary atmosphere is orders of magnitude smaller than the one emitted by the host star. For this reason, little or no atmospheric spectra of temperate terrestrial exoplanets have been measured so far. Nevertheless, since a new generation of spectrographs is currently being constructed with high enough resolution to observe an Earth analog around cooler stars, this milestone is within reach in the next few years.

Now more than ever, a theoretical approach to the study of atmospheres of terrestrial exoplanets is necessary, to provide cross-reference tools and to set the ground for the upcoming observations.

This thesis aims to contribute to the construction of a robust theoretical framework for the study of the atmospheres of terrestrial exoplanets, with a particular focus on the detection of life signatures on such targets.

To do this, some preliminary statistical studies are needed to understand the currently known features of the sample of confirmed Super Earths. On that basis, it is possible to build grids of simulations to find the pressure-temperature equilibrium profile for atmospheres of various chemical compositions that are likely to exist on those planets.

The results of these simulations can be then used as input conditions for other theoretical studies: in particular, qualitative studies on the ozone production on oxygen-rich atmospheres may lead to interesting conclusions on the surface habitability, as well as information on the detection of ozone itself, one of the atmospheric indicators of life on a planet.

Simulations are helpful also for laboratory studies, which connect astronomy, biology, and engineering for the research on the survival rate of terrestrial life forms on alien environments. Being able to reproduce plausible atmospheres based on the theoretical results, irradiating them with stellar simulators, and exposing biological samples to the altered conditions, would produce interesting information concerning the likelihood of finding life on exotic planets, as well as provide quantitative data on the products of their life cycle.

That being said, the Ph.D. project touched various aspects of the general topic of the habitability of Super Earths, building up different tools to achieve different milestones. A graphic overview of the topics that were included in the Ph.D. projects is shown in Figure 1.1.

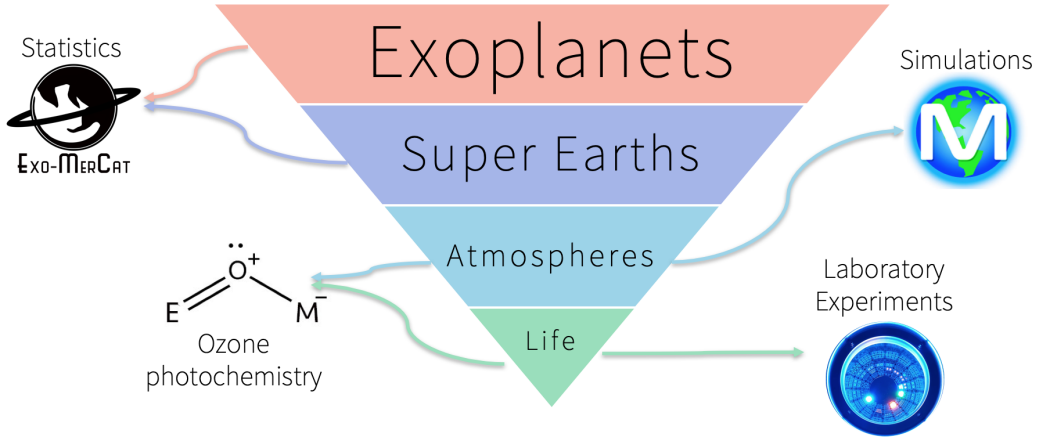


Fig. 1.1: Graphic overview of the general topics treated in the thesis, and more in particular of the four Ph.D. projects, with the logos of the corresponding tools developed for those purposes.

First of all, I studied the properties of the exoplanetary population and I developed a software program, Exo-MerCat, that collects all data available from the most important online databases and arranges the information while uniforming the current knowledge about the available targets.

In a framework of a standard way of displaying, collecting, and providing to the user the available data, Exo-MerCat obeys to the International Virtual Observatory Alliance (IVOA) standards and it is a pioneering resource for any VO-aware tool for exoplanetary data. In this way, the catalog produced by Exo-MerCat can be easily read and customized by any software that agrees to the IVOA policy for the standardization of nomenclature and data treatment.

Besides all its general functionalities, Exo-MerCat has been useful for this thesis since it allowed the retrieval of the sample of currently known Super-Earths. Data belonging to this sample were used as inputs for the simulations and the experiments that constitute the remainder of this thesis.

This leads us to the second project: the theoretical modeling of terrestrial atmospheres.

With the current technologies, the atmospheric modeling software programs can be extremely powerful and can reproduce a wide variety of environments, even simulating tridimensional dynamics – albeit with a few approximations. However, a simplified, 1D treatment of terrestrial atmospheres can be still useful to assess the influence of every parameter in the global energy balance, exploring a wider range of parameter space in a shorter time compared to more complex 3D General Circulation Models. For this reason, MAGRATHEA was created.

MAGRATHEA is an Italian effort that started a decade ago and that has achieved already the accuracy that is provided by currently available open-source codes but focusing mainly on temperate terrestrial exoplanets, rather than gaseous giants. This software allows the user to simulate a wide range of different atmospheres producing the corresponding 1D pressure-temperature profiles.

Approximately 20000 simulations were made using MAGRATHEA, both to study its potential and to retrieve scientifically-relevant results. Mass and radius measurements belonging to the sample of Super Earths retrieved by Exo-MerCat, as well as information concerning their host stars, were used to define Earth- or Mars-like plausible atmospheres that could be present on the currently known terrestrial planets and that could be potentially observed with future characterization facilities.

The results of the simulations were then used as input for the two remaining projects: a toy model of ozone photochemistry (the `Exoplanet Ozone Model`, or `EOM`), and the laboratory experiments held in the Biology Department of the University of Padua.

The `EOM` is a project still at its infancy, but it already allowed the retrieval of the ozone profiles that corresponded to the radiative-convective equilibrium profiles of the currently known Super Earths as retrieved by `MAGRATHEA`. The code was built to be extremely flexible and to take into account further features. The general aim of the `EOM` project would be to reproduce the behavior of the ozone profile in different atmospheres, with a particular focus on planets orbiting active M stars, on which the presence of atmospheric ozone could be essential for the survival of any life form on the surface. This effort will be continued in the next years, on the basis of the validation of the code presented in this thesis.

More quantitative information on the behavior of biological samples could be retrieved through laboratory experiments, such as the ones held in the Biology Department of the University of Padua, within the "Atmosphere in a Test Tube" project. Biologists are studying terrestrial cyanobacteria that produce oxygen using photosynthesis in exotic environments. The biological samples are exposed to different stellar sources, reproduced by a stellar simulator built by engineers and astronomers of CNR Institutes for Photonics and Nanotechnologies and INAF Astronomical Observatory of Padua, and to altered atmospheric mixtures, whose pressure, temperature, and composition are determined by the simulations performed by `MAGRATHEA`.

These various projects are indeed very diverse and each one of them requires a deep knowledge of various aspects of the exoplanetology. However, during these three years, I tried to connect every piece, filling a more complex puzzle of knowledge that it is not provided by current modeling efforts. The direct connection between data archives, simulations, and laboratory experiments will produce a set of information, instruments, and facilities, that will allow a quantitative and more reliable understanding of the incoming data concerning the atmospheres of terrestrial exoplanets. By knowing the processes that are present in an exoplanetary atmosphere, as well as the very complex interaction with the biosphere, the astronomer would be able to detect and understand biosignatures (i.e. signatures of chemical species which are tracers of life), when we will finally be able to characterize the atmosphere of a terrestrial, potentially habitable exoplanet.

Exoplanets

2

> *Onde possiamo stimare che de stelle innumerabili sono altre tante lune, altre tanti globi terrestri, altre tanti mondi simili a questo; circa gli quali par che questa terra si volte, come quelli appaiono rivolgersi ed aggirarsi circa questa terra.*
>
> ---
>
> *"Wherefore we may suppose that of the innumerable stars some are moons, some terrestrial globes, some worlds like our own, and around them our earth appeareth in their eyes to revolve just as they appear to us to revolve and to take their course around the earth."*
>
> **— Giordano Bruno (1548-1600)**
> (Italian Dominican friar, philosopher, mathematician, cosmological theorist)

This chapter describes the current state of the art in exoplanetology, with a focus on the methods of detection and on some statistical discussion on the sample of the known exoplanets.

2.1 Overview

The thissearch for exoplanets has been growing as a bold branch of astrophysics for the past two decades, as surveys and missions reveal day by day new candidates to enrich the list of all the currently known exoplanets, as well as their orbital and bulk parameters, and the properties of their host stars. By these observations, we are aware of the fact that exoplanets differ from one another in many of these physical parameters, and in the future, we should be able to fully understand the diversity of

chemical compositions, atmospheric processes, internal structures and formation conditions of the different classes [Madhusudhan et al., 2014].

Astrophysicists faced an analogue situation in the 20th century, when they had to find a classification for the many stars observed: it was found that a star's structure was determined only by its mass, initial chemical composition, and age; similarly, Russell and Hertzsprung understood that the stellar brightness was strictly linked to the perceived color of the star itself. As a consequence, from the observational data, one could derive key parameters such as temperature and radius, allowing the classification of the observed target.

This time, however, the knowledge of the planetary mass, radius, temperature, and chemical composition can provide very few constraints about a planet, as they are loosely correlated with one another and depend from the initial conditions, history, and interaction with the host star [Tinetti, Encrenaz, et al., 2013]. Even in our Solar System, there are no planets with the same characteristics, and as we look further than the Kuiper belt boundaries, things become even more complex.

The first exoplanet orbiting a main-sequence star ever discovered (51 Peg b, by Mayor and Queloz (1995)) was a gaseous giant orbiting much nearer to the star than Mercury: because of this peculiarity, 51 Peg b and some of the other exoplanets later found, are now known as "Hot Jupiters". This brought to discussion the "nebular theory" of planetary formation, invoking migration mechanisms which may have happened in the Solar System also.

Since then, planets were found around every class of stars, including pulsars and binaries, and nearly the 60% of them follow extremely eccentric orbits [Tinetti, Encrenaz, et al., 2013]; they also span a large range in mass and effective temperature, so that we can indicatively define three classes of mass:

1. Jupiters ($M > 20\ M_\oplus$) — H- and He-rich gaseous planets;
2. Neptunes ($10\ M_\oplus < M < 20\ M_\oplus$) — H- and He-rich gaseous planets;
3. Small Exos ($M < 10\ M_\oplus$) — known as "Earths" and "Super Earths", rocky planets;

And five classes of temperature:

1. Very hot ($T > 2000$ K);
2. Hot ($800\ K < T < 2000\ K$);
3. Warm ($350\ K < T < 800\ K$);

4. Temperate (250 K < T < 350 K);
5. Cold (T< 250 K).

Many of these planets have no analog in the Solar System in terms of physical parameters, but sometimes they share similar chemical composition: knowing those, we could perhaps gain information about planetary formation and evolution [Madhusudhan et al., 2014].

2.2 Observational methods

The study of exoplanetary characteristics cannot rely on *in situ* measurements, but only on remote observations [Tinetti, Encrenaz, et al., 2013], based on:

- Time-varying signals, involved in indirect detection methods such as radial velocities, transit detection, astrometry, pulsar timing, transit timing variation, gravitational microlensing;
- Spatially resolved signals, involved in direct detection methods such as high contrast imaging and interferometry.

These two categories of observational techniques are often combined and can produce many observables, from which it is possible to derive parameters about the interior and atmospheric composition (see Fig. 2.2). Thanks to these techniques, more than 5000 exoplanets or candidates are now known (see Fig. 2.1).

Three techniques that have achieved major results until now are better explained in the following paragraphs.

2.2.1 Radial Velocity Measurements

The first method historically used to identify an exoplanet is the radial velocities (RV) technique: this was the technique used to discover the very first exoplanet orbiting a main-sequence star, 51 Pegasi b. It is an indirect method since it is based on the detection of the reflex motion of the host star induced by the presence of a planet [Benatti, 2018].

A star + planet system can be imagined as a spectroscopic binary with an invisible companion [Fischer et al., 2014]: by identifying a Doppler signal (a periodical shift of the lines due to the motion of the host star) on the stellar spectra, it is possible to

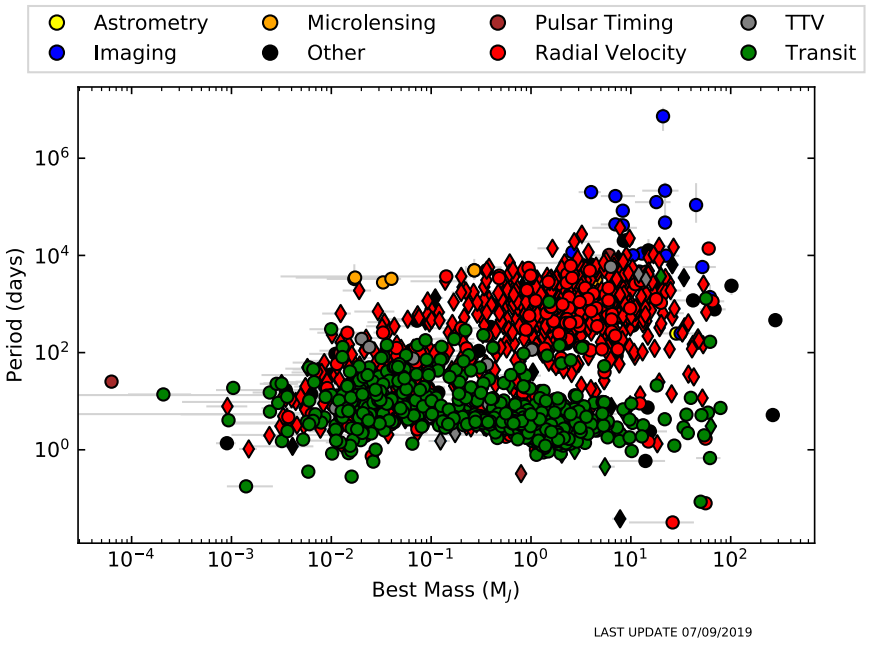


Fig. 2.1: Mass-Period diagram of currently known exoplanets obtained with `Exo-MerCat` (see Section 5).

retrieve information about the orbital period and the minimum mass of the planet $M_P \sin i$.

The radial velocity amplitude of the star K can be retrieved as follows

$$K\ [m\ s^{-1}] = \frac{28.4329\ [m\ s^{-1}]}{\sqrt{1-e^2}} \frac{M_P \sin i}{M_J} \left(\frac{M_\star + M_p}{M_\odot}\right)^{-2/3} \left(\frac{P}{yr}\right)^{-1/3} \tag{2.1}$$

Where M_P is the planetary mass in Jovian mass units, $M_\star$ is the stellar mass in solar mass units, i and e are the inclination and the eccentricity of the orbit respectively, and P is the revolution period of the planet in years.

This method has a higher sensibility for close-in planets with short orbital periods since the increase in gravitational force depends both on the planetary mass and on its semi-major axis.

However, the signal induced by the planet on a star is still fairly small: a Jovian-like planet orbiting a Sun-like star at 1 AU would cause a signal whose amplitude is about 30 $m\ s^{-1}$, while the Earth seen by another astronomer in the galaxy would cause a signal of about 10 $cm\ s^{-1}$ on the Sun. Also, many planetary periods need to

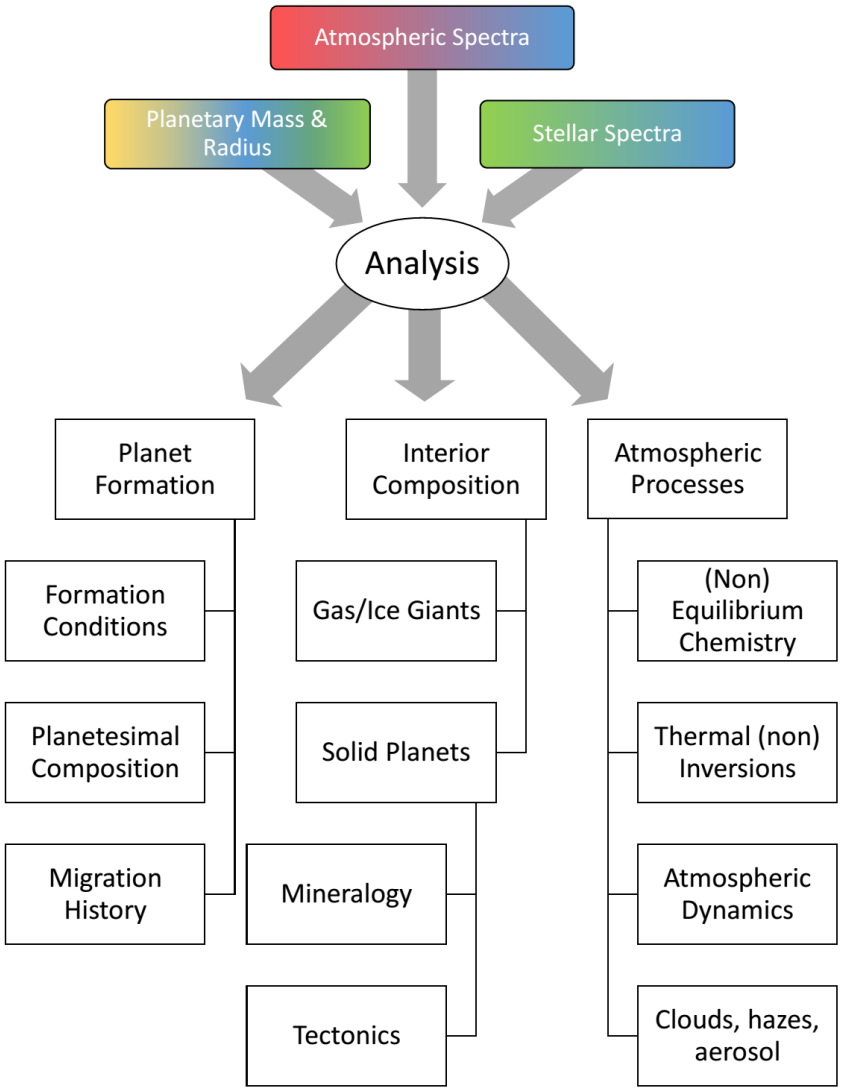


Fig. 2.2: Schematic diagram of exoplanet characterization. A different technique corresponds to a colour: in red, direct imaging; in blue, transit detection; in green, radial velocities; in yellow, astrometry.

be observed to be able to distinguish the nature of the signal, which could be on the other hand caused by aliasing or by stellar activity pulsation.

Stellar activity is, unfortunately, a noise source which could bring to false positives since it induces an intrinsic RV signal with similar or larger amplitude with respect to a Keplerian signal. Most of the activity signals are periodic or quasi-periodic and they cause time-dependent changes in the lines profiles as well. The most important stellar noise sources are [Benatti, 2018; Wright, 2018]:

- magnetic activity ($\tau \approx$ tens of days): some stars can establish a strong magnetic field that can cause inhomogeneities in brightness (spots and faculae);
- granulation ($\tau \approx$ hours): the surface of solar-like stars shrinks and expands as a result of the surface convective motions;
- global pulsations ($\tau \approx$ tens of minutes): observed in the Sun as well, the stars shrink and expand to maintain equilibrium;
- non-periodic stellar jitter.

To avoid the pollution that the host star can cause, radial velocity surveys can either focus on quieter targets (old, G-K dwarfs) whose jitter can be neglected [Wright, 2018], or try to mitigate these effects.

A few indicators could be used to discern true Keplerian signals from stellar noise, such as the Ca II H and K lines, as well as the Hα and the Na I D lines, widely used in asteroseismology [Kjeldsen et al., 1995]: if there is a strong magnetic field, the gas is heated and will have a further emission that fills the cores of such lines.

Furthermore, if there is a correlation between the radial velocity signal and the bisector velocity span (BVS, the difference between the average velocity values in regions at the top and the bottom of a typical line profile) the signal is likely to be due to stellar activity [Benatti, 2018]. This approach, however, requires high precision and intense monitoring of the target.

This issue is particularly present in the visible spectral range rather than the near-infrared [ibid., and references therein]. For this reason, the simultaneous observation of both bands could be essential to distinguish a Keplerian signal from a false positive since an activity-related signal would be wavelength dependent, while the one caused by the presence of a planet should be the same regardless of the wavelength observed band.

Some of the radial-velocity facilities currently in use for exoplanetary surveys and follow-up operations are HARPS (High Accuracy Radial Velocity Planet Searcher)

at ESO 3.6m telescope, HARPS-N (High Accuracy Radial velocity Planet Searcher for the Northern hemisphere) [Cosentino et al., 2012] and GIANO-B [Oliva et al., 2012] at Telescopio Nazionale Galileo (TNG), HIRES (High Resolution Echelle Spectrometer) [Vogt and Donald Penrod, 1988] at Keck Observatory, ESPRESSO (Echelle Spectrograph for Rocky Exoplanet- and Stable Spectroscopic Observations) [Pepe et al., 2010] at the Very Large Telescope (VLT), HERMES (High-Efficiency and high-Resolution Mercator Echelle Spectrograph) [Raskin et al., 2014] at the 1.2-m Mercator telescope on La Palma, TRES (Tillinghast Reflector Echelle Spectrograph) [Szentgyorgyi et al., 2007] at Fred Lawrence Whipple Observatory (FLWO) 1.5m telescope, LEVY [Vogt, Radovan, et al., 2014] at the Lick Observatory Automated Planet Finder, PEPSI (Potsdam Echelle Polarimetric and Spectroscopic Instrument) [Strassmeier et al., 2015] at the Large Binocular Telescope (LBT), HPF (Habitable Zone Planet Finder) at 10-meter Hobby-Eberly Telescope (HET) [Mahadevan et al., 2012] at McDonald Observatory.

More and more spectrographs are being developed to accomplish multi-wavelength observations, as well as to improve the precision of the measurements: at the time of writing, the most important ones are CARMENES [Quirrenbach et al., 2014] at 3.5 m Telescope at Calar Alto and GIARPS (GIANO-B+HARPS-N) [Claudi, Benatti, et al., 2017] at Telescopio Nazionale Galileo.

Future facilities include the simultaneous use of NIRPS (Near Infra-Red Planet Searcher) [Wildi et al., 2017] and HARPS (High-Accuracy Radial velocity Planetary Searcher) [Mayor, Pepe, et al., 2003] at ESO New Technology Telescope (NTT) as well, G-CLEF (GMT Consortium Large Earth Finder[1]) and GMTNIRS (GMT Near-IR Spectrograph[2]) at the Giant Magellan Telescope (GMT), HIRES (High Resolution Spectrograph) [Marconi et al., 2016] at the European Extremely Large Telescope (E-ELT), SPIRou (SpectroPolarimètre Infra-Rouge) [Parès et al., 2012] at the Canada-France-Hawaii Telescope (CFH).

2.2.2 Transit Detections

A *transit* (Fig. 2.3) takes place when a celestial body crosses the path of a more distant object with a larger angular diameter; an *occultation* takes place when the distant body has a smaller angular diameter. When a planet passes in front of its host star, the event is called *primary/direct transit*; when it passes behind the star, it is called *secondary/indirect transit*, or *eclipse* [Tinetti, Encrenaz, et al., 2013].

[1]https://www.gmto.org/resources/visible-echelle-spectrograph-g-clef/
[2]https://www.gmto.org/resources/ir-echelle-spectrograph-gmtnirs/

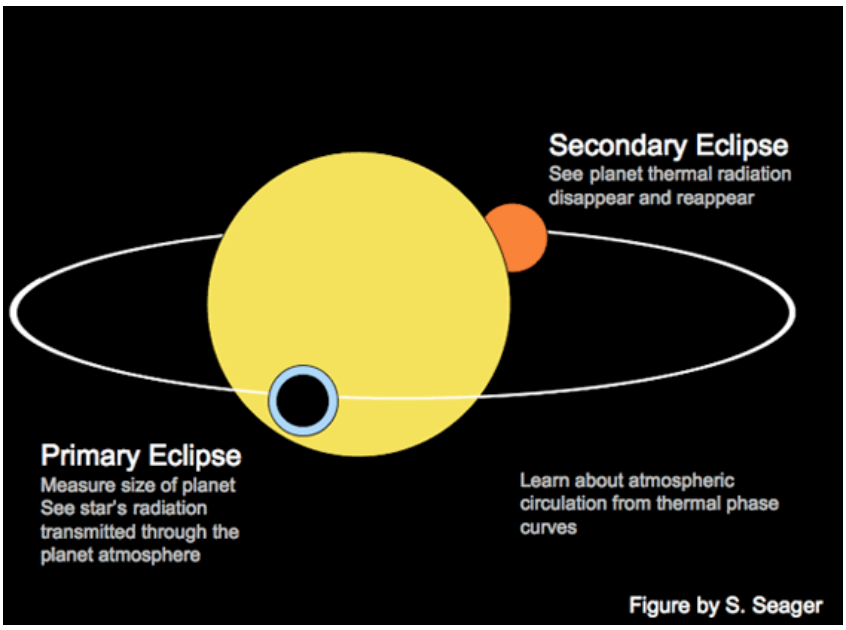


Fig. 2.3: Diagram of a transiting exoplanet going in front and behind its parent star [Seager and Deming, 2010].

The concept of using time-dependent flux variation of an object as a detection of a fainter companion was proposed in the early 50s; in the last decade, this method has been extremely successful in finding the majority of candidates [Cameron, 2016, and references therein], from the most massive planets to terrestrial size ones.

When observing a transit, the astronomer observes a decrease in flux due to the transiting planet which obscures a portion of the hotter stellar photosphere. As shown in Figure 2.4, the duration of the transit is the time occurred between contact points 1 and 4, 2 and 3 being the end of the ingress and the beginning of the egress from the stellar disk. As the planet revolves around the star, variable portions of its surface can be observed and contribute to the global flux, which increases quasi sinusoidally the closer the planet approaches the occultation. In that period, the star hides the planetary reflected/emitted flux, and only the stellar flux is visible.

This method is particularly effective in finding a measurement of the planetary radius R_P since it depends on the relative flux variation $\Delta F/F$ caused by the planet transiting in front of its host star whose radius is $R_\star$:

$$\frac{\Delta F}{F} \approx \frac{R_P^2}{R_\star^2} \tag{2.2}$$

If the limb darkening of the stellar photosphere is not considered, this leads to an overestimate of the transit depth. This is caused by the temperature gradient that is

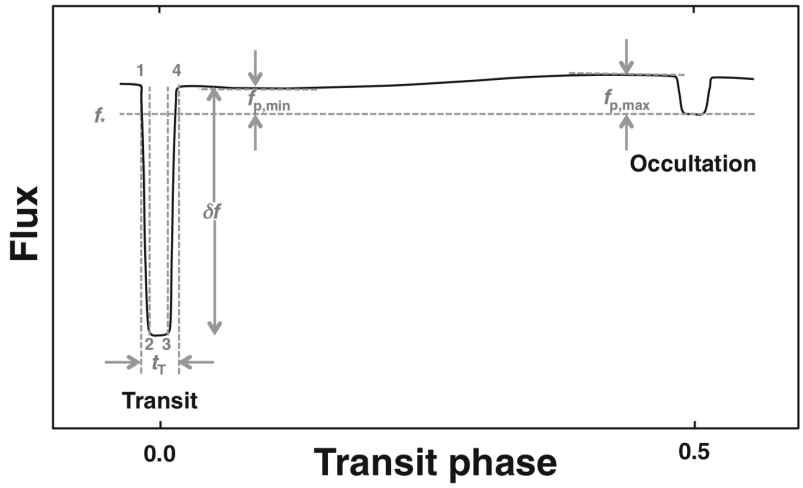


Fig. 2.4: Sketch of a typical transit light curve [Cameron, 2016], plotted as the flux variation with respect to the transit phase. $f_\star$ is the stellar flux, t_τ the transit duration, δf the transit depth, $f_{p,min}$ and $f_{p,max}$ the minimum and maximum of the planetary flux, the first being the flux coming from the coolest part of the nightside and the latter being the flux coming from the hottest hemisphere.

established from the core to the atmosphere of the star. When observing the center of the disk, the portion of photosphere that can be observed before reaching an optical depth of $\tau = 2/3$ allows to observe the radiation coming from the deeper, hotter regions; observing the edge of the star, the line of sight enters the photosphere at a steeper inclination, thus allowing to detect only the radiation coming from cooler zones of the stellar photosphere.

From the observer's point of view, the sky-projected distance between the center of the stellar disc and the center of the planetary disc at mid-transit can be expressed in terms of a dimensionless *impact parameter* b:

$$b = \frac{a \cos i}{R_\star} \tag{2.3}$$

where a is the semi-major axis, and i the inclination (see Figures 2.5 and 2.6).

The probability of a transit to be detectable over interstellar distances depends on the inclination of a planet, as well as the fraction of the celestial sphere that the planet's shadow sweeps out.

In order to have a grazing transit, the value of $a \cos i$ must lie between $R_\star + R_P$ and $R_\star - R_P$; to have a full transit, the inequality $a \cos i \leq R_\star - R_P$ must be satisfied.

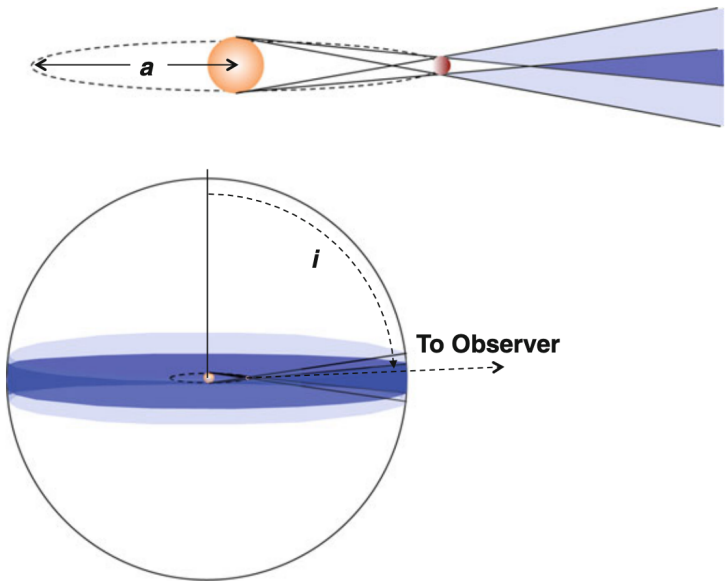


Fig. 2.5: Geometry of a full (dark shadow) and a grazing (light shadow) transit [Cameron, 2016].

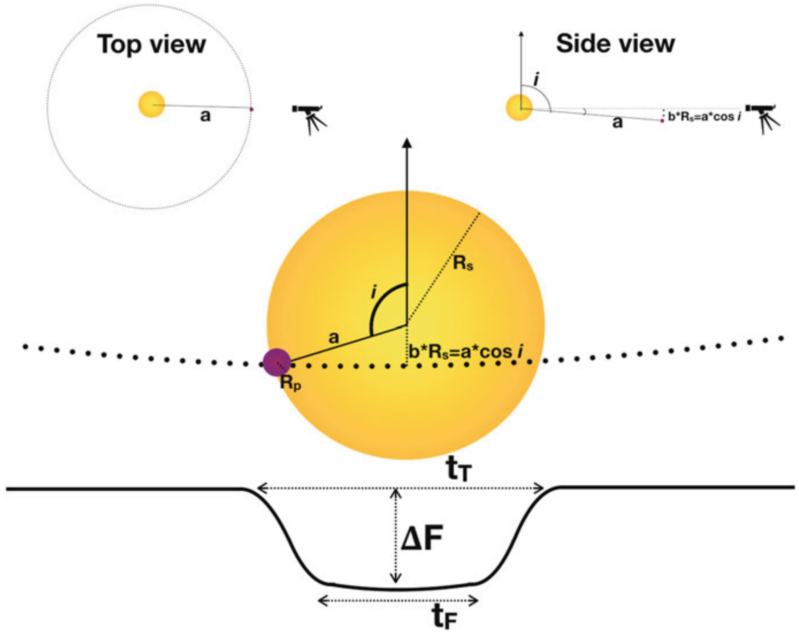


Fig. 2.6: Geometry of the impact factor definition [Deeg et al., 2018].

The fraction of the solid angle on the celestial sphere that is spanned by the shadow of a planet whose unknown inclination lies between i and $i+1$ is:

$$\frac{d\Omega}{4\pi} = \frac{2\pi \sin i di}{4\pi} = \frac{d(\cos i)}{2} \tag{2.4}$$

The probability of detecting a transit (whether grazing or full) is, therefore:

$$Pr\left(\cos i < \frac{R_\star + R_P}{a}\right) = \frac{1}{2}\int_{-(R_\star+R_P)/a}^{(R_\star+R_P)/a} d(\cos i) = \frac{R_\star + R_P}{a} \simeq \frac{R_\star}{a} \tag{2.5}$$

From the previous equation, it is clear that the detection of close-in planets is more probable. The planetary radius also plays an important role in the discovery of such kind of planets: as can be seen in Eq. 2.2, the greater the planetary radius, the higher the flux variation on the stellar flux (for a given stellar radius). This means that hot, massive planets can be identified with this technique.

If the stellar radius decreases, it could be easier to detect less massive planetary companions, because of the favorable radii ratio.

Information on the atmosphere of a transiting exoplanet can be retrieved from the flux difference of the star-planet system before, during and after the event: therefore, a very large sensibility is needed, as the planet to star flux contrast is, at best, from 10^{-3} to 10^{-6}, depending on the examined wavelength.

During a transit event, the presence of an atmosphere (at zero-order approximation it looks like an annulus around the planet) determines the planet's measured radius: it can vary from one wavelength to another, according to the spectral absorption signatures of the atomic or molecular species within the atmosphere.

The amplitude of the absorption caused by the presence of an atmosphere is [Brown, 2001]:

$$\left(\frac{\delta A}{A}\right)_{atm} = \frac{2R_P}{R_\star^2}\frac{k_B T}{\mu g} \tag{2.6}$$

Where $k_B = 1.38{\cdot}10^{-23}\ m^2\ kg\ s^{-2}\ K^{-1}$ is the Boltzmann constant, T the atmospheric temperature, g the planet's gravity, μ mean molecular weight, R_P the planet radius and $R_\star$ the star radius.

A more effective approach to the study of atmospheres would be spectroscopy, rather than photometry in various wavelength ranges. Indeed, it allows covering a wider range of wavelengths compared to photometric band filters.

For transiting exoplanets three atmospheric measurements are possible [Sing, 2018]: transmission spectra – essential to detect atmospheric absorption features during a transit, emission spectra – to detect the day-side average emission during eclipses, and phase curves – to map the spectral emission of the planet following it around its orbit around the star.

Transmission spectra

When analyzing a spectrum of a transiting planet, we observe the absorption of the incoming stellar radiation caused by the cooler planetary atmosphere. The change in intensity is, in general [Rybicki et al., 1981]:

$$\delta I_S = L I_S (B_{ul} N_u - B_{lu} N_l) + L A_{ul} N_u \tag{2.7}$$

Where L is the path length through the atmosphere; the B coefficients are the Einstein B coefficients for absorption from the lower to the upper state and vice-versa; similarly, A is the Einstein coefficient for spontaneous emission, and $N_{u/l}$ are the number densities in the upper and lower quantum states.

With this method it is possible to observe many phenomena such as escape processes, winds, and hazes, and it should be possible to make a description of the atmosphere; however, this could be particularly challenging because the instrumental or telluric variability generates a time-depending noisy signal that is hard to remove, or because the host star is an active one with a time-depending flux. Errors like those can distort the shape of the transmission spectrum [Madhusudhan et al., 2014].

Furthermore, transit spectra refer to the terminator of the planet, where considerable variation in temperature, density and cloud condition could appear. Clouds, in particular, present themselves as a challenge as well since they flatten the spectrum masking or damping all features of the lower regions of the atmosphere, thus making harder the characterization of the atmosphere itself (see Figure 2.7).

Emission Spectra

Secondary eclipses are used to obtain the planet's emission/reflection spectrum, by simply recording the difference between the star + planet spectrum before (or after) and during the eclipse, where only the stellar flux appears. Doing so, it is possible to distinguish particular molecular signatures in emission or absorption (mainly in

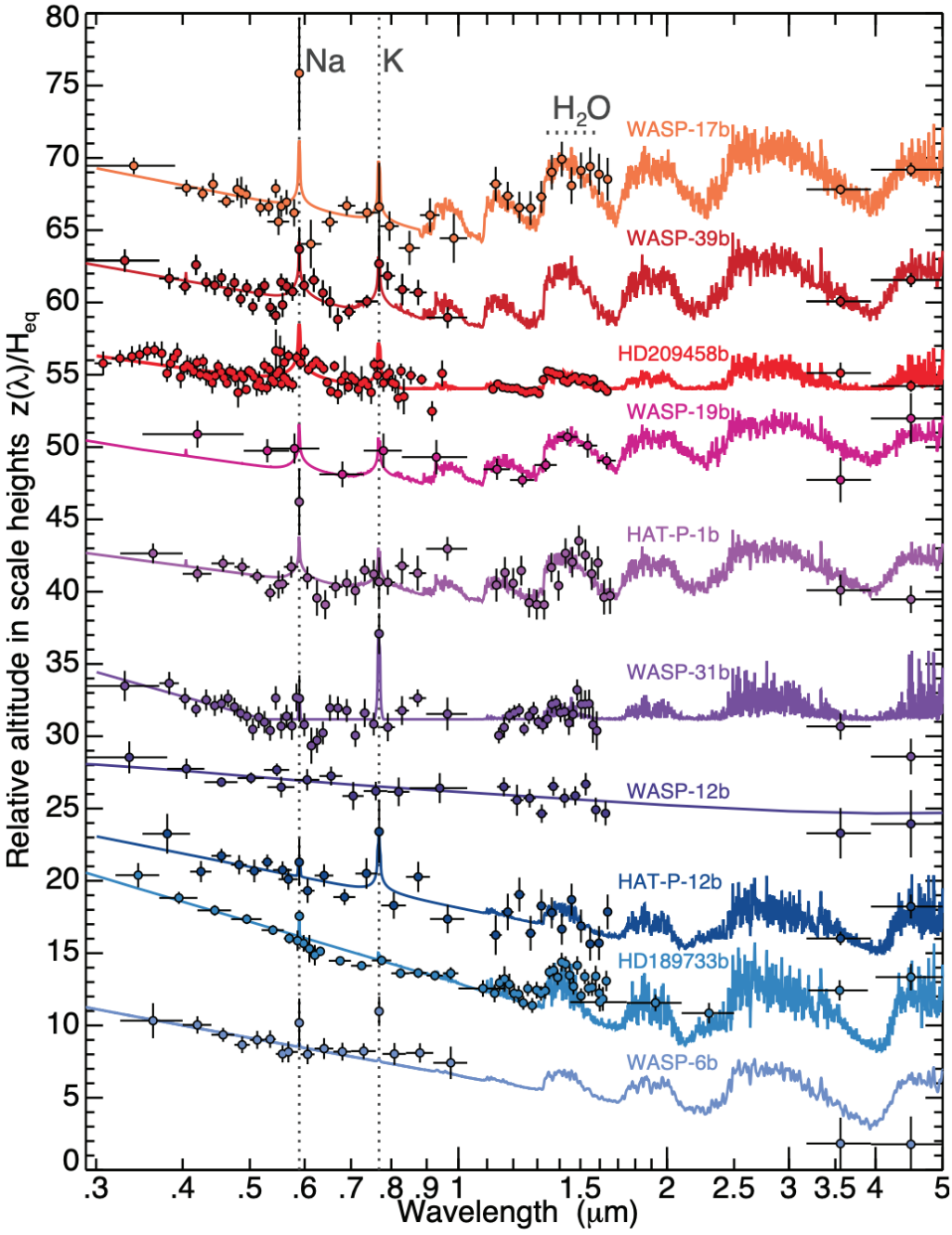


Fig. 2.7: HST/Spitzer transmission spectral sequence of hot-Jupiter survey targets [Sing et al., 2016].

the near- and mid-infrared) on the planet's dayside, depending on their formation region. By observing the phenomenon in the visible range, it could be possible to have information about reflected or scattered light. This technique has revealed to be effective especially for close-in planets, which are more likely to be in a tidally locked configuration: in this way, the phase should be invariant during subsequent eclipses [Alonso, 2018]. From the study of secondary eclipses, it is possible to have information concerning the eccentricity of the orbit, as well as the thermal dayside structure of the atmosphere.

High-resolution analysis of the ingress and egress from the eclipse allows the study of the spatial structure of the planet. This technique is called eclipse mapping and it has already allowed a 2-D study of a few candidates, retrieving information concerning the substellar point and how it can be shifted due to the presence of advective processes in the atmosphere (see e.g. Fig.2.8).

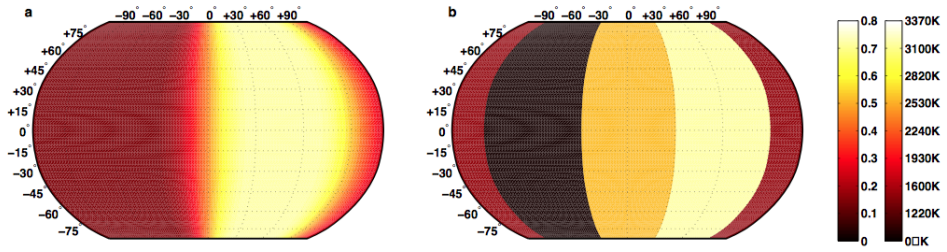


Fig. 2.8: Longitudinal brightness maps of 55 Cancri e [Demory et al., 2016].

Phase Curves

The analysis of planetary targets in a synchronous rotation during every phase of its orbit is particularly interesting since it allows the observer to analyze all sides of the planet as it revolves around the star. A complete phase curve can be then produced and used to infer information concerning the atmospheric composition, the presence of clouds and the nature of advective processes that occur in the atmosphere itself [Deming et al., 2019]. The optimal wavelength range changes depending on the observables [ibid.].

Secondary eclipses are indeed more interesting in the infrared band, to have information concerning the planetary emission (the visible part of the spectrum being generally dominated by the reflected light from the host star). The expected spectrum should be composed by absorption and emission features, depending on the chemical and physical composition of the atmospheric region that produces them. Low, hot, thick regions of the atmosphere are most likely to emit in the infrared, while cooler, thinner regions are more likely to absorb radiation.

For what concerns transit spectroscopy, it relies on the star as the main light source, so in general the optical/near-infrared band is preferred. An exception is due to the cooler M dwarfs, which emit massively in the infrared region, so the spectral range of an instrument that aims to observe planetary companions around such host stars should be changed accordingly. Since the planetary atmosphere is supposed to be much cooler than the host star, only absorption features are expected in transmission spectra. If emission features are present, then a thermal inversion may have occurred in the atmosphere [Madhusudhan, 2018].

There is a non-null chance of detecting false positives in this technique, in the case when other objects may produce similar transit-like events. These can be produced by:

- grazing binaries, companions that have similar masses and that are inclined at peculiar angles with respect to the line of sight to exhibit grazing stellar eclipses. This causes a signal whose intensity is comparable to the one expected from a transiting planet, but the phase curve is more V-shaped. Such cases can be identified by studying the value of the impact parameter, as well as the self-luminosity of the eclipsing object (by multi-band photometry);

- low-mass (sub)stellar companions that transit in front of the main star. This produces similar flat-bottomed transits.

- background eclipsing binaries appearing randomly in the field of view.

- hierarchical triples, where there is a faint eclipsing binary gravitationally bound to a brighter star. If the distance between the companions is short enough, the eclipse could be misinterpreted as a planetary transit in front of the brighter star.

- stellar activity could induce some variations in the observed flux. Uncovered, rotating spots on the stellar disk may mimic the effect of a transit, as well as reduce the stellar flux at shorter wavelengths, thus giving the misleading idea of the presence of hazes on the exoplanetary atmosphere; on the other hand, if the planet passes in front of a spot, the global luminosity is perceived to vary less due to the lower luminosity of the covered region, thus resulting in a bump in luminosity in the phase curve.

In any case, to confirm a transiting candidate, further follow-up observations are due. Radial velocities prove themselves to be a valuable technique to achieve this goal. However, sometimes the stellar targets can be too faint to be observed with

ground-based spectrographs, and therefore the confirmation of a candidate could be challenging.

Many transit surveys were accomplished in the last few years: starting from CoRoT (COnvection, ROtation et Transits planétaires) [Viard et al., 2006], then the Kepler Space Telescope [Borucki et al., 2010] (and later K2 [Howell et al., 2014]), Super WASP at the Observatorio del Roque de Los Muchachos on the island of La Palma in the Canary Islands, and at the Sutherland Station of the South African Astronomical Observatory [Pollacco et al., 2006], HATNet (Hungarian Automated Telescope Network) [Bakos et al., 2002], KELT (Kilodegree Extremely Little Telescope) [Pepper et al., 2007] at Winer Observatory, XO [Crouzet, 2018], TRAPPIST (TRAnsiting Planets and PlanetesImals Small Telescope) at ESO's La Silla Observatory and Oukaimden Observatory in Morocco [Jehin et al., 2011].

Present and future facilities are TESS (Transiting Exoplanet Survey Satellite - launched in 2018) [Ricker et al., 2015], CHEOPS (Characterizing Exoplanets Satellite - to be launched in 2019) [Broeg et al., 2013], PLATO (PLAnetary Transits and Oscillations of stars - to be launched in 2024) [Rauer, Aerts, et al., 2016], ARIEL (Atmospheric Remote-sensing Infrared Exoplanet Large-survey - to be launched in mid 2028) [Tinetti, Drossart, et al., 2016], and JWST (James Webb Space Telescope - to be launched in 2021) [Gardner et al., 2006].

2.2.3 Direct Imaging

This technique is based on space-resolved imaging, thus allowing the astronomer to observe the planetary companion directly and not using its reflex motion on the host star. This allows measuring size, physical and chemical properties of the planet, as well as orbital and rotational parameters. It is particularly sensitive to massive ($M > 1000\ M_{\oplus}$) planets in wide orbits ($a > 5$ AU), to have a wide angular separation from their stars.

Strong thermal emission from the planet is required to be able to detect a planet with this technique: thus, young planets are privileged. It allows as well to analyze protoplanetary disks, to provide substantial information concerning the formation and evolution of exoplanetary systems.

When attempting to detect an exoplanet via direct imaging, many issues are to be solved: first of all, to detect an object which is by various orders of magnitude fainter than its host star, the light coming from the main source should be somehow blocked, as well as any diffraction pattern that could be generated by the instrumentation;

furthermore, if the observing telescope is ground-based, all scattered light from the atmosphere needs to be suppressed as well. Such atmospheric noise sources are known as *speckles*.

For these purposes, advanced technologies such as adaptive optics and coronagraphy are applied. Wavefront detectors identify the shape of the wavefront disturbed by the atmosphere, instructing deformable mirrors to assume shapes which should allow the wavefront to be parallel, just as it was before crossing the atmospheric medium.

Other techniques can be used, such as the Angular Differential Imaging (ADI) for which the telescope is rotated about the line of sight during contiguous observations. Speckles do not vary their position even though the field of view rotates, so they can be identified and removed. The Simultaneous Spectral Differential Imaging (SSDI) can be used to remove those noise sources since speckles are wavelength-dependent as well. By comparing the acquired images at a different wavelength, it is thus possible to discern the atmospheric noise and the scientific objects (i.e. stellar or planetary companions): as a matter of fact, speckles change their distance from the star depending on the wavelength, while real objects don't [Traub et al., 2010].

The light from the parent star can be removed with coronagraphs, devices which block light from the center of the telescope beam while permitting light from surrounding sources to pass through. Coronagraphs normally use a Lyot architecture, which was developed for studies on the solar corona. Phase/amplitude masks are introduced on the input pupil plane, on the intermediate focal plane, and on the output plane, being shaped and optimized to be most effective exactly in the region around the star where the planetary companion is expected [Guyon, 2018].

Direct imaging allows photometric and spectrometric measurement of the observed planet and has currently discovered many planetary systems.

New imagers are currently in a developing phase, in order to improve sensitivity. Atacama Large Millimeter Array (ALMA)[3] has allowed the study of protoplanetary disks and could possibly reveal structures and "proto-planets"; Spectro-Polarimetric High-Contrast Exoplanet Research (SPHERE) at VLT has recently seen its first light and it is starting to locate some candidates [Beuzit et al., 2019]; the high-contrast Gemini Planet Imager (GPI) has achieved contrast performance that is an order of magnitude better than conventional adaptive optics imagers [Macintosh et al., 2014]. In a few years, JWST will also contribute to other discoveries, as well as

[3] http://www.almaobservatory.org/

SHARK (System for coronagraphy with High order Adaptive optics from R to K band) VIS-NIR imagers at LBT [Farinato et al., 2018].

2.2.4 Microlensing

Gravitational microlensing detects planets by analyzing their gravitational influence on the photons coming from a farther background star [Tsapras, 2018].

This is in agreement with Einstein's Theory of General Relativity, for which gravity can create a curvature in the space-time, deflecting the path of light rays that pass close to a massive object. This effect is called *gravitational lensing*. As a consequence, the astronomer may observe multiple, distorted images of the same source object which is instead farther back in space with respect to the object that acts as a lens. If the source and the lens are in perfect alignment, all images merge to form a bright ring around the lens, the Einstein ring.

Gravitational microlensing appears when both the lens and the source are stars, and the angular distance between the images is of the order of milli-arcseconds. Thus, with current instruments, only the variation in brightness of the source can be observed, rather than spatially resolved images (see Figure 2.9).

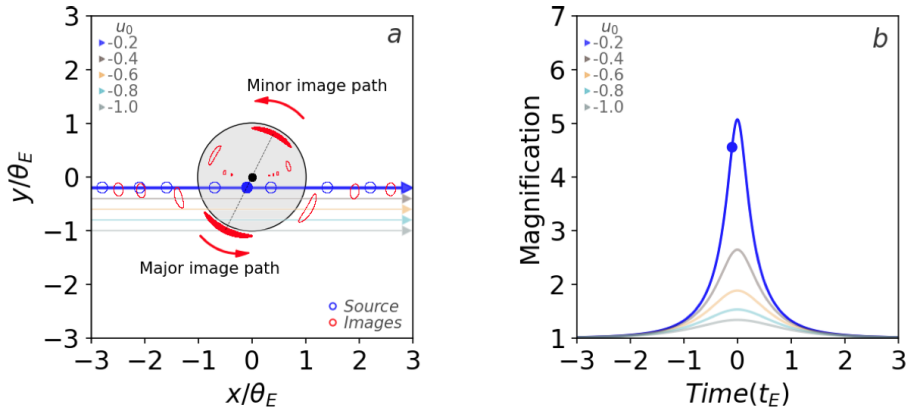


Fig. 2.9: *Panel a*: Geometry of a face-on microlensing event. In black, the lens star; the edge of the grey circle is the Einstein ring; arrows in various colors, different source trajectories; red figures, produced images. *Panel b*: total magnification in time for the same trajectories shown in Panel a. Source: Tsapras (2018).

If a planet orbits the lens, it could be revealed when the image generated by the lensing effect sweeps past its location, generating binary-lens perturbations on the light curves, called *planetary anomalies*. Because of the geometry of the phenomenon, microlensing is indeed effective in detecting planets around the Einstein ring, which

is proportional to the distances of both lens and sources from the observer, as well as the mass of the lens object itself. Typically, this corresponds to semi-major axes of about 1-10 AU [ibid., and references therein].

Microlensing events normally last a few weeks to a few months, depending on the proper motions of the lens and the source stars. Planetary anomalies caused by the orbiting planets happen on shorter timescales, days or hours, but causing substantial variations. It is, however, a rare phenomenon, with only one star in every million experiencing a microlensing event in the Galaxy: for this reason, microlensing surveys focus on the Galactic bulge, where is the highest density of stars per square degree. It is thus possible to observe 3-5 new microlensing events every year. Since they are all one-off occurrences, a network of complementary observation needs to be set up to be able to sample the event when first detected.

This is a powerful technique, insensitive to the host star type and cheap in terms of resources: even 1m-class telescopes can be suitable for high-quality observations.

Past surveys such as OGLE (Optical Gravitational Lensing Experiment) [Udalski et al., 1994] at Las Campanas Observatory in Chile, MOA (Microlensing Observations in Astrophysics) [Hearnshaw et al., 2006] at Mt John University Observatory in New Zealand, and KMTNet (Korea Microlensing Telescope Network) [Kim et al., 2016] at the Cerro-Tololo Inter-American Observatory (CTIO) in Chile were efficient in finding some exoplanets in the interesting region of the Jupiter-sized masses around the snow line threshold. Prolonged usage of such facilities may help have a higher understanding of how gaseous planets are formed.

In 2005, a cold Super Earth (OGLE-2005-BLG-390L b) around an M-dwarf was discovered with this technique [Beaulieu et al., 2006]. Such class of planets is indeed interesting for exoplanetary systems evolution studies as well, but not ideal candidates for the quest for life in the universe. Due to its unique target (Super Earths and Neptunes around the snowline of the host star), further missions include a microlensing program, such as NASA WFIRST [Spergel et al., 2015].

2.2.5 Transit Timing Variations

A transiting planet generally transits at perfectly periodic times. If another (possibly non-transiting) planet is present in the system, then the transit periodicity is perturbed by the gravitational tug of the companion.

In this way, further planets orbiting a system with a known transiting planet can be discovered. Furthermore, precise information concerning the masses of the planetary candidates and their relative eccentricities can be determined with great precision.

2.2.6 Pulsar Timing

Pulsar timing allowed Wolszczan et al. (1992) to find the very first two exoplanets, two Super-Earths named PSR B1257+12 A and B (and later on, a Moon-sized planet PSR B1257+12 C). Those candidates orbit around a pulsar, a compact object that continuously rotates with a specific, extremely precise period.

Such class of planets can be detected by studying the irregularities in the arrival time of the pulses of the host compact objects, otherwise extremely precise.

2.2.7 Astrometry

Astrometry aims to measure the star's position in the sky over a long time span, thus detecting the wobble caused by the motion of the primary star around the center of mass of a potential stellar-planetary (or stellar-brown dwarf) system.

The apparent semi-major axis of the host star orbit induced by a planet is given by the following equation (see Malbet et al. (2018)):

$$\Delta a = 0.33 \left(\frac{a}{1AU}\right) \left(\frac{M_P}{1M_\oplus}\right) \left(\frac{M_\star}{1M_\odot}\right)^{-1} \left(\frac{d}{10pc}\right)^{-1} \mu as \tag{2.8}$$

Where a is the planetary semi-major axis, M_P the planetary mass, $M_\star$ the stellar mass, d the distance from the observer. A Jupiter orbiting a Sun-like star at 10 parsecs would yield an astrometric signal of nearly 500 μas, while a hot Jupiter orbiting the very same star would cause a much smaller variation ($\approx$ 10 μas). The magnitude of the signal dramatically decreases for telluric planets up to fractions of μas. Due to the inverse dependence with the distance from the observer, this method could be effective only for nearby stars.

Instruments such as STEPS (Stellar Planet Survey) [Pravdo et al., 2003] at the Palomar telescope, CAPSCam camera [Boss et al., 2019] on Las Campanas telescope, and FORS2 (FOcal Reducer/low dispersion Spectrograph 2) [Appenzeller et al., 1998] at VLT revealed a few nearby brown dwarfs. Apart from that, astrometry has proven to be still extremely challenging for current technologies. However, the

Gaia space mission [Gaia Collaboration, Prusti, et al., 2016] is starting to give some results in this field as well, other than further characterize the proper motions and orbital parameters of known planet-bearing stars.

2.3 Exoplanet Occurrence

At the time of writing, about 4000 exoplanets are currently confirmed, and almost 3000 planetary candidates are being analyzed. Recently launched missions such as TESS are discovering new candidates every day, thus making the exoplanetology field a mature branch of astrophysics. Due to the larger sample, statistical studies on exoplanets are more and more robust.

A few decades ago, when only the Solar System planets were known, the astronomical community assumed that massive, gaseous planets could be formed far from their host star, beyond the snow line. This was proved wrong by the discovery of the first exoplanet orbiting a main-sequence star, 51 Peg b [Mayor and Queloz, 1995], a Hot Jupiter orbiting at 0.05 AU from its star.

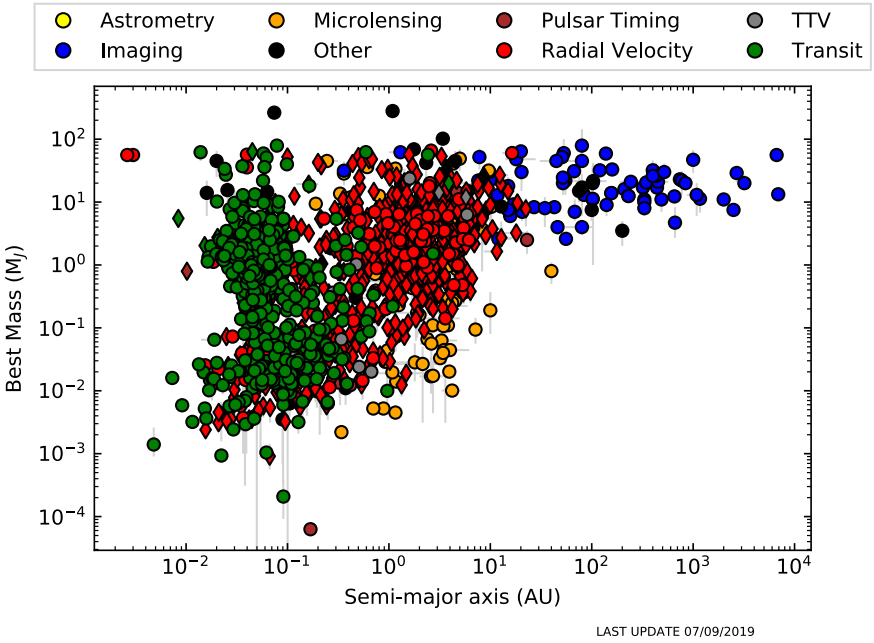


Fig. 2.10: Mass vs. Semi-major axis diagram of currently known exoplanets obtained with `Exo-MerCat` (see Section 5).

Since then, a large fraction of exoplanets with similar features was discovered, as can be seen in Figure 2.10. The most effective methods for the detection of such targets are radial velocity and transit surveys, which indeed represent the most massive samples. Microlensing has proven to be effective for Neptune-like planets at 1-10 AU, while the direct imaging technique is essential to detect young gaseous planets at a farther distance from the host star.

The absence of detection in the Jovian and sub-Jovian range at large distance from the star, similar to the gaseous planets in our system, is most likely due to observational biases: such planets would cause a variation in radial velocity amplitude or transit depth too small to be detected with current facilities, or they could be much less bright with respect to their host stars to be detected with direct imaging techniques. We must then rely on the next generation of instruments, whose increase in sensibility should allow to cover this population as well.

The occurrence of Jupiters appears to fall for masses above 10 M_J and distances lower than a few AU: the lack of detected planets in the range 10-80 M_J is often known as *brown dwarf desert*. Few Jovian objects are found in the 0.1-1 AU semi-major axis range, thus forming a *period valley* which is probably caused by the different formation and/or migration mechanisms occurring in hot and temperate/cold giant planets [Santerne, 2018]. There appears to be a desert of short-period low-mass giants as well, probably the result of a dramatic evolution or migration of giant exoplanets [ibid.].

Statistically, 15% of the Sun-like stars host one or more planets with a minimum mass between 3 and 30 Earth masses orbiting within 0.25 AU [Fischer et al., 2014]. Small stars (M-dwarfs) have the higher probability of hosting at least one terrestrial planet; with current facilities, it is possible to find candidates orbiting in the habitable zone of such stars (see Section 4). It is, on the other hand, unlikely to find giant planets orbiting M stars, at least for periods shorter than a few years – while their occurrence is instead 2-3 times higher for FGK stars.

Within the Kepler sample, nearly a fourth shows evidence for more than a single transiting planet. The majority of the currently known systems lie on coplanar orbits as well, with mutual inclinations of less than 3 degrees: this is in agreement with a formation scenario lacking massive dynamical perturbations in the protoplanetary disk. Many multiple-planet systems are closely spaced and often in mean-motion resonances as well [Winn, 2018].

There appears to be also a planet-metallicity correlation, for which metal-rich stars are more likely to host giant planets within 5 AU; close-in terrestrial planets, with

periods up to 10 days, appear to be more abundant in metal-rich environments: this correlation is probably caused by the higher efficacy in assembling a massive solid core in metal-rich protoplanetary disks.

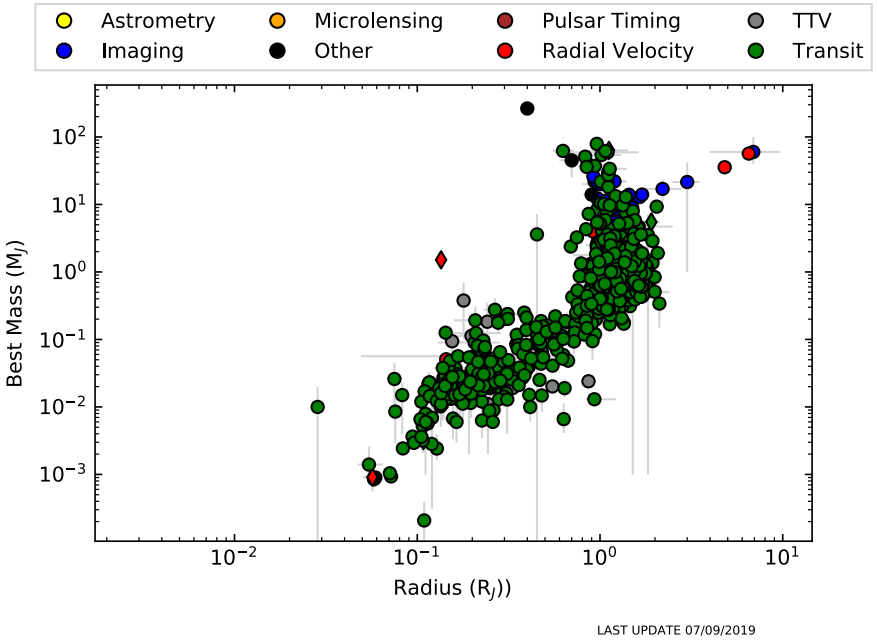


Fig. 2.11: Mass vs. radius diagram of currently known exoplanets obtained with `Exo-MerCat` (see Section 5). Circles show the mass measurements, diamonds the minimum mass estimates; various colors label the planet discovery method.

From Figure 2.11, it is clear that a new population of planets whose size is intermediate between Earth and Neptune (i.e. the Super Earths) is fairly common in extrasolar systems, even though it is absent in the Solar System. On the other hand, Neptunian planets in the 0.1-0.2 M_J range are fairly uncommon. This may be caused by formation mechanisms that lead to runaway accretion of planetesimals of this size.

The sample of masses and radii of currently known planets distributes on a diagonal band, from low-mass/small-size to high-mass/large-size. Some massive planets are largely inflated because of the effect of poorly understood heating mechanisms, causing their radii to be larger than the values predicted by the simple hydrogen model [Fischer et al., 2014].

With the Doppler technique is possible to retrieve information about the orbital eccentricity of the planet (Figure 2.12): it has been found that many of the close-in planets have eccentricities lower than 10^{-3}, while farther giant planets can have

eccentricities up to 0.8-0.9; also, multiplanetary systems are compatible with small planets and low eccentricities because of dynamical stability reasons [Winn and Fabrycky, 2015, and references therein]. This method has also allowed the discovery of many planets orbiting around one or both stars in binary systems.

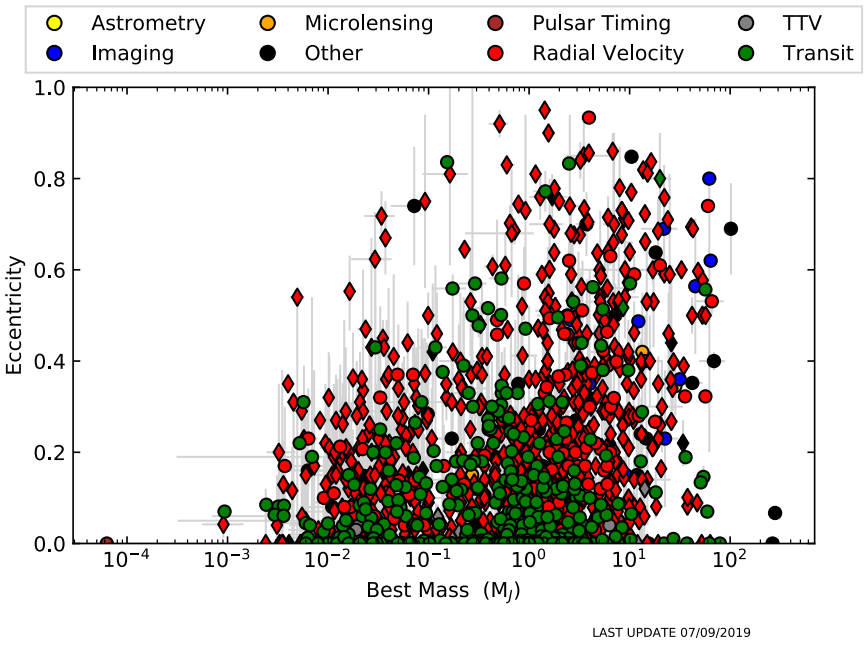


Fig. 2.12: Mass vs. eccentricity diagram of currently known exoplanets obtained with `Exo-MerCat` (see Section 5). Circles show the mass measurements, diamonds the minimum mass estimates; various colors label the planet discovery method.

As visible in Figure 2.13, the planets that lie closer to their host stars are more likely to have very small eccentricities. This suggests the presence of tidal circularization mechanisms, which create tidally locked planets (i.e. planets facing the same hemisphere to the star throughout the orbit). The tidal interaction induced by the host star is more effective in a shorter timescale the closer the planet is. Farther planets are, on the other hand, likely to possess non-null eccentricities, possibly caused by migration effects or perturbing stellar or planetary companions.

Indeed, the statistical analysis is strongly conditioned by observational biases: as previously mentioned, some regions of the mass-period parameter space can hardly be measured with current facilities, because of the faint signal that such exoplanets may produce, or reflect on their star. However, future ground- and space-based missions are expected to improve our knowledge concerning the occurrence rate of exoplanets in the universe. The next decade will be pivotal in the statistical study of this ever-growing sample.

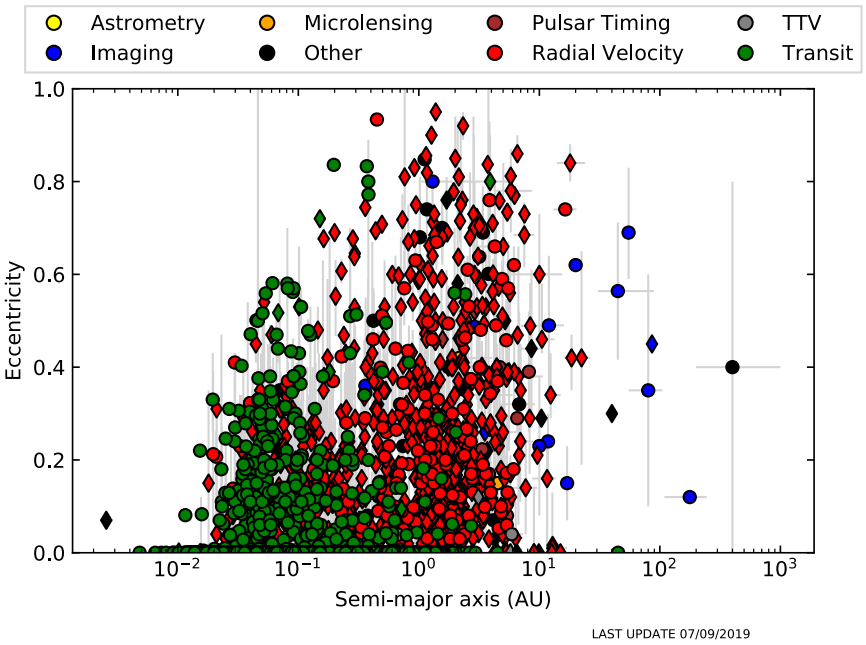


Fig. 2.13: Eccentricity vs. Semi-major axis diagram of currently known exoplanets obtained with `Exo-MerCat` (see Section 5). Circles show the mass measurements, diamonds the minimum mass estimates; various colors label the planet discovery method.

3 Super Earths

> *To get a sense of these distant planets as worlds we'd really like to know something about their atmospheres [...] and we can actually do that, but we have to do something kind of terrible: we have to take a sunset, the most beautiful, ineffable phenomenon celebrated in countless works of art, and we have to try and quantify it and simplify it in a few numbers! I am really sorry about this, but... not that sorry.*
>
> — **Zachory Berta-Thompson**
> (American astrophysicist)

This chapter summarizes the current knowledge about Super Earths. A particular focus is given to the atmospheres of terrestrial exoplanets, the processes that occur and how they are likely to evolve – on the solid grounds of Earth's atmospheric studies.

3.1 Scientific Interest

Extrasolar Super Earths (1-10 $M_\oplus$, 0.8-2 $R_\oplus$) have been discovered by recent surveys such as Kepler and are particularly interesting since there is no analog in our Solar System: this is intriguing, because many of the observed exoplanetary systems contain at least a Super Earth, so the lack of it could question the robustness of the current formation models applied to the Solar System itself.

As visible in Figure 3.1, many detected Earths have relatively short orbital periods and are therefore very close to their host stars: they thus experience high irradiations, as well as tidal torques, which may pose some constraints on the formation processes, as well as the potential habitability of such targets.

These planets are predicted to have large surface gravities (25 $m\,s^{-2}$ for 5 $M_\oplus$ - Miller-Ricci et al. (2009) and references therein) and are likely to exist within a wide range

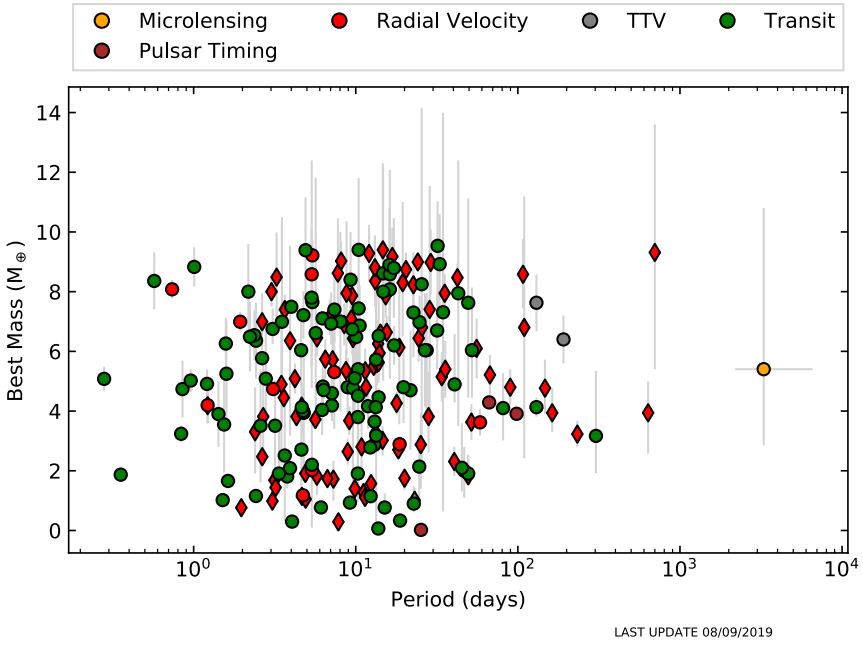


Fig. 3.1: Period-Mass diagram for currently known Super Earths.

of atmospheres: some of them could be able to retain a thick H-rich atmosphere, whereas others could have a stronger resemblance to Earth itself with heavier molecules in their atmospheres. A third possibility could be an atmosphere with a moderate abundance of hydrogen due to its escape and/or molecular hydrogen outgassing Miller-Ricci et al. (2009).

The bulk composition of Super Earths could help understand planet formation, while the study of their atmospheres can give information about few of the most fascinating questions in human minds: are we alone in the universe, or are there other life forms on other planets? How could we gather "life fingerprints" when observing an exoplanet?

Current technologies allow detection of Super Earths, but it is still improbable to spot Earths. However, Kepler discovered Kepler-186 f, a planet with a radius similar to the Earth one and probably in the habitable zone (Section 4) of its host star, while Anglada-Escudé et al. (2016) recently announced the discovery of Proxima b, a planet with a minimum mass of 1.3 $M_{\oplus}$ orbiting Proxima Centauri, the Sun's closest stellar neighbor and one of the best-studied low-mass stars.

Another notable discovery was the system of seven terrestrial planets orbiting the M dwarf TRAPPIST-1, some of them within the habitable zone of their star as well.

Furthermore, water vapor was recently detected in the spectrum of a Super Earth (K2-18 b, by Tsiaras et al. (2019) and Benneke et al. (2019)).

Future facilities will allow an easier detection of such targets, both indirectly (via transit detection and radial velocities) and directly (via imaging techniques): it could be therefore probable to spot an exoplanet very similar to Earth in the near future, and perhaps be able to assess the presence of life on its surface.

The Chapter will be thus structured: Section 3.2 deals with the mechanisms that allow a Super Earth to be formed; Section 3.3 describes the predicted interior compositions of such planets. Finally, the focus of the Chapter will undoubtedly be on Super Earths Atmospheres in Section 3.4, with discussion on the vertical structure, the energy balance and transport, the chemical composition, condensation, escape, and evolution.

3.2 Formation

When a protostar is formed by gravitational collapse in a molecular cloud, the surrounding clump may turn into a circumstellar disk, because of the conservation of angular momentum. The circumstellar "protoplanetary" disk itself is believed to be the birthplace of planets.

Initially, dust and gas are well-mixed but, later on, the first tends to settle into the disk midplane [Izidoro et al., 2018, and references therein]. By the first 3 Myr from the its formation, the disk loses mass by XUV-driven photoevaporation, as well as accretion onto the central star. This poses a threshold in the formation timescale. Because of this, many theories are currently being considered for planetary formation, their effectiveness estimated and corroborated by observed data.

The formation process can be divided into steps: different physical and chemical processes drive the various stages of growth, from dust to fully formed planets. The four main steps are the following [Haghighipour, 2013; Izidoro et al., 2018]:

1. *From particles to pebbles*.

   The solid component of a disk is not uniform throughout the radial direction: only species with condensation/sublimation temperatures below the local temperature can exist as solids at any given radius. The cooling of the disk in time changes the condensation lines of each species; in general, the inner

part of the disk is rich in iron and silicates, while the outer part shows a higher abundance of ice and other volatiles.

At this point, micrometer-sized dust grains undergo random motions, with velocities strictly coupled to the gas and subject to the pressure gradient. When such particles collide with small relative velocities, they stick together forming greater structures, held together by adhesive Van der Waals forces. Dust particles rich in volatiles experience larger adhesive forces and grow rapidly.

2. *From pebbles to planetesimals*.

   Collisional processes are not effective in growing dust grains up to meter- and kilometer-size scales: on the contrary, larger colliding dust particles may bounce off each other or fragment in smaller pebbles instead of growing. This is known as the "meter-size barrier" or "bouncing barrier".

   However, larger planetesimals do exist. The process that explains this growth is still debated. Energetic impacts may lead to mass transfer and allow the production of decimeter- to meter-sized pebbles, but this will lead to an increase in the experienced gas-driven drag and would cause such objects to spiral inward rapidly, falling towards the star. Another plausible option would be the gravitational collapse of clusters of small pebbles, but such locally concentrated environments would be probably disrupted by turbulent diffusion. Different mechanisms may cooperate to concentrate particles locally to cause gravitational collapse. Alternatively, there may be pressure bumps in the disk, higher pressure traps which could generate a positive gradient and accelerate the gas up to super Keplerian speeds, while slowing the drift inwards of the pebbles (*streaming instability*).

3. *From planetesimals to embryos*.

   At this stage, interactions are merely gravitational. Planetesimals may grow in two different regimes, by collisions or accretion.

   In the first regime (planetesimal-planetesimal growth), close encounters increase the speed of the planetesimals, deflecting their orbits and causing them to collide more often. Depending on the rate of growth of larger and smaller planetesimals, it is possible to have "runaway growth" (i.e. larger planetesimals grow faster than smaller ones), "orderly growth" (i.e. planetesimals grow at the same rate), or "oligarchic growth" (an intermediate scenario).

   In the second regime (pebble accretion), planetesimals may accrete dust grains from the disk. This process can be much faster than planetesimal accretion

and can lead to sufficiently large planetary embryos (up to 100 times the sizes of the individual planetesimals). The growing body opens a gap in the disk and creates a pressure bump outside the orbit that eventually stops them from growing further.

4. *From embryos to planets*.

   At this stage, the protoplanetary disk has dissipated, removing the gas dynamical friction. Massive embryos accrete or scatter nearby planetesimals and create large gaps in the dissipating disk. There are still violent collisions between embryos, and the planetesimal population is dramatically reduced.

   At large distances rotational velocities are small, so impacts among planetesimals are likely to result in accretion; furthermore, the bulk material of these objects is primarily ice, which increases the efficiency of their sticking. In these zones nebular gas is still present, so the forming object could attract gas from the surroundings and form a thick envelope: that is how a gaseous giant planet is formed.

   In the inner parts of the disk, collisions may also lead to the formation of larger bodies, but planetesimals are more likely to lose their surface ices and other volatiles during their motion, so coalescence may not be as efficient as the previous case. Simulations show that the results of these collisions may be several Moon- or Mars-sized bodies, instead of big objects such as the cores of giant planets [Izidoro et al., 2018]; many of them could also have eccentric orbits due to the frequent collisions and perturbations by farther giant planets. In this scenario, the formation of an Earth-like planet with a thin atmosphere could take up to several hundred million years.

During the growth of a planetary embryo, tidal torques between the planet and the gas elements can modify the angular momentum of the planet and cause it to migrate [Alibert et al., 2018]. There are two main types of migration:

- *type-I migration* is the migration of low-mass planets that do not modify the disk structure (typically, planetary masses below the Neptune mass). This migration is generally inward and fast, but some simulations show that such planets could migrate outwards depending on the planetary mass, disk density, and temperature profiles;

- *type-II migration* is the migration of massive planets that form gaps in the disk (masses above the Saturn mass). When the gaseous material moves away from

the growing embryo, tidal forces lower their magnitude and, as a result, the planet slowly migrates inwards.

Migration has an important effect on planetary formation and composition. If the migration is slow enough, the accreting planet has more time to collect material from the disk during the inwards spiralization, so larger solids can be accreted. Secondly, the composition may also change depending on the formation environment and location: a planet formed far from the water snowline (the radial location from the star where the temperature allows water to become ice) is more likely to be water-rich, whereas a planet formed close to its star has more chances to be rocky [Martin et al., 2016].

It is still not clear whether Super Earths were formed at large distances as failed cores of giant planets and then migrated to their present locations, or were created in their actual close-in orbits following a similar process as the formation of terrestrial planets. The most trusted theory favors the migration of such class of planets as well: Super Earths should form very fast because of the short dynamical timescales of the disk, and therefore they should migrate and not form *in situ* [Izidoro et al., 2018]. The lack of Super Earths in the Solar System is probably due to the Jupiter barrier, which could have hindered the migration of Neptune and Uranus.

The quantity of volatiles contained in their structure could be essential for habitability purposes. One may think that the more water on the surface or in the crust of the planet, the higher the chance of developing life, provided the right conditions of irradiation from the host star. This is not the case, since too much water may lead to water-planets whose pressure at the seafloor may induce the formation of an icy layer at the surface of the mantle. This layer may forbid the formation of a carbonate-silicate cycle, which is essential for the stabilization of the climate in geological timescales. The large solubility of carbon dioxide in water may also lead to destabilizing CO_2 cycles on ocean planets [Alibert et al., 2018].

Terrestrial-like exoplanets may obtain atmospheres from three sources: the capture of nebular gases, outgassing during accretion, and outgassing from tectonic processes. While the first source is important for giant planets, outgassing is the most efficient phenomenon for Super Earths, because the gaseous nebula is likely to be dissipated by the time of final planetary accretion.

The chemical composition of the forming atmosphere depends on the initial composition of the primitive material. Plausible compositions are those of the meteorites fallen to Earth; they can be divided into two main classes: the primordial *chondrites*,

which contain silicates, iron, nickel and sometimes water- or carbon-rich veins, and the differentiated *achondrites*, which lack metallic iron and water, originated by fragmentation of planetesimals [Elkins-Tanton et al., 2008]. While accreting, they could outgas hydrogen, water, and carbon-oxygen compounds, that undergo subsequent interrelated atmospheric processes such as atmospheric escape, photolysis, and chemical reactions. Some first-order models were made by Elkins-Tanton et al. (ibid.), as shown in Fig. 3.2: reducing conditions during accretion could produce a planet with a metallic core and a water-rich atmosphere, while oxidizing conditions could produce a planet without a metallic core and with a hydrogen-rich atmosphere; also, the initial atmospheric mass could range from 1% to 20% of the planet's mass. Generally, water/carbon compounds should dominate the atmosphere, with traces of hydrogen, nitrogen, and helium.

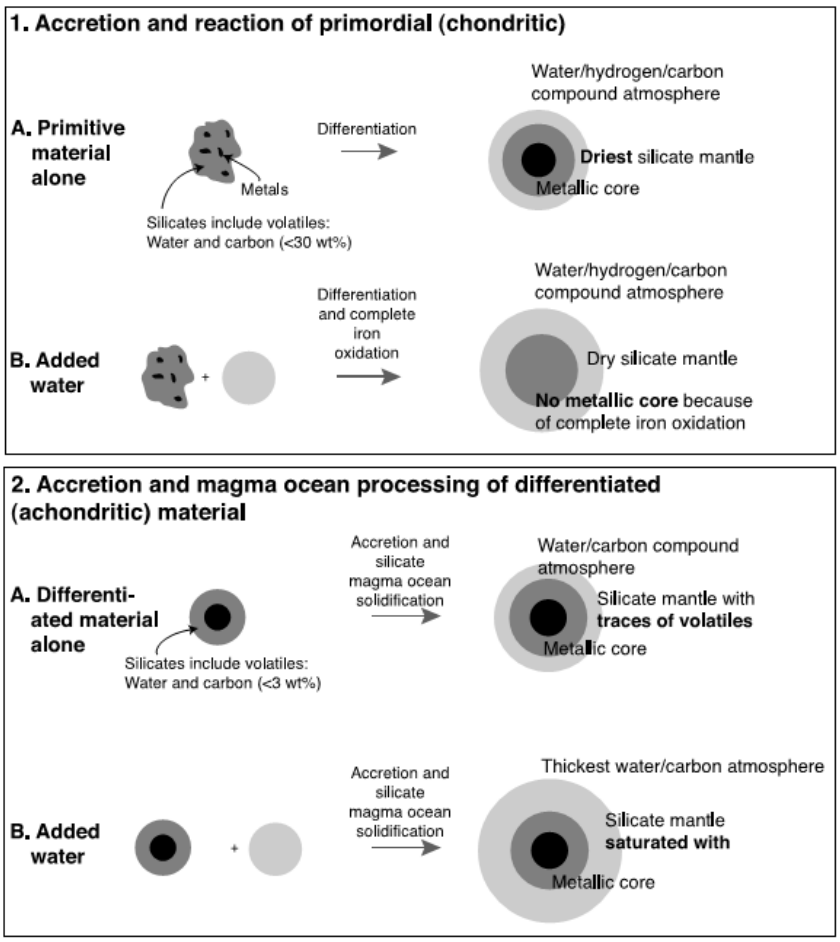


Fig. 3.2: Schematic of the four atmosphere-outgassing models in Elkins-Tanton et al. (2008).

3.3 Interior Composition

Since Super Earths can exist in different environments, the nature of the problem of the interior composition and dynamics of this class of planets is highly degenerate: rocky surfaces as well as ocean worlds are predicted and, actually, already observed (Fig. 3.3). Consequently, discrimination between the cases is needed, implying extremely highly accurate measurements.

The internal structure of Super Earths should be differentiated: it should have a dense hot core of pure iron or Fe compounds; a mantle composed of different silicate shells; a crust or an additional water/ice layer.

Ocean worlds are believed to be formed from icy Neptunes that migrate inwards and melt; the ice composition, therefore, should be similar to that of comets (90% H_2O, 5% NH_3, 5% CO_2) [Vasquez et al., 2010, and references therein].

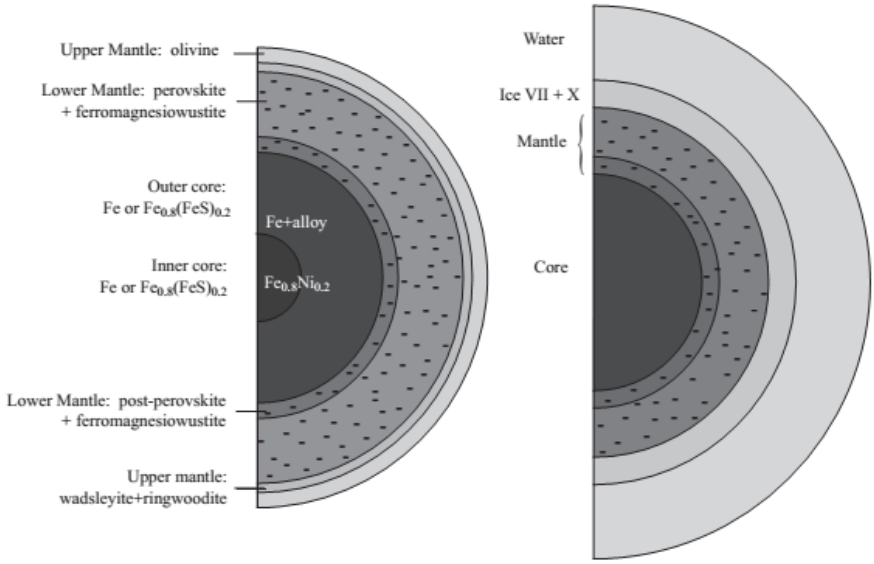


Fig. 3.3: Varieties of Super Earths: rocky and ocean worlds [Valencia, Sasselov, and O'Connell, 2007].

Valencia, Sasselov, and O'Connell (2007) showed that, given a planetary mass, there exists a maximum planetary radius. If observations retrieve values of the radius above this threshold, the target planet either formed beyond the snow line acquiring large amounts of water or was created by the accretion of wet planetesimals. The thickness of the H_2O, mantle and core regions depend on the ice mass fraction (IMF) and core mass fraction (CMF) assumed for a planet. The percentage amount of these three zones can be related to one another using ternary diagrams: each vertex of a triangle means 100% of a particular component and data on the opposite side

means 0%. As shown in Fig. 3.4, minimum and maximum radius exist for a planet: the minimum radius is achieved with the largest core possible from solar nebula composition (60% iron core, 40% silicate mantle) and no water; the maximum radius corresponds to a planet with no core and largest amount of water possible from solar nebula constraints (80% water, 20% mantle). It can be seen that the radius is less sensitive to the amount of mantle mass (points B and D have similar radii in spite of the different percentage of mantle mass).

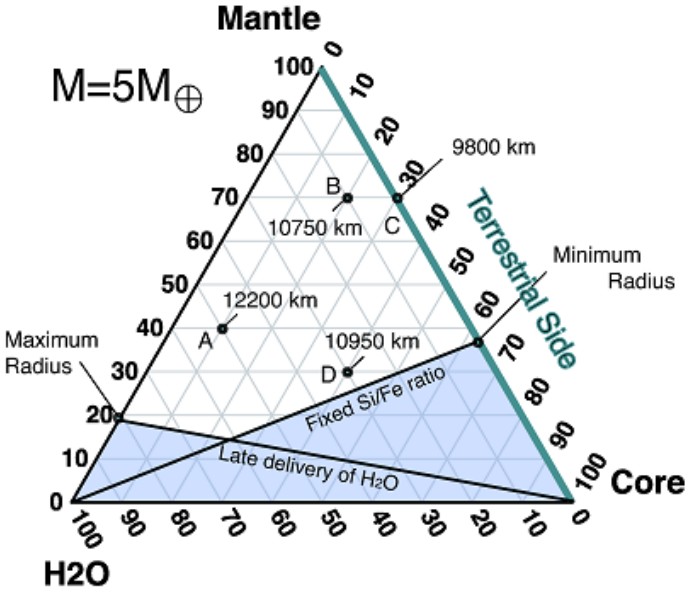


Fig. 3.4: Ternary diagram of a 5 $M_\oplus$ Super Earth. Few examples (A, B, C, D) and corresponding radii are labelled [Valencia, Sasselov, and O'Connell, 2007].

The radius of a Super Earth seems to be linked to the planet mass by a power law relation:

$$R = aR_\oplus \left(\frac{M}{M_\oplus}\right)^\beta \tag{3.1}$$

With different values of a and β deriving from various models and depending from composition: a (~ 1) increases almost linearly with H_2O content; β (~ 0.27) decreases as the amount of H_2O increases [ibid.]. Different internal compositions imply different relations, as can be seen in Fig. 3.5.

In our Earth, most of the energy has two sources: firstly, the potential energy acquired during the accretion process and via impacts; secondly, the radioactive decay of elements such as ^{40}K, ^{238}U and ^{232}Th.

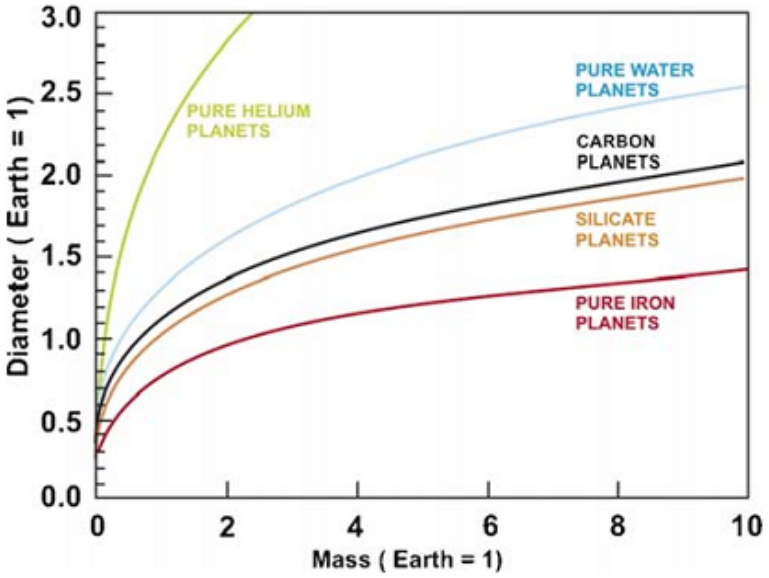


Fig. 3.5: Mass-radius relation for solid planets. (Vasquez et al. (2010) adapted from Seager, Kuchner, et al. (2007)).

The energy equation for the transfer of energy per unit volume is [Vasquez et al., 2010, and references therein]:

$$C\frac{\partial T}{\partial t} = Q_C + Q_H - Q_{conv} \tag{3.2}$$

Where

- C is the heat capacity of the Earth;
- $Q_C = -\kappa\nabla^2 T$ is the heat transfer by conduction (where κ is the thermal diffusivity);
- $Q_{conv} = c_P v \nabla T$ is the heat transfer by convection (where v is the velocity and c_P the heat capacity at constant pressure);
- $Q_H = 9.5 \cdot m(^{238}\text{U}) + 2.7 \cdot m(^{232}\text{Th}) + 3.6 \cdot m(^{40}\text{K})$ is the local heat generation by radioactive decays; it depends on the mass of the radioactive isotopes in units of 10^{17} kg and it is measured in $terawatts$.

Typically, mantle materials are poor conductors of heat and the radioactive heat is usually less than 30% of the one transferred by convection, which is the most effective process involved.

The Rayleigh number[1] for the mantle is:

$$Ra = \frac{g\alpha(T_m - T_s)(R_m - R_c)^3}{\kappa\nu} \tag{3.3}$$

Where g is the gravitational acceleration, α the coefficient of thermal expansion, ν the viscosity, κ the thermal diffusivity, R_m and R_c the outer and the inner radii of the mantle, T_s the surface temperature and T_m the mantle temperature. This number is far greater than 10^3, which is the threshold for convection to exist: for the mantle, values range between 10^6 and 10^8.

Heat is then released on the surface primarily via plate tectonics: the lithosphere is broken in plates continuously moving and interacting with one another, whether diverging and creating new crust or subducting and melting in the mantle. Earth is the only solar rocky planet where this process has been observed, as Venus and Mars have a stagnant lid - a single plate covering the entire planet.

Plate tectonics should be effective (if not more effective, due to bigger size) also in Super Earths, as it is supported by various models [Tackley et al., 2012]. Massive terrestrial planets have larger convective driving forces that should be capable to deform the crust, which in turn could be weaker: this effect contributes to the subduction of the lithosphere and, therefore, to plate tectonics enhancement.

This process has been identified as a mechanism that has enabled the settlement of life on Earth. Indeed, it facilitates long-term volatile cycling between the atmosphere/ocean and solid interior, developing a negative feedback mechanism which stabilizes the surface temperature over geological timescales: water, carbon dioxide, and other volatiles are outgassed by volcanism, while ocean plates subduct and carry these compounds with them back in the mantle. As this effect is probably amplified on Super Earths, these planets could be good habitable candidates.

[1] The Rayleigh number is a dimensionless number associated with the heat transfer for the fluid: when it is below a critical value, heat is transferred via conduction; when it exceeds the critical value, heat transfer is primarily in form of convection.

3.4 Super Earths Atmospheres

3.4.1 Vertical Structure

The vertical temperature structure is fundamental to determine if a planet is potentially habitable. Furthermore, it is the starting point for computing equilibrium and non-equilibrium chemistry to understand how an atmosphere is composed.

Since Earth's atmosphere is the most accessible atmosphere to us, it is possible to take that as an example of a typical terrestrial atmosphere. It is mainly composed by nitrogen (78%) and oxygen (21%) with traces of argon, carbon dioxide, water vapor, neon, helium, methane, nitrous oxide, hydrogen, xenon, and ozone. Such compounds are produced by various biotic and abiotic processes, as shown in Table 3.1.

Tab. 3.1: Earth's atmospheric gases to the ppb level, adapted from Seager, Bains, et al. (2016). We used the abbreviations for *parts-per-million*, ppm, 10^{-6}, and ppb, *parts-per-billion*, 10^{-9}.

Compound	**Formula**	**Concentration**	**Primary terrestrial source**
Nitrogen	N_2	78%	Biology, outgassing
Oxygen	O_2	21 %	Photochemistry, biology
Water	H_2O	1-4%	Evaporation, biology
Argon	Ar	9340 ppm	Outgassing
Carbon Dioxide	CO_2	350 ppm	Outgassing, biology, anthropogenic
Neon	Ne	18.18 ppm	Outgassing
Helium	^4He	5.24 ppm	Outgassing
Methane	CH_4	1.7 ppm	Outgassing, biology
Krypton	Kr	1.14 ppm	Outgassing
Hydrogen	H_2	0.55 ppm	Photochemistry, biology
Nitrous oxide	N_2O	320 ppb	Biology
Carbon monoxide	CO	125 ppb	Photochemistry, biology
Xenon	Xe	87 ppb	Outgassing
Ozone	O_3	10-100 ppb	Photochemistry
Hydrogen chloride	HCl	$\sim$ 1 ppb	Sea salt, biology
Isoprene	C_5H_8	1-3 ppb	Biology
Ethane (hydrocarbons)	C_2H_6	0.2-3 ppb	Fires, oceans, anthropogenic
Benzene (aromatics)	C_6H_6	0.1-1 ppb	Antropogenic, biology
Ammonia	NH_3	0.1-3 ppb	Biology
Nitric Acid	HNO_3	0.04-4 ppb	Photochemistry, biology

It can be divided in different layers, according to the different processes taking place in them (Fig. 3.6):

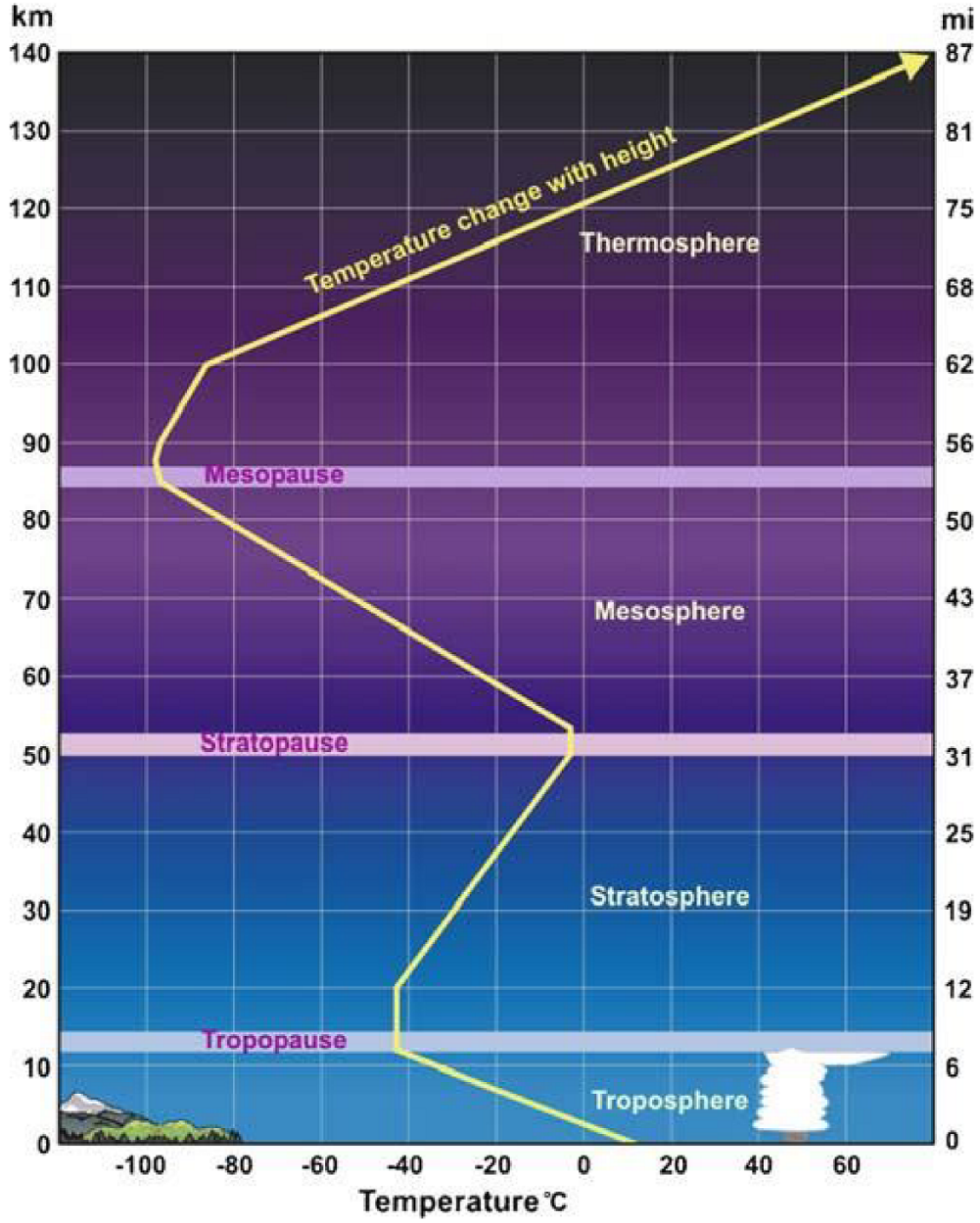


Fig. 3.6: Thermal profile of the atmosphere [Vasquez et al., 2010].

- *Troposphere*: it extends from the surface up to 11 km; it is warmed by the infrared radiation coming from the surface. The heat exchange between surface and troposphere happens by evaporation of water and the formation of clouds, and the greenhouse effect restrains the loss of energy into space. Greenhouse gases (water vapor, carbon dioxide, methane, ozone) absorb heat

radiation more strongly than the rest of the atmosphere [Vasquez et al., 2010, and references therein], contributing to the increase of the surface temperature (see Section 3.4.3.5). In this layer pressure decreases with height with an exponential rate. At first order it is possible to assume hydrostatic equilibrium for an ideal gas

$$\frac{dP(z)}{dz} = -\frac{P(z)\mu(z)g(z)}{k_B T(z)} = -\frac{P(z)}{H(z)} \tag{3.4}$$

Where H the so-called *scale height* thus defined:

$$H(z) = \frac{k_B T}{\mu g} \tag{3.5}$$

In the previous equations k_B is the Boltzmann constant, T the temperature, g the planet's gravity and μ mean molecular weight, all depending from the geometrical altitude z. These equations automatically lead to:

$$P(z) = P_{gr} e^{-z/H} \tag{3.6}$$

Or, inverting for the altitude:

$$z(i) = H \log \frac{P_{gr}}{P_i} \tag{3.7}$$

Where P_{gr} is the ground pressure, while P_i is the pressure at layer i.

In the troposphere, the temperature decreases with height (-10 K per increase of 1 km in altitude). This causes almost all of the meteorological processes to happen: strong convective currents and constant winds are produced by this gradient, enhancing meteorological perturbations.

- *Stratosphere*: this layer extends between 10 and 50 km altitude. Due to the ozone absorption of solar UV radiation, the temperature increases with altitude: a thermal inversion happens.

- *Mesosphere*: this layer extends between 50 and 90 km altitude. Here the temperature reaches a minimum (between -70 °C and -90 °C) so that water vapor freezes and produces ice clouds. Oxidation processes and penetration of UV radiation cause the dissociation of polyatomic molecules: therefore, there are more light gases, such as H and He, than heavy gases.

- *Thermosphere*: this layer extends up to 700 km altitude. UV radiation and photoionization by X-rays cause another temperature increase: the air is so thin that even a small amount of deposited energy causes strong effects in temperature.

- *Exosphere*: in this layer particles have enough kinetic energy to escape from the terrestrial atmosphere, so the atmosphere gradually dissolves into space.

The number densities of atoms and molecules in an atmosphere depend on chemistry and physics, as the abundance of each element varies with respect to temperature and pressure.

In order to better characterize an atmosphere it is useful to define the *volume mixing ratio*, which is the ratio of the number density of a gas to the total number density of an atmosphere:

$$X_{gas} = \frac{n_{gas}}{n_{total}}. \tag{3.8}$$

It is usually altitude dependent when gases are not well mixed throughout the atmosphere, as it is on Earth (Fig. 3.7).

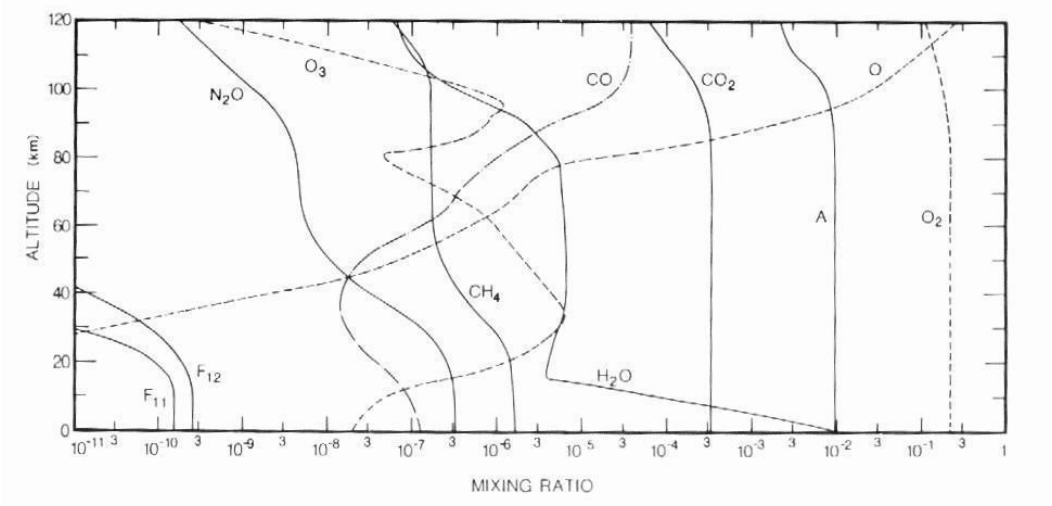


Fig. 3.7: Vertical profiles of the mixing ratios of selected species at the equinox [Goody et al., 1989].

3.4.2 Energy Balance

Planetary atmospheres receive radiation from planets' host stars in the form of photons, which interact with the molecules by absorption, emission, and scattering.

Considering an area dA normal to a given ray of radiation, the *specific intensity* I_ν is the amount of energy crossing dA, within a differential solid angle $d\Omega$, per frequency interval, per unit time [Rybicki et al., 1981]:

$$I_\nu \, [Wm^{-2}sr^{-1}Hz^{-1}] = \frac{dE}{d\Omega dA d\nu dt}. \tag{3.9}$$

Considering now a radiation field and choosing a particular direction $\hat{n}$, the area dA being normal to it, the differential flux in a direction $\hat{k}$ (so that $\hat{n} \cdot \hat{k} = \cos\theta$, θ being the angle between the two vectors) from the solid angle $d\Omega$ is:

$$dF_\nu = I_\nu \cos\theta d\Omega \tag{3.10}$$

And the *net specific flux* F_ν in the direction $\hat{n}$ is obtained by integrating all over the solid angles:

$$F_\nu \, [Wm^{-2}Hz^{-1}] = \int_\Omega I_\nu \cos\theta d\Omega. \tag{3.11}$$

To study exoplanets, it is useful to define other quantities. First of all, we need to specify that a Lambert surface is a surface that scatters intensity isotropically. The apparent surface brightness is the same from any viewing angle: to a variation in the viewing angle corresponds a variation in the number of photons traveling in that direction.

The *surface flux* F_S is the flux emerging from a particular location of the planet "surface", which is the layer where most of the radiation originates without further interactions with the particles in the atmosphere. For a terrestrial planet, "surface" refers to the solid external crust, while for a gaseous planet it might refer to the optically thick atmospheric layer. Using spherical polar coordinates, the surface flux is:

$$F_S(\nu, t) = \int_0^{2\pi} \int_0^{\pi/2} I_S(\theta, \phi, \nu, t) \cos\theta \sin\theta d\theta d\phi. \tag{3.12}$$

If the intensity is uniform on the surface, the flux is:

$$F_S(\nu, t) = \pi I_S(\nu, t). \tag{3.13}$$

Furthermore, it is possible to retrieve the wavelength-independent flux:

$$F_S(t)\ [Wm^{-2}] = \int_o^\infty F_S(\nu, t) d\nu. \tag{3.14}$$

The *luminosity* L(t) is the rate at which the planet radiates energy in all directions:

$$L(t)\ [W] = \int_A F_S(t) dA = 4\pi R_P^2 F_S(t). \tag{3.15}$$

Nevertheless, the observer on Earth cannot spatially resolve the flux received from a planet; all the radiation coming from the source is averaged into a single value of flux, called the *observed flux*:

$$F_\oplus(\nu, t) = \left(\frac{R_P}{D_\oplus}\right)^2 F_S(\nu, t) \tag{3.16}$$

Which is a first-order approximation for a very distant planet (this means that $R_P \ll D_\oplus$, where $D_\oplus$ is the planetary distance from Earth).

Irradiation, or *incident flux*, is the amount of energy per unit time that the planet receives from the host star. If the stellar intensity is uniform across the star's surface, the incident flux at the substellar point is:

$$F_{inc}(\nu, t) = \left(\frac{R_\star}{a}\right)^2 \pi I_{S,*}(\nu, t) = \left(\frac{R_\star}{a}\right)^2 F_{S,*}(\nu, t) \tag{3.17}$$

Where $R_\star$, $I_{S,*}$ and $F_{S,*}$ are the stellar radius, the surface intensity and the surface flux of the star, while a is the semi-major axis. Similarly to what happens on Earth, the poles of the planet receive less radiation than the substellar point; considering this, it is necessary to integrate over the surface area of one hemisphere (only half of the planet is brightened by the star, not necessarily the same through time), obtaining the *incident energy* per unit time:

$$E_{inc}(\nu, t) = \pi R_P^2 F_{inc}(\nu, t) \tag{3.18}$$

And, subsequently:

$$E_{inc}(t) = \pi R_P^2 F_{inc}(t). \tag{3.19}$$

For the Earth, the average F_{inc} is known as the "solar constant S_0", and its value is 1367 $W\ m^{-2}$. For any other star, it is common to generalize the term as *stellar constant*. By measuring the flux, one should define the planet's temperature. First of all, it is important to notice that in the planet atmosphere there is no source or sink of energy: all the radiation comes either from the host star or from the planetary interior. Thus, the *energy balance equation* per unit time is:

$$E_{out}(t) = (1 - A_B)E_{inc}(t) + E_{int}(t) \tag{3.20}$$

Where E_{out} is the energy per unit time (or power) leaving the planet, E_{inc} is the stellar energy per unit time and E_{int} is the energy per unit time coming from the planetary interior. Not all the incident radiation heats the atmosphere and radiates back at longer wavelengths: a fraction A_B is scattered back into space – that is the Bond albedo and will be soon explained. Supposing a circular orbit (so that power can be considered time-independent) and using the definitions of luminosity and stellar flux, the equation is:

$$4\pi R_P^2 F_{S,P} = (1 - A_B)F_{S,*}\left(\frac{R_\star}{a}\right)^2 \pi R_P^2 + L_{P,int}. \tag{3.21}$$

For giant planets whose atmospheres are mainly composed by H and He, $L_{P,int}$ is the loss of residual gravitational potential energy; on Earth-like planets, it comes partly from the same source, but mostly from the decay of radioactive isotopes. This term can be ignored when the stellar flux is many orders of magnitude higher, i.e. when the planet is really close to its star.

The temperature of a planet varies with latitude, altitude and time (from day to night or from season to season). Therefore, many different temperatures can be used in different contexts:

- The *effective temperature* T_{eff} is used as a global temperature of the planet atmosphere: more specifically, it is the temperature of a blackbody of the same shape, distance and total flux (over all wavelengths) as the planet

$$F_S = \pi \int_0^\infty B(T, \nu)d\nu = \sigma_B T_{eff}^4 \tag{3.22}$$

With $\sigma_B = 5.67 \cdot 10^{-8}\ W\ K^{-4}\ m^{-2}$ being the Stefan-Boltzmann constant. T_{eff} is referred to the top of the atmosphere, where the radiation leaves the planet: therefore, it can be different from the temperature on the planetary surface (as in the case of Venus, heated by greenhouse gases).

- The *equilibrium temperature* T_{eq} is a theoretical number of the effective temperature for a planet with $L_{P,int} = 0$: physically, it is the effective temperature of an isothermal planet at equilibrium with the incoming radiation of its host star. It can be derived by substituting

$$F_S = \sigma_B T_{eq}^4 \tag{3.23}$$

To the energy balance equation (modified by a parameter f which counts for the section of the planet effectively measured):

$$\frac{4\pi R_P^2 F_{S,P}}{f} = (1 - A_B) F_{S,*} \left(\frac{R_\star}{a}\right)^2 \pi R_P^2 = (1 - A_B) S_0 \pi R_P^2 \tag{3.24}$$

Therefore

$$T_{eq} = T_{eff,*} \left(\frac{R_\star}{a}\right)^{1/2} \left[\frac{f}{4}(1 - A_B)\right]^{1/4} = \left[\frac{f}{4\sigma_B}(1 - A_B) S_0\right]^{1/4} \tag{3.25}$$

The parameter f requires further discussion: it is needed because we can observe only one portion of the planet at a time. Obviously, if the radiation is uniformly distributed around the planet, then $f = 1$; if this isn't the case, the values should be different: for example, tidally locked planets always show the same hemisphere to the star; thus, T_{eq} is different from the day-side to the night-side and $f = 8/3$.

- The *brightness temperature* T_b is the temperature of a blackbody of the same shape, distance, and flux of the planet in a particular frequency range. It can be derived from the Stefan-Boltzmann Law:

$$\sigma_B T_b^4(\nu) = F_S(\nu) = F_\oplus \left(\frac{D_\oplus}{R_P}\right)^2. \tag{3.26}$$

This is the only temperature we're capable to observe because we measure the various fluxes at different frequency ranges. To retrieve the effective temperature from the various brightness temperatures, a computer model atmosphere is needed.

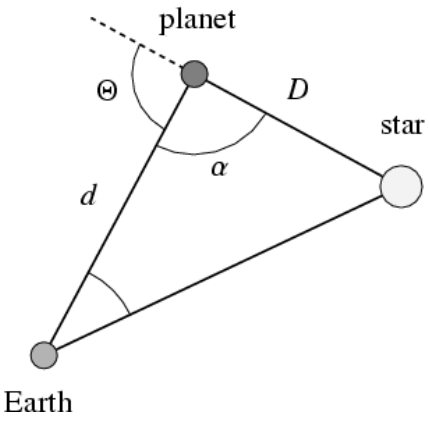


Fig. 3.8: Distances and angles involved when observing an extrasolar planet [Stam et al., 2004].

The phase angle is the star-planet-observer angle (Fig. 3.8): $\alpha = 0°$ means full phase, $\alpha > 170°$ means crescent phase, $\alpha = 180°$ means that the planet is not illuminated at all. When observing the Solar System from Earth, Mercury and Venus show illumination phases, while the outer planets are most of the time at full phase. For exoplanets, illumination phases depend from the inclination of the planet-star system: a transiting exoplanet will show phases, while a "face–on" system (where the orbit plane is almost perpendicular to the observer) will show a constant "half" phase.

The *phase law* or *integral phase function* is the fraction of scattered radiation at any phase angle Ψ_α, normalized to the flux at $\alpha = 0°$ ($\Psi_{\alpha=0}$). For a Lambert sphere it has an analytical form and is [Traub et al., 2010]:

$$\Phi_\alpha(\nu) = \frac{\Psi_\alpha(\nu)}{\Psi_{\alpha=0}(\nu)} = \frac{1}{\pi}[\sin\alpha + (\pi - \alpha)\cos\alpha]. \tag{3.27}$$

The planetary albedo is a measure of the planet's reflectivity: it is the ratio of the scattered light to the received light. It gives information about clouds in the atmosphere and other surface conditions; as previously seen, it is also important for the energy balance of the planet.

- The *single scattering albedo* $\tilde{\omega}$ is the fraction of incident light scattered by a particle in the atmosphere; it is wavelength-dependent and it also depends on the kind of particle.

- The *geometric albedo* $A_g(\nu)$ is the ratio of the planet's flux at full phase to the flux from a Lambert disk at the same distance and the cross-sectional area (the

surface that subtends the same solid angle) as the planet, which is the same as the incident flux from the star to the substellar point.

- The *Bond albedo* A_B is the light received from the star which is scattered back into space. By definition, it is ≤ 1. The *spherical albedo* $A_S(\nu)$ is the albedo at a specific frequency. Therefore:

$$A_B = \int_0^\infty A_S(\nu) d\nu. \tag{3.28}$$

  Both quantities depend on the stellar spectrum.

- The *apparent albedo* is the ratio between the scattered flux in the direction of the observer and the scattered flux from a Lambert sphere with the same size, phase angle, and distance as the planet. This is the only quantity that could be measured from a planet of known size.

A stellar spectrum can be considered as a blackbody curve with a single peak at a certain wavelength (thus, at a certain temperature); a planetary spectrum, instead, has two peaks (Fig. 3.9). The short-wavelength peak is due to the light of the star scattered by the atmosphere: therefore, it has the same wavelength as the star; the long-wavelength peak comes from the thermal emission of the planet. The ratio between planet and star fluxes at all wavelengths is essential for detecting exoplanets.

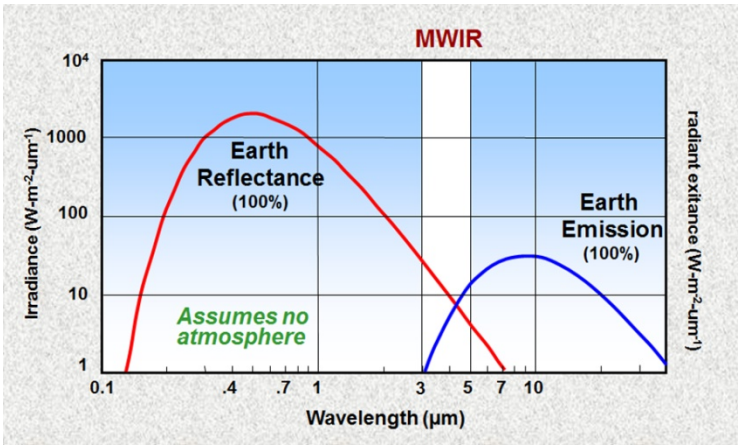


Fig. 3.9: Emitted and reflected flux of the Earth. (http://www.markelowitz.com/)

Recalling the expression for the observed flux of a planet/star from the Earth, it is easy to show that, generally:

$$\frac{F_{\oplus,P}(\nu)}{F_{\oplus,*}(\nu)} = \frac{F_{S,P}(\nu)}{F_{S,*}(\nu)} \frac{R_P^2}{R_\star^2}. \tag{3.29}$$

The thermal emission flux ratio comes directly from the energy balance equation. Approximating the stellar and planet fluxes as blackbody fluxes, where the temperatures are the brightness temperatures of the planet and the star, and assuming that these fluxes are normally peaked in the Rayleigh-Jeans region ($h\nu \ll k_B T$, where $h = 6.62607004 \cdot 10^{-34} m^2\ kg\ s^{-1}$ is the Planck constant, so that $F(T,\nu) = \pi B(T,\nu) \approx \pi \frac{2\nu^2 k_B T}{c^2}$), then:

$$\frac{F_{\oplus,P}(\nu)}{F_{\oplus,*}(\nu)} \approx \frac{T_{b,P}}{T_{b,*}} \frac{R_P^2}{R_\star^2}. \tag{3.30}$$

The scattered light flux ratio depends directly from the planetary geometrical albedo and inversely from the semi-major axis:

$$\frac{F_{\oplus,P}(\nu)}{F_{\oplus,*}(\nu)} = \frac{F_{S,P}(\nu)}{F_{S,*}(\nu)} \frac{R_P^2}{R_\star^2} = A_g(\nu) \frac{R_P^2}{a^2}. \tag{3.31}$$

A more detailed discussion of all these quantities can be found in Seager (2010).

3.4.3 Energy Transport

In a planetary atmosphere, energy can be transported mainly by radiation and convection; thermal conduction can be neglected due to the relatively low density of the atmosphere. The most efficient method depends on the considered environment: in regions where the opacity is low, energy is more likely to be transported by photons; if the opacity is high, instead, energy is more likely to be transported by convection. The radiative and convective transfer will be explained respectively in Sections 3.4.3.1 and 3.4.3.2; the criteria to discriminate these two scenarios will be explained in Sections 3.4.3.3.

3.4.3.1 Radiative Transfer

The radiative transfer equation describes the change in a beam of radiation as it travels some distance s through a volume of gas: more specifically, it considers absorption and emission processes from atoms, molecules and solid particles dispersed in a gas. Assuming a static atmosphere, the radiative transfer equation is:

$$\frac{dI(x,\hat{n},\nu)}{ds} = -\kappa(x,\nu)I(x,\hat{n},\nu) + \varepsilon(x,\hat{n},\nu) \tag{3.32}$$

Where:

- $\kappa(x,\nu)I(x,\hat{n},\nu)$ is due to losses from the beam. The extinction coefficient κ $[m^{-1}]$ includes all processes that remove energy from the beam, such as true absorption by a particle (denoted by the *true absorption coefficient* $\alpha(x,\nu)$), and scattering of photons towards other directions (denoted by the *scattering coefficient* $\sigma_s(x,\nu)$). Therefore:

$$\kappa(x,\nu) = \alpha(x,\nu) + \sigma_s(x,\nu). \tag{3.33}$$

  κ is bound to the mean free path $l = \frac{1}{n\sigma}$, since the drop of intensity in a beam of radiation is:

$$dI = -\kappa I dz = -n\sigma I dz = \frac{I dz}{l} \tag{3.34}$$

  So, $l = \frac{1}{\kappa}$.

- $\varepsilon(x,\hat{n},\nu)$ is due to additions to the beam. The monochromatic emission coefficient ε $[J\ m^{-3}\ sr^{-1}\ s^{-1}\ Hz^{-1}]$ includes both thermal emission and scattering.

  The first term can be retrieved applying the *local thermal equilibrium (LTE)*: it means that in a local area of the atmosphere any temperature, pressure, or chemical gradients are small compared to the photon free mean path, assuming that all properties of the matter depend only on local kinetic temperature and density (or pressure), with no bound to the unknown radiation field. Applying Kirchhoff's Law, ε_{therm} is obtained:

$$\varepsilon_{therm}(x,\nu) = \alpha(x,\nu)B(x,\nu) \tag{3.35}$$

  Where $B(x,\nu)$ is blackbody radiation. The second term is:

$$\varepsilon_{scat}(x,\hat{n},\nu) = \sigma_s(x,\nu)\frac{1}{4\pi}\int_{\Omega} P(\Theta)I(x',\hat{n}',\nu)d\Omega' \tag{3.36}$$

  Where Θ is the scattering angle and $P(\Theta)$ the dimensionless single scattering function. If the scattering is isentropic, it reduces to

$$\varepsilon(x,\nu) = \sigma_s(x,\nu)J(x,\nu) \tag{3.37}$$

Where $J(x,\nu)$ is the mean intensity. Therefore,

$$\varepsilon(x,\hat{n},\nu) = \alpha(x,\nu)B(x,\nu) + \sigma_s(x,\nu)J(x,\nu). \tag{3.38}$$

Now, it is possible to define the *source function* as the ratio of the emission coefficient and the extinction coefficient:

$$S(x,\hat{n},\nu) = \frac{\varepsilon(x,\hat{n},\nu)}{\kappa(x,\nu)} = \frac{\alpha(x,\nu)B(x,\nu) + \sigma_s(x,\nu)J(x,\nu)}{\alpha(x,\nu) + \sigma_s(x,\nu)}. \tag{3.39}$$

It is useful to introduce the *optical depth scale* τ_ν in order to simplify the radiative transfer equation and its solutions: it describes how opaque a part of the planetary atmosphere is to radiation traveling through it [Seager, 2010]: if $\tau_\nu \ll 1$, the medium is optically thin and a photon travels a long distance without being absorbed or scattered; if $\tau_\nu > 1$, the medium is thick and a photon interacts easily with it. The optical depth scale is dimensionless and it is related to the extinction coefficient, as shown below:

$$\tau_\nu(z,\nu) = -\int_{z=0}^{z_{max}} \kappa(z',\nu)dz' \equiv \int_{z_{max}}^{z=0} \kappa(z',\nu)dz' \tag{3.40}$$

Where z is the distance from the top of the atmosphere (or *TOA*); here, a plane-parallel approximation was used, so that $\tau(z=0,\nu) = 0$ at the TOA.

With the plane-parallel atmosphere assumption, which is valid when the radial depth of the atmosphere is much smaller than the planetary radius, it is possible to model an atmosphere as a stratified plane with each layer having homogeneous properties, dropping two of the three dimensions and adopting z as the (1D) vertical coordinate (Fig. 3.10).

From Fig. 3.10 it can be easily noticed that, being θ the angle between the surface normal and a beam of intensity,

$$\frac{dz}{ds} = \cos\theta \equiv \mu \tag{3.41}$$

So that the plane-parallel radiative transfer equation is:

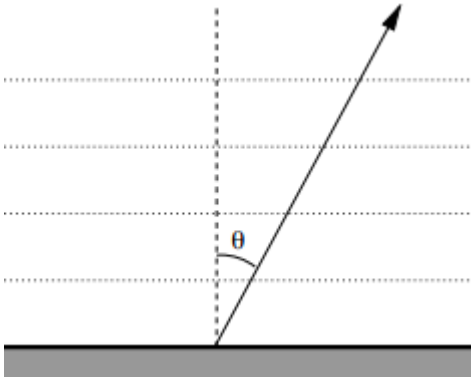


Fig. 3.10: Schematic description of a 1D plane-parallel atmosphere.

$$\mu \frac{dI(z,\mu,\nu)}{dz} = -\kappa(z,\nu)I(z,\mu,\nu) + \varepsilon(z,\mu,\nu). \tag{3.42}$$

This is an integro-differential equation, since the scattering term of ε contains angle integrals of I. Using the optical depth distance scale ($d\tau_\nu = \kappa dz$) and the definition of the source function, the equation becomes:

$$\mu \frac{dI(\tau_\nu,\mu,\nu)}{d\tau_\nu} = -I(\tau_\nu,\mu,\nu) + S(\tau_\nu,\mu,\nu). \tag{3.43}$$

The upper boundary condition is the stellar radiation incident on the TOA as a function of μ:

$$I(0,\mu,\nu) = I_\star(0,\mu_0,\nu) \qquad (-1 \leq \mu \leq 0). \tag{3.44}$$

The lower boundary condition is the intensity coming from the planet interior

$$I(\tau_{max,\nu},\mu,\nu) = I_{int}(\tau_{max,\nu},\mu_0,\nu) \qquad (0 \leq \mu \leq 1). \tag{3.45}$$

The solution is obtained by integrating from an initial optical depth $\tau_{\nu,i}$ to a final optical depth $\tau_{\nu,f}$:

$$I(\tau_{\nu,f},\mu,\nu) = I(\tau_{\nu,i},\mu,\nu)e^{-(\tau_{\nu,i}-\tau_{\nu,f})/\mu} - \frac{1}{\mu}\int_{\tau_{\nu,i}}^{\tau_{\nu,f}} S(\tau'_\nu,\mu,\nu)e^{-(\tau'_\nu-\tau_{\nu,f})/\mu}d\tau'_\nu. \tag{3.46}$$

The first term on the right side describes the initial intensity diminished by exponential attenuation of absorption; the second term describes the emission from the atmosphere.

It is possible to simplify the equation in a few simple scenarios:

- *Transmission of radiation alone*. There is no emission or scattering, therefore $\varepsilon(z, \mu, \nu) = 0$ and $S(z, \mu, \nu) = 0$. So, integrating from $\tau_{\nu,i} = 0$ to $\tau_{\nu,f} = \tau_\nu$:

$$I(\tau_\nu, \mu, \nu) = I(0, \mu, \nu)e^{-\tau_\nu/\mu}. \tag{3.47}$$

  This equation is commonly known as the Beer-Lambert's Law for the transmitted intensity. This is typically the case of a transiting planet, when the starlight crosses the planetary atmosphere and it is partially absorbed by the atoms and molecules in it.

- *Thermal emission*. The radiation from the planet is not due to scattering, but originates from absorbed stellar radiation or from the planetary interior. It usually occurs at infrared wavelengths near the peak of the planetary blackbody radiation curve, where scattering can be neglected (therefore $\varepsilon(\tau_\nu, \mu, \nu) = \varepsilon_{therm}(\tau_\nu, \mu, \nu)$, $\kappa(\tau_\nu, \nu) = \alpha(\tau_\nu, \nu)$ and $S_{therm}(\tau_\nu, \mu, \nu) = B(\tau_\nu, \nu)$):

$$\mu \frac{dI(\tau_\nu, \mu, \nu)}{d\tau_\nu} = -\kappa(\tau_\nu, \nu)I(\tau_\nu, \mu, \nu) + B(\tau_\nu, \nu) \tag{3.48}$$

  Whose solution is

$$I(0, \mu, \nu) = \int_0^\infty \frac{1}{\mu} B(\tau_\nu, \nu)e^{-\tau_\nu/\mu} d\tau_\nu \tag{3.49}$$

- *Constant Source Function*. The source function doesn't change with altitude:

$$I(\tau_{\mu,min}, \mu, \nu) = I(\tau_{\mu,max}, \mu, \nu)e^{-(\tau_{\nu,max} - \tau_{\nu,min})} + S(\nu)(1 - e^{-(\tau_{\nu,max} - \tau_{\nu,min})}). \tag{3.50}$$

  In the optically thick regime $e^{\tau_\nu} \to 0$, so the equation becomes:

$$I(\tau_{\mu,min}, \mu, \nu) = S(\nu). \tag{3.51}$$

This is valid deep in the atmosphere, so spectral feature detected on spectra cannot be formed at high optical depths. In the optically thin regime, higher on the planetary atmosphere, $e^{\tau_\nu} \approx (1 - \tau_{\nu,max})$:

$$I(0, \mu, \nu) = I(\tau_{\nu,max}, \mu, \nu) - \tau_{\nu,max}[I(\tau_{\nu,max}, \mu, \nu) - S(\nu)]. \tag{3.52}$$

So, if $I(\tau_{\nu,max}, \mu, \nu) > S(\nu)$, the initial intensity will be diminished and there will be absorption lines in the emergent spectrum; if $I(\tau_{\nu,max}, \mu, \nu) < S(\nu)$ instead, the initial intensity will be enhanced and there will be emission lines in the emergent spectrum.

- *Radiative diffusion*. Photons slowly move from one region to another; it can be applied on optically thick regions, so deep in an exoplanetary atmosphere. Absorption and scattering are efficient, and collisions dominate over radiative rates, so that LTE is valid ($S(\tau_\nu, \mu, \nu) = B(\tau_\nu, \nu)$). The source function changes slowly over a distance τ_ν, so it is possible to consider a Taylor expansion:

$$S(\tau'_\nu, \mu, \nu) \approx B(\tau_\nu, \nu) + \frac{dB(\tau_\nu, \nu)}{d\tau_\nu}(\tau'_\nu - \tau_\nu) \tag{3.53}$$

And the solution becomes

$$I(\tau_\nu, \mu, \nu) \approx B(\tau_\nu, \nu) - \mu\frac{dB(\tau_\nu, \nu)}{d\tau_\nu}. \tag{3.54}$$

Therefore, applying Eq. 3.12, the flux is:

$$F(\tau_\nu, \nu) \approx \frac{4\pi}{3}\left[\frac{dB(\tau_\nu, \nu)}{d\tau_\nu}\right] = -\frac{4\pi}{3}\left[\frac{1}{\kappa_R(\tau_\nu, \nu)}\frac{dB(\tau_\nu, \nu)}{dT}\right]\left(\frac{dT}{dz}\right) \tag{3.55}$$

Where κ_R is the Rosseland mean opacity, defined as:

$$\frac{1}{\kappa_R} = \frac{\int_0^\infty \frac{1}{\kappa}\frac{dB}{dT}d\nu}{\int_0^\infty \frac{dB}{dT}d\nu}. \tag{3.56}$$

The total radiative flux is obtained integrating over all frequencies:

$$F(z) = -\frac{16}{3}\frac{\sigma_B T^3}{\kappa_R(\tau_\nu, \nu)}\frac{dT}{dz}. \tag{3.57}$$

3.4.3.2 Convection

Convection is the transport of energy by bulk motions of matter in the vertical direction [Seager, 2010]. It is a process driven by temperature gradient and enabled by the gravity field.

Considering a parcel of dry air initially at rest that is slightly pushed upwards by upslope wind or buoyancy, it is possible to approximate its motion as a dry adiabat: while not exchanging heat with the surrounding (adiabatic conditions), it adapts to the (lower) atmospheric pressure of the upper layers and therefore expands and cools itself, up to temperatures that are possibly different from the surroundings.

In a framework where the hydrostatic equilibrium and perfect gas law equations are valid, it is possible to write the change in enthalpy H $[J]$ for the parcel [North et al., 2009]:

$$dH = Mc_p dT = Vdp = M\frac{RT}{p}dp \tag{3.58}$$

where M $[kg]$ is the mass of the parcel, c_P $[J\ kg^{-1}\ K^{-1}]$ the specific heat capacity at constant pressure, V $[m^3]$ the volume, p $[kg\ m^{-1}\ s^{-2}]$ the pressure, T $[K]$ the temperature, and $R = k_B/\mu$ $[J\ kg^{-1}\ K^{-1}]$ the specific gas constant. Dividing each side by Mdz and using the hydrostatic equation (Eq. 3.4 and the perfect gas law, we obtain:

$$c_p\frac{dT}{dz} = -g \tag{3.59}$$

The temperature gradient of the parcel during the ascent is therefore:

$$\left(\frac{dT}{dz}\right)_{ad,d} = -\frac{g}{c_P}. \tag{3.60}$$

It is however common in terms of the opposite of this variation, that is the *dry adiabatic lapse rate* $\Gamma_{ad,d}$ $[K\ km^{-1}]$:

$$\Gamma_{ad,d} = -\left(\frac{dT}{dz}\right)_{ad,d} = \frac{g}{c_P} \tag{3.61}$$

This parameter depends on the chemical composition of the atmosphere, which defines the specific heat capacity at constant pressure, and on the surface gravity,

but it is not related to the actual temperature of the atmosphere. On Earth, the critical lapse rate of a dry adiabat is approximately 10 $K\ km^{-1}$ (i.e. the temperature decreases of 10 degrees every km).

In the case of a parcel containing moisture, the process of lifting to upper layers of the atmosphere may cause condensation of the water vapor in the parcel itself. Therefore, the motion of a moving air mass would not be entirely adiabatic, since the release of latent heat from condensation must be considered. This additional heating from condensation causes the temperature to decrease at a slower rate compared to the dry adiabatic process.

As the parcel absorbs heat that the condensing vapour releases, the enthalpy variation becomes:

$$dH \approx Mc_p dT = Vdp + dQ = -Mgdz - MLd\omega_s \tag{3.62}$$

where L $[J\ kg^{-1}]$ is the specific latent heat for the vapor, and ω_s is saturation mixing ratio (kg of condensed water vapor per kg of dry air).

Dividing both sides by $Mc_p dz$, we obtain:

$$\frac{dT}{dz} = -\frac{L}{c_p}\frac{d\omega_s}{dz} - \frac{g}{c_p} \tag{3.63}$$

Thus, the *moist adiabatic lapse rate* $\Gamma_{ad,m}$ $[K\ km^{-1}]$ is:

$$\Gamma_{ad,m} = \Gamma_{ad,d} + \frac{L}{c_p}\frac{d\omega_s}{dz} \tag{3.64}$$

The derivative $\frac{d\omega_s}{dz}$ can be retrieved using the hydrostatic equation and the Clausius-Clapeyron equation. Rearranging the lapse rate equation, we obtain:

$$\Gamma_{ad,m} = \Gamma_{ad,d} \times \left(\frac{1 + \omega_s L/R_d T}{1 + \omega_s L^2/c_P R_w T^2}\right) \tag{3.65}$$

Where R_d is the specific gas constant for dry air, and R_w is the specific gas constant for water vapor.

If the water droplets are retained by the air parcel during its motion upwards, the process is called a *moist* adiabatic process. If the droplets precipitate, the process is no longer reversible, and it is rather defined as a *pseudo-adiabatic* process.

The difference between the dry and moist adiabatic lapse rates can be significant and that determines the variable proportion between dry and moist convection: on Earth, the value of Γ_m is typically 5 $K\ km^{-1}$.

3.4.3.3 Criteria

The study of the temperature gradient dT/dz helps quantify which energy transport mechanism dominates.

When a dry parcel moves upwards in the atmosphere, it changes its temperature according to the values of its dry adiabatic lapse rate. At the same time, the environment surrounding the air parcel changes its temperature with its own *atmospheric temperature gradient* $(dT/dz)_{atm}$, corresponding to an *atmospheric lapse rate*:

$$\Gamma_{atm} = -\left(\frac{dT}{dz}\right)_{atm} \tag{3.66}$$

Fig. 3.11 represents a sketch of a typical sounding curve of the atmosphere in a qualitative T-z plot. The segments that are tangents to points A, B, and C are portions of a dry adiabat: any parcel of dry air lifted upwards from those points will be cooled down following these segments. At any altitude, the temperature comparison of the parcel with respect to the sounding triggers any convective mechanism. Three scenarios could happen:

- $\Gamma_{atm} > \Gamma_{ad,d}$ (A): the atmosphere is *unstable* with respect to convection; when lifted upwards of an infinitesimal altitude dz, the parcel becomes warmer than the environment and will experience a buoyant force upwards, so convection takes place;
- $\Gamma_{atm} = \Gamma_{ad,d}$ (B): the atmosphere is *neutral*. For an infinitesimal perturbation in altitude dz, the parcel is at the same temperature as the atmosphere, so it will stop and stay at its place;
- $\Gamma_{atm} < \Gamma_{ad,d}$ (C): the atmosphere is *stable* with respect to convection. For an infinitesimal shift in altitude dz, the parcel becomes colder than the atmosphere and will sink back, generating small oscillations; no convection will occur.

If there is water vapor in the atmosphere and, on a particular layer, atmospheric lapse rate is higher than the moist adiabatic lapse rate but lower than the dry adiabatic lapse rate, *conditional instability* may occur: this could cause turbulence in

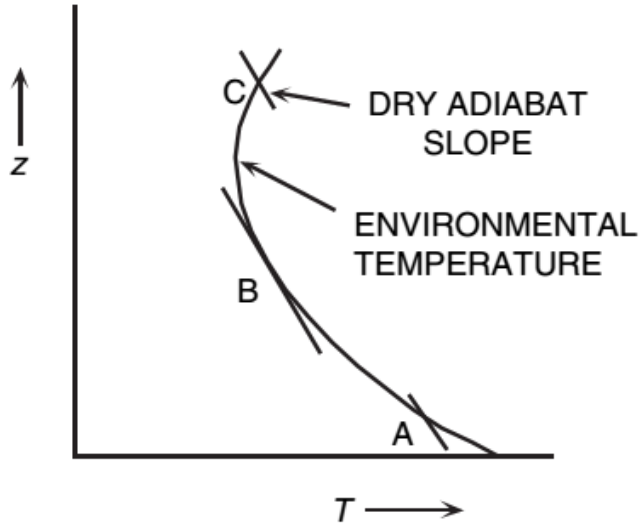


Fig. 3.11: A sounding curve of the atmosphere in a T-z plot; straight line segments represent dry adiabats passing through the points A (unstable), B (neutral) and C (stable) [North et al., 2009].

that layer of the atmosphere, since convection may happen if the air parcels become saturated.

So, in this case, the stability criteria are the following:

- $\Gamma_{atm} > \Gamma_{ad,d}$: the atmosphere is *absolutely unstable* with respect to convection;
- $\Gamma_{ad,d} > \Gamma_{atm} > \Gamma_{ad,m}$: the atmosphere is *conditionally unstable* with respect to convection;
- $\Gamma_{atm} < \Gamma_{ad,m}$: the atmosphere is *absolutely stable* with respect to convection.

Therefore, if an air parcel in an unstable atmospheric layer is hotter than the surroundings, it travels upwards; while doing this, it experiences a lower ambient pressure, it expands and cools down until it reaches a point when it is cooler than the environment, so it sinks towards a level with higher pressure and temperature. Here, the parcel shrinks and heats up again and rises once more, in a closed energy cycle.

On the other hand, if the atmospheric layer is stable, convection is not efficient in transferring energy, which will be on the other hand exchanged via radiative transfer.

The energy will be transported by the most efficient method, and that is the one with the lowest temperature gradient: usually, in optical thick regions of an atmosphere

convection almost always sets in, since the opacity is high and increases the radiative temperature gradient; on the contrary, in optical thin layers heat is mainly transferred via radiation.

3.4.3.4 Advection

Circulation is the large-scale movement of gas in an atmosphere that distributes energy all through the atmosphere itself: it acts to minimize temperature gradients and to make the atmosphere uniform. It is responsible for many large- and small-scale phenomena, such as winds, jets, hurricanes, and tornadoes [Seager, 2010], especially in tidally locked planets (i.e. planets that show always the same hemisphere to the star, resulting in a strong temperature gradient between "night" and "day"). It is a very complex process that requires numerical calculations and many simplifying assumptions.

It is possible to estimate the typical timescales of the radiative and advective processes in an atmosphere: their comparison should tell if atmospheric dynamics is negligible in a planetary atmosphere or not.

The radiative timescale τ_{rad} is the time for absorbed stellar energy to be reemitted:

$$\tau_{rad} \sim \frac{\Delta P}{g} \frac{c_P}{4\sigma_B T^3} \tag{3.67}$$

Where ΔP is the pressure thickness of the considered atmospheric layer, T the temperature, g the surface gravity, c_P the specific heat capacity, and σ_B the Stefan-Boltzmann constant.

The advective timescale τ_{adv} is the time for the absorbed stellar energy to circulate around the planet. It can be estimated by the planet radius R_P and the characteristic wind speed U, usually derived from atmospheric circulation models:

$$\tau_{adv} \sim \frac{R_P}{U}. \tag{3.68}$$

If $\tau_{rad} \ll \tau_{adv}$, nearly all of the absorbed stellar energy will be reemitted before being advected around the planet: therefore, strong latitudinal and longitudinal temperature gradients are expected to arise.

In contrast, if $\tau_{rad} \gg \tau_{adv}$ the temperature should be much more uniform because heat will be redistributed over the entire planet.

3.4.3.5 Greenhouse effect

The global average surface temperature of the Earth is 288 K. However, if we insert the proper values of stellar temperature, distance, planetary radius, and albedo in the Eq. 3.25, we would obtain only 255 K. To better understand this discrepancy, the spectrum of Earth's infrared spectrum as observed by the Nimbus 4 satellite and compared with some blackbody curves at various temperatures in Figure 3.12 comes in handy.

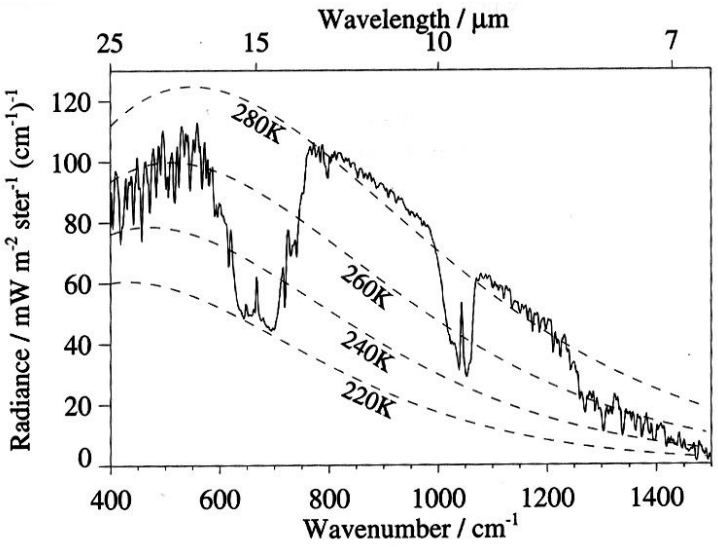


Fig. 3.12: Spectrum of infrared emission escaping to space, as observed from outside the Earth's atmosphere by the Nimbus 4 satellite. Dashed lines represent the blackbody spectrum at different temperatures (Data from Hanel et al. (1972)).

The average infrared emission temperature is about 250 K, well in agreement with the equilibrium temperature estimate. However, this value is indeed a balance between some cooler spectral regions and some hotter ones. In particular, between 12 and 17 μm, at 9.6 μm and at wavelengths less than 8 μm there are some emission-temperature dips, at much lower temperatures than the other regions, whose blackbody emission temperature would be otherwise well in agreement with the measured Earth surface temperature.

These low-temperature wavelength regions correspond to the infrared bands of CO_2, O_3, and H_2O respectively: up to a certain layer in the atmosphere, such gases absorb the infrared radiation that comes from the surface and emit radiation upwards at their (cooler) temperature, thus generating the features in the spectrum. The hotter regions show, on the other hand, the direct surface emission radiation, not absorbed by any gas throughout the crossing of the atmosphere [Wayne, 2000].

In the visible range, where the solar radiation is more abundant, the atmosphere is overall transparent; on the other hand, it is more effective in absorbing the infrared radiation, which is where the blackbody emission radiation of the surface is peaked.

The radiation absorbed from the atmosphere is directed downward, as well as upward, so that the surface receives both the stellar radiation and the incoming infrared radiation emitted from the atmosphere and reaches higher temperatures: for this reason, this phenomenon is often called *greenhouse effect*.

On a first-order approximation, one can interpret the atmosphere as one single layer at average temperature T_a, assuming a planar geometry and an atmosphere that is completely opaque to infrared radiation (see Figure 3.13 for a sketch of the considered quantities).

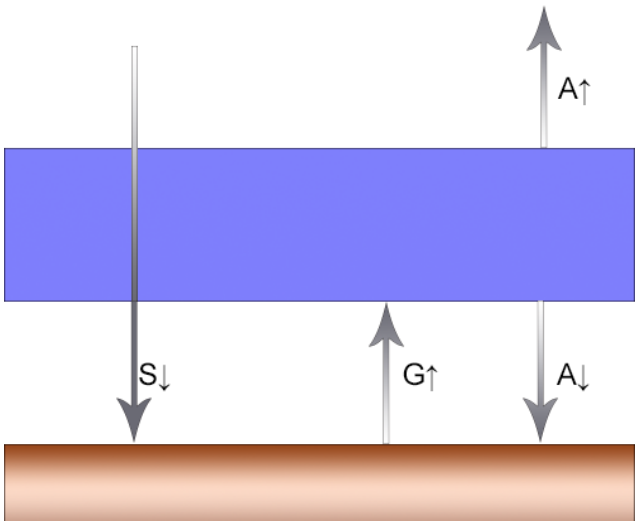


Fig. 3.13: Sketch of a simple greenhouse atmosphere. The brown box represents the ground, the pale blue box the atmosphere, approximated to a single layer. The quantities that label the arrows are explained in the text.

Recalling Eq. 3.19 for the incident energy per unit time and taking into account the planetary albedo, the average incoming solar energy per unit area on Earth's surface is the ratio between the intercepted incoming radiation and Earth's surface area:

$$S_{\downarrow} = \frac{\pi R_{\oplus}^2 (1 - A_B)\, S_0}{4\pi R_{\oplus}^2} = \frac{(1 - A_B)\, S_0}{4} \tag{3.69}$$

This is equal to (see Eq.3.25):

$$S_{\downarrow} = \sigma_B T_{eq}^4 \tag{3.70}$$

On the other hand, the radiation re-emitted by the atmosphere to space is equal to:

$$A_{\uparrow} = \sigma_B T_a^4 \tag{3.71}$$

In equilibrium, the incoming radiation at the top of the atmosphere must balance the emitted counterpart. Therefore:

$$S_{\downarrow} = \sigma_B T_{eq}^4 = \sigma_B T_a^4 = A_{\uparrow} \tag{3.72}$$

So the atmosphere is, as a matter of fact, at the equilibrium temperature. Since the atmosphere is at equilibrium, the downward component of the emitted atmospheric radiation is equal to its upward component: $A_{\uparrow} = A_{\downarrow}$.

At the surface, the incident flux is composed by both the stellar radiation and the downward atmospheric re-emitted radiation contribute, so the incoming flux on the ground is:

$$S_{\downarrow} + A_{\downarrow} = \sigma_B T_{eq}^4 + \sigma_B T_a^4 = 2\sigma_B T_{eq}^4 \tag{3.73}$$

While the surface flux radiating upward is:

$$G_{\uparrow} = \sigma_B T_S^4 \tag{3.74}$$

Since this layer too must be at equilibrium, its net flux must be zero.

$$S_{\downarrow} + A_{\downarrow} = 2\sigma_B T_{eq}^4 = \sigma_B T_S^4 = G_{\uparrow} \tag{3.75}$$

Therefore, the surface temperature is:

$$T_S = 2^{\frac{1}{4}} T_{eq} \tag{3.76}$$

In other words, the presence of an absorbing atmosphere heats the surface of a factor $2^{1/4} = 1.19$. This factor gives a surface temperature, for the Earth, of 303 K, showing that this treatment is an oversimplification of what actually happens in the atmosphere.

A slightly more realistic model has to account the atmospheric absorption and its emission: in this case, we are considering a *leaky greenhouse model* (see Figure 3.14

for a sketch of the considered quantities). Now, the atmosphere is not completely opaque to the infrared radiation but has its own absorptivity/emissivity[2] ϵ.

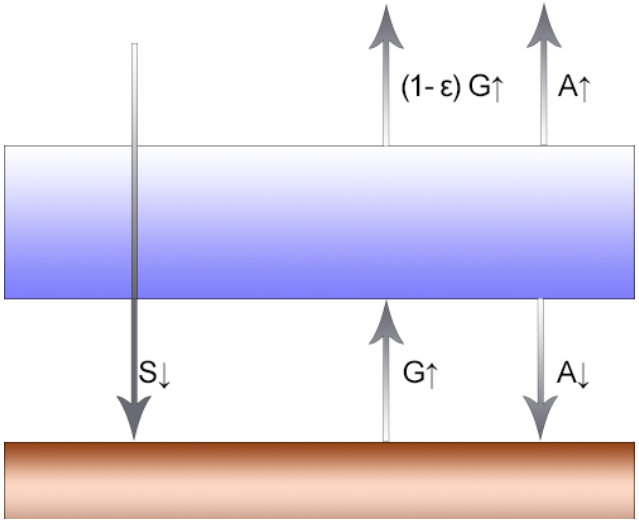


Fig. 3.14: Sketch of a leaky greenhouse atmosphere. The brown box represents the ground, the pale blue box the atmosphere, approximated to a single layer. The quantities that label the arrows are explained in the text.

So, only a fraction ϵ of the atmosphere absorbs the infrared radiation coming from the surface. The remaining surface upward radiation $(1-\epsilon)G_{\uparrow}$ needs to be taken into account in the energy budget. The balance at the top of the atmosphere must now be:

$$S_{\downarrow} = \sigma_B T_{eq}^4 = A_{\uparrow} + (1-\epsilon)G_{\uparrow} \tag{3.77}$$

Because, at equilibrium, $A_{\downarrow} = A_{\uparrow}$, Eq. 3.75 becomes:

$$S_{\downarrow} + S_{\downarrow} - (1-\epsilon)G_{\uparrow} = 2\sigma_B T_{eq}^4 - (1-\epsilon)\sigma_B T_S^4 = \sigma_B T_S^4 = G_{\uparrow} \tag{3.78}$$

Which leads to

$$T_S = \left(\frac{2}{2-\epsilon}\right)^{1/4} T_{eq} \tag{3.79}$$

[2]Invoking Kirchoff's law, the emittance of a body (the ratio of the emitted flux compared to the one emitted by a blackbody at the same temperature) is equal to its absorptance.

In the limit of $\epsilon \to 0$ (transparent/no atmosphere), the surface temperature is equal to the equilibrium temperature; on the other hand, for $\epsilon \to 1$, the surface temperature is equal to its simple greenhouse value $T_S = 2^{\frac{1}{4}} T_{eq}$.

The atmospheric temperature can now be calculated considering the fraction of radiation that is absorbed by the atmosphere itself. Thus:

$$A_{\uparrow} = A_{\downarrow} = \epsilon \sigma_B T_a^4 \tag{3.80}$$

Substituting the previous equations, we obtain:

$$T_a = \left(\frac{1}{1-\epsilon}\right)^{\frac{1}{4}} T_{eq} = \left(\frac{1}{2}\right)^{\frac{1}{4}} T_S \tag{3.81}$$

Therefore, the atmosphere of a leaky greenhouse is always cooler than the surface.

This treatment leads, for $\epsilon = 0.78$, to $T_S = 288.3\ K$ and $T_a = 242.5\ K$, fairly in agreement with the measured values.

However, clouds and hazes do play a role in the total energy budget, and inhomogeneities in the chemical composition of the atmosphere (and therefore, the profile of the greenhouse gases with respect to altitude) make the treatment of this process, in truth, very complicated.

3.4.4 Chemical composition

Solar System planets have, up to date, quite a detailed description of their atmospheric composition: this was possible by way of *in situ* measurements and probes sent from Earth to orbit those planets. This is impossible for an exoplanet, so it is hard to have a complete characterization of its gaseous envelope.

A typical atmosphere is composed by different compounds, but only a few of them dominate its spectrum: for example, CH_4 dominates the reflectance spectrum of a giant planet atmosphere, while CO_2 has a very strong absorption feature in a typical terrestrial planet [Seager, 2010].

On Super Earths and smaller terrestrial exoplanets, the atmospheric composition depends on many mechanisms, such as the kinetics of the reactions among atmospheric components and their vertical mixing (Section 3.4.4.1), photochemistry (Section 3.4.4.2), condensation in clouds (Section 3.4.5) and atmospheric escape (Section

3.4.6). Many further processes are involved, making the detailed description of a planetary atmosphere undoubtedly complex; for this purpose, it is important to compute synthetic models - that require some approximation - to help to guess which chemical compositions match the observed spectra.

Following Seager (2010), the discussion about this subject starts with the easiest process and adds complexity little by little.

3.4.4.1 Chemical Equilibrium

This situation occurs when number densities remain constant through time in a closed system: chemical reactions still occur, but without net change. Usually, to have equilibrium in a reaction a certain amount of time is needed: therefore, the basic assumption is that an atmospheric layer must have had enough time to reach chemical equilibrium.

A thermodynamic state function widely used in the field of atmospheric science is the *Gibbs free energy* (or *Gibbs function*): studying that, it can be determined whether a particular reaction has reached chemical equilibrium and, if so, what are the abundances of the compounds involved [North et al., 2009]. The Gibbs energy is defined as:

$$G = H - TS = U + PV - TS \tag{3.82}$$

Where $H = U + PV$ is the enthalpy, U the internal energy, P the pressure, V the volume, T the temperature, and S is the entropy. It can be written in differential form:

$$dG = dU + d(PV) - d(TS) = VdP - SdT \tag{3.83}$$

Since $dU = TdS - PdV$ is the differential form of the first law of thermodynamics. From Eq. 3.83, it can be seen that G is a natural function of temperature and pressure: in a phase transition, where T and P remain constant, the Gibbs energy doesn't change either. In other terms, $dG = 0$ for a reversible process where $dP = dT = 0$.

However, irreversible processes happen in nature as well. Thus, invoking the second law of thermodynamics:

$$dS \geq \frac{dQ}{T} \tag{3.84}$$

So that

$$TdS \geq dU + PdV \tag{3.85}$$

$$dU \leq TdS - PdV. \tag{3.86}$$

Eq 3.83 becomes:

$$dG \leq VdP - SdT. \tag{3.87}$$

For a spontaneous change at constant pressure and temperature, $dG \leq 0$. At equilibrium $dP = dT = 0$, therefore $dG = 0$: this is a minimum point of the Gibbs function. All the processes with $dG \geq 0$ are non-spontaneous (see Fig. 3.15).

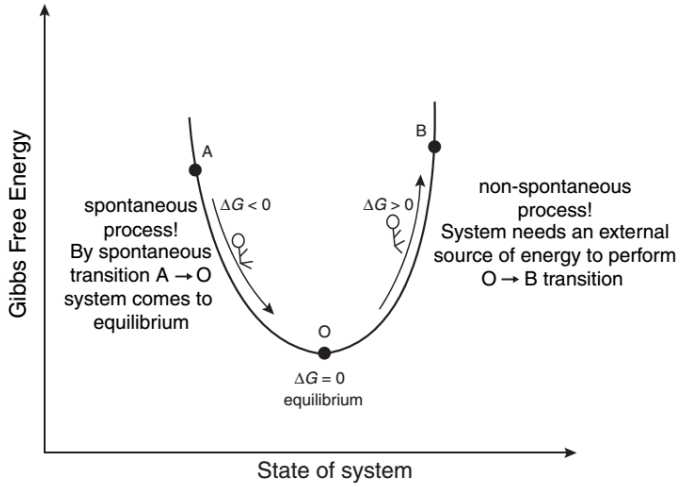


Fig. 3.15: Qualitative graph of the Gibbs function [North et al., 2009].

Reactions proceed faster at higher pressure and temperature, so it is likely to have chemical equilibrium in the deeper and hotter part of an atmosphere; in contrast, reactions in the upper part of an atmosphere hardly reach equilibrium, so *non-equilibrium chemistry* develops.

Such processes generally involve nitrogen and carbon compounds reactions, as for example:

$$N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$$
$$CO + 3\,H_2 \rightleftharpoons CH_4 + H_2O$$

The reason those reactions are out of chemical equilibrium (especially in H-rich atmospheres) lies in the strength of the triple bond of N_2 and the double bond in CO, that are therefore difficult to break apart. Also, the equilibrium timescale τ_{chem} is often higher than the vertical mixing timescale τ_{mix}: new amounts of those molecules are dredged up from the bottom of the atmosphere (where equilibrium is established, as $\tau_{chem} \leq \tau_{mix}$), resulting in an overabundance of N_2 or CO in spite of their products. The level where $\tau_{chem} = \tau_{mix}$ is called *quench level*.

3.4.4.2 Photochemistry

In the upper layers of a planetary atmosphere, molecules absorb UV and visible radiation from the host star and could be photodissociated [Hu et al., 2012]. Photochemistry is essential in the description of an Earth-like atmosphere since the very light photochemically-produced elements (such as H and He) are likely to escape the atmosphere itself, due to the low surface gravity. Moreover, low pressure and temperature hardly allow reformation of molecules split apart by the UV radiation.

The 1D continuity-transport equation that governs the abundance is, for each compound:

$$\frac{\partial n_i(z)}{\partial t} + \frac{\partial \Phi_i(z)}{\partial z} = P_i(z) - n_i(z)L_i(z) \tag{3.88}$$

Where

- $n_i(z)$ $[cm^{-3}]$ is the number density of a species i at altitude z;
- $\Phi_i(z)$ $[cm^{-2}\ s^{-1}]$ is the vertical flux of a species i at altitude z;
- $P_i(z)$ $[cm^{-3}\ s^{-1}]$ is the production rate term, that collects all the sources of $n_i(z)$; it could include also surface fluxes generated by biochemical processes;
- $L_i(z)$ $[s^{-1}]$ is the loss term, that collects all the sinks of $n_i(z)$.

As sources and sinks involve other chemical species, these equations are strictly coupled. On a first approximation, it would suffice to neglect the vertical motion of each species ($\frac{\partial \Phi_i(z)}{\partial z} = 0$), as well as neglecting the variation of the number density in time, supposing photochemical equilibrium ($\frac{\partial n_i(z)}{\partial t} = 0$).

For example, considering this network of photochemical reactions (following Seager (2010)):

$$(A_2 + h\nu) \xrightarrow{J} (A + A)$$
$$(A + A + M) \xrightarrow{K} (A_2 + M)$$

Where A is a generic atom, A_2 a diatomic molecule, M a third atom or molecule, J the photodissociation rate and K the chemical reaction rate, the corresponding rate equations in photochemical equilibrium are (square brackets denote concentrations):

$$\frac{\partial A}{\partial t} = +J[A_2] - K[A][A] = 0 \tag{3.89}$$

$$\frac{\partial A_2}{\partial t} = -J[A_2] + K[A][A] = 0. \tag{3.90}$$

Knowing J, K, and assuming that the number density of the species must be conserved, it is possible to solve for the unknown number densities A and A_2.

The photodissociation rate J $[s^{-1}]$ can be expressed in terms of the *spectral actinic flux* $F_{act}(\nu)$ $[photons\ m^{-2}\ s^{-1}\ m^{-1}]$, i.e. the flux incident on the molecule from all directions, and of the absorption cross section $\sigma(\nu)$:

$$J(z) = \int_0^\infty \sigma(\nu) F_{act}(\nu) d\nu \tag{3.91}$$

Where z is altitude in the vertical direction and ν is the frequency.

On Earth, this description suits the case of ozone production and destruction by solar radiation; on exoplanetary atmospheres, by contrast, it is impossible to measure the concentrations of individual species, so that a full photochemical network model could include over one hundred reaction rates [see e.g. Hu et al., 2012].

Now, taking back into account vertical mixing among the different layers of the atmosphere, it is possible to define $\Phi_i(z)$ as:

$$\Phi_i(z) = n_i(z) v_i = \kappa_E \frac{\partial n_i}{\partial z} \tag{3.92}$$

Where v_i is the velocity in the vertical direction and κ_E the *eddy diffusion coefficient* $[m^2\ s^{-1}]$, so that the term in Eq. 3.88 becomes:

$$\frac{\partial \Phi_i(z)}{\partial z} = -\frac{\partial}{\partial z}\left[\kappa_E \frac{\partial n_i}{\partial z}\right] \tag{3.93}$$

Further processes can be added when relevant, but a robust modeling routine is still needed to achieve reliable results, often taking into account many simplifying assumptions.

The knowledge of the chemistry of an exoplanet atmosphere is essential mostly for the study of the surface habitability: biochemically-produced chemical species are most likely to react with radiation and other compounds in the atmosphere and their products, when spectroscopically detected, could be a quantitative hint towards the discovery of life on other worlds (see Section 4 for further information).

On Earth's atmosphere, photochemistry is involved in the ozone production and destruction, a cycle of reactions which is extremely important for habitability purposes on our planet. Ozone (or trioxygen O_3) is present in the atmosphere only in trace amounts; its distribution is characterized by a peculiar dependence on altitude, with a maximum abundance at an altitude of 25-30 km. The majority of the ozone ($\sim 10^{12}\ molecule\ cm^{-3}$) is contained in the so-called *ozone layer*, about 20 km thick in the stratosphere: indeed, it is the ozone layer itself that causes the stratosphere to exist. Its presence allows the life-threatening UV solar radiation up to ~230 nm to be absorbed in the stratosphere: it is the only compound in the atmosphere capable of absorbing photons at such wavelengths, which are dangerous for proteins and nucleic acids – the building blocks of all living cells.

Ozone is formed by photolysis of molecular oxygen (mainly biologically produced) caused by ultraviolet radiation; its loss is due both to ozone photolysis and reactions with other trace gases, such are oxides of nitrogen, which are also biologically produced. This is why ozone is interpreted as a biosignature, as will be explained in Section 4.3. However, it can be also produced by anthropogenic processes in the troposphere.

The energy absorbed by ozone converts into heating of the stratosphere, which is responsible for the temperature inversion experienced in this layer. When disrupting ozone, molecular oxygen is formed, which is the main source of ozone itself: this is one of the most important *cyclic processes* in the atmosphere.

The ozone cycle was at first resolved by Sydney Chapman [Chapman, 1932]. It considers four reactions for the production and distruction of ozone, which in-

volve molecular oxygen and any molecule M not directly involved in the reactions themselves, and act as a third body:

$$O_2 + h\nu \xrightarrow{J_{O2}} O + O \tag{3.94a}$$

$$O + O_2 + M \xrightarrow{k_1} O_3 + M \tag{3.94b}$$

$$O_3 + h\nu \xrightarrow{J_{O3}} O + O_2 \tag{3.94c}$$

$$O + O_3 \xrightarrow{k_2} O_2 + O_2 \tag{3.94d}$$

The first reaction (3.94a) takes place when UV photons (of wavelengths less than 242 nm) dissociate molecular oxygen. The oxygen atoms react with a third molecule M (reaction 3.94b), so ozone is produced. The O_3 thus formed absorbs radiation in the wavelength range of 240-320 nm producing the Chappuis and Hartley bands (reaction 3.94c), and then decomposes back to O_2 (reaction 3.94d).

These reactions define a set of three differential equations for the corresponding species concentrations $[O_i]$:

$$\frac{d[O]}{dt} = -k_1[O][O_2][M] - k_2[O][O_3] + J_{O_3}[O_3] + 2J_{O_2}[O_2] \tag{3.95a}$$

$$\frac{d[O_2]}{dt} = -k_1[O][O_2][M] + 2k_2[O][O_3] + J_{O_3}[O_3] - J_{O_2}[O_2] \tag{3.95b}$$

$$\frac{d[O_3]}{dt} = +k_1[O][O_2][M] - k_2[O][O_3] - J_{O_3}[O_3] \tag{3.95c}$$

$$\tag{3.95d}$$

Where the reactions rates k_1 and k_2 are variable with altitude and depend on the temperature at each height z:

$$k_1(z)\ [cm^6\ s^{-1}] = 6 \cdot 10^{-34} \left(\frac{T(z)}{298}\right)^{-2.3} \tag{3.96}$$

$$k_2(z)\ [cm^3\ s^{-1}] = 8 \cdot 10^{-12} \exp\left(\frac{-2060}{T(z)}\right) \tag{3.97}$$

And the photolysis rates J_{O_2} and J_{O_3} for each layer are given by Eq. 3.91.

The ozone abundance is directly related to the latitude of the planet, based on the variation of the sunlight intensity and the temperature of the layer [Seinfeld et al., 2016].

The Chapman mechanism predicts, however, too much ozone. Catalytic cycles must be taken into account. A generic catalytic cycle can be represented by the pair of reactions [Wayne, 2000]:

$$\begin{aligned} X + O_3 &\longrightarrow XO + O_2 \\ XO + O &\longrightarrow X + O_2 \\ \hline O + O_3 &\longrightarrow O_2 + O_2 \end{aligned} \tag{3.98}$$

The reactive species X does not change abundance in the process and the net resulting reaction is the destruction of odd-oxygen species to form molecular oxygen.

The most important catalytic species are H, OH, NO, Cl, Br, and I. However, members of the different oxide families can react with one another, making the treatment of ozone production still hard to understand [ibid.].

Anthropogenic chlorofluorocarbons have an impact on the ozone depletion on the poles as well, generating the "Antartic ozone hole" [Farman et al., 1985], which is alarming for the survival of life on our planet.

If production and loss rates are at equilibrium, which is the case for short timescales that allow assuming the steady-state approximation, the concentration of the oxygen and ozone do not vary: however, since the oxygen abundance varied during the evolution of the Earth because of the appearance of life on the surface, ozone is supposed to have been very different in time. A different star-planet scenario could also play a role in the ozone abundance (triggering its production if the environment is rich of UV photons, or dampening it otherwise): the study of ozone abundance on exoplanets will be treated theoretically in Chapter 8.

In H_2-dominated atmospheres (i.e., reducing atmospheres) the main reactive agent is atomic hydrogen, produced by water vapor dissociation with hydroxyl radical OH. The latter reacts with H_2 to reform H_2O and H; therefore, the abundance of atomic hydrogen is much higher than that of OH [Hu et al., 2012].

H is more efficiently produced in this way, rather than direct photodissociation of H_2, which requires radiation in wavelengths less than 85 *nm*. The abundances of H and

OH, therefore, depend on the CO_2 mixing ratio, according to temperature. Atomic hydrogen is then removed by recombination to H_2O, enhanced by the presence of CO. Also, in this class of atmosphere, large quantities of O_2 and CO are supposed to exist, in addition to CO_2.

In N_2-dominated atmospheres (i.e., oxidizing atmospheres) both reducing radicals (H) and oxidizing radicals (O and OH) are relatively abundant: as in the latter case, H abundance is many orders of magnitude higher than OH abundance because water can be easily reformed [ibid.]. Despite its lower abundance, OH is important in removing CO, producing CO_2. Therefore, the reaction between H and CO is inefficient. Methane is photodissociated and oxidized slowly into methanol (CH_4O), producing atomic hydrogen also.

In CH_4-dominated atmospheres (i.e., highly oxidizing atmospheres) carbon dioxide is photodissociated and produces CO, O, and O_2. H is removed by O_2 and OH is removed by CO. As a result, O is the most abundant reactive species, together with an overabundance of CO and CH_4. Also, a small fraction of the volcanic SO_2 is transported in the atmosphere and converted into sulfuric acid aerosols.

3.4.5 Clouds

Clouds are supposed to be present on most of the planets with an atmosphere. They contribute to the albedo, reflecting at certain wavelengths and absorbing at others, thus changing the color and adding variability to the planet. Studying cloud formation could provide clues to the atmospheric behavior, such as the temperature variation or the planetary rotation rate. On the other hand, their presence weakens other spectral features and potential biosignatures.

Even if they are often used as synonyms, there is a difference between clouds, aerosols and hazes [Seager, 2010]:

- A cloud is a mass of liquid or solid particles suspended in an atmosphere. It can be produced by condensation either of atmospheric gases or photochemically produced compounds.
- Aerosols are particles of any size and kind floating in an atmosphere. They can be also smoke, pollutants or bacteria. This term is generally used to refer to very small particles that do not precipitate out of the atmosphere.
- Hazes are very small, opaque particles that are mostly photochemically produced.

Clouds composition varies depending on the atmosphere: on Earth, clouds are made of liquid water droplets or frozen water crystals, whereas on Hot Jupiters they could be composed of liquid iron and solid silicates. Also, planets could have different cloud layers with a different composition, depending on the condensation temperature of each compound. Condensation happens when the gas partial pressure $P_g(T)$ (the product of the gas mixing ratio X_g and the total gas pressure $P(T)$) exceeds the saturation vapour pressure $P_{V_g}(T)$ (Fig. 3.16):

$$P_g(T) = X_g P(T) \geq P_{V_g}(T). \tag{3.99}$$

The temperature dependence of the saturation vapour pressure follows the Clausius-Clapeyron equation [Sánchez-Lavega et al., 2004]:

$$\frac{dP_V}{dT} = \frac{L}{T(V_2 - V_1)} \tag{3.100}$$

Where T is the temperature, $V = 1/\rho$ is the specific volume (phase 1 = liquid/solid phase; phase 2 = vapour phase), and L [$J\ kg^{-1}$] is the specific latent heat of the phase transition. For an ideal gas, whose specific volume in phase 1 is small compared to that of the vapour phase, it becomes:

$$\frac{dP_V}{dT} = \frac{LP_V}{R_V T^2} \tag{3.101}$$

Where R_V is the specific vapor constant for the gas.

The integration of this equation provides information about the cloud vertical extent from a cloud base temperature T_c to a temperature T at a greater altitude in the atmosphere:

$$P_V(T) = P_V(T_c) exp\left[-\frac{L(T_c - T)}{R_V T T_c}\right]. \tag{3.102}$$

Making some further approximations, such as $L \approx const$, $T \approx T_c$, and assuming a convective atmosphere (see 3.4.3.2) so that $dT/dz = -g/c_P$ (where c_P is the specific heat capacity at constant pressure), it becomes:

$$P_V(T) = P_V(T_c) e^{-(z-z_c)/H_c} \tag{3.103}$$

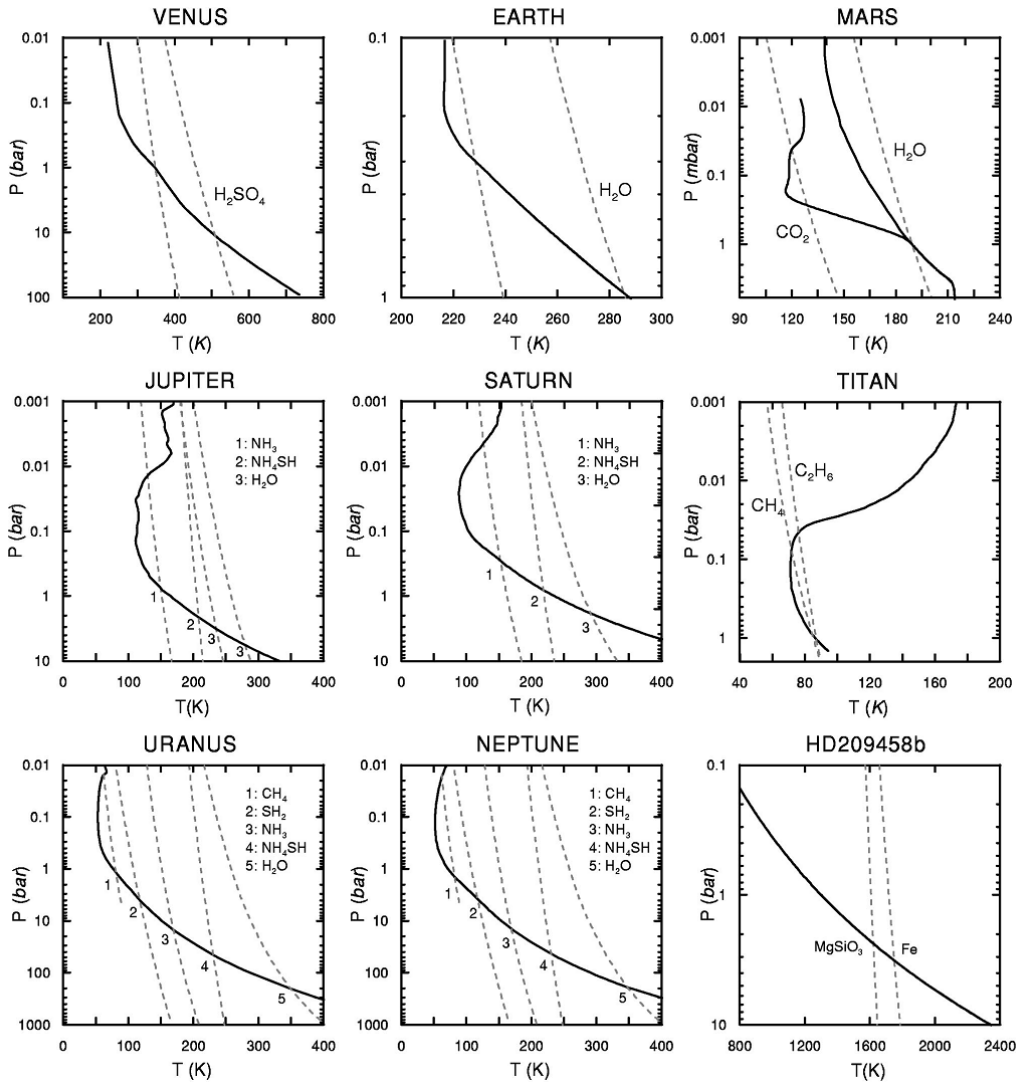


Fig. 3.16: Saturation vapour pressure curves (dashed lines) compared to planet vertical temperature profiles. Intersections among the lines mark the cloud base [Sánchez-Lavega et al., 2004].

Where z_c is the altitude corresponding to the temperature T_c, z the one corresponding to T, while H_c is the *cloud vertical scale height*, that is equal to

$$H_c = \frac{R_V T_c^2 c_P}{gL}. \tag{3.104}$$

A useful parameter is the ratio H_c/H, in order to better compare the cloud vertical extent among different planets:

$$\frac{H_c}{H} = \frac{R_V}{R}\frac{c_P T_c}{L}. \tag{3.105}$$

Generally, $H_c/H \approx 0.05 - 0.2$, so cloud layers are relatively thin compared to the height of the atmosphere.

The validity of the Clausius-Clapeyron relation is essential to allow cloud formation, but in reality, many processes concur and interact with one another to form a wide range of cloud droplets. These are nucleation, condensation/evaporation, sedimentation, coagulation, and coalescence [Montmessin et al., 2018]: many of such processes are still not very well considered in the treatment of exoplanetary atmospheres, thus causing the modeling of cloud formation and dynamics still challenging.

3.4.6 Atmospheric Escape

Atmospheric escape is a key process on every planet history: it is believed, for example, that early Earth lost abundant quantities of hydrogen during its cooling, thus allowing life to exist; also, some hot Super Earths orbiting close to their host stars lost the entire atmosphere and became more massive analogs of Mercury [Seager, 2010].

This process involves three major stages: firstly, the transport of gases from the lower to the upper atmosphere, where atmospheric escape could happen; secondly, conversion from the atmospheric gas (usually in a molecular form) to the escaping form (usually atomic or ionic); finally, the actual escape.

Escape could happen in three main ways, depending on the planet's mass, the upper atmosphere composition, the distance from the star, its activity, and the presence of a magnetic field:

- *Thermal hydrostatic escape*: an atom/molecule will escape a planet's atmosphere if its thermal velocity exceeds the escape velocity of the planet. So:

$$\sqrt{\frac{2k_B T_{exo}}{m}} > \frac{1}{6}\sqrt{\frac{2GM_P}{R_P}} \tag{3.106}$$

Where k_B is the Boltzmann constant, T_{exo} is the temperature of the exosphere (the uppermost layer of the atmosphere, where it becomes collisionless), m the atomic/molecular mass, $G = 6.674 \cdot 10^{-11}\ m^3\ kg^{-1}\ s^{-2}$ the gravitational constant, M_P and R_P the planet mass and radius. The factor $1/6$ gives an estimate of the escape based only on the velocities in the high-energy tail of the velocity distribution.

A more accurate description can be made using "Jeans escape" view. In this scenario, velocities follow a Maxwellian distribution:

$$f(v)dv = \frac{4n}{\sqrt{\pi}}\left(\frac{m}{2k_B T}\right)^{3/2} v^2 e^{-mv^2/2k_B T} dv \tag{3.107}$$

Where n is the number density of particles, that at constant temperature and in hydrostatic equilibrium approximation is equal to:

$$n(r) = n(r_0)e^{-r/H_0} \tag{3.108}$$

r being the radial distance from the planet center and $H_0 = \frac{k_B T}{\mu m g(r)}$ the scale height in r. The Jeans escape flux can be obtained by integration:

$$\Phi_{Jeans}[cm^{-2}\ s^{-1}] = \int_{v_{esc}}^{\infty}\int_{0}^{2\pi}\int_{\pi/2}^{0} f(v)\cos\theta\sin\theta d\theta d\phi \tag{3.109}$$

$$= \frac{n_c}{2\sqrt{\pi}} B\sqrt{\frac{2k_B T_c}{m}}(1+\lambda_c)e^{-\lambda_c} \tag{3.110}$$

Where the subscript c refers to the properties of the base of the exosphere (*exobase*) and the factor $B \approx 0.5 - 0.8$ accounts for the slow repopulation of the energetic tail of the distribution. Here, λ_c is the *escape parameter*:

$$\lambda_c = \frac{GM_p m}{k_B T r_c} = \frac{1/2mv_{esc}^2}{k_B T_c} = \frac{r_c}{H}. \tag{3.111}$$

- *Thermal hydrodynamic escape*: the atmosphere behaves like a fluid expanding radially outward, as particles are so energetic they cannot be held by the gravitational field. It typically happens on planets very close to stars with large EUV flux.

  The Jeans escape parameter can be used to divide hydrostatic and hydrodynamic regimes: for $\lambda_c \gg 1$ a hydrostatic process occurs, while $\lambda_c \ll 1$ is necessary but not sufficient to drive hydrodynamical flow. It is a complex process that requires numerical models and that has still many uncertainties, due to the various approximations that are made.

- *Nonthermal escape*: there are collisional processes between charged species that produce energetic atoms, which escape from the atmosphere. These collisions enable the escape of heavier elements such as N, C, and O. It could also refer to sputtering and stellar wind erosion, mainly on planets and asteroids with a very thin atmosphere or none at all. Also, charge exchange processes occur, as EUV radiation photodissociates atoms creating ions; they could be then accelerated by the magnetosphere or recombine, adding excess kinetic energy to the newly formed molecule/atom.

3.4.7 Atmospheric evolution

The present configuration of the Earth, for what concerns atmosphere, internal composition, hydrosphere, and biosphere, has not been always the same throughout the lifetime of our planet. In the research of Earth-like exoplanets, it is very likely to find planets at any stage of their evolution: the history of the Earth can, therefore, be a guide to characterize such planets [Vasquez et al., 2010].

The main geological period of the Earth are the Precambrian (composed by Hadean, Archean, and Proterozoic) from 4500 to 550 Ma (i.e. million years ago) and the Phanerozoic age (composed by Paleozoic, Mesozoic, and Cenozoic), from 550 Ma to the present.

During the Hadean era (4500-3800 Ma), the Earth was formed via runaway accretion in the solar protoplanetary disk; the internal structure then differentiated. Further collisions then enriched the atmospheres of volatiles and silicates, which eventually precipitated on the surface, leaving a residual hot atmosphere rich in carbon dioxide and water vapor. The gradual cooling of the atmosphere caused the water vapor to precipitate as well, forming the oceans (about 4.4 Ga, i.e. Gigayears ago). After a period of quietness concerning impacts with other planetesimals, during which

life may have emerged, around 3.85 Gigayears ago a new period of strong impacts (the so-called "Late Heavy Bombardment" or "LHB") caused abrupt changes in the atmosphere, truncating the evolution on any life form. It is only after the LHB that life appeared on the Earth.

During the Archaean Era, the internal heating was more intense and lead to stronger plate tectonics. Cyclic processes of carbon dioxide started to be effective, cooling the atmosphere and allowing life to evolve in three groups (Bacteria, Archaea, and Eucaryotes) from a common ancestor. Bacteria and Archaea compose the unicellular life-forms and started to appear about 3.5 Gigayears ago, when the temperatures were presumably in the 55-85 °C range. At this point, the Solar luminosity would have been about 30% weaker than the present time and should have triggered a climatic unhabitable *snowball Earth*, but this presumably didn't happen (this is often referred to as the *faint Sun paradox*). The most likely explanation to this apparent contradiction could be the presence of larger abundances of CO_2 that increased the efficacy of the greenhouse effect and allowed the surface to be habitable.

The development of the carbon dioxide cycle, that translated the outgassed CO_2 from the atmosphere to the crust and mantle via weathering and metamorphism, acted as a thermostat of the global climate and the habitability of the Earth [ibid.].

The presence of life has also contributed to the thermal evolution of the atmosphere as well, producing a large amount of methane, another powerful greenhouse gas. It is destroyed both via photolysis and reacting with the OH radical, and it is also likely to form aerosols. For it to be so abundant in the primordial atmosphere, a biological source was required.

Now, the atomic oxygen was produced from the photolysis of water vapor and carbon dioxide and in turn, could produce O_2 and O_3. Molecular oxygen would, however, be removed by weathering or red-ox reactions, so the overall concentration was about 10^{-4}% of the current abundance (hereafter "Present Atmospheric Level" or "PAL").

At the beginning of the Proterozoic era (2500-500 Ma), the development of *cyanobacteria*, photosynthetic prokaryotes, caused a massive increase in atmospheric oxygen through photosynthesis, up to 1% PAL: this change is often called the "Great Oxygenation Event". Short after that, the ozone layer should have been formed.

During this age, two global glaciations may have been occurred, caused by the reduced solar luminosity and the increase in O_2 which caused the depletion of methane and carbon dioxide. Life, however, survived in the bottom of the oceans.

At the end of the Proterozoic, the oxygen had reached 10% of the current abundance. The presence of ozone and its shielding effect from the harmful UV radiation allowed life to develop on the surface.

The Phanerozoic era (600 Ma-present) begins with a flourishing of multicellular life forms (known as Cambrian Explosion). Carbon dioxide was efficiently buried in the mantle and the oxygen increased as a consequence. At this age, the tectonic planets broke and moved, thus causing the drift of continents. The change in the location of ocean basins and of the continents, which produced variable amounts of CO_2 [Kasting and Catling, 2003], caused the climate to alternate between hot and col scenarios with a periodicity of nearly 140 million years.

Different biological species evolved, some becoming extinct and others emerging; mass extinctions could have been caused by the climatic periodicity, or a period of intense asteroid impacts. Periodic giant volcanic events may have had an impact on these catastrophic events.

From the last ice age, about 11 thousand years ago, the climate has been fairly stable. The current geological age is known as Anthropocene and it is characterized by the effects that the presence of the humankind has on the climate and the ecology of the planet. The most important event is the industrial era, that caused the production of carbon dioxide and, as a consequence, an increase in atmospheric average temperature. The anthropogenic effects may cause a variation on the climate for the next thousands of years.

The future increase of solar luminosity is likely to drive the atmosphere towards higher temperatures, possibly causing the establishment of a runaway greenhouse effect, causing the oceans to evaporate (see Section 4). At the same time, the weathering could reduce the CO_2 abundance up to levels that forbid the survival of plants. So, photosynthetic life may soon disappear – but a chance of survival may be given to organisms that use chemical energy for their metabolism, possibly underground [Vasquez et al., 2010]. In 5.5 billion years from now, the Sun will become a red giant and expand its diameter reaching 0.77 AU, marking the end at least of the terrestrial planets of our Solar System.

4 Habitability

> *Il creato che contempla l'astronomo non è un semplice ammasso di materia incandescente: è un prodigioso organismo in cui dove cessa l'incandescenza della materia incomincia la vita.*
>
> ---
>
> *"What the astronomer admires is not the simple clustering of incandescent matter: it is a prodigious organism in which the incandescence of matter ceases and life starts."*
>
> — **Padre Angelo Secchi (1818-1878)**
> Italian Jesuit and astronomer

This chapter introduces to the current knowledge about the so-called "habitable zone" and the factors that affect the habitability on a planet. Finally, a review of the main tracers of life in a planet's atmosphere is summarized.

4.1 Habitable Zone

At the core of exoplanet search is the hunt for habitable environments beyond Earth [Gelman et al., 2011]. Life on Earth requires "environments [that] must provide extended regions of liquid water, conditions favorable for the assembly of complex organic molecules, and energy resources to sustain metabolism" (NASA Astrobiology Roadmap, Goal 1).

Life structures on Earth are based on complex organic molecules, mainly composed by carbon: this element is particularly versatile since it can form up to four covalent bonds at a time. Most of the carbon-biomolecules are composed of hydrogen, oxygen, nitrogen, in addition to carbon itself: they are jointly called "CHON".

Liquid water is essential since [Vasquez et al., 2010]:

- it is the main dipolar solvent for CHON molecules;
- it remains in a liquid state over a large temperature range;
- its freezing point decreases when salts are dissolved in it;
- its polarity results in the formation of hydrogen bonds, especially with OH, NH and SH groups, in addition to carbon;
- its large dipolar momentum favors the dissociation of ionizable groups and improves their solubility;
- it conducts heat more easily than any liquid (except mercury);
- it has a neutral pH;
- it has a high surface tension, so it sticks to the sides of tiny vertical structures, facilitating capillarity;
- it has high specific heat and this allows organisms to regulate their body temperature;
- being ice less dense than liquid water and thus floating on oceans or lakes, life could survive in such extreme environments.

Many authors [Vasquez et al., 2010, and references therein] support the universal nature of our biochemistry, but much exotic chemistry-based life (such as silicon with liquid nitrogen or liquid methane as solvent) have been proposed, even if they should be less common than the carbon-based life.

The region around a star in which life-supporting planets can exist has been termed the *"habitable zone"* or "HZ" [Kasting, Whitmire, et al., 1993] (see Figure 4.1): this definition regards all water-based life forms, so not all planets would make suitable homes for humans.

This definition came up well before the discovery of the first exoplanet orbiting a main-sequence star, and much progress has been done since then in the field. In general, theoretical models are required to define the actual boundaries of the habitable zone, and different estimates have been calculated based on the various processes included in such models: as a matter of fact, the atmosphere plays a substantial role in the habitability of the planet, and very little is acknowledged in the treatment of more complex phenomena happening in it. Moreover, the stars undergo many variations in luminosity during their lifetimes, so much that the

habitable zone boundaries change with time. Many other processes do have an impact on the habitable zone as well (see Section 4.2).

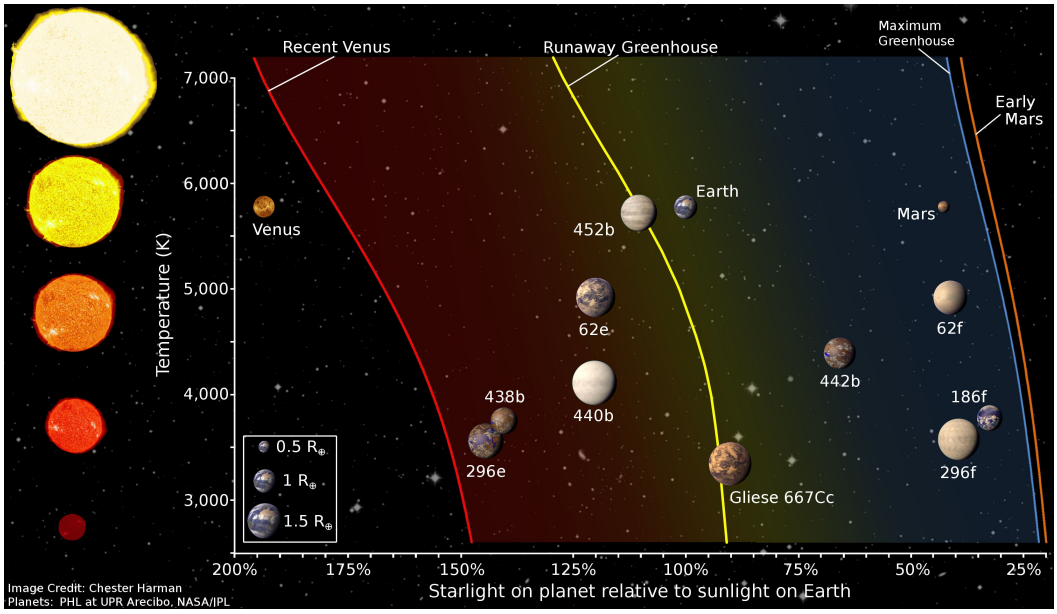


Fig. 4.1: The habitable zone boundaries as a function of the stellar temperature and the incoming stellar radiation.

In general, two main processes define the inner and outer boundary of the HZ [Kopparapu, 2018]:

- The inner boundary is defined either by the *runaway greenhouse effect* or the *moist greenhouse effect*. The first one occurs when the surface temperature increases rapidly due to the high outgoing thermal radiation, causing the oceans to evaporate. The latter occurs when water vapor in the stratosphere becomes so abundant that the planet starts to lose hydrogen by photolysis of water vapor, so much that water cannot form anymore. In general, water vapor plays a role in the determination of the inner boundary, either by increasing the greenhouse effect or by being destroyed by the upcoming photons, so Earth-like planets close to this boundary are expected to have a water-rich atmosphere or to have lost their water reservoir to space [Kaltenegger et al., 2009].

- The buildup of carbon dioxide in an atmosphere is linked to the volcanic activity and can be regulated mainly by silicate weathering, which is more effective if the surface temperature is above freezing. On planets that receive lower irradiations, the weathering may not destroy enough CO_2 in the atmosphere, which eventually would start to condense in clouds. As a consequence, the

Rayleigh scattering of the atmosphere would be increased and overcome the greenhouse effect that the carbon dioxide itself produces. The greenhouse effect is maximized at the so-called *maximum greenhouse effect* boundary: at a farther distance, CO_2 would cause surface cooling.

Other empirical limits may be defined. For the inner boundary it is possible to determine the "recent Venus" limit, based on the inference that Venus has not had liquid water on the surface for the last 1 billion years at least; on the other hand, an "early Mars" limit, based on the inference that Mars had liquid water on the surface 3.8 billion years ago can be taken as an outer boundary.

The most common inner and outer boundaries of the habitable zone for a Solar-like star are shown in Table 4.1.

Tab. 4.1: Available estimates for the inner and outer boundaries of the habitable zone of a Solar-like star, with the respective references.

Inner Boundary (AU)	Outer Boundary (AU)	Reference
0.88 – 0.912		Rasool et al. (1970)
0.95	1.01	Hart (1979)
	3.0	Fogg (1992)
0.725	1.24	Dole (1964)
	1.385 – 1.398	Budyko (1969)
0.75		Abe et al. (2011)
0.95	1.37	Kasting, Whitmire, et al. (1993)
	2.0	Spiegel et al. (2010)
	10	Pierrehumbert et al. (2011)
0.95		Leconte et al. (2013)
0.38		Zsom et al. (2013)
0.77 – 0.87	1.02 – 1.18	Vladilo et al. (2013)
0.99	1.70	Kopparapu et al. (2013)
0.95	2.4	Ramirez et al. (2017)

4.2 Factors Affecting Habitability

As we have seen, the survival of life on Earth is based on three factors: the energy source, the CHON elements, and water. The first two are fairly common in the universe, while the presence of liquid water on the surface of a planet can be challenging to assess [Meadows et al., 2018].

The definition of the habitable zone is basically linked to the orbital distance of a planet and to the effective temperature of the host star. This definition does not

take into account any other process that may alter the surface enviroment of the planet itself: the atmosphere could have been eroded during the evolution, or the planet could have been formed with little or no water at all. So, an accurate study of the habitability of a planet must consider many other factors. A summary of the planetary, stellar, and orbital properties that do contribute to the presence of liquid water on the surface is shown in Figure 4.2.

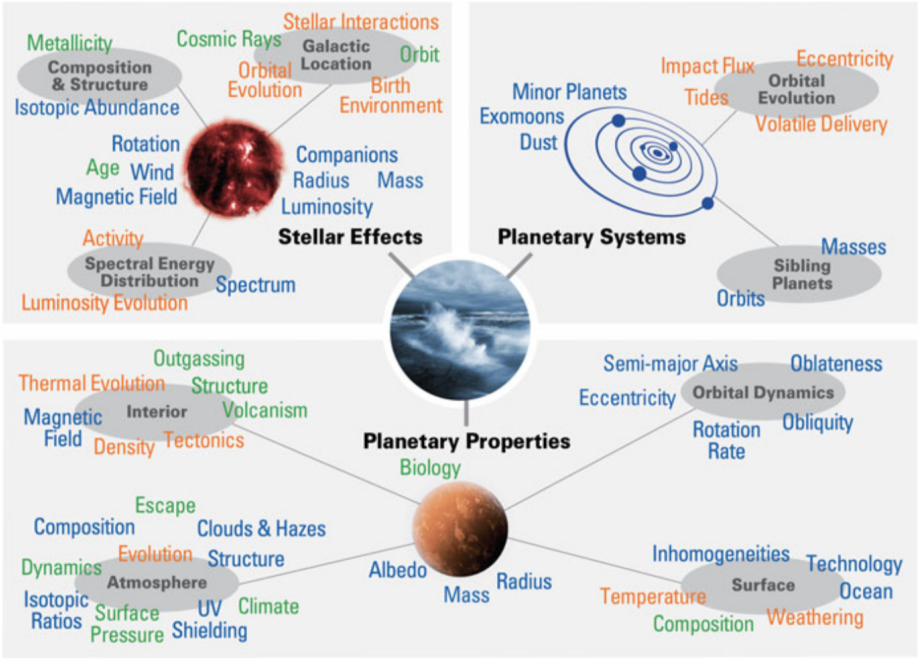


Fig. 4.2: Factors affecting habitability. In blue: the characteristics that could be observed directly with sufficiently powerful telescopes; in green: the characteristics that require modeling interpretation constrained by observations; in orange: all information that can be only accessible via modeling [Meadows et al., 2018].

Planetary Properties First of all, a planet needs to be terrestrial and to possess sufficient initial volatile content to be able to form an ocean. It should also lose its primary atmosphere but retain a secondary atmosphere over time, as well as some kind of water reservoir. To do that, the geological activity must be present to expel new gases in the atmosphere and favor climatic evolution.

The mass of the planet impacts the internal heating from radionuclides and enhance tectonics. It also impacts the ability of the planet to retain an atmosphere, by determining the escape velocity on the upper layers of the atmosphere itself (see Eq. 3.106).

The presence of magnetic fields prevents the atmosphere to extend too much under the action of the stellar radiation, thus inhibiting mass loss by the stellar wind. Magnetic fields are linked to a convective dynamo in the iron-rich core, a process that is influenced by the planetary rotation, the internal composition, and the core-cooling rate (enhanced by plate tectonics, if present).

The orbital parameters (semi-major axis, eccentricity, obliquity) drive the radiation received by the planet and, thus, the climate. The planets that are tidally locked (i.e. they show the same hemisphere towards the star throughout the whole orbit, since the rotational period is the same as the orbital period) are most likely to experience extreme climates, with larger temperature differences between the day and the night side. At this point, if advective processes take place, the terrestrial conditions could still be favorable to life, but catastrophic events may occur if the radiative timescale is shorter than the advective one: the atmosphere may be stripped from the day hemisphere, and could condense and collapse in the night side. Different resonance motions may contribute to complicating things further.

The presence of an atmosphere indeed can favor or preclude habitability. The atmosphere can provide, as a matter of fact, the pressure required to maintain liquid water on the surface, and subsequently changes the range of temperatures for which water can exist in its liquid form: as visible in Figure 4.3, the phase changes curves of any chemical species are strongly dependent on the values of pressure and temperature. Therefore, higher atmospheric pressures should imply larger ranges of liquidity for water on the surface; for the same reason, the upper layers of the atmosphere (whose pressure is lower) may experience condensation of water vapor (or other species, as carbon dioxide) into droplets or ice particles – a phenomenon that is indeed common in our atmosphere, and impacts on the albedo of the Earth.

One may, therefore, think that the higher the pressure, the more favorable the conditions for life to appear and survive may be. This is not always the case, since atmospheres that are too dense may not lose the primordial H-H_2 envelope, compromising the overall habitability. The right atmospheric pressure has also an impact on the survival of the atmosphere, especially in planets that are synchronously rotating.

The chemical composition of the atmosphere plays a role as well: a large abundance of greenhouse gases, clouds, and hazes, may modify the environmental conditions in opposite ways – either enhancing favorable conditions for life (e.g. by shielding stellar UV radiation), or forbidding life to appear at all (e.g. by increasing the temperature up to the evaporation boundary) [Meadows et al., 2018].

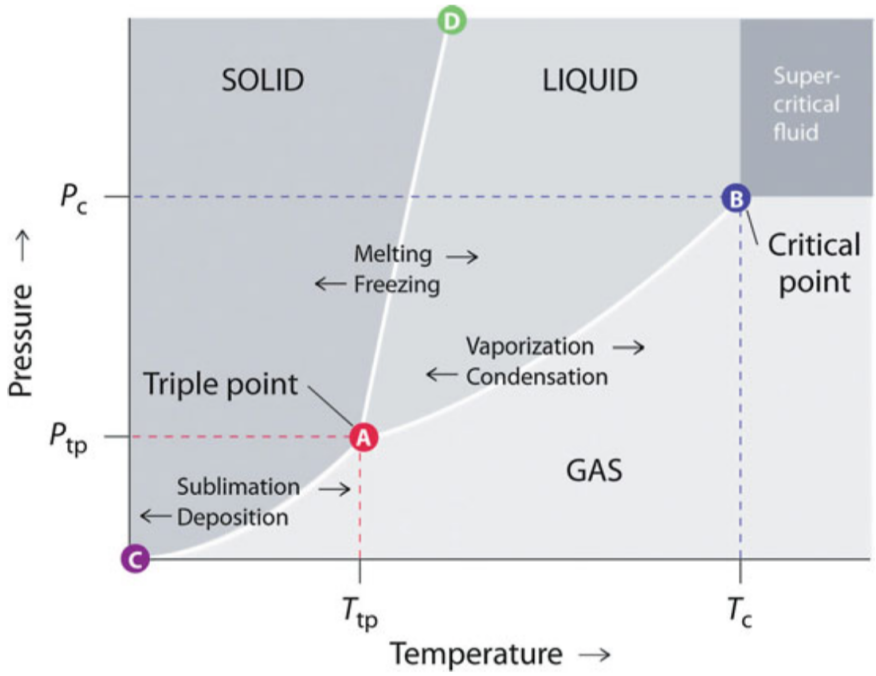


Fig. 4.3: A schematic of a typical phase diagram showing the behaviour of the boundaries between the solid, liquid, and gas phases with respect to temperature and pressure variations. The triple point A is where all phases coexist together (for water, $P_{tp} = 611.657\ Pa$ and $T_{tp} = 273.16\ K$); the critical point B is the point after which liquid and gaseous phase become indistinguishable and the species is a supercritical fluid (for water, $P_{tp} = 22.064\ MPa$ (217.75 atm) and $T_c = 647.096\ K$). The line that connects A and D is the solid-liquid phase boundary; the line that connects A and B is the liquid-gas phase boundary; the line that connects A and C is the solid-gas phase boundary [Montmessin et al., 2018].

Photochemistry will also contribute to varying the atmospheric composition, either destroying or creating greenhouse gases, or enhancing the hydrogen escape after the photolysis of water vapor. Photochemically-produced hazes may also impact the albedo and influence the climate.

A stable and favorable climate is the key to habitability, but it strongly depends on the irradiation, the stellar effective temperature, the composition of the atmosphere and dynamical processes, which are still hard to be modeled properly.

Stellar Effects The stellar environment changes the chemistry of the atmosphere and, subsequently, contributes to control the surface conditions. The effective temperature of the photosphere (and thus the shape of the spectrum), as well as the level of activity, can enhance escape and cause abrupt changes in the atmospheric evolution.

The mass and temperature of the host star is also a limiting factor for the appearance of life on any orbiting planet: hot, massive stars would remain in the main sequence phase for too short a time to allow the planet to form, cool and develop a consistent biosphere. Cooler M dwarfs are, on the other hand, expected to stay in the hydrogen-burning phase for as long as 10000 Gyrs and, for this reason, such targets could be ideal ones to look for life in the universe. However, some M dwarfs are very active and produce UV and shorter wavelengths that could be detrimental to any form of life on the surface.

The activity is linked to the stellar rotation and evolves throughout time, slowly decreasing as long as the star ages. The spindown times are however still not clear from current observations.

The luminosity also changes in time, pre-main-sequence stars being brighter at first, and then cooling up to their main-sequence status. The habitable zone, therefore, changes and shrinks towards shorter distances: any forming planet which now resides in the habitable zone of its main-sequence host star may have been originally subject to much higher irradiations that could have caused the evaporation of volatiles – thus lacking liquid water even if the observed temperature now could be favorable.

A strong stellar magnetic field may also enhance the atmospheric escape via EUV/XUV photochemistry ad impacts, as well as reduce the planetary magnetic field, exposing, even more, the atmosphere to erosion by the stellar wind.

Gravitational effects with both the star and sibling planets, if any, may cause a variation in the orbit (if the eccentricity is non-null) and therefore in the magnitude of the incoming radiation. The tidal deformation caused by the star-planet interaction may heat the planetary interior, enhance the tectonic activity, change the obliquity, and reduce the magnetic field. Tidal forces may also cause the production of a magma ocean for very close-in planets before the circularization of their orbits, mining all chances of survival of any life-form on the surface.

The galactic environment may also play a role in the determination of the habitability: cosmic rays may contribute to the atmospehric erosion; close encounters with nearby stars can influence the orbital motion of the planet, or even induce migration or ejecting planets from their systems (producing *rogue planets*).

Many host stars can be binary or multiple systems, thus varying the habitability zone of any planet orbiting either one or both companions.

Planetary Systems The presence of other planets orbiting the star may affect the abundance of volatiles during the formation of a potentially habitable planet, as well as cause resonance motions. Minor bodies may also vary the water reservoir on a planet, exchanging volatiles via impacts with the planet itself. Exomoons can cause tidal motions and obliquity oscillations that could be challenging for habitability purposes.

All those processes are strongly interconnected and contribute massively to the actual habitability of a planet. The definition of a "habitable zone" works for first-order filtering, but eventually, other mechanisms are to be considered: some of them may rely on future observations (planetary and orbital parameters, stellar composition and characterization...), but the determination of the others must require a theoretical approach.

4.3 Biosignatures

Life alters the chemical composition of a planetary atmosphere, by producing peculiar fingerprints on the trasmission/emission spectrum. As visible in Figure 4.4, there is a substantial difference from the emission spectrum of the Earth if compared to the other (most likely lifeless) terrestrial planets of the Solar System: carbon dioxide

is present in all planets, but absorption features of ozone and methane are present uniquely in our planet's spectrum, enhanced by the presence of life on the surface.

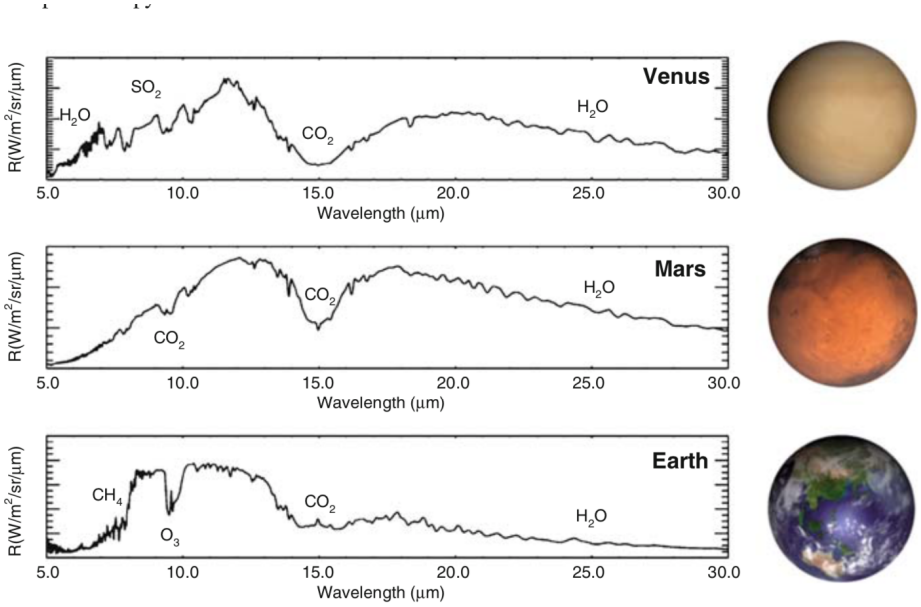


Fig. 4.4: Infrared spectral radiance for different planets of the Solar System [Vasquez et al., 2010].

One should expect to find similar features in any life-bearing planet. For this reason, when looking for transmission spectra of terrestrial exoplanets, an astrophysicist should look for the *biosignatures*: these are currently defined as "objects, substances, and/or patterns whose origin specifically requires a biological agent" [Kiang et al., 2018, and references therein]. Therefore, this definition includes biosignature gases in the atmosphere produced or enhanced by photosynthetic activity, such as oxygen O_2, ozone O_3, methane CH_4, nitrous oxide NO_2, and methyl chloride CH_3Cl, as well as a surface biosignature (the so-called *red edge*), which is the unique reflectance spectrum from plants or photosynthetic bacteria on the surface of the planet. Another signature of life on Earth is the *redox disequilibrium* of its atmosphere, i.e. the presence of simultaneously reducing and oxydizing species (such as methane and molecular oxygen), which would be removed in short timescales from the atmosphere if not for the presence of life [Grenfell, 2018].

These spectral fingerprints change throughout the evolution of life on a planet: indeed, during the Archean Eon of the Earth ($\approx$ 4.0-2.5 billion years ago) there was no trace of oxygenic photosynthetic bacteria, but anaerobic microbes were present. Moreover, different irradiance conditions may change the reflectance vegetation spectrum, as different photosynthetic processes may appear to be sensitive to different wavelengths: so, bacteria on a planet orbiting an M star could be

responsive to near-infrared wavelengths to be able to gain energy for their survival. One could even imagine terrestrial planets whose vegetation would be purple, rather than green, if anoxygenic phototrophs are present [Kiang et al., 2018, and references therein]. One may, therefore, define another category of biosignatures: the *temporal* biosignatures, which account for seasonal/climatic changes in gases or surface biosignatures. In other words, this kind of biosignatures would be modulations in measurable quantities that could be linked to the actions and time-dependent patterns of the biosphere [Schwieterman et al., 2018].

However, the majority of the features that can be identified on an observed spectrum are likely to be caused by abiotic processes, completely unrelated to biospheric activity. It is, therefore, preferable to consider them as "potential biosignatures", rather than actual evidence of life on the surface [Schwieterman, 2018]. When addressing the topic of biosignatures, we refer to terrestrial-like forms of life: this does not preclude the existence of evidence of other life-forms, but the carbon-based life is what we are able to recognize and compare.

4.3.1 Gaseous biosignatures

The features of the main biosignatures are the following [Grenfell, 2018; Claudi and Alei, 2019]:

- **Oxygen (O_2)**: O_2 is a rather inert atmospheric species that is present in modern Earth's atmosphere with a volume mixing ratio of 0.21 up to ≈80 km altitude. At higher levels, it is photolyzed and recombines (generating the "oxygen airglow" of the atmosphere). Its main source is the photosynthetic activity coupled with the burial of organic material into the mantle; a weaker source is the photolysis of water vapor followed by the hydrogen escape. Its main sinks are the reactions with volcanic gases and surface weathering. On Earth's history, no substantial molecular oxygen was present in the atmosphere during the Archaean period (but a strongly reducing atmosphere); then, ≈2.5 Gyr ago, an initial rise of O_2 took place – the so-called "Great Oxygenation Event" (GOE), followed by a second one (the "Second Oxygenation Event" or SOE) about 0.6 Gyr ago. Possible explanations of these events could be a faster burial rate or a decrease in volcanic emissions.

  Since the abundance of molecular oxygen is driven both by biotic and abiotic processes, it is necessary to rule-out any false positive. Burial of material in the mantle is linked to plate tectonics, whose efficiency is still uncertain for

the Super Earths case. On the other hand, the O_2 source due to the hydrogen escape could be more efficient on such exoplanets with respect to the terrestrial case. On the other hand, methane abundance (an oxygen sink) could be up to 1000 times higher on Earth-like planets orbiting M-dwarfs. As of today, the sinks for oxygen (and ozone) are yet poorly understood, so modeling studies about these species still lack quantitative and reliable results.

The most important spectral features of O_2 are the so-called "A band" at 0.76 μm, the "B band" at 0.68 μm, as well as some near-infrared features at $\approx$1.3 μm. A strong UV band at around 0.12 μm is also present [Bétrémieux et al., 2013].

- **Ozone (O_3)**: On current Earth's atmosphere, the ozone layer extends from 20 to 50 km with a maximum abundance of $\approx$10 parts per million (ppm) at about 30 km. It is photochemically produced from molecular oxygen, and at the same time photochemically destroyed to form back O_2 (the Chapman cycle, see Section 3.4.4.2). It is also destroyed by certain families of gases, such as HOx, ClOx, NOx, by means of catalytic cycles. It is mainly produced at lower latitudes and then transported towards the poles; it efficiently shields the surface by absorbing UV radiation, which in turn heats up the stratosphere. While the Chapman mechanism is efficient in producing stratospheric ozone, an alternative mechanism (the "smog mechanism") is responsible for its tropospheric production, which constitutes typically about the 10% of the total column density of the species.

  Due to its dependence from molecular oxygen, it is generally expected that the ozone production proceeded rather at the same time as the GOE. Due to the nature of its source reactions, the current ozone abundance should have been reached in a very short time after the Great Oxygenation Event.

  On exoplanets, ozone could be a potentially significant biosignature, especially on planets orbiting active M-dwarf stars. Its main spectral feature occurs at 9.6 μm, but other ones are present in the visible and ultraviolet range, such as the Chappuis band (0.5-0.7 μm) and the Hartley band for wavelengths shorter than 0.3 μm. It could be still challenging to observe such features even with future facilities, due to the overlapping of ozone and carbon dioxide bands, weakening by clouds, and possible interferences due to the presence of an exomoon.

- **Nitrous oxide (N_2O)**: This species is mainly produced by biological activity and it has a current volume mixing ratio of $3.3 \cdot 10^{-7}$. It is destroyed by pho-

tolysis and reactions with excited oxygen atoms. It is also a strong greenhouse gas, with infrared absorption bands, and it has a long chemical lifetime of several hundred Earth years [Lammer et al., 2019]. Its origin during the Earth's evolution is still debated, probably caused by lightning or cosmic rays. It can be sensitive to the ozone abundance, which regulates the UV reaching the surface. Nitrous oxide produces weak spectral features at 7.8, 4.5, and 3.7 μm. which overlap with methane and water bands. On low-UV environments, these features could become more prominent, while high UV irradiances would cause destruction via photolysis.

Further information on this species may be found in Segura et al. (2005), Rauer, Gebauer, et al. (2011), Rugheimer, Kaltenegger, Zsom, et al. (2013), Rugheimer, Kaltenegger, and Sasselov (2015), and Airapetian et al. (2017).

- **Methane (CH_4)**: Its volume mixing ratio is currently $\approx 1.8 \cdot 10^{-6}$, mainly associated with the respiration of methanogenic bacteria (and some anthropogenic and abiotic sources as well). Minor sources are geological as well. Its loss is driven by the reaction with the hydroxyl (OH) radical, as well as dry deposition and photolysis. It is unreactive in the troposphere and its main lifetime is $\approx$ 8 years: for this reason, it is a tracer of tropospheric dynamics as well.

  In the past, it may have been 500 times more abundant, because of a massive production by methanogenic bacteria and an increased lifetime due to the reducing atmosphere. Enhanced volcanic activity may have contributed as well to this overabundance.

  Exoplanets orbiting M-dwarfs could have up to 1000 times more methane with respect to current Earth, especially for low-UV environments that lead to less abundant OH. Spectral features of methane appear at $\approx$3.4 and $\approx$7.7 μm, but they could be mixed with water vapor bands at low spectral resolutions. Higher abundances of this species may lead to additional bands at $\approx$1.1, $\approx$1.4, $\approx$1.7 and $\approx$2.4 μm.

  Further information may be found in Krasnopolsky et al. (2004), Rugheimer, Kaltenegger, Zsom, et al. (2013).

- **Chloromethane (CH_3Cl):** This species is mainly produced by biological activity (e.g. ocean plankton, fungi, wood rotting) and destroyed by the OH radical and biological degradation. It may be produced by abiotic sources as well. Its features (e.g. 1.3. μm) are weak in the Earth atmosphere but could be more prominent in planets orbiting inactive M-dwarf stars.

Further information on this species may be found in Segura et al. (2005) and Rugheimer, Kaltenegger, and Sasselov (2015).

- **(Organo)Sulfur Gases (DMS, DMDS, CH_3SH)**: These species are direct or indirect products of metabolism [Schwieterman et al., 2018]. They can be destroyed easily by ultraviolet photons, so it is unlikely to spot them in UV-rich environments. Some output products such as C_2H_6 may, however, be tracers of the presence of sulfur gases in the atmosphere – but this compound may be also produced by abiotic sources (such as volcanism). The strongest features of such gases are in the mid-infrared range; the bands notably overlap with the O_3 ones. Further information on these species may be found in Pilcher (2003), and Domagal-Goldman et al. (2011).

The atmospheric redox disequilibrium, i.e. the presence of two species with differing redox states, one oxidizing as molecular oxygen, and one reducing as methane, is most likely to be caused by the presence of life: as a matter of fact, these two species would not be present in such abundant concentrations, being extremely reactive with each other, if a constant replenishment was not present [Kiang et al., 2018].

Simultaneous detection of an oxidizing and a reducing species could be considered, thus, a biosignature. These species are associated with cyanobacteria (that produce O_2) and methanogenic bacteria (that produce CH_4): without life, these species would be easily removed by fast reactions in the atmosphere. However, the biotic production of oxygen and methane seldom have been equally efficient at the same time in Earth's history: it may, therefore, be hard to spot both signature gases on an exoplanetary spectrum.

Water and CO_2 are not biosignatures since they could not be considered as a sign of life by themselves: they are, however, essential for life, being by-products of many reactions generating biosignatures as well as greenhouse gases extremely important for planetary habitability purposes [Claudi and Alei, 2019].

The simultaneous detection of CO_2, O_2, and H_2O are considered to be very important for the assessment of life on an exoplanet: these species produce strong features in our spectrum and are sometimes referred to as the *triple fingerprint*. Detecting a similar pattern on an exoplanet may be another potential clue of the presence of life on the surface.

However, the chemicals produced by life on Earth are hundreds of thousands (estimated from plant, microbial, and marine natural products), but only a subset of hundreds are volatile enough to enter the atmosphere at more than trace concentrations.

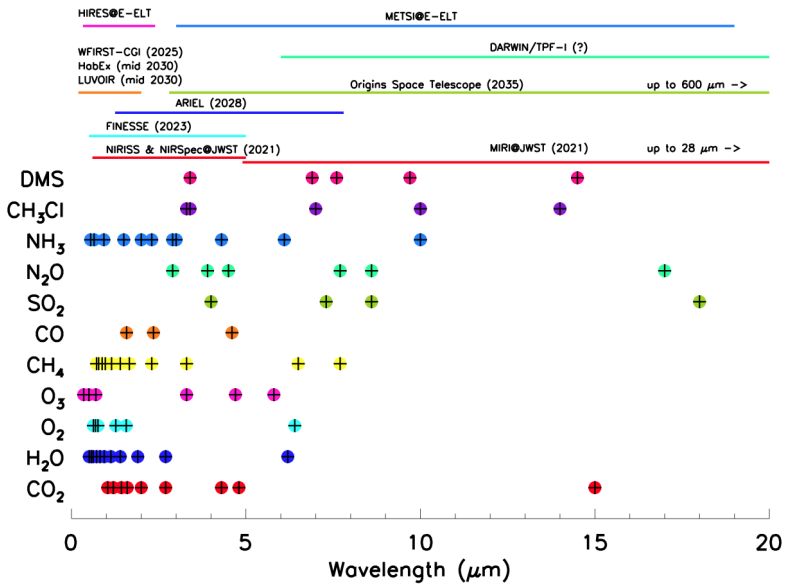


Fig. 4.5: The potentiality of the future instrumentation (ground- and space-based) in detecting biosignatures in optical, near and mid infrared. The colored filled circles identify the central wavelength of the absorption band of the molecules named on the vertical axis. On top of the plot, the wavelength range of the considered instruments is indicated by the corresponding horizontal colored solid line. The use of colors has the aim to make the plot clearer [Claudi and Alei, 2019].

Future generation facilities will open new paths to the characterization of the new worlds discovered by the demographic space missions (e.g. Kepler, TESS and PLATO). Extremely large telescopes, in particular, would allow the detection and characterization of Earth-like candidates in the habitable zone via the direct imaging technique.

The potentiality in the biosignature detection of all of these future ground based instruments and space missions is summarized in Figure 4.5. In the figure, the central wavelength of each absorption band of the different molecule is listed. In the upper part of the plot, the wavelength ranges of all planned future missions/instruments are plotted [ibid.].

4.3.2 Surface biosignatures

A surface biosignature comes from living material that causes an inferable spectral or polarization marker on reflected, transmitted, or scattered light [Schwieterman, 2018].

The survival of vegetation on our planet relies on photosynthesis: this is the process that allows the biosphere to convert photons to energy, which can be used to produce biomass. Due to the direct access with the solar photons, photosynthetizers are the most productive form of life on Earth and, therefore, it is plausible to assume that, if present on an exoplanet, they would be the most detectable ones as well.

Although the photosynthetic process may be different depending on the environment, it generally relies on a redox reaction that uses photons to excite electrons from protein-pigment complexes [Schwieterman, 2018]. The electrons are transported to acceptor molecules and replaced by electron donors found in the environment.

A general equation for photosynthesis would be:

$$CO_2 + 2\,H_2A + h\nu \longrightarrow (CH_2O)_{organic} + H_2O + 2\,A \tag{4.1}$$

Where H_2A is a reducing agent (electron donor), $h\nu$ the photon energy that causes the reaction, CH_2O the produced biomass, and A a generic species that is produced by the reaction. In the case of oxygenic photosynthesis, the previous equation would be:

$$CO_2 + 2\,H_2O + h\nu \longrightarrow (CH_2O)_{organic} + H_2O + O_2 \tag{4.2}$$

Other forms of photosynthesis that do not involve molecular oxygen do exist. Such processes are within the broader category termed “anoxygenic photosynthesis” since they produce other compounds instead of O_2 as a waste product: an example could be methanogenic photosynthesis that was previously mentioned, which released high levels of CH_4 in the history of Earth’s atmosphere.

The absorption of light is facilitated by photosynthetic pigments.

Oxygenic photosynthesizers (plants, algae, cyanobacteria) use different types of chlorophyll (Chl *a*, *b*, *c*, *d*, and *f*)): the first two pigments are characterized by absorption peaks in the red and the blue (Chl *a* peaks at 430 and 632 nm, while Chl *b* peaks at 453 and 642 nm), with a less efficient absorption of the green part of the visible spectrum. For this reason, a healthy photosynthesizer typically appears to have a green color. The other pigments are common in some species of algae and cyanobacteria and absorb light at different wavelengths.

Anoxygenic photosynthesizers (green sulfur bacteria, purple bacteria, filamentous anoxygenic phototrophs, heliobacteria) use other kinds of pigments: the bacteriochlorophylls (Bchls), which have near-infrared wavelength absorption peaks.

In addition to the pigments, phototrophs use carotenoids to capture higher energy photons. A summary of the absorbance spectra of the main chlorophylls, bacteriochlorophylls, and carotenoids is shown in Figure 4.6.

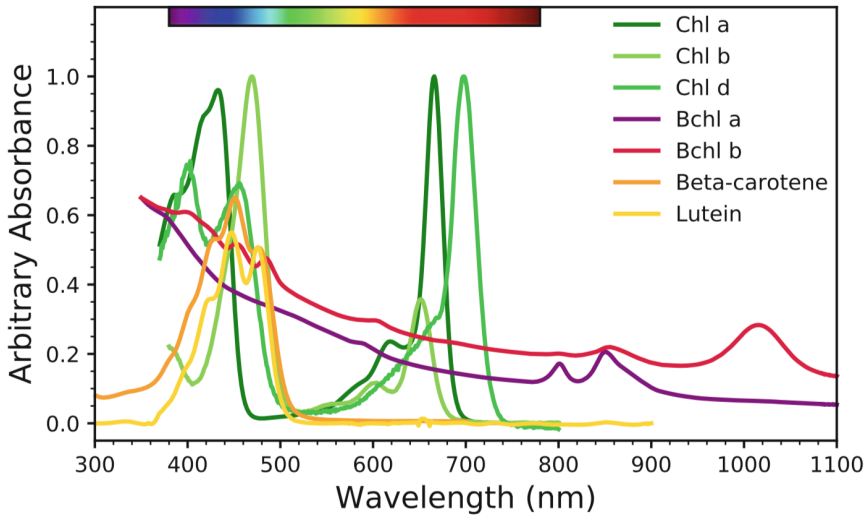


Fig. 4.6: Absorbance spectra in arbitrary units of some categories of chlorophylls, bacteriochlorophylls, and carotenoids [Schwieterman, 2018].

The Vegetation Red-Edge (VRE) on the Earth's emission spectrum is by far the most well-studied surface biosignature. It is the increase in reflectance of oxygenic photosynthesizers at about 700 nm, produced from the contrast between chlorophyll absorption (at 650-700 nm) and the scattering properties of the cell at near-infrared wavelengths (750-1100 nm) [ibid.]. It is a particularly strong feature, which produces sensible variation in a spectrum of a terrestrial planet (see e.g. Figure 4.7), increasing the near-infrared spectral albedo.

The strong feature of the VRE is used to study vegetation on Earth's surface as well. To characterize vegetated areas, the *normalized difference vegetation index* (or *NDVI*) was defined. It is calculated as:

$$NDVI = \frac{NIR - VIS}{NIR + VIS} \tag{4.3}$$

Where NIR is the reflectance (albedo) in the near-infrared range, and VIS is the reflectance in the visible band. This indicator covers the range [-1:1], with larger

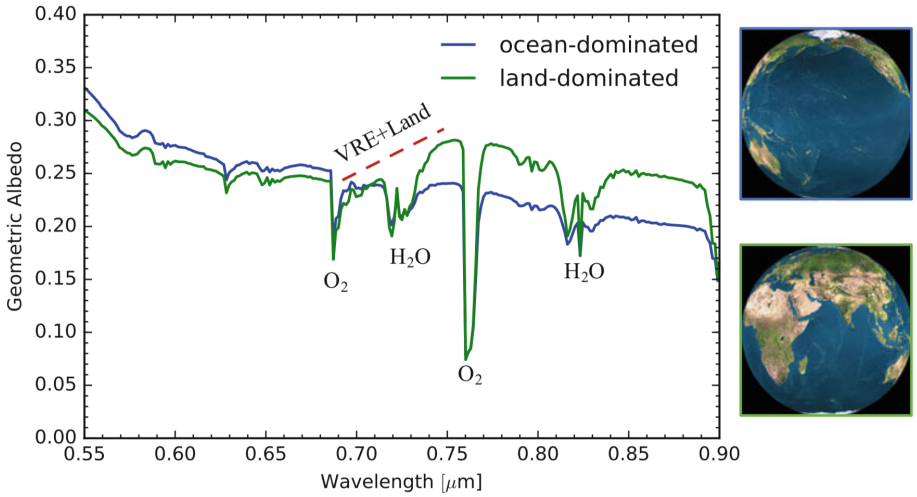


Fig. 4.7: Synthetic disk-averaged spectra of Earth at ocean (blue) and land-dominated (green) views. The reflectance increase of the latter case is caused by the higher reflectivity of the soil, as well as the VRE [Schwieterman, 2018].

values corresponding to vegetated areas, small positive values corresponding to bare soil, and negative values for snow and cloud decks.

The NDVI does not exactly distinguish the contribution of vegetation and of rocks or metals, so it could not be the unique tracer of life on a terrestrial exoplanet. It could nevertheless give qualitative results that would need to eventually be coupled with additional data. Its calculation requires a fairly low resolution, so NDVI could be used to filter at first order approximation the planets that could bear life.

However, it is important to point out that the VRE varied in time according to the evolution of life on our planet. Moreover, we should expect that different stellar temperature would determine a greater abundance of photons on one particular band (shifted to shorter and longer wavelengths), which could drive the evolution of the pigment absorption to the most efficient wavelength: therefore, the "edge" may be shifted accordingly. For these reasons, this feature may still be challenging to detect on exoplanetary spectra.

4.3.3 Temporal biosignatures

A temporal biosignature is a modulation of an observable quantity that can be linked to a living process (such as a seasonal change in strength or location of a biosignature).

On Earth, the abundance of atmospheric gases such as carbon dioxide, oxygen and methane experience a seasonal cycle: for the first two compounds, the life cycle of land vegetation is responsible of such cyclic variations, while methane abundance depends on the interactions with the OH ions, which in turn depend on the tropospheric water volume mixing ratio, strongly variable according to seasonal temperature. Such changes are hemisphere- and latitude-dependent, depending on the composition of the surface and on the average irradiation [Schwieterman, 2018].

The surface signatures may also vary with respect to insolation, water vapor abundance, or other observable parameters: the NDVI index does vary in time even when observing the Earth, due to the different color the leaves acquire during the seasonal changes – which is nothing more than a change in the absorption and reflection features due to desiccation and lack of chlorophyll pigments. Such features may be very hard to detect on an exoplanetary spectrum. However, other environments may cause slower, enhanced seasonal signatures.

On active environments - such as planets orbiting flaring M-dwarf stars - fluorescence features may appear. These signatures could be caused by the reprocessing of high-energy absorbed photons into less energetic emitted photons and could be a defense mechanism that photosynthesizers apply to survive bursts in harmful radiation. The detection of such signatures may, however, be complicated, due to the probable presence of gaseous absorption features and auroral activity.

5

Exo-MerCat: a catalog of exoplanets

> *That is a big question we all have: are we alone in the universe? And exoplanets confirm the suspicion that planets are not rare.*
>
> — **Neil deGrasse Tyson (1958-present)**
> (American astrophysicist, and science communicator)

In the following chapter, the development of the Merged Exoplanet Catalog (Exo-MerCat) code is summarized, as well as its performances and the treatment of some peculiar cases. Then, the subsample of all the well-characterized Super Earths currently known is retrieved using Exo-MerCat and described. This sample of targets will be then used as a basis for the radiative-convective atmospheric simulations in Chapter 7 and for the ozone photochemistry modeling in Chapter 8. Some interesting cases are currently being reproduced in the laboratory, in the framework of the "Atmosphere in a Test-Tube" project, described in Chapter 9.

5.1 Introduction

To date, the most used online catalogs are the *Exoplanets Encyclopaedia*[1] [Schneider et al., 2011], the *NASA Exoplanet Archive*[2] [Akeson et al., 2013], the *Open Exoplanet Catalogue*[3] [Rein, 2012] and *The Exoplanet Data Explorer*[4] [Wright et al., 2011].

Such databases (DBs) are built considering different criteria to include planets in the collection, whether statistical thresholds or physical properties of the suitable targets. They may provide a rich reference set connected to every single planet, allowing the retrieval of the original information and the method used by the various research

[1] http://exoplanet.eu/
[2] https://exoplanetarchive.ipac.caltech.edu/
[3] http://www.openexoplanetcatalog.com/
[4] http://exoplanets.org/

groups to obtain the data. If multiple parameter sets are available for each planet, some of the catalogs can provide a historical archive of the knowledge of the planet parameters as they evolve with time.

These archives should be also responsive to confirmation or rejection of candidates due to follow-up observations. As previously seen, the stellar activity may induce false positives in the detection of candidates; then, an analysis of the NIR radial velocity time series of some claimed planetary companions resulted in discharging the planetary hypothesis, confirming instead the activity nature of the signal, e.g. TW Hya and BD +20 1790b, by Figueira, Marmier, et al. (2010), Figueira, Pepe, et al. (2010), and Carleo et al. (2018). The removal of such targets from the different DBs depends on the frequency of the catalog update, which changes based on the research groups that maintain the catalogs.

Archives are particularly useful for identifying and examining the broader population of exoplanets. To perform robust population analyses, it is necessary to examine carefully the selection effects and biases in the creation of the catalog.

Bashi et al. (2018) analyzed from a statistical point of view the impact of the differences among the catalogs, concluding that, although statistical studies are unlikely to be significantly affected by the choice of the DB, it would be desirable to have one consistent catalog accepted by the general exoplanet community as a base for exoplanet statistics and comparison with theoretical predictions.

The Exoplanet Merged Catalog software (or `Exo-MerCat`) was created to respond to this need, and it will be described in this Chapter. In Section 5.2, the four online catalogs characteristics are described, then compared in Section 5.3. All the necessary operations to extract the `Exo-MerCat` catalog, the quality check procedures, the standardization, and the treatment of the critical cases are described in Section 5.4, while its performances are analyzed in Section 5.5. Some science cases are discussed in Section 5.6 and Section 5.7. Section 5.8 describes the catalog update procedure as a workflow and its deployment as a Virtual Observatory resource. Section 5.9 describes the dedicated Graphic User Interface, while Section 5.10 deals with the scheduled features to be added in the future.

Finally, a sample of the known Super Earths was retrieved from the `Exo-MerCat` catalog and studied in Section 5.11.

Box 5.1 - Naming of exoplanets

The issue concerning the naming of objects in the sky has always been present in human history. During ancient times Romans, Greeks, Arabs (among many others) gave names to constellations and the brightest stars, as well as deity names for the moving planets. Even today, Solar System planets are known with names based on Roman gods and goddesses (*Mars*, the god of war, *Venus*, the goddess of beauty, *Neptune*, the god of the sea...).

As more and more stars were discovered, due to the development of instrumentation that enhanced the capability of naked-eye to detect fainter objects, a naming convention was developed by Johannes Bayer (1572-1625), using Greek letters based on the relative brightness within a constellation (so that the brightest star of the constellation would be labeled as α).

This method was soon outdated since more and more faint stars were discovered. These were known with the name that reflected their position within the survey catalog (e.g. HD 20794 is the 20794th star in the *Draper Catalogue of Stellar Spectra* HD). Different surveys were then made for different wavelengths so that the very same star could have different names depending on the catalog number of each survey.

In modern times, many objects are known in the catalogs by their J2000 coordinates (see e.g. 2MASS catalog), but the upcoming surveys such as Gaia are expected to contain billions of objects, so there may be some ambiguity in this notation technique as well. To complicate things further, many known stars happen to be a system of two or more stars dynamically connected. Depending on the nature of the binary system, different notations were used: for eclipsing binaries, an uppercase Roman letter was appended to the system name (e.g. Sirius A and Sirius B); for spectroscopic binaries, lowercase Roman letters were appended instead. Combinations of suffixes were possible, as well.

Since the first historically most effective technique to discover exoplanets was radial velocities rather than transits, it was natural to use lowercase Roman suffixes to label the first batch of exoplanets: the central star was assumed to be the *a* component and the planetary system around that was labeled with *b*, *c*, *d* (...) letters in order of discovery.

The discovery of directly imaged planets, as well as transiting ones, would suggest in principle the use of the uppercase naming convention, but the well-established notation for exoplanets lead to use lowercase even in these cases.

With the advent of Kepler satellite, there was a need to label the large number of planetary candidates as well: the potential candidate host star received a *KOI* (Kepler Objects of Interest) number, and the planetary candidates were labeled with a two-decimal number in order of discovery. If confirmed, each candidate receives its official Kepler name.
When the exoplanet orbits a companion of a binary/multiple system, the uppercase and lowercase Roman letters should be appended to the system's name, to underline the dynamical hierarchy (see e.g. Figure 5.1), but this rule is still to be officially accepted by the IAU working group. In any case, gravitationally bound systems formed by known stars are currently being discovered: in such cases, a change in the name of the binary component (and any exoplanet orbiting that target), could lead to a higher chance of misunderstanding, rather than achieve the mission of clarifying the dynamical hierarchies.
In general, four rules should be followed when naming an exoplanet [Hessman, 2018]:

1. The formal name of an exoplanet is obtained appending the appropriate suffix to the formal name of the host star (i.e. 30 Ari B b);

2. If the leading capital letter of the host star is missing, an implicit "A" is assumed (this is the case for planets orbiting single stars i.e. 51 Peg *A* b = 51 Peg b);

3. If there is a more complex hierarchical relationship, the names of the higher-order system are placed in parentheses, followed by the suffix for a lower-order system (this is the case for circumbinary planets i.e. HIP 79098 (AB) b);

4. If a different name has not been set in literature, the astronomer should express all dynamically distinct (sub-)systems in order of dynamical relevance.

For these reasons, aliases and alternative nomenclature for both planets and stars can be an issue, so any catalog should provide an updated list of names for which one target can be known.

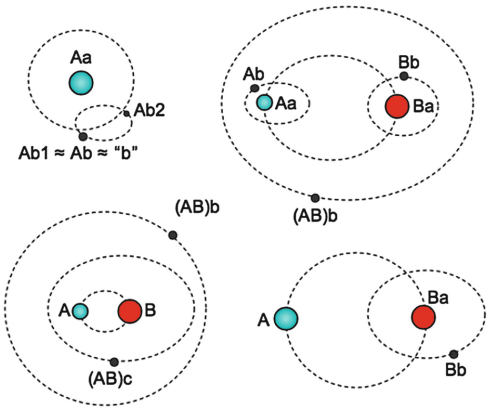


Fig. 5.1: Examples of different exoplanet name suffixes in single and binary systems [ibid.]. *Top left*: a single exoplanet around a single star (e.g., 51 Peg) plus a moon. *Top right*: double star, each with a planet (e.g. HD 41004), plus a circumbinary planet. *Bottom left*: two circumbinary planets (e.g., NN Ser). *Bottom right*: a planet around the secondary star in a binary (e.g., HD 178911).

5.2 Current state-of-art

Since the first discoveries, several online tables were built with the results of the different radial velocity and transit surveys. These catalogs, e.g. California and Carnegie Planet search table [Butler et al., 2006] and the Extrasolar Planets catalog which is the ancestor of *Exoplanets Encyclopaedia*, were workhorse catalogs in which first-hand data from observers were stored.

In 2011, with the creation of *Exoplanets Encyclopaedia* by Schneider et al. (2011), the list of discovered planets became a real catalog with planets discovered not only by radial velocity and transit surveys, but also by astrometry, direct imaging, microlensing, and timing, taking into account also unconfirmed or problematic planets. After that, other groups began to maintain general purpose exoplanet catalogs as well. In this section, the characteristics, the requirements, and criteria that characterize each of the main catalogs that are available online today are described.

5.2.1 NASA Exoplanet Archive

NASA Exoplanet Archive [Akeson et al., 2013] is a database and a toolset funded by NASA to support astronomers in the exoplanet community. It includes interactive tables containing properties of all published exoplanets and candidates, as well as stellar parameters, light curves, spectra and radial velocity measurements from both space missions and literature. Since data values are sorted by reference, it allows the user to compare stellar and planetary physical and orbital values published by different detection methods. Values for both new exoplanets and updated parameters are weekly updated by monitoring submissions via the journal pages and the LANL astro-ph server. Users are provided with an interactive table of confirmed planets, containing over 75 planetary and stellar parameters with uncertainties and limits. Furthermore, an overview page for each confirmed planetary system shows the default values (included in the interactive table) as well as any additional published values that the Archive has collected. In the case of multiple sets of values available in the literature for a given target, the NExScI (NASA Exoplanet Science Institute) scientists decide which reference to set as the default one, depending on the uncertainties and the completeness of the published data sets.

This archive includes and classifies all objects whose mass or minimum mass is less than 30 Jupiter masses and all those objects that have sufficient follow-up observations and validation, to avoid false positives. Furthermore, free-floating planets are excluded. All datasets show orbital/physical properties that appear in peer-reviewed publications.

The dataset of all confirmed planets can be easily downloaded either by browsing or through the corresponding API (application program interface). The table can be downloaded in multiple formats and both rows and columns can be filtered, selecting only the ones the user is interested in.

In this dataset, some KOI objects may however appear. Those are the ones which were at first published as candidates and then confirmed - and their name changed to a Kepler-NNN notation. When the confirmation of a target happens, the Archive does not update the name of the target itself, but the planet is included in the confirmed planets dataset. The updated name is stored in the alias column. KOI objects and candidate planets are stored in a separate table and are subject to further analysis: their status is then updated and, if necessary, the confirmed catalog is updated.

5.2.2 Exoplanet Orbit Database

The Exoplanet Orbit Database [Wright et al., 2011; Han et al., 2014] (hereafter ORG) includes robustly-detected orbital, transit, stellar, and other parameters of exoplanets reported in the peer-reviewed literature; also, Kepler Objects of Interests (KOIs) are present, provided they are not false positives. The website also hosts the Exoplanet Data Explorer (EDE), an interactive table with plotting tools for all planets included in the database. It retains all planets whose planet-star mass ratio is smaller than 0.023 (24 M_J for Solar-like stars). The archive aims to provide the highest quality orbital parameters of exoplanets rather than providing a complete presentation of every claimed target. The authors require that the period measurement has to be certain to at least 15%: this justifies the overall lower number of confirmed planets included in the catalog.

In this database M is often set equal to $M \sin i$ when the inclination is not known; if neither $M \sin i$ nor M are known, mass is calculated using the mass-radius relation explained in Han et al. (2014):

$$M_P = \begin{cases} M_{WM} & R < R_\oplus \\ (1-x)M_{WM} + xM_L & R_\oplus \leq R < 4\ R_\oplus \\ M_L & 4\ R_\oplus \leq R < 6\ R_\oplus \\ \min\left(M_M, M_J\right) & 6\ R_\oplus \leq R < b \\ M_J & R \geq b \end{cases} \tag{5.1}$$

Where:

- M_{WM} (in Earth masses), which has the following dependence on the radius (in Earth radii):

$$M_{WM} = \begin{cases} 0.615 \cdot (0.171 + R_P) \cdot {R_P}^3 \\ 2.69 \cdot {R_P}^{0.93} \end{cases} \tag{5.2}$$

- M_L (in Earth masses), whose dependence on the radius (in Earth radii) is given by:

$$M_L = {R_P}^{2.06} \tag{5.3}$$

- M_M (in Earth masses) follows the equation (R_P in Earth radii):

$$\log_{10}\left(\frac{M_M}{M_0}\right) = -w\left(\frac{b}{R_P} - 1\right)^{1/p} \tag{5.4}$$

The values of the parameters M_0, w, b, and p vary with respect to the semi-major axis (see Table 5.1):

Tab. 5.1: Considered values of the parameters in the retrieval of M_M.

Parameter	$a/AU \leq 0.1$	$a/AU > 0.1$
M_0 $[M_\oplus]$	1756.7	1308.7
b $[R_\oplus]$	11.684	11.858
w	1.646	1.635
p	2.489	2.849

In case of inconsistent host star names, the maintainers choose proper names, Bayer designations of Flamsteed numbers if available, or rather give ranked priority to GJ numbers, HD numbers, HD numbers, or HIP numbers. The planet name is then composed by the combination of stellar name and planet letter.

KOI objects are stored in the catalog and, when validated, their name is replaced by the official Kepler ID. The old KOI notation is stored in the `OTHERNAME` column. For most candidates, no coordinates are available, most likely because of strict disclosure policies concerning those targets.

5.2.3 Exoplanet Encyclopaedia

The Exoplanet Encyclopaedia [Schneider et al., 2011] (hereafter EU) includes planet detections already published or submitted to professional journals, or announced by professional astronomers in professional conferences: for this reason, this catalog contains more data than the previous one. Planets are sorted in four categories (Confirmed, Candidate, Retracted, and Controversial). In this database, every detected planet whose mass is lower than 60 Jupiter Masses up to 1 sigma uncertainties is included, since they could be disentangled by stars employing density-mass-radius relationships.

It considers also candidates without any estimate of the mass value but with a known radius: they are included in the Unconfirmed planets category. Exoplanet Encyclopaedia provides tools for histograms and graphs, as well as Virtual Observatory services.

5.2.4 Open Exoplanet Catalogue

The Open Exoplanet Catalogue [Rein, 2012] (hereafter OEC) is an archive based on small `.xml` files. Because of its structure, it can easily display planets orbiting a binary (or multiple) star system, and straightforwardly handle exomoons. The catalog is community-driven and open-source, downloadable from GitHub[5] and editable at will. The maintainer periodically checks the validity of all updates and only credible data are added. The database is also accessible on a website, the *Visual Exoplanet Catalogue*[6], and it is used by the iOS *Exoplanet* app [7].

Tab. 5.2: Selection criteria for all catalogs.

Catalog	Criteria
NASA Exoplanet Archive (NASA)	• Mass $< 30\ M_J$ • peer-reviewed publications
Exoplanet Orbit Database (ORG)	• $M_P/M_\star < 0.024$ • peer-reviewed publications • high quality datasets (orbital measurements)
Open Exoplanet Catalogue (OEC)	• Open-source
Exoplanet Encyclopaedia (EU)	• Mass or Msini $< 90\ M_J + 1$ sigma • published, submitted, announced references

5.3 Catalog Comparison

As reported in Bashi et al. (2018), the four catalogs are indeed similar, but not equal.

In Table 5.2 a summary of the selection criteria of each catalog is shown. This is the amount of information collected by reading the various documentation links provided by the websites. Often, the files lack update, or the main documentation is represented by the release paper itself, which may not consider all of the actual features of the various websites, nor any modification to the DBs themselves. For this reason, the actual boundaries of the various mass and radius parameters for all catalogs were carefully studied, shown in Table 5.3. It is clear that the OEC and EU archives do not follow their selection criteria since at least one target exceeds

[5] `https://github.com/`
[6] `http://exoplanet.hanno-rein.de/`
[7] `http://exoplanetapp.com/`

Tab. 5.3: Statistics for all catalogs. Update: September 19, 2019.

Query	NASA	ORG	OEC	EU
Free Floating?	No	No	Yes	No
Candidates?	No	Yes	Yes	Yes
Stellar Mass	89 %	92 %	97 %	94 %
Stellar Radius	90 %	97 %	97 %	94 %
Stellar Temperature	93 %	97 %	92 %	93 %
Stellar Metallicity	77 %	90 %	81 %	89 %
Stellar Distance	-	44 %	54 %	47 %
Stellar Age	58 %	-	-	18 %
U Magnitude	-	-	1 %	-
B Magnitude	-	-	35 %	-
V Magnitude	-	19 %	37 %	27 %
I Magnitude	-	-	10 %	41 %
J Magnitude	-	99 %	48 %	70 %
H Magnitude	-	99 %	48 %	69 %
K Magnitude	-	99 %	48 %	70 %
Spectral Type	-	-	34 %	23 %
Mass Minimum (M_J)	$6 \cdot 10^{-5}$	0	$8 \cdot 10^{-6}$	$2 \cdot 10^{-6}$
Mass Maximum (M_J)	30	22.62	263	280
Msini Minimum (M_J)	$9 \cdot 10^{-4}$	0	$8 \cdot 10^{-6}$	$5 \cdot 10^{-4}$
Msini Maximum (M_J)	55.59	22.62	263	63.3
Radius Minimum (R_J)	$3 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$2 \cdot 10^{-3}$	0
Radius Maximum (R_J)	6.9	9730	6	4332.12

the maximum value of (minimum) mass prescribed by the criteria themselves. Also, extremely high values of planetary radii are present (in ORG and EU archives in particular). These are due to human errors or discrepancies in the measurements.

In any case, this discrepancy between theoretical and actual boundary in the catalogs is probably linked to the ongoing discussion concerning the mass threshold for which the object is no longer a planet, but a brown dwarf (see Section 5.7).

For what concerns the amount of stellar data present in the various archives, shown in Table 5.3, the overall information about the host star's mass, radius, temperature, and metallicity is fairly complete for all catalogs. All archives but NASA have also magnitudes measurements, even though not all wavelength bands are uniformly filled by the various DBs. Distances, spectral types, and ages information are not uniformly provided by all catalogs. This is, in any case, not so important, since the main goal of such archives is to provide suitable information concerning exoplanets rather than their host stars. Such lack in stellar data can be overcome, however, by looking for more specific and trustworthy data into dedicated catalogs.

A more important analysis can be made on the available planetary measurements for all catalogs since, because of the different philosophy on the consistency of the datasets, the amount of data available for each target may be different and lead to substantially different records for a single planet.

As shown in Table 5.4, ORG and EU catalogs include a massive amount of candidates, and therefore appear to be much larger than the NASA and OEC archives. The number of confirmed planets is similar for NASA and EU catalogs, while OEC and ORG archives show fewer items, due either to selection criteria or lack of update.

In the ORG and EU catalogs, large importance is given to radius and period measurements, while the EU catalog alone seems to be the most complete for what concerns mass and minimum mass.

In all catalogs, simultaneous values of mass and minimum mass appear for the same target; on the other hand, the majority of the planets having at least one mass-related measurement and a non-null radius value, has a non-null period measurement as well. In facts, all transiting targets fall in this subset. By counting all unique host star names in the various archives, it is possible to estimate the number of planetary systems as well. This value is not the same for all catalogs, but it reflects the difference in the number of entries in each archive, due to the presence of candidates in some catalogs rather than others.

Tab. 5.4: Available measurements for various combinations of parameters in the four catalogs (see Bashi et al. (2018) for comparison). Update: September 19, 2019.

Query	NASA	ORG	OEC	EU
All planets	4055	5747	3793	6759
Confirmed	4055	3236	3674	4069
Candidates	0	2511	108	2681
False Positives	0	0	11	9
With radius	3142	4999	2917	5393
With mass	877	456	1120	1161
With minimum mass	769	29	273	1017
With period	3941	5733	3678	6517
With mass or minimum mass	1613	480	1393	2047
With mass and minimum mass	33	5	0	131
With mass and minimum mass and radius	10	2	0	43
With mass or minimum mass and radius	717	420	537	783
With mass or minimum mass, and radius and period	707	420	525	763
All systems	3027	4715	2795	5458

In Figure 5.2, the distribution of all planetary parameters for the four catalogs is shown. Different behaviors can be seen from panel to panel. Due to the presence of candidates, the EU and ORG catalogs show higher values of period, semi-major axis, and radius – which are indeed the first measurable parameters in transiting candidates. This often masks features in the distribution that appear in the other histograms (see e.g. the semi-major axis panel on the top right). The ORG catalog shows also many values of mass, due to the presence of theoretical values.

The EU is the only catalog that includes measurements of the interstellar objects 1I/'Oumuamua and C/2019 Q4 Borisov, as visible from the eccentricity values higher than 1 in the bottom right subplot of Figure 5.2.

No substantial difference is seen in the other graphs. A few uncommon values for the inclination were found in the OEC catalog, probably due to unreliable measurements or theoretical values.

The large difference in the number of mass measurements visible in the center-left panel of Figure 5.2 reflects also in the mass-radius plots in Figure 5.3. Even though overall, there seems to be a good agreement among the various measurements, there is indeed a fraction of planets not belonging to all catalogs. While the region around 1-10 Jupiter Masses and 1 Jupiter Radii seems to be more or less equally populated by the four DBs, the area around 1-10 Jupiter masses and 0.1 Jupiter radii is not uniformly covered. On the other hand, the ORG catalog provides a few targets at low masses and Jupiter-like radii, which are absent in the other DBs. Also, a clear

trend determined by all mass values retrieved from the theoretical M-R relationship is present in the ORG data. The masses indeed follow the trend determined by observed values, except for the strong vertical at 1 M_J showing that for radii larger than the Jupiter radius, the relation is out of its range of validity.

From these plots, it is clear that any attempt to fully merge the four catalogs is impossible. It was therefore necessary to provide the four datasets with a greater uniformity, which may lead to a more effective association among the various targets and a higher statistical significance on the measurements, creating a catalog that would cross-match at best the four archives.

5.4 Exo-MerCat: development

Exo-MerCat is a program written in Python 3.6 that merges the exoplanet catalogs described in the previous sections. To merge the exoplanets catalogs some preliminary operations are necessary, among which the standardization of the four data sets to be able to compare each entry of a catalog with those of the others. This task is very difficult to do automatically and it was necessary to choose in a very accurate way the software tools more suitable for the purposes.

One of the biggest challenges was to hunt for the aliases and check the coordinates of host stars. Most of the aliases problems derived by discrepancies in the notation of both stars and planets in the different catalogs. The different phases of the Exo-MerCat workflow are reported schematically in Figure 5.4 and discussed in details in the following sections.

A Graphical User Interface is provided to all users and it allows the filtering of the catalog, as well as the automatic plotting of some interesting plots.

5.4.1 Libraries and Tools

To be operative, the software needs a few Python packages in addition to the default ones. The package pandas[8] allows flexibly manipulating large datasets, by storing data in Series (1-D arrays) or DataFrames (2-D arrays) structures. It also allows data grouping and merging, as well as quick operations between rows and columns, and hierarchical indexing.

[8] https://pandas.pydata.org/

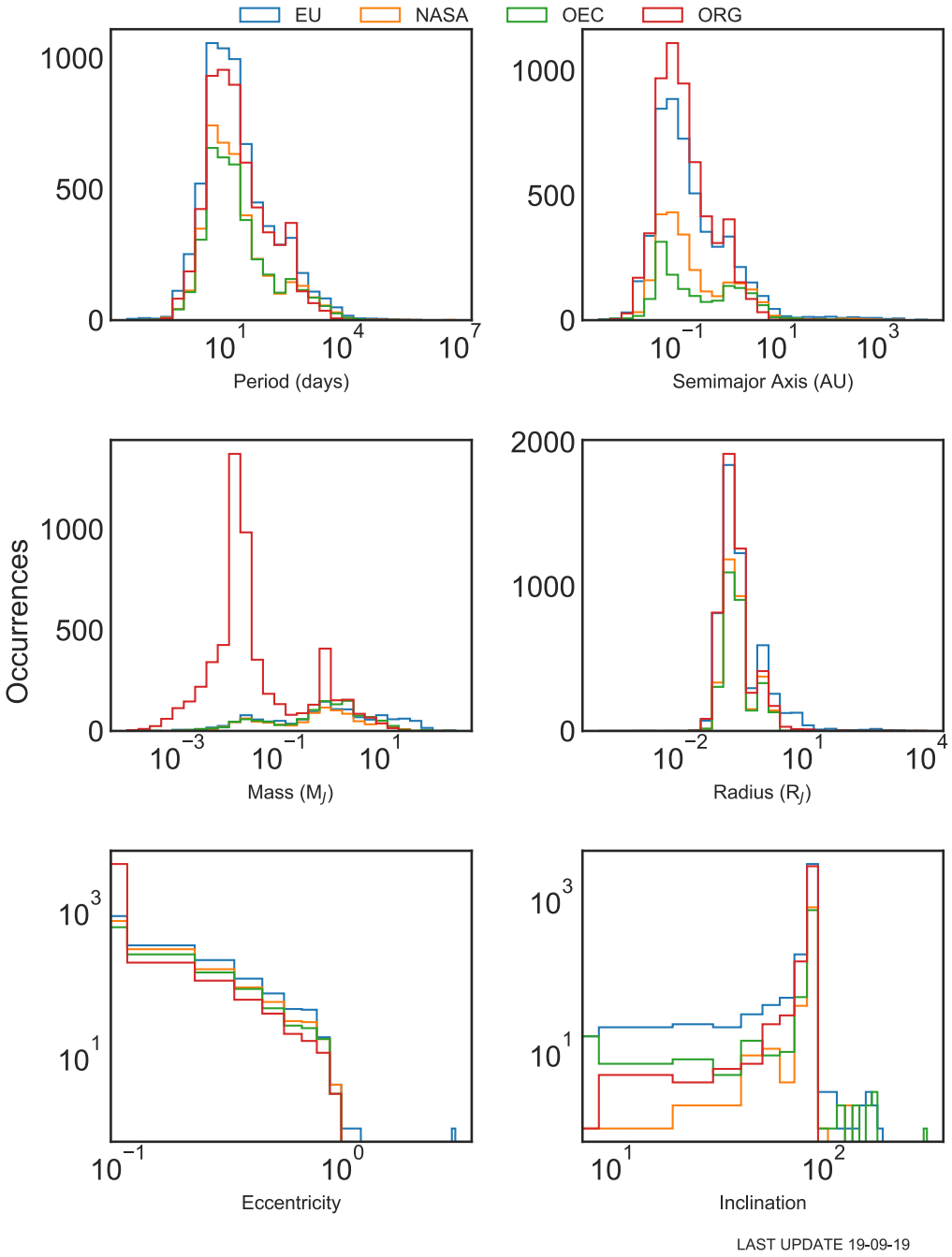


Fig. 5.2: Period, semi-major axis, mass, radius, eccentricity, inclination histograms for each of the input catalogs. As shown in the legend, the blue histograms refer to the Exoplanet Encyclopaedia, the orange histograms to the NASA Exoplanet Archive, the green histograms to the Open Exoplanet Catalogue, and red ones to the Exoplanet Orbit Database.

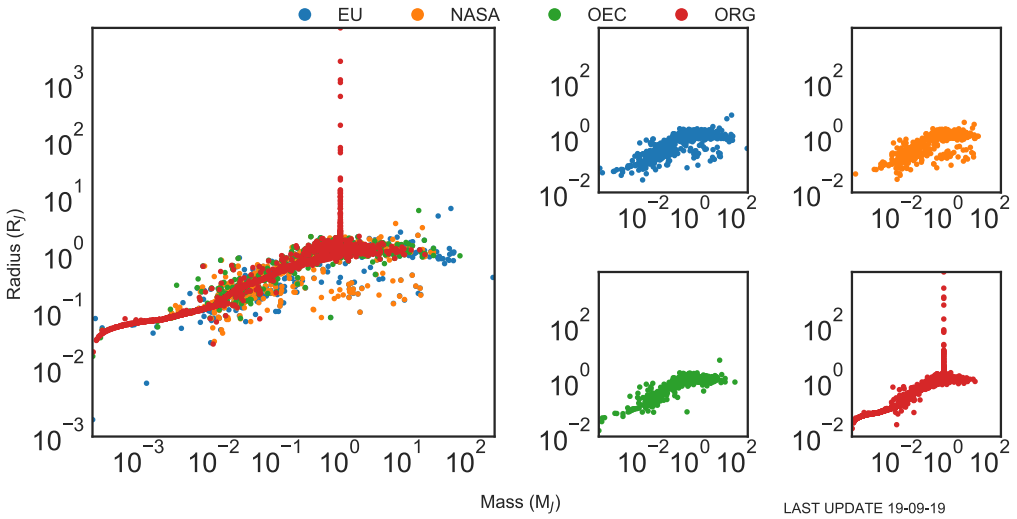


Fig. 5.3: Mass-Radius plot for the raw catalogs. As shown in the legend, blue dots refer to the Exoplanet Encyclopaedia, the orange dots to the NASA Exoplanet Archive, the green dots to the Open Exoplanet Catalogue, and the red ones to the Exoplanet Orbit Database.

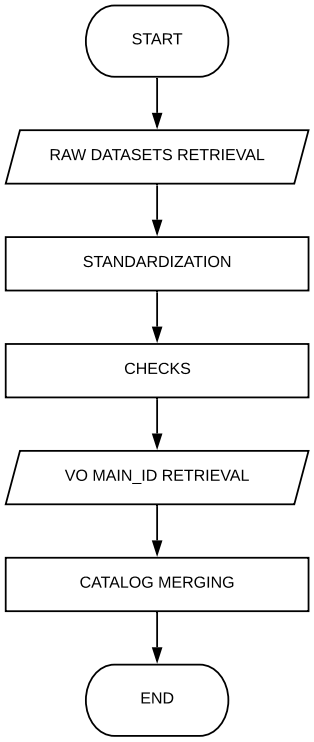


Fig. 5.4: Workflow of the main script.

The package `astropy`[9] is already included in the `Anaconda Python Distribution`, a community-developed core Python package for Astronomy [Astropy Collaboration et al., 2013; Price-Whelan et al., 2018]. In our case, this package was used to treat astronomical coordinates and to properly convert the various parameters. Also, we used `astroquery`, an `astropy` affiliated package, to access and download the original ORG catalog.

For what concerns the Open Exoplanet Catalogue, an `.xml` reader package is needed. This is by default available in Python, while the retrieval code (which converts an `.xml` file to a `pandas` Series) was adapted from the default ones, available at the original website[10].

All the other VO queries were performed using `pyvo`[11], an `astropy` affiliated package, which implements general methods for discovery and access of astronomical data available from archives complying with the standard protocols defined by the International Virtual Observatory Alliance (IVOA).

The software makes extensive use of the Table Access Protocol [TAP, Dowler et al., 2010], an IVOA standard designed to provide access to relational table sets specifically annotated for astrophysical usage. The queries posted to TAP compliant services can be specified using the Astronomical Data Query Language [ADQL, Osuna et al., 2008, another IVOA standard]. The SQL-like queries built in ADQL and posted to TAP services allow catalog filtering using lists of astronomical targets, as well as spatial cross-matching functions among various catalogs and general custom manipulation of the content of each catalog.

5.4.2 Initial datasets retrieval

There is no uniform retrieval of the raw datasets since not all catalogs allow the same service to download the source file. Therefore, `Exo-MerCat` starts its run by downloading the various files using different methods.

For the Exoplanet Orbit Database, the Exoplanet Encyclopaedia, and the NASA Exoplanet Archive, a simple call to command-line instruction `wget` allows downloading a comma-separated value file. The code selects specific columns when making the `wget` call, to reduce the amount of downloaded data and to be sure that all necessary columns are correctly considered.

[9] `http://www.astropy.org/`
[10] `https://github.com/hannorein/open_exoplanet_catalogue`
[11] `https://pyvo.readthedocs.io/`

The Open Exoplanet Catalogue is on the other hand composed by a set of separate `.xml` files, which can be downloaded from the GitHub repository. In this case, the code needs to download the latest updates of the repository itself, and then to convert the `.xml` files into a unique `.csv` file.

The various input datasets are stored in four `pandas DataFrame` objects.

This operation requires about two minutes on a standard MacBook Pro (Mid-2015), but it relies on the speed of the Internet connection. It should not require, however, more than five minutes for slower connections.

5.4.3 Standardization

The raw datasets present themselves as very different, so any sort of merging at this stage would be impossible. For this reason, the DataFrame objects need to be carefully standardized to the desired, common output.

For every single catalog, a dedicated function within the software can process the following operations:

- First of all, only part of the available columns were considered: at this first step, all information about the host stars was discarded since it could be easily retrieved by connecting to the most important stellar catalogs. However, precise measurements of stellar parameters are not always present in exoplanets' paper references, so columns in the raw databases concerning those are often far from completion.
- Selected columns were then renamed to ease the subsequent merging. For each parameter `X` (mass, minimum mass, radius, period, semi-major axis, eccentricity, inclination), the code creates a new column called `X-REF` to store the link to the bibliography in which the measurement first appeared. This was not possible for the "EU" and "OEC" catalog, which do not provide information concerning the reference in the first place; the software keeps track of those by filling the `X-REF` columns with a string displaying either EU or OEC respectively.
- A column dedicated to the aliases of a single target is created, displaying them as a comma-separated string. The list of aliases is seldom complete, since most reference papers report up to two aliases per target, despite being it known with other identifiers as well.
- All double and unnecessary white spaces were removed.

- All target names were checked and standardized: all Kepler-like entries were labeled as "Kepler-X" (with X as a 1-4 ciphers integer with no leading zeros), the Greek letters for some stars were displayed as three-character strings (α as alf, β as bet...). The host constellations were displayed as three-character strings too. A dictionary of all abbreviations for constellations was retrieved from the IAU official list[12] and use to make coherent replacements. This step was necessary to allow the merging of stars belonging to a known constellation, but stored in the various databases using a slightly different notation. A suitable example would be the host star Algieba, Gamma Leonis. The planet orbiting that star is labeled "gamma 1 Leo b" in the Exoplanet Encyclopaedia, "gamma Leo A b" in the Exoplanet Orbit DataFrame, "gam 1 Leo b" in the NASA Exoplanet Archive, and "Gamma Leonis b" in the Open Exoplanet Catalogue. For a human being, it could be easy to assume that the four entries represent the same target, but software programs that could only compare them as strings would recognize them as undoubtedly different.

- On a regular basis, a planet is labeled as the name of its host star, plus a letter (b to h) to rank planets within the same stellar system based on the year of discovery. To retrieve the host star name from the catalog, it is necessary to strip the last letter from each target name. On the other hand, unconfirmed Kepler Objects of Interest have a different notation compared to confirmed exoplanets: they are usually displayed as KOI-NNNN.DD where KOI-NNNN represents the host star, while the last two digits DD unambiguously identify each target within the same system (where 01 is the first discovered planet, 02 the second one, etc.). In this case, the last three characters ".DD" were removed from the planet name to retrieve its host star name.

  Another exception to handle was represented by the very first exoplanets discovered orbiting the pulsar PSR 1257+12 [Wolszczan et al., 1992]. This system was originally labeled as PSR 1257+12 A, PSR 1257+12 B, and PSR 1257+12 C, but since then a massive variation in common notation happened, so it was necessary to change those names to a more standardized PSR 1257+12 b, PSR 1257+12 c, and PSR 1257+12 d, so that the planets could be labeled uniformly throughout the final catalog.

- For what concerns the labeling of planets orbiting binary systems, the four catalogs behave differently. NASA and EU catalogs provide the letter labeling the host binary companion as a substring in the string displaying the name of the planet; all planets on a P-type orbit (i.e. circumbinary planets) show the

[12] `https://www.iau.org/public/themes/constellations/`

substring AB in the name string. The ORG catalog provides a BINARY column that indicates whether the host star is supposed to be part of a binary/multiple system, but it does not provide information concerning the orbit type (whether S- or P-type). This information is on the contrary provided by the OEC catalog within the binaryflag (which is 2 if the planet is on an S-type orbit, 1 if it is on a P-type orbit, 0 otherwise), but little information is given with respect to S-type planets, since it is not known which stellar companion is the actual host star. However, the OEC and ORG catalogs provide as well the substring labeling the binary star (or both of them if circumbinary), but that is often not coherent with the flags. In particular, the ORG catalog provides the letter of the binary companion that hosts a planet only 15 times throughout the whole catalog, but the binary flag indicates that more than 700 planets orbit a binary star (corresponding to nearly 500 unique host star names). On the other hand, the OEC catalog provides about 200 non-null binaryflag values, but less than half of the sample displays the binary substring within the name of the planet.

The ideal setup for the four DataFrames, to provide a correct match in the following functions, would be to have both information concerning the orbit of the planet, and the binary system that hosts him. This was not possible with the data provided by the catalogs. The software collects as much information as it can from the original datasets by stripping the label letter(s) from the host string, when available and storing this substring in a dedicated column binary, whose value is left empty if no information is provided, assuming that in that case the host star is a single one. The flag values for these two catalogs were not taken into account, due to the incompleteness of the information provided. The only exception was the circumbinary sample in the OEC catalog (binaryflag=2), for which the binary value was forced to be "AB".

- The letter labeling each planet is stored in a dedicated column that will allow hierarchical indexing of the DataFrame. For Kepler Objects of Interest, the software converts the ciphers DD in letters (01 as b, 02 as c...), to keep a uniform notation among confirmed and unconfirmed objects.

- In the input catalogs, calculated values for mass and radius can be identified by a flag in dedicated columns. These values could be either retrieved by theoretical mass/radius relations, or by assuming a typical value for the unknown inclination. We chose to set to undefined all values that were calculated or theoretical, thus retaining only actual measured values for these parameters.

- Finally, the names of the retrieval methods were standardized, since the various catalogs adopted different notations.

This step required about 80 seconds for each one of the four original DataFrames on a standard MacBook Pro (Mid-2015). The higher amount of data contained in the ORG and EU DataFrames, due to the presence of candidates, corresponds to a 20% higher computational time, which is however negligible from the total running time point of view.

5.4.4 KOI Objects Status

It may be possible that some additional candidates or false positives are included in the current archives, due to lack of updates or human error. A check on the status of each target (especially for Kepler ones, since they represent the majority of known exoplanets) is due.

NASA Exoplanet Archive and Mikulski Archive for Space Telescopes[13] (MAST) provide an updated table of all Kepler Objects of Interest both belonging to Kepler and K2 missions, periodically updated to show the status of each target, whether confirmed by follow-up observations, still candidate or retracted as false positive. For all confirmed planets, a Kepler-like identifier is given to replace the original KOI- or KIC/EPIC-like notation.

The software downloads the table and cross-matches it with the four DataFrames, updating any KOI name with the official Kepler identifier, if present. Then, it stores the various information concerning the status of each target in a column named `status`, filling it with `CONFIRMED`, `CANDIDATE`, or `FALSE POSITIVE` strings.

In the best possible scenario, we should not see variations in the number of confirmed, candidates, and false positives as reported in Table 5.3, thus meaning that the original status of every target in each catalog is correctly updated. This was unfortunately not the case: in the NASA catalog 15 candidates and 1 false positive appeared; in the ORG catalog 1 confirmed planet was actually still a candidate; in the OEC catalog 10 confirmed planets were still candidates or false positives; finally, in the EU catalog nearly 500 false positives were contained in the candidate sample.

[13] `https://archive.stsci.edu/index.html`

This could have been caused by delays in the update of the single catalogs, or either misinterpretation of reference papers. In any case, the NASA Exoplanet Archive appears to be the most updated catalog from this point of view.

In the case where no coordinates for the Kepler candidates are available (i.e. for the ORG catalog), the crossmatch among the exoplanetary DataFrames and the MAST KOI table is useful in retrieving the missing information. The function successfully retrieved all coordinates of the ORG candidates, about 2500 at the time of writing.

This routine will be modified as soon as more TESS candidates will be confirmed by follow-up observations. Provided that a KOI-like table is available for TOI (TESS Objects of Interest) objects, a similar feature will help to treat such targets.

5.4.5 Alias and Coordinates Check

By trying to merge the four DataFrames, it should be likely to find a large number of targets in two or more catalogs. It may be possible, however, that some targets are labeled differently despite being the very same objects, since the various catalog maintainers may have chosen a different alias to represent the host stars, and thus the orbiting planets. In this way, a code that performs a match among strings would not be effective, not considering all the occurrences of a given planet (see Section 5.5).

For this reason, this function stores all the available host star default names and aliases from the four original DataFrames and attempts to find if (and when) the same host star is saved with an alias. In order to be more coherent and to ease the way for the operations to come, it would be preferable to retain a host star name which can be easily recognized by SIMBAD[14] [Wenger et al., 2000]; furthermore, a more exhaustive list of identifiers for which each stellar target is known would undoubtedly lead to more effective results in this subroutine and the following ones.

All host star names retrieved from the four DataFrames (dropping all duplicated strings) are therefore queried to SIMBAD through a VO ConeSearch [Plante et al., 2008] query. For each queried string, the ConeSearch returns a string listing all available aliases, as well as a single string labeling the main identifier for which each target is known in the Archive. At the time of writing, starting from a list of about 6200 host star strings (which may contain duplicates of the same physical target labeled differently), only 550 queries were unsuccessful. This was probably

[14] `http://simbad.u-strasbg.fr/simbad/`

caused by an unconventional notation displayed in this target, mainly the usage of unknown aliases.

SIMBAD can recognize many of the aliases under which a star is known and all of them points to the same target, identified by a unique name. We, therefore, expect that the result of a ConeSearch of the same target queried under different aliases should return the very same results (i.e. main identifier and list of all known aliases). This feature allowed us to further identify duplicates within the list of host star targets.

When such issues happen, the function chooses a common identifier for each host star and overwrites the host star name in the original DataFrames when necessary. At the time of writing, 330 duplicates within the total list of host star names were found.

Many of these were KOI-like objects: as a matter of fact, for the Kepler systems in which one or more planets are confirmed (and thus renamed in a Kepler-like notation) while others are still candidates, the host star is by construction named differently. This function, therefore, helps to correct and to uniform such cases, too.

At this point, the identifier list retrieved for each successful target by SIMBAD was completed if necessary with the available aliases for each star available in the catalogs.

For the host stars for which the VO ConeSearch was unsuccessful, the code performs a less effective yet useful check. For all available identifiers of a target, whether belonging to SIMBAD or saved by the original catalogs, the function queries the Host column of the DataFrame for other occurrences. If an alias in that column is found (i.e. that entry is the same host star but labeled in an alternative way), the host star name is uniformed. At the time of writing, about 19 further corrections were made.

Subsequently, the software checks for the consistency of coordinates, to avoid mismatches when merging the catalogs. Indeed, it may be possible to have coordinate values which are not correctly updated with new measurements, or either sign errors may occur.

On the other hand, J2000 coordinate differences can be very important in correctly identifying any planet orbiting a binary, especially for those cases in which no label was provided by default. In particular, the same binary companion can appear with the same `host` name in more than one catalog, but in some cases the `binary` string

would be null (i.e. no information concerning the fact that the host star was part of a binary/multiple system was given by one or more catalogs): in this case, a code which compares strings would interpret the various entries as different targets, even though the actual planet would be the same. Whenever possible, then, this check identifies all targets having different values of the `binary` string, for each system in the catalog. The software creates subgroups depending on the value of `binary` (typically "A", "B", "AB", or null) and checks if each pair of coordinates of the null subsample can match any of the coordinates of the other subgroups. In this way, most of the originally null `binary` values are fulfilled with the correct value, thus allowing the following functions to perform correct operations among targets.

Sometimes, the difference in coordinates can be high enough (greater than 0.005 degrees) to forbid an automatic match between the various entries. In this case, the flag `MismatchFlagHost` (see Table 5.6 for further information) was set as 1 for all the involved targets to warn the user about this issue.

Furthermore, it may happen that within the same system, S-type and P-type orbiting planets existed simultaneously, depending on the original value of `binary`, which often belongs to different catalogs. This is a somewhat difficult problem for what concerns the dynamics of the system that needs to be studied carefully. In such cases, it is highly probable that the different entries are in truth the same planet, but two or more catalogs were not in agreement for what concerns the orbital type. In this case, the software fills the `MismatchFlagHost` flag value as 2.

At the time of writing, about 75 `binary` corrections were successfully made. Two planetary systems had the `MismatchFlagHost` flag set to 1 (HD 106906, Kepler-420), and one to 2 (HD 131399). These targets will be analyzed in Section 5.6.

For all planets not showing issues with the `binary` flag, the code performs a simple check on the coordinates, to find out if all entries for a single target are consistent with one another.

The code retrieves the mode (i.e. the most common value) of the coordinates for each host star entry and checks if there are inconsistent values (whose difference from the mode is greater than 0.005 degrees). In that case, the wrong value is replaced by the mode.

If no mode is found (i.e. there is no most common value), no replacement is made: any inconsistency will be solved at a later point in the process. The software sends a warning to the user, reporting that the four catalogs are not in agreement for what concerns either right ascension, declination, or both. This automated check helped

us find errors within the original catalogs and warn the catalog maintainers about certain issues.

At the time of writing the code successfully found 200 inconsistent coordinates, most of which (about 110) replaced with the mode value. About two-thirds of these errors concerned the declination value. In some cases, especially for the lower values of declination (less than 1 degree), a plus/minus sign difference appeared among the various datasets. This is caused by the inner uncertainties of such coordinates. Gaia could improve accuracy by retrieving more precise coordinates and proper motions.

This step required about 2 minutes on a standard MacBook Pro (Mid-2015).

5.4.6 Main identifier retrieval

Despite all efforts made up to this point by the previous functions, in some cases, the host identifier for the same target could be different in the four catalogs, so any merging by host star name would still be inefficient. Also, it could be useful to provide a link to the most important stellar catalogs for future analysis on the stellar-planet systems.

To accomplish this, the code performs a series of ADQL (Astronomical Data Query Language) queries to multiple TAP services such as SIMBAD and VizieR[15] [Ochsenbein et al., 2000], in order to collect all useful data (in our case, identifier and coordinates) from the most important catalogs such as Kepler [Kepler Mission Team, 2009] and K2/EPIC Input Catalogs [Huber et al., 2017], as well as Gaia DR2 [Gaia Collaboration, Prusti, et al., 2016; Gaia Collaboration, Brown, et al., 2018].

First of all, the four DataFrames are concatenated to create a global DataFrame with more than 20000 entries belonging to the four catalogs. Indeed, the majority of such entries should be duplicate datasets belonging to different catalogs.

The code loads the SIMBAD TAP service and queries it via `pyvo`. The first query looks for an exact match between the name of the host star as assigned in the global catalog, and the known identifiers in the SIMBAD Archive. All successful results from the query are stored in the corresponding `main_id` and official coordinates `ra_off`, `dec_off` columns for all occurrences of each host star.

For all host star names for which the exact string match was unsuccessful, a new query is made by considering all the known aliases contained in the `alias` column.

[15] `http://vizier.u-strasbg.fr/viz-bin/VizieR`

These queries are indeed effective in finding the appropriate main identifier for most of the targets: the number of missing main identifiers at this stage is reduced to about 400 entries, from an original number of more than 20300 elements (the concatenation of all four DataFrames). At this stage, all main identifiers found in the previous queries are unequivocally linked to the original denomination, being based on an exact match of strings.

For all unsuccessful targets, another query to SIMBAD is then made by cross-matching the coordinates of each target with all sources within the online Archive. These sources are considered to be potential matches with the corresponding target if their coordinates fall inside a circle of radius 0.0005 degrees from the coordinates provided by the considered exoplanet catalog. This value was chosen to account for the average precision of the right ascension and declination values which are available from the input catalogs.

In general, it may be possible that multiple sources are found within the circle, so the software calculates the angular separation from the original coordinates with `astropy`. Only the source with the shortest angular separation from the center is stored in the main identifier and default coordinates columns.

In this case, all successful matches have very small angular separation and from a quick view it was possible to witness the fact that all identifiers linked correctly and the main identifier string was indeed similar to the original one, but the notation of the latter was unconventional and was not recognized by the previous query.

These steps sort out the vast majority of targets, leaving 160 entries in the general catalog still without a main identifier at the time of writing.

Switching the TAP service from SIMBAD to VizieR, the code is able to query the other catalogs by coordinates, in a similar way it previously did. The code queries the Kepler Input Catalog and the K2/EPIC Input Catalog since most of the known candidates are included in the Kepler surveys. In this way, only about 130 entries have still a missing identifier.

At this point, the software connects to the ARI-GAIA[16] TAP service to query Gaia DR2 Archive. This proves to be effective, leaving 89 targets with no identifier at tolerance 0.0005 degrees.

Since there are still targets without their main identifier, the code increases the tolerance of the query (i.e. the radius of the circle around the original coordinates)

[16] `http://gaia.ari.uni-heidelberg.de/`

and tries the same queries again until all remaining items acquire the corresponding main identifier.

At tolerance 0.0025 degrees the queries to SIMBAD and K2/EPIC are still effective, leaving only 24 items with no identifier. At the same tolerance, ARI-GAIA finds 13 of them. Any further increase of tolerance seems to be effective only for Gaia DR2, which finds all other targets by a maximum tolerance of 0.0175 degrees.

At the time of writing, the current amount of targets in all source catalogs have correctly been taken care of. It is however impossible to exclude the need for some adjustment in the cross-match radius in the future, depending on future discoveries and their treatment in the original databases.

This step required about 4 minutes on a standard MacBook Pro (Mid-2015).

At this point, since the main identifier column allows to easily group all occurrences, we performed a check to find multiple entries of the same planet within the same source catalog. This was, unfortunately, the case for a few targets, mainly for the EU (73 duplicated entries) ORG (63 duplicated entries) and OEC (17 duplicated entries) catalogs. These planets were included in the catalog with both their provisional candidate name and with their confirmed one. Such issues are automatically identified by the software, and stored in a log file that could be sent to the catalog maintainers. The NASA Archive has no duplicated entries at all.

5.4.7 Catalog retrieval

The cumulative catalog can be hierarchically indexed by the tuple `main_id`, `binary` (if present) and `letter`. This is supposed to be more effective after the previous treatment on the homogeneity of notations.

We expect to have up to four entries for each planet and the code has to collapse them to one single entry, based on the precision of the measurement.

For each parameter (mass, minimum mass, radius, period, semi-major axis, inclination, eccentricity) the code calculates the relative error X_{rel}, defined as:

$$X_{rel} = \frac{max(err^X_{min}, err^X_{max})}{X} \tag{5.5}$$

Where X is the value of the considered parameter, while err^X_{min} and err^X_{max} the absolute values of the lower and upper error.

For every single parameter, the code selects the dataset (value and errors) with the smallest relative error, and it stores the reference paper in which the chosen dataset first appeared.

Until now, the column describing the planet name was left unchanged from its raw value, while the host star names were rearranged and standardized. It could be then possible that multiple planet names appear as default names for the same tuple `(main_id, binary, letter)`. The code selects the string that contains the commonly known name, by privileging Kepler, WASP, Gliese, K2, HD, Hipparcos, CoRoT identifiers when available.

Aliases for the same target are stored in a cumulative list of strings.

At this point, each group of duplicates was collapsed into one single row. The fingerprint of the original group remains within the column `catalog`, which shows a list of all catalogs in which the item was found. This list could be then composed of four elements, if the planet was present in all input catalogs, rather than three, two or a single element.

The status of the single target is stored in a string which retains information concerning the original label from the source catalogs. The strings follow the pattern `AXDXEXCX` where A represents the NASA **A**rchive, D the Exoplanet Orbit **D**atabase, E the Exoplanet **E**ncyclopaedia, and C the Open Exoplanet **C**atalog, while X is an integer from 0 to 3, where 0 means that the target is not present in the catalog (represented by the previous letter), 1 if it is labeled as a false positive, 2 if it is a candidate and 3 if it is confirmed.

To ease further analysis, the code also provides the most probable status of the planet: if the previous string is composed by 0 or 3 only (the planet may be not present in every catalog, but when present it is labeled as confirmed); if at least a 2 is present, the planet is labeled as candidate; if at least a 1 is present, the target is a false positive.

Once the final row for each target is ready to be concatenated to the rest of the output catalog, the code stores the values of the best mass, by choosing between mass and minimum mass the measurement with the smallest relative error. The fingerprint of the original measurement is stored in the column `mass_prov`: it could, therefore, contain `mass` strings, if the most precise measurement was the mass, or `msini` strings otherwise. This will be useful to plot any value of the mass, eventually choosing a different marker for minimum mass or mass of the planet (as described in Section 5.9).

5.5 Performance

In this section, we will try to assess the performance of the code, and in particular, we will be focused on the improvements on the final catalog that the various functions allow.

First of all, we merged the initial databases as they were downloaded from their sources. Of course, in order to be able to compare and sort the four datasets, they needed to be standardized, so that all interesting columns could have the same string as a header. No treatment whatsoever with filters, aliases, main identifiers, and coordinates was made.

The function which retrieves the merged catalog was then executed, interpreting the `Host` column as the main identifier. In the remainder of the section, we will refer to this particular run as a "Simple" among the catalogs.

We then performed a full run of the `Exo-MerCat` software, executing all of the aforementioned functions. We will refer to this run as an "EMC" match among the catalogs.

Therefore, we performed twice the same merging of the four catalogs: while the first time the original state of the four archives was preserved, the second run showed the full potential of the software. By comparing the results, we are able to analyze the improvements that this code allows to achieve.

The results are shown in Table 5.5. The M-R plot with all quadruple, triple, double and single matches is shown in Figures 5.5 and 5.6.

First of all, the Simple run causes the whole catalog (confirmed planets, candidates, and false positives) to be composed of about 10200 elements. On the other hand, for the EMC run the final amount of planets is less than 7500 (nearly 4200 confirmed targets, 2800 candidates, and 500 false positives). This means that the software indeed is effective in finding a large number of duplicates.

For the Simple run, nearly half of the sample is present in the Single Match subgroup. This does not happen for the EMC run, which on the other hand shows nearly half of the same amount of candidates in the Double Matches subgroup: this means that nearly 3000 duplicate/multiple candidates were indeed present in the sample, but were not recognized in the Simple run because of the difference in their notation – and were therefore categorized as single occurrences. Most of the items of this subgroup belonged either to the EU or the ORG archive (which at present are the only two catalogs that provide a substantial amount of candidates by default), and

Tab. 5.5: Results of the merged catalog with and without corrections. Update: September, 19, 2019.

Sample	EMC RUN	SIMPLE RUN
All Planets	7481	10185
All Confirmed/Candidate Planets	7002	10166
Quadruple Matches	3194	2929
CONFIRMED	3131	2901
CANDIDATE	62	28
FALSE POSITIVE	1	0
Triple Matches	362	437
CONFIRMED	299	412
CANDIDATE	55	24
FALSE POSITIVE	4	1
Double Matches	2414	581
CONFIRMED	446	525
CANDIDATE	1962	53
FALSE POSITIVE	6	3
Single Match	1500	6238
CONFIRMED	341	1041
CANDIDATE	691	5182
FALSE POSITIVE	468	15

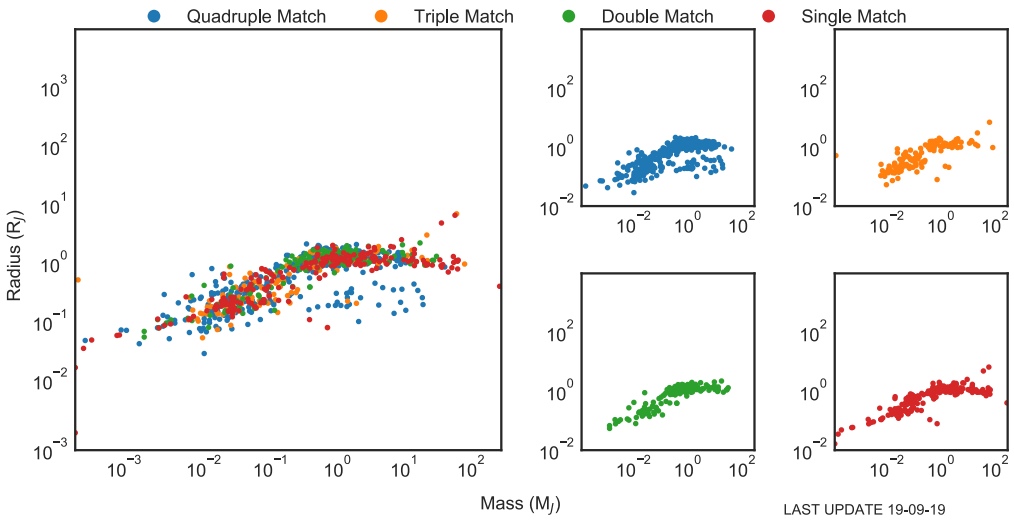


Fig. 5.5: Mass-Radius plot for the Simple Run. As shown in the legend, blue dots refer to all quadruple matches (i.e. planets found in all databases), the orange dots to the triple matches (i.e. planets found in three over four databases), the green dots to the double matches (i.e. planets found in two over four databases), and the red ones to the single planets (i.e. planets found only in one database).

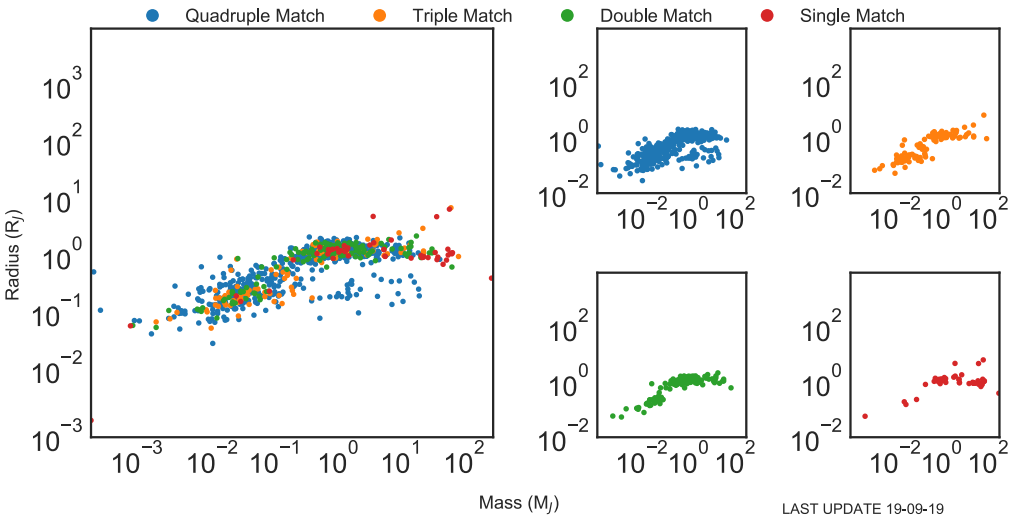


Fig. 5.6: Mass-Radius plot for the EMC run. As shown in the legend, blue dots refer to all quadruple matches (i.e. planets found in all databases), the orange dots to the triple matches (i.e. planets found in three over four databases), the green dots to the double matches (i.e. planets found in two over four databases), and the red ones to the single planets (i.e. planets found only in one database).

in those cases the notation used for the KOI objects is different (KOI-NNNN.DD for the ORG, KNNNN.DD for the EU): this was the cause of the low efficiency in the match, which on the other hand can be corrected by the `Exo-MerCat` catalog.

The amount of quadruple matches for the two cases shows a difference of nearly 300 planets, giving us more confidence concerning the effectiveness of the software. The presence of 60 candidates in this subgroup shows that there is a slight amount of candidates which is common for all catalogs, maybe due to a lack of updates.

By comparing the M-R plots, we notice a much smaller amount of data contained in the subgroup of planets appearing in only one catalog (red sample) for the EMC run compared to the Simple run. On the other hand, the subgroup of planets appearing in all catalogs (blue sample) appears to contain more data. This depends both on the effectiveness of the match for the various targets, but also on the data selection, which allows the final catalog to have, in the end, a higher number of measurements belonging to different reference papers. The trend determined by the theoretical relation present in the ORG catalog is a priori removed in the `Exo-MerCat` software.

5.6 Binary Host Stars

As stated before, it is impossible to figure out a complete sample of binary host stars due to the lack of information from the original catalogs. This is caused by a well-known ambiguity in the nomenclature of such host stars since the discovery of binary companions can often be tricky. Sometimes, most of the primary companions are already known in the community and a change in the notation due to the subsequent discovery of a companion would be confusing.

Other times, the two stars are already known with different names well before discovering that they are gravitationally bound, so the names cannot vary to avoid the chance of mismatch in the literature.

For these and many other reasons, it is no wonder that even the exoplanet databases suffer from such discrepancy. For our purposes, however, this issue does not preclude the possibility of correctly comparing the various targets in the vast majority of cases, since it would suffice to create uniform strings to be compared.

We managed to do that in the `Exo-MerCat` code, as stated previously. However, it is essential to keep in mind that the current version of the code cannot provide a complete sample of all planets orbiting one or more binary companions, since the value of the `binary` string is not indicative of the effective architecture of the system in most cases, but it is only useful to allow the software to correctly compare planets belonging to complex systems.

At the time of writing, the final catalog had 193 non-null values of the `binary` column: within this sample, 68 planets were circumbinary. The unique host star identifiers in the binary sample were 146. The targets displaying a non-null `MismatchFlagHost` flag were 4: therefore, the software failed in recognizing any duplicate entries among these targets.

The remainder of the section will be devoted to a more detailed study of these targets.

- HD 106906 b is a planet discovered by the imaging technique, found in NASA, OEC, and EU catalogs. In the OEC Database, this planet is labeled as circumbinary, while there is no information concerning the architecture of the planet in the other archives. When checking the coordinates for all entries in the global DataFrame, a difference in coordinates up to 2 degrees appears in the EU catalog and this forbids any correction for this row. The coordinates stored in NASA and OEC archives are, on the other hand, in good agreement

and for this reason, the correction of the `binary` label can be made for this couple. In the final catalog, therefore, two rows for this planet, with different coordinates and different values of the `binary` string. In any case, SIMBAD successfully retrieves the correct coordinates since the host star name is well known within the archive. The ambiguity between the catalogs is probably due to the recent discovery of the binary nature of HD 106906, which belongs to the Lower Centaurus Crux (LCC) group in ScoCen association, which was claimed a few years after the discovery of the planetary companion [Bailey et al., 2013]. The latter is on a wide orbit (with a period of about 3000 years), while the binary stars are very close, with an orbital period of 100 days [Rodet et al., 2017], so it is highly probable that the planet is on a P-type orbit. For this reason, we forced the merging of the two duplicated rows, discarding the set of coordinates belonging to the Exoplanet Encyclopaedia and setting the `binary` string to "AB".

- HD 131399 b is a directly imaged candidate, found in EU and OEC catalogs. According to the EU catalog, it orbits an S-type orbit around the primary companion (its `binary` string is "A"), while the OEC catalog suggests that it follows a circumbinary orbit (its `binary` string is "AB"). The architecture of the host system is indeed complex, HD 131399 supposed to be a system composed by a primary A and a couple of close binaries BC. The planetary candidate was claimed as a companion of the primary star [Wagner et al., 2016], later found to be a close binary system as well [Lagrange et al., 2017]. This is probably the source of the mismatch between the `binary` strings. Further studies by Nielsen et al. (2017) assessed that the target should be a background star, thus being a false positive. A check on the consistency of the two duplicated rows, the measurements of all parameters being the same, allowed us to assume that the target was indeed the same in both archives and we forced the merging, then assigning the flag "FALSE POSITIVE" to its status.

- Kepler-420 b is a transiting planet discovered with Kepler and confirmed with follow-up radial velocities observations. This target is present in the four catalogs, with different names: in the EU and ORG catalog, the original string was either Kepler-420 A b or Kepler-420 b; in NASA and OEC catalogs, the KOI notation was present (KOI-1257 b). This target had a non-null binary flag in all source DBs except for the NASA Archive.

  As stated by Santerne et al. (2014), this target is a planet that is most likely to orbit the primary companion of a binary/multiple group, so the expected `binary` string should be A. The check in coordinates is not effective in this case

to identify the four entries as the same target, since the value of the declination shows about half a degree difference in the EU catalog with respect to the other ones, despite being the other parameters consistent in all the archives. The main identifier is however easily found in SIMBAD.

We, therefore, had to simply force the value of the `binary` string to be A for all entries to ensure a perfect match among the various entries.

During the tests and the validation of the software, also ROXs-42 B b had a non-null potential mismatch flag. This is a planet on a circumbinary orbit around the binary ROXs-42 B (AB) in the Ophiucus Molecular Cloud. This target is found in NASA, EU, and OEC archive. For the latter, the code managed to retrieve information about the P-type orbit, while no information is retained in the remaining two. The software, at first, interpreted the B in the planet name string and assigned the value "B" to the `binary` cell. When comparing the various entries during the coordinate check, for this target a disagreement between two non-null values of `binary` was present. In this case, the issue originated from the very name of the stellar system, which is labeled B not because of its binary nature, but because it is the second-brightest counterpart of the X-ray source ROXs-42. This source is too largely separated from ROXs-42 A and C to be gravitationally bound to those [Kraus et al., 2013], but it is a close binary system by itself [Simon et al., 1995].

Further updates on the original databases automatically solved this issue, allowing all available entries to be in agreement concerning the nature of this planet-star system.

5.7 Brown Dwarfs

As shown in Table 5.2 and 5.4, the four original databases do not seem to have similar constraints over the mass and radius values of the objects to include in the sample. Sometimes, the values that are present in the database do not follow the selection criteria reported in the official documentation. This could be surely caused by errors or lack of updated version of either the database itself or the ancillary documentation files, but it underlines an issue that is still debated, concerning the nature of the brown dwarf sample.

These are in fact intermediate objects with masses so low to cause the electron degeneracy pressure to forbid hydrogen burning in their cores, which is an essential requirement to be a low-mass star. For a review on these object see e.g. Luhman

(2012). Because of this process, the upper boundary for a brown dwarf's mass is supposed to be less than 72 M_J (the hydrogen-burning mass limit, see e.g. Schneider (2018)).

Earlier in their evolution, brown dwarfs are able to burn deuterium (or lithium, for more massive objects – about 60 M_J). This process should determine the lower boundary of 13 M_J as the threshold which divides brown dwarfs from massive exoplanets.

Formation by itself should provide a boundary between planets and brown dwarfs: the first class of objects is expected to gain mass by accretion of planetesimals from the surrounding dust disk, while the other is expected to form by gravitational collapse from the original gas cloud.

However, formation models still provide a great degeneracy on the expected mass of both classes of objects, depending on the initial condition of the system which is no longer observable: indeed, different formation processes could lead to higher or smaller masses, making the 13 M_J boundary less relevant.

Since the brown dwarfs cool progressively with time and observables such as temperature, luminosity, and mass often overlap with young massive exoplanets, or extremely old small stars [Faherty, 2018].

Because of this degeneracy in the observables, the classification of such objects is somewhat still arbitrary.

The Exoplanet Encyclopaedia, for example, follows the arguments supported by Hatzes et al. (2015), for which the mass-density relation follows a well-defined trend up to 60 M_J, after that a dramatic change in the slope happens, leading to much smaller densities and larger masses, typical for stellar objects. This theory, however, suffers from the lack of objects in the 30/60 M_J (the so-called brown dwarf desert), as well as the difficulties in measuring radii for the available objects. Large errors in the mass measurement are also probable, especially for direct imaging candidates, since they rely only on photometry and models [Schneider, 2018].

Major improvements on this topic are foreseen with future radial velocity surveys as well as astrometric data from Gaia.

At the time of writing, 222 targets whose best mass is higher than 13 M_J are present in the merged catalog (see Figure 5.7). Most of them belong to the Exoplanet Encyclopedia and are assumed as confirmed (167), while 53 are labeled as candidates and 2 as false positives. Of this sample, only 15% has a measurement of the radius. Overall, errors on the mass increase up to 80%.

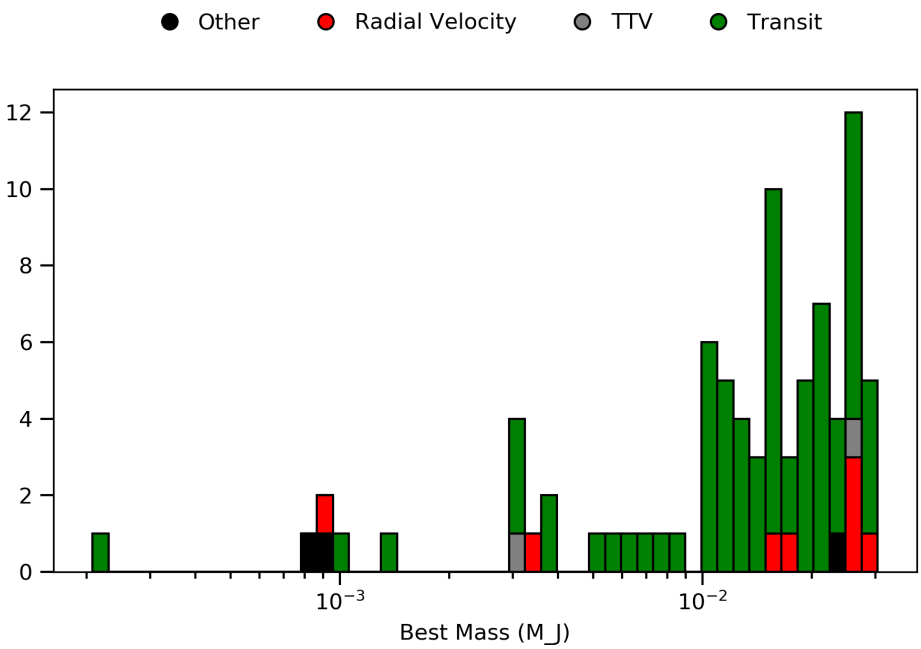


Fig. 5.7: Distribution of the best masses of brown dwarf sample.

The sample of objects whose mass is higher than 60 M_J reduces to 16 targets, the majority of which are objects at slightly higher masses with large enough errors to be under the boundary. The two false positives belong to this sample as well, four are candidates and the remaining ones are confirmed targets. In this case as well, nearly all objects belong to the Exoplanet Encyclopaedia.

5.8 Catalog update and VO interoperable access

The catalog produced by `Exo-MerCat` is both susceptible of updates (due to changes in the originating catalogs and improvements in the knowledge of exoplanets) and useful when used in combination with other astrophysical databases. The former feature makes it clear that subsequent runs of the code that generates it are needed, the latter brings in the interoperability scenario and thus the idea of having it exposed as a VO resource.

The Exo-MerCat workflow is described with the Common Workflow Language (CWL[17]), a standard specification designed purposely for describing analysis workflows and tools with YAML (YAML Ain't a Markup Language) structured text files.

Since CWL is a standard, it can provide a high level of interoperability and portability across different hardware environments. CWL is excellent for jobs that must be run periodically, as Exo-MerCat, because all input parameters are stored in YAML files, which can be versioned. Finally, the wide flexibility of CWL in workflow description is optimal to keep the whole workflow description updated for any future upgrade of Exo-MerCat software.

Once the Exo-MerCat catalog is available as a table in a database system, it becomes easy to annotate it with proper metadata and register it as a resource in the VO ecosystem. To do that, first the Exo-MerCat table was included and described within a TAP service; the catalog was then registered as a VO resource, together with the aforementioned TAP service[18].

Such metadata description and resource registration allow for the catalog to be visible and consumable by all the VO-aware TAP-enabled client applications, like TOPCAT [Taylor, 2005]. This will improve catalog visibility and interoperability of the data resource. As an example of the use of TOPCAT, in Figure 5.8 the distribution of the objects contained in the Exo-MerCat catalog in the Galaxy is known. The knob of Kepler objects is easily recognizable.

Table 5.6 reports the columns the catalog is composed of, with descriptions and data type domain. This forms the basis from which metadata information has been derived for the inclusion of the catalog as a VO resource.

Tab. 5.6: Default column headers, meaning, and type.

Header	Meaning	Type
name	The name of the planet.	STRING
host	The name of the host star.	STRING
letter	The letter labeling the planet.	STRING
ra_off	The official right ascension in degrees, as provided by SIMBAD/K2/EPIC/Gaia catalogs.	FLOAT

[17] https://www.commonwl.org

[18] The Exo-MerCat IVOID is *ivo://ia2.inaf.it/catalogues/exomercat*, while the TAP service IVOID is *ivo://ia2.inaf.it/tap/projects*. The former identifier should persistently represent the Exo-MerCat catalog and its evolution in time, the latter the TAP service that currently deploys Exo-MerCat content.

Tab. 5.6: (continued)

Header	Meaning	Type
dec_off	he official declination in degrees, as provided by SIMBAD/K2/EPIC/Gaia catalogs.	FLOAT
mass	The mass of the planet in Jovian masses.	FLOAT
mass_max	The positive error on the mass measurement in Jovian masses.	FLOAT
mass_min	The negative error on the mass measurement in Jovian masses.	FLOAT
mass_url	The bibcode of the reference paper in which the mass value first appeared.	STRING
msini	The minimum mass of the planet in Jovian masses.	FLOAT
msini_max	The positive error on the minimum mass measurement in Jovian	FLOAT
msini_min	The negative error on the minimum mass measurement in Jovian masses.	FLOAT
msini_url	The bibcode of the reference paper in which the minimum mass value first appeared.	STRING
bestmass	The most precise value between mass and minimum mass of the planet in Jovian masses.	FLOAT
bestmass_max	The positive error on the best mass measurement in Jovian masses.	FLOAT
bestmass_min	The negative error on the best mass measurement in Jovian masses.	FLOAT
bestmass_url	The bibcode of the reference paper in which the mass/minimum mass value first appeared.	STRING
mass_prov	A string labeling if the Best Mass is the mass itself, or the minimum mass.	STRING
p	The period of the planet in days.	FLOAT
p_max	The positive error on the period measurement in days.	FLOAT
p_min	The negative error on the period measurement in days.	FLOAT

Tab. 5.6: (continued)

Header	Meaning	Type
p_url	The bibcode of the reference paper in which the period value first appeared.	STRING
r	The radius of the planet in Jovian radii.	FLOAT
r_max	The positive error on the radius measurement in Jovian radii.	FLOAT
r_min	The negative error on the radius measurement in Jovian radii.	FLOAT
r_url	The bibcode of the reference paper in which the radius value first appeared.	STRING
a	The semi-major axis of the planet in au.	FLOAT
a_max	The positive error on the semi-major axis measurement in au	FLOAT.
a_min	The negative error on the semi-major axis measurement in Jovian masses.	FLOAT
a_url	The bibcode of the reference paper in which the semi-major axis value first appeared.	STRING
e	The eccentricity of the planet (between 0 and 1).	FLOAT
e_max	The positive error on the eccentricity measurement in Jovian masses.	FLOAT
e_min	The negative error on the eccentricity measurement in Jovian masses.	FLOAT
e_url	The bibcode of the reference paper in which the eccentricity value first appeared.	STRING
i	The inclination of the planet in degrees.	FLOAT
i_max	The positive error on the inclination measurement in degrees.	FLOAT
i_min	The negative error on the minimum mass measurement in degrees.	FLOAT
i_url	The bibcode of the reference paper in which the inclination value first appeared.	STRING
main_id	The main identifier of the host star, as provided by SIMBAD/K2/EPIC/Gaia catalogs.	STRING
binary	String labeling the binary host star, if any.	STRING
discovery_method	The discovery method of the planet.	STRING

Tab. 5.6: (continued)

Header	Meaning	Type
Status	The string AXDXEXCX showing the status of the planet in all source catalogs.	STRING
Status_string	Most probable status of the planet.	STRING
confirmed	Number of 3 values in the Status column.	INTEGER
yod	Year of the discovery of the planet.	INTEGER
alias	Known aliases for the host star.	STRING
catalog	List of catalogs in which the target appears.	STRING
MismatchFlagHost	Flag displaying the probable binary duplicate.	INTEGER

At present, the Exo-MerCat workflow is configured to be executed once a week to integrate updates of the four source catalogs.

Further information on the VO registration can be found in Alei et al. (2019).

5.9 Graphical User Interface

The Graphical User Interface can be downloaded from a public GitHub repository[19]. It is a Python 3.6 script, which requires pyvo to query the Exo-MerCat TAP service. In addition to that, matplotlib, numpy, pandas, and guizero need to be installed. This can be easily done by installing via command line using the provided file requirements.txt.

The package guizero[20] allows the user to create a highly customizable interface that offers the user the option to filter the catalog. The user can choose the upper or lower limit for any of the following parameters: mass, radius, period, semi-major axis, eccentricity, and inclination. It is furthermore possible to exclude candidates and false positives and to select one or both mass measurements (mass and minimum mass), if available. It is also possible to select all discovery methods or just a few.

By default, the filtered catalog and the plots will be stored in a folder named as the current date in format YYYYMMDD. If not already present, the folder is created by the GUI code itself. The user can, however, change the name of the folder. The path can

[19] https://gitlab.com/eleonoraalei/exo-mercat-gui
[20] https://lawsie.github.io/guizero/about/

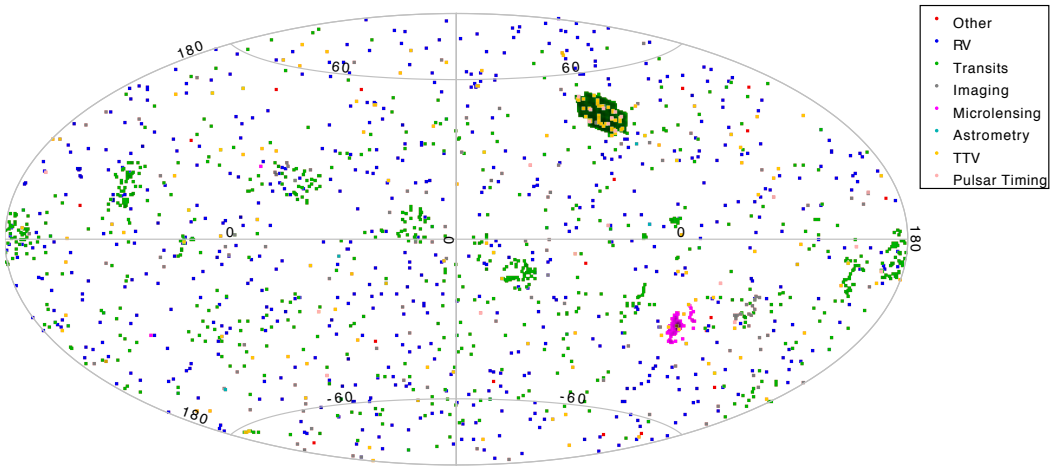


Fig. 5.8: Distribution of the coordinates of `Exo-MerCat` objects (color-coded depending on the discovery method) as shown by TOPCAT (Aitoff projection of equatorial coordinate system).

be changed by specifying its relative position to the enclosing folder (i.e. the folder where the GUI script is saved).

The script checks if the threshold values the user provides are correct and adequate. In particular, when trying to press any of the two buttons at the bottom of the interface, a function is called to check whether a minimum is greater than the corresponding maximum, and/or if a value is set as a nonphysical negative, for each of the selected parameters. If at least one error appears, an alert window is displayed. Any incorrect text is colored in red. The user can correct each value until everything appears to be correct. If this is the case, an infobox pops up with a summary of all selected filters.

The values are stored in the corresponding variables, which are then used to filter the catalog.

By clicking the PLOT button, the script checks the threshold values, filters the catalog and automatically produces a set of standard plots, which are stored in the chosen folder:

- *Radius ($R_\oplus$) vs. Best Mass ($M_\oplus$)*
- *Distance (`pc`) vs. Best Mass ($M_\oplus$)*
- *Period (`days`) vs. Best Mass ($M_\oplus$)*
- *Semi-major axis (AU) vs. Best Mass ($M_\oplus$)*
- *Eccentricity vs. Best Mass ($M_\oplus$)*

- *Eccentricity vs. Semi-major axis (AU)*

And histograms: *Best Mass ($M_\oplus$)*, *Radius ($R_\oplus$)*, *Semi-major axis (au)*, *Period (days)*, *Eccentricity*, *Inclination (deg)*.

The script sorts data depending on the discovery method of each target, thus displaying the items in different colors. A string is also added on the bottom right corner of each plot to show the latest update of the graphs.

If both mass measurements are selected, the plot shows the value of the best mass measurement i.e. the one with the lowest relative error. If the best mass is the minimum mass itself, the target marker is a diamond instead of a circle. For the mass histogram, the best mass is the plotted parameter.

By default, all axes in plots are displayed in logarithmic scale, except those plots concerning the eccentricity. Also, all error bars are shown. All axes in the histograms are also displayed in logarithmic scale, except the horizontal axes concerning eccentricity and inclination.

However, depending on the filtering, it may be useful to switch on and off the logarithmic scale for any of the axes. This can be done by clicking the ADVANCED PLOT button. In this case, a second panel is opened to show a set of checkboxes (the default values will appear as already selected). The user can deselect an entire plot, so that it won't be created, and/or determine the scale of each axis. For greater clarity, error bars can be deselected.

5.10 Future Prospects

The effort is still ongoing with further improvements and the development of new features, such as the possibility to query for one or more older versions of the catalog: this is essential to allow any astronomer to dig deeper into the history of a certain target, by studying the variation of any measurement in time; it could be furthermore useful to retrieve an old version of the catalog, corresponding to the sample of targets known up to a specific date in time, to compare it with the current sample.

Other catalogs may be connected to this one, possibly linking the targets to the available observed data, whether raw or already refined by some data analysis. A more automated connection with the host star catalogs could be also established, to allow the user to retrieve useful information concerning the star.

We should point out that this script would be nothing more than a cross-match among different sources if only the currently available exoplanet catalogs were fully VO-aware, with a common Unified Content Descriptor (UCD) for each parameter. This process of database standardization is starting to be more and more common for the most important stellar catalogs, which can now be easily handled by any VO-aware tool.

Due to the youth of this field, this standardization is still not so straightforward for the exoplanets. For this reason, we started to work (under the EU H2020 ASTERICS project) on the delineation of a specific Data Model for this class of targets, which will pick up model components from the IVOA specifications and attach new ones when needed. We expect that, soon, this new Data Model would be fully integrated into `Exo-MerCat`, and that many other sources would choose to follow the path towards standardized labeling of all planetary-related parameters.

5.11 Super Earths Sample

Using the GUI to filter the `Exo-MerCat` final catalog, we selected the subset composed by confirmed Earths and Super Earths, by applying the following filters for mass and radius:

- $M > 0.00315\ M_J = 1\ M_\oplus$ and $M < 0.0315\ M_J = 10\ M_\oplus$;
- $R > 0.071\ R_J = 0.8\ R_\oplus$ and $R < 0.223\ R_J = 2.5\ R_\oplus$.

We thus included both Earth-like and mini-Neptune planets, selecting all targets that had both a measurement of mass and radius.

At the time of writing, the sample was composed of 72 confirmed planets, nearly all of them appearing in the four original catalogs. The list of such targets, as well as some parameters, are shown in Table 5.7.

Those targets were discovered mainly through transit detection (66 items), plus a planet (Kepler-138 d) found with the transit timing variation. The rest of the sample was discovered with the radial velocities technique.

However, due to the filter applied, it is most likely that both transits and radial velocities techniques were used to characterize the various measurements. As a matter of fact, due to the lack of precise radius measurements, Proxima Cen b cannot be included in the sample.

As visible in Table 5.7, the orbital parameters are not available for all targets: the semi-major axis is available for the 80% of the sample; data on inclination is shown only for the 40% of the sample (even though it is still possible to infer that the inclination would be somewhat similar to 90 degrees in order to detect the planet with the transiting technique); the eccentricity is available only for the 40% of the planets. Period measurements appear to be available for all the sample, as well as the mass and radius ones (required by the filters).

The values in Table 5.7 are rounded to the significant digits of the corresponding errors; the apex is the upper error, while the subscript is the lower one. When no errors are available, the measurement is left without errors.

The `Exo-MerCat` GUI provided also some plots and histograms which are indeed useful to better analyze the features of the subsample:

- The mass values (Figure 5.9) are roughly evenly distributed within the filtered range, with an overabundance of targets at about 5 Earth masses. The planets discovered via radial velocities are more massive ($M > 5\ M_{\oplus}$), while the one discovered via TTV belongs to the Earth analogs subsample.

- The radii distribution (Figure 5.10) appears to be peaked at around 1.7 Earth radii, as well as at 1.2 and 2.2 Earth radii, suggesting a probable difference in the internal composition of the targets.

- The presence of different kinds of planets in the sample is visible employing a mass-radius plot (Figure 5.11), in which the colored lines represent the mass-radius relationship for a few particular internal compositions. These correspond to a planet fully composed by iron, silicates, or ice, and are calculated by Seager, Kuchner, et al. (2007) through a modified equation of state (that will be later applied for the retrieval of simulated models, see Section 7). The targets appear to have densities well in agreement with one or more theoretical trends, confirming qualitatively the presence of different planetary interiors. It is not possible to infer more than this, due to the large errors in the measurements.

- Figure 5.12 shows that the majority of the sample has eccentricity values very close to zero, thus suggesting that most of the targets follow a circular orbit around their host stars. Because of the short distances, as seen in Figure 5.13, it is likely that part of the sample could be composed by tidally-locked planets on circularized orbits, or at least by planets in some sort of orbital resonance.

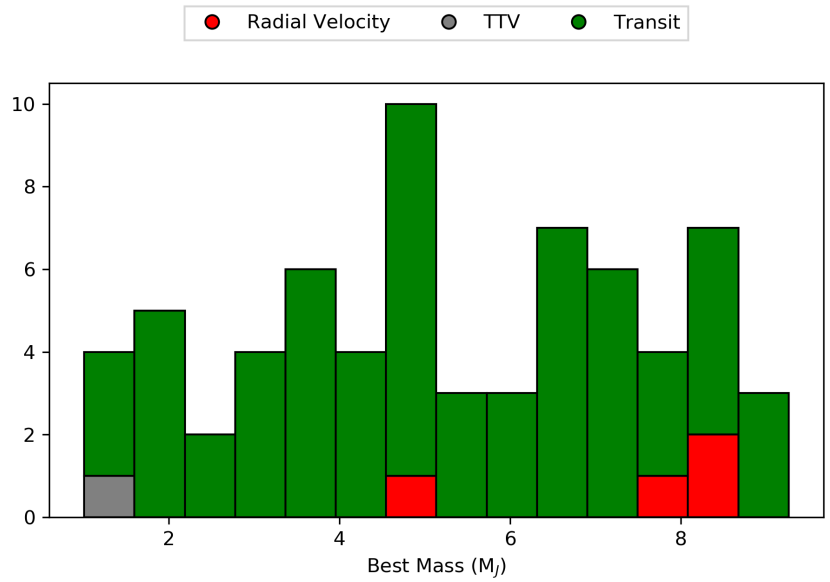


Fig. 5.9: Histogram for the available mass values in the confirmed Earths and Super Earths with available mass and radius measurements sample.

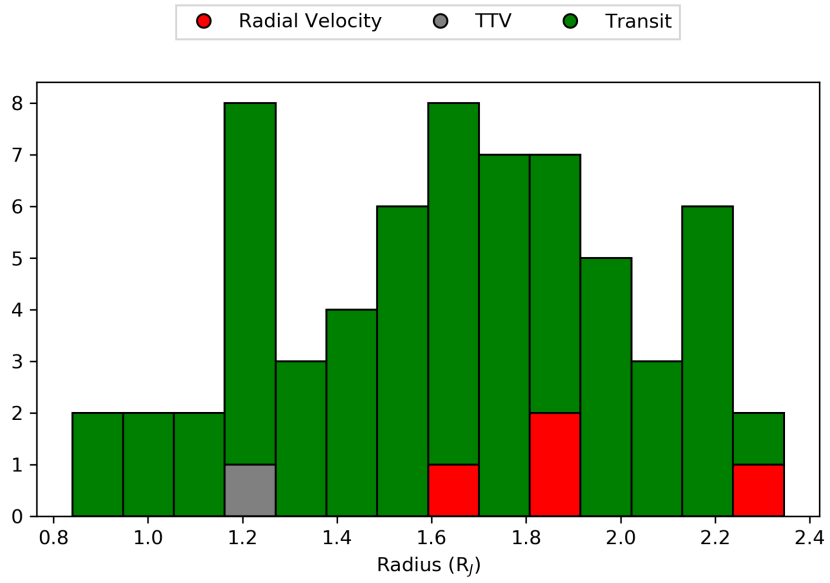


Fig. 5.10: Histogram for the radius values in the confirmed Earths and Super Earths with available mass and radius measurements sample.

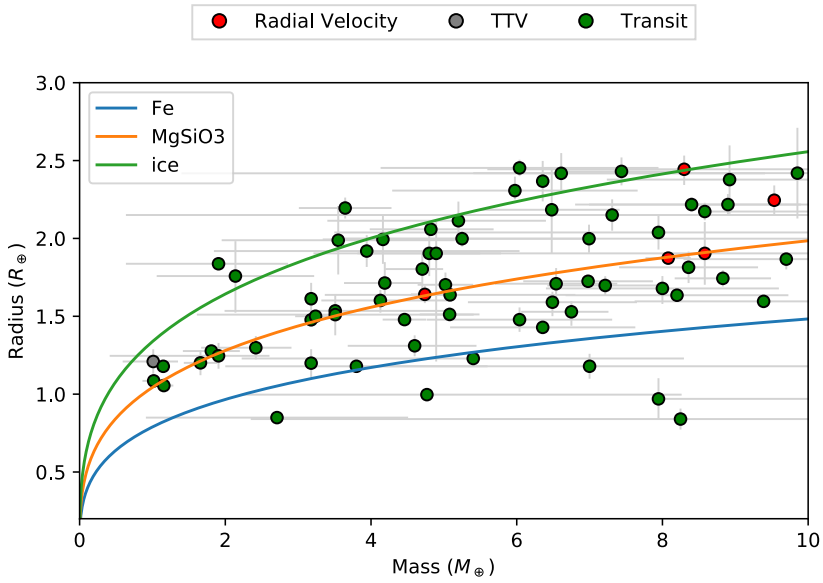


Fig. 5.11: Mass vs. radius for all confirmed Earths and Super Earths with available mass and radius measurements. The lines show the theoretical mass-radius relationships for a few proxy planets: in blue, the mass-radius relationship for a planet composed by iron only; in orange, a planet composed by $MgSiO_3$ perowskite only; in green, a planet composed by H_2O ice only [Seager, Kuchner, et al., 2007].

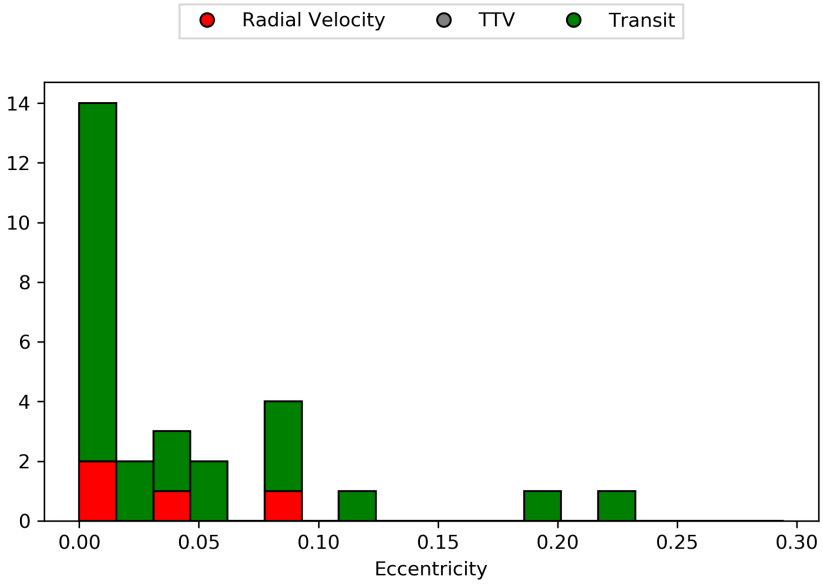


Fig. 5.12: Histogram for the available eccentricity values in the confirmed Earths and Super Earths with available mass and radius measurements sample. In blue: mass-radius theoretical relationship for a pure iron planet; in orange: mass-radius theoretical relationship for a pure $MgSiO_3$ planet; in green: mass-radius theoretical relationship for an icy planet.

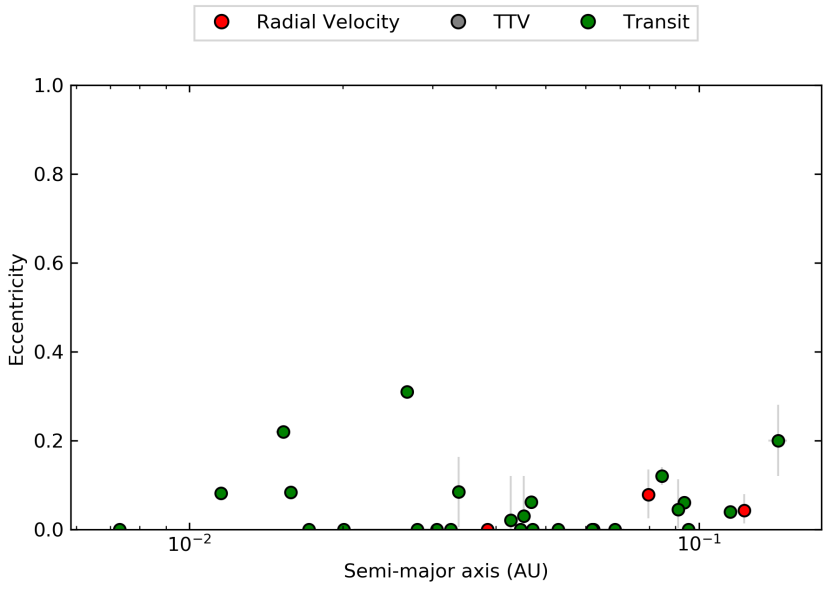


Fig. 5.13: Eccentricity vs. distance for all confirmed Earths and Super Earths with available mass and radius measurements.

- The inclination values (Figure 5.14) are clustered at around 90 degrees, as expected when observing a transiting planet.

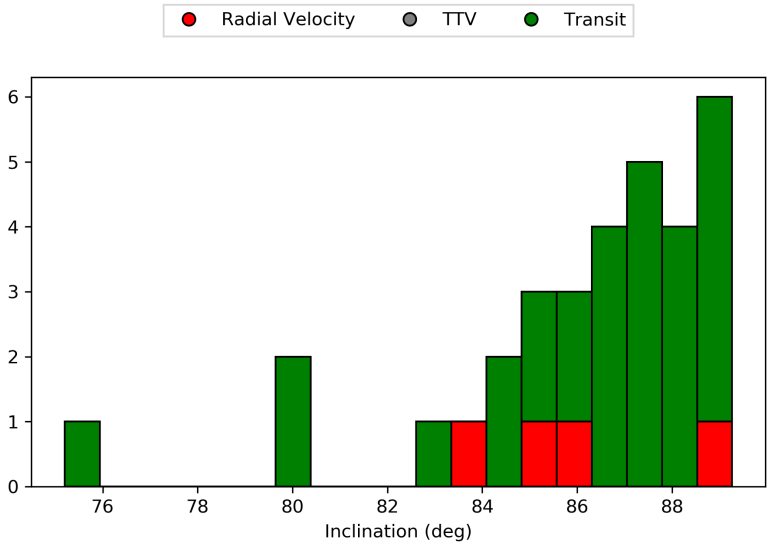


Fig. 5.14: Histogram for the available inclination values in the confirmed Earths and Super Earths with available mass and radius measurements sample.

- The periods histogram (Figure 5.15) shows a peak at periods up to 5 days, which reflects in the semi-major axes histogram (Figure 5.16) lower than 0.05 AU. This is caused by biases in the observation techniques, which are at present sensitive mostly to close-in planets.

High irradiation from the host stars is expected for many of these targets, as clearly visible in Figure 5.17: none of the planets in the sample appears to be at a distance that, in principle, could allow having liquid water on the surface (within the minimum greenhouse-runaway greenhouse calculated by Kopparapu et al. (2013)).

However, different values of albedo, as well as the presence of an atmosphere could change the surface conditions. The study of theoretical atmospheres on such planets may thus yield interesting results concerning habitability (see Section 7).

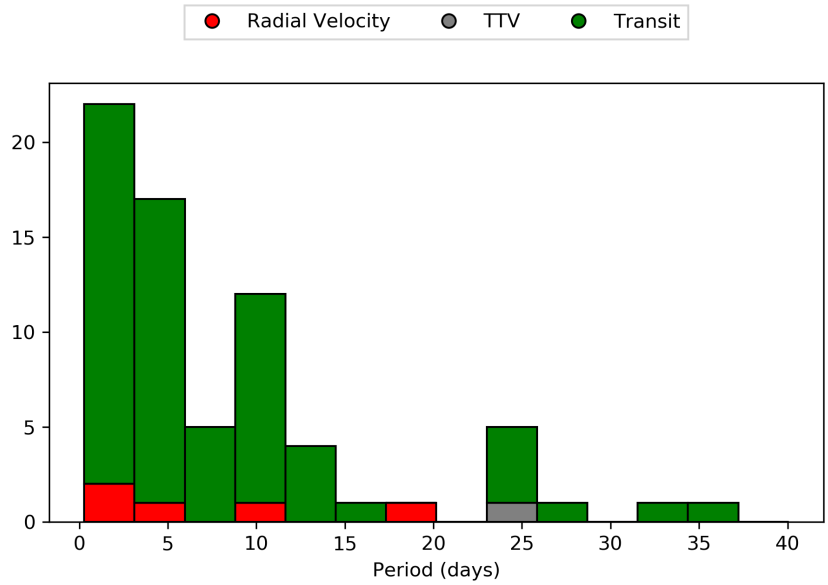


Fig. 5.15: Histogram for the available period values in the confirmed Earths and Super Earths with available mass and radius measurements sample.

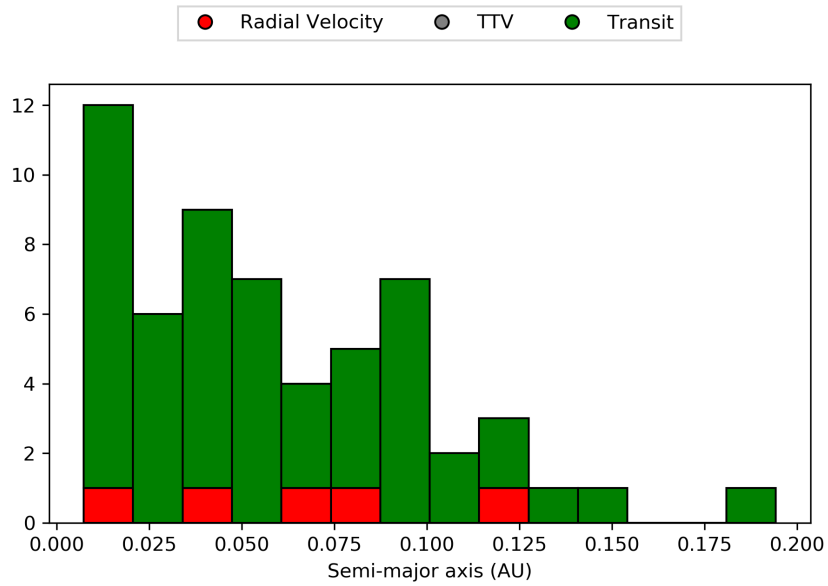


Fig. 5.16: Histogram for the available semi-major axis values in the confirmed Earths and Super Earths with available mass and radius measurements sample.

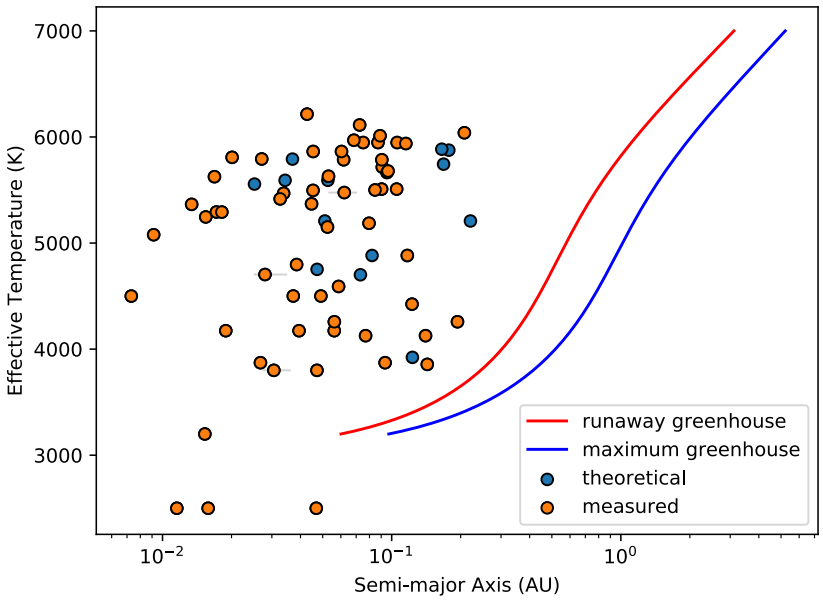


Fig. 5.17: Effective stellar temperature vs. semi-major axis for all confirmed Earths and Super Earths with available mass and radius measurements. Orange dots represent retrieved semi-major axes measurements (from period and eccentricity measurements); blue dots represent a theoretical calculation from the third Kepler's law, supposing a null eccentricity. The red line is the runaway greenhouse boundary, while the blue line is the maximum greenhouse boundary, from Kopparapu et al. (2013).

Tab. 5.7: Masses, radii, orbital periods, semi-major axis and year of discovery of currently known Super Earths, according to `Exo-MerCat`. For greater clarity, error values are not displayed if not available or less than the round-off error.

Name	ra (deg)	dec (deg)	M ($M_\oplus$)	R ($R_\oplus$)	P (day)	a (AU)	i (deg)	e	Method	Status	Year
55 Cnc e	133.1492	28.3308	$8.1^{+0.3}_{-0.3}$	$1.87^{+0.03}_{-0.03}$	$0.736547^{+1.00e-06}_{-1.00e-06}$	$0.01544^{+1.00e-05}_{-1.00e-05}$	$83.6^{+0.4}_{-0.5}$		RV	A3D3E3C3	2004
CoRoT-7 b	100.9561	−1.063	3.1782	$1.6^{+0.1}_{-0.1}$	$0.85359163^{+6.00e-09}_{-6.00e-09}$	$0.0172^{+0.0003}_{-0.0003}$	$80.1^{+0.3}_{-0.3}$	0.0	Transit	A3D3E3C3	2009
LHS 1140 b	11.2472	−15.2715	$7.0^{+0.9}_{-0.9}$	$1.72^{+0.03}_{-0.03}$	$24.73696^{+8.00e-05}_{-8.00e-05}$	$0.094^{+0.002}_{-0.002}$	$89.89^{+0.03}_{-0.05}$	0.06	Transit	A3D3E3C3	2017
HD 3167 b	8.7397	4.3815	$5.0^{+0.4}_{-0.4}$	$1.7^{+0.08}_{-0.08}$	$0.95963^{+1.00e-05}_{-1.00e-05}$	$0.0182^{+0.0002}_{-0.0002}$	$85.0^{+5.0}_{-4.0}$		Transit	A3D3E3C3	2016
HD 97658 b	168.6382	25.7104	9.5346	$2.25^{+0.09}_{-0.1}$	$9.491^{+0.002}_{-0.002}$	$0.08^{+0.001}_{-0.001}$	$89.1^{+0.4}_{-0.4}$	$0.08^{+0.05}_{-0.06}$	RV	A3D3E3C3	2010
Kepler-68 b	291.0324	49.0403	$6.0^{+2.0}_{-2.0}$	$2.31^{+0.09}_{-0.06}$	$5.398753^{+1.00e-06}_{-1.00e-06}$	$0.0617^{+0.0006}_{-0.0006}$	$87.6^{+0.9}_{-0.9}$	0.0	Transit	A3D3E3C3	2011
Kepler-21 b	287.3618	38.714	$5.0^{+2.0}_{-2.0}$	$1.64^{+0.01}_{-0.02}$	$2.78578^{+3.00e-05}_{-3.00e-05}$	0.0427	$83.2^{+0.3}_{-0.3}$	$0.0^{+0.1}_{-0.1}$	Transit	A3D3E3C3	2011
HD 219134 b	348.3207	57.1684	$4.7^{+0.2}_{-0.2}$	$1.64^{+0.06}_{-0.06}$	$3.09293^{+1.00e-05}_{-1.00e-05}$	$0.038474^{+8.00e-07}_{-8.00e-07}$	$85.06^{+0.08}_{-0.08}$	0.0	RV	A3D3E3C3	2015
K2-18 b	172.5605	7.5884	$9.0^{+2.0}_{-2.0}$	$2.4^{+0.2}_{-0.2}$	$32.9396^{+0.0001}_{-0.0001}$	$0.143^{+0.006}_{-0.006}$	$89.578^{+0.009}_{-0.008}$	$0.2^{+0.08}_{-0.08}$	Transit	A3D3E3C3	2015
K2-3 b	172.335	−1.4548	$6.5^{+0.9}_{-1.0}$	$2.2^{+0.3}_{-0.3}$	$10.05463^{+1.00e-05}_{-1.00e-05}$	$0.077^{+0.004}_{-0.004}$	$89.6^{+0.4}_{-0.4}$		Transit	A3D3E3C3	2015
K2-3 c	172.335	−1.4548	$2.0^{+1.0}_{-1.0}$	$1.8^{+0.2}_{-0.2}$	$24.64658^{+4.00e-05}_{-4.00e-05}$	$0.14^{+0.007}_{-0.007}$	$89.7^{+0.2}_{-0.2}$		Transit	A3D3E3C3	2015
Kepler-107 b	297.0282	48.2086	$4.0^{+2.0}_{-2.0}$	$1.535^{+0.002}_{-0.002}$	$3.180022^{+3.00e-06}_{-3.00e-06}$	$0.0454^{+0.0004}_{-0.0004}$	$89.0^{+0.7}_{-0.7}$		Transit	A3D3E3C3	2014
Kepler-107 c	297.0282	48.2086	$9.0^{+2.0}_{-2.0}$	$1.6^{+0.03}_{-0.03}$	$4.90145^{+1.00e-05}_{-1.00e-05}$	$0.0604^{+0.0005}_{-0.0005}$	$89.5^{+0.4}_{-0.4}$		Transit	A3D3E3C3	2014
Kepler-48 b	299.1392	40.949	$4.0^{+2.0}_{-2.0}$	$1.9^{+0.1}_{-0.1}$	4.778				Transit	A3D3E3C3	2011
Kepler-48 d	299.1392	40.949	$8.0^{+4.0}_{-4.0}$	$2.0^{+0.1}_{-0.1}$	42.8961				Transit	A3D3E3C3	2014
Kepler-114 b	294.1214	48.3495	$7.0^{+3.0}_{-3.0}$	$1.18^{+0.08}_{-0.08}$	$5.18855^{+1.00e-05}_{-1.00e-05}$	$0.059^{+0.001}_{-0.001}$			Transit	A3D3E3C3	2013
Kepler-11 b	297.1151	41.9091	$2.0^{+1.0}_{-1.0}$	$1.84^{+0.02}_{-0.02}$	$10.3038^{+0.0002}_{-0.0002}$	$0.091^{+0.001}_{-0.001}$	$89.6^{+0.1}_{-0.1}$	$0.04^{+0.04}_{-0.07}$	Transit	A3D3E3C3	2010
Kepler-307 b	297.7951	40.4177	$7.4^{+0.9}_{-0.9}$	$2.43^{+0.09}_{-0.09}$	$10.4208^{+0.0008}_{-0.0009}$	$0.09^{+0.002}_{-0.002}$			Transit	A3D3E3C3	2013
Kepler-307 c	297.7951	40.4177	$3.6^{+0.6}_{-0.6}$	$2.2^{+0.07}_{-0.07}$	$13.073^{+0.001}_{-0.001}$	$0.105^{+0.002}_{-0.002}$			Transit	A3D3E3C3	2013
Kepler-60 b	288.9612	42.265	$4.2^{+0.5}_{-0.6}$	$1.7^{+0.1}_{-0.1}$	$7.13162^{+9.00e-05}_{-9.00e-05}$	$0.075^{+0.001}_{-0.001}$			Transit	A3D3E3C3	2012
Kepler-60 c	288.9612	42.265	$5.0^{+1.0}_{-1.0}$	$1.9^{+0.1}_{-0.1}$	$8.9193^{+0.0001}_{-0.0001}$	$0.087^{+0.001}_{-0.001}$			Transit	A3D3E3C3	2012
Kepler-60 d	288.9612	42.265	$4.2^{+0.8}_{-0.8}$	$2.0^{+0.2}_{-0.2}$	$11.9016^{+0.0002}_{-0.0002}$	$0.106^{+0.002}_{-0.002}$			Transit	A3D3E3C3	2012
Kepler-68 c	291.0324	49.0403	$5.0^{+3.0}_{-3.0}$	$1.0^{+0.02}_{-0.02}$	$9.60498^{+4.00e-05}_{-4.00e-05}$	$0.0906^{+0.0008}_{-0.0008}$	$86.9^{+0.4}_{-0.4}$		Transit	A3D3E3C3	2013
Kepler-454 b	287.4785	38.2289	6.3564	$2.4^{+0.1}_{-0.1}$	$10.57^{+7.00e-08}_{-7.00e-08}$	$0.095^{+0.001}_{-0.001}$	$87.1^{+0.2}_{-0.2}$	0.0	Transit	A3D3E3C3	2015

Tab. 5.7: (continued)

Name	ra (deg)	dec (deg)	M ($M_\oplus$)	R ($R_\oplus$)	P (day)	a (AU)	i (deg)	e	Method	Status	Year
Kepler-36 b	291.2502	49.2318	$4.5^{+0.3}_{-0.3}$	$1.48^{+0.03}_{-0.03}$	$13.8399^{+0.0006}_{-0.0008}$	$0.115^{+0.002}_{-0.002}$	$90.0^{+0.7}_{-0.7}$	0.039	Transit	A3D3E3C3	2011
Kepler-131 c	288.5309	40.9423	$8.0^{+6.0}_{-6.0}$	$0.84^{+0.07}_{-0.07}$	25.5169				Transit	A3D3E3C3	2014
Kepler-92 c	289.0861	41.563	$6.0^{+2.0}_{-2.0}$	$2.45^{+0.05}_{-0.05}$	$26.7231^{+0.0002}_{-0.0002}$				Transit	A3D3E3C3	2013
Kepler-406 b	291.8481	44.9683	$6.0^{+1.0}_{-1.0}$	$1.43^{+0.03}_{-0.03}$	2.4263				Transit	A3D3E3C3	2014
Kepler-406 c	291.8481	44.9683	$3.0^{+2.0}_{-2.0}$	$0.85^{+0.03}_{-0.03}$	4.6233				Transit	A3D3E3C3	2014
Kepler-80 d	296.1126	39.9788	$6.8^{+0.5}_{-0.7}$	$1.53^{+0.07}_{-0.09}$	$3.07219^{+1.00e-05}_{-1.00e-05}$	$0.0372^{+0.0005}_{-0.0005}$	$88.0^{+2.0}_{-1.0}$		Transit	A3D3E3C3	2012
Kepler-80 e	296.1126	39.9788	$4.1^{+1.0}_{-0.8}$	$1.6^{+0.07}_{-0.08}$	$4.64539^{+1.00e-05}_{-1.00e-05}$	$0.0491^{+0.0007}_{-0.0007}$	$88.8^{+1.0}_{-0.8}$		Transit	A3D3E3C3	2012
Kepler-197 c	295.2264	50.559	$5.0^{+3.0}_{-3.0}$	$1.23^{+0.04}_{-0.04}$	$10.3497^{+4.00e-05}_{-4.00e-05}$	$0.089^{+0.001}_{-0.001}$	$89.3^{+0.7}_{-0.7}$		Transit	A3D3E3C3	2014
Kepler-93 b	291.4183	38.6723	3.1782	$1.48^{+0.02}_{-0.02}$	$4.72674^{+1.00e-06}_{-1.00e-06}$	$0.053^{+0.002}_{-0.002}$	$89.18^{+0.04}_{-0.04}$	0.0	Transit	A3D3E3C3	2014
Kepler-102 d	281.4827	47.208	$4.0^{+2.0}_{-2.0}$	$1.18^{+0.04}_{-0.04}$	10.3117				Transit	A3D3E3C3	2014
Kepler-102 e	281.4827	47.208	$9.0^{+2.0}_{-2.0}$	$2.22^{+0.07}_{-0.07}$	$16.14569^{+3.00e-05}_{-3.00e-05}$	$0.117^{+0.003}_{-0.003}$	$89.6^{+0.6}_{-0.6}$		Transit	A3D3E3C3	2013
Kepler-19 b	290.4208	37.8518	$8.0^{+2.0}_{-2.0}$	$2.22^{+0.04}_{-0.04}$	$9.286994^{+9.00e-06}_{-9.00e-06}$	$0.085^{+0.001}_{-0.001}$	$89.9^{+0.4}_{-0.2}$	$0.12^{+0.02}_{-0.02}$	Transit	A3D3E3C3	2011
Kepler-10 b	285.6794	50.2413	$3.2^{+0.3}_{-0.3}$	$1.5^{+0.02}_{-0.03}$	$0.8374907^{+2.00e-07}_{-2.00e-07}$	$0.0168^{+0.0001}_{-0.0001}$	$84.0^{+2.0}_{-1.0}$		Transit	A3D3E3C3	2011
Kepler-105 c	287.8873	46.2762	$4.6^{+0.8}_{-0.9}$	$1.31^{+0.07}_{-0.07}$	$7.12594^{+2.00e-05}_{-2.00e-05}$	$0.072^{+0.001}_{-0.001}$			Transit	A3D3E3C3	2014
Kepler-18 b	298.0795	44.7463	$7.0^{+3.0}_{-3.0}$	$2.0^{+0.09}_{-0.09}$	$3.50472^{+3.00e-05}_{-3.00e-05}$	$0.0447^{+0.0006}_{-0.0006}$	$84.9^{+0.3}_{-0.3}$	0.0	Transit	A3D3E3C3	2011
Kepler-20 b	287.698	42.3387	$10.0^{+1.0}_{-1.0}$	$1.87^{+0.03}_{-0.07}$	$3.6961152^{+9.00e-07}_{-1.00e-06}$	$0.0454^{+0.0008}_{-0.0008}$	$86.5^{+0.3}_{-0.4}$	$0.03^{+0.03}_{-0.09}$	Transit	A3D3E3C3	2011
Kepler-289 b	297.4653	42.8829	$7.0^{+7.0}_{-7.0}$	$2.2^{+0.1}_{-0.1}$	$34.5439^{+0.0002}_{-0.0002}$	$0.208^{+0.003}_{-0.003}$	$89.6^{+0.5}_{-0.3}$		Transit	A3D3E3C3	2014
Kepler-78 b	293.7417	44.4483	3.1782	$1.2^{+0.09}_{-0.09}$	$0.35500745^{+6.00e-08}_{-6.00e-08}$	$0.0091^{+0.0002}_{-0.0002}$	$75.0^{+2.0}_{-3.0}$		Transit	A3D3E3C3	2013
Kepler-9 d	285.574	38.4009	$5.0^{+2.0}_{-2.0}$	$2.0^{+0.05}_{-0.05}$	$1.59285^{+4.00e-05}_{-4.00e-05}$	$0.0271^{+0.0001}_{-0.0001}$			Transit	A3D3E3C3	2010
GJ 1132 b	153.7157	−47.1567	$1.7^{+0.2}_{-0.2}$	$1.2^{+0.08}_{-0.08}$	$1.62893^{+3.00e-05}_{-3.00e-05}$	$0.0153^{+0.0005}_{-0.0005}$		0.22	Transit	A3D3E3C3	2015
TRAPPIST-1 b	346.6224	−5.0414	$1.0^{+0.1}_{-0.2}$	$1.085^{+0.003}_{-0.003}$	$1.5108764^{+4.00e-07}_{-4.00e-07}$	$0.01154775^{+6.00e-08}_{-6.00e-08}$	$89.6^{+0.2}_{-0.2}$	0.081	Transit	A3D3E3C3	2016
TRAPPIST-1 c	346.6224	−5.0414	$1.2^{+0.1}_{-0.1}$	$1.055^{+0.003}_{-0.003}$	$2.4218075^{+9.00e-07}_{-9.00e-07}$	$0.0158151^{+2.00e-07}_{-2.00e-07}$	$89.7^{+0.2}_{-0.2}$	0.083	Transit	A3D3E3C3	2016
TRAPPIST-1 g	346.6224	−5.0414	$1.1^{+0.1}_{-0.1}$	$1.18^{+0.03}_{-0.03}$	$12.35447^{+2.00e-05}_{-2.00e-05}$	$0.0468769^{+3.00e-07}_{-3.00e-07}$	$89.71^{+0.02}_{-0.02}$	0.061	Transit	A3D3E3C3	2017
pi Men c	84.2912	−80.4691	$4.8^{+0.9}_{-0.8}$	$2.06^{+0.03}_{-0.03}$	$6.2683^{+0.0002}_{-0.0002}$	$0.0684^{+0.0005}_{-0.0005}$	$87.27^{+0.07}_{-0.07}$	0.0	Transit	A3D0E3C3	2018
GJ 9827 b	351.7702	−1.2863	$8.0^{+2.0}_{-2.0}$	$1.64^{+0.02}_{-0.02}$	$1.20898^{+8.00e-06}_{-8.00e-06}$	$0.0189^{+0.0003}_{-0.0003}$	$86.0^{+1.0}_{-1.0}$		Transit	A3D3E3C0	2017
GJ 9827 c	351.7702	−1.2863	$2.4^{+0.5}_{-0.8}$	$1.3^{+0.07}_{-0.07}$	$3.64808^{+6.00e-05}_{-6.00e-05}$	$0.0394^{+0.0006}_{-0.0006}$	$87.8^{+0.6}_{-0.6}$		Transit	A3D3E3C0	2017

Tab. 5.7: (continued)

Name	ra (deg)	dec (deg)	M ($M_\oplus$)	R ($R_\oplus$)	P (day)	a (AU)	i (deg)	e	Method	Status	Year
GJ 9827 d	351.7702	−1.2863	$5.0^{+1.0}_{-2.0}$	$2.1^{+0.1}_{-0.1}$	$6.20147^{+6.00e-05}_{-6.00e-05}$	$0.0562^{+0.0009}_{-0.0009}$	$87.4^{+0.2}_{-0.2}$		Transit	A3D3E3C0	2017
K2-106 b	13.0798	10.7947	$8.0^{+1.0}_{-1.0}$	$1.8^{+0.1}_{-0.1}$	$0.57129^{+1.00e-05}_{-1.00e-05}$	$0.0134^{+0.0002}_{-0.0002}$	$86.0^{+4.0}_{-4.0}$		Transit	A3D0E3C3	2016
K2-111 b	59.8897	21.2987	$9.0^{+4.0}_{-4.0}$	$1.9^{+0.2}_{-0.2}$	$5.3512^{+0.0006}_{-0.0006}$	$0.062^{+0.008}_{-0.009}$	$86.0^{+1.0}_{-1.0}$	0.0	RV	A3D0E3C3	2017
K2-38 c	240.0336	−23.1893	$10.0^{+4.0}_{-4.0}$	$2.4^{+0.3}_{-0.3}$	$10.561^{+0.0009}_{-0.0009}$	$0.096^{+0.002}_{-0.002}$	$89.0^{+2.0}_{-1.0}$		Transit	A3D0E3C3	2015
Kepler-106 d	300.864	44.3376	$7.9455^{+8.0}$	$1.0^{+0.1}_{-0.1}$	23.9802				Transit	A3D0E3C3	2014
Kepler-113 c	287.9979	50.9444	$8.5811^{+9.0}$	$2.17^{+0.06}_{-0.06}$	8.9251				Transit	A3D0E3C3	2014
Kepler-97 b	287.3266	48.6734	$4.0^{+2.0}_{-2.0}$	$1.5^{+0.1}_{-0.1}$	2.5866				Transit	A3D3E0C3	2014
Kepler-98 b	285.6616	37.9645	$4.0^{+2.0}_{-2.0}$	$2.0^{+0.2}_{-0.2}$	1.5417				Transit	A3D3E0C3	2014
Kepler-99 b	297.354	41.3001	$6.0^{+1.0}_{-1.0}$	$1.48^{+0.08}_{-0.08}$	4.6036				Transit	A3D3E0C3	2014
K2-141 b	350.9165	−1.1893	$5.1^{+0.4}_{-0.4}$	$1.51^{+0.04}_{-0.04}$	$0.280323^{+1.00e-06}_{-1.00e-06}$	$0.0073^{+0.0001}_{-0.0001}$	$86.0^{+4.0}_{-3.0}$	0.0	Transit	A3D3E3C0	2017
K2-216 b	11.4803	6.347	$8.0^{+2.0}_{-2.0}$	$1.68^{+0.08}_{-0.1}$	$2.17479^{+4.00e-05}_{-4.00e-05}$	$0.028^{+0.007}_{-0.003}$	$87.0^{+4.0}_{-2.0}$	0.0	Transit	A3D0E3C3	2018
K2-265 b	342.0315	−14.4947	$6.5^{+0.8}_{-0.8}$	$1.7^{+0.1}_{-0.1}$	$2.36917^{+9.00e-05}_{-9.00e-05}$	$0.0338^{+0.0002}_{-0.0002}$	$88.0^{+2.0}_{-2.0}$	$0.08^{+0.08}_{-0.08}$	Transit	A3D0E3C0	2018
K2-266 e	157.9354	0.9376	$8.0^{+2.0}_{-3.0}$	$2.44^{+0.09}_{-0.1}$	$19.482^{+0.001}_{-0.001}$	$0.123^{+0.002}_{-0.002}$	$89.7^{+0.2}_{-0.2}$	$0.04^{+0.03}_{-0.04}$	RV	A3D0E3C0	2018
LHS 1140 c	11.2472	−15.2715	$1.8^{+0.4}_{-0.4}$	$1.28^{+0.02}_{-0.02}$	$3.777931^{+3.00e-06}_{-3.00e-06}$	$0.0268^{+0.0007}_{-0.0007}$	$89.92^{+0.09}_{-0.06}$	0.31	Transit	A3D0E3C0	2018
K2-291 b	76.4458	21.5486	$6.0^{+1.0}_{-1.0}$	$1.59^{+0.07}_{-0.09}$	$2.22518^{+7.00e-05}_{-7.00e-05}$	$0.0326^{+0.0004}_{-0.0004}$	$85.2^{+0.2}_{-0.2}$	0.0	Transit	A3D0E3C0	2019
K2-155 b	65.4687	21.3536	$4.7^{+0.3}_{-0.5}$	$1.8^{+0.1}_{-0.2}$	$6.342^{+0.002}_{-0.002}$	$0.056^{+0.001}_{-0.001}$	$88.0^{+2.0}_{-1.0}$		Transit	A3D0E3C0	2018
K2-155 d	65.4687	21.3536	$4.9^{+0.6}_{-2.0}$	$1.9^{+0.2}_{-0.7}$	$40.718^{+0.005}_{-0.005}$	$0.194^{+0.006}_{-0.006}$	$89.6^{+0.5}_{-0.3}$		Transit	A3D0E3C0	2018
Kepler-138 d	290.3815	43.293	$1.0^{+0.3}_{-0.4}$	$1.21^{+0.08}_{-0.08}$	$23.0893^{+0.0007}_{-0.0007}$				TTV	A3D0E3C0	2014
TOI-402 b	36.8682	−27.6352	$7.2^{+0.8}_{-0.8}$	$1.7^{+0.06}_{-0.06}$	$4.7564^{+0.0002}_{-0.0002}$	$0.0524^{+0.0007}_{-0.0008}$	$88.4^{+0.3}_{-0.3}$		Transit	A0D0E3C0	2019
HD 213885 b	338.9846	−59.8645	$8.8^{+0.7}_{-0.7}$	$1.74^{+0.05}_{-0.05}$	$1.00804^{+2.00e-05}_{-2.00e-05}$	$0.0201^{+0.0001}_{-0.0001}$	$80.1^{+0.5}_{-0.5}$	0.0	Transit	A0D0E3C0	2019
TOI-270 b	68.4155	−51.9562	$1.9^{+0.7}_{-1.0}$	$1.25^{+0.08}_{-0.08}$	$3.36008^{+7.00e-05}_{-7.00e-05}$	$0.031^{+0.006}_{-0.003}$	$89.0^{+1.0}_{-1.0}$	0.0	Transit	A0D0E3C0	2019
TOI-270 c	68.4155	−51.9562	$7.0^{+3.0}_{-5.0}$	$2.4^{+0.1}_{-0.1}$	$5.66017^{+3.00e-05}_{-3.00e-05}$	$0.047^{+0.003}_{-0.003}$	$89.5^{+0.4}_{-0.4}$	0.0	Transit	A0D0E3C0	2019

6

MAGRATHEA: modeling of terrestrial atmospheres

> *[...] And thus were created the conditions for a staggering new form of specialist industry: custom-made luxury planet building. The home of this industry was the planet Magrathea, where hyperspatial engineers sucked matter through white holes in space to form it into dream planets - gold planets, platinum planets, soft rubber planets with lots of earthquakes - all lovingly made to meet the exacting standards that the Galaxy's richest men naturally came to expect. [...]*
> *Magrathea is a myth, a faery story! It's what parents tell their children about at night when they want them to grow up to become economists!*
>
> **— Douglas Adams**
> Hitchhiker's Guide to the Galaxy

In this chapter, the development and the validation of the 1D radiative-convective transfer code MAGRATHEA are described. The code is used to run two sets of simulations, as will be mentioned in Chapter 7: the first grid of models will be composed by theoretical Super Earths atmospheres, while the second grid will be based on the sample of known Super Earths retrieved by Exo-MerCat in Chapter 5. Some of the results will be used for both simulations of ozone photochemistry (see Chapter 8) and laboratory experiments (see Chapter 9).

6.1 Overview

The large quantity of processes described in the past chapters makes the characterization of an atmosphere undoubtedly hard; moreover, the lack of observations does not contribute to this task. A theoretical modeling effort is required to rule out the processes that can be more efficient and the parameters that play a role in differentiating the atmospheres, paving the way for the incoming observed data. Many numerical models were therefore built to solve the various coupled differential equations and to infer constraints and parametrizations about all kinds of atmosphere that could probably be observed in the near future.

At first, photochemical models for exoplanets were adapted from Jupiter chemical schemes, including all those endothermic reactions that could be neglected in Jupiter's atmosphere. Later on, kinetic networks and associated reaction rates have been constructed and many other kinds of atmospheres could be modeled [Venot et al., 2015]. Many reaction databases have been developed for common use, such as KIDA[1] [Wakelam et al., 2012], as well as compilations of spectroscopic parameters such as HITRAN[2].

Until recently, most of the studies on terrestrial exoplanets have been performed with 1D steady-state radiative-convective models, that evaluate the global mean conditions on a planetary atmosphere insulated from its host star [Forget et al., 2014]; some of them can include haze and cloud layers and are also able to produce synthetic spectra of the considered atmospheres.

Such models are often not sufficient to predict the actual state of a planet, its formation, its variability due to rotation and revolution around its host star, as well as cloud formation and dynamics. For this purpose, the computation of very complex 3D GCMs (*General Circulation Models* or *Global Climate Models*) is required. Those software programs are, on the other hand, constrained by the lack of observations and by the relatively low achievable computational power [ibid.]. Predicting the climate and the atmospheric dynamics of a planet is still affected by instability, positive feedbacks such as runaway glaciation and greenhouse effect, and nonlinear behavior: those processes are poorly understood, so they are currently missing in the models, enhancing the probability to have inaccurate results.

Besides, the difficulty of assessing the habitability of a planet in alien environments depends mostly on the reliability of simulations in reproducing inherently 3D pro-

[1] http://kida.obs.u-bordeaux1.fr/
[2] http://hitran.org/

cesses, having only a rough estimate of the planet's past irradiation, and orbital history.

For this reason, the use of 1D models, albeit fairly less sensitive than high dimensional models, is useful to test the effects of some processes, or to explore a parameter space too broad for heavy 3D modeling, such as e.g., the impact of differences in the water vapor content (even in clear-sky conditions) on the inner edge of the habitability zone [Petralia et al., 2019, and references therein].

In this simple case, one possibility is to consider the energy balance along single air vertical columns, as implemented e.g., in Radiative-Convective (RC) models, in which the only dimension is the altitude.

This class of models assumes that the vertical structure of an atmosphere is determined by the convection and radiation within the simulated atmospheric column.

RC solutions to temperature profiles are computed numerically, frequently using the so-called convective adjustment (a technique that will be explained in Section 6.6). There is no explicit atmospheric circulation, although it is present implicitly. RC models may provide useful general indications of the impact on the planetary surface temperature of parameters such as ground pressure and CO_2 mixing ratio.

Many unknowns are still on the table, since the initial conditions are fairly uncertain, as well as the processes that couple the surface, the interior, and the atmosphere, as well as biological activity. Many computational improvements can be done in this field, but the comparison with future data and subsequent fine-tuning of the models is still necessary to make a bigger leap forward in the treatment of atmospheres.

Within a collaboration among the INAF Observatories of Padua, Palermo, and Cagliari, we developed `MAGRATHEA`, a 1D radiative-convective code able to simulate temperate martian and terrestrial atmospheres. The effort started more than a decade ago, well before the appearance of many open-source programs, with the construction of this software program from scratch. This allowed to have a deeper understanding of the functionalities of the code and of the processes that were gradually included.

The workflow of the code is shown in Figure 6.1 and the various features of the software will be described in the following sections. Section 6.2 describes the correlated *k* distribution table, Section 6.3 describes the input setup; the chemistry update will be explained in Section 6.4, the radiative transfer in Section 6.5 and the convective adjustment in Section 6.6. Information about the convergence criterion and the handling of the errors are reported in Sections 6.7 and 6.8 respectively.

Finally, the results of MAGRATHEA validations are shown in Section 6.9. For the purpose of this project, MAGRATHEA has been used to simulate a large number of atmospheric models of terrestrial atmospheres, considering all combinations of a vast physical and chemical parameter space. The results of the simulations will be outlined and discussed in Chapter 7.

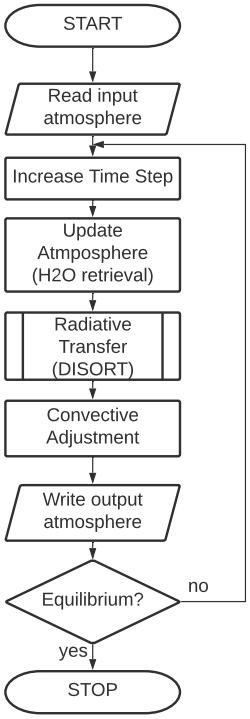


Fig. 6.1: MAGRATHEA workflow.

6.2 Correlated *k* distributions table

When studying the features of an irradiated atmosphere, one should account for broadening effects on the absorption spectra, such as Doppler and pressure broadening [Rybicki et al., 1981]. The Doppler (thermal) broadening in a turbulence-free medium of a line whose natural frequency is ν_0 shape has a width:

$$\alpha_D = \frac{\nu_0}{c}\sqrt{\frac{2k_B T}{m}} \tag{6.1}$$

Where k_B is the Boltzmann constant, and m is the mass of the species involved in the transition at a temperature T. The pressure broadening of a species X of volume mixing ratio ϵ_X at pressure p is modelled using a Lorentzian line shape of width:

$$\alpha_L = p[(1-\epsilon_x)\alpha_f + \epsilon_x \alpha_s] \left(\frac{T_{ref}}{T}\right)^n \tag{6.2}$$

Where n is the temperature exponent; α_f and α_s are the foreign and self-broadening, the first being caused by all the remaining species in the atmosphere and the latter being caused by the species X involved in that transition. These coefficients can be provided by numerical codes at a particular reference temperature T_{ref} and are normally stored in databases such as HITRAN[3] [Gordon et al., 2017], HITEMP[4] [Rothman et al., 2010], and EXOMOL[5] [Tennyson et al., 2016].

The convolved Voigt profile of the line composed by a thermal core and Lorentzian wings is:

$$\Phi(\nu) = \frac{1}{\pi^{-0.5}\alpha_D} H(a,u) \tag{6.3}$$

Where:

$$H(a,u) = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{-y}}{a^2+(u-y)^2} dy \tag{6.4}$$

$$a = \Gamma/(4\pi\alpha_D) \tag{6.5}$$

$$u = (\nu - \nu_0)/\alpha_D \tag{6.6}$$

and where $\Gamma = \gamma + 2\nu_{col}$ is the damping rate (γ being the spontaneous decay rate, and ν_{col} the atomic collision rate).

So, absorption spectra at low pressures are narrow, but they become smooth at high pressures. The dampening caused by pressure comes with a different magnitude within the same atmosphere: here, each line differs from level to level, depending on the layer's pressure. As an example of this broadening effect, Fig. 6.2 shows the simulated absorption spectrum at 15 μm (where a prominent CO_2 band is present) for a given atmosphere at various pressures.

[3] https://hitran.org
[4] https://hitran.org/hitemp/
[5] http://exomol.com

A radiative code needs to consider all contributions from the atmospheric layers: a realistic treatment could be achieved only through *line-by-line calculations*, for which every single wavelength in the spectrum needs to be taken into account.

Considering an atmosphere composed by only four species (CO_2, H_2O, N_2, and O_2 - the chemical database treated by `MAGRATHEA` at the time of writing), the line-by-line absorption spectrum of an atmospheric layer of temperature T and pressure P is built as follows:

$$
\begin{aligned}
k_\nu = & \; k_{CO_2} + k_{H_2O} + k_{N_2} + k_{O_2} = \\
& = \epsilon_{CO_2} \left[\sum_{\delta_{\nu_0} < 25\, cm^{-1}} k_\nu^{CO_2} + (1 - \epsilon_{CO_2}) \left(\frac{p}{p_{ref}} \right) \left(\frac{T_{ref}}{T} \right) k_f^{CO_2} + k_\nu^\star \right] + \\
& + \epsilon_{H_2O} \left[\sum_{\delta_{\nu_0} < 25\, cm^{-1}} k_\nu^{H_2O} + (1 - \epsilon_{H_2O}) \left(\frac{p}{p_{ref}} \right) \left(\frac{T_{ref}}{T} \right) k_f^{H_2O} + \epsilon_{H_2O} \left(\frac{p}{p_{ref}} \right) \left(\frac{T_{ref}}{T} \right) k_s^{H_2O} \right] + \\
& + \epsilon_{N_2} \sum_{\delta_{\nu_0} < 25\, cm^{-1}} k_\nu^{N_2} + \epsilon_{O_2} \sum_{\delta_{\nu_0} < 25\, cm^{-1}} k_\nu^{O_2}
\end{aligned}
\tag{6.7}
$$

where ϵ_i is the volume mixing ratio of the $i-$th species, while $\delta_{\nu_0} = |\nu - \nu_0|$.

The line profiles are Voigt functions, opportunely modified for H_2O and CO_2 to take into account collision-induced absorption ($k_\nu^\star$), as well as foreign (subscript f) self-continuum (subscript s) absorption [Petralia et al., 2019, and references therein].

This treatment increases the computation cost of this kind of calculations, which makes this approach feasible for limited applications. To achieve the goal of a shorter computational time for the radiative transfer calculation, approximations are due.

The effects of atmospheric pressure broadening could be accounted for by scaling the absorber amount with pressure (*single-parameter scaling*), to guarantee computational speed, but it can lead to large errors in the calculation of atmospheric heating and cooling rates, particularly in the stratosphere and upper troposphere [Lacis et al., 1991].

On the other hand, the procedure proposed in Lacis et al. (ibid.) aims to reduce the points that represent the spectral absorption: this can be achieved by sorting a "*correlated k distribution*" of absorption coefficients strengths within a frequency interval and then performing a least-squares fit over that smoother function.

The effects of pressure broadening and general vertical inhomogeneity are modeled through the implicit wavelength correlation of k distributions at different pressure

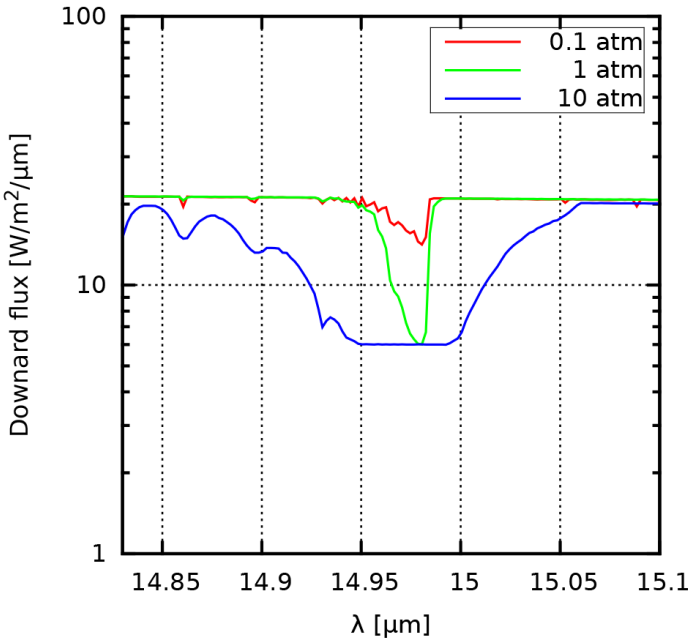


Fig. 6.2: The 15 μm CO_2 band at different atmospheric pressures.

levels, presuming to be the same for all atmospheric layers: in other words, for a given spectral interval $[\nu_1, \nu_2]$, the monotonic ordering by strength of absorption coefficients retains the relative spectral alignment of absorption lines among different levels Lacis et al. (ibid.). However, the presence of different absorbers on the line of sight does not affect the resulting fluxes by more of the 0.2% [Fu et al., 1992].

For a homogeneous light path (where the monochromatic absorption coefficient k_ν remains constant) of length z within a spectral interval $[\nu_1, \nu_2]$, the spectrally integrated transmission function can be written as:

$$T(z) = \frac{1}{\Delta\nu} \int_{\Delta\nu} e^{-k_\nu z} d\nu. \tag{6.8}$$

An absorption spectrum such as the one depicted in the left panel of Fig. 6.3 is subdivided into a number of subintervals (generally $N \sim 10^1 - 10^2$ per decade of absorption coefficient strength) of specified width Δk_i, while the frequency band $\Delta\nu$ is subdivided into $M \sim 10^3 - 10^5$ intervals of width $\Delta\nu_j$, for which it is possible to establish the corresponding Δk_i. Therefore, the probability distribution function is:

$$f(k_i) = \frac{1}{\Delta\nu} \sum_j^M \frac{\Delta\nu_j}{\Delta k_i} W(k_i, k_i + \Delta k_i) \tag{6.9}$$

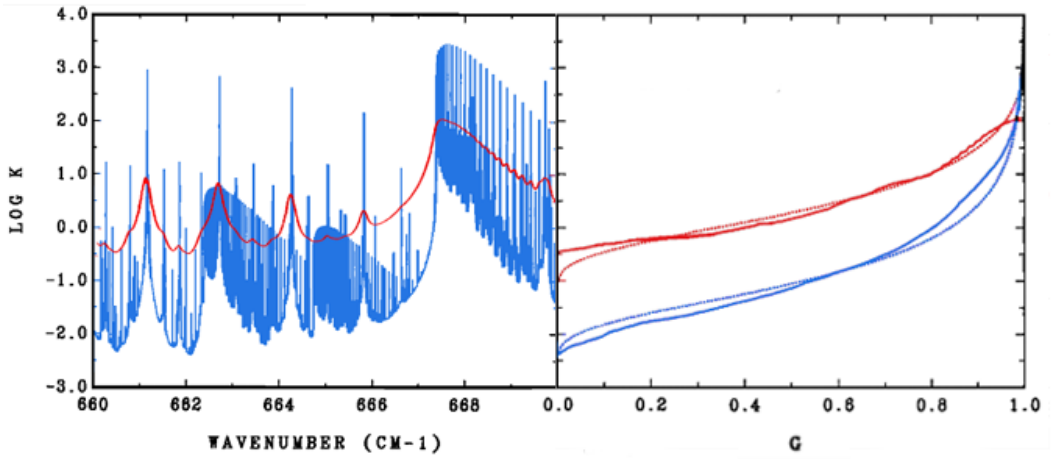


Fig. 6.3: Absorption coefficient spectrum for the 660-670 cm^{-1} portion of the 15 μm CO_2 band [Lacis et al., 1991]. *Left:* Line-by-line absorption coefficient for a pressure of 10 *mbar* and temperature of 240 K interval and correspondent (cyan), and for 1 *bar* and 296 K (red). *Right:* k distributions for both cases (solid lines); Malkmus band model equivalents obtained to provide best fit to line-by-line transmission (dashed lines).

Where $W(k_i, k_i + \Delta k_i)$ is a window function, equal to zero everywhere except in the interval between k_i and $k_i + \Delta k_i$, where it is equal to 1. Integrating $f(k)$ over k, the cumulative probability function takes the form:

$$g(k) = \int f(k)dk. \tag{6.10}$$

The inverse function $k(g)$ is the k distribution: it is a pseudo-absorption line profile that has the same nongray transmission properties as the original absorption coefficient spectrum within the interval $\Delta\nu$.

The transmission function is then:

$$T(z) = \int_0^1 e^{-k(g)z} dg. \tag{6.11}$$

The advantage of this formulation comes with the possibility of an approximation via Gaussian quadratures, since the integral is smooth. Therefore:

$$T(z) = \sum_{i=1}^{N_P} \omega_i e^{-k_i z} \tag{6.12}$$

Where ω_i are the weights associated to the k_i values and N_P are the quadrature points. To achieve enough precision on the overall spectrum it is sufficient to choose 14 spectral bands and, within each band, 32 quadrature points (see e.g. Mischna et al. (2012)).

In the end, the transmission can be calculated iterating through a few hundreds of points instead of millions, making data storage and handling faster and simpler. Furthermore, spectra at different temperatures and pressures are now degraded at the same resolution and can be matched.

The optical depth in the $\Delta\nu$ interval is now:

$$\tau = -ln(T(z)). \tag{6.13}$$

The main advantages of the correlated k distribution are its numerical efficiency and its computational speed in modeling nongray absorption in an inhomogeneous atmosphere: for a spectral interval of arbitrary width, only least-squares parameters fitted at a given pressure and temperature are needed to describe all the spectral information.

As a first step, the retrieval of line-by-line spectra is still mandatory and necessary for a large combination of parameters (pressure, temperature, and chemical composition - see Table 6.1).

For this work, it was necessary to compute Eq. 6.7 within the interval 0.24 -1000 μm ($10 < \nu < 42000\ cm^{-1}$). The original spectrum was divided in 14 reference spectral bands (see Table 6.2), similarly to Mischna et al. (ibid.), selecting the resolution for each spectrum to one-fifth of the narrowest line width, so that the narrowest line is described by 5 points. This leads to resolutions $\approx 10^5 - 10^7$ depending on composition, pressure, temperature, and wavelength. The portion of the spectrum between 0.24 and 4.6 μm is generally termed *solar*, while the remaining part extending to the far-infrared simply *IR*.

At this stage, all the absorption spectra are substantially different, as some of them may be formed by millions of points (especially at low pressures). By calculating the correlated k distributions of any of these theoretical spectra, the amount of data to store is massively reduced. In the end, this look-up table (composed by files containing the 32 weights for the quadrature points for the 14 bands, for any combination of the input parameters) is used by `MAGRATHEA` to retrieve the absorption for each layer of the atmosphere at each time step.

Tab. 6.1: Look-up correlated k distributions table used by MAGRATHEA. In italic: the grid points that the group aims to explore in the near future.

Parameters	Value
Pressure (atm)	*1E-6, 1e-5*,1E-4, 1E-3, 2.5E-3, 1E-2, 2.5E-2, 1E-1, 2.5E-1, 1, 10, 100
Temperature (K)	100, 150, 200, 250, 300, 350, 400, 450, *500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000*
X_{CO_2} (%)	0, 1E-4, 1E-3, 1E-2, 1E-1, 1, 10, 95
X_{H_2O} (%)	0, 1E-4, 1E-3, 1E-2, 1E-1, 1, 10, 95
X_{O_2} (%)	0, 2, 20
X_{CH_4} (%)	0, *0.1, 1, 10*
X_{CO} (%)	0, *0.1, 1, 10*

Tab. 6.2: Spectral bands and corresponding identifiers.

BAND ID	wavenumber (cm^{-1})	wavelength (μm)
1	10-166	60.24-100
2	166-416	24.04-60.24
3	416-625	16.00-24.04
4	625-710	14.00-16.00
5	710-833	12.00-14.00
6	833-1250	8.00-12.00
7	1250-2222	4.5-8.00
8	2222-3087	3.24-4.5
9	3087-4030	2.48-3.24
10	4030-5370	11.86-2.48
11	5370-7651	1.30-1.86
12	7651-12500	0.80-1.30
13	12500-24960	0.40-0.80
14	24960-41666	0.24-0.40

6.3 Initial setup

The code needs to be fed with some input values from an input file which must store in a specific order the physical, chemical, and computational parameters (see Listing 6.1 for a proxy input file seized on the Earth atmosphere).

```
!INPUT FOR MAGRATHEA [Aresu-Alei-Petralia 2016-2019]
!Physical Parameters
1.                  ::: PLANET MASS [Earth Masses]
1.                 ::: PLANET RADIUS [Earth Radii]
1.                  ::: DISTANCE [AU]
5778.0              ::: STAR TEMPERATURE [K]
1                 ::: STAR RADIUS [Solar radii]
0.3                  ::: ALBEDO
0.5              ::: SOLAR FACTOR [day-night]
60.0              ::: SOLAR ZENITH ANGLE
!Atmospheric Parameters
0                   ::: CLR_MODE [0] Physical [1] Fixed
0                   ::: If fixed specify [K/km]
1                   ::: WATER MODE [0] DRY [1] WET
60.0              ::: If wet, relative humidity
1.                  ::: GROUND PRESSURE [atm]
1e-4              ::: CO2 VMR
0.20             :::    O2 VMR
0                   ::: EMISSION FACTOR
!Computational Parameters
4.0              ::: INTEGRATION TIME [hours]
0.01              ::: CONVERGENCE CRITERION
30                  ::: CONSECUTIVE CONVERGENCE
1                   ::: CONVECTIVE CORRECTION [0] disabled [1] enabled
1                   ::: WATER VAPOR UPDATE [0] fixed [1] variable
```

Listing 6.1: Sample of an input file read by `MAGRATHEA`: Earth case.

- *Physical Parameters*: these include the mass of the planet in Earth masses, the radius of the planet in Earth radii, the distance from the star in AU, the temperature in K of the star and the stellar radius in solar radii. The albedo of the planet (from 0 to 1) must be defined, as well as the solar zenith angle in degrees (i.e. the angle between the incoming photons irradiation and the zenith). The day/night factor determines the average irradiation on the atmosphere caused by rotation: a value 0.5 means that the column is irradiated for half of the rotation period; for tidally locked planets, the 1D model would

be able to reproduce an atmospheric column on the dayside with a value equal to 1.

- *Atmospheric Parameters*: the first two items in this category deal with the calculation of the critical lapse rate for the convective adjustment: this value can be calculated within the code (by setting 0 in CLR_MODE), or set as fixed (CLR_MODE=1): in the latter case, the fixed lapse rate must be specified in the following row. The presence or absence of water vapor must be specified in WATER MODE (set as 0 for a dry atmosphere, 1 for a wet one): if H_2O is present in the atmosphere, the relative humidity at the ground level must be specified on the following row. The volume mixing ratios of CO_2 and O_2 must be specified as well. The emission factor should be a value from 0 to 1 that takes into account the emissivity of the top layer of the atmosphere (see Section 3.4.3.5).

- *Computational parameters*: these include the integration time in hours (i.e. the time span at which the temperature variation must be calculated at each iteration), the convergence criterion (i.e. the relative variation of the average energy balance of the atmosphere from one iteration to the following), the consecutive convergence value (i.e. the consecutive iterations for which the convergence criterion must be satisfied). It is also possible to disable the convective correction, to take into account the pure radiative equilibrium. In this case, the CONVECTIVE CORRECTION switch must be set to 0; otherwise, the default value 1 will consider both radiative transfer and convection. The variation of water vapor mixing ratio according to temperature can be disabled by changing the WATER VAPOR UPDATE switch.

From these input values, the code retrieves the starting temperature at which all atmospheric layers must be set. We chose it to be equal to the surface temperature of a leaky greenhouse model (see Section 3.4.3.5):

$$T_s = [1/(2-\epsilon)]^{1/4} T_{eq} = [1/(2-\epsilon)]^{1/4} T_\star (1-A_B)^{1/4} \sqrt{R_\star / 2d_P} \qquad (6.14)$$

Box 6.3 - Notation: layers and boundary layers

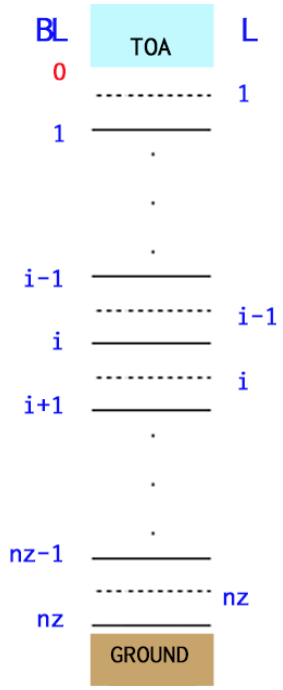


Fig. 6.4: Notation used in the code.

In the description of the code, as well as in its infrastructure, there is a difference between the terms "boundary layer" (BL) and "layer" (L).

The proxy atmosphere used by `MAGRATHEA` (see Fig. 6.4 for clarity) is composed by nz layers (at the time of writing $nz = 100$), which are in facts portions of atmosphere enclosed between two boundary layers.

For each *BL* the quantities defined by a `MAGRATHEA` run are the height z, the pressure p, the temperature T, the chemical abundances, the upward flux F_u and the downward flux F_d.

On each layer L, the code defines the averaged specific heats for all species c_p, the critical lapse rates Γ, the scattering parameters ω, the optical depths τ, and the heating rates H. These are generally performed by applying a weighted mean over pressure on the values defined in the *BL*s.

To retrieve the averaged values of the very first layer, the values at the "top of the atmosphere" boundary layer (TOA - in red) must be retrieved first. This is done by interpolation, as mentioned in Section 6.4.

`MAGRATHEA` reads another file as well, which constitutes the chosen pressure grid template, i.e. the points (*boundary layers*, see Box 6.3 for notation conventions), at which the atmospheric temperature/chemical composition must be evaluated in time. The template is scaled on a standard atmosphere whose ground pressure is 10300 Pa = 1 atm; the 100 boundary layers are evenly log-spaced.

We chose to let the code read this second input file rather than constructing the arrays in an hard-coded way to ensure the possibility to change the structure of the sampling grid, as well as to allow the modification of the distribution of the chemical species in the atmosphere, or to introduce an already determined temperature profile. In this case, we instructed the code to read the pressure profile only, while replacing the temperature, CO_2, O_2, and N_2 profiles with the values specified in the first input file. However, this step can be easily customized due to the possibility to modify the grid template.

At this point, the atmosphere is dry and composed only by carbon dioxide, oxygen and nitrogen, whose abundance is retrieved as $X_{N_2} = 1 - X_{CO_2} - X_{O_2}$: in other words, its abundance is the one that allows the mixture to have a total volume mixing ratio of 1.

The partial pressures of these species are then retrieved by multiplying the volume mixing ratio by the input pressure; these arrays are always fixed during the run.

In the occurrence of a forced interruption of the run, the code is able to restart from the last saved atmosphere, by reading back all input parameters and useful variables, periodically stored in a `.txt` file: the presence of this service file is essential since some quantities calculated in an iteration i need to be compared to the ones relative to the iteration $i - 1$ (see e.g. Section 6.7).

The atmosphere is periodically stored as well, since both temperature and water vapor profile may vary in time. Thus, in the case of a restart of the same model, the atmosphere is left unchanged from the read value – since it becomes the starting point of a new run. This approach allowed to save much time, avoiding the interrupted models to start from the very beginning. This feature has proven especially useful when dealing with a large number of models running and an external queuing system.

6.4 Atmospheric Chemistry Update

On every iteration, it is indeed necessary to calculate the layer-averaged atmospheric parameters (such as the critical lapse rate, and the specific heat at constant pressure for all species - all of them extremely temperature-dependent) to be able to approach the subsequent steps coherently.

The water vapor abundance needs to be updated as well if the atmosphere is supposed to be moist by input. At present, the relative humidity at the ground boundary layer is the input value that allows `MAGRATHEA` to estimate the H_2O abundance throughout the boundary layers of the atmosphere at all iterations: no time-dependency (i.e. increasing relative humidity in time) is included in the software at the time of writing. The actual abundance is however dependent on the temperature, thus it is most likely to have a variation in the volume mixing ratio of water vapor throughout the iterations, at the expense of nitrogen abundance (which changes accordingly, to keep the sum of all mixing ratios equal to 1). The increase

in water vapor abundance is, at any point in time, the one that ensures the relative humidity profile at all levels to be constant.

The humidity value $h(i)$ at every boundary layer $i = 1...nz$ is defined as (following the approach proposed by Manabe and Wetherald (1967)):

$$h(i) = RH \frac{Q(i) - 0.02}{1.00 - 0.02} \tag{6.15}$$

Where the factor RH is the user-provided relative humidity at the surface. For this work, we used a standard value of the relative humidity $RH = 60\%$.

The array $Q(i) = \frac{P_0(i)}{P_0(nz)}$ for $i = 1...nz$ is the ratio between the input (dry) global pressure of the layer and the input (dry) ground pressure, fixed for the entire run. Any value of $Q(i)$ lower than 0.02 (higher altitudes) is forced to be equal to 0.02 itself, thus resulting in dry layers at low pressures.

The partial pressure of water vapor $P_{\mathrm{H_2O}}(i)$ at every boundary layer $i = 1...nz$ is then retrieved according to the Arden Buck equations [Buck, 1981] for the saturation vapor pressure of moist air:

$$\begin{aligned} P_{H_2O}(i) &= h(i)\ P_{sat}(i)\ = \\ &= h(i) \cdot a \exp\left[\left(b - \frac{\widetilde{T}(i)}{c}\right)\left(\frac{\widetilde{T}(i)}{\widetilde{T}(i) + d}\right)\right] \end{aligned} \tag{6.16}$$

Where $\widetilde{T}(i) = T(i) - 273.15$ is the temperature in Celsius degrees and the constants a, b, c, and d are temperature dependent (see Table 6.3), if the vapor pressure is calculated over water or ice.

Tab. 6.3: Constants used to retrieve the saturation vapor pressure of water vapor according to Buck (1981).

$\widetilde{T}$ (°C)	a	b	c	d
$\widetilde{T} < 0\ °C\ (ice)$	6.1115	23.036	333.7	279.82
$\widetilde{T} > 0\ °C\ (water)$	6.1121	18.678	234.5	257.14

If the code runs in fixed chemistry mode, the partial pressure of the water vapor P_{wat} is retrieved when initializing the run and then held fixed; if the code runs in variable chemistry mode, the partial pressure is calculated at each iteration, and the

abundance of water is therefore simply determined by the local temperature at each boundary layer for each iteration.

If the temperature increase is such that the water vapor calculation leads to negative (nonphysical) values of molecular nitrogen in at least one boundary layer of the atmosphere, the code is forced to stop: the modeled atmosphere is not stable in those conditions.

To perform the convective adjustment at a later stage, the average values of the specific heats, as well as the layer-averaged critical lapse rate, are needed. In order to do that, all values of temperature and chemical abundances need to be evaluated in the layers, rather than the boundary layers.

In general, the mean of any variable x for a given layer i is weighted over the pressure values at the respective boundary layers i (lower boundary at higher pressure) and $i-1$ (higher boundary at lower pressure):

$$x_L(i) = \frac{x_{BL}(i) \cdot P(i) + x_{BL}(i-1) \cdot P(i-1)}{P(i) + P(i-1)} \tag{6.17}$$

For what concerns the very first layer, the code needs to retrieve the values at the top of the atmosphere, which constitute the boundary conditions. MAGRATHEA assumes the values of temperature and chemical abundances to be the same as the ones for the first boundary layer, while the pressure is calculated by subtracting the thickness of the second layer (i.e. $\Delta P = P(2) - P(1)$) to the pressure at the first boundary layer $P(1)$.

Then, the quantities in the first level can be calculated as usual.

The isochoric and isobaric specific heats $c_{V,Y}(i)$ and $c_{P,Y}(i)$ for all species Y and for all layers $i = 1...nz$ of the atmosphere are calculated interpolating the temperature values averaged on the layers $T_L(i)$ with empirical values, downloaded from the Engineering ToolBox website.

The global specific heats $C_P(i)$ and $C_V(i)$ for every layer $i = 1...nz$, are thus calculated:

$$C_{V/P}(i) = \sum_Y c_{V/P,Y}(i) \cdot X_{Y,L}(i) \tag{6.18}$$

Where $X_{Y,L}(i)$ is the volume mixing ratio of a species Y (CO_2, H_2O, O_2, or N_2) averaged on the layer i, while $c_{V/P,Y}(i)$ the corresponding isochoric/isobaric specific

heats. Then, the code calculates the gas constant for each layer i in the atmosphere:

$$R(i) = C_P(i) - C_V(i) \tag{6.19}$$

And the mean molecular weight for each layer i in the atmosphere:

$$\mu(i) = \sum_Y m_Y \cdot X_{Y,L}(i) \tag{6.20}$$

Where m_Y is the molecular weight of the species Y. The same quantities are calculated also for the respective *dry* atmosphere, taking into account all species except water vapor.

The code then retrieves the mean values of temperature and partial pressures, which are used to evaluate once again the specific heats at constant pressure (this time, averaged over the whole atmosphere), which will be used only to retrieve an estimate of the scale height of the atmosphere (see Eq. 3.5), and thus the height of the atmosphere by using the hydrostatic equation (see Eq.3.7).

At this point, the routine has all useful variables to calculate the critical lapse rate in each layer of the atmosphere.

If the case of a physical calculation of the critical lapse rate, the following operations are performed; this is, of course, skipped when running a model with fixed, user-provided critical lapse rate.

In the first scenario, both dry and moist adiabatic lapse rates are retrieved ($\Gamma_d(i)$ – see Eq. 3.61, and $\Gamma_m(i)$ – see Eq. 3.65).

The actual critical lapse rate of each layer $\Gamma_{ad}(i)$, which will be used in the iteration, is a result of a balance between the stabilizing effect of upward heat in moist and dry convection and the destabilizing effect of radiative transfer. It is supposed to be a linear combination weighted with the humidity $h_L(i)$ defined in the layer:

$$\Gamma_{ad}(i) = (1 - h_L(i)) \cdot \Gamma_d(i) + h_L(i) \cdot \Gamma_m(i) \tag{6.21}$$

This is a plausible approach: in the Earth's case (for which the theoretical values of the dry and moist lapse rate are $\Gamma_d \approx 9.8\ K\ km^{-1}$ and $\Gamma_m \approx 5\ K\ km^{-1}$) the average lapse rate retrieved with Eq. 6.21 is $\Gamma_d \approx 6.9\ K\ km^{-1}$, very similar to the observed

Earth tropospheric critical lapse rate of temperature is $\Gamma_{ad} \approx 6.5\ K\ km^{-1}$. The small difference is due to the variation of the relative humidity on the surface of the Earth in time, due to climatic processes.

6.5 Radiative Transfer

The radiative subroutine can calculate up and down fluxes at each atmospheric layer using absorption spectra previously adapted via the *k* distribution method (see Section 6.2).

The *k* distributed weights are read from the look-up table and interpolated, according to the required values of pressure, temperature, and chemical composition. The optical depth for each band in each layer of the atmosphere is calculated with Eq. 6.12 and Eq. 6.13, as well as the albedo (which will be considered in the visible wavelength range of the absorption spectrum), and the wavelength-dependent Rayleigh scattering absorption coefficients.

Such values, as well as the temperature profile and the user-provided emissivity of the top boundary, are given in input to DISORT4 module.

DISORT (DIScrete Ordinate Radiative Transfer) [Stamnes et al., 1988] considers the transfer of monochromatic unpolarized radiation in a vertically inhomogeneous, nonisothermal, plane-parallel medium. It includes physical processes such as thermal emission, scattering, absorption, and bidirectional reflection and emission at the lower boundary. It can be customized easily, by changing the input settings to allow the user to reproduce a wide range of astrophysical problems.

This software was the first to numerically handle the discrete ordinate method for radiative transfer, that is a discretization of the integral of 2N Legendre polynomials that compose the scattering phase function of the source term of a beam incident from a specific direction, made with the Gaussian quadrature rule.

In short terms, DISORT solves for every quadrature point in every band of the spectrum and every layer in the atmosphere a set of 2N coupled differential equations, thus taking into account integrated values over a range of wavelengths over which the medium properties are invariant (see Stamnes et al. (ibid.) for further information).

DISORT retrieves, for every boundary layer and every quadrature point of every band, the direct and diffuse downward intensity $F_d(i, j, k)$ and the upward diffuse

intensity $F_u(i,j,k)$ caused by the interaction of the light beam with the gaseous mixture at the boundary layer.

To retrieve the total values of F_d and F_u for every layer i (where i iterates over the 101 boundary layers including the TOA), it is necessary to iterate j over the 14 spectral bands, and k over each of the 32 quadrature points, using the same quadrature weights ω_k:

$$F_{d/u,i} := F_{d/u}(i) = \sum_{j=1}^{14} \sum_{k=1}^{32} \omega_k F_{d/u}(i,j,k) \tag{6.22}$$

The heating rates for every layer are then calculated from the fluxes difference at each layer i, by use of the following equation:

$$H_i = \left(\frac{dT}{dt}\right)_i^{(heating)} - \left(\frac{dT}{dt}\right)_i^{(cooling)} = \frac{g}{c_P(i)} \frac{(F_{d,i} + F_{u,i+1}) - (F_{u,i} + F_{d,i+1})}{p_{i+1} - p_i} \tag{6.23}$$

where g is the surface gravity, and $c_P(i)$ the isobaric specific heat for the layer i. The pressure p, the upward fluxes F_u, and downward fluxes F_d are evaluated at $(i-1)$-th and i-th boundary layers. Therefore, the temperature variation in time of a layer i caused purely by radiative transfer is:

$$\left(\frac{\partial T_i}{\partial t}\right)_{RAD} = \frac{p_{i-1} H_{i-1} + p_i H_i}{p_{i-1} + p_i} \tag{6.24}$$

At each time step $\tau + \Delta t$, the temperature at each layer i is then updated:

$$T_{i,\tau+\Delta t} = T_i + \left(\frac{\partial T_i}{\partial t}\right)_{RAD,\tau} \Delta t \tag{6.25}$$

Where T_τ is the temperature at the τ-th time step, and Δt the time interval, $(\partial T_i/\partial t)_{RAD,\tau}$ is the rate of change of T_i at that particular time step.

Pure radiative equilibrium is considered to have been reached when $T_{\tau+1} - T_\tau$ at each layer is less than a small value δ.

6.6 Convective Adjustment

To better simulate a terrestrial atmosphere, the convective transport needs to be taken into account.

Manabe and Strickler (1964) introduced the simple process of a *convective adjustment* to approximate the upward heat transfer by atmospheric motions. The base concept is really simple: whenever the lapse rate of temperature exceeds a critical value while approaching the final steady-state, it is forced to assume the critical value itself. This simulates convective heat transfer from the planet's surface into the lower and upper troposphere, and thereby allows to retrieve more realistic temperatures.

It is now possible to include a convective adjustment, to set the lapse rate to the critical lapse rate whenever it exceeds its value in the course of the numerical integration.

The requirements for *thermal equilibrium* are now:

- At the top of the atmosphere, the net incoming solar radiation has to be equal to the net outgoing longwave radiation;
- At the surface, the excess of net downward solar radiation over net upward longwave radiation has to be equal to the net integrated radiative cooling of the atmosphere;
- Whenever the lapse rate is subcritical, local radiative equilibrium has to be satisfied.

Following the approach described in Manabe and Wetherald (1967), the following procedures are executed at each time step:

1. Set the temperature on the nz-th layer is initially as the same temperature of the previous time-step: $T_{nz}^{(0)} = T_{nz}$.
2. Compute the radiative equilibrium temperature T_{eq} for the ground such that it satisfies:
$$\sigma_B T_{eq}^4 = F_{d,nz} \tag{6.26}$$
where σ_B is the Stefan-Boltzmann constant.
3. Compute $T_{nz}^{(1)}$ at the ground layer so that it satisfies the relationship:
$$\frac{c_P(p_{nz} - p_{nz-1})}{g}(T_{nz}^{(1)} - T_{nz}^{(0)}) = \sigma_R \Delta t[T_{eq}^4 - (T_{nz}^{(1)})^4] \tag{6.27}$$

4. If $T_{nz}^{(1)} - T_{nz-1}^{(0)} > \Gamma_{nz}(z_{nz-1} - z_{nz})$ where $\Gamma_{nz} = \Gamma_{ad}(nz)$ is the critical lapse rate of the ground layer nz (which is in direct contact with the ground), this layer is unstable for convection. So, two new values of T_{nz} and T_{nz-1} have to be calculated: those will be called respectively $T_{nz}^{(2)}$ and $T_{nz-1}^{(1)}$ and need to satisfy the relationships:

$$T_{nz}^{(2)} - T_{nz-1}^{(1)} = \Gamma_{nz}(z_{nz-1} - z_{nz}) \tag{6.28}$$

$$\frac{c_P}{g}[(p_{nz} - p_{nz-1})(T_{nz}^{(2)} - T_{nz}^{(1)}) + (p_{nz-1} - p_{nz-2})(T_{nz-1}^{(1)} - T_{nz-1}^{(0)})] = \\ = \sigma_R \Delta t[(T_{nz}^{(1)})^4 - (T_{nz}^{(2)})^4]. \tag{6.29}$$

   If the layer is stable for convection (the lapse rate is lower than the critical lapse rate), then T_{nz} and T_{nz-1} remain unchanged (so $T_{nz}^{(2)} = T_{nz}^{(1)}$ and $T_{nz-1}^{(1)} = T_{nz-1}^{(0)}$).

5. For $i = nz - 1, \ldots, 2$, only if $T_i^{(1)} - T_{i-1}^{(0)} > \Gamma_i(z_{i-1} - z_i)$ (unstable layer), compute $T_i^{(2)}$ and $T_{i-1}^{(1)}$ such that they satisfy the following relations:

$$T_i^{(2)} - T_{i-1}^{(1)} = \Gamma_i(z_{i-1} - z_i) \tag{6.30}$$

$$(p_i - p_{i-1})(T_i^{(2)} - T_i^{(1)}) + (p_{i-1} - p_{i-2})(T_{i-1}^{(1)} - T_{i-1}^{(0)}) = 0. \tag{6.31}$$

6. Replace every new value into the original array $T^{(0)}$ and repeat processes 4 and 5.

7. Repeat process 6 until every layer of supercritical lapse rate is completely removed.

Convective adjustment is executed after 40 purely radiative iterations so that the modeled atmosphere can assess the biggest variations caused by the radiative process before starting to define the contribution of convection.

On the other hand, adjustment is not performed at all if the user is interested in purely radiative equilibrium (so CONVECTIVE CORRECTION in the input file is set to 0).

6.7 Convergence Criterion

An atmosphere is considered to be at equilibrium if the net upward flux of longwave radiation is equal to the net downward flux of stellar radiation both at the surface and at the ground level: those conditions are important to guarantee the radiative equilibrium of both the planet's surface and atmosphere.

At each iteration τ, the energy flux density is related to the upward and downward fluxes in the visible and infrared range. The net energy flux densities can be calculated at the top of the atmosphere and at the ground boundary layer:

$$E_{\tau,0} = |F^{tot}_{\downarrow,0} - F^{tot}_{\uparrow,0}| = |(F^{VIS}_{\downarrow,0} - F^{VIS}_{\uparrow,0}) - (F^{IR}_{\uparrow,0} - F^{IR}_{\downarrow,0})| \tag{6.32}$$

$$E_{\tau,nz} = |F^{tot}_{\downarrow,nz} - F^{tot}_{\uparrow,nz}| = |(F^{VIS}_{\downarrow,nz} - F^{VIS}_{\uparrow,nz}) - (F^{IR}_{\uparrow,nz} - F^{IR}_{\downarrow,nz})| \tag{6.33}$$

The averaged energy budget of the whole atmosphere E_τ at iteration τ is then the difference between these two quantities:

$$E_\tau = (E_{\tau,0} - E_{\tau,nz}) \tag{6.34}$$

The mean of the various E_τ is then calculated over all the elapsed iterations n_{it}:

$$\mu_{n_{it}} = \frac{\Sigma_{\tau=1}^{n_{it}} E_\tau}{n_{it}} \tag{6.35}$$

The relative variation of μ from one iteration to the following is then:

$$\mu = \frac{\mu_{n_{it}} - \mu_{n_{it}-1}}{\mu_{n_{it}-1}} \tag{6.36}$$

where $\mu_{n_{it}-1} = \frac{\Sigma_{\tau=1}^{n_{it}-1} E_\tau}{n_{it}-1}$ is the mean of the first $n_{it} - 1$ iterations, not considering the current one.

The model is supposed to have reached convergence if μ is lower than a user-provided threshold δ for a number N of subsequent days (user-provided information). For all the following simulations, we set $d = 0.01$ and $N = 30$.

6.8 Error handling

MAGRATHEA is forced to abort execution whenever it reaches either unstable atmospheres or temperatures hotter than 500 K in at least one boundary layer.

Unstable Atmospheres. The profile is unstable (within the assumptions considered in MAGRATHEA) if the abundance of N_2 is supposed to be negative - an unrealistic value. This can happen whenever the abundance of water vapor (calculated at each iteration based on the current temperature of each layer) exceeds the maximum value prescribed by the remaining components of the atmosphere: since the sum of all volume mixing ratios of the species in the mixture must be 1, the maximum abundance of water vapor for each layer should be equal to $1 - X_{CO_2}$, which would lead to an absence of nitrogen in the mixture. Since this last compound is calculated as filling gas, higher values of X_{H_2O} would lead to negative nitrogen abundances, which would not agree with the mass conservation principle. This error happens because, at the time of writing, no complete water cycle is accounted for: the water vapor present in the atmosphere is not able to condense in clouds or precipitate. The main assumption is therefore that there is a surface infinite water reservoir and that water abundance continues to grow up to its maximum value, determined by the abundance of carbon dioxide, at the expense of nitrogen, the filling gas. In other words, if a model fails because of the negative abundance of nitrogen, the atmosphere cannot be stable with the prescribed pressures, temperatures, and chemical composition.

Exceedingly hot temperatures. With the current MAGRATHEA configuration, it is not possible to accurately model atmospheres whose temperature exceeds 500 K in any point of their pressure profile: this threshold is the last available point if the look-up k distribution table, over which the absorption coefficients would be extrapolated, forbidding the model to provide an accurate treatment of the radiative transfer. For such models, MAGRATHEA is not able to provide further information concerning the stability and the structure of the profiles. The failed models that belong to this category can, however, be studied with a larger grid of k-correlated distributions, an upgrade that is already scheduled in the near future.

6.9 Validation

To validate the radiative transfer subroutine, we considered an hypothetical early Mars atmosphere distributed in $nz = 100$ boundary layers, with an input temperature profile that decreases rather linearly with respect to height from a ground temperature $T_{nz} = 250$ K to ~ 170 K at a pressure of 100 mbar, and then it is let constant (see Mischna et al. (2012)). At the ground level, the surface is set to be $P_{nz} = 500$ mbar.

The atmosphere is composed of 95% CO_2, $\approx$ 5% N_2, with trace amounts of water vapor. The water abundance follows closely the temperature profile: the water volume mixing ratio is $X_{H_2O} = 1.5 \times 10^{-3}$ at the ground, and rises linearly up to 10^{-7} at 100 mbar, where it remains constant for lower pressures.

We set the bond albedo to 0 and we consider an irradiation perpendicular to the horizontal plane, i.e. $\cos(\varphi) = 1$, φ being the irradiation zenith angle. The incoming irradiation is set to 220 $W\ m^{-2}$, i.e. the expected illumination of a young Mars that takes into account a day/night factor of 0.5.

To validate the radiative routine we compare our results with those obtained using a similar 1D $k-$distribution RT model developed at NASA Ames [Toon et al., 1989]. We also incorporate the absorption coefficients calculated from Mischna et al. (2012) into our radiative transfer module and show the resulting outcomes in Figure 6.5.

The different cross-sections and the different radiative transfer modules do not introduce important differences in the downward fluxes. There is however a $\sim 2.5\ W\ m^{-2}$ difference in the upward solar flux, where Rayleigh scattering dominates: this is due to the different resolutions at which calculations are performed by the different models. Contrary to our model, in both NASA Ames and Mischna et al. (ibid.) model opacities the Rayleigh scattering cross-section in a given spectral range is calculated at the center of the band. Because the Rayleigh scattering depends steeply on wavelengths, low-resolution calculations tend to underestimate the cross-sections.

The coupling of radiative transfer and convective adjustment routines can be now validated. We reproduced a present-day Mars-like atmosphere (see Table 6.4 for some calculation parameters).

The atmosphere is composed of a mixture of CO_2 (95%) and N_2 (5%), with no water vapor. The surface pressure is set to $P_{nz} = 6.72$ mbar. We ran four models taking all combinations of two different solar zenith angles (20 or 45 degrees) and physical or fixed (2.5 $K\ km^{-1}$) critical lapse rate.

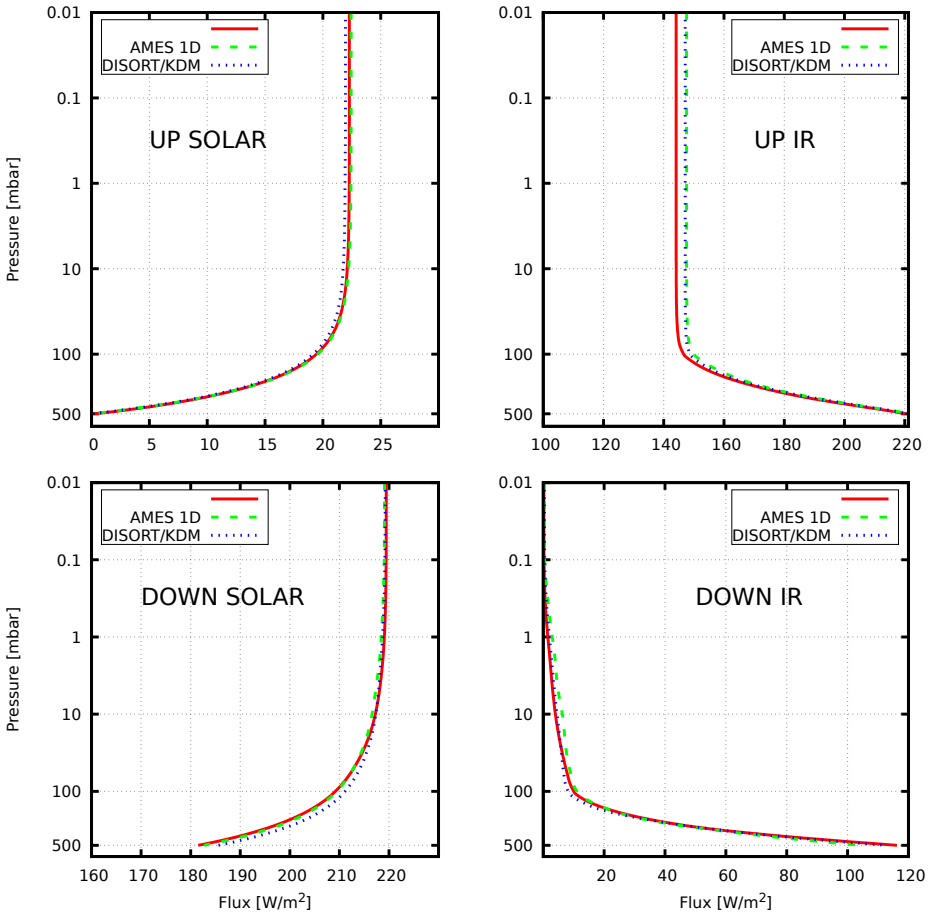


Fig. 6.5: Upward and downward fluxes in the solar and IR bands. Solid red lines: `MAGRATHEA` radiative transfer subroutine; dashed green lines: NASA AMES 1D model; dotted blue lines: `MAGRATHEA` radiative transfer subroutine using Mischna et al. (2012) absorption coefficients.

Tab. 6.4: Calculation parameters for the present-day Mars simulation.

Parameter		Value
Stellar Temperature	$T_\star$ (K)	5778
Stellar Radius	$R_\star$ $(R_\odot)$	1.0
Stellar Irradiation	$S_\star$ (W m^{-2})	586
Planet Mass	M_P $(M_\oplus)$	0.107
Planet Radius	R_P $(R_\oplus)$	0.531
Bond Albedo	A_B	0.1
Orbital distance	d_P (AU)	1.523
Day/night factor	d/n	0.5
Zenith Angle	φ (degrees)	20 - 45
Emissivity	ϵ	0

The necessity to set the critical lapse rate fixed in some runs is due to the fact that the calculated one (using Eq. 6.21) would be $\approx 5\ K\ km^{-1}$, while the observed one is on average 2.5 $K\ km^{-1}$. This discrepancy is produced by heating processes caused by the suspended dust particle absorption, as well as circulation phenomena in the martian atmosphere, which lower the temperature variation with respect to height.

In those runs where the critical lapse rate is calculated at each iteration, its value evolves through time due to the temperature dependence of the atmospheric isobaric specific heat on temperature. Initial and final critical lapse rate profiles are shown in Figure 6.6.

The models start from an initial isothermal profile at the temperature of 180 K. Using a Solar Zenith Angle (SZA) $\varphi = 45°$ (approximately the *Viking-2* landing site latitude) and a fixed critical lapse rate, we find results consistent with data acquired during the descent of some of the major martian landers (see Figure 6.7 – dashed blue line compared with the dots). Some differences are indeed present, mostly in the upper part of the profile: this is caused by photochemistry, condensation, and circulation mechanisms that are not taken into account in `MAGRATHEA`, at present – processes which cause a high variability in the composition of the atmosphere above 80 km. However, the tropospheric trend of the profile is in agreement with the observed data.

The profiles obtained using a calculated critical lapse rate show a slight change (a few degrees) in the ground pressure, as well as a change in the inclination of the tropospheric section, which translates into cooler stratospheres with respect to the fixed lapse rate situation. This is a straightforward consequence of the change in critical lapse rate according to which the convective adjustment is activated.

The profiles obtained using $\varphi = 20°$, similar to *Pathfinder* and *Viking-1* landing sites' latitudes, are also fairly in agreement with the data. There is a larger discrepancy in the tropospheric profile compared to the previous models (about 20 K at the surface), due to the higher irradiation of the latter cases.

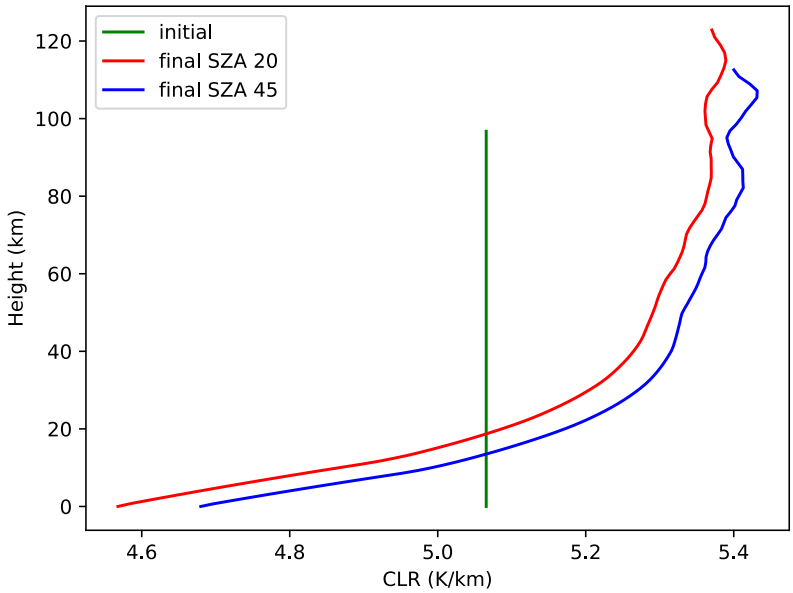


Fig. 6.6: CLR as a function of the altitude. Initial value: green line; final profiles: $\varphi = 20°$, red line; $\varphi = 45°$, blue line.

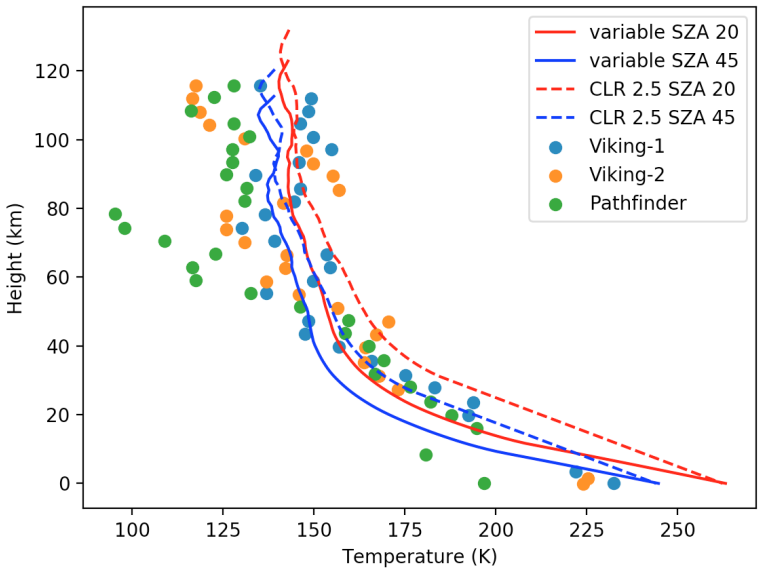


Fig. 6.7: Synthetic (solid and dashed lines) and retrieved (dots) thermal profile of Mars. Red ($\varphi = 20°$) and blue ($\varphi = 45°$) lines: solid, variable critical lapse rate from Eq. (6.21); dashed, constant critical lapse rate, $\Gamma_c = 2.5\ K\ km^{-1}$. Blue dots: *Viking-1*; orange dots: *Viking-2*; green dots: *Pathfinder*. Data from *Viking-1* and *Viking-2* are taken in Nier et al. (1976), while the ones from *Pathfinder* in Magalhães et al. (1999).

7 MAGRATHEA Runs

> *Looking back, we were the luckiest people in the world. There was no choice but to be pioneers; no time to be beginners. When answers could not be found, we had to invent them.*
>
> **— Margaret Hamilton (1936-present)**
> American computer scientist and systems engineer, developer of NASA's Apollo on-flight software

With MAGRATHEA, two grids of models were run. The first grid, composed by about 18000 models, spans a range of physical and chemical parameters of plausible, yet imaginary, atmospheres of terrestrial planets orbiting at various distances different main-sequence stars. The second grid, composed by about 2400 models, is based on the true available data on planetary mass and radius, semi-major axis, and stellar effective temperature as retrieved from the known Super Earths sample previously queried in Exo-MerCat (see Chapter 5). Both of these grids will be described and analyzed in this chapter. Some of the successful pressure-temperature profiles of the second grid will be then used for further simulations (using the ozone photochemistry software program, as explained in Chapter 8) and for the laboratory experiments within the "Atmosphere in a Test-Tube" project (described in Chapter 9).

7.1 Aim

The key aim of this study is to assess the contribution of the various physical and chemical parameters/variables to the stability of temperate terrestrial planets, with a particular focus on habitability on Super Earths. To gain such information qualitatively, a large number of models needs to be run and a parameter space as wide as possible (albeit plausible). Later on, other choices of the input parameters could be made, possibly disregarding the ones that were considered not as decisive as others, or exploring subsets of parameters in a more detailed way.

During the whole duration of the project, many simulations were run to test the various features that were gradually implemented in MAGRATHEA: about 100000 models were executed in three years, exploring a wide range of parameters and methods to include more and more processes. In time, the code has evolved from its serial configuration to an MPI-parallelized version, that allows running multiple models at the same time, to fill large grids of models in a few weeks.

The optimization of the code and its conversion to FORTRAN 90 helped reduce the computing time of a single model, which now reaches convergence on an average time of 10 hours – rather than days, as in the case of the very first MAGRATHEA prototype.

In the remainder of the chapter, we chose to focus on the very last grids of models, the ones that were reproduced with the current, more updated, software version.

The following sections will describe both the theoretical grid of Super Earths (Section 7.2) and the observed sample of Super Earths (Section 7.3). Every section will describe the input retrieval, the methods, and the analysis of the two input grids.

7.2 Theoretical Super Earths

7.2.1 Input

Amongst others, a grid of 17280 models was run at CINECA Galileo cluster, within the *Accordo Quadro INAF-CINECA (2017)* agreement. The grid was formed by all combinations of the input parameters reported in Table 7.1. The chemical database available at the time of writing is composed by CO_2, O_2, H_2O, and N_2 as filling gas.

Tab. 7.1: Input parameters of the dry grid of models. See the text for the meaning of S_{maxg}, S_{rg}, and α.

Parameters	**#**	**Notes/Values**
Planetary Mass ($M_\oplus$)	3	1, 5, 10
$T_\star$ (K)	9	$\Delta T = 500\ K$ from 3000 to 7000 K
Distances (AU)	4	From 0.8 S_{maxg} to 1.2 S_{rg}
X_{CO_2}	5	0.0001, 0.01, 0.1, 0.5, 0.9
Pressure Factors α	4	0.3,1,3,10
Albedo	2	0.1, 0.3
Emissivity	2	0, 2/3
Water vapor	2	dry, wet (RH 60%)

The input files for each model were automatically written by a dedicated Python script. The derived values were also calculated for each combination within the same software: those were the planetary radius, the ground pressure, the stellar radius, and a more specific calculation of the distance.

- PLANETARY RADIUS AND DENSITY

  The planetary radius is retrieved from the planetary mass using the input density, following the approach presented in Seager, Kuchner, et al. (2007). The dimensionless mass-radius relationship is:

$$\log_{10} R_s = k_1 + \frac{1}{3}\log_{10} M_s - k_2 M_s^{k_3} \tag{7.1}$$

  where $k_1 = -0.20945$, $k_2 = 0.0804$, and $k_3 = 0.394$ while $M_s = M_{pl}/m_1$ and $R_s = R_{pl}/r_1$.

  M_{pl} is the planetary mass provided by the user, while the constants m_1 and r_1 are related to the equation of state of the planet, which is in turn linked to its density. For our purposes, we have used the values provided by the paper for a 0.3% Fe - 0.7% $MgSiO_3$ composition: this corresponds to $m_1 = 6.41$ and $r_1 = 3.19$ Seager, Kuchner, et al. (ibid.). The code is then built to retrieve directly R_{pl} from these parameters. In Table 7.2 the corresponding radii and bulk densities for each planet are shown.

Tab. 7.2: Masses, radii, and corresponding densities.

M_{pl} ($M_\oplus$)	R_{pl} ($R_\oplus$)	**Density** (g/cm^3)
1	1.0	5.0
5	1.9	4.0
10	2.5	3.4

- GROUND PRESSURE AND ATMOSPHERIC MASS

  In order to consider plausible ground pressures when modeling theoretical exoplanets, we chose to determine the ground pressure of the atmosphere on an exoplanet basing our calculations on the ratio of the Earth's atmospheric mass to the planetary mass. Considering $M_{atm,\oplus} \approx 5.15 \cdot 10^{18}\ kg$ and $M_\oplus \approx 5.97 \cdot 10^{24}\ kg$, this ratio r is:

$$r = M_{atm,\oplus}/M_\oplus = 8.6 \cdot 10^{-7} \tag{7.2}$$

When changing the planetary mass, we imposed that the mass of the planetary atmosphere should change accordingly to keep the ratio r constant. Thus:

$$M_{atm,pl} = r \cdot M_{pl} \tag{7.3}$$

This leads to $M_{atm,5\ M_\oplus} \approx 2.57 \cdot 10^{19}\ kg$ and $M_{atm,10\ M_\oplus} \approx 5.15 \cdot 10^{19}\ kg$.

The four atmospheric mass factors (hereafter α_i) can be now considered to retrieve the desired ground pressures for each planetary mass, following this equation:

$$P_{gr,i}\ [Pa] = \alpha_i \cdot \frac{M_{atm,pl}\ [kg]\ g_{pl}\ [m\ s^{-2}]}{4\pi (R_{pl}\ [m])^2} \tag{7.4}$$

The obtained ground pressures for each planets are reported in Table 7.3 (displayed in atm units, where $1\ atm = 1.01325 \cdot 10^5\ Pa$).

Tab. 7.3: Masses, atmospheric factors, and corresponding ground pressures.

Mass ($M_\oplus$)	α	**Ground Pressure (atm)**
1	0.3	0.258
	1	0.859
	3	2.58
	10	8.59
5	0.3	0.55
	1	1.83
	3	5.50
	10	18.33
10	0.3	0.70
	1	2.36
	3	7.07
	10	23.57

- STELLAR RADIUS

To be able to calculate the blackbody radiation within MAGRATHEA and the semi-major axis at which the theoretical planets must be put, the stellar radius needs to be provided. It is retrieved theoretically from the stellar temperature $T_\star\ [K]$, following the approach proposed by Boyajian et al. (2012):

$$R\ [R_\odot] = a + b\ T_\star + c\ T_\star^2 + d\ T_\star^3 \tag{7.5}$$

where $a = -10.8828$, $b = 7.18727 \cdot 10^{3}$, $c = -1.50957 \cdot 10^{-6}$, $d = 1.07572 \cdot 10^{-10}$. The obtained radii for each stellar temperature are reported in Table 7.4.

Tab. 7.4: Input stellar temperatures and corresponding calculated radii.

$T_\star$ (K)	$R_\star$ ($R_\odot$)
3000	0.182
3500	0.391
4000	0.595
4500	0.690
5000	0.755
5500	0.873
6000	1.122
6500	1.585
7000	2.341

- DISTANCE FROM THE STAR

  A recent calculation by Kopparapu et al. (2013) retrieved some parametric inner and outer HZ boundaries around different classes of stars by calculating the effective stellar flux $S_{eff} = S/S_0$, i.e. the value of stellar constant S (see Section 3.4.2) required to maintain a given surface temperature, in units of solar constant S_0.

  The authors used a 1D radiative-convective cloud-free model with fixed temperature profile and varying abundances of water vapor (for the inner limit) and carbon dioxide (for the outer limit). Their results are shown in Figure 7.1).

  In general, the equation to be used for all boundaries is:

$$S_{eff} = S_{eff,\odot} + a(T_\star - 5778) + b(T_\star - 5778)^2 + c(T_\star - 5778)^3 + d(T_\star - 5778)^4 \quad (7.6)$$

  where $T_\star$ [K] is the stellar temperature and the various coefficients change depending on the scenario. For our purposes, we chose the runaway greenhouse limit $S_{rg} := S_{eff,rg}$ and the maximum greenhouse limits $S_{maxg} := S_{eff,maxg}$ for the inner and outer limit to the habitable zone (see Section 4.1). The corresponding values of the constants are displayed in Table 7.5.

  The code defines two points at $S_{eff} = 1.2S_{rg}$ (shortest distance - highest irradiation) and $S_{eff} = 0.8S_{maxg}$ (higher distance - lowest irradiation) for each

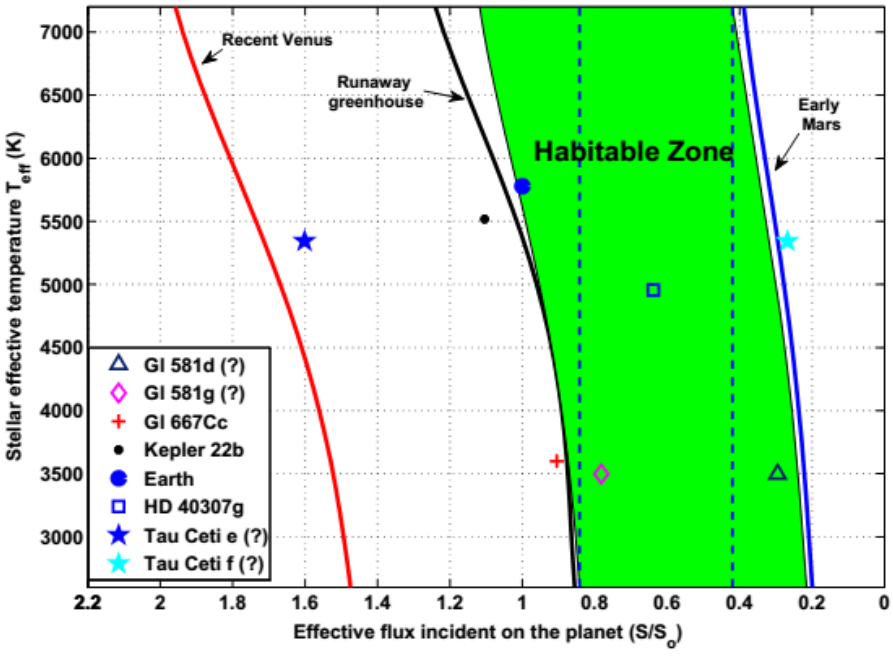


Fig. 7.1: Various cloud-free habitable zone flux boundaries for stars with different T_{eff}. The boundaries of the green-shaded region are determined by the water loss (inner edge) and maximum greenhouse (outer edge) [Kopparapu et al., 2013].

Tab. 7.5: Constants used for the retrieval of the stellar constants at runaway greenhouse and maximum greenhouse boundaries [Kopparapu et al., 2013].

Constant	Runaway g.h.	Maximum g.h.
$S_{eff,\odot}$	1.0512	0.3438
a	$1.3242 \cdot 10^{-4}$	$5.8942 \cdot 10^{-5}$
b	$1.5418 \cdot 10^{-8}$	$1.6558 \cdot 10^{-9}$
c	$-7.9895 \cdot 10^{-12}$	$-3.0045 \cdot 10^{-12}$
d	$-1.8328 \cdot 10^{-15}$	$-5.2983 \cdot 10^{-16}$

stellar temperature, and then selects other N-2 points equally spaced between the two limits (where N is the amount of different distance points desired by the user) in order to consider various distances within the habitable zone.

The distances at each effective stellar flux are then retrieved by solving the formula (where $T_\star$ is in K, and $R_\star$ is in $R_\oplus$ units):

$$d\ [AU] = \sqrt{\frac{L_\star}{S_{eff}}} = \sqrt{\frac{(T_\star/5778)^4 \cdot R_\star^2}{S_{eff}}} \tag{7.7}$$

In the present case, we aimed at modeling planets at four distances for each host star ($N = 4$), one being within the inner limit of the habitable zone, two well inside the habitable zone, and one farther from the outer limit.

By construction, only half of the models are theoretically habitable according to the limits imposed by Kopparapu et al. (2013). However, since `MAGRATHEA` can update the atmospheric temperature (and composition, to some extent), which was, on the contrary, held fixed by the authors, some slightly different results are foreseen.

Other parameters used by the code were kept constant for all models:

- The solar zenith angle was set to 60 degrees (a common value in 1D radiative transfer models, used to model a temperate, zonally averaged atmosphere);
- The solar factor was set as 0.5 (considering, therefore, the average flux for a point on a rotating planet irradiated for half of the duration of the day);
- No O_2 was present in the atmosphere;
- we chose a time interval ΔT of 4 hours, a convergence criterion of $\delta = 0.01$, and a number of consecutive iterations below the convergence criterion of $N = 30$;
- Water vapor, when available in the modeled atmosphere, is set to be updated at each iteration based on the temperature of the layer (WATER VAPOR UPDATE=1).

Considering all combinations of parameters, the script creates the directories where all models will be saved, as well as a set of input files, one for each model, stored in the parent directory `ALL_IN`.

For practical reasons, the larger grid was divided into two subgrids (dry and wet models), each one composed by 8640 models. Nine grids (960 models each) for each subgrid were then set up in the cluster and executed simultaneously.

7.2.2 Analysis and results

The theoretical grid of Super Earths was run first, to validate the code and to study its performances in a large parameter space. This allowed us to retrieve qualitative information on the various processes involved, to identify critical points and to schedule further improvements in the future.

However, the analysis of this large number of models can be challenging, since it could be hard to visualize the whole sample of results without applying some sort of filtering on the data.

Qualitative analysis can be indeed useful for an initial study on the output pressure-temperature profiles (see Section 7.2.2.1), but one cannot dismiss a more thorough analysis on the whole sample, as will be described in Section 7.2.2.2. The analysis of the failed models (Section 7.2.2.3) provides further information concerning both the validity ranges of the code and the stability of some simulated models. Then, on the successful models, some preliminary studies concerning the surface habitability can be made (Section 7.2.2.4).

7.2.2.1 Qualitative Analysis

To have a brief overview of the effect that each input parameter has on the shape of the modeled profile, we selected a subset of models to be visually displayed in Figure 7.2. To do that, the physical and chemical input parameters (planetary mass, stellar temperature, distance factor, albedo, pressure factor, emissivity, carbon dioxide volume mixing ratio, and relative humidity) were set to a precise value, when possible corresponding to a medium value with respect to the input grid. Then, for each subplot in the figure, a single parameter ranges for all its input values, while keeping the remaining ones constant (see Table 7.6 for the selected parameters' values). In this way, it is possible - albeit qualitatively - to see what direct variation on the model is generally caused by the analyzed parameter.

From Figure 7.2, it is possible to retrieve some information:

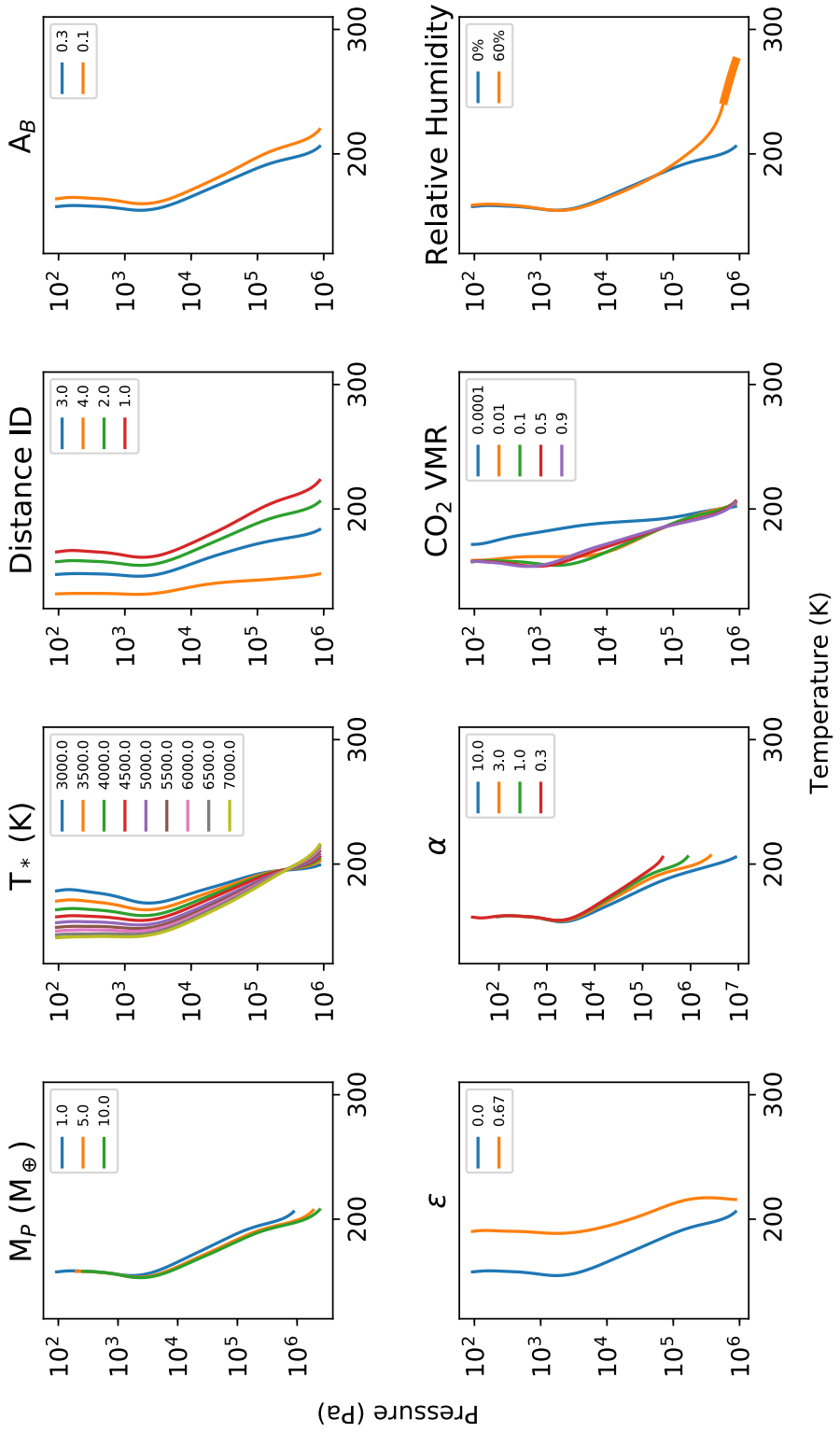


Fig. 7.2: Set of subplots with varying parameters, based on Table 7.6 (dry atmospheres).

Tab. 7.6: Subset of selected parameters for the qualitative analysis.

M_P ($M_\oplus$)	$T_\star$ (K)	a ID	A_B	ϵ	α	X_{CO_2}	RH (%)
5	4500	2	0.3	0	1	0.1	0

- *Planetary mass variation*. This parameter does not seem to cause a variation in the shape of the profile, apart from an obvious shift towards higher pressures. This is explained by the higher atmospheric mass that needs to be taken into account to keep the ratio r constant (see Eqs. 7.3 and 7.4). In particular, as seen in Table 7.3, the ground pressures for the three planets range from less than 1 atm to more than 2 atm. The height of the tropopause appears to be at the same pressure level regardless of the ground pressure; the stratospheres, as well, do not experience variation from one model to the other, but they are sampled up to variable height because of the change in pressure (by construction, the pressure difference between the ground layer and the top of the atmosphere must be the same). A slightly more prominent change in inclination is visible in the troposphere: this is due to the change in gravity, which causes the critical (dry) lapse rate to change in return.

- *Stellar Temperature variation*. The stellar temperature increase from 3000 to 7000 K contributes to change the atmospheric profile. In particular, the upper layers of the atmosphere are inversely correlated to the increase of stellar temperature, and therefore they are hotter for planets orbiting cooler stars. The trend is inverted, on the other hand, in the lower layers of the atmosphere. As a matter of fact, when varying the stellar temperature, the peak of the blackbody radiation shifts, according to the Wien's displacement law. The carbon dioxide absorbs more effectively infrared photons, subsequently heating the atmospheric layer, and therefore, for the same amount of CO_2 in the atmosphere, a planet orbiting an M star at 3000 K normally has hotter stratospheres compared to the same planet orbiting a solar-like star with similar incoming flux. Since, in the case of cooler stars, most of the light is absorbed by the upper layers of the atmosphere, the downward direct flux decreases rapidly and little is left to heat the troposphere as well; the opposite effect for hotter effective temperatures leads to hotter tropospheres. Planets orbiting cooler stars experience also a small temperature inversion on their stratospheres.

- *Distance factor (ID) variation*. In this case, a small clarification is due. Recalling that, for each stellar temperature in the sample, a script retrieved the four distances corresponding to precise values of stellar irradiation, it is possible to

rank the four modeled planets by distance from the smallest semi-major axis to the largest – no matter the actual value of the four positions, which is indeed different based on the stellar temperature. In short, for any given modeled stellar temperature, the closest modeled planet has a distance factor of 1, while the farthest one has a distance factor of 4. Unsurprisingly, a decrease in distance from the host star heats the whole atmosphere. The heating is however not uniform: while the stratospheres vary more or less evenly, the tropospheres double the variation. This is presumably caused by the increasing thermal contribution of the upper layers of the atmosphere, which is higher the hotter their temperature. This produces more downward infrared radiation, which is more effectively absorbed by the carbon dioxide in the lower layers.

- *Bond Albedo variation*. The albedo is effective in reflecting a portion of incoming visible flux, thus reducing the global stellar flux in all the atmosphere. Increasing the value of the albedo, the whole atmosphere shifts towards lower temperatures, while keeping overall the same shape. For a more complete treatment, however, we expect the stellar temperature to a role, as previously seen, as well as the emissivity of the top layer of the atmosphere.

- *Emissivity variation*. A change in emissivity causes similar changes in the overall temperature as the previous case. The upper layer is much hotter if the emissivity is non-null since the layer absorbs and emits the radiation more effectively than a thin, nearly transparent atmosphere. The effect is mitigated towards the ground layer, with similar temperatures at the surface. In this case as well, the output is strongly coupled with the effective temperature of the host star and with the albedo, causing the value of the difference between an emitting and a non-emitting atmosphere to be extremely subject to other conditions. The chosen values are, however, extreme ones: we expect the real value of an atmospheric emissivity to lie between these two values. Therefore, a more realistic pressure-temperature profile would lie between those two boundaries.

- *Pressure factor α variation*. This factor determines the atmospheric mass to be considered in the model and, as a result, the ground pressure. The effects on the profiles are indeed the same ones as the case of varying planetary mass since both are caused by an increase in ground pressure. There is, therefore, no substantial change in the ground temperature, but rather an increase in the thickness of the troposphere.

- *CO_2 abundance variation*. A change in the carbon dioxide abundance causes a variation in the optical depth of each layer of the atmosphere. Therefore, for

the same incoming flux (same stellar temperature, semi-major axis, albedo, and emissivity), the atmosphere reaches the same optical depth value at lower pressures for increasing carbon dioxide. The flux is then completely absorbed by the upper layers of the atmosphere up to a certain level, after which the thermal emission of the atmosphere alone start to heat the lower layers. This effect is the cause of the temperature inversion, which takes place the deeper in the atmosphere, the lower the values of carbon dioxide in the mixture (see the close-up of this plot in Figure 7.3). For the lowest case of carbon dioxide, which is 100 times lower than the previous point of the input grid, the atmosphere does not use up all the flux and therefore appears to be hotter than the models at higher abundance of CO_2: as a matter of fact, on the ground boundary layer the downward direct flux for the model with a volume mixing ratio of CO_2 equal to $1 \cdot 10^{-4}$ is 30% higher than the same quantity for the model with the maximum abundance of carbon dioxide. The difference in temperature in the upper layers of the atmosphere for the model that considers the lowest CO_2 volume mixing ratio with respect to the profiles at increasing carbon dioxide abundance is due to the minor efficiency of the scattering at the top of the atmosphere: less radiation is scattered, so more is available to heat the region.

- *H_2O abundance variation*. The presence or absence of water vapor in the atmosphere leads to dramatically different results. The largest differences are found in the troposphere, where water has its peak abundance; little or no differences appear in the upper layers of the atmosphere, which are dry for all models. The presence of water strongly impacts the ground temperature as well as the height of the troposphere, due to the variation in critical lapse rate caused by the moist adiabat term in Eq. 6.21. The wet model experiences convection (the levels that are unstable for convection being highlighted by the thicker portion of the profile), while the dry one doesn't.

From this set of plots it is clear that the most prominent process driving the ground temperature (apart from the obvious variation in distance and hence stellar flux) is the presence of water vapor in the atmosphere, while the temperature of the upper layers is driven by the carbon dioxide abundance, whose absorption is driven more or less effectively by the stellar temperature. To prove such statement, we repeated the same plots (Figure 7.4) but choosing the corresponding moist subset of models, instead of selecting the ones without water. Thus, referring to Table 7.6, we changed the value of the relative humidity at the surface from 0% to 60% (or, in MAGRATHEA-friendly language, selecting those models that ran with WATER MODE equal 1 instead of 0).

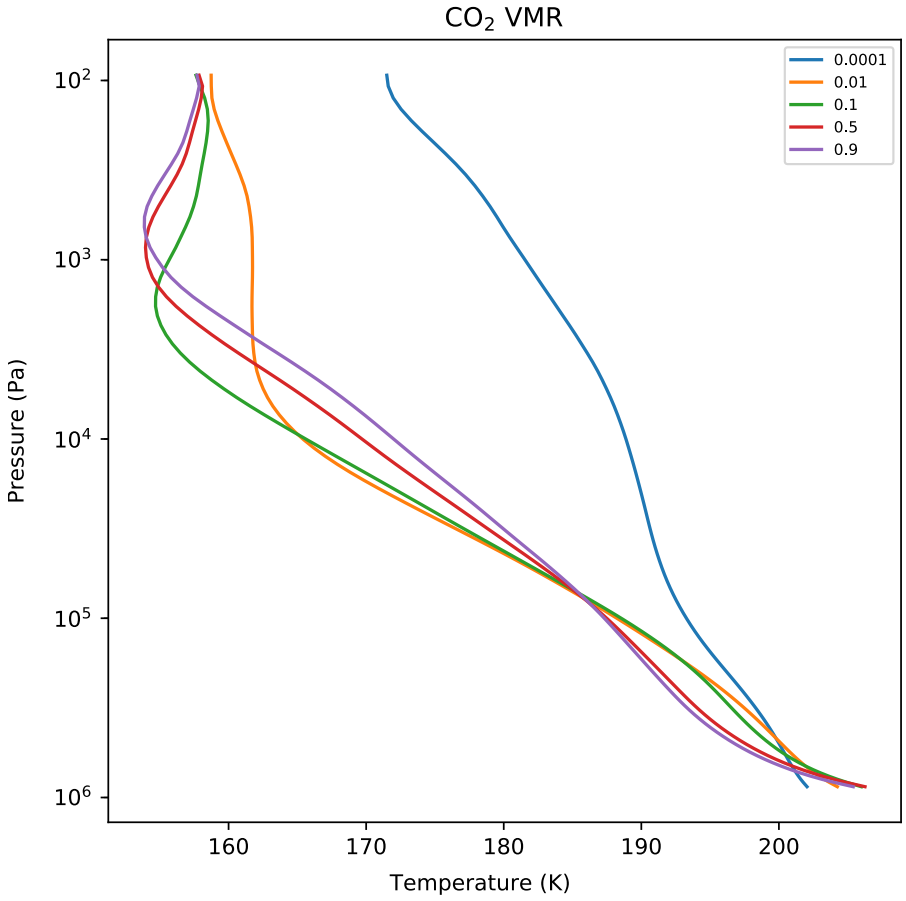


Fig. 7.3: Set of models with varying values of carbon dioxide, based on Table 7.6 (dry atmospheres).

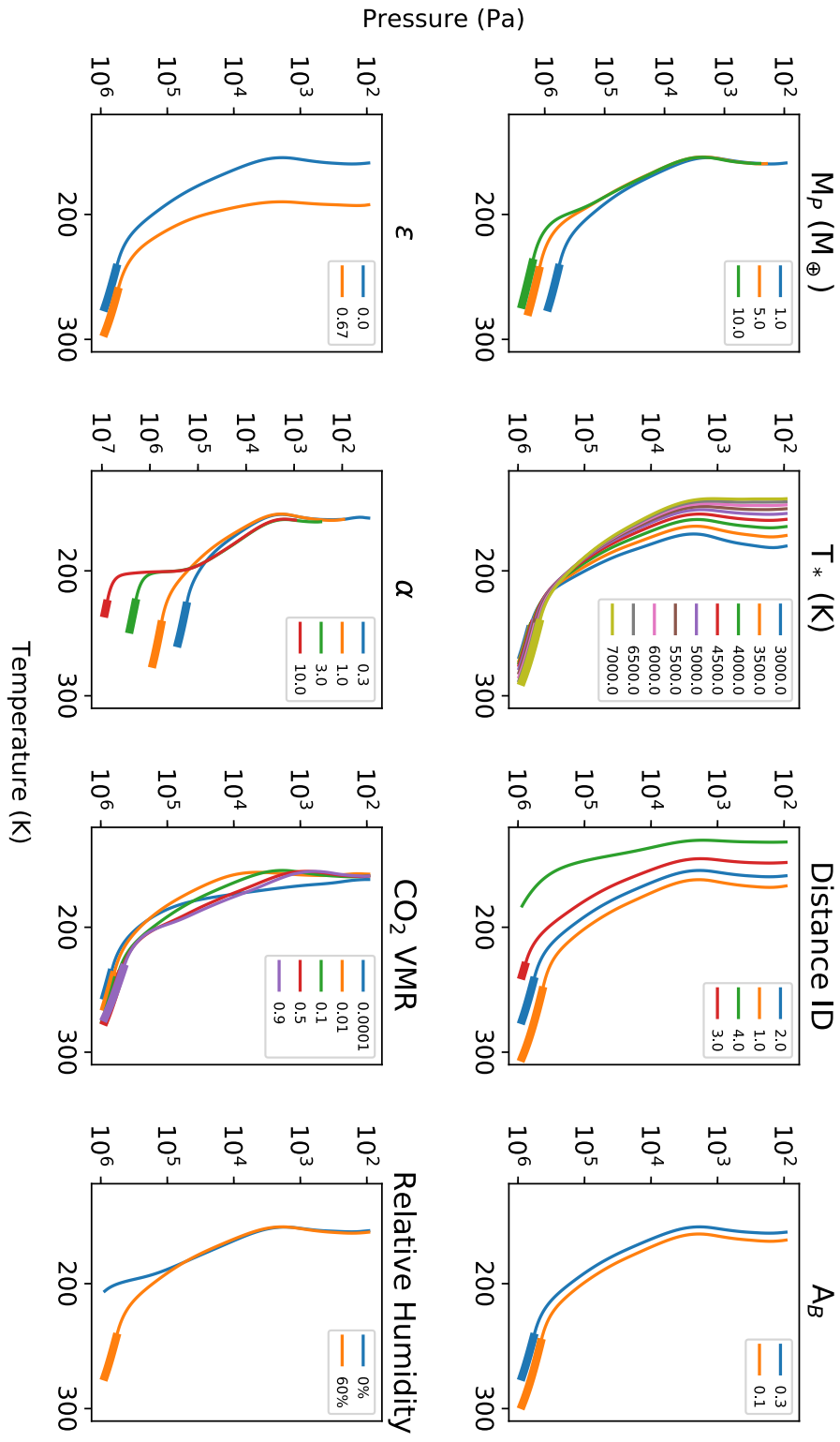


Fig. 7.4: Set of subplots with varying parameters, based on Table 7.6 (wet atmospheres).

The presence of water in the lower layers of the atmosphere acts in the same way and it is sensitive most exclusively to the variation in flux intensity (and thus distance): as a matter of fact, water acts as a “grey absorber”, nearly equally sensitive to all wavelengths. A change in pressure causes an increase of temperature in the middle layers of the atmosphere caused by the presence of water and, because of that, the troposphere on high-pressure models appears to be cooler than the ones at lower pressures, while in the dry models no sensible variation in ground pressure caused by an increase in pressure was detected.

A more detailed view of the atmospheric composition affects the ground temperature at the end of a `MAGRATHEA` run can be seen in Figure 7.5.

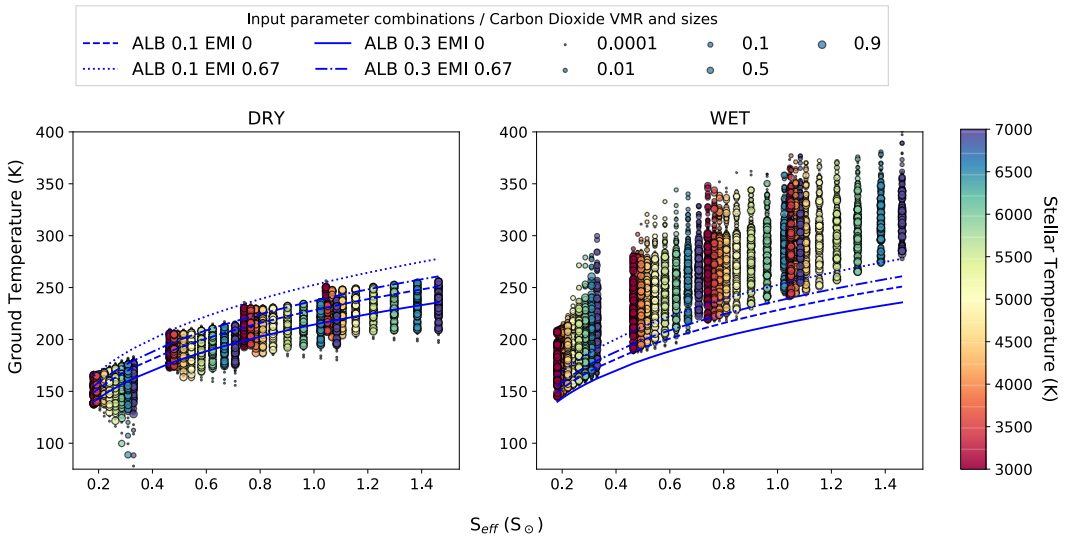


Fig. 7.5: Ground temperature of all models vs stellar irradiation. *Left:* The subset of all the dry models on the theoretical grid. *Right:* The subset of all the wet models of the theoretical grid. Solid, dashed, dash-dotted, and dotted lines represent the trends of the initial temperatures determined by different combinations of the albedo and emissivity values. The size of the dots represents carbon dioxide abundance. The dots are color-coded depending on the stellar temperature.

First of all, every model starts its run with an isothermal temperature that changes according to the albedo, the emissivity of the atmosphere, and the distance from the star (which depends on the stellar temperature by construction). The initial temperature trends are displayed in the four lines in both graphs, corresponding to the four combinations of albedo values and plotted as a function of stellar irradiation (higher values of S_{eff} corresponding to the planets closer to their stars). Generally, the various combinations cause the initial temperature to have slightly different values, but the average difference between one combination and the others is no more than 35 K.

Then, the models evolve up to their radiative-convective equilibrium situation: in particular, the ground temperature moves in the vertical direction of the graph during the run, up to its equilibrium value, displayed in the two subplots. It should be no wonder that the dataset appears to be clustered in four groups, corresponding to the four distance points; the width of the groups depends on the stellar temperature, which causes the effective stellar fluxes to vary according to the definition of the grid (because of the varying habitable zone boundaries).

The two subplots show quite a different behavior of the ground temperature: for the dry subgrid, the output ground temperature is basically independent of the incoming radiation. An increase in stellar temperature does not produce hotter tropospheres as well: instead, the range of temperatures is nearly the same for all the sampled host stars. This is a result of a balance between the slightly increasing stellar irradiation from cooler to hotter stars, and the effect of the albedo on the visible part of the spectrum. The peak of the stellar spectrum shifts from the infrared to the visible with increasing effective temperature, so the albedo is more effective in cutting out the flux for hotter stars. In addition to that, since carbon dioxide is mainly an infrared absorber, the less infrared photons are available, the less absorptive the atmosphere is. Because of that, the atmosphere of planets orbiting hotter stars is slightly cooler than the initial temperature (i.e. the leaky greenhouse temperature).

A slight differentiation depending on the chemical composition of the atmosphere is visible, as models with a lower amount of carbon dioxide tend to be cooler than the ones with a higher CO_2 volume mixing ratio, because of a lower amount of greenhouse absorbers in the atmosphere.

Models with the lowest stellar irradiation show a steep decrease in temperature, pushing the ground temperature to values lower than 100 K. However, an appropriate analysis of such models cannot be made yet, since at least one of the layers in those atmospheres are most likely to enter the regime of CO_2 condensation, which hasn't been implemented in the software yet: for this reason, such models could not be as accurate as the others and should be repeated once some treatment on condensable species is included in `MAGRATHEA`.

When adding water to the mixture, an enhancement of the ground temperature range much wider than the previous case appears. For each modeled stellar irradiation, there is a differentiation between models at low and high carbon dioxide volume mixing ratios, the latter being generally cooler than the first – but in any case hotter than the initial temperature. This effect is caused by the presence of unstable (failed) models at higher abundances of carbon dioxide, which do not allow very high abundances of water vapor – abundances which instead should be higher the current

temperature is. For this reason, only models with low CO_2 abundance manage to retain the required massive amount of water vapor without falling into the negative N_2 abundance error. For a more detailed treatment of failed models, see Section 7.2.2.3.

The effect of water vapor as a grey absorber appears in the increase of ground temperature based on the stellar irradiance: the higher the stellar irradiance (hence, the shorter the distance and the higher the stellar temperature), the hotter the ground temperature. In this case, the trend experienced by the dry subsample is overcome by the presence of water vapor and it is a sign of a runaway greenhouse effect: the higher the temperature, the higher the water vapor abundance in the atmosphere, the higher the heating of the atmosphere.

`MAGRATHEA` seems to be successful (albeit considering a few approximations) in finding a clear relationship between the initial temperature (which depends only on physical parameters), and the ground temperature, which instead accounts for the chemical composition and mass of the modeled atmosphere as well (see Figure 7.6).

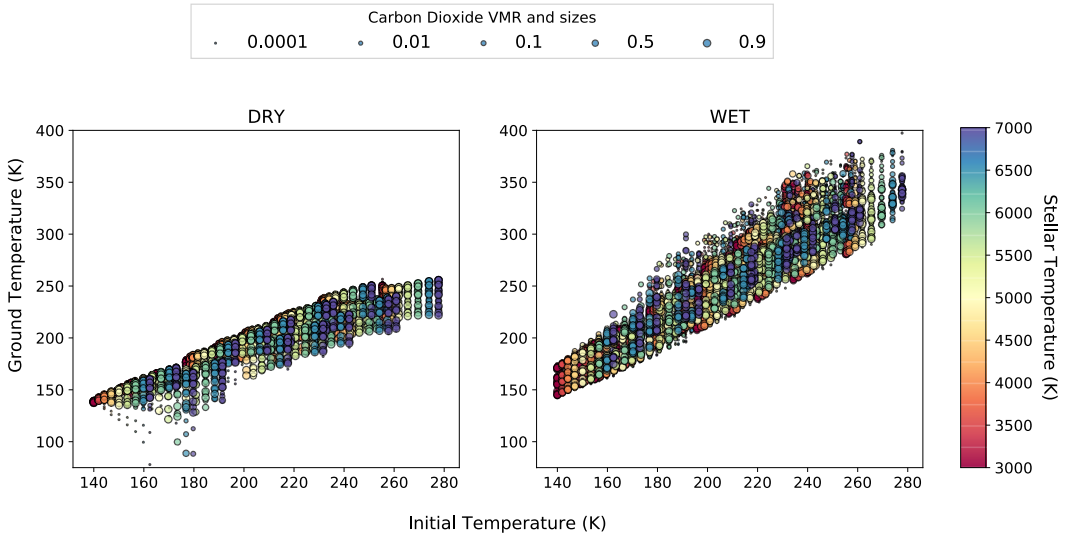


Fig. 7.6: Ground temperature of all models vs initial temperature. *Left:* The subset of all the dry models on the theoretical grid. *Right:* The subset of all the wet models of the theoretical grid. The size of the dots represents the carbon dioxide abundance. The dots are color-coded depending on the stellar temperature.

From what emerges from the two subplots, from an initial temperature it is possible to define a more or less small range of ground temperatures that take into account various combinations of atmospheric composition and mass.

These relationships could be particularly useful to look-up when in need of some realistic information concerning the ground temperature of an observed target: provided that the host star temperature and the distance are known, it is possible to gain information concerning the surface temperature (and thus, the planetary habitability) even without any detailed information on the atmospheric composition. Of course, the more detailed the characterization of the planet is (in particular, if an estimate of the planetary mass is known, see e.g. Figure 7.7), the smaller the uncertainties on the surface temperature could be. For this reason, we are encouraged to run more models in the future, expanding the grid of models to be more specific in providing information on the surface conditions – which hardly can be measured with current facilities.

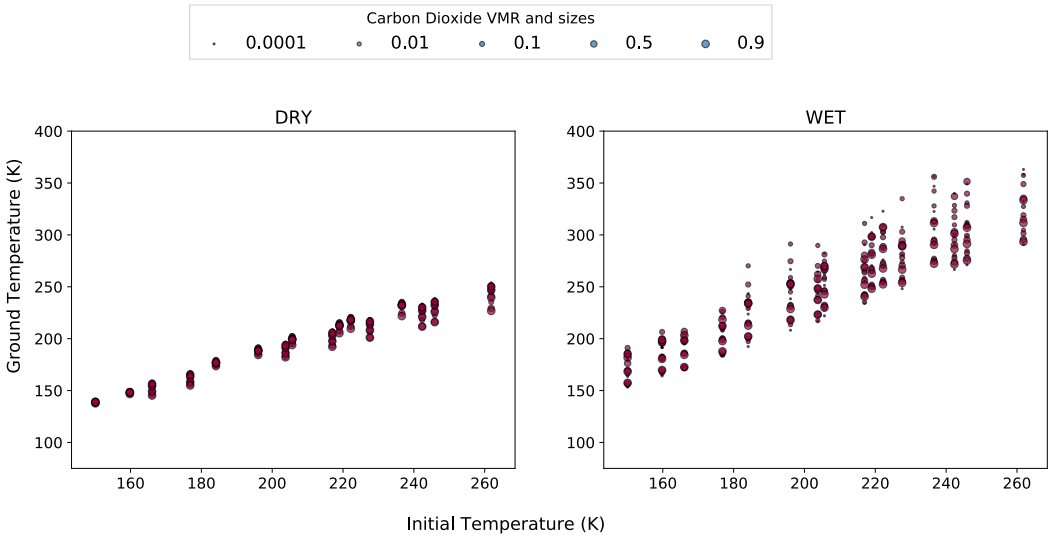


Fig. 7.7: Ground temperature of the models with planetary mass equal to 5 $M_{\oplus}$ vs initial temperature. *Left:* The subset of the dry models of the sample. *Right:* The subset of the wet models of the sample. The size of the dots represents the carbon dioxide abundance.

However, both albedo and emissivity depend on the chemical composition of the atmosphere and on the surface reflectivity to be able to provide more rigorous results. This is especially true when considering cloud decks or non-uniform cloud distributions to the atmospheres. These features are currently being implemented in the software and will be tested soon.

This qualitative analysis highlighted a few important information concerning the grid of models and its potential. In any case, as previously seen with the analysis of Figures 7.2 and 7.4, a more detailed treatment of the output profiles claims a study based on derived parameters rather than direct input parameters, to properly

infer what are the quantities that best describe the features of the modeled grids as a whole.

7.2.2.2 Principal Component Analysis

When analyzing large datasets that depend on a large number of interrelated variables, it is common to adopt a dimensionality-reduction method such as the Principal Component Analysis (PCA). This technique aims at transforming a large set of variables into smaller ones while retaining most of the information present in the original dataset. Such new variables are the Principal Components (PCs): these are uncorrelated and ordered so that the first ones represent most of the information present in all the original values. In this way, the new dataset could be much easier to be explored and visualized, at the expense of negligible loss in accuracy.

The PCs are linear combinations of the original data which are uncorrelated from one another and that are linked to the largest possible variance in the dataset. There are as many principal components as there are variables in the data, but the first few PCs are then expected to be the ones that carry the majority of the information about the dataset. In general, the first few PCs will consider most of the variation in the original variables, while the last few PCs will identify near-constant linear relationships among the original variables [Joliffe, 2002]. For further information concerning the Principal Component Analysis, see Box 7.2.2.2.

In the case of the atmospheric profiles that we are about to analyze, the original dataset would be composed by a 2D array of shape $M \times N$, where M is the number of models to be analyzed, and N the number of points at which the temperature is evaluated (in this case, N is equal to 100, since 100 are the boundary layers considered by `MAGRATHEA`). In other words, the whole dataset is a matrix where the rows are the final atmospheric temperature profiles, evaluated at the 100 boundary layers, which are the columns.

To find PCs mathematically, it is necessary to standardize the variables, to convert them to the same scale, so that it would be possible to consider normalized variances for the dataset. Each temperature array must be therefore normalized by subtracting to each value the mean over the 100 boundary layers and dividing by the standard deviation of the profile.

Box 7.2.2.2 - Principal Component Analysis

The variance of the dataset is related to the distance of each variable from the average value. As shown in Figure 7.8, the maximum variance of a PC can be graphically interpreted as the largest average of the squared distance of data projection points with respect to the origin, on the direction of the PC itself: the larger the variance carried by the line, i.e. the larger the dispersion of data projection points, the more information is carried out by the first PC. The second PC would then be on a perpendicular direction with respect to the first PC, with the second largest variance that characterizes the sample.

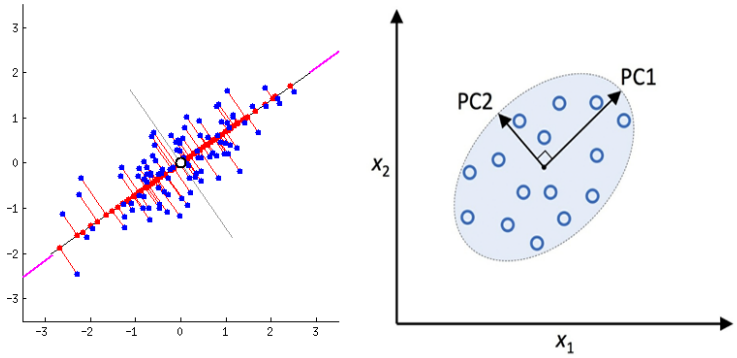


Fig. 7.8: *Left:* Proxy scatter plot of a random dataset (blue points); in red, the projections of the dataset over the first principal component, whose direction is determined by the black and purple lines. The orthogonal line represents the direction of the second PC (adapted from Towards Data Science Website [Jaadi, 2005]). *Right:* sketch of a random scatter plot where x_1 and x_2 are the original variables, while PC1 and PC2 are the principal components [Raschka, 2015, adapted from].

Supposing a dataset composed by a 2D array of shape $M \times N$, the covariance matrix can be then defined as the symmetric $N \times N$ matrix, whose (i, j)-th element is the covariance between the ith and jth elements x_i and x_j (both composed by M variables x_i^k and x_j^k, where k goes from 1 to M):

$$Cov_{i,j} = \frac{1}{M} \sum_{k=1}^{M} (x_i^k - \mu_i)(x_j^k - \mu_j) \tag{7.8}$$

Keeping in mind that $Cov(j, i) = Cov(i, j)$ and $Cov(i, i) = Var(i)$, and that mean values μ_i and μ_j are null if the dataset is standardized, the covariance matrix can be expressed as follows:

$$\mathcal{C} = \begin{bmatrix} Var(1) & Cov(1,2) & \cdots & Cov(1,N) \\ Cov(1,2) & Var(2) & & Cov(2,N) \\ \vdots & & \ddots & \vdots \\ Cov(1,N) & Cov(2,N) & \cdots & Var(N) \end{bmatrix} \tag{7.9}$$

The covariances express the correlations between two elements. It is particularly important to study the sign of the covariance: when positive, the two elements increase or decrease together (i.e. are directly correlated); when negative, one element decreases if the other increases and viceversa (i.e. they are inversely correlated).
The eigenvectors of the covariance matrix represent the principal components, whereas the corresponding eigenvalues will define their magnitude. An eigenvector $\tilde{v_i}$, by definition, must satisfy the condition:

$$\mathcal{C}\tilde{v_i} = \lambda_i \tilde{v_i} \tag{7.10}$$

Where λ_i is the (scalar) eigenvalue. In principle, N eigenvalues/eigenvectors can be retrieved from the covariance matrix: these are the principal components. By ordering the values of the eigenvalues λ_i, the PCs can be ranked.
It is then possible to reduce the dimensionality of the dataset by selecting a subset of PCs that contains most of the variance and to transform the dataset onto the new PCs axes, by constructing a projection matrix $\mathcal{W}$ composed by the chosen eigenvectors, and applying it to the original dataset.
When plotting the transformed dataset, adding labels or coloring data to help visualization, it is possible to detect clusters of similar features, or pattern in the data.

From the PCA we find that the first 4 components describe a cumulative variance of 99.7%, retaining nearly all information on the whole grid.

In particular, the first principal component $\tilde{v_1}$ explains the 86.7% of the variance, the second one $\tilde{v_2}$ the 10.7%, the third one $\tilde{v_3}$ the 1.9% of the variance, and the fourth one $\tilde{v_4}$ the 0.5%. These four eigenvectors are shown in Figure 7.9.

Here, it is possible to notice as well that the ranking of the components is linked to the sign inversion of the eigenvector magnitudes: the first component has always the same sign, the second one encounters once the zero, the third components crosses the origin twice, and the fourth one three times.

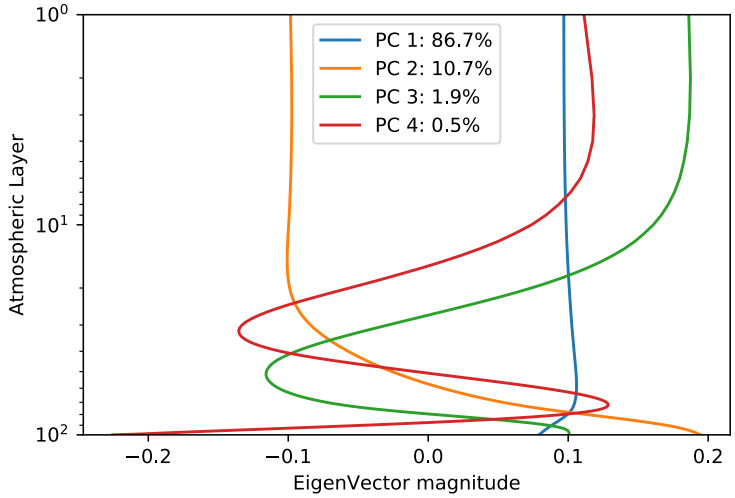


Fig. 7.9: First four principal components (and corresponding variance) of the theoretical grid as a function of the atmospheric layer, from first modeled layer (NL = 1) to the ground level (NL = 100).

A linear combination of these four eigenvectors $\tilde{v_i}$, taking the corresponding eigenvalues λ_i as weights, can reproduce the general behavior of any pressure-temperature profile of the theoretical grid. As an example, we selected a random model and tried to reconstruct it with a varying number of principal components, by use of the equation:

$$\tilde{T} = \sum_{i=1}^{k} \lambda_i \tilde{v_i} \tag{7.11}$$

where $\tilde{T}$ is the normalized temperature profile, which then needs to be multiplied back by the standard deviation of the actual profile and adding the mean temperature over all layers, to be compared with the actual modeled atmospheric profile. The resulting profiles, compared with the actual data (blue solid line), are shown in Figure 7.10. The values of k, i.e. the numbers of principal components considered in the linear combination, are shown in the legend and color-linked to the corresponding output profiles.

It is clear that the first component (orange line, $k = 1$) determines the overall behavior of the profile, by approximating the higher levels of the atmosphere, but leading to much lower temperatures at the surface. This region is, on the contrary, well modeled by the combination of the first and second components (green line, $k = 2$), up to the slight temperature inversion at the 30th atmospheric layer, after which the profile shows cooler temperatures than the real profile. The linear

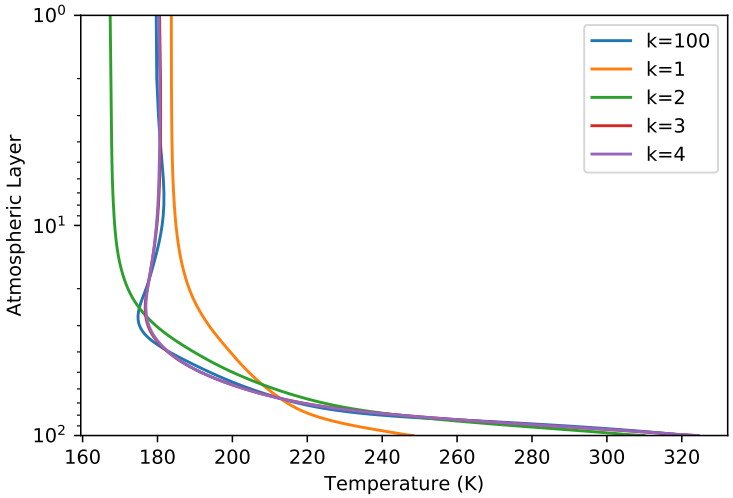


Fig. 7.10: *Blue solid line*: normalized pressure-temperature profile of a 5 $M_\oplus$ Super Earth orbiting a 4000 K star at a distance of 0.276 au, with a ground pressure of 1.83 atm, a uniform 1% carbon dioxide abundance and 3.8% water vapor at the surface, with albedo $A_B = 0.1$, and emissivity $\epsilon = 0$. *Orange solid line*: normalized profile derived from the linear combination of the first eigenvector and corresponding eigenvalue. *Green solid line*: normalized profile derived from the linear combination of the first two eigenvectors and corresponding eigenvalues. *Red solid line*: normalized profile derived from the linear combination of the first three eigenvectors and corresponding eigenvalues. *Purple solid line*: normalized profile derived from the linear combination of the first four eigenvectors and corresponding eigenvalues.

combinations of the first three and four components (red and purple line, $k = 3$ and $k = 4$ respectively), are basically overlapping on all layers, reproducing more accurately the shape of the actual profile, with negligible differences of a few K at most.

It appears therefore clear that the more principal components are considered in the linear combination, the more accurate the reconstruction of each profile would be.

An analysis of the eigenvalues of the first four components (each one of them being composed by a total of 16934 elements, one for each successful model of the grid) would allow a more accurate determination of all the parameters, either input quantities or some combination of them, that are more relevant in shaping the various profiles.

The reduction in dimensionality aims at a better visualization of the whole dataset. Four dimensions are, however, still hard to be visually represented. We, therefore, chose to reproduce 2D slices of the 4D sample, each point representing one model and being determined by four coefficients which are the eigenvalues corresponding to the four principal directions (the principal components/eigenvectors), plotting scatterplots of all combinations of the first four eigenvalues.

All plots are color-coded based on some interesting parameters. The results are shown in Figures from 7.11 to 7.18.

These plots, in other words, show the projection of the profiles along the principal components: by using the colors, it is possible to distinguish any sort of grouping in each subplot. If differently colored clusters appear and are well separated by a particular set of eigenvalues, it means that the parameter that determines the color-coding is dominant in determining the weights in the reconstruction of the P-T profile in the direction (i.e. principal component) corresponding to the set of eigenvalues itself. If, on the contrary, the color-coding does not suggest any particular clustering for a given eigenvalue, that parameter does not influence that component.

Since the principal components are ranked based on the variance of the whole sample, we should expect that all parameters that determine the first components are the ones that cause the largest variation in the shape of any given profile. The parameters that are effective in clustering the data on the second principal component produce “second-order” variations, and so on for all the remaining components.

From the analysis of the shown plots it is possible to infer that:

- The incoming stellar flux at the top of the atmosphere in units of Wm^{-2}, defined as

$$F_{TOA} = (1 - A_B) \cdot T_\star^4 \left(\frac{R_\star}{a}\right)^2 \tag{7.12}$$

(Where A_B is the albedo, $T_\star$ is the stellar temperature in K, $R_\star$ the stellar radius in meters, and a the semi-major axis in meters) determines the first principal component. As a matter of fact, in the top row of subplots in Figure 7.11, negative values of λ_1 identify with the models with lower values of flux at the top of the atmosphere, while higher fluxes correspond to positive values of λ_1. No clear distinction appears on the remaining plots, indicating that there is no correlation between this parameter and the other principal components. This information confirms what previously seen in Figure 7.4, where the albedo and the distance (and therefore the incoming flux) operate a translation of the whole profile towards lower or hotter temperatures.

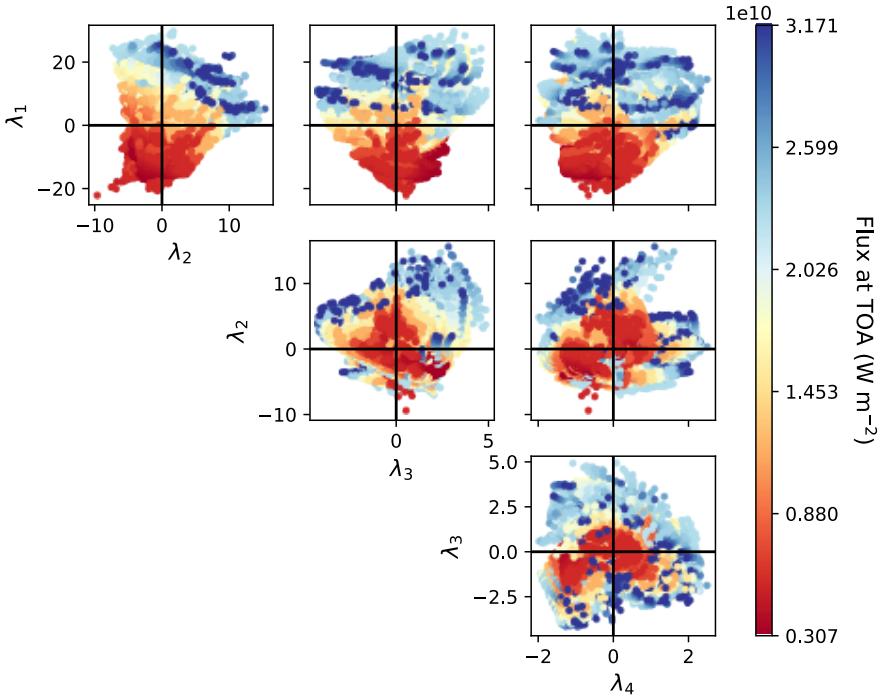


Fig. 7.11: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the flux at the top of the atmosphere.

- The peak wavelength of the host star in μm, which is determined by Wien's displacement law in this way:

$$\lambda_{max} = \frac{b}{T_\star} \tag{7.13}$$

Where $b = 2.8977685 \cdot 10^{-3}\ m\ K$ is Wien's displacement constant, and $T_\star$ is the stellar temperature in K, contributes to determine the second and third component (Figure 7.12). This is also visible from the previous analysis since the stellar effective temperature changes the shape of the profile but does not interfere massively with the mean temperature. Indeed, the most important difference when treating the radiative transfer of an atmospheric layer is the numeric value of the incoming flux, determined by distance and albedo. This value is determined to be nearly the same for each one of the four distances by construction (given the similar S_{eff} for the four distances); the differences in the shape of the blackbody radiation due to the variation of stellar temperature determine the absorption of the layer (and thus the stratospheric features, as seen in Figure 7.4) but not its overall heating.

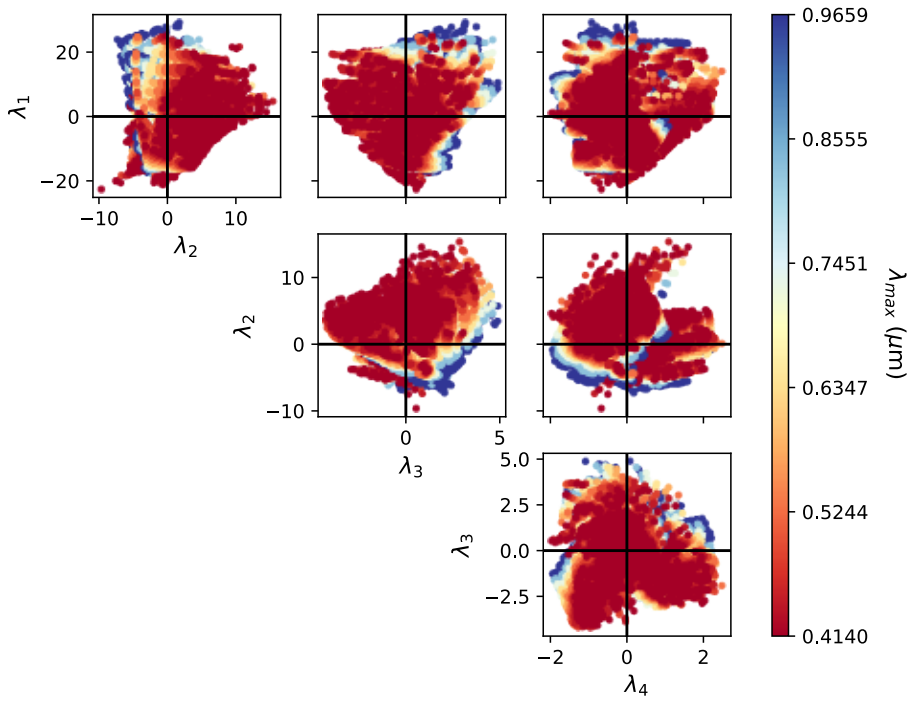


Fig. 7.12: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the peak wavelength determined by Wien's displacement law for all stellar effective temperatures.

- The emissivity of the top layer (Figure 7.13) contributes to determining the second component. The initial temperature, which is a linear combination of the flux and the emissivity, unsurprisingly shows that the two quantities, entwined, contribute to both the first and the second component (Figure 7.14). This information confirms what previously found, as well.

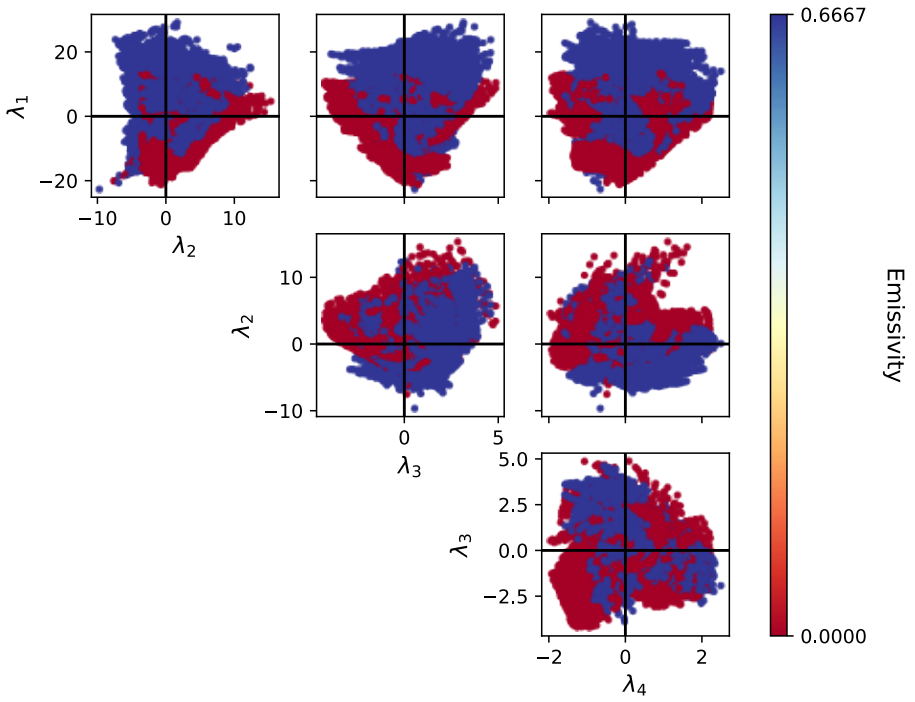


Fig. 7.13: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the emissivity at the top of the atmosphere.

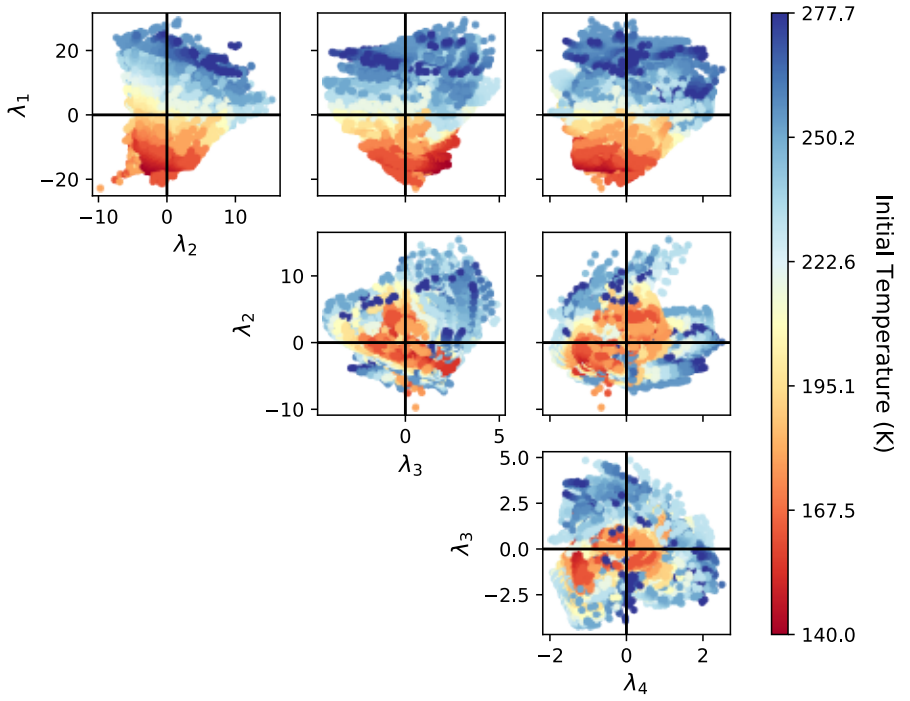


Fig. 7.14: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the initial temperature.

- The presence or absence of water vapor (in Figure 7.15 displayed in terms of the *water switch*, which is 1 for all wet models and 0 otherwise) contributes to control the second and fourth component. For the subset of wet models, we plotted the principal component distribution in Figure 7.16, color-coded by the volume mixing ratio of water vapor in logarithmic scale: for those models, water vapor controls the first and second components. As a matter of fact, models with negative values of λ_1 are the one with the lowest amount of water vapor in the atmosphere. This effect is strictly coupled with the incoming flux because of the variability of water vapor abundance throughout a `MAGRATHEA` run: the volume mixing ratio of such species is triggered by (and triggers in return) the increase of temperature, which is more likely to happen for close-in planets.

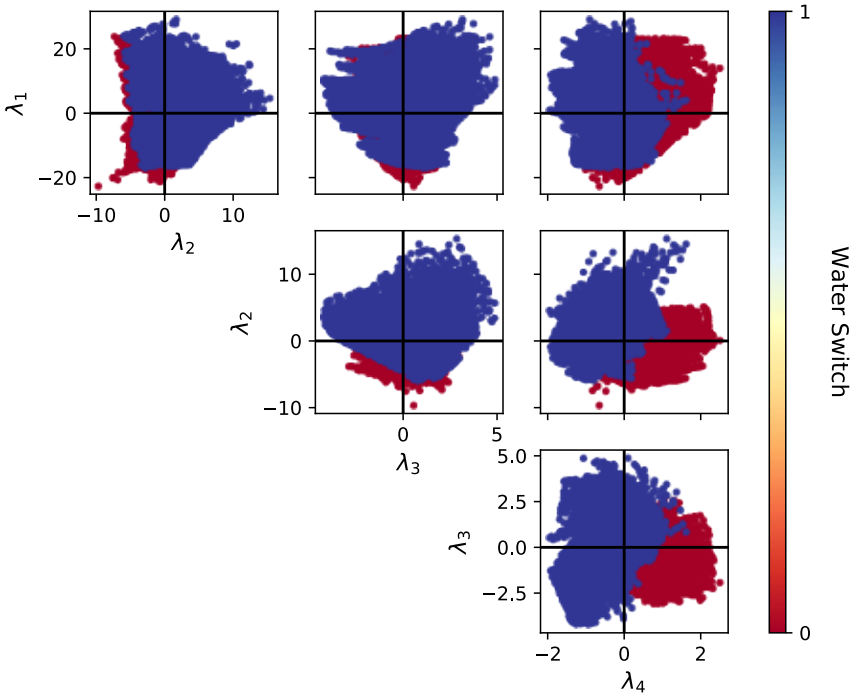


Fig. 7.15: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the water vapor switch (0 for the models with no water vapor in the atmosphere, 1 for the models with water vapor in the atmosphere).

- The carbon dioxide abundance contributes to the determination of the second and third component, as shown in Figure 7.17. This is in agreement with what previously noticed, where the variation of CO_2 abundance caused only peculiar features in the models to change.

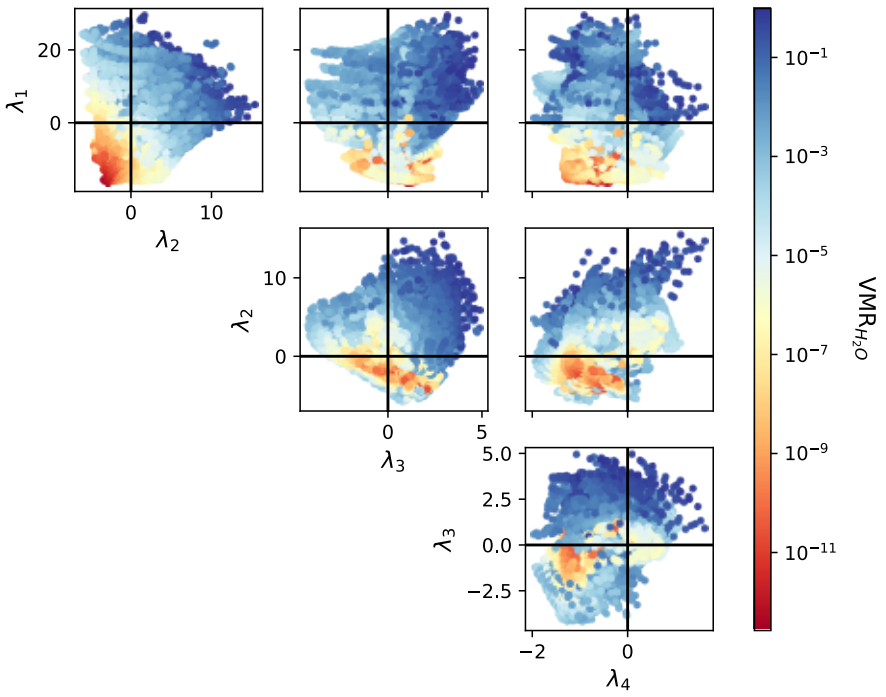


Fig. 7.16: Scatter plot of all combinations of the eigenvalues of the first four principal components **for the subsample of wet models**. The colors are determined by water vapor abundance at the ground boundary layer in logarithmic scale.

- The abundance of the greenhouse gases, the sum of carbon dioxide and water vapor, has a strong impact on the second component, thus causing second-order variations on the models. This is, of course, a combination of both the contributions of the two species, which impact the features of the upper layers of the modeled atmospheres (carbon dioxide in particular) and the lower levels (mainly water vapor).

The other parameters do not appear to infer any of the first four components and are therefore not shown. This is foreseeable since input parameters such as the planetary mass and the pressure factor cause a shift in the vertical (pressure-driven) direction in the P-T profiles, but the PCA cannot take into account such component because the atmospheric profiles need to be normalized. Information about the pressure is however still accounted for in the volume mixing ratio of water, which is calculated from the ratio between the partial pressure of that species and the total pressure of each boundary layer.

By performing the Principal Component Analysis, we restricted the most important input parameters of the grid and ranked them by importance: the first component is mainly driven by the incoming flux (and therefore the semi-major axis and the

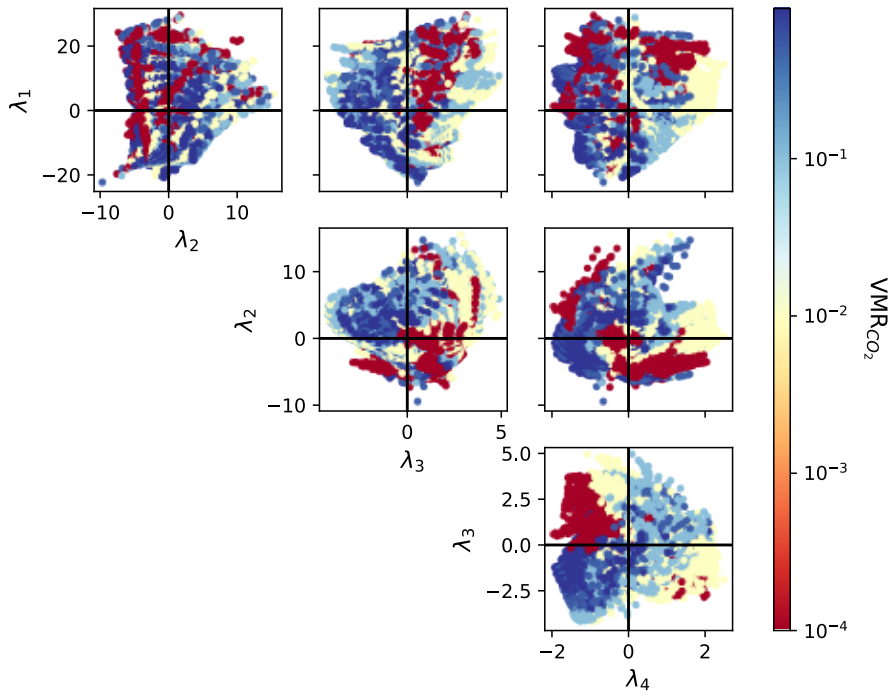


Fig. 7.17: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the abundance of carbon dioxide in logarithmic scale.

albedo) and water vapor (for wet models only); the second component is determined by the stellar temperature, which drives the magnitude of the absorption features of both carbon dioxide and water vapor, and by the emissivity of the top layer as well; specific features caused by carbon dioxide and water vapor influence the third and fourth component.

To have a hands-on understanding of the results of this analysis, we can look at the profiles that show large variations on each one of the four principal components and check which parameters effectively change massively in the two cases. To take into account the contribution of lower-ranked PCs, it is, however, necessary to fix the values of the upper ones, and therefore reducing at each step the number of elements in the sample, as more conditions are applied to the original distribution.

For the first PC, we selected the profiles corresponding to the minimum and the maximum values of the eigenvalue λ_1: these are shown in Figure 7.19. In this case, we expect the contribution of all the PCs to contribute to the overall profile, the greatest variance being represented by the first principal component.

The blue profile (whose λ_1 value is the minimum of the population) showing a planet orbiting a 7000 K star with a relatively small incoming flux. No water is present in

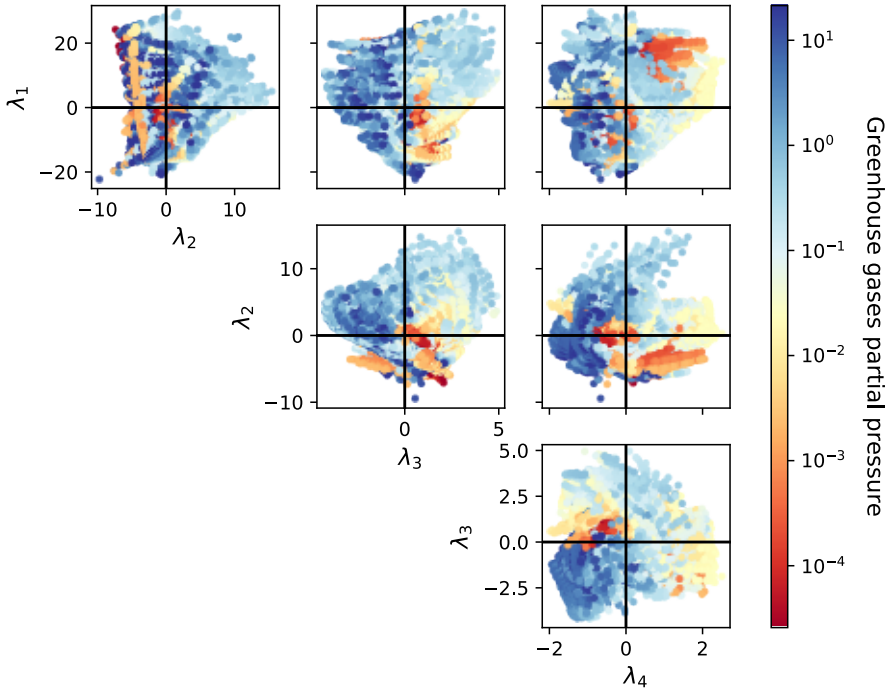
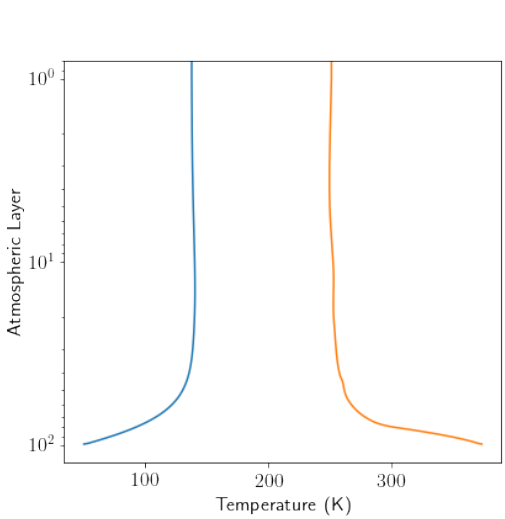


Fig. 7.18: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the total volume mixing ratio of the greenhouse gases in the atmosphere (CO_2 and H_2O) in logarithmic scale.



	minimum	maximum
λ value	-22.68	28.92
T_* (K)	7000.0	3000.0
S_{eff}	0.33	1.04
A_B	0.3	0.1
ϵ	0.67	0.67
F_{TOA}(W m^{-2})	5.56E+09	2.27E+10
T_{in} (K)	179.72	255.34
P (atm)	23.57	0.71
CO2 VMR	5.00E-01	1.00E-04
H2O VMR	0.00E+00	8.55E-01
CO2+H2O VMR	5.00E-01	8.55E-01

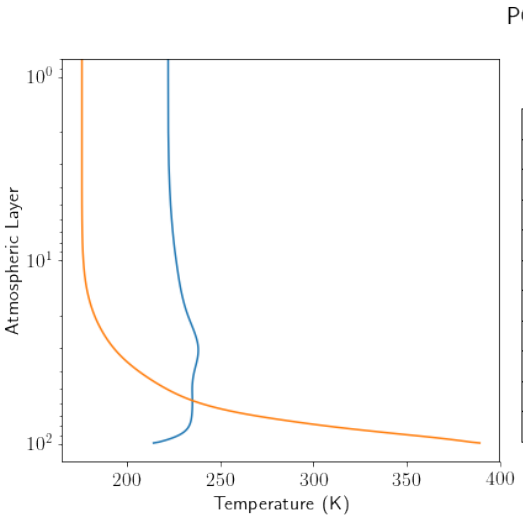
Fig. 7.19: Profiles corresponding to the minimum and maximum values of the first PC. Input details of such profiles are shown in the table on the right of the figure.

the lower layers and the carbon dioxide does not efficiently absorb throughout all layers: as a matter of fact, the input temperature is higher than the output, showing a cooling of the atmosphere throughout the run.

Quite the opposite behavior appears in the orange profile (whose λ_1 value is the maximum of the population): the stellar temperature is much lower and the shorter distances cause an enhancement in the incoming flux. Water vapor is present in the atmosphere and carbon dioxide, although less abundant than the previous case, is more efficient in absorbing the incoming flux because of the peak of the blackbody radiation towards the infrared. The input temperature is higher than the previous case, and the output profile is overall hotter than the input.

From this Figure it is clear that the flux magnitude and the presence of water have a strong influence on the overall shape of the profiles, thus determining first-order variations on the outputs.

Some secondary effects can be observed by setting the value of the principal component to a common value of $\overline{\lambda_1}$ and then selecting the minimum and maximum of the λ_2 eigenvalue. In this way, we are minimizing the first-order variations caused by the first component, and taking into account the smaller ones, in particular, the second PC that retains the largest residual variance of the sample. For the present analysis, we set $\overline{\lambda_1} \approx 14$. The results are shown in Figure 7.20.



	minimum	maximum
λ value	-6.27	12.31
T_* (K)	3000.0	7000.0
S_{eff}	0.74	1.46
A_B	0.1	0.3
ϵ	0.67	0.67
F_{TOA}(W m^{-2})	1.61E+10	2.47E+10
T_{in} (K)	234.31	260.83
P (atm)	8.59	18.33
CO2 VMR	9.00E-01	1.00E-02
H2O VMR	0.00E+00	5.61E-02
CO2+H2O VMR	9.00E-01	6.61E-02

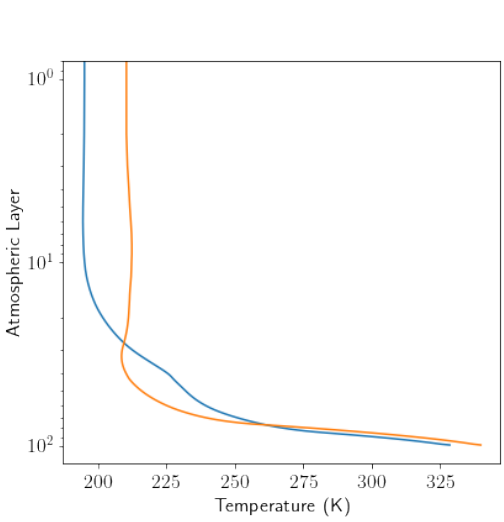
Fig. 7.20: Profiles corresponding to the minimum and maximum values of the second PC. Input details of such profiles are shown in the table on the right of the figure.

Here, the two profiles receive more similar fluxes than the previous case and started their run with a similar input temperature. The blue profile (whose λ_2 value is the minimum of the subsample) is dry and has a massive amount of carbon dioxide in

the atmosphere, which is effective in absorbing the infrared portion of the stellar blackbody radiation, driven by the effective temperature of 3000 K. The upper layers of the atmosphere are warmed up because of the CO_2 absorption, which uses all the flux and leaves the lower layers cooler.

On the contrary, the orange profile (whose λ_2 value is the maximum of the subsample) retains water vapor in the lower layers, while having a lower abundance of carbon dioxide. The presence of water (whose numeric density is extremely abundant due to the high pressure at those lower levels) causes a massive increase in the tropospheric temperature, enhanced by the lower effectiveness of CO_2 in absorbing the incoming flux. So, even though the global mixing ratio of the greenhouse species is lowest for the orange profile with respect to the blue one (only 6% compared to 90%), their coupled effect leads to more extreme results.

For the analysis of third-order variations, we had to set both values of λ_1 and λ_2: we chose $\overline{\lambda_1} \approx 14$ and $\overline{\lambda_2} \approx 4.5$. The results are shown in Figure 7.21.



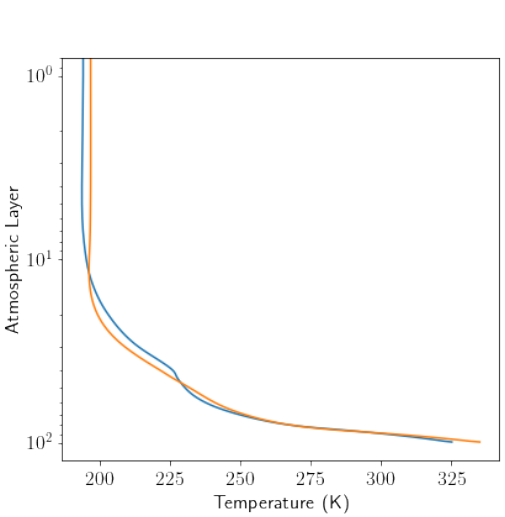
	minimum	maximum
λ value	-0.48	3.98
T_* (K)	6000.0	4000.0
S_{eff}	1.3	1.07
A_B	0.1	0.3
ϵ	0.67	0.67
F_{TOA}(W m^{-2})	2.81E+10	1.80E+10
T_{in} (K)	269.57	241.08
P (atm)	2.58	0.26
CO2 VMR	5.00E-01	1.00E-01
H2O VMR	3.66E-02	6.21E-01
CO2+H2O VMR	5.37E-01	7.21E-01

Fig. 7.21: Profiles corresponding to the minimum and maximum values of the third PC. Input details of such profiles are shown in the table on the right of the figure.

The resulting profiles show a similar difference in incoming flux as the previous case (which shouldn't wonder, since that parameter controls mainly the first component, which is still the same one from the previous plot). There is a variation (although lower than the before) in the stellar effective temperature, as well as in the chemical composition of the two models – both containing water vapor in their troposphere. The variations in the profiles are therefore caused by these three parameters, with a cooler stratosphere in the blue profile (a less effective CO_2 absorption due to the

shape of the blackbody radiation) and hotter troposphere for the orange profile (a combined effect caused by CO_2 and H_2O absorption).

The effects on the fourth PC can be seen when fixing the other three components eigenvalues: we therefore set $\overline{\lambda_1} \approx 14$, $\overline{\lambda_2} \approx 4.5$, and $\overline{\lambda_3} \approx 0$. Figure 7.22 shows the results.



	minimum	maximum
λ value	-0.69	-0.1
T_* (K)	6000.0	6000.0
S_{eff}	1.3	1.3
A_B	0.1	0.1
ϵ	0.67	0.67
F_{TOA}(W m^{-2})	2.81E+10	2.81E+10
T_{in} (K)	269.57	269.57
P (atm)	2.58	2.58
CO2 VMR	9.00E-01	1.00E-01
H2O VMR	3.11E-02	4.98E-02
CO2+H2O VMR	9.31E-01	1.50E-01

Fig. 7.22: Profiles corresponding to the minimum and maximum values of the fourth PC. Input details of such profiles are shown in the table on the right of the figure.

First of all, we notice that both values of λ_4 are negative, underlying that such combination of the other principal components would lead to extremely similar profiles – information that, after all, shows that the first components do describe the general behavior of a profile with a high degree of accuracy.

In this case, there is no difference at all in the incoming flux (nor ground pressure) and the slight variations in the profiles are caused by the varying abundances of greenhouse gases in the atmosphere, the hotter troposphere corresponding to the model with higher H_2O volume mixing ratio.

These graphs are successful in convincing us that the principal component analysis was properly performed.

7.2.2.3 Failed models

As previously mentioned when describing `MAGRATHEA`, the code is prescribed to stop when some features of the atmosphere fall in the regime of instability or trespass the validity range of the code itself (see Section 6.8).

Very few models in the grid of theoretical Super Earths failed (only 346), and all of them failed because of the instability of the atmosphere, rather than achieving hotter temperatures than 500 K. This is, in some ways, an effect of the grid construction: all those models considered rather small irradiations (the maximum one being about 1.5 times the solar constant), so one should expect rather temperate atmospheres.

However, it is interesting to study under which conditions the modeled atmospheres are more likely to be unstable. In Figure 7.23 a set of plots showing the normalized frequency of successful (green, crossed diagonal hatching sample) and unstable (yellow, plain sample) models with respect to various input parameters are shown. For all subplots, each normalized bin is based on the same number of models: in particular, for the most model-dependent parameters (semi-major axis, effective stellar flux, initial temperature, and ground pressure), the margins of the bins are set in order to have the same number of models for each bin. For the other parameters, their initial values were taken. The normalization is therefore simply used to better visualize the fraction of models belonging to each category.

All parameters appear to play some role in the success or failure of the models. The largest impact on the stability of the model is, unsurprisingly, the water vapor abundance: only the wet models are the ones who may fail. The rate of failure is linked to the abundance of carbon dioxide as well, which is the one that determines the maximum abundance of water vapor. Higher initial temperatures also appear to impact on the success rate: the hotter the atmosphere is from the very beginning, the more water vapor is likely to be calculated and, in turn, even hotter temperatures are generated – another demonstration of the water/temperature positive feedback mechanism.

Higher S_{eff} values (and shorter semi-major axes), as well as higher emissivities and lower albedo values, confirm this trend, being the parameters involved in the determination of the initial temperature.

Smaller values of the ground pressure have a slightly higher chance of failing, because they cause higher volume mixing ratios of H_2O: for the same atmospheric temperature in any layer, the very same partial pressure of water vapor is produced; then, when calculating the mixing ratio of that species by dividing the value of the partial pressure to the total pressure, lower pressures return higher volume mixing ratios by definition.

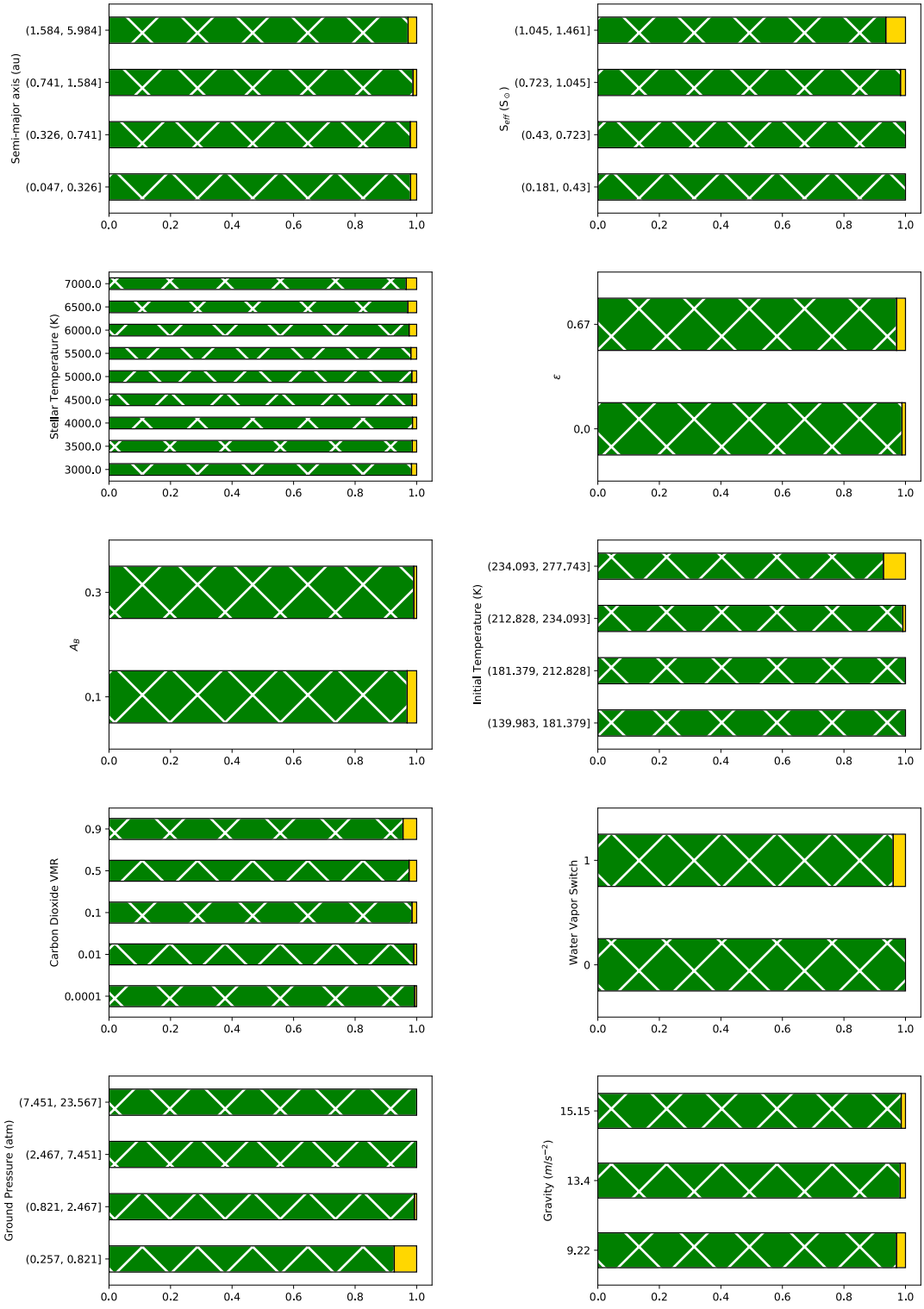


Fig. 7.23: Stacked histograms for the normalized frequency of successful (green, crossed diagonal hatching sample) or unstable (yellow, plain sample) models with respect to a few input parameters.

7.2.2.4 Habitability

Recalling Figures 7.5 and 7.6, some considerations concerning the surface habitability of the modeled planets can be made.

First of all, it is possible to notice that the majority of the initial temperatures for all combinations of albedo and emissivity values (the blue solid/dotted/dashed/dash-dotted lines) lie below the freezing point of water at standard pressure and temperature, even though half of the modeled planets are, by construction, within the maximum and runaway greenhouse limits of the habitable zone. This should not concern much: after all, the limits calculated by Kopparapu et al. (2013) do take into account the contribution of the atmosphere, which is not included in the retrieval of the initial temperature – but will indeed be considered throughout the run. It is, therefore, more important to focus on the habitability of the converged runs to perform a more coherent analysis.

However, due to the different approaches used by the two software programs, the results may be different: even though `MAGRATHEA` is more self/consistent in retrieving the atmospheric profile than the Kopparapu et al. (ibid.) model, which uses a fixed temperature profile, the constant values of albedo and emissivity (quantities that are, actually, extremely dependent on the chemical composition and therefore wavelength) may cause an underestimate of the incoming flux in the present case, with respect to the other model.

On the other hand, it is also important not to incur in the opposite situation, overestimating the irradiation that an atmosphere receives, since this may lead to abrupt changes in the chemical composition due to a higher production of water vapor.

The initial conditions are indeed important for the reliability of the results, so varying their retrieval when adding new features to the software is very likely to lead to different results. However, adding complexity little by little is important to understand the impact of each update.

At this present stage, even with a few simplifying assumptions, `MAGRATHEA` can indeed provide some qualitative results on the habitability of the modeled atmosphere.

It is possible to define the *thermodynamic habitability* of a planet, based on the pressure-dependent temperature range of liquidity of water on the surface. Therefore, recalling the phase diagram of a chemical species in Figure 4.3, a planet is habitable if this relationship is valid:

$$T_{sol}(P_{gr}) < T(P_{gr}) < T_{vap}(P_{gr}) \tag{7.14}$$

Where T_{sol} is the temperature of solidification to ice at ground pressure, T_{vap} is the temperature of vaporization to water vapor at ground pressure, and $T(P_{gr})$ the temperature of the ground boundary layer as calculated from MAGRATHEA.

The values of T_{sol} and T_{vap} can be thus retrieved:

$$T_{sol}(P) = T_{TP} \cdot \exp \frac{P - P_{TP}}{H_{melt}\ V_{melt}} \tag{7.15}$$

$$T_{vap}(P) = \frac{H_{vap} T_{TP}}{H_{vap} - R\ T_{TP} \cdot log(P/P_{TP})} \tag{7.16}$$

Where T_{TP} is the temperature of the triple point (273.16 K), P_{TP} is the pressure of the triple point (611.73 Pa), H is the molar enthalpy of water during the melting process ($H_{melt} = 5.98\ kJ/mol$) and during the vaporization process ($H_{vap} = 44.9\ kJ/mol$), V the molar volume of water vapor during the melting process ($V_{melt} = -1.634\ cm^3/mol$) and the vaporization process ($V_{vap} = 22050\ cm^3/mol$), while R is the universal gas constant ($R = 8.314\ J/(mol \cdot K)$).

This approach is, in any case, quite simplified: there are many different types of ice crystals, which define various triple points temperatures and pressures. In this case, we interpreted the solid phase as composed by hexagonal ice (ice Ih), which constitutes all the natural snow on Earth.

The pressure-dependence of the liquidity range allows to properly consider the habitability of the surface while taking into account the atmosphere. This is important especially for high pressures, for which the range of liquidity extends over 400 K, allowing more models to be considered habitable.

However, this estimate cannot consider diurnal or seasonal variations because of the 1D nature of the radiative-convective model, which only takes into account averaged day/night fluxes and, at present, not tilted planets.

The majority of the models in the theoretical grid are still too cold to be habitable: the ground temperature is lower than the freezing point ($T_{sol}(P_{gr}) > T(P_{gr})$), most likely due to the cold trap of the carbon dioxide and the effective shielding in the visible range by the albedo. About 3200 models are habitable according to Eq. 7.14, while only 51 exceeded the temperature of vaporization.

A more detailed study on the parameters that have an impact on the habitability may be done by studying Figure 7.24. Here, the blue (crossed diagonal hatching) sample represents the normalized frequency of the cooler models, while the green (diagonal hatching sample) sample is the frequency of habitable models. Barely noticeable are the hotter models, whose frequency is shown in red (crossed hatching). In grey, the unsuccessful models are shown, for the sake of completeness. The choice of the bins was made to ensure the same number of models in each bin, as in the previous set of plots; the normalization was made simply to ease visualization.

Only the wet subgrid appears to have a portion of habitable (and warm) planets in the sample, while the dry subgrid reports only cool atmospheres. This was already noticeable in Figures 7.5 and 7.6 and it is clear from the "water vapor switch" subplot (fourth row, right subplot of the figure). This is, of course, a confirmation of what previously seen: the presence of water vapor heats the troposphere and therefore pushes the ground temperature towards warmer temperatures.

The semi-major axis (left subplot of the first row) seems to be loosely linked to habitability. It is, however, important to keep in mind that planets orbiting similar distances may receive very different irradiations, depending on the stellar temperature, so the variation of effective flux (right subplot of the first row) provides more interesting information: as foreseeable, higher values lead to a higher rate of habitable, and even hotter, models.

The stellar temperature (left subplot of the second row) has an impact on the habitability rate: more habitable planets appear around hotter stars. This is due to the larger absorption of the stratosphere in the infrared, which leaves less radiation to heat up the lower layers (as visible in the temperature subplot of Figure 7.4) for atmospheres of planets orbiting cooler stars. On the other hand, the effect of the albedo on the visible range avoids the runaway greenhouse effect of water vapor by reducing the incoming flux on hotter stars, thus allowing to reach convergence in the habitable range.

A non-null atmospheric emissivity (right subplot of the second row) increases the frequency of habitable planets, by making the ordinary greenhouse effect more effective. To this category belong the hotter models, too, barely noticeable in the frequency histogram, showing that sometimes the greenhouse heating provided by the atmosphere can lead to hotter tropospheres as well.

Lower albedo values (left subplot of the third row) increase the incoming flux and therefore contribute to heat a larger portion of atmosphere to habitable surface temperatures.

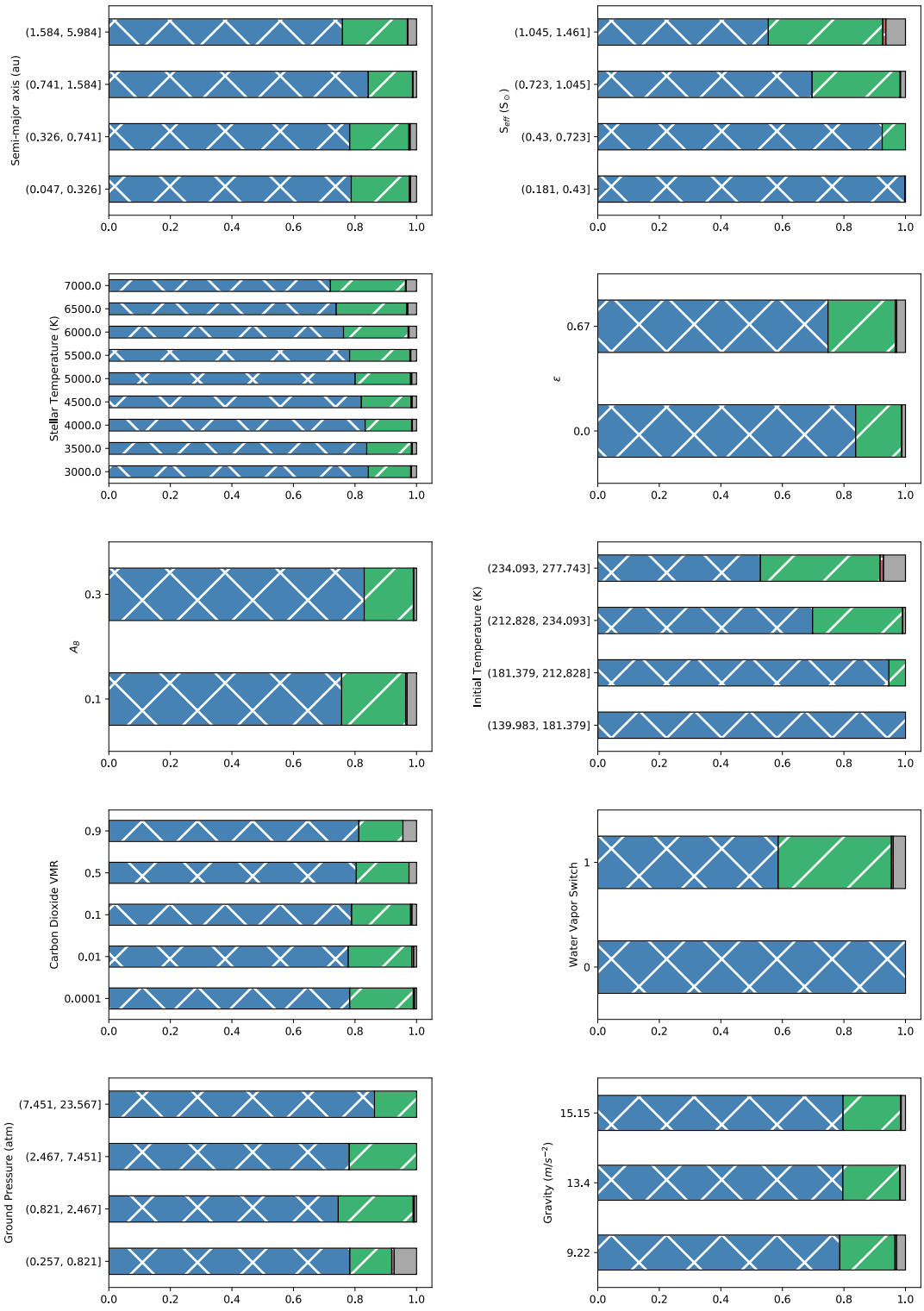


Fig. 7.24: Stacked histograms for the normalized frequency of habitable (green, diagonal hatching sample), cooler (blue, crossed diagonal hatching sample) or hotter (red, crossed sample) models with respect to a few input parameters. The unsuccessful models are shown in grey.

The initial temperature contributes a lot to the habitability, as it collects all the aforementioned effects of the emissivity, albedo, and stellar incoming flux. Since it controls the abundance of water vapor from the very first iteration, it is very likely to find that cooler initial temperatures are basically dry (even when water vapor is in the atmosphere) and carbon dioxide has the burden of heating the troposphere all on its own. As previously seen, the carbon dioxide alone has a larger impact on the stratosphere of an atmosphere and often leads to an average cooling. One should then not be surprised that, the hotter the initial temperature, the more likely it is to have habitable planets. On the other hand, this leads to a higher production of water vapor which may push the temperatures to higher values.

Carbon dioxide by itself (left subplot of the fourth row) does not contribute massively to the habitability. A slightly higher rate of habitable planets (and hot ones, as well) is linked to lower values of CO_2 since the atmosphere is less effective in absorbing the incoming flux, which can heat up the lower layers and produce more water vapor, whose impact is indeed stronger. The carbon dioxide effect should be more massive in producing cooler planets if CO_2 condensation is included, varying the albedo values accordingly and thus allowing to simulate a Snowball Earth-like situation.

Variation in pressure (left subplot of the fifth row) determines the definition of the thermodynamic habitability: at higher pressures, it is more likely to have cooler planets rather than hotter ones, since the liquidity range is translated towards higher temperatures. No sensible variation appears in the last subplot (the surface gravity), instead.

Overall, the large number of cooler planets suggests that the atmosphere does contribute to determining the habitability of the planet, in a careful balance between thermal regulation and runaway variations. In any case, the treatment of the albedo and the emissivity as constants and not as wavelength-dependent variables may have lead to a slight underestimate of the incoming flux, which is responsible for the translation of the profiles towards hotter temperatures.

7.3 Observed Super Earths

7.3.1 Input

In Section 5.11 a sample of currently known terrestrial exoplanets with an accurate measurement of mass and radius was retrieved. Such targets are predicted to have fairly high irradiations so that in principle none of them could be suitable for life.

The effect of atmospheric composition on real planets could, however, set more refined boundaries concerning the actual habitability of these targets. As of now, there are very few observed spectra of terrestrial exoplanets, so a statistical analysis could not help but using theoretical simulations.

Furthermore, such studies are essential to be able to compare theory and observations, since new data are going to be acquired by new-generation facilities.

For this Super Earths grid, the very same sample analyzed in Section 3 is used as a basis. All physical parameters (planetary and stellar dimensions and semi-major axes) are set equal to their measured values. Semi-major axes values that were not available were theoretically retrieved from the period measurements considering a null eccentricity.

The stellar temperatures were retrieved through a TAP query on SIMBAD and Gaia DR2 catalogs, using either the main identifier provided by `Exo-MerCat` or the coordinate tuples. From those values, the stellar radii were retrieved using Eq. 7.5.

On the other hand, it was possible to select a few atmospheric parameters, to model many combinations of atmospheric compositions, as shown in Table 7.7. In particular, for every simulated planet, three different atmospheric masses were considered, the corresponding ground pressures being retrieved by Eq. 7.4. Two Bond albedo values were considered, corresponding to a Mars-like and Earth-like albedo (0.1 and 0.3 respectively).

Furthermore, both dry and moist atmospheres were modeled, with varying carbon dioxide abundances. Chemical compositions of primordial atmospheres were considered, with no oxygen in the mixtures, as well as more Earth-like atmospheres, with a fixed molecular oxygen abundance of 20%.

Tab. 7.7: Atmospheric parameters of the observed Super Earths grid of models.

Parameters	**#**	**Notes/Values**
X_{CO_2}	5	0.0001, 0.001, 0.01, 0.1, 0.9
X_{O_2}	2	0, 0.2
Pressure Factors α	3	0.1,1,10
Albedo	2	0.1, 0.3
Emissivity	2	0, 2/3
Water vapor	2	dry, wet (RH 60%)

As previously seen, MAGRATHEA stops whenever the temperature exceeds 500 K in any atmospheric layer, or when the water vapor exceeds its limiting abundance determined by the volume mixing ratios of the remaining components.

Since the Super Earths in the sample receive high irradiations from their host stars due to their short distances, shown in Figure 5.17, it is likely to have equilibrium temperatures which would exceed by far the 500 K threshold from the very first iteration.

To spare as much computational time as possible, not all planets in the sample were selected for the modeling grid: the ones whose input temperature (Eq. 6.14) exceeded 500 K for any combination of emissivity and albedo were excluded from further analyses.

The subsample of observed Super Earths remaining after the application of this filter (i.e. the ones that showed initial temperature less than 500 K for at least one combination of emissivity and albedo values) was composed by eleven targets: LHS 1140 b, K2-18 b, K2-3 b, Kepler-48 b, GJ 1132 b, TRAPPIST-1 b, TRAPPIST-1 c, TRAPPIST-1 g, K2-266 e, K2-155 d, and Kepler-138 d.

Also, keeping in mind that the combination of 0.9 CO_2 volume mixing ratio and 0.2 O_2 volume mixing ratio (for a total of 264 occurrences) would cause the model to fail from the very beginning, being the total volume mixing ratio of such models higher than 1, even without taking into account water vapor or nitrogen, we excluded such models *a priori*.

Therefore, the grid of atmospheric models is composed of 2376 models.

7.3.2 Analysis and results

Despite the preliminary selection on the initial temperature, this grid of models shows many more unsuccessful models than the previous one. This is expected since all modeled planets reside within the inner boundary of the habitable zone and are therefore expected to receive much higher irradiations from their host stars.

The presence of various abundances of carbon dioxide, water vapor, as well as oxygen, is likely to drive towards even hotter temperatures the modeled atmospheric profiles. We, therefore, expected to find out models that exceeded the threshold of 500 K (which were absent from the previous grid of models, more temperate by construction), as well as atmospheres that would not be stable within the current assumptions of the software.

For these reasons, only a third of the grid reached convergence (the number of converged models for each planet - out of a total of 216 per target initially set up - is shown in Table 7.8, while a graphic representations of the frequency of successful models per planets in a stellar temperature versus irradiation or semi-major axis plot are shown in Figures 7.25 and 7.26.

Tab. 7.8: Atmospheric parameters of the observed Super Earths grid of models.

Exoplanet	# Successful
TRAPPIST-1 g	216
TRAPPIST-1 c	139
TRAPPIST-1 b	121
K2-18 b	120
K2-155 d	109
Kepler-138 d	73
K2-3 c	61
Kepler-48 d	25
LHS 1140 b	14
GJ 1132 b	0
K2-266 e	0

Two planets (GJ 1132 b and K2-266 e) failed with every combination of parameters: these were, in any case, the most irradiated planets, with an input temperature higher than 400 K.

We will proceed to analyze the output models with a similar approach as the previous grid: the qualitative analysis of the results is described in Section 7.3.2.1, the principal component analysis in Section 7.3.2.2.

In this case, the analysis of the failed models (Section 7.3.2.3) will be extremely useful to understand under which combinations of initial parameters the modeled atmospheres can be stable or unstable.

Information on the habitability of the successful models is then treated in Section 7.3.2.4.

7.3.2.1 Qualitative Analysis

This time, the physical parameters such as planetary mass, stellar temperature, and distance are set by observational constraints: so, this grid of models cannot handle any uniform variation of such parameters. The possibility to consider many combinations of chemical composition, albedo, and emissivity, can, however, be

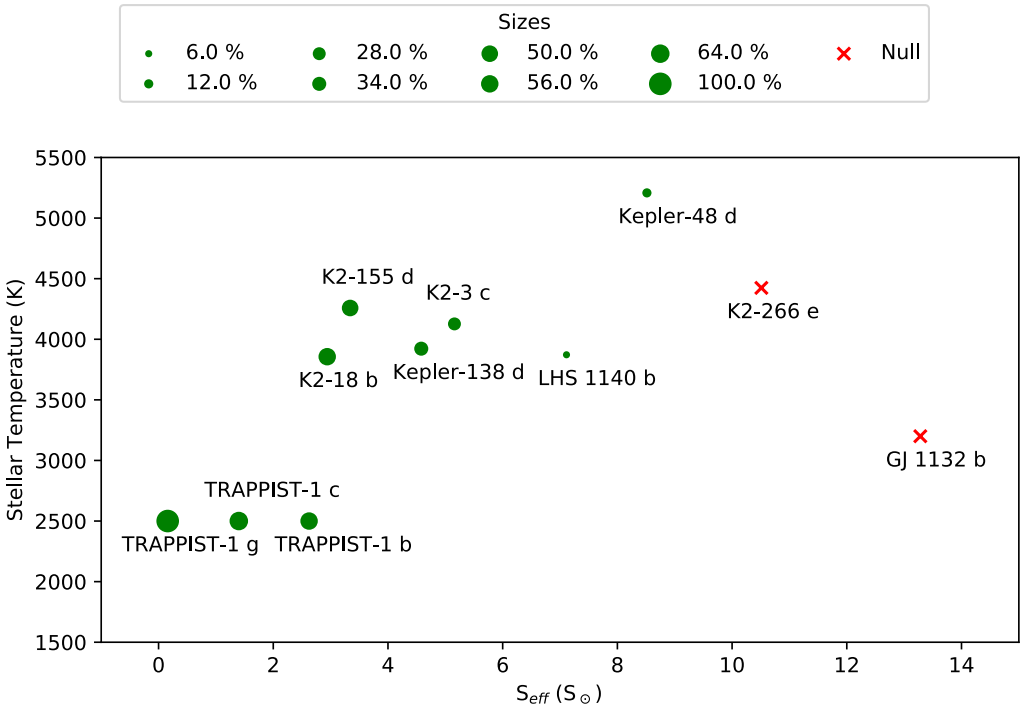


Fig. 7.25: Stellar temperature vs effective stellar flux for the eleven modeled planets. The size of the dots show the frequency of successful models (out of a total of 216). The red *x* markers represent the planets for which no models were successful.

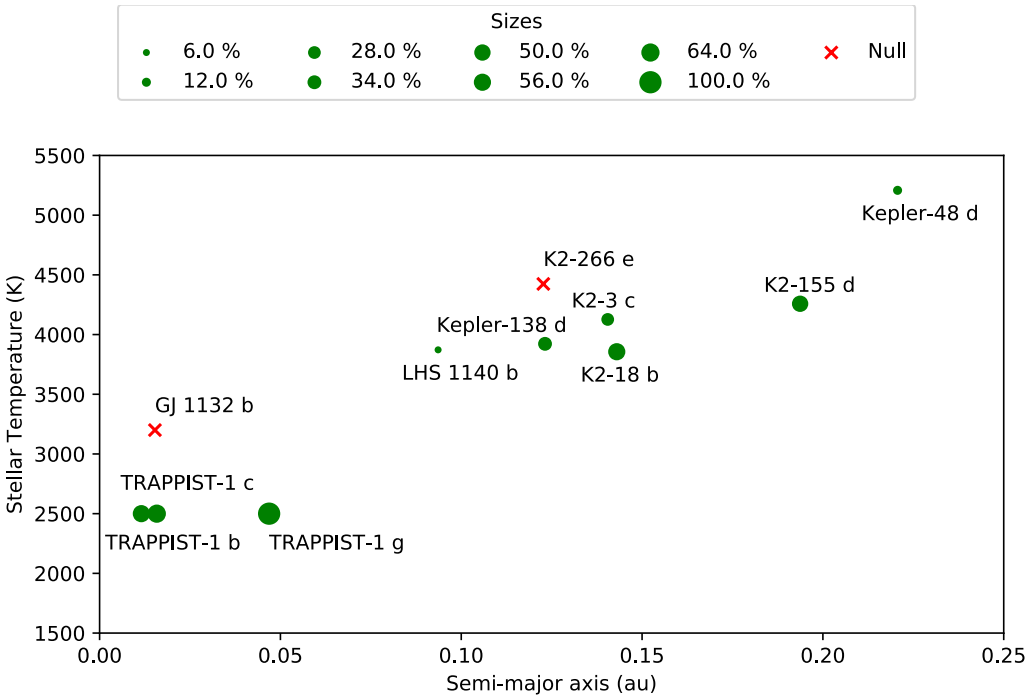


Fig. 7.26: Stellar temperature vs semi-major axis for the eleven modeled planets. The size of the dots shows the frequency of successful models (out of a total of 216). The red *x* markers represent the planets for which no models were successful.

useful both to test the performances of MAGRATHEA on a different regime than the previous case and to infer some information that would be useful for future characterization on such observed targets, or similar ones.

To perform a qualitative analysis of the successful models, as we previously did in Section 7.2.2.1, we, therefore, need to choose first of all a modeled planet – to set the physical parameters in the grid. We selected TRAPPIST-1 b, the third most successful planet, with 121 converged models (out of 216). We preferred this target, rather than the most successful one (TRAPPIST-1 g) because of its higher irradiation, which is most likely to lead to hotter profiles, which would magnify the contribution of water vapor. This species is, as stated previously, highly temperature dependent, so modeling wet profiles at lower irradiation would reduce the overall abundance of this compound, whose effect would not be detectable in the profile.

Then, we fixed the remaining parameters to have a subset of successful models (all shown in Table 7.9).

Tab. 7.9: Subset of selected parameters for the qualitative analysis (target: TRAPPIST-1 b).

M_P ($M_\oplus$)	$T_\star$ (K)	a (AU)	A_B	ϵ	α	X_{CO_2}	RH (%)	X_{O_2}
1.0	2500	0.012	0.3	0	10	0.001	0	0

The subsample of models, divided by varying parameter, is shown in Figure 7.27. Of course, being the actual subgrid composed only by combinations of the values of albedo, emissivity, pressure, and chemical volume mixing ratios abundances, only those subplots are shown.

Overall, this picture seems to confirm the already analyzed behavior of MAGRATHEA, especially for what concerns the albedo, the emissivity, the pressure factor, and the carbon dioxide abundance.

Even though this subset is composed by dry models, many of the profiles are unstable for convection in the lower layers of their atmospheres, a process that did not appear in the analog plot of the previous grid (Figure 7.2). The convection is activated even in absence of oxygen (which changes the specific heat of the dry components) and of water vapor (which introduces a new term in the retrieval of the critical lapse rate).

It is, however, important to keep in mind that the two grids are covering different regions of the stellar temperature vs irradiation plot: TRAPPIST-1 b receives a stellar irradiation of nearly 2.5 times the solar constant, while the irradiation corresponding to the most similar combination modeled by the theoretical grid was barely equal

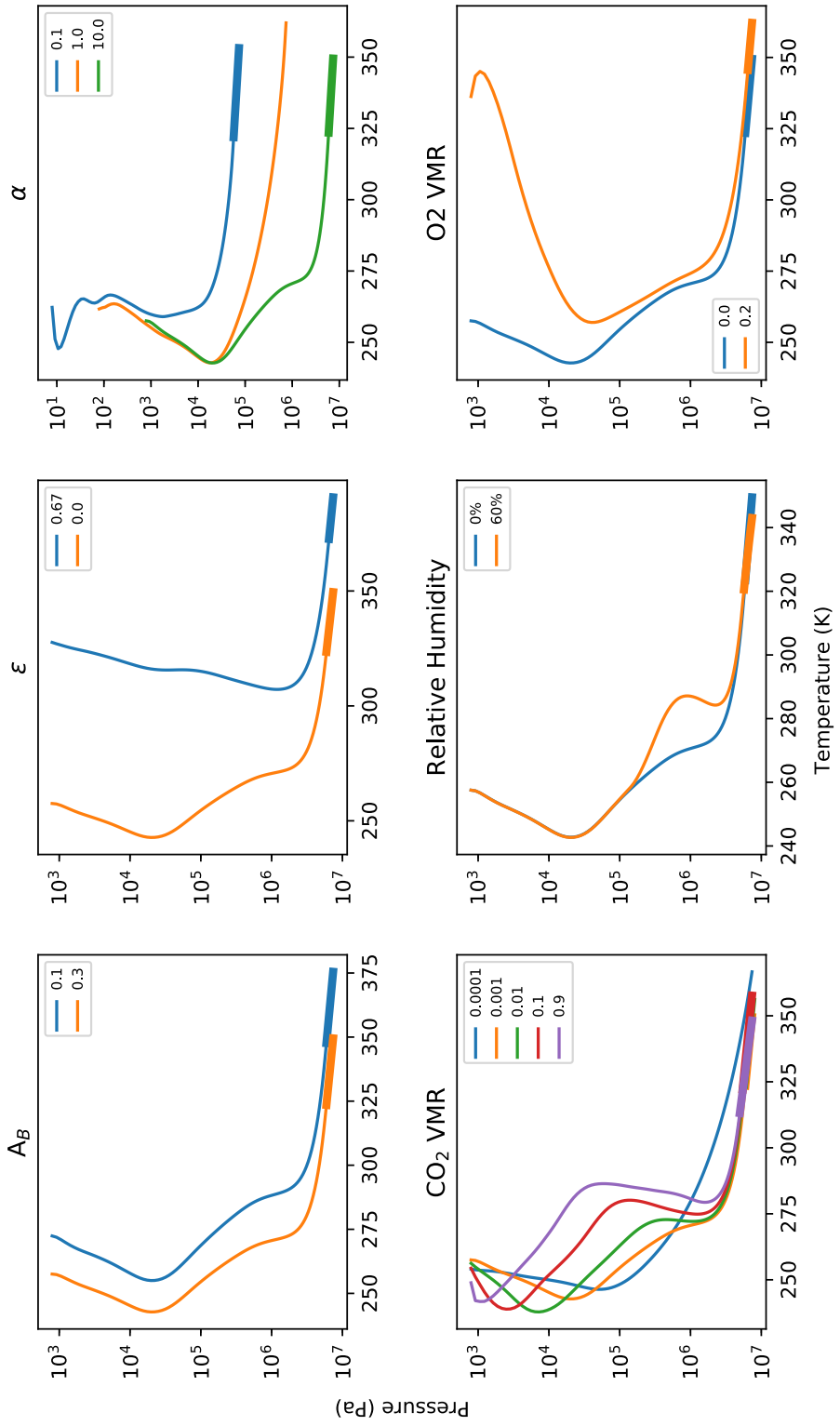


Fig. 7.27: Set of subplots with varying parameters, based on Table 7.9 (TRAPPIST-1 b).

to one solar constant. Furthermore, the lower effective temperature of TRAPPIST-1 corresponds to a peak of the blackbody radiation towards longer wavelengths, with a more efficient absorption by carbon dioxide.

The introduction of water vapor in the mixture (third subplot of the second row in Figure 7.27) does not produce the massive change in the tropospheric temperature as it appeared in the previous case, but has a similar behavior to that of the CO_2: in the first wet layers of the atmosphere, a high amount of water is produced and absorbs the majority of the incoming flux, leading to slightly lower ground temperatures.

The effect produced by molecular oxygen in this large amount in the atmosphere leads to a higher absorption on the upper layers, forming a much more pronounced temperature inversion, whose maximum almost equals the ground temperature. The radiative flux is exhausted at about half of the atmosphere, then leading to thermally-heated lower layers.

7.3.2.2 Principal Component Analysis

A principal component analysis can be performed on the sample of successful models. A similar approach to the one previously explained (see Section 7.2.2.2) is used, to reduce the dimensionality of the sample and to be able to understand the parameters that explain the majority of the variance. The expected results should be, overall, in agreement with the ones already found; in addition to those, however, this grid of models allows to understand the contribution caused by molecular oxygen (not present in the atmospheres of the theoretical set), as well as the impact that larger values of irradiation can have on the shape of the profiles.

The smaller number of successful models (878 out of 2376, which translates into 878 points in the space of the principal components) is very likely not to uniformly cover the parameter space: for this reason, another theoretical grid spanning this region of parameter values should be useful to determine more precise results in the future.

In this case as well, the first four principal components describe the majority of the variance for the observed Super Earths sample, as shown in Figure 7.28: the first component describes the 87.9% of the variance, the second one the 8.2%, the third one the 2.5% and the fourth one the 1% of the total variance. The shape of each component is very similar to the ones depicted in Figure 7.9.

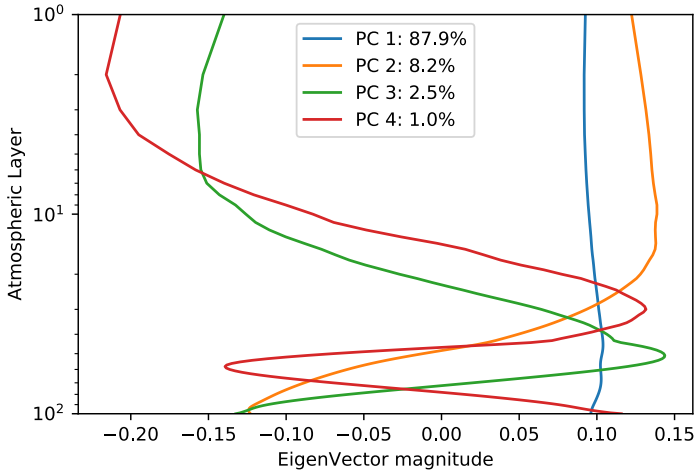


Fig. 7.28: First four principal components (and corresponding variance) of the theoretical grid as a function of the atmospheric layer, from first modeled layer (NL=1) to the ground level (NL = 100).

We represent the sample in the principal components space with the usual 2D corner plots. The sample is color-coded depending on a few interesting parameters, as shown in Figures from 7.29 to 7.37.

A similar study on the parameter-depending clustering of the dataset can be now made. It is possible to infer the following information:

- The incoming stellar flux at the top of the atmosphere in units of Wm^{-2} determines the first and the second principal components (thus describing nearly the 96% of the variance). This is visible in the top left subplot of Figure 7.29, in which the models experiencing lower fluxes are clustered in the upper left part of the graph. It is important to notice the difference in order of magnitude of the maximum flux value with respect to the analog plot of Figure 7.11: indeed, these models can experience up to six times more flux with respect to the previous case; the minimum-maximum difference, in this case, spans two orders of magnitude, compared to only one in the theoretical grid.

- The peak wavelength of the host star in μm contributes to determining the second component (Figure 7.30). This is caused by the larger efficiency in absorbing infrared photons by the atmosphere (further enhanced if O_2 is present in the mixture). However, this second-order dependence ensures that even a higher absorption does not cause extreme variations in the average temperature of the profile (which is determined by the flux at the top of the atmosphere), but rather on the shape of the profile. The larger number of

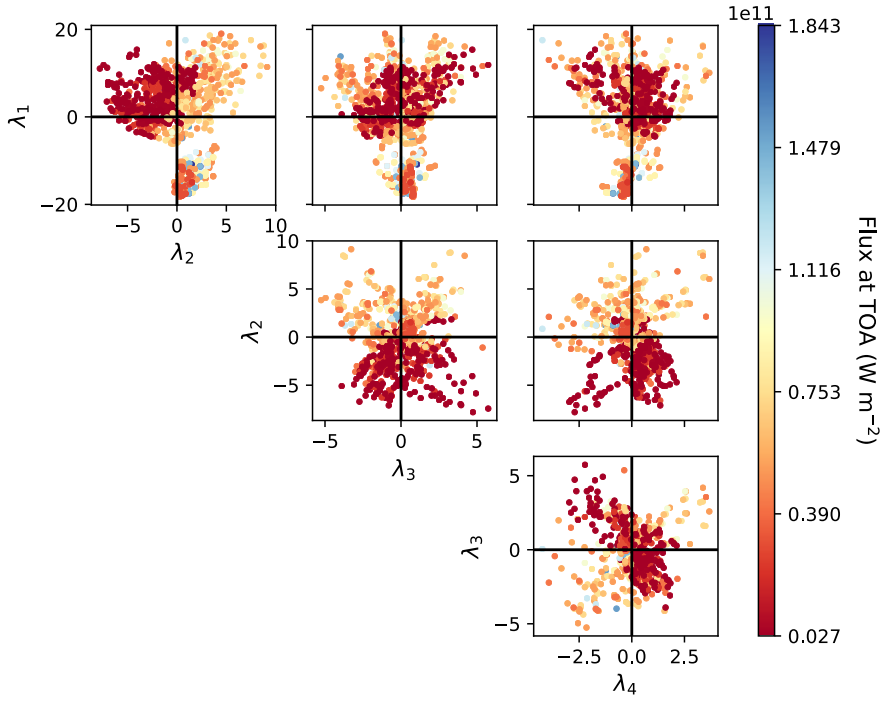


Fig. 7.29: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the flux at the top of the atmosphere.

modeled atmospheres on planets orbiting cooler stars (the blue cluster) over the ones orbiting hotter stars (the orange-red cluster) is due to the higher success rate of the first set of atmospheres, as previously seen.

- The emissivity of the top layer (Figure 7.31) does not appear to influence any component. On the other hand, there is a lower number of successful models at non-null emissivity, which underlines some effect of the emissivity value on the rate of success: therefore, this parameter influences the overall sample of models, but the information cannot be traced in the principal component analysis. The study of the failure rate in Section 7.3.2.3 addresses this issue as well.

- The initial temperature (Figure 7.32), which takes into account the flux at the top of the atmosphere and the emissivity of the atmosphere itself, echoes the behavior of the first parameter, determining both the first and the second component. Much higher input temperatures are explored in this sample with respect to the previous grid, as expected from the larger values of stellar irradiation.

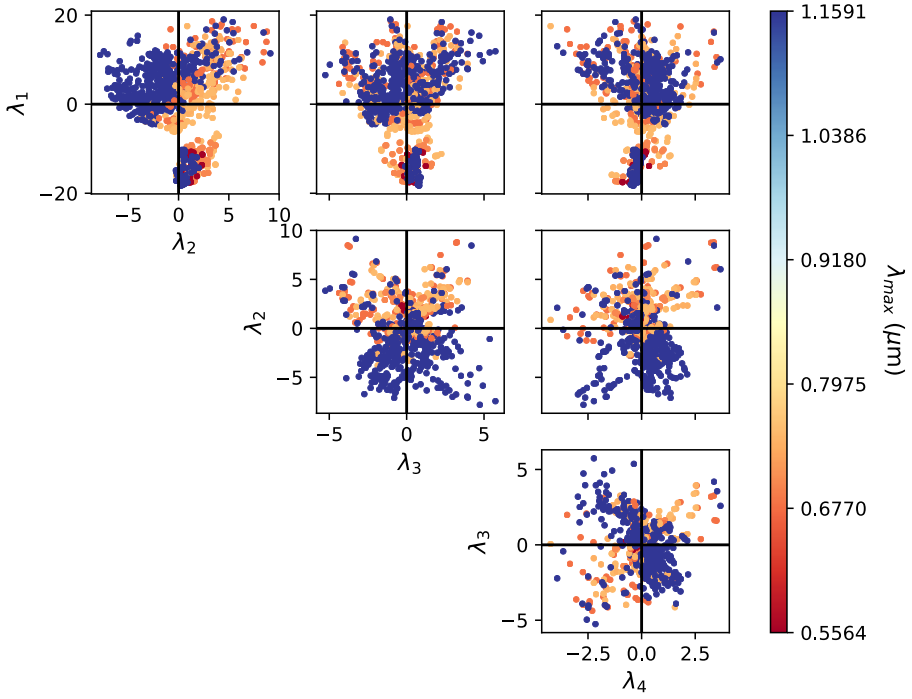


Fig. 7.30: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the peak wavelength determined by Wien's displacement law for all stellar effective temperatures.

- The presence or absence of water vapor (in Figure 7.33 displayed in terms of the *water switch*, which is 1 for all wet models and 0 otherwise) does not appear to contribute to any component. In this case, a similar bias as the one noticed in the emissivity color-coded plot appears: the majority of successful models is composed of dry atmospheres, while the moist models are but a smaller fraction. So, water vapor as well contributes to the success or failure of the atmospheres, most likely by driving the temperature towards higher values.

  Studying only the portion of models with water vapor in their atmospheres (Figure 7.34), the effect of water on the first and second component can be appreciated. Similarly to the theoretical grid of models, water vapor abundance is driven by the incoming flux, which increases the temperature of the atmosphere. Positive feedback is expected to be triggered in these cases as well, enhanced by the larger irradiance, leading to much more extreme results (and possibly failing, because of that).

- The carbon dioxide abundance (Figure 7.35) does not appear to contribute to any component either. However, only the models at low abundances of

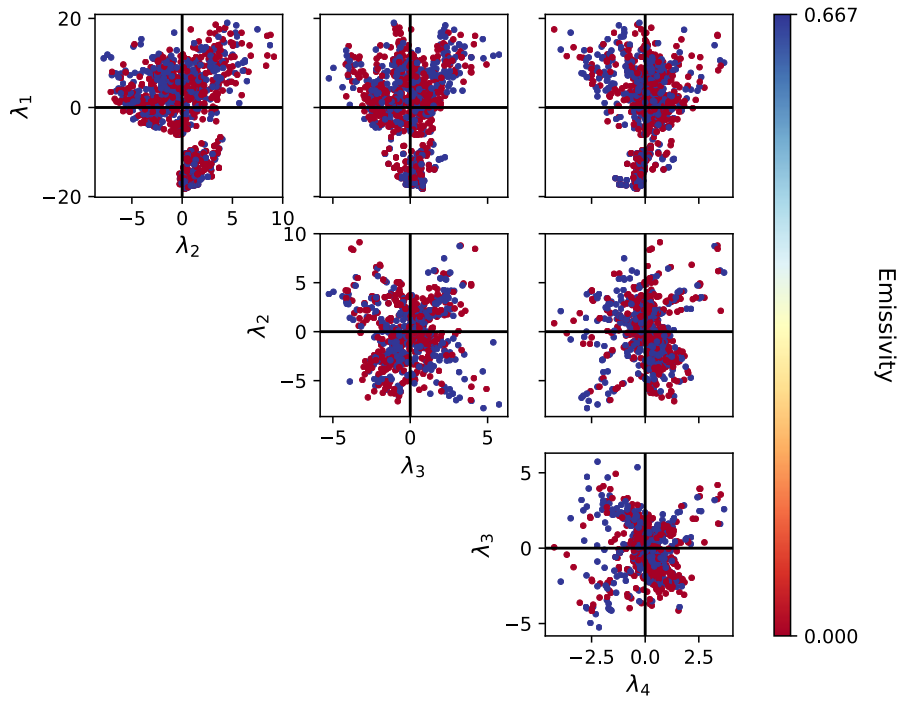


Fig. 7.31: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the emissivity at the top of the atmosphere.

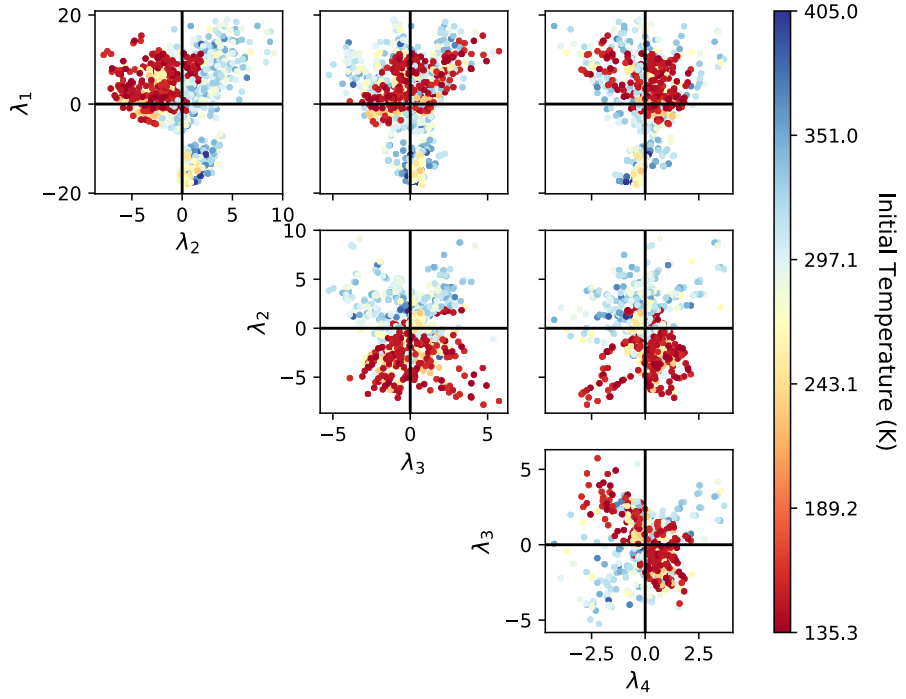


Fig. 7.32: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the initial temperature.

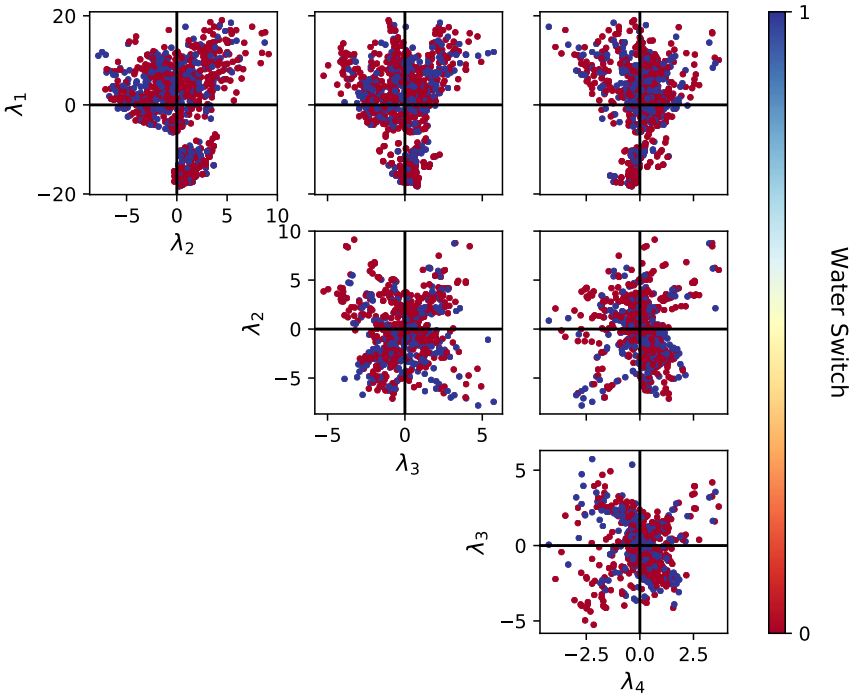


Fig. 7.33: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the water vapor switch (0 for the models with no water vapor in the atmosphere, 1 for the models with water vapor in the atmosphere).

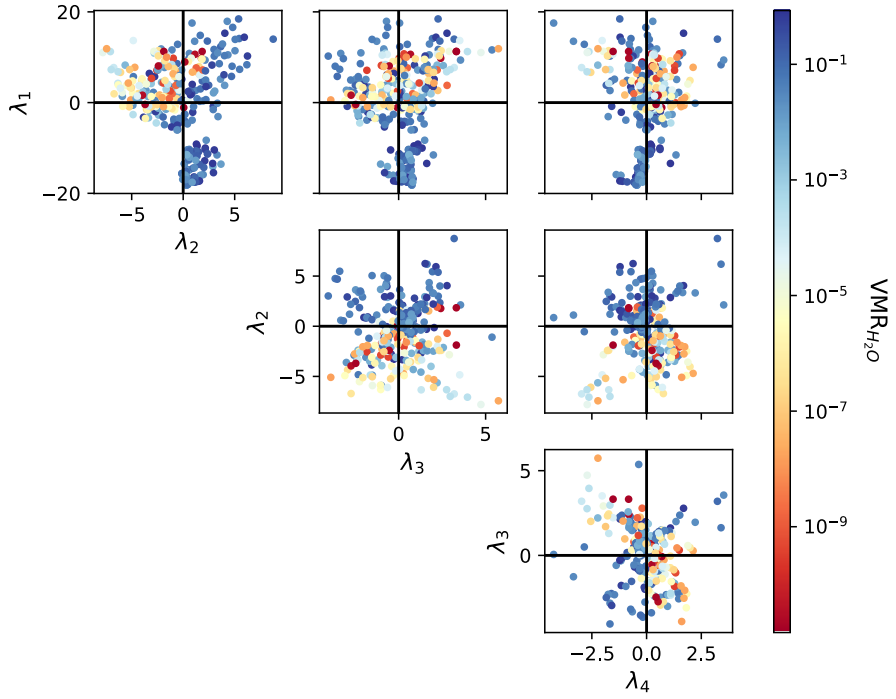


Fig. 7.34: Scatter plot of all combinations of the eigenvalues of the first four principal components **for the subsample of wet models**. The colors are determined by water vapor abundance at the ground boundary layer in logarithmic scale.

carbon dioxide are more likely to have succeeded. This is another aspect of the failed models issue: higher CO_2 abundances lead to a smaller tolerance on the maximum abundance of water vapor and therefore on a higher chance for the atmosphere to be unstable. Moreover, a portion of models at 90% CO_2 and 20% O_2 could not be run because they would have lead to instantaneous instability: the models at a high abundance of carbon dioxide are, by construction, fewer than the other cases.

For the successful models at lower carbon dioxide volume mixing ratio, this value does not influence the overall variance of the sample.

- The oxygen abundance (Figure 7.36) influences the second principal component. In this case, both subsamples with and without O_2 in their atmospheres (the first being the blue cluster, and the latter being the red one), appear to be much more populated than the previous cases, even though the first is slightly less populated than the latter. Being the nitrogen volume mixing ratio calculated, in this case, as $X_{N_2} = 1 - X_{CO_2} - X_{O_2} - X_{H_2O}$, the molecular oxygen contributes as well to reduce the tolerance on the maximum abundance of

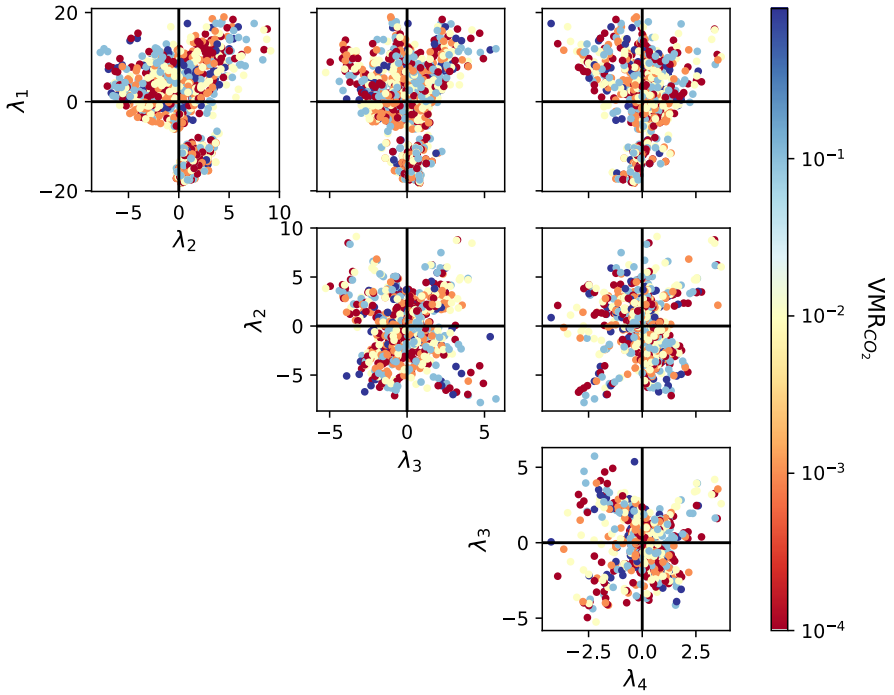


Fig. 7.35: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the abundance of carbon dioxide in logarithmic scale.

water vapor. This, however, does not appear to produce dramatic results in this case, with respect to the other species.

- The abundance of all the absorbing gases (the sum of carbon dioxide, molecular oxygen, and water vapor) influences the second and the fourth component, as visible in Figure 7.37. In this case as well, there is a lack of successful models at a high abundance of absorbing gases. A cluster of intermediate values ($X_{abs} \approx 0.2 - 0.4$, orange sample) appears and is differentiated from the models at lower abundances (red sample): this should represent the contribution of the oxygen in the mixture.

Overall, this principal component analysis underlines not only what is clearly driving the shape and the average temperature of the successful models, but also - and perhaps more importantly - what parameters appear to affect the success rate of such irradiated models. A more detailed study of the whole sample is extremely useful and will be treated in the next Section.

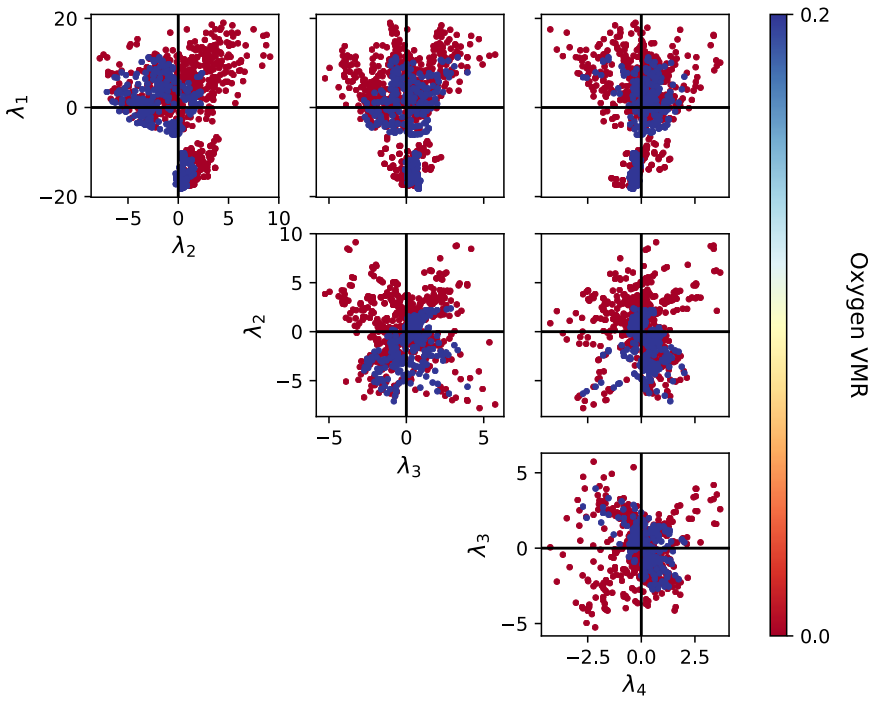


Fig. 7.36: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the total volume mixing ratio of O_2.

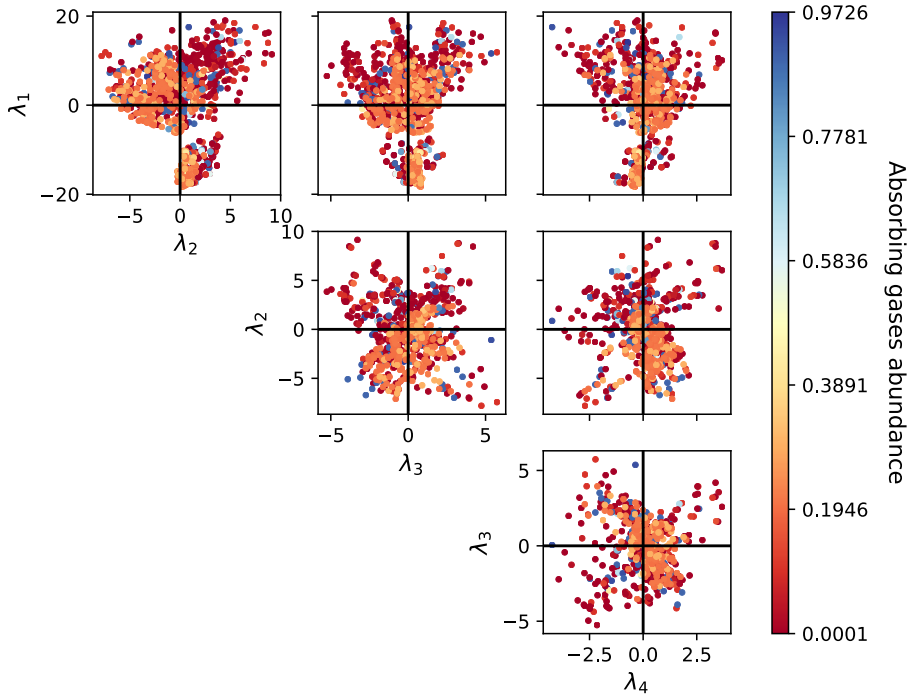


Fig. 7.37: Scatter plot of all combinations of the eigenvalues of the first four principal components. The colors are determined by the total volume mixing ratio of the absorbing gases.

7.3.2.3 Instability

In this grid of observed Super Earths, the majority of the models appear to have failed, contrary to the previous case. It is therefore interesting to understand which parameter range is best for the stability of the radiative-convective profiles or influence their temperature range.

A large number of models failed not only because of instability (negative expected nitrogen abundance), as in the previous grid, but also because of the exceeding temperature (trespassing the 500 K threshold): this issue was indeed foreseeable, because of the much higher initial temperatures, as well as irradiances, which were likely to trigger a runaway production of water vapor in the troposphere, rapidly leading either to unstable atmospheres (if the expected H_2O volume mixing ratio breaks the mass conservation), or higher temperatures which cannot at present be modeled by `MAGRATHEA`. If water vapor is present in such models, in principle, it is still likely that this class of models will nevertheless be unstable due to the large abundance of water vapor, but nothing quantitative can be said about them at the time of writing.

In Figure 7.38 the normalized frequencies of successful models (green, crossed diagonal hatching sample), the ones exceeding the 500 K threshold (red, diagonal hatching sample), and the unstable ones (yellow, plain sample), with respect to some input parameters are shown. Because of the necessity of keeping the number of models per bin equal for each subplot, and being this grid not uniform by definition, we report also the models that couldn't be stable from the very first iteration, showing them with the crossed hatching portion of the yellow sample.

The variation of the semi-major axis (top left subplot) does not appear to influence uniformly the rate of success or failure. It is essential to keep in mind that, however, this sample compares a much larger range of irradiances, so similar values of the semi-major axis correspond to extremely different fluxes at the top of the atmosphere (as visible by comparing Figures 7.25 and 7.26).

The variation of effective stellar flux (top right subplot) is indeed responsible for a larger number of failed models: for values higher than 8 $S_\odot$, the majority of the models are too hot to be modeled. The percentage of unstable models (yellow sample) also increases with higher irradiances: the hotter the temperatures, the more water vapor abundance produced, the more likely to overcome the maximum abundance for an atmosphere to be stable. On the other hand, the largest rate of success appears for the lowest irradiances, which also refer to planets nearer to the habitable zone range of their host stars.

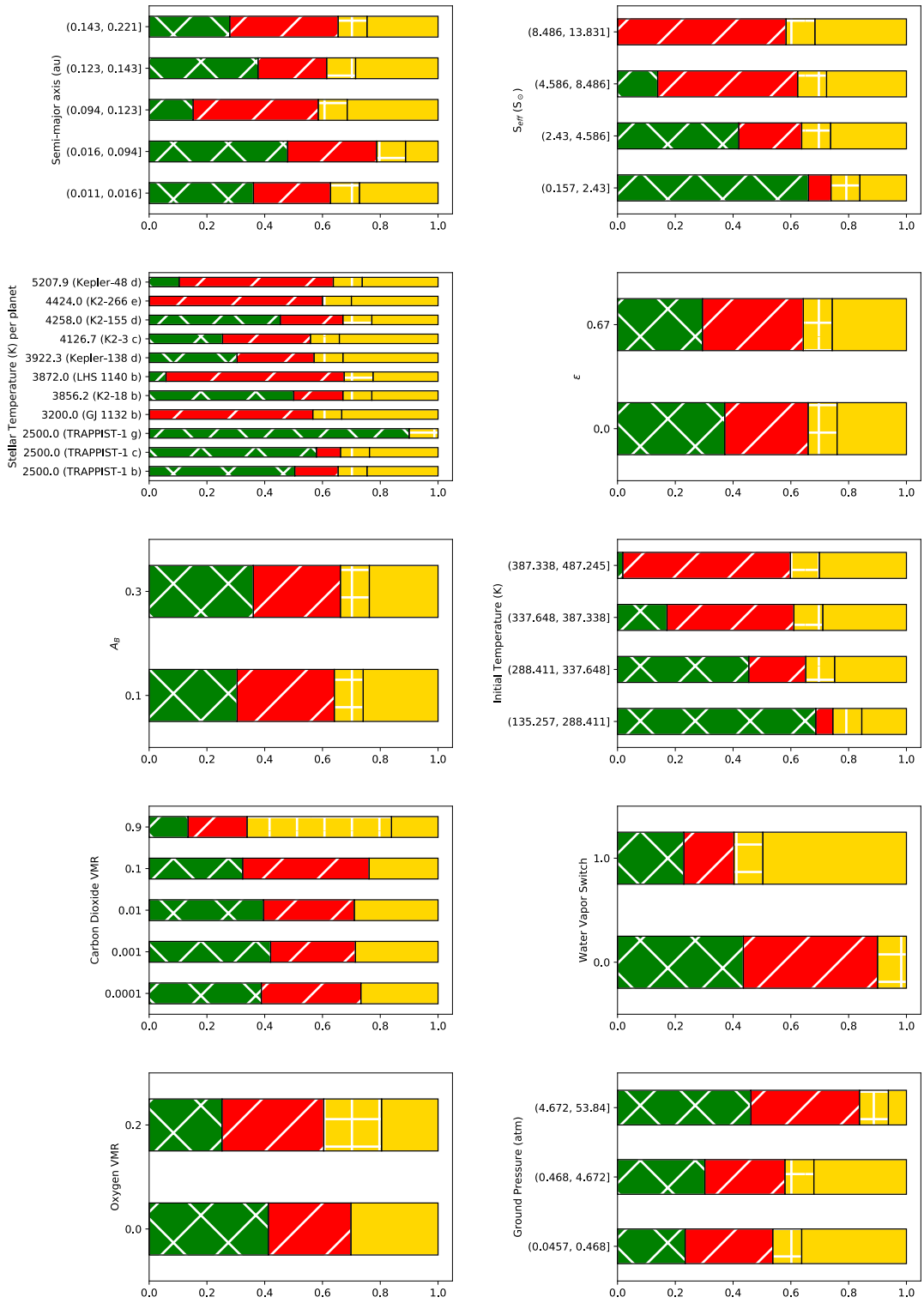


Fig. 7.38: Stacked histograms for the normalized frequency of successful (green, crossed diagonal hatching sample), hotter (red, diagonal hatching sample), or unstable (yellow, plain sample) models with respect to a few input parameters. The models that were unstable from the first iteration are shown in yellow, crossed hatching sample.

The stellar temperature variation (left subplot of the second row) does not seem to play a considerable role: differently from the previous grid, the large variation in irradiance can hardly allow a comparison between the incoming fluxes, and even less the change in wavelength peak of the blackbody spectrum, which influences lower-order variations. This is visible when studying the three TRAPPIST-1 planets: for the same star ($T_\star = 2500\ K$), a different rate of successful and unsuccessful models appears, based on the semi-major axes of the three planets.

The change in emissivity of the planet from null to 2/3 (right subplot of the second row) causes an increase in the rate of hotter models at the expenses of the successful ones: this is caused by the higher absorption of the atmosphere which increases its average temperature. The effect does not, however, produce a high variation in the frequency of unstable models. A similar variation can be seen in the albedo normalized histogram (left subplot of the third row), which shows a higher number of successful planets for the models at albedo 0.3, which therefore receive less radiation and cannot be heated as much as the other case.

The initial temperature (right subplot of the third row) sums up all of the previous effects since it takes into account both the effective stellar flux, the emissivity, and the albedo. Lower temperatures lead to more successful models: this is indeed based on the lack of water vapor in the atmosphere, whose abundance in this range of temperatures is extremely low. These models are very unlikely to be unstable and to reach much higher temperatures since no runaway effect takes place. However, this should pose some issues concerning the habitability of such planets, as will be studied in Section 7.3.2.4. Higher temperatures, as foreseeable, cause the models to be too hot to be modeled or unstable due to the massive abundance of H_2O.

The carbon dioxide abundance (left subplot of the fourth row) does not seem to have a strong impact on the rate of success or failure, apart from the largest abundance which allows less tolerance for the water vapor abundance.

The presence of water vapor in the atmosphere (right subplot of the fourth row) produces a larger number of unstable models since this issue involves exclusively the wet models by construction. Nearly half of the wet models are unstable. On the other hand, half of the dry model failed because of an excessive temperature: this underlines the contribution of carbon dioxide and oxygen in heating the atmosphere.

The presence of oxygen in the atmosphere (left subplot of the fifth row) contributes to the heating of the atmosphere itself, leading to a higher chance of falling in the category of the atmosphere too hot to be modeled.

The variation of atmospheric mass (the pressure factor in the right subplot of the fifth row) causes a variation in the rate of unstable models. This derives from the calculation of the water vapor mixing ratio from the partial pressure of water vapor, which leads to higher values for lower pressures. By considering only the remaining models, for all values of pressure nearly half of them appear successful, while the other half overcomes the 500 K threshold in at least one boundary layer.

7.3.2.4 Habitability

Using the technique explained in Section 7.2.2.4, it is still possible to retrieve information concerning the habitability of the modeled atmospheres. This is even more interesting than before, since this, however small, grid spans a larger range of stellar fluxes, which drive the largest variations in the profiles.

The presence of oxygen in the mixture is also interesting for habitability purposes: this species is one of the most important biosignatures and therefore is expected to be observed on a potentially habitable planet. However, the presence of molecular oxygen does contribute to the variation of the equilibrium profile and, by consequence, to the habitability.

The frequency of habitable planets (green, diagonal hatching sample), cool planets (blue, crossed diagonal hatching sample), and hot planets (red, plain sample) with respect to some input parameters is shown in Figure 7.39, while the unsuccessful models are represented in grey. For better visualization of only the successful models, a magnification of this subsample is shown in Figure 7.40: it is important to keep in mind that here, the number of models per bin over which the normalization is performed is no longer the same.

This time, many more habitable and hotter planets are present, due to the higher irradiations that the planets in this grid experience: this is visible in the right subplot of the first row (effective stellar flux), which shows a totality of unhabitable, cool planets at very low values of S_{eff}, then a higher rate of habitable planets for irradiances similar to the solar constant. A larger number of warm models increases with the increase in stellar constant. This effect is not clearly visible in the first plot (semi-major axis), as previously seen, because of the varying irradiations that reach planets at similar distances from their stars. As a matter of fact, the presence of habitable simulated planets is not straight-forward, since the values of the stellar effective flux S_{eff} are all too high to be included in the habitable zone, as defined by Kopparapu et al. (2013): this underlines the strong impact that a more reliable

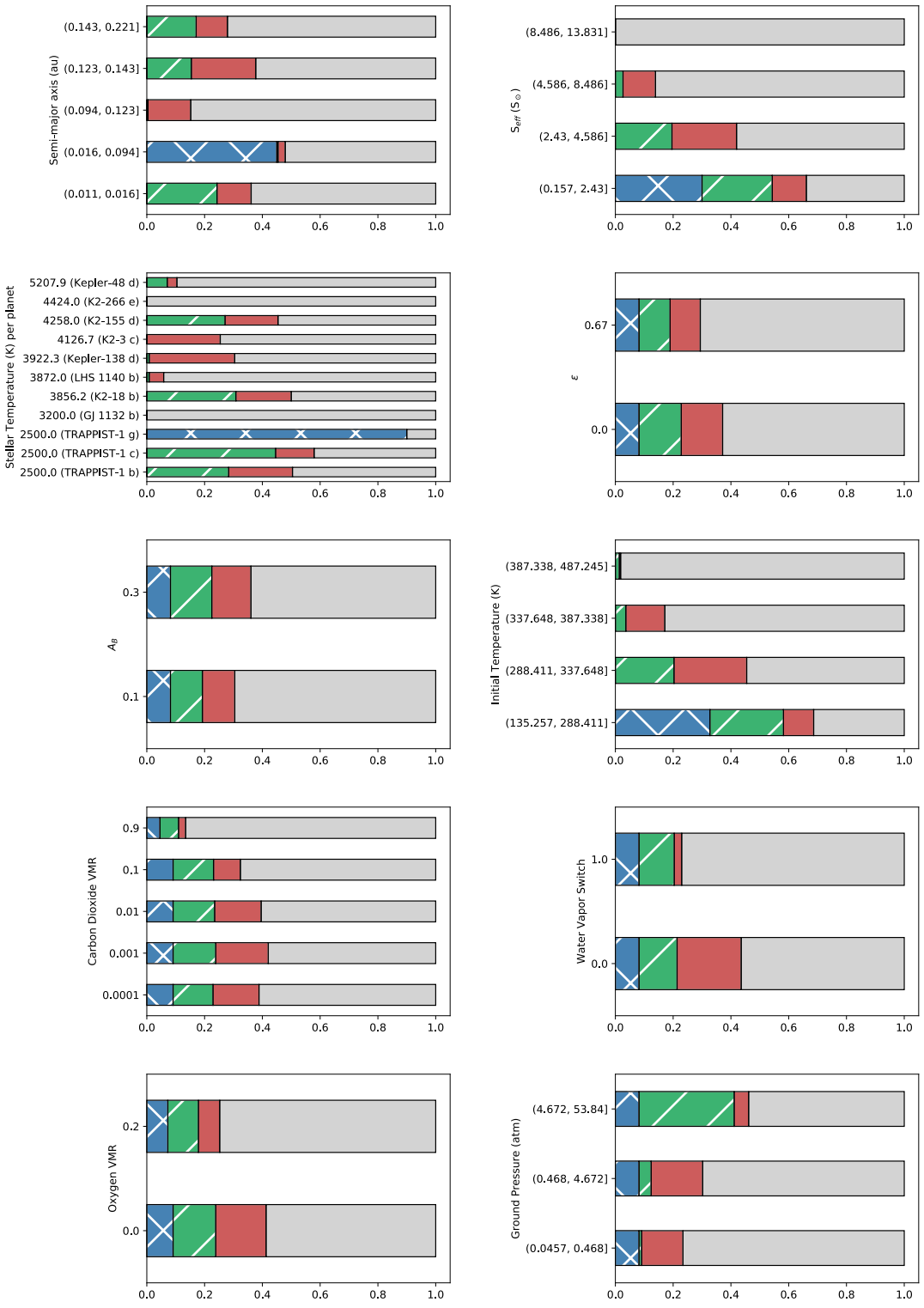


Fig. 7.39: Stacked histograms for the normalized frequency of habitable (green, diagonal hatching sample), cooler (blue, crossed diagonal hatching sample) or hotter (red, plain sample) models with respect to a few input parameters. The unsuccessful models are shown in grey.

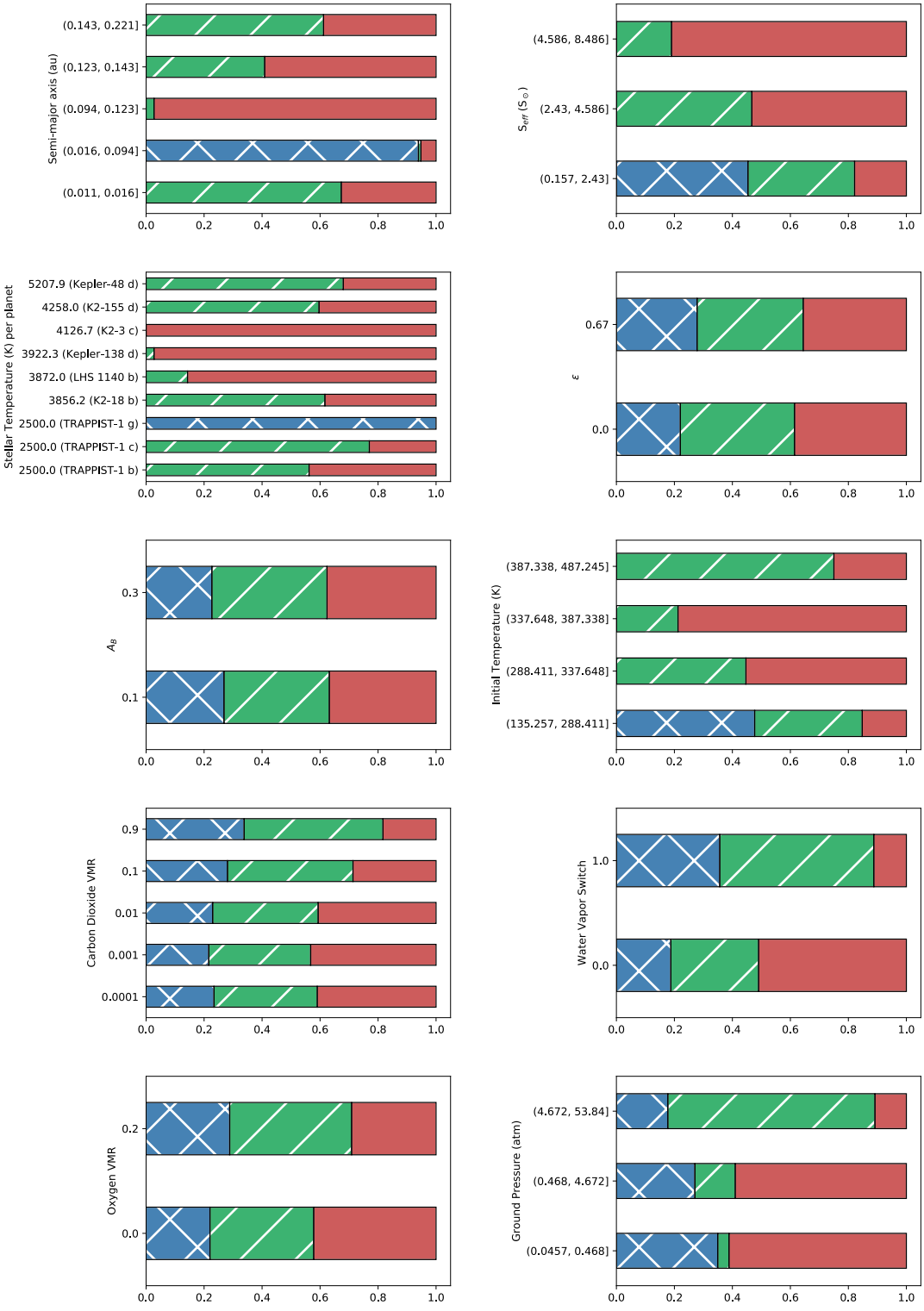


Fig. 7.40: Stacked histograms for the normalized frequency of habitable (green, diagonal hatching sample), cooler (blue, crossed diagonal hatching sample) or hotter (red, plain sample) models with respect to a few input parameters.

modeling of the atmosphere of a planet may have on habitability studies of such interesting target.

The stellar temperature does not seem to be as closely related to habitability as well: the largest difference is caused by the varying S_{eff}. Only the coolest star (TRAPPIST-1, at 2500 K) bears a fraction of cool, warm, and temperate atmospheres. TRAPPIST-1 g, despite being the most successful planet in the grid with no failed models, is too cool to be habitable for all the combinations of the input parameters. The other two planets in the system can have either habitable or warmer atmospheres.

Most of the modeled atmospheres appear to be too hot to be habitable (and in particular K2-3 c, which is not habitable for all combinations of input parameters). Therefore, even if the modeled atmospheres reached the radiative-convective equilibrium with `MAGRATHEA`, in the case of TRAPPIST-1 g and K2-3 c it is impossible to have liquid water on the surface.

Emissivity (right plot, second row) and albedo values (left plot, third row) do not appear to produce a massive change in the rates of habitability. As a consequence, the initial temperature (fourth row, right plot) is overall driven purely by the variation in stellar incoming flux. Initial temperatures cooler than 160 K are more likely to produce cool atmospheres, while an increase towards higher temperatures reduces the frequency of habitable planets, pushing the tropospheric temperature towards higher temperatures.

An increase in carbon dioxide abundance (left plot of the fourth row) does not show substantial changes in the rate of habitable planets. The lower number of successful models for the highest carbon dioxide abundance is linked to the higher rate of failed models: a few of them, we recall, were unstable from the very beginning because of the unphysical value of 90% carbon dioxide and 20% O_2 volume mixing ratios. Looking at the analog subplot in Figure 7.40, however, the fraction of successful habitable models is even bigger in this case with respect to the other abundances, normalized over a larger number of models.

The presence of water vapor (right plot of the fourth row) leads to a higher rate of unsuccessful models, as previously seen. However, the frequency of successful habitable models does not change in the dry or wet models. Dry models have a lower rate of failure, but in any case, a higher sample of warm atmospheres, underlying the effect that the other components of the mixture have on the temperature, enhanced by high irradiations.

The oxygen abundance subplot (left plot of the fifth row) suffers from the lack of successful models caused by the input instability, in combination with high values

of carbon dioxide. Among the successful models, the frequency of hotter planets is reduced if oxygen is present in the mixture, most likely because the models trespass the 500 K limit and fail. However, a "shielding" effect on the upper layers of the atmosphere could be produced when considering the photolysis of oxygen which forms ozone: a further study on the ozone production of these successful models will be described in Section 8.

The ground pressure (right plot, fifth row) increases the range of habitability, reducing both the number of cooler and hotter models. On the contrary, for lower values of pressure, the vaporization temperature decreases up to 325 K, thus reducing the range of water liquidity.

7.4 Take-home message

The two grids explore different regions of the parameters' space, and in particular different values of the effective stellar flux S_{eff}. Oxygen is also present in the grid of observed Super Earths, contrary to the previous one.

In general, the largest impact on the atmospheric profile is caused by the irradiance: the more photons an atmosphere receives from the star, the hotter it will be. Secondary changes in the shape of the profile are caused by the chemical composition of the atmosphere and are further enhanced by the stellar temperature (which changes the wavelength peak of the blackbody radiation). CO_2 and O_2 are effective in shielding the surface from the radiation, absorbing or scattering the majority of the flux in the upper layers, allowing the tropospheres to be more temperate. This may lead to ground temperatures well below the range of habitability, experiencing a scenario similar to the maximum greenhouse, even without including the formation of clouds in the atmosphere.

Water vapor is set to be more abundant, by construction, in the lower layers of the atmosphere: here, it can produce extreme variations in the shape of the profile compared to the modeled dry analog, and in particular, it increases the temperature at the ground boundary layer. This contributes to an increase of the frequency of unstable models (which would require more water vapor than it is physically possible) or surface temperatures hotter than 500 K, threshold over which `MAGRATHEA` loses its realiability. For habitability purposes, the positive feedback between water vapor and the temperature is a demonstration of the runaway greenhouse process.

If the emissivity of the atmosphere is non-null, the greenhouse effect of carbon dioxide and water vapor is even more effective in heating the whole atmosphere.

Any variation in pressure does not change dramatically the shape of the profile, nor its ground temperature. This value, on the other hand, plays a role in defining the habitability range of each modeled atmosphere. For lower pressures, it is more likely to have unstable atmospheres, since less water is required to effectively reach the maximum volume mixing ratio allowed by the other species.

As seen from the results of the modeled atmospheres on observed Super Earths, despite being the irradiances often much higher than what prescribed by the definition of the Habitable Zone by Kopparapu et al. (2013), with some combinations of atmospheric parameters the observed targets may retain a habitable atmosphere.

To be able to compare the results of these radiative-convective models with observed spectra of even warmer Super Earths, an extension towards hotter temperatures is required and, in fact, already scheduled for the next future.

It may also be necessary to loosen the requirement that imposes N_2 as a filling gas, adding the partial pressure of water vapor to the ones of the other dry components. This would, of course, increase the mass of the atmosphere in time, and its pressure subsequently: in this case, to avoid abrupt changes in pressure from one boundary layer to the subsequent caused by the increasing water vapor, a dynamic sampling of the boundary layers (at which the temperature must be evaluated) from one iteration to the following must be adopted.

For some cooler profiles, the condensation of both water vapor and carbon dioxide can no longer be neglected, so more accurate modeling of the outer boundary of the habitable zone should require the treatment of cloudy atmospheres. This feature is currently being implemented and validated, so the team will soon be able to repeat some interesting models.

As a consequence, an update of the albedo and the emissivity of cloudy layers may lead to more accurate results when calculating the incoming flux. The treatment of such parameters as wavelength-dependent would lead to much more reliable results.

`MAGRATHEA` should be also coupled with a routine that allows to vary the chemical composition both via equilibrium and non-equilibrium processes. This could allow a more consistent variation of the composition, as well as the production of more species that could be tracers for life. To do this, it is also necessary to include more species in the *k*-correlated table.

At this point, the pressure-temperature profiles could be fed to a retrieval software to generate theoretical emission spectra for many combinations of input parameters: this would allow to finally couple theory and data.

8 Exoplanet Ozone Model: ozone photochemistry

> *Our planet is a lonely speck in the great enveloping cosmic dark. In our obscurity – in all this vastness – there is no hint that help will come from elsewhere to save us from ourselves. It is up to us.*
>
> **— Carl Sagan (1934-1996)**
> (American astronomer, and science communicator)

This chapter deals with the third software developed during the doctoral fellowship, the `Exoplanet Ozone Model`. Its aim is to simulate qualitatively the ozone abundance in exoplanetary atmospheres. The code is presented and validated in the following sections. About 300 simulations were then executed and analyzed, as described in the remainder of the chapter. The input temperature profiles selected for these simulations belong to the second grid of models run by `MAGRATHEA` and analyzed in Chapter 7: these are therefore probable atmospheres of observed Super Earths belonging to the sample retrieved by `Exo-MerCat` (as seen in Chapter 5.4.7).

8.1 Overview

As seen in Sections 3.4.4.2 and 4.3, the presence of ozone is essential for habitability purposes, as well as for the chance of detecting life on an exoplanet.

A preliminary study on the production of ozone in exoplanetary atmospheres was made as a parallel branch of the modeling setup. The `Exoplanet Ozone Model`, or `EOM`, is the result of this effort. It is presently a standalone software, but coupling with `MAGRATHEA` is foreseen in the near future for more consistent results.

The main aim of the software is to study the abundance profile of ozone depending on the incoming star radiation and the atmospheric composition, given a fixed pressure-temperature profile. It considers a network of chemical reactions still fairly small, taking into account the Chapman reactions (reactions 3.94a to 3.94d), as well as some catalytic cycles (Cl and Br).

The software, written in Python 3, is extremely versatile, allowing the user to model many orbital situations, with non-null eccentricities, planetary axes inclinations, as well as changing the distance from the star and the stellar temperature. Its architecture, however, allows to easily implement new features for what concerns both physical and chemical processes.

At the time of writing, the atmospheric temperature profile is set by default to a user-provided value (which can be set in different ways) and cannot be changed during the run.

The portion of incoming radiation that is used in the `EOM` is composed by only a wavelength bin $[200, 400]$ *nm*: photons belonging to this (easily modified) range are the ones most efficient in causing the photolysis of molecular oxygen and ozone. The rest of the radiation, which indeed contributes to the heating of the atmosphere, is at the time of writing discarded.

These two approximations will, of course, be released in the future, when coupling the `EOM` with `MAGRATHEA` to coherently take into account for the heating produced both by the ozone photochemistry and by the other species in the mixture, which has a massive impact on the temperature profile.

The present study is, however, still extremely interesting to have some clues concerning the habitability of terrestrial exoplanets and to understand how effective is the absorption of UV photons provided by the ozone, which is essential for the survival of life.

This chapter is organized as follows: first, the `EOM` is described in Section 8.2; then, a validation study is performed in Section 8.3; finally, the ozone profiles for the subset of successful atmospheres of observed Super Earths containing oxygen (studied in Section 7.3) are retrieved and discussed in Section 8.4.

8.2 Exoplanet Ozone Model: description

The workflow of this software is shown in Figure 8.1. A general description of the code is provided in the following paragraphs.

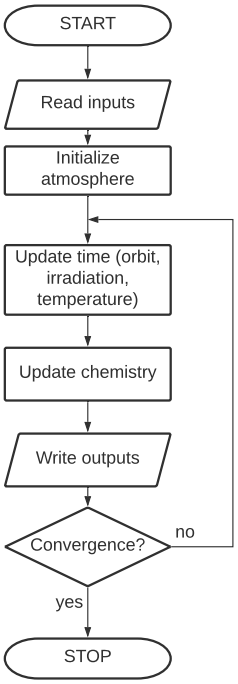


Fig. 8.1: Workflow of the `Exoplanet Ozone Model`.

Inputs. The user needs to provide an input file with all the desired settings for the model. The inputs are divided into categories (labeled by a comment line starting with a hash #) and can be added in the file regardless of the order. However, the information contained in the same category must be sorted to allow the code to properly read and assign the parameters. A sample of the input file is shown in Listing 8.1.

```
#ATMOSPHERE
1.0          (P/PEarth)
0.21         (O2 mixing ratio)
28.97        (Average molar mass of the atmosphere)

#PLANET
1            (AU)
1.0          (rPlanet/rEarth)
1.0          (mPlanet/mEarth)
0            (Eccentricity)
```

```
365          (nDaysInYear)
35           (SZA)

#STAR
5778         (Tstar)
1            (rStar/rSun)

#TEMPERATURE
equilibrium (input, isothermal, equilibrium,avgscaled)
0.3          (Albedo)
0.           (Emissivity)

#CHEMISTRY
backwardEuler   (simple or backwardEuler)

#TSTEP
180

#STEADYSTATE
T
0.005

#USEPHOTODATA
F
input/sorce_13.dat

#TSTART
2018     (year)
3        (month)
21       (day)
00       (hours)
00       (minutes)
00       (seconds)

#TEND
2019     (year)
01       (month)
13       (day)
00       (hours)
00       (minutes)
00       (seconds)

#OUTPUT
43200 (seconds)
1
ALL
```

Listing 8.1: Sample of an input file read by the `Exoplanet Ozone Model`: Earth case.

The categories are:

- **ATMOSPHERE**: Here, some information concerning the atmosphere must be set. The first float number must be the ground pressure of the atmosphere in *atm* units. Then, since the code deals only with the chemistry of oxygen-related compounds, the only volume mixing ratio to be provided is the one of the molecular oxygen (from 0 to 1); the rest of the atmosphere would be then composed by any other species that is inert in this simplified framework. The chemical composition needs, however, to be taken into account when specifying the average molar mass of the atmosphere μ in the following row.

- **PLANET**: The fundamental parameters to calculate the incoming radiation on the top of the atmosphere. They must be, in order: distance in astronomical units; planetary radius in Earth radii; planetary mass in Earth masses; eccentricity; the period in days; solar zenith angle in degrees.

- **STAR**: Here, the stellar temperature in K and the radius of the star in Solar radii must be specified, to allow the calculation of the blackbody stellar radiation.

- **TEMPERATURE**: The temperature can be calculated in different ways, each one requiring its own set of ordered parameters. For the `isothermal` option, the temperature values are replaced by the same, user-provided value. For the `equilibrium` option, the atmospheric temperature is set to a uniform value, equal to the equilibrium temperature of a leaky greenhouse model (Eq. 6.14); it requires one to specify the values of the albedo and the emissivity of the atmosphere.

  For the `avgscaled` option, the temperature is loaded from a file containing the U.S. Standard Atmosphere 1976 (hereafter USSA); the code compares the value of the average temperature of the USSA (237.4 K) with the equilibrium temperature of the leaky greenhouse model, retrieving the difference between the two estimates, which is then subtracted from every temperature point in the USSA profile. This is useful to shift the average temperature according to the distance of the planet while keeping a shape of the profile similar to the terrestrial one.

  For the `input` option, the profile is read from a custom atmospheric file, whose path needs to be specified as a parameter.

- **CHEMISTRY**: The chemical solver can be selected: if the option is `simple`, the photochemical equilibrium for certain species is assumed; if the option

is `backwardEuler`, the backward implicit Euler solver is used, with no photochemical equilibrium assumed by default.

- **USEPHOTODATA**: If the logical `T` is inserted, the code uses user-provided observed stellar irradiances at the top of the atmosphere, read at the path provided as an input at the following row; if `F`, the code calculates the blackbody radiation, ignoring the irradiance file.

- **STEADYSTATE**: If `T` is inserted, the model runs until a steady-state has been reached, ignoring any information concerning the ending time (see below); the stop criteria (in percent) needs to be provided in this case. Otherwise, the value `F` means that the code will stop when the time reaches the ending time specified in the TEND category.

- **TSTEP, TSTART, and TEND**: These are respectively the time step at which the chemical abundances needs to be evaluated during the run, the initial, and the final time of the simulation. TSTEP requires the time step in seconds as a parameter; for TSTART and TEND the parameters must be inserted in this order: year, month, day, hours, minutes, seconds. The information about TEND is ignored if the model runs in steady-state mode.

- **OUTPUT**: It concerns all information on the outputs provided by the code. The first integer is the number of seconds between writing output; the second integer is the number of outputs to write. After that, the name of all the required outputs can be listed. Those can be: `ALL` to write the altitude values (in km) and the concentrations of O, O_2, and O_3 at each altitude point (in $molecule\ cm^{-3}$); `PHOTO` to write the photodissociation rates J_{O_2} and J_{O_3} for every altitude point; `USER` to plot any user-provided variable, to be specified within the `outputuser` Python function in the `user` module.

The `EOM` reads the input file and stores all useful variables. In doing this, checks concerning the presence and the validity range of any parameter in the file are performed: if something is not valid or absent, an error is raised and the execution is halted.

Initializing the atmosphere and other useful values From the input values the atmospheric pressure profile is derived, as well as the corresponding numerical density profile in perfect gas approximation and hydrostatic equilibrium:

$$\xi_i = \xi_{NL} e^{-z_i/H} \tag{8.1}$$

with ξ being either pressure or numerical density, z_i being the altitude, and H the scale height of the atmosphere. Depending on the user-provided volume mixing ratio, the molecular oxygen profile is also retrieved. The other components are set to Earth-like values: in particular, the abundances of chlorine are set to volume mixing ratios of $3 \cdot 10^{-9}$ and $2 \cdot 10^{-11}$ respectively. The code allows customizing easily the chemical network, by setting either a constant or an altitude-dependent abundance profile.

If needed, the irradiance is read from the file whose path was provided in input: the wavelength-dependent values are integrated for the considered wavelength bins.

The O_2 and O_3 cross section values read from the files retrieved from [Huebner et al., 2015] and stored in the `input` folder. They are then averaged in the wavelength bin that the `EOM` considers (200-400 nm).

Time step update. The software starts to iterate over time from the initial time specified by TSTART and adding TSTEP at each iteration. The code calculates the orbital distance of the planet from the star (which may vary if the eccentricity is non-null). It then updates the irradiance value at the top of the atmosphere according to the updated distance, either scaling the stellar blackbody radiation or interpolating between two observed values of irradiation at two orbital distances. The temperature profile is also updated, particularly in the `equilibrium` case and in the case of non-null eccentricity, for which the value is updated according to the varying semi-major axis variation.

Chemical composition update. The optical depth τ is calculated depending on the abundance of the major absorbers of the theoretical atmosphere in the considered wavelength range (O_2 and O_3). The values of τ are calculated for each atmospheric layer i (for $i = 1...NL$ where NL is the ground level) as follows:

$$\tau(i) = \Delta z \sum_{X=O_2,O_3} [\mathrm{X}](i-1) \cdot \sigma_X \tag{8.2}$$

Where Δz is the thickness of the layer i, σ_X is the average cross-section of each species in the wavelength bin and $[X](i-1)$ is the abundance of the species X in the upper layer $i-1$.

From the Beer-Lambert law, it is possible to calculate the attenuated flux in the wavelength bin for each layer i by multiplying the incoming radiation from the upper

layer $i-1$ by the absorption coefficient due to the absorption within the layer of thickness Δz, over a slanted path defined by the solar zenith angle α:

$$I(i) = I(i-1) \cdot \exp\left[-\frac{\tau(i-1)}{\cos\alpha}\right] \tag{8.3}$$

This is equal to:

$$I(i) = I_0 \cdot \exp\left[-\frac{1}{\cos\alpha}\sum_{j=1}^{i-1}\tau_j\right] \tag{8.4}$$

Where I_0 is the value of incoming radiation at that orbital distance previously calculated.

As a consequence, $i = 1$ is the level at the TOA, where $\tau = 0$ and $I = I_0$. In the previous equations, we discarded the contribution of the abundance of atomic oxygen in the absorption coefficient and we approximated the abundances of O_2 and O_3 in the layer i, still unknown, to the ones of the layer $i-1$.

For each layer i the photodissociation rates of molecular oxygen and ozone are thus retrieved:

$$J_X(i) = \sigma_X I(i) \tag{8.5}$$

Now, the abundance of ozone in each layer needs to be updated.

At the time of writing, the following reactions are included in the network, in addition to the four ones that characterize the Chapman mechanism (reactions from 3.94a to 3.94d):

$$Cl + O_3 \xrightarrow{k_3} ClO + O_2 \tag{8.6}$$

$$Br + O_3 \xrightarrow{k_4} BrO + O_2 \tag{8.7}$$

These are part of the catalytic cycles of chlorine and bromine, which are an ozone sink.

If the chemical solver is set to `simple`, the abundance of O_2 is held fixed throughout the run and O is set to be at photochemical equilibrium with the other species.

Now, the reaction rates for all the other reactions involved are updated according to the temperature of each layer i. The values of k_1 and k_2 are calculated using Eqs. 3.96 and 3.97; the values of k_3 and k_4 are thus retrieved:

$$k_{3,i}\ [cm^3\ s^{-1}] = 2.8 \cdot 10^{-11} \exp\left(\frac{-2100}{T_i}\right) \tag{8.8}$$

$$k_{4,i}\ [cm^3\ s^{-1}] = 1.7 \cdot 10^{-11} \exp\left(\frac{-6600}{T_i}\right) \tag{8.9}$$

The updated equation for the variation of ozone in the iteration for each layer i is:

$$[O_3]'_i := \frac{d[O_3]_i}{dt} = +k_{1,i}[O]_i[O_2]_i[M]_i - k_{2,i}[O]_i[O_3]_i - \\ -k_{3,i}[Cl][O_3]_i - k_{4,i}[Br][O_3]_i - J_{O_3,i}[O_3]_i \tag{8.10}$$

And therefore, the new ozone abundance at the time step $t + \Delta t$ for each layer i is:

$$[O_3]_i(t + \Delta t) = [O_3]_i(t) + \Delta t \frac{d[O_3]_i}{dt} \tag{8.11}$$

If the chemical solver is set to `backwardEuler`, the full time-dependent sources and losses of all reactions for the three oxygen-related species are calculated explicitly, obtaining the system of three ordinary differential equations already seen in Eqs. 3.95a, 3.95b, and 3.95c, with the addition of two terms that account for the catalytic cycles. The partial derivatives of each equation with respect to the three oxygen-related species can be now calculated. For example, the partial derivatives of Eq. 8.10 are:

$$\frac{\partial[O_3]'_i}{\partial[O]_i} = +k_{1,i}[O_2]_i[M]_i - k_{2,i}[O_3]_i \tag{8.12}$$

$$\frac{\partial[O_3]'_i}{\partial[O_2]_i} = +k_{1,i}[O]_i[M]_i \tag{8.13}$$

$$\frac{\partial[O_3]'_i}{\partial[O_3]_i} = -k_{2,i}[O]_i - k_{3,i}[Cl] - k_{4,i}[Br] - J_{O_3,i} \tag{8.14}$$

A Jacobian 3x3 matrix can be defined, where each element is the partial derivative of each equation for each species.

$$\mathcal{J}_i = \begin{bmatrix} -A - B & 2J_{O_2,i} - C & -D + J_{O_3,i} \\ -A + 2B & -J_{O_2,i} - C & 2D + J_{O_3,i} + E + F \\ +A - B & +C & -J_{O_3,i} - D - E - F \end{bmatrix} \tag{8.15}$$

Where the values of A, B, C, D, E, and F are:

$$A = k_{1,i}[O_2]_i[M]_i \tag{8.16}$$
$$B = k_{2,i}[O_3]_i \tag{8.17}$$
$$C = k_{1,i}[O]_i[M]_i \tag{8.18}$$
$$D = k_{2,i}[O]_i \tag{8.19}$$
$$E = k_{3,i}[Cl]_i \tag{8.20}$$
$$F = k_{4,i}[Br]_i \tag{8.21}$$

This Jacobian is then used to retrieve the abundance of the three species according to the inverse Euler method for every layer in the atmosphere, repeating the steps if necessary until convergence.

Output writing and Convergence criteria. The code writes the desired output periodically and stores in the `data` folder. If the steady-state option is set to False, the code runs until the end time is reached.

On the other hand, if the models must reach the steady-state, the software calculates the maximum difference between the ozone profile corresponding to a given iteration to the one produced by the following one. If the value is less than the user-provided tolerance, the model has converged. The `EOM` stores the previous difference values and it is able to warn the user if the solution is diverging (i.e. if the difference is increasing in time) and stops the model from running further. This may be the case, especially when considering the simple chemical solver with a large time step.

Once either convergence or final time is reached, the code writes the last set of output adding the label `final` to the string of the file name.

8.3 Validation

The EOM project is still at its first steps, so many approximations are present. We cannot, therefore, expect quantitatively accurate results for the ozone profile. Besides, the abundance of O_3 and its distribution on the atmosphere vary massively depending on seasonality, latitude, stellar environment, as well as anthropogenic effects.

Nevertheless, the output profile produced by the validation run is coherent with the average abundance of O_3 in Earth's atmospheres (a volume mixing ratio of a few parts per million, or a concentration of the order of 10^{12} $molecules\ cm^{-3}$). As visible from Figure 8.2, the peak of the maximum concentration is shifted towards slightly higher altitudes with respect to the altitude of the average terrestrial ozone layer (about 25-30 km). This effect can be due to multiple reasons. First of all, the Chapman cycle is known to produce slightly higher peaks in the stratosphere, and the presence of catalytic cycles have too weak an impact for it to vary: this could be changed in the future when adding more reactions to the chemical network. Furthermore, considering only a reduced wavelength bin in radiation underestimates the I_0 value, which has an impact on the determination of the J coefficients that control the altitude of the peak, as will appear in the remainder of the section.

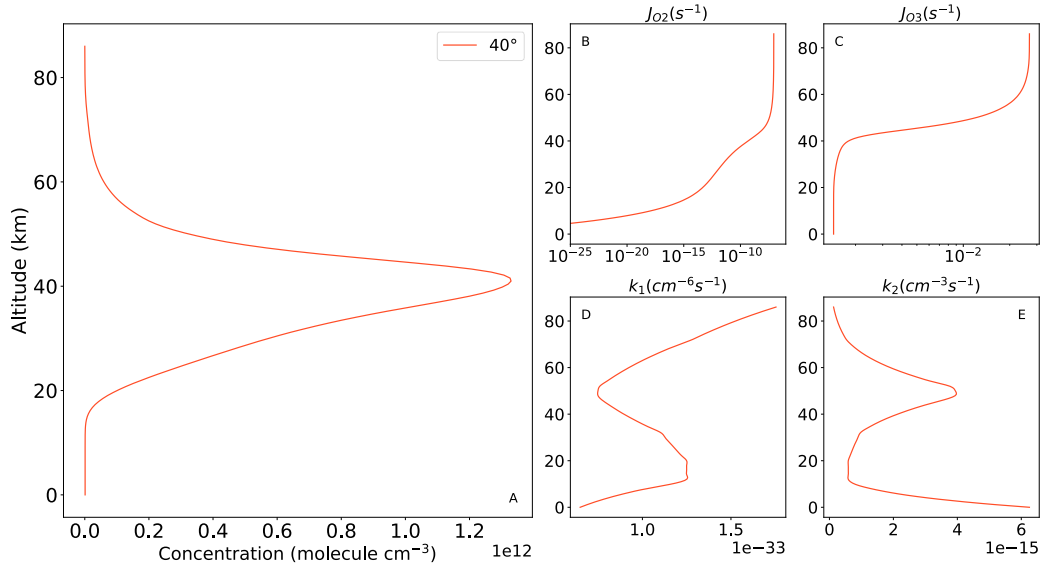


Fig. 8.2: *Subplot A*: Ozone abundance in a terrestrial atmosphere at a solar zenith angle of 20 degrees using the U.S. Standard Atmosphere; *Subplot B*: Photodissociation rate of molecular oxygen for the modeled atmosphere; *Subplot C*: Photodissociation rate of ozone for the modeled atmosphere; *Subplot D*: Thermal reaction rate of reaction 3.94b for the modeled atmosphere; *Subplot E*: Thermal reaction rate of reaction 3.94d for the modeled atmosphere.

The model is however qualitatively accurate, within the current assumptions considered by the `EOM`. The other features were separately tested in a few other models. Some useful graphs are shown in this section as well, since the study of the effects of the separate parameters may be instructive for both the study of the potential of the code and the identification of all the improvements to be made in the future.

Chemical solver variation. Varying the chemical solver does produce some variation on the output, as shown in Figure 8.3. The values of J_{O_2}, J_{O_3}, k_1, and k_2 are the same for both models, but the profiles show some difference around the peak. The Backward Euler chemical solver is indeed more accurate in taking into account the variation of all the oxygen-related species, so it is preferable to select this routine when running a model. However, the Backward Euler model reaches the steady-state in 19440 iterations, compared to the 3525 iterations of the `simple` chemical solver, using the same time step of 10 seconds and the same steady-state convergence criterion of 0.005%. In terms of computing time, this translates into a run of a dozen minutes compared to 50 seconds.

So, the choice of the solver depends on the purpose of the run: a small, maybe negligible, difference in the shape of the profile may pay out in terms of computational speed – extremely valuable when running a wide batch of models in a short time.

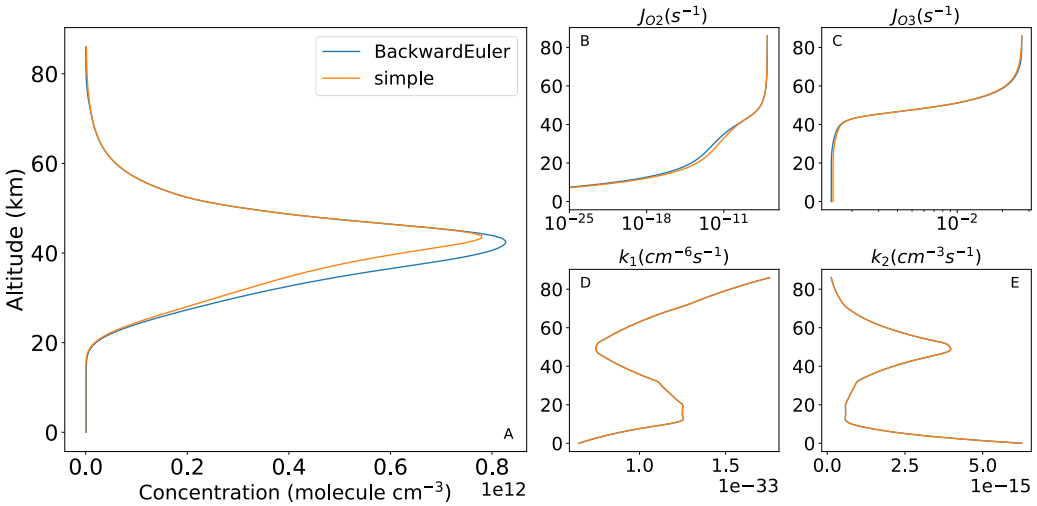


Fig. 8.3: *Subplot A*: Ozone abundance in a terrestrial atmosphere at a solar zenith angle of 60 degrees using the U.S. Standard Atmosphere and the `Backward Euler` (blue line) and `simple` (orange line) solvers; *Subplot B*: Photodissociation rates of molecular oxygen for the modeled atmospheres; *Subplot C*: Photodissociation rates of ozone for the modeled atmospheres; *Subplot D*: Thermal reaction rates of reaction 3.94b for the modeled atmospheres; *Subplot E*: Thermal reaction rates of reaction 3.94d for the modeled atmospheres.

Solar zenith angle variation. We then modeled the Earth's atmosphere, sampling it at various solar zenith angles, considering a backward Euler chemical solver with a time step of 10 seconds and a steady-state convergence criterion of 0.005%. The results are shown in Figure 8.4.

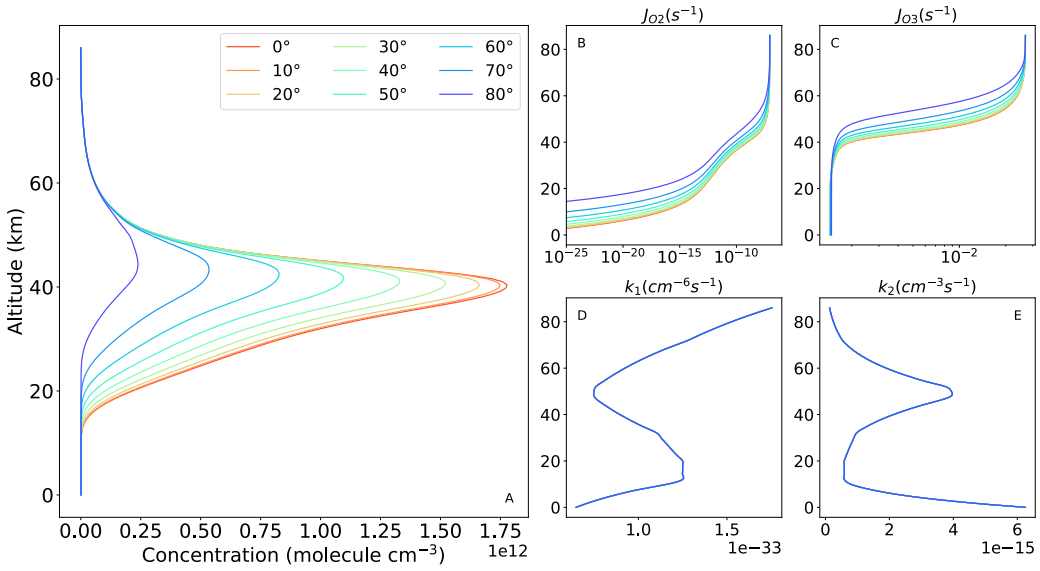


Fig. 8.4: *Subplot A*: Ozone abundance in a terrestrial atmosphere at varying solar zenith angles; *Subplot B*: Photodissociation rate of molecular oxygen in a terrestrial atmosphere at varying altitudes (reaction 3.94a); *Subplot C*: Photodissociation rate of ozone in a terrestrial atmosphere at varying altitudes (reaction 3.94c); *Subplot D*: Thermal reaction rate of reaction 3.94b; *Subplot E*: Thermal reaction rate of reaction 3.94d.

Since the planet is not tilted at the time of writing, one can assume the solar zenith angle as an estimate of the latitude at which the atmospheric column is sampled.

The models show a decrease in the abundance of ozone for increasing latitudes; the abundance near the poles is massively reduced, up to two orders of magnitude: due to the larger solar zenith angle, a very small amount of radiation reaches the poles and can produce ozone in the first place. This variation has an impact on the photodissociation factors, according to equation 8.4, by varying the exponent coefficient which depends on the solar zenith angle α. The main difference is in the height at which the value of J_{O_3} is sensibly reduced, marking the end of the regime in which the ozone photolysis is efficient. Ozone can then be built more easily and forms the ozone layer, whose peak concentration altitude is very similar to the one corresponding to the steep reduction of the photodissociation rate, which happens at higher altitudes depending on the solar zenith angle. Physically, this is due to the higher atmospheric mass crossed by the photons, which reduces the overall available

flux. The height of the peak, by consequence, increases with increasing values of α, while the maximum abundance is inversely related to that quantity.

The reaction rates k_1 and k_2, on the other hand, have no reason to be different since they only depend on the temperature, whose profile is the same for every model in this set.

Such models cannot take into account the complex dynamical motion that drives ozone from the equator, where it is mostly produced, to the poles, as indeed is the case for the Earth. Therefore, the `EOM` cannot accurately model the higher abundance of ozone towards the poles, but simply because that is not the goal of the software itself: the Exoplanet Ozone Model aims at the study and the understanding of the production of the ozone, with little concern about its motion throughout the various latitudes. In any case, coupling with a 3D GCM model may produce further results on this fascinating topic.

Temperature profile variation. The choice of the input temperature in the `TEMPERATURE` category causes the ozone profile to vary as well. As visible in Figure 8.5, both the maximum concentration and the height at which it occurs changes when considering the same atmosphere (Earth's atmosphere around the Sun) using the U.S. Standard Temperature (with `TEMPERATURE` mode equal to `input`, orange line) and using an isothermal temperature with all layers set at the equilibrium temperature of a leaky greenhouse model (`TEMPERATURE` mode equal to `equilibrium`, blue line), which is very similar to the average temperature of the U.S. Standard Temperature profile.

The differences are produced by the k rates, which are temperature-dependent and therefore are fairly different from one case to the other. The small differences in the value of J_{O_3} are a reflection of the variation in concentration produced by the thermal reactions, which vary the optical depth of each layer.

In any case, both profiles give comparable results, with negligible differences when considering the order of magnitude. For this reason, in the case where no specific pressure-temperature profile is available, the use of isothermal atmospheres may be still useful. Of course, a more coherent analysis could be made when introducing a customized profile: as will be discussed in Section 8.4, `MAGRATHEA` could be ideal to provide such inputs.

Distance variation. To test the performance of the code at varying distances, three further atmospheres were calculated, supposing an Earth-like atmosphere on an Earth-like planet around the Sun at Martian, terrestrial, and Venusian distances.

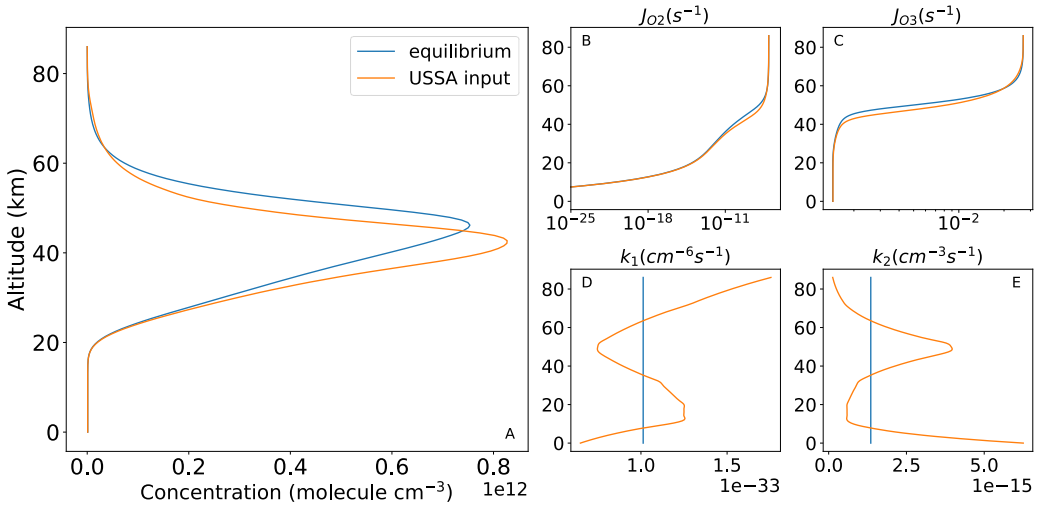


Fig. 8.5: *Subplot A*: Ozone abundance in a terrestrial atmosphere at a solar zenith angle of 60 degrees using the U.S. Standard Atmosphere (orange line) and an isothermal atmosphere at the equilibrium temperature of a leaky greenhouse model (blue line); *Subplot B*: Photodissociation rates of molecular oxygen for the modeled atmospheres; *Subplot C*: Photodissociation rates of ozone for the modeled atmospheres; *Subplot D*: Thermal reaction rates of reaction 3.94b for the modeled atmospheres; *Subplot E*: Thermal reaction rates of reaction 3.94d for the modeled atmospheres.

At first, these three models were calculated considering the equilibrium temperature of the leaky greenhouse model at $A_B = 0.3$ and $\epsilon = 0$ scaled depending on the values of the average semi-major axes of Venus ($\approx 0.7\ AU,\ T_a = 279.7\ K$), Earth ($1\ AU,\ T_a = 237.3\ K$), and Mars ($\approx 1.52\ AU,\ T_a = 192.4\ K$). The results are shown in Figure 8.6. The same models were then repeated with the `avgscaled` temperature option, considering the translated U.S. Standard Temperature at Martian and Venusian distances (whose average temperatures are still consistent with the previous case). These results are shown in Figure 8.7.

The most significant difference between the models is given by the change in the width of the ozone layer. The maximum abundance is comparable, but the altitude at which the maximum concentration appears is shifted more and more upwards the farthest the planet is from the star. This effect is linked to the variation in the I_0 component of the irradiation. The opacity of each layer (which contributes to the exponent of Eq. 8.4) is invariant depending on distance, so the value of I_0 is the one that determines the number of available photons for the photochemical reactions. Therefore, the ozone photolysis loses efficiency higher in the atmosphere for the atmosphere orbiting a planet at 1.5 AU compared to closer planets.

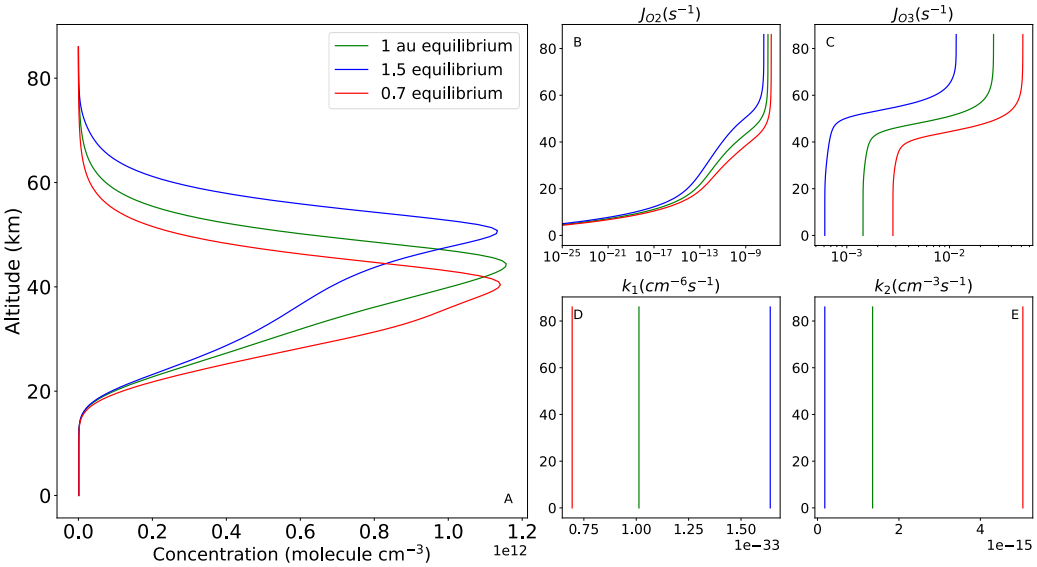


Fig. 8.6: *Subplot A*: Ozone abundance in a terrestrial atmosphere at varying distances, using an isothermal temperature. In green, the profile corresponding to the modeled planet is at 1 AU (Earth-Sun, distance) from the star; in blue, the profile corresponding to the modeled planet is at 1.5 AU (Mars-Sun, distance) from the star; in red, the profile corresponding to the modeled planet is at 0.7 AU (Venus-Sun, distance) from the star. *Subplot B*: Photodissociation rates of molecular oxygen for the modeled atmospheres; *Subplot C*: Photodissociation rates of ozone for the modeled atmospheres; *Subplot D*: Thermal reaction rates of reaction 3.94b for the modeled atmospheres; *Subplot E*: Thermal reaction rates of reaction 3.94d for the modeled atmospheres.

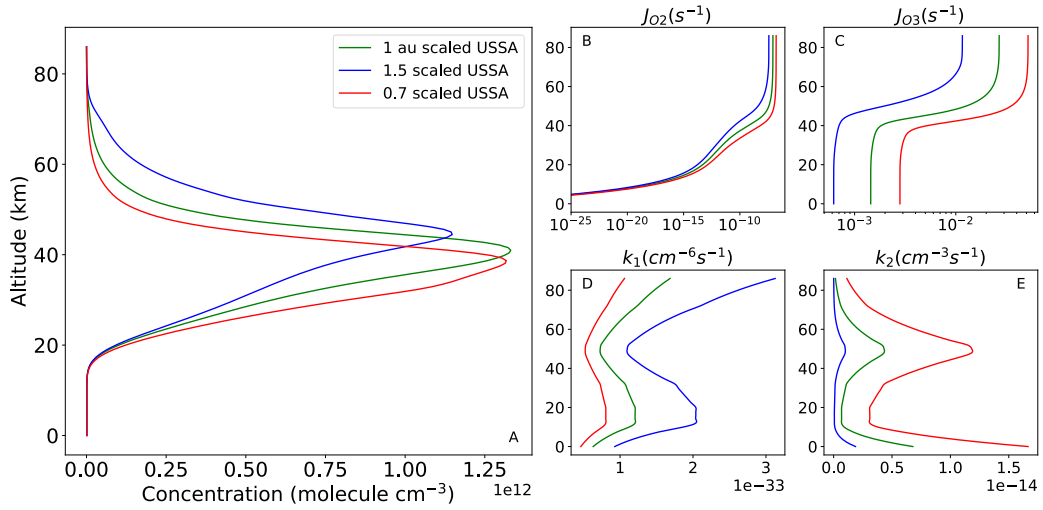


Fig. 8.7: *Subplot A:* Ozone abundance in a terrestrial atmosphere at varying distances, using a scaled U.S. Standard Profile. In green, the profile corresponding to the modeled planet is at 1 AU (Earth-Sun distance) from the star; in blue, the profile corresponding to the modeled planet is at 1.5 AU (Mars-Sun distance) from the star; in red, the profile corresponding to the modeled planet is at 0.7 AU (Venus-Sun distance) from the star. *Subplot B*: Photodissociation rates of molecular oxygen for the modeled atmospheres; *Subplot C*: Photodissociation rates of ozone for the modeled atmospheres; *Subplot D*: Thermal reaction rates of reaction 3.94b for the modeled atmospheres; *Subplot E*: Thermal reaction rates of reaction 3.94d for the modeled atmospheres.

If a non-uniform temperature profile is introduced, the values of k_1 and k_2 start to vary as well, causing a variation in the concentration of ozone, which appears to be lower for the less irradiated planet.

Temperature variation From what previously seen, the variation in the distance leads to changes in the thickness of the profile and the abundance of ozone. The stellar temperature is expected to play a role as well, determining the amount of available radiation in the considered wavelength range. Selecting Proxima Centauri (3042 K, 0.154 $R_\oplus$) and imagining an Earth orbiting at 0.041 AU, so that the effective stellar flux would be comparable to the terrestrial one ($\approx 1\ S_\oplus$), the ozone abundance decreases with respect to the solar case, as shown in Figure 8.8, but both the thickness of the layer and the altitude of the maximum concentration are similar.

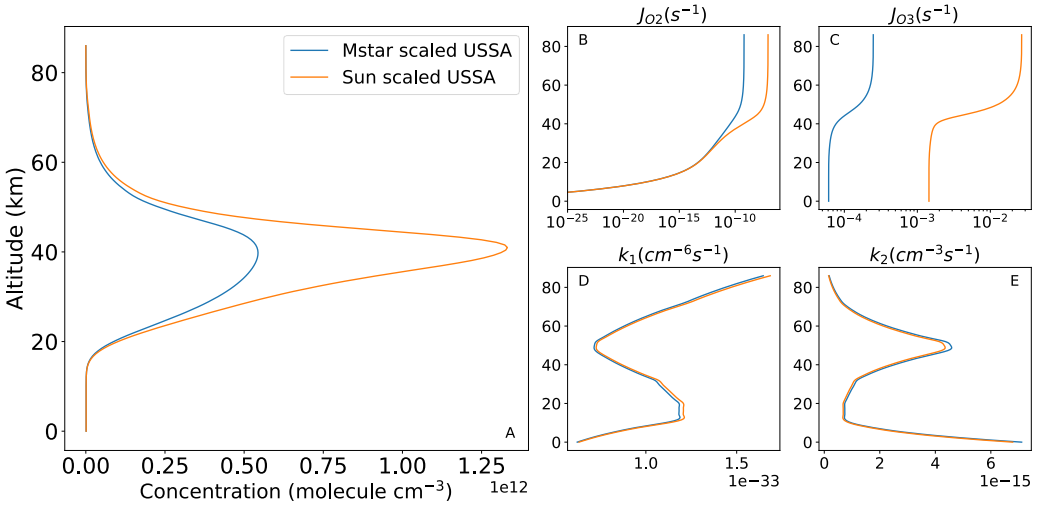


Fig. 8.8: *Subplot A*: Ozone abundance in a terrestrial atmosphere at varying stellar temperatures (but similar irradiation). In blue, the profile corresponding to a planet orbiting an M star (3042 K) at 0.0041 AU; in orange, the profile corresponding to the Earth; *Subplot B*: Photodissociation rates of molecular oxygen for the modeled atmospheres; *Subplot C*: Photodissociation rates of ozone for the modeled atmospheres; *Subplot D*: Thermal reaction rates of reaction 3.94b for the modeled atmospheres; *Subplot E*: Thermal reaction rates of reaction 3.94d for the modeled atmospheres.

The variation is caused by the smaller value of I_0, caused by the depletion of UV photons for the M star scenario (whose blackbody radiation is shifted towards the infrared). This reduces the efficiency of both photochemical production and destruction of ozone, which is, therefore, lower than the solar case.

At this point, a clarification is due: the currently available results consider the stellar irradiation in the UV range of interest (200-400 nm) in its blackbody approximation. In this framework, cooler stars provide fewer UV photons compared to hotter objects.

This is however not the case for some M stars in particular, which can be active and therefore particularly strong ultraviolet emitters. Such scenarios are extremely interesting and will be studied in the future.

8.4 Ozone production on Super Earths

From the subsample of successful atmospheres on observed Super Earths described in Section 7.3, the ones with a non-null oxygen abundance were selected, for a total of 333 models. For each atmospheric profile, an `EOM` run was made to calculate the ozone abundance that would correspond to those conditions of chemical composition, pressure, and temperature. The input files were built based on the `MAGRATHEA` input file, to take into account the very same parameters. Details on the chemical composition were not compulsory since the `EOM` is interested only in the abundance of molecular oxygen (the rest of the atmosphere being composed by any kind of molecule that contributes to acting as an unspecified third body).

The atmospheric profile, as retrieved by `MAGRATHEA`, was given in input to the `EOM`, by setting the mode `input` in the TEMPERATURE category. The chemical solver was set to `backwardEuler`, with a time step of 180 seconds.

All models but 55 reached convergence: the failed ones corresponded to some models at the lowest pressure, for which the same number density of ozone would correspond to higher values of volume mixing ratio, and fairly high irradiations: it would be easier for these models to exceed the mass conservation limit even for low ozone abundances.

The analysis of the results suffers from the non-uniform distribution of the physical parameters and on the very low number of models to be considered: the results are expected to be biased by the large variability of such initial conditions. As a preliminary study on the topic, however, this analysis may still interesting to test the performance of the software and to set the ground for future runs.

Physical parameters such as irradiation and stellar temperature have already proven to be important for the production of ozone in a few ad-hoc cases. This scenario will likely occur again when analyzing this set of models.

The chemical composition of the atmosphere, on the other hand, is not directly included in the EOM but is indeed taken into account by the input temperature profiles, which were obtained through a MAGRATHEA run. The differences in temperature and shape of the profile are likely to affect the abundance of ozone, most certainly by varying the thermal reaction rates k_1 and k_2.

The variation in pressure is very likely to play a role as well, since it determines the optical depth τ which is involved in the retrieval of the photodissociation rates J_{O_2} and J_{O_3}.

For a qualitative approach to the problem, we considered only three particular features of each modeled ozone profile: the maximum abundance of this species, the altitude at which the maximum abundance occurs in the atmosphere (corresponding to the average altitude of the ozone layer), and the thickness of the ozone layer (calculated as the difference between the two altitudes at which the ozone abundance starts being greater than the 5% of the maximum value). Similarly, the information on the input profile was reduced to a single parameter, the average temperature of the atmosphere.

The contribution of the average temperature, the irradiation, the stellar temperature, and the pressure, is shown in Figures from 8.9 to 8.14. Figures from 8.9 to 8.11 show the variation of the maximum O_3 concentration of ozone (Figure 8.9), the altitude of the maximum ozone concentration (Figure 8.10), and the thickness of the layer (Figure 8.11) with respect to the average temperature in a colormap determined by the irradiation S_{eff}. Figures from 8.12 to 8.14 represent the very same plots, but in a colormap determined by the stellar effective temperature T_*.

A change in atmospheric temperature towards higher values is related to an increase in ozone abundance up to an order of magnitude. In particular, given the solar zenith angle of 60 degrees fixed for all models, which corresponds to about 10^{12} $molecules\ cm^{-3}$ at an average temperature of $\sim$237 K in the terrestrial case (see Figure 8.4), it is possible to notice that slightly higher concentrations of O_3 are foreseen for planets orbiting some stars cooler than the Sun, provided that they receive high enough irradiation.

The different behavior is caused by the variation of all reaction rates involved in the ozone chemistry, the thermal ones being affected by the temperature of each layer, and the photodissociation ones related to the change in stellar irradiation – which, as previously described, is the parameter that determines the overall heating of the atmosphere, thus its temperature.

It is indeed the stellar effective flux the parameter that produces the largest variations. The colormap in Figure 8.9 represents graphically this trend: the most irradiated planets can produce much more ozone than their counterpart at high average temperatures but lower irradiations. On the other hand, for lower irradiations (e.g. the TRAPPIST-1 system, which have the lowest S_{eff} in the sample and whose models are shown in dark blue, the lower part of the figure), an increase in temperature does produce a general increase in concentration, but the magnitude of the effect is much smaller.

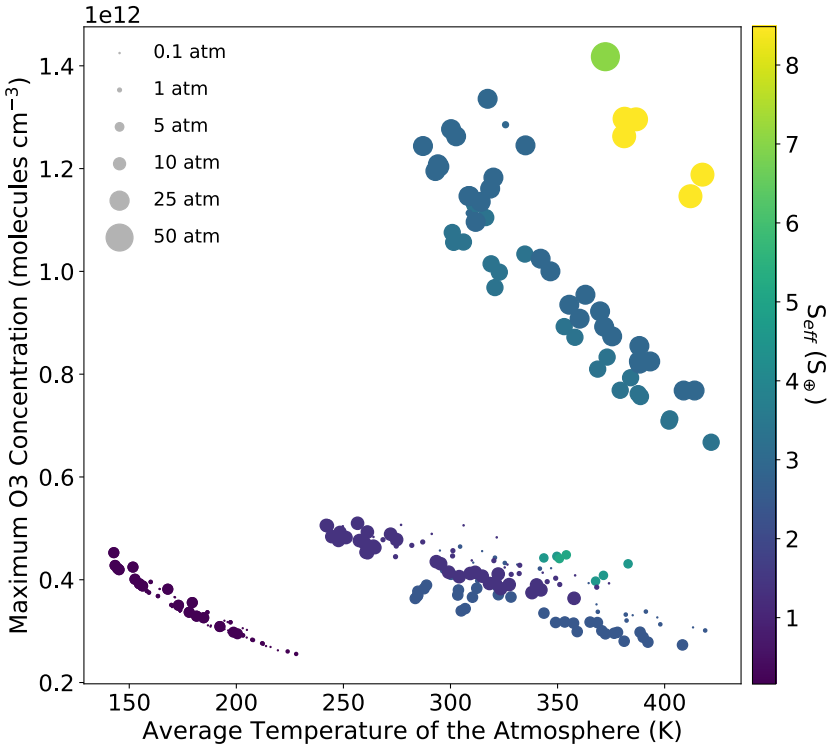


Fig. 8.9: Maximum ozone concentration in *molecules* cm^{-3} vs average atmospheric temperature in K. The colormap shows varying stellar irradiations; the size of the dots show varying ground pressures in atm.

This trend is noticeable in Figure 8.10 as well: higher values of average temperatures are related to wider ranges of possible altitudes of the maximum ozone concentration. The spread of the altitude range appears to be even enhanced by pressure effects (graphically represented by the size of the dots).

As shown in Section 7.2.2.1, varying the pressure of the atmosphere translates into a variation in the density of each layer, but not a large change in the average temperature of the profiles. The same conditions of average temperature and incoming radiation with a change in pressure lead to different values of the photodissociation rates, due to the change in optical depth τ and, thus, of the exponent of Eq. 8.4. On

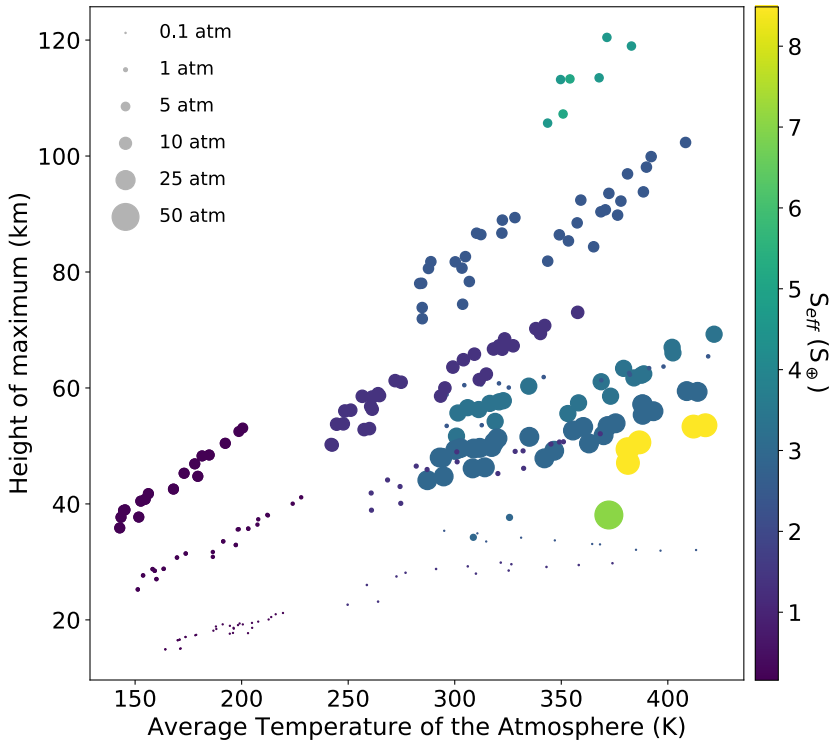


Fig. 8.10: Altitude in km corresponding to the maximum ozone concentration vs average atmospheric temperature in K. The colormap shows varying stellar irradiations; the size of the dots show varying ground pressures in atm.

the other hand, the values of the thermal reaction rates are comparable, because of the similarity of the input temperature profiles.

For low irradiations, not a large variation is reflected in the ozone concentration because of pressure, but a clear distinction appears when plotting the height of the maximum value, which is shifted towards higher values the higher the pressure. Physically, the value of J_{O_2} and J_{O_3} at the top of the atmosphere is the same for the same model at the three values of ground pressure. Then, the photodissociation rates decrease more rapidly the higher the density of the atmosphere, determining the altitude of the maximum abundance.

This trend is inverted for higher temperatures and higher irradiations, for which higher pressure values do translate to higher peak abundances and lower heights. The models experiencing this reverse effect are, as a matter of fact, the ones with the highest values of ground pressures ($P_{gr} > 30\ atm$) and correspond to the most massive planets on the sample. The higher surface gravitational acceleration experienced by these atmospheres shrinks them – and, in particular, the mean altitude of the ozone layer is translated to lower values.

The discussion of the pressure effect may be however particularly biased by the absence of many low-pressure models at high irradiations.

A similar behaviour with respect to pressure is visible in Figure 8.11, studying the thickness of the ozone layer. The more irradiated models at the highest values of pressure are the ones with the thinnest ozone layer (about 35 km). Lowering the pressure value translates into an increase in the thickness of the ozone layer to its maximum values (about 80 km).

The models at lower irradiations, on the other hand, do not experience any noticeable pressure effect in varying the thickness of the ozone layer, possibly because of the lower mass of these planets. Broader ozone layers appear, on the other hand, for increasing atmospheric temperatures.

As previously stated, the change in stellar temperature within the `EOM` simply provides varying values of UV radiation in the wavelength range considered, since the atmosphere is not heated by the rest of the incoming radiation and the temperature profile is held fixed. The effect produced by this parameter must be related to the value of I_0, which impacts the photodissociation rates.

For similar average temperatures (moving in the vertical direction of Figure 8.12), the change in stellar temperature produces up to an order of magnitude more ozone, showing that quiet M stars are more likely to produce less ozone compared to hotter

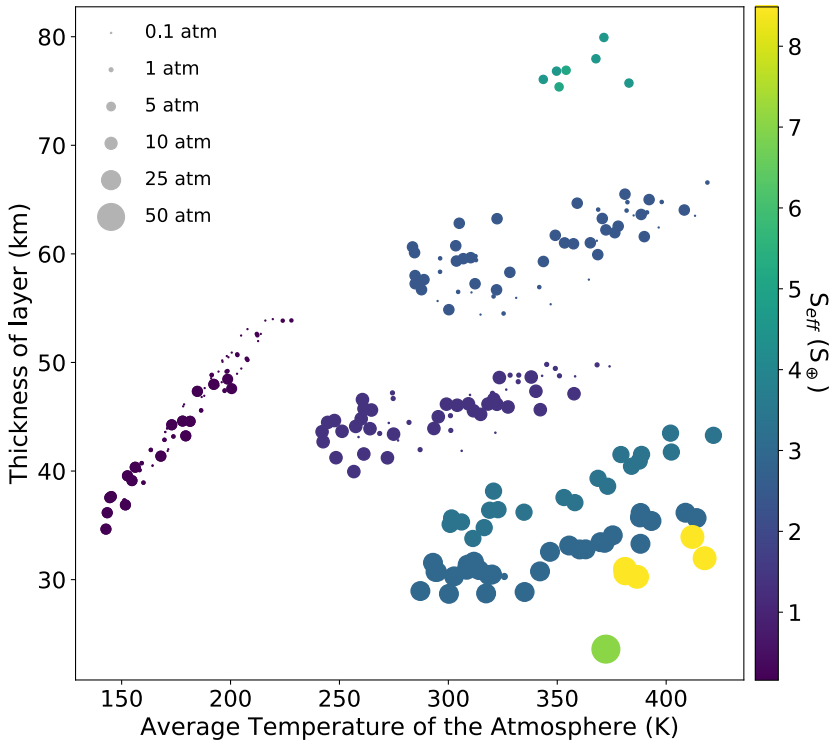


Fig. 8.11: Thickness of the ozone layer in km vs average atmospheric temperature in K. The colormap shows varying stellar irradiations; the size of the dots show varying ground pressures in atm.

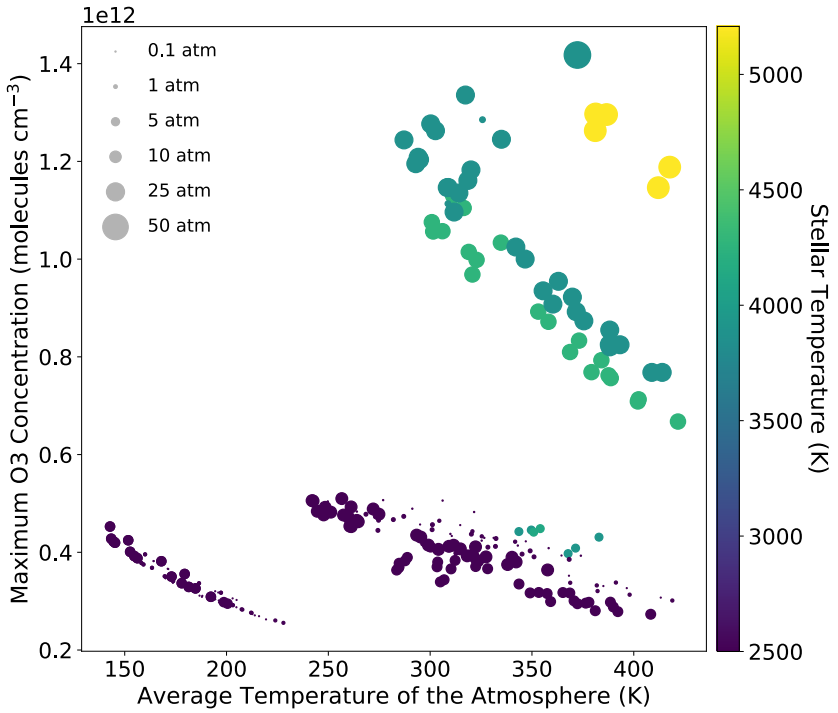


Fig. 8.12: Maximum ozone concentration vs average atmospheric temperature. The colormap shows varying stellar temperatures; the size of the dots show varying ground pressures.

stars because of the lack of UV photons. By consequence, it is probable to expect that active stars would cause the production of more ozone with respect to their quiet counterparts since the radiation in the wavelength bin between 200 and 400 nm would increase – similarly to what happens for hotter stars.

The height of the maximum of ozone (Figure 8.13) and the thickness of the layer (Figure 8.14) reflect what previously noticed: the ozone layer is thinner and lies at lower altitudes in the atmosphere for the bigger planets orbiting hotter stars (which are also highly irradiated). For the TRAPPIST-1 system (the dark blue population) the stellar temperature cannot produce any variation, being obviously the same for all planets in the system: the differences in altitude and thickness therefore rely on the variation of effective flux and on the atmospheric temperature previously explored.

The lack of regularity in the construction of the input grid of models forbids us from retrieving more accurate information. However, what emerges from this qualitative analysis is still important: the combination of irradiation, stellar temperature and planetary pressure does contribute to all the features of the ozone profile, enhancing its abundance and setting the width and the altitude of the ozone layer. This may

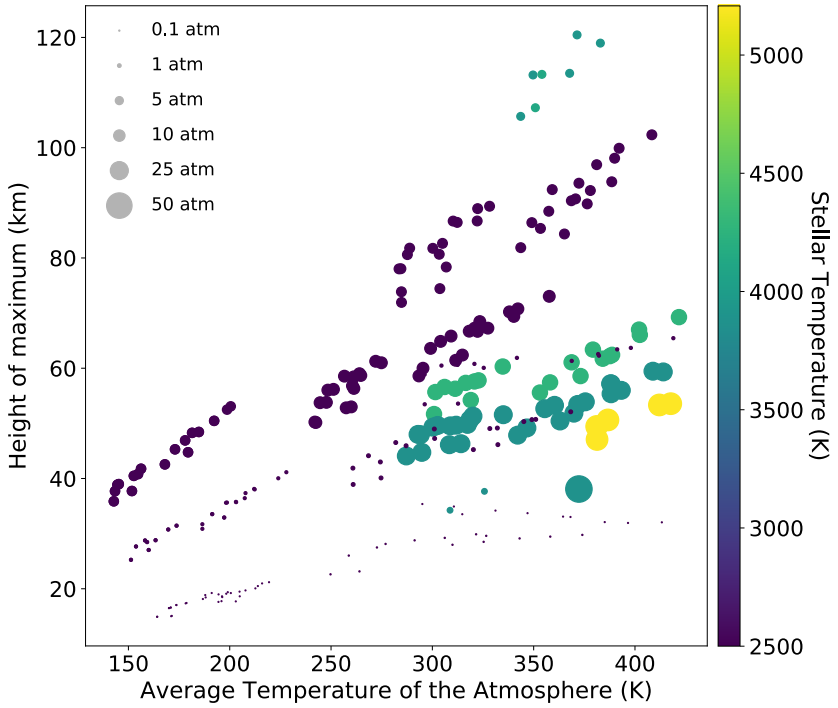


Fig. 8.13: Altitude in km corresponding to the maximum ozone concentration vs average atmospheric temperature in K. The colormap shows varying stellar temperatures in K; the size of the dots show varying ground pressures in atm.

produce some interesting effects that could favor the habitability on the surface of these massive planets, but at the same time threaten the survival of any life form because of the increased greenhouse effect.

To have a clearer overview of the ozone production in exoplanets, a more systematic grid of models is required. This will be performed in the future, allowing more precise conclusions to be drawn concerning this topic.

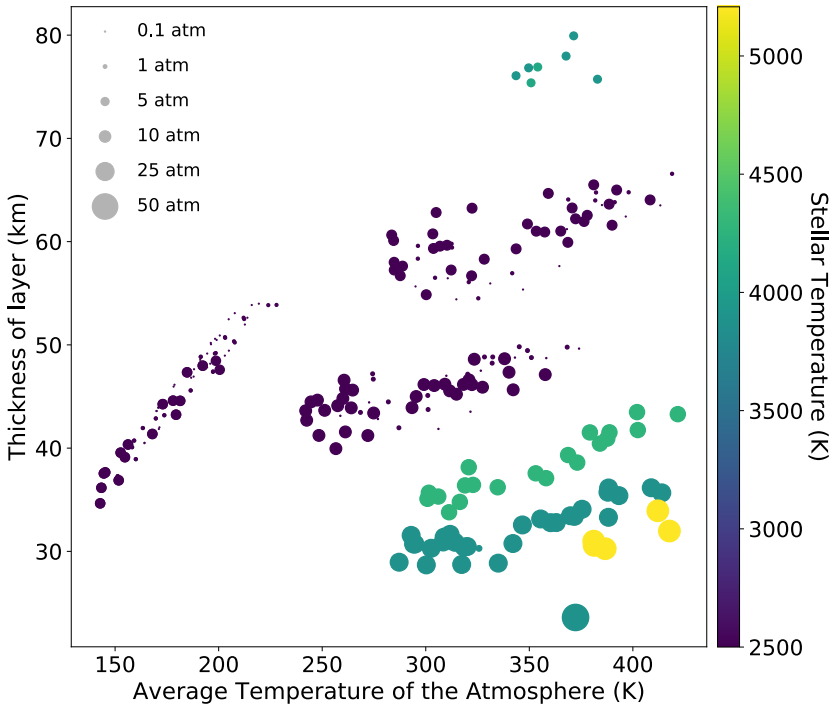


Fig. 8.14: Thickness of the ozone layer in km vs average atmospheric temperature in K. The colormap shows varying stellar temperatures in K; the size of the dots show varying ground pressures in atm.

9 “Atmosphere in a Test-Tube”: laboratory experiments

> *Tonight the hand of man reaches out*
> *To throw light on how life came about,*
> *Computer is reckoning an all-time high,*
> *The future is beckoning onward and onward we fly.*
>
> **— Brian May (1947-present)**
> English musician, singer, songwriter, and astrophysicist

The theoretical set of tools developed and described up to this point are useful for laboratory experiments as well. In this short chapter, the “Atmosphere in a Test-Tube” project is outlined. In particular, some of the simulations of the terrestrial atmospheres run with `MAGRATHEA` (as seen in Chapter 7) are a perfect guideline for the input pressure, temperature, and chemical composition conditions of the environments that are reproduced in the laboratory by biologists of the Biology Department of the University of Padua. The measurement of oxygen production by biological samples exposed to these altered environments should improve the theoretical treatment of the biosphere-atmosphere interaction processes.

This constitutes, therefore, a privileged example of a direct connection between simulations and experiments.

9.1 Aim

The Atmosphere in a Test Tube (“ATM_ITT”) project is a tight collaboration, lead by the Observatory of Padua, between astrophysicists, biologists and engineers of Padua, Florence, Rome and Palermo Universities and Research Centers teaming up to investigate in laboratory the main modification paths of chemistry and physical parameters of simulated exoplanets and Solar System planets due to the surrounding

environment, as well as the physical extremes that constrain life under non-Earth conditions and the methods for characterization by spectroscopic techniques. It represents the first Italian attempt to build an interdisciplinary collaboration in such an innovative research field.

The main aims of such collaboration are:

- to build a database of planetary atmospheres and surface reflections;
- to define procedures that could be used in *in situ* exploration and remote characterization of terrestrial exoplanets;
- to provide new tools for the detection and recognition of biosignatures;
- to broaden the knowledge concerning the adaptability of terrestrial extremophiles to spatial and atmospheric environments different from an Earth around a Sun-like star;
- to plan, integrate, and test new irradiation sources and tools to be used in laboratory experiments;
- to find strategies, methodologies, and tools for the development of instrumentation for new ground- and space-based facilities, and for the interpretation of the already available data for the characterization of planetary atmospheres.

The "ATM_ITT" group in Padua involves people from INAF – Astronomical Observatory, the Department of Biology of the University of Padua, and CNR IFN LUXOR. The main goal is to perform laboratory experiments to understand the behavior of photosynthetic biota in the presence of exotic environments, studying the impact that life has on changing the atmosphere of a planet. The research focuses also on the study of the VRE feature and its variation according to a change in stellar effective temperature [Claudi, La Rocca, et al., 2018].

To do that, a stellar simulator and an environmental chamber were built (see Section 9.2). Bacteria are put in the chamber, then the atmospheric composition of the proxy atmosphere inside the chamber itself is changed, as well as their temperature and pressure. Then, the chamber is positioned under the stellar simulator, so that the proxy biosphere-atmosphere system could evolve in time.

The project is composed of three phases: firstly, a fiduciary experiment measures the photosynthetic bacteria O_2 production in terrestrial conditions using the solar simulator; then, the same measure is repeated after changing the light source to an M7 star simulator; finally, the atmosphere inside the environmental chamber is

changed according to the results of the radiative-convective software MAGRATHEA to simulate Super-Earth-like atmospheres orbiting both solar-like and M7 stars.

9.2 Setup

A scheme of the experimental setup is shown in Figure 9.1. It is composed by a starlight simulator (formed by the lighting source and the feedback spectrometer) on top of an environmental chamber, where biota used for the experiments are hosted, attached to pipes that allow the modification of the atmosphere and instruments that monitor the O_2 and CO_2 production, or control the temperature within the chamber. The entire system is isolated in a dark container cooled by two fans and can be remotely inspected through a webcam [Salasnich et al., 2018].

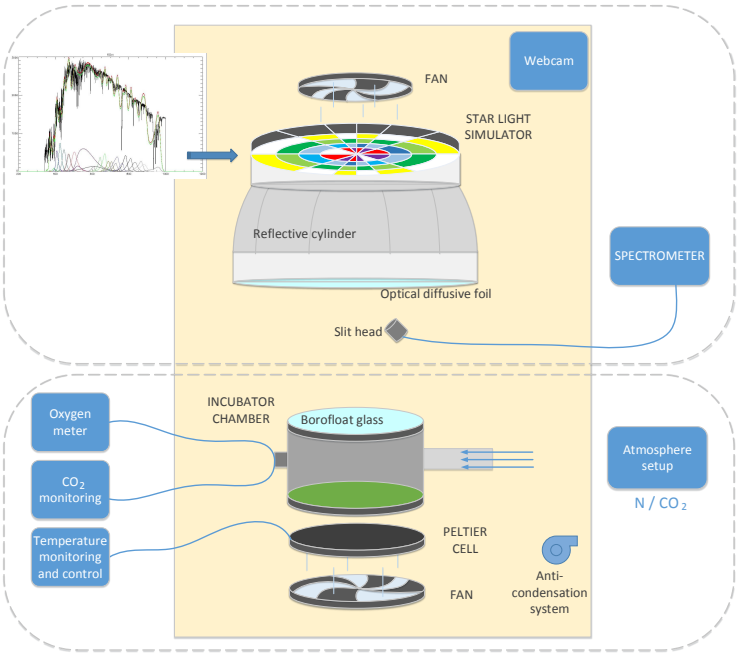


Fig. 9.1: Schematic overview of the experimental setup [Salasnich et al., 2018].

The multi-channel led starlight simulator was built to reproduce the radiation of stars of F, G, K, and M spectral types in the 365-940 nm range. It is composed by twenty-five led channels and a total of 273 diodes arranged in five concentric rings [Erculiani et al., 2015]. The modular design allows quick and easy maintenance in case of damage. The disposition of the diodes has been designed to reduce the

non-uniformity of the flux; plus, a reflective cylinder and an optical diffusive foil were mounted to enhance uniformity.

The control software allows the user to vary the intensity of each class of LEDs to best reproduce the desired spectra. Some default spectra of different stellar fluxes (representing different stars and various semi-major axes) are stored in a library folder; the user can choose the amount of desired radiation in the *photosynthetically active radiation* "PAR" interval, from about 400 to 700 nm (the band that is mainly used by terrestrial photosynthesizers) to simulate any form of shielding by rocks or atmosphere [Salasnich et al., 2018]. The feedback spectrometer detects the output spectra produced by the simulator and allows the software to verify and update the intensity of each type of led to reproduce at best the desired stellar spectrum.

The simulator can be used even for photo-bioreactors, microscopy, yeasts growth and colorimetric applications. Other prototypes are scheduled as well, to improve the performances and adapt the device to further purposes [Erculiani et al., 2015].

The incubator chamber is a steel cylinder in which light enters through a borosilicate glass, thermally resistant and not filtering the radiation. The atmosphere in the chamber can be controlled by a gas line which fluxes carbon dioxide up to a user-determined value. The chamber is kept at a constant temperature employing a Peltier cell which supports the cylinder.

Oxygen and carbon dioxide sensors are connected to the chamber to measure the variation of these species at any point in time. A second spectrograph is inserted into a special structure that can be put over the borofloat glass and allows only a portion of the incoming light to irradiate the sample. The corresponding reflectance spectrum is retrieved by the spectrograph allowing remote measurements of the NDVI index without opening the chamber.

9.3 Experiments

Many different organisms are considered for the experiments, both model and atypical. Some of them are characterized by the presence of chlorophylls (Chl a and b) and an *in vivo* absorption major peak around 680 nm. Other peculiar photosynthetic microorganisms chosen for such experiments can extend their *in vivo* absorption to the near-infrared, due to particular rearrangement of the Chl a in their photosystems, or to the presence of other chlorophyll forms such as Chl d and f [La Rocca et al., 2015].

The validation experiments, performed by biologists in the Biology Department at the University of Padua, already gave some interesting results, as shown in Figure 9.2: the stellar radiation is changed throughout the experiment and the growth of the sample is measured in time via remote NDVI measurements, which appear to be in agreement with the initial and final measurements of optical density of the sample. The production of oxygen and the decrease in carbon dioxide is effectively measured as well. A more detailed analysis of this and many other experiments will appear in Battistuzzi [2020, Ph.D. thesis, in prep.].

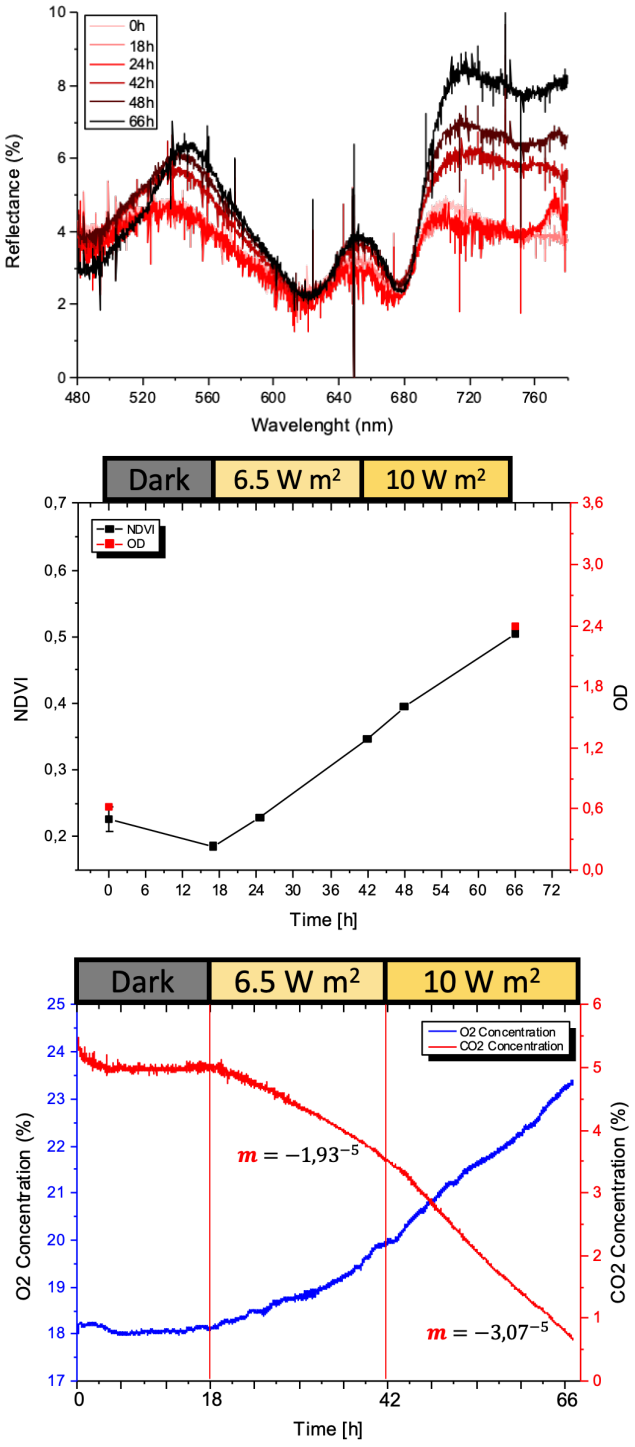


Fig. 9.2: Data acquired during an experiment of bacteria exposed to Solar-like spectrum at varying intensities in the PAR range (top: reflectance spectra; center: optical density and NDVI measurements in time; bottom: O_2 and CO_2 abundances in time); the atmospheric average parameters are T = 303 K, P = 1 atm, atmospheric composition: 75 %N_2, 20 %O_2, 5 %CO_2 [Battistuzzi et al., 2018].

10 Conclusions

In an era where the discovery of life in the universe may be within reach, a theoretical approach to the habitability of terrestrial exoplanets can be very useful.

At present, it is still challenging to assess quantitatively the impact that each parameter – or process – has on the surface conditions, essential for the development and survival of life forms. Determining the presence of a surface can be difficult in the first place since the current mass and radius measurements of Super Earths and Earth-like planets suffer from large errors. The internal composition of the known targets may, therefore, be extremely different, from ocean worlds with light envelopes to iron-rich cores with a very thin atmosphere.

On the other hand, the modeling of the atmospheres of such exoplanets is prone to extreme complication, due to the coupling among all concurrent processes, which include radiative transfer, convection, dynamics, escape, thermal and photon-driven chemistry, as well as cyclic reactions that involve both the surface and the atmosphere. These determine the evolution of a planet, by controlling the constituents of the atmosphere itself.

The stellar environment plays a determining role in the chemical composition and structure as well, enhancing the formation of photochemical hazes as well as possibly favoring the escape of the lighter elements from the atmosphere, or stripping the whole atmosphere from the planet leaving bare rocky Mercury-like cores. If life appears and evolves, its impact on the other "spheres" cannot be neglected: it is indeed the presence of life to determine the actual composition of the atmosphere.

This Ph.D. project has approached the fascinating problem of the habitability of Super Earths atmospheres, concerning both statistics, theoretical modeling, and laboratory experiments.

Chapter 1 summarizes the aim of the thesis, the tools and the methods. Then, after an overview of the most efficient methods for detecting exoplanets in Chapter 2, the current knowledge concerning the formation, the internal structure, and the atmospheres of Super Earths has been reported in Chapter 3. Chapter 4 dealt with the habitability of a planet, defining the habitable zone, the factors that affect

habitability, and the spectral fingerprints that could suggest the presence of life when observing an atmosphere.

The first milestone accomplished during the Ph.D. was the construction of `Exo-MerCat` (described in Chapter 5), a Python code that collects and selects the most precise measurement for all interesting planetary and orbital parameters contained in the four most used online databases, accounting for the presence of multiple aliases for the same target. `Exo-MerCat` can correct as many issues that prevent a direct correspondence between multiple items in the four databases as possible, with the available data. It is also able to download information about the host star by use of Virtual Observatory ConeSearch connections to the major archives such as SIMBAD and those available in VizieR. A Graphical User Interface is provided to filter data based on the user's constraints and generate automatic plots that are commonly used in the exoplanetary community.

With `Exo-MerCat`, we retrieved a unique catalog that merges information from the four main databases, standardizing the output and handling notation differences issues. The catalog is available for all VO-aware TAP-enabled client applications and it is periodically updated. It relies on VO tools and standards, in a perspective of the more and more common usage of such technologies in the future, to ease datasets availability, maintenance, and coherent analysis.

Using `Exo-MerCat`, a sample of all currently known Super Earths to date was collected and statistically described. From this sample, a set of targets was selected for the atmospheric simulations.

A large part of the Ph.D. project dealt with the construction and validation of `MAGRATHEA` (described in Chapter 6), a 1D radiative-convective code that is supposed to model the average temperature profile of the atmosphere of terrestrial exoplanets, exploring a wide range in pressures, as well as temperatures from 100 to 500 K. The chemical database currently included is composed of carbon dioxide, molecular oxygen, water vapor, and nitrogen: it is thus possible to model Mars-like atmospheres, as well as terrestrial ones.

The code treats the absorption of each atmospheric layer through the k correlated distribution, which is much faster than the usual line-by-line approach. This allowed us to model a large set of theoretical Super Earths (described in Chapter 7), selecting various combinations of planetary masses, atmospheric compositions, stellar temperatures, and semi-major axes, for a total of about 18000 models. A principal component analysis was then performed on the sample, to understand the parameters that have the largest effects on the shape on the pressure-temperature

profiles. Further considerations concerning the instability of some atmospheres and the thermodynamic habitability of these models were also made.

The same analysis was carried for a sample of nearly 2400 atmospheres considering some observed Super Earths (in Chapter 7 as well), previously retrieved by `Exo-MerCat`. Albeit more models failed due to the higher irradiances experienced by those planets, which led to higher temperatures and therefore a higher chance to trespass the 500 K boundary and/or to produce too much water vapor, a substantial portion of models were well within the thermodynamic range of surface liquid water.

Most interestingly, K2-18 b, in whose atmosphere water vapor has been recently detected, was included in this sample of potentially habitable terrestrial exoplanets, and `MAGRATHEA` could, therefore, provide further information about this target in particular, as well as the already-exploited planets of the TRAPPIST-1 system.

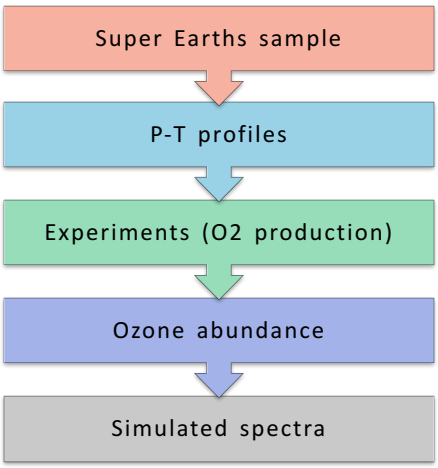
A subset of this sample (all the successful models with molecular oxygen in their atmospheres) was then used to calculate the ozone profile corresponding to the radiative-convective equilibrium profile produced by `MAGRATHEA`.

To do this, the `Exoplanet Ozone Model` was built (as described in Chapter 8). The development of this software is still at its first steps, but a few interesting results have been retrieved already, shortly after its validation.

The abundance of ozone is strictly linked to the ultraviolet flux, suggesting that Solar-like stars may be producing more ozone in comparison with cool dwarfs. The activity of the star has not yet been taken into account, so the reality of nature is very likely to be more complicated.

This doctoral work is included in the "Atmosphere in a Test-Tube" project, whose aim is described in Chapter 9. In the Biology Department of the University of Padua, experiments on cyanobacteria are currently in progress to study the behavior of such species in altering their photochemical apparatus and producing O_2 in exotic conditions, such as those that could probably happen on the surfaces of terrestrial exoplanets.

A graphical overview of the four Ph.D. projects, the tools developed and their status is shown in Figure 10.1; in addition to those, the last item on the infographic represents the main topic of the postdoctoral fellowship, which will focus on the modeling of theoretical spectra of terrestrial exoplanetary atmospheres.



Tools	Status
Exo-MerCat	Completed
MAGRATHEA	Completed
ATM_ITT	In progress
EOM	Validated
T.B.D.	W.I.P.

Fig. 10.1: Infographic of the four Ph.D. projects, the tools developed for each purpose (Exo-MerCat, MAGRATHEA, EOM, and ATM_ITT) and their status. The last project will be treated during the postdoctoral fellowship.

10.1 Future Work

As visible from Figure 10.1, more studies on the various topics can be made, together with improvements on the various tools.

The theoretical work must rely on the observations, even more so in such an engaging field as exoplanetology. All software must be constantly adapted to reproduce the data with a higher degree of accuracy. Only after a fine-tuning with what it is currently known, it could be possible to have higher confidence in what needs to be extrapolated by numerical modeling.

First of all, the implementation of new features on `Exo-MerCat` has already been scheduled. A history log of the various updates on the catalog will be made to ensure the possibility to query for older versions of the catalog, allowing the study of the time-dependent variation of the most precise measurements of the parameters. A more automated link to the major stellar catalogs is also foreseen, to provide updated information on the host star. Other catalog sources may be added, e.g. TEPCat[1] [Southworth, 2011], and direct connection to data should be provided as well.

At the same time, we are spreading the awareness concerning the need for a Data Model for exoplanets in the Virtual Observatory community, to improve the methodology used by the current archives.

For what concerns `MAGRATHEA`, the performed models underlined the necessity to enlarge the look-up table of k-correlated distributions, aiming at the study of more irradiated atmospheres, possibly with a wider variety of compositions. A change in assumptions concerning the albedo and the emissivity may also be essential to achieve more realistic results.

A major improvement would be the treatment of clouds. We are currently working on stochastic modeling of clouds, assuming that along a ray the casual alternation of the gaseous/cloudy phase is a stochastic process, with markovian transitions between states. This feature could allow modeling uneven decks of clouds, which are common in the atmosphere of the Earth.

The `Exoplanet Ozone Model` is at present validated, but the treatment is extremely simplified: as previously seen, the heating of the atmosphere is not taken into account and the thermal profile is set as fixed throughout the entire run. This is obviously not realistic since the thermal heating of the atmosphere and the photochemistry of

[1] `https://www.astro.keele.ac.uk/jkt/tepcat/`

ozone happen simultaneously and interact with each other. Indeed, as a first step, some strong assumptions needed to be made, but these can be released in the future, coupling the EOM with MAGRATHEA to take into account both processes at the same time.

The stellar activity should be also considered for the production of this biosignature, allowing the Exoplanet Ozone Model to be more reliable when modeling planets orbiting cool M-dwarfs. A variation of the stellar intensity in time has to be implemented as well, to simulate flaring events.

The "Atmosphere in a Test Tube" project is characterized by some interesting future milestones, concerning the testing and the validation of other laboratory facilities for biological, geophysical, and astrophysical purposes. All these projects are supposed to work together towards a better characterization of both the surface conditions on terrestrial exoplanets and the detectability of gaseous, surface, and temporal biosignatures in future observations.

What essentially links these theoretical and laboratory projects with the reality of nature is then the possibility to detect hints of life in the observed exoplanetary spectra. These are expected in the next few years: the theoretical determination of the spectral features that could be traceable to life is, therefore, essential now more than ever.

Many open-source software exist to produce simulations of exoplanetary spectra at high resolution (TAU [Hollis et al., 2013], BART [Cubillos et al., 2016], ExoTransmit [Kempton et al., 2017] just to mention a few). To achieve this goal, a simplified pressure-temperature profile is often taken into account, discarding any impact that the composition of the atmosphere, the dynamics, and the stellar environment have on the profile itself.

The usage of MAGRATHEA, coupled with one of these retrieval codes, could provide more realistic spectra, even filling up a grid of theoretical transmission spectra that span a wide range of parameters such as planetary dimensions and atmospheric compositions (abundances, clouds, albedo), as well as varying conditions of the star-planet system (distance, stellar class).

Then, it would be possible to simulate observation with current and future facilities: this could be achieved by convolving a theoretical model with the line-spread function (LSF) of any instrument, then accounting for resolution and wavelength normalization (see e.g. Pino et al. (2018)). Such spectra would be useful for the analysis of observations, allowing us to have information about a real spectrum by matching or interpolating it with simulated ones.

The milestones of this project, which deals with the production of transmission spectra, have yet to be determined. Nevertheless, these are of the utmost importance to allow such theoretical studies to be useful for observations and to be constantly adjusted in return according to those.

As a matter of fact, a code built to retrieve information about an unknown reality should be constantly adjusted to take into account anything we have learned from the ongoing observation. Only then, a database of known possible and realistic scenarios could allow the correct analysis of terrestrial exoplanets atmospheres - and of biosignatures in particular - to be easier and achievable.

Useful Values and Constants

Universal	
Gravitational constant G	$6.673 \cdot 10^{-11}\ N\ m^2\ kg^{-2}$
Universal Gas Constant ($R*$)	$8.3145\ J\ K^{-1}\ mol^{-1}$
Boltzmann's constant k_B	$1.381 \cdot 10^{-23}\ J\ K^{-1}\ molecule^{-1}$
Stefan-Boltzmann constant σ_B	$5.671 \cdot 10^{-8}\ W\ K^{-4}\ m^{-2}$
Wien's displacement constant b	$2.898 \cdot 10^{-3}\ m\ K$
Planck's constant h	$6.626 \cdot 10^{-34}\ m^2\ kg\ s^{-1}$
Solar system	
Solar Mass ($M_\odot$)	$1.989 \cdot 10^{30}\ kg$
Solar Radius ($R_\odot$)	$6.960 \cdot 10^{8}\ m$
Jupiter Mass (M_J)	$1.898 \cdot 10^{27}\ kg$
Jupiter Radius (R_J)	$6.991 \cdot 10^{7}\ m$
Astronomical Unit (AU)	$1.496 \cdot 10^{11}\ m$
Earth	
Mass ($M_\oplus$)	$5.983 \cdot 10^{24}\ kg$
Mean Radius ($R_\oplus$)	$6.367 \cdot 10^{6}\ m$
Mean gravitational acceleration (g)	$9.8067\ m\ s^{-2}$
Solar constant (S_0)	$1370\ W\ m^{-2}$
Dry air gas constant (R_d)	$287.0\ J\ K^{-1}\ kg^{-1}$
Dry air molecular weight (M_d)	$28.97\ g\ mol^{-1}$
Dry air specific heat at constant pressure (c_p)	$1004\ J\ K^{-1}\ kg^{-1}$
Dry air specific heat at constant volume (c_v)	$717\ J\ K^{-1}\ kg^{-1}$
Water vapor gas constant (R_w)	$461.5\ J\ K^{-1}\ kg^{-1}$
Water molecular weight	$18.015\ g\ mol^{-1}$

Publication List

Accept. **Alei, E.**; Claudi, R.; Bignamini, A.; Molinaro, M. (2019) *Exo-MerCat: a merged exoplanet catalog with Virtual Observatory connection* accepted by *Astronomy and Computing*.

Subm. Petralia, A.; **Alei, E.**; Aresu, G.; Locci, D.; Cecchi-Pestellini, C.; Micela, G.; Claudi, R.; Ciaravella, A. (2019) *A systematic study of CO_2 planetary atmospheres and their link to the stellar environment* submitted to *Monthly Notices of the Astronomical Society*.

Subm. **Alei, E.**; Claudi, R.; Bignamini, A.; Molinaro, M. (2019) *Exo-MerCat and Virtual Observatory: towards a Data Model for exoplanets.* submitted to *ADASS Conference Proceedings*.

In prep. **Alei, E.**; Claudi, R. (2019) *1D radiative-convective models of observed temperate Super Earths* to be submitted to *The Astrophysical Journal Letters*.

In prep. Claudi, R; **Alei, E.**; D'Orazi, V.; Marzari, F. (2019) *Core mass function of observed Super Earths* to be submitted to *Astronomy and Astrophysics*.

In prep. Zinzi, A.; Turrini, D.; Verrecchia, F.; **Alei, E.** (2019) *SSDC ExoplAn3T: a webtool helping unveiling connections between extrasolar systems* to be submitted to *Astronomy and Computing*.

In prep. **Alei, E.**; Locci, D.; Petralia, A.; Cecchi-Pestellini, C.; Micela, G.; Claudi, R.; Ciaravella, A. (2019) *Radiative-convective models of exoplanetary atmospheres around low-mass stars: water vapor treatment and habitability* to be submitted to *Astronomy and Astrophysics*.

In prep. Claudi, R.; Erculiani, M. S.; Barbisan, D.; Cocola, L; Trivellin, N.; **Alei, E.**; Battistuzzi, M.; Farisato, G; La Rocca, N.; Poletto, L.; Salasnich, B.; Pace, E. (2019) *Multi – Channel LED Starlight Simulator for Laboratory Experiments* to be submitted to *Review of Scientific Instruments*.

In prep. Carleo, I.; Malavolta, L.; Lanza, A.F.; Damasso, M.; Mallon, M.; **Alei, E.**; + GAPS Team. (2019) *The GAPS Programme with GIARPS at TNG I. Observing young stars: first disclaimed and confirmed exoplanets* to be submitted to *Astronomy and Astrophysics*.

Sept 2019 Claudi, R; **Alei, E.** (2019) *Biosignatures Search in Habitable Planets*, *Galaxies*, 7(4), 82, 10.3390/galaxies7040082.

July 2018 Salasnich, B.; Claudi, R.; **Alei, E.**; Barbisan, D.; Baruffolo, A.; Cocola, L.; Erculiani, M. S.; La Rocca, N.; Pace, E.; ,Poletto, L.; Trivellin, N. (2018) *Control software for the Multi-Channel Led starlight simulator*, SPIE 10707, Software and Cyberinfrastructure for Astronomy V, 107071I, 10.1117/12.2311436.

Dec 2018 Molinaro, M., **Alei, E.**, Benatti, S., Bignamini, A., Bonnarel, F., Damasso, M., Louys, M., Maris, M., Nascimbeni, V. (2018) *Starting up a Data Model for Exoplanetary Data*, Astronomical Society of the Pacific Conference Series Proceedings.

Aug 2018 **Alei, E.** (2018) *CINECA INA17_C2BB0: report and preliminary results* accepted by ICT and Science Data management website.

Bibliography

Abe, Y., Abe-Ouchi, A., Sleep, N. H., et al. (June 2011). “Habitable Zone Limits for Dry Planets”. In: *Astrobiology* 11.5, pp. 443–460 (cit. on p. 86).

Airapetian, V. S., Jackman, C. H., Mlynczak, M., et al. (Nov. 2017). “Atmospheric Beacons of Life from Exoplanets Around G and K Stars”. In: *Scientific Reports* 7, 14141, p. 14141 (cit. on p. 95).

Akeson, R. L., Chen X.and Ciardi, D., Crane, M, et al. (2013). “The NASA Exoplanet Archive: Data and Tools for Exoplanet Research”. In: *Publications of the Astronomical Society of the Pacific* 125.930, pp. 989–999 (cit. on pp. 103, 108).

Alei, E., Claudi, R., Bignamini, A., et al. (2019). “Exo-MerCat: a merged exoplanet catalog with Virtual Observatory connection.” submitted (cit. on p. 141).

Alibert, Y., Ataiee, S., and Venturini, J. (2018). “Planet Formation, Migration, and Habitability”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 2879–2895 (cit. on pp. 35, 36).

Alonso, R. (2018). “Characterization of Exoplanets: Secondary Eclipses”. en. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–26 (cit. on p. 18).

Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. (2016). “A terrestrial planet candidate in a temperate orbit around Proxima Centauri”. In: *Nature* 536.7617, pp. 437–440 (cit. on pp. 1, 32).

Appenzeller, I., Fricke, K., Fürtig, W., et al. (Dec. 1998). “Successful commissioning of FORS1 - the first optical instrument on the VLT.” In: *The Messenger* 94, pp. 1–6 (cit. on p. 24).

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. (Oct. 2013). “Astropy: A community Python package for astronomy”. In: *A&A* 558, A33, A33. arXiv: 1307.6212 [astro-ph.IM] (cit. on p. 118).

Bailey, V., Meshkat, T., Reiter, M., et al. (Dec. 2013). “HD 106906 b: A planetary-mass companion outside a massive debris disk”. In: *The Astrophysical Journal* 780.1. arXiv: 1312.1265, p. L4 (cit. on p. 134).

Bakos, G. Á., Lázár, J., Papp, I., et al. (Sept. 2002). “System Description and First Light Curves of the Hungarian Automated Telescope, an Autonomous Observatory for Variability Search”. In: *Publications of the Astronomical Society of the Pacific* 114.799, pp. 974–987. arXiv: astro-ph/0206001 [astro-ph] (cit. on p. 20).

Bashi, D., Helled, R., and Zucker, S. (2018). “A Quantitative Comparison of Exoplanet Catalogs”. In: *Geosciences* (cit. on pp. 104, 111, 114).

Battistuzzi, M., Claudi, R., Cocola, L., et al. (2018). “An experimental setup to study by remote sensing analyses cyanobacteria growth and photosynthetic performances under non-terrestrial simulated environments”. In: *18th EANA Conference*. Ed. by European Astrobiology Network Association (cit. on p. 282).

Beaulieu, J. P., Bennett, D. P., Fouqué, P., et al. (Jan. 2006). “Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing”. In: *Nature* 439.7075, pp. 437–440. arXiv: astro-ph/0601563 [astro-ph] (cit. on p. 23).

Benatti, S. (Aug. 2018). “Multi-Wavelength High-Resolution Spectroscopy for Exoplanet Detection: Motivation, Instrumentation and First Results”. en. In: *Geosciences* 8.8, p. 289 (cit. on pp. 7, 10).

Benneke, B., Wong, I., Piaulet, C., et al. (Sept. 2019). “Water Vapor on the Habitable-Zone Exoplanet K2-18b”. In: *arXiv e-prints*, arXiv:1909.04642, arXiv:1909.04642. arXiv: 1909.04642 [astro-ph.EP] (cit. on p. 33).

Bétrémieux, Y. and Kaltenegger, L. (Aug. 2013). “Transmission Spectrum of Earth as a Transiting Exoplanet from the Ultraviolet to the Near-infrared”. In: *Astrophysical Journal Letters* 772.2, L31, p. L31. arXiv: 1307.0416 [astro-ph.EP] (cit. on p. 94).

Beuzit, J. L., Vigan, A., Mouillet, D., et al. (Feb. 2019). “SPHERE: the exoplanet imager for the Very Large Telescope”. In: *arXiv e-prints*, arXiv:1902.04080, arXiv:1902.04080. arXiv: 1902.04080 [astro-ph.IM] (cit. on p. 21).

Borucki, W. J., Koch, D., Basri, G., et al. (Feb. 2010). “Kepler Planet-Detection Mission: Introduction and First Results”. In: *Science* 327.5968, p. 977 (cit. on p. 20).

Boss, A. P., Astraatmadja, T. L., and Weinberger, A. J. (Jan. 2019). “Ground-based Astrometric Detection of Exoplanets with CAPSCam: Current Status”. In: *American Astronomical Society Meeting Abstracts #233*. Vol. 233. American Astronomical Society Meeting Abstracts, p. 408.08 (cit. on p. 24).

Boyajian, T. S., von Braun, K., van Belle, G., et al. (2012). “Stellar Diameters and Temperatures II. Main Sequence K & M Stars”. In: (cit. on p. 186).

Broeg, C., Fortier, A., Ehrenreich, D., et al. (Apr. 2013). “CHEOPS: A transit photometry mission for ESA’s small mission programme”. In: *European Physical Journal Web of Conferences*. Vol. 47. European Physical Journal Web of Conferences, p. 03005. arXiv: 1305.2270 [astro-ph.EP] (cit. on p. 20).

Brown, T. M. (May 2001). “Transmission Spectra as Diagnostics of Extrasolar Giant Planet Atmospheres”. In: *The Astrophysical Journal* 553.2, pp. 1006–1026 (cit. on p. 15).

Buck, A. L. (Dec. 1981). “New Equations for Computing Vapor Pressure and Enhancement Factor”. In: *Journal of Applied Meteorology* 20.12, pp. 1527–1532 (cit. on p. 169).

Budyko, M. I. (Oct. 1969). “The effect of solar radiation variations on the climate of the Earth”. In: *Tellus Series A* 21.5, pp. 611–619 (cit. on p. 86).

Butler, R. P., Wright, J. T., Marcy, G. W., et al. (July 2006). “Catalog of Nearby Exoplanets”. In: *ApJ* 646, pp. 505–522. eprint: astro-ph/0607493 (cit. on p. 107).

Cameron, A. C. (2016). “Extrasolar Planetary Transits”. en. In: *Methods of Detecting Exoplanets*. Ed. by Valerio Bozza, Luigi Mancini, and Alessandro Sozzetti. Vol. 428. Springer International Publishing, pp. 89–131 (cit. on pp. 12–14).

Carleo, I., Benatti, S., Lanza, A. F., et al. (May 2018). “Multi-band high resolution spectroscopy rules out the hot Jupiter BD+20 1790b. First data from the GIARPS Commissioning”. In: *A&A* 613, A50, A50. arXiv: 1805.01281 [astro-ph.EP] (cit. on p. 104).

Chapman, S. (1932). “Discussion of memoirs. On a theory of upper-atmospheric ozone”. In: *Quarterly Journal of the Royal Meteorological Society* 58.243, pp. 11–13 (cit. on p. 72).

Claudi, R. and Alei, E. (Sept. 2019). “Biosignatures Search in Habitable Planets”. In: *Galaxies* 7.4, p. 82 (cit. on pp. 93, 96, 97).

Claudi, R., Benatti, S., Carleo, I., et al. (Aug. 2017). “GIARPS@TNG: GIANO-B and HARPS-N together for a wider wavelength range spectroscopy”. In: *The European Physical Journal Plus* 132.8, p. 364 (cit. on p. 11).

Claudi, R., La Rocca, N., Poletto, L., et al. (Sept. 2018). “M Dwarfs, Super Earths and photosynthetic bacteria: a mix for laboratory studies”. In: *European Planetary Science Congress*, EPSC2018–228 (cit. on p. 278).

Cosentino, R., Lovis, C., Pepe, F., et al. (Sept. 2012). “Harps-N: the new planet hunter at TNG”. In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Vol. 8446. SPIE Proceedings, p. 84461V (cit. on p. 11).

Crouzet, N. (2018). “Small Telescope Exoplanet Transit Surveys: XO”. In: *Handbook of Exoplanets, ISBN 978-3-319-55332-0. Springer International Publishing AG, part of Springer Nature, 2018, id.129*, p. 129 (cit. on p. 20).

Cubillos, P., Blecic, J., Harrington, J., et al. (Aug. 2016). *BART: Bayesian Atmospheric Radiative Transfer fitting code*. Astrophysics Source Code Library. ascl: 1608.004 (cit. on p. 288).

Deeg, H. J. and Alonso, R. (2018). “Transit Photometry as an Exoplanet Discovery Method”. en. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–25 (cit. on p. 14).

Deming, D., Louie, D., and Sheets, H. (Jan. 2019). “How to Characterize the Atmosphere of a Transiting Exoplanet”. In: *Publications of the Astronomical Society of the Pacific* 131.995. arXiv: 1810.04175, p. 013001 (cit. on p. 18).

Demory, B., Gillon, M., de Wit, J., et al. (Apr. 2016). “A map of the large day-night temperature gradient of a super-Earth exoplanet”. In: *Nature* 532.7598, pp. 207–209. arXiv: 1604.05725 [astro-ph.EP] (cit. on p. 18).

Dole, S. H. (1964). *Habitable planets for man* (cit. on p. 86).

Domagal-Goldman, S. D., Meadows, V. S., Claire, M. W., et al. (June 2011). “Using Biogenic Sulfur Gases as Remotely Detectable Biosignatures on Anoxic Planets”. In: *Astrobiology* 11, pp. 419–441 (cit. on p. 96).

Dowler, P., Rixon, G., and Tody, D. (Mar. 2010). *Table Access Protocol Version 1.0*. Tech. rep., p. 327 (cit. on p. 118).

Elkins-Tanton, L. T. and Seager, S. (2008). “Ranges of Atmospheric Mass and Composition of Super-Earth Exoplanets”. In: *The Astrophysical Journal* 685, pp. 1237–1246 (cit. on p. 37).

Erculiani, M. S., Claudi, R., Barbisan, D., et al. (Sept. 2015). “A tunable integrated system to simulate colder stellar radiation”. In: (cit. on pp. 279, 280).

Extension to the Standard Atmosphere, United States Committee on (1976). *U.S. standard atmosphere, 1976*. National Oceanic and Amospheric [sic] Administration : for sale by the Supt. of Docs., U.S. Govt. Print. Off.

Faherty, J. K. (2018). “Spectral Properties of Brown Dwarfs and Unbound Planetary Mass Objects”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–12 (cit. on p. 136).

Farinato, J., Bacciotti, F., Baffa, C., et al. (Aug. 2018). “SHARK-NIR, the coronagraphic camera for LBT, moving toward construction”. In: *arXiv e-prints*, arXiv:1808.00364, arXiv:1808.00364. arXiv: 1808.00364 [astro-ph.IM] (cit. on p. 22).

Farman, J. C., Gardiner, B. G., and Shanklin, J. D. (May 1985). “Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction”. In: *Nature* 315.6016, pp. 207–210 (cit. on p. 74).

Figueira, P., Marmier, M., Bonfils, X., et al. (Apr. 2010). “Evidence against the young hot-Jupiter around BD +20 1790”. In: *A&A* 513, L8, p. L8. arXiv: 1003.3678 [astro-ph.EP] (cit. on p. 104).

Figueira, P., Pepe, F., Santos, N. C., et al. (Apr. 2010). “TW Hya under CRIRES light: evidence against the presence of a hot Jupiter”. In: *EAS Publications Series*. Ed. by K. Gożdziewski, A. Niedzielski, and J. Schneider. Vol. 42. EAS Publications Series, pp. 125–129 (cit. on p. 104).

Fischer, D. A., Howard, A. W., Laughlin, G. P., et al. (2014). “Exoplanet Detection Techniques”. In: *Protostars and Planets VI*, p. 715 (cit. on pp. 7, 26, 27).

Fogg, M. J. (Jan. 1992). “An Estimate of the Prevalence of Biocompatible and Habitable Planets”. In: *Journal of the British Interplanetary Society* 45, pp. 3–12 (cit. on p. 86).

Forget, F. and Leconte, J (2014). “Possible climates on terrestrial exoplanets.” In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 372.2014, p. 20130084 (cit. on p. 156).

Fu, Q. and Liou, K. N. (1992). “On the Correlated k-Distribution Method for Radiative Transfer in Nonhomogeneous Atmospheres”. In: *Journal of the Atmospheric Sciences* 49.22, pp. 2139–2156. eprint: https://doi.org/10.1175/1520-0469(1992)049<2139:OTCDMF>2.0.CO;2 (cit. on p. 161).

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. (Aug. 2018). “Gaia Data Release 2. Summary of the contents and survey properties”. In: *A&A* 616, A1, A1. arXiv: 1804.09365 (cit. on p. 126).

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. (Nov. 2016). “The Gaia mission”. In: *A&A* 595, A1, A1. arXiv: 1609.04153 [astro-ph.IM] (cit. on pp. 25, 126).

Gardner, J. P., Mather, J. C., Clampin, M., et al. (Apr. 2006). "The James Webb Space Telescope". In: *Space Science Reviews* 123.4, pp. 485–606. arXiv: astro-ph/0606175 [astro-ph] (cit. on p. 20).

Gelman, S. E., Elkins-Tanton, L. T., and Seager, S. (2011). "Effects of Stellar Flux on Tidally Locked Terrestrial Planets: Degree-1 Mantle Convection and Local Magma Ponds". In: *The Astrophysical Journal* 735.2, p. 72 (cit. on p. 83).

Goody, R. M. and Yung, Y. L. (1989). *Atmospheric Radiation: Theoretical Basis*. Oxford University Press, p. 519 (cit. on p. 45).

Gordon, I. E., Rothman, L. S., Hill, C., et al. (Dec. 2017). "The HITRAN2016 molecular spectroscopic database". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 203, pp. 3–69 (cit. on p. 159).

Grenfell, J. L. (2018). "Atmospheric Biosignatures". In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 3159–3172 (cit. on pp. 92, 93).

Guyon, O. (2018). "Imaging with Adaptive Optics and Coronographs for Exoplanet Research". en. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–20 (cit. on p. 21).

Haghighipour, N. (2013). "The Formation and Dynamics of Super-Earth Planets". In: *Annual Review of Earth and Planetary Sciences* 41, pp. 469–495 (cit. on p. 33).

Han, E., Wang, S. X., Wright, J. T., et al. (Sept. 2014). "Exoplanet Orbit Database. II. Updates to Exoplanets.org". In: *Publications of the Astronomical Society of the Pacific* 126, p. 827. arXiv: 1409.7709 [astro-ph.EP] (cit. on p. 109).

Hanel, R. A., Conrath, B. J., Kunde, V. G., et al. (1972). "The Nimbus 4 infrared spectroscopy experiment: 1. Calibrated thermal emission spectra". In: *Journal of Geophysical Research (1896-1977)* 77.15, pp. 2629–2641. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/JC077i015p02629 (cit. on p. 63).

Hart, M. H. (Jan. 1979). "Habitable Zones about Main Sequence Stars". In: *Icarus* 37.1, pp. 351–357 (cit. on p. 86).

Hatzes, A. P. and Rauer, H. (Sept. 2015). "A Definition for Giant Planets Based on the Mass-Density Relationship". In: *Astrophysical Journal* 810.2, L25, p. L25. arXiv: 1506.05097 [astro-ph.EP] (cit. on p. 136).

Hearnshaw, J. B., Abe, F., Bond, I. A., et al. (Jan. 2006). "The MOA 1.8-metre alt-az Wide-field Survey Telescope and the MOA Project". In: *The 9th Asian-Pacific Regional IAU Meeting*. Ed. by W. Sutantyo, P. W. Premadi, P. Mahasena, et al., p. 272. arXiv: astro-ph/0509420 [astro-ph] (cit. on p. 23).

Hessman, F. V. (2018). "The Naming of Extrasolar Planets". In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 85–93 (cit. on pp. 106, 107).

Hollis, M. D. J., Tessenyi, M., and Tinetti, G. (Oct. 2013). "TAU: A 1D radiative transfer code for transmission spectroscopy of extrasolar planet atmospheres". In: *Computer Physics Communications* 184, pp. 2351–2361 (cit. on p. 288).

Howell, S. B., Sobeck, C., Haas, M., et al. (Apr. 2014). "The K2 Mission: Characterization and Early Results". In: *Publications of the Astronomical Society of the Pacific* 126.938, p. 398. arXiv: 1402.5163 [astro-ph.IM] (cit. on p. 20).

Hu, R., Seager, S., and Bains, W. (2012). "Photochemistry in Terrestrial Exoplanet Atmospheres. I. Photochemistry Model and Benchmark Cases". In: *The Astrophysical Journal* 761.2, p. 166 (cit. on pp. 70, 71, 74, 75).

Huber, D., Bryson, S. T., and et al. (Sept. 2017). "VizieR Online Data Catalog: K2 Ecliptic Plane Input Catalog (EPIC) (Huber+, 2017)". In: *VizieR Online Data Catalog*, IV/34, pp. IV/34 (cit. on p. 126).

Huebner, W. F. and Mukherjee, J. (2015). "Photoionization and photodissociation rates in solar and blackbody radiation fields". In: *Planetary and Space Science* 106, pp. 11–45 (cit. on p. 255).

Izidoro, A. and Raymond, S. N. (2018). "Formation of Terrestrial Planets". In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 2365–2423 (cit. on pp. 33, 35, 36).

Jehin, E., Gillon, M., Queloz, D., et al. (Sept. 2011). "TRAPPIST: TRAnsiting Planets and PlanetesImals Small Telescope". In: *The Messenger* 145, pp. 2–6 (cit. on p. 20).

Joliffe, I. T. (2002). *Principal Component Analysis*. Springer-Verlag (cit. on p. 201).

Kaltenegger, L., Selsis, F., Fridlund, M., et al. (2009). "Deciphering Spectral Fingerprints of Habitable Extrasolar Planets". In: p. 17 (cit. on p. 85).

Kasting, J. F. and Catling, D. (2003). "Evolution of a Habitable Planet". In: *Annual Review of Astronomy and Astrophysics* 41.1, pp. 429–463. eprint: https://doi.org/10.1146/annurev.astro.41.071601.170049 (cit. on p. 82).

Kasting, J. F., Whitmire, D. P, and Reynolds, R T (1993). *Habitable zones around main sequence stars.* (Cit. on pp. 84, 86).

Kempton, E. M. R., Lupu, R., Owusu-Asare, A., et al. (Apr. 2017). "Exo-Transmit: An Open-Source Code for Calculating Transmission Spectra for Exoplanet Atmospheres of Varied Composition". In: *Publications of the Astronomical Society of the Pacific* 129.974, p. 044402. arXiv: 1611.03871 [astro-ph.EP] (cit. on p. 288).

Kepler Mission Team (Nov. 2009). "VizieR Online Data Catalog: Kepler Input Catalog (Kepler Mission Team, 2009)". In: *VizieR Online Data Catalog*, V/133, pp. V/133 (cit. on p. 126).

Kiang, N. Y., Domagal-Goldman, S., Parenteau, M. N., et al. (2018). "Exoplanet biosignatures: At the dawn of a new era of planetary observations". In: *Astrobiology* 18.6, pp. 619–629 (cit. on pp. 92, 93, 96).

Kim, S., Lee, C., Park, B., et al. (Feb. 2016). "KMTNET: A Network of 1.6 m Wide-Field Optical Telescopes Installed at Three Southern Observatories". In: *Journal of Korean Astronomical Society* 49.1, pp. 37–44 (cit. on p. 23).

Kjeldsen, H. and Bedding, T. R. (Jan. 1995). "Amplitudes of stellar oscillations: the implications for asteroseismology." In: *A&A* 293, pp. 87–106. eprint: astro-ph/9403015 (cit. on p. 10).

Kopparapu, R. K. (2018). “The Habitable Zone: The Climatic Limits of Habitability”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–13 (cit. on p. 85).

Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. (2013). “Habitable Zones Around Main-Sequence Stars: New Estimates”. In: *The Astrophysical Journal* 765.1993, p. 16 (cit. on pp. 86, 149, 151, 187–189, 219, 242, 247).

Krasnopolsky, V. A., Maillard, J. P., and Owen, T. C. (Dec. 2004). “Detection of methane in the martian atmosphere: evidence for life?” In: *Icarus* 172, pp. 537–547 (cit. on p. 95).

Kraus, A. L., Ireland, M. J., Cieza, L. A., et al. (Dec. 2013). “Three Wide Planetary-Mass Companions to FW Tau, ROXs 12, and ROXs 42B”. In: *The Astrophysical Journal* 781.1. arXiv: 1311.7664, p. 20 (cit. on p. 135).

La Rocca, N., Claudi, R., Erculiani, M. S., et al. (Sept. 2015). “Microorganisms suitable for studying biomarkers within the atmosphere in a test tube project”. In: *5th Workshop of the Italian Astrobiology Society, Life in a Cosmic Context* (cit. on p. 280).

Lacis, A. and Oinas, V. (1991). “A Description of the Correlated k Distribution Method for Modeling Nongray Gaseous Absorption , Thermal Emission , and Multiple Scattering in Vertically Inhomogeneous Atmospheres to N2). Comparison cooling”. In: *Journal of Geophysical Research* 96, pp. 9027–9063 (cit. on pp. 160–162).

Lagrange, A.M., Keppler, M., Beust, H., et al. (Dec. 2017). “Discovery of a stellar companion to HD 131399A”. en. In: *Astronomy & Astrophysics* 608, p. L9 (cit. on p. 134).

Lammer, H., Sproß, L., Grenfell, J. L., et al. (July 2019). “The Role of N_2 as a Geo-Biosignature for the Detection and Characterization of Earth-like Habitats”. In: *Astrobiology* 19.7, pp. 927–950. arXiv: 1904.11716 [astro-ph.EP] (cit. on p. 95).

Leconte, J., Forget, F., Charnay, B., et al. (Dec. 2013). “Increased insolation threshold for runaway greenhouse processes on Earth-like planets”. In: *Nature* 504.7479, pp. 268–271. arXiv: 1312.3337 [astro-ph.EP] (cit. on p. 86).

Luhman, K. L. (Sept. 2012). “The Formation and Early Evolution of Low-Mass Stars and Brown Dwarfs”. In: *Annual Review of Astronomy and Astrophysics* 50, pp. 65–106. arXiv: 1208.5800 [astro-ph.GA] (cit. on p. 135).

Macintosh, B., Graham, J. R., Ingraham, P., et al. (2014). “First light of the Gemini Planet Imager”. In: *Proceedings of the National Academy of Sciences* 111.35, pp. 12661–12666 (cit. on p. 21).

Madhusudhan, N. (2018). “Atmospheric Retrieval of Exoplanets”. en. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–30 (cit. on p. 19).

Madhusudhan, N., Knutson, H., Fortney, J., et al. (2014). “Exoplanetary Atmospheres”. In: *Protostars and Planets VI*, pp. 739–762 (cit. on pp. 6, 7, 16).

Magalhães, J. A., Schofield, J. T., and Seiff, A. (1999). “Results of the Mars Pathfinder atmospheric structure investigation”. In: *Journal of Geophysical Research: Planets* 104.E4, pp. 8943–8955. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1998JE900041 (cit. on p. 182).

Mahadevan, S., Ramsey, L., Bender, C., et al. (Sept. 2012). “The habitable-zone planet finder: a stabilized fiber-fed NIR spectrograph for the Hobby-Eberly Telescope”. In: *SPIE Proceedings*. Vol. 8446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 84461S. arXiv: 1209.1686 [astro-ph.EP] (cit. on p. 11).

Malbet, F. and Sozzetti, A. (2018). “Astrometry as an Exoplanet Discovery Method”. en. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–16 (cit. on p. 24).

Manabe, S. and Strickler, R. F. (1964). “Thermal Equilibrium of the Atmosphere with a Convective Adjustment”. In: *Journal of the Atmospheric Sciences* 21.4, pp. 361–385 (cit. on p. 174).

Manabe, S. and Wetherald, R. T. (1967). *Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity* (cit. on pp. 169, 174).

Marconi, A., Di Marcantonio, P., D’Odorico, V., et al. (Aug. 2016). “EELT-HIRES the high-resolution spectrograph for the E-ELT”. In: *Ground-based and Airborne Instrumentation for Astronomy VI*. Vol. 9908. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 990823. arXiv: 1609.00497 [astro-ph.IM] (cit. on p. 11).

Martin, R. G. and Livio, M. (Mar. 2016). “on the Formation of Super-Earths With Implications for the Solar System”. In: *The Astrophysical Journal* 822.2, p. 90 (cit. on p. 36).

Mayor, M., Pepe, F., Queloz, D., et al. (Dec. 2003). “Setting New Standards with HARPS”. In: *The Messenger* 114, pp. 20–24 (cit. on p. 11).

Mayor, M. and Queloz, D. (1995). “A Jupiter-mass companion to a solar type star”. In: *Nature* 378.1993, pp. 355–359 (cit. on pp. 6, 25).

Meadows, V. S. and Barnes, R. K. (2018). “Factors Affecting Exoplanet Habitability”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–24 (cit. on pp. 86–88).

Miller-Ricci, E., Seager, S., and Sasselov, D. (2009). “The Atmospheric Signatures of Super-Earths: how to distinguish between Hydrogen-rich and Hydrogen-poor atmospheres”. In: *The Astrophysical Journal* 690.2, pp. 1056–1067 (cit. on pp. 31, 32).

Mischna, M., Lee, C., and Richardson, M. I. (2012). “Development of a fast, accurate radiative transfer model for the Martian atmosphere, past and present”. In: *Journal of Geophysical Research* 117.E10, pp. 1–22 (cit. on pp. 163, 178, 179).

Montmessin, F. and Määttänen, A. (2018). “Temperature, Clouds, and Aerosols in the Terrestrial Bodies of the Solar System”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–29 (cit. on pp. 78, 89).

Nielsen, E. L., De Rosa, R. J., Rameau, J., et al. (Nov. 2017). “Evidence that the Directly-Imaged Planet HD 131399 Ab is a Background Star”. In: *The Astronomical Journal* 154.6. arXiv: 1705.06851, p. 218 (cit. on p. 134).

Nier, A. O., Hanson, W. B., Seiff, A., et al. (1976). "Composition and Structure of the Martian Atmosphere: Preliminary Results from Viking 1". In: *Science* 193.4255, pp. 786–788. eprint: https://science.sciencemag.org/content/193/4255/786.full.pdf (cit. on p. 182).

North, G. R. and Erukhimova, T. L. (2009). *Atmospheric Thermodynamics*. Cambridge (cit. on pp. 58, 61, 68, 69).

Ochsenbein, F., Bauer, P., and Marcout, J. (Apr. 2000). "The VizieR database of astronomical catalogues". In: *Astronomy and Astrophysics Supplement Series* 143, pp. 23–32. arXiv: astro-ph/0002122 [astro-ph] (cit. on p. 126).

Oliva, E., Origlia, L., Maiolino, R., et al. (Oct. 2012). "The GIANO spectrometer: towards its first light at the TNG". In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Ed. by Ian S. McLean, Suzanne K. Ramsay, and Hideki Takami. SPIE (cit. on p. 11).

Osuna, P., Ortiz, I., Lusted, J., et al. (Oct. 2008). *IVOA Astronomical Data Query Language Version 2.00*. Tech. rep., p. 1030 (cit. on p. 118).

Parès, L., Donati, J.-F., Dupieux, M., et al. (Sept. 2012). "Front end of the SPIRou spectropolarimeter for Canada-France Hawaii Telescope". In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Vol. 8446. SPIE Proceedings, 84462E (cit. on p. 11).

Pepe, F. A., Cristiani, S., Rebolo Lopez, R., et al. (July 2010). "ESPRESSO: the Echelle spectrograph for rocky exoplanets and stable spectroscopic observations". In: *Ground-based and Airborne Instrumentation for Astronomy III*. Vol. 7735. Proceedings of the SPIE, 77350F (cit. on p. 11).

Pepper, J., Pogge, R. W., DePoy, D. L., et al. (Aug. 2007). "The Kilodegree Extremely Little Telescope (KELT): A Small Robotic Telescope for Large-Area Synoptic Surveys". In: *Publications of the Astronomical Society of the Pacific* 119.858, pp. 923–935. arXiv: 0704.0460 [astro-ph] (cit. on p. 20).

Petralia, A., Alei, E., Aresu, G., et al. (2019). submitted (cit. on pp. 157, 160).

Pierrehumbert, R. and Gaidos, E. (June 2011). "Hydrogen Greenhouse Planets Beyond the Habitable Zone". In: *Astrophysical Journal Letters* 734.1, L13, p. L13. arXiv: 1105.0021 [astro-ph.EP] (cit. on p. 86).

Pilcher, C. B. (Nov. 2003). "Biosignatures of Early Earths". In: *Astrobiology* 3, pp. 471–486 (cit. on p. 96).

Pino, L., Ehrenreich, D., Wyttenbach, A., et al. (Apr. 2018). "Combining low- to high-resolution transit spectroscopy of HD 189733b. Linking the troposphere and the thermosphere of a hot gas giant". In: *Astronomy & Astrophysics* 612, A53, A53. arXiv: 1709.09678 [astro-ph.EP] (cit. on p. 288).

Plante, R., Williams, R., Hanisch, R., et al. (Feb. 2008). *Simple Cone Search Version 1.03*. Tech. rep., p. 222 (cit. on p. 123).

Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. (Oct. 2006). "The WASP Project and the SuperWASP Cameras". In: *Publications of the Astronomical Society of the Pacific* 118, pp. 1407–1418. eprint: astro-ph/0608454 (cit. on p. 20).

Pravdo, S. H. and Shaklan, S. B. (Jan. 2003). “Stellar Planet Survey-STEPS”. In: *Scientific Frontiers in Research on Extrasolar Planets*. Ed. by Drake Deming and Sara Seager. Vol. 294. Astronomical Society of the Pacific Conference Series, pp. 107–110 (cit. on p. 24).

Price-Whelan, A. M., Sipócz, B. M., Günther, H. M., et al. (Sept. 2018). “The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package”. In: *ApJ* 156, 123, p. 123 (cit. on p. 118).

Quirrenbach, A., Amado, P. J., Caballero, J. A., et al. (July 2014). “CARMENES instrument overview”. In: *Ground-based and Airborne Instrumentation for Astronomy V*. Vol. 9147. SPIE Proceedings, 91471F (cit. on p. 11).

Ramirez, R. M. and Kaltenegger, L. (Mar. 2017). “A Volcanic Hydrogen Habitable Zone”. In: *Astrophysical Journal Letters* 837.1, L4, p. L4. arXiv: 1702.08618 [astro-ph.EP] (cit. on p. 86).

Raschka, S. (2015). *Python Machine Learning*. Packt Publishing (cit. on p. 202).

Raskin, G. and Winckel, H. (Jan. 2014). “HERMES at Mercator, competitive high-resolution spectroscopy with a small telescope”. In: *Astronomische Nachrichten* 335 (cit. on p. 11).

Rasool, S. I. and de Bergh, C. (June 1970). “The Runaway Greenhouse and the Accumulation of CO_2 in the Venus Atmosphere”. In: *Nature* 226.5250, pp. 1037–1039 (cit. on p. 86).

Rauer, H., Aerts, C., Cabrera, J., et al. (Sept. 2016). “The PLATO Mission”. In: *Astronomische Nachrichten* 337, p. 961 (cit. on p. 20).

Rauer, H., Gebauer, S., Paris, P. V., et al. (2011). “Astronomy & Astrophysics Potential biosignatures in super-Earth atmospheres I. Spectral appearance of super-Earths around M dwarfs”. In: *A&A* 529 (cit. on p. 95).

Rein, H. (Nov. 2012). “A proposal for community driven and decentralized astronomical databases and the Open Exoplanet Catalogue”. In: *arXiv e-prints*, arXiv:1211.7121, arXiv:1211.7121. arXiv: 1211.7121 [astro-ph.EP] (cit. on pp. 103, 111).

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. (Jan. 2015). “Transiting Exoplanet Survey Satellite (TESS)”. In: *Journal of Astronomical Telescopes, Instruments, and Systems* 1, 014003, p. 014003 (cit. on p. 20).

Rodet, L., Beust, H., Bonnefoy, M., et al. (May 2017). “On the origin of the wide-orbit circumbinary giant planet HD 106906: A dynamical scenario and its impact on the disk”. In: *Astronomy & Astrophysics* 602. arXiv: 1703.01857, A12 (cit. on p. 134).

Rothman, L. S., Gordon, I. E., Barber, R. J., et al. (Oct. 2010). “HITEMP, the high-temperature molecular spectroscopic database”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 111, pp. 2139–2150 (cit. on p. 159).

Rugheimer, S., Kaltenegger, L., and Sasselov, D. (Jan. 2015). “Spectral Fingerprints of Earth-like Planets Orbiting Other Stars”. In: *American Astronomical Society Meeting Abstracts #225*. Vol. 225. American Astronomical Society Meeting Abstracts, p. 124.02 (cit. on pp. 95, 96).

Rugheimer, S., Kaltenegger, L., Zsom, A., et al. (Mar. 2013). “Spectral Fingerprints of Earth-like Planets Around FGK Stars”. In: *Astrobiology* 13, pp. 251–269. arXiv: 1212.2638 [astro-ph.EP] (cit. on p. 95).

Rybicki, G. B. and Lightman, A. P. (May 1981). *Radiative Processes in Astrophysics*. Ed. by George B. Rybicki and Alan P. Lightman. Vol. 25. 4. Weinheim, Germany: Wiley-VCH Verlag GmbH, p. 432 (cit. on pp. 16, 46, 158).

Salasnich, B., Claudi, R., Alei, E., et al. (2018). “Control software for the Multi-Channel Led starlight simulator”. In: *Proc. SPIE*. Vol. 10707 (cit. on pp. 279, 280).

Sánchez-Lavega, A., Pérez-Hoyos, S., and Hueso, R. (2004). “Clouds in planetary atmospheres: A useful application of the Clausius–Clapeyron equation”. In: *Am. J. Phys.* 72.6, p. 767 (cit. on pp. 76, 77).

Santerne, A. (2018). “Populations of Extrasolar Giant Planets from Transit and Radial Velocity Surveys”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–23 (cit. on p. 26).

Santerne, A., Hébrard, G., Deleuil, M., et al. (Nov. 2014). “SOPHIE velocimetry of Kepler transit candidates - XII. KOI-1257 b: a highly eccentric three-month period transiting exoplanet”. en. In: *Astronomy & Astrophysics* 571, A37 (cit. on p. 134).

Schneider, J. (2018). “Definition of Exoplanets and Brown Dwarfs”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–6 (cit. on p. 136).

Schneider, J., Dedieu, C., Le Sidaner, P., et al. (2011). “Defining and cataloging exoplanets: the exoplanet.eu database”. In: *Astronomy & Astrophysics* 532, A79 (cit. on pp. 103, 107, 110).

Schwieterman, E. W. (2018). “Surface and Temporal Biosignatures”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–29 (cit. on pp. 93, 97–101).

Schwieterman, E. W., Kiang, N. Y., Parenteau, M. N., et al. (June 2018). “Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life”. In: *Astrobiology* 18.6, pp. 663–708. arXiv: 1705.05791 [astro-ph.EP] (cit. on pp. 93, 96).

Seager, S. (2010). *Exoplanet Atmospheres: Physical Processes*. Princeton University Press, p. 243 (cit. on pp. 52, 54, 58, 62, 67, 68, 70, 75, 78).

Seager, S., Bains, W., and Petkowski, J. (2016). “Toward a List of Molecules as Potential Biosignature Gases for the Search for Life on Exoplanets and Applications to Terrestrial Biochemistry.” In: *Astrobiology* 16.6, pp. 1–21 (cit. on p. 42).

Seager, S. and Deming, D. (2010). “Exoplanet atmospheres”. In: *Annual Review of Earth and Planetary Sciences* 48, pp. 631–72 (cit. on p. 12).

Seager, S., Kuchner, M. J., Hier-Majumder, C. A., et al. (2007). “Mass-Radius Relationships for Solid Exoplanets”. In: *The Astrophysical Journal* 669, p. 1279 (cit. on pp. 40, 145, 147, 185).

Segura, A., Kasting, J. F., Meadows, V., et al. (2005). “Biosignatures from Earth-Like Planets Around M Dwarfs”. In: *Astrobiology* 5.6, pp. 706–725 (cit. on pp. 95, 96).

Seinfeld, J. and Pandis, S. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Wiley (cit. on p. 74).

Simon, M., Ghez, A. M., Leinert, C., et al. (Apr. 1995). “A lunar occultation and direct imaging survey of multiplicity in the Ophiuchus and Taurus star-forming regions”. In: *The Astrophysical Journal* 443, pp. 625–637 (cit. on p. 135).

Sing, D. K. (2018). “Observational Techniques With Transiting Exoplanetary Atmospheres 1 Background and History of Exoplanet Atmosphere Observations”. In: arXiv: 1804.07357v1 (cit. on p. 16).

Sing, D. K., Fortney, J., Nikolov, N., et al. (Jan. 2016). “A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion”. In: *Nature* 529.7584. arXiv: 1512.04341, pp. 59–62 (cit. on p. 17).

Southworth, J. (2011). “Homogeneous studies of transiting extrasolar planets - IV. Thirty systems with space-based light curves”. In: *Monthly Notices of the Royal Astronomical Society* 417.3, pp. 2166–2196 (cit. on p. 287).

Spergel, D., Gehrels, N., Baltay, C., et al. (Mar. 2015). “Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report”. In: *arXiv e-prints*, arXiv:1503.03757, arXiv:1503.03757. arXiv: 1503.03757 [astro-ph.IM] (cit. on p. 23).

Spiegel, D. S., Raymond, S. N., Dressing, C. D., et al. (Oct. 2010). “Generalized Milankovitch Cycles and Long-Term Climatic Habitability”. In: *The Astrophysical Journal* 721.2, pp. 1308–1318. arXiv: 1002.4877 [astro-ph.EP] (cit. on p. 86).

Stam, D. M., Hovenier, J. W., and Waters, L. B. F. M. (2004). *Using polarimetry to detect and characterize Jupiter-like extrasolar planets* (cit. on p. 50).

Stamnes, K., Tsay, S. C., Wiscombe, W., et al. (June 1988). “Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media.” In: *Applied optics* 27.12, pp. 2502–2509 (cit. on p. 172).

Strassmeier, K. G., Ilyin, I., Järvinen, A., et al. (May 2015). “PEPSI: The high-resolution échelle spectrograph and polarimeter for the Large Binocular Telescope”. In: *Astronomische Nachrichten* 336, p. 324. arXiv: 1505.06492 [astro-ph.IM] (cit. on p. 11).

Szentgyorgyi, A. H. and Furész, G. (June 2007). “Precision Radial Velocities for the Kepler Era”. In: *Revista Mexicana de Astronomia y Astrofisica Conference Series*. Ed. by S. Kurtz. Vol. 28. Revista Mexicana de Astronomia y Astrofisica Conference Series, pp. 129–133 (cit. on p. 11).

Tackley, P. J., Ammann, M. M., Brodholt, J. P., et al. (2012). “Habitable Planets: Interior Dynamics and Long-Term Evolution”. In: *Proceedings of the International Astronomical Union* 8.293, pp. 339–349 (cit. on p. 41).

Taylor, M. B. (Dec. 2005). “TOPCAT & STIL: Starlink Table/VOTable Processing Software”. In: *Astronomical Data Analysis Software and Systems XIV*. Ed. by P. Shopbell, M. Britton, and R. Ebert. Vol. 347. Astronomical Society of the Pacific Conference Series, p. 29 (cit. on p. 138).

Tennyson, J., Yurchenko, S., Al-Refaie, A., et al. (2016). “The ExoMol database: Molecular line lists for exoplanet and other hot atmospheres”. In: *Journal of Molecular Spectroscopy* 327. New Visions of Spectroscopic Databases, Volume II, pp. 73–94 (cit. on p. 159).

Tinetti, G., Drossart, P., Eccleston, P., et al. (July 2016). “The science of ARIEL (Atmospheric Remote-sensing Infrared Exoplanet Large-survey)”. In: *SPIE Proceedings*. Vol. 9904. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 99041X (cit. on p. 20).

Tinetti, G., Encrenaz, T., and Coustenis, A. (2013). “Spectroscopy of planetary atmospheres in our Galaxy”. In: *Astron Astrophys Rev* 21 (cit. on pp. 6, 7, 11).

Toon, O. B., McKay, C. P., Ackerman, T. P., et al. (Nov. 1989). “Rapid calculation of radiative heating rates and photodissociation rates in inhomogeneous multiple scattering atmospheres”. In: *Journal of Geophysical Research* 94.89, pp. 287–301 (cit. on p. 178).

Traub, W. and Oppenheimer, B. R. (2010). “Direct Imaging of Exoplanets”. In: *Exoplanets*, pp. 111–156 (cit. on pp. 21, 50).

Tsapras, Y. (Sept. 2018). “Microlensing Searches for Exoplanets”. en. In: *Geosciences* 8.10. arXiv: 1810.02691, p. 365 (cit. on pp. 22, 23).

Tsiaras, A., Waldmann, I. P., Tinetti, G., et al. (Sept. 2019). “Water vapour in the atmosphere of the habitable-zone eight Earth-mass planet K2-18 b”. In: *arXiv e-prints*, arXiv:1909.05218, arXiv:1909.05218. arXiv: 1909.05218 [astro-ph.EP] (cit. on p. 33).

Udalski, A., Szymanski, M., Stanek, K. Z., et al. (Apr. 1994). “The Optical Gravitational Lensing Experiment. The Optical Depth to Gravitational Microlensing in the Direction of the Galactic Bulge”. In: *Acta Astronomica* 44, pp. 165–189. arXiv: astro-ph/9407014 [astro-ph] (cit. on p. 23).

Valencia, D., Sasselov, D., and O’Connell (2007). “Radius and Structure Models of the First Super-Earth Planet”. In: *The Astrophysical Journal* 656, p. 545 (cit. on p. 38).

Valencia, D., Sasselov, D., and O’Connell, R. J. (2007). “Detailed Models of Super-Earths: How Well Can We Infer Bulk Properties?” In: *The Astrophysical Journal* 665.2, pp. 1413–1420 (cit. on pp. 38, 39).

Vasquez, M., Pallé, E., and Montanes Rodrigues, P. (2010). *The Earth as a Distant Planet* (cit. on pp. 38, 40, 43, 44, 80–84, 92).

Venot, O. and Agúndez, M. (2015). “Chemical modeling of exoplanet atmospheres”. In: *Experimental Astronomy* 40.2-3, pp. 469–480 (cit. on p. 156).

Viard, T., Mathieu, J. -C., Fer, Y., et al. (Nov. 2006). “The CoRoT Telescope”. In: *The CoRoT Mission Pre-Launch Status - Stellar Seismology and Planet Finding*. Ed. by M. Fridlund, A. Baglin, J. Lochard, et al. Vol. 1306. ESA Special Publication, p. 187 (cit. on p. 20).

Vladilo, G., Murante, G., Silva, L., et al. (Apr. 2013). “The Habitable Zone of Earth-like Planets with Different Levels of Atmospheric Pressure”. In: *The Astrophysical Journal* 767.1, 65, p. 65. arXiv: 1302.4566 [astro-ph.EP] (cit. on p. 86).

Vogt, S. S. and Donald Penrod, G. (1988). “HIRES: A High Resolution Echelle Spectrometer for the Keck 10-Meter Telescope”. In: *Instrumentation for Ground-Based Optical Astronomy*. Ed. by Lloyd B. Robinson. New York, NY: Springer New York, pp. 68–103 (cit. on p. 11).

Vogt, S. S., Radovan, M., Kibrick, R., et al. (Apr. 2014). “APF—The Lick Observatory Automated Planet Finder”. In: *PASP* 126.938, p. 359. arXiv: 1402.6684 [astro-ph.IM] (cit. on p. 11).

Wagner, K., Apai, D., Kasper, M., et al. (Aug. 2016). “Direct Imaging Discovery of a Jovian Exoplanet Within a Triple Star System”. In: *Science* 353.6300. arXiv: 1607.02525, pp. 673–678 (cit. on p. 134).

Wakelam, V., Herbst, E., Loison, J.C., et al. (2012). “A KInetic Database for Astrochemistry (KIDA)”. In: *Physics* (cit. on p. 156).

Wayne, R. P. (2000). *Chemistry of atmospheres : an introduction to the chemistry of the atmospheres of earth, the planets, and their satellites*. Oxford University Press, p. 775 (cit. on pp. 63, 74).

Wenger, M., Ochsenbein, F., Egret, D., et al. (Apr. 2000). “The SIMBAD astronomical database. The CDS reference database for astronomical objects”. In: *A&As* 143, pp. 9–22. eprint: astro-ph/0002110 (cit. on p. 123).

Wildi, F., Blind, N., Reshetov, V., et al. (Sept. 2017). “NIRPS: an adaptive-optics assisted radial velocity spectrograph to chase exoplanets around M-stars”. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 10400. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 1040018 (cit. on p. 11).

Winn, J. N. (2018). “Planet Occurrence: Doppler and Transit Surveys”. In: *Handbook of Exoplanets*. Ed. by Hans J. Deeg and Juan Antonio Belmonte. Cham: Springer International Publishing, pp. 1–18 (cit. on p. 26).

Winn, J. N. and Fabrycky, D. C. (2015). “The Occurrence and Architecture of Exoplanetary Systems”. In: *Annual Review of Astronomy and Astrophysics* 53.1, pp. 409–447 (cit. on p. 28).

Wolszczan, A. and Frail, D. A. (Jan. 1992). “A planetary system around the millisecond pulsar PSR1257 + 12”. In: *Nature* 355, pp. 145–147 (cit. on pp. 24, 120).

Wright, J. T. (2018). “Radial Velocities as an Exoplanet Discovery Method”. In: *arXiv:1707.07983 [astro-ph]*. arXiv: 1707.07983, pp. 619–631 (cit. on p. 10).

Wright, J. T., Fakhouri, O., Marcy, G. W., et al. (2011). “THE EXOPLANET ORBIT DATABASE”. In: *Astronomy & Astrophysics* (cit. on pp. 103, 109).

Zsom, A., Seager, S., de Wit, J., et al. (Dec. 2013). “Toward the Minimum Inner Edge Distance of the Habitable Zone”. In: *The Astrophysical Journal* 778.2, 109, p. 109. arXiv: 1304.3714 [astro-ph.EP] (cit. on p. 86).

Webpages

Jaadi, Zakaria (2005). *A step by step explanation of Principal Component Analysis*. URL: https://towardsdatascience.com/a-step-by-step-explanation-of-principal-component-analysis-b836fb9c97e2 (cit. on p. 202).

ToolBox, Engineering (2005a). *Carbon Dioxide - Specific Heat*. URL: https://www.engineeringtoolbox.com/carbon-dioxide-d_974.html.

– (2005b). *Nitrogen - Specific Heat*. URL: https://www.engineeringtoolbox.com/nitrogen-d_977.html.

– (2005c). *Oxygen Gas - Specific Heat*. URL: https://www.engineeringtoolbox.com/oxygen-d_978.html.

– (2005d). *Water Vapor - Specific Heat*. URL: https://www.engineeringtoolbox.com/water-vapor-d_979.html.