

Exoplanet Spectroscopy: A review of the detection, modeling, and analysis of exoplanet spectra

Ary Cheng

September 2022

Abstract

One of the most successful methods to characterize an exoplanet is through the use of high-resolution spectroscopy ($R > 25,000$). Light from the exoplanet is collected and analyzed to determine properties about the planet, including the presence of certain molecules in its atmosphere. However, there are many factors that can modify the spectrum of light that is detected by our telescopes, and the resulting spectrum often has noticeable differences compared to an unaltered spectrum. In this paper, a review of the most prominent factors that can affect the observed spectrum is presented, such as the planet’s radial velocity (RV), rotation, and temperature. The current methods that are used to detect and analyze exoplanet spectra are also discussed. Finally, a Python code that can both computationally implement these effects to generate a model spectrum, and analyze a given spectrum to determine these parameters is developed.

I Introduction

As of March 2022, more than 5000 extrasolar planets (exoplanets) have been discovered, with the closest one to Earth being Proxima Centauri b at a distance of 4.24 light-years [1]. However, it is difficult to characterize these planets in detail due to their distance from Earth, and most of the obtainable information comes only from the light we can receive from their atmosphere or surface [2]. Even so, it is possible to obtain information about the planet’s formation and evolution, its properties such as rotation and temperature, and even the possibility of extraterrestrial life through the existence of biosignatures such as oxygen (O_2) or ozone (O_3), among many others. This information is collected as a spectrum, which consists of a measurement of flux or intensity as a function of the wavelength or frequency of light.

In this paper, the current status of the field of exoplanet science is briefly reviewed, and an first-order model for current methods used by researchers in this field is presented. The paper is structured as follows: A brief overview of the methods used to detect and characterize exoplanets based on measurements we can obtain is presented in Section I. A first-order model for generating model spectra is presented in Section II, and the specific techniques used to compare observed spectra with model spectra is introduced in Section III. Sample results from the models presented here are available in the Appendix.

A Exoplanet Detection Methods

There are four main methods for exoplanet detection today: radial velocity, gravitational microlensing, transits, and direct imaging [3, 4]. A brief overview of the principles and challenges for each of these methods is given in this section (I-A).

A.1 Radial Velocity (RV)

As a planet orbits its host star, its gravitational pull will cause the star to “wobble” back and forth. This will induce a periodic Doppler shift on the starlight, from which the orbital period of the exoplanet can be determined. If the mass of the star is known, then the mass of the planet can be determined within a factor of 2 roughly 87% of the time [3]. Around 20% of all known exoplanets have been discovered using this method [4].

However, this method primarily targets larger (more massive) exoplanets, as the gravitational pull

between the star and planet would be stronger, making the blue/redshift in the starlight more noticeable. Additionally, planets with a shorter orbital period would produce results that can be detected across a shorter amount of time. By contrast, it is significantly more difficult to identify smaller planets orbiting at a further distance from their host star, and stellar activity may sometimes produce noise that can be misinterpreted as a planet as well [3].

A.2 Gravitational Microlensing

According to Einstein’s general theory of relativity, massive objects (such as a star or planet) will bend light traveling near them due to their gravity. This effect is known as *gravitational lensing*, and is usually observed only when the light source, lens, and observer are approximately collinear. Thus when a host star acts as a lens for a background source, it bends the light as detected by an aligned observer. If the star has a companion planet, it will have a similar though much smaller effect and can thus be detected as a variation over the star’s lensing. Since the planet’s lensing effect is much weaker compared to its host star, this is called gravitational *microlensing*. While this method is less prone to false positives [3], lensing effects are not periodic and therefore cannot be verified with other means of detection [4].

This method is similar to the radial velocity method in that both methods indirectly infer the existence of an exoplanet. That is, the effects of an exoplanet on our measurements are seen instead of seeing the planet itself. While some characteristics, such as the mass of the planet, can be inferred, it often provides little information about other characteristics, such as the properties of its atmosphere and by extension, its habitability.

A.3 Transits

When a planet passes between its host star and the observer, some of the starlight is blocked. This is known as a transit, and is observed as a dip in the amount of light received from the host star. Unlike the methods introduced above, the planet can be directly observed. Additionally, information about the planet’s radius, inclination of the plane of the orbit, and the semi-major axis can be determined as well [4].

This method comes with its own set of limitations. The host star, planet, and observer must be approximately collinear, and the probability of detecting a planet in transit depends on the planetary orbit as well as the stellar radius. For an Earth-sized planet in an Earth-like orbit around solar-type stars, the probability that the planet *can* be observed in transit is merely 0.5% [3]. Planetary transits can often be difficult to distinguish from starspots as well.

A.4 Direct Imaging

As the name suggests, this method detects exoplanets by directly recording the light originating from them. This method makes it much easier to characterize the planet’s composition and atmosphere through high-resolution spectroscopy, and is the most relevant method to the model presented here. However, this is by far the most technically challenging way to detect and characterize an exoplanet, as a high angular resolution is needed to distinguish a planet and its host star from a very far distance. Additionally, light from a host star often drowns out any signal from a planet.

In particular, the flux of the brightest planet in our solar system would be roughly one billion times smaller than the flux of the Sun as seen from 30 light-years away [3]. It would also be extremely difficult to single out the light coming from a planet, as residual starlight will be detected along with the planet’s own light. Consequently, technology such as coronagraphs are often employed to block out only the star’s light while leaving the planet’s light relatively untouched. While this allows for increased sensitivity towards Earth-like planets in the habitable zone [5], the method generally works better for young self-luminous planets orbiting at a further distance from their host star.

A.5 Overview

The methods summarized above are, for the most part, complementary; that is, each method is effective for only a specific subset of exoplanets around a host star. For instance, the RV method and transits can be used to discover more massive planets orbiting closer to their host star, while direct

imaging and gravitational microlensing are more effective for planets that orbit further away from the star. Additionally, multiple methods can be used independently to verify the existence of an exoplanet [4].

Out of the four main methods introduced above, only transits and direct imaging reveal information about the planet’s atmospheric composition and other properties, since light passing through the planet’s atmosphere is directly observed and can be collected using a spectrograph. In simple terms, a spectrograph separates the incoming light by its wavelength and records the intensity of each individual wavelength of light. The ability of a spectrograph to distinguish between light of different wavelengths is measured by the resolving power R , which is defined as follows:

$$R = \frac{\lambda}{\Delta\lambda} \quad (1)$$

where $\Delta\lambda$ represents the smallest difference in wavelengths that can be distinguished at a wavelength of λ . This quantity ($\Delta\lambda$) is also known as the spectral resolution of the spectrograph. The best instruments to date can achieve a resolving power of $R > 100,000$ in the near-mid infrared region [6].

B Types of Spectra

There are two types of spectra: emission and absorption. The general principle behind both are the same; atoms and molecules have specific energy states, and a specific amount of energy is associated with each transition between energy states. An emission spectrum arises when a sample is heated, and the particles emit light at specific frequencies to return to a lower energy state. The spectrum usually consists of bright “lines” at these frequencies and is dark elsewhere. Absorption spectra arise from the passage of light through a sample, where the particles absorb light at specific frequencies to reach a higher energy state. The spectrum usually consists of dark “lines” at these frequencies. The specific frequencies that these spectral features occur at is unique to each compound, and thus can be used to characterize the sample. For the purposes of this paper, we will be focusing on absorption spectra. This is because in direct-imaging spectroscopy, the light observed passes through and is modulated by the planet’s atmosphere, and is observed as an absorption spectrum.

C Spectrum Modeling

One of the most crucial aspects of exoplanet spectroscopy is the atmospheric modeling that allows the data to be interpreted [7]. In brief terms, measurements by instruments such as those on the James Webb Space Telescope (JWST) are compared against models with generated spectra to confirm the existence of certain compounds within an exoplanet’s atmosphere. There are many high-fidelity models for this purpose [8, 9, 10], but differences in the databases they use along with the different effects that are taken into account can often result in differences exceeding the experimental uncertainties of measurements [7]. Nonetheless, the models have been successfully used to detect properties of multiple exoplanets. For example, three exoplanets orbiting the star HR 8799 were found to have both carbon monoxide (CO) and water (H₂O) present in their atmosphere [11, 12, 13].

II The Forward Model

The most accurate way to model molecular absorption spectra is a *line-by-line* model, where the overall spectrum is given by the product of all individual absorption lines [14]. However, this process is not as straight-forward as it sounds as there are many factors that can affect how the final absorption lines are observed, and accurate model spectra need to make adjustments to account for these effects. One such model is the LinePak™ library, which is used by NASA and other research institutions worldwide [15]. The model presented here will only take into account the most prominent effects and should not be interpreted as a high-fidelity model, but rather a first-order approximation for educational purposes.

A Blackbody Radiation

Any object above the temperature of absolute zero will emit continuous electromagnetic radiation known as *blackbody radiation*. The spectral radiance of an object at temperature T at wavelength λ is

given by Planck’s radiation law:

$$B = \frac{2c^2h}{(e^{kT\lambda} - 1) \cdot \lambda^5} \quad (2)$$

As this blackbody radiation passes through an exoplanet’s atmosphere, it is absorbed at specific wavelengths by the different compounds present in the atmosphere. Thus the spectrum that is observed by our instruments will consist of “dips” in a background curve described by Planck’s radiation law.

B Radial Velocity (RV)

Light is an electromagnetic wave, and as such, is subject to the Doppler effect when the source is moving relative to the observer. This is known as a *radial velocity (RV)* shift, where RV represents the relative velocity of the source along the line of sight of the observer. For non-relativistic values of RV ($RV \ll c$), a light wave emitted by the source at wavelength λ will be shifted to a new wavelength λ' given by:

$$\lambda' = \lambda \cdot \left(1 + \frac{RV}{c}\right) \quad (3)$$

For relativistic values of RV, it is necessary to introduce the Lorentz factor $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$, where v represents RV in this case. This yields the following equation with a bit of algebra:

$$\lambda' = \lambda \cdot \left(1 + \frac{RV}{c}\right) \cdot \gamma = \lambda \cdot \sqrt{\frac{1 + RV/c}{1 - RV/c}} \quad (4)$$

By convention, RV is chosen to be positive when the source is moving away relative to the observer.

C Line Width

One of the most fundamental principles of quantum mechanics is that energy is quantized, and the energy of a photon with wavelength λ is described by:

$$E = \frac{hc}{\lambda} \quad (5)$$

As a result, we may expect each absorption line to have effectively zero width, since each energy transition has a single, definite value, and therefore it must correspond to a single, definite wavelength. However, this is not the case.

C.1 Natural Broadening

While each absorption line does indeed correspond to a single energy transition, the *value* of this energy transition is inherently uncertain as a result of the Heisenberg uncertainty principle:

$$\Delta E \Delta t \geq \frac{h}{4\pi} \quad (6)$$

where Δt represents the lifetime of a decaying state, and ΔE represents the uncertainty in its energy [16]. In other words, a quantum state with a short lifetime, such as an excited electron, cannot have a definite energy [17, 18]. As a result, all absorption lines will have a definite width due to this inherent uncertainty in its corresponding energy transition. However, the magnitude of this effect is negligible compared to the other broadening factors discussed in this section. A typical atomic energy state has a lifetime on the order of 10^{-8} seconds, corresponding to an energy uncertainty on the order of 10^{-27} Joules and a line width on the order of 10^{-5} nm for the near-mid infrared region.

C.2 Collision (Pressure) Broadening

A significant source of line broadening comes from molecular collisions, which can disturb the molecules’ energy states and create a wide range of energy transitions, which is then observed as a broadened

absorption line. This broadening is described by a Lorentzian distribution [17, 19]. For a Lorentzian distribution function P around a central wavelength λ_0 :

$$P(\lambda) = \frac{\alpha}{(\lambda - \lambda_0)^2 + \alpha^2} \quad (7)$$

where α represents the half-width at half-maximum (HWHM) of the distribution. This parameter is impossible to obtain without experimental measurements, as it varies from compound to compound. Furthermore, parameters like temperature and pressure determine the frequency of molecular collisions, which affects the broadening parameter as well. In practice, models such as the Spectral Calculator [14] have to make adjustments to experimental results at initial conditions P_0 and T_0 when modeling spectra at a different set of conditions [17]:

$$\alpha = \alpha_0 \cdot \frac{P}{P_0} \left(\frac{T_0}{T} \right)^n \quad (8)$$

where n is the temperature dependence exponent, and also varies depending on the identities of the colliding molecules. It can also only be obtained through experimental measurements. For planets with atmospheric conditions similar to Earth, typical line widths due to pressure broadening are on the order 10^{-1} nm to 10^{-2} nm for the near-mid infrared region.

C.3 Doppler Broadening

Another source of broadening comes from the random thermal motion of gas molecules along the observer's line of sight [19]. Each molecule effectively has its own RV shift, and the cumulative effect across all molecules is a broadened line following a Gaussian (normal) distribution [20]. Note that this effect is intrinsic to any system with a temperature above absolute zero. By contrast, a true RV shift is observer-dependent and doesn't represent intrinsic properties of the planet.

For an ideal gas, where all collisions between gas molecules are assumed to be perfectly elastic, the velocity distribution of the molecules follows a Maxwell-Boltzmann distribution [21]. The full-width at half-maximum (FWHM) and standard deviation σ of the resulting Gaussian distribution around a central wavelength of λ_0 can then be shown to be:

$$\text{FWHM} = 2\lambda_0 \sqrt{2 \ln 2} \frac{kT}{mc^2} \quad (9)$$

$$\sigma = \frac{\text{FWHM}}{2\sqrt{2 \ln 2}} = \lambda_0 \sqrt{\frac{kT}{mc^2}} \quad (10)$$

A derivation of these results by Prof. Rod Nave of Georgia State University can be found online [22]. Typical line widths are on the order 10^{-2} nm to 10^{-3} nm for the near-mid infrared region at a temperature of 300 K.

C.4 Rotational Broadening

Similar to Doppler broadening, rotational broadening also comes from the cumulative effect of RV shifts from individual gas molecules. When an exoplanet is rotating in the observer's line of sight, half of the atmosphere will be moving towards the observer, while the other half will be moving away. This also results in a Gaussian distribution. For most celestial objects, we can only infer the *projected* rotational velocity as seen by an observer, defined as $v \cdot \sin(i)$, where v is the rotational velocity and i is the inclination angle between the rotational axis and the observer's line of sight [23]. The resulting FWHM and standard deviation σ of the Gaussian distribution around a central wavelength of λ_0 are [24]:

$$\text{FWHM} = 2\lambda \sqrt{2 \ln 2} \cdot \frac{v \sin(i)}{c} \quad (11)$$

$$\sigma = \frac{\text{FWHM}}{2\sqrt{2 \ln 2}} = \lambda \cdot \frac{v \sin(i)}{c} \quad (12)$$

C.5 Line Shape

The final shape of an absorption line is a combination of all of the above effects. In particular, the convolution of Gaussian distributions due to Doppler broadening and rotational broadening with a Lorentzian distribution due to pressure broadening will give a distribution function known as a Voigt profile. The Voigt profile has no simple analytic form, although recent studies have shown that it can be expressed simply with the use of the Humbert function [25].

D Zeeman Splitting

When placed in an external magnetic field, the energy of an atom will change because of its magnetic moment. For a singlet state (all electrons are paired), the net spin of all electrons will be 0, so the angular momentum of the atom comes solely from the orbital angular momentum. An electron with magnetic quantum number m_l will have its energy changed by:

$$\Delta E = m_l \mu_B B \quad (13)$$

where B represents the strength of the magnetic field and μ_B is the Bohr magneton. The selection rules for an electron transition ($\Delta m_l = 0, \pm 1$) ensures that each absorption line will be split into 3 absorption lines of equal intensity, with energies varying by $\pm \mu_B B$. This is called the *normal Zeeman effect*.

For non-singlet states (not all electrons are paired), both electron spin and orbital angular momentum will contribute to the total angular momentum of the atom. This results in an absorption line being split into many more lines and is called the *anomalous Zeeman effect*. This behavior is much more difficult to predict, so it will not be discussed in-depth here. Additionally, Zeeman splitting in exoplanet spectroscopy is almost entirely negligible, as most planets are unable to produce strong magnetic fields that can actually “split” an absorption line into 3 separate lines. Most of the time, the Zeeman effect is simply observed as a slightly widened absorption line. In fact, astronomers often use the Hanle effect to detect the presence of weak magnetic fields in exoplanets instead of the Zeeman effect [26]. In brief terms, the Hanle effect concerns the polarization of light due to a magnetic field and is much more sensitive to the presence of a magnetic field than the Zeeman effect [27, 28].

E Other Effects

As described in the previous section, the presence of a magnetic field can change the magnitude of an energy transition. The presence of an electric field can also cause a similar change in the molecular energy levels, and this effect is known as Stark broadening [29]. However, this effect is much more prominent for stellar spectra, since stars are made of ionized particles in the form of plasma, while exoplanet atmospheres have very few charged particles in comparison. For this reason, it will be completely ignored in the model presented here as well.

Another effect that can potentially modulate an observed spectrum is limb darkening. In brief terms, limb darkening describes a decrease in the light received from the edges of an object. However, this effect is usually second-order and only noticeable in stellar spectra [30]. For direct imaging, since the light observed originates not from the star but from the planet itself, this effect is negligible and is therefore ignored in the model.

The intensity of an absorption line also depends on the pressure and temperature, but these effects are entirely negligible as well [14, 15]. For an absorption line in the near-mid infrared region, this effect will usually cause the wavelength to shift by an amount on the order of 10^{-3} nm. In addition to the ones described in this section, there are many other subtle factors that can modulate a spectrum, but they will all be ignored as their effects are unnoticeable by our instruments today.

F A Simple Model

A link to the simple model developed here can be found [here](#). As stated before, this model is far from being accurate, utilizing many approximations and neglecting many factors in the interests of computation time and simplicity. It should only serve as a conceptual introduction.

All of the data about absorption lines is taken from the HITRAN database [31]. Sample spectra generated by this model along with a comparison to other models can be found in Appendix A.

F.1 Modeling the shape of absorption lines

One of the most challenging aspects of generating a spectrum comes from determining the shape of each absorption line. As mentioned before, each absorption line follows a Voigt profile, which has no analytic form. Furthermore, the pressure broadening parameters can only be determined experimentally. Recall the equation for adjusting pressure broadening to fit different conditions:

$$\alpha = \alpha_0 \cdot \frac{P}{P_0} \left(\frac{T_0}{T} \right)^n$$

This is only part of the picture. A molecule colliding with another molecule of the same compound will affect the spectrum differently than if it collided with a molecule of a different compound. A more realistic model looks something like this [14, 15]:

$$\alpha = (1 - q)\alpha_a^0 \cdot \frac{P}{P_0} \left(\frac{T_0}{T} \right)^{\gamma_a} + q\alpha_s^0 \cdot \frac{P}{P_0} \left(\frac{T_0}{T} \right)^{\gamma_s} \quad (14)$$

Where α represents the HWHM of the Lorentzian distribution, α_a^0 and α_s^0 represent the air- and self-broadened HWHM respectively at conditions P_0 and T_0 , γ_a and γ_s represent the air- and self-broadening temperature dependence respectively, and q represents the relative abundance of the compound in question. All of these parameters¹² are available in HITRAN.

The *voigt_profile* function in the Python package *scipy.special* is then used to compute a Voigt profile given the standard deviation of the Gaussian distribution and the HWHM of the Lorentzian distribution. Here, a step to simplify computation is made. Since the total broadening function is the convolution of each individual broadening effect, the Gaussian distributions resulting from rotational and Doppler broadening are convolved first, resulting in another Gaussian distribution, which is then convolved with the Lorentzian distribution resulting from pressure broadening. Note that this is not the sequence in which the actual broadening is applied; in reality, rotational broadening only arises because the measurement is taken outside of the exoplanet’s atmosphere, so the resulting Gaussian distribution should be convolved with the Voigt profile due to Doppler and pressure broadening instead. However, since convolution is a commutative operation, the order in which it is applied does not affect the final result.

F.2 Line Intensity

The depth of each absorption line is a combination of the line’s intensity and the compound’s relative abundance. All values for line intensity from HITRAN are already adjusted to account for the relative abundances of the isotopes that create those absorption lines [31]. These values are then normalized such that the largest value is 1, then multiplied by the relative abundance of the compound, which is inputted by the user. Also recall that the normal Zeeman effect will split each absorption line into 3, so the code will divide the intensity of each line by 3 while adding 2 new lines to model the splitting.

F.3 Implementation

Recall that an absorption spectrum consists of a background blackbody curve with “dips” to represent the absorption lines. The Voigt profile for an absorption line is first normalized such that the peak value is 1. It is then scaled by its relative intensity, subtracted from 1, and multiplied to the blackbody curve. This process is repeated for each absorption line.

Finally, all wavelength values will be changed according to equation (4) to account for an RV shift between the planet and the observer. This is to ensure that the code will continue to function accurately even for relativistic values of RV.

¹ γ_s is not listed for certain compounds. In these cases, it is assumed to be equal to γ_a .

²Since q is not a property of the compound, it won’t be included either. It must be inputted manually by the user.

G Limitations

Perhaps one of the most striking issues with the approach in this simple model is the way that it models the exoplanet’s atmosphere. In particular, a single value for pressure and temperature are used, effectively collapsing all of the atmospheric conditions into a single set of conditions. This is called a 0-dimensional (0-D) model, as the atmosphere is modeled as a single point. While this makes computation extremely fast and efficient, it is certainly not a realistic representation. By contrast, a theoretically perfect model would have to be 3-D to accurately simulate the full complexity and spatial diversity of a planet’s atmosphere, and the best models to date are 2-D models developed as recently as 2020 [32, 33].

Additionally, this model uses air-broadening parameters to calculate pressure broadening, which only works for exoplanets with Earth-like atmospheres and completely breaks down for a different atmosphere like that of Venus [17]. Potential noise from detectors or Earth’s atmosphere are also entirely ignored.

III The Reverse Process

As described in Section I-C, the spectra generated by high-fidelity models must be compared to the observed spectra to recover information about the planet’s properties and the chemical composition of its atmosphere. However, unlike the forward process, the recovery of these parameters cannot be done sequentially, as multiple parameters can have similar effects on the spectrum and it is often impossible to distinguish one from another. Instead, a technique known as cross-correlation is used [12, 13].

A Cross-correlation

The cross-correlation function (CCF) is commonly used in signal processing as a way to measure the time difference between two correlated signals. The cross-correlation function $c(\tau)$ between signals $f(t)$ and $g(t)$ is defined as follows:

$$c(\tau) = \int_{-\infty}^{\infty} f(t)g(t + \tau) dt \quad (15)$$

where τ represents the time delay, or *lag*, between the two signals. The CCF $c(\tau)$ has a maximum, or “peak”, when τ is precisely equal to the time delay between the two signals. In general, the CCF will have a higher value when the two signals have a higher similarity to each other.

In exoplanet spectroscopy, a CCF is used to find the RV shift between the observed and model spectra. The detection strength is directly related to the peak value of the CCF expressed as a function of RV [34], so the value of RV that returns the highest CCF peak is likely to be the true RV shift of the observed spectrum. With this method, the planet’s RV can be known up to an accuracy on the order of 1 km/s [35].

For other parameters such as the planet’s temperature and rotation, a grid of model spectra with different values for these parameters is used due to the high computational inefficiency of generating new model spectra each time [13]. Essentially, the grid is a database of pre-generated model spectra that can then be compared against the observed spectrum. The set of parameters corresponding to the model that is most similar to the observed spectrum (highest CCF peak) is then most likely to correspond to the observed spectrum as well.

A.1 Limitations

However, the disadvantages of this method are quite clear as well; it is limited by the number of parameters that can be varied, as well as the sampling size of the grid. For instance, if the real values for these parameters don’t land on the grid of model spectra, then it will be impossible to perfectly recover the values to arbitrarily high degrees of precision. As a result, some recovered parameters have relatively high error; for the aforementioned HR 8799 star system, the rotational speeds of some planets had error bars up to 25% [12].

Another fundamental limitation of this method is its inability to distinguish between different sets

of parameters that have roughly the same effects on the spectrum. For instance, both a higher pressure and a faster rotational speed will cause absorption lines to be broadened. In these cases, it is up to physical analysis to decide what is most reasonable. For instance, an upper bound on the rotational speed of a planet can be set at its break-up velocity, and reasonable values for atmospheric pressure and temperature can be determined from other characteristics such as planetary mass and radius. However, this is something that cannot be done by software and requires human intuition.

B Implementation

As mentioned in Section III-A, most methods to recover planetary information from an observed spectrum involve the use of a grid of pre-generated model spectra to save computation time. The model presented here will instead generate new spectra every time due to a lack of these grids. Additionally, the Pearson product-moment correlation coefficient (PPMC) instead of the full CCF will be used in the interests of computational efficiency. The PPMC is essentially just the value of the normalized CCF at one specific value for the time lag τ [36, 37], and it will be implemented using the *corrcoef* function in the Python package *numpy*. The correlation coefficient r between data sets x and y is defined as follows:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (16)$$

where \bar{x} and \bar{y} represent the mean values of data sets x and y respectively. The value of r ranges from -1 to 1, with 1 representing perfect correlation while -1 represents perfect anti-correlation. Due to the implementation of RV in the forward model, a CCF will also not be needed to find the RV shift. Instead, it will be calculated directly from the amount that the wavelength vector was stretched by according to Eq. (4).

There will also be many steps taken to increase computational efficiency. The model will essentially be narrowing down the range of values for each parameter by first finding a broad region with the highest correlation coefficients, then a smaller region within this broad region with high correlation coefficients, and so on. This significantly reduces runtime by only devoting computation to the range of parameters that match most closely with the observed spectral features.

Some sample outputs from this model can be found in Appendix B. The code is available here.

IV Conclusion

In this paper, a quick overview of the state of the field of exoplanet science is presented, including detection and spectral modeling techniques. A model to demonstrate the most prominent effects that are accounted for by today's spectrum modeling libraries is also provided. In particular, effects due to the rotation, temperature, RV, magnetic field, and pressure of the planet's atmosphere are modeled, while other factors are ignored due to their negligible effect on the spectrum. A brief explanation and demonstration of the techniques used to recover parameters from an observed spectra is also presented. The rest of this section will be used to briefly mention recent advancements in the field.

Spectroscopic analysis on exoplanets is inherently extremely difficult; we are measuring light coming from hundreds of light-years away that is likely to be significantly modulated by both atmospheric and planetary properties. Additionally, direct imaging spectroscopy involves singling out the planet's spectrum next to its host star, which can be up to one billion times brighter. A popular analogy to describe the incredible measurements that our instruments can procure is that it is akin to measuring the light of a firefly on a lighthouse from across the country. Yet it is possible to definitively characterize some of these planets' properties through this miniscule amount of light, and infer the existence of certain molecules within their atmospheres.

There have also been a number of recent advancements in this field. Most notably, the launch of the James Webb Space Telescope (JWST) in 2021 allowed direct imaging spectroscopy at wavelengths up to 28 microns, which was previously impossible with ground-based instruments or other telescopes such as the Hubble Space Telescope. In fact, NASA's Jet Propulsion Laboratory announced the first

exoplanet directly imaged by the JWST’s instruments in September 2022 [38]. Additionally, as mentioned in Section II-G, the best spectrum modeling libraries are now capable of simulating exoplanet atmospheres in 2-D, one step closer to a fully accurate 3-D description [33].

V Acknowledgements

I would like to thank Dan Echeverri for his guidance and support throughout this challenging project, from helping me find a suitable spectral database to debugging my code for spectral analysis.

VI Appendices

A Generating Spectra

The spectra generated for multiple different compounds are shown here, and are compared with spectra described in previous literature when available.

A.1 Carbon Monoxide (CO)

One of the defining features of the CO absorption spectrum in the infrared region is the first overtone band, a collection of deep absorption lines beginning at a wavelength of around 2300 nm. A series of spectra at different pressures are generated and shown in Figs. (1) – (3). Values for the other parameters are as follows: $T = 303$ K, $B = 30 \mu\text{T}$, $v \sin(i) = 0$ km/s, $RV = 0$ km/s. These values are chosen to simulate Earth’s atmospheric conditions as closely as possible as seen by an observer on Earth. Additionally, the reference spectrum [39] (Fig. (4)) is measured in terms of wavenumber, which is defined as the reciprocal of wavelength. In the interests of an easier comparison, the generated spectrum will also be in terms of wavenumber. The conversion between wavelength λ measured in nm and wavenumber $\tilde{\nu}$ measured in cm^{-1} is given by:

$$\tilde{\nu} = \frac{10^7}{\lambda} \quad (17)$$

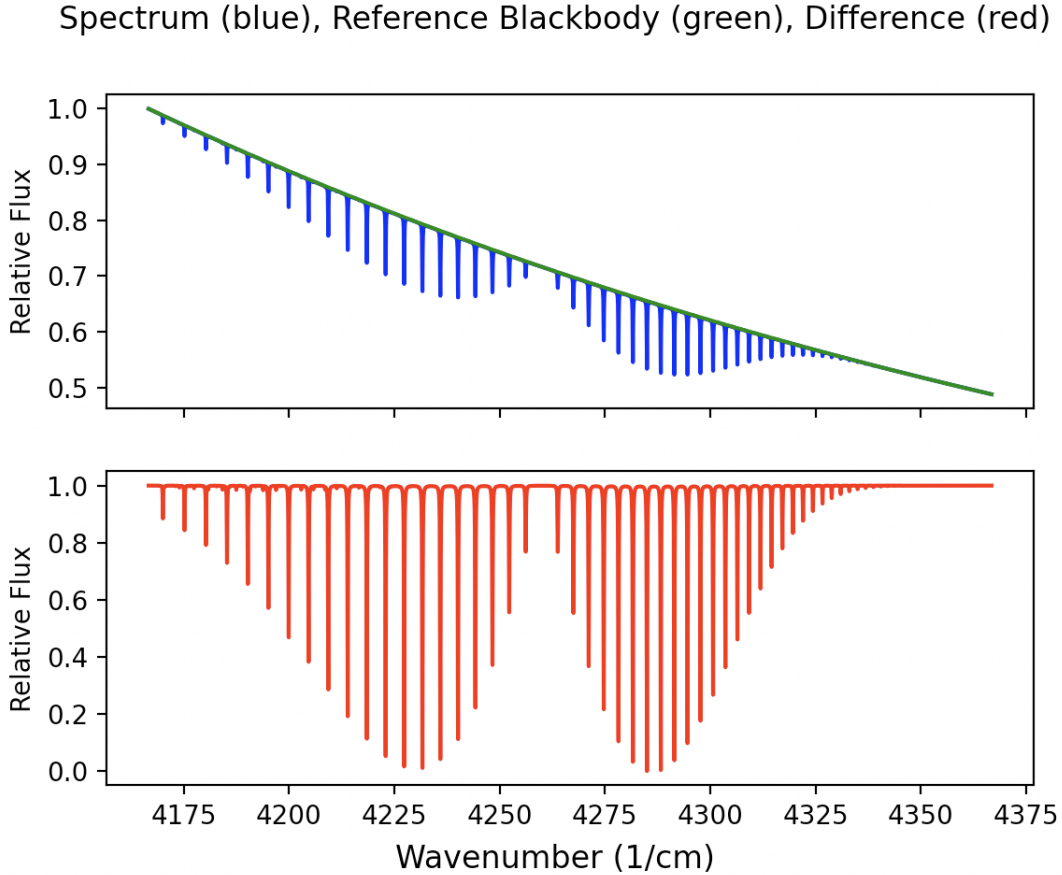


Figure 1: Calculated CO spectrum from 4170 cm^{-1} to 4350 cm^{-1} . Data for absorption lines taken from Ref. (31). $T = 303$ K, $B = 30 \mu\text{T}$, $v \sin(i) = 0$ km/s, $RV = 0$ km/s, $P = 1$ atm. The blue line represents the spectrum on a background blackbody curve, the green line represents the background blackbody curve with no spectral features, and the red line represents the difference between the two.

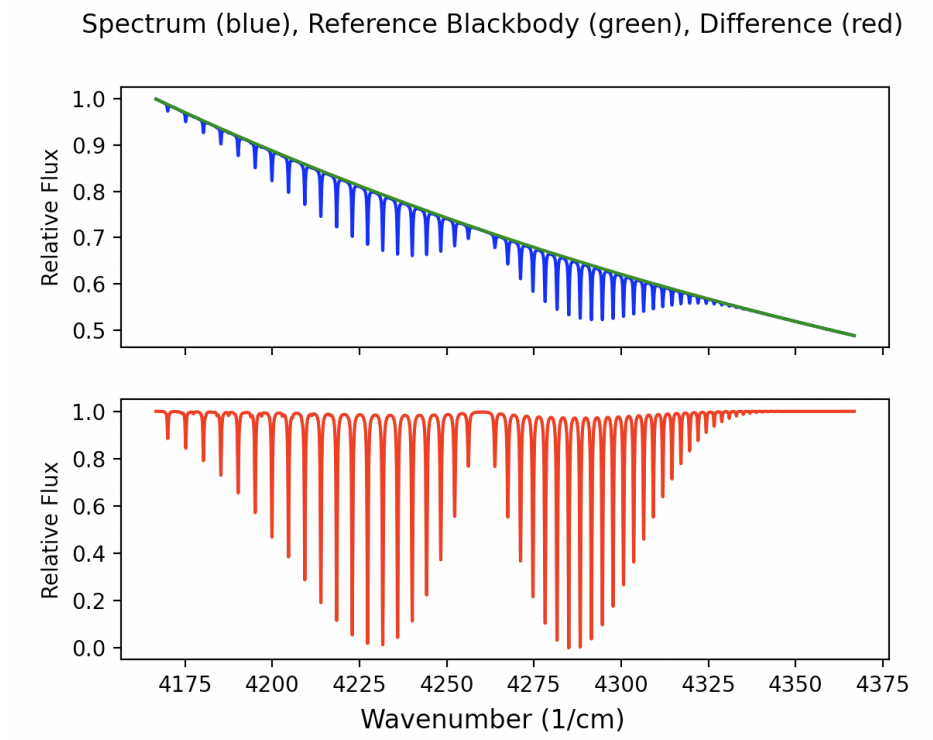


Figure 2: Calculated CO spectrum from 4170 cm^{-1} to 4350 cm^{-1} . Data for absorption lines taken from Ref. (31). $T = 303 \text{ K}$, $B = 30 \text{ } \mu\text{T}$, $v \sin(i) = 0 \text{ km/s}$, $RV = 0 \text{ km/s}$, $P = 3 \text{ atm}$. The blue line represents the spectrum on a background blackbody curve, the green line represents the background blackbody curve with no spectral features, and the red line represents the difference between the two.

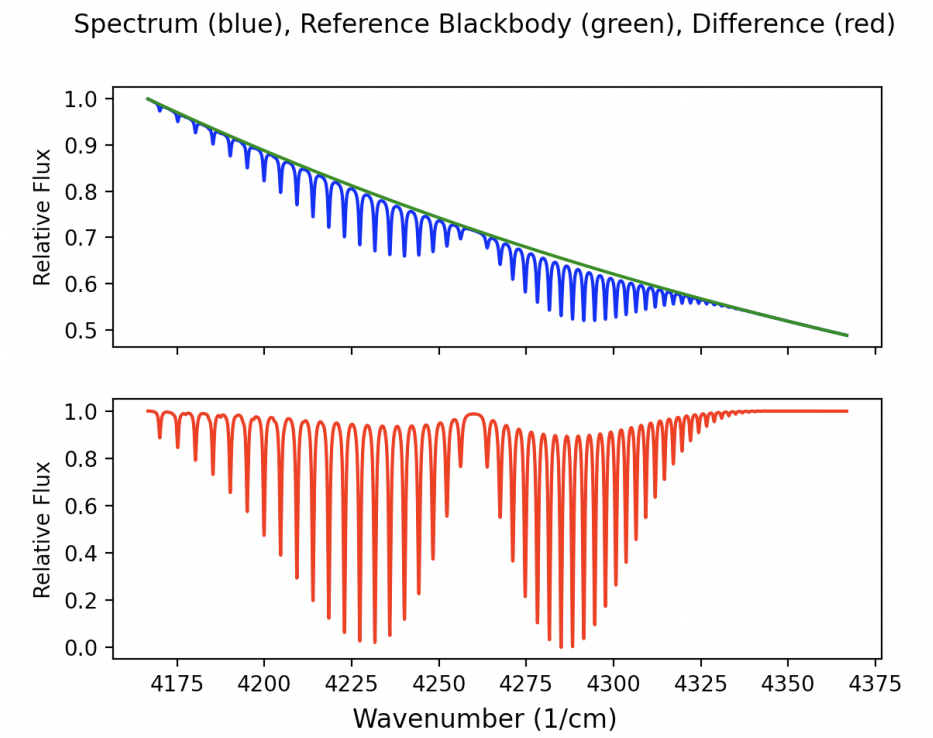


Figure 3: Calculated CO spectrum from 4170 cm^{-1} to 4350 cm^{-1} . Data for absorption lines taken from Ref. (31). $T = 303 \text{ K}$, $B = 30 \text{ } \mu\text{T}$, $v \sin(i) = 0 \text{ km/s}$, $RV = 0 \text{ km/s}$, $P = 6 \text{ atm}$. The blue line represents the spectrum on a background blackbody curve, the green line represents the background blackbody curve with no spectral features, and the red line represents the difference between the two.

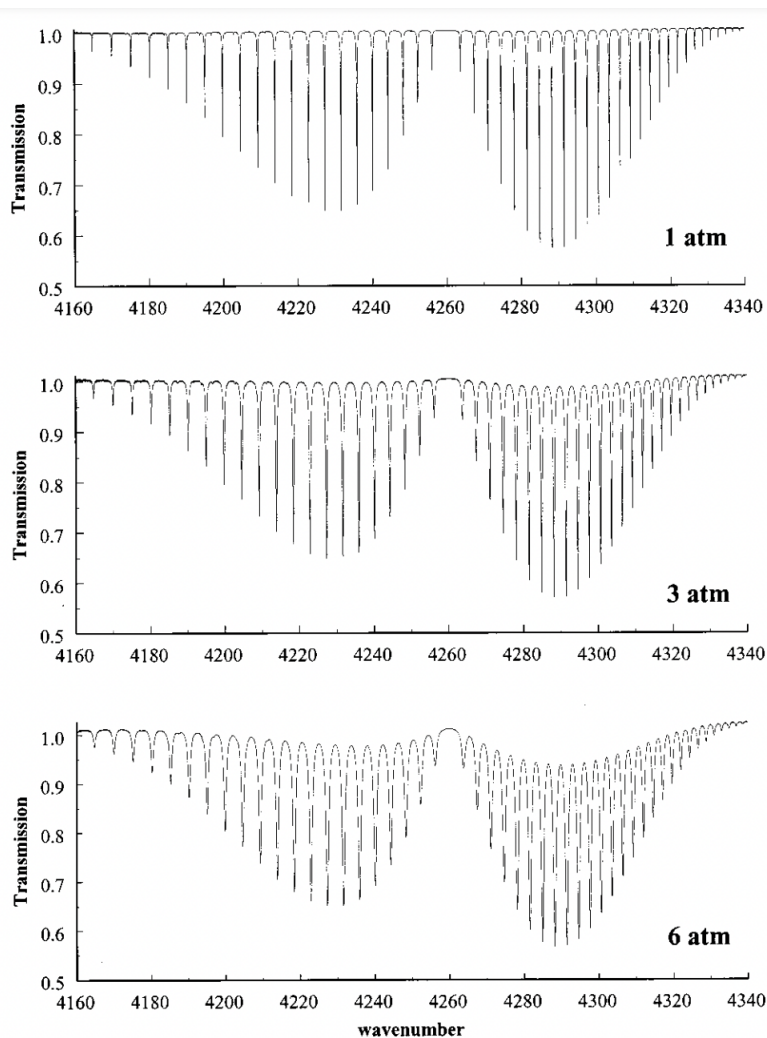


Figure 4: CO spectrum from 4170 cm^{-1} to 4350 cm^{-1} for different values of pressure. Source: Predoi-Crossa et al., *The Journal of Chemical Physics* **113**, 158 (2000). (Ref. (39))

Recall Eq. (14), which shows that a higher pressure would lead to broader absorption lines. This is observed as lines with wider “bases” that all seem to merge together. The spectra generated are consistent with this result, and a qualitative analysis shows that they are consistent with the literature [39] as well. The only noticeable difference is in the y-axis, but this is simply due to the way that the generated data sets were normalized.

A.2 Other Compounds

Some other generated spectra for various compounds are shown in Figs. (5) – (8). The values chosen for the various parameters are completely arbitrary. Again, a reference for what these spectra should look like is shown in Fig. (9) [5].

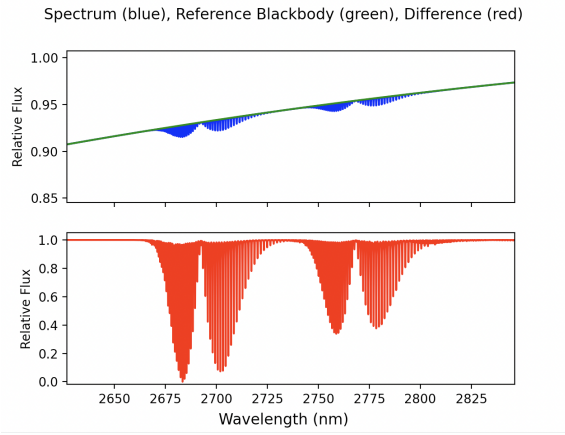


Figure 5: Calculated spectrum of CO_2 from 2650 nm to 2850 nm. Data for absorption lines taken from Ref. (31). $T = 900 \text{ K}$, $B = 50 \mu\text{T}$, $v \sin(i) = 5.5 \text{ km/s}$, $\text{RV} = 40 \text{ km/s}$, $P = 1.2 \text{ atm}$.

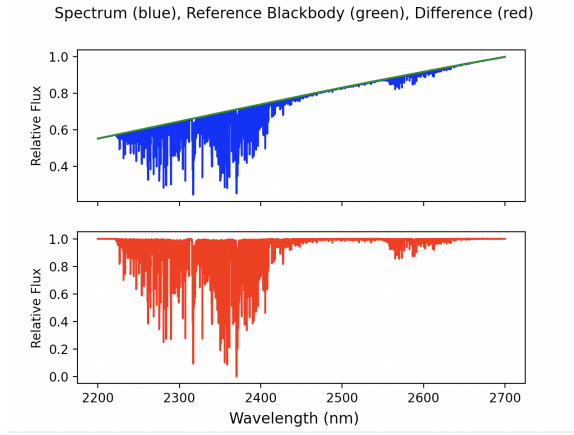


Figure 6: Calculated spectrum of CH_4 from 2200 nm to 2700 nm. Data for absorption lines taken from Ref. (31). $T = 750 \text{ K}$, $B = 1 \text{ mT}$, $v \sin(i) = 3.4 \text{ km/s}$, $\text{RV} = -20 \text{ km/s}$, $P = 0.6 \text{ atm}$.

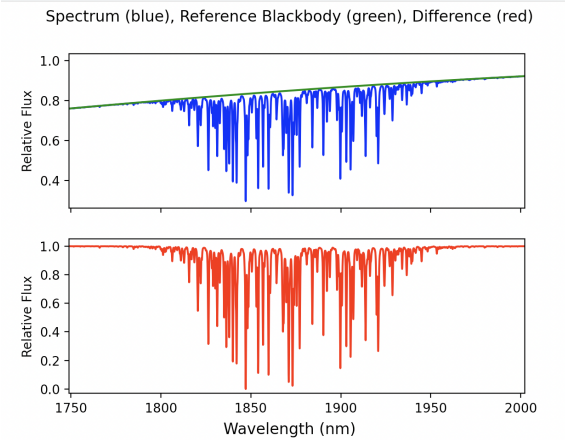


Figure 7: Calculated spectrum of H_2O from 1750 nm to 2000 nm. Data for absorption lines taken from Ref. (31). $T = 1200 \text{ K}$, $B = 250 \mu\text{T}$, $v \sin(i) = 20 \text{ km/s}$, $\text{RV} = 5 \text{ km/s}$, $P = 2.5 \text{ atm}$.

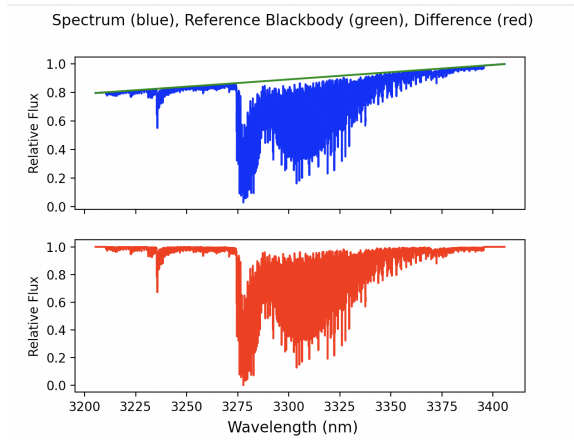


Figure 8: Calculated spectrum of O_3 from 3200 nm to 3400 nm. Data for absorption lines taken from Ref. (31). $T = 500 \text{ K}$, $B = 5.0 \text{ mT}$, $v \sin(i) = 1 \text{ km/s}$, $\text{RV} = 500 \text{ km/s}$, $P = 0.3 \text{ atm}$.

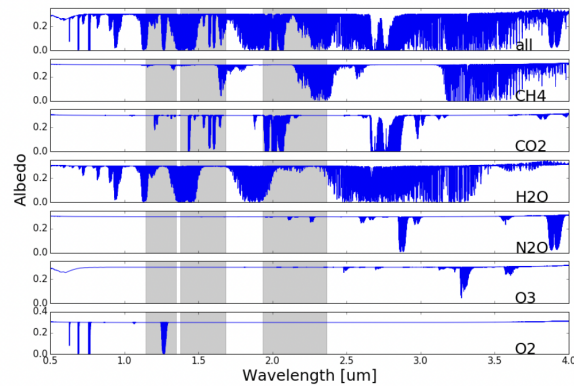


Figure 9: Reference near-mid infrared spectrum for a variety of compounds. Source: Wang et al., *The Astronomical Journal* **153**, 183 (2017). (Ref. (5))

B Cross-correlation and Parameter Recovery

For this section, a spectrum will be generated with certain arbitrary parameters. This serves as the “observed” spectrum, and the values for the parameters recovered using the cross-correlation method are compared against the input values for the “observed” spectrum. For the following trials, the CO spectrum from 2200 nm to 2400 nm are used.

B.1 On the Grid

As described in Section III-A, cross-correlation usually relies on a pre-generated “grid” of spectra to compare against the observed spectrum. The input values of the observed spectrum for this trial are chosen to lie precisely on the grid. The results are as follows:

	Input Value	Test Range	Output Value
RV	40 km/s	N/A	40.0 km/s
$v \sin(i)$	4.1 km/s	1 km/s - 6 km/s	4.10 km/s
Temperature	490 K	100 K - 600 K	490 K
B-field	0.012 T	0.01 T - 0.02 T	0.012 T
Pressure	2.1 atm	0.5 atm - 2.5 atm	2.1 atm

Table 1: Trial with all input parameters (input value) lying on the grid of generated spectra. The range of values tested (test range) and the recovered parameter (output value) are given as well. Note that the RV is not recovered using the cross-correlation method, so it does not have a range of values to be tested.

As expected, the parameters were recovered perfectly. The CCF between the observed spectrum and the spectrum generated using all of the recovered values is given below:

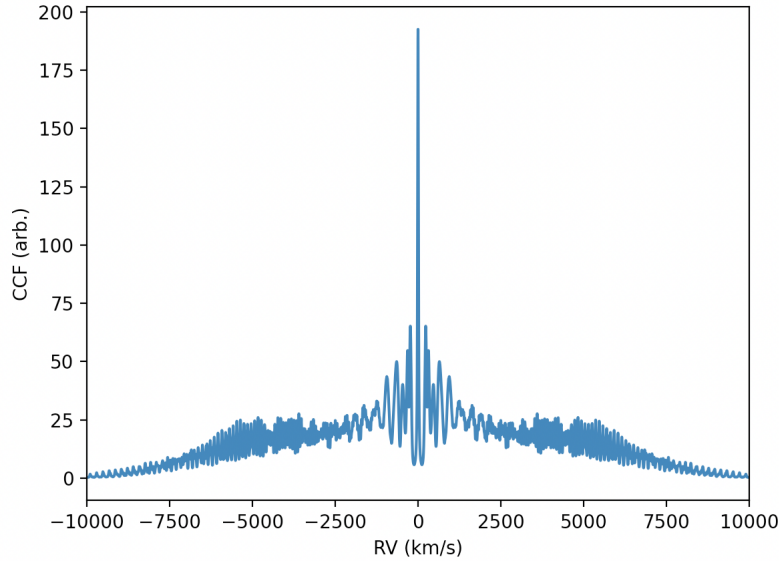


Figure 10: CCF between the observed spectrum and generated spectrum. The extremely high peak at $RV = 0$ indicates a very strong correlation between the two spectra.

B.2 Off the Grid

The input values of the observed spectrum for this trial are chosen to not lie on the grid, as is often the case with real measurements. The results are as follows:

	Input Value	Test Range	Output Value
RV	40 km/s	N/A	40.0 km/s
$v \sin(i)$	3.46 km/s	1 km/s - 6 km/s	3.4 km/s
Temperature	485.2 K	100 K - 600 K	484 K
B-field	0.0146 T	0.01 T - 0.02 T	0.02 T
Pressure	1.92 atm	0.5 atm - 2.5 atm	1.94 atm

Table 2: Trial with no input parameters (input value) lying on the grid of generated spectra. The range of values tested (test range) and the recovered parameter (output value) are given as well. Note that the RV is not recovered using the cross-correlation method, so it does not have a range of values to be tested.

The parameters were all recovered to within $\pm 1\%$ of the actual value, except for the value of the magnetic field. However, as described in Section II-D, the Zeeman effect is very weak compared to other effects, so the inaccurate value is likely an artifact of trying to compensate for slight inaccuracies with the values of the other parameters. Again, the CCF between the observed spectrum and the spectrum generated using the recovered parameters is shown:

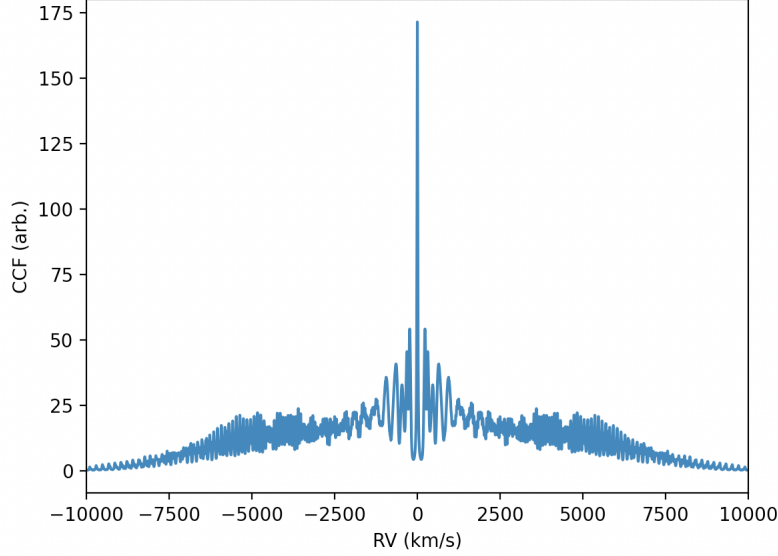


Figure 11: CCF between the observed spectrum and generated spectrum. The extremely high peak at $RV = 0$ indicates a very strong correlation between the two spectra.

References

- ¹G. Anglada-Escudé et al., “A terrestrial planet candidate in a temperate orbit around Proxima Centauri”, *Nature* **536**, 437–440 (2016).
- ²J. Kasting et al., “Exoplanet Characterization and the Search for Life”, *Astro2010: The Astronomy and Astrophysics Decadal Survey*, *Science White Papers* **2010**, 151 (2009).
- ³A. D. Fischer et al., “Exoplanet Detection Techniques”, *Protostars and Planets* **6**, 715–737 (2014).
- ⁴Z. Dai et al., “Five Methods of Exoplanet Detection”, *Journal of Physics: Conference Series* **2012**, 012135 (2021).
- ⁵J. Wang et al., “Observing Exoplanets with High Dispersion Coronagraphy. I. The Scientific Potential of Current and Next-generation Large Ground and Space Telescopes”, *The Astronomical Journal* **153**, 183 (2017).
- ⁶D. Mawet et al., “High-resolution Infrared Spectrograph for Exoplanet Characterization with the Keck and Thirty Meter Telescopes”, *Bulletin of the AAS* **51** (2019).
- ⁷J.-L. Baudino et al., “Toward the Analysis of JWST Exoplanet Spectra: Identifying Troublesome Model Parameters”, *The Astrophysical Journal* **850**, 150 (2017).
- ⁸P. Tremblin et al., “Fingering Convection and Cloudless Models for Cool Brown Dwarf Atmospheres”, *The Astrophysical Journal Letters* **804**, L17 (2015).
- ⁹J. L. Baudino et al., “Interpreting the photometry and spectroscopy of directly imaged planets: a new atmospheric model applied to β Pictoris b and SPHERE observations”, *Astronomy and Astrophysics* **582**, A83 (2015).
- ¹⁰P. Mollière et al., “Model Atmospheres of Irradiated Exoplanets: The Influence of Stellar Parameters, Metallicity, and the C/O Ratio”, *The Astrophysical Journal* **813**, 47 (2015).
- ¹¹Q. M. Konopacky et al., “Detection of Carbon Monoxide and Water Absorption Lines in an Exoplanet Atmosphere”, *Science* **339**, 1398–1401 (2013).
- ¹²J. J. Wang et al., “Detection and Bulk Properties of the HR 8799 Planets with High-resolution Spectroscopy”, *The Astronomical Journal* **162**, 148 (2021).
- ¹³J.-B. Ruffio et al., “Deep Exploration of the Planets HR 8799 b, c, and d with Moderate-resolution Spectroscopy”, *The Astronomical Journal* **162**, 290 (2021).
- ¹⁴*Calculation of Molecular Spectra with the Spectral Calculator*, <https://www.spectralcalc.com/info/CalculatingSpectra.pdf>, (Accessed 20 August 2022).
- ¹⁵L. L. Gordley, B. T. Marshall, and D. A. Chu, “Linepak: Algorithms for modeling spectral transmittance and radiance”, *Journal of Quantitative Spectroscopy and Radiative Transfer* **52**, 562–580 (1994).
- ¹⁶J. S. Briggs, “A derivation of the time-energy uncertainty relation”, *Journal of Physics: Conference Series* **99**, 012002 (2008).
- ¹⁷C. Griffith, *Line Broadening*, https://www.lpl.arizona.edu/classes/Griffith_656A/RT5-linewidth.pdf, (Accessed 20 August 2022).
- ¹⁸*The Heisenberg Uncertainty Principle*, [https://phys.libretexts.org/Bookshelves/University_Physics/Book%5C%3A_University_Physics_\(OpenStax\)/University_Physics_III_-_Optics_and_Modern_Physics_\(OpenStax\)/07%5C%3A_Quantum_Mechanics/7.03%5C%3A_The_Heisenberg_Uncertainty_Principle](https://phys.libretexts.org/Bookshelves/University_Physics/Book%5C%3A_University_Physics_(OpenStax)/University_Physics_III_-_Optics_and_Modern_Physics_(OpenStax)/07%5C%3A_Quantum_Mechanics/7.03%5C%3A_The_Heisenberg_Uncertainty_Principle), (Accessed 20 August 2022).
- ¹⁹C. Hedges and N. Madhusudhan, “Effect of pressure broadening on molecular absorption cross sections in exoplanetary atmospheres”, *Monthly Notices of the Royal Astronomical Society* **458**, 1427–1449 (2016).
- ²⁰C. Hill, S. N. Yurchenko, and J. Tennyson, “Temperature-dependent molecular absorption cross sections for exoplanets and other atmospheres”, *Icarus* **226**, 1673–1677 (2013).
- ²¹X. Su, A. Fischer, and F. Cichos, “Towards Measuring the Maxwell–Boltzmann Distribution of a Single Heated Particle”, *Frontiers in Physics* **24**, 669459 (2021).
- ²²R. Nave, *Broadening of Spectral Lines*, <http://hyperphysics.phy-astr.gsu.edu/hbase/Atomic/broaden.html>, (Accessed 20 August 2022).

- ²³J. O. Sundqvist et al., “The rotation rates of massive stars”, *Astronomy and Astrophysics* **559**, L10 (2013).
- ²⁴M. Richmond, *Using spectra to derive motions*, <http://spiff.rit.edu/classes/phys301/lectures/doppler/doppler.html>, (Accessed 20 August 2022).
- ²⁵S. Chib and A. Belafhal, “Simple Analytical Expression of the Voigt Profile”, *Quantum Reports* **4**, 010004 (2022).
- ²⁶A. Oklopčić et al., “Detecting Magnetic Fields in Exoplanets with Spectropolarimetry of the Helium Line at 1083 nm”, *The Astrophysical Journal* **890**, 88 (2020).
- ²⁷R. Ignace, K. H. Nordsieck, and J. P. Cassinelli, “The Hanle Effect as a Diagnostic of Magnetic Fields in Stellar Envelopes. I. Theoretical Results for Integrated Line Profiles”, *The Astrophysical Journal* **486**, 550–570 (1997).
- ²⁸R. Ignace, K. H. Nordsieck, and J. P. Cassinelli, “The Hanle Effect as a Diagnostic of Magnetic Fields in Stellar Envelopes. II. Some Theoretical Results for Resolved Line Profiles”, *The Astrophysical Journal* **520**, 335–346 (1999).
- ²⁹H. R. Griem, “Stark Broadening”, *Advances in Atomic and Molecular Physics* **11**, 331–359 (1976).
- ³⁰S. Csizmadia et al., “The effect of stellar limb darkening values on the accuracy of the planet radii derived from photometric transit observations”, *Astronomy and Astrophysics* **549**, A9 (2013).
- ³¹L. S. Rothman et al., “The HITRAN 2008 molecular spectroscopic database”, *Journal of Quantitative Spectroscopy and Radiative Transfer* **110**, 533–572 (2009).
- ³²A. S. Burrows, “Spectra as windows into exoplanet atmospheres”, *Proceedings of the National Academy of Sciences* **111**, 12601–12609 (2014).
- ³³Y. K. Feng, M. R. Line, and J. J. Fortney, “2D Retrieval Frameworks for Hot Jupiter Phase Curves”, *The Astronomical Journal* **160**, 137 (2020).
- ³⁴J.-B. Ruffio et al., “Radial Velocity Measurements of HR 8799 b and c with Medium Resolution Spectroscopy”, *The Astronomical Journal* **158**, 200 (2019).
- ³⁵M. Brogi and M. R. Line, “Retrieving Temperatures and Abundances of Exoplanet Atmospheres with High-resolution Cross-correlation Spectroscopy”, *The Astronomical Journal* **157**, 114 (2019).
- ³⁶Statistics How To, *Cross correlation*, <https://www.statisticshowto.com/cross-correlation/>, (Accessed 10 September 2022).
- ³⁷K. Briechle and U. D. Hanebeck, “Template matching using fast normalized cross correlation”, *Optical Pattern Recognition XII* **4387**, 95–102 (2001).
- ³⁸NASA’s *Webb Takes Its First-Ever Direct Image of Distant World*, <https://www.jpl.nasa.gov/news/nasas-webb-takes-its-first-ever-direct-image-of-distant-world>, (Accessed 12 September 2022).
- ³⁹A. Predoi-Crossa et al., “Broadening, shifting, and line asymmetries in the 2,0 band of CO and CO-N₂ : Experimental results and theoretical calculations”, *The Journal of Chemical Physics* **113**, 158 (2000).