

The Search For Habitable Worlds:

How Would We Know One If We Saw One?

Dr Victoria Meadows

Introduction

Motivated by the discoveries of over a hundred extrasolar giant planets, NASA and ESA have initiated a series of mission concept studies for space-based astronomical observatories that will search for and study extrasolar Earth-sized planets, searching for signs of life. The search for life outside our solar system is just one facet of a new scientific field called "astrobiology." Astrobiology is the study of life in the Universe, and with life as its focus, covers a broad range of mostly interdisciplinary topics. These topics include but are not limited to: the astronomical and planetary processes and conditions that are conducive to life, the origin of life, the probability of life elsewhere and its likely environments, and the search for life in our own solar system and beyond. So, by finding and understanding the properties of extrasolar terrestrial planets, these planned missions will address one of astrobiology's fundamental questions, "Are we alone?"

In the previous chapter we focused on how we might *detect* planets outside our solar system. In this chapter we will focus on how we might *characterize* these distant planets. To do this, we must learn how to recognize worlds that might have habitable conditions, and to discriminate between planets with and without life, *using only remotely-sensed information*. This improved understanding will help us to optimize the designs and search strategies for planet detection and characterization missions, and to ultimately interpret the remote sensing observations that they return.

Habitable Worlds

So what *is* a habitable world? In its most conservative definition, a habitable world is a solid-surfaced world, either a planet or a moon, that can maintain liquid water on its surface. Although this definition has the advantage of describing worlds that would be more *detectable* as habitable over enormous distances, it is considered by some to be too limited. Why not life in the clouds of a giant planet, or life that is not carbon-based, and that could exist in different temperature regimes? Why not life that originates or survives in a solvent other than water? However, in drafting our "conservative" definition, we have concentrated first on what we know to be true for life on our own planet. Water is the one common constituent used by the enormous array of life forms on Earth. If we adopt liquid water as an important component for life, then our habitable world is limited by two required conditions: the presence of water, and a range of planetary temperatures and pressures that are conducive to keeping water in its liquid state.

To achieve these conditions, the habitability of a planet is in part determined by its formation heritage and early history, which will set its mass, orbit, spin and chemical composition. Habitability will subsequently depend upon the evolution of the planet's interior and its crust, and the evolution of planetary volatiles after they reached the surface of the planet. This is because the temperature of a planet's surface depends on the brightness of the parent star, the planet's distance

from that star and its orbital shape, the planet mass, the reflectivity of the planet's surface, the density and composition of its atmosphere, and the availability of internal (geothermal) energy.

So even our conservative definition of habitability still encompasses a vast array of potential worlds that could be considered habitable, without being at all like the current Earth. For example, let's imagine a planet more massive than Earth, with a dense carbon dioxide atmosphere, and on a slightly non-circular orbit that takes it further from its parent star than the Earth is from the Sun. This planet's dense atmosphere would provide a "greenhouse effect" which would help to maintain liquid water on its surface by smoothing over the effects of any changes in received sunlight due to its orbit. This planet would not be at all "Earth-like", but by our definition, it would be habitable.

Would The *Real* Earth Please Stand Up?

In trying to search for examples of the range of possible extrasolar planets that might be out there, one need look no further than our own home planet. When one thinks of an "Earth-like" planet, most people envisage a sparkly blue-green world, populated with plants and animals, and with a nitrogen dominated atmosphere and abundant oxygen. However, the Earth has undergone many dramatic changes in its 4.6 billion year history, and had very different characteristics over this period. So much so, that you might have difficulty in picking it out of a line-up! For example, the Earth has only had abundant free oxygen in its atmosphere for the last 2.5 billion years. For the first 2 billion years, there was no detectable oxygen at all. However, as different as it appeared, in most of these incarnations the Earth was not only still habitable, but teeming with life. In this way, the Earth can provide us with geological and other constraints for a range of potential habitable and inhabited worlds, without even having to set foot outside the Solar System!

The Family of Earths

To understand the diversity of worlds that the Earth has represented throughout its history, let's briefly recap some of the things that we understand about these different periods.

The Hadean (4.6-4.0 Gya)

Relatively little is known about this "hellish" period of the Earth's history, but it is likely that it was characterized by dynamic and violent change. 4.6 to 4.5 billion years ago (Gya) the forming Earth was hit by a Mars-sized object, and formed the Earth - Moon system. As the molten Earth cooled, it formed a crust, with sea-floor crust forming first, and continental crust forming from subducted oceanic crust by about 4.0 Gya. Because the creation of new continental crust is an ongoing process, there was much less continental crust present during the first half of the Earth's history, and probably very little continental material at 4.0Gya. Since the oceans were present by about 4.0Gya, the Earth may have appeared even more ocean-dominated than it is today. On the other hand, topographic relief may have been present even though there were no "true" continental masses, as is the case on Mars and Venus, which have mountains and basins, despite having no classic continental crust material.

The presence of carbonate rocks at 3.8Gya, which must be formed in the presence of liquid water, may indicate both the presence of widespread oceans at that time, and an early atmosphere that was abundant in carbon dioxide. This carbon dioxide may have helped to warm the surface of the Hadean Earth to temperatures as high as 85C. Volcanism would have been far more active than it is today, as the Earth lost its internal energy through a thin crust.

Studies of lunar craters suggest that large impacts were also far more common and may have catastrophically modified the climate and composition of the atmosphere. The larger impact events

during this period would likely have heated the Earth's surface to temperatures as high as 2000K locally, vaporizing the oceans and producing a thick "steam" atmosphere. The ensuing water vapor greenhouse effect may have kept the Earth uncomfortably warm for several thousand years after the impact. Even the smaller impacts during this period would have potentially vaporized the uppermost several hundred meters of any ocean that was present, destroying any surface life, if it had existed. This impact-besieged Hadean Earth, although periodically habitable, was not continuously habitable, and to date has revealed no signs that it was ever inhabited.

To a distant observer, this planet might have appeared as a water-dominated world, with a periodic steam atmosphere and high temperatures. But even that may have been difficult to tell from a distance, because of the relatively dense dust and rock disk in which the planet would have been embedded.

The Archean (4.0-2.5 Gya)

This period of the Earth's history is characterized by abundant microbial life, and an environment with little resemblance to the modern Earth. The impact deluge experienced during the Hadean continued through to the end of the Late Heavy Bombardment at 3.8Gya. At this time, as has been determined from cratering rates observed on the Moon, the number of impacts dropped dramatically to a few percent of its Hadean value, and decreased even further with time. The oldest evidence for life dates to 3.5Gya, and life is believed to have originated somewhere between 4.0 and 3.5Gya. Interestingly, the time period between when the Earth stopped being constantly sterilized by impacts, and the first evidence for life, is relatively short, only a few hundred million years, and there is some (albeit controversial) evidence, that it may have been as short as 100 million years or less. As soon as the Earth *became* habitable, it *was* inhabited in relatively short order!

During the Archean period microbial life was abundant and widespread, and survived in an atmosphere that was relatively rich in carbon dioxide and nitrogen, but initially very poor in oxygen. Life in this period may have consisted of archeobacteria called "methanogens" which metabolised hydrogen and acetate to produce methane. The Archean was also possibly a waterworld, with the majority of the present-day continental material not being formed until the end of the period, at about 2.5Gya. This event also likely marked the birth of modern-day plate tectonics.

Throughout most of the Archean, a distant observer would have seen a planet with a surface that would have appeared water-dominated, and the atmosphere would have been rich in methane, carbon dioxide and water vapor, but with little or no detectable oxygen or ozone.

However, towards the end of the Archean, rising atmospheric oxygen levels (to 0.2% of the Earth's present day value) heralded the development of oxygen producing life. The oxygen was most likely produced via the method of photosynthesis, which was employed by bacteria and plants. However, the breakdown via sunlight of the methane that was produced by early life, may have enhanced the availability of free hydrogen in the atmosphere. This would have increased its subsequent loss to space, which would have also contributed to the subsequent oxidation of the atmosphere. For a distant observer, this brief period would be an interesting time to catch, as it would provide an opportunity to monitor the evolution of the planet's environment in the presence of life.

The Proterozoic (2.5-0.6 Gya)

This period of the Earth's history was characterized by the development of an oxygen-rich atmosphere, the emergence of multicellular life forms, and large changes in global climate. Based on evidence from the geological records, between 2.5 and just before 2.0 Gya, the Earth saw widespread glaciation, with the land masses covered with ice down to the mid-latitudes. In this period, the atmospheric oxygen concentration increased to about half of its present-day value. This rapid rise probably does not indicate the onset of photosynthesis, but rather the time at which the Earth's ability to store or "sequester" the oxygen away in the crust or the oceans was exhausted, allowing it to build up in the atmosphere instead. At this time, the Earth from space would have shown ice-dominated continents, and the signature of carbon dioxide, oxygen and ozone in its atmosphere.

As oxygen rose, atmospheric CO₂ amounts decreased, with a sharp decrease occurring around 1Gya. CO₂ is delivered continuously into the atmosphere from the Earth's interior via volcanic eruptions. So the observed reduction was due to a more efficient sequestering process. This process was most likely carbonate formation in the Earth's oceans. The amount of CO₂ that can be dissolved in the ocean to react to form carbonates is very sensitive to ocean temperature. As the organisms using photosynthesis and aerobic respiration prospered, the anaerobes that had dominated the Archean died.

The appearance of oxygen resulted in an explosion of life, by enabling both aerobic respiration, a highly efficient means of generating energy from the environment, and the formation of a planetary ozone shield. The oldest evidence for multicellular life was found in leaf-like fossils dated 1.7 Gya. Between 900 and 600 Mya, the Earth experienced a further series of ice ages, some so severe that ice sheets may have extended below 25 degrees latitude. Several tens of millions of years before the end of this era there is fossil evidence of soft-bodied creatures with many diverse body types, few of which survived in the creatures we know today. Toward the end of this period in the Earth's evolution, a distant observer would have seen a planet that looks much more like the world that we know today.

The Phanerozoic (0.6Gya to the present)

The beginning of this era (the Vendian and the Cambrian periods) is marked by the appearance of fossils of many different forms of life, and many which apparently existed on land. Land plants are seen in the geological record 100 million years later. As viewed from space, this modern Earth now has oceans, continents, and new surface signatures of chlorophyll and leaf reflectance from plant life. The atmosphere is dominated by nitrogen and oxygen, with a strong ozone shield. Carbon dioxide and methane from land plants and animals are also seen in the atmosphere.

When we search for planets around other stars, it is important to remember this "family of Earths" and recognize that a habitable planet, and in fact one that is already inhabited, does not necessarily have to look like our modern-day Earth.

Habitable Zones

As we examine the history of our planet, it is clear that after the late heavy bombardment, the Earth was habitable for very long periods of time, and probably continuously, at least in some environments. Today, our Earth is a beautiful oasis for life, with moderate temperatures and a liquid ocean. But on either side, it is flanked by far less lovely worlds: the arid, freezing desert of Mars, and the hellish, furnace-like desert of Venus. Earth's somewhat miraculous condition is governed by many factors, but the most important factor is the placement of our Earth in its solar system. The regions around a star in which a planet is habitable at a given time, and can be kept

habitable over a continuous period of time are known as the “instantaneous habitable zone”, and the “continuously habitable zone” respectively.

The Instantaneous Habitable Zone

Within this distance range from its parent star, a planet is able to maintain liquid water on its surface. The boundaries of this zone are determined in scientific computer models by taking an Earth-like planet, and moving it closer to its parent star. Just before the planet would become so hot that its oceans would boil in a process called “the runaway greenhouse” we reach the inner edge of the habitable zone. If we move the Earth-like planet away from its parent star, then just before the surface freezes catastrophically in a process sometimes called “runaway glaciation” then we have reached the outer edge of the habitable zone. For our own Solar System the habitable zone is estimated to be between 0.93 and 1.37 times the distance from the Earth to the Sun (which is 1 Astronomical Unit, or AU). For reference, Venus has a mean distance of 0.7 AU from the Sun, and Mars has a mean solar distance of 1.52 AU.

Figure 1: The Habitable Zone around stars of different spectral type. The x-axis gives the distance from the parent star, and the y-axis gives the mass of the star, which is an indication of the spectral type and temperature of the star. Our own Sun is of spectral type “G”, and the position of the Solar System planets are shown relative to the habitable zone. The ribbon shape in this plot shows the extent of the habitable zone around stars of different spectral type. The inner edge of the habitable zone is essentially bounded by the distance at which a planet would boil its oceans, and the outer edge is bounded by the distance at which a planet would freeze surface water. For lower mass stars that are cooler than our Sun (spectral types K and M) the habitable zone is much closer to the star. For higher mass, hotter stars (spectral type F and A) the habitable zone is much further from the parent star. The tidal lock radius line shows the distance a planet would have to be from its parent star to avoid being tidally locked (having one face of the planet always turned towards the star). For planets closer to their stars than the tidal lock radius, it might be difficult to have a habitable planet, even if it within the habitable zone, because the atmosphere may eventually all freeze out on the far side of the planet.

If we have discovered a planet around another star, then determining whether or not it is in the habitable zone for its planetary characteristics is of interest, because it helps us understand whether the planet is likely to be habitable. However, if we are interested in determining whether it is likely that the planet supports life, we need to know how long the planet could have been kept habitable, so that life had an opportunity to originate and develop. It is estimated that our Solar System has had a CHZ spanning 0.95-1.15AU in the past 4.6 Gy. However, if the Sun becomes 10% brighter in the next 1.1 billion years, the Earth’s surface temperature may increase to the point where we leave the CHZ in another 500-900 million years!

So now we understand a little bit more about what it means to be habitable. But how would we actually recognize whether an extrasolar terrestrial planet is habitable? And what are the signatures of life on a planet that might be visible from space, and so could potentially be observed and understood by a distant observer in another planetary system? To answer these questions, we need to understand what to look for, and how to measure and classify the global-scale properties of terrestrial planets.

Characterization of Extrasolar Terrestrial Planets

In the next decade, we expect to have space-based instrumentation that will allow us to detect and characterize extrasolar terrestrial (Earth-sized) planets. By necessity, we will study these planets from a distance (remote-sensing), because planets found around even the few closest stars to our Solar System will be impossible to travel to for more detailed, close-up examination. So once an extrasolar terrestrial planet has been detected and resolved from its parent star, all the information about its environment and characteristics will arrive as electromagnetic radiation. The information carried in spectra by these electromagnetic photons can be decoded using remote-sensing techniques.

The new field of extrasolar terrestrial planet characterization is theory-based, because existing observing techniques are not yet sensitive enough to directly detect and gather information on Earth-sized planets around other stars. This new field seeks to improve our understanding of the potential range of characteristics for terrestrial planets in our galaxy, and the observational signatures that we are likely to encounter.

How Can We Tell If A Planet Is Habitable?

To understand whether a planet is habitable, we must learn as much about it as possible using remote-sensing techniques, many of which have already been developed for studying the Earth and other planets in our solar system. Some of these techniques include (i) determining the brightness and color of the planet, and whether there are any changes with time, (ii) analysis of spectra to determine atmospheric and surface composition and physical parameters, and (iii) monitoring spectral features to look for diurnal or seasonal variations in surface reflectivity or atmospheric composition. However, studying extrasolar planets using these techniques poses special challenges. This is because the planet will appear as an unresolved point source, a single dot, and there will be no direct constraints on size, or any spatial details.

The Planetary System Environment

When an extrasolar terrestrial planet is first detected via direct astronomical observation, we will initially characterize it based on its solar system environment, that is, what sort of a system is it embedded in? Astronomical observations can be used to determine the spectral type and luminosity of its parent star, the planet's orbital characteristics and its placement in the solar system relative to the star, the type and relative position of other planets in the system, and whether or not the newly discovered planet has rings or moons, if this can be determined.

Photometry and photometric variability

Once the solar system environment is understood, additional information about an extrasolar planet can be extracted by looking at the total brightness of the planet (photometry), and how that brightness changes with time (the photometric variability, or "lightcurve"). Photometric observations have provided a wealth of information about planets in our solar system. This experience is important for understanding what might be learned from photometric observations of extrasolar planets. The planet's apparent brightness in one or more wavelength ranges can be used to determine its effective temperature, and its "color", the ratio of the brightness of the planet in two defined wavelength ranges, could provide clues to the composition of its surface and atmosphere. Time-resolved photometric observations acquired as the planet rotates may reveal variability in brightness or color that could indicate the presence of clouds or surface variations. Photometric

observations acquired as the planet moves around its parent star may also provide a way to discriminate between daily and seasonal variations in its environment.

Remote Sensing Spectroscopy

Although photometric measurements could yield valuable information about an extrasolar terrestrial planet, by far the most powerful technique available for retrieving the characteristics of planetary surfaces and atmospheres is spectroscopy. As ultraviolet (UV), visible, and near infrared (NIR) radiation from the parent star is *scattered* and *reflected* from the planet, the material that makes up the planet's surface and atmosphere interact with the incoming solar radiation. Similarly, at typical planetary temperatures, significant thermal infrared (IR – wavelengths longward of $5\mu\text{m}$) and microwave radiation is also *emitted* by the planet. As this radiation escapes to space, it also interacts with the planetary atmosphere. These interactions produce distinctive absorption and emission lines in the observed spectrum that provide clues to the composition and physical properties of the atmosphere and surface (Figure 2).

Figure 2: Terrestrial Planet Spectra. The mid-infrared spectra of Venus, Earth and Mars. All three terrestrial planets show a strong absorption at $15\mu\text{m}$ that is due to absorption by CO_2 in their atmospheres. Earth, however, distinguishes itself from the other two planets in this wavelength range by also showing strong absorption features from O_3 and H_2O in the atmosphere. If seen at high-enough spectral resolution, the shape of the CO_2 band can be used to retrieve information about the temperature structure of the atmosphere (c.f. Figure 3).

Because the spectral position of each absorption or emission line corresponds to the internal energy difference between two vibration-rotation states of that particular molecule, it provides a unique spectral “fingerprint” that can be used to infer the composition of the atmosphere. In addition, the temperature and pressure environment of the molecule can affect the shape and width of the line (the “line profile”). “Pressure broadening” of the line width is determined primarily by the frequency of collisions with other molecules, and this effect dominates near the surface of a planet where the pressures (and collision rates) are high. “Doppler broadening” of the line is a function of the mean velocity of the molecules, and this effect dominates at lower pressures, higher in the atmosphere. The observed line broadening can be used to “retrieve” information about the temperature and pressure of the atmosphere.

In addition to looking for changes in the width of a line, at thermal wavelengths, atmospheric conditions can also be inferred from the shape of the line. Whether a feature is in absorption or emission, or a composite of the two, can give clues to the temperature structure along the line of sight, and the atmospheric environment between the emitting molecule and the observer. For the Earth, water vapour distributed between the warm surface and cold tropopause predominately absorbs thermal radiation from the surface, producing absorption features in a thermal spectra recorded from space. Ozone in the cold lower stratosphere also predominately absorbs thermal radiation from the warm surface and lower atmosphere. However, ozone in the warm upper stratosphere emits enough thermal radiation to reduce the depth of the observed absorption feature (Figure 3).

Figure 3. Emission and absorption in a planetary atmosphere. This figure shows the effect on mid-infrared spectral features observed at the top of the Earth's atmosphere, due to the interplay between the vertical position of an atmospheric constituent and its surrounding temperature environment. The x-axis shows atmospheric

temperature, and the y-axis gives atmospheric altitude. The atmospheric temperature structure is shown a black line relative to the vertical distributions of tropospheric water vapor (orange band), and stratospheric ozone (blue band). Water vapor, which is concentrated near the surface, absorbs upwelling thermal radiation from the warm, sun-heated ground and it shows an absorption feature in a planetary spectrum (orange side panel). The signature of ozone is more complicated. Ozone in the lower stratosphere absorbs upwelling thermal radiation from the surface, producing an absorption feature, but ozone also emits thermal radiation from the warm upper stratosphere, which has no cold atmospheric layer above it to absorb, so also produces an emission feature. The ozone feature seen from the top of the atmosphere is thus a composite of the lower stratospheric absorption and the upper stratospheric emission (three blue side panels). This composite spectral feature can be used to infer the vertical temperature structure of the upper atmosphere. Carbon dioxide (not shown) is however the best probe of temperature structure at these wavelengths, because unlike O_3 it is evenly mixed throughout the entire atmosphere.

Bulk and Trace Atmospheric Constituents

Using disk-averaged spectra, it is possible to detect a planetary atmosphere, and determine its bulk (principal) and trace (smaller component) gas composition. Ironically, Earth's principal atmospheric constituent, gaseous nitrogen (N_2), has no strong signature throughout most of the spectrum, except in the far UV. From the UV to the visible, this gas also affects the scattering properties of the atmosphere, but is otherwise undetectable by direct methods! However, in some cases the presence of a bulk component that is radiatively inactive (like N_2), can also be inferred if it contributes significantly to the pressure broadening of the spectral absorption of other gases in the atmosphere.

Most other constituents of terrestrial atmospheres, such as H_2O , CO_2 , O_3 , N_2O , NO_2 , NO , CO , CH_4 , O_2 , etc., do produce distinctive spectral signatures at UV, visible, and/or infrared wavelengths, that can be directly detected, and several, like CO_2 , can be easily detected. (Figure 2).

Atmospheric Mass

Modeling the effects of broadening seen in infrared bands of absorbing gases as a result of the amount of atmosphere the gas is in, can be used to infer the atmospheric pressure at the emitting level in an atmosphere. For transparent atmospheres (e.g. Earth, Mars) radiation escapes from the surface and upward through the entire atmosphere, allowing the total mass of the atmosphere to be derived from the planet's spectrum. For opaque atmospheres, (e.g. Venus or the Jovian planets), only the total atmospheric mass above a cloud deck or other opaque surface can be determined, because that is the limiting depth from which we can see atmospheric radiation escape. Observations at UV, visible, and IR wavelengths will probe to different levels of the atmosphere, and will indicate different effective pressures, providing additional constraints on how pressure varies with altitude in an atmosphere. This information is important both for studies of the planet's climate (which can be used to infer surface temperature) and for quantifying the trace gas mixing ratios for studies of the atmospheric chemistry.

Surface and Atmospheric Temperature

Measurements of the surface temperature of an extrasolar planet are essential to characterize its near-surface environment and habitability. This information can be retrieved from thermal infrared or microwave spectra acquired at wavelengths where the atmosphere is relatively transparent. For an Earth-like terrestrial planet whose atmospheric absorption is dominated by H_2O , CO_2 , and O_3

(e.g. Earth, Mars), surface temperature data can be collected most easily in the spectral window at wavelengths between 8 and 12 μm (Figure 4). At these wavelengths, the ubiquitous water vapor continuum absorption and strong, narrow O_3 band near 9.6 μm are the only significant sources of atmospheric absorption. At shorter and longer thermal wavelengths, strong absorption by H_2O and CO_2 largely preclude surface observations. Note though, that the accuracy of the surface temperatures derived from these measurements rely on understanding the planet's surface composition, which could be anything from ice to molten rock, as well as the abundance of water vapour, which produces weak continuum absorption throughout this spectral range. If the surface composition, and most importantly, its emissivity, is not known, then it is very difficult to retrieve an accurate temperature.

Figure 4. The Earth in the Mid-Infrared. This plot shows synthetic spectra generated for lines of sight from the top of the Earth's atmosphere over different surface types (forest, ocean, desert) and for two cloud types (high cirrus cloud and lower stratus cloud, both over ocean). These spectra were generated with the Virtual Planetary Laboratory's spectral mapping atmospheric radiative transfer (SMART) model (see below) and are validated against satellite observations of the Earth. In each case, the sun is 60° from the zenith, and the atmospheric thermal structure and the abundance profiles for H_2O , CO_2 , O_3 , N_2O , CO , CH_4 , and O_2 are provided for a mid-latitude summer atmosphere. As can be seen from this plot, the majority of this mid-infrared spectral range is not sensitive to radiation from the surface. However, the range 8-13 μm , is an "atmospheric window." Within this wavelength range it is possible to see to the surface of the planet, if the view is not obstructed by clouds or strong molecular absorption. The strong absorption band seen at 9.6 μm is due to ozone, and the strong band at 15 μm is due to CO_2 . Other species, such as water vapour and CH_4 absorb near 6.3 and 7.7 μm respectively, but their spectral signatures are far more subtle than O_3 and CO_2 .

Atmospheric temperature information can be retrieved from the spectrum of a well-mixed gas with a well-characterized absorption features. For the Earthlike terrestrial planets in our solar system (Venus, Earth, and Mars), observations of the CO_2 15 micron band provide the best available constraints on the atmospheric thermal structure. For planets with H_2 dominated atmospheres (Jupiter, Saturn, Uranus, Neptune), the atmospheric temperature structure can be retrieved from the hydrogen continuum absorption at wavelengths longer than 20 microns.

Because infrared spectral features seen in the spectra of planets are sensitive to both the amount of constituent in the atmosphere, and the temperature structure of the atmosphere between the constituent and the observer, atmospheric temperatures must also be measured before we can quantify trace gas amounts from thermal radiances. For example, on Earth, accurate estimates of the atmospheric temperature distribution are essential for quantifying the atmospheric concentrations of gases such as water vapour and ozone.

Like determination of atmospheric pressure, global scale constraints on the atmospheric thermal structure are valuable for studies of the planet's climate and chemical equilibrium, and estimates of the globally-averaged surface temperature would be useful for determining directly whether liquid water is stable on the surface. Methods to derive temperatures from remote-sensing spectra have already been developed for routine use in cloud-free conditions on Earth and for the Jovian planets. More advanced methods, for use in strongly scattering atmospheres (e.g. Venus and Mars) are also under development. Methods specifically adapted for the analysis of full-disk spectra of planets with mixed surface composition, such as the Earth, are also needed.

Surface Composition

One advantage the optical wavelength regime has over the mid-IR is that although the mid-IR region is sensitive to both surface and atmospheric temperatures and trace gas abundances, it is extremely insensitive to underlying surface composition. This is due in part to the relatively small "atmospheric windows" to the surface in the MIR range (see Figure 4), but is principally due to the lack of spectral features and the uniformly high emissivity of virtually all surface types (ocean/land/vegetation) at these wavelengths. (In analysis of mid-IR remote-sensing data of the Earth, the emissivity of virtually all surfaces is routinely set to 1.0 as an acceptable approximation!). However, as shown in Figure 5 the optical and near-infrared regions of the spectrum display a rich array of spectral features associated with surface composition. Some of these include the reflectance spectrum of the ocean, mineralogical features from rock and sand, and reflectance features from ice. With spectra, or with observations taken at a set of discrete wavelengths, we might be able to use optical observations to distinguish between a world dominated by oceans, sand, rock or ice.

Figure 5: The Earth in the Visible and Near-Infrared. In a similar manner to Figure 4, this plot shows the visible and near-infrared synthetic spectra generated for multiple lines of site from the top of the Earth's atmosphere with different underlying surface types (ocean, desert, forest and cloud). However, unlike the mid-infrared, this plot shows that the surface type contributes significantly to variations in the observed spectrum at visible and near-infrared wavelengths. Atmospheric absorption from molecules of interest such as ozone ($<0.3 \mu\text{m}$), water vapor, and the O_2 A-Band at $0.67 \mu\text{m}$ are marked.

Phase and Seasonal Variations in Spectral Features

In a similar manner to the photometry, understanding how different spectral features change with time may help us infer spatial information in disk-averaged observations (e.g the presence of continents or oceans), and may also reveal variations in surface and atmospheric composition that are linked to day-night or seasonal variations (e.g seasonal ice caps or dust storms). Variations in the retrieved atmospheric temperature or pressure variations could also yield information about global scale weather systems, or seasonal variations in the total atmospheric mass, like those that characterize the Martian atmosphere. Information on the surface type, temperature variations and climate over the course of an orbit are all important for understanding the potential habitability of a planet.

Although the mid-IR and the optical/Near-IR spectral regions have many different characteristics, neither region yet shows a clear advantage for detection of planetary characteristics. In fact, it would be highly desirable to characterize planets in more than one wavelength range, to reduce spectral confusion due to overlapping spectral features associated with different constituents in the system.

How Can We Tell If A Planet Is Inhabited?

In addition to characterizing the planet for potential habitability, it is of extreme interest to also look for signs that the planet is in fact, already inhabited. In astrobiology parlance, signs of life, either past or present, are called "biosignatures". Signs of life that can be inferred from remote-sensing or astronomical measurements are called "astronomical", or "remote-sensing" biosignatures. A

definitive list of biosignatures does not yet exist, and the field of astrobiology is working towards developing these. Generally speaking though, remote-sensing biosignatures could take the form of photometric or spectral signatures that are seen alone, or in certain combinations, or that exhibit time-varying behavior that is considered indicative of life. Biosignatures could also exist anywhere across the electromagnetic spectrum.

The Signs of Life on a Planetwide Scale

The search for remote-sensing biosignatures is based on the understanding that widespread life will modify the atmosphere and surface of its planet, and that these modifications will be detectable on a global scale. On our own planet, this thesis is demonstrably true, with life-induced changes to our planet's surface and atmosphere that are global, and visible from space. Widespread life interacts with the surface and atmosphere, affecting many atmospheric and geological processes, as well as affecting the albedo and spectral properties of planetary surfaces.

Initial work on the detectability of remote-sensing biosignatures focused principally on biosignatures detectable in a narrow range of wavelengths in the mid-infrared. This was driven by initial concepts for missions like NASA's Terrestrial Planet Finder, which were constrained by astronomical and technical implementation considerations to favor observations in the mid-infrared where the contrast between the planet and its star was at its lowest, and the planet would be most visible. However, new studies indicate that coronagraphic techniques operating in the visible and near-IR can also provide the required sensitivity to see Earth-sized planets around their stars. Therefore, potential signatures from the visible through the mid-infrared will be discussed here.

Atmospheric Signatures

An atmospheric signature of life could either be the detection of a single atmospheric constituent in a sufficiently large quantity that it cannot be explained as a product of a non-biological process, or a combination of atmospheric constituents whose simultaneous presence, or inferred abundance ratio are again unlikely via non-biological processes. For example, in our own atmosphere, the abundant oxygen (20.95% of the total atmosphere) is difficult to produce via geological or photochemical processes. Instead, the oxygen in the Earth's atmosphere is produced principally by photosynthetic organisms such as bacteria and plants. Oxygen can be detected most easily in the Earth's spectrum at visible wavelengths near 0.76microns, the so-called oxygen "A-band" (see Figure 5). In the mid-infrared, molecular oxygen has no prominent spectral features, and the presence and concentration of O₂ must be inferred from detection of ozone at 9.6 μ m (see Figure 4). This determination of oxygen concentration based on the abundance of ozone is done via atmospheric chemistry models which indicate significant concentrations of ozone even at relatively low oxygen levels (Kasting and Donahue, 1980). While this can be an advantage, because in some ways ozone is a more sensitive indicator of oxygen than oxygen itself, inferring O₂ abundance from O₃ may or may not be robust in atmospheres with different chemical composition and incoming solar flux to our own. In addition, our ability to detect ozone is also dependent on the temperature structure of the atmosphere. The combination of the non-linear drop off in ozone abundance with reduction in oxygen abundance, and the importance of the atmospheric temperature structure, which can change markedly with decreasing oxygen and ozone, can lead to the ozone absorption feature remaining strong, even as the oxygen level decreases (Figure 6).

Figure 6: Synthetic spectra showing the depth and shape of the 9.6 μ m ozone band for different amounts of atmospheric oxygen in Earth-like atmospheres. "PAL" denotes "present atmospheric level" of O₂ (about 21% of the atmosphere). Note that

it is not until the O₂ concentration is reduced by a factor of 1000 from PAL that we see a significant change in the absorption depth of the O₃ band.

It is also important to remember that oxygen as a large fraction of the total atmosphere has only been characteristic of the Earth's spectrum since the Proterozoic, i.e. for about half the time that the Earth has supported life (Holland, 1994). Prior to that time, the dominant microbial life-forms produced different gas products, such as methane. Remote-sensing biosignatures detectable before oxygen became a major constituent of our atmosphere are currently not well understood, and it cannot be assumed that all planets with life have evolved down the same path taken by our Earth.

Another type of atmospheric biosignature is the simultaneous presence of strongly oxidized and reduced gases that are not in chemical equilibrium (e.g. O₂ and CH₄ in the Earth's atmosphere, Lovelock and Margulis, 1974). This type of biosignature is thought to be robust for many different kinds of planetary atmospheres. However, to be interpreted correctly, some understanding of the planetary environment is required to set the different components in context. This is especially important because on another planet, or on Earth earlier in its history, the biosignature pair may not necessarily be O₂ and CH₄, but other combinations of gases that must be assessed relative to the rest of the atmosphere to determine the equilibrium state. Also, these robust two-component indicators are generally much harder to detect via remote-sensing techniques. In the case of the Earth, the oxygen is relatively readily detectable, at least in the visible, but the methane is at much lower concentration and is spectrally most active at near-infrared wavelengths near 2.2 μ m (a wavelength region that is not currently being considered for the first generation planet finding instruments) and at thermal wavelengths, in a spectral range that includes strong water vapor bands. Consequently the methane is much harder to detect, and requires higher sensitivity and spectral resolution to disentangle its contribution to the spectrum from that of water vapor.

Surface Signatures

As described previously, the optical and near-infrared regions of the spectrum display a rich array of spectral features associated with surface composition, such as rock, ice or water. However, the Earth also displays a spectral signature due to the plant life that covers much of its surface. Chlorophyll absorbs strongly in the UV and blue (<0.5 μ m) and in the red (0.6-0.7 μ m), and has slightly less absorption in the green (0.55 μ m). However, the most detectable spectral feature of plants is produced by the change in refractive index between air and the internal leaf structure, which makes plants highly reflective just beyond the visible range (>0.7 μ m). This feature, combined with the chlorophyll absorption just shortward of 0.7 μ m, results in a strong discontinuity in plant reflectance at ~0.7 μ m, which is known as "the red edge" (see, for example, Elachi, 1987; Short, 1982) (Figure 7). This property of plants is widely used for remote sensing of the Earth via satellite, and can be used to monitor vegetation coverage over particular portions of the Earth. However, it has also been shown that it is visible in the Earth's global spectrum, (although very faintly) by observing spectra of Earth light reflected from the dark side of the Moon (e.g. Arnold et al., 2002).

Figure 7: Synthetic spectrum of a line of site through the Earth's atmosphere over a conifer forest, with chlorophyll absorption and the red-edge reflectivity marked. Chlorophyll, a potentially important biosignature, has strong absorption in the UV and blue (<0.5 μ m) and in the red (0.6-0.7 μ m – marked in green), and slightly less absorption in the green (0.55 μ m). Due to changes in the refractive index between air and the internal leaf structure, plants are also highly reflective just beyond the visible

range ($>0.7 \mu\text{m}$), resulting in a prominent discontinuity (marked in red) known as "the red edge" (see, for example, Elachi, 1987; Short, 1982).

Temporal Signatures

Much of the life on Earth is photosynthetic, or relies on photosynthetic life for its existence. Consequently, many life processes and life cycles on our planet are tied to the amount of sunlight received at the surface over a day, or throughout the seasons. Consequently, instead of a sign of life being either a single photometric brightness, spectral feature or combination of features, it could instead be a time-varying behavior of either one of these features, or a different feature entirely. For example, a "snapshot" spectrum of the Earth would show the presence of CO_2 and CH_4 . Observed separately, and not in the presence of oxygen or ozone, it would be hard to conclude that these gases are produced by biology, since they can be produced by photochemistry and geological processes also. However, sensitive spectroscopic observations of the Earth taken over a period of time and examined as a set would reveal periodic variations in the atmospheric CO_2 and CH_4 abundance. This behavior could be shown to be correlated with season (highly unlikely for a geological process!) and on our Earth this observed behaviour is known to be linked to seasonal variations in the respiration and photosynthesis of land plants. However, these seasonal variations are quite small, and would require a very sensitive instrument to detect them. This is perhaps beyond the ability of the first generation of planet detection and characterization missions. Another time-variable sign of life might be vegetation coverage as a function of season, which might be detected spectrally or photometrically. One must be cautious, however. Not all time-variable surface signatures are due to life. Numerous astronomers from the late 19th and early 20th century attributed seasonal albedo variations on Mars to variations in vegetation, when the true cause was the seasonal cycle of dust storm activity.

Sensitivity to Cloud Cover

Another factor that must be considered when attempting to characterize a planet is the potential loss of information due to persistent cloud cover. Typically, clouds are associated with convection and condensation of a volatile species, like the water ice clouds seen in the atmosphere of the Earth, or the CO_2 clouds seen on Mars. However, "clouds" can also result from photochemical processes, like the planetwide haze layers that dominate the atmospheres of Venus and Titan. As we show in Figure 4, cirrus clouds high in the Earth's atmosphere can obliterate even a strong signal due to ozone at mid-IR wavelengths, although clouds at lower levels still allow the detection of O_3 . On the other hand, in the visible, the strong A-band of O_2 is visible in oxygen rich atmospheres, even in the presence of high cloud, although its contrast is somewhat reduced (see Figure 5).

Although the photochemical hazes that shroud Venus and Titan are opaque at visible wavelengths, they display "atmospheric windows" at near-IR wavelengths (Meadows and Crisp, 1996; Smith et al., 1996), which allow penetration and remote-sensing of the underlying planetary surface. For Venus, thermal radiation from the hot surface and lower atmosphere escapes through the clouds and can be detected only on the night side of the planet. In the case of Titan, the haze is sufficiently transparent at near-IR wavelengths the surface can be detected even when the satellite is fully illuminated.

Simulating Planets and their Spectra

To better understand what we might encounter when we search remotely for signs of life on planets in other solar systems, we must try to understand the range of terrestrial planets that might exist. Once we better understand this, we can determine what we ought to look for, by understanding which characteristics of these planets we might be able to detect remotely. Since we will not have any direct observations of extrasolar terrestrial planets for some time, the only way we can gain more insight at this time is to use computer modeling to explore the possibilities.

The spectrum of a planet is the product of the complex interplay of a range of environmental components and processes. Hence, to generate realistic spectra of a range of plausible extrasolar terrestrial planets we must simulate planetary environments that include these factors. The models used to generate these conditions must be consistent with known physical, chemical, and biological processes. The basic components of such a model therefore include: incident solar/stellar flux; thermal structure and composition of the atmosphere, including gases, clouds and aerosols, and surface properties (land/ocean/ice/biology). In addition, the evolution and equilibrium state of a planet's environment are governed by a series of coupled physical, chemical, and biological processes. These processes include atmospheric and surface heating and cooling rates, atmospheric chemistry, impacts and atmospheric escape, the hydrological cycle (oceans, clouds), geological processes such as volcanism, tectonics and weathering, and life processes such as respiration and photosynthesis. (Figure 8).

Figure 8: Planetary Processes: Diagram demonstrating the basic components and planetary processes that contribute to the disk-averaged spectrum of a planet.

The Virtual Planetary Laboratory

To incorporate all these processes, you need a fairly complicated computer model! This model, dubbed the "Virtual Planetary Laboratory" (VPL) is currently under development at the Jet Propulsion Laboratory/California Institute of Technology. The VPL is a suite of computer models that will incorporate all of the components and processes outlined above, to simulate a broad range of planetary environments both with and without life, and to determine the spectral signature of these environments. The interrelationship of the various VPL computer models is shown in Figure 9.

Figure 9: The structure of the Virtual Planetary Laboratory. The suite of radiative transfer, climate, chemical, geological, and biological component models are shown as boxes, and their interactions with each other are shown by the arrows. The information transferred between these component models is labeled at each interface. The order in which these component models are coupled to each other during the course of this proposed work is specified by the Task number. The principal product at each Task development stage is a suite of synthetic spectra which can be used to identify potential biosignatures and derive required capabilities for astronomical instrumentation to observe extrasolar terrestrial planets.

To simulate an alien world, one must first start with a set of "initial conditions" i.e. what we think the planet is like. These initial conditions will include things like the size of the planet, the type of star it orbits, how far away from the star it is, the bulk composition of the atmosphere, whether or not it has plate tectonics, and its surface and trace atmospheric composition. Based on these initial conditions we can model the physics involved in the interaction of the star's radiation with the

atmosphere and surface using a “radiative transfer” model, and model the resulting temperature balance throughout the atmosphere using a “climate” model. Looking at a particular atmospheric temperature balance, we can determine whether it would speed up or slow down chemical reactions at different levels of the atmosphere, and then use a chemical model to determine whether it would appreciably change the composition of the atmosphere by doing this. If it does, then we take the new chemical state and use the radiative transfer and climate models to determine how this new chemical state affects the temperature balance.

As a specific example of how chemical composition and the temperature of an atmosphere are coupled, the production of ozone in the Earth’s stratosphere causes that part of the atmosphere to warm up, because ozone absorbs a lot of UV photons. So by taking the laws of chemistry and physics into account through the atmospheric chemistry and radiative balance, we can iteratively come to a state we call “radiative-convective and chemical equilibrium”. In doing so, we will have created a planet that *might* exist, because at least it hasn’t been disqualified for violating the known laws of physics and chemistry!

The core of the VPL then, is a coupled radiative transfer/climate/chemistry model, which is being assembled using three existing atmospheric models that have already been validated and used to address key scientific problems in planetary and Earth sciences. However, although this core model can self-consistently describe the state of the planet’s atmosphere, we would still have to specify the “boundary conditions”, i.e. what is happening at the surface of the planet and at the boundary where the atmosphere meets outer space. So the coupled-climate-chemistry model will be augmented by interchangeable boundary modules that will characterize material lost and gained at the upper and lower boundaries of a planetary atmosphere, and so will consist of geological, impact influx, atmospheric escape, and life processes.

The integrated VPL will be validated using data derived from terrestrial planets in our own solar system. It will then be used to explore the plausible range of atmospheric compositions and thermal structures, and to generate disk-averaged spectra for extrasolar planets and for early Earth. These models will be run with and without biological processes to improve our understanding of the effects of life on a planet’s atmospheric composition and spectrum. They will also be used to create a spectral catalog that can be used as a statistical sample space to explore the optimum wavelength range, spectral resolution, and instrument sensitivity required to characterize extrasolar terrestrial planets.

While the VPL is still under development, current work on this project includes using a self-consistent coupled climate-chemical model, and a radiative transfer model to explore the change in the spectrum of an Earth-like planet as a function of oxygen abundance in that atmosphere. Figure 10 shows the changes throughout an optical/NIR/MIR spectrum when the oxygen abundance of an Earth-like atmosphere is severely depleted. Note that coupled changes due to the change in composition and the corresponding change in temperature structure produce effects that are more involved than a simple reduction in oxygen or ozone abundance. It is hoped that by exploring a different range of Earth-like planets like these, we will have a better understanding of what to search for with the planned planet detection and characterization missions.

Figure 10: Difference spectrum generated by subtracting the spectrum of an Earth-like planet with the present atmospheric level (PAL) of oxygen, and an Earth-like planet that has one thousandth of PAL. Because the oxygen has been changed in the atmosphere, some of the most obvious differences in the spectra are due to O₂ spectral features. However, the chemical and temperature structure changes induced in the atmosphere by the reduced O₂ can also be seen. For example, changes in CH₄

and N₂O can be seen in the MIR, and the changes in the stratospheric temperature structure result in changes in the band profile for CO₂, strongly affecting appearance of the 15μm band between the two cases.

The Missions

As described in the previous chapter, technology is currently being planned and developed to support missions that will be able to directly detect and characterize extrasolar terrestrial planets. The first generation of these missions, such as the European Space Agency's Darwin mission, and NASA's Terrestrial Planet Finder (TPF) mission, will be able to detect terrestrial planets within 45 light years and perform limited characterization of them with low spectral resolution. TPF will search for atmospheres, and is planned to be able to pick out either the oxygen A-band in the visible, or the 9.6μm ozone band in the mid-infrared, as an indicator of life, along with other features such as water vapor and CO₂, which may denote habitability. The final concept will be chosen for this mission in 2006, and TPF should launch in about 2015. Table 1 gives the anticipated time we will have to look at a planetary system to discover or characterize the planet. As you can see, once we find a terrestrial planet, it is so faint that we will have to stare at it for days or even months to get even a simple spectrum over a small wavelength range.

Table 1: Time Requirements for Various Configurations of TPF to Observe Terrestrial Planets (Table 6.1 from NASA's TPF Book). This table shows the anticipated integration times required to detect important characteristics of a planet, for different configurations of a spacecraft (number of collecting areas and their diameters) that uses nulling interferometry to observe the planet.

A more detailed characterization of extrasolar terrestrial planets will probably have to wait until we fly the successor to the TPF mission, the NASA Life Finder mission. Life Finder has not even been designed yet, but is envisioned to provide enhanced spectral resolution to study in far greater detail the planets found by TPF. Life Finder is envisaged to have the spectral resolution required to identify features as faint as the present day Earth's CH₄ band on an extrasolar planet, as well as providing an unambiguous determination of the planet's atmospheric composition and temperature structure. Life Finder will give us our most definitive answer on whether extrasolar terrestrial planets harbor life. Even if that life is still at the microbial level!

Summary

By 2015, humanity will have the capability to directly detect and characterize extrasolar terrestrial planets. Some of the first questions asked will be "What is the planet like? Is it habitable like our Earth? Does it show any signs of life?" To answer these questions we will have to understand more about the possible range of habitable planets in the Universe, including the many different habitable worlds manifested by the Earth during its 4.6Gyrs of history. The characterization of extrasolar terrestrial planets is therefore an emerging theoretical field that gains a great deal from existing techniques and tools used by the astronomy, Earth-observing and planetary science communities. These include remote-sensing techniques (photometry, atmospheric thermal structure and composition, surface types, clouds, aerosols, etc.) and environmental models, including atmospheric chemistry, climate, carbon cycle, hydrological cycle, and biospheric models. These techniques and models can be used to provide a rigorous scientific basis for studies of the habitability and detection of life in a broad range of plausible planetary environments.

References and Further Reading

- Arnold, L., Gillet, S., Lardiere, O., Riaud, P., and Schneider, G., 2002, *A&A*, 392, 231-237.
- Beichman, C.A., & Velusamy, T., 1999, *ASP Conf. Ser.*, 194, 405.
- Chahine M. T., 1968, *J Opt. Soc. Am.*, 58(12), 1634.
- Conrath B. J., 1972, *J. Atmos. Sci.*, 29(7), 1262.
- Conrath B. J., Gierasch P. J., & Ustinov E. A., 1998, *Icarus*, 135(2), 501.
- DesMarais D., and 10 others, 2001, *Biosignatures and Planetary Properties to be Investigated by the TPF Mission*, JPL Publication 01-008.
- Elachi, C., 1987, *Introduction to the Physics and Techniques of Remote Sensing*, John Wiley and Sons, New York.
- Ford E. B., Seager S., & Turner E. L., 2001, *Nature*, 412(6850), 885.
- Holland H.D., Beukes, N. J., 1990, *AM J SCI*, 290A, 1.
- Kasting, J. F., & Donahue, T. M. 1980, *J. Geophys. Res.* - OC, 85(NC6), 3255-3263.
- Lovelock J. E., & Margulis L., 1974, *Origins Life Evol.*, B5(1-2), 93.
- Meadows, V. S., 2001, <http://nai.arc.nasa.gov/library/downloads/II.pdf>
- Meadows, V. S., & Crisp, D., 1996, *J. Geophys. Res.-Planet*, 101(E2), 4595.
- Rodgers, C. D., 1976, *Rev. Geophys.*, 14(4), 609.
- Sagan C., Thompson W.R., Carlson R., Gurnett D., & Hord C., 1993, *Nature*, 365(6448), 715.
- Santee, M., & Crisp, D., 1993, *J. Geophys. Res.-Planet*, 98(E2), 3261.
- Short, N., 1982, *The Landsat Tutorial Workbook: Basics of Satellite Remote Sensing*, NASA Ref. Publ. 1078, U.S. Gov. Printing Office, Washington, D.C.
- Smith P.H., Lemmon M.T., Lorenz R.D., 1996, *Icarus*, 119(2), 336.

Figure 1

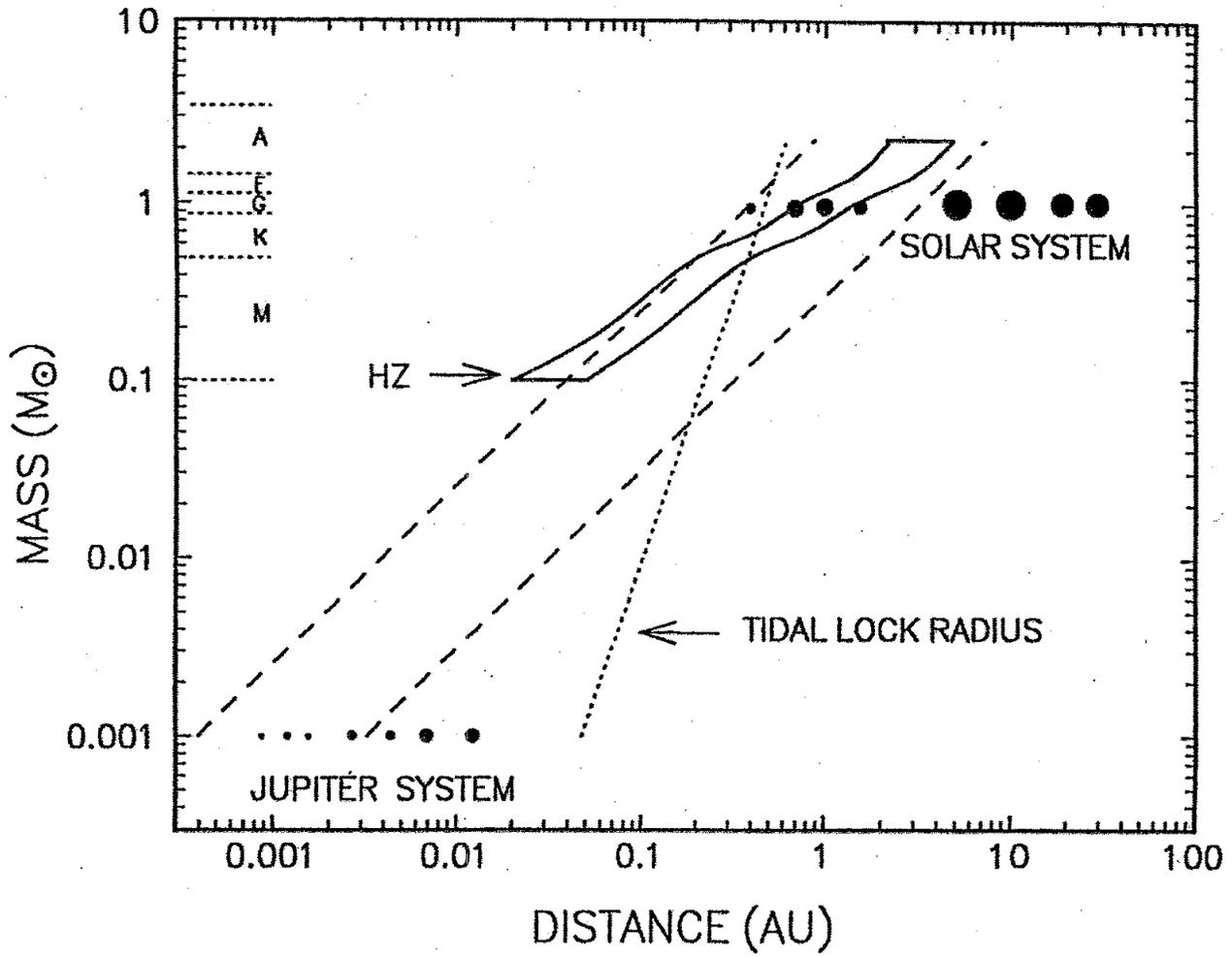


Figure 2

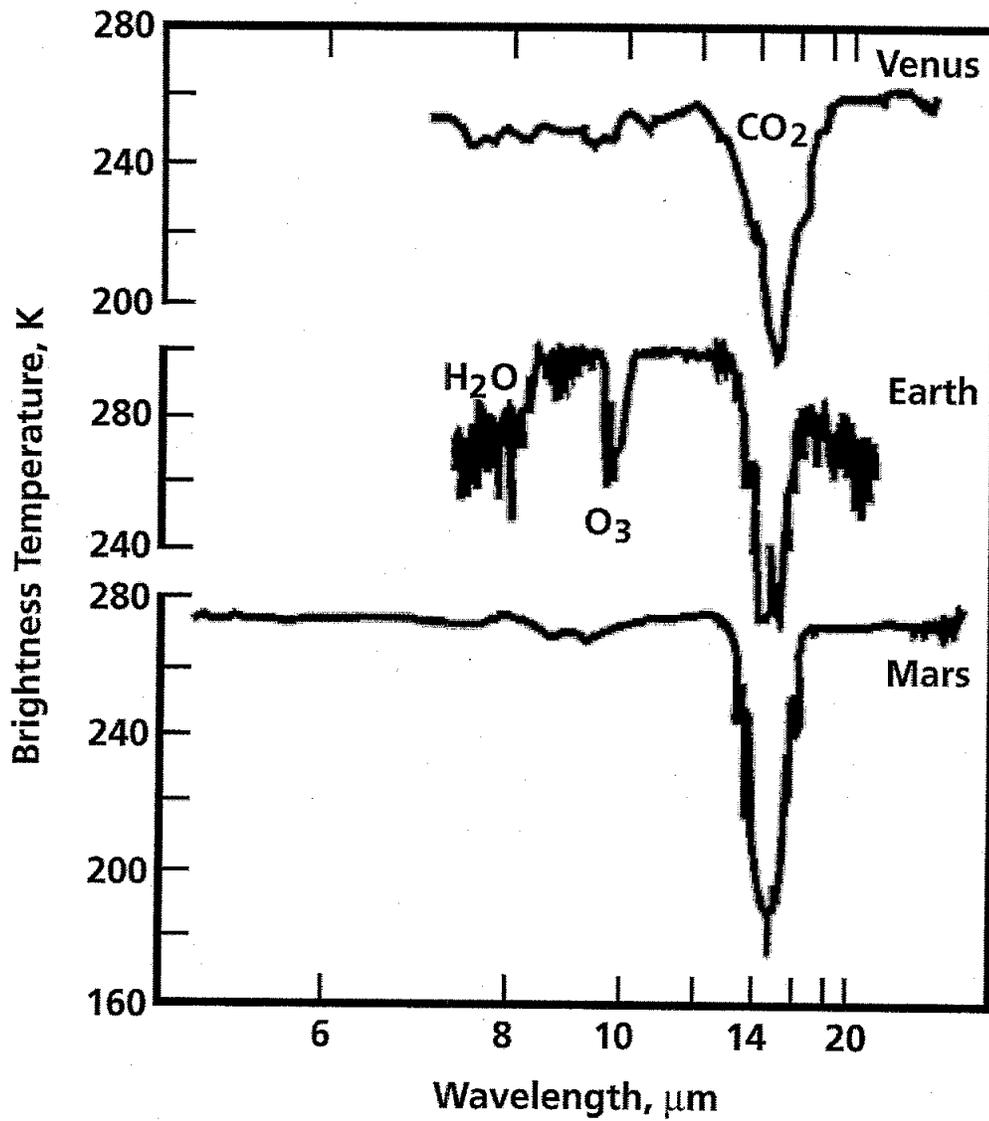


Fig 3

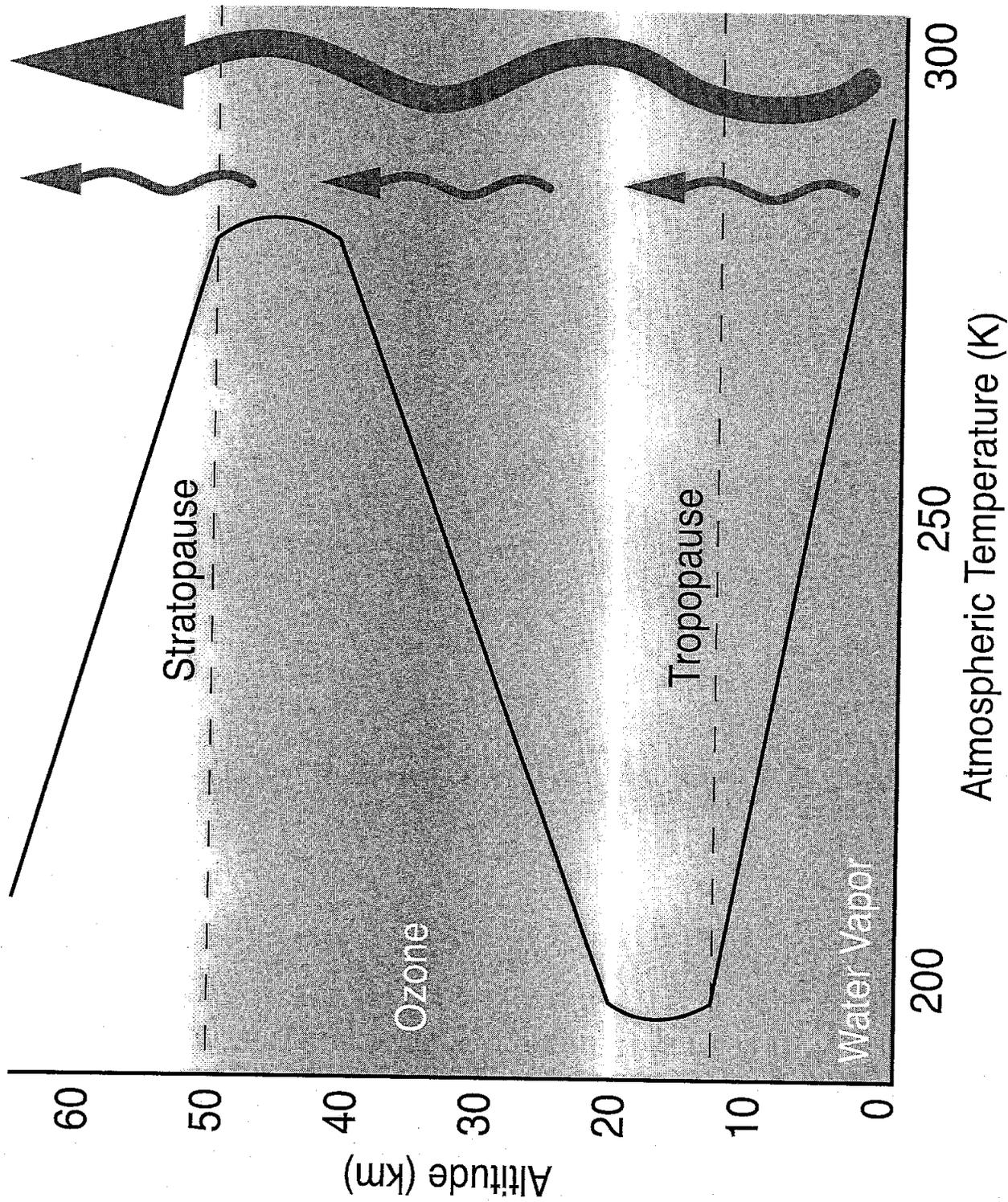


Figure 4

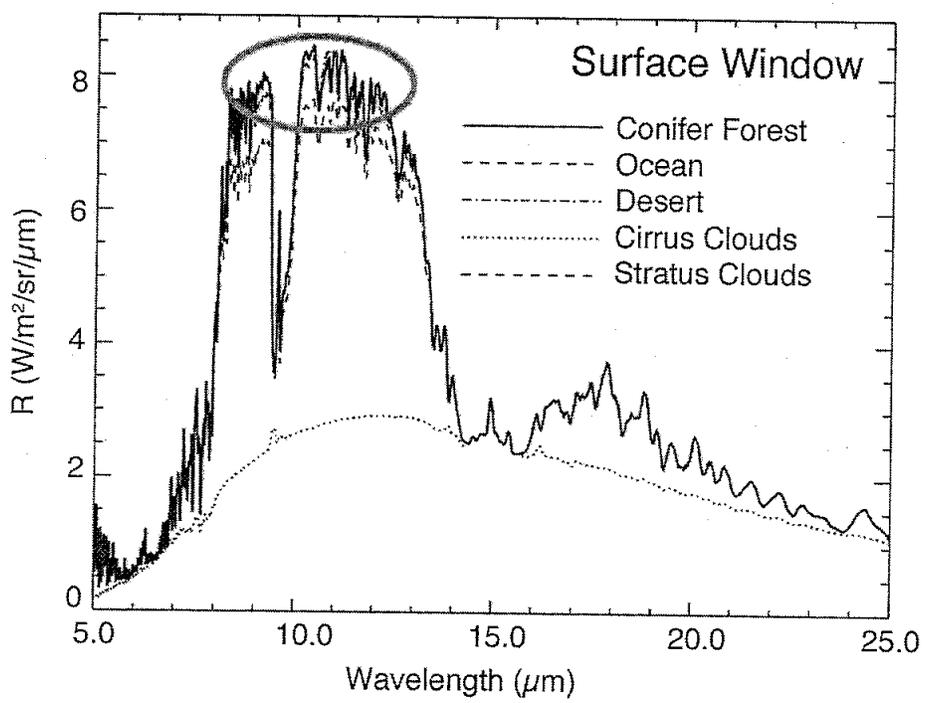


Figure 5

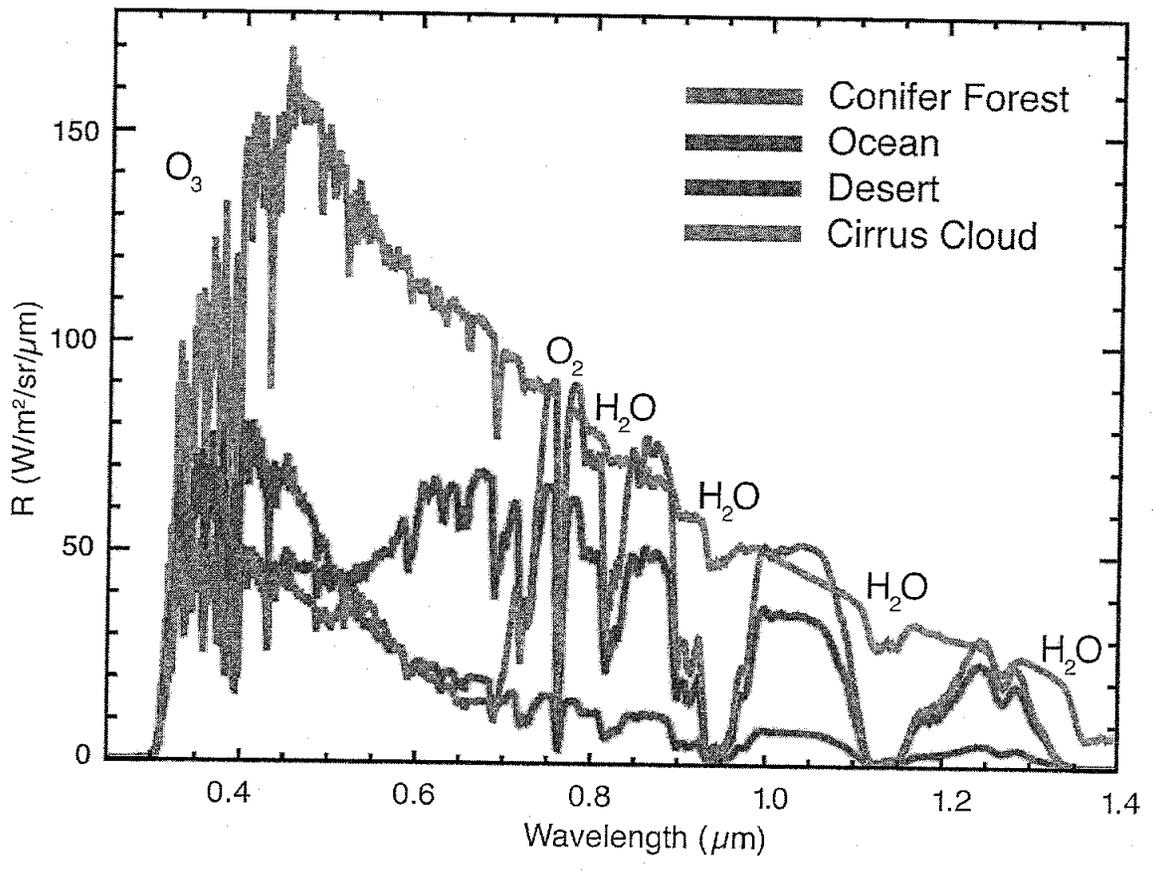
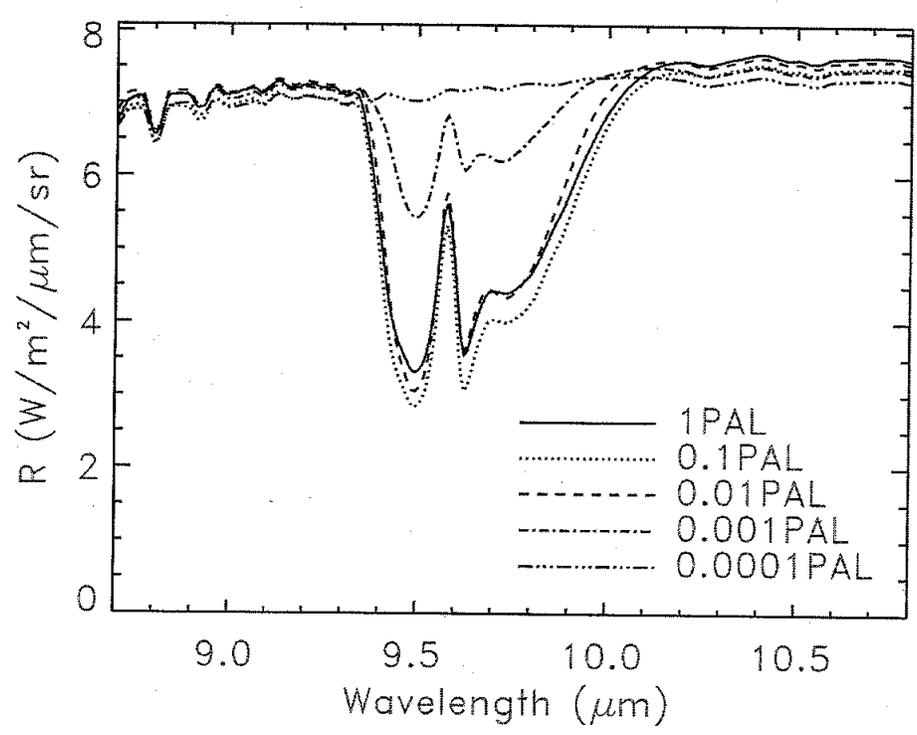


Figure 6.



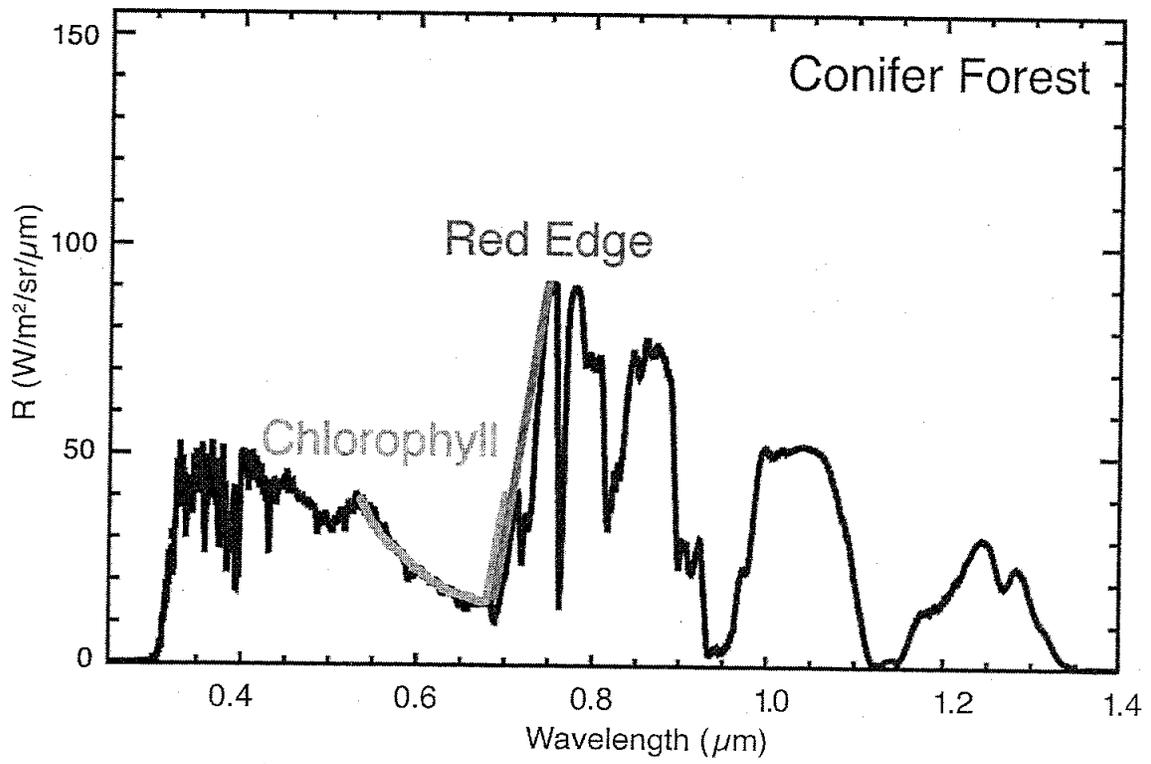


Fig 8

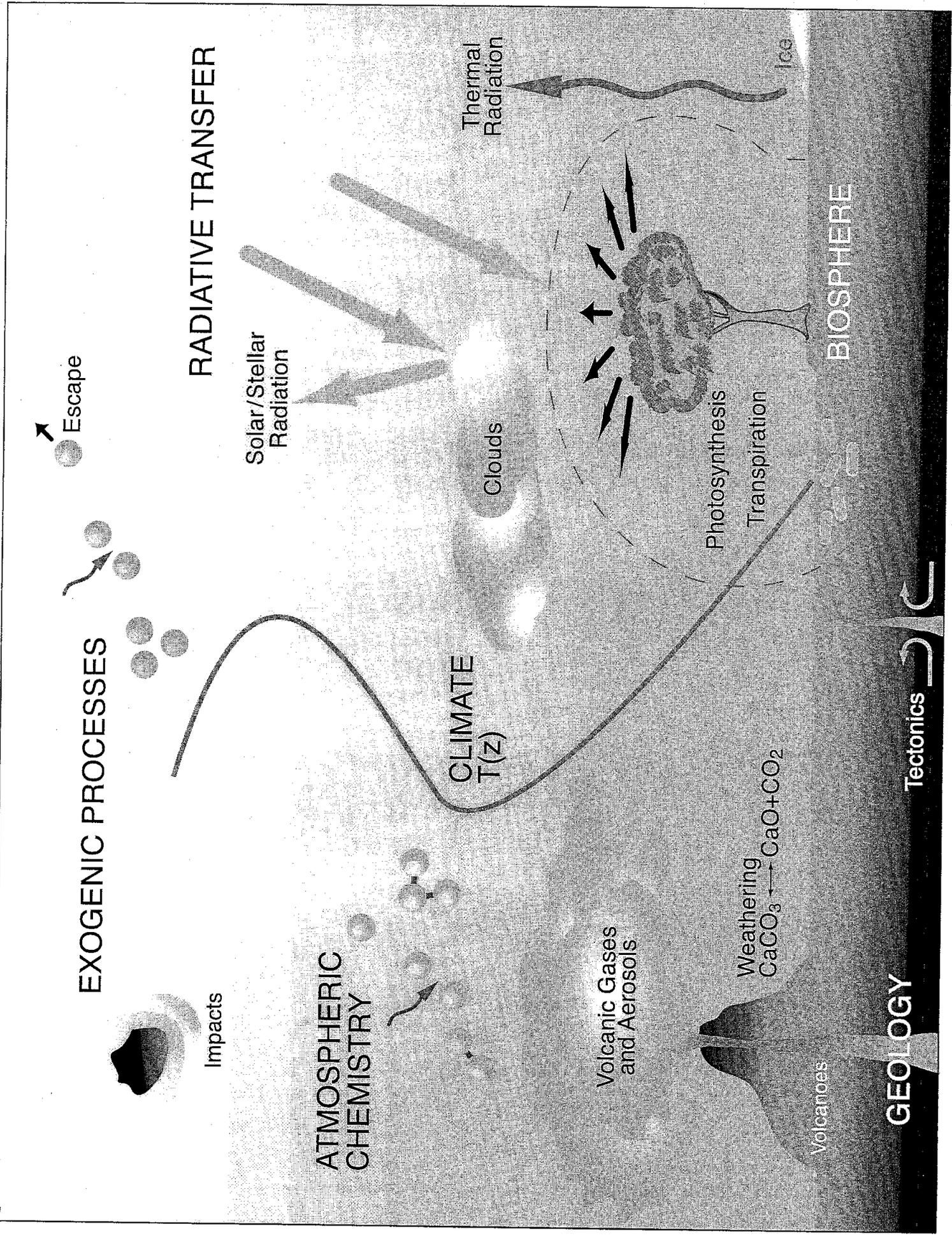


Fig 10

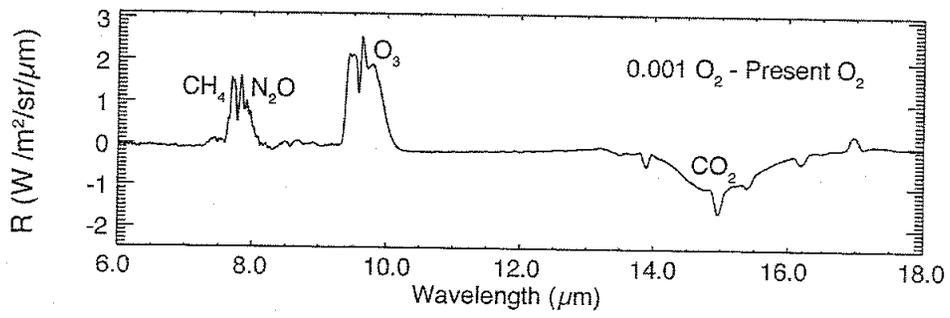
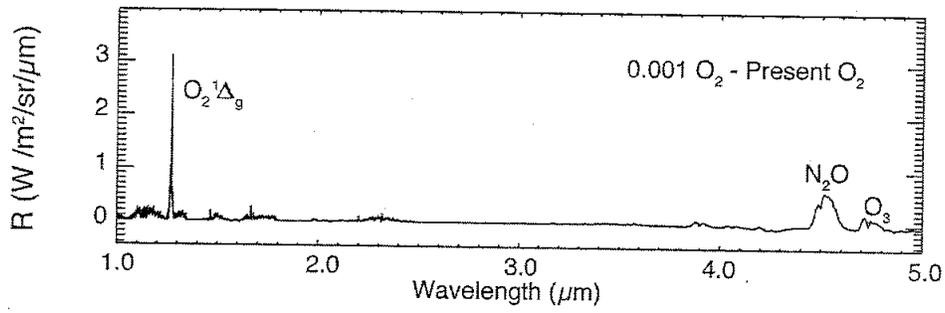
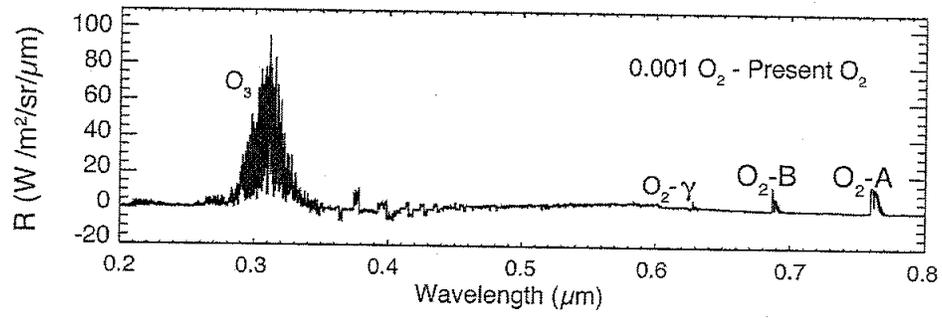


Table 1

Time Requirements for Various Configurations of TPF to Observe Terrestrial Planets

Science Goal	12 μ m observation of an Earth at 10pc	4x2 m (5 AU)	4x.085 m (1 AU)	4x2 m (1 AU)	4x2.7 m (1 AU)	4x3.5 m (1 AU)
Detect Planet	Spectral Resolution R=3	1.4 hr	470 hr	15.3 hr	5.1 hr	2.0 hr
	Signal to Noise SNR=5					
Detect Atmosphere CO ₂ , H ₂ O	R=20/SNR=10	2.4 day	-	18.1 day	5.9 day	2.3 day
Habitable? Life? O ₃ , CH ₄	R=20/SNR=25	15.0 day	-	-	-	14.7 day