Characterizing Exoplanet Habitability
Astro2020 Science White Paper

Thematic Area: Planetary Systems


1Lead Author; Northern Arizona University; tyler.robinson@nau.edu; 928-523-0350,
2University of Washington, 3NASA Goddard Space Flight Center, 4Planetary Science Institute, 5New York University Abu Dhabi, 6Carnegie Institution for Science, 7NASA Ames Research Center, 8Arizona State University, 9SETI Institute, 10University of Arizona, 11The University of Texas at Austin, 12Subura Telescope, 13University of Crete, 14Princeton University, 15University of California, Santa Cruz, 16University of California, Berkeley, 17The Ohio State University, 18NASA Exoplanet Science Institute, IPAC, Caltech, 19University of Hawaii, 20Jet Propulsion Laboratory, California Institute of Technology, 21University of Vienna, 22University of California, Riverside, 23Leiden Observatory, 24University of Chicago, 25Johns Hopkins University Applied Physics Laboratory, 26Harvard-Smithsonian Center for Astrophysics, 27University of Heidelberg, 28National Research Council Canada, 29Earth-Life Science Institute, 30W.M. Keck Observatory, 31Georgia Institute of Technology, 32University of Oxford, 33Massachusetts Institute of Technology, 34Goddard Institute for Space Studies / Columbia University, 35Space Telescope Science Institute, 36Vanderbilt University, 37University of Toronto, 38The Adler Planetarium
1 Introduction

Recent observations have uncovered a handful of Earth-sized exoplanets that are known to orbit within the habitable zone of their host star (e.g., Quintana et al., 2014; Anglada-Escudé et al., 2016; Gillon et al., 2017) — implying that, with the right atmospheric composition, these worlds could maintain liquid water on their surface (i.e., be habitable). In the near future, characterizing exoplanets for signs of habitability will provide the first observational tests for the prevalence of truly habitable worlds, and will also enhance our ability to interpret signs of life. In fact, habitability and biosignature investigations must go forward hand-in-hand if we are to know how rare (or common) the origin of life may be. Should the search for life beyond Earth find many habitable worlds and few inhabited ones, we would begin to understand how uncommon the origin of life was for our planet. Conversely, if our hunt yields many worlds that are both habitable and inhabited, then we would learn that life is a relatively common consequence of habitability.

Alongside answering fundamental questions about life’s origins, observational studies aimed at characterizing exoplanets for signs of habitability can address other basic questions in planetary science and astrobiology. For example, given the extended pre-main sequence lifetimes of M dwarfs, is it even possible for cool stars to host habitable worlds (Tian and Ida, 2015; Ramirez and Kaltenegger, 2014; Luger and Barnes, 2015)? What are the true limits of the N$_2$-CO$_2$-H$_2$O habitable zone (Kasting et al., 1993; Kopparapu et al., 2013), and can alternative atmospheric compositions lead to habitable worlds far outside this zone (Pierrehumbert and Gaidos, 2011; Ramirez and Kaltenegger, 2018)?

2 Detecting and Constraining Habitability

Habitability indicators fall into two broad categories — direct and indirect. The former category includes the direct detection of surface liquid water, thereby demonstrating liquid water stability and planetary surface habitability. Here, specular reflectance from surface liquid water — leading to “glint” reflectance and a polarization signature (especially for scattering near the Brewster angle) — can indicate surface habitability (McCullough, 2006; Williams and Gaidos, 2008). A final direct constraint on surface habitability relies on remotely characterizing both the surface temperature ($T$) and surface pressure ($p$) for a rocky exoplanet, which then indicates habitability via a comparison to the water phase diagram.

Indirect habitability indicators are atmospheric and/or surface characteristics that are likely associated with habitable environments, but which do not directly demonstrate that a given exoplanet is habitable. Foremost amongst these indicators is the presence of atmospheric water vapor (which does not prove habitability as water vapor is expected to be a component of many terrestrial and non-terrestrial atmospheres). Additionally, and excepting the inner-most edge of the habitable zone, the presence of atmospheric carbon dioxide can also signal a habitable world. Water clouds are also likely to be common across habitable worlds, although remotely constraining cloud properties is challenging (e.g., Feng et al., 2018). Mapping of heterogeneous exoplanetary surfaces from photometric lightcurves might also reveal structure due to contrasting oceans and continents (Cowan et al., 2009, 2011). Finally, a stand-alone measurement of surface pressure or temperature can help argue for (or against) habitability (again, via comparison to a water phase diagram). Aspects of many of these constraints were first explored using Galileo flyby data for Earth in Sagan et al. (1993).
In reality, direct habitability indicators may suffer from measurement uncertainties or ambiguities in interpretation. Here, a habitability assessment could be strengthened via combined measurements of several direct and indirect habitability indicators. With sufficient constraints on the atmospheric and surface environment of a terrestrial exoplanet, a probabilistic argument for (or against) habitability could then be derived.

3 Requisite Observing Capabilities

The paragraphs below highlight key direct and indirect habitability indicators. For each, a summary of the relevant observables and associated observational approaches (transit spectroscopy, secondary eclipse spectroscopy, and reflected-light direct imaging) is provided. Requisite wavelength range(s), spectral resolving power ($\mathcal{R} = \lambda/\Delta\lambda$, where $\lambda$ is wavelength), and signal-to-noise ratio (SNR) for detection are also given and are summarized in Table 1.

Glint: The reflectivity of a water-bearing planet with partial cloud coverage (like Earth) increases dramatically towards crescent phases due to specular reflectance from surface oceans. Not only do the oceans become more reflective at glancing solar incidence angles, but the relative size of the glint spot on the crescent sliver also increases (Williams and Gaidos, 2008). The result, for Earth, is a dramatic multi-fold increase in reflectivity for red and near-infrared wavelengths (where Rayleigh scattering does not preclude surface sensitivity; Figure 1). Due to this significant increase in brightness as compared to a non-glinting planet, and because glint affects continuum wavelengths, a crescent-phase glint signature can be detected (at $2\sigma$) using photometry or low-resolution spectroscopy with a characteristic continuum SNR of $\gtrsim 6$ at star-planet-observer (i.e., phase) angles ($\alpha$) between 135–165° (Robinson et al., 2010). Glint detections are relevant to exoplanet direct imaging in reflected light, where only a subset ($\sim 70\%$ at $\alpha = 135^\circ$) of directly imaged exoplanets would have appropriately inclined orbits such that the crescent phases are accessible. Additionally, observations of a planet at crescent phase require access to a smaller projected separation between planet and star compared to quadrature phase observations (e.g., 0.71 AU for an Earth-Sun analog observed at $\alpha = 135^\circ$, corresponding to 71 mas at 10 pc), which then relates to inner working angle requirements for high contrast imaging.

Polarization: Scattering of sunlight from a liquid surface viewed at/near the Brewster angle leads to light polarization. For an Earth twin this polarization signature would be maximized at a phase angle near 110°. This is distinct from the maxima due to Rayleigh scattering (at 90°) and water cloud scattering “rainbows” (at 40°; Bailey, 2007; Stam, 2008; Zugger et al., 2010). The polarization effect due scattering from surface oceans is also least contaminated by polarized Rayleigh scattering at red and near-infrared wavelengths (Zugger et al., 2011). Using polarimetric observations of reflected Earth-light from the observable portion of the unilluminated Moon (i.e., Earthshine), Sterzik et al. (2019) demonstrated a polarization maximum of 10–20% for Earth, broadly centered near $\alpha = 110$–120°. Detecting this signature (at $2\sigma$) would require that the individual perpendicular and parallel polarized fluxes be measured at SNRs of 15–30 (Robinson, 2018). While not as strict as for glint, measuring a polarization curve for a distant Earth twin would place requirements on instrument IWAs and would be largely unaffected by orbital inclination limitations. Using aperture polarimetry to measure the combined light from the planet and star could reveal planetary
Table 1: Key Habitability Observables and Constraints for Earth Twins

<table>
<thead>
<tr>
<th>Observable</th>
<th>Technique$^1$</th>
<th>Wavelength</th>
<th>Noise Req.</th>
<th>Add’l Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>glint</td>
<td>direct imaging</td>
<td>0.7–2 µm</td>
<td>SNR $\gtrsim$ 6</td>
<td>broadband; IWA</td>
</tr>
<tr>
<td>polarization</td>
<td>direct imaging</td>
<td>0.7–2 µm</td>
<td>SNR $\gtrsim$ 15–30</td>
<td>broadband</td>
</tr>
<tr>
<td></td>
<td>aperture</td>
<td>5–20 µm</td>
<td>0.01–0.1 ppm</td>
<td>noise characterization</td>
</tr>
<tr>
<td>mapping</td>
<td>direct imaging</td>
<td>0.4–1 µm</td>
<td>SNR $\gtrsim$ 6</td>
<td>broadband; 1 hr cadence; IWA</td>
</tr>
<tr>
<td>surface $p$ &amp; $T$</td>
<td>direct imaging</td>
<td>0.4–1 µm</td>
<td>SNR $\gtrsim$ 5</td>
<td>$\mathcal{R} \gtrsim$ 70; no $T$ constraint</td>
</tr>
<tr>
<td></td>
<td>secondary eclipse</td>
<td>5–20 µm</td>
<td>0.5(3) ppm</td>
<td>M5V(M9V); $\mathcal{R} \gtrsim$ 20</td>
</tr>
<tr>
<td>H$_2$O &amp; CO$_2$</td>
<td>transit</td>
<td>1–5 µm</td>
<td>$&lt; 10(60)$ ppm</td>
<td>M5V(M9V); $\mathcal{R} \gtrsim$ 40</td>
</tr>
<tr>
<td></td>
<td>transit</td>
<td>5–20 µm</td>
<td>$&lt; 5(30)$ ppm</td>
<td>M5V(M9V); $\mathcal{R} \gtrsim$ 40</td>
</tr>
</tbody>
</table>

$^1$Direct imaging emphasizes Sun-like hosts at $\sim 10^{-10}$ contrast. Secondary eclipse and transit emphasize M dwarf hosts.

polarization effects, but would require sub-ppm sensitivity (Kopparla et al., 2018) and good characterization of the star and other noise sources (Bott et al., 2018).

**Mapping:** Modulations in photometric reflected-light time series observations induced by planetary rotation offer an opportunity to study the heterogeneity of a planetary surface and/or cloud distribution (Ford et al., 2001). Fitting such data at multiple wavelengths could reveal a longitudinal color map of the rotating planet, which could indicate distributions of continents and oceans (Cowan et al., 2009). Such observations require a direct imaging mission capable of achieving a temporal cadence markedly shorter than the planetary rotational period (e.g., a cadence of $\sim$1 hr for Earth’s 24 hr rotational period). Recently, Lustig-Yaeger et al. (2018) combined the rotational mapping and glint techniques to show that mapping at crescent phases can increase the robustness of ocean detection by identifying a glinting component of the planet and a non-glinting component of the planet (i.e., diffusely/Lambertian scattering continents). This method has the combined observational requirements of glint detection and rotational mapping, requiring SNR $\gtrsim$ 6 on $\sim$1 hr cadence photometric time series observations at red or near-infrared wavelengths for at least one rotation of the planet at both crescent and quadrature phases. The need for modest-SNR photometry within a short exposure time impacts the expected mapping yield as a function of aperture size, which is explored by Lustig-Yaeger et al. (2018). Finally, observations of orbit-resolved infrared phase curves for rotationally-locked planets can reveal maps of atmospheric thermal structure and clouds (e.g., Stevenson et al., 2017) but, to achieve a sufficient contrast ratio, this approach would be limited to Earth-like planets orbiting M dwarf hosts.
Surface $p$ and $T$: Measured together surface pressure and temperature directly indicate surface habitability and, measured on their own, a pressure or temperature markedly below the water triple point (at 611 Pa and 273 K) would exclude surface habitability. Remote sensing of the surface environment for Worlds with substantial fractional cloud coverage (like Earth) is made more challenging by the blending of cloudy and cloud-free regions of the planetary disk in unresolved observations, implying clouds can significantly impact constraints on surface pressure and temperature. The widths of pressure-broadened gas bands, a Rayleigh scattering slope, and pressure-induced absorption features can all yield constraints on surface pressure (provided the planetary surface is not fully obscured). Thermal emission (e.g., at secondary eclipse) is directly sensitive to temperature, which can also be constrained using the sensitivity of transmission spectra to the atmospheric scale height for transiting exoplanets. Here, signal sizes typically limit this approach to Earth-like planets orbiting M dwarf hosts, although, even here, transit spectra have extremely limited surface sensitivity due to clouds, gas opacity, and refraction (Robinson, 2018; Misra et al., 2014). For simulated visible wavelength (0.4–1.0 µm) direct imaging observations of a realistically cloudy Earth in reflected light at $R = 70$ and 140, Feng et al. (2018) could constrain the retrieved surface pressure to be above the triple point (at $> 2\sigma$) for a V-band SNR of $\geq 5$. Here, temperature cannot be directly constrained. Using simulated thermal infrared (5–20 µm) observations of a cloud-free Earth at $R = 20$, von Paris et al. (2013) could retrieve the surface pressure to be above the triple point pressure (at $> 2\sigma$) for SNRs of 10–20. Surface temperature was more challenging to remotely characterize, requiring an SNR of 20 to to be constrained above the freezing point of water at $2\sigma$ (i.e., $295 \pm 7$ K). (Note that the absolute uncertainty on temperature can be larger for warmer planets while still confidently constraining the surface to be above 273 K.) These SNRs of 10/20 can be translated into secondary eclipse noise requirement of 1/0.5 ppm at 10 µm for an Earth twin orbiting an M5V host, and 6/3 ppm for an M9V host — although further improved sensitivities are likely required when considering realistically cloudy Earth-like planets. Finally, mid-infrared high contrast imaging from
ground-based ELTs also has the potential to detect 10 µm thermal emission from Earth-like exoplanets for a handful of nearby Sun-like stars (see white paper by Currie et al.).

**Water Vapor and Carbon Dioxide:** A non-detection of atmospheric water vapor using direct imaging or secondary eclipse techniques (which are more sensitive to the deep atmosphere) would be potential evidence against surface habitability. Conversely, a detection of strong water vapor absorption for a transiting potentially Earth-like planet viewed in transmission (which is more sensitive to the upper atmosphere) may argue against habitability, as this could indicate a world in a moist/runaway greenhouse state that is undergoing water photolysis and atmospheric loss (Lincowski et al., 2018). In all cases, the detection of carbon dioxide for a potential exo-Earth could help constrain the greenhouse effect and, beyond the inner edge of the habitable zone, would be at least consistent with the habitability-stabilizing carbonate-silicate cycle (Walker et al., 1981). Note, though, that the carbonate-silicate cycle has requirements beyond atmospheric carbon dioxide and spectra of many non-habitable terrestrial planets are also expected to show spectral features due to CO₂, implying care must be taken when interpreting carbon dioxide as an indirect habitability indicator. The aforementioned study from Feng et al. (2018) showed water vapor detections at low SNRs (i.e., 5), and constraints on the water vapor mixing ratio to better than an order of magnitude (at 1σ) with $R \geq 140$ and V-band SNR $\gtrsim 15–20$. (Similar constraints could likely be achieved at lower spectral resolutions, as the water vapor bands in the 0.4–1.0 µm are broad.) A pair of relatively weak carbon dioxide features are apparent in Earth’s reflected light spectrum near 1.59 µm, and the shape/depth of these combined features indicate that measurement would require a spectral resolution of at least 11 (Des Marais et al., 2002) and an SNR (in the nearby continuum) of at least 10. For transit spectra of an Earth twin orbiting an M5V(M9V) host, the water vapor bands at 1.9, 2.6, and 6.3 µm rise 10(60), 10(60), and 5(30) ppm above the continuum, respectively (Figure 1). Note that, of these bands, the 1.9 µm band is narrowest, arguing for a spectral resolution of at least 40. Similarly, the 4.3 and 15 µm carbon dioxide bands both rise roughly 25(160) ppm above their surrounding continua. Taken altogether, these argue for a requisite sensitivity of better than 10(60) ppm in the near-infrared, and 5(30) ppm in the mid-infrared, for M5V(M9V) hosts.

**4 Conclusions**

- Characterizing habitable exoplanets is a critical, aspirational goal for astronomy and astrophysics. Excitingly, transiting and directly imaged Earth-like exoplanets provide a range of observables that could probabilistically constrain surface habitability.

- Key observables include glint and polarization signatures from ocean specular reflectance, surface maps, surface pressure and temperature, and atmospheric water vapor and carbon dioxide. Key observational requirements are given in Table 1.

- The diversity of habitability indicators span many observational approaches with challenging observational needs, highlighting next generation space telescopes as the key to characterizing habitable environments outside our Solar System.
References


