Potential Habitability as a Stellar Property: Effects of Model Uncertainties and Measurement Precision

Noah W. Tuchow and Jason T. Wright

Department of Astronomy & Astrophysics and Center for Exoplanets and Habitable Worlds and Penn State Extraterrestrial Intelligence Center, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

Received 2021 December 5; revised 2022 March 29; accepted 2022 April 4; published 2022 May 5

Abstract

Knowledge of a star’s evolutionary history combined with estimates of planet occurrence rates allows one to infer whether a star would be a good target in a search for biosignatures, and to quantify this intuition using long-term habitability metrics. In this study, we analyze the sensitivity of the biosignature yield metrics formulated by Tuchow & Wright to uncertainties in observable stellar properties and to model uncertainties. We characterize the uncertainties present in fitting models to stellar observations by generating a stellar model with known properties and adding synthetic uncertainties in the observable properties. We scale the uncertainty in individual observables and observe the effects on the precision of properties such as stellar mass, age, and our metrics. To determine model uncertainties, we compare four well-accepted stellar models using different model physics and see how they vary in terms of the values of our metrics. We determine the ability of future missions to rank target stars according to these metrics, given the current precision to which host star properties can be measured. We show that obtaining independent age constraints decreases both the model and systematic uncertainties in determining these metrics and is the most powerful way to improve assessments of the long-term habitability of planets around low-mass stars.

Unified Astronomy Thesaurus concepts: Exoplanets (498); Exoplanet evolution (491); Astrobiology (74); Biosignatures (2018); Direct imaging (387); Planet hosting stars (1242); Habitable zone (696); Stellar properties (1624)

1. Introduction

In the coming decade, astronomers strive to develop instruments with the sensitivity to search for biosignatures in the atmospheres of earthlike planets. One of the top recommendations of the 2020 Decadal Survey on Astronomy is the construction of a large IR/Optical/UV space telescope capable of directly imaging Earth-sized planets in the habitable zones of their stars, and to obtain spectra to infer their atmospheric compositions (National Academies of Sciences & Medicine 2021). In planning for such a mission, astronomers will need to select target stars to survey in order to search for planets. For direct imaging missions, only nearby bright stars would be viable candidates to host planets that are luminous enough for the instruments to obtain high enough resolution spectra. Furthermore, after planet candidates are discovered, future missions will need to prioritize which of these targets to investigate for follow-up observations.

Planning for a future mission to detect biosignatures is inherently a very challenging prospect. Even if next-generation telescopes are able to obtain sufficiently precise spectra for Earth-sized planets, it is nontrivial to infer whether a planet hosts life based on its chemical composition. In Earth’s atmosphere, molecules such as O\(_2\) and O\(_3\) are among the strongest biosignatures, and the detection of these chemical species in the atmospheres of other planets could suggest the presence of photosynthetic life (Kiang et al. 2018). However, in several cases, biosignatures such as oxygen can be ambiguous and could potentially originate from abiotic sources (Meadows et al. 2018). To determine whether a potential biosignature constitutes evidence for life requires an understanding of its interaction with other chemical species in the atmosphere and any abiotic sources and sinks present in the environment. Due to the potential ambiguity of detections of biosignatures in the spectra of individual planets, an alternative strategy is to take a statistical comparative planetology approach and search for overarching signs of life for a population of exoplanets (Bean et al. 2017). One example of a statistical biosignature is the planetary age–oxygen correlation proposed by Bixel & Apai (2020), where if one assumes that atmospheric oxygen concentrations increase in time, as they did on Earth after the emergence of photosynthetic life, older exoplanets with life should exhibit higher oxygen concentrations. Future direct imaging missions aim to measure the spectra of habitable zone exoplanets, and obtain a large enough sample to allow astronomers to test hypotheses about the population.

The Decadal Survey’s proposed large IR/O/UV observatory is influenced by LUVIOR and HabEX, the two direct imaging mission concepts proposed to the decadal survey (The LUVIOR Team 2019; Gaudi et al. 2020). These mission concepts proposed observation strategies that focused on obtaining the largest yield of spectra of habitable zone Earth-sized planets. They would select target stars for a future mission based on a variety of factors such as whether stars are known planet hosts, the amount of exozodiical dust, and the stellar multiplicity. While all of these properties are certainly essential in determining where one could obtain the spectra of earthlike exoplanets, if one of the goals of a future direct imaging mission is to search for biosignatures, then target prioritization should also incorporate our prior knowledge for where we would expect life to emerge and develop.
While usually not explicitly stated, astronomers have a qualitative understanding of which stars would be the best candidates to host planets with biosignatures. Sunlike stars, as well as other FGK stars on the main sequence are often preferred, as we know for a fact that life developed on Earth around a main-sequence G star. Our understanding of the evolution of life on Earth suggests that older stars may be better candidates, as complex life took billions of years to develop on Earth, and planets around older stars would have had more time to develop biosignatures. Other types of stars are often excluded due to their potential obstacles to habitability. These include massive stars, which have too short lifetimes to allow for the development of complex life on earthlike timescales, and giant evolved stars that evolve too quickly to allow their planets to be habitable for an extended duration. Furthermore, neutron stars and white dwarfs may be unlikely to host life on their planets as the dramatic end phases of stellar evolution are likely to sterilize planets of life and may lead to the complete destruction of planets. To quantify this qualitative understanding, Tuchow & Wright (2020) developed a mathematical framework to determine which stars would be the strongest candidates in a search for biosignatures.

In this paper, we will directly continue the earlier work of Tuchow & Wright (2020) and make use of their mathematical framework. This framework assesses which stars have the highest probability of hosting planets that have been in the habitable zone for a sufficient duration to allow for the development of complex life and biosignatures. Tuchow and Wright defined a series of biosignature yield metrics, $B$, based on the evolution of a star’s habitable zone in time, which relies on stellar effective temperature and luminosity. Thus for any stellar model that produces an evolutionary track and for any habitable zone formulation, one can compute a metric for biosignature yield in the form

$$B = \int H(a, t)\Gamma(a, R_p)da\,dR_p, \quad (1)$$

The formulation of these metrics $B$ depends on two functions: $H(a, t)$ and $\Gamma(a, R_p)$. $H(a, t)$ is defined as the probability that a planet of age $t$ at a distance $a$ from its star hosts biosignatures. Lacking any empirical constraints on the emergence of biosignatures, $H$ encapsulates our prior knowledge and assumptions about the emergence of life and biosignature on planets around other stars. The function $\Gamma$ in Equation (1) is the distribution of habitable planets in orbital distance. Similar to $H$, it is poorly constrained in the habitable zones of sunlike stars, though future missions aim to obtain better constraints for $\Gamma$. Therefore our knowledge of the distribution of planets in the habitable zone relies on either extrapolation to larger orbital distances (Petigura et al. 2013), or inferences based on limited data (Burke et al. 2015). Since $\Gamma$ and $H$ are not tightly constrained, there are several reasonable assumptions for both of them, comprising a family of $B$ metrics for different combinations of forms of $\Gamma$ and $H$.

The Tuchow & Wright (2020) study tested various assumptions for $\Gamma$ and $H$ and determined which populations of stars were preferred by each. This study suggested that the most physical assumptions for $H$ depend on the time that planets spend in the habitable zone, which they define as the habitable duration, $\tau$. When considering the evolution of habitable zones in time and the time spent in the habitable zone, $\tau$, one encounters the open question as to whether planets that form outside of habitable zone and enter it as their host star evolves can actually be considered habitable. Tuchow & Wright (2021) refer to the region of the habitable zone occupied by these planets as the Belatedly Habitable Zone (BHZ) and stress that the habitability of these planets remains unknown. BHZ planets may face greater obstacles to habitability than those that have remained in the habitable zone from their formation. For instance, planets forming interior to the habitable zone will likely have water loss and desiccated atmospheres by the time they enter the habitable zone (Luger & Barnes 2015), whereas planets forming exterior to the habitable zone will be globally glaciated with high ice albedos, making them difficult to thaw without directly transitioning to a runaway greenhouse state (Yang et al. 2017).

Since BHZ planets will make up a large fraction of planets around low-mass stars and older stars, determining whether they can be habitable is a rich area of future research. If BHZ planets are unable to support habitability, then long-term habitability would be limited to planets within the continuously habitable zone (CHZ), as defined in Tuchow & Wright (2020). Note that previous studies, such as Kasting et al. (1993) and Tuchow & Wright (2020), referred to the class of planets that form exterior to habitable zone and enter it as their host star becomes more luminous as “cold start” planets. However, since the term “cold start” has often been used in the context of giant planet formation, we shall instead use the term Outer BHZ Planets to refer to them. In this study, we only consider BHZ planets that enter the habitable zone due to the evolution of their host stars, but we note that migrating planets will face similar obstacles to habitability, albeit with incident stellar fluxes changing on different timescales. For further discussion of the formulation of relative biosignature yield metrics, as well as additional background on exoplanet direct imaging and the evolution of habitable zones in time, we direct the reader to Sections 1 and 2 of Tuchow & Wright (2020).

Tuchow & Wright (2020) demonstrated how biosignature yield metrics changed as a function of stellar mass and age. Now we would like to investigate how precisions in these metrics depend on the uncertainties on stellar parameters for a typical star. A given stellar model can be thought of in the form $y = f(p, \theta)$. Here $p$ is a set of fundamental stellar input parameters such as mass, age, and metallicity, $\theta$ are a variety of parameters that govern the model physics, and $y$ are the model outputs. Usually, stellar mass, age, and other properties in $p$ are not directly observable. Instead, stellar models are typical fit to observed output properties $y$, such as apparent magnitudes, effective temperatures, surface gravities, surface metallicities, and parallax measurements. Some stars may have measurements of additional observable properties such as asteroseismic modes of oscillation, interferometric radii, dynamical masses from binaries, and gyrochronological ages inferred from stellar rotation periods. These additional properties allow stellar models to be better constrained, and future measurements of these observables may allow astronomers to obtain more precise estimates of the fundamental properties of target stars.

When determining input properties $p$ and derived properties, such as $B$ metrics, via fitting stellar models to observables, there are two main sources of uncertainty: systematic and model uncertainties. Systematic uncertainties arise from the fact that all observable properties have some measurement uncertainty, while model uncertainties are due to the fact that stellar models themselves are uncertain. In this study, we will
analyze the sensitivity of our metrics to both systematic and model uncertainties. The first part of our analysis will focus on systematic uncertainties caused by fitting stellar models to observables. Past studies, such as Bellinger et al. (2019), have investigated how the uncertainty in observable properties affects the precision and accuracy of model-derived stellar properties, such as masses, ages, and radii. Our analysis will differ from previous work in that, in addition to fundamental stellar properties like mass and age, we will observe the sensitivity of our long-term habitability metrics to uncertainty in observable properties. We will start with fiducial stellar models with known outputs. Then we will inject synthetic uncertainties in the output parameters and see how the spread in values of the recovered input parameters and derived properties change in response. We would like to determine the effects of obtaining more precise measurements of stellar properties and to determine which properties our metrics are most sensitive to.

The second component of this study will focus on model uncertainties. While stellar structure and evolution are generally well understood for main-sequence stars, stellar models often differ in terms of the physical and chemical processes they incorporate. For instance, models disagree about the chemical abundances of the solar photosphere. This is a reflection of what is known as the solar abundance problem, where newer solar abundance measurements result in a worse match to the observed sound speed and density profiles obtained via asteroseismology (Serenelli et al. 2009). Stellar models also vary in terms of the equations of state, opacities, and nuclear reaction networks that are used (Buldgen et al. 2019). Most stellar models are one-dimensional to make evolution calculations over billions of years computationally feasible. However, this leads to uncertainty in a variety of input physics such as their treatment of convection, convective overshoot, rotation, and element diffusion. Given that stellar models vary substantially in terms of the model physics they incorporate, one would like to determine the uncertainties introduced by the choice of stellar model. In a recent study, Tayar et al. (2022) developed a means to quantify stellar model uncertainties. Rather than individually varying the myriad of parameters influencing model physics, they instead compared the differences in the outputs of four well-accepted stellar models that vary in terms of the model physics used. The work of Tayar et al. (2022) characterizes the typical uncertainty in stellar fundamental properties, such as masses and ages, as well as how the uncertainty in these properties varies as a function of a star’s location in temperature–luminosity space. For our analysis, we will apply the methodology of Tayar et al. (2022) to determine the model uncertainties in our metrics for biosignature yield. We will assess which regions of parameter space have the greatest differences between the four stellar models in terms of the calculated values for biosignature yields.

This sensitivity analysis is organized as follows. In Section 2, we describe our methodology, including the metrics that we will use and the stellar model grids and interpolation scheme used in this study. Sections 3 and 4 analyze the sensitivity of biosignature yield metrics to systematic and model uncertainties, respectively. In Section 5, we consider the realistic case where our metrics are subject to the combined effects of systematic and model uncertainties. We discuss additional sources of uncertainty in Section 6. Lastly, in Section 7 we discuss the precision to which direct imaging target stars can reasonably be prioritized according to these biosignature yield metrics, and we address which future measurements could best improve the precision in these rankings and reduce ambiguity.

2. Methods

2.1. Relative Biosignature Yield Metrics

In this study, we will make use of the relative biosignature yield metrics, $B$, defined by Tuchow & Wright (2020). These metrics are formulated in terms of assumptions about the distribution of habitable exoplanets, $\Gamma$, and the emergence of detectable biosignatures, $H$, as seen in Equation (1). Higher values for $B$ correspond to a higher likelihood of hosting planets in the habitable zone (described by $\Gamma$) that are able to host detectable biosignatures (described by $H$).

In regions of parameter space where the distribution function for exoplanets, $\Gamma$, is sufficiently well constrained, many studies show that it is well fit to either a power law (Cumming et al. 2008) or split power law (Kopparapu et al. 2018; Dulz et al. 2020). While we do not have many empirical constraints on the distribution function of earthlike planets in the habitable zones of sunlike stars, for this study, we will make use of metrics that assume that planets are distributed as a power law. We will use a form of $\Gamma$ that is uniform in log semimajor axis, $\ln(a)$, which is equivalent to having a planetary distribution function in the form $Ca^\beta$ with $\beta = -1$. This choice of power-law exponent is close to those used in past estimates for the distribution of planets around sunlike stars (see Table 1 in Dulz et al. 2020, but note their different formulation of $\Gamma$ in which our power law would correspond to $\beta = 0.0$), and it implies that exoplanets would be distributed in distance similarly to the observed distribution for planets in the solar system (i.e., “Bode’s Law”). This choice of power-law exponent has been selected for concreteness, but other realistic values of $\beta$ will lead to similar conclusions. One should note that the planetary distribution function is almost certainty dependent on host star mass and spectral type. From past studies, one can clearly see that planet occurrence rates vary between M stars (Hsu et al. 2020) and FGK stars (Hsu et al. 2019). Due to the relatively small sample sizes of long-period planets found, constraining the mass or spectral type dependence of planet occurrence rates has been limited to very large mass bins, such as including all FGK stars together. Therefore, in our study we will use a form of $\Gamma$ that does not depend on stellar mass, as the range of masses in the model grids we use largely overlaps with the FGK spectral types, which have often been grouped together in past occurrence rate studies. Furthermore, the function $H$ can also incorporate the spectral type and mass of a host star, describing one’s prior knowledge about the obstacles to habitability and the emergence of biosignatures caused by different types of stars. However, for the mainly FGK stars in this study, we shall limit ourselves to considering simple forms of $H$ that depend only on the habitable duration, $\tau$.

For our assumptions about the emergence of biosignatures, we use three physically motivated or otherwise commonly used forms for $H$, the probability that a planet with known age and semimajor axis hosts biosignatures (Tuchow & Wright 2020). All of these forms of $H$ depend on the duration that a planet spends in the habitable zone, $\tau$. The first assumption, which can be framed in terms of the quantity $H$, is to assume that all planets in the CHZ have an equal chance of hosting biosignatures. While in this study, we use the term CHZ to refer to the region of the habitable zone that remains habitable.
from the formation of planets to the current day, other studies often use a different formulation for the CHZ. In this commonly used alternative formulation, the CHZ is defined as a region occupied by planets that remains in the habitable zone longer than a given amount of time. Values for the fixed age required for sustained habitability range in the literature from 2 Gyr (Truitt et al. 2020) to 4 Gyr (Hart 1979). If one were to use this alternative formulation for the CHZ and select 2 Gyr as the habitable duration required to host biosignatures, then \( H \) would have the functional form:

\[
H(a, t) = \begin{cases} 
\text{constant}, & \text{if } \tau(a, t) \geq 2 \text{ Gyr} \\
0, & \text{otherwise.}
\end{cases}
\] (2)

Such assumptions about the emergence of biosignatures are somewhat physically motivated based on the timescales on which complex life and biosignatures developed on Earth, but having such a hard age cutoff between systems may have undesired consequences when prioritizing which stars to search for life around.

Another form of \( H \) we will consider is to assume that, on a habitable planet, biosignatures have a constant chance of developing, \( b \), per unit time. This would have the functional form

\[
H(a, t) = 1 - e^{-bt(a,t)}.
\] (3)

In the case where there is a low chance of biosignatures emerging per unit time, and \( b \) is very small, then \( H \) can be approximated as \( H(a, t) \propto \tau(a, t) \). We will assume that \( b \) is small and therefore \( H \) depends linearly on \( \tau(a, t) \), but note that there are actually two possible forms of \( \tau \), the duration that planets are habitable, based on our assumptions about whether BHZ planets can be habitable. One could assume that BHZ planets, originating outside the habitable zone can eventually become habitable, in which case \( \tau \) as a function of distance has a plateau shape, ramping up until it flattens in the CHZ (see Figure 2 in Tuchow & Wright 2020). We will also consider the alternative assumption that only planets in the CHZ can host biosignatures. In this case, \( H \) has the functional form described in Equation (3) within the CHZ and is zero outside of it.

With the different options for \( \Gamma \) and \( H \) discussed above, we have a total of three metrics that we will use in our sensitivity analyses. All of these metrics have the same assumption for \( \Gamma \)—that \( \Gamma \) is uniform in \( \ln(a) \), but they differ in their assumptions about \( H \). First we have a metric that uses a form of \( H \) that is constant for habitable durations greater than 2 Gyr in the form of Equation (2). We will refer to this metric as the 2 Gyr metric, or \( B(2 \text{Gyr}) \). We will refer to the metric with \( H \) proportional to \( \tau \) including BHZ planets as the BHZ metric or \( B(\text{BHZ}) \). Finally, the metric with a form of \( H \) that is proportional to \( \tau \) in the CHZ, and zero outside of it, will be referred to as the CHZ metric or \( B(\text{CHZ}) \). The values of these metrics will be normalized to correspond with the solar value of each metric. This means that a value of 1.00 for a metric corresponds to the value obtained from a model of the present-day Sun. The normalization of biosignature yield metrics is somewhat arbitrary, as metrics are compared relative to each other, but normalizing to solar values allows one to see the relative merit of a star, relative to the probability that a solar analog would host planets with biosignatures.

Calculation of these metrics requires knowledge of how habitable zones change over the course of a star’s lifetime. Using the temperatures and luminosities in stellar model grids discussed in the next section, one can calculate the habitable zone boundaries. In this study we will use the Ramirez & Kaltenegger (2018) formulation of the classical \( \text{N}_2-\text{CO}_2-\text{H}_2\text{O} \) habitable zone. This is similar treatment of the habitable zone to that of Kopparapu et al. (2013), but it is applicable for a wider range of stellar effective temperatures between 2600 and 10,000 K. The inner edge of this habitable zone is defined by the Leconte et al. (2013) runaway greenhouse limit, while the outer edge is given by the maximum \( \text{CO}_2 \) greenhouse heating.

### 2.2. Model Grids

We shall make use of several different stellar models in the multiple components of our sensitivity analysis. In our analysis of the sensitivity of derived model parameters to measurement uncertainties in observational parameters, we shall use a grid of MESA Isochrones and Stellar Tracks (MIST) stellar models (Choi et al. 2016; Dotter 2016). MIST is a grid of stellar models computed using the MESA stellar structure and evolution code ( Paxton et al. 2011, 2013, 2015, 2018, 2019). The MIST models we are using in Section 3 of our analysis include the effects of stellar rotation with a maximum rotational velocity on the zero-age main sequence of \( v/v_{\text{crit}} = 0.4 \). In Section 4, we will also consider MIST model grids that do not include the effects of stellar rotation. MIST gives a grid of stellar models in mass, equivalent evolutionary phase (EEP), and metallicity. EEP is a dimensionless quantity that defines the evolutionary phase of a star ( Dotter 2016). It is defined relative to several physically motivated “primary EEPs” or significant phases in stellar evolution, such as the pre-main sequence (EEP = 1), zero-age main sequence (EEP = 202), terminal age main sequence (EEP = 454), and the tip of the red giant branch (EEP = 605). Between primary EEPs are several “secondary EEPs,” uniformly spaced according to a metric function (see Section 2.2 in Dotter 2016). EEPs are used as a proxy for stellar ages, as stellar lifetimes vary widely over the range of masses, and if models were sampled uniformly in age, late phases of stellar evolution, where stars evolve more quickly, would have inadequate time resolution ( Dotter 2016). To infer the properties of stars between the grid points, we will use the interpolation scheme in the Isochrones python package ( Morton 2015).

For the sensitivity analysis for model uncertainties, we are faced with a challenge in that there are so many options to tweak for stellar model physics. Rather than varying each of the hundreds of model parameters, we will instead use the approach of Tayar et al. (2022), comparing the results of four widely used and well-established stellar models that incorporate different model physics. We will make use of their Klaushokus python package ( Claytor et al. 2020a, 2020b) and their model grids expressed in terms of mass, EEP, and metallicity. We will use the following model grids:

1. Yale Rotating Evolution Code ( YREC; Pinsonneault et al. 1989; Tayar et al. 2022).
2. MESA Isochrones and Stellar Tracks (MIST; Choi et al. 2016; Dotter 2016).
3. Dartmouth Stellar Evolution Program (DSEP; Dotter et al. 2008).

\[\text{Note that this model grid is for MIST models not including the effects of stellar rotation.}\]
4. Garching Stellar Evolution Code (GARSTEC; Weiss & Schlattl 2008; Serenelli et al. 2013).

For a comprehensive summary of the input physics used in each of these stellar models, see Table 1 of Tayar et al. (2022).

### 3. Systematic/Fitting Uncertainties

#### 3.1. Markov Chain Monte Carlo Fitting Uncertainty

When determining the derived properties of a star, such as masses and ages, much of the uncertainty is due to uncertainties in the observed stellar properties. To obtain a stellar model for a given star, one will typically fit a model to a set of observed target properties. It is important to have several properties to fit a stellar model to, as too few constraints can lead to ambiguities in stellar parameters. For instance, if one only has measurements of luminosity and effective temperatures of stars (i.e., positions on the HR Diagram), it can lead to many degeneracies between stars of different masses, ages, and metallicities (Godoy-Rivera et al. 2021). For this reason, it is useful to fit stellar models to additional measured properties such as \(\log (g)\) and surface \([\text{Fe/H}]\) to break the degeneracy. In this section, we would like to show how improving the precision of target values for model fitting affects the precision in derived properties such as masses, ages, and our metrics for biosignature yield.

For this analysis, we will start by using fiducial stellar models of known input properties. We regard these input properties as the “true” properties of the stars, and we will try to recover these values by fitting to observed properties. To simulate observations, we take the model output values and add synthetic measurement uncertainties to them. In this sensitivity analysis, we consider three fiducial target stars (see Table 1). These example stars are chosen heuristically based on our prior assumptions for which stars would be likely or unlikely to host biosignatures on their planets. Later in Section 5, we will compare the values of biosignature yield metrics between these stars and see whether they replicate our intuition for the relative merits of these target stars. First we consider a star that is a solar analog with a mass of \(1.0\, M_\odot\), an age of 4.6 Gyr, and a metallicity of \([\text{Fe/H}] = 0.0\). This star serves as a point of comparison with other stellar models, and we would expect it to have a good chance of hosting planets with biosignatures, as Earth is our one example of a planet with life. Note that this model is slightly different than the Sun in that the Sun’s metallicity is not the same as its observed surface value. Setting \([\text{Fe/H}] = 0.0\) implies that the net stellar iron abundance is the same as the surface iron abundance observed for the Sun. We then consider an example model for a lower-mass K star in the early to middle stages of its main-sequence evolution, which we suspect may be a better candidate in a search for biosignatures than our Sun, as it evolves less quickly on the main sequence, and planets in its habitable zone would remain temperate for a longer duration. This fiducial model has a mass of \(0.7\, M_\odot\), an age of 5 Gyr, and \([\text{Fe/H}] = 0.0\). Finally we consider a higher-mass F-star model with a mass of \(1.25\, M_\odot\), an age of 1 Gyr, and \([\text{Fe/H}] = 0.0\), in the middle of the main-sequence phase of its evolution. This is an example of a star we believe is unlikely to host planets with biosignatures, as its age is younger that the timescale on which detectable biosignatures developed on Earth. For all of these stars, we assume they are observed at a distance of \(d = 50.0\, \text{pc}\) with a visual extinction of \(A_V = 0.0\).

For each of these stars, we assume that we have measurements of the GAIA band magnitudes, parallax, \(\log (g)\), \(T_{\text{eff}}\), and surface \([\text{Fe/H}]\). In a later section, we will also consider the case where independent age constraints are available. These “measurements” are derived from the rotating MIST model’s outputs with added uncertainties typical of each measurement (see Table 2). For several spectroscopic quantities, we obtain typical uncertainties from the work of Brewer et al. (2016). We adopt their typical uncertainties as \(25\, \text{K}\) in \(T_{\text{eff}}\), \(0.010\) in \([\text{M/H}]\), and \(0.028\) in \(\log (g)\). Note that the precision in these spectroscopic measurements may be better that those of typical field stars, but they are a good estimate for the uncertainty in the properties of a well-characterized, bright, nearby target star for future direct imaging missions. Stars of different spectral types will almost certainly vary in terms of the precision of their observable properties. However, in this study we will analyze the effects of scaling these standard uncertainties to higher or lower values. This means that, for a star of a given spectral type, realistic values for the uncertainties in its measurable properties will fall within the range of uncertainties we are testing. Therefore we will use the same standard values for uncertainties for all of the stars, even though they have different spectral types. To obtain the uncertainties in parallax and GAIA band magnitudes, we consulted the GAIA Data Release 2 (Brown et al. 2018). This gives an uncertainty in parallax of 0.04 mas, while the GAIA \(G\), \(G_R\), and \(G_B\) have very small uncertainties on the order of 0.1–1 mmag (for bright stars with \(G < 13\)).

Since the measurement uncertainties in band magnitudes are so small, the dominant source of uncertainty actually comes from the bolometric correction. The models that we are using utilize the MIST bolometric correction grid. These bolometric corrections are computed for the surface conditions present in the MIST models. When one applies this bolometric correction to other model grids (which we shall do in later sections), there is additional uncertainty introduced. For a different but similar bolometric correction grid, Casagrande & VandenBerg (2018) estimated the GAIA band uncertainties introduced by the bolometric correction. We will use the 0.02 mag uncertainty they obtained as a rough estimate of the uncertainty in GAIA band magnitudes caused by bolometric corrections.
To recover stellar fundamental properties from measurements of stellar observables, we use the affine-invariant ensemble Markov Chain Monte Carlo (MCMC) sampler in the *emcee* python package (Foreman-Mackey et al. 2013). We use a $\chi^2$ log-likelihood to compare model outputs to observations, summing the squared differences between outputs and observations divided by their uncertainties. We chose to use Gaussian priors for all fundamental stellar properties except $A_V$, centered near the “true” values of the properties. In the case of $A_V$, we used an exponential distribution for the prior as the example stars we are considering are all nearby and are expected to have close to zero visual extinction. We generated starting points for the ensemble of random walkers used by the *emcee* package’s MCMC sampler by drawing from the prior distribution. For fiducial models for each of the three example stars, priors were adjusted to be sufficiently tight to ensure that walkers did not get lost in regions of parameter space distantly removed from the maxima of the posterior distribution, while being broad enough to not impact the final posterior distribution we obtain. We ran the MCMC chains using 32 walkers for 20,000 iterations, discarding the first 600 iterations as a burn-in step to allow the MCMC chains to find the maxima. To roughly assess the convergence of MCMC chains, we reran the chains using different priors and verified that they resulted in the same underlying posterior distribution.

Using the standard values for the uncertainties in the target properties (shown in Table 2), we investigate the effects of scaling the uncertainty of a single observable property to higher and lower values, while the other parameters’ uncertainties remain at their standard values. This amounts to varying the likelihood used for the MCMC chains, which depends on the uncertainties of observable parameters, while the priors we use remain the same. We vary the uncertainties in $T_{\text{eff}}$, surface $[\text{Fe}/\text{H}]$, and $\log(g)$ individually, and then scale the uncertainty in all parameters together. From the spread in values sampled by the MCMC chains, we observe how stellar properties such as mass, age, and the three biosignature yield metrics vary in response to varying the uncertainty in these target values. Figure 1 illustrates the results of increasing precision in various target properties for a solar model. One can observe that at the standard values for the uncertainties in target values (total uncertainty scale of 1.00 in the top-left panel), different derived properties vary in terms of their typical uncertainties. Stellar ages and the habitable duration dependent metrics, $B(\text{BH}Z)$ and $B(\text{CH}Z)$, have the largest uncertainties between around 5% and 10%. On the other hand, stellar masses and the $B(2\text{Gyr})$ metric have much smaller uncertainties.

One can see in Figure 1 that precision in derived properties is more sensitive to certain observables that others. This figure is, for the case of the solar analog G star, but the same trends...
discussed below are present in the results of the other fiducial target stars. Scaling the uncertainty in specific observables for the solar model, we observe the following trends:

1. Increasing the measurement precision of all observables in tandem directly corresponds to increased precision in derived properties. This relationship appears to be close to a power law with similar exponents for each property.
2. Increasing the precision in $T_{\text{eff}}$ appears to improve precision in derived properties until an uncertainty of roughly 10 K, beyond which there are diminishing returns.
3. Increasing precision in $[\text{Fe}/\text{H}]$ measurements does not seem to play a major role in constraining ages and our $B$ metrics, but more imprecise measurements than the current fiducial uncertainty of 0.010 dex will lead to a decrease in the precision of derived properties.
4. Measurement precision in $\log(g)$ does not have as strong of an effect on constraining derived properties, but increasing the precision of $\log(g)$ measurements appears to continually and gradually increase precision in ages and $B$ metrics, down to very precise $\log(g)$ values.

Note that in this figure, the uncertainties are scaled to unobtainably small levels. We include these small uncertainties to illustrate the effects of obtaining more precise observations, even if such precision in observations may not be attainable. If the uncertainty in properties such as $[\text{Fe}/\text{H}]$ is unrealistically small, it starts having a chaotic effect where it dominates the goodness of fit, and best-fit models are found in very different regions of parameter space. Future measurements of these properties are unlikely to approach the extreme precisions where this becomes a problem, but this serves to illustrate the negative effects of having one observable measured much more precisely than the others.

3.2. Adding Age Constraints

We then consider the case where one has independent constraints on stellar ages. Independent age constraints are usually difficult to obtain (Soderblom 2010), but may be immensely useful to fit models to stars and reduce systematic uncertainties. Among the different methods to obtain independent estimates on stellar ages, gyrochronology may be one of the best methods to determine the ages of main-sequence stars that will comprise the majority of target stars for future exoplanet detection missions. Gyrochronology is able to constrain the ages of stars by observing the empirical relation that older stars have longer stellar rotation periods (Barnes 2003). When gyrochronological relations are used in addition to stellar isochrone fitting, Angus et al. (2019) found that it improved the precision in FGK star ages by a factor of three beyond the precision given by isochrone fitting alone. It should be noted that these empirical gyrochronological relations have inherent uncertainty due to the spread in rotational periods for stars of the same age and spectral type. However, recent observational work has shown the mechanisms responsible for slowing stellar rotation may stall during some phases of stellar evolution only to resume later (Curtis et al. 2020). Accounting for the duration that stars of different masses spend in this stalled state may allow for much more precise ages to be determined via gyrochronology. Gyrochronology infers the ages of stars based on their rotational periods, so in order to use it to constrain stellar ages, one requires measurements of stellar rotational periods, which are only available for certain target stars. Periods are often inferred by measuring the photometric variability of stars caused by features on the star’s surface such as starspots (Soderblom 2010). As stars with measurable rotational periods have a certain level of activity, this implies that it may be more difficult to detect planets around such stars.

In addition to the target values described in the previous section, for the three example stars of different masses, we include an independent age constraint. We modify the precision of the age constraint and, after running MCMC chains for each value, we report the uncertainty in derived properties. The results of adding this age constraint can be seen in Figure 2. Obviously, adding an age constraint decreases the uncertainty in model ages for all of the stars, as one can see in the figure. Less obvious is the fact that the sensitivity of $B$(BHZ) and $B$(CHZ) closely follows that of model ages, specifically in the case for the 0.7 and 1.25 $M_\odot$. Interestingly, the 1.0 $M_\odot$ star appears to have lower-percent uncertainties in age, $B$(BHZ), and $B$(CHZ) compared to the other stars. It is unclear why this is the case, but it may have to do with the fact that the solar analog is at a slightly later evolutionary phase than the other.
stars. One observes that for the solar analog, and to a lesser extent the 1.25 $M_\odot$ star, decreasing the uncertainty in age to a certain extent causes the uncertainty in $B(\text{BHZ})$ and $B(\text{CHZ})$ to saturate at a fixed value. However, the age uncertainty at which this occurs is so small as to not be reasonably obtainable. Mass and $B(2\text{Gyr})$ appear to be relatively insensitive to constraints in age, as they can be determined precisely without them. Note that for the case of the 1.25 $M_\odot$ star, the model ages are all less than 2 Gyr, so the $B(2\text{Gyr})$ values are all zero with very low variability, so they do not appear on the plot.

4. Model Uncertainties

While systematic uncertainties in fitting stellar models to stars contribute a large portion of uncertainty in derived parameters, another major source of uncertainty comes from the fact that stellar models themselves are uncertain. In this section, we will investigate how differences between stellar models affect the calculated values for our biosignature yield metrics.

As described in Section 2.2, we use the Kiauhoku package (Claytor et al. 2020a) to compare the outputs of four widely used stellar models with varying model physics (Tayar et al. 2022). For all of the model grids, we calculated the values of our $B$ metrics for each point in mass, EEP (equivalent evolutionary phase), and metallicity space. Then we used the dataframe interpolator in the Isochrones package (Morton 2015) to interpolate our metrics between the points given in the model grids. For each of the three treatments of $B$, we plotted our metrics as a function of mass and EEP, and calculated the maximum difference in values between the model grids. In this analysis, we focus on the region of parameter space that overlaps between the different model grids, and that is of interest in a search for biosignatures. We chose to consider stars in a mass range between 0.6 and 1.5 $M_\odot$ due to the fact that some model grids do not extend to lower masses than 0.6 $M_\odot$, and stars more massive than 1.5 $M_\odot$ would evolve too quickly and have too short lifetimes to be preferential locations in a search for life. Note that while M dwarfs could potentially be ideal candidates for the search for life, they are not considered in this section of our analysis due to the fact that they fall outside the overlapping regions of these stellar model grids. Future studies to analyze the model uncertainties for M dwarfs will need to consider different stellar models that are valid in that region of parameter space. For the range of EEP values, we consider the evolutionary phases of interest to be between the zero-age main sequence (EEP = 202) and the tip of the red giant branch (EEP = 605).

In Figures 3–5, we show the values of the $B$ metrics for the four different models in the left panels, and the right panels show the maximum differences in $B$ between models. Each of these figures shows models with a solar metallicity of $[\text{Fe}/\text{H}] = 0.0$. This is due to the fact that many stellar model grids have fairly low resolution in metallicity space, but all of them include grid points at $[\text{Fe}/\text{H}] = 0.0$. Therefore we will compare models with solar metallicity to avoid the additional complication of
interpolation over a roughly sampled model grid. Note that these different models may disagree about what the true metal mass fraction for the Sun is, or what solar abundance pattern to use. However, these models agree that the solar metallicity at $[\text{Fe/H}] = 0.0$ matches the observed surface values for the Sun, even if they disagree as to what those values are. The color bars in these figures show the values of the metrics relative to those of the present-day Sun.

The contours in the left panels of these figures represent stellar ages in gigayears. In these subplots, ages greater than 10 Gyr have been masked out since nearby stars, which are the primary targets for future direct imaging surveys, are unlikely to have ages exceeding 10 Gyr, the age of the Galactic disk (Carraro 2000). Masking stars past this age has the consequence of excluding the regions of parameter space occupied by highly evolved low-mass stars that would have very high B values. Since these stars would have ages exceeding the age of the galactic disk and would not be found among typical field stars, masking them allows one to more clearly see features in the regions of the color maps that are more relevant to future surveys. One can see very large differences between the age contours of different models in these plots. This represents the fact that stellar models disagree on the age of a star corresponding to a given mass and evolutionary phase.

On the right panels of these plots are the maximum differences in $B$ between models. For the $B(CHZ)$ and $B(BHZ)$ metrics, one can see that there are major differences between models during the earlier part of the main sequence, corresponding roughly to EEPs of 225–350. This corresponds to differences in model ages, and since these metrics are dependent on time spent in the habitable zone, they are heavily dependent on stellar ages. The mismatch between model ages is greatest for low-mass stars, and thus we see the largest model uncertainties for $B(CHZ)$ and $B(BHZ)$ in this region of parameter space. Additionally, for $B(CHZ)$ there is a slight model disagreement near the end of the main sequence at EEPs of 400–450. This is likely due to differences in stellar evolutionary tracks for the four models and disagreements about how quickly the stars evolve on the late main sequence. Since the $B(CHZ)$ metric depends on the time spent in the CHZ, the models disagree about where $B(CHZ)$ goes to zero, i.e., when there ceases to be a region of the habitable zone where planets have been continuously habitable from their formation. The $B(BHZ)$ model differences plot has a similar, but more prominent secondary maximum on the subgiant branch, at EEPs around 475–550. The models disagree regarding how long regions of the habitable zones of subgiants have stayed habitable and how quickly subgiants evolve in temperature and luminosity. The $B(BHZ)$ metric allows planets that start outside the habitable zone to be considered habitable, and depends on time spent in the habitable zone, so there is a large model discrepancy in this region.

The $B(2\text{Gyr})$ metric is formulated differently from the other metrics, in that it primarily depends on the width of the region that spends more that 2 Gyr in the habitable zones, and $H$ does not depend linearly on the habitable duration. This causes the plot of model differences in Figure 5 to appear different than those of the other metrics. In the left panel, one can see that the $B(2\text{Gyr})$ drops off to zero for stars with ages less than 2 Gyr, and then abruptly jumps up for ages greater than 2 Gyr. However, one can observe that the four stellar models disagree about the position of the 2 Gyr age contour. Note that in the left panels of Figure 5, there is a roughness in the color maps around the 2 Gyr contour due to the resolution of the model grids in mass and EEP spaces. The MIST model grid, with the highest resolution, does not have the same problems as the other grids, emphasizing that this roughness is due to resolution and not an inherent uncertainty in the models themselves. Nonetheless, resolution in model grids is a concern to take into account when a given star falls between grid points, and it can contribute to uncertainty. The right panel of Figure 5, representing model differences in $B(2\text{Gyr})$, shows a large strip near the bottom of the plot representing the disagreement in the 2 Gyr contour. Furthermore, there seems to be an additional region of high disagreement between models on the subgiant branch. Models disagree about when regions of the habitable zone that have remained habitable for more than 2 Gyr leave the habitable zone, and about how quickly stars evolve on the subgiant branch.

From these results, we can see that a large portion of model uncertainties arises due to uncertainties in the ages of stars. For all three of these metrics, one can see that the largest values are typically obtained for older low-mass stars, but stars in this region of parameter space also appear to have the largest model uncertainties. This is due to the fact that low-mass stars evolve very slowly and it is difficult to constrain their ages as evolutionary states billions of years apart appear very similar. The $B(2\text{Gyr})$ metric, which has a sharp cutoff at 2 Gyr, is particularly sensitive to these uncertainties, as models disagree about which evolutionary phase corresponds to an age of 2 Gyr. There are also model disagreements for later phases of stellar evolution, during the late main-sequence and subgiant phases. While uncertainty in stellar ages may play a role here, the
adding a 10% independent age constraint. Horizontal dotted lines represent the true values of observables for each of the fiducial stellar models. Models without stellar age constraints are shown as the filled markers, while the unfilled markers show the effects of adding a 10% independent age constraint.

In a real scenario, when one fits models to stellar observations, model predictions will be subject to both model and systematic uncertainties. Using the three example stars described in Section 3, we will briefly discuss how the derived values for these stars vary when both these sources of uncertainty are taken into account.

We start with the “true” values for these stellar derived properties, based on the model outputs from the rotating MIST model used in Section 3. Rotating MIST models were used to generate “true” values for the stars as these models are state of the art, and are different from the four stellar models we are comparing (though the nonrotating MIST grid shares much of the model physics besides rotation). We would like to see whether the four different stellar models can recover the “true” values of our metrics, and whether the relative ranking of the stars in terms of $B$ values is the same for all stellar models and if the uncertainties are small enough to unambiguously determine which stars have higher metric values. We will consider two cases with and without independent constraints on stellar age. For each stellar mass and each stellar model, we run MCMC chains and fit the models to the simulated observables.

The results of these model fits can be seen in Figure 6. We plot the recovered values of the $B$(BHZ) metric, as this metric behaves very similarly to $B$(CHZ) and age, and plots of those quantities would appear almost identical. (The $B$(2Gyr) metric behaves differently and will be discussed later.) In this plot, the columns on the $x$-axis represent the different stellar models, and the different colored markers represent the recovered values of the $B$(BHZ) metric. Filled markers represent the case where one does not have independent age constraints, while the unfilled markers show the results if one has a 10% constraint on stellar ages. The $y$-axis gives the value of $B$(BHZ) in solar units, relative to the values computed for a solar model using the rotating MIST model. Horizontal dotted lines on this plot represent the “true” values of $B$(BHZ) for the fiducial model, and are shown as a point of comparison.

Without age constraints, one can see that values for $B$(BHZ) appear to be quite variable between different stellar models. While for the 1.25 $M_\odot$ and solar-mass cases, the values are pretty well constrained even if they may be a bit offset from the “true” values, the $B$(BHZ) values for the low-mass 0.7$M_\odot$ case appear to vary wildly and with large uncertainties. Only the recovered value for the nonrotating MIST model matches the “true” values, but even then, the uncertainties are large enough to overlap with those of the 1.0$M_\odot$ model. Still, even if these values for $B$(BHZ) do not match the “true” values, one can still see that in all cases except for the MIST model, one can...
unambiguously rank the 0.7 $M_\odot$ star as being a better candidate for biosignatures than the 1.0 $M_\odot$ star, which in turn is a better candidate than the 1.25 $M_\odot$ star.

The differences in $B(BHZ)$ between different models stem from the fact that $B(BHZ)$ is heavily dependent on a star’s age, and the different models can strongly disagree on stellar ages. The disagreement is highest for low-mass stars, so it makes sense that the 0.7 $M_\odot$ star has the largest model difference in $B(BHZ)$ values. Adding a 10% independent constraint on age drastically improves the agreement between the derived $B(BHZ)$ values and the “true” values, especially for the case of lower-mass stars. One can observe that not only is there a greater agreement between models, but there is improved precision on the derived values of $B(BHZ)$. This demonstrates that while independent age constraints are difficult to obtain, they greatly improve our ability to constrain our values for biosignature yield metrics.

The results for the derived properties of $B(CHZ)$ and stellar age have very similar forms to those of Figure 6. However, the other derived properties, mass and $B(2\text{Gyr})$, behave differently from these primarily age-dependent quantities. The derived stellar masses largely match those of the fiducial models, and have relatively low uncertainties. Without age constraints, there are a few small offsets from the fiducial values, but adding age constraints generally improves the agreement. For the $B(2\text{Gyr})$ metrics, the derived values for all of the models without age constraints correspond well to the “true” values. The uncertainties for $B(2\text{Gyr})$ values are generally very small, but for Markov chains that explore the sharp boundary around 2 Gyr, such as some of the chains for the 1.25 $M_\odot$ star, the uncertainties can be quite large. Adding age constraints greatly reduces uncertainties in these values.

When both model and systematic uncertainties are included in our analysis, one can observe that stars of masses equal to or greater than that of the Sun have small enough uncertainties in the values of $B(BHZ)$ metrics to allow them to be unambiguously ranked in terms of their potential to host biosignatures. Low-mass stars generally vary greatly between models and have large uncertainties, but in many cases, the relative ranking between stars of different masses and ages is preserved. The incorporation of independent age constraints greatly lowers the model disagreements and uncertainties in metric values for all stars, but the effect is most prominent for the case of low-mass stars. One can see that, with or without age constraints, the ranking of the three example stars compared in this study replicates our qualitative understanding that the older K star should be a stronger candidate for hosting planets with biosignatures than the solar analog, which itself is a better candidate than the younger F star. Our results are consistent with the observations of Tuchow & Wright (2020) that metrics with forms of H that depend on the duration spent in the habitable zone and power-law forms of $\Gamma$ favor older low-mass stars. This makes sense, as low-mass stars evolve more slowly on the main sequence, so their habitable zones remain stable for longer periods of time, and planets found in their habitable zones would have a higher chance of developing life. Furthermore, assuming planets are distributed according to a power-law form of $\Gamma$, habitable zones around lower-mass stars may have a higher chance of hosting rocky planets. However, this probability is offset by the fact the more massive stars have wider habitable zones located farther from their stars, so even if planets have lower occurrence rates at larger semimajor axes, there may still be comparable probabilities of hosting planets in the habitable zone (see the bottom-right panel of Figure 5 in Tuchow & Wright 2020, where metrics that depend only on the number of planets in the habitable zone are nearly uniform in mass and age). As the stars with the highest chances of hosting planets with biosignatures also have the greatest systematic and model uncertainties, it is of critical importance to obtain independent age constraints for these stars, allowing one to obtain their fundamental properties with much higher precision.

6. Additional Sources of Uncertainty

Our study is primarily concerned with investigating how our metrics for long-term habitability are affected by systematic and model uncertainties. However, one should note that these are not the only sources of uncertainty that $B$ values are subject to. Here we will briefly discuss the additional sources of uncertainty that would affect one’s ability to precisely rank target stars according to our metrics.

Recall from Equation (1) that our metrics $B$ were formulated in terms of two quantities: $H$ and $\Gamma$. In our analysis, we have been mainly concerned with $H$, the probability that a planet hosts biosignatures. The forms of $H$ we considered in this study were dependent on the duration spent in the habitable zone, and our sensitivity analysis has focused on the uncertainties these quantities caused by uncertainty in stellar evolutionary tracks. Much of our sensitivity analysis for $B$ has actually been an investigation as to how stellar evolutionary tracks are affected by systematic and model uncertainties. Beyond the effects of stellar evolution, $H$ is also subject to uncertainties from other sources, namely determining the boundaries of the habitable zone. The habitable zone is formulated based on models of planetary atmospheres, determining the range of stellar fluxes where a rocky planet could support liquid water on its surface (Kasting et al. 1993). However, models of planetary atmospheres vary substantially and can yield vastly different predictions for the inner and outer limits of the habitable zone. In this study, we limited ourselves to considering one particular formulation of the “classical” habitable zone that assumes that CO$_2$ and H$_2$O are the dominant greenhouse gases in planetary atmospheres (Ramirez & Kaltenegger 2018), but using other formulations of the habitable zone will likely change the ranking of stars according to our metrics. For instance, some habitable zone formulations use different greenhouse gases such as CH$_4$ (Ramirez & Kaltenegger 2018) or H$_2$ (Pierrehumbert & Gaidos 2011), while others consider the case of planets with low atmospheric water concentrations (Abe et al. 2011). Future work is needed to investigate how our metrics perform when other formulations of the habitable zone are used.

The present-day habitable zone (as calculated via a given formulation) can be determined based solely on the effective temperature and luminosity of the host star. Because these stellar observable properties are subject to measurement uncertainties, the boundaries of the habitable zone also have an associated uncertainty (Kane 2014). Furthermore, properties used to determine the habitable zone boundaries, such as stellar luminosities, have additional uncertainty introduced by the uncertainty in stellar distance measurements (Kane 2018). These effects are largely included in our study of systematic uncertainties, and our work expands on previous studies, analyzing the uncertainty in quantities derived from the CHZ and the habitable durations of exoplanets.
Up until this point, we have only discussed uncertainties in $B$ caused by phenomena that influence the function $H$. The other quantity used to calculate our $B$ metrics, the planetary distribution function, $\Gamma$, has been held at a fixed functional form for our sensitivity analysis, but we note that $\Gamma$ will also be subject to large uncertainties. As mentioned in Section 1, due to the limited sample sizes of known planets in many regions of parameter space, the planetary distribution function is inherently uncertain, especially in the range of semimajor axes around the habitable zones of sunlike stars. Furthermore, as we discuss in Section 2.1, while $\Gamma$ is almost certainly dependent on additional factors such stellar mass, this dependence is extremely difficult to constrain due to the small sample sizes involved. In future work, we would like to investigate how our biosignature yield metrics are sensitive to uncertainties in $\Gamma$, and determine how different forms of $\Gamma$ affect the ranking of target stars. Beyond the two contrasting forms of $\Gamma$ compared in Tuchow & Wright (2020), we would like to investigate how using power laws with different exponents, or using split power laws, affects the relative values of $B$.

7. Results and Conclusions

In this study, we have analyzed the sensitivity of the biosignature yield metrics developed by Tuchow & Wright (2020) to both model uncertainties and systematic fitting uncertainties. These metrics serve to quantify one’s prior knowledge about planetary habitability and assumptions about the emergence of life. Using this mathematical framework allows one to assess whether a star would be a good target to host biosignatures on its planets, and these metrics would be potentially useful for target prioritization for future missions to search for biosignatures. In order to apply this framework to rank which stars would be the best targets, we needed to determine the precision to which we could determine the values of our metrics, $B$, and whether their uncertainties were small enough to allow individual stars to be unambiguously ranked. We looked at how several biosignature yield metrics with different models for $H(a, t)$, the probability of biosignature emergence, varied depending on the choice of stellar model and the precision of stellar measurements.

The $B(2\text{Gyr})$ metric treats all planets that spend more than 2 Gyr in the habitable zone as equally likely to host biosignatures. This metric therefore depends primarily on the size of the region that spends more than 2 Gyr in the habitable zone and only depends on age insofar as it affects the width of that region. $B(2\text{Gyr})$ has relatively small systematic uncertainties when fitting models to stars even when stellar observables are poorly constrained. However, this metric can have large model uncertainties in some regions of parameter space, such as around the 2 Gyr age cutoff, as models disagree about the exact ages of stars, and the timescales for different phases of stellar evolution. The sharp cutoff where $B(2\text{Gyr})$ drops to zero for habitable durations less than 2 Gyr makes this metric and other metrics with fixed-age formulations less physical in nature and more biased by assumptions of the timescales required for the emergence of life and detectable biosignatures.

The metrics $B(\text{CHZ})$ and $B(\text{BHZ})$ depend on the amount of time planets spend in the CHZ and the entire habitable zone, respectively. The form of $H$ used by $B(\text{CHZ})$ incorporates the assumption that only planets in the CHZ, which have remained in the habitable zone for the age of the star, would be suitable locations to search for biosignatures. Meanwhile, $B(\text{BHZ})$ uses a more optimistic form of $H$, which assumes that in addition to CHZ planets, BHZ planets that enter the habitable zone due to the star’s evolution could also host biosignatures. Since these metrics are proportional to the habitable duration, it makes sense that they are heavily dependent on stellar ages, and the precision in $B(\text{CHZ})$ and $B(\text{BHZ})$ depends on the precision to which the stellar ages can be determined. Without independent constraints on ages, stellar ages can be well determined for F and G spectral types, but for K stars and stars with lower masses, stellar ages can be very uncertain and vary wildly between stellar models. This in turn means that $B(\text{CHZ})$ and $B(\text{BHZ})$ have the greatest discrepancy for low-mass stars on the main sequence. For these metrics, there are also disagreements later in a star’s lifetime. $B(\text{CHZ})$ has high model uncertainties near the end of the main sequence, while $B(\text{BHZ})$ has large discrepancies between models on the subgiant branch. As older low-mass stars have the highest values for these metrics, but also the largest uncertainties, one would like to better constrain their evolutionary tracks by obtaining independent constraints on their ages.

Although independent constraints on stellar ages can be very difficult to obtain, we have shown that they can play a major role in improving the precision of biosignature yield metrics. For the $B(\text{BHZ})$ and $B(\text{CHZ})$ metrics, we find that incorporating independent age constraints when fitting isochrones to stars can greatly decrease both the model and systematic uncertainties. Furthermore, the precision on the age constraint applied is strongly correlated to the precision in $B(\text{BHZ})$ and $B(\text{CHZ})$ metrics. Because of the role that age constraints play in decreasing systematic and model uncertainties for these metrics, we find that they make the ranking of target stars according to biosignature yield less ambiguous.

In preparation for future space-based direct imaging missions, it makes sense to place emphasis on obtaining independent stellar age constraints. Current and future missions, such as TESS and PLATO, will be able to precisely measure the photometric variability of stars and obtain their rotational periods and asteroseismic modes of oscillation. Among the stars that will be surveyed by the TESS and PLATO missions are bright, nearby stars that are the best targets for future direct imaging missions. With independent age constraints via gyrochronology and asteroseismology, ages are able to be more precisely constrained than via isochrone fitting alone (Angus et al. 2019; Bellinger et al. 2019). This would allow these direct imaging target stars to be more effectively prioritized according to their long-term habitability and would allow future missions to test hypotheses involving stellar ages, such as the planetary age–oxygen correlation proposed as a statistical biosignature (Bixel & Apai 2020). If obtaining independent age constraints is prioritized for the population of direct imaging target stars, future missions will be able to better assess the evolutionary tracks and long-term habitability of lower-mass stars, which may be the best candidates in a search for biosignatures.

Even without age constraints, the long-term habitability metrics that we have tested in this analysis will be useful for prioritizing direct imaging observations. These metrics are determined probabilistically, based on planet occurrence rates and the probability of biosignature emergence, so the exact value of a metric for a star is less informative than the relative values between stars when assessing the merit of target stars in a search for biosignatures. While one may have difficulty
determining the relative ranking of individual low-mass stars according to these metrics, ultimately what is most important is determining which populations have higher metric values relative to each other. Future missions will be able to select optimal target lists according to these long-term habitability metrics regardless of whether we have the precision to determine an unambiguous ranking of specific stars according to these metrics. By using these metrics as a factor for target prioritization, future missions can ensure that they have the highest chance of observing biosignatures in planetary atmospheres.

N.W.T. and J.T.W. acknowledge James Kasting and Ravi Kopparapu for their assistance and guidance on this project. This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. The computations for this study made use of the following python packages: isochrones (https://github.com/timothydmorton/isochrones), Kiauhoku (https://github.com/zclaytor/kiauhoku), and emcee (https://github.com/dfm/emcee). The Center for Exoplanets and Habitable Worlds and the Penn State Extraterrestrial Intelligence Centers are supported by the Pennsylvania State University and the Eberly College of Science.

**ORCID iDs**

Noah W. Tuchow @ https://orcid.org/0000-0003-3989-5545  
Jason T. Wright @ https://orcid.org/0000-0001-6160-5888

**References**

Abe, Y., Abe-Ouchi, A., Sleep, N. H., & Zahnle, K. J. 2011, *AsBio*, 11, 443
Angus, R., Morton, T. D., Foreman-Mackey, D., et al. 2019, *AJ*, 158, 173
Barnes, S. A. 2003, *ApJ*, 586, 464
Bean, J. L., Abbot, D. S., & Kempton, E. M.-R. 2017, *ApJL*, 841, L24
Bellinger, E. P., Hekker, S., Angelou, G. C., Stokholm, A., & Basu, S. 2019, *A&A*, 622, A130
Bixel, A., & Apai, D. 2020, *ApJ*, 896, 131
Brewer, J. M., Fischer, D. A., Valenti, J. A., & Piskunov, N. 2016, *ApJS*, 225, 32
Brown, A., Vallenari, A., Prusti, T., et al. 2018, *A&A*, 616, A1
Buldgen, G., Salmon, S., & Noels, A. 2019, *FrASS*, 6, 42
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, *ApJ*, 809, 8
Carraro, G. 2000, *The Evolution of The Milky Way* (Berlin: Springer), 335

Casagrande, L., & VandenBerg, D. A. 2018, *MNRAS*, 479, L102
Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102
Claytor, Z. R., van Saders, J. L., Santos, A. R., et al. 2020a, *ApJ*, 888, 43
Claytor, Z. R., van Saders, J. L., Santos, A. R. G., et al. 2020b, kiauhoku:
Stellar Model Grid Interpolation, Astrophysics Source Code Library, ascl:2011.027
Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, *PASP*, 120, 531
Curtis, J. L., Agüeros, M. A., Matt, S. P., et al. 2020, *ApJ*, 904, 140
Dotter, A. 2016, *ApJS*, 222, 8
Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, *ApJS*, 178, 89
Dulz, S. D., Plavchan, P., Crepp, J. R., et al. 2020, *ApJ*, 893, 122
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv:2001.06683
Godoy-Rivera, D., Tayar, J., Pinsonneault, M. H., et al. 2021, *ApJ*, 915, 19
Hart, M. H. 1979, *Icar*, 37, 351
Hsu, D. C., Ford, E. B., Raggozine, D., & Ashby, K. 2019, *AI*, 158, 109
Hsu, D. C., Ford, E. B., & Terrien, R. 2020, *MNRAS*, 498, 2249
Kane, S. R. 2014, *ApJ*, 782, 111
Kane, S. R. 2018, *ApJL*, 861, L21
Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, *Icar*, 101, 108
Kiang, N. Y., Domagal-Goldman, S., Parenteau, M. N., et al. 2018, *AsBio*, 18, 619
Kopparapu, R. K., Hebrard, E., Belikov, R., et al. 2018, *ApJ*, 856, 122
Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, 765, 131
Leconte, J., Forget, F., Charnay, B., Wordsworth, R., & Pottier, A. 2013, *Natur*, 504, 268
Luger, R., & Barnes, R. 2015, *AsBio*, 15, 119
Meadows, V. S., Reinhard, C. T., Arney, G. N., et al. 2018, *AsBio*, 18, 630
Morton, T. D. 2015, isochrones: Stellar Model Grid Package, Astrophysics Source Code Library, ascl:1503.010
National Academies of Sciences, E., & Medicine 2021, *Pathways to Discovery in Astronomy and Astrophysics for the 2020s* (Washington, DC: The National Academies Press)
Paxton, B., Bildsten, L., Dotter, A., et al. 2011, *ApJS*, 192, 3
Paxton, B., Cantiello, M., Arras, P., et al. 2013, *ApJS*, 208, 4
Paxton, B., Marchant, P., Schwab, J., et al. 2015, *ApJS*, 220, 15
Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, *ApJS*, 234, 34
Paxton, B., Smolec, R., Schwab, J., et al. 2019, *ApJS*, 243, 10
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, *PNAS*, 110, 19273
Pierrehumbert, R., & Gaidos, E. 2011, *ApJL*, 734, L13
Pinsonneault, M., Kawaler, S. D., Sofia, S., & Demarque, P. 1989, *ApJ*, 338, 424
Ramirez, R. M., & Kaltenegger, L. 2018, *ApJ*, 858, 72
Serenelli, A. M., Basu, S., Ferguson, J. W., & Asplund, M. 2009, *ApJL*, 705, L123
Serenelli, A. M., Bergemann, M., Ruchti, G., & Casagrande, L. 2013, *MNRAS*, 429, 3645
Soderblom, D. R. 2010, *ARA&A*, 48, 581
Tayar, J., Claytor, Z. R., Huber, D., & van Saders, J. 2022, *ApJ*, 927, 31
The LUVIOIR Team 2019, arXiv:1912.06219
Truitt, A. R., Young, P. A., Walker, S. I., & Spacek, A. 2020, *AJ*, 159, 30
Tuchow, N. W., & Wright, J. T. 2020, *ApJ*, 905, 108
Tuchow, N. W., & Wright, J. T. 2021, *RNAAS*, 5, 194
Weiss, A., & Schlattl, H. 2008, *Ap&SS*, 316, 99
Yang, J., Ding, F., Ramirez, R. M., et al. 2017, *NatGe*, 10, 556