The utility of astrometry as a precursor to direct detection

Dmitry Savransky, N. Jeremy Kasdin, Brandt A. Belson" 

Princeton University, Princeton, NJ USA

ABSTRACT

A recent source of debate in the exoplanet community has been the question of whether an astrometry ‘precursor’ mission is required in order for a direct detection mission to succeed. Using an existing framework for the evaluation of direct detection missions, we address this question by incorporating data which may be generated by an astrometry mission. We present results for cases where the astrometry mission is able to resolve which target stars have planets, where it is able to fit a subset of the orbital parameters of discovered planets, and where the astrometric data is good enough to fit complete orbits. Each of these is evaluated assuming perfect performance on the part of the astrometric instrument, and with varying levels of error.

Keywords: exoplanets, planet-finding, precursors, astrometry

1. INTRODUCTION

As interest grows in dedicated space-based observatories for the direct detection of exoplanets, a question has arisen in the planet-finding community of whether such a mission should be flown before a precursor mission has identified target stars with planets, and begun characterizing their orbits. The Exoplanet Task Force has stated in its 2008 report that “the locations and times of maximum star-planet angular separations determined by an astrometric mission [would] make the follow on direct detection planet characterization missions more cost-effective and observationally efficient”.1 While it seems a matter of common sense that any additional data about exoplanets will necessarily improve the science yield of a direct detection mission, it is actually unclear whether the costs of a precursor are worth the benefits it provides for a follow-up direct detection planet imager.

In this paper, we systematically model the types of information we can reasonably expect from precursor investigations, considering both the ideal and non-ideal cases (i.e., the possibility of errors in precursor results). We then incorporate this data into an existing framework for the simulation of direct detection planet-finding missions and evaluate how it affects mission science yield as measured by a group of metrics. In doing so, we wish to stress that we make absolutely no judgements or claims as to the inherent science value of any potential precursor mission. Our only purpose is to evaluate the effects of precursor data on direct detection missions, and to see whether the ability of an instrument to act as a precursor should be used as a further justification to construct and operate that instrument.

2. SYSTEM FORMULATION

Let the vector \( \mathbf{r}_p \) be the position of a planet with respect to the barycenter of the planetary system to which it belongs, and \( \mathbf{r}_s \) the position of the star with respect to the system barycenter, so that \( \mathbf{r}_{p/s} = \mathbf{r}_p - \mathbf{r}_s \) is the position of the planet with respect to the star. We will assume a dextral coordinate system whose origin is at the system barycenter, with the \( \hat{z} \) unit vector parallel to our line of sight (i.e., the vector between our observatory and the system barycenter can always be written as \( d\hat{z} \) for some positive distance \( d \)). The directions of the other two unit vectors are arbitrary, and for simplicity, are chosen such that \( \hat{x} \) corresponds to the direction of the observatory’s orbital velocity at the time of observation. Thus, if \( \mathbf{r}_p \) and \( \mathbf{r}_s \) in this coordinate system, at the time of an observation are given by \( (x_p, y_p, z_p) \) and \( (x_s, y_s, z_s) \), respectively, then a direct detection instrument will return values for \( (x_p - x_s, y_p - y_s) \), an astrometric instrument∗ will return values for \( (x_s, y_s) \), and a radial

Send correspondence to Dmitry Savransky dsavrans@princeton.edu

*In reality, a narrow-angle astrometric survey will actually give the positions of stars with respect to a centroid determined by a set of reference stars, which will also move in time due to the proper motions of the target and reference stars, and parallax effects due to the observatory’s orbit. For simplicity, we assume that all of these effects can be factored out with a sufficiently large and high fidelity data set, reducing the observation to the target star’s position with respect to its system barycenter.2
velocity measurement will return values for \( \dot{z}_s \), where the dot operator is used to represent differentiation with respect to time.

Assuming our planetary system can be well described using the Newtonian formulation \((\phi \ll c^2)\), the equations of motion for a system of \( n \) planets are given by

\[
\mathbf{r}_j = -\sum_{k \neq j} \frac{\mu_k \mathbf{r}_{k/j}}{r_{k/j}^3}, \quad j = 1, \ldots, n
\]

where \( \mathbf{r}_{k/j} = \mathbf{r}_k - \mathbf{r}_j \) and \( \mu_k \) is the gravitational parameter of the \( k \)th object. Thus, given a state vector composed of the positions and velocities of a set of planets and a target star at some reference epoch, we can numerically integrate to find the state at any future time. While this formulation is a (nearly) exact representation of a planetary system, we also find it useful to be able to represent each orbit in terms of a set of average orbital elements. We can find a set of Keplerian orbital elements: \((a, e, \tau, \omega, I, \Omega)\) which represent the least squares fit of an ellipse to available orbital data. These elements are the semi-major axis, eccentricity, time of periapsis passage, argument of pericenter, inclination, and longitude of the ascending node (the last three elements can be considered to be Euler angles representing a \( \mathbf{z} - \mathbf{x} - \mathbf{z} \) body rotation, with the original orbit oriented in the \( \mathbf{z} - \mathbf{y} \) plane with periapsis in the \( -\mathbf{z} \) direction.\(^3\)

Having described a target planetary system, we must consider any restrictions on when an observation can be made. All of the detection methods described here cannot operate without a clear line of sight to the target, and direct detection and astrometry are further restricted by the location of the sun and other bright sources in the field of view, so that observing seasons are determined by the orbit of the observatory (for ground-based radial velocity measurements, the orbit and rotation of the Earth determine when observations can be made). During an observing season, the success of an observation is determined by the specific instrument’s ability to resolve the desired signal from the background noise. For direct detection instruments we can define a simple set of parameters which will determine whether a planet is observable at a given point in time. The first of these is the inner working angle (IWA) - a minimum angular separation between a planet and its star at which the planet can be resolved. Thus, a planet is observable only if the apparent separation \((s = \|(x_p - x_s, y_p - y_s)\|)\) is greater than the projected IWA \((\text{IWA} \times d)\) at the time of observation. The second requirement for direct detection is that the planet is sufficiently bright, which we represent as a maximum difference in brightness between the planet and its star, generally stated as the difference between the star and planet magnitude \((\Delta\text{mag})\). This value is given by

\[
\Delta\text{mag} = -2.5 \log \frac{F_p}{F_s}
\]

where \( F_p \) and \( F_s \) are the spectral fluxes of the planet and star, respectively. Since a direct detection system is essentially a mechanism for suppressing starlight while collecting a maximum amount of planet light, we can posit a limiting \( \Delta\text{mag} \) due to systematic limitations, past which the remaining noise of the suppressed starlight completely obscures the planet signal.\(^4\)

As an illustration of these parameters and restrictions, we can consider a system composed of a star identical to the sun, and one planet identical to the Earth, on the same size (semi-major axis and eccentricity) orbit as the Earth, observed with a direct detection instrument with a projected IWA of 0.75 AU (e.g., a 75mas IWA instrument observing a star at 10pc) and a limiting \( \Delta\text{mag} \) of 26. On average, for all possible orientations of this orbit, the probability that the planet will be observable is about 58%.\(^4\) Figure 1 shows a typical example of this kind of orbit, with a randomly selected orientation \((\omega = 97^\circ, I = 50^\circ, \Omega = -60^\circ)\) that makes the planet observable on 53% of the orbit. The figure shows components of apparent separation, star position and radial velocity over the course of one full orbit, indicating when the planet can be observed by the assumed direct detection instrument (the observable epochs).

Extrema in the radial velocity tell us when the star is moving the fastest towards or away from our observatory, which (for an edge-on orbit) would indicate the epochs of periapsis and apoapsis, at which points direct observation of a planet is more likely. Similarly, extrema in the astrometric position (the magnitude of the target star’s separation from the system barycenter) indicate these points for a face-on orbit. Neither of these data sets is a perfect guarantee of finding the maximum orbital separation, since that depends on the exact orientation of
Figure 1. For a 1 solar luminosity star with an Earth mass and radius planet on Earth’s orbit (1 AU semi-major axis and 0.0167 eccentricity), oriented as described in the text with \( \omega = 97^\circ, I = 50^\circ, \Omega = -60^\circ \), this figure shows: (a) Components of apparent separation. The dashed portions represent times when this planet would be unobservable by a direct detection instrument with a projected IWA of 0.75 and a limiting \( \Delta \text{mag} \) of 26. (b) Components of star’s position with respect to system barycenter, in the plane of the sky. (c) Radial velocity of the star.

the orbit with respect to our line of sight, and there are cases when a planet is not sufficiently well illuminated at apoapsis for detection to occur. Still, as we can see from the figure, in this case one year’s worth of noise-free (or low noise) astrometry or radial velocity data would be very helpful, since peaks in the astrometry and radial velocity data correspond to portions of the orbit where the planet is directly observable.

We can also consider another ‘Earth-twin’ on an orbit drawn from this star’s habitable zone (semi-major axis between 0.7 and 1.5 AU, and eccentricity between 0 and 0.35). Figure 2 shows the same orbital components as in Figure 1, but for an orbit with a semi-major axis of 0.82 AU and an eccentricity of 0.3. The planet is also visible on about 52% of this orbit, but, in this case, the extrema in the astrometry and radial velocity signal do not correspond to observable points on the orbit. That is, the only way that precursor data would help in direct detections on this system would be if we had low-noise astrometry and radial velocity data which could be combined into a highly accurate orbital fit, and prior knowledge of the planet’s average albedo and radius. Barring that, partial precursor data would not actually increase the probability of detection of this planet. Lest this be considered an outlier, we can look at the whole population of orbits. For the 97.3% of these orbits on which an Earth-twin can be observed at some point, 10% have extrema in the astrometric position which do not
correspond to a directly observable epoch, and 17% have extrema in the radial velocity (maximum elongations) which do not correspond to directly observable epochs. If we consider the same population of orbits, but with maximum eccentricity of 0.1, now Earth-twins on 96% of the orbits are observable at some point in their year, and 1 and 5% of orbits have extrema in astrometric position and radial velocity, respectively that do not map to directly observable epochs. These values were found via Monte Carlo system generation described in detail in Ref. 6.

Figure 2. For a 1 solar luminosity star with an Earth mass and radius planet on a 0.8 AU semi-major axis, and 0.3 eccentricity orbit, oriented as described in the text with \( \omega = 164^\circ, I = 60^\circ, \Omega = 88^\circ \), this figure shows: (a) Components of apparent separation. The dashed portions represent times when this planet would be unobservable by a direct detection instrument with a projected IWA of 0.75 and a limiting \( \Delta \text{mag} \) of 26. (b) Components of star’s position with respect to system barycenter, in the plane of the sky. (c) Radial velocity of the star.

We can also use this simple set of examples to explore how errors in our knowledge of the planetary orbit will affect our knowledge of when the planet will be detectable. Again taking an arbitrarily oriented copy of Earth’s orbit (this time with \( \omega = -22^\circ, I = 45^\circ, \Omega = 190^\circ \)), we generate a set of orbital elements based on an initial position and velocity. Introducing a 0.1% error in the semi-major axis (equivalent to about 1% error in period - which would be very good for astrometry or radial velocity derived orbital fits for Earth-sized planets), we propagate both orbits forward by five years and compare the planet’s actual observable epochs with those predicted by the approximate orbital elements. As shown in Figure 3, the predicted observable epochs do not correspond at all to the true ones. Even if our knowledge of the orbital period for this planet was precise to within
1% at the beginning of a direct detection mission, by it’s end, this knowledge would be worthless (assuming no updates to the original estimates). For this population of orbits (1 AU semi-major axis and 0.0167 eccentricity), this level of error in orbital parameters yields incorrect predictions of when the planet will be observable over more than half of the orbit for 5% of all possible orbital orientations and over 20% of the orbit in 57% of all orientations.

Figure 3. For a 1 solar luminosity star with an Earth mass and radius planet on Earth’s orbit (1 AU semi-major axis and 0.0167 eccentricity), oriented as described in the text with $\omega = -22^\circ$, $I = 45^\circ$, $\Omega = 190^\circ$, this figure shows the components of apparent separation for (a) The original planet position propagated forward by five years. (b) The planet position derived from average orbital elements with an initial 0.1% error in the semi-major axis. The dashed portions represent times when this planet would be unobservable by a direct detection instrument with a projected IWA of 0.75 and a limiting $\Delta$mag of 26.

These examples are meant to illustrate that, depending on the planetary population of interest, there may exist non-negligible probabilities that even perfect (noise-free) precursor data, short of a complete orbital fit, may not improve our knowledge of when to attempt direct planetary observations. Even full orbital fits cannot tell us when a planet will be sufficiently bright, since this is a function of the planet’s reflectivity and/or emissivity, which cannot be known a priori. Furthermore, relatively small errors in estimated orbital parameters can quickly lead to very large errors in predicted planet position and observability. Nevertheless, it can still be argued that this data does indicate points on a planet’s orbit when it can be directly observed in a large number of cases, and would thus be very useful for direct detection missions. To evaluate this, we need to consider this data in terms of the scheduling of such a mission. Because there are multiple complex constraints on when a given target may be observed, a true test of the utility of precursor data must include a simulated mission timeline, and employ a pool of target stars which are actually available for observation in our galaxy.

3. PRECURSOR DATA IN MISSION SIMULATION

For this analysis, we use the mission simulation framework described in Ref. 6, which generates ensembles of mission simulations (scheduled observations and their simulated outcomes) over a distribution of planets, using a pool of real target stars and a description of a direct-detection planet-finding observatory. The science yield of a mission is evaluated based on the metrics of unique planets found, total number of detections, total number of spectral characterizations, and number of target stars observed. The observatory used here will be the THEIA planet-finding mission concept$^8,9$ (selected since it is the design for which we currently have the most complete...
This planet-finding system is composed of a 4m diameter circular on-axis telescope and a 40m diameter starshade, which blocks light from the target star with no decrease in throughput of planet light down to angular separations of 75mas. The two spacecraft fly in formation, separated by 55000km to cover the 250–700nm spectral band, and by 35000km when covering the 700–1000nm band. The starshade is propelled using high $I_{sp}$ NEXT ion thrusters, and is capable of carrying up to 2930kg of propellant for its 5 year primary mission.

Figure 4. Simulation results for THEIA, for Earth-twin planetary populations on habitable zone orbits. The cases considered are: no precursor knowledge; a precursor modeled as an ideal classifier of which stars have planets; a precursor modeled as an ideal classifier which also produces exact orbital fits; and, a precursor which produces exact orbital fits for a subset of all targets set by $\eta_\oplus$, but does not classify the remaining targets. The lines represent median values for 100 simulations at each assumed value of $\eta_\oplus$. Error-bars represent the one-sided deviations of each distribution. The top left plot shows the total number of planets found (including multiple detections of the same planet), the top right plot shows the number of unique planets found, the bottom left plot shows the number of unique target systems visited during the mission, and the bottom right plot shows the total number of complete spectra (250-1000nm) acquired.

We begin by considering the case of a precursor which acts as an ideal classifier of which stars have planets. This is equivalent to a mission in which you know that every star on your target list has a planet, but not when the planet will be observable. In the simulation framework, knowing which systems have planets is identical to
filtering the target list to some subset whose size is determined by the assumed planetary frequency ($\eta$). For example, if we wish to evaluate the effects of a precursor which perfectly identifies which systems have planets for a mission with a target list of 100 stars in a universe with $\eta = 0.1$, for each simulation, we include only 10 randomly selected stars from the target pool and give each of them a planet. Figure 4 shows the resulting science metrics with and without precursor data for missions simulated on a universe populated with Earth-twins at various assumed frequencies ($\eta_{\oplus}$), with a target list of 117 stars.†

The biggest impact of the precursor knowledge is in cases of low $\eta_{\oplus}$ and is most evident in the number of total detections. Since the mission with the precursor knowledge does not waste any time observing systems without planets, it gets many more total detections, but the number of unique planets discovered does not increase proportionately. This is due to the fact that a large portion of planets will be on orbits where they are only observable for short periods and thus quite difficult to detect. This is demonstrated by the simulation results when $\eta_{\oplus}$ is set to 1—fewer than one third of the 117 planets (one per star) in this case are detected, on average. These represent the ‘easily detectable’ planets - those that are visible on most of their orbits, and orbit stars with longer observing seasons. With the precursor knowledge modeled as a classifier only, we still find the easily detectable planets, but are only marginally more likely to find the ‘difficult’ ones. The precursor knowledge also does not help improve the number of complete spectra taken, since this is more a function of when a planet is detected, and how long it typically stays observable.

Next, we consider the case of a precursor which not only identifies exactly which subset of our target list has planets, but also produces exact orbits for these planets (i.e., we can predict exactly where the planet indefinitely into the future). The simulation framework is further modified so that the scheduling algorithm only considers targets whose planets will be observable at the time of the observation (i.e., when the starshade can be moved into position). As shown in Figure 4, the incorporation of this knowledge makes a large difference in the science yields. The number of total detections is almost constant for all values of $\eta_{\oplus}$ greater than 0.3, indicating that the majority of visits lead to detections, and the number of unique planets found is consistently about 50% of all of the planets within the target pool. The number of spectral characterizations is also greatly increased, since the orbital knowledge lets the scheduling algorithm maximize the available time for integration. The only metric that suffers as the result of this incorporated knowledge is the portion of the target list visited. Because so much more time is now spent on spectral integration, the five year mission is not sufficient to observe even all of the stars which are known to have planets when $\eta_{\oplus}$ is greater than 0.2. However, this problem could be addressed with an extended mission or an instrument with shorter re-targeting times (like an internal coronagraph with the same IWA). We can also simulate the case where exact orbital knowledge is available for a subset of the target list, but it is not known whether the other targets have planets (i.e., the precursor is no longer an ideal classifier). This case is essentially equivalent to the classifier with orbital knowledge when $\eta_{\oplus} = 1$, and to the no precursor case for low $\eta_{\oplus}$, since the scheduler is dominated by targets about which nothing is known. This is clearly reflected in how the science yield metrics vary over increasing $\eta_{\oplus}$.

Finally, we modify the ideal classifier, orbit fitting precursor so that it no longer produces exact orbital fits. Instead, we introduce a 0.1% error in the estimate of semi-major axis, and an error in eccentricity and inclination (at either 1 or 5% for each) at the beginning of the mission (these quantities are harder to constrain from radial velocity and astrometry, so we assume higher errors). From the results in Figure 5, we see that at the lower error rate, the orbit knowledge still produces significant increases in all of the science yields over the ‘no precursor’ case, although these are smaller than those achieved with exact orbit knowledge (by 15% for unique detections and 20% for spectra). At the 5% error rate in eccentricity and inclination, the benefit of having orbital knowledge is almost completely gone, and the science yields are nearly identical to those achieved by the ideal classifier. We can verify that the increase in the science yield in this case is due only to the classifier aspect of the precursor, and not the partial orbit knowledge, by tracking the rate of null detections. If orbit knowledge is making a significant impact, then the number of null detections should be minimized in this case, since the ideal classifier knowledge means that we are never observing a system without a planet. We find that the rate of null detections for the 5% error case is the same as for the classifier only case when averaged over the mission lifetime, indicating

†Note that at $\eta_{\oplus} = 1$, the case where every star has an Earth-twin, the simulations with and without precursor knowledge should produce the same results, which they do, within the margin of error associated with any Monte Carlo with a limited number of trials.
4. CONCLUSIONS

We have considered three basic types of precursor data—the classification of exactly which stars on the direct detection mission’s target list have planets, exact knowledge of planetary positions for all time, and approximate orbital fits. We find that classification alone does not significantly improve the science yield of a direct detection mission, although if the number of targets with planets is small (low $\eta$), it provides more observing time for any other instruments sharing the observatory with the planet-finder. As expected, incorporation of exact knowledge
of the planets' orbits, along with known albedo and planetary radii, significantly improve direct detection mission performance. In the simulations presented here, the THEIA mission concept, which can only find approximately 1/3 of all available planets with no precursor knowledge, was able to detect almost half of all simulated planets with exact knowledge of their orbits. However, these benefits immediately decrease when errors are introduced into the orbit knowledge. We still see great improvement over the 'no precursor' performance when the orbits are very well constrained (0.1% error in semi-major axis and 1% error in inclination and eccentricity), but as soon as the errors on eccentricity and inclination are raised to 5%, the benefits of orbital knowledge vanish.

The precursor knowledge modeled here is actually highly optimistic. There is no such thing as an ideal classifier of which stars have planets. Ground based radial velocity surveys have proven very good at finding gas giants on short period orbits, but detections of smaller planets become progressively more difficult. It is probable that a space-based astrometric survey will be able to more accurately classify stars as having planetary systems, but to benefit a direct detection mission, the astrometric survey would have to cover the direct detection mission's target list. Furthermore, orbital fitting from discrete, noisy measurements is very, very difficult. The error rates assumed here are tiny compared with the confidence intervals assigned to currently existing orbital fits for exoplanets. Finally, it has been suggested that while astrometric planet-finding instruments will be able to constrain orbital periods very well, they may not be able to accurately extract the planetary mass or orbit semi-major axis from their data.\textsuperscript{10} If this turns out to be the case, then the orbital knowledge derived from these missions will not be useful for direct detection missions. Putting all of this together, we must conclude that it has not yet been demonstrated that a precursor is a requirement, or even necessarily useful for direct detection planet-finders. While many of the proposed instruments have other, very good scientific justifications, these do not include their ability to improve the efficiency or yield of direct detection missions.

REFERENCES