

## NEW COMPLETENESS METHODS FOR ESTIMATING EXOPLANET DISCOVERIES BY DIRECT DETECTION

ROBERT A. BROWN AND RÉMI SOUMMER

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Received 2009 December 2; accepted 2010 March 22; published 2010 April 26

### ABSTRACT

We report on new methods for evaluating realistic observing programs that search stars for planets by direct imaging, where observations are selected from an optimized star list and stars can be observed multiple times. We show how these methods bring critical insight into the design of the mission and its instruments. These methods provide an estimate of the outcome of the observing program: the probability distribution of discoveries (detection and/or characterization) and an estimate of the occurrence rate of planets ( $\eta$ ). We show that these parameters can be accurately estimated from a single mission simulation, without the need for a complete Monte Carlo mission simulation, and we prove the accuracy of this new approach. Our methods provide tools to define a mission for a particular science goal; for example, a mission can be defined by the expected number of discoveries and its confidence level. We detail how an optimized star list can be built and how successive observations can be selected. Our approach also provides other critical mission attributes, such as the number of stars expected to be searched and the probability of zero discoveries. Because these attributes depend strongly on the mission scale (telescope diameter, observing capabilities and constraints, mission lifetime, etc.), our methods are directly applicable to the design of such future missions and provide guidance to the mission and instrument design based on scientific performance. We illustrate our new methods with practical calculations and exploratory design reference missions for the *James Webb Space Telescope (JWST)* operating with a distant starshade to reduce scattered and diffracted starlight on the focal plane. We estimate that five habitable Earth-mass planets would be discovered and characterized with spectroscopy, with a probability of zero discoveries of 0.004, assuming a small fraction of *JWST* observing time (7%),  $\eta = 0.3$ , and 70 observing visits, limited by starshade fuel.

*Key words:* methods: statistical – planetary systems – planets and satellites: detection

### 1. INTRODUCTION

Various proposals for direct-imaging searches for Earth-like planets are now on the table (Guyon et al. 2008; Cash et al. 2008; Spergel et al. 2009; Soummer et al. 2009a). This paper presents methods for guiding the development of the new instrumentation that such searches will require. They can also help optimize observing programs by scheduling observations to maximize their impact. Using these new methods, we can predict the outcome of a single observation or an entire observing program and interpret the observational results. These tools should be widely useful for studying, comparing, and optimizing alternative direct-search concepts.

The completeness of a direct limiting search observation (LSO) is the fraction of all possible planets of interest (POIs) that satisfy the detection criteria. By definition, an LSO has sufficient exposure time,  $t_{\text{exp},j}$  on the  $j$ th star, to reach the systematic limit of the instrument. For optical missions, an LSO is assumed to achieve the desired photometric signal-to-noise ratio (S/N) on the limiting source, which has magnitude  $\text{mag}_j + \Delta\text{mag}_0$ , where  $\text{mag}_j$  is the stellar magnitude and  $\Delta\text{mag}_0$  is the limiting magnitude difference with the star (flux contrast), expected to be determined by speckle instability (Brown 2005).

In the simplest picture, an LSO discovers any and all POIs that satisfy two criteria at the time of the observation:

$$\Delta\text{mag} < \Delta\text{mag}_0 \quad (1)$$

and

$$s > \text{IWA}, \quad (2)$$

where  $s$  is the angular separation between a planet and a star and IWA is the inner working angle (angular radius of the real or effective central field obscuration). This simplified picture of a sharp dividing line between detectable and undetectable planets has proven useful for measuring search power to first order. Fidelity could be improved, if necessary, by adding more complex detection criteria, such as a detection probability that varies in a more complicated and realistic way over the field of view.

In a common treatment, which we follow here, the POIs are body-twins of Earth with orbits in the habitable zone.

Let a fraction  $\eta$  of all stars in the universe have a POI. The list of candidate stars for the observing program is a subset of all stars, and within this subset, the expected fraction is  $\eta$ , but the actual fraction will vary. Also, we have limited knowledge of  $\eta$ , a number that is both interesting scientifically and useful operationally to schedule observations to maximize discoveries. Therefore, we will draw distinctions when necessary between the true value ( $\eta_{\text{true}}$ ), which is unknown, the assumed value for science operations ( $\eta_{\text{ops}}$ ), which is used to estimate discovery probabilities, and the value estimated from the results of a search program ( $E(\eta)$ ).

Our new methods extend the original concept of direct-search completeness for exoplanets (Brown 2004, 2005) for purposes of optimizing the timing of revisits to previously searched stars, increasing the realism of instrument comparisons, and providing an estimate of  $\eta$  from the results of observing programs. Meanwhile, Brown (2009a, 2009b) has extended completeness studies to indirect detection by reflex astrometry, photometric detection in the case of no occultations, and the estimation of orbital parameters from Keplerian data sets.

## 2. EVOLUTION AND ESTIMATION OF COMPLETENESS

### 2.1. Four Types of Completeness

Four types of completeness pertain to a program of LSOs: virgin, dynamic, accumulated, and ultimate.

Until now, the scheduling of LSOs in mission studies of direct-search power has been based on “virgin,” time-independent, first-visit completeness  $c_{1,j}$ , which is the completeness of the first LSO of star  $j$  (Brown 2004; Agol 2007; Savransky et al. 2010). In this study, we include the possibility of multiple LSOs of any target star.

A non-detection by an LSO rules out some fraction of possible planets: those with sufficient angular separation and brightness at the time  $t$  of the LSO to be detectable—if they existed. After the LSO, a pool of possible planets may remain, comprising planets that had not been ruled out by previous LSOs and also were not detectable by the most recent LSO. As time goes on, each planet in this pool moves along its unique orbit and may become detectable at some future time. In this way, the fraction of all possible planets that is detectable on the  $i$ th visit to the  $j$ th star at time  $t$ —which is the dynamic completeness,  $c_{i,j}(t)$ —depends on the elapsed time after each of the  $i - 1$  preceding LSOs.

Accumulated completeness ( $C_{i,j}$ ) is the sum-total completeness of  $i$  LSOs of star  $j$ :

$$C_{i,j} \equiv \sum_{l=1}^i c_{l,j}(t_l). \quad (3)$$

$C_{i,j}$  increases monotonically with  $i$ .

Ultimate completeness ( $C_{\infty,j}$ ) is the maximum value of  $C_{i,j}$ . It is the value of completeness that would be accumulated from an arbitrarily high number of LSOs spread arbitrarily over time:

$$C_{\infty,j} = \sum_{l=1}^{\infty} c_{l,j}(t_l). \quad (4)$$

$C_{\infty,j} < 1$  whenever some POI orbits are permanently fainter than  $\Delta\text{mag}_0$  or permanently obscured inside IWA—or never brighter than  $\Delta\text{mag}_0$  and located outside IWA at the same time.

### 2.2. Estimating Dynamic Completeness $c_{i,j}$

All types of completeness are derived from  $c_{i,j}$ , and all results (probability of discovery and mission outcomes) are ultimately based on this quantity.

Because the planetary position is determined by a transcendental equation (Kepler’s Equation),  $c_{i,j}$  must be estimated by Monte Carlo trials. In these trials, we represent the universe of POIs by a large random sample of  $N_0$  particular POIs, each of which is defined by randomly chosen values for 10 parameters:  $\{a, e, M_0, i, \omega_p, \Omega, T, R_p, q, \Phi\}$ , where  $a$  is the semimajor axis,  $e$  is the orbital eccentricity,  $M_0$  is the mean anomaly at some definite time,  $i$  is the inclination angle,  $\omega_p$  is the argument of periastron,  $\Omega$  is the position angle of the ascending node,  $T$  is the orbital period,  $R_p$  is the planetary radius,  $q$  is the geometric albedo, and  $\Phi$  comprises the necessary subparameters for defining the phase function.

In this paper, the POIs are Earth-twins on habitable-zone orbits. The particular values of six parameters are drawn from random deviates:  $0.7\sqrt{L} \leq a \leq 1.5\sqrt{L}$ ,  $0 \leq e \leq 0.35$ , and  $0 \leq M_0 \leq 2\pi$  (uniformly distributed;  $L$  is the stellar luminosity);  $i$ ,  $\omega_p$ , and  $\Omega$  uniformly distributed on the sphere.

Three parameters are delta functions:  $R_p = R_{\oplus}$ ,  $q$  depends on the filter passband, and  $\Phi$  is the Lambertian phase function.  $T$  is given by  $a$  and the stellar mass via Kepler’s Third Law.

When estimating  $c_{1,j}$ , for the first LSO,  $\Delta\text{mag}$  and  $s$  are computed for all  $N_0$  POIs at, say,  $t = 0$ , when the orbital phases of all POIs in the sample are equal to  $\{M_0\}$ . Thereafter, record is kept of the epoch of each LSO,  $t_{i>1}$ , only for the POIs that have not yet been eliminated and are still in play.

For the  $i$ th LSO, we identify and count the number of POIs that satisfy Equations (1) and (2),  $N_{i,j}$ , producing

$$c_{i,j} = \frac{N_{i,j}}{N_0}. \quad (5)$$

Based on Equation (5), all types of completeness can be computed with any required precision by appropriately choosing the value of  $N_0$ .

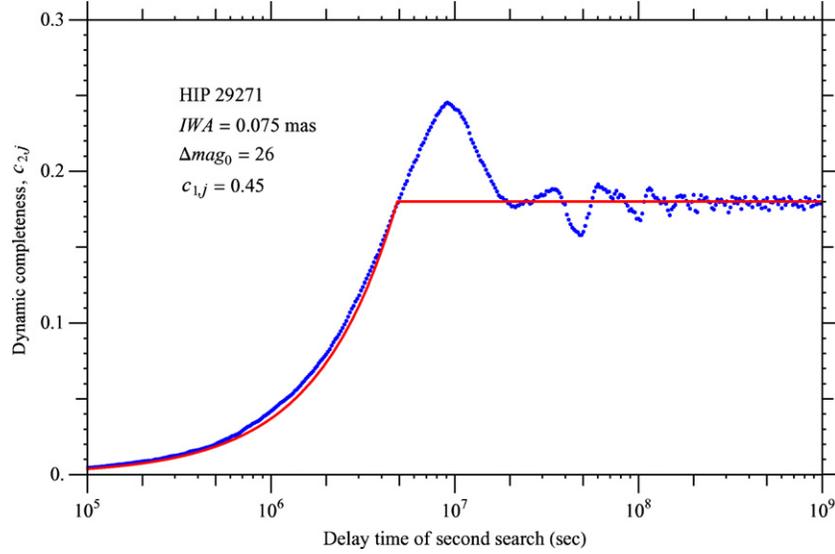
### 2.3. Rebound of $c_{i,j}$

$c_{i+1,j}(t)$  rebounds following the  $i$ th LSO.  $c_{i+1,j}(t_i + \epsilon) = 0$ , where  $\epsilon$  is a diminishingly small increment of time, and then it rebounds toward a constant, asymptotic value,  $c_{i+1,j}(\infty)$ , as the orbits of still-possible POIs lose orbital phase coherence.

The blue points in Figure 1 show how  $c_{2,j}$  rebounds after a first LSO at  $t = 0$  of HIP 29271 for IWA = 0.075 arcsec (*THEIA*; Savransky et al. 2010),  $q = 0.26$ , and  $\Delta\text{mag}_0 = 26$  (typical value; Brown 2005). The rebound resembles the response of the output voltage for an underdamped, series, LRC circuit after a step change in the input voltage. After a linear rise (curved on a logarithmic plot),  $c_{2,j}$  undergoes damped oscillations. The details—rise time, damping time, and asymptotic value  $c_{2,j}(\infty)$ —depend on the star’s particular mix of habitable-zone orbits resolved by IWA. For different stellar mass, luminosity, and distance, for other values of IWA and  $\Delta\text{mag}_0$ , and for other definitions of POIs, the rebound of  $c_{2,j}$  will be qualitatively similar to Figure 1, but with different details and numbers. For example, other factors being equal, a star at larger distance would have a longer rebound time, because the central field obscuration (IWA) would limit the observation to planets with wider separations, and therefore a longer time would be necessary to loose orbit coherence.

When the next observation is being planned, we need to know  $c_{i,j}(t)$  for each star in play, and where  $i - 1$  observations of star  $j$  have already been performed. When  $i = 1$ , we would use tabulated values of time-independent virgin completeness, and no real-time computation would be required. When  $i > 1$ , however, we may need an efficient function or procedure for computing the values of  $c_{i,j}(t)$ .

The most accurate, brute-force method would perform a blue-point-type calculation (see Figure 1) for every star in play every time a new observation is planned. The number of times would be of order the number of stars times the number of observations. For example, the number of blue-point-type calculations would exceed  $10^5$  for a program of 100 stars and 1000 LSOs, typical for a 4 m class instrument with IWA = 0.075 arcsec. Monte Carlo full-mission studies would be impractical, as each of the 400 blue points in Figure 1 took  $\sim 5$  s to compute on a 3 GHz Intel Xenon processor running MATHEMATICA 6. Therefore, we must look at two approximate functions for  $c_{i,j}(t)$ , one of which may be perfectly adequate for first-order scheduling studies. They demand only one or four blue-point-type calculations performed a number of times that is of the order of the number of observations.



**Figure 1.** Typical rebound of dynamic completeness  $c_{2,j}(t)$  following the first limiting search observation (LSO) at  $t = 0$ . Blue: values calculated from Equation (5) with  $N_0 = 20,000$ , for Earth-like, habitable-zone planets around HIP 29271, assuming IWA = 0.075 arcsec,  $q = 0.26$ , and  $\Delta\text{mag}_0 = 26$ . Red: the linear approximation. Here,  $c_{2,j}(\infty) = 0.18$  and brektime =  $5 \times 10^7$  s.

The first alternative approximation is a linear function,  $c_{i,j}(t) = \text{slope} \times t$  for  $t < \text{brektime}$  and  $c_{i,j}(t) = c_{i,j}(\infty)$  for  $t > \text{brektime}$ . This is illustrated by the red curve in Figure 1.

We used the following algorithm to find the parameters slope and brektime  $\equiv c_{i,j}(\infty)/\text{slope}$ . The algorithm comprises four blue-point-type computations of  $c_{i,j}(t)$ . First, we estimate  $c_{i,j}(\infty)$  by computing  $c_{i,j}(10^{10}$  s). Second, we compute  $c_{i,j}(10^{5.5}$  s) and use it to make a first estimate of the slope,  $\text{slope1} \equiv c_{i,j}(10^{5.5}$  s)/ $10^{5.5}$  s. (The starting point  $t = 10^{5.5}$  is somewhat arbitrary. It should be large enough to afford an accurate value of  $c_{i,j}$  according to the counting statistics of Equation (5), but also confidently smaller than the true value of brektime.) Third, we compute a first estimate of brektime,  $\text{brektime1} = c_{i,j}(10^{10}$  s)/ $\text{slope1}$ , and compute  $c_{i,j}(\text{brektime1})$ , which produces second estimates of the brektime and slope:

$$\text{brektime2} \equiv \text{brektime1} - \frac{c_{i,j}(\text{brektime1}) - c_{i,j}(10^{5.5} \text{ s})}{\text{slope1}}, \quad (6)$$

and  $\text{slope2} \equiv c_{i,j}(10^{10})/\text{brektime2}$ . Fourth, we compute  $c_{i,j}(\text{brektime2})$ —the fourth and last blue-point-type computation—and use it to produce the final estimates

$$\text{brektimeFinal} \equiv \text{brektime2} - \frac{c_{i,j}(\text{brektime2}) - c_{i,j}(\text{brektime1})}{\text{slope2}}, \quad (7)$$

and  $\text{slopeFinal} \equiv c_{i,j}(10^{10} \text{ s})/\text{brektimeFinal}$ . (The linear function required about  $\sim 20$  s to calculate on a 3 GHz Intel Xenon processor running MATHEMATICA 6.)

The second alternative approximation is a step function:  $c_{i,j}(t) = 0$  for  $t < \text{brektime}$  and  $c_{i,j}(t) = c_{i,j}(\infty)$  for  $t > \text{brektime}$ , where  $\text{brektime} \sim 10^7$  s for habitable-zone orbits and popular instrument concepts. For first-order investigations, accurate knowledge of the completeness rebound may not be important to the outcome. The important thing is to avoid the mistake that brektime is zero or too small, an error which may

cause bogus observations to pile up on the high-completeness, low-exposure-time stars. (The step function required  $\sim 5$  s to calculate on a 3 GHz Intel Xenon processor running MATHEMATICA 6.)

In Figure 2, the first (linear) approximation of  $c_{i,j}(t)$  is used to illustrate a hypothetical program of five LSOs to HIP 29271 starting at absolute time  $t = 10^6$  s. (The abscissa here is now linear.) Such approximations would be used only in the scheduling process of a design reference mission (DRM), where it may be necessary to compute dynamic completeness for many stars on the fly. After the decision to observe a particular star, Equation (5) would be used to produce an accurate value for the record. Note that as the number of LSOs of a star increases, the accumulate completeness converges on the ultimate completeness,  $C_{\infty,j}$ .

### 3. THE PROBABILITY OF A DISCOVERY BY THE NEXT LSO, $\mathcal{P}$

After  $i - 1$  unproductive LSOs of star  $j$ ,  $(1 - C_{i-1,j})$  is the fraction of all possible POIs that have not been ruled out and are still in play. The fraction of the *remaining* POIs that the next ( $i$ th) LSO at time  $t$  would detect is

$$\mathcal{K}_{i,j}(t) = \frac{c_{i,j}(t)}{(1 - C_{i-1,j})}. \quad (8)$$

The probability of a discovery on the next LSO of star  $j$  is

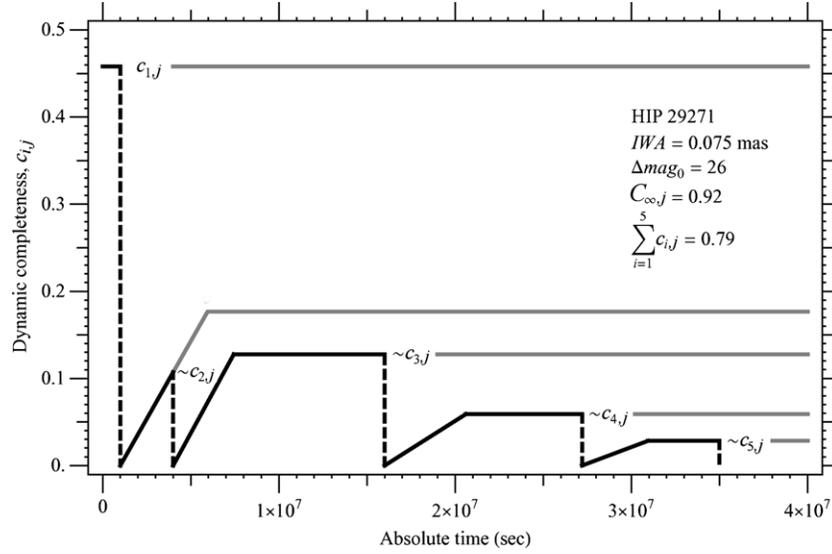
$$\mathcal{P}_{i,j} = \eta'_j \mathcal{K}_{i,j}(t), \quad (9)$$

where  $\eta'_j$  is the Bayesian correction of the occurrence probability  $\eta$ .  $\eta'_j$  is the probability that star  $j$  possesses a POI after taking into account the contrary evidence of  $n_j$  previous unproductive LSOs with accumulated completeness  $C_{i-1,j}$ .

Bayes's theorem states

$$\eta'_j \equiv P(\mathcal{H}|\mathcal{E}) = \frac{P(\mathcal{E}|\mathcal{H})P(\mathcal{H})}{P(\mathcal{E})} = \frac{(1 - C_{i-1,j})\eta}{1 - \eta C_{i-1,j}}, \quad (10)$$

where the hypothesis  $\mathcal{H}$  is that star  $j$  has a POI; the evidence  $\mathcal{E}$  is the lack of a discovery so far;  $P(\mathcal{E}|\mathcal{H}) = (1 - C_{i-1,j})$  is the



**Figure 2.** Dynamic completenesses of five hypothetical LSOs of HIP 29271 at  $t = 10^6$ ,  $4 \times 10^6$ ,  $1.6 \times 10^7$ ,  $2.7 \times 10^7$ , and  $3.5 \times 10^7$  s. The virgin completeness  $c_{1,j} = 0.46$  is an accurate value, independent of time and computed in advance of simulations, using Equation (5). The symbols  $c_{2-5,j}$  are preceded by tildes to indicate they are approximations. Such approximations would be used only in the selection process of a DRM, where it may be necessary to compute dynamic completeness for many stars on the fly. After the decision to observe a particular star, Equation (5) would be used to produce an accurate value for the record. Note that as the number of LSOs of a star increases, the accumulate completeness converges on the ultimate completeness,  $C_{\infty,j}$ .

conditional probability of  $\mathcal{E}$  if  $\mathcal{H}$  is true;  $P(\mathcal{H}) = \eta$  is the prior probability of  $\mathcal{H}$ ; and the marginal probability of  $\mathcal{E}$  is

$$\begin{aligned} P(\mathcal{E}) &= P(\mathcal{E}|\mathcal{H})P(\mathcal{H}) + P(\mathcal{E}|\overline{\mathcal{H}})P(\overline{\mathcal{H}}) \\ &= (1 - C_{i-1,j})\eta + (1 - \eta) = 1 - \eta C_{i-1,j}, \end{aligned} \quad (11)$$

where  $\overline{\mathcal{H}}$  is the hypothesis that star  $j$  does not have a POI. ( $P(\mathcal{E}|\overline{\mathcal{H}}) = 1$  and  $P(\overline{\mathcal{H}}) = 1 - \eta$ .)

The result for  $\mathcal{P}_{i,j}$  is

$$\mathcal{P}_{i,j} = \frac{\eta c_{i,j}(t)}{1 - \eta C_{i-1,j}}. \quad (12)$$

For the case of no prior searches ( $i = 1$ ,  $C_{0,j} \equiv 0$ ), the probability of a discovery on the first search is  $\eta c_{1,j}$ , as expected.

We want to confirm numerically that Equation (12) accurately estimates the probability of a discovery by the next LSO for random values of the parameters, for example,  $\eta = 0.272673$ ,  $C_{1,j} = 0.437671$ , and  $c_{2,j} = 0.506385$ , for which  $\mathcal{P}_{2,j} = 0.156789$ . For this verification, we conduct  $N_{\text{obs}} = 200,000$  independent LSOs, each involving  $N_0 = 200,000$  possible POIs. Each LSO involves the following computational steps.

1. Randomly pick the serial number of the “real” POI:  $n = \mathcal{B}(\eta)\mathcal{I}(1, N_0)$ , where  $\mathcal{B}$  is a Bernoulli random deviate with probability  $\eta$ , which yields the value 0 or 1, and  $\mathcal{I}$  is a uniform random deviate producing an integer in the range 1 to  $N_0$ . (If the serial number is zero, it means the “star” being observed does not have a POI.)
2. Perform a first LSO by selecting  $N_1 = \text{ROUND}(C_{1,j}N_0)$  random integers in the range 1 to  $N_0$ , where the ROUND function yields the closest integer to the argument. If  $n$  is one of these  $N_1$  integers, then go back and repeat Steps 1 and 2, because we only want cases where the first LSO does not make a “discovery.”
3. Perform a second LSO by randomly selecting  $N_2 = \text{ROUND}(c_{2,j}N_0)$  integers from the  $N_0 - N_1$  integers defined by excluding the  $N_1$  integers in Step 2 from the set of all integers 1 to  $N_0$ . If  $n$  is equal to one of these  $N_2$  integers, then we have a discovery; otherwise, not.

4. Repeat Steps 1 to 3  $N_{\text{obs}}$  times and count the number of discoveries,  $N_{\text{disc}}$ .

5. Compute the empirical probability,  $N_{\text{disc}}/N_{\text{obs}}$ , which was 0.156845, 0.157050, and 0.156695 in three runs we performed using the parameters above. These values compare well with the theoretical value,  $\mathcal{P}_{2,j}$  and confirm Equation (12). (As a benchmark, one run required 20,000 s on a 3 GHz Intel Xenon processor running MATHEMATICA 6.)

#### 4. APPLICATIONS OF $\mathcal{P}$ TO OBSERVING PROGRAMS

Two applications of  $\mathcal{P}$  and Equation (12) must be sharply distinguished. The first application is in the scheduling algorithm for real or simulated observing programs, where we use the discovery rate,

$$\mathcal{Z}_{i,j} = \frac{\mathcal{P}_{i,j}}{t_{\text{exp}} + \text{OH}}, \quad (13)$$

as a merit function or science benefit/cost metric for optimizing the observing program for discovery. In this application,  $\eta$  in Equation (12) is  $\eta_{\text{ops}}$ . (OH is any observational overhead time that will be charged to the program, such as for calibration or alignment.)

The second application of Equation (12) is in estimating the probability distribution of discoveries for simulated observing programs, as discussed below. In this application,  $\eta = \eta_{\text{true}}$ , where  $\eta_{\text{true}}$  is the “true” value. In this context,  $\eta_{\text{true}}$  is a control parameter. If  $\eta_{\text{ops}} \neq \eta_{\text{true}}$ , the number of planets discovered by an observing program may be less than optimal, because the scheduling algorithm may not always choose, as the next star to observe, the qualified star with the highest “true” value of the merit function.

In both applications of Equation (12),  $C_{i-1,j}$  is the accumulated completeness from all LSOs of star  $j$  prior to the  $i$ th, each contribution computed as accurately as desired from Equation (5).

#### 4.1. Probability Distribution of the Number of Discoveries, pdf( $m$ )

No matter what star  $j$ , nor what search  $i$  of that star, the  $k$ th LSO in the overall observing program discovers  $u$  planetary systems, where  $u \in \{0, 1\}$  is a Bernoulli random variable with probability  $\mathcal{P}_k$ , as given in Equation (12) using  $\eta_{\text{true}}$  and the indices  $i$  and  $j$  corresponding to the  $k$ th LSO. (For each star ( $j$ ), multiple visits ( $i$ ) are possible, so both  $i$  and  $j$  define the  $k$ th LSO.) The probability density function (pdf) of  $u$  for the  $k$ th LSO is

$$\text{pdf}_k(u) = (1 - \mathcal{P}_k)\delta(u, 0) + \mathcal{P}_k\delta(u, 1), \quad (14)$$

where  $\delta(i, j) \equiv 1$  for  $i = j$ , and zero otherwise, is the Kronecker delta.

An entire observing program, consisting of  $n_{\text{total}}$  LSOs, where

$$n_{\text{total}} = \sum_{j=1}^{n_{\text{stars}}} n_j, \quad (15)$$

where  $n_j$  is the total number of LSOs of star  $j$  and  $n_{\text{stars}}$  is the total number of stars observed, discovers  $m$  planets, where  $m \in \{0, 1, \dots, n_{\text{total}}\}$  is the sum of  $n_{\text{total}}$  Bernoulli random variables, each with  $\text{pdf}_k$  given by Equation (14). Therefore, the pdf of  $m$  is the convolution ( $\star$ ) of  $\text{pdf}_k$  for all  $k$ :

$$\text{pdf}(m) = \text{pdf}_1(u) \star \text{pdf}_2(u) \star \dots \star \text{pdf}_{n_{\text{total}}}(u), \quad (16)$$

where each successive convolution has the form

$$(\text{pdf}_k \star \mathcal{F})(n) \equiv (1 - \mathcal{P}_k)\mathcal{F}(n) + \mathcal{P}_k\mathcal{F}(n - 1). \quad (17)$$

Equation (16) offers a practical advantage for estimating the outcome of a search program. Starting from a *single* simulated observing program, it allows us to estimate theoretically the pdf of the total number of discoveries. The alternative—running many full simulations to build up an empirical estimate of the pdf from the discovery results—is much less efficient.

#### 4.2. Estimation of $\eta$

The pertinent record of a real or simulated observing program is a set of  $n_{\text{total}}$  data triplets re-indexed from  $i$  (LSO) and  $j$  (star) to  $k$  (observation):

$$\{c_k, C_k, u_k\} \equiv \{c_{i,j}, C_{i-1,j}, u_{i,j}\}, \quad (18)$$

for  $1 \leq k \leq n_{\text{total}}$ , where  $u_{i,j} = 0$  or  $1$  is the number of discoveries by the  $i$ th LSO of the  $j$ th star. (If we assume that we stop searching after a discovery,  $u_{i,j} = 1$  for at most one value of  $i$  for any  $j$ .)

The logarithmic likelihood function  $\mathcal{L}$  is the logarithm of the probability of the set  $\{u_k\} \equiv \{u_{i,j}\}$  as a function of  $\eta$ :

$$\begin{aligned} \mathcal{L}(\{u_k\}|\eta) &= \sum_{k=1}^{n_{\text{total}}} \ln p_{u_k} = \sum_{k=1}^{n_{\text{total}}} \ln \left( \frac{c_k \eta}{1 - C_k \eta} u_k \right. \\ &\quad \left. + \left( 1 - \frac{c_k \eta}{1 - C_k \eta} \right) (1 - u_k) \right). \end{aligned} \quad (19)$$

The maximum-likelihood estimate of the occurrence probability,  $E(\eta)$ , is the  $\eta$ -root of the equation

$$\frac{\partial \mathcal{L}(\{u_k\}|\eta)}{\partial \eta} = 0. \quad (20)$$

The minimum variance bound (MVB) is the inverse of the Fisher information near  $E(\eta)$ :

$$\text{MVB}(E(\eta)) = \left( - \frac{\partial^2 \mathcal{L}(\{u_k\}|\eta)}{\partial \eta^2} \Big|_{\eta=E(\eta)} \right)^{-1}. \quad (21)$$

We want to confirm that Equations (19)–(21) accurately estimate  $\eta$  and its variance. To this end, we performed a numerical experiment simulating 100,000 missions of 100 LSOs, according to the following steps.

1. Generated  $n_{\text{total}} = 100$  random data triplets for the left side of Equation (18) as follows:

$$c = \mathcal{R}, \quad (22)$$

$$C = (1 - c)\mathcal{R}, \quad (23)$$

$$u = \mathcal{B} \left( \frac{\eta c}{1 - \eta C} \right), \quad (24)$$

where  $\mathcal{R}$  is a uniform random deviate on the range 0–1,  $\mathcal{B}(p)$  is a Bernoulli random deviate with probability  $p$ , and  $\eta = 0.10$ .

2. Compute  $E(\eta)$  and  $\text{MVB}(E(\eta))$  using Equations (19)–(21).

For the sample of 100,000 trials, we found

$$\langle E(\eta) \rangle = 0.101, \quad (25)$$

$$\sigma_\eta = 0.039, \quad (26)$$

and

$$\langle \sqrt{\text{MVB}(E(\eta))} \rangle = 0.042. \quad (27)$$

Equation (25) is the mean value of  $\eta$  found using Equation (20) in the 100,000 simulated missions. Equation (26), the standard deviation of those values of  $\eta$ , is the empirical estimate of the scatter in  $\eta$  determined by Equation (20). The square of Equation (26) is the empirical variance. The Cramér–Rao theoretical limit on the variance of any estimator of  $\eta$  is the MVB given by Equation (21). We computed the MVB for each simulated mission, and Equation (27) gives the mean value of the square root of the MVB for the suite of 100,000 simulated missions—computed for direct comparison with the empirical value in Equation (26).

These results illustrate that the maximum-likelihood estimator accurately recovers  $\eta$  from a record of the results of observing programs and that the accuracy of this estimator appears to approach the Cramér–Rao limit ( $\sigma^2 \approx \text{MVB}$ ). (This Monte Carlo experiment required 270 s on 56 2.66 GHz Intel Xenon processors operating in parallel.)

## 5. ILLUSTRATIVE DESIGN REFERENCE MISSIONS (DRMs)

The purpose of a DRM is to gauge the science operations of a mission concept. To illustrate the new completeness methods introduced in this paper, we now describe a ministudy using simple DRMs to explore and measure the power of the *James Webb Space Telescope* (JWST) to discover and

**Table 1**

Spectrophotometric Parameters for Calculating Completenesses and Exposure Times

Instrument	Mode	Nominal $\lambda$ (nm)	Resolving Power	Zero Point (Jy)	Earth Geometric Albedo ( $q$ )
	F070W	700	4	3043	0.232
	F115W	1150	4	1766	0.187
NIRCam	F140M	1400	10	1324	0.021
	F150W	1500	4	1188	0.103
	F162M	1625	10	1045	0.179
NIRSpec	Prism	1150	31.7	1766	0.260

**Notes.** Zero points interpolated from values for *VJHK* filters (Leinert et al. 1998). We adjusted Earth’s effective albedo for prism spectroscopy at 1150 nm.

characterize Earth-like extrasolar planets using a starshade to suppress scattered starlight (Cash et al. 2009; Soummer et al. 2009b). In this scenario, *JWST* and the starshade revolve in coordinated orbits around the second Earth–Sun Lagrange point, L2. The starshade operates on a  $\sim 70,000$  km sphere centered on *JWST*. In a 3-year planet-finding mission, we assume enough propulsion to slew the starshade 70 times to take up new positions between *JWST* and target stars. We want the DRMs to tell us about the science, for example, how many discoveries to expect if we optimize the observing program, assuming  $\eta_{\text{true}} = \eta_{\text{ops}} = 0.3$ , say.

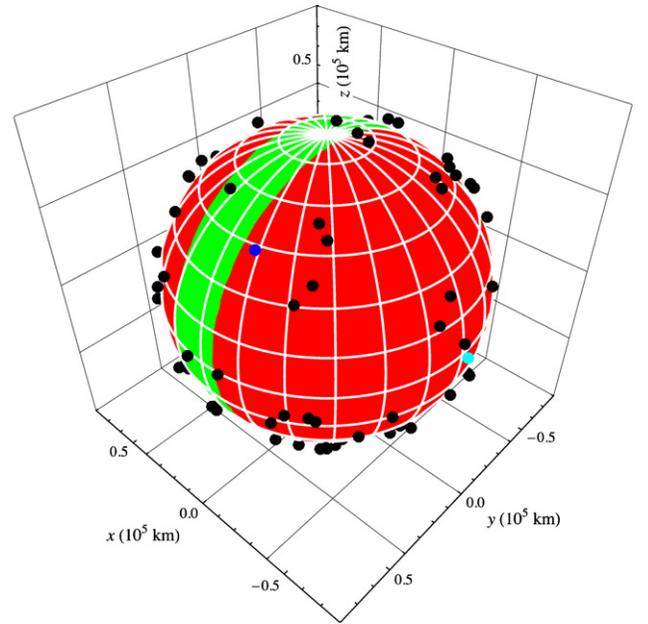
Other DRM inputs include: a science strategy; a definition of POIs (same as Section 2.2, with  $q$  depending on filter as given in Table 1); IWA = 0.085 arcsec,  $\Delta\text{mag}_0 = 26$ , and pointing restrictions  $\gamma_1 = 85^\circ$  (solar avoidance) and  $\gamma_2 = 105^\circ$  (starshade bright-side avoidance); an input catalog of stars; exposure time calculators; typical overheads OH; and a merit function—in this case, the discovery rate  $\mathcal{Z}$ —which we use to select the next star to search.

The science strategy is simple: we perform an LSO, and if a likely POI is discovered, we immediately perform additional spectrophotometry to characterize the body. Such immediate follow-up reduces the risk of a newly discovered POI becoming undetectable before it can be characterized and avoids the difficulty of trying to recover it with inadequate knowledge of its orbit (Brown et al. 2007). If we discover a POI, we cease further LSOs of that star. If we do not find one, we move the starshade to the next target star, but we return to a star already searched if and when it once again offers the highest value of  $\mathcal{Z}$ .

The LSO is a deep image using whichever NIRCam filter offers maximum  $\mathcal{Z}$ . We call this filter the preferred filter, and it varies from star to star. The possible filters are listed in Table 1.

After an LSO finds a potential POI, we obtain images through the four *non-preferred* NIRCam filters for that star, and take a low-resolution spectrum with NIRSpec. We use exposure time calculators for NIRCam and NIRSpec, based on parameters from the instrument teams, to achieve  $S/N = 5$  on a source of magnitude  $\text{mag}_j + \Delta\text{mag}_0$  for the LSOs, and  $\text{mag}_j + \Delta\text{mag}_{\text{median}}$  for the follow-up filter photometry and spectroscopy, where  $\text{mag}_j$  is the apparent magnitude of star  $j$ , and  $\Delta\text{mag}_{\text{median}}$  is the median magnitude difference between the star and the universe of possibly detected POIs for that star. (If  $\Delta\text{mag}_{\text{median}} > \Delta\text{mag}_0$ , we use  $\Delta\text{mag}_0$ .)

At this stage of the study we do not try to refine our understanding of the exposure time calculation beyond the current estimates by the instrument teams (M. Rieke et al. 2009, private communication; P. Jakobsen et al. 2009, pri-



**Figure 3.** Sphere of starshade operations, centered on *JWST*, shown here on the vernal equinox. Green: permitted pointings for  $\gamma_1 = 85^\circ$  and  $\gamma_2 = 105^\circ$ . Red: forbidden pointings. As seen from above and as time passes, stars revolve on the starshade sphere around the  $+\hat{z}$  axis (north ecliptic pole) in the clockwise direction. Depending on a star’s ecliptic latitude  $b$ , it may be observable for one or two periods per year, or for the entire year ( $b > 85^\circ$ ). Blue: a typical *TPF-C* target star from Brown (2005, HIP 92043,  $b = +43^\circ.4$ ). Black: other *TPF-C* stars. Cyan: the Sun, which is fixed in this L2 coordinate system ( $-\hat{x}$  toward the Sun,  $+\hat{y}$  in the direction of Earth’s orbital motion).

vate communication). We use standard parameters based on instrument requirements. There may be better observing modes for this particular application (e.g., involving detector sub-arrays).

We interpolate standard stellar magnitudes and zero points to the effective wavelengths of each NIRCam filter and the NIRSpec prism, starting from the *VJHK* magnitudes from NStED ([nsted.ipac.caltech.edu/](http://nsted.ipac.caltech.edu/)) and the *VJHK* zero points from Leinert et al. (1998). We used a near-infrared spectrum of Earth calculated by S. Seager (2010, private communication) to estimate the effective geometric albedo  $q$  of Earth at the wavelengths and resolving powers of each instrument modes. These parameters are listed in Table 1.

As illustrated in Figure 3, each potential target star is continuously observable only for a limited period of time, once or twice a year, depending on its ecliptic latitude ( $b$ ). The total time costs of observing a star—itemized in Table 2—must fit into a single observability period for that star. Some 26 of the 117 target stars (Brown 2005) used in an earlier study of the coronagraphic *Terrestrial Planet Finder* (*TPF-C*) are qualified according to this criterion. These stars constitute the input catalog for these DRMs (Table 3).

To determine the preferred filter for LSOs, we use Equation (5) and the procedures of Section 2.2 to calculate virgin completeness  $c_{1,j}$  for each of the 26 qualified stars using samples of  $N_0 = 40,000$  POIs. We do this separately for each of the five filters, because of the dependence of  $q$  on wavelength and resolving power. Next, we use Equations (12) and (13) to compute  $\mathcal{P}_{1,j}$  and  $\mathcal{Z}_{1,j}$ , using OH = 10 hr as our estimate of the time cost of the fine alignment of the starshade, which we assume is incurred by each new observation with a different instrument, or of a different star. For each star  $j$ , we select the

**Table 2**  
Observing Sequence and Typical Time Costs for a Single Visit of the Starshade to a Target Star

Step	Activity	Observing Time Cost	Clock Time Cost
1	Final alignment of <i>JWST</i> , starshade, and target star	10 hr	10 hr
2	LSO through preferred NIRCcam filter	$4.8 \times 10^4$ s	$4.8 \times 10^4$ s
3	Analyze data for discovery		7 days
4	Final alignment of <i>JWST</i> , starshade, and target star	10 hr	10 hr
5	Images through four non-preferred NIRCcam filters	$1.4 \times 10^5$ s	$1.4 \times 10^5$ s
6	Analyze data to prepare for spectroscopy		7 days
7	Final alignment of <i>JWST</i> , starshade, and target star	10 hr	10 hr
8	Spectrum using NIRSpec	$1.9 \times 10^5$ s	$1.9 \times 10^5$ s
	Total with discovery	$4.9 \times 10^5$ s	$1.7 \times 10^6$ s
	Total without discovery	$8.4 \times 10^4$ s	$6.9 \times 10^5$ s

**Notes.** Overhead costs (1, 3, 4, 6, 7) are fixed. Exposure times (2, 5, 8) are median values for the 26 stars in the input catalog. With no discovery in step 3, only steps 1–3 are executed during a visit to a target star. With a discovery, all the eight steps are executed. During the analysis steps (3, 6), *JWST* conducts observations for other programs; only the starshade remains aligned with the target star for this program.

**Table 3**  
The Input Catalog of Target Stars, Ranked in Descending Order of the Discovery Rate on the First LSO ( $\mathcal{Z}_1$ )

Disc. Rate	HIP	Type	$L$	Distance (pc)	$b(^{\circ})$	LSO Filter	LSO $t_{\text{exp}}$	Max Time	Max Cont. Obs.	$c_1$	$C_{\infty}$	$\mathcal{P}_1$	$\mathcal{Z}_1$
1	71681	K1 V	0.61	1.35	−43	F070W	3.2	6.12	6.38	0.8	1	0.24	−5.19
2	8102	G8 V	0.47	3.65	−25	F115W	4.03	6.13	6.29	0.76	1	0.23	−5.31
3	71683	G2 V	2.2	1.35	−43	F070W	2.43	6.12	6.38	0.55	1	0.17	−5.34
4	3821	G0 V	1.2	5.95	47	F115W	3.94	6.14	6.41	0.59	1	0.18	−5.4
5	99240	G6/8 IV	1.5	6.11	−45	F115W	4.12	6.16	6.4	0.56	1	0.17	−5.47
6	108870	K4/5 V	0.2	3.63	−41	F150W	4.48	6.13	6.37	0.65	1	0.19	−5.53
7	22449	F6 V	2.6	8.03	−15	F115W	3.88	6.15	6.26	0.39	0.96	0.12	−5.57
8	19849	K0/1 V	0.41	5.04	−28	F115W	4.69	6.16	6.3	0.64	1	0.19	−5.64
9	15510	G8 III	0.71	6.06	−58	F115W	4.68	6.18	6.53	0.63	1	0.19	−5.65
10	2021	G1 IV	3.9	7.47	−65	F070W	4.2	6.15	6.63	0.34	0.92	0.1	−5.7
11	27072	F6.5 V	2.3	8.97	−46	F115W	4.47	6.25	6.4	0.41	0.98	0.12	−5.73
12	1599	G0 V	1.2	8.59	−58	F115W	4.71	6.25	6.53	0.5	1	0.15	−5.76
13	64394	G0	1.3	9.15	33	F115W	4.81	6.3	6.32	0.47	1	0.14	−5.85
14	57757	F8	3.4	10.9	0.69	F115W	4.33	6.23	6.24	0.27	0.88	0.08	−5.85
15	12777	F8	2.2	11.2	32	F115W	4.64	6.31	6.32	0.36	0.99	0.11	−5.87
16	96100	G9 V	0.41	5.77	81	F115W	4.99	6.2	7.42	0.56	0.97	0.17	−5.9
17	105858	F7 V	1.4	9.22	−47	F115W	4.91	6.33	6.41	0.47	1	0.14	−5.92
18	73184	K4 V	0.26	5.91	−4.4	F150W	5.05	6.2	6.25	0.34	0.76	0.1	−6.17
19	70497	F8	3.9	14.6	60	F115W	4.75	6.41	6.56	0.17	0.81	0.051	−6.26
20	23693	F6/7 V	1.4	11.7	−79	F115W	5.18	6.53	7.42	0.33	0.91	0.098	−6.29
21	77952	F0 III/IV	9.6	12.3	−42	F070W	4.3	6.23	6.38	0.073	0.4	0.022	−6.41
22	29271	G6 V	0.83	10.2	−82	F115W	5.41	6.53	7.43	0.32	0.81	0.095	−6.49
23	114622	K3 V	0.28	6.53	55	F150W	5.34	6.27	6.49	0.27	0.68	0.081	−6.5
24	50954	F2/3 IV/V	5.0	16.2	−68	F115W	4.98	6.57	6.7	0.11	0.68	0.032	−6.61
25	40702	F5 V	6.8	19.4	−75	F070W	5.29	6.66	7.41	0.076	0.61	0.023	−7.01
26	86614	F5	5.5	22.00	84	F070W	5.6	6.93	7.46	0.048	0.67	0.014	−7.48

**Notes.** Times in log seconds. “Max time” is the time cost of the full observing sequence in Table 1. “Max cont. obs.” is the maximum continuous observing time for ecliptic latitude  $b$  for  $\gamma_1 = 85^{\circ}$  and  $\gamma_2 = 105^{\circ}$ .  $\mathcal{P}_1$  is the probability of a discovery on the first LSO, from Equation (12), and  $\mathcal{Z}_1$  is  $\mathcal{P}_1$  divided by the sum of LSO  $t_{\text{exp}}$  and 10 hr for alignment, from Equation (13). The value of  $\mathcal{Z}_1$  is given in log discoveries per second.

filter with the highest value of  $\mathcal{Z}_{1,j}$  as the preferred filter for LSOs of that star (listed in Table 3).

Next, we determine the universe of possible LSOs. For the preferred filter only, we continue to compute  $c_{i>1,j}$  using Equation (5) until the sample of  $N_0$  POIs is effectively exhausted. This yields 26 lists of dynamic completenesses, in sequence, one list for each of the 26 stars—some 2075 values of  $c_{i,j}$  in all. Again using Equations (12) and (13), we convert these lists of  $c_{i,j}$  into a full list of possible LSOs in the form of vectors  $\{\text{HIP}_k, i_k, \mathcal{P}_k, \mathcal{Z}_k\}$ , where  $1 \leq k \leq 2075$  is the index for LSOs introduced in Section 4.1, and the items are the *Hipparcos*

number of the star, the number of this visit to that star, and the discovery probability and rate for that visit.

At the start of a DRM, the prioritized observing program is the list of 2075 LSO vectors sorted in descending order of  $\mathcal{Z}_k$ . (Table 4 lists the top 80 LSOs for the current illustration.) We expect to perform only 70 LSOs—but we do not know which ones. How far down the list a DRM reaches is determined by the random discoveries as the DRM unfolds. That is, we determine the outcome of each LSO in turn—discovery, yes or no, with possible follow-on observations and alternative time costs in Table 2—by interrogating a Bernoulli random deviate with

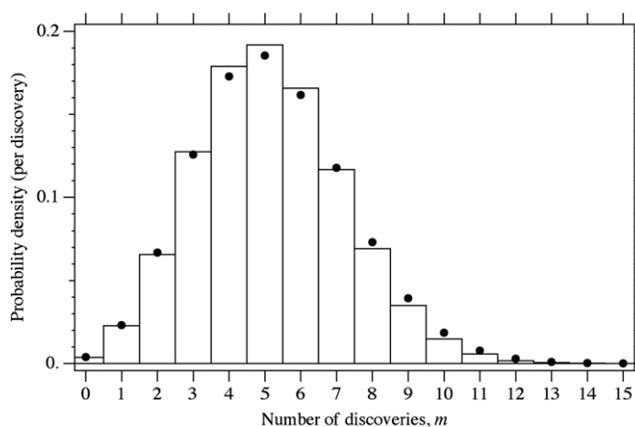
**Table 4**  
Record of a Typical DRM

$k$	$k_{\text{DRM}}$	HIP	$i$	$\mathcal{P}_k$	$\mathcal{Z}_k$	Discovery	$k$	$k_{\text{DRM}}$	HIP	$i$	$\mathcal{P}_k$	$\mathcal{Z}_k$	Discovery
1	1	71681	1	0.24	-5.19	No	41	37	57757	3	0.039	-6.17	No
2	2	8102	1	0.23	-5.31	Yes	42	38	105858	2	0.075	-6.19	No
3	3	71683	1	0.17	-5.34	No	43	...	71683	4	0.022	-6.22	...
4	4	3821	1	0.18	-5.40	No	44	39	12777	3	0.046	-6.24	No
5	5	99240	1	0.17	-5.47	Yes	45	40	2021	4	0.029	-6.26	No
6	6	108870	1	0.19	-5.53	No	46	41	70497	1	0.051	-6.26	No
7	7	22449	1	0.12	-5.57	No	47	42	23693	1	0.098	-6.29	No
8	8	19849	1	0.19	-5.64	No	48	43	57757	4	0.029	-6.30	No
9	9	15510	1	0.19	-5.65	No	49	44	1599	3	0.043	-6.31	No
10	10	71683	2	0.079	-5.66	Yes	50	45	96100	2	0.065	-6.31	No
11	11	2021	1	0.10	-5.70	No	51	46	108870	3	0.032	-6.32	No
12	12	27072	1	0.12	-5.73	No	52	...	8102	3	0.022	-6.32	...
13	13	1599	1	0.15	-5.76	No	53	47	27072	4	0.030	-6.34	No
14	14	3821	2	0.077	-5.77	No	54	48	4394	3	0.045	-6.35	No
15	15	22449	2	0.070	-5.79	No	55	49	71681	3	0.017	-6.35	No
16	...	99240	2	0.079	-5.79	...	56	...	99240	4	0.022	-6.35	...
17	16	71681	2	0.054	-5.84	No	57	50	22449	5	0.019	-6.36	No
18	17	64394	1	0.14	-5.85	No	58	51	70497	2	0.040	-6.37	No
19	18	57757	1	0.080	-5.85	No	59	...	3821	4	0.019	-6.37	...
20	19	12777	1	0.11	-5.87	No	60	52	15510	3	0.035	-6.38	No
21	...	8102	2	0.059	-5.90	...	61	53	12777	4	0.033	-6.39	No
22	20	96100	1	0.17	-5.90	No	62	54	2021	5	0.020	-6.41	Yes
23	21	2021	2	0.063	-5.91	No	63	55	77952	1	0.022	-6.41	No
24	22	105858	1	0.14	-5.92	No	64	56	105858	3	0.045	-6.42	No
25	23	27072	2	0.073	-5.96	No	65	57	19849	3	0.032	-6.42	No
26	...	71683	3	0.040	-5.96	...	66	58	57757	5	0.020	-6.45	No
27	24	22449	3	0.044	-6.00	No	67	59	70497	3	0.031	-6.48	No
28	25	108870	2	0.065	-6.01	No	68	60	29271	1	0.095	-6.49	No
29	26	57757	2	0.055	-6.02	No	69	61	73184	2	0.048	-6.49	No
30	27	1599	2	0.075	-6.06	No	70	62	22449	6	0.014	-6.49	No
31	28	12777	2	0.069	-6.07	No	71	...	71683	5	0.012	-6.50	...
32	29	15510	2	0.071	-6.07	No	72	63	23693	2	0.060	-6.50	No
33	30	3821	3	0.038	-6.08	Yes	73	64	114622	1	0.081	-6.50	No
34	...	99240	3	0.041	-6.08	...	74	65	27072	5	0.021	-6.50	No
35	31	2021	3	0.042	-6.09	No	75	66	1599	4	0.026	-6.52	No
36	32	19849	2	0.068	-6.09	No	76	67	77952	2	0.017	-6.52	No
37	33	64394	2	0.074	-6.13	No	77	68	12777	5	0.022	-6.56	No
38	34	27072	3	0.047	-6.15	No	78	69	57757	6	0.016	-6.56	No
39	35	22449	4	0.030	-6.16	No	79	70	64394	4	0.028	-6.56	No
40	36	73184	1	0.10	-6.17	No	80	...	2021	6	0.014	-6.57	...

**Notes.**  $k$ : order of top 80 LSOs in terms of discovery rate  $\mathcal{Z}$ ;  $k_{\text{DRM}}$ : order of 70 LSOs executed in this example DRM;  $i$ : number of this LSO of this star;  $\mathcal{P}_k$ : probability of discovering a POI in this LSO;  $\mathcal{Z}_k$ : log discovery rate of this LSO; ellipses indicate that this LSO was not performed in this DRM due to a previous discovery.

probability  $\mathcal{P}$  and interpreting “1” as a discovery and “0” as no discovery. Each discovery deletes from the observing program all the pending LSOs of a star—meaning those with  $i$  greater than the visit number  $i$  of the LSO that produced the discovery. These deletions promote the lower priority LSOs of other stars into higher positions on the list. To investigate this behavior and its ramifications, we conduct a Monte Carlo experiment of 500,000 DRMs.

Table 4 provides an example DRM in the form of the LSOs actually executed in one DRM run and the discoveries actually made—five in this case. We use it to illustrate the method of deriving the theoretical probability distribution of  $m$  from the probabilities  $\mathcal{P}_k$  of a single DRM. We collect the 70 values of  $\mathcal{P}_{k_{\text{DRM}}}$  for  $1 \leq k_{\text{DRM}} \leq 70$  in Table 4 and follow the recipe in Section 4.1 to obtain the theoretical pdf( $m$ ) represented by the dots in Figure 4. For comparison, the histogram in Figure 4 shows the empirical pdf( $m$ ) based on the actual values of  $m$  from 500,000 DRMs.



**Figure 4.** Comparison of the theoretical and empirical pdf( $m$ ). Dots: the theoretical pdf estimated from the 70 values of  $\mathcal{P}_k$  of the actual LSOs in the single example DRM in Table 4. Histogram: the empirical pdf from the actual results for  $m$  in 500,000 DRMs. (The histogram has been normalized to 1.0.)

**Table 5**

DRM Results for  $\langle m \rangle$ ,  $\sigma_m$ , and  $\sigma_{\langle m \rangle}$  for the Expected Number of POIs Discovered

DRM	Empirical			Theoretical		
	$\langle m \rangle$	$\sigma_m$	$\sigma_{\langle m \rangle}$	$\langle m \rangle$	$\sigma_m$	$\sigma_{\langle m \rangle}$
Example DRM	5.132	...	...	5.21	2.15	...
500,000 DRMs	5.132	2.06	0.003	5.129	0.09	0.0001

**Notes.** Empirical results are based on counting the number of discoveries in Monte Carlo experiments where discoveries were decided by Bernoulli random deviates. Theoretical results use the methods of Section 4.1 to estimate pdf( $m$ ) from the individual LSO probabilities ( $\mathcal{P}_k$ ), and the results are—or are derived from—that distribution. (The highest precision is achieved by the mean of the 500,000 individual theoretical results for  $\langle m \rangle$ .)

Table 5 compares the means, standard deviations, and standard deviations of the means— $\langle m \rangle$ ,  $\sigma_m$ , and  $\sigma_{\langle m \rangle}$ —for the results ensuing from Table 4, which encompass both the example DRM and the 500,000 DRMs computed from the same suite of potential or actual LSOs ( $1 \leq k \leq 2075$  or  $1 \leq k_{\text{DRM}} \leq 70$ ). The highest precision is achieved by the mean of the 500,000 individual theoretical results for  $\langle m \rangle$ . We see that our new method produces accurate theoretical estimates of  $\langle m \rangle$  and  $\sigma_m$  from a single DRM and avoids the onerous alternative, which is to conduct a large number of DRMs to obtain empirical results.

The 500,000 DRMs ensuing from the prioritized LSOs in Table 4 offer additional insights into the science operations of the *JWST* starshade mission. The most likely number of unique stars searched is 23 (ranges 22–25). The median number of visits per star is 3. The total *JWST* observing time used by a DRM is  $(7.6 \pm 1.3) \times 10^6$  s, and the total number of LSOs is 70 (here, this is the limiting parameter). Note that the total observing time corresponds to only  $\sim 7\%$  of the total *JWST*

observing time. The most likely probability of zero discoveries is 0.004 (ranges 0.003–0.006).

Table 6 gives the data triplets defined in Equation (18), which we use to estimate  $\eta_{\text{true}} = 0.30 \pm 0.12$ , where the quoted error is the square root of the MVB.

*A last note on this ministudy.* We have not treated the recovery times of  $c_{i,j}$  discussed in Section 2.3 because these example DRMs are “dilute,” meaning the LSOs account for only a small fraction of all the observations of *JWST*. In a “dense” DRM, with every LSO competing to be the next observation of the telescope, recovery time is important. Here, however, we can assume that a buffer of at least  $\sim 10^7$  s can conveniently pad the time between any two visits of the same star, which ensures adequate  $c_{i,j}$  recovery.

## 6. SUMMARY

In this paper, we have extended completeness-based metrics and algorithms for direct exoplanet searches to include multiple visits, estimating the probability distribution of search results and the occurrence rate of extrasolar planets. These extensions open the way for improved scheduling decisions, more realistic expectations for candidate instruments and missions, and enhanced science returns.

Preliminary DRMs for a starshade mission to enable Earth-like planet searches with *JWST* suggest a viable program for  $< 10^7$  s of observing time. About five discoveries and spectral characterizations are expected if  $\eta = 0.3$ . Soon, we hope, the *Kepler* mission will provide a first estimate of the true value of  $\eta$ .

The authors thank Jay Anderson, Massimo Robberto, Elizabeth Barker, Marsha Rieke, and Doug Kelly for help with the NIRCcam parameters; Jeff Valenti, Jason Tumlinson, Peter

**Table 6**  
Data Triplets for Estimation of  $\eta$  from Results of the Typical DRM

$k$	$c_k$	$C_k$	$u_k$	$k$	$c_k$	$C_k$	$u_k$	$k$	$c_k$	$C_k$	$u_k$
1	0.8038	0	0	25	0.1759	0.6484	0	48	0.1203	0.6833	0
2	0.7574	0	1	26	0.1681	0.2672	0	49	0.04000	0.9398	0
3	0.5537	0	0	27	0.2123	0.5008	0	50	0.04858	0.7929	0
4	0.5950	0	0	28	0.2043	0.3599	0	51	0.1264	0.1697	0
5	0.5560	0	1	29	0.1932	0.6266	0	52	0.08750	0.8198	0
6	0.6484	0	0	30	0.09530	0.8049	1	53	0.08603	0.6926	0
7	0.3886	0	0	31	0.1171	0.5334	0	54	0.05295	0.7275	1
8	0.6446	0	0	32	0.1827	0.6446	0	55	0.07265	0	0
9	0.6266	0	0	33	0.2121	0.4711	0	56	0.1190	0.6811	0
10	0.2199	0.5537	1	34	0.1264	0.6184	0	57	0.07973	0.8273	0
11	0.3442	0	0	35	0.07820	0.7147	0	58	0.05505	0.6267	0
12	0.4061	0	0	36	0.3361	0	0	59	0.09337	0.2960	0
13	0.5008	0	0	37	0.1120	0.4353	0	60	0.3181	0	0
14	0.2099	0.5950	0	38	0.2154	0.4657	0	61	0.1448	0.3361	0
15	0.2067	0.3886	0	39	0.1284	0.5642	0	62	0.03500	0.8414	0
16	0.1360	0.8038	0	40	0.07705	0.6505	0	63	0.1806	0.3261	0
17	0.4711	0	0	41	0.1697	0	0	64	0.2701	0	0
18	0.2672	0	0	42	0.3261	0	0	65	0.05180	0.8227	0
19	0.3599	0	0	43	0.07945	0.5473	0	66	0.06625	0.8257	0
20	0.5590	0	0	44	0.1126	0.7131	0	67	0.05485	0.07265	0
21	0.1893	0.3442	0	45	0.1813	0.5590	0	68	0.05650	0.7786	0
22	0.4657	0	0	46	0.07942	0.8242	0	69	0.04180	0.6818	0
23	0.2123	0.4061	0	47	0.07785	0.7449	0	70	0.06988	0.8035	0
24	0.1193	0.5953	0								

**Notes.**  $k$  here is  $k_{\text{DRM}}$  in Table 4.  $\mathcal{P}_k$  in Table 4 can be recovered from  $c_k$  and  $C_k$  here using Equations (12) and (18) and  $\eta = \eta_{\text{ops}} = 0.3$ .

Jakobsen, and Pierre Ferruit for help with the NIRSspec parameters; and Webster Cash, Amy Lo, and Tiffany Glassman for help with starshade parameters. They thank Sara Seager for discussions and providing the atmospheric models used to calculate the albedos of Earth. Thanks to Christopher Burrows and Dmitry Savransky for discussions. Appreciation to Marc Postman and Matt Mountain for encouraging this investigation. Gratitude to Christian Lallo for his help keeping R.A.B.'s computers running. Kudos to Sharon Toolan for preparing the manuscript with such great skill.

#### REFERENCES

- Agol, E. 2007, *MNRAS*, 374, 1271  
Brown, R. A. 2004, *ApJ*, 607, 1003  
Brown, R. A. 2005, *ApJ*, 624, 1010  
Brown, R. A. 2009a, *ApJ*, 699, 711  
Brown, R. A. 2009b, *ApJ*, 702, 1237  
Brown, R. A., Shaklan, S. B., & Hunyadi, S. L. 2007, JPL 07-02 7/07, in Proc. TPF-C Workshop (9/28–29/2007), ed. W. A. Traub (Pasadena, CA: JPL), 53  
Cash, W., Oakley, P., Turnbull, M., Glassman, T., Lo, A., Polidan, R., Kilston, S., & Noecker, C. 2008, *Proc. SPIE*, 7010, 70101Q  
Cash, W., Spergel, D. N., Soummer, R., Mountain, M., Hartman, K., Polidan, R., Glassman, T., & Lo, A. 2009, The New Worlds Probe: A Starshade with *JWST* (Request for Information to Astro2010: The Astronomy and Astrophysics Decadal Survey, June 2009; Boulder, CO: Univ. Colorado Boulder), [http://newworlds.colorado.edu/documents/Decadal/cash\\_jwststarshades\\_ppp.pdf](http://newworlds.colorado.edu/documents/Decadal/cash_jwststarshades_ppp.pdf)  
Guyon, O., et al. 2008, *Proc. SPIE*, 7010, 70101Y  
Leinert, Ch., et al. 1998, *A&AS*, 127, 1  
Savransky, D., Kasdin, N. J., & Cady, E. 2010, *PASP*, 122, 401  
Soummer, R., Levine, M., & Exoplanet Forum Direct Optical Imaging Group 2009a, in Exoplanet Community Report, ed. P. R. Lawson, W. A. Traub, & S. C. Unwin (Pasadena, CA: JDL), 47 (<http://exep.jpl.nasa.gov/documents/ExoplanetCommunityReport.pdf>)  
Soummer, R., et al. 2009b, *Proc. SPIE*, 7440, 74400A  
Spergel, D. N., et al. 2009, *BAAS*, 41, 362