

# A Step Toward Eta-sub-Earth

Wesley A. Traub

Jet Propulsion Laboratory, California Institute of Technology,  
M/S 321-100, 4800 Oak Grove Dr, Pasadena, CA 91109, USA  
email: [wtraub@jpl.nasa.gov](mailto:wtraub@jpl.nasa.gov)

**Abstract.** The Kepler mission observed exoplanet transits for 4 full years (greater than its expected lifetime of 3.5 years) until it became inoperable for its original purpose, as a result of a reaction wheel failure. Kepler was spectacularly successful in its goal of observing exoplanet transits of host star disks for the purpose of measuring the statistics of such transits in its target star sample. The Kepler data, when fully analyzed, will determine the statistics of planets in the underlying population, and in particular the expected number of terrestrial planets in habitable zone orbits per solar-type star, the quantity known as eta-sub-Earth. This report is an initial examination of Kepler's third catalog (Feb. 2012) of planets and candidate planets. I find that the apparent projected value of eta-sub-Earth is several times smaller than I had found from the second catalog, but that the data are now approaching the point where intrinsic biases can be uncovered. When all bias factors are eventually found, it is likely that the true value of eta-sub-Earth will be substantially greater than its current apparent value.

**Keywords.** stars: planetary systems, methods: statistical

---

## 1. Introduction

In this paper I will use the symbol  $\eta_{\oplus}$ , pronounced eta-sub-Earth, for the ratio of the number of terrestrial planets in habitable-zone orbits to the number of solar-like stars. Here, terrestrial planets are those with radii in the range 0.5 to 2.0 Earth radii (corresponding roughly to masses in the range 0.1 to 10 Earth masses), the habitable zone is the planet-star distance range within which liquid water can exist on a planet's surface (nominally, 0.80-1.80 AU for the Sun), and solar-like stars include main-sequence dwarf F, G, K spectral classes. The value of  $\eta_{\oplus}$  is relevant for planning future direct-imaging missions that could detect and characterize nearby Earth-like planets, and search for signs of life on those planets.

The first release of Kepler planets and candidates (hereafter simply planets) was in June 2010 (Borucki *et al.*, 2011a), containing 312 planets and covering 43 days (labeled "quarter" 1) of Kepler data.

The second release was in February 2011 (Borucki *et al.* 2011b), containing 1235 planets and covering 13 months (quarters 1-6) of data. This data was analyzed by Catanzarite & Shao (2011), who found a projected value of  $\eta_{\oplus} \simeq (1 - 3)\%$ . The data was also analyzed by Traub (2012) who found  $\eta_{\oplus} \simeq (34 \pm 14)\%$ . The difference between these estimates, both derived from the same data base, is mainly owed to the different assumptions of the respective authors concerning which of the sample planets should be considered, and in their assumptions about how to project the existing short-period data to longer (habitable-zone) periods. At present, it is not clear which of these values is closest to the truth. The present paper is a step in the direction of uncovering that truth.

The third release was in February 2012 (Batalha, 2012), containing 2321 planets and covering 16 months (quarters 1-7) of observations; for this release the algorithms had

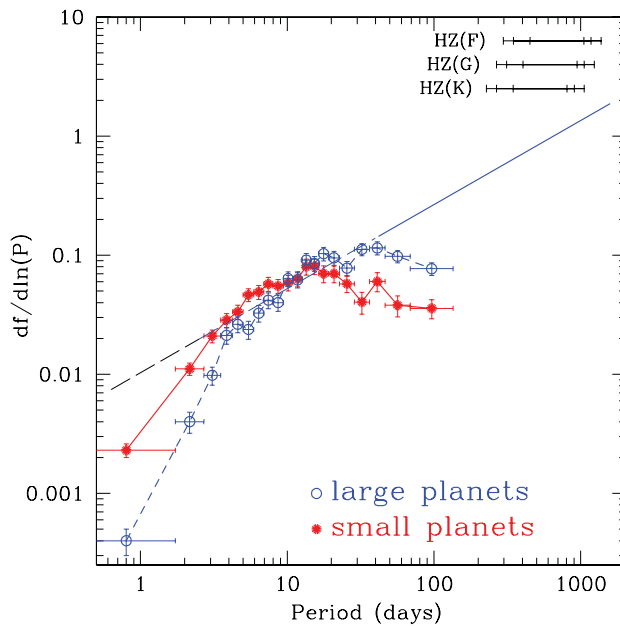
advanced to the point where the individual quarters of data could be stitched together, improving the detectability of longer-period planets.

## 2. First Stage of Analysis

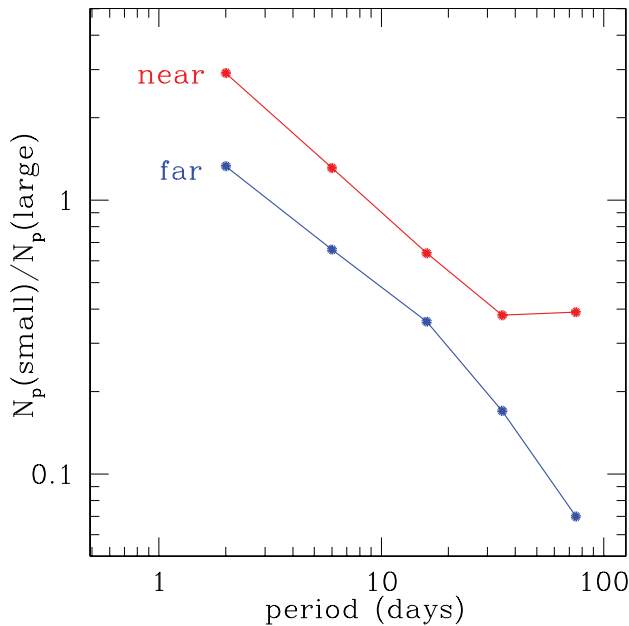
For this paper, I first made several cuts to the data, as follows: targets need to have been observed for at least 5 of the available 7 quarters; host star effective temperature less than 7000 K; host star  $\log(g)$  greater than 4.00; host star radius in the range 0.45 to 1.95; planet period in the range 0.37 to 137 days; planet radius in the range 0.34 to 19 Earth radii; and signal to noise in the range 4 to 5900. These cuts reduced the number of planets from 2321 to 1926.

I binned the resulting sample into 10 bins of host star magnitude, and 10 bins of host star effective temperature, adjusting the binning intervals so that approximately equal numbers of planets fell into each magnitude or temperature bin. I did the same for all target stars, not just the planet-hosting ones. In both cases the distributions show a striking correlation of parameters, such that cool stars tend to be faint, and warm stars tend to be bright. This is not an astrophysical surprise, but it is a warning that we should be very careful about drawing conclusions from the data, since there is an absence of cool, bright stars, and also warm, faint ones. This holds for the planet hosts as well as all target stars.

For each planet in the sample, I estimated the number of actual planets in the underlying population by dividing by the geometric probability of detection (star radius divided by orbital radius). I binned the resulting number of planets in the population into 20 bins as a function of  $\ln(\text{period})$ , choosing bin sizes such that there were approximately equal numbers of sample planets per bin, making each bin comparable to the others in



**Figure 1.** Populations of small and large planets are shown as a function of period. Note how small planets dominate at short periods, but are conspicuously absent at long periods. We find evidence that this effect is, at least in part, a result of a bias against detecting small planets around faint stars, when the correlation between brightness and spectral type is removed.



**Figure 2.** The ratio of small to large planets is shown separately for targets that are effectively “near” and “far”, plotted in 5 bins in period, chosen to give good statistics in each bin. If there were no bias in the Kepler data against small planets around faint stars, the “near” and “far” curves would lie on top of each other. But they do not, and in fact the “far” group is about a factor of 2 lower, indicating that small planets are being missed in that group. This analysis assumes that the “near” and “far” planetary systems are identical except for their distance from the Sun.

terms of signal to noise ratio. I divided the number of planets in the population, in each  $\ln(\text{period})$  bin, by the total number of target stars, giving a fraction  $f$  of planets per star in each bin. I separated the ensemble into two groups of approximately equal numbers of sampled planets, based on planet radius. I designated the resulting values as  $df/d\ln(P)$ , the ratio of number of planets per star in each logarithmic bin in period.

These values are plotted in Fig. 1, with open circles for large planets, and filled circles for small planets. It is immediately obvious that small planets dominate by roughly 2-10 sigmas at short periods, less than 10 days. There are about equal numbers of planets per star in the range 10-20 days. And at long periods, 30 days and larger, small planets are notably absent, by roughly 4 sigma.

If we simply extrapolate the small-planet population to longer, habitable-zone periods, we will get a value of  $\eta_{\oplus}$  that is perhaps 30 or more times smaller than the value we would get from the straight-line extrapolation in Fig. 1, which line is copied from the corresponding figure in Traub (2012). This leads us to ask, is it valid to simply extrapolate the data at hand, or are there any potential biases in the data that could mislead us? This is the question that I address in the next section.

### 3. Second Stage of Analysis

In this second stage of analysis I show that the existing data appear to show that small planets are not being detected as efficiently as large planets. This is precisely what we might naively expect, given that small planets have a weaker transit signal than large ones.

For this stage of analysis I assumed that, for any given spectral type, the probability that a star would host a given size or period of planet is the same for nearby as for distant stars. I estimated the distance to each target star by using the JHK magnitudes, effective temperature, and radius. Reddening was not included. Gratifyingly, in a plot of distance vs. Kepler magnitude, for each stellar temperature, the stars fall off in magnitude just as expected from an inverse square law. To avoid cool and warm stragglers, I kept only those stars in the range 5000 - 6000 K, and split these into nearly equal numbers of “near” and “far” stars, keeping nearly the same balance of warm and cool stars in each group. The resulting “near” and “far” groups are therefore roughly homogeneous in numbers of stars of each spectral type, removing any bias from that parameter. The main distinguishing parameter of the two groups is simply distance for each spectral type.

To search for a bias against small planets, I form the ratio of numbers of small planets to large planets in the population in each “near” and “far” group, as a function of period. These data are shown in Fig. 2, where I have chosen the 5 periods to give roughly comparable statistics in each bin. It is dramatically clear from this plot that the “far” group systematically has fewer small planets compared to large ones than the “near” group, by a factor of about 2. The effect is even more dramatic for the longest-period group, where small planets are about a factor of 6 depleted in the “far” group compared to the “near” one.

#### 4. Conclusions

It is clear from this preliminary analysis that the Kepler survey is not detecting small planets as efficiently as large ones. It remains to model this effect in a way that can be used to correct the apparent detections. It is fortunate that the Kepler spacecraft survived long enough to give us four full years of observing time, in addition to its initial 3.5-year mission which was completed on 14 November 2012. The additional data will provide more transits of a given period and planet size, which will in turn make it easier to use the data itself to calibrate the inevitable signal-to-noise bias terms.

I thank the Kepler team for providing such abundant and precise data. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

#### References

- Batalha, N. M., Rowe, J. F., Bryson, S. T., *et al.* 2012, *ApJS*, 204, article id. 24
- Borucki, W. J., Koch, D. G., Basri, G., *et al.* 2011a, *ApJ*, 728, 117
- Borucki, W. J., Koch, D. G., Basri, G., *et al.* 2011b, *ApJ*, 736, 19
- Catanzarite, J. & Shao, M. 2012, *ApJ*, 738, 151
- Traub, W. A. 2012 *ApJ*, 745, 20