

# ExoPAG SAG13: Exoplanet Occurrence Rates and Distributions

Ruslan Belikov  
NASA Ames Research Center  
on behalf of SAG13



# SAG13 members

Belikov, Ruslan (Chair, [rulsan.belikov@nasa.gov](mailto:rulsan.belikov@nasa.gov))

Stark, Christopher (Co-chair)

Batalha, Natalie (Steering Committee)

Burke, Chris (Steering Committee)

Angerhausen, Daniel

Apai, Daniel

Bendek, Eduardo

Bennett, David

Blackwood, Gary

Boss, Alan

Brown, Robert

Bryden, Geoff

Cahoy, Kerri

Catanzarite, Joe

Ciardi, David

Clanton, Christian

Cowan, Nick

Danchi, William

Domagal-Goldman, Shawn

Dressing, Courtney

Farr, Will

Foreman-Mackey, Daniel

Fressin, Francois

Gaudi, Scott

Ge, Jian

Gould, Andy

Hogg, David W

Howard, Andrew

Kane, Stephen

Kasting, Jim

Kopparapu, Ravi

Macintosh, Bruce

Mandell, Avi

Mendez, Abel

Meyer, Michael

Morgan, Rhonda

Mulders, Gijs

Nielsen, Eric

Petigura, Erik

Ragozzine, Darin

Roberge, Aki

Rogers, Leslie

Savransky, Dmitry

Serabyn, Gene

Shabram, Megan

Shao, Mike

Solmaz, Arif

Sparks, William

Stahl, Philip

Stapelfeldt, Karl

Stark, Christopher

Still, Martin

Suzuki, Daisuke

Swain, Mark

Traub, Wes

Turnbull, Margaret

Unwin, Stephen

Vanderbei, Bob

Walkowicz, Luzianne

Weiss, Lauren M.

Wolfgang, Angie

Youdin, Andrew



# Charter

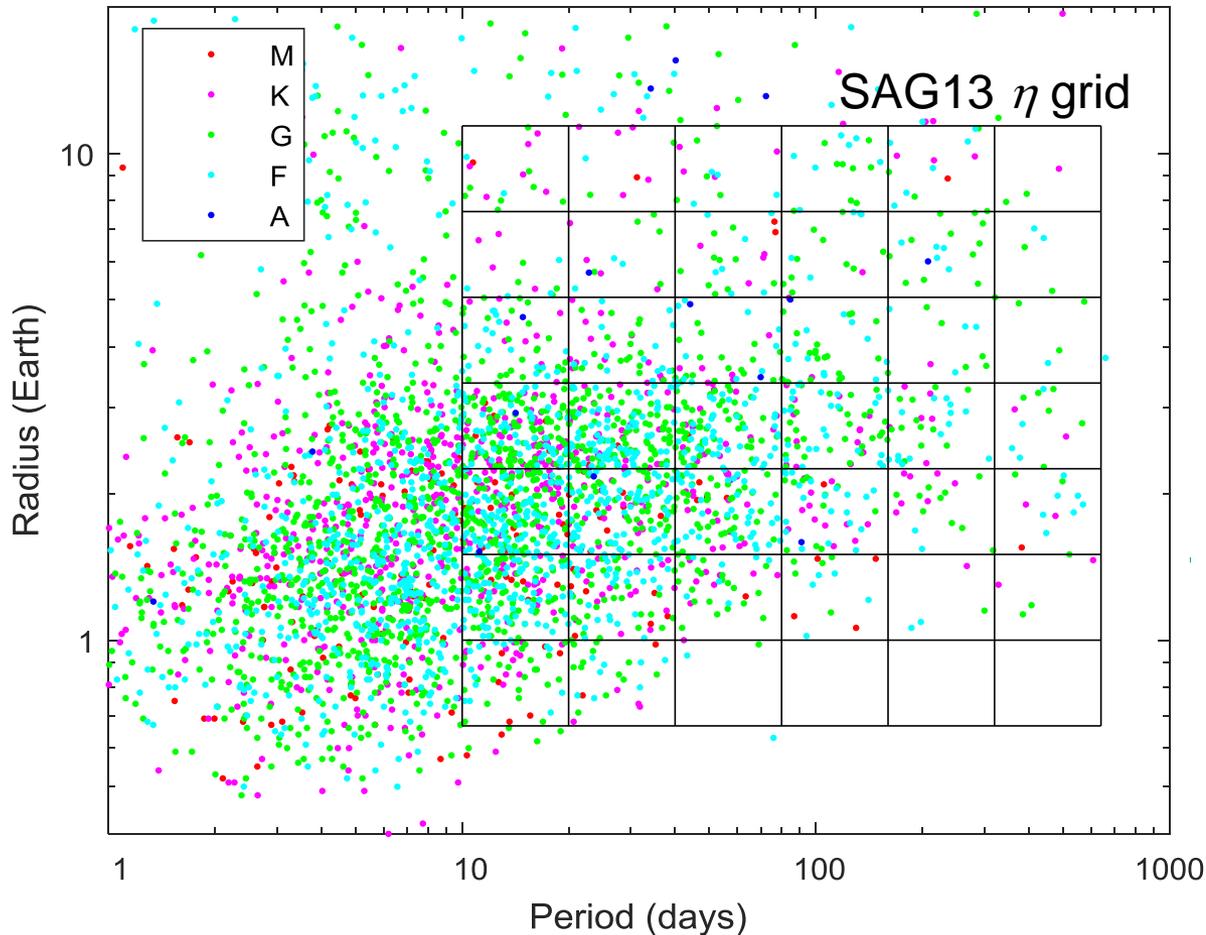
Over 5000 exoplanets and exoplanet candidates have been discovered to date. Many studies have been published and are on-going to determine exoplanet occurrence rates and distributions, particularly for potentially habitable worlds. These studies employ different statistical and debiasing methods, different definitions of terms such as  $\eta_{\text{Earth}}$  and habitable zone, different degrees of extrapolation, and present distributions in different units from each other. The primary goal of this SAG is to evaluate what we currently know about planet occurrence rates, and especially  $\eta_{\text{Earth}}$ , by consolidating, comparing, and reconciling discrepancies between different studies. A secondary goal is to establish a standard set of occurrence rates accepted by as much of our community as possible to be used for mission yield estimates for missions to be considered by the decadal survey.

## Key objectives and questions:

- Completed → 1. Propose standard nominal conventions, definitions, and units for occurrence rates/distributions to facilitate comparisons between different studies.
- Current activity → 2. Do occurrence estimates from different teams/methods agree with each other to within statistical uncertainty? If not, why?
- Current activity → 3. For occurrence rates where extrapolation is still necessary, what values should the community adopt as standard conventions for mission yield estimates?



# Standardized eta grid



11 community sourced occurrence tables

Batalha, Natalie (2)

Belikov, Rus

Burke, Chris

Catanzarite, Joe

Dressing, Courtney\*

Farr, Will

Foreman-Mackey, Daniel\*

Kopparapu, Ravi

Mulders, Gijs

Petigura, Erik\*

Traub, Wes\*\*

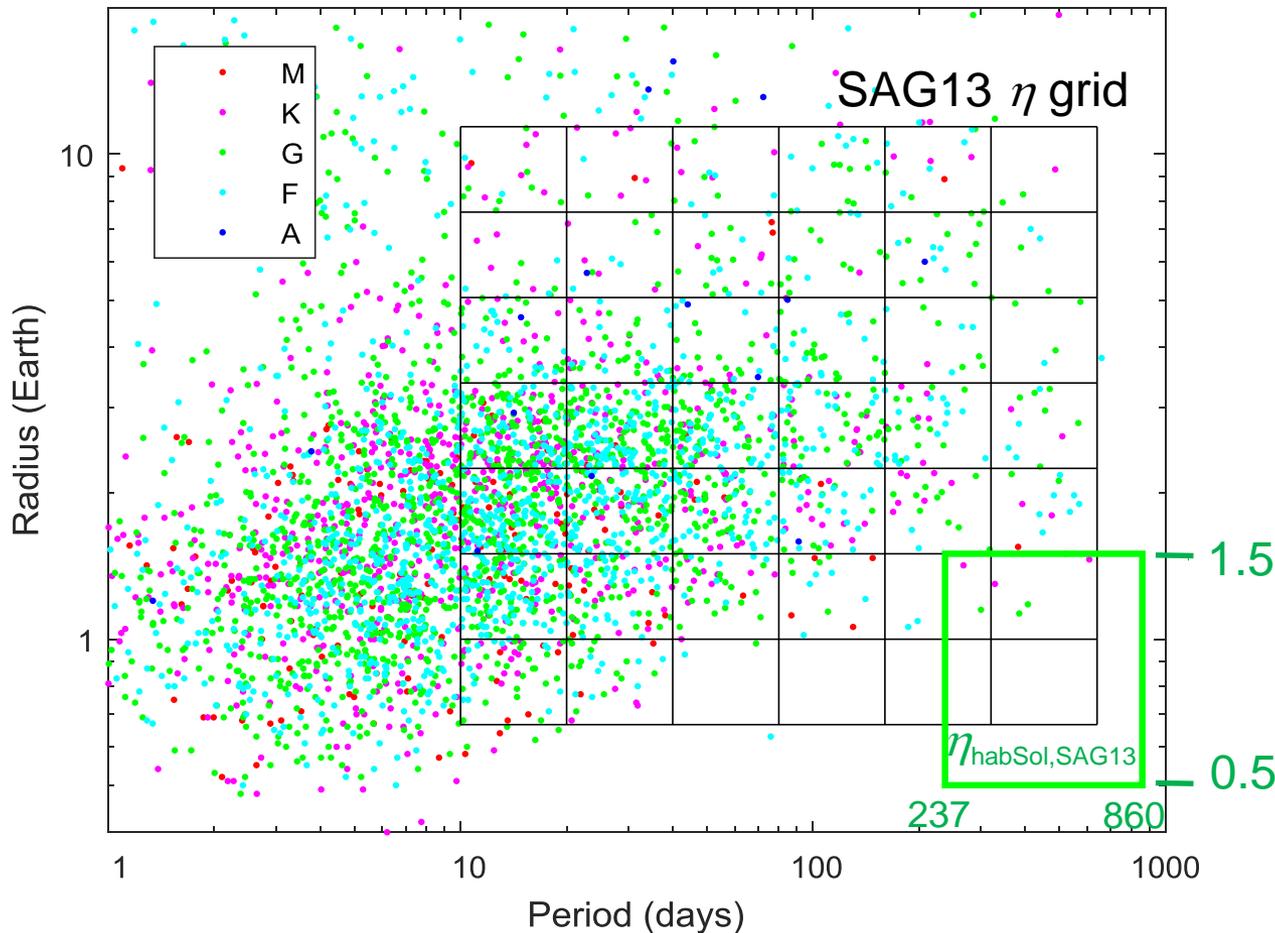
\*dataset was based on prior publications and re-integrated across SAG13 bins by Burke  
\*\* expected soon

All datasets and documents can be found on SAG13 repository:  
<https://drive.google.com/drive/folders/0B520NCfkP4aOQUJYdmUzQTJkdkE>

- Dots are an example catalog: Kepler candidates from Q1-Q17, dr24
  - SAG participants are free to choose any catalog and method



# Standardized eta grid



11 community sourced occurrence tables

Batalha, Natalie (2)

Belikov, Rus

Burke, Chris

Catanzarite, Joe

Dressing, Courtney\*

Farr, Will

Foreman-Mackey, Daniel\*

Kopparapu, Ravi

Mulders, Gijs

Petigura, Erik\*

Traub, Wes\*\*

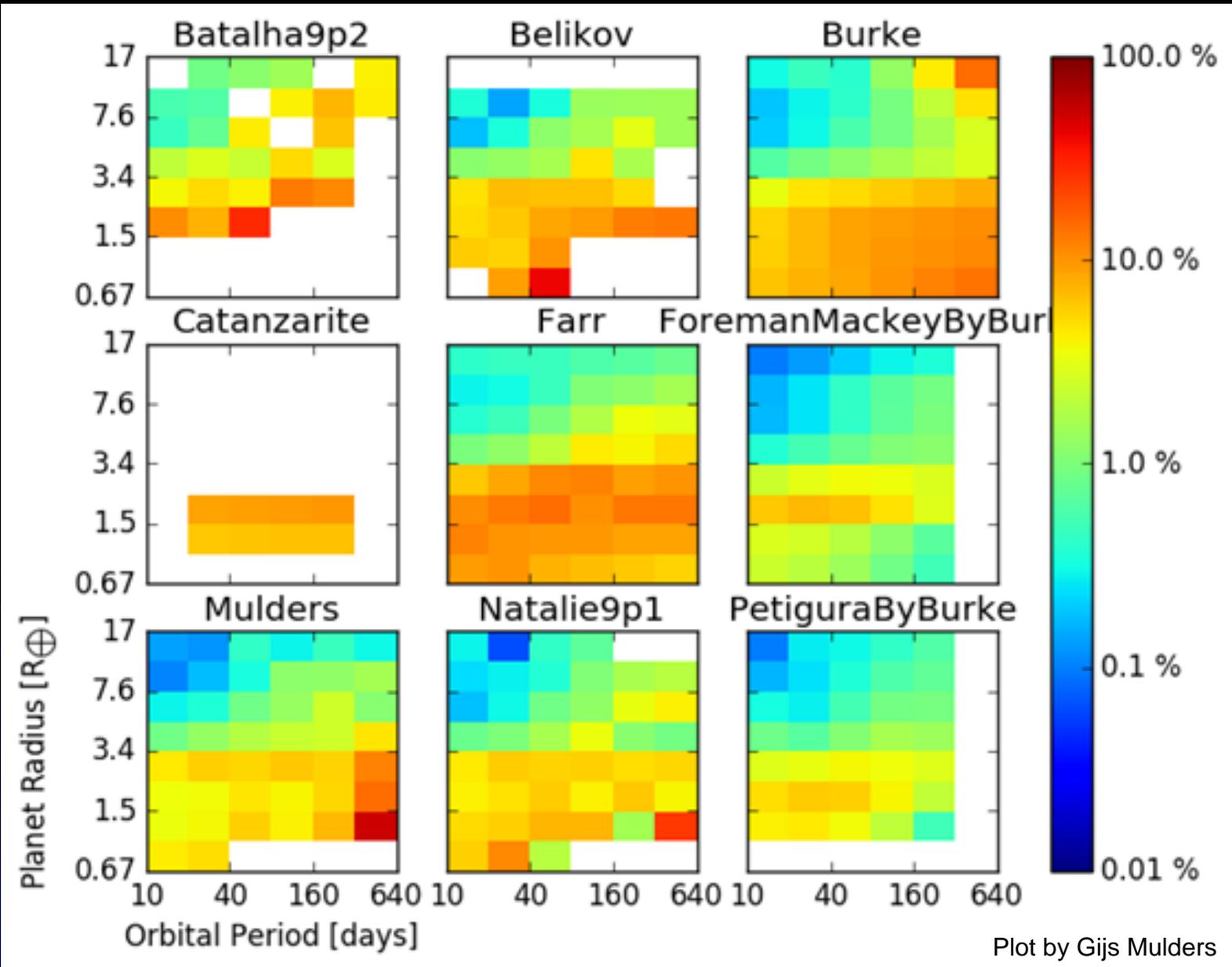
\*dataset was based on prior publications and re-integrated across SAG13 bins by Burke  
\*\* expected soon

All datasets and documents can be found on SAG13 repository:  
<https://drive.google.com/drive/folders/0B520NCfkP4aOQUJYdmUzQTJkdkE>

## $\eta_{habSol,SAG13}$

- $R = [0.5 - 1.5]$ ,  $P = [237 - 860]$  (Kopparapu extended HZ for Sol twin)
- This is not exactly  $\eta_{Earth}$ , just a tentative rough representation of a potentially habitable region

 Example: submitted occurrence rates for G-dwarfs





# How do we combine different submissions into one occurrence table?

Full accounting:  
Only “independent” submissions  
are averaged

Accounting for “dependency”  
between submissions

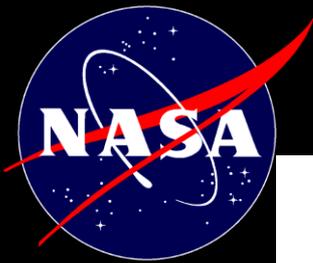
No accounting:  
Simply average all  
submissions



- Best for producing an actual scientific measurement
- Measuring “dependency” is not trivial (and may be impossible in principle)
- Consensus on method can be challenging
- Psychological biases are challenging to identify and control

- Will not generate a scientific measurement, but possibly best for predictions?
- Simple method
- Easier consensus: all submissions are automatically fairly represented
- Crowdsourcing / Prediction market philosophy: psychological biases are in theory averaged out

The question of which method is “correct” is possibly philosophical  
Will probably do both, explicitly describe the process, and leave interpretation to the reader  
Feedback on our strategy is welcome and encouraged

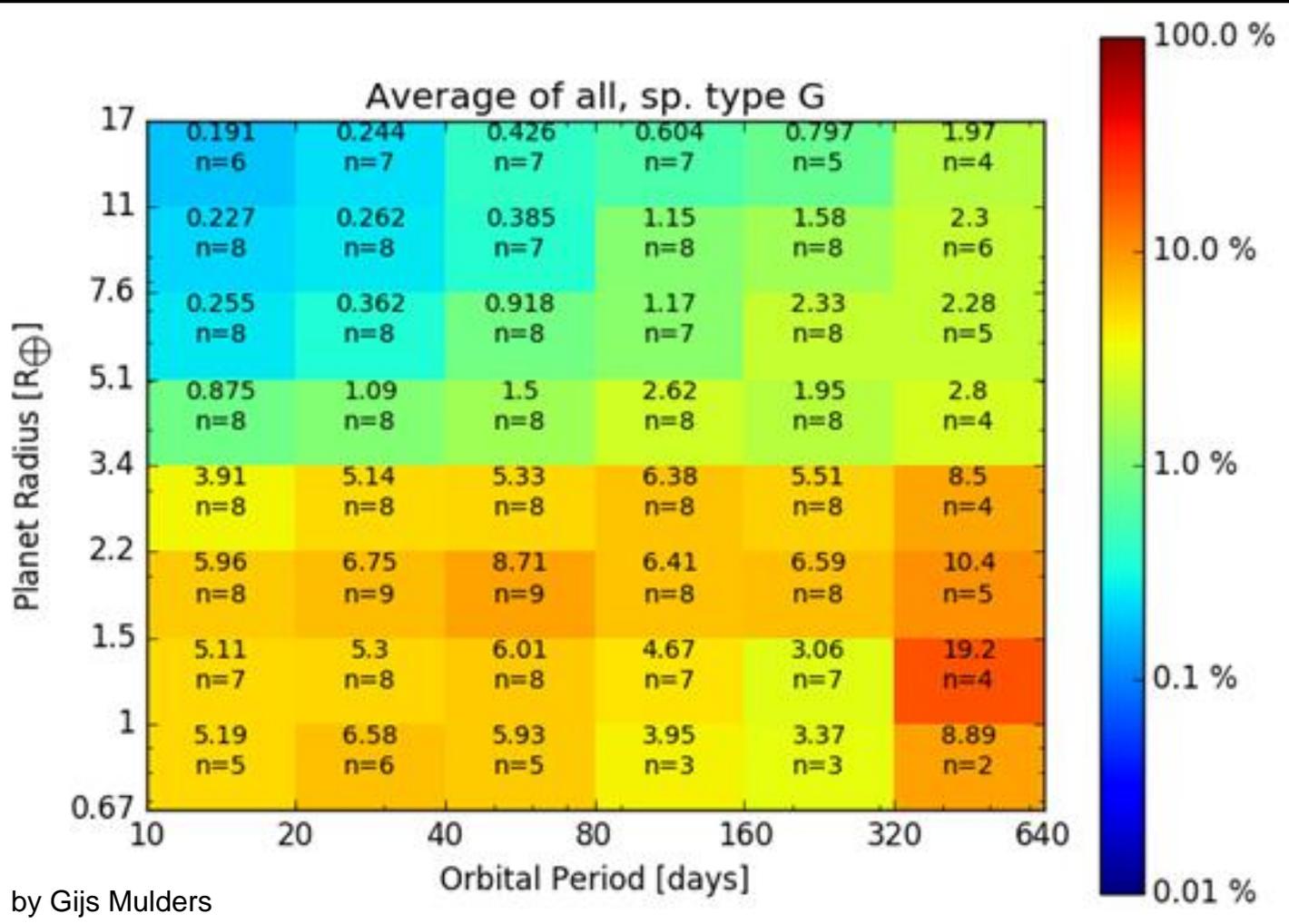


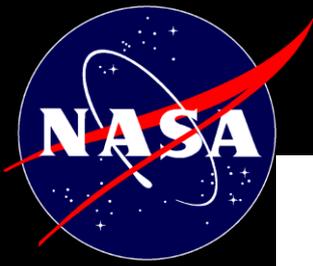
# Closer look at G-dwarf average

## legend

% occurrence

# of submissions



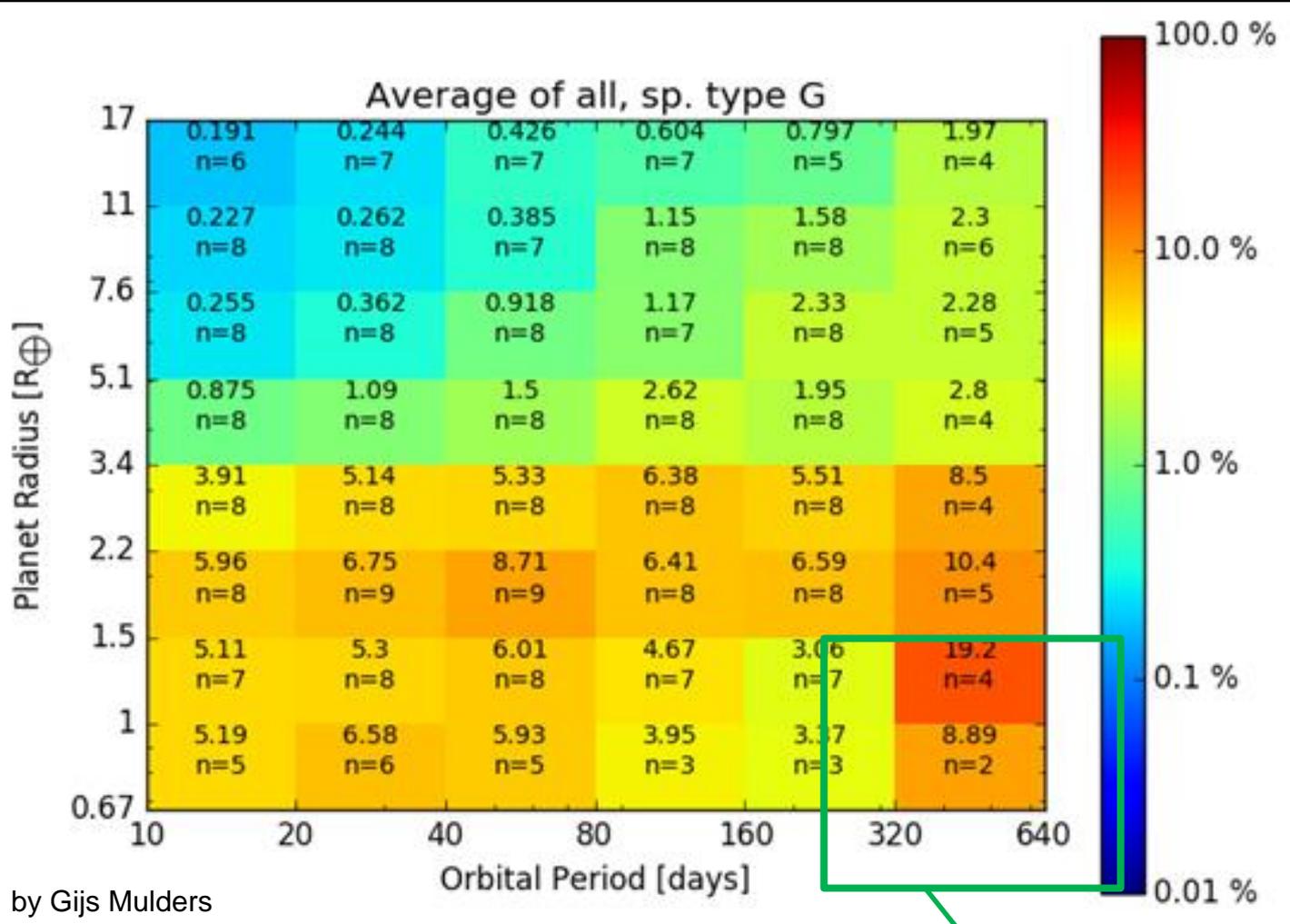


# Closer look at G-dwarf average

## legend

% occurrence

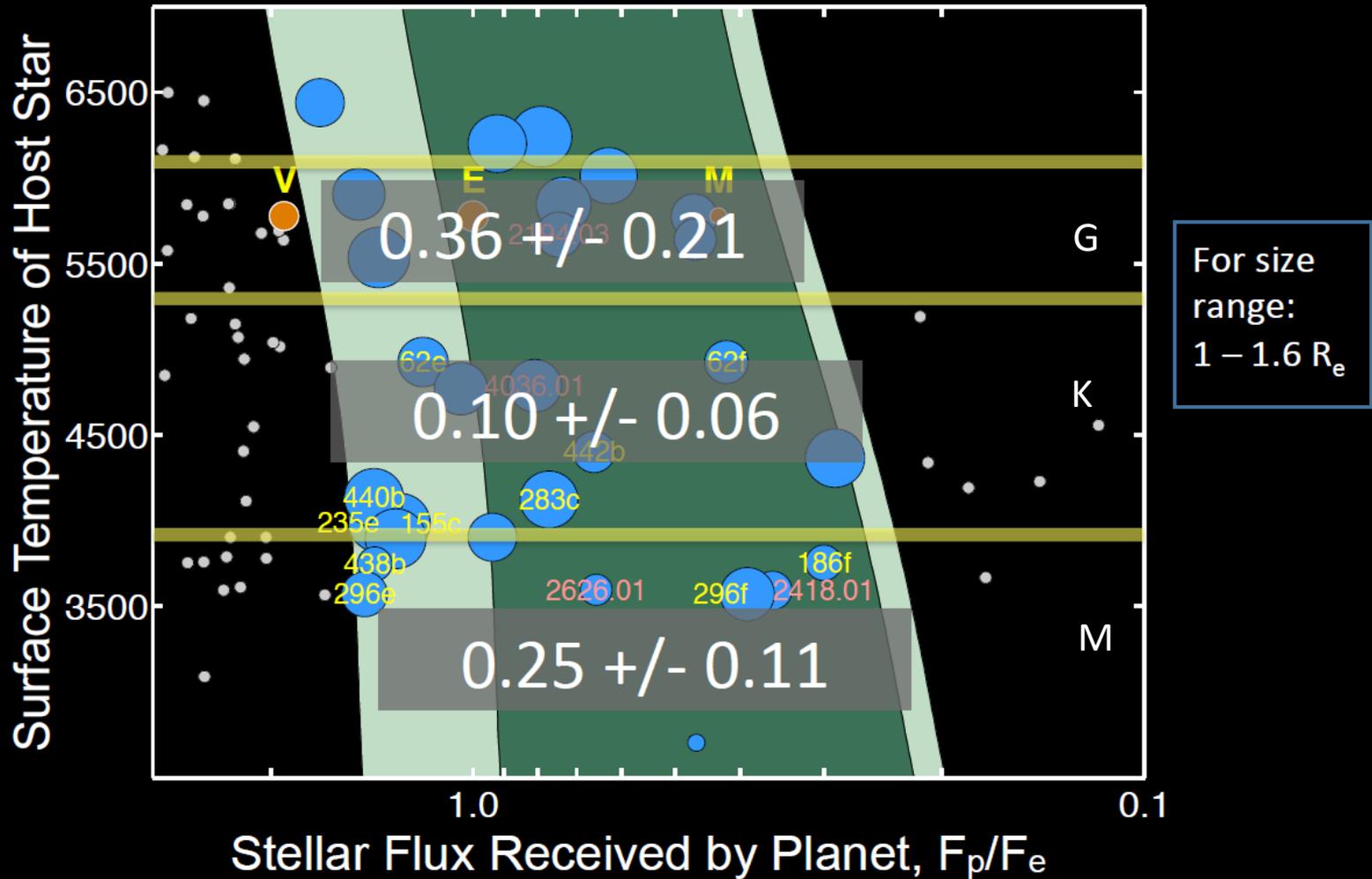
# of submissions



Average occurrence per SAG13 bin in green box: ~10%  
 Area of green box: 5x SAG13 bins  
 Uniform extrapolation implies green box occurrence of **50%**  
 Note: this number is \*not\* an official SAG13 value for  $\eta_{\text{habSol,SAG13}}$ , it is just an example of a simple extrapolation.

# Small ( $< 2 R_e$ ) Planets in the HZ: 4 yr

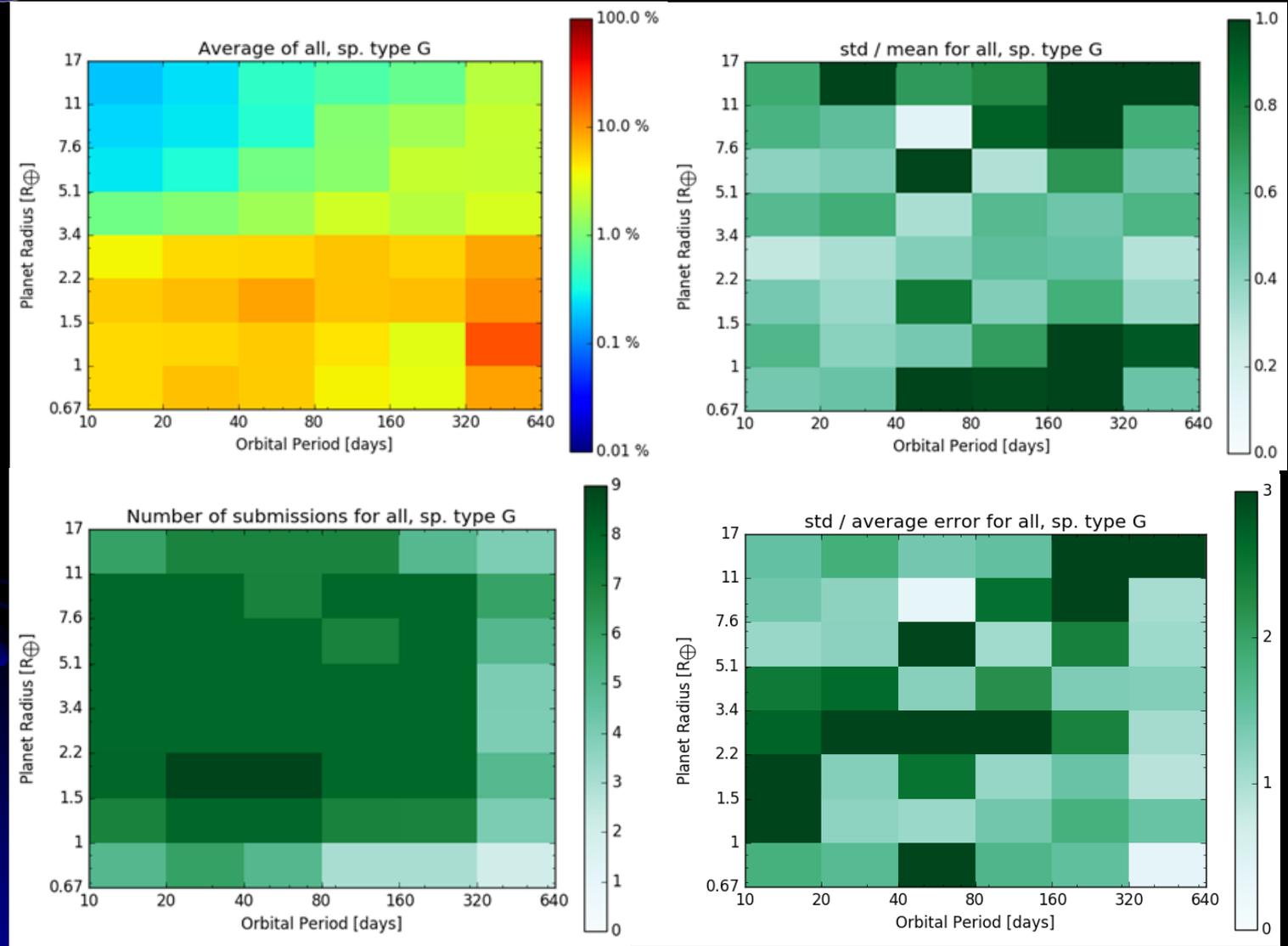
Empirical HZ
  Narrow HZ
  In HZ, symbol scaled to size of Earth



Note: for planet size range of  $0.5 - 1.6 R_e$ , expected # of planets may be a factor of  $\sim 2-3$  higher (based on extrapolation)



# Analysis of variations in submissions (for G-dwarfs)



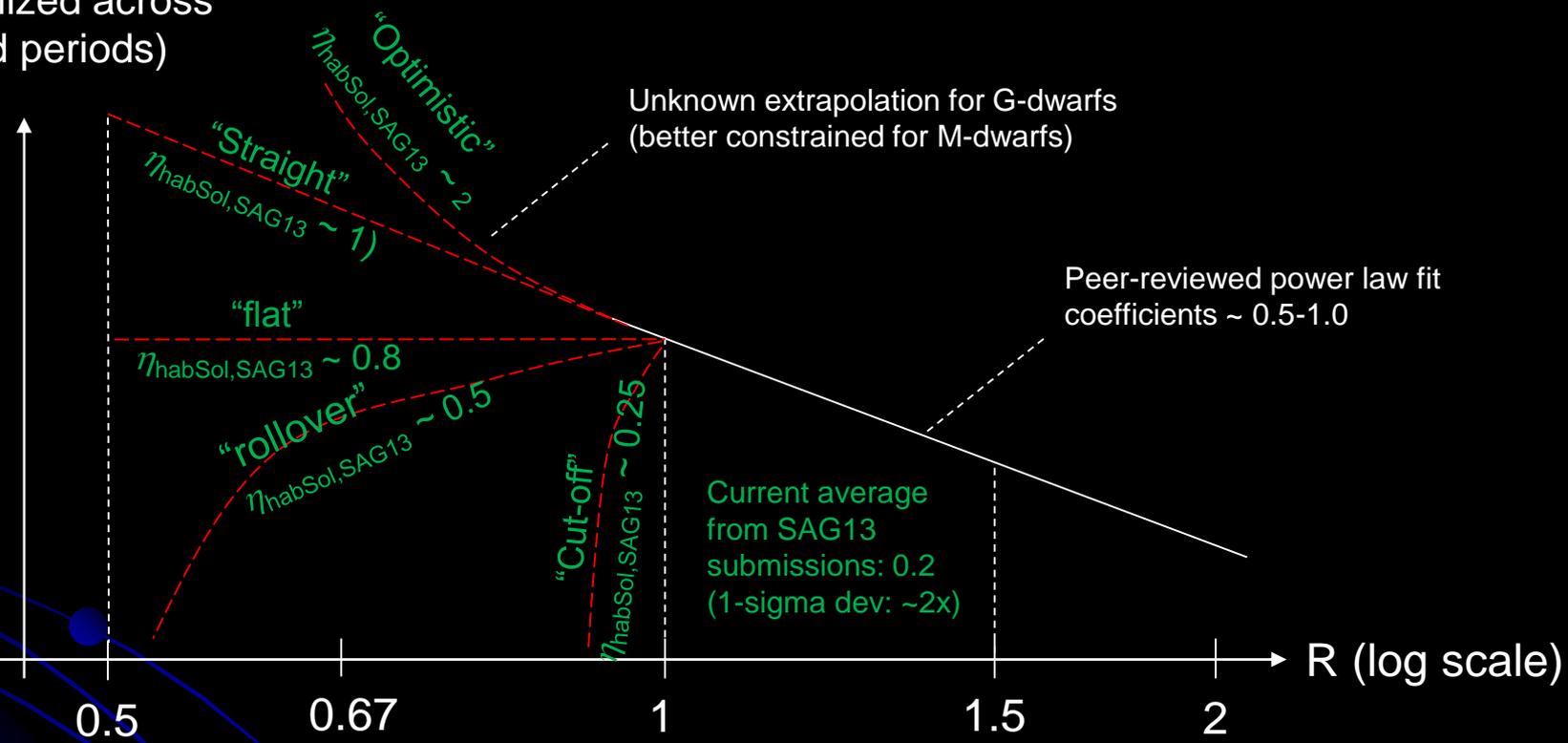
Crowdsourced standard deviation normalized to submitted occurrence uncertainty

by Gijs Mulders



# Importance of 0.5-1.0 Earth size bin

$dN / d\ln(R)$   
(marginalized across  
237-860d periods)



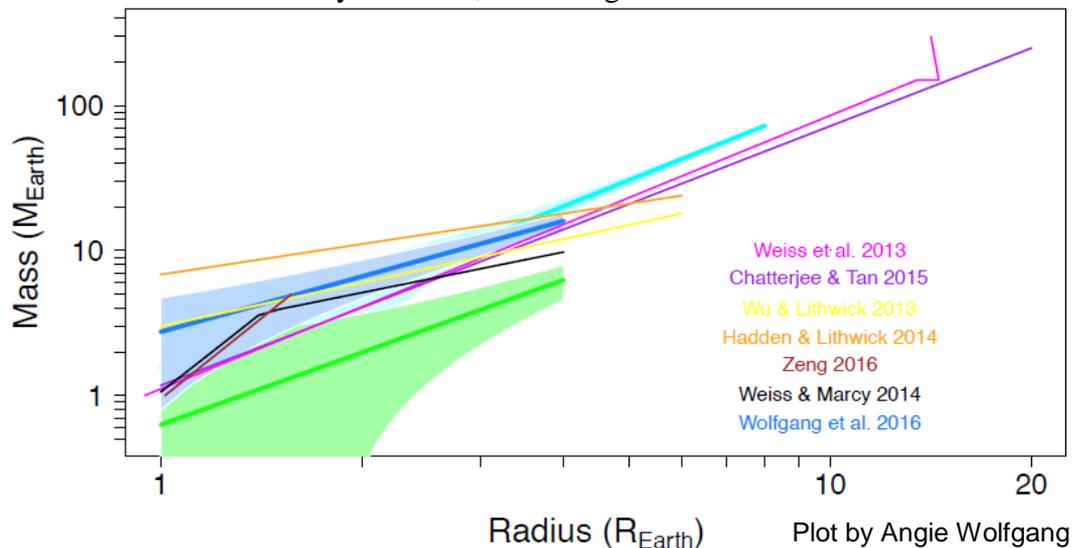
- Any estimate of  $\eta_{Earth}$  should always very clearly specify:
  - Whether 0.5-1.0 bin is included or not
  - What extrapolation assumption was made
- Many discrepancies in  $\eta_{Earth}$  estimates can be traced to inclusion or exclusion of 0.5-1.0 bin
- Mission study teams may want to consider the possibility of a large number of potentially habitable planets in the 0.5-1.0 bin



# New focus group: Mass-Radius relationship led by Angie Wolfgang and Lauren Weiss

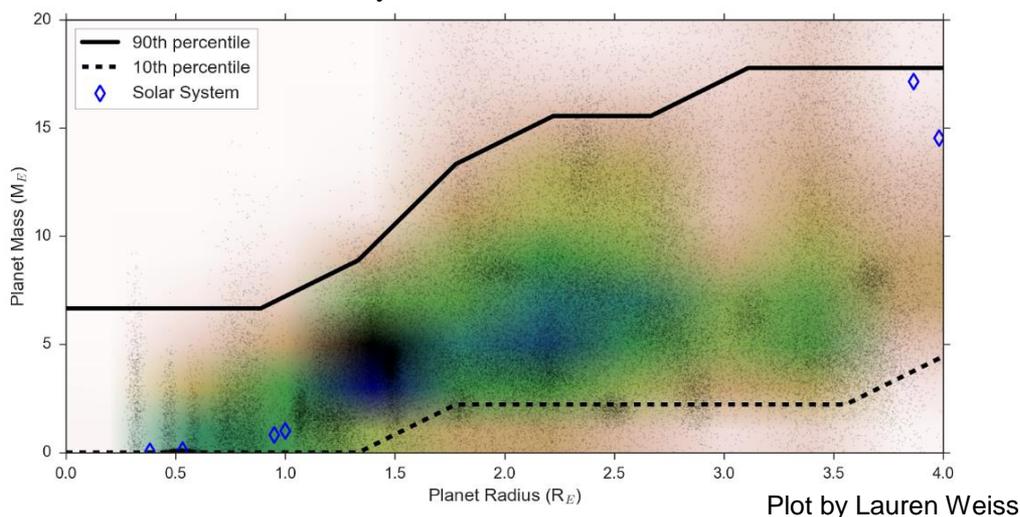
Previous M-R relations in the literature:

wide variety of radius, mass ranges and datasets used

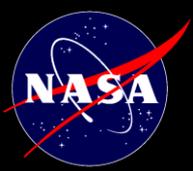


- Purpose: enable SAG13 occurrence rate submissions based on RV planets
- M-R relationship is fundamentally not a 1-1 map (e.g.  $M = f(R)$ ), but a correlation (e.g. density function  $C(M,R)$ )
- M-R focus group deliverables
  - an estimate of this correlation based on open community input
  - analysis of uncertainties and dependency on period and other parameters

Preliminary estimate of M-R correlation

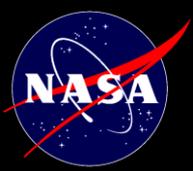


- Notes about plots / methods
  - TTV data is included
  - Black dots: MC posterior simulation accounting for uncertainties on currently known M-R planets
  - Color map: estimate of the 2D correlation density function (using Gaussian kernel density estimator)

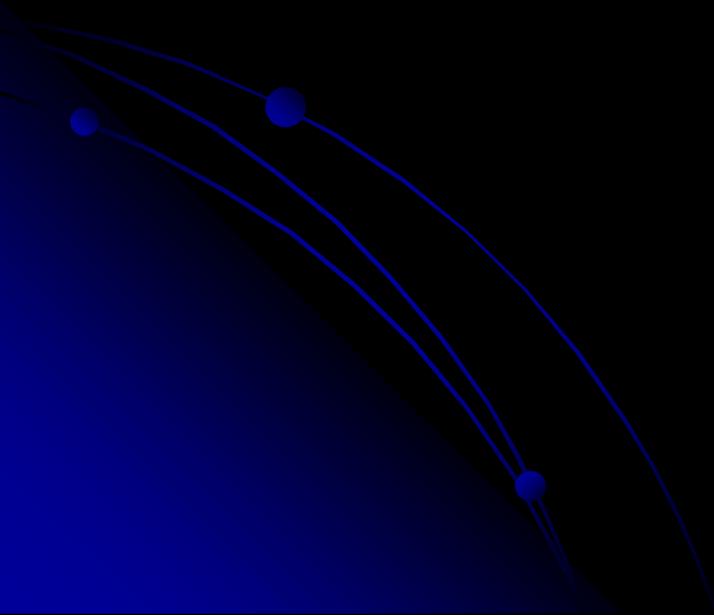


# Conclusions

- Completed products
  - Proposed conventions for binning planet periods, sizes, and star temperatures
  - Living repository of occurrence rate datasets submitted by scientists
  - Code to visualize them and compute statistics (first version done, evolving)
  - All can be found on <https://drive.google.com/drive/folders/0B520NCfkP4aOQUJYdmUzQTJkdkE>
- Expected products (by start of 2017)
  - More thorough analysis of submissions
  - Final 3D matrix of SAG13 bin values representing a combination of all submissions
    - With uncertainties and analysis of deviations
  - Recommendation for a standard assumption(s) of parametrized multi-variate distribution(s) for missions
    - With a discussion of uncertainties and method
    - Estimates of occurrence rates relevant to habitable planets based on this distribution
- Latest estimates of occurrences of potentially habitable planets seem to be converging (at least to a factor of  $\sim 2-3$ ), and explanations for discrepancies are starting to clarify



# Backup slides



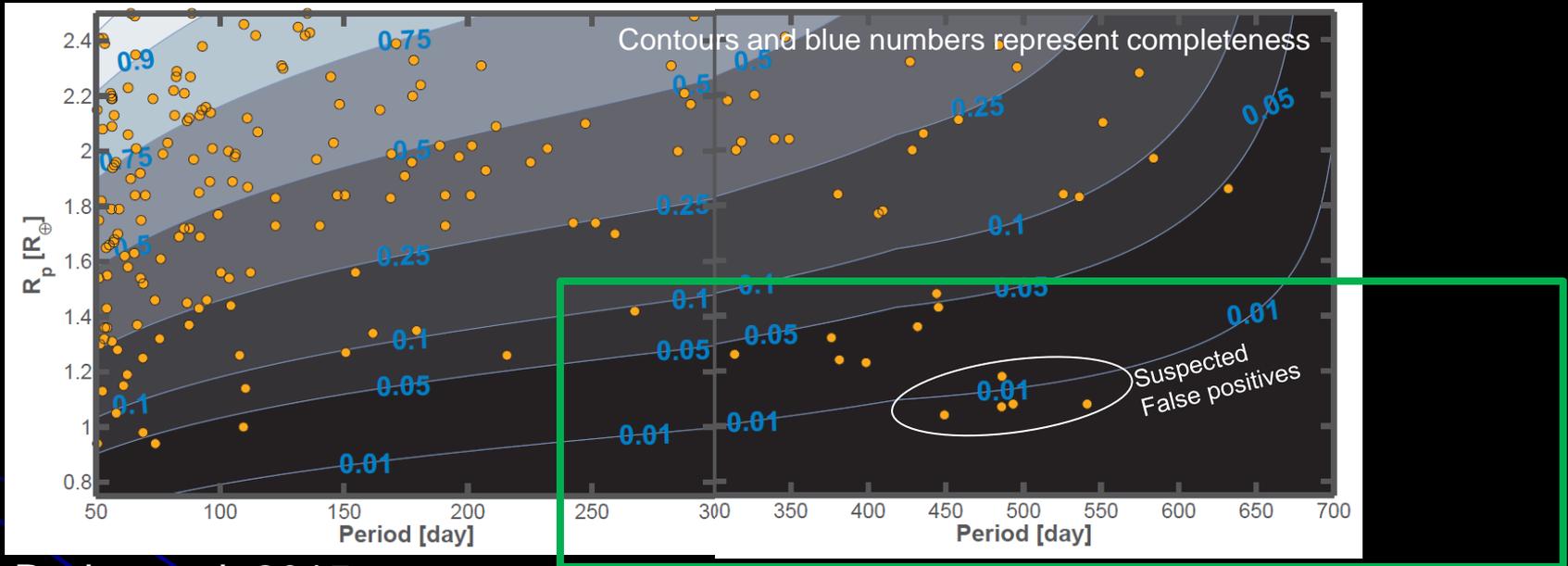


# Extrapolation vs. using long-period candidates

[potential slide, meant to show actual planets and thus better visualize Poisson uncertainty]

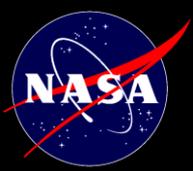
Shorter periods, more reliable

Longer periods, less reliable



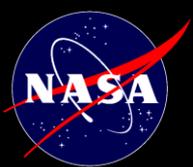
Burke et al. 2015

0.5-1.5 Earth size  
237-860 days (Kopparapu extended HZ for Sun)



# Coordination with ExEP Standards Committee

- Schedule
  - Standards team needs to have final consensus by Aug 2017
  - Standards committee product by end of 2016
  - August 2016
    - Define what the product is going to contain
- How do we extrapolate to long periods
- Mass-radius relationship
- Two versions of the green box
  - One that does not need extrapolating
  - One that does
- Pick a milestone date where the Kepler team thinks there would be no more updates



# Variance in submissions

