ExoPAG SAG13:
Exoplanet Occurrence Rates and Distributions

Belikov, Ruslan (Chair, 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
Bryson, Stephen
Cahoy, Kerri
Caldwell, Douglas
Callas, John L.
Catanzarite, Joe
Ciardi, David
Clanton, Christian
Cowan, Nick
Danchi, William
Domagal-Goldman, Shawn
Dressing, Courtney
Farr, Will
Foreman-Mackey, Daniel
Eric B. Ford
Fressin, Francois
Fulton, BJ
Gaudi, Scott
Ge, Jian
Gould, Andy
Hogg, David W
Howard, Andrew
Hsu, Danley
Kane, Stephen
Kasting, Jim
Kopparapu, Ravi
Macintosh, Bruce
Mamajek, Eric
Mandell, Avi
Mendez, Abel
Meyer, Michael
Morgan, Rhonda
Moustakas, Leonidas A.
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
Still, Martin
Suzuki, Daisuke
Swain, Mark
Thompson, Susan
Traub, Wes
Turnbull, Margaret
Unwin, Stephen
Vanderbei, Bob
Walkowicz, Luzianne
Weiss, Lauren M.
Wolfgang, Angie
Youdin, Andrew
Ziemer, John K.

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_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 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:

1. Propose standard nominal conventions, definitions, and units for occurrence rates/distributions to facilitate comparisons between different studies.

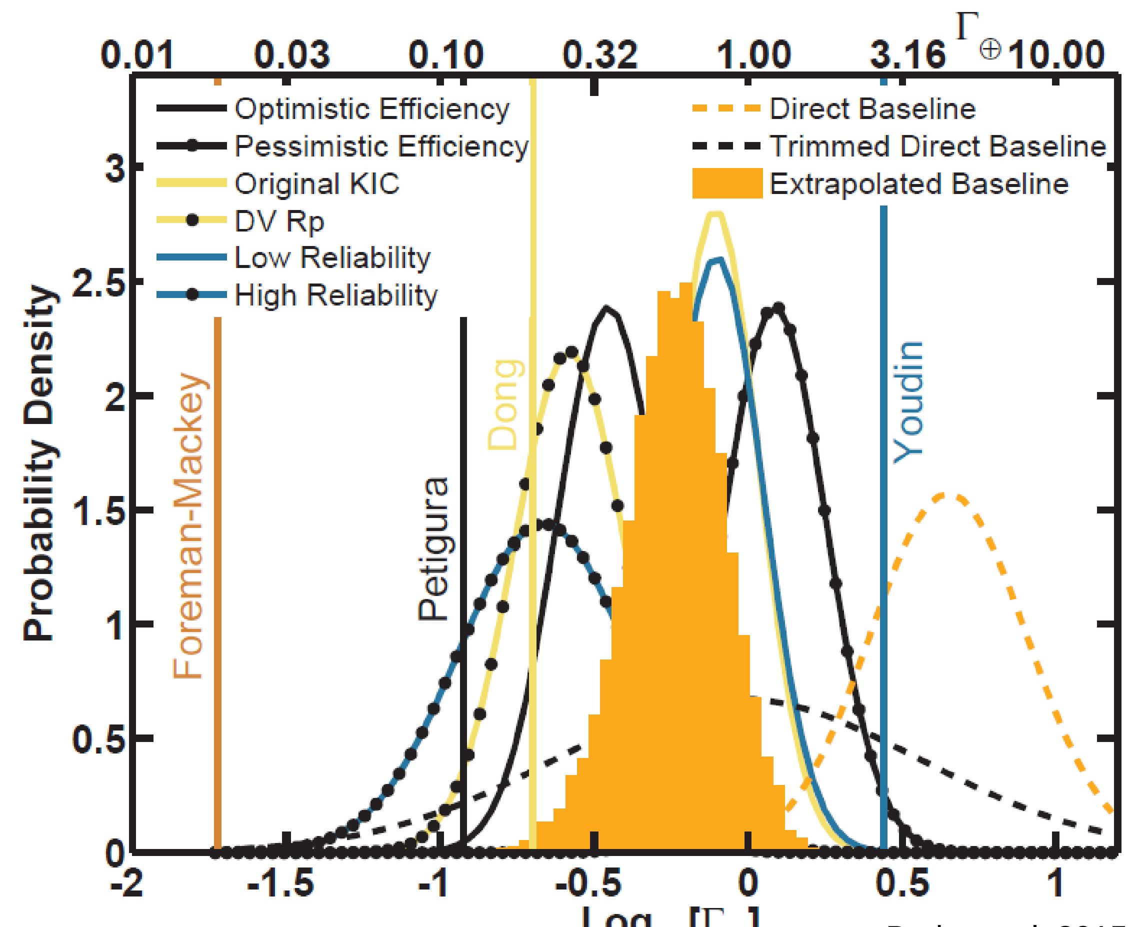
2. Do occurrence estimates from different teams/methods agree with each other to within statistical uncertainty? If not, why?

3. For occurrence rates where extrapolation is still necessary, what values should the community adopt as standard conventions for mission yield estimates?

Outline / Summary of Products

Key objective / question	Answer / Product
1. Propose standard nominal conventions, definitions, and units for occurrence rates/distributions to facilitate comparisons between different studies.	• Document proposing a standard grid of bins and other definitions https://exoplanets.nasa.gov/system/internal_resources/details/original/578_SAG13_standard_eta_definitions_v5.pdf
2. Do occurrence estimates from different teams/methods agree with each other to within statistical uncertainty? If not, why?	• Tables of occurrence rates and uncertainties from different studies, integrated across the same standard grid. • MATLAB and python code to plot and compare tables (https://drive.google.com/drive/folders/0B520NCfkP4aOOW1SWDg2cHpYOVE) • Analysis and explanations for some differences
3. For occurrence rates where extrapolation is still necessary, what values should the community adopt as standard conventions for mission yield estimates?	• Preliminary parametric model, already being used in mission yield simulations by ExEP standards team • Will be updated at ~end of Summer Once implications of DR25 are better understood
	• Extra credit: • living online repository, to enable continued community meta-analysis https://drive.google.com/drive/folders/0B520NCfkP4aOOW1SWDg2cHpYOVE • online tool to compute SAG13 occurrence rates http://www.princeton.edu/~rvdb/SAG13/SAG13.html • Study of MR relationships to help link to RV and other studies

Comparisons of Γ_{earth}



Burke et al. 2015

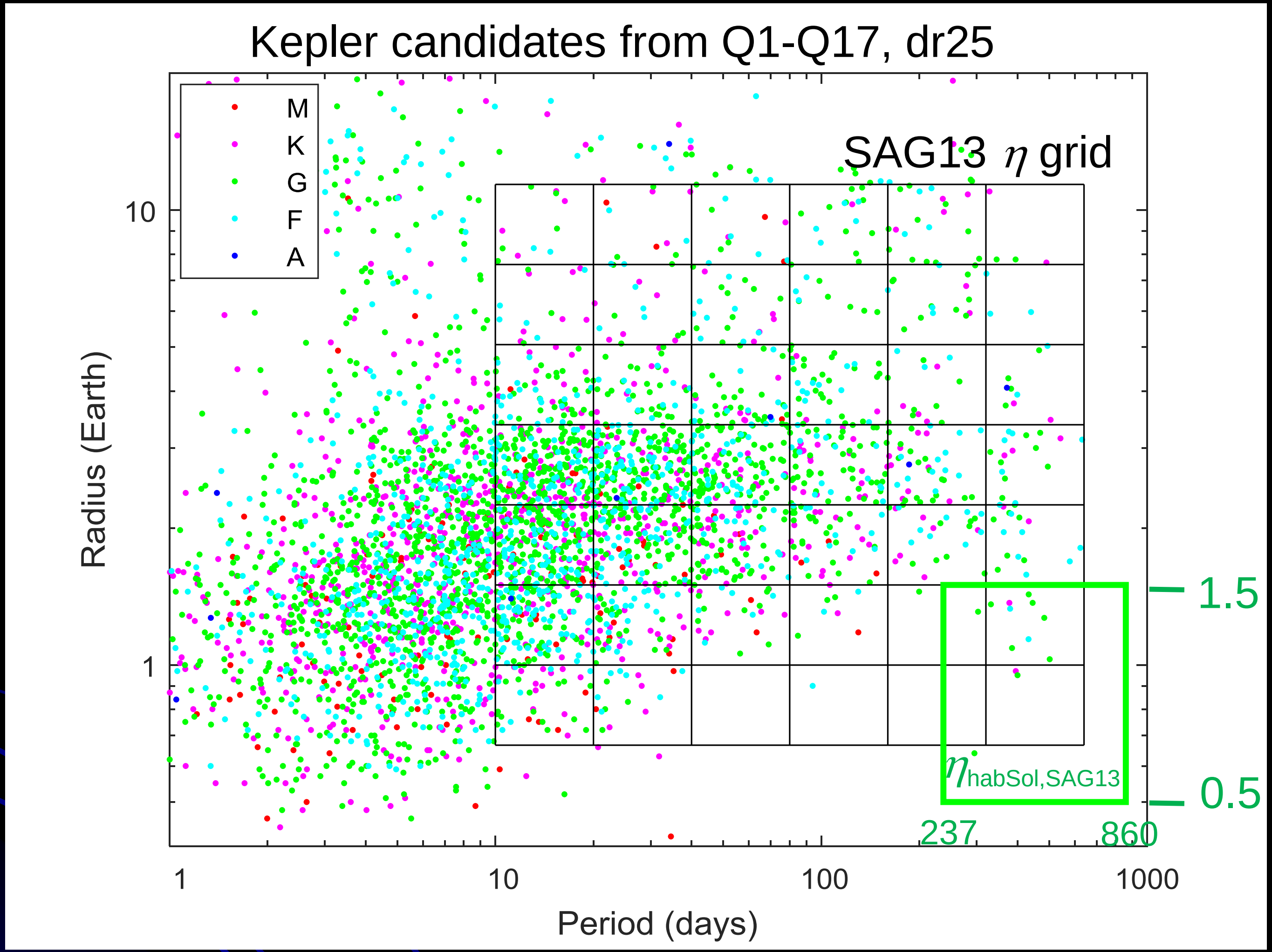
$$\Gamma_{earth} = \left.\frac{\partial^2 N(R,P)}{\partial \ln R \; \partial \ln P}\right|_{R=1,P=1y}$$

Γ_{earth} is independent of definitions of HZ or habitable size range

For most definitions of η_{Earth}, $\Gamma_{earth} \sim \eta_{Earth}$

Burke et al. 2015: "We generally find higher planet occurrence rates and a steeper increase in planet occurrence rates towards small planets than previous studies of the Kepler GK dwarf sample"

Standardized eta grid



12+ community sourced occurrence tables

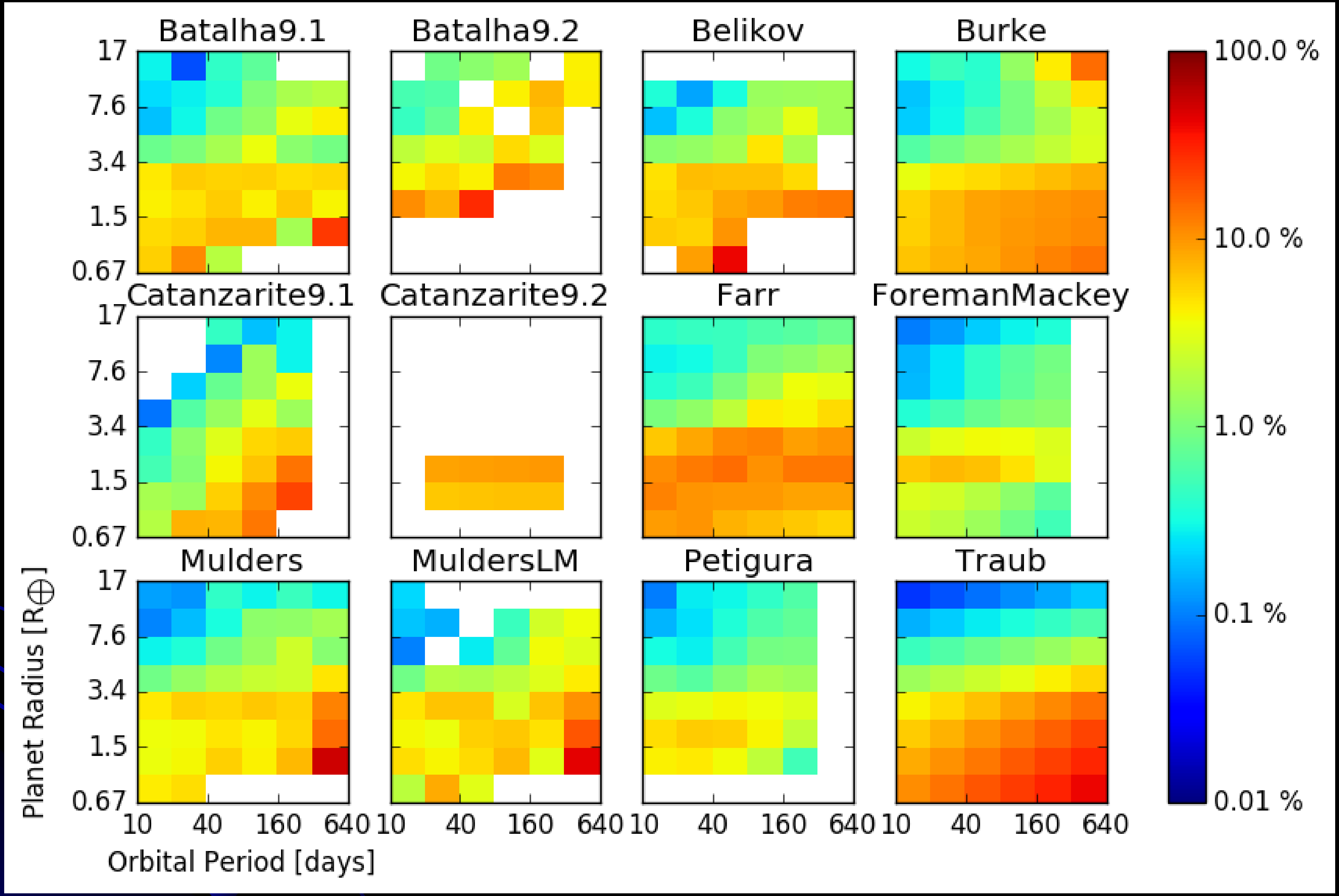
Batalha, Natalie (2)
Belikov, Rus
Burke, Chris
Catanzarite, Joe (2)
Dressing, Courtney*
Farr, Will
Foreman-Mackey, Daniel*
Fulton, BJ
Kopparapu, Ravi
Mulders, Gijs (2)
Petigura, Erik*
Traub, Wes*

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

All datasets and documents can be found on SAG13 repository: https://drive.google.com/drive/folders/0B520NCfkP4aOOW1SWDg2cHpYOVE

- $\eta_{habSol,SAG13}$
  - R = [0.5 – 1.5], P = [237 860] (Kopparapu optimistic HZ for Sol twin)
  - This is not exactly η_{Earth} , just a rough representation

Example: selected submitted occurrence rates for G-dwarfs



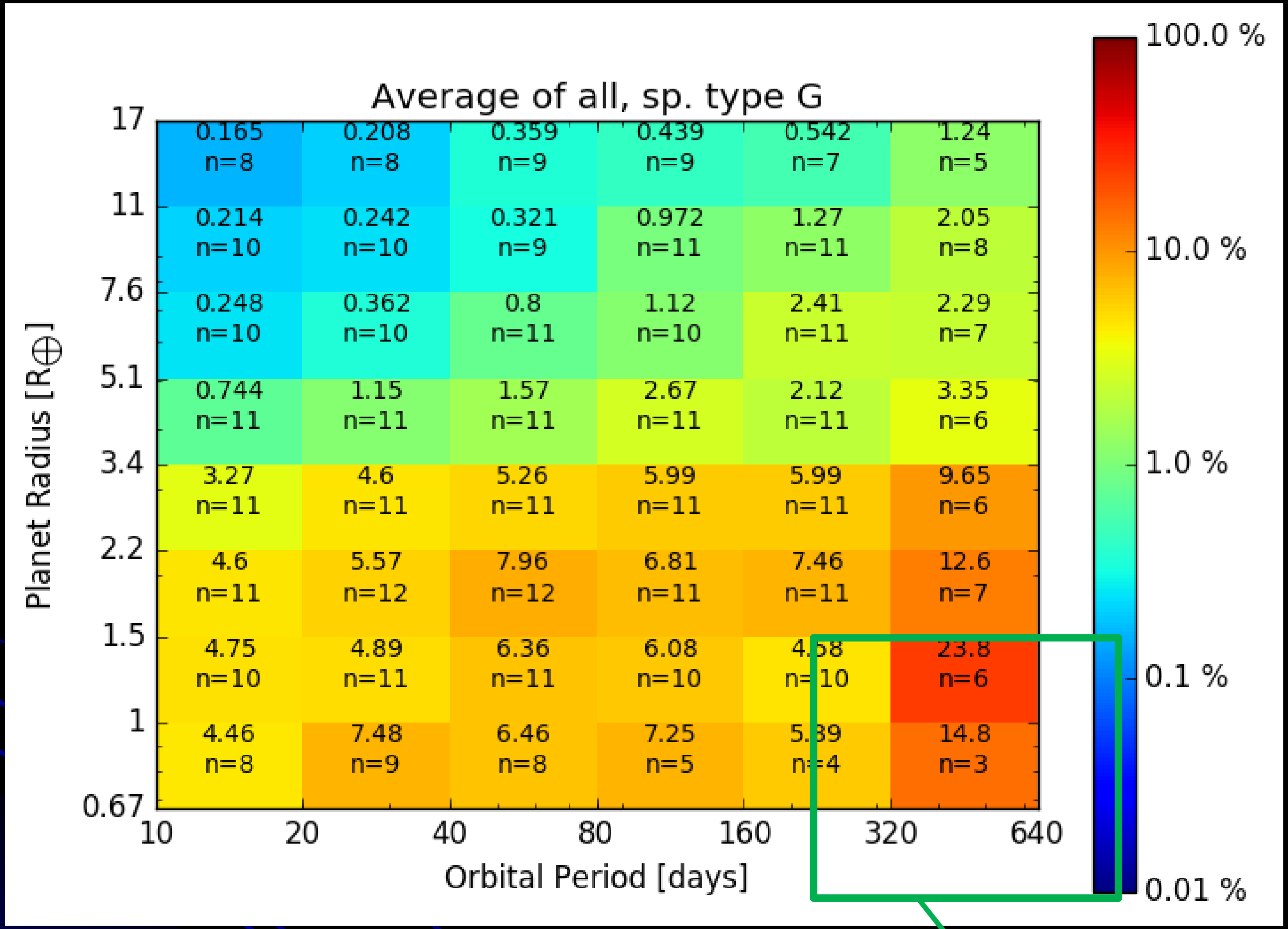
Plots and analysis are generated with the make_plots.py scriptby Gijs Mulders.
Full data and plots available online at https://drive.google.com/drive/folders/0B520NCfkP4aOOW1SWDg2cHpYOVE

G-dwarf average

legend

% occurrence

# of submissions



$\eta_{habSol,SAG13}$ ~ 0.58
(based on best 2-piece power law fit)

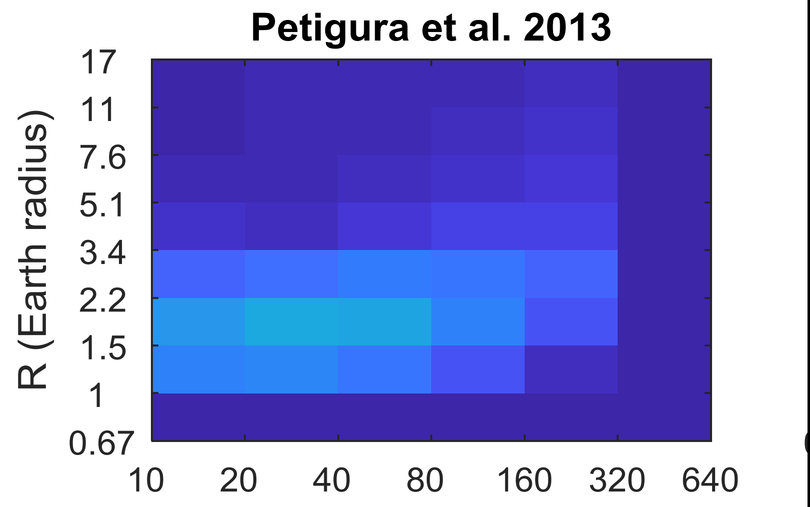
- Simple geometric average across submissions
- Scientifically not very meaningful because it does not account for e.g. dependencies between submissions.
  - majority of submissions were based on DR24, so the result is close to Burke et al. 2015
  - A formal meta-analysis requires more resources, but more meaningful combinations will be included in the written report.
- Value is primarily in being a standard assumption that a community can (perhaps) agree to for mission studies, while we wait for a formal scientific value

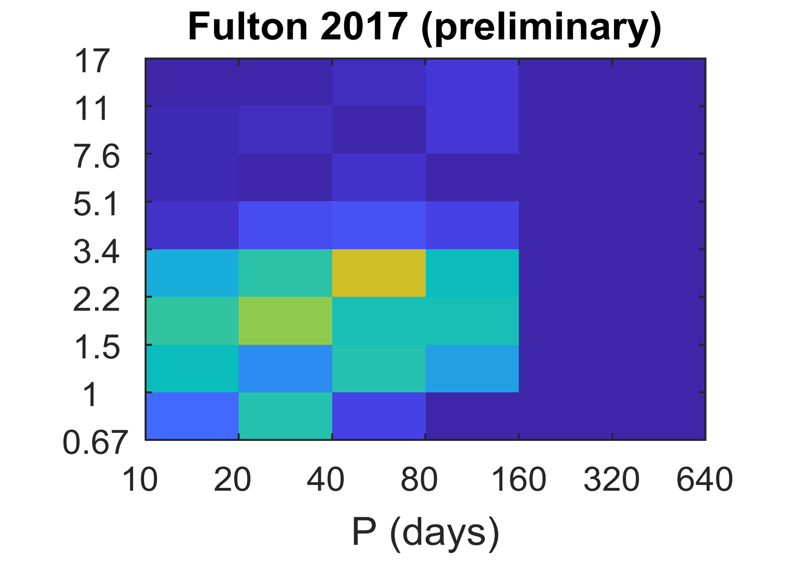
Plots and analysis are generated with the make_plots.py script in the SAG13 Google drive, code by Gijs Mulders.

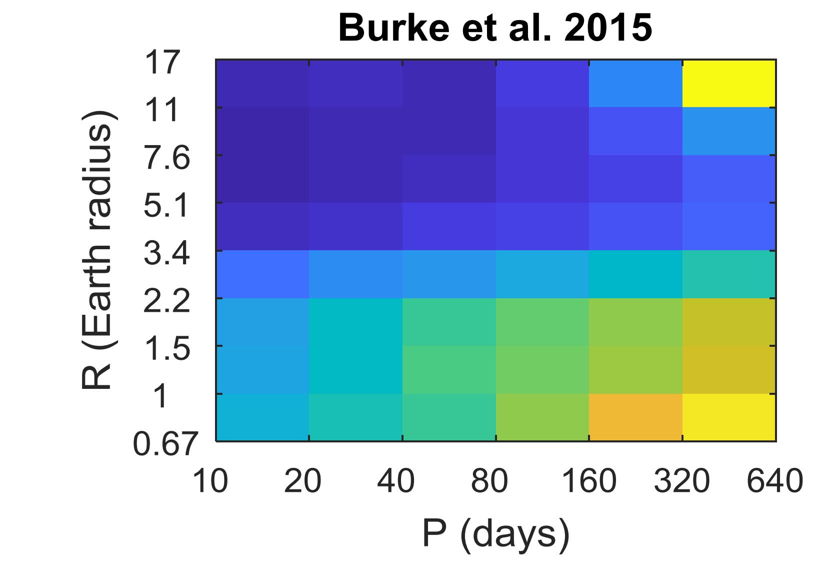
Example comparison of selected occurrence rates

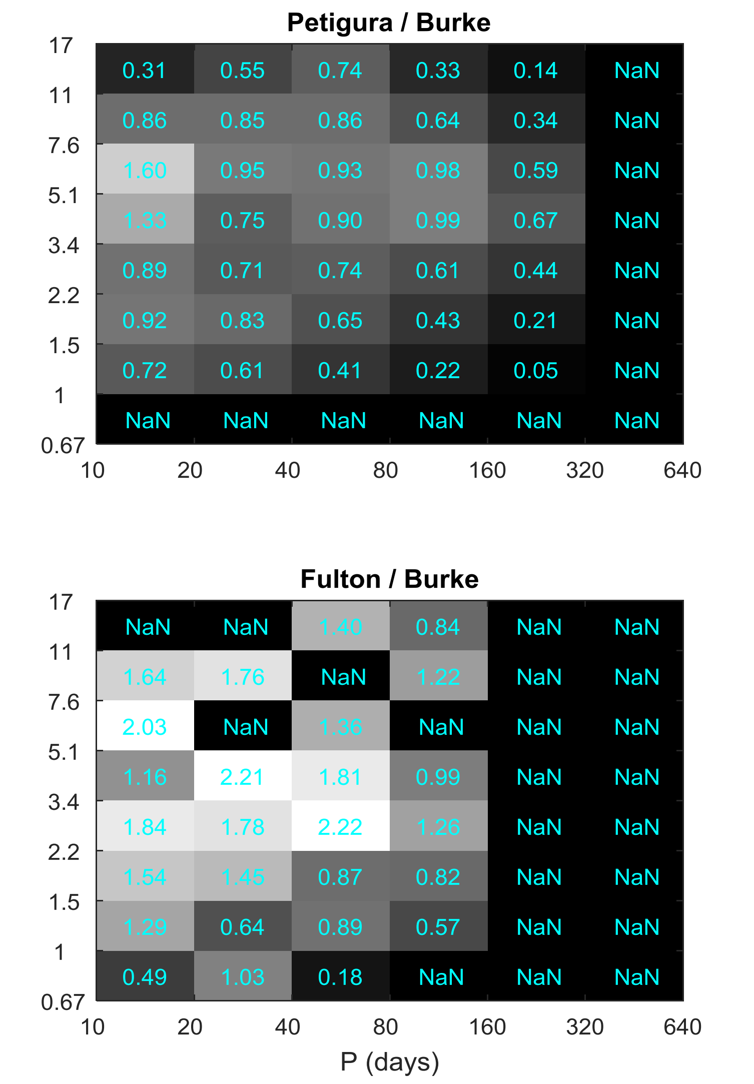
Occurrence rates for G-dwarfs from different studies

Ratios









Closing the factor of ~4 gap between Petigura 2013 and Burke 2015

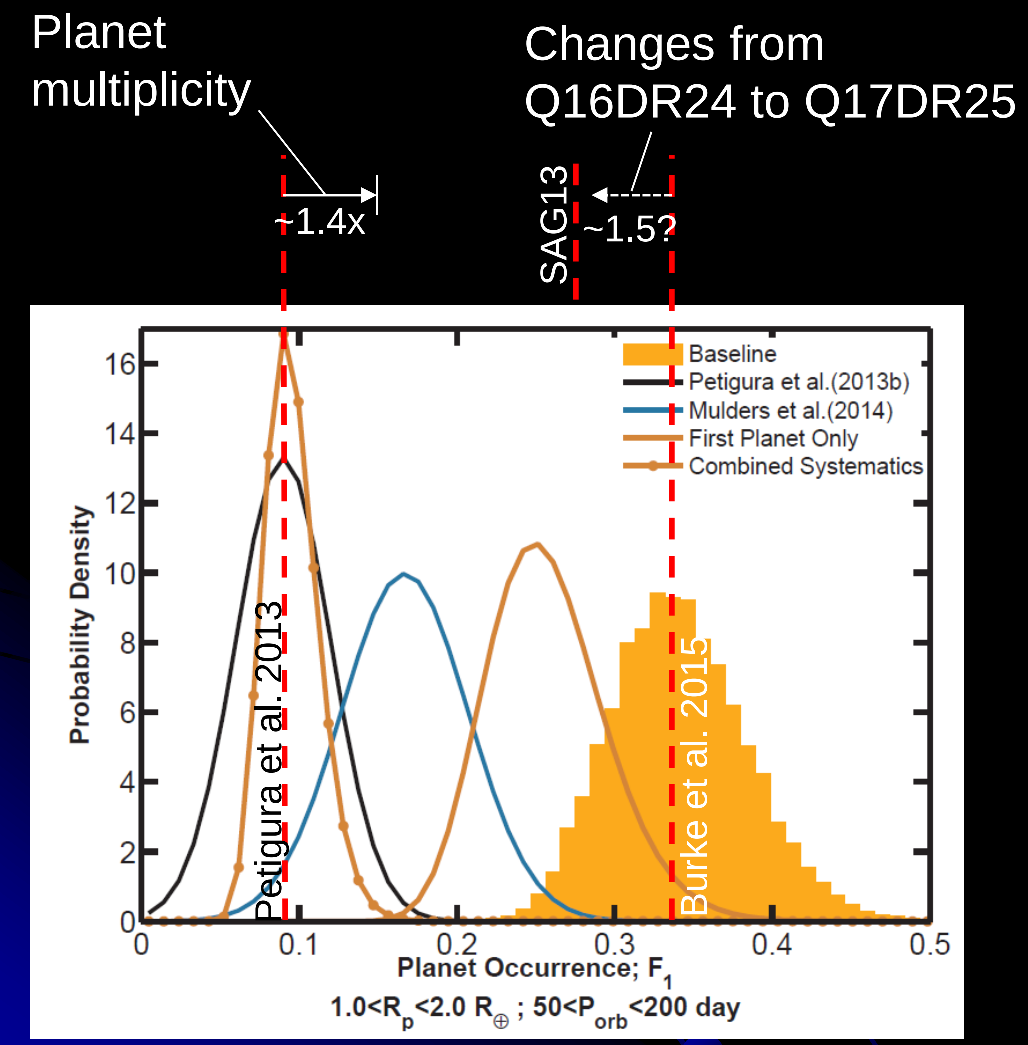


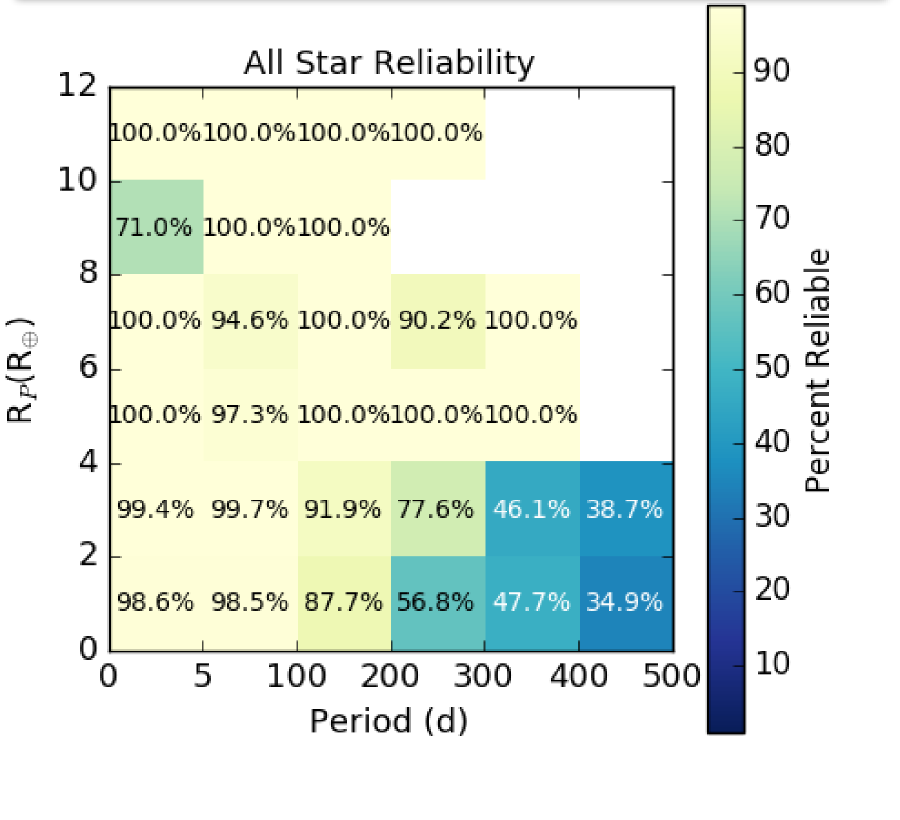
figure from Burke et al. 2015

- Petigura 2013 counted the largest planet in the system, while Burke 2015 considered all planets (a factor of ~1.4 difference)
- Changes from DR24 to DR25 may slightly decrease the rates in Burke et al. 2015:
  - Star sizes are slightly larger, hence planets are slightly larger
    - # of {50<P<300, 0.75<R<2.5} planet candidates decreased from 156 to 118
    - Stellar models in Petigura 2013 may have been “closer” to DR25 b/c they were partially based on spectroscopy
  - Detection contours have slightly better recovery than Q16
- Remaining factor of 2 gap remains unexplained
  - Reliability is a potentially critical source of differences, not yet fully explored

Reliability

- For Rp < 4 Re, P > 100 days you must account for reliability
  - Some PCs are not real planets
- DR25 is the first catalog to measure reliability
  - *Inverted* and *Scrambled* data measure instrumental reliability
  - Offset and EB injections provide insight into which astrophysical false positives are undetectable
- FPP table measures astrophysical reliability
- Accounting for reliability in occurrence rate estimates is an open problem

DR25 measured instrumental reliability

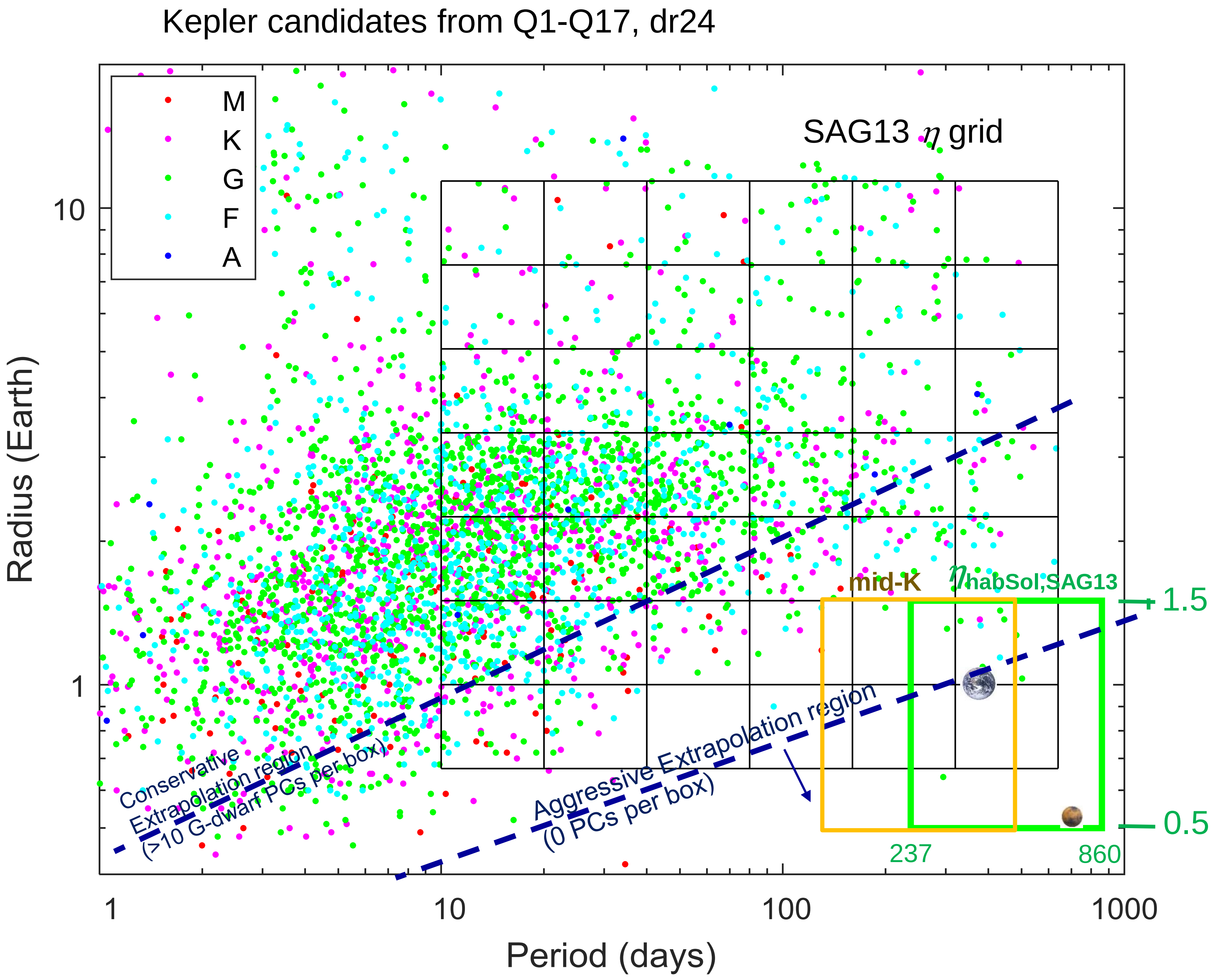


Summary of sources of differences in small, long period planet occurrence rates

- Differences in bin boundaries and units!
  - Especially the small size boundary

- Large systematic differences (> factor of 2 in occurrence rates) mostly traced to differences in:
  - Catalog and other data products (completeness curves, etc.)
    - DR24 lead to systematically higher numbers than many prior studies (~3-4x)
    - DR25 is likely to be a little bit lower than DR24 (perhaps ~1.5x)

- The following typically cause only small systematic differences (< factor of 2)
  - estimation method
  - details of the estimation code
  - extrapolation


Kepler candidates from Q1-Q17, dr24
M
K
G
F
A
SAG13 η grid
Radius (Earth)
10
1
Period (days)
1
10
100
1000
mid-K
$\eta_{habSol,SAG13}$
1.5
0.5
237
860
Conservative
Extrapolation region
(>10 G-dwarf PCs per box)
Aggressive Extrapolation region
(0 PCs per box)

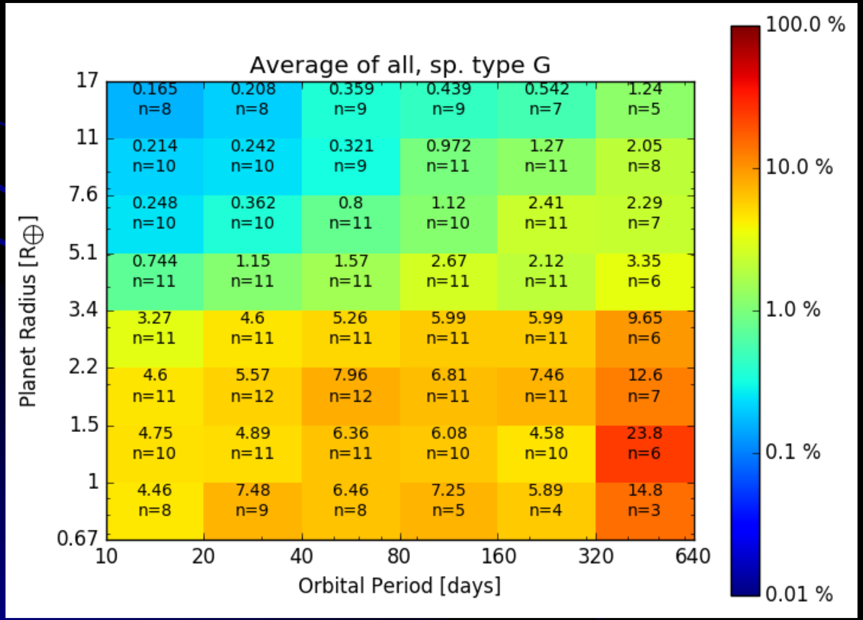
Preliminary parametric fit (for G-dwarfs)

$$\frac{\partial^2 N(R,P)}{\partial \ln R \, \partial \ln P} = \Gamma_i R^{\alpha_i} P^{\beta_i} \qquad \text{in region } R_{i-1} \le R < R_i$$

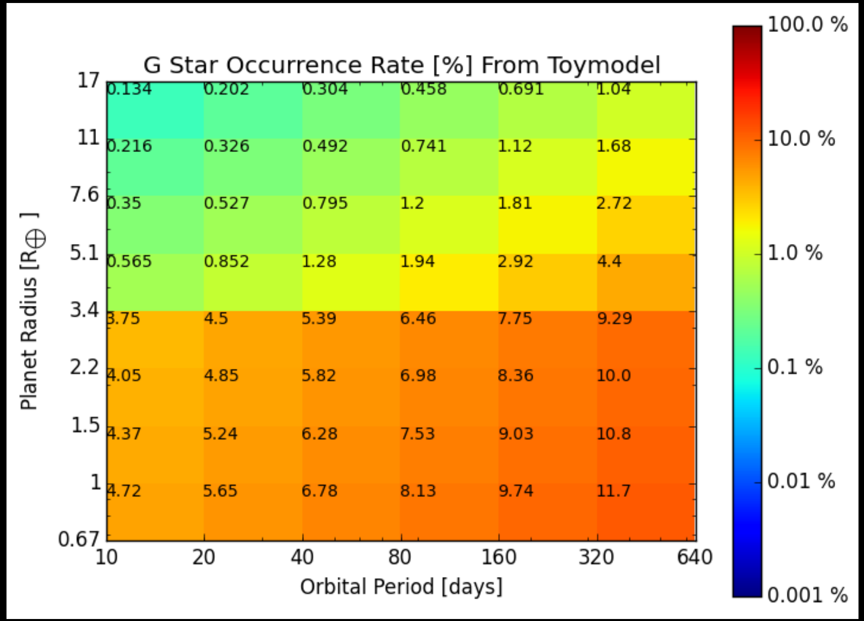
(R in Earth radius, P in years)

Γ_i	α_i	β_i	R_i
0.38	**-0.19**	**0.26**	**3.4**
0.73	-1.18	0.59	Inf

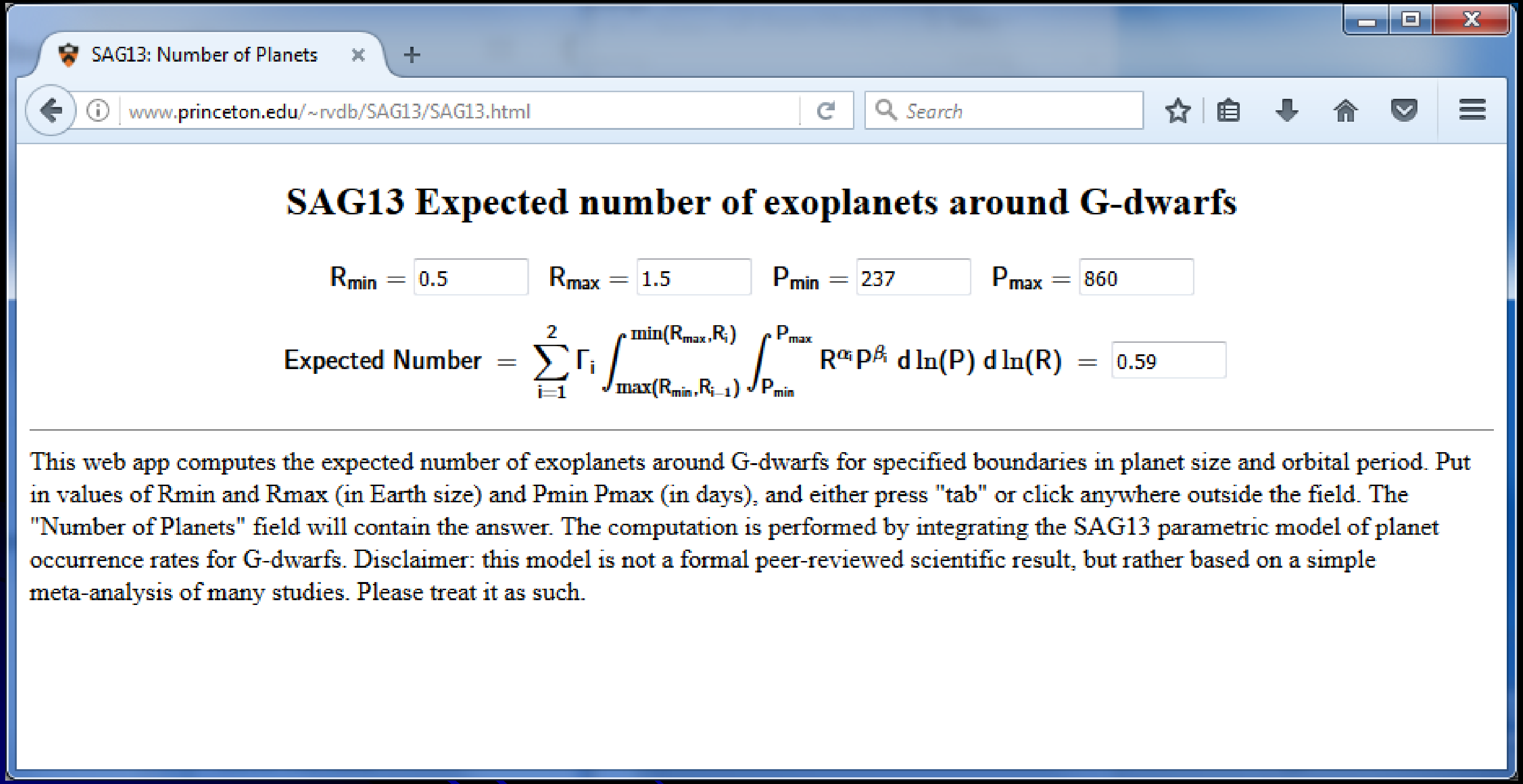
Submission average



Parameteric fit (integrated across bins)



Online occurrence rate calculator



- Preliminary online implementation (by Bob Vanderbei)
- If there is interest, other SAG13 tools and code can be deployed as web apps
- Disclaimer: the SAG13 model used in this tool is NOT a formal peer-reviewed scientific result, but rather based on a simple meta-analysis of many studies. Please treat it as such.

Calculations of habitable occurrence rates (example for G-dwarfs)

Integrating SAG13 parametric fit
web app: http://www.princeton.edu/~rvdb/SAG13/SAG13.html

		Habitable Zone*	
		Conservative	Optimistic
Planet radius range	1.0-1.5	$\mathbf{0.14^{+0.12}_{-0.04}}$	$\mathbf{0.2^{+0.18}_{-0.06}}$
	0.5-1.5	$\mathbf{0.40^{+0.48}_{-0.14}}$	$\mathbf{0.58^{+0.7}_{-0.2}}$

(uncertainties correspond to 1-sigma equivalent deviations across submissions)

$\eta_{\mathrm{habSol,SAG13}}$

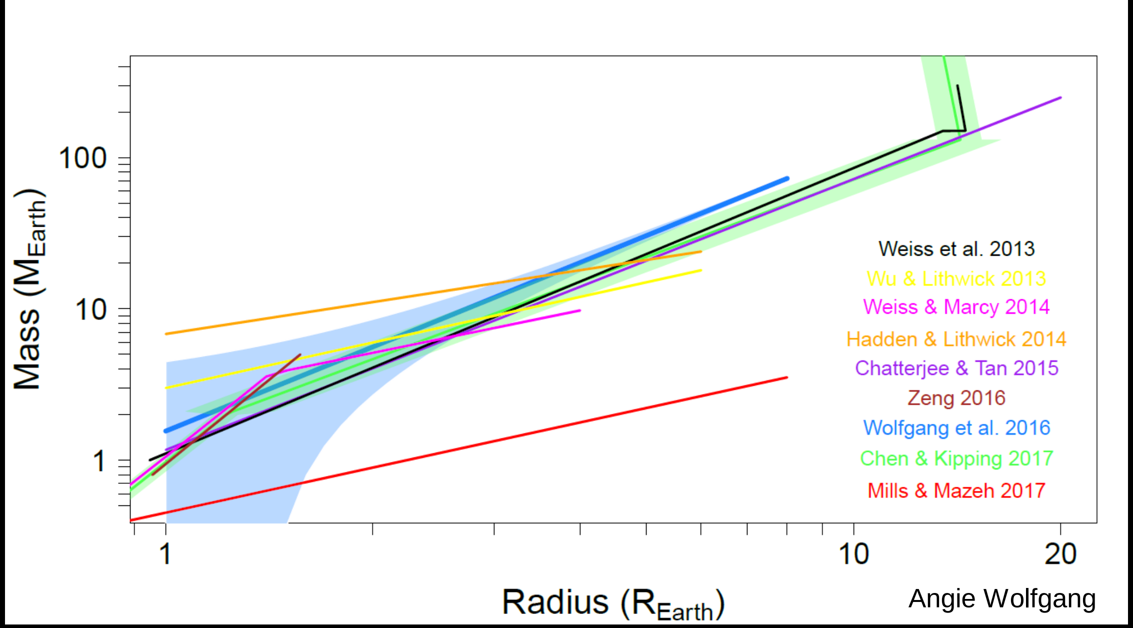
Using Burke et al. 2015 posterior tool
https://github.com/christopherburke/KeplerPORTs

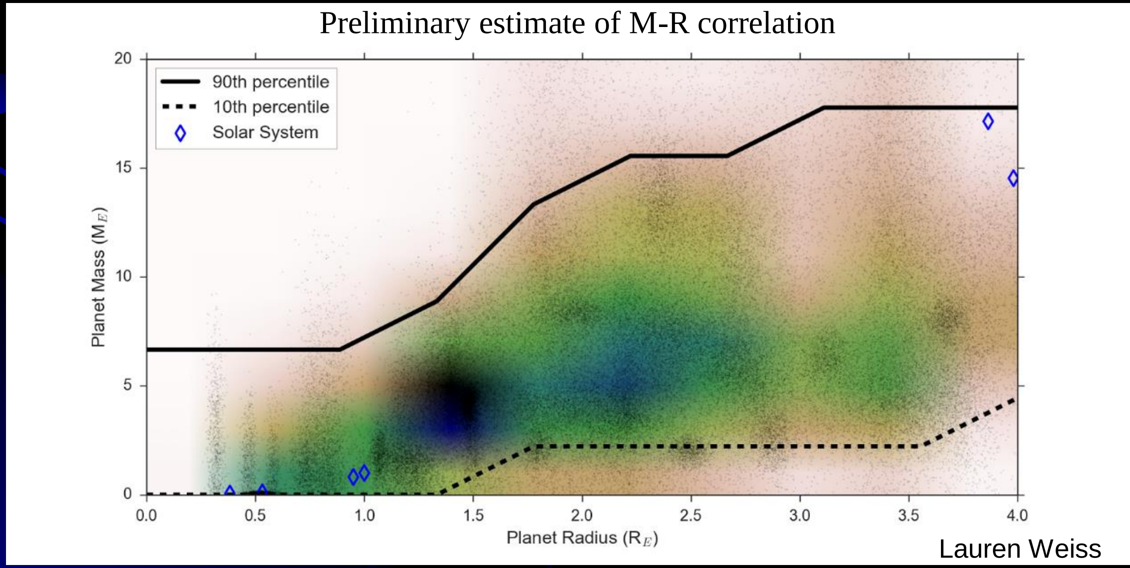
		Habitable Zone*	
		Conservative	Optimistic
Planet radius range	1.0-1.5	$\mathbf{0.21^{+0.08}_{-0.08}}$	$\mathbf{0.31^{+0.1}_{-0.1}}$
	0.5-1.5	$\mathbf{0.5^{+0.4}_{-0.2}}$	$\mathbf{0.73^{+0.6}_{-0.3}}$

*Habitable zone definitions are from Kopparapu 2013 for Solar twin
Conservative: 338-792 days; Optimistic: 237-864 days

Converting between Mass and Radius

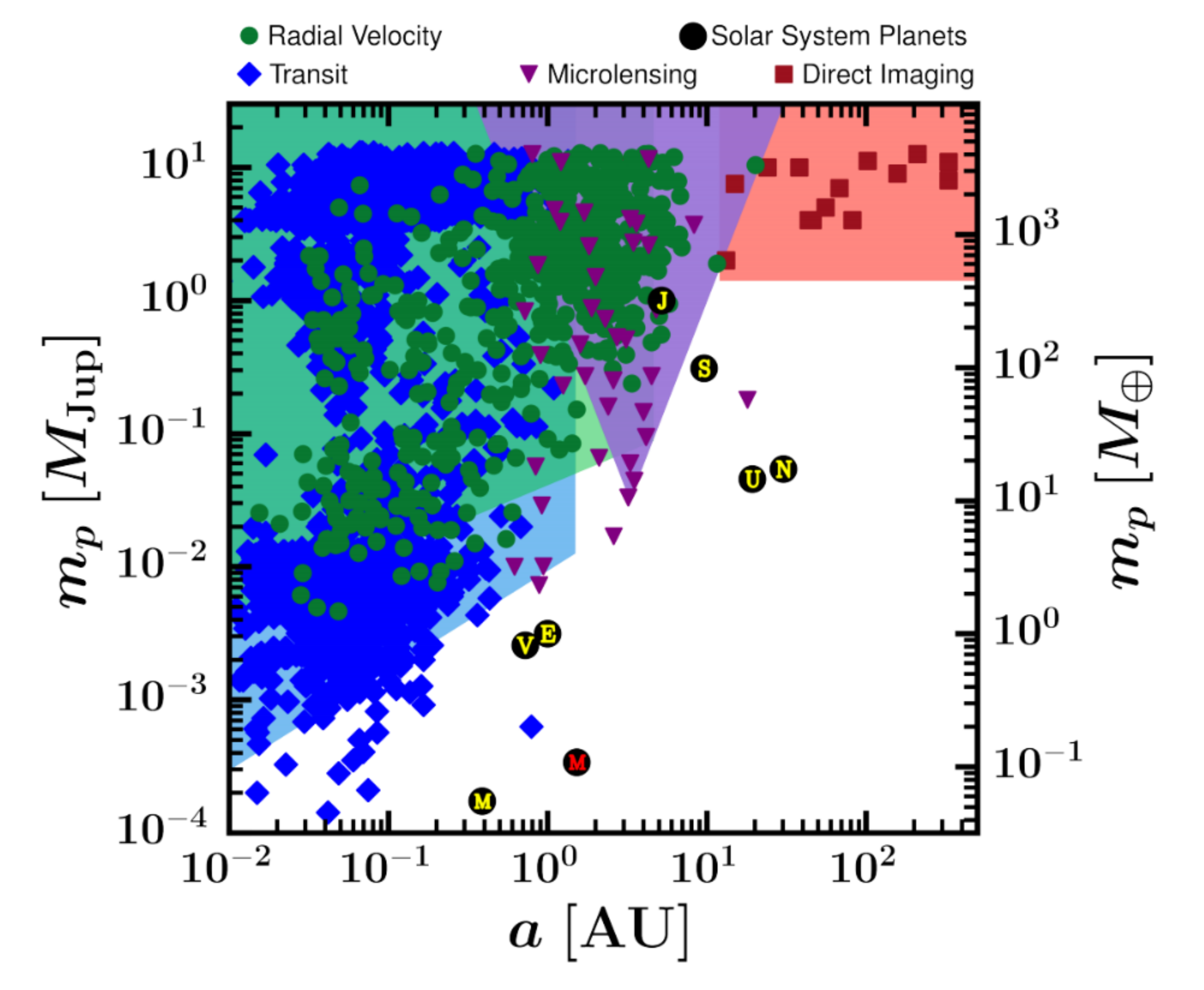
(focus group led by Angie Wolfgang and Lauren Weiss)





- 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
- 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)

Linking to results from non-Transit techniques (Christian Clanton)



Conclusions

- The average SAG13 occurrence rates for potentially habitable planets may be higher than what has been commonly adopted in the past
  - $\eta_{habSol,SAG13}$ ~ 0.6 (for 0.5-1.5 Earth size, 237-864 days)
  - Slightly lower than the latest peer-reviewed estimate from the Kepler team (Burke et al. 2015)
  - Current SAG13 model represents a point in time and not a formal scientific result; DR25 may lower occurrence rates

- Future work is still necessary to reduce systematics and uncertainties (outside the scope of SAG13)
  - Rigorous estimate based on DR25
  - Reliability remains a concern

- Summary of SAG 13 products:
  1. Proposed standard grid of bins and other definitions
  2. Tables of combined occurrence rates and uncertainties from different studies across that grid.
  3. Analysis of differences between studies and some known explanations
  4. Parametric model to be used for mission yield simulations
  5. Online tools to plot SAG13 tables and compute occurrence rates