ExoPAG SAG13: Exoplanet Occurrence Rates and Distributions

Ruslan Belikov
on behalf of SAG13

Note: some results in this presentation are still preliminary

ExoPAG 15, Grapevine, TX, 1/2/2017

SAG13 members

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
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
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_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.

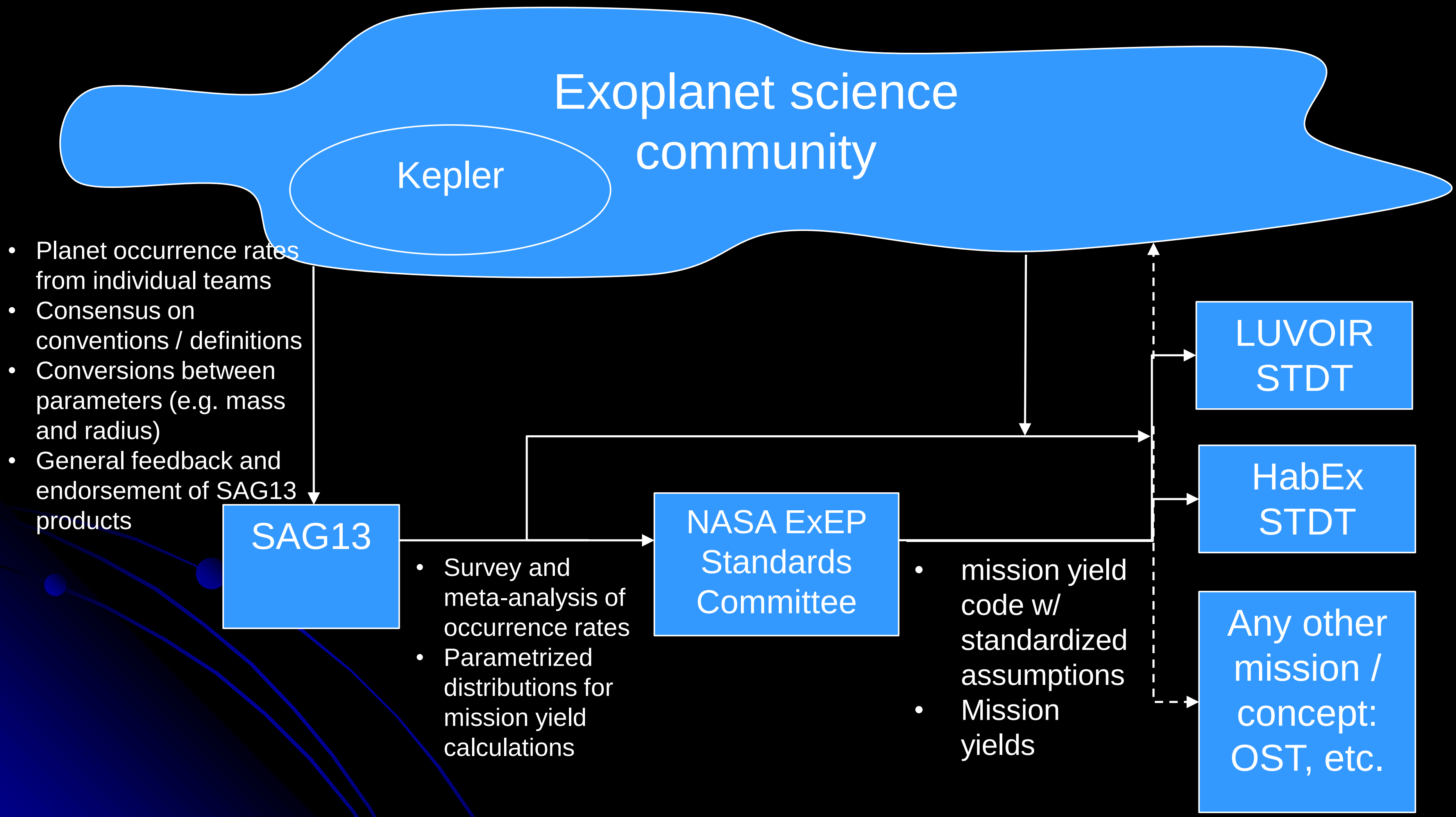
Key objectives and questions:

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?

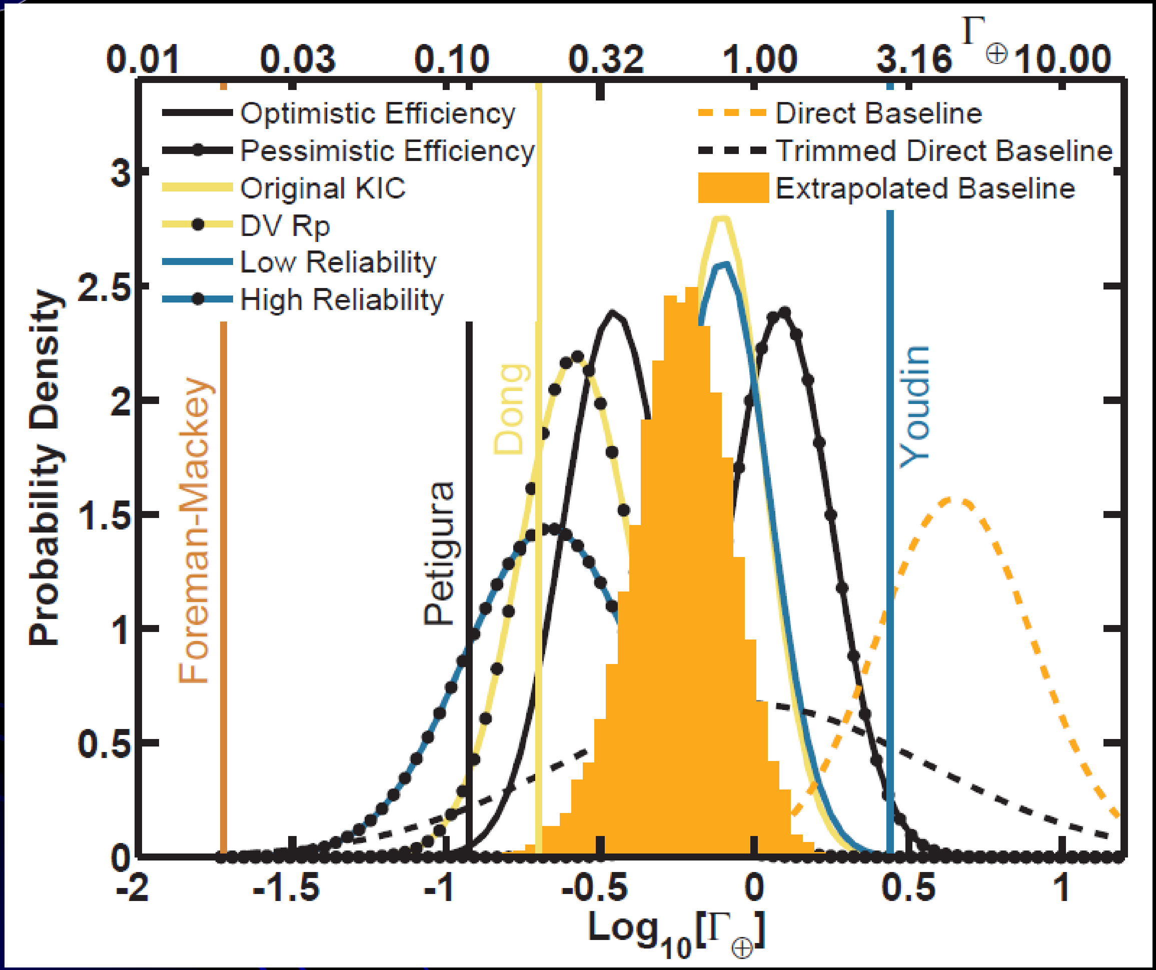
Current activity →
3. For occurrence rates where extrapolation is still necessary, what values should the community adopt as standard conventions for mission yield estimates?

SAG13 role



Exoplanet science community
Kepler
• Planet occurrence rates from individual teams
• Consensus on conventions / definitions
• Conversions between parameters (e.g. mass and radius)
• General feedback and endorsement of SAG13 products
SAG13
• Survey and meta-analysis of occurrence rates
• Parametrized distributions for mission yield calculations
NASA ExEP Standards Committee
• mission yield code w/ standardized assumptions
• Mission yields
LUVOIR STDT
HabEx STDT
Any other mission / concept: OST, etc.

From Burke et al. 2015



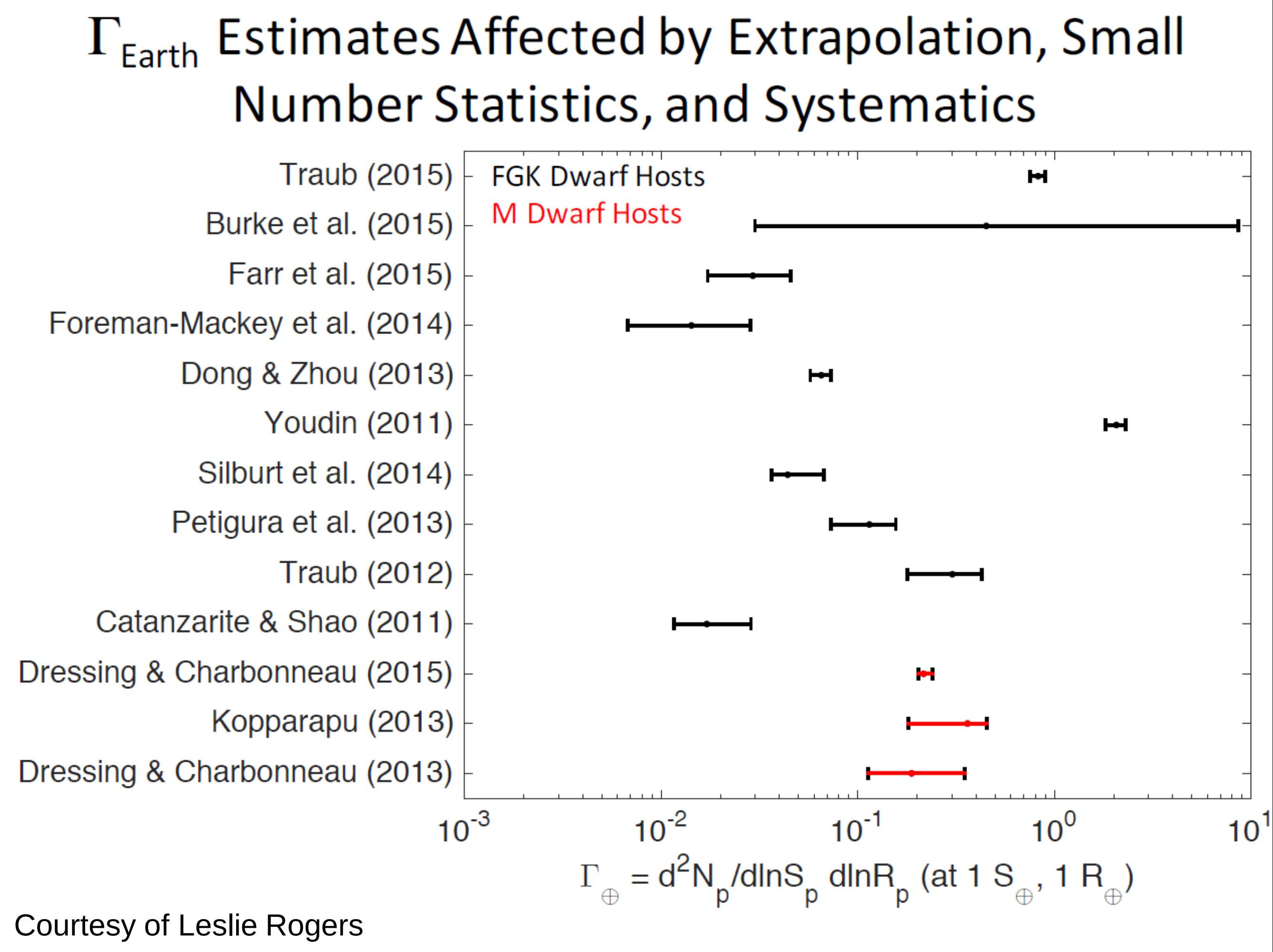
$$\Gamma_{\text{earth}} = \left.\frac{\partial^2 N(R,P)}{\partial \ln R \, \partial \ln P}\right|_{R=1,P=1y}$$

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

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

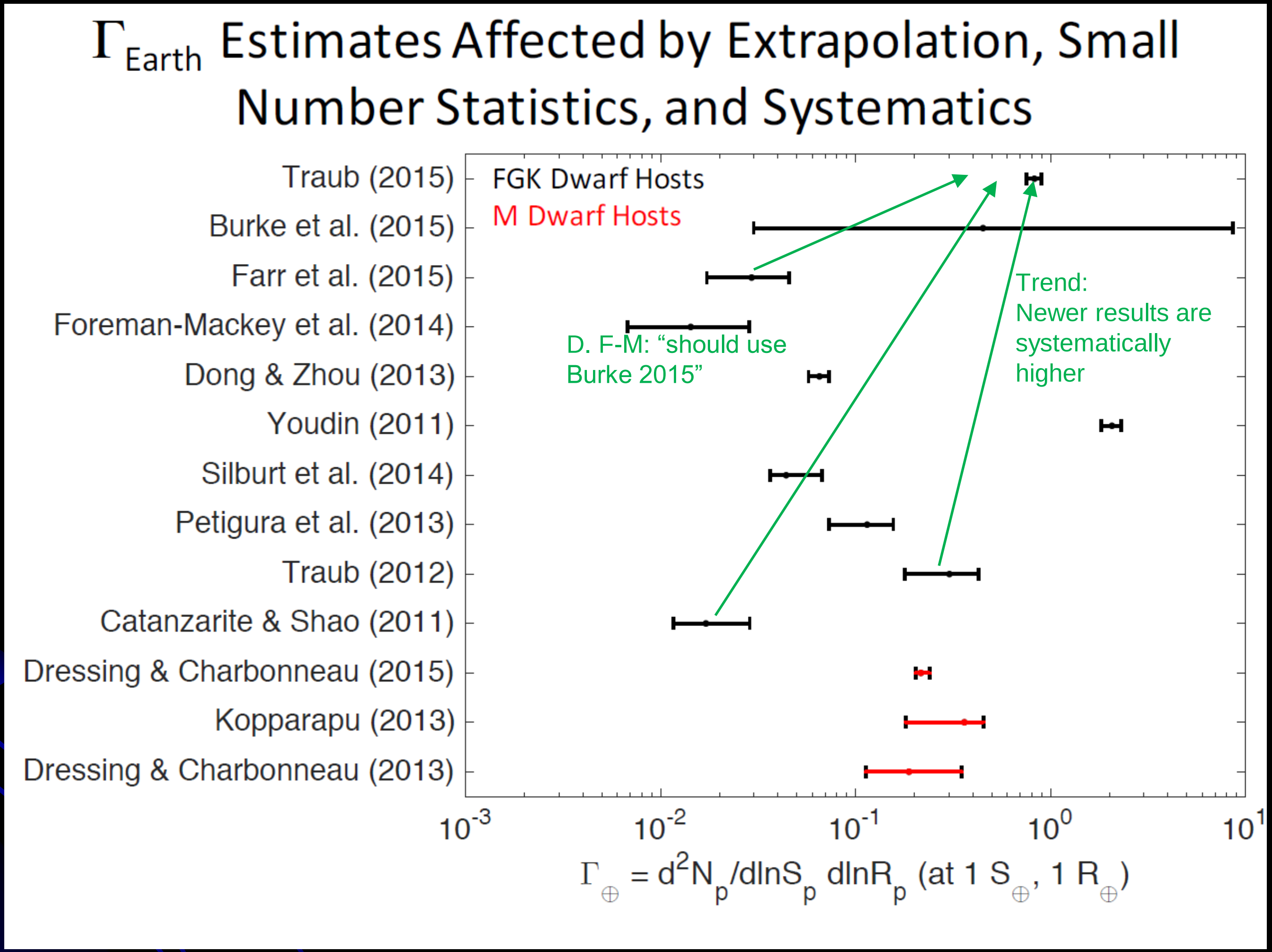
"We generally find higher planet occurrence rates and a steeper increase in planet occurrence rates towards small planets than previous studies of the Kepler GK dwarf sample"

Comparison of Γ_{earth} from different publications



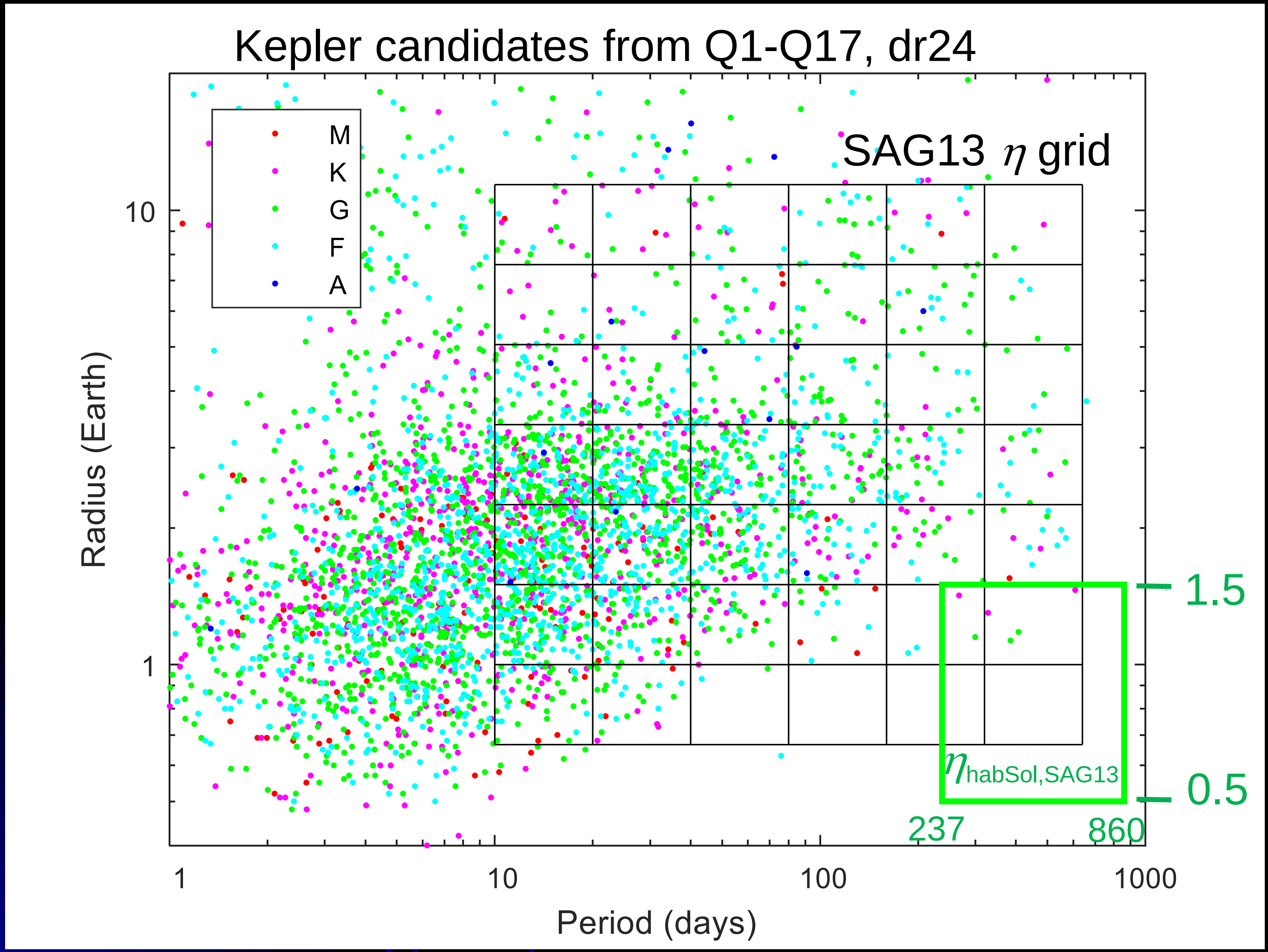
- Initially, it appears that the possible range of Γ_{earth} spans 2-3 orders of magnitude
- This is true, but extremely conservative: only the middle ~couple of octaves are "likely"

Comparison of Γ_{earth} from different publications



- Initially, it appears that the possible range of Γ_{earth} spans 2-3 orders of magnitude
- This is true, but extremely conservative: only the middle ~couple of octaves are "likely"

Standardized eta grid



12 community sourced occurrence tables

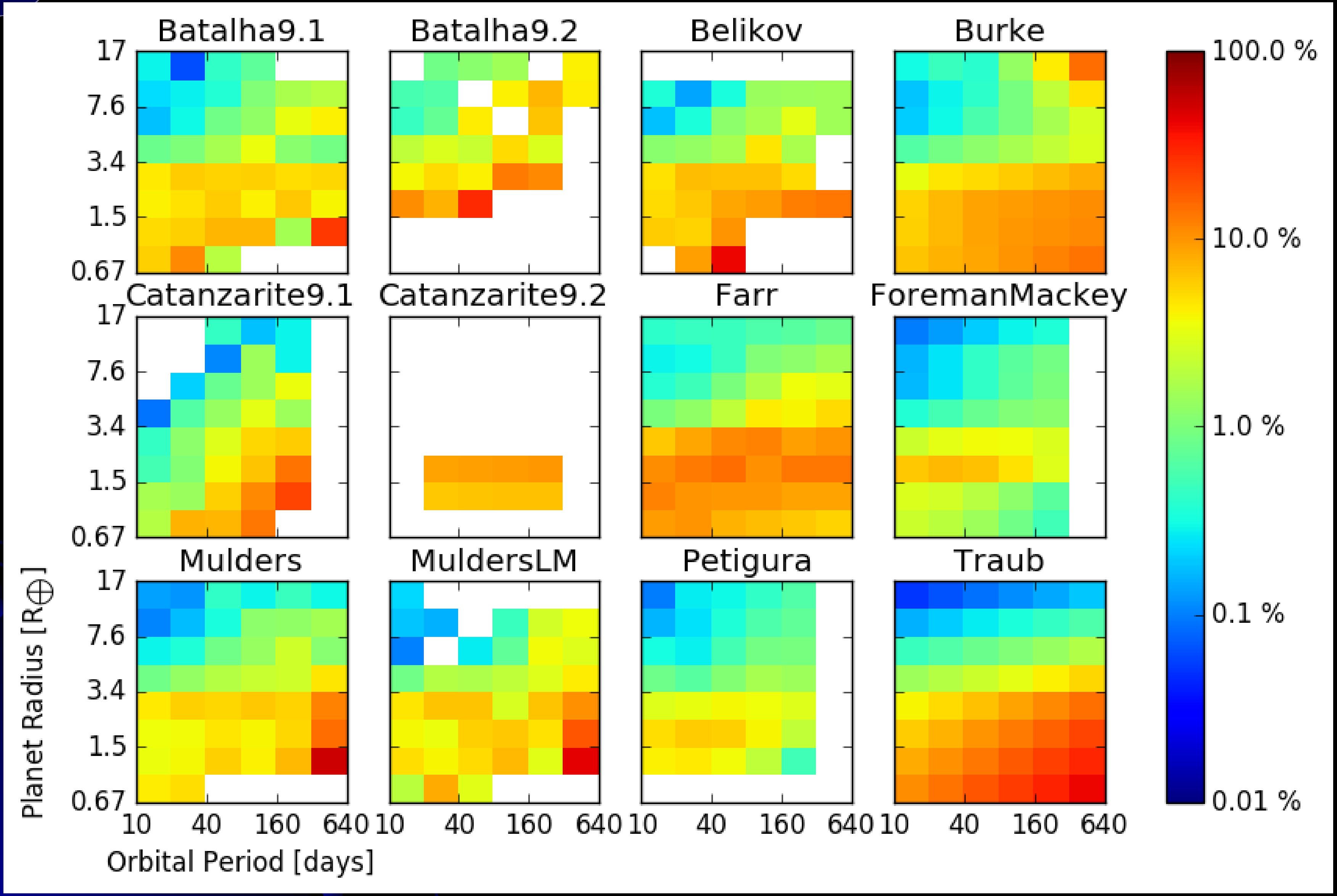
Batalha, Natalie (2)
Belikov, Rus
Burke, Chris
Catanzarite, Joe
Dressing, Courtney*
Farr, Will
Foreman-Mackey, Daniel*
Kopparapu, Ravi
Mulders, Gijs
Petigura, Erik*
Traub, Wes*

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

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 optimistic HZ for Sol twin)
  - This is not exactly η_{Earth} , just a tentative rough representation of a potentially habitable region

Example: submitted occurrence rates for G-dwarfs



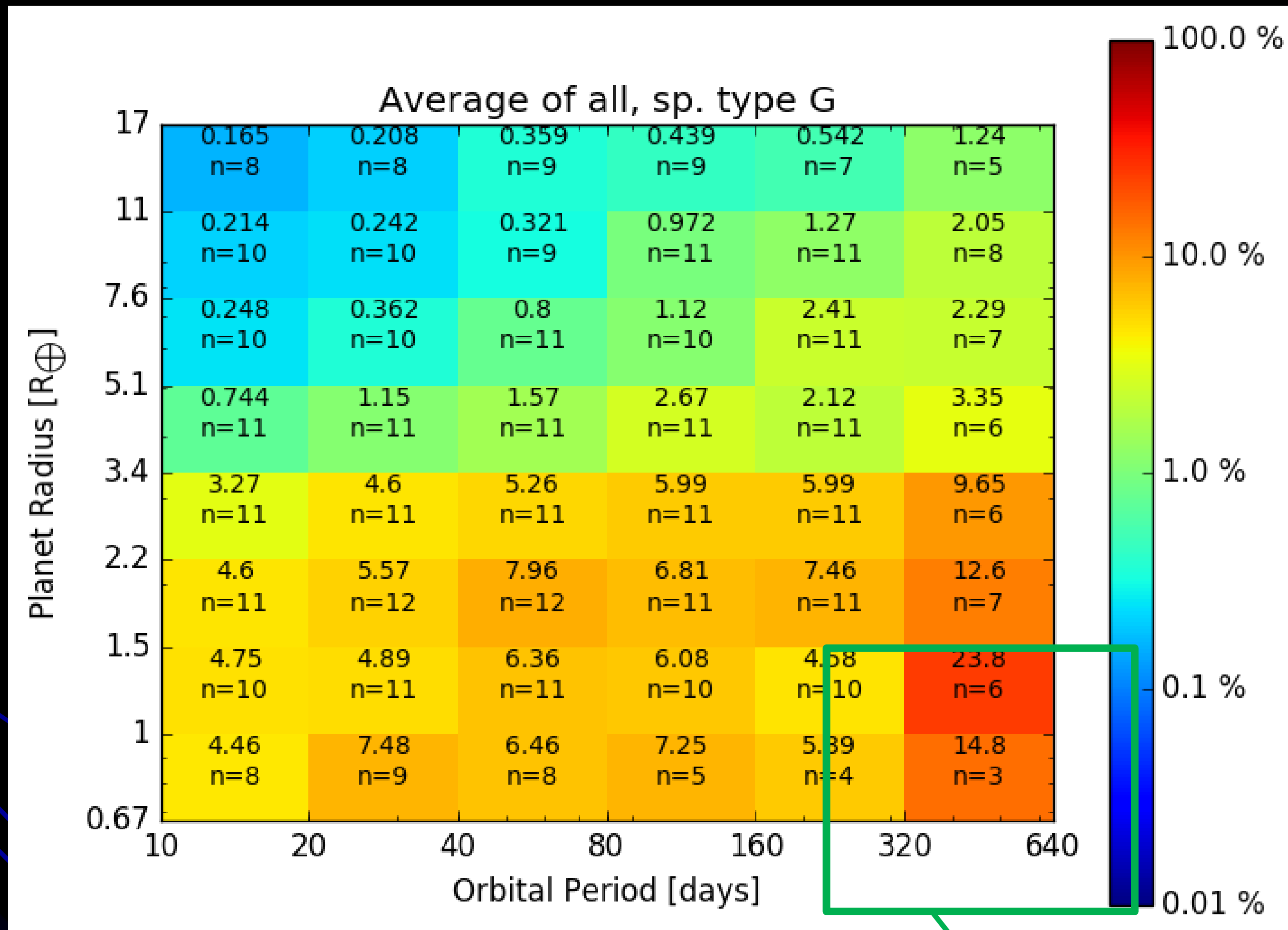
Plots and analysis are generated with the make_plots.py script in the SAG13 Google drive, code by Gijs Mulders.

Closer look at G-dwarf average

legend

% occurrence

# of submissions



$\eta_{habSol,SAG13}$ ~ 0.58

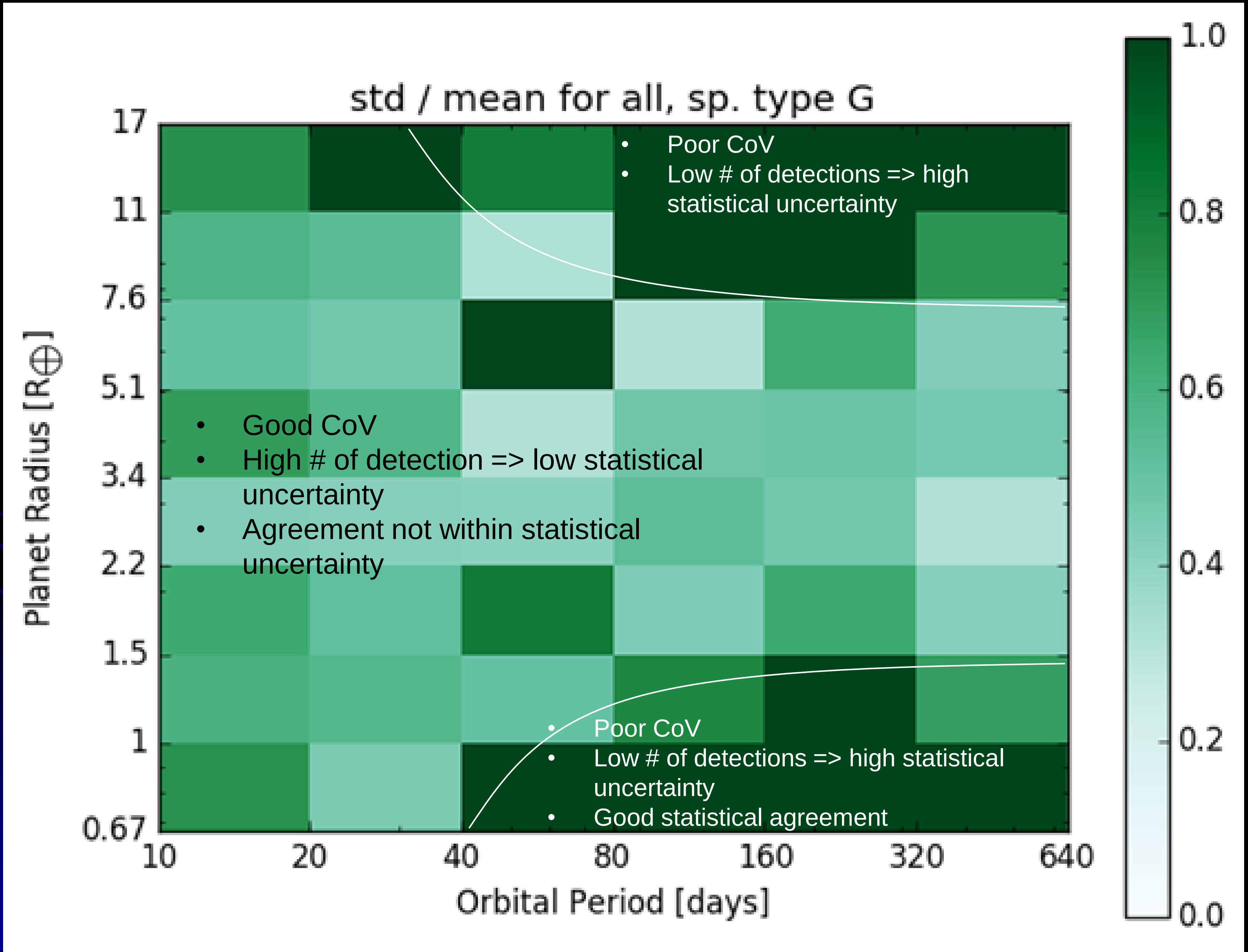
(based on best power law fit)

Note: this is a simple average across submissions
More sophisticated combination methods are being explored, such as weighting by quoted uncertainties and/or accounting for dependencies

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

Coefficient of Variation

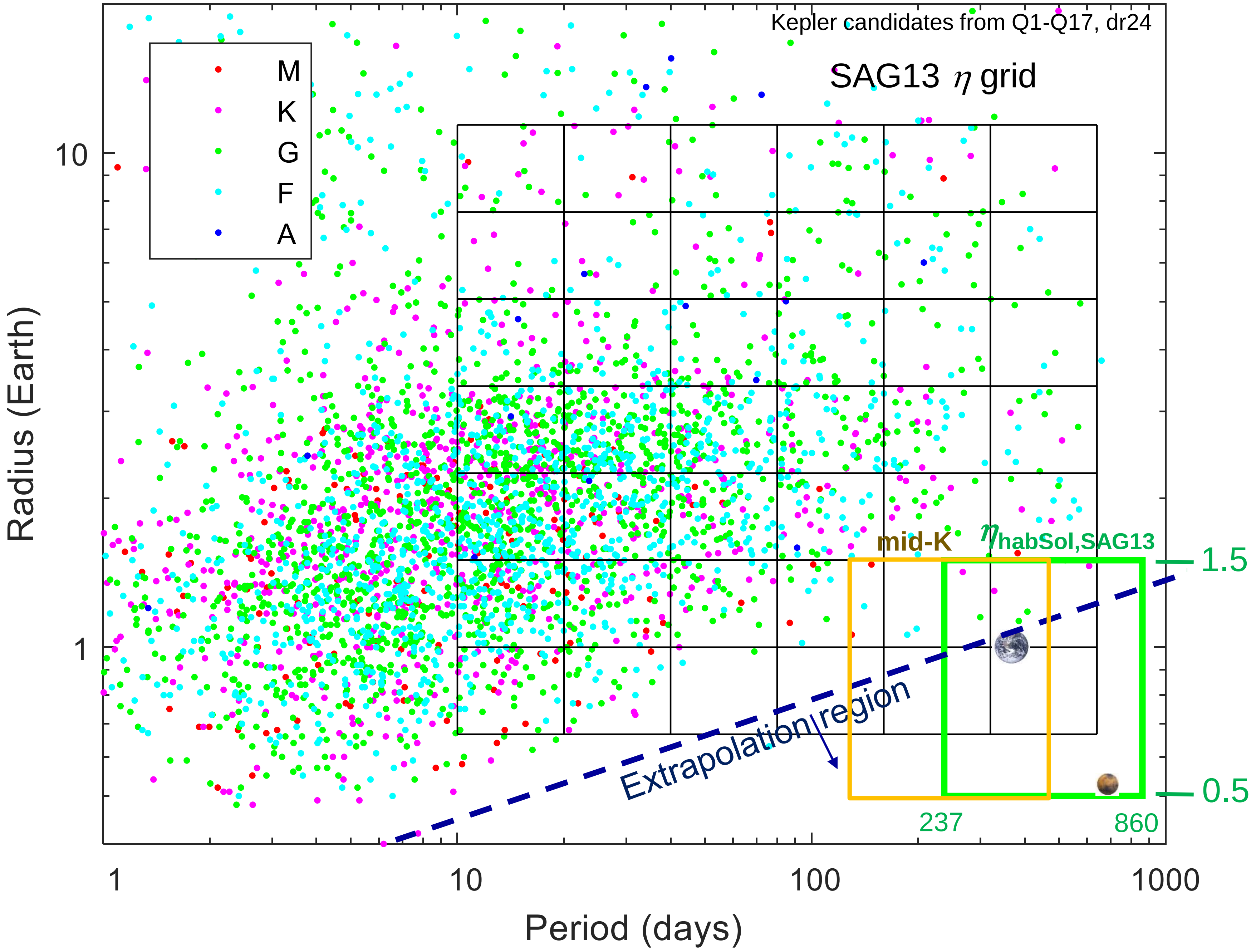
(aka relative standard deviation = std / mean)



Sensitivity of occurrence rates to methodologies and assumptions

- Completeness curves and catalog seem to make the largest systematic differences
  - More recent completeness curves and catalogs seem to lead to systematically higher numbers
- Other things (estimation method, details of the code, extrapolation) usually result in occurrence rates that are consistent to better than a factor of 2, usually much better

Extrapolation and importance of 0.5-1.0 R_{Earth} bin



Parametric fit (for G-dwarfs)

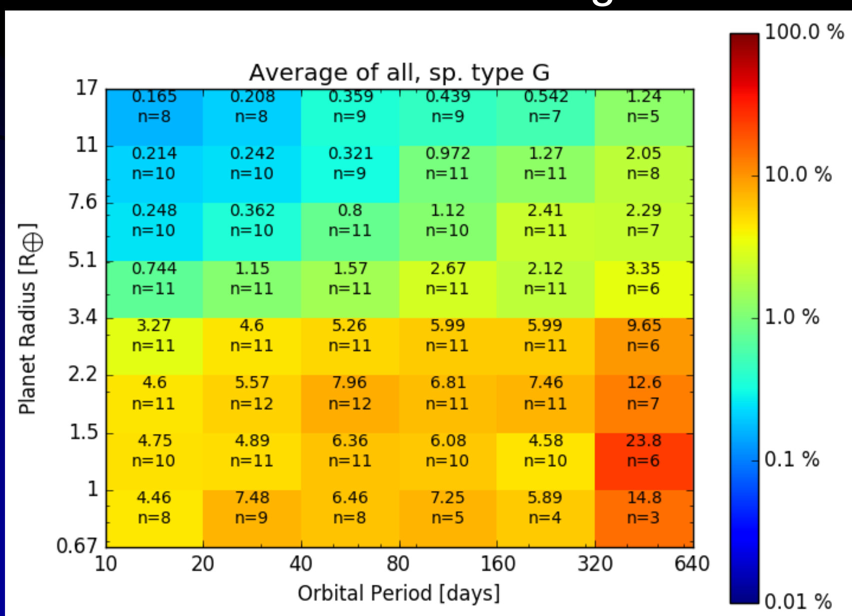
$$\frac{\partial^2 N(R,P)}{\partial \ln R \, \partial \ln P} = \Gamma_i R^{\alpha_i} P^{\beta_i} \qquad \text{in region } R_{i-1} \leq R < R_i$$

(R in Earth radius, P in years)

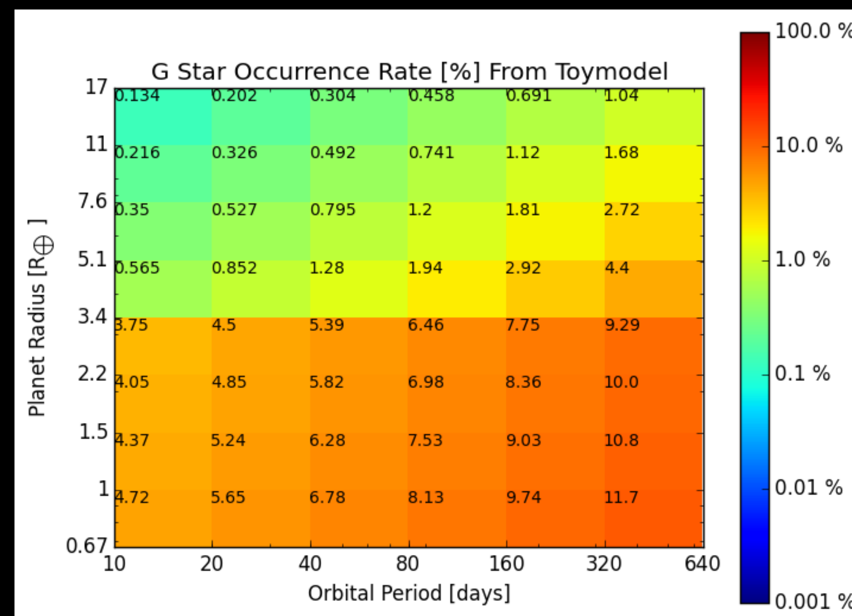
Two-piece broken power law →

i	Γ_i	α_i	β_i	R_i
1	**0.38**	**-0.19**	**0.26**	**3.4**
2	0.73	-1.18	0.59	Inf

Submission average



Parameteric fit (integrated across bins)



Calculations of habitable occurrence rates

Integrating SAG13 parametric fit

		HZ (from Kopparapu 2013)	
		Conservative	Optimistic
Planet radius range	1.0-1.5	**0.14**	**0.20**
	0.5-1.5	**0.40**	**0.58**

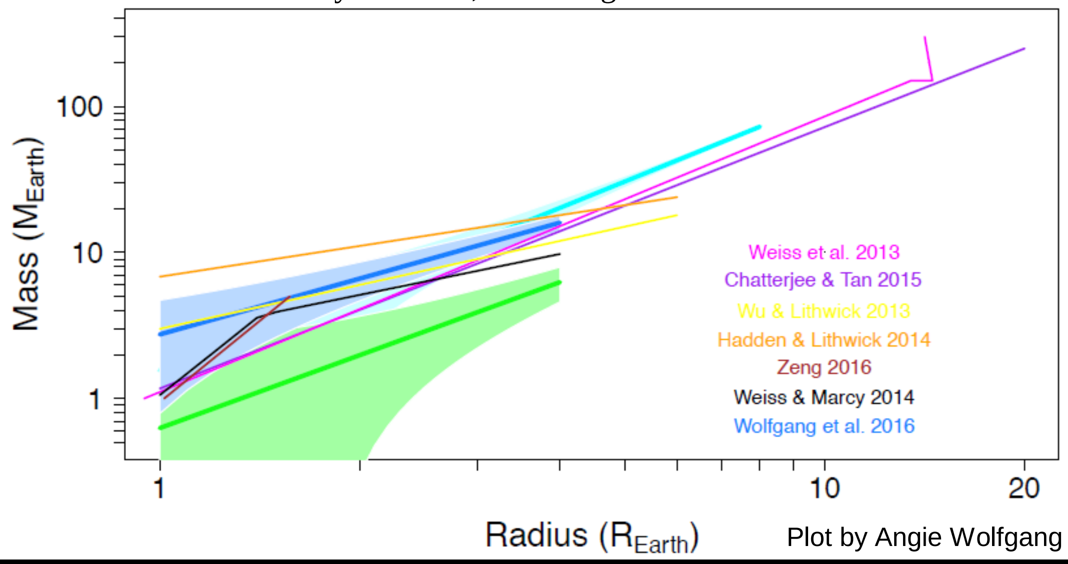
$\eta_{\text{habSol,SAG13}}$

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

		HZ (from Kopparapu 2013)	
		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	$\mathbf{0.5^{+0.4}_{-0.2}}$	$\mathbf{0.73^{+0.6}_{-0.3}}$

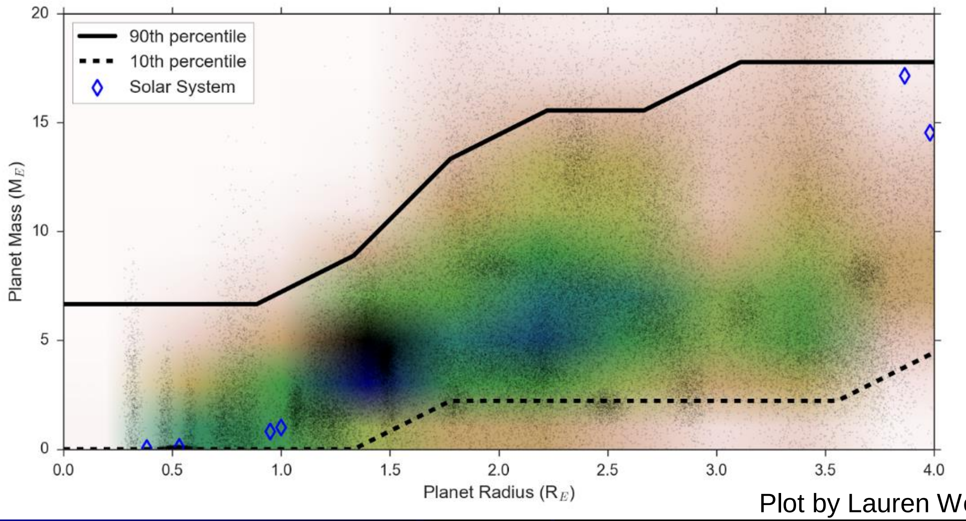
Converting between Mass and Radius
(focus group led by Angie Wolfgang and Lauren Weiss)

'Previous M-R relations in the literature:
wide variety of radius, mass ranges and datasets used



Plot by Angie Wolfgang

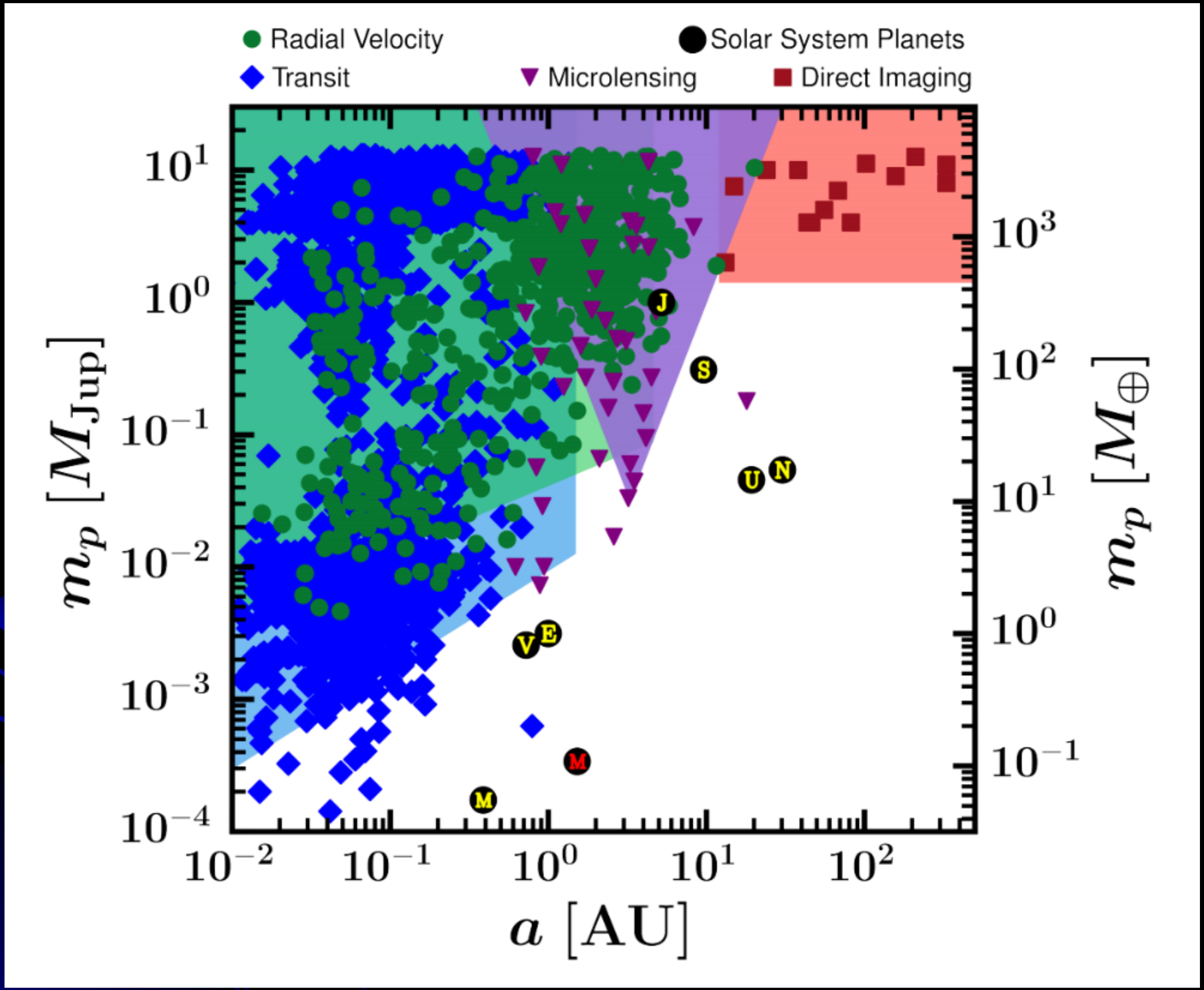
Preliminary estimate of M-R correlation



Plot by 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

- η_{Earth} may be significantly higher than many older estimates, especially if going down to 0.5 Earth size ($\eta_{habSol, SAG13}$ ~ 0.6)
  - Cannot yet rule out lower values (e.g. < ~0.3), but values >1 appear to be more likely than < 0.3
  - Burke 2015 η_{Earth} is even higher
  - Caveat: SAG13 products are not formal scientific results, but rather a meta-analysis to achieve consensus on “most likely” assumptions for mission studies. The upcoming Kepler closeout will yield a formal scientific result.

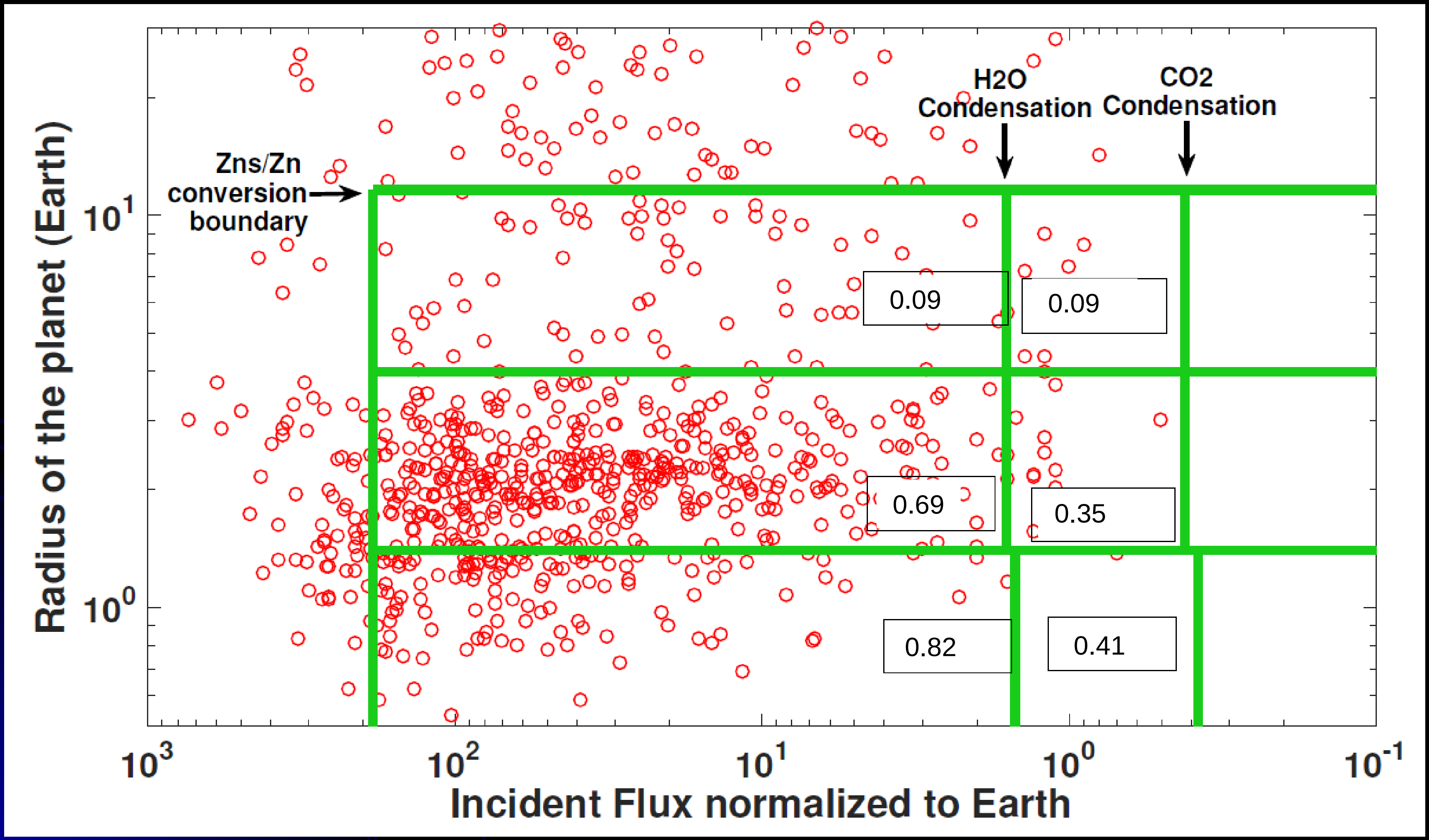
- Although many orders of magnitude of Γ_{Earth} (or η_{Earth}) are possible, only a small range (~ 1-2 octaves) within that is “likely”

- Tentative parametrized distributions are available from SAG13 to use with mission yield calculation codes (or any other purpose)
  - Based on input from the entire exoplanet community
  - See slide 14

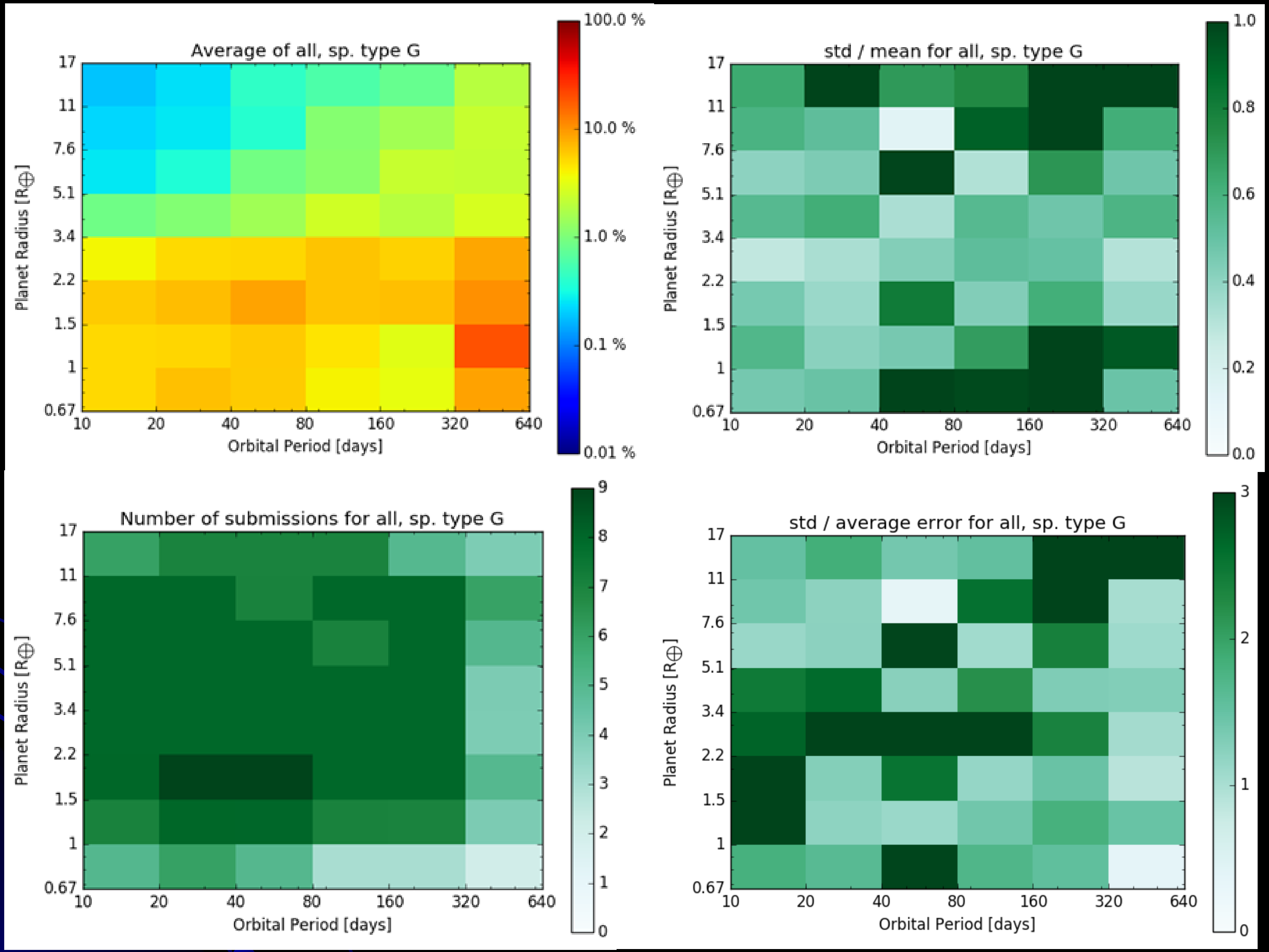
Backup slides

Occurrence rates for new proposed planet classification

(from Kopparapu, Domagal-Goldman, et al., in prep)
Numbers based on integrating SAG13 parametric fit

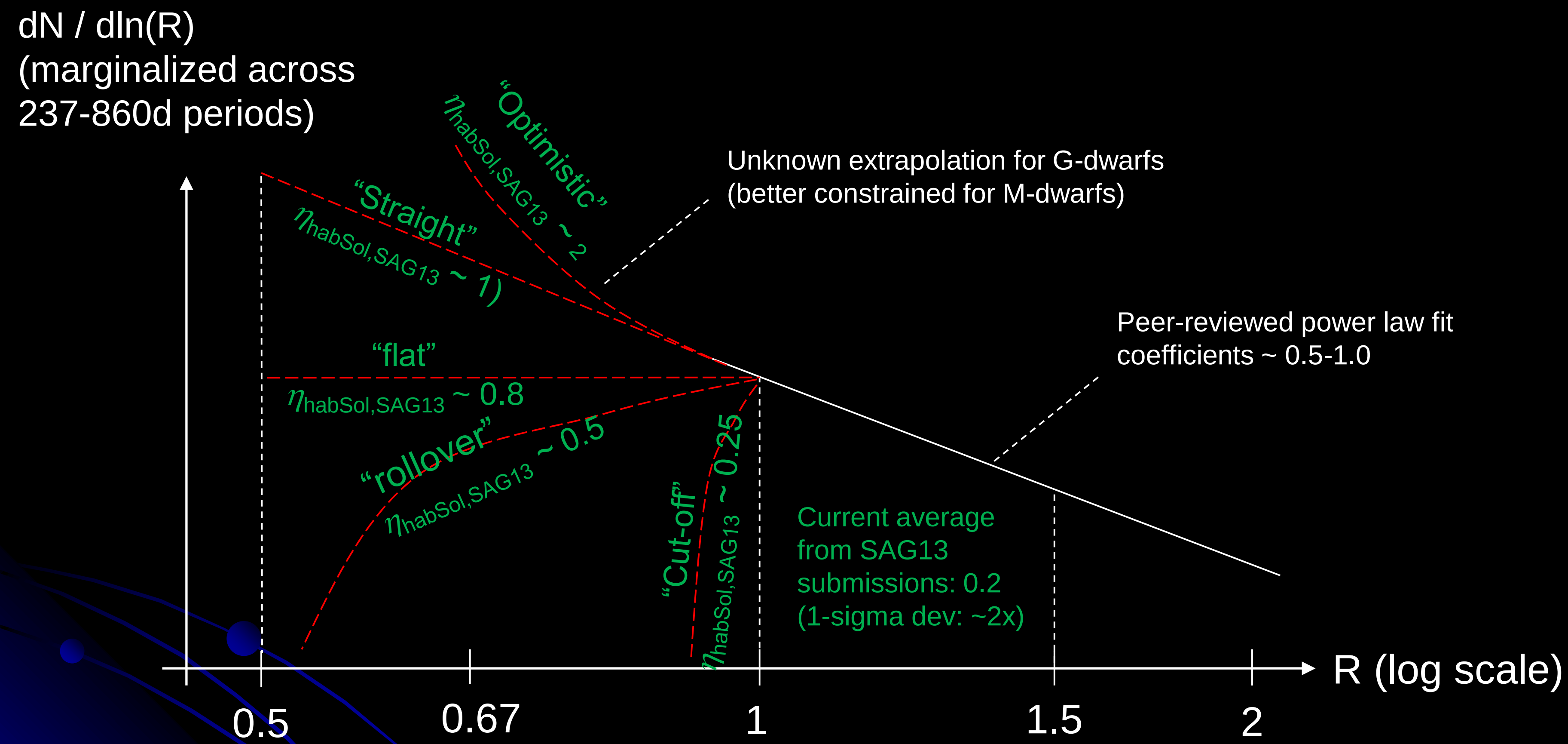


Analysis of variations in submissions (for G-dwarfs)



by Gijs Mulders

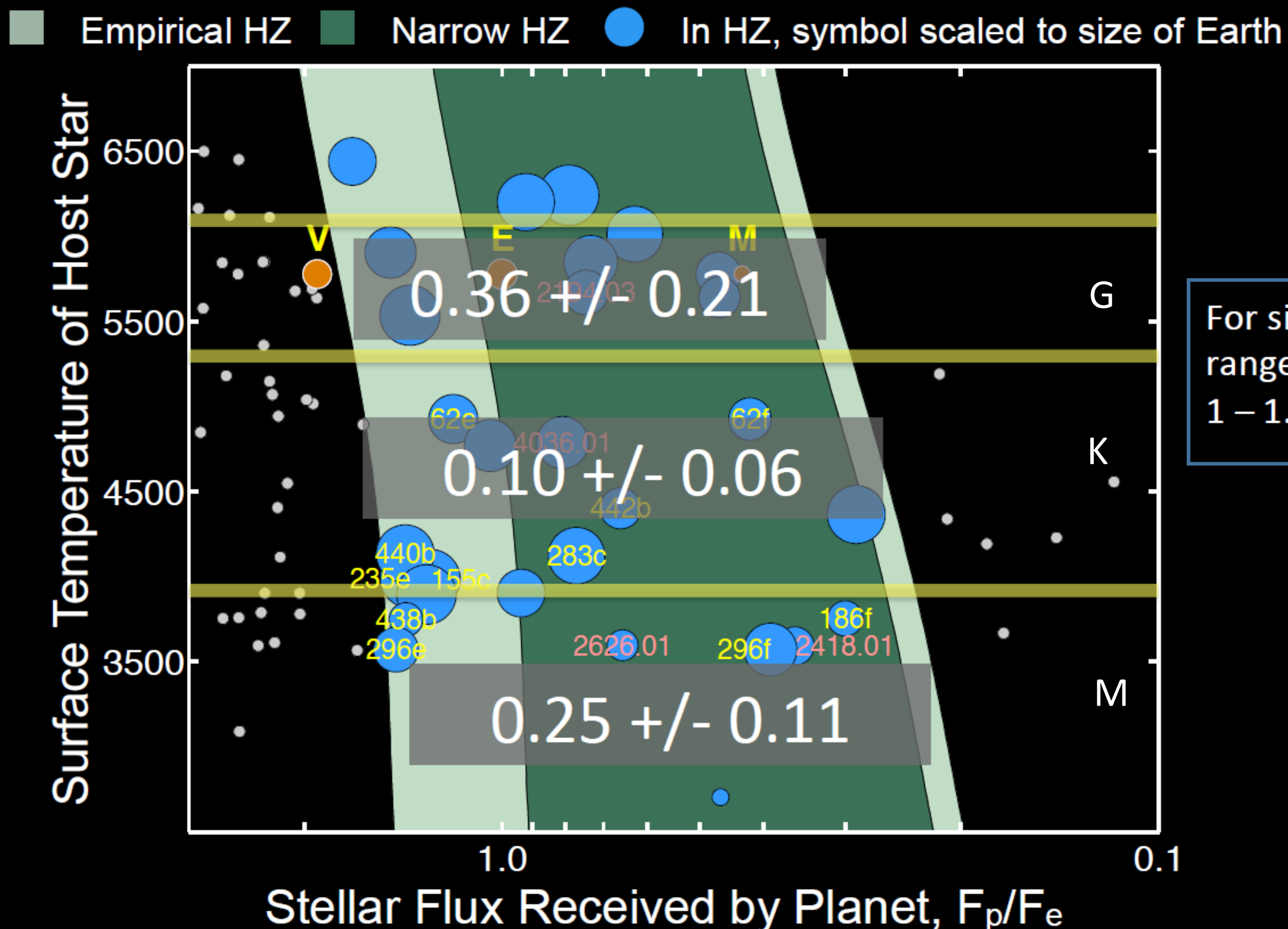
Importance of 0.5-1.0 Earth size bin



- 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

Courtesy of Natalie Batalha

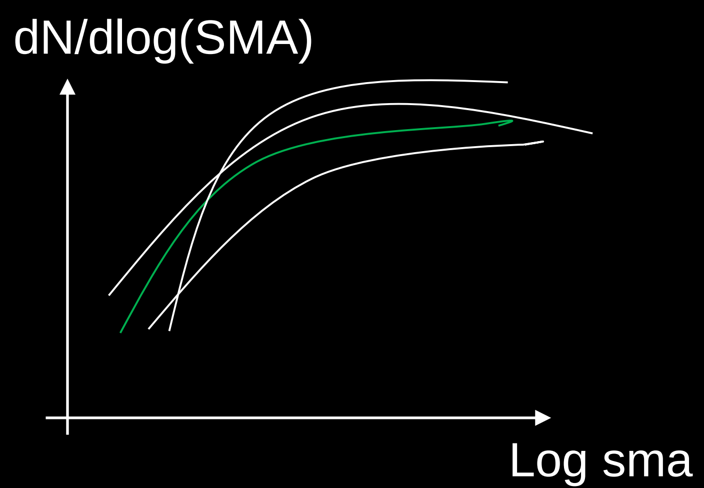
Small (< 2 R_e) Planets in the HZ: 4 yr

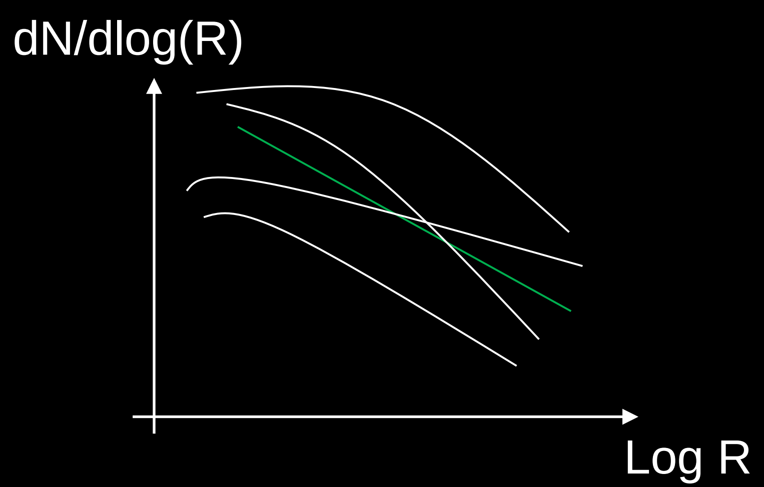


For size range: 1 – 1.6 R_e

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

Variances between individual parameterized distributions



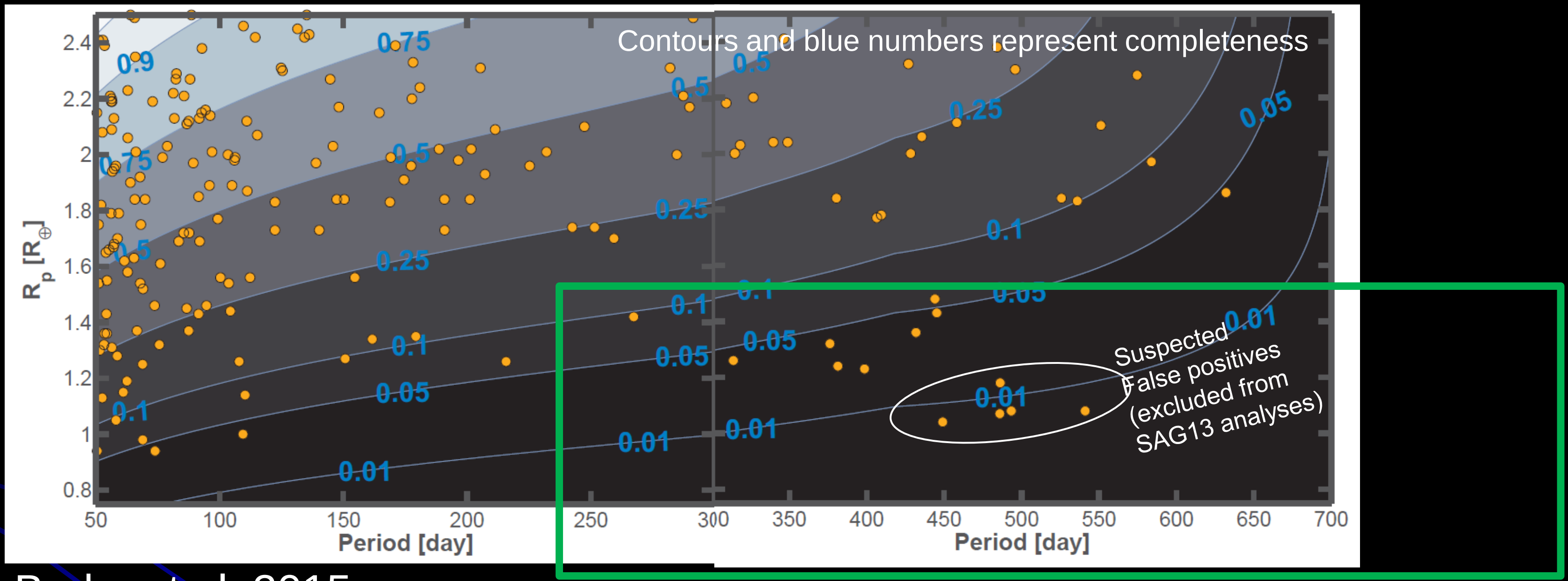


Current edge of planet candidates

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

Shorter periods, more reliable

Longer periods, less reliable

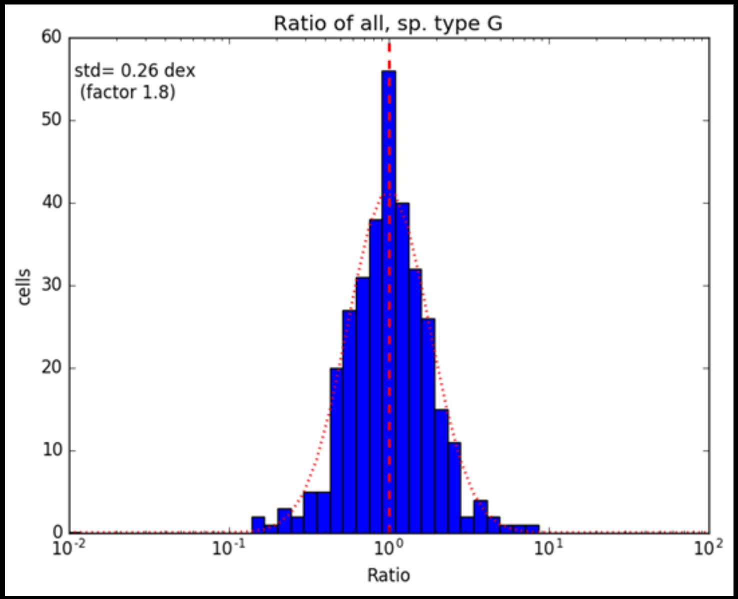


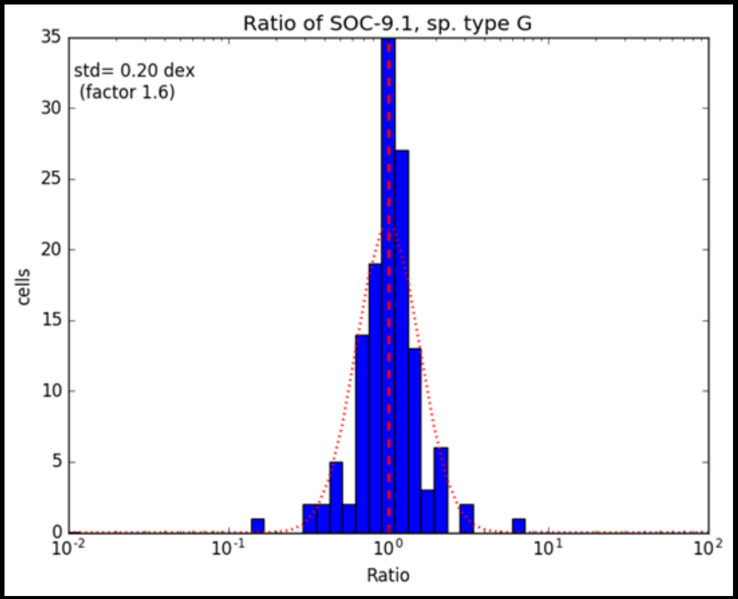
Burke et al. 2015

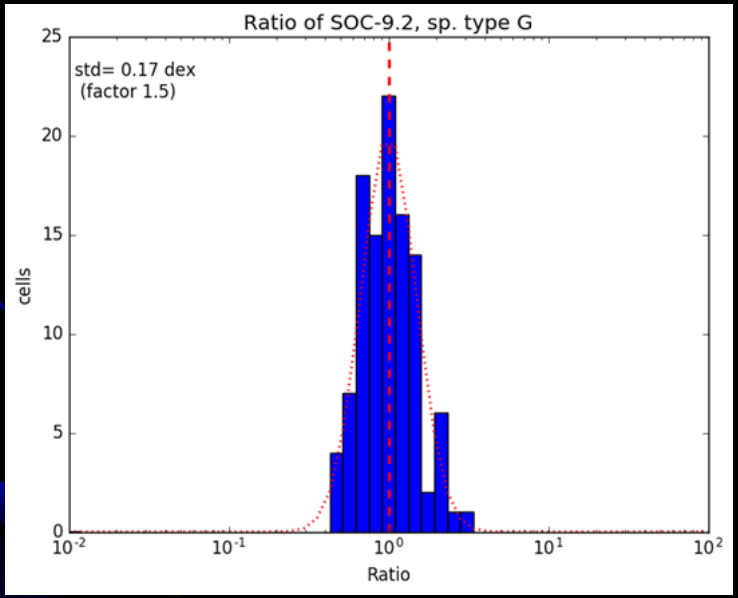
0.5-1.5 Earth size

237-860 days (Kopparapu extended HZ for Sun)

Variance in submissions





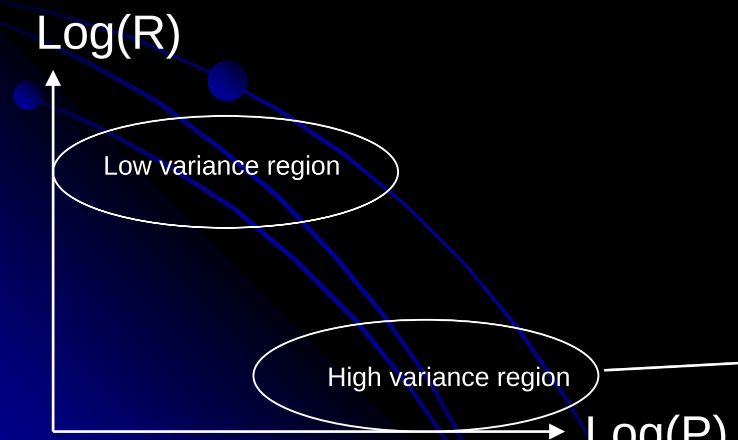


Courtesy of Gijs Mulders

Slide which shows any key correlations we found between variances / outliers and submission parameters (catalog, method, etc.)

[Goal is to show status and any key preliminary patterns we found in the most clear and concise way but emphasize that this is still a work in progress]

Rough idea for visualization:



Value of Γ_{earth}	catalog	Completeness ?	Methodology ?
Lowest value	Early catalog ?		
Highest value	More recent catalog ?		

[Note – for now, table entries are purely illustrative, not necessarily ones that we will have in the final slide]

ROUGH DRAFT

Details of submitted rates

	Catalog	Filters	Completeness model	Vetting efficiency	Reliability	Methodology	Value of Γ_{earth}
Batalha, Natalie (2)							
Belikov, Rus							
Burke, Chris							
Catanzarite, Joe							
Dressing, Courtney*							
Farr, Will							
Foreman-Mackey, Daniel*							
Kopparapu, Ravi							
Mulders, Gijs							
Petigura, Erik*							
Traub, Wes**							

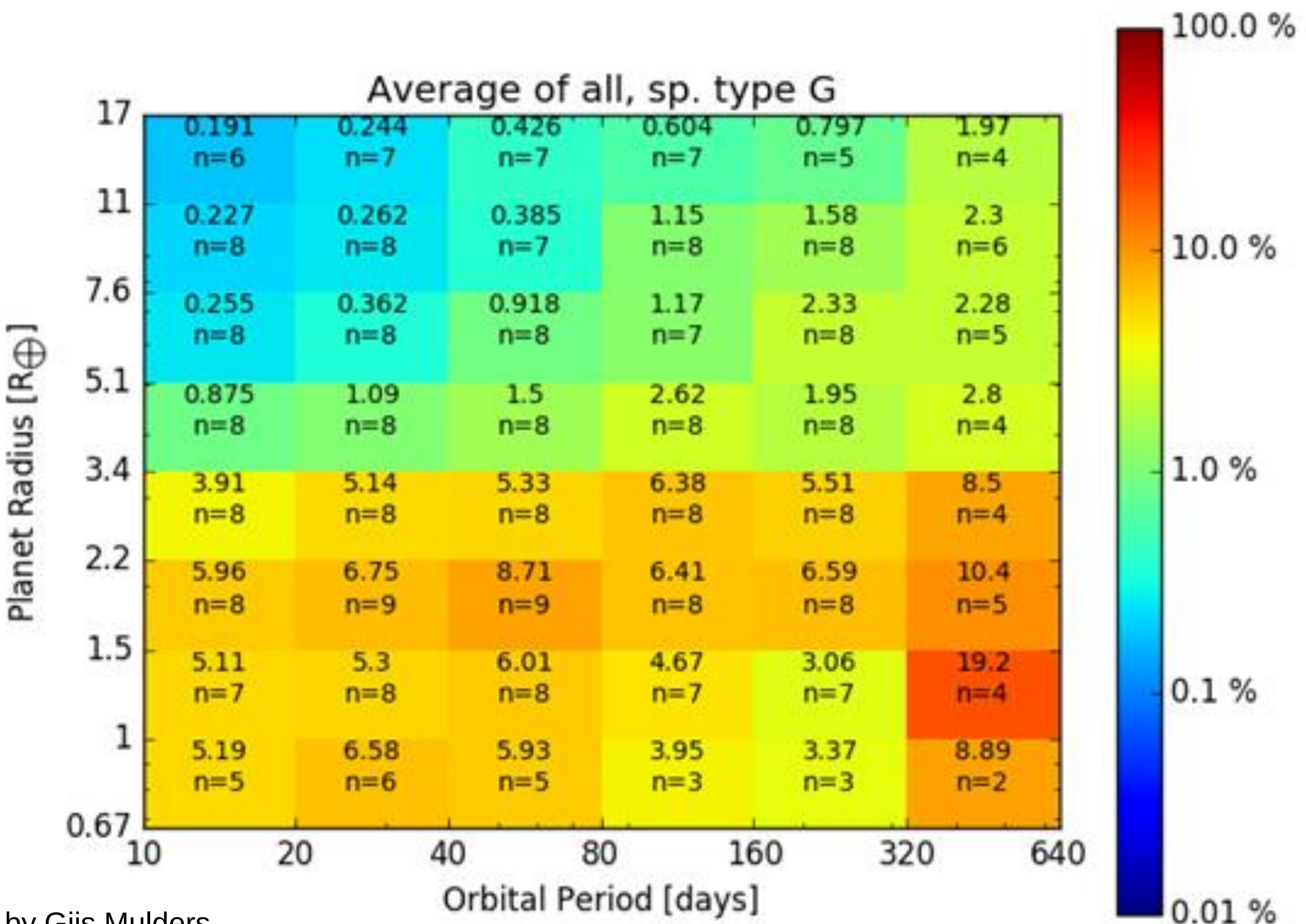
ROUGH DRAFT / SLIDE IDEA

Closer look at G-dwarf average

legend

% occurrence

# of submissions



by Gijs Mulders

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