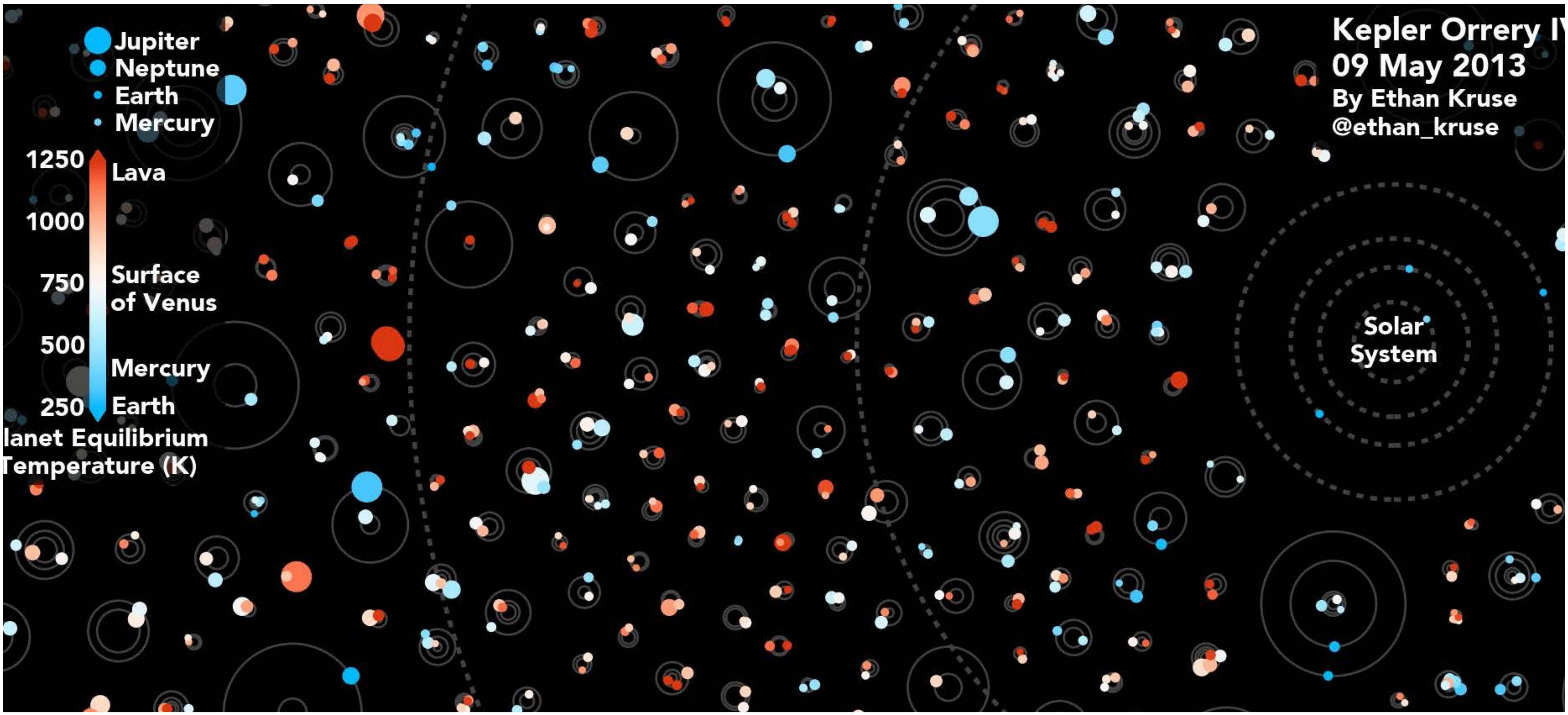
Hierarchical Bayesian Modeling of Planet Populations



Angie Wolfgang
NSF Postdoctoral Fellow, Penn State

The Big Picture

We've found > 3000 planets, and counting.

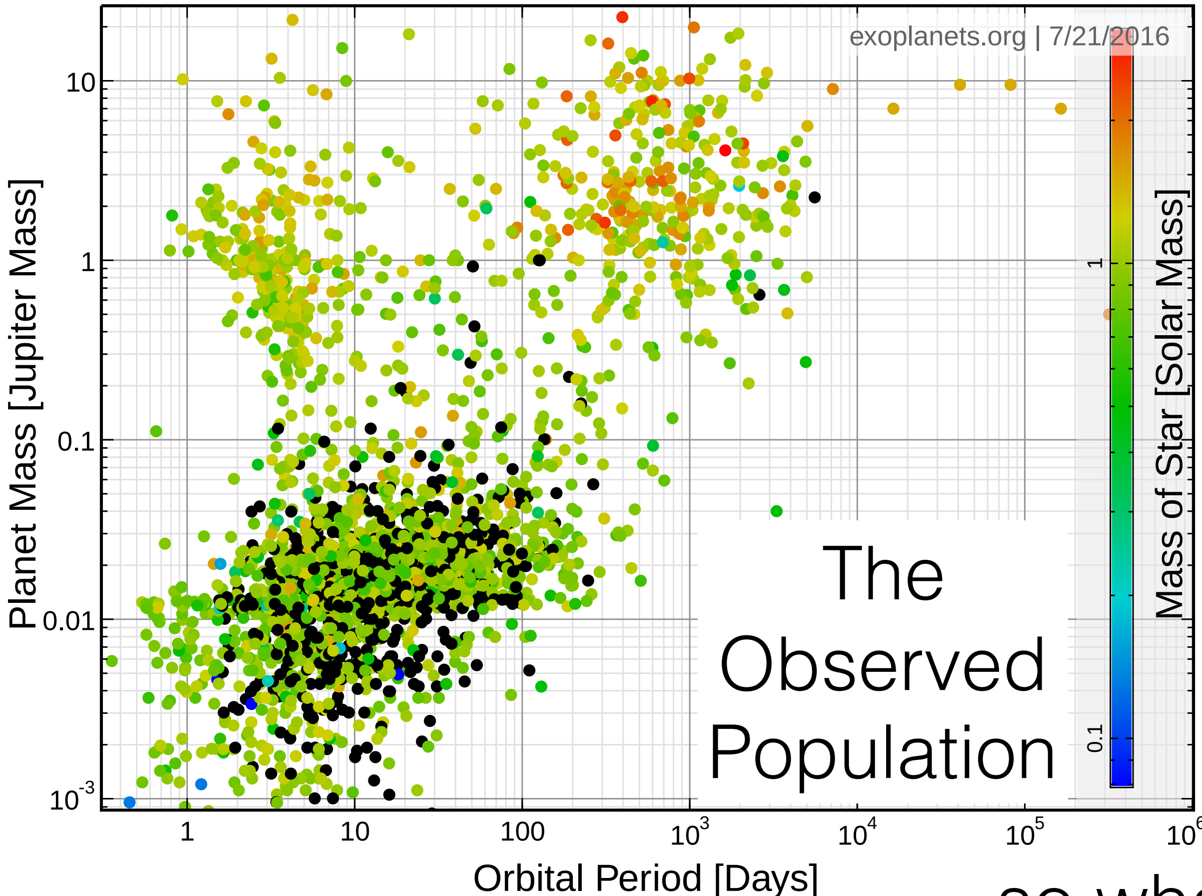
Earth's place in the Universe . . .

We're measuring their properties. (Mass, radius, atmospheres, orbits, host stars, …)

Habitable exoplanets?

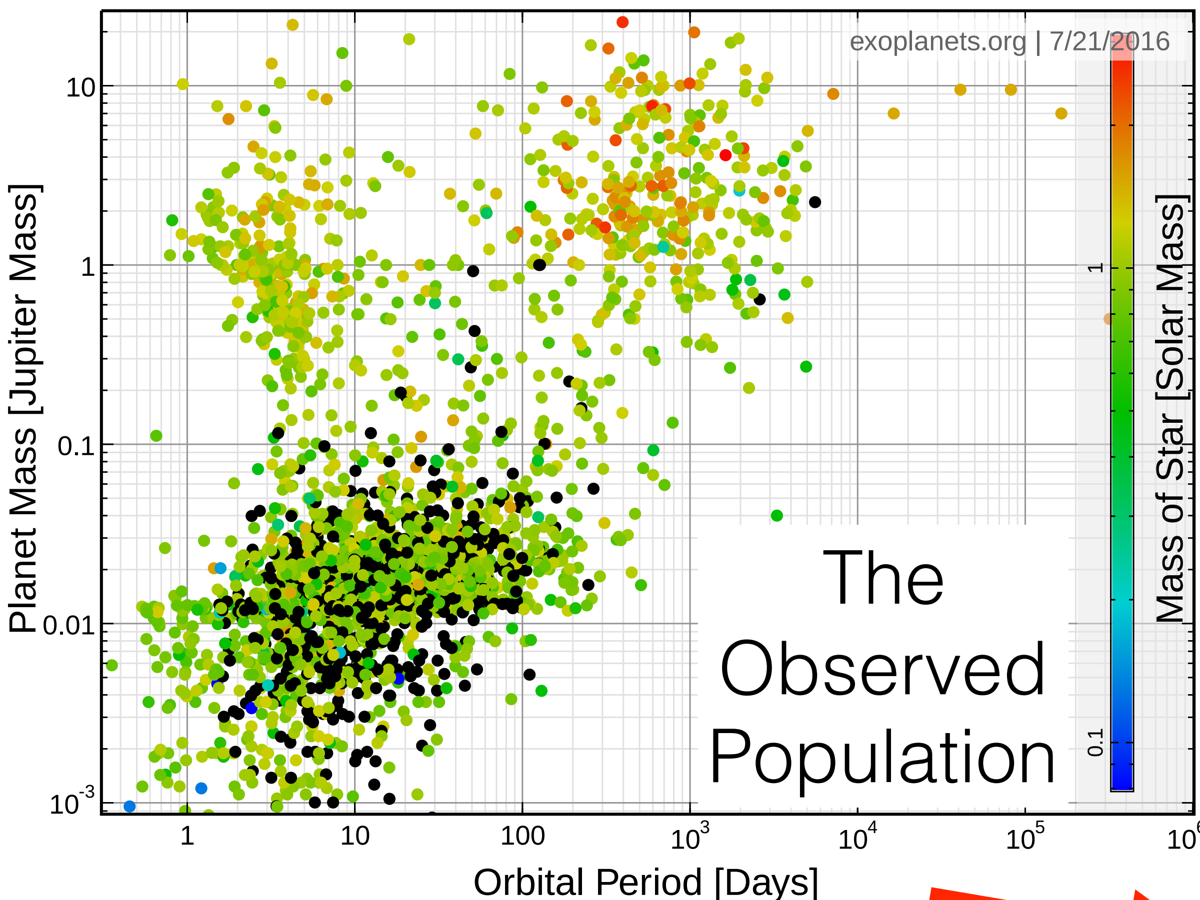
. . . so what's the physics involved? (How do they form, interact with their surroundings, change via which physical processes?)

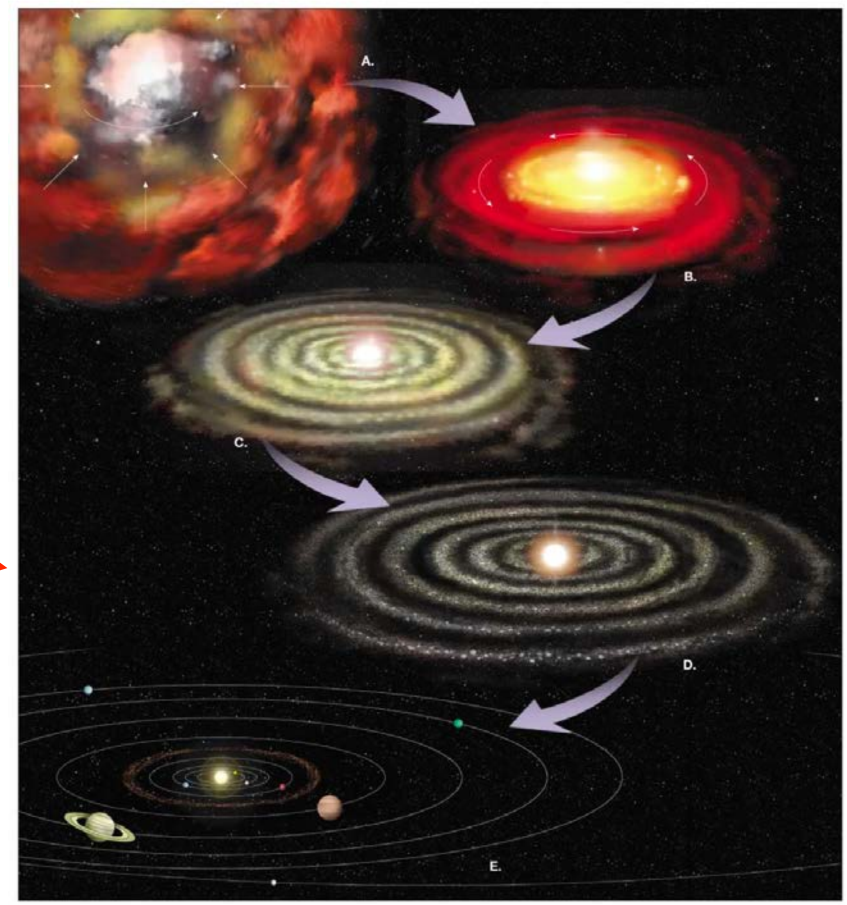
The Big Picture



. . . so what's the physics involved?
(How do they form, interact with their surroundings, change via which physical processes?)

The Big Picture

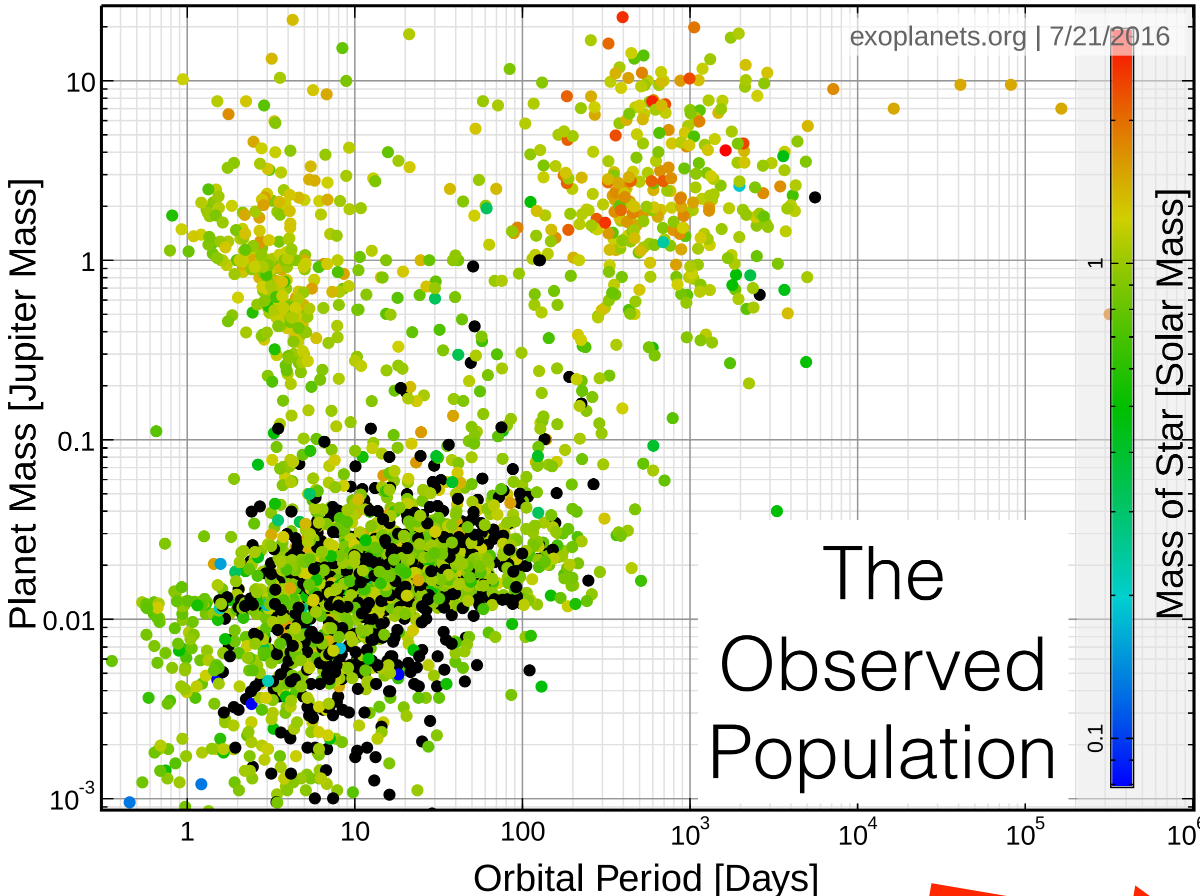




Deterministic planet formation model with physical parameters $\boldsymbol{\alpha}$ (disk mass, viscosity…):

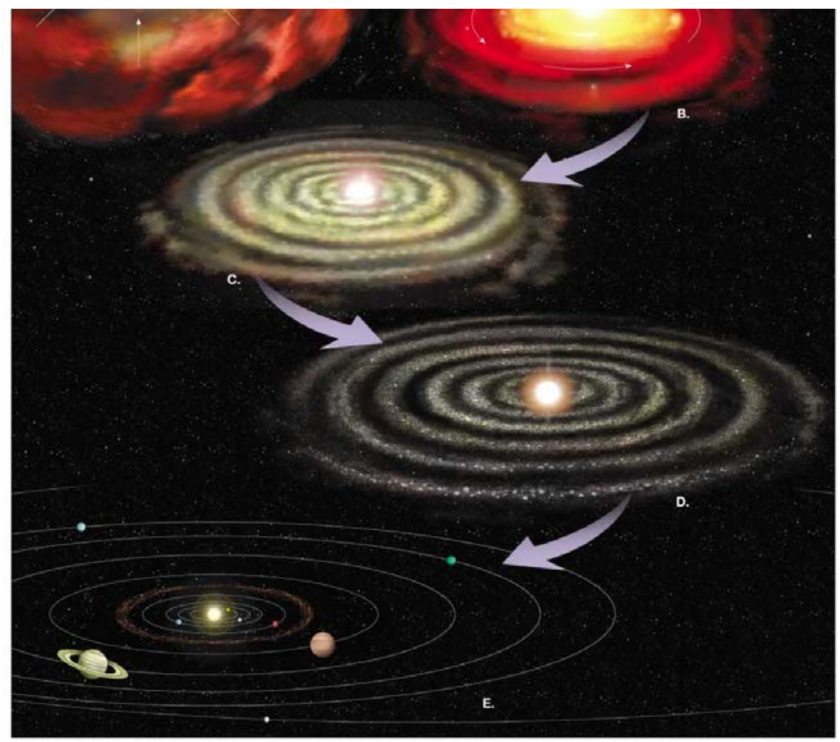
$f(M,P,\ldots|\boldsymbol{\alpha})$

The Big Picture



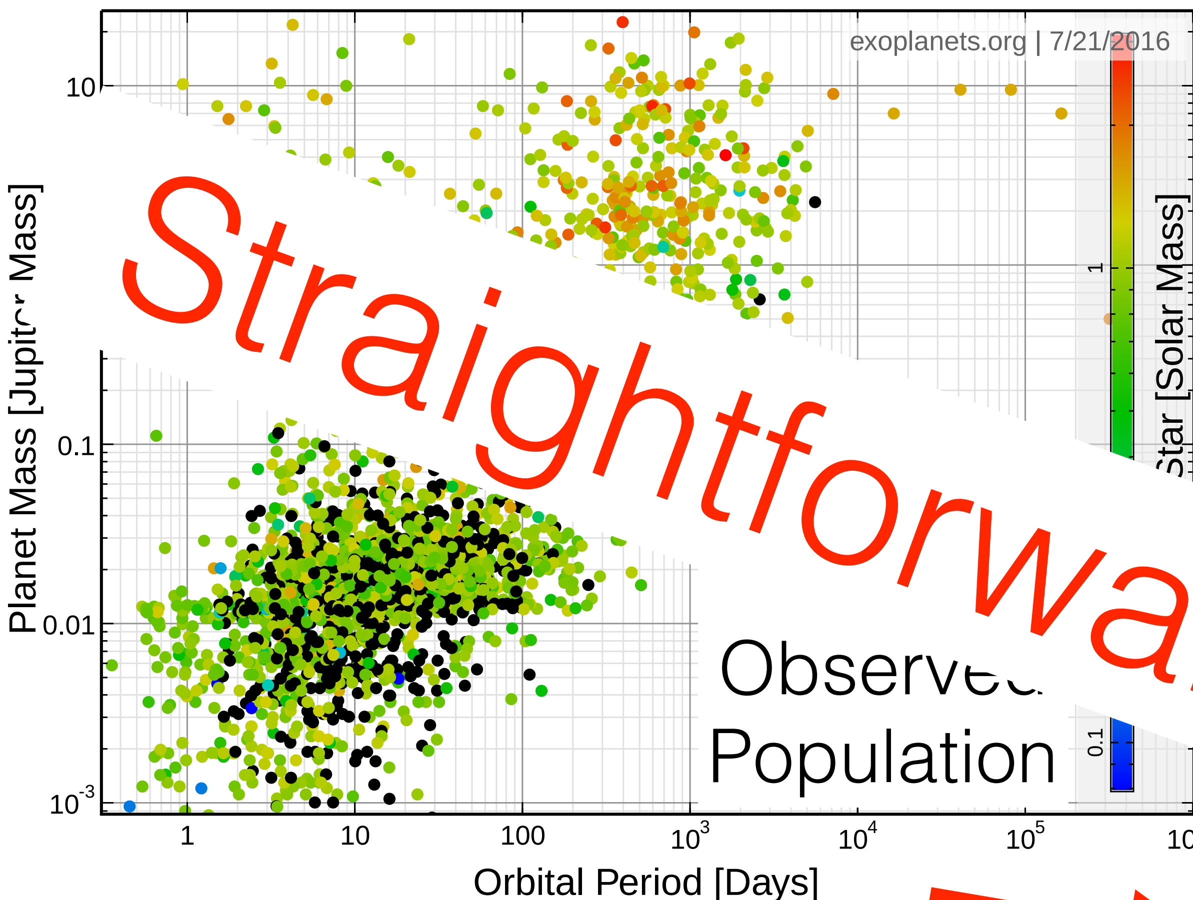
From the outset:

1) distributions of planet properties,
2) inference on $\boldsymbol{\alpha}$,
3) model comparison (f_1 vs f_2 vs f_3).



Deterministic planet formation model with physical parameters $\boldsymbol{\alpha}$ (disk mass, viscosity…): $f(M,P,\ldots|\boldsymbol{\alpha})$

The Big Picture



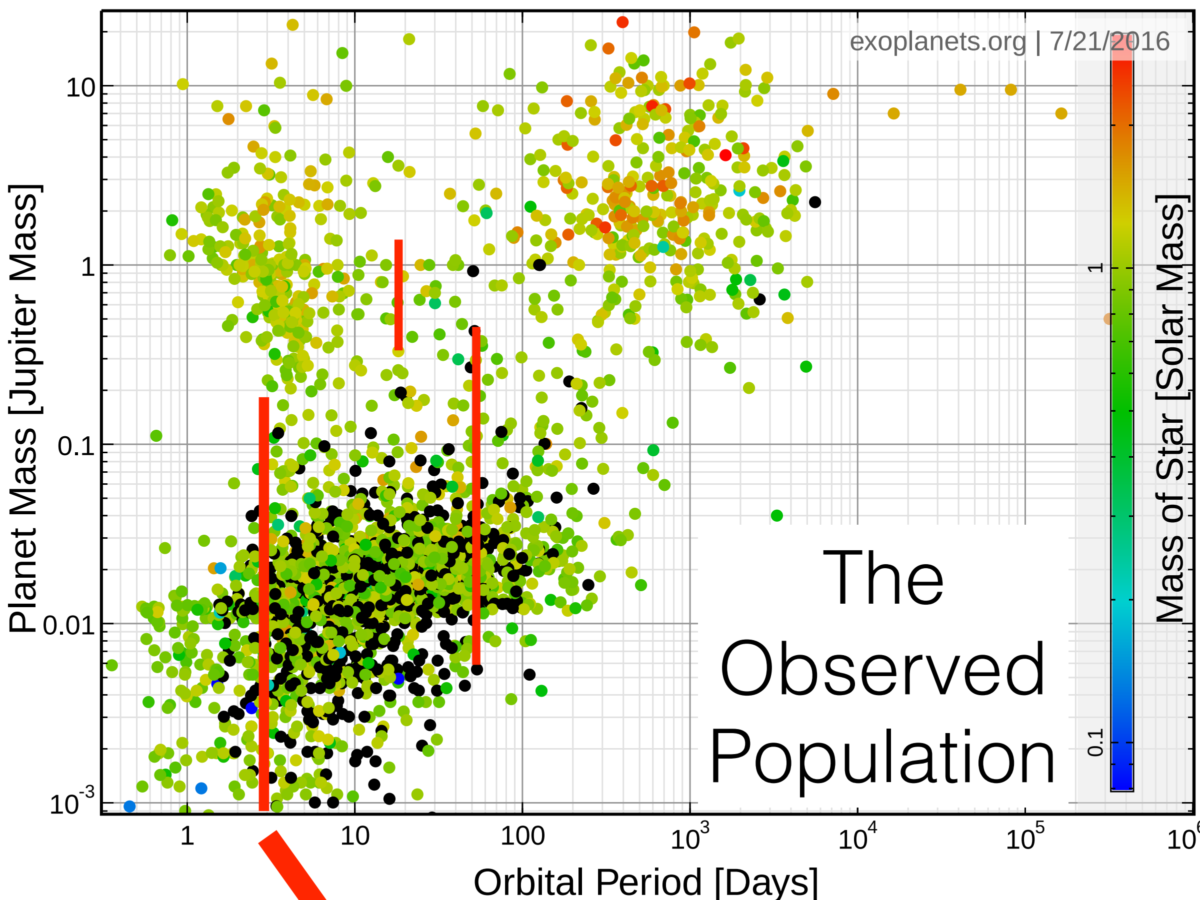
From the outset:

1) distributions of planet properties,
2) inference on **α**,
3) model comparison (f_1 vs f_2 vs f_3).

Straightforward, right?

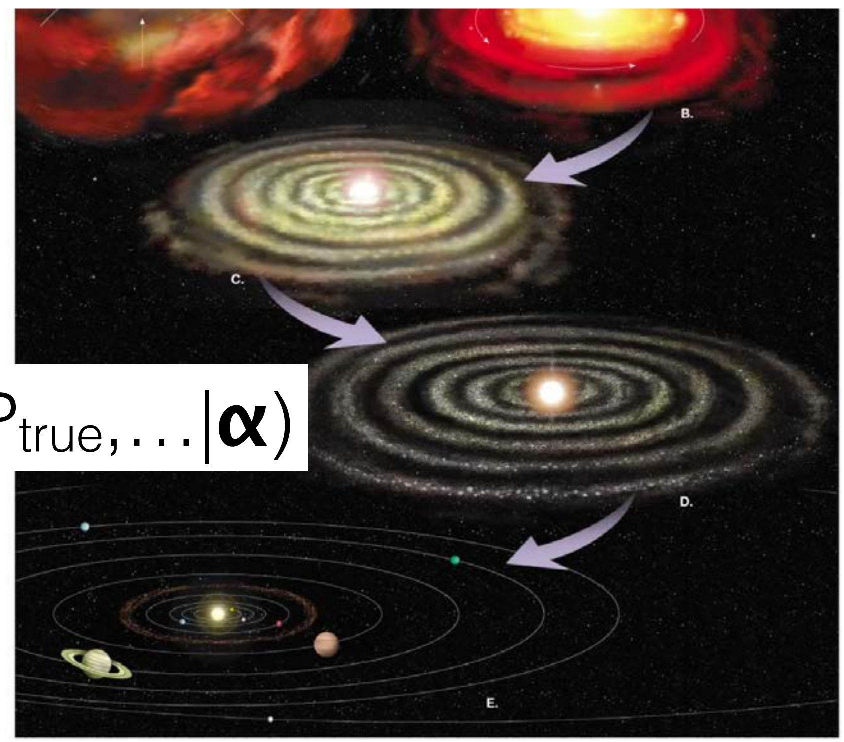
Deterministic planet formation model with physical parameters **α** (disk mass, viscosity…): f(M,P,…|**α**)

Large Measurement Uncertainty



From the outset:

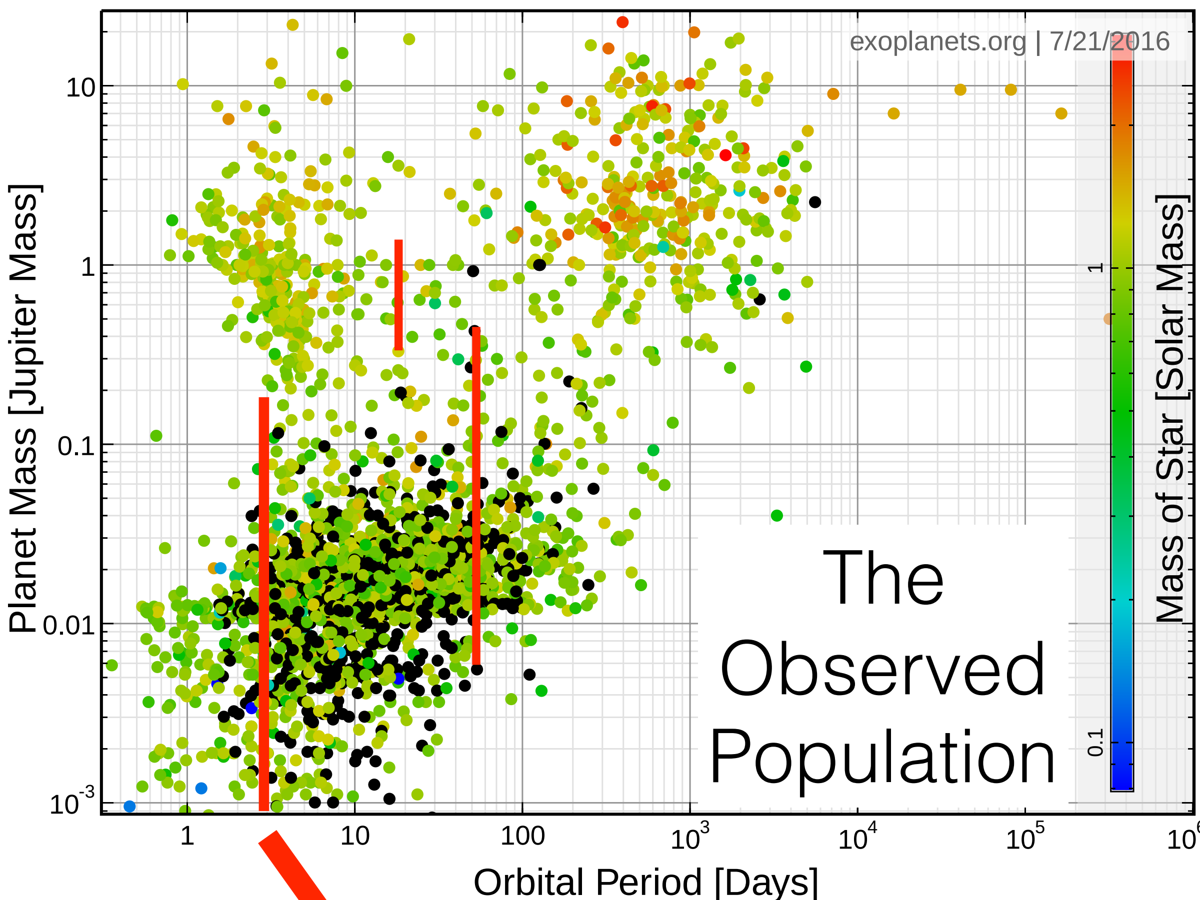
1) distributions of planet properties,
2) inference on $\boldsymbol{\alpha}$,
3) model comparison (f_1 vs f_2 vs f_3).



M_{true}, P_{true}, …

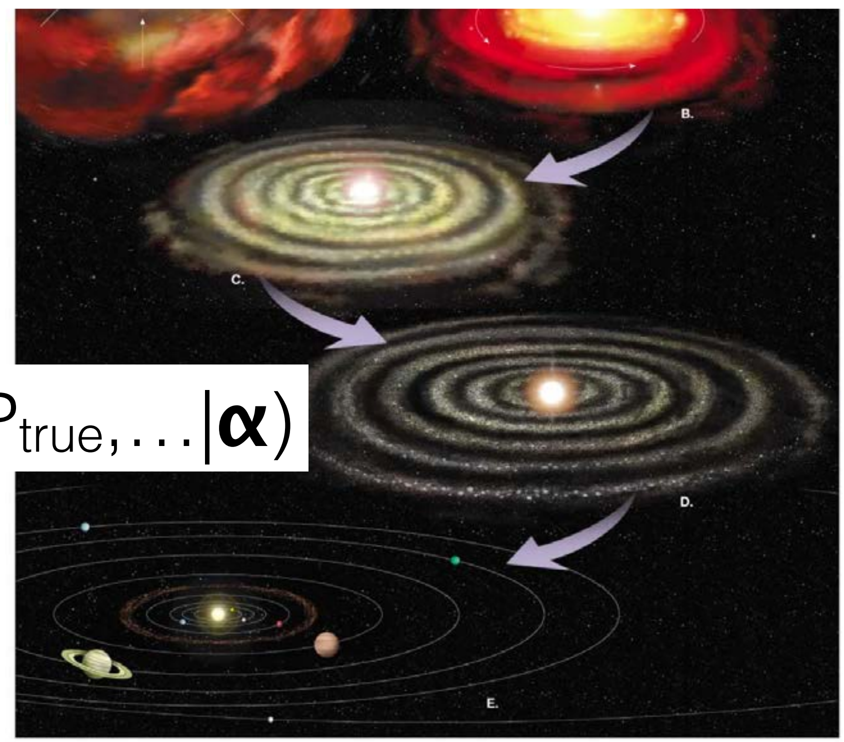
$f(M_{true}, P_{true}, \ldots | \boldsymbol{\alpha})$

Large Measurement Uncertainty



From the outset:

1) distributions of planet properties,
2) inference on **α**,
3) model comparison (f_1 vs f_2 vs f_3).

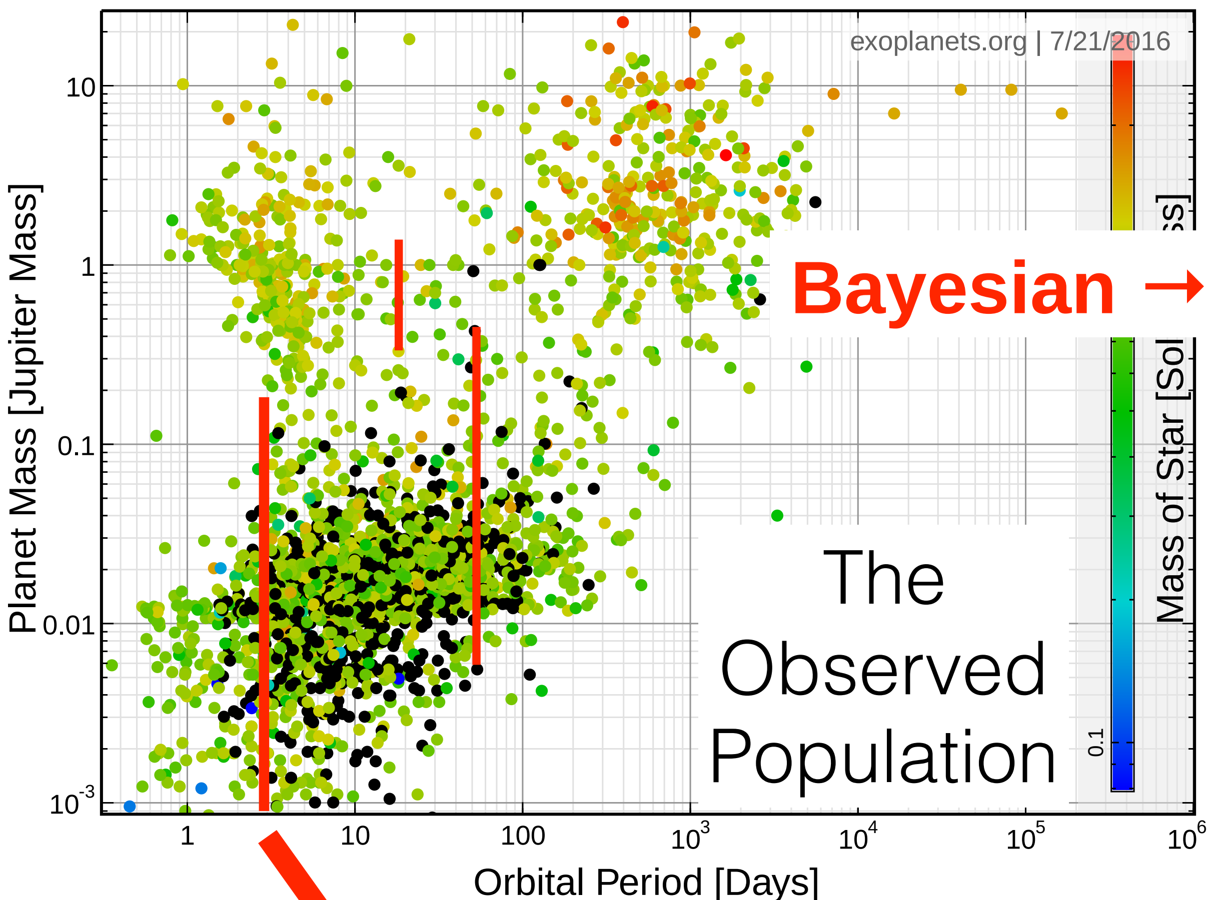


M_{true}, P_{true}, …

$f(M_{true}, P_{true}, \ldots | \boldsymbol{\alpha})$

Multiple levels → **hierarchical modeling**

Large Measurement Uncertainty



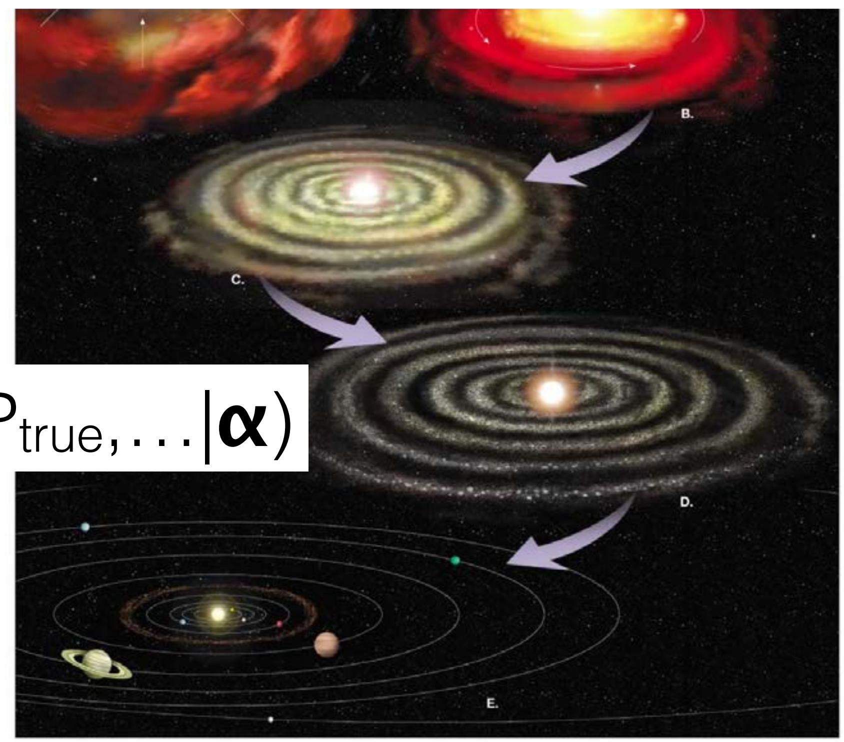
The Observed Population

From the outset:

1) distributions of planet properties,

Bayesian → 2) inference on **α**,

3) model comparison (f_1 vs f_2 vs f_3).



M_{true}, P_{true}, …

$f(M_{true}, P_{true}, \ldots | \boldsymbol{\alpha})$

Multiple levels → **hierarchical modeling**

What is Hierarchical Bayesian Modeling (HBM)?

Arises naturally when want to make **scientific inferences** about a **population** based on many **individuals**.

"Regular" Bayes:

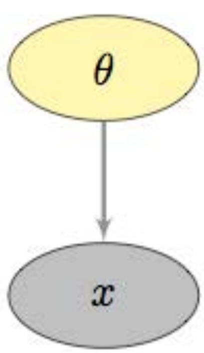
$$p(\theta|x) \propto p(x|\theta)\, p(\theta)$$

posterior likelihood prior

"What is the probability that θ has some value, given the data?"

Parameters

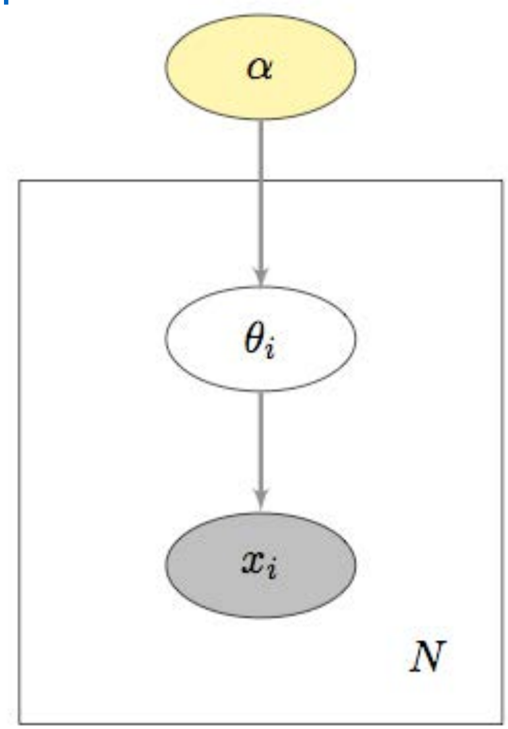
Observables



Hierarchical Bayes:

$$p(\alpha,\theta|x) \propto p(x|\theta,\alpha)\, p(\theta|\alpha)\, p(\alpha)$$

posterior likelihood prior



Population Parameters

Individual Parameters

M_{true}, P_{true}, …

Observables

M_{obs}, P_{obs}, …

HBM for Exoplanets

Hogg et al. 2010
(orbital eccentricities)

Morton & Winn, 2014
Campante et al. 2016
(angle between stellar spin & planet's orbit)

Foreman-Mackey et al, 2014
(Kepler occurrence rates)

Demory 2014
(geometric albedos)

Rogers 2015
(rocky-gaseous transition)

Wolfgang & Lopez, 2015
(super-Earth composition distribution)

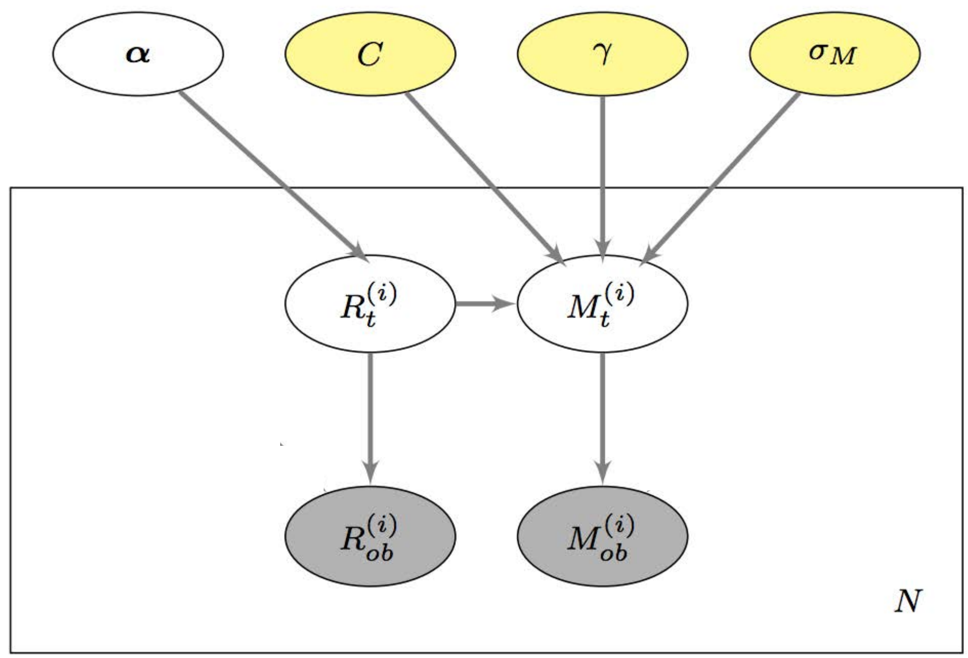
Shabram et al. 2016
(short-period eccentricity distribution)

Wolfgang, Rogers, & Ford, 2016
Chen & Kipping, submitted
(mass-radius relationship)

2013 SAMSI workshop on analyzing Kepler data

All are some variation on this 3-level structure:

(from Wolfgang et al. 2016)



HBM for Exoplanets

Hogg et al. 2010
(orbital eccentricities)

Morton & Winn, 2014
Campante et al. 2016
(angle between stellar spin & planet's orbit)

Foreman-Mackey et al, 2014
(Kepler occurrence rates)

Demory 2014
(geometric albedos)

Rogers 2015
(rocky-gaseous transition)

Wolfgang & Lopez, 2015
(super-Earth composition distribution)

Shabram et al. 2016
(short-period eccentricity distribution)

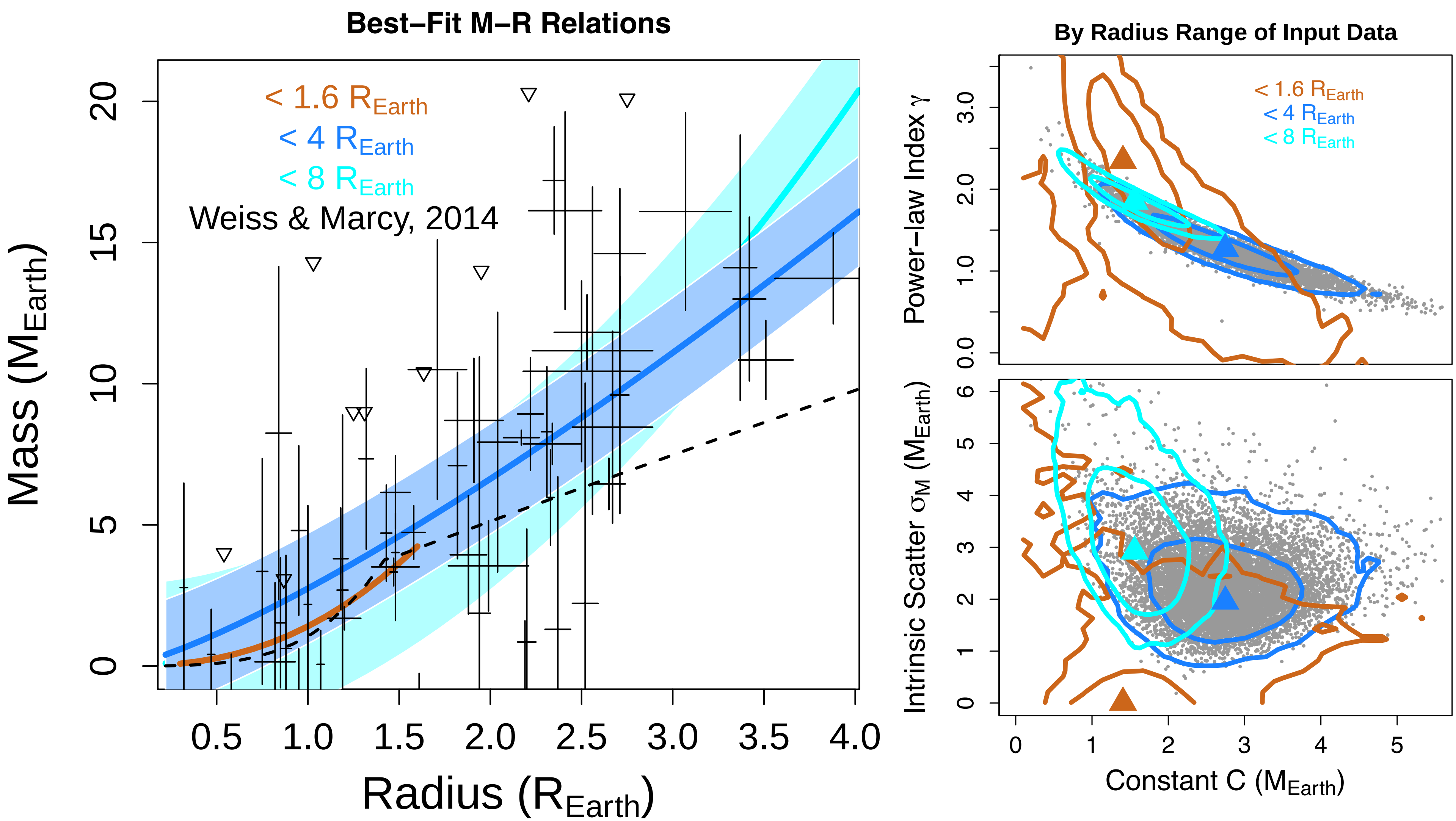
Wolfgang, Rogers, & Ford, 2016
Chen & Kipping, submitted
(mass-radius relationship)

Already have posteriors for the observables? Can use importance sampling in multi-level models (Hogg et al. 2010)

More recently: full HBM using JAGS or own hierarchical MCMC code; many people moving to STAN

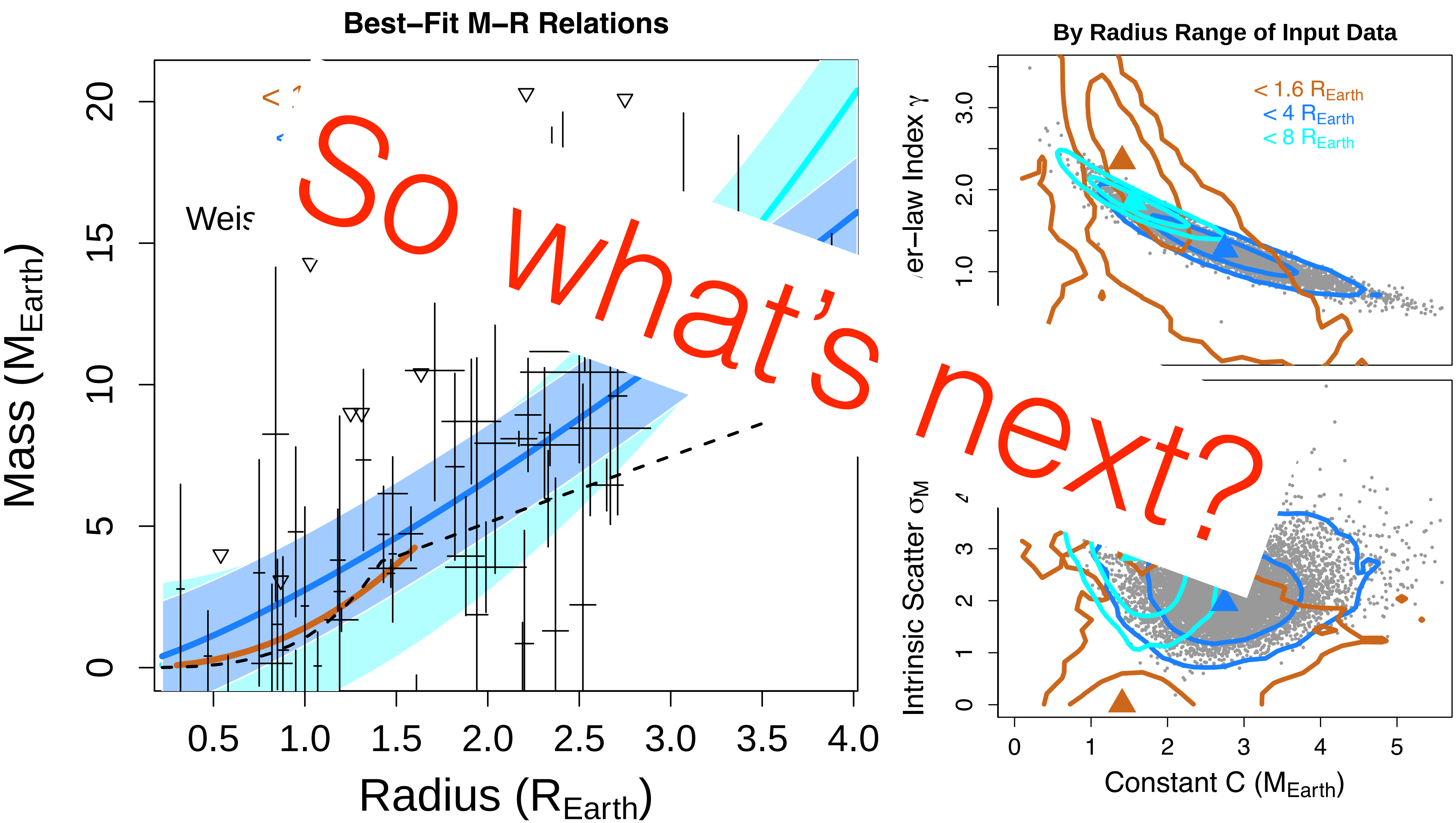
Example: Planet HBM Results

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016



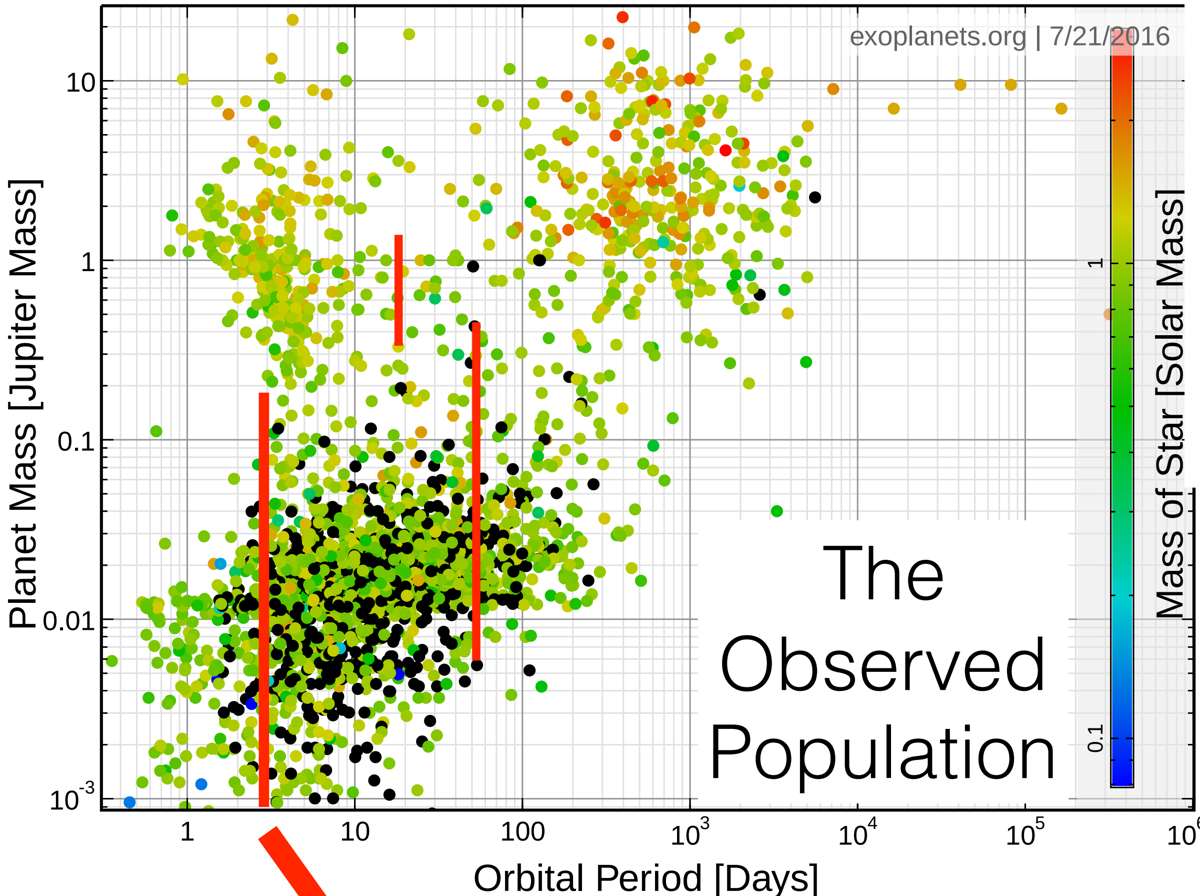
Example: Planet HBM Results

Mass-radius “relation”: Wolfgang, Rogers, & Ford, 2016

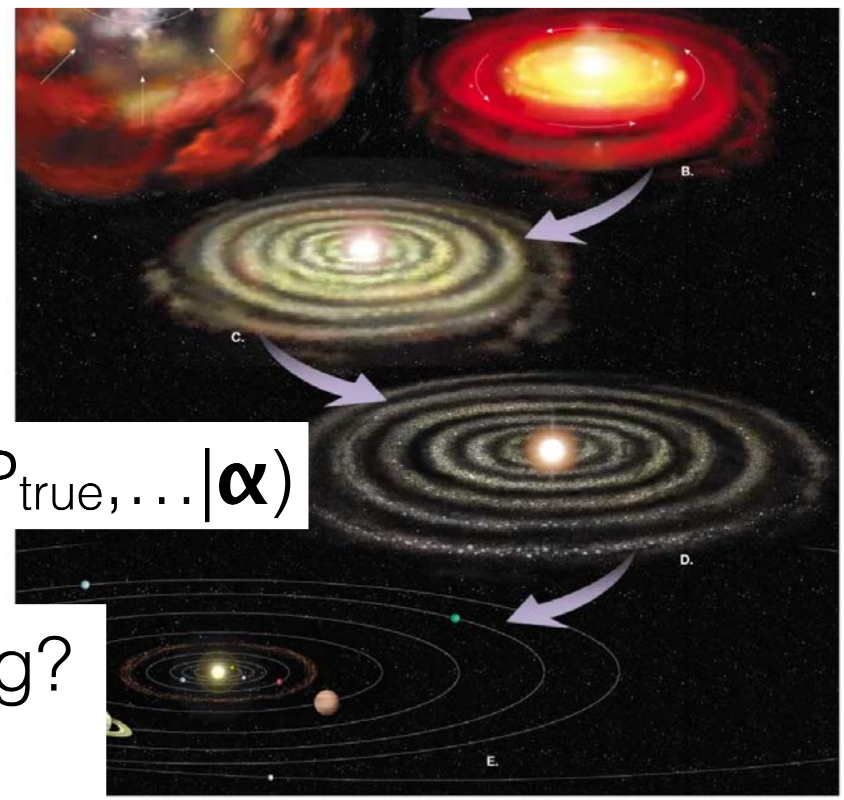


So what’s next?

On the Theory Side



There are many competing theories; like to quantitatively compare which is a better fit to data.



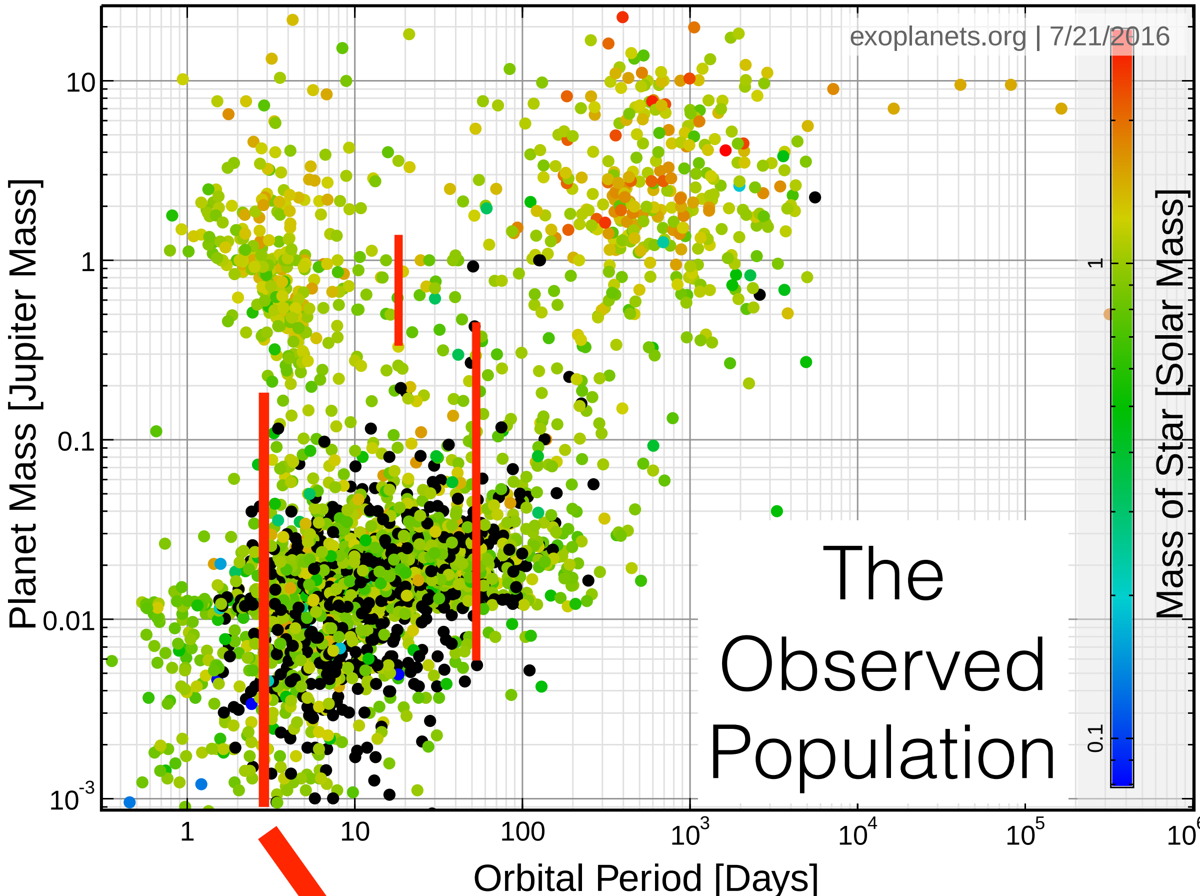
M_{true}, P_{true}, ...

$f(M_{true}, P_{true}, \ldots | \boldsymbol{\alpha})$

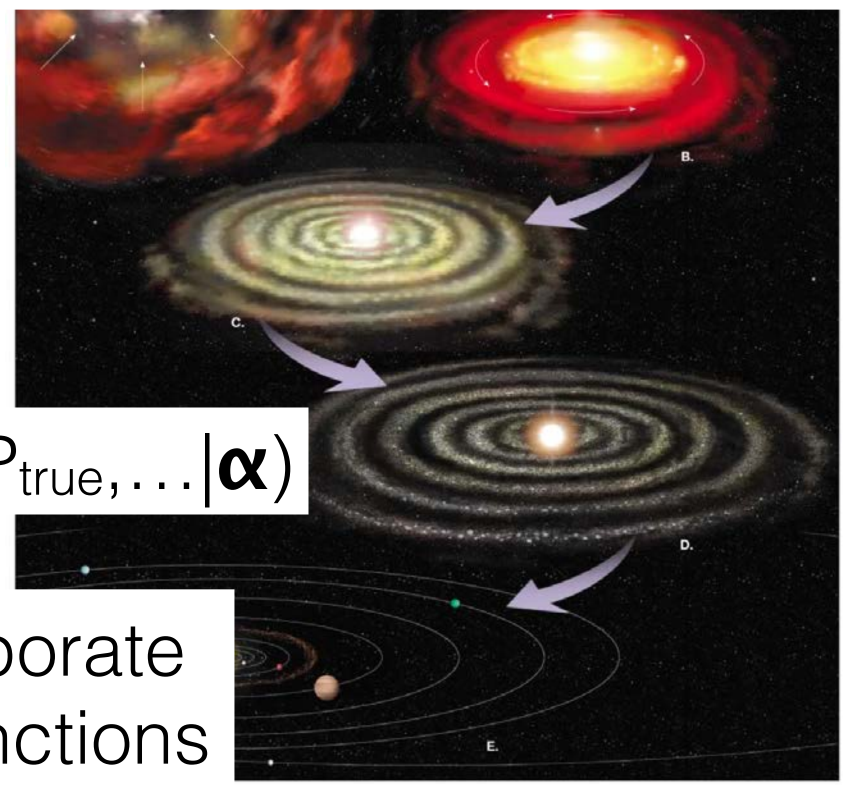
Now: BIC, qualitative

Next: Nested Sampling? Need your help!

More on the Theory Side



Planet formation is complicated, and $f(M\ldots|\boldsymbol{\alpha})$ involves expensive computer simulations.



M_{true}, P_{true}, …

$f(M_{\mathrm{true}}, P_{\mathrm{true}}, \ldots|\boldsymbol{\alpha})$

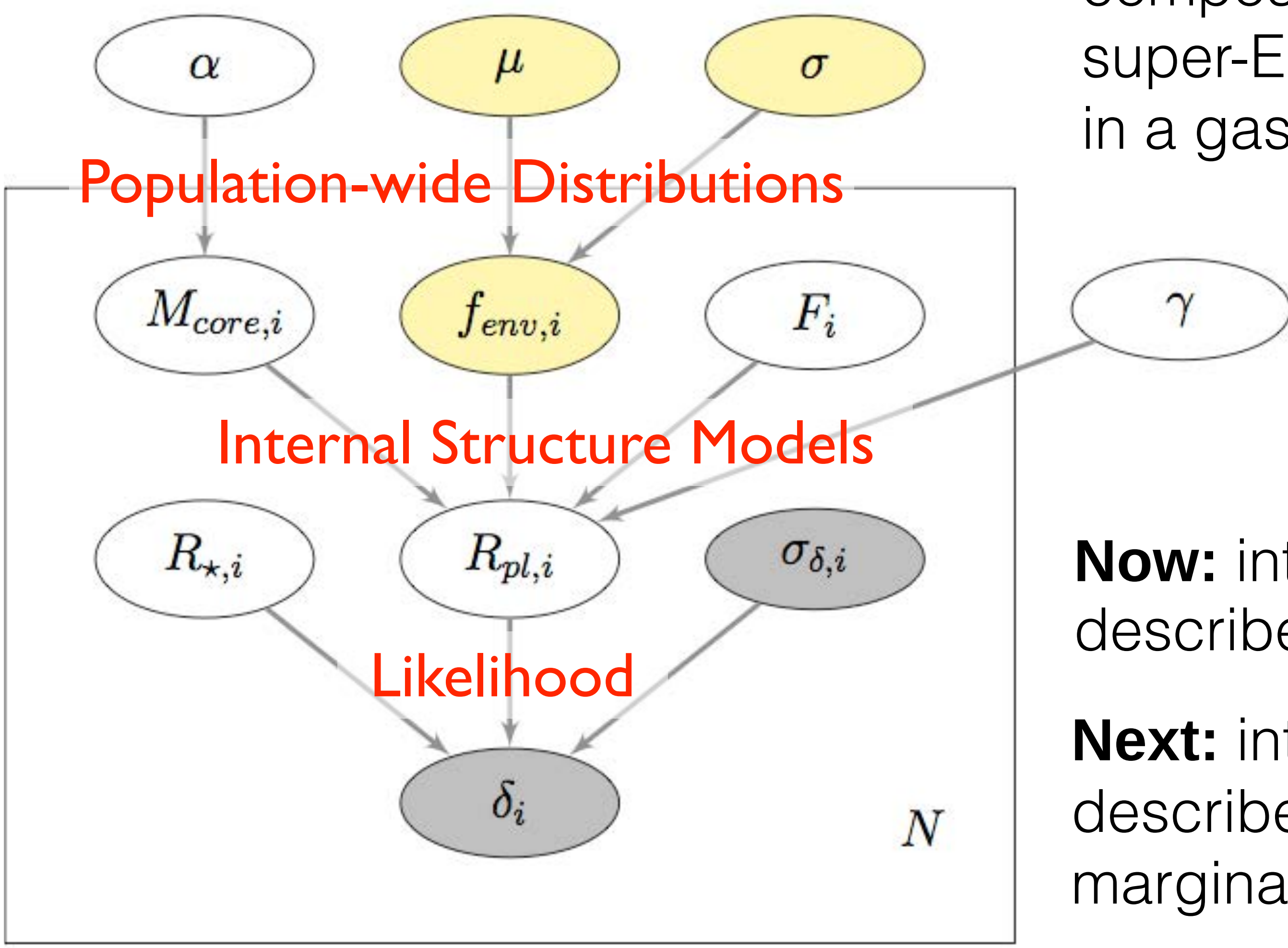
Now: parametric functions, little theory

Next: incorporate emulator functions

Emulators: An example

Sub-Neptune compositions: Wolfgang & Lopez, 2015

Wanted to understand BOTH:

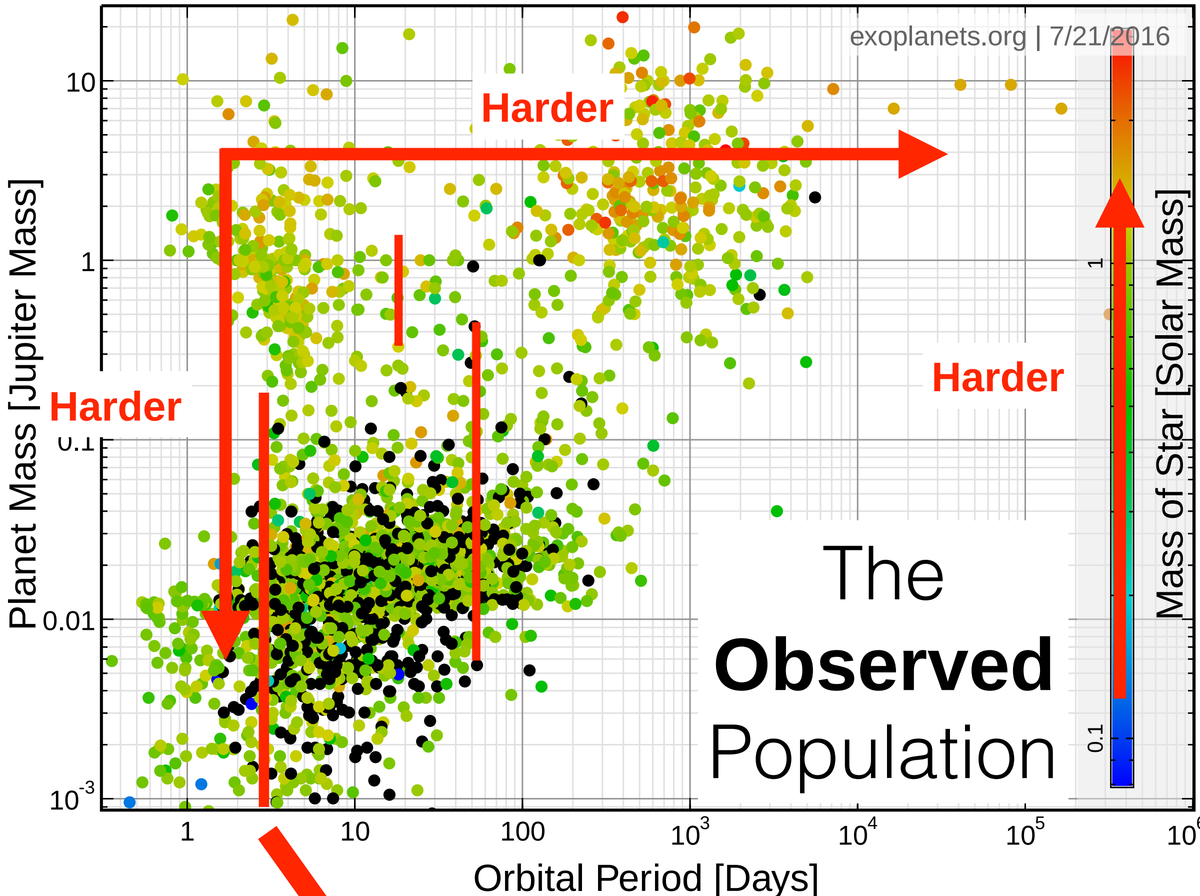


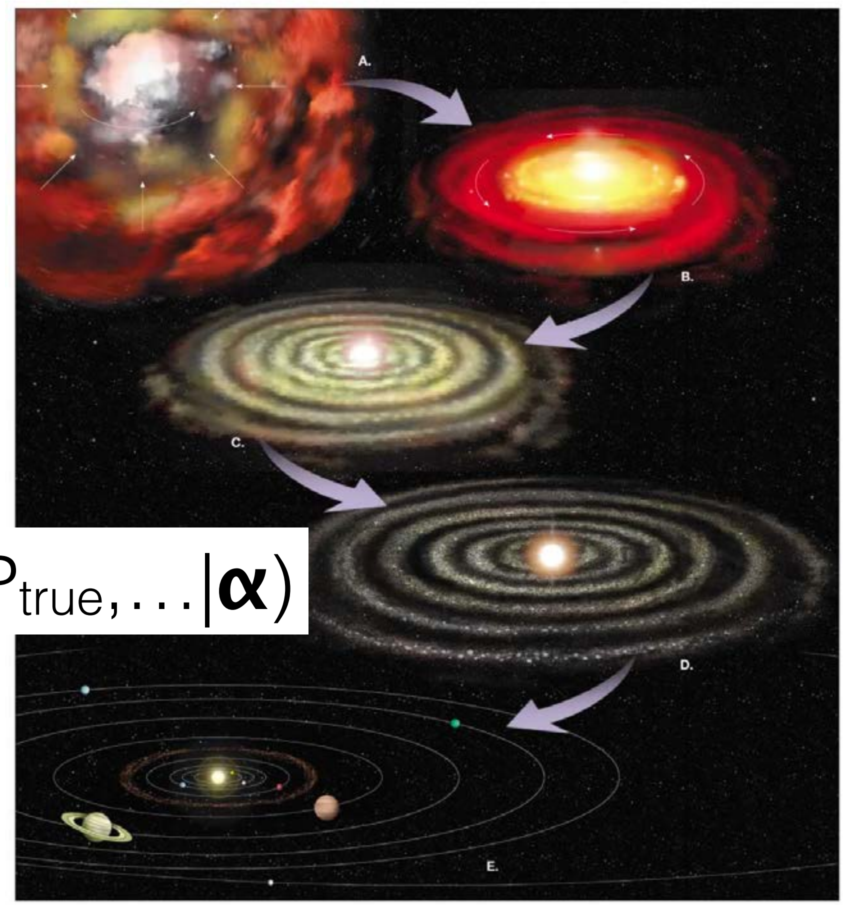
- compositions of individual super-Earths (fraction of mass in a gaseous envelope: f_{env})
- the distribution of this composition parameter over the Kepler population (μ, σ).

Now: internal structure models described by power laws

Next: internal structure models described by nonparametric/ marginally parametric distributions

On the Data Side

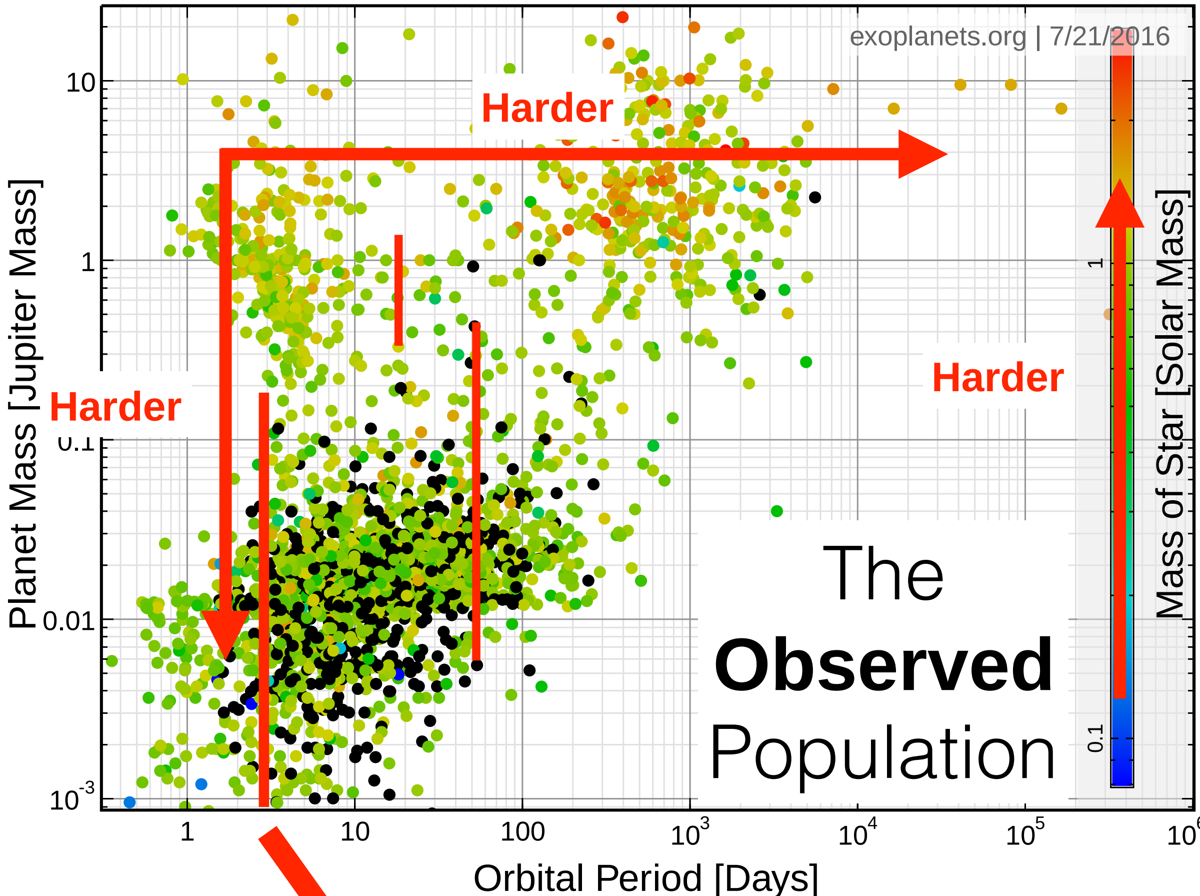




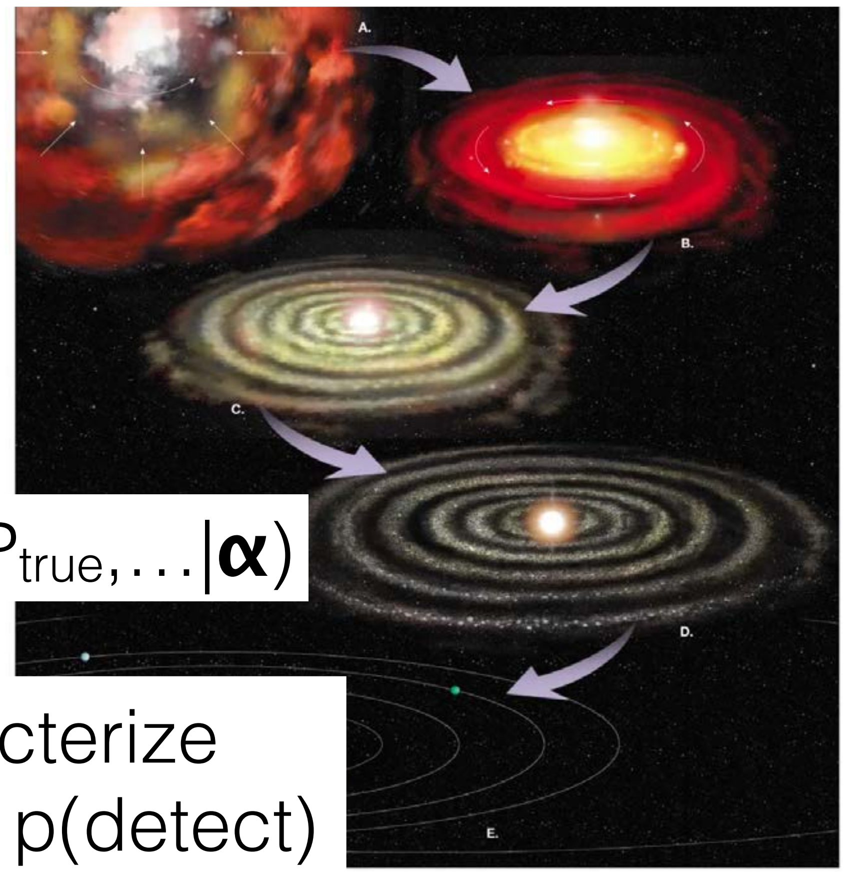
$M_{true}, P_{true}, \ldots$

$f(M_{true}, P_{true}, \ldots | \boldsymbol{\alpha})$

On the Data Side



Non-trivial detection functions are present in the observed population



M_{true}, P_{true}, …

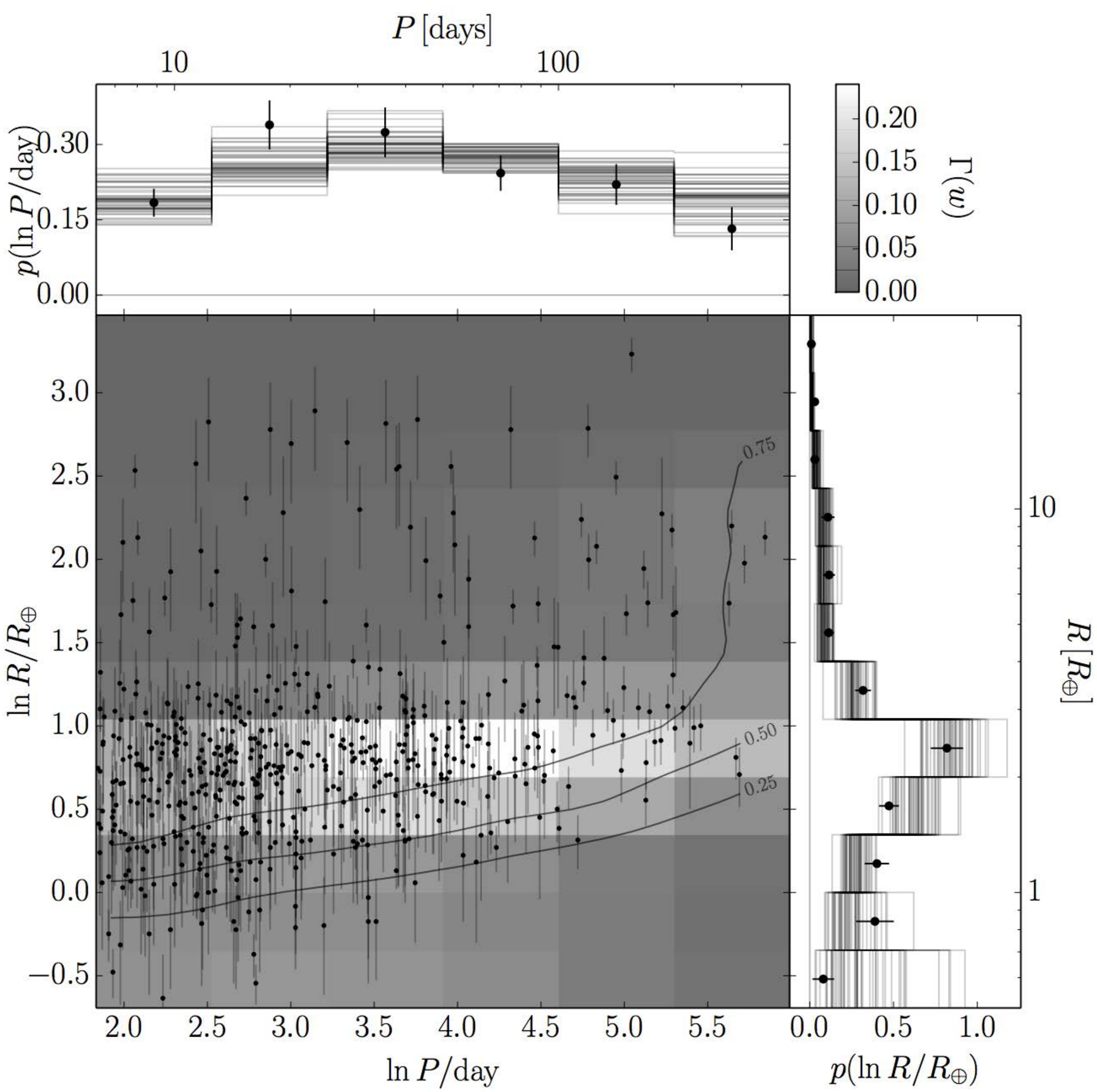
$f(M_{true}, P_{true}, \ldots | \boldsymbol{\alpha})$

Now: ignore p(detect) or cut stellar sample

Next: characterize and include p(detect)

Example: Including p(det)

Kepler occurrence rates: Foreman-Mackey et al. 2014



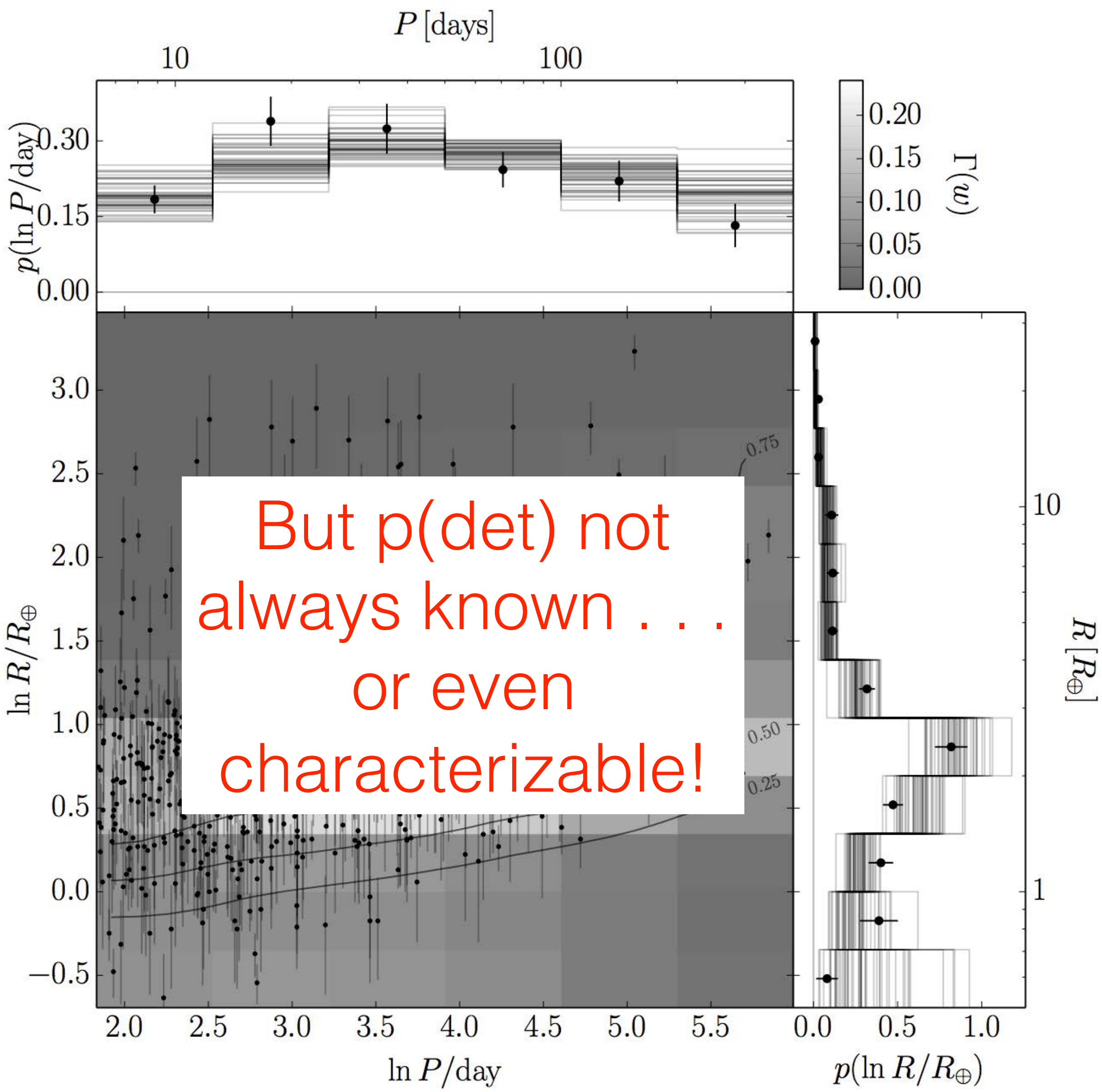
p(det) characterized by injecting synthetic transit signals in data and running detection algorithm on them (Petigura et al. 2014)

Grid of recovery fraction vs. radius and period

Incorporated with inferred occurrence rate (Poisson point process)

Example: Including p(det)

Kepler occurrence rates: Foreman-Mackey et al. 2014



p(det) characterized by injecting synthetic transit signals in data and running detection algorithm on them (Petigura et al. 2014)

↓

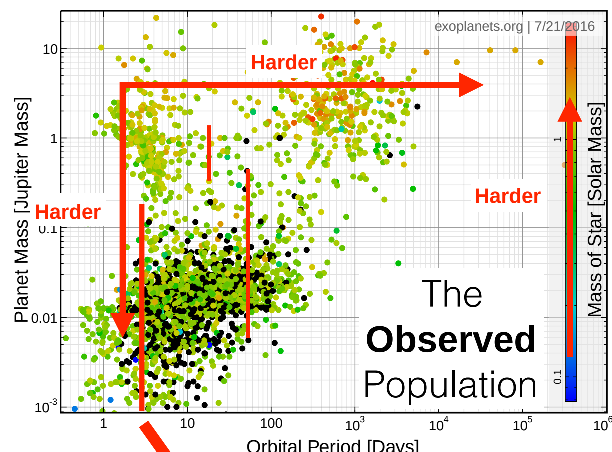
Grid of recovery fraction vs. radius and period

↓

Incorporated with inferred occurrence rate (Poisson point process)

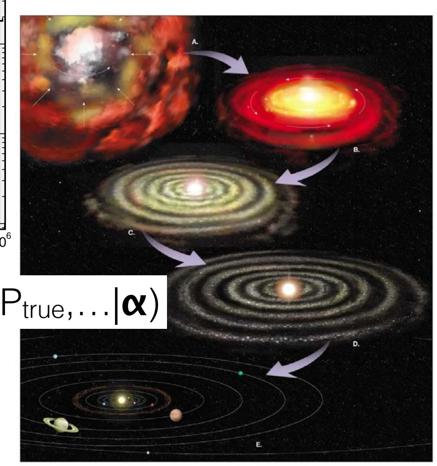
More on the Data Side

Mass, radius, period is not what we actually observe, and current likelihoods $p(M_{obs}|M_{true})$ very simple

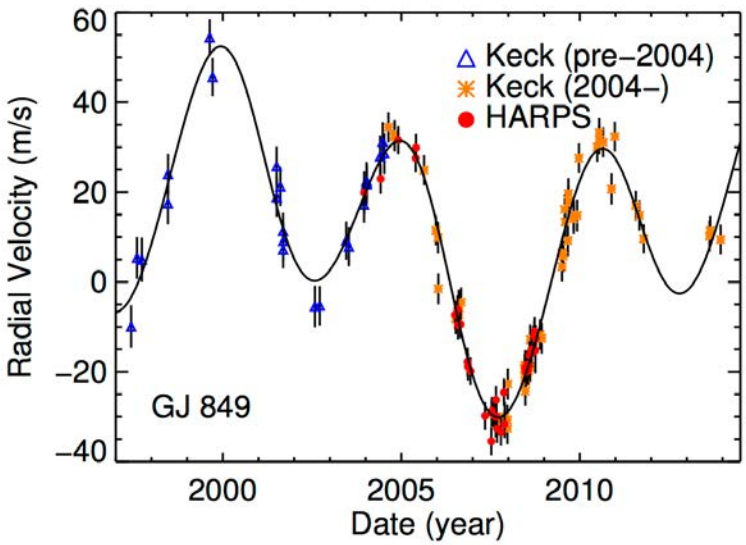


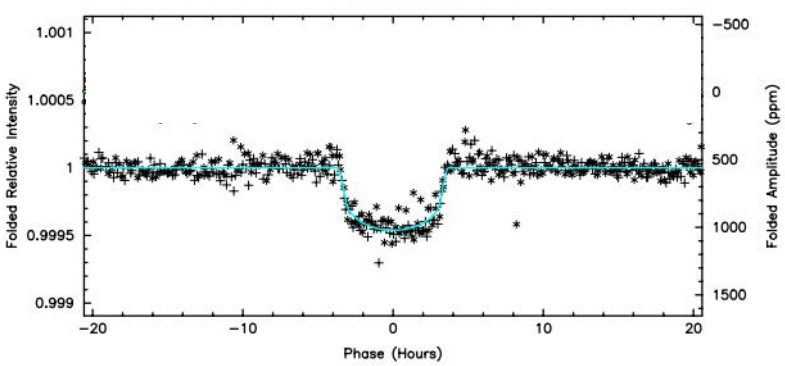
M_{true}, P_{true}, …

$f(M_{true}, P_{true}, \ldots|\boldsymbol{\alpha})$

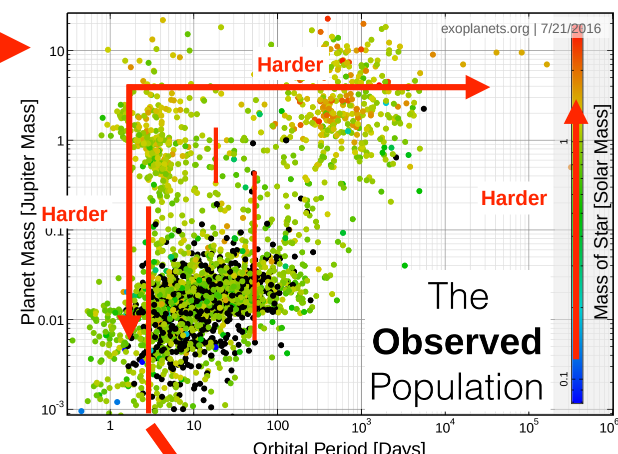


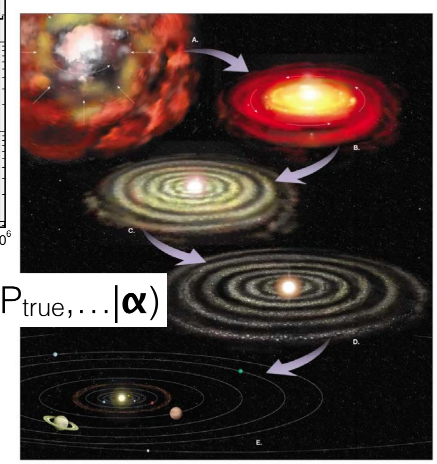
More on the Data Side





Mass, radius, period is not what we actually observe, and current likelihoods $p(M_{obs}|M_{true})$ very simple





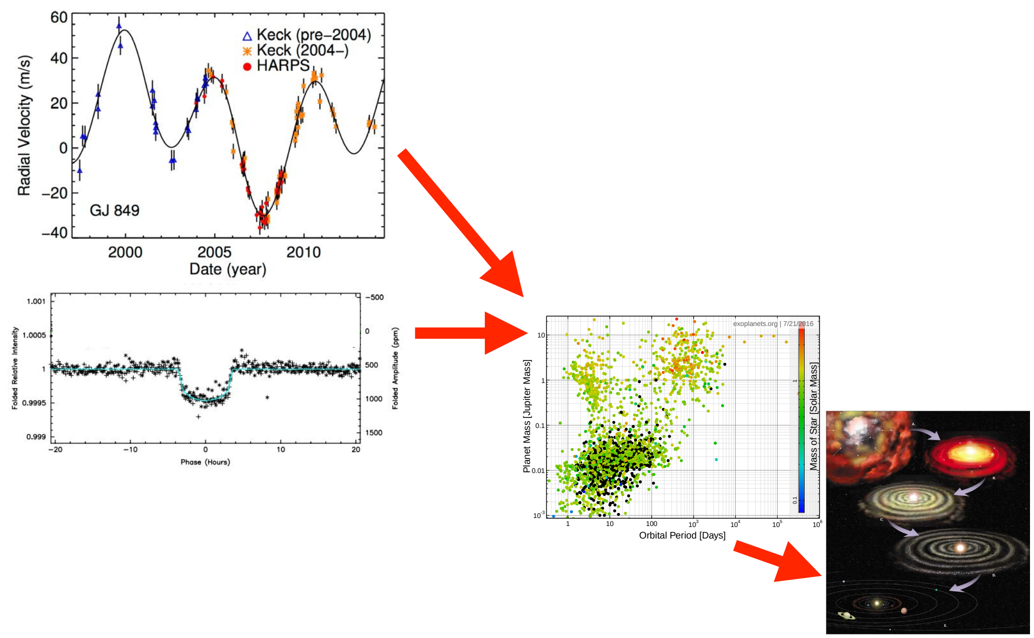
M_{true}, P_{true}, …

$f(M_{true}, P_{true}, \ldots|\boldsymbol{\alpha})$

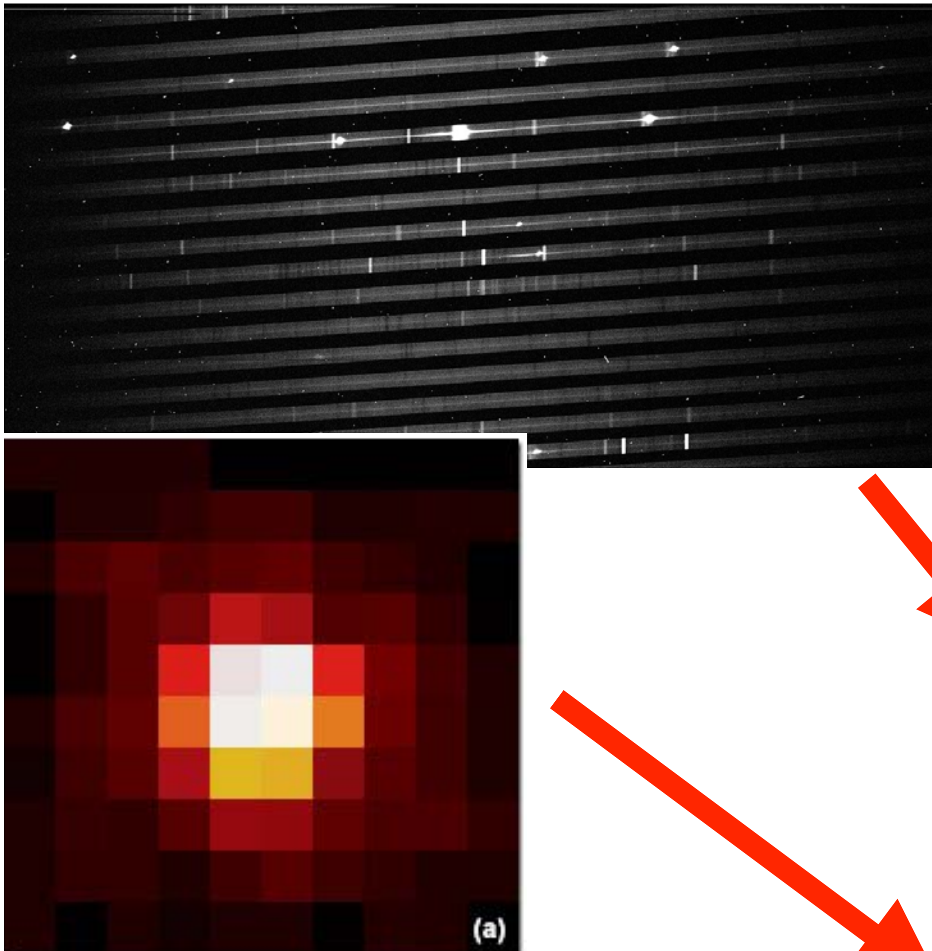
Next: inference on population directly from RVs vs. time, flux vs. time

Even more on the Data Side

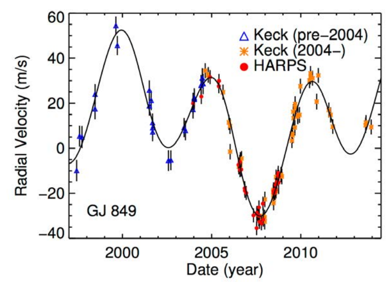
But we don't actually observe RVs vs. time or flux vs. time either . . . our real data is light on a detector

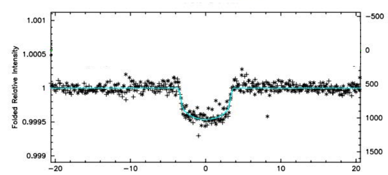


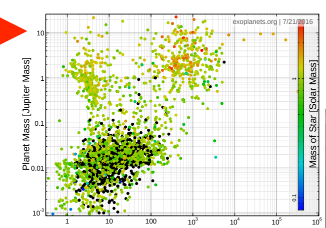
Even more on the Data Side

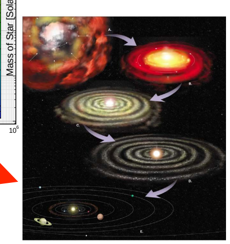


But we don't actually observe RVs vs. time or flux vs. time either . . . our real data is light on a detector









Next next: > 3 level HBMs, inference straight from the actual data

Next: Super-Earth Compositions

Understanding selection effects in mass-radius space:
Wolfgang, Jontof-Hutter, & Ford, in prep.

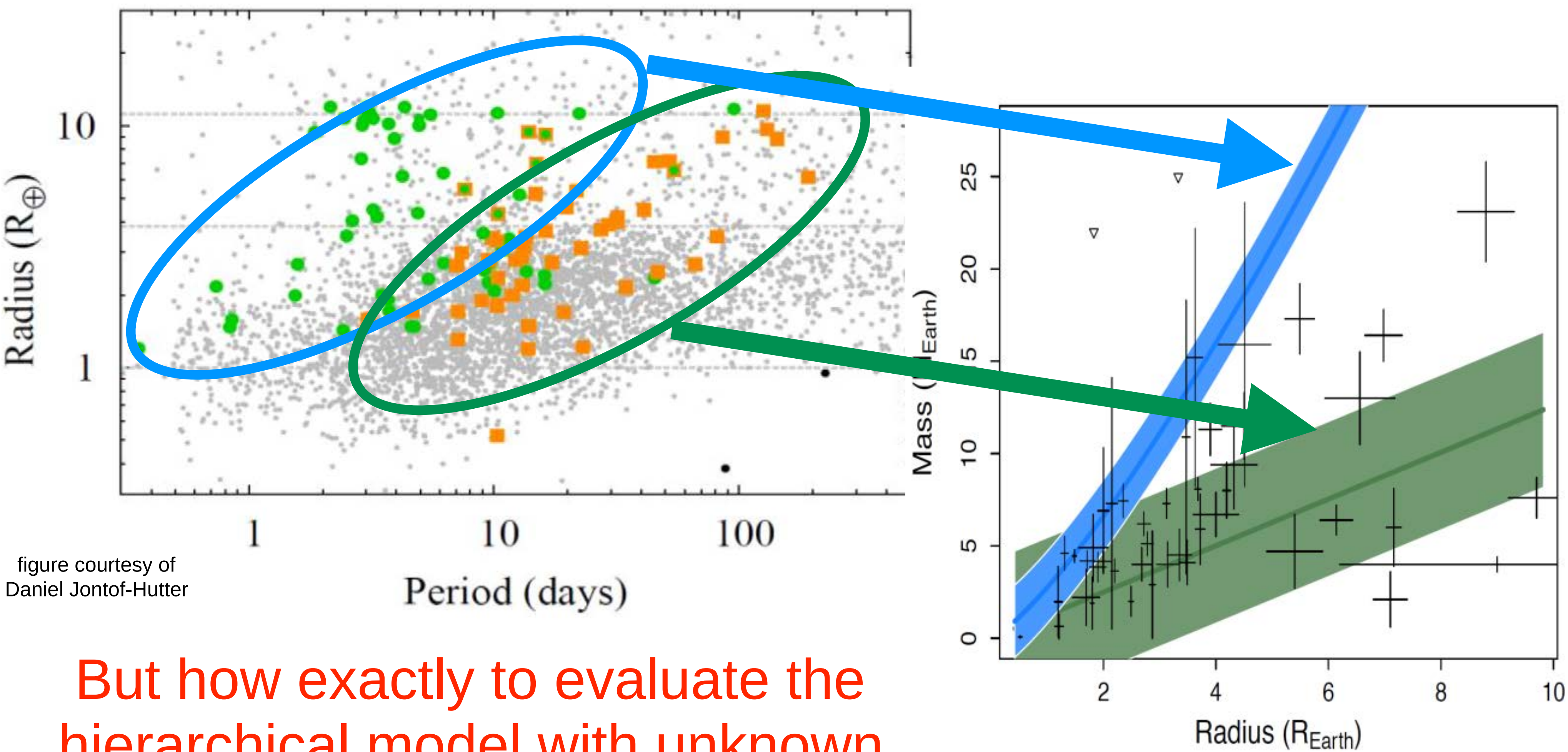
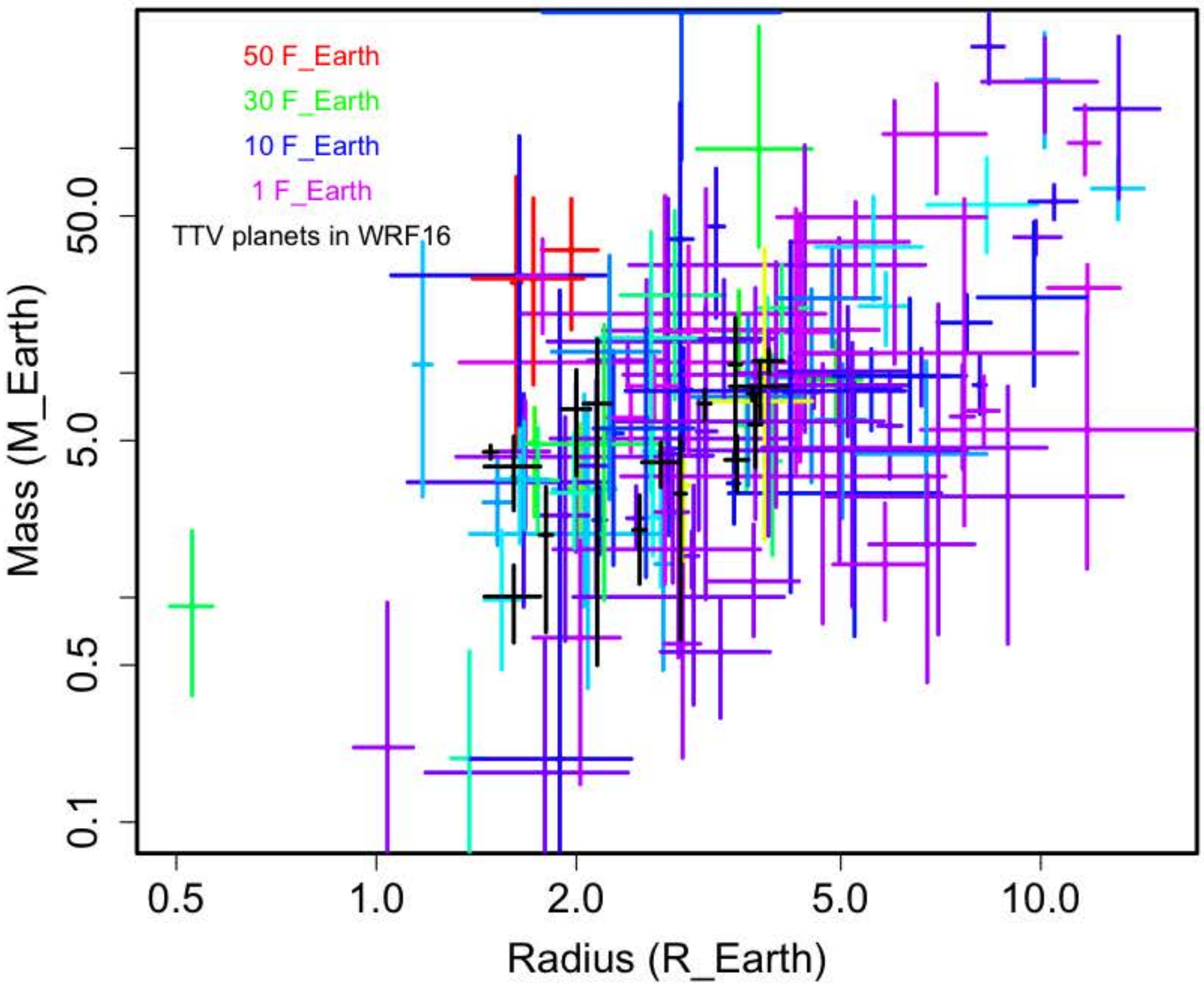


figure courtesy of
Daniel Jontof-Hutter

But how exactly to evaluate the hierarchical model with unknown number of non-detections?

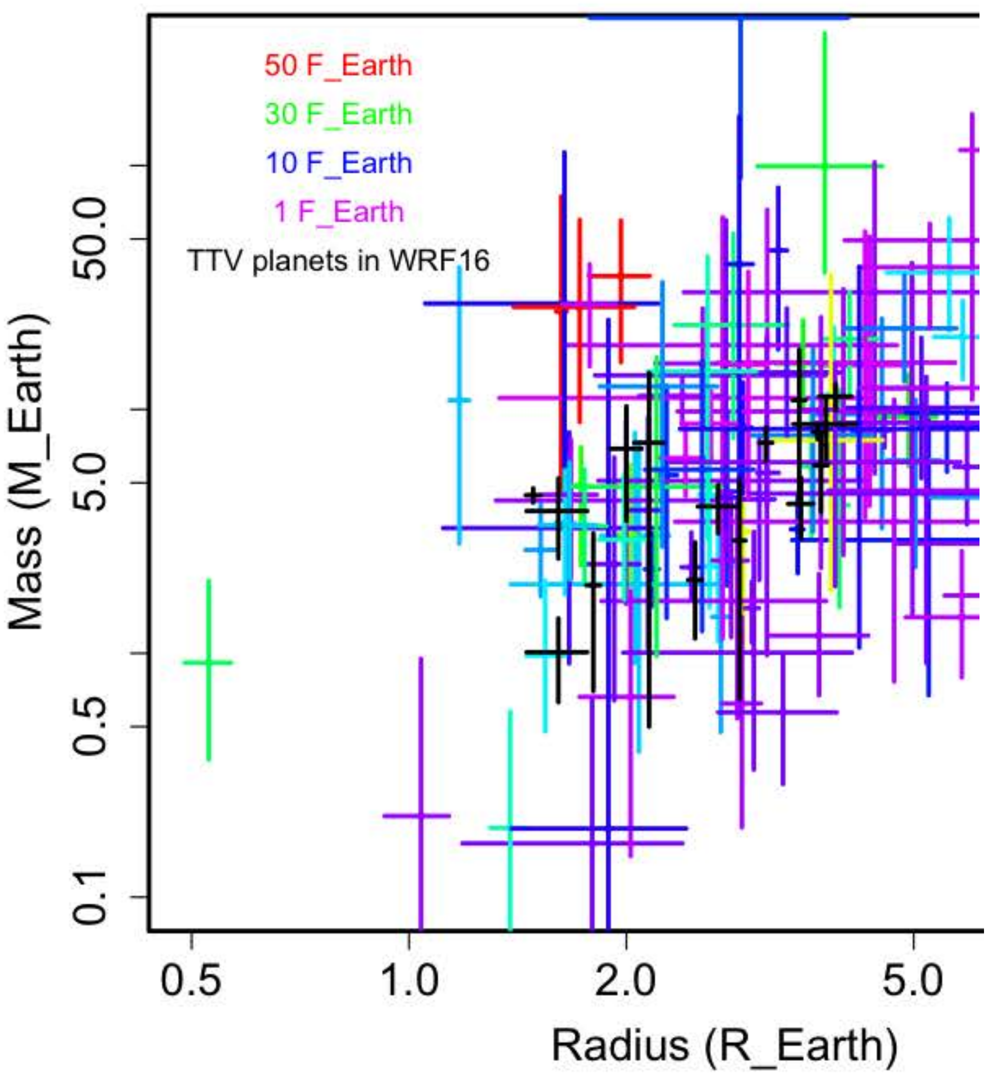
Next: Super-Earth Compositions

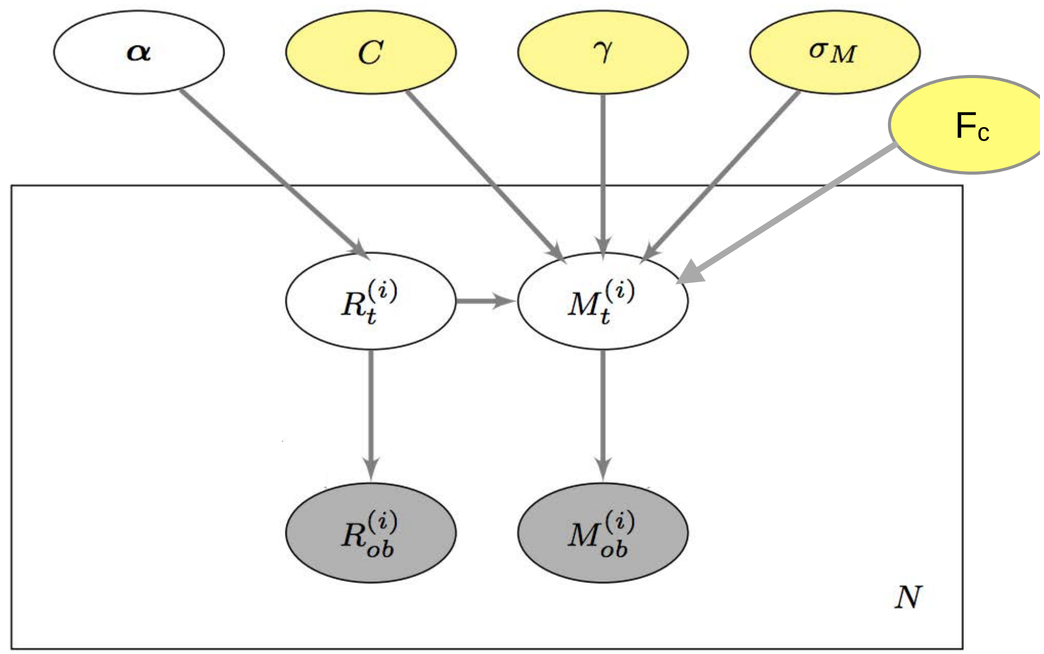
Characterizing Joint Mass-Radius-Flux Distribution:
Wolfgang, Jontof-Hutter, Rogers & Ford, in prep.



Next: Super-Earth Compositions

Characterizing Joint Mass-Radius-Flux Distribution:
Wolfgang, Jontof-Hutter, Rogers & Ford, in prep.

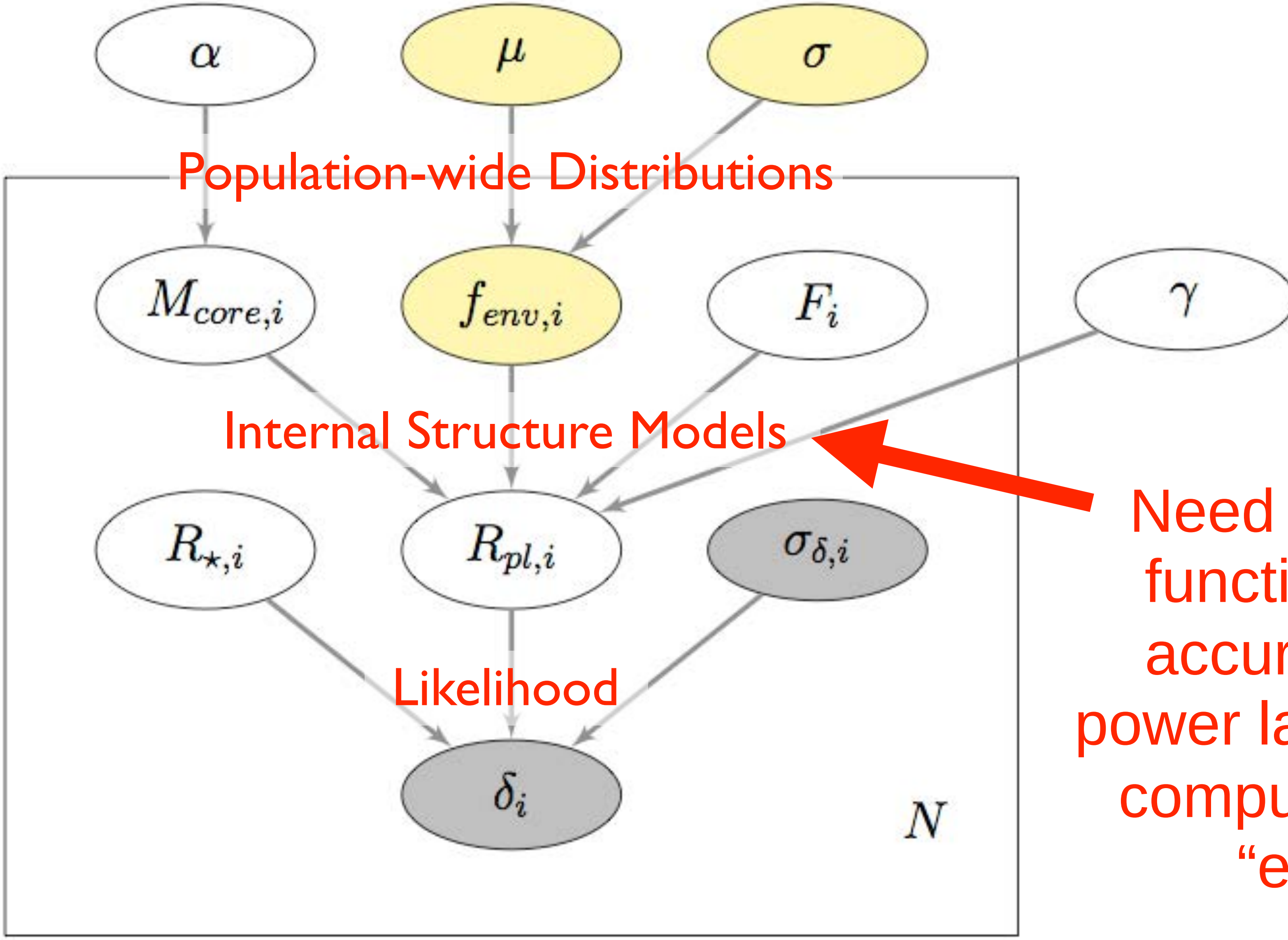




But many options for parameterizations … hierarchical model comparisons?
Poisson point process?

Next: Super-Earth Compositions

Initial Sub-Neptune compositions: Wolfgang & Lopez, in prep.



Need emulator function more accurate than power laws but still computationally "easy".

Summary:

Where we are:

~ a dozen exoplanet astronomers
working on very simple hierarchical models
describing distributions of planet properties

Where we can go this year at SAMSI:

1) Incorporate survey detection efficiency
2) Develop emulator functions to include computationally expensive theoretical simulations directly into HBM
3) Compare different theoretical models via hierarchical model comparison
4) Implement more realistic likelihoods: inference from lower-level data