

**The LSST Informatics and
Statistics Science Collaboration:
*Activity at the
SAMSI ASTRO Program***

Tom Loredo (Cornell U.) and Ashish Mahabal (Caltech)
(see Ashish's companion presentation)

The ISSC

<https://issc.science.lsst.org/>



[PUBLIC & SCIENTISTS](#)

[PROJECT TEAM](#)

[LSST CORPORATION](#)

[Home](#)

[Activities](#)

[Documents](#)

[Members](#)

[Apply](#)

[Contact](#)

LSST Informatics & Statistics Science Collaboration

Welcome to the web site for the LSST Informatics and Statistics Science Collaboration (ISSC).

The ISSC consists of over [40 data scientists](#) devoted to developing tools for use with large astronomical surveys. Our team includes astronomers, statisticians, computer scientists, and machine learning researchers, with a shared objective of addressing the inference challenges facing LSST as it works to meet its scientific goals.

The ISSC is led by a core team consisting of the following members:

Jogesh Babu

Tamas Budavari

Eric Feigelson

Tom Lored, co-chair

Chad Schafer, co-chair

Sam Schmidt

Robert Wolpert

Purpose of the ISSC

To help coordinate interdisciplinary interactions and pursue data analysis methodology research in collaboration with other Science Collaborations (SCs) and the broader LSST user community.

— 2009 ISSC proposal

ISSC objectives

- Identify promising areas of methodological overlap between applications and disciplines
- Bring existing methods to bear on these overlap problems, via education and consultation
- Build alliances pursuing research on new methods where needed

Key features

- **Interdisciplinary:** 60% astronomers, 40% information scientists
- **Dual roles:** Members undertake research (often in association with specific partner SCs), and serve as consultants
- Focus on areas where research alliances cutting across astronomy application areas can benefit LSST science
- A *core team* of 7 members (including the two co-chairs) coordinates ISSC efforts; broader membership has "regular" and "affiliate" categories (this is evolving)

Activities

<https://issc.science.lsst.org/projects>



[PUBLIC & SCIENTISTS](#)

[PROJECT TEAM](#)

[LSST CORPORATION](#)

[Home](#)

[Activities](#)

[Documents](#)

[Members](#)

[Apply](#)

[Contact](#)

[Home](#) » [Informatics & Statistics Science Collaboration](#) [Activities](#)

Informatics & Statistics Science Collaboration Activities

Activities

2016-2017 SAMSI Program on Astrostatistics

Image:



The [Statistical and Applied Mathematical Sciences Institute](#), located in the Research Triangle of North Carolina, will host a year-long program entitled "[Statistical, Mathematical and Computational Methods for Astronomy](#)." Among a handful of planned working groups is one focused on LSST statistical challenges. Specific research areas are under discussion; please contact the organizers if you have suggestions.

[Read more](#)

Current/planned activities

- **Research**
- **Tutorials:** Coordinated by Feigelson, with IAU/AAS/ASA/IAA participation
- **Curators:**
 - ▶ Provide a highly visible contact to help promulgate useful techniques and tools
 - ▶ Avoid duplication of effort
 - ▶ Initial areas: Event class'n; photo-z
- **Data challenges:** Transient class'n...

SAMSI

Statistical and Applied Mathematical Sciences Institute

<https://www.samsi.info/>



Research Technology Park, NC, USA

SAMSI

- Part of the Mathematical Sciences Institutes program of the NSF DMS
- In Research Triangle Park, NC, partnered with Duke U., UNC Chapel Hill, NC State U., and the National Institute of Statistical Sciences (NISS)
- **Mission:** "Forge a synthesis of the statistical sciences and the applied mathematical sciences with disciplinary science to confront the very hardest and most important data- and model-driven scientific challenges."
- **Mechanism:** Sponsor & host two or three *9-month interdisciplinary research programs* per year, and shorter summer programs

SAMSI ASTRO program

*Program on Statistical, Mathematical
and Computational Methods for Astronomy*

<http://samsi.info/astro>



Jogesh Babu, leader
CAST @ PSU



Sujit Ghosh, SAMSI liaison
NCSU

+16 co-organizers

Program elements

- Opening Workshop, Aug 2016



- *Working groups*
- Mid-program workshops: ExoPop, Transients, UQPop...
- Transition Workshop, May 2017
- Astrostat course: James Long
- 3 postdocs: David Jones, David Stenning, Hyungsuk Tak
- Long-term visitors: Babu, Feigelson, Long; several ~1-mo visitors in Spring 2017

ASTRO working groups

- **WG1:** Uncertainty Quantification and Astrophysical Emulation — Ilya Mandel, Derek Bingham
- **WG2:** Synoptic Time Domain Surveys — Ashish Mahabal, Jogesh Babu
- **WG3:** Multivariate and Irregularly Sampled Time Series — Ben Farr, Soumen Lahiri
- **WG4:** Astrophysical Populations — Eric Ford, Jessi Cisewski
- **WG5:** Statistics, computation, and modeling in cosmology — Elisabeth Krause, Joe Guinness

WG organization/operation

- Weekly telecons, mostly via Webex
- Most have subgroups with a small team trying to address specific topics, reporting occasionally to the full WG
- WGs organize the mid-term workshops
- WGs have fluid membership; some will partially "reset" after the New Year; new members welcome (contact leaders)

WG3: Multivariate and Irregularly Sampled Time Series

Science drivers/subgroups

- LIGO: How to use diverse multivariate data
- Pulsar timing arrays: Outliers
- AGN, exoplanets: Irreg. sampled TS

Statistical themes

- Beyond kernel-based Gaussian processes: continuous time, non-Gaussian, heavy-tailed
- Functional data analysis (FDA): populations of time series/spectra

WG4: Astrophysical populations

Science drivers/subgroups

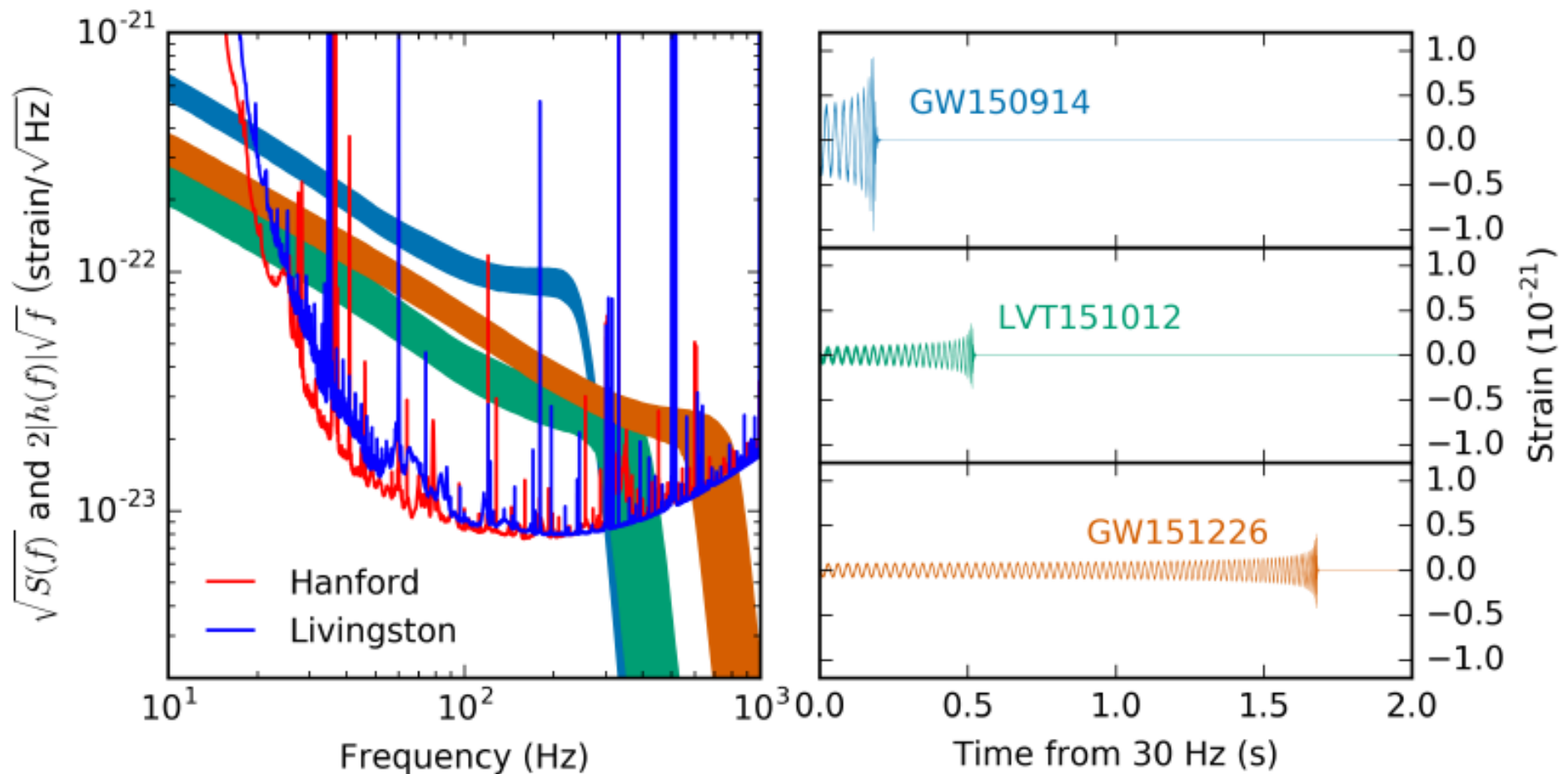
- LIGO: Occurrence rates, distributions
- Exoplanets: Occurrence rates, distributions, hi-res spectra with stellar activity
- Time-variable stellar spectra: Young stars, spectroscopic binaries

Statistical themes

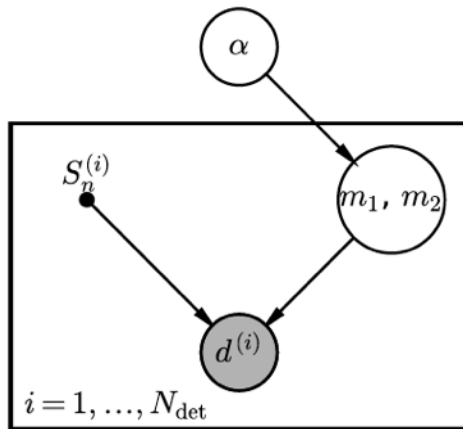
- Hierarchical Bayes & graphical models for demographics
- Nonparametric models for latent point processes, dynamic spectra
- Spectral decomposition and dimension reduction (PCA, diffusion maps...)
- Approximate Bayesian Computation (ABC)
- Functional data analysis

Cosmic demographics

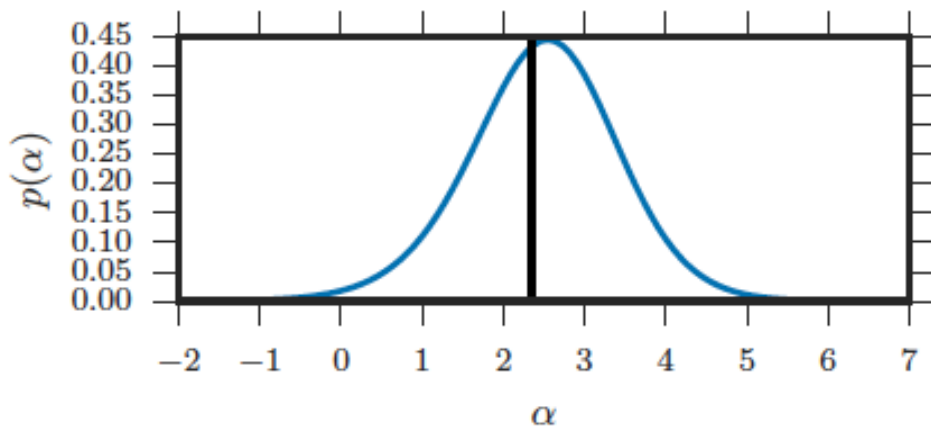
LIGO: Binary black holes (Farr, Mandel...)



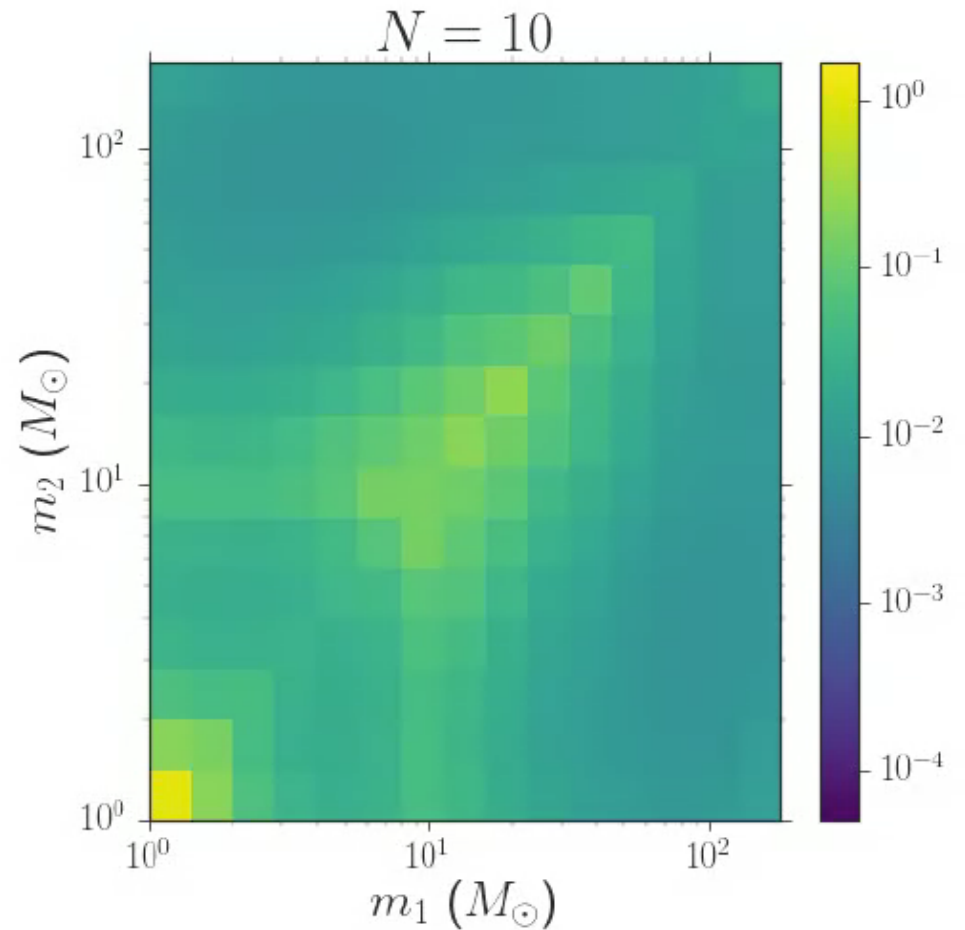
Current: Parametric HB models



$$m_1 \sim m_1^{-\alpha}$$
$$m_2 \sim U(5 M_{\odot}, m_1)$$



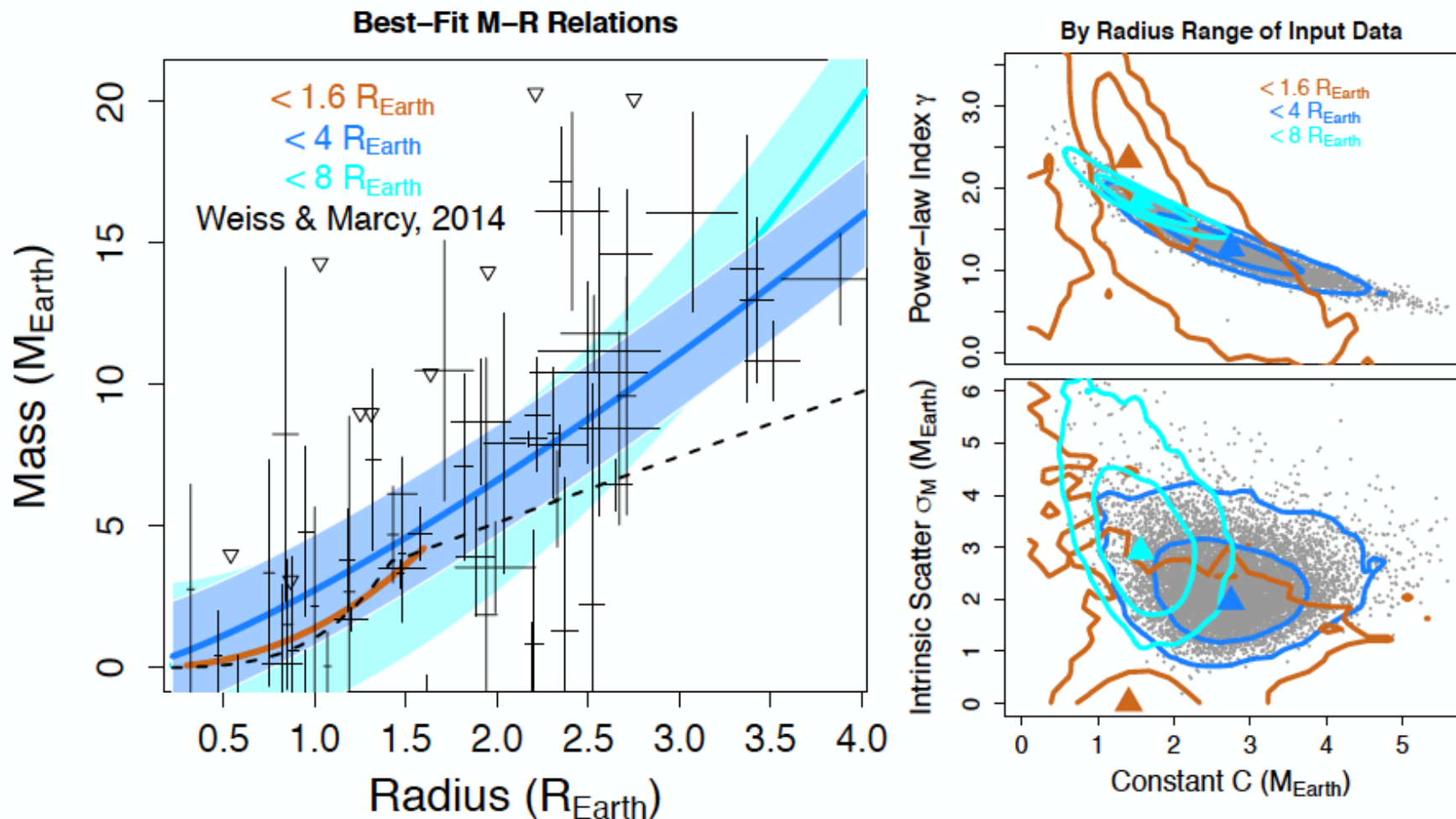
Future: ~Nonparametric models



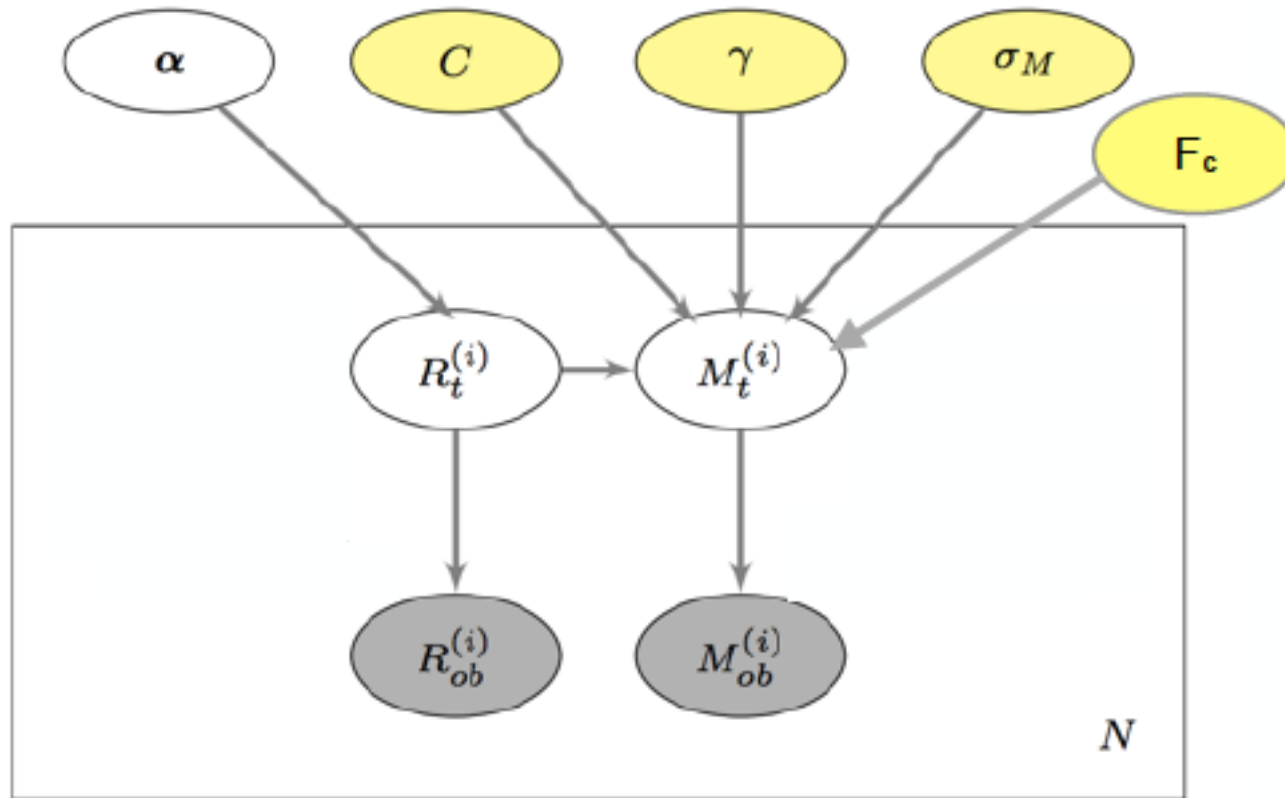
Exoplanets: Occurrence rate, properties

(main topic of ExoPop workshop, Oct 2016)

Wolfgang, Rogers, & Ford, 2016



Exoplanet Hierarchical Bayes conditional density estimation



New(ish) HB directions

Farr, Mandel (for LIGO):

Where Does Selection
Function Go?

$$p(d, m_1, m_2 | \alpha) = \frac{p(d | m_1, m_2) p(m_1, m_2 | \alpha)}{\beta(\alpha)},$$

$$\beta(\alpha) = \int_{\{d|f(d)>f_0\}} dm_1 dm_2 dd p(d | m_1, m_2) p(m_1, m_2 | \alpha),$$

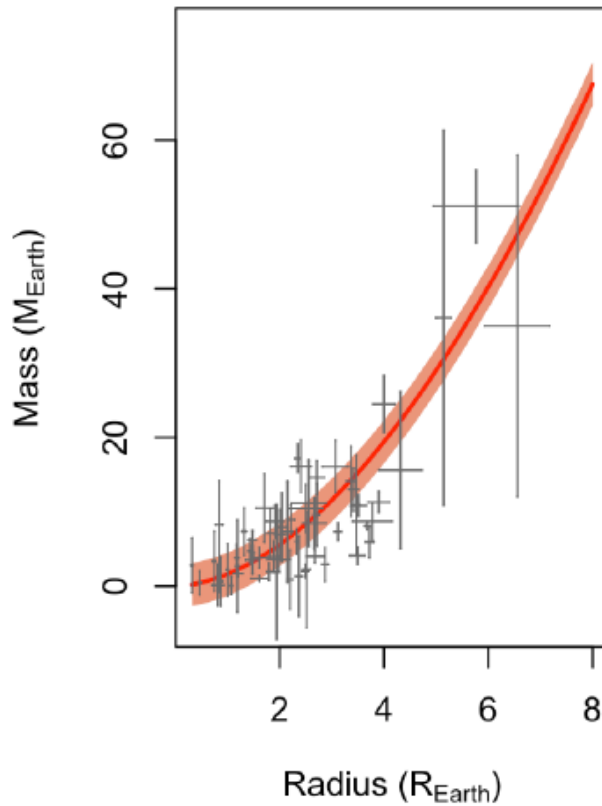
$$\int_{\{d|f(d)>f_0\}} dd p(d | m_1, m_2) p(m_1, m_2 | \alpha) = p(m_1, m_2 | \alpha) P_{\text{det}}(m_1, m_2)$$

Answer: into the normalisation!

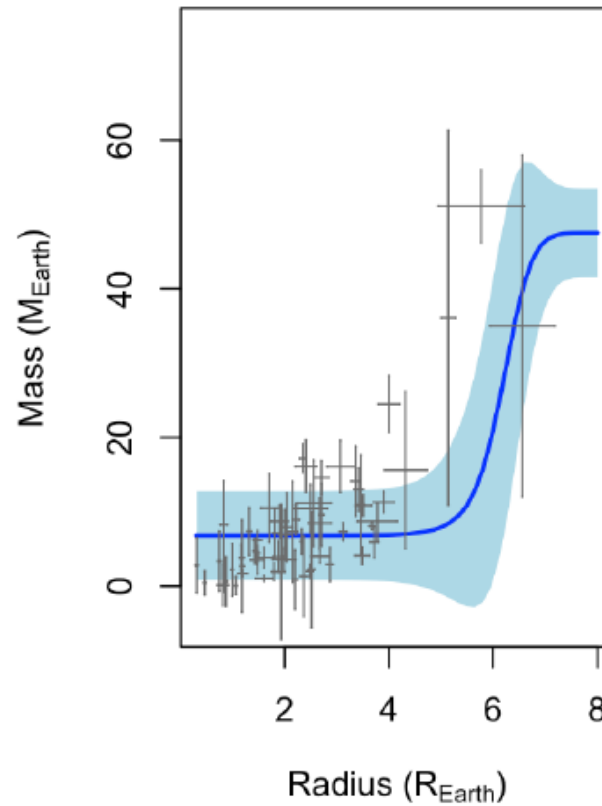
See Loredo (2004), Mandel, Farr & Gair (2016)

Nonparametric HB density estimation with Bernstein polynomials (Ghosh, Bo, Wolfgang...)

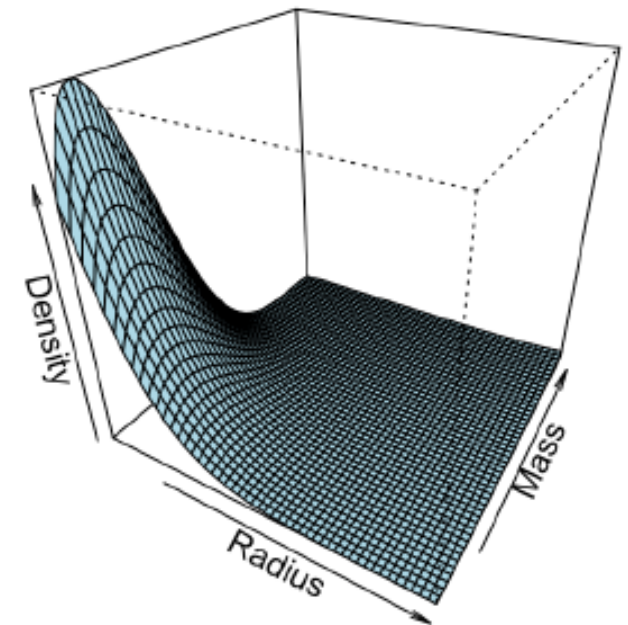
Power Law Fit M-R Relation



Nonparametric Model Fit M-R Relations

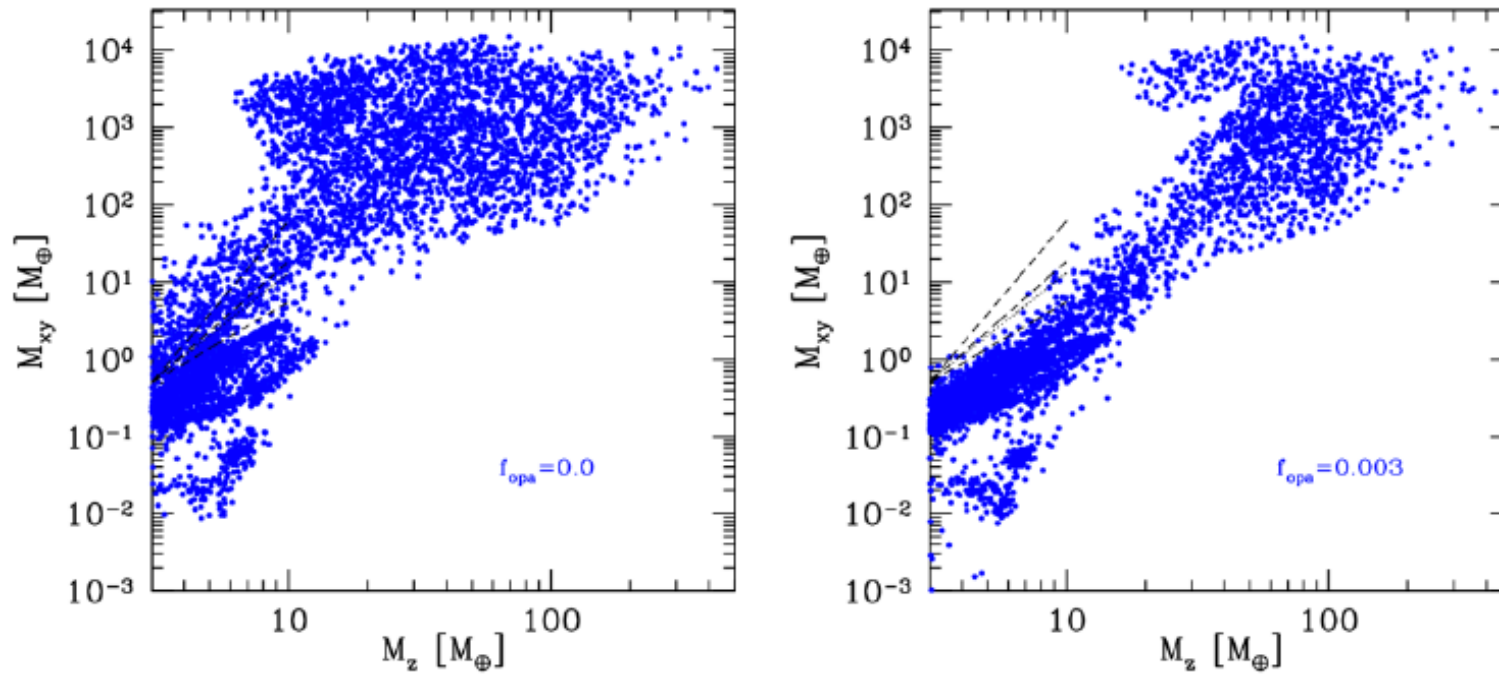


Joint Mass and Radius density plot



Population inference with simulation-based models (LIGO, exoplanets)

$M_{XY}-M_c$ From Planet Population Synthesis



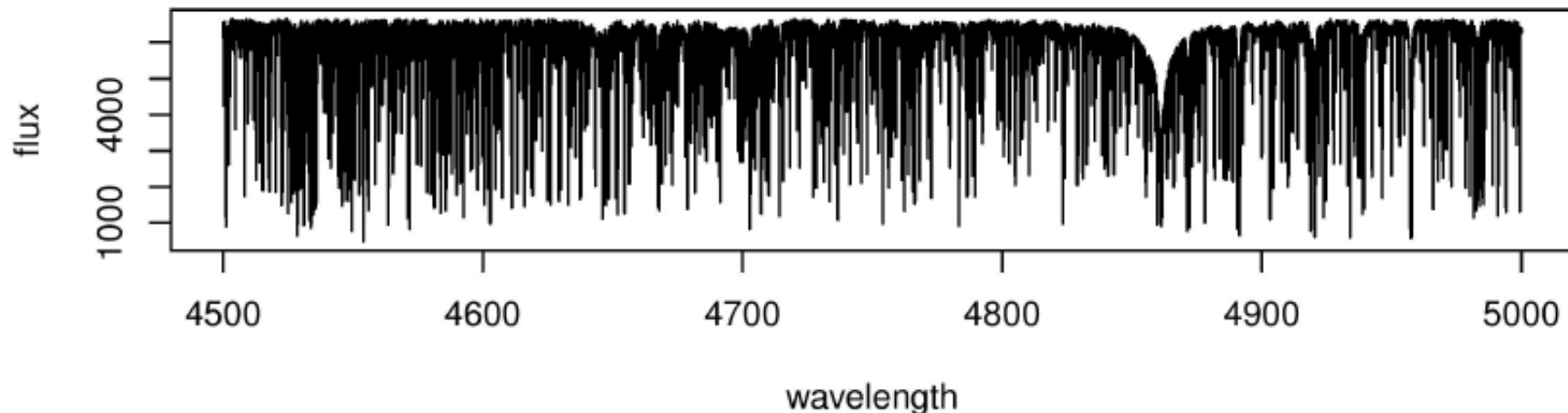
- Approximate Bayesian Computation (ABC)
- Bayesian emulation of simulators

Workshop: Astrophysical Population Emulation and Uncertainty Quantification, April 3-7, 2017

Modeling ensembles of hi-res spectra

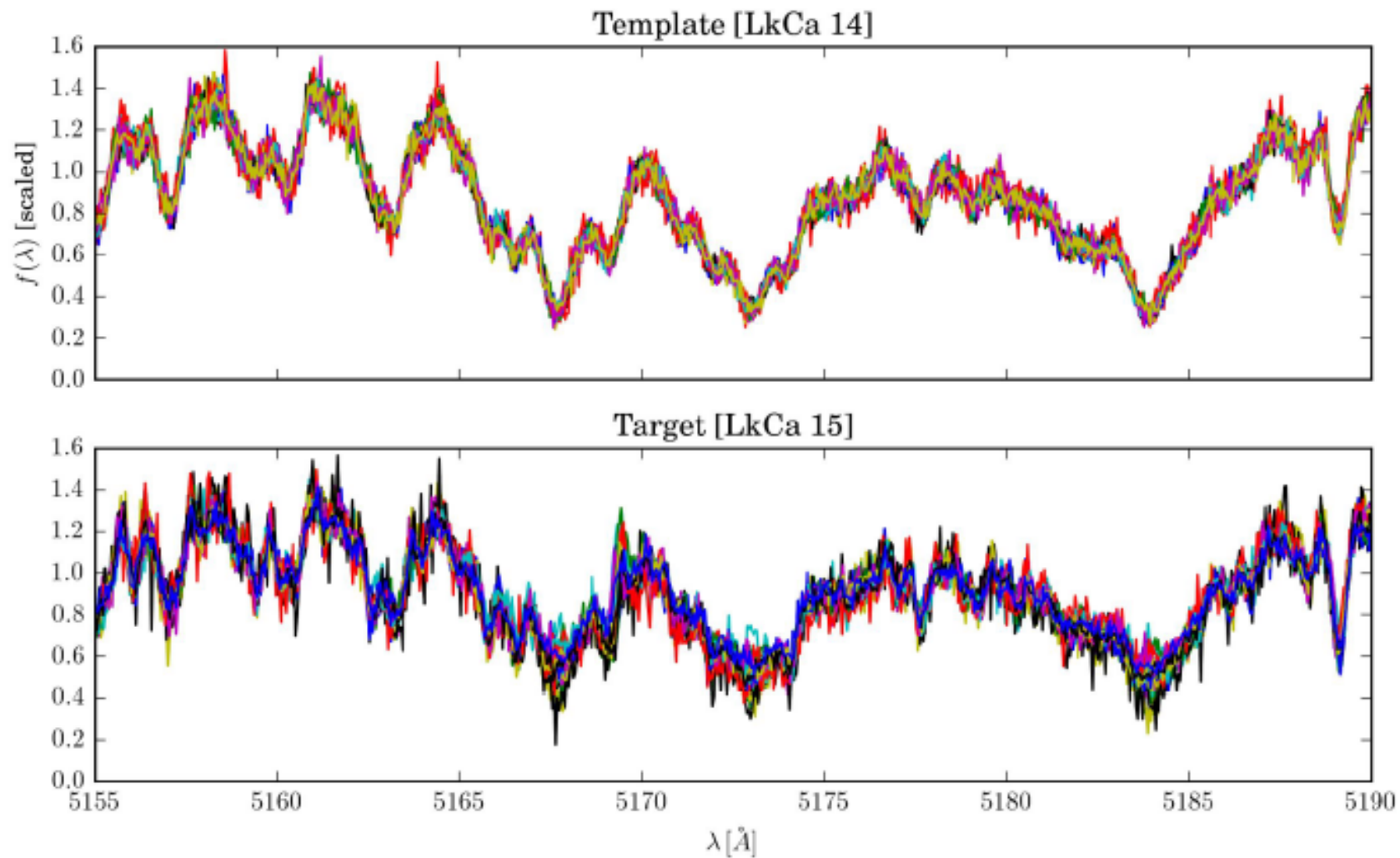
Dumusque et al. (2014): Spot Oscillation and Planet (SOAP) 2.0 simulation software

- ▶ Used to simulate the RV variations induced by active regions
- ▶ Settings: one large spot at fixed inclination (relative to Earth) and spot latitude
- ▶ Simulated spectra at 25 phases (-0.5 to 0.46 in steps of 0.04)



Ford, Cisewski, Stenning, Jones, Wolpert, Loredó...

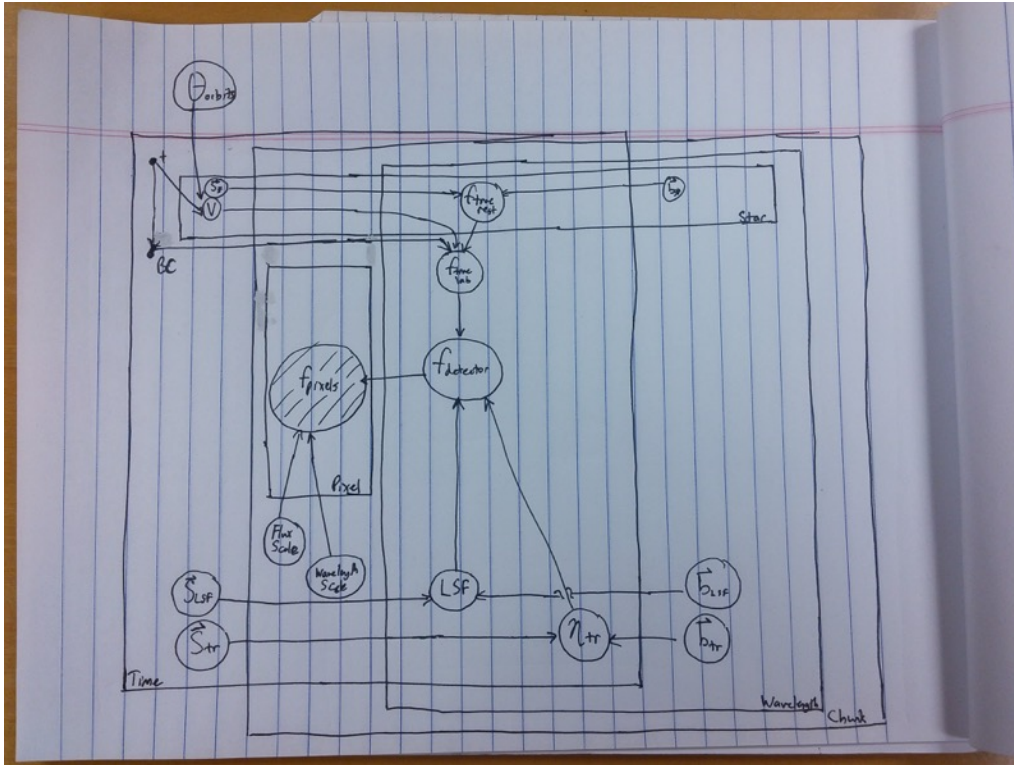
Young stars: Stellar atmosphere + accretion



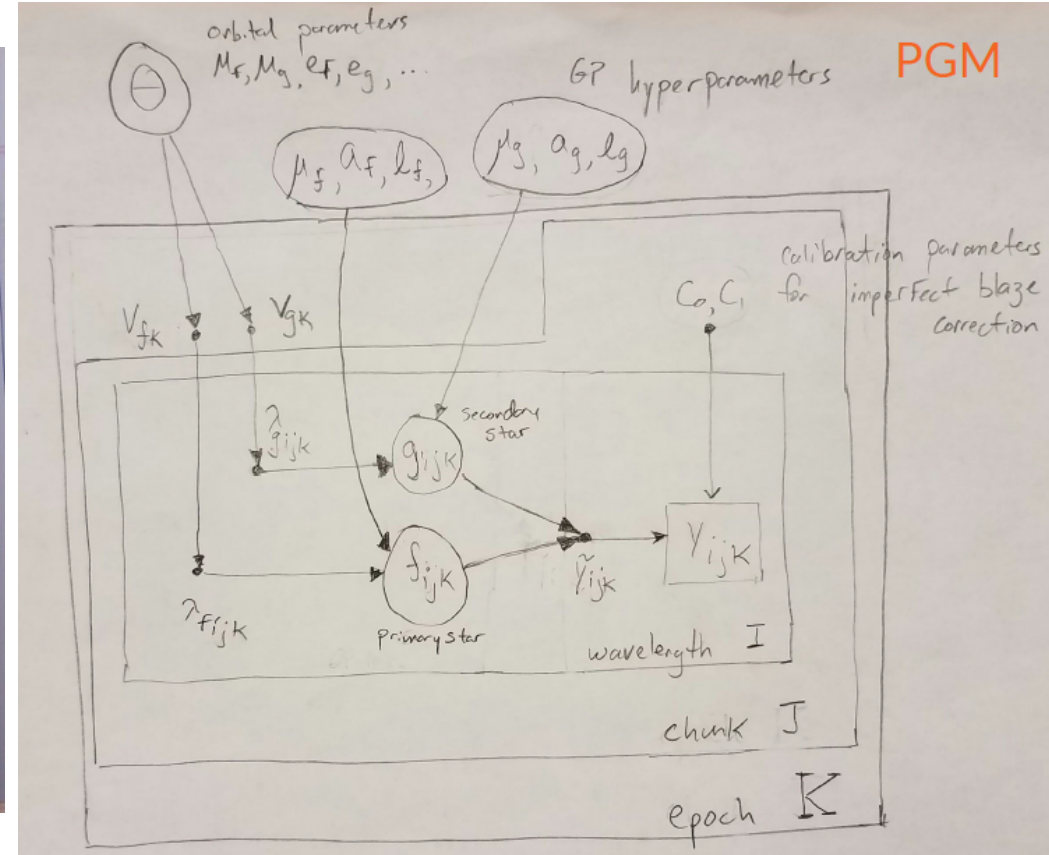
Ian Czekala, with exoplanet team

DAGs for hi-res spectra

Exoplanets



Stars



DAG ingredients

- Multivariate Gaussian processes
- Empirical basis functions: constrained PCA, sparse functional PCA...
- Diffusion maps

LSST connection: Modeling galaxy SEDs for photo-z