Astrostatistics
Part III Mathematical Tripos

CMS MR12, Tue, Thu, Sat 12pm

Dr Kaisey S Mandel
University Lecturer on Astrostatistics
Institute of Astronomy
(Kavli Institute for Cosmology, Office 03)
&
Statistical Laboratory
Department of Pure Mathematics & Mathematical Statistics
(CMS Pavilion D, Office 1.07)

kmandel@statslab.cam.ac.uk

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics

What is Astrostatistics?

- The application of statistics to astronomy, astrophysics & cosmology
- A research field: the interdisciplinary intersection of astronomy & statistics
- How do we properly interpret and analyse increasingly large and complex astronomical datasets?
- Developing and applying advanced statistical and computational methods to meet the unique challenges of astronomical data
- There is no “theory” of astrostatistics, the field is application-driven; we will focus on case studies

Scope & Goals

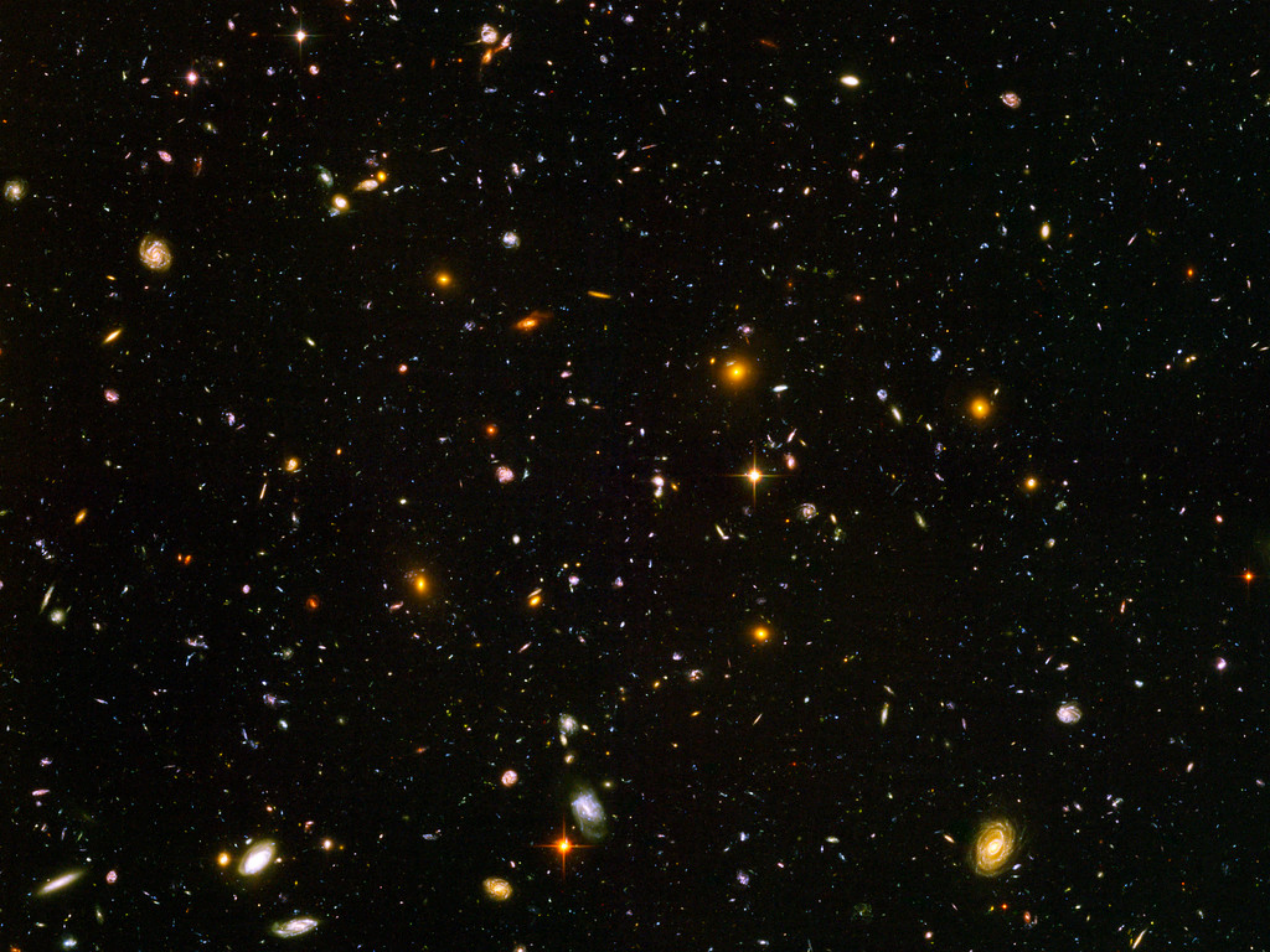
- Avoid thinking of statistics as a “bag of tricks”: with an ad-hoc recipe to run for each particular data analysis problem in astronomy
- It will be impossible to address every potential statistical task in astronomy in 8 weeks
- Instead, we will focus on general principles to help you think about how to analyse data in your specific cases. Where do the data come from? What is the model? What are the assumptions?
- Be pragmatic, and very applied. Examine real applications

For Astronomers

- Goal is to help you think critically about your data, rather than blindly applying canned black-box methods
- Understanding your statistical methods is crucial to interpreting their results. When can they go wrong?
- Often your data may be uniquely complex, and may require you to develop a data analysis method optimally suited to your inference problem - this is research!
- Astrostatistics is a creative endeavour!
- Get Jobs! (data-intensive astronomy, or industry)

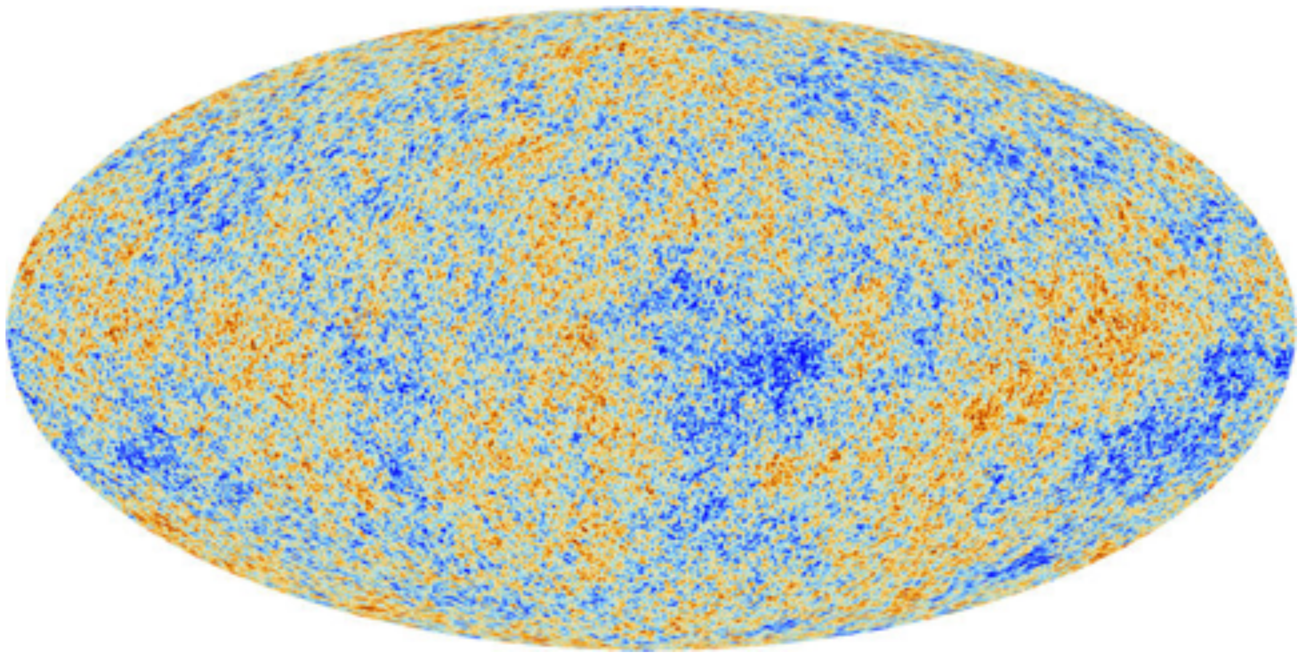
For Statisticians

- Astronomers have complex data sets with unique and challenging inference problems
- Astronomy is an observational science - (usually) no lab experiments
- Measurement errors, selection effects/biased samples, small sample size (One Universe), Big Data (billions of galaxies), combining heterogenous datasets over multiple wavelengths, EM/Gravitational Waves, multi-messenger astronomy
- Billions of $/£/Eur spent on Space Missions, how to get the most scientific value out of them? e.g. LSST, Euclid, WFIRST, TESS, …
- Optimal use of data requires developing, applying, and understanding new statistical approaches
- Play a critical role in answering Deep Questions about the Universe!
- Astronomy Jargon for Statisticians: http://hea-www.harvard.edu/AstroStat/astrojargon.html

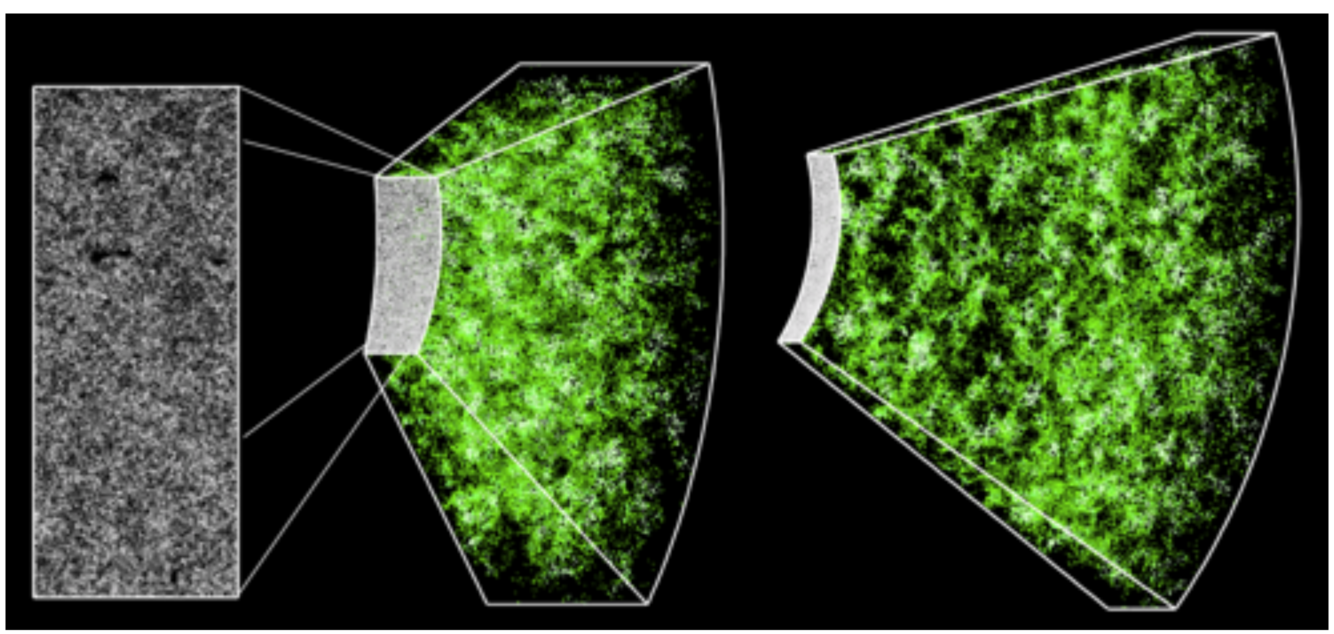


Cosmology

Planck CMB



Type Ia Supernovae



SDSS Baryonic Acoustic Oscillations

Scolnic et al. 2018

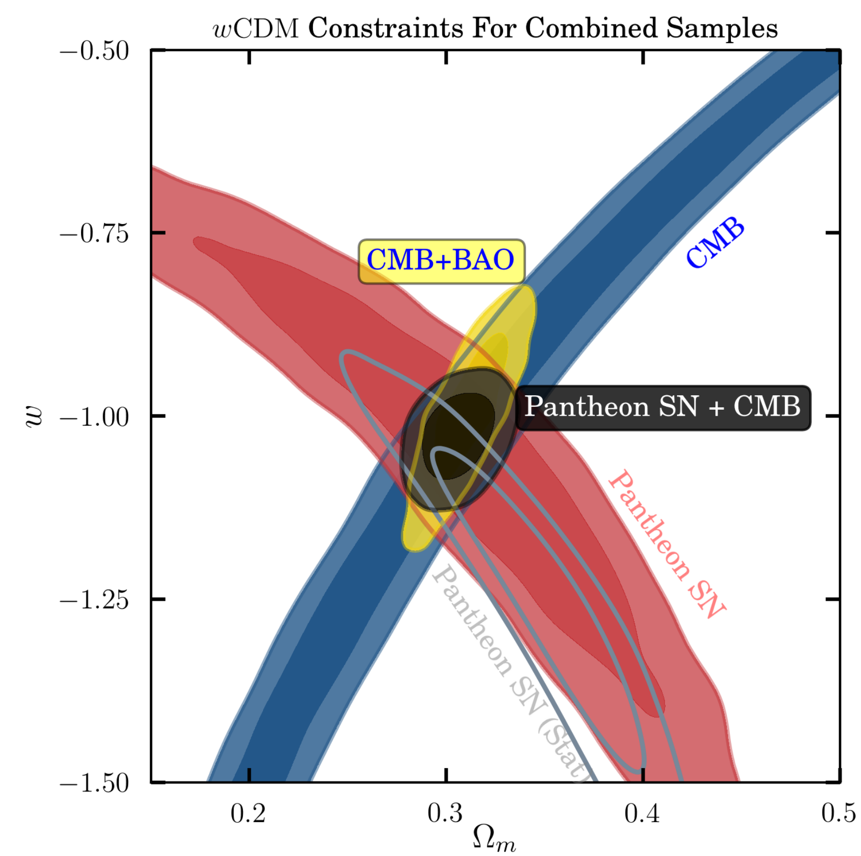


Figure 20. Confidence contours at 68% and 95% for the Ω_m and w cosmological parameters for the wCDM model. Constraints from CMB (blue), SN - with systematic uncertainties (red), SN - with only statistical uncertainties (gray-line), and SN+CMB (purple) are shown.

Extrasolar Planets

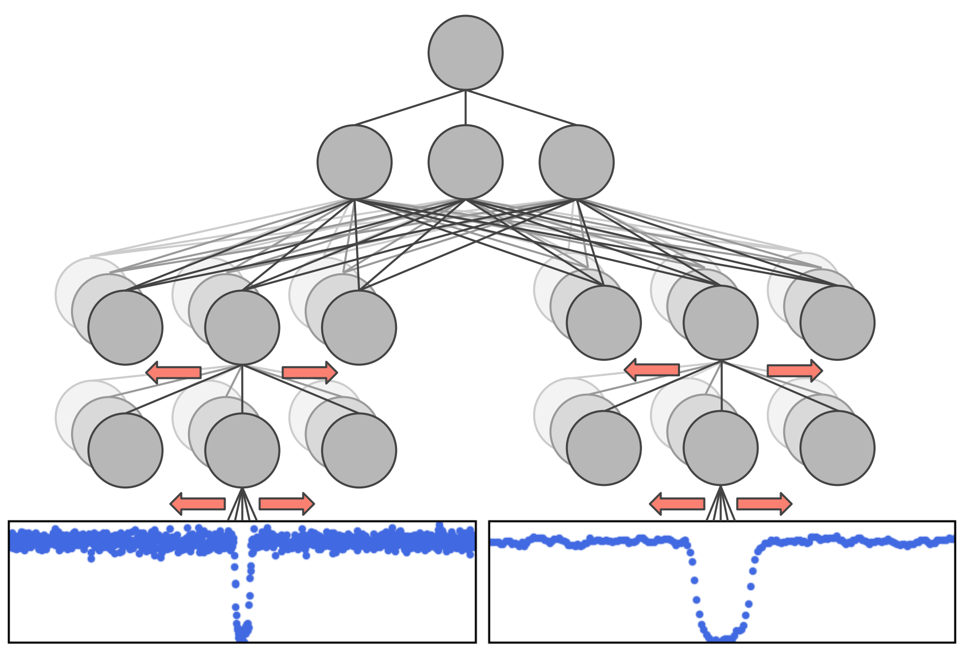


FIG. 5.— Convolutional neural network architecture for classifying light curves, with both global and local input views.

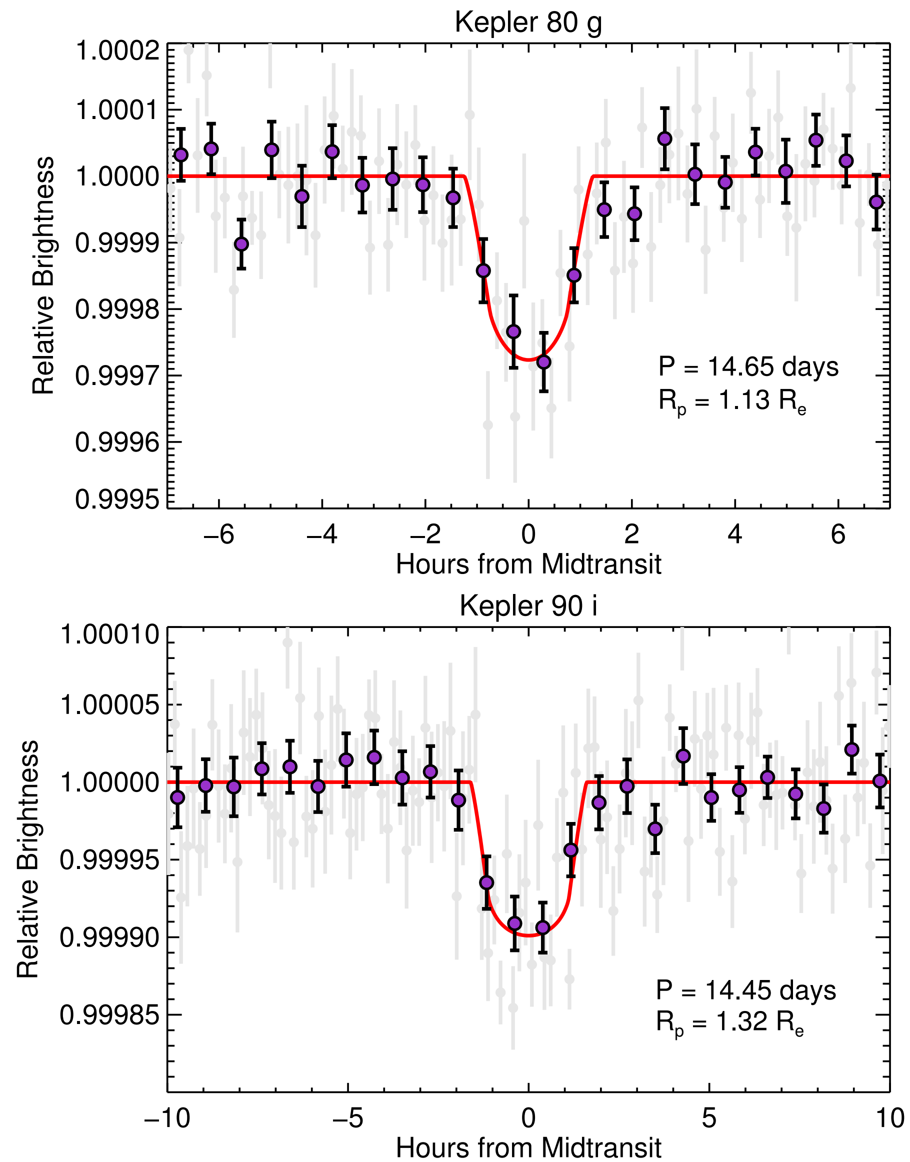
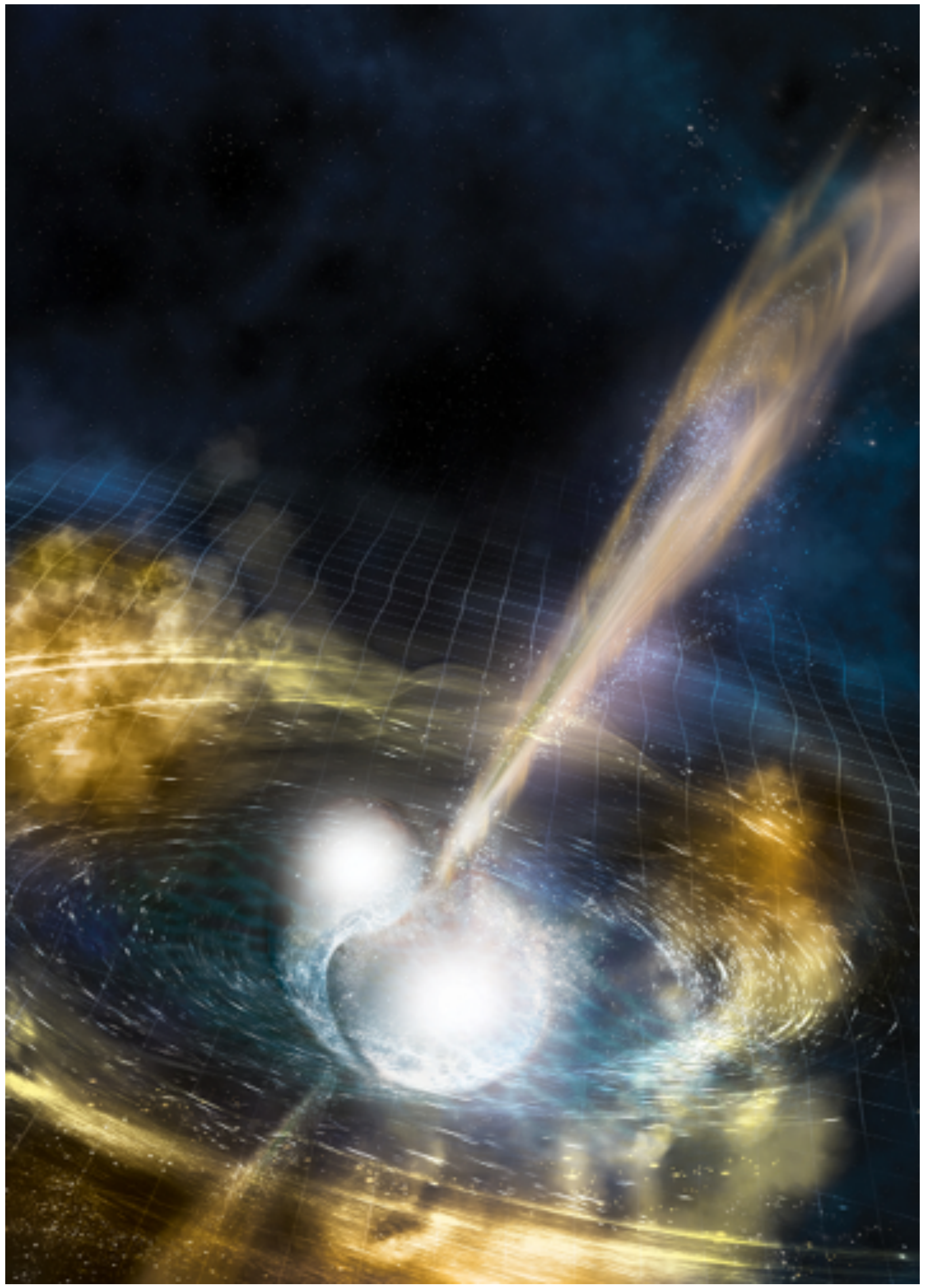


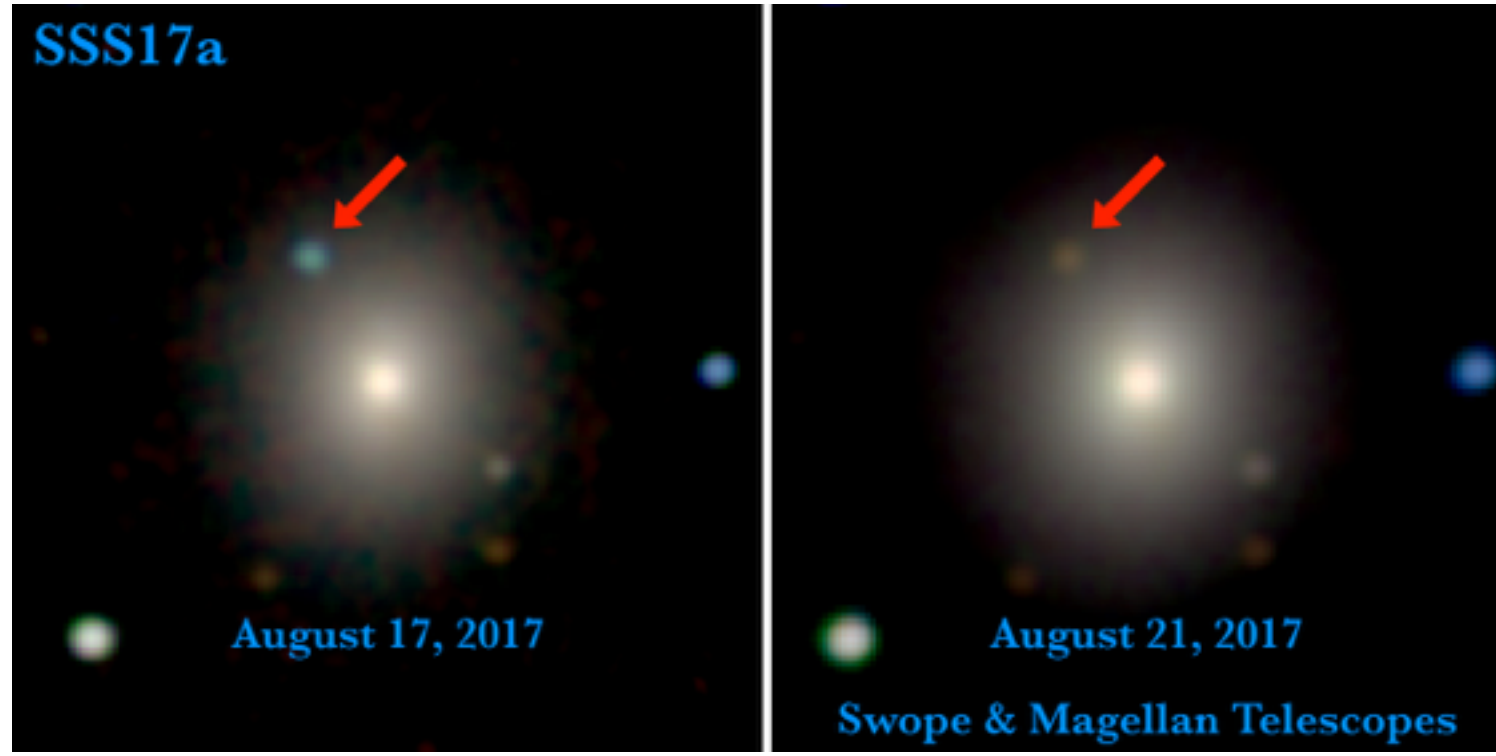
FIG. 12.— Transit light curves and best-fit models for the newly discovered planets around *Kepler*-80 and *Kepler*-90. In these plots, the grey points are robust averages of bins with width of approximately 10 minutes. The purple points are robust averages of bins with size about 1/4 the transit duration of the planet (bins of about 30 minutes for *Kepler*-80 g and about 45 minutes for *Kepler*-90 i.

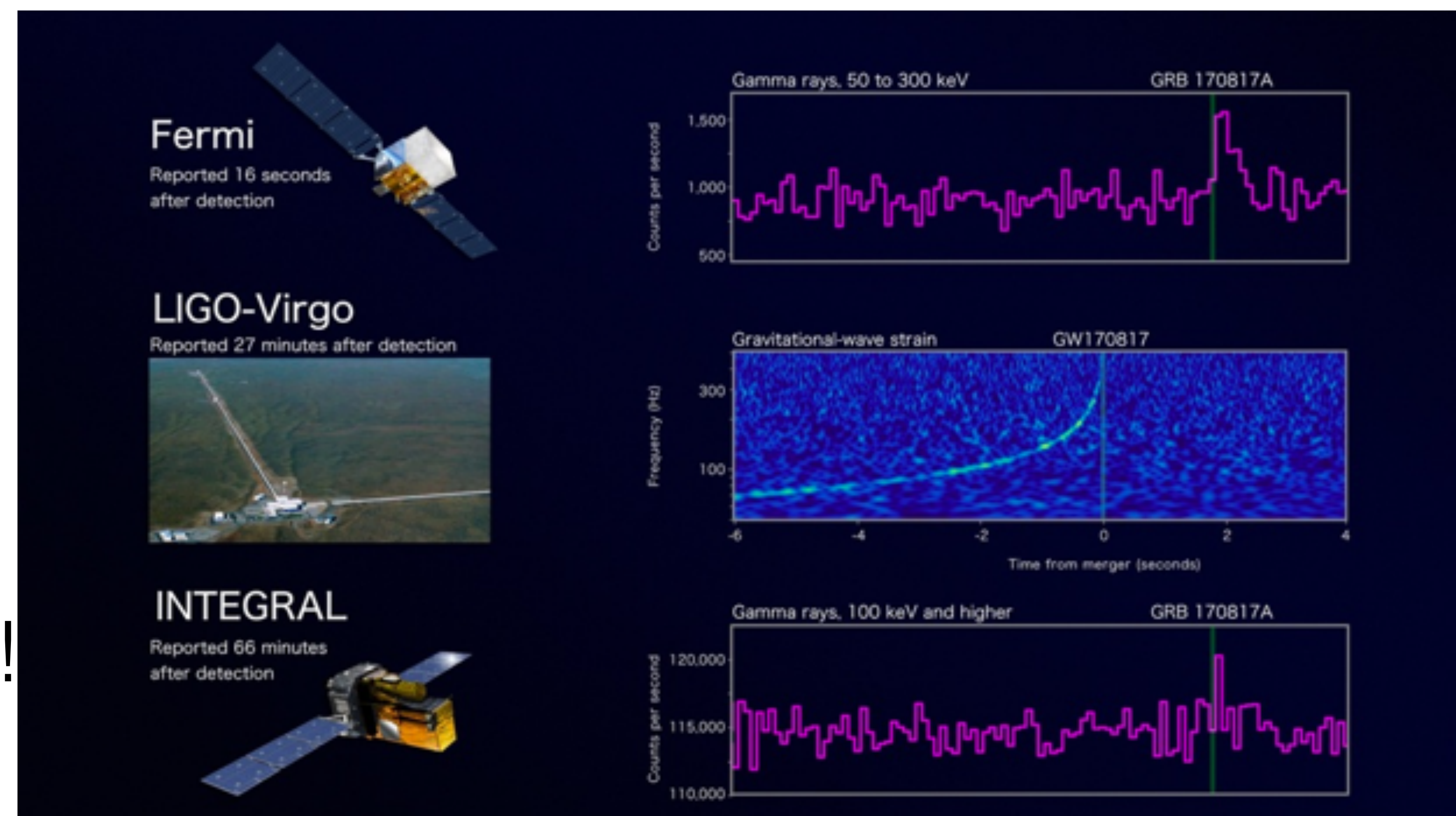
Deep Learning to Find New Exoplanets
(Shallue & Vanderberg 2017)

Gravitational Waves



Colliding Neutron Stars!





Intro to Astrostatistics

Short Articles

Roberto Trotta. **Astrostatistics is a field full of opportunities right now**
http://www.statisticsviews.com/details/feature/10741983/Astrostatistics-is-a-field-full-of-opportunities-right-now-An-interview-with-Rob.html

Long & de Souza. **Statistical methods in astronomy.**
https://arxiv.org/abs/1707.05834

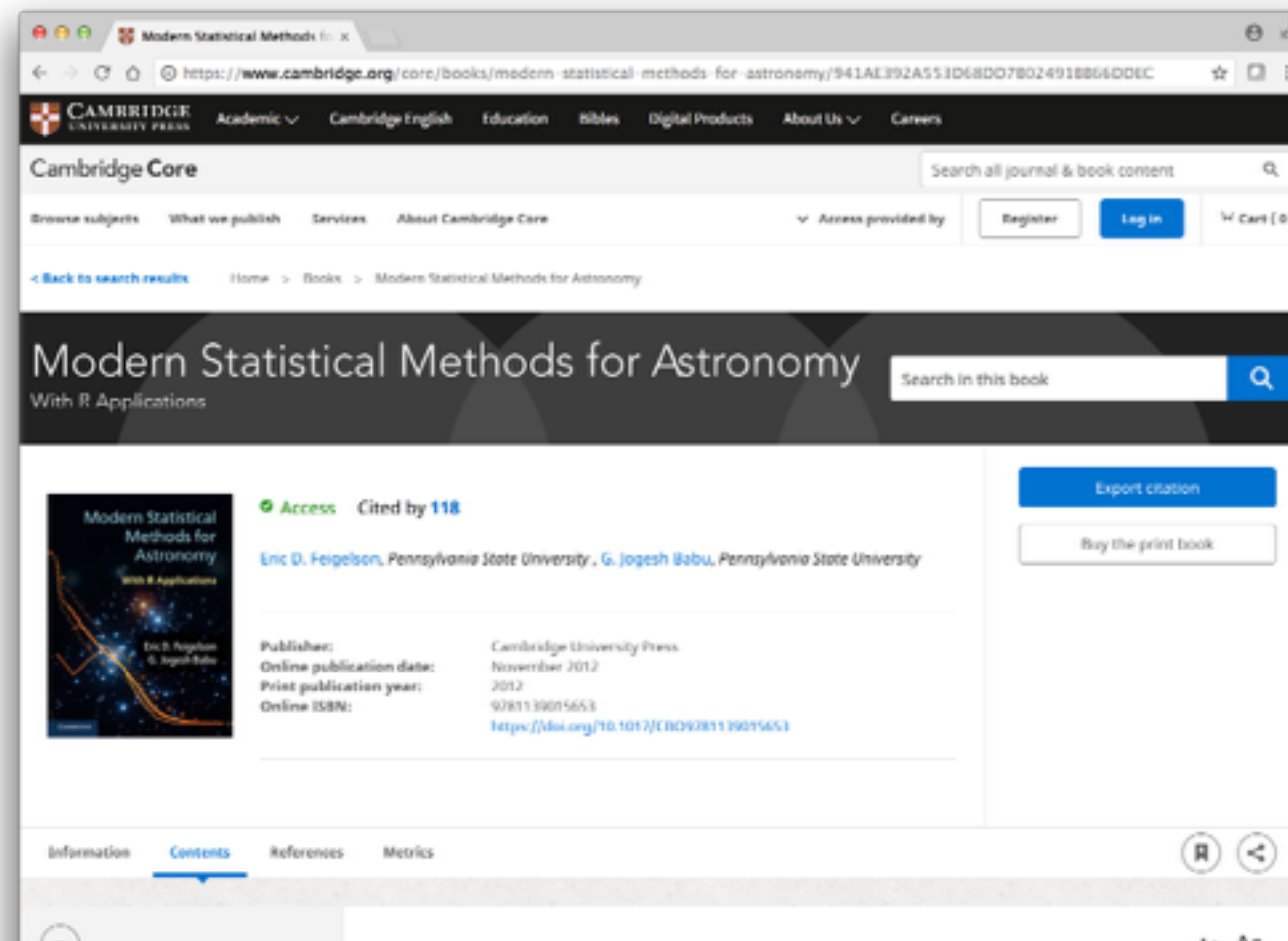
C. Schafer. **A Framework for Statistical Inference in Astrophysics.** 2015, Annual Review of Statistics and Its Application, 2: 141-162

Astrostatistics Texts

E. Feigelson and G. Babu. **Modern statistical methods for astronomy: with R applications**. CUP, 2012.
[Free Cambridge access: search at
https://www.cambridge.org/core/ or
http://idiscover.lib.cam.ac.uk/ or go to the library]

An overview of statistical
Methods for astronomers.
R code available

Recommended Reading:
Chapters 1-2
Intro to Statistics in Astronomy
Review of Probability

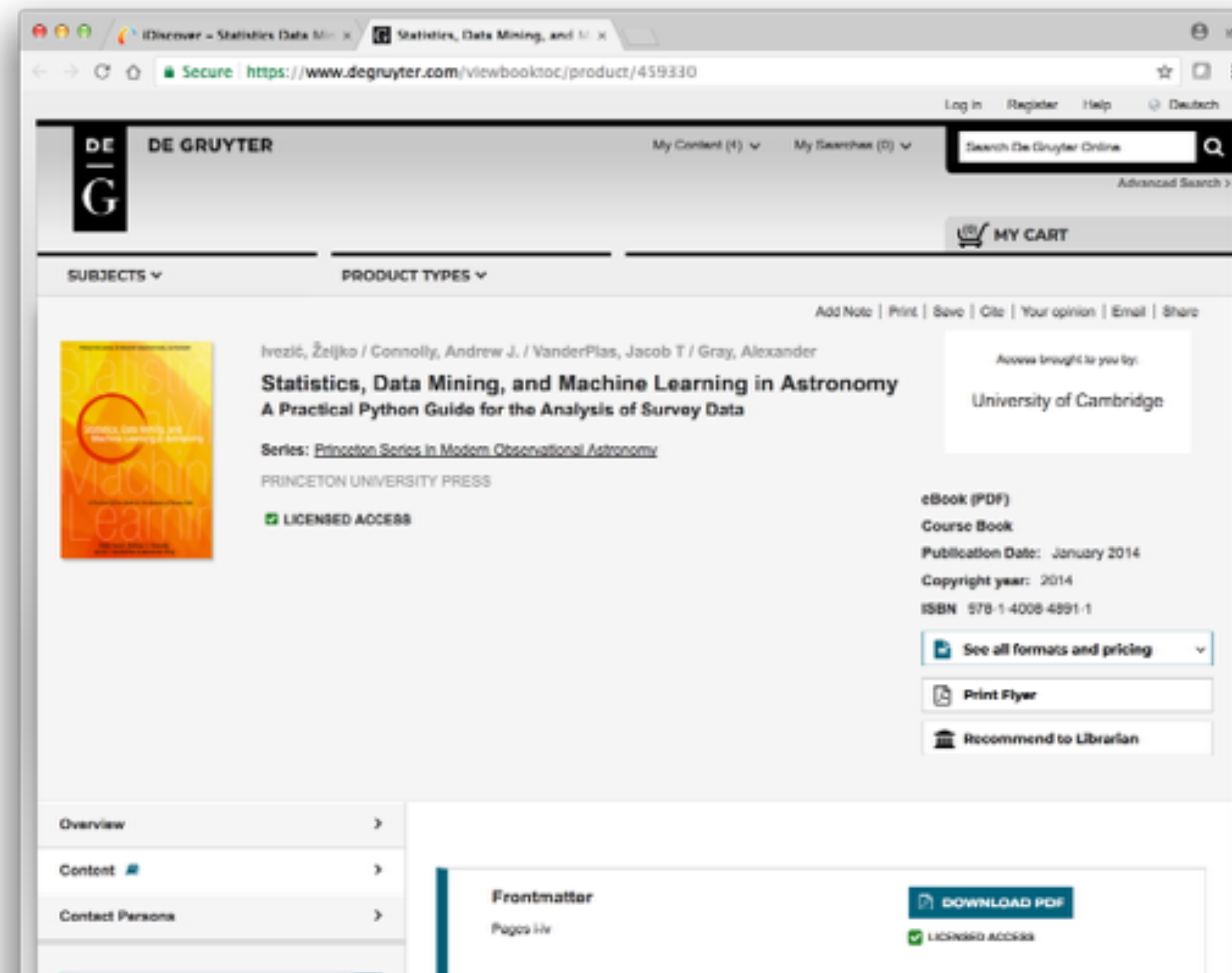


Astrostatistics Texts

Z. Ivezic et al. **Statistics, Data Mining, and Machine Learning in Astronomy**. PUP, 2014.
[Cambridge Library Online Access: Search at http://idiscover.lib.cam.ac.uk/, hard copies also available in library]

A Machine Learning bent. Python package AstroML and datasets to play with. Try it!

Recommended Reading:
Chapters 1-3
Introduction and Basic review of Probability & Statistics



Statistics & Machine Learning

Going Deeper…

Gelman et al. **Bayesian Data Analysis**, 3rd Edition, 2013
[Hard copy in library]

Bishop et al. **Pattern Recognition and Machine Learning**, 2006
[Hard copy in Moore Library, on reserve]

MacKay, D. **Information Theory, Inference, and Learning Algorithms**, 2003
http://www.inference.org.uk/itila/ [FREE online]

Hastie, Tibshirani and Friedman. **The Elements of Statistical Learning** (2nd Ed), 2009.
https://web.stanford.edu/~hastie/ElemStatLearn/ [FREE online]

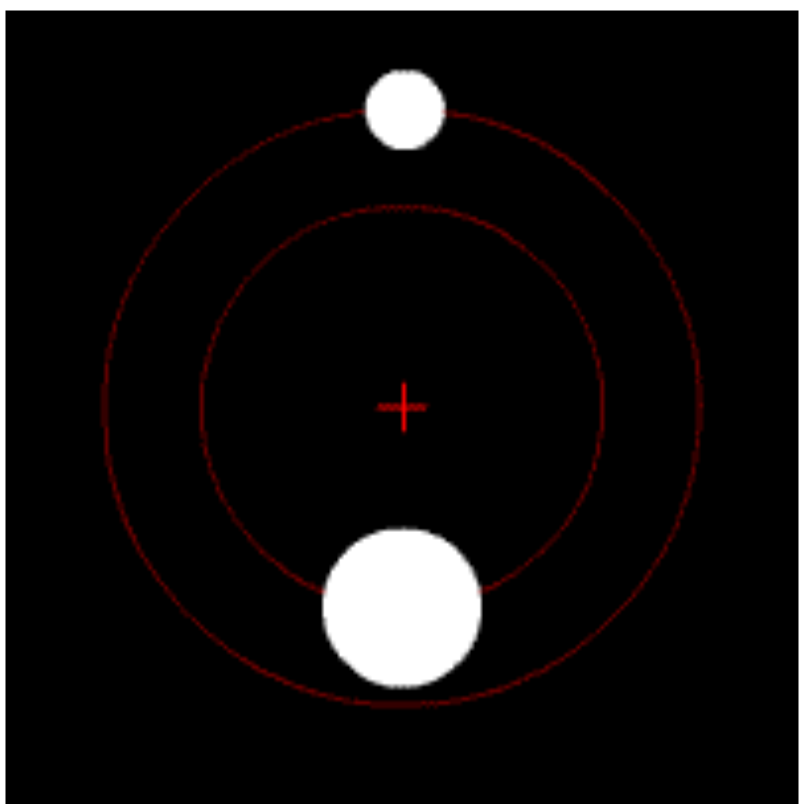
Rasmussen & Williams. **Gaussian Processes for Machine Learning**, 2006
http://www.gaussianprocess.org/gpml/ [FREE online]

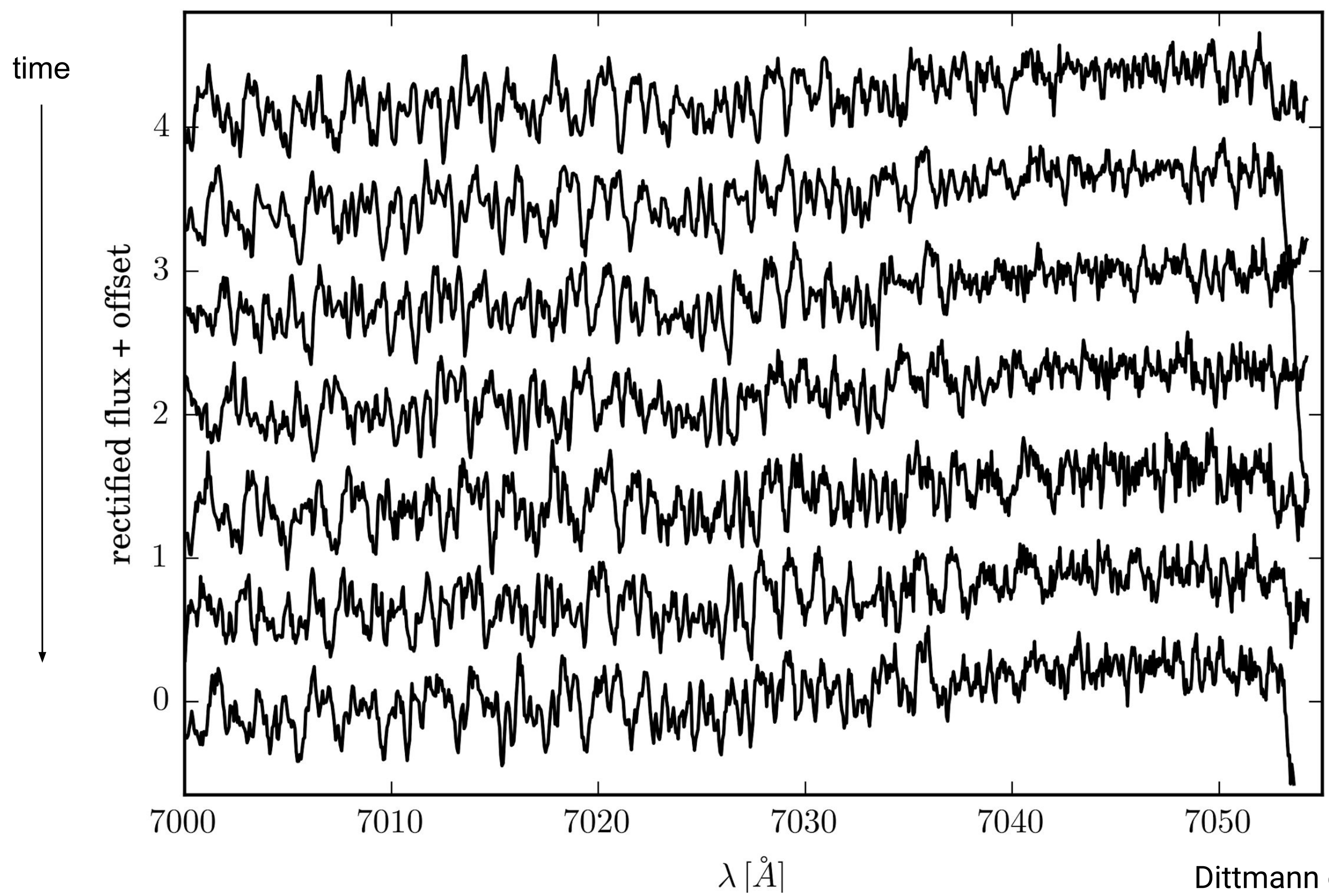
Topics

- Preliminaries / Statistics & Astronomy Background
- Regression / Fitting Models to Astronomical Data
- Generative / Forward Modeling
- Bayesian Inference
- Gaussian Processes / Nonparametric Bayes
- Time Series Analysis
- Hierarchical Bayesian Modeling
- Probabilistic Graphical Models
- Statistical Computation:
  - Markov Chain Monte Carlo,
  - Approximate Bayesian Computation (ABC)
  - Nested Sampling
- Model Selection
- Machine Learning / Classification

Astrostatistics Case Studies:
Disentangling Time Series Spectra with Gaussian Processes: Applications to Radial Velocity Analysis (Czekala et al. 2017)

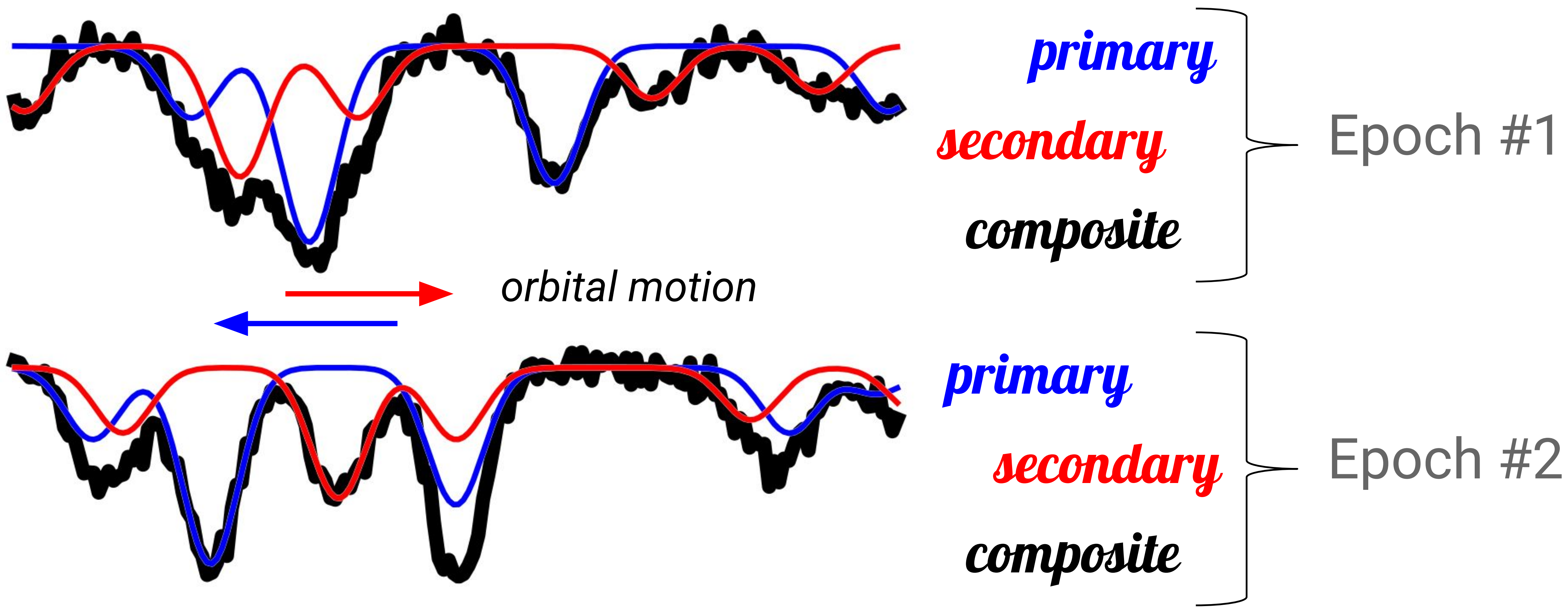
Raw Observations of the LP661-13 M4 Binary



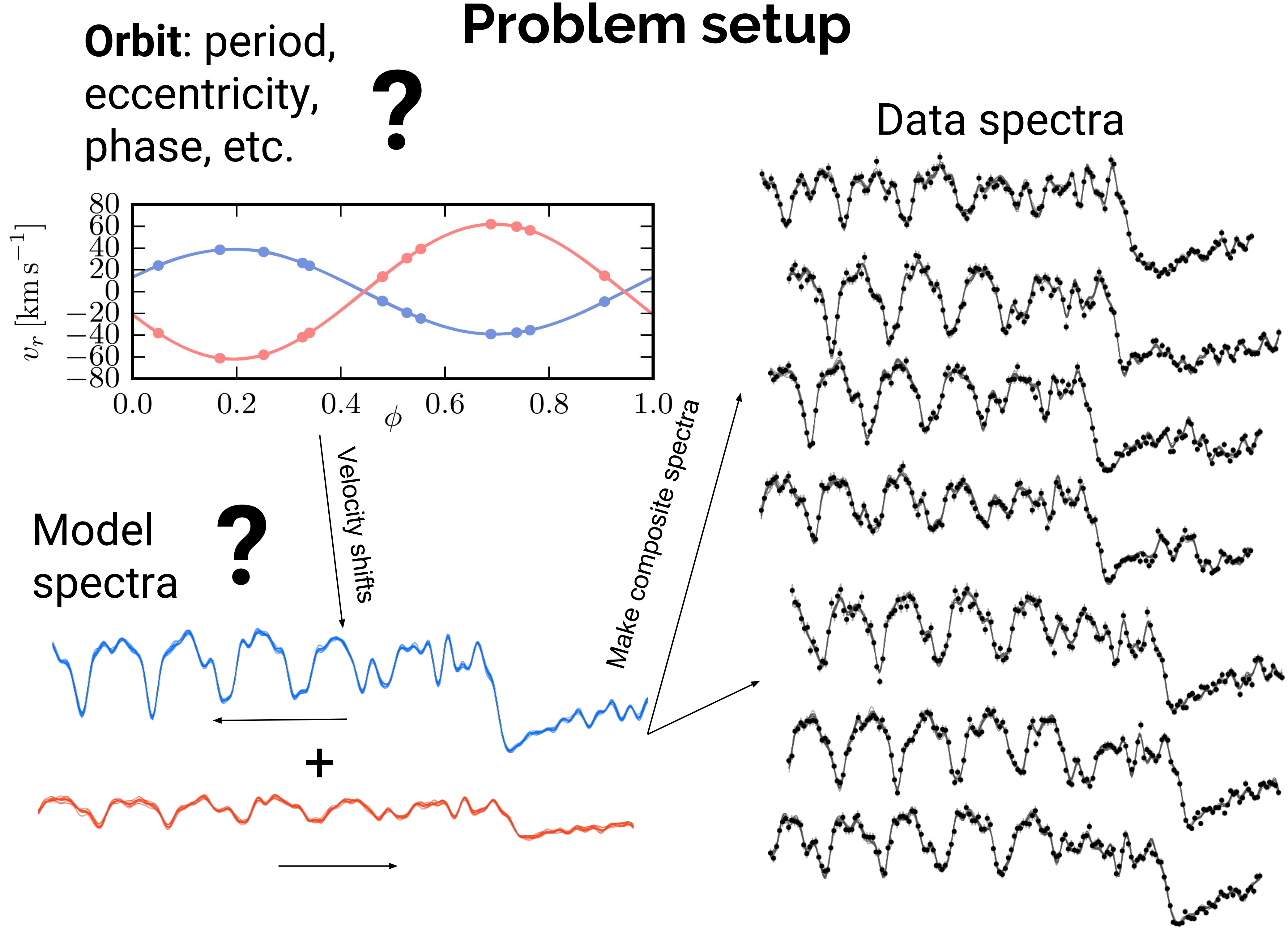


Dittmann et al. 17
Czekala et al. 17a

Spectroscopic Binary Stars



Problem setup



https://www.youtube.com/watch?v=kHjN42ft6aU

Goal: Go Backwards and Infer the Component Spectra & Orbital Parameters from noisy, observed (composite) spectra time series

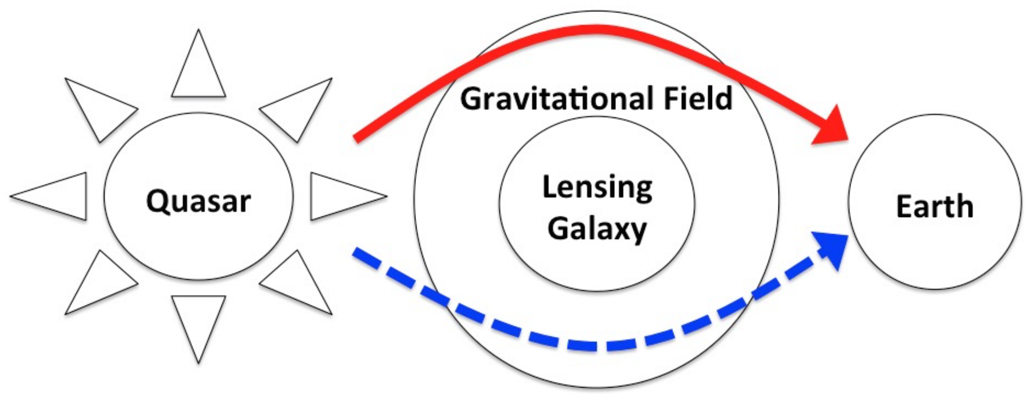
Astrostatistics Case Study 1: Disentangling Time Series Spectra with Gaussian Processes: Applications to Radial Velocity Analysis (Czekala et al. 2017, arXiv:1702.05652)

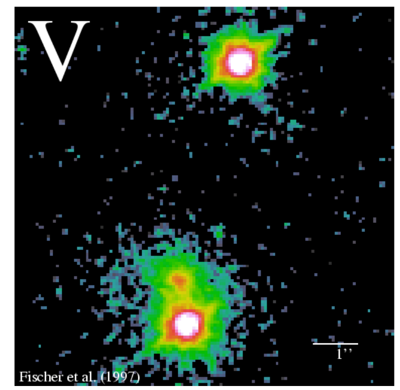
http://psoap.readthedocs.io/en/latest/

- Statistics:
  - Parametric Modelling (Orbit)
  - Nonparametric Modelling (Gaussian Process Spectrum)
  - Bayesian Inference
  - Markov Chain Monte Carlo
- Astronomy:
  - Applications to Radial Velocity Analysis of Stars/Exoplanets

Astrostatistics Case Study 2:

Bayesian Estimates of Astronomical Time
Delays Between Gravitationally Lensed Stochastic Light Curves
(Tak et al. 2017, Annals of Applied Statistics, arXiv:1602.01462)





Estimating time delays between noisy, irregularly sampled, gappy astronomical time series —> determine expansion rate of Universe (H_0)

- Bayesian Inference
- Stochastic Processes
- MCMC
- Gibbs Sampling

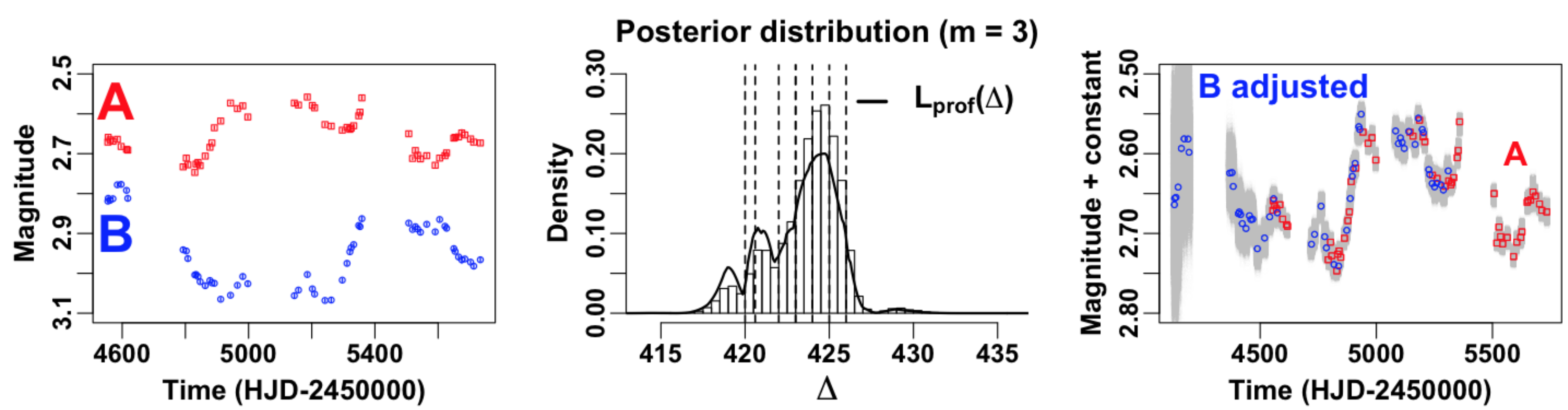
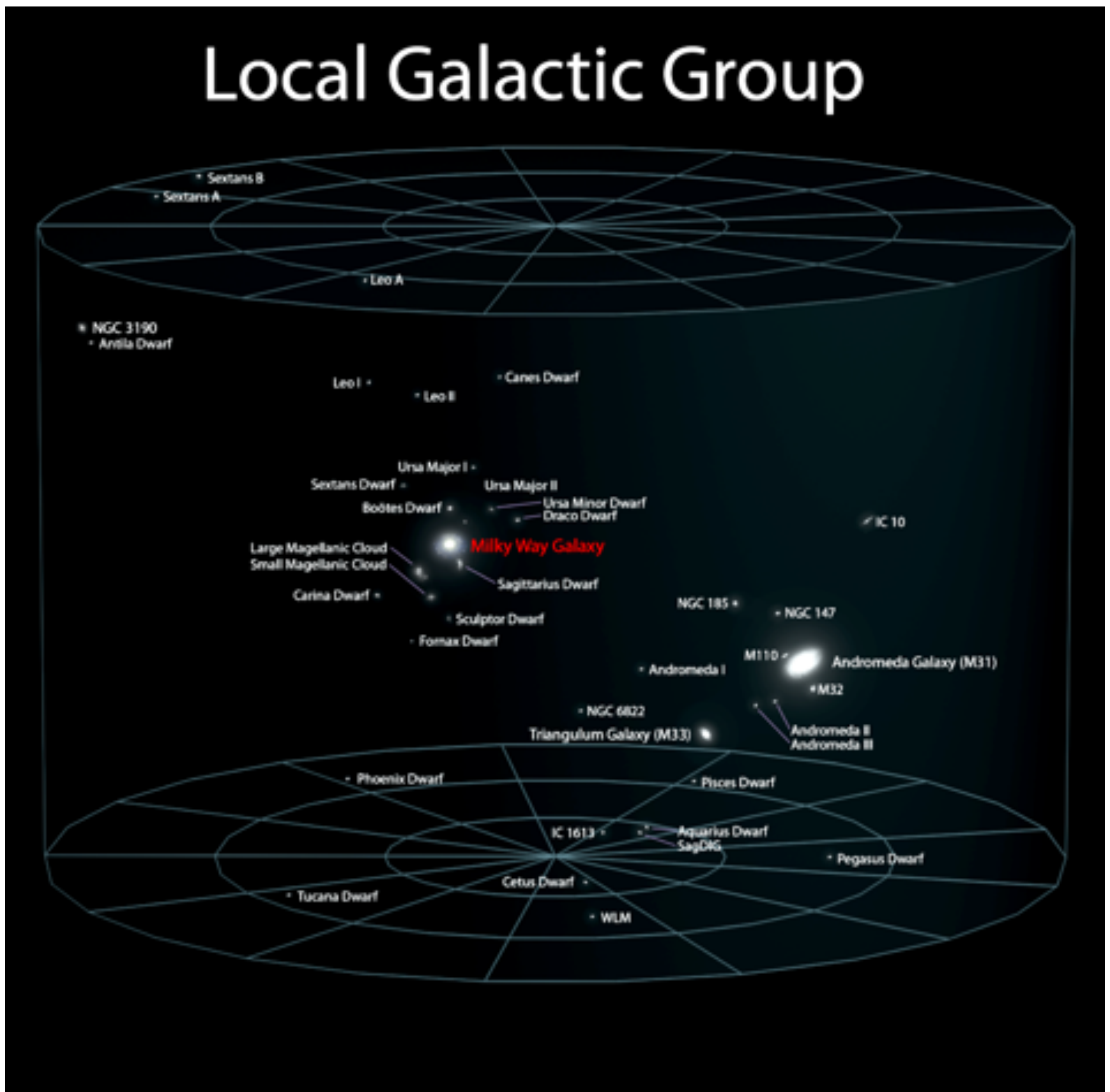


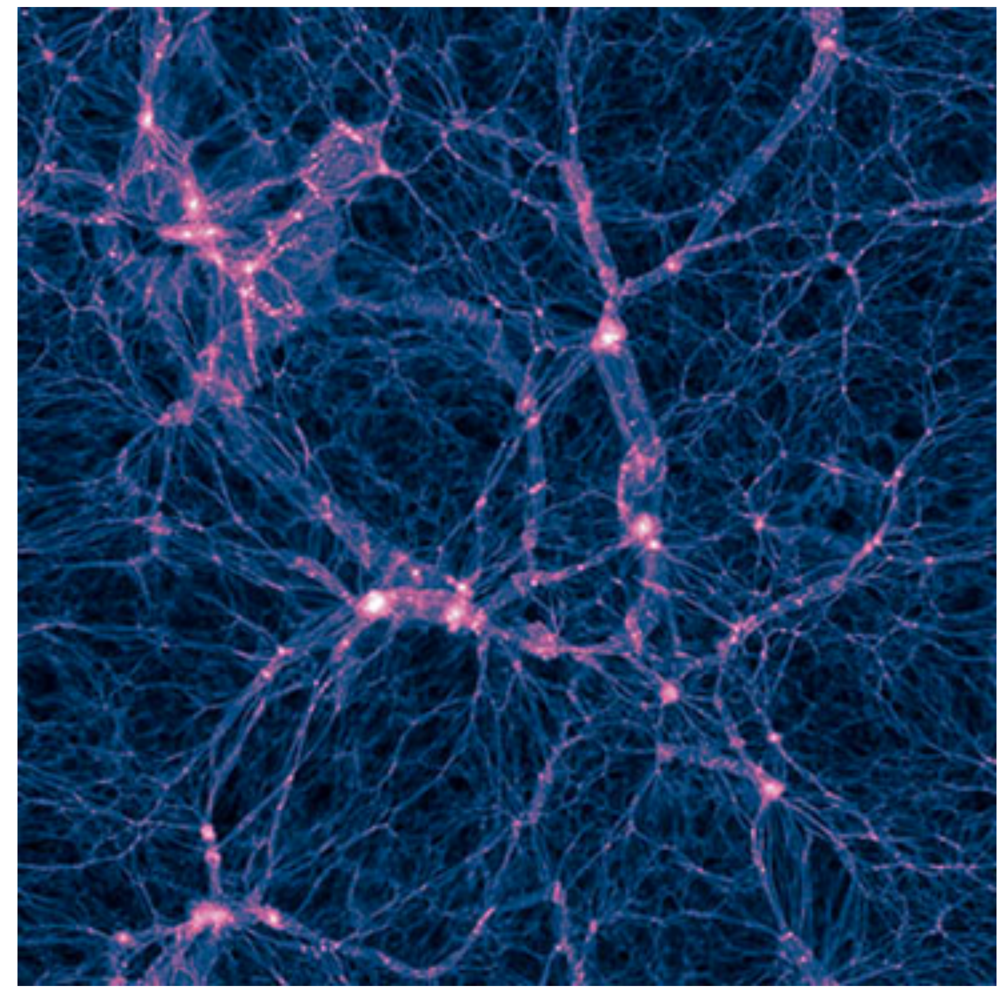
FIG 13. *Observations of Quasar Q0957+561 from Hainline et al. (2012) are plotted in the first panel. The second panel exhibits the marginal posterior distribution of* Δ *with*

Astrostatistics Case Study 3:

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations

(Patel et al. 2017, arXiv:1703.05767)



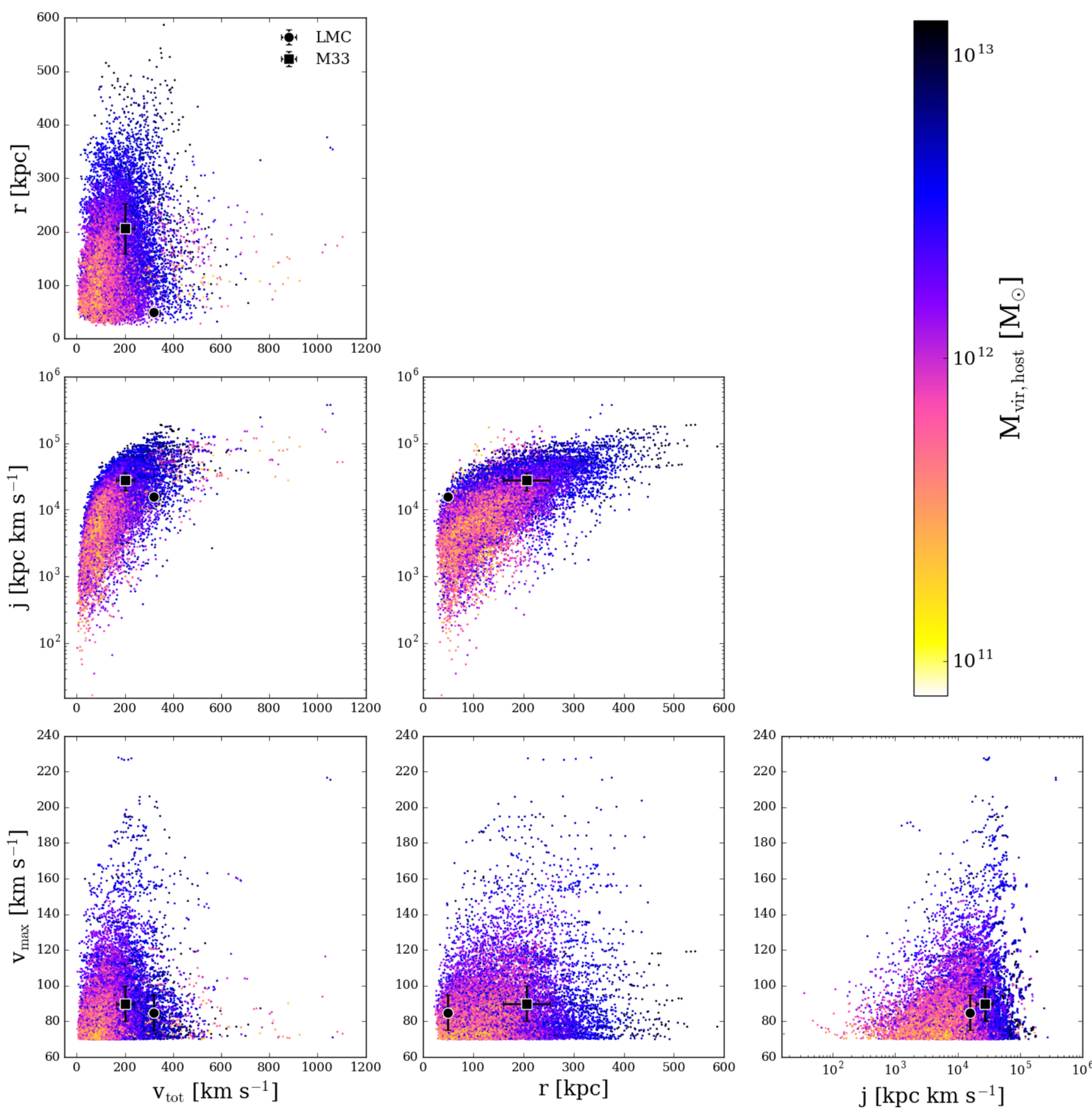


Illustris

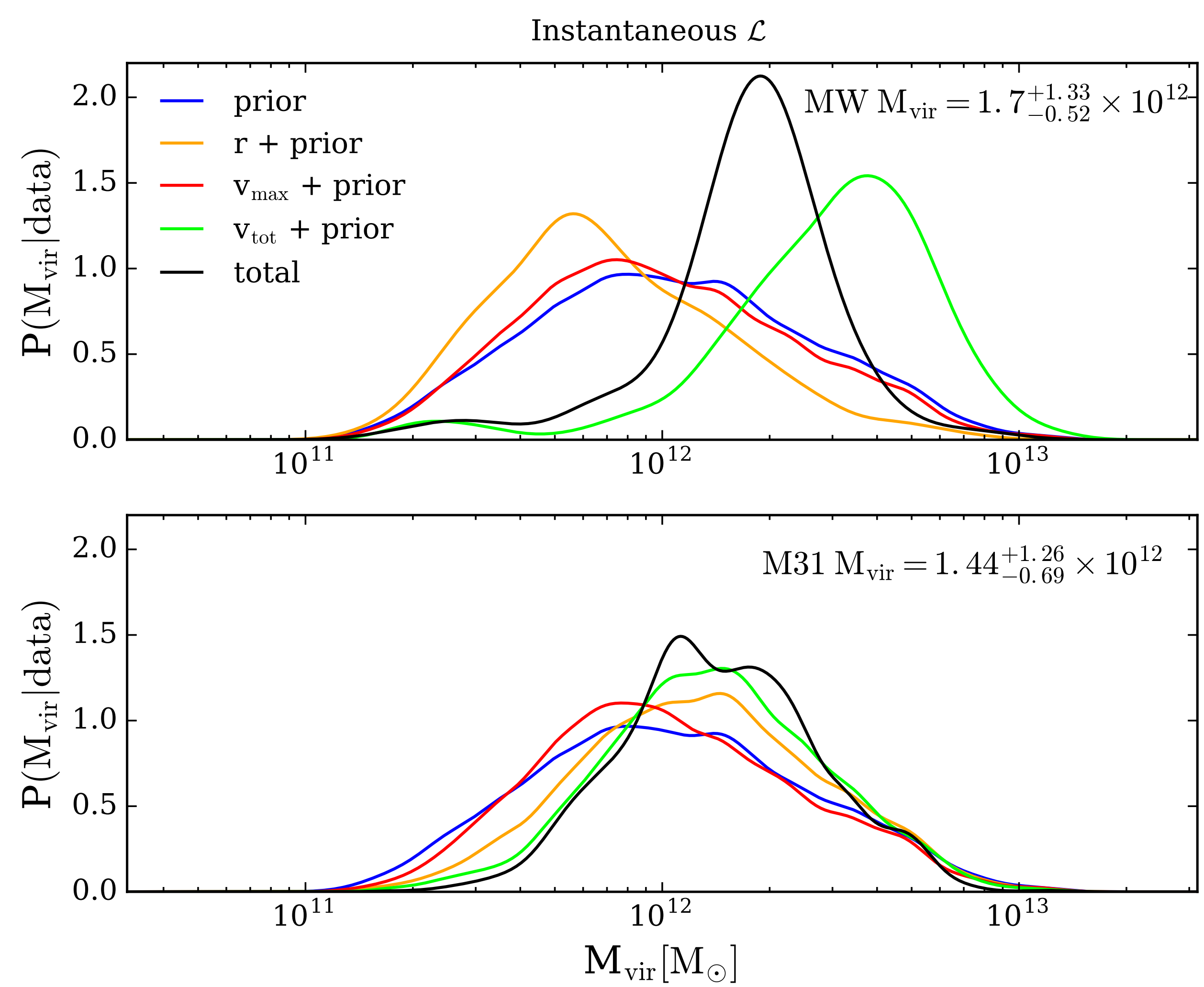
Cosmological Simulation of Galaxy Formation

Astrostatistics Case Study 3:

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations (Patel et al. 2017, arXiv:1703.05767)



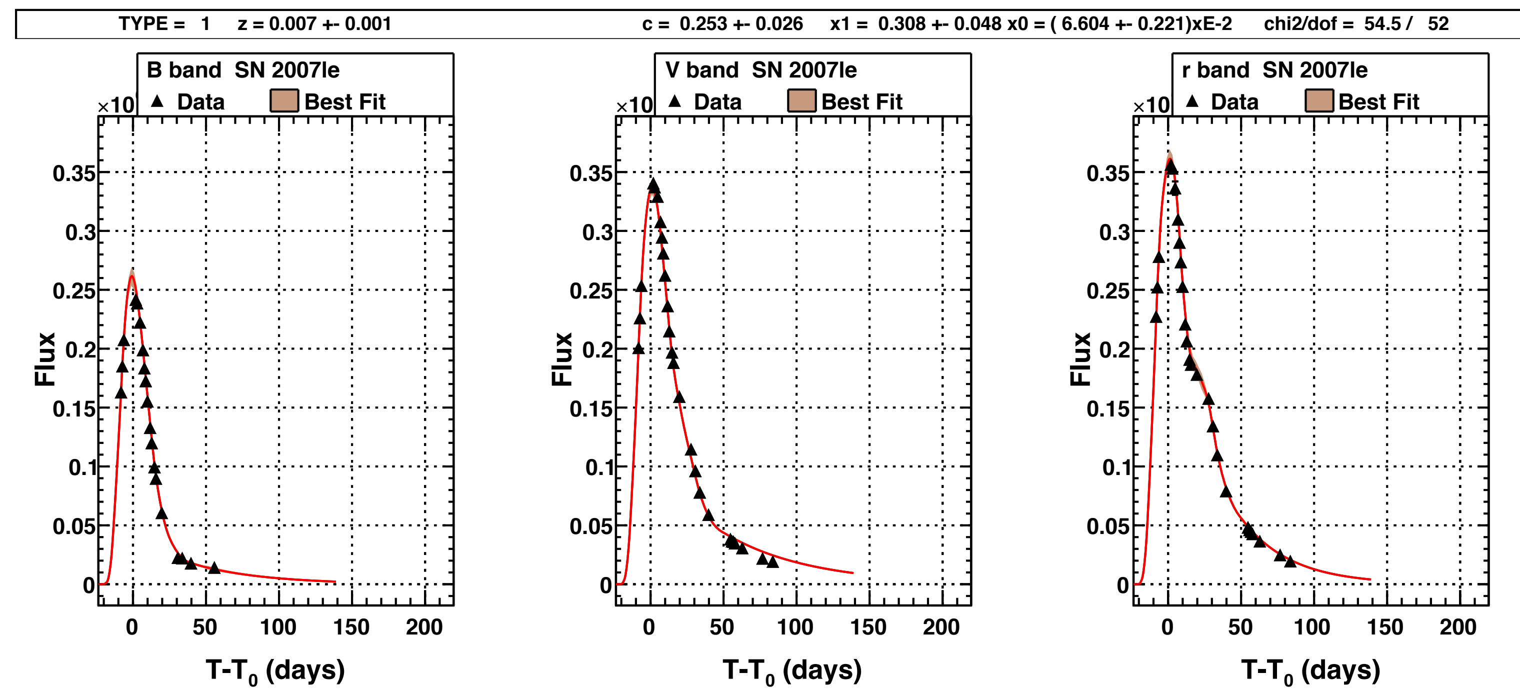
Simulation —> Prior

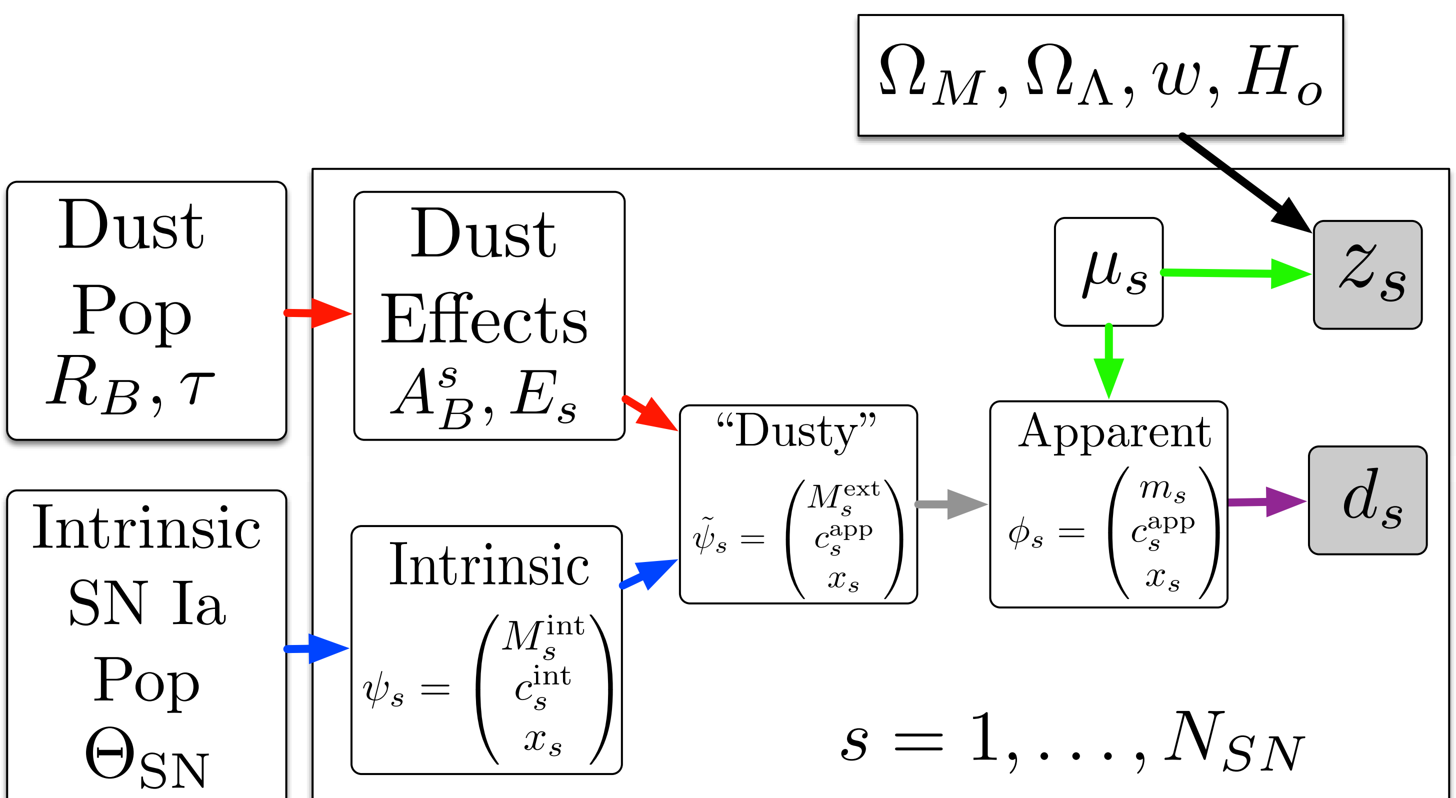


- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

Astrostatistics Case Study 4:

Hierarchical Bayesian Models for Supernovae (Mandel et al. 2017)





- Hierarchical Bayes / Multilevel Models
- Probabilistic Graphical Models
- Latent Variable Modelling
- Time Series Analysis
- MCMC