Astrostatistics

Saturday, 20 January 2017

Recommended Reading:
Feigelson & Babu: Chapters 1-4
Ivezic: Chapters 1-5
Schafer article

Intro to Statistics in Astronomy
Review of Probability & Statistics Foundations
Classical & Bayesian Statistical Inference

kmandel@statslab.cam.ac.uk

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

Today

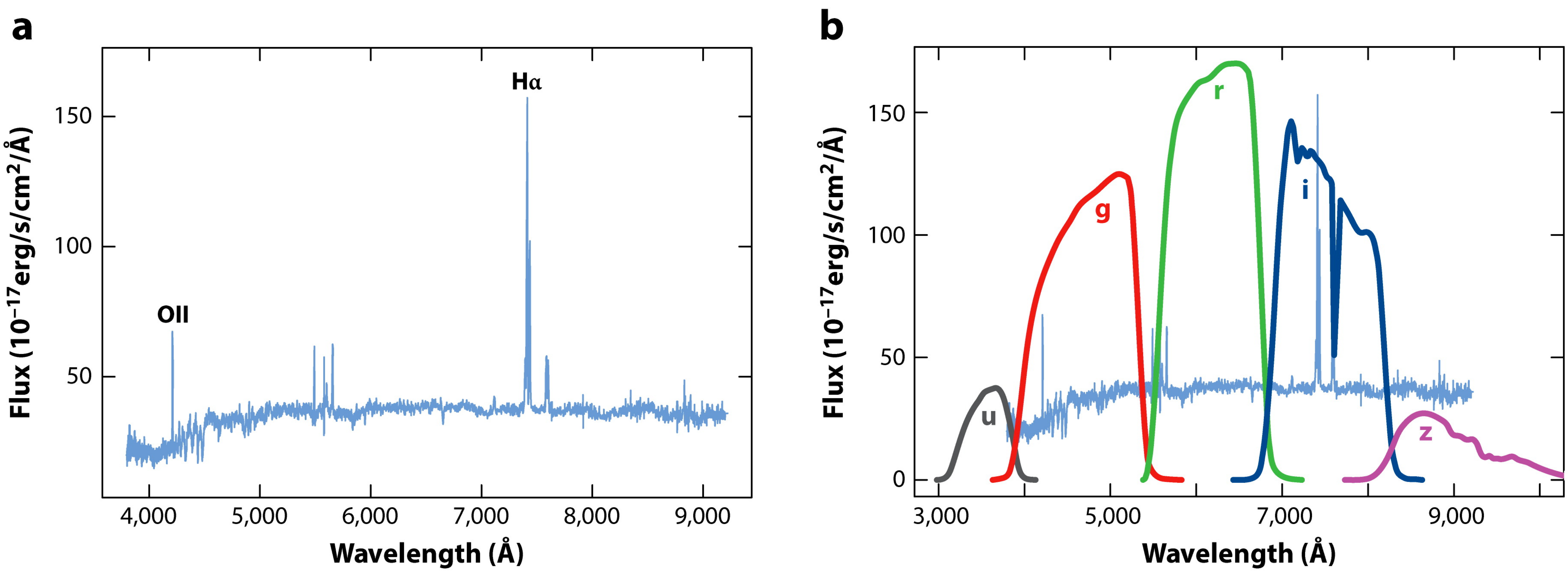
- Introduction to Astronomical Data Types (for statisticians)

- Continue Motivation: Radial Velocity Case Study

What astronomers measure

- Astrometry (angular position on sky, e.g. Gaia)
- Photometry (how bright is it?)
- Spectroscopy (brightness versus wavelength)
- Time Series: Transients & Variables (e.g. stars, quasars, supernovae, exoplanets), Moving objects (e.g. asteroids)
- Spatial Variation (images, maps)
- Combinations of the above

Spectroscopy and Photometry



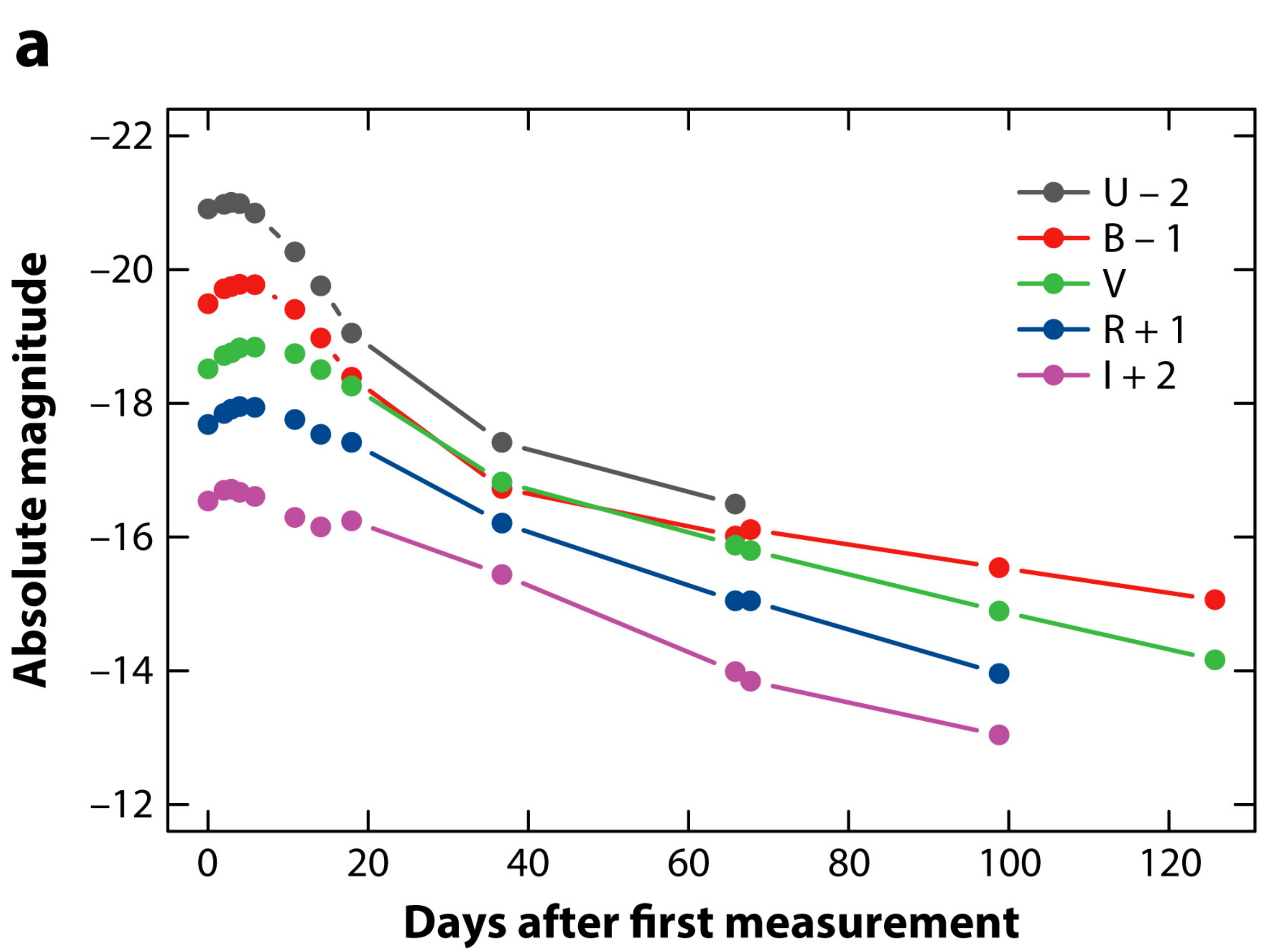
Schafer CM. 2015.
Annu. Rev. Stat. Appl. 2:141–62

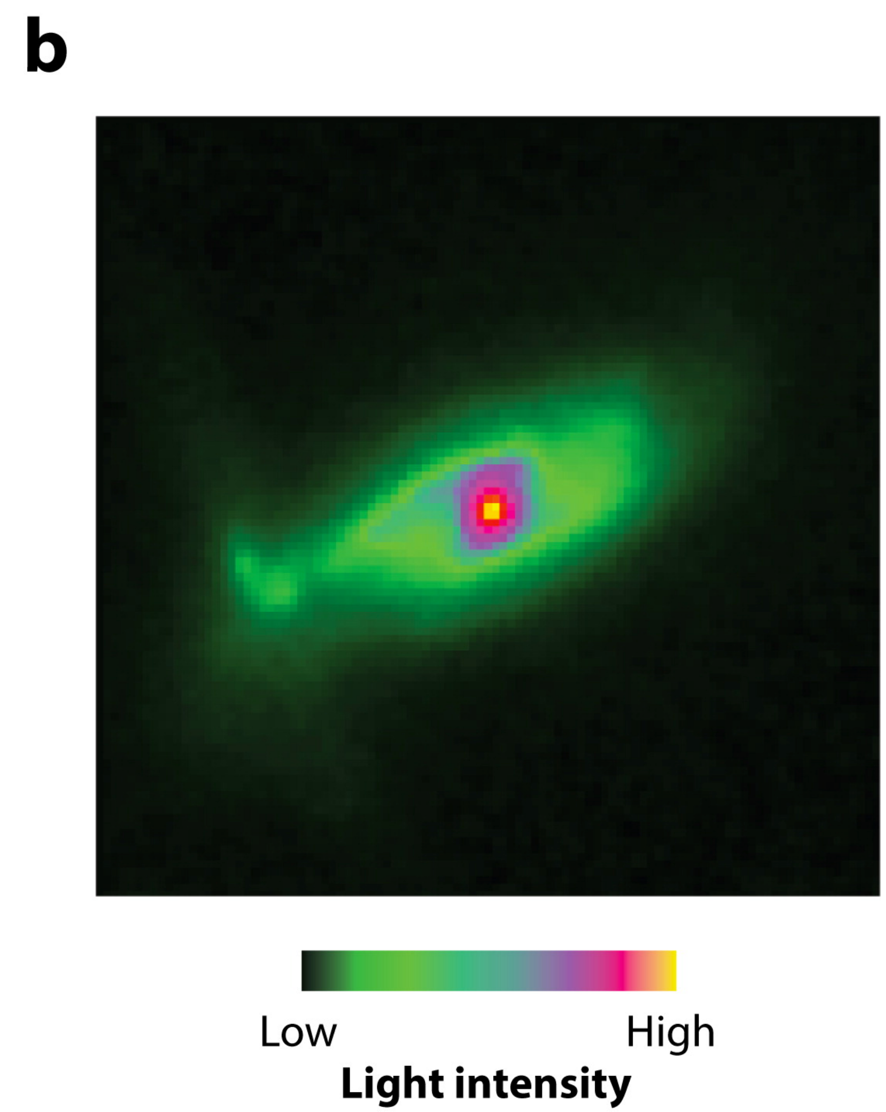
Galaxy Spectrum

Galaxy Photometry

(Flux / Magnitude)

Temporal & Spatial Variation



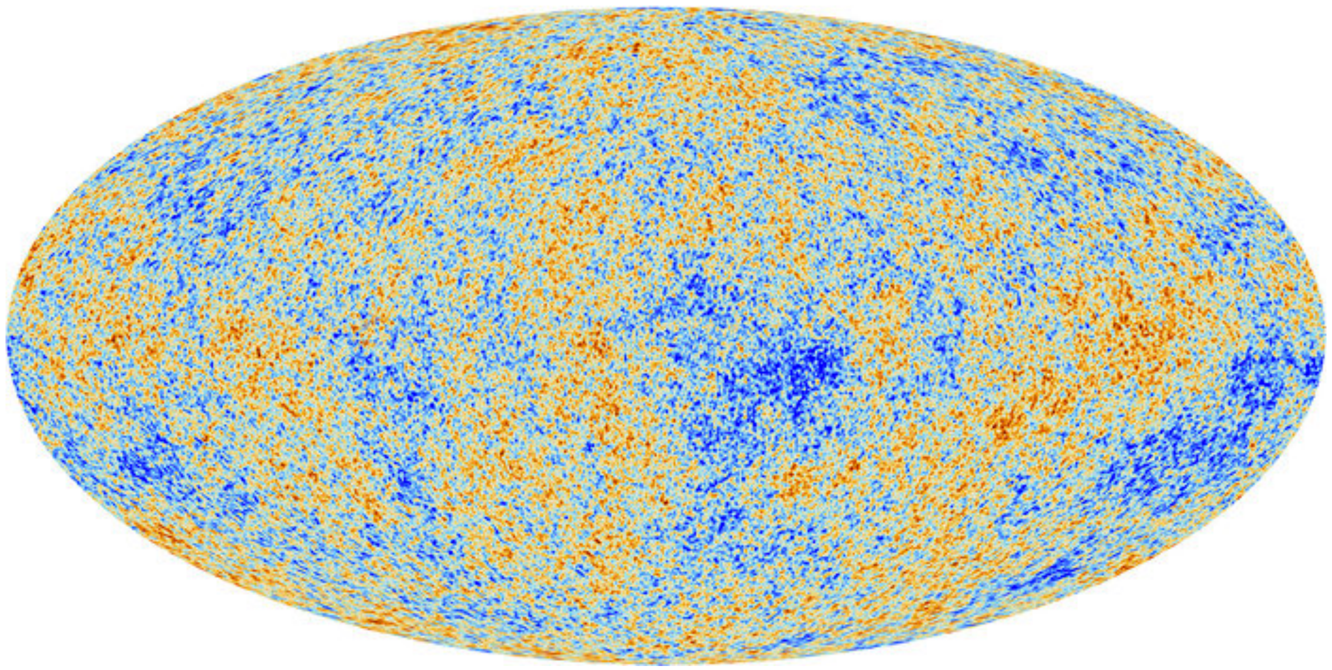


Schafer CM. 2015.
Annu. Rev. Stat. Appl. 2:141–62

Time Series (Light Curve)
Supernova

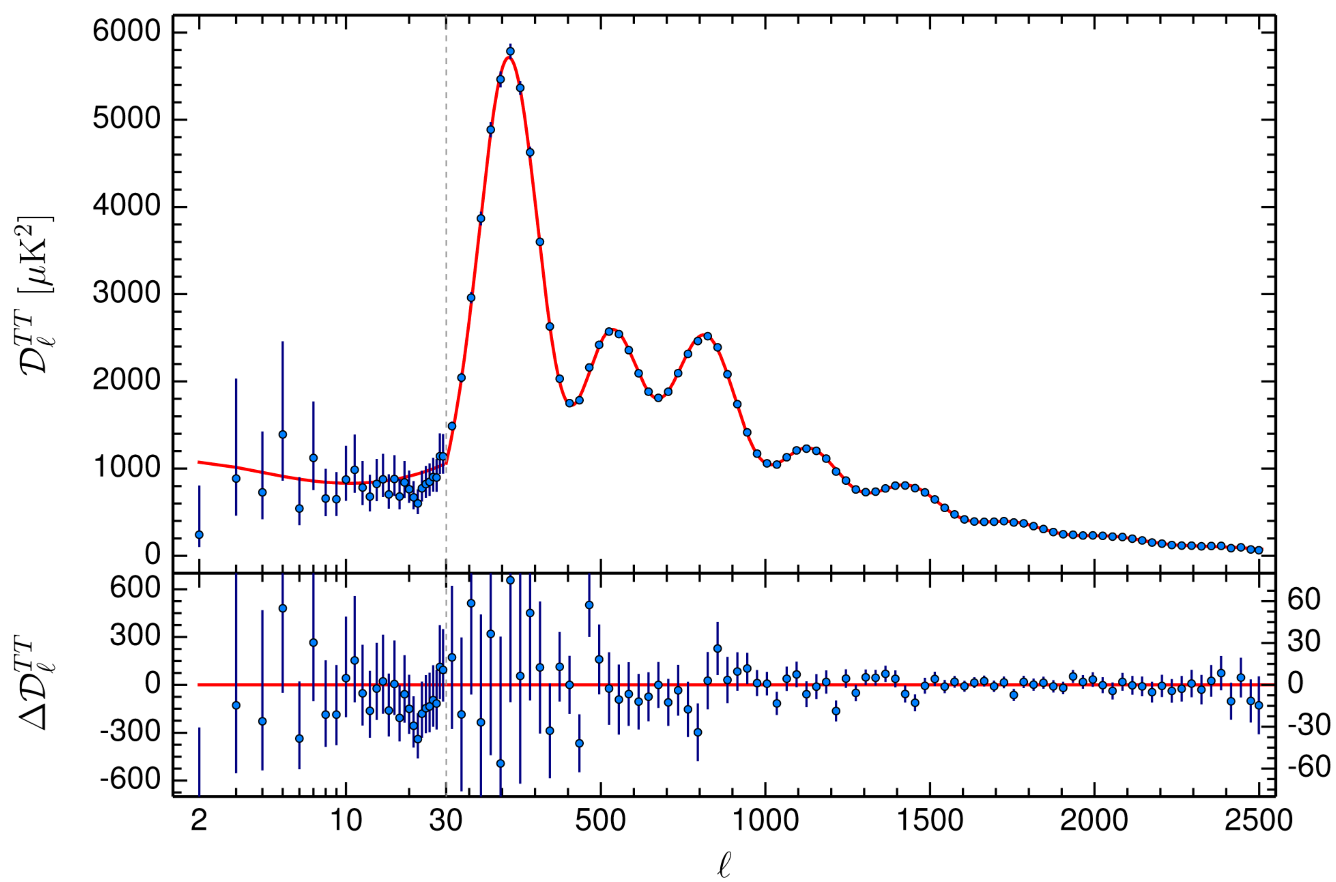
Galaxy Image
(Intensity Map)

Spatial Variation of Intensity



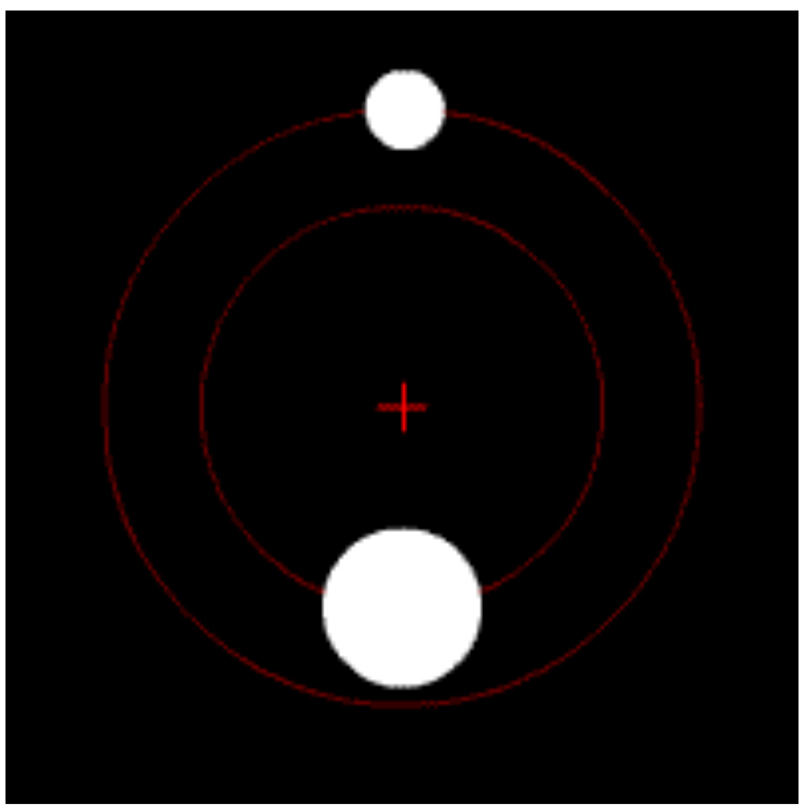
Cosmic Microwave Background (Planck) ~ Gaussian Random Field (mean = 2.7 K, std dev ~ 10^{-5})

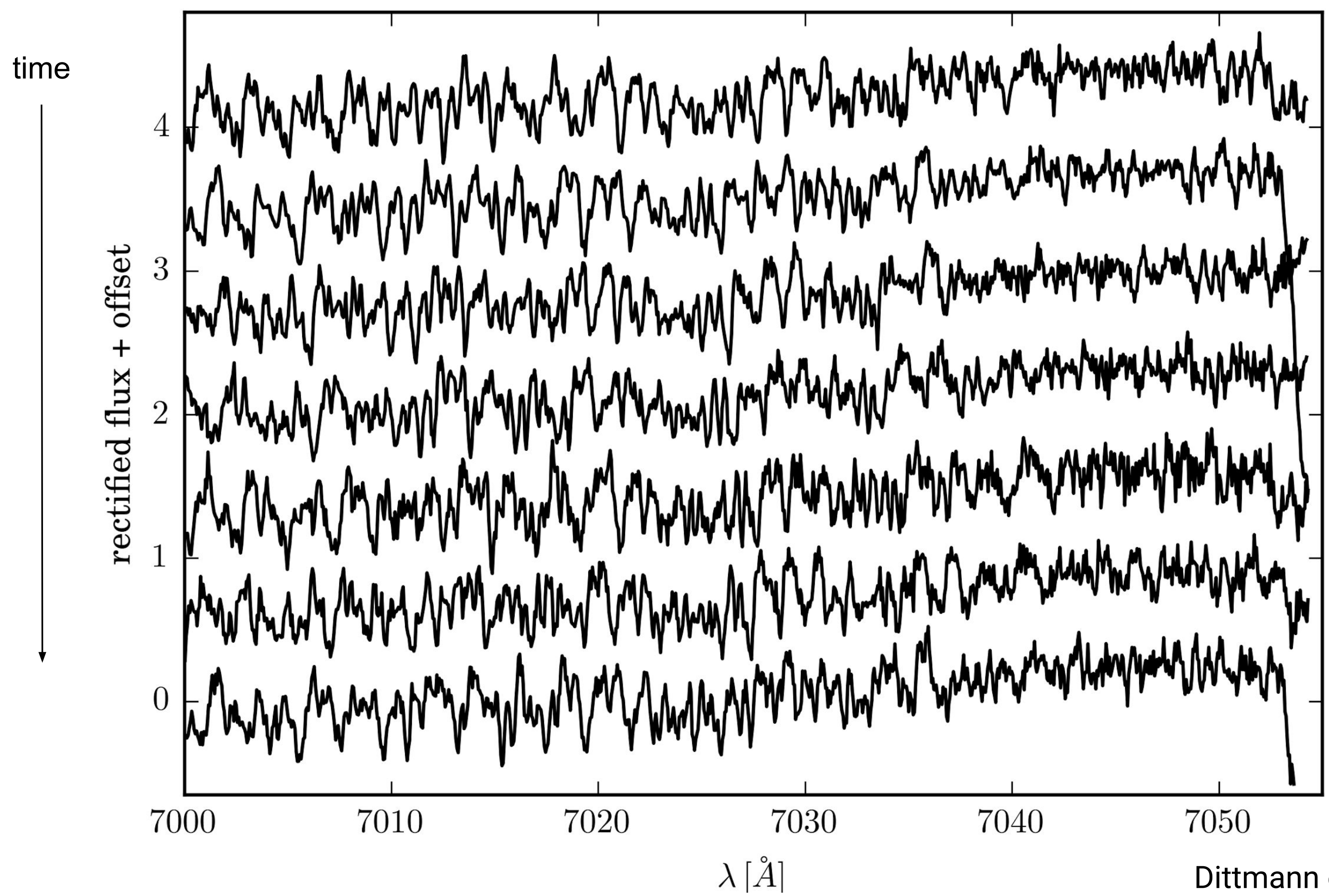
Power Spectrum (~Fourier Transform of Correlation Function)



Astrostatistics Case Studies:
Disentangling Time Series Spectra with Gaussian Processes: Applications to Radial Velocity Analysis
(Czekala et al. 2017, ApJ, 840, 49. arXiv:1702.05652)

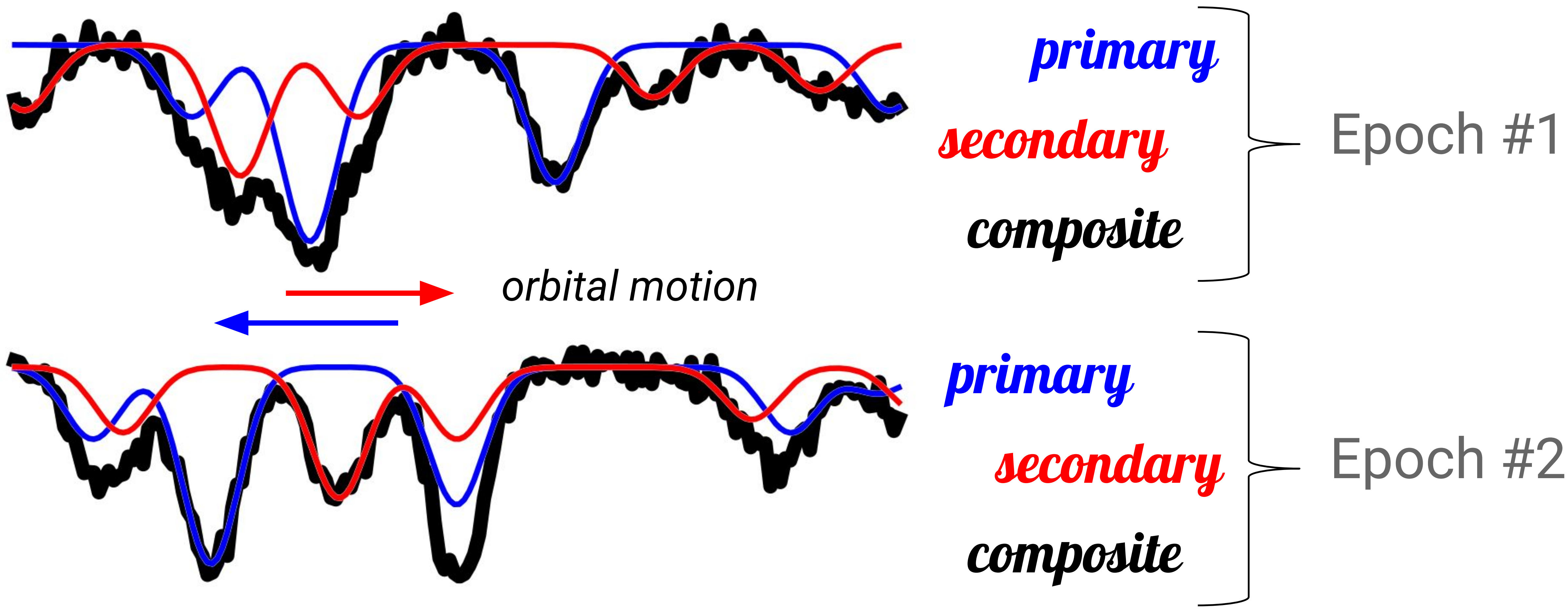
Raw Observations of the LP661-13 M4 Binary





Dittmann et al. 17
Czekala et al. 17a

Spectroscopic Binary Stars

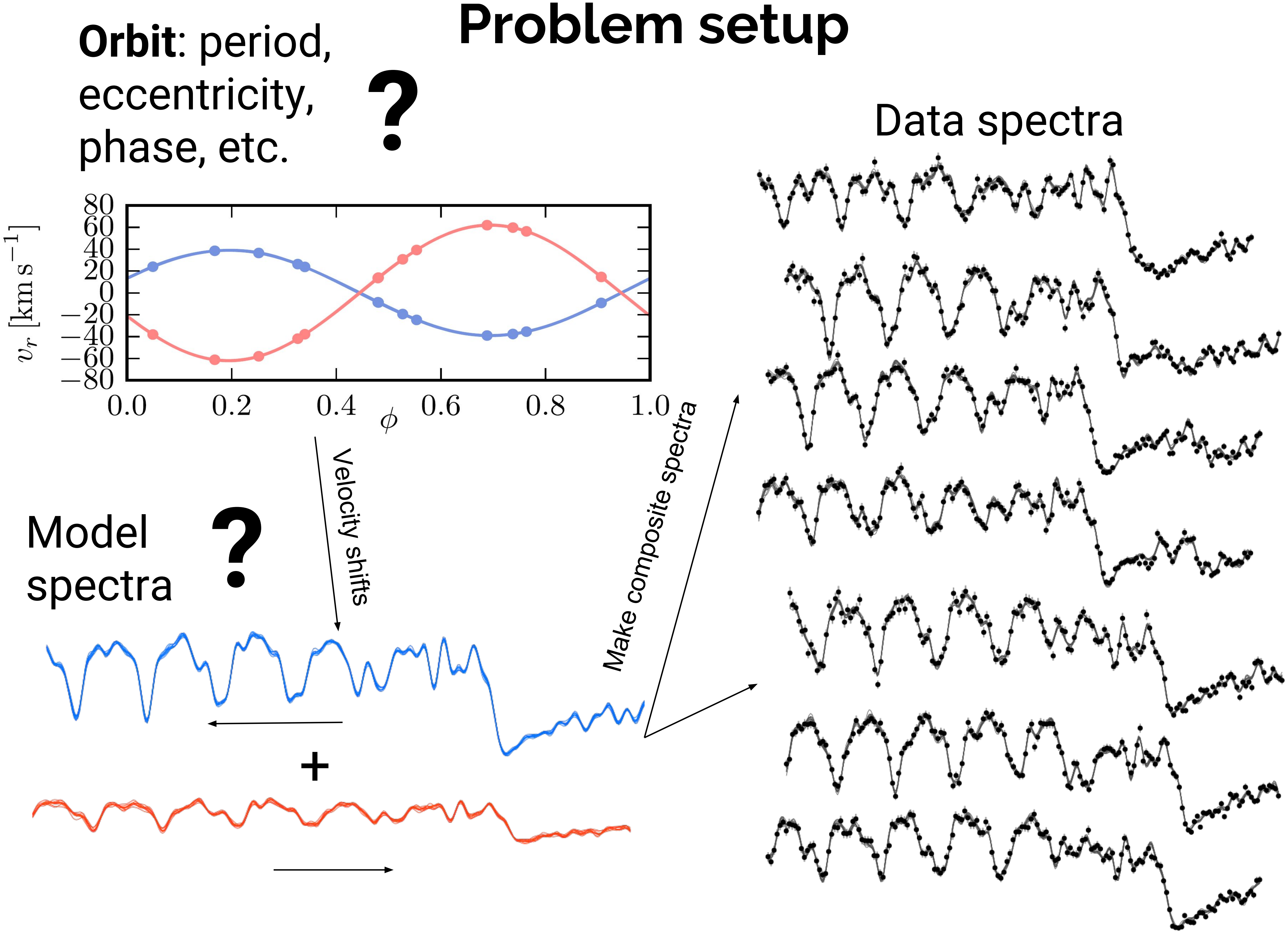


(Epoch = Time)

We only observe the “noisy” sum of two (latent) spectra.
Latent (underlying) spectra are unknown functions
Observed spectrum = Measured Data

Forward Model = Generates Data

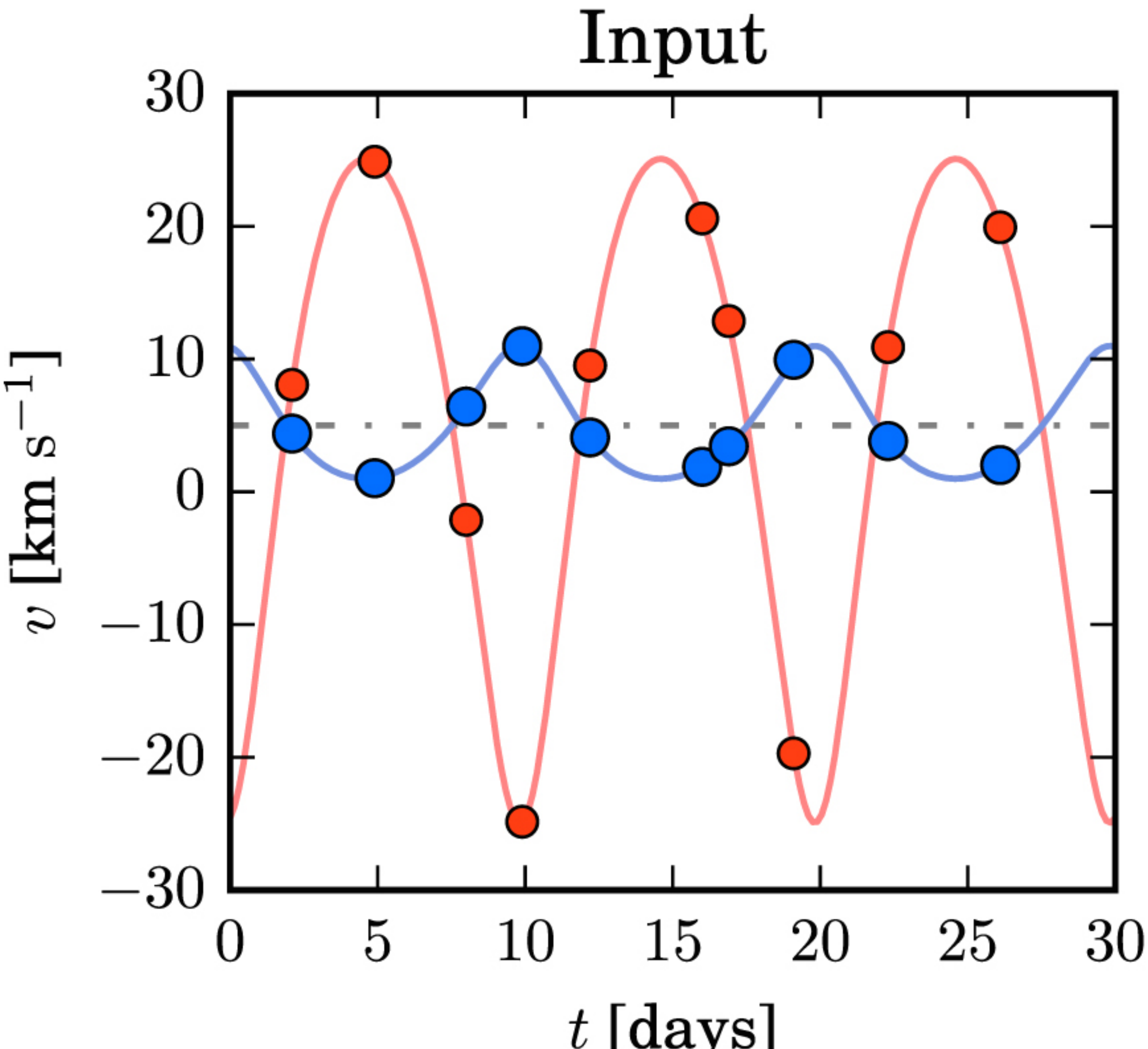
Problem setup



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

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

Orbital Parametric Model



Seven Parameters:

- Mass Ratio
- Velocity Amplitude
- eccentricity
- Arg of Periastron
- Epoch of Periastron
- Orbital Period
- Systemic Velocity

Nonparametric Bayes

Gaussian processes

We will model the latent stellar spectrum f_λ as a Gaussian process

$$f_\lambda \sim \mathrm{GP}(\mu(\lambda),\, k(\lambda, \lambda'))$$

A function is said to have a Gaussian process if for any collection of inputs the random vector ***f*** has a multivariate Gaussian distribution with mean ***mu*** and covariance matrix given by *k* evaluated over ***lambda***

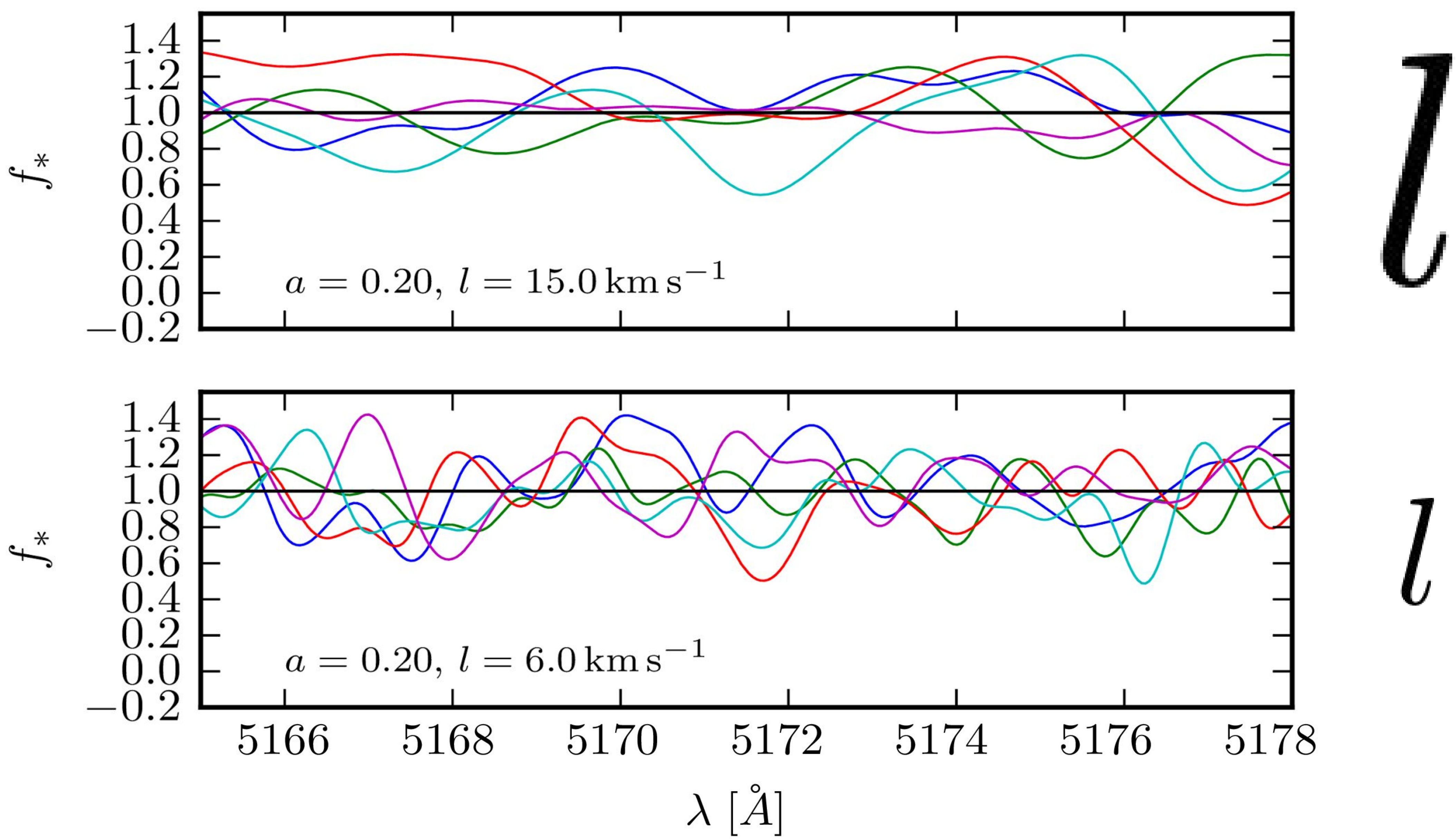
For a covariance kernel, we will use the commonly used squared exponential kernel, which relates pixels in the spectrum based upon their distance in log-wavelength (∝ velocity)

$$k_{ij}(r_{ij} \,|\, a, \boxed{l}) = a^2 \exp\left(-\frac{r_{ij}^2}{2l^2}\right)$$

Gaussian Process = a prior on functions (latent spectra)

Gaussian Process model for a single, stationary star

(Zoomed) draws from the prior

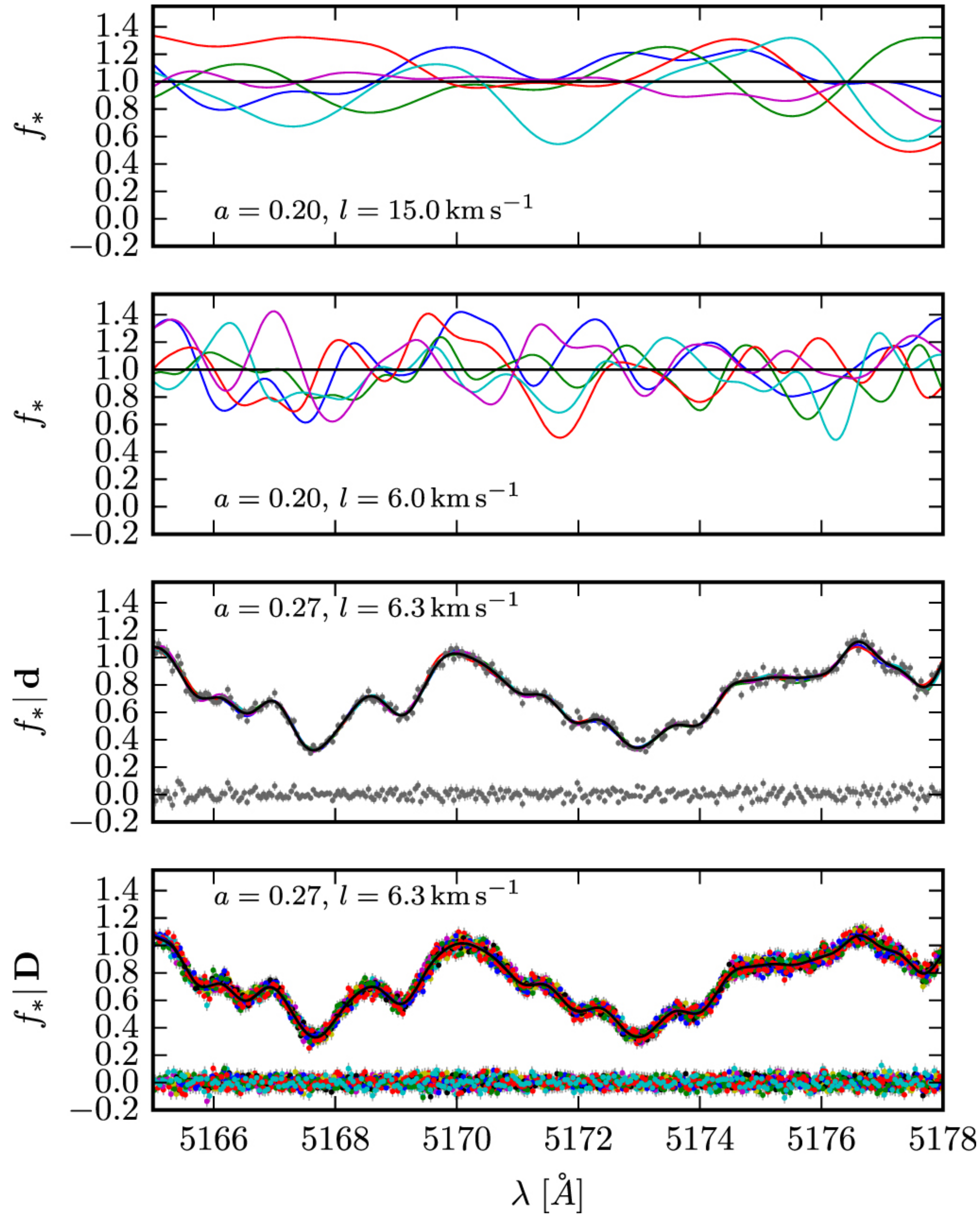


Inference = Which function is most consistent with the data?

Gaussian Process: Priors & Posteriors

GP prior
(long/short length scales)

GP Posterior
(conditioned on data spectrum d)
Inference of latent spectrum



Known Unknowns

$$\Theta$$

7-dim Orbital Parameters = Period, Phase, eccentricity, Velocity Amplitude

$$f(\lambda), g(\lambda)$$

(∞-dim) Latent Functions = the unobserved component spectra of the primary (f) and secondary (g) stars

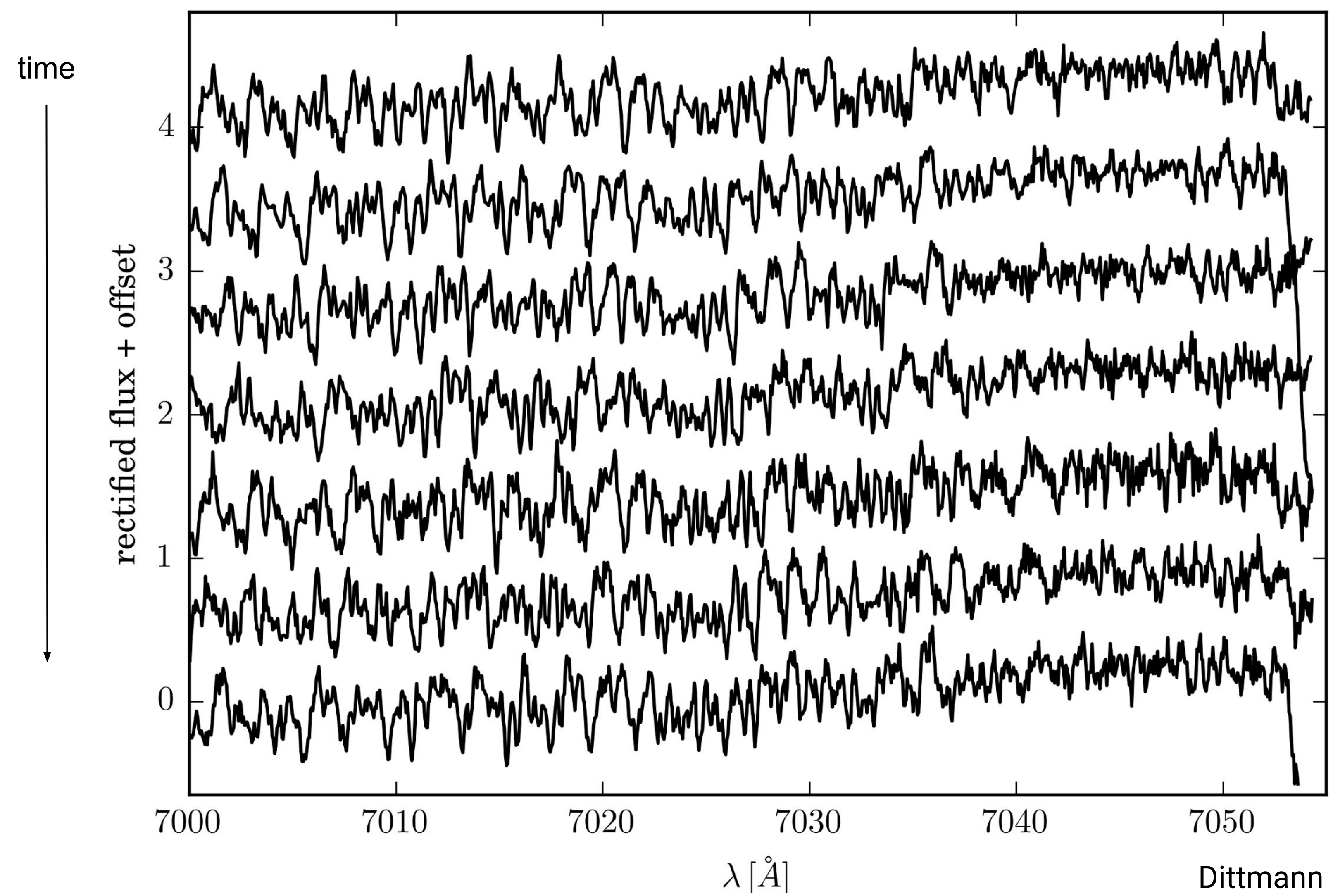
$$\boldsymbol{\alpha} = (a_f, l_f, a_g, l_g)$$

4-dim GP hyperparameters = controlling the amplitude and smoothness of Gaussian Process prior on latent spectra

Knowns (Data)

Raw Observations of the LP661-13 M4 Binary

$$\boldsymbol{D} =$$



Dittmann et al. 17
Czekala et al. 17a

Bayesian Inference

$$P(A|\,B) = \frac{P(B|A)P(A)}{P(B)}$$

In this case:

$$P(\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha}|\boldsymbol{D}) \propto$$
$$P(\boldsymbol{D}|\,\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha}) \times P(\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha})$$

a probability density on (4+7+∞)-dim parameter space

Bayesian Computation

1. Run Markov Chain Monte Carlo (MCMC) (e.g. *emcee* affine-invariant ensemble sampler) on the 4+7 small dimensional marginal posterior

$$P(\boldsymbol{\Theta}, \boldsymbol{\alpha} | \boldsymbol{D}) = \int d\boldsymbol{f} \int d\boldsymbol{g}\, P(\boldsymbol{\Theta}, \boldsymbol{f}, \boldsymbol{g}, \boldsymbol{\alpha} | \boldsymbol{D})$$

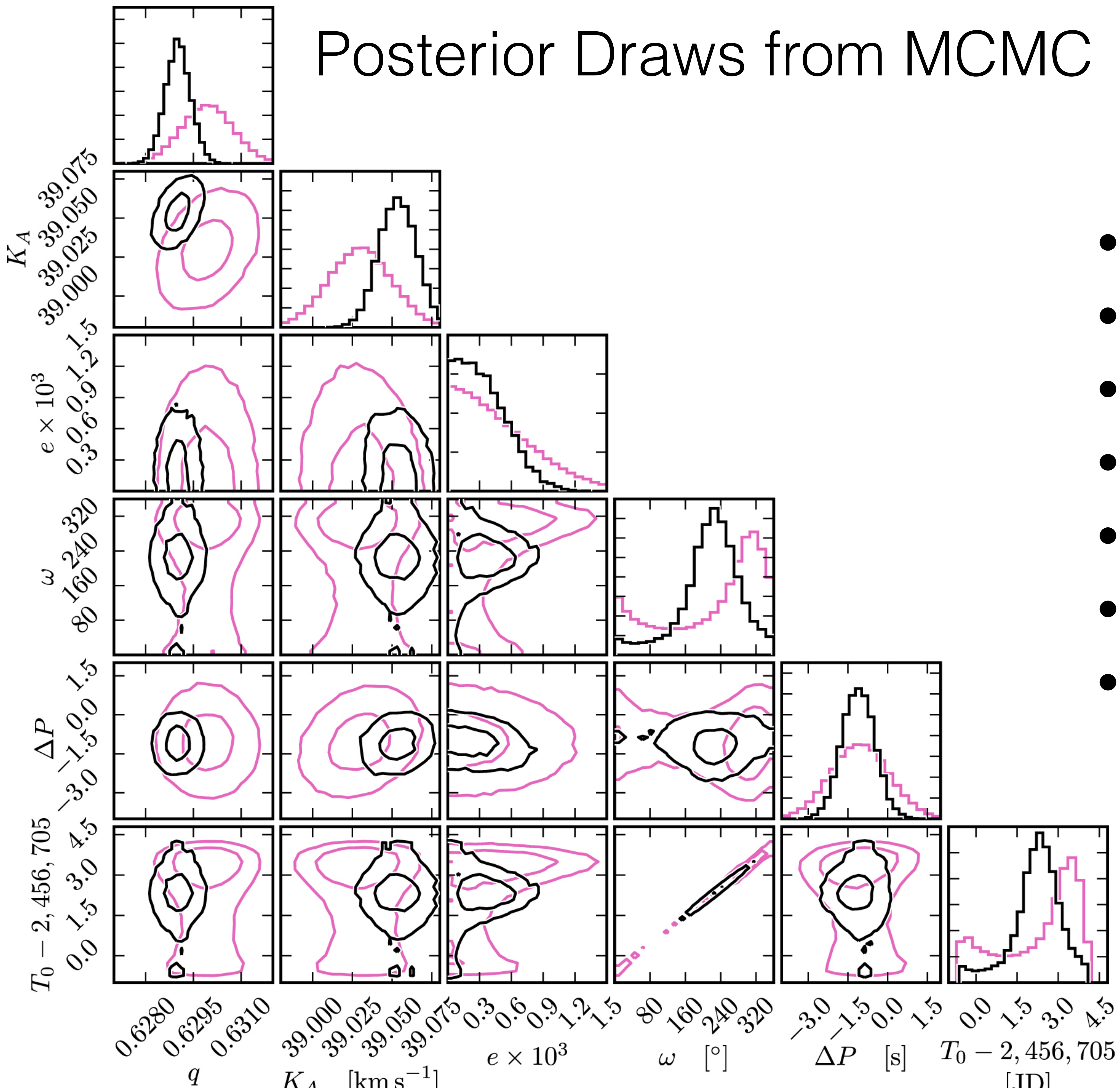
MCMC generates samples: $\boldsymbol{\Theta}_i, \boldsymbol{\alpha}_i \sim P(\boldsymbol{\Theta}, \boldsymbol{\alpha} | \boldsymbol{D})$

2. Draw high-dim (**f**, **g**) spectra from the posterior predictive distribution

$$\boldsymbol{f}_i, \boldsymbol{g}_i \sim P(\boldsymbol{f}, \boldsymbol{g} | \boldsymbol{\Theta}_i, \boldsymbol{\alpha}_i, \boldsymbol{D})$$

Application to the Mid-M-Dwarf Binary LP661-13

Posterior Draws from MCMC

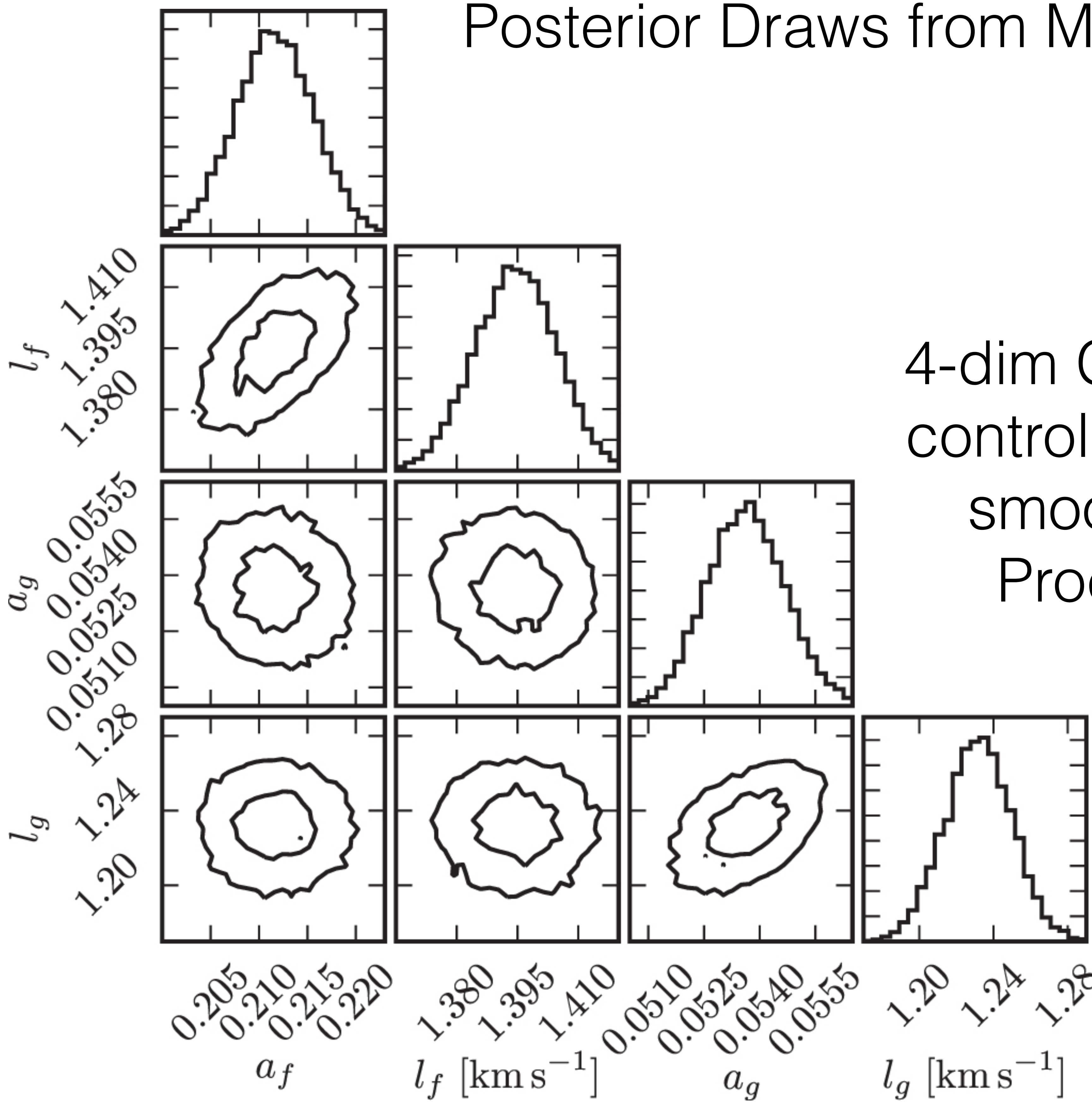


Seven Orbital Parameters:

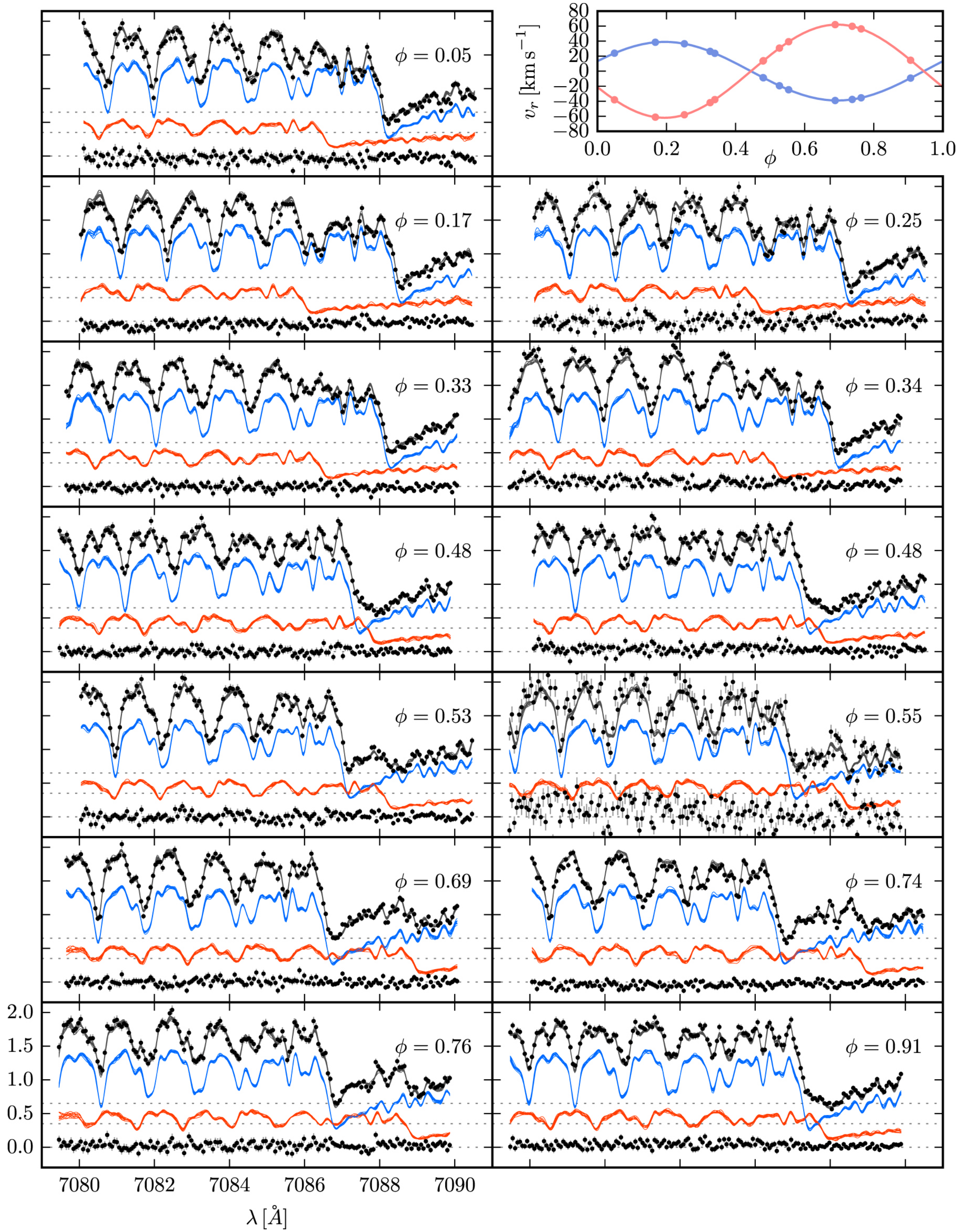
- Mass Ratio
- Velocity Amplitude
- eccentricity
- Arg of Periastron
- Epoch of Periastron
- Orbital Period
- Systemic Velocity

Application to the Mid-M-Dwarf Binary LP661-13

Posterior Draws from MCMC $\boldsymbol{\alpha} = (a_f, l_f, a_g, l_g)$



4-dim GP hyperparameters = controlling the amplitude and smoothness of Gaussian Process prior on latent spectra



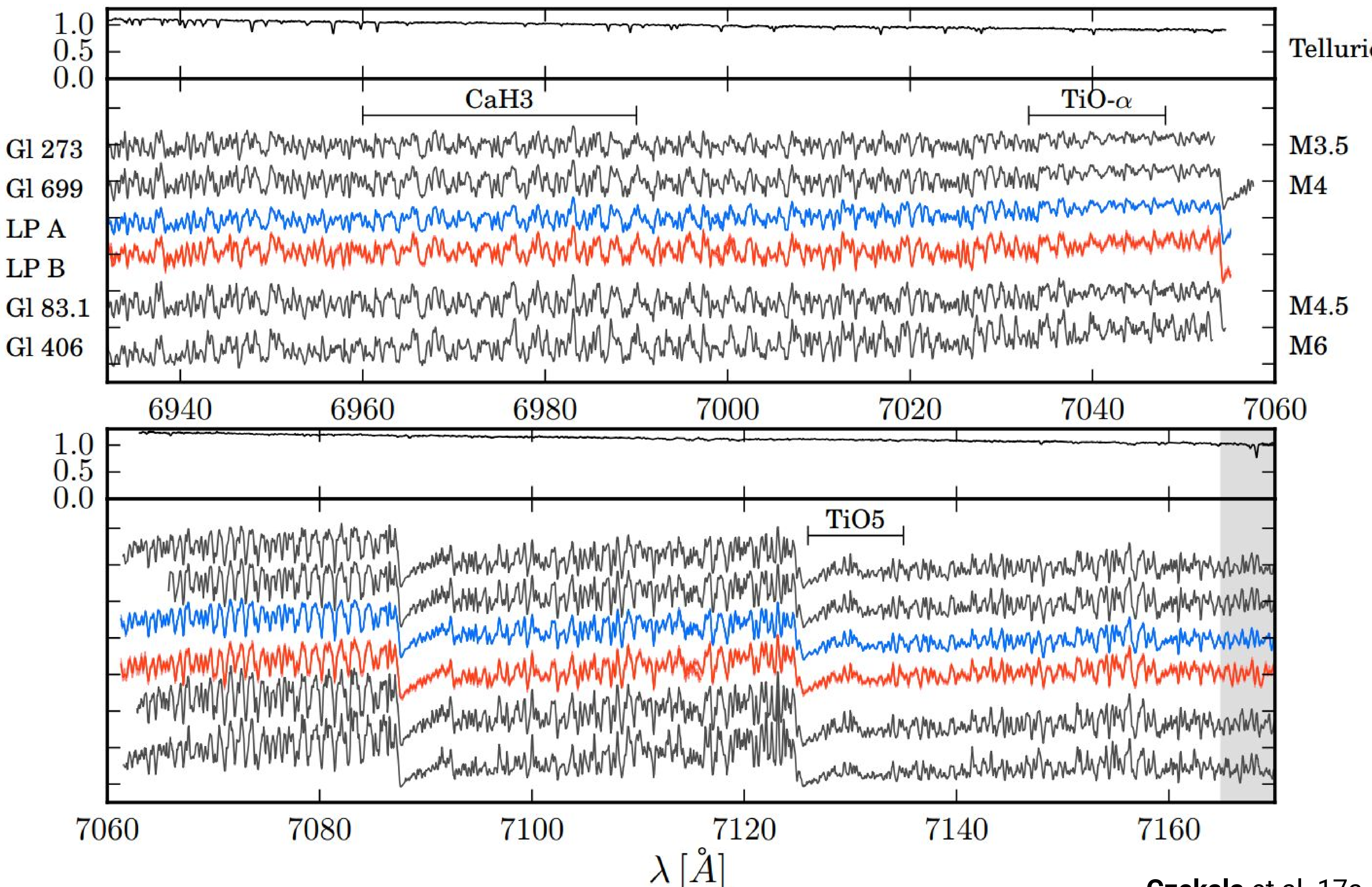
Posterior Inference of Component Spectra (**f**, **g**)

Compared to 10 epochs of observed spectra (**data**)

Model Checking!
Checking Fit against Data

Model Checking!
Checking Fit against Domain Knowledge (astrophysics)!

Disentangled spectra match other single standard stars



Czekala et al. 17a

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