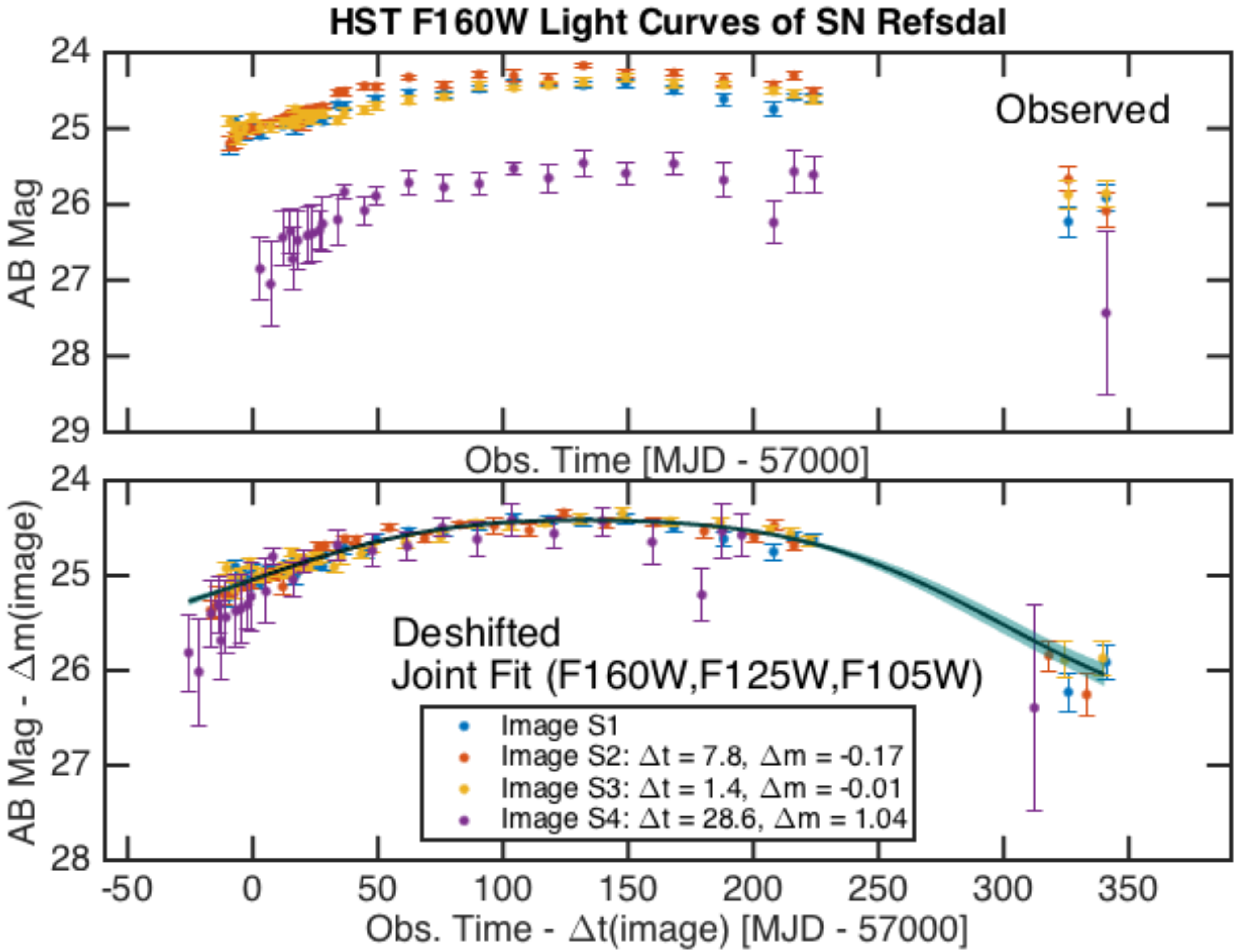
Astrostatistics: 08 Mar 2019

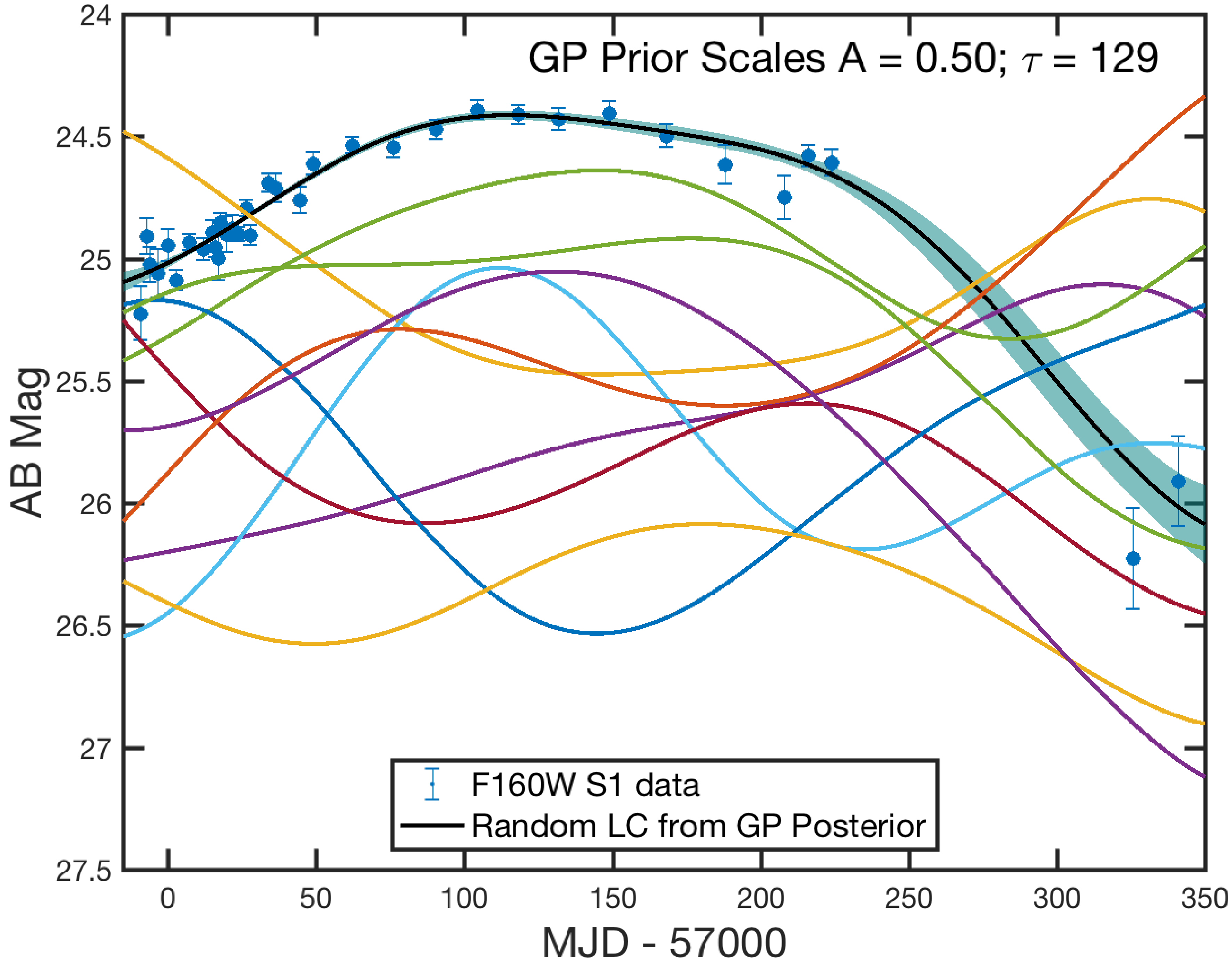
https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2019

- Next Week scheduling
- Make-up Lecture & Example class?
  - Tue 12 Mar, 2pm, Thu 14, 11am or Fri 15 Mar 1pm MR13 ?
- Finish Gaussian Processes in Astrophysics
- Probabilistic Graphical Models (Bishop, Ch 8) & Hierarchical Bayes

Metropolis-within-Gibbs results: 4 chains
Deshifting the data



What is the form for the fitted GP?



Nonparametric form implicit in the posterior calculations as a linear combination of the observed data

$$\begin{pmatrix} \boldsymbol{y}_o \\ \boldsymbol{f}_* \end{pmatrix} \sim N\left(\begin{bmatrix} \mathbf{1}c \\ \mathbf{1}c \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{t}_o, \boldsymbol{t}_o) + \boldsymbol{W} & \boldsymbol{K}(\boldsymbol{t}_*, \boldsymbol{t}_o) \\ \boldsymbol{K}(\boldsymbol{t}_o, \boldsymbol{t}_*) & \boldsymbol{K}(\boldsymbol{t}_*, \boldsymbol{t}_*) \end{bmatrix} \right)$$

Now can calculate function prediction at unobserved points

$$\boldsymbol{f}_* | \boldsymbol{y}_o \sim N(\mathbb{E}[\boldsymbol{f}_* | \boldsymbol{y}_o], \mathrm{Var}[\boldsymbol{f}_* | \boldsymbol{y}_o])$$

Using Gaussian Conditional Properties:

$$\mathbb{E}[\boldsymbol{f}_* | \boldsymbol{y}_o] = \mathbf{1}c + \boldsymbol{K}(\boldsymbol{t}_*, \boldsymbol{t}_o)[\boldsymbol{K}(\boldsymbol{t}_o, \boldsymbol{t}_o) + \boldsymbol{W}]^{-1}(\boldsymbol{y}_o - \mathbf{1}c)$$

$$\mathrm{Var}[\boldsymbol{f}_* | \boldsymbol{y}_o] = \boldsymbol{K}(\boldsymbol{t}_*, \boldsymbol{t}_*) - \boldsymbol{K}(\boldsymbol{t}_*, \boldsymbol{t}_o)[\boldsymbol{K}(\boldsymbol{t}_{\boldsymbol{o}}, \boldsymbol{t}_o) + \boldsymbol{W}]^{-1}\boldsymbol{K}(\boldsymbol{t}_, \boldsymbol{t}_*)$$

Other covariance functions (R&W Chapter 4)

Squared Exponential gives very smooth curves.

Ornstein Uhlenbeck Process (Damped Random Walk)
Exponential Covariance Function

$$k(t,t') = A^2 \exp(-|t-t'|/\tau)$$

(Stationary = time translation invariant, Symmetric)

$$df(t) = \tau^{-1}\underbrace{[\mu - f(t)]}_{\text{Mean-Reversion Drag Term}}dt + \underbrace{\sigma dW_t}_{\text{Random Walk}}$$

$\tau =$ mean-reversion timescale

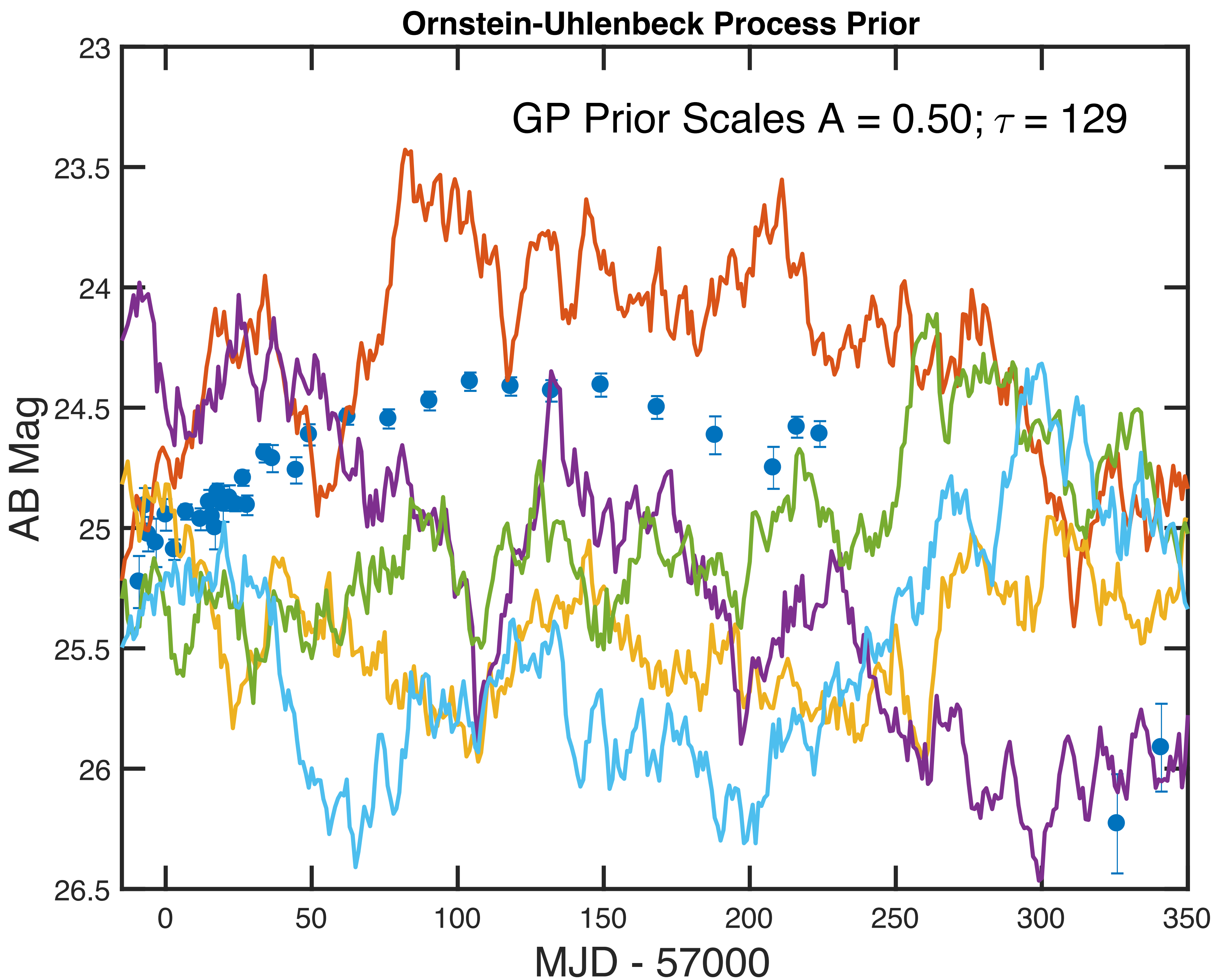
$\mu =$ long-term mean

$\sigma =$ volatility

$$A^2 = \tau\sigma^2/2$$

Ornstein Uhlenbeck Process (Damped Random Walk)
Exponential Covariance Function

$$k(t, t') = A^2 \exp(-|t - t'|/\tau)$$



Everywhere continuous but not differentiable

Other covariance functions (R&W Chapter 4)

Ornstein Uhlenbeck Process (Damped Random Walk)
Exponential Covariance Function

$$k(t,t') = A^2 \exp(-|t-t'|/\tau)$$

v = 1/2 Special Case of Matern kernel:

The Matérn Class of Covariance Functions

The *Matérn class* of covariance functions is given by

$$k_{\mathrm{Matern}}(r) \;=\; \frac{2^{1-\nu}}{\Gamma(\nu)}\Big(\frac{\sqrt{2\nu}r}{\ell}\Big)^{\nu} K_{\nu}\Big(\frac{\sqrt{2\nu}r}{\ell}\Big), \tag{4.14}$$

with positive parameters ν and ℓ, where K_ν is a modified Bessel function

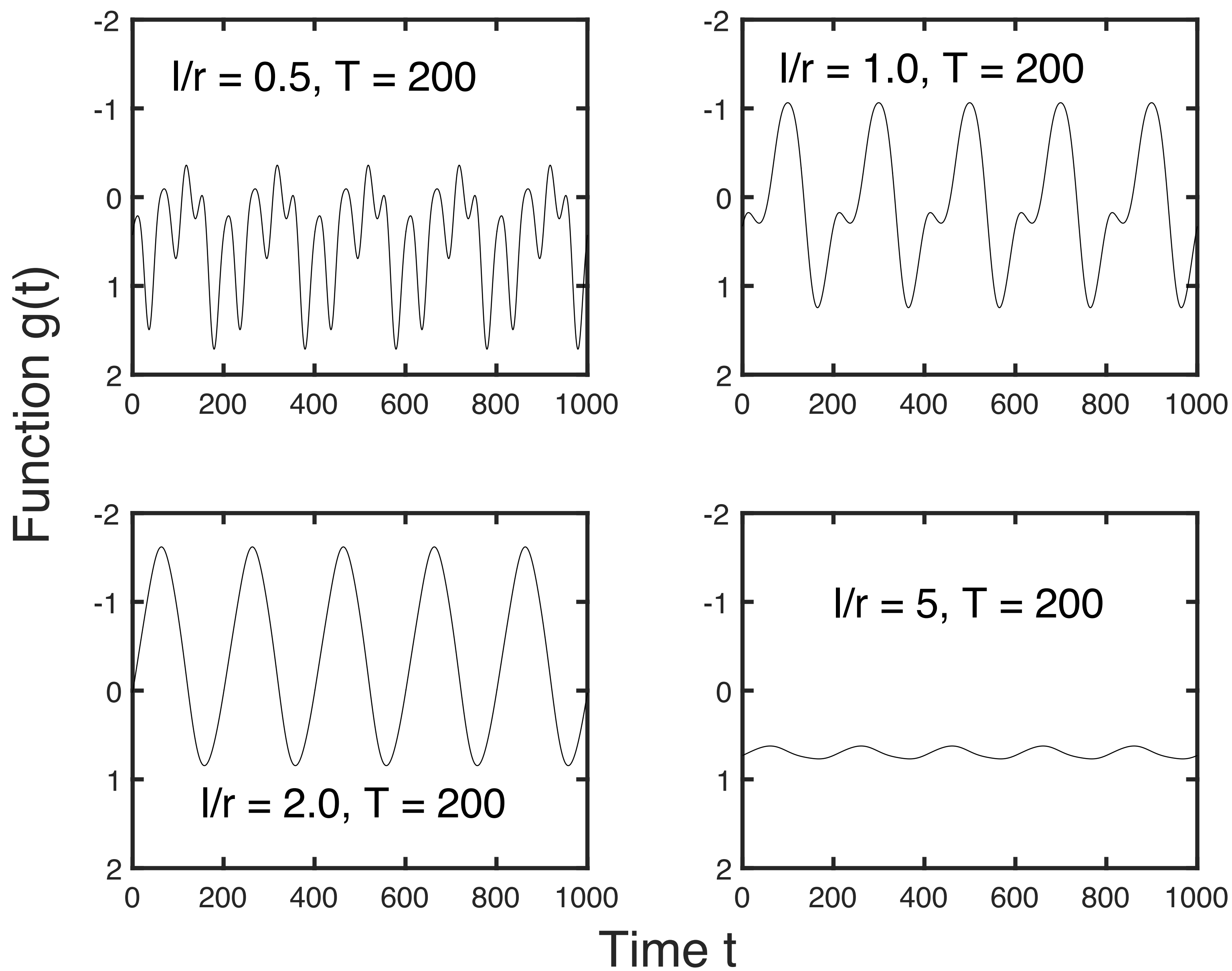
When v = p + 1/2 is half-integer: Exponential x p-Polynomial

$$k_{\nu=p+1/2}(r) \;=\; \exp\Big(-\frac{\sqrt{2\nu}r}{\ell}\Big)\frac{\Gamma(p+1)}{\Gamma(2p+1)}\sum_{i=0}^{p}\frac{(p+i)!}{i!(p-i)!}\Big(\frac{\sqrt{8\nu}r}{\ell}\Big)^{p-i}$$

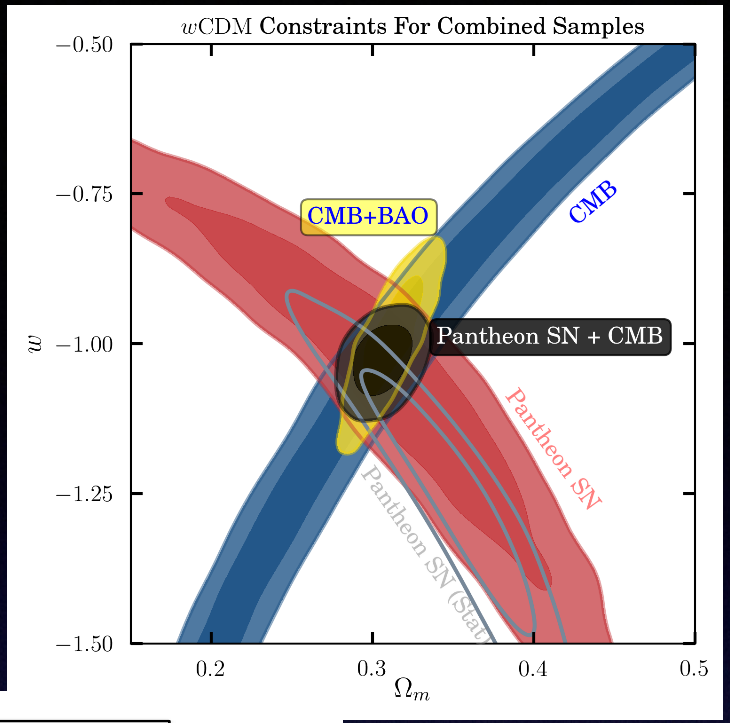
Periodic Covariance Functions

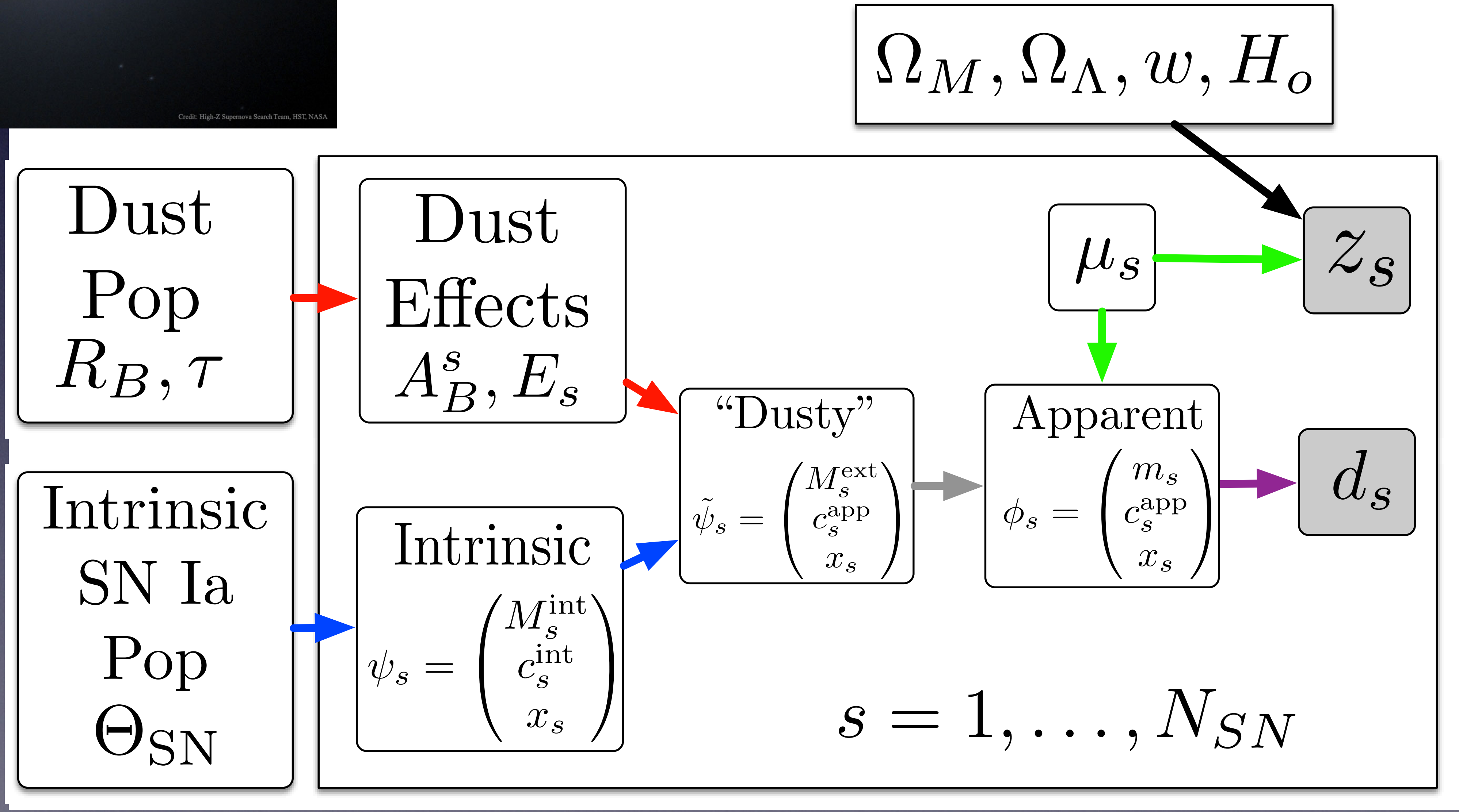
$$k(t,t') = A^2 \exp\left(-\frac{2r^2}{l^2} \sin^2(\pi(t-t')/T) \right)$$

(Example Sheet: Periodic kernel can be derived for any Gaussian process on **R²**)



Hierarchical Bayes, Huh?
What is it good for?
ABSOLUTELY EVERYTHING!



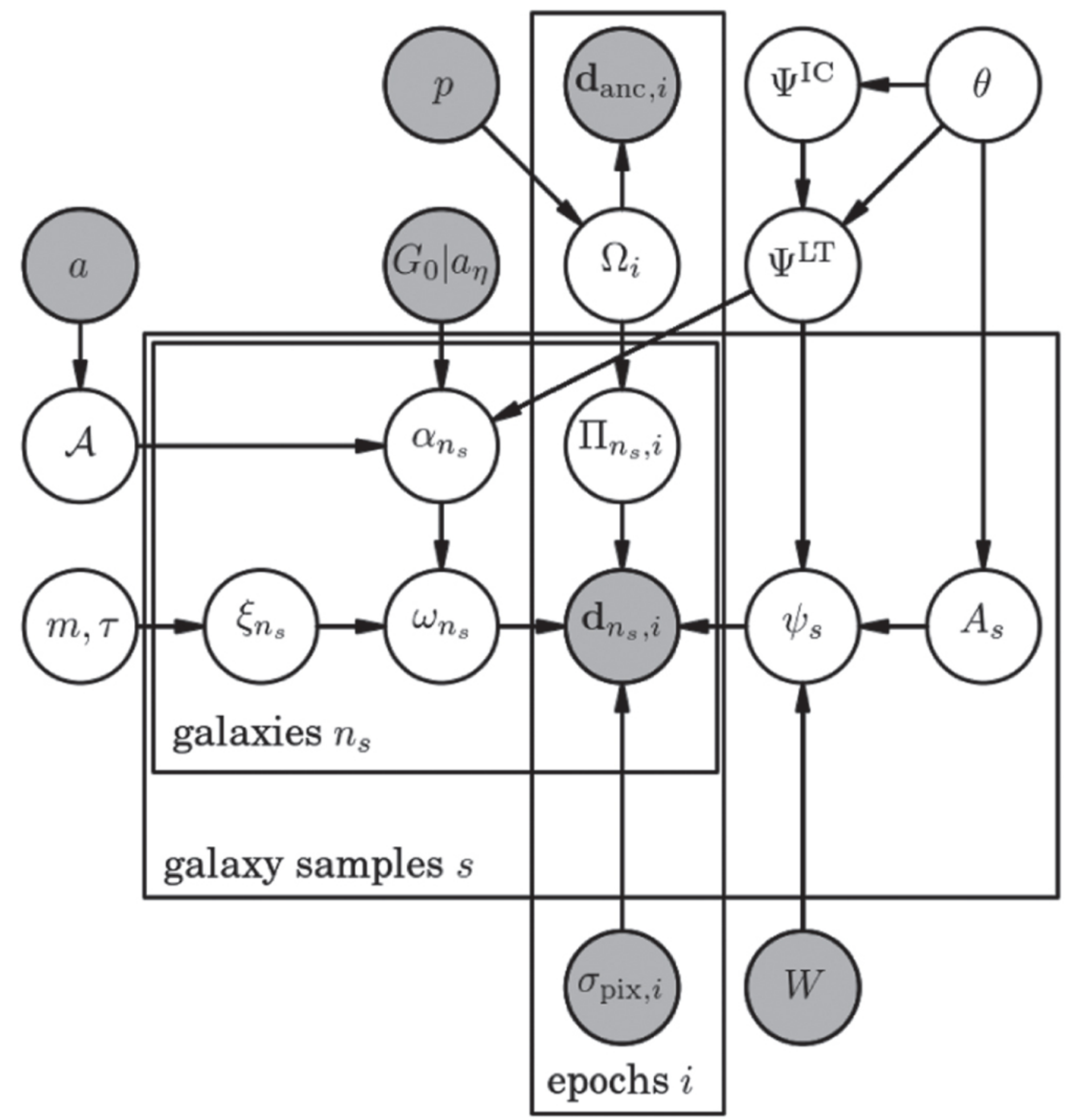


Mandel et al. 2017

Hierarchical Models in Astrophysics

- M. Schneider et al. 2015, ApJ 807, 87, "Hierarchical Probabilistic Inference of Cosmic Shear". (weak gravitational lensing)

- "Hierarchical Probabilistic" = "Hierarchical Bayesian" = "Bayesian Hierarchical" = "Multilevel Model (MLM)"



Probabilistic Graphical Model

Generative Model for the Sky: Regier et al. "Learning an Astronomical Catalog of the Visible Universe through Scalable Bayesian Inference"

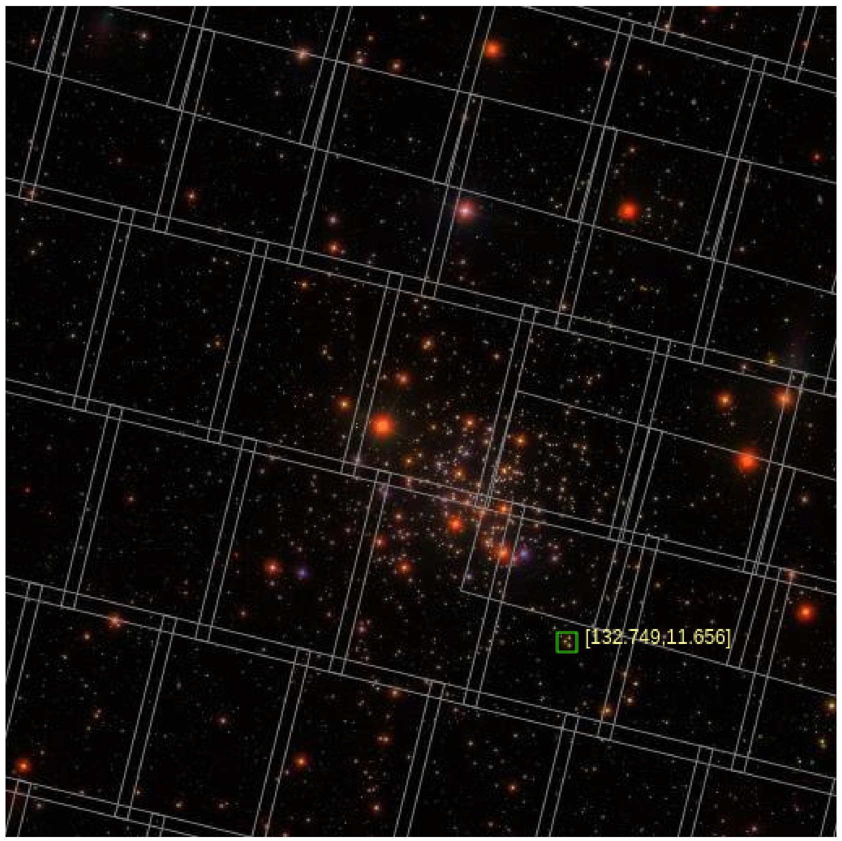


Figure 1: SDSS image boundaries. Some images overlap substantially. Some light sources appear in multiple images that do not overlap. Celeste uses all relevant data to locate and characterize each light source whereas heuristics typically ignore all but one image in regions with overlap. credit: SDSS DR10 Sky Navigate Tool.

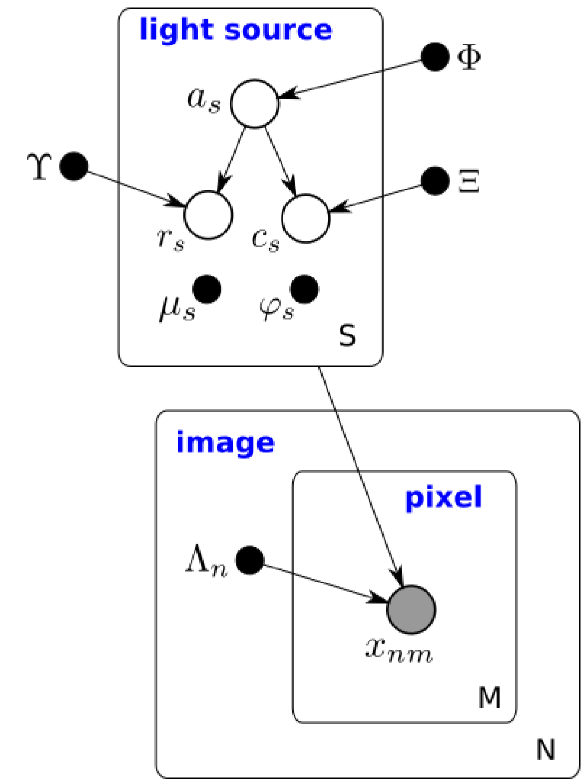


Figure 2: The Celeste graphical model. Shaded vertices represent observed random variables. Empty vertices represent latent random variables. Black dots represent constants. Constants denoted by uppercase Greek characters are fixed throughout our procedure. Constants denoted by lowercase Greek letters are inferred, along with the posterior distribution of the latent random variables. Edges signify permitted conditional dependencies. Plates (the boxes) represent independent replication.

Probabilistic Graphical Model

Common Problems in Astronomy

- Want to learn about a population of objects from a finite sample of individuals, each measured with error
- Observed Data is actually a combination of uncertain astrophysical & instrumental & selection effects. Need to model them to infer the "intrinsic" properties of the object or population of objects ("deconvolve")

What is Hierarchical Bayes?

Simple Bayes: $\mathcal{D}|\,\theta \sim \mathrm{Model}(\theta)$

Posterior (Bayes' Theorem): $P(\theta|\mathcal{D}) \propto P(\mathcal{D}|\,\theta)P(\theta)$

Hierarchical Bayes: $\theta_i:$ Parameter of Individual

α, β: Hyperparameter of Population

$$\mathcal{D}_i|\,\theta_i \sim \mathrm{Model}(\theta_i)$$

$$\theta_i|\,\alpha, \beta \sim \mathrm{PopModel}(\alpha, \beta)$$

Joint Posterior:

$$P(\{\theta_i\}, \alpha, \beta|\{\mathcal{D}_i\}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i|\,\theta_i)P(\theta_i|\alpha, \beta)\right]P(\alpha, \beta)$$

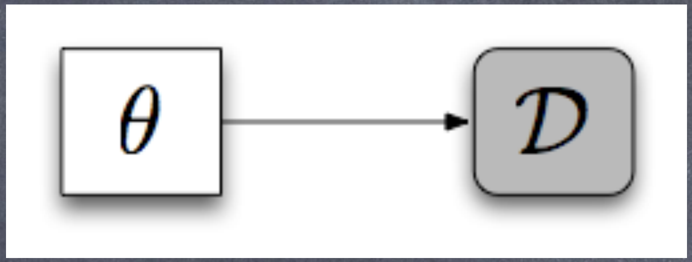
Build up complexity by layering conditional probabilities

Probabilistic Graphical Models:

a visual way to understand complex statistical models

Simple Bayes:

$$\mathcal{D}|\,\theta \sim \mathrm{Model}(\theta)$$

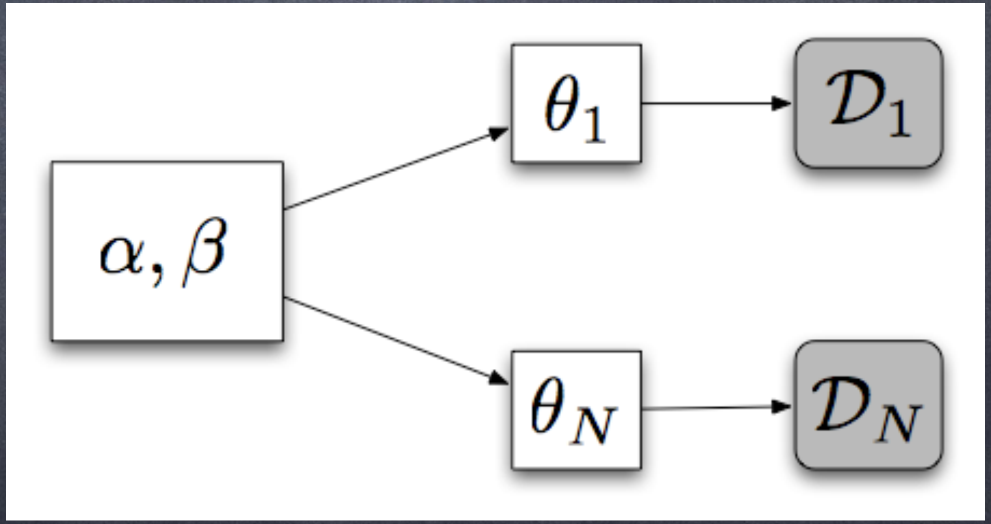


$$P(\theta|\,\mathcal{D}) \propto P(\mathcal{D}|\,\theta)P(\theta)$$

Hierarchical Bayes:

$$\mathcal{D}_i|\,\theta_i \sim \mathrm{Model}(\theta_i)$$

$$\theta_i|\,\alpha,\beta \sim \mathrm{PopModel}(\alpha,\beta)$$



$$P(\{\theta_i\},\alpha,\beta|\{\mathcal{D}_i\}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i|\,\theta_i)P(\theta_i|\alpha,\beta)\right]P(\alpha,\beta)$$

Build up complexity by layering conditional probabilities

Probabilistic Graphical Models

Forward Model:

$$\theta_i | \, \alpha, \beta \sim \mathrm{PopModel}(\alpha, \beta)$$

$$\mathcal{D}_i | \, \theta_i \sim \mathrm{Model}(\theta_i)$$

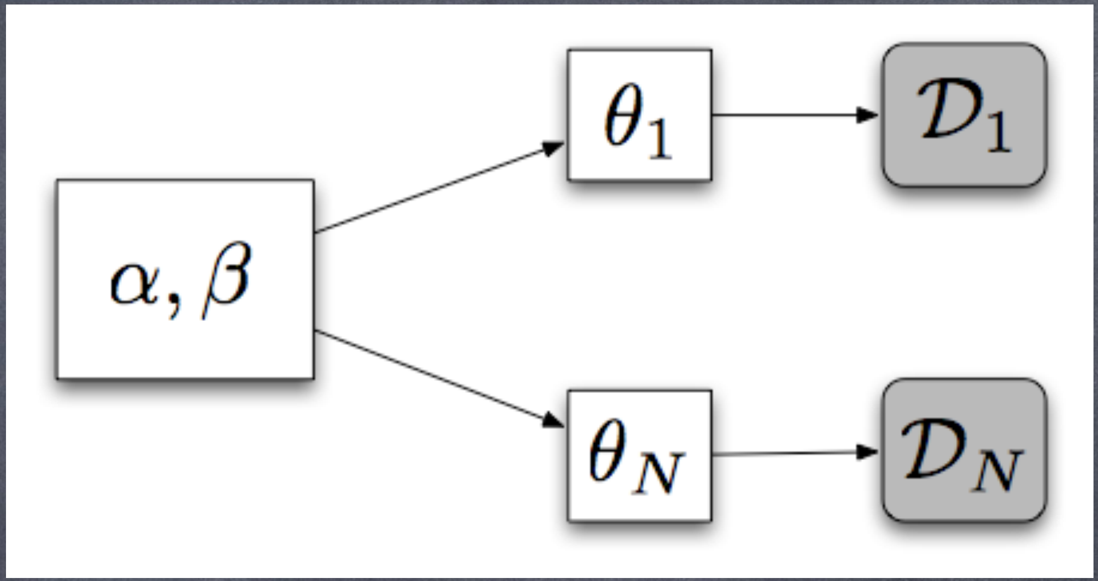
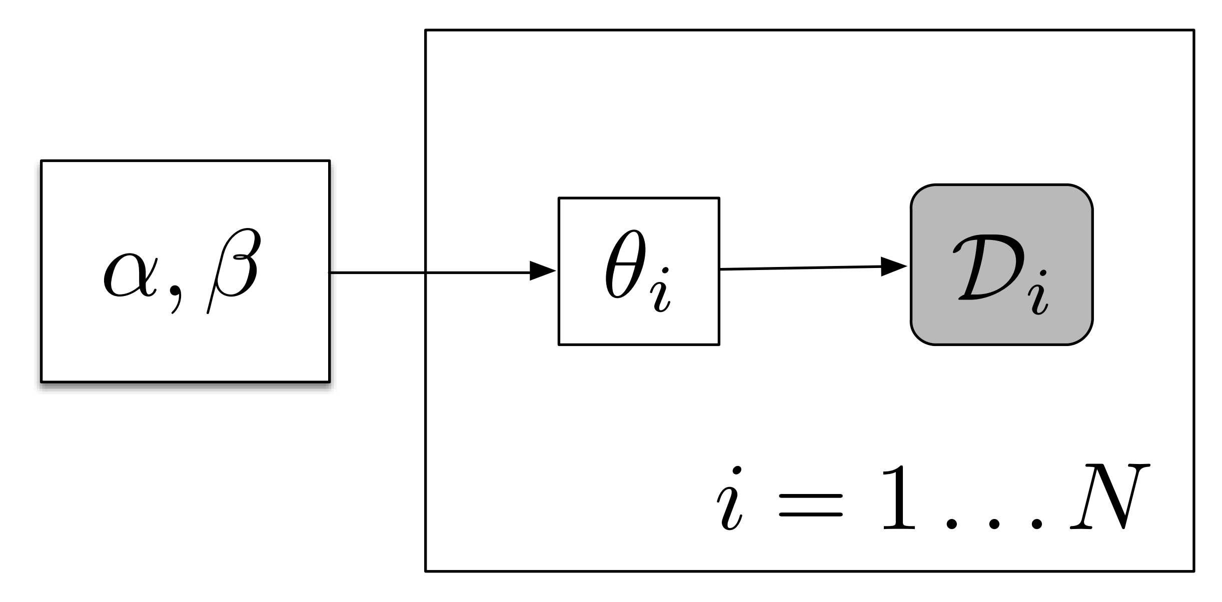


Plate Notation:
(loop over individuals in sample)



$$P(\{\theta_i\}, \alpha, \beta | \{\mathcal{D}_i\}) \propto \left[\prod_{i=1}^{N} P(\mathcal{D}_i | \, \theta_i) P(\theta_i | \alpha, \beta) \right] P(\alpha, \beta)$$

Build up complexity by layering conditional probabilities

Advantages of Hierarchical Bayesian Models

- Common Problem in Astronomy: Infer properties of population from finite sample of individuals with noisy measurements
- Incorporate multiple sources of randomness & uncertainty as "latent variables" with distributions underlying the data
- Express structured probability models adapted to data-generating process ("forward model")
- Bayesian: Full (non-gaussian) probability distribution = Global, coherent quantification of uncertainties
- Completely Explore & Marginalize Posterior trade-offs/degeneracies between parameters/hyperparameters

Switch to PRML Slides on properties of Graphical Models (got up until Cond'l independence 1)

PGM Slides:
lecture21-22_prml_slides-8_abridged.pdf

Slides on: Bayesian Networks, Curve Fitting, Generative Models, Conditional Independence, D-Separation