Cepheid light curve demography via Bayesian functional data analysis

Toward better Cepheid luminosities

Tom Loredo,[1] Martin Hendry,[2] Dan Kowal,[3] and David Ruppert[3]

[1]Center for Radiophysics & Space Research, Cornell University; [2]SUPA School of Physics & Astronomy, University of Glasgow; [3]Dept. of Statistical Science, Cornell University

1. Introduction

Motivation

- ***The dark energy equation of state:***
  To learn the DE-EOS—e.g., the w parameter—we need to know the Hubble parameter H_0 to 1% accuracy or better.
- ***Cepheid variables:***
  The main H_0 measurement bottleneck is the limited precision of the Cepheid variable period-luminosity relation (PLR), a foundation for the cosmic distance ladder.

Our project

- ***Goal:*** Improve our ability to infer Cepheid luminosities from light curve observations
- ***Objective:*** Move beyond standard PLR analyses
  - Explicitly model light curve diversity about the "average" PLR
  - Generalize the PLR to more flexible light curve-luminosity relations (LCLRs)
- ***Approach:*** Bayesian functional data analysis (FDA)
  - *FDA:* "Function demographics" (vs. demographics of scalars or multivariate vectors)—an important emerging area of statistics
  - *Bayesian:* Probabilistic modeling of light curves as stochastic processes

2. Multilevel modeling

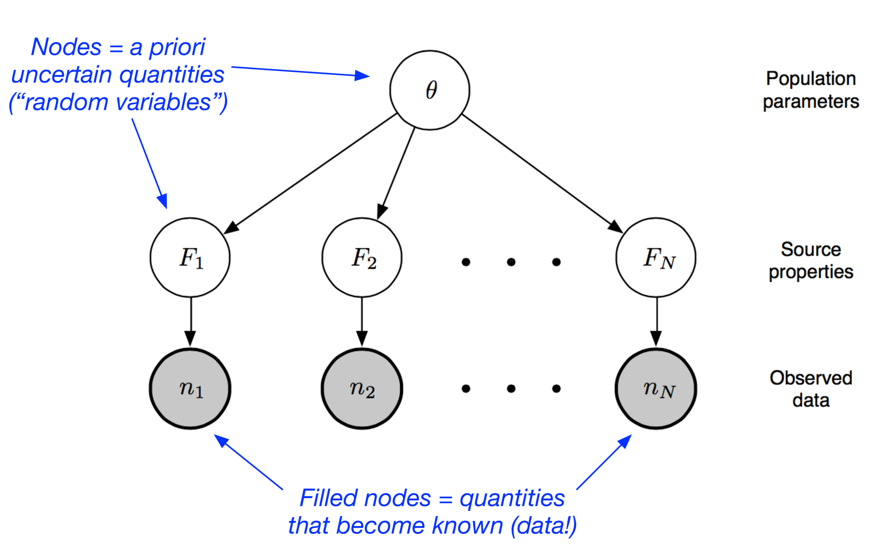
A key notion underlying Bayesian FDA is **multilevel modeling** (a.k.a. *hierarchical modeling* or *graphical modeling*)

- Accounts for multiple sources of randomness—population diversity, measurement error…
- Grouping of data ⇒ "borrowing strength"—indirect sharing of information between population members

Illustrative example: Infer the **number–size distribution** ("log(N)–log(S)") for a population of sources from noisy measurements of their fluxes.

- Population model: Gamma distribution (like Schechter function) $\propto F^{\alpha-1}e^{-F/s}$
- Observation model: Photon counting data modeled with the Poisson distribution

Graphical model



Slope & scale; Mean & standard dev'n

$$\theta = (\alpha, s) \text{ or } (\mu, \sigma)$$

$$\pi(\theta) = \mathrm{Flat}(\mu, \sigma)$$

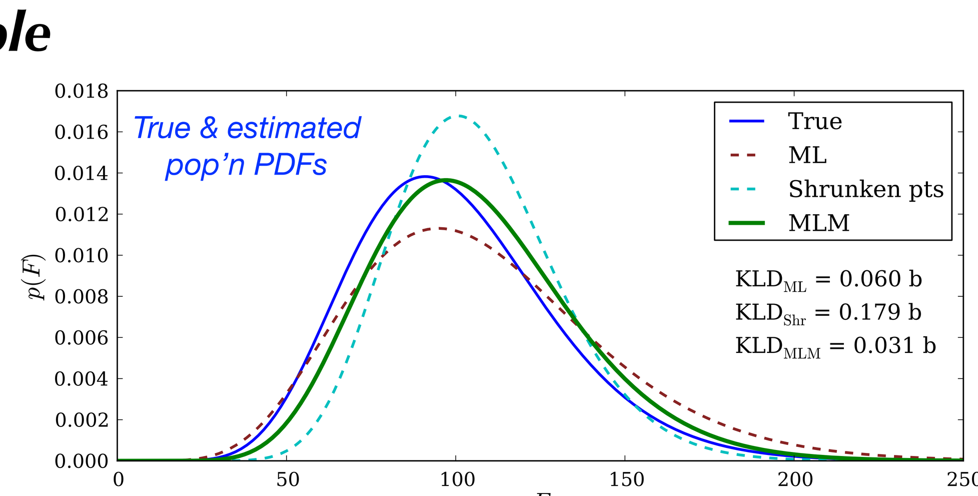
$$p(F_i|\theta) = \mathrm{Gamma}(F_i|\theta)$$

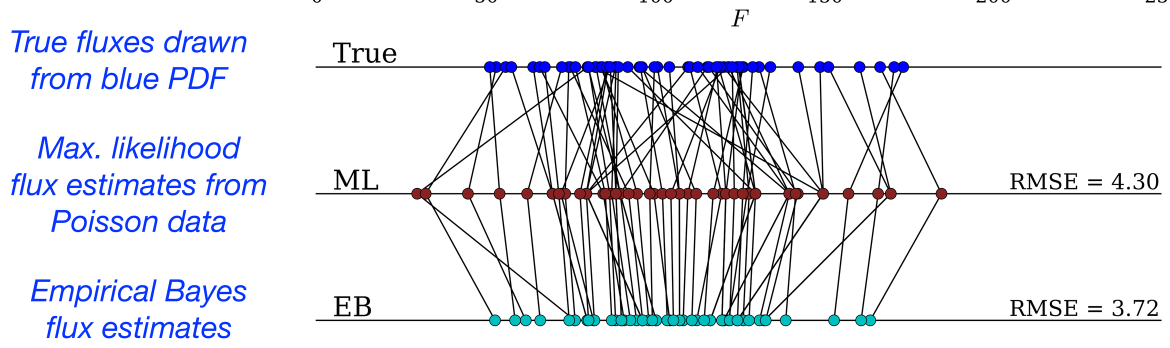
$$p(n_i|F_i) = \mathrm{Pois}(n_i|\epsilon_i F_i)$$

Simulated example

60 sources from gamma PDF with $\langle F\rangle = 100$ and $\sigma_F = 30$

Exposures spanning dynamic range of ×16





Take-aways:

- Using best-fit fluxes as surrogates for true fluxes is bad!
- Multilevel modeling implements an adaptive "Eddington bias" correction
- Joint modeling of the population and measurements:
  - Produces improved member estimates via shrinkage/borrowing strength
  - Produces population-level/demographic inferences that account for measurement errors (and selection effects when necessary)

3. Bayesian FDA for Cepheids

Conventional Cepheid calibration

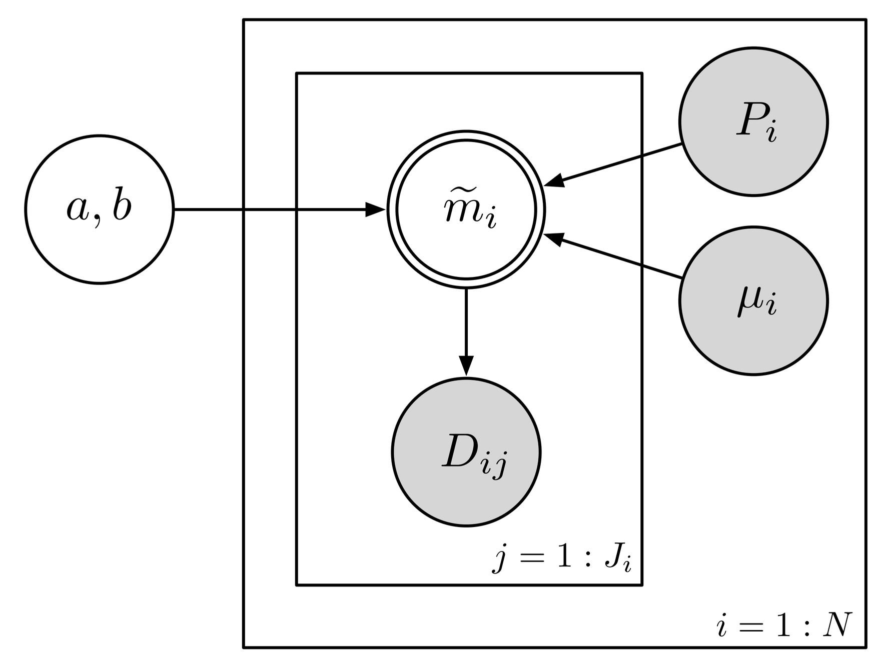
- Estimate periods from light curves via periodograms/harmonic fitting; treat as precise
- Estimate the average Cepheid magnitude, $\widetilde{m}$, from light curve data D
- For Cepheids with known distance moduli, perform linear regression to estimate Leavitt law parameters

Leavitt law & observation model:

$$M_i = a + b\log\left(\frac{P_i}{10\mathrm{d}}\right)$$

$$\hat{\widetilde{m}}_i = M_i + \mu_i + \epsilon_i$$

Graph for conventional approach



Bayesian FDA Cepheid calibration

- Model each light curve as the sum of two components ("mixed effects model"):
  - Shared template ("fixed effect"), governed by the Leavitt law PLR
  - Peculiar component ("random effect"), reflecting diversity among Cepheids with a common period
- Estimate the template PLR parameters, the peculiar component distribution, and each peculiar component *jointly*

Toy-model demonstration

- Generate "true" light curves for 250 Cepheids, based on the Fourier light curve fits of Pejcha and Kochanek (2012) and a "true" Leavitt law with $(a,b) = (-4.05, -2.43)$
- Simulate ~100 observations of each, using cadences and typical noise levels from various Cepheid surveys
- Estimate (a,b) with conventional regression and mixed effects approaches, using Fourier models for the template and peculiar components

Fourier light curve model:

$$m_{ij} = M_i + \mu_i + \sum_{h=1}^{H}\left[c_{h,i}\cos(h\phi_{ij}) + s_{h,i}\sin(h\phi_{ij})\right]$$

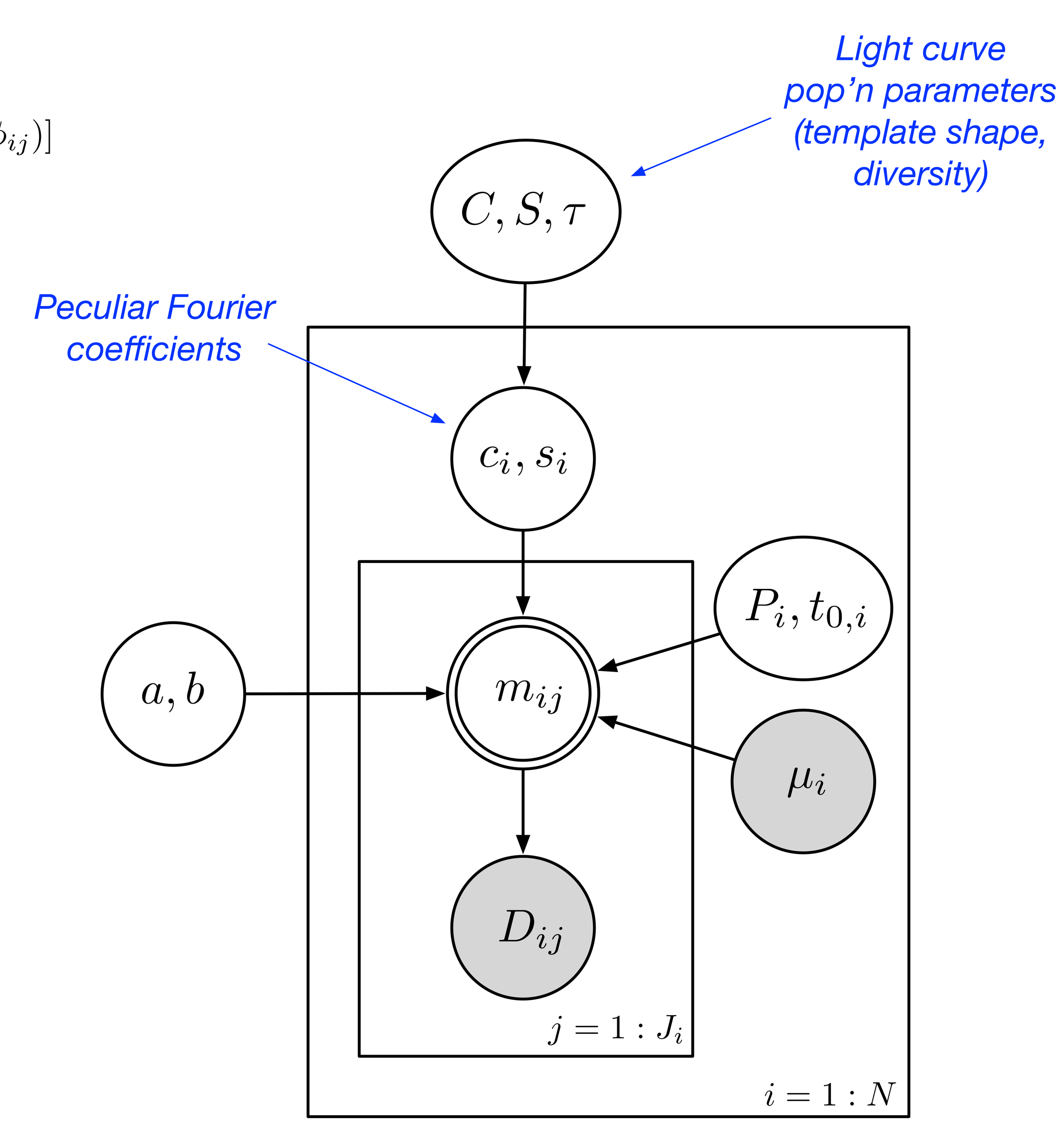
$$\phi_{ij} = \frac{2\pi}{P_i}(t_{ij} - t_{0,i})$$

Light curve population model:

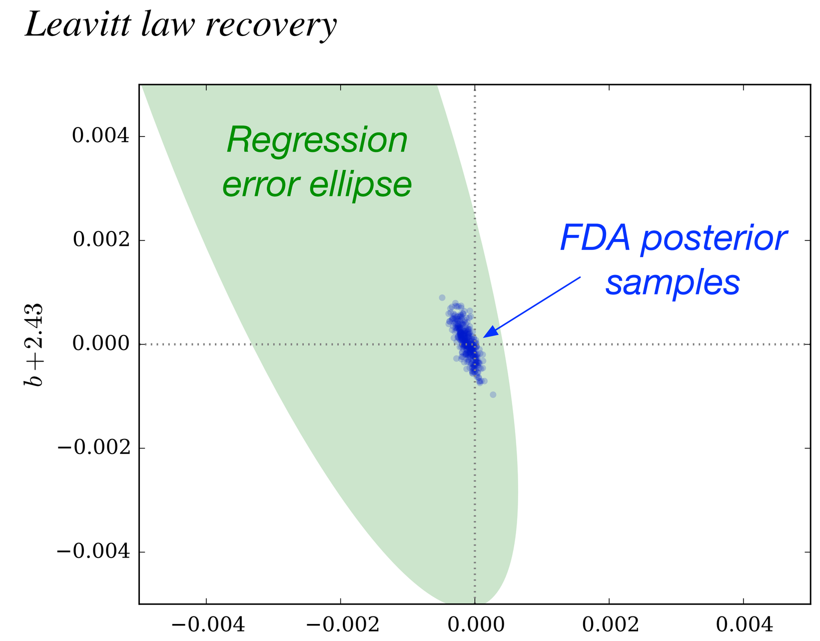
$$c_{h,i} \sim \mathrm{Norm}(C_h, \tau_h)$$

$$s_{h,i} \sim \mathrm{Norm}(S_h, \tau_h)$$

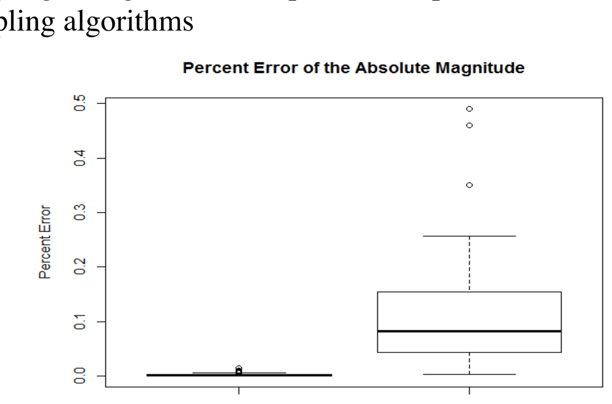
Means ⇒ template; Variance ⇒ diversity



Leavitt law recovery



FDA implemented via Metropolis-within-Gibbs MCMC sampling, using Robust Adaptive Metropolis (RAM) and slice sampling algorithms



- FDA precisely and accurately recovers the PLR, separating template and peculiar components of light curves
- Individual Cepheid luminosity accuracy is improved via shrinkage
- *Work in progress:*
  - Multiband extension, including treatment of extinction & reddening
  - Beyond the PLR: Mining for structure amidst Cepheid diversity and exploiting it to improve Cepheid calibration---building a generalized Cepheid ***light curve–luminosity relation***

This work is funded by the NSF via grant AST-1312903