Flexible and Scalable Methods for Time-Series Characterization

Dr. Andrew Becker
Associate Research Professor
University of Washington
Department of Astronomy

Motivation

Motivation

LSST

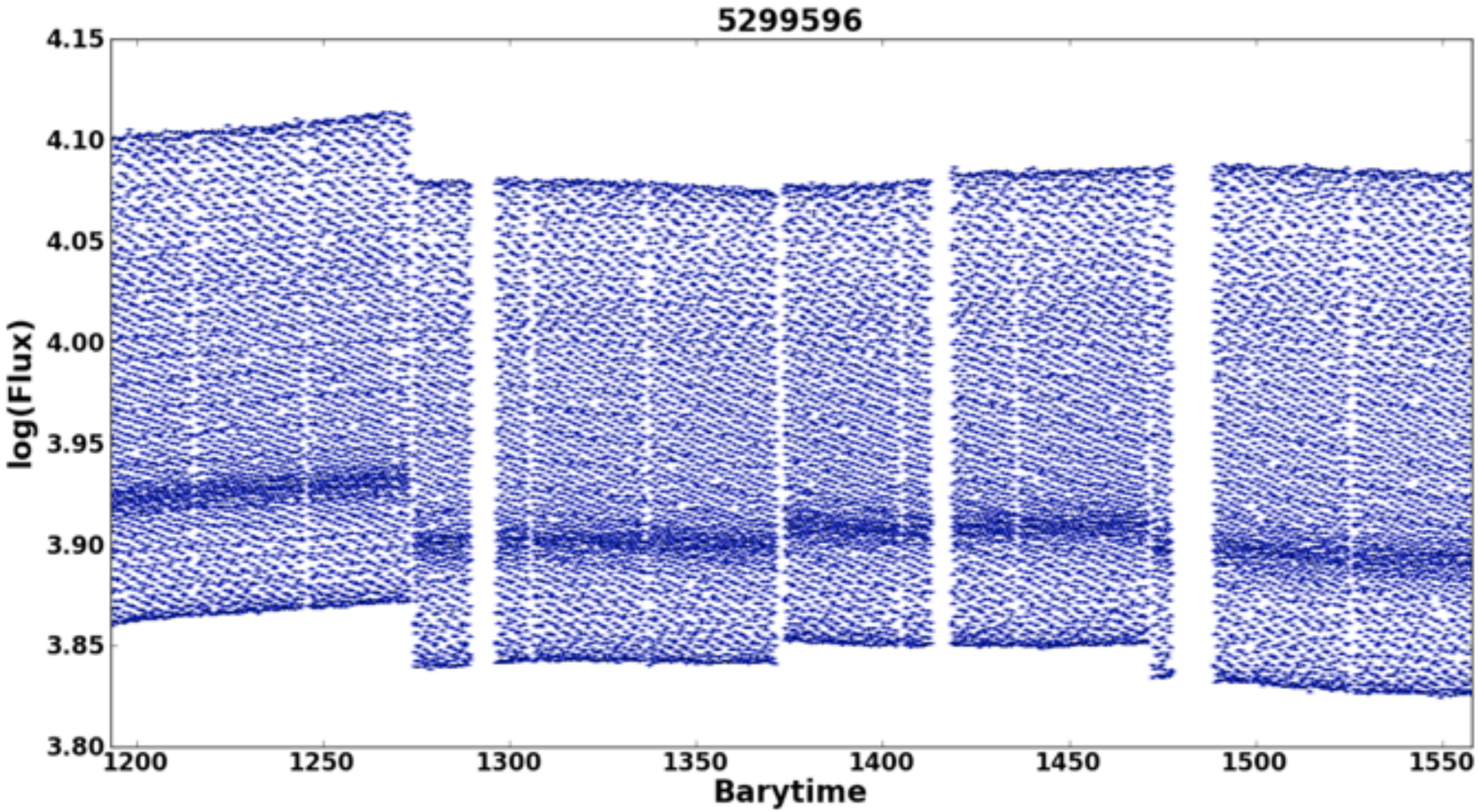
Large Synoptic Survey Telescope

LSST Data Volumes/Rates

Survey Property	Quantity
Data Rate	6.4 GB every 15s
Image Size	3.2 gigapixels
Visits per Night	1000
Data Rate per Day	15 TB
Processing Requirement	Latency of 60 seconds
Transients Rate	10k / visit; 10M / night
Objects Detected	24B galaxies, 14B stars
Photometric Measurements	32 Trillion
Final Data Volume	60 PB raw; 15 PB catalog

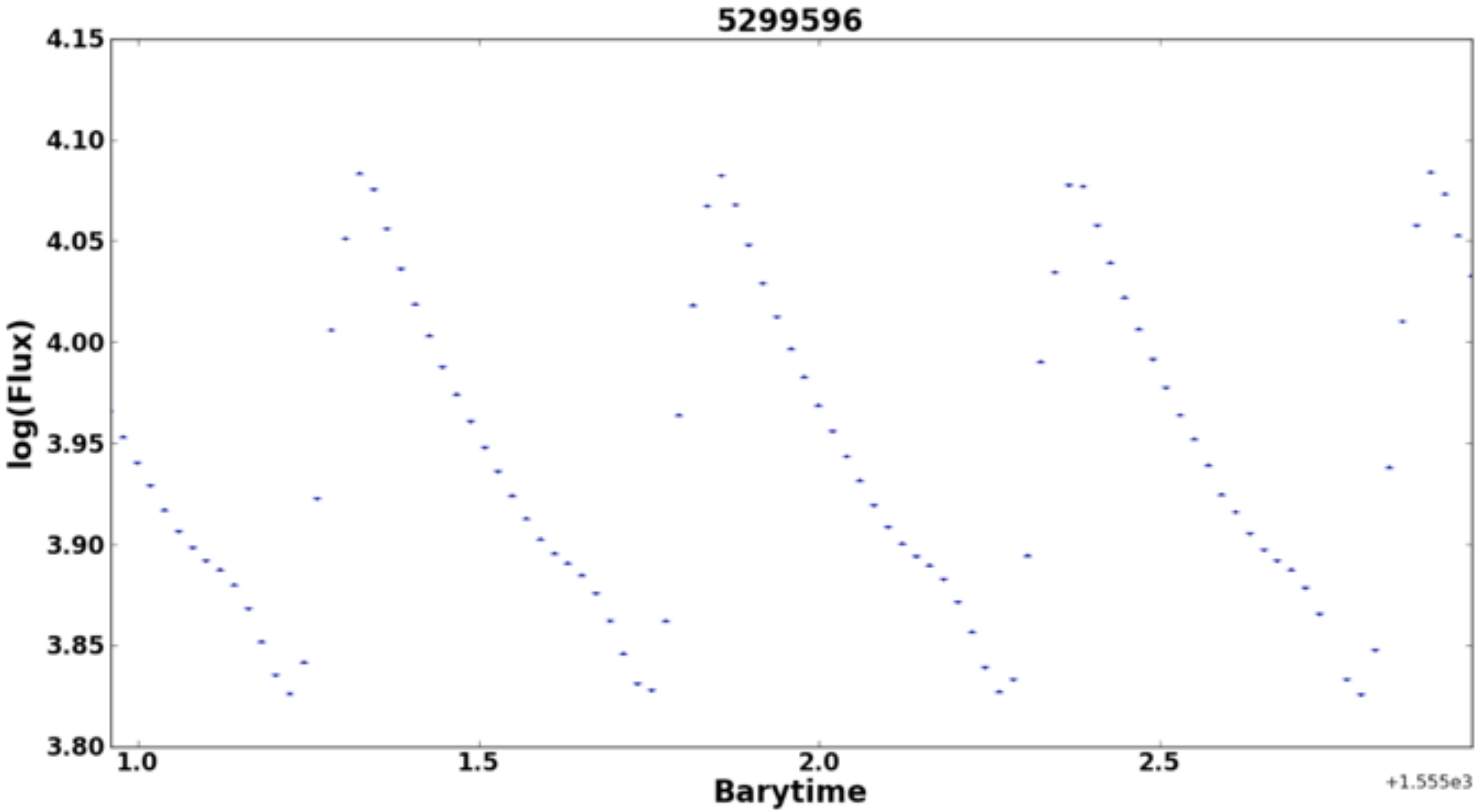
Astronomical Time-Series Data
Brightness vs. Time

1 year of data on variable star from Kepler spacecraft



Astronomical Time-Series Data
Brightness vs. Time

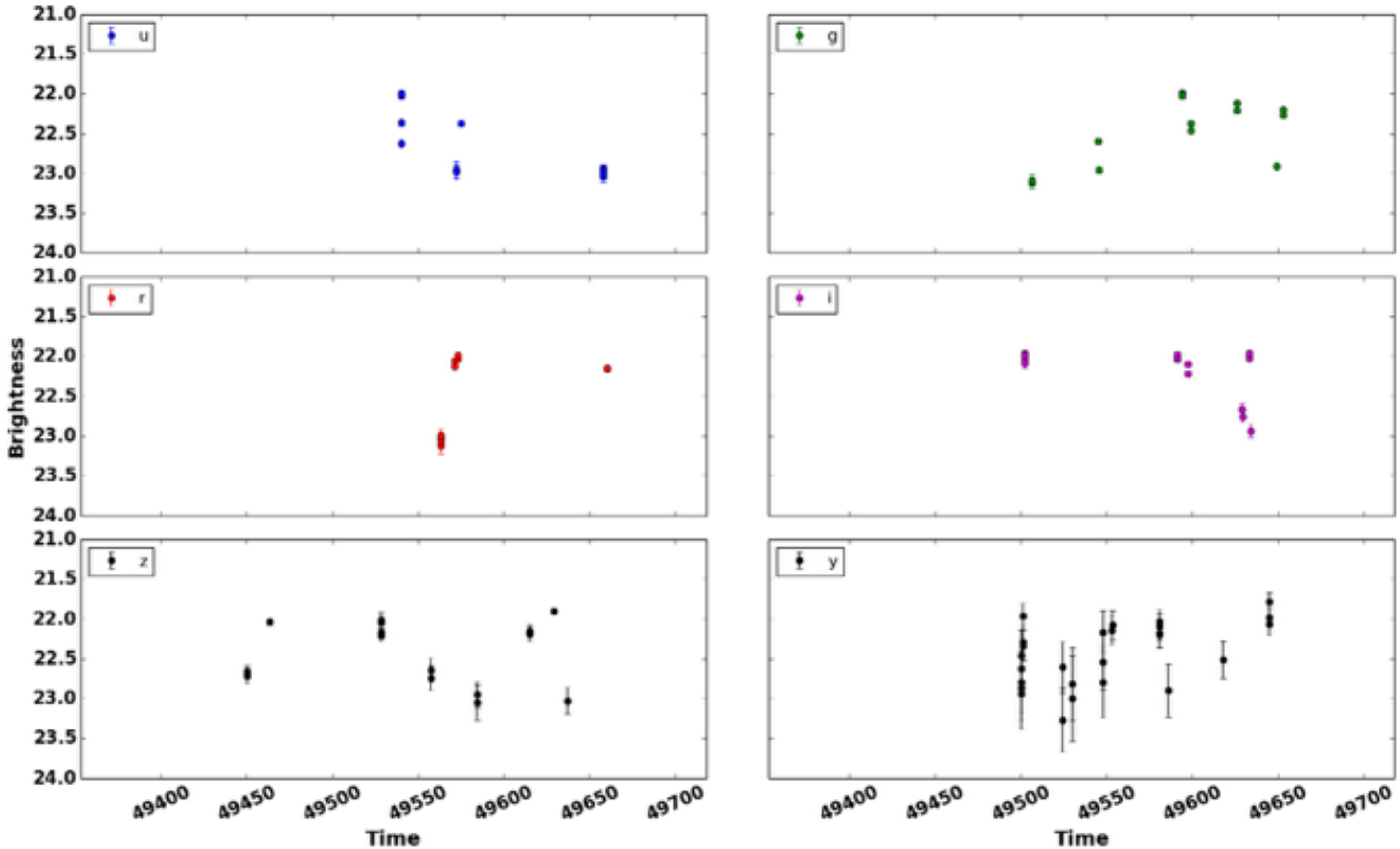
2 days of data from Kepler



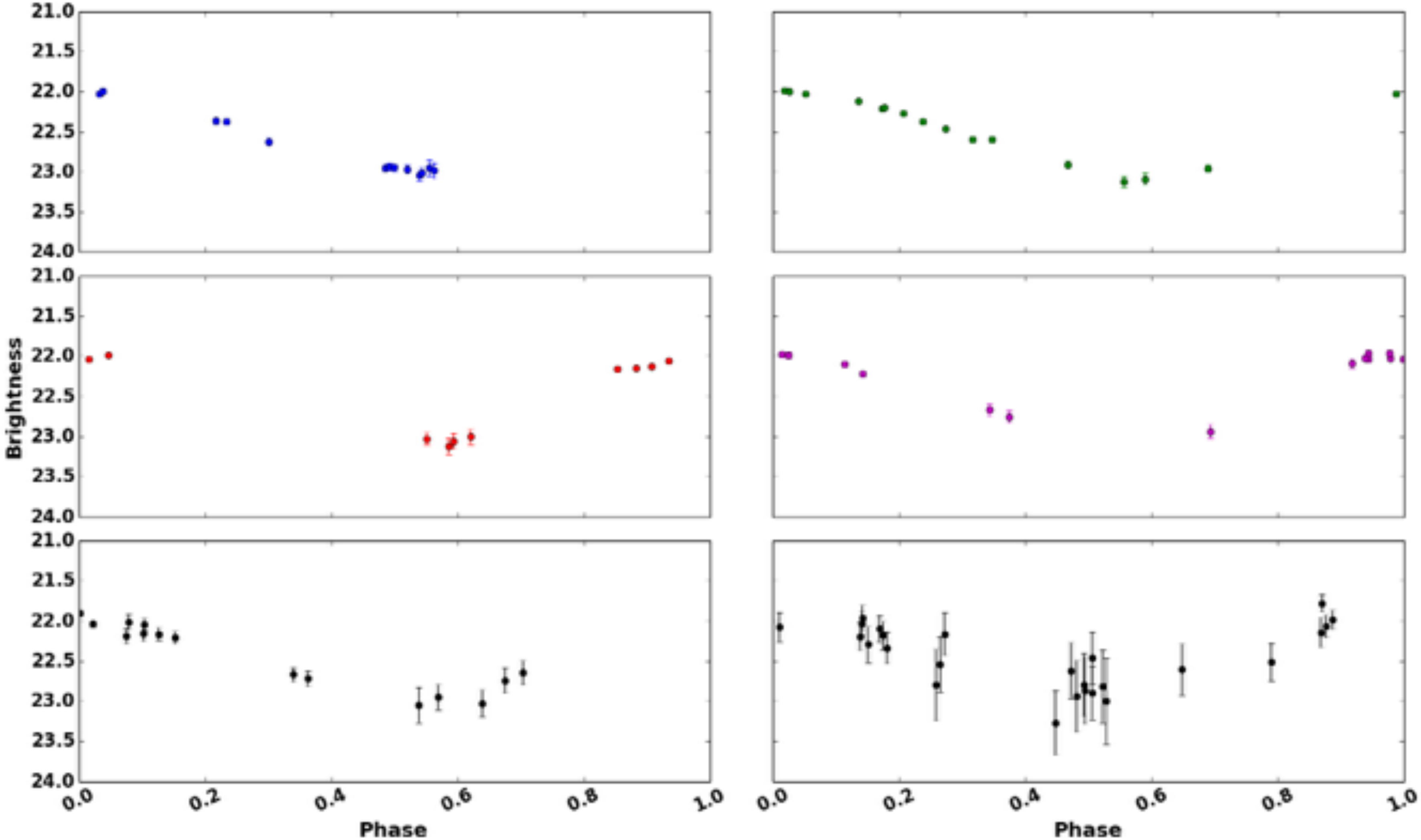
Astronomical Time-Series Data
Brightness vs. Time

1 year of data on same variable star from LSST

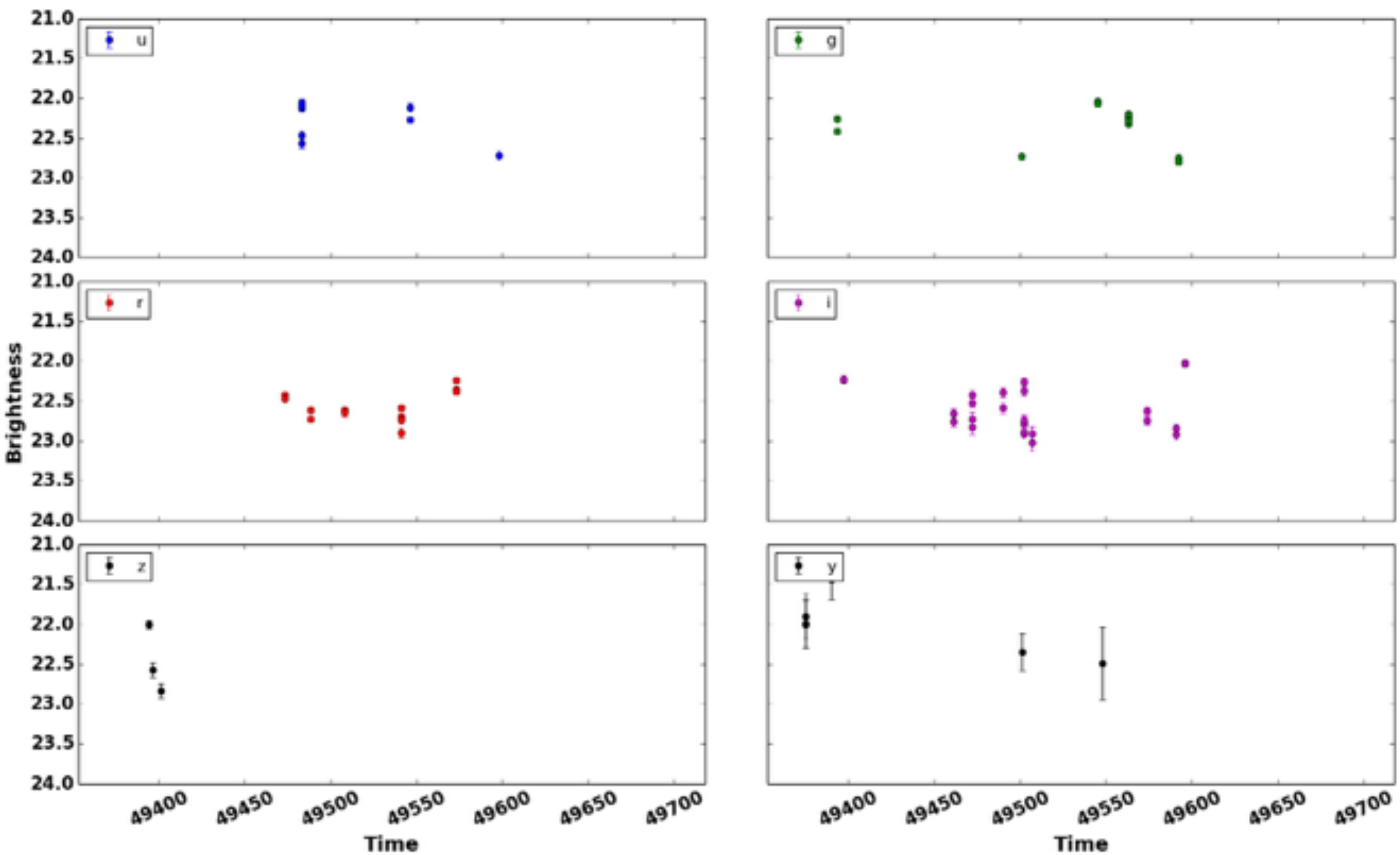
Sampling far less regular, and distributed across 6 channels



How to learn that folding the data on the correct period yields smooth lightcurve?



How to tell how similar that sparse lightcurve is to this one?



Lightcurve Classification

We need **machine learned classification** of light curves to optimally make use of the LSST Data Stream:

Detection and discovery of events in real time:
filtering of the LSST data stream into its transient components

Efficiencies, rates, number counts of populations:
Eclipsing binary stars (star formation and evolution)
RR Lyrae-type variables (Galactic structure)
Supernovae (cosmology; star formation)

Recognition of unexpected phenomena:
Outlier detection to find rare or new classes enabled by the survey volume

Driving autonomous follow-up networks:
Allowing machines to decide, in real-time, what is "important" and to study it in detail

J Richards

Lightcurve Classification

We need
optimally

Detection and discovery of events in real time:
filtering of the LSST data stream into its transient components

Efficiencies, rates, number counts of populations:
Eclipsing binary stars
RR Lyrae-type variables
Supernovae (cosmology; star formation)

Recognition of unexpected phenomena:
Outlier detection to find rare or new classes enabled by the survey volume

Driving autonomous follow-up networks:
Allowing machines to decide, in real-time, what is "important" and to study it in detail

Discovery on massive data streams is not assured

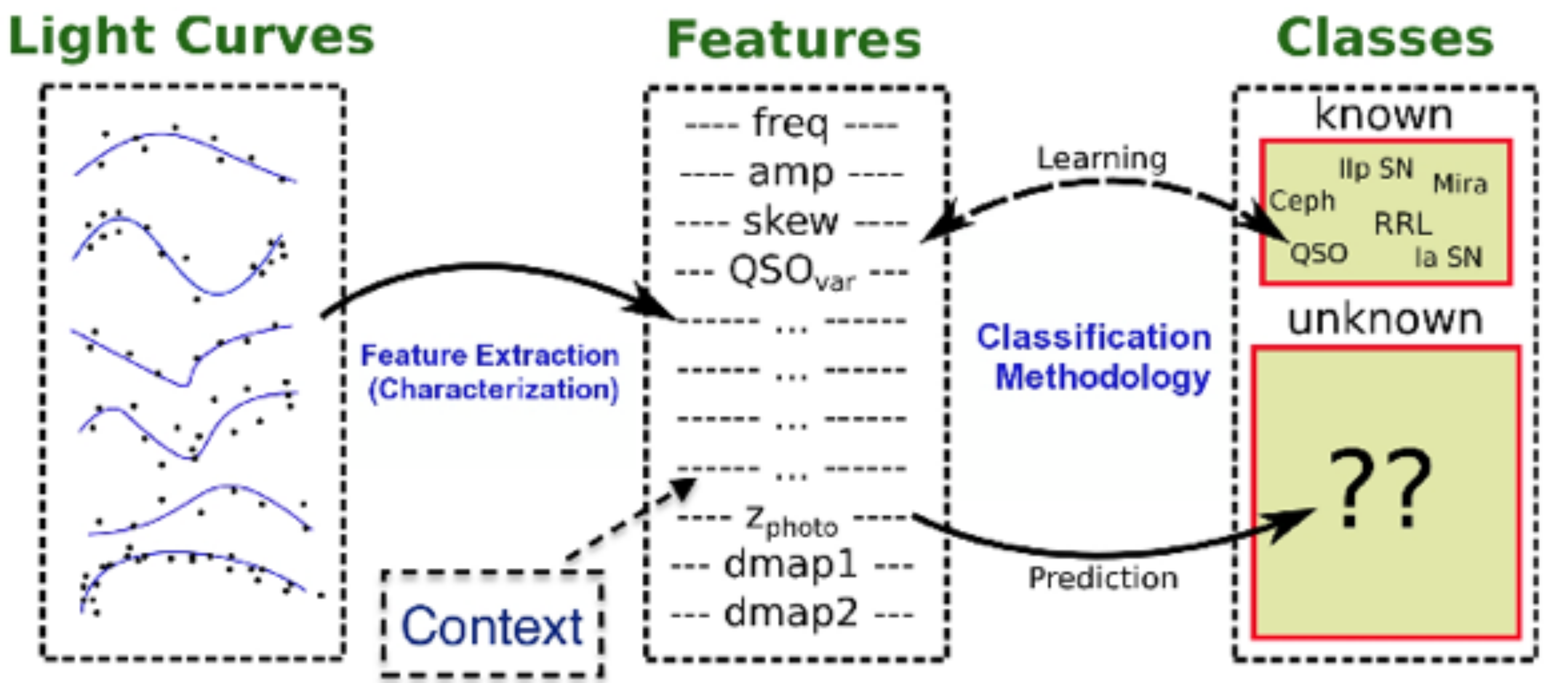
J Richards

Astrophysics as Applied Machine Learning

Single-band lightcurve classification

All techniques are based on "features" calculated from lightcurves
Features are ranked in importance, fed to binary decision trees or Bayesian Networks
Typically trained using vetted data sets (supervised learning). Domain knowledge is key

Roadmap for lightcurve classification



Richards et al., 2011

Astrophysics as Applied Machine Learning

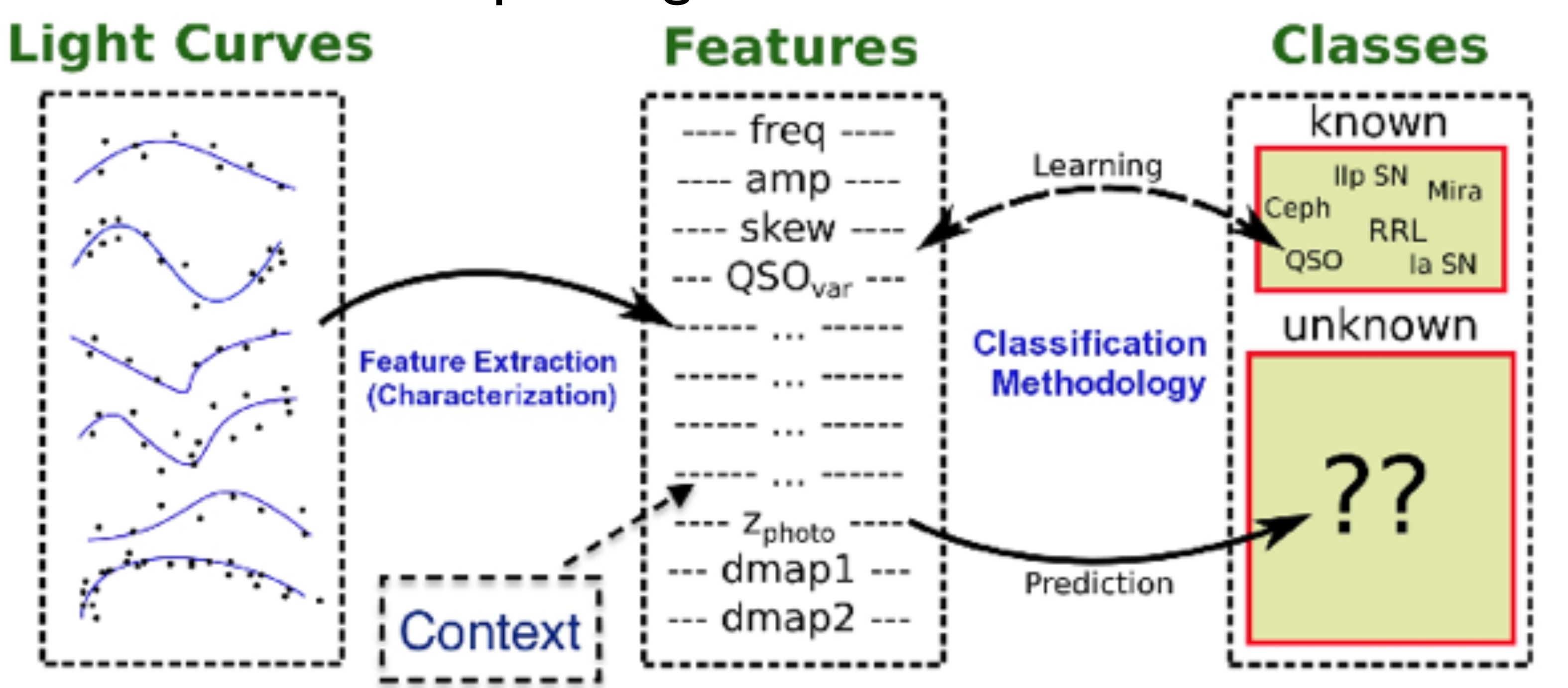
Single-band lightcurve classification

All techniques are based on "features" calculated from lightcurves

Features are ranked in importance, fed to binary decision trees or Bayesian Networks

Typically trained using vetted data sets (supervised learning). Domain knowledge is key

Roadmap for lightcurve classification



Richards et al., 2011

LSST: 10000-D space with 40 billion sources

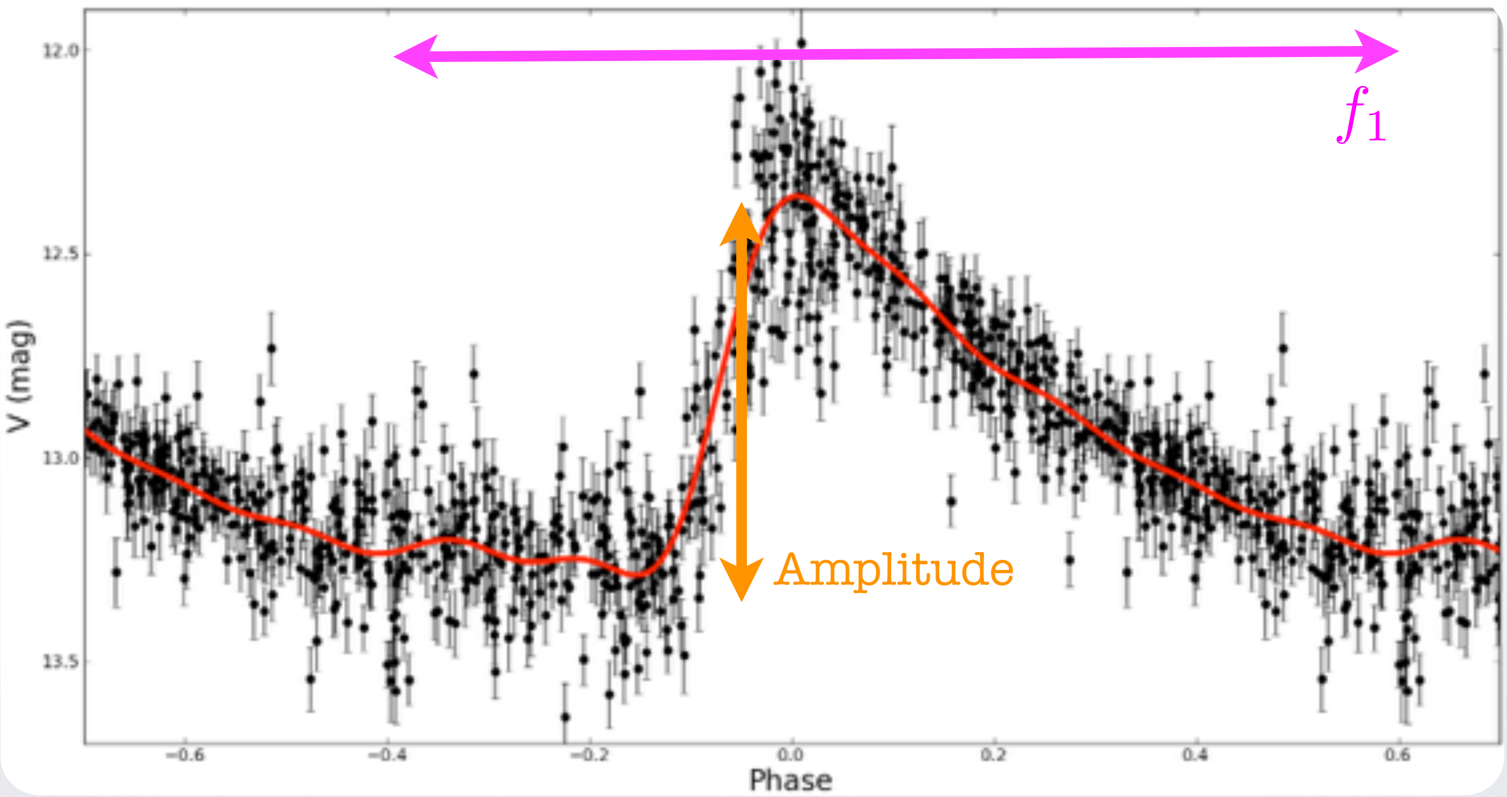
Single-band lightcurve classification

All techniques are based on “features” calculated from lightcurves

Features are ranked in importance, fed to binary decision trees or Bayesian Networks

Typically trained using vetted data sets (supervised learning). Domain knowledge is key

Feature Selection



Courtesy of Adam Miller

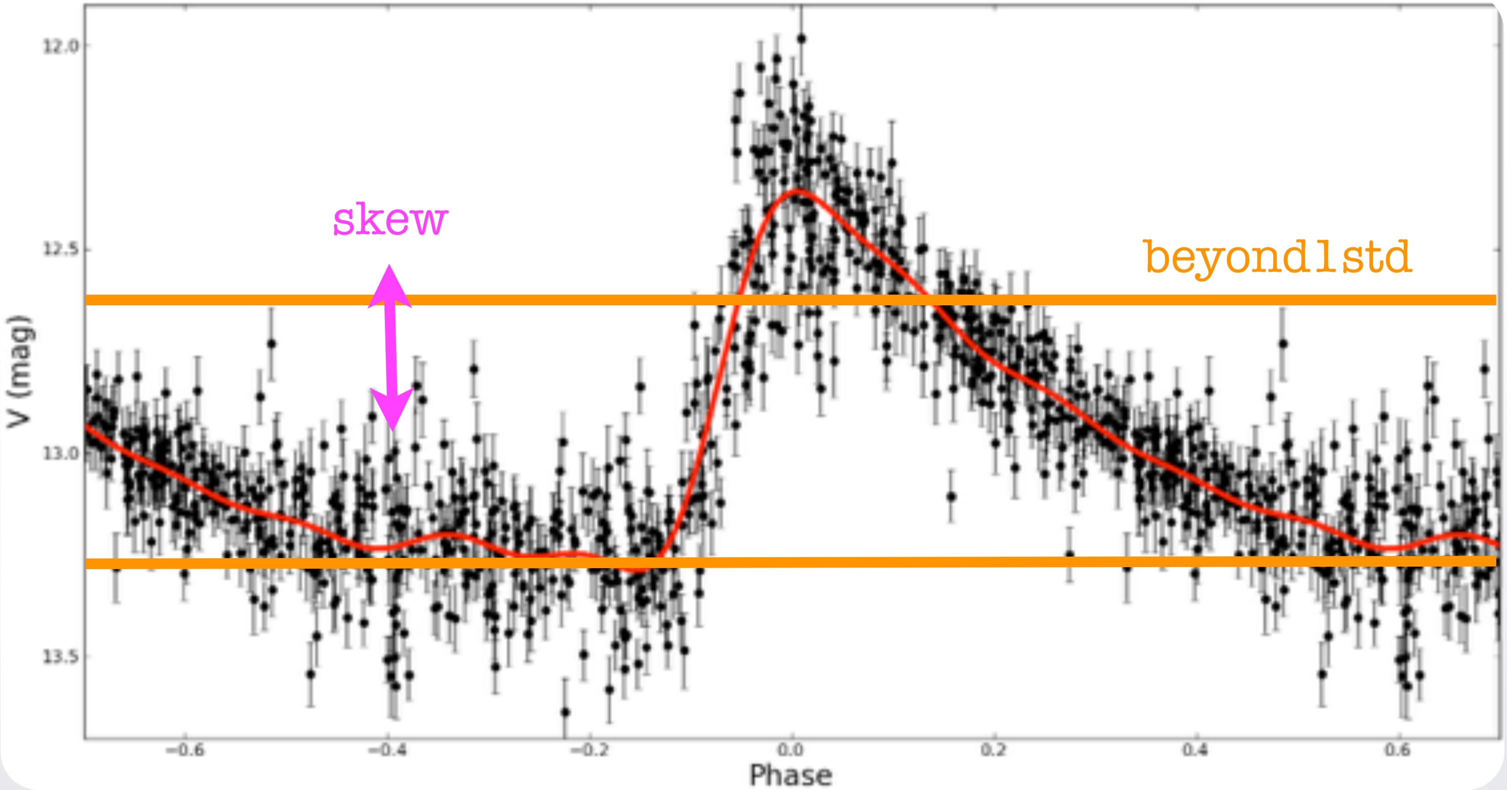
Single-band lightcurve classification

All techniques are based on "features" calculated from lightcurves

Features are ranked in importance, fed to binary decision trees or Bayesian Networks

Typically trained using vetted data sets (supervised learning). Domain knowledge is key

Feature Selection

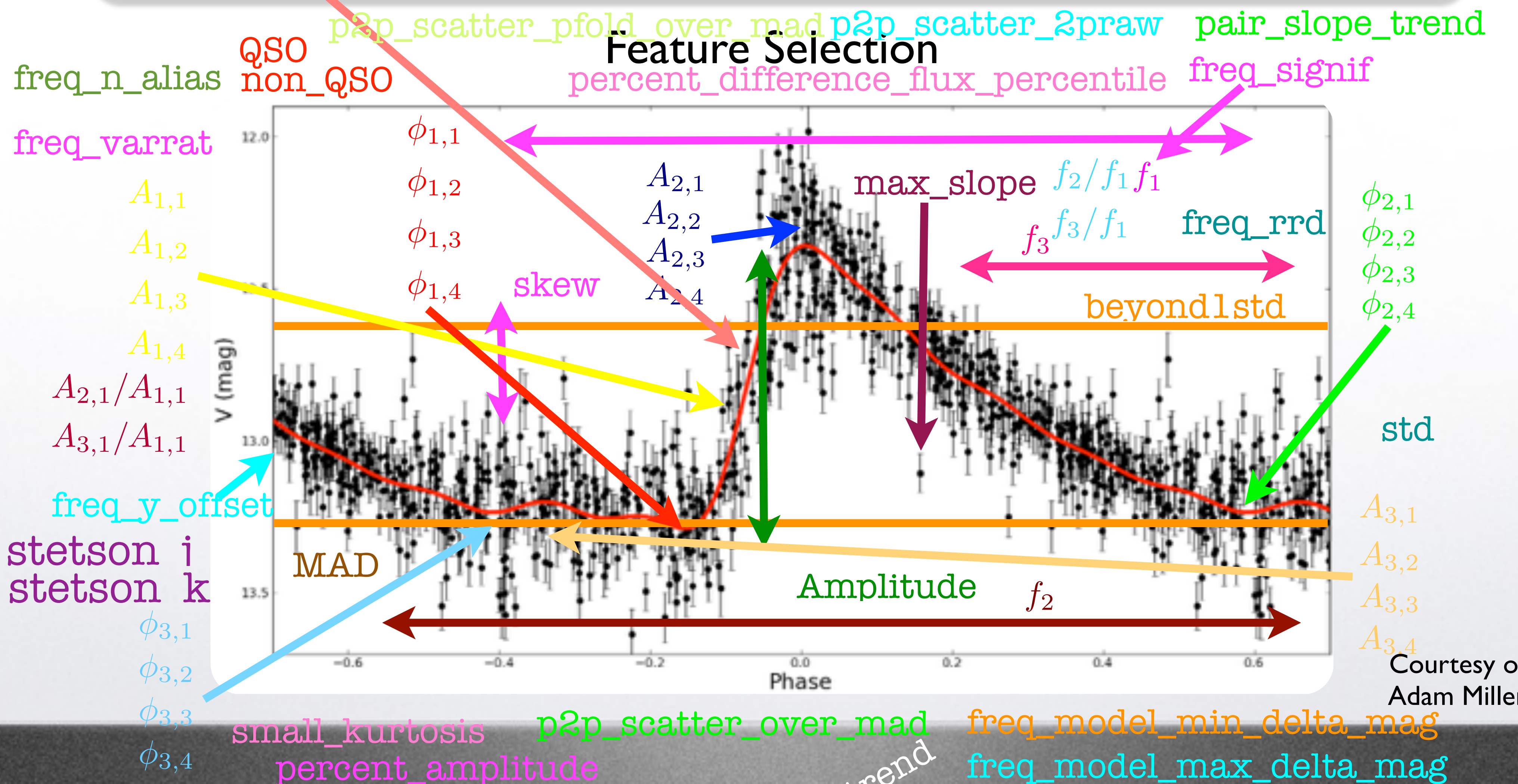


Courtesy of Adam Miller

Astrophysics as Applied Machine Learning

Single-band lightcurve classification

- All techniques are based on "features" calculated from lightcurves
- Features are ranked in importance, fed to binary decision trees or Bayesian Networks
- Typically trained using vetted data sets (supervised learning). Domain knowledge is key



Courtesy of Adam Miller

Not Extensible to LSST Scale

- Want unique or minimal set of features

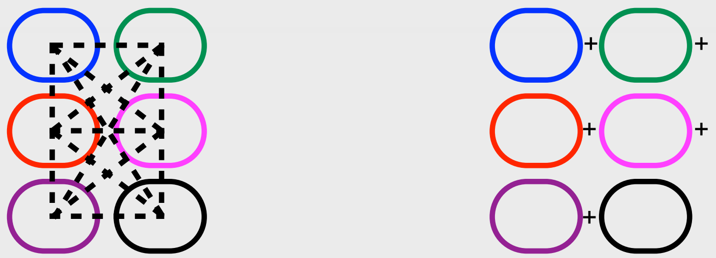
  Focus compute resources

- No techniques in place for irregularly sampled data in multiple channels

  Is tonights data in the g-passband deviant, given last nights data in the r-passband?

Need Models that can Accommodate this Complexity

One 6-passband model vs. six 1-passband models



Challenges for Time-Series Modeling

- How quickly can you recognize the class of object
- How quickly can you recognize the uniqueness of object
- How quickly can you recognize when the object changes its behavior
- How often are you wrong, and what are the consequences
- How often are you right, and what are the consequences

Challenges for Time-Series Modeling

Fast = Tight coupling of [software, data, hardware] and scalable algorithms

- How often are you wrong, and what are the consequences
- How often are you right, and what are the consequences

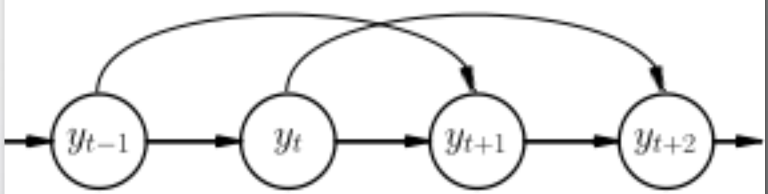
Challenges for Time-Series Modeling

Fast = Tight coupling of [software, data, hardware] and scalable algorithms

Right = Calibrated probabilities and forecasting uncertainties

General Time-Series Models

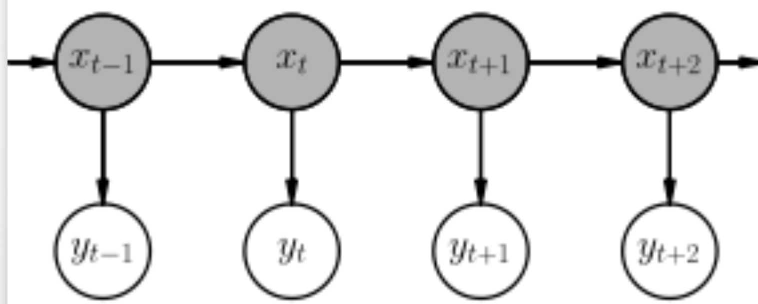
Historically, probabilistic time-series stochastic models are based on **autoregressive** and/or **state-based** methods.



(a) Autoregressive model

General Time-Series Models

Historically, probabilistic time-series stochastic models are based on **autoregressive** and/or **state-based** methods.



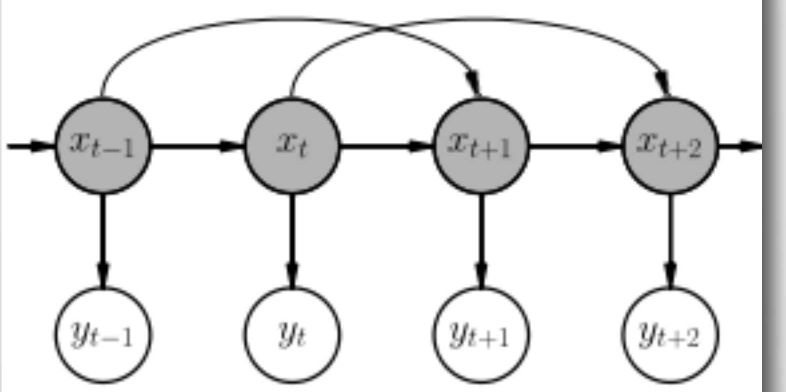
(b) State space model

Latent Variable

Observed Variable

General Time-Series Models

Historically, probabilistic time-series stochastic models are based on **autoregressive** and/or **state-based** methods.



(c) Autoregressive state space model

Latent Variable

Observed Variable

CARMA Models

Continuous Autoregressive Moving Average

A zero-mean CARMA(p, q) process y(t) is defined to be a solution to the stochastic differential equation:

$$\frac{d^p y(t)}{dt^p} + \alpha_{p-1}\frac{d^{p-1}y(t)}{dt^{p-1}} + \ldots + \alpha_0 y(t) = \beta_q \frac{d^q \epsilon(t)}{dt^q} + \beta_{q-1}\frac{d^{q-1}\epsilon(t)}{dt^{q-1}} + \ldots + \epsilon(t)$$

α are the autoregressive coefficients to order p

β are the moving average coefficients to order q

ε is a zero-mean normal (white noise) innovation

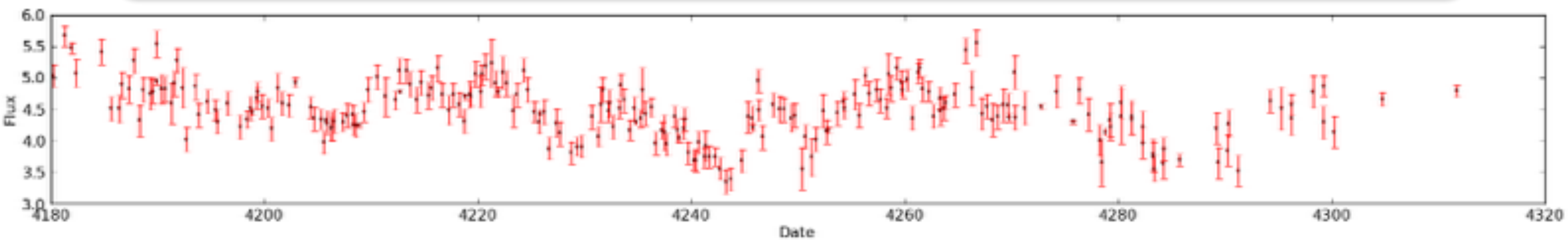
Kelly, Becker, et al., 2014

Models lightcurve as sum of (deterministic) autoregression plus (random) stochastic noise

Gaussian Process with an autoregressive covariance matrix

Features fully describe lightcurve

Evaluation of likelihood function scales as number of data points N



CARMA Models

Continuous Autoregressive Moving Average

A zero-mean CARMA(p, q) process y(t) is defined to be a solution to the stochastic differential equation:

$$\frac{d^p y(t)}{dt^p} + \alpha_{p-1}\frac{d^{p-1} y(t)}{dt^{p-1}} + \ldots + \alpha_0 y(t) = \beta_q \frac{d^q \epsilon(t)}{dt^q} + \beta_{q-1}\frac{d^{q-1}\epsilon(t)}{dt^{q-1}} + \ldots + \epsilon(t)$$

α are the autoregressive coefficients to order p

β are the moving average coefficients to order q

ε is a zero-mean normal (white noise) innovation

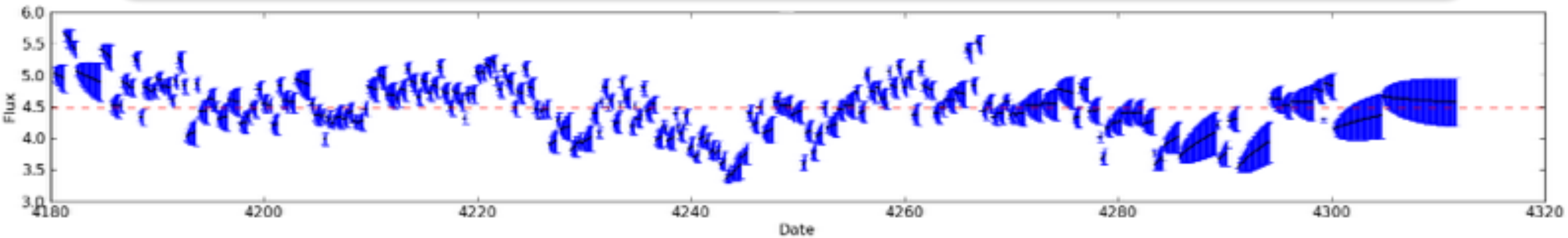
Kelly, Becker, et al., 2014

Models lightcurve as sum of (deterministic) autoregression plus (random) stochastic noise

Gaussian Process with an autoregressive covariance matrix

Features fully describe lightcurve

Evaluation of likelihood function scales as number of data points N



Extending Models Using Bayesian Inference

- We are able to fully define the likelihood function of a CARMA model for a measured time series
- With a suitable set of priors, this enables us to perform Bayesian inference
  - ln(posterior) = ln(prior) + ln(likelihood)
  - Provides a probability distribution of the CARMA process, given data
  - Allows a rigorous assessment of uncertainties by looking at multiple models drawn from the posterior distribution

Measurement Level Model :

$$y_i = \beta x_i + \delta_i$$

State/Population Level Model :

$$x_i = e^{\alpha (t_i - t_{i-1})} x_{i-1} + \eta_i$$

Hyperprior Level :

Priors on $\alpha, \beta, \delta, \eta$

Markov Chain Sampling of Posterior

Important for non-Gaussian or multi-modal posteriors

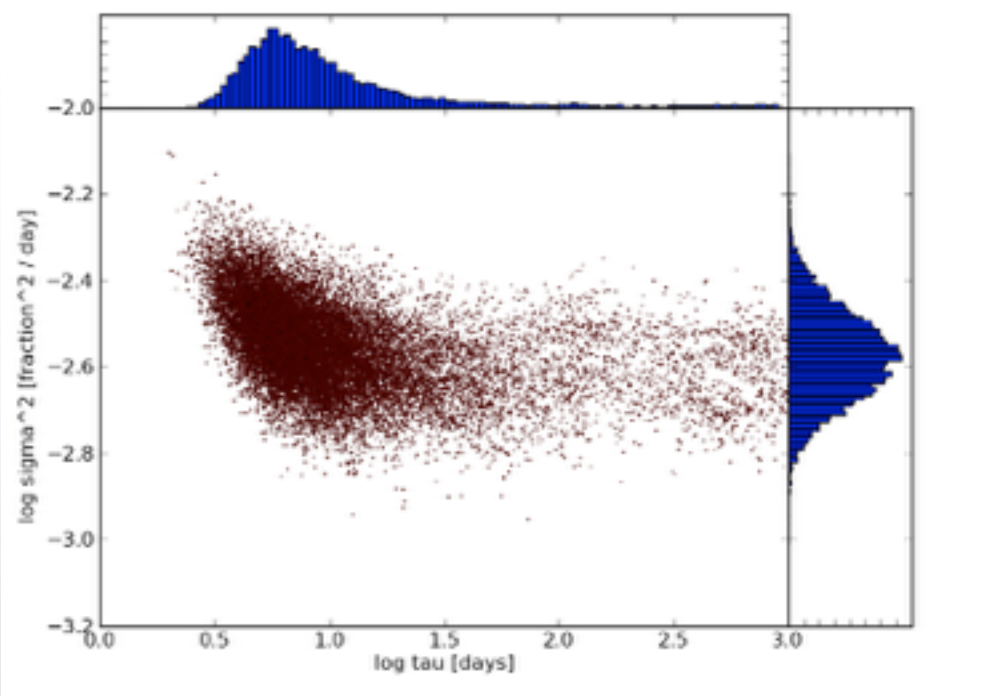
Sample from posterior using random walk algorithm

Acceptance rule designed to converge to stationary distribution

Parallel tempered MCMC to sample from multiple modes

Trivial to marginalize over nuisance parameters

CARMA(1,0) Parameters

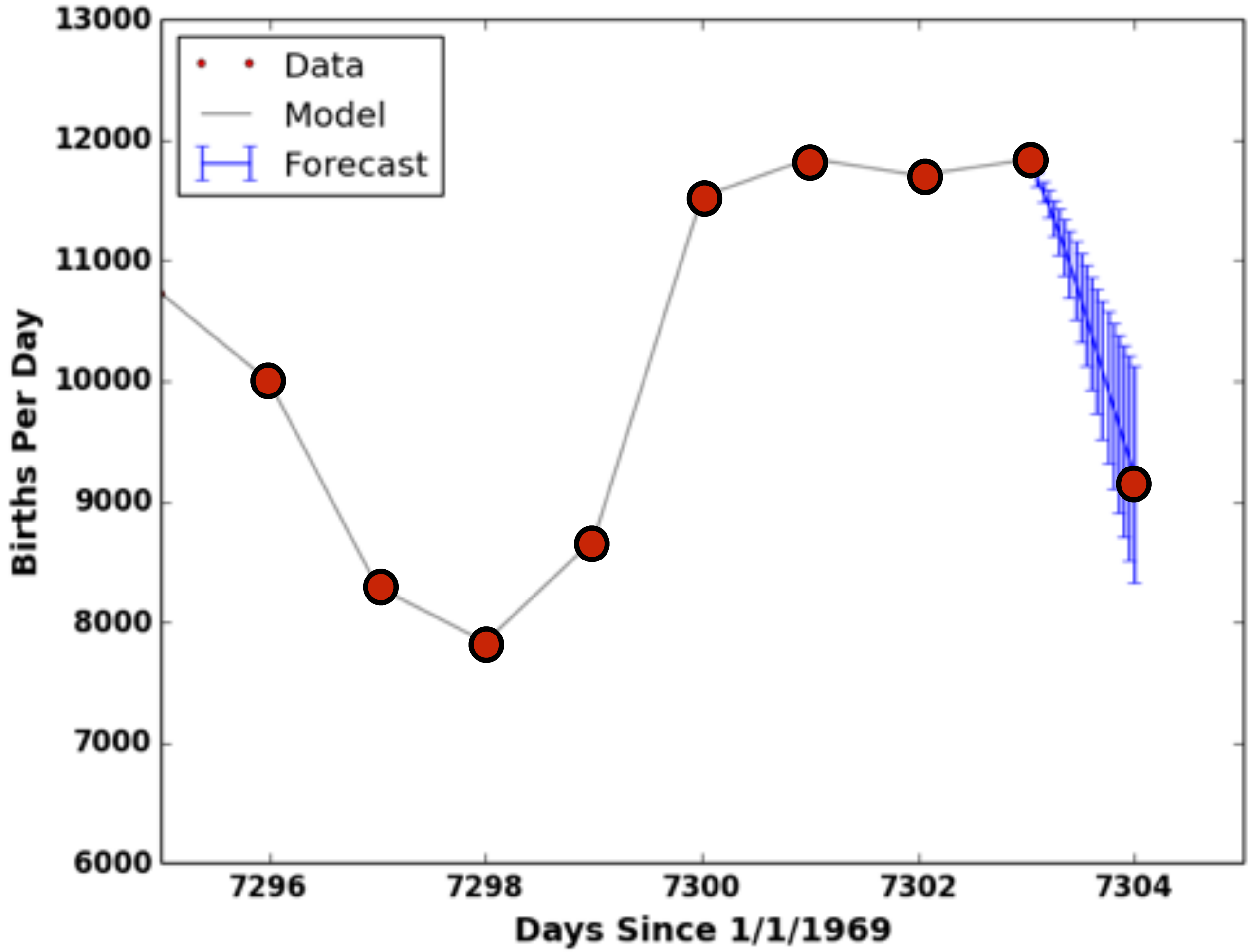


Computationally expensive: order $10^{5\text{-}6}$ function evaluations

Less sensitive to overfitting

Forecasting Options

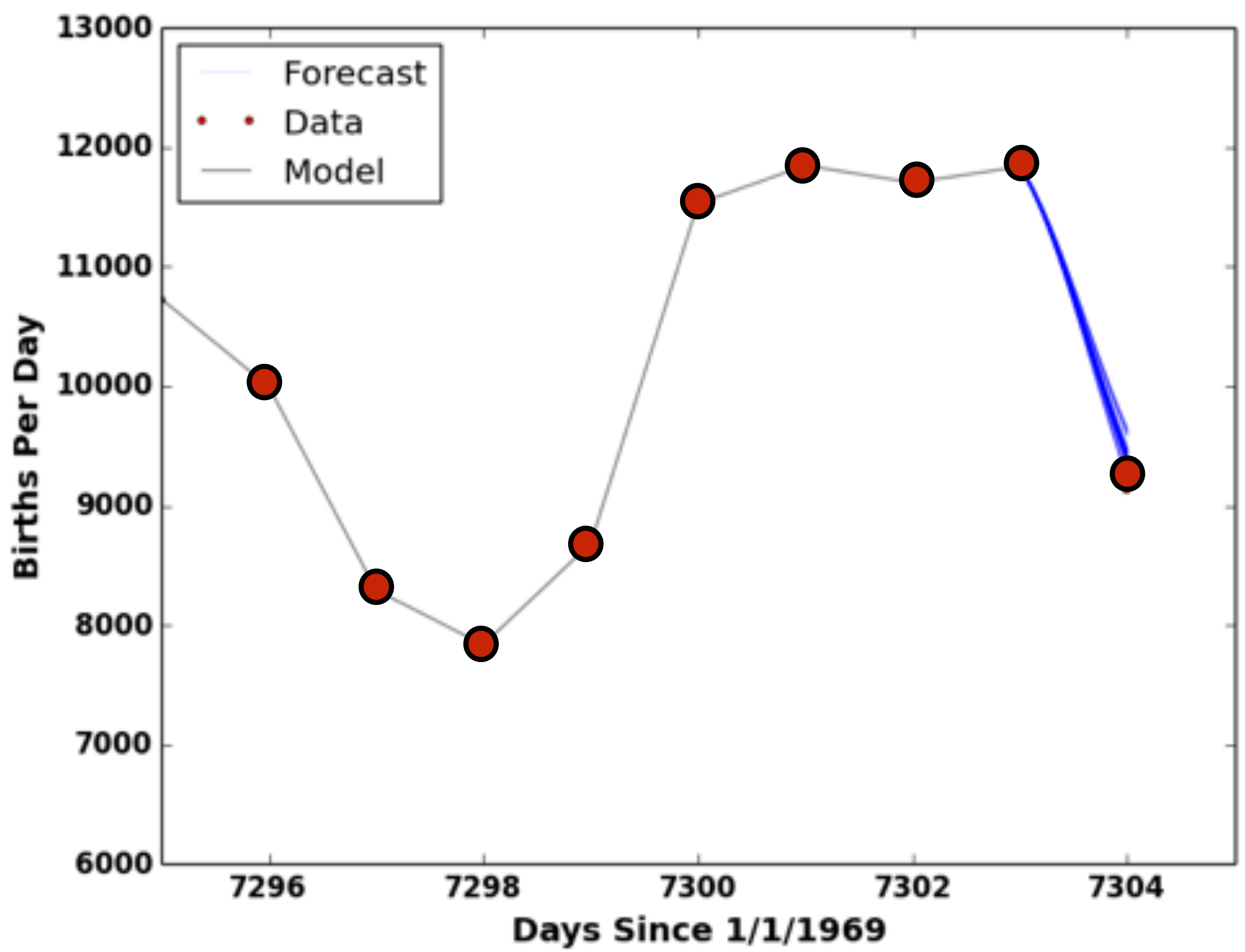
Conditional on data *and* maximum-likelihood model parameters



Looking at "best fit" model

Forecasting Options

Conditional on data *alone* ; draw 100 models from posterior

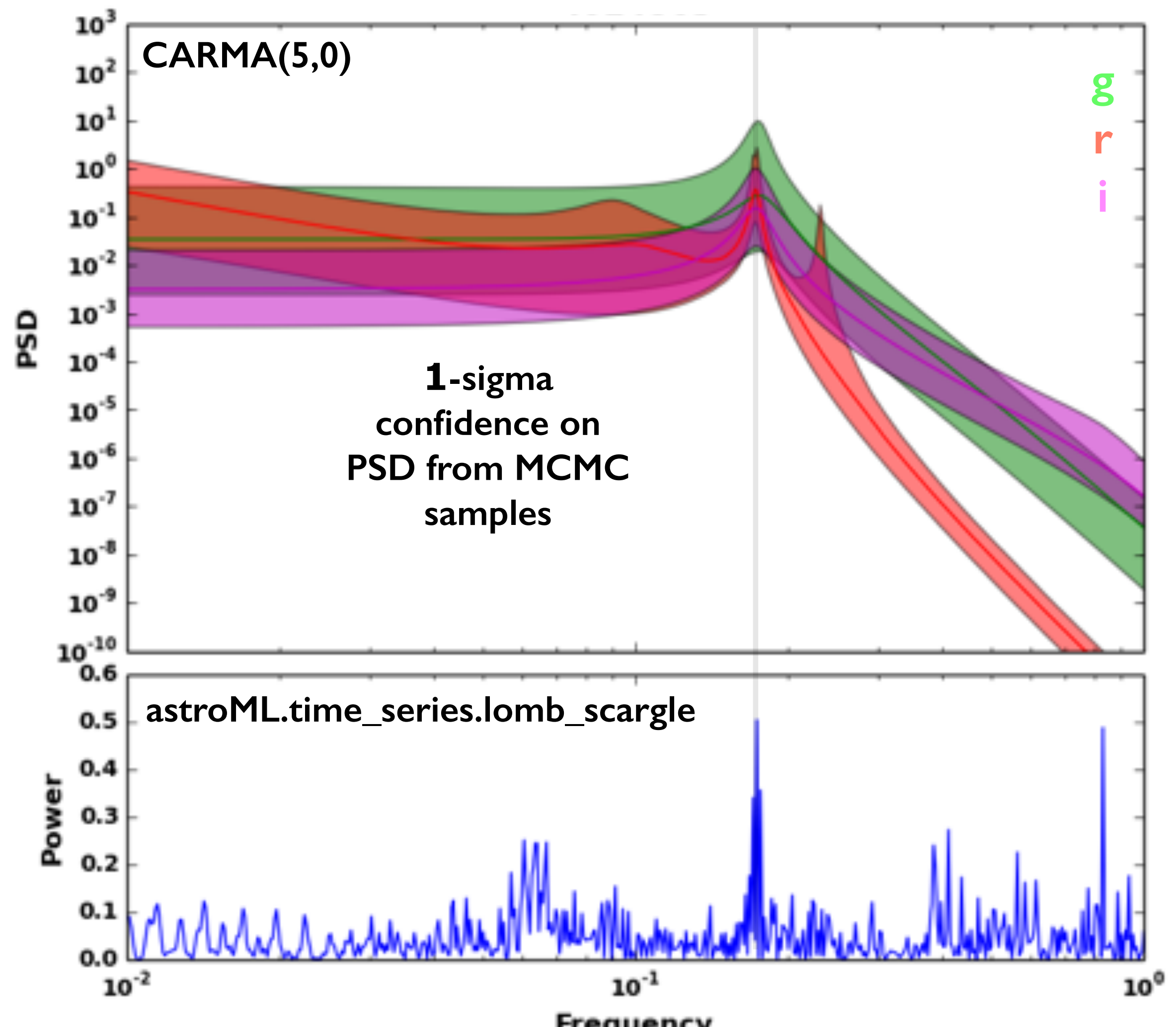


Sampling from posterior of CARMA process

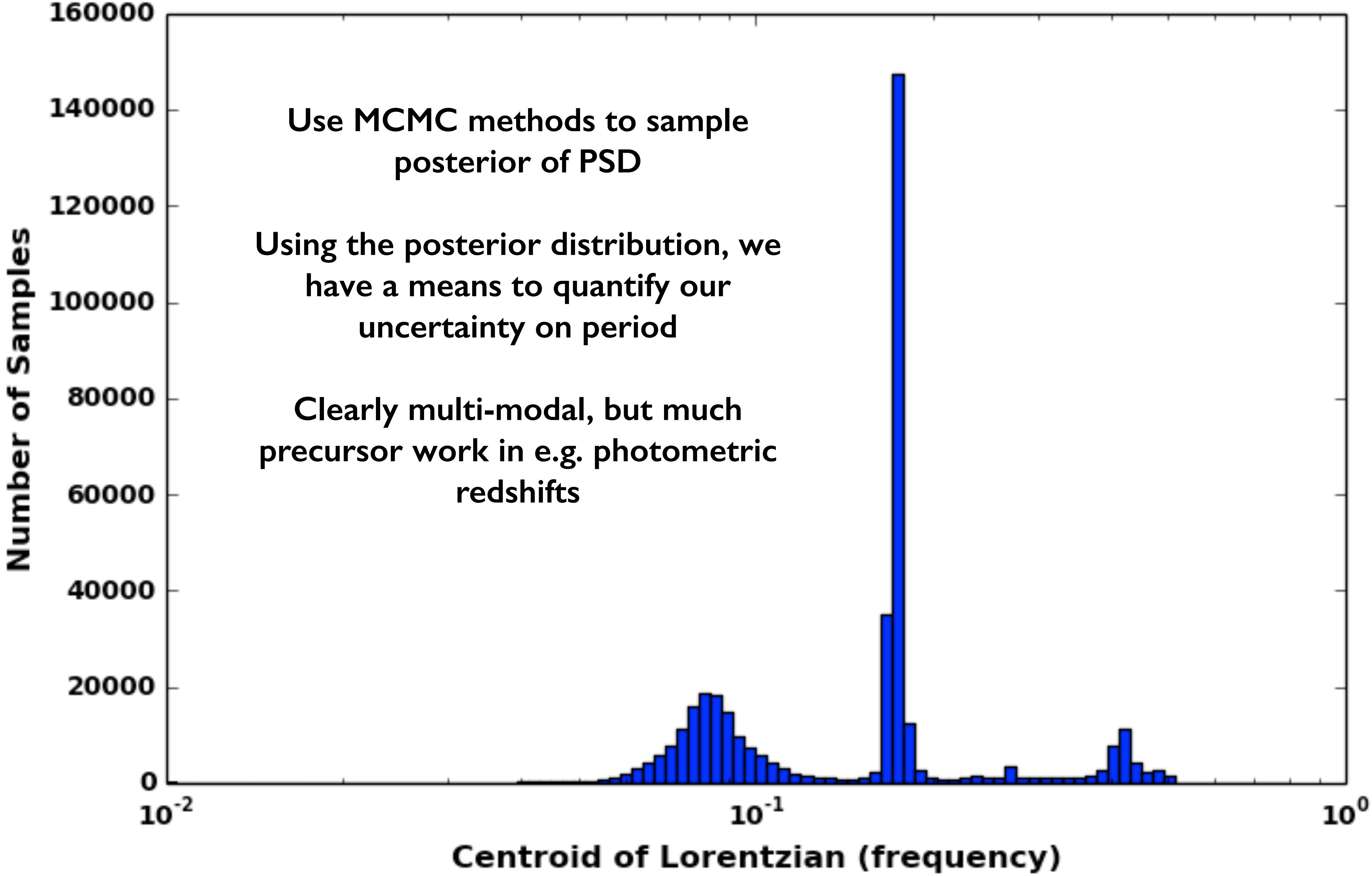
CARMA PSD vs. Periodogram

CARMA model parameters can be directly used to describe the power spectral density of the lightcurve.

CARMA on lightcurve of known periodically variable star



Posterior Probability Distribution on Period

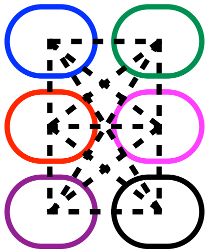


Use MCMC methods to sample posterior of PSD

Using the posterior distribution, we have a means to quantify our uncertainty on period

Clearly multi-modal, but much precursor work in e.g. photometric redshifts

ARMA Family Capabilities



Modeling formalism flexible enough to include correlations between channels, not just within channels

VARMA(p, q): Vector autoregression with a moving average

MCARMA(p,q): Multivariate CARMA(p,q)

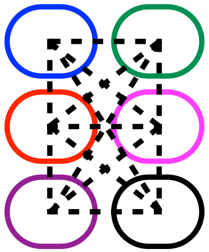
α,β of CARMA model are matrices vs. scalars

	u	g	r	i	z
u	uu	ug	ur	ui	uz
g		gg	gr	gi	gz
r			rr	ri	rz
i				ii	iz
z					zz

Also $PARMA_{\nu}(p,q)$: Periodic with period ν

ARMA Family Capabilities

Single phenomenological model using all available data



Flexible through choice of autoregessive and moving average orders

Scalable through order(N) scaling of likelihood function, and online model update

Calibrated forecasting using Bayesian techniques

Uncertainties come from distribution of models, not just a single model