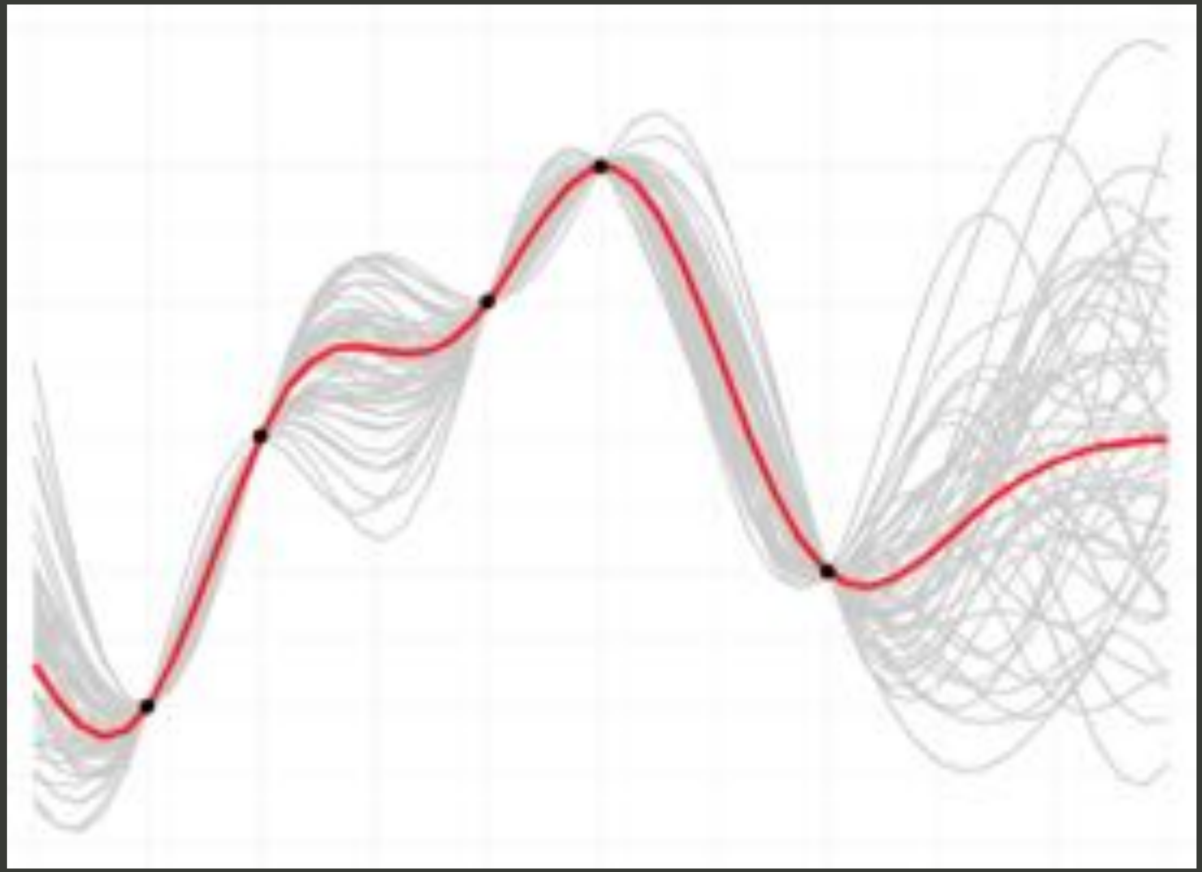
Machine Learning+Time-Series and Gaussian Process Regression



credit: datascience.stackexchange

Adam A Miller

Northwestern/Adler Planetarium

2018 IDEAS Course

7 Mar 2018

CIERA
CENTER FOR INTERDISCIPLINARY EXPLORATION AND RESEARCH IN ASTROPHYSICS

ADLER PLANETARIUM

LSST

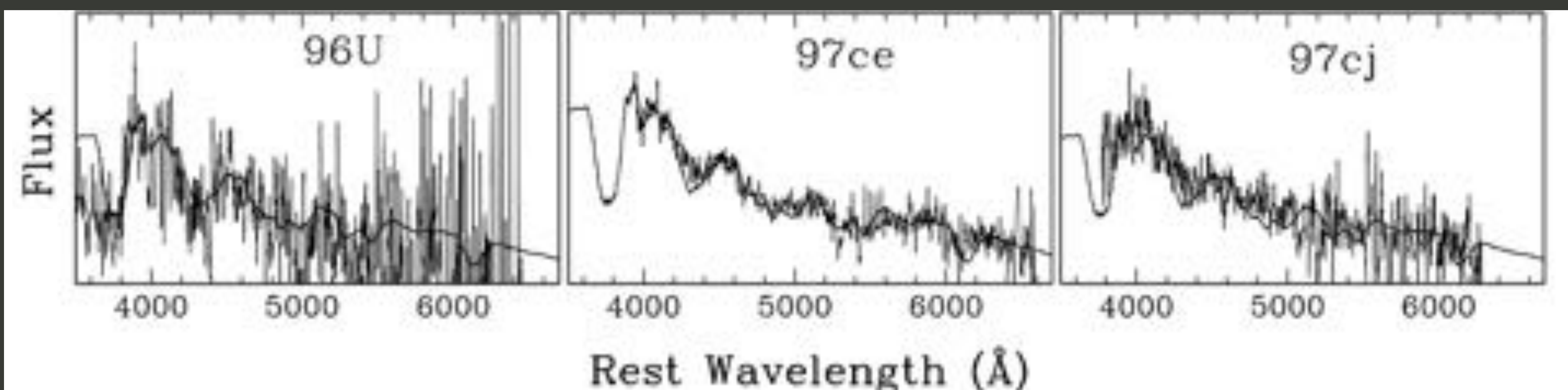
SNe with LSST

LSST will discover ~2000 new SNe per night

Vast majority will be faint (m > 23 mag)

~1000 hr/night on 8-m class telescopes needed for spec

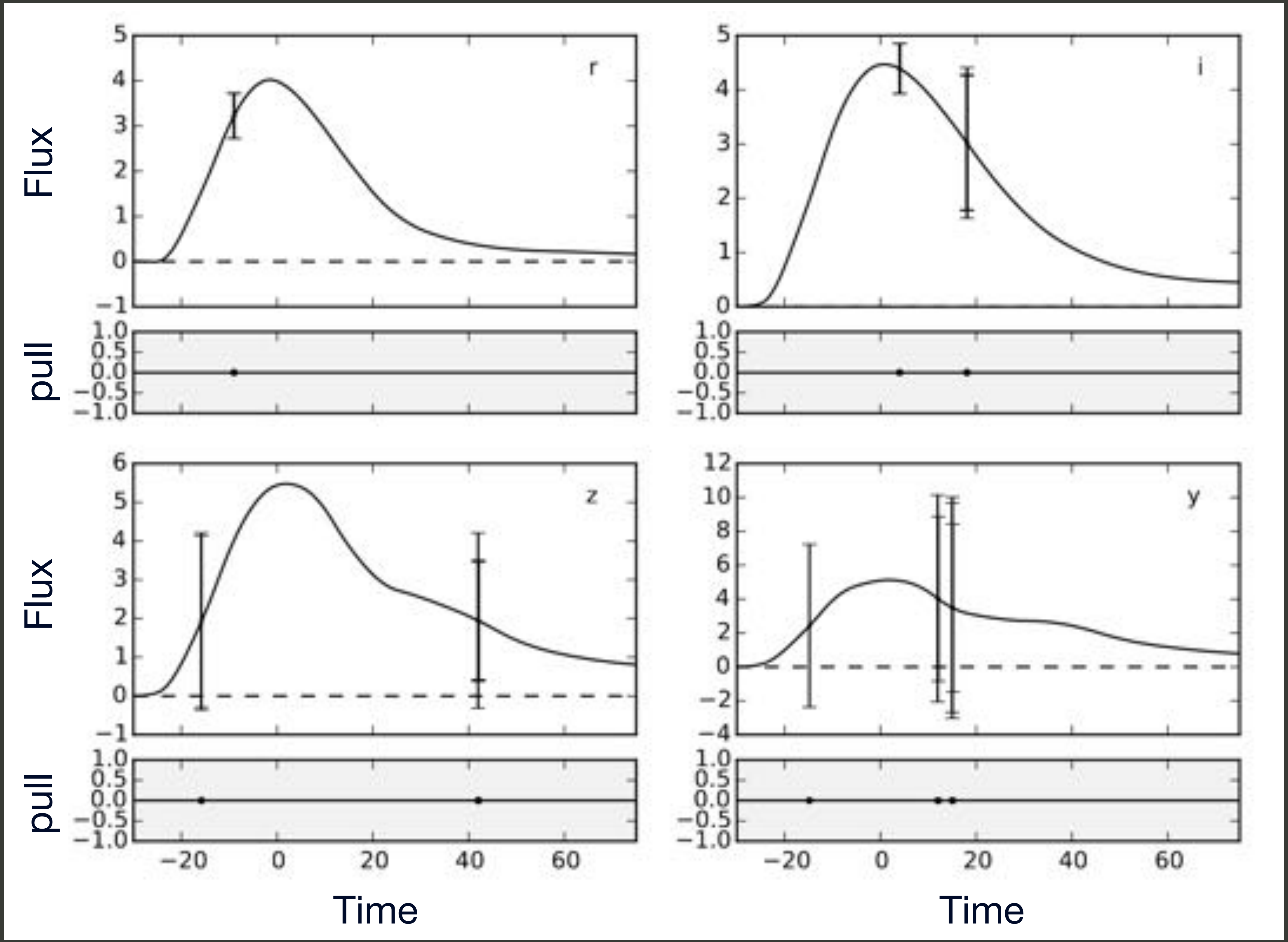
~100 hr available on 8-m class telescope



Riess+98

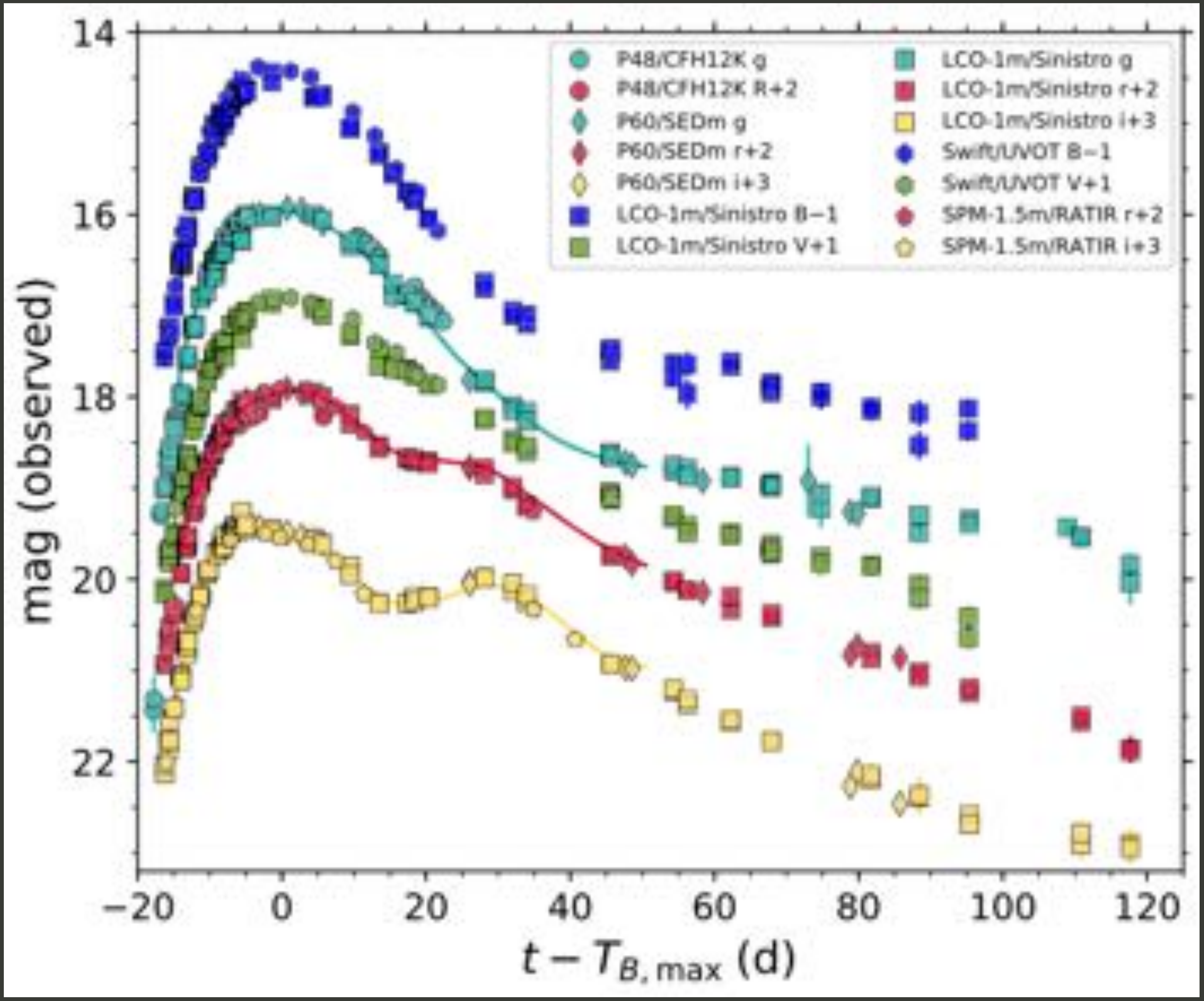
LSST will primarily be a photometric-only transient survey

SNe with LSST



Marshall+16

SNe with LSST



AAM+18

Real-Time Data from LSST

Application Layer -
Generates open, accessible data products with fully documented quality

Processing Cadence	Image Category (files)	Catalog Category (database)	Alert Category (database)
Nightly "Level 1"	Raw science image Calibrated science image Subtracted science image Noise image Sky image Data quality analysis	Source catalog (from difference images) Object catalog (from difference images) Orbit catalog Data quality analysis	Transient alert Moving object alert Data quality analysis
Data Release (Annual) "Level 2"	Stacked science image Template image Calibration image RGB JPEG Images Data quality analysis	Source catalog (from calibrated science images) Object catalog (optimally measured properties) Data quality analysis	Alert statistics & summaries Data quality analysis

table credit: L. M. Walkowicz

Real-Time Data from LSST

Large Synoptic Survey Telescope

Data Products

http:// ls.st / dpdd

Alerts: 1-10 million/night, issued in 60 sec Orbits for 6 million solar system objects	Level 1 *Nightly*
Catalogs: ~37 billion objects (20B galaxies, 17b Stars); ~7 trillion "sources", ~30 trillion "forced sources" Deep co-added images	Level 2 *Annual*
Services/computing resources at Data Access Centers Software & APIs to enable development of analysis codes	Level 3 *Community*

table credit: L. M. Walkowicz

Real-Time Data from LSST

position

flux, size, and shape

light curves in all bands
(up to a ~year; stretch: all)

variability characterization
(e.g. low-order light-curve moments,
probability the object is variable)
cut-outs centered on the object
(template, difference image)

credit: L. M. Walkowicz

Real-Time Data from LSST

Identification ≠ Discovery

For LSST identification is not sufficient without classification

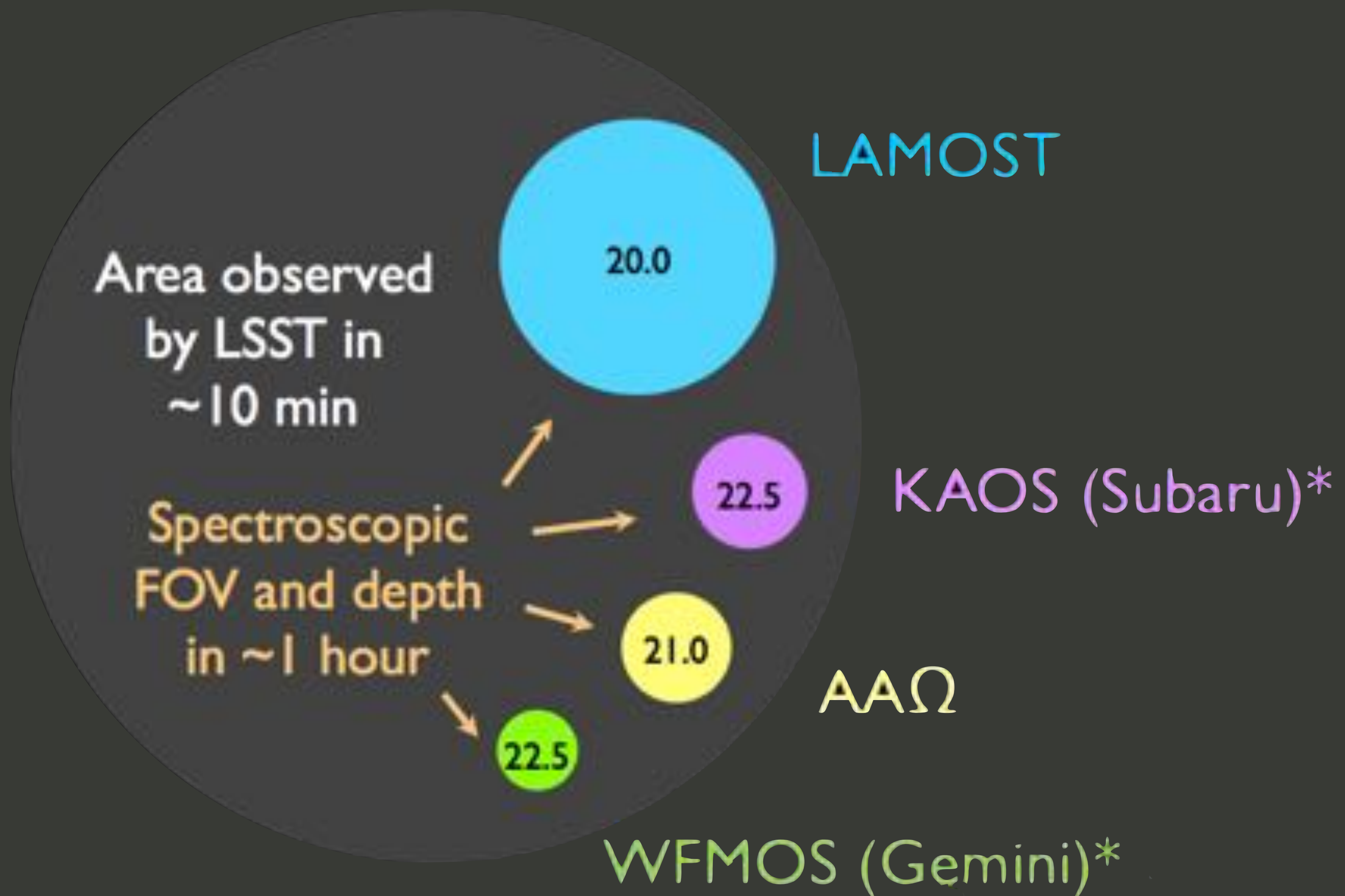
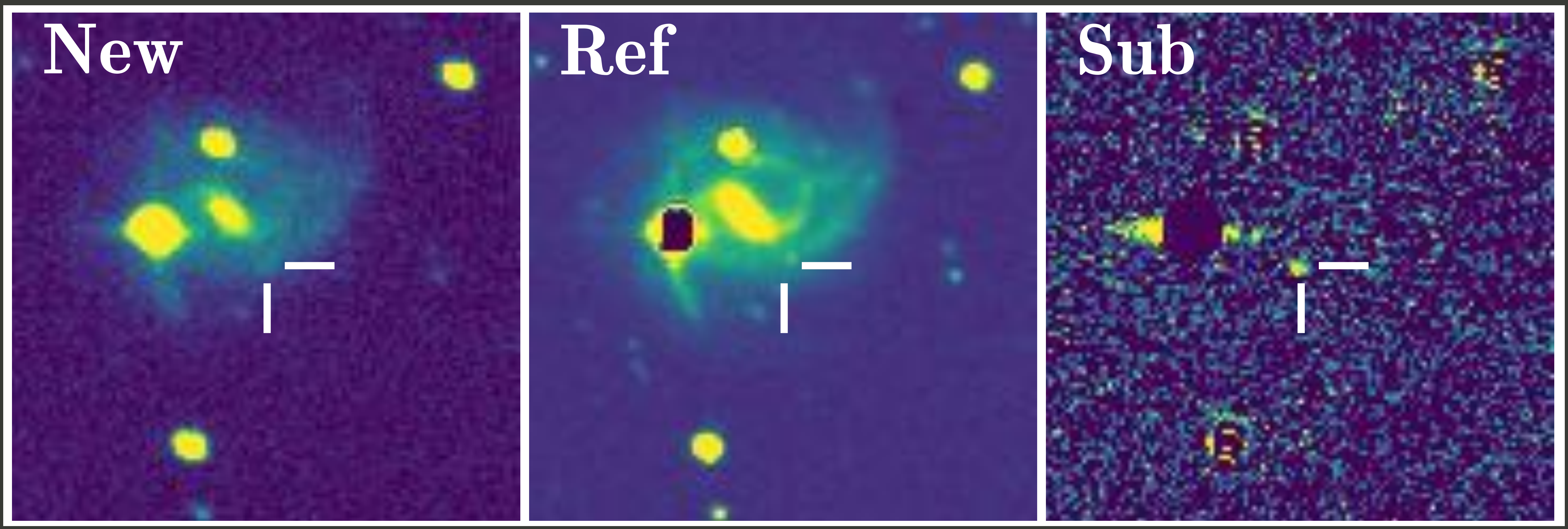


figure credit: Adam Myers

How Can ML Help LSST?

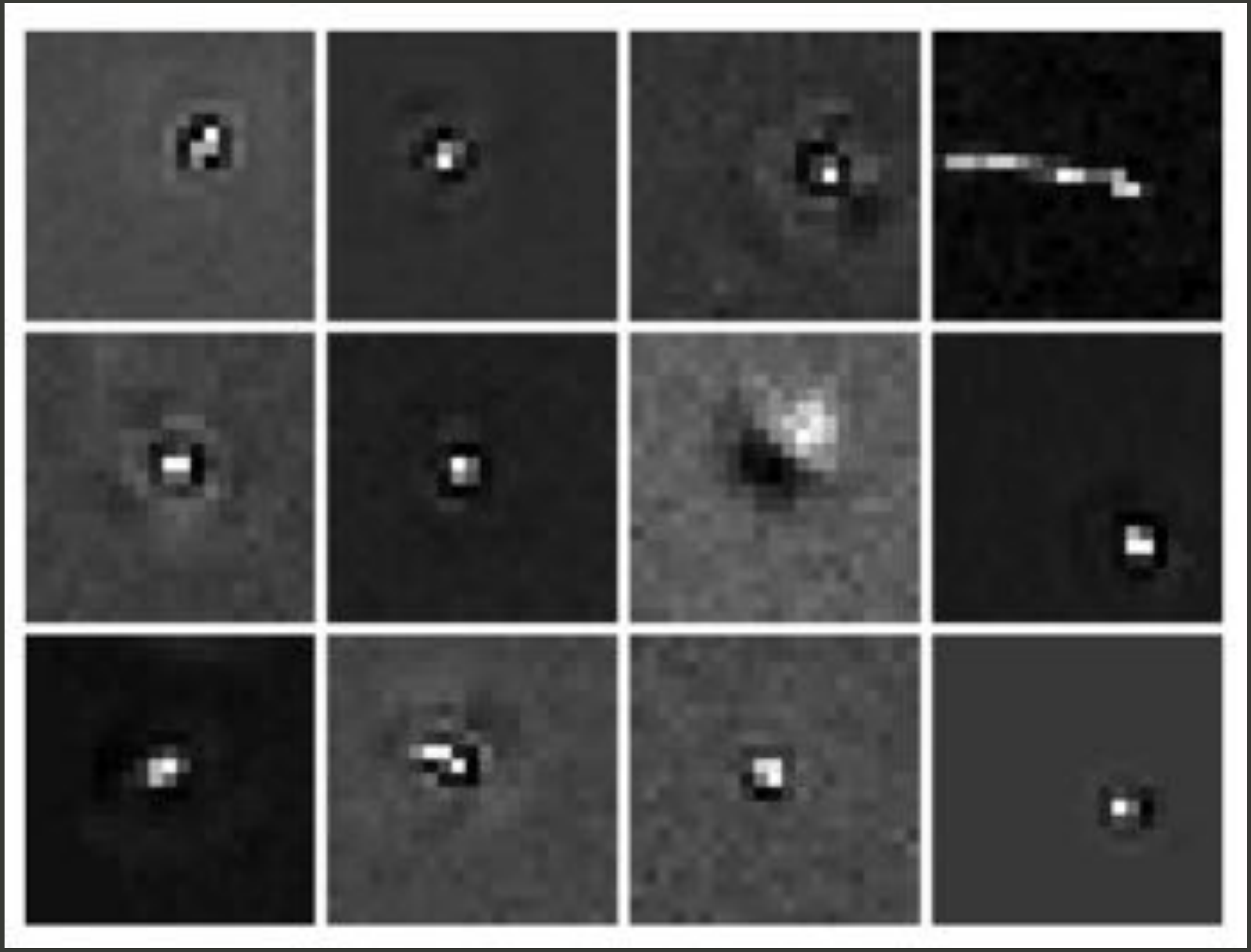
How are alerts generated?



AAM+17b

How Can ML Help LSST?

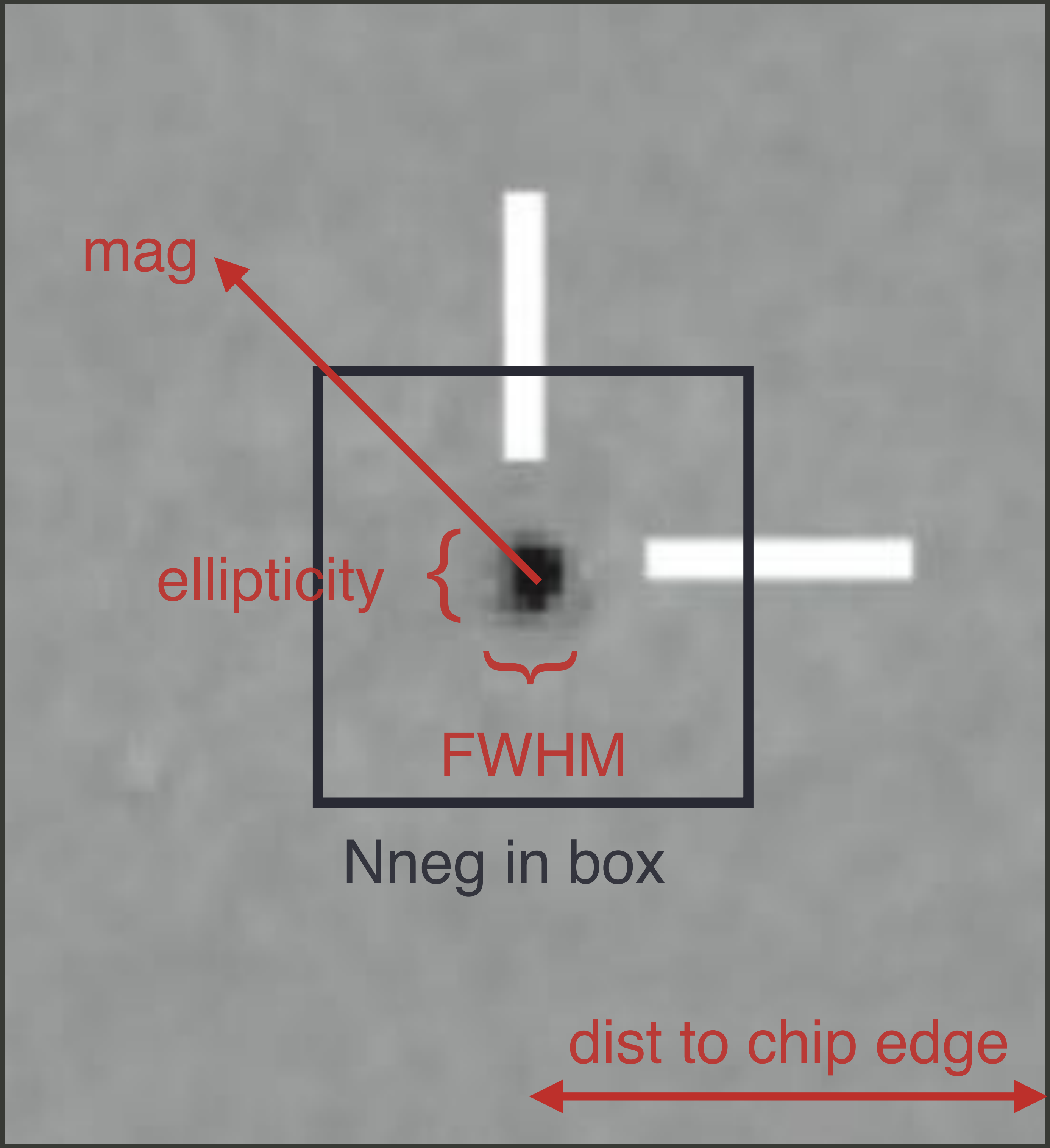
Image subtraction produces lots of artifacts



Brink+13

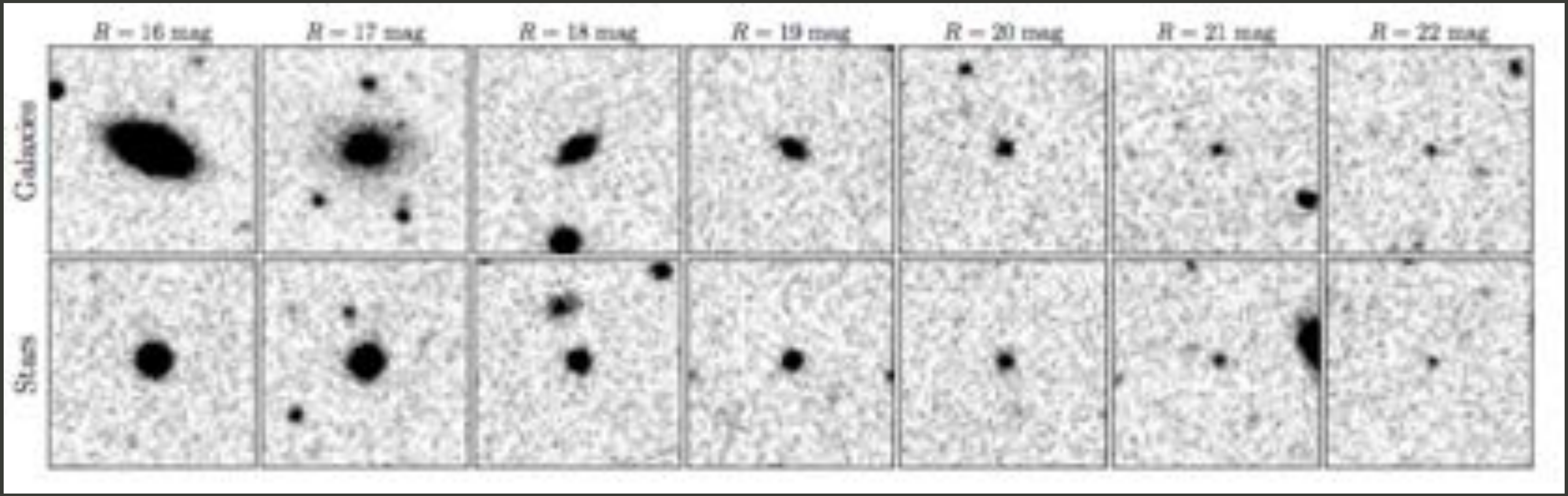
How Can ML Help LSST?

Real/Bogus



How Can ML Help LSST?

Star/Galaxy Separation



AAM+17a

How Can ML Help LSST?

Real/Bogus helps *characterize*

What is the probability this particular alert is real?

Star/Galaxy helps *characterize*

Is this new event in the Milky Way Galaxy or not?

Ultimately want **classification**

This alert is for a dying massive star that has lost all its Hydrogen

How Can ML Help LSST?

Real/Bogus helps *characterize*

What is the probability this particular alert is real?

Star/Galaxy helps *characterize*

Is this new event in the Milky Way Galaxy or not?

Ultimately want **classification**

This alert is for a dying massive star that has lost all its Hydrogen

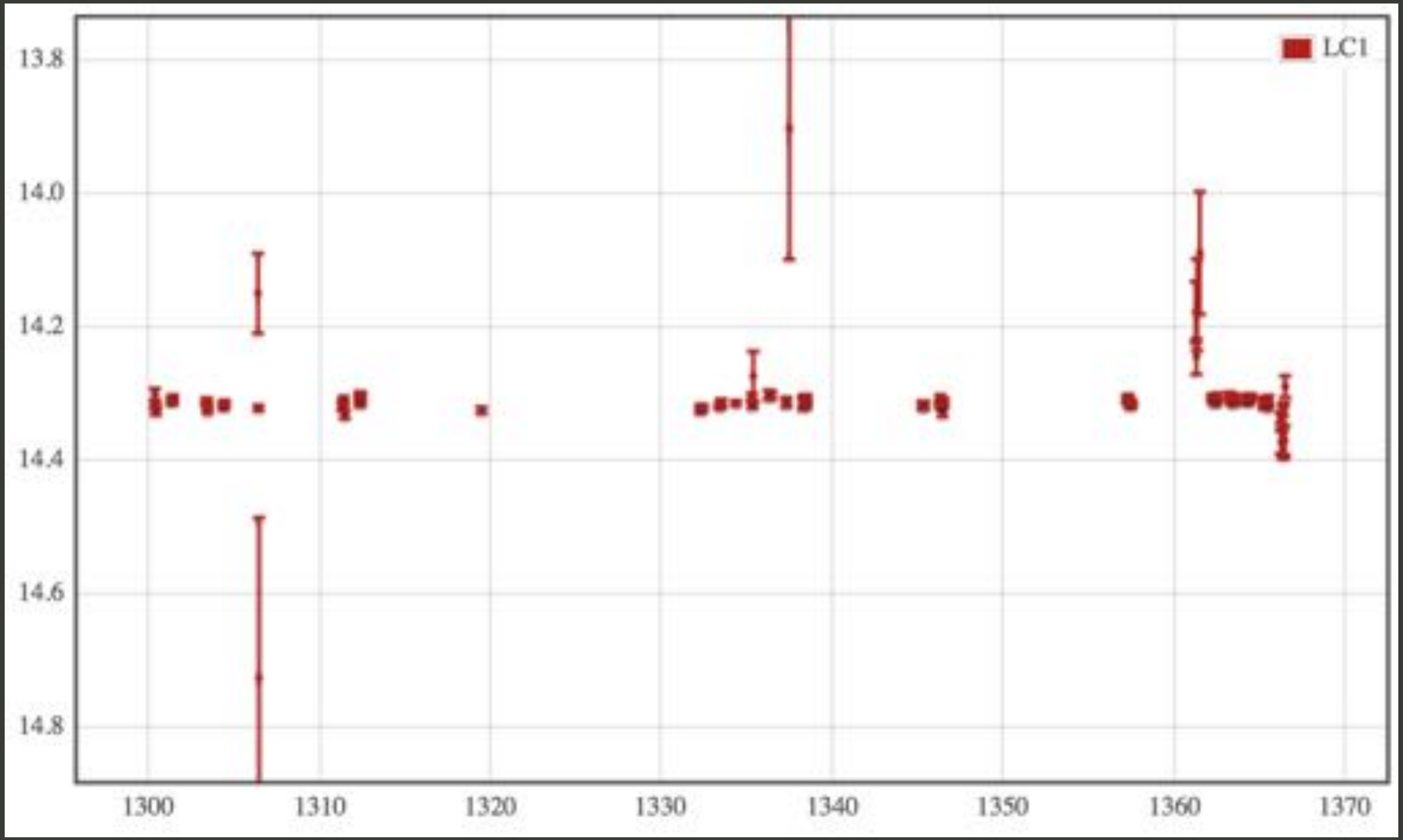
Machine Learning is all about classification

(see previous lecture)

How Can ML Help LSST?

Light curve classification challenges

Photometry provides an incomplete (noisy) picture



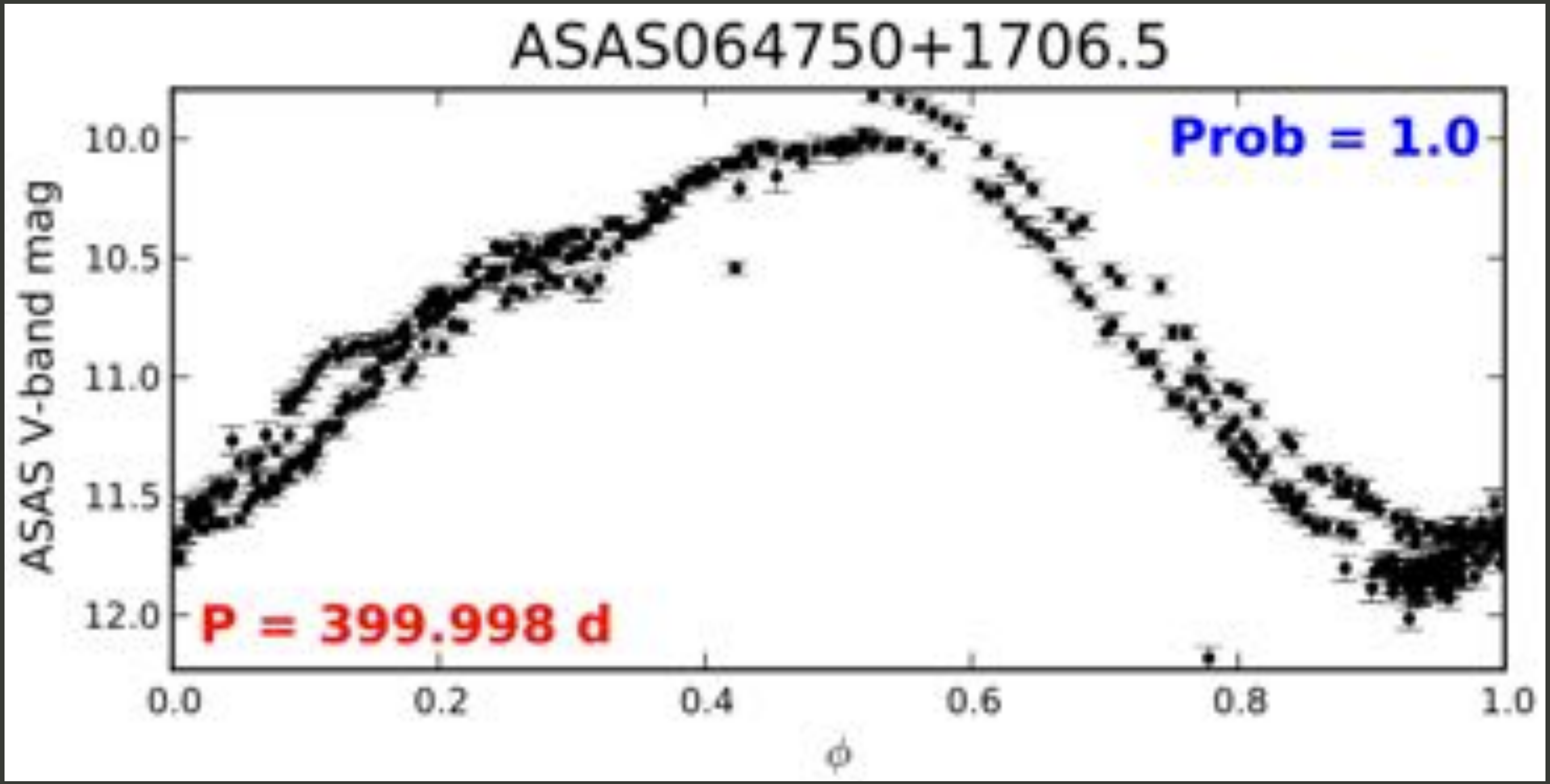
brightness

Time (d)

How Can ML Help LSST?

Light curve classification challenges

Photometry provides an incomplete (noisy) picture

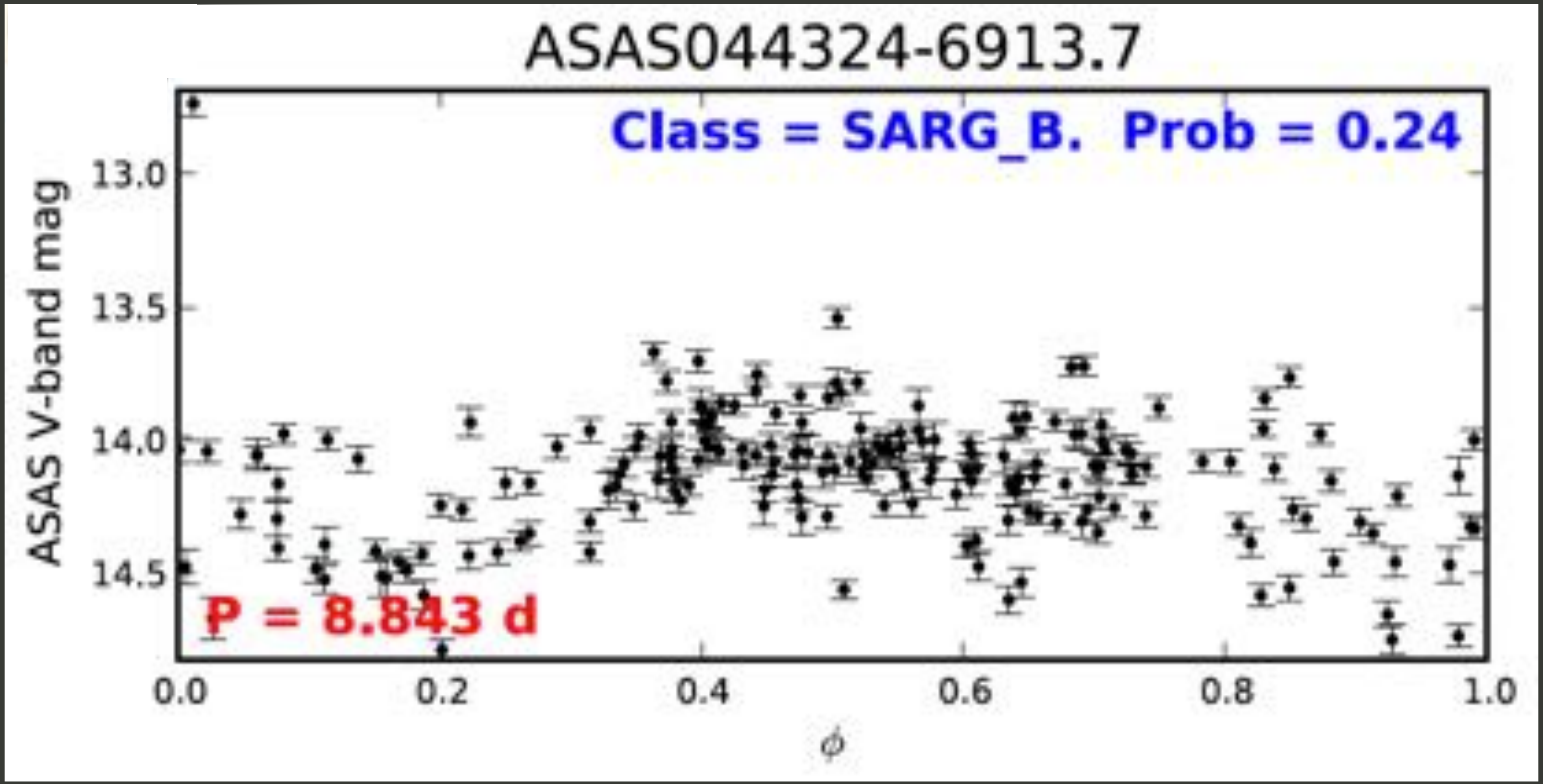


Richards+AAM+12b

How Can ML Help LSST?

Light curve classification challenges

Photometry provides an incomplete (noisy) picture



Richards+AAM+12b

How Can ML Help LSST?

Light curve classification challenges

Photometry provides an incomplete (noisy) picture

Data volumes are becoming enormous

follow up is expensive (faint, N large)

requires good algorithms

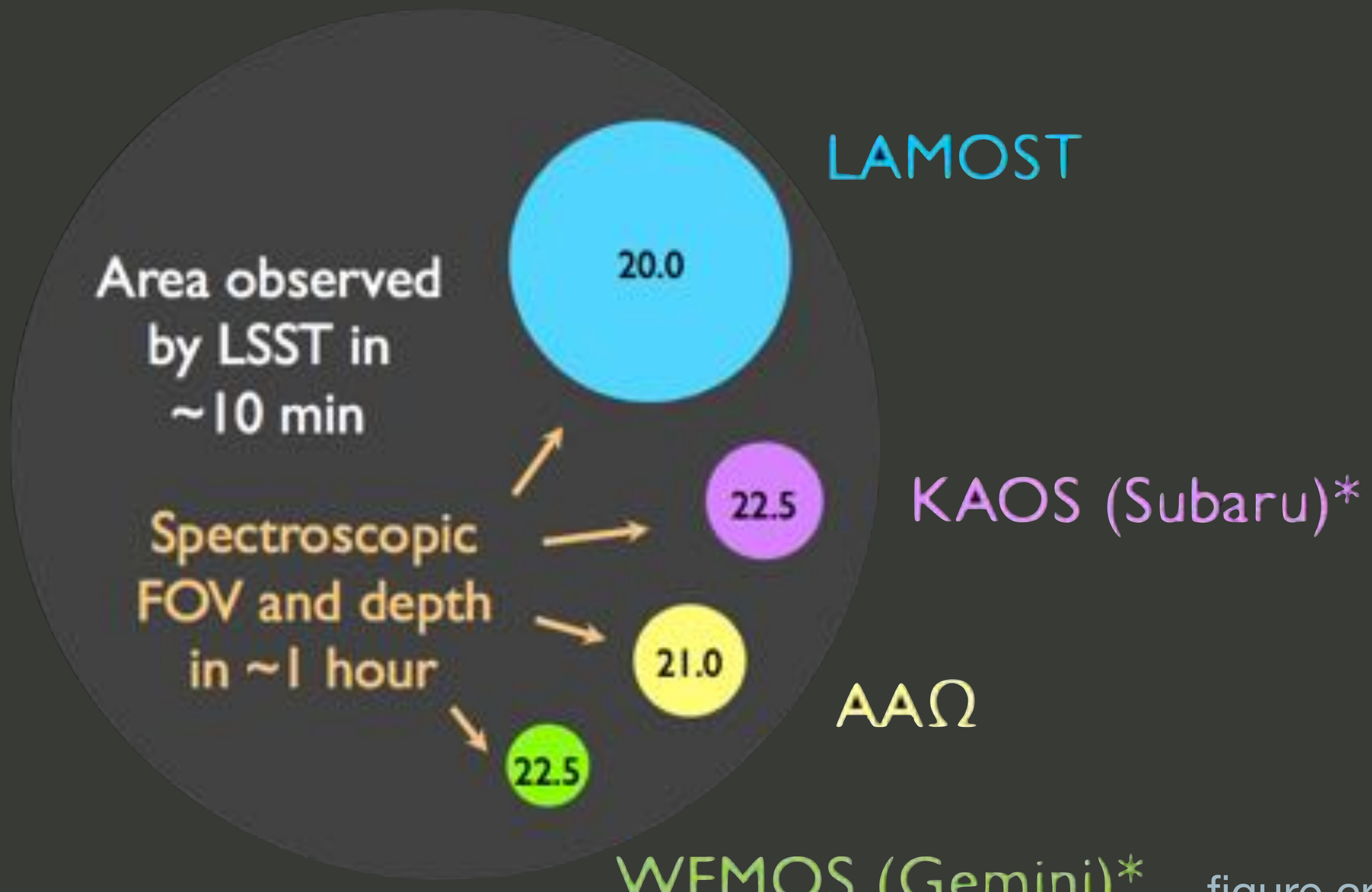


figure credit: Adam Myers

How Can ML Help LSST?

Light curve classification challenges

Photometry provides an incomplete (noisy) picture

Data volumes are becoming enormous

- follow up is expensive (faint, N large)
- requires good algorithms

Different parameter space for different surveys

How Can ML Help LSST?

Light curve classification challenges

Photometry provides an incomplete (noisy) picture

Data volumes are becoming enormous

follow up is expensive (faint, N large)

requires good algorithms

Different parameter space for different surveys

$$\mathrm{LSST}: r_{\mathrm{lim}} \approx 24.5\,\mathrm{mag},\, A \approx 15\mathrm{k}\,\mathrm{deg}^2$$

$$\mathrm{ZTF}: r_{\mathrm{lim}} \approx 20.5\,\mathrm{mag},\, A \approx 30\mathrm{k}\,\mathrm{deg}^2$$

$$\mathrm{ASAS}: r_{\mathrm{lim}} \approx 13.5\,\mathrm{mag},\, A \approx 28\mathrm{k}\,\mathrm{deg}^2$$

How Can ML Help LSST?

Light curve classification challenges

Photometry) picture

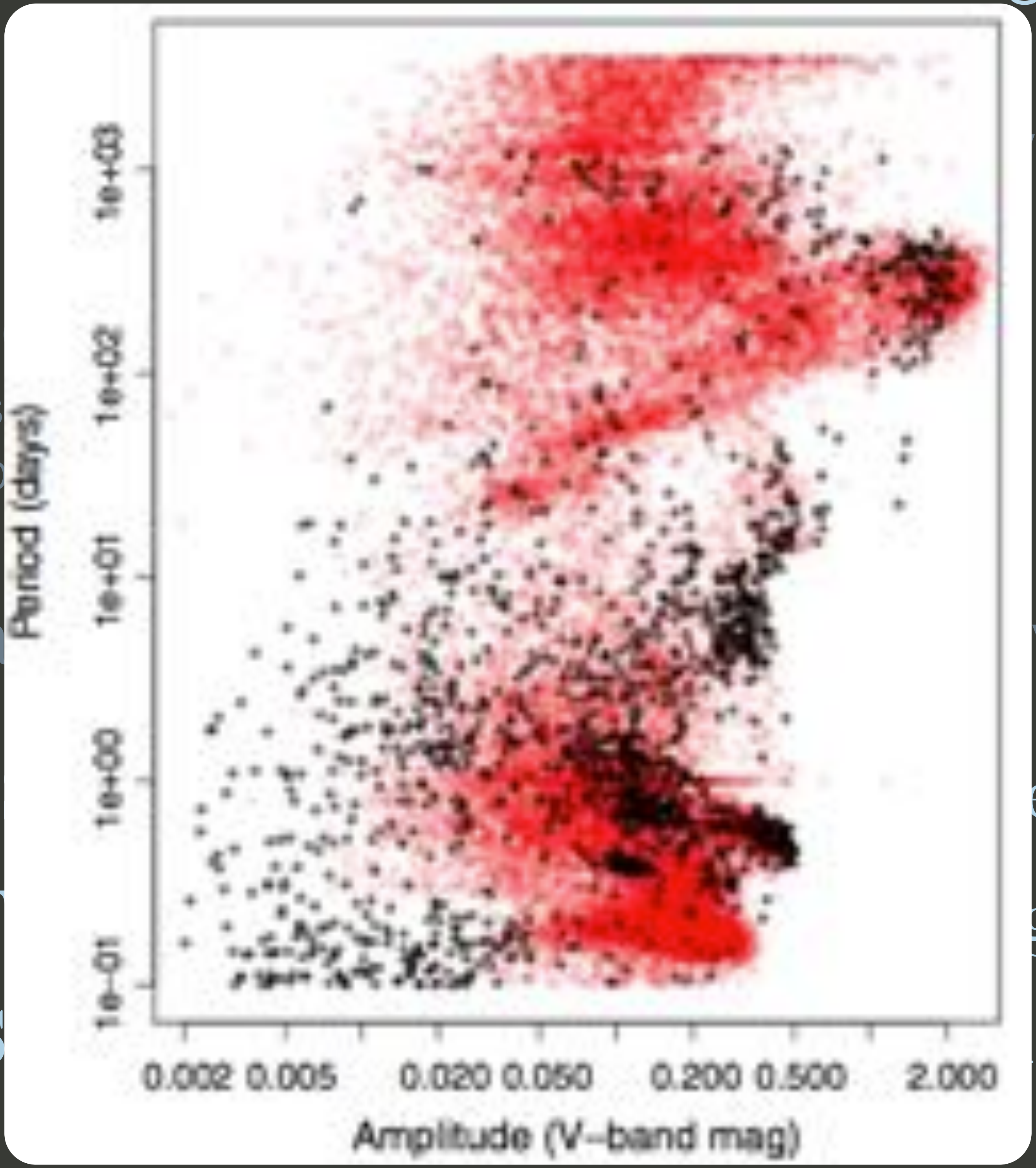
Data volum
follow up is
requires go

Different pa veys

LS eg^2

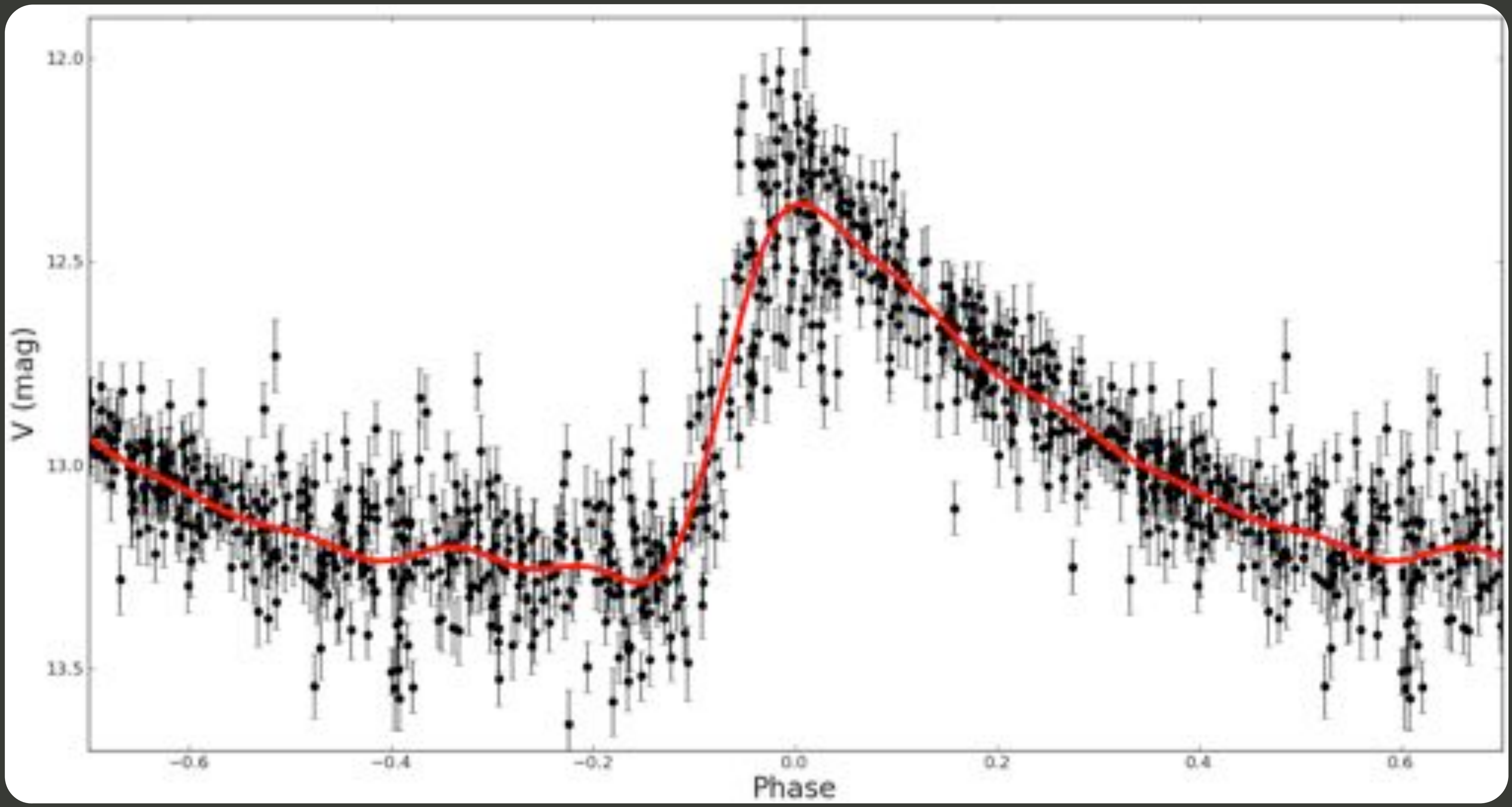
ZT g^2

AS eg^2



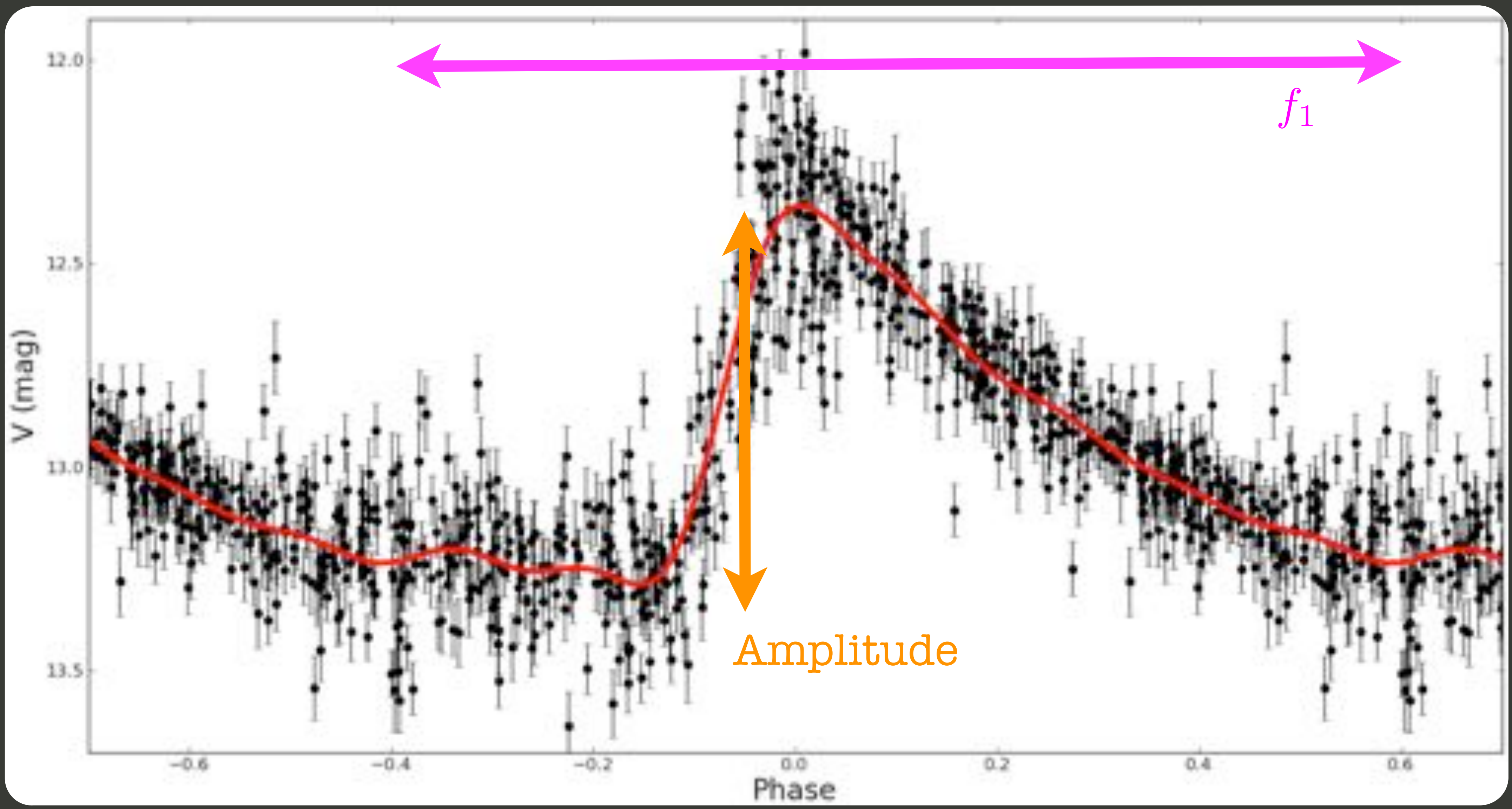
Richards+AAM+12a

How Can ML Help LSST?



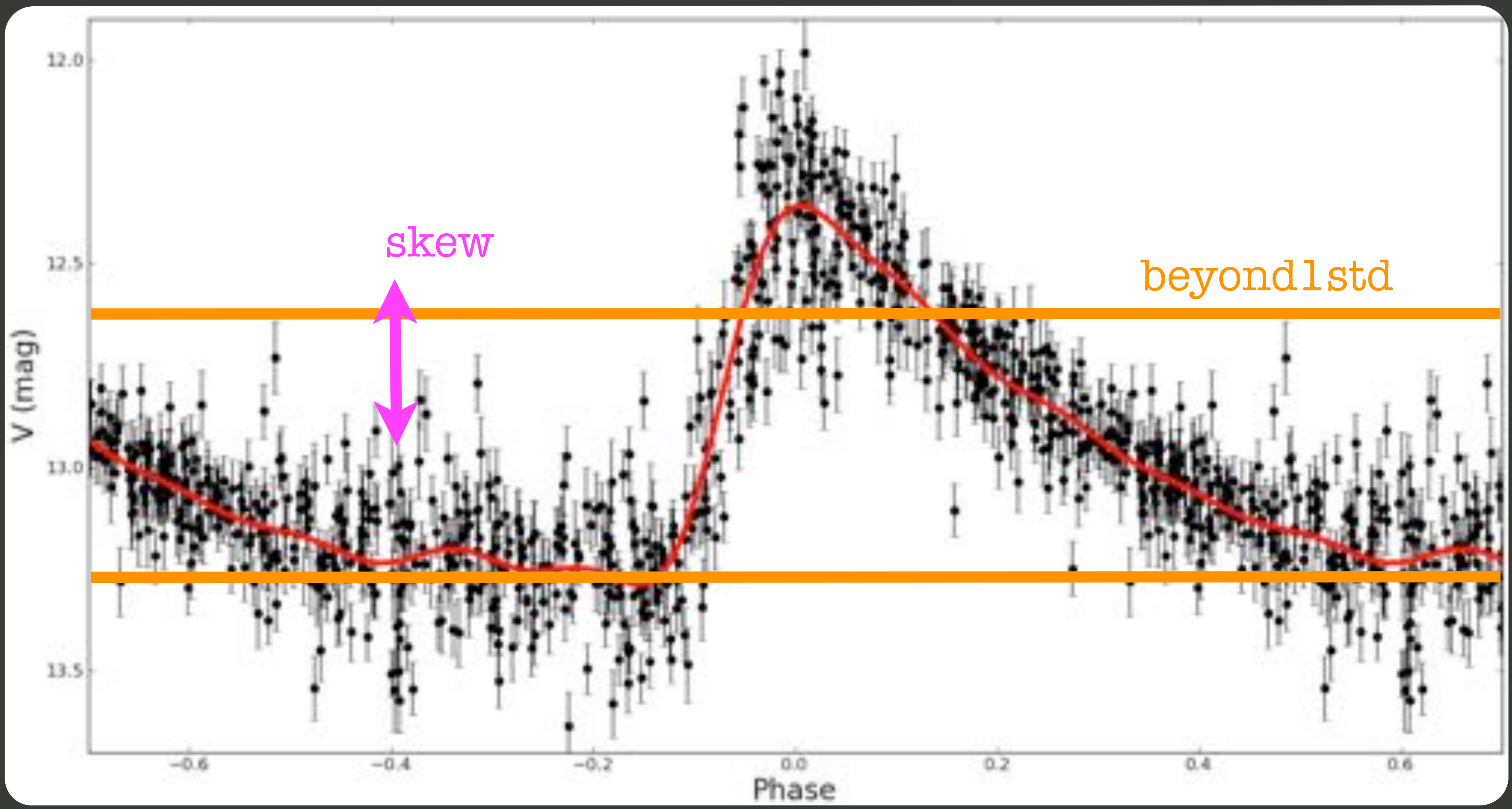
How Can ML Help LSST?

Measure features (metrics) for all light curves

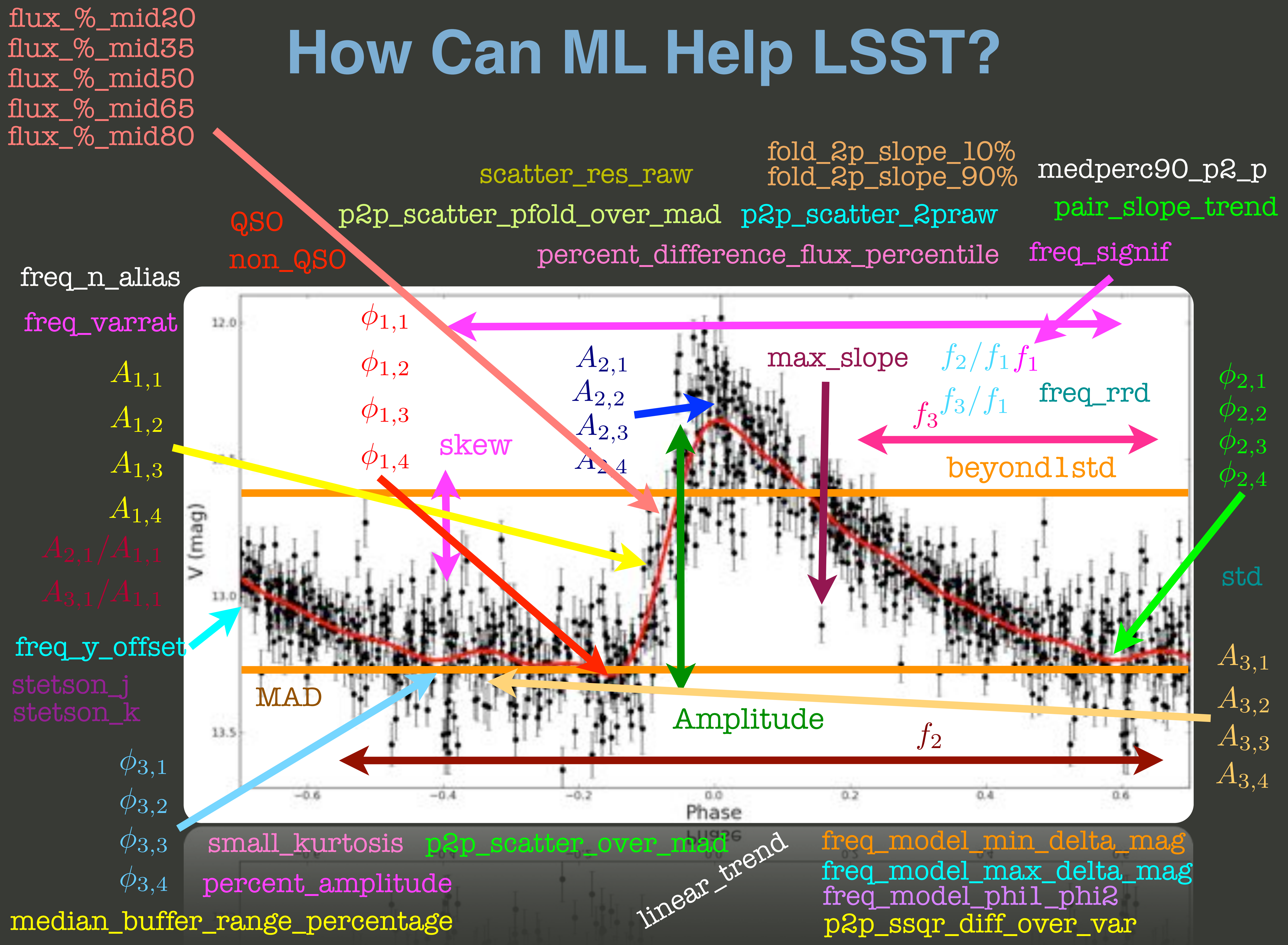


How Can ML Help LSST?

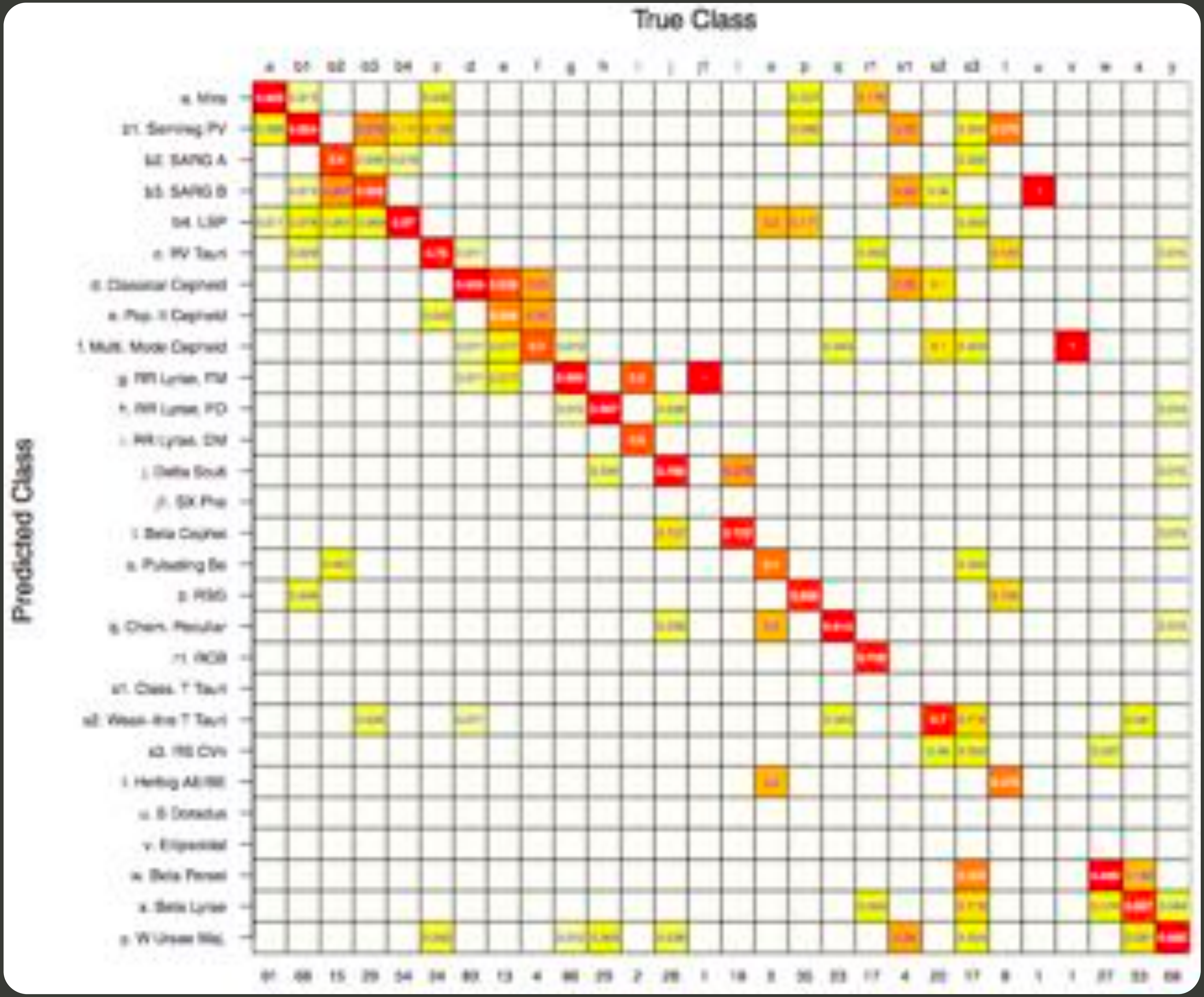
Measure features (metrics) for all light curves



How Can ML Help LSST?



How Can ML Help LSST?



Richards+AAM+12b

How Can ML Help LSST?

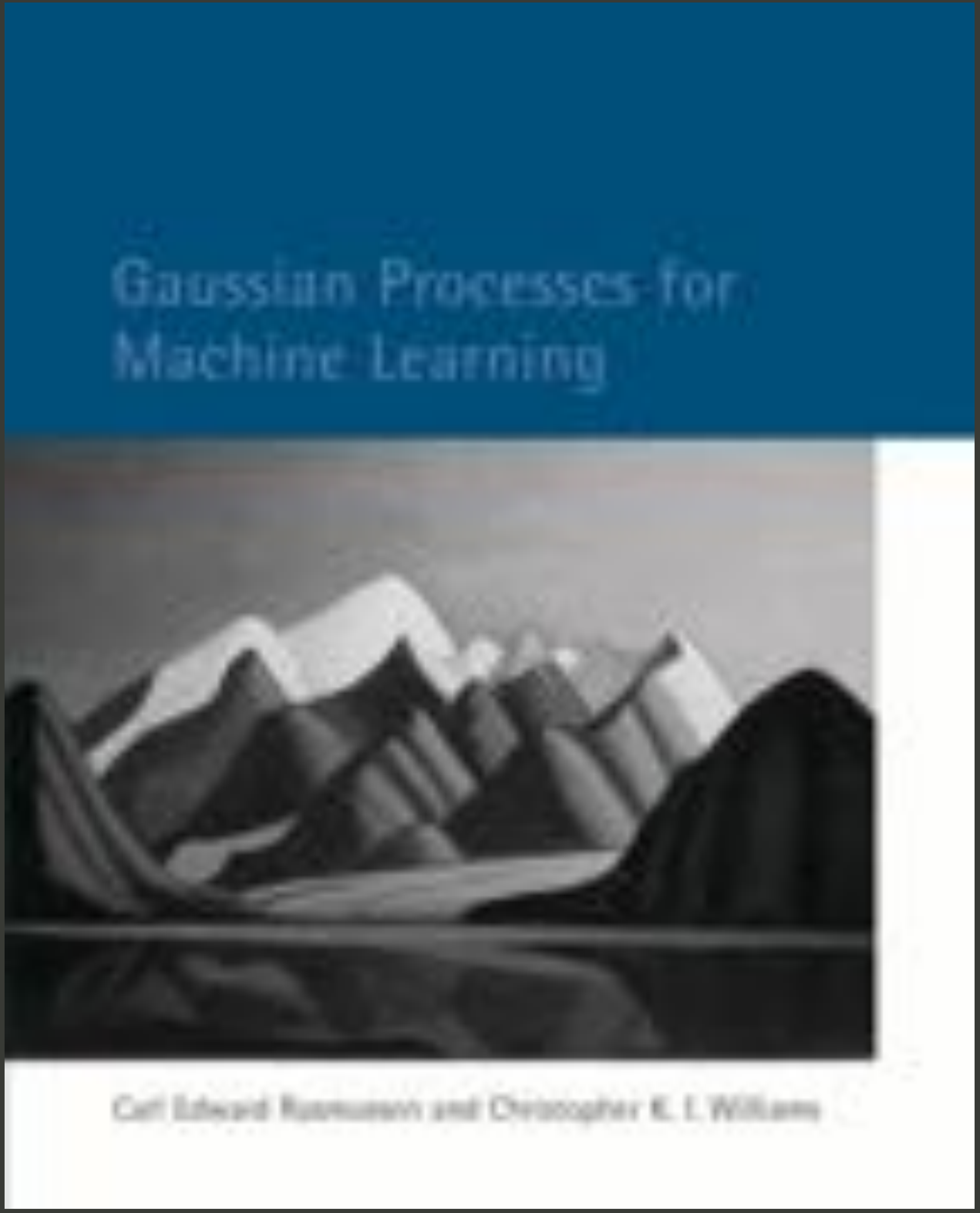
What if obs gaps bias feature measurements?

Hint - they absolutely do

Then ML may not be the best answer…

Gaussian Processes

The ultimate expression of model flexibility…



Gaussian Processes

A quick word of warning…

Gaussian Processes

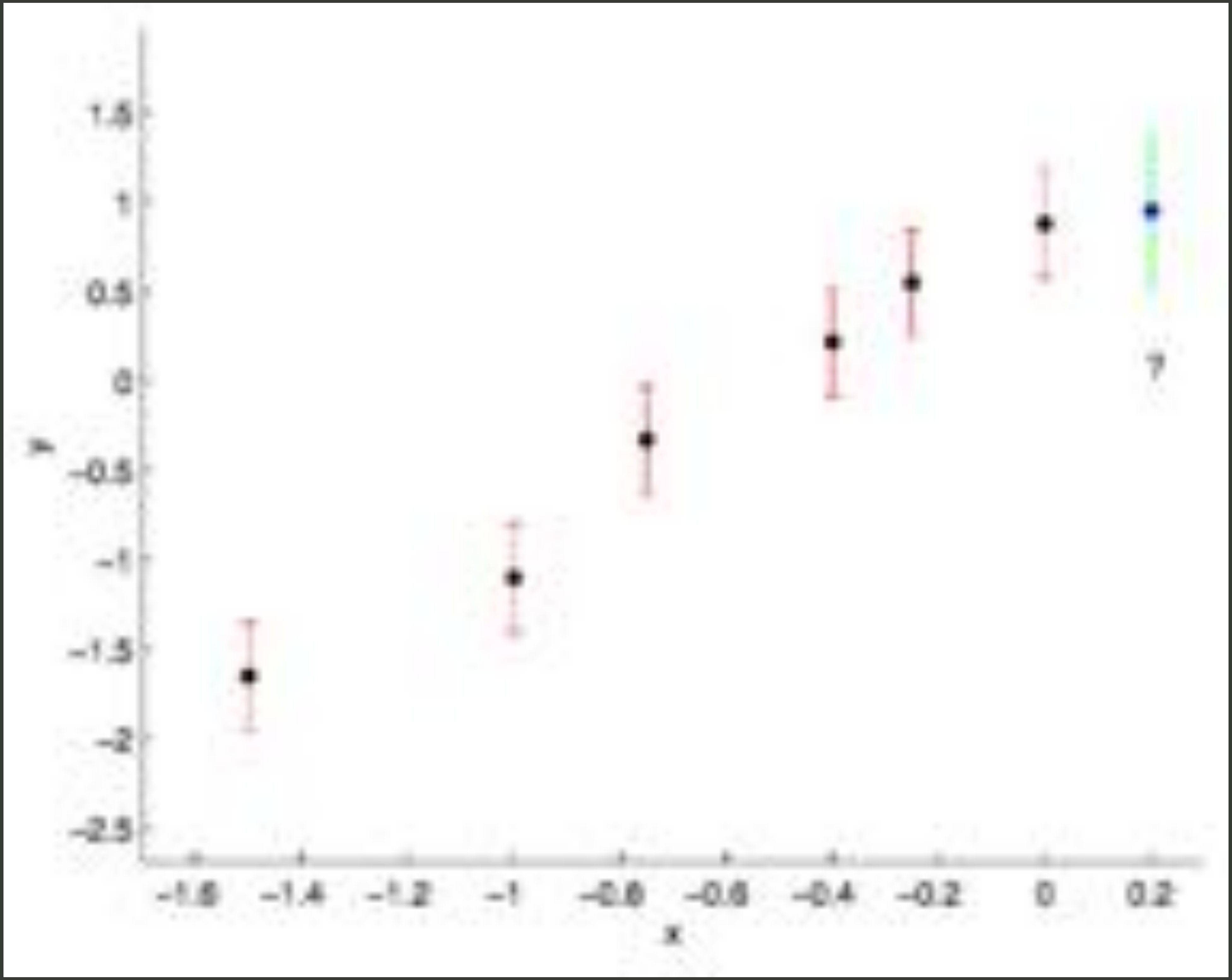


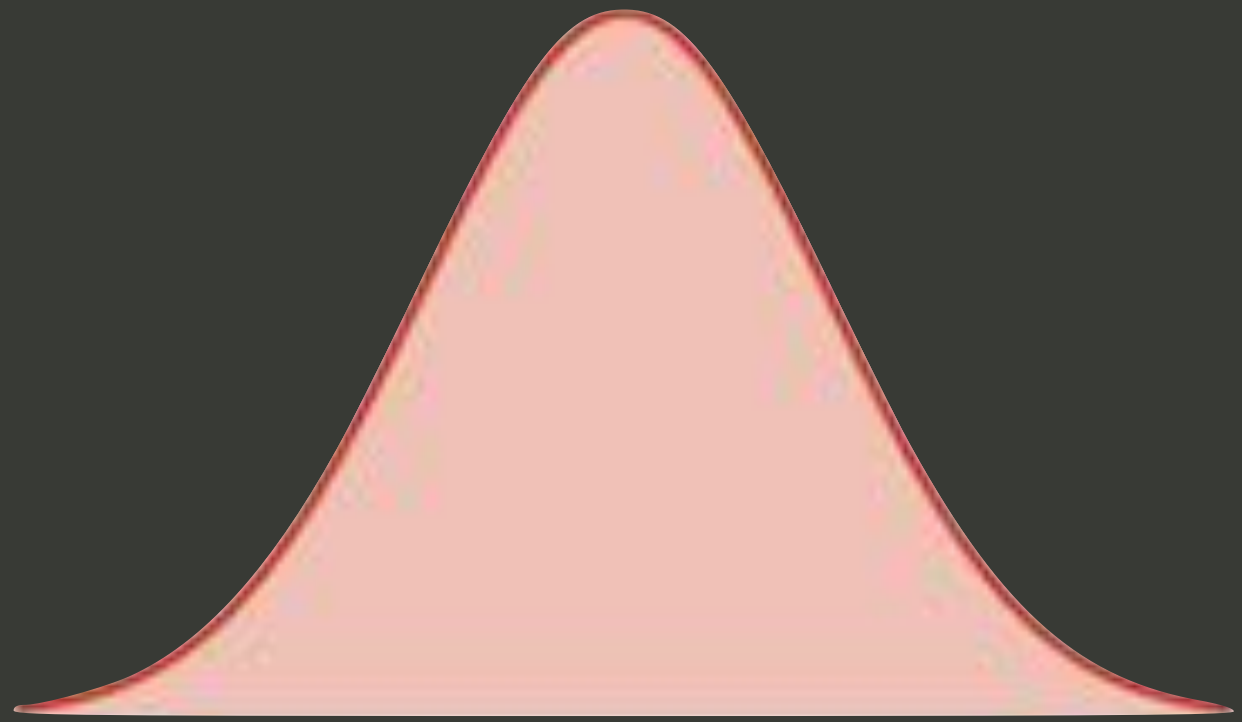
image credit: Mark Ebden

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right]$$

$$y \sim \mathcal{N}(\mu, \sigma^2)$$



Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & C \\ C & \sigma_2^2 \end{bmatrix}\right)$$

$$C = \mathrm{cov}(y_1, y_2)$$

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

marginal distribution of y_1:

$$p(y_1) = \mathcal{N}(\mu_1, \sigma_1^2)$$

if y_2 known, conditional distribution on y_1:

$$p(y_1 \mid y_2) = \mathcal{N}\left(\mu_1 + C(y_2 - \mu_2)/\sigma_2^2, \sigma_1^2 - C^2\sigma_2^2\right)$$

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

consider N variables drawn from multivariate Gaussian:

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}, K)$$

where:

$$y = (y_1, y_2, \ldots, y_N)^T$$

$$\boldsymbol{m} = (m_1, m_2, \ldots, m_N)^T$$

$$K_{ij} = \mathrm{cov}(y_i, y_j)$$

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

what choice for K? squared exponential is common

$$K_{ij} = k(x_i, x_j) = \sigma_f^2 \exp\left[\frac{-(x_i - x_j)^2}{2l^2}\right]$$

a GP is fully specified by mean function and covariance function

Gaussian Processes

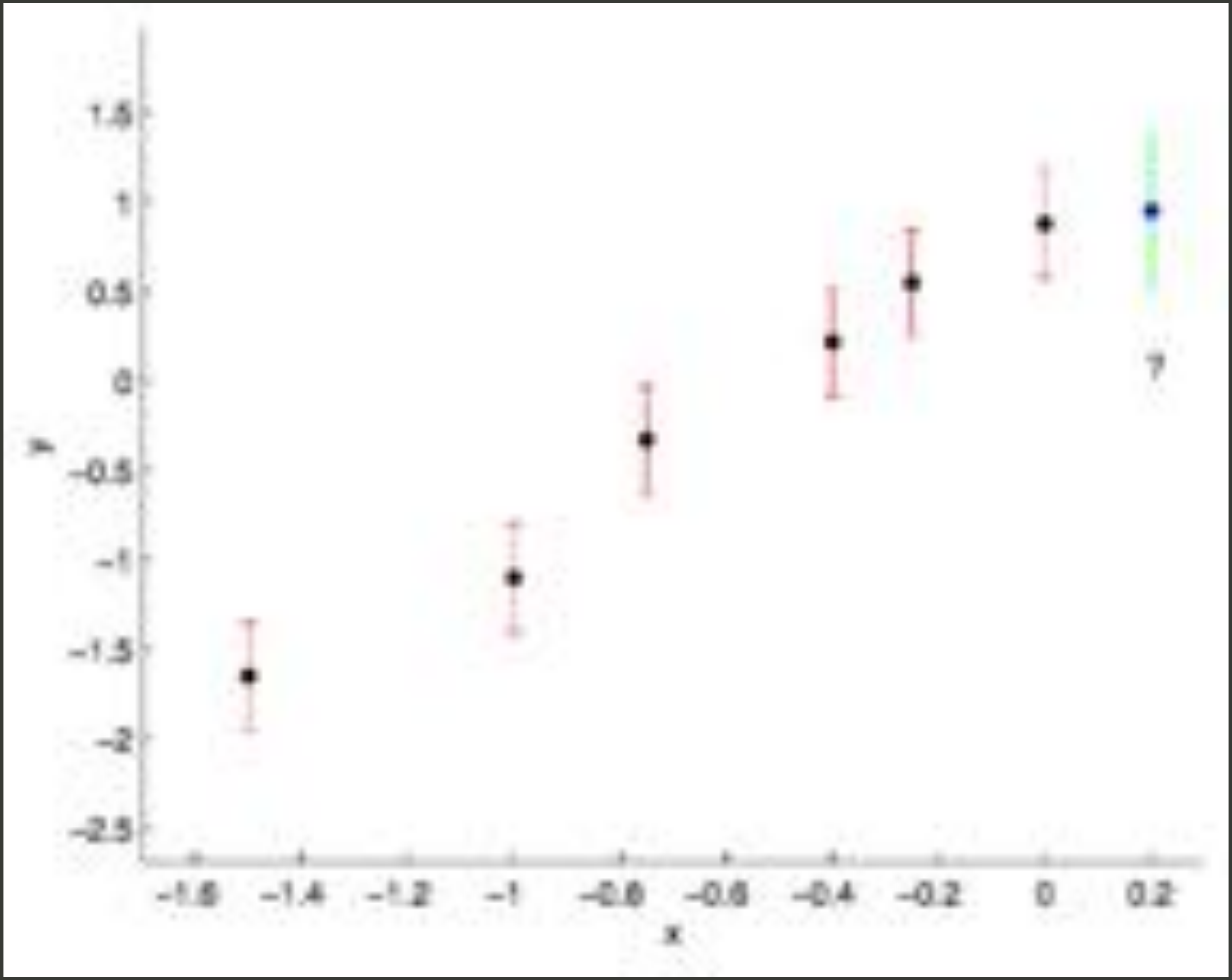


image credit: Mark Ebden

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

we know (x,y) for several points, now determine y_* at x_*

$$\mathrm{p}\left(\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_* \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m} \\ \boldsymbol{m}_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}\right)$$

which yields conditional probability

$$\mathrm{p}(\boldsymbol{y}_* \mid \boldsymbol{y}, k) = \mathcal{N}(K_*^T K^{-1} \boldsymbol{y}, K_{**} - K_*^T K^{-1} K_*)$$

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

in real life there is also noise:

$$\mathrm{cov}(y_i, y_j) = k(x_i, x_j) + \delta_{ij}\sigma^2$$

which yields conditional probability

$$\mathrm{p}(\boldsymbol{y}_* \mid \boldsymbol{y}, k) = \mathcal{N}(K_*^T(K + \sigma^2\mathbb{I})^{-1}\boldsymbol{y}, K_{**} - K_*^T(K + \sigma^2\mathbb{I})^{-1}K_*)$$

Gaussian Processes

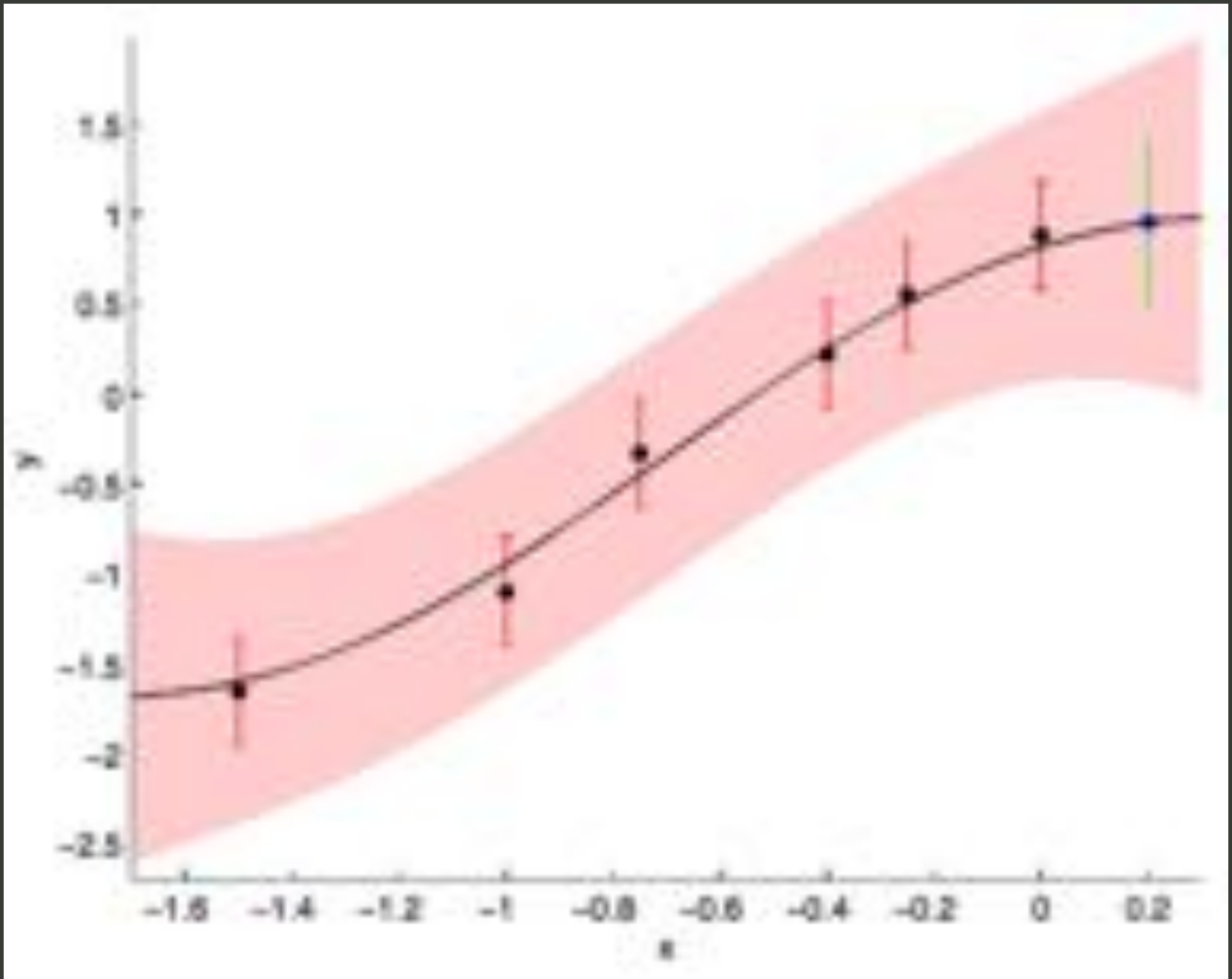


image credit: Mark Ebden

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

but what about σ_f, l, σ

$$\mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}, \theta) = \mathcal{N}(\boldsymbol{y} \,|\, \mathbf{0}, K + \sigma^2 \mathbb{I})$$

which is just a likelihood, which can be maximized

$$\log p(\boldsymbol{y}|\boldsymbol{x}, \theta) = -\frac{1}{2}\boldsymbol{y}^T K^{-1} \boldsymbol{y} - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi$$

Gaussian Processes

A GP is a collection of random variables,
in which any finite subset has a multivariate gaussian distribution

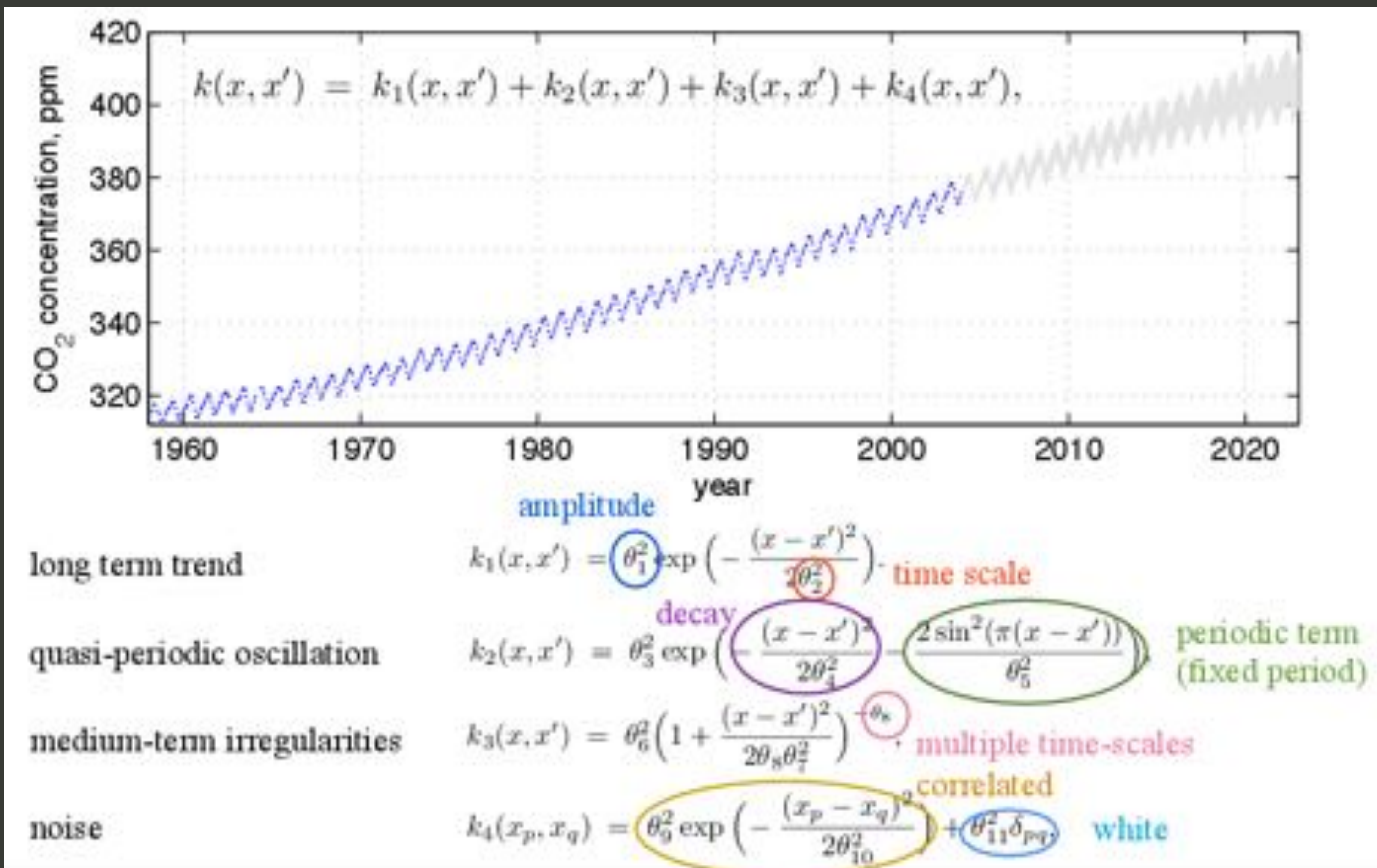
but what about Bayes

$$\mathrm{p}(\theta \,|\, \boldsymbol{x}) = \frac{\mathrm{p}(\boldsymbol{x} \,|\, \theta)\, \mathrm{p}(\theta)}{\mathrm{p}(\boldsymbol{x})}$$

marginalize over hyper-parameters using favorite MCMC

Gaussian Processes

what if the kernel is more complicated?



Rasmussen & Williams