Generalized Multiple Importance Sampling for Source Characterization in Multi-epoch Imaging

CASIS Workshop

Michael D. Schneider
Physics Division

May 15, 2016

In collaboration with:
William Dawson (LLNL)
Joshua Meyers (Stanford)

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

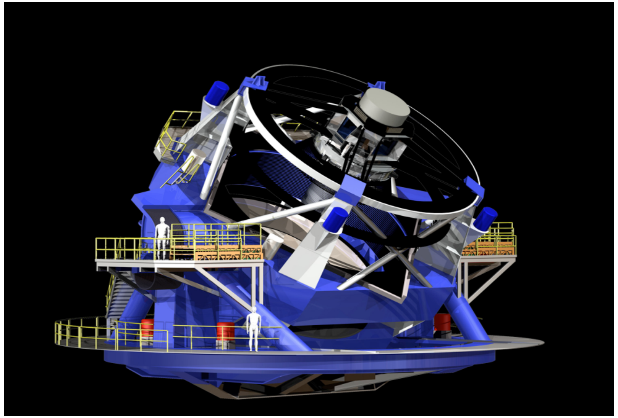
Lawrence Livermore National Laboratory

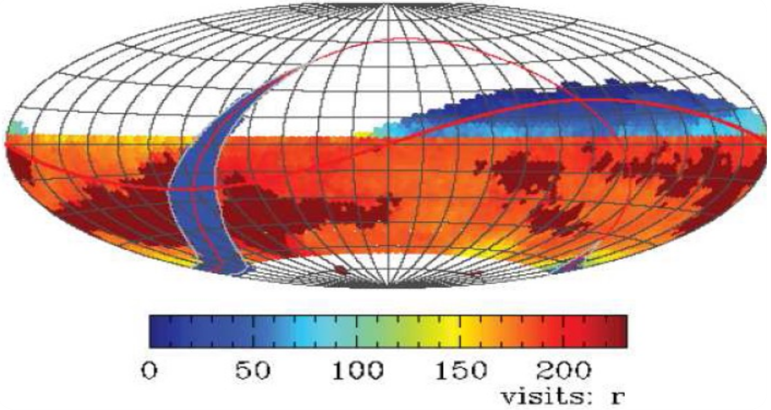
The Large Synoptic Survey Telescope (LSST)

Construction start: 2014 First light: 2020 Survey end: 2030

8.4m telescope 18,000+ deg^2 10 maarcsec astrometry r<24.5 (<27.5@10yr)

6 broad optical bands 0.5-1% photometry

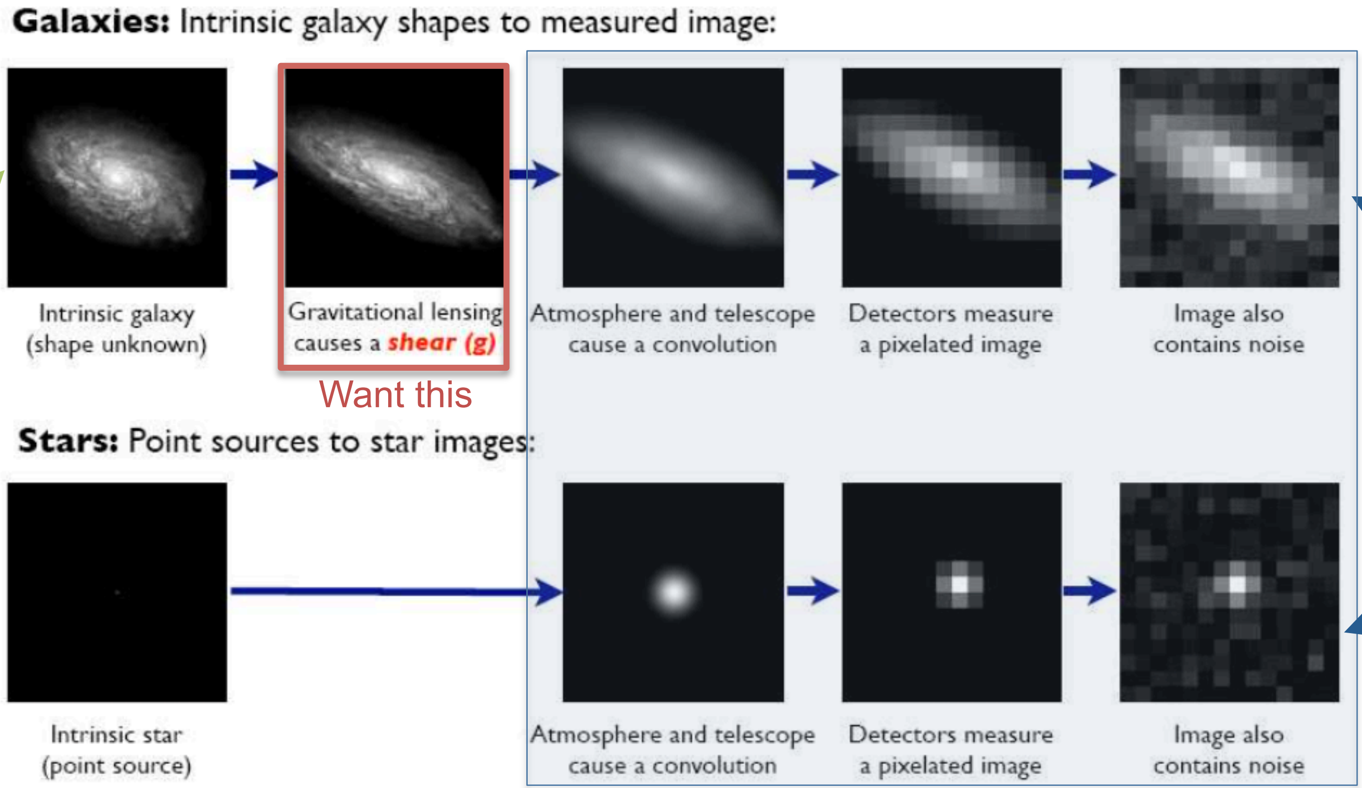




3.2Gpix camera 2x15sec exp/2sec read 15TB/night 20 B objects

Imaging the visible sky, once every 3 days, for 10 years (**825 revisits**)

Science use case: Weak gravitational lensing of galaxies



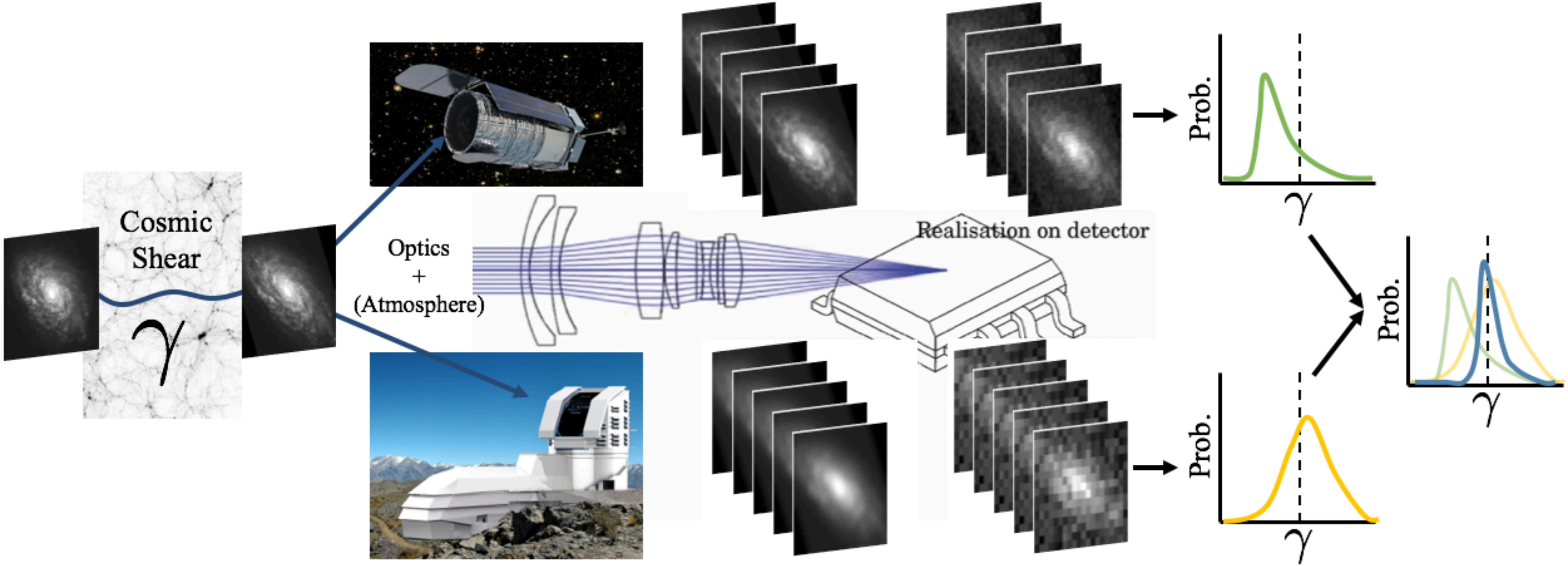
Unknown & dominates signal

Want this

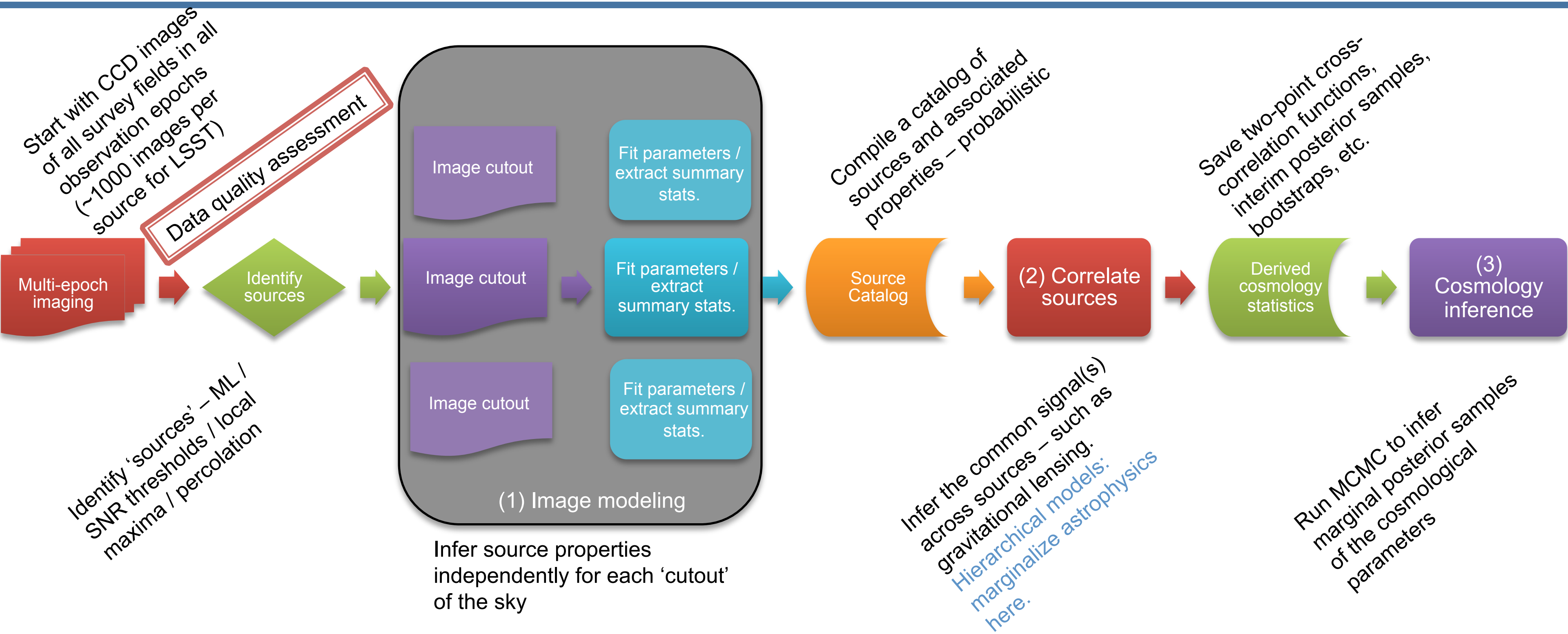
Marginalize

Constrained by

Unlike in the past, we will have many observations of the same sources that must be combined, while marginalizing distinct systematic errors
– A new processing paradigm

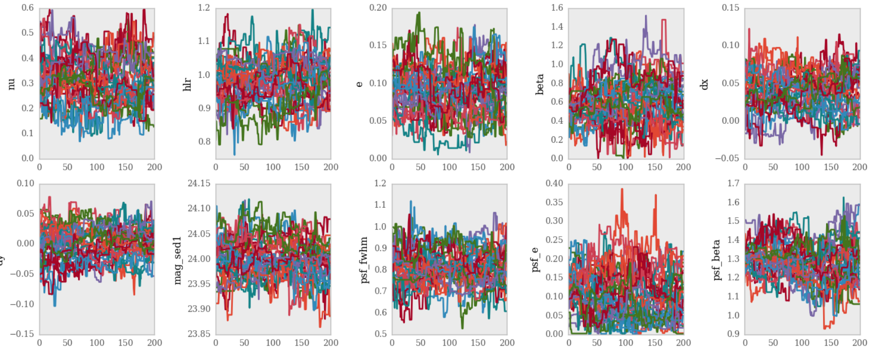


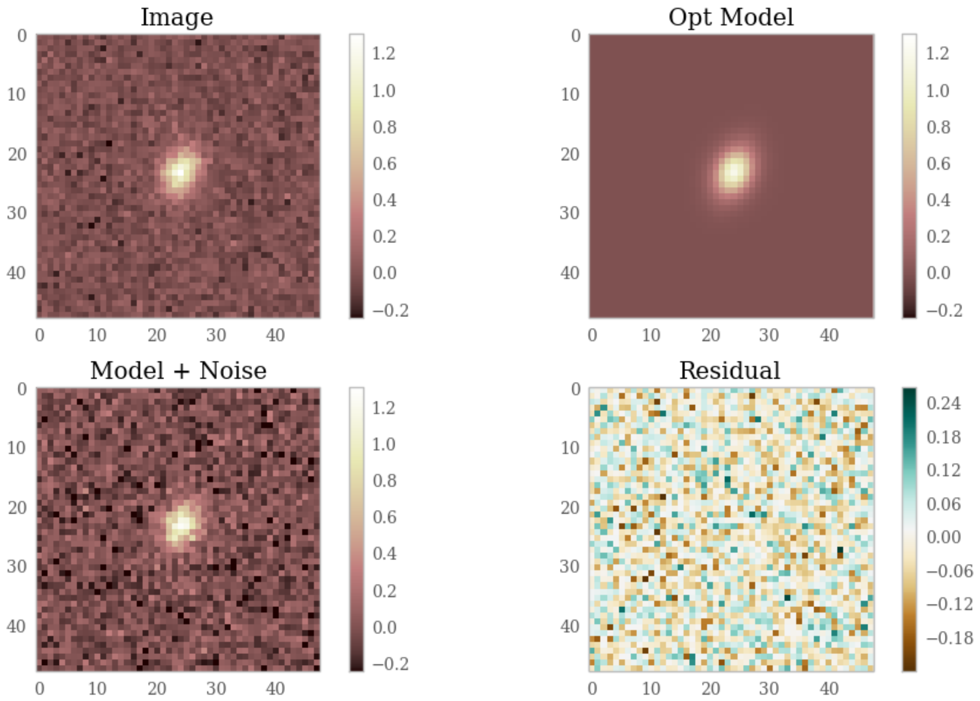
The data management & analysis pipeline components for gravitational weak lensing and galaxy clustering with LSST



Source characterization via probabilistic image modeling

Infer image model parameters via MCMC, then marginalize as part of a hierarchical model for the distributions of galaxy and systematics parameters

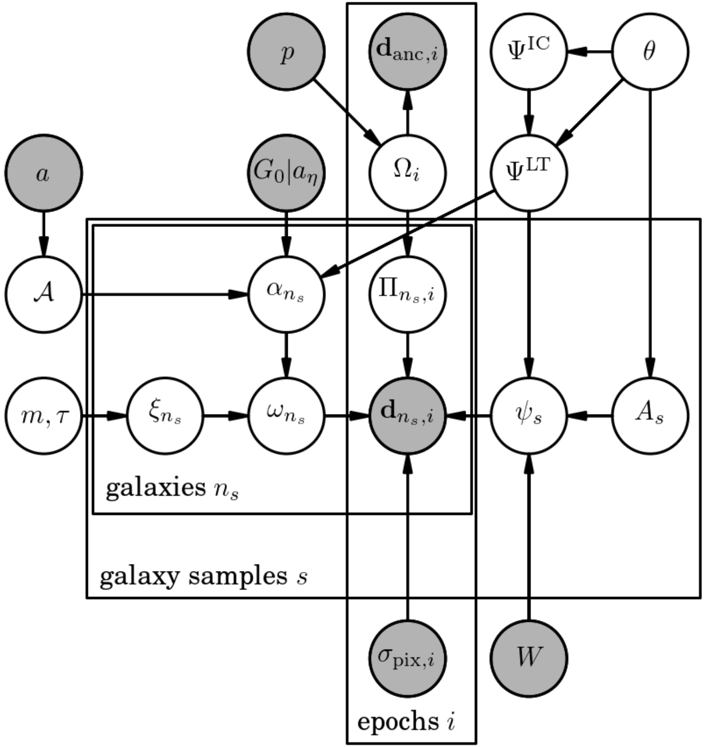




The complete statistical model for cosmic shear

arXiv:1411.2608

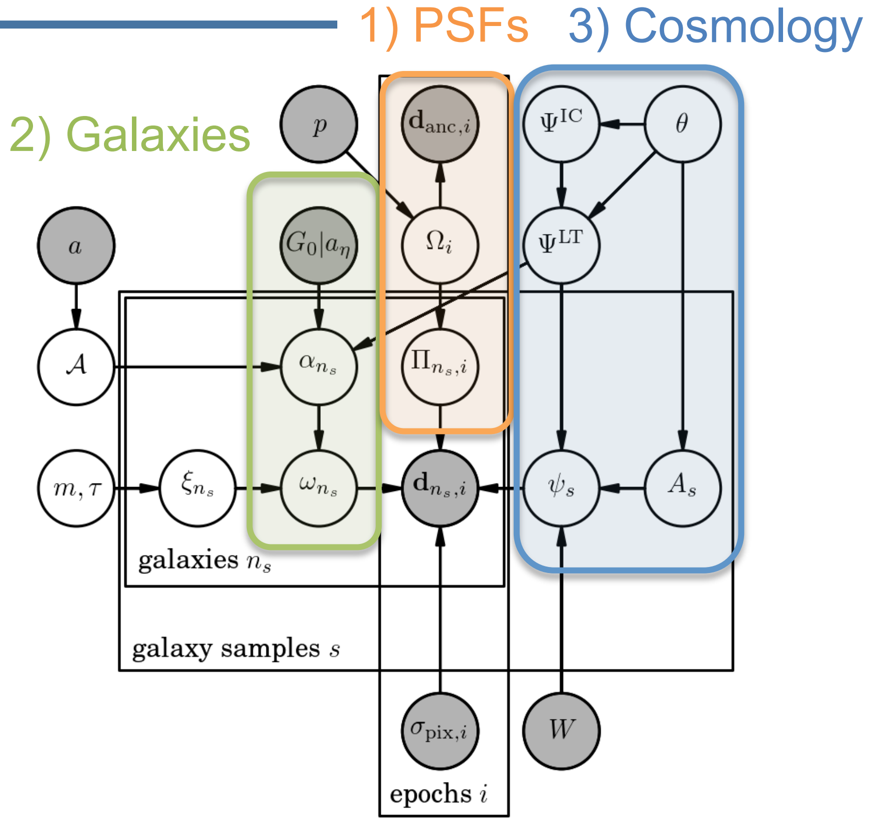
Parameter	Description
$\boldsymbol{\theta}$	Cosmological parameters
$\Psi^{\rm IC}$	Initial conditons for the 3D gravitational potential
$\Psi^{\rm LT}$	Late-time 3D gravitational potential
ψ_s	2D lens potential (given source photo-z bin s)
A_s	Parameters for the line-of-sight source distribution
$\Pi_{n_s,i}$	PSF for galaxy n_s observed in epoch i
Ω_i	Observing conditions in epoch i
$\{\omega_{n_s}\}$	Galaxy model parameters; $n_s = 1, \ldots, n_{{\rm gal},s}$
$\{\alpha_{n_s}\}$	Parameters for the distribution of $\{\omega_{n_s}\}$
$\{\xi_{n_s}\}$	Scaling parameters for $\{\omega_{n_s}\}$
m, τ	Hyperprior parameters for $\{\xi_{n_s}\}$
$\mathcal{A}$	Hyperparameter for $\{\alpha_{n_s}\}$ classifications
$\{\mathbf{d}_{n_s,i}\}$	Pixel data for galaxies $n_s = 1, \ldots, n_{{\rm gal},s}$ in epoch i
$G_0 \vert a_\eta$	Prior specification for $\{\alpha_{n_s}\}$
s	Source sample (e.g., photo-z bin)
W	Survey window function
$\mathbf{d}_{{\rm anc},i}$	Ancillary data for PSF in epoch i
p	Prior params. for observing conditions
a	Prior params. for $\mathcal{A}$
$\sigma_{{\rm pix},i}$	Pixel noise r.m.s. in epoch i
I	Model selection assumptions



The complete statistical model for cosmic shear

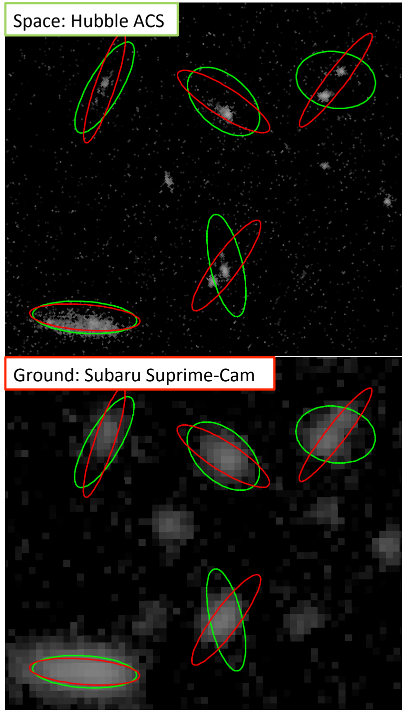
arXiv:1411.2608

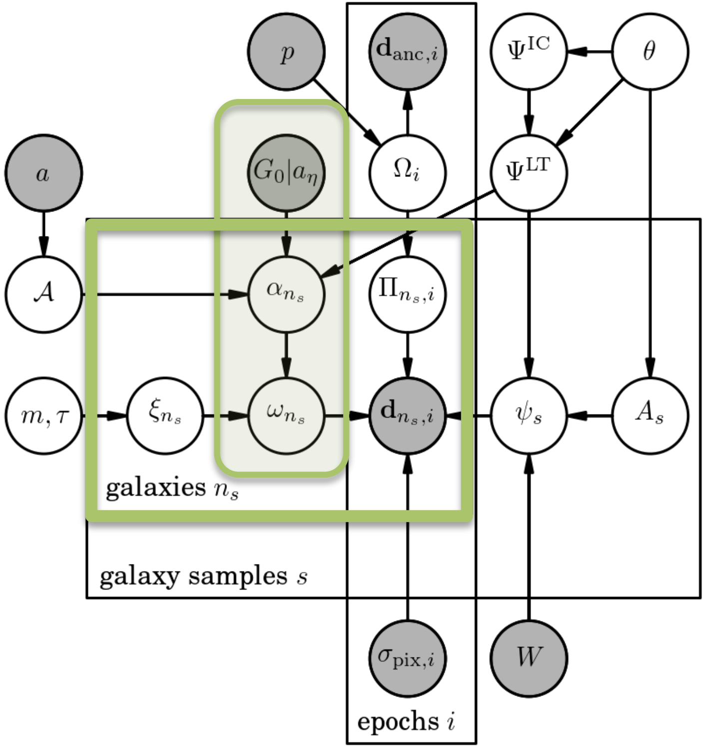
Parameter	Description
$\boldsymbol{\theta}$	Cosmological parameters
$\Psi^{\rm IC}$	Initial conditons for the 3D gravitational potential
$\Psi^{\rm LT}$	Late-time 3D gravitational potential
ψ_s	2D lens potential (given source photo-z bin s)
A_s	Parameters for the line-of-sight source distribution
$\Pi_{n_s,i}$	PSF for galaxy n_s observed in epoch i
Ω_i	Observing conditions in epoch i
$\{\omega_{n_s}\}$	Galaxy model parameters; $n_s = 1, \ldots, n_{\mathrm{gal},s}$
$\{\alpha_{n_s}\}$	Parameters for the distribution of $\{\omega_{n_s}\}$
$\{\xi_{n_s}\}$	Scaling parameters for $\{\omega_{n_s}\}$
m, τ	Hyperprior parameters for $\{\xi_{n_s}\}$
$\mathcal{A}$	Hyperparameter for $\{\alpha_{n_s}\}$ classifications
$\{\mathbf{d}_{n_s,i}\}$	Pixel data for galaxies $n_s = 1, \ldots, n_{\mathrm{gal},s}$ in epoch i
$G_0 \vert a_\eta$	Prior specification for $\{\alpha_{n_s}\}$
s	Source sample (e.g., photo-z bin)
W	Survey window function
$\mathbf{d}_{\mathrm{anc},i}$	Ancillary data for PSF in epoch i
p	Prior params. for observing conditions
a	Prior params. for $\mathcal{A}$
$\sigma_{\mathrm{pix},i}$	Pixel noise r.m.s. in epoch i
I	Model selection assumptions



Naïve approach is intractable

1. Galaxy models be joint fitted to all available epochs i

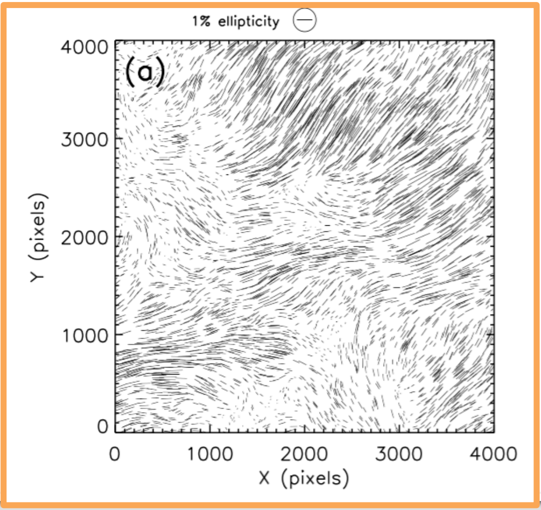


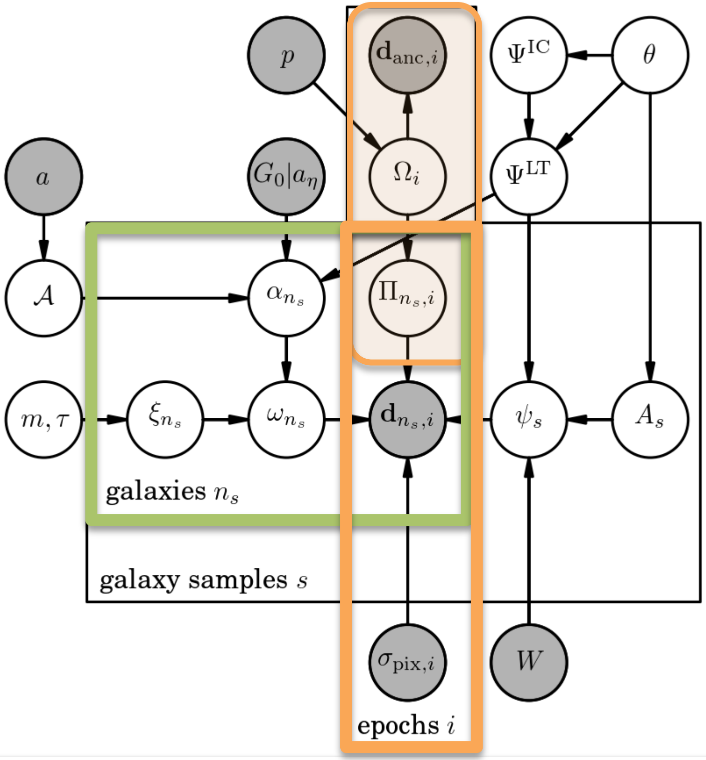


Naïve approach is intractable

1. Galaxy models be joint fitted to all available epochs i

2. PSF models must be joint fitted to all galaxies in an exposure n_s





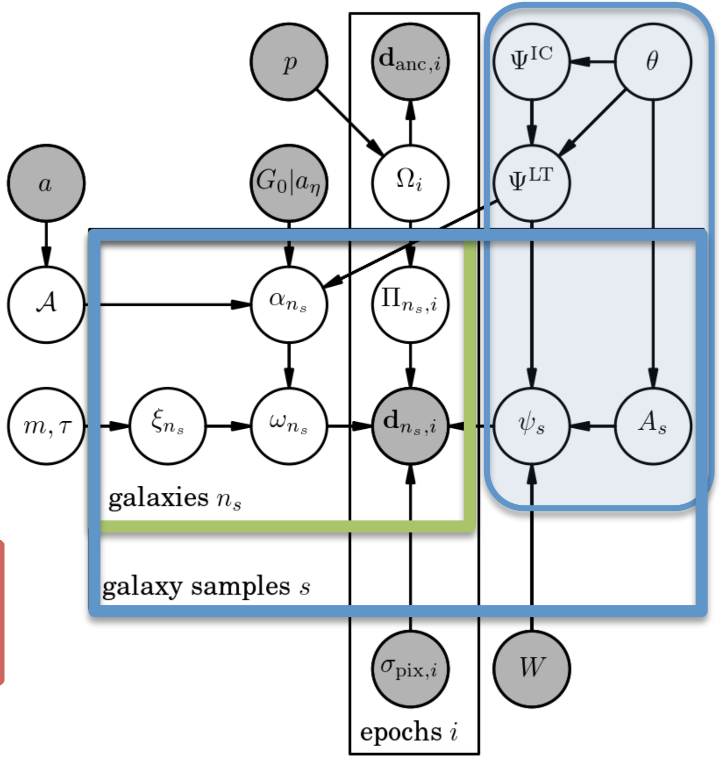
Naïve approach is intractable

1. Galaxy models be joint fitted to all available epochs i

2. PSF models must be joint fitted to all galaxies in an exposure n_s

3. Cosmology must be joint fitted to all galaxy samples & epochs

The principled inference requires fitting all pixels of all surveys simultaneously



Importance sampling to separate the sampling of individual galaxies: The pseudo-marginal likelihood

Want:

Galaxy dist. Likelihood

$$\Pr(\mathbf{d}|\alpha) \propto \prod_{n=1}^{n_{\rm gal}} \int d\omega_n \Pr(\omega_n|\alpha) \Pr(\mathbf{d}_{n,i}|\omega_n)$$

Have samples from:

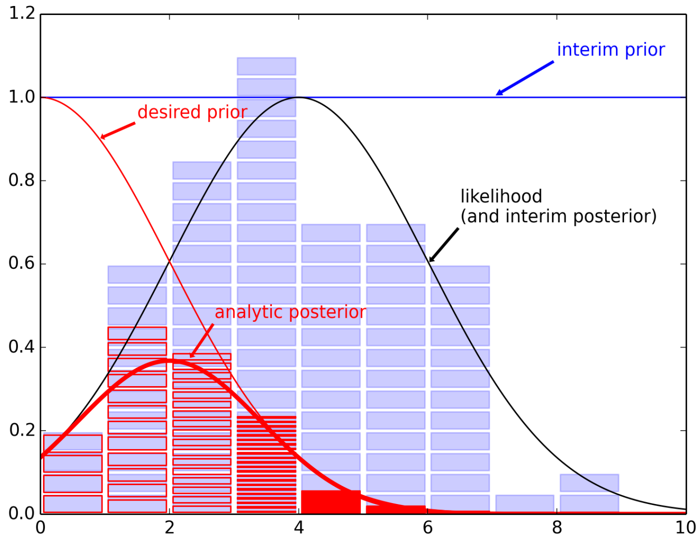
$$\Pr(\omega_n|\mathbf{d}_n, I_0)$$

'Interim prior' specification

Importance sampling:

$$\Pr(\mathbf{d}_n|\alpha) \approx \frac{Z_n}{K} \sum_k \frac{\Pr(\omega_{nk}|\alpha)}{\Pr(\omega_{nk}|I_0)},$$

$$\Pr(\mathbf{d}|\alpha) = \prod_{n=1}^{n_{\rm gal}} \Pr(\mathbf{d}_n|\alpha).$$



Credit: J. Meyers

How do we combine multiple observations of the same galaxy? Naïvely we must joint fit all epochs simultaneously

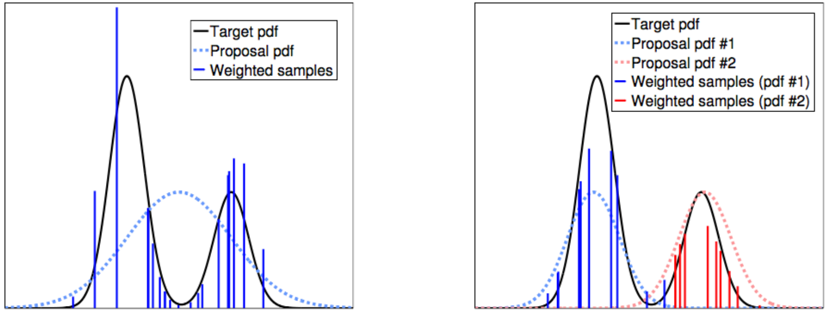
Problem: Imagine we have fit pixel data from LSST year 1.
How do we incorporate year 2 observations without redoing (expensive) calculations?

$$\Pr(\mathbf{d}_n|\alpha, \{\Pi_i\}) = \int d\omega_n \Pr(\omega_n|\alpha) \prod_{i=1}^{n_{\rm epochs}} \Pr(\mathbf{d}_{n,i}|\omega_n, \Pi_i)$$

Solution: Consider single-epoch samples as draws from a multi-modal importance sampling distribution:

$$q(\omega_n) = \frac{1}{n_{\rm epochs}} \sum_{i=1}^{n_{\rm epochs}} \Pr(\omega_n|\mathbf{d}_{n,i}, \Pi_i, I_0)$$

arXiv:1511.03095
Generalized Multiple Importance Sampling
Elvira, Martino, Luengo, & Bugallo



(a) Single proposal pdf (standard IS). (b) Two proposal pdfs (MIS).

Fig. 1: Approximation of the target pdf, $\pi(\mathbf{x})$, by the random measure χ.

Generalized multiple importance sampling (MIS) weights

MIS sampling distribution: sample from the conditional posterior for each epoch individually

$$q(\omega_n) = \frac{1}{n_{\text{epochs}}} \sum_{i=1}^{n_{\text{epochs}}} \Pr(\omega_n | \mathbf{d}_{n,i}, \Pi_i, I_0)$$

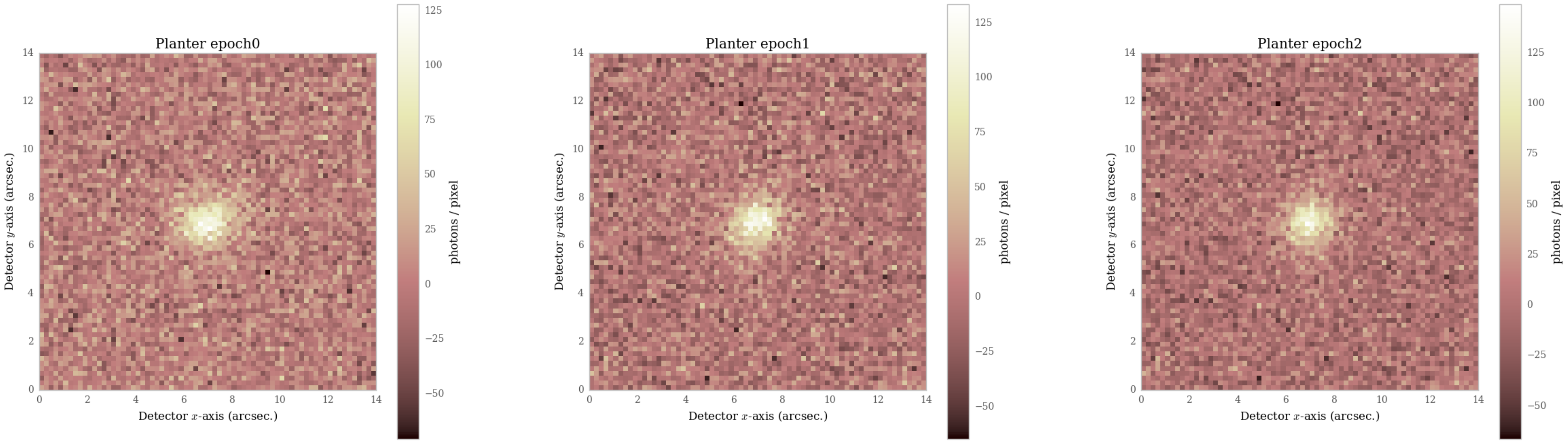
MIS weights: Evaluate the ratio of the conditional posterior for each epoch i to that of the MIS sampling distribution

$$w_i = \frac{\Pr(\mathbf{d}_{n,i}|\omega_n, \Pi_i)\Pr(\omega_n|\alpha)}{\sum_{i=1}^{n_{\text{epochs}}} \Pr(\mathbf{d}_{n,i}|\omega_n, \Pi_i)\Pr(\omega_n|I_0)}$$

'*cross-pollination*' needed:
Evaluate the likelihood of epoch i given model parameter samples from epoch j, for all combinations of i, j.
A standard scatter / gather operation

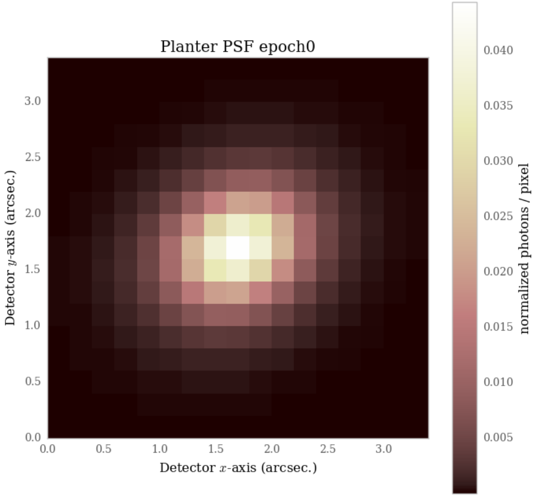
$$\Pr(\mathbf{d}_{n,i}|\omega_n^{(j)}\Pi_i)$$

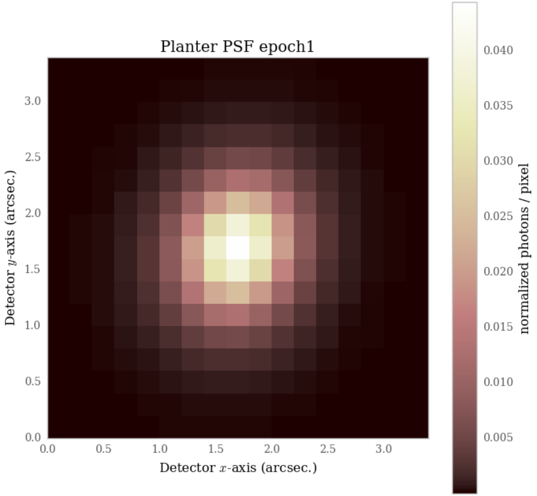
Example: 1 galaxy, 3 epochs – fit the galaxy model parameters

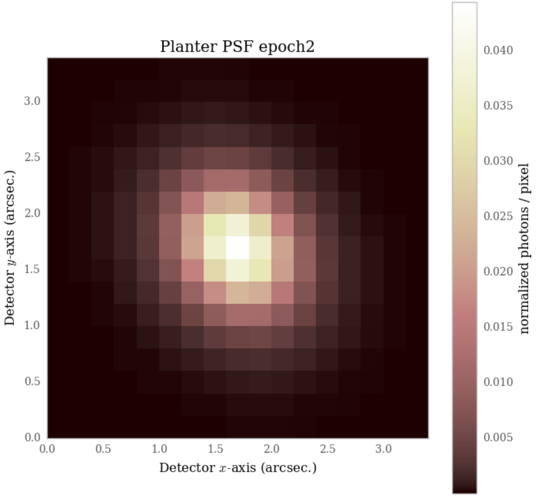


Each epoch has highly elliptical PSFs (|e| = 0.1) of same size, but different orientations

The PSF FWHM also matches the galaxy HLR making the single-epoch inferences noticeably different from each other. There is therefore a large gain of information in combining epochs.

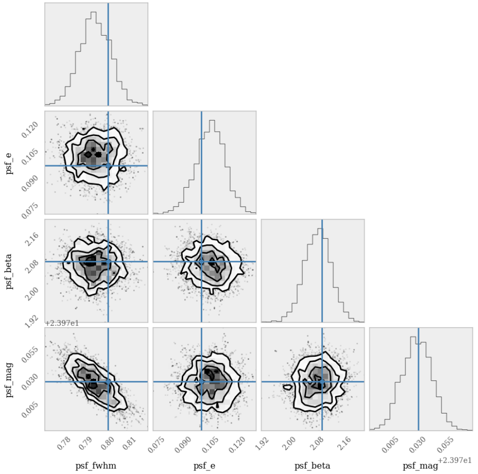




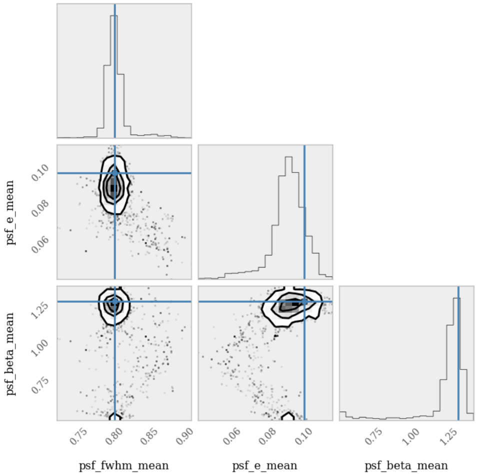


Interim posterior samples at each stage of the PSF hierarchical model

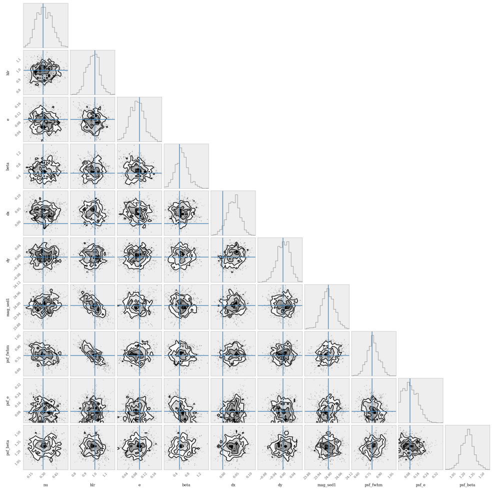
1) Fit stars



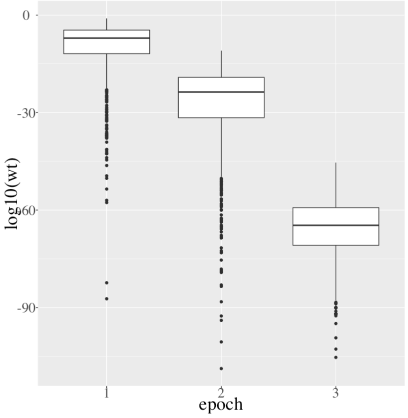
2) Constrain PSF model



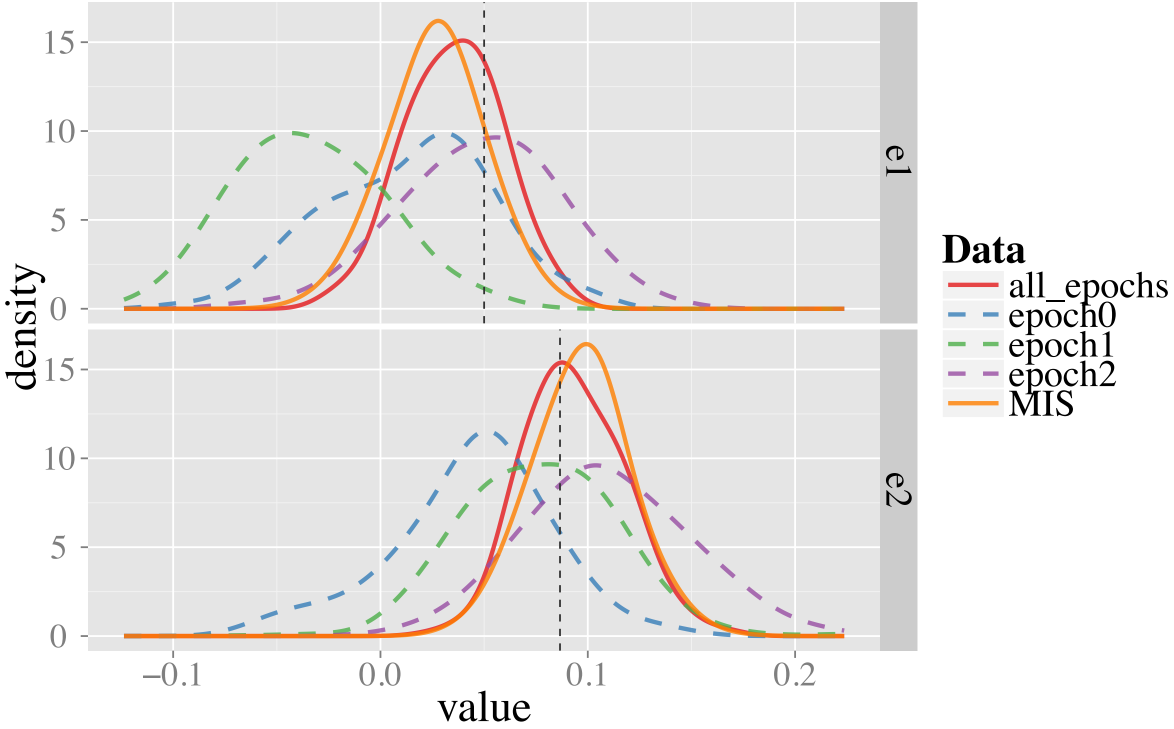
3) Fit galaxies & PSFs



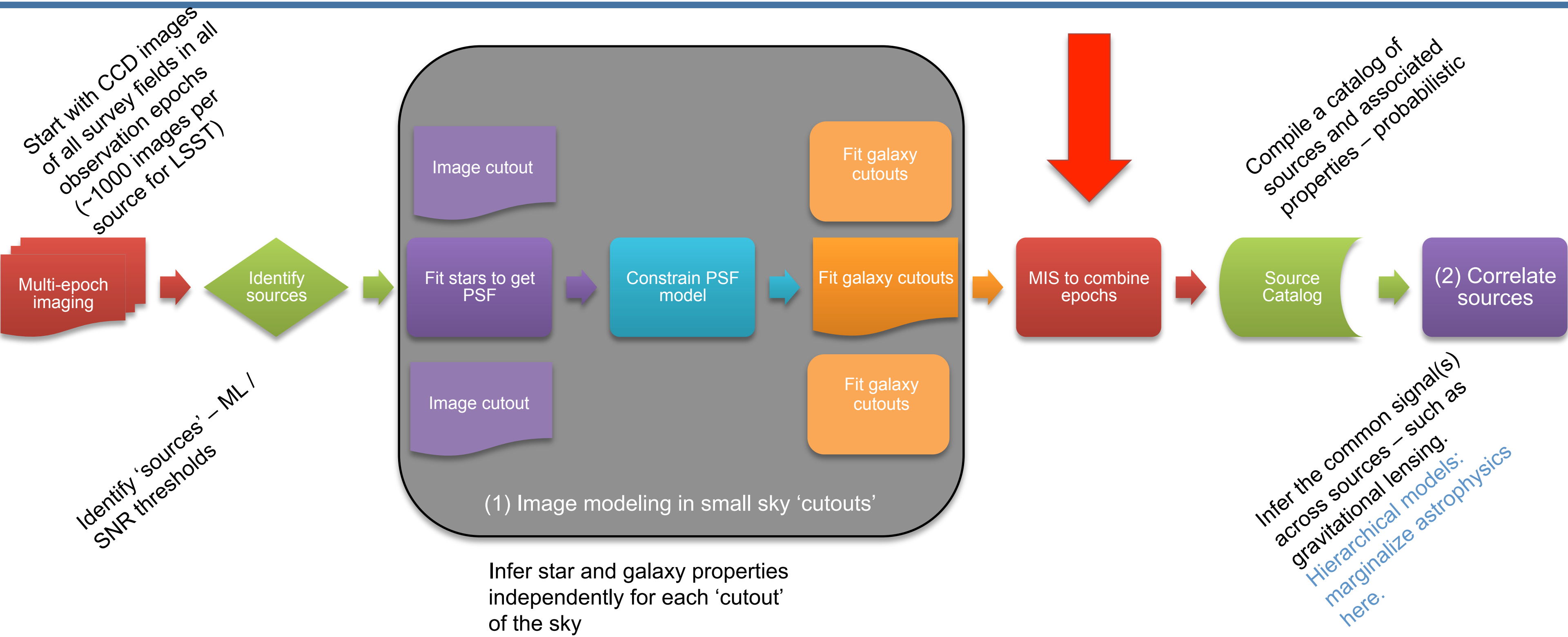
4) Calculate MIS weights to combine epochs



Comparison of single-epoch and combined epochs marginal posteriors



Data management and analysis pipeline: MIS-enabled



Infer star and galaxy properties independently for each 'cutout' of the sky

Conclusions

- Cosmic shear is systematics limited & signal is dominated by PSF and astrophysics
  - A probabilistic approach is warranted to infer a small signal and mitigate biases
- Importance sampling methods allow tractable approaches to a probabilistic forward model of LSST imaging
  - With billions of galaxies and hundreds of epochs per galaxy modeling LSST imaging requires an approach to separating analyses of data subsets, even though statistically correlated
- We are able to sample from a probabilistic model with multiple hierarchies to marginalize both correlated image systematics and astrophysical properties of galaxies

Lawrence Livermore National Laboratory

Outline

- The science: cosmic shear in LSST
- Probabilistic pipeline introduction
- The problem of multiple epochs and MIS solution
- MIS example
- Pipeline redux in light of MIS
- Conclusions

Marginalizing PSFs: MIS makes this tractable

- LSST will have ~200 epochs per object per filter
  - We aim to marginalize the PSF $\prod_{n,i}$ in every epoch
  - The marginalization is constrained by:
    - Consistency of PSF realizations over the focal plane for each epoch
    - Consistency of the underlying source model across epochs
- Simplest approach (statistically, not computationally): Infer galaxy models given all epoch imaging simultaneously
  - "Interim" samples are of size: ~10 galaxy params + 200 * ~4 PSF params = ~1k parameters!