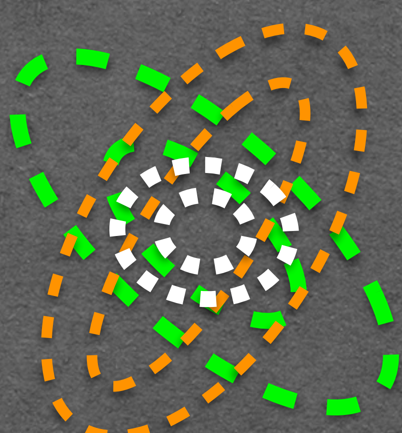
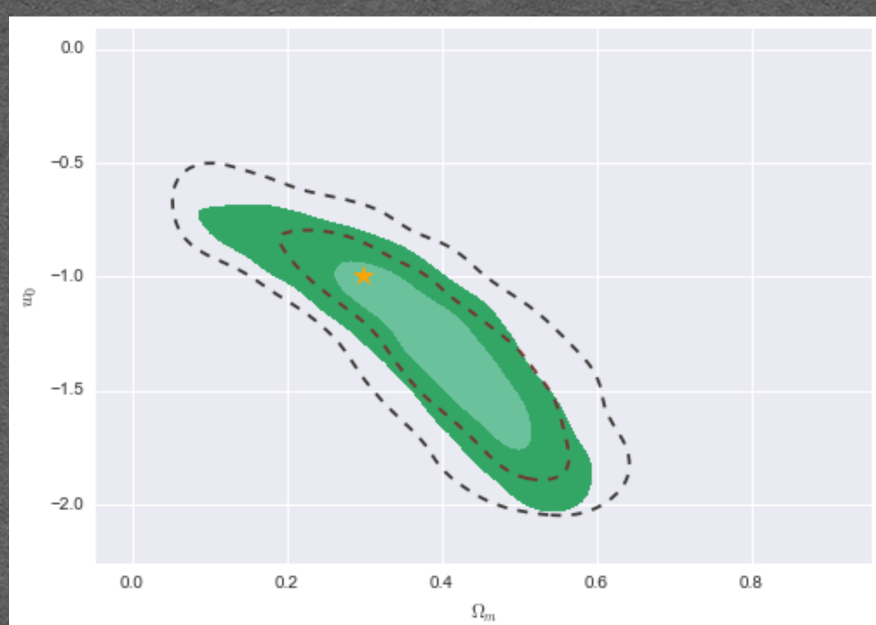
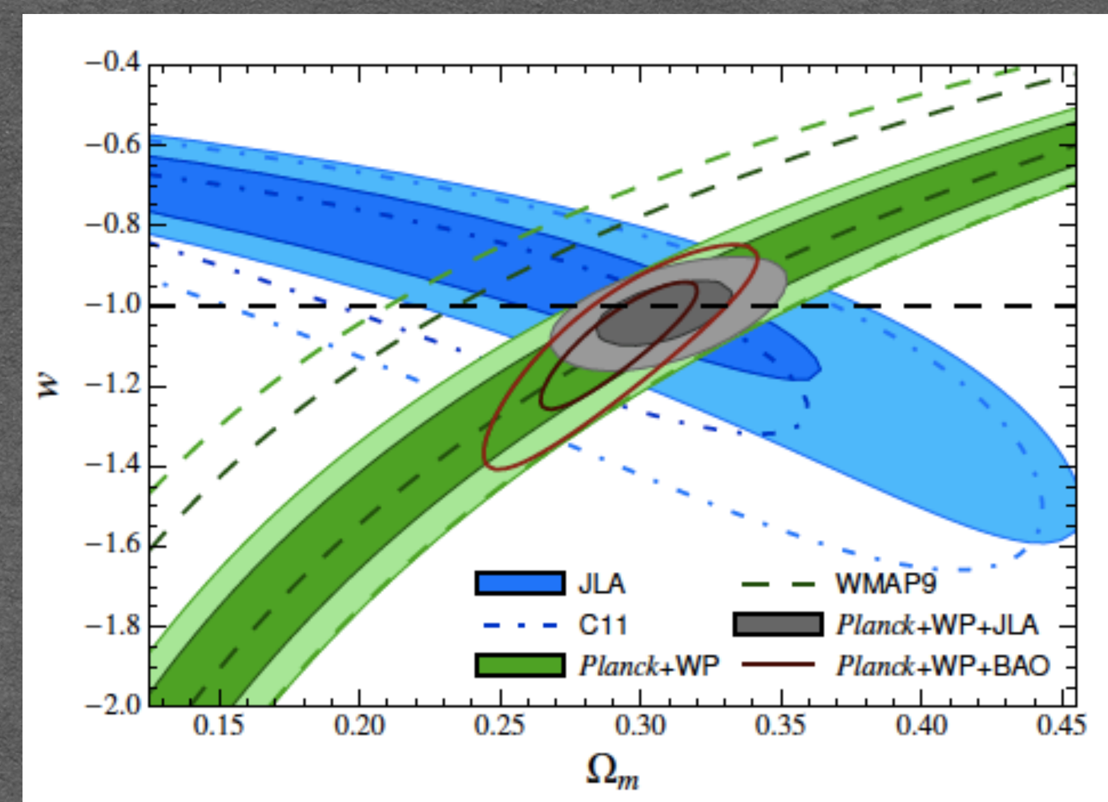
New statistical frameworks & methods for precision cosmology

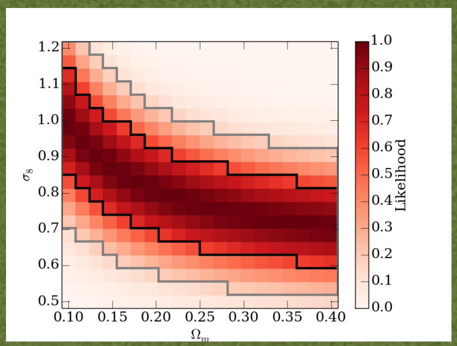
Elise Jennings



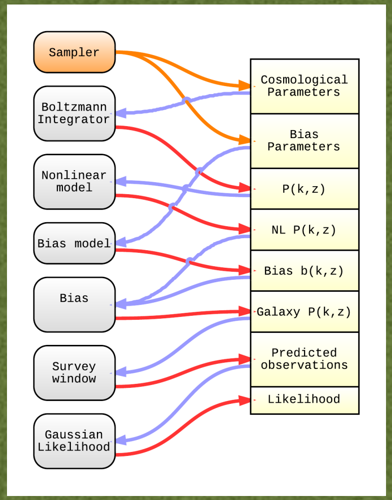




Outline



CosmoSIS



-1-2% precision on cosmological parameter constraints,multi probe approach

modularity is the key to addressing these challenges!

Advanced statistical techniques Sampling beyond Metropolis-Hastings

-Particle/Walkers in Monte Carlo methods
-Approximate Bayesian Computation

Classical estimation

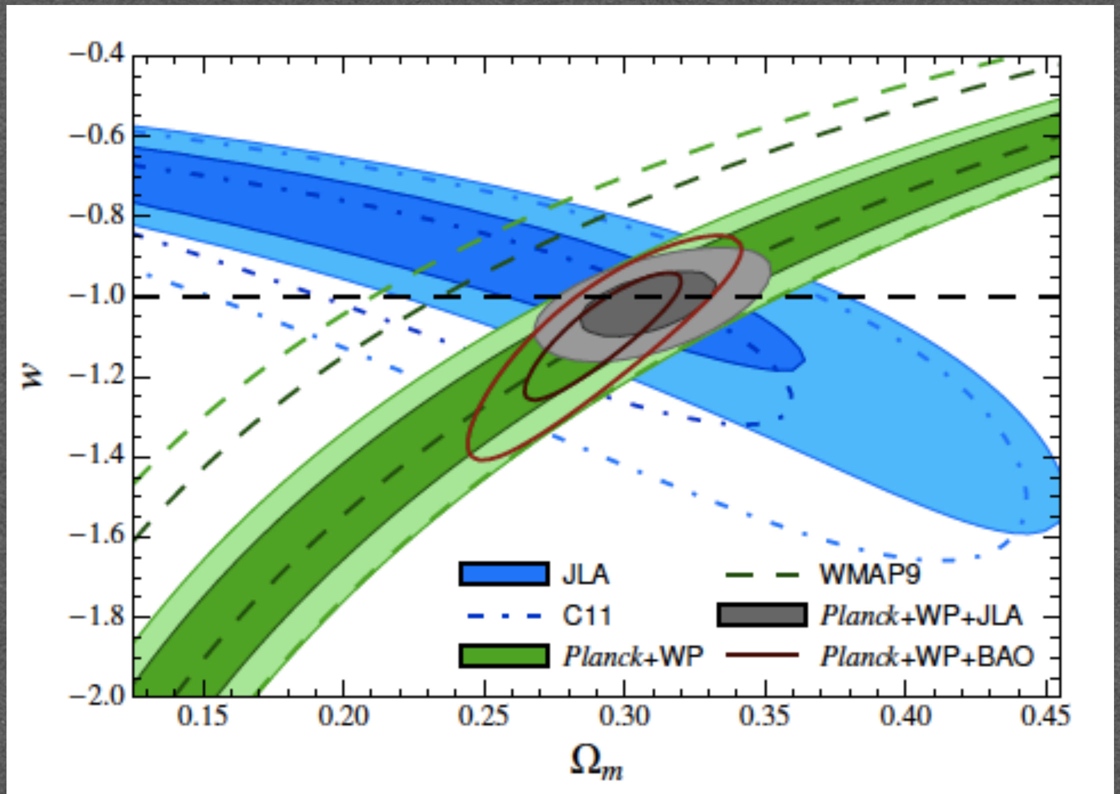
Likelihood

prior
probability

$$p(\theta|\mathbf{D}) = \frac{\mathcal{L}(\mathbf{D}|\theta)\pi(\theta)}{\mathbf{p}(\mathbf{D})}$$

Posterior probability
distribution

- Given some data and model with free parameters
- Figure out max likelihood values or confidence intervals for parameters
- Markov Chain Monte Carlo: efficient estimation of likelihood in multi-dim space



Betoule
et al 2014

Current issues for precision cosmology with LSST:

- Many, sometimes correlated, **observables**:
  - -CMB, lensing, galaxy clustering, supernovae, clusters…..
  - -most powerful constraints come from combining 2 or more

Current issues for precision cosmology with LSST:

- Many, sometimes correlated, **observables**:
  - -CMB, lensing, galaxy clustering, supernovae, clusters…..
  - -most powerful constraints come from combining 2 or more

- Different **theoretical models**:
  - -e.g. Supernovae light curve fitters, bias models for galaxy clustering

Current issues for precision cosmology with LSST:

- Many, sometimes correlated, **observables**:
  - -CMB, lensing, galaxy clustering, supernovae, clusters…..
  - -most powerful constraints come from combining 2 or more

- Different **theoretical models**:
  - -e.g. Supernovae light curve fitters, bias models for galaxy clustering

- Different **parameters, systematics** in each model:
  - -how to sample over each in an MCMC chain?

Current issues for precision cosmology with LSST:

• Many, sometimes correlated, **observables**:
  - CMB, lensing, galaxy clustering, supernovae, clusters…..
  - most powerful constraints come from combining 2 or more

• Different **theoretical models**:
  - e.g. Supernovae light curve fitters, bias models for galaxy clustering

• Different **parameters, systematics** in each model:
  - how to sample over each in an MCMC chain?

• Complicated, possibly multimodal, **Posterior/ Likelihoods**:
  - sampling choice may impact results, estimate/model covariances

Current issues for precision cosmology with LSST:

- Many, sometimes correlated, **observables**:
  - -CMB, lensing, galaxy clustering, supernovae, clusters…..
  - -most powerful constraints come from combining 2 or more

- Different **theoretical models**:
  - -e.g. Supernovae light curve fitters, bias models for galaxy clustering

- Different **parameters, systematics** in each model:
  - -how to sample over each in an MCMC chain?

- Complicated, possibly multimodal, **Posterior/ Likelihoods**:
  - -sampling choice may impact results, estimate/model covariances

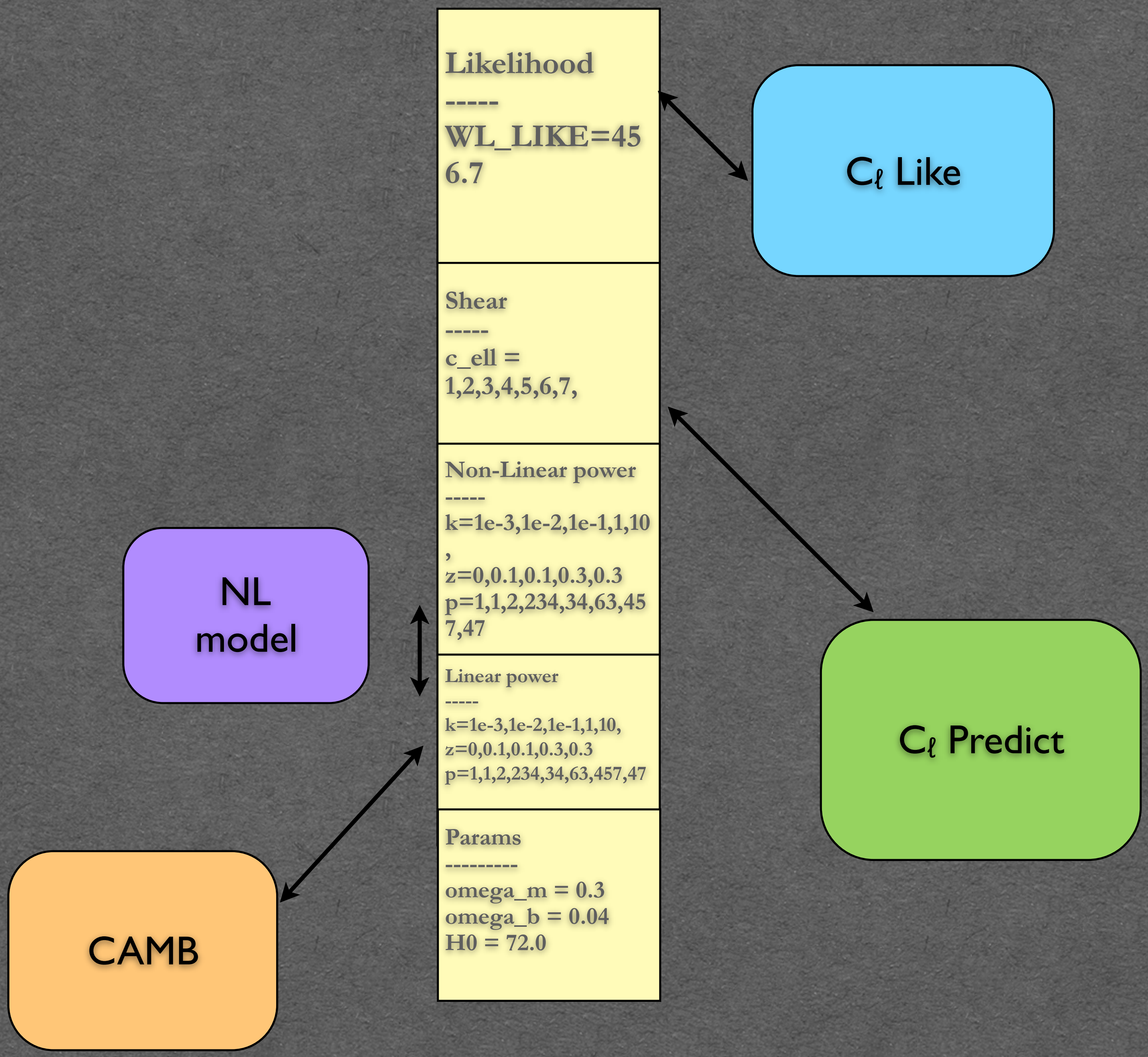
- Large **collaborations**:
  - -how to track contributions, ensure reproducibility & consistency
  - -how to use wealth of exisiting code/data without wasting PhD deciphering it all, learning new coding language…

CosmoSIS

<u>Development Team</u>

Joe Zuntz, Marc Paterno, Elise Jennings, Doug Rudd, Alessandro Manzotti, Scott Dodelson, Sarah Bridle, Saba Sehrish, Jim Kowalkowski

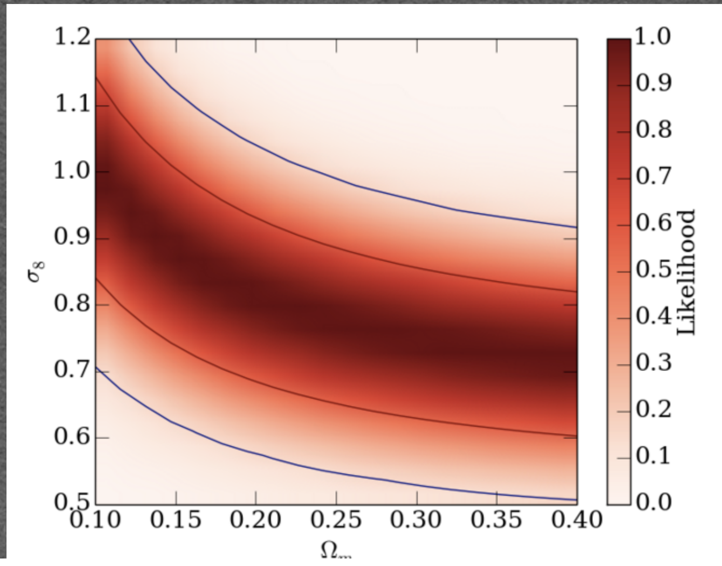
2015 break-out paper already cited 18 times, and over ~100 users have downloaded the framework and signed up to mailing list.

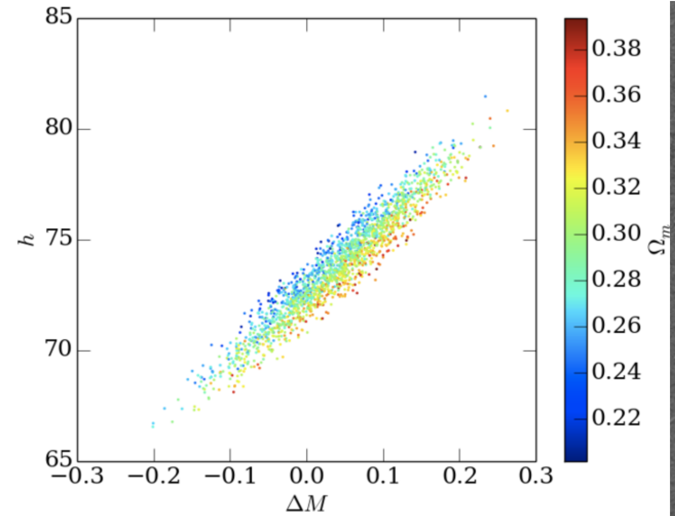


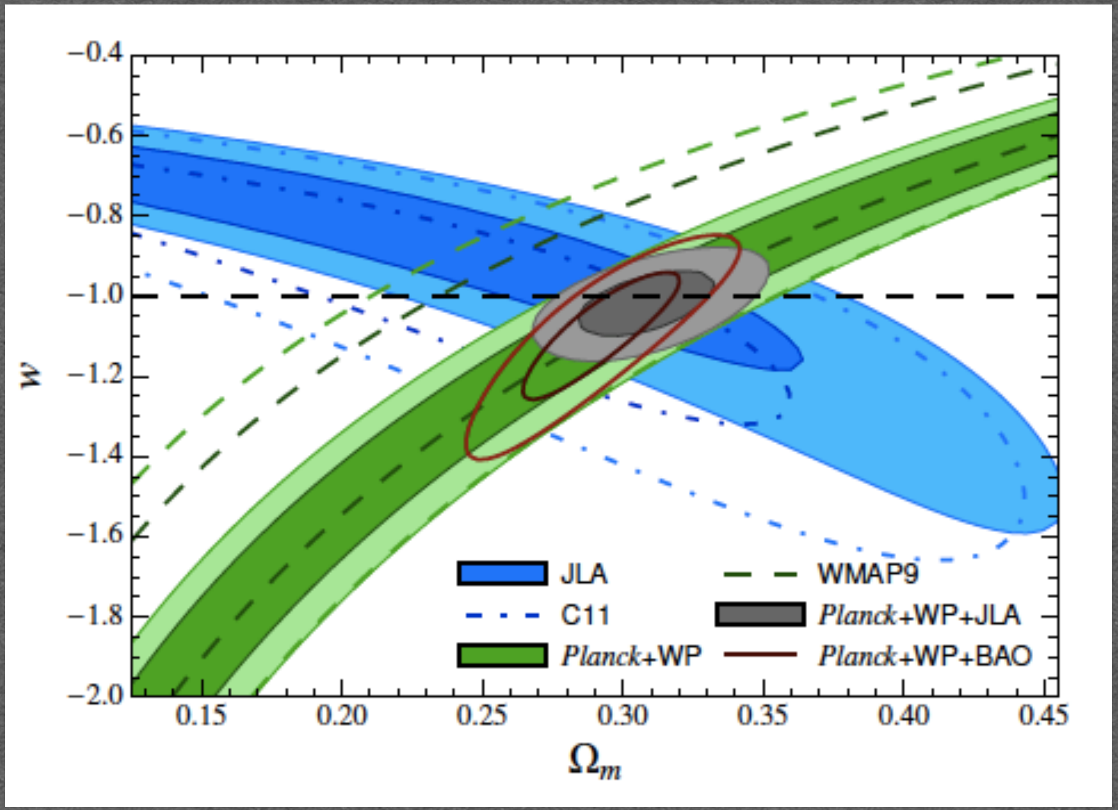
CosmoSIS was designed to address each of the issues facing future surveys

CosmoSIS is a new cosmology parameter estimation code focus on *modularity*

- Open source code which community actively contributes to
- multi - language modules: Python, C++, C, Fortran
- choice of physics & likelihood modules
- collection of samplers - mostly in python
- nice python plotting functions







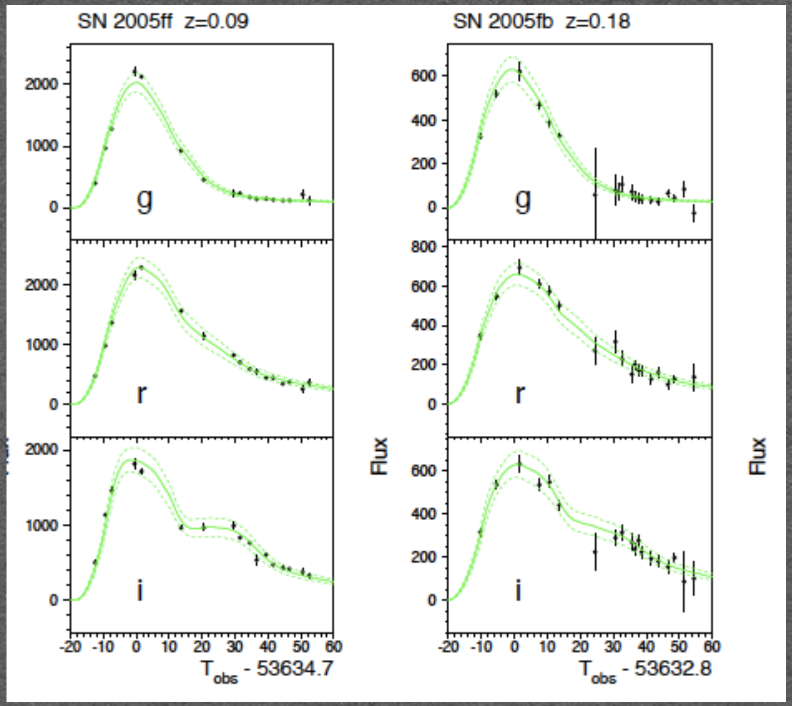
Betoule
et al 2014

Let's say you work on Supernovae but want to make this plot showing joint constraints from SN, Planck, WMAP, BAO

How much work is involved?

$$-2\ln\mathcal{L} = \sum_{\mathrm{i}} \frac{\mu_{\mathrm{i}}^{\mathrm{data}} - \mu_{\mathrm{i}}^{\mathrm{model}}(\mathrm{z_i}, \mathrm{w}, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, \mathrm{H_0})}{\sigma_{\mu}^2} + \chi^2_{\mathrm{BAO}} + \chi^2_{\mathrm{CMB}}$$

Supernovae



-decide on light curve fitter:
SALT2, MLCS2K2
(different parameters for each which could be varied!)

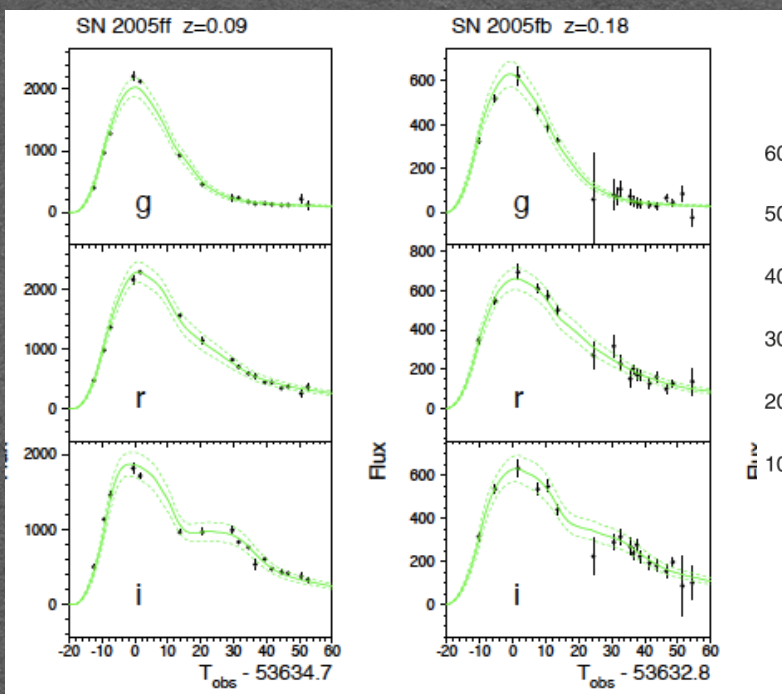
-e.g. JLA Likelihood (C++)

-simulations? SNcosmo (python), SNANA (C/Fortan)

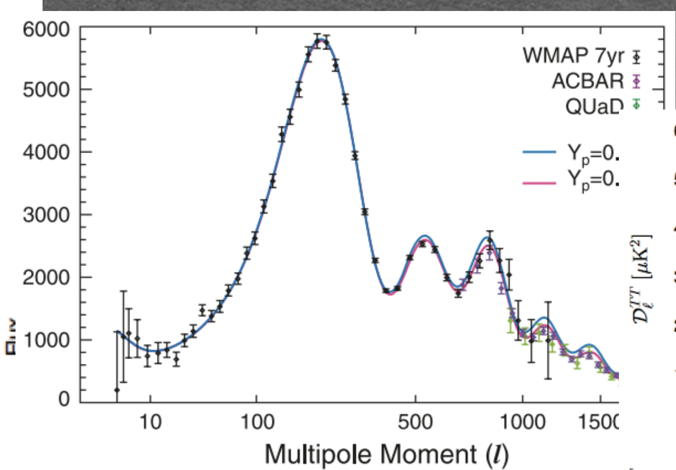
-Dynamical DE params? priors on parameters?

$$-2\ln\mathcal{L} = \sum_{\mathrm{i}} \frac{\mu_{\mathrm{i}}^{\mathrm{data}} - \mu_{\mathrm{i}}^{\mathrm{model}}(\mathrm{z_i}, \mathrm{w}, \Omega_{\mathrm{m}}, \Omega_\Lambda, \mathrm{H_0})}{\sigma_\mu^2} + \chi^2_{\mathrm{BAO}} + \chi^2_{\mathrm{CMB}}$$

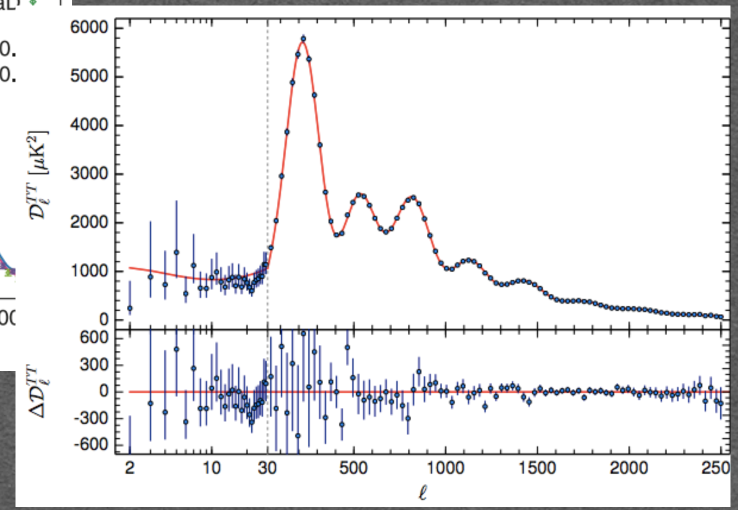
Supernovae



WMAP



Planck



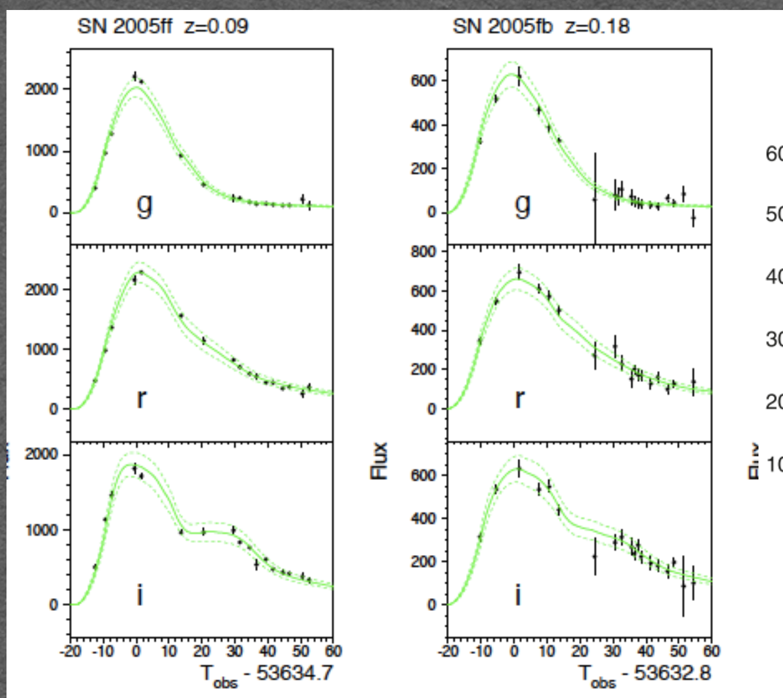
-Boltzmann code, CAMB?,CLASS?, MGCAMB?
Different parameters for each of these

-WMAP (Fortran), Planck likelihood (C), Planck2013/2015? WMAP7 or WMAP9?

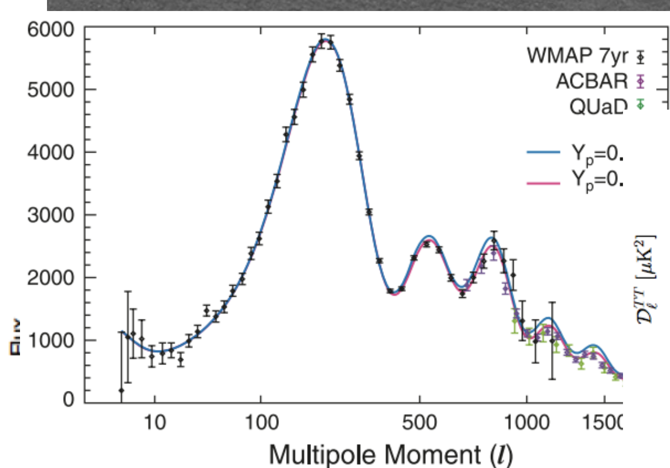
-code up something just for shift parameter?

$$-2\ln\mathcal{L} = \sum_{\mathrm{i}} \frac{\mu_{\mathrm{i}}^{\mathrm{data}} - \mu_{\mathrm{i}}^{\mathrm{model}}(\mathrm{z_i}, \mathrm{w}, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, \mathrm{H_0})}{\sigma_\mu^2} + \chi^2_{\mathrm{BAO}} + \chi^2_{\mathrm{CMB}}$$

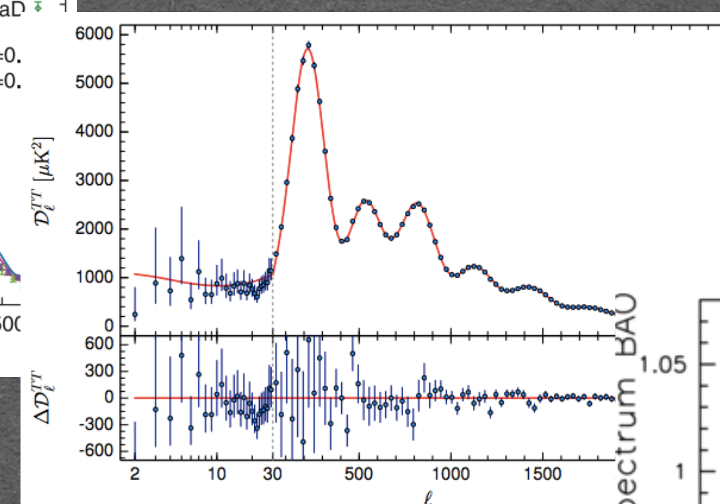
Supernovae



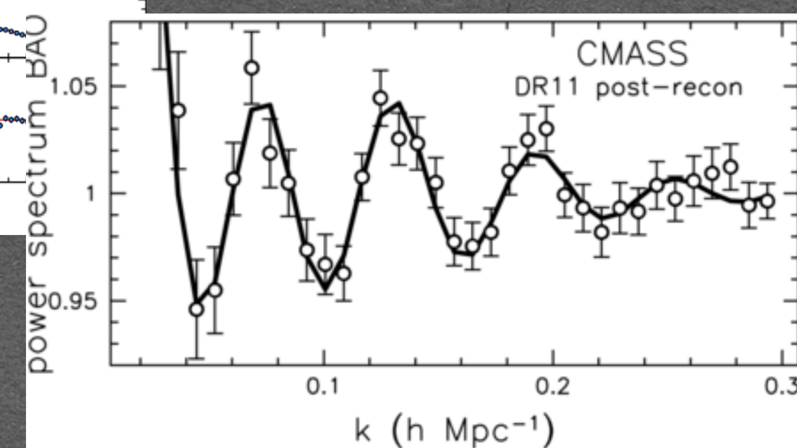
WMAP



Planck



BAO



-Find or code up something to estimate

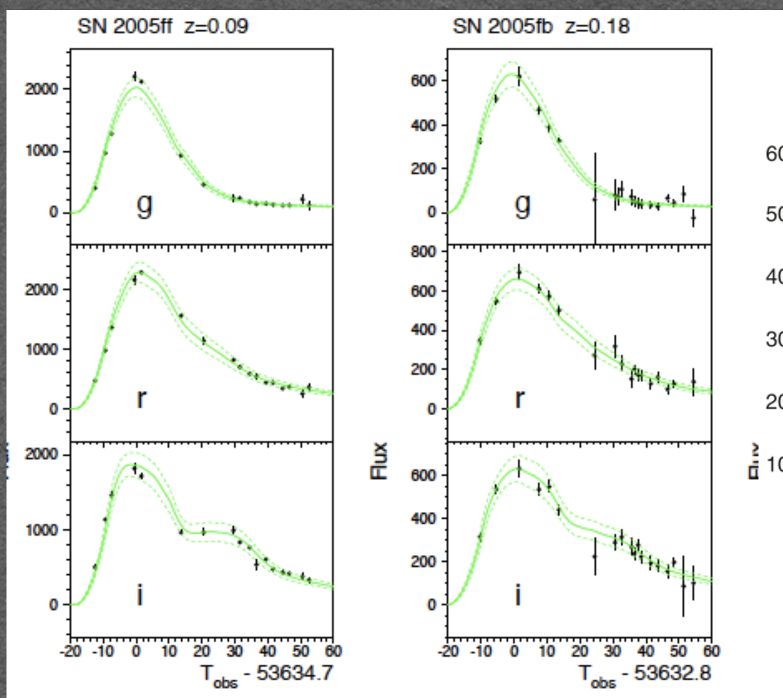
$P(k), \xi(r), D_A(z),$

$f(z), f\sigma_8$

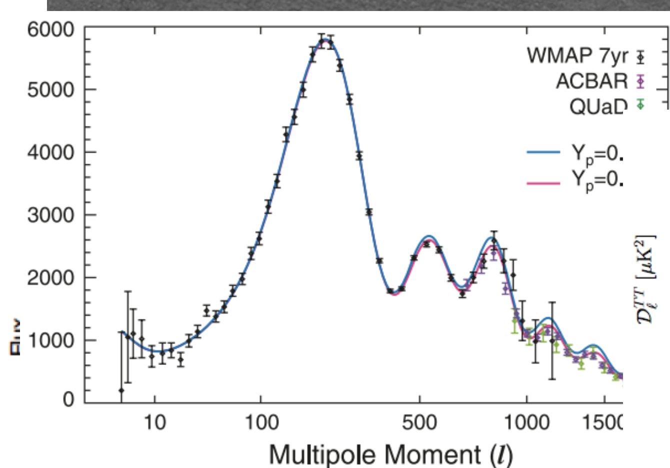
-model for bias, RSD?

$$-2\ln\mathcal{L} = \sum_{\rm i} \frac{\mu_{\rm i}^{\rm data} - \mu_{\rm i}^{\rm model}(\rm z_i, w, \Omega_m, \Omega_\Lambda, H_0)}{\sigma_\mu^2} + \chi^2_{\rm BAO} + \chi^2_{\rm CMB}$$

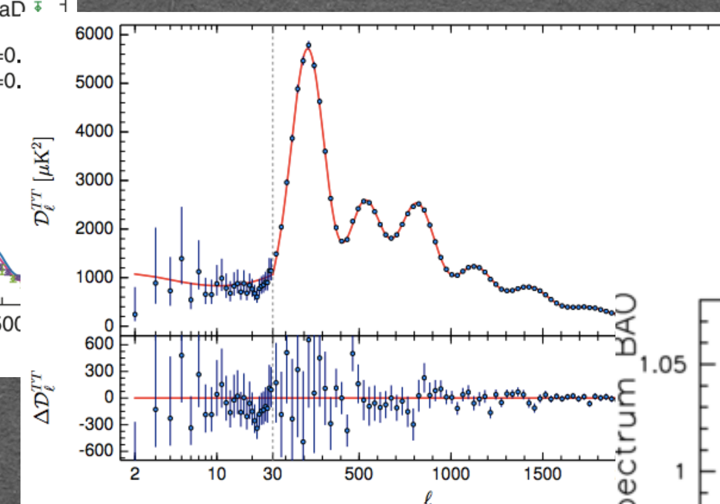
Supernovae



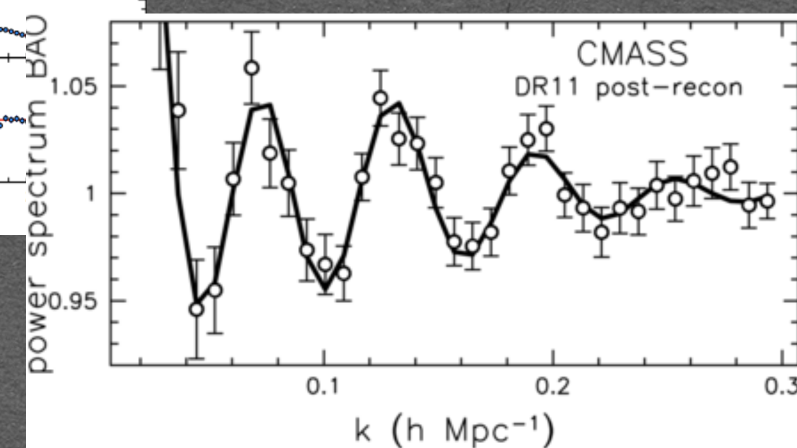
WMAP



Planck



BAO



After gathering all the pieces together you would still have to:

-write sampling algorithm or implement public sampler
-make sure shared cosmology parameters are treated properly
-decide on how to treat nuisance parameters for each observable, priors?
-decide on covariance matrix? from simulations?

How to do this in CosmoSIS…

use single **ini file** create a pipeline of modules

```
[pipeline]
modules = consistency camb jla riess11 wmap_shift planck growthfunction boss
values = demos/values.ini
extra_output =
likelihoods = jla riess wmap_shift planck2015 boss
quiet=T
debug=F
timing=F
```

- All of these modules currently in CosmoSIS!
- choose CAMB here, or could replace with CLASS/MGCAMB w/o recompiling
- Same cosmological parameters, shared systematic parameters across pipeline
- values, prior ranges for parameters specified in values.ini file
- swop in/out modules with ease
- add your own module easily in any of C, C++, Fortran and python

use **ini file** to specify the sampler you want

```
[runtime]
sampler = multinest

[multinest]
max_iterations=50000
live_points=500
multinest_outfile_root=
```

OR

```
[runtime]
sampler = metropolis

[metropolis]
nsteps=10
random_start=F
samples=100000
covmat=examples/covmat_a.txt
```

- many different samplers available!
- specify settings for individual modules

```
[camb]
file = cosmosis-standard-library/boltzmann/camb/camb.so
mode=all
lmax=2550
feedback=2
do_tensors=T
do_lensing=T
do_nonlinear=T
high_ell_template=cosmosis-standard-library/boltzmann/camb/camb_Jan15/
HighLExtrapTemplate_lenspotentialCls.dat
```

CosmoSIS standard library

- Boltzman: CAMB, CLASS, Halofit, Halofit_Takahashi, MGCAMB, ISITGR
- Intrinsic alignments
- Likelihoods: WiggleZ, WMAP 4.1 & 5, Planck 2013/2015, BBN, Bicep, BOSS, DES, CFHTLens, cluster masses, fgas, Riess SN
- Mass function, luminosity function
- Number density: window functions for photometric n(z), photoz bias
- Shear: intrinsic aligments, compute correlation functions
- Strong lensing
- Supernovae: JLA likelihood, SALT2 fitter
- Large scale structure: Pk emulator, FrankenEmu, growth rate/factor

Getting started with CosmoSIS - users

Try some short demos

These demos illustrate basic cosmosis concepts in action.

- Demo 1: Get standard cosmological functions for a given cosmology
- Demo 2: Get the Planck and BICEP likelihood for a given cosmology
- Demo 3: Get a slice through the BICEP likelihood in r_T
- Demo 4: Find the best fit cosmological parameters for a Planck likelihood
- Demo 5: Running an MCMC analysis of supernova data
- Demo 6: Get the CFHTLenS likelihood for a given cosmology
- Demo 7: Get a 2D likelihood grid from BOSS DR9 measurements of f*sigma_8
- Demo 8: Compare CAMB with Eisenstein & Hu; get growth factors; tweak plots
- Demo 9: Get Bayesian evidence with Multinest; make scatter and custom plots
- Demo 10: Compare constraints with two different halofit models for nonlnear power, sampling sigma_8 instead of A_s
- Demo 11: Slice through constraints on modified gravity with isItGR and CFHTLenS; save full data for all grid points to file
- Demo 12: Get the PDF for extreme cluster masses using the Tinker mass function
- Demo 13: A fast grid sampling of the JLA supernovae using the Snake sampler
- Demo 14: The Kombine sampler applied to cluster gas fraction measurements; advanced postprocessing options
- Demo 15: More Galaxy Two-Point Correlation Functions
- Demo 16: Robust maximum-posteriors with the Minuit sampler
- Demo 17: A Fisher Matrix for the Dark Energy Survey SV cosmic shear

Longer demos

These longer examples show realistic parameter constraint pipelines.

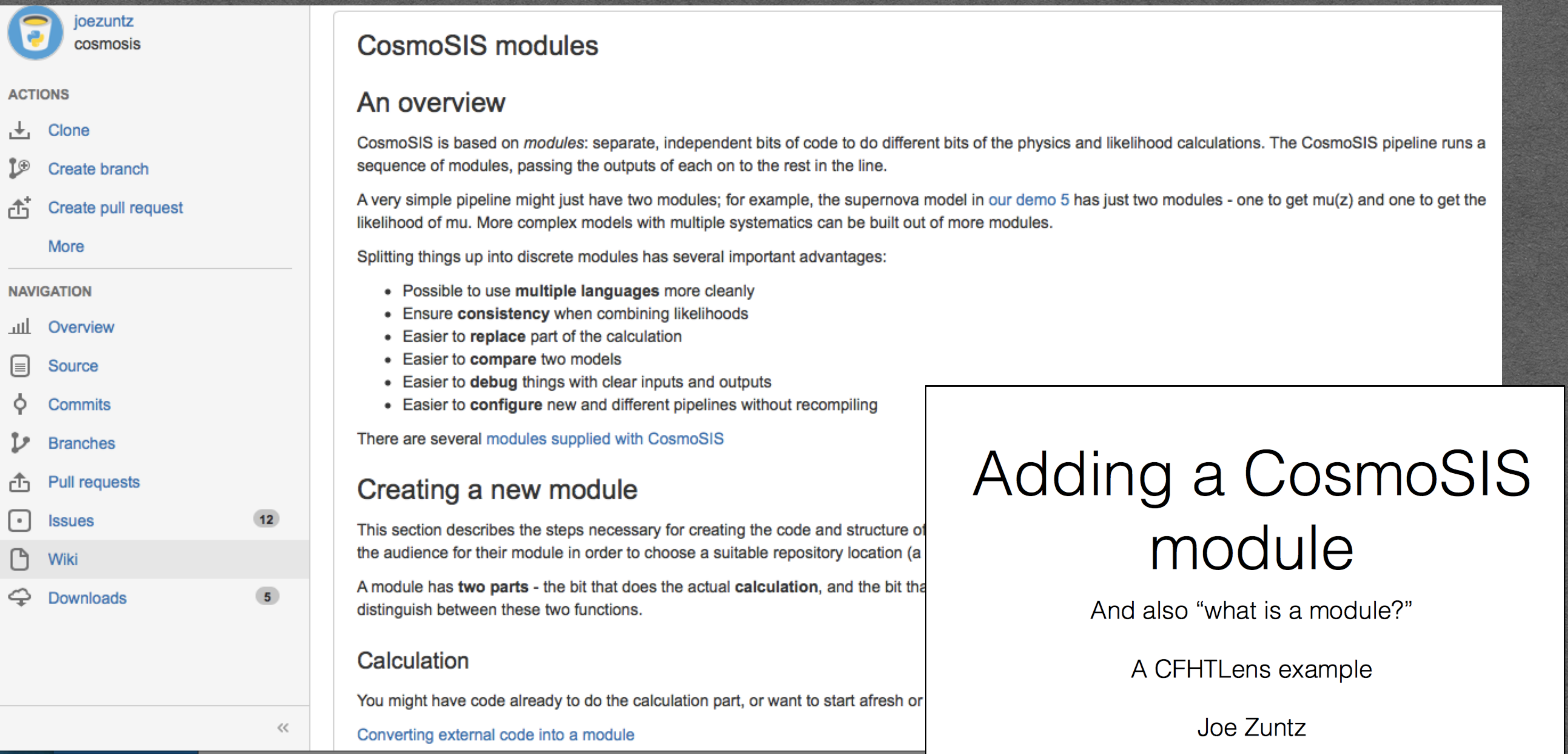
- Example A: Simple CosmoMC-like Metropolis-Hastings analysis for WMAP9

setup-cosmosis script uses UPS commands to activate appropriate software and build CosmoSIS environment

https://bitbucket.org/joezuntz/cosmosis/wiki/Home

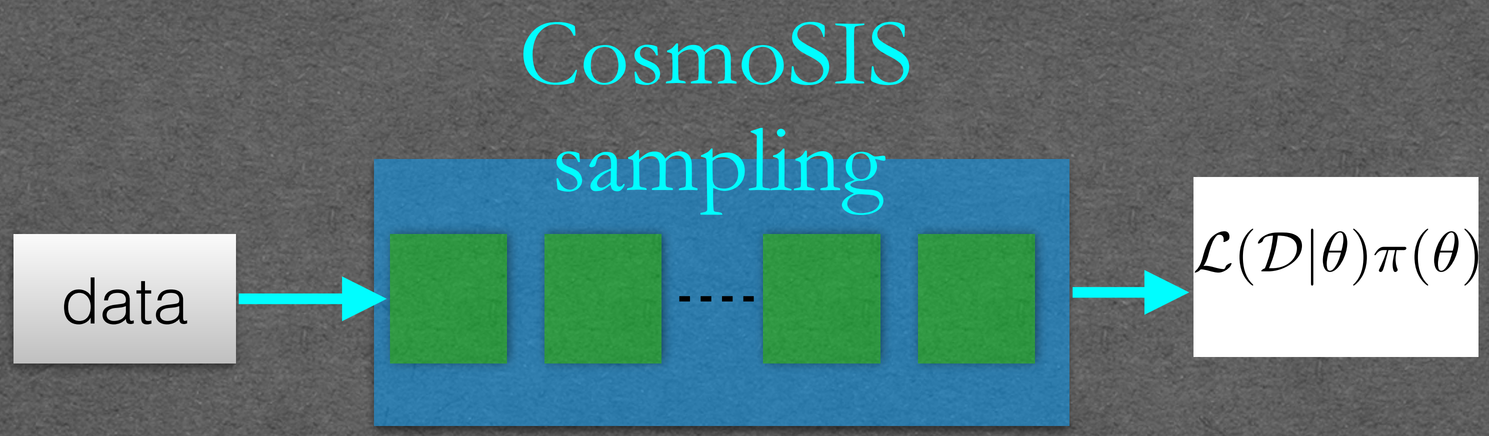
Getting started with CosmoSIS - developers

- We invite contributors to the CosmoSIS Standard Library
- Clone repo into a new directory or use existing DES repo, LSST-DESC repo
- sign up to mailing list on wiki



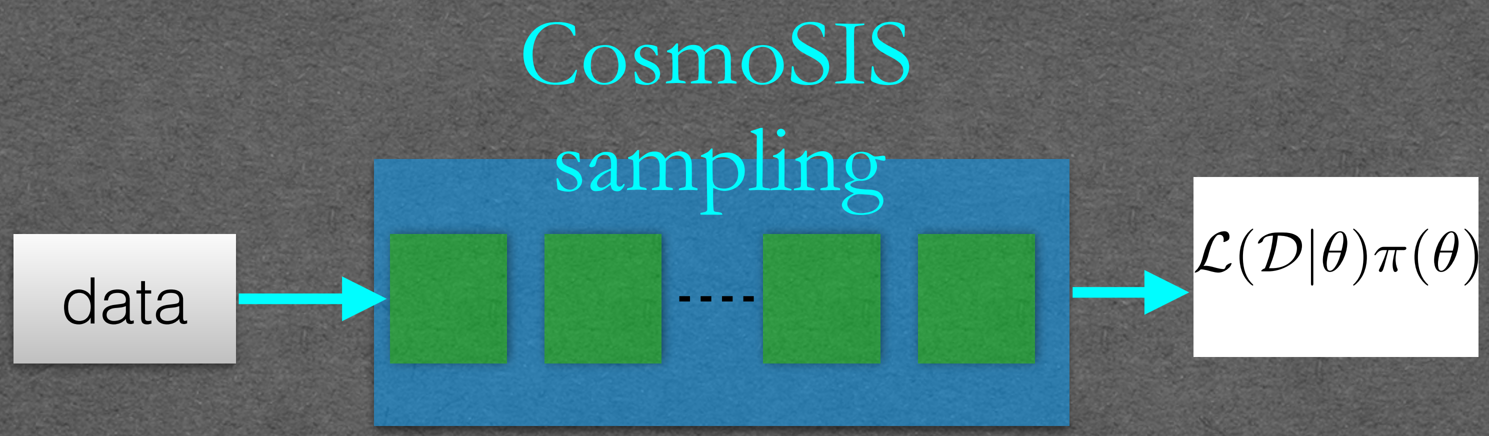
CosmoSIS
getting ready for LSST

- Currently people use CosmoSIS like this



CosmoSIS
getting ready for LSST

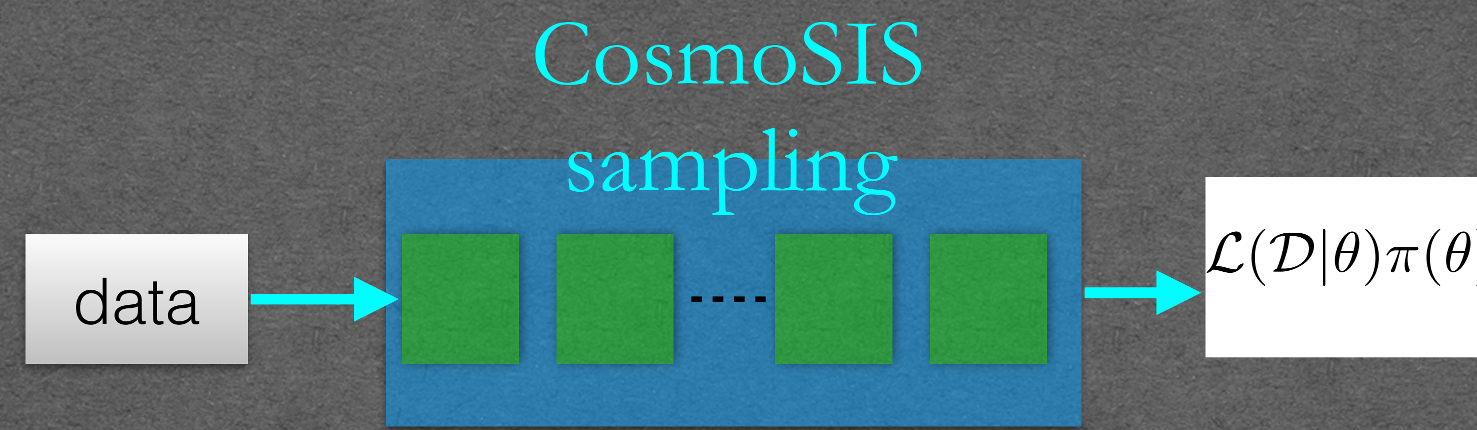
- Currently people use CosmoSIS like this



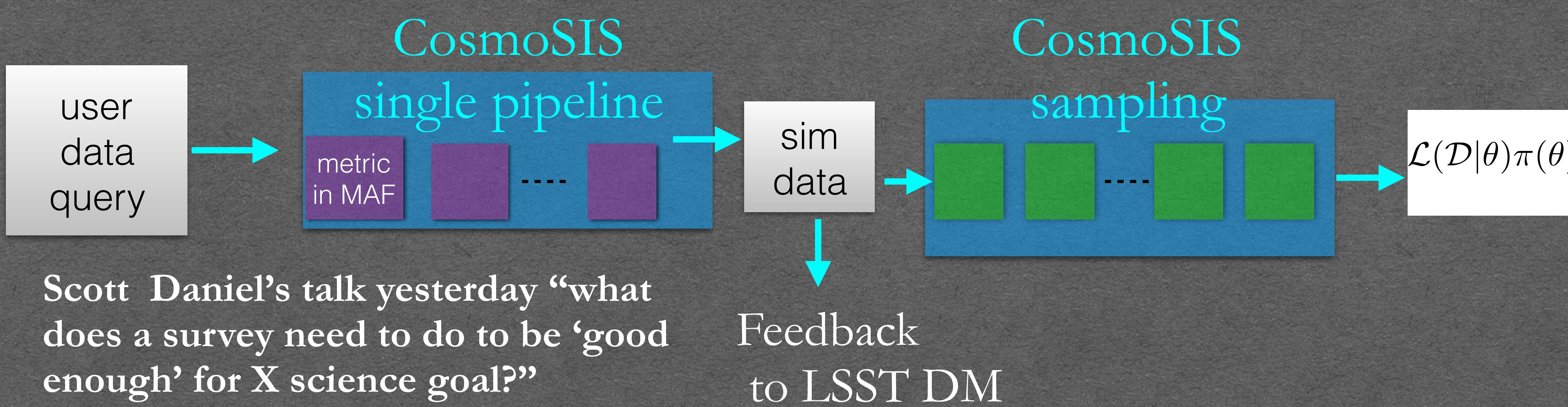
- Infrastructure can be used without sampling to call CatSim, ImSim to test the LSST Data Management pipeline. Scott Dodelson co-convenor of LSST DESC Computing and Infrastructure Working Group

CosmoSIS
getting ready for LSST

- Currently people use CosmoSIS like this



- Infrastructure can be used without sampling to call CatSim, ImSim to test the LSST Data Management pipeline.



Scott Daniel's talk yesterday "what does a survey need to do to be 'good enough' for X science goal?"

CosmoSIS
getting ready for LSST

- Connecting with Informatics and Statistics Science Collaboration, Chad Schafer, co-chair

Informatics & Statistics
Science Collaboration

- Platform independent distribution via VM:
Docker images, easy install on NERSC

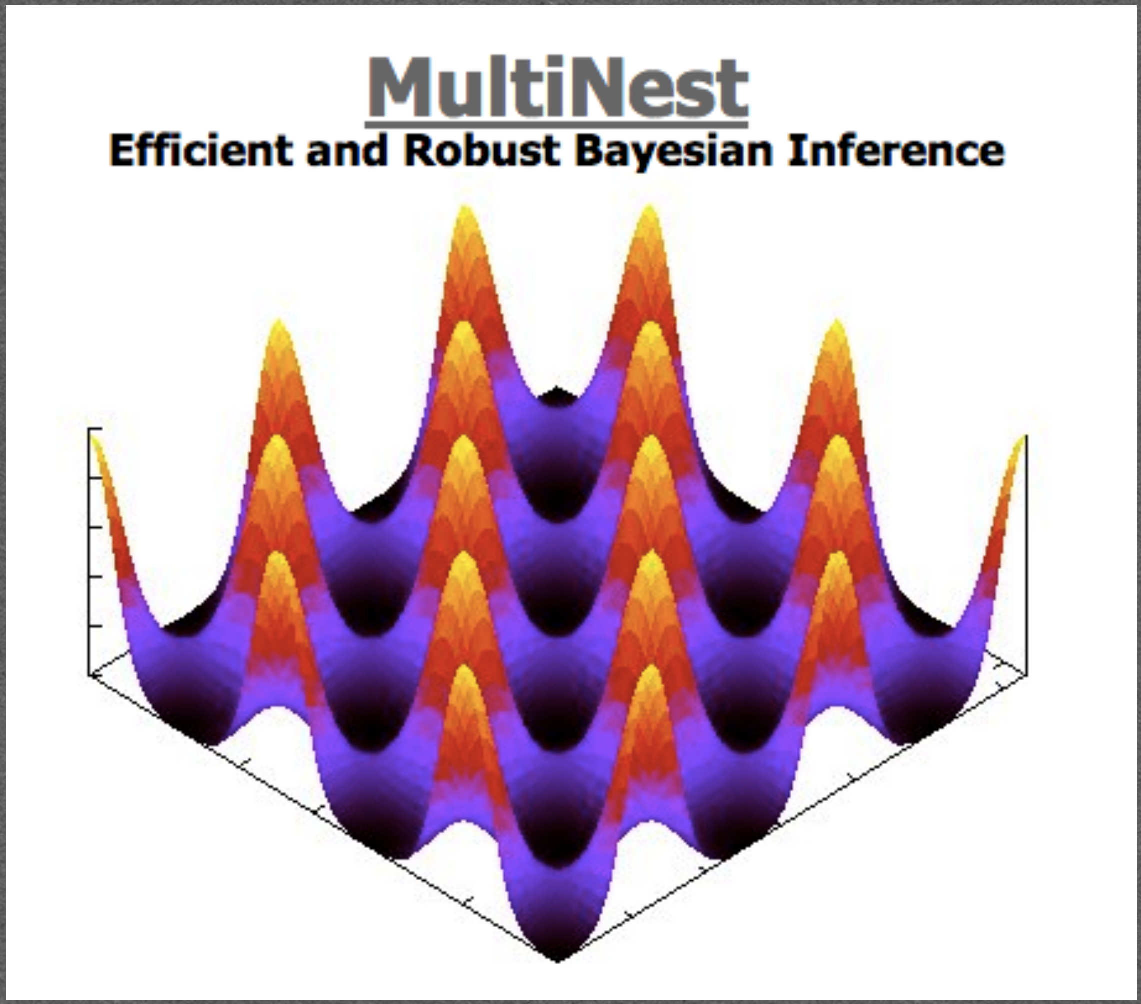
- Fermilab LSST DESC hack week April 2017

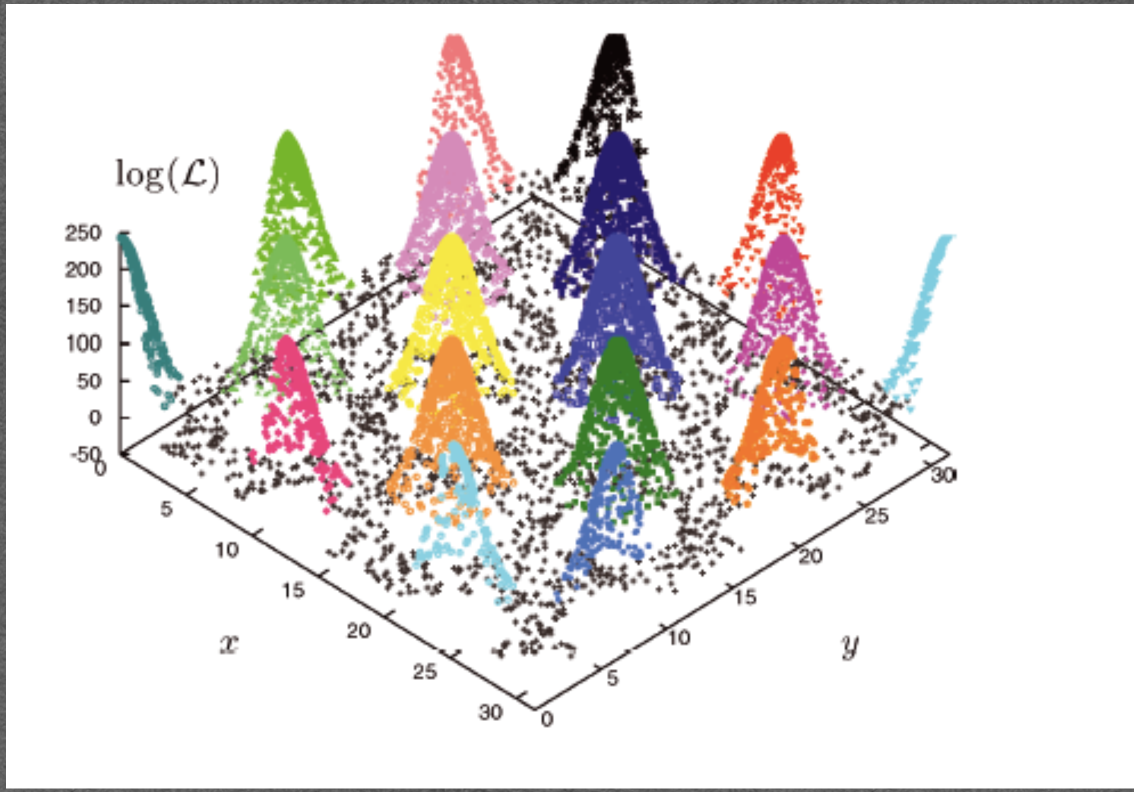
Advanced statistical techniques Sampling beyond Metropolis-Hastings

Some interesting features of alternative samplers

- Nested sampling (Skilling 2004) -> **MultiNest** http://ccpforge.cse.rl.ac.uk/gf/project/multinest/
- Ensemble sampling -> e.g. in **emcee** http://dan.iel.fm/emcee/current/
- Clustering algorithms -> **Kombine**
- Parallel tempering for initial distribution
- Hamiltonian Monte Carlo
- Population Monte Carlo PMC
- Adaptive MCMC -> e.g. in **PyMC** https://github.com/pymc-devs/pymc,
- Fast/Slow sampling -> e.g. in **CosmoMC**
- **Snake, Minuit, Maxlike, grid**
- Approximate Bayesian Computation, **ABC**

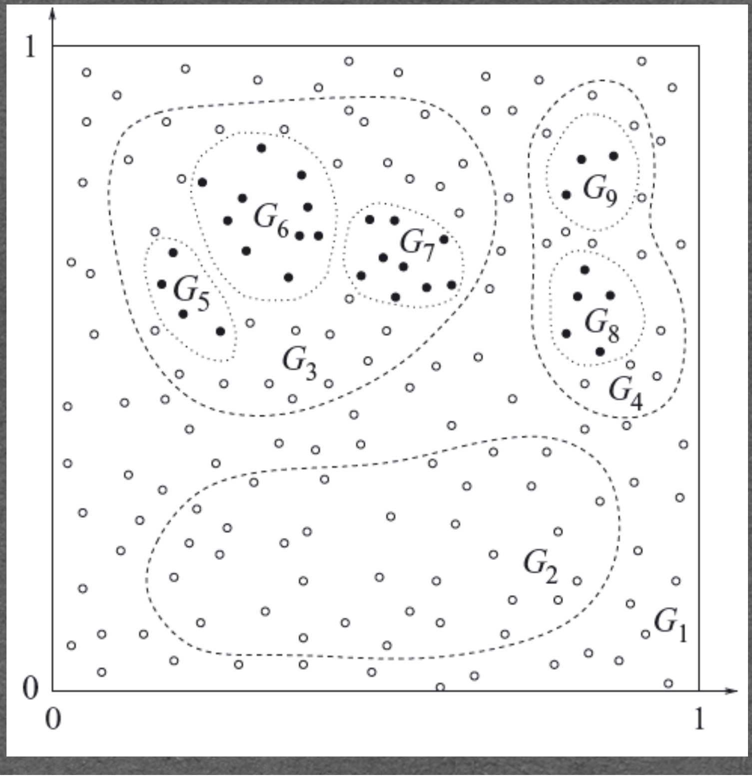
In CosmoSIS !

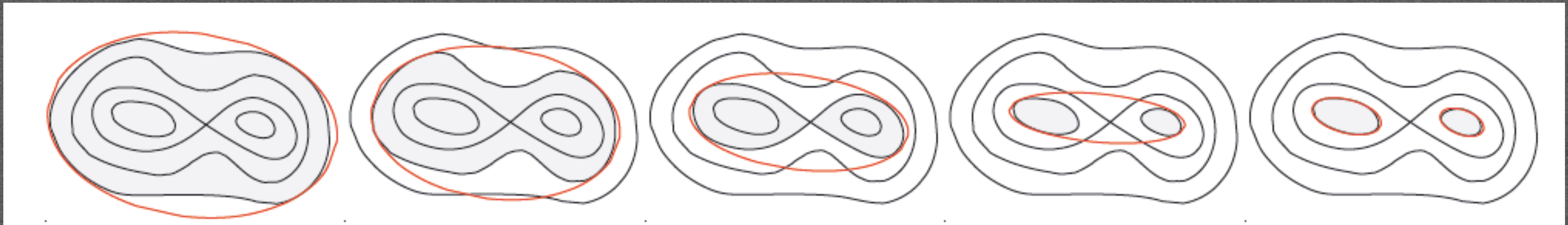




- Multiple peaks in the posterior identified and isolated

- iso-likelihood contours approximated by cov matrix of active points

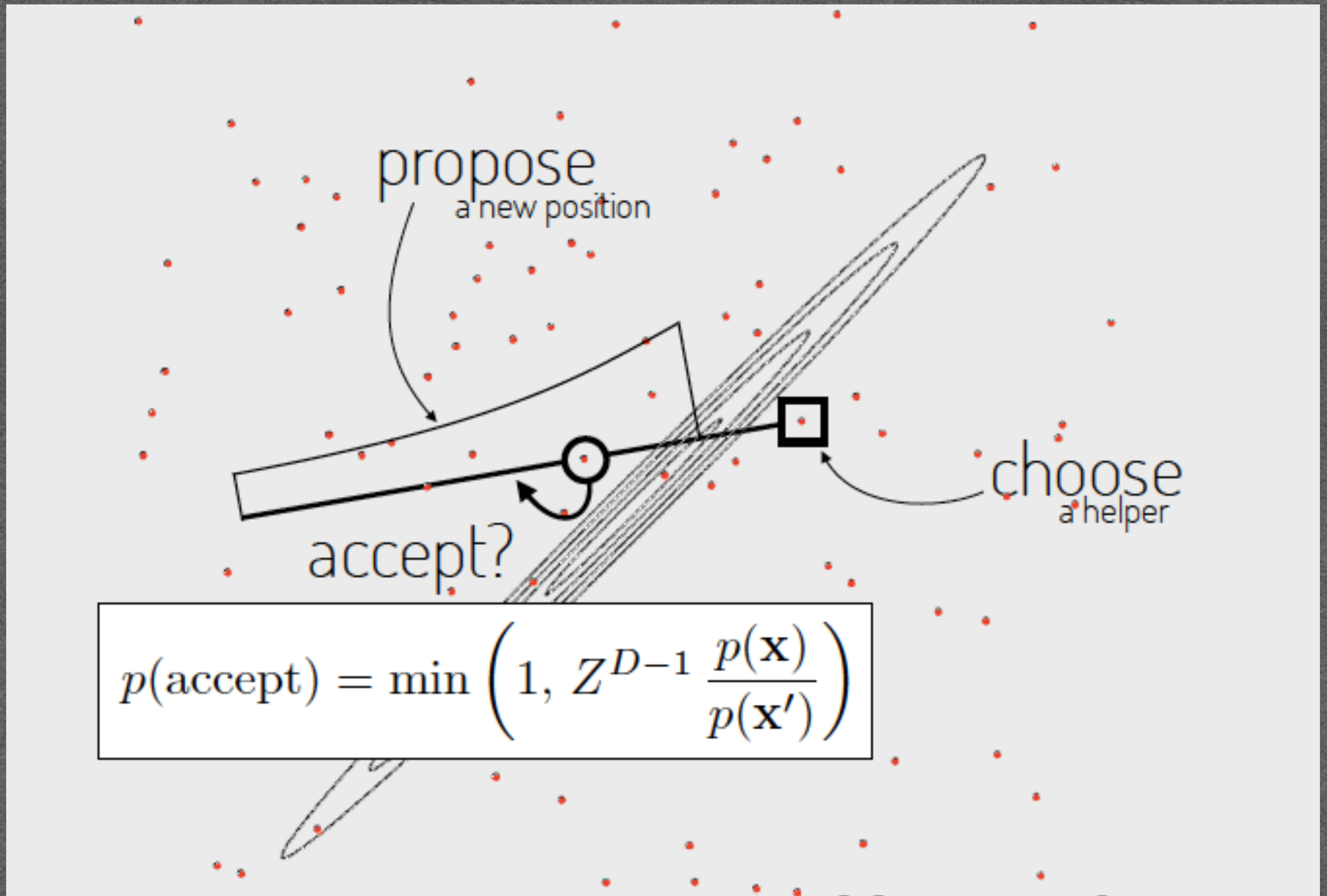




emcee
The MCMC Hammer

Foreman-Mackey et al 2014

- Ensemble sampler, uses “walkers” to probe parameter space
- Affine invariant



Slide from Foreman-Mackey’s talk at Vanderbilt

Dan Foreman-Mackey

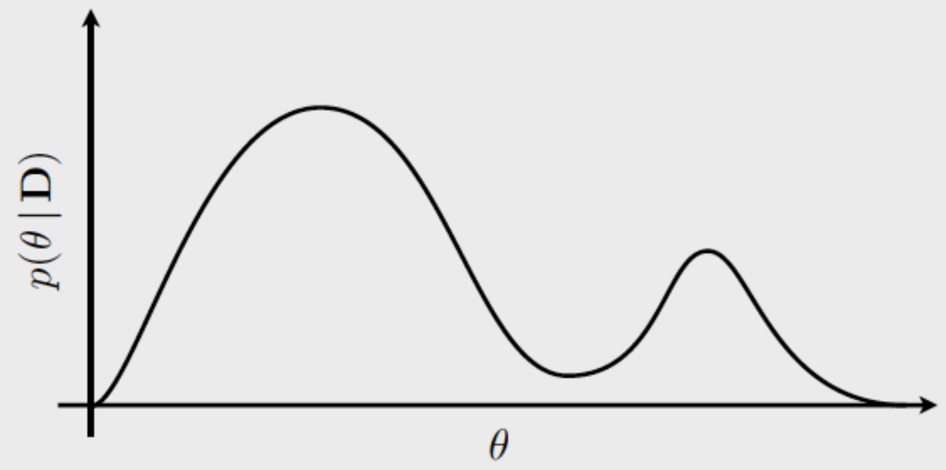
dan.iel.fm

Remember:

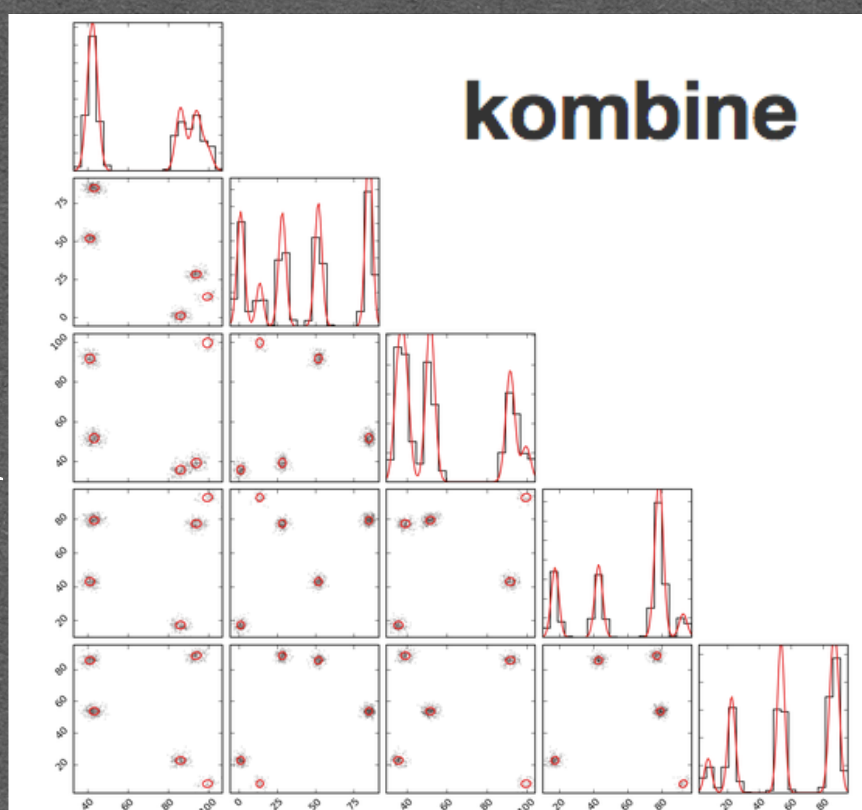
emcee isn't always The Right Choice™

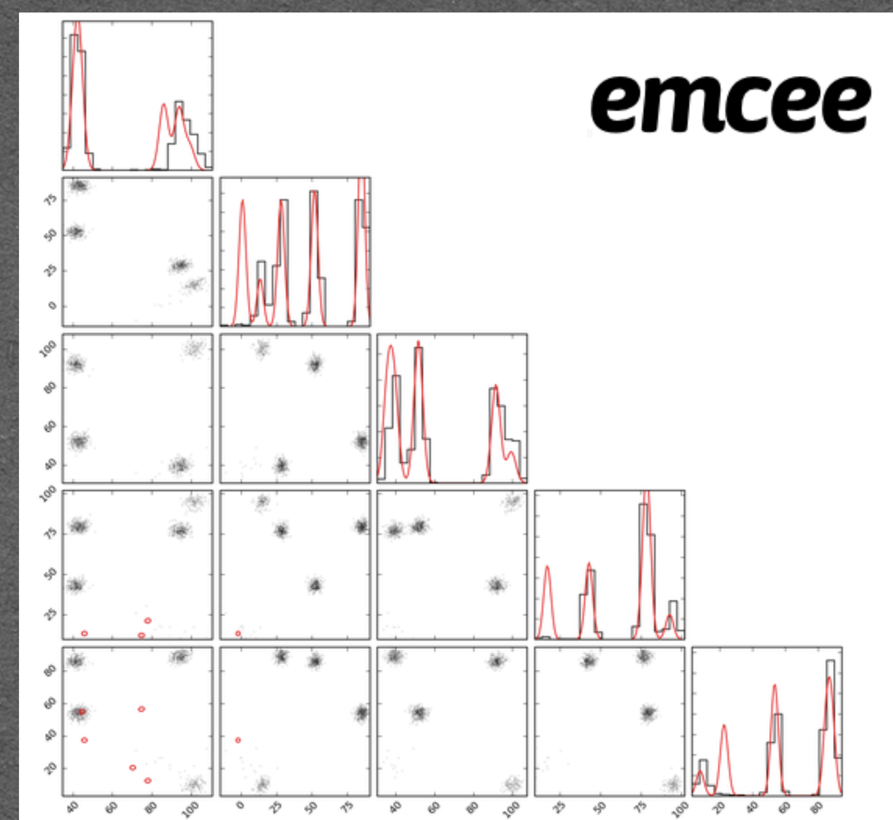
Slides from Foreman-Mackey's talk at Vanderbilt

what about multimodal densities?



5D 5mode MultiGaussian example comparison





Approximate Bayesian Computation

- Likelihood free parameter estimation
- works by simulating in a forward process

Approximate Bayesian Computation

Don't assume anything for Likelihood $P(\theta|\mathbf{x_i}, \mathbf{z_i}) \propto \cancel{\mathcal{L}(\mathbf{x_i}}|\theta, \mathbf{z_i})\pi(\theta, \mathbf{z_i})$

-Need to be able to simulate data at every point in parameter space

- Analytic formula for all systematic uncertainties is non-trivial
Incorporate into the simulation all of the statistical & systematic uncertainties using models and priors

– Draw θ^* from $\pi(\theta)$

– Simulate $\mu_{sim}(\theta^*)$

– Accept θ^* if $\rho(\mu_{sim} - \mu_{obs}) < \epsilon$

distance metric

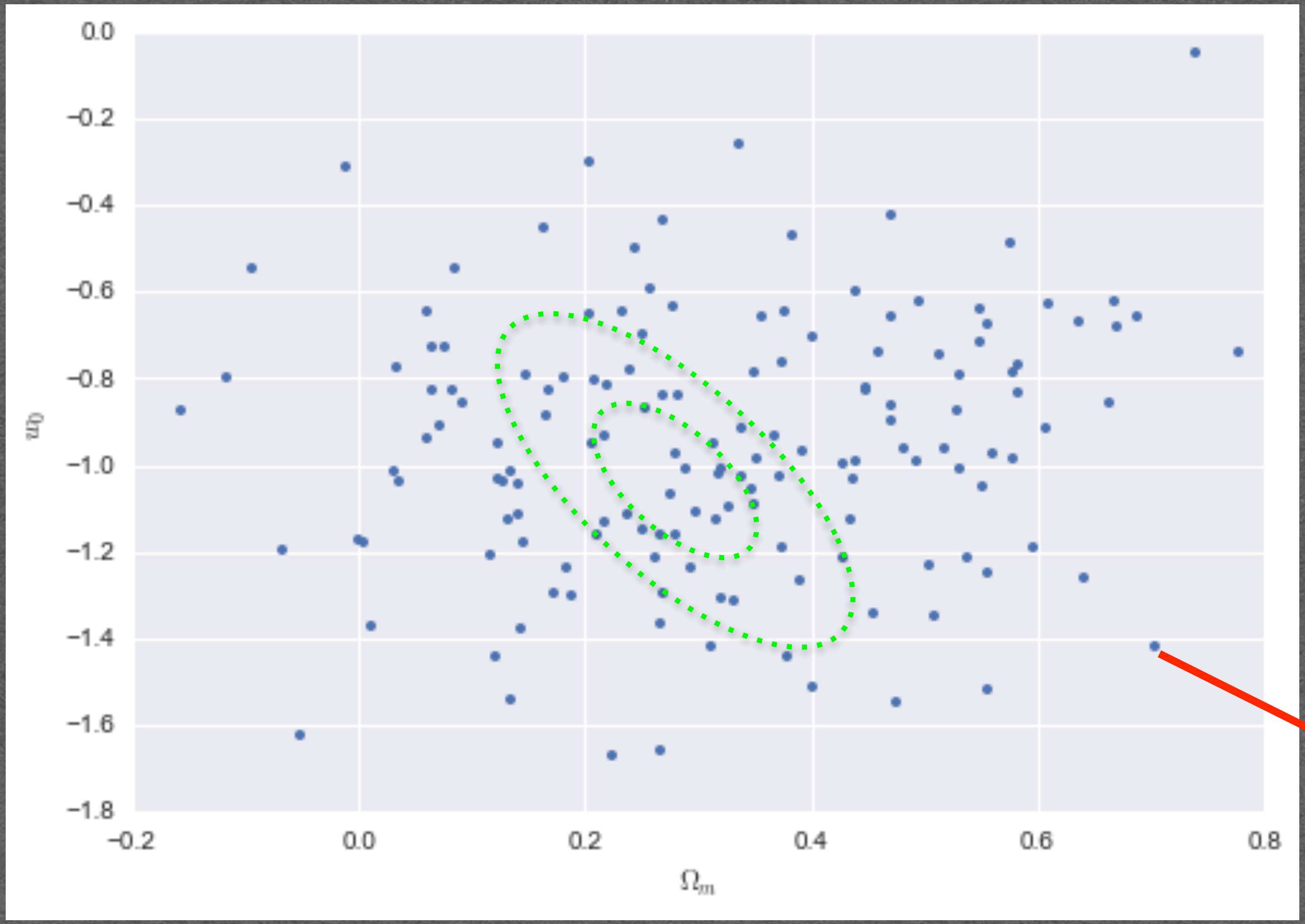
threshold

Finding a distance metric which is sensitive to cosmology is crucial

Sequential Monte Carlo ABC

T iteration levels with threshold levels

$$\epsilon = \{\epsilon_0, \epsilon_1, \cdots | \epsilon_{i+1} < \epsilon_i\}$$



t=0: initialize N points in parameter space from prior

At every point run simulation.

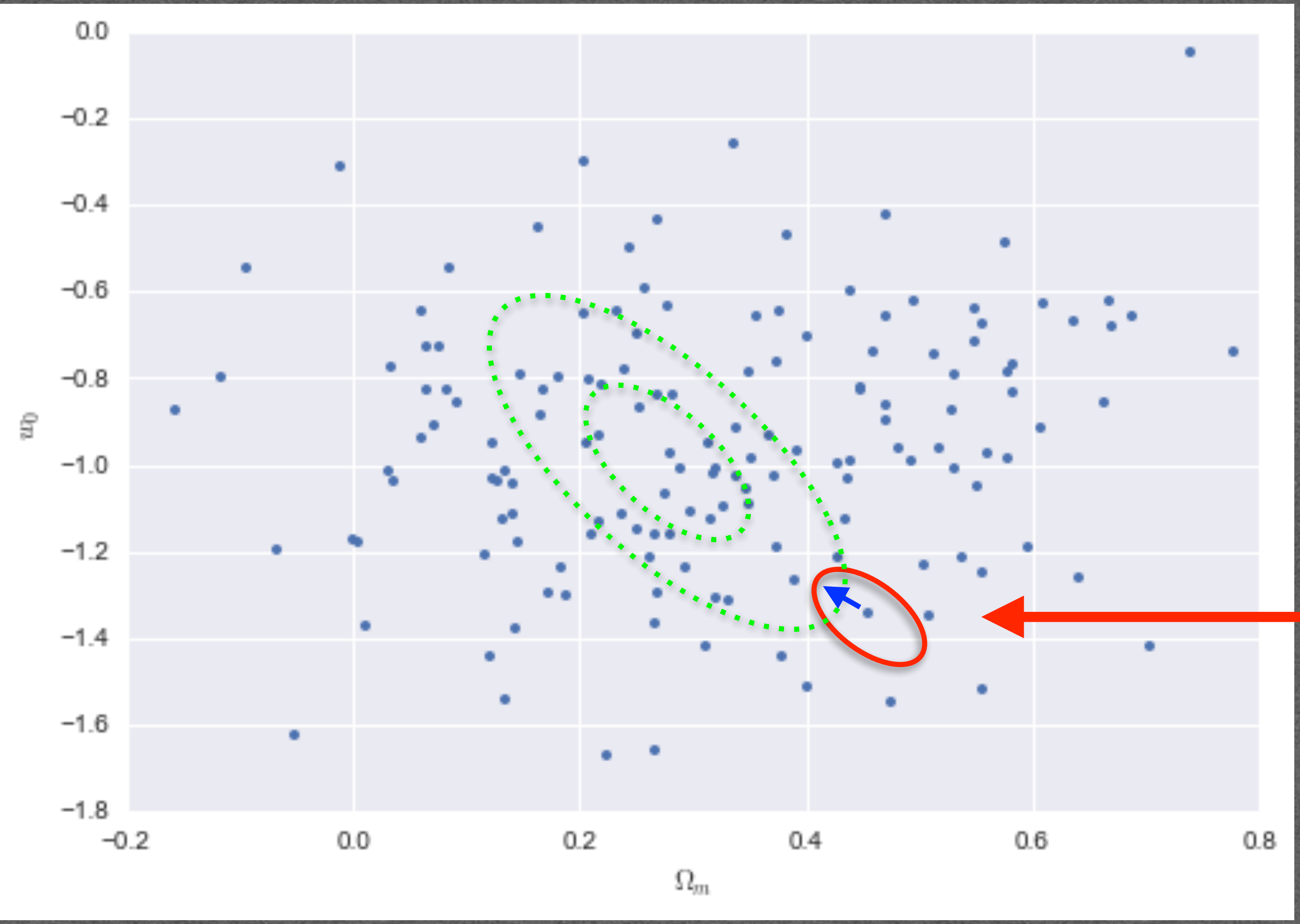
points are only accepted if

$$\rho < \epsilon_0$$

Sequential Monte Carlo ABC

T iteration levels with threshold levels

$$\epsilon = \{\epsilon_0, \epsilon_1, \cdots | \epsilon_{i+1} < \epsilon_i\}$$



t=1: perturb a random particle, using covariance amongst other particles

run simulation at new proposal point

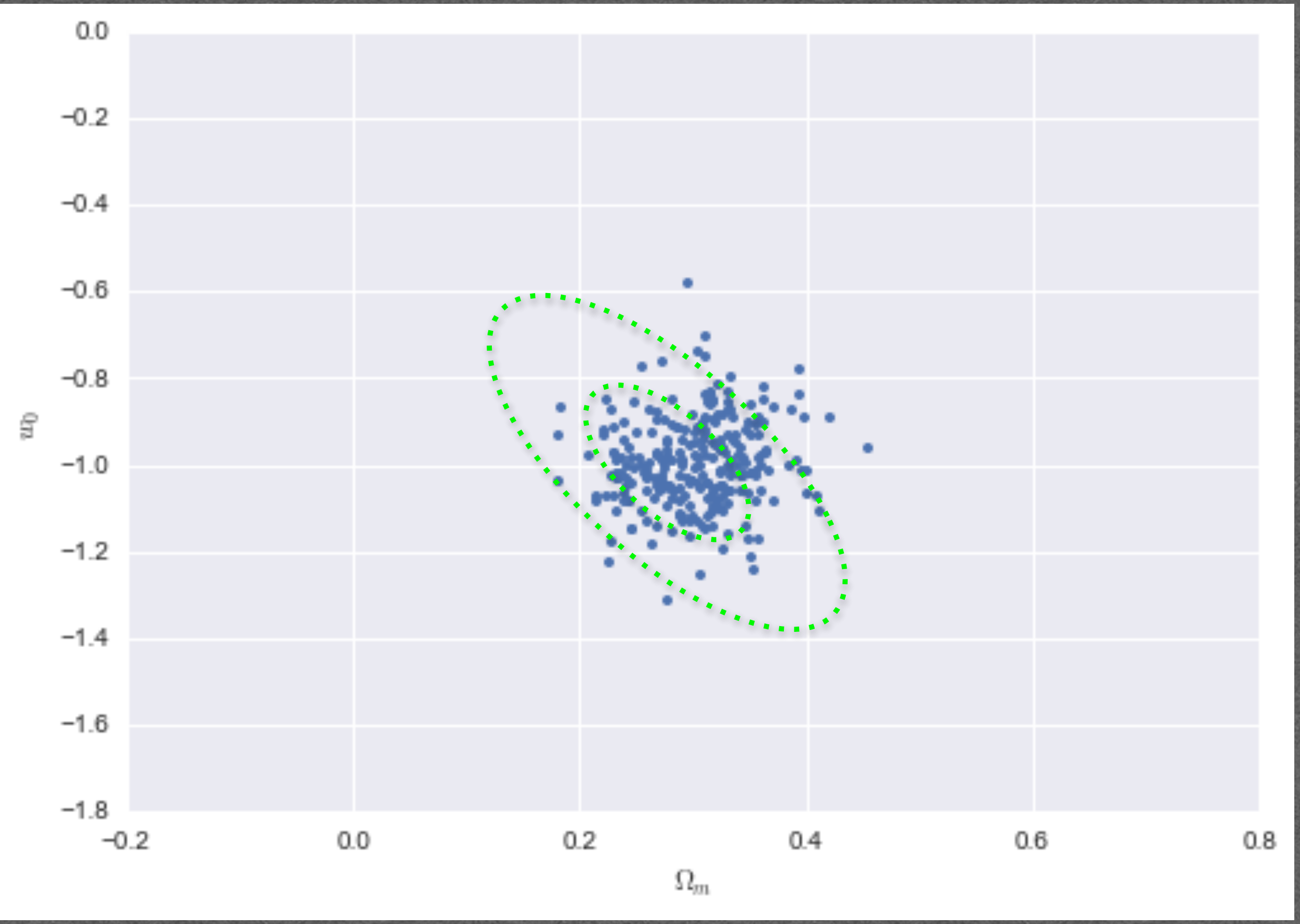
Accept if

$$\rho < \epsilon_1$$

Sequential Monte Carlo ABC

T iteration levels with threshold levels

$$\epsilon = \{\epsilon_0, \epsilon_1, \cdots | \epsilon_{i+1} < \epsilon_i\}$$



t=T: particles trace "true" contours

MCMC and ABC are very different methods…

ABC pros…

- ABC - free from any assumption for likelihood important for complicated physical processes (N.B. for multiprobe analysis)
- For MCMC need accurate covariance C_i,j which includes all systematics. Unbiased precision matrix. For ABC you don't need this.

ABC cons…

- ABC can be slow and inefficient compared to MCMC - > if you are *certain* of the Likelihood don't use ABC !
- Need fast accurate simulations for ABC - not always possible

ABC has similar pitfalls/arbitrary user choices to MCMC

- ABC metric is user defined — this *seems* hand wavy
  -in MCMC you choose a proposal distribution or some update criteria for walkers - can also slow convergence/bias results if you use incorrect choice

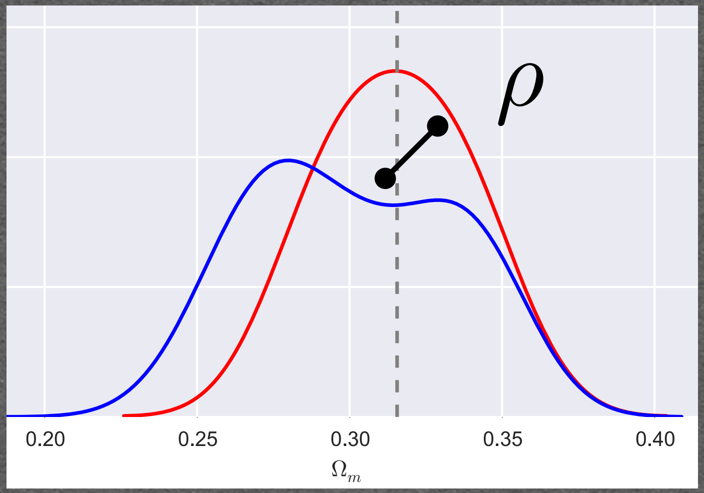
- ABC Summary statistics of data — this *seems* like a loss of information
  -techinically these should be "sufficient" statistics
  Think of Gaussian data reduced to two point function,
  no information loss

- When has my run converged? Both methods have no clear answer….

- ABC Metric is user defined
  using chi^2 is *not* the same as MCMC analysis where logLike is chi^2 !

ABC

MCMC



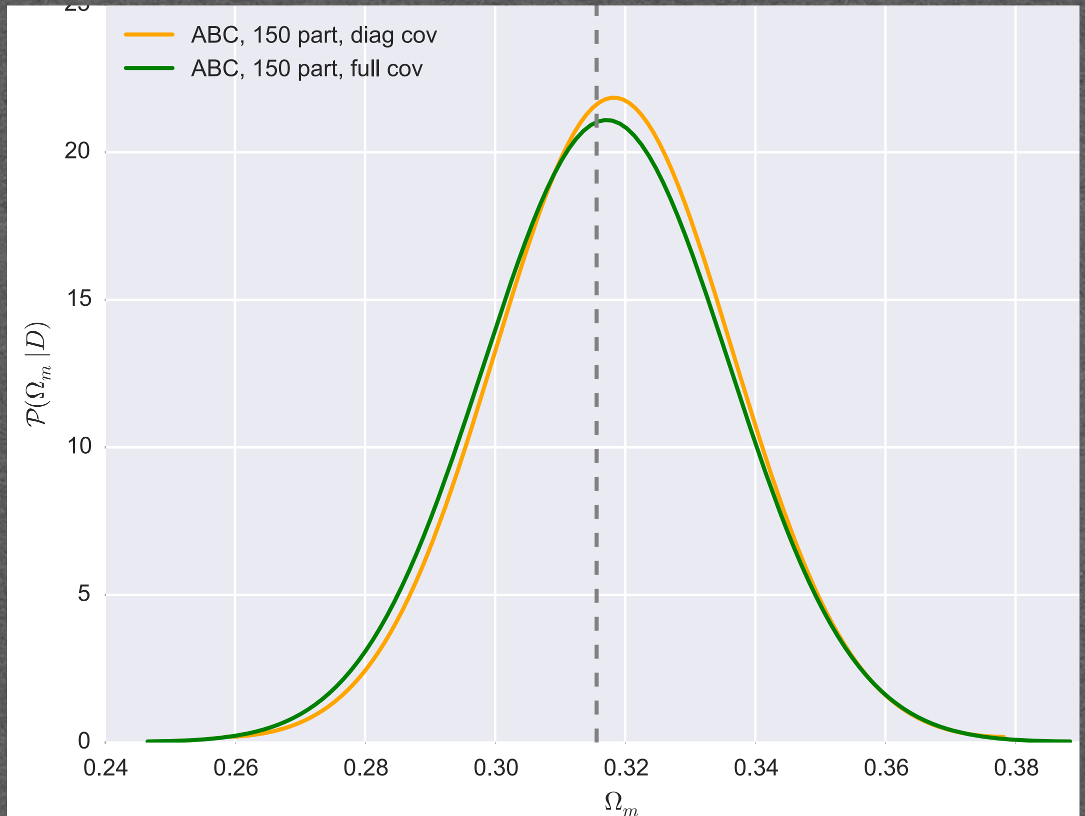
metric is distance between Proposed Posterior and the True posterior

e.g. when a Gaussian Likelihood is assumed

$$\text{data} \sim \mathcal{N}(\mu^{\mathbf{model}}, \mathbf{C_{i,j}})$$

$$\rho = (\mathcal{S}_{data} - \mathcal{S}_{sim})\mathcal{C}^{-1}(\mathcal{S}_{data} - \mathcal{S}_{sim})$$

this doesn't have to be the full Covariance matrix!



ABC is new tool to play around with
but important fact is that

any Cosmology results should not depend
on the sampler or internal sampler choices

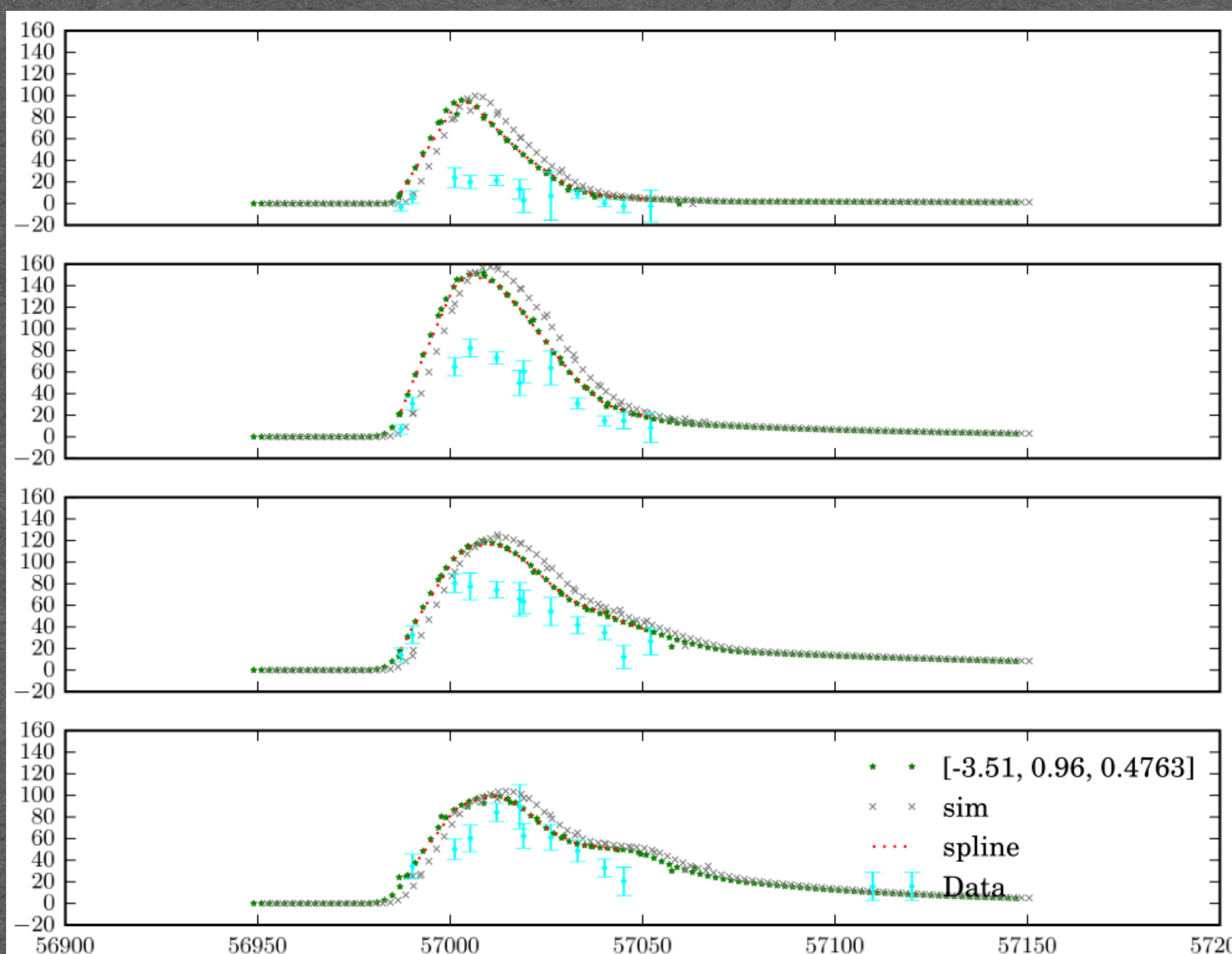
Using multiple methods is good check for robust results !

ABC for Type Ia SN for The Dark Energy Survey

EJ, Rick Kessler & Dan Sconic in prep

DARK ENERGY SURVEY

New idea: simulate light curves and use these in the ABC metric directly, no assumption of likelihoods



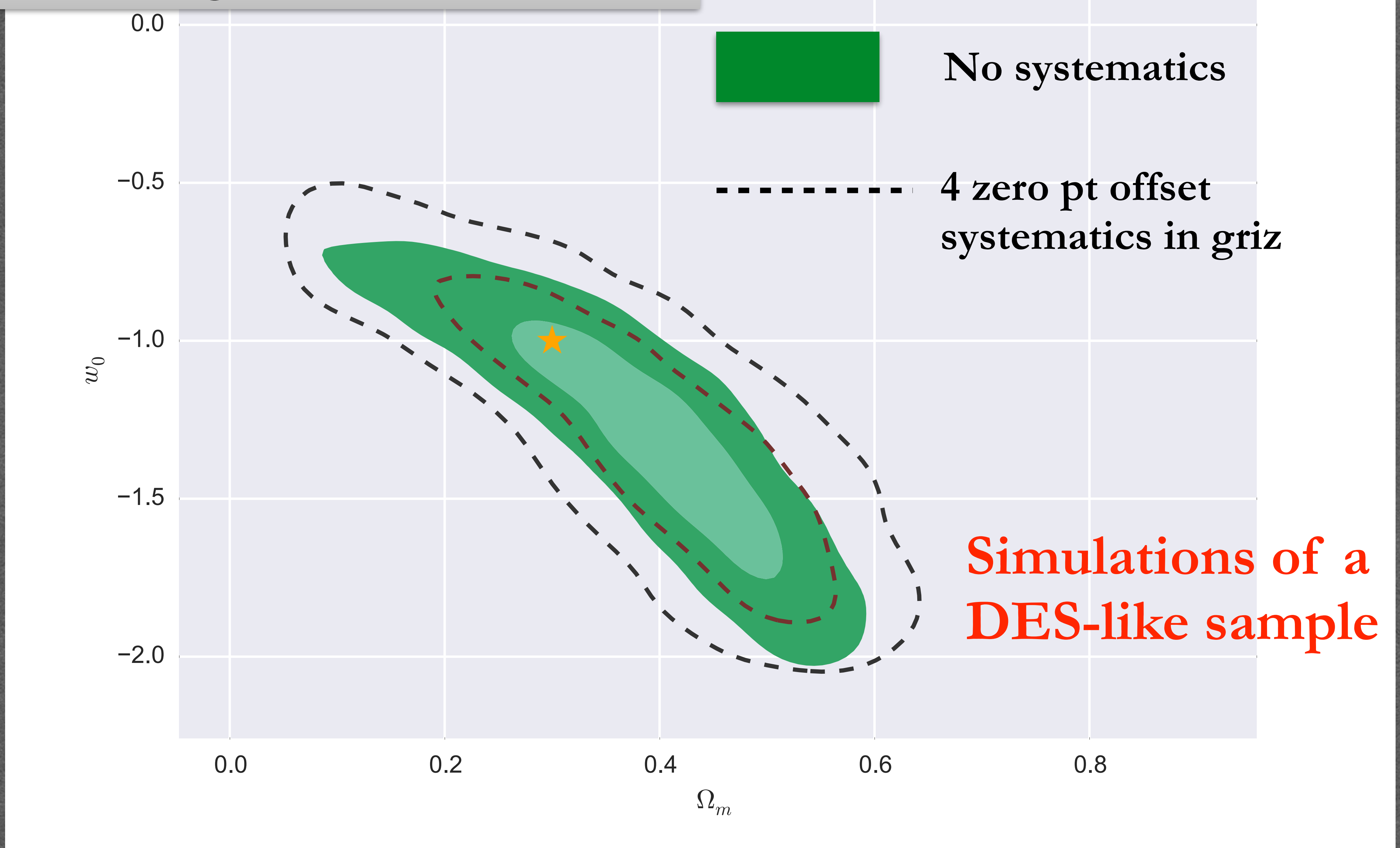
Match simulation and data light curves by running SNANA at
every point in parameter space during sampling, systematics naturally included

ABC with Type Ia SN for DES

No sys error contours match standard MCMC approach when M0 fixed

Zero point offsets naturally included in the simulations - no way to fully do this in standard MCMC

NB: sys effects amplified from expected DES 0.01 mag for effect !



EJ, Rick Kessler & Dan Sconic in prep

Summary

Parameter estimation

- modular approach is key: **CosmoSIS**
- facilitate collaboration, provenance tracking, consistency and reproducibility.
- advanced sampling methods needed to handle:
  correlations, large number of nuisance parameters.
- ABC method interesting alternative for including systematics when likelihood is complicated