Bayesian Hierarchical Models for SNIa Cosmology

Collaborators: Hik Shariff, Xiyun Jiao, David van Dyk (Imperial)
(astro) (stats) (stats)

ICIC
Imperial Centre
for Inference & Cosmology

Roberto Trotta - AACC 2015
16/07/15 - www.robertotrotta.com

Imperial College
London

Roberto Trotta
@R_Trotta

SNIa cosmology is broken

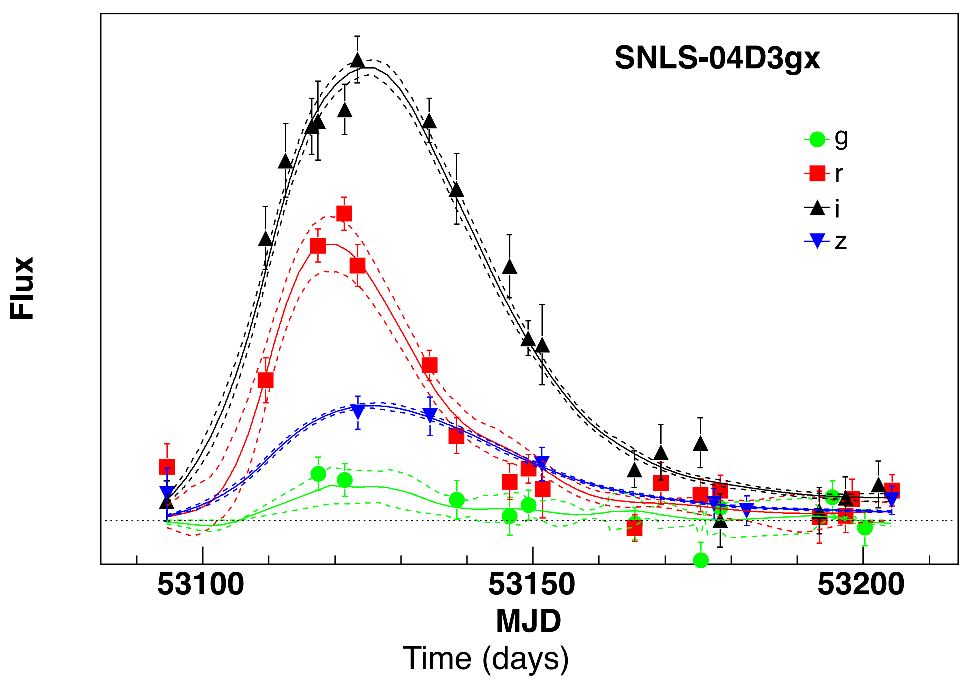
Roberto Trotta
@R_Trotta

SNIa cosmology is broken

@R_Trotta

Observations

SNLS

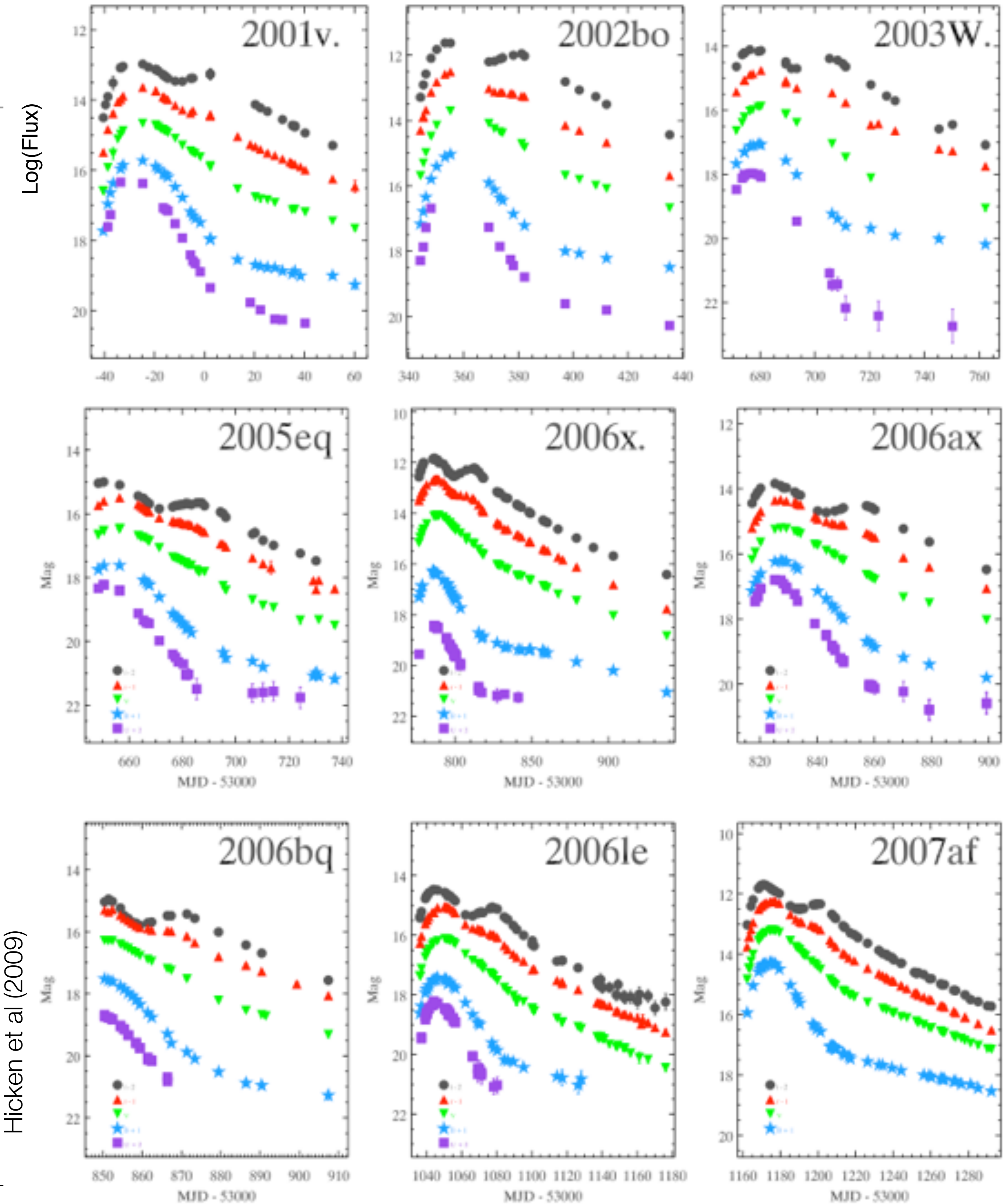


Guy et al (2007)

SALT2

$$\hat{x}_{1i}, \hat{c}_i, \hat{m}^*_{B,i} + \hat{C}_i$$

CfA3
185 multi-band optical nearby SNIa



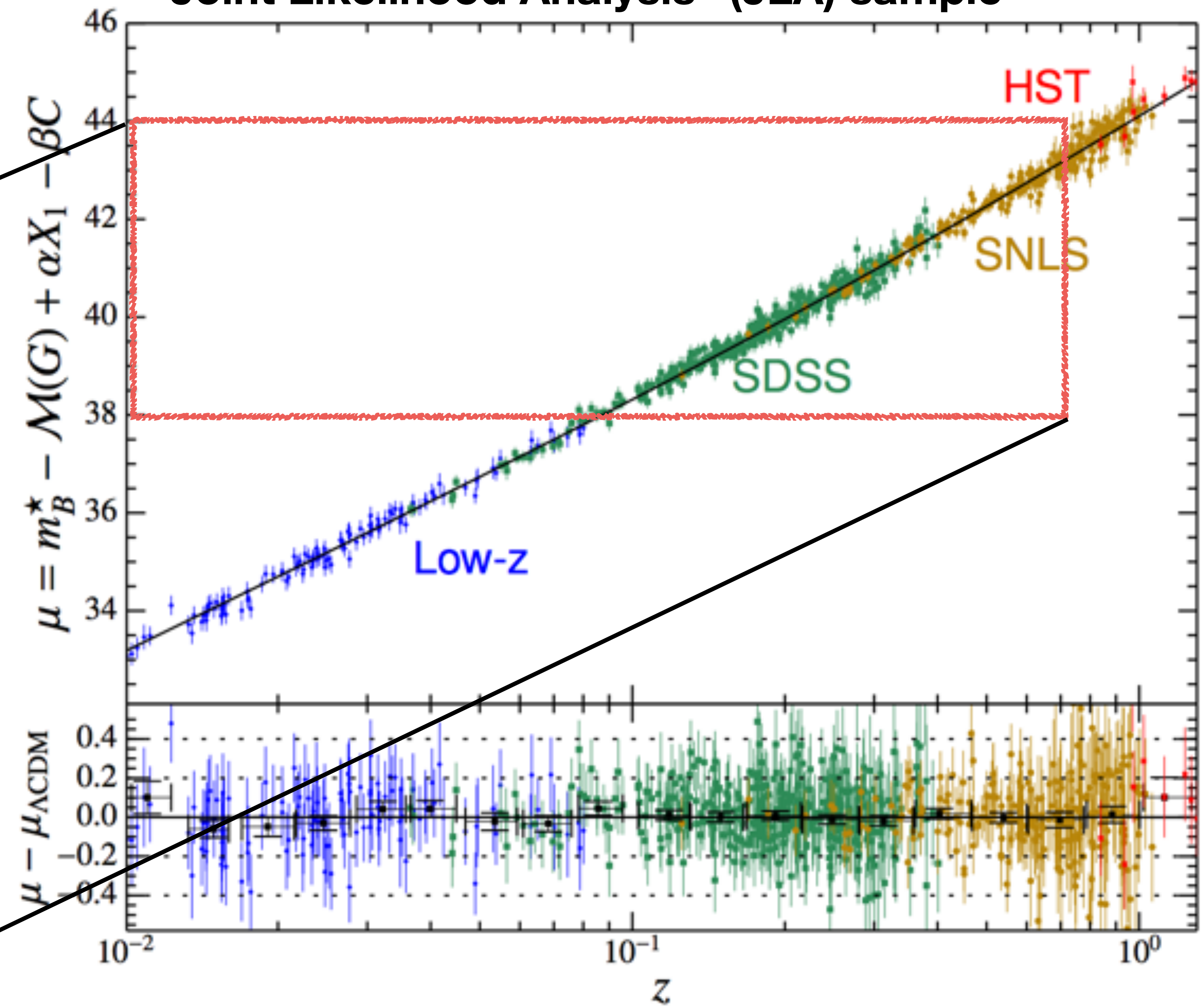
Hicken et al (2009)

ICIC

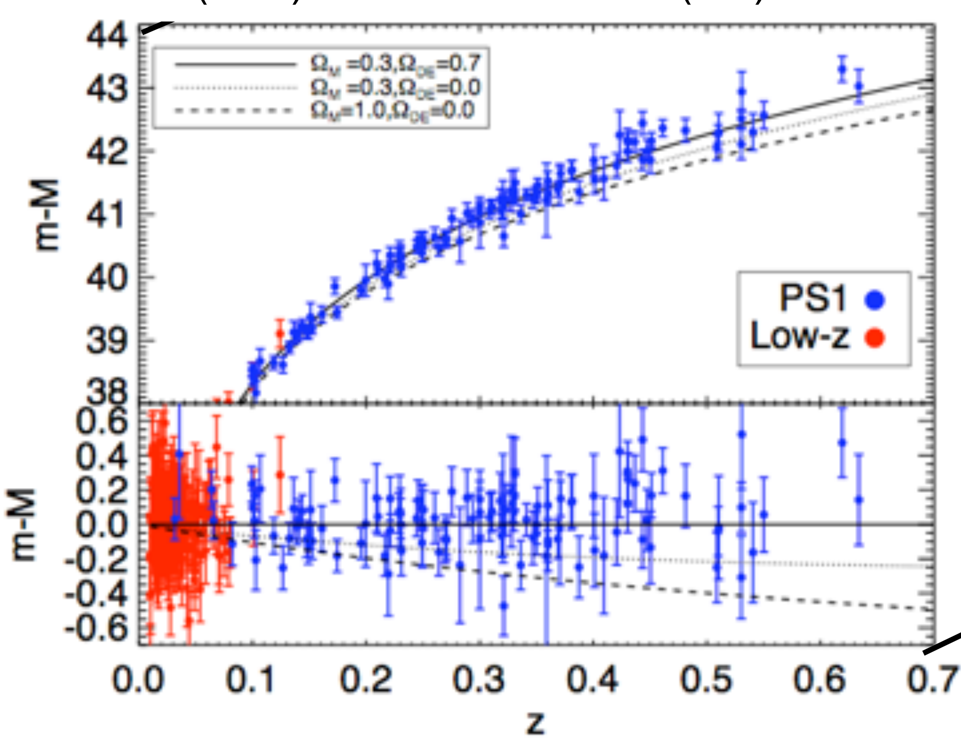
SNIa sample

Betoule+ 2014: 740 spectroscopically confirmed SNIa, out to z ~ 1.4, with joint re-analysis of LC fits

"Joint Likelihood Analysis" (JLA) sample



Rest+ 2013: 112 PS1 at high-z (blue) + 201 low-z SNIa (red)



The statistical problem

Apparent rest-frame B-band magnitude

Luminosity distance

Cosmological parameters

Quantities of interest
Ω_M, Ω_{DE}, w, w(z),
H_0 (degenerate with M),

$$\mu = m_B - M = 5\log_{10}\left(\frac{d_L(z,\mathcal{C})}{1\ \mathrm{Mpc}}\right) + 25$$

Distance modulus

Absolute magnitude

Unknown, but
IF ~ constant
unimportant
("standard candle")

Needs to be corrected via empirical correlations with other observables

M → M + linear corrections

Redshift

Measured via spectrum

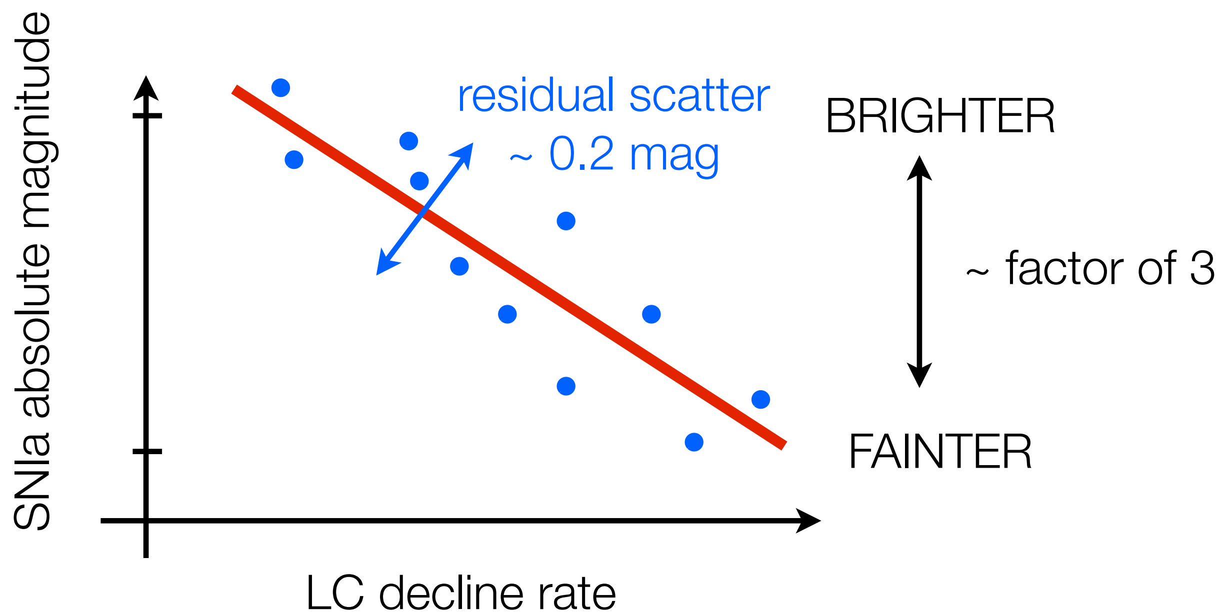
Goal: From the measured multi-band light curves and redshift, infer constraints on the cosmological parameters.

Bayesian approach: March, RT+11; March+14; Nielsen+15; Rubin+15; Jiao,RT+(in prep); Shariff, RT+(in prep)

Brightness-width relationship

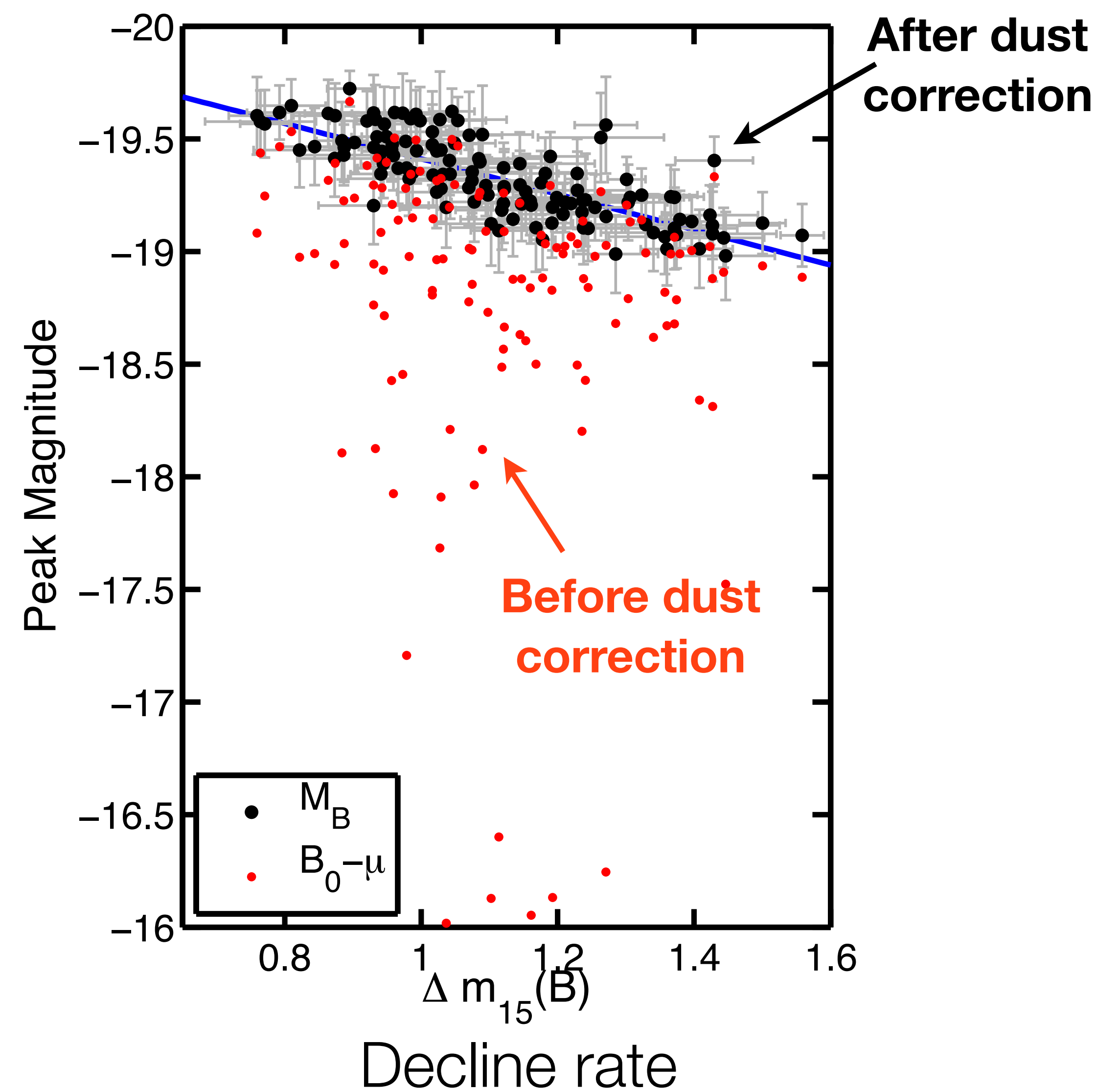
Imperial College London

Peak magnitude scatter can be reduced by exploiting phenomenological correlations with the shape (and colour) of the LC (Phillips93,Riess+99,...)



Brighter SNIa are slow decliners

Low-z calibration sample



Mandel et al (2011)

This can be understood is terms of the progenitor ^{56}Ni mass which powers the LC

SNIa cosmology error budget is already dominated by "systematics":

- X Flux calibrations
- X Selection effects
- X Outliers
- ✓ Redshift evolution of Phillips corrections
- ✓ Non-linear corrections
- ✓ Multiple populations
- ✓ Dust extinction modeling
- ✓ Environmental properties

Explicit modeling of "systematics" transforms them into manageable statistical errors

Standard Chi^2 fits of SALT2 output

Standard analysis minimizes an arbitrarily defined Gaussian likelihood (typically, C minimized with α, β fixed, then α, β minimized with C fixed):

observed values (LC fits)

parameters

$$-2\log\mathcal{L} = \chi^2 = \sum_i \frac{(\mu(z_i,\mathcal{C}) - [\hat{m}_{B,i} - M + \alpha\hat{x}_{1,i} - \beta\hat{c}_i])^2}{\sigma^2_{\rm int} + \sigma^2_{\rm fit}}$$

$$\sigma^2_{\rm fit} = \sigma^2_{m_B} + \alpha^2\sigma^2_{x_1} + \beta^2\sigma^2_c + \text{correlations}$$

$\sigma^2_{\rm int}$ represents the "intrinsic" (residual) scatter

determined by requiring Chi^2/dof ~ 1

Problems of the standard analysis

$$-2\log\mathcal{L} = \chi^2 = \sum_i \frac{\left(\mu(z_i,\mathcal{C}) - [\hat{m}_{B,i} - M + \alpha\hat{x}_{1,i} - \beta\hat{c}_i]\right)^2}{\sigma^2_{\rm int} + \sigma^2_{\rm fit}}$$

$$\sigma^2_{\rm fit} = \sigma^2_{m_B} + \alpha^2\sigma^2_{x_1} + \beta^2\sigma^2_c + \text{correlations}$$

- X Likelihood is Gaussian *in the data*!
- X Normalization
- X Unknown parameters appear in the variance
- X Chi^2/dof = 1 enforced: no model checking
- X Incorrect likelihood prevents use of powerful MCMC/ evidence calculations

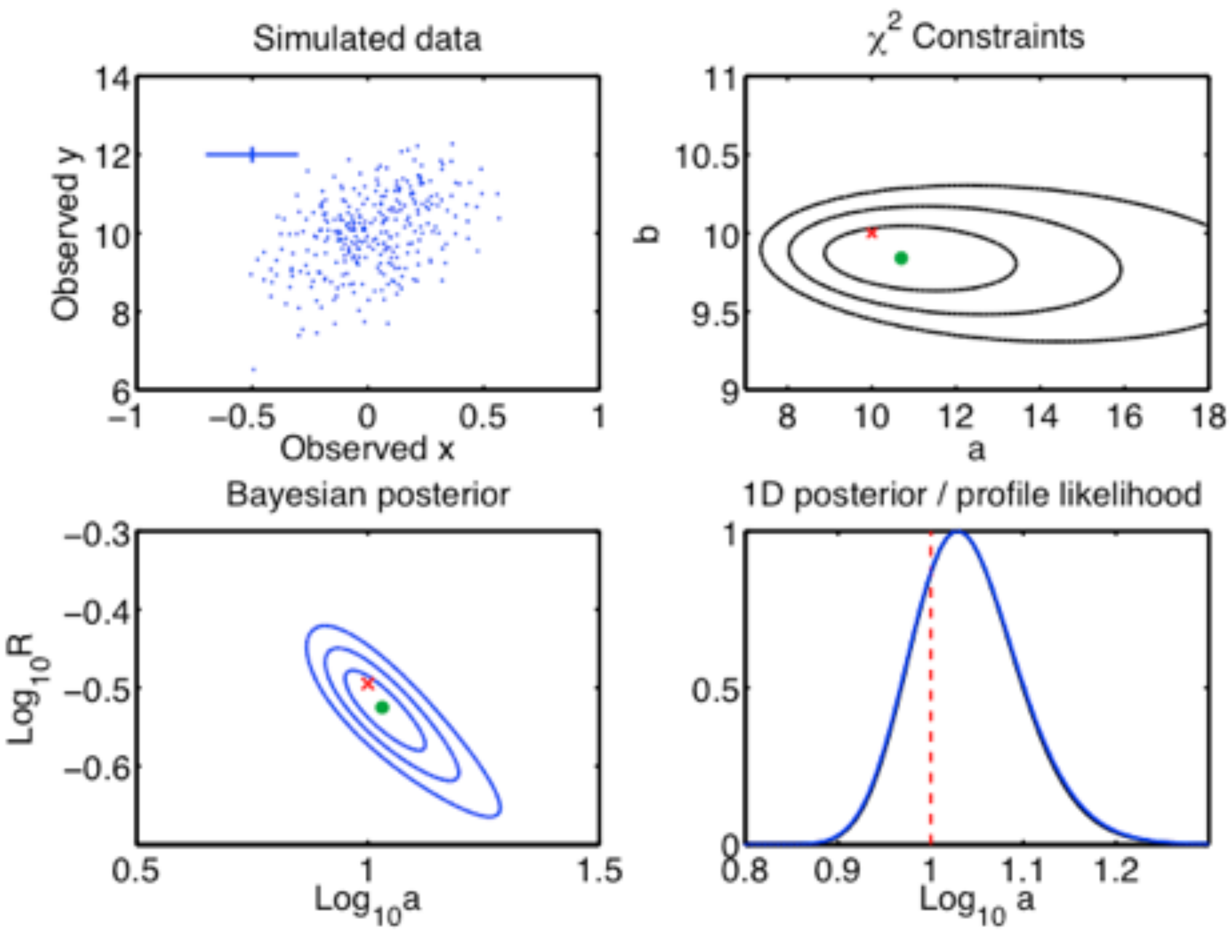
March, RT+11; Rubin+15

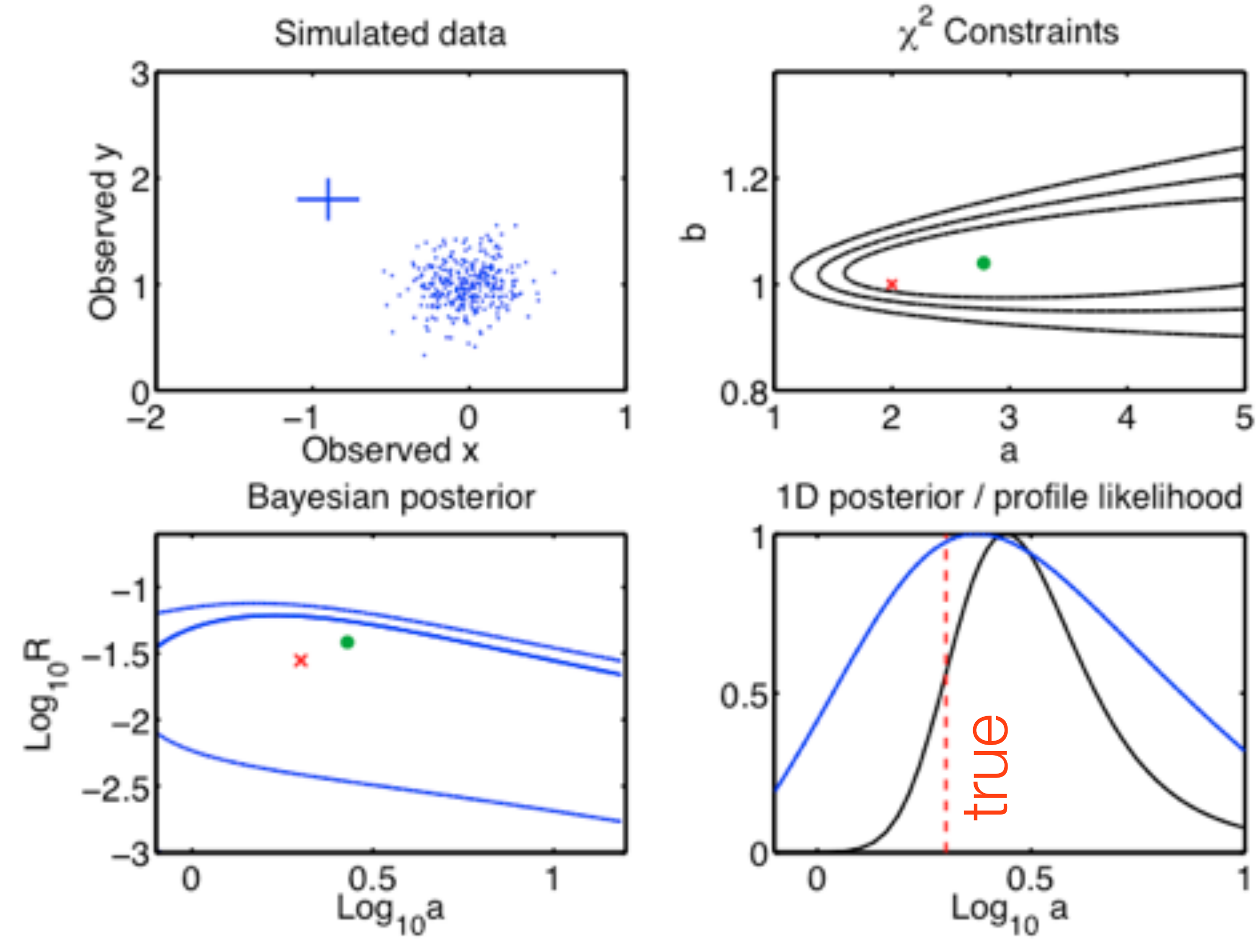
The key parameter is noise/population variance $\sigma_x\sigma_y/R_x$

$\sigma_x\sigma_y/R_x$ small

$$y_i = b + ax_i$$

$\sigma_x\sigma_y/R_x$ large





Bayesian marginal posterior identical to profile likelihood

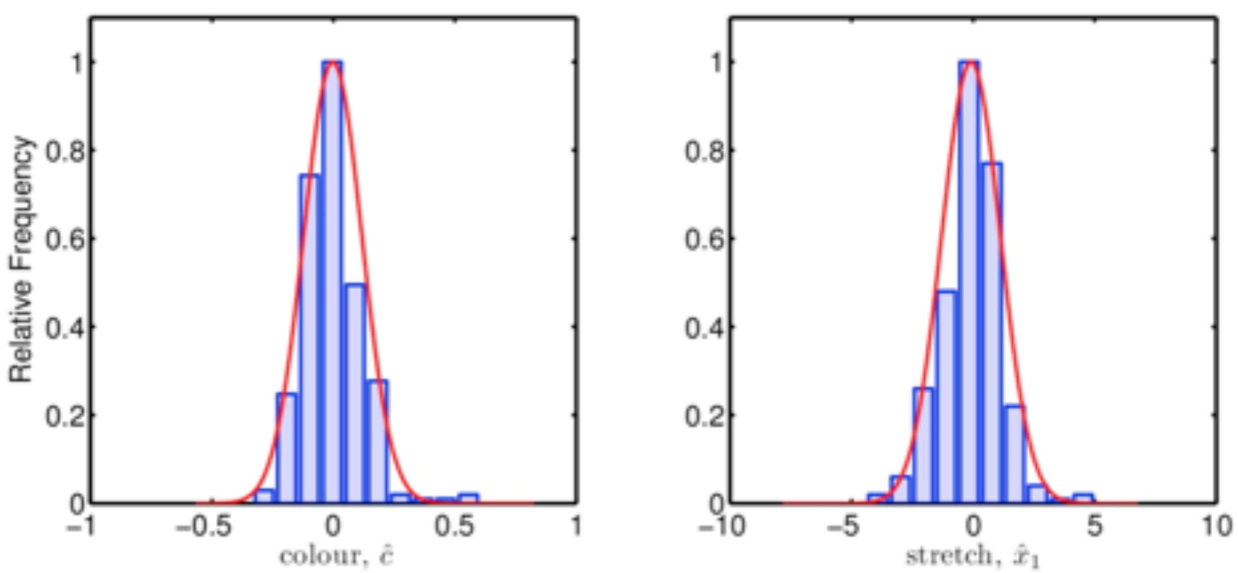
Bayesian marginal posterior broader but less biased than profile likelihood

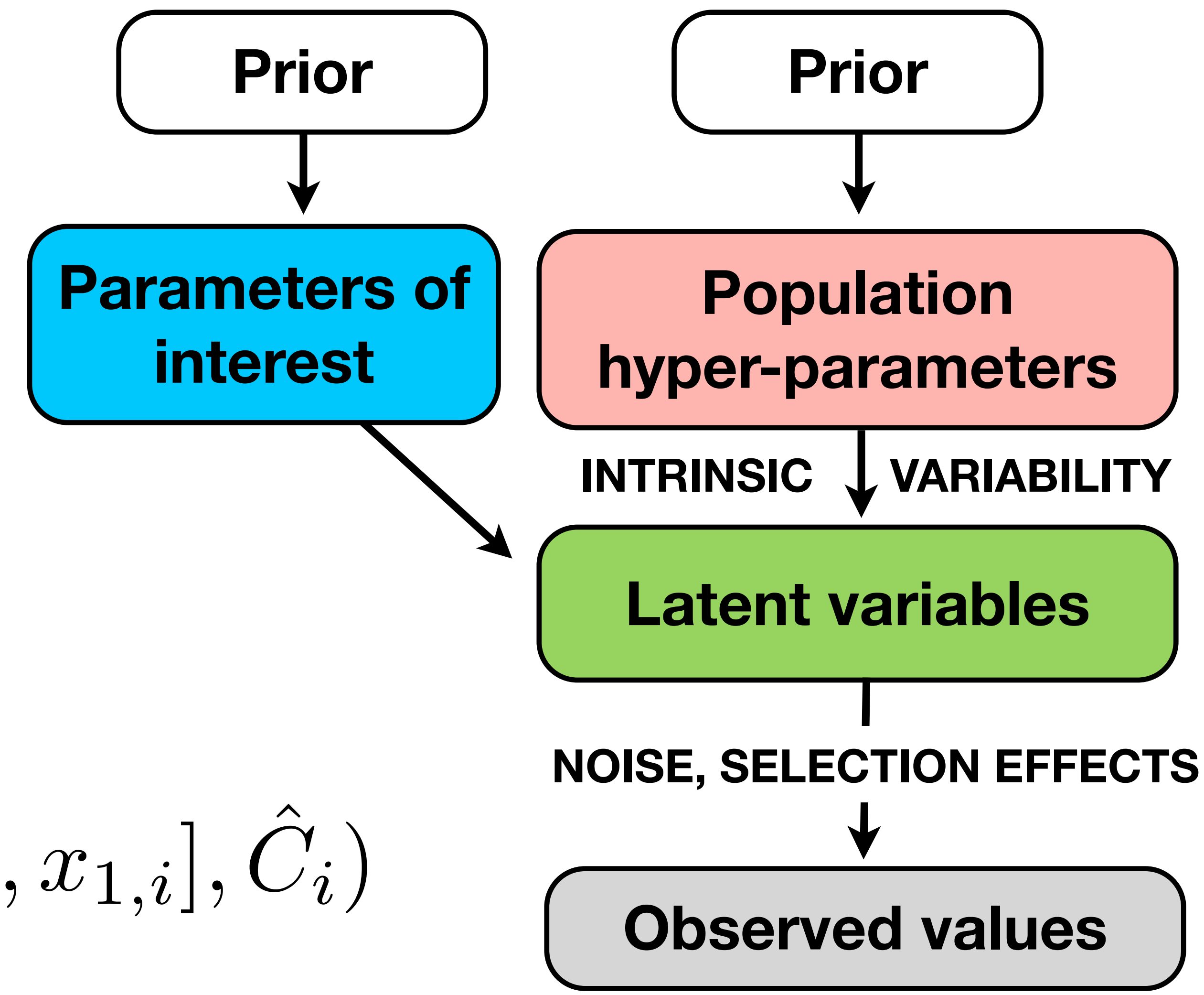
Gull89; Kelly10; March,RT+11

Bayesian hierarchical model

For each SNIa, this relation holds **exactly** between **latent** (unobserved) variables:

$$x_{1,i} - \beta c_i$$





$$c_i \sim \mathcal{N}(c_\star, R_c)$$

$$x_{1,i} \sim \mathcal{N}(x_\star, R_x)$$

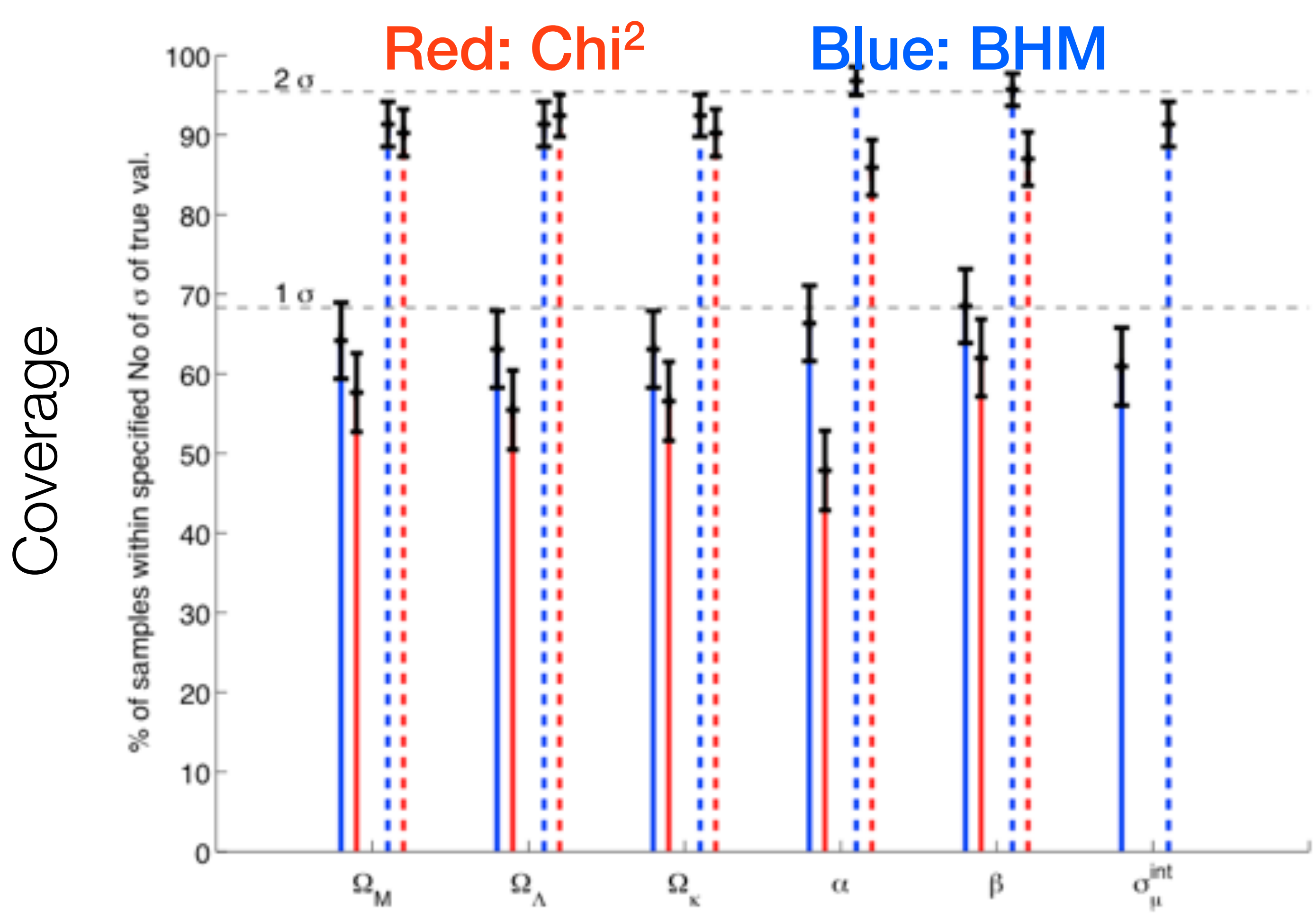
$$[\hat{m}_{B,i}, \hat{c}_i, \hat{x}_{1,i}] \sim \mathcal{N}([m_{B_i}, c_i, x_{1,i}], \hat{C}_i)$$

Improvements from Bayesian approach

Imperial College London

100 mock Union datasets to compare the performance of Chi^2 vs Bayesian (BHM):

- Coverage
- Bias (expected distance from the truth)
- Mean Square Error (MSE = bias^2 + Var)



Results:

✓**Coverage:** BHM improves coverage

✓**Bias:** BHM reduces bias by a factor ~ 2-3 for most parameters

✓**MSE:** BHM reduces MSE by a factor 1.5-3.0 for all parameters

March,RT+11

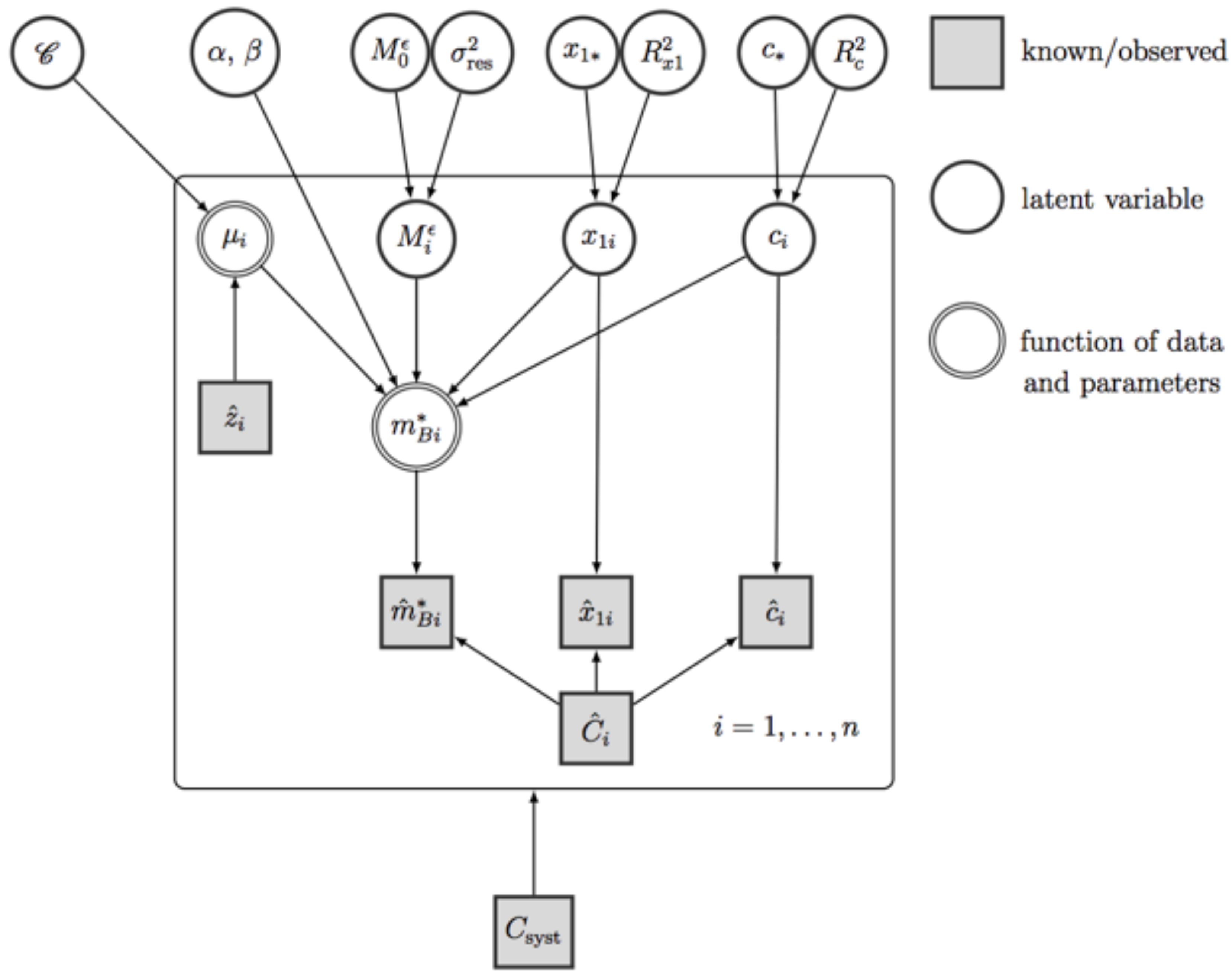
ICIC

Posterior sampling

Imperial College London

The full hierarchical model (including ~2200 latent variables) is sampled via Metropolis-Hastings with a Partially Collapsed Gibbs sampler (vanDyk&Park08) and Ancillarity-Sufficiency Interweaving Strategy (Yu&Meng11) to improve convergence

(Jiao, RT,+ submitted)

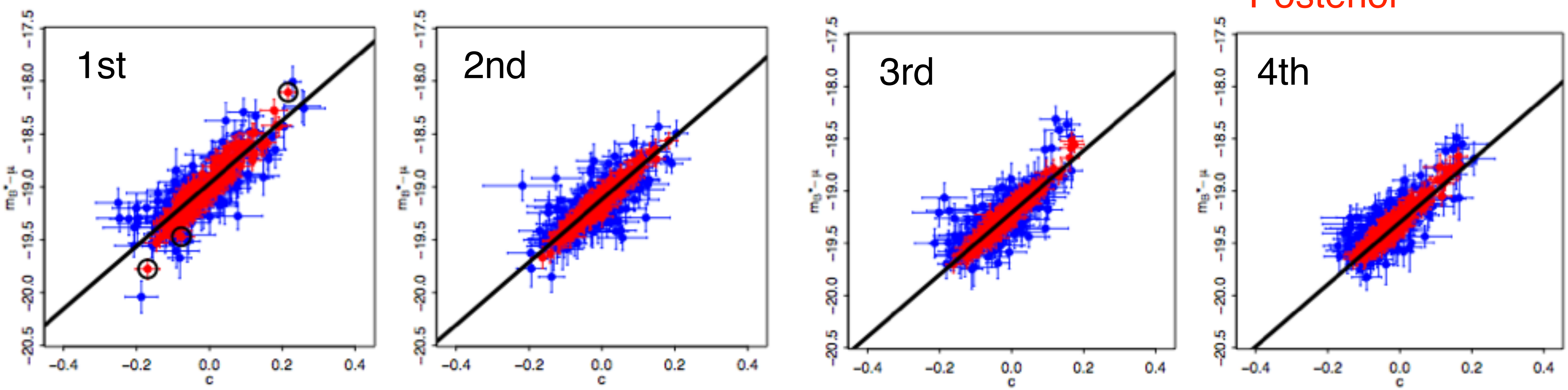


ICIC

Borrowing of strength

$\hat{x}_{1i}$ quartile:

Likelihood
Posterior



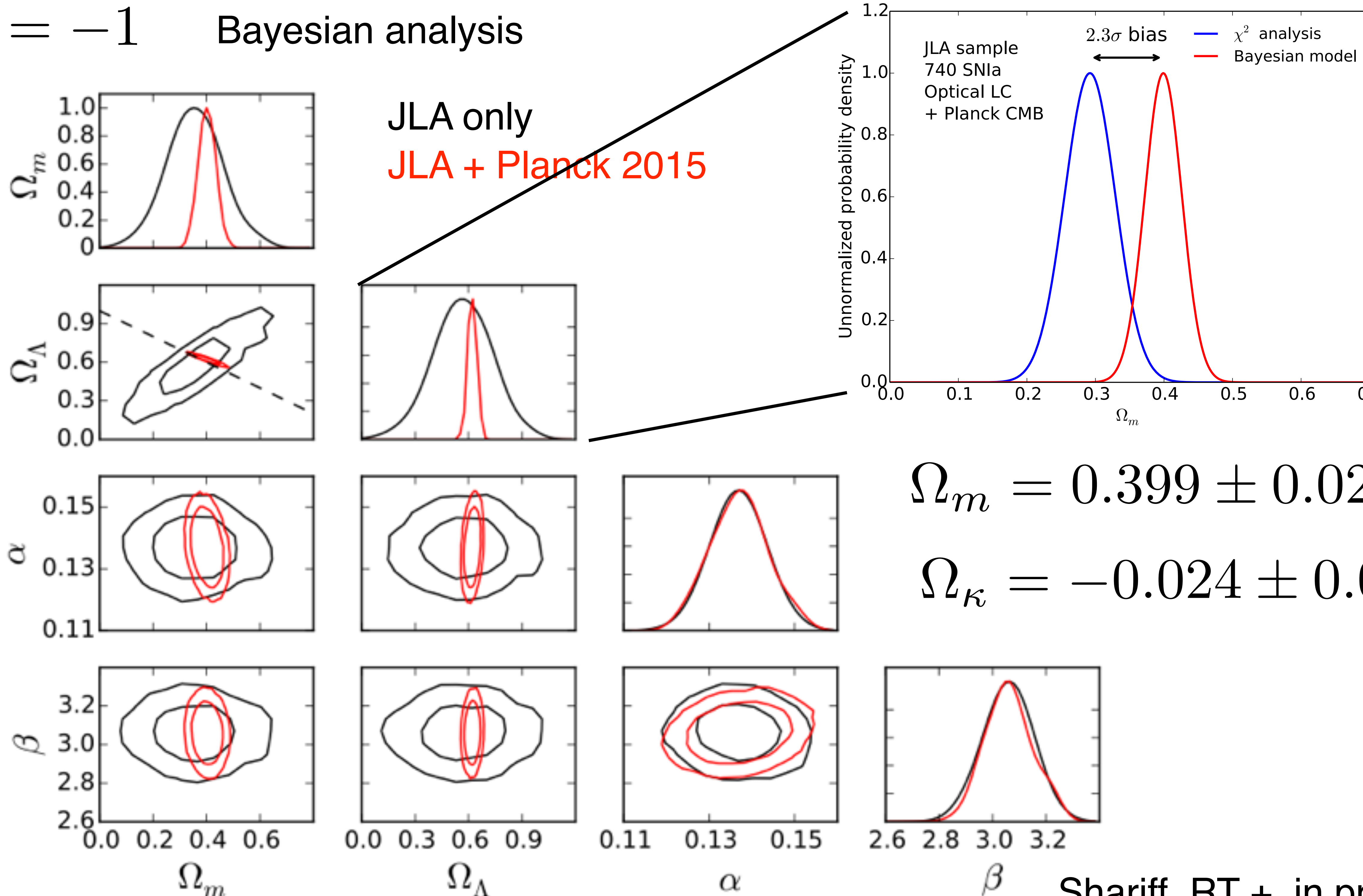
The posterior estimates (red) exhibit smaller residual scatter when compared to the likelihood (blue), around the regression line: "borrowing of strength" from the structure of the hierarchical model

Shariff, RT +, in prep

Cosmological parameters from JLA

Imperial College London

$w = -1$ Bayesian analysis



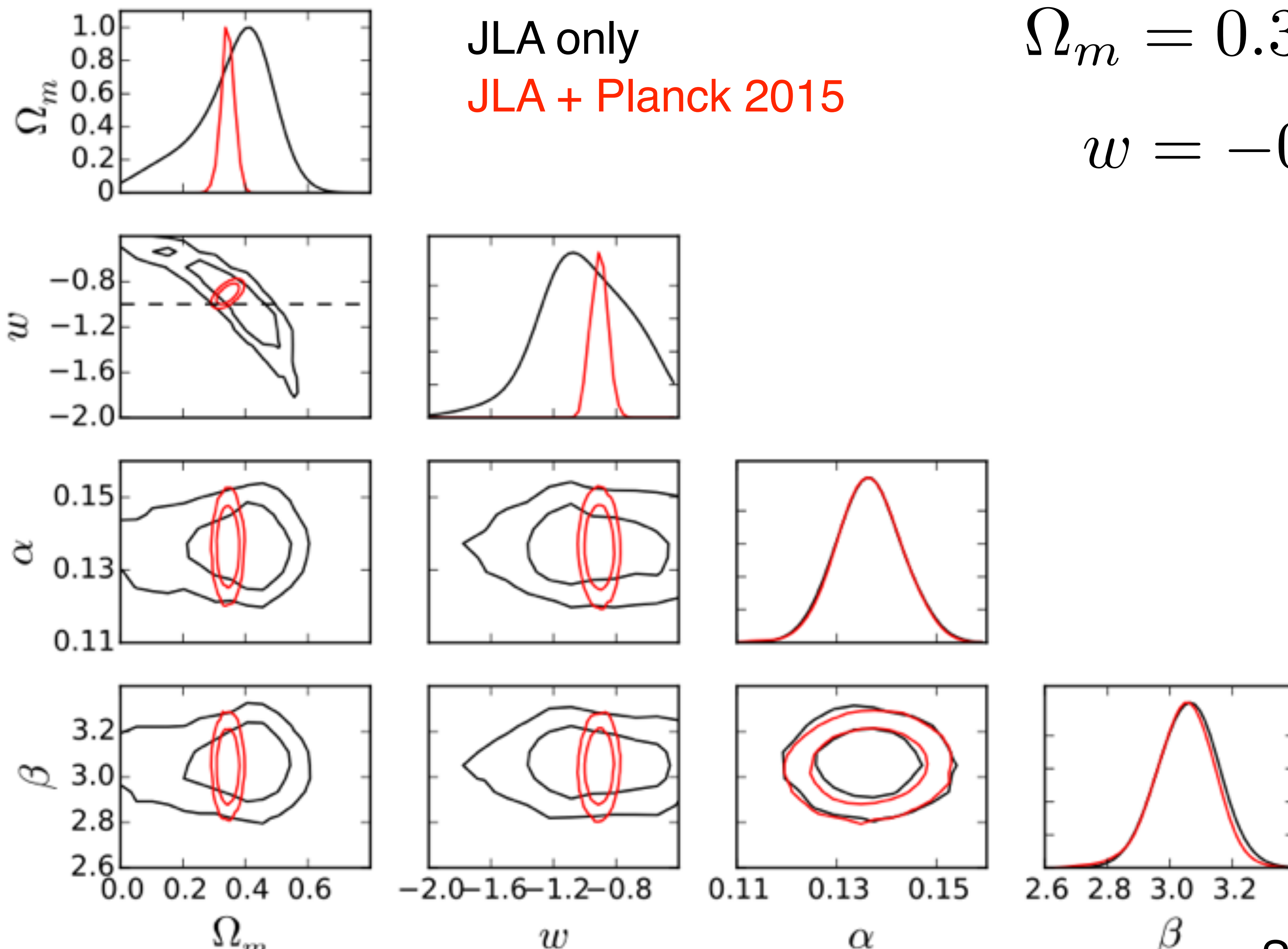
$$\Omega_m = 0.399 \pm 0.027$$

$$\Omega_\kappa = -0.024 \pm 0.010$$

Shariff, RT +, in prep

Cosmological parameters from JLA

$\Omega_\kappa = 0$ Bayesian analysis



$$\Omega_m = 0.343 \pm 0.019$$

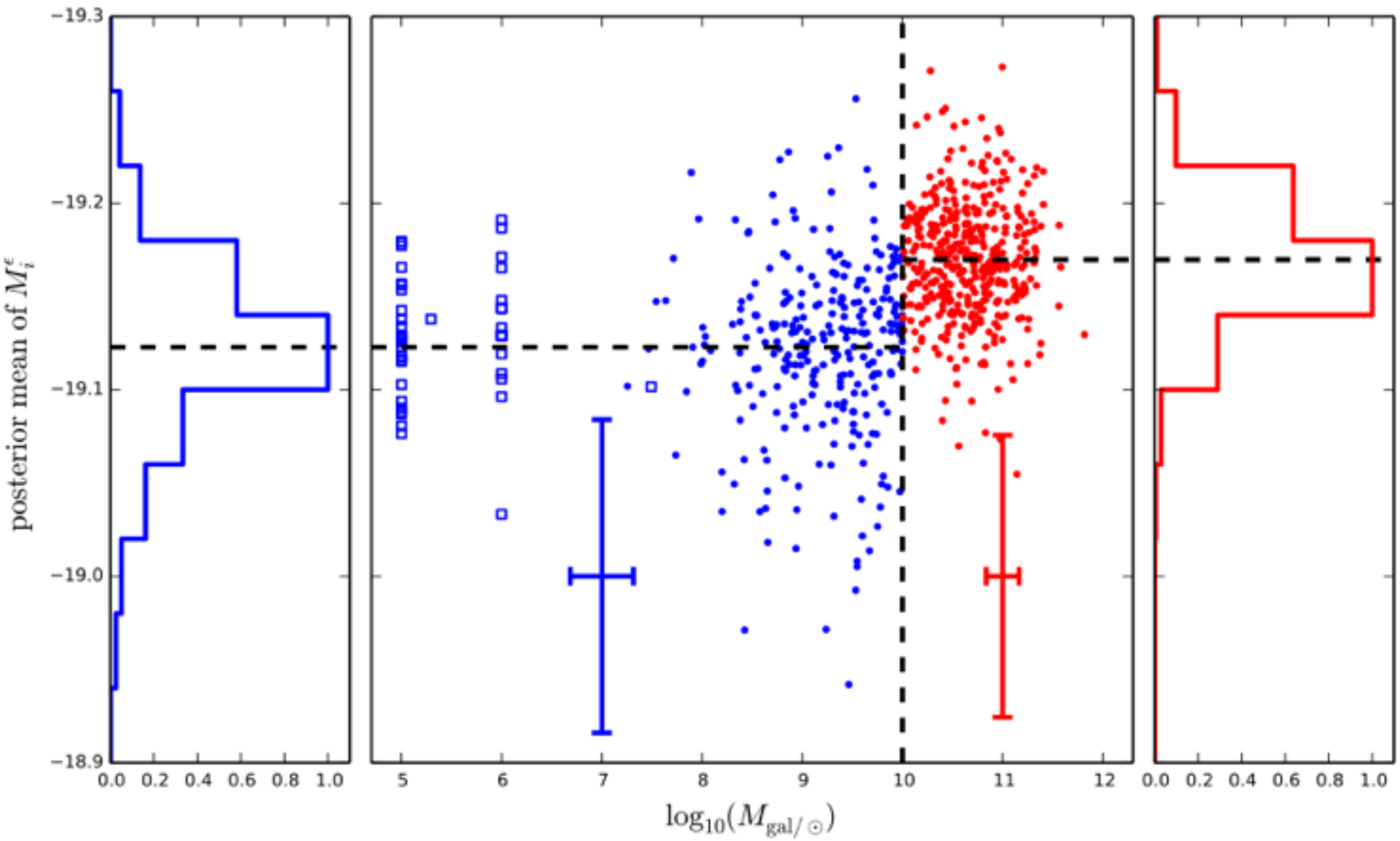
$$w = -0.90 \pm 0.05$$

Shariff, RT +, in prep

Host-galaxy mass as additional covariate

Imperial College London

2-sigma preference for host-galaxy environmental segregation (Kelly+10, Sullivan +10, Rigault+13, Rigault+15) does not change cosmological inferences



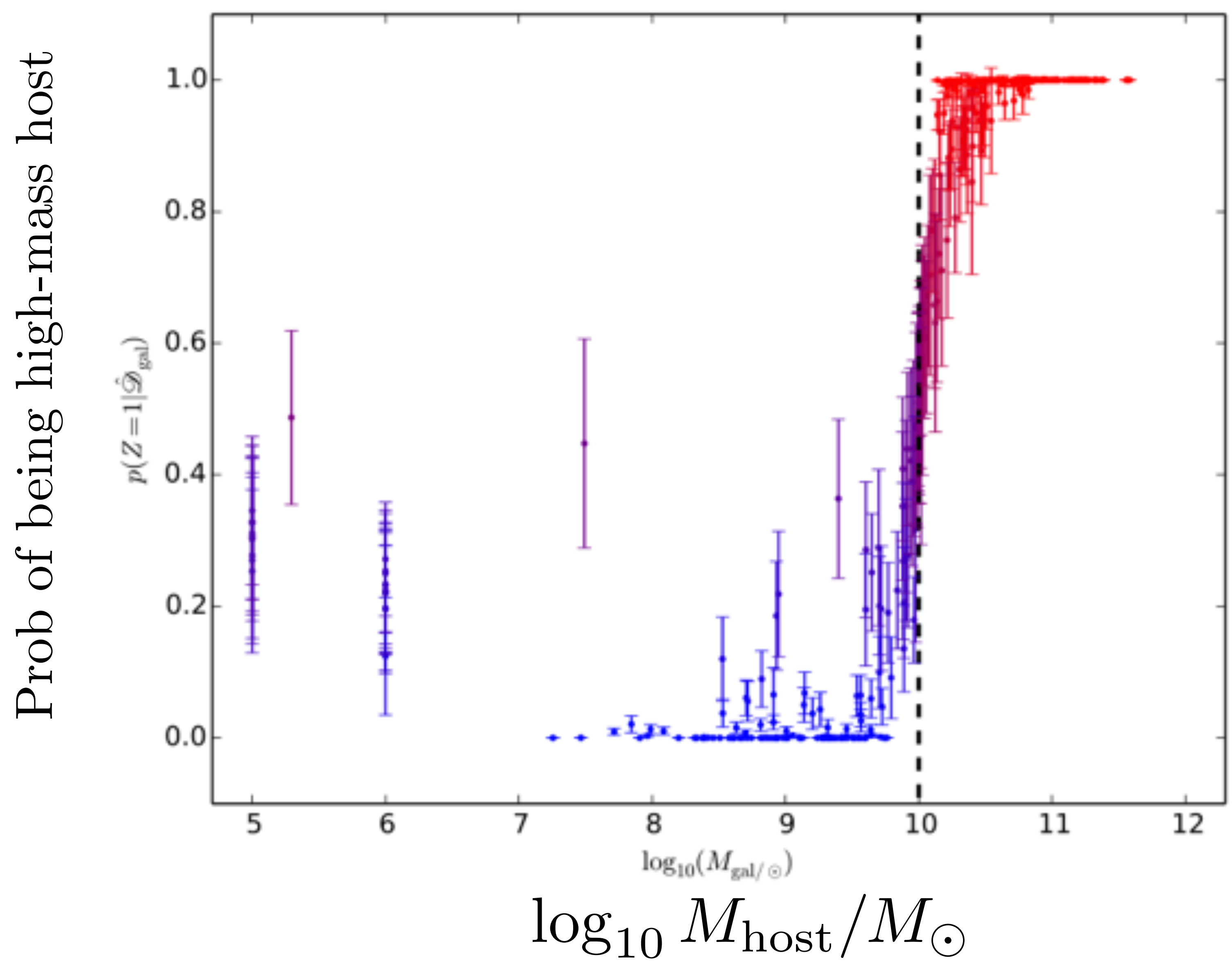
$\log_{10} M_{\mathrm{host}}/M_{\odot}$

Shariff, RT +, in prep

Probabilistic classification of hosts

Imperial College London

Low/High host-galaxy mass preference survives a probabilistic treatment of the host-mass measurement:



Shariff, RT +, in prep

z-evolution of colour correction

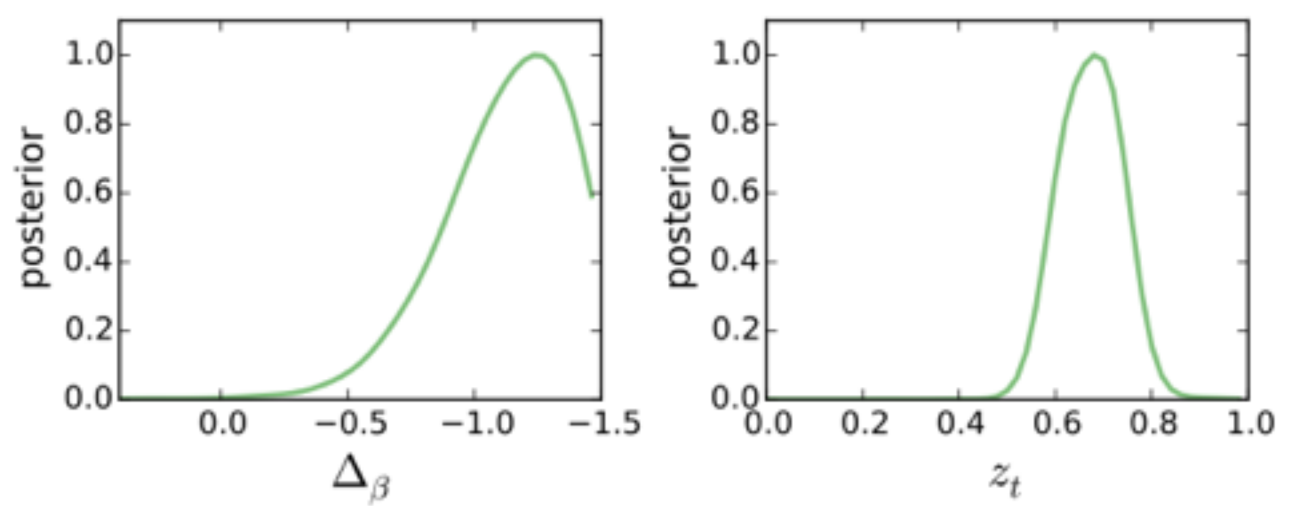
Imperial College London

Strong preference (3-sigma) for a transition from β ~ 3 to β ~ 2 at z ~ 0.7.

This however does not change cosmological inferences.

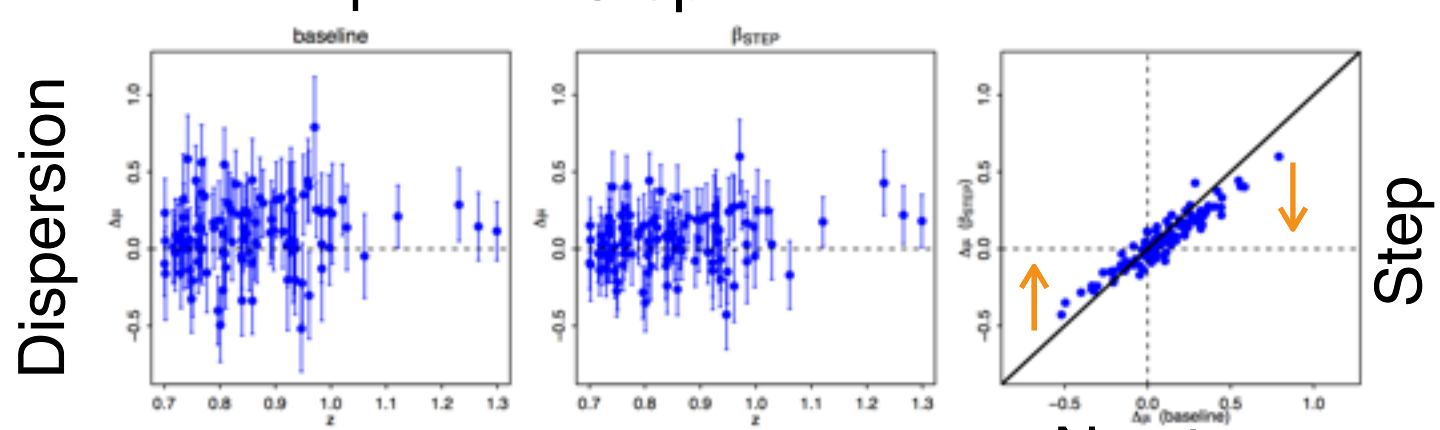
Change in β

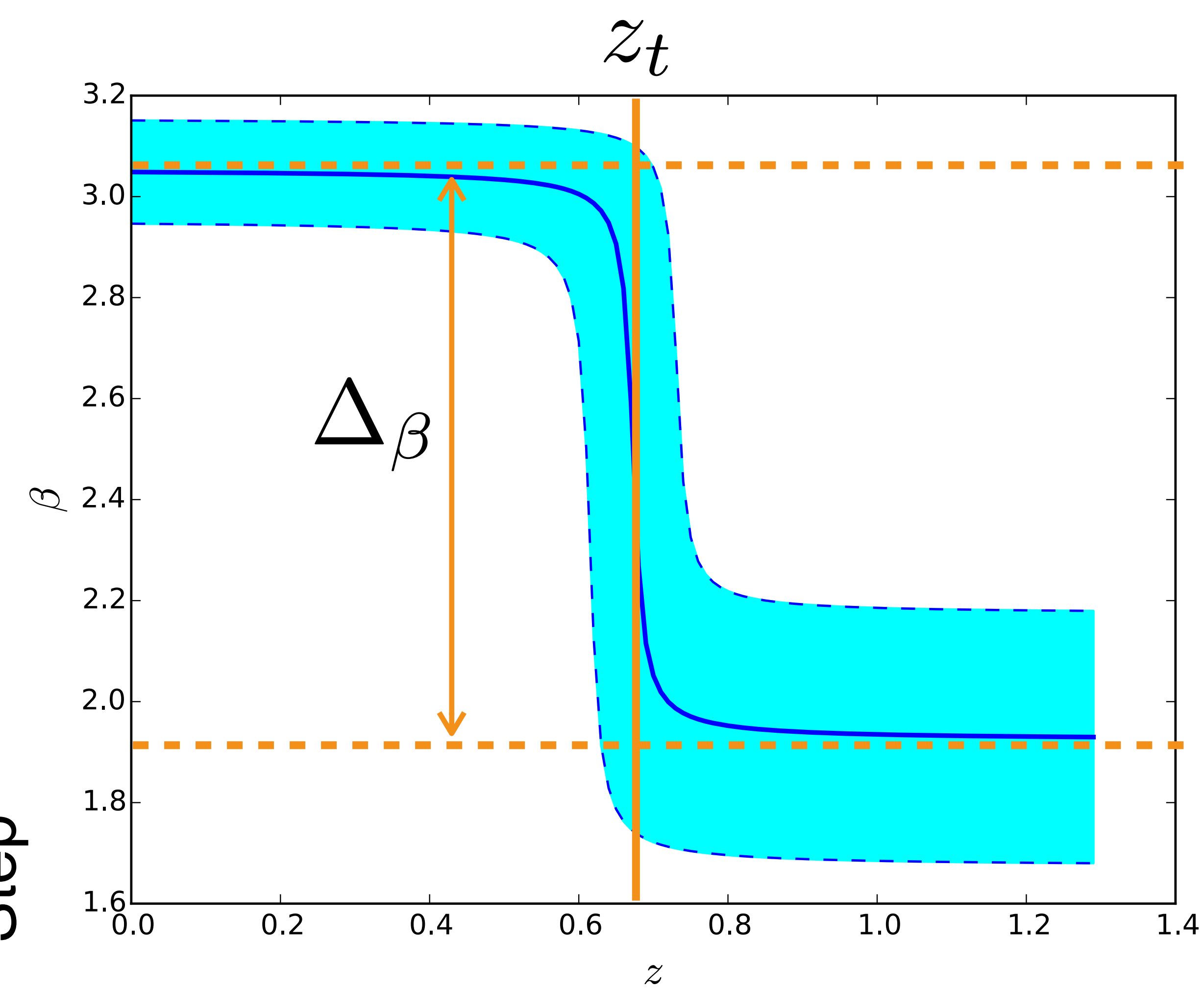
Redshift of transition



No step

Step





Shariff, RT +, in prep

Summary

@R_Trotta

- Bayesian approach allows to model explicitly and consistently many "systematics" of SNIa cosmology
- Correct likelihood function unlocks the power of MCMC in thousands of dimensions
- We find ~ 2-sigma discrepancies in w, Ω_m, Ω_κ wrt standard analysis of JLA data
- Strong evidence for z-evolution of colour correction
- Who is right? Possible solution: re-fit light-curve data using a Bayesian method (RT+, in prep)

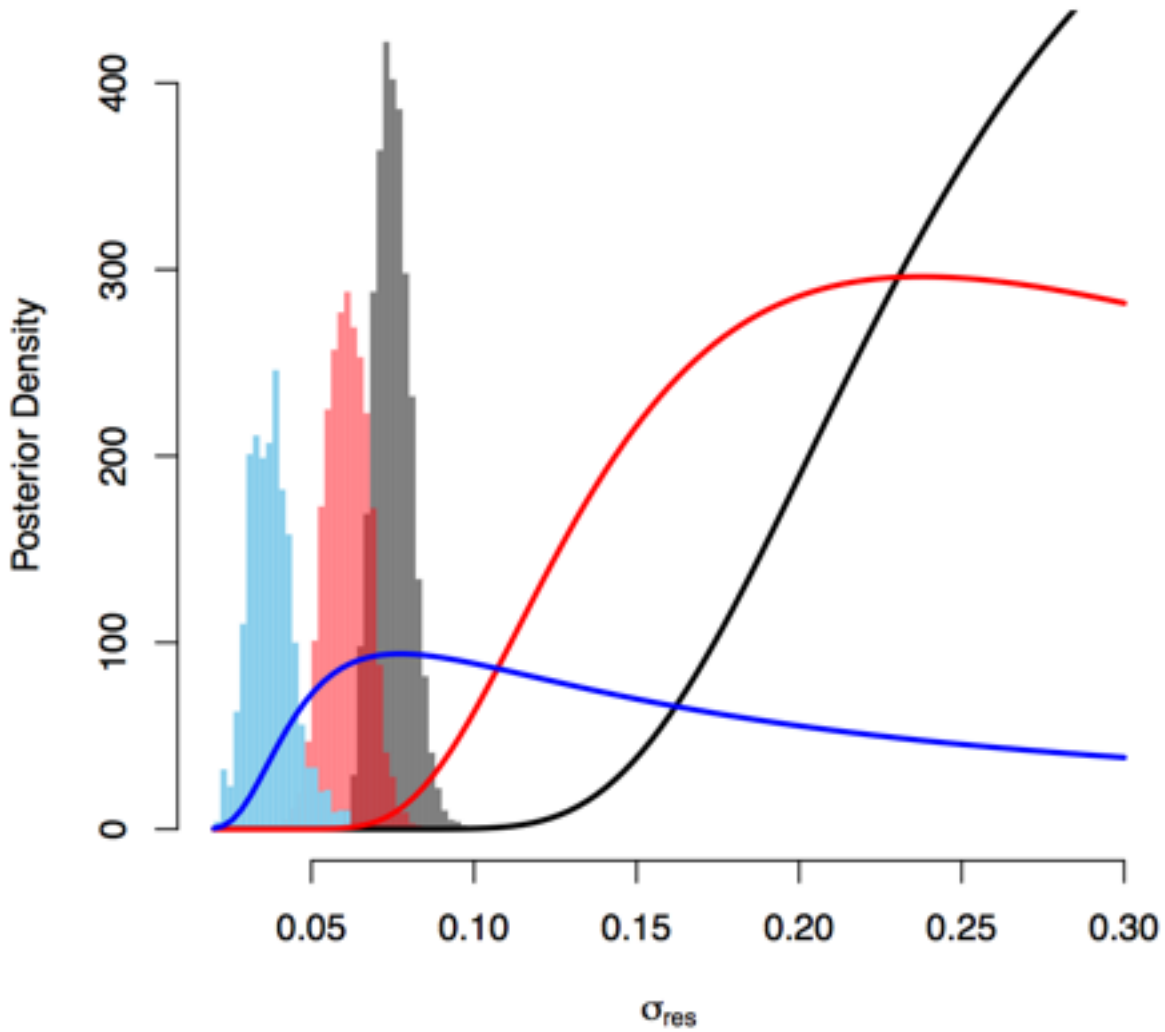
Roberto Trotta
@R_Trotta

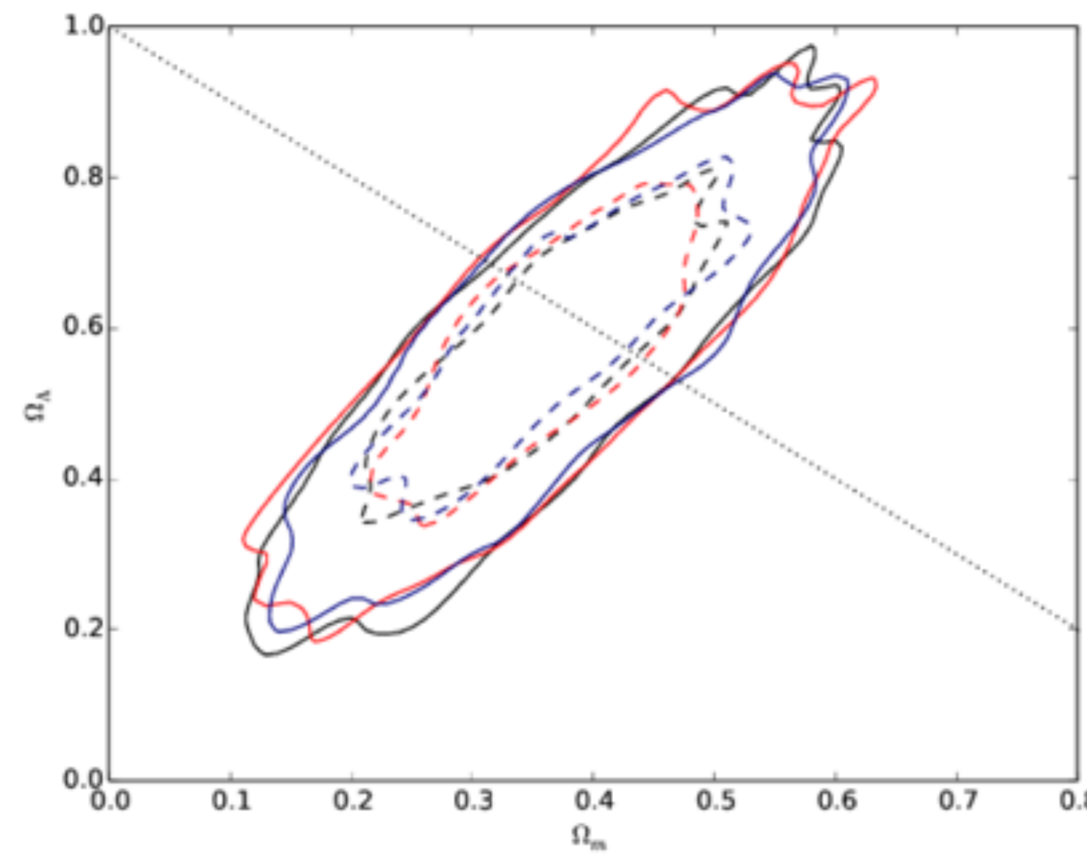
SNIa cosmology is ~~broken~~ fixable

BACKUP SLIDES

Prior robustness

Imperial College London

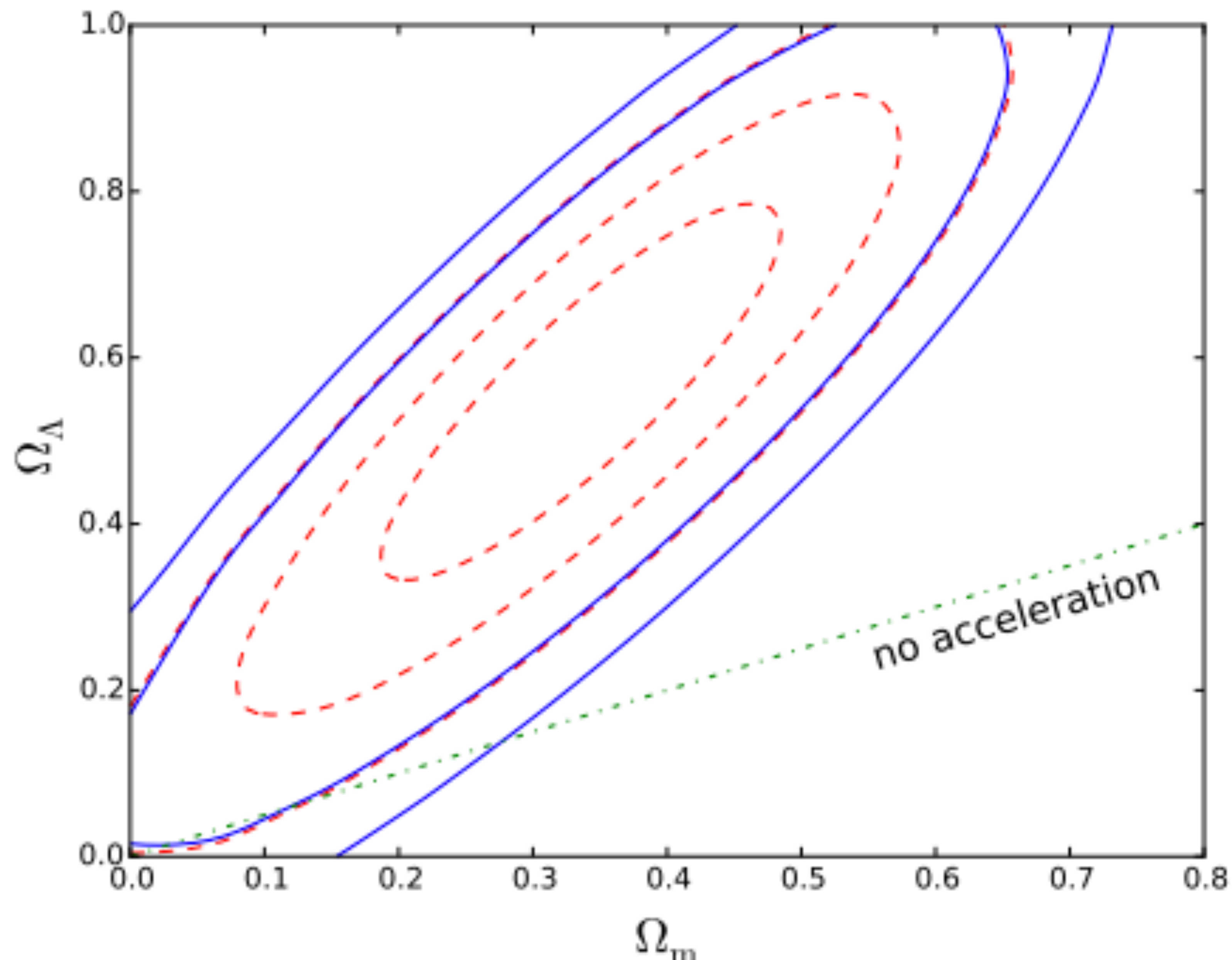




Nielsen+15

Imperial College London

Profile likelihood analysis of an effective likelihood (similar to BHM) claims only a ~3-sigma preference for non-zero acceleration (red/dashed):



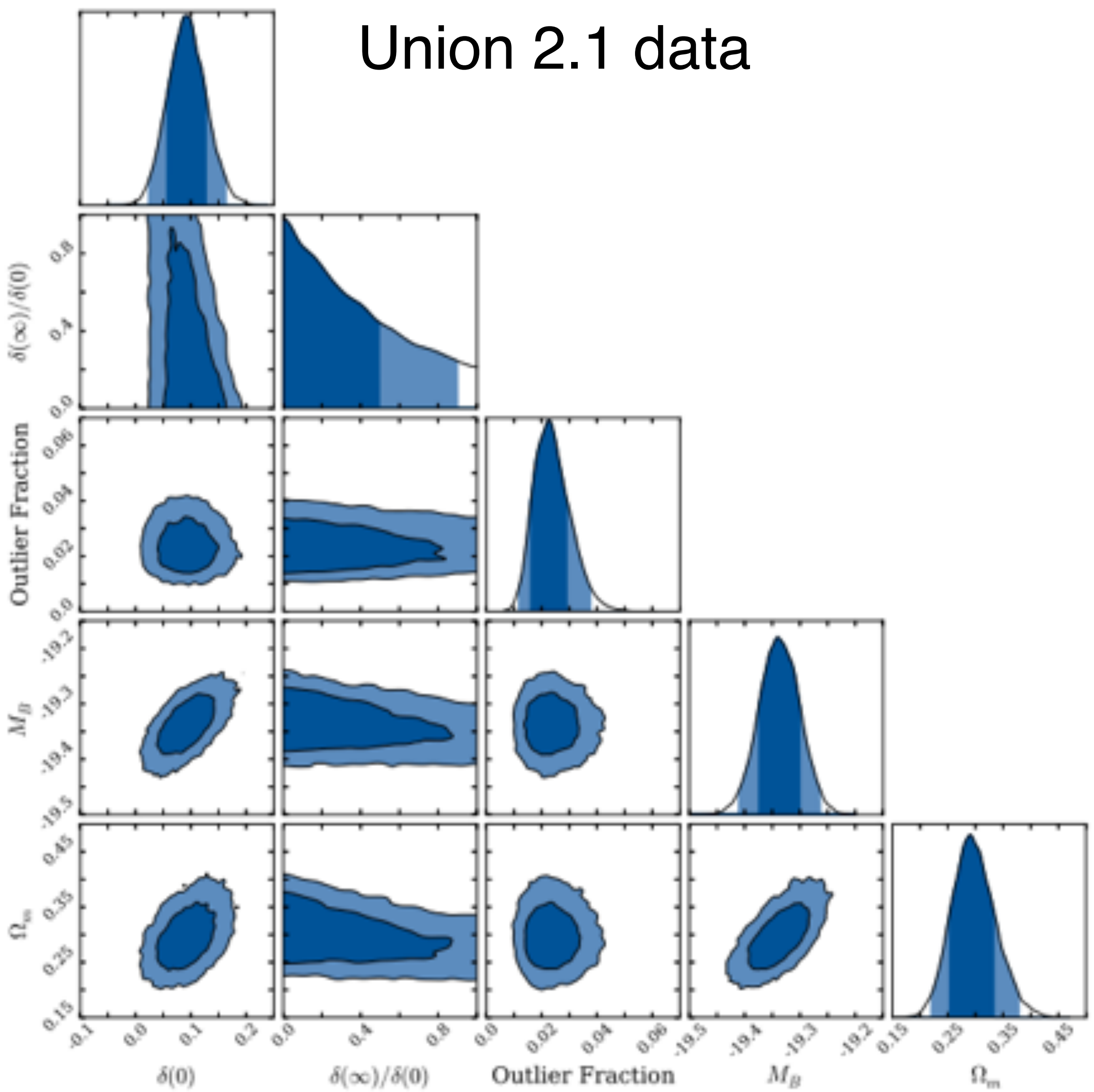
Nielsen+15, 1506.01354

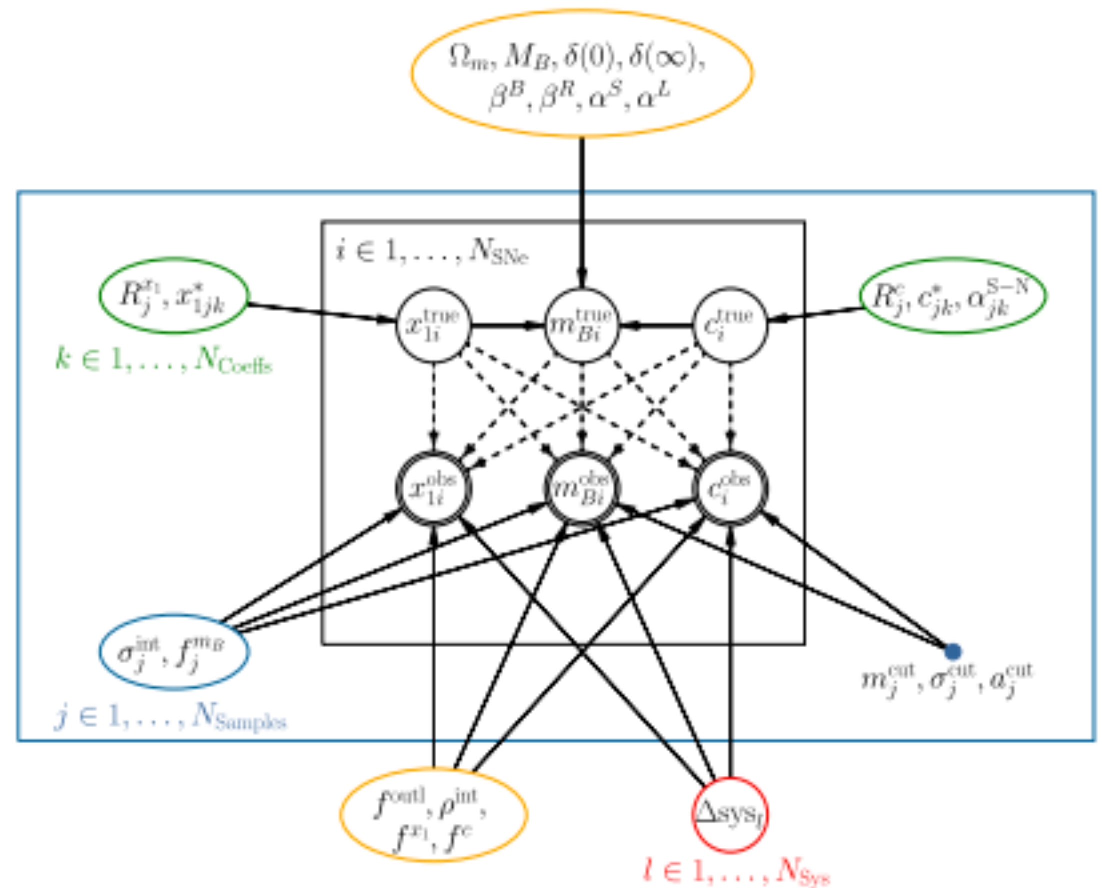
UNITY (Rubin+15)

Imperial College London

Extension of the Bayesian method of March+11 to include outliers, selection effects and host-galaxy mass:

Union 2.1 data

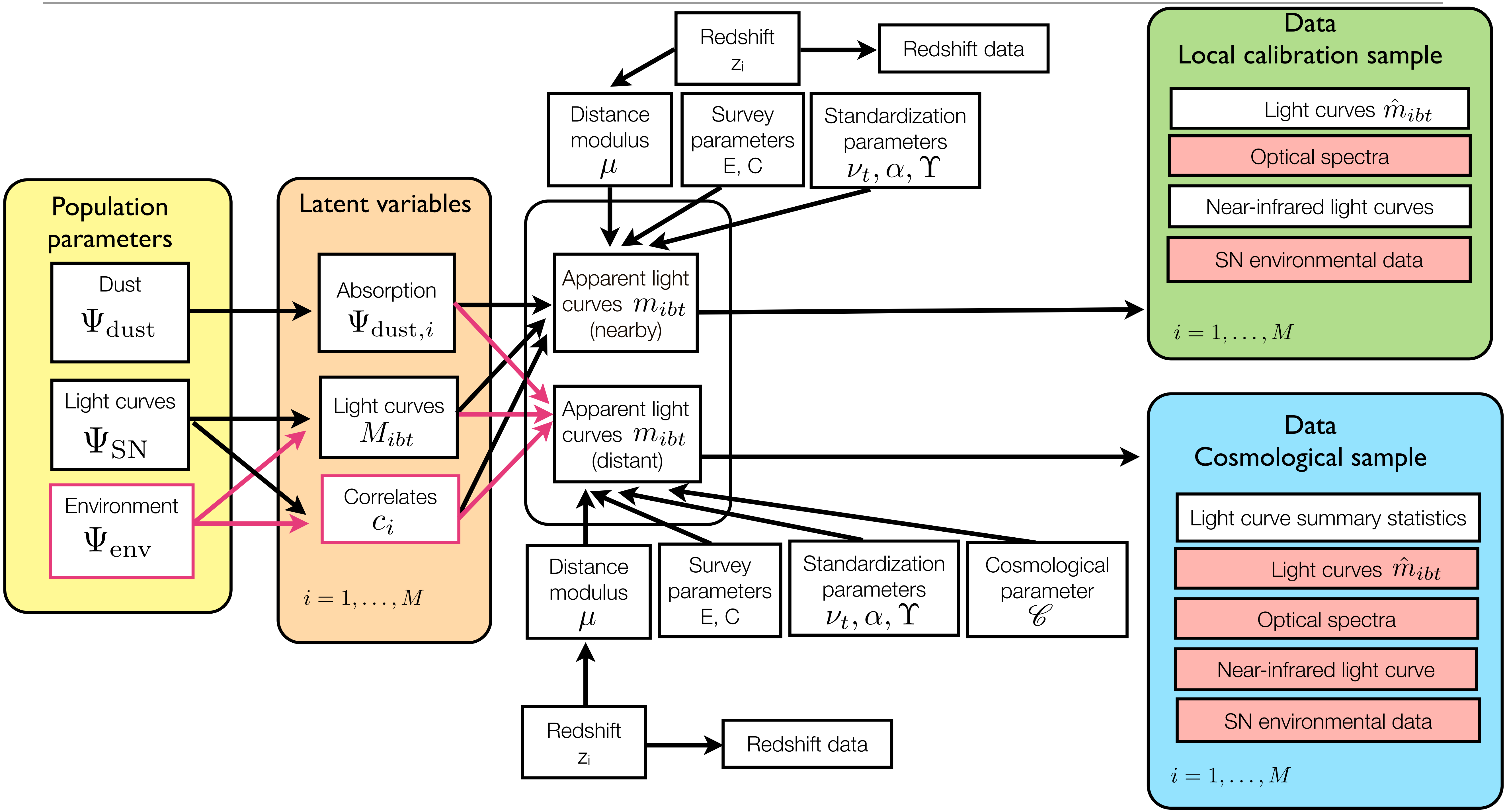




Rubin+15 (SCP) 1507.01602

ICIC

The complete hierarchical model



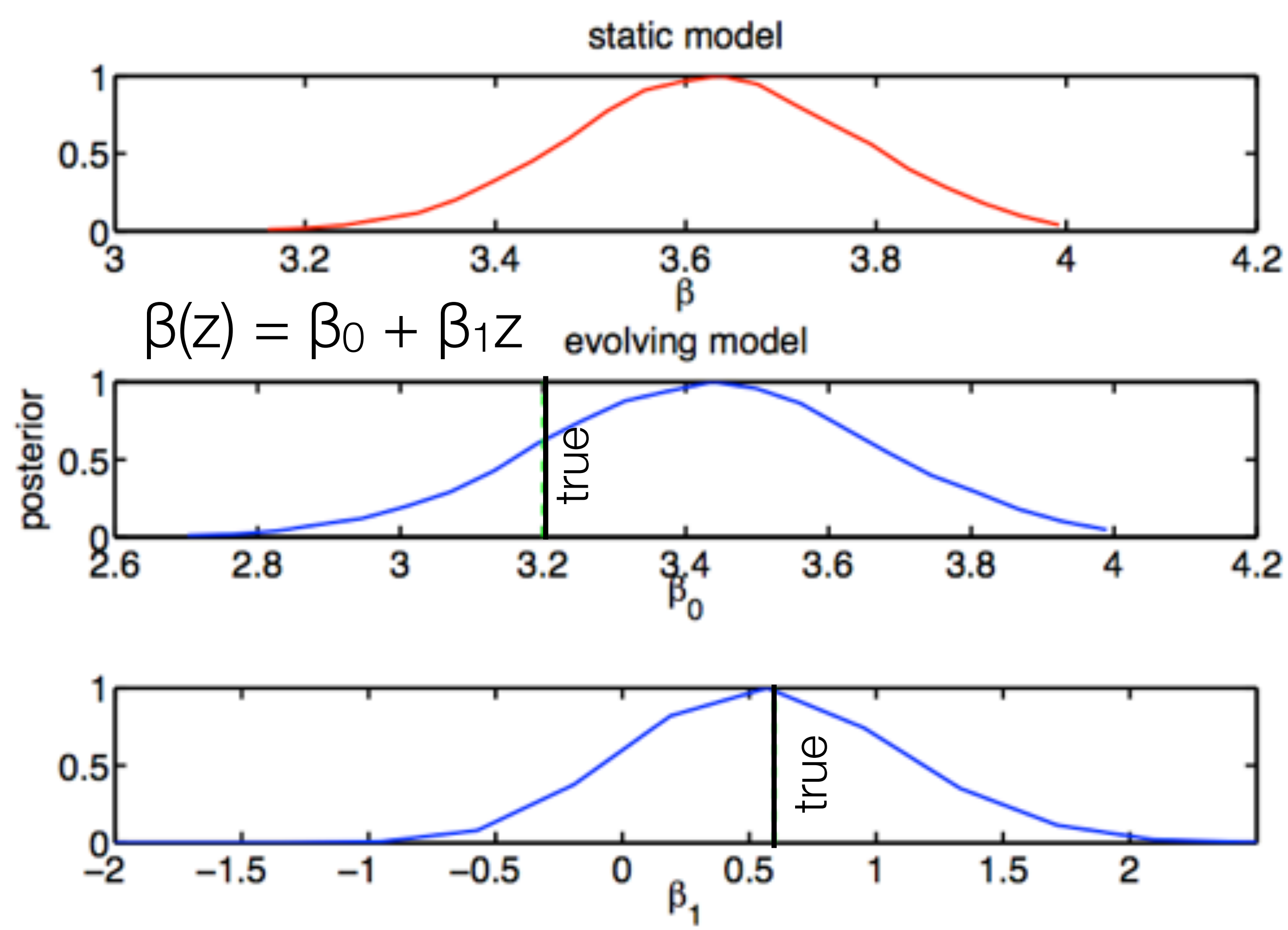
Red arrows/boxes indicate elements/data that have never been explored before in such a multi-level setting

Conclusions and future work

Imperial College London

- Powerful Bayesian methods can take SNIa cosmological inference to a next quantitative step: **reduced systematics thanks to better modeling**
- Required to deal with future large (~ 3000) samples

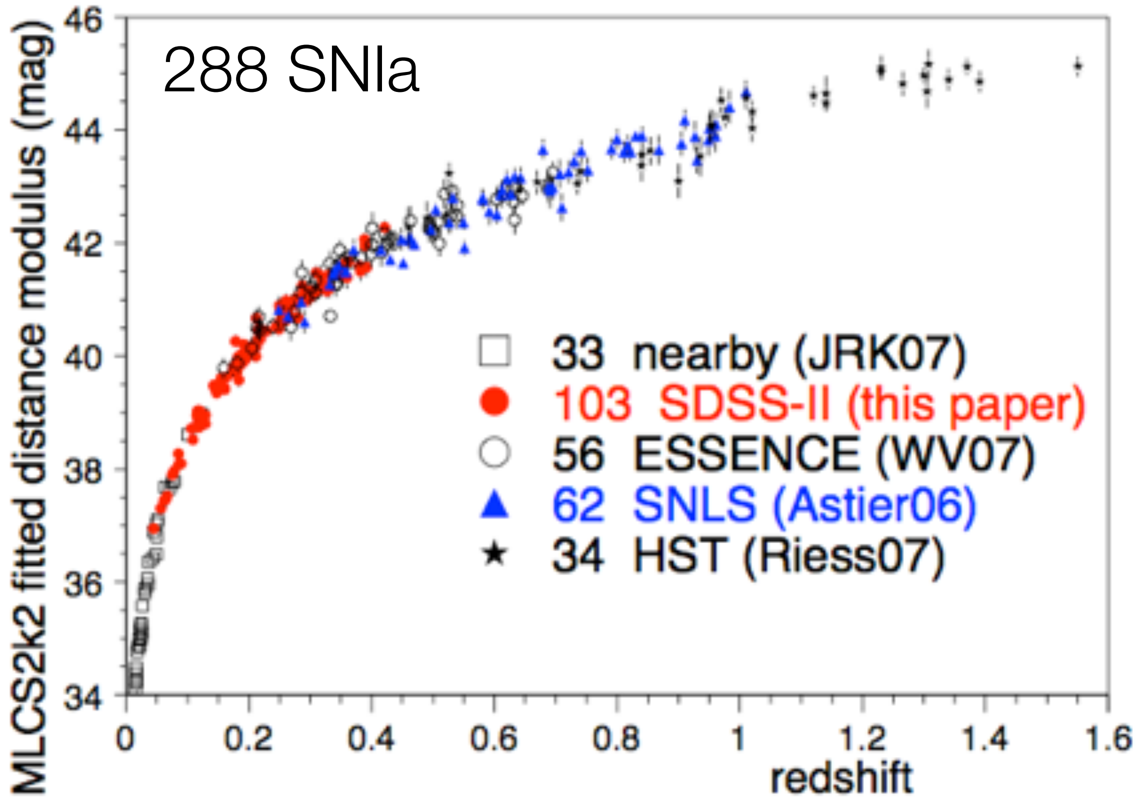
- Extension of our method to include survey selection effects ongoing
- Inclusion of multiple SNIa population, possible redshift-dependence of SNIa properties, correlation with other observables (galaxy mass, metallicity, spectral lines, etc) straightforward
- Bayesian model comparison (LCDM vs modified gravity)



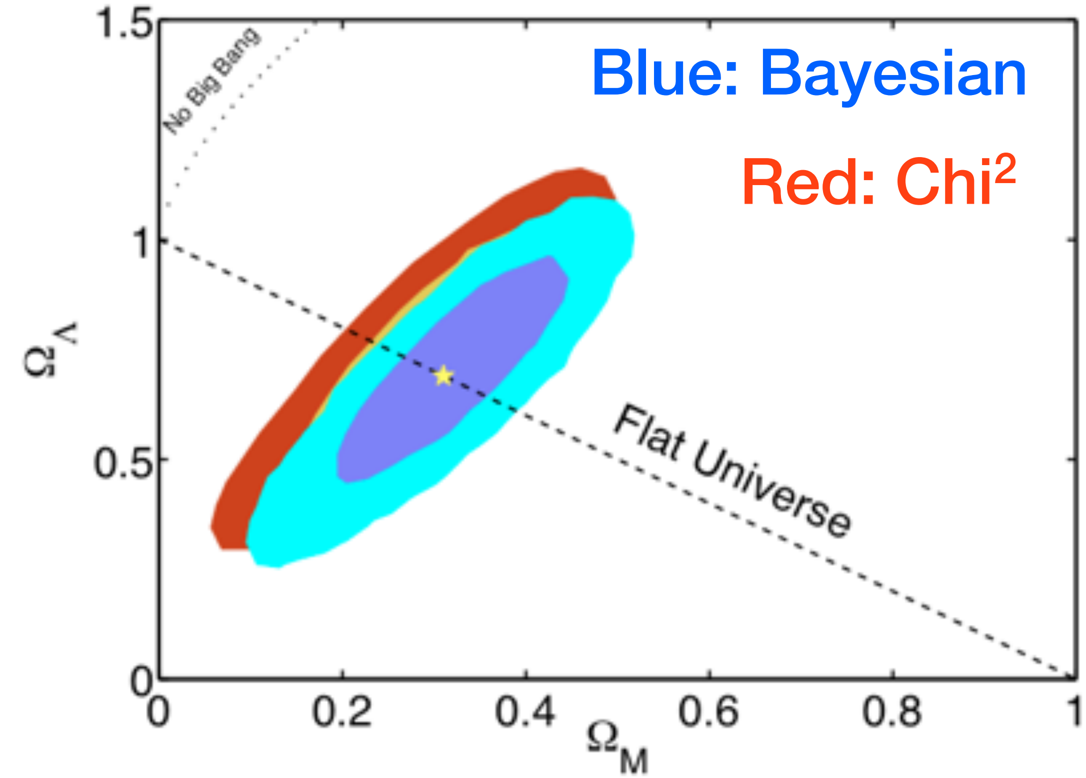
Cosmology results (Union)

Combined sample

Kessler et al (SDDS collaboration) (2010)

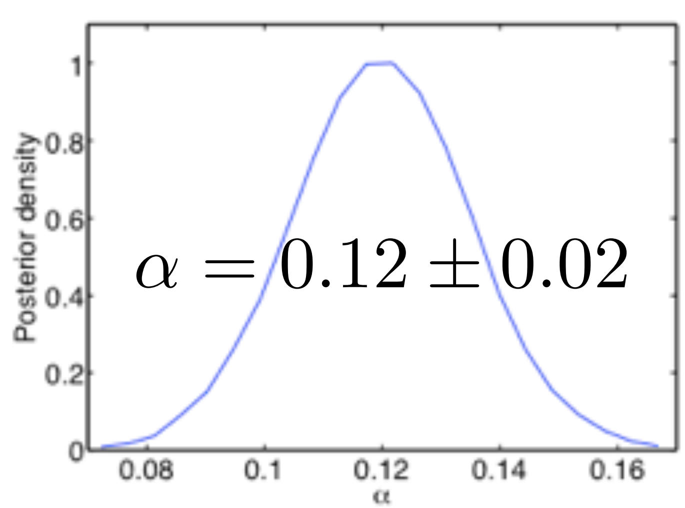


w = 1

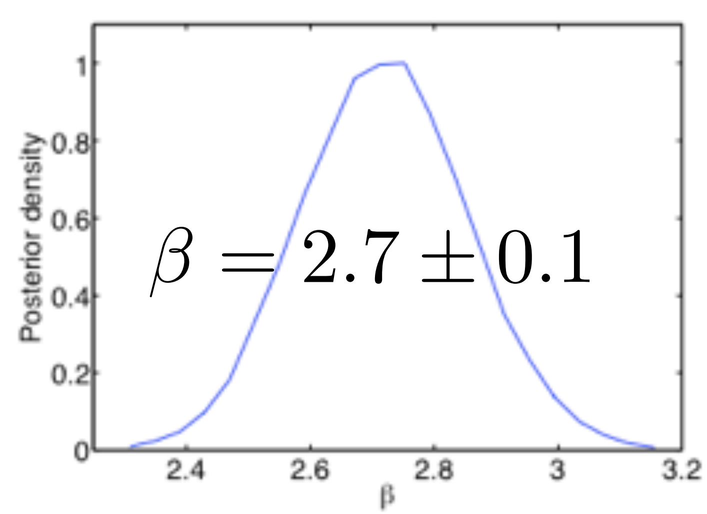


March et al (2011)

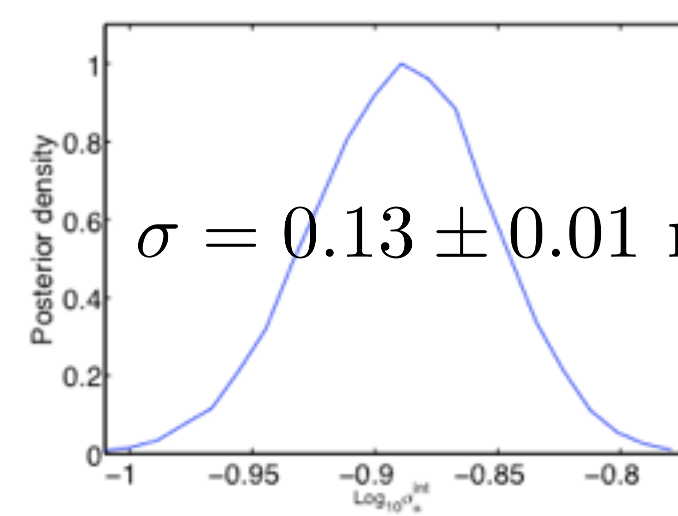
Marginal posteriors



$$\alpha = 0.12 \pm 0.02$$



$$\beta = 2.7 \pm 0.1$$

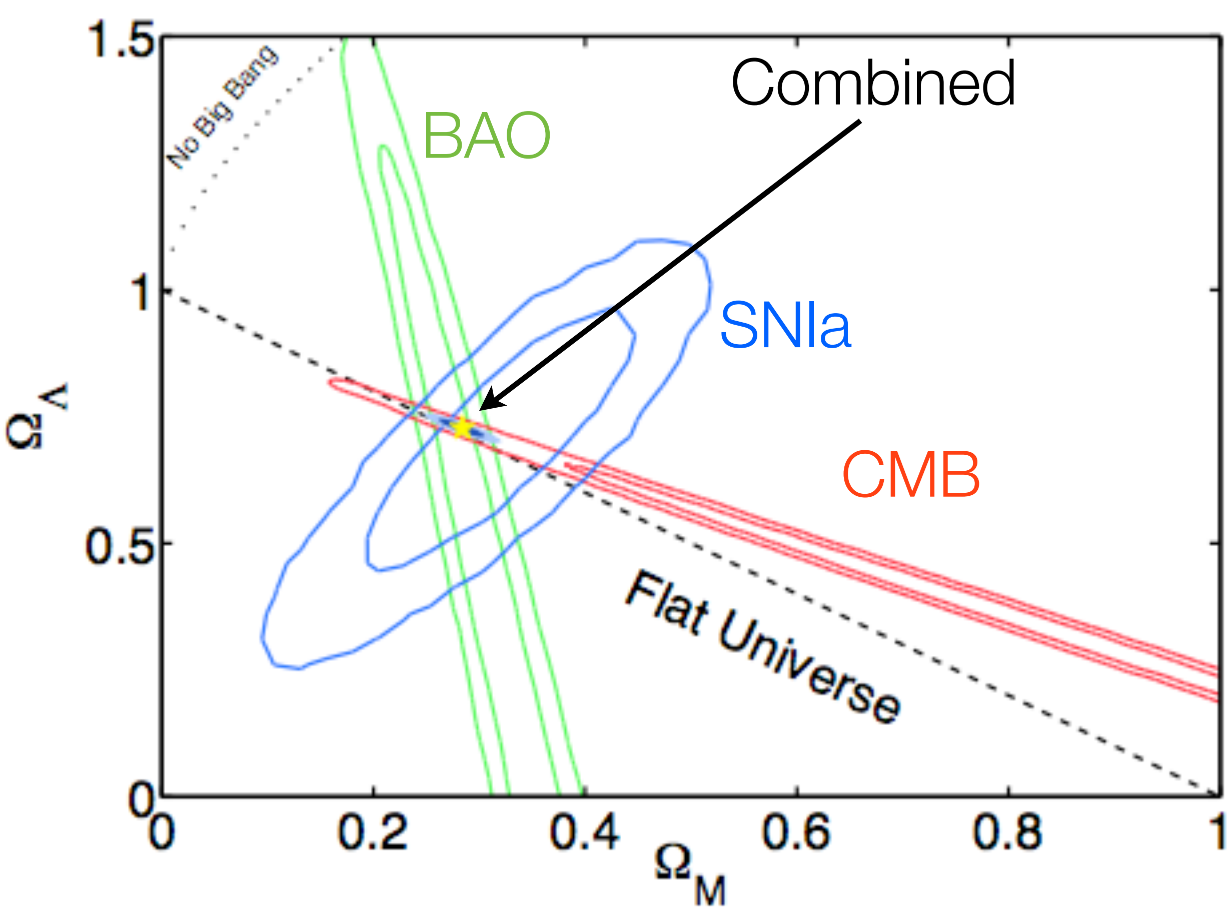


$$\sigma = 0.13 \pm 0.01 \text{ mag}$$

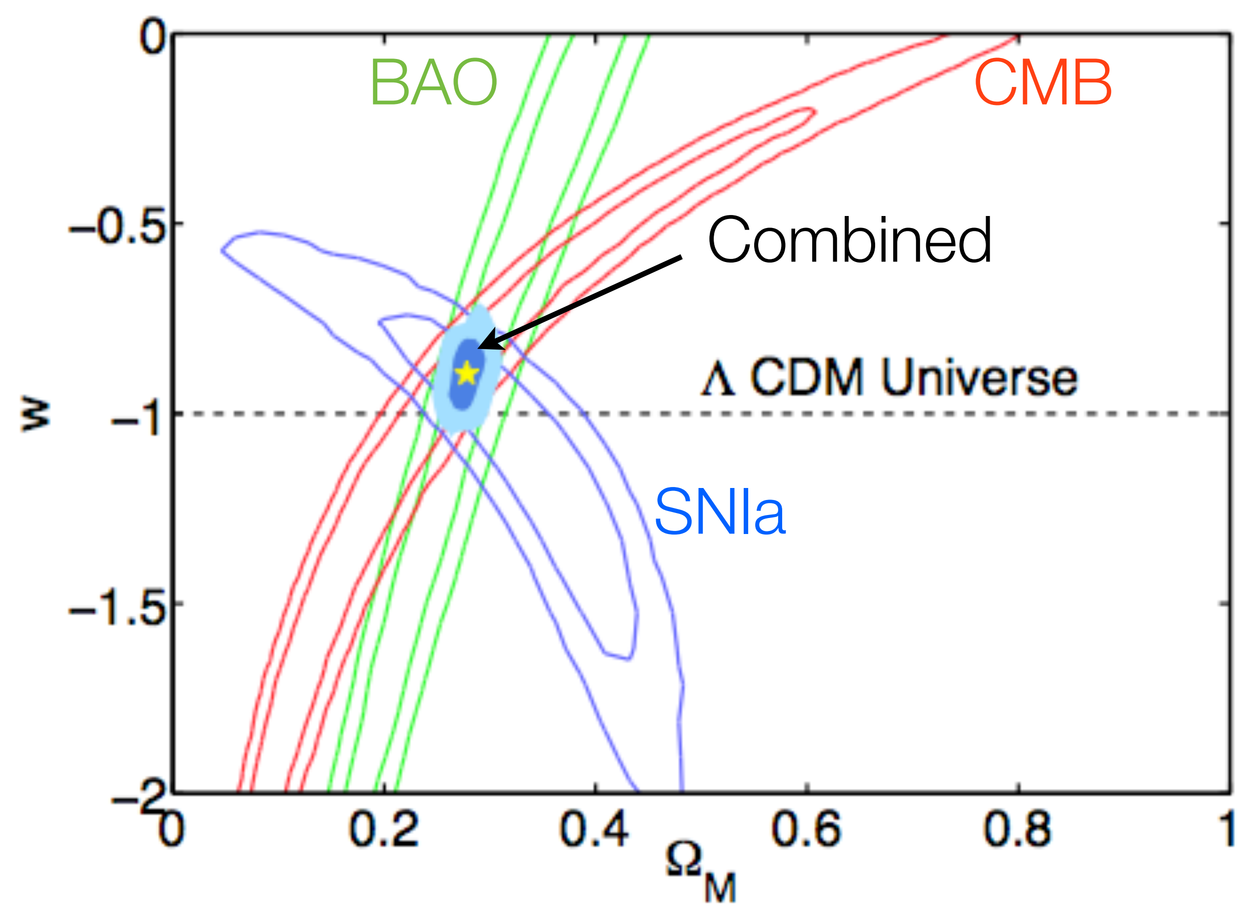
Combined constraints

- Combined cosmological constraints on matter and dark energy content:

w = 1



$\Omega_K = 0$

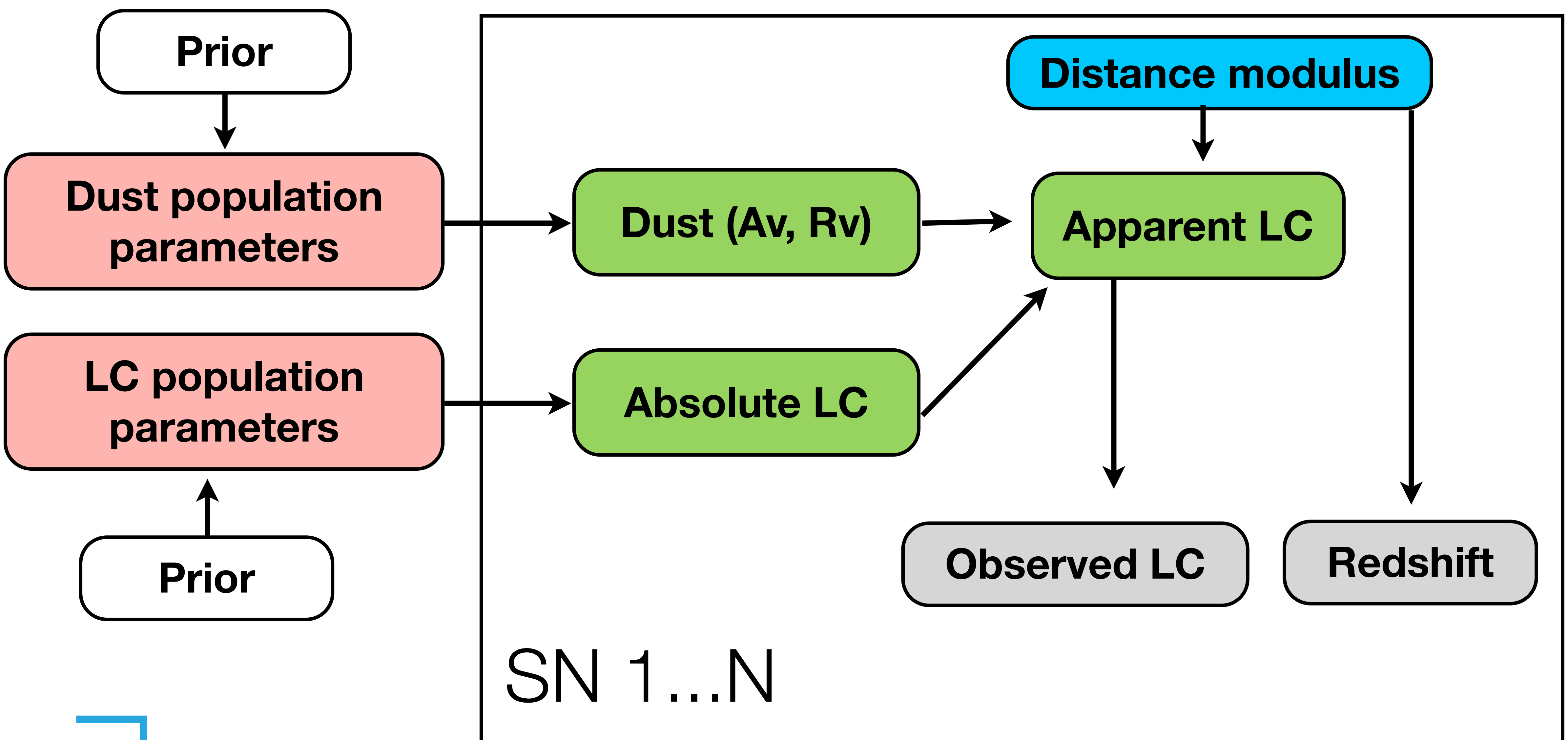


March, RT et al (2011)

The BayeSN approach

Imperial College London

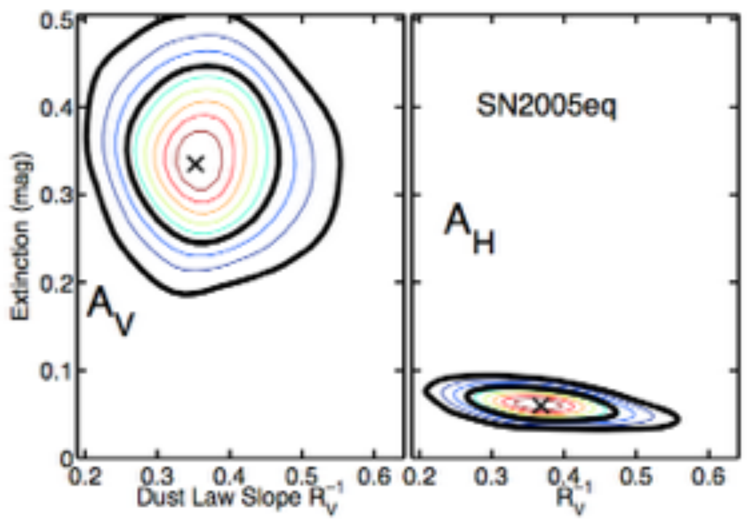
- Developed by K. Mandel (Mandel et al, 2009, 2011) and collaborators: fully Bayesian approach to LC fitting, including random errors, population structure, intrinsic variations/correlations, dust extinction and reddening, incomplete data



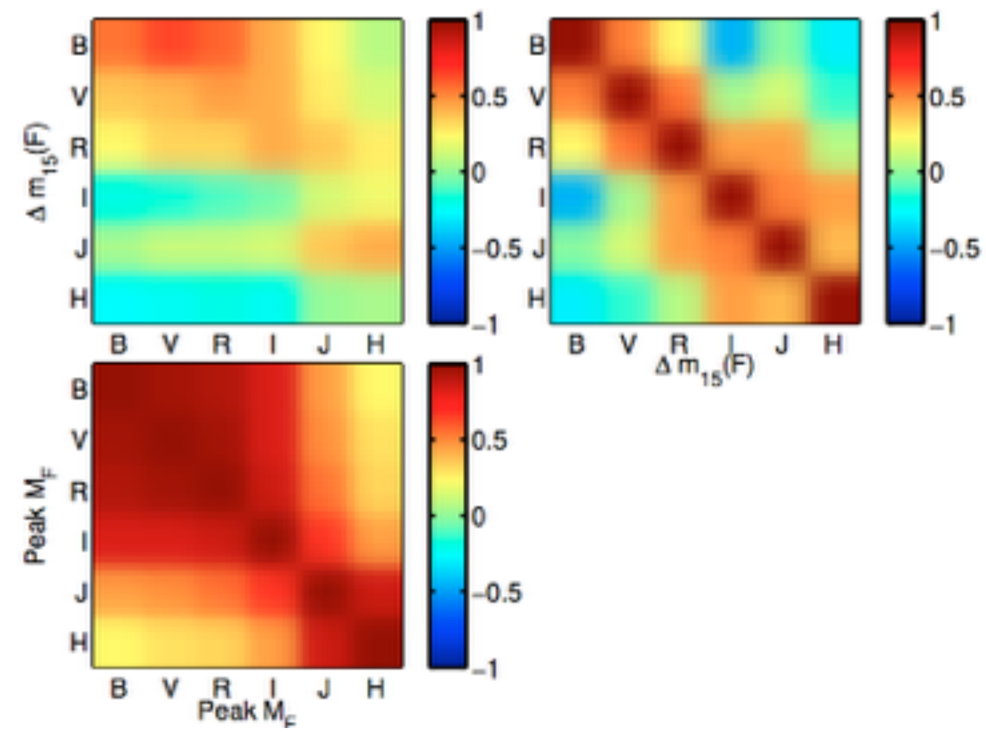
Some results from BayeSN

Mandel et al (2011)

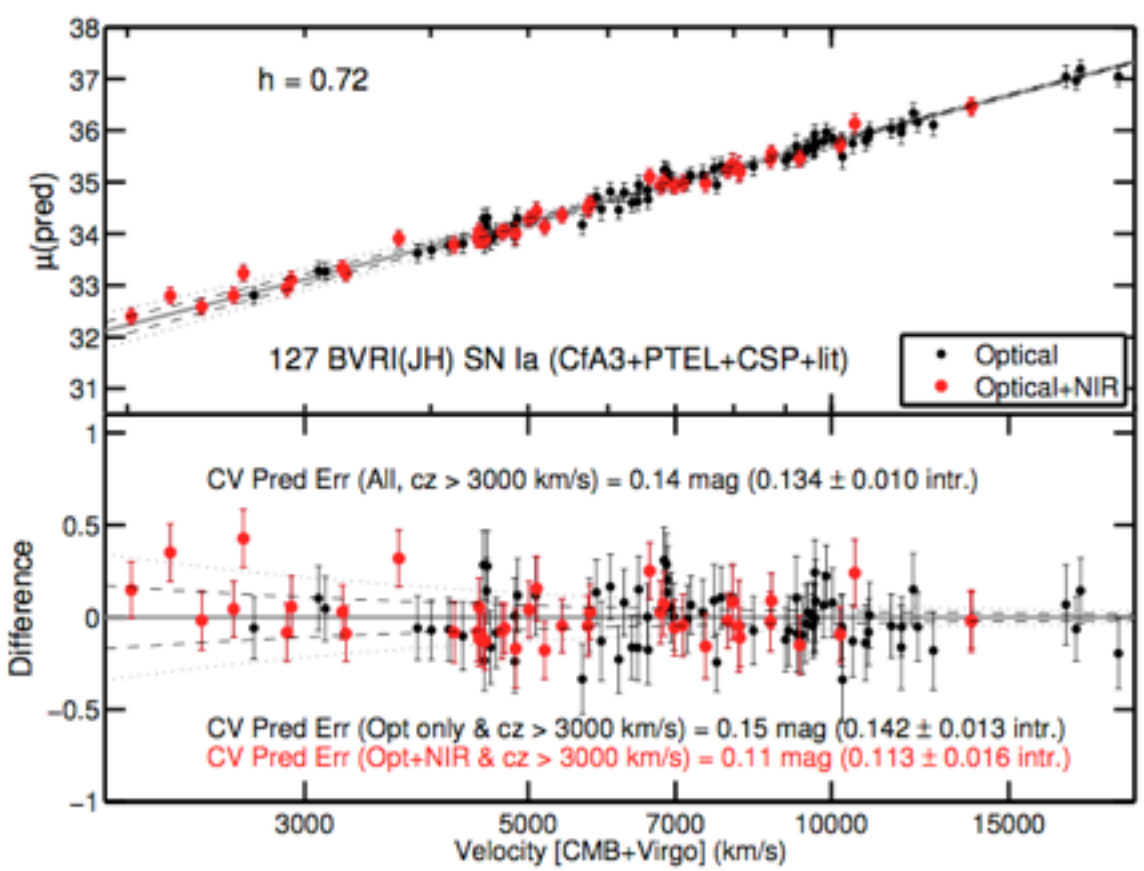
Dust absorption for each SNIa



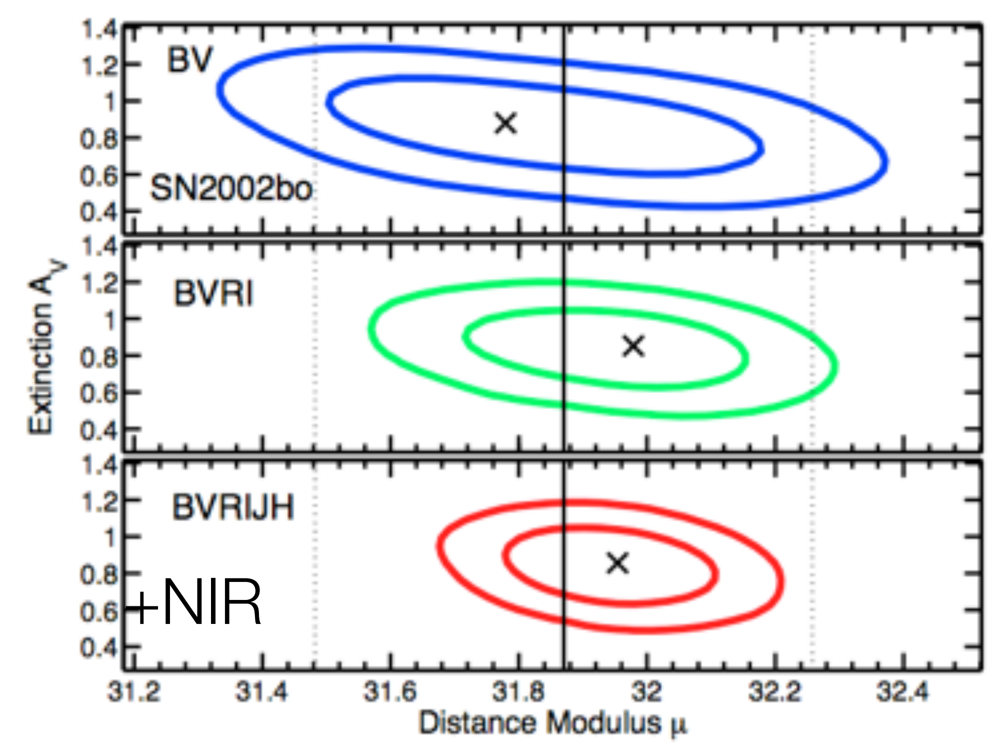
Population level analysis of correlations



Hubble diagram: residual scatter reduced by ~2 using optical+NIR LC



Inclusion of NIR LC



From lightcurves to distances

Imperial College London

- There are a few different lightcurve (LC) fitters on the market, with different philosophies/statistical approaches:
- **MLCS2k2** (Jha et al, 2007): color (A_V) and LC shape (Δ) parameters fitted simultaneously with cosmology. Color correction includes a dust extinction law correction.
- **SALT/SiFTO/SALT2** (Guy et al, 2007): LC shape (x_1) and colour (c) correction extracted from LC alongside apparent B-band magnitude (m_B) + covariance matrix. The distance modulus

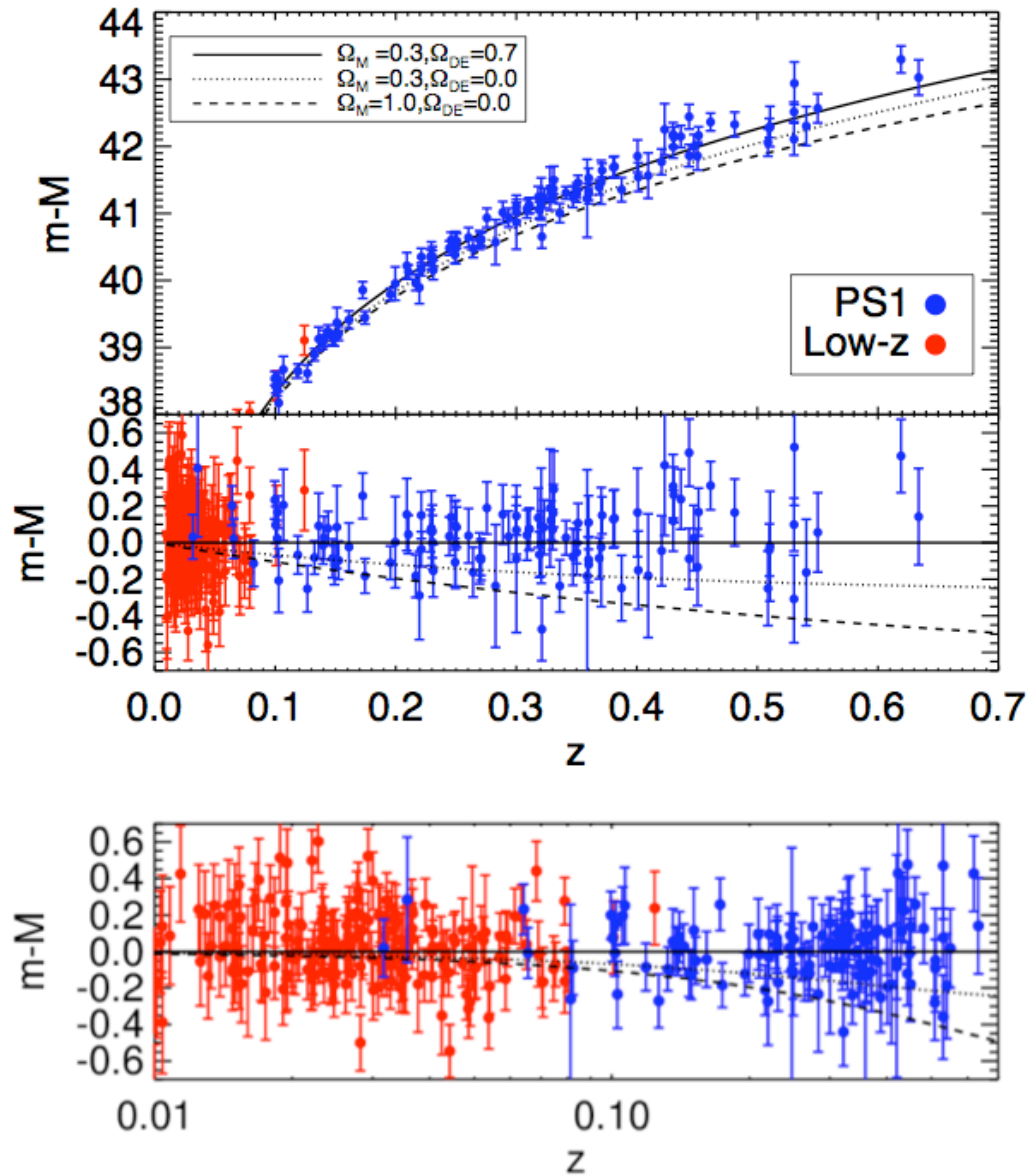
$$\mu = m_B - M + \alpha \times \mathrm{width} - \beta \times \mathrm{colour}$$

  is subsequently estimated with cosmological parameters and remaining "intrinsic" scatter.
- **BayeSN** (Mandel et al, 2009, 2011): Fully Bayesian hierarchical modeling of LC, including population-level distributions (see later).

@R_Trotta

PS1 data

- Most recent data set from PAN-STARRS1 survey
- 146 spectroscopically confirmed SNIa
- Cosmological fit: 112 PS1 at high-z (blue) + 201 low-z SNIa (red)



At the heart of the method...

Imperial College London

- ... lies the fundamental problem of **linear regression** in the presence of measurement errors on both the dependent and independent variable and intrinsic scatter in the relationship (e.g., Gull 1989, Gelman et al 2004, Kelly 2007):

$$\mu_i = m_{B,i} - M_i + \alpha x_{1,i} - \beta c_i$$

anagolous to

$$y_i = b + a x_i$$

$$x_i \sim p(x|\Psi) = \mathcal{N}_{x_i}(x_\star, R_x)$$

POPULATION DISTRIBUTION

$$y_i|x_i \sim \mathcal{N}_{y_i}(b + a x_i, \sigma^2)$$

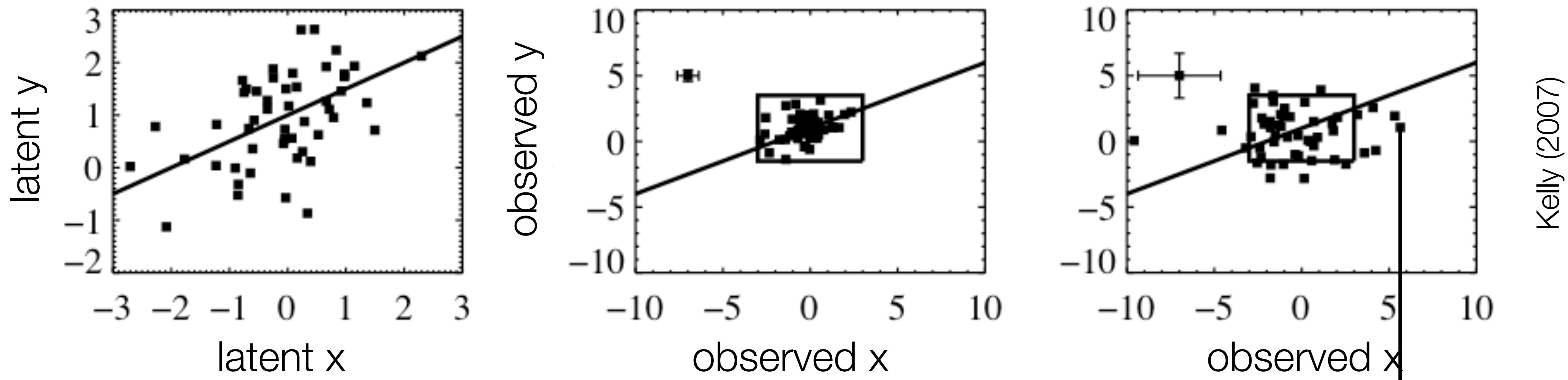
INTRINSIC VARIABILITY

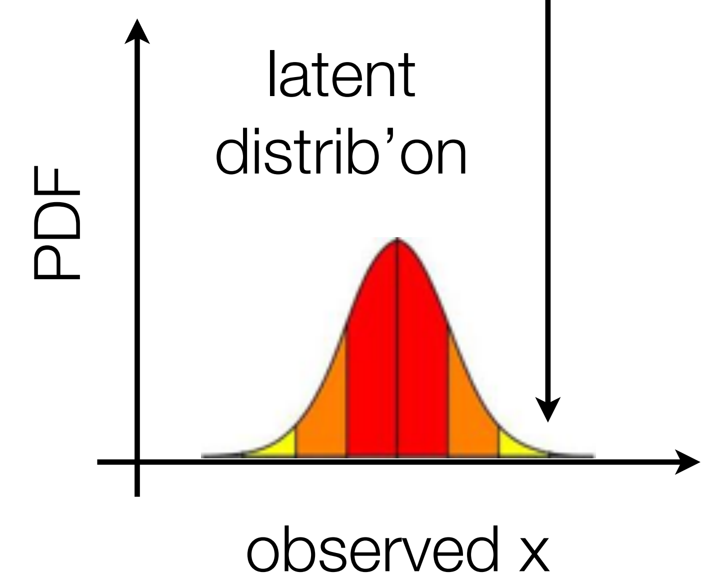
$$\hat{x}_i, \hat{y}_i | x_i, y_i \sim \mathcal{N}_{\hat{x}_i, \hat{y}_i}([x_i, y_i], \Sigma^2)$$

MEASUREMENT ERROR

ICIC

INTRINSIC VARIABILITY **+ MEASUREMENT ERROR**





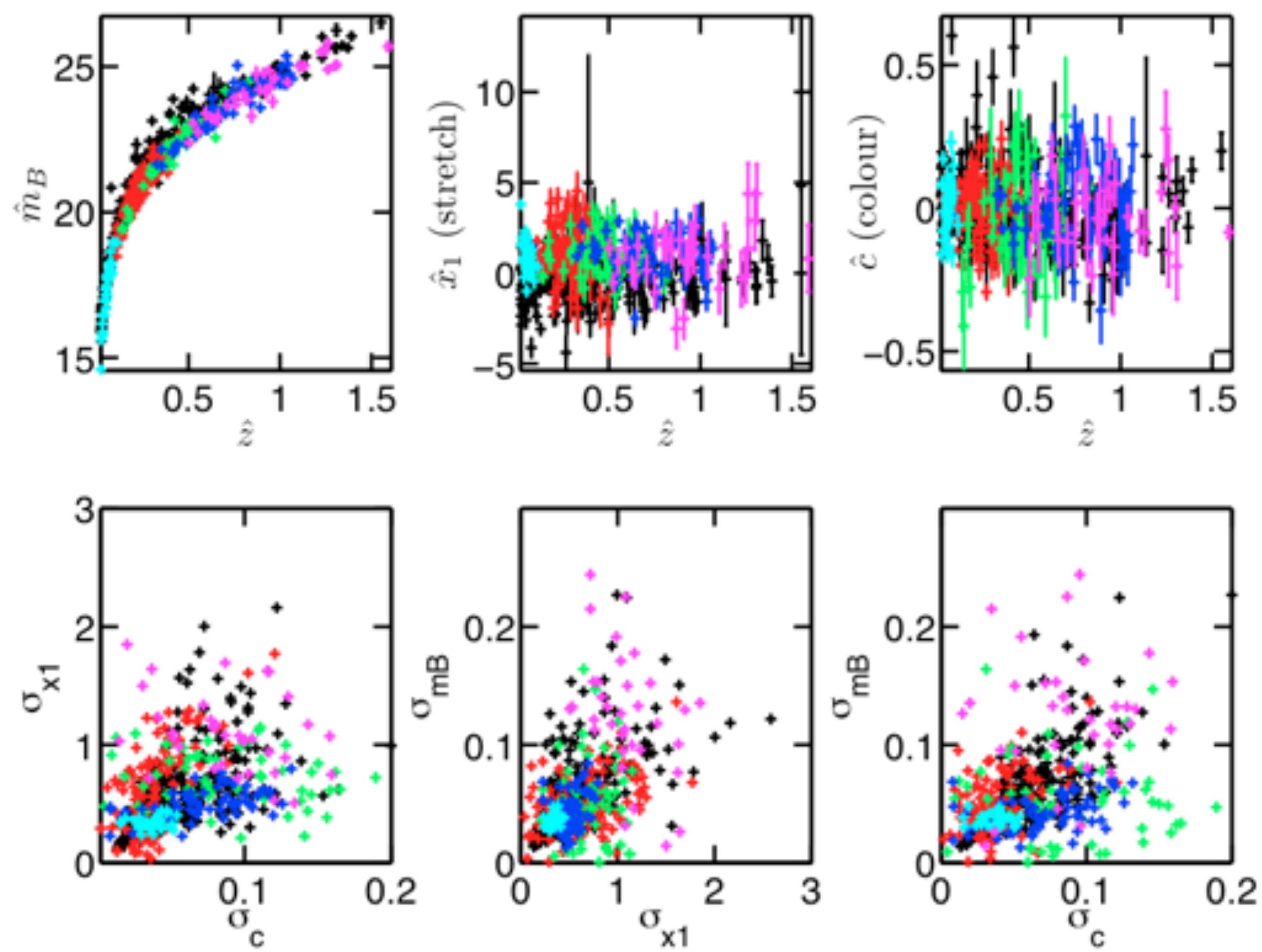
- Modeling the latent distribution of the independent variable accounts for “Malmquist bias”

- An observed x value far from the origin is more probable to arise from up-scattering (due to noise) of a lower latent x value than down-scattering of a higher (less probable) x value

Tests on simulated SNIa data

- Simulated N=288 SNIa with similar characteristics as SDSS +ESSENCE+SNLS+HST +Nearby sample
- Reconstruction of cosmological parameters over 100 realizations, comparing Bayesian hierarchical method with standard Chi^2.

Simulated SNIa realization (colour coded according to "survey")



March et al (2011)

Posterior sampling

- In the Bayesian hierarchical approach, we have
  - 3 cosmological parameters: H_0, Ω_M, Ω_K (w=1) or H_0, Ω_M, w (Ω_K =0)
  - 2 stretch/colour correction parameters: α, β
  - 6 population-level parameters: M_0, σ^2, x^*, R_x, c^*, R_c
  - 3N (=864) latent variables M_i, x_{1i}, c_i
- Analytical marginalization over all latent variables and linear population-level parameters is possible in Gaussian case (no selection effects). Sampling of the remaining parameters via MultiNest.
- Alternatively, Gibbs sampling can be used to sample over all parameters (conditional distributions are Gaussian in the absence of selection effects. Including them introduces additional accept/reject step).

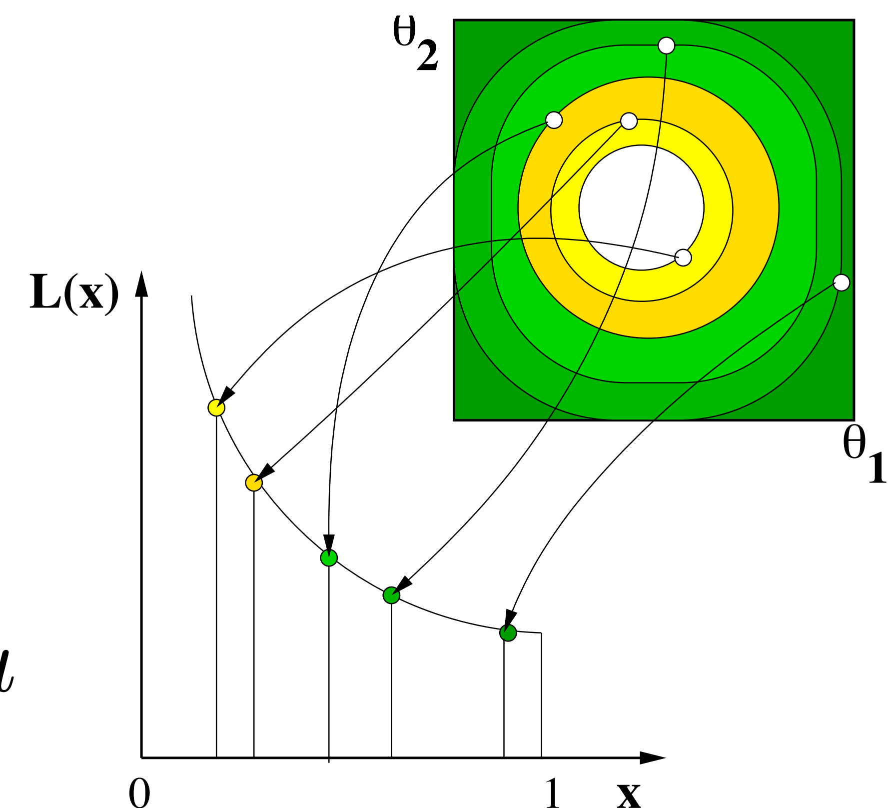
The Nested Sampling algorithm

Imperial College London

- Skilling (2006) introduced Nested Sampling as an algorithm originally aimed at the efficient computation of the model likelihood (Skilling, 2006).
- The idea is to map a multi-dimensional integral onto a 1D integral which is easy to compute numerical
- **The method requires to sample uniformly from the fraction of the prior volume X(μ) above the iso-likelihood level μ**

$$X(\mu) = \int_{L(\theta)>\mu} P(\theta) d\theta$$

$$P(d) = \int d\theta L(\theta) P(\theta) = \int_0^1 X(\mu) d\mu$$

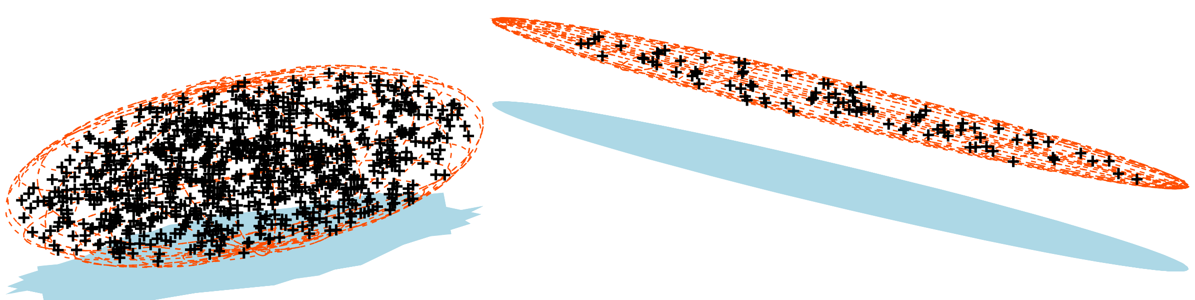


Feroz et al (2008), *arxiv: 0807.4512*
Trotta et al (2008), *arxiv: 0809.3792*

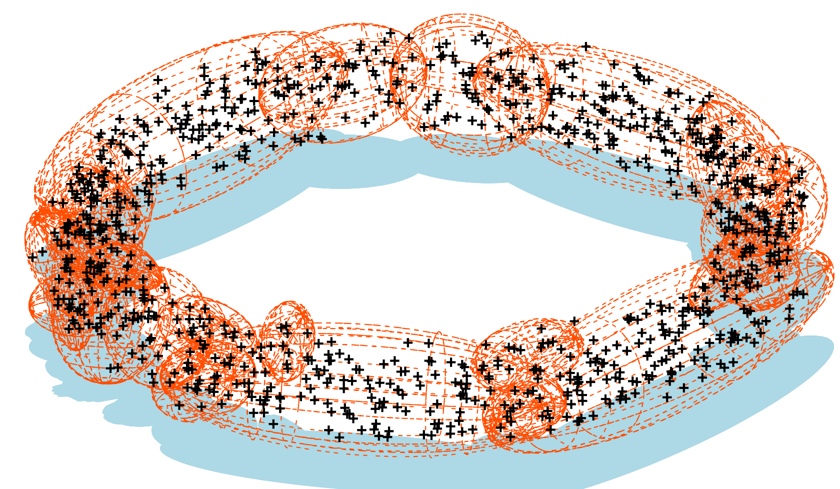
The MultiNest ellipsoidal sampling

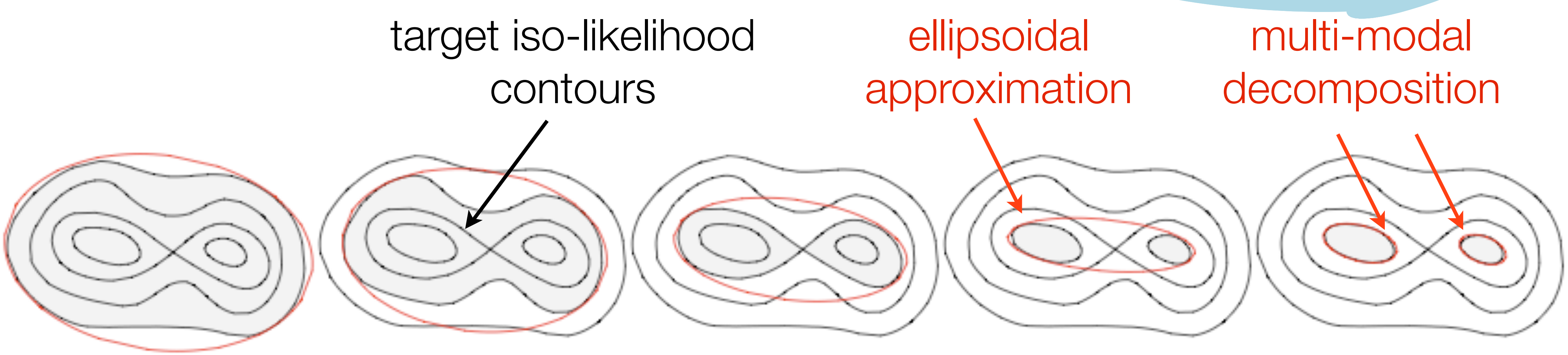
- The MultiNest algorithm (Feroz & Hobson, 2007, 2008) uses a multi-dimensional ellipsoidal decomposition of the remaining set of "live points" to approximate the prior volume above the target iso-likelihood contour.

Multimodal likelihood



Highly degenerate likelihood



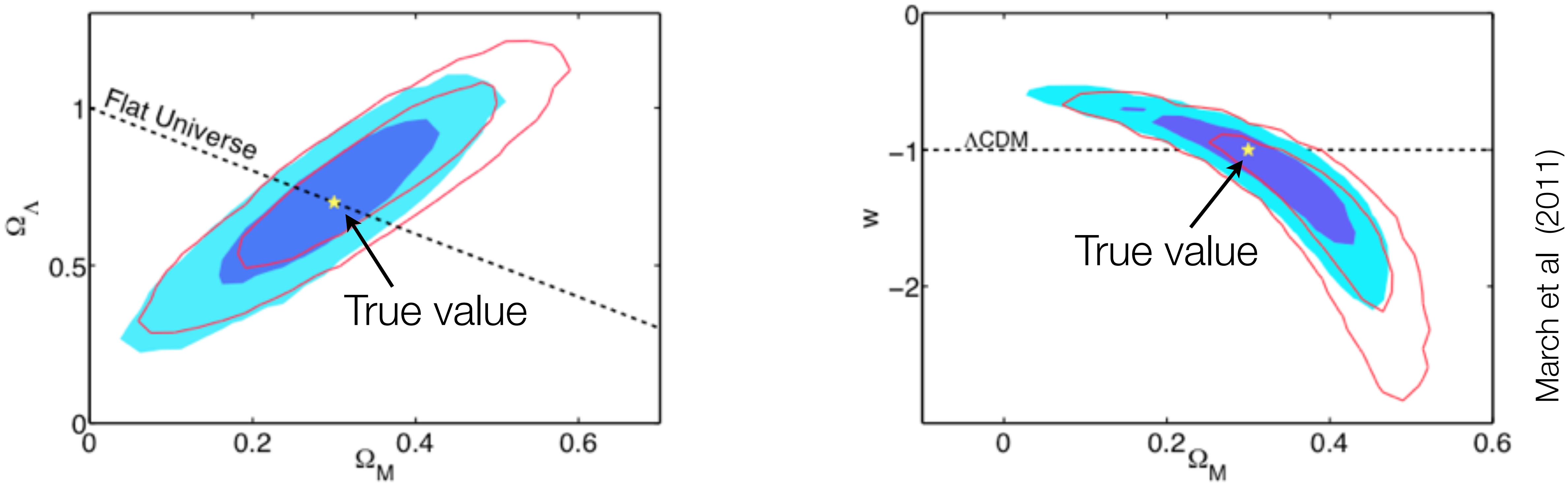


Decreasing prior fraction X

Marginal posterior (simulated data)

Imperial College London

$w = 1$ $\Omega_K = 0$



Red/empty: Chi^2 (68%, 95% CL)

Blue/filled: Bayesian (68%, 95% credible regions)

Bayesian posterior is noticeably different from the Chi^2 CL: which one is “best”?

ICIC