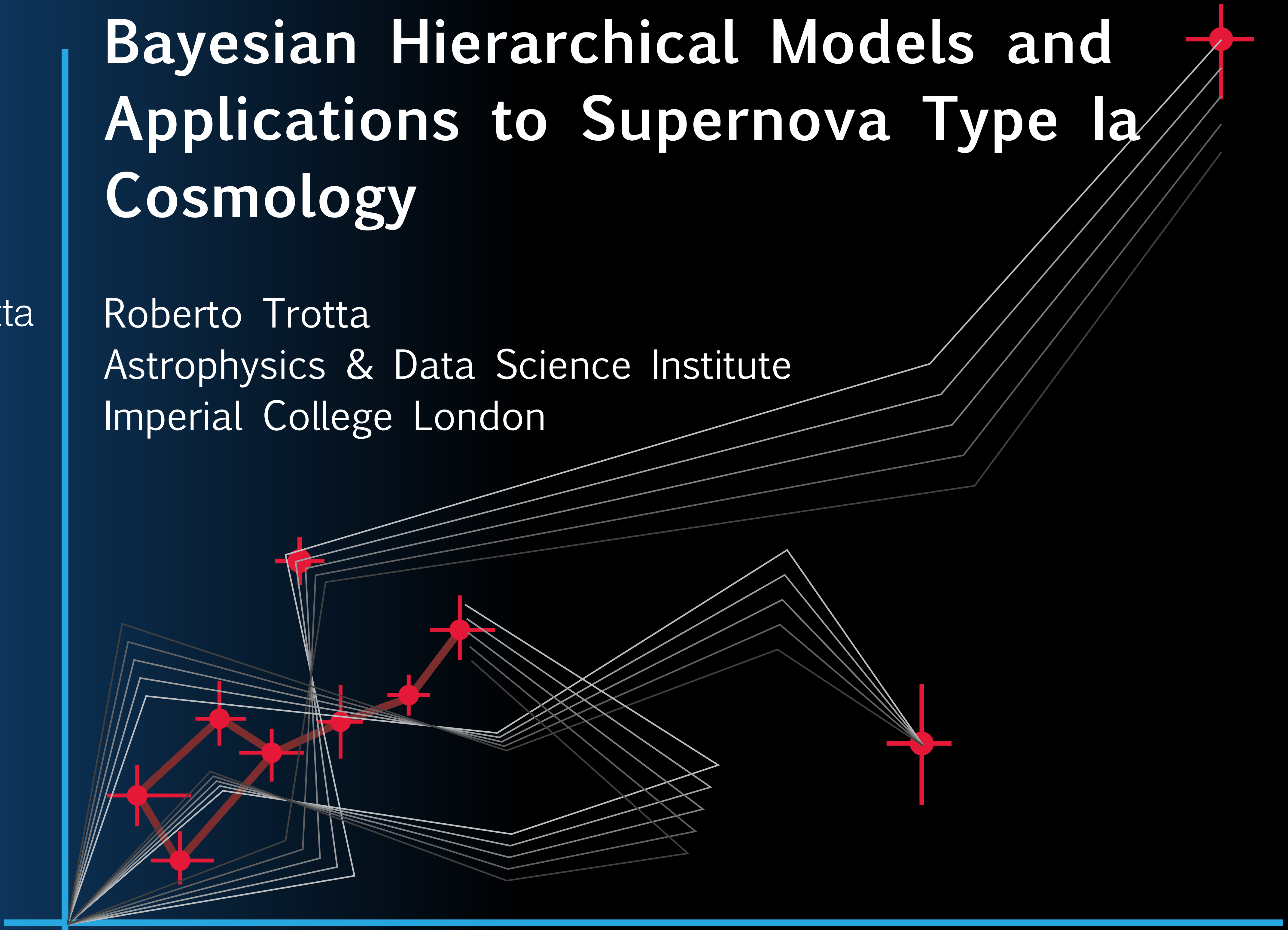
Bayesian Hierarchical Models and Applications to Supernova Type Ia Cosmology

@R_Trotta

Roberto Trotta
Astrophysics & Data Science Institute
Imperial College London



ICIC
Imperial Centre
for Inference & Cosmology

www.robertotrotta.com

Imperial College
London

The cosmological concordance model

The ΛCDM cosmological concordance model is built on three pillars:

1. **INFLATION:**
   A burst of exponential expansion in the first ~10^{-32} s after the Big Bang, probably powered by a yet unknown scalar field.
2. **DARK MATTER:**
   The growth of structure in the Universe and the observed gravitational effects require a massive, neutral, non-baryonic yet unknown particle making up ~25% of the energy density.
3. **DARK ENERGY:**
   The accelerated cosmic expansion (together with the flat Universe implied by the Cosmic Microwave Background) requires a smooth yet unknown field with negative equation of state, making up ~70% of the energy density.

The FRW Universe

The simplest model is that of a spatially flat, homogeneous and isotropic expanding Universe (the FRW model):

$$ds^2 = c^2 dt^2 - a^2(t)\left[dx^2 + dy^2 + dz^2\right]$$

where a(t) is the **"scale factor"**.

Light from distant objects is **"redshifted"**:

$$\frac{\lambda_{\rm obs}}{\lambda_0} = 1 + z$$

Redshift gives the amount by which the scale factor has grown:

$$1 + z = \frac{1}{a}$$

(normalization: a = 1 today, when z=0)

The cosmological parameters

Imperial College London

General Relativity gives a differential equation (Friedmann Equation) for the evolution of the scale factor as a function of the “cosmological parameters”:

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left(\frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \frac{\Omega_\kappa}{a^2} + \Omega_\Lambda\right)$$

Cosmological parameters:

Ω_m: Matter parameter
Ω_r: Radiation parameter
Ω_κ: Curvature parameter (= 0 for a flat Universe)
Ω_Λ: Cosmological constant parameter

H_0 (km/s/Mpc): Hubble parameter

(further parameters describe the initial conditions)

Goal:
to observationally determine the cosmological parameters

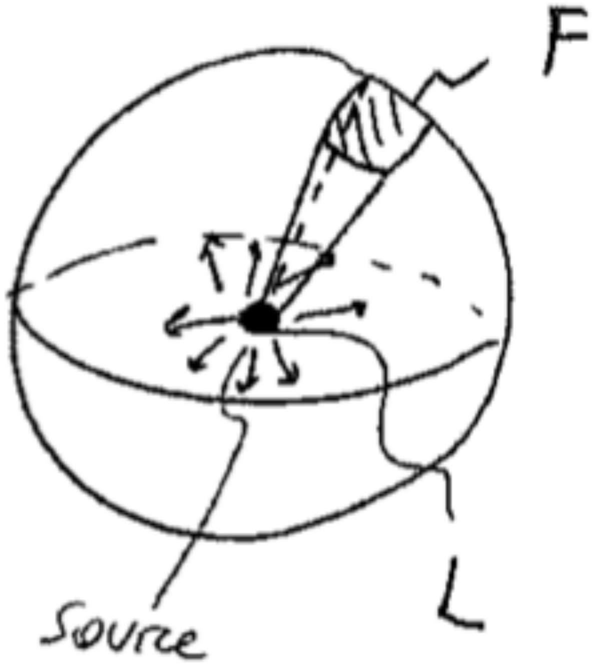
Distances in the Universe

Two different (observationally based) definitions for “distance”:

LUMINOSITY DISTANCE

d_L

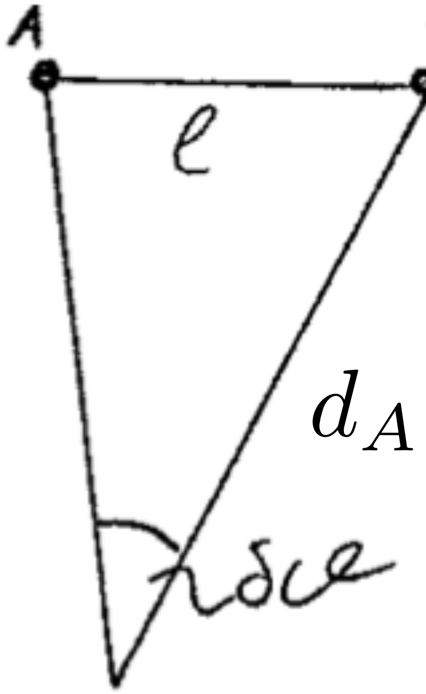
$$F = \frac{L}{4\pi d_L^2(z)}$$



ANGULAR DIAMETER DISTANCE

d_A

$$\delta\vartheta = \frac{\ell}{d_A(z)}$$



For any metric theory of gravity: $d_L(z) = (1+z)^2 d_A(z)$

Distance-redshift relation

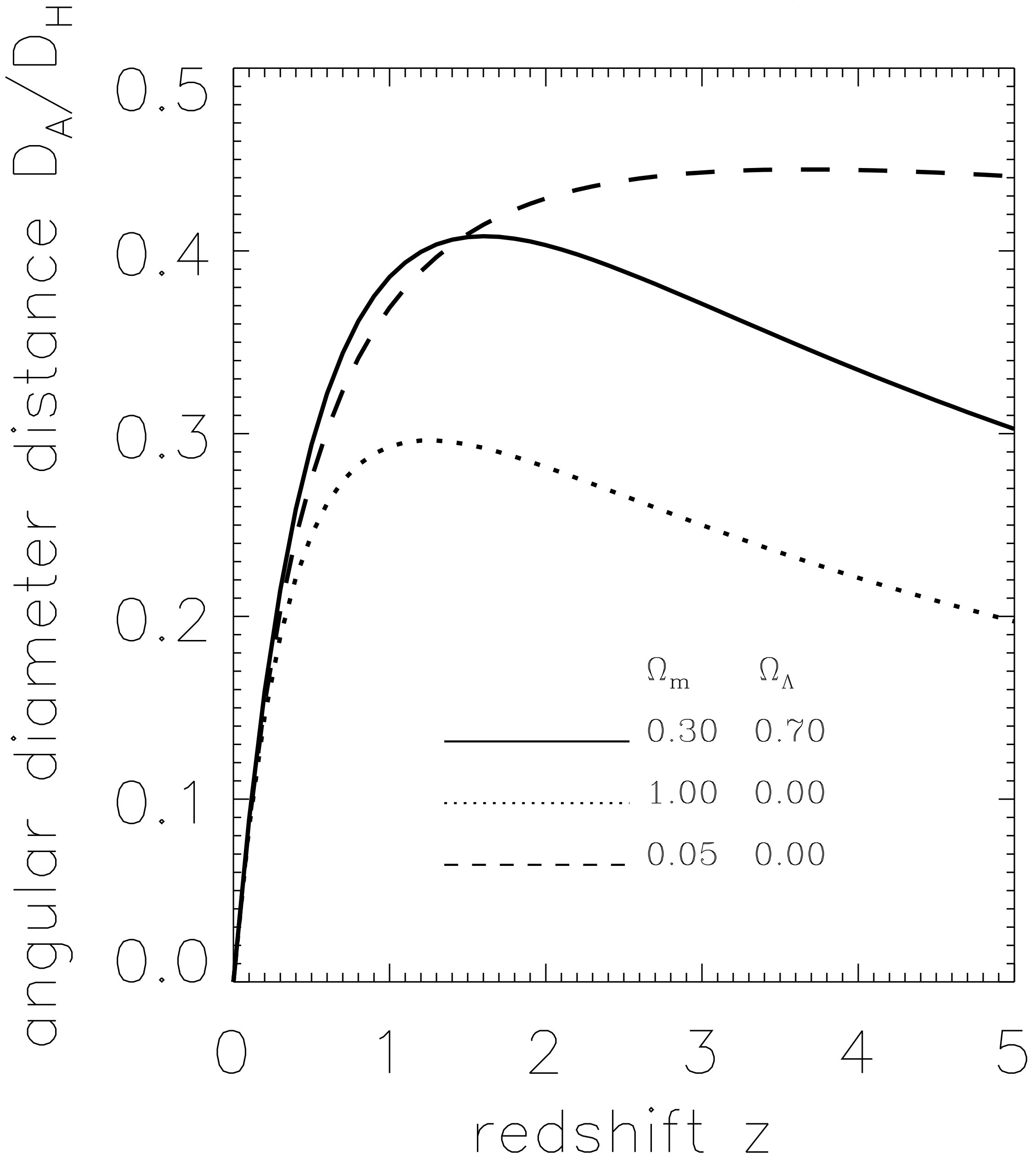
The distance-redshift relation depends on the cosmological parameters

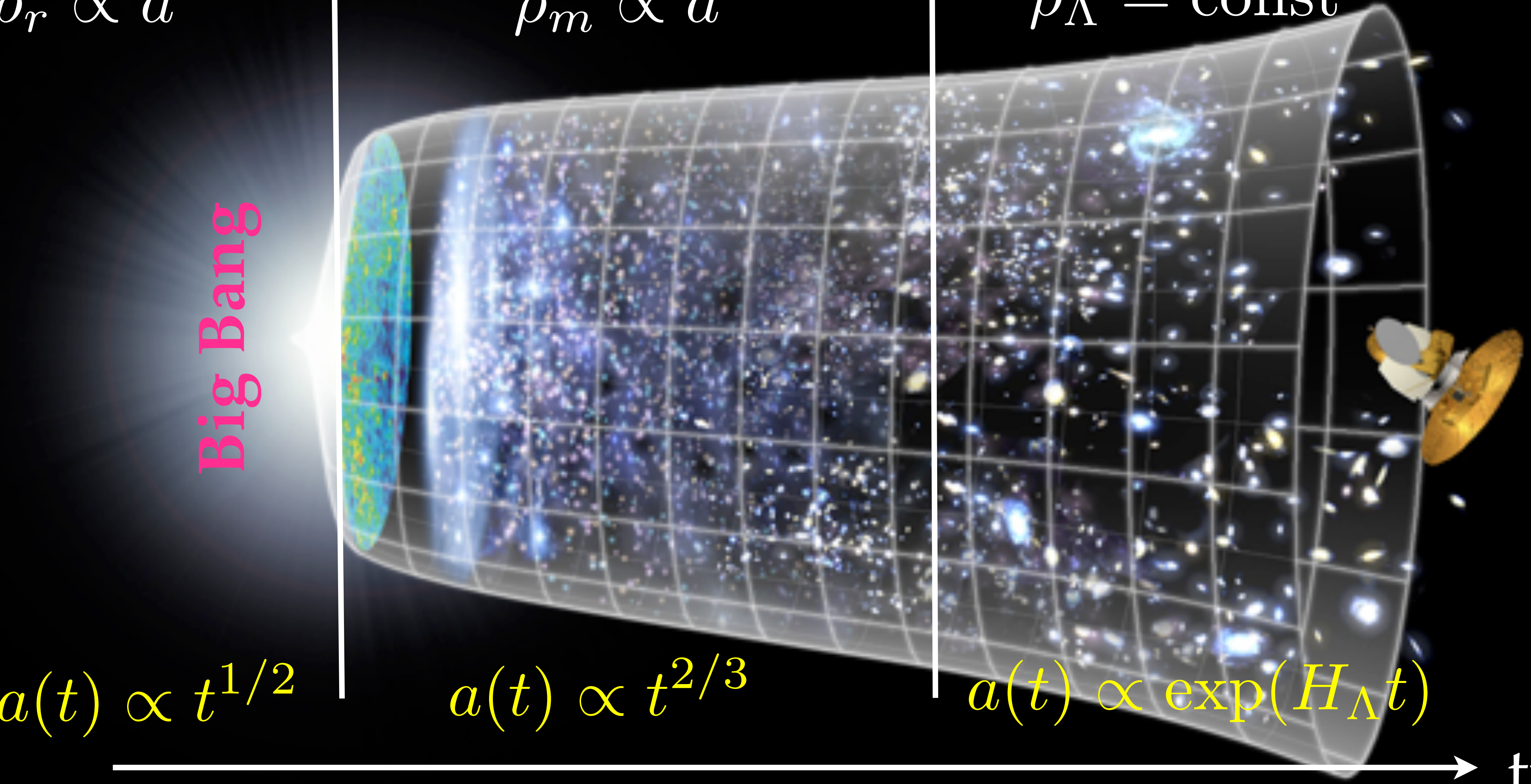
Strategy:
measure redshift (~"easy") and distance (hard), to infer the cosmological parameters controlling the redshift-distance relationships

In this example:

the matter content: Ω_m
the dark energy content: Ω_Λ
the Hubble parameter:
$H_0 = 100h$ km/s/Mpc

In units of $c/H_0 = 2997/h$ Mpc





Radiation

~50k yrs

$\rho_r \propto a^{-4}$

Dark matter

~7bn yrs

$\rho_m \propto a^{-3}$

Dark energy

last ~6bn yrs

$\rho_\Lambda = \mathrm{const}$

Big Bang

TODAY

$a(t) \propto t^{1/2}$

$a(t) \propto t^{2/3}$

$a(t) \propto \exp(H_\Lambda t)$

time

~ 10

~ 1.5

Supernovae Type Ia

redshift

1,100

~ 0.7

0

z

NASA/WMAP Science

Precision cosmology

Imperial College London

Comparison of theoretical model to CMB+BAO+SNIa data gives sub-percent accuracy on Cosmological Parameters:

$$\Omega_m = 0.316 \pm 0.009$$

$$H_0 = 67.8 \pm 0.9 \text{ km/s/Mpc}$$

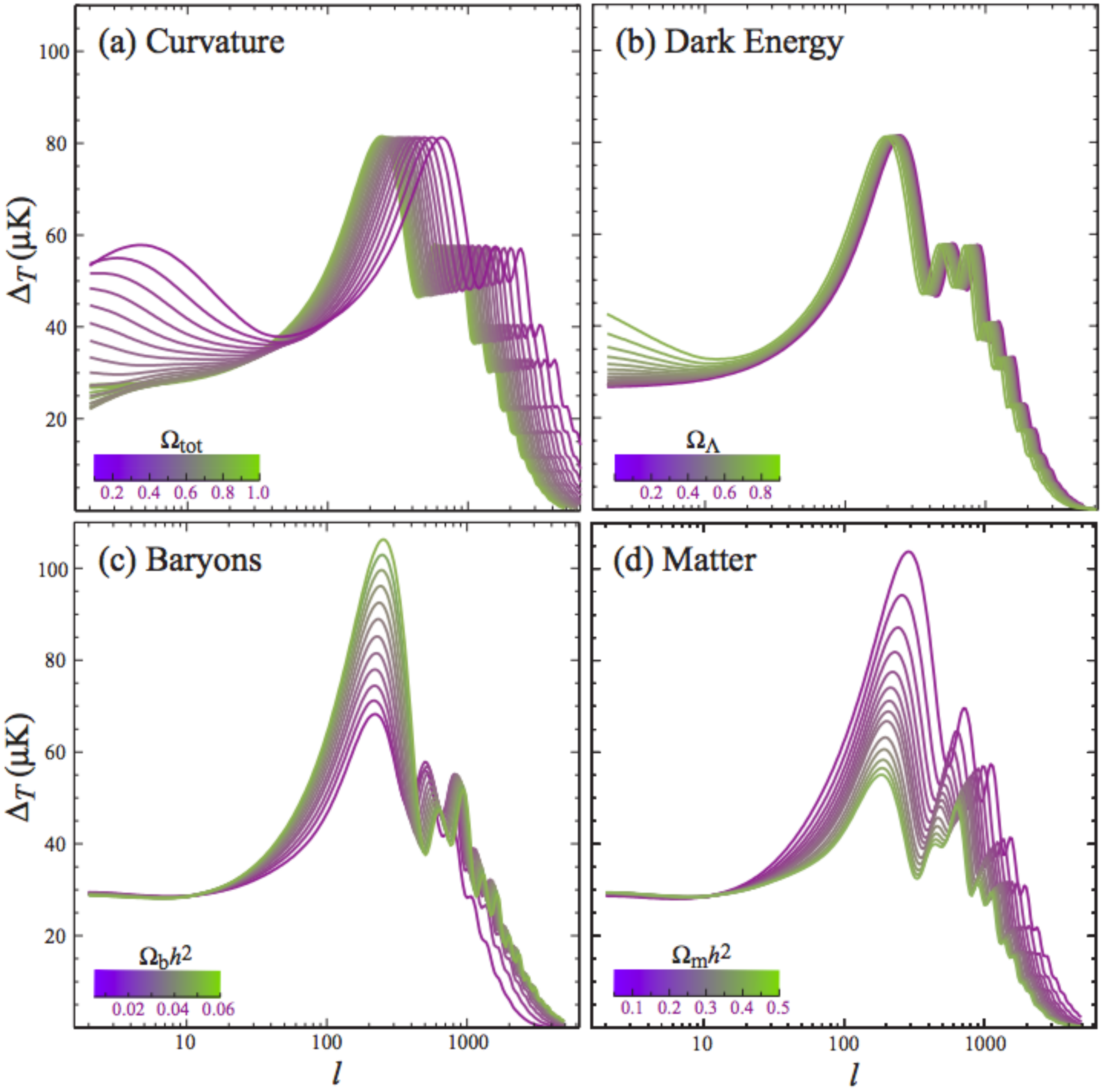
$$\Omega_\Lambda = 0.684 \pm 0.009$$

$$|\Omega_\kappa| < 0.005$$

$$\text{Age } = 13.796 \pm 0.029 \text{ Gyr}$$

Planck Collaboration (2015)

How accurate is this?

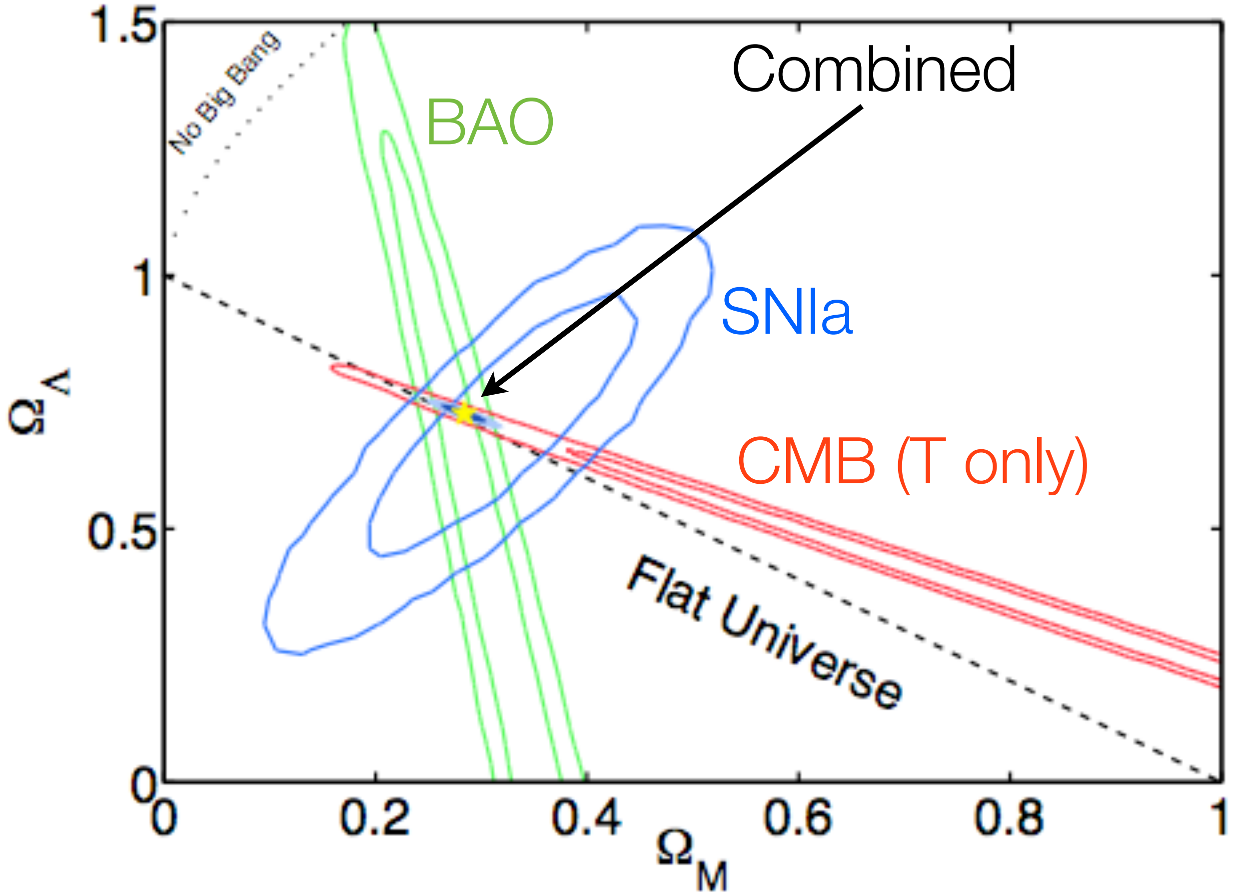


Hu & Dodelson (2002)

Cosmological constraints

Imperial College London

Combination of multiple probes **is required** to determine the cosmological parameters to high precision:
Baryonic Acoustic Oscillations (BAO) and Cosmic Microwave Background (CMB) are standard rulers.



Ref: *March et al, MNRAS 418(4):2308-2329, (2011)*

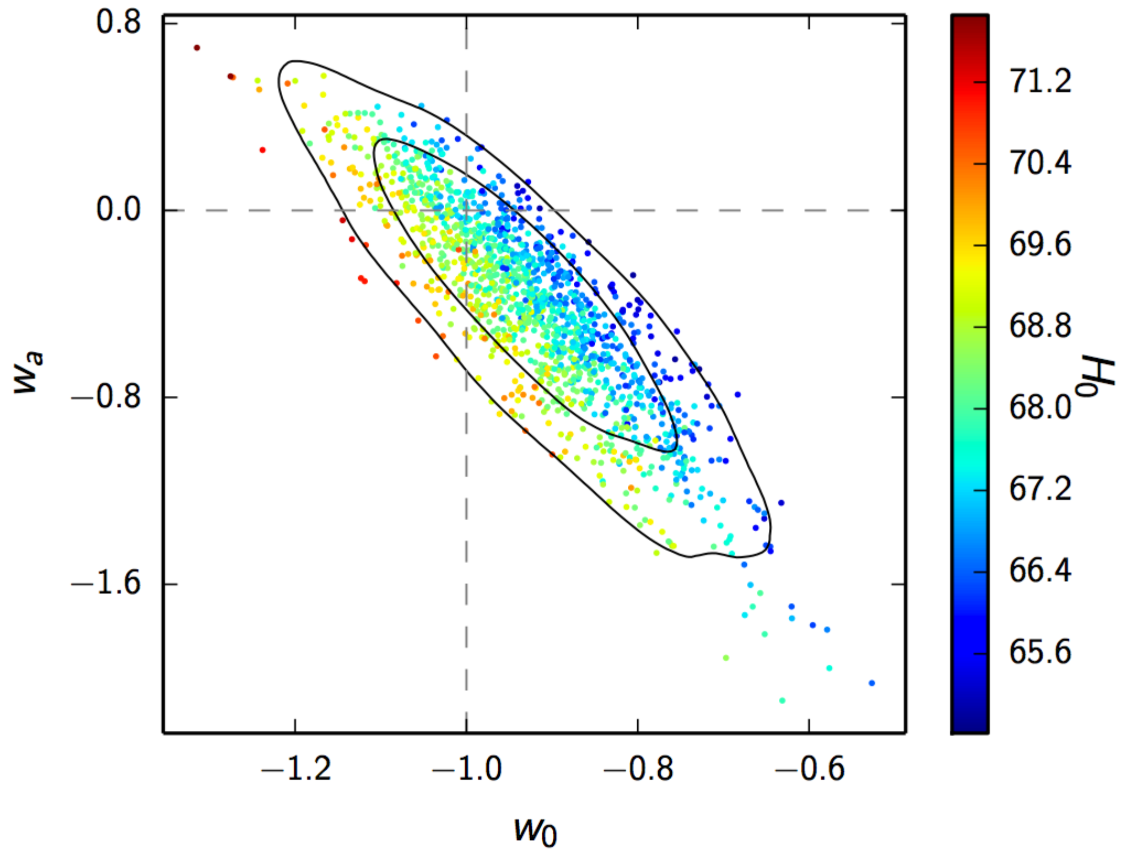
$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho(t) + 3\frac{p}{c^2}\right)$$

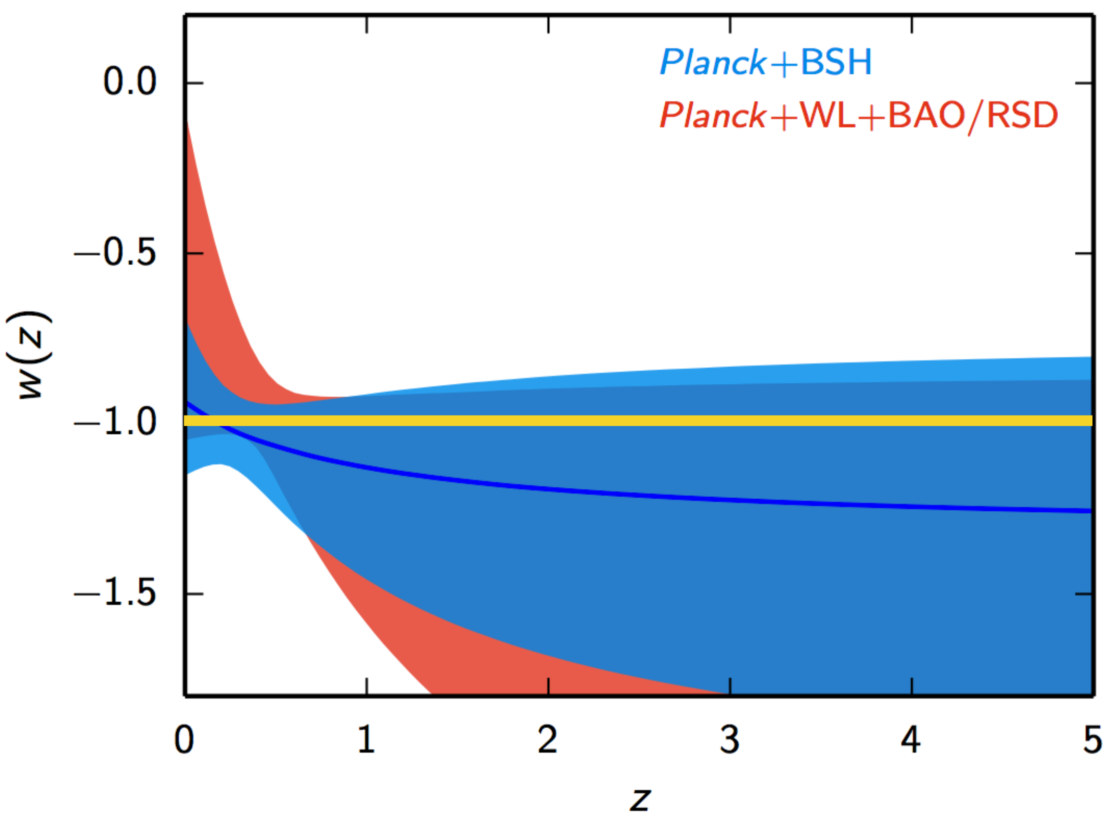
$$\ddot{a} > 0 \text{ for } w = \frac{p}{\rho c^2} < -\frac{1}{3}$$

w = -1 (cosmological constant): accelerating expansion

Constraints on dark energy

Constraints on the parameter (w_0, w_a) for the dark energy equation of state (w = P/rho) parameterization: $w(z) = w_0 + z/(1+z)\, w_a$
The cosmological constant corresponds to
(w_0, w_a) = (-1, 0) and has w(z) = const

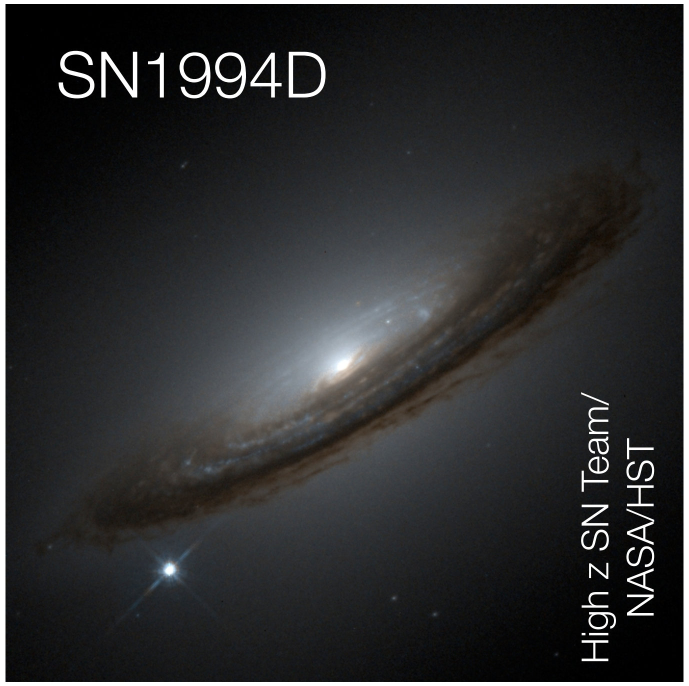




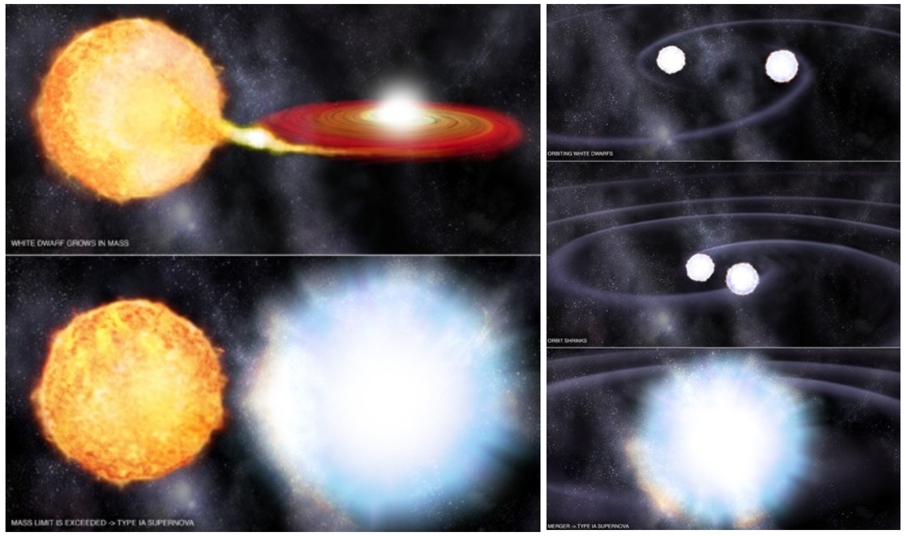
Left: Planck Collaboration 2015: Cosmological parameters, arXiv: 1502.01589.
Right: Planck Collaboration 2015: Dark energy and modified gravity, arXiv: 1502.01590

Type Ia supernovae

- **CO White Dwarf (WD):** compact CO remnant of star ($M < 2.5 M_{Sun}$), supported by electron degeneracy pressure. $M \sim 0.6 M_{Sun}$, radius ~ Earth, density ~ 1 tonne/cm^3
- **Supernovae type Ia (SNIa):** No H, Si lines. Probably the runaway thermonuclear explosion of a CO WD accreting mass approaching the Chandrasekhar limit (~1.4 solar masses), thus igniting C fusion. "Standard" candles?
- **Progenitor models:**
  Single Degenerate (WD + Main sequence or Red giant or a He star companion) vs
  Double Degenerate (WD + WD merger)
- Possibly, a mixture of both



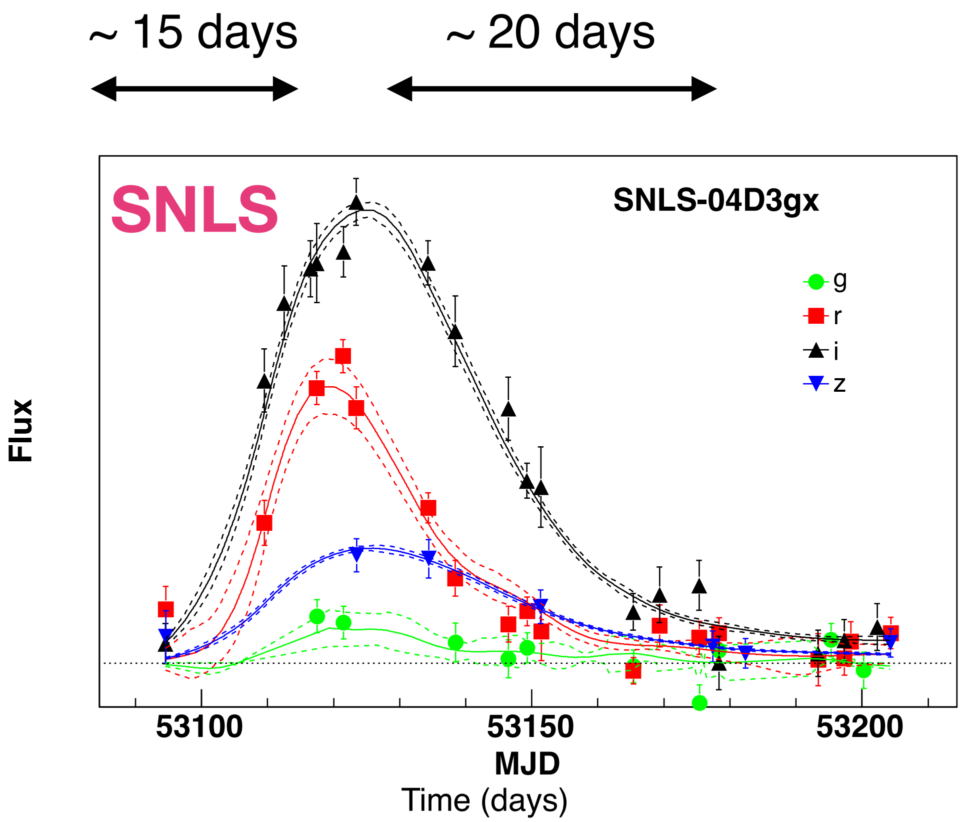
High z SN Team/ NASA/HST



NASA/CXC/M. Weiss

@R_Trotta

Light-curves

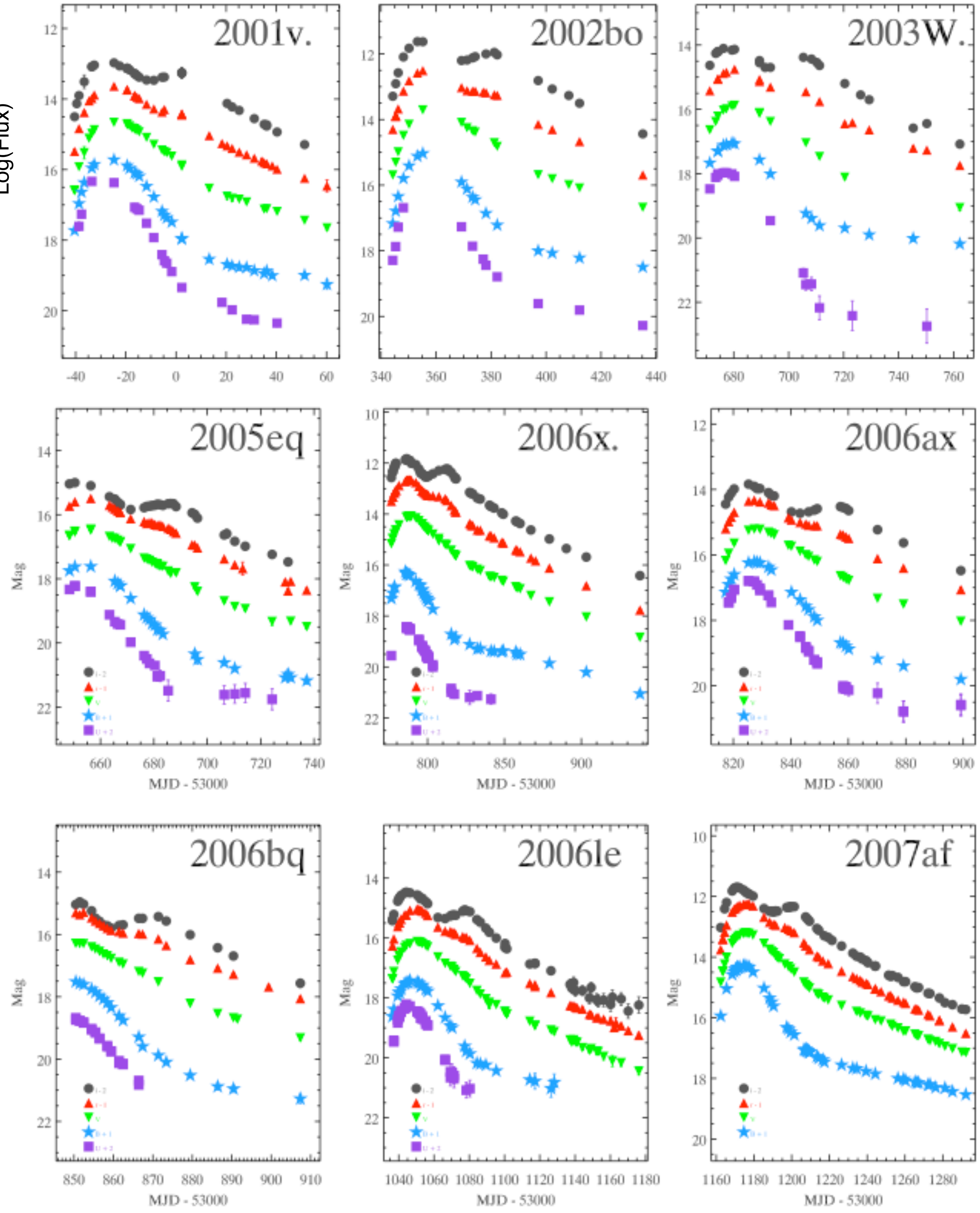


Guy et al (2007)

Optical filters ("bands"):

B = 450 nm
g = 520 nm
r = 670 nm
i = 790 nm
z = 910 nm

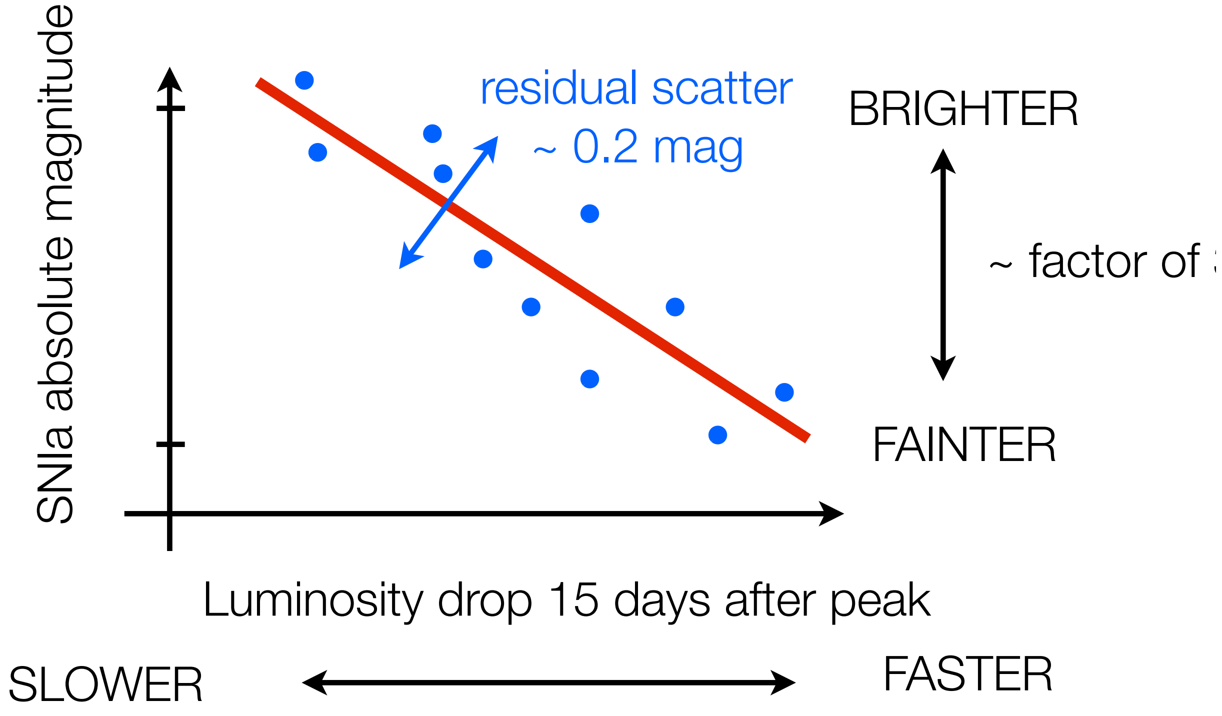
CfA3
185 multi-band optical nearby SNIa

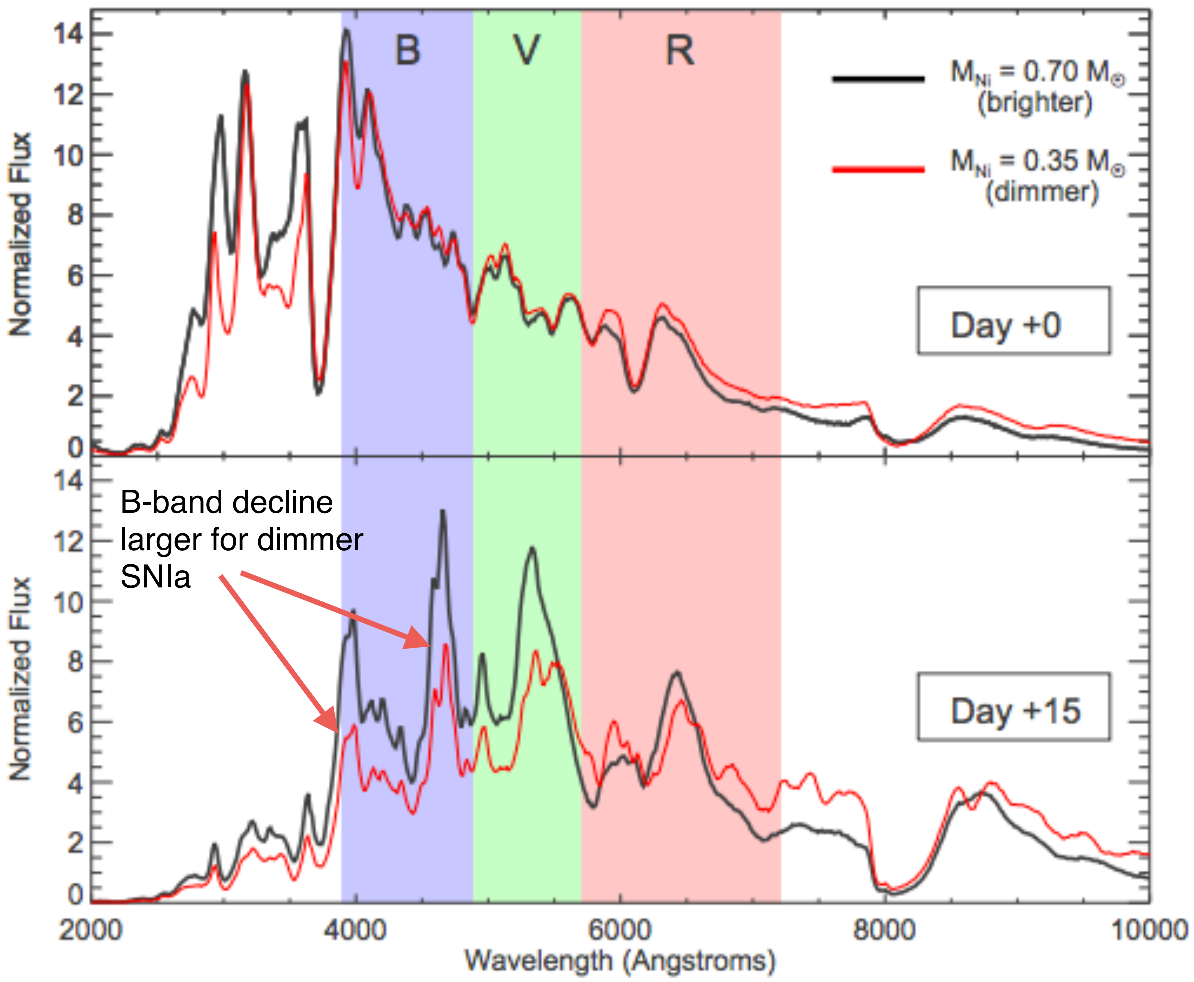


Hicken et al (2009)

Brightness-width relationship

Peak magnitude scatter can be reduced by exploiting phenomenological correlations with the shape (and colour) of the LC (Phillips93,Tripp98, Riess+96,...)





B-band decline larger for dimmer SNIa

Kasen & Woosley 2006

Brighter SNIa are slow decliners

Dimmer SNIa are cooler → Earlier recombination → earlier colour transfer from B-band to red via line blanketing of iron-group elements → faster B-band decline

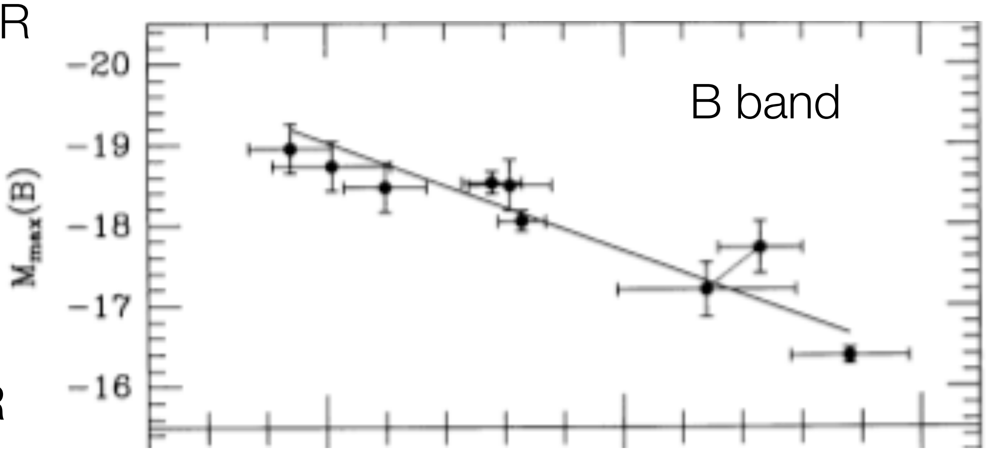
"Stretch correction"

"Brighter SNIa are slow decliners"

BRIGHTER

↕

FAINTER



Drop from peak after 15 days

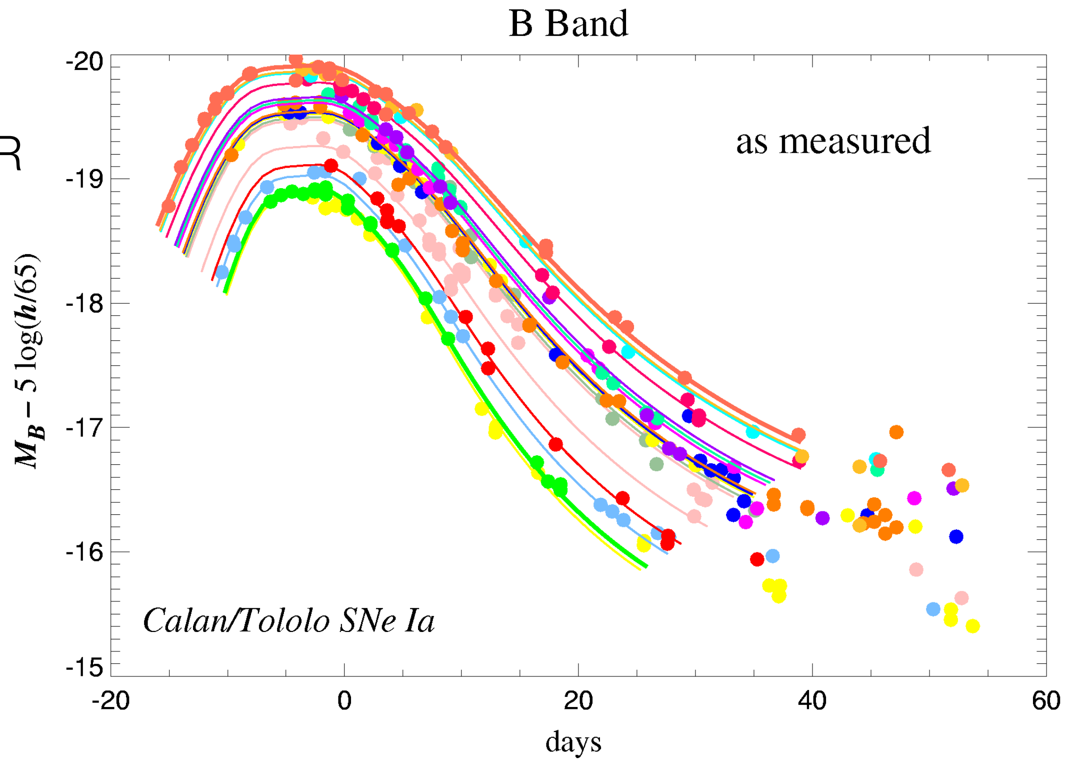
SLOWER ⟷ FASTER

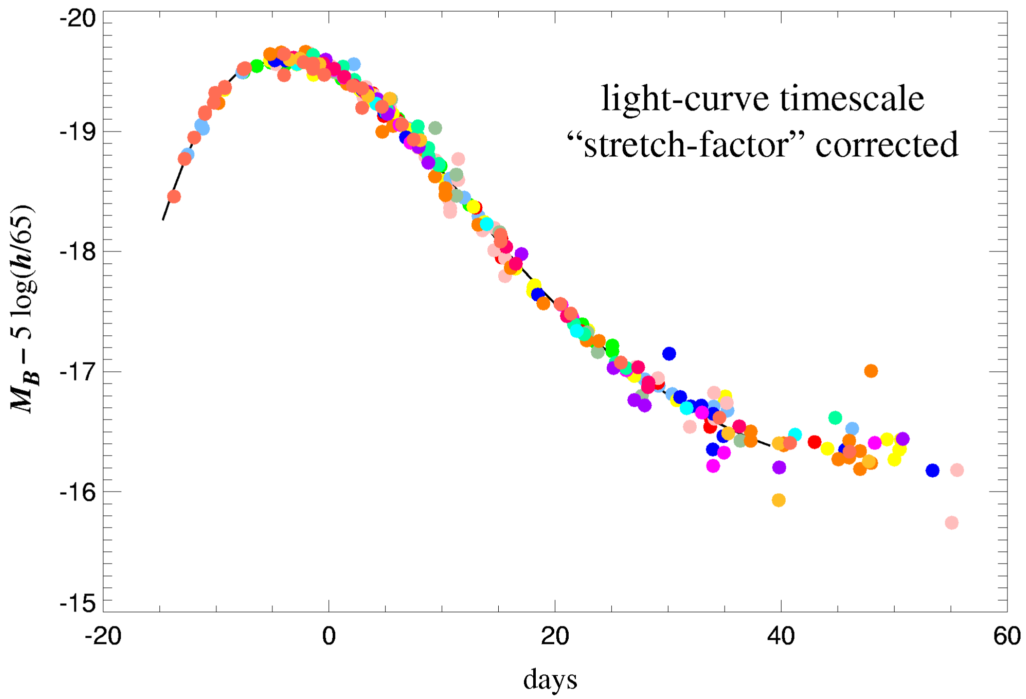
Phillips, ApJ 413 (1993) L105-L108

BRIGHTER

↕

FAINTER

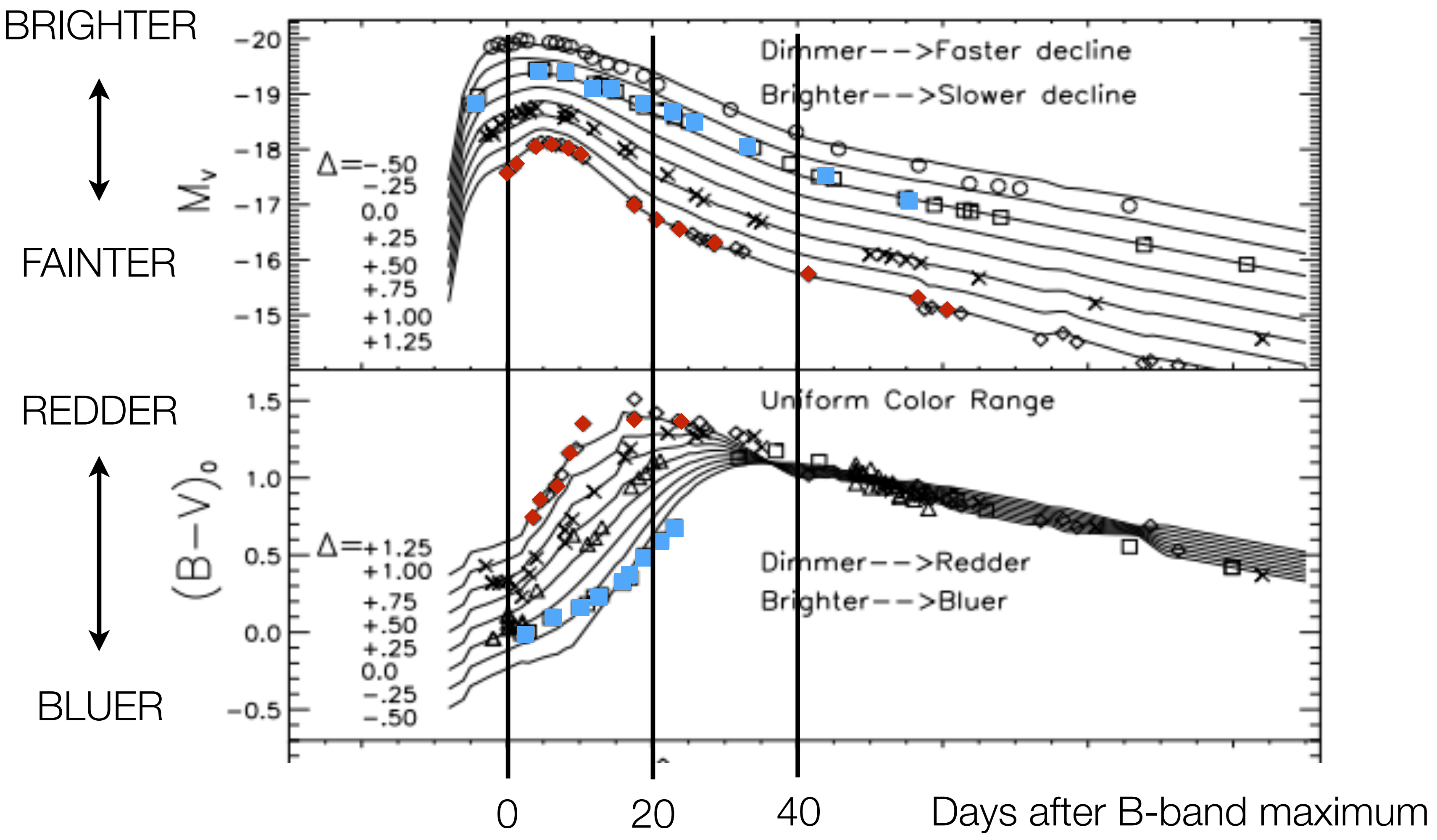




Kim, *et al.* (1997)

"Color correction"

"Brighter SNIa are bluer in colour"



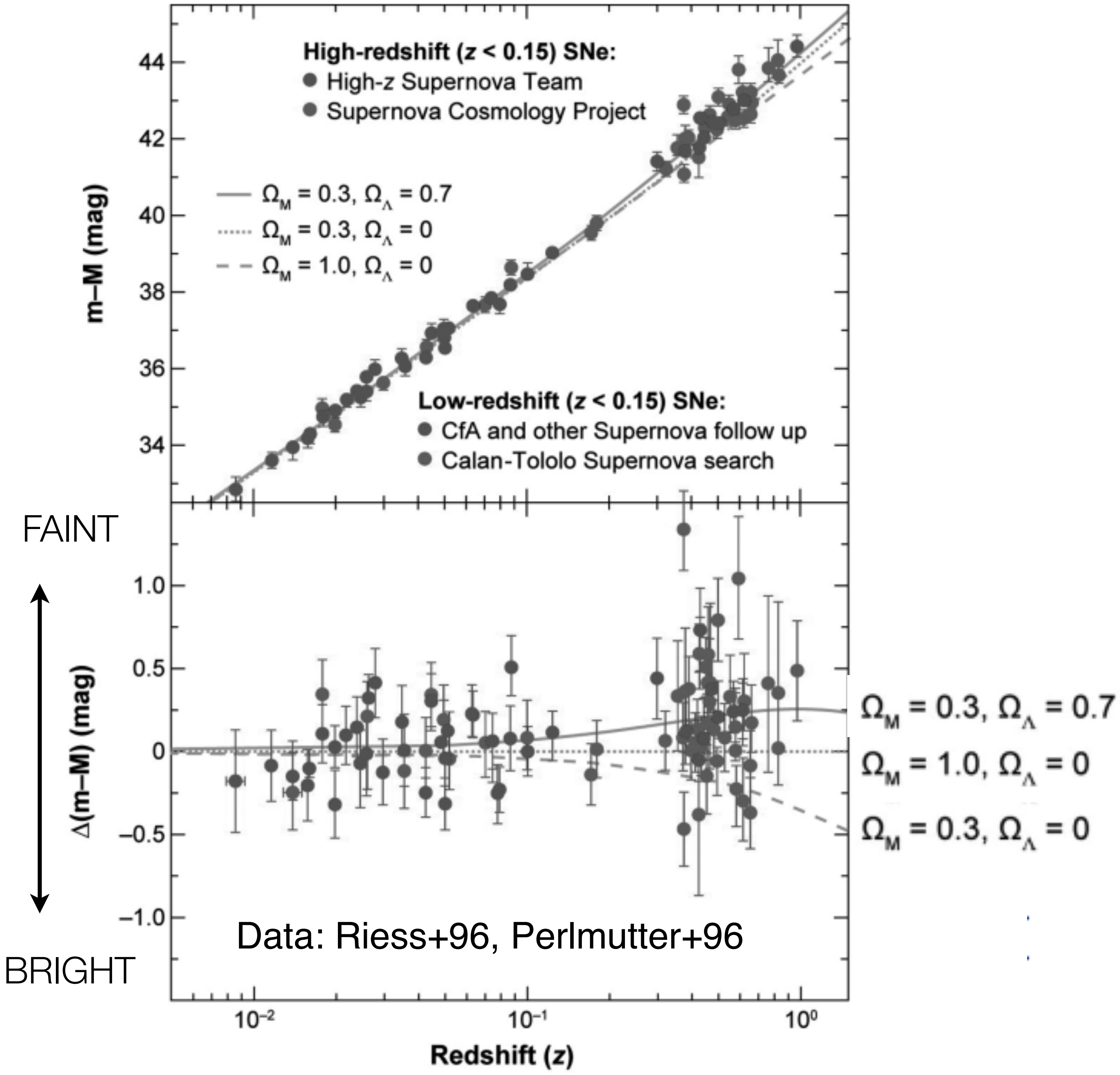
Higher host galaxy dust content → larger extinction (dimmer) → larger transfer to longer wavelength→ redder

Typical values for the slope of the dust absorption law with wavelength ($R_V \sim 1.1$-2.5) are in tension with values for Milky Way dust ($R_V = 3.1$)

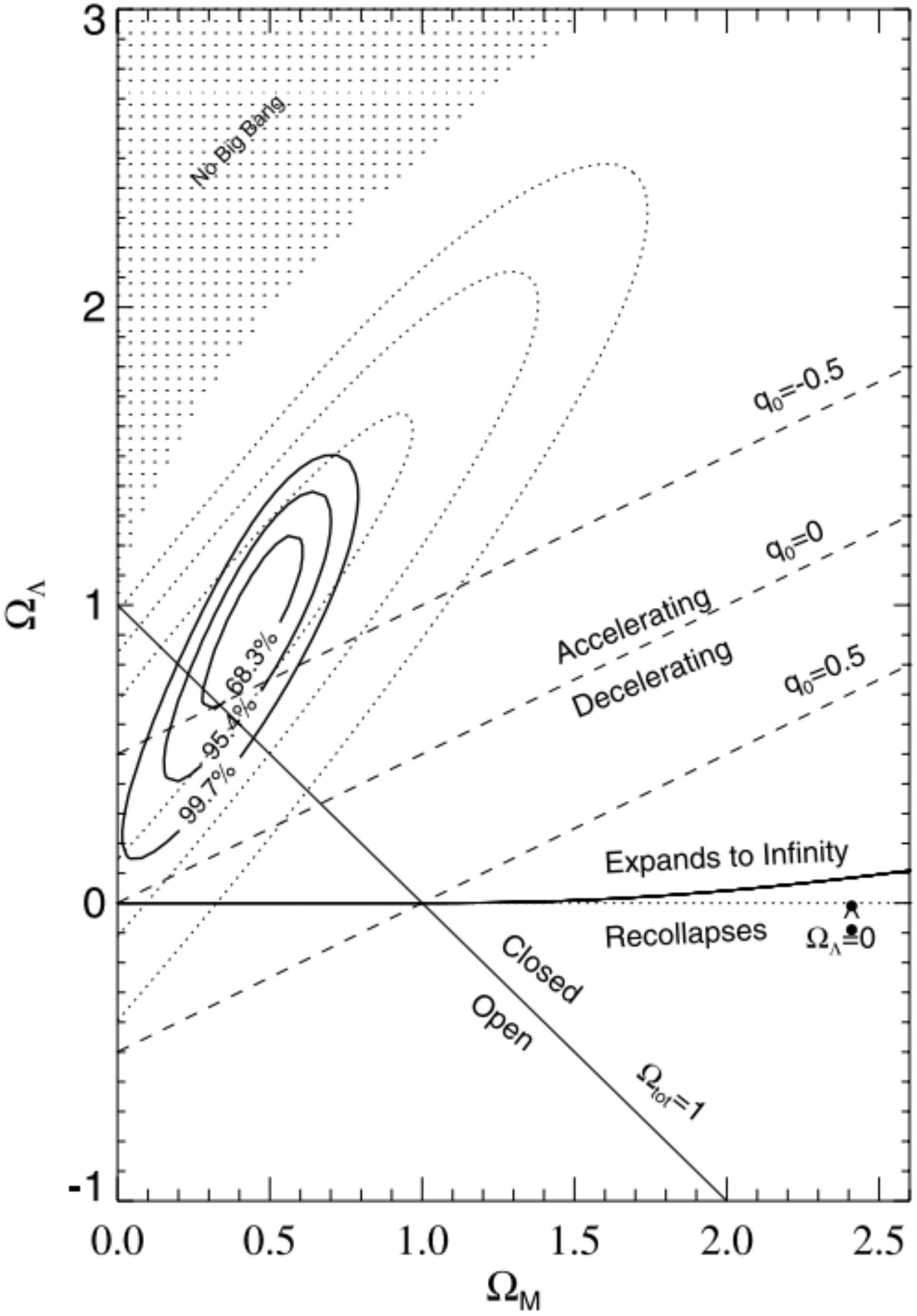
Mandel et al (1609.04470) show that this is due to the colour correction not being exactly linear. Idea: split intrinsic colour variability from dust reddening/dimming

Riess+1996

Cosmic acceleration



Frieman+08

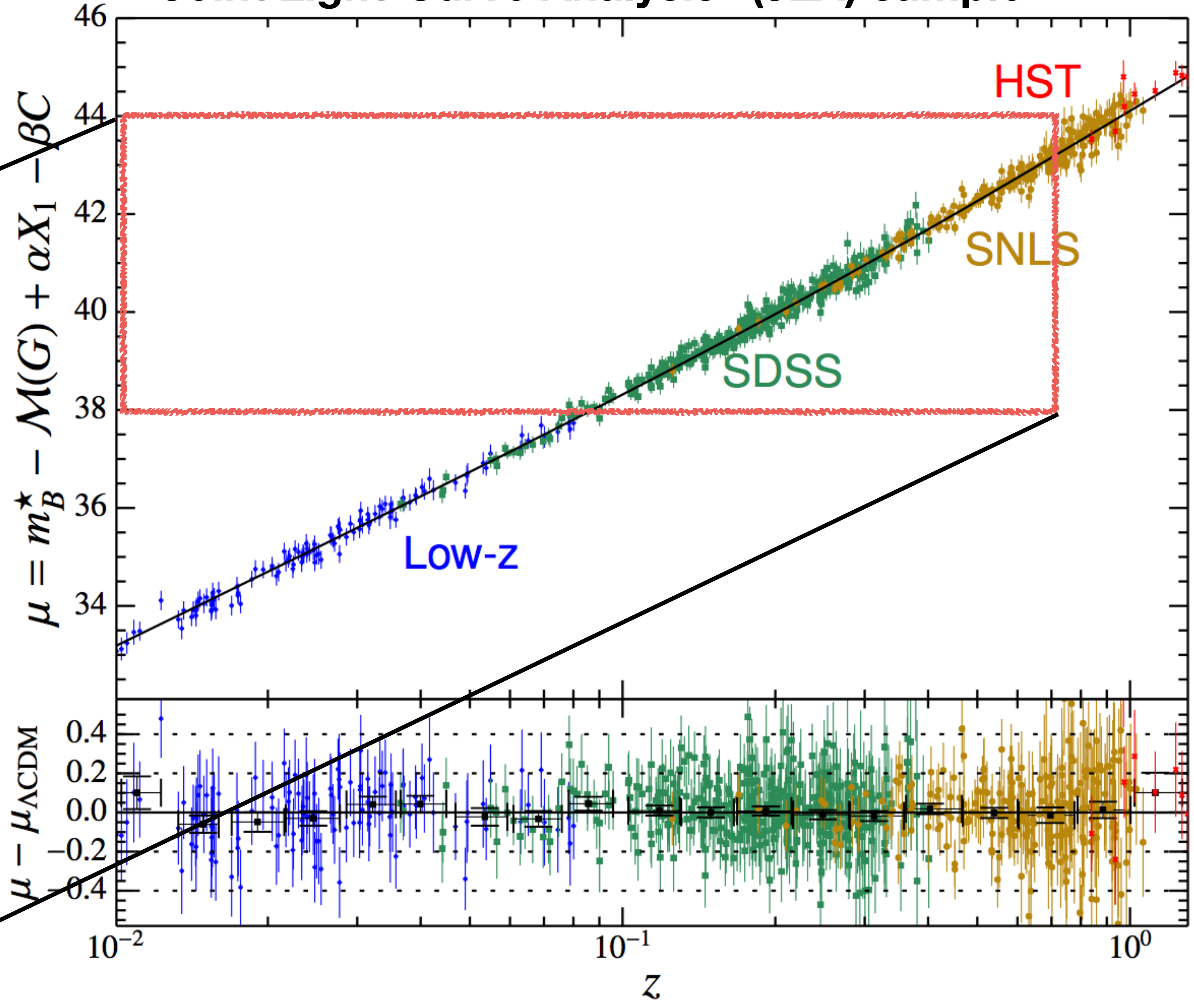


Riess et al, ApJ, 607:665-687 (2004)

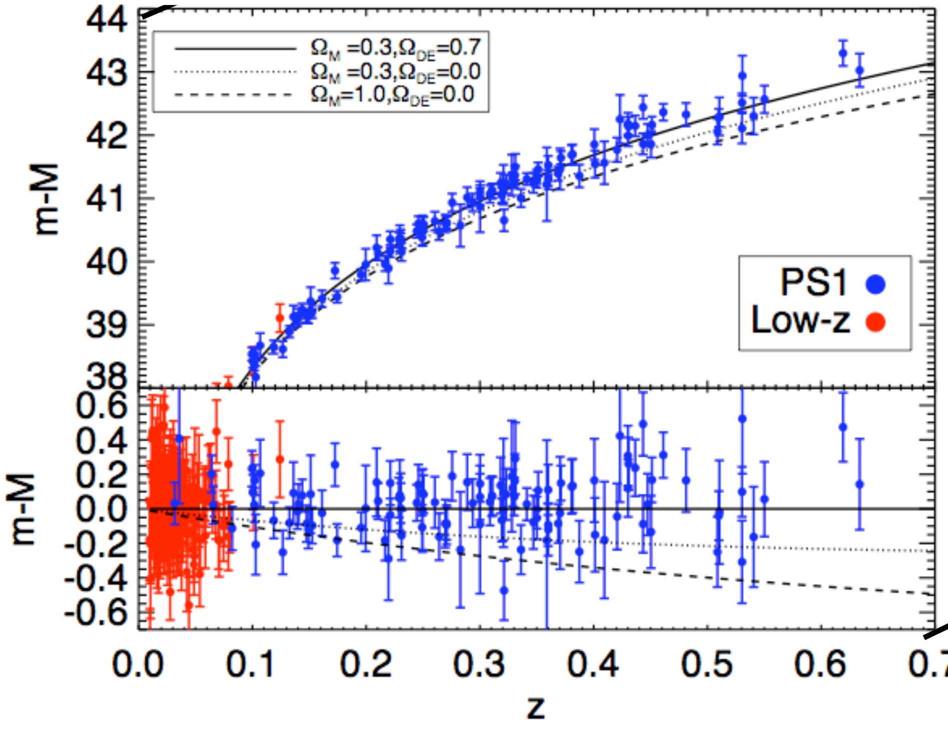
SNIa sample

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

"Joint Light-Curve Analysis" (JLA) sample



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



Cosmological fit

Standard analysis minimizes the Chi² between the theoretical distance modulus (left) and the observed one (after colour and stretch corrections, from SALT2):

$$\mu^{\rm theo}(z_i, \overset{\text{unknown}}{\mathcal{C}}) \quad \xleftrightarrow{\text{Chi}^2} \quad \mu_i^{\rm obs} = \underset{\overset{\uparrow}{\text{apparent mag}}}{m_i} - \overset{\overset{\text{absolute mag}}{\downarrow}}{M_i} + \overset{\text{unknown}}{\alpha}\,\underset{\overset{\uparrow}{\text{stretch}}}{x_{1,i}} - \underset{\text{unknown}}{\beta}\,\overset{\overset{\text{colour}}{\downarrow}}{c_i}$$

Error budget contains:

- Statistical errors (measurements of mag, stretch and colour corrections)
- Systematic errors (flux calibration, peculiar velocities, lensing, ...)
- "Residual dispersion" $\sigma_{\rm int}$: everything else, including intrinsic variability in SNIa

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

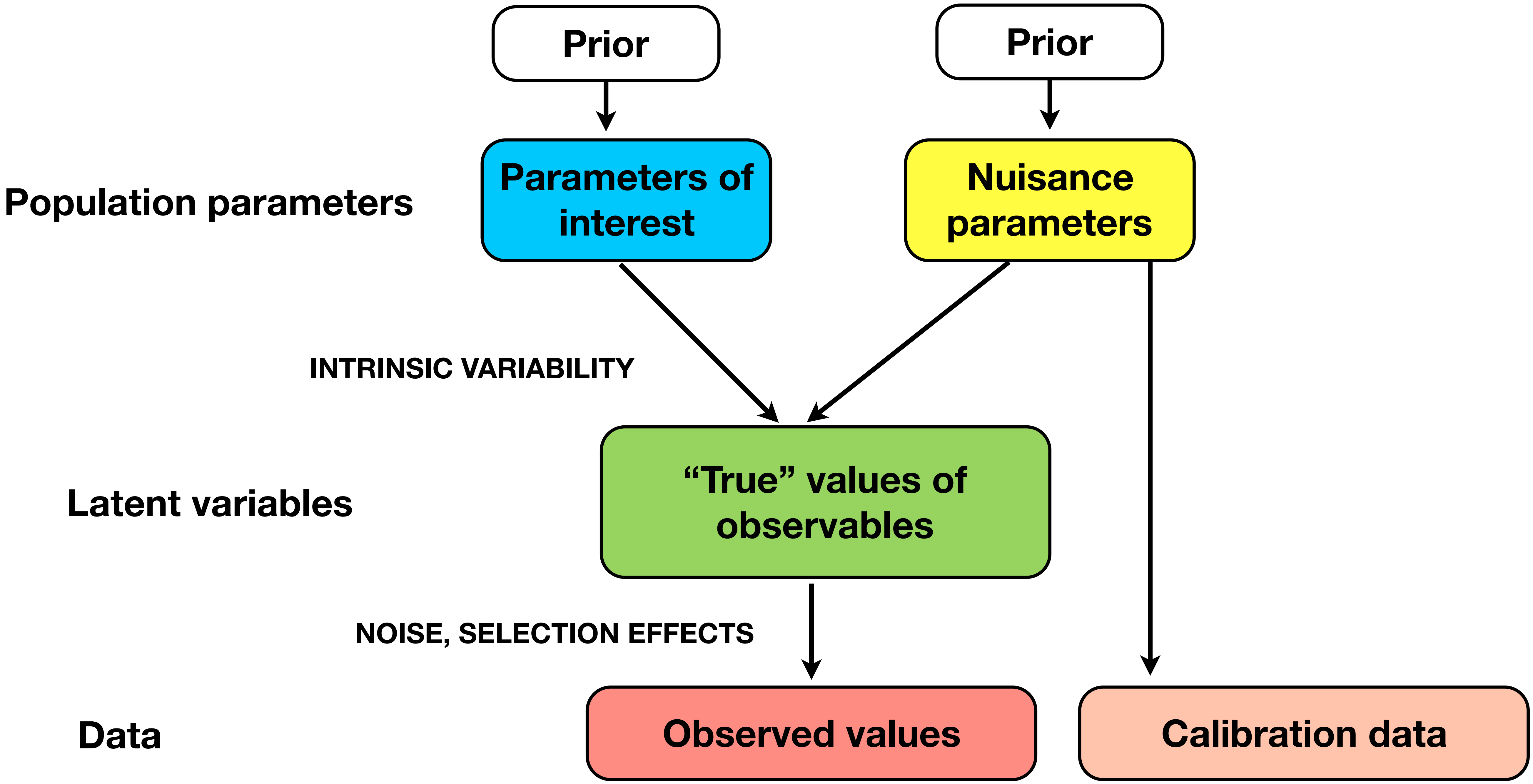
March, RT+11; Rubin+15

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

- X Flux calibrations (Betoule+14, JLA paper)
- X Selection effects (Rubin+15 for a Bayesian approach)
- X Contamination by non-Ia's (Kunz+07, Kessler & Scolnic16)
- ✓ 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

Bayesian hierarchical models



Prior

Prior

Population parameters

Parameters of interest

Nuisance parameters

INTRINSIC VARIABILITY

Latent variables

“True” values of observables

NOISE, SELECTION EFFECTS

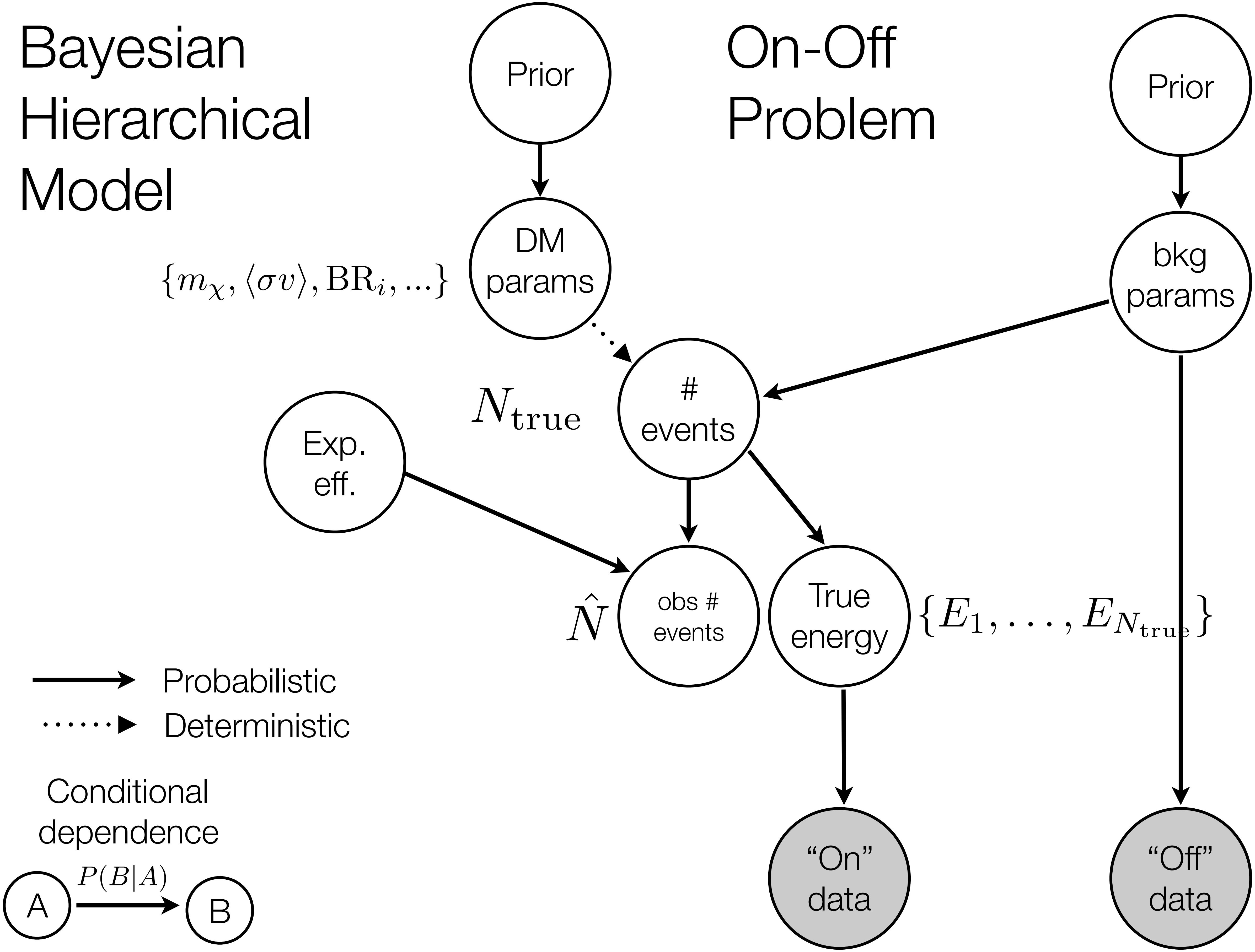
Data

Observed values

Calibration data

Bayesian Hierarchical Model

On-Off Problem



Prior

DM params

$\{m_\chi, \langle\sigma v\rangle, \mathrm{BR}_i, ...\}$

events

$N_{\mathbf{true}}$

Exp. eff.

obs # events

$\hat{N}$

True energy

$\{E_1, \ldots, E_{N_{\mathbf{true}}}\}$

"On" data

Prior

bkg params

"Off" data

Probabilistic

Deterministic

Conditional dependence

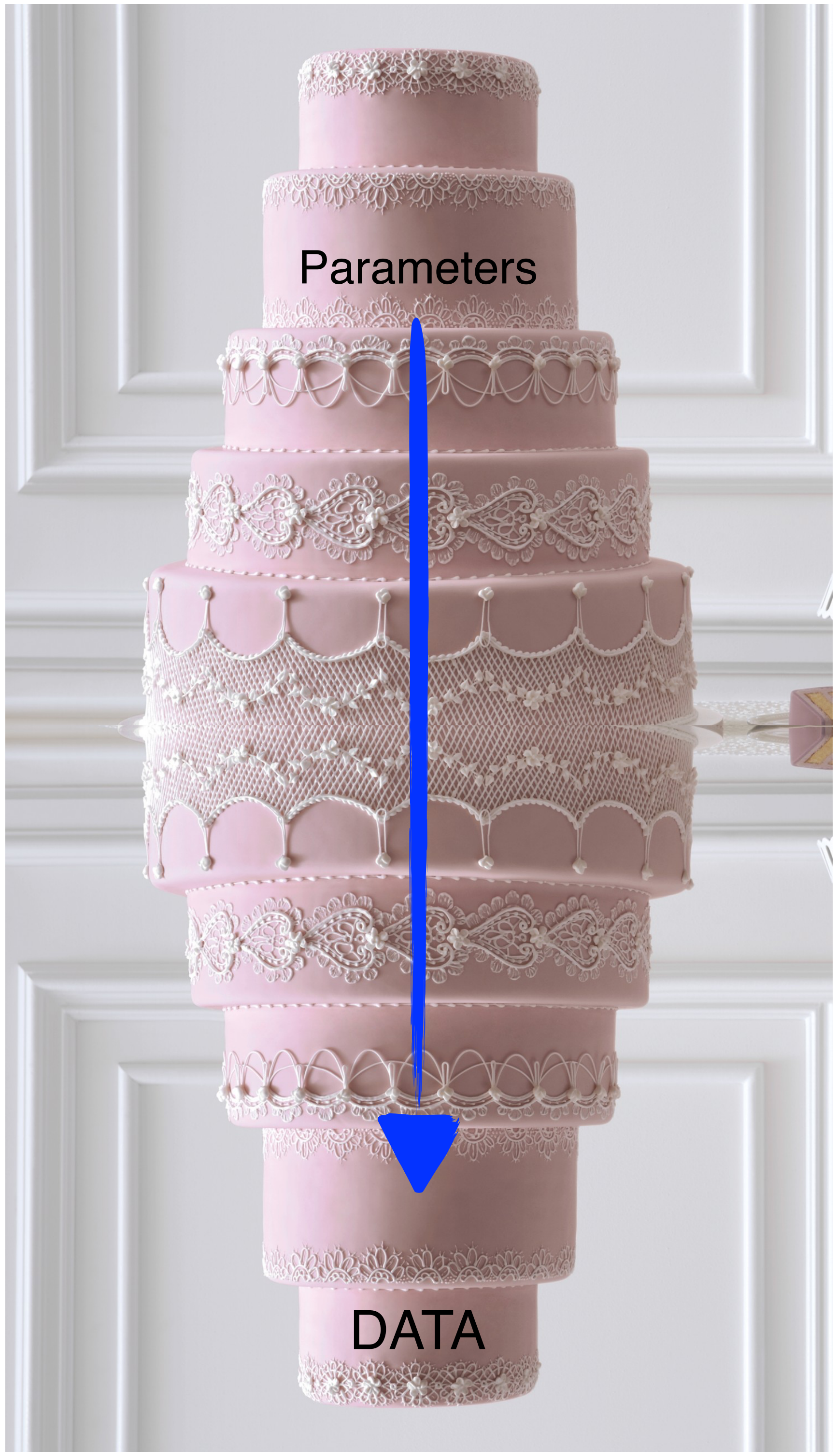
A

$P(B|A)$

B

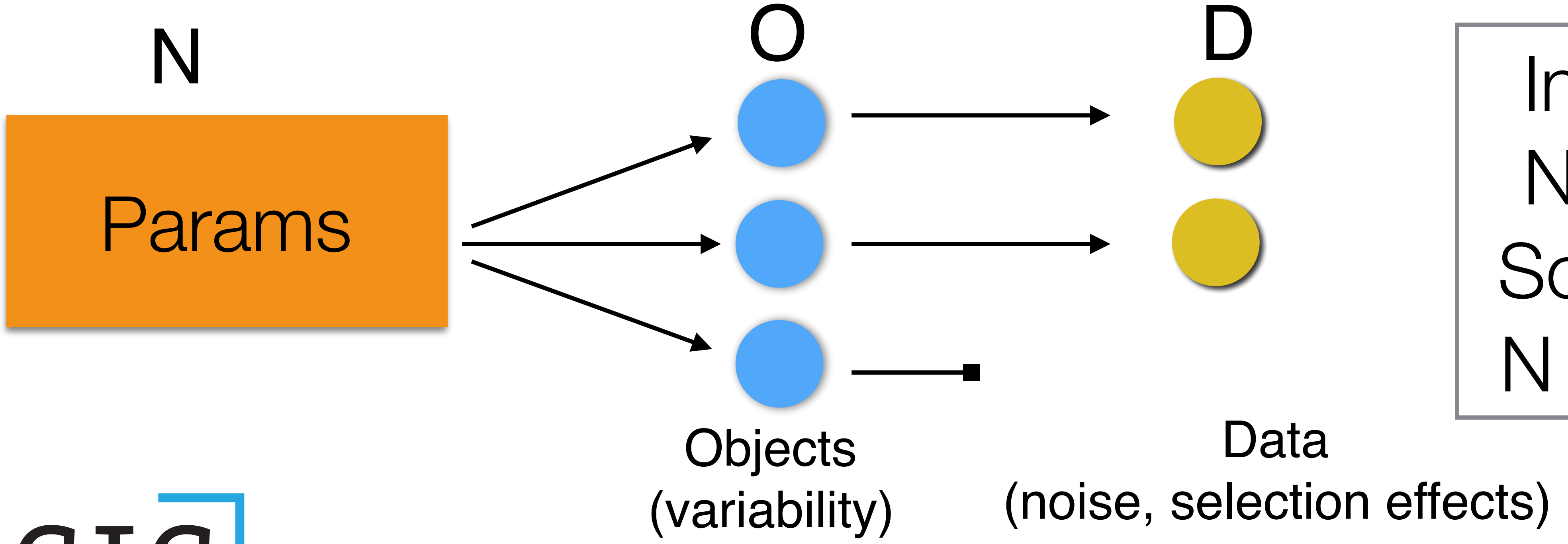
Why "Hierarchical"?

- In cosmology, we have many problems of interest where the "objects" of study are used as tracers for underlying phenomena
- Eg:
  - SNIa's to measure d_L
  - Galaxies to measure velocity fields, BAOs, growth of structure, lensing, …
  - Galaxy properties to measure scaling relationships
  - Stars to measure Milky Way gravitational potential/dark matter
  - …
- In many cases, we might or might not be interested in the objects themselves — insofar as they give us accurate (and unbiased) tracers for the physics we want to study



Why “Models”?

- By “model” in this context I mean a probabilistic representation of how the measured data arise from the theory
- We always need models: They incorporate our understanding of how the measurement process (and its subtleties, e.g. section effects) “filters” our view of the underlying physical process
- The more refined the model, the more information we can extract from the data: measurement noise is unavoidable (at some level), but supplementing our inferential setup with a probabilistic model takes some “heavy lifting” away from the data
- The key is to realise that there is a difference between “measurement noise” and intrinsic variability — and each needs to be modelled individually



In general:
$N + O > D$
Sometimes:
$N + O >> D$

Mathematical formulation

The posterior distribution can be expanded in the usual Bayesian way:

$$\text{p(params | data)} \propto \text{p(data | params)p(params)}$$

$$\text{p(data|params)} \propto \int \text{p(data, true, pop | params) dtrue dpop}$$

$$= \int \boxed{\text{p(data | true)}}\,\boxed{\text{p(true | pop)}}\,\boxed{\text{p(pop)}}\ \text{dtrue drop}$$

🟥 Measurement errors

🟦 Intrinsic variability

🟩 Population-level priors

Gaussian linear model

- Intuition can be gained from the “simple” 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):

$$y_i = b + a x_i$$

Model: unknown parameters of interest (a,b)

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

POPULATION DISTRIBUTION

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

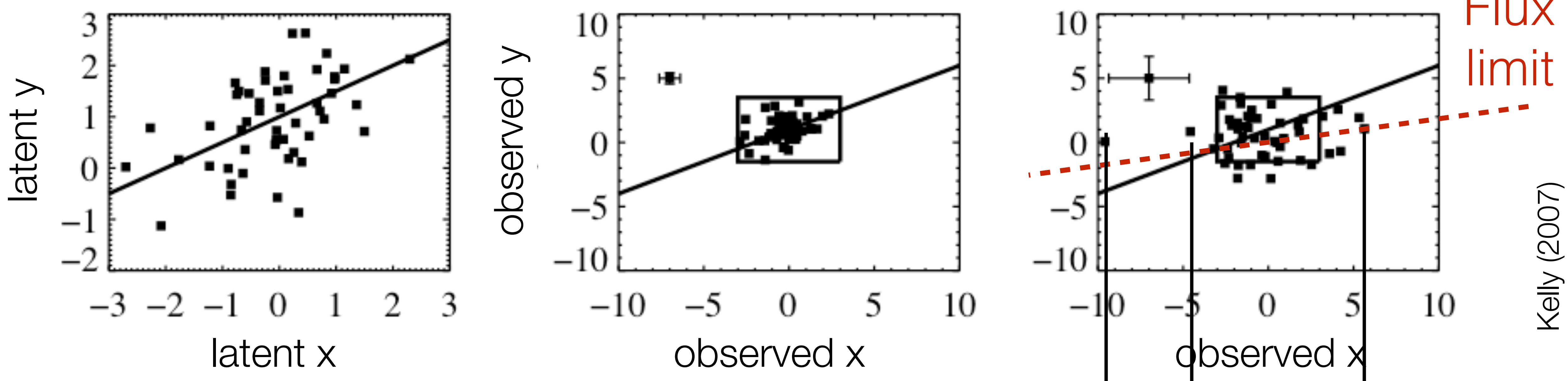
INTRINSIC VARIABILITY

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

MEASUREMENT ERROR

INTRINSIC VARIABILITY

+ MEASUREMENT ERROR



Flux limit

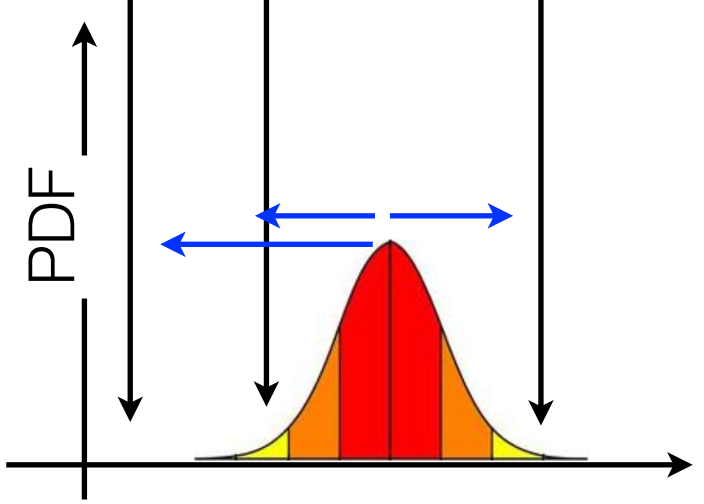
Kelly (2007)

TRUE VALUES

“SMALL” ERRORS

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

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



latent distrib’on

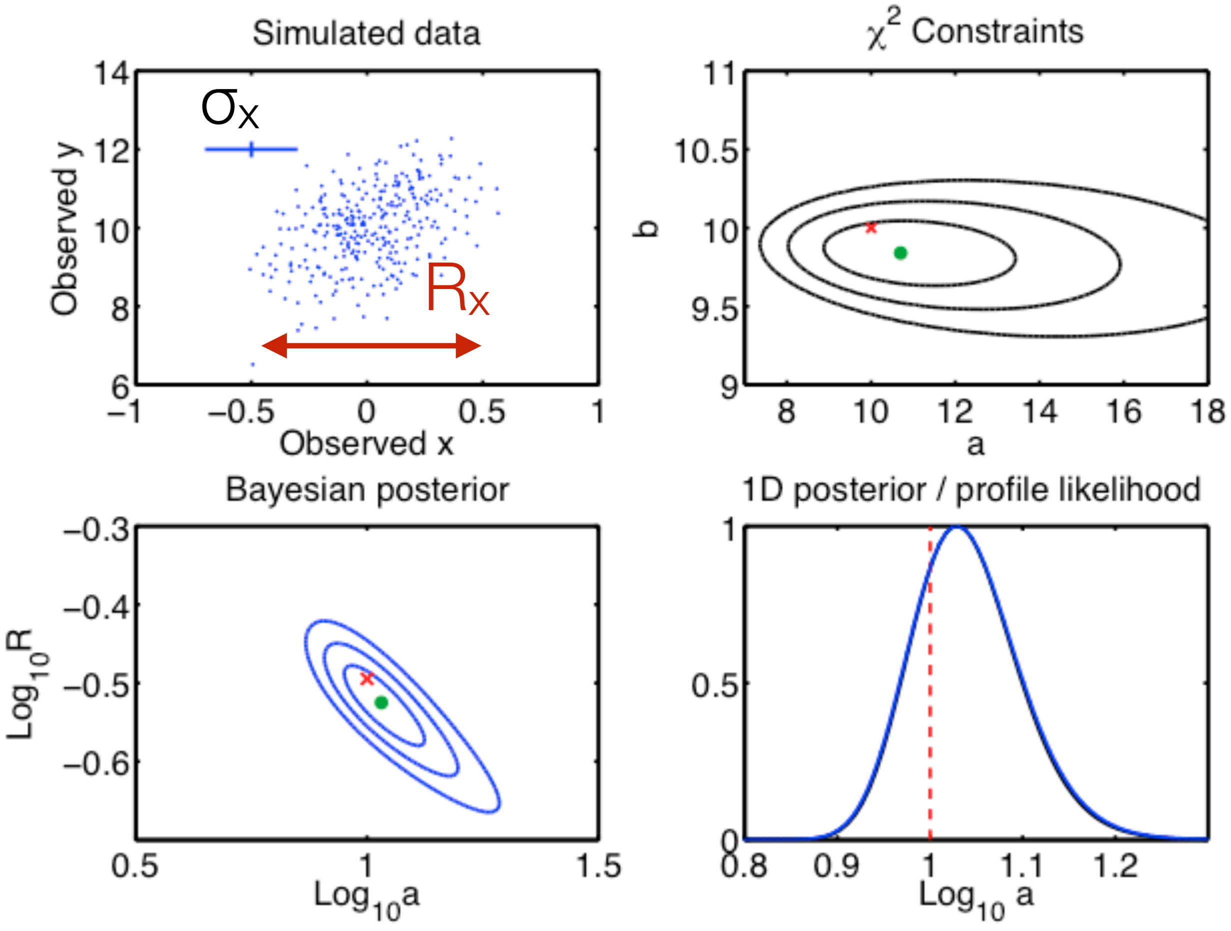
“LARGE” ERRORS

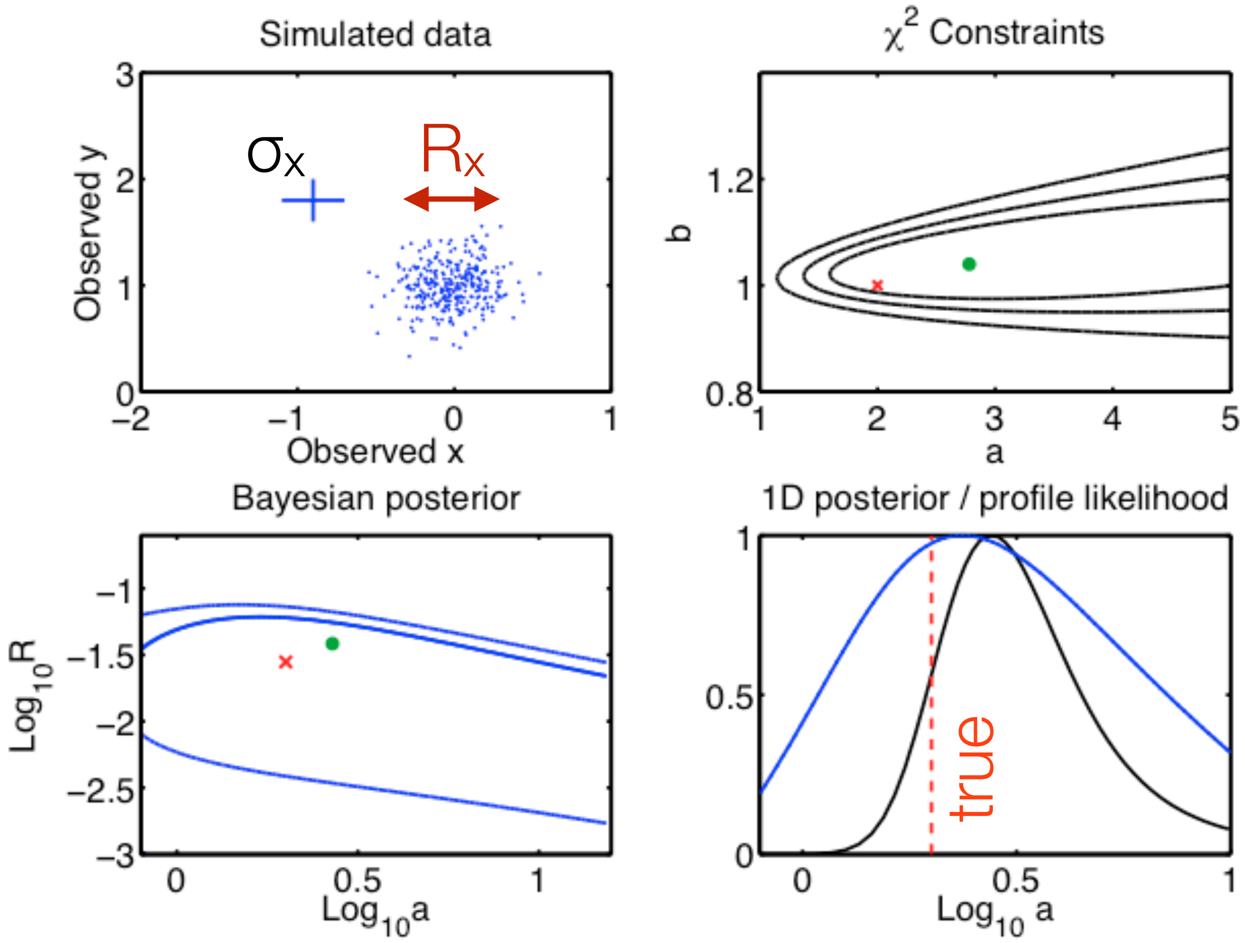
The key parameter is noise (σ_x) to population (R_x) characteristic variability scale ratio

$\sigma_x/R_x << 1$

$$y_i = b + a x_i$$

$\sigma_x/R_x \sim 1$





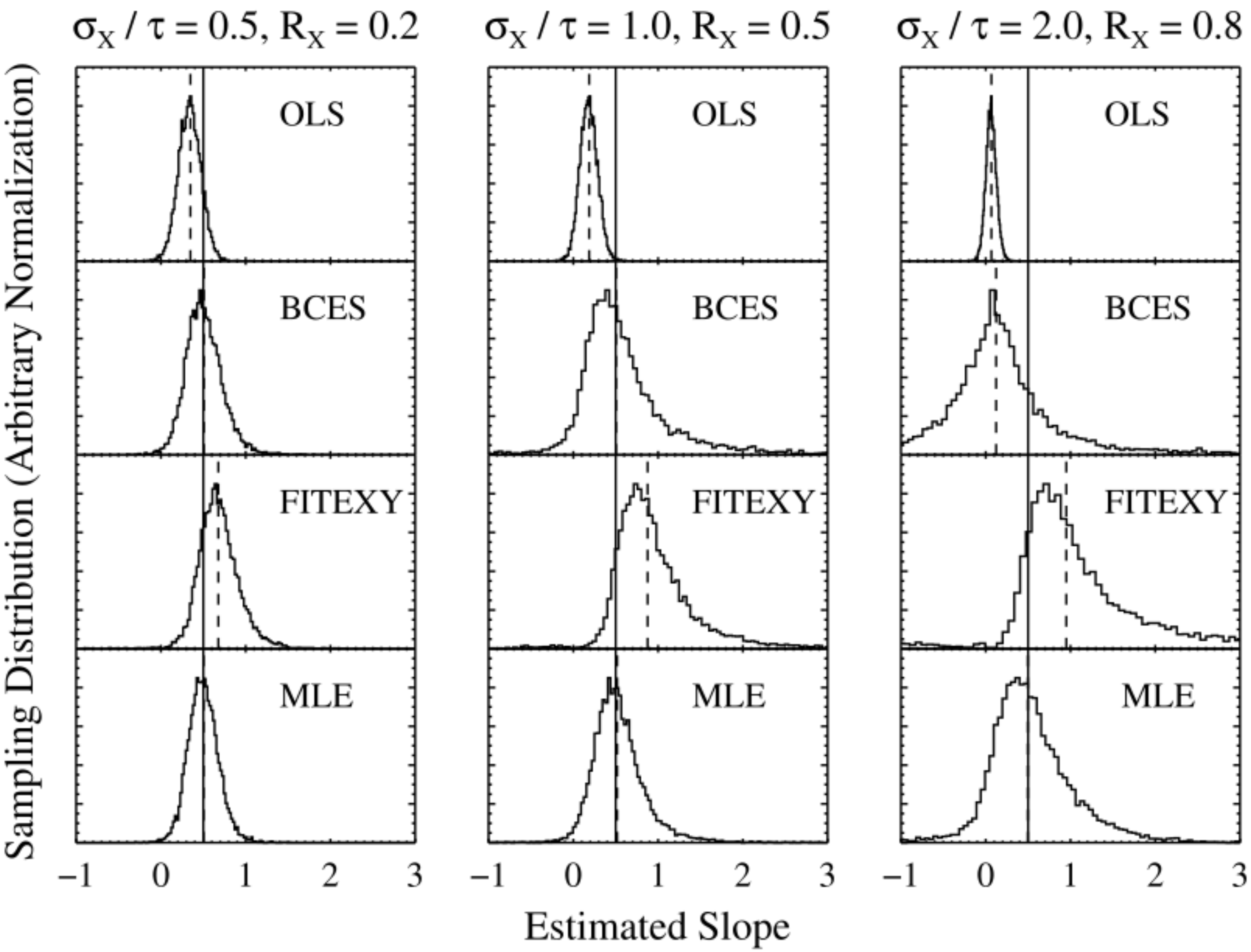
Bayesian (black) marginal posterior identical to Chi-Squared (blue)

Bayesian marginal posterior broader but less biased than Chi-Squared

March, RT et al (2011)

Slope reconstruction

$R_x = \sigma_x^2/\mathrm{Var}(x)$: ratio of the covariate measurement variance to observed variance



Ordinary Least Square

BIASSED LOW

Chi-Square incl variance

BIASSED HIGH

Maximum Likelihood

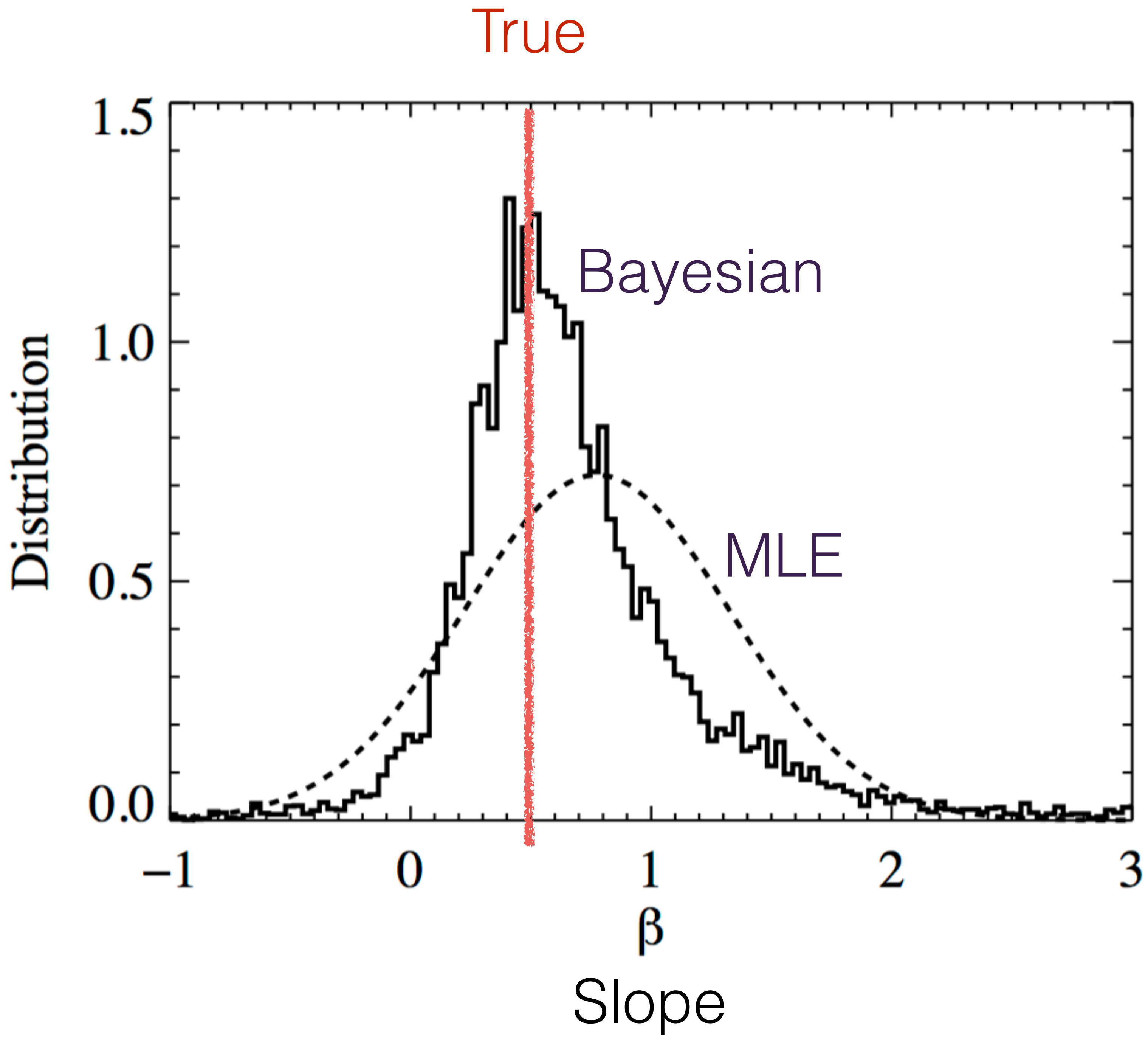
APPROX UNBIASSED

Kelly, Astr. J., 665, 1489-1506 (2007)

Why should you care?

Imperial College London

$R_x = \sigma_x^2/\mathrm{Var}(x) = 1$ in this example: Comparing the MLE (dashed) with the Bayesian Hierarchical Model Posterior (histogram)



Standard MLE (or Least Squares/ Chi-Squared) fits are biased!

(even if you artificially inflate the errors to get Chi-Squared/dof ~ 1)

Kelly, Astr. J., 665, 1489-1506 (2007)

Supernovae Type Ia Cosmology example

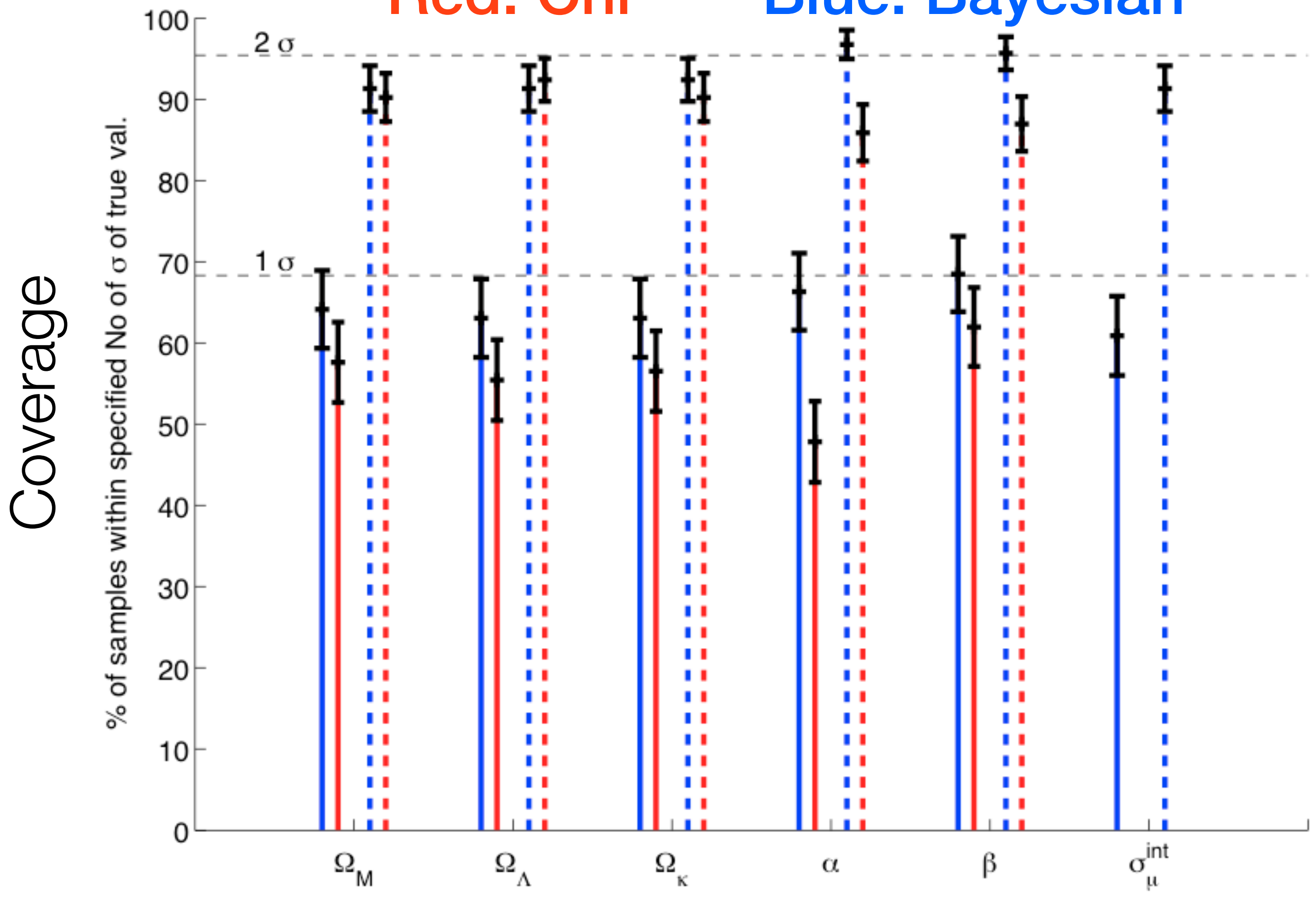
- Coverage of Bayesian 1D marginal posterior CR and of 1D Chi² profile likelihood CI computed from 100 realizations
- Bias and mean squared error (MSE) defined as

$$\text{bias} = \langle \hat{\theta} - \theta_{\text{true}} \rangle$$

$$\text{MSE} = \text{bias}^2 + \text{Var}$$

$\hat{\theta}$ is the posterior mean (Bayesian) or the maximum likelihood value (Chi²).

Red: Chi²   Blue: Bayesian



Results:

Coverage: generally improved (but still some undercoverage observed)

Bias: reduced by a factor ~ 2-3 for most parameters

MSE: reduced by a factor 1.5-3.0 for all parameters

BAHAMAS

BAyesian **H**ier**A**rchical **M**odeling for the **A**nalysis of **S**upernova cosmology *ArXiv: 1510.05954*

Cosmo/Standardization

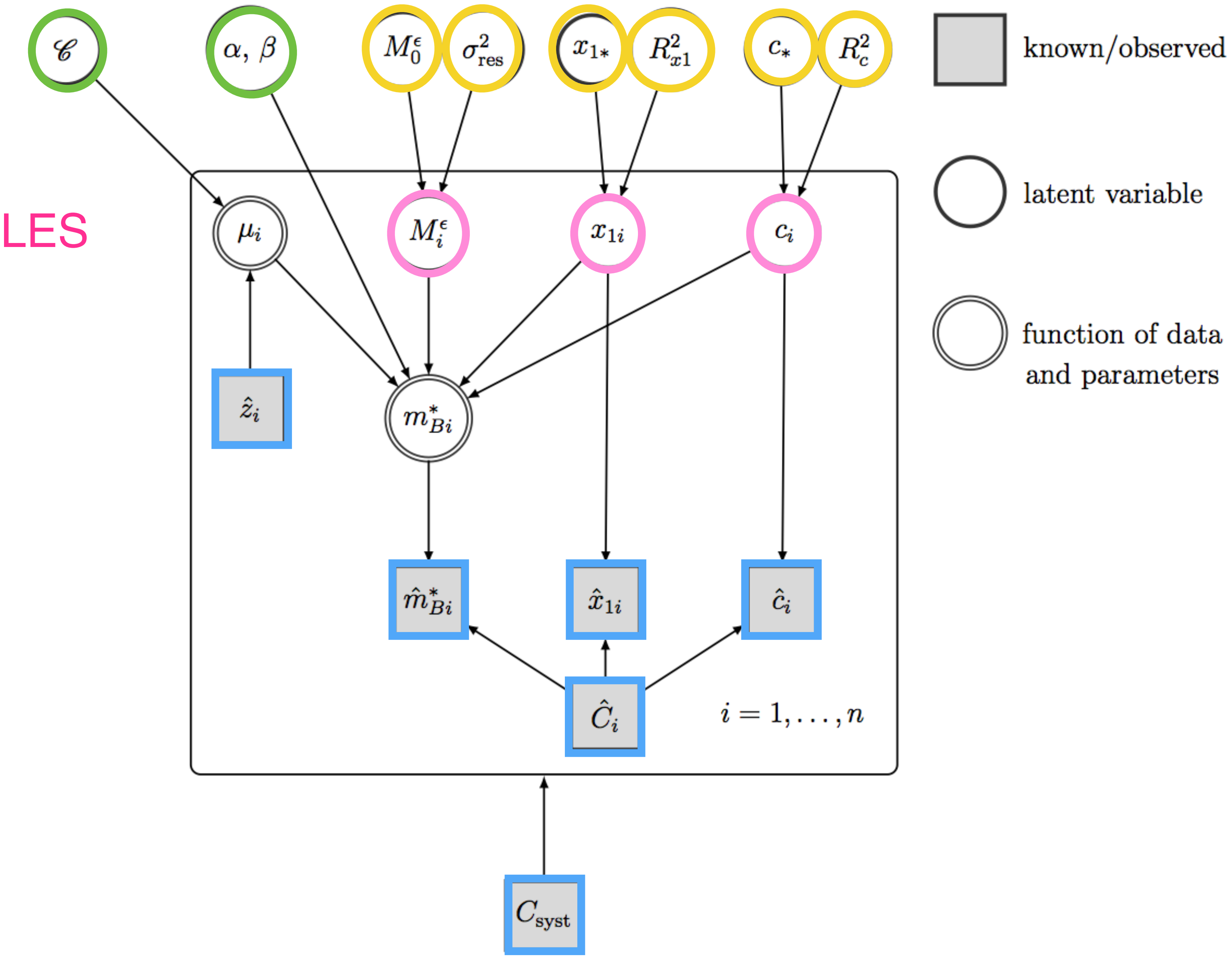
POPULATION PARAMETERS

Intrinsic variability

LATENT VARIABLES

Observational noise

DATA



All variables are fit simultaneously using powerful Gibbs-type samplers (~ 3000 dimensions)

March,RT+11; Jiao, RT,+15; Shariff, RT+15

Sampling strategy

- Introduction of an explicit layer of latent variables increases massively the dimensionality of the parameter space that needs to be sampled (eg: SNIa , 3 x N, N ~ 700, the number of SNIa's observed; >> for BayesSN, ~ 50,000)

- For Gaussian linear models, latent variables can be marginalised out analytically

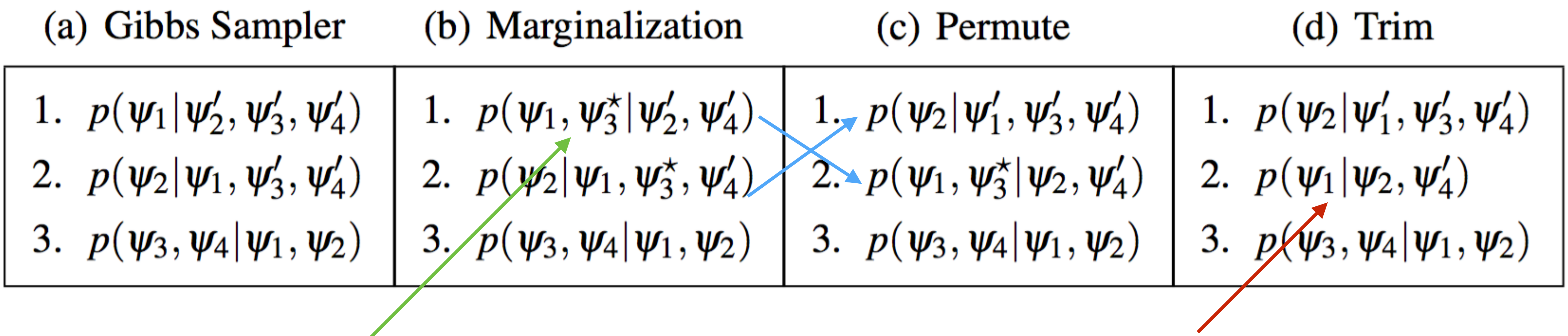
- But: selection effects break Gaussianity, hence require an explicit numerical sampling scheme

Sampling scheme	Ok?	Note
Metropolis-Hastings	✗	Hopeless efficiency
Gibbs	✓	Exploits conditional structure
Hamiltonian MC	✓	Good but requires gradient
Nested sampling	✗	Too large dimensionality

Partially Collapsed Gibbs

- The Partially Collapsed Gibbs (PCG) sampler replaces complete conditional distributions in Gibbs by **complete conditionals of marginal distribution** of the joint target. This reduction of conditioning can improve convergence.

Construction of the PCG Sampler :



(a) Gibbs Sampler	(b) Marginalization	(c) Permute	(d) Trim
1. $p(\psi_1 \mid \psi_2', \psi_3', \psi_4')$	1. $p(\psi_1, \psi_3^\star \mid \psi_2', \psi_4')$	1. $p(\psi_2 \mid \psi_1', \psi_3', \psi_4')$	1. $p(\psi_2 \mid \psi_1', \psi_3', \psi_4')$
2. $p(\psi_2 \mid \psi_1, \psi_3', \psi_4')$	2. $p(\psi_2 \mid \psi_1, \psi_3^\star, \psi_4')$	2. $p(\psi_1, \psi_3^\star \mid \psi_2, \psi_4')$	2. $p(\psi_1 \mid \psi_2, \psi_4')$
3. $p(\psi_3, \psi_4 \mid \psi_1, \psi_2)$	3. $p(\psi_3, \psi_4 \mid \psi_1, \psi_2)$	3. $p(\psi_3, \psi_4 \mid \psi_1, \psi_2)$	3. $p(\psi_3, \psi_4 \mid \psi_1, \psi_2)$

Jointly updated

Steps permutation

Analytical marginalisation

ASIS sampling scheme

- In *BAHAMAS*, we use an Ancillarity-Sufficiency Interweaving Strategy (ASIS) and a Partially Collapsed Gibbs (PCG) sampler to improve efficiency (Jiao&vanDyk16, vanDyk&Park08; Yu&Meng11)

- **ASIS:** A special **Data Augmentation scheme:** we introduce "missing data" (aka "messenger field") $Y_{\mathrm{mis,S}}$ or $Y_{\mathrm{mis,A}}$ so that:

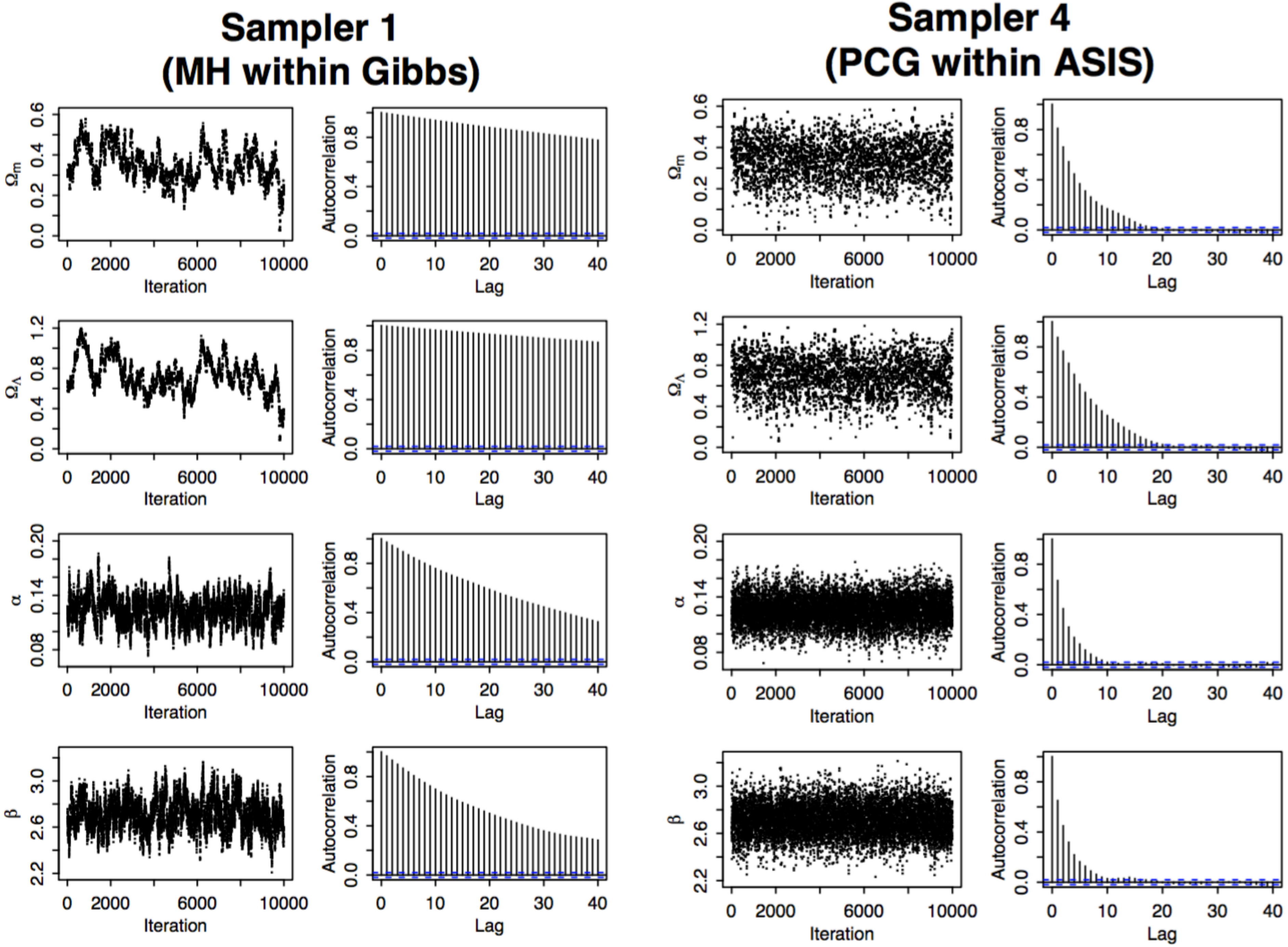
$$p(Y_{\mathrm{obs}}|Y_{\mathrm{mis},S},\theta) \text{ is free of } \theta \qquad p(Y_{\mathrm{mis},A}|\theta) \text{ does not depend on } \theta$$

ASIS Sampler, with target p(θlYobs) :

$$\begin{aligned}
&\text{Step 1:} \quad Y_{\mathrm{mis},S}^{(t+1)} \sim p(Y_{\mathrm{mis},S}|\theta^{(t)}, Y_{\mathrm{obs}}),\\
&\text{Step 2:} \quad \theta^{(t+1/2)} \sim p(\theta|Y_{\mathrm{mis},S}^{(t+1)}, Y_{\mathrm{obs}});\; Y_{\mathrm{mis},A}^{(t+1)} = \mathscr{F}_{\theta^{(t+1/2)}}(Y_{\mathrm{mis},S}^{(t+1)}),\\
&\text{Step 3:} \quad \theta^{(t+1)} \sim p(\theta|Y_{\mathrm{mis},A}^{(t+1)}, Y_{\mathrm{obs}}),
\end{aligned}$$

Performance improvement

- Adopting both ASIS and PCG improved mixing while noticeably reducing correlation length for all variables



Sampler 1 (MH within Gibbs)

Sampler 4 (PCG within ASIS)

Trace plot

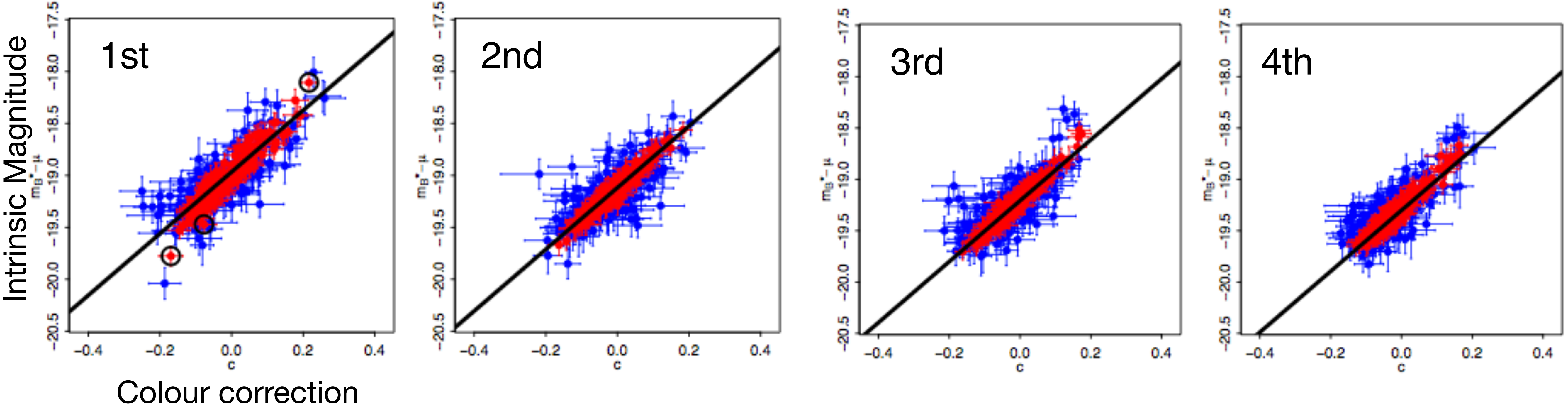
Auto-correlation

Borrowing of strength

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

Likelihood

Posterior



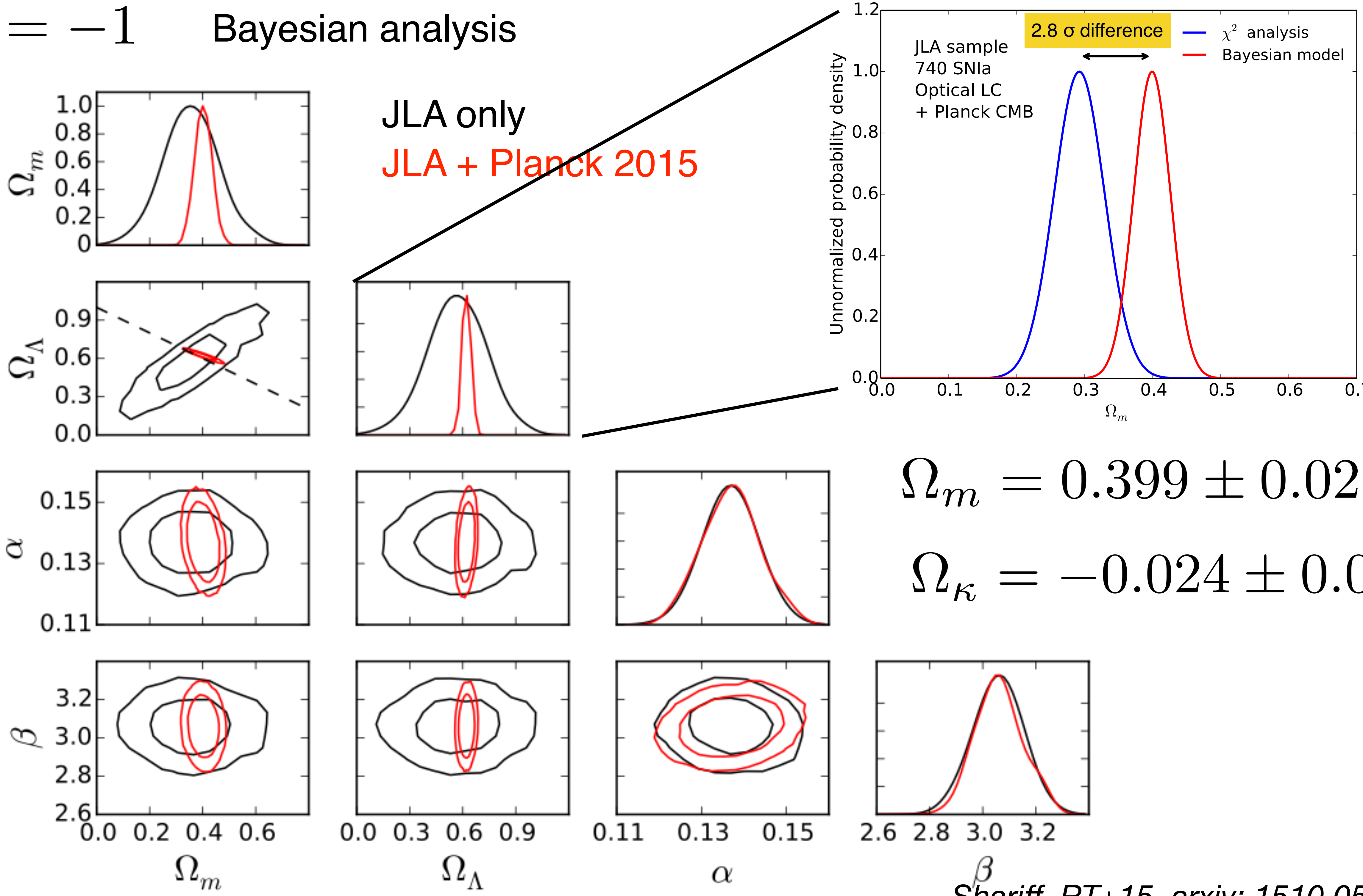
The posterior estimates (red) exhibit smaller residual scatter when compared to the likelihood (blue), around the regression line: "borrowing of strength" from the structure of the hierarchical model.

The Bayesian hierarchical model (*BAHAMAS*) has smaller bias, smaller MSE, better coverage than the standard Chi^2 (March,RT+11)

Shariff, RT+15, arxiv: 1510.05954

Cosmological parameters from JLA

$w = -1$ Bayesian analysis



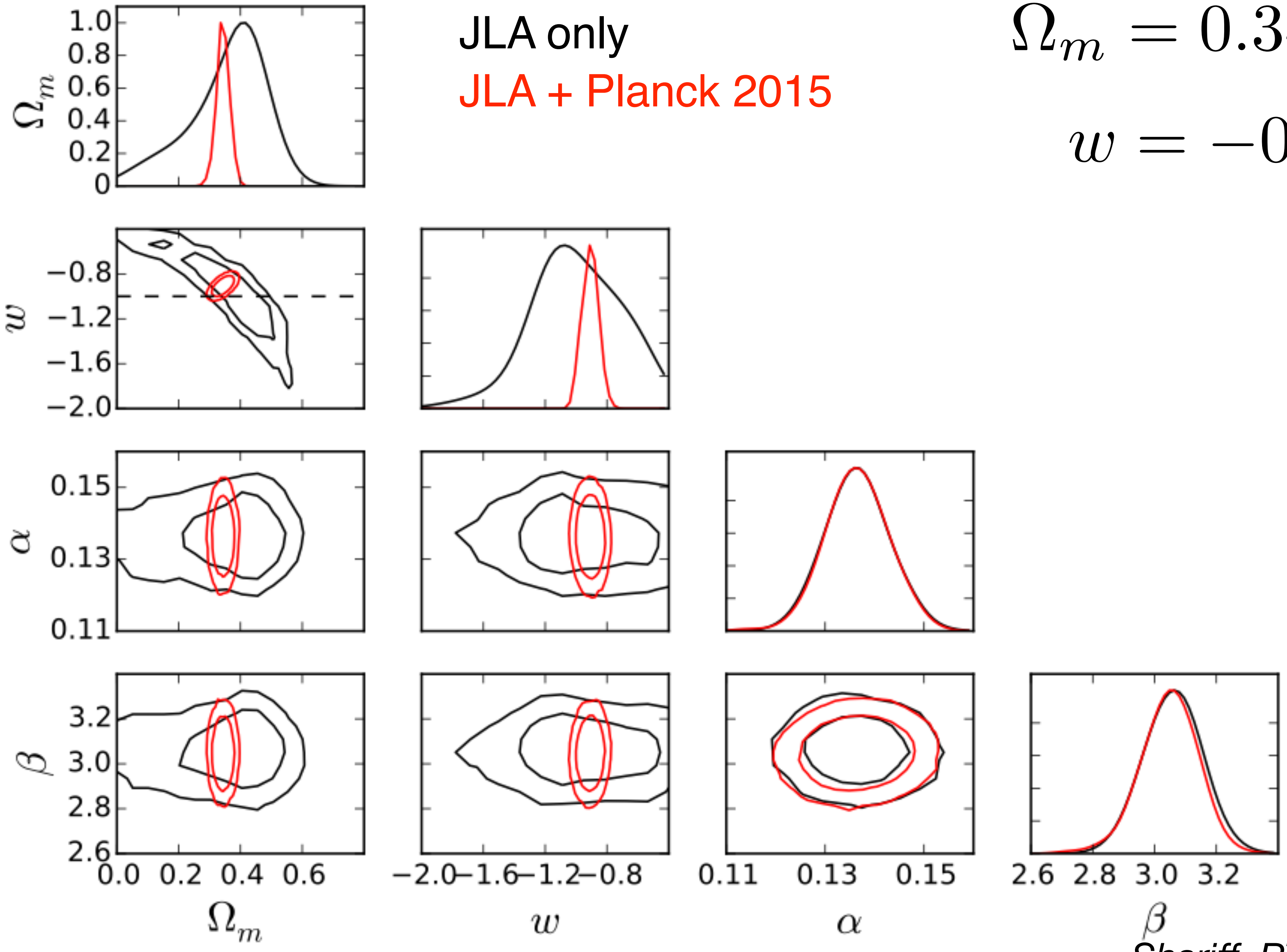
$$\Omega_m = 0.399 \pm 0.027$$

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

Shariff, RT+15, arxiv: 1510.05954

Cosmological parameters from JLA

$\Omega_\kappa = 0$ Bayesian analysis



JLA only

JLA + Planck 2015

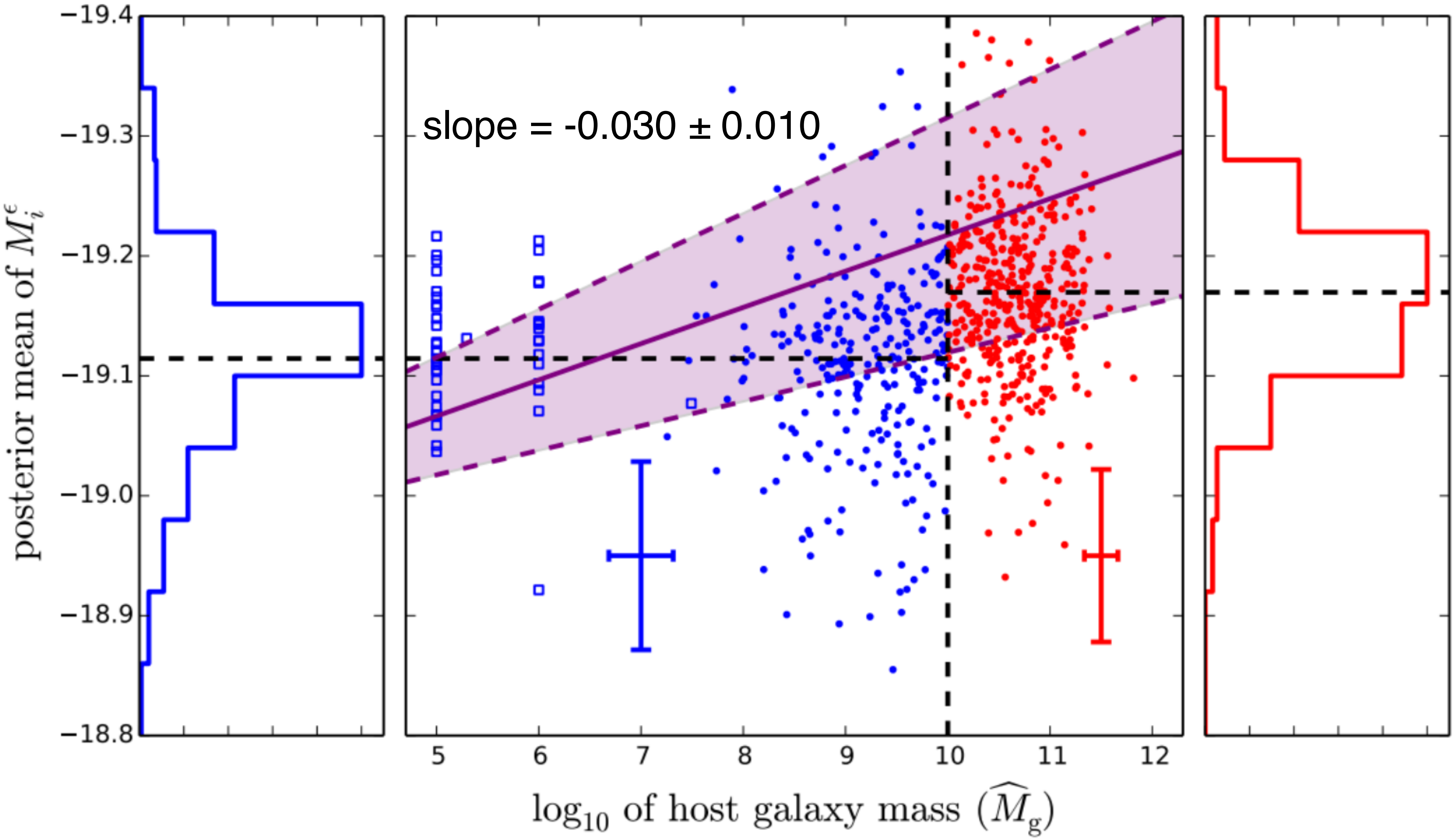
$$\Omega_m = 0.343 \pm 0.019$$

$$w = -0.90 \pm 0.05$$

Shariff, RT+15, arxiv: 1510.05954

Host-galaxy mass as additional covariate

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

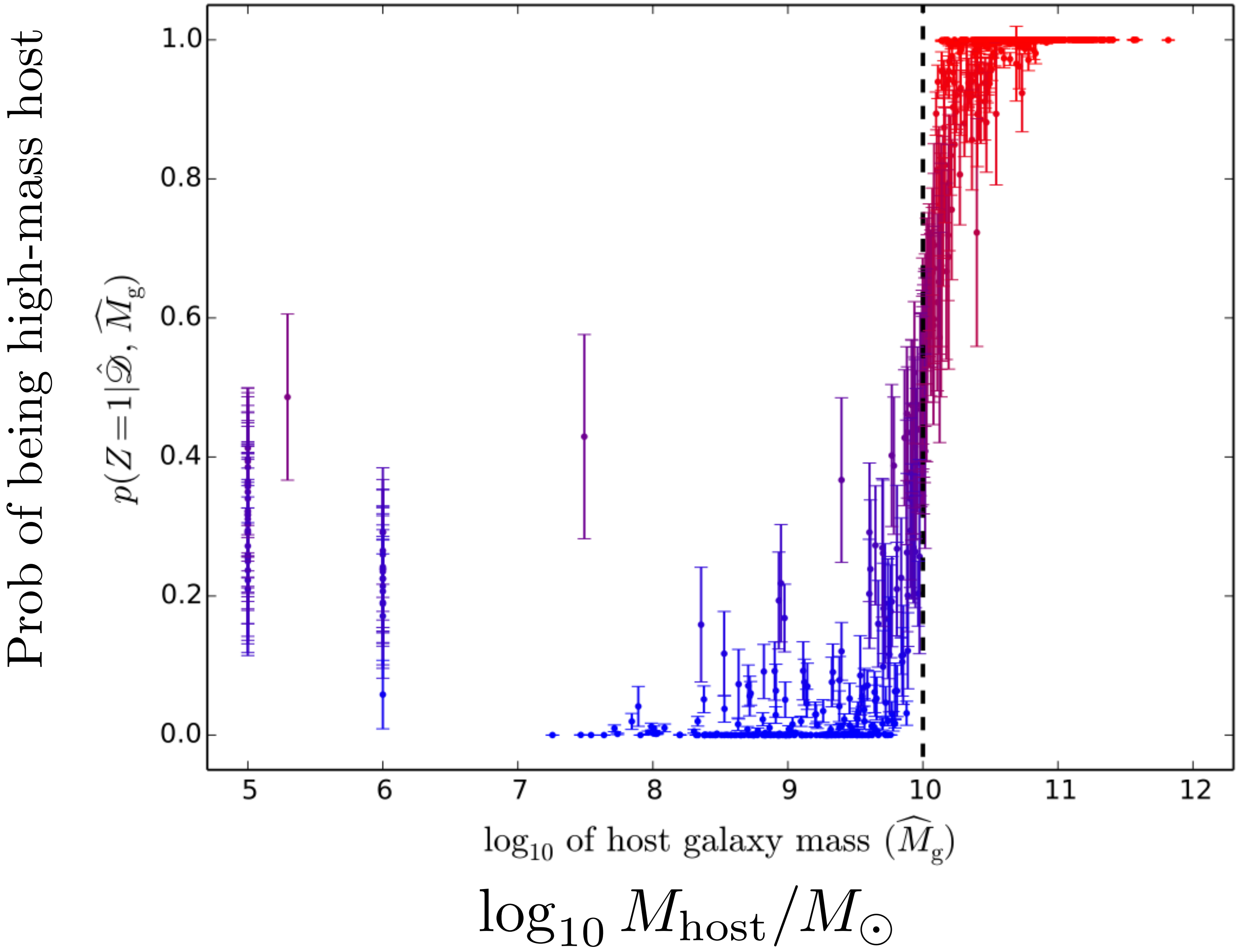


$\log_{10} M_{\text{host}}/M_\odot$

Shariff, RT+15, arxiv: 1510.05954

Probabilistic classification of hosts

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



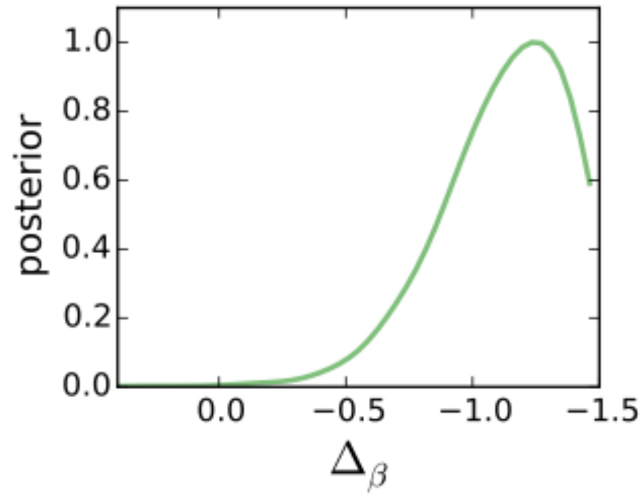
Shariff, RT+15, arxiv: 1510.05954

z-evolution of colour correction

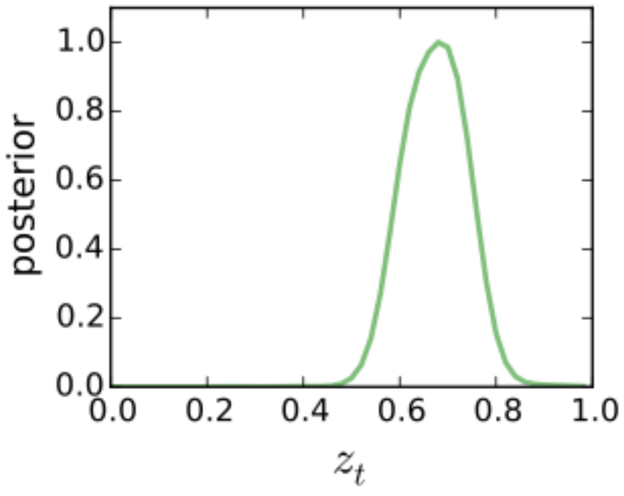
Strong preference (~ 4-sigma) for a transition from β ~ 3.1 to β ~ 2 at z ~ 0.7.

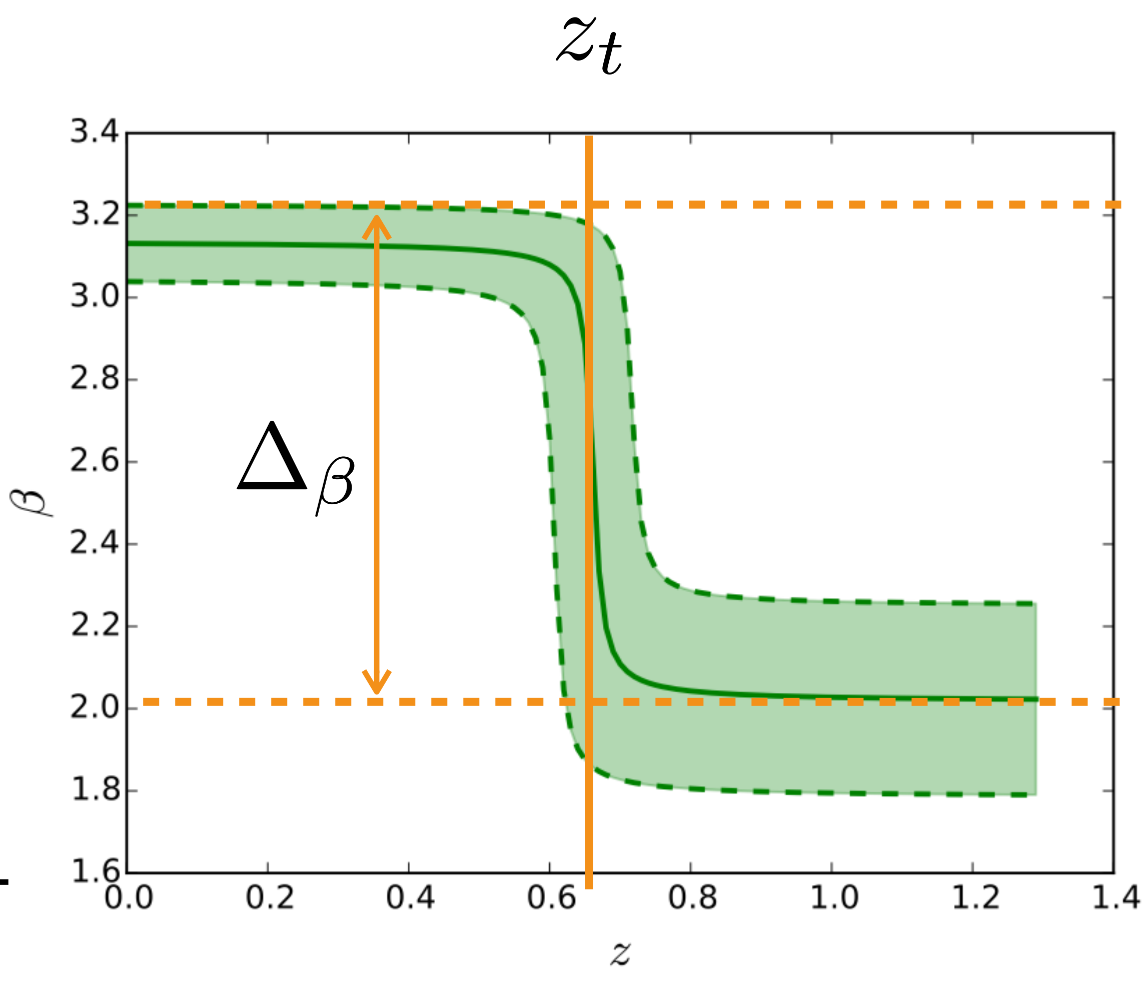
This however does not change cosmological inferences.

Change in β



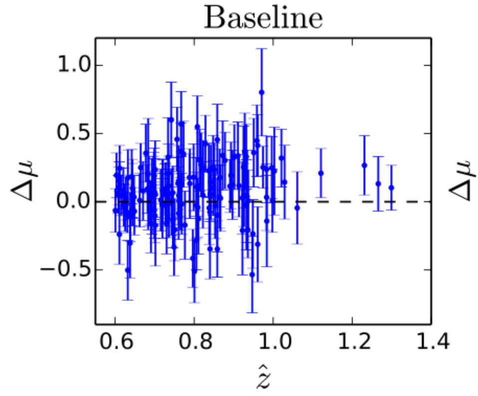
Redshift of transition



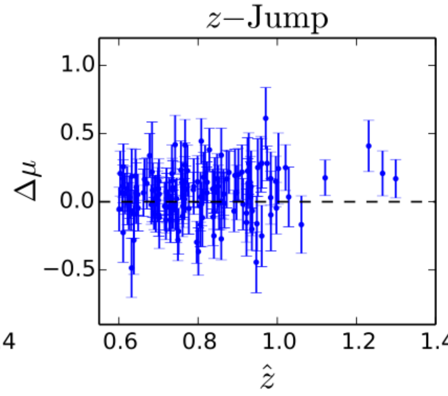


Res scatter

No step

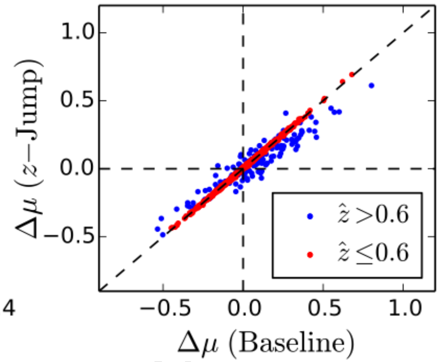


Step



Red: z < 0.6

Blue: z > 0.6



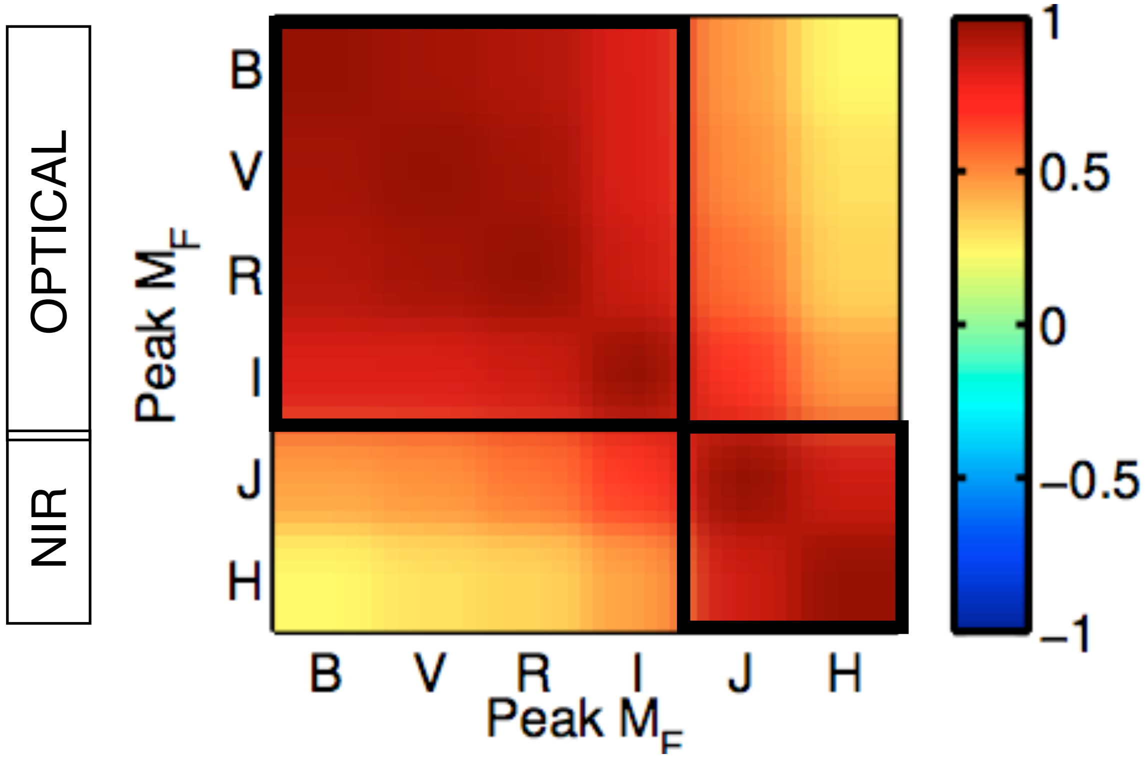
Step

No step

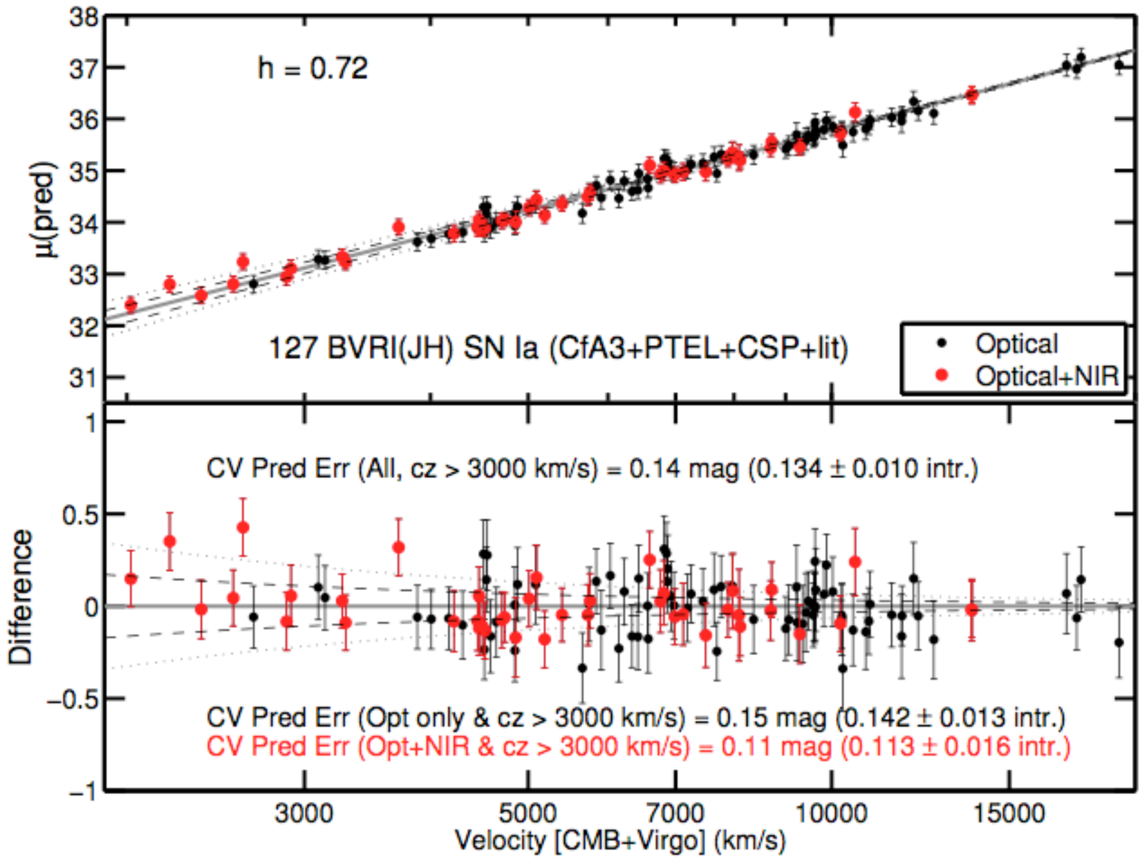
Shariff, RT+15, arxiv: 1510.05954

The advantage of NIR data

NIR and optical uncorrelated: extra information in the NIR!



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

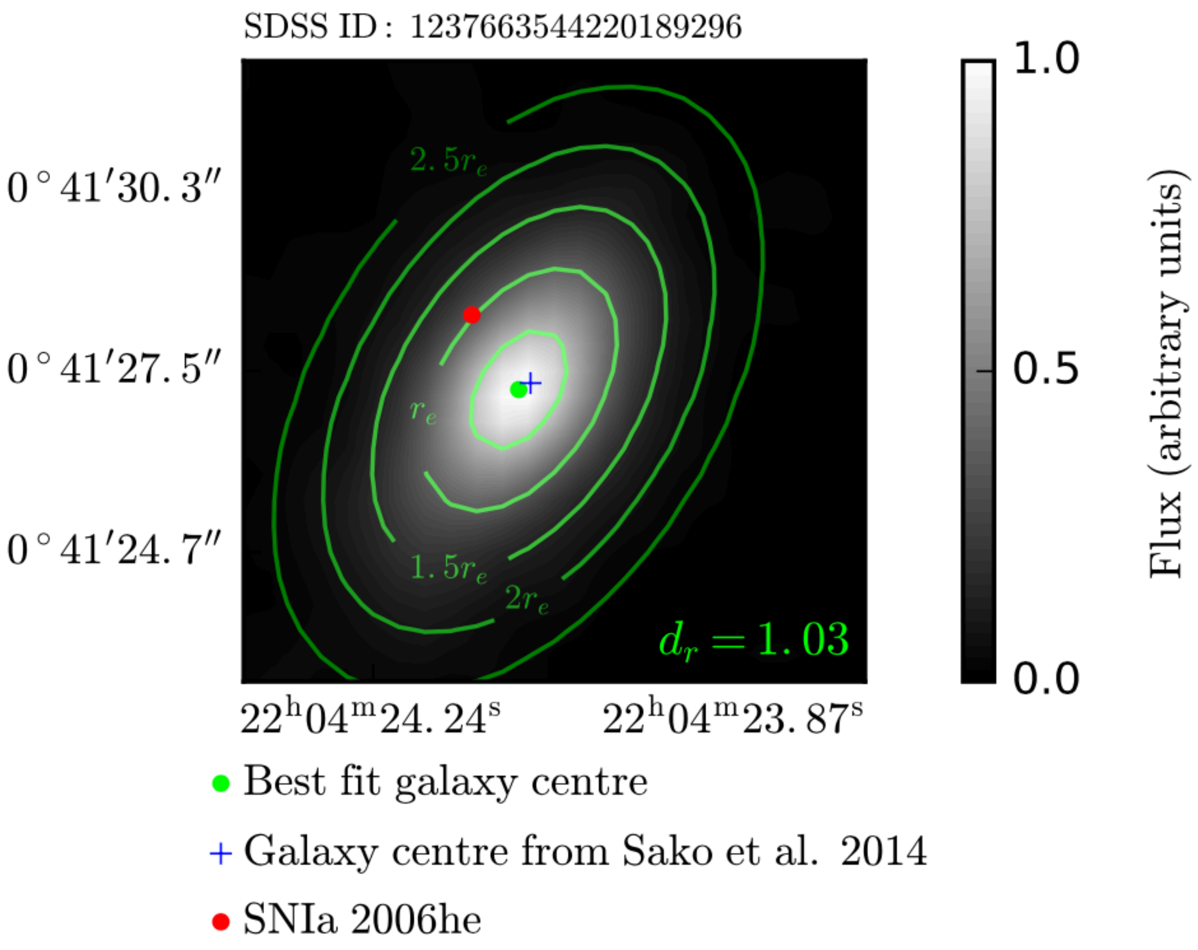


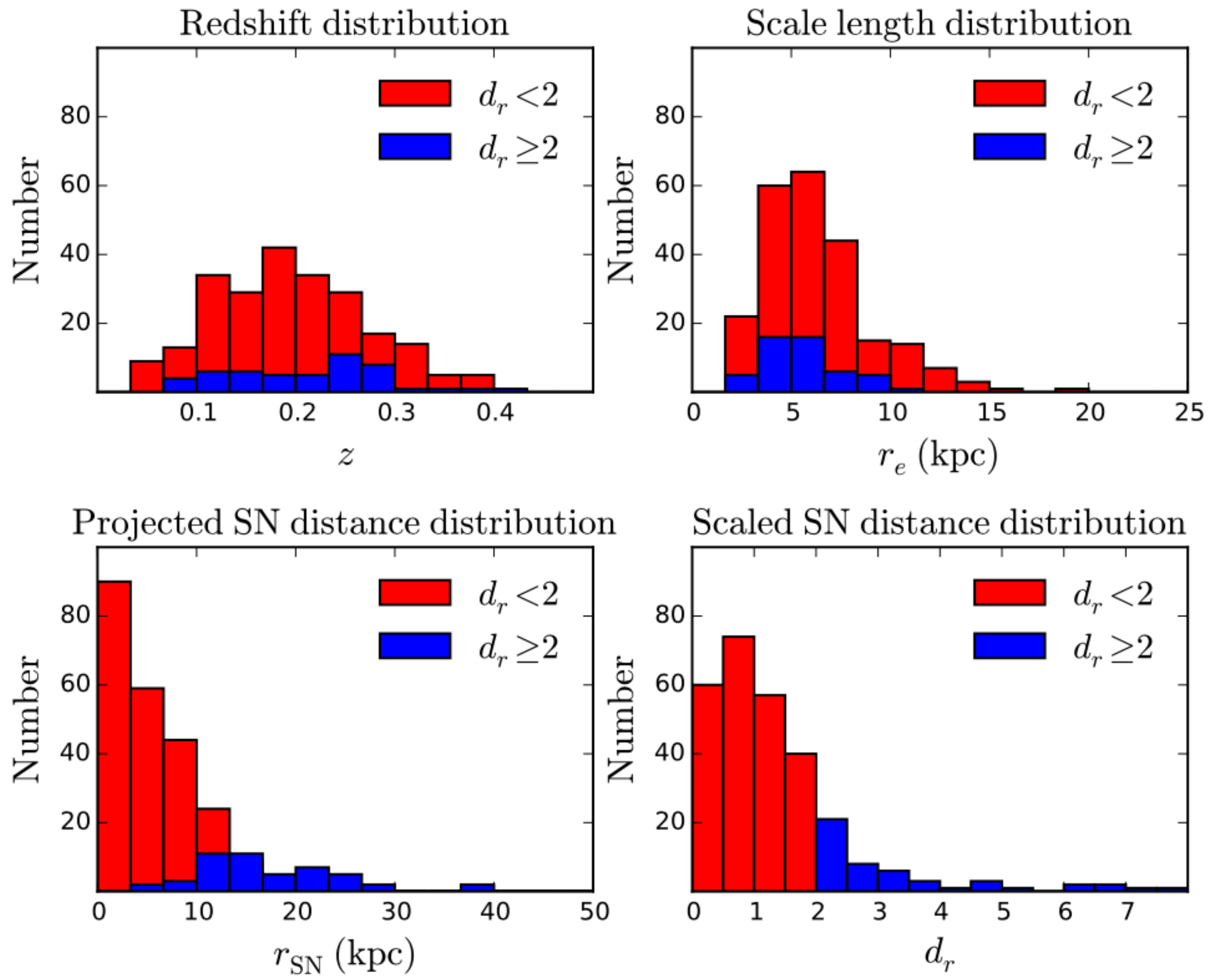
But: You need to go to space to get z~1 rest-frame NIR! (RAISIN/RAISIN 2 with HST, Kirshner+)

Mandel+09,11

SNIa's in the outskirts of galaxies are better standard candles

Looking for SNIa in the outskirts (= less dusty/more homogeneous) regions of galaxies (280 SDSS host galaxies fitted): measured the galactocentric SN distance normalised to the host's scale length, d_r





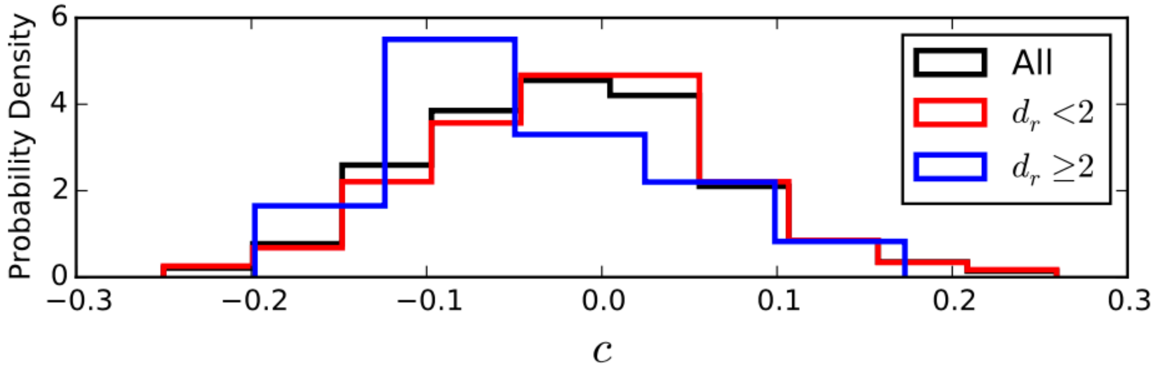
Hill, RT+ 2018 (1612.04417), Galbany+12

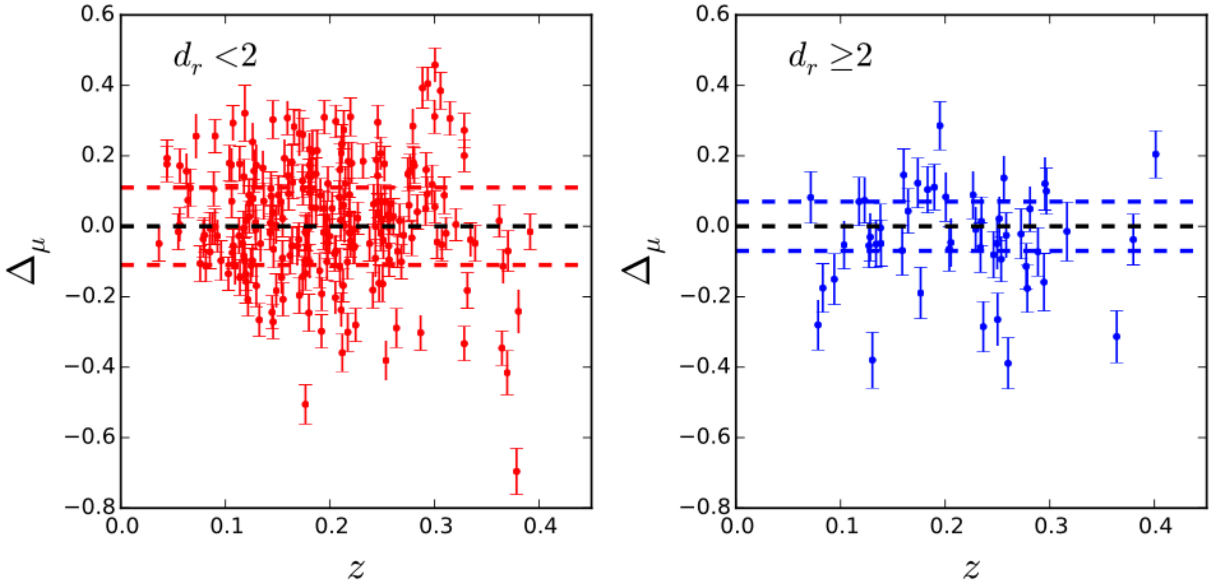
Segregation by distance to host centre

Projected distance to the host centre in units of half-light radius (d_r)

Significant difference (95% CL) in the colour correction between the two samples: high d_r SNIas are bluer.

Residual scatter in the high-distance (d_r >2) sub-sample is reduced by ~30%

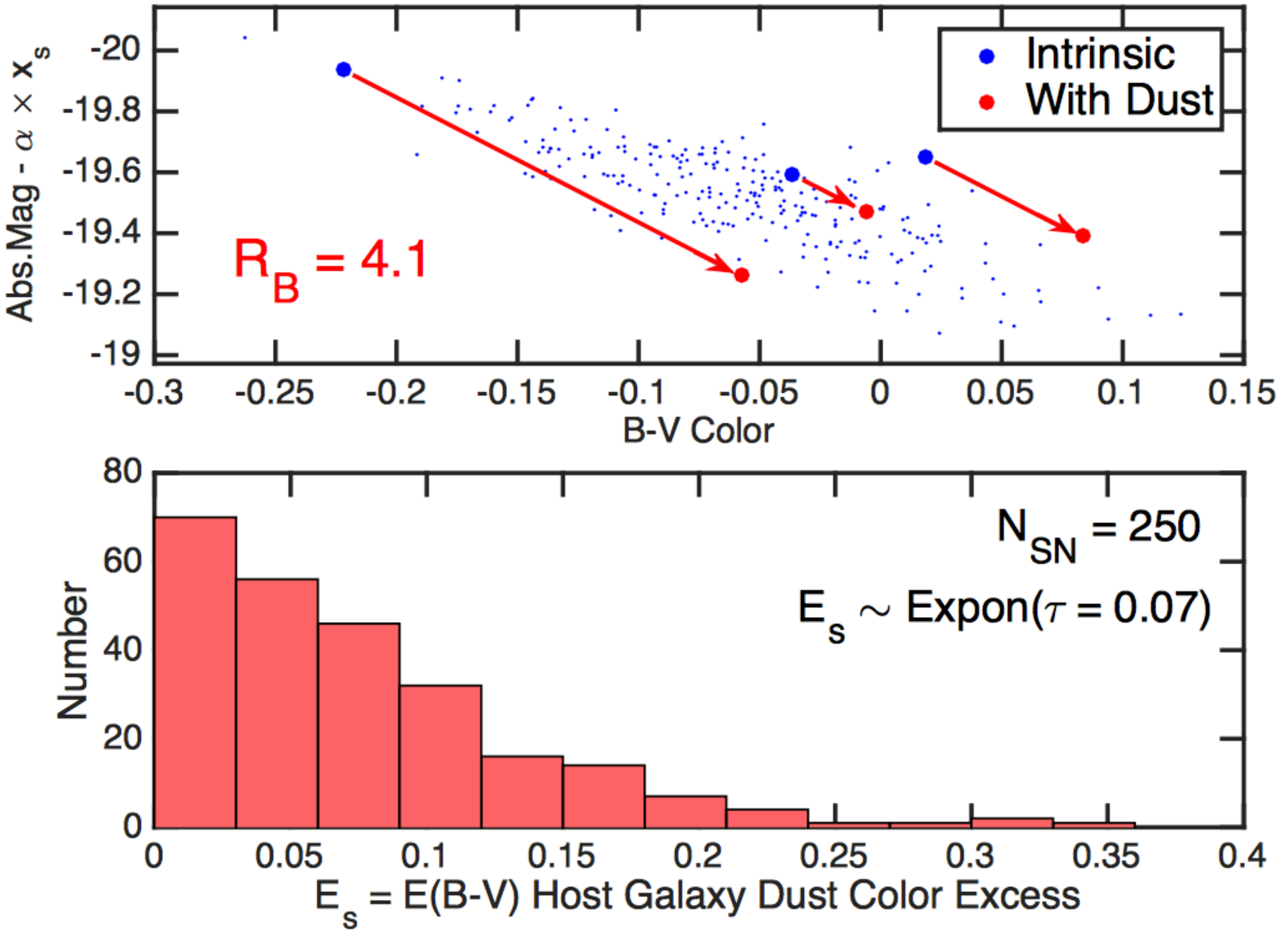


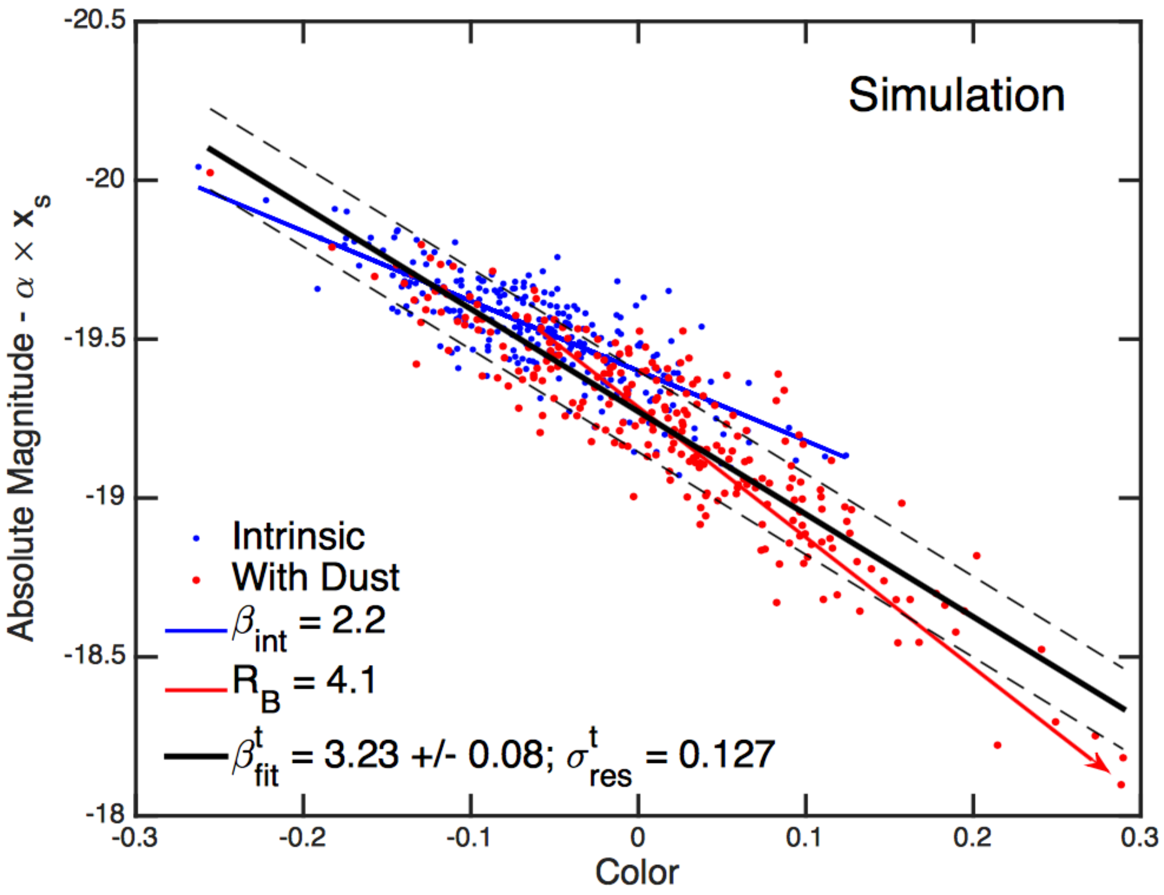


Mandel+16

Take advantage of the fact that dust only absorbs and reddens to split colour into intrinsic and dust-related. The usual linear fit returns a slope that is not the intrinsic slope, nor the dust slope, but an average between the two.

Simulated data



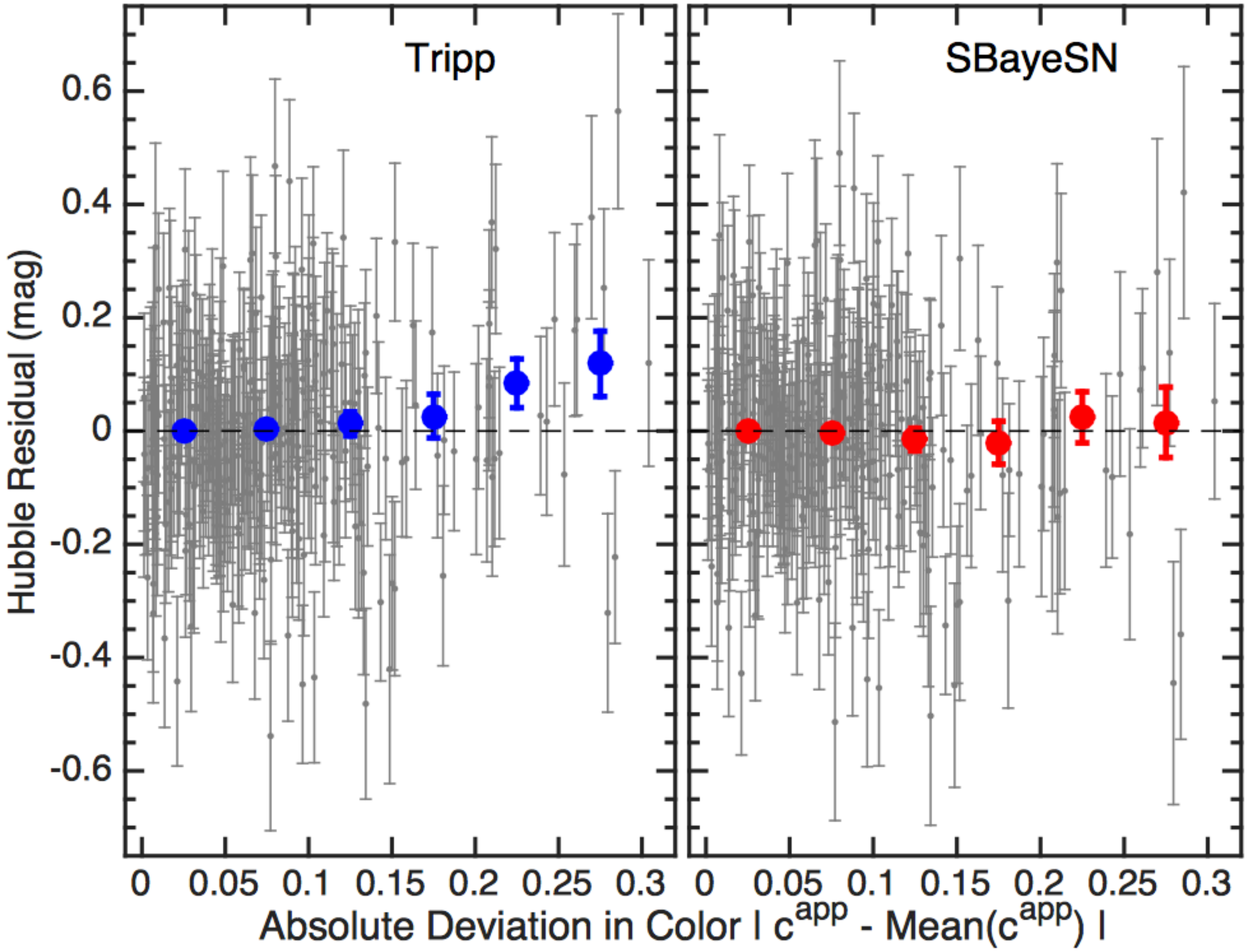


Mandel+16, 1609.04470

Mandel+16: Implications

The non-linear colour fit reduces Hubble residuals for very red/very blue objects wrt the conventional Tripp linear formula. This might have important implications for cosmological fits, with systematic corrections of up to ~ 0.1 mag.

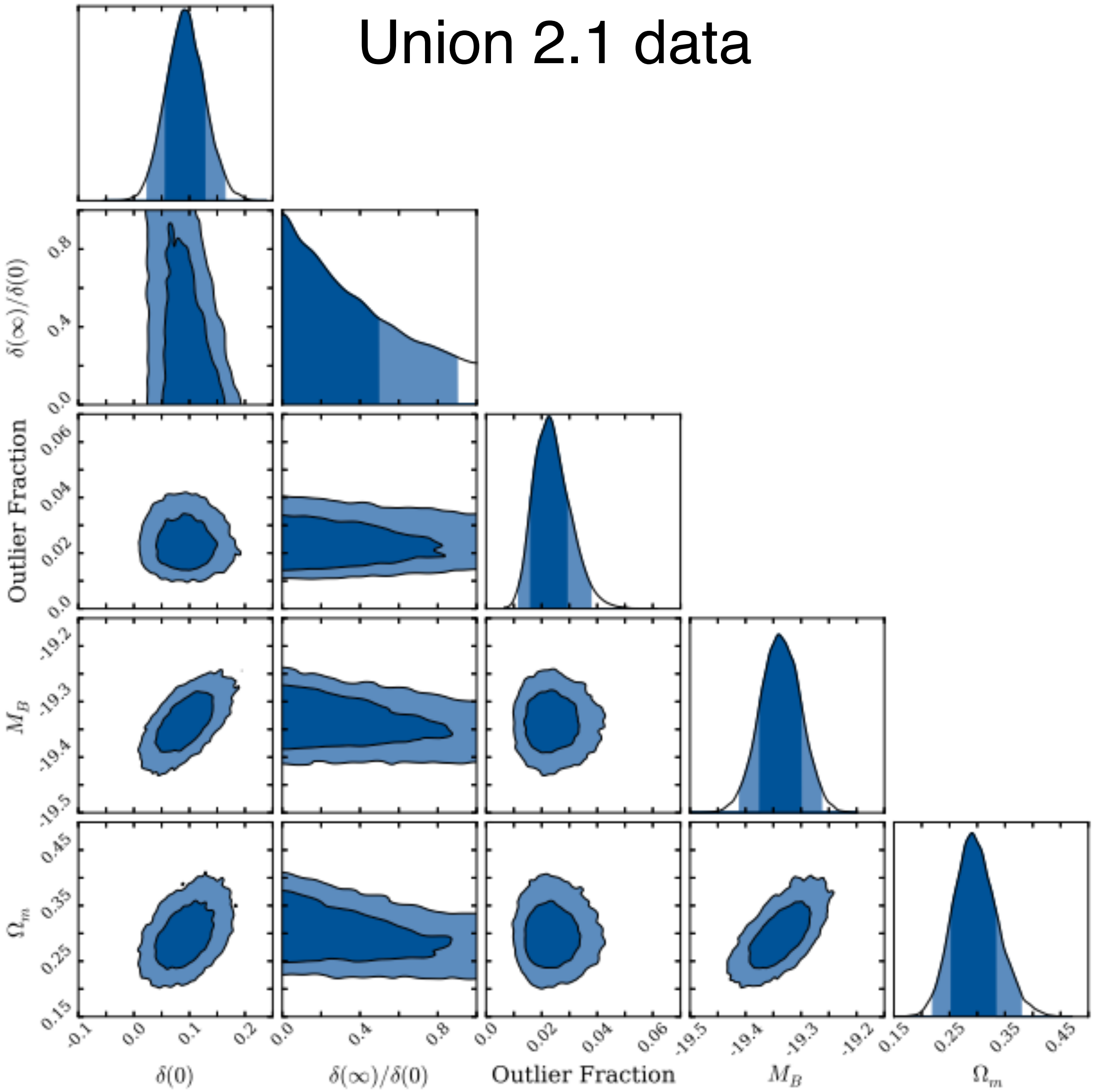
CfA data (215 low-z SNIa)



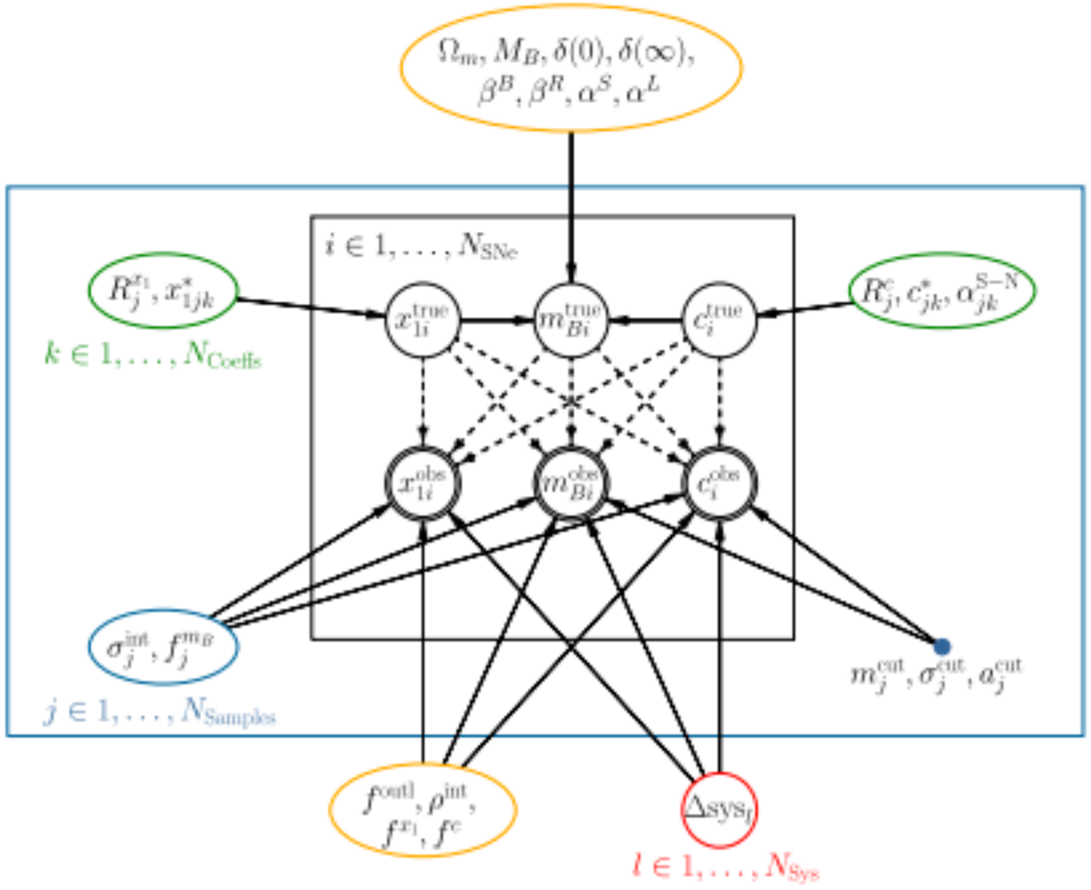
Mandel+16, 1609.04470

UNITY (Rubin+15)

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



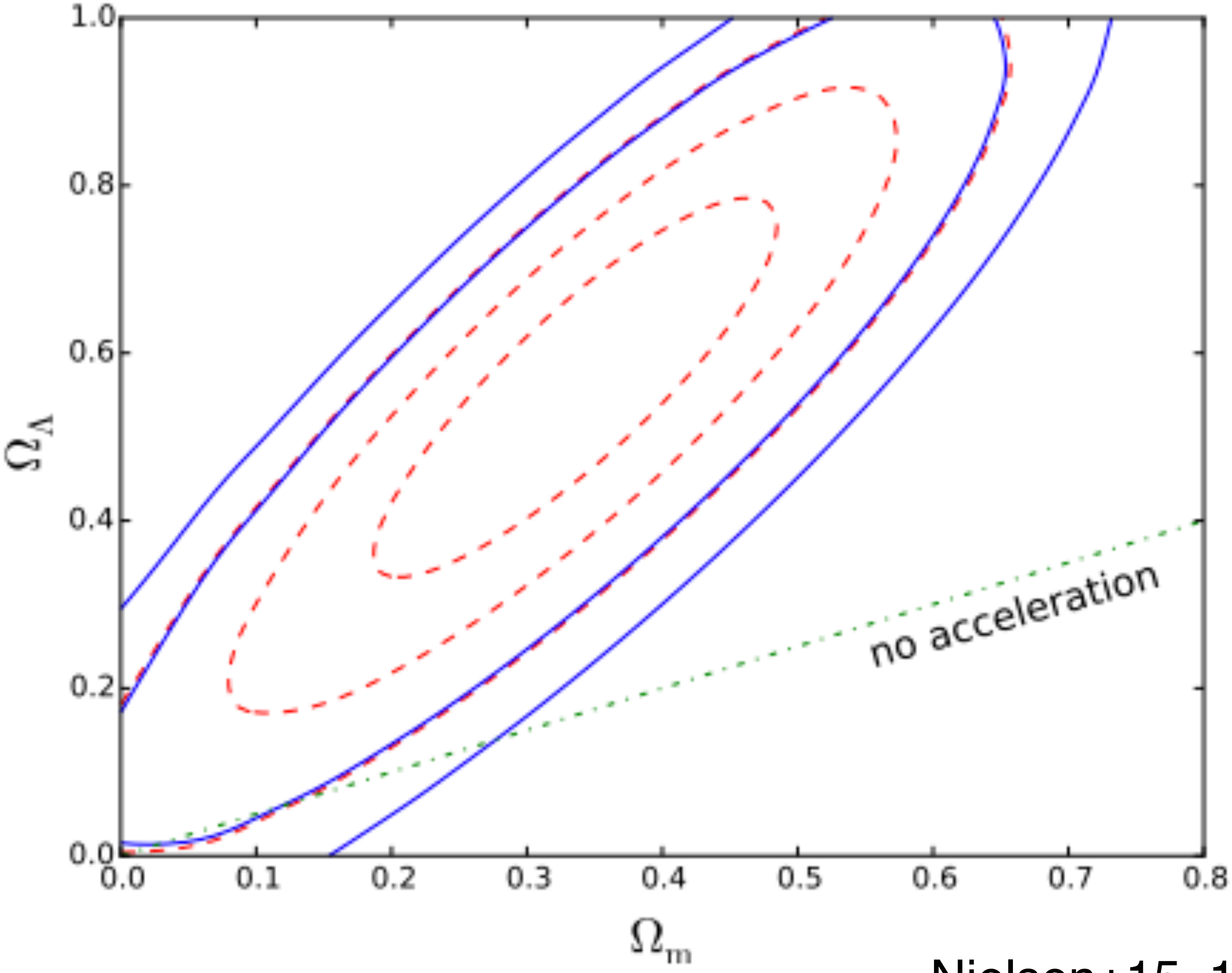
Union 2.1 data



Rubin+15 (SCP) 1507.01602

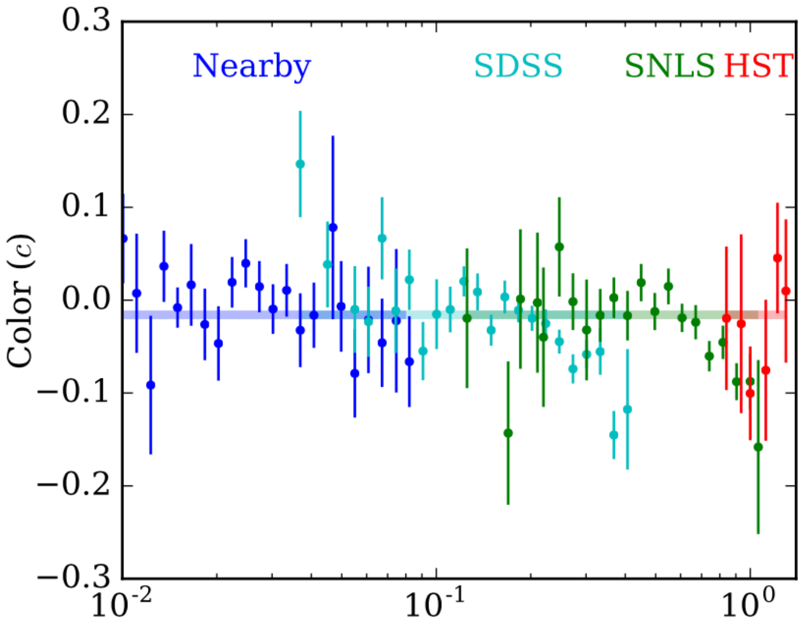
Nielsen+15

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

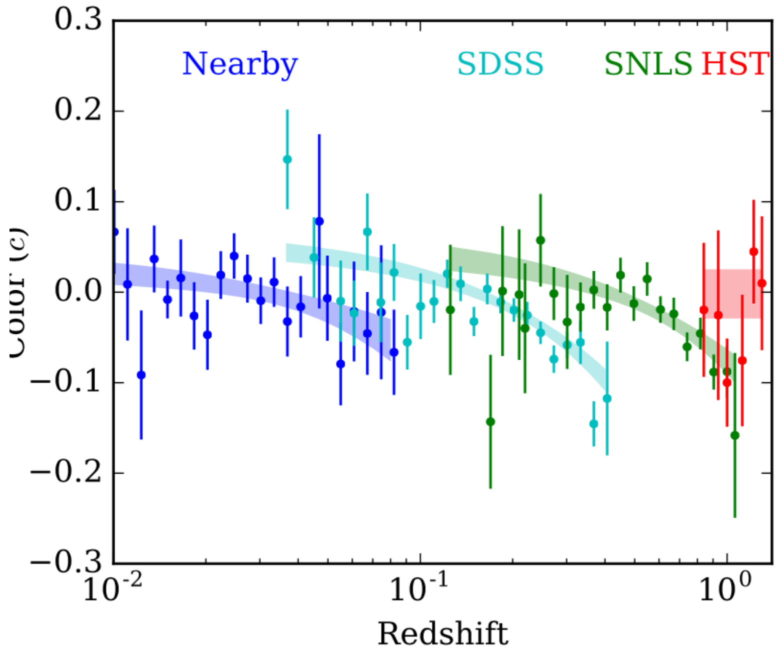


Nielsen+15, 1506.01354

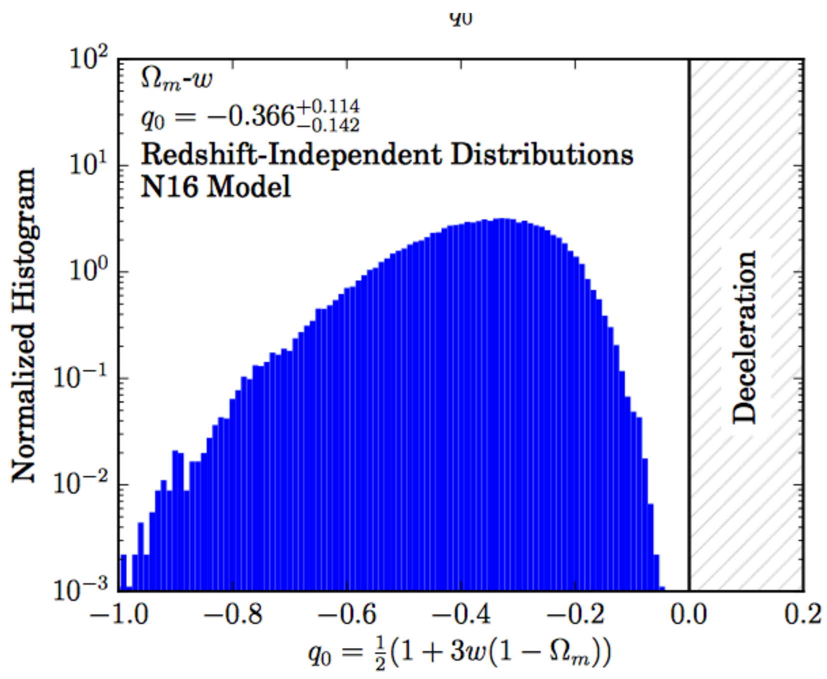
Rubin & Hayden16



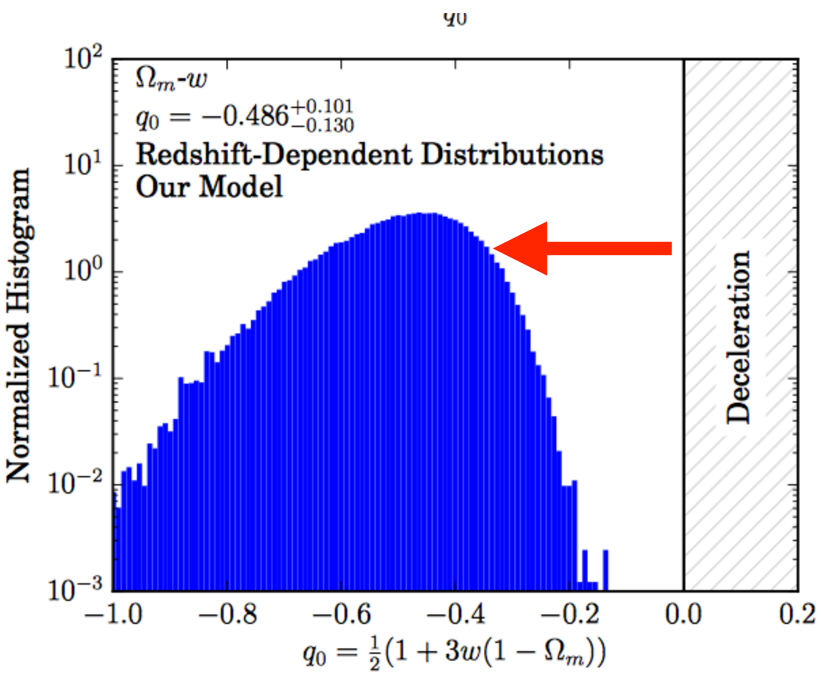
Redshift



Claim that modelling the redshift drift of colour (due to selection effect) moves the cosmological constraints from JLA back to the “standard” values



Deceleration parameter



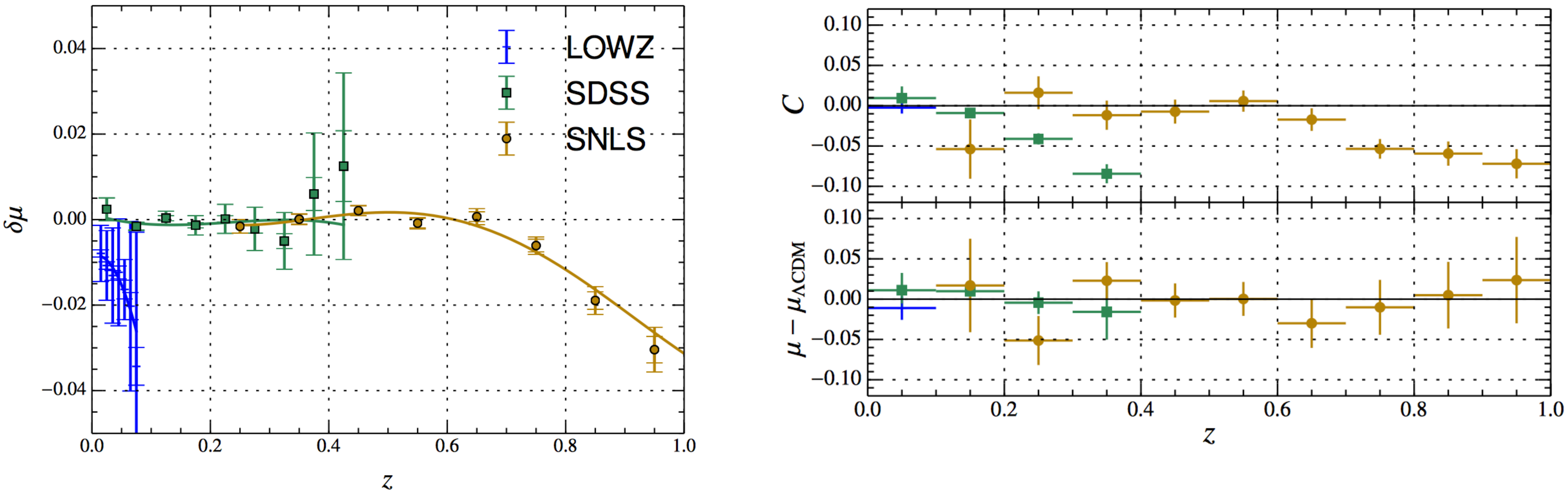
Deceleration parameter

Rubin & Hayden16, 1610.089

Selection Effects

Due to magnitude-limited observations, slow declining and bluer SNIa (i.e., brighter) are observed more easily → Malmquist (1925) bias (i.e., truncation)

Current solution: MC the selection bias, then “correct back” magnitudes to compensate for it



Magnitude

Colour

Betoule et al, Astron.Astrophys. 568 (2014) A22

Bayesian Modelling of Selection Effects

Partition data y into “observed” and “missed”:

$$y = \{y_{\rm obs}, y_{\rm mis}\}$$

$$y_{\rm obs} = \{y_i | I_i = 1, i \in [1, \dots, N]\}$$

$$y_{\rm mis} = \{y_i | I_i = 0, i \in [1, \dots, N]\}$$

The observed data likelihood is obtained by integrating over the missed observations in the complete data likelihood

$$p(y_{\rm obs}, I | \theta, \phi) = \int dy_{\rm mis} p(y, I | \theta, \phi)$$

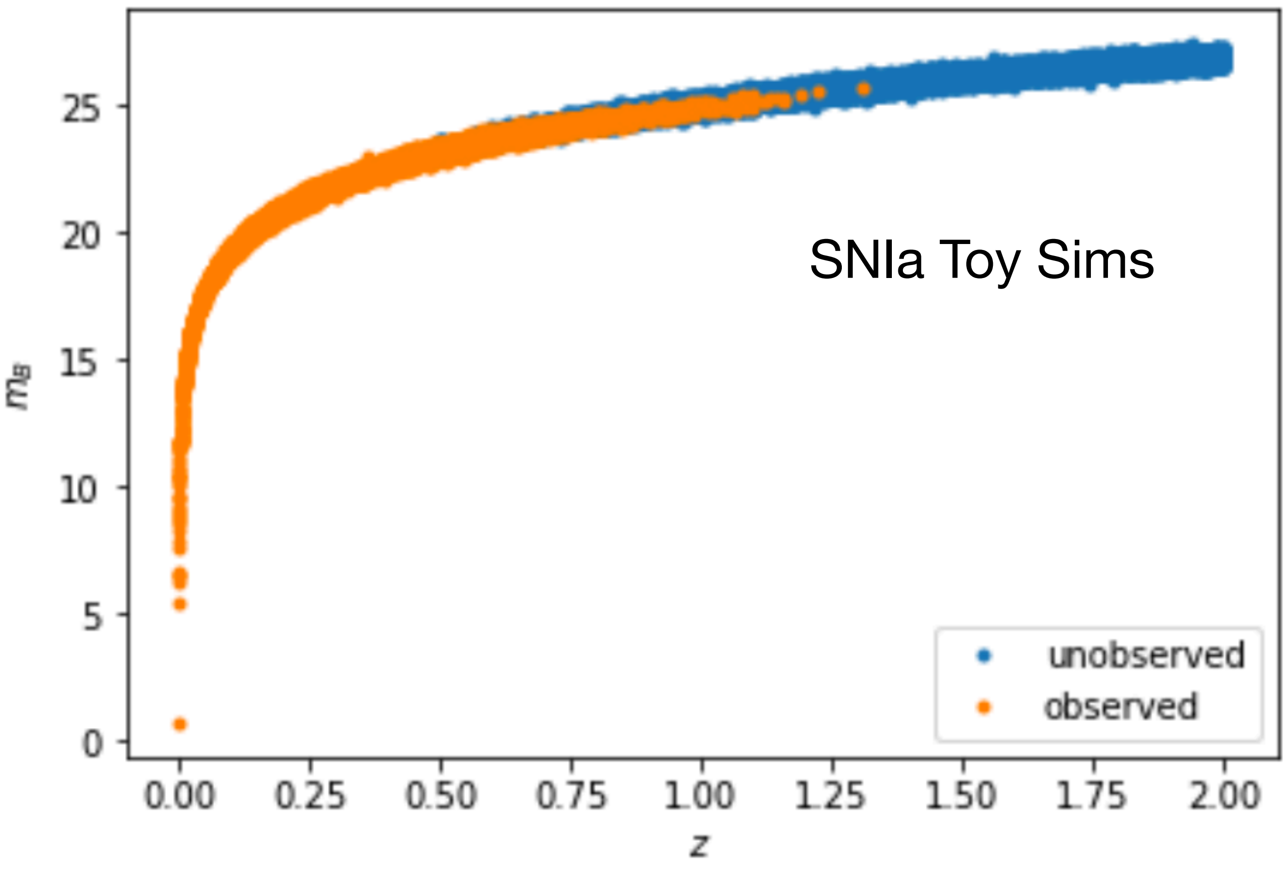
θ=parameters of interest
ϕ=parameters of the data collection procedure

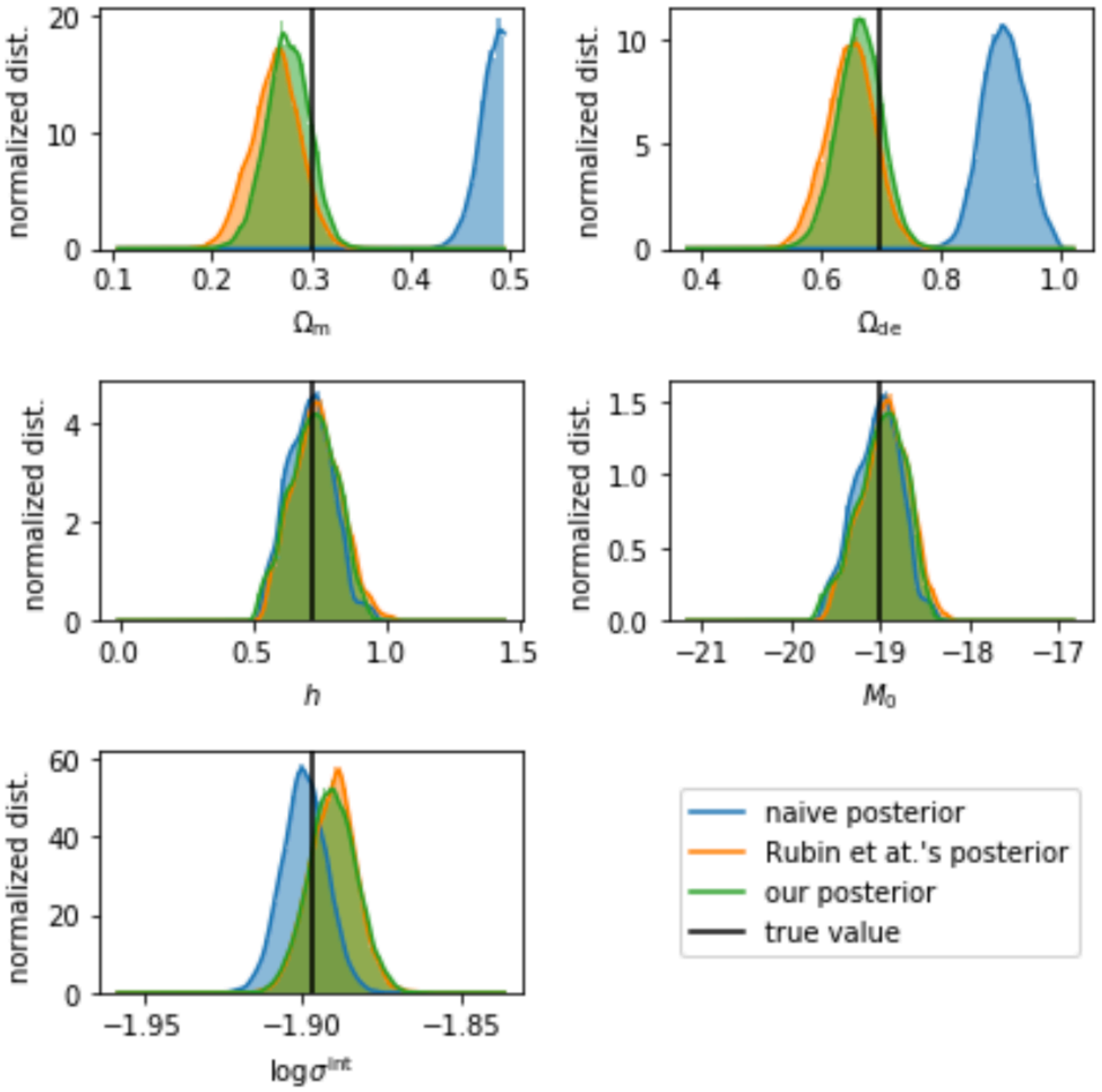
Posterior conditional on the observed data:

$$p(\theta | y_{\rm obs}, I) \propto \int d\phi p(\theta, \phi) \int dy_{\rm mis} p(y|\theta) p(I|y, \phi)$$

Selection Function Exactly Known

If we know the selection function, the procedure of Rubin et al (2015) leads to correct results (so does ours, of course):



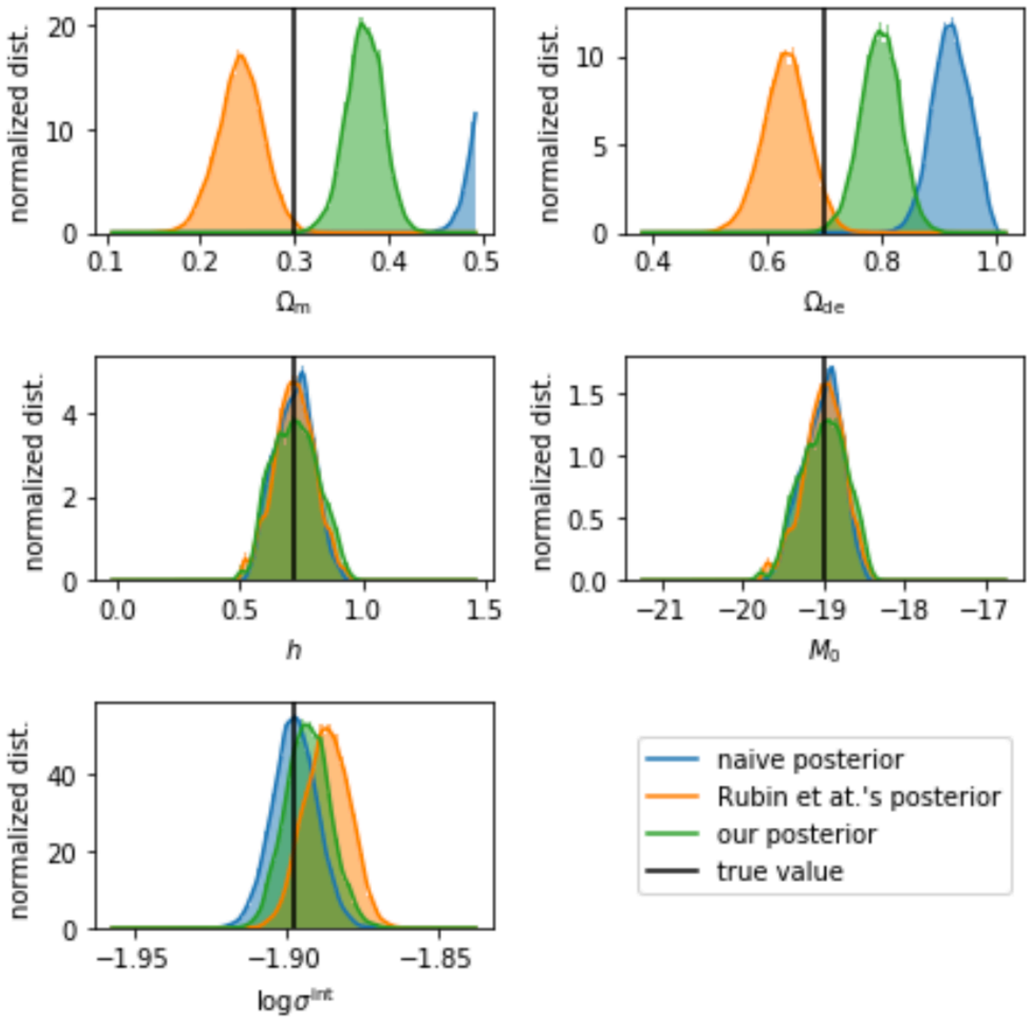


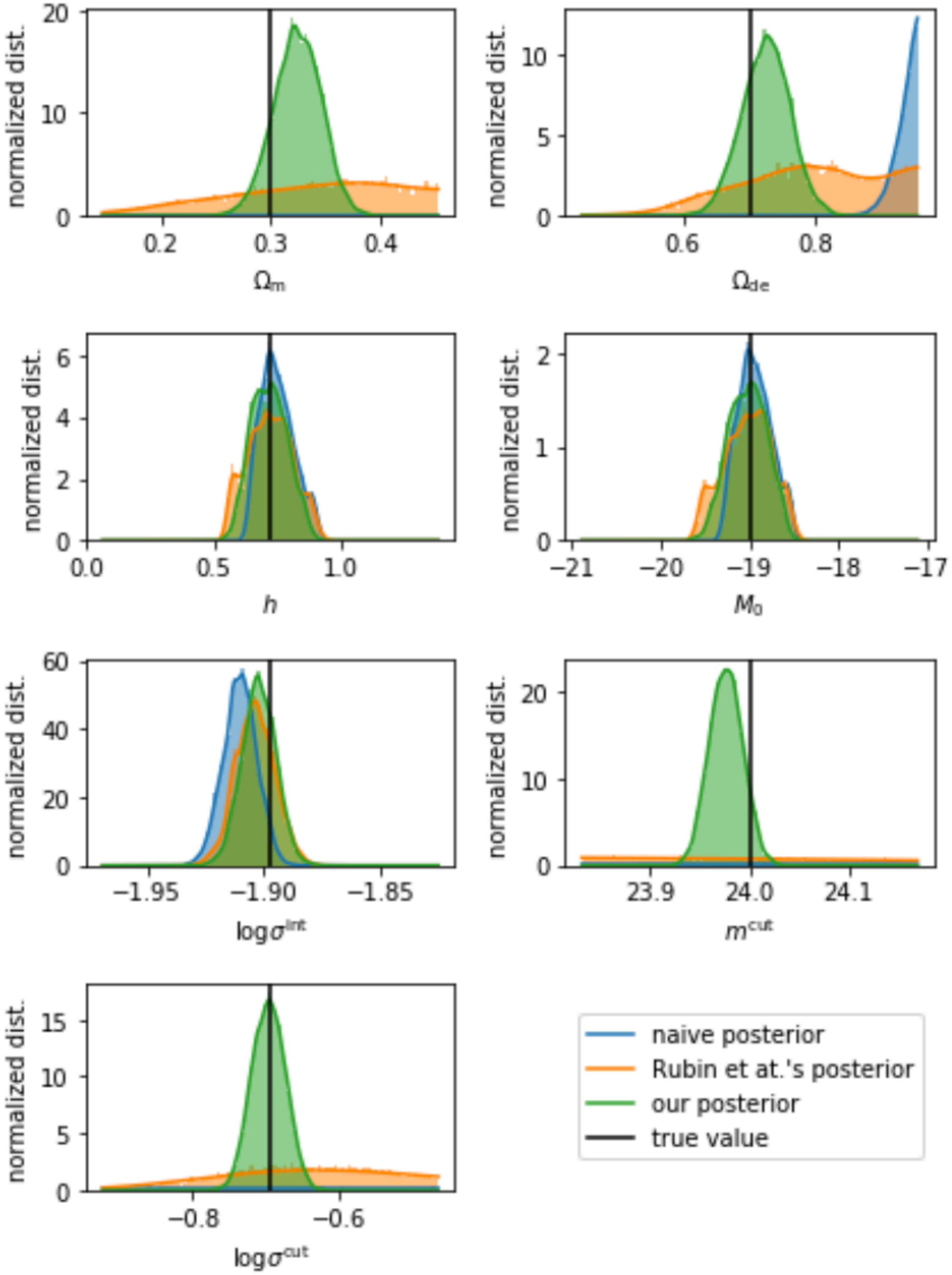
Chen, RT et al (in prep)

Chen, RT et al (in prep)

Selection Function Needs Inferring

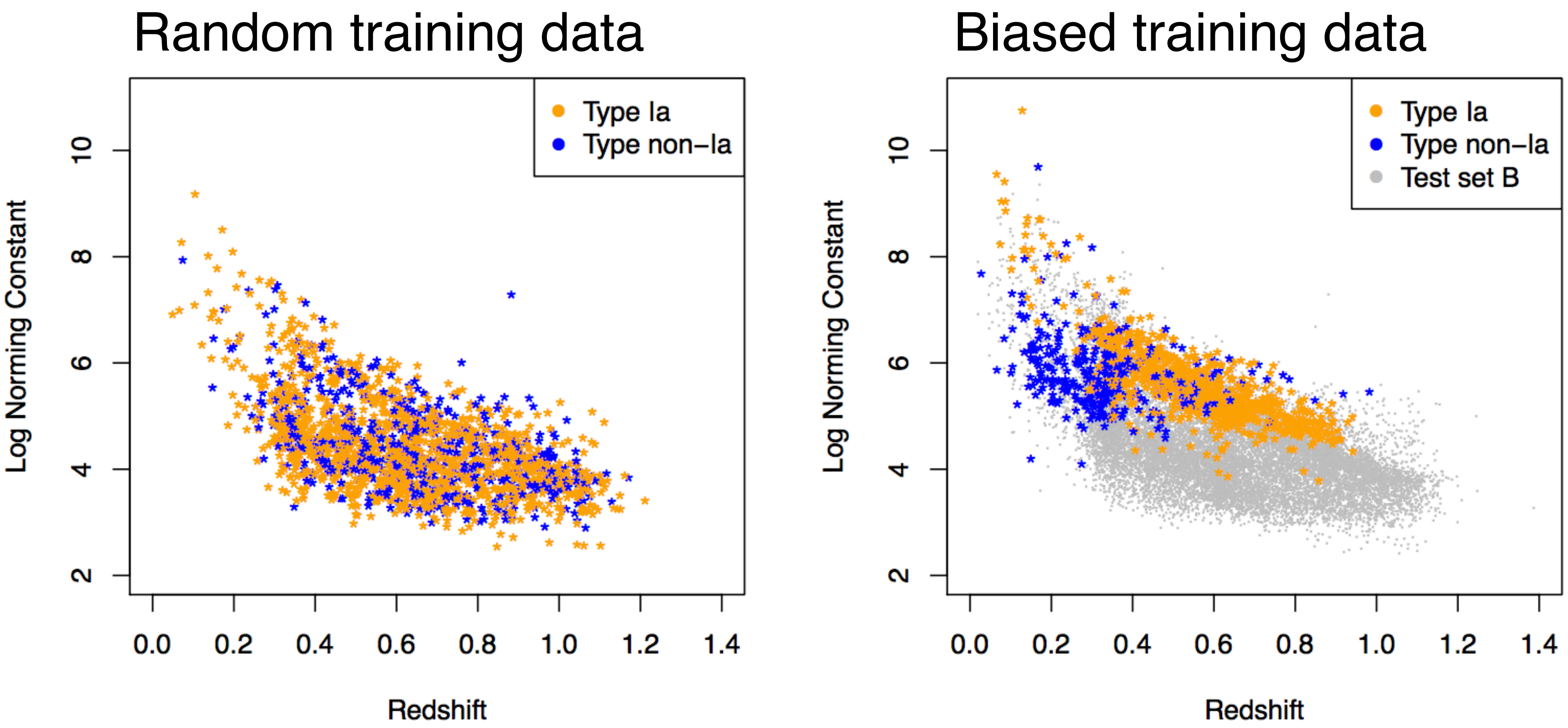
Using wrong parameters of the selection function leads to systematic bias in the cosmological parameters (left).
Solution: Infer the selection function simultaneously (right)





The future: photometric SNIa cosmology

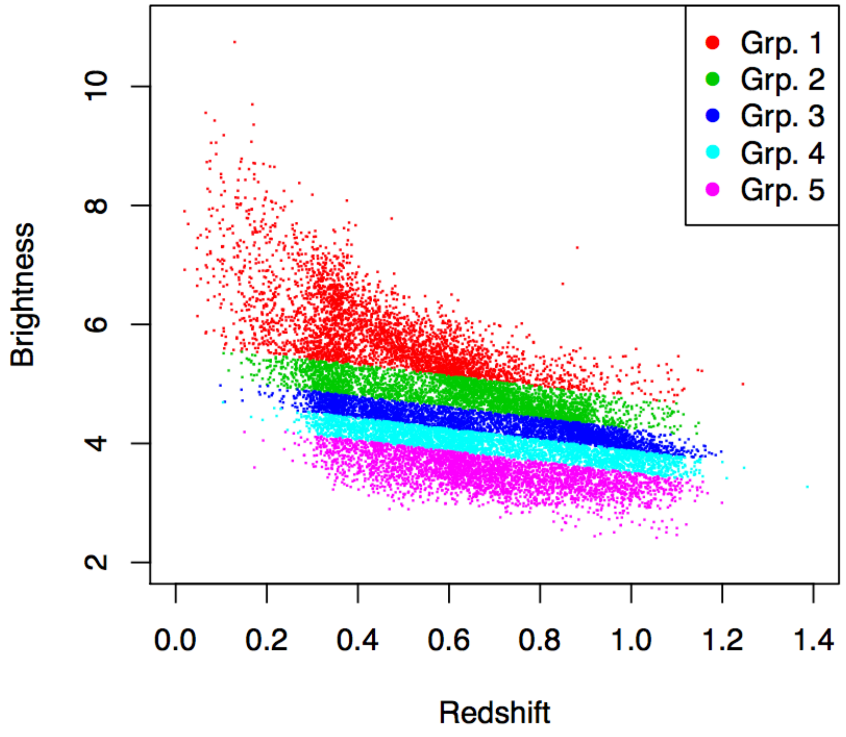
- SNIa identification relies on observationally expensive spectroscopy
- In the future, we won't have spectra for all SNIa candidates (DES: 3000 SNIa over 5 yrs; LSST: 10,000 SNIa/yr)
- SNIa classification will be needed based on multi-band imaging alone -> "SN Classification Challenge" (Kessler+10)
- **Problem:** Training set is biased. Especially at high z, more SNIa's than in the population, hence non representative



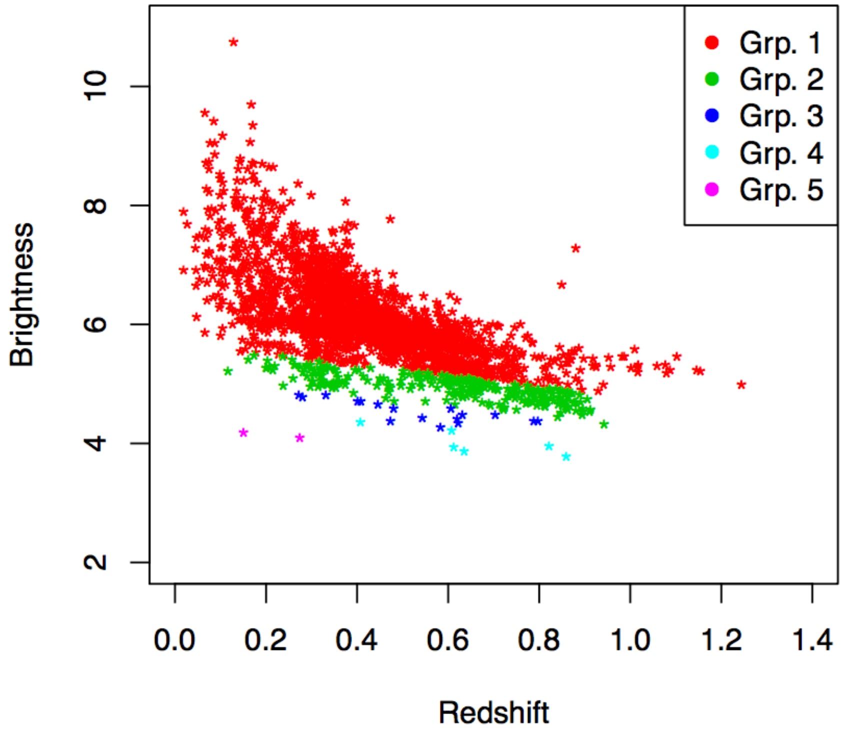
STACCATO

- Our solution (Revsbech, RT, van Dyk in prep): **S**yn**T**hetically **A**ugmented Light **C**urve **C**lassific**AT**i**O**n proceeds as follows:
  - Fit light curve with Gaussian Process (GP)
  - Compute Diffusion Map (to quantify similarities between LCs), Richards+12
  - Perform Random Forest Classification
  - **New:** Group SNs according to Propensity Score (probability of belonging to the training set) to reduce bias between training and test set

Test data partitioning



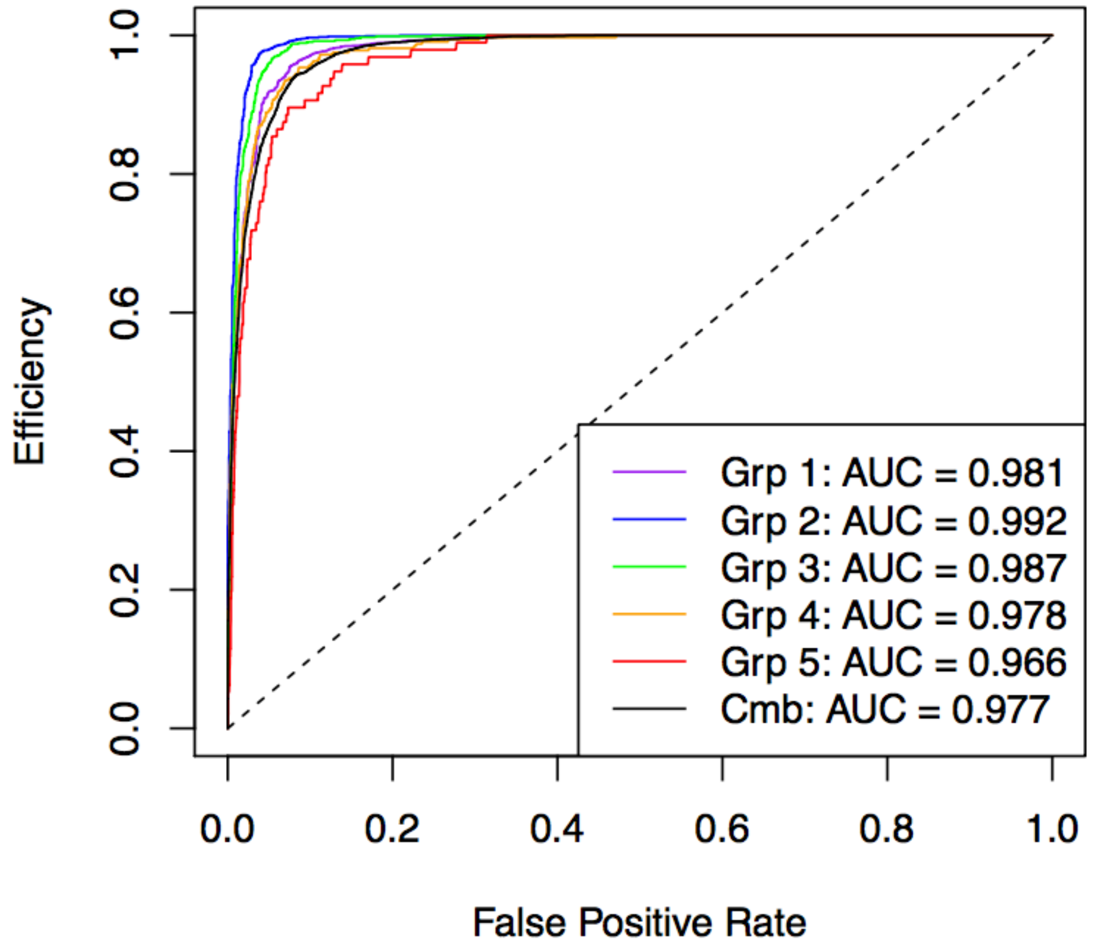
Training set partitioning



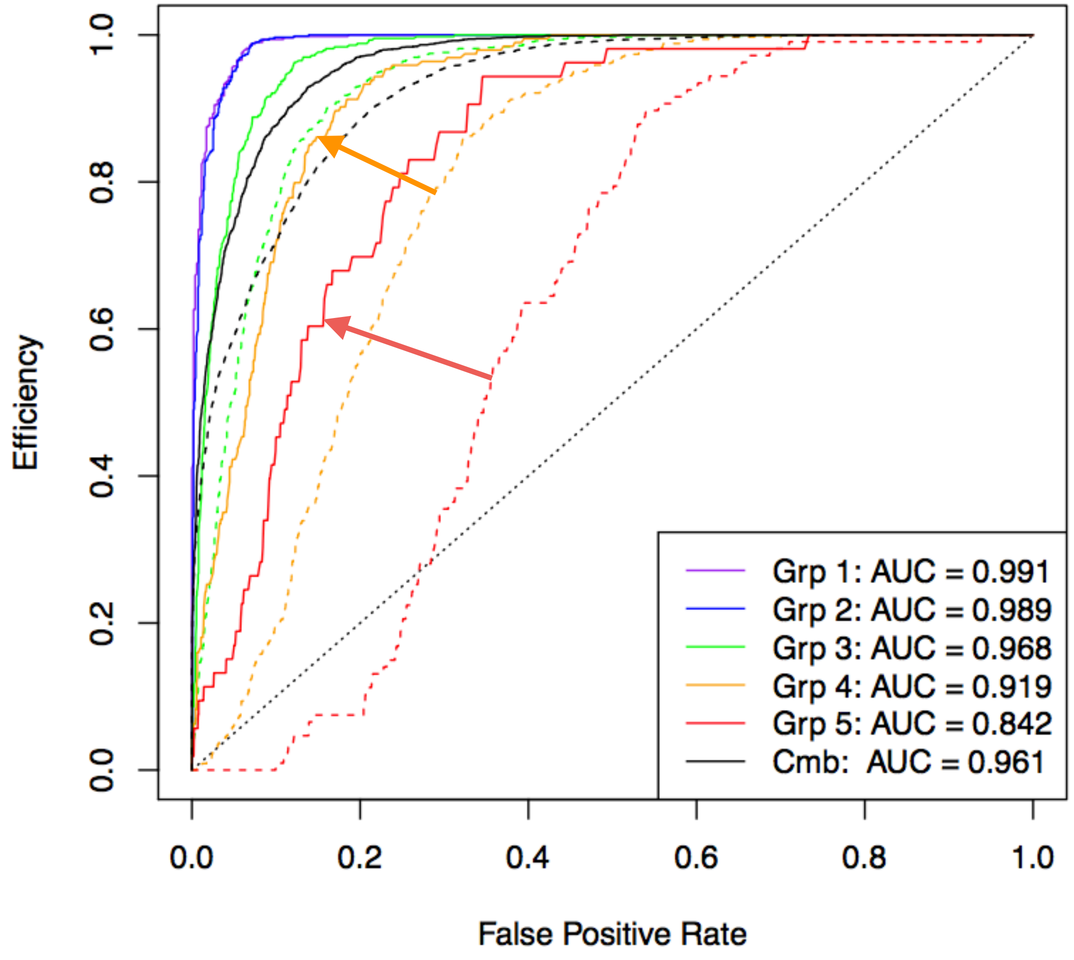
Augmenting LCs via GP resampling

- The final step in STACCATO is to augment the training set by synthetically sampling LC’s from the fitted GP according to the Propensity Scores.
- Evaluated using Area under the ROC Curve (AUC):
  - ‘Gold Standard’ (unbiased training set) = **0.977** vs STACCATO = **0.961**

Gold Standard



STACCATO
dashed (solid) w/o (w) augmentation

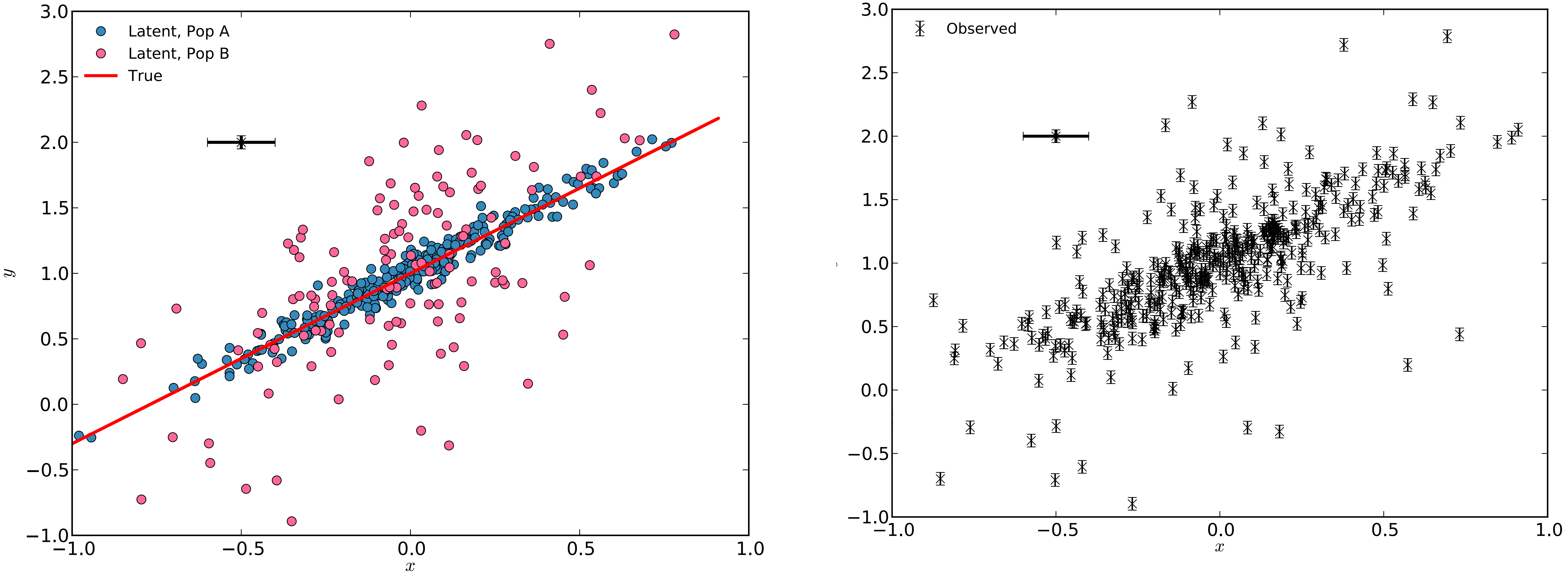


Object-by-object classification: example

- “Events” come from two different populations (with different intrinsic scatter around the same linear model), but we ignore which is which:

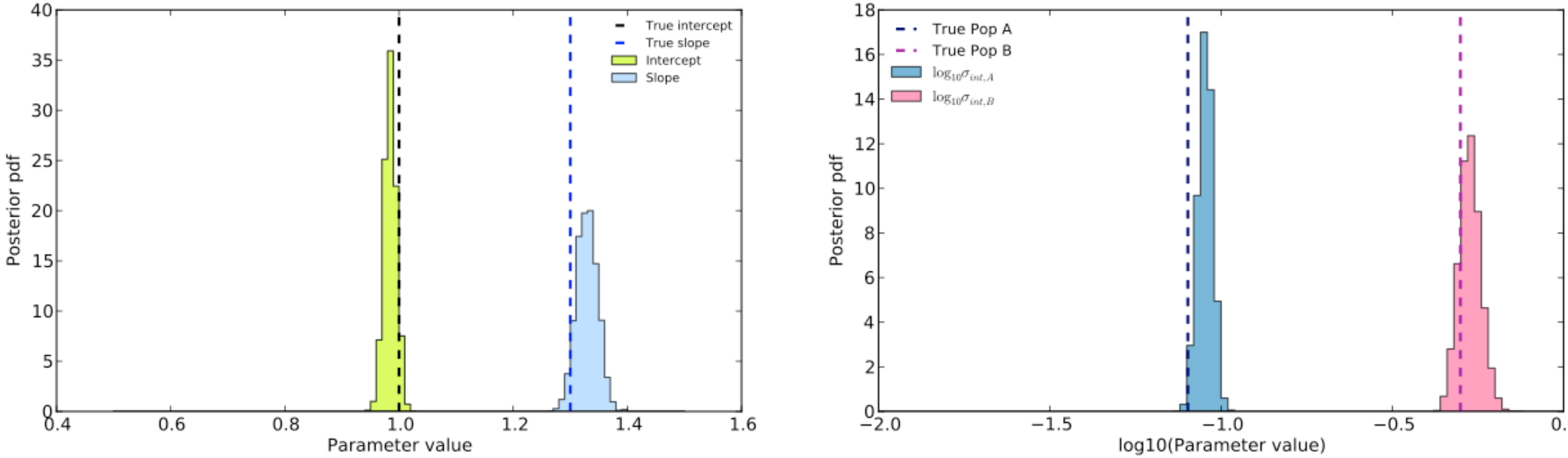
LATENT

OBSERVED

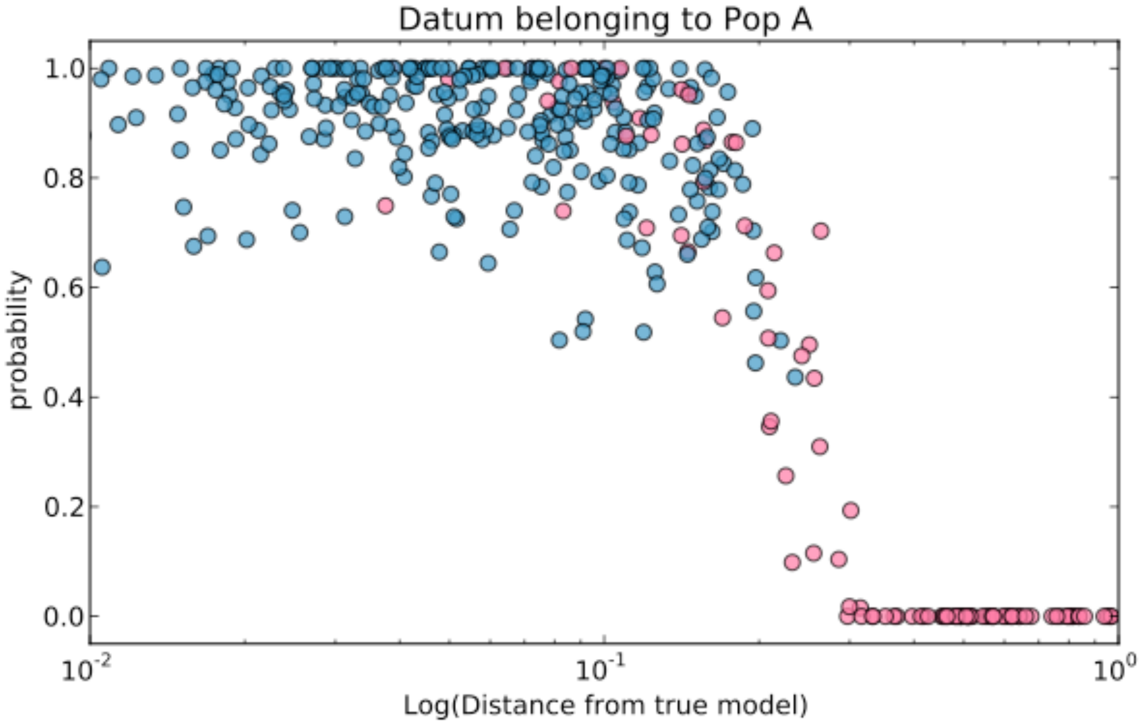


Hierarchical model reconstruction

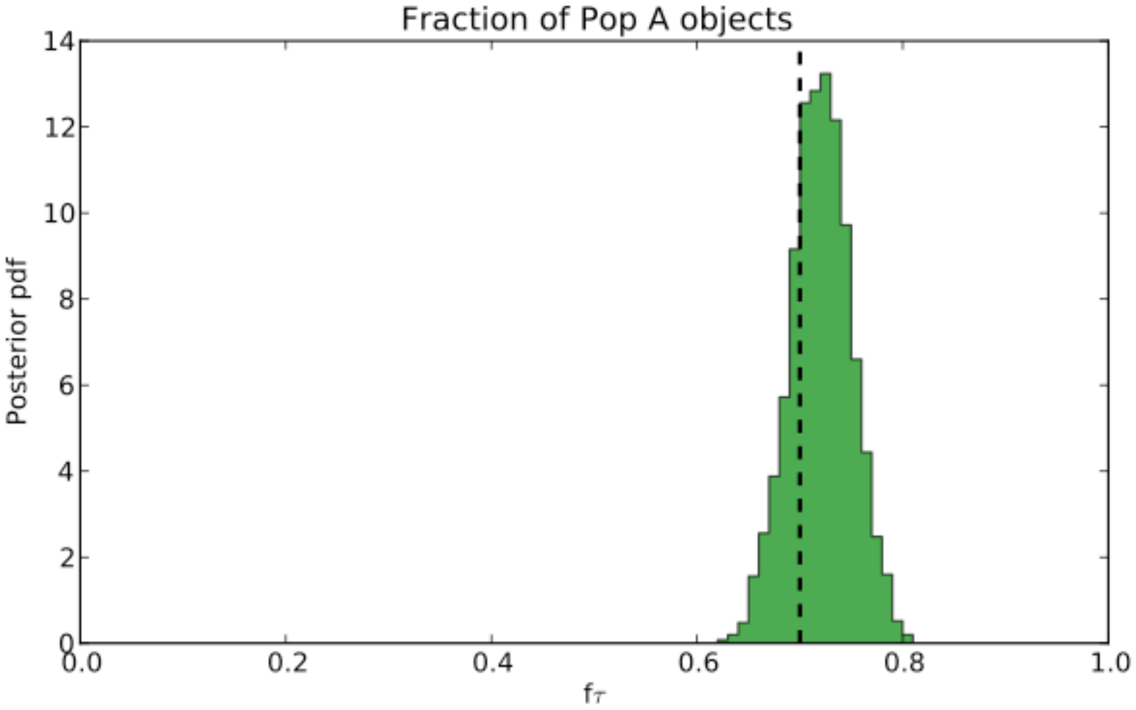
Parameters of interest



Classification of objects



Population-level properties



SUMMARY

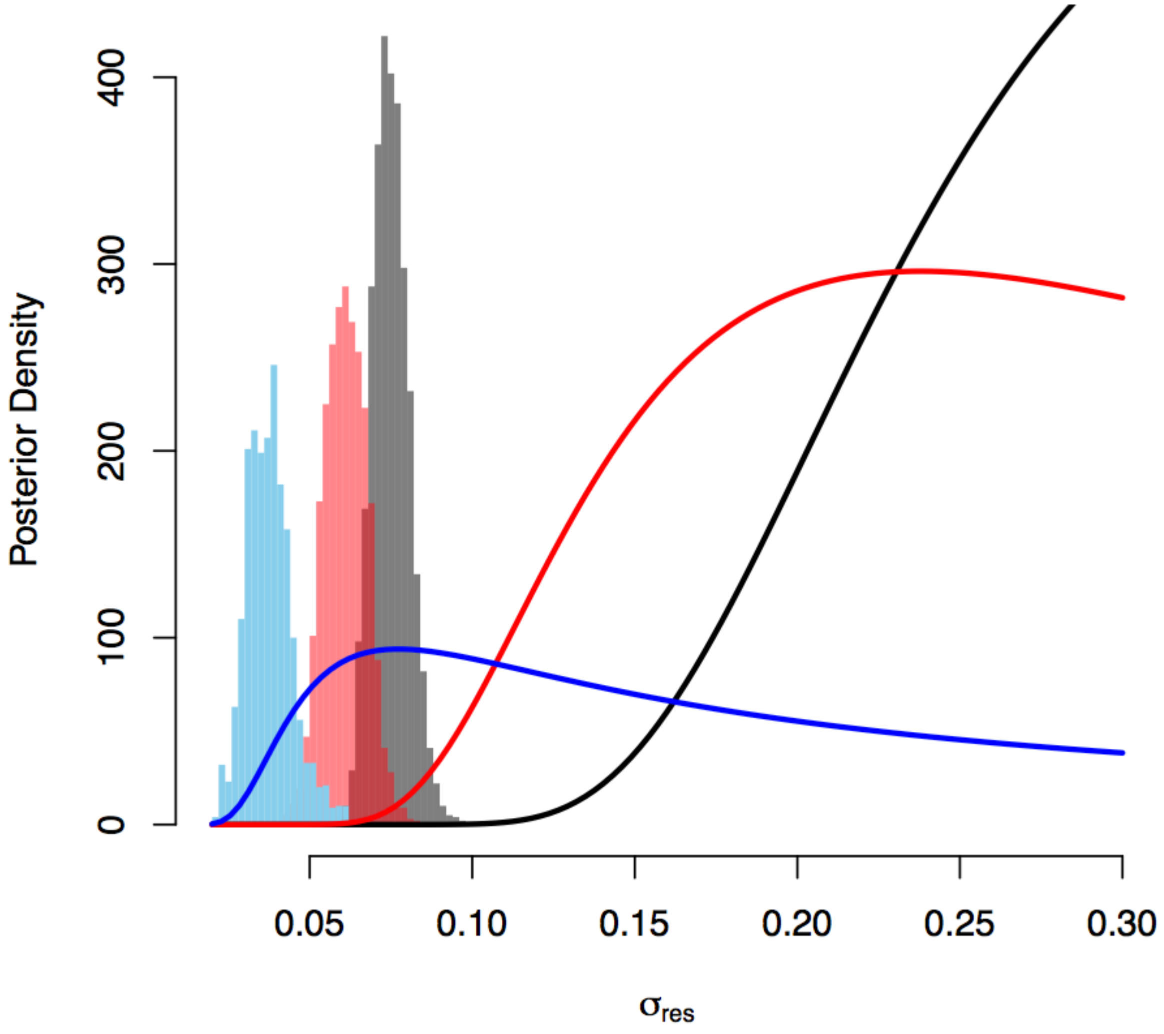
@R_TROTTA

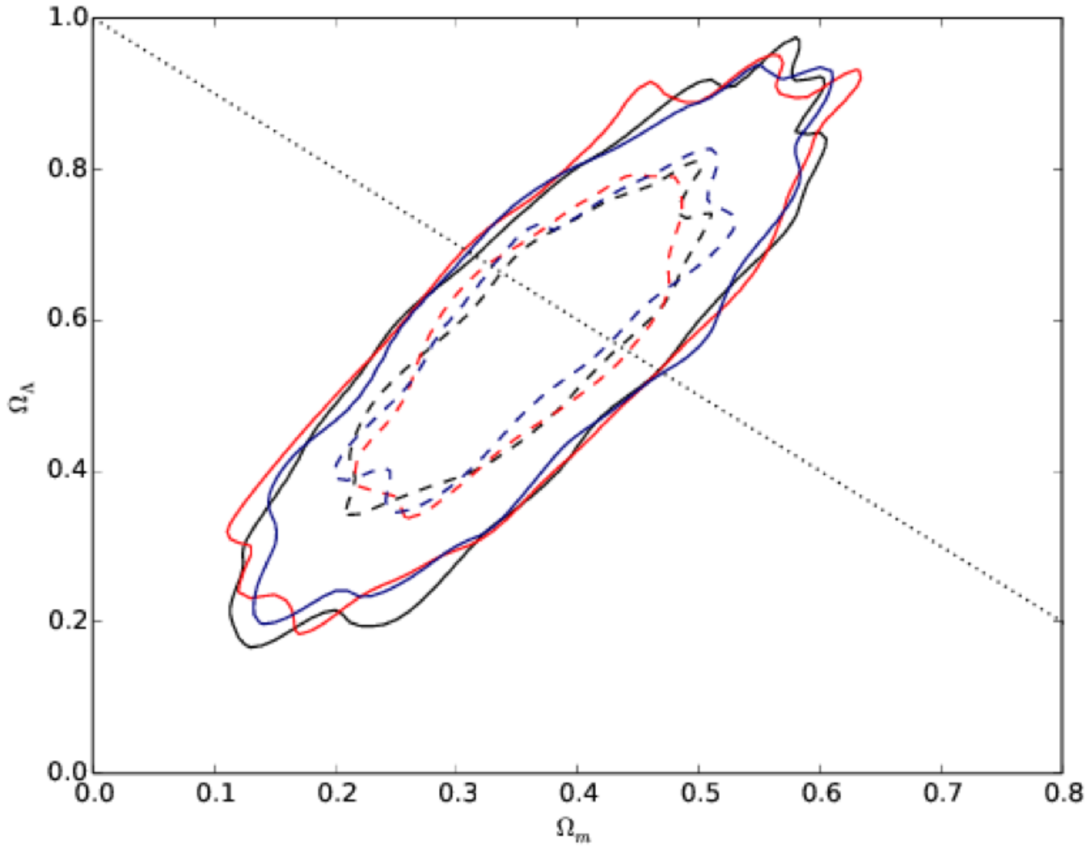
THANK YOU!

- SNIa cosmology has reached maturity: Further advances hindered by "systematics" that need to be modelled explicitly. From Precision to Accuracy.
- A Bayesian hierarchical model approach (BAHAMAS) yields strangely discrepant results wrt the standard analysis
- We find ~ 2-sigma discrepancies in w, Ω_m, Ω_κ
- SNIa's in the outskirts of galaxies are potentially better standard candles
- The future of SNIa's cosmology requires more sophisticated statistical approaches being implemented

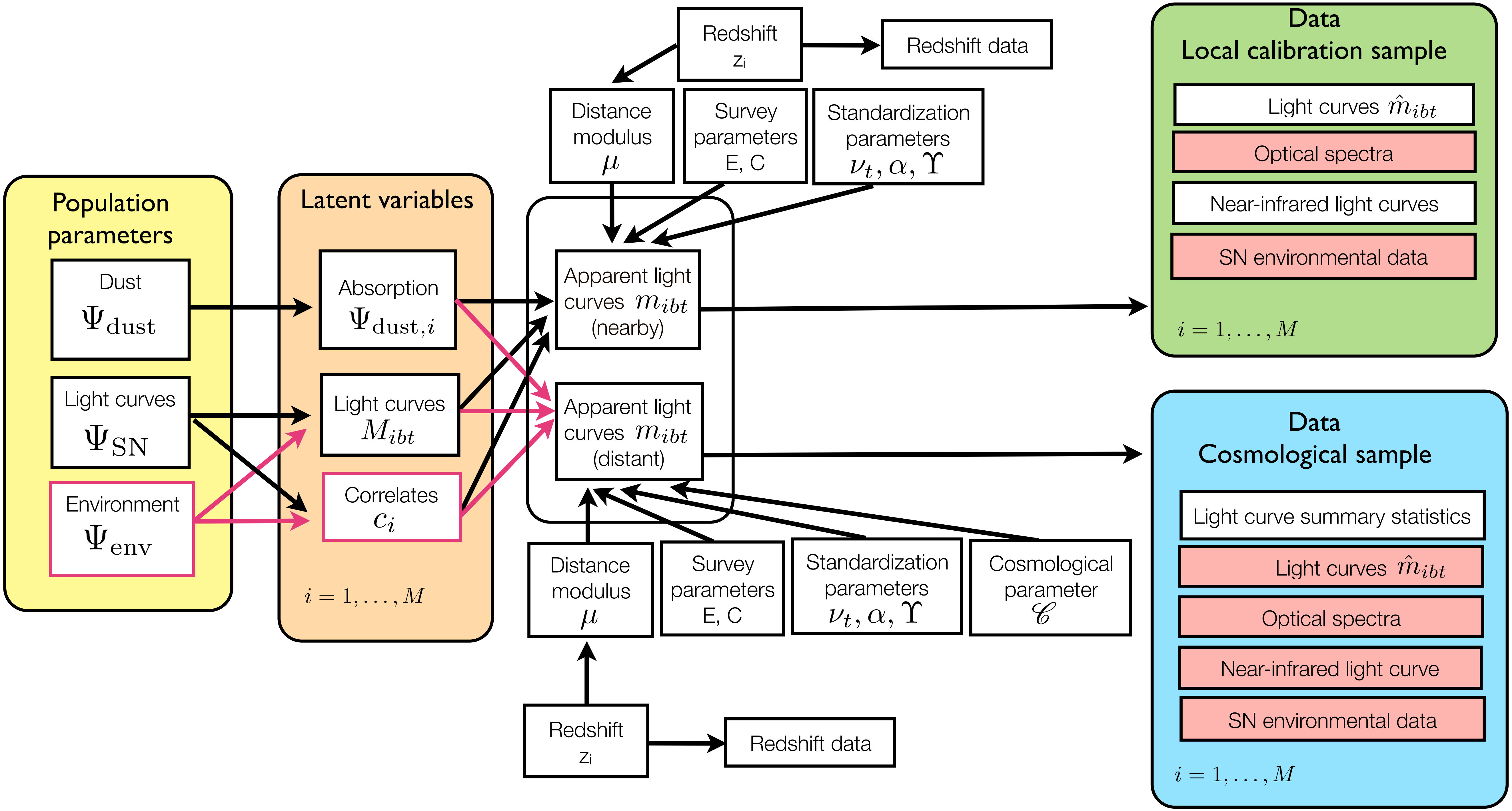
BACKUP SLIDES

Prior robustness





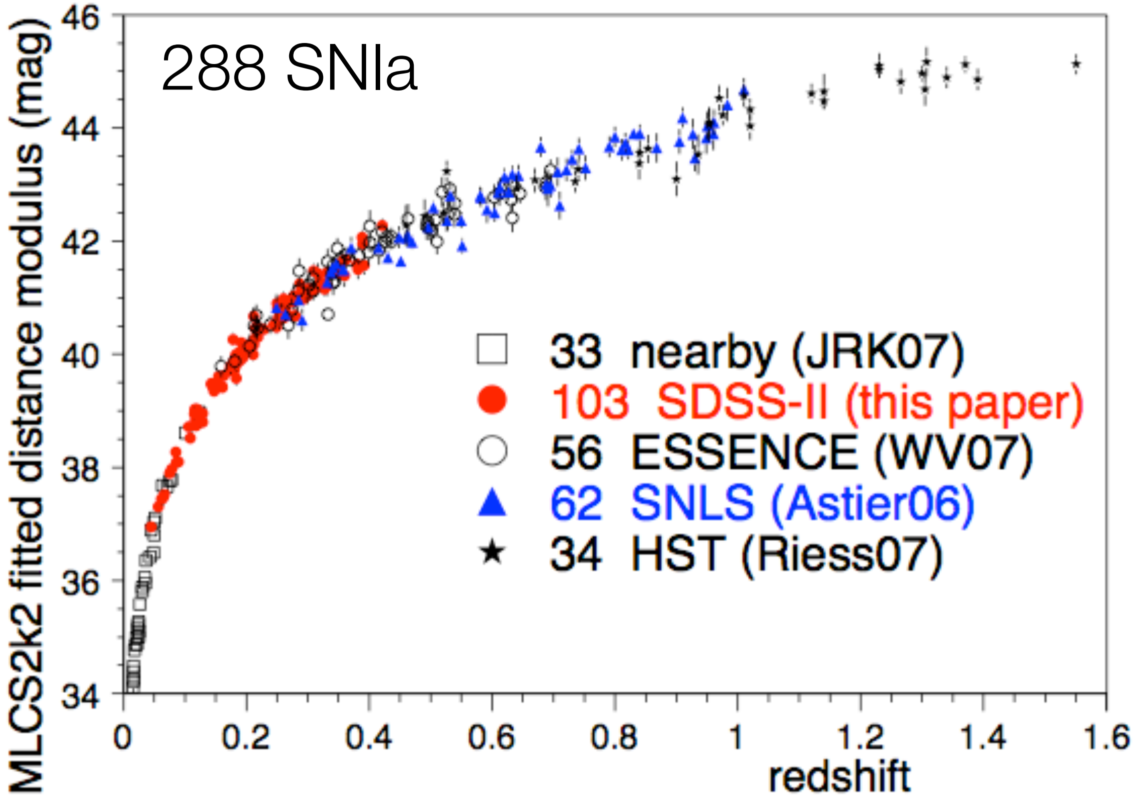
The complete hierarchical model



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

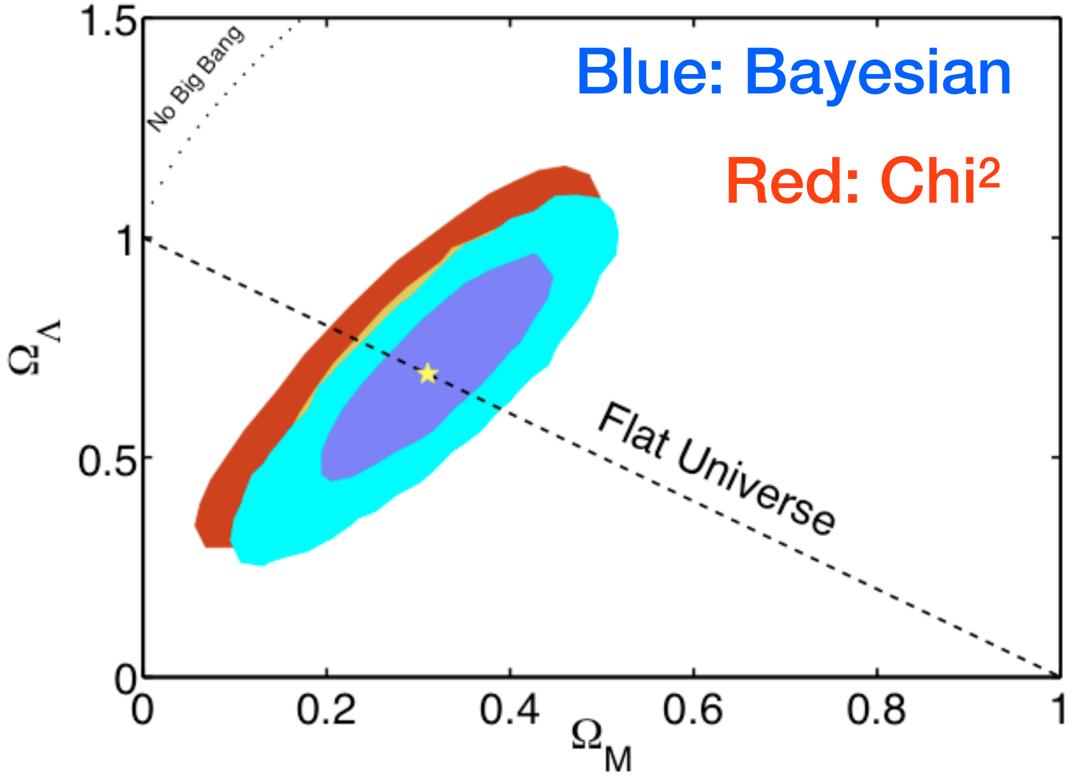
Cosmology results (Union)

Combined sample



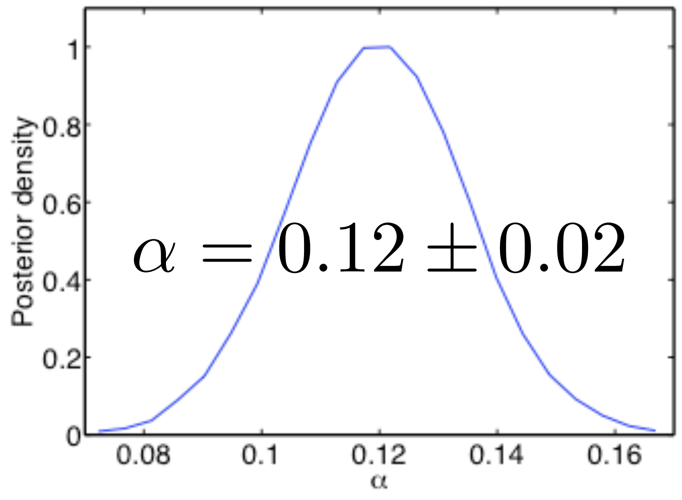
Kessler et al (SDDS collaboration) (2010)

w = 1

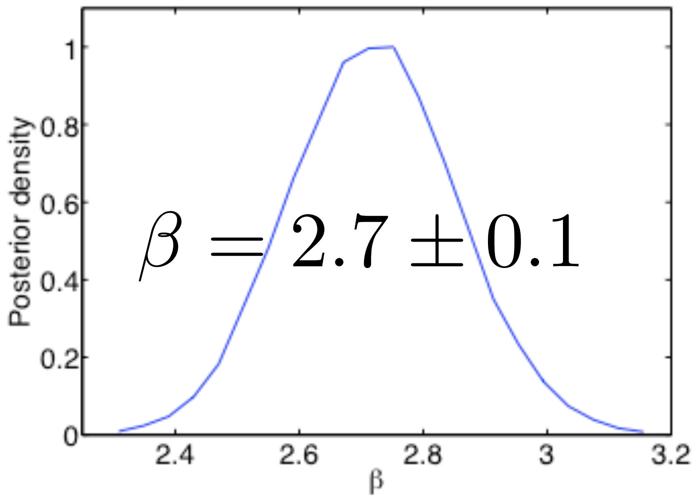


March et al (2011)

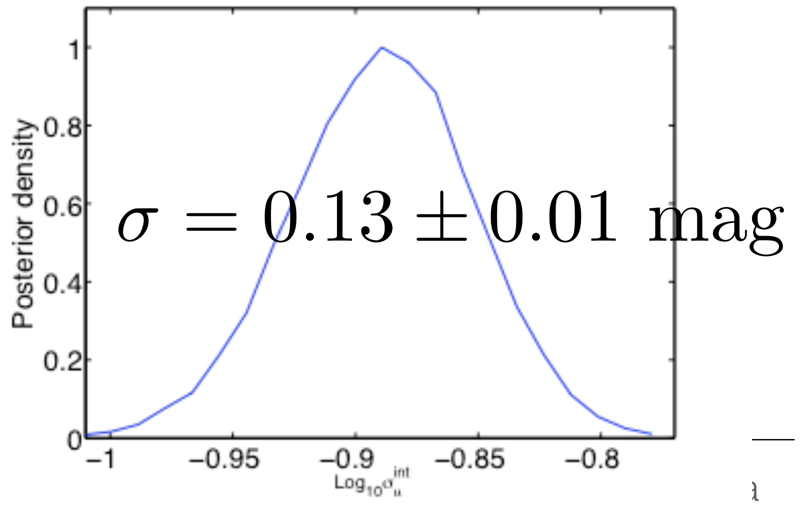
Marginal posteriors



$\alpha = 0.12 \pm 0.02$



$\beta = 2.7 \pm 0.1$

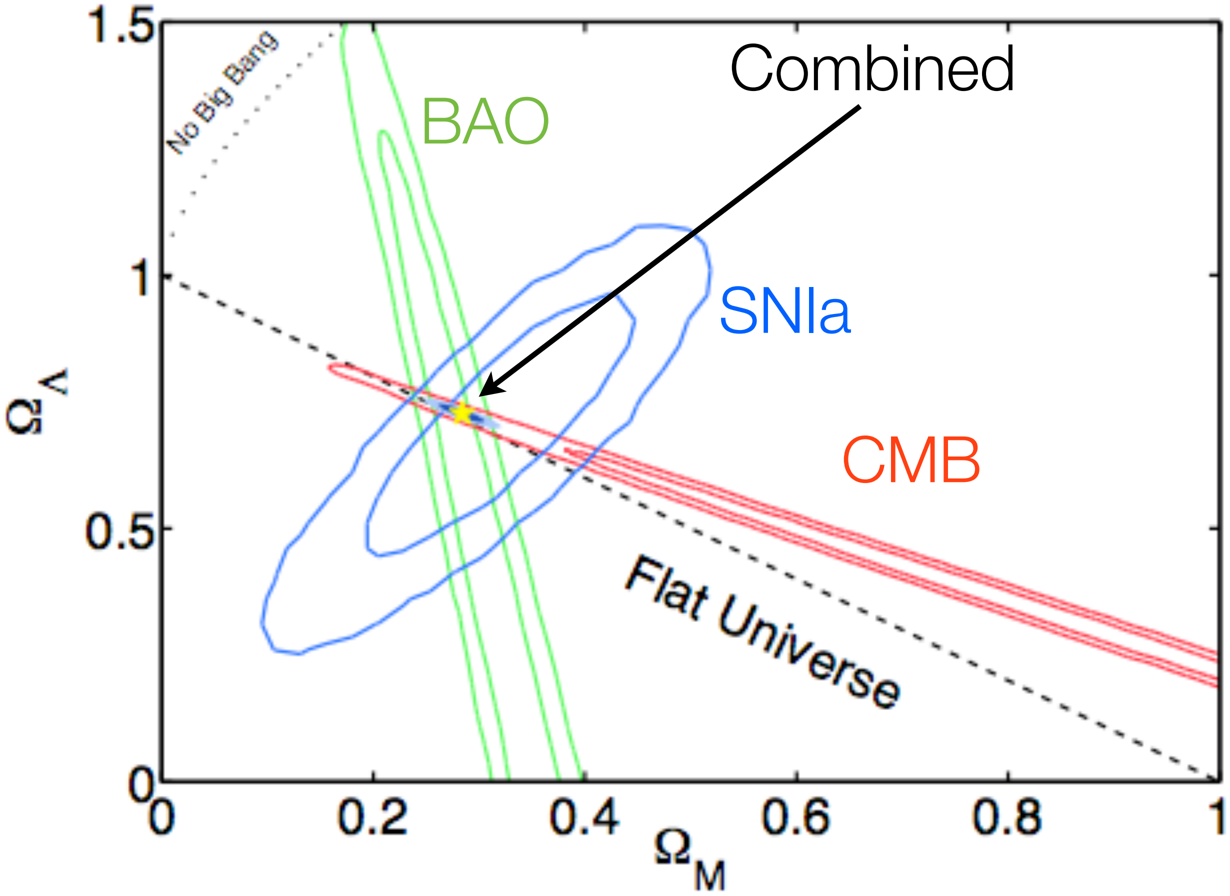


$\sigma = 0.13 \pm 0.01$ mag

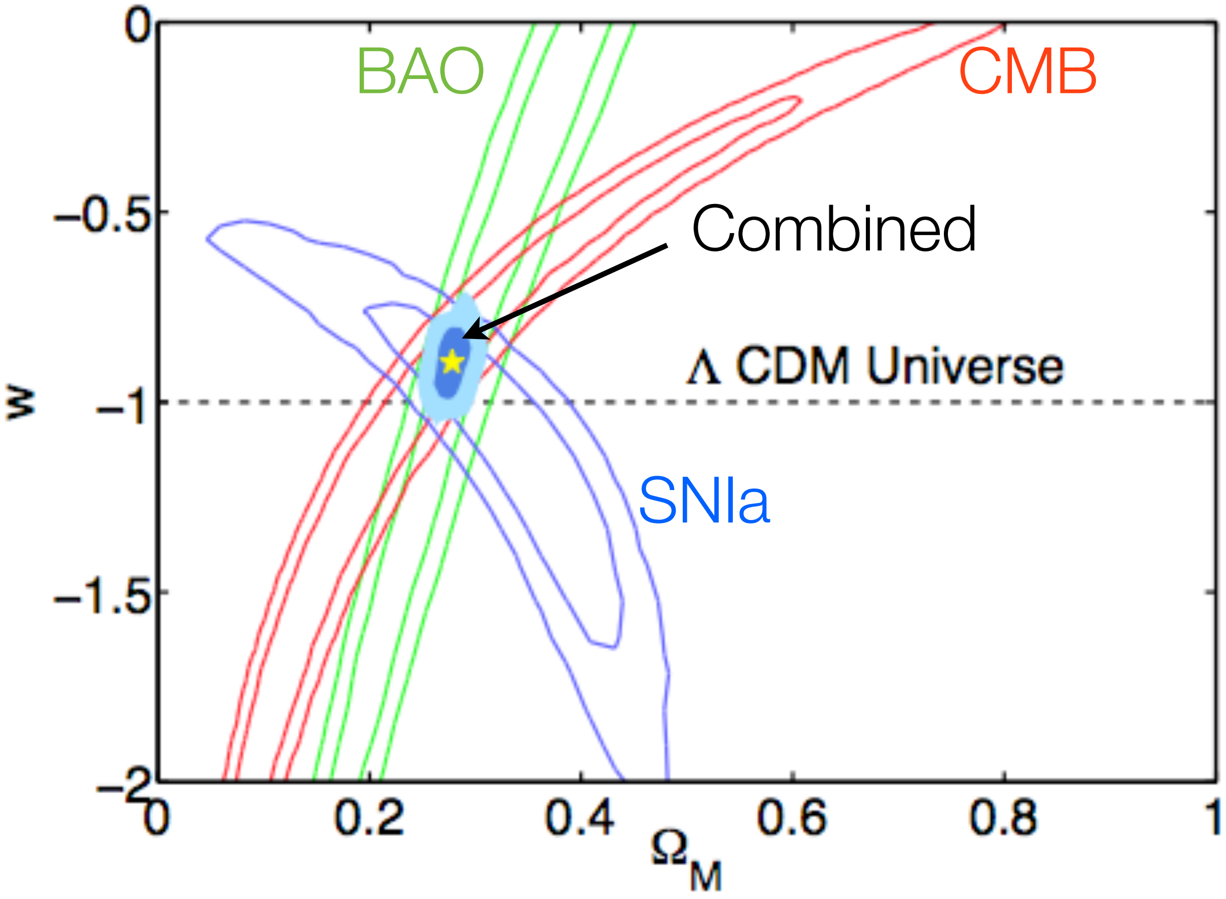
Combined constraints

- Combined cosmological constraints on matter and dark energy content:

w = 1



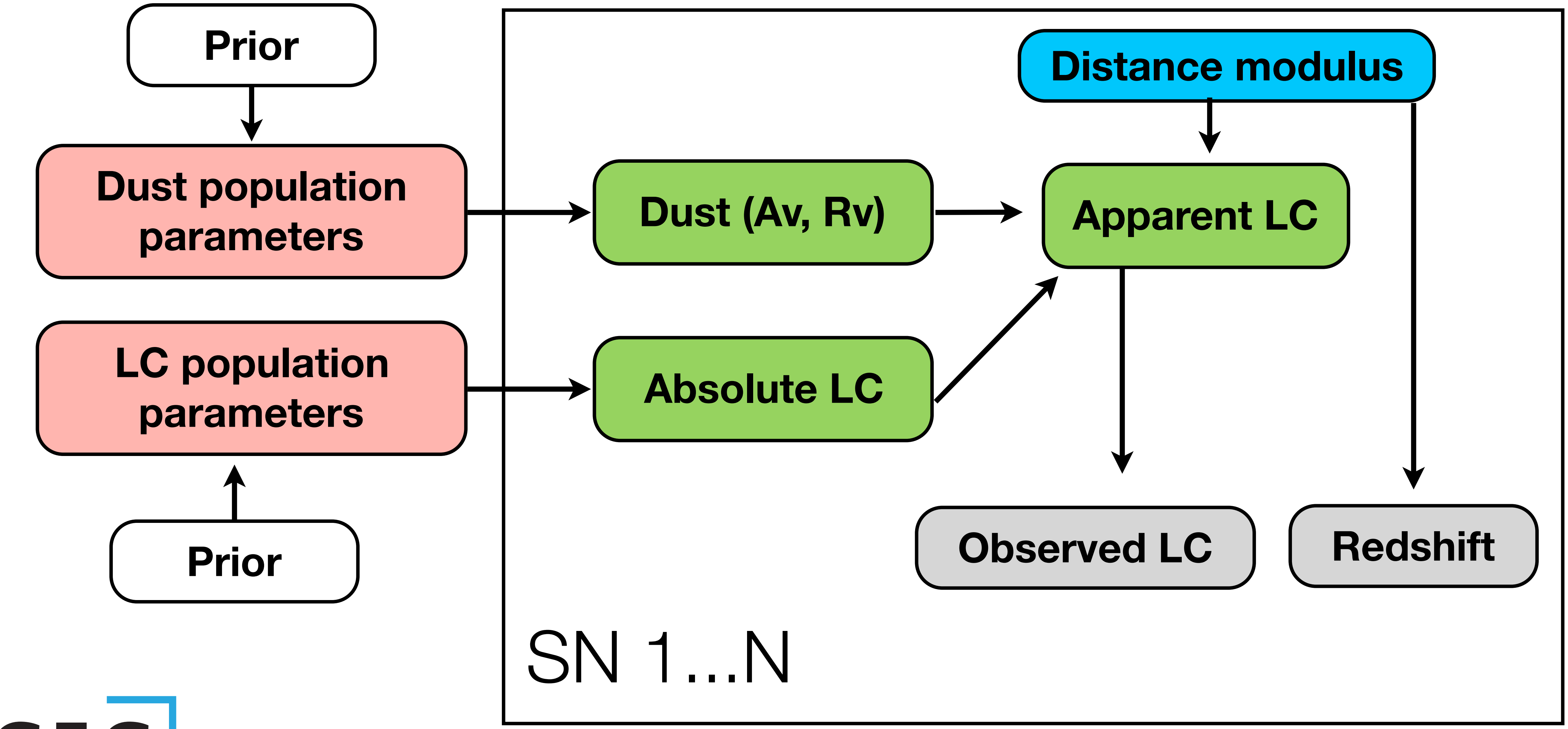
$\Omega_K = 0$



March, RT et al (2011)

The BayeSN approach

- 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



Prior

Dust population parameters

LC population parameters

Prior

Distance modulus

Dust (Av, Rv)

Apparent LC

Absolute LC

Observed LC

Redshift

SN 1...N

From lightcurves to distances

- 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.

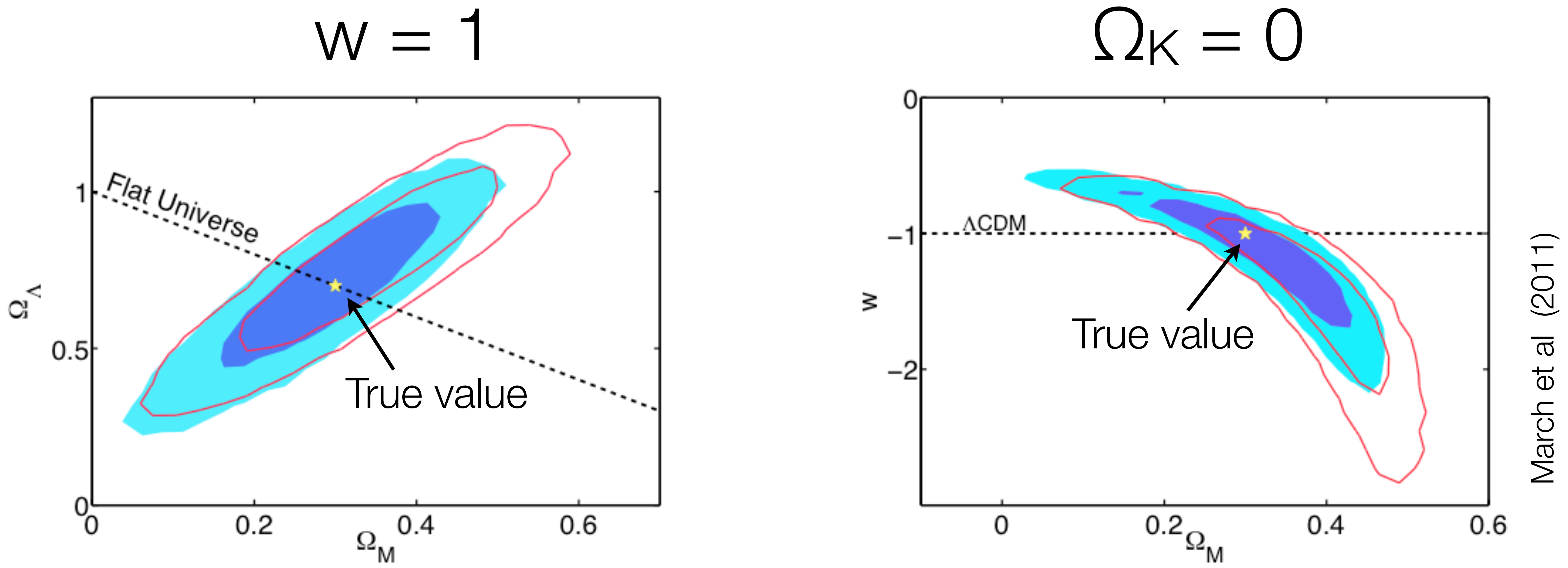
- **SALT/SiFTO/SALT2** (Guy et al, 2007): LC shape (x_1) and colour (c) correction extracted from LC alongside apparent B-band magnitude (m_B) + covariance matrix. The distance modulus

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

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).

Marginal posterior (simulated data)



March et al (2011)

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"?