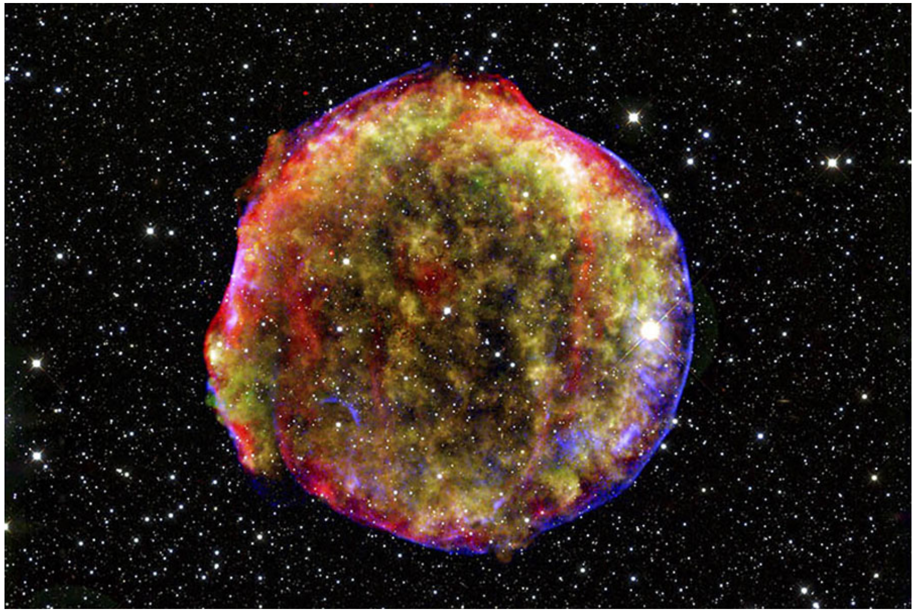
Bayesian cosmological parameter inference from supernovae type Ia data



Marisa Cristina March

with

R. Trotta, P. Berkes, G.D. Starkman and P.M. Vaudrevange

Imperial College London

marisa.march06@imperial.ac.uk

arXiv:1102.3237

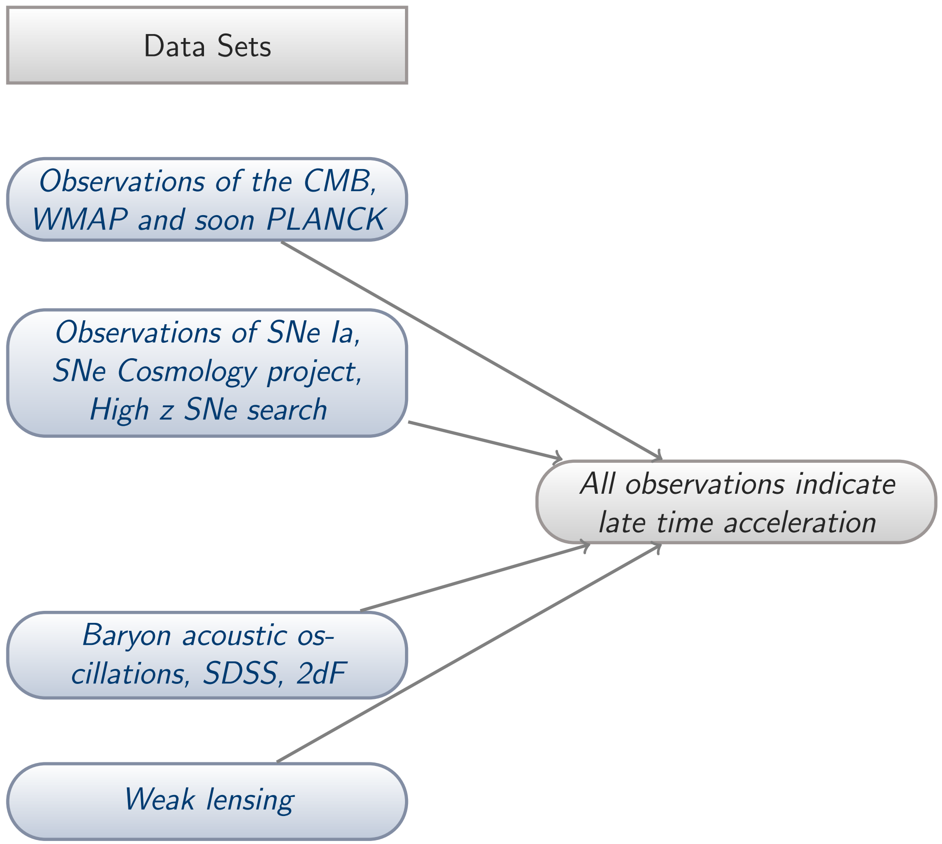
Motivations for this work

We desire a Bayesian method for cosmological parameter inference from supernovae type Ia in order to:

1. Be able to use the supernova data to discriminate between different cosmological models using Bayesian model selection.
2. Provide a statistically well motivated framework for evaluating the posterior probabilities of the cosmological parameters.
3. Obtain a probability density function for the unknown intrinsic dispersion of the absolute magnitudes of the SNe Ia.

marisa.march06@imperial.ac.uk
arXiv:1102.3237

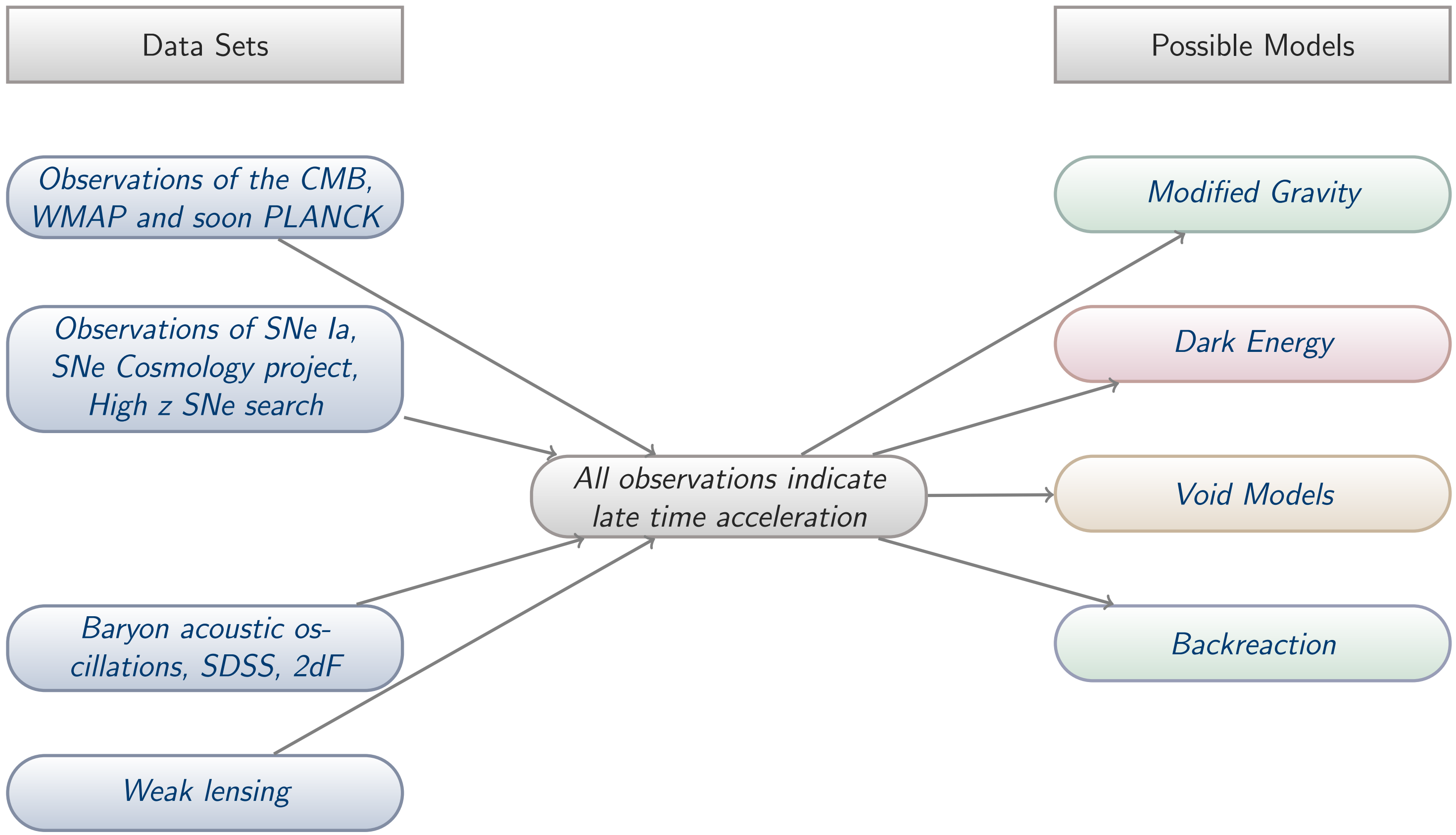
Which is the best model, given the data?



marisa.march06@imperial.ac.uk

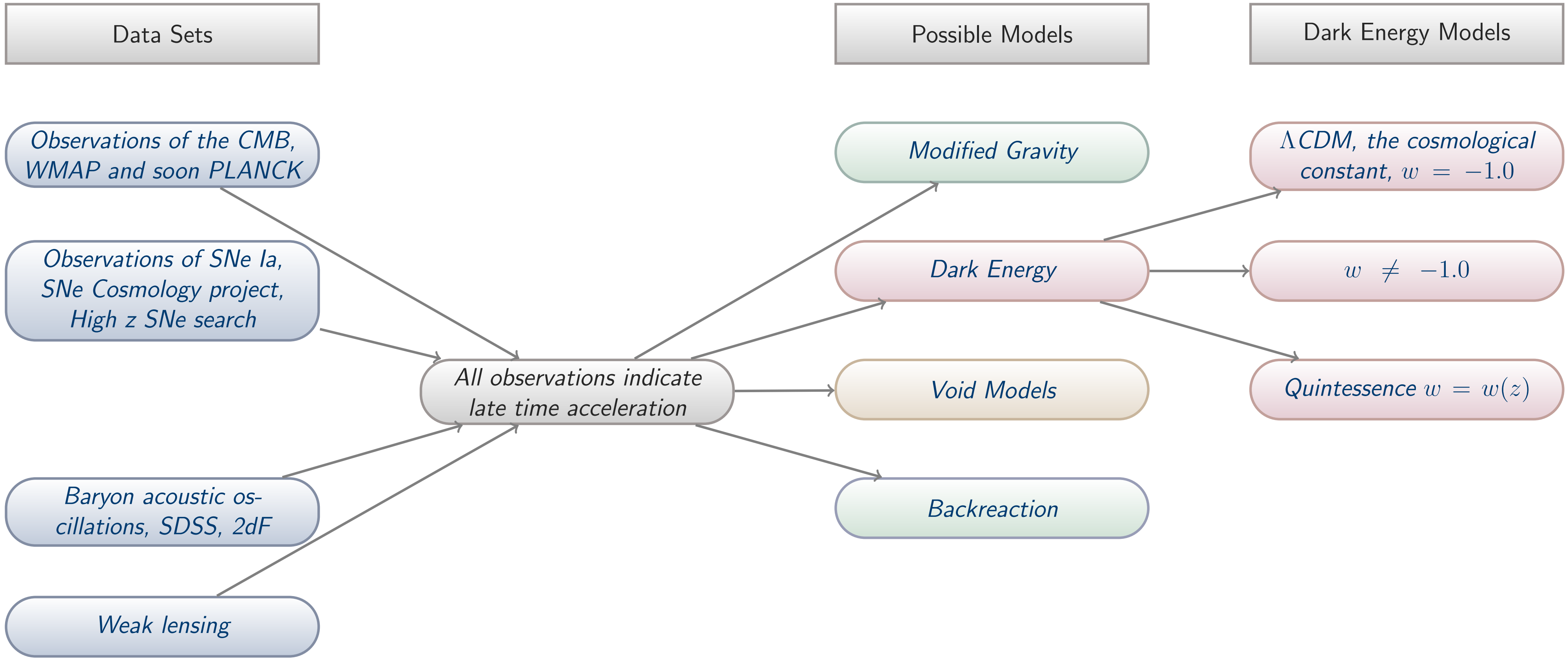
arXiv:1102.3237

Which is the best model, given the data?



marisa.march06@imperial.ac.uk
arXiv:1102.3237

Which is the best model, given the data?



marisa.march06@imperial.ac.uk

arXiv:1102.3237

Reminder: Bayesian model selection

For a given set of known models, $\{\mathcal{M}_1 \ldots \mathcal{M}_j \ldots \mathcal{M}_N\}$;

$$p(\mathcal{M}_j|d) = \frac{p(d|\mathcal{M}_j)p(\mathcal{M}_j)}{p(d)}$$

To compare models with equal priors, look at the Bayes factor:

$$B_{ab} = \frac{p(d|\mathcal{M}_a)}{p(d|\mathcal{M}_b)}$$

The relative goodness of model can be interpreted using Jeffreys' scale:

$\lvert \ln B_{ij} \rvert$	Odds	Strength of evidence
< 1.0	$\lesssim 3:1$	Inconclusive
1.0	$\sim 3:1$	Weak evidence
2.5	$\sim 12:1$	Moderate evidence
5.0	$\sim 150:1$	Strong evidence

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Reminder: Bayes' theorem

Given a data set d what is the posterior probability of the parameters θ for a specific model $\mathcal{M}_j$?

posterior likelihood prior

$$p(\theta|d,\mathcal{M}_j) = \frac{p(d|\theta,\mathcal{M}_j)p(\theta|\mathcal{M}_j)}{p(d|\mathcal{M}_j)}$$

evidence

The model likelihood is also known as the Bayesian evidence

evidence

$$p(d|\mathcal{M}_j) = \int p(d|\theta,\mathcal{M}_j)p(\theta|\mathcal{M}_j)\mathrm{d}\theta$$

Bayes' theorem can be applied to problems of parameter inference and also problems of model selection.

marisa.march06@imperial.ac.uk

arXiv:1102.3237

Overview

Supernovae Type Ia can be used as standard candles to infer the cosmological parameters.

- Standard candles and their use in inferring the cosmological parameters.
- How supernovae Ia can be made into standard candles.
- How to get from supernovae data to the cosmological parameters.
- The standard method and its shortfalls.
- Our new method.
- Numerical trials and preliminary results.

marisa.march06@imperial.ac.uk

Standard Candles

Standard candles are any class of object of which all members have the same absolute magnitude, M

$$\mu \equiv m_B - M$$

$$\mu = 5\log_{10}\left(\frac{D_L(z)}{1Mpc}\right) + 25$$

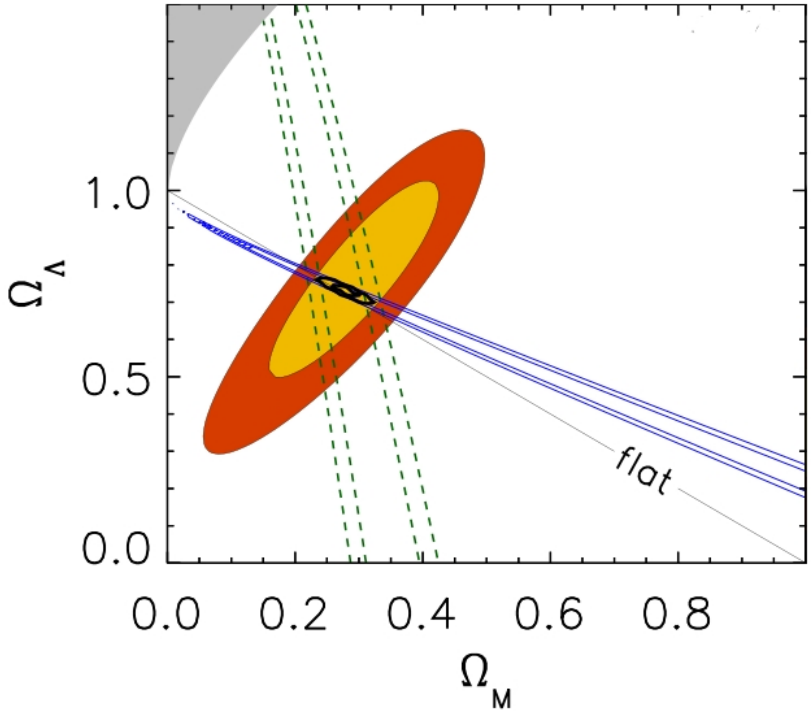
The distance modulus μ is the difference between the apparent magnitude m_B and absolute magnitude M of a standard candle

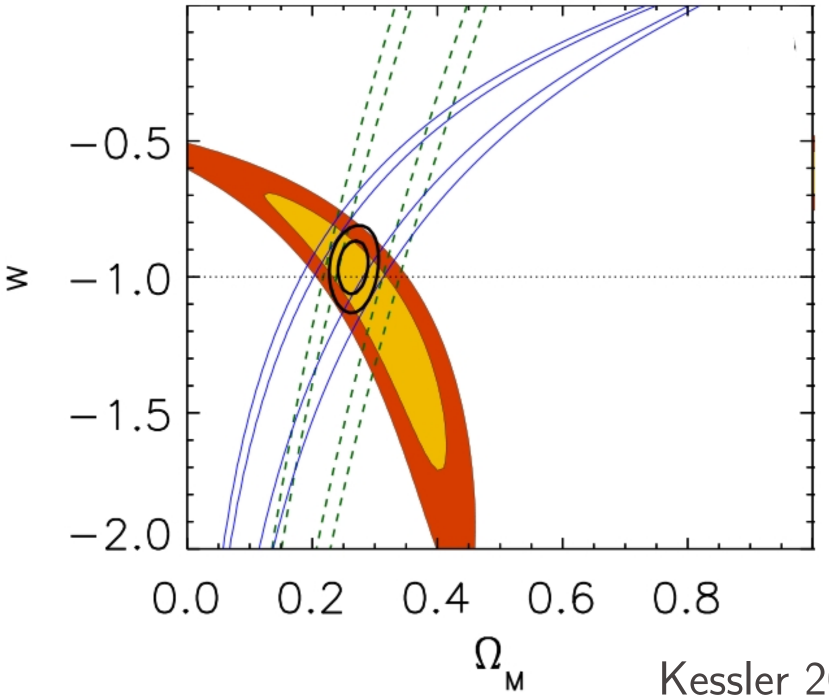
Luminosity distance

$$D_L(z) = \frac{c(1+z)}{H_0\sqrt{-\Omega_\kappa}} \sin\left(\sqrt{-\Omega_\kappa}\int_0^z \frac{H_0}{H(z)}\mathrm{dz}\right)$$

marisa.march06@imperial.ac.uk
arXiv:1102.3237

The make-up of the universe





Kessler 2009

Friedmann equation relates cosmological parameters to Hubble rate:

$$H(z)^2 = H_0^2 \left(\Omega_m^0 (1+z)^3 + \Omega_r^0 (1+z)^4 + \Omega_\kappa^0 (1+z)^2 + \Omega_{DE}^0 \exp \left[3 \int_0^z \frac{1+w(z)}{1+z} dz \right] \right)$$

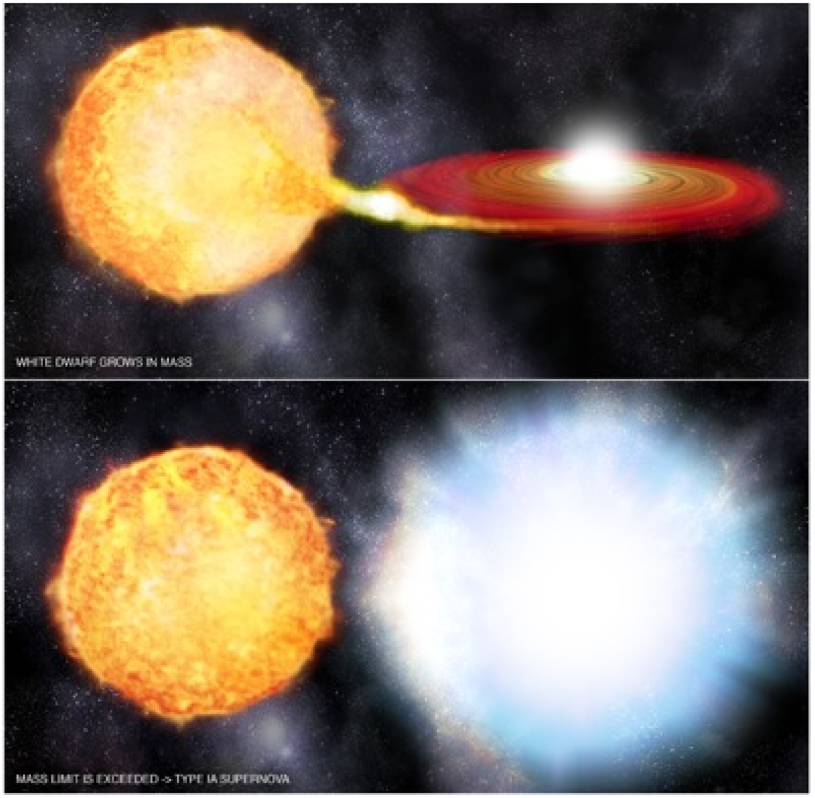
$w(z)$ is the dark energy equation of state

marisa.march06@imperial.ac.uk

arXiv:1102.3237

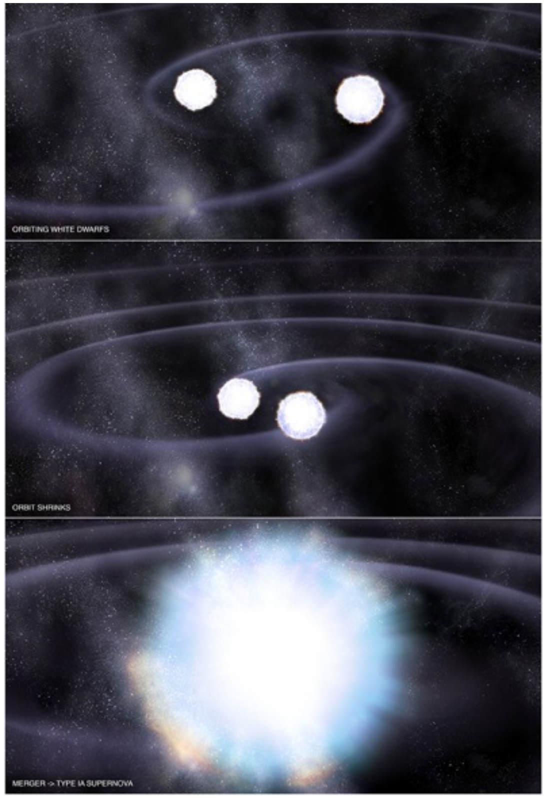
Progenitors of Type 1a Supernovae?

Accretion



Images:NASA/CXC/M Weiss.

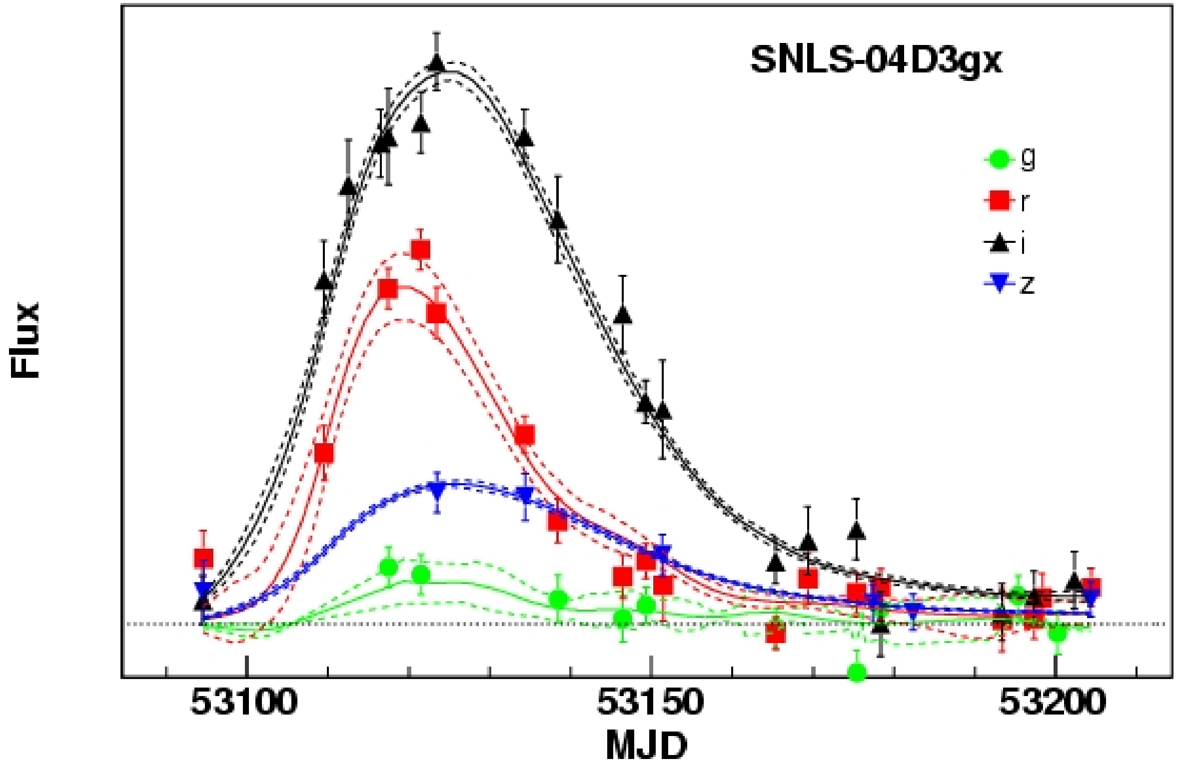
Merger



bimodal population?

marisa.march06@imperial.ac.uk
arXiv:1102.3237

From SNe Ia to the cosmological parameters



Guy et. al. 2007

find SN Ia candidates

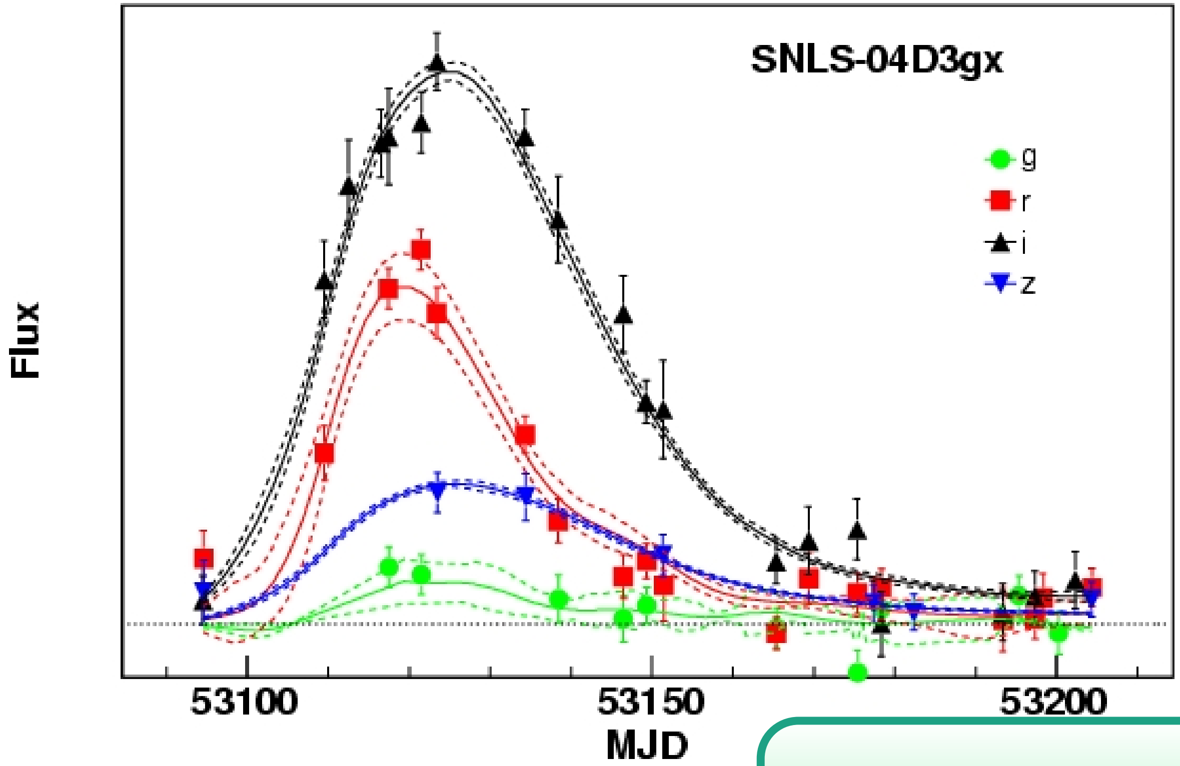
follow up and identify by spectra

make observations over a few weeks

do light curve fitting: SALT2 $\hat{m}^*_{Bi}$, $\hat{x}_{1i}$,$\hat{c}_i$

cosmology fitting: Ω_M,Ω_Λ,w

From SNe Ia to the cosmological parameters



Guy et. al. 2007

find SN Ia candidates

follow up and identify by spectra

make observations over a few weeks

do light curve fitting: SALT2 $\hat{m}^*_{Bi}$, $\hat{x}_{1i}$,$\hat{c}_i$

cosmology fitting: Ω_M,Ω_Λ,w

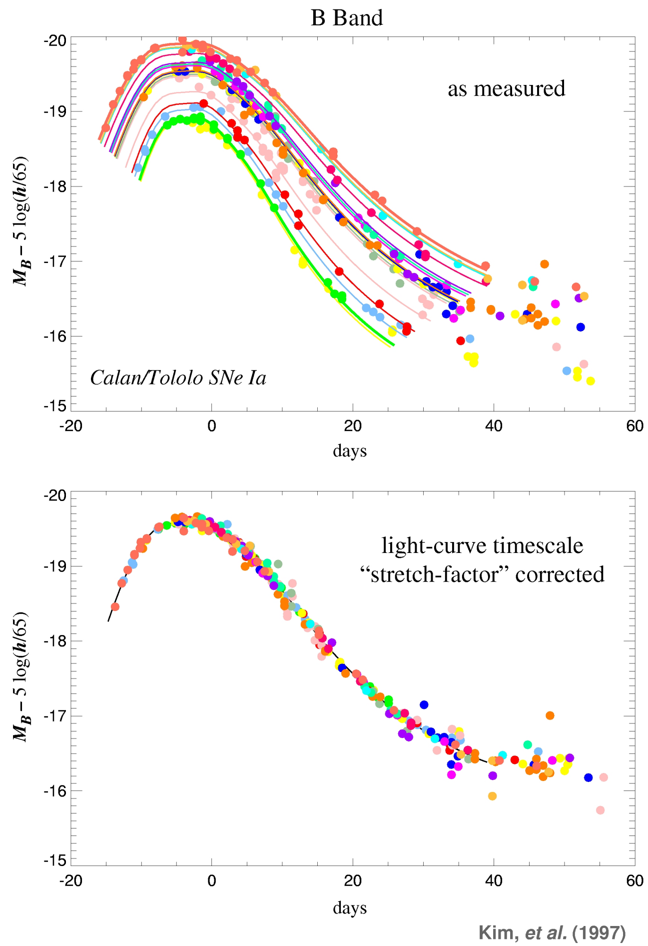
our work

SNe Ia as standardizable candles

Supernovae type Ia may be made into standard candles by applying empirical corrections for stretch and colour:

$$\mu_i = m^*_{Bi} - M_0 + \alpha x_{1,i} - \beta c_i$$

- Estimates for m^*_{Bi}, $x_{1,i}$, c_i come from SALT2 lightcurve fitting.
- M_0,α,β are unknown nuisance parameters.
- The corrections reduce the scatter in the absolute magnitudes of the SNe Ia.
- Even after correction, a dispersion in the absolute magnitudes remains, σ_μ^{int}.



Sources of systematic uncertainty σ_μ^{int}

Physical reasons: (our main interest)

- Variations in intrinsic colour of SN Ia.
- Variations in host galaxy reddening: spiral vs elliptical.
- Possible bimodal population: single degenerate progenitor system vs double degenerate progenitor system.
- Evolution of supernovae with redshift $\longrightarrow$ early SNe population may be different from late SNe population.

Technical reasons:

- Instrumentation, detector sensitivity, calibration of light curve fitter.

marisa.march06@imperial.ac.uk

Sources of systematic uncertainty σ_μ^{int}

Physical reasons: (our main interest)

- Variations in intrinsic colour of SN Ia.
- Variations in host
- Possible bimodal p
system vs double d
- Evolution of super
population may be

Technical reasons:

- Instrumentation, d
curve fitter.

- How do these systematics affect our ability to recover the cosmological parameters?
- What is σ_μ^{int} ?
- How can we model the systematic errors?

marisa.march06@imperial.ac.uk
arXiv:1102.3237

The standard χ^2 method

- χ^2 is minimised by fitting for cosmological and supernovae parameters simultaneously
- α and β appear in both numerator and denominator
- σ_μ^{int} is unknown

$$\chi^2_\mu = \sum_i \frac{\left[\mu_i(\Omega_M, \Omega_\Lambda, \Omega_\kappa, w, H_0) - \mu_i^{obs}\right]^2}{\sigma^2_{\mu i}}$$

$$\mu_i^{obs} = \hat{m}^*_{Bi} - M_0 + \alpha \hat{x}_{1i} - \beta \hat{c}_i$$

$$\sigma^2_{\mu i} = (\sigma^{fit}_{\mu i})^2 + (\sigma^z_{\mu i})^2 + (\sigma^{int}_\mu)^2$$

$$(\sigma^{fit}_{\mu i})^2 = \sigma^2_{\hat{m}^*_{Bi}} + \alpha^2 \sigma^2_{x_1} + \beta^2 \sigma^2_c$$

(uncorrelated case)

marisa.march06@imperial.ac.uk
arXiv:1102.3237

The standard χ^2 method

- χ^2 is minimised by fitting for cosmological and supernovae parameters simultaneously
- α and β appear in both numerator and denominator
- σ_μ^{int} is unknown

- σ_μ^{int} is chosen such that $\chi^2/\text{dof} \sim 1$ $\longrightarrow$ assumes model is a good fit.
- Only gives single value for σ_μ^{int}.
- No justification for χ^2 distribution.
- Cannot apply Bayesian model selection or use with MCMC.

$$\chi_\mu^2 = \sum_i \frac{\left[\mu_i(\Omega_M, \Omega_\Lambda, \Omega_\kappa, w, H_0) - \mu_i^{obs}\right]^2}{\sigma_{\mu i}^2}$$

$$\mu_i^{obs} = \hat{m}_{Bi}^* - M_0 + \alpha \hat{x}_{1i} - \beta \hat{c}_i$$

$$\sigma_{\mu i}^2 = (\sigma_{\mu i}^{fit})^2 + (\sigma_{\mu i}^z)^2 + (\sigma_\mu^{int})^2$$

$$(\sigma_{\mu i}^{fit})^2 = \sigma_{\hat{m}_{Bi}^*}^2 + \alpha^2 \sigma_{x_1}^2 + \beta^2 \sigma_c^2$$

(uncorrelated case)

Imperial College
London

marisa.march06@imperial.ac.uk

The standard χ^2 method

- χ^2 is minimised by fitting for cosmological and supernovae parameters simultaneously
- α and β appear in both numerator and denominator
- σ_μ^{int} is unknown

$$\chi^2_\mu = \sum_i \frac{\left[\mu_i(\Omega_M, \Omega_\Lambda, \Omega_\kappa, w, H_0) - \mu_i^{obs}\right]^2}{\sigma^2_{\mu i}}$$

$$\mu_i^{obs} = \hat{m}^*_{Bi} - M_0 + \alpha \hat{x}_{1i} - \beta \hat{c}_i$$

$$\sigma^2_{\mu i} = (\sigma^{fit}_{\mu i})^2 + (\sigma^z_{\mu i})^2 + (\sigma^{int}_\mu)^2$$

$$(\sigma^{fit}_{\mu i})^2 = \sigma^2_{\hat{m}^*_{Bi}} + \alpha^2 \sigma^2_{x_1} + \beta^2 \sigma^2_c$$

(uncorrelated case)

- σ_μ^{int} is chosen such that $\chi^2/\text{dof} \sim 1$ $\longrightarrow$ assumes model is a good fit
- Only gives single value for σ_μ^{int}
- No justification for χ^2 distribution
- Cannot apply Bayesian model selection or use with MCMC.

A new and better method for parameter estimation is required!

marisa.march06@imperial.ac.uk
arXiv:1102.3237

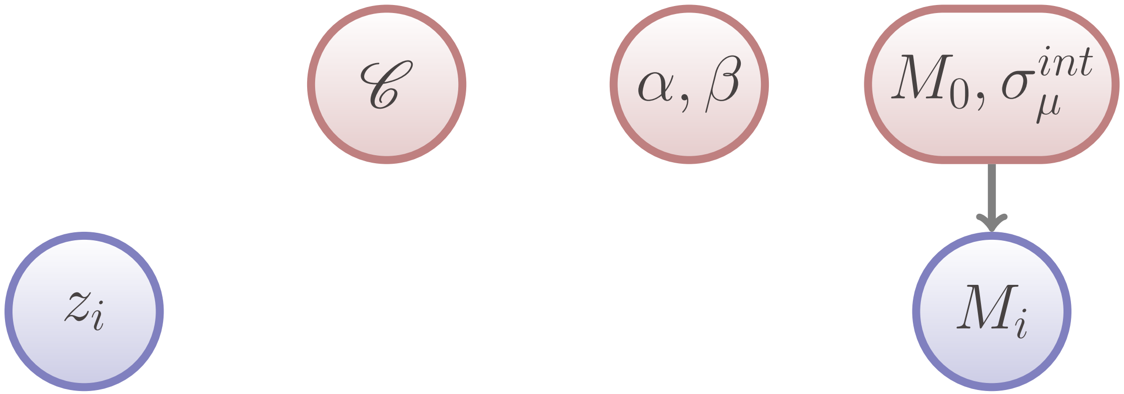
Bayesian hierarchical network

...each supernova has a redshift z_i and (an un-observed) absolute magnitude M_i...

marisa.march06@imperial.ac.uk

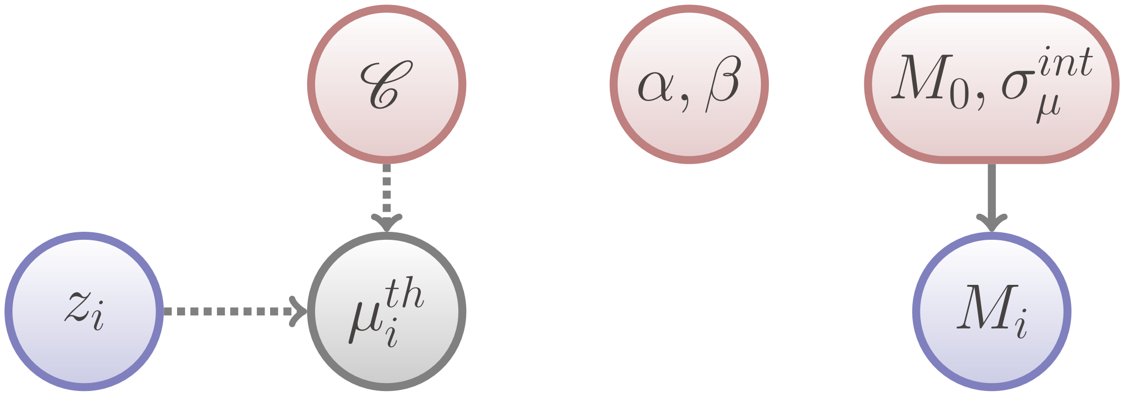
...M_i is drawn from a Gaussian distribution of mean M_0 and standard deviation σ_μ^{int}...

marisa.march06@imperial.ac.uk
arXiv:1102.3237



..the cosmological parameters $\mathscr{C} = \{\Omega_M, \Omega_\Lambda, w, H_0\}$ are unknown to us..
....the supernovae parameters α and β are also unknown...

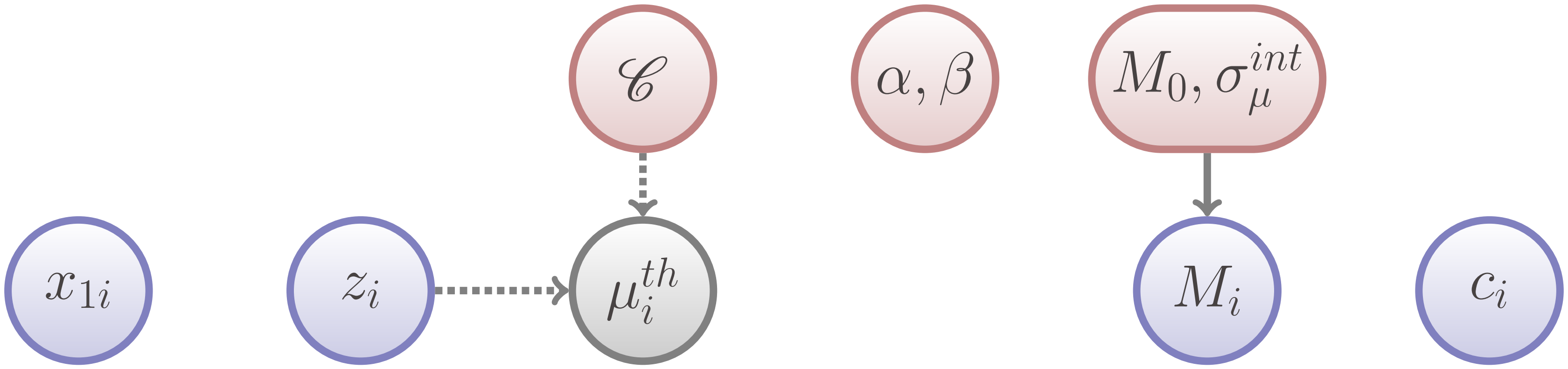
marisa.march06@imperial.ac.uk
arXiv:1102.3237



...$\mathscr{C}$ and z_i deterministically specify the theoretical distance modulus μ_i^{th}...

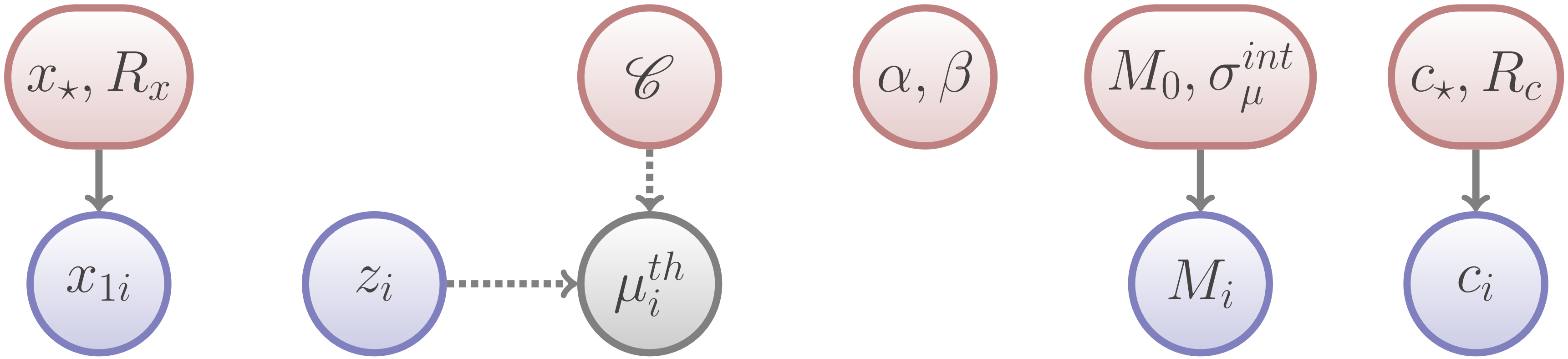
marisa.march06@imperial.ac.uk

arXiv:1102.3237



...each supernova has a true stretch x_{1i} and colour c_i parameter...

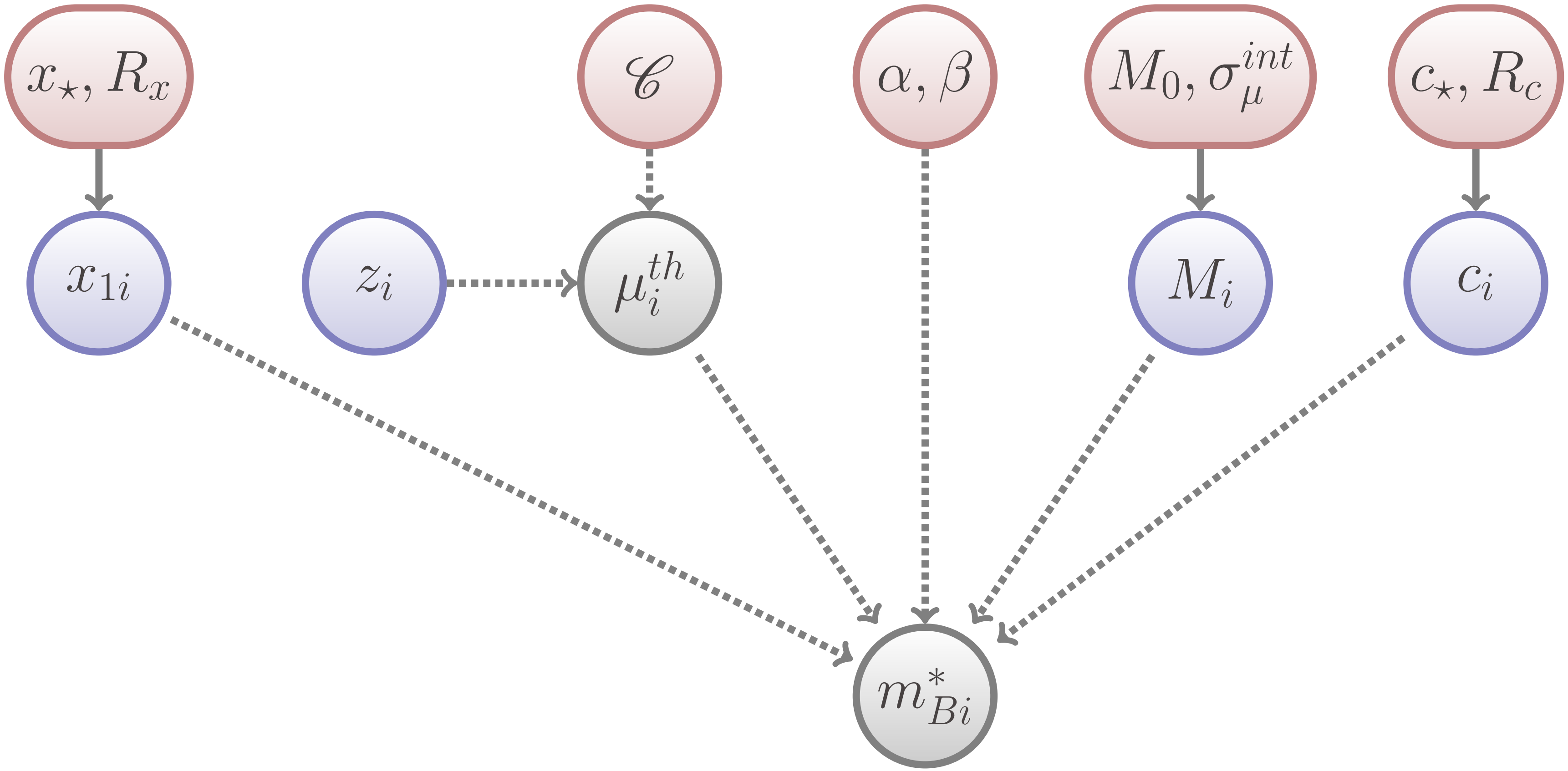
marisa.march06@imperial.ac.uk
arXiv:1102.3237



...each x_{1i} and c_i are drawn from their respective Gaussian distributions...

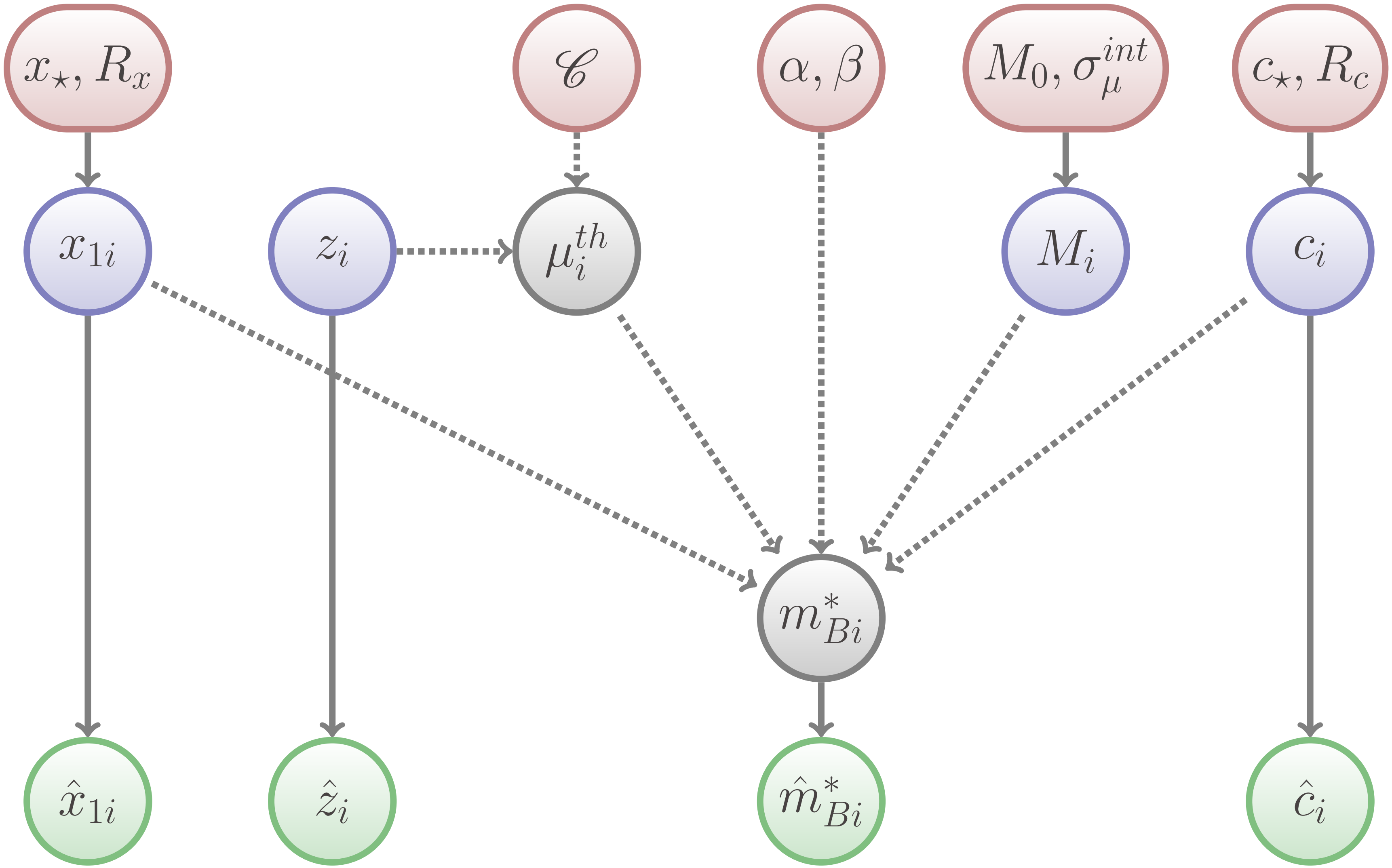
marisa.march06@imperial.ac.uk

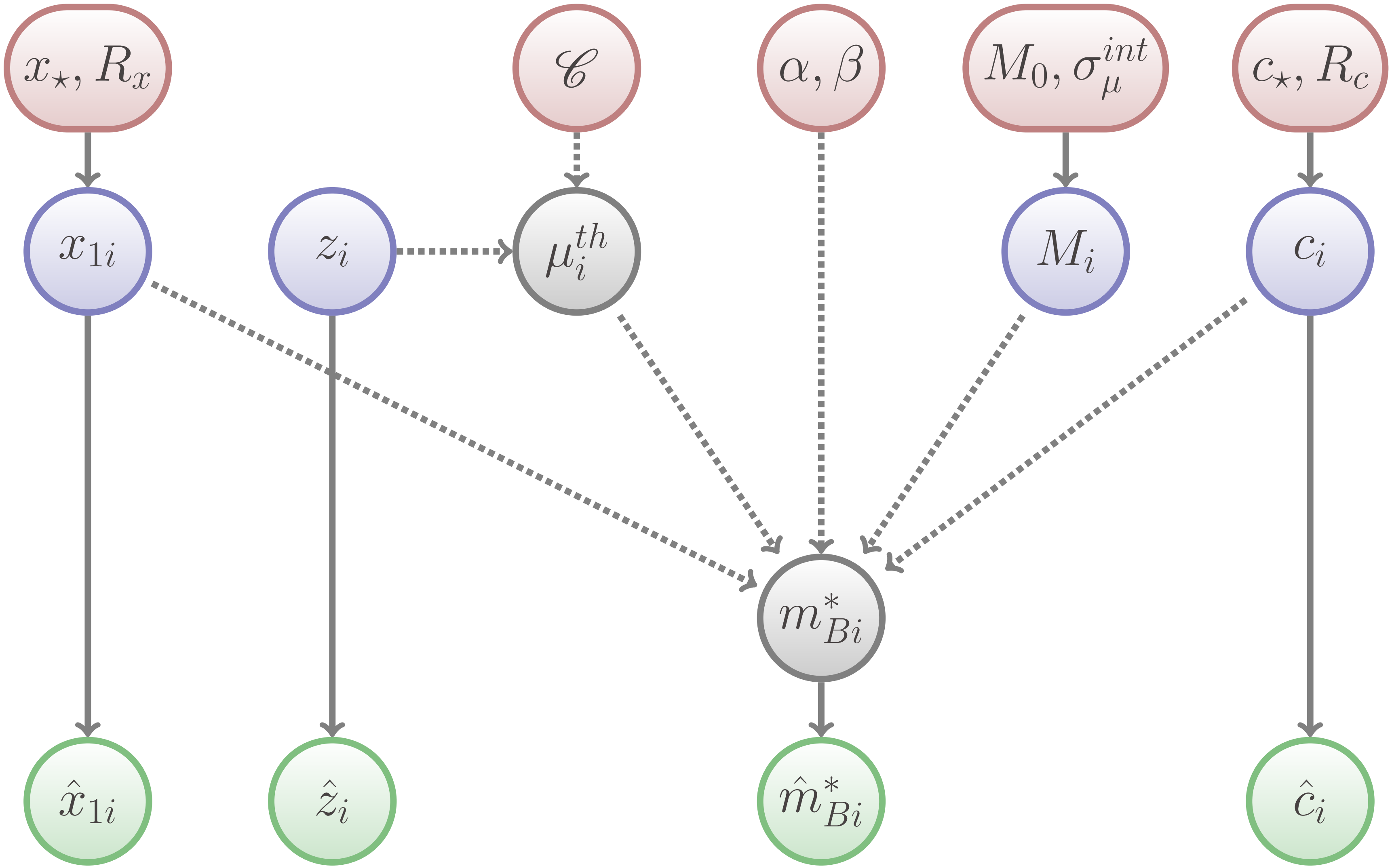
arXiv:1102.3237



...m^*_{Bi} is specified deterministically...

...we do not observe the true values; our observations are subject to a further Gaussian noise...

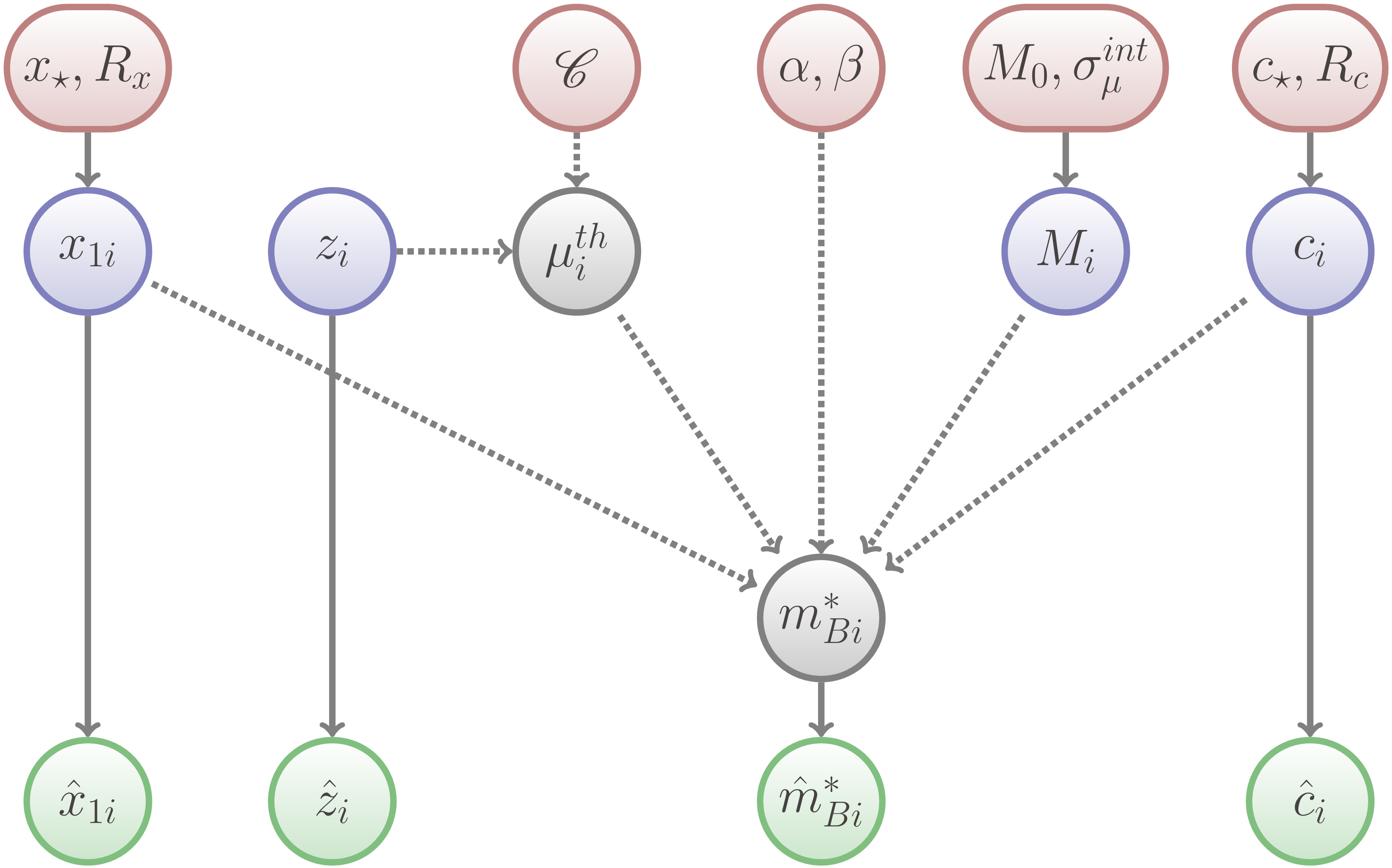




marisa.march06@imperial.ac.uk
arXiv:1102.3237

Posterior:

$$p(\Theta|\hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}) \propto p(\hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}|\Theta)p(\Theta) \quad \text{where} \quad \Theta = \{\mathscr{C}, \alpha, \beta, M_0, \sigma_\mu^{int}\}$$



marisa.march06@imperial.ac.uk
arXiv:1102.3237

Calculating the posterior pdf:

Bayes theorem:

$$p(\Theta|d) = \frac{p(d|\Theta)p(\Theta)}{p(D)}$$

Where the parameters of interest are: $\Theta = \{\mathscr{C}, \alpha, \beta, M_0, \sigma_\mu^{\text{int}}\}$

prior:

$$\begin{aligned} p(\Theta) &= p(\mathscr{C}, \alpha, \beta)p(M_0, \sigma_\mu^{\text{int}}) \\ &= p(\mathscr{C}, \alpha, \beta)p(M_0|\sigma_\mu^{\text{int}})p(\sigma_\mu^{\text{int}}) \end{aligned}$$

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Expanding out the likelihood alone:

$$\begin{aligned} p(d|\Theta) &= p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}^*_B|\Theta) \\ &= \int \mathsf{d}\underline{c}\ \mathsf{d}\underline{x}_1\ \mathsf{d}\underline{M}\ p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}^*_B|\underline{c}, \underline{x}_1, \underline{M}, \Theta)p(\underline{c}, \underline{x}_1, \underline{M}|\Theta) \\ &= \int \mathsf{d}\underline{c}\ \mathsf{d}\underline{x}_1\ \mathsf{d}\underline{M}\ p(\underline{\hat{c}}, \underline{\hat{x}}_1, \underline{\hat{m}}^*_B|\underline{c}, \underline{x}_1, \underline{M}, \Theta)p(\underline{c}, \underline{x}_1|\Theta)p(\underline{M}|\Theta) \end{aligned}$$

where:

$$\begin{aligned} \underline{M} &\sim \mathcal{N}(\underline{M}_0, \Sigma_\Delta) \\ \underline{c} &\sim \mathcal{N}(c_\star \cdot \mathbf{1}_n, \mathsf{diag}\left(R_c^2 \cdot \mathbf{1}_n\right)) = p(\underline{c}|c_\star, R_c) \\ \underline{x}_1 &\sim \mathcal{N}(x_\star \cdot \mathbf{1}_n, \mathsf{diag}\left(R_x^2 \cdot \mathbf{1}_n\right)) = p(\underline{x}_1|x_\star, R_x) \end{aligned}$$

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Expanding out the likelihood alone:

Priors on $\underline{x}_1$ and $\underline{c}$:

$$p(\underline{c}, \underline{x}_1|\Theta) = \int \mathrm{d}R_c \, \mathrm{d}R_x \, \mathrm{d}c_\star \, \mathrm{d}x_\star \; p(\underline{c}|c_\star, R_c) p(\underline{x}_1|x_\star, R_x) p(R_c) p(R_x) p(c_\star) p(x_\star).$$

The expression for the likelihood becomes:

$$\begin{aligned} p(\underline{\hat{c}}, \underline{\hat{s}}, \underline{\hat{m}}^*_B|\Theta) = & \int \mathrm{d}\underline{c} \, \mathrm{d}\underline{x}_1 \, \mathrm{d}\underline{M} \; p(\underline{\hat{c}}, \underline{\hat{x}}_1|\underline{c}, \underline{x}_1) p(\underline{\hat{m}}^*_B|\underline{c}, \underline{x}_1, \underline{M}, \Theta) p(\underline{M}|\Theta) \\ & \times \int \mathrm{d}R_c \, \mathrm{d}R_x \, \mathrm{d}c_\star \, \mathrm{d}x_\star \; p(\underline{c}|c_\star, R_c) p(\underline{x}_1|x_\star, R_x) p(R_c) p(R_x) p(c_\star) p(x_\star). \end{aligned}$$

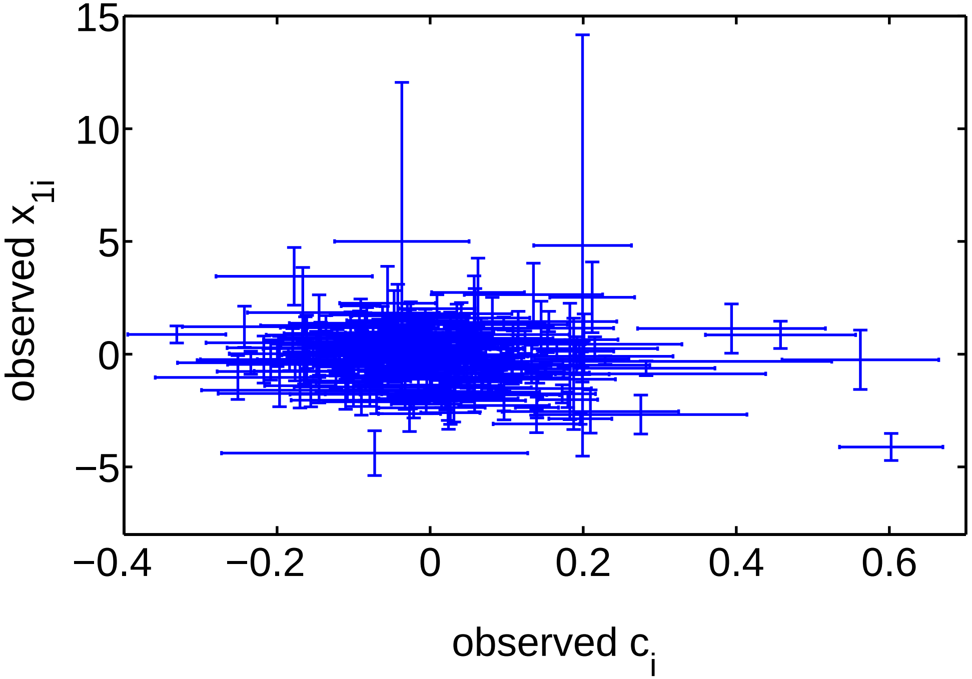
Final expression for the effective likelihood is:

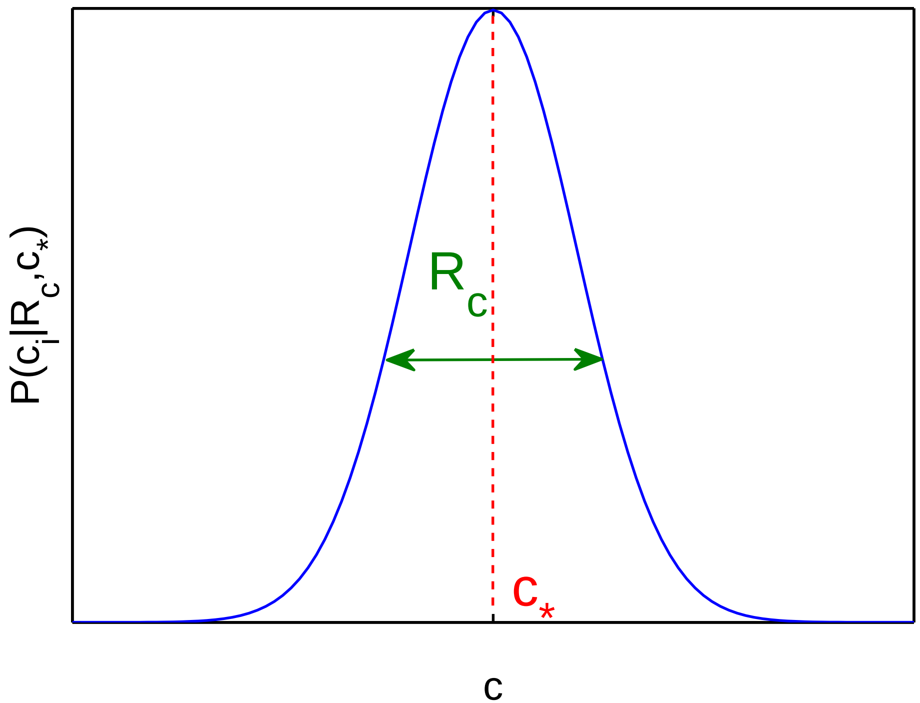
$$\begin{aligned} p(\underline{\hat{c}}, \underline{\hat{s}}, \underline{\hat{m}}^*_B|\Theta) = & \int \mathrm{d}\log R_c \, \mathrm{d}\log R_x \; |2\pi(\Sigma_m + \Sigma_\Delta)|^{-\frac{1}{2}} |2\pi\Sigma_R|^{-\frac{1}{2}} |2\pi\Sigma_0|^{-\frac{1}{2}} |2\pi Y|^{\frac{1}{2}} |2\pi\Sigma_b|^{\frac{1}{2}} |2 \\ & \times |2\pi\Sigma_h|^{-1/2} |2\pi W|^{\frac{1}{2}} |2\pi\Sigma_h|^{\frac{1}{2}} \exp\left[-\frac{1}{2} \left(\underline{\hat{g}}^T (\Sigma_g + \Sigma_b)^{-1} \underline{\hat{g}} + \underline{h}_0^T \Sigma_h^{-1} \underline{h}_0 + \underline{\delta}_0^T \Delta \underline{\delta}_0 - \underline{u}^T W \right.\right. \end{aligned}$$

Imperial College
London

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Why priors on the latent c_i, x_{1i}?





- Size of the error bars on $\hat{x}_{1i}$ and $\hat{c}_i$ is comparable with range of $\hat{x}_{1i}$ and $\hat{c}_i$.
- If not properly accounted for, this leads to a bias in the recovery of the supernovae parameters α and β.
- Solution is to put a prior on the range of the latent $x_1 i$ and c_i. [Gull, 1989].

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Overview of new method

1. Introduce latent variables $\underline{c}$, $\underline{x}_1$ and $\underline{M}$.
2. Apply prior to $\underline{c}$, $\underline{x}_1$ and $\underline{M}$.
3. Marginalize over $\underline{M}$ analytically.
4. Marginalize over $c_\star$ and $x_\star$ analytically.
5. Marginalize over $\underline{c}$ and $\underline{x}_1$ analytically.
6. Marginalize over other nuisance parameters numerically.

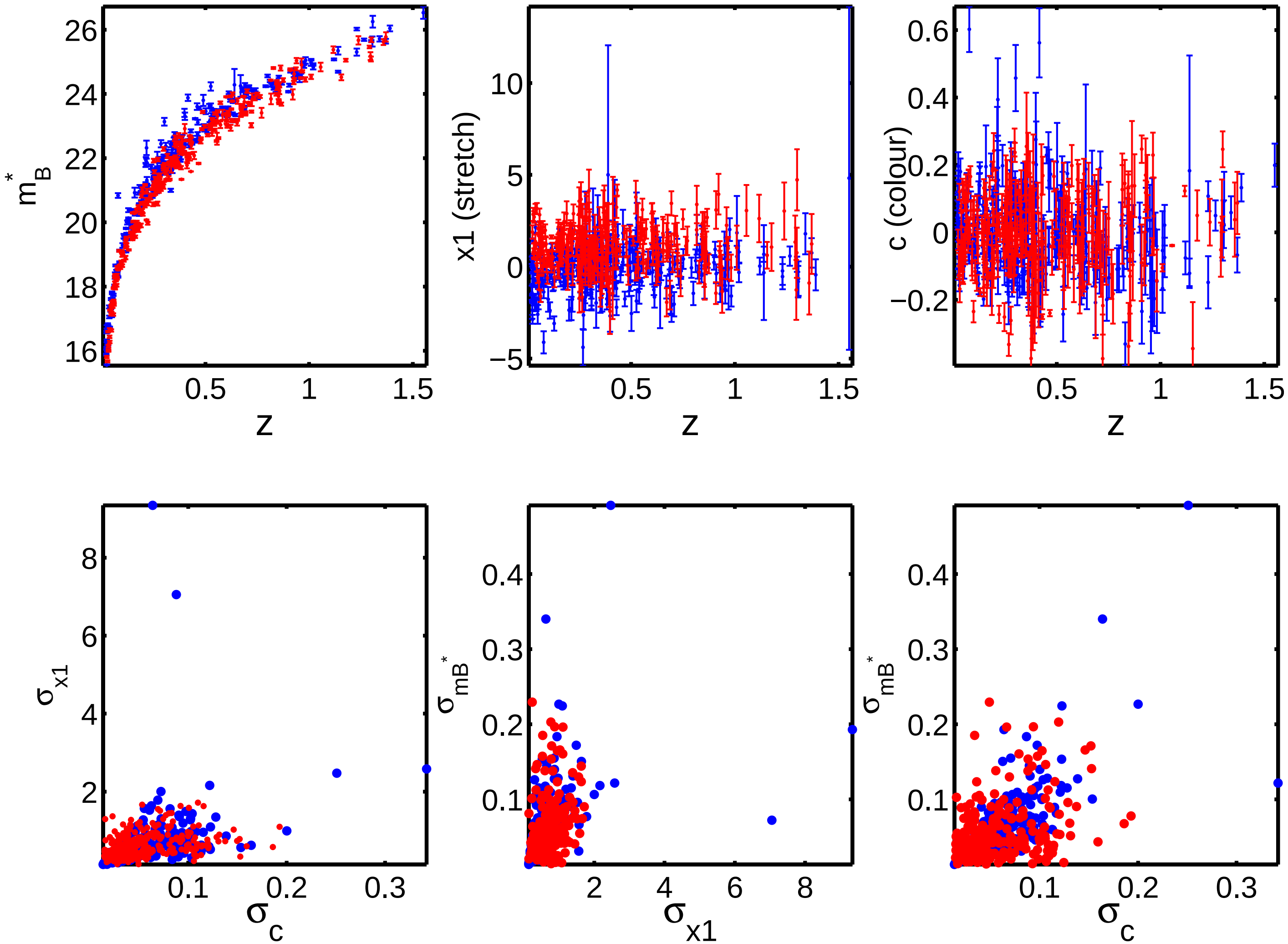
Posterior pdf:

$$p(\Theta|\hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}) = \frac{1}{Z} p(\hat{c}_i, \hat{x}_{1i}, \hat{m}^*_{Bi}|\Theta)p(\Theta)$$

A fully Bayesian method which can be used in problems of Bayesian model selection, and can be used with MCMC methods.

marisa.march06@imperial.ac.uk
arXiv:1102.3237

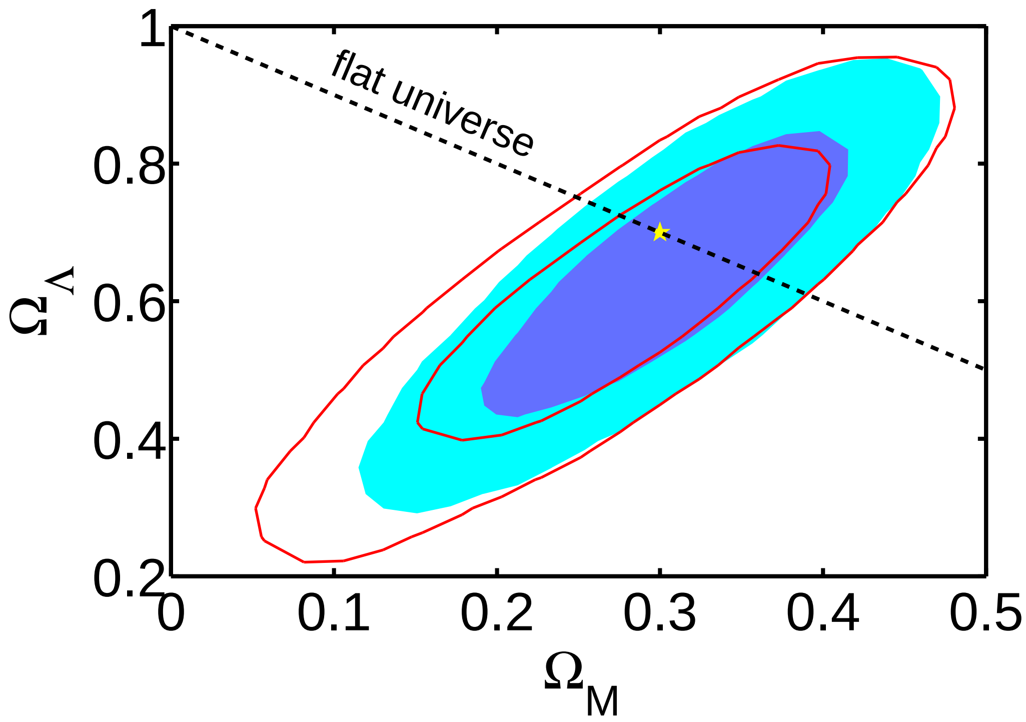
Numerical trials with simulated data

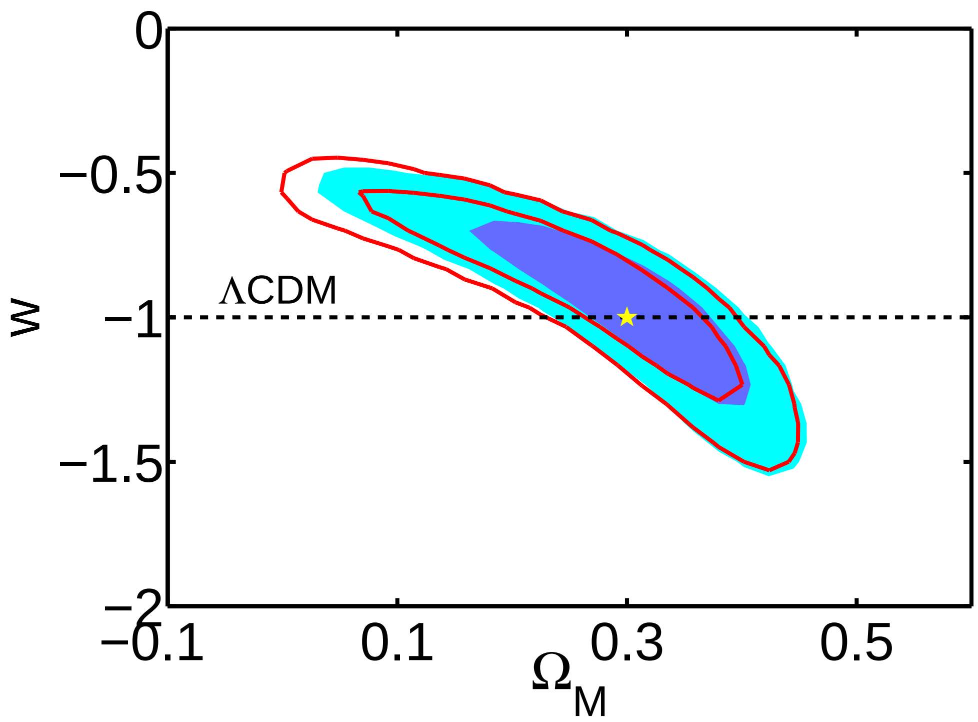


- Simulated data (red) compared with real data (blue);[Kessler,2009]
- True parameters in simulated data are:
  $\Omega_M = 0.3$
  $\Omega_\Lambda = 0.7$
  $\Omega_\kappa = 0.0$
  $w = -1.0$
  $H_0 = 70$ km/s/Mpc
  $M_0 = -19.3$ mag
  i.e. flat ΛCDM

marisa.march06@imperial.ac.uk
arXiv:1102.3237

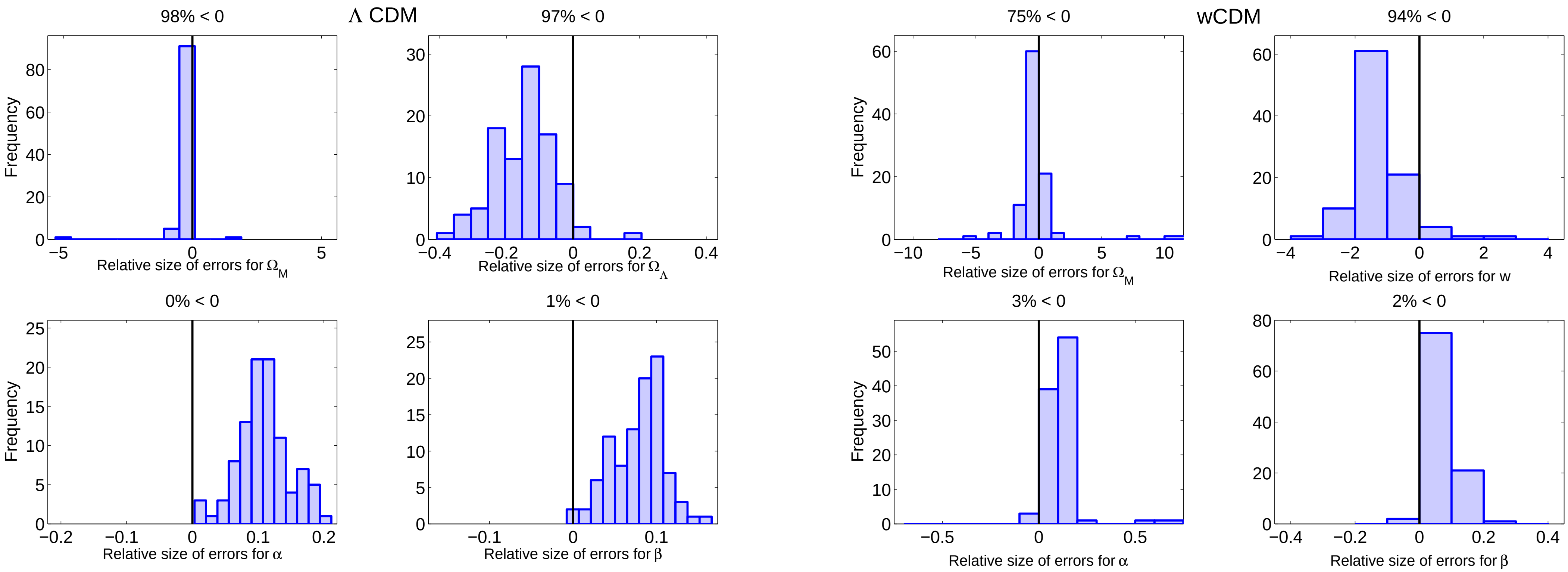
Posterior plots: simulated data





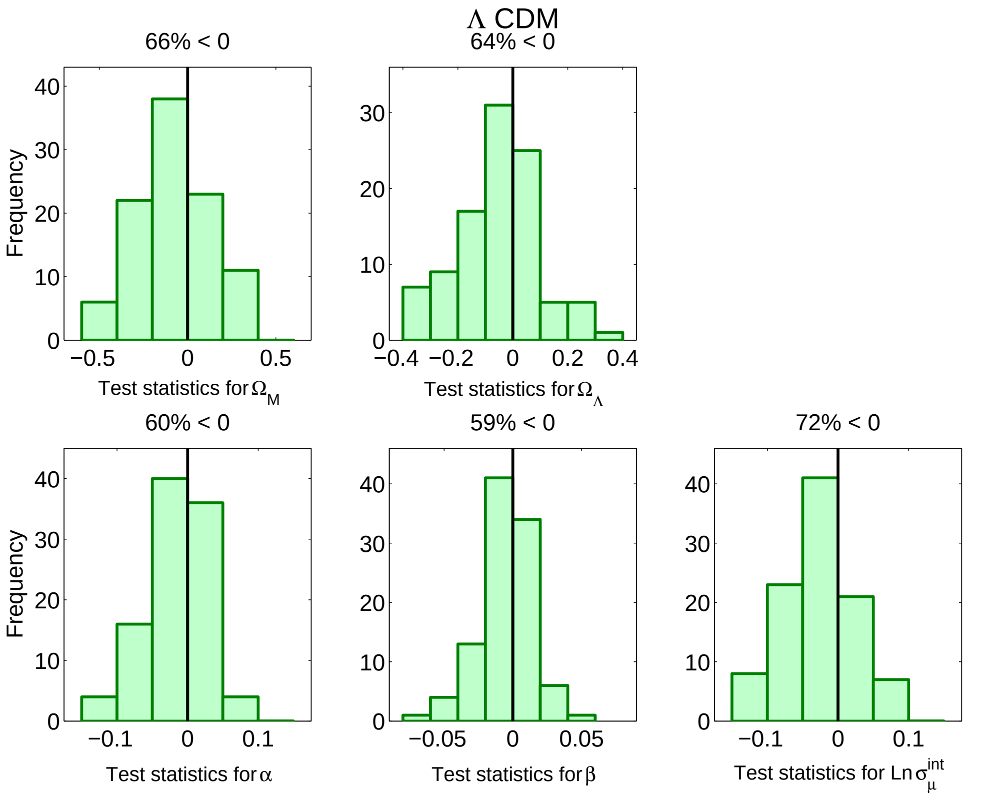
- Contours delimit 95% and 68% regions. Blue is new Bayesian method, red is chisquare method.
- Yellow star indicates true value of parameters.
- H_0 prior is: $H_0 = 70 \pm 8$ km/s/Mpc
- Bayesian method places tighter constraints on parameters.
- Value of parameter recovered by Bayesian method is closer to true value in 60-70 % of trials, compared with chisquare method. (100 trials)

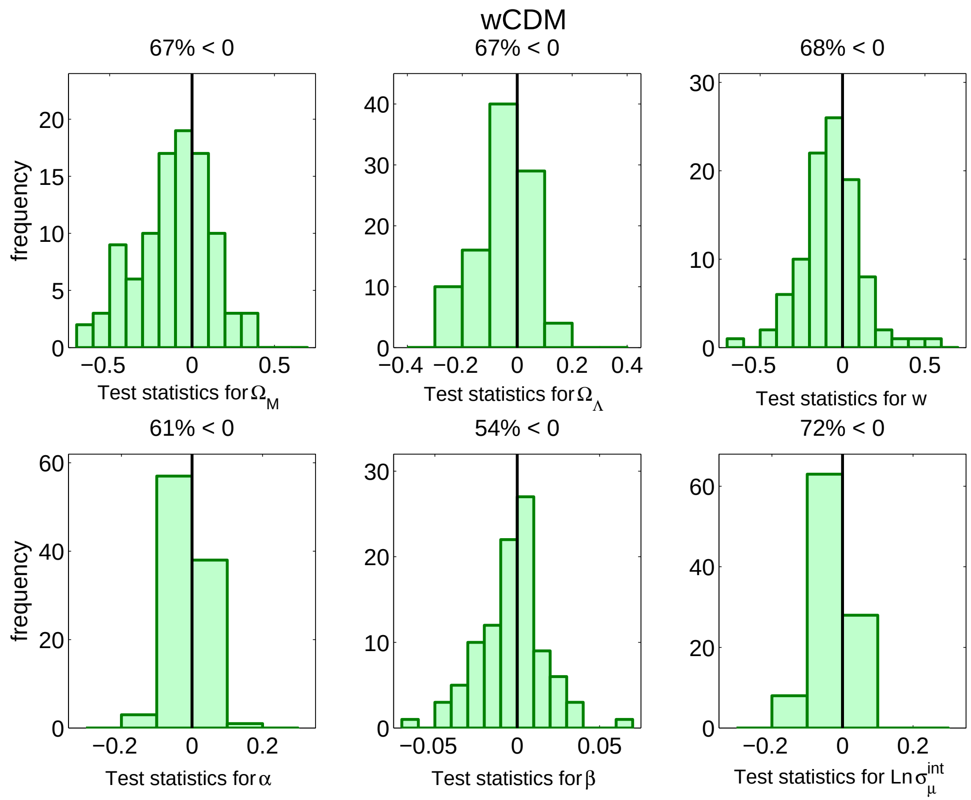
Comparison over 100 trials



- Comparison of size of errorbars on each parameter from each method.
- A value < 0 indicates smaller error bars from our method.
- Our method generally delivers smaller errors on the cosmological parameters (top row), but larger errors on the SALT II correction parameters (lower row)

Comparison over 100 trials

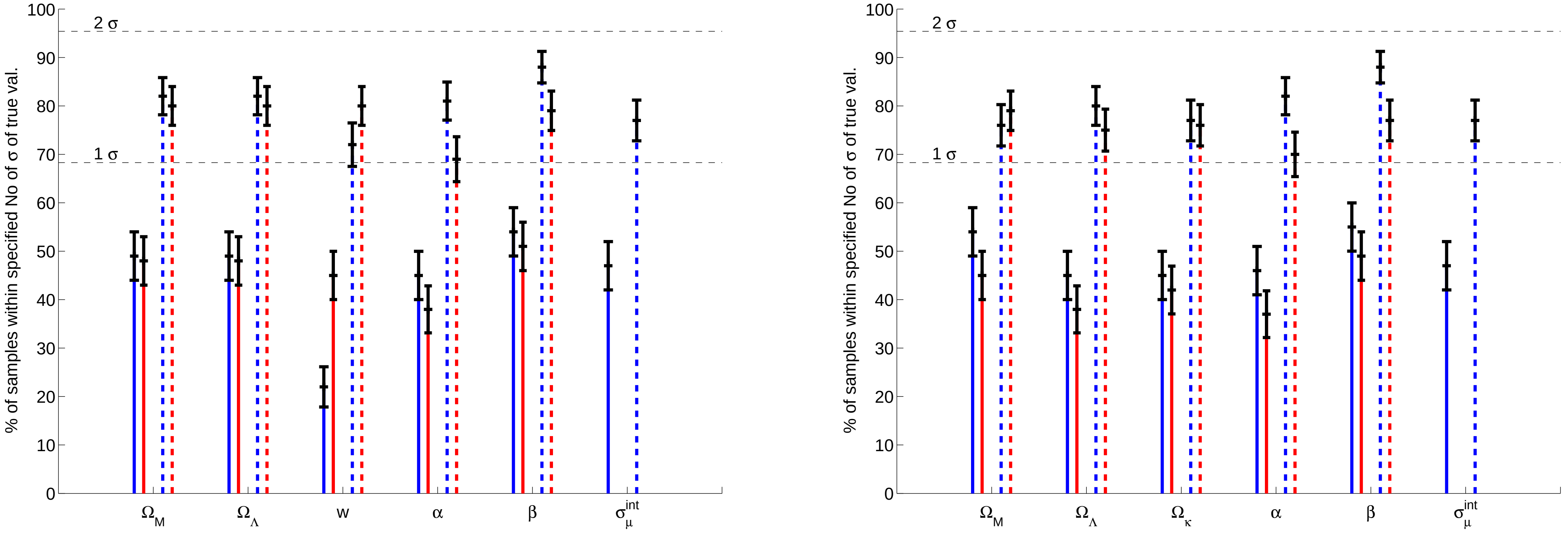




- A predominantly negative value of the test statistics means that our method gives a parameter reconstruction that is closer to the true value than the usual chisquare method.
- For the cosmological parameters (top row), our method outperforms χ^2 about 2 times out of 3.

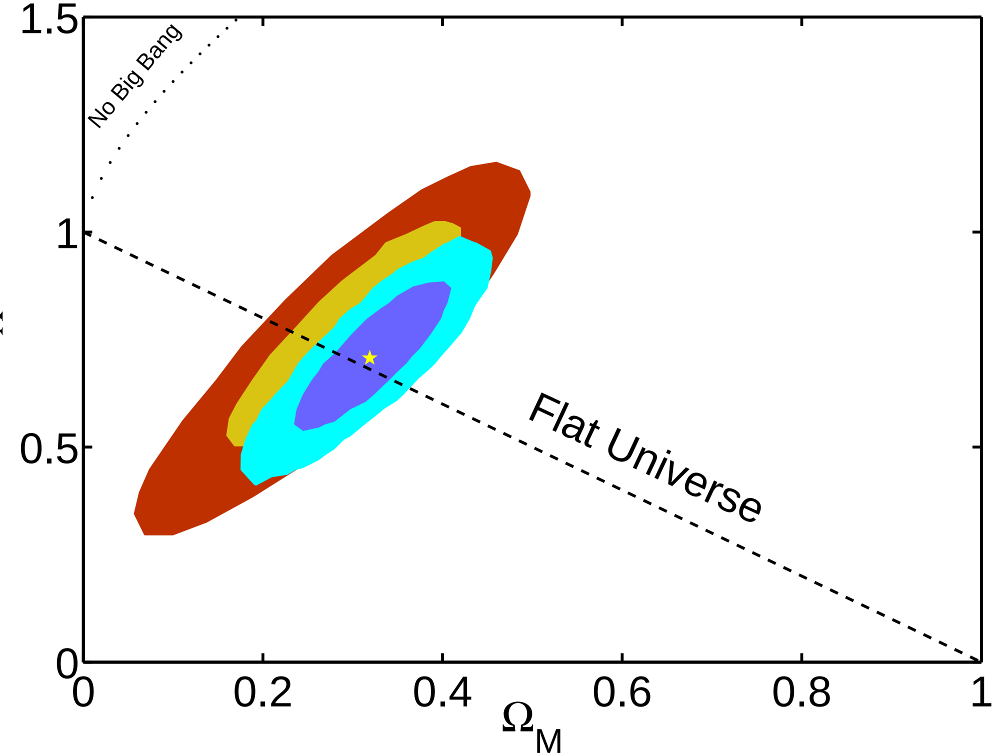
marisa.march06@imperial.ac.uk
arXiv:1102.3237

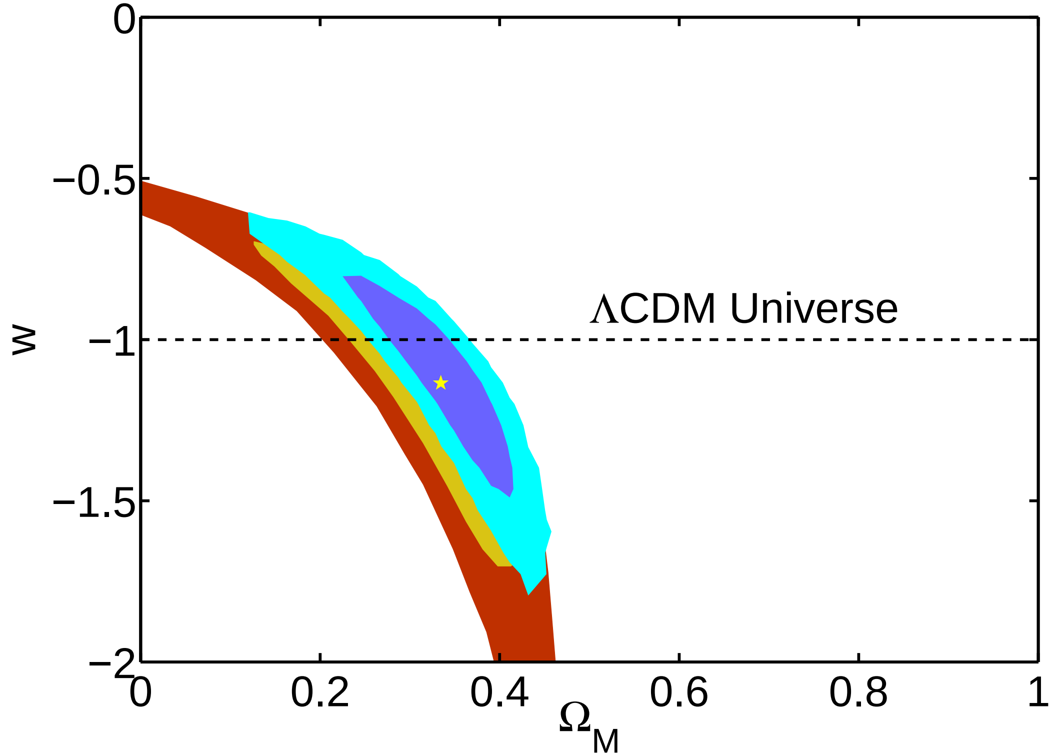
Coverage over 100 trials



- Our method (blue) and chisquare method (red)
- For 68% intervals (solid) and 95 % intervals (dashed)
- Both methods show significant undercoverage - to be investigated further.
- To what extent should Bayesians care about coverage?

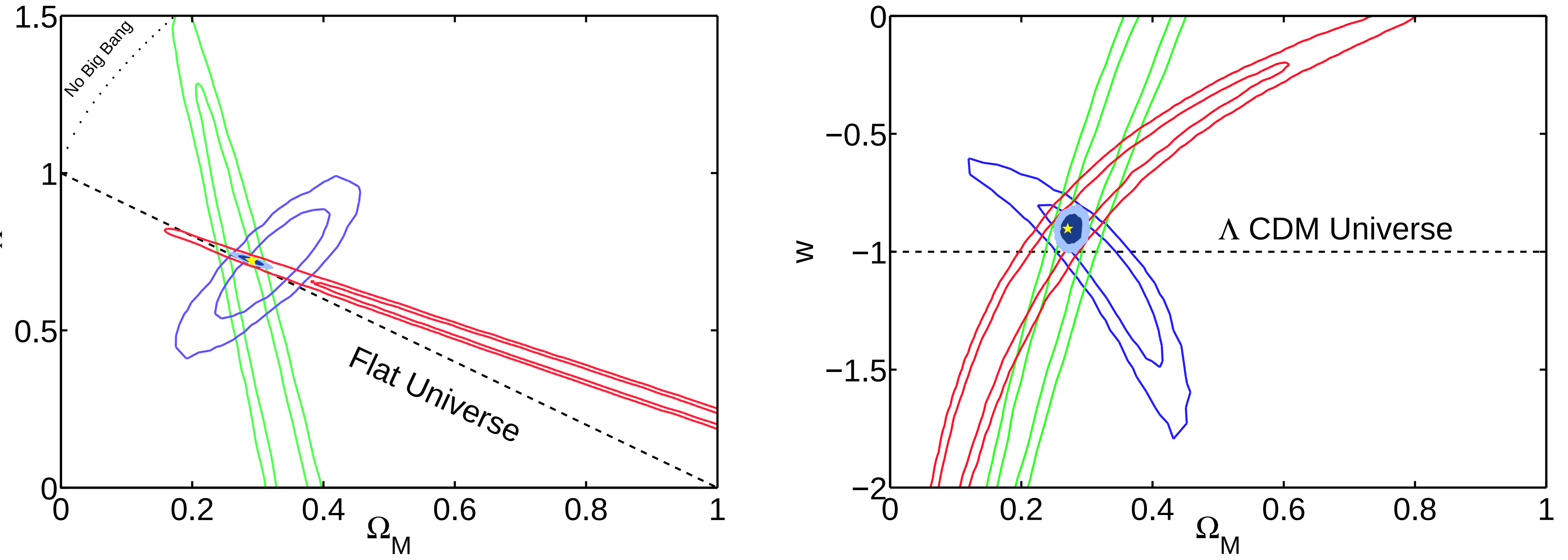
Preliminary results: real data





- Plots adapted from [Kessler,2009] arXiv:0908.4274 288 Supernovae (Nearby+SDSS+ESSE+SNLS+HST)
- Contours: Red/yellow is Kessler 2009, blue is new Bayesian method.
- Yellow star is mean value of posterior.
- Bayesian method: $\sigma_{\mu}^{\text{int}} = 0.12 \pm 0.01$
- Caveat: currently $\sigma_{m_B x_1}$ and $\sigma_{m_B C}$ not fully treated in new Bayesian method.

Preliminary results: real data

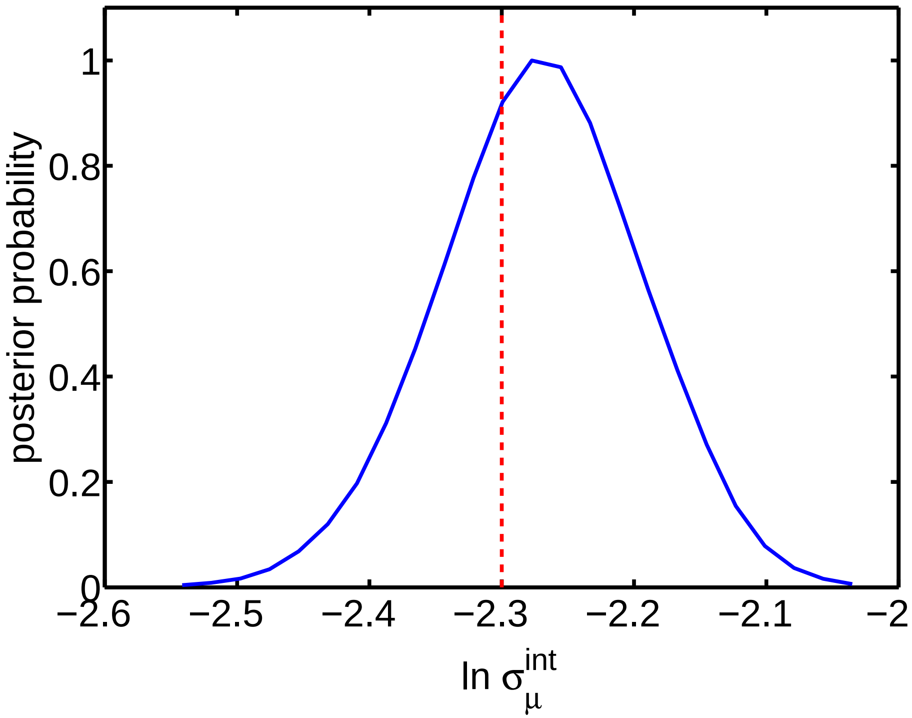


- Data adapted from [Kessler,2009] arXiv:0908.4274 288 Supernovae (Nearby+SDSS+ESSE+SNLS+HST)
- Contours: Shaded = Combined; Blue = SNe; Red = CMB; Green = BAO.
- Yellow star is mean value of posterior.
- Caveat: currently $\sigma_{m_B x_1}$ and $\sigma_{m_{BC}}$ not fully treated in new Bayesian method.

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Preliminary results: real data

- Outperforms chisquare method for parameter inference:
  - More accurate in 60-70 % of trials.
  - Tighter constraints on parameters
- Gives a distribution for σ_{μ}^{int}.
- Provides a framework for modeling for the systematic errors.



- Fully Bayesian $\longrightarrow$ can be used with other Bayesian techniques: Bayesian model selection.
- All parameters treated in the same way $\longrightarrow$ can sample posterior using MCMC / nested sampling methods.

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Applications:

- Bayesian model selection: can now validly use SNIa data to discriminate between ΛCDM, wCDM, flat models, curved models, w_0, w_a models, void models....
- Do SNIa evolve with redshift? Extend parameter space to test for different $\sigma_\mu^{int}(z)$ and different $\beta(z)$.
- Are there different SNIa populations? Differentiate by host galaxy type and recover parameters.
- What is the effect of outliers on the ability to recover the cosmological parameters? Can we detect outliers through their effect on the pdf of σ_μ^{int}?
- Extract systematic error for each survey: σ_μ^{int}(survey)

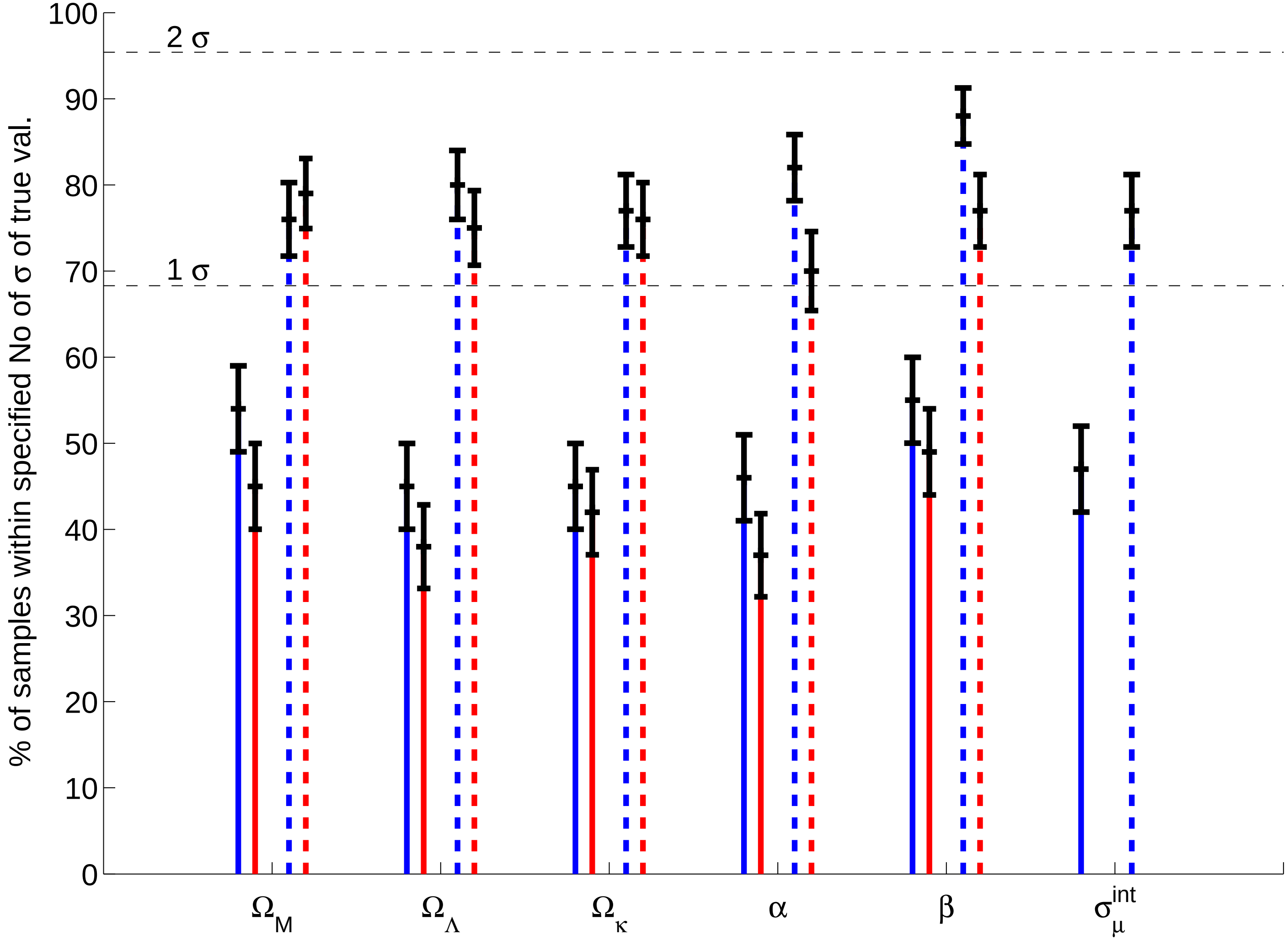
marisa.march06@imperial.ac.uk
arXiv:1102.3237

Work in progress

- Currently testing our method with ΛCDM and flat wCDM models with HST prior on H_0, and with flat prior on H_0.
- Testing with real data, including full treatment of correlations.
- Further testing with more realistic SNANA data.

marisa.march06@imperial.ac.uk
arXiv:1102.3237

Coverage plot



- Results from 100 realizations of simulated data.
- Blue shows our results; red shows results from chisquare method.
- Our method recovers true parameter as well as chisquare method.
- H_0 prior is: $H_0 = 70 \pm 8$ km/s/Mpc

marisa.march06@imperial.ac.uk
arXiv:1102.3237