FERMILAB-SLIDES-18-079

THE DARK ENERGY SURVEY

# Bayesian Hierarchical Models for parameter inference with missing data: Supernova cosmology case study.

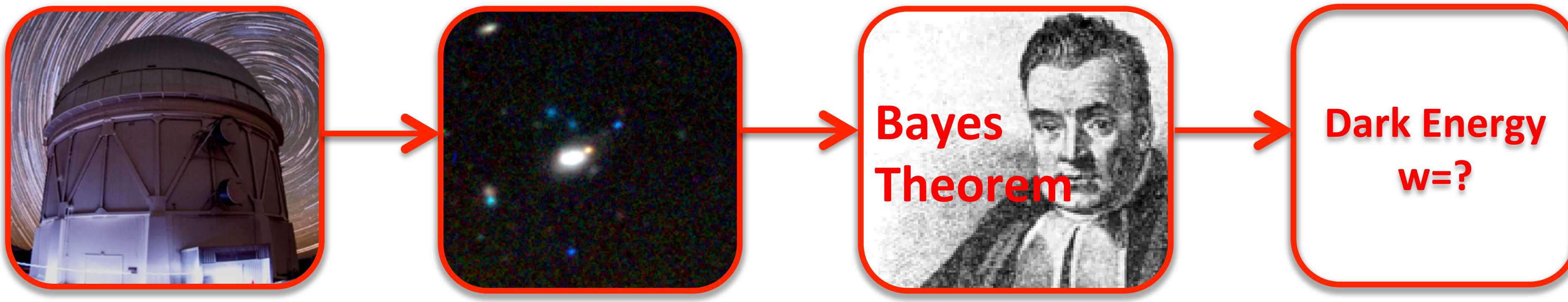
Marisa Cristina March, University of Pennsylvania

New PerspecĀves, Fermilab 18th-19th June 2018

This document was prepared by [DES Collaboration] using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.

Penn
UNIVERSITY of PENNSYLVANIA

# Mo) vaĀon & Overview




## Mission: To understand the nature of dark energy

- Using the **Dark Energy Camera** to search for Supernovae Ia
- Using **Bayes Theory** to do the staĀs) cal analysis in order to understand the nature of **dark energy**.
- Specific challenge addressed in this talk:
  - How to deal with **missing data** (magnitude limited survey) in a **Bayesian** way, in order to:
    - Use Supernovae Ia to do **Bayesian Model Selection**
    - Understand and reduce **systematics**

# Physics concept: Using standard candles to measure dark energy

**If you have objects of a standard brightness, you can work out how far away they are based on how bright they appear to be.**

Define the 'observed' distance modulus, to be the difference between the apparent (observed) and absolute magnitudes (brightness) of your standard object:

$$\mu^{\text{observed}} = m_B - M_0$$

absolute magnitude (← $M_0$)

apparent magnitude (→ $m_B$)

The theore) cal distance modulus depends on the redshiX and the cosmological parameters:

$$\mu^{\text{theory}} = f\{z, \Omega_m, \Omega_\kappa, \Omega_\Lambda, w(z)\}$$

dark energy equaĀon of state (→ $w(z)$)

redshift (→ $z$)

maĀer density (→ $\Omega_m$)

curvature density (→ $\Omega_\kappa$)

dark energy density (→ $\Omega_\Lambda$)

**Recipe:**

(1) Measure the apparent magnitude.

(2) Measure the redshiX.

(3) Work out what values the cosmological parameters must be to get:

$$\mu^{\text{theory}} = \mu^{\text{observed}}$$

THE DARK ENERGY SURVEY

# Evidence for Cosmic AcceleraĀon

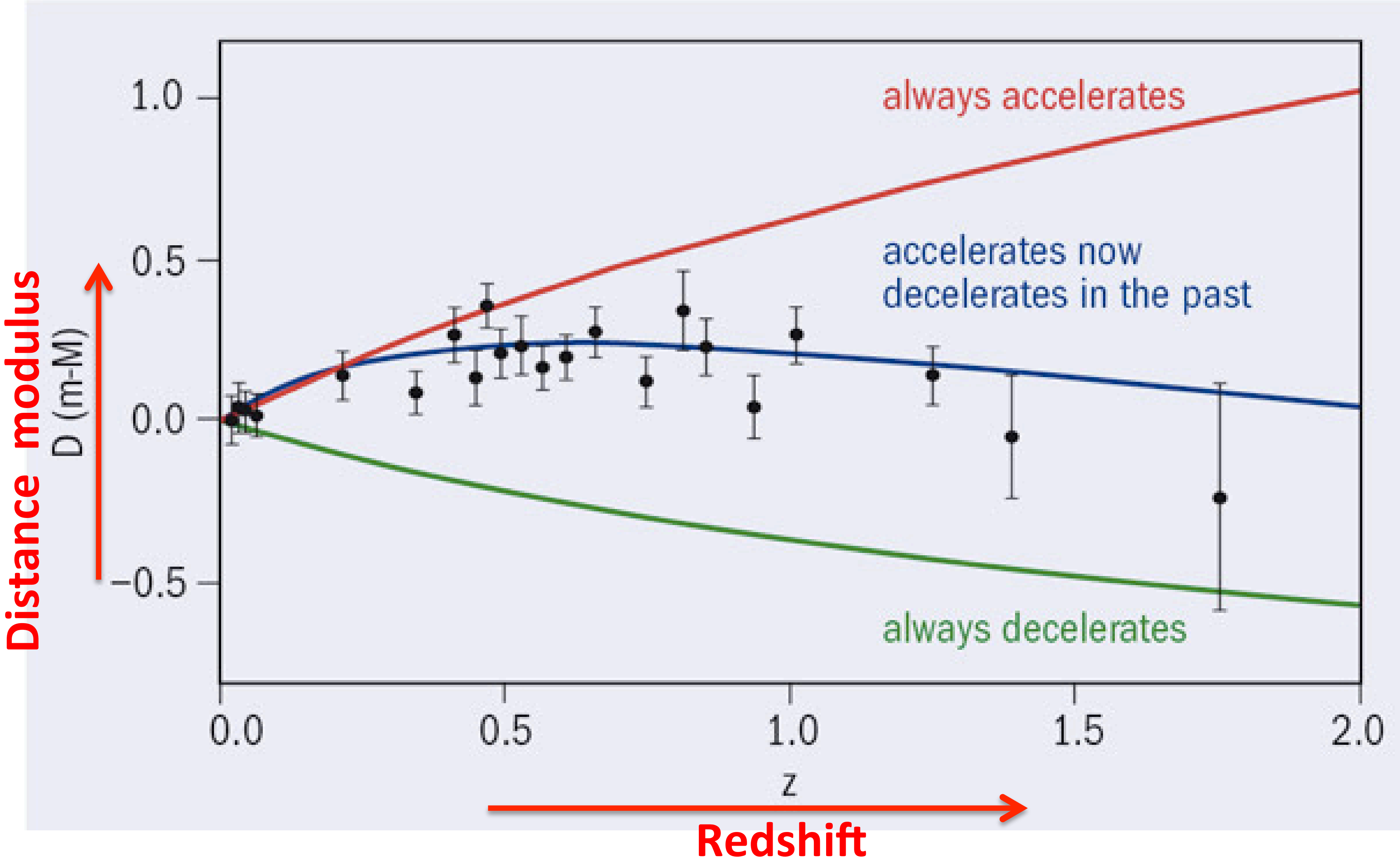



Image Credit: (A Riess et .).
Michael Turner, conference summary; Turner and Huterer 2007.

# Using Supernovae Type Ia as Standard Candles

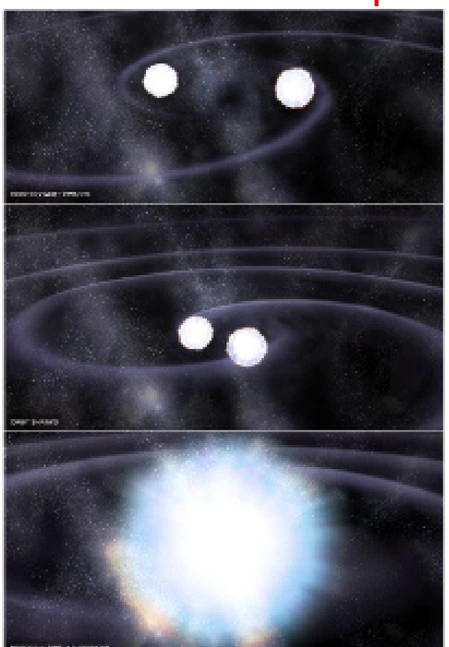
Use the stretch and color of the SNe light curves to apply small correc) ons to (i.e. to standardize) their brightness.

$$\mu^{\text{observed}} = m_B - M_0 + \alpha x_1 - \beta c$$

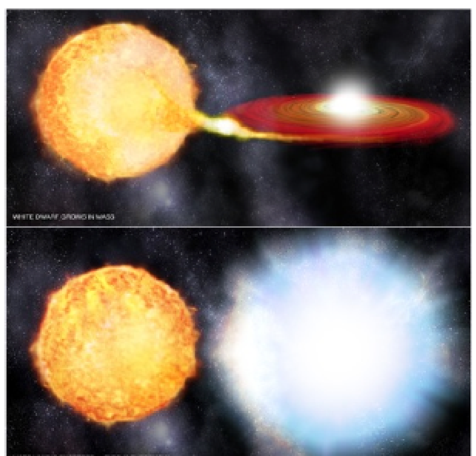
stretch

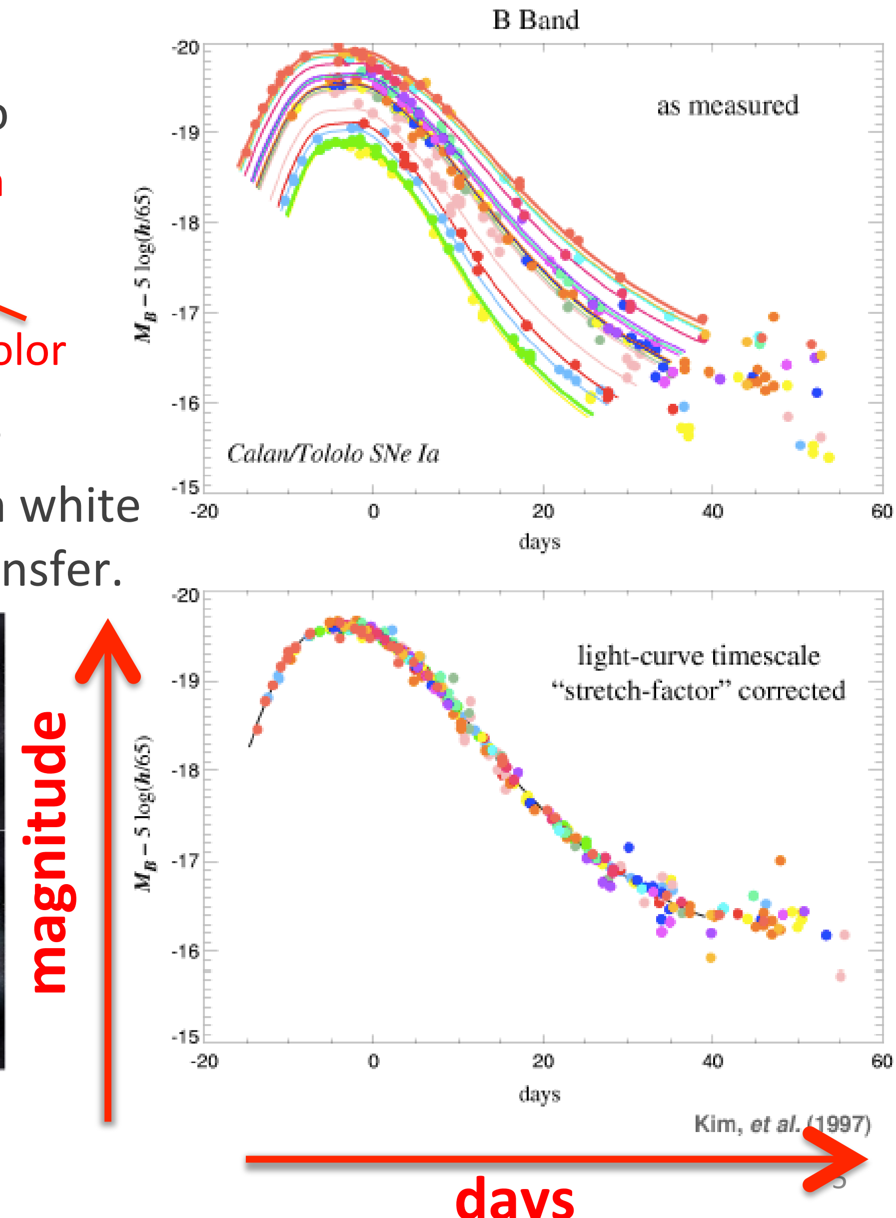
color

nuisance parameters



SNe Ia thermonuclear explosions come from white dwarf binary mass transfer.



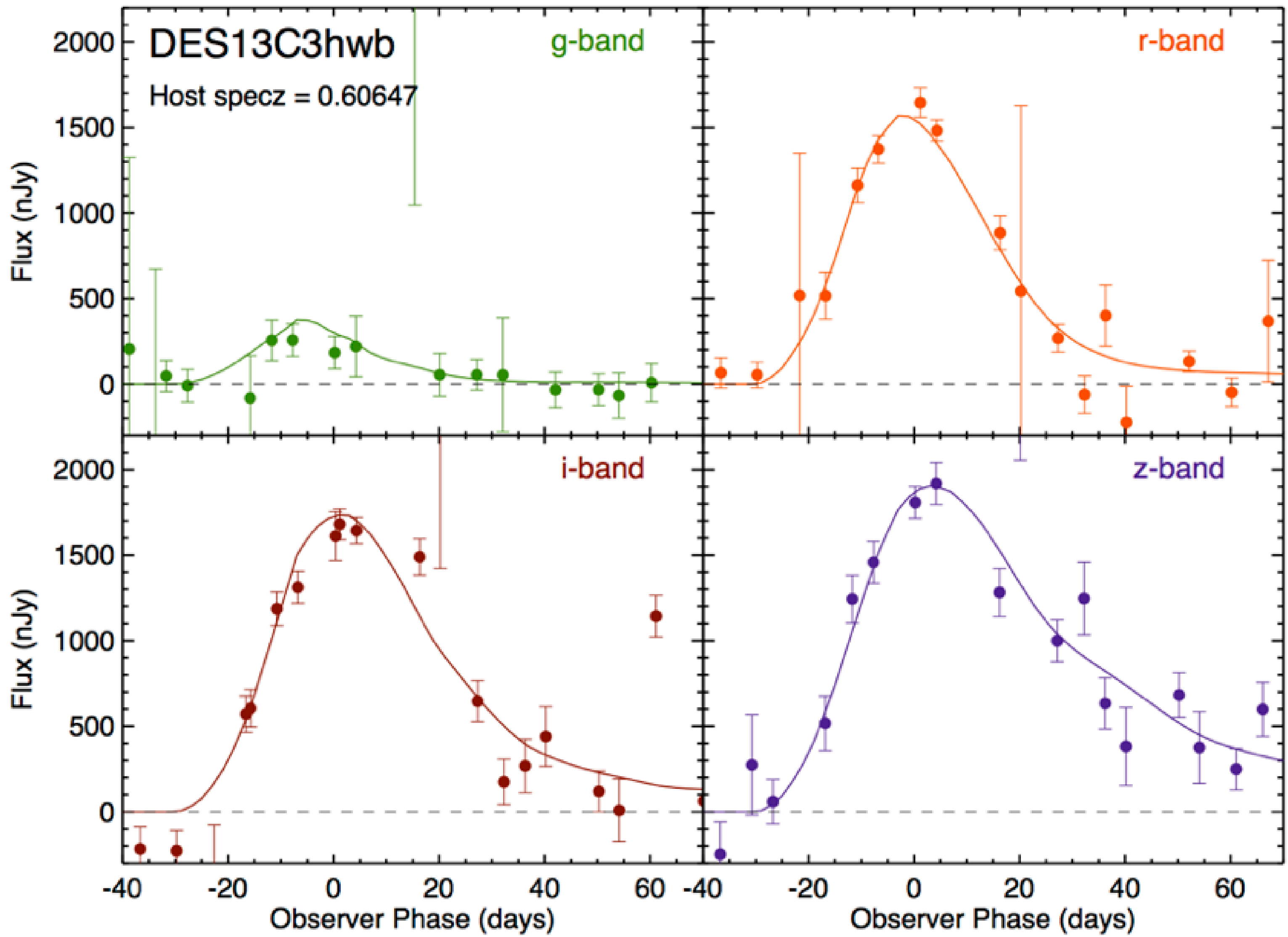



Kim, *et al.* (1997)



THE DARK ENERGY SURVEY

# Data: Supernova Light Curves




Plot credit: Chris D'Andrea

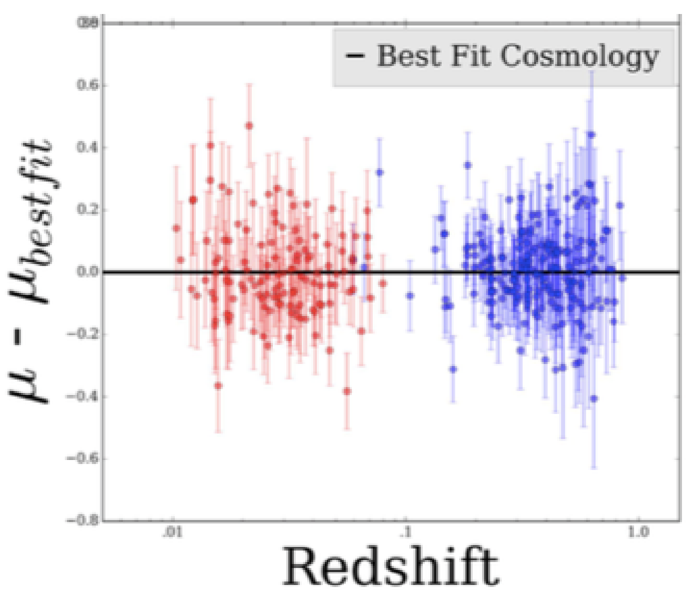
# Prelim DES Results! Flat wCDM

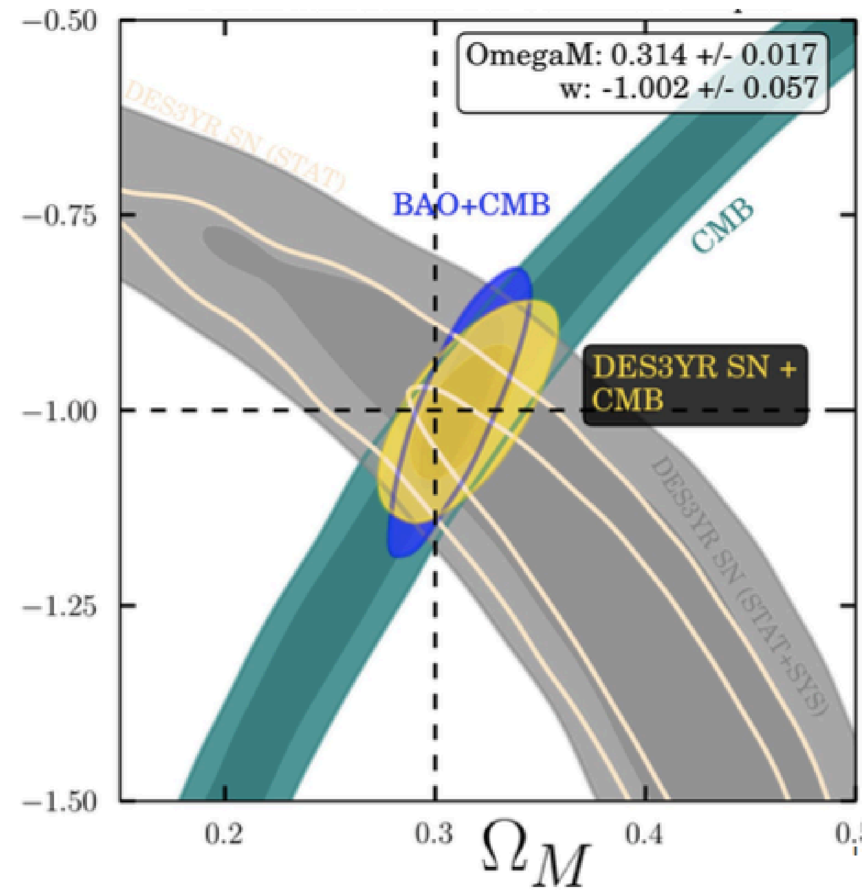
$$w = -1.002 \pm 0.057$$

$$\sigma_w = 0.041(STAT), 0.040(SYS)$$

The beginning of an era dominated by systematic uncertainties

$$\Omega_M = 0.314 \pm 0.017$$



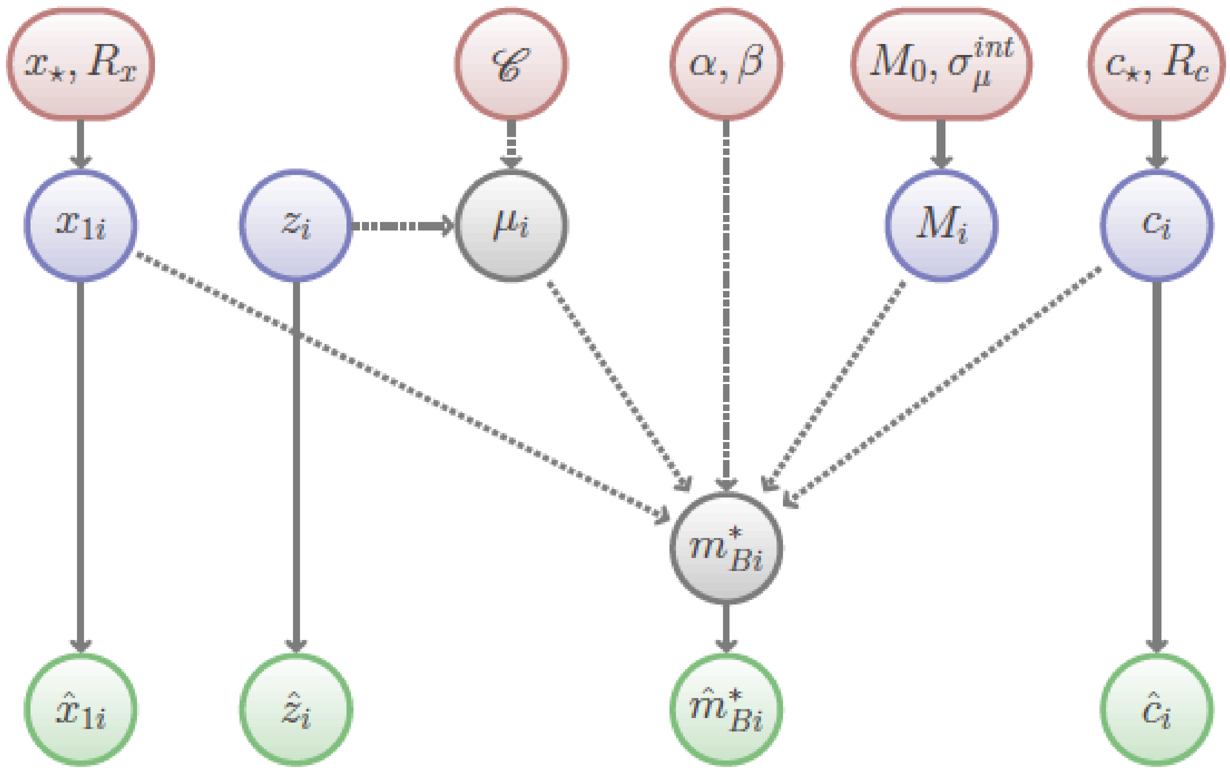

Slide & Plot Credit: Thanks to Dillon Brout!

*Beyond* the preliminary results:

- SystemaĀcs?
- Model Selec) on?

***Use Bayes Theory!***

# Supernova Bayesian Hierarchical Model



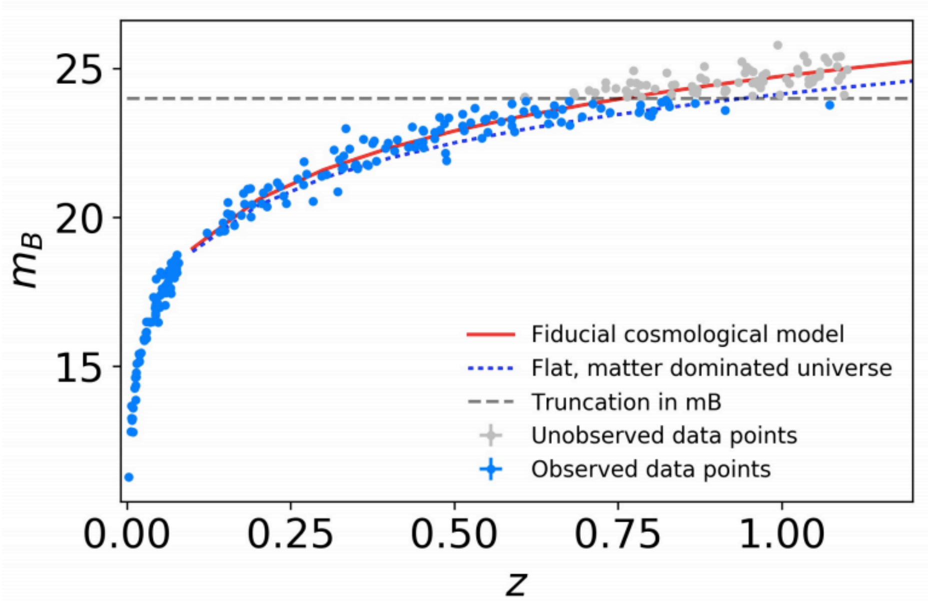

MM. et al. 2011

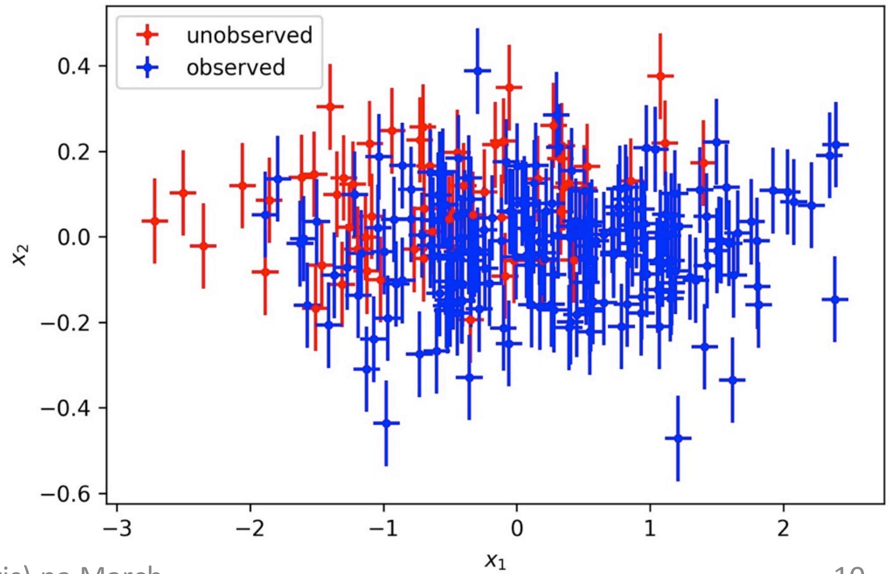
Allows use of Supernova data for Bayesian Model Selec) on.

- **Which model best explains dark energy?** LCDM, Modified Gravity? Scalar Field? Chameleon Field?
- Uses **latent** or hidden variables and **priors** to model observaĀonal data.

# Truncated data sets and Malmquist bias in SN cosmology

- Problem is that supernova data sets are **incomplete** in magnitude space. Limit of magnitude is set by instrument and environmental condionĀs.
- One solu) on is to **discard data** below a magnitude threshold. Disadvantage is **loss of information.**
- Another solu) on is to **simulate** surveys and "**correct**" mB data points to recover correct cosmology. Disadvantage is that this cannot be used for Bayesian model selec) on.
- **Alternative way:** Bayesian Hierarchical Model.



# Analy) c solu) on for Malmquist bias (missing data) in Supernova Bayesian Hierarchical Model

$$p(\mathscr{C}, \alpha, \beta | x_1^{\rm obs}, c^{\rm obs} m_B^{\rm obs}, z^{\rm obs}, m_B^{\rm thresh} I, M)$$

$$\propto \int_{N_{\rm obs}}^{\rm inf} dN \iint d\mathbf{R_x}\, d\mathbf{x}_\star \frac{1}{N} \binom{N}{N_{\rm obs}} \prod_i^{N_{\rm obs}} |2\pi \Sigma_{v,i}^{\rm obs}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left((\hat{\mathbf{w}}_{\mathbf{i}}^{\rm obs} - \mathbf{q}_i)^T \Sigma_{v,i}^{\rm obs} (\hat{\mathbf{w}}_{\mathbf{i}}^{\rm obs} - \mathbf{q}_i)\right)\right)$$

$$\times \prod_i^m \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{m_B^{\rm thresh}}^{+\infty} \int_0^{\hat{z}^{\rm max}} d\hat{x}_{1i}^{\rm mis} d\hat{c}_i^{\rm mis} d\hat{m}_{Bi}^{\rm mis} d\hat{z}_i^{\rm mis} |2\pi \Sigma_{v,i}^{\rm mis}|^{-\frac{1}{2}}$$

$$\times \exp\left(-\frac{1}{2}\left((\hat{\mathbf{w}}_{\mathbf{i}}^{\rm mis} - \mathbf{q}_i)^T \Sigma_{v,i}^{\rm mis} (\hat{\mathbf{w}}_{\mathbf{i}}^{\rm mis} - \mathbf{q}_i)\right)\right)$$

$$\times p(R_x, x_\star | I, M) p(\mathscr{C}, \alpha, \beta | I, M)$$

Posterior probability of parameters in a truncated data set

$$\mathbf{x}_\star = [x_{1,\star} c_\star] \in \mathbb{R}^2$$

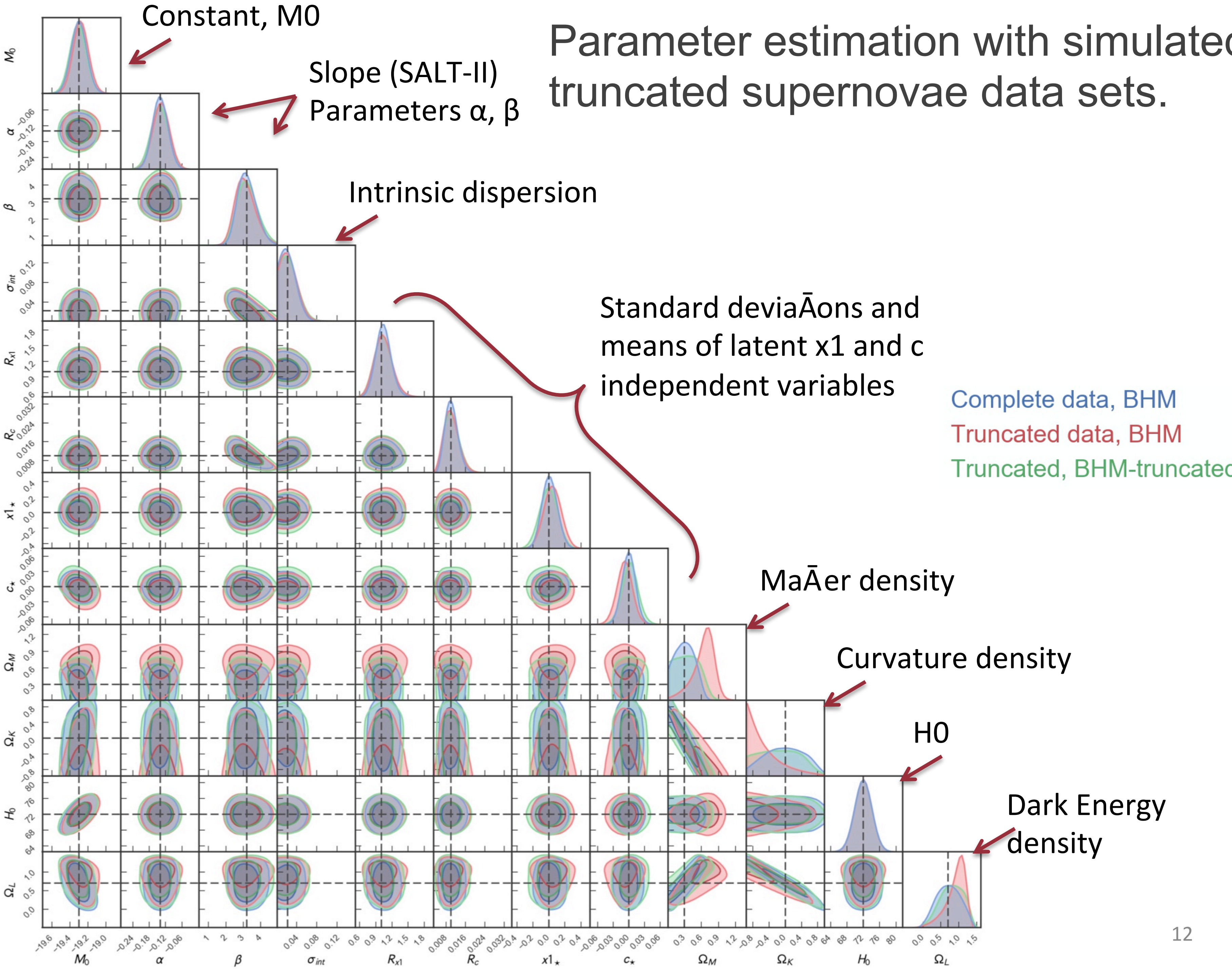
$$\mathbf{R_x} = \begin{bmatrix} R_{x1} & 0 \\ 0 & R_c \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

$$\hat{\mathbf{w}}_{\mathbf{i}} = \begin{bmatrix} \hat{m}_{B,i} \\ \hat{x}_{1,i} \\ \hat{c}_i \end{bmatrix} \in \mathbb{R}^3$$

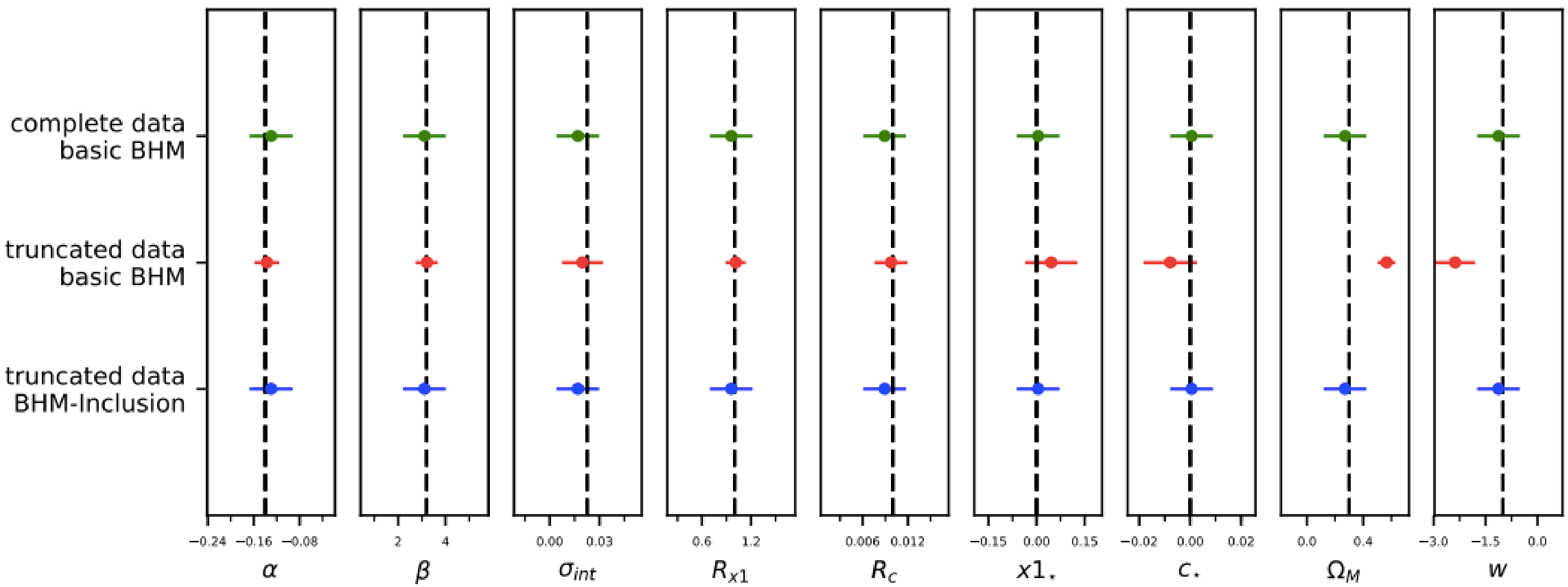
$$\Sigma_{c,i} = \begin{bmatrix} \sigma_{mi}^2 & \sigma_{mi,x1i} & \sigma_{mi,ci} \\ \sigma_{mi,x1i} & \sigma_{x1i}^2 & \sigma_{c,x1i} \\ \sigma_{mi,ci} & \sigma_{x1i,ci} & \sigma_{ci}^2 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

arXiv:1804.02474

# Parameter estimation with simulated truncated supernovae data sets.



# Summary & Next Steps




**Summary:**

- If you want to do Bayesian **Model selection**, you need to have the correct Bayesian **Posterior.**
- How do you account for **missing data** in a Bayesian way?
- See arXive: **1804.02474**

**Done:**

Analy) c solu) on for missing data.
Tested on basic simulaĀons.

**Next Steps:**

- Include refined selec) on func) on, test on SNANA DES like simulaĀons.
- Account for uncertainty in **typing.**