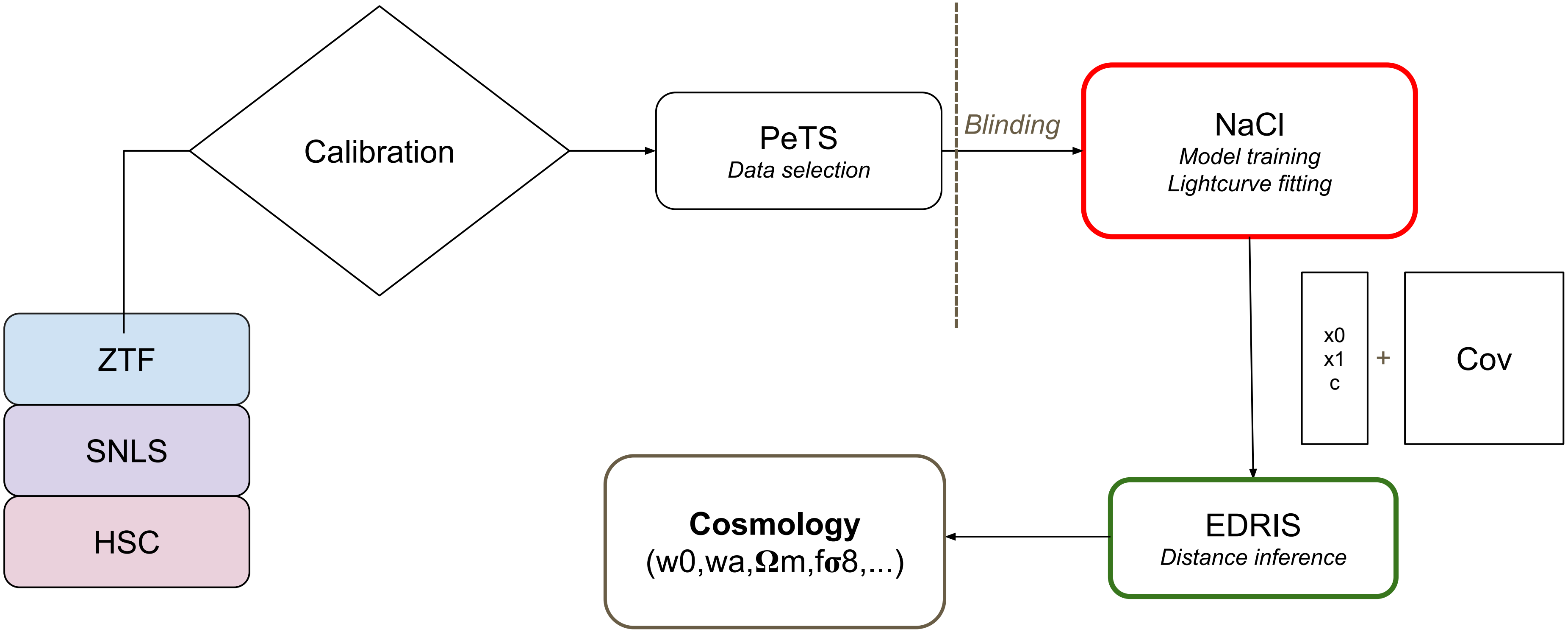
EDRIS : concept and application

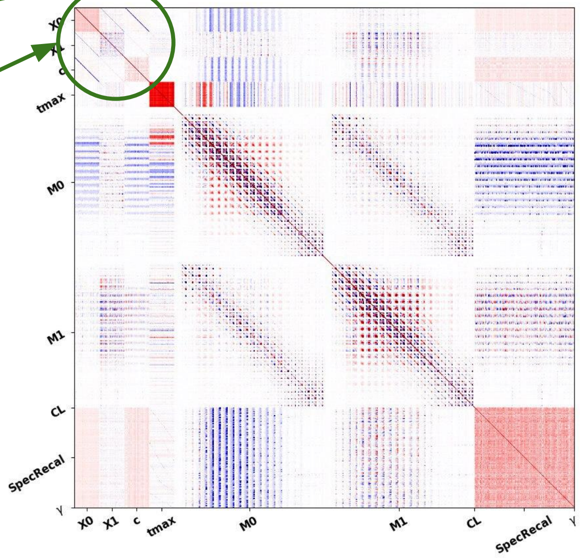
Dylan Kuhn
ZTF meeting Barcelona 2024

The Lemaitre analysis pipeline



Cosmological inference

$$\begin{bmatrix} x_0 \\ x_1 \\ c \\ t_{max} \\ M_0 \\ M_1 \\ CL \end{bmatrix}$$



1- Standardization

$$\mu(z,\theta) = -2.5\log_{10}(x_0) - M + \alpha x_1 + \beta c$$

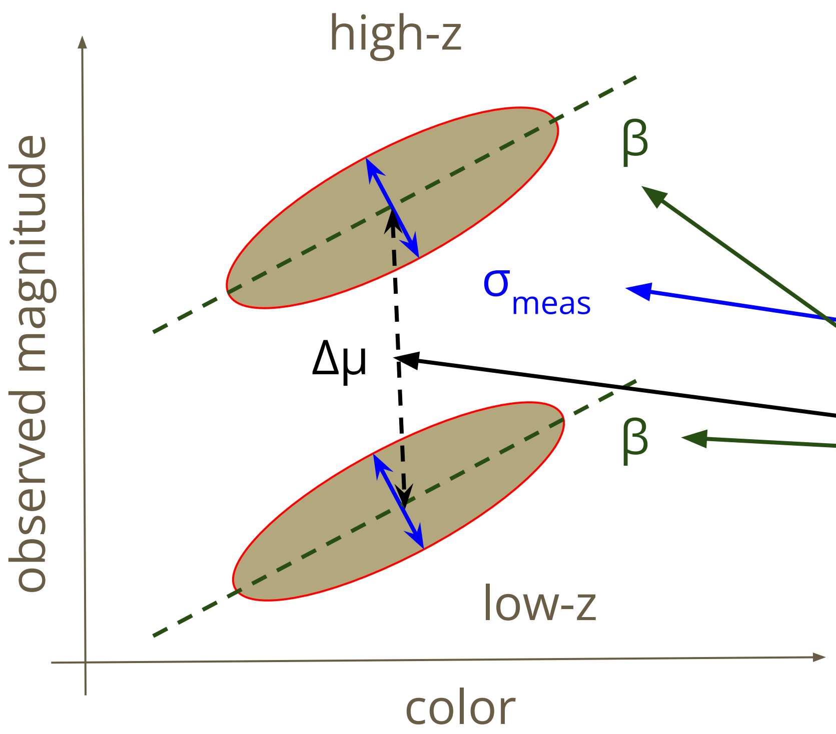
~45% dispersion at max

~15% dispersion at max

2- Estimation of residual dispersion σ

3- Estimation and correction of intrumental selection bias Δμ(σ)

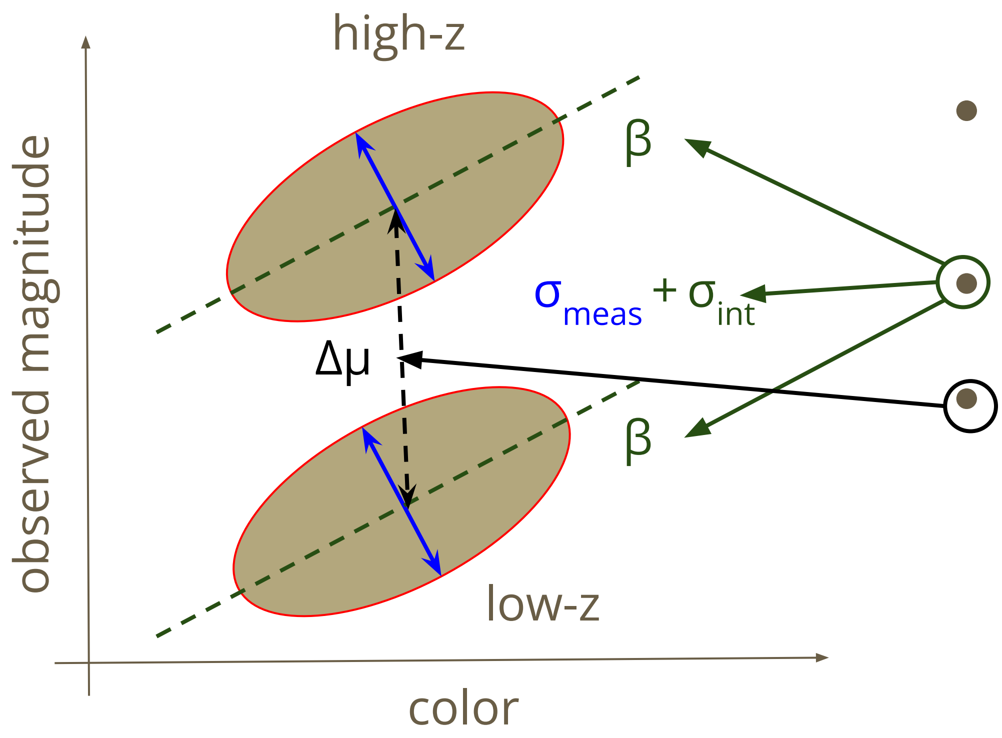
Instrumental selection bias: the “Malmquist bias”



1 survey, 2 populations (low-z, high-z)

- known measurement noise
- *estimation* of β, Δμ

Instrumental selection bias: the "Malmquist bias"



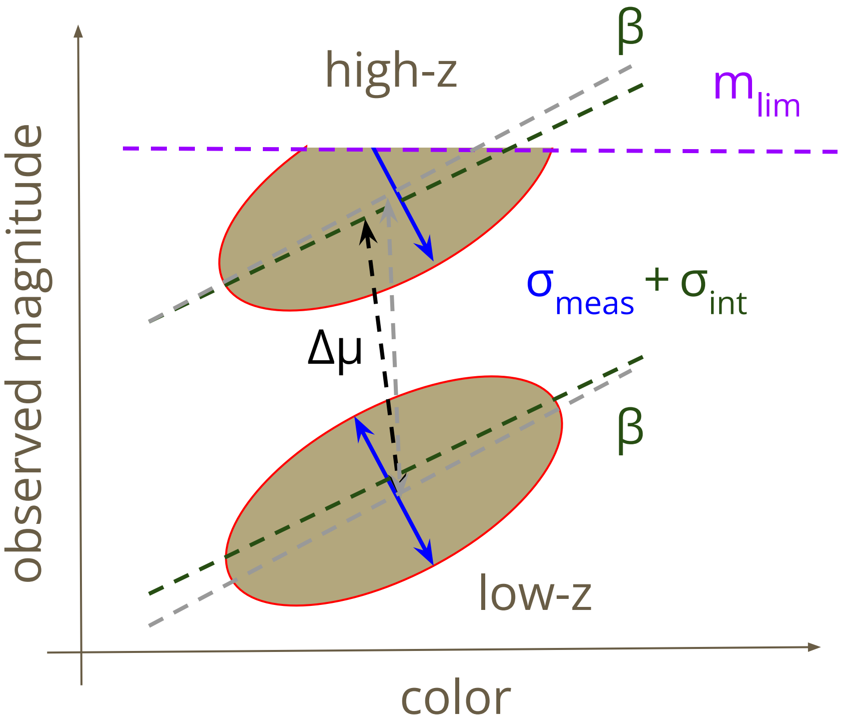
- known measurement noise
  + *unknown intrinsic noise*

biased estimation of β, σ_{int}

estimation of Δμ

Not well defined problem but cosmology does not change

Instrumental selection bias: the "Malmquist bias"



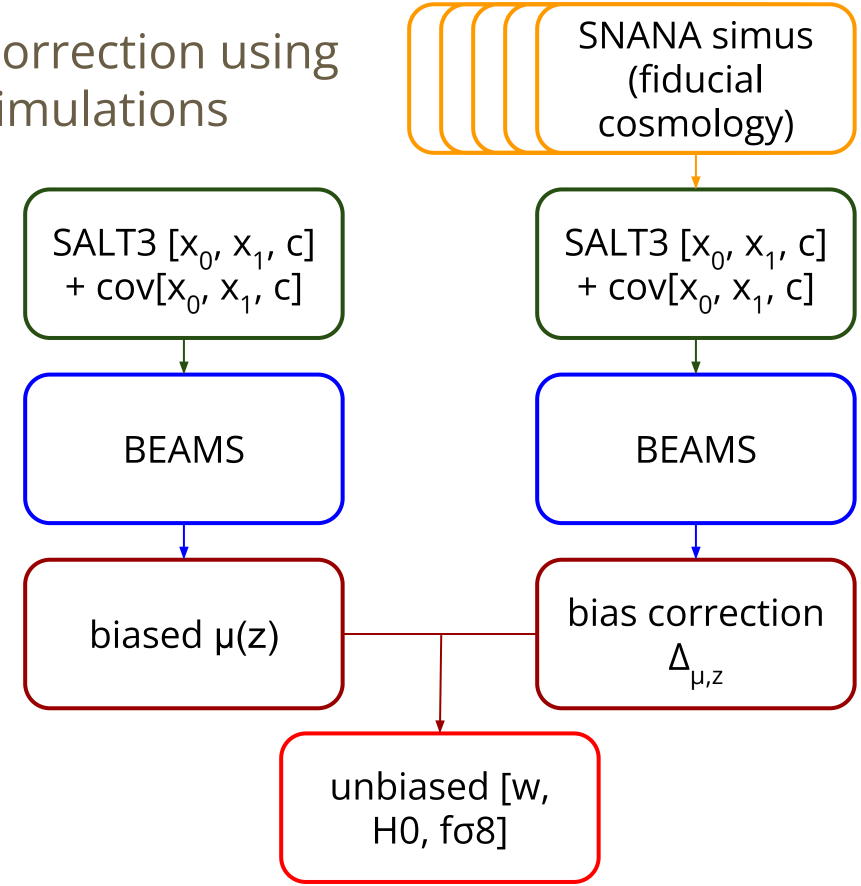
In practice, only the *intrinsically brightest supernovae* are detected:

- *truncation of data* by m_{lim}
- *biased* estimation of β, σ_{int}, Δμ

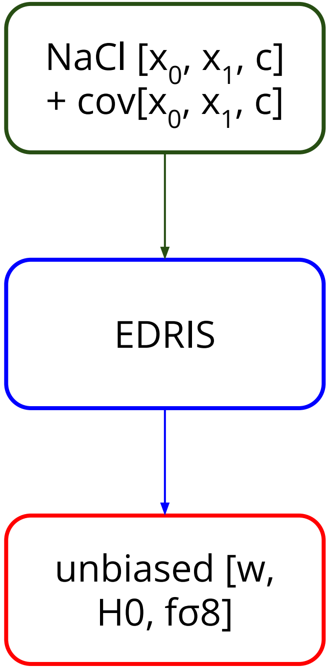
Not well defined problem and cosmology is biased by truncation

How to tackle this issue

Correction using simulations



Account for the selection effects in the statistical model



Our approach: NaCl + EDRIS

EDRIS:

- cosmology from NaCl [x_0, x_1, c]

- includes selection in statistical model (simulations are only needed to test the pipeline)

$$m_{obs,i} = m_{obs,i}^{*} + \eta_i \; if \; m_{obs,i}^{*} \leq m_{lim} + \kappa_i$$

$$with \; \eta_i \sim \mathcal{N}\left(0, C_i\right) \; and \; \kappa_i \sim \mathcal{N}\left(0, \sigma_{m_{lim}}^{2}\right)$$

$$m_{obs,i} \; is \; unobserved \; otherwise$$

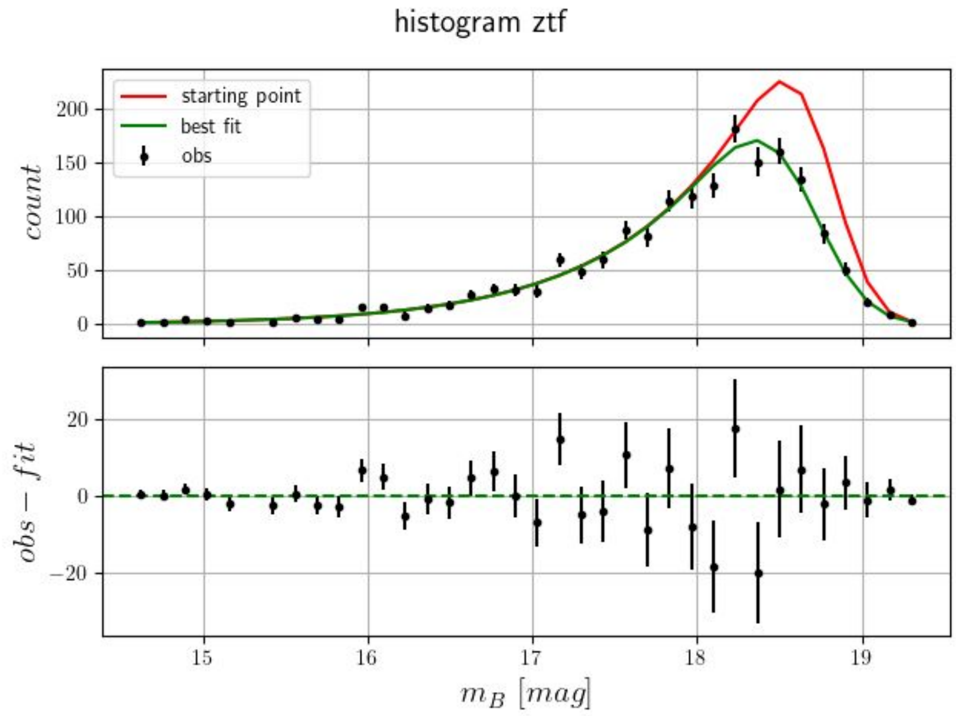
Two-step estimator:

- estimation of the selection functions [m_{lim}, σ_{mlim}] from m_{obs} histograms

- standardization & estimation of distances

Estimation of the selection function

Estimation of [m_{lim}, σ_{mlim}] for each survey from observed magnitudes histogram



Model of the selection function

Density of SNeIa

$$\rho(z,m) = R\frac{\partial V_c}{\partial z}(z)\,\Phi\left(\frac{m - m_{lim}}{\sigma_{m_{lim}}}\right)\frac{1}{\sqrt{2\pi}\sigma_m}e^{-\frac{1}{2}\left(\frac{m-\mu(z)-M}{\sigma_m}\right)^2}$$

SNeIa rate (supposed constant), one parameter per survey

Selection function

$$\sigma_m = \sqrt{\sigma^2 + \alpha^2\sigma_{x_1}^{\;2} + \beta^2\sigma_c^{\;2}}$$

For now, [α, β, σ, M, cosmology] are fixed to realistic values
→ later, volume will be replaced with a smooth generic function (polynomial) with shape parameters for each survey
→ uncertainties on dVc correctly propagated

Estimation of the x_1 and c distributions

→ Estimation on measured x1 and c
→ Per survey
→ Cut at completion redshift to avoid biases

$$\begin{bmatrix} \overline{x_1} \\ \overline{c} \\ \sigma_{x_1} \\ \sigma_c \end{bmatrix} + cov\left(\sigma_{x_1}, \sigma_c\right)$$

Parameters of interest
→ Will be used as a prior

Estimation of the selection function

Number of SNeIa in a bin

$$N_b = \int_0^{\infty} \int_{m_b}^{m_{b+1}} \rho(z, m)\, \mathrm{d}z\, \mathrm{d}m$$

Poisson likelihood

$$\mathscr{L}_{poisson} = \prod_i \frac{N_{b,i}^{\;N_{obs,i}}}{N_{obs,i}!} e^{-N_{b,i}}$$

$$\mathscr{L}_{selection} = \sum_s \mathscr{L}_{poisson,s} + \mathscr{L}_{prior,distribution}$$

$$\begin{bmatrix} R \\ m_{lim} \\ \sigma_{m_{lim}} \end{bmatrix} + cov\left(m_{lim}, \sigma_{m_{lim}}\right)$$

Parameters of interest → will be used as a prior for next step

Standardization model

Noise
- intrinsic dispersion ε
- measurement, calibration, … η

Standardization coefficients (α, β, …)

Distances

$$\begin{pmatrix} m_i \\ Y_{1,i} \\ Y_{2,i} \\ \vdots \\ Y_{n,i} \end{pmatrix} = \begin{pmatrix} \mu_i(z,\theta) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_n \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X^\star_{1,i} \\ X^\star_{2,i} \\ X^\star_{3,i} \\ \vdots \\ X^\star_{n,i} \end{pmatrix} + \begin{pmatrix} \epsilon_i \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \eta_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

$$\eta \sim \mathcal{N}(0, Cov(m, Y))$$

LCs parameters (m_B, x_1, c, …)

Latent parameters (x_1*, c*, …)

Standardization & estimation of distances

term that takes into account the truncation of data

classic likelihood for multivariate normal distributions

$$\mathscr{L}_{cosmo} = -\ln\left(\left|C(\sigma)^{-1}\right|\right) + r^t C(\sigma)^{-1} r + \sum_i 2\ln\left(\Phi\left(\frac{m_{lim} - M^* - \mu_i - \alpha x_{1,i}^* - \beta c_i^*}{\sqrt{\sigma^2 + \sigma_{m_{lim}}^2}}\right)\right) - 2\ln\left(\Phi\left(\frac{m_{lim} - m_{obs,i}}{\sqrt{\sigma_{m_{lim}}^2 + f(C_i)}}\right)\right)$$

$$\textit{with } \Phi(z) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right) \textit{ and } r = Y_i - Y_i^*$$

Acceleration of the computation

Likelihood function computed in $O(N^2)$
2700 SN → model evaluated in 165 ms
Use of JAX for auto differentiation → efficient minimization

$$W = \begin{pmatrix} C_{mm} + \sigma^2 I_N & C_1 \\ C_1{}^t & C_2 \end{pmatrix}^{-1} \longrightarrow S^{-1} = Q\left(\Lambda + \sigma^2 I_N\right)^{-1} Q^t$$

Schur complement of C_2 *in* $C = W^{-1}$

At the end, only matrix-to-vector products

$$-\ln(|W|) = \ln(|C_2|) + \sum_i \ln\left(\Lambda_i + \sigma^2\right)$$

$$r = \begin{pmatrix} r_1 & r_2 \end{pmatrix} \longrightarrow r^t W r = r_1{}^t S^{-1} r_1 - 2 r_1{}^t S^{-1} C_1 C_2{}^{-1} r_2 + r_2{}^t C_2{}^{-1} r_2 + r_2{}^t C_2{}^{-1} C_1{}^t S^{-1} C_1 C_2{}^{-1} r_2$$

Standardization & estimation of distances

$$\mathscr{L}_{EDRIS} = \mathscr{L}_{\mathrm{cosmo}} + \mathscr{L}_{prior,\, selection} \rightarrow \begin{bmatrix} \mu_{bins} \\ \alpha \\ \beta \\ \sigma \\ x_1^* \\ c^* \end{bmatrix} + cov(\mu_{bins}) \rightarrow \begin{bmatrix} \Omega_m \\ w \\ \vdots \end{bmatrix} + cov(\Omega_m, w, \cdots)$$

→ Are the μ_{bins} biased ?

Lemaitre pre-DC2 simulations : simulation strategy

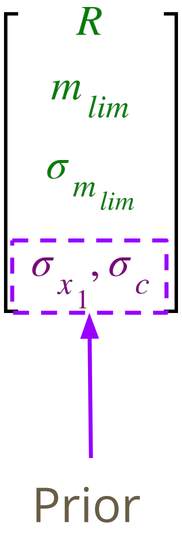
Goal : consistency check

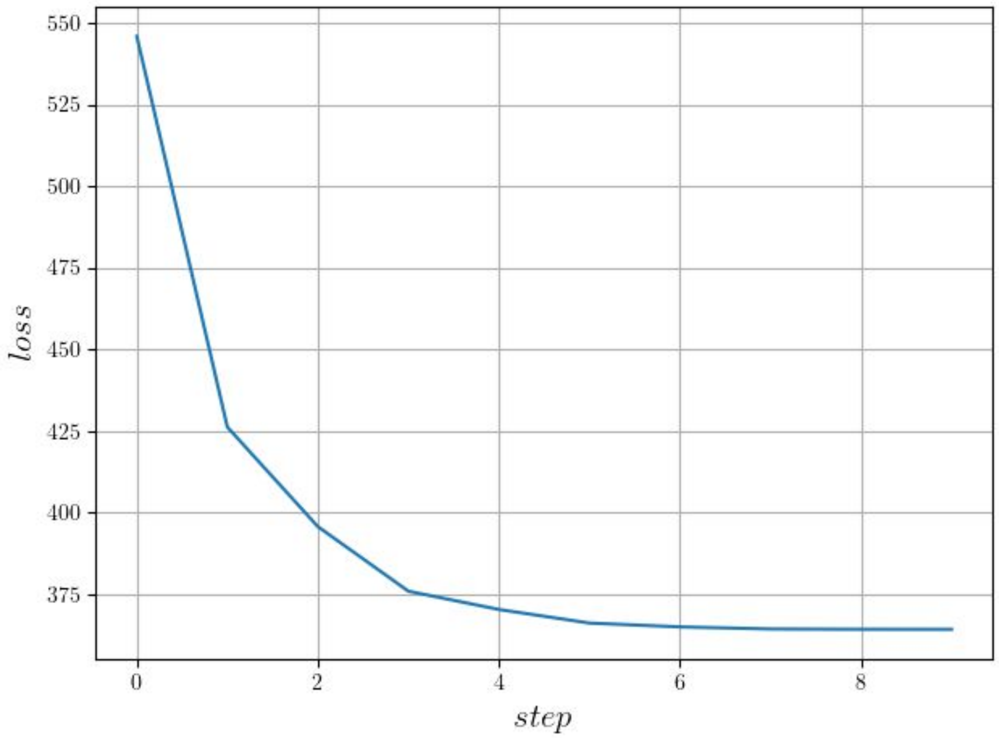
- From a DC1 NaCl fit, infer an error model for (x_0, x1, c) as a function of x_0

- With SkySurvey, draw (z, x_0, x1, c) for the three survey and draw errors (+ covmat) from the error model (x_1 and c are gaussian for now, covmat is purely diagonal)

- Do a selection on the observed B-band magnitudes to be consistent with the analysis model

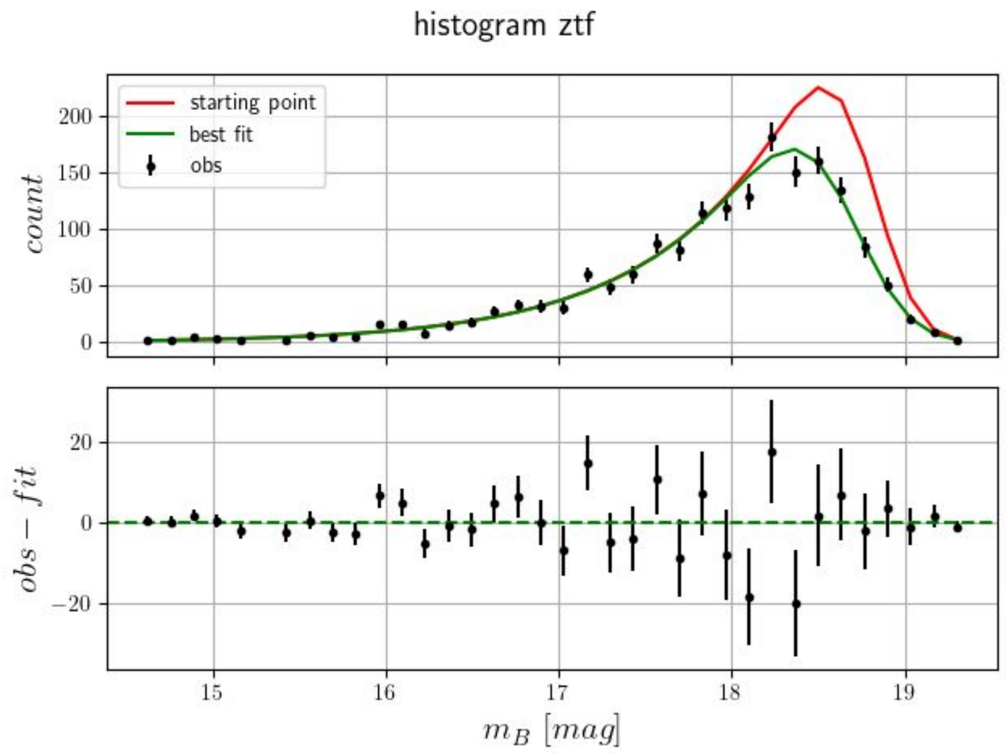
Lemaitre-like pre-DC2 simulation : EDRIS step 1



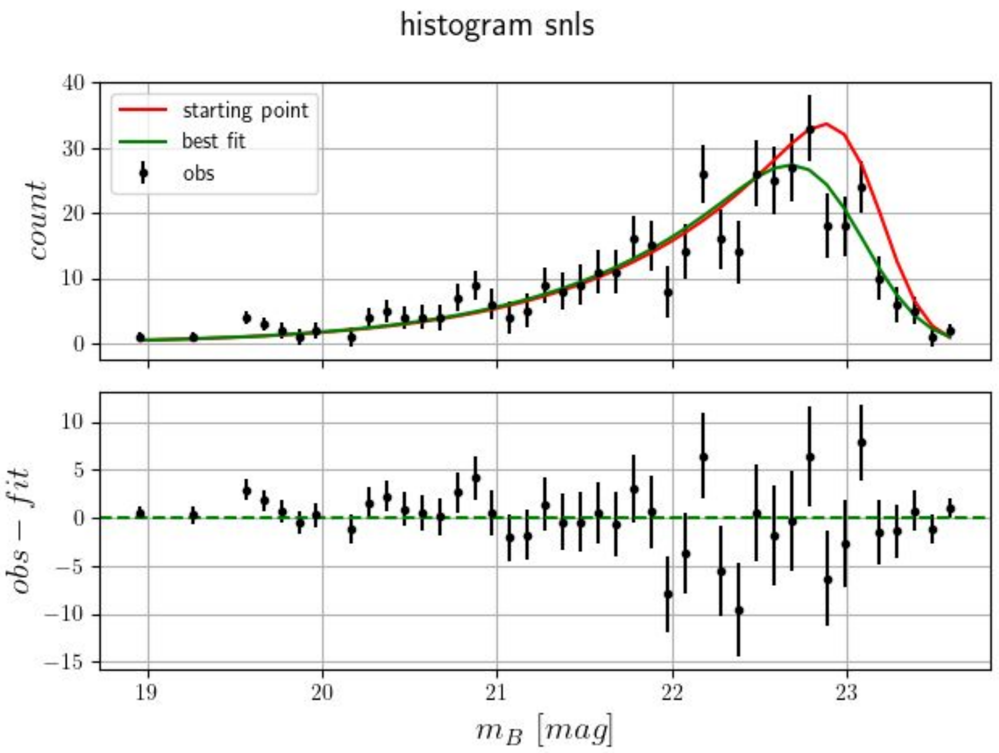


9 iterations
~ 3min

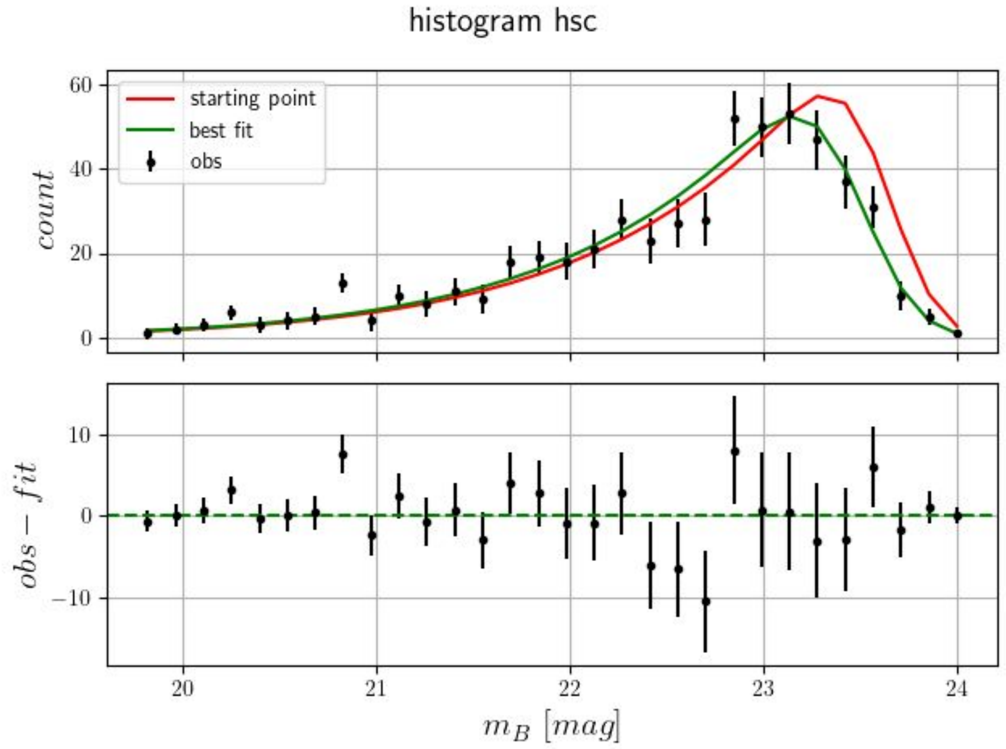
Lemaitre-like pre-DC2 simulation : ZTF selection fit



Lemaitre-like pre-DC2 simulation : SNLS selection fit



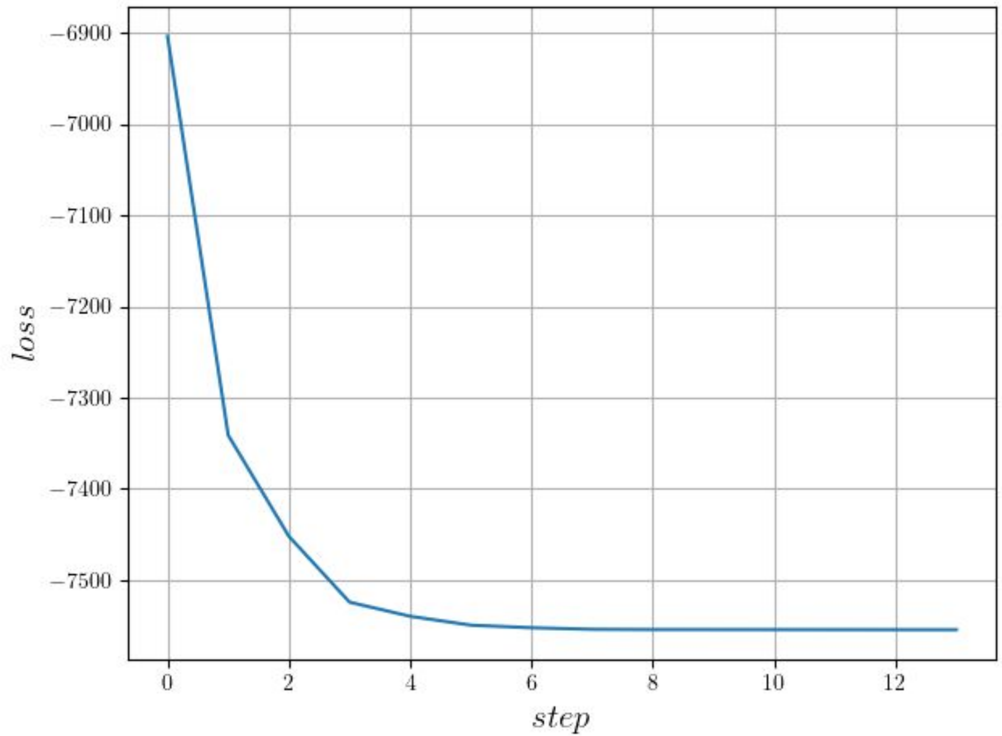
Lemaitre-like pre-DC2 simulation : HSC selection fit



Lemaitre-like pre-DC2 simulation : EDRIS step 2

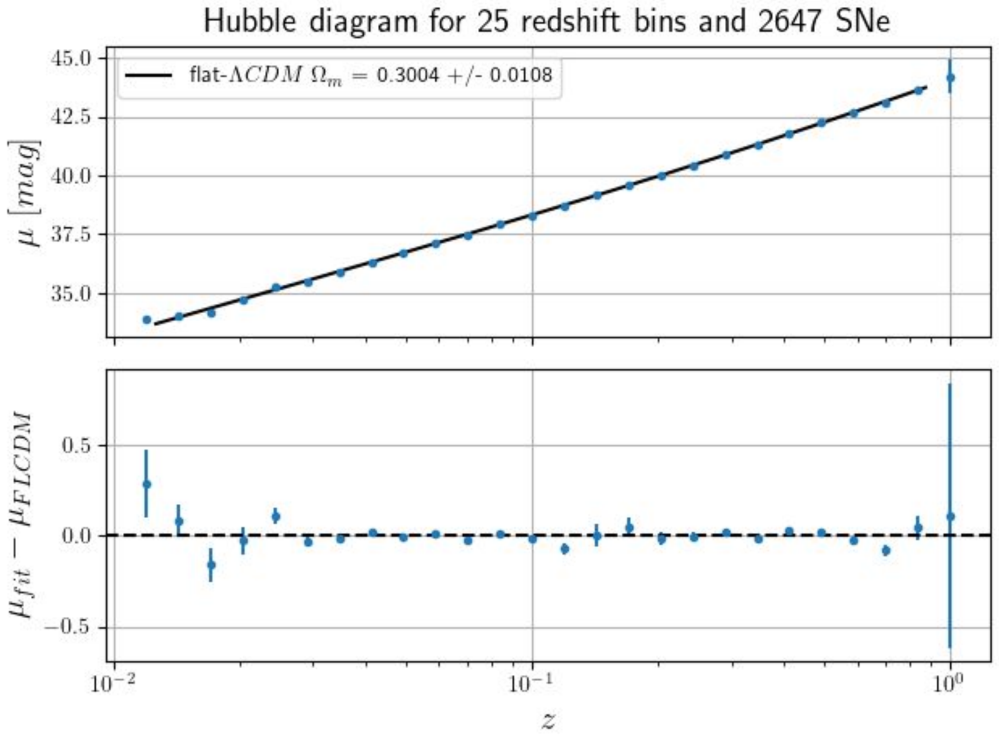
$$\begin{bmatrix} \mu_{bins} \\ \alpha_i \\ X_i^* \\ \sigma \\ m_{lim} \\ \sigma_{m_{lim}} \end{bmatrix}$$

Prior

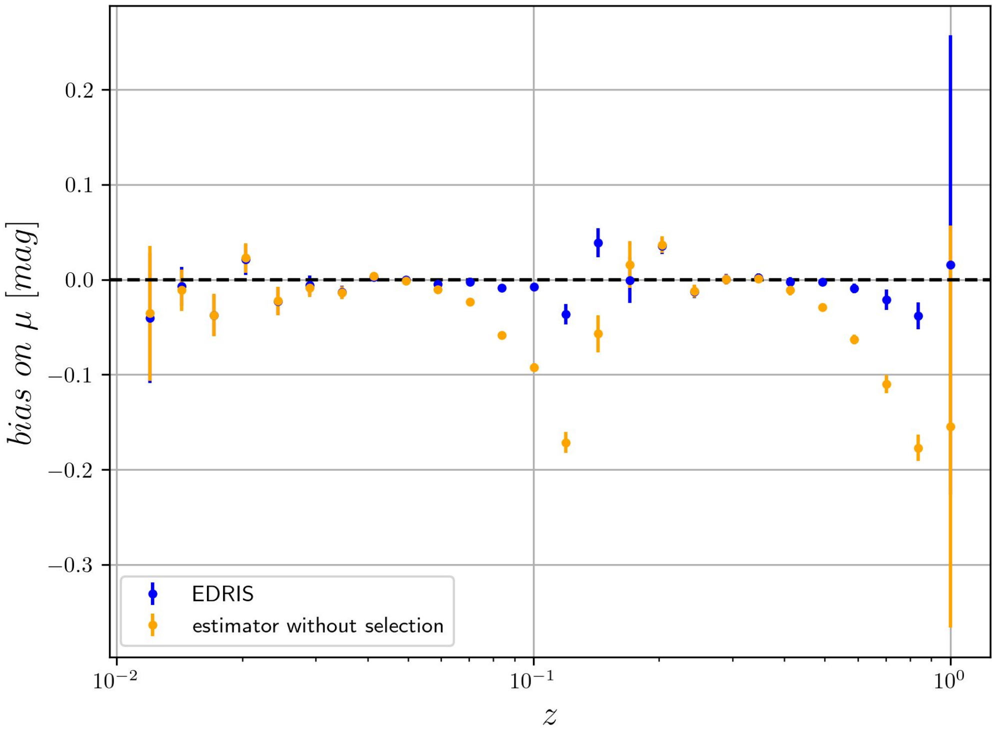


13 iterations
~ 30s

Lemaitre-like pre-DC2 simulation : binned Hubble diagram



Lemaitre-like pre-DC2 simulation : bias on μ_{bins}



Monte-Carlo with 10 pre-DC2 simulations

No bias on μ_{bins} when using EDRIS

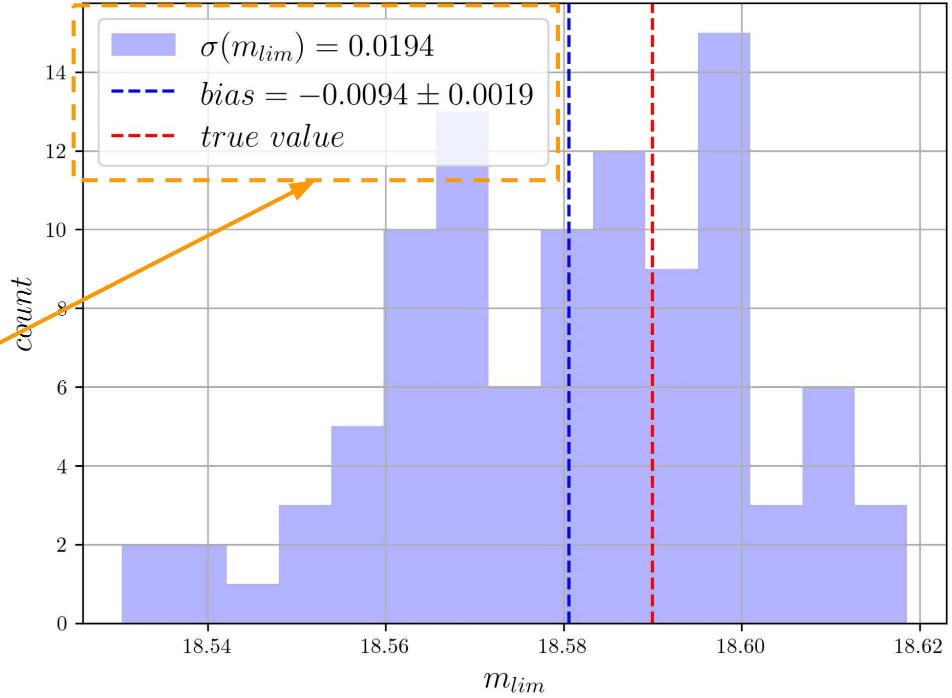
Effect of realistic x1 and c distributions on selection function measurement

100 ZTF-like simulations

→ Bimodal x_1, exponential tail c

→ Bias on m_{lim} 2x smaller than $\sigma(m_{lim})$

→ Effect on binned distances : work in progress



What's next for EDRIS ?

- pre-DC2 : overall encouraging results
- DC2 : key date → validation of the method on realistic simulations (several open questions)

  → what happens when NaCl is trained on truncated dataset ?

  → is EDRIS able to reconstruct unbiased cosmology ?

- DC3 : adding outliers rejection (contamination)
- DC4 : adding astrophysical effects (broken alpha, ...)
- DR2.5 paper : methodology paper