Unfolding Procedure

Recap

Andrea Carlo Marini

Massachusetts Institute of Technology

on behalf of the CMS Collaboration

LHCSW

Follow up …

Issues on the unfolding were previously discussed in the meeting of the 24/6/2015 [link]

Problems and issues specific to the unfolding method in the Higgs measurements, particularly to the $H \to \gamma\gamma$:

- Why Unfold, how and when …
- Signal Extraction
  - Large background to the analysis (e.g. in $H \to \gamma\gamma$, the $\gamma\gamma$ continuum)
  - Mass: profile or not profile

Importance of Unfolding

- Cross sections are computed using:
  - N_T events observed
  - N_O events coming from out-of-acceptance (fakes)
  - efficiency, acceptance and luminosity

$$\sigma = \frac{N_T - N_O}{\varepsilon \mathcal{L} \mathcal{A}}$$

- Errors are propagated: N_T is Poisson (data),
- N_O non-knowledge is modelled by systematics (or by other data bins)

$$\Delta\sigma = \frac{\Delta N_T}{\varepsilon \mathcal{A} \mathcal{L}} \qquad \frac{\Delta\sigma}{\sigma} = \frac{\Delta N_T}{N_T - N_O}$$

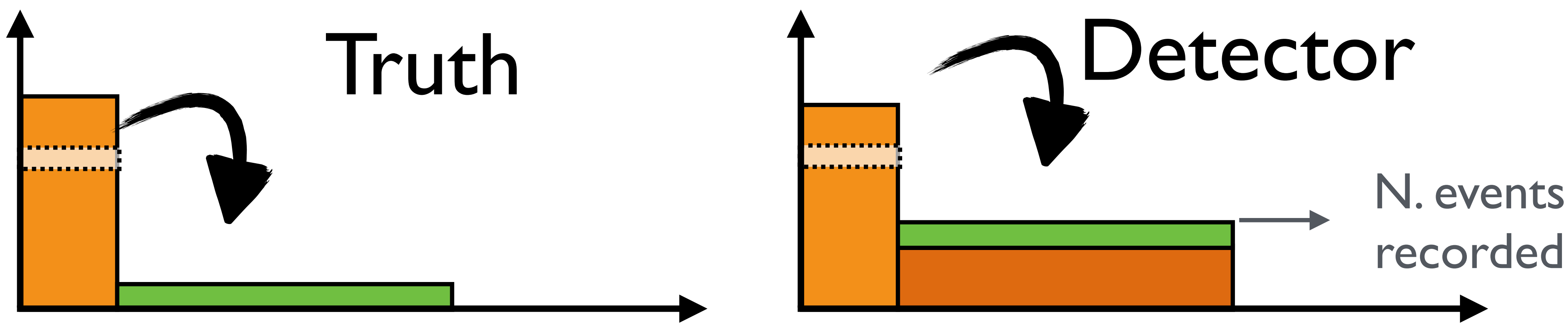
- For example if N_T=100 and N_O = 50 $\boldsymbol{\Delta\sigma/\sigma}$ = 10/(100-50) = 20%

Multiplicative factors underestimate the errors :

- $\boldsymbol{\Delta\sigma}$ = 10/100 = 10%

Importance of Unfolding II

- The same point can be obtained in migrations:



very well predicted (data/theory)
interesting / new physics

Statistical Propagation to the new bin need to take into account the precision of the "very well predicted" events.

If = 226 = 30 → Poisson error is = $\sqrt{256}$ =16

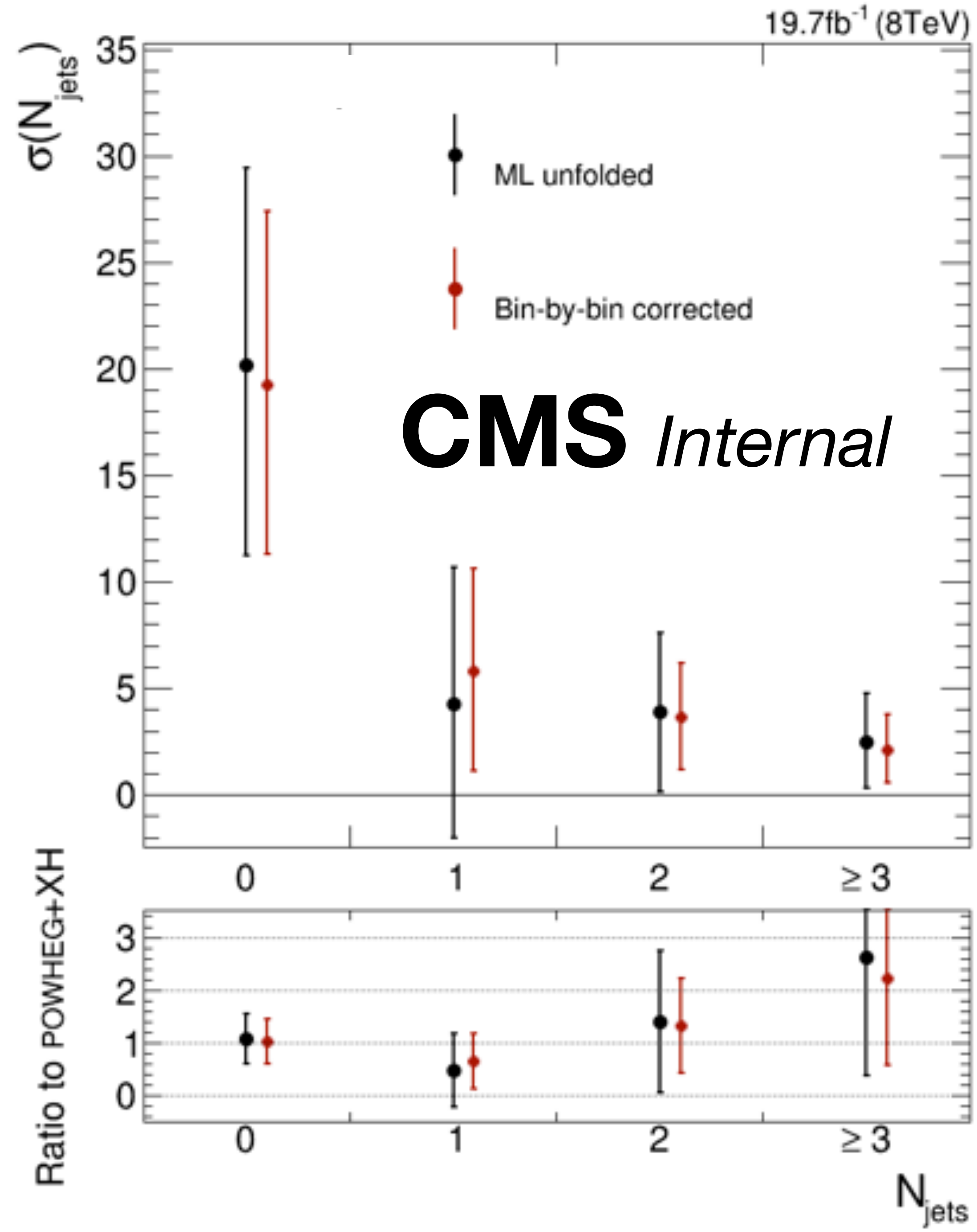
Statistical error on green is $\Delta\sigma/\sigma$ = 16 / 30 = **53%**

Perfect detector: $\Delta\sigma/\sigma$ = $\sqrt{30}/30$ = **18%**

Bin-by-bin: f = 30/256 σ = f * 256 $\Delta\sigma/\sigma$ = 16/256 = **6.25%** **WRONG**

Importance of Unfolding III

- Bin-by-Bin is a biased estimation (smaller uncertainties).
  - Also in **real life**
- **Out-of-acceptance:**
  - A out-of-acceptance shape should be **subtracted** from the fiducial results
- **Bin Migration** can be important:
  - change the best fit values
  - change the confidence intervals!
- p_T differences in the statistical uncertainties are small (up to few percent)
- N_{jets} differences in the statistical uncertainties can be big (up to 30%)
  - jet resolution induces important migrations
- data can pull the best-fit values in the different bins



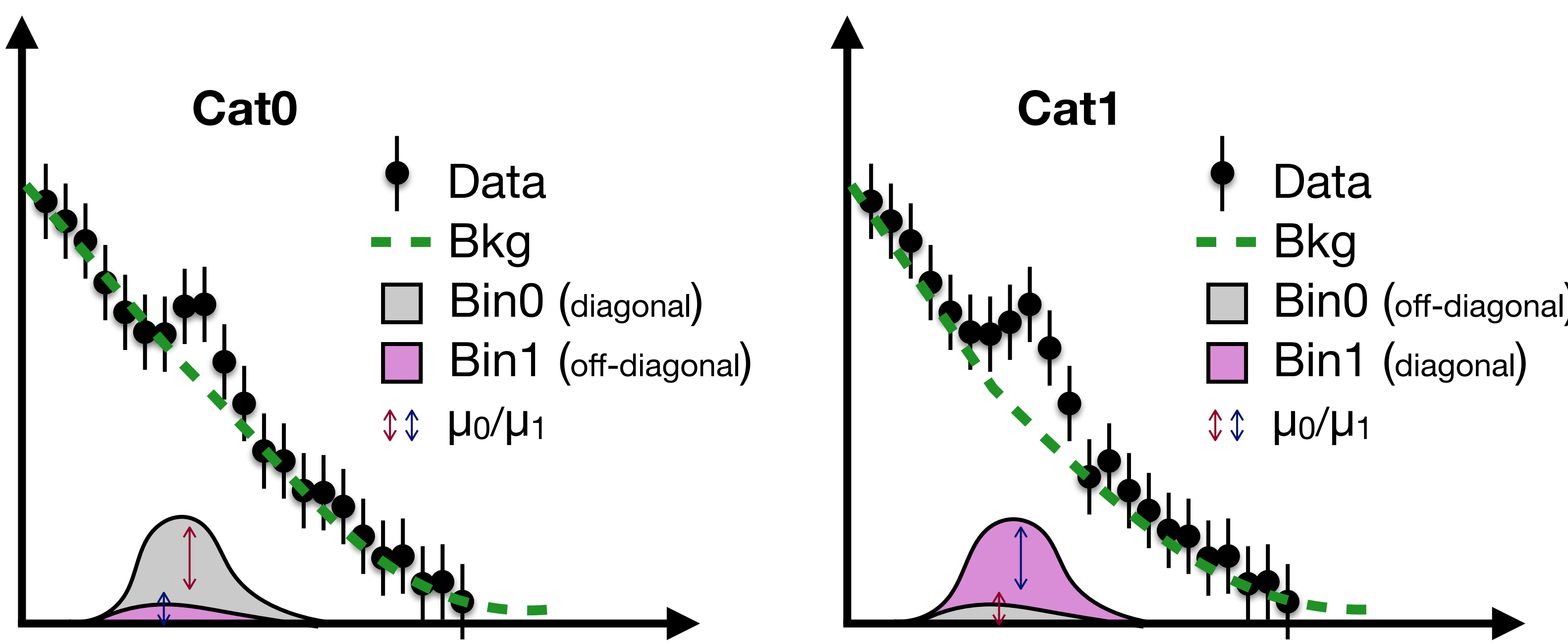
Proposed method

- Method we are and plat to use is base on the ML estimator
  - takes into account: asymm errors, small stat, background functions, nuisances, ..
  - can include regularization

Same method used for μ production channel (and not $\mu_{dijetCat} * f_{VBF}$)

Same method will be used for pseudo cross-sections

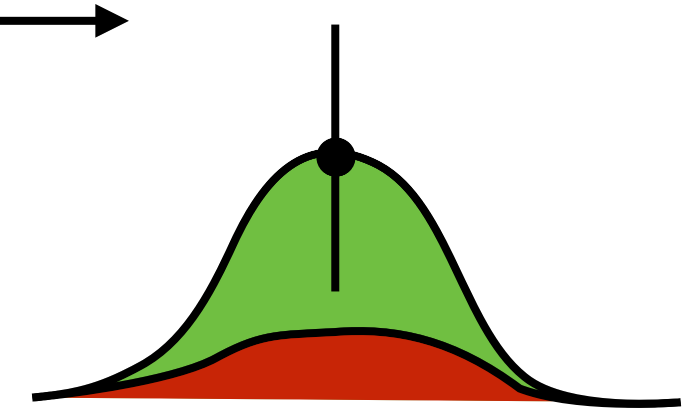
$$\mathscr{F} = -2\log\mathscr{L}(\mathbf{A}\vec{\mu}|\vec{y}) + \delta||\mathbf{L}\vec{\mu}||^2$$



fit simultaneously in cat0/cat1 to get the Bin strength modifiers $\mu=(\mu_0,\mu_1)$

Out of acceptance is subtracted:
- fixing it to MC
- or fixing it to the total xSec

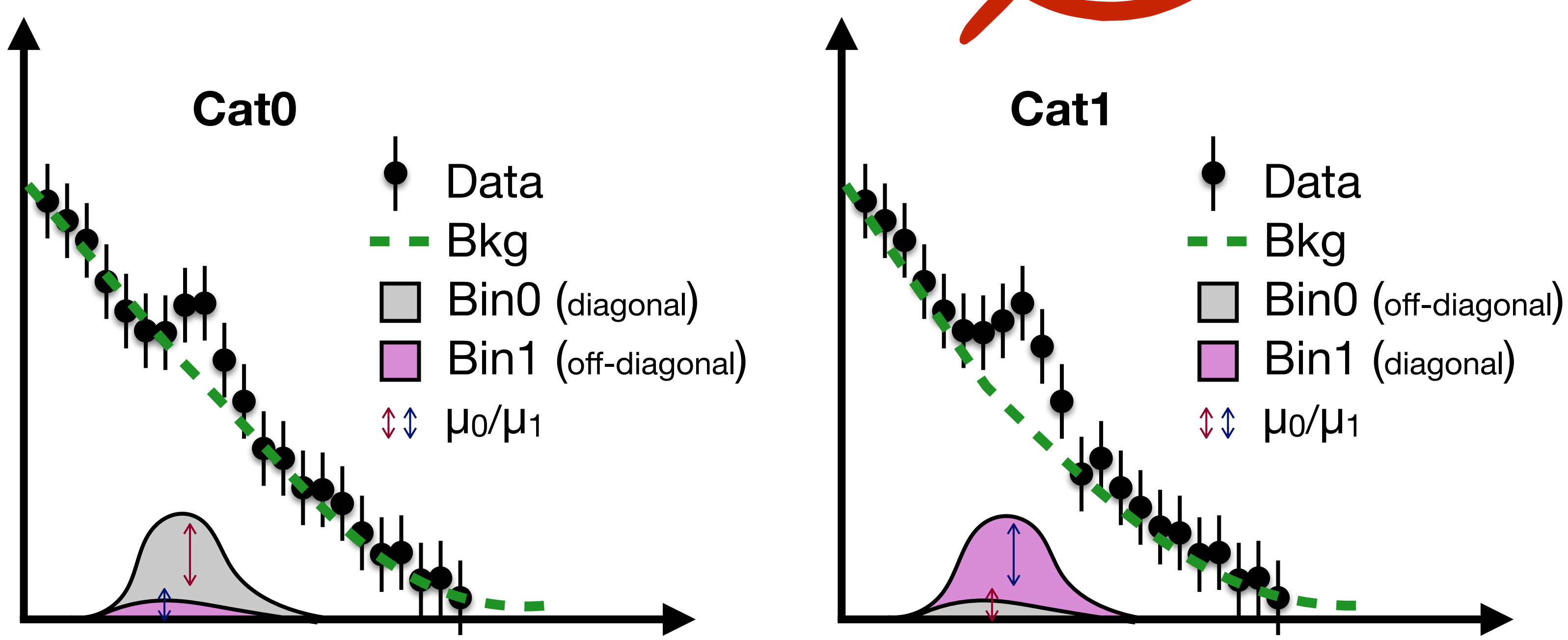
Floating it **coherently** with the signal, reduce the signal error (slide 2)



Regularization

- In Run I we didn't applied it
- In Run II we should think if we should do
- Can be applied a posteriori with the covariance matrix (next slide)

$$\mathscr{F} = -2\log\mathscr{L}(\mathbf{A}\vec{\mu}|\vec{y}) + \delta\|\mathbf{L}\vec{\mu}\|^2$$



fit simultaneously in cat0/cat1 to get the Bin strength modifiers $\mu=(\mu_0,\mu_1)$

Post extraction regularization

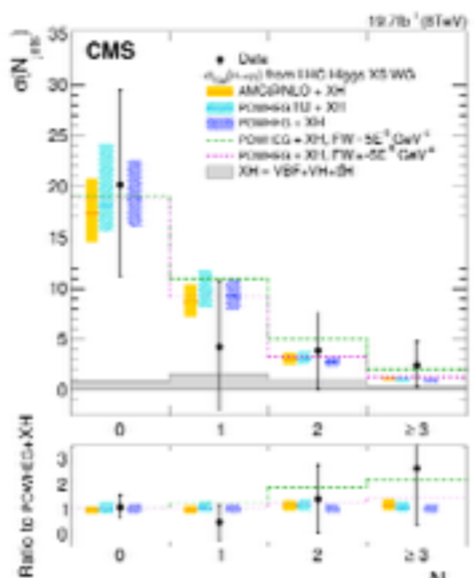
- Example of Tikhonov regularization
  - using the published data points
  - and the **covariance matrix**

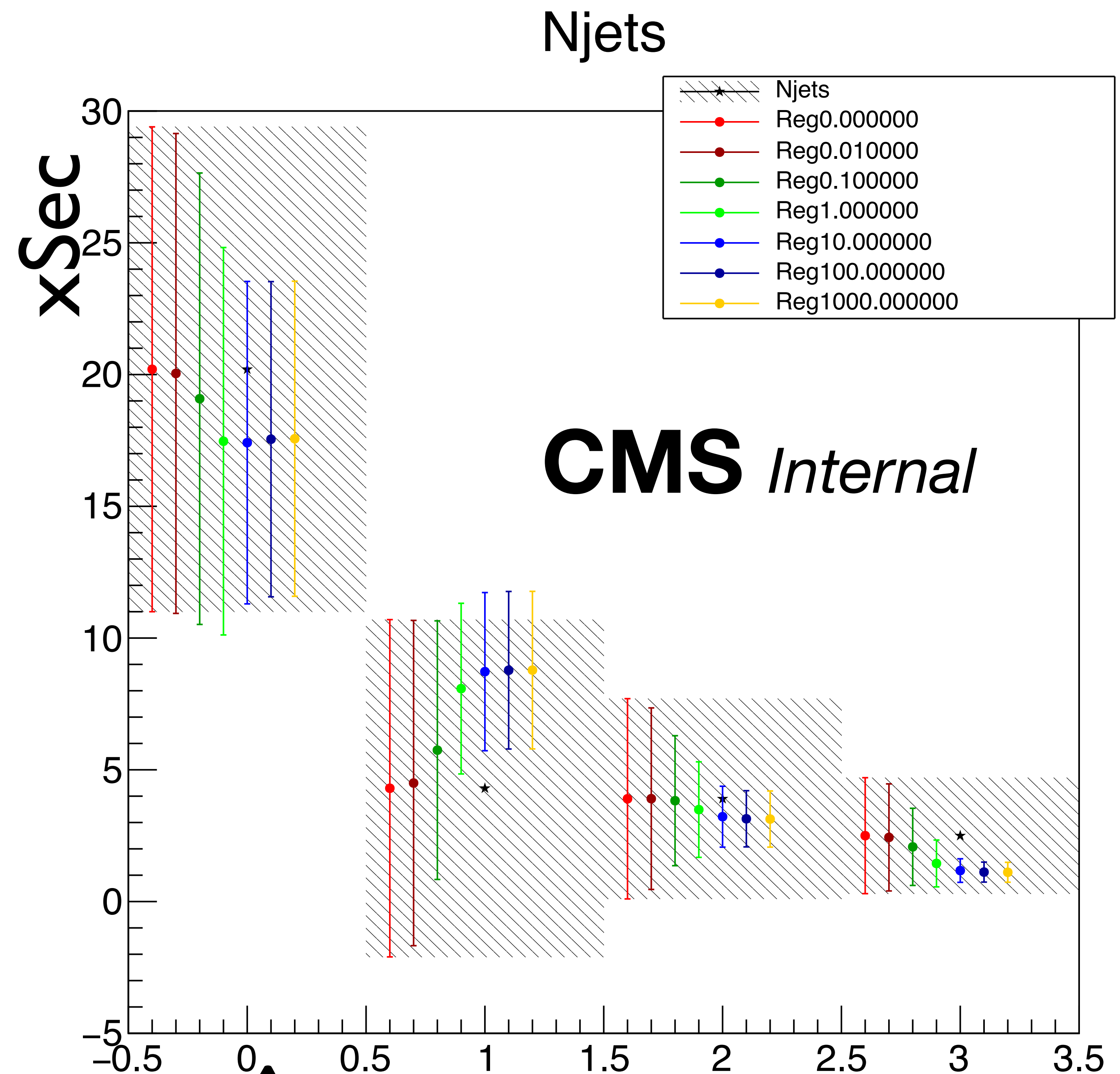
$$\mathcal{F} = \chi^2 + \delta\|\mathbf{L}\cdot\mu\|^2$$

- Effect of regularization are:
  - bias (towards MC)
    (kernel of the regularization operator)
  - reduce of “large” variance in the distributions

- Study of the regularization parameter, bias … is needed

- Done **a posteriori** assuming gaussian errors (with correlation)



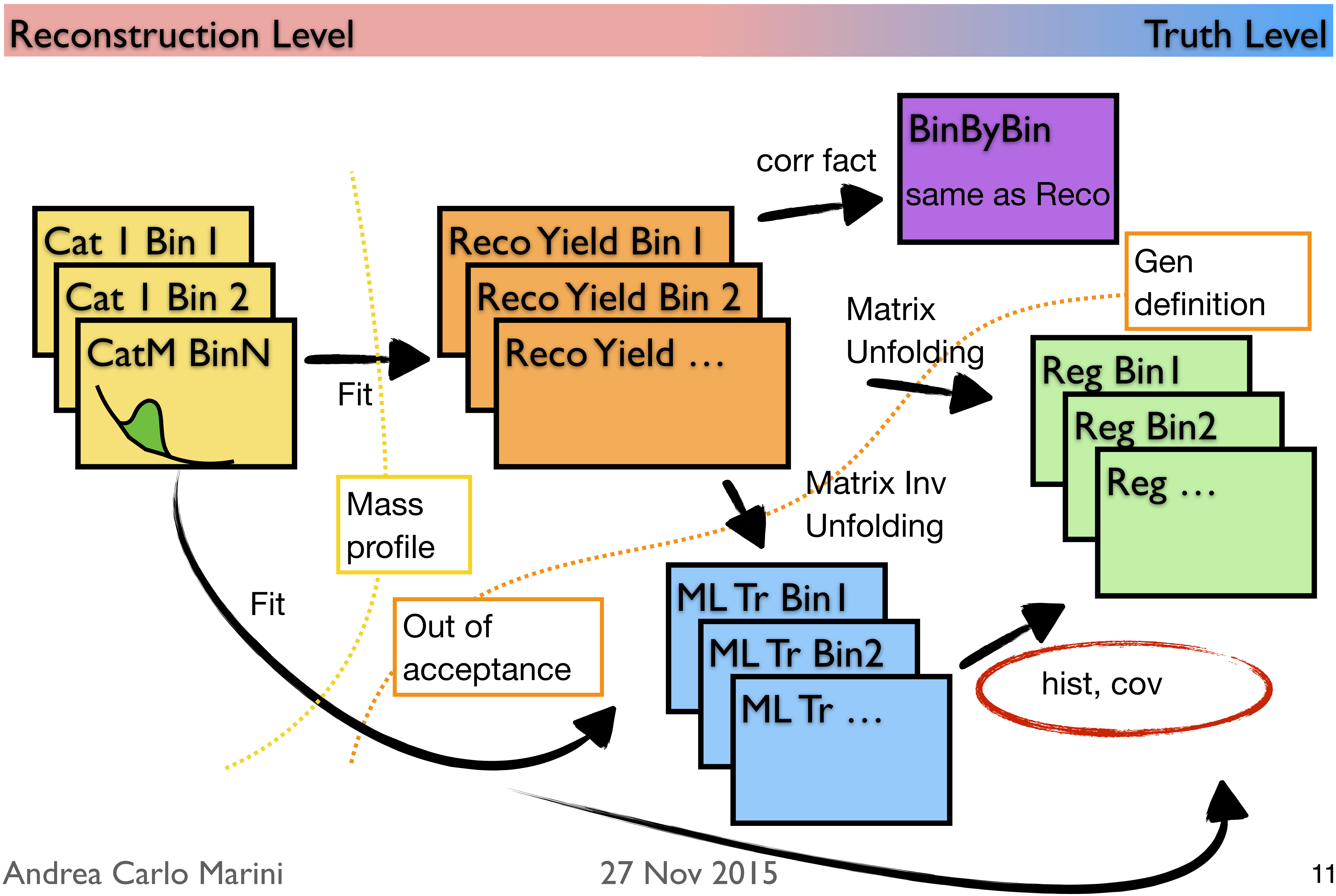


Summary & Conclusions

- Bin-by-Bin correction provides wrong statistical error estimation
  - these can be easily wrong of 20 – 30%
- ML provides a way to construct estimators
  - take into account error propagation
  - include nuisances, systematics, categories …
- Signal Extracted detector yields can be unfolded using other standards techniques (e.g. RooUnfold)
- We use already this technology for the couplings
  - same arguments holds
- Regularization can be added in the likelihood or a posteriori

Backup

Possible paths



Adding regularization

- Adding Tickhonov regularization to the likelihood

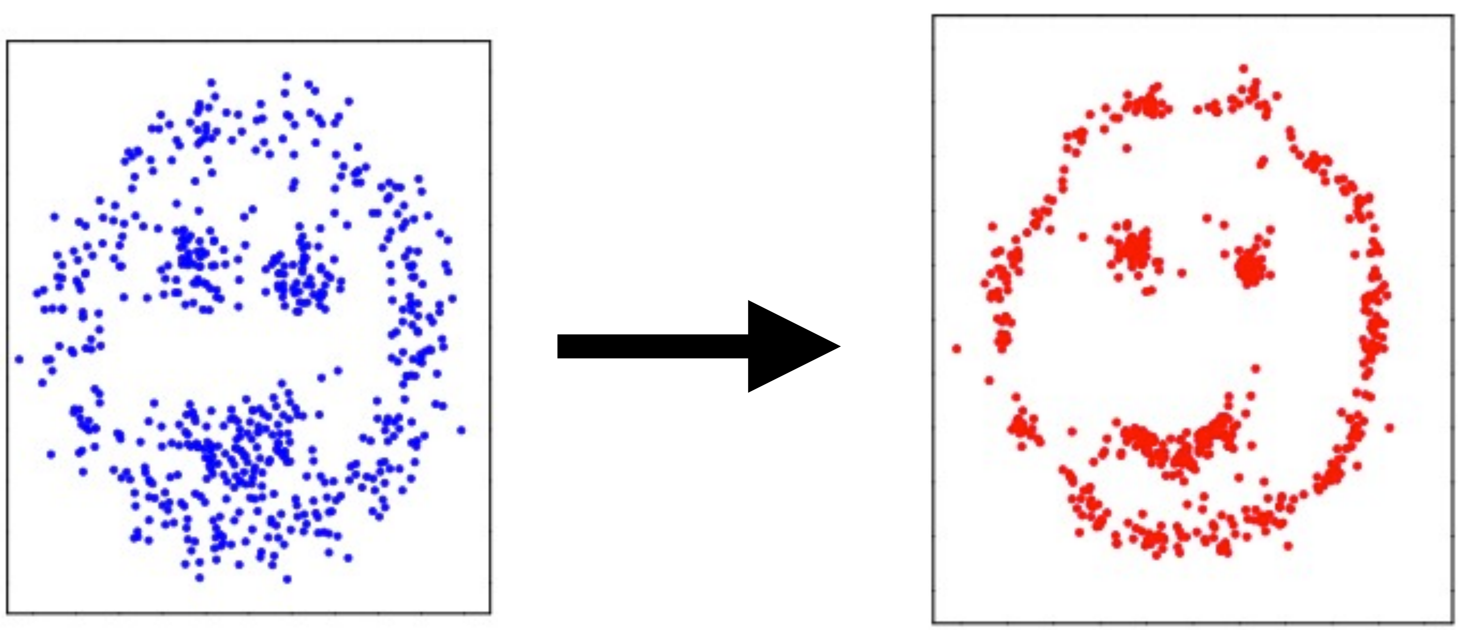
$$\mathscr{F} = -2\log\mathscr{L}(\mathbf{A}\vec{\mu}|\vec{y}) + \delta||\mathbf{L}\vec{\mu}||^2$$

A certain number of choices (L, delta) …

- it’s not trivial to keep under control these parameters with the current statistics.

The goal of the regularization is to give a not distorted spectrum

- use the additional fact that distributions are continuous



Categories, Signal and Literature

- **Categories** (SVD):
  - SVD can be extended with categories

$$\vec{y}_{\text{reg}} = \underline{0} \qquad \mathbf{B} = \left(\hat{\mathbf{A}}^{\mathrm{T}}\Sigma^{-1}\hat{\mathbf{A}}\right)^{+}\hat{\mathbf{A}}^{\mathrm{T}}\Sigma^{-1}$$

$$\hat{\mathbf{A}}_{\text{reg}} = \sqrt{\delta}\mathbf{L} \qquad \vec{x}_T = \mathbf{B}\vec{y}$$

$$\Delta\vec{y}_{\text{reg}} = \underline{1} \qquad \Sigma' = \mathbf{B}\Sigma\mathbf{B}^{\mathrm{T}}$$

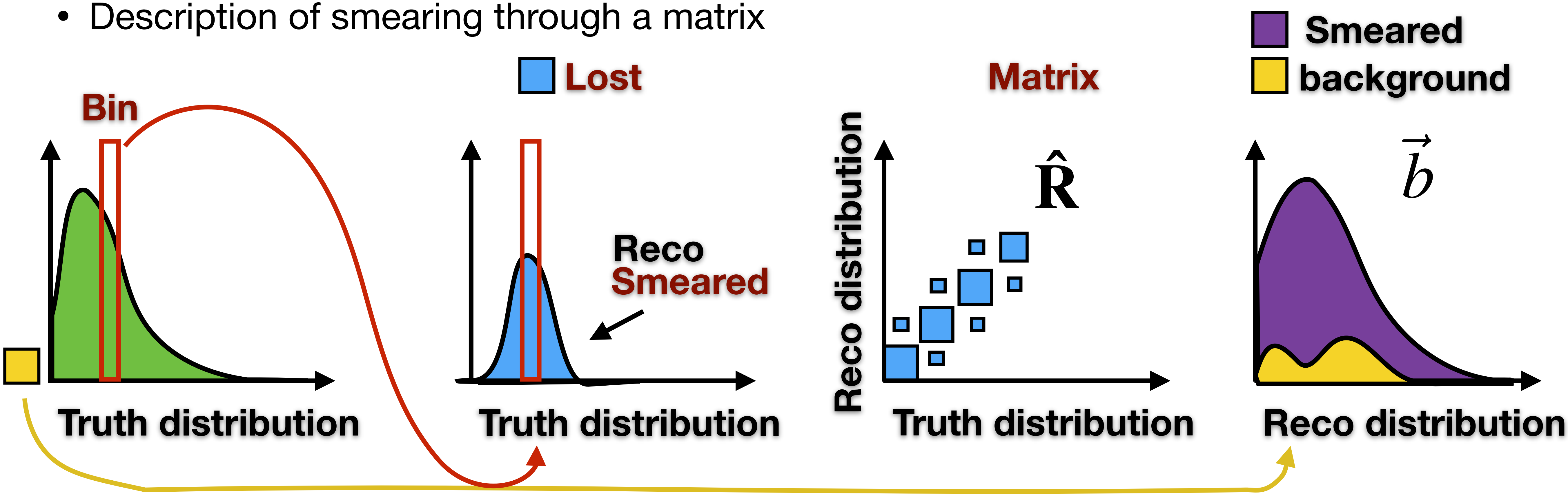
  but signal extraction must be performed before.
  - Bayes:
    - cannot use the "built-in" categories due to the very non-poissonian errors of the mgg continuum:
    - Each category should be unfolded separately and results re-combined later

Signal Extraction:

- These methods wants that signal extraction is performed before
- Systematics and nuisances (eg, m_H) will be just approximations
- Covariance matrix approximation for low yields

Unfolding 1

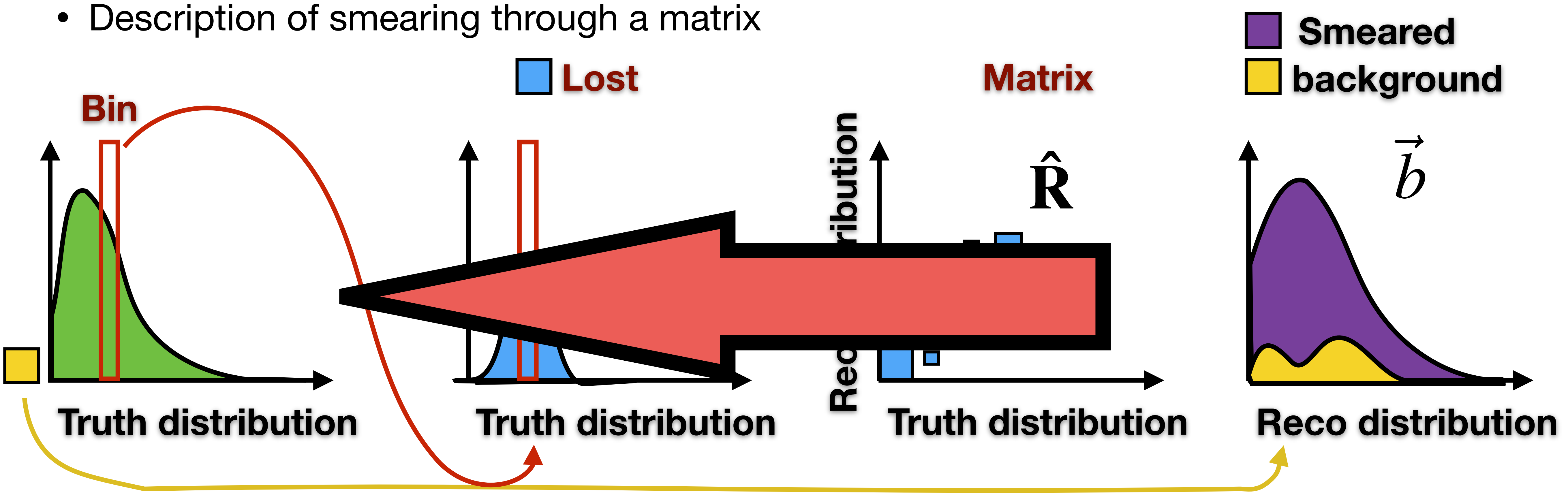
- Undo detector effects
- based on linearity assumption
  - Description of smearing through a matrix



$$x_M^i = \hat{R}^{ij} x_T^j + b^i$$

Unfolding I

- Undo detector effects
- based on linearity assumption
  - Description of smearing through a matrix



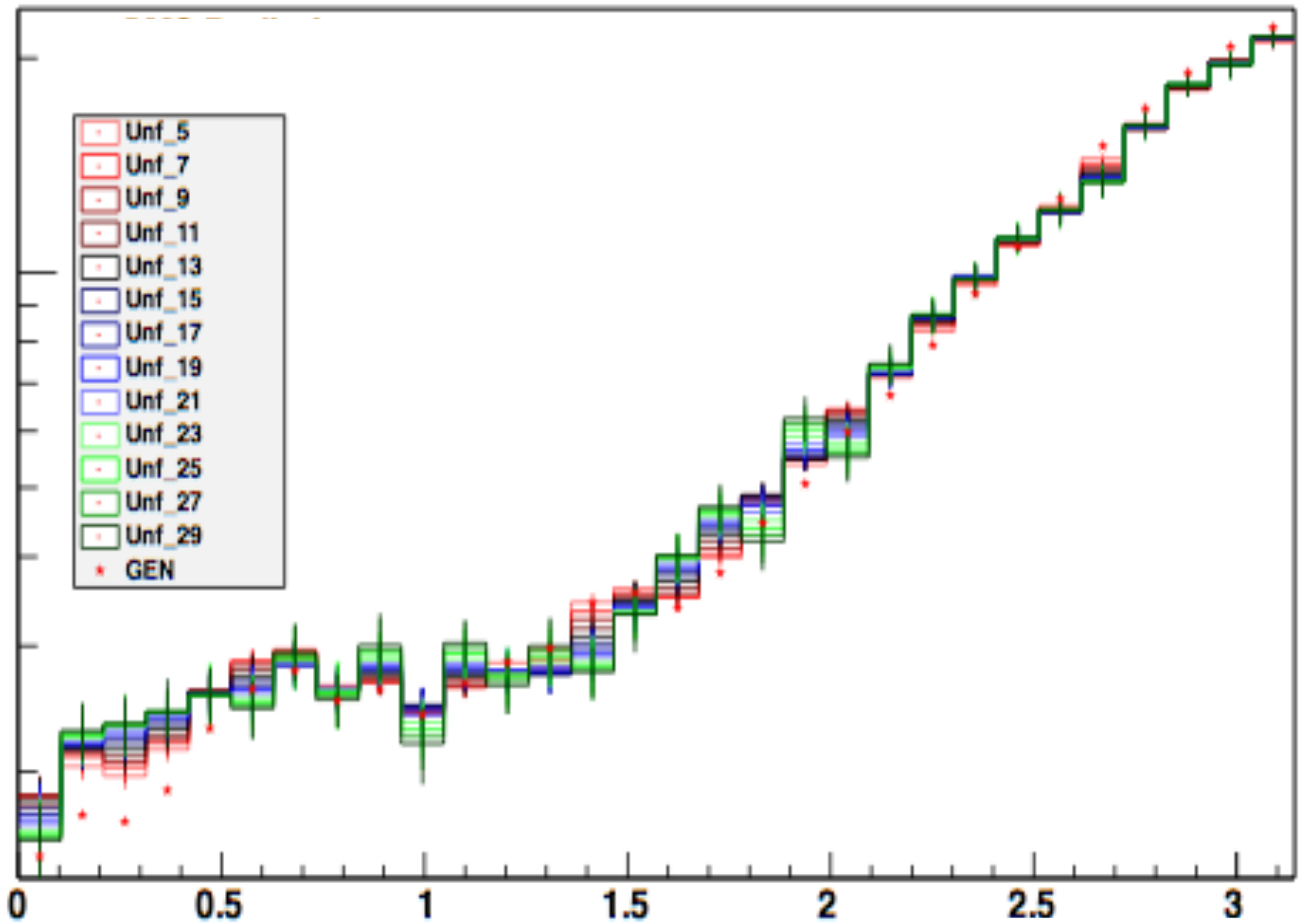
$$x_M^i = \hat{R}^{ij} x_T^j + b^i$$

Regularization & Unfolding

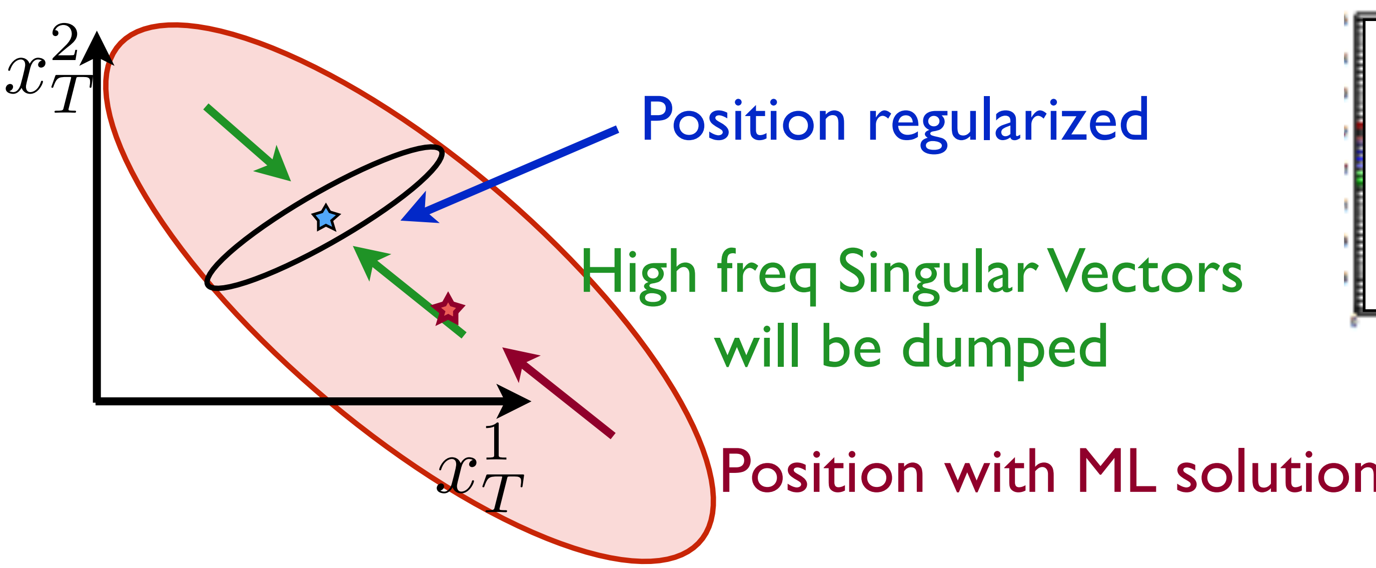
- What is regularization doing ?
- Penalize high fluctuating solutions
  - bias in the "minimum search"

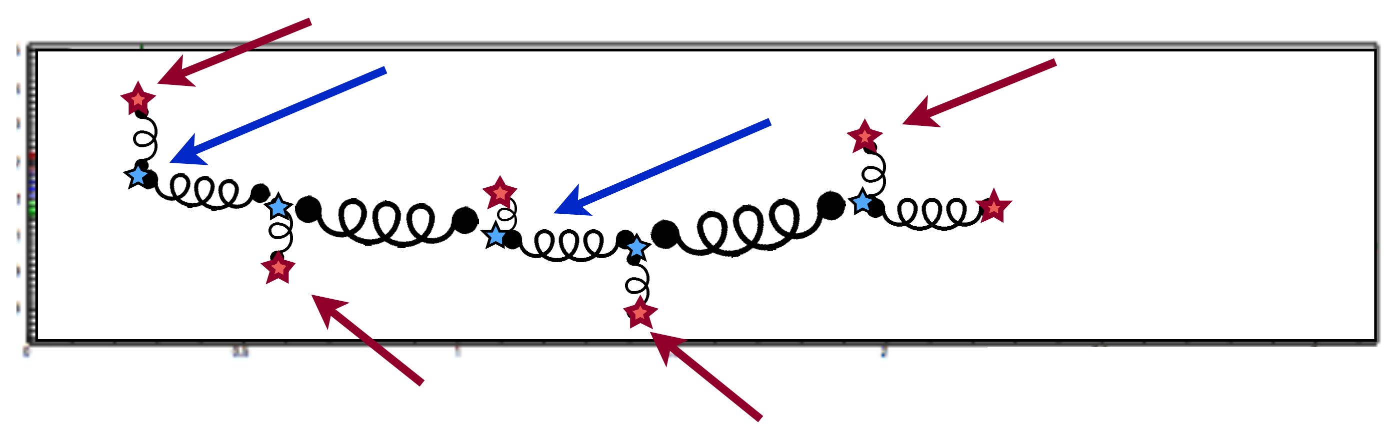
$$\min_{\mu} ||\vec{x}_M - \vec{b} - \mathbf{R} \cdot \vec{\mu}||^2 + \delta ||\mathbf{L} \cdot \vec{\mu}||^2$$

$$\vec{x}_T = \hat{\mathbf{R}}^{-1} (\vec{x}_M - \vec{b})$$



- Reduce variance of the final distribution





- Binning is an other way of "regularize"