Unfolding ATLAS data

IMPRS Colloquium | March 13th 2025
Elia Schmidt

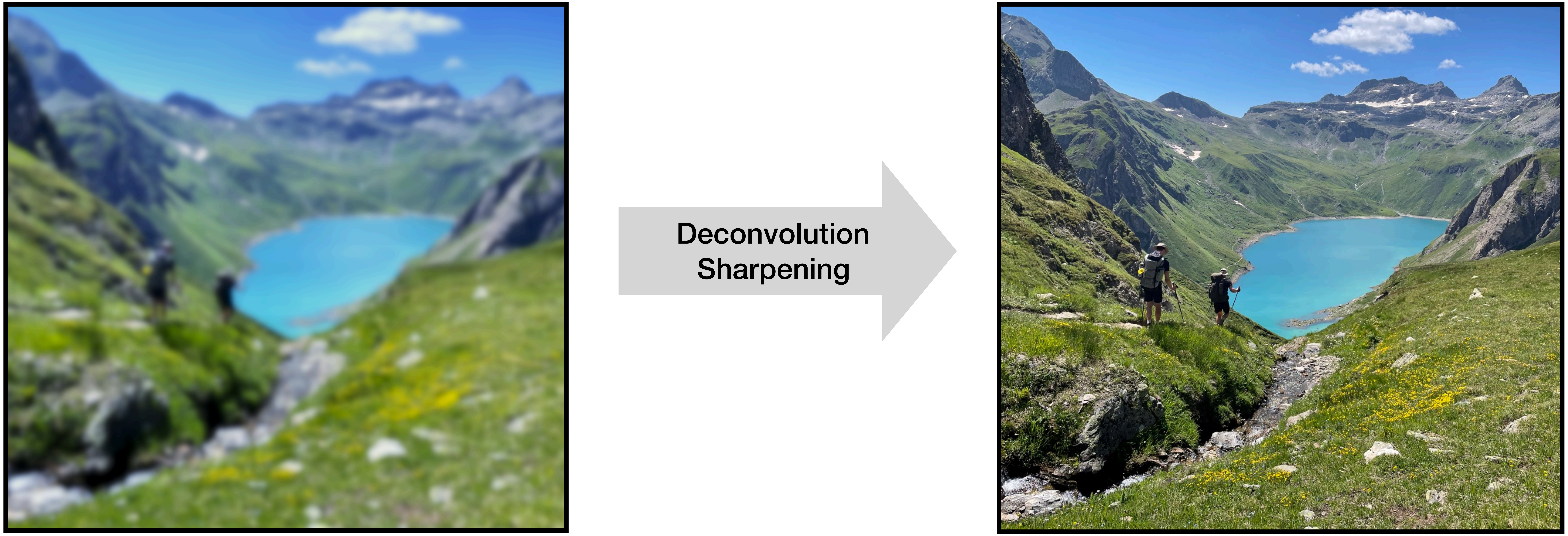
1. What is unfolding?

1. We can never directly access the quantities we want to measure

2. The measurement device necessarily introduces some distortion effects

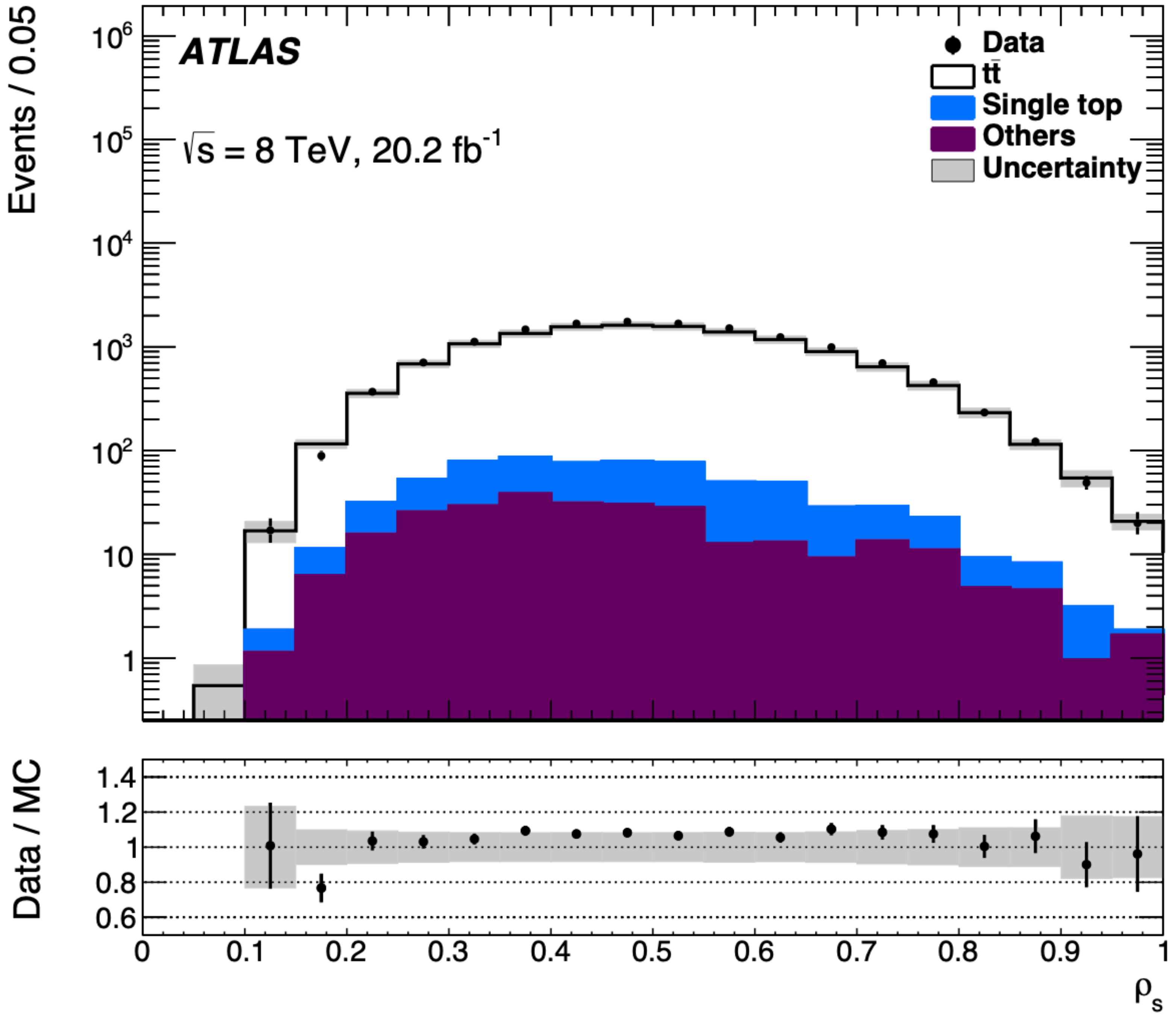
3. The reversal of these distortion effects is **unfolding**

1. What is unfolding?



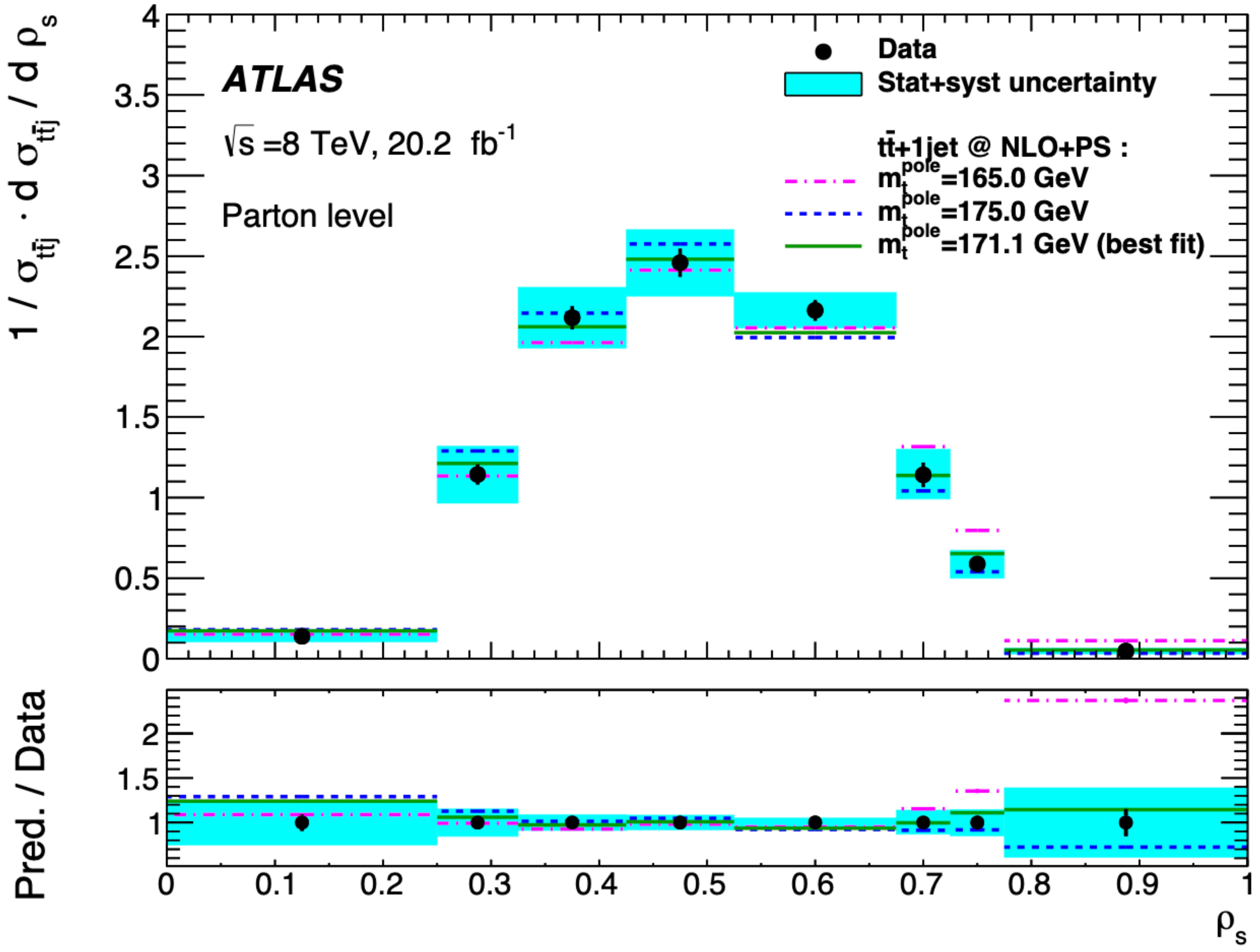
Deconvolution
Sharpening

1. What is unfolding?



Distribution of the ρ_s-variable after the final selection at detector level (top pole mass extraction)

Unfolding

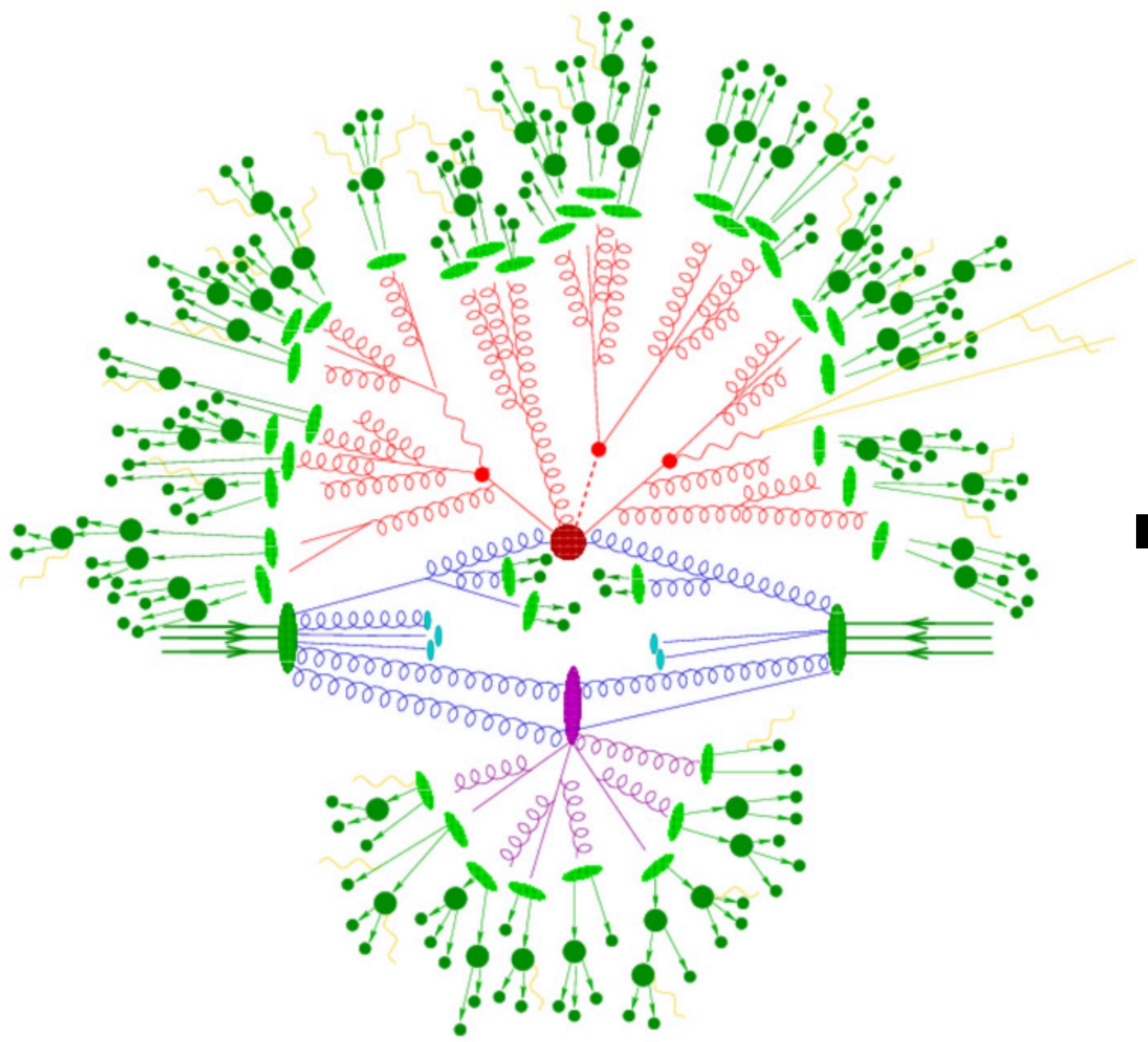


Differential cross section wrt. ρ_s

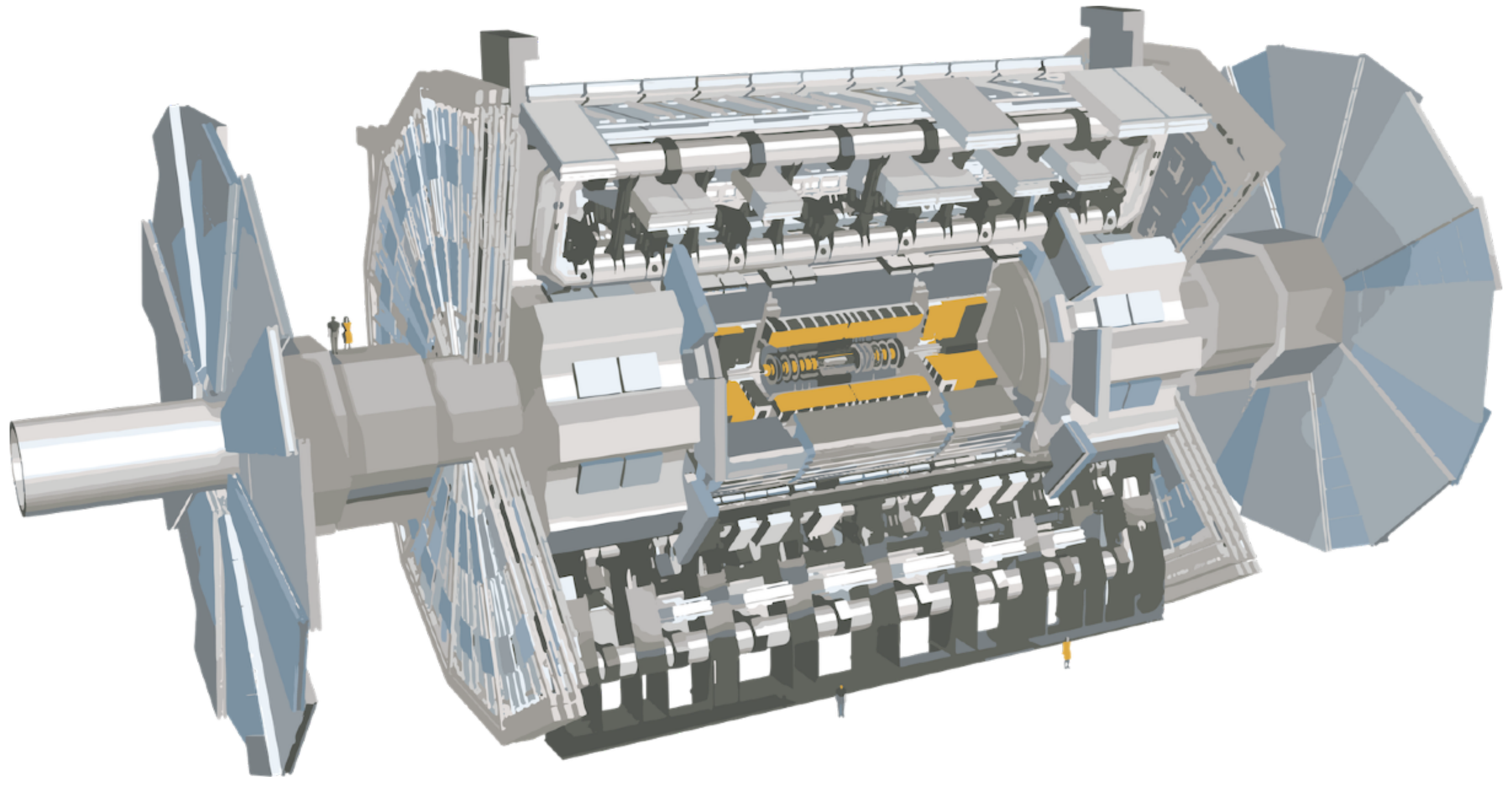
2. When do we need to unfold?

- We always need **Monte Carlo** simulations to validate our results and to estimate our uncertainties
- 3 Components:

Event generation



Showering and Hadronisation

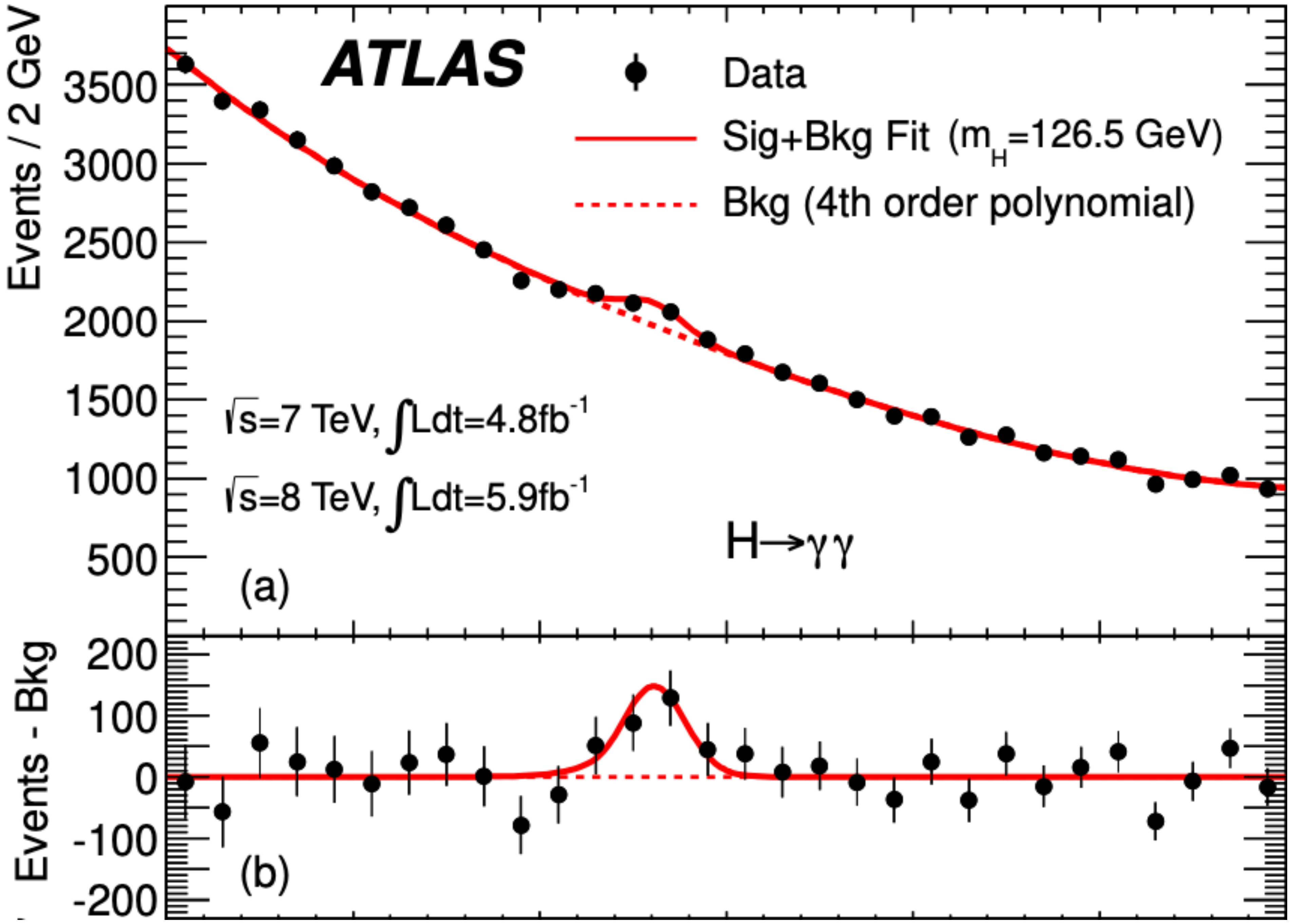


Detector simulation

2. When do we need to unfold?

- Sometimes detector-level quantities are fine
- For example, the Higgs discovery involved no data unfolding
- Unfolding itself introduces uncertainties

- Unfolded results are detector-independent
- Unfolded results can be easily used for reinterpretations using new models
- Some parameters that are well-defined at unfolded level become unclear after showering and hadronisation



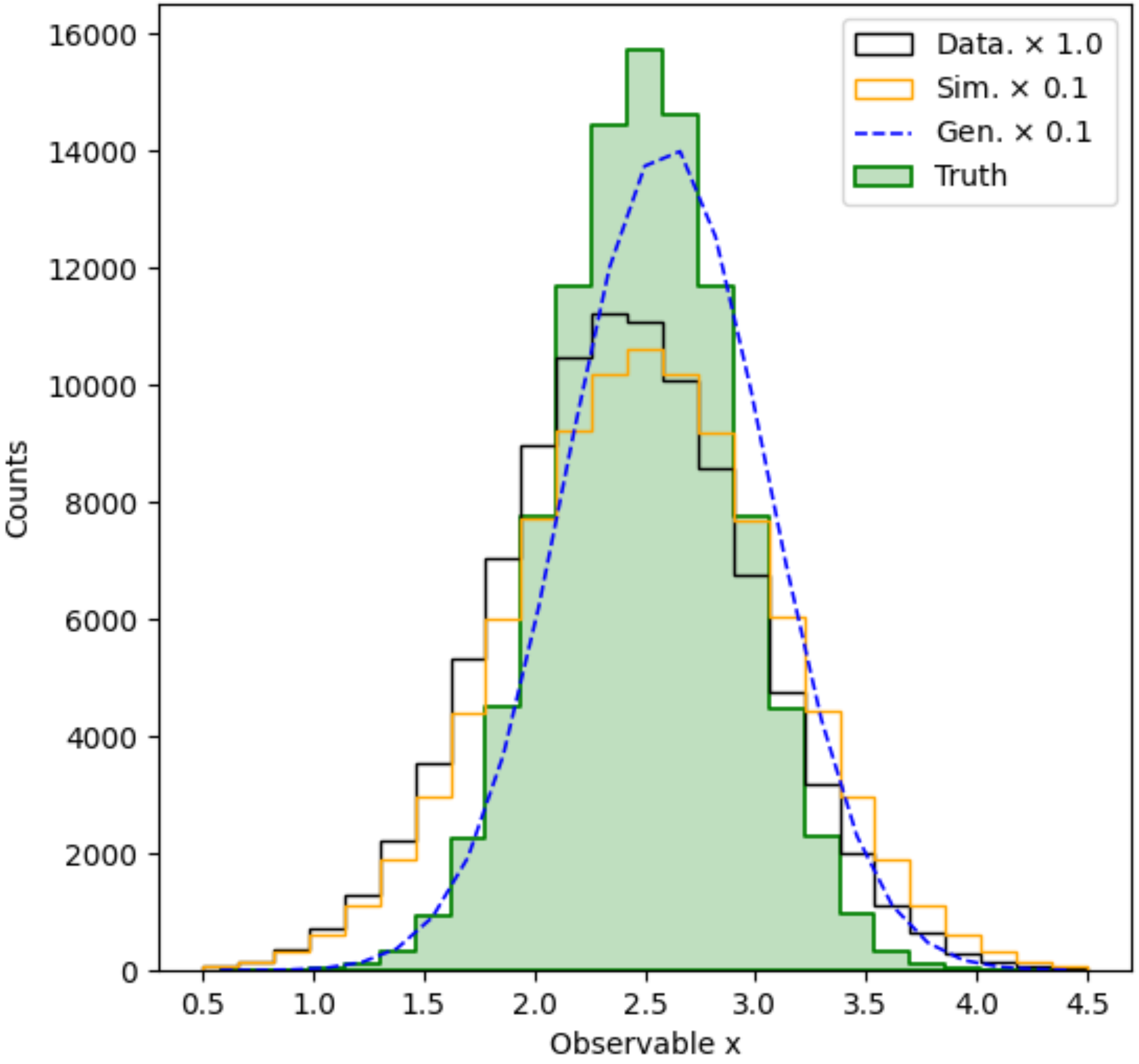
Propaganda plot from the original Higgs discovery paper

3. So what's the big deal?

Unfolding = **Inverse Problem**
(What is the *cause* behind a measured *effect?*)

- Naive idea: Apply **bin-by-bin correction factors**

3. So what's the big deal?

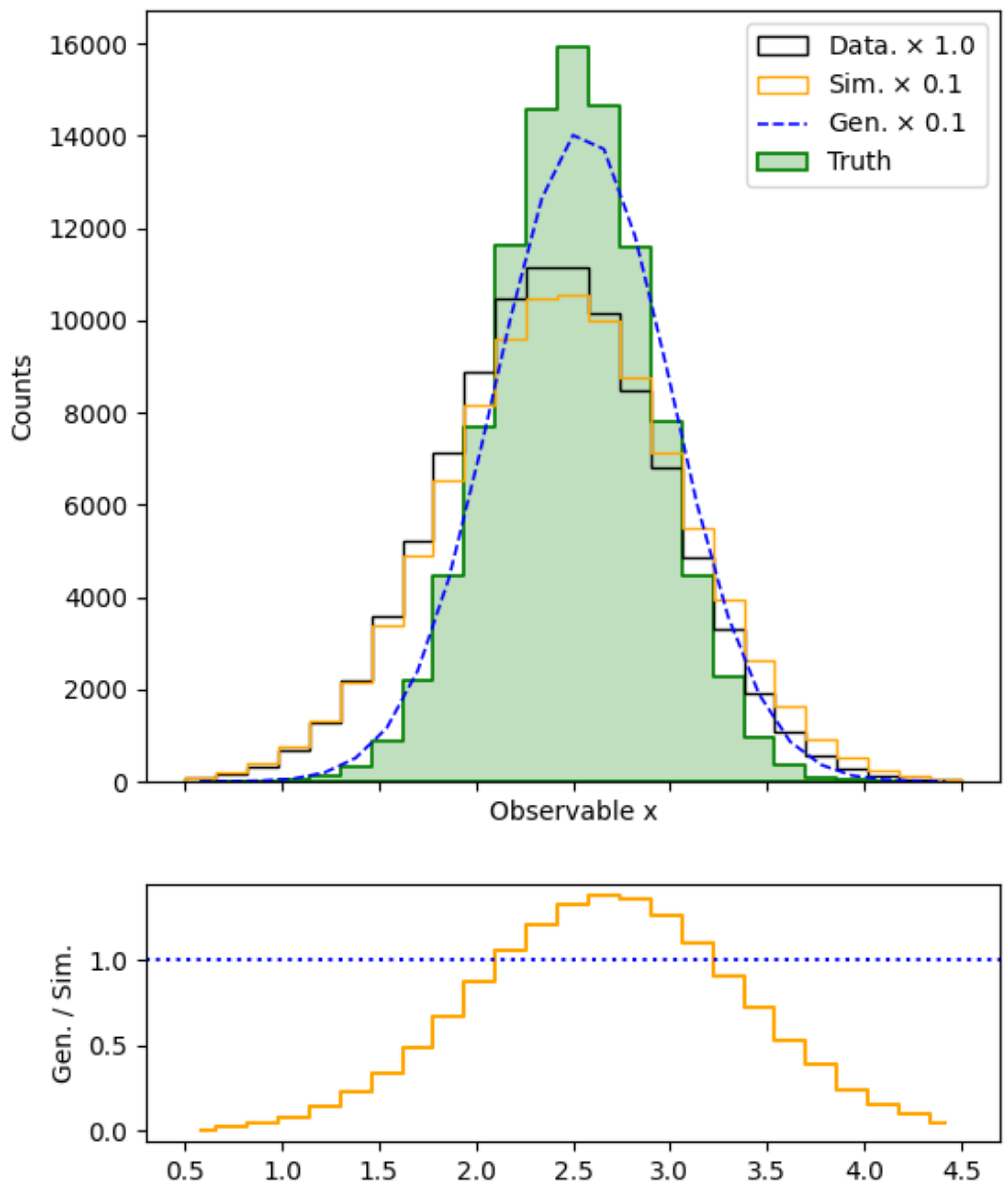


- Toy study:
  - **Truth**: Where we want to get to
  - **Data**: What we measure
  - **Gen**: Monte-Carlo generation
  - **Sim**: MC after detector simulation
- Goal:

Data → **Truth**

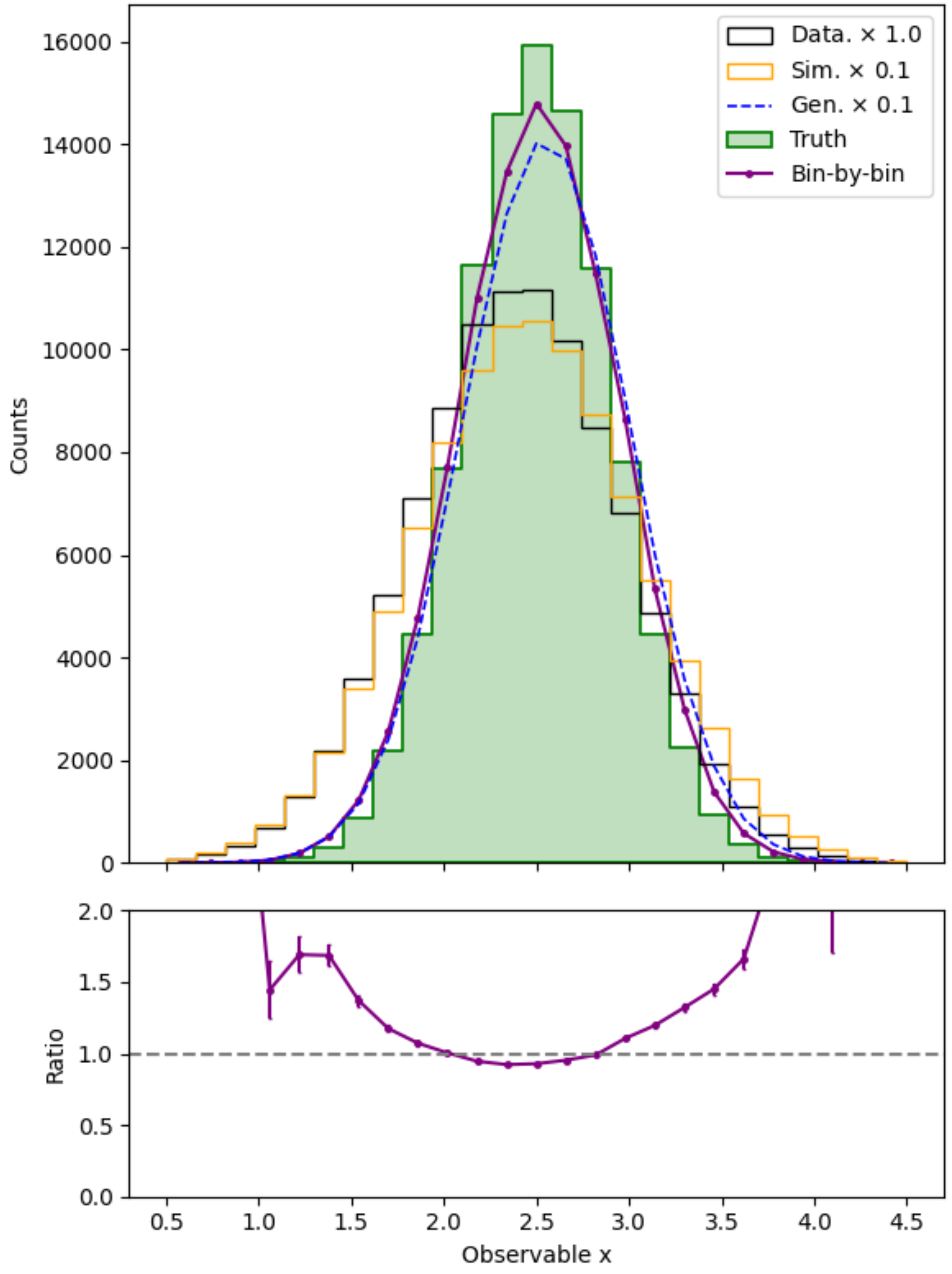
3. So what's the big deal?

- Naive idea: Apply **bin-by-bin correction factors**
  - Calculate bin-by-bin correction factors in simulated sample



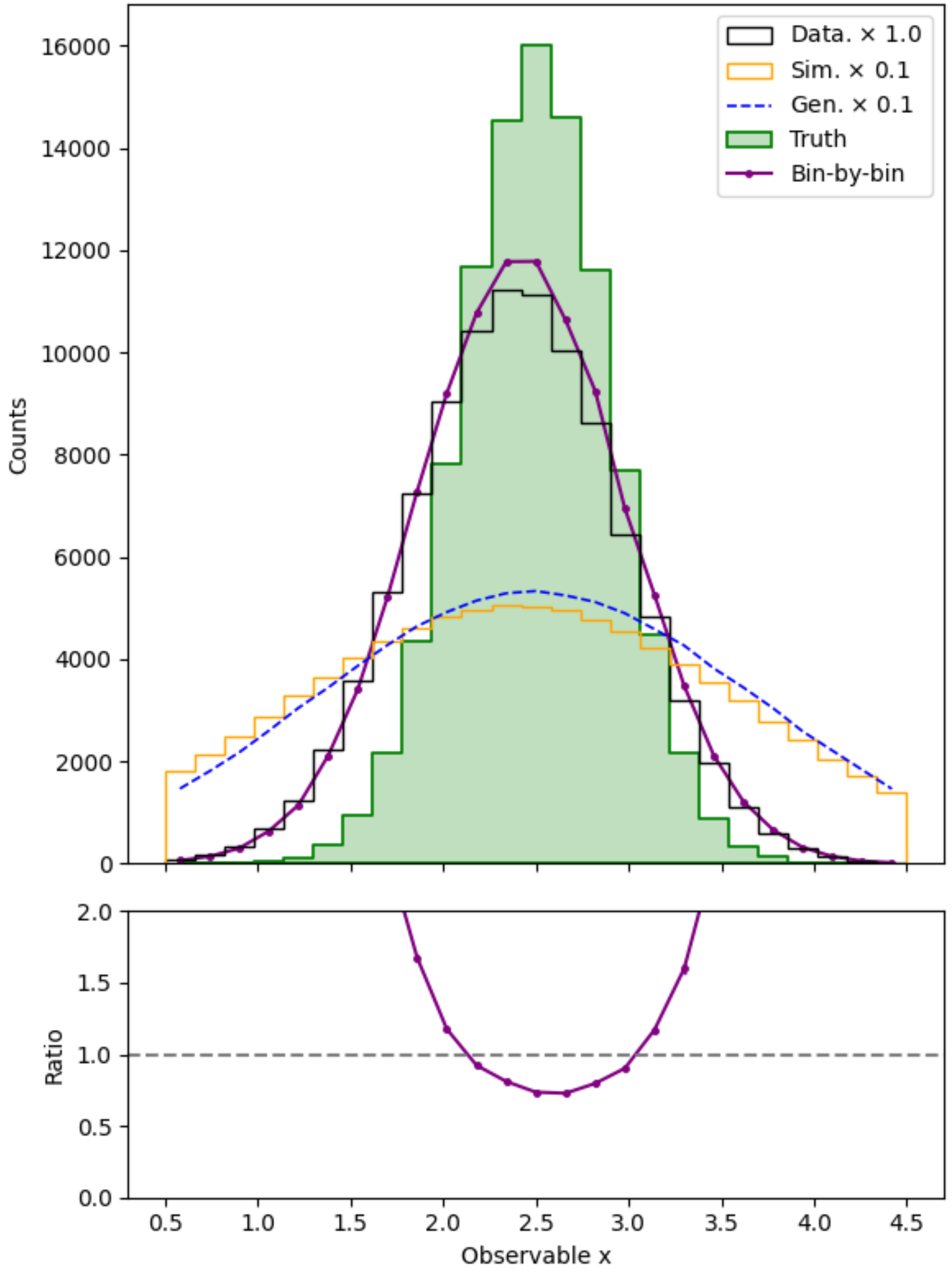
3. So what's the big deal?

- Naive idea: Apply **bin-by-bin correction factors**
  - Calculate bin-by-bin correction factors in simulated sample
  - Apply correction factors to the measured data



3. So what’s the big deal?

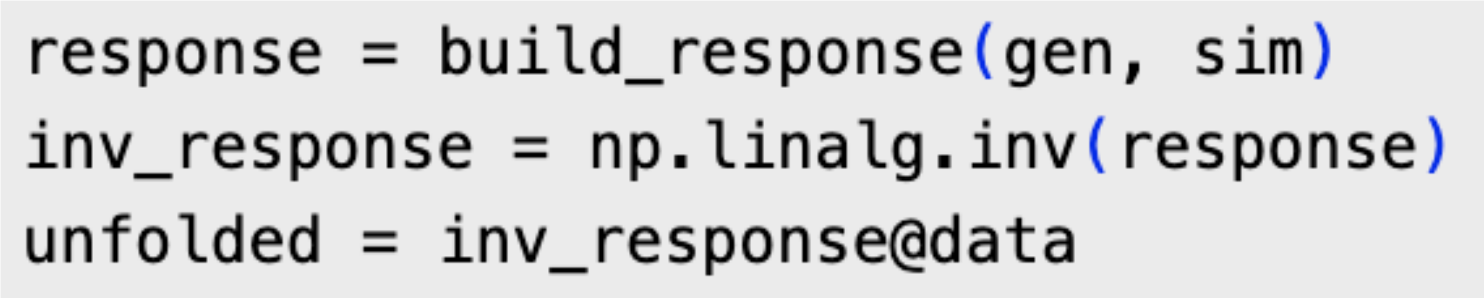
- Naive idea: Apply **bin-by-bin correction factors**
  - Calculate bin-by-bin correction factors in simulated sample
  - Apply correction factors to the measured data
  - Works only if the simulation is very close to reality
  - Works only if the detector smearing is slight
  - Works only if truth- and detector-level have same bins
  - No model-agnostic unfolding possible



3. So what's the big deal?

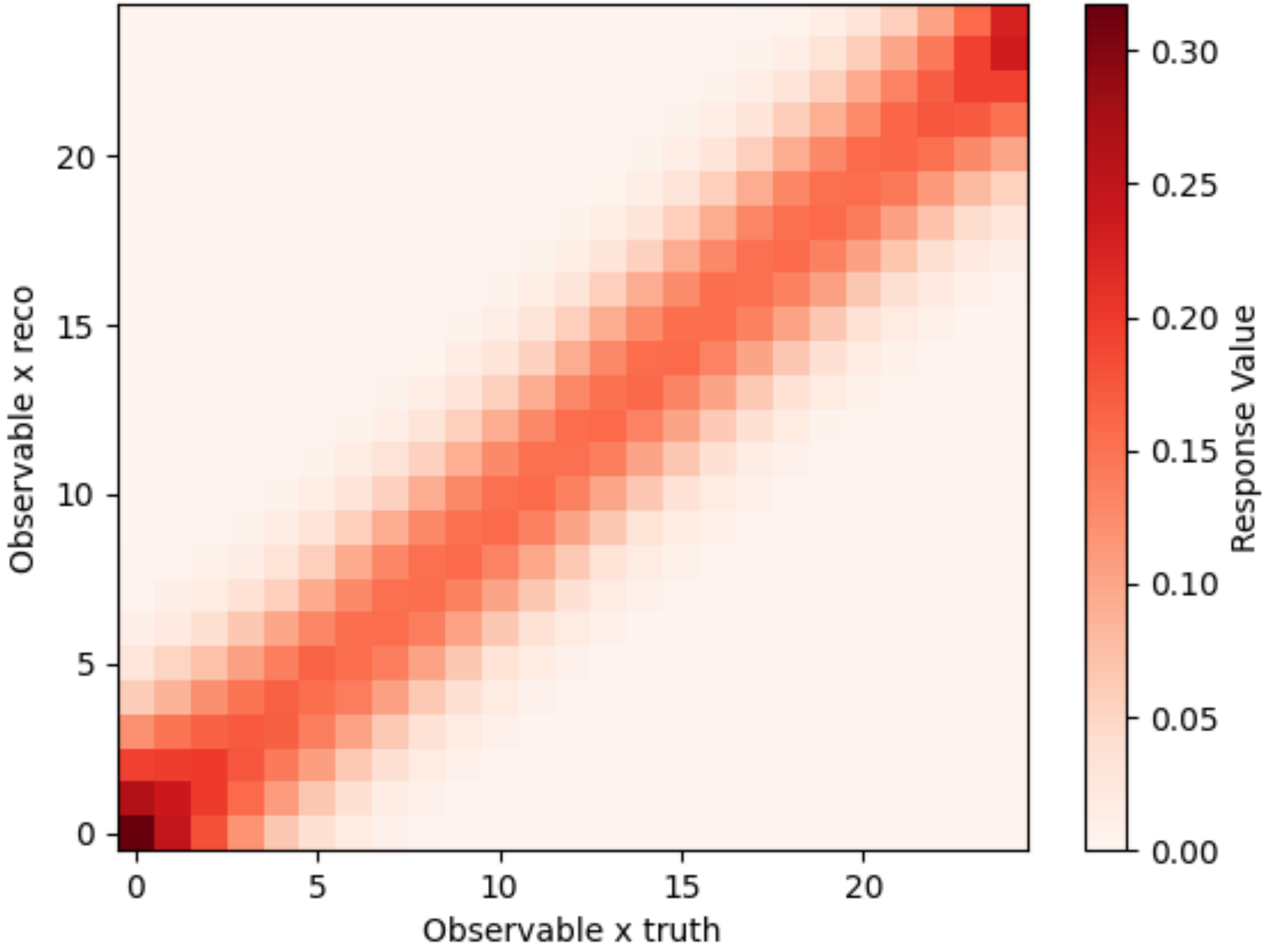
- **Matrix-based unfolding**
- Detector response: $\sum_i A_{ij} t_i = m_j$
- Solve $(At - m)^T (At - m) = \min$
- Slightly less naive but still naive idea: **Numerical matrix inversion**

$$\sum_i A_{ij}^{-1} m_i = t_j$$



```
response = build_response(gen, sim)
inv_response = np.linalg.inv(response)
unfolded = inv_response@data
```

- Works regardless of the binning scheme



Detector response matrix

3. So what's the big deal?

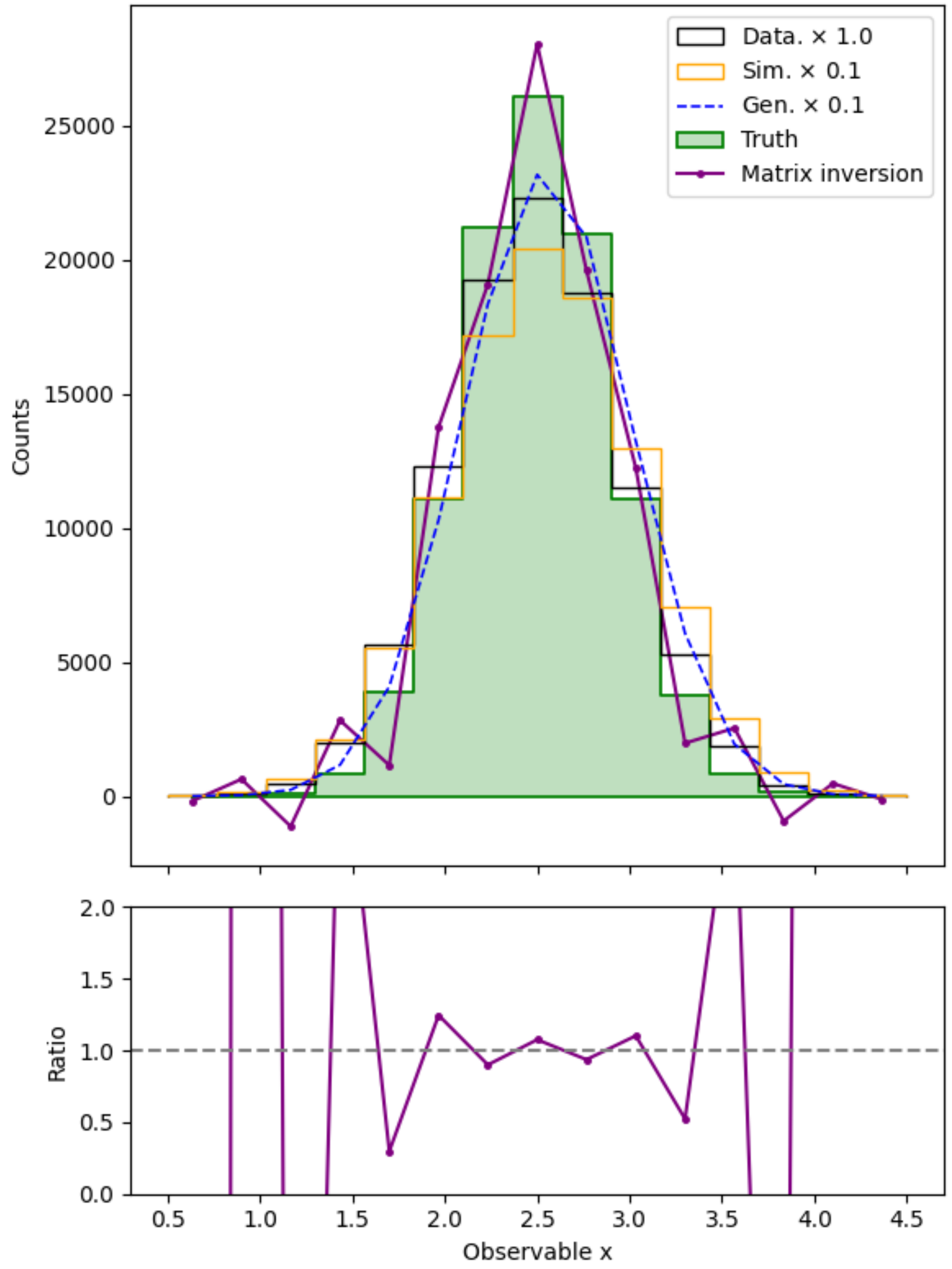
- Slightly less naive but still naive idea:
  Numerical matrix inversion

$$\sum_i A_{ij}^{-1} m_i = t_j$$

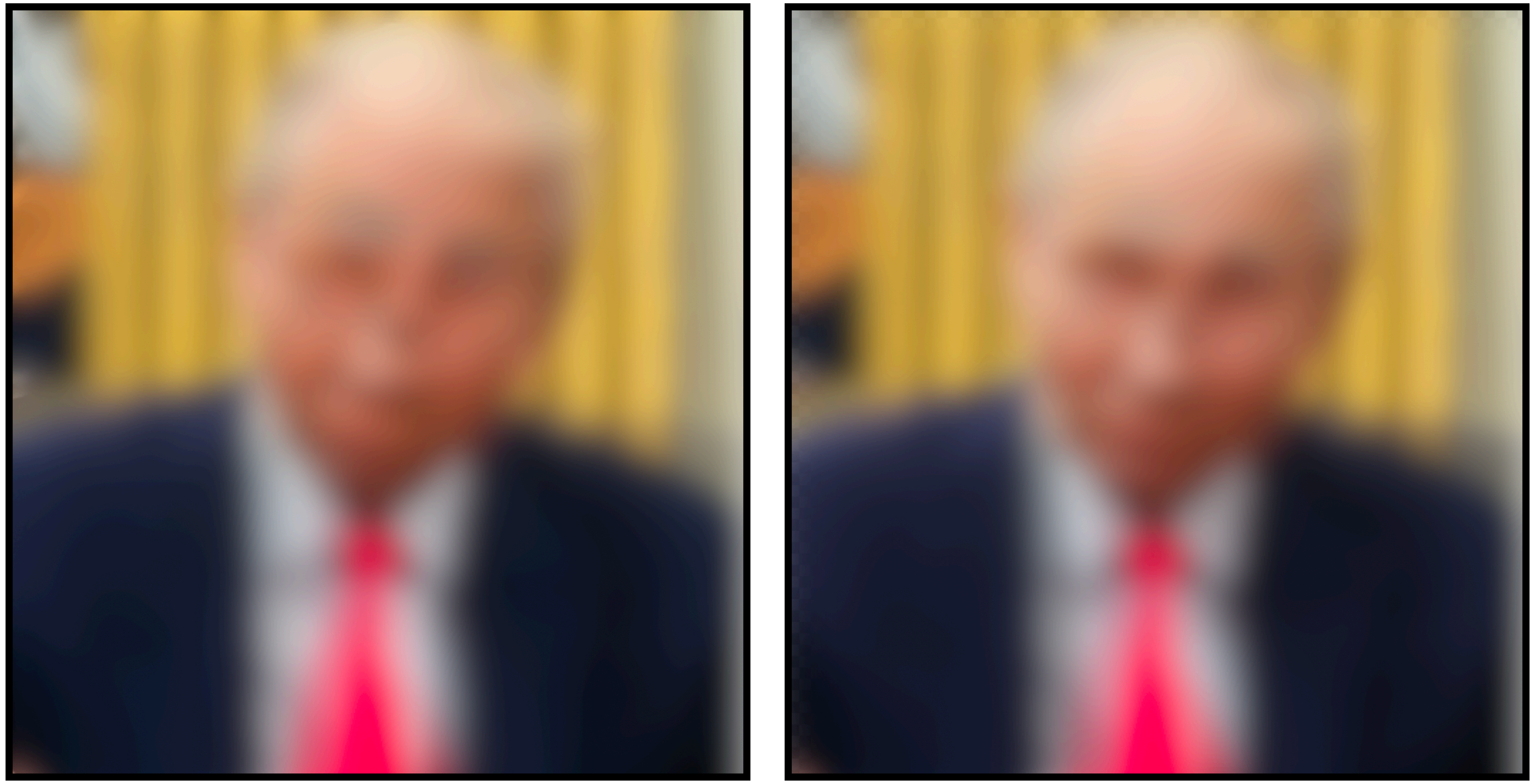
- Numerical instabilities in matrix inversions
- Statistical fluctuations blow up

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & \epsilon \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\epsilon^{-1} \\ 0 & 0 & \epsilon^{-1} \end{pmatrix}$$

- **Need to disentangle real physics from stat. fluctuations**



3. So what’s the big deal?



Sometimes similar images…

3. So what’s the big deal?

…have different truths

4. Solution: Singular Value Decomposition

- Singular Value Decomposition (**SVD**) of the response matrix $A = U\Sigma V^T \quad \Sigma = \mathrm{diag}(\sigma_i)$
  - U: orthogonal matrix
  - Σ: diagonal matrix containing singular values (**in descending order**)
  - V^T: orthogonal matrix
- Matrix inversion:

$$A^{-1} = V\Sigma^{-1}U^T \quad \Sigma^{-1} = \mathrm{diag}(\sigma_i^{-1})$$

- Example from earlier:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & \epsilon \end{pmatrix} = U \cdot \mathrm{diag}\left(\sqrt{2},\ 1,\ \epsilon/\sqrt{2}\right) \cdot V^T$$

4. Solution: Singular Value Decomposition

- **Tichonov Regularisation**
- Constrain solution by using some a priori knowledge about the solution
- Introduce a term proportional to the curvature of the result
- Curvature:

$$\sum_i \left[(t_{i+1} - t_i) - (t_i - t_{i-1})\right]^2 = (Ct)^T Ct$$

$$C = \begin{pmatrix} -1 & 1 & 0 & 0 \dots \\ 1 & -2 & 1 & 0 \dots \\ 0 & 1 & -2 & 1 \dots \\ 0 & 0 & 1 & -2 \dots \\ & & \dots & \end{pmatrix}$$

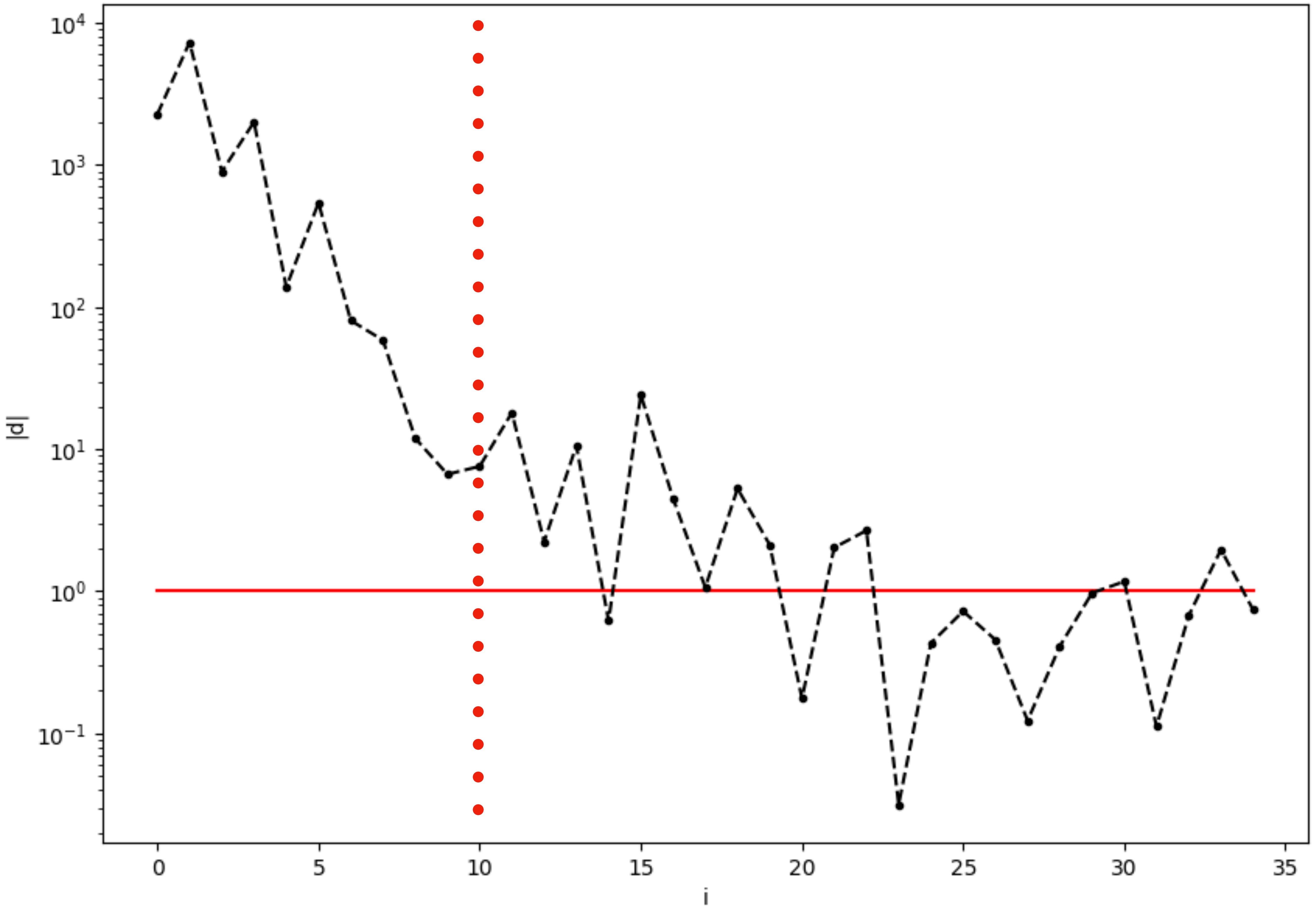
- Solve:

$$(At - m)^T (At - m) + \tau (Ct)^T Ct = \min$$

- Choice of regularisation strength τ

4. Solution: Singular Value Decomposition

Significant k Random fluctuations

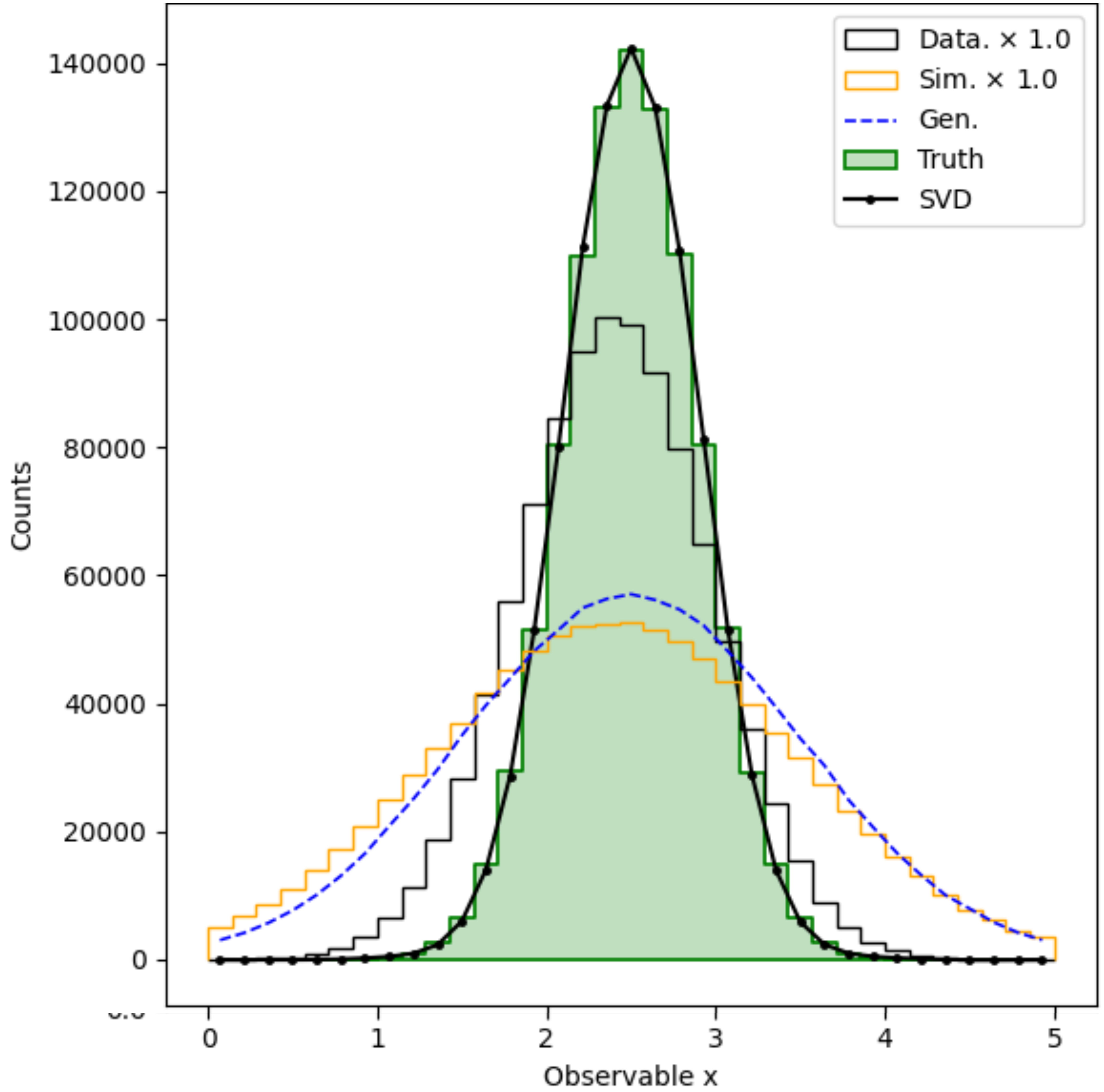


Components of the input vector in the SVD basis

Solution: Cutoff for singular values $< \tau = \sigma_k^2$.

Smooth cutoff: $d_i^\tau = d_i \frac{\sigma_i^2}{\sigma_i^2 + \tau}$

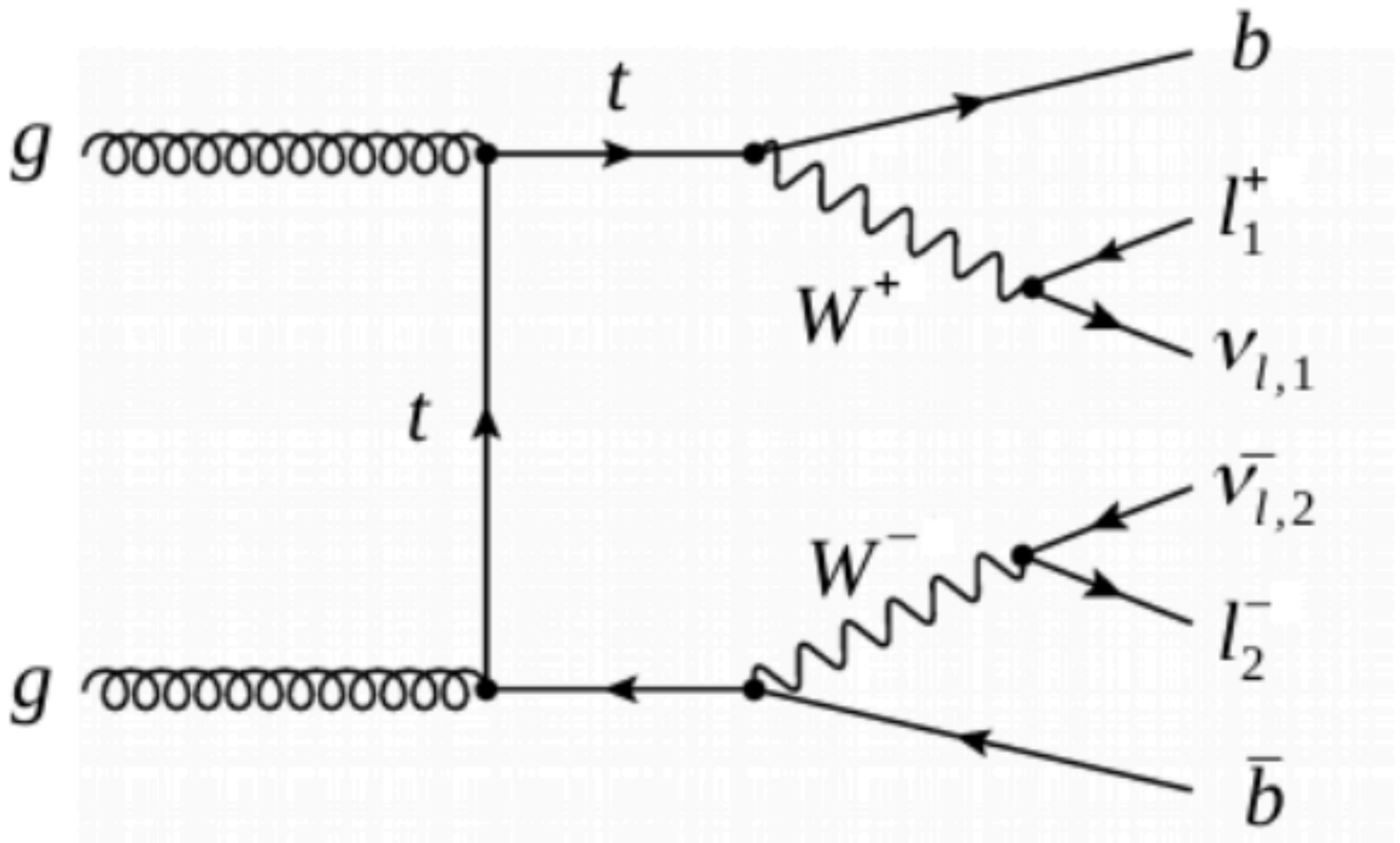
4. Solution: Singular Value Decomposition



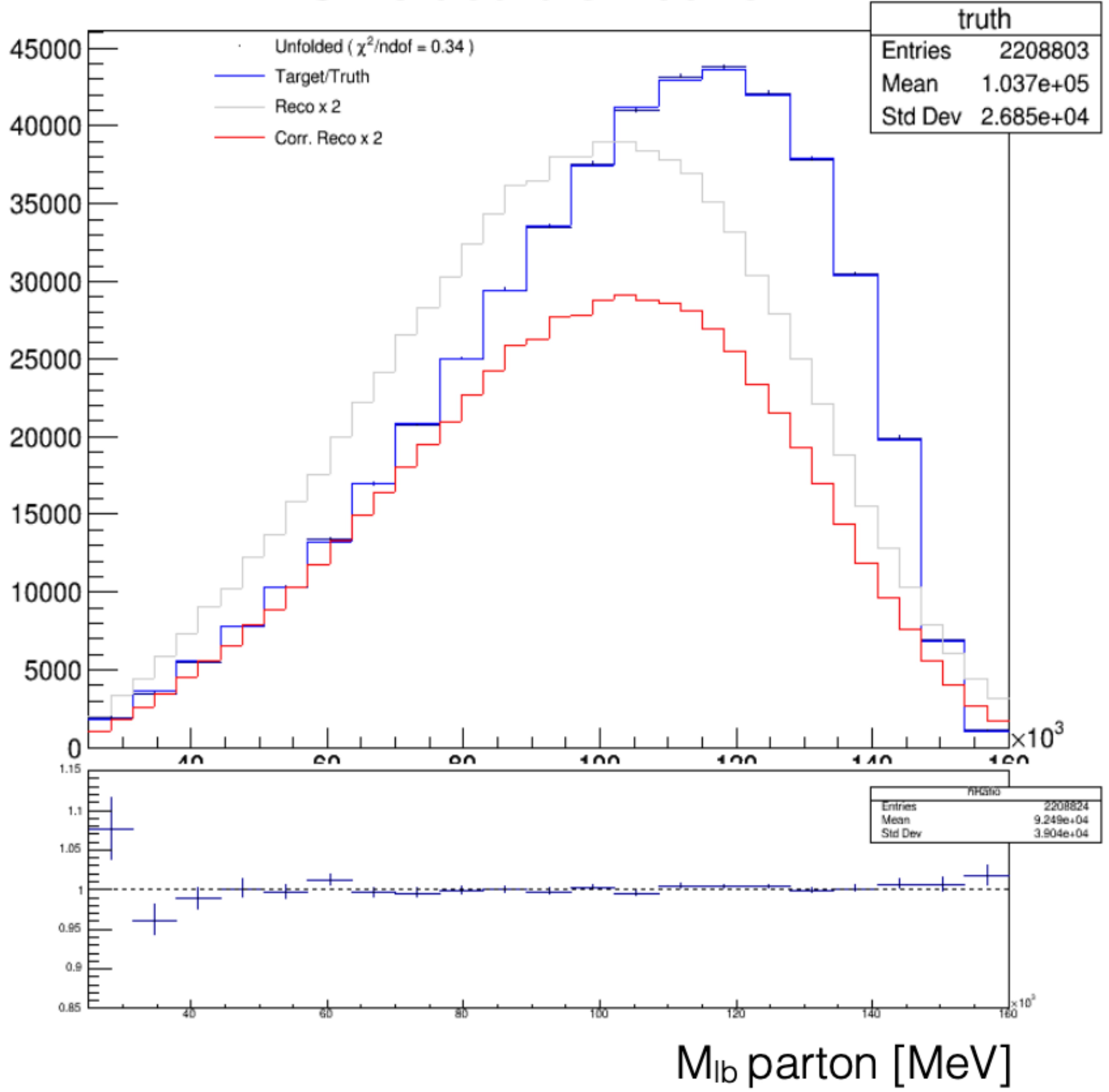
- Other methods:
  - Iterative Bayesian Unfolding
  - ML-based algorithms

5. My use case

- Information on the top-quark *pole* mass
- Channel: $pp \to t\bar{t} \to l^- b l \bar{b} (\nu\nu)$



Unfolded distribution



- Observable: Invariant mass of the lepton-b-system
- Use SVD to unfold m_{lb} from detector to parton level
- Extract $m_{\text{top}}^{\text{pole}}$ from comparison with fixed-order QCD predictions

Unfolding ATLAS data

IMPRS Colloquium | March 13th 2025

Elia Schmidt