Unfolding procedures

Kerstin Tackmann (DESY)

summarizing studies and thoughts from ATLAS

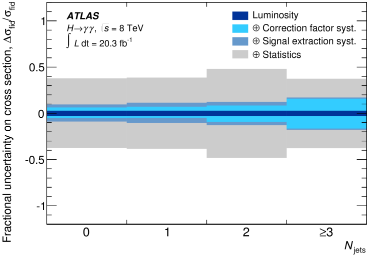
2nd LHCHXS meeting on fiducial cross sections

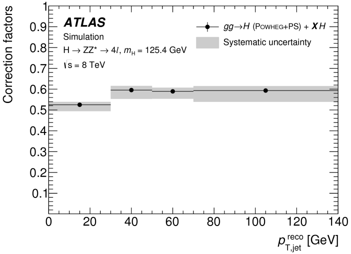
November 27, 2015

Recap of ATLAS unfolding procedure.

Detailed talk on ATLAS unfolding procedure was shown by Bijan Haney at the last meeting

- Bin-by-bin unfolding correcting for resolution, efficiency and acceptance effects: $\sigma_i = N_i^{\text{signal}}/(\mathcal{L}C_i)$
  - ★ Cross-checked against iterative unfolding
- Uncertainties on correction factors evaluated by varying composition of production modes, standard perturbative and Pdf uncertainties, ..., reweighting MC to match better measured distributions, ...
- With the present statistics, systematic uncertainties are small compared to statistical uncertainties





A few words on biases in unfolding.

- In general, unfolding introduces biases from simulation not perfectly describing the physics in data, and more important for larger bin-to-bin migrations
  - $\rightarrow$ more important for jet-related variables than photon- or lepton-related variables

Biases introduced through...

- physics/model dependence of correction factors/detector response matrix
  - ⋆ Bin-by-bin unfolding, “CMS method”, ...
- regularization procedure (usually uses MC truth distribution)
  - ⋆ Iterative unfolding, SVD, IDS, ...

Different methods are different methods and do not necessarily have identical uncertainties (statistical and systematic), biases and correlations

$\rightarrow$ Choice of method is a trade-off between biases and variance, and depends on the problem (usually no “single right solution”)

- ⋆ Of course for any method biases need to be estimated properly

Estimating uncertainties/unfolding biases.

Relevant sources of uncertainties/biases for differential Higgs cross sections

- relative contributions of the different production modes
- shape of individual spectra (missing higher-order corrections, pdf choice, ...)

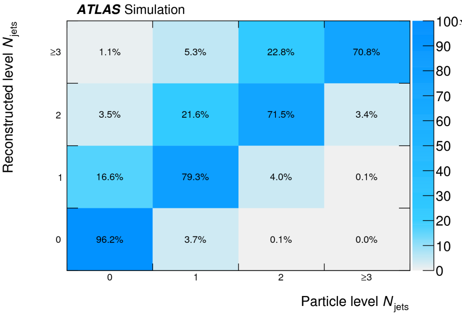
$\rightarrow$ Need to be varied in large enough range to estimate uncertainties from unfolding procedure

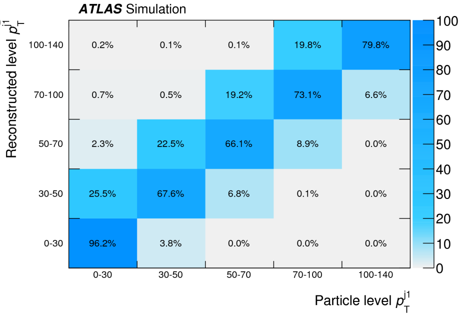
NB: Estimating these biases is an important part of determining proper regularization strength for regularized unfolding

Determination of binning.

- Binning is chosen to limit migration effects
  - ⋆ $p > 60\%$, where $p = N_i^{\text{fid+reco}} / N_{\text{reco}}$
  - ⋆ ...not a very strict requirement...
- On the other hand, larger bins can enlarge physics/model dependence

$P(\text{truth bin}|\text{reco bin})$

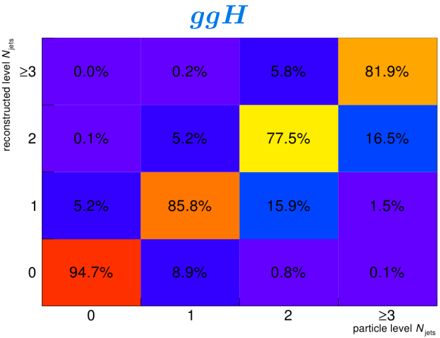


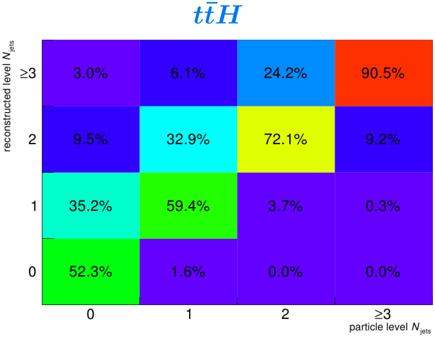


Physics/model dependence of detector response.

- Not just bin-by-bin correction factors are physics/model dependent, but also detector matrices

→ Also methods relying on the full detector response matrix need to estimate corresponding uncertainties/biases

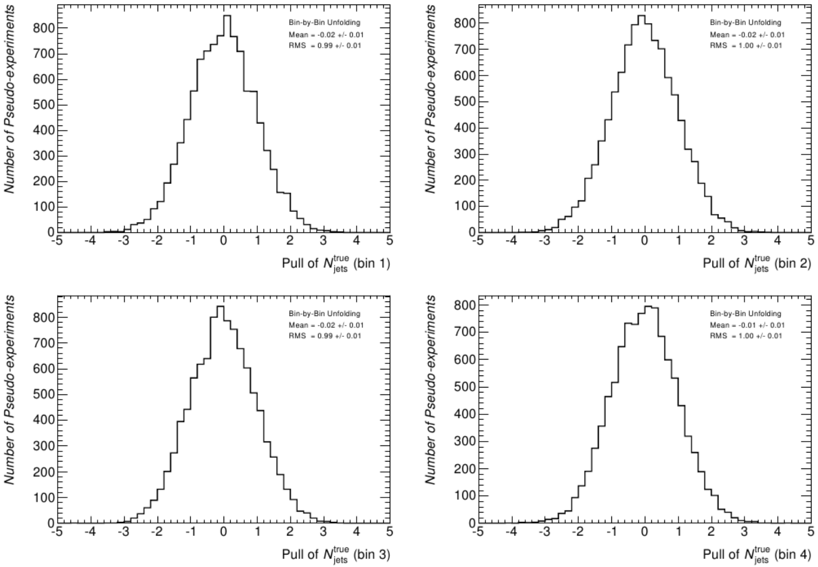




"Toy detector resolution matrices" – thanks to Dag Gillberg!

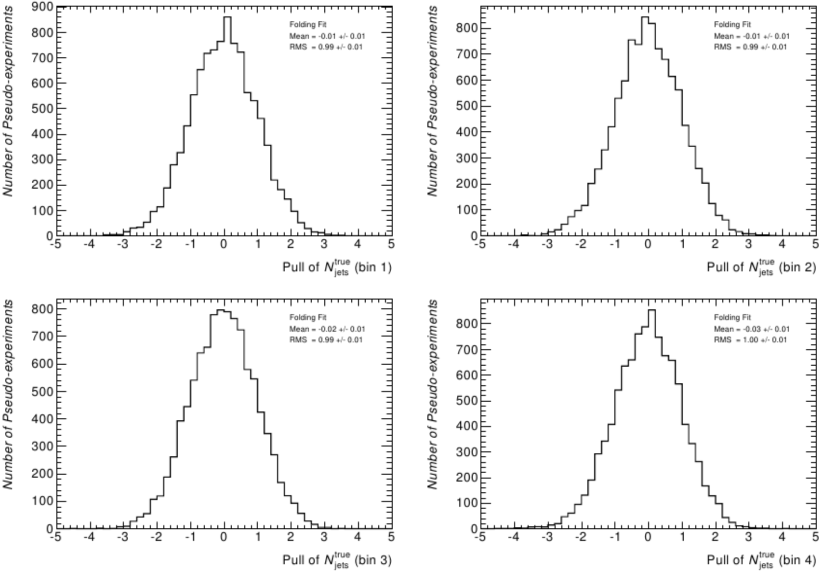
Simple unfolding tests.

- Same underlying truth distribution in “data” and “MC”



Simple unfolding tests.

- Same underlying truth distribution in “data” and “MC”



Summary.

- Simple toy tests ok: unbiased and ok coverage
- Ongoing: extend study to different underlying truth distributions in “data” and “MC” to check biases and coverage
  - ⋆ Expect to see biases with any method
  - ⋆ Corresponding uncertainties need to be estimated and should not dominate uncertainties

Backup

Bin-by-bin unfolding.

$$\frac{1}{C_i} = \frac{N_i^{\mathrm{Fid}}}{N_i^{\mathrm{Reco}}} = \frac{P_i}{\epsilon_i} \qquad P_i = \frac{N_i^{\mathrm{Fid\&Reco}}}{N_i^{\mathrm{Reco}}} \qquad \epsilon_i = \frac{N_i^{\mathrm{Fid\&Reco}}}{N_i^{\mathrm{Fid}}}$$

- Bin-by-bin correction factors, c_i, are calculated from MC simulations in order to correct for detector effects.
  - N^{Fid} is the # of truth level MC events after event selection within a fiducial volume.
  - N^{Reco} is the # of MC events after event selection with detector effects (e.g. gaps in the detector, Jet reconstruction efficiency, other smearing effects, etc.)
  - $N^{\mathrm{Fid\&Reco}}$ are events that pass the Higgs event selection under both circumstances.
- Purity, P_i, accounts for the number of fakes in a bin.
- Efficiency, ϵ_i, accounts for poor object reconstruction and identification.

[Slide from Bijan]

Unfolding uncertainties.

There is uncertainty in both N^{Reco} and N^{fid} because the generators themselves may not match reality. There is extra uncertainty specifically in N^{Reco} because the MC smearing may not match reality.

- Generator Modeling and Uncertainty
  - Alternative MC generators were used and their envelope was taken as an uncertainty.
    - eigenvector variations of the baseline CT10 PDF.
    - central values of alternative MSTW2008NLO and NNPDF2.3 PDFs.
  - Signal composition of the production modes was varied.
    - VBF+WH+ZH production XS were doubled and halved.
    - ttH production XS was multiplied x5 and x0.
  - Varying the renormalization and factorization scales by double and a half.
  - Reweighing was applied to the MC to make it more closely reflect the observed distribution of data.
    - The unfolded data distributions of p_T and |y| were compared to fiducial MC predictions. Reweighing functions from data/MC were used to correct the fiducial p_T and |y| spectrum.
    - Data tend to have harder Higgs p_T, and more forward |y|.

[Slide from Bijan]