# H1 Style Calibration From Data

Rob Duxfield

University of Sheffield

# Outline

- Technique for obtaining H1-style (or similar) weights from data.
- $E_T^{miss}$-significance is minimised with respect to the weights in events with jets and little intrinsic $E_T^{miss}$ ($\gamma/Z$ + jets, multijets).
- So far just a proof-of-principle study. Not yet integrated with existing software.

- Motivation – why data-driven? why use $E_T^{miss}$?
- Details of the principle of the technique and its implementation.
- Some performance results ($E_T^{miss}$ only so far – still working on jets).

- Similar idea suggested by Andrei : http://indico.cern.ch/conferenceDisplay.py?confId=46967

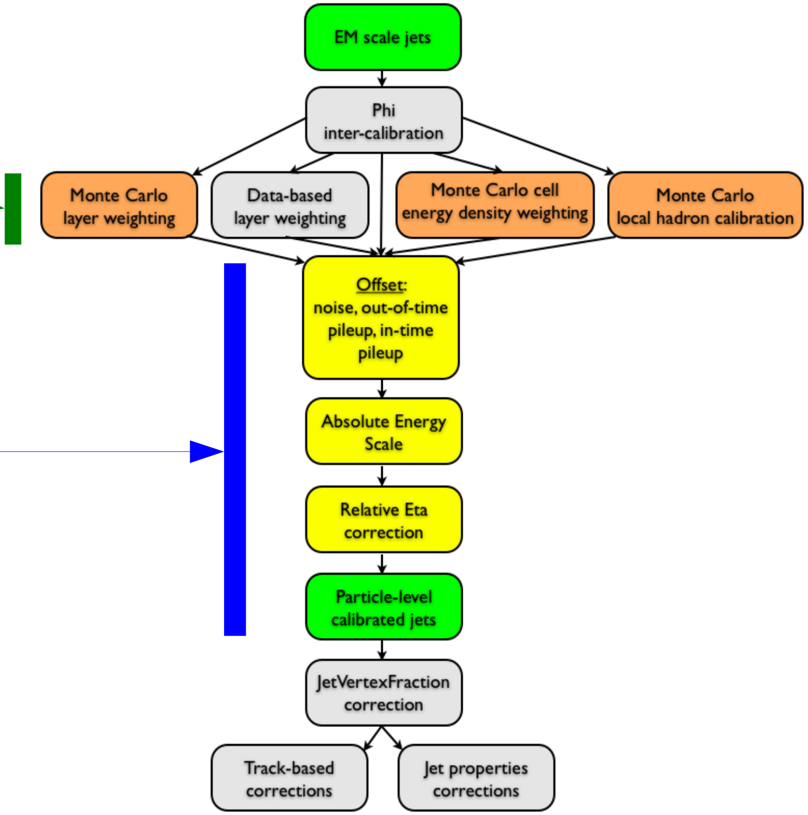
# Motivation

## Why data-driven?

- Data driven calibration is likely to provide the best possible performance with early data (ie. before MC is validated).
- Can also provide a reference for validation/understanding of MC results for later running.

## Why use $E_T^{miss}$ (not jets)?

- Ideally want cell weights to correct to hadronic scale and be independent of jet algorithm.
- Algorithm dependent corrections then applied to jets as a function of $p_T$, η (preferably obtained from data).
- Also –
  - Eliminates sentitivity to ISR, FSR.
  - No need for hard back-to-back cut.
    - Best use of available statistics



# The Technique – I

- The basic idea is to select events with a well measured, isolated photon and minimise the mean value of $S^2$ over all selected events with respect to the H1 weights, where $S$ is the $E_T^{\text{miss}}$–significance, given by

$$S = \frac{E_T^{\text{miss}}}{\sqrt{E_T^{\text{sum}}}}$$

- The denominator factors out the dependence of the $E_T^{\text{miss}}$ resolution on the $p_T$ scale of the event.

- Have used the same geometric and cell $E/V$ binning scheme as for the rel–12 H1 weights – see, eg. CaloUtils/H1WeightToolCSC12.cxx.
- 166 + 1 (cyostat) bins in total.
- So, using label $i$ for H1 bin number (and neglecting the cryostat and muon terms),

$$E_X = \sum_i w_i^{\text{H1}} \sum_{\text{cells in } i} E_X^{\text{cell}} = \sum_i w_i^{\text{H1}} E_X^i,$$

and similarly for $E_Y$ and $E_T^{\text{sum}}$, where the cell energy is measured at the EM scale.

# The Technique – II

- For the purposes of minimisation the sums $\{E_X^i\}$, $\{E_Y^i\}$ and $\{E_T^i\}$ are calculated from cells *not associated with the photon*.
- The $E_T^{miss}$–significance is then calculated from

$$\begin{array}{ccccc} & \text{hadronic} & & \text{photon} & & \text{muon} \\ -E_{X(Y)}^{miss} = & \sum_i w_i^{H1} E_{X(Y)}^i + w_{cryo}^{H1} E_{X(Y)}^{cryo} & + & E_{X(Y)}^{\gamma} & + & E_{X(Y)}^{MET\_MuonBoy} \\ E_T^{sum} = & \sum_i w_i^{H1} E_T^i + w_{cryo}^{H1} E_T^{cryo} & & & & \end{array}$$

- The hadronic component is assumed to dominate the variance in the reconstruction of $E_T^{miss}$, so only this component is used in the calculation of $E_T^{sum}$.
- The function

$$f\left(w_i^{H1}, w_{cryo}^{H1}\right) = \sum_{events} S^2$$

can then be minimised with respect to the weights to obtain the optimal values.

# The Technique – III

- The technique described on the previous slides uses $\gamma$ + jets events to obtain the absolute values of the cell weights.
- However is possible to obtain relative cell weights using multijet events.
- The procedure is almost identical to that for the $\gamma$ + jets case, except during the minimisation stage where one of the weights must be kept fixed (at 1.0, say).
- The advantage in using multijet events is that more statistics will be available at high $p_T$ (high cell $E/V$) allowing more weights to be determined.
- If the fixed bin is chosen such that it typically contains cells with a significant portion of the activity in the event for a wide range of event $p_T$ scales, $\gamma/Z$ + jets events can be used to determine the absolute scale.

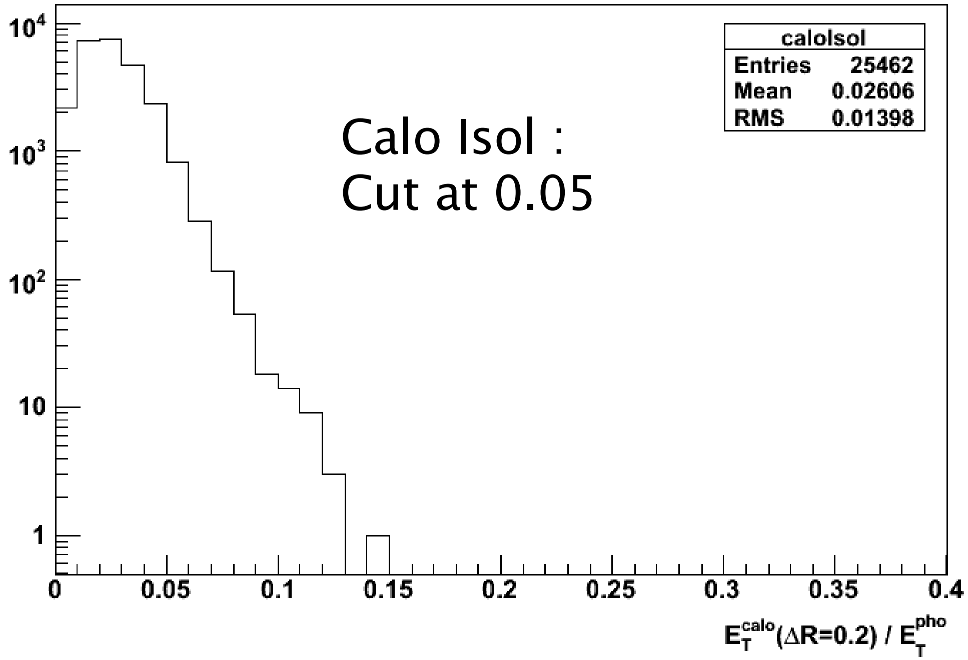
# Event Preselection + Data Format

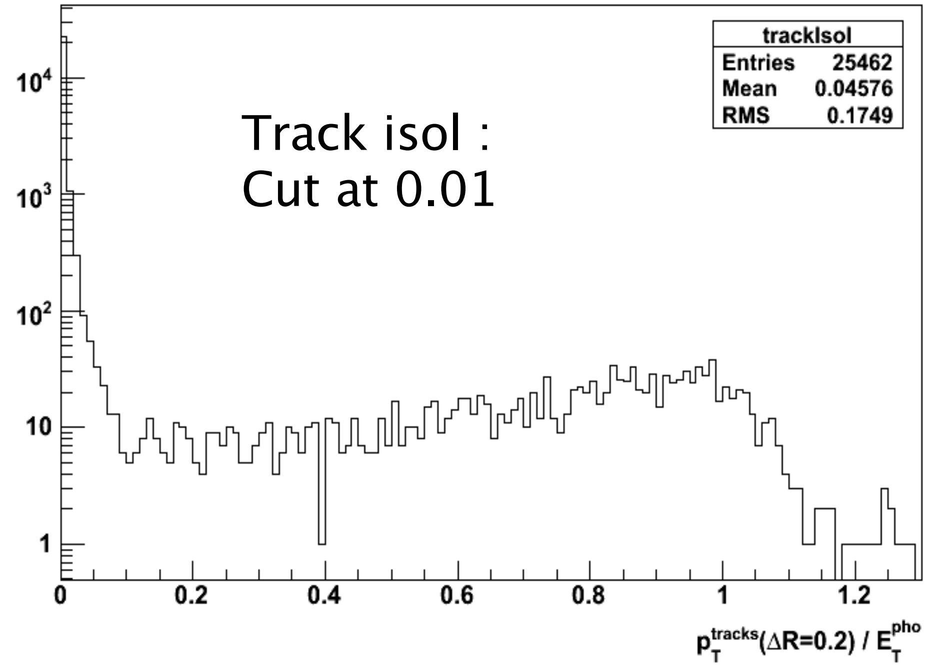
- The samples used for this study were

  mc08.10800?.PythiaPhotonJet?.recon.ESD.e344_s456_r545 (J2 – J4)
  mc08.10501?.J?_pythia_jetjet.recon.ESD.e344_s479_r451 (J2 – J5)
- Event preselection was performed in an ATHENA algorithm and event data was written to a ROOT ntuple using a common format for both samples.
- Key information written to the ntuples is the EM-scale cell-sum vectors $\{E_X^i\}$, $\{E_Y^i\}$ and $\{E_T^i\}$.
- Only cells from CaloTopoCluster collection used (with geo. weights applied).
- Samples used have displaced vertex, so correction to cell $E_T$ was made using the reconstructed primary vertex (corrected and uncorrected sums written out).
- $\gamma$ + jets events were required to pass the g25 trigger and to have a 'tight' photon and at least one jet with $p_T > 20$GeV.
- Leading jet also required to be back-to-back with the photon: $\pi - |\Delta\phi(j, \gamma)| < 1.0$
- No preselection cuts were applied to the dijet sample.
- Ntuple size ~3k / event passing preselection.

# Selection of Training Sample – I

- Used standalone root-based c++ to select training sample and perform minimisation.
- To avoid biasing towards high response due to requirement of a 20GeV jet in preselection, additionally required photon to have $p_T > 50$GeV.
- To ensure good photon energy reconstruction (and in anticipation of QCD background), cut on calorimeter and track isolation (cone $\Delta R = 0.2$).



# Selection of Training Sample – II

- Samples used have HEC section switched off ($-3.2 < \eta < 1.5$, $-\pi/2 < \phi < 0$).
- Events containing any jet pointing into this region were vetoed.
- Any events with $-\pi/2 < \phi(E_T^{miss}) < 0$ AND $E_T^{sum}(EME) / E_T^{sum}(all) > 5\%$ were also vetoed.

- The *z* position of the reconstructed primary vertex was required to be within 15cm of the nominal interaction point.

- To avoid sensitivity to errors in muon reconstruction, events with MET_MuonBoy_EtSum > 100MeV were also vetoed.

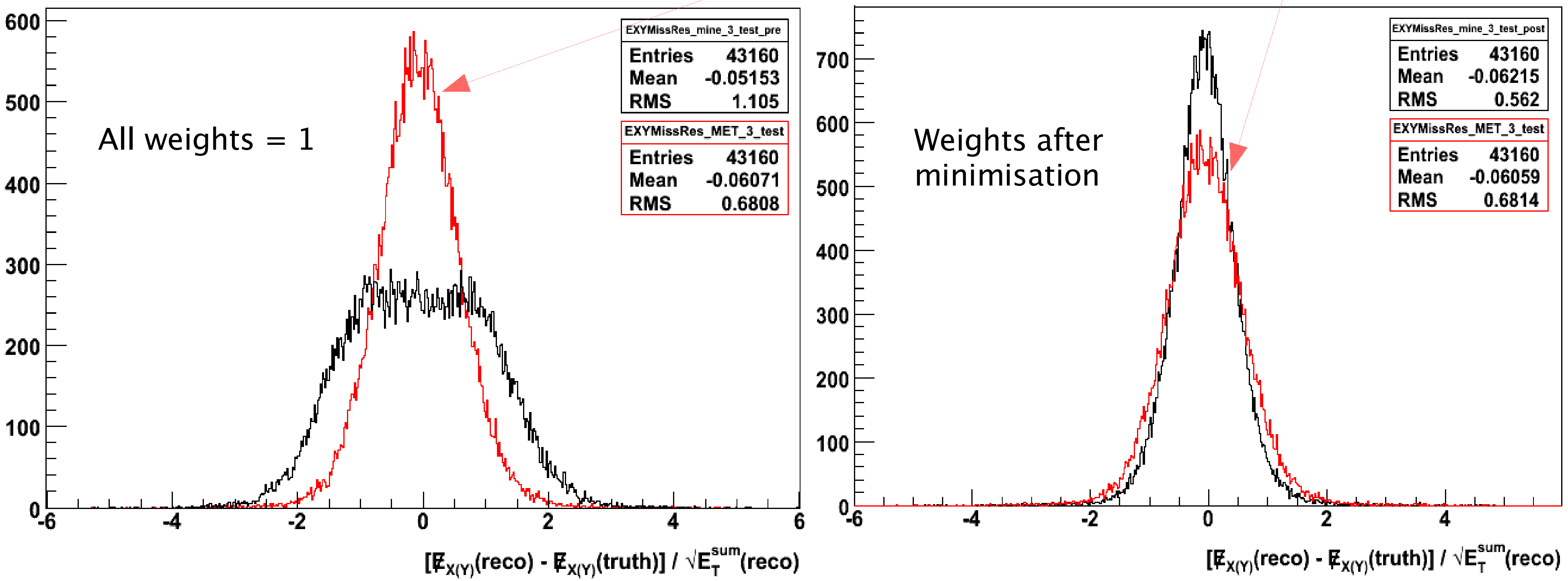
- Left with 19004 events from $10.4pb^{-1}$ sample.

# Minimisation Procedure

- Function *f* must be minimised with respect to all parameters simultaneously.
- Heavily CPU intensive function with 167 parameters.
- Used MINUIT, but had to break up the problem to get it to converge...
  - Obtain initial values by assigning a single weight to each geometric bin and minimising wrt these weights.
  - Select subset of training sample with all jets in the barrel and minimise *f* wrt all EMB, Tile and Cryostat weights.
  - Select events with leading jet in the endcap and minimise *f* wrt EME, HEC weights.
  - Select events with leading or second jet in forward region and minimise *f* wrt FCAL weights.
  - Perform full minimisation wrt all weights.
- Using this procedure, the minimisation converges in a total of ~12h.

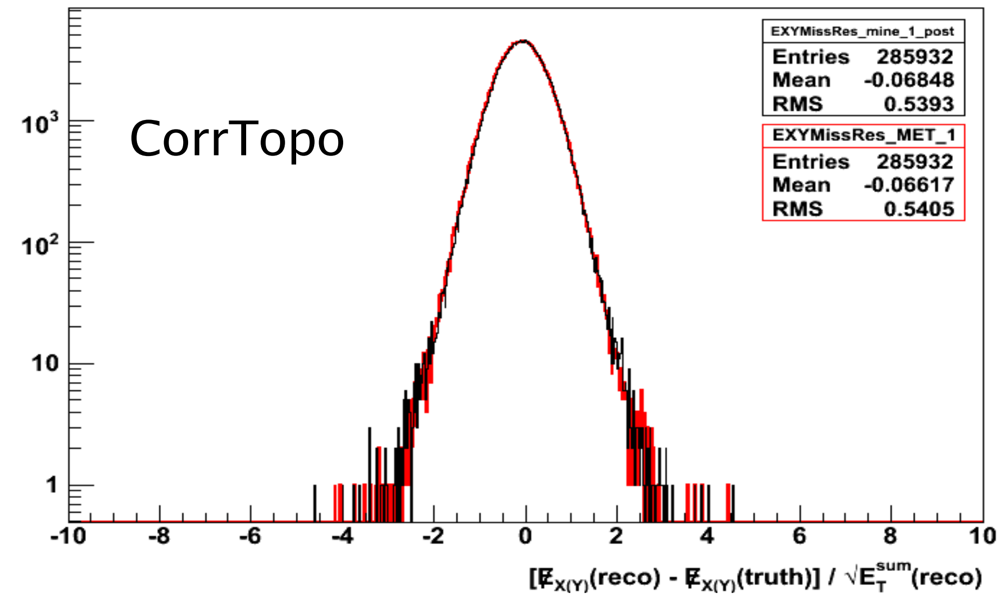
# Minimisation Results : $\gamma$ + jets

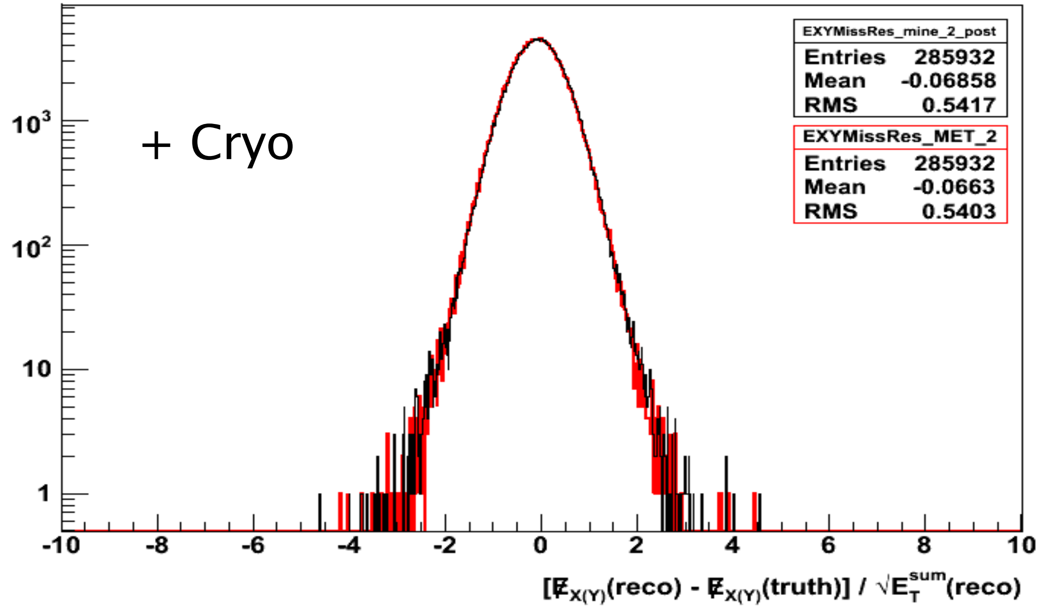
- $E_T^{miss}$ resolution before and after minimisation (no HEC or MET_Muon cut).
- All cells (including photon) calibrated with hadronic weights.
- MET_CorrTopo + MET_Cryo + MET_MuonBoy calibrated with CSC12 weights for comparison.
- Vertex correction used for minimisation, but not for comparison with existing calibration (small effect anyway).

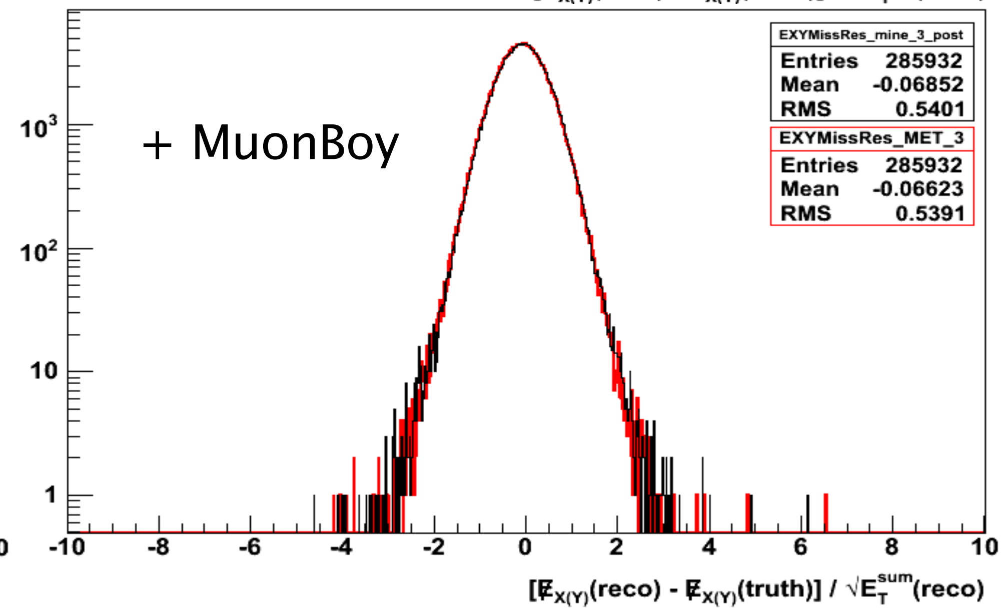


# $E_T^{miss}$ Performance : J2 dijets

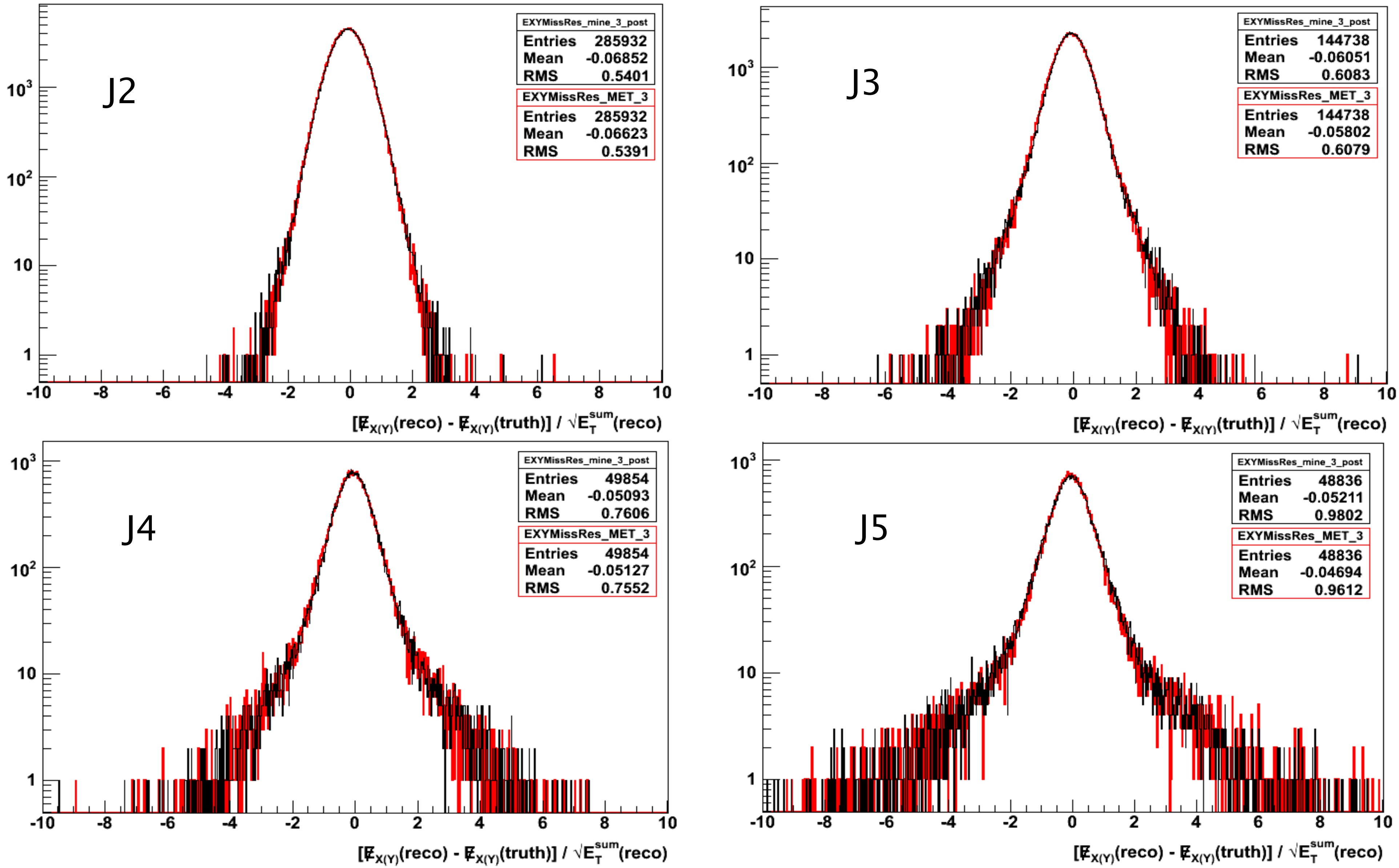
- Weights obtained from γ + jets analysis applied to dijet events.
- Again, equivalent distributions using standard CSC12 weights are shown for comparison.
- See similar performance to CSC12 weights in J2 dijets.
- Only cut is $N_J \geq 2$.



# $E_T^{miss}$ Performance : J2–J5 dijets



# Conclusions & Future Plans

- Method seems to work well for obtaining an $E_T^{miss}$ calibration and may provide a robust alternative to MC methods with early data, but...
- Need to test performance on jets.
- Want to look at using multijets to obtain weights and test on a wider range of samples (see slide 6).
- Need to consider how to handle hot/dead cells etc. (similar to missing HEC section?).
- Investigate systematic effects of cell selection (clustered v. all, only using cells within given $\Delta R$ of a jet).
- Look at weight distributions and compare with MC derived weights.
- Move towards 'proper' implementation – integrate with existing software and use available tools as far as possible.