

Camille Bélanger-Champagne  
*for the Uppsala University group*



UPPSALA  
UNIVERSITET

# Toolkit for MC tuning on first data

ATLAS Tau Workshop  
NBI, Copenhagen, April 15<sup>th</sup> 2009



# Why this tool?

- Our MC will not match the first data perfectly.
  - Physics process modelling ( $p_T$  distribution, jet shapes, etc), Sample Mixing (QCD), Detector modelling
- Traditional resolution:
  - Improve theoretical cross-section calculations, PYTHIA, tune, GEANT representation of detectors, etc.
  - Regenerate MC samples
- Very long procedure: months up to year(s)
- Adjustments to e.g. tau trigger selection need to be done much faster
- Toolkit provides basis to make informed decisions while in-depth work ongoing



# What does it do?

- General tool to derive an event-by-event weight

$$W(v_1, v_2, v_3, \dots, v_N) = W_1(v_1) * W_2(v_2) * \dots * W_N(v_N)$$

by simultaneously optimising 1D histograms of the variables  $v_1, v_2, \dots, x_1, x_2, \dots$

- Optimisation can be min/maximisation of

- Log(Poisson likelihood)
- Log(Kolmogorov-Smirnow likelihood)
- $\chi^2$

of MC distributions with respect to data.



# What does it do?

- Effective “tuning” of MC sample to data.
  - Study of effect of changes in algorithms/cuts
  - Adjust the relative composition of the sample (combining QCD samples)
  - Derive smearing coefficients
- Can use one set of variables  $v_1, v_2, \dots$  for event weights and another set of variables  $x_1, x_2, \dots$  to “tune”
  - Learn about correlations between variables
  - Identify critical variables (see example 1)
  - Identify minimal set of independent variables necessary for data tuning



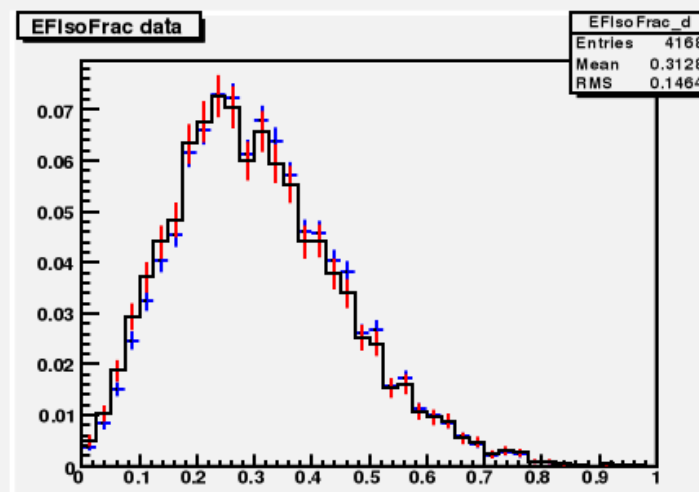
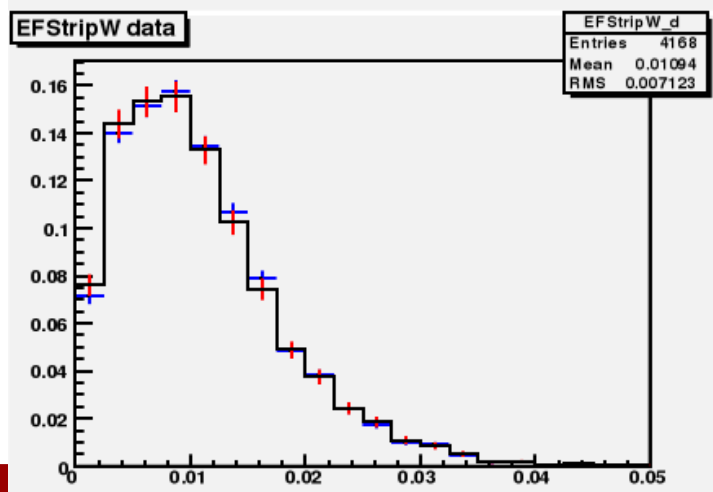
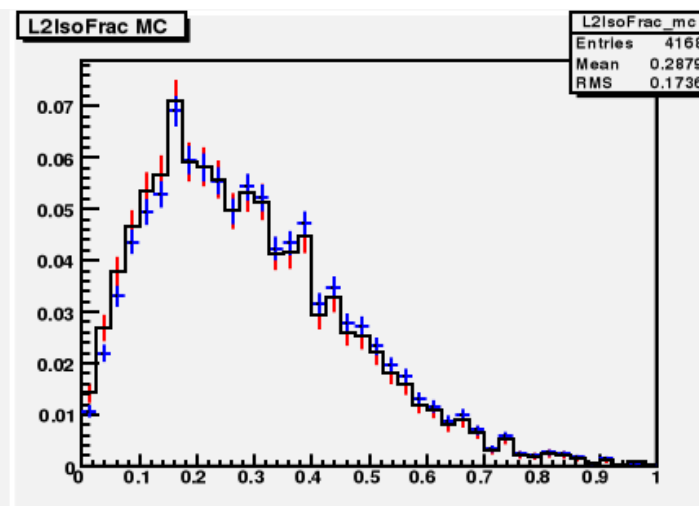
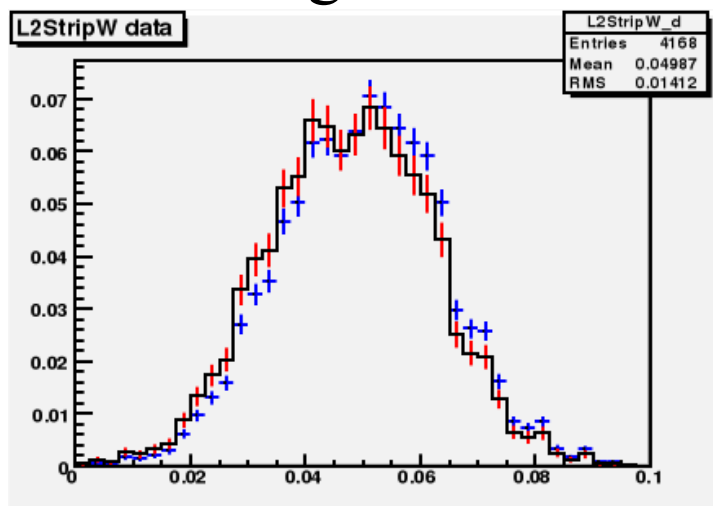
# Some words of caution...

- Toolkit provides event-by-event weight
  - Not the same as a scaling/smearing factor
  - Events are used as is with attached weight, no need to re-reco
  - Many variables can be used to determine the unique event weight, but scaling/smearing has to be done for each variable/detector subsystem
- Not a magic black box: requires some brain involvement at all steps of the process:
  - Choice of weight function
  - Inclusion of variables, etc

# Example 1: Critical variables

- One variable is imperfectly modelled: effect identical to having a small linear weight applied to all events, based on this variable.
- Resulting distributions:

Reference sample "data"  
To-be-weighted sample "MC"  
Weighted sample, here all  $w=1$ , identical to red sample



So, which variable is imperfectly modelled?  
Statistical tests do not help...

	Chi2	KS
L2SW	117.25	2.40E-008
L2IsoF	28.72	0.05
EFSW	6.63	0.94
EFIsoF	30.67	0.05
Combined	183.27	5.97E-011

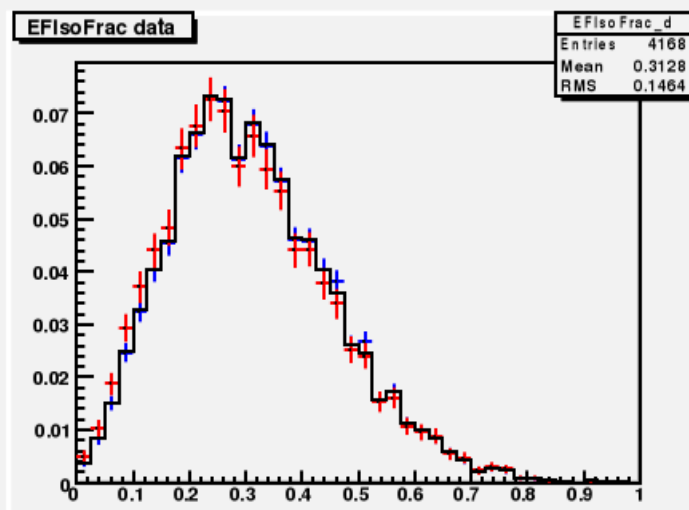
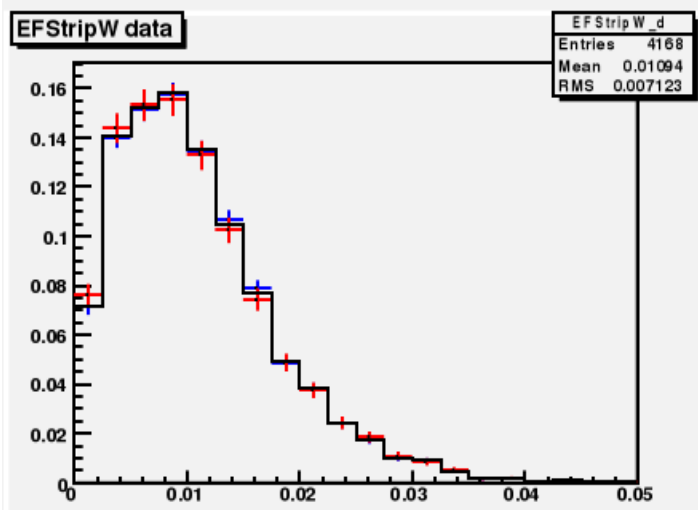
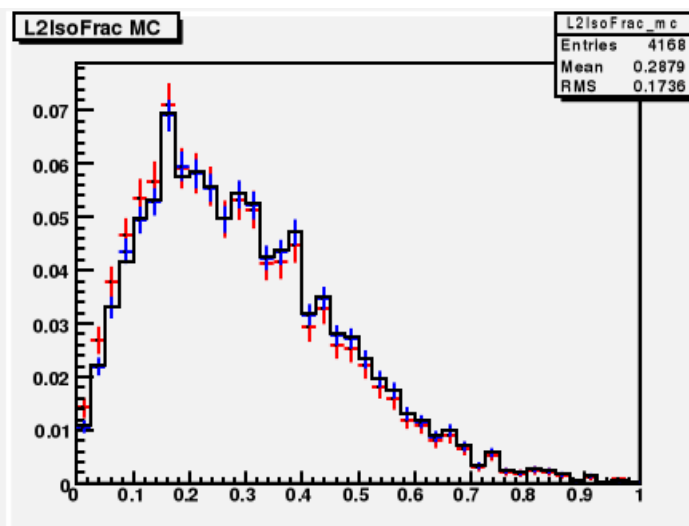
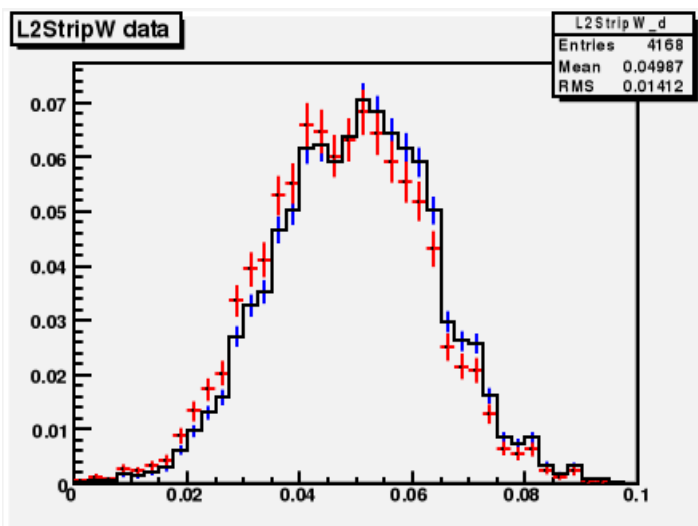


# Example 1

- Reweight according to the variable that is shifted (L2StripWidth).

Reference sample “data”  
To-be-weighted sample “MC”  
Weighted sample

All variables do match, as expected!...  
Reflected in improved chi square, KS



	Chi2	KS
L2SW	9.45E-005	1
L2IsoF	1.1	1
EFSW	0.55	1
EFIsoF	1.72	1
Combined	3.37	1

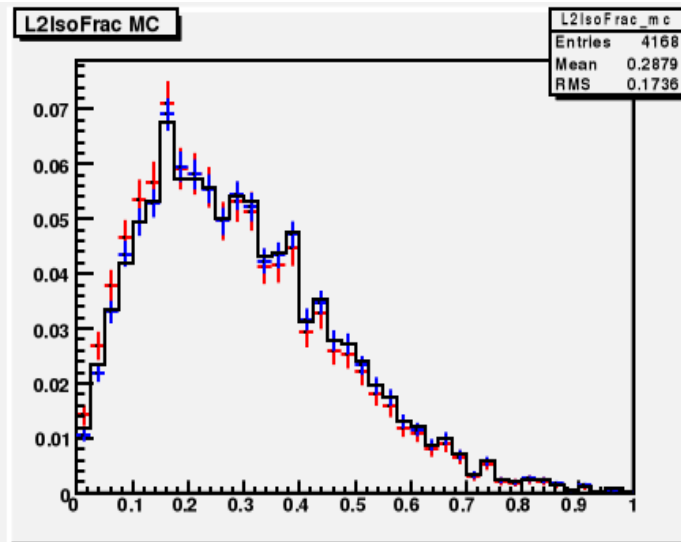
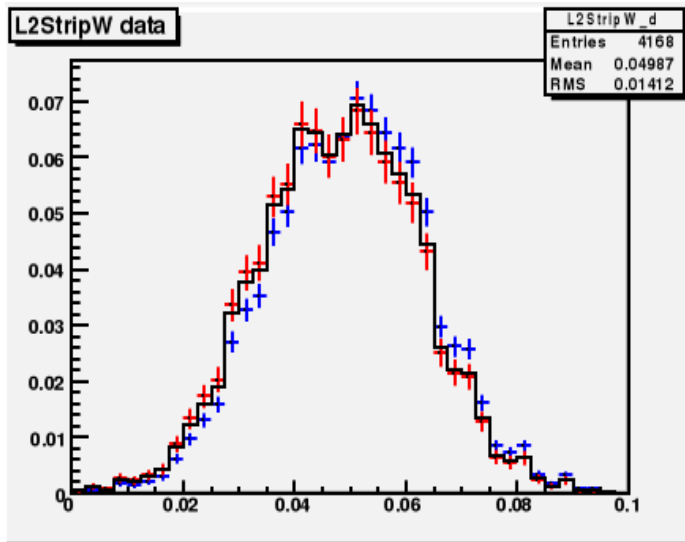


# Example 1

- Reweight according to any other variable (here L2IsolationFraction).

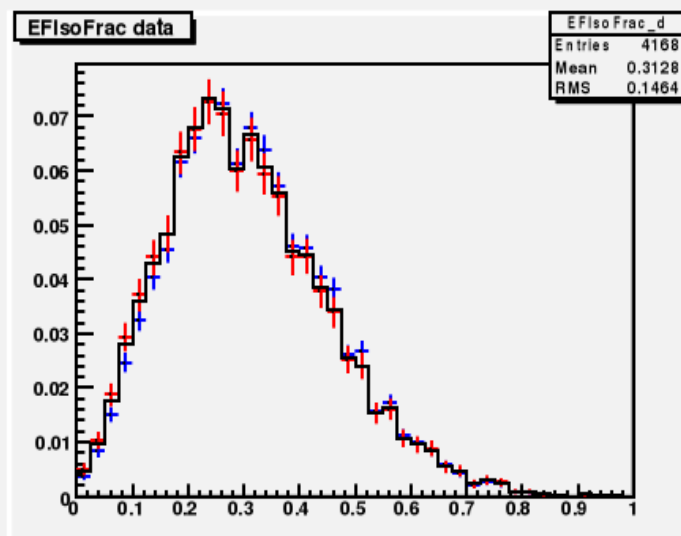
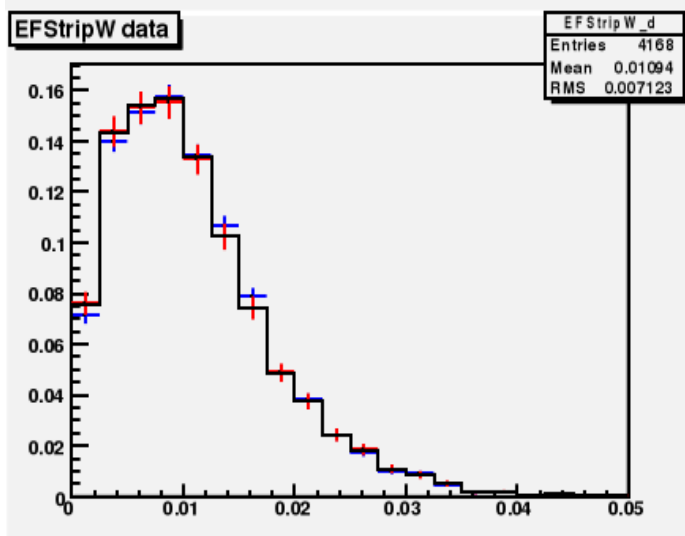
Reference sample “data”  
To-be-weighted sample “MC”  
Weighted sample

Shift still present in other variables...



	Chi2	KS
L2SW	76.71	2.61E-005
<b>L2IsoF</b>	<b>76.71</b>	<b>1</b>
EFSW	4.95	0.97
EFIsoF	17.81	0.23
Combined	103.05	5.82E-006

Similarly for the other 2 other variables



	Chi2	KS
L2SW	106.31	1.55E-007
L2IsoF	25.15	0.09
<b>EFSW</b>	<b>2.27</b>	<b>1</b>
EFIsoF	24.43	0.11
Combined	158.16	1.52E-009

	Chi2	KS
L2SW	79.06	1.39E-005
L2IsoF	15.82	0.37
EFSW	3.83	0.999998
<b>EFIsoF</b>	<b>3.84</b>	<b>1</b>
Combined	102.56	5.19E-006

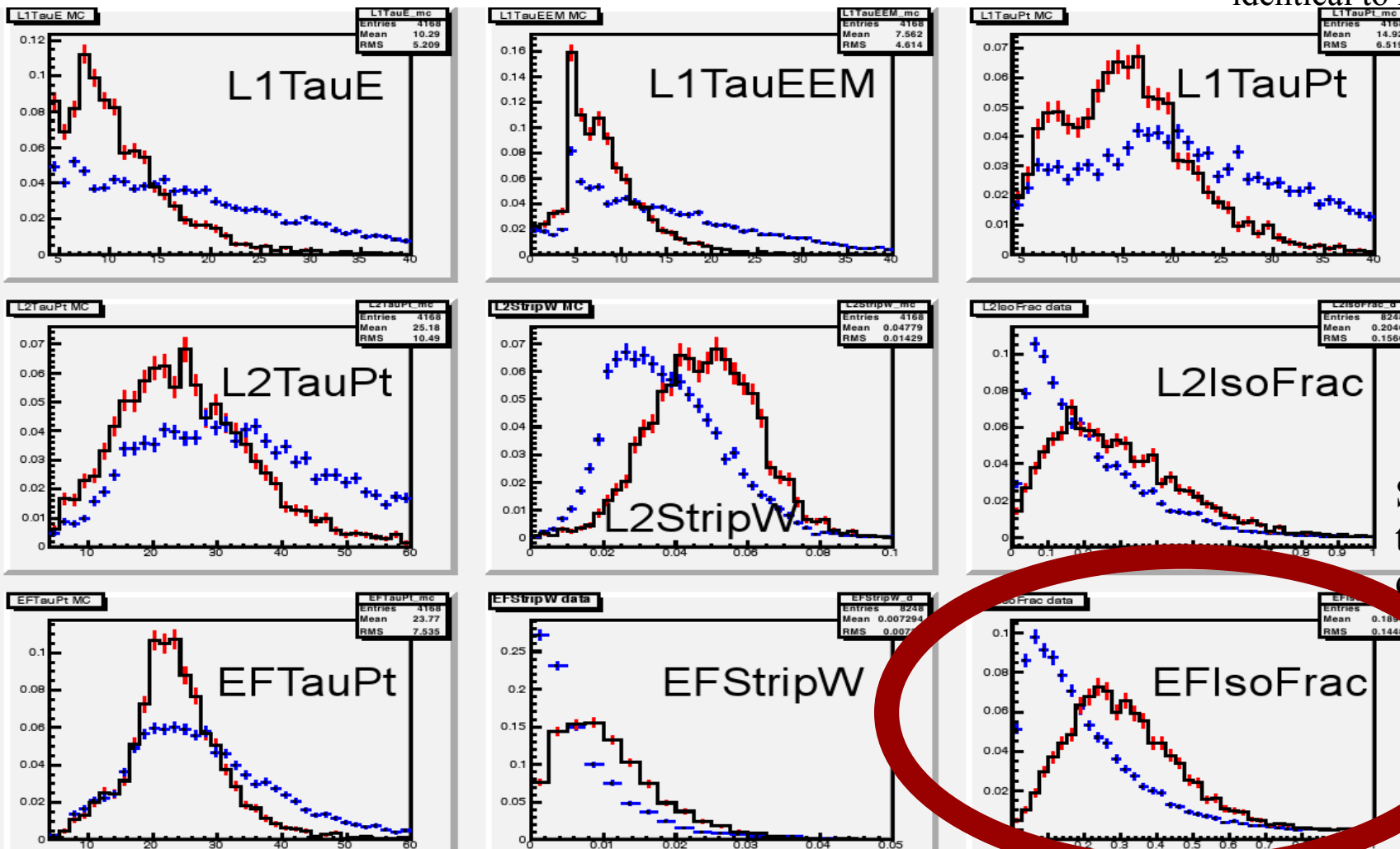




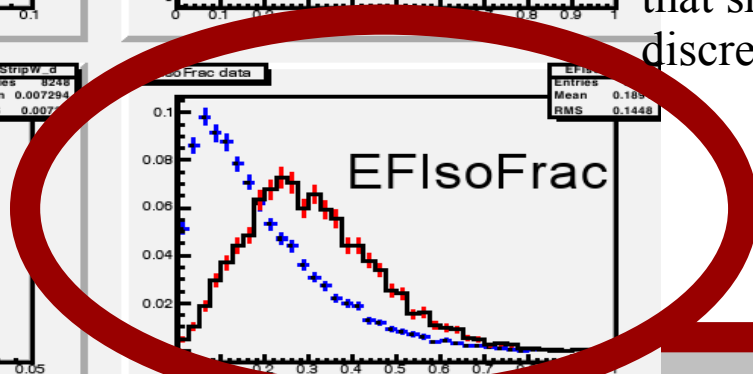
# Example 2: Overall mismatch

- Reweighting multiple variables, pathological case: make real taus from Z events look like QCD tau candidates

Reference sample “data”  
 To-be-weighted sample “MC”  
 Weighted sample, here all  $w=1$ , identical to red sample



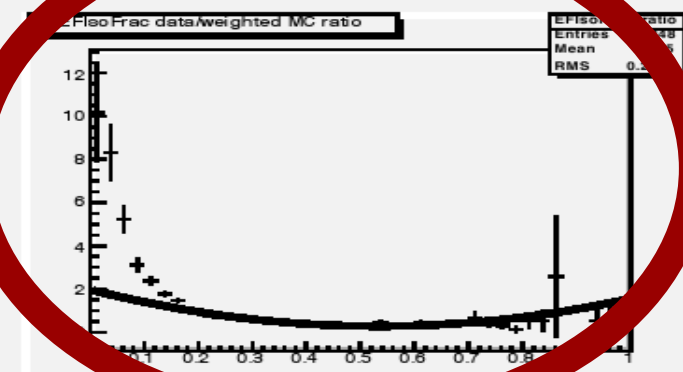
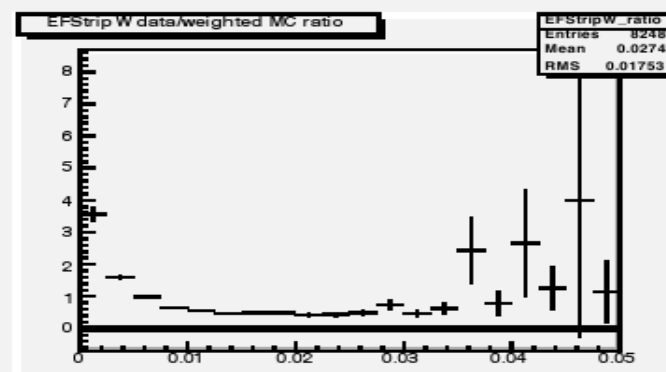
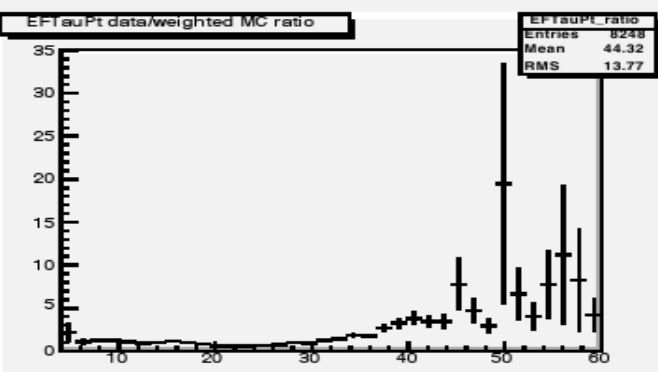
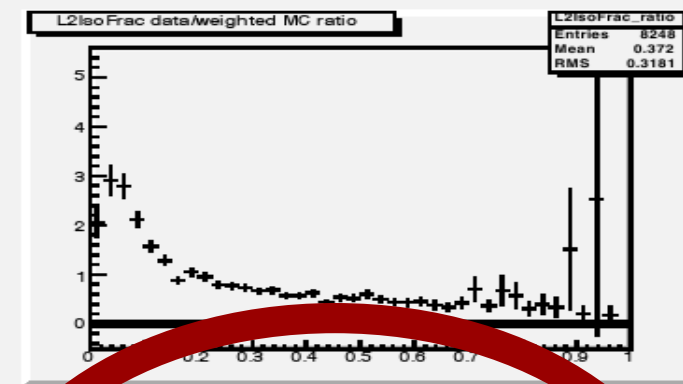
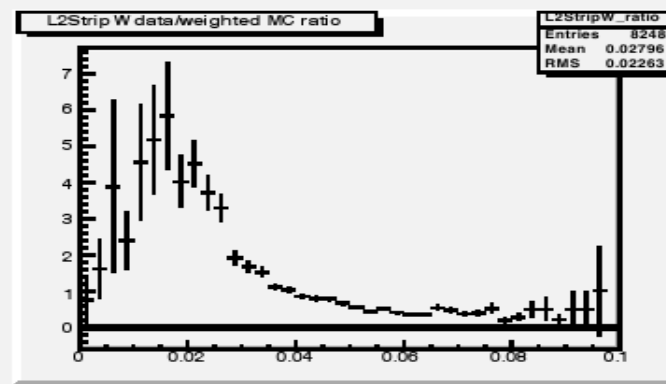
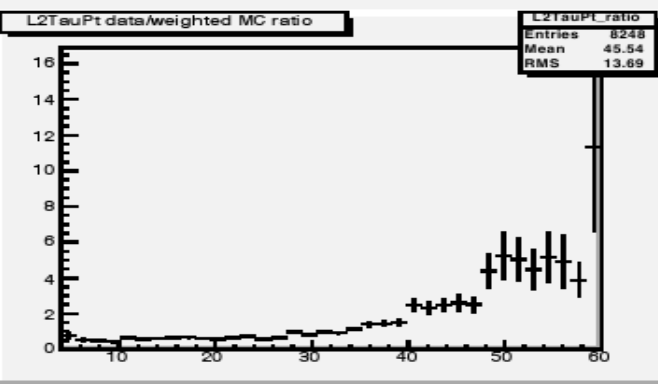
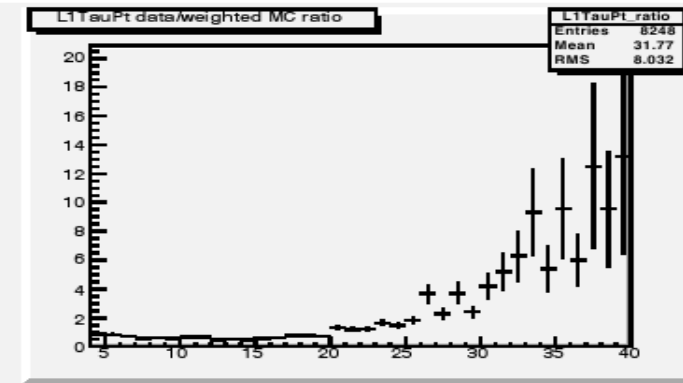
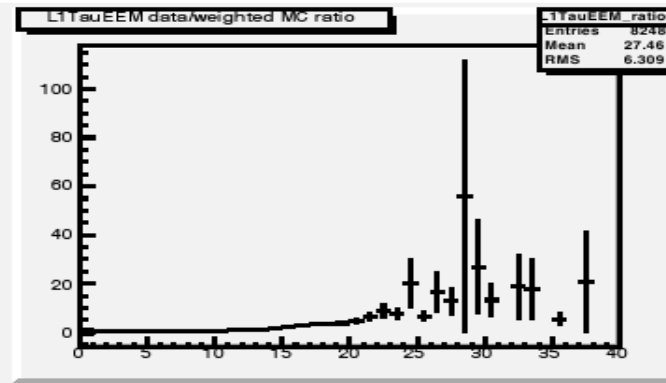
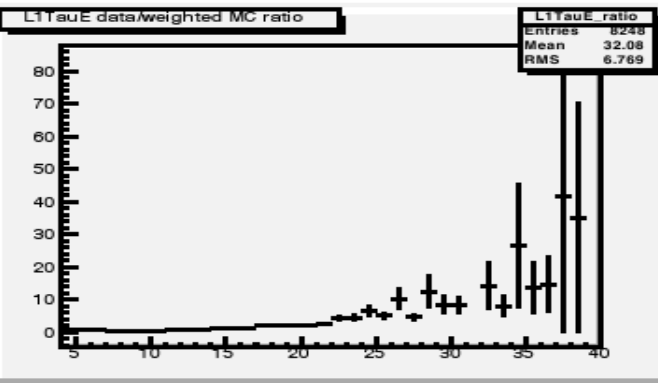
Start with variable that shows largest discrepancy!





# Example 2

- Starting point: ratio of “MC” to “data”:

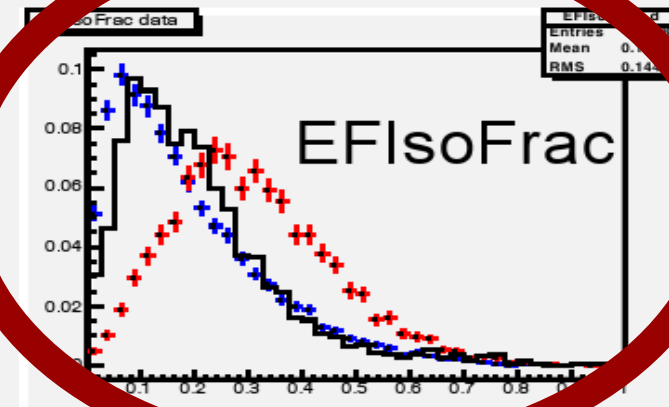
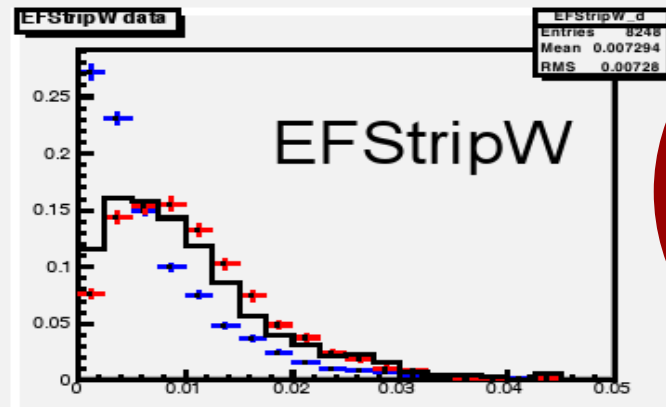
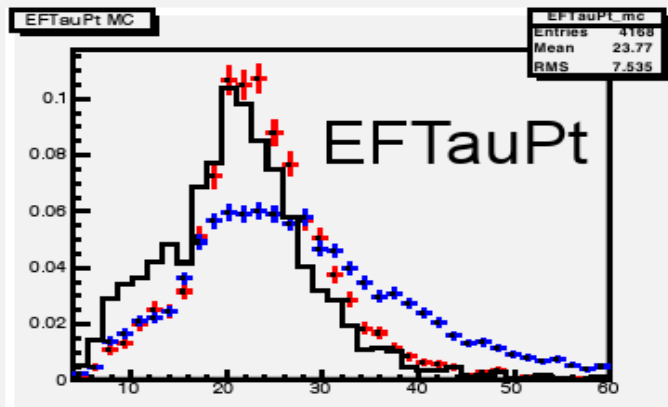
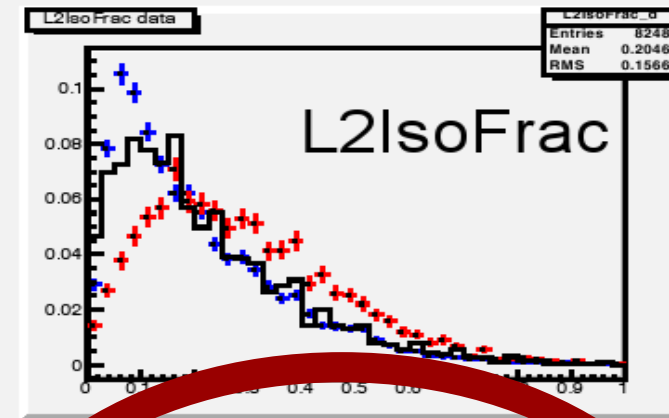
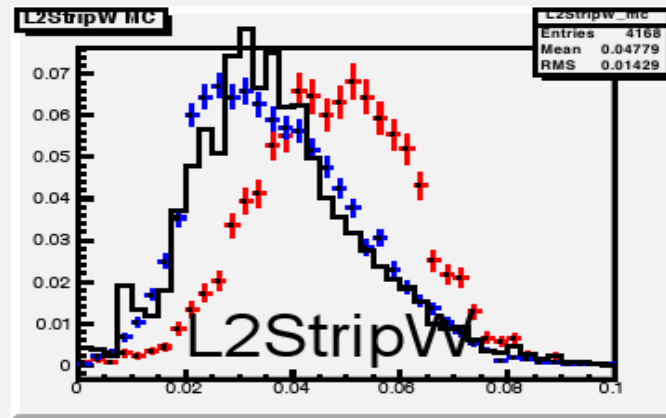
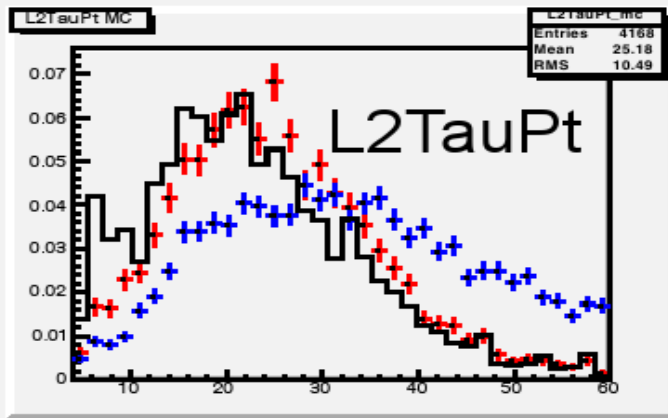
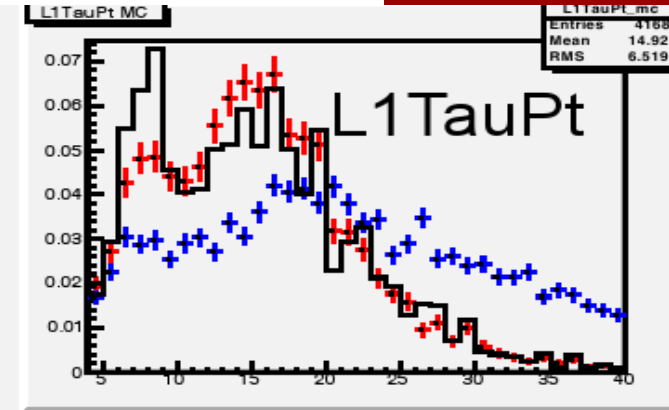
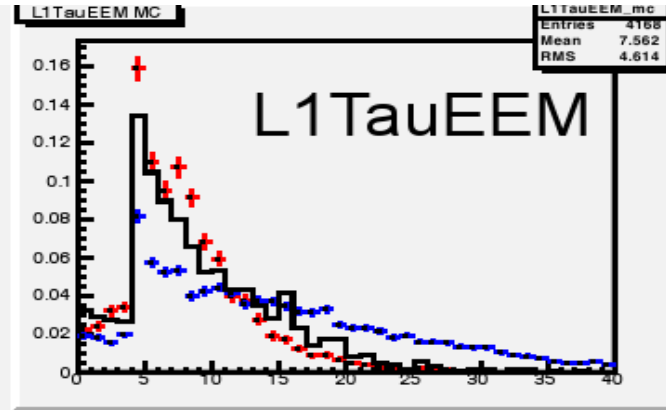
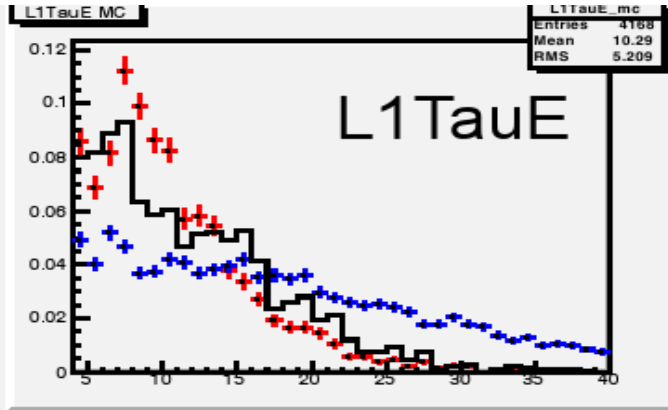


# 1 variable

Reference sample “data”  
To-be-weighted sample “MC”  
Weighted sample



UPPSALA  
UNIVERSITET

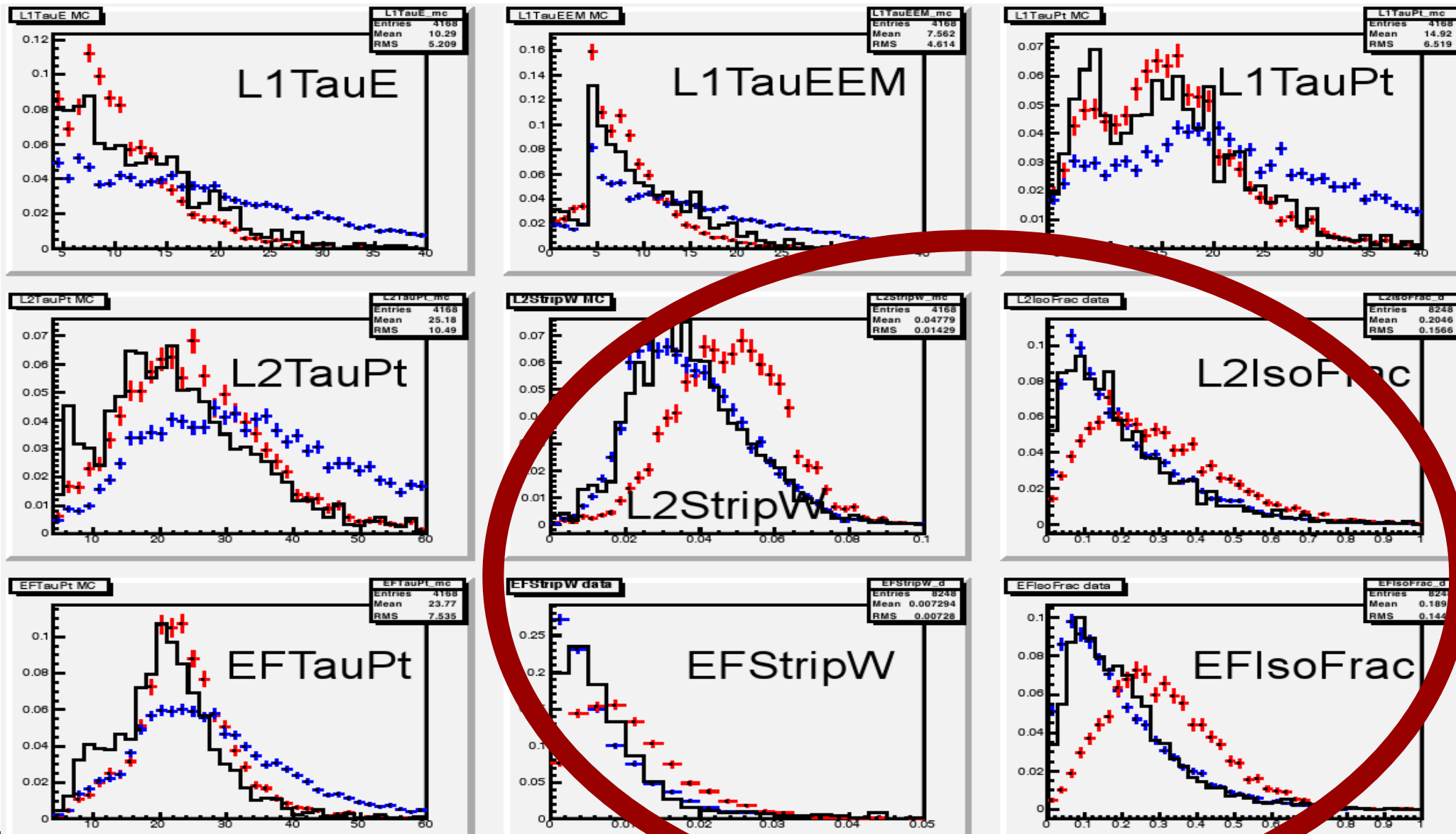


# Simultaneous reweighting with 4 variables



UPPSALA  
UNIVERSITET

Reference sample "data"  
To-be-weighted sample "MC"  
Weighted sample





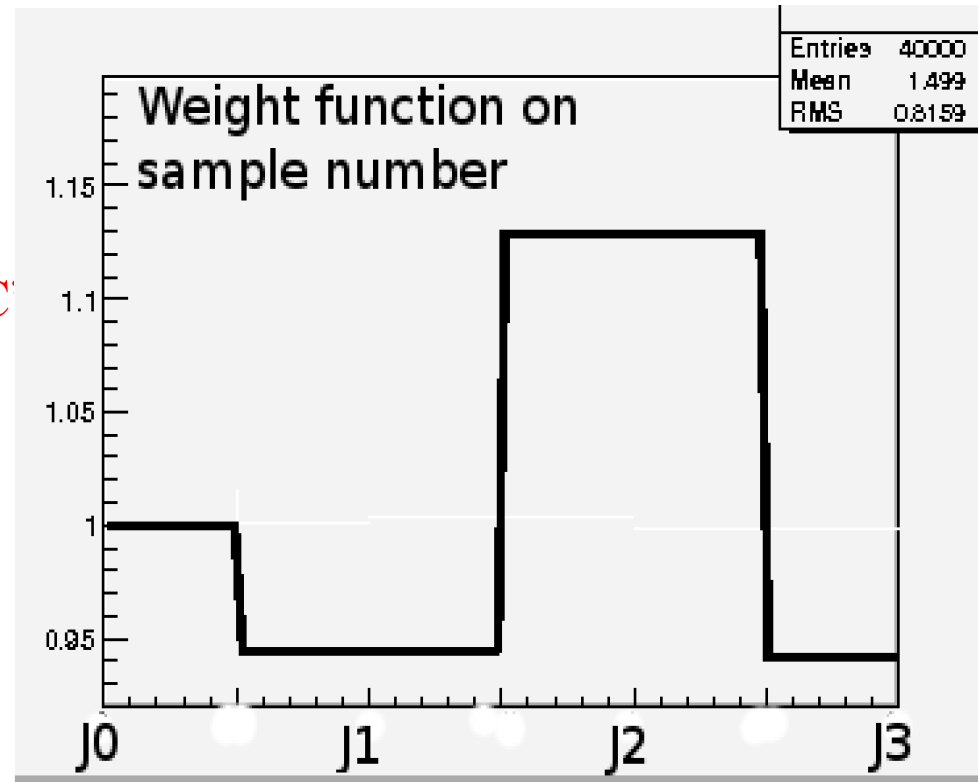
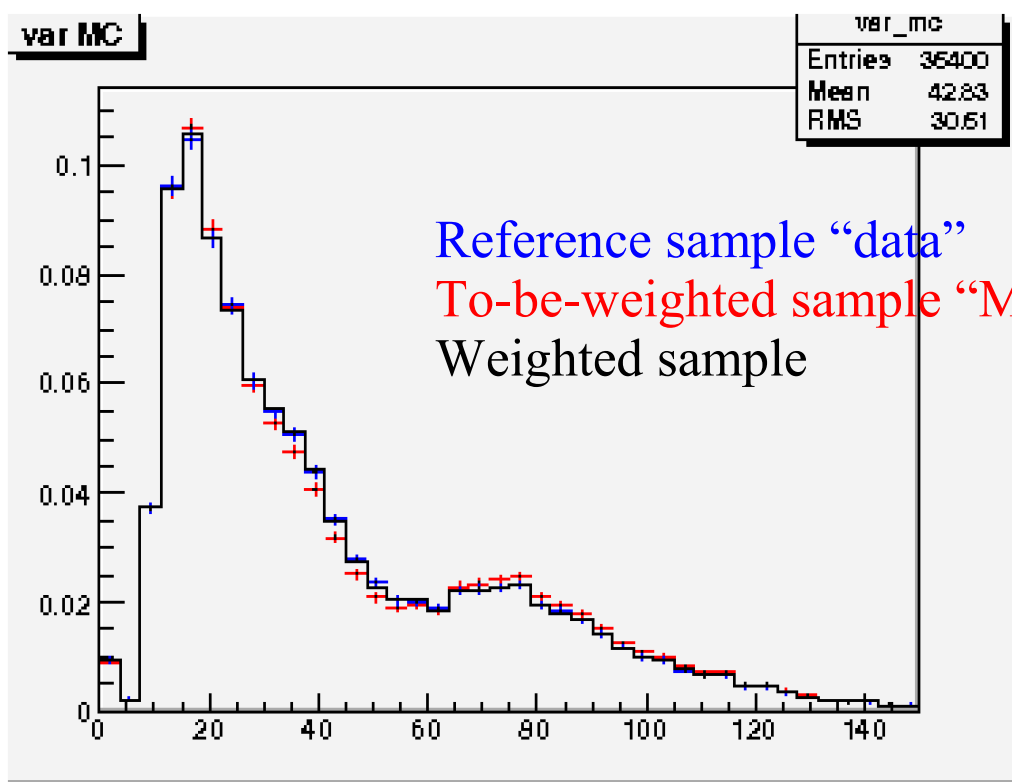
## Example 3: Sample composition

- QCD samples at ATLAS are binned in  $p_T$ .
- Designated as “J0”, “J1”, “J2”, “J3” in increasing momentum.
- Can use values 0, 1, 2, 3 as an event variable in MC sample
- Info not available in data, assign value at random (uniform) to data events
- Determine weight function on the “sample number” variable but tuning on other event characteristics

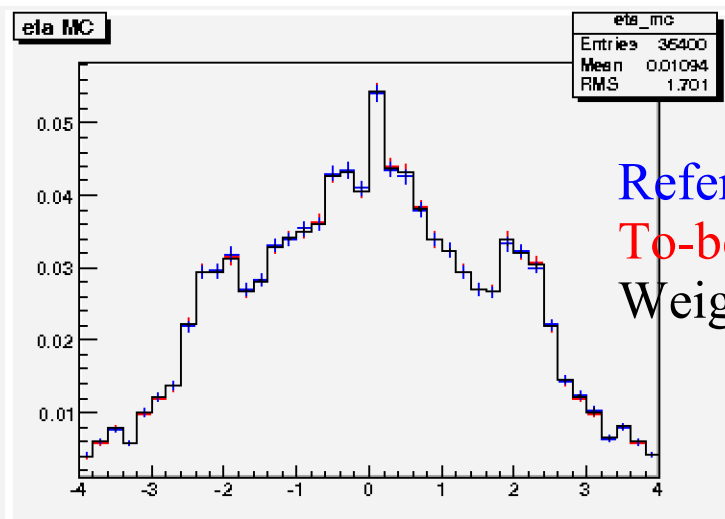
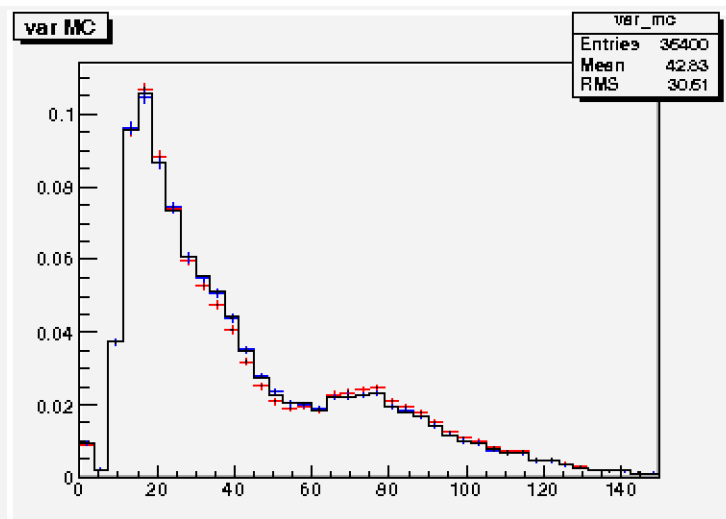


# Example 3

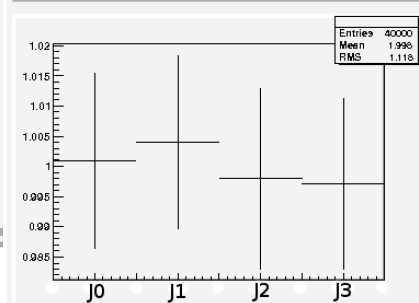
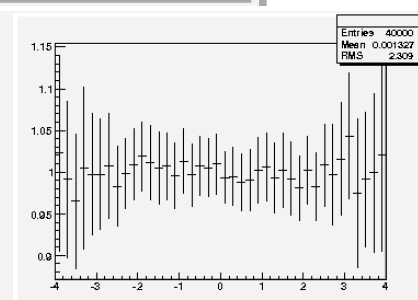
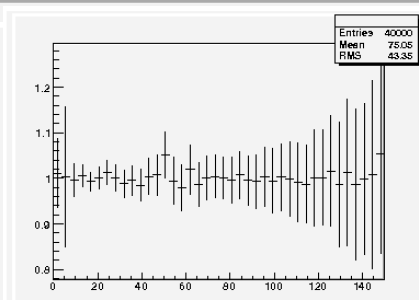
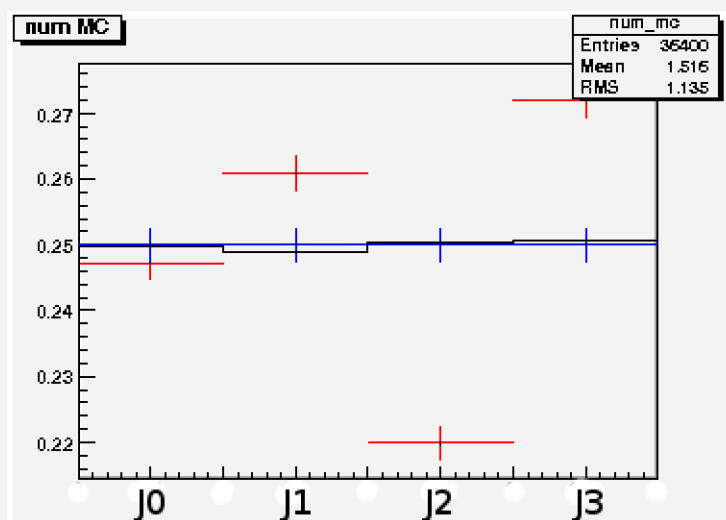
- Weight function on sample number from jet  $p_T$ ,  $\eta$  in 40k QCD events.



# Using $p_T$ and $\eta$ to adjust QCD sample composition



Reference sample "data"  
To-be-weighted sample "MC"  
Weighted sample





# Summary

- A toolkit to match data samples through event-by-event weight is currently under active development.
- Current input via specially prepared ASCII text files, developing interface to TTP and DPD formats.
- A wide range of use cases has been considered.
- Intrigued? Interested? We would love to hear your comments, ideas and feature requests.