Estimate b-tag efficiency from top decays

Towards a new method

Joris Maes

IIHE top quark meeting

6th may 2009

Outline

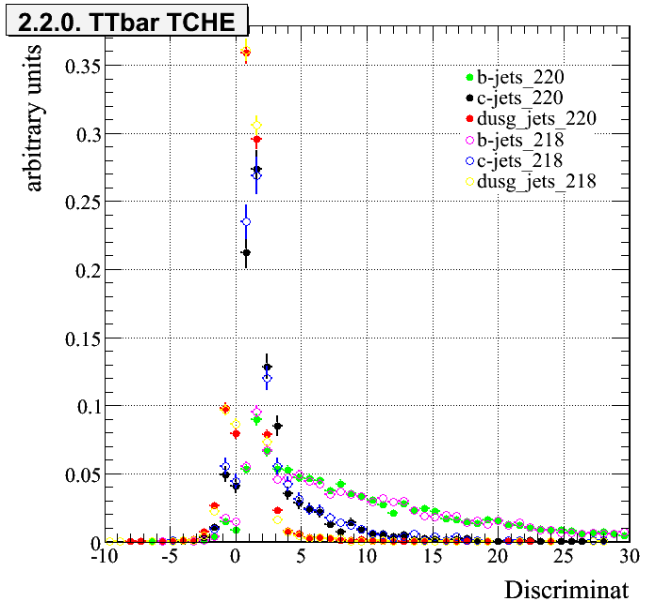
In this talk I will present the ideas and briefly discuss the current status for a new method to estimate the b-tag efficiency.

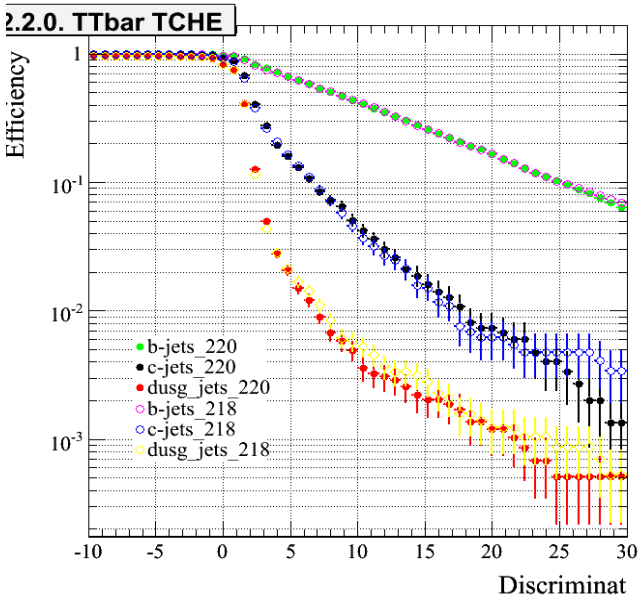
The goal is to find a method to estimate the b tag efficiency in top quark decays with a data driven control of the background contribution

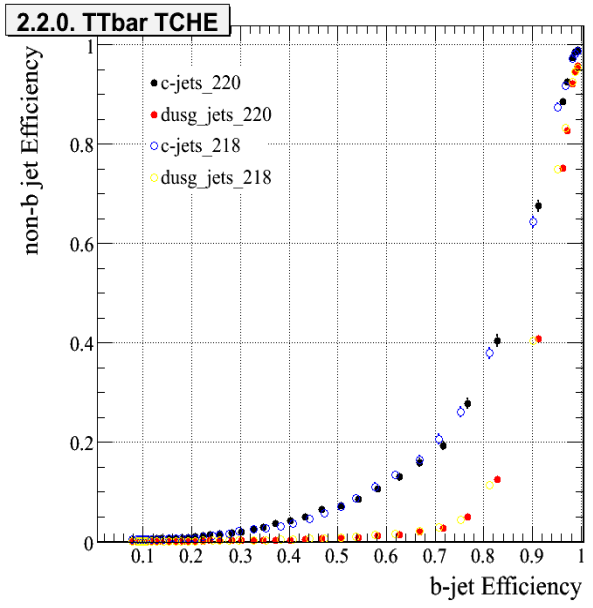
- Aim of the method
- LR method
- 'ATLAS' method
- Towards a new method
- status

Aim of the method

- Aim of the method: reconstruct in a data driven way the distribution of the various b-tag discriminators for pure b-jets
- This by using the ttbar semi-leptonic muon events
- This distribution allows us to estimate the b-tag efficiency for each threshold on the b-tag discriminator







LR method

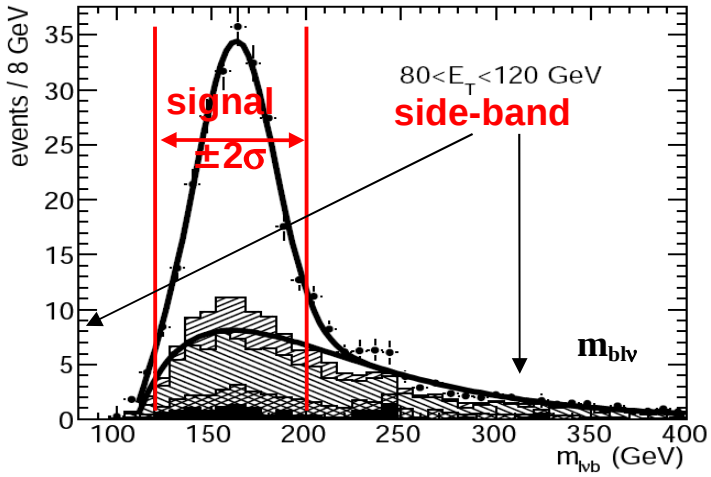
- The first attempt was to extend the Likelihood Ratio (LR) method in (CMS note 2006/013).
- The main idea in this analysis is to use a Likelihood Ratio to obtain a jet sample with a b-jet purity of ~80%
- On this sample one applies a b-tagger and by counting the number of tagged jets one can obtain the b-tag efficiency at a certain threshold (taking into account the impurity of the sample)

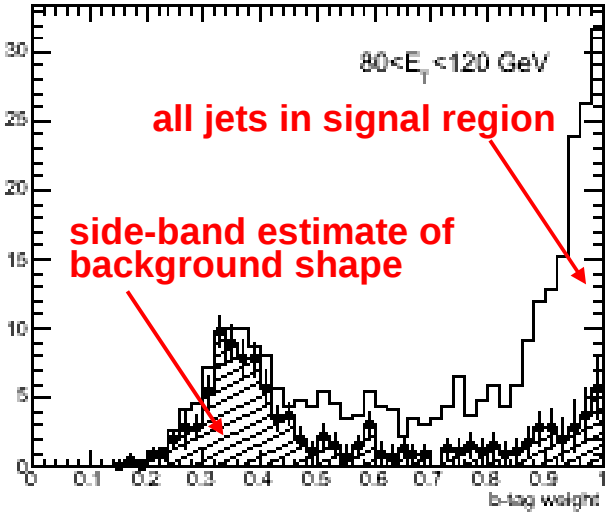
- Problems & shortcomings
  - The method was relying heavily on the Monte Carlo model used, e.g ratio between signal and background (b vs. non b jets)
  - To reduce the influence of background (non b jets) the purity needs to be rather high but this will bias the b-tag measurement

- To reduce the influence of the MC one needs to get control on the background (non b jets) contribution to the b-tag discriminator directly from data.

'ATLAS' method

- The next attempt was to look at the method studied in the ATLAS collaboration
- The *Signal* region is constructed by the 2σ interval around the top mass in the leptonic top distribution and contains both b and non-b jets
- To obtain the b-tag distribution in the signal region one needs to subtract the b-tag distribution for non b jets.
- They obtain the b-tag distribution for non b jets from the side-band
- The scale factor for the amount of non b jets in the *signal* region is obtained by fitting the top mass (see further)

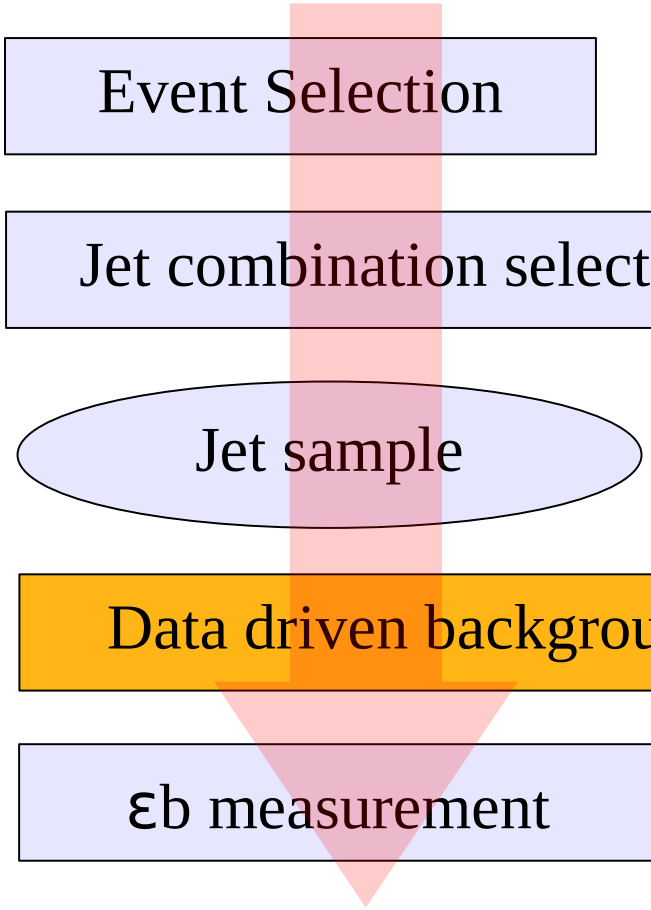




- This only works if the side-band region is completely b-jet free, but this is not the case so one will be correcting the shape for both b and non b jets

Analysis flow

I'm implementing the code in CMSSW_2_2_7 using the Summer08/Fall08 Samples, (+PAT, +TopTree)



Use cuts on lepton and jets to select the ttbar semi-muon decay channel (might need also to tag one jet to suppress W+jets)

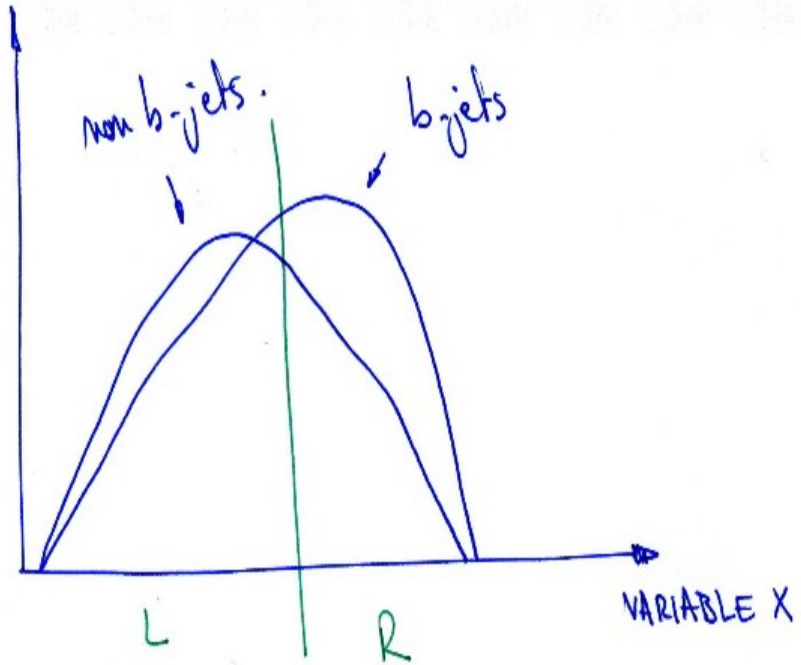
Not yet settled, but should be simple

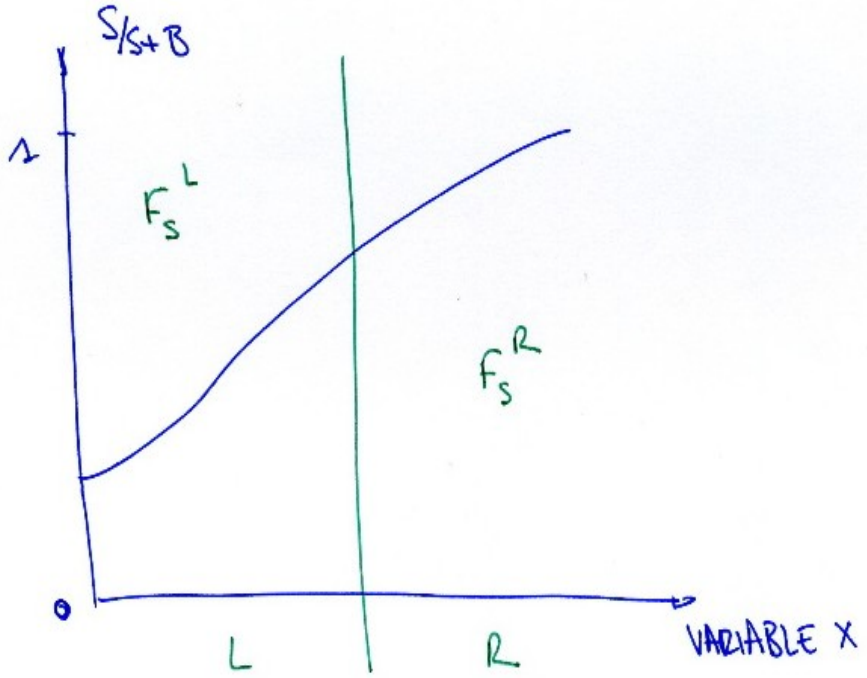
One of the 2 jets assigned as b jets will be selected to form the jet sample with enriched b jet content

Current work

method

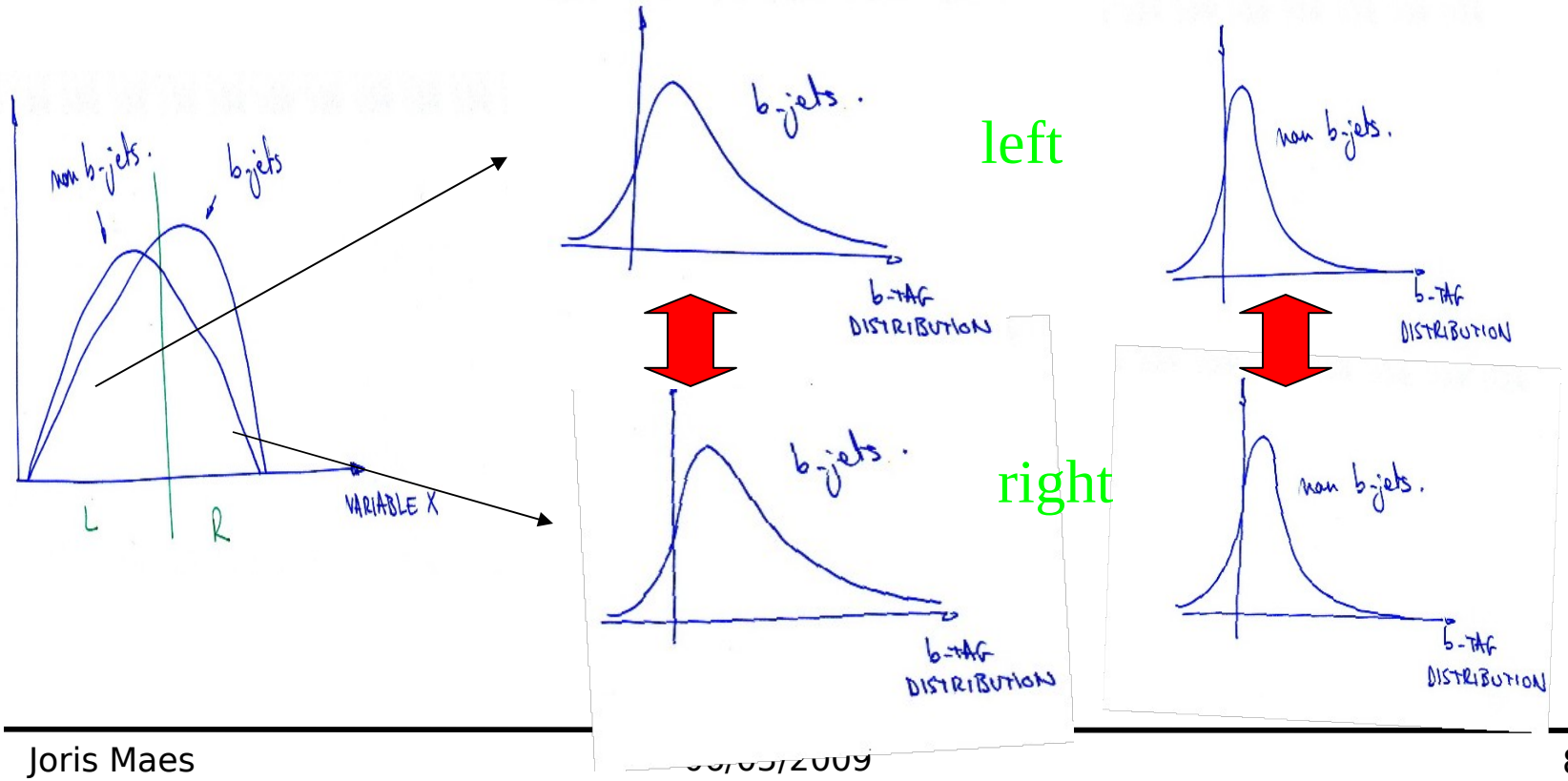
- Once the jet sample is defined, we divide the sample in bins depending on the p_T and eta of the jet, for each bin the b-tag distribution will be derived.
- The main thing we need is a variable X which has a different shape for the events where the selected jet is a b jet w.r.t the shape for a non b jet selected
- We should have a reasonable difference in signal fraction if we divide the variable in 2 regions, left (L) and right (R).





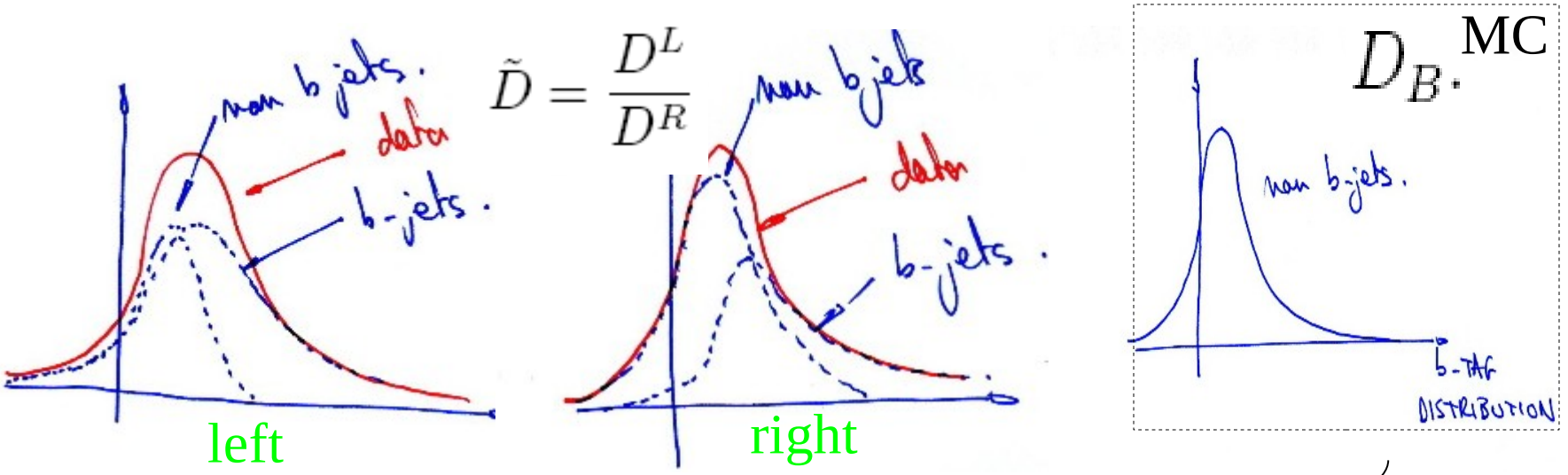
method

- To use the method we would like to have that the shape of the distribution of the b-tag discriminator for both b and non-b jets is respectively the same in the left and right region
- This is needed to safely subtract one from another
- The similarity of the shape will be tested by a chi² test
- This code is almost in place, within a few days I will be able to test a lot of variables (top mass, H_T, angles)



method

- To obtain now the b-tag distribution for pure b-jets we look at the relative difference in contribution to the b discriminant distribution for the left and right region, in a bin by bin way



$$\tilde{D} = \frac{D^L}{D^R}$$

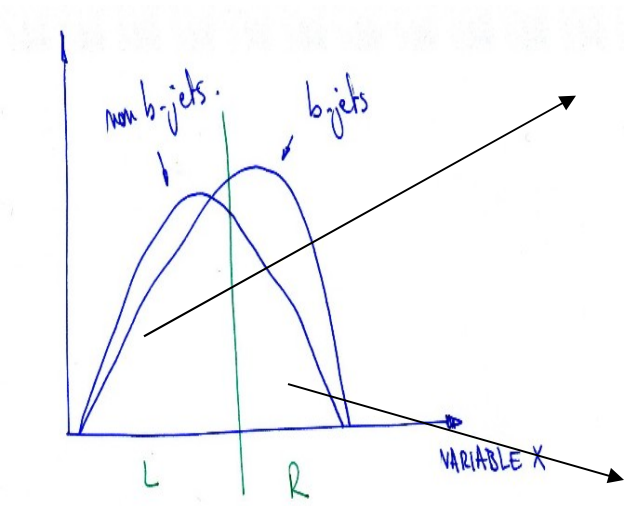
$$D_S = \frac{(1 - F_S^R)N^R\tilde{D} - (1 - F_S^L)N^L}{F_S^L N^L - F_S^R N^R \tilde{D}} D_B.$$

From data

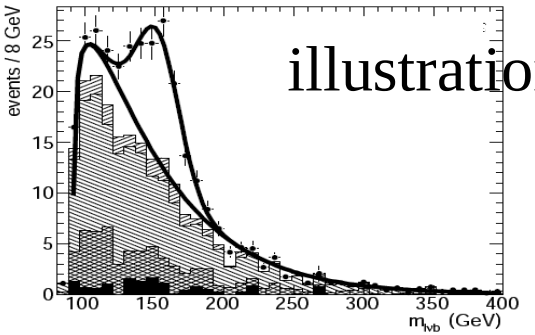
- One needs to estimate the fraction of b jets in the left and right region from data, then the only input from MC would be the b-tag distribution for non b jets

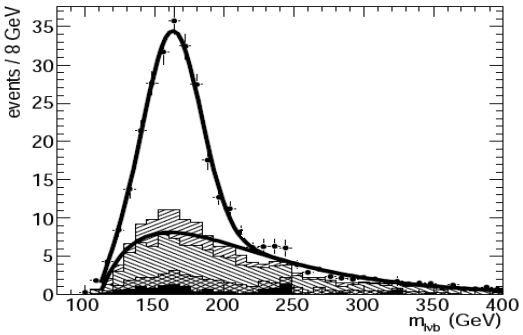
method

- How to estimate the b-jet fraction?
  - In the ATLAS method they fit the leptonic top mass with a gaussian and a function for the background.
  - Can I do it with the same variable? To be tested
  - Another way to estimate the S/B fraction?



illustration





status

- Samples
  - The Summer08/Fall08 samples are available in pat and toptree format
  - They are still lacking the tqaf layer 2 information (e.g. info on the jet combination)
  - No systematic samples are yet in pat format

- Code
  - The code to divide the sample in bins of p_T and eta is in a good shape, minor things are missing (e.g. correct adding up all samples + treating the weights and uncertainties)
  - This class can be easily extended to contain all the needed histograms to study the variables
  - Timescale of having first plots of the chi^2 is 1-2 weeks
  - To proof that the method is working some things like the b to non b fraction estimate are still missing