Combining b-tagging calculations in ATLAS

Abstract

The ATLAS collaboration calibrates its b-tagging algorithms using a number of data-driven methods. The b-tagging algorithms, used to separate b-quark jets from light-flavor jets, are widely used in top, Higgs, and Exotics analyses in ATLAS. The algorithm’s performance is measured for light-flavor jets, charm jets, and bottom jets. In some cases multiple methods can be used to measure the performance; methods using dijet data and a pure sample of top-antiop top are being used to measure the bjet performance. These methods have complementary strengths – the dijet methods tend to have the smallest errors at low jet energy, and the $\ttbar$ methods tend to have smaller errors at high jet energy. We used a minimum likelihood method to fit the results, profiling common systematics, to combine the results. For results on common data we have also understood the statistical correlation of our methods. This paper will describe the details of the combination.

Tools

- RooFit – a fitting package, part of the ROOT distribution, is used for the fitting. Heavy use is made of the RooGaussian object. The Gausians shown below in this poster are then combined using RooProduct and RooAddition. MINUIT is used to do the final fit and find the maximum and the best values of the scale factors. Some tuning was required to the fitting process to make sure there was convergence in as many cases as reasonable.
- Spirit – A parser-combiner library that is part of the boost library distribution. parser-combiner libraries allow you to construct a grammar using C++ method calls and for a small language are much more approachable and usable than a full blown parser like Bison. Used to parse the result files from each of the calibrations. And kept the inputs more readable than XML.
- C++ was heavily used to glue everything together. A better choice, in retrospect, would have been python because a great deal of string manipulation was required.

Calibration Methods

ATLAS is using two primary calibration methods for the efficiency and two for the fake rate. These are described in detail in ATLAS public notes and other posters here at PLHC. The fake rate methods are not described here (see the Problems section below). The efficiency methods are both based on dijet data:

- The $p_T^{rel}$ method – About 15% of bottom quark decays result eventually in a muon. The relative $p_T$ of the muon to the jet axis $(p_T^{rel})$ is a function of the mass of the object the muon decayed from. Bottom quarks (at 5 GeV) are much more massive than other objects that produce muons. Thus this distribution can be used as an independent way of measuring the efficiency. It is dependent on MC templates.
- The $\ttbar$ method – Two uncorrelated taggers, two data samples, and 8 equations with 8 unknowns that relate the data samples; one can solve for the tagging efficiency. The straight forward method needs very little in the way of input from MC, though a thorough understanding of the systematics does. Each measurement comes with a long list of correlated and uncorrelated systematic errors.

Since both methods are based on the same dijet events and similar triggers, we must account for statistical correlation as well as systematic correlations.

Statistical Correlations

The two calibrations methods are statistical correlated because they are based on the same underlying jet sample. In order to combine them a study was done to determine the extent of the statistical correlation.

- In each kinematic bin the overlap in number of jets used in each calibration method was studied. Due to differences in the triggers, in some bins the overlap was close to 100%, and in other bins it was almost non-existent.
- For bins where the overlap was not negligible, a toy Monte Carlo was used to determine the correlation.
- The numbers of events in the two samples in System 8 along with the flavor composition and the b-tagging efficiency were taken for a particular tagger and working point.
- The expected $p_T^{rel}$ distribution was taken for similar samples from MC.
- Using these numbers and distributions, one can now run a toy MC.
- Re-fit both $p_T^{rel}$ and System 8 for the efficiency on each iteration. The scatter plot contains the correlation between the two taggers. The statistical error is then split into a correlated part and uncorrelated component and added to the fit above. The uncorrelated part is the new statistical error ($\delta SF_{stat}^{corr}$) and the correlated part becomes a new systematic error.

Getting the statistical correlations correct with this method is a bit tricky. Our first attempt we mis-estimated how much data was shared in some of the bins resulting in the bins being more correlated than they should have been. This can lead to invalid results when results from the two methods are averaged.

Comments & Future Work

RooFit does not always converge. In cases where the fit did not work we took the best measurement in the bin and used that as the final result. Final results, with complete systematic errors, are written out to a ROOT files in the form of histograms which is used directly by ATLAS analyzers.

Future work is mostly driven by including new measurements in the combination. Specifically those from $\ttbar$. The scale factors from $p_T^{rel}$ and System 8 are most powerful in the low jet $p_T$ bins. We expect $p_T^{rel}$ to be most powerful in the higher bins. We also expect to do further tuning to the algorithms to improve our ability to combine edge cases - like the fake rates.

We will also study the compatibility between the measurements – some regions will have 3 or more measurements.