Distinguishing quark and gluon jets at 13 TeV in the CMS experiment

Posters@LHCC, CERN, Geneva, Switzerland, 22nd February 2017

Giorgia Rauco (UZH), giorgia.rauco@cern.ch
Paolo Azzurri, Christian Contreras Campana, Tom Cornelis, Andrea Carlo Marini, Francesco Pandolfi, John Strologas, Yuta Takahashi,
on behalf of the CMS Collaboration

Abstract

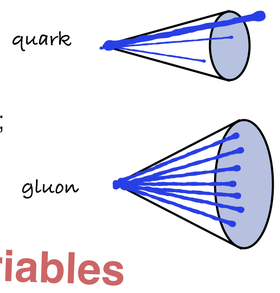
At the LHC, many physics analyses study processes characterized by a signal with jets originating from quarks, while the jets in the background are gluon-like. Based on observables sensitive to fundamental differences in the fragmentation properties of gluons and quarks, a likelihood discriminant is built. The performance of the tagger is evaluated on data using Z+jets and dijet events produced in proton-proton collisions at a centre-of-mass energy of 13 TeV, recorded in 2015 by the CMS experiment at the LHC. A comparison between the discrimination obtained in Pythia8 and Herwig++ generators is also presented.

theoretical background and motivation

jets from light-flavor quarks ≠ jets from gluons

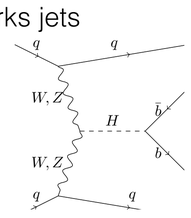
main differences:

- the **particle multiplicity**
- the **fragmentation function**;
- **collimation**.



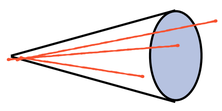
the capability of distinguishing between quark and gluon is **very important at analysis level**: often analysis' signal final states include quarks jets

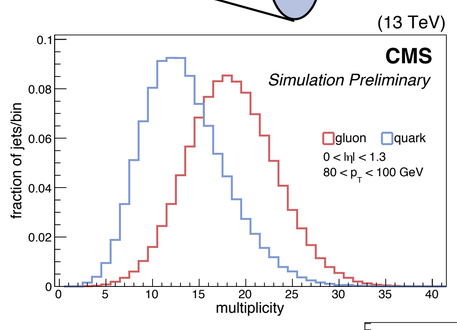
ex. VHF H(bb)

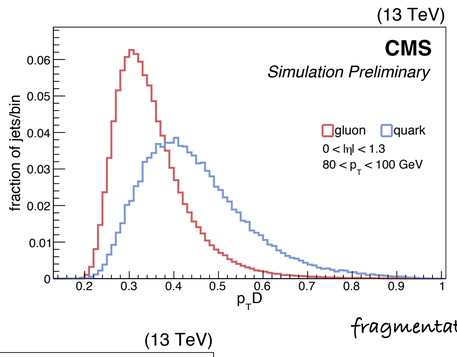


discriminating variables

multiplicity

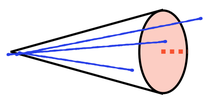


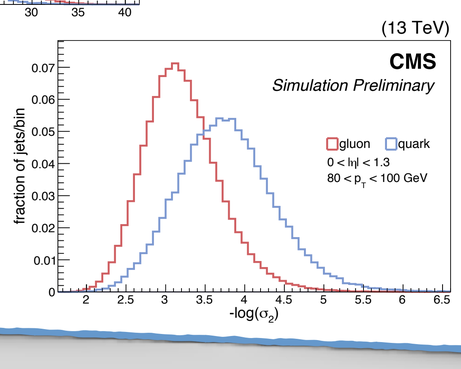




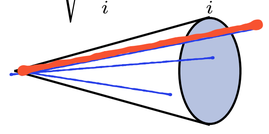
fragmentation (ptD)

ellipse minor axis





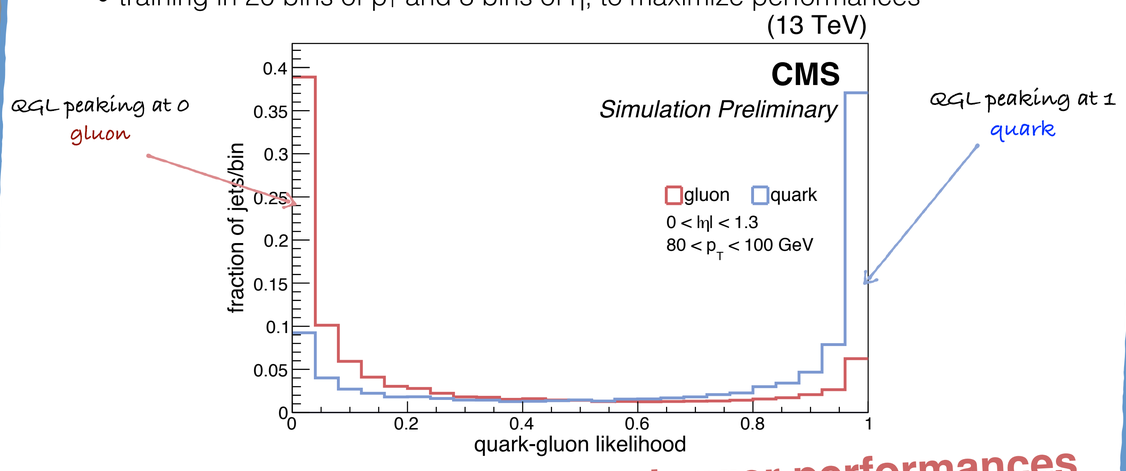
$$p_T D = \sqrt{\sum_i p_{T,i}^2} / \sum_i p_{T,i}$$



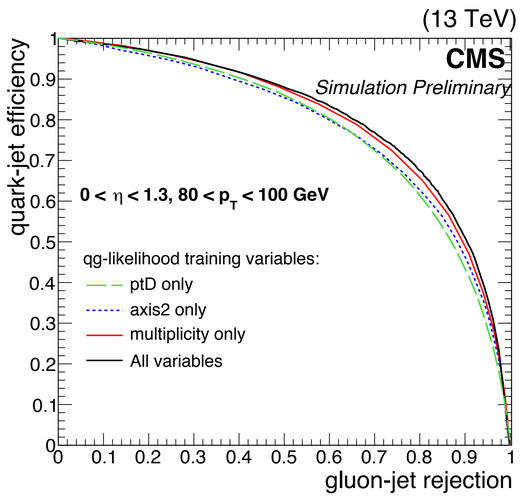
building the discriminator

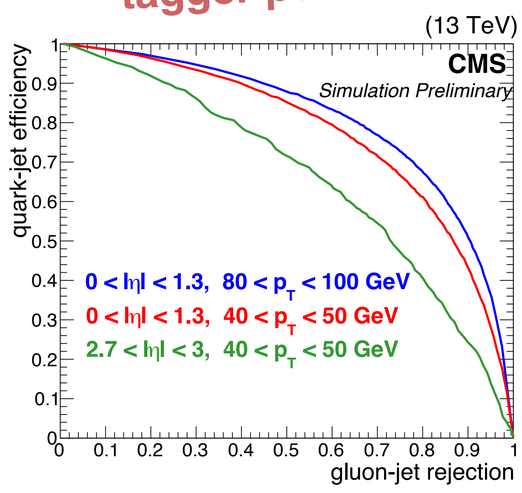
likelihood-discriminator (QGL) built using the variables PDFs from jets in QCD events

- training in 20 bins of p_T and 8 bins of η, to maximize performances



tagger performances



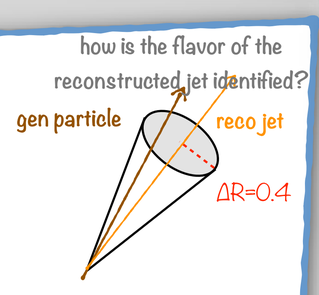


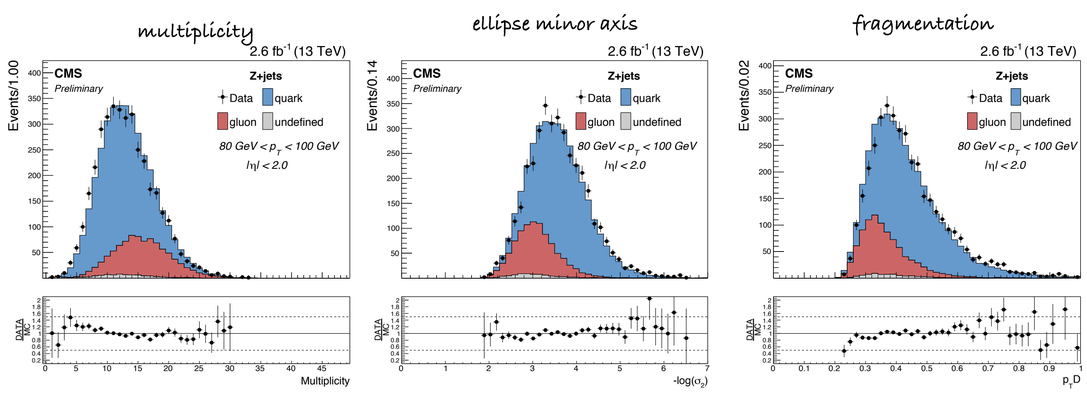
the full 2015 dataset is analyzed

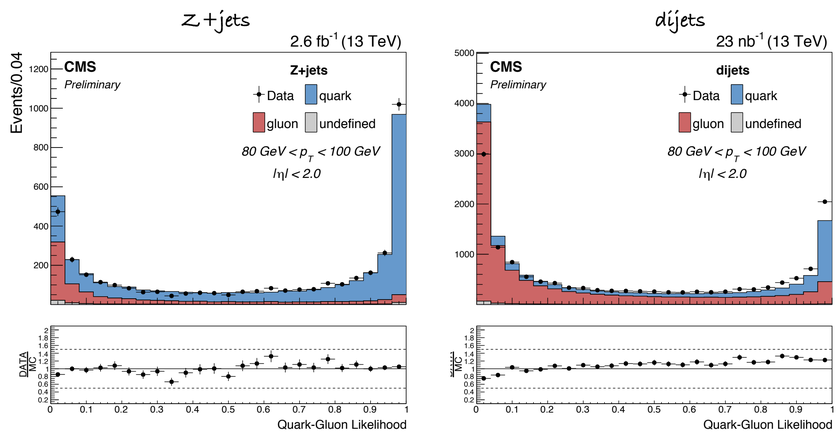
validation in data

tagger correct functioning verified on data in 2 samples:

- **Z+jets events**, quark-rich
- **dijet events**, gluon-rich



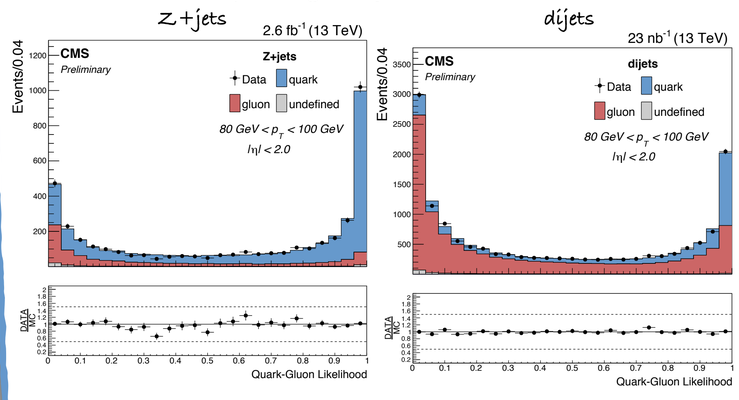




data driven reweighting

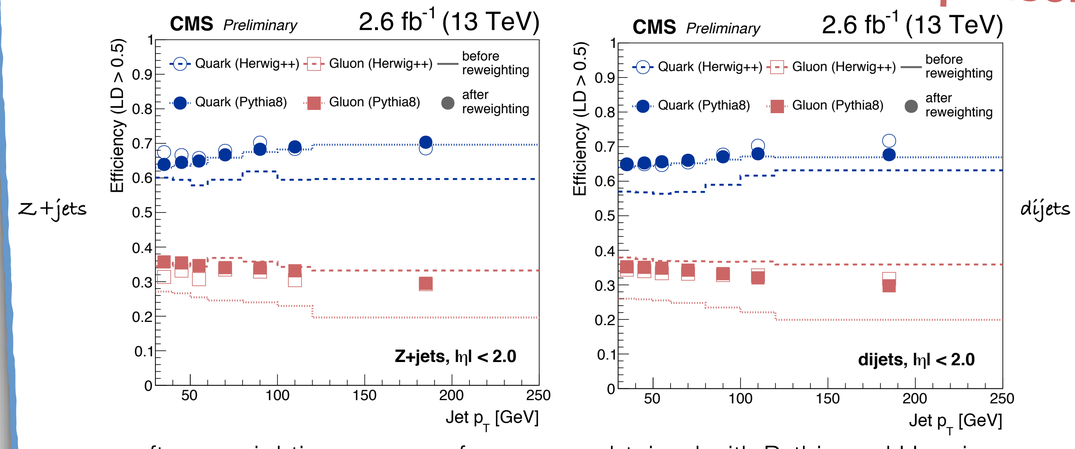
$$N_{data}^{DY} = \alpha_g N_{MC,gluons}^{DY} + \alpha_q N_{MC,quarks}^{DY} + N_{MC,undef}^{DY}$$
$$N_{data}^{QCD} = \alpha_g N_{MC,gluons}^{QCD} + \alpha_q N_{MC,quarks}^{QCD} + N_{MC,undef}^{QCD}$$

to improve the agreement between data and MC, **data-driven weights** are derived for each QGL bin exploiting the yields in data and MC for the two parton flavor components



MC components reshaped to agree with Data

generator comparison



after reweighting same performances obtained with Pythia and Herwig++

Currently 2016 data are being analyzed and validated. In parallel there is an ongoing effort to build a quark-gluon tagger using a Boosted Decision Tree.

conclusions

A discriminant has been developed to separate jets originating from the hadronization of gluons or light-quarks. It is built as the likelihood-product of the PDFs of three input variables and it is trained separately in several kinematic regions. The discriminator input variables and output distributions have been validated in data and, in order to reproduce better the observed data discrimination, Monte Carlo data-driven weights have been derived. Finally, a comparison between Pythia8 and Herwig++ has been performed and it resulted in similar performances after

References

CMS Collaboration, *Performance of quark/gluon discrimination in 13 TeV data*, 2016, **CMS-DP-2016-070**

CMS Collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data, 2013*, **CMS-PAS-JME-13-002**