Status report

July 22, 2022

Jieun Choi

HYU / IP2I

Investigating feature importance

Machine Learning Group

- L3: Knowledge, Production, Innovation
- Meetings are in videoconference

Feature importance in CMS (cms-ml/documentation)

- Newly updated from the last discussion in Knowledge Group (18/07, indico)
- Most feature importance methods fall into one of three broad categories:
  - filter methods, embedding methods, wrapper methods
  - Using SciKit-Learn python package

Recommended techniques in simple DNN ? (CMSTalk)

A1. cms-ml documentation (above)

A2. Taylor expansion (arXiv, ML-JC Review)

- Talyor expansion of output function (up to second order) with respect to the input features
- Have been used it in several Higgs analysis (1,2,3)
- Example is available (jupyter)

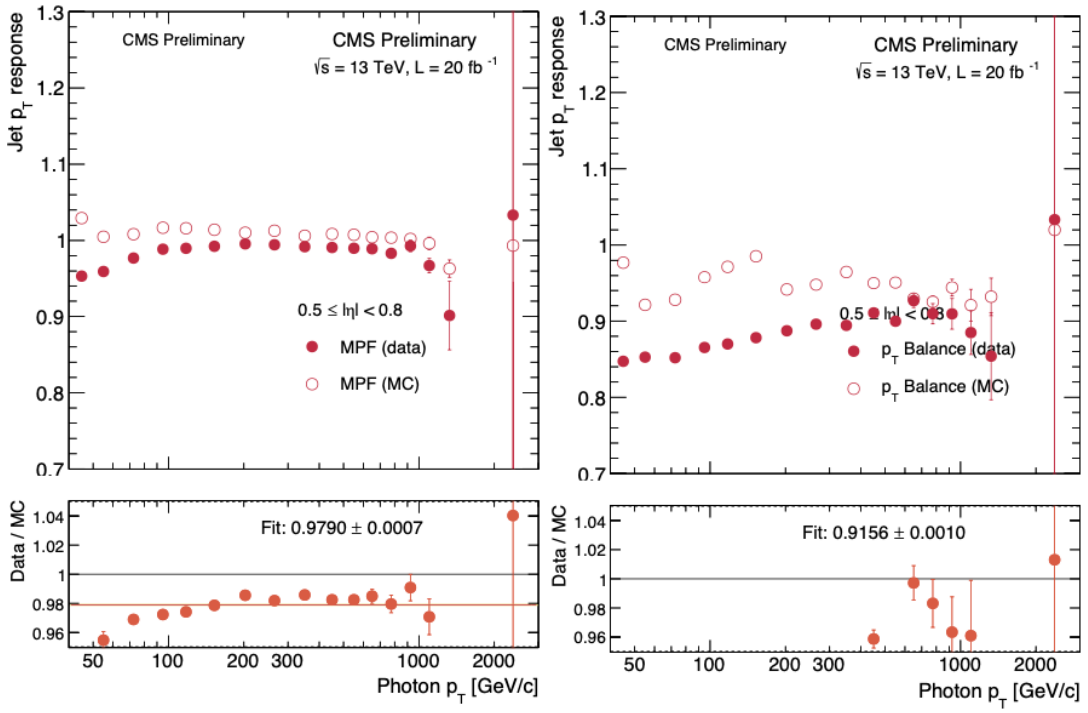
HLT path for JES in gamma+jet

Review

- Deployed on online from Run 355815 (July 18, ~17:00)
- Prescale: 15
- With only stable collisions run (low deadtime)
- 355872 → ~ 13.26 Hz, 338872 Counted (The "Good Run") (PU ~ 32, Fill 8017)
- 355892 → ~ 12.86 Hz, 25490 Counted (PU ~ 32, Fill 8018)
- 355912 → ~ 12.74 Hz, 48913 Counted (PU ~ 32, Fill 8019)
- 355913 → ~ 12.23 Hz, 30005 Counted (PU ~ 32, Fill 8019)
- 355921 → ~ 12.85 Hz, 121087 Counted (PU ~ 32, Fill 8020)
- 355933 → ~ 16.54 Hz, 7329 Counted (PU ~ 32, then 0 Hz at lowPU ~ 1 (prescaled 0), Fill 8022)
- 355942 → ~ 12.05 Hz, 7990 Counted (PU ~ 29, then 0 Hz at lowPU ~ 1 (prescaled 0), Fill 8023)
- 355991 → ~ 20.65 Hz, 6689 Counted (PU ~ 50, then 0 Hz for test (prescaled 0), Fill 8027)
- 356003 → ~ 16.40 Hz, 103664 Counted (without CSC/CTTPS, PU ~ 40, Fill 8027)
- 356004 → ~ 14.96 Hz, 6667 Counted (without CTTPS PU ~ 40, Fill 8027)
- 356005 → ~ 14.99 Hz, 62840 Counted (PU ~ 40, Fill 8027)

- With PU ~ 50 run (for a short time) rate is expected as ~ 92.9 (20.65 * 4.5) Hz

JES in gamma+jet



Alpha < 0.4 Closure
2016 APV
Run BCDEF

Understanding plots (Ref: Closure: G+Jets from Stephanie)…

- Plots with 1) Jet pT response, 2) Alpha value, for photon pT per eta ranges
- Jet pT response = R, response estimator
- Ideal case: converged into 1 for DATA and MC
- R for Balance method:

1)

$$R_{bal} = \frac{p_{T,reco}^{1^{er}jet}}{p_T^{\gamma}}$$

2) $$\alpha = \frac{p_T^{2^{e}jet}}{p_T^{\gamma}}$$

- R for MPF:

$$R_{MPF} = 1 + \frac{\vec{p_T}^{\gamma} \cdot \vec{\not{E}}_T}{|\vec{p_T}^{\gamma}|^2}$$

(Hugues's thesis, CMS-CR-2012-188*)

Gjets sample comparison on DR

Introduction

- A direct photon is defined as one having ΔR > 0.4 with respect to the nearest parton
- QCD and GJets_DR-0p4 MC are stitched together at ΔR=0.4 to coincide with the gen-level cut on the GJets_DR-0p4 sample
- Standard sample has ΔR > 0.05 cut
- Going to check if there is difference in photon / (sub)leading jet pT

Gjet samples for JES study

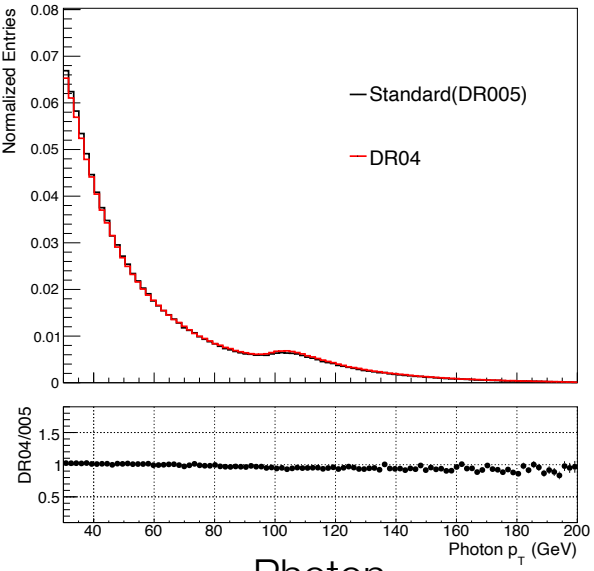
- Standard samples:
  `Gjets_HT-*` (RunIISummer19UL*NanoAODv2, DAS)
- Samples with larger ΔR cut:
  `Gjets_DR-0p4_HT-{100To200, 200To400, 400To600, 600ToInf}_*` (RunIISummer20UL*NanoAODv9, DAS)
- Compared 2018 samples with the same range of HT bins

Selections

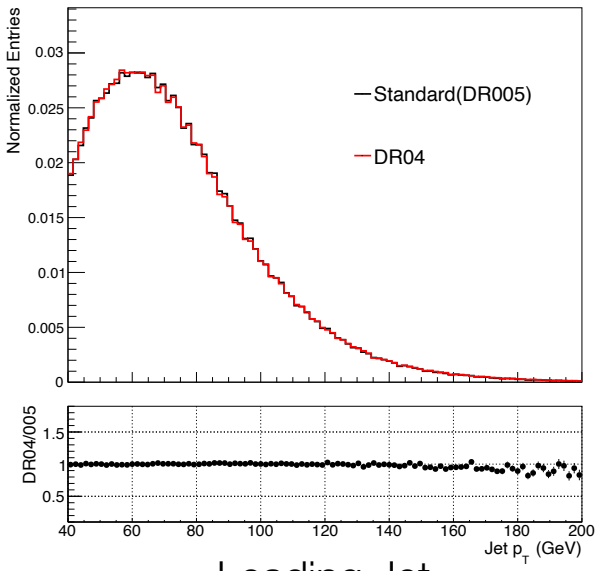
- Naïve selections from HLT_Photon30EB efficiency estimation (from MET study in gamma+jet)
- MET Filters (twiki)
- Photon:
  - == 1 Photon with pT > 30 GeV && |eta| < 1.44, TightID/Iso (CutBased ID (Fall17 V2))
- Jet:
  - >= 1 Jet with pT > 40 GeV && |eta| < 2.5, Tight JetID
- Lepton Veto:
  - Any loose lepton (muon/electron) with pT > 10 GeV

Comparison on pT / eta distribution

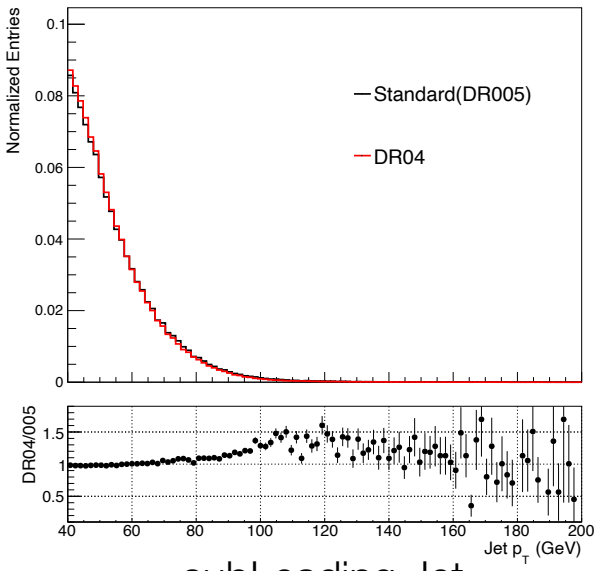
- HT-100To200 samples



Photon



Leading Jet



subLeading Jet