1 Sample Status

Downloading :

- Yu : 2016 preAPV and 2016 postAPV
- Fabio : 2018

1.1 What is not found on DAS

- 2016 preAPV
  - GluGluHToBB_M-125_TuneCP5_13TeV-powheg-pythia8 : no entry on McM
  - GluGluHToGG_M125_13TeV-sherpa : no entry on McM
  - all HH samples
  - ...
- 2016 postAPV
  - ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8 : monitor page
  - all HH samples
  - ...
- 2017
  - GluGluHToBB_M-125_TuneCP5_13TeV-powheg-pythia8 : no entry on McM
  - all HH samples
  - ...
- 2018
  - all HH samples
  - JetHT Run2018D : monitor page
  - ...

What SM Higgs do we need? All production modes and all decay modes?
I will check the availability of SM Higgs samples once we decide what to include...

- ggH, VBF Higgs $\rightarrow b\bar{b}$
- ggH, VBF Higgs $\rightarrow \gamma\gamma$
- ggH, VBF Higgs $\rightarrow \tau\tau$
- ggH, VBF Higgs $\rightarrow WW \rightarrow 1l, 2l$

- ggH, VBF Higgs $\rightarrow ZZ \rightarrow 2l, 4l$
- VH Higgs $\rightarrow nonbb$
- ttH Higgs $\rightarrow bb, nonbb$

2 Selection

2.1 GoldenJSON

- 2016 : /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions16/13TeV/Legacy_2016/Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt
- 2017 : /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions17/13TeV/Legacy_2017/Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt
- 2018 : /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions18/13TeV/Legacy_2018/Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt

2.2 METFilter

- 2016
  - Flag_goodVertices
  - Flag_globalSuperTightHalo2016Filter
  - Flag_HBHENoiseFilter
  - Flag_HBHENoiseIsoFilter
  - EcalDeadCellTriggerPrimitiveFilter
  - Flag_BadPFMuonFilter
  - Flag_BadPFMuonDzFilter
  - Flag_eeBadScFilter
- 2017 and 2018
  - Flag_goodVertices
  - Flag_globalSuperTightHalo2016Filter
  - Flag_HBHENoiseFilter
  - Flag_HBHENoiseIsoFilter
  - Flag_EcalDeadCellTriggerPrimitiveFilter
  - Flag_BadPFMuonFilter
  - Flag_BadPFMuonDzFilter
  - Flag_hfNoisyHitsFilter (optional)
  - Flag_eeBadScFilter
  - Flag_ecalBadCalibFilter

2.3 Trigger

- UL2016 preAPV :
  - HLT_PFHT450_SixJet40_BTagCSV_p056
  - OR HLT_PFHT400_SixJet30_DoubleBTagCSV_p056
  - OR HLT_PFJet450
- UL2016 postAPV :
  - HLT_PFHT450_SixJet40_BTagCSV_p056
  - OR HLT_PFHT400_SixJet30_DoubleBTagCSV_p056
  - OR HLT_PFJet450
- UL2017 : not decided, only possible triggers in nanoAOD
  - HLT_PFHT380_SixJet32_DoubleBTagCSV_p075
  - HLT_PFHT430_SixJet40_BTagCSV_p080
- UL2018 : not decided, only possible trigger in nanoAOD
  - HLT_PFHT380_SixPFJet32_DoublePFBTagCSV_2p2
  - HLT_PFHT380_SixPFJet32_DoublePFBTagDeepCSV_2p2
  - HLT_PFHT430_SixPFJet40_PFBTagCSV_1p5
  - HLT_PFHT430_SixPFJet40_PFBTagDeepCSV_1p5

2.4 Object

2.4.1 Electron

- $Electron\_pt > 10\ GeV$
- $|Electron\_eta| < 2.5$
- Electron_mva17V2Iso_WP90
- $|Electron\_dxy| < 0.05$, $|Electron\_dz| < 0.1$, $|Electron\_sip3d| < 4$
- Electron_convVeto&&Electron_lostHits==0

2.4.2 Muon

- $Muon\_pt < 10\ GeV$
- $|Muon\_eta| < 2.5$
- Muon_mediumId
- $|Muon\_dxy| < 0.05$, $|Muon\_dz| < 0.1$, $|Muon\_sip3d| < 4$

- Muon_miniIsoId>=3 (or Muon_miniPFRelIso_all/Muon_pt<0.1?)

Which isolation should be used? Muon_miniIsoId or Muon_miniPFRelIso_all/Muon_pt<¡0.1. The only difference is the rho value. Muon_miniIsoId uses "fixedGridRhoFastjetCentralNeutral". Muon_miniPFRelIso_all uses "fixedGridRhoFastjetAll". Muon_miniIsoId is strictly tighter than Muon_miniPFRelIso_all. In ttH AN and BSMFramework, "fixedGridRhoFastjetAll" is used.

The muon isolation efficiency is listed below.
tttt
eff of miniIsoId>=3 : 0.496609
eff of miniPFRelIso_all/pt<0.1 : 0.798705
ratio(miniPFRelIso/miniIsoId) : 1.60832
ttbar_2l
eff of miniIsoId>=3 : 0.776747
eff of miniPFRelIso_all/pt<0.1 : 0.935742
ratio(miniPFRelIso/miniIsoId) : 1.20469
ttbar_1l
eff of miniIsoId>=3 : 0.641226
eff of miniPFRelIso_all/pt<0.1 : 0.880348
ratio(miniPFRelIso/miniIsoId) : 1.37291
ttbar_0l
eff of miniIsoId>=3 : 0.0451004
eff of miniPFRelIso_all/pt<0.1 : 0.621495
ratio(miniPFRelIso/miniIsoId) : 13.7802

Compared to Muon_miniIsoId, Muon_miniPFRelIso_all will select more muons. I am not sure what is the impact on the signal and background yield in the categories.

2.4.3 Tau

- $Tau\_pt > 25\ GeV$
- $|Tau\_eta| < 2.3$
- $|Tau\_dz| < 0.2$
- Tau_decayMode!=5&&Tau_decayMode!=6
- Tau_DeepTau2017v2p1VSjet Medium
- Tau_DeepTau2017v2p1VSe VVVLoose
- Tau_DeepTau2017v2p1VSmu VLoose

2.4.4 Jet

- $Jet\_pt > 25\ GeV$

- $|Jet\_eta| < 2.4$
- Jet_jetId>0 (loose is not anymore supported. 2 means tight, 6 means tightLepVeto)

2.4.5 B-jet

Medium Working Point

- 2016 preAPV : Jet_btagDeepFlavB>0.2598
- 2016 postAVP : Jet_btagDeepFlavB>0.2489
- 2017 : Jet_btagDeepFlavB>0.3040
- 2018 : Jet_btagDeepFlavB>0.2783

2.5 Event selection

- 1Tau0L : 1 tight tau, 0 lepton, >= 8 jets, >= 2 b-jets
- 1Tau1L : 1 tight tau, 1 lepton, >= 6 jets, >= 2 b-jets
- 1Tau2L : 1 tight tau, 2 lepton, >= 4 jets, >= 2 b-jets
- 1Tau3L : 1 tight tau, 3 lepton, >= 2 jets, >= 2 b-jets
- 2Tau0L : 2 tight tau, 0 lepton, >= 6 jets, >= 2 b-jets
- 2Tau1L : 2 tight tau, 1 lepton, >= 4 jets, >= 2 b-jets
- 2Tau2L : 2 tight tau, 2 lepton, >= 2 jets, >= 2 b-jets

2.6 BDT optimization

3 Theoretical Uncertainty

3.1 PDF

Please refer to BSM3G_TNT_Maker/src/EventInfoSelector.cc

3.2 Parton Shower

There is a branch called "PSWeight" in NanoAOD. Float_t PS weights (w_var / w_nominal); [0] is ISR=2 FSR=1; [1] is ISR=1 FSR=2[2] is ISR=0.5 FSR=1; [3] is ISR=1 FSR=0.5;

3.3 Renormalization and Factorization aka Scale Uncertainty

There is a branch called "LHEScaleWeight" in NanoAOD. Float_t LHE scale variation weights (w_var / w_nominal); [0] is MUF="0.5" MUR="0.5"; [1] is MUF="1.0" MUR="0.5"; [2] is MUF="2.0" MUR="0.5"; [3] is MUF="0.5" MUR="1.0"; [4] is MUF="1.0" MUR="1.0"; [5] is MUF="2.0" MUR="1.0"; [6] is MUF="0.5" MUR="2.0"; [7] is MUF="1.0" MUR="2.0"; [8] is MUF="2.0" MUR="2.0". We should take the envelope as the uncertainty.

4 Experimental Uncertainty

4.1 Prefire

There are branches "L1PreFiringWeight_Dn,L1PreFiringWeight_Nom,L1PreFiringWeight_Up".

4.2 Pile-up

The official PU profile in data can be found in

- 2018 : /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions18/13TeV/PileUp/UltraLegacy/
- 2017 : /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions17/13TeV/PileUp/UltraLegacy/
- 2016 : /afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions16/13TeV/PileUp/UltraLegacy/

The MC scenarios can be found here

- 2016 : https://github.com/cms-sw/cmssw/blob/master/SimGeneral/MixingModule/python/mix_2016_25ns_UltraLegacy_PoissonOOTPU_cfi.py
- 2017 : https://github.com/cms-sw/cmssw/blob/master/SimGeneral/MixingModule/python/mix_2017_25ns_UltraLegacy_PoissonOOTPU_cfi.py
- 2018 : https://github.com/cms-sw/cmssw/blob/master/SimGeneral/MixingModule/python/mix_2018_25ns_UltraLegacy_PoissonOOTPU_cfi.py

We should take the ratio of data/mc after normalization and apply the ratio on "Pileup_nTrueInt" in MC. The central, up and down value should be from PileupHistogram-goldenJSON-13tev-*-69200ub-99bins.root, PileupHistogram-goldenJSON-13tev-*-72400ub-99bins.root and PileupHistogram-goldenJSON-13tev-*-66000ub-99bins.root.

4.3 Trigger

Provided by Fabio.

4.4 Electron

https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018
We do not apply inefficiency scale factor for lepton veto.

4.4.1 RECO

egammaEffi_ptBelow20.txt_EGM2D_UL*.root and egammaEffi_ptAbove20.txt_EGM2D_UL*.root are used. The last pt bin is [100, 500]. If pt > 500 GeV, the SF from last pt bin should be used. There are 6 uncertainty sources. I add the variation from 6 histograms in quadrature and treat it as the total uncertainty of electron reco SF.

- 2016 preAPV :
  - egammaEffi_ptAbove20.txt_EGM2D_UL2016preVFP.root
  - egammaEffi_ptBelow20.txt_EGM2D_UL2016preVFP.root
- 2016 postAPV :
  - egammaEffi_ptAbove20.txt_EGM2D_UL2016postVFP.root
  - egammaEffi_ptBelow20.txt_EGM2D_UL2016postVFP.root
- 2017 :
  - egammaEffi_ptAbove20.txt_EGM2D_UL2017.root
  - egammaEffi_ptBelow20.txt_EGM2D_UL2017.root
- 2018 :
  - egammaEffi_ptAbove20.txt_EGM2D_UL2018.root
  - egammaEffi_ptBelow20.txt_EGM2D_UL2018.root

4.4.2 ID

We are using mav17FallV2Iso_WP90. The treatment of pt bins and uncertainty is same as reco SF.

- 2016 preAPV : egammaEffi.txt_Ele_wp90iso_preVFP_EGM2D.root
- 2016 postAPV : egammaEffi.txt_Ele_wp90iso_postVFP_EGM2D.root
- 2017 : egammaEffi.txt_EGM2D_MVA90iso_UL17.root
- 2018 : egammaEffi.txt_Ele_wp90iso_EGM2D.root

4.4.3 ISO

We do not apply any other isolation cut than the isolated ID WP.

4.5 Muon

See

- https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2016
- https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2017
- https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018

4.5.1 RECO

The RECO SF is derived from Jpsi sample for low pt muon ($3 < pT < 40$ GeV). For higher momentum muon, it is suggeste to use the json file from low pt muon.

- 2016 preAPV : Efficiency_muon_generalTracks_Run2016preVFP_UL_trackerMuon.json
- 2016 postAPV : Efficiency_muon_generalTracks_Run2016postVFP_UL_trackerMuon.json
- 2017 : Efficiency_muon_generalTracks_Run2017_UL_trackerMuon.json
- 2018 : Efficiency_muon_generalTracks_Run2018_UL_trackerMuon.json

4.5.2 ID

The ID SF could be derived from Jpsi sample for low pt muon ($3 < pt < 40$ GeV) or from Z sample for medium pt muon ($15 < pt < 120$ GeV). Our selection is $pT > 10$ GeV and medium ID.

For $10 < pt < 15$, we have two options:

- use SF from Jpsi sample.
- use SF from Z sample in the bin pt in [15, 20] GeV.

I would prefer the latter one.

- 2016 preAPV : Efficiencies_muon_generalTracks_Z_Run2016_UL_HIPM_ID.root
- 2016 postAPV : Efficiencies_muon_generalTracks_Z_Run2016_UL_ID.root
- 2017 : Efficiencies_muon_generalTracks_Z_Run2017_UL_ID.root
- 2018 : Efficiencies_muon_generalTracks_Z_Run2018_UL_ID.root

4.5.3 ISO

We agree to use mini isolation, but miniIsoId or miniPFRelIso_all? They have different efficiencies and we need to compute the scale factors by ourselves.

4.6 Hadronic Tau

See https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendationForRun2

4.6.1 Correction on genuin tau

- Efficiency correction on DeepTau2017v2p1VSjet discriminator
  - We are using medium VSjet discriminator.
  - pT dependent and decay mode dependent SFs are available.
  - The uncertainty should be treated uncorrelated across years and pT bins.
  - Since we are using very loose VSe and VSmu working point, we need to assign additional3% uncertainty to tau with pt<100 GeV and 15% uncertainty to tau with pt>100 GeV.
  - 2016 preAPV : TauID_SF_pt_DeepTau2017v2p1VSjet_UL2016_preVFP.root
  - 2016 postAPV : TauID_SF_pt_DeepTau2017v2p1VSjet_UL2016_postVFP.root
  - 2017 : TauID_SF_pt_DeepTau2017v2p1VSjet_UL2017.root
  - 2018 : TauID_SF_pt_DeepTau2017v2p1VSjet_UL2018.root
- Energy scale correction
  - The decay mode dependent SF is provided.
  - The uncertainty of each decay mode is pt-binning
    * pt<34 GeV : taken from low pt measurement
    * $34 < pt < 170$ GeV : interpolation of low pt measurement and high pt measurement
    * pt>170 GeV : taken from high pt measurement
  - In UL, pt pt measurement is performed. The high pt measurement from ReReco is recommended now.
  - Low pt measurement
    * 2016 preAPV : TauES_dm_DeepTau2017v2p1VSjet_UL2016_preVFP.root
    * 2016 postAPV : TauES_dm_DeepTau2017v2p1VSjet_UL2016_postVFP.root
    * 2017 : TauES_dm_DeepTau2017v2p1VSjet_UL2017.root
    * 2018 : TauES_dm_DeepTau2018v2p1VSjet_UL2018.root
  - High pt measurement
    * 2016 : TauES_dm_DeepTau2017v2p1VSjet_2016Legacy_ptgt100.root
    * 2017 : TauES_dm_DeepTau2017v2p1VSjet_2017Legacy_ptgt100.root
    * 2018 : TauES_dm_DeepTau2017v2p1VSjet_2018Legacy_ptgt100.root

4.6.2 Correcton on fake tau from electron

- Efficiency correction on DeepTau2017v2p1VSe discriminator
  - We are using VVVLoose working point.
  - The eta dependent SFs are provided.

sample version	global tag	JEC version	JER
Summer20UL16NanoAODAPVv9	106X_mcRun2_asymptotic_preVFP_v11	Summer19UL16APV_V7_MC_AK4PFchs	JR_Summer20UL16APV_JRV3_MC
Summer20UL16NanoAODv9	106X_mcRun2_asymptotic_v17	Summer19UL16_V7_MC_AK4PFchs	JR_Summer20UL16_JRV3_MC
Summer20UL17NanoAODv9	106X_mc2017_realistic_v9	Summer19UL17_V5_MC_AK4PFchs	-
Summer20UL18NanoAODv9	106X_upgrade2018_realistic_v16_L1v1	Summer19UL18_V5_MC_AK4PFchs	JR_Summer19UL18_JRV2_MC

  - We can use the VVLoose SFs and assign additional 3% uncertainty.
  - The SFs are uncorrelated across years and eta bins.
    * 2016 preAPV : TauID_SF_eta_DeepTau2017v2p1VSe_UL2016_preVFP.root
    * 2016 postAPV : TauID_SF_eta_DeepTau2017v2p1VSe_UL2016_postVFP.root
    * 2017 : TauID_SF_eta_DeepTau2017v2p1VSe_UL2017.root
    * 2018 : TauID_SF_eta_DeepTau2018v2p1VSe_UL2018.root

- Energy scale correction
  - The decay mode dependent SFs are available from ReReco dataset
  - No significance changes are observed between ReReco and UL. It is recommended to use ReReco SFs.
  - The scale factor should be use for tau pt>20 GeV.
    * 2016 : TauFES_eta-dm_DeepTau2017v2p1VSe_2016Legacy.root
    * 2017 : TauFES_eta-dm_DeepTau2017v2p1VSe_2017Legacy.root
    * 2018 : TauFES_eta-dm_DeepTau2017v2p1VSe_2018Legacy.root

4.6.3 Correction on fake tau from muon

- Efficiency correction on DeepTau2017v2p1VSmu discrminator
  - eta dependent SF and uncertainty. The uncertainty should be treated uncorrelated across years and eta bins.
  - 2016 preAPV : TauID_SF_eta_DeepTau2017v2p1VSmu_UL2016_preVFP.root
  - 2016 postAPV : TauID_SF_eta_DeepTau2017v2p1VSmu_UL2016_postVFP.root
  - 2017 : TauID_SF_eta_DeepTau2017v2p1VSmu_UL2017.root
  - 2018 : TauID_SF_eta_DeepTau2017v2p1VSmu_UL2018.root
- Energy scale correction
  - no need to correct
  - assign a 1% uncertainty

4.7 Jet energy scale and resolution

The GlobalTag and the corresponding JEC version used in NanoAODv9 are listed below. The global tag is from McM page and the JEC and JER versions are from conddb.

4.8 Jet energy resolution

The recommendation is from https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution. Check whether a truth jet is matched with the reco jet by

$$\Delta R < R_{cone}/2, |p_T^{reco} - p_T^{truth}| < 3\sigma_{JER} \times p_T^{reco}$$

R_{cone} is the size of jet e.g. 0.4 for AK4Jet. σ_{JER} is the relative momentum resolution measured in simulation.

If the reco jet is matched to a truth jet, the four-momentum of the reco jet should be corrected by

$$c_{JER} = 1 + (s_{JER} - 1)\frac{p_T^{reco} - p_T^{truth}}{p_T^{reco}}$$

s_{JER} is the data-to-simulation core resolution scale factor. c_{JER} is truncated at zero.

If the reco jet is not matched to any truth jet, the four-momentum of the reco jet should corrected by

$$c_{JER} = 1 + \mathcal{N}(0, \sigma_{JER})\sqrt{max(s_{JER}^2, 0)}$$

The s_{JER} and σ_{JER} are documented in https://github.com/cms-jet/JRDatabase/tree/master/textFiles.

- 2016 preAPV :
  - PtResolution : Summer20UL16APV_JRV3_MC_PtResolution_AK4PFchs.txt
  - SF : Summer20UL16APV_JRV3_MC_SF_AK4PFchs.txt
- 2016 postAPV :
  - PtResolution : Summer20UL16_JRV3_MC_PtResolution_AK4PFchs.txt
  - SF : Summer20UL16_JRV3_MC_SF_AK4PFchs.txt
- 2017
  - PtResolution : Summer19UL17_JRV3_MC_PtResolution_AK4PFchs.txt
  - SF : Summer19UL17_JRV3_MC_SF_AK4PFchs.txt
- 2018
  - PtResolution : Summer19UL18_JRV2_MC_PtResolution_AK4PFchs.txt
  - SF : Summer19UL18_JRV2_MC_SF_AK4PFchs.txt

4.8.1 Jet energy scale

See this code https://github.com/miquork/jecsys/blob/master/test/drawUncertainty.C. I save a pt-eta dependent "TF2" function for the SFs, when input a proper txt file.

- UL2016 preAPV : Summer19UL16APV_V7_MC_UncertaintySources_AK4PFchs.txt
- UL2016 postAPV : Summer19UL16_V7_MC_UncertaintySources_AK4PFchs.txt
- UL2017 : Summer19UL17_V5_MC_UncertaintySources_AK4PFchs.txt
- UL2018 : Summer19UL18_V5_MC_UncertaintySources_AK4PFchs.txt

4.9 B-tagging

It seems we need two kinds of scale factors.

- Fixed working poing scale factor : since we are cut on number of medium b-jets
- b score shape correction : since we want to use b score in MVA

See :

- https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL16preVFP
- https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL16postVFP
- https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL17
- https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation106XUL18

4.9.1 FixedWP scale factor

We should follow this page BTagSFMethods.

We need to compute the btag efficiency for b-jet/c-jet/light jet by ourselves. Then

- pT-dependent SF for tagged jet by a specific WP.
- pT-dependent inefficiency SF for un-tagged jet $(1 - SF * \epsilon)/(1 - \epsilon)$, ϵ is the b-tagging efficiency in the given pT bin for a given truth jet flavour.
- 2016 preAPV : wp_deepJet_106XUL16preVFP_v2.csv
- 2016 postAPV : wp_deepJet_106XUL16postVFP_v3.csv
- 2017 : wp_deepJet_106XUL17_v3.csv
- 2018 : wp_deepJet_106XUL18_v2.csv

The uncertainties are splitted into two categories

- up_correlated, down_correlated : correlated across years
- up_uncorrelated, down_correlated : uncorrelated across years

4.9.2 BShapeCalibration

- 2016 preAPV : reshaping_deepJet_106XUL16preVFP_v2.csv
- 2016 postAPV : reshaping_deepJet_106XUL16postVFP_v3.csv
- 2017 : reshaping_deepJet_106XUL17_v3.csv
- 2018 : reshaping_deepJet_106XUL18_v2.csv