Higgs DPS plots

Contacts: cms-phys-conveners-hig@cern.ch
Twiki page: https://twiki.cern.ch/twiki/bin/view/CMS/Dec15Jamboree
Ref: DPS-XX-YYY
Date: 11.12.2015
$H \rightarrow \gamma \gamma$
Trigger efficiency for the L1 electromagnetic objects (red $E_T > 10$ GeV, black $E_T > 22$ GeV) used to seed the High Level Trigger of the $H \rightarrow \gamma \gamma$ analysis. The efficiency is computed as a function of the photon $p_T$ using the tag-and-probe method on electrons from the decay of Z bosons, and obtained with respect to analysis pre-selections ($p_T > 30$ (20) GeV on the leading (sub-leading) $p_T$ photon and loose requirements on the candidate shower shapes and isolation).
High Level Trigger efficiency for the filters used in the H→γγ analysis for events passing the L1-trigger requirements of at least one L1 single EG with E_T > 10 GeV and at least one L1 single EG with E_T > 20 GeV. The leading p_T (black) photon trigger selection has a p_T threshold at 30 GeV, while the sub-leading (red) has one at 18 GeV. The H→γγ analysis employs mass-dependent scaling p_T thresholds equal to m_γγ/3 and m_γγ/4 for the leading and sub-leading p_T photon respectively. The efficiency is computed as a function of the photon p_T using the tag-and-probe method on electrons from the decay of Z bosons and is obtained with respect to analysis preselections (p_T > 30 (20) GeV on the leading(sub-leading) p_T photon and loose requirements on the candidate shower shapes and isolation).
High Level Trigger efficiency for the filters used in the $H\rightarrow\gamma\gamma$ analysis for events passing the L1-trigger requirements of at least one L1 single EG with $E_T > 10$ GeV and at least one L1 single EG with $E_T > 20$ GeV. The leading $p_T$ photon (black) trigger selection has a $p_T$ threshold at 30 GeV, while the sub-leading (red) has one at 18 GeV. The $H\rightarrow\gamma\gamma$ analysis employs mass-dependent scaling $p_T$ thresholds equal to $m_{\gamma\gamma}/3$ and $m_{\gamma\gamma}/4$ for the leading and sub-leading $p_T$ photon respectively. The efficiency is computed as a function of the number of reconstructed vertices using the tag-and-probe method on electrons from the decay of Z bosons and is obtained with respect to analysis pre-selections ($p_T > 30$ (20) GeV on the leading(subleading) $p_T$ photon and loose requirements on the candidate shower shapes and isolation).
Vertex identification

Fraction of $Z \rightarrow \mu\mu$ events for which the vertex has been correctly assigned by the $H \rightarrow \gamma\gamma$ identification algorithm, as a function of the transverse momentum of the dimuon system for data (red) and MC simulated events (black) at 13 TeV and their ratio. The muon tracks have been ignored during the vertex reconstruction process to mimic a diphoton system. The vertex is considered as correctly identified if its z coordinate lies within 10 mm of the z coordinate of the vertex tagged by the muon tracks. The simulated events are weighted such that the distribution of the per event vertex multiplicity matches the one in data.
Internal notes: vertex identification

Simulations:
Drell-Yan sample MadGraph + Pythia 8

Data:
- events selected with a double muon trigger

H→γγ vertex BDT inputs:
- squared transverse momenta of the charged particle tracks associated with the vertex,
- and two variables that quantify the vector and scalar balance of \( p_T \) between the diphoton system
  and the charged particle tracks associated with the vertex

Procedure:
- Muon selection:
  - Tight ID criteria (removing vertex related quantities) + tight tracker based relative isolation
  - Dimuon selection: 70 GeV < \( M_{\mu\mu} \) < 110 GeV
- Select as true vertex the one reconstructed using the two muon tracks
- Ignore the muons from the tracks collection entering the vertex fit
- Produce a new vertex collection obtained by refitting the vertices ignoring the muon tracks
- Select the vertex obtained using H→γγ vertex identification algorithm on the vertex collection
  without muon tracks.
- If \( \Delta z(\text{Selected Vertex}, \text{True Vertex}) < 1.0 \text{ cm} \), the selected vertex is considered to be correct
Fraction of γ+jet events with a converted photon for which the vertex has been correctly assigned by the H→γγ identification algorithm, as a function of the transverse momentum of the photon-plus-jet system. The tracks associated with the jet have been ignored during the vertex identification process to mimic a diphoton system. The simulated events are weighted such that the distribution of the per event vertex multiplicity matches the one in data. The vertex is considered as correctly identified if its z coordinate lies within 10 mm of the z coordinate of the vertex tagged by the jet. (top plot) data (in red) and simulated events (in black) at 13 TeV and (bottom plot) their ratio are shown.
Internal notes: vertex identification

Simulations:
Photon + jets sample MadGraph + Pythia 8

Data:
events selected with a single photon prescaled trigger: HLT_Photon50

H→γγ vertex BDT inputs:
- squared transverse momenta of the charged particle tracks associated with the vertex,
- two variables that quantify the vector and scalar balance of p_T between the diphoton system and the charged particle tracks associated with the vertex
- compatibility between the longitudinal position of the reconstructed vertex and the one estimated using conversions track(s)

Procedure:
- Select photon (converted) with pT > 55 GeV, passing loose identification criteria
- Require jets pT > 30 GeV
- Pileup re-weighting is performed based on the number of vertices in the events
- Apply the H→γγ vertex identification algorithm for photon+jet objects to obtain the selected vertex
- Tracks with ΔR(jet, track) < 0.4 are excluded from the calculation of the Vertex ID
- The vertex with the highest sum p_T^2 is taken as the ‘Jet Tagged Vertex’
- If Δz(Selected Vertex, Jet Tagged Vertex) < 1.0 cm, the selected vertex is considered to be correct
The vertex finding efficiency in $H \rightarrow \gamma \gamma$ simulated events at 13 TeV as a function of the transverse momentum of the diphoton system. The efficiency is computed as the fraction of events where the $z$ coordinate of the vertex selected by the vertex identification algorithm lies within 10 mm of the $z$ coordinate of the vertex that is closest to the simulated true vertex. The vertex finding efficiency for a Higgs boson of mass 125 GeV, integrated over its $p_T$ spectrum, is computed to be about 81%. The distribution is reweighted to the 2015 average pileup measured in data, but not rescaled for the data to MC ratio obtained from the $Z \rightarrow \mu \mu$ validation.
Internal notes: vertex identification

Simulations:
- $H \rightarrow \gamma \gamma$ gluon fusion sample at 125 GeV Powheg

$H \rightarrow \gamma \gamma$ vertex BDT inputs:
- squared transverse momenta of the charged particle tracks associated with the vertex,
- two variables that quantify the vector and scalar balance of $p_T$ between the diphoton system and the charged particle tracks associated with the vertex
- compatibility between the longitudinal position of the reconstructed vertex and the one estimated using conversions track(s)

Procedure:
- Events preselected requiring $p_T > m_{\gamma \gamma}/3$ ($m_{\gamma \gamma}/4$) on the leading (subleading) and loose requirements on the candidate shower shapes and isolation
- Selected vertex obtained from the reconstructed vertex collection using the vertex identification algorithm for the diphoton system
- True vertex obtained from the generator level information
- If $\Delta z(\text{Selected Vertex, True Vertex}) < 1.0$ cm, the selected vertex is considered to be correct
Photon identification BDT score of the lower-scoring photon of diphoton pairs with an invariant mass in the range $100 < m_{\gamma\gamma} < 180$ GeV, for events passing the preselection in the 13 TeV dataset (points), and for simulated background events (cyan histogram). Histograms are also shown for different components of the simulated background, in which there are either two, one, or zero prompt candidate photons. The distribution of the sum of all the simulated background events is scaled to data preserving the relative ratio of the single components, generated at leading order. The red histogram corresponds to simulated Higgs boson signal events.
Photon identification: internal notes

Simulations:
  Background:
    diphoton sample SHERPA
    photon + jet and QCD samples Pythia 8 preselected to increase the fraction of events
      with jets having a large electromagnetic fraction
  Signal:
    H→γγ gluon fusion sample at 125 GeV Powheg

Data:
  events are selected with a double photon trigger with E_T > 30 (18) GeV on the
    leading(subleading) photon candidate and applying shower shape requirements and
    m_{γγ} > 95 GeV

BDT inputs:
  - shower shapes
  - isolation variables
  - median energy density in the event
  - pseudorapidity and energy of the superclusters corresponding to the photons
Comparison of diphoton BDT classifier score distributions for data (points) and MC (stacked histograms), for events passing the preselection and in the mass range 100 GeV $< m_{\gamma\gamma} < 180$ GeV. The MC backgrounds are split into photon-photon (green), photon-jet (red) and jet-jet (blue) components. The analysis employs a BDT classifier to discriminate between signal-like and background-like diphoton pairs. The distribution of the sum of all the simulated background events is scaled to data preserving the relative ratio of the single components, generated at leading order.
Diphoton MVA: internal notes

Simulations:
  Background:
    diphoton sample SHERPA
    photon + jet and QCD samples Pythia 8 preselected to increase the fraction of events
    with jets having a large electromagnetic fraction
  Signal:
    H→γγ gluon fusion sample at 125 GeV (Powheg)

Data:
  events are selected with a double photon trigger with \( E_T > 30 \) (18) GeV on the leading(subleading)
  photon candidate and applying shower shape requirements and and \( m_{\gamma\gamma} > 95 \) GeV

Diphoton BDT inputs:
  per-event estimate of the diphoton mass resolution, the identification BDT scores of both photons,
  and the kinematic properties of the diphoton system, except for \( m_{\gamma\gamma} \). To avoid any dependence on \( m_H \),
  the transverse momenta and resolutions are divided by \( m_{\gamma\gamma} \).

Procedure:
- Events preselected requiring \( p_T > m_{\gamma\gamma}/3 \) (\( m_{\gamma\gamma}/4 \)) on the leading (subleading) and loose requirements on
  the candidate shower shapes and isolation
- photon candidates selected by requiring a value of the photon identification MVA which retains 99% of
  the signal photons
- \( m_{\gamma\gamma} \in [100, 180] \) GeV
Comparison of diphoton BDT classifier score distributions for $Z \rightarrow ee$ events with electrons reconstructed as photons in data (points) and MC (histograms). Events in the mass range 70-120 GeV with leading $p_T > m_{\gamma\gamma}/3$ and sub-leading $p_T > m_{\gamma\gamma}/4$ are required to pass the full analysis preselection, inverting the veto condition on the electrons. The BDT classifier trained on diphoton MC samples is applied to an independent di-electron sample for validation.
Simulations:
  Drell-Yann sample aMC@NLO + Pythia 8

Data:
  events selected with a single electron trigger with $p_T > 27$ GeV and loose electron identification requirements

Diphoton BDT inputs:
  per-event estimate of the diphoton mass resolution, the identification BDT scores of both photons, and the kinematic properties of the diphoton system, except for $m_{\gamma\gamma}$. To avoid any dependence on $m_H$, the transverse momenta and resolutions are divided by $m_{\gamma\gamma}$.

Procedure
- events preselected requiring $p_T > 30$ (20) GeV on the leading (subleading) and loose requirements on the candidate shower shapes and isolation
- the electron veto condition is inverted
- loose photon identification requirement is applied
- events in the mass range 70-120 GeV with leading (subleading) photon $p_T > m_{\gamma\gamma}/3$ ($m_{\gamma\gamma}/4$) are considered
H \rightarrow WW
Control region definition for ggH

WW control region:
• one electron and one muon with opposite charge
• lepton transverse momentum $p_T > 20/20$ GeV *- extra lepton veto ($p_T > 10$)
• $m_{ll} > 50$ GeV (to suppress Higgs contribution)
• $p_{Tll} > 45$ GeV
• $m_{et} > 20$ GeV
• $n_{jet}=0$ (jet $p_T > 30$ GeV)
• $b$-veto for jets between 10 and 30 GeV (loose working point)
WW Control Region. Invariant mass spectrum reconstructed using di-leptons (eμ final state).
WW Control Region. Transverse mass obtained combining lepton and met information.
Control region definition for ggH

Top control region:
• one electron and one muon with opposite charge
• lepton transverse momentum $p_T > 20/15$ GeV
• extra lepton veto ($p_T > 10$)
• $m_{ll} > 50$ GeV
• $p_{Tll} > 45$ GeV
• $M_{ET} > 20$ GeV
• $n_{jet}=0$ (jet $p_T > 30$ GeV)
• at least one jet b-tagged (loose working point) with $p_T$ between 10 and 30 GeV
Top Control Region. Invariant mass spectrum reconstructed using di-leptons (eµ final state).
Top Control Region. Transverse mass obtained combining lepton and met information.
ttH
The ttH multilepton analysis uses a MVA lepton identification algorithm to better reject non-prompt leptons mostly from b-jets in ttbar. The same overall strategy as in the Run-1 analysis is used with a revised choice of input variables, benefiting from the developments done in leptonic SUSY searches. Data/MC comparisons of the output discriminator of the lepton MVA in two control regions.

**Control region definition for ttH**

Pure control region:
- Dilepton events (same sign and opposite sign)
- Third lepton veto
- For opposite-sign events (targeting prompt leptons from dileptonic ttbar decays):
  - at least two jets
  - at least one b-jet satisfying the medium WP, or at least two satisfying the loose WP
- For same-sign events (targeting semileptonic ttbar decays with one non-prompt lepton):
  - exactly 3 or 4 jets
  - exactly one b-jet satisfying the medium WP

Mixed control region:
- Trilepton events
- one Z candidate (same-flavour opposite-sign pair, $60 < m(ll) < 120$ GeV)
- $E_T^{\text{miss}} > 20$ GeV
- relaxed requirements on the 3rd lepton ($p_T > 10$, loose id/iso/sip, no lepton mva cut)
Pure control region enriched in prompt leptons, opposite sign leptons. Lepton MVA score of the trailing lepton, after tight selection requirements applied on leading lepton. Simulation normalized to data yield. Statistical uncertainty due to the size of the simulated samples represented as a hatched area.
Pure control region enriched in non-prompt leptons, same sign leptons. Lepton MVA on the trailing lepton, after tight selection requirements applied on leading lepton. Simulation normalized to data yield. Statistical uncertainty due to the size of the simulated samples represented as a hatched area.
Mixed control region. Lepton MVA for the third lepton, tight selection requirements applied on leading lepton. Simulation normalized to data for the three components from fits to the dilepton mass for the Z candidate (for Z- vs non-Z backgrounds) and the transverse mass of the third lepton plus missing energy (for Z+X vs VZ+X). Statistical uncertainty due to the size of the simulated samples and statistic uncertainty on fitted normalizations for the three MC components represented as a hatched area.
Supporting material

Control regions where normalization factors for the mixed sample plot are fitted:
Supporting material

Mixed sample, split in electrons and muons
Data/MC comparisons for the main input variables in the pure samples: significance of the 3D impact parameter. (left) prompt enriched (right) non-prompt enriched
Data/MC comparisons for the main input variables in the pure samples: lepton mini Isolation ($p_T$ dependent isolation cone). (left) prompt enriched (right) non-prompt enriched
Data/MC comparisons for the main input variables in the pure samples built from the jet containing the lepton: $p_T$ rel. with respect to jet. (left) prompt enriched (right) non-prompt enriched.
Data/MC comparisons for the main input variables in the pure samples built from the jet containing the lepton: ratio between $p_T$ lepton and $p_T$ jet. (left) prompt enriched (right) non-prompt enriched.
Data/MC comparisons for the main input variables in the pure samples built from the jet containing the lepton: b-tagging discriminator of the jet. (left) prompt enriched (right) non-prompt enriched.
$H \rightarrow \tau \tau$
Observed and predicted visible di-τ mass distribution in events requiring the presence of an opposite-charge $\mu\tau_h$ pair and in which both leptons are required to pass identification and isolation criteria. The electroweak background contribution includes events from W + jets, diboson, and single-top-quark production. The data sample is collected with a single muon trigger and corresponds to 2.2 fb$^{-1}$ of $\sqrt{s} = 13$ TeV pp collisions data.
Observed and predicted visible di-τ mass distribution in events requiring the presence of an opposite-charge $e\tau_h$ pair and in which both leptons are required to pass identification and isolation criteria. The electroweak background contribution includes events from W + jets, diboson, and single-top-quark production. The data sample is collected with a single electron trigger and corresponds to 2.2 fb$^{-1}$ of $\sqrt{s} = 13$ TeV pp collisions data.
$H \rightarrow b\bar{b}$
Number of b-tagged jets after applying the b-tag and mis-tag scale factors derived from the shape-based b-tag discriminator reweighting method for events with exactly one tight lepton, at least four jets, and at least one jet passing the CSVv2 medium working point in the ttH analysis. The shape-based calibration of the b-tag discriminator uses a tag-and-probe technique to correct the per-jet b-tag discriminator distribution for both heavy- and light-flavour jets.
Performance of b-jet regression on event $p_T$ balance

Distribution of the ratio between the $p_T(jj)$ and the $p_T$ of the dilepton system on data versus MC using anti-$k_T$ 0.4 jets. Data and simulated events selected requiring two opposite sign same flavor leptons with $p_T > 20$ GeV and two central jets with $p_T > 20$ GeV and CSVv2 Loose WP. Only JEC and PU correction applied to MC. The average of the data distribution is 0.96.

Distribution of the ratio between the $p_T(jj)$ and the $p_T$ of the dilepton system on data versus MC after applying the jet energy regression to the anti-$k_T$ 0.4 jets. Data and simulated events selected requiring two opposite sign same flavor leptons with $p_T > 20$ GeV and two central jets with $p_T > 20$ GeV and CSVv2 Loose WP. Only JEC and PU correction applied to MC. The average of the data distribution is 0.99.
Performance of b-jet regression on simulated $ZH$ signal

Simulated ZH to $llbb$ events with $m_H=125$ GeV, selected requiring two opposite sign same flavor leptons with $p_T > 20$ GeV and two central jets with $p_T > 20$ GeV and CSVv2 Loose WP and $Z(ll) p_T > 100$ GeV. Invariant mass distribution of the reconstructed Higgs to bb candidate, with (red) and without (black) b-jets specific corrections are applied. A fit to both distributions has been performed using the Bukin function.
$H \rightarrow ZZ \rightarrow 4l$
The following plots are obtained using the following Monte-Carlo samples:
SM Higgs bosons: all SM production modes included, \( m(H) = 125 \) GeV
Irreducible backgrounds: \( q\bar{q}ZZ \) (Powheg), \( ggZZ \) (background only, MCFM)

The following cross sections for the irreducible backgrounds have been used:
\( q\bar{q}ZZ \): the NLO cross section from Powheg multiplied by a (NNLO/NLO) \( k \)-factor of 1.1
\( ggZZ \): the LO cross section from MCFM multiplied by an approximate (NLO/LO) \( k \)-factor of 1.7

These are additionally scaled for the ZZ measured signal strength (0.99).
The reducible background contribution is estimated from data.
Figures for $70 < m_{4l} < 110$ GeV

Distribution of the four-lepton reconstructed mass $m_{4l}$ in the $[70, 110]$ GeV range. Points with error bars represent the data and stacked histograms represent expected distributions. The ZZ backgrounds are normalized to the SM expectation, the Z+X background to the estimation from data.
Figures for $70 < m_{4l} < 110$ GeV

Distribution of the kinematic discriminant $D_{\text{kin}}^{\text{bkg}}$, for events in the $70 < m_{4l} < 110$ GeV region. Points with error bars represent the data and stacked histograms represent expected distributions. The ZZ backgrounds are normalized to the SM expectation, the $Z+X$ background to the estimation from data.
Figures for $70 < m_{4\ell} < 110$ GeV

Distribution of the kinematic discriminant $D^{\text{kin}}_{bkg}$ versus the four-lepton reconstructed mass $m_{4\ell}$ in the $70 < m_{4\ell} < 110$ GeV region. The color scale represents the expected event yield (not including the $Z+X$ background contribution), and the points represent the data.
Figures for $m_{4l} > 150$ GeV

Distribution of the kinematic discriminant $D_{\text{bkg}}^{\text{kin}}$, for events in the $m_{4l} > 150$ GeV region. Points with error bars represent the data and stacked histograms represent expected distributions. The ZZ backgrounds are normalized to the SM expectation, the Z+X background to the estimation from data.
Distribution of the kinematic discriminant $D_{\text{bkg}}^{\text{kin}}$ versus the four-lepton reconstructed mass $m_{4\ell}$ in the $m_{4\ell} > 150$ GeV region. The color scale represents the expected event yield (not including the $Z+X$ background contribution), and the points represent the data.
Figure for $m_{4l} > 70$ GeV, excluding the [110, 150] GeV region

Distribution of the four-lepton reconstructed mass $m_{4l}$ in the full mass range. The [110, 150] GeV region is blinded. Points with error bars represent the data and stacked histograms represent expected distributions. The $H^0(125)$ signal and the ZZ backgrounds are normalized to the SM expectation, the Z+X background to the estimation from data. No events are observed with $m_{4l} > 800$ GeV.
$H \rightarrow ZZ \rightarrow 2\ell 2\nu$
For all following plots all backgrounds are estimated using only Monte Carlo predictions, also the DY contribution where a data-driven method is preferred. An electroweak singlet signal with standard model couplings \(m_H=400\) and \(1000\) GeV produced by gluon fusion (full line) and by vector boson fusion (dashed line) is superimposed.
The invariant mass spectrum reconstructed using di-leptons (ee/μμ) with a transverse momentum greater then 20 GeV, and leptons that satisfy the tight ID and isolation requirements.
Transverse momentum of di-leptons pairs (ee/µµ) reconstructed using all the events where leptons pass tight ID and isolation requirements, and the reconstructed di-lepton invariant mass satisfies \[m_{ll} - 91\] < 15 GeV
Events selected requiring: exactly two leptons coming from the Z boson, the $Z \, p_T > 55$ GeV, a veto on the third lepton, no b-jets must be present in the event. Missing transverse energy obtained using Particle Flow Met (Type 1). The region above 100 GeV is kept blind, as shown in the plot.
Events selected requiring: exactly two leptons coming from the Z boson, the $Z p_T > 55$ GeV, a veto on the third lepton, no b-jets must be present in the event. Distribution of the transverse mass obtained using kinematic information on leptons & neutrinos. The region above 325 GeV is kept blind, as shown in the plot.
$H \rightarrow \text{invisible}$
Trigger efficiencies:

**VBF trigger** requires (at HLT) 2 PF jets with $p_T>40$ GeV, $\Delta \eta_{jj}>3.5$ and $M_{jj}>600$ GeV in addition to PFMET ignoring muons (METnoMU) $>140$ GeV and at L1 MET $>60$ GeV

**MET-only trigger** requires (at HLT) PFMET $>170$ GeV and at L1 MET $>60$ GeV
Efficiency of VBF Higgs to invisible trigger and MET only trigger in single muon data as a function of MET ignoring muons (METnoMU). The denominator of the efficiency is the number of events passing a single muon trigger which have two jets with $p_T > 80\,\text{GeV}$, $M_{jj} > 600\,\text{GeV}$ and $\Delta\eta_{jj} > 3.6\,\text{GeV}$.
Efficiency of VBF Higgs to invisible trigger in single muon data as a function of sub-leading jet \( p_T \). The denominator of the efficiency is the number of events passing a single muon trigger which have a leading jet with \( p_T > 80 \) GeV, METnoMU > 300 GeV, \( M_{jj} > 600 \) GeV and \( \Delta\eta_{jj} > 3.6 \) GeV.
Efficiency of VBF Higgs to invisible trigger in single muon data as a function of dijet mass ($M_{jj}$). The denominator of the efficiency is the number of events passing a single muon trigger which have two jets with $p_T > 80$ GeV, METnoMU $> 300$ GeV and $\Delta \eta_{jj} > 3.6$ GeV.
Efficiency of VBF Higgs to invisible trigger in single muon data as a function of dijet $\Delta \eta$. The denominator of the efficiency is the number of events passing a single muon trigger which have two jets with $p_T > 80$ GeV, $\text{METnoMU} > 300$ GeV and $M_{jj} > 600$ GeV.