Vertex identification in the search for the Higgs boson decaying to two photons in CMS

Inna Kucher on behalf of the CMS Collaboration
CEA Saclay IRFU/SPP (France)

A search for the Higgs boson decaying to two photons has been performed by the CMS Collaboration at the LHC experiment using pp collisions at a center-of-mass energy of 13 TeV with an integrated luminosity of 2.7 fb^{-1}. In the decay of the Higgs boson into two photons, the unconverted final state photons are not detected in the tracker. Moreover, the CMS electromagnetic calorimeter (ECAL) has no longitudinal segmentation, it is thus not a pointing calorimeter. As a result, the primary vertex determination of the Higgs boson decaying to two photons is not trivial. The vertex identification algorithm used in the search for the Higgs boson decaying to two photons is described, together with its performance.

1 Introduction

The Higgs boson decaying into two photons channel has a very small branching ratio, about 0.2%, but at the same time it has a very clean final state signature, two high-energy isolated photons. Vertex identification is an important part of the $H \to \gamma\gamma$ analysis [1], it has a direct impact on diphoton mass resolution and on the photons' isolation calculation with respect to tracks.

In the decay of the Higgs boson into two photons, unconverted final state photons are not detected in the tracker, so the determination of the primary vertex associated with the signal is not trivial. Information from the recoiling tracks can be used to determine the vertex position, and, when at least one of the photons is converted in the tracker, the conversion tracks can be used in addition. For LHC Run II, the tracking algorithm has changed [2], and the energy and pile-up conditions are different. Thus the vertex identification algorithm for the $H \to \gamma\gamma$ analysis was reoptimized.

The principle of the vertex identification used for CMS $H \to \gamma\gamma$ analysis for the LHC Run II is described in Section 2.1, where the distributions of the main discriminating variables from 13 TeV Monte Carlo simulation are shown and the performance estimated on simulation is quoted. The vertex identification performance validation on 13 TeV data and Monte Carlo simulation using $Z \to \mu\mu$ and γ+jets events are depicted in Section 2.2. A second algorithm is used to estimate per-event probability to choose the right vertex, allowing to fully benefit from the excellent CMS ECAL resolution. The principle of the probability determination is summarized in Section 3 with its performance at 13 TeV.

2 Vertex identification

The vertex identification algorithm exploits the correlation between the recoiling tracks attached to a given vertex and the diphoton system. In events where at least one of the two photons converts into an e^+e^- pair, the conversion tracks can be reconstructed and linked to the photon supercluster. All this information is used to build an optimal discriminating variable. In each event, the vertex with the most signal-like value of this variable is chosen as the primary one.

2.1 Algorithm

In presence of unconverted photons, the three most discriminating variables for the vertex identification were found to be: $\sum_i |\vec{p}_T^{\,i}|^2$, $-\sum_i(\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|})$, $(|\sum_i \vec{p}_T^{\,i}| - p_T^{\gamma\gamma})/(|\sum_i \vec{p}_T^{\,i}| + p_T^{\gamma\gamma})$, where $\vec{p}_T^{\,i}$ is the transverse momentum of the *i-th* track associated with a given vertex, $\vec{p}_T^{\gamma\gamma}$ is the transverse momentum of the diphoton pair.

The production vertex for a Higgs boson is likely harder than minimum bias interaction vertices, so $\sum_i |\vec{p}_T^{\,i}|^2$ tends to take higher values for the true vertex than for the wrong ones. The negative sum of projections of the $\vec{p_T}$ of the tracks on the diphoton $\vec{p_T}$, $-\sum_i(\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|})$, tends to be positive for the true vertex, as tracks recoil against the diphoton pair, and centered at 0 for wrong vertices. For the same reason, the asymmetry between the total transverse momentum of the tracks attached to a given vertex and the modulus of the diphoton p_T tends to have higher values for the true vertex and to peak at -1 for wrong vertices. The distributions of the discriminating variables for 13 TeV $H \to \gamma\gamma$ Monte Carlo samples are shown for right (H(125) vertex) and wrong (random wrong vertex) vertices in Figure 1.

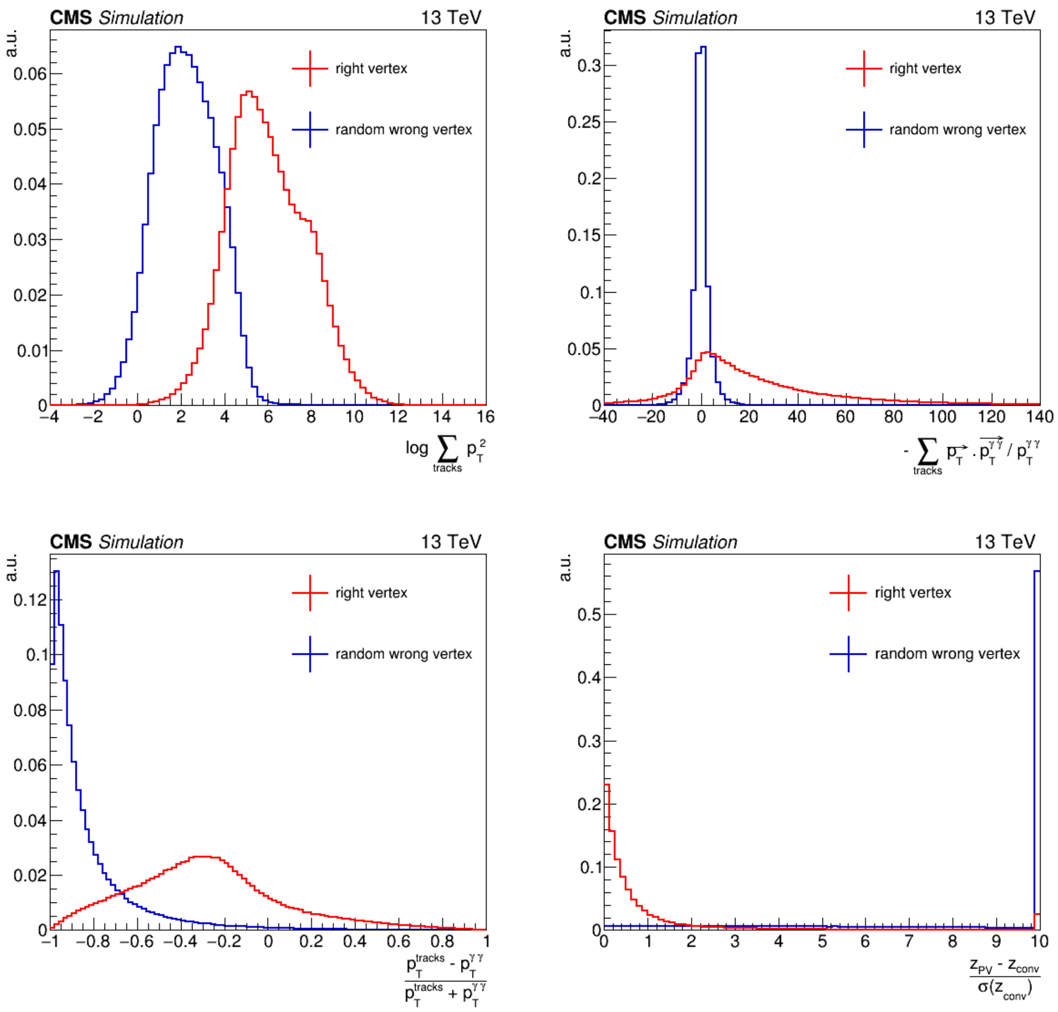

Figure 1 – Discriminating input variables distributions at 13 TeV for right (H(125) vertex) and wrong (random wrong vertex) vertices.

In the presence of conversions, i.e. photons producing electron-positron pairs due to the interaction with the tracker material, electron tracks can be exploited to determine the longitudinal coordinate of the primary interaction vertex where the Higgs boson is produced. Two complementary methods have been developed, which give different performance depending on where the photon conversion occurs. Both methods exploit the knowledge of the converted photon direction extracted from the conversion reconstruction. Once the direction is known, it is extrapolated to the beam line to obtain the estimate of the z position of the primary interaction vertex, z_V^{conv}. In the first method, the photon direction is calculated using the angle between the conversion momentum and the z-axis. The conversion momentum is evaluated from the track pair refitted with the vertex constraint. In the second method, the direction of the converted photon is instead determined by combining the information on the conversion vertex position and the position of the ECAL supercluster.

The variables used in the analysis in presence of converted photons are the number of converted photons ($N_{conv} = 0, 1, 2$) and $Pull_{conv} = \frac{|z_{PV} - z_V^{conv}|}{\sigma_{conv}}$, where z_{PV} is the z position of the tested primary vertex, z_V^{conv} is the z position of the primary vertex estimated from one of the two algorithms described above, and σ_{conv} is the resolution on z_V^{conv}, for the tracker part where the conversion vertex is reconstructed. The $Pull_{conv}$ variable provides a good separation between the right (H(125) vertex) and wrong (random wrong vertex) vertices which is shown in Figure 1.

All discriminating variables are combined, using a multivariate technique based on boosted decision trees. In each event, the vertex with the most signal-like value of this variable is chosen.

The vertex identification algorithm was optimized for LHC Run II using simulation for all Higgs production modes weighted according to their expected production cross-sections. The efficiency is defined as the fraction of events where the z position of the chosen vertex is located within 1 cm of the true one, in this case the photons' opening angle makes a negligible contribution to the diphoton mass resolution. It is about 83% from $H \to \gamma\gamma$ simulation. All production modes are included and reweighted to their expected production cross-sections to compute the efficiency, and the simulation is reweighted to the pile-up distribution in the data.

2.2 Performances validation

Since the algorithm was optimized on simulation, it has to be validated on data.

The performance of the $H \to \gamma\gamma$ vertex identification algorithm is validated for the case of unconverted photons using $Z \to \mu\mu$ events. The two muon tracks are removed from the event and vertices are re-reconstructed, mimicking the diphoton topology. In computing the discriminating, the diphoton kinematic properties are replace with those of the dimuon. The efficiency of finding the good vertex is estimated both in data and simulation. The vertex is selected within the vertex collection without muon tracks. The true "good" vertex is determined from the two muons tracks in the collection of vertices including the muon tracks. For the case of converted photons the performance of the algorithm is validated using γ+jet events. In this case, a photon-jet system is created by pairing a photon and a jet while removing the tracks associated with the jet during the process of vertex identification, in order to mimic a diphoton system. The selected vertex is then compared to the vertex that is associated to the jet in order to calculate the efficiency to choose the correct vertex.

The efficiency of selecting the correct vertex with the $H \to \gamma\gamma$ vertex identification algorithm in $Z \to \mu\mu$ and γ+jet events in data and simulation, along with their ratio, are shown in Figure 2 as a function of the p_T of the system. In both cases the efficiency in data and simulation agree to within. The $Z \to \mu\mu$ data/simulation efficiencies ratio versus p_T is used to correct the efficiencies in $H \to \gamma\gamma$ simulation, and varied within uncertainties to estimate the associated systematics.

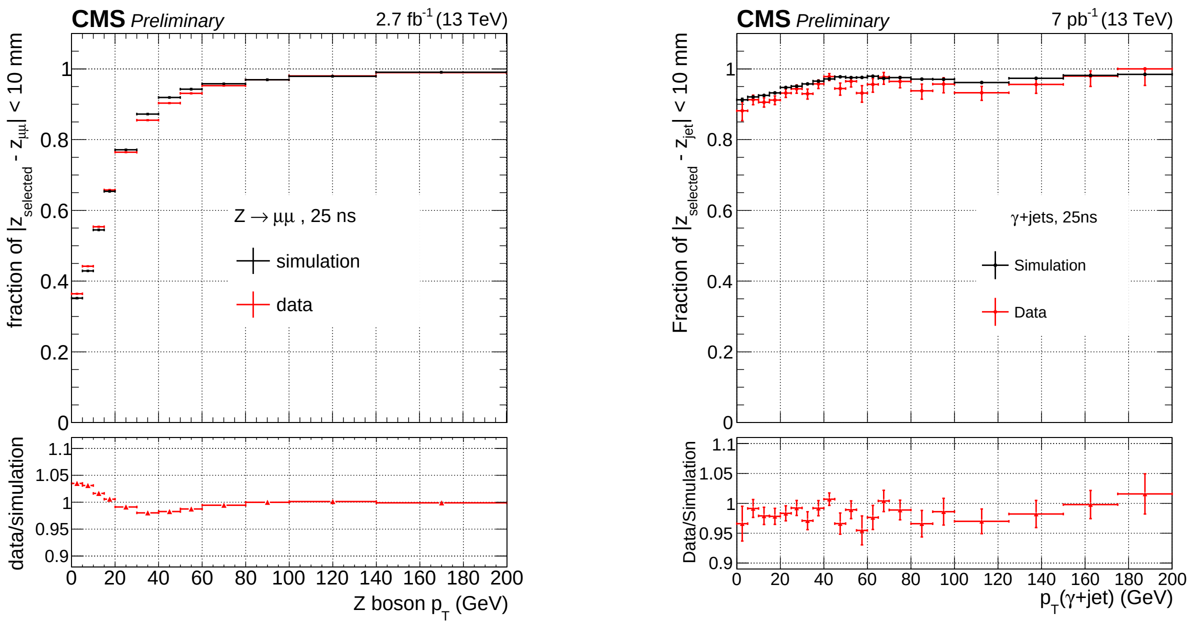

Figure 2 – Efficiency as a function of p_T to find the vertex within 1 cm of the true one, using $Z \to \mu\mu$ (left) and γ+jet (right) events for data and simulation.

3 Per-event probability to choose the correct vertex

A second vertex-related multivariate discriminant was designed to estimate, event-by-event, the probability for the vertex assignment to be within 1 cm of the diphoton interaction point. This allows a per-event estimate of the contribution of the vertex choice to the diphoton mass resolution. The method was optimized, using simulated $H \to \gamma\gamma$ events, to separate events where the chosen vertex lies within 1 cm of the generated interaction point. The inputs of the method are:

- the p_T of the diphoton system,
- the number of vertices in each event,
- the values of the per-vertex BDT discriminant for the three best vertices,
- the Δz between the best vertex and the second-best and third-best,
- the number of converted photons (0, 1, or 2).

The per-event probability to choose the vertex within 1 cm of the true one is parametrized as a function of the BDT output with a 4th order polynomial separately for events with converted photons and with only unconverted photons. The efficiency of choosing the vertex within 1 cm of the true vertex from Higgs to two photons simulation and the average vertex probability estimate are shown as a function of the p_T of the diphoton pair in Figure 3 (left). The probability estimate is modeling well the true efficiency.

The performance of the $H \to \gamma\gamma$ vertex probability estimate algorithm is validated using $Z \to \mu\mu$ events where the vertices are refitted without the muon tracks to mimic a diphoton system. The normalised distribution of the per event probability of correct diphoton vertex is shown in Figure 3 (right) separately for the vertices correctly selected and for mis-assigned vertices in $Z \to \mu\mu$ events in data and simulation. The correct and mis-assigned vertices are well separated, and the agreement between data and simulation is fairly good.

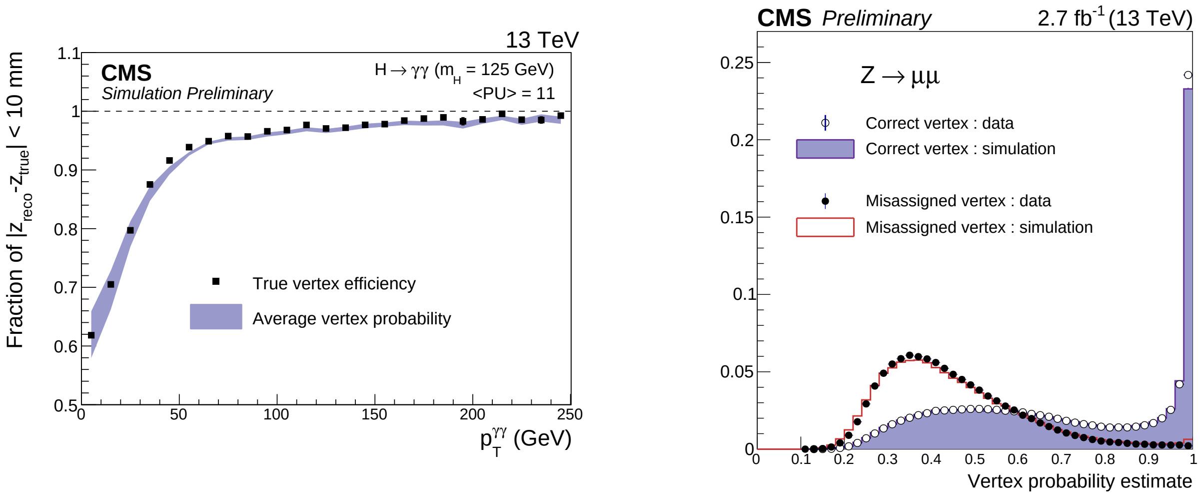

Figure 3 – Left: Comparison the true vertex identification efficiency and the average estimated vertex probability as a function of the reconstructed diphoton p_T. All production modes are included. Events are reweighted according to cross-sections of different production modes and with respect to pile-up in data. Right: Normalised distributions of the per event probability of correct diphoton vertex for correctly selected vertices in data (open black circles) and simulation (purple histogram) and for mis-assigned vertices in data (closed black circles) and simulation (open red histogram) in $Z \to \mu\mu$ events.

4 Conclusion

The vertex identification algorithm and good vertex probability estimate were optimized for LHC Run II and validated using data collected by CMS experiment in 2015. It is an important part of the $H \to \gamma\gamma$ analysis. The expected performance of the vertex identification algorithm in 2016 is about 79% for an average pile-up 20.

References

1. The CMS Collaboration *First measurements of the Higgs boson production in the diphoton decay channel at* $\sqrt{s} = 13\ TeV$, CMS Physics Analysis Summary, 2016.
2. Lorenzo Viliani for the CMS Collaboration, *CMS Tracker Performance and Readiness for LHC Run II*, CMS Conference report, 2015.