



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

CMS 上通过双光子末态测量希格斯性质和寻找额外希格斯

作者姓名: Muhammad Aamir Shahzad

指导教师: 陈国明研究员

学位类别: 理学博士

学科专业: 粒子物理与原子核物理

培养单位: 中国科学院高能物理研究所

2023 年 06 月



**Measurements of Higgs boson and search for additional Higgs
boson in the di-photon final state at CMS**

**A dissertation submitted to
University of Chinese Academy of Sciences
in partial fulfillment of the requirement
for the degree of
Doctor of Philosophy
in Particle and Nuclear Physics**

By

Muhammad Aamir Shahzad

Supervisor: Professor Chen Guoming

Institute of High Energy Physics, Chinese Academy of Sciences

June, 2023

中国科学院大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。承诺除文中已经注明引用的内容外，本论文不包含任何其他个人或集体享有著作权的研究成果，未在以往任何学位申请中全部或部分提交。对本论文所涉及的研究工作做出贡献的其他个人或集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日 期：

中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关收集、保存和使用学位论文的规定，即中国科学院大学有权按照学术研究公开原则和保护知识产权的原则，保留并向国家指定或中国科学院指定机构送交学位论文的电子版和印刷版文件，且电子版与印刷版内容应完全相同，允许该论文被检索、查阅和借阅，公布本学位论文的全部或部分内容，可以采用扫描、影印、缩印等复制手段以及其他法律许可的方式保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘要

概述

2012 年 7 月, 大型强子对撞机 (LHC) 的 ATLAS 和 CMS 实验分别发现了一个与标准模型 (SM) 希格斯玻色子性质相符合的粒子。到目前为止对该粒子的最新实验测量结果, 都与 SM 对希格斯玻色子的预测相一致。然而, SM 模型并不能完全回答所有的问题包括暗物质、暗能量和正反物质不对称性等一些最基本的物理问题, 不能排除存在超标准模型 (BSM) 的可能性。

用 LHC 收集的更多数据对观测到的希格斯玻色子进行精确测量, 可以帮助我们探索与标准模型的偏差, 看是否存在超出标准模型的新物理。随着数据的增加, 测量结果将变得更加精确, 这将促使理论家们更加精确地计算 SM 的预期。精确测量希格斯玻色子的性质最终将提供一个更深入的研究探针, 如果精确的理论计算与希格斯玻色子实验测量结果有所不同, 就表明可能存在新物理。

一些超标准模型, 例如次最小超对称模型 (NMSSM) 和广义的两希格斯双重态模型 (2HDM), 预言了存在多个希格斯玻色子的有趣现象。其中一些希格斯玻色子的质量, 可能低于观测到的 SM 希格斯玻色子的质量即 125 GeV。因此, 寻找额外的希格斯玻色子是研究 BSM 物理的最直接和有效的方法之一。

本论文的主要部分是通过双光子末态致力于寻找额外的、SM 类似的希格斯玻色子。该分析利用 CMS 13 TeV 数据来寻找双光子不变质量谱上的超出, 分别给出了 2016 年 (36.3 fb^{-1})、2017 年 (41.5 fb^{-1}) 和 2018 年 (54.4 fb^{-1}) 数据以及 2016 至 2018 年全部 Run2、 132.2 fb^{-1} 积分亮度的数据的统计结果。联合统计分析结果表明, 未观察到明显的超出。观测到的最大信号显著性为质量 95.4 GeV 处的大约 2.9 倍标准偏差。本论文分别给出了 2016 年、2017 年和 2018 年数据以及全部 Run2 数据 95% 置信水平上信号截面与分支比乘积的预期和观测的上限值。全部 Run2 数据 95% 置信水平上观测到的上限范围从 15 fb 到 73 fb 不等。

本论文也利用 CMS 探测器从 2016 年至 2018 年收集的 13 TeV 质子-质子对撞、 137 fb^{-1} 的积分亮度的数据, 通过双光子衰变道, 测量了希格斯玻色子的性质, 包括信号强度、信号强度修正因子、产生截面、与其他粒子的耦合、基准产生截面和微分基准产生截面。测量到的信号强度为 $\mu = 1.12^{+0.06}_{-0.06}(\text{theo})^{+0.03}_{-0.03}(\text{syst})^{+0.07}_{-0.06}(\text{stat})$, 基准截面为 $\sigma_{\text{fid}} = 73.40^{+5.4}_{-5.3}(\text{stat})^{+2.4}_{-2.2}(\text{syst}) \text{ fb} = 73.40^{+6.1}_{-5.9} \text{ fb}$ 。所有测量结果均与 SM 的预期相符合。

研究动机和背景

粒子物理学的“标准模型”(Standard Model, 简称 SM) 是迄今为止最成功的粒子物理理论, 它成功地描述了基本粒子在很宽的能量范围内的电弱和强相互作用 [1, 2]。欧洲核子研究中心 (CERN) 大型强子对撞机 (LHC) 实验 [3] 是目前世界高能物理最重要的粒子物理实验前沿, 是本世纪初人类认识微观世界最重要的物理实验之一, 实验研究结果将进一步揭示物质运动最根本的规律。LHC 运行第一阶段 (Run1, 2010-2012 年) 期间, 得到了希格斯粒子发现的重大成果 [4, 5], 这是近几十年来粒子物理学最重要的里程碑, 开启了探索宇宙中一切基本粒子质量起源的新时代, 成为 21 世纪粒子物理学的重要转折点。虽然标准模型能够解释目前所有粒子物理实验结果, 但存在许多不足之处, 它无法解释引力相互作用、中微子质量、规范等级、宇宙中的暗物质和暗能量以及物质与反物质不对称等问题, 探索能更好反映自然界作用规律的超出标准模型的新物理将会是粒子物理研究的主要方向。

用 LHC 收集的更多数据对观测到的希格斯玻色子进行精确测量, 可以帮助我们探索与标准模型的偏差, 看是否存在超出标准模型的新物理。随着数据的增加, 测量结果将变得更加精确, 这将促使理论家们更加精确地计算 SM 的预期。精确测量希格斯玻色子的性质最终将提供一个更深入的研究探针, 如果精确的理论计算与希格斯玻色子实验测量结果有所不同, 就表明可能存在新物理。通过双光子衰变道, 测量希格斯玻色子的性质包括信号强度、信号强度修正因子、产生截面、与其他粒子的耦合、基准产生截面和微分基准产生截面, 并且基于实验测量结果与理论进行精细比较以寻找差异, 是本论文的一部分。

同时, 考虑到强子对撞机上比较大的实验误差, 还不能排除这个希格斯玻色子属于其他粒子物理理论模型的可能性, 比如超对称模型 (SUSY) 等。一些超标准模型的理论, 比如次最小超对称模型 (NMSSM) [6–8] 或者双希格斯二重态模型 (2HDM) [9, 10] 或者 Georgi-Machacek (GM) 模型 [11, 12], 在保证存在一个质量为 125GeV 并且其属性与标准模型希格斯或者 LHC 希格斯测量结果相吻合的希格斯粒子的同时, 在小于 125GeV 的低质量区间可以存在额外的小质量希格斯粒子, 值得在实验上进行寻找。LEP 实验 2003 年寻找 Higgs 的研究结果表明在质量 98GeV 处有 2.3σ 的超出 [13]。LHC 上的 CMS 合作组利用 Run1 8TeV 数据研究的结果显示在质量为 97.5GeV 附近有大约 2σ 的超出, 而利用 Run2 2016 年 13TeV 的数据结果显示在质量为 95.3GeV 处有 2.9σ 的超出 [14]。LEP 和 LHC CMS 实验公布的 $95\text{--}98\text{GeV}$ 质量区间 $2\text{--}3\sigma$ 的超出结果, 急需 LHC 更多的数据积

累和进一步分析来进行检验。通过双光子末态寻找额外的、SM 类似的低质量希格斯玻色子，是本论文的主要研究内容。

标准模型及其相关物理

粒子物理学的“标准模型”是迄今为止最成功的粒子物理理论，它成功地描述了基本粒子在很宽的能量范围内的电弱和强相互作用。SM 是一种量子场论，它为物质的组成成分即基本粒子及其相互作用提供了现代式的理解。它可以被描述为一个扩展的数学模型，在量子场论的共同理论框架下来统一自然界四种基本力或相互作用中除重力外的三种，即电磁相互作用、弱相互作用和强相互作用 [15–17]。

1933 年恩里科·费米 (Enrico Fermi) 是第一个通过理解被称之为费米相互作用的 β 衰变来揭示弱相互作用的人 [18]。1961 年，Sheldon Glashow 提出电磁相互作用与弱相互作用相似，并以对称群为电磁相互作用和弱相互作用实现了理论上的统一。这个对称群分别为弱相互作用和电磁相互作用提供了载体，即 Z^0 和 $W^\pm$ 玻色子是电弱相互作用的载体而光子 (γ) 是电磁相互作用的载体。基本粒子通过 Robert Brout、François Englert、Peter Higgs 等人基于自发对称性破缺的想法提出的机制（统称希格斯机制） [19, 20] 获得质量，并将希格斯机制引入的新的基本粒子命名为希格斯玻色子。Abdus Salam 和 Steven Weinberg 随后将这一想法与 Glashow 的模型联系起来，从而完成了 SM 中电弱部分的描述。

与此同时，SM 的量子色动力学 (QCD) 部分也在开发中。60 年代末，在不同的散射实验中都有夸克存在的证据。1974 年粲夸克偶素 J/ψ 介子被两个互相之间独立的研究组 (Friedman 领导的研究组以及 Richter 与丁肇中等人) 分别发现，这进一步证实了夸克的存在。底夸克和反底夸克分别于 1976 年和 1977 年被发现，后来又发现了 τ 粒子 (第三代轻子)。费米实验室在 1995 年发现了顶夸克。与其他力相比，强相互作用的力具有非常高的能量量级，但还是不可能找到孤立的夸克。因此，碰撞中产生的夸克通过强子化产生包含大量粒子的喷注。强相互作用可以用 QCD 来解释，QCD 将夸克组成为强子。分别由 “uud” 和 “udd” 夸克组成的质子和中子是公认的最广泛提及的强子，它们是物质的原子核的组分。颜色中性的强子，包括由两个夸克 (夸克-反夸克对) 组成的介子，或者三个夸克结合成的重子。部分子是唯一表现出色电荷未消失的基本粒子。所有这些发现都为 SM 的预测提供了证据。最后，2012 年 7 月，欧洲核子研究中心的 CMS 和 ATLAS 合作组分别独立发现了希格斯玻色子 [4, 5]。2013 年的诺贝尔物理学

奖授予了 Peter Higgs 和 François Englert，以表彰他们对这种规范玻色子-希格斯玻色子-的理论预测。

希格斯到双光子的衰变模式为希格斯性质测量和新粒子寻找提供了一个干净的末态拓扑结构，可以通过双光子精确地重建希格斯玻色子的质量。双光子的衰变末态，是 2012 年 Higgs 粒子发现的主要贡献过程之一。胶子熔合产生 (ggH) 是希格斯玻色子的主要产生模式。另外的希格斯玻色子的产生模式包括相对较小截面的来自矢量玻色子聚变 (VBF) 过程，与矢量玻色子 (VH, $V=W$ 或 Z 玻色子) 关联产生过程以及与 $t\bar{t}$ 对关联产生 ($t\bar{t}H$) 过程。质量为 125GeV 的希格斯粒子衰变到双光子的分支比较小，大约是千分之二左右。双光子末态物理分析的主要的背景或本底，包括质子-质子 (pp) 对撞直接产生的不可约的双光子过程，由于喷注中假光子导致的可压制的 $pp \rightarrow \gamma + jet$ (光子加喷注) 和 $pp \rightarrow jet + jet$ (双喷注及以上) 过程。

实验背景 (LHC 实验和 CMS 探测器)

欧洲核子研究中心 (CERN) 大型强子对撞机 (LHC) 实验是目前世界高能物理最重要的粒子物理实验前沿，是本世纪初人类认识微观世界最重要的物理实验之一，实验研究结果将进一步揭示物质运动最根本的规律。LHC 位于法国和瑞士的边界，是一个 27 公里长的强子对撞机，其设计的质心能量高达 14 TeV [21]，用来精确地探测电弱尺度的物理。LHC 上主要有四个探测器实验，其中一个紧凑型缪子螺线管探测器 (CMS)。其他三个是 ATLAS (环形 LHC 装置)、ALICE (大型离子对撞机实验) 和 LHCb (LHC B 物理)。LHC 由 1232 个主偶极子组成，产生高达 8.33T 的磁场，以测量粒子进入探测器的曲线轨迹。386 个四极、360 个六极和 336 个八极也在那里用于控制束流的稳定性。2008 年，质子束首次被注入大型强子对撞机。2009 年 11 月，LHC 实现了每束 1.18 TeV 的能量，击败了 Tevatron 的 0.98 TeV，成为世界上最强的粒子加速器。2010 年，LHC 实现了质心能量高达 7TeV (每束质子束 3.5TeV) 的对撞。LHC 在 2011 年和 2012 年分别提供了 7TeV 和 8TeV 质子-质子对撞数据。LHC 在 2013-2014 年期间关闭，来升级其对撞机和探测器的主要部件。它于 2015 年以 13TeV 的质心对撞能量再次启动，对撞质心能量接近其设计值 (14TeV)。LHC 机器在第二次运行阶段成功取数后，于 2018 年关闭，进行了几次重大升级。在 2019 年至 2021 年的第二次长停运期 (LS2) 期间，升级了所有 LHC 实验的几个探测器组件。这些升级包括，为 1200 多个磁体改进了二极管的电绝缘，还安装了新的准直器，用于 LHCb

实验的中性束吸收器，以及更换了 22 个超导磁体等。在 LS2 之后，LHC 已经从 2022 年开始了第三阶段的运行，获取了 13.6TeV 的质子-质子对撞数据。

LHC 的主要物理目标之一是发现希格斯玻色子并精确测量标准模型。2012 年，LHC 实现了一个粒子物理世上的一个里程碑，ATLAS 和 CMS 合作组分别宣布发现了一种类似于 SM 希格斯玻色子的新粒子 [4, 5]。希格斯玻色子被认为是 SM 的最后缺失的一个部分。LHC 上分析了这个新发现粒子的各种产生和衰变模式，证实了它就是希格斯玻色子。因此，整个空间存在希格斯场的想法被人们强烈接受。希格斯玻色子发现之后，精确测量希格斯玻色子的性质是 LHC 的主要实验目标。同时，对 SM 进行了许多严格的测试以观察测量结果与理论预测的差异，以及寻找可能表明存在超出 SM 范围的物理即超标准模型 (BSM)。在本论文中，讨论了在 LHC 中希格斯玻色子的各种产生和衰变模式，这是本论文的主要物理分析的研究基础，包括标准模型希格斯玻色子性质测量的分析以及寻找低质量 SM 类似的额外希格斯玻色子的分析。

CMS 是 LHC 上的主要探测器之一，其实验设计的主要目标是要寻找希格斯粒子 (Higgs) 以及超出标准模型的其他新粒子、探清自然界的电弱破缺机制以及寻找暗物质等。它同时可以与 LHC 另外一个通用探测器 ATLAS 的实验结果，进行相互验证。CMS 和 ATLAS 具有相同的物理目标，但具有不同的几何形状。CMS 能够以高速率来处理高堆积事例。CMS 探测器及其电子设备能承受高辐射剂量和高通量，它对带电粒子的各参数具有显著的分辨率和轨迹重建效率。特别地，CMS 在很大的动量范围内具有良好的 μ 子识别和动量重建分辨率，并能够正确确定其电荷。CMS 在 τ 轻子和 b 夸克上也具备良好的动量分辨率、重建和触发。CMS 对电子和光子具有非常好的能量分辨率以及孤立性重建，同时保持对电子和光子较高的接受度。这些特性对于在双光子衰变通道中测量希格斯玻色子的性质以及寻找额外的新粒子是非常重要的。CMS 具有出色的喷注重建，以及有效的缺失横能量 (MET) 计算和质量分辨率。寻找超对称性的各种分析都认为 MET 是关键因素之一。CMS 能够探测各种各样的其他物理领域，例如在从几个 GeV 到 TeV 尺度的能量范围内搜索暗物质和额外维度 [22]。CMS 探测器的最终目标是精确测量带电粒子的动量，因此磁场配置在 CMS 中起着关键作用。220 吨的超导螺线管保持在 4.5K(-268.5°C)，这是 CMS 探测器的核心特征。磁场的方向分别是朝向螺线管内部 (径迹室) 和螺线管外部 (μ 子室)，这允许对 μ 子进行两次动量测量 [23, 24]。CMS 超导螺线管磁体直径 6 米，长度 12.5 米，产生 3.8T 的磁场，比地球磁场强 10 万倍。

寻找低质量 SM 类似的额外希格斯玻色子的分析策略

本论文的主要内容是利用双光子衰变末态来寻找一个不变质量低于 110 GeV 的额外希格斯玻色子。该分析工作是利用 CMS 探测器第二运行阶段 (Run2) 即 2016 年、2017 年和 2018 年收集到的 13 TeV 数据的基础上开展的分析, 然后将每年的结果进行联合统计分析。这项研究, 致力于在平稳下降的本底的双光子不变质量谱上寻找超出的共振峰。本论文所研究的双光子不变质量 $m_{\gamma\gamma}$ 的范围为 $70\text{GeV} < m_{\gamma\gamma} < 110\text{GeV}$ 。分析中, 多变量分析技术 (MVA) 方法, 被广泛的应用在光子鉴别、信号事例的鉴别和分类中, 以获得最佳的分析灵敏度。最后, 通过对双光子不变质量谱进行拟合, 来提取信号。

由于每一年数据在取数时采取的事例触发方案中的选择条件都不一样, 物理分析时所选取的事例选择条件每一年都不尽相同。分析时, 分别对每一年的分析要素包括能量的刻度、蒙特卡洛 (MC) 模拟的纠正以及事例选择及其效率进行了仔细研究。在光子鉴别和双光子信号事例的鉴别中, 分别使用到了多变量分析技术。基于双光子信号鉴别的输出值或者得分, 把事例划分成不同的事例类, 来提高分析的灵敏度。2017 年和 2018 年分析中, 还额外引入了 VBF 标识的事例类, 在经过初步的双喷注的动力学要求之后, 通过机器学习以及决策树方法 (BDT) 训练所谓的双喷注 BDT 来区分 VBF 信号事例和 ggH 信号事例以及本底事例, 以及双光子加双喷注结合的 BDT 来区分 VBF 信号和本底以及事例分类, 这是与之前 2016 年数据发表的结果 [14] 的不同与创新。本课题分析的一个难点, 是处理残存的 Drell-Yan(DY) 事例。在 2016 年数据发表结果 [14] 分析中使用到的方法即要求硅像素探测器中不存在光子对应的着火点来排除电子的基础上, 通过额外引入两个新的选择条件来压制 DY 事例, 这也是本论文中全部 Run2 数据分析中的创新。在对残余 DY 事例的双光子质量分布的描述时, 在 double-sided Crystal Ball 函数 (2016 年数据发表结果 [14] 分析中使用到的方法) 的基础上引入了额外的幂指数函数, 使得拟合函数能够更好的描述残余 DY 事例的双光子质量分布。和 2016 年数据发表结果 [14] 的分析一样, 对本底质量谱的描述采用了 CMS $H \rightarrow \gamma\gamma$ 物理分析小组通用的所谓的“envelope”方法或者“discrete profiling”方法, 而对信号质量分布的参数化描述则使用多个高斯函数通过对不同质量假设的 SM-类似的希格斯 MC 事例的双光子不变质量分布进行“同时信号拟合”的方法来得到。最后, 通过 CMS 合作组通用的统计分析软件包“combine”进行统计分析, 得到截面测量上限以及信号显著性等统计分析结果。因为 DY 本底压制的分析策略的改进, 本论文中重新分析了 2016 年的数据即较新的 Legacy 重建数

据, 在包括 DY 本底抑制以及谱形描述等新的分析策略下重新对所有基本要素包括事例触发效率、MC 模拟的修正、使用多变量分析技术对光子的鉴别和信号事例的鉴别、选择效率等, 进行了重新研究。最后, 对 2016-2018 年 3 年数据进行了联合统计分析, 得到了全部 Run2 数据低质量双光子共振态寻找的实验寻找结果。

标准模型希格斯玻色子性质测量的分析策略

本论文还通过双光子末态、利用全部 Run2 数据测量了希格斯粒子的性质, 包括利用 STXS 框架 [25] 测量希格斯玻色子的信号强度、产生截面、与其他粒子的耦合强度等, 以及基准截面和微分基准截面的测量 [26]。本文所包括的希格斯玻色子性质测量的两个分析, 使用了完整的 Run 2 数据集, 2016 年、2017 年和 2018 年数据的积分亮度分别为 35.9 fb^{-1} 、 41.5 fb^{-1} 和 59.4 fb^{-1} , 即总的积分亮度为 137 fb^{-1} 。模拟的信号样本是在考虑所有主要希格斯玻色子产生模式的情况下 MC 模拟产生。信号提取策略以及信号和本底的参数化描述技术, 对于这两种分析都是相同的, 与上面低质量寻共振态寻找中的技术是一样的。信号事例的提取方法, 也是通过对双光子不变质量谱进行拟合来提取。

希格斯粒子的性质测量的第一部分研究内容主要是在简化模板横截面(STXS)框架内, 在不同事例分类(bin)里面测量了希格斯粒子的截面。STXS 框架提供了一种在减少了理论不确定性的同时允许使用先进的实验方法和技术来测量希格斯玻色子耦合或截面的方法。这些处理技术首先在参考文献 [27] 中引入, 随后根据实验结果开发了后续的新版本 [28]。在 STXS 框架中, 基于希格斯玻色子的产生模式, 定义了由理论驱动的运动学区域。这些简化的事例 bin 在不同分析阶段(stage)定义是不一样的。在信号 bin 不断增加即不断细化下, 实验测量结果有助于为各种理论解释提供更多的信息。因此, 不断细化的 STXS 测量增强了发现可能的 BSM 信号的灵敏度。LHC 不断取数增加了更多积分亮度的数据, 从而提高了希格斯性质测量的实验灵敏度, 这也会导致数据分析时采用 STXS 框架的最新定义的 stage。在之前的分析 HIG-18-029 中 [29] 中, 采用 STXS stage1 策略, 已经测量了基于 ggH 和 VBF 的动力学来分 bin 的截面。现在最新的 STXS 框架是 stage 1.2, 而按照此 stage 1.2 的策略来更细致的分 bin, 是与之前分析的最大不同。

除了 STXS 框架测量截面, 本论文也包含了对基准截面和微分基准截面的测量, 得到最小模型依赖性的实验结果。使用希格斯到双光子衰变模式可以高精度重建双光子不变质量, 更精确的进行微分产生截面的测量。与之前的微分截面

研究结果相比,本次测量研究了希格斯截面随许多新的观测值的分布。基准截面的测量,目的是减少提取信号结果时对理论假设的依赖性。而基准区域的定义,是尽可能减少探测器接受度的影响。事例的分类是通过重建级别上的双光子质量分辨率的来优化划分的。基准微分截面,是基于重建水平上双光子系统以及与双光子一起产生的喷注等事例层级的变量分 bin 测量得到的。这些经典变量,包括双光子的横动量、喷注的个数以及能量最高喷注的横动量等。在这一次的测量中,额外地测量了基准截面随着用来探索低双光子横动量区间性质的相角变量 ϕ_η^* 以及喷注赝快度加权的喷注横动量 τ_C^j [26] 的微分分布。此外,微分基准截面的测量,同时在 2 维观测量的相空间进行了研究。论文中包括了微分基准截面随双光子横动量和喷注数的 2 维微分截面,以及随双光子横动量和 τ_C^j 的 2 维微分截面。

结果总结和展望

在利用双光子衰变末态来寻找一个不变质量低于 110 GeV 的额外希格斯玻色子的研究中,本论文分别给出了 2016-2018 每一年以及 3 年联合分析的分析结果。在 2016 年的分析中,观察到一个质量位于 95.7 GeV 的局部信号显著性 3.3σ (全局信号显著性 2.1σ) 的超出,而在 2017 年和 2018 年没有观察到这么显著的超出。在 95% 置信水平上,测量到的产生截面与分支比乘积的最小(最大)上限分别为 2016 年质量为 87.2 (95.7) GeV 的约 30 (154) fb、2017 年质量为 108.8 (70) GeV 处的约 34 (159) fb 以及 2018 年质量为 110 (74.6) GeV 处的 24 (94) fb。最后,将 2016 年、2017 年和 2018 年数据集的结果在质量范围为 $70 < m_{\gamma\gamma} < 110$ GeV 内进行了联合统计分析,给出了信号显著性和在 95% 置信水平上产生截面与分支比乘积相对于 LHCHGXS WG [30] 给出的 SM 类似的 BSM 预言的上限以及绝对上限。全部 Run2 数据观测到的最大信号显著性为质量 95.4 GeV 处的大约 2.9 倍标准偏差,在 95% 置信水平上观测到的产生截面与分支比乘积的绝对上限范围从 15 fb 到 73 fb 不等。

利用 2016-2018 年收集的 13 TeV 质子-质子对撞、 137 fb^{-1} 的积分亮度的数据以及双光子衰变道测量了希格斯玻色子的性质,包括信号强度、信号强度修正因子、产生截面、与其他粒子的耦合、基准产生截面和微分基准产生截面。在 STXS 框架下,测量到的总信号强度为 $\mu = 1.12_{-0.06}^{+0.06}(\text{theo})_{-0.03}^{+0.03}(\text{syst})_{-0.06}^{+0.07}(\text{stat})$,单一衰变道的测量精度达到大约 8%。同时,论文中也给出了不同信号产生模式的信号强度。接着给出了两种不同分 bin 情况下,即保证每个 bin 的期望误差小于

SM 预言的 150% 的所谓的“最大化”合并 bin 的情况以及确保拟合参数（每个 bin 的截面）的关联性小于 90% 的所谓的“最小化”合并 bin 的情况，stage1.2 截面的测量结果，它们均与 SM 预言的截面相符合。测量到的信号强度修正因子以及与其他粒子的耦合，均没有观测到与 SM 预言的偏离。测量到的基准截面为 $\sigma_{\text{fid}} = 73.40^{+5.4}_{-5.3}(\text{stat})^{+2.4}_{-2.2}(\text{syst})\text{fb} = 73.40^{+6.1}_{-5.9}\text{fb}$ 。该基准截面测量结果以及所有的 1 维和 2 维基准微分截面的测量结果，均与 SM 的预期相符合。

展望未来，随着 LHC 不断取数即数据量的不断增加，对希格斯粒子性质的测量会越来越精确。如果可以看到与 SM 预期的偏差，将会是对标准模型的一个重大突破。与此同时，基于 BSM 物理模型的寻找工作一直在进行中，这也可能使我们不断获得新的观测结果。关于低质量 $H \rightarrow \gamma\gamma$ 的寻找，更多的数据量（如 LHC 正在进行的第三阶段 Run 3 的数据采集）可以用来对我们 Run2 看到的超出进行定论。同时，对低质量额外希格斯粒子的寻找，可以利用 ZH 生产模式（Z 衰变到双轻子）以双轻子（双缪或者双电子）或者单电子的高阶触发来拓展双光子不变质量区间的寻找范围。同时，低质量 $H \rightarrow \gamma\gamma$ 的寻找结果，可以对 BSM 模型的可能限制 [8, 31, 32] 等提供最新的实验依据。

关键词: CMS 实验，标准模型，超标准模型，希格斯玻色子，双光子末态。

Abstract

In July 2012, a particle with properties compatible with the Standard Model Higgs boson was discovered at the Large Hadron Collider (LHC) by ATLAS and CMS experiments. The measurements showed that the new Higgs boson so far in accordance with those predictions for the SM Higgs boson. However, this model is not fully able to address several fundamental physics issues such as the dark matter, the dark energy and the asymmetry of matter and anti-matter. So, we can not exclude the possibility of physics Beyond Standard Models (BSM).

Precision measurements of the observed Higgs boson with more data collected at the LHC, could lead us to probe deviations from the Standard Model. With more data, the measurements would become even more precise which will push theorists to calculate their anticipations more precisely as well. It will eventually provide a deeper probe to see, if a more-precise prediction diverges from the current Higgs boson experimental measurements which could direct to new physics.

Some of the beyond Standard Models, e.g. the Next-to-Minimal Supersymmetric Model (NMSSM), and generalized two-Higgs-doublet models (2HDM) gives rise to an interesting phenomenology including several Higgs bosons, some of which could have masses below the observed mass of SM Higgs boson i.e. 125 GeV. Therefore, searching for the additional Higgs bosons is one of the most direct and effective ways to study the BSM physics.

The main part of this dissertation is dedicated to the probe for an additional SM-like Higgs boson in the diphoton final state. This analysis searches for an excess in the diphoton invariant mass spectrum over a smoothly falling background in the range $70 < m_H < 110$ GeV with CMS at 13 TeV. The results are presented for 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}), 2018 (54.4 fb^{-1}) data and 2016 to 2018 (full Run 2) combination corresponding to an integrated luminosity of 132.2 fb^{-1} . No significant excess is observed. One modest excess with approximately 2.9σ local (1.3σ global) significance is observed for a mass hypothesis of 95.4 GeV. The expected and observed upper limits with 95% confidence level on the product of cross section times branching ratio into two photons for each of the analysis with 2016 data set, 2017 data set, 2018 data set and the

statistical combinations of the full Run 2 data set, are addressed. The observed upper limit for the combined data set ranges from 73 fb to 15 fb.

This thesis also includes some of the latest results for the measurements of Higgs boson properties in the diphoton decay channel including the signal strengths, signal strength modifiers, production cross sections, its couplings to other particles, inclusive and differential fiducial production cross sections. Events are selected from a sample of proton-proton collisions at 13 TeV collected by the CMS detector at the LHC from 2016 to 2018, corresponding to an integrated luminosity of 137 fb^{-1} . The observed signal strength is $\mu = 1.12_{-0.06}^{+0.06}(\text{theo})_{-0.03}^{+0.03}(\text{syst})_{-0.06}^{+0.07}(\text{stat})$. The measured fiducial cross section is $\sigma_{\text{fid}} = 73.40_{-5.3}^{+5.4}(\text{stat})_{-2.2}^{+2.4}(\text{syst})\text{fb} = 73.40_{-5.9}^{+6.1}\text{fb}$. All results are found to be in agreement with the SM expectations.

Key Words: CMS experiment, Standard Model, Beyond Standard Model, Higgs boson, diphoton final state.

Contents

Chapter 1: Introduction	1
Chapter 2: Standard Model and LHC physics	5
2.1 The Standard Model	5
2.1.1 Theoretical Overview	5
2.1.2 Elementary Particles	6
2.1.3 Fundamental Interactions	8
2.1.4 Electroweak Unification	10
2.1.5 Spontaneous symmetry breaking and Higgs mechanism	11
2.2 Higgs boson properties	15
2.2.1 Higgs boson production modes at LHC	15
2.2.2 Higgs boson decay modes	17
2.3 Review of Higgs boson property measurements and additional Higgs searches at LHC	18
Chapter 3: LHC and Compact Muon Solenoid (CMS) Experiment	23
3.3.1 Introduction	23
3.3.2 The LHC layout	23
3.3.3 Machine operation and performance	24
3.3.4 Luminosity and beam parameters	26
3.3.5 Proton-proton collisions at LHC	27
3.3.6 Full Run 2 (2016-2018) data taking and High Luminosity LHC	29
3.4 The Compact Muon Solenoid (CMS) experiment	31
3.4.1 CMS detector overview	31
3.4.2 Coordinate System of CMS	32
3.4.3 The CMS Solenoid	33
3.4.4 Tracker System	34
3.4.5 Pixel Detector	35
3.4.6 Silicon Strip Detector	36
3.4.7 Electromagnetic Calorimeter (ECAL)	36
3.4.8 Hadronic Calorimeter (HCAL)	38
3.4.9 Muon System	40
3.4.10 Resistive Plate Chambers (RPC's)	42

3.4.11 Cathode Strip Chambers (CSC's)	43
3.4.12 Drift Tubes (DT's)	44
3.4.13 CMS Trigger System	46
3.4.14 CMS Level 1 (L1) Trigger	46
3.4.15 High Level Trigger (HLT) system	46
3.4.16 LHC Computing Grid	48
Chapter 4: Physics Objects	49
4.5 Particle flow	49
4.6 Photons	50
4.6.1 Definition of variables	51
4.6.2 Photon energy	52
4.6.3 Photon identification	55
4.6.4 Diphoton Vertex	59
4.7 Other objects	61
4.7.1 Electrons reconstruction and identification	61
4.7.2 Muons reconstruction and identification	62
4.7.3 Jet reconstruction	63
Chapter 5: Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state with full Run 2 dataset	65
5.8 Introduction	65
5.9 Data and simulation samples	67
5.10 Event selection and reconstruction	72
5.10.1 L1 trigger and HLT	72
5.10.2 Photon reconstruction	76
5.10.3 Photon preselection	80
5.10.4 Photon identification	85
5.10.5 Additional selections for Drell-Yan suppression	93
5.10.6 Diphoton vertex identification	98
5.10.7 Event classification	99
5.10.8 Diphoton BDT classifier	100
5.10.9 Boundary optimization	108
5.11 VBF selection strategy	111
5.11.1 Jet definition	111
5.11.2 VBF preselection	112

5.11.3	Dijet MVA and combined MVA categorization	112
5.11.4	Validations	115
5.12	Signal parametrization	120
5.13	Background estimation	125
5.14	Systematic uncertainties	137
5.15	Results	151
5.15.1	Full Run 2 combination	155
5.16	Conclusion	158
 Chapter 6: Higgs properties measurement in the diphoton decay channel		
6.17	Introduction	161
6.18	Signal strengths and STXS with full Run 2 dataset at 13 TeV	162
6.18.1	The STXS framework	162
6.18.2	Analysis strategy	164
6.18.3	Events categorisation	167
6.18.4	Data and MC samples	172
6.18.5	Signal and background modelling	173
6.18.6	Systematic uncertainties	174
6.18.7	Results and conclusion	175
6.19	Inclusive and differential fiducial cross-sections with full Run 2 dataset at 13 TeV	184
6.19.1	Introduction	184
6.19.2	Data and MC samples	185
6.19.3	Event reconstruction and categorization	185
6.19.4	Mass resolution estimator	188
6.19.5	Decorrelated mass resolution categorization	189
6.19.6	Observables and fiducial phase space	190
6.19.7	Systematic uncertainties	194
6.19.8	Results and Conclusion	194
 Chapter 7: Summary and Outlook		
7.20	Summary	203
7.21	Outlook	205
Bibliography		209
Acknowledgements		217

Author's resume, academic papers and other related academic achievements published during the degree program	219
--	-----

List of Figures

Figure 2-1 Elementary particles in three generations, gauge bosons and the Higgs boson.	7
Figure 2-2 Representaion of Higgs potential.	12
Figure 2-3 Feynman diagrams representing the four main Higgs production mechanism at tree level. a) Gluon-Gluon fusion, b) Vector Boson Fusion, c) Associated production W/Z and d) Associated top quark production.	16
Figure 2-4 Production cross sections of Higgs. As function of mass illustrated left and on right, as function of centre of mass energies.	16
Figure 2-5 Branching ratios of Higgs boson for various channels.	17
Figure 2-6 Signal strength parameters extracted for various production modes (left) and decay channels (right) [33].	19
Figure 2-7 Coupling modifiers measurements [33].	20
Figure 2-8 Limits on the production of Higgs boson pairs [33].	21
Figure 2-9 Observed and expected 95% CL upper limits of CMS (upper) and ATLAS (bottom).	22
Figure 3-10 Layout of LHC and its experiments in the CERN accelerator complex. The PS and the SPS also included through which the proton pass before reaching the LHC ring.	24
Figure 3-11 Cross Section values of several processes at different center-of-mass energy ($\sqrt{s}$) of p-p collisions.	26
Figure 3-12 Diagram of proton-proton collision at high energies.	28
Figure 3-13 LHC delivered luminosities shown in blue and the ones recorded by CMS are in yellow for 2015 (top left), 2016 (top right), 2017 (bottom left) and 2018 (bottom right).	30
Figure 3-14 Planned operation schedule of LHC and high-luminosity LHC upgrade.	31
Figure 3-15 Design of the CMS detector.	32
Figure 3-16 CMS coordinte system.	33
Figure 3-17 CMS tracker system layout and lines show the detector modules. .	34
Figure 3-18 Diagram of pixel detector. Lower half represent the old pixel detector and the upper part is the new pixel detector.	35
Figure 3-19 Pseudo-rapidity coverage of CMS ECAL.	37
Figure 3-20 Overall layout of CMS ECAL.	38
Figure 3-21 Layout of CMS HCAL.	39

Figure 3-22 Muon system layout of the CMS during Run2 operations.	41
Figure 3-23 Schematic diagram (updated) of CMS muon system with GEM chambers.	42
Figure 3-24 RPC chamber's cross sectional layout.	43
Figure 3-25 CSC functional layout.	43
Figure 3-26 Electronic boards of a CSC chamber and the readout system.	44
Figure 3-27 Cross sectional layout of DT chambers (one of the 5 wheels).	45
Figure 3-28 Diagram of drift cell.	45
Figure 3-29 The L1 trigger configuration.	47
Figure 4-30 Positron and Photon pattern in the detector.	53
Figure 4-31 Invariant mass distribution of $Z \rightarrow ee$ events shown for and 2017 endcap (left) and 2018 barrel (right) electrons.	54
Figure 4-32 Invariant mass distribution of $Z \rightarrow \mu\mu\gamma$ events shown for 2017 barrel (left) and 2018 endcap (right) photons selected from data and simulation.	54
Figure 4-33 Probe electron R_9 distribution in 2017 (top) and 2018 (bottom) data and simulation in ECAL barrel (left) and endcaps (right).	57
Figure 4-34 Distribution of photon ID MVA output with the photons selected from $Z \rightarrow \mu\mu\gamma$ events in selected with the L2016 (top), UL2017 (middle) and UL2018 (bottom) data and simulation in ECAL barrel (left) and endcaps (right). The systematic uncertainty applied to the shape of the MC distribution is shown by the hashed region.	58
Figure 4-35 Distribution of the ρ -corrected PF photon isolation for 2016 (top) and $\sigma_{in\eta}$ for 2018 (bottom) data and simulation in ECAL barrel (left) and endcaps (right).	59
Figure 4-36 (left) $H \rightarrow \gamma\gamma$ vertex identification algorithm validation using $Z \rightarrow \mu^+\mu^-$ events, omitting the muon tracks with 2017 data and simulation. (Right) Comparison of the true vertex ID efficiency and average estimated vertex probability as a function of reconstructed diphoton p_T . ..	61
Figure 4-37 A general of a collimated jet in p-p collision.	63
Figure 5-38 Pileup profiles for data taken in 2017 with the full integrated luminosity of 41.5fb^{-1} (black points) and with the prescaled HLT path (red points) in the left. For 2018 data with the full integrated luminosity of 59.7fb^{-1} (black points) and with the 54.4fb^{-1} from the low-mass HLT path (red points) used in the main analysis is in the right.	68
Figure 5-39 Layout of 2016 HLT path showing the list, position and logic of filters.	73
Figure 5-40 Layout of 2017 HLT path showing the list, position and logic of filters.	73

Figure 5-41 Layout of 2018 HLT path showing the list, position and logic of filters.	74
Figure 5-42 (2016) Seeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP technique. Efficiency w.r.t. offline probe p_T with the “OR” path (left) and “AND” path (right).	75
Figure 5-43 (2016) Final trigger efficiency of the “unseeded” leg, measured as the product of the efficiency of the unseeded leg of the standard path up to the mass filter (ϵ_i^{STD}), the ratio between the “OR”/“AND” path efficiency of the unseeded leg up to the pixel veto, the “STD” path efficiency of the unseeded leg just after the H/E filter (SF_i), the pixel-veto efficiency (ϵ_{PV}). The “OR” path results are shown on the left, while the “AND” path shown on right.	75
Figure 5-44 (2017) Seeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.	76
Figure 5-45 (2017) Unseeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.	76
Figure 5-46 (2018) Seeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.	77
Figure 5-47 (2018) Unseeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.	77
Figure 5-48 Distributions of raw σ_E/E output for 2016 (top), 2017 (middle) and 2018 (bottom) for the EB (left) and EE (right) cases. The electrons from $Z \rightarrow ee$ are reconstructed as photons.	78
Figure 5-49 Distributions of σ_E/E of photons in top (2016), middle (2017) and bottom (2018) with EB on left and EE on right with photons selected from $Z \rightarrow \mu\mu\gamma$ events. The bottom panel in each figure shows the data to MC ratio together with the systematic uncertainty.	79
Figure 5-50 (2017) p_T , supercluster η shown for barrel(left), and endcap(right), background photons distributions are shown in red, signal photons are in blue and the signal photon ones reweighted with 2D $p_T - \eta_{SC}$ weights are shown in green.	86
Figure 5-51 Full 5x5 R_9 (left) and ϕ width (right) for $Z \rightarrow ee$ events in data and simulation in ECAL barrel for probe candidates for 2016 (top), 2017 (middle) and for 2018 (bottom).	88

Figure 5-52 2017 distributions of the photon $Cov_{ini\phi}$ (left) and ϕ width (right) in ECAL barrel with 2017 (top) and for 2018 (bottom). Points are data and the green filled histogram is simulation.	89
Figure 5-53 Photon ID MVA output distribution for $Z \rightarrow ee$ (left) and $Z \rightarrow \mu\mu\gamma$ (right) in ECAL barrel for 2016 (top), 2017 (middle) and for 2018 (bottom).	91
Figure 5-54 Photon ID MVA output distribution for $Z \rightarrow ee$ (left) and $Z \rightarrow \mu\mu\gamma$ (right) in ECAL endcaps for 2016 (top), 2017 (middle) and for 2018 (bottom).	92
Figure 5-55 Photon identification BDT score of the lower-scoring photon of diphoton pairs in an event in 2017 (left) and 2018 (right).	93
Figure 5-56 Comparisons of the $\log(\Sigma p_T^2)$ distributions of the relic DY MC events and the signal events from 90 GeV ggH production mode (left), $\log(\Sigma p_T^2)$ versus $p_T^{\gamma\gamma}$ for the signal events from 90 GeV ggH production mode (middle) and the relic DY MC events (right). The boundary of the linear cut, $\log(\Sigma p_T^2) \leq 0.016 \times p_T^{\gamma\gamma} + 6.0$, is shown as the white line in each 2D plot. Top plots are for 2016, middle plots are for 2017 and the bottom ones are for 2018.	95
Figure 5-57 Selection efficiencies of the linear cut, “ $N_e^{Matched} = 0$ ” and the combination of these two requirements on the relic DY events (left) and 90 GeV ggH signal events (right). Top plots are for 2016, middle plots for 2017 and the bottom ones for 2018.	96
Figure 5-58 Efficiencies measured using $Z \rightarrow ee$ events in data (top left) and MC (top right), and the scale factors (bottom) of the linear cut with 2016, 2017 and 2018 datasets.	97
Figure 5-59 Comparisons on the input variables of 2016 legacy diphoton BDT between the signal and background. The blue distribution corresponds to the signal whereas the red one corresponds to the background.	102
Figure 5-60 2016 (left) and 2017 (right) diphoton BDT outputs of training and testing events in the signal and background samples.	103
Figure 5-61 2016 (left) 2017 (right) diphoton BDT classifier ROC curve.	103
Figure 5-62 Comparison between 2018 analysis diphoton BDT using 2017 analysis weight file and 2017 analysis diphoton BDT for the signal and background distribution with the same normalization across years (left) and for ROC (right).	103
Figure 5-63 Comparisons of data and simulation for the p_T distribution (top), $\cos(\Delta\phi)$ (middle) and diphoton BDT output score distribution (bottom) on electrons reconstructed as photons from Z boson decay for 2016 (left) and 2017 (right).	105

Figure 5-64 Comparisons of data and simulation for the p_T distribution (1a), $\cos(\Delta\phi)$ (b) and diphoton BDT output score distribution (c) on electrons reconstructed as photons from Z boson decay for 2018.	106
Figure 5-65 Distributions of diphoton invariant mass and diphoton BDT scores for 2016 (top), 2017 (middle) and 2018 (bottom) data sets.	107
Figure 5-66 Figure of merit based on the p-value as a function of the number of classes n_{cat} for 2016 (left) and 2017 (right).	110
Figure 5-67 Events distribution based on BDT output value. Background MC events (red), right vertex (green) and wrong vertex (blue) signal MC events. Black lines shows the class boundaries. The plots are shown for 2016 (left) and 2017 (right).	110
Figure 5-68 Comparison of training and the testing is on left which indicates no obvious sign of over-training and the ROC curve of dijet MVA training is on the right.	113
Figure 5-69 Comparison of training and the testing is on left which indicates no obvious sign of over-training and the ROC curve of combined MVA training is on the right.	114
Figure 5-70 The significance, ‘p-value’ symbolized as Z_0 as a function of various cut values on combined MVA score for 2017 (left) and 2018 (right) using the same weight files from 2017 analysis.	115
Figure 5-71 ‘ROC’ comparisons between 2017 and 2018 for the dijet MVA (left) and the combined MVA (right) with the same weights files from 2017 used in 2018 analysis.	116
Figure 5-72 Invariant mass of diphoton candidates, dijet BDT and combined BDT in $Z \rightarrow ee$ plus 2jets events selected from 2017 data and MC (left), and in 2018 data and MC (right). MC is normalized according to number of events in data for each year.	117
Figure 5-73 m_{jj} , $C_{\gamma\gamma}$ and $p_T^{\gamma\gamma}/m_{\gamma\gamma}$ in $Z \rightarrow ee$ plus 2jets events selected from 2017 data and MC (left), and 2018 data and MC (right). MC is normalized according to number of events in data for each year.	118
Figure 5-74 The distributions of m_{jj} and $C_{\gamma\gamma}$ in data and MC (without QCD) 2017 (left) and in 2018.	119
Figure 5-75 The distributions of dijet and combined MVA in data and MC (without QCD) 2017 (left) and in 2018.	120
Figure 5-76 (2016 (top), 2017 (middle) and 2018 (bottom)) Simultaneous fit (left) and the interpolation (right) on the simulated signal events for all classes combined. Different production modes are summed according to their relative SM cross sections.	122

- Figure 5-77 Full parameterized signal shape integrated over all event classes in 90 GeV Higgs boson Monte Carlo at $\sqrt{s} = 13$ TeV in 2016 (top left), 2017 (top right) and 2018(bottom). The black points are the weighted Monte Carlo events and the blue lines are the corresponding models. The effective σ values and corresponding intervals are also shown. 123
- Figure 5-78 Signal efficiency $\times$ acceptance in (top left) 2016, (top right) 2017 and (bottom) 2018 for the analysis as a function of mass hypothesis. 124
- Figure 5-79 Double-sided Crystal Ball function fits to the diphoton invariant mass distributions from 2016 Drell-Yan double-fake events for Class 0 (top left), Class 1 (top right) and Class 2 (bottom). 128
- Figure 5-80 Double-sided Crystal Ball function fits to the diphoton invariant mass distributions from 2017 Drell-Yan double-fake events for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). 129
- Figure 5-81 Double-sided Crystal Ball function fits to the diphoton invariant mass distributions from 2018 Drell-Yan double-fake events for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). 130
- Figure 5-82 A set of functions chosen to fit the background using the discrete profiling method for each of the inclusive classes for 2016. Class 0 (top left), Class 1 (top right) and Class 2 (bottom). 131
- Figure 5-83 A set of functions chosen to fit the background using the discrete profiling method for each of the inclusive and VBF classes for 2017. Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). 132
- Figure 5-84 A set of functions chosen to fit the background using the discrete profiling method for each of the inclusive and VBF classes for 2018. Class 0 (top left), Class 1 (top right) and Class 2 (bottom left) and VBF Class (bottom right). 133
- Figure 5-85 (2016) Background model fits using the chosen "best-fit" parametrization to data. The corresponding signal model for $m_H = 90$ GeV, multiplied by 10, is also shown for Class 0 (top left), Class 1 (top right) and Class 2 (bottom). The difference between the data and the best-fit model is shown in the lower panels. 134
- Figure 5-86 (2017) Background model fits using the chosen "best-fit" parametrization to data. The corresponding signal model for $m_H = 90$ GeV, multiplied by 10, is also shown for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). The difference between the data and the best-fit model is shown in the lower panels. 135

- Figure 5-87 (2018) Background model fits using the chosen "best-fit" parametrization to data. The corresponding signal model for $m_H = 90$ GeV, multiplied by 10, is also shown for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). The difference between the data and the best-fit model is shown in the lower panels. 136
- Figure 5-88 (2016: Class 0) Fit of the single-fake events for Data (top left), DY MC (top right) and $MC_{allwoQCD}$ (bottom). 141
- Figure 5-89 (2017 (left) / 2018(right) Class 0): Fit of the single-fake data (top), DY MC (middle) and $MC_{allwoQCD}$ (bottom) events. 142
- Figure 5-90 Full Run 2 unblinded impact plots of NP on signal strength for 70 GeV (top) and 90 GeV (bottom) signal mass, where θ_0 is the NP before the fit and θ is the NP after the fit. The red band shows the impact on the signal strength when θ reaches its $+1\sigma$ while the blue one shows the impact when θ reaches its -1σ value. 149
- Figure 5-91 Full Run 2 unblinded impact plots of NP on signal strength for 95 GeV (top) and 110 GeV (bottom) signal mass, where θ_0 is the NP before the fit and θ is the NP after the fit. The red band shows the impact on the signal strength when θ reaches its $+1\sigma$ while the blue one shows the impact when θ reaches its -1σ value. 150
- Figure 5-92 The expected number of signal events per event class and the percentage breakdown per production mode in 2016 (top) 2017 (middle) and 2018 (bottom) analysis. The effective σ is also provided as an estimate of the $m_{\gamma\gamma}$ resolution in that category and compared directly to the full-width at half maximum. The ratio of the number of signal events (S) to the number of signal plus background events (S+B) is shown on the right hand side. ... 152
- Figure 5-93 (2016) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation. 153
- Figure 5-94 (2017) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation. 154
- Figure 5-95 (2018) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation. 154

Figure 5-96 (Run 2) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation.	155
Figure 5-97 (Run 2) Expected and observed exclusion limits (95% CL, in the asymptotic approximation) for the ggH plus $t\bar{t}H$ (top left) and VBF plus VH (top right) processes, 100% VBF production hypothesis (bottom left) and the VH production hypothesis (bottom right). The inner and outerbands indicate the regions containing the distribution of limits located within ± 1 and $\pm 2\sigma$ of the expectation under the background-only hypothesis.	156
Figure 5-98 (Run 2) Observed local p -values for an additional SM-like Higgs boson as a function of m_H from 2016, 2017, 2018 and the full Run 2 combination.	157
Figure 5-99 Full Run 2 channel compatibility per event class with the categories belong to each corresponding year.	157
Figure 5-100 Full Run 2 channel compatibility per year.	158
Figure 5-101 Events in all classes of the combined 13 TeV dataset together with the fit of signal-plus-background model, under a mass hypothesis of 95.4 GeV. The left plot is unweighted and in right plot, each event is weighted by the ratio $S/(S+B)$ for its event class, where S and B are the numbers of expected signal and background events.	158
Figure 6-102 Full layout of STXS stage 1.2 bins.	168
Figure 6-103 Parametric signal shapes for each year of simulation as well as for the sum of all years.	174
Figure 6-104 Each analysis category composition in terms of a reduced set of STXS bins.	176
Figure 6-105 Data points in black and fit on the signal plus background model for all categories summed are shown. Each category is weighted by $S/(S+B)$	177
Figure 6-106 Observed results of the fit to four production mode signal strength modifiers.	178
Figure 6-107 Observed results of the maximal merging scheme STXS fit.	180
Figure 6-108 Observed results of the minimal merging scheme STXS fit.	180
Figure 6-109 Observed correlations in the "maximal" merging fit between the 17 parameters.	181
Figure 6-110 Observed correlations in the "minimal" merging fit between the 27 parameters.	182
Figure 6-111 Two dimensional likelihood scan (observed) in the κ -framework κ_v -vs- κ_f in the resolved κ model.	183

Figure 6-112 Two dimensional likelihood scan (observed) in the κ -framework κ_g -vs- κ_γ in the unresolved κ model.	183
Figure 6-113 A general layout of chained approach for some input variables for quantile BDT's. The variables to be corrected are also given.	187
Figure 6-114 The distribution of isolation sum of photon in data (black dots). The histogram in green shows uncorrected distribution and after equalizing number of events having zero isolation in MC with data is in orange colour. The distribution in purple is the one after the application of equalizing step and CQR method to tail part.	187
Figure 6-115 Photon ID MVA distribution for probe candidate in tag and probe ($Z \rightarrow ee$) sample for data and the MADGRAPH5_AMC@NLO simulation with barrel (left) and endcaps (right).	188
Figure 6-116 Signal models for fiducial cross section measurement for the best and worst resolution categories in 2018.	190
Figure 6-117 The event yields summed across all resolution categories divided by the total $H_{\gamma\gamma}$ cross section multiplied by the integrated luminosity for the bins in particle level and reconstruction level observables for the year 2018.	193
Figure 6-118 Likelihood scan for $H \rightarrow \gamma\gamma$ cross section in the fiducial region.	195
Figure 6-119 Invariant mass distribution of diphoton with all of the categories combined and used for the measurement of inclusive fiducial cross section.	195
Figure 6-120 Cross sections of $H \rightarrow \gamma\gamma$ in the selected regions of fiducial phase space.	196
Figure 6-121 Correlation matrix for the cross sections per particle level bin for $p_T^{\gamma\gamma}$, extracted by performing simultaneous maximum likelihood fit for cross sections.	197
Figure 6-122 Correlation matrix for the cross sections per particle level bin for n_{jets} , extracted by performing simultaneous maximum likelihood fit for cross sections.	198
Figure 6-123 Differential fiducial cross sections for $p_T^{\gamma\gamma}$ and n_{jets} . The black dots represents the observed values of differential fiducial cross sections and full uncertainty is shown with vertical bars.	198
Figure 6-124 Differential fiducial cross sections for $ y^{\gamma\gamma} $, $ \cos\theta^* $, $ \phi_\eta^* $ and τ_C^j observables. The black dots represents the observed values of differential fiducial cross sections and full uncertainty is shown with vertical bars. ...	199
Figure 6-125 Measured double-differential fiducial cross section in bins $p_T^{\gamma\gamma}$ and n_{jets}	200
Figure 6-126 Measured double-differential fiducial cross section in bins $p_T^{\gamma\gamma}$ and τ_C^j	201

Figure 7-127 (Unweighted) The likelihood scan of the expected Higgs boson mass in diphoton decay mode for an integrated luminosity of 3000 fb^{-1} . The solid lines are for the full likelihood scan including all systematic uncertainties, while the dashed lines shows the same with the statistical uncertainty only [34]. 206

List of Tables

Table 2-1 Various properties of the gauge bosons.	8
Table 2-2 Standard Model particles and related parameters.	9
Table 3-3 Parameters values designed for LHC operations.	27
Table 4-4 Criteria for tag and probe electrons selection for validation (all years).	56
Table 4-5 HLT paths for the tag electrons.	56
Table 5-6 Data samples used in the analysis, with their corresponding integrated luminosity per run-periods.	70
Table 5-7 Background samples used for training and validation purpose in 2017 and 2018 along with their relative number of events, cross-section (pb) and Equiv. Lumi. (pb^{-1}).	71
Table 5-8 HLT Trigger paths employed in 2016 data taking.	72
Table 5-9 HLT Trigger paths employed in 2017 and 2018 data taking.	72
Table 5-10 2016 preselection requirements.	80
Table 5-11 2017 preselection requirements.	81
Table 5-12 2018 preselection requirements.	81
Table 5-13 (2016) Preselection efficiencies measured in the three η , R_9 cate- gories using tag and probe with $Z \rightarrow ee$ events (for all preselection criteria except electron veto).	82
Table 5-14 (2016) Loose photon ID MVA cut efficiencies measured in the three η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all ID MVA criterion except electron veto).	82
Table 5-15 (2017) Preselection efficiencies measured in the four η , $R_9 <$ cate- gories using tag and probe with $Z \rightarrow ee$ events (for all preselection criteria except electron veto).	82
Table 5-16 (2017) Loose photon ID MVA cut efficiencies measured in the four η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all ID MVA criterion except electron veto).	83
Table 5-17 (2018) Preselection efficiencies measured in the three η , R_9 cate- gories using tag and probe with $Z \rightarrow ee$ events (for all preselection criteria except electron veto).	83
Table 5-18 (2018) Loose photon ID MVA cut efficiencies measured in the three η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all ID MVA criterion except electron veto).	83

Table 5-19 (2016) Efficiency of the pixel-seed electron veto, measured in the three photon categories using tag and probe with 2016 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.	84
Table 5-20 (2017) Efficiency of the pixel-seed electron veto, measured in the four photon categories using tag and probe with 2017 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.	84
Table 5-21 (2018) Efficiency of the pixel-seed electron veto, measured in the three photon categories using tag and probe with 2018 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.	84
Table 5-22 (2016) Efficiency of the “ $N_e^{Matched} = 0$ ” requirement, measured in the three photon categories using tag and probe with 2016 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.	98
Table 5-23 (2017) Efficiency of the “ $N_e^{Matched} = 0$ ” requirement, measured in the four photon categories using tag and probe with 2017 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.	98
Table 5-24 (2018) Efficiency of the “ $N_e^{Matched} = 0$ ” requirement, measured in the three photon categories using tag and probe with 2018 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.	98
Table 5-25 (2016) Fit parameter values for the double-fake Drell-Yan events. .	127
Table 5-26 (2017) Fit parameter values for the double-fake Drell-Yan events. .	127
Table 5-27 (2018) Fit parameter values for the double-fake Drell-Yan events. .	127
Table 5-28 (2016) Fitted values of μ (GeV) from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 3 event classes.	143
Table 5-29 (2016) Fitted values of σ from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 3 event classes.	143
Table 5-30 (2017) Fitted values of μ (GeV) from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.	143
Table 5-31 (2017) Fitted values of σ from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.	144
Table 5-32 (2018) Fitted values of μ (GeV) from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.	144
Table 5-33 (2018) Fitted values of σ from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.	144
Table 5-34 (2016) Values of μ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$	145
Table 5-35 (2016) Values of σ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class.	145

Table 5-36 (2017) Values of μ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$	145
Table 5-37 (2017) Values of σ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$	145
Table 5-38 (2018) Values of μ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$	146
Table 5-39 (2018) Values of σ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$	146
Table 5-40 Fit values of α_L and α_R while floating in the fitting of the double-fake DY events.	147
Table 5-41 (2016) Central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R	147
Table 5-42 (2017) Central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R	147
Table 5-43 (2018) Central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R	148
Table 5-44 The expected number of SM-like Higgs boson signal and background events ($m_H = 90$ GeV) per event class and the corresponding percentage breakdown per production process.	151
Table 6-45 Description and the method used for defining the categories of events.	166
Table 6-46 The efficiencies of the photon ID MVA and σ_m^D boundaries using signal sample for 2016, 2017 and 2018 data taking.	190
Table 6-47 Fiducial phase space definition where labels 1, 2 refer to the leading and subleading photon.	191
Table 6-48 Dominant systematic uncertainties and impacts.	194

Chapter 1: Introduction

Fundamental structure of matter and its interactions has been efficiently described by standard model (SM) of particle physics in the context of the Quantum Field Theory [1, 2]. The electromagnetic and weak forces are unified by the SM and position them in a same framework along with the strong force. The SM has been unprecedentedly successful, with its predictions verified to extraordinary precision. The electro-weak symmetry is broken spontaneously by an additional scalar field (the Higgs field) and Higgs mechanism is also a key aspect of SM. It explains how particles gain mass and predicts the existence of a fundamental particle known as the Higgs boson. The Higgs boson was discovered experimentally on 4th July 2012 by the ATLAS and CMS Collaborations [4, 5] at the Large Hadron Collider (LHC) [3].

The Standard Model can not be considered as a complete theory of nature. It does not identify the suitable candidate of dark matter which is a key part to explain certain astrophysical phenomena [35] as well as does not include any description regarding the force of gravity. Neutrinos are considered to be massless in SM but they are not, as confirmed by the experimental measurements of neutrino oscillations [36]. Many of these phenomena have been proposed by the wide range of Beyond Standard Model (BSM) theories. Entirely new particles are predicted by these theories and eventually amend the predictions of SM [37].

In this thesis, the latest measurements of the properties of the Higgs boson in the diphoton final state are presented. The objective of the measurements of Higgs boson property is to check how compatible these are w.r.t. the SM predictions. Any deviations from the SM predictions will indicate new physics beyond the Standard Model. Events are selected from the data of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the CMS detector at the LHC during the Run 2 data taking period in years 2016, 2017 and 2018, corresponding to an integrated luminosity of 137.1 fb^{-1} . Relative to the standard model (SM) prediction, the total signal strength of the Higgs boson is measured to be 1.12 ± 0.09 [25]. Other measured properties include SM signal strength modifiers, production cross section and its coupling to other particles. The simplified template cross section (STXS) framework provides a reasonable approach to perform precision

measurement of the Higgs boson. The main motivation is to minimise the theory dependence of Higgs boson measurements and reduce the direct impact of SM predictions on the results [25]. The inclusive and differential fiducial cross section measurements in diphoton decay mode of the Higgs boson, are also reported with the data collected at 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1} . The inclusive fiducial cross section is measured as $\sigma_{\text{fid}} = 73.40^{+5.4}_{-5.3}(\text{stat})^{+2.4}_{-2.2}(\text{syst}) \text{ fb}$ which is in agreement with the expectation of standard model $75.44 \pm 4.1 \text{ fb}$. The estimations are also carried out in fiducial regions for different production mechanisms w.r.t. observables that describe the system of pair of photons, count of the additional jets and event level quantities. No significant deviations are observed relative to the standard model expectations [38].

The main part of this thesis is to search for an additional SM-like Higgs boson in the diphoton final state. This work is to probe for a new low-mass Higgs boson in the invariant mass range of $70 < m_H < 110 \text{ GeV}$ with 2018 (54.4 fb^{-1}) data and full Run2 combination (2016, 2017, 2018) corresponding to an integrated luminosity of 132.2 fb^{-1} . The analysis searches for a localized excess of diphoton events over a smoothly falling background as a result of prompt diphoton production and to events with at least one jet misidentified as a photon. It extends the method developed in CMS for the observation and the measurement of the properties of the SM-like Higgs boson discovered in 2012 [39]. In order to achieve the best sensitivity, the events are separated into classes and Multivariate analysis (MVA) techniques are used both for photon identification and event classification. The observed 95% confidence level upper limits on the product of the cross section times branching ratio into two photons at 13 TeV, ranges from 15 fb at 108.9 GeV to 73 fb at 95.4 GeV. No significant excess is observed.

The structure of this thesis is as follows:

- **Chapter 2:** Theoretical background of the Standard Model Higgs boson and LHC physics is described.
- **Chapter 3:** Overview of the Large Hadron Collider (LHC) and Compact Muon Solenoid (CMS) detector is addressed.
- **Chapter 4:** Physics objects reconstruction, identification and resolution of photons and some other objects in the CMS detector is explained.
- **Chapter 5:** Probe for an additional SM like Higgs (low mass) boson and results in the diphoton decay channel are presented.

- **Chapter 6:** Higgs boson properties measurement and results in the diphoton decay channel are reported.
- **Chapter 7:** Summary of the dissertation and outlook.

Chapter 2: Standard Model and LHC physics

2.1 The Standard Model

The best theory in explaining high-energy experimental data at fundamental level is known as the standard model (SM) of particle physics. It unifies the electromagnetic, weak and strong forces in a single framework. The SM is a type of quantum field theory which provides modern understanding for the fundamental constituents of matter and their interactions with each other. Many of the SM projections have been verified experimentally to remarkable precision [40]. The question which remains open is about the origin of the masses of the W and Z boson. It has been answered in the SM by the mechanism of electroweak symmetry breaking by introducing an additional scalar field whose quantum is the Higgs boson [19, 20] which is now experimentally observed [4, 5].

2.1.1 Theoretical Overview

The elementary particles of SM are all known and so are the electromagnetic, weak and strong interactions between them [15–17]. It can be described as a mathematical model expanded in context of the Quantum Field Theory (QFT) under a common theoretical framework to unify three of the four fundamental forces of nature i.e. the electromagnetic, strong and the weak forces with the exception of gravity.

A collaborative work spanned over the 20th century formulated standard model of particle physics which is a quantum field theory represented by the gauge symmetry group as follows:

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \quad (2-1)$$

Enrico Fermi in 1933 was the first to reveal the weak interaction by understanding of beta decay called Fermi's interaction [18]. In 1961, from Glashow, it was proposed that weak and electromagnetic interactions are alike and provide the theoretical unification of these forces to the electroweak interactions represented by $SU(2)_L \otimes U(1)_Y$ symmetry group. The theory from this symmetry group was derived by Glashow [15], Weinberg [16] and Salam [17] which provides the existence and behaviour of the electroweak mediators is (the Z^0 and $W^\pm$ bosons) and the carrier of electromagnetic force,

the photon (γ). The $SU(3)_c$ term is used to derive the strong force carriers. Glashow, Salam and Weinberg were awarded with the Physics Nobel Prize for work in electroweak unification. The fundamental particles acquire mass with a mechanism proposed by Brout, Englert, Higgs *et al.* based on the idea of spontaneous symmetry breaking and named the new elementary particle known as Higgs boson. Weinberg and Salam then connected the idea to Glashow's model ($SU(2)_L \otimes U(1)_Y$) and eventually completed the electroweak sector of standard model.

At the same time, Quantum Chromodynamics (QCD) area of the standard model was evolving and there were signs of the presence of quarks at various scattering experiment in 60's. The existence of J/ψ meson further confirmed the presence of quarks. Friedman, Richter and Ting were awarded with the Nobel prizes after the discoveries of quark and charmonium. The bottom and anti-bottom quarks were discovered in 1976 and after that, a τ particle was also discovered. Fermilab finally discovered the top quark in 1995.

Strong force has very high magnitude compared to others and because of this, isolated quarks are impossible to find at very high energies. Thus, the quarks produced in the collisions are considered as jets of particles originated by the hadronization process. Strong interaction are explained by QCD which confines quarks into hadrons. The most generally recognized hadrons are the protons and neutrons which are responsible for making nuclei of matter. The protons and neutrons are made up of quarks i.e. uud and udd respectively. Mesons ($q_1 \bar{q}_2$) are made by the combination of quark-antiquark pair. The baryons are formed by the combination of three quarks ($q_1 q_2 q_3$), to make color-neutral hadrons. Partons are the only elementary particles which exhibit non-vanishing color charges. All these discoveries provided a proof for the SM predictions. Finally the Higgs bosons which remained un-discovered was reported in July 2012 by CMS and ATLAS collaborations at CERN [4, 5]. The Physics Nobel Prize in 2013 was awarded to Higgs and Englert for their theoretical predictions about this gauge boson, the Higgs boson.

2.1.2 Elementary Particles

In the SM, the fundamental particles except Higgs boson are categorized into spin-half fermions and spin-one bosons. Fermions are the elementary constituents of mat-

ter and can interact via electromagnetic and weak forces (leptons), or electromagnetic, weak and strong forces (quarks). Fermions obey Fermi-Dirac statistics and bosons are characterized by Bose-Einstein statistics. Leptons and quarks exist in three well defined generations. In each family, three particles have very same properties except for mass, which increases over generations. Additionally, every particle posses an antiparticle which has identical mass but with opposite charge and parity to that of the original particle. Leptons are given the number "1" and their antiparticles receive the number "-1," respectively. This quantity (lepton number) is conserved in SM processes.

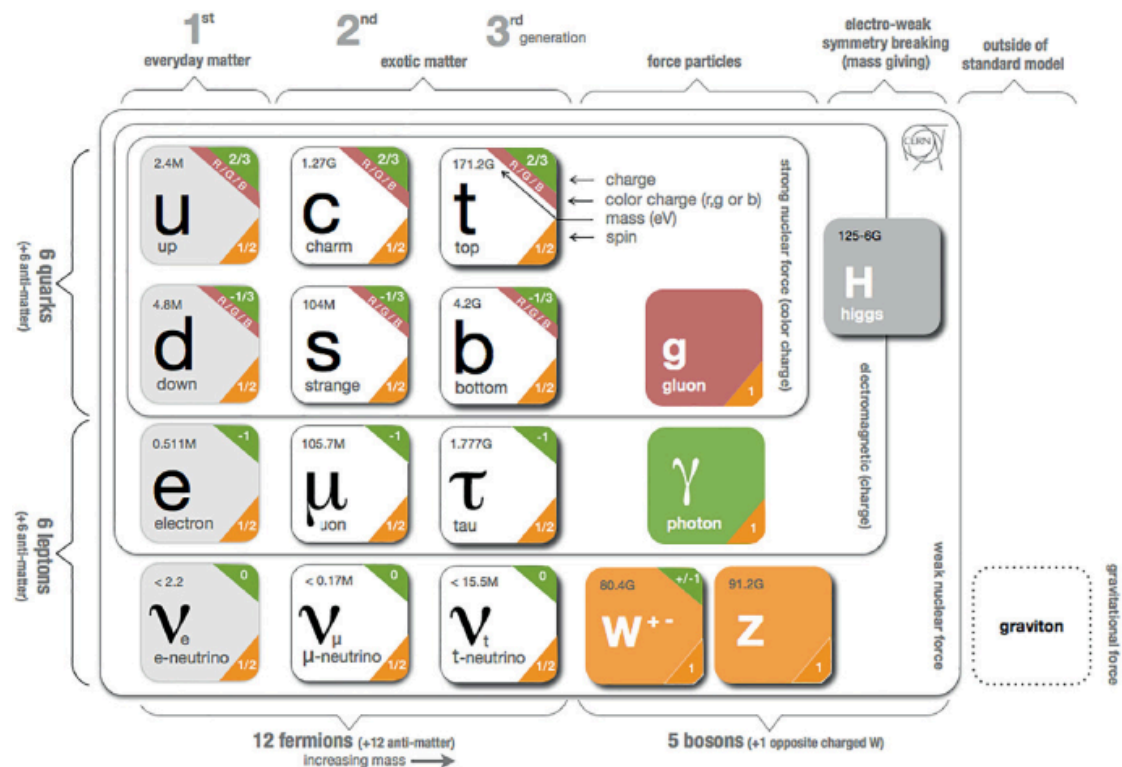


Figure 2-1 Elementary particles in three generations, gauge bosons and the Higgs boson.

Quarks have fractional charges. It includes an "up" quark having electric charge $+\frac{2}{3}$ and a "down" carrying electric charge $-\frac{1}{3}$ in the first generation. They also carry color charge (red, green, or blue) through which they interact strongly. The quarks in ascending mass order are "up," "down," "strange," "charm," "bottom," and "top". Heavy unstable second and third generation quarks can be seen only in high energy collisions and decay rapidly into 'up' and 'down' quarks. The color charge is also a characteristic of these quarks which is a conserved quantity in all of the standard model processes.

Bosons are the fundamental force carriers and are responsible for the interactions between fermions. Three forces, the electromagnetic force, weak force and the strong force are represented by the SM gauge bosons. Massless photon is the mediator of the electromagnetic force while the weak interactions occurs via exchange of W^+ , W^- , and Z bosons. Gluons, the mediators of strong force, are massless. The SM Higgs boson, is the only spin zero (scalar) particle in the theory and is a massive particle arises as a result of spontaneous symmetry breaking in the electroweak sector. Its existence is necessary to explain the masses of both bosons and fermions. The graphical representation of fundamental particles is shown in Figure 2-1 [41].

2.1.3 Fundamental Interactions

Fundamental interactions commonly known as fundamental forces, are considered to be the forces which mediate these interactions and can not be moved to more basic level. There are four basic forces i.e. the strong force, weak force, electromagnetic force and the gravitational force.

Table 2-1 Various properties of the gauge bosons.

Force	Electromagnetic	Weak	Strong
Mediator	photon (γ)	$W^\pm$ and Z	Gluon (g)
Mass (GeV)	0	80.4, 91.2 GeV	0
Coupling constant	$\alpha = \frac{1}{137}$	$G_F \approx 1.2 \times 10^{-5} \text{ GeV}^{-2}$	$\alpha_s \approx 0.1$
Interaction range	∞	10^{-16} cm	10^{-13} cm

The electromagnetic and weak interactions are described by electroweak theory and strong interactions are described by Quantum Chromodynamics (QCD) in the Standard Model of particle physics. Neutrinos are neutral electrically with no color charges and only exhibit weak interactions. Although the SM is considered to be theoretically self-consistent and demonstrated remarkable success in giving experimental predictions, but it does not provide the full theory of gravitation.

During the understanding of beta decay by Enrico Fermi in 1933, weak interaction was first revealed which is later on considered as one of the four fundamental forces. Mediators of the weak force are $W^\pm$ and Z bosons. The force which is able to alter quarks flavour is the weak force and is rarely known as quantum flavordynamics (QFD).

Table 2-2 Standard Model particles and related parameters.

Family and generation	Name	Symbol	Mass	Charge[e]	Interactions
leptons (1 st)	electron	e	0.511 MeV	-1	weak, EM
leptons (1 st)	electron neutrino	ν_e	$<2 \text{ eV}$	0	weak
leptons (2 nd)	muon	μ	105.658 MeV	-1	weak, EM
leptons (2 nd)	muon neutrino	ν_μ	$<0.19 \text{ MeV}$	0	weak
leptons (3 rd)	tau	τ	1776.82 MeV	-1	weak, EM
leptons (3 rd)	tau neutrino	ν_τ	$<18.2 \text{ MeV}$	0	weak
quarks (1 st)	up	u	2.3 MeV	$+\frac{2}{3}$	EM, weak, strong
quarks (1 st)	down	d	4.8 MeV	$-\frac{1}{3}$	EM, weak, strong
quarks (2 nd)	charm	c	1.275 GeV	$+\frac{2}{3}$	EM, weak, strong
quarks (2 nd)	strange	s	95 MeV	$-\frac{1}{3}$	EM, weak, strong
quarks (3 rd)	top	t	173.07 GeV	$+\frac{2}{3}$	EM, weak, strong
quarks (3 rd)	bottom	b	4.18 GeV	$-\frac{1}{3}$	EM, weak, strong

The symmetry group $SU(2)$ is used to represent the weak interaction in terms of a field theory with three gauge bosons (W^+ , W^- , Z). As these bosons ($W^\pm$, Z) are massive, so weak force is a very short range force and the Higgs boson and Fermions interact only through weak force. A milestone was achieved in 1970s by Sheldon Glashow, Abdus Salam, and Steven Weinberg as they unified the electromagnetic and weak forces and they were awarded with the Physics Nobel Prize for their role in electroweak unification. The experimental existence of electroweak interactions was fulfilled in two steps. In 1973, the presence of neutral current was reported for the first time in neutrino scattering by Gargamelle experiment, which later on leads to the discovery of $W^\pm$ and Z bosons in proton-antiproton collisions by UA1 and UA2 collaborations in 1983 [42–44].

In SM, the theory of strong interactions which is known as quantum chromodynamics (QCD), is represented by the $SU(3)$ symmetry group. These symmetries can be identified by traceless 3×3 unitary matrices which implies that, there are eight independent matrices of the group [45]. The covariant derivative is written as

$$D_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} A_\mu^a \quad (2-2)$$

where λ^a are the Gell-Mann matrices and the eight fields A_μ^a corresponds to the massless gluons, the carriers of the strong force. The colour charge g_s is the QCD equivalent of electric charge. The only SM fermions which possess colour charge are quarks. Ad-

ditionally, $SU(3)$ is a non-Abelian group, which means that its transformations do not commute.

2.1.4 Electroweak Unification

As it was proved by Glashow that electromagnetic and weak interactions are represented as $SU(2)_L \otimes U(1)_Y$ group. ‘ L ’ stands for left handed chiral fermionic field states and ‘ Y_W ’ stands for weak hyper charge which is related to the electric charge (Q) and weak-isospin (T) as:

$$Q = T_3 + \frac{Y_W}{2} \quad (2-3)$$

The equation (2.3) is famous Gell-Mann-Nishijima relation. It can be applied to right and left-handed fermions. The left-handed fermions are represented by $SU(2)_L$ doublets ψ_R while the right-handed fermions are singlets ψ_L . This was done in the Glashow-Weinberg-Salam model by constructing the $SU(2)_L \otimes U(1)_Y$ group with weak isospin and weak hypercharge Y_W as the respective charges. Three vector fields W_μ and the neutral field B_μ couples with strength $\frac{g'}{2}$ to weak hyper-charge. The covariant derivative, in order to establish Lagrangian, is stated as:

$$D_\mu = \partial_\mu + ig\mathbf{T} \cdot \mathbf{W}_\mu + \frac{ig'}{2}Y B_\mu \quad (2-4)$$

For the following transformation, the covariant derivative can be operated on the left-handed and right-handed states:

$$D_\mu \chi_L \rightarrow e^{i\alpha(x)\cdot\mathbf{T} + i\beta(x)Y} D_\mu \chi_L \quad (2-5)$$

$$D_\mu \psi_R \rightarrow e^{i\beta(x)Y} D_\mu \psi_R \quad (2-6)$$

The outcome of this transformation, $\bar{\psi} D_\mu \psi$ is invariant:

$$\mathbf{W}_\mu \rightarrow \mathbf{W}_\mu - \frac{1}{g} \partial_\mu \alpha - \alpha \times \mathbf{W}_\mu \quad (2-7)$$

$$B_\mu \rightarrow B_\mu - \frac{1}{g'} \partial_\mu \beta. \quad (2-8)$$

The Lagrangian is as follows:

$$\begin{aligned}\mathcal{L}_{EWK} = & \bar{\chi}_L(x)i\gamma^\mu(\partial_\mu + ig\mathbf{T} \cdot \mathbf{W}_\mu + \frac{ig'}{2}YB_\mu)\chi_L(x) + \\ & \bar{\psi}_R(x)i\gamma^\mu(\partial_\mu + i\frac{ig'}{2}YB_\mu)\psi_R(x) + \\ & \frac{1}{4}W_{\mu\nu}W^{\mu\nu} + \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\end{aligned}\quad (2-9)$$

the kinetic energy terms of the gauge fields are represented by $W_{\mu\nu}$ and $B_{\mu\nu}$ and their equation are referred as: 2-10 and 2-11 respectively.

$$W_{\mu\nu} \equiv \partial_\mu W_\nu - \partial_\nu W_\mu - gW_\mu \times W_\nu \quad (2-10)$$

$$B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \quad (2-11)$$

where $\mathbf{T}$ is $SU(2)$ generator and linked to Pauli matrices:

$$T^a = \frac{1}{2}\sigma^a; a = 1, 2, 3 \quad (2-12)$$

Electroweak lagrangian (in 2-9) report massless lepton. The mass term is chosen as zero to remain invariant under transformations in the equations 2-7 and 2-8 because of the fact that, the theory remain unable to report massive gauge bosons kinematics. Later on, Salam and group, reported a method to incorporate mass term in the electroweak Lagrangian (in equation 2-9).

2.1.5 Spontaneous symmetry breaking and Higgs mechanism

The mechanism of spontaneous symmetry breaking address the process to give mass to the particles which are gauge invariance of the lagrangian. The proposal of Glashow for the unification of EM and weak forces describe the kinematics, only for the massless particles as explained in Section 2.1.4. The SM gives mass to the ($W^\pm$, Z boson) via spontaneous symmetry breaking in the electro-weak sector which is known as the Higgs mechanism. The Higgs mechanism introduces the existence of a new scalar field known as Higgs field [46–51]. This method establish a complex scalar field doublet:

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix} \quad (2-13)$$

and is replaced in the lagrangian as:

$$\mathcal{L}_{EWSB} = (D_\mu \phi)^\dagger (D_\mu \phi) + V(\phi^\dagger \phi) \quad (2-14)$$

where potential term is denoted by V and D_μ is expressed in the equation 2-4. The lagrangian (Eq. 2-14) is gauge invariant under $SU(2)_L \otimes U(1)_Y$ transformation. Potential depends only upon the product $\phi^\dagger \phi$ and the kinetic energy term is written in terms of the covariant derivatives. ϕ , the scalar field is stated in the form of quantum numbers i.e electric charge (Q), hypercharge (Y) and the isospin (I_3).

	I_3	Y	Q
$\begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

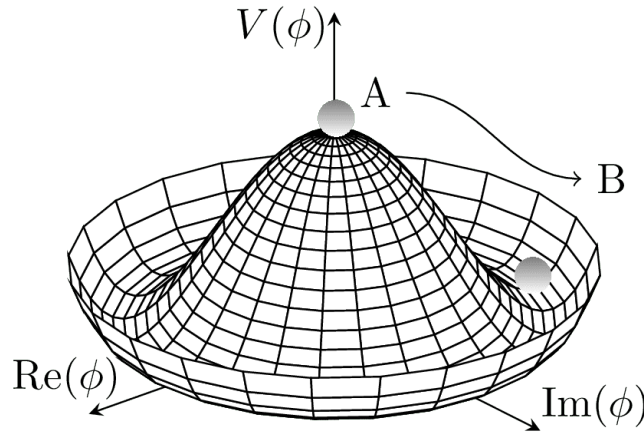


Figure 2-2 Representaion of Higgs potential.

The potential term in Equation 2-14 can also be written as

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2-15)$$

here μ and λ denotes the mass term and interaction strength respectively. The potential shape is like an interesting ‘mexican hat’ because $\mu^2 < 0$ and $\lambda > 0$ is core part for Higgs mechanism illustrated in the Figure 2-2.

$$\phi^\dagger \phi = \frac{1}{2}(\phi_1^\dagger + \phi_2^\dagger + \phi_3^\dagger + \phi_4^\dagger) = \frac{-\mu^2}{2\lambda} \equiv \frac{v^2}{2} \quad (2-16)$$

The arbitrary vacuum state that was selected is out of the symmetry center. So, there is no single value of ϕ for the minimum but with non-zero values. The $(\phi^\dagger \phi^0)$ consideration which is connected with arbitrary ground state and in the plane of $(\phi^\dagger \phi^0)$, the

considered point is not invariant which is a well known phenomena referred as spontaneous symmetry breaking. The ground state on ϕ^0 axis was settled down with the VEV (vacuum expectation value) of ϕ field is:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} = \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 = -\frac{\mu^2}{\lambda}. \quad (2-17)$$

Based on the vacuum expectation value, in term of generic gauge, the ϕ field below can be restated as:

$$\phi = \frac{1}{\sqrt{2}} e^{\frac{i}{v} \phi^a t_a} \begin{pmatrix} 0 \\ h(x) + v \end{pmatrix}, \quad a = 1, 2, 3 \quad (2-18)$$

ϕ_a and $\phi_4 = h(x) + v$ refers to the Goldstone fields. These fields are the massless fields. As a result of the polarizations of vector bosons $W^\pm$ and Z , latest degrees of freedom are defined because they are scalar and massless. A unitary gauge transformation in the field is

$$\phi' = e^{-\frac{i}{v} \phi^a T_a} \phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h(x) + v \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \phi^4 \end{pmatrix}$$

The expectation value is zero for Higgs field. from Equation 2-14, the Lagrangian of Higgs is stated as:

$$\begin{aligned} \mathcal{L}_{BEH} = & \frac{1}{4} g^2 v^2 W_\mu^+ W^{-\mu} - \frac{1}{2} (\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+) (\partial^\mu W^{-\nu} - \partial^\nu W^{-\mu}) + \\ & \frac{1}{8} (g^2 + g'^2) v^2 Z_\mu Z^\mu - \frac{1}{4} (\partial_\mu Z_\nu - \partial_\nu Z_\mu) (\partial^\mu Z^\nu - \partial^\nu Z^\mu) - \\ & \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} \partial_\mu h \partial^\mu h - \mu^2 h^2 + \dots \end{aligned} \quad (2-19)$$

the terms for W 's can be formulated as:

$$W^\pm = \left(\frac{W^1 \mp iW^2}{\sqrt{2}} \right) \quad (2-20)$$

These are for the oppositely charged W bosons. Their mass is:

$$m_W = \frac{1}{2} v g \quad (2-21)$$

for neutral Z boson, another term is:

$$Z_\mu = \frac{-g' B_\mu + g W_{3\mu}}{\sqrt{g^2 + g'^2}} \quad (2-22)$$

and mass of Z boson is:

$$m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \quad (2-23)$$

Using W_μ^3 and B_μ , the orthogonal combination to Z_μ can be restated as:

$$A_\mu = \frac{g' B_\mu + g W_\mu^3}{\sqrt{g^2 + g'^2}} \quad (2-24)$$

For the photon, having a condition, above equation is:

$$\frac{g' g}{\sqrt{g^2 + g'^2}} = e \quad (2-25)$$

Using equations 2-21 and 2-23, we can get,

$$\frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2 + g'^2}} \quad (2-26)$$

Suppose, the θ_W is the weak mixing angle,

$$g' = g \tan \theta_W \quad (2-27)$$

Equations 2-22 and 2-24 can be rewritten by using Equation 2-27 as:

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W \quad (2-28)$$

$$A_\mu = B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W \quad (2-29)$$

Spontaneous symmetry breaking (SSB) is giving masses to the gauge bosons (W, Z) via interaction with Brout-Englert-Higgs (BEH) or Higgs field. After SSB, Higgs fields emerges with a boson particle. Its potential term (Lagrangian) can be stated as:

$$\mathcal{L}_V = \frac{1}{4}\lambda(v + h(x))^4 - \frac{1}{2}\lambda v^2(v + h(x))^2 \quad (2-30)$$

$M_H^2 = 2\lambda v^2$ gives the first order mass of Higgs boson. λ is regarded as a free parameter in the SM.

If an additional term is added to electroweak lagrangian, fermions get masses by Higgs field interactions via Yukawa couplings. The term in the lagrangian is as follows:

$$\mathcal{L}_{Yukawa} = -G_l[\bar{\psi}_R(\phi^\dagger \chi_L) + (\bar{\chi}_L \phi)\psi_R] \quad (2-31)$$

With the help of ϕ and other parameters, the Higgs field provide fermion masses (m_f) as well as couplings (g_{Hf}):

$$m_f = \frac{G_l}{\sqrt{2}}v \quad (2-32)$$

$$g_{Hf} = \sqrt{G_l}\sqrt{2} = \frac{m_f}{v} \quad (2-33)$$

2.2 Higgs boson properties

One of the core physics objectives of the Large Hadron Collider (LHC) at CERN is to discover the Higgs boson and observe the various elements of the standard model. In 2012, LHC at CERN, achieved a milestone after the discovery announcement of a new particle by CMS and ATLAS collaborations. This particle resembles to the SM Higgs boson [4, 5]. Higgs boson was considered as the last missing piece of the SM. Various production and decay modes of the Higgs boson were analysed which confirmed the existence of the Higgs boson and therefore, the idea of Higgs field all over the space is strongly accepted. After that, to measure Higgs boson precise properties is the main experimental goal. Many stringent tests on the SM were then made in order to either observe the discrepancies from its predictions which might show the existence of the physics beyond SM scope i.e. BSM (Beyond Standard Model). In this section various productions and decay modes of Higgs boson at LHC are addressed.

2.2.1 Higgs boson production modes at LHC

The LHC produces many interesting particles as a result of collision between two circulating proton beams. The Higgs boson production can begin in various ways i.e. cross sections which based on the centre-of-mass energy ($\sqrt{s}$) of the p-p collision and mass of the Higgs boson. As we move higher towards the center of mass energies, remarkable increase in production cross section of Higgs is observed.

Gluon fusion (ggH) is the dominant production mode for collisions at 13TeV and for 125 GeV Higgs boson mass as shown in Figure 2-3. This (ggH) final state only

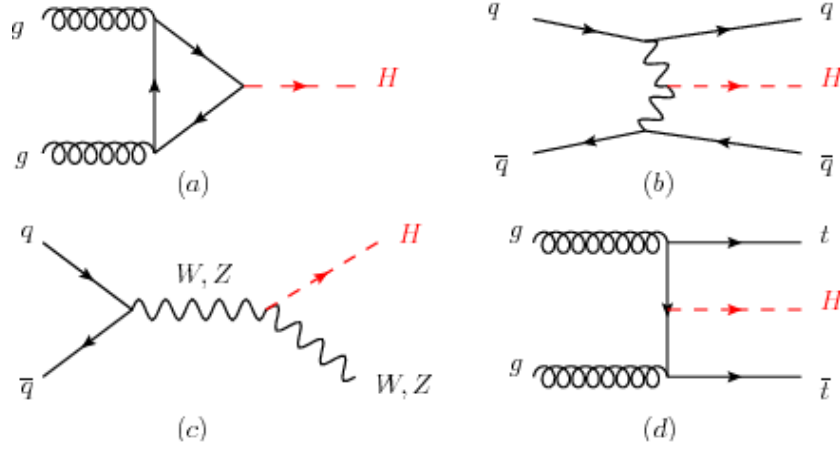


Figure 2-3 Feynman diagrams representing the four main Higgs production mechanism at tree level. a) Gluon-Gluon fusion, b) Vector Boson Fusion, c) Associated production W/Z and d) Associated top quark production.

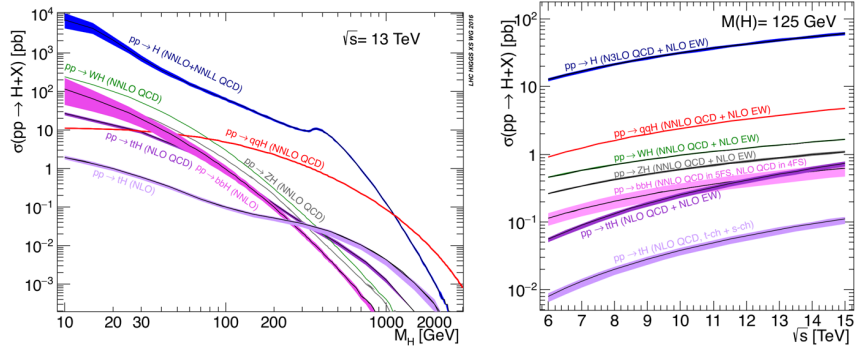


Figure 2-4 Production cross sections of Higgs. As function of mass illustrated left and on right, as function of centre of mass energies.

contains the Higgs boson [52]. Vector boson fusion (VBF) is the second most important production mode in which a Higgs boson is produced along with two quarks which hadronize in the form of jets. These jets help in the discrimination of Higgs boson signal from background which means, events exhibit similar detector signatures as that of the Higgs boson but comes from other SM processes. Its cross section is about 10 times smaller than ggH for 125 GeV Higgs boson and becomes comparable at $m_H = 1$ TeV.

In VH (W/Z associated production mechanism), Higgs boson is emitted via highly energetic W or Z bosons. The least possible process of Higgs production is, when two gluons interact and produce a pair of top antitop which gives Higgs in association with $t\bar{t}$. The ggH and VBF modes were observed during Run 1 of the LHC [53–55] and the

VH and $t\bar{t}H$ modes are being observed in the Run 2 [56–58].

2.2.2 Higgs boson decay modes

As predicted by the SM, the Higgs boson has very short lifetime [59]. Therefore, it can never be directly observed, but only from its decay products. There are several decay modes of the Higgs boson as shown in Figure 2-5 [60]. It can decay to the massive (onshell) particles as its coupling strength is proportional to the mass of the particle in which the Higgs boson is decaying.

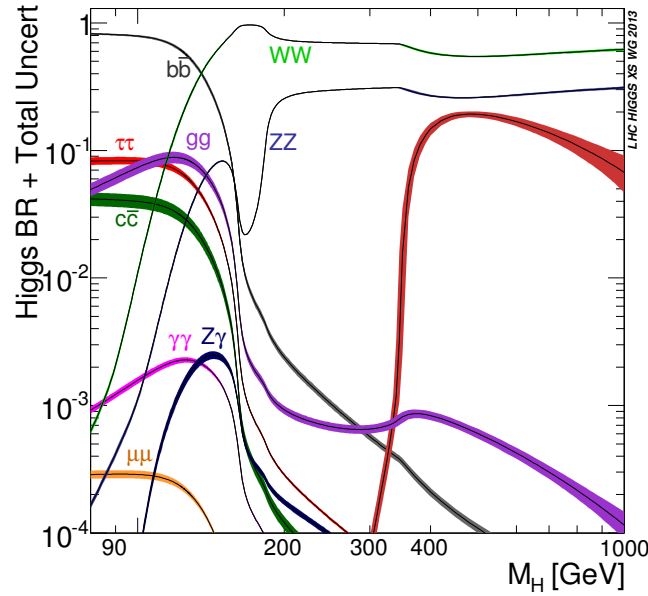


Figure 2-5 Branching ratios of Higgs boson for various channels.

Higgs boson can directly decay to any massive particle as well as loop diagrams allow it to decay in to massless particles like photons and gluons. The largest branching fraction, as can be seen from the Figure 2-5 [60] is for decay to $b\bar{b}$. The utilisation of this channel is quite difficult experimentally because of the large hadronic background at the LHC. The greatest sensitive channels but relatively low branching fractions, are the diphoton $\gamma\gamma$ and ZZ^* decays (where Z^* is off mass-shell, Z boson). These channels has good sensitivity because of relatively low background and the ability to reconstruct precisely the Higgs boson mass from its decay products. $H \rightarrow \gamma\gamma$ is one of the most suitable decay channel because of its clean final state with two highly energetic photons with remarkable mass resolution.

At the time of discovery of Higgs boson, $\gamma\gamma$ and ZZ^* decay channels were the important ones and because of this reason, favorable to probe its properties. For the Higgs boson coupling to the other particles like top quark and the W boson, diphoton decay channel is highly sensitive. As for the 125 GeV Higgs mass, its decay to WW , ZZ boson are also accessible, where one or both of the bosons can be off-shell. The WW further decays to neutrinos, so it's not possible to fully reconstruct the final state.

2.3 Review of Higgs boson property measurements and additional Higgs searches at LHC

In July 2012, after the discovery of Higgs boson of mass about 125 GeV by the ATLAS and CMS Collaborations at the CERN, many searches were triggered to measure its properties precisely. Almost eleven years later with the availability of data with thirty times larger number of Higgs boson led us to learn more about its properties. Several fermionic and bosonic decay modes have been probed by the CMS experiment to established its spin parity quantum numbers, to measure its mass and production cross sections precisely. This section presents the most up-to-date results on the Higgs boson properties by CMS Collaboration and also includes the stringent limit on the production cross section for the pair of Higgs bosons based on 13 TeV data [33].

At the time when Higgs boson was discovered [4, 5], with the combination (7 TeV plus 8 TeV) of data revealed the observed (obs.) significance of 5.0σ with an expected (exp.) significance of 5.8σ with the CMS experiment. Independently, in the most sensitive modes i.e. $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4l$ provided 4.1σ obs. (2.8σ exp.) and 3.2σ obs. (3.8σ exp.) respectively.

In Run 1 data, the separately observed significances in the bosonic decay channels was 6.5σ for $H \rightarrow ZZ \rightarrow 4l$, 5.6σ for $H \rightarrow \gamma\gamma$, 4.7σ for $H \rightarrow WW$ and 3.8σ for the fermionic decay channel $H \rightarrow \tau\tau$ [61]. The first results were reported in the Higgs boson decay to fermions in Ref. [62] with a significance of 3.8σ after the combination of $H \rightarrow \tau\tau$ and $H \rightarrow bb$ decay channels. The measured mass precision was $\approx 0.2\%$ [61].

With the help of Run 2 data, in the tau leptons pair decay of Higgs boson, CMS observed a significance of 5.9σ [63] and in the decay to a pair of bottom quarks with significance of 5.6σ [64]. For the $t\bar{t}H$ production mode it is 5.2σ [65]. The Higgs boson has been also seen in its decay to muons with significance of 3σ [66]. The Higgs boson

mass has been observed to be 125.38 ± 0.14 GeV using the decay channels $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4l$ [67]. The measured fiducial cross section at CMS (13 TeV) in $H \rightarrow \gamma\gamma$ decay channel is $\sigma_{\text{fid}} = 73.40^{+5.4}_{-5.3}(\text{stat})^{+2.4}_{-2.2}(\text{syst})$ fb and in ATLAS (13 TeV), it is $\sigma_{\text{fid}} = 67 \pm 6$ fb. For the results related to the STXS framework, the ATLAS collaboration has reported measurements at stage 1 [68–70] and then the enhancements in the framework were done leading to the most recent STXS bins definition i.e. stage 1.2 as shown in Figure 6-102. CMS has addressed the first results with STXS stage 1.2 cross sections [71].

The agreement between SM expectations and the signal yields (observed) can be estimated by fitting of data with the help of a model which introduces the parameters of signal strength. These are normally referred as μ . Figure 2-6 shows the extracted signal strength parameters of various production mechanism and decay modes.

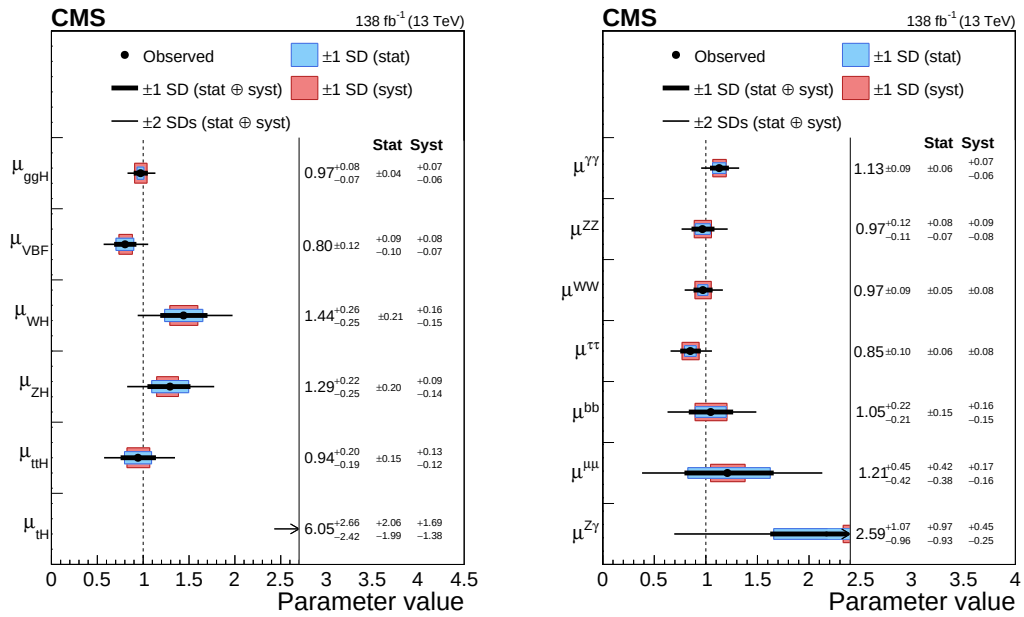


Figure 2-6 Signal strength parameters extracted for various production modes (left) and decay channels (right) [33].

The physics beyond SM is expected to influence in the production modes and decay channels if they are processed by identical interactions. To look deeper into such alterations from the SM expectations, the κ -framework [72] has been utilized. To the Higgs boson couplings, a first fit in this domain introduces two parameters, κ_f and κ_b , which scales the couplings of Higgs boson to massive fermions and gauge bosons respectively. At discovery time, such fit shows first indications for both types of couplings.

The sensitivity has been improved with the availability of present data and both of the coupling modifiers are observed to be in accordance with the SM expectations, within an uncertainty of 10%.

In the extensions of SM for new particles, the loop induced decays may get more contributions. In this context, a general fit has been performed to probe the deviations in the couplings of Higgs boson by establishing further modifiers for the Higgs boson coupling to gluons (κ_g), photons (κ_γ) and Z_γ (κ_{Z_γ}). The results for such fit are presented in Figure 2-7. The coupling modifiers are explored at uncertainty of 10%, except for κ_b and κ_μ which is $\approx 20\%$ and κ_{Z_γ} which is $\approx 40\%$. All of these measurements are in agreement with the SM expectations. The precision level of these measurements has been increased by a factor of ≈ 5 relative to which was possible with the discovery data set.

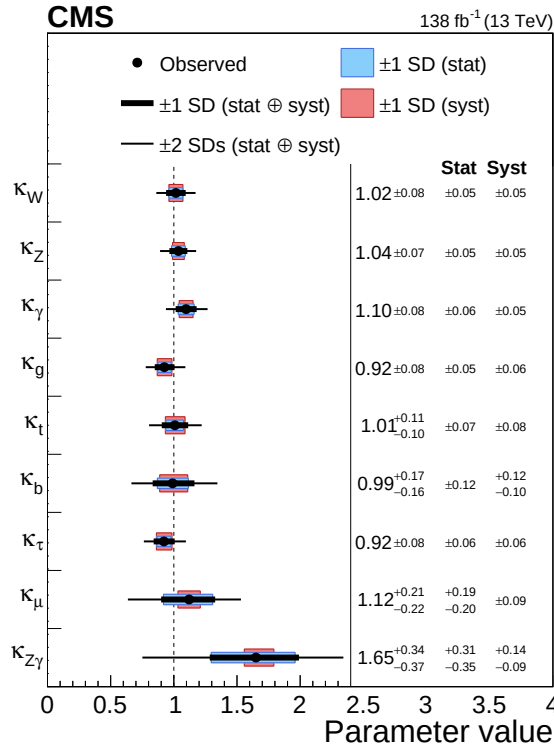


Figure 2-7 Coupling modifiers measurements [33].

The pair production cross section for the Higgs boson in SM is very small and that's why, so far, it's not detected in the LHC. Therefore, the results of the search are presented as the upper limit on production cross section. Figure 2-8 address the observed and expected limits on the production of Higgs boson pair relative to the predictions of SM. With the help of current data and from the combination of all modes and channels,

the cross section of Higgs boson pair production is observed to be 3.4 times less than the expectation of SM at 95% CL.

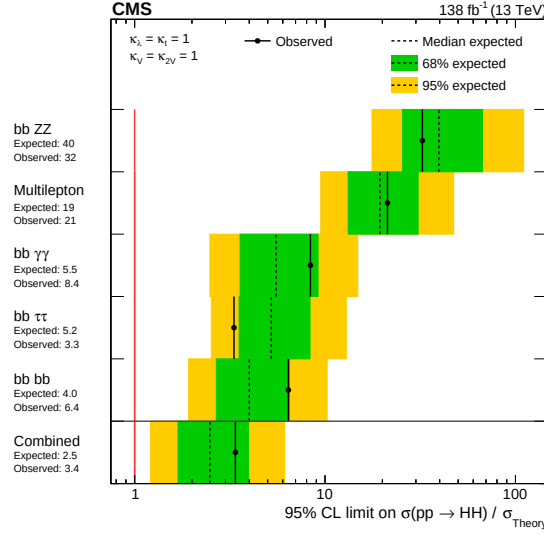


Figure 2-8 Limits on the production of Higgs boson pairs [33].

In the extensions of SM, several additional Higgs bosons may appear. The focal point is on the 2-Higgs-doublet models (2HDMs) where the main case of Higgs sector i.e. the minimal super-symmetric extension of the SM (MSSM). In the MSSM, the important production mechanism for a Higgs boson (charged) is the top quark decays if allowed kinematically, or in connection with a top quark. The neutral Higgs bosons are foreseen via gluon fusion dominantly. Non standard production and decay mechanism become authoritative in other 2HDMs. BSM Higgs sectors may offer more, like the NMSSM (next to MSSM) anticipate two neutral Higgs bosons in addition to MSSM which can be relatively light. Many searches have been evolved after SM Higgs bosons discovery for its precise measurement as well to probe for the additional Higgs boson in the low-mass regime. CMS has shown a variety of results by looking into several decay mechanism such as $H \rightarrow a_1 a_2 \rightarrow \mu\mu bb(\tau\tau bb)$ in the mass range $12 \text{ GeV} < m_{a1} < 62.5 \text{ GeV}$, $H \rightarrow \gamma\gamma$ in the mass range $70 \text{ GeV} < m_H < 110 \text{ GeV}$ and SM-like exotic $X \rightarrow e\mu$ ($110 \text{ GeV} < m_H < 160 \text{ GeV}$). These are the searches which has been recently presented with full Run 2 data of CMS.

In the current thesis, one chapter (Chapter 5) is dedicated to the low-mass $H_{\gamma\gamma}$ search. In the 13 TeV CMS data (35.9 fb^{-1}) in the year 2016 and 8 TeV data (19.7 fb^{-1}) in year 2012, a small excess was observed with a local significance of 2.8σ and global significance 1.3σ for a mass of 95.3 GeV [14]. The upper limit on the product of cross

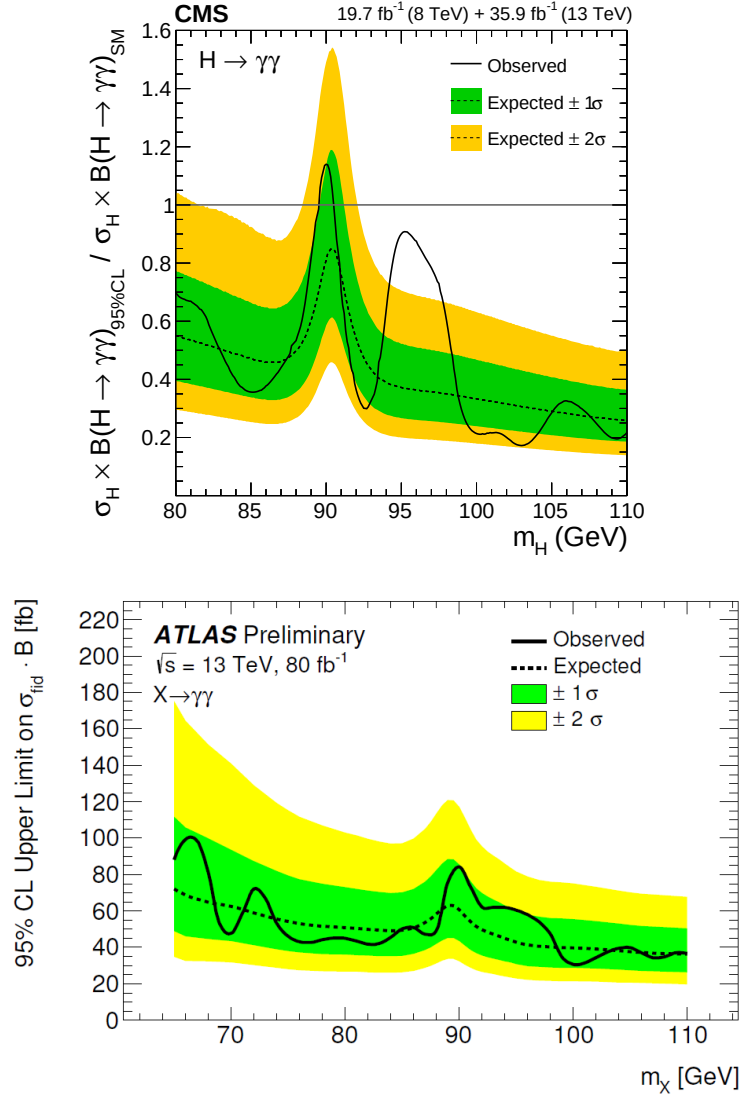


Figure 2-9 Observed and expected 95% CL upper limits of CMS (upper) and ATLAS (bottom).

section branching ratio observed by ATLAS using 80 fb^{-1} with 2016 and 2017 [73] is higher compared to the corresponding upper limit observed by CMS. The expected limit from ATLAS's result is $\approx 185 \text{ fb}$ at 70 GeV and $\approx 73 \text{ fb}$ at 110 GeV with 80 fb^{-1} data [74]. With the 2016 (legacy) and 2017 combination corresponding to 78.1 fb^{-1} which is close to ATLAS's 80 fb^{-1} data, we get the expected upper limit of about 70 fb at 70 GeV and about 31 fb at 110 GeV. So CMS has much better sensitivity from each year of Run 2 and the combinations, than ATLAS's results with 80 fb^{-1} data. No evidence so far, has been found for the existence of extra Higgs bosons.

Chapter 3: LHC and Compact Muon Solenoid (CMS) Experiment

3.3.1 Introduction

The LHC at CERN is located at France-Swiss border. It is a 27 km hadron accelerator, collider and designed for the center-of-mass energy of up to 14 TeV [21]. This collider (LHC) is one of its kind which is providing collisions at sufficiently high energies to probe physics precisely at the level of electroweak scale. There are mainly four detectors are present in the LHC ring and one of them is the Compact Muon Solenoid (CMS). The other three are the ATLAS (A Toroidal LHC Apparatus), ALICE (A Large Ion Collider Experiment) and LHCb (LHC-beauty). CMS is designed as a general-purpose detector for the observation of wide variety of physics objects such as SM precision measurements, dark matter searches and to probe electroweak symmetry breaking with the presence of Higgs boson. This chapter will mainly focus on the design and operation of LHC and the CMS detector.

3.3.2 The LHC layout

The main goal of the LHC design was to establish a new hadron collider in the 27 km long tunnel which exists already and was being used by the Large Electron Positron collider (LEP) [75] 100 meter underground at CERN near Geneva. The idea to utilise the existing tunnel gives rise to the possibility to reuse the part of the infrastructures already present i.e. the pre-accelerators etc. The LHC consist of 1232 main dipoles generating up to 8.33 T magnetic field to observe the particles into curvilinear trajectories. 386 quadrupoles, 360 sextupoles and 336 octupoles are also there for stability control. A schematic layout of LHC accelerating mechanism and the data it has collected so far is given below the Figure 3-10 [76].

The proton beams circulate in two separate rings in clockwise and counterclockwise directions. These rings are equipped with separate vacuum chambers and magnetic fields. The intersections of these rings only occur at the points where the detectors are installed. The ATLAS and CMS, detectors of LHC, are constructed exactly opposite to each other (at P1 and P5 respectively). As shown in the Figure 3-10 [76], there are two other important detectors, ALICE and LHCb, are installed (at P2 and P8 respectively)

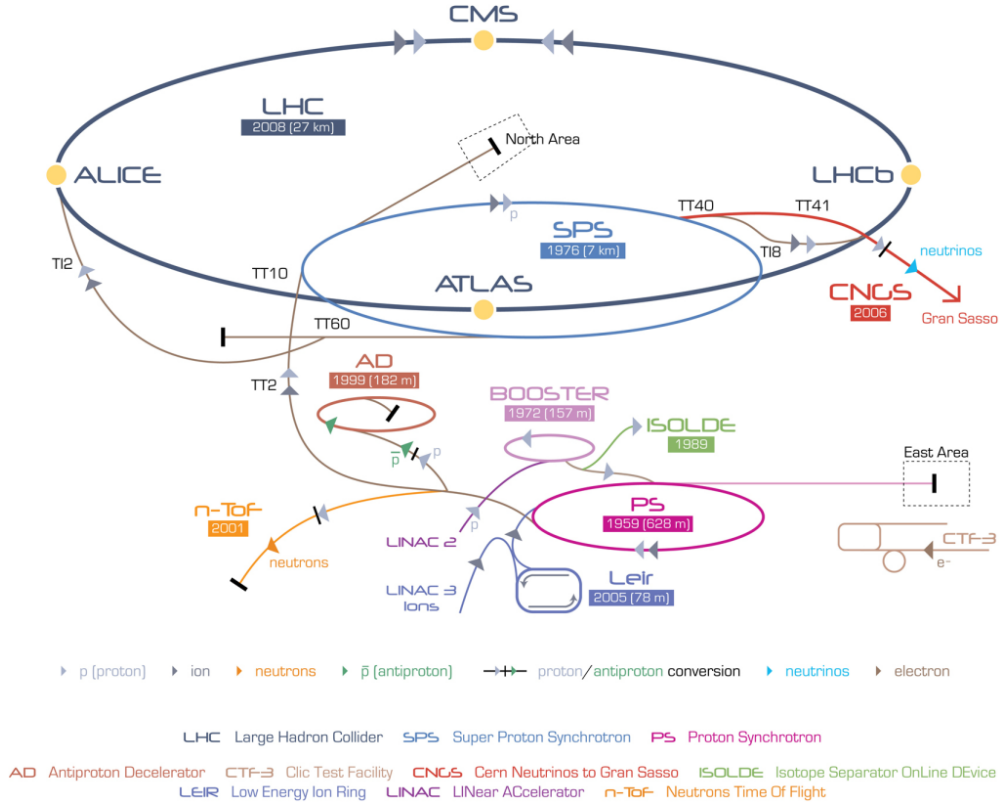


Figure 3-10 Layout of LHC and its experiments in the CERN accelerator complex. The PS and the SPS also included through which the proton pass before reaching the LHC ring.

and dedicated to probe heavy ion Physics and B Physics (b quark) respectively. Three smaller experiments i.e. MoEDAL, TOTEM and LHCf [77–79] are designed to perform studies of forward physics and magnetic monopoles searches.

3.3.3 Machine operation and performance

The design of LHC allows to collide the protons at a maximum centre-of-mass energy of 14 TeV. The beams of proton are accelerated in many stages before they enter the collision rings. A strong electric field is utilized in order to obtain the protons by stripping the electrons from hydrogen atoms and then they are accelerated up to 50 MeV by LINAC 2 (Linear Accelerator 2). In Proton Synchrotron Booster (PSB), the protons reaches to an energy of 1.4 GeV before the beam is passed to the Proton Synchrotron (PS). Here the beam is accelerated to 25 GeV and being passed to Super Proton Synchrotron (SPS). The SPS is the last step before the beam enters into the LHC tunnel and then the LHC is responsible for the final acceleration from 540 GeV to 6.5 TeV per beam for the collisions. The interaction occurs at four points where the detectors are

located.

In 2008, beam of protons were injected into LHC for the first time and the test to circulate the beams were completed on the same day. After a small repair on the bending magnets, in November 2009, LHC achieved 1.18 TeV energy each beam and became the World's strongest particle accelerator thus defeating the Tevatron's 0.98 TeV. In 2010, the center of mass energy achieved by LHC was up to 7 TeV (3.5 TeV each beam). 7 TeV and 8 TeV data was delivered by LHC in 2011 and 2012 later on respectively. LHC machine was shutdown during 2013-2014 for upgrade of its major parts of LHC machine and detectors. It started again in 2015 at 13 TeV which is close to its full expected center of mass energy (14 TeV). LHC machine was shut down in 2018 for several major upgrades after successfully delivering data during Run 2. During the long shutdown period 2 (LS2) from 2019 to 2021, several detector components for all of the LHC experiments were upgraded. Some of the upgradations includes, the improved diodes electrical insulation for more than 1200 magnets, new collimators has also been installed, neutral beam absorber for LHCb experiment and the replacement of 22 superconducting magnets etc. After the LS2, the LHC has started taking data for Run 3 from 2022 with 13.6 TeV proton-proton collisions.

The ultimate goal of LHC is to precisely test the SM Physics and hunt for new physics up to 14 TeV. In the interaction process, number of events (N) per unit time (called event rate) for any kind of process produced is given as

$$\frac{dN}{dt} = L \times \sigma_{process} \quad (3-1)$$

where L is the instantaneous machine luminosity and $\sigma_{process}$ is the cross section of the process. Time integral gives the total number of events of the process,

$$N = \int L dt \times \sigma_{process} = \mathcal{L} \sigma_{process} \quad (3-2)$$

Here $\mathcal{L}$ is *integrated* luminosity of the LHC. The cross sections of the possible processes produced during proton-(anti) proton collisions at different center-of-mass energy ($\sqrt{s}$) are shown in Figure 3-11. It is understandable that by increasing the luminosity and center-of-mass energies would obviously lead us to get more statistics of Higgs boson.

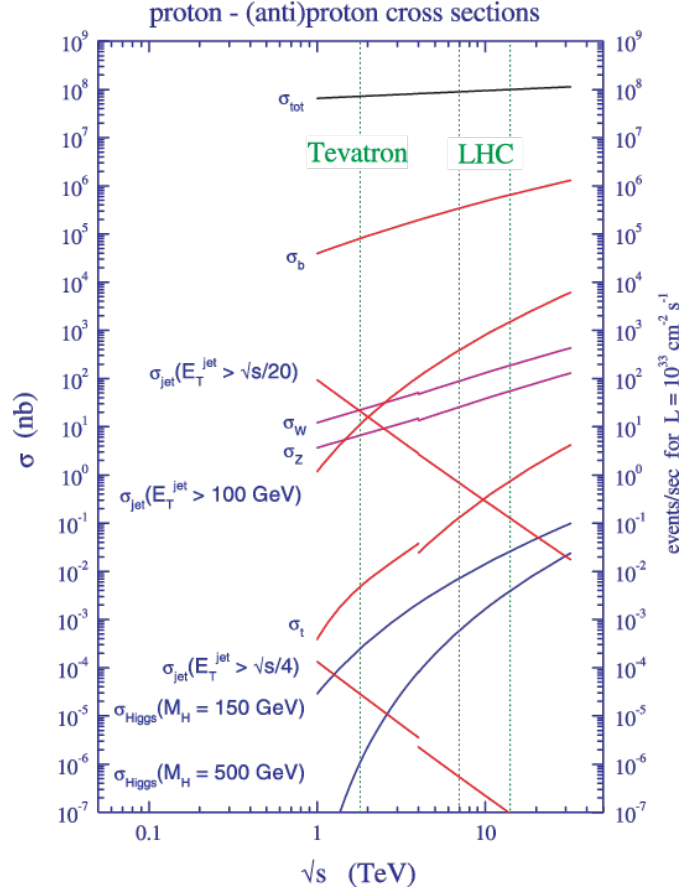


Figure 3-11 Cross Section values of several processes at different center-of-mass energy ($\sqrt{s}$) of p-p collisions.

The designed luminosity for CMS and ATLAS experiment is $\mathcal{L} = 10^{34} \text{cm}^{-2} \text{s}^{-1}$ and the center of mass energy of LHC is $\sqrt{s} = 14 \text{ TeV}$.

3.3.4 Luminosity and beam parameters

The beam parameters corresponds to the luminosity can be written as follows:

$$\mathcal{L} = \frac{\gamma_r N_b^2 f_{rev} k_b}{4\pi \epsilon_n \beta^*} F \quad (3-3)$$

where γ_r is relativistic factor, k_b corresponds to bunches number in a beam, N_b is the intensity of bunch, f_{rev} shows the revolution frequency, β^* is beta function at collision point and F denotes the luminosity depletion factor because of the crossing angle at a point where the beam interacts. The factor F is:

$$F = 1 + \left(\frac{\theta_c \theta_X}{2\sigma^*} \right)^2 \quad (3-4)$$

In the above equation, θ_c is crossing angle at interaction point, θ_z is root mean square bunch length and σ^* is transverse root mean square beam size at the point of interaction.

In luminosity relation 3-3, a few variables are utilised to enhance LHC luminosity but some are permanent, such as f_{rev} i.e. LHC tunnel perimeter. For circulating beams, this formula is appropriate having identical parameters. The nominal luminosity can be

Table 3-3 Parameters values designed for LHC operations.

Parameter	Value
Circumference	26.7 km
Energy ($\sqrt{s}$)	14 TeV
Collision momentum	7 TeV/s
Nominal luminosity	$10^{34} cm^{-2} s^{-1}$
Bunch spacing	25 ns
Distance between bunches	7 m
Bunches in ring	2808
Protons in a bunch	1.15×10^{11}
Beta function at point of interaction (β^*)	0.55 m
Transverse RMS beam size at point of impact (σ^*)	$16.3 \mu m$
Max. magnetic field of dipole	8.33 T
Lifetime of beam	10 h
main dipole current	11800 A
Magnets stored energy	11 GJ
Number of dipoles	1232
Number of quadrupoles	858
Dipole magnet operational temperature	1.9 K

achieved by using the parameters summarized in table 3-3.

LHC has successfully delivered $150 fb^{-1}$ of data till the end of the Run 2. In order to fully probe the SM and the Higgs sector particularly, higher luminosity is needed by physicist which is also helpful to explore the physics beyond SM. For this purpose, $5 \times 10^{34} cm^{-2} s^{-1} - 10 \times 10^{34} cm^{-2} s^{-1}$ luminosity is planned for High luminosity LHC.

3.3.5 Proton-proton collisions at LHC

The robust and huge colliding experiment ever built is LHC with the motivation to test the SM of particle physics. The collision of the protons at the center-of-mass energy

in LHC, actually involves the interaction of their constituents. These constituents are called partons i.e. quarks and gluons, and each parton only carries a fraction of the proton's momentum. Scattering experiments are normally quite helpful to probe e.g. production cross section of a certain process and to compare the results with the QCD predictions.

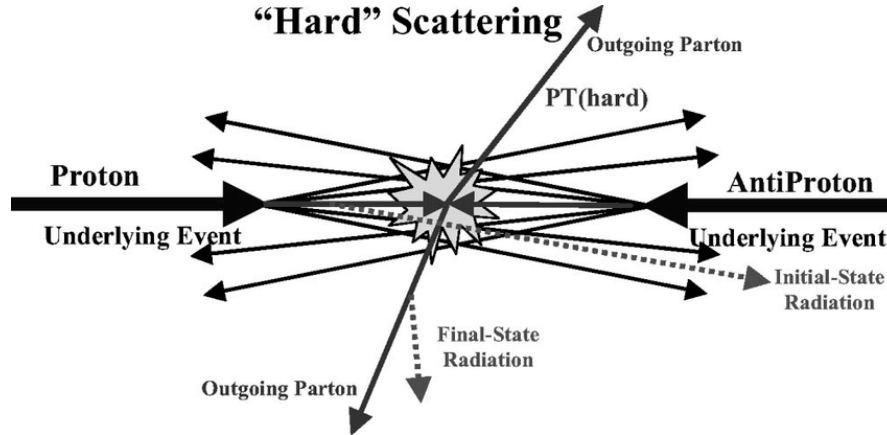


Figure 3-12 Diagram of proton-proton collision at high energies.

In collider experiment, large momentum transfer from the interacting partons can occur. This is hard scattering process, which originate various particles e.g. Higgs. Occurrence of a certain process probability is called its cross section. Hard scattering process in proton proton collisions is illustrated in the Figure 3-12. In Figure 2-4 [30], there are production cross sections shown for various processes in p-p collisions at LHC.

In a hard interaction, non-collision partons typically travel straight and can participate in a variety of low momentum transfer processes. These are known as underlying events. If in a certain proton-proton collision, more than one interactions happen then it is called as multi-parton interactions (MPI). There is another process which usually happen in accelerator experiment referred as “Bremsstrahlung” in which charged partons can radiate photons or gluons. If this radiation (Bremsstrahlung) is connected to the incoming partons then it is referred to as initial state radiation (ISR) and if originating from the outgoing parton, it is referred as final state radiation (FSR).

The hadronization (form color singlet) process occurs when quarks and/or gluons bind together in a cone like structure consisting of hadron bunches are referred as ‘jets’. For the hard scattering processes, cross section describes the probability that a particular

process will take place. Factorization theorem is used to estimate this probability. In this theorem, a factorization scale (μ_F) is used to disintegrate the cross section into soft and hard processes. Lets consider the collision of two protons (A and B). The final state S is produced from the hard scattering of the partons a and b . The relation used to write the total cross section of the process would be:

$$\sigma_{AB} = \int dx_a dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) \hat{\sigma}_{ab \rightarrow S} \quad (3-5)$$

where x_a is parton a momentum fraction. Similary, x_b is parton b momentum fraction. Factorization scale (μ_F) is entire momentum transfer. The parton distribution functions (PDFs) are $f_{a/A}(x_a, \mu_F)$ and $f_{b/B}(x_b, \mu_F)$.

3.3.6 Full Run 2 (2016-2018) data taking and High Luminosity LHC

In 2013 and 2014, the LHC was shutdown because of the upgrades for its machine and detectors, after successful data taking during Run 1. Run 1 has 7 TeV in 2010 which continued till 2011 with the same centre of mass energy with total 6.1 fb^{-1} of data. In 2012, 23.3 fb^{-1} data was collected after increasing the energy to 8 TeV. Higgs boson [4, 5] was discovered in the analyses based on Run1 data.

LHC was operational (for Run 2) after the shutdown in 2015 and collected data at a constant centre of mass energy 13 TeV till 2018. For majority of its operations in 2018, LHC has the luminosity of $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. In addition to the collisions of interest, certain unwanted additional p-p collision occur that are known as pileup, which increases with the instantaneous luminosity. In 2016, 23 pileup events were present per bunch crossing, and in both 2017 and 2018, the mean number of pileup events was 32. the luminosity delivered by LHC in 2016 is 41.0 fb^{-1} , 49.8 fb^{-1} in 2017 and 67.9 fb^{-1} in 2018 as shown in the Figure 3-13. The recorded data by the CMS detector and delivered by LHC can be seen in the figures for 2015, 2016, 2017 and 2018 run periods.

LHC was shutdown again for its long shutdown 2 (LS2) period by the end of 2018, after delivering data for the Run2 period successfully. Each of LHC experiments has been upgraded during LS2. CMS has altered the photodetectors of endcaps and barrel to silicon photo-multipliers. Stainless steel made beam pipe has been replaced with the material composed of Aluminium. Because of radiation damage in endcaps of HCAL, scintillating tiles were changed. A finer granularity tracker has been added in the upgrade having increased coverage in the forward region. As this tracker is much lighter,

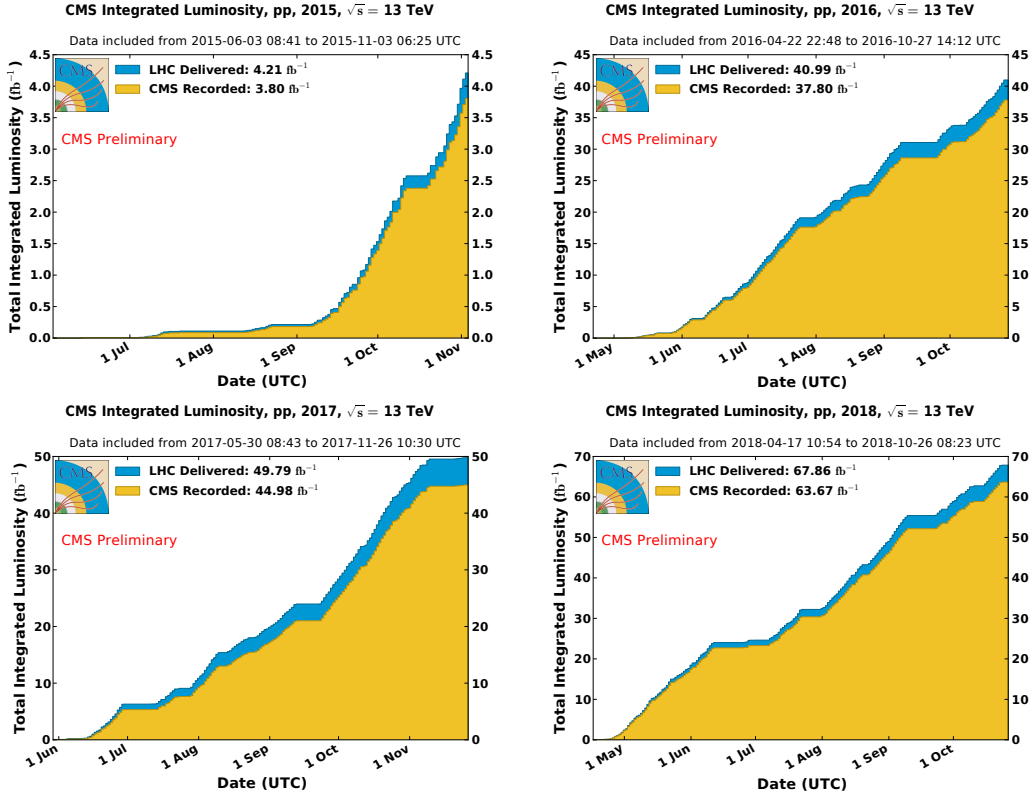


Figure 3-13 LHC delivered luminosities shown in blue and the ones recorded by CMS are in yellow for 2015 (top left), 2016 (top right), 2017 (bottom left) and 2018 (bottom right).

eventually material budget reduced and along with this, the latency is enhanced for the L1 trigger at Phase 2 to $12.5 \mu\text{s}$. Additionally, new front end (FE) electronics is being employed for HL-LHC preparations in CSC muon chambers.

Muon small wheels has been built by ATLAS for the integration with trigger system. For the efficient identification and reconstruction of physics objects, the Inner tracker system (ITS) has also been upgraded by ATLAS. ALICE has also made various upgrades for its Central Trigg, Time Projection Chamber, Inner Tracking System (ITS), Data Acquisition system (DAQ), Muon Forward Tracker (MFT), High Level Trigger (HLT) etc. In the same manner as other LHC experiments, LHCb had also enhanced its readout from 1 MHz to 40 MHz with full software based trigger by changing the FE electronics and efficient computing infrastructure.

During its LS2 period, various upgrades were made to the detectors that is committed for High Luminosity LHC afterwards (HL-LHC). The mean pileup per bunch crossing at HL-LHC is to be predictably 140. In the HL-LHC design, the mean pileup values will go up to 200. This would result a major change, and the conditions would be



Figure 3-14 Planned operation schedule of LHC and high-luminosity LHC upgrade.

remarkably harsher compare to what we have in the current LHC conditions. Furthermore, major upgrades for HL-LHC would take place during LS3 when the Run 3 period of LHC would be done [80]. LINAC2 will be replaced by LINAC4. It took about 10 years to built because it (LINAC4) is around 90 meters in length. It's capability is to accelerate negative hydrogen ions (H^-) to the energy up to 160 MeV. Operation schedule for LHC and plan for the HL-LHC is shown in Figure 3-14.

3.4 The Compact Muon Solenoid (CMS) experiment

3.4.1 CMS detector overview

CMS is a general purpose detector of LHC designed to observe the wide range of physics objects. This is an experimental facility based on the latest technology in order to probe new physics as well as validate the results from the LHC experiments like ATLAS. CMS and ATLAS have same physics goals but different geometries. The apparatus is able to deal with high pileup at high rate, around 1000 particles are produced every 25 ns. The detector and its electronics should have to withstand high radiation doses and fluence. It has remarkable resolution and track reconstruction efficiency of charged particle parameters. The CMS detector and its subsystems are shown in Figure 3-15.

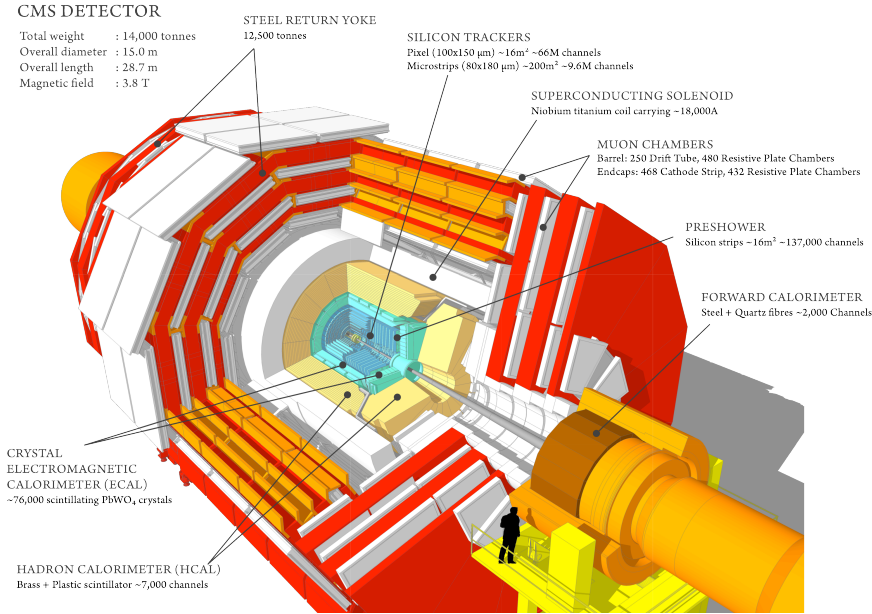


Figure 3-15 Design of the CMS detector.

The CMS, in particular has excellent resolution for identification of muons, their momentum across a wide range of dimensions and ability to correctly determine their charges. This is necessary for the measurement of Higgs boson, in the ZZ decay channel particularly. CMS also exhibit good momentum resolution, reconstruction and trigger system on τ leptons and b quarks. Excellent reconstruction of the jets is required in the majority of physics measurements. CMS has very good energy resolution as well as isolation for electrons and photons while keeping high acceptance. These characteristics are really essential for Higgs boson measurements in the diphoton decay channel. For dijet objects, CMS has an effective missing energy calculation and mass resolution. Various analyses in search of supersymmetry consider missing energy as the key element. The CMS is capable of probing wide variety of other physics domains, such as search for dark matter and extra dimensions in the energy range from a few GeV to TeV scale [22].

3.4.2 Coordinate System of CMS

The coordinate system of the CMS detector is cylindrical with its centre as the nominal proton-proton interaction point. The z – axis is the beam axis as shown in the Figure 3-16. The x -axis is horizontal in the direction towards the center of LHC ring and y axis is directed vertically upwards. The radial component in transverse plane is

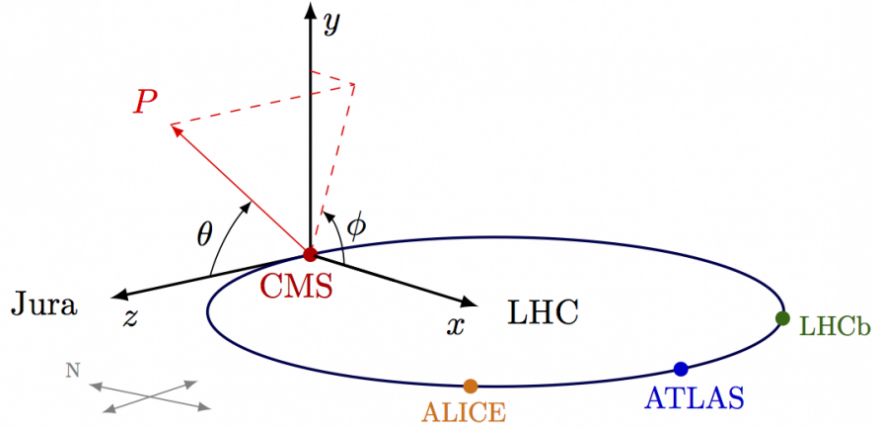


Figure 3-16 CMS coordinte system.

represented by r . θ being polar angle, is measured from z -axis and the azimuthal angle ϕ is measured from the x -axis (perpendicular to the z -axis). Particle's position in the CMS detector is stated as a function of (η, ϕ) . The $\eta = -\ln \tan(\theta/2)$ is the pseudorapidity which is invariant under the longitudinal boosts. Based on the pseudorapidity, barrel regions from the endcap regions of sub detectors can be differentiated. The barrel region in terms of pseudorapidity is $|\eta| < 1.5$ whereas $1.5 < |\eta| < 3.0$ corresponds to the endcap region. The region of the detector referred as the forward region is the one where we have high pseudorapidity values the direction close to the beam pipe.

3.4.3 The CMS Solenoid

The ultimate goals of the CMS detector is to precisely measure the momentum of the charged particle and because of this, magnetic field configuration plays key role in CMS. Superconducting solenoid with 220 tonnes cold mass is maintained at 4.5K (-268.5°C), is the central feature of CMS detector. The direction of the magnetic field is opposite in the tracker (towards inside) and in the muon system (towards outside) which permits to have an additional muon momentum measurement [23, 24]. CMS superconducting solenoid magnet is creating 3.8T magnetic field that is 100,000 time robust than earth's magnetic field. The relation used to describe the motion of a charged particle is stated as:

$$\frac{dp}{p} \propto \frac{p}{Br^2} \quad (3-1)$$

The path length r is the charged particle's length travelled within the magnetic field. From the relation, it is obvious that to attain efficient momentum resolution, a

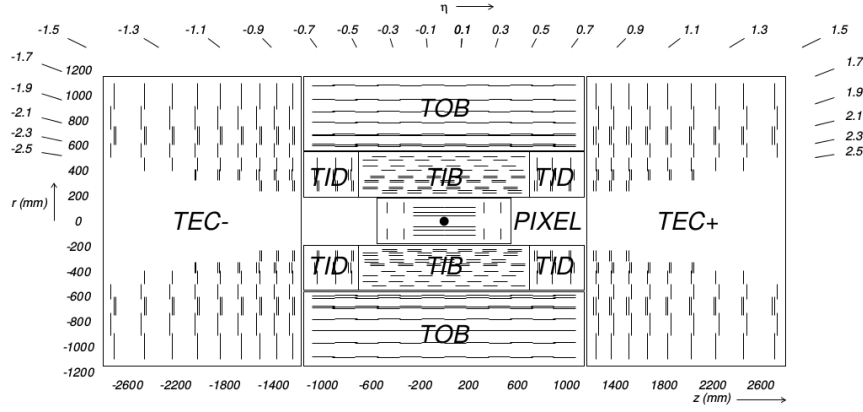


Figure 3-17 CMS tracker system layout and lines show the detector modules.

solenoid magnet was required with stronger magnetic field and then the CMS was made capable of that. The CMS magnet precisely measure the tracks of charged particles and has excellent momentum resolution because of its extremely high bending power.

3.4.4 Tracker System

The inner most sub-detecting system of CMS detector is the silicon tracker which is known as inner tracking system (ITS). Its diameter is about 2.5 m and length about 5.8 m with $|\eta|$ coverage up to 2.5. As being very close to the beam pipe, the tracker must withstand a high radiation dose throughout its expected 10-year lifetime without reducing the performance. Its purpose to reconstruct the charged particles curved tracks when they are deflected by the magnetic field. As being the inner most sub-detector, the tracking system is under the influence of a strong magnetic field of CMS solenoid. In order to reject information of extra interactions and objects, the CMS tracking system also provides the position information of vertices (Primary and secondary).

An average of more than 20 proton-proton interactions when the LHC is functioning on nominal luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, are called pileup which were above 30 for 2018 data. These events are 1000 on average after each bunch crossing i.e. 25 ns. The inner tracking system of the CMS is made up of silicon and is super efficient for being in the region of extremely high flux of particles directly emerging from the interaction point. In the Figure 3-17 [81], a schematic diagram of CMS inner tracking system can be seen. The ITS is kept at -20° C in order to reduce the radiation damaging. The full inner tracking system of the CMS is made up of the combination of silicon strip tracker

and the pixel detector.

3.4.5 Pixel Detector

The pixel detector is used to provide precise 2D space measurements in the ϕ and z directions. In the pixel detector, each silicon module has a pixel cell of size $100 \times 150 \mu\text{m}^2$. It provides help to reconstruct the vertex and precisely give three spatial points throughout its geometry [22]. In each of the endcap, two disc layers are facilitated in order to ensure the presence of three tracking points at least covering almost full $|\eta|$ range. It provides position resolution of 20 and 10 mm in the longitudinal and transverse direction respectively.

As a result of some inefficiencies that were reported during Run 1 regarding the first barrel layer of pixel sub-detector because of the limited readout bandwidth. For this reason, in March 2017, a new pixel detector was installed [82] with another layer of barrel and a forward disk additionally. The internal layer, in updated pixel detector is much near the point of interaction and outermost layer is a bit away from it as can be seen in the Figure 3-18.

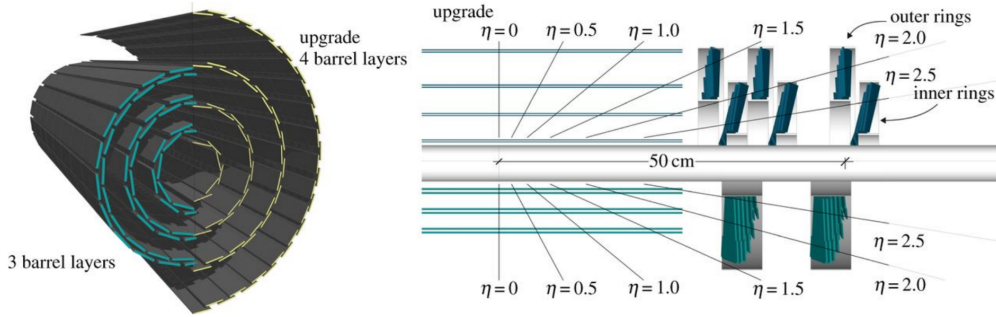


Figure 3-18 Diagram of pixel detector. Lower half represent the old pixel detector and the upper part is the new pixel detector.

New Readout Chips were installed in order to mimic the loss of data at high rate of collision relative to Phase 1 operations. The power converters were placed also in a way to utilize current cables and fibers. Enhanced and optimized pixel detector was capable to provide enhanced rate and channel numbers, resulting improved tracking efficiency and fake rate reduction. 66 million pixels are present in a detector of 1 m^2 size.

3.4.6 Silicon Strip Detector

The layer installed nearest to the pixel detector is Silicon Strip Tracker, at a radial distance from 20 to 116cm. Both sides of the pixel detector is surrounded by its tracker inner barrel (TIB) with 4 layers and tracker inner disk (TID) with 3 layers. It is extended up to 20-50cm having silicon micro strips of thickness $320\mu\text{m}$. Tracker outer barrel (TOB) have 6 layers and 220 cm length which provides 118 cm coverage. In barrel, the pitch of the strips changed from $80\mu\text{m}$ to $120\mu\text{m}$, and in endcap region, it changes from $100\mu\text{m}$ to $141\mu\text{m}$.

The outer most part of CMS inner tracking system, is tracker outer barrel (TOB) which is installed radially from 55-116 cm. It consists of $500\mu\text{m}$ thick 6 layers in barrel with $183\mu\text{m}$ and $122\mu\text{m}$ strip pitch. At $z=118$ cm, tracker endcaps are present. It normally covers $22.5\text{ cm} < |r| < 113.5\text{ cm}$ and $124\text{ cm} < |z| < 282\text{ cm}$ and also able to provide pseudorapidity coverage of $|\eta| < 2.5$. It can achieve the momentum resolution more than 2% for high p_T (100 GeV) charged particles. The reconstruction efficiency of muon tracks is expected to be close to 100%. For the hadrons and electrons, it is expected to have slightly worse efficiencies i.e 95% in the central region to 80% in the forward region. The impact parameter resolution of reconstructed tracks would be an important factor for efficient identification of final states including b quarks and tau leptons. In the Higgs to diphoton studies, the tracker plays crucial role in identifying the interaction vertex.

3.4.7 Electromagnetic Calorimeter (ECAL)

The CMS ECAL is a homogeneous calorimeter which is cylindrical in shape and built with PbWO_4 (lead tungstate) crystals. In the barrel (EB), 61,200 PbWO_4 crystals are present and give pseudorapidity $|\eta| < 1.479$ coverage. Endcap (EE) area is made up of 7,324 crystals providing coverage of pseudorapidity up to $|\eta| < 3$. The choice of (PbWO_4) was made for being a dense, active material and hard for radiation damaging. The fast response time was needed to collect the energy deposit before the crossing of next bunch. The pseudorapidity region ($1.4442 < |\eta| < 1.56$) is not considered in the analysis as being the cracked region between endcaps and barrel as a clearly visible in Figure 3-19 taken from [22].

When the collision happen, the high energy particle are produced. The electron

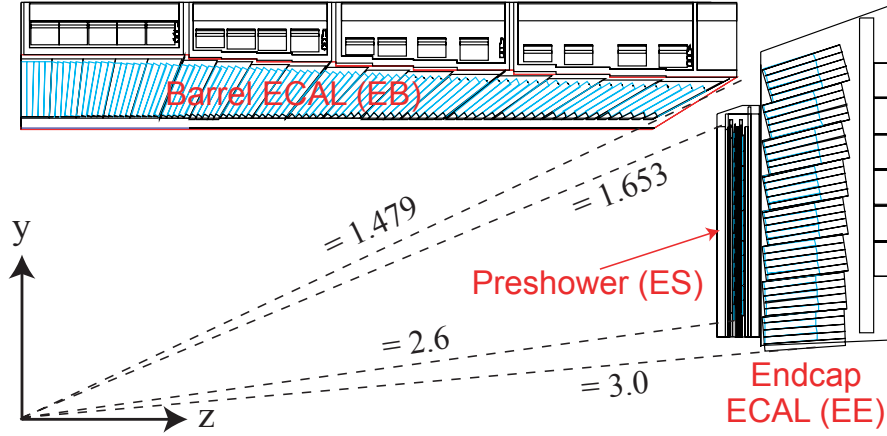


Figure 3-19 Pseudo-rapidity coverage of CMS ECAL.

or photon enters ECAL and cause cascade of particles through bremsstrahlung and pair production. This is known as EM showering, which continues until ionization influence electrons and photon energy falls below the minimal threshold needed for pair production. In the barrel region, Avalanche photodiodes (APDs) are installed and the endcap region is benefited with the Vacuum phototriodes (VPTs) in order to work efficiently under high magnetic field [83]. The ECAL response arises with the light emission which is directly proportional to the energy of incident particle. The overall structure of ECAL is illustrated in Figure 3-20. In order to distinguish between a nearly diphoton candidate such as from decay of neutral pion ($\pi^0 \rightarrow \gamma\gamma$) and the single photon, a preshower detector is installed in the endcap regions. This preshower detector ensures the pseudorapidity coverage of $1.653 < |\eta| < 2.6$ which is quite helpful to effectively identify electron and photon positions.

The ECAL is made compact by using lead tungstate crystals with high granularity. Moreover, lead tungstate aid to greatly reduce the pileup while measuring the energy because it exhibit small Molière radius. Additionally, (PbWO_4) crystals has scintillation time of about 25 ns which is coherent with the bunch crossing time of LHC. The temperature of ECAL operation is about 18°C and the scintillation (light emission) process depend on the temperature. So, for this purpose (to maintain the required temperature), a water circulation (via Aluminium pipes) based cooling system is positioned there and is made capable to effectively sustain the detector's temperature of up to $18 \pm 0.05^\circ\text{C}$. The ECAL's energy resolution i.e. the ability to accurately measure of incident radiation

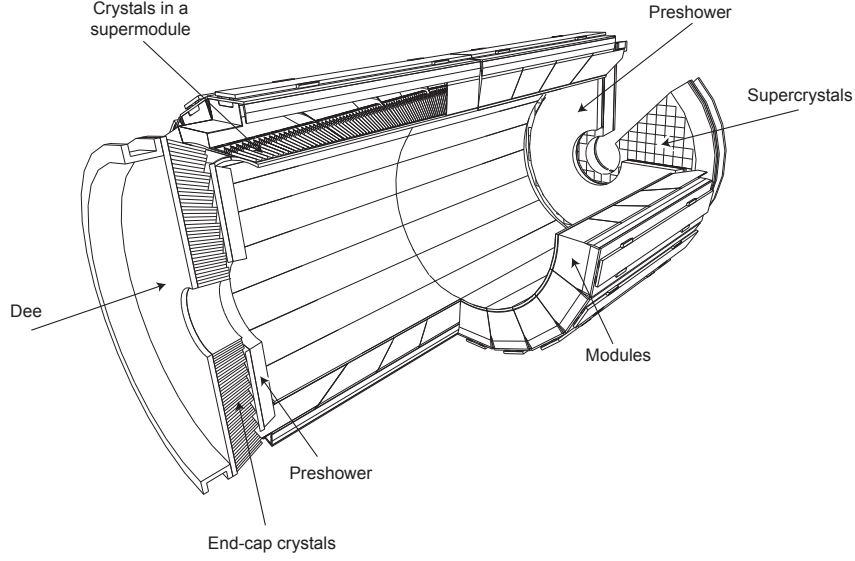


Figure 3-20 Overall layout of CMS ECAL.

energy is formulated as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_n}{E}\right)^2 + c^2 \quad (3-2)$$

The S being stochastic term, is the preshower energy deposits and the photo(electron)-statistics contributions. The noise term of the electronics is σ_n and c is a constant in order to covers the constant factors i.e. non-uniformity in detector for the collection of light etc. Furthermore, all of the energy is usually not contained within a crystal and due to this, some calibrations are needed to estimate the shower energy accurately. It has been verified from the offline data analysis, that the trigger to be $> 99\%$ efficient for $E_T > 40\text{GeV}$, ensuring full efficiency for the $H \rightarrow \gamma\gamma$ search. The remarkable performance of ECAL was important for the Higgs boson discovery in CMS, particularly in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(*)} \rightarrow 4l$ channels. The ECAL data have enabled the precise reconstruction of electrons and photons up to the TeV scale which is mandatory for several searches and phenomena beyond the standard model.

3.4.8 Hadronic Calorimeter (HCAL)

The part of CMS that surrounds the ECAL is known as HCAL which is employed to find and measure the energy of particles which interacts strongly like quarks, gluons hadronic jets etc. HCAL is crucial for measuring missing transverse energy (MET),

which is primarily connected to invisible particles like neutrinos that are important for the studies related to top quark. Hadron calorimeter lies between the solenoid magnet of CMS and ECAL at a distance from 1.77 m to 2.95 m as it is visible in the Figure 3-15. A general layout of HCAL is shown in the Figure 3-21 [22].

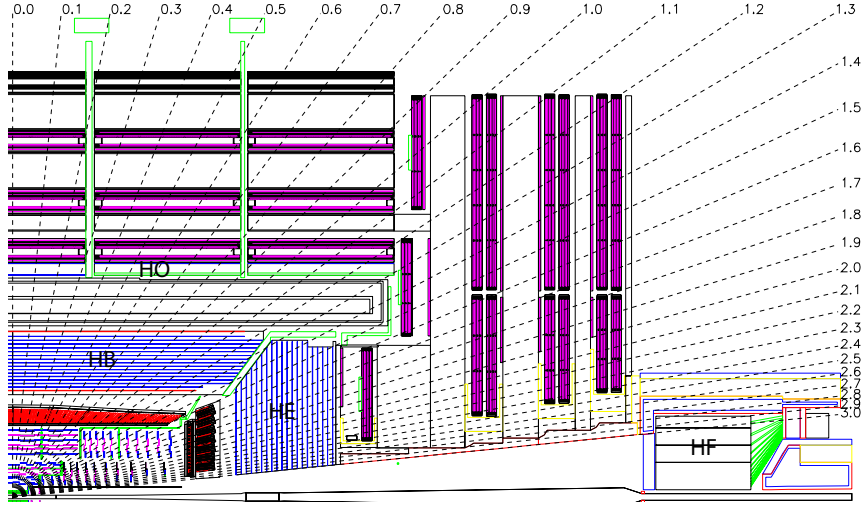


Figure 3-21 Layout of CMS HCAL.

Brass material is used in making of HCAL because of its short interaction length which acts as an absorber material. It is a cost effective and non-magnetic material. It give the pseudorapidity coverage upto $|\eta| < 3.0$. The distance of HCAL from the interaction point is 11.2 m. A froward Hadron Calorimeter (HF) is made on each side of detector to ensure the coverage upto $|\eta| < 5.2$. This HF provide help to detect particles produced at small angles relative to the beam axis in the p-p collisions. Good performance of the HF should be sustained as it plays crucial role in forward jet reconstruction which are of great importance in VBF $H \rightarrow \gamma\gamma$ events. The HCAL is divided into two regions i.e. barrel and endcaps just like the ECAL. HB is made up of 32 towers that make sure the coverage of pseudorapidity of $|\eta| < 1.4$ for HCAL barrel and HCAL endcap of $1.3 < |\eta| < 3$ for HE. As a result, 2304 towers in a segmentation of $\Delta\eta \times \Delta\phi = 0.87 \times 0.87$. HB has splitted into two halves i.e. HB^+ and HB^- consisting of 36 identical wedges and each one has weight of 26 tonnes. In order to avoid the shower tail of hadrons do not enters into ECAL and/or magnet from the rear part of calorimeters, an additional calorimeter i.e. Hadron Outer (HO) is placed next to the solenoid to act as an absorber. It (HO) has 10 mm thick scintillator which ultimately improves the

resolution of hadron calorimeter for MET. As the HCAL have to suffer huge flux of the particles, sampling material has been made up of steel/quartz fibre. Quartz fibres are then responsible for Cerenkov radiations which are then channeled to Photo-multiplier Tubes (PMTs). The alternate layers design of HCAL is consist of with of brass material and plastic which act as the absorber and scintillator respectively. It provide helps to accurately measure position, arrival time and energy of the produced particle. When a particle traverse through the scintillating material, the emitted light is collected by the optical and convey it to readout box. The HCAL energy resolution can be stated as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{90\%}{\sqrt{E}}\right)^2 + (a)^2 \quad (3-3)$$

and for HF

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{170\%}{\sqrt{E}}\right)^2 + (b)^2 \quad (3-4)$$

where a is 0.045 and b is 0.09.

In 2017, few upgrades were made in HCAL such as photomultiplier tubes were replaced with multi-anode PMTs and electronics was updated for better differentiation of muon hits [84]. The HE part of the HCAL was also upgraded by replacing the photo detectors with silicon photo multipliers (SiPMs) in order to eliminate the noise and enhanced detection efficiency.

3.4.9 Muon System

One of the most important part of the CMS detector is its muon system. In order to precisely identify the muons which are originating from the point of collision point is a big challenge. Muons exhibit a longer lifetime upto 2.19×10^{-6} and in relativistic regime its much higher. They can traverse the full detector without being decaying. Muon are approximately 200 times the mass of electron and because of this, they lose very small amount of energy through bremsstrahlung. However, by depositing low amount of energy, they leave track in the calorimeters. So, because of these features of muons, a dedicated muon system was installed in the outermost part of the CMS [85]. Figure 3-22 shows a general layout of CMS muon system during the Run 2 of LHC. The dedicated muon system of the CMS detector is made up of three different strategies to perform measurements (identification, momentum etc) on muons.

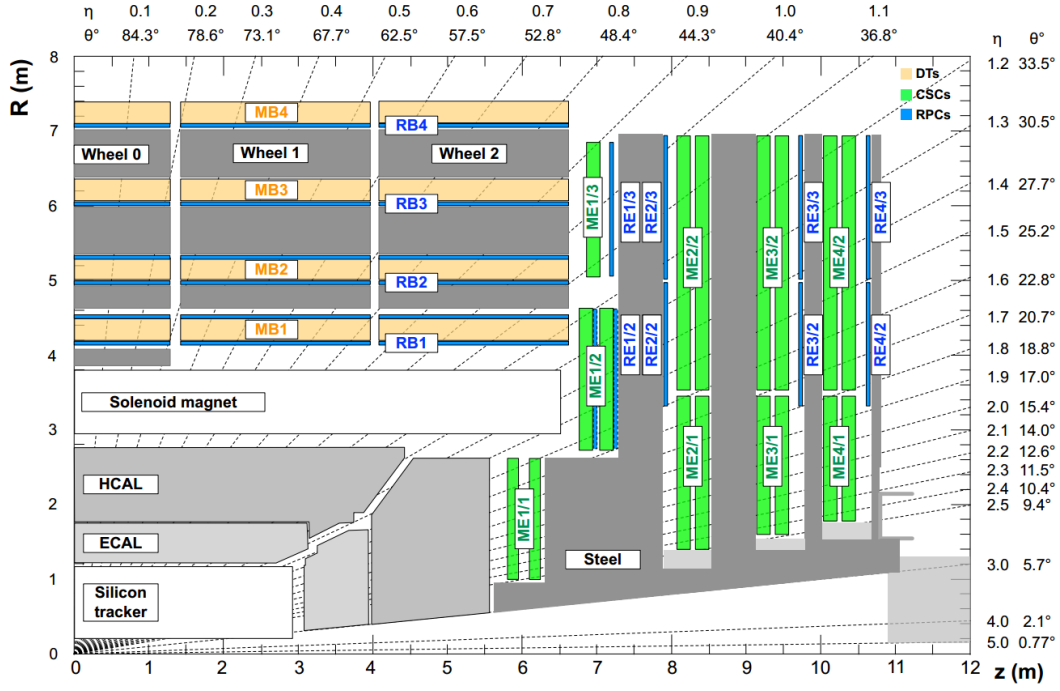


Figure 3-22 Muon system layout of the CMS during Run2 operations.

The detection of muons is crucial for various physics analysis specially for the Higgs (SM) when it decays to ZZ . This channel is believed as a golden channel for properties measurement of Higgs boson. Precise reconstruction and identification of muons is also required by some theories beyond SM.

In order to detect muons in the CMS, muon system is mainly based on the gas detection technology which is cylindrical in shape and provide pseudorapidity upto $|\eta| < 2.4$. Resistive Plate Chamber (RPC's) are installed in the barrel and endcap regions. Drift Tubes (DT's) covers the barrel region and in the endcaps region, Cathode Strip Chamber (CSC's) are present. The main motive of the CMS muon system includes muon identification, momentum resolution, triggering and to bear the high radiation.

During the long shutdown 2 (LS2), various upgrades were made for physics programs. CSC readout of ME1/1, ME2/1, ME3/1, and ME4/1 were improved. 180 chambers were reconditioned and cathode front end boards (CFEBs) replaced with updated digital CFEBs. One of the major upgrade for the muon system was the installation of Gas Electron Multiplier (GEM) detector in forward regions. It was needed because of the upcoming high collision rate in Run 3 and after that, to enhance the muon system capabilities of triggering, tracking and to work under extreme conditions which could be achieved with GEM. Two muon stations would have the assist of GE1/1 and GE2/1

($1.5 < |\eta| < 2.4$) layers for efficient precise measurements. Additionally, ME0 station is planned, which would enable to extend the pseudorapidity coverage of muon system upto $|\eta| < 3$.

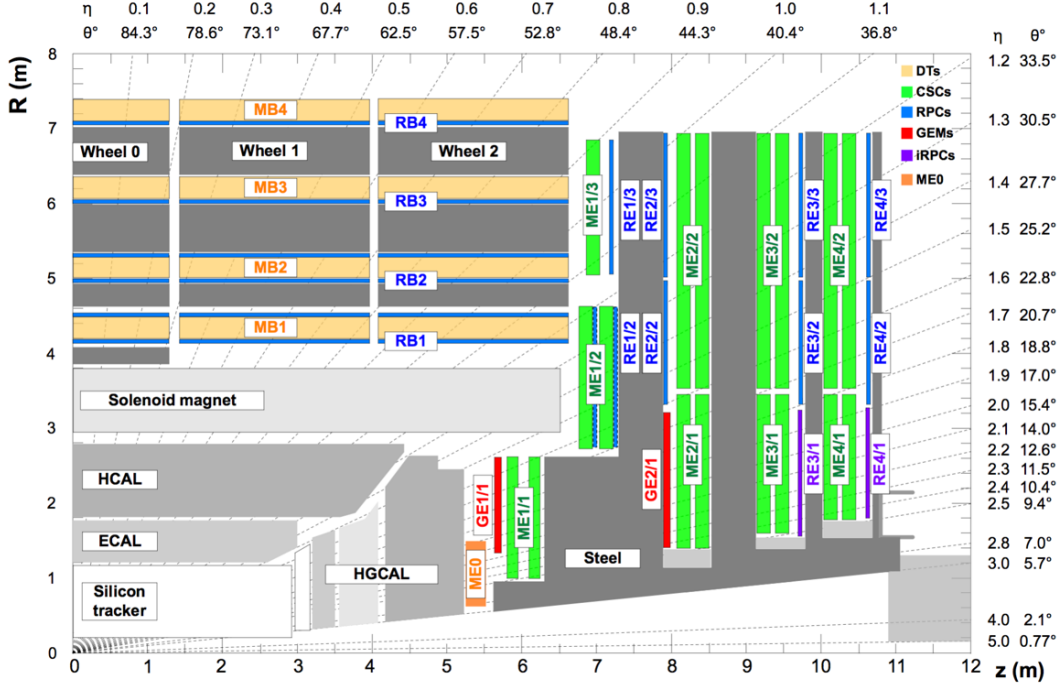


Figure 3-23 Schematic diagram (updated) of CMS muon system with GEM chambers.

GE1/1 is taking part in Run 3 operations whereas GE2/1 and ME0 will be placed after Run 3 when the LHC would be stopped for long shutdown 3 (LS3). A schematic figure of upgraded CMS muon system with GEM chambers is shown in Figure. 3-23 [86, 87] where the positioning of these detectors can be well understood.

3.4.10 Resistive Plate Chambers (RPC's)

An important part of the CMS muon system is the Resistive plate chamber which are the parallel plate gaseous detectors settled in both barrel and endcap regions. Along with the efficient time and spatial resolutions, RPC's does exhibit a very fast responsive which is used for the triggering process. A schematic diagram of a RPC chamber is shown in the Figure. 3-24 [88].

These gaseous detectors have positive and negative plastic charged plates, placed in parallel to each other and filled with gas (95% of $C_2H_2F_4$ and 5% of C_4H_{10}). The readout system of RPC's becomes operational in the avalanche mode, when a muon

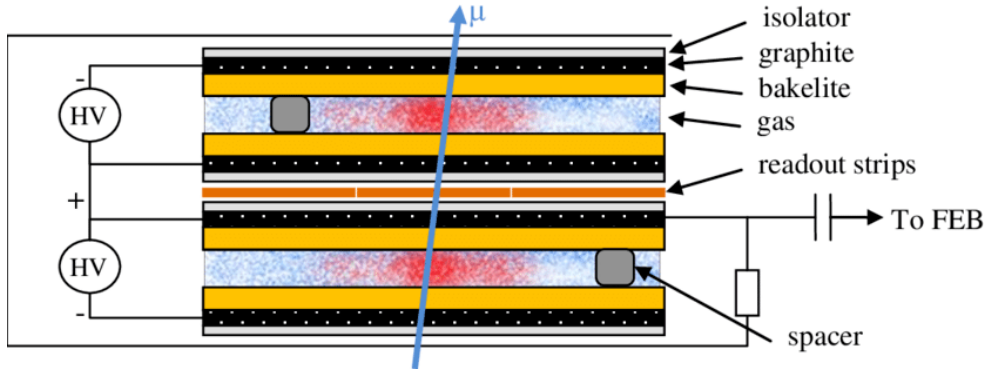


Figure 3-24 RPC chamber's cross sectional layout.

traverse by liberating electrons during the ionization (primary and secondary) of gas molecules. RPC's have outstanding time resolution.

3.4.11 Cathode Strip Chambers (CSC's)

Another important part of the CMS muon system are the cathode strip chambers (CSC's). As shown in the Figure 3-22, it is providing the pseudorapidity range up to $0.9 < |\eta| < 2.4$ which enables it to provide the tracking information and muon triggering in this region with the aid of RPC's. It operates efficiently at high event rate. A total of 540 CSC modules are placed in 4 CSC stations (ME1, ME2, ME3 and ME4). Each of this station is divided in rings (2 or 3) with from 18 to 36 CSC's installed on each of the rings. In each of the assembly, it has 6 independent layers with 6 gaps of gas (Ar 50%, 40% CO_2 , 10% CF_4). Each of the layer has cathode strips in the radial direction

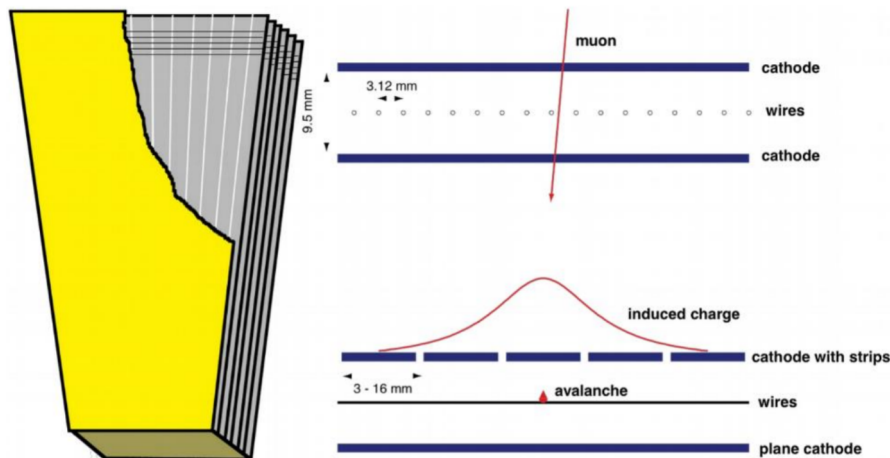


Figure 3-25 CSC functional layout.

and anode wires in transverse direction. It also operates on the ionization of the gas

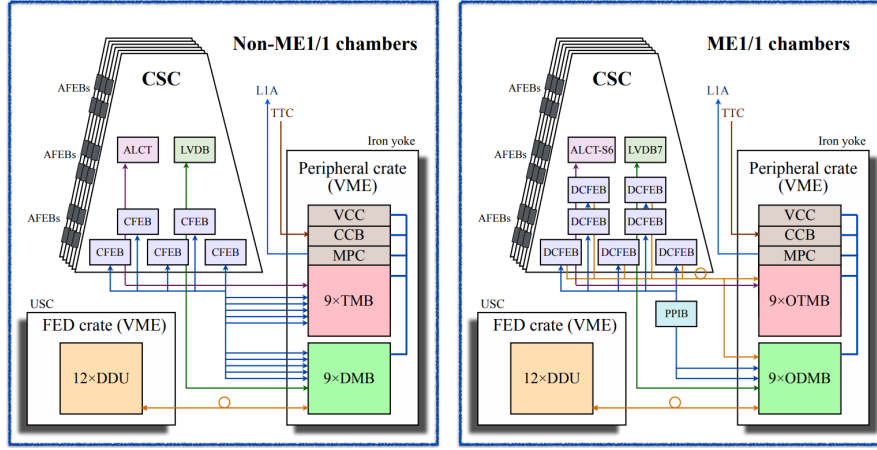


Figure 3-26 Electronic boards of a CSC chamber and the readout system.

molecules and produce the avalanches (when muon traverse) of electron which then move towards the anode wires. Gaseous ions move towards the cathode strips for the image charge detection. The readout system then collect these signals, amplify them and transmit to the electronic on chamber. Muon hit information from each of the chamber measured through its track in the 6 layers. In order to generate a trigger, 4 layers hits are essential out of 6. By the fitting of successive hits, precision in the measurement muon's momentum can be achieved. A general functional layout of CSC can be seen in the Figure 3-25 and overview of the chamber's electronics is shown in Figure 3-26.

3.4.12 Drift Tubes (DT's)

The drift tubes covers the central part (along with RPC's) of the CMS muon system and plays a vital role in tracking as shown in the Figure 3-22. It is providing pseudo-rapidity range of $|\eta| < 1.2$. It is installed in 4 stations i.e. MB1, MB2, MB3 and MB4, in outward direction at 4.0, 4.9, 5.9 and 7.0 m respectively from the interaction point. There are 250 DT's chamber (each composed of 12 layers of Aluminium) are assembled along z-axis on 5 wheels. At high transverse momentum, the DT's alone has over 95% identification efficiency. The planes (12) in a DT chamber are of average size 2m×2.5m. Apart from the MB4, they are built having 3 units known as super-layer (SL). A combination of two SLs determine the coordinates in (ϕ) plane and the third one aid to track z coordinates [89]. Layout of DT on one of the wheel chambers can be seen in Figure 3-27.

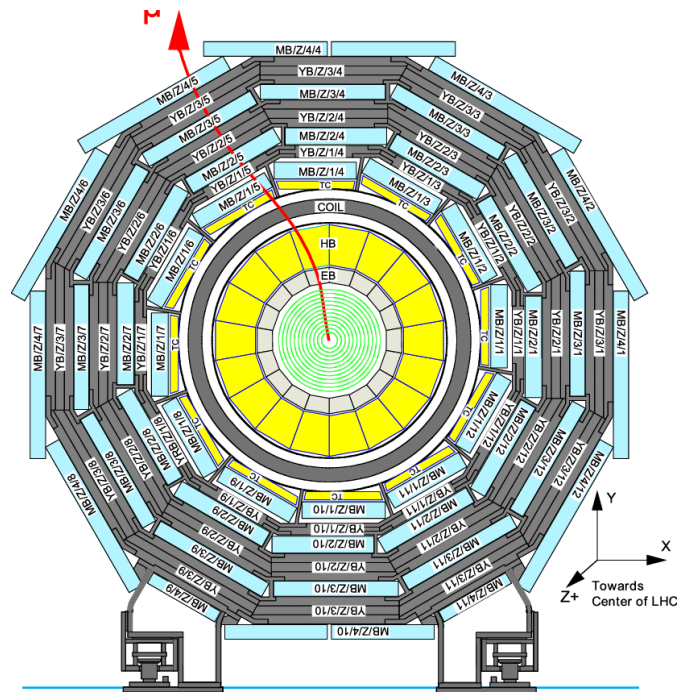


Figure 3-27 Cross sectional layout of DT chambers (one of the 5 wheels).

Each of the DT cell has a steel anode wire ($25\mu\text{m}$) and walls of cell acting as cathodes producing the electric field for drift. Every cell is filled with CO_2 (15%) Ar (85%) mixture which is cost effective. A general diagram of a drift cell is shown in Figure 3-28.

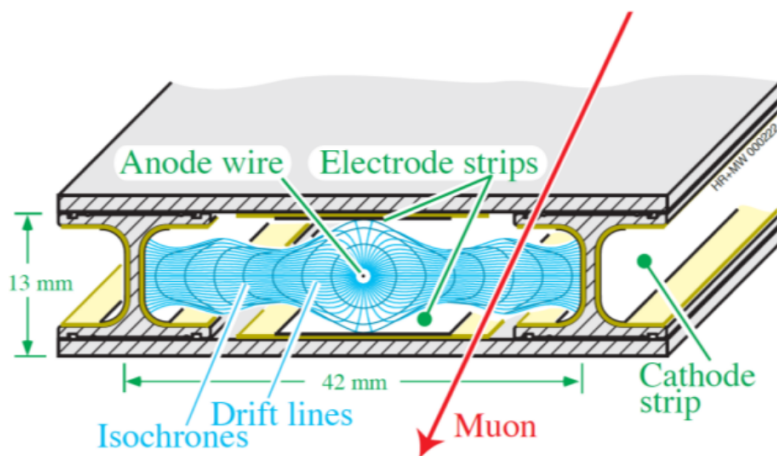


Figure 3-28 Diagram of drift cell.

3.4.13 CMS Trigger System

CMS is huge general purpose detector and it is producing extremely high rate of information at each bunch crossing. LHC is producing about 40 million collisions per second (40 MHz). All of the collisions are not of great interest for physics studies. Additionally, the data acquisition system (DAQ) is not much effective to look into every event produced at this rate. Storing of all the events which are not of interest also need big storage. Eventually, in order to reduce the rate of events to be saved, a trigger system was needed which sort events only on the basis of interest related to dedicated physics purpose. In the CMS, the trigger system is a divided into two sequential stages i.e. the Level 1 (L1) trigger system and a high level trigger (HLT).

3.4.14 CMS Level 1 (L1) Trigger

As the maximum output rate of the electronics is 100 kHz. The L1 trigger is a hardware level trigger which takes calorimeter's as well as CMS muon system information to make decisions [88]. It is capable to reduces event rate to 100 kHz from 40 MHz. The L1 trigger must make a decision for an event to be saved or not within $3.2\mu s$. The full understanding of the event is buffered before making the final decision of trigger. It is built with reprogrammable electronics (Field Programmable Gate Arrays (FPGAs)) which uses energy information (from HCAL and ECAL) as well as momentum information (CSC, DT and RPC) of electrons, photons, muons, jets and MET etc. It (L1 trigger) is further divided into calorimeter and muon triggers. A schematic of L1 trigger can be seen in the Figure. 3-29.

ECAL and HCAL (calorimeters) has independent triggers which transmit information to the global Calorimeter Trigger (GCT). Similarly, muon has local triggers which correspond to CSC, DT and RPC. These local triggers are later on combined to get a global muon trigger (GMT). Global Trigger (GT) made up of GCT and GMT in order to make final decision that whether to accept or reject an event. Once an event pass the L1 trigger selections, it is moved to the HLT (High Level Trigger).

3.4.15 High Level Trigger (HLT) system

HLT is a software level trigger and select events in same way that we usually do in offline events processing. The main motive of the HLT is to reduce the rate (100kHz coming from L1 trigger) to 1 kHz. This job is basically done in various stages. For an

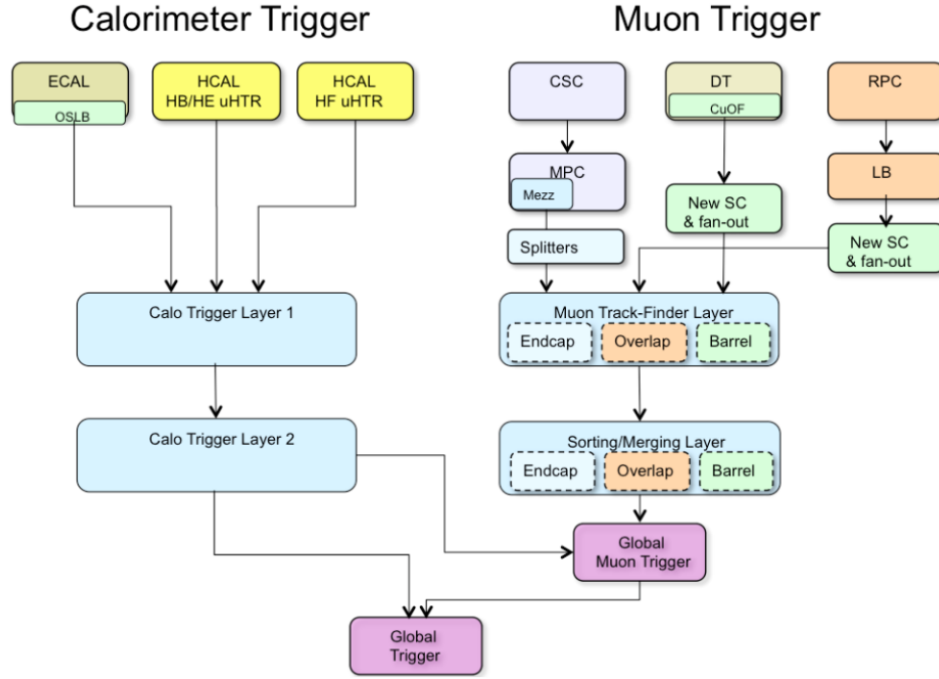


Figure 3-29 The L1 trigger configuration.

event of interest, in order to identify and reconstruct electron, muon or photon, a criteria is placed over each of the event. It is built with several single processor computers (event filter farm) which join pieces of an individual events together using builder units in order to make final events. This data is then transferred to filter unit and after that, the information (raw) is changed to data structure format. This information is detector specific and do reconstruction of the events [88].

Offline reconstruction algorithms are there in order to exploit the precise data coming from trigger. The CPU power enable to process events at rate of 100 kHz for an average processing per event time *viz* 90 ms. The performance of HLT is actually based on specific "HLT paths". These paths are made of various set of algorithms to be performed in a prescribed order to determine the physics objects and a reconstruction step is to be followed after that. Each of the HLT path execute in increasing order of complexity since simple algorithms need to run first to perform the basic selections. Events selected by the L1 trigger are transmitted to HLT which have several assigned paths for decision making. Finally, if the event is accepted by HLT, the full readout is saved to local disk and transfer to CMS Tier-0 for permanent storage. These events are then processed with the help of CMS offline software and the objects are produced for physics analysis.

3.4.16 LHC Computing Grid

It was the need of four experiments of LHC (CMS, ATLAS, ALICE, LHCb) to save the data collected from LHC interactions. The further process of this data and to analyse it in order to extract physics results is quite challenging. LHC interactions are actually producing approximately 15 petabyte data per year while its operating. In 2002, CERN council approved a multi-tiered system to deal with this extremely large amount of data. This infrastructure is well known Worldwide LHC Computing Grid (WLCG) and about 170 computing centers are linked in more than 40 countries. This computing grid providing organized facilities for computing to all of the four main experiments of LHC (CERN). About 150,000 processor cores are connected together with the ability to store 50 petabytes of data on the disks [\[90\]](#).

Chapter 4: Physics Objects

As explained in the HLT section that events after being organized by the HLT decisions are saved to the permanent storage system of CMS. The stored data is in RAW form and require to be reconstructed fully. A reconstruction procedure is applied on the stored data in order to associate them to the physical particles such as electrons, photons and muons etc. These particles then aid to probe various interesting physics studies such Higgs decay to two photon or Higgs decays to four lepton channel for the measurement of its properties such as cross sections, mass, parity and spin etc. For photons, an efficient reconstruction and identification is required and to keep the rate of misidentified photons rate as minimum as possible.

4.5 Particle flow

In CMS, the global event description is established by exploiting the particle flow (PF) algorithm [91]. The aim of the PF is to integrate the information of all CMS sub-detectors and to enable the best identification and measurements of energy for various objects. The inputs of PF algorithm are the tracks coming from the tracker, from the muon system and clusters of calorimeter (from ECAL and HCAL). The CMS is capable of benefiting from PF approach because of its powerful magnetic field as well as fine segmentation of the tracker, calorimeters and muon system.

The combinatorial track-finding procedure is used to reconstruct the tracks utilizing the hits information from the tracker by having several iterations [92]. The iteration procedure is, firstly track seeds having two or three hits are considered by defining the parameters of initial track and then an extrapolation is done from the paths of expected track by adding additional hits information. After this, a re-estimation of track parameters is done and the track candidate collection is cut-off depending upon quality criteria. All of the selected hits are taken off from the consideration for following iterations. In this method, the identification of the most obvious tracks i.e. with high p_T and close to point of interaction is done in the first iteration. As a results, the complexity of next iteration is decreased as several hits are removed. Therefore, it allows lower thresholds can be used and tracks having lower p_T to be found. Additionally, an independent procedure

has been employed for reconstruction of muon tracks utilizing the hits information from the muon system.

A clustering algorithm, in the calorimeters, is exploited to collect deposits of energy corresponding to each shower. This method in the ECAL is explained here as it is a crucial input for the $H \rightarrow \gamma\gamma$ studies. Clustering algorithm starts by choosing cluster seeds having energy above a certain level and larger than any neighbouring crystal [93]. The topological clusters composed repeatedly by including deposits which shares a side or corner with the one already within the cluster, provided that the energy is more than a threshold. If a crystal is included in more than one topological cluster then its energy is shared between them by considering a Gaussian shower shape. At the end, the topological clusters are combined into the so-called superclusters (SCs). This step is outlined to reacquire the energy lost due to bremsstrahlung. The radiated showers normally have similar η values but have spread out in ϕ direction due to the magnetic field.

Based on these tracks and calorimeter clusters as inputs, PF algorithm forms the group of candidates for five kinds of particle i.e. muons, electrons, photons, neutral hadrons and charged hadrons. As a result of above mentioned procedure, the central software of CMS offers analyses with the physics objects which are ready to be used in studies.

4.6 Photons

In order to use the photons in $H \rightarrow \gamma\gamma$ analysis, they are selected from Particle Flow (PF) set of photon candidates. From the super clusters (SC), the energy is estimated for each of the photon candidate. The SC contains the energy deposits from EM showers of various particles. Sometimes, photons can interact with the detector materials in front of ECAL producing electron-positron pairs which are called as converted photons. These converted photons make energy deposition in pre-shower detector of ECAL, which does include in the estimation of SC energy. A multivariate regression technique is then employed for the correction of this SC energy. Scale and smearing corrections are used after that for photon energies in order to make the data and MC agreement better. The explanation of the procedure to correctly estimate the energy is in Section 4.6.2. Furthermore, a pre-selection criteria is applied also on the set of photons to be used in physics analysis, once photon energy is fully established. An-

other requirement is also placed on the MVA trained for photon identification in order to mimic the contamination from other objects usually lessen the real photons.

4.6.1 Definition of variables

Reconstruction of the photons usually based upon various set of variables which are categorized as isolation and shower shape variables. The shower shape information from the ECAL is utilized for the energy correction as well as to discriminate the real photon from the fake ones. Isolation variables plays important role in reducing the misidentification of objects resemble to a real photon. The variables and their definitions are described below in this section.

Shower Shape variables:

- $Cov_{\eta\phi}$: The covariance of the single crystal η and ϕ in terms of crystal cells within the 5x5 crystals centred on the SC seed crystal.
- $\sigma_{\eta\eta}$: The lateral extension of the EM shower measured in terms of crystal cells.
- R_9 : $E_{3\times3}/E_{RAW}$, where $E_{3\times3}$ is the energy sum of the 3 by 3 crystals surrounding the SC seed crystal and E_{RAW} is the raw energy sum of the SC.
- σ_{η} : The logarithmic energy weighted standard deviation of single crystal η within the SC.
- σ_{ϕ} : The logarithmic energy weighted standard deviation of single crystal ϕ within the SC.
- $Preshower \sigma_{RR}$: The standard deviation of the shower spread in the x and y planes of the preshower (only defined in the endcap).

Isolation variables:

- *PF Photon Isolation (Iso_{ph})*: Transverse energy sum associated with all particles identified as photons by the particle flow algorithm, falling inside a cone size $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around photon candidate direction with ρ correction applied.
- *PF Charged Isolation (selected vertex) (Iso_{ch})*: Transverse energy sum associate with all particles identified as charged hadrons falling inside a cone size $R=0.3$ around the photon candidate direction with respect to the selected vertex;
- *PF Charged Isolation (worst vertex) ($Iso_{ch,worst}$)*: Transverse energy sum associated with all particles identified as charged hadrons falling inside a cone size $R=0.3$ around the photon candidate direction with respect to the worst vertex. The advantage

of this variable is due to the fact that generally a prompt photon is isolated with respect to any vertex and fake photon is not.

- *Charged Hadron Isolation ($Iso_{ch,had}$)*: The energy in HCAL from charged particles inside a cone ($R < 0.3$) around the photon candidate.

- *H/E*: The energy collection by the HCAL towers within a cone of $R = 0.15$ centered on the SC position, divided by the SC energy.

- *Tracker Isolation in a hollow cone (Iso_{track})*: Transverse momentum sum associated with all tracks falling in a cone size of $R = 0.3$ around the photon candidate direction. Tracks falling in an inner cone of size $R = 0.04$ are not included in the p_T sum.

Other variables:

- *Electron veto*: It is set to true if SC corresponds to PF reconstructed electron.
- *Pixel-seed electron veto*: It is set to true if its SC is found to match an electron track with least one hit in the pixel detector.

- ρ (*rho*): The energy median density per unit area in the event.
- *Supercluster η* : The η of the SC corresponding to the reconstructed photon.
- *Supercluster E_{RAW}* : The raw energy of the SC corresponding to the reconstructed photon.

4.6.2 Photon energy

The energy of photon SC is computed from the calibration of the ECAL SC but this estimation of energy is not perfect for various reasons. First of all is that, the SC might not be able to capture all the energy of EM shower and which eventually cause underestimation the photon energy. It happen when a photon interacts with the material and begin making EM shower before reaching ECAL. As a result, the energy is lost within the material causing a dispersed shower which is may not be contained by the supercluster. Additionally, leakage may occur between crystals and modules in the ECAL, and can also through the back of the ECAL because shower may star near the crystal's end. The energy deposition by the electrons in the EM calorimeter is the same as photons do, so they are considered as reference to photons. A general pattern of electron/photon in the ECAL when they came out of the ITS is shown in Figure 4-30. A multivariate regression can improve significantly the estimate of uncorrected supercluster (E_{SC}). An additional correction of the energy to each of the reconstructed photon is applied so that its energy may agree well in simulation with the real energy of photon. The aim to train

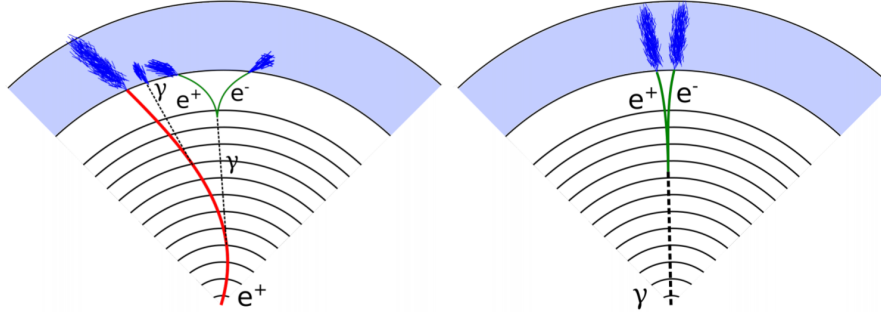


Figure 4-30 Positron and Photon pattern in the detector.

a regression, is to know the parameters of E_{true}/E_{SC} distribution function, which are parameterised as Gaussian having power law tails on either side.

In order to account for different effects, a large number of inputs variables are provided to the regressor. These may add the information of shower shapes, the SC position and the energy density for that specific event. Position and the shower shape variables allows us to correct the difference in contaminated showers due to changed in ECAL geometry and showers which caused by the material. The seed crystal position variables in the SC and energy ratios estimated using the seed are responsible for the shower containment. When the regression is made, some additional corrections are there to make the data and MC agreement reasonable.

Firstly, data energy scales are corrected. Detector response may be effected because of the radiation damaging while LHC operation and recovery during LHC shut-down time. It may alter the energy measured and transform discontinuously with the passage of time. With the help of $Z \rightarrow ee$ events, where electrons are reconstructed as photons, a set of corrections are derived by using the well known Z boson mass. These additional corrections are expressed differentially in time and binned in photon R_η and η with two bins in barrel and endcap. The magnitude of these derived corrections changes from 0.1% to 0.3% (barrel) and from 0.2% to 2.0% (Endcaps). In the simulation, the energy resolution is also corrected to have a better agreement to that in data. These are implemented actually as a Gaussian smearing with the same set of binning as that of the energy corrections.

After applying all of the corrections, the energy resolution of the photons from 1% to 2.5% (barrel) and in endcaps it range from 2.5% to 3.5%. Energy validations after applying all of the corrections shown in Figure 4-31 are done by using $Z \rightarrow ee$ event,

where electrons are reconstructed as photons [94]. For the validation of unbiased

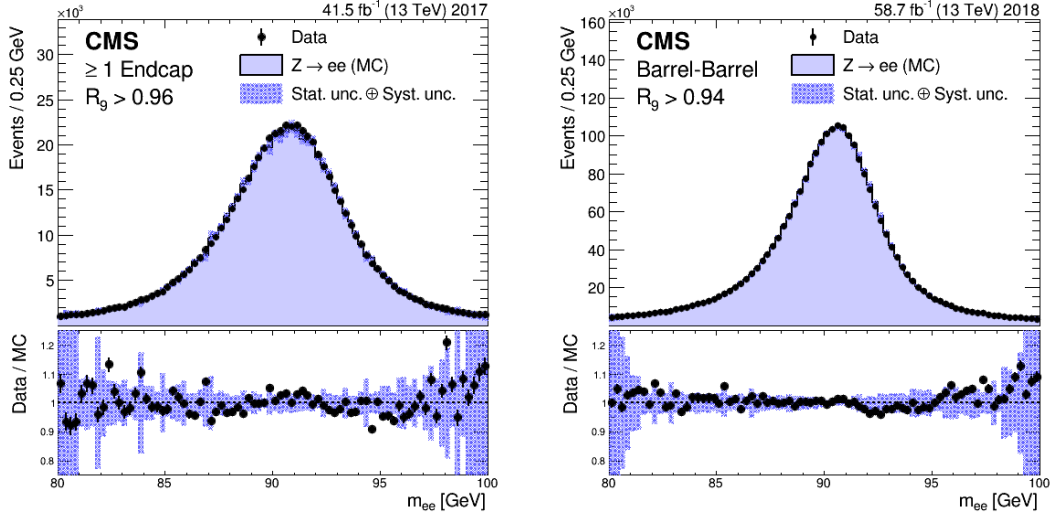


Figure 4-31 Invariant mass distribution of $Z \rightarrow ee$ events shown for 2017 endcap (left) and 2018 barrel (right) electrons.

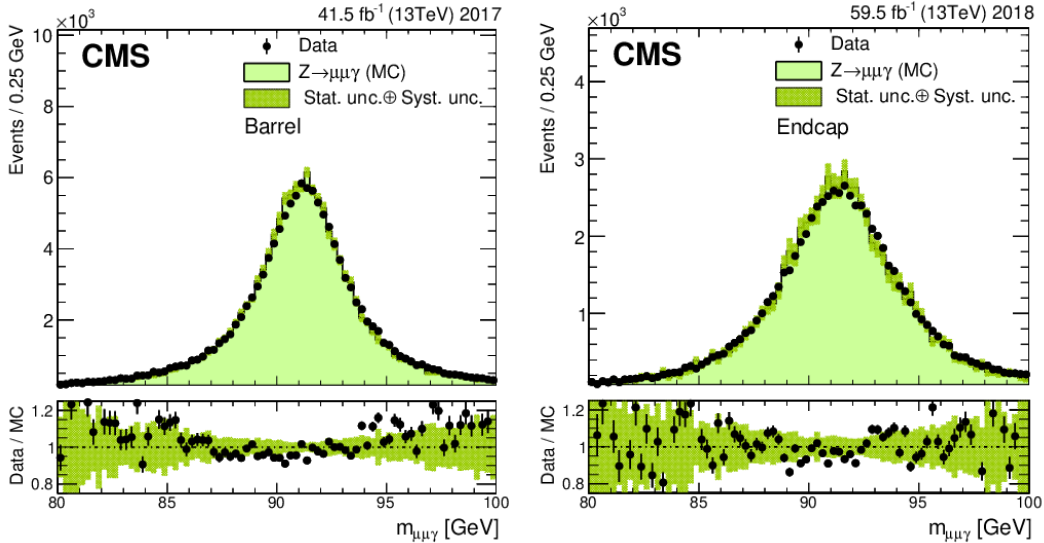


Figure 4-32 Invariant mass distribution of $Z \rightarrow \mu\mu\gamma$ events shown for 2017 barrel (left) and 2018 endcap (right) photons selected from data and simulation.

photons, sample of $Z \rightarrow \mu\mu\gamma$ events used are selected from data having 99% photon purity. The standard dimuon trigger requirements are necessary to be satisfied for both data and MC events. An event is considered if at least two muons passed the tight muon ID requirements [95], with $p_T > 30$ and 10 GeV and $|\eta| < 2.4$. The two muons must be oppositely charged with mass ($m_{\mu\mu}$) > 35 GeV. After successfully identifying the

dimuon system, a photon in the event be reconstructed which must have $|\eta| < 2.5$ and $p_T > 20$ GeV but outside EB and EE transition region. The $\mu\mu\gamma$ system is selected with the requirement that, photon is within $\Delta R = 0.8$ of at least one of muons. Figure 4-32 shows the invariant mass distribution of the $Z \rightarrow \mu\mu\gamma$ system ($m_{\mu\mu\gamma}$) attained after the scale and smearing corrections applied which were derived with electrons from Z boson decay. For both data and MC, the energy scale of photon is extracted by using mean of the distribution of a per event level estimator [96] defined as $s = (m_{\mu\mu\gamma}^2 - m_Z^2)/(m_Z^2 - m_{\mu\mu}^2)$ where m_Z is the Z boson mass. The energy scale difference between data and MC, is corrected by using energy scale corrections which are derived with the electrons originating from the decays of Z boson, both from $Z \rightarrow m_{\mu\mu\gamma}$ events, is less than 0.1% for the photons in EB or in EE. The correction is within the quadratic sum of uncertainties (statistical and systematic) connected with the process of scale extraction which include the uncertainty (systematic) of scale and spreading, along with the systematic uncertainties because of the corrections applied to muon momenta [95].

4.6.3 Photon identification

In order to properly identify the photons, an MVA is trained using the Boosted Decision Tree, so called photon ID BDT, to distinguish the real photons from the ones (non-prompt photons) coming from other objects like jet fragments etc. Non-prompt photons are the ones originating from the neutral meson decays where both photon (prompt and non-prompt) shares the same SC and reconstructed as a single photon candidate. Whereas photons coming from quarks are considered as prompt photons. The training of photon ID BDT is done by using $\gamma + jet$ MC samples with the events passing some preselection criteria. The real photons are treated as signal and fake ones are background. The inputs of this BDT are the same as the ones used in energy regression which includes shower shape and isolation variables. Some other variables are also used as inputs of this BDT such as total number of interaction vertices and the kinematic variables of photons etc. A loose selection on the photon ID BDT is also used i.e. -0.9, in the analyses of Higgs boson property measurements and the search for an additional Higgs boson in the diphoton decay channel.

This BDT is then validated that whether the input and output variables of photon ID MVA are reproduced well. Taking the SM Higgs to diphoton analyses for example, it is done using the $Z \rightarrow ee$ events, where the electrons are constructed as photons

(i.e. neglecting tracks) selected in data and simulation. It's so-called the Tag and Probe (TnP) method. The purity of Z boson decay makes it the most suitable channel for this study over electrons. The $Z \rightarrow ee$ events are selected with an inverted electron veto which is applied as a part of the preselection. The TnP selections are described in the Table 4-4.

Table 4-4 Criteria for tag and probe electrons selection for validation (all years).

Probe electrons:	$p_T > 25 \text{ GeV}$, $ \eta < 2.5$ not in EB-EE gap ($1.4442 \leq \eta \leq 1.566$), passes preselection cuts, has matched pixel seed
Tag electrons:	$p_T > 40 \text{ GeV}$, $ \eta < 2.1$ not in EB-EE gap ($1.4442 \leq \eta \leq 1.566$), $R_9 > 0.8$, photon IDMVA > 0.4 , has matched pixel seed, $Iso_{ch} < 20 \text{ GeV}$, $Iso_{ch}/p_T \leq 0.3$

Alongwith these selections, the tag electrons are also required to undergo the HLT mentioned in the Table 4-5. The validation of the input and output variables of photon

Table 4-5 HLT paths for the tag electrons.

HLT path 2016:	HLT_Ele27_WPTight_Gsf_v
HLT path 2017:	HLT_Ele35_WPTight_Gsf_v
HLT path 2018:	HLT_Ele35_WPTight_Gsf_v

ID BDT are also compared in data and simulation with the help of $Z \rightarrow \mu\mu\gamma$ events. The $Z \rightarrow ee$ validation of the variable Full 5x5 R_9 (one of the shower shape variables) are shown in Figure 4-33.

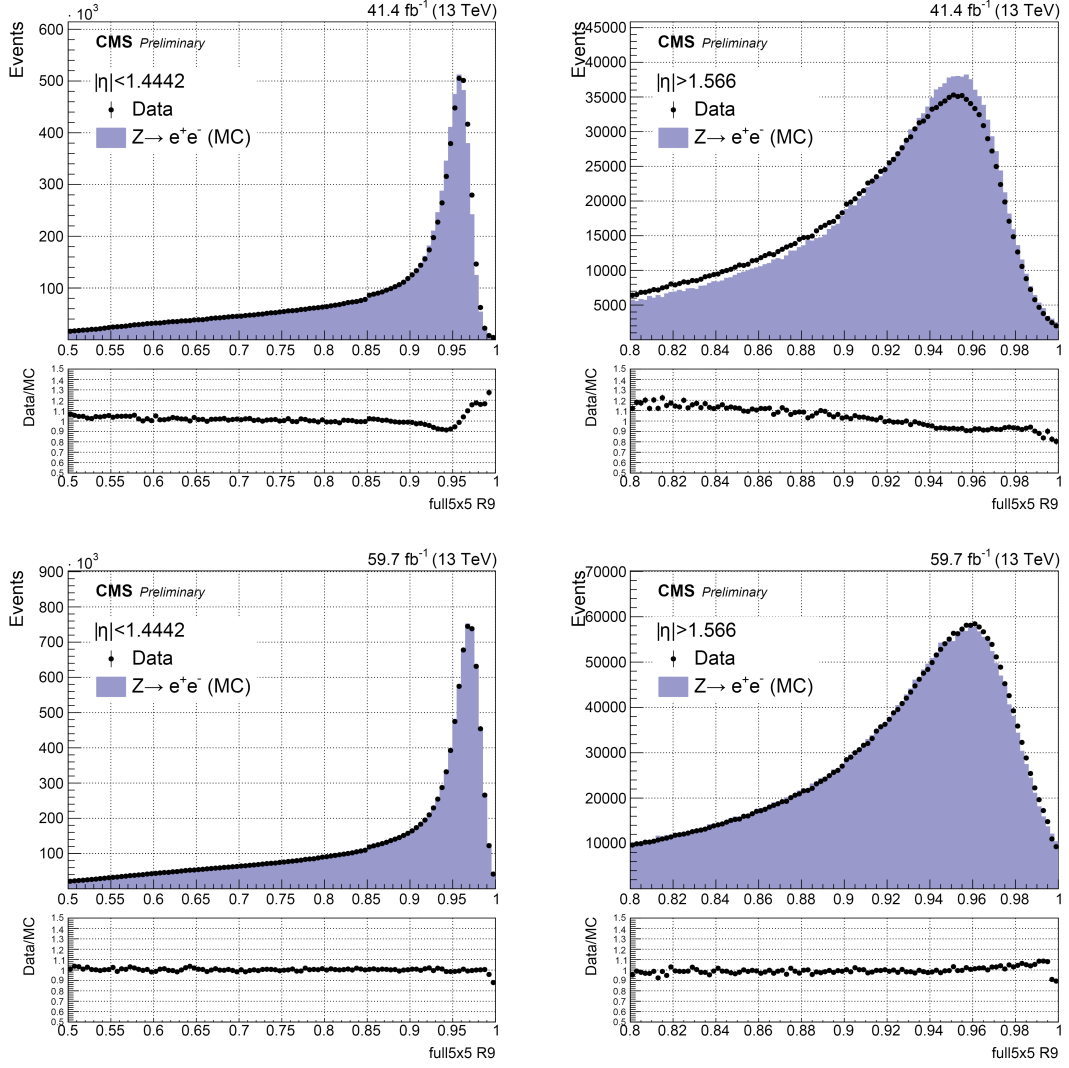


Figure 4-33 Probe electron R_9 distribution in 2017 (top) and 2018 (bottom) data and simulation in ECAL barrel (left) and endcaps (right).

The photon ID MVA output with photons from $Z \rightarrow \mu\mu\gamma$ events are shown in Figure 4-34. The $Z \rightarrow \mu\mu\gamma$ validation of the PF photon isolation and full 5x5 $\sigma_{i\eta i\eta}$ (one of the shower shape variables) are shown in Figure 4-35.

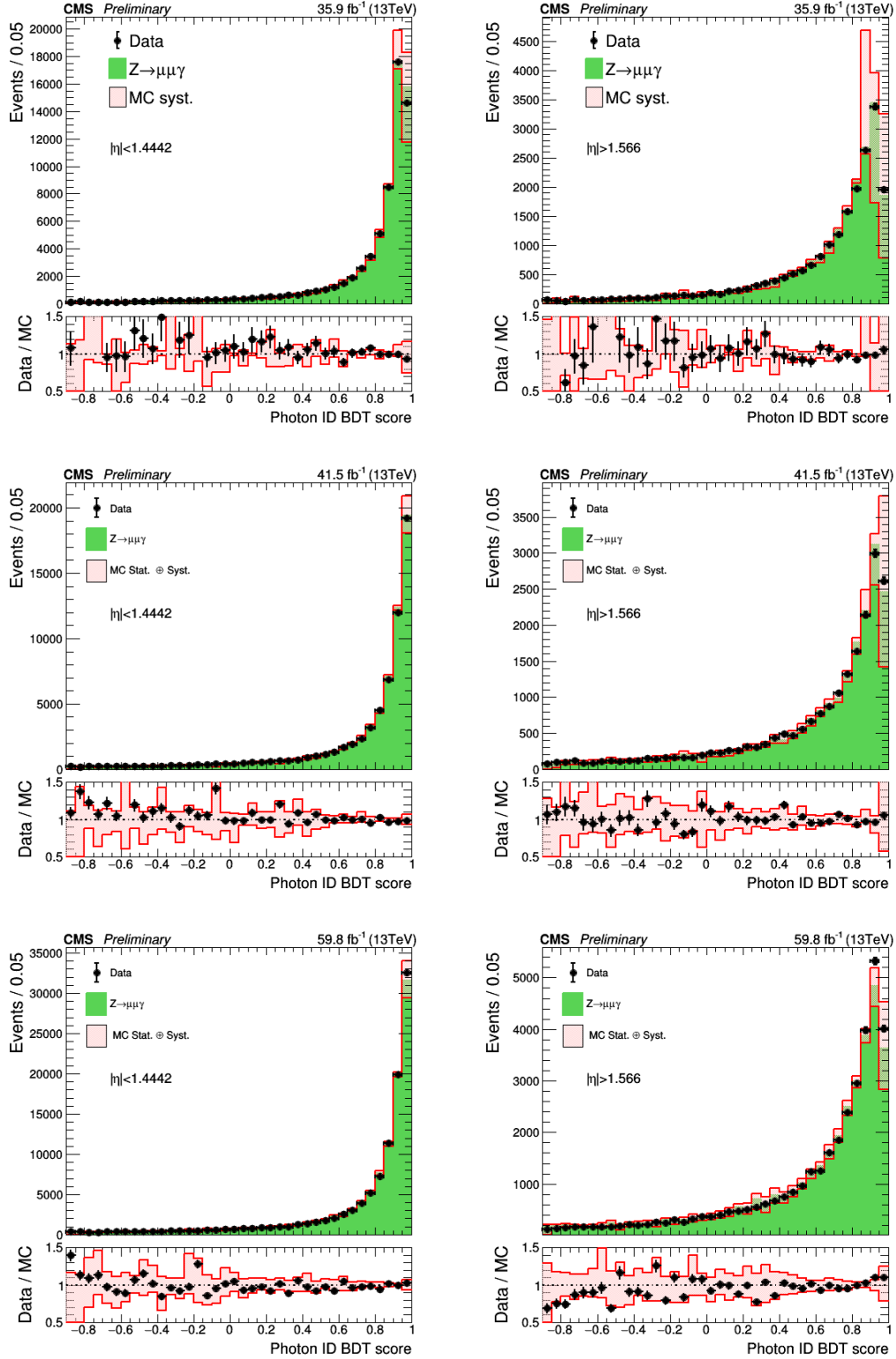


Figure 4-34 Distribution of photon ID MVA output with the photons selected from $Z \rightarrow \mu\mu\gamma$ events in selected with the L2016 (top), UL2017 (middle) and UL2018 (bottom) data and simulation in ECAL barrel (left) and endcaps (right). The systematic uncertainty applied to the shape of the MC distribution is shown by the hashed region.

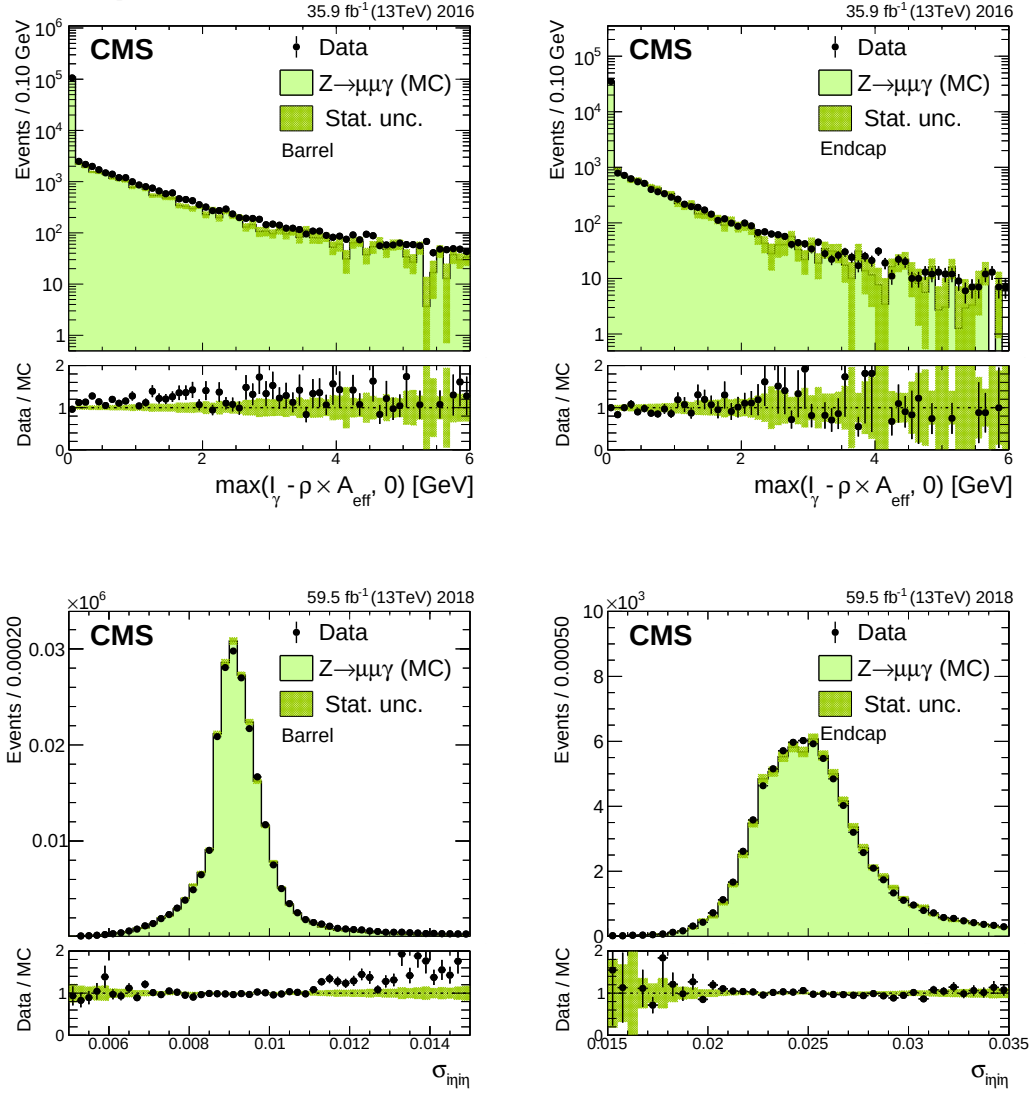


Figure 4-35 Distribution of the ρ -corrected PF photon isolation for 2016 (top) and $\sigma_{in\eta}$ for 2018 (bottom) data and simulation in ECAL barrel (left) and endcaps (right).

4.6.4 Diphoton Vertex

The mass resolution of the diphoton decay channel mainly depend upon two factors i.e. the energy resolution of the photon and the opening angle of two photons. It's also determined by two things such as the position of the SC and vertex location. Due to the extremely good granularity of the ECAL, the position of the SC can be well measured. So, it's crucial to correctly select the primary vertex among the pileup vertices. The photon opening angle resolution does make a tiny contribution to the mass resolution relative to the energy resolution of ECAL as long as the interaction point is considered to be within 1 cm. The mass resolution can be achieved by correctly determining the

reconstructed photons to the interaction vertices reconstructed using charged tracks with the help of CMS algorithm. The methodology used in the study of diphoton decay channel is to choose the best primary vertex by training an MVA using the kinematics of vertex tracks and their correspondence with kinematics of diphoton. Additionally, the information from the tracker for converted photons is also required [81].

In the Higgs boson decay to two photons, the unconverted photons in final states are the ones that are not detected in the ITS and as a result, the primary vertex determination correspond to the signal is not trivial. The recoiling tracks information can be utilized to access the correct vertex position with at least one of the converted photon belongs to the tracker. For Run 2 LHC operations, the CMS tracking algorithm was optimized to better cope up with the expected higher pile-up conditions as the energy was about to change from 8 TeV to 13 TeV.

With the presence of unconverted photons, for the vertex identification, the most discriminating variables are:

$$sumpt^2 = \sum_i |\vec{p}_T^i|^2 \quad (4-1)$$

$$pt_{bal} = - \sum_i (\vec{p}_T^i \cdot \frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|}) \quad (4-2)$$

$$pt_{asym} = (|\sum_i \vec{p}_T^i| - p_T^{\gamma\gamma}) / (|\sum_i \vec{p}_T^i| + p_T^{\gamma\gamma}) \quad (4-3)$$

where $\vec{p}_T^i$ is momentum of the i -th track which corresponds to a given vertex. $\vec{p}_T^{\gamma\gamma}$ is diphoton pair momentum. The sum is over all of the PF candidate having an electric charge, with $p_T > 10$ GeV and attached with the chosen vertex are considered. If one or more than one photons are converted in to electrons within the tracker, two additional variables for the vertex MVA training along with the ones defined above are used:

- Total number of conversions,
- the pull $|z_{vtx} - z_e|/\sigma_z$ between the longitudinal position of the constructed vertex (z_{vtx}) and the vertex (longitudinal position) constructed using conversion track (z_e). The uncertainty on z_e is denoted by variable σ_z .

The probability of choosing the vertex within 1 cm of the real vertex is then used which is computed from the output of a another MVA (vertex probability BDT) using a BDT method same used for vertex identification. The input variables used for this second MVA are:

- The diphoton system p_T ,
- number of vertices,
- the values of vertex ID MVA discriminant for the three best vertices,
- the Δz among the best vertex, the second and third choices,
- converted photons number (0, 1, or 2).

The performance validation of the vertex ID BDT is using $Z \rightarrow \mu^+ \mu^-$ events where the muon tracks are ignored to mimic the diphoton system. Good agreement between data and MC is observed which is shown on the "left" side of Figure 4-36. The vertex choice efficiency, alongwith the estimated vertex probability is shown on the "right" side of Figure 4-36. The performance in MC is shown as a function of both diphoton p_T . It can be seen from the figure that, average vertex probability estimate resembles the true vertex efficiency [97].

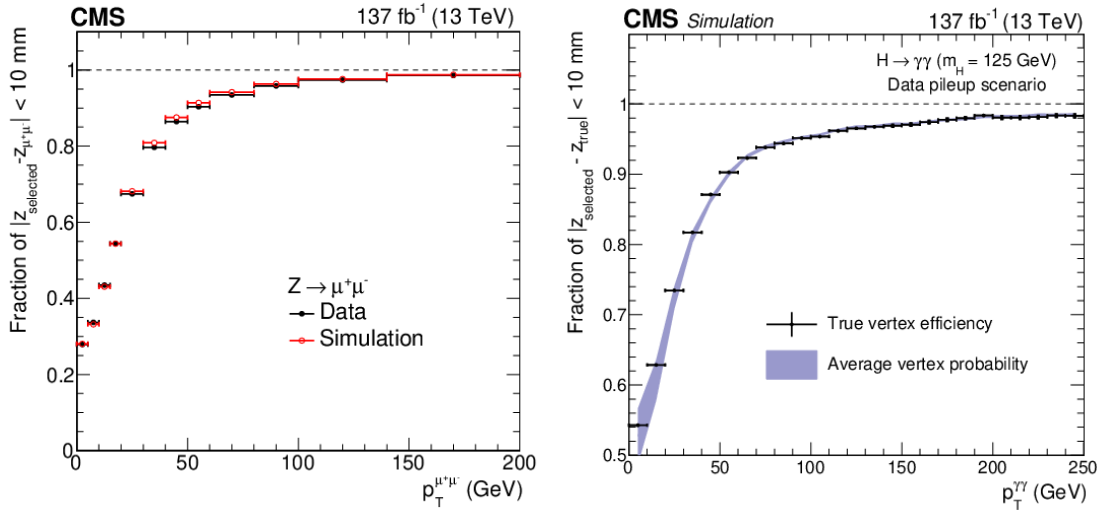


Figure 4-36 (left) $H \rightarrow \gamma\gamma$ vertex identification algorithm validation using $Z \rightarrow \mu^+ \mu^-$ events, omitting the muon tracks with 2017 data and simulation. (Right) Comparison of the true vertex ID efficiency and average estimated vertex probability as a function of reconstructed diphoton p_T .

4.7 Other objects

4.7.1 Electrons reconstruction and identification

Electrons and photons interact electromagnetically, thus can be reconstructed and identified effectively in the CMS. Photons having no charge and deposits its energy

in the EM calorimeter with no tracks in the CMS tracking system. Contrary to this, as electrons exhibit charge so they show tracks in the ITS. A general behaviour of electrons and photons to the detector is shown in Figure 4-30. Reconstruction and identification of the electrons can be done when its supercluster is matched to a track in ITS of the CMS. For electron track reconstruction, the seeding of the ECAL is good for $p_T > 10$ and for lower p_T , seeding from the tracker is (using BDT method for tracker clusters) is more effective. The electron p_T is measured as it traverse through the tracker and produce hits. As it is a charged particle, it exhibit curved trajectory due the presence of magnetic field caused by superconducting magnet and radiate photons. Additionally, a Gaussian Sum Filter (GSF) algorithm is considered in order to report the bremsstrahlung loss, and is exploited for track reconstruction.

After having the loose cut on track-cluster matching, preselection of the electron candidates is made. It is needed actually to maintain highest selection efficiency of electrons and remarkable rejection of QCD background. Electron reconstruction details can be accessed in Ref. [98]. There is a dedicated group in the CMS which is responsible for the reconstructions (electrons/photons) known as E/gamma Physics Object Group (EGM POG). The main goal EGM POG is to develop and validate the tools in order to correctly identify and reconstruct photons and electrons based on the available information from the detector. This information is the later on used for both offline studies and also for the online selection of events. The identification and isolation of electrons is done by training a new BDT which uses the EM cluster information, matching of tracks as well as PF isolation sums.

4.7.2 Muons reconstruction and identification

The information from the CMS muon system is exploited for the reconstruction of muons as well as the track from the tracker (ITS). They are reconstructed as global muons. They can also be reconstructed with the help of at least one muon track created in CSC or DT hits. Muons are also produced in the W or Z bosons decays.

The identification criteria to ensure the muon isolation, is exploited by considering the p_T tracks and deposits in the calorimeter are coherent with the muon track p_T . This way the chances of the mis-identification with the other charged hadrons are reduced. The muon energy can be estimated from inner track curvature for $p_T < 200$ GeV. For the values above 200 GeV, more information is needed from muon tracks is correctly

fitting of the tracks and eventually improves the p_T measurement precision. If a heavy resonance decay into two Z bosons, both of them are highly boosted. The leptons coming from these boosted Z are collinear and the particle flow ID loses its efficiency. As the muons are separated with $\Delta R < 0.4$, they usually originate from Z bosons having p_T greater than 500 GeV. A detailed information related to the reconstruction process of muons can be checked in Reference [99].

4.7.3 Jet reconstruction

The production of particles due to the hadronisation of quarks/gluons which are collimated and usually known as jets. In high energy physics experiments such as at LHC, these jets are produced by making a structure like cones. The existence of quarks and the gluons is also confirmed because of the presence of jets. The jets possess signatures in the components of detector i.e. tracker, EM calorimeter and HCAL. The reconstruction of the jet is done through some jet algorithms. The Vector Boson Fusion (VBF) process of Higgs production differ from the other production mechanisms as the jets are involved in VBF. A general layout of a jet is shown in Figure 4-37 taken from Reference [100]. The jets are also important for the $H \rightarrow \gamma\gamma$ analysis and the studies presented in this dissertation.

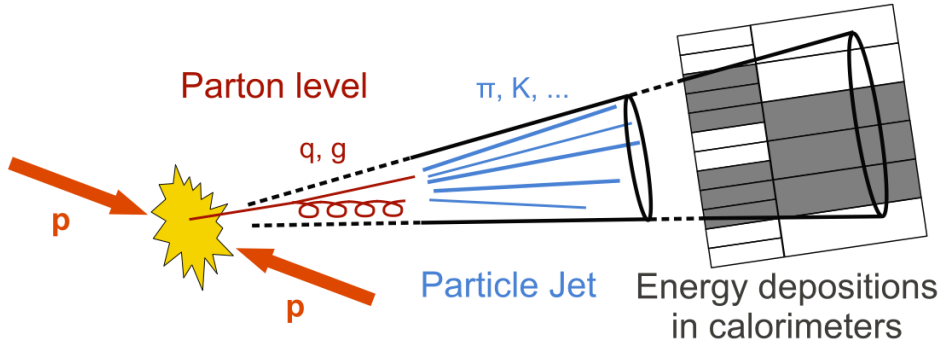


Figure 4-37 A general of a collimated jet in p-p collision.

The jets in the CMS are reconstructed with the help of anti- k_T clustering algorithm with the distance parameter $R = 0.4$ by charged hadrons rejection which are connected with a primary vertex (pileup) [101]. These algorithms perform iterations which create a distance metric based upon the p_T and angular parameters of energy clusters and the combine these clusters together starting with two nearest objects. As soon as the

minimum distance is achieved between an object and beampipe, the process quit and determine these clustered objects as a jet. The process keeps on iterating until all the clusters are composed as jets.

As the tracker coverage is up to only $|\eta|=2.5$, so the jets forward than this do not have removed pileup. After all the jets are in place, a set of corrections are made, in order to correct the energy scale and the resolution of jet in data and MC. It is estimated as function of p_T and η to include the variation throughout the detector. The procedure how these corrections are applied can be accessed in Ref. [102]. Additionally, the jets from pileup are removed with the help of pileup jet ID standards provided by the JetMET physics object group.

The jets from the pileup are normally more dispersed than real ones, as they are usually made up of multiple jets overlapping having low p_T values. A pileup jet ID BDT is trained utilizing the shape variables of the jets and also the track variables corresponding to jet interaction vertex. The BDT output score is actually used for pileup rejection with p_T and η threshold of jets. In simulation, L1FastJet, L2Relative, and L3Absolute corrections have been applied. In data, after the three correction levels, L2L3Residual corrections are applied additionally. In the current analysis, the jets are considered if they pass the pileup jet ID standards and have $p_T > 20$ GeV with $|\eta| < 4.7$.

Chapter 5: Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state with full Run 2 dataset

5.8 Introduction

In the standard model (SM) of particle physics, particle acquire masses with the phenomena of spontaneous electroweak symmetry breaking (EWSB). It is implemented through the Brout-Englert-Higgs mechanism [46]. In 2012, the ATLAS [4] and CMS [5] experiments at CERN LHC, claims the existence of new particle called the Higgs boson with an invariant mass of 125 GeV. The properties of Higgs boson till date are compatible with the SM Higgs boson.

The physics beyond the SM (BSM) is capable of providing a Higgs state that is in accordance with the observed 125 GeV boson and also the extended space of some BSM models like, Next-to-Minimal Supersymmetric Model (NMSSM) [6–8], the generalized two-Higgs-doublet models (2HDM) [9, 10] and the Georgi-Machacek (GM) model [11, 12]. These model give rise to an interesting phenomenology containing several Higgs bosons and some of them may have masses below 125 GeV. Therefore, it is well-motivated to extend the lower search range of the Higgs boson mass as much as possible below $m_H = 110$ GeV.

The $H \rightarrow \gamma\gamma$ decay mode furnish a clean final state topology which permit the mass of Higgs boson to be reconstructed precisely. Gluon fusion is the primary production mode of the Higgs boson and an additional contributions, relatively smaller from vector boson fusion (VBF), associated production with a W or Z boson, or with a $t\bar{t}$ pair. The dominant background sources are irreducible direct production of diphoton, reducible $pp \rightarrow \gamma + \text{jet}$ and $pp \rightarrow \text{jet} + \text{jet}$ processes.

The study presented in this chapter is the probe for an additional Higgs boson with an invariant mass lower than 110 GeV in the diphoton decay mode. This search is performed on full dataset taken by the CMS detector at 13 TeV in 2016, 2017 and 2018. The results from these years are then combined statistically.

In this study we search for an excess of diphoton events over a smoothly falling background because of prompt production of diphoton and events with one jet which is

misidentified as a photon. An extended method is developed as compare to what used in the observation of properties of a SM Higgs boson discovered in 2012 [103]. The mass range investigated is $70 \text{ GeV} < m_{\gamma\gamma} < 110 \text{ GeV}$. MVA methods are used for the photon event classification and identification in order to gain the best sensitivity. A fit is performed on the background to extract the signal.

As compared to previous published results with 2016 data [14] various elements have been updated and optimized, such as diphoton BDT and photon ID MVA in the current studies with the full Run2 data. The suppression of relic DY events having a mass peak not very far from the observed excess in the previous results with 2016 data, is done by introducing two additional cuts as our additional DY suppression strategy for the studies with full Run 2 dataset. The trigger pre-firing is considered for 2016 and 2017 datasets which is same as the standard $H \rightarrow \gamma\gamma$ analysis which was not taken in account in the previously published studies with 2016 dataset. The scale factors are also studied for two additional cuts of DY suppression with events in data and MC for all of the three years. The "Simultaneous Signal Fitting" method is used for the signal modelling which was not used in the previous low-mass analysis results with 2016 data. An exponential is introduced to the double-sided Crystal Ball (DCB) function to well model the relic Drell-Yan events in the background modelling. In some of the event classes of 2016, 2017 and 2018 analyses, significant bias was observed because of the relic DY, for the signal masses closer to the fitted peak of DY events in each event class. The issue was resolved by inflating the normalization uncertainties of DY background and the bias was reduced to some acceptable values.

The isolation and differential shower shape corrections are used as what explained in the standard $H \rightarrow \gamma\gamma$ analysis. The photon ID MVA and the systematic uncertainty for MC photon ID BDT score (EB and EE) is exploited from the standard $H \rightarrow \gamma\gamma$ studies with the 2016 legacy data. The diphoton BDT has been retrained using MC samples in the low mass region are used. The optimization of boundaries is redone based on the retrained diphoton BDT.

In the studies with 2017 dataset, various elements are updated or optimized, as compare to what used in the previously published results with 2016 dataset. The trigger requirements were different in the 2017 HLT path. So there is a difference in preselection of the events 2017 relative to 2016 analysis. The photon ID BDT and diphoton BDT was

retrained for 2017 analysis using MC samples of low mass region and the boundaries of categories (Untagged classes) were also optimized based on retrained diphoton BDT. An additional VBF tagged event class is introduced with the training method same as used in the standard $H \rightarrow \gamma\gamma$ studies [104] in order to enhance the sensitivity of analysis. The dijet and combined (dijet plus diphoton) BDT is trained using the samples of low mass region. The studies with VBF tagged events are also of much interest for various BSM models.

The analysis strategy with 2018 data set is inherited from 2017 analysis. The weights files for photon ID MVA, diphoton BDT, dijet and combined (diphoton plus dijet) BDT used for VBF class are exploited in 2018 analysis after the validation of weights with data and simulation from 2018. The class boundaries (Untagged and VBF) are same as what used in 2017 analysis. The tighter preselection in 2018 applied which is different from 2017 because of different HLT in 2018 and as a result, the trigger efficiencies are not same as the 2017 analysis. Secondly, in 2018 studies we use differential (shower shape and isolation) corrections as used in the standard $H \rightarrow \gamma\gamma$ analysis. The systematic uncertainty in simulation of photon ID BDT score in EB and EE (doubled) is taken directly from standard $H \rightarrow \gamma\gamma$ analysis. For the analysis with 2018 data, the elements which are inherited from 2017 are used after obtaining the satisfactory validations results which are also explained later in this chapter to demonstrate that the performance is reasonable in 2018 even after using 2017 weights.

5.9 Data and simulation samples

The data samples mainly used for the studies, corresponds to 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}) and 2018 (54.4 fb^{-1}) datasets collected by CMS in p-p interactions at 13 TeV center-of-mass energy. The 2018 trigger (HLT) path, `HLT_Diphoton30_18_R9IdL_AND_HE_AN_IsoCaloId_NoPixelVeto`, was not placed at the start of 2018 data taking, so almost 5 fb^{-1} data could not be used for the studies. As shown in the right side plot of Figure 5-38, the pileup for the true interactions passing HLT in data is somewhat different from full integrated luminosity of 59.7 fb^{-1} . The pileup in 2017 data passing prescaled HLT is not same as the full 2017 integrated luminosity of 41.5 fb^{-1} , as displayed at the left side of Figure 5-38.

The DoubleEG dataset is used for signal for 2016 and 2017 studies as well as Dou-

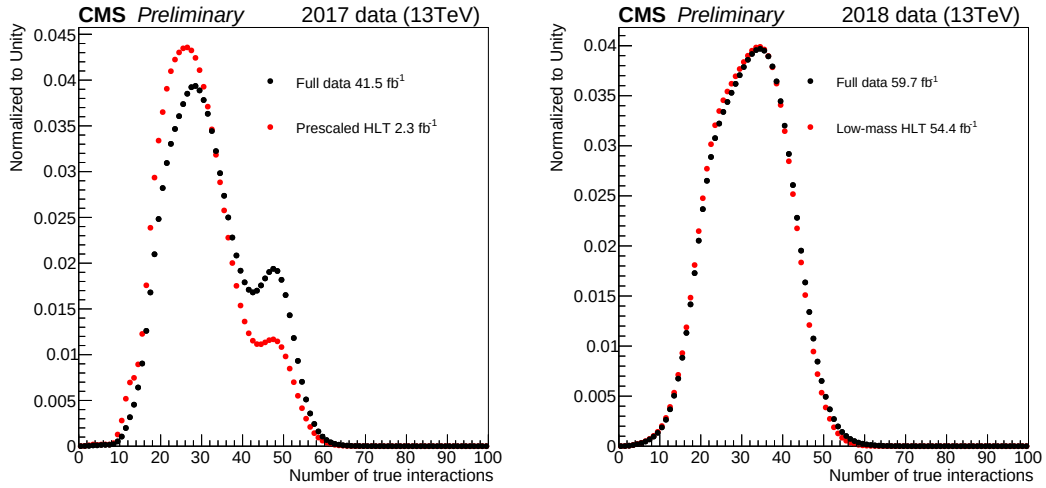


Figure 5-38 Pileup profiles for data taken in 2017 with the full integrated luminosity of 41.5fb^{-1} (black points) and with the prescaled HLT path (red points) in the left. For 2018 data with the full integrated luminosity of 59.7fb^{-1} (black points) and with the 54.4fb^{-1} from the low-mass HLT path (red points) used in the main analysis is in the right.

bleMuon and SingleElectron datasets are exploited for efficiency calculation, calibration and control samples. In 2018, EGamma and DoubleMuon datasets are used for the corresponding studies and for related calibrations and efficiency calculation respectively. In Table 5-6, the data samples used for the years 2016, 2017 and 2018 are placed along with their integrated luminosity.

Signal samples corresponding to four major Higgs production mechanism are produced with POWHEG and MADGRAPH_aMC@NLO v2.4.2 generators. However, the MADGRAPH_aMC@NLO samples are used in order to carry out the studies presented here for the low-mass search analysis. The samples at the parton level are interlinked to PYTHIA8.230 for the purpose of parton showers and hadronization. The recommended cross-sections and branching ratios by LHCXSWG are used [59]. The signal samples are generated with 5 GeV granularity from 65 GeV to 115 GeV Higgs mass values which are to be used for signal modeling in all of the three years. In order to increase the number of events for dijet and combined (diphoton plus dijet) MVA training used for VBF categorization, the 60 GeV signal MC samples are also used. The simulated **background samples** are produced using various generators. The matrix-element generator MADGRAPH is interlinked with PYTHIA8 and SHERPA which also generate particle showers. SHERPA (v.2.2.4) is used to model the prompt-prompt background for diphoton. It also includes the born processes having up to 3 jets at lowest-order

(LO) along with the box processes at LO. The simulation of DY samples is done by using MADGRAPH_aMC@NLO. In Table 5-7, all of the background samples are listed used in the current studies with their corresponding events counts and cross-section, for 2017 and 2018 analyses. The samples used in 2016 analysis which are actually the same background samples but generated using 2016 pileup conditions. All of the background events are generated using same conditions as that of the SM Higgs boson events. These background events are exploited in calculation of shower shape corrections, energy scales and spreading corrections, setting the preselection criteria, photon ID efficiencies and MVA classifier training, estimations of systematic errors and also for validation purposed. The background estimation is done by using the data directly [105].

Table 5-6 Data samples used in the analysis, with their corresponding integrated luminosity per run-periods.

Dataset name	Run-range	$\mathcal{L}$ [fb ⁻¹]
/DoubleEG/Run2016B-17Jul2018_ver2-v1 ¹		
/DoubleEG/Run2016C-17Jul2018-v1 ¹		
/DoubleEG/Run2016D-17Jul2018-v1 ¹		
/DoubleEG/Run2016E-17Jul2018-v1 ¹		36.3
/DoubleEG/Run2016F-17Jul2018-v1 ¹		
/DoubleEG/Run2016G/H-17Jul2018-v1 ¹		
/DoubleEG/Run2017B-31Mar2018-v1/MINIAOD ¹	297047–299329	4.79
/DoubleEG/Run2017C-31Mar2018-v1/MINIAOD ¹	299368–302029	9.63
/DoubleEG/Run2017D-31Mar2018-v1/MINIAOD ¹	302030–302663	4.24
/DoubleEG/Run2017E-31Mar2018-v1/MINIAOD ¹	303818–304797	9.30
/DoubleEG/Run2017F-31Mar2018-v1/MINIAOD ¹	305040–306460	13.54
/EGamma/Run2018A-17Sep2018-v2/MINIAOD		
/EGamma/Run2018B-17Sep2018-v1/MINIAOD	314472–325175	54.4 ^{2,3}
/EGamma/Run2018C-17Sep2018-v1/MINIAOD		59.4 ^{2,4}
/EGamma/Run2018D-22Jan2019-v2/MINIAOD		
/DoubleMuon/Run2016B-17Jul2018_ver2-v1 ¹		
/DoubleMuon/Run2016C-17Jul2018-v1 ¹		
/DoubleMuon/Run2016D-17Jul2018-v1 ¹		
/DoubleMuon/Run2016E-17Jul2018-v1 ¹		36.3
/DoubleMuon/Run2016F-17Jul2018-v1 ¹		
/DoubleMuon/Run2016G/H-17Jul2018-v1 ¹		
/DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD ¹	297047–299329	4.79
/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD ¹	299368–302029	9.63
/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD ¹	302031–302663	4.24
/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD ¹	303824–304797	9.30
/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD ¹	305040–306462	13.54
/DoubleMuon/Run2018A-17Sep2018-v2/MINIAOD		
/DoubleMuon/Run2018B-17Sep2018-v1/MINIAOD	314472–325175	59.5 ^{2,5}
/DoubleMuon/Run2018C-17Sep2018-v1/MINIAOD		
/DoubleMuon/Run2018D-PromptReco-v2/MINIAOD		
/SingleElectron/Run2016B-17Jul2018_ver2-v1 ¹		
/SingleElectron/Run2016C-17Jul2018-v1 ¹		
/SingleElectron/Run2016D-17Jul2018-v1 ¹		
/SingleElectron/Run2016E-17Jul2018-v1 ¹		36.3
/SingleElectron/Run2016F-17Jul2018-v1 ¹		
/SingleElectron/Run2016G/H-17Jul2018-v1 ¹		
/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD ¹	297047–299329	4.79
/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD ¹	299368–302029	9.63
/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD ¹	302030–302663	4.24
/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD ¹	303818–304797	9.30
/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD ¹	305040–306460	13.54

Table 5-7 Background samples used for training and validation purpose in 2017 and 2018 along with their relative number of events, cross-section (pb) and Equiv. Lumi. (pb^{-1}).

2017 Samples	Events	σ	Equiv. Lumi.
/GJet_Pt-20to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/ MINIAODSIM	17058859	220.0	77540.2
/GJet_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/ MINIAODSIM	77049626	850.8	90561.4
/GJet_Pt-20toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/ MINIAODSIM	42042387	3216.0	13072.9
/QCD_Pt-30to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/ MINIAODSIM	14501828	22110.0	655.9
/QCD_Pt-30toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/ MINIAODSIM	41954304	241400.0	173.8
/QCD_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/ MINIAODSIM	20341160	118100.0	172.2
/DiPhotonJetsBox_MGG-40to80_13TeV-Sherpa/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/ MINIAODSIM	3849864	303.2	12697.4
/DiPhotonJetsBox_MGG-80toInf_13TeV-Sherpa/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/ MINIAODSIM	49605719	84.4	587745.5
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/ MINIAODSIM	27413121	5765.4	3222.1
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/ RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_ext1-v1/ MINIAODSIM	182100467	5765.4	21405.1
2018 Samples*	Events	σ	Equiv Lumi
/GJet_Pt-20to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v2/ MINIAODSIM	14366641	220.0	65302.9
/GJet_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10205533	850.8	11995.2
/GJet_Pt-20toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15_ext1-v2/ MINIAODSIM	6072002	3216.0	1888.1
/QCD_Pt-30toInf_DoubleEMEnriched_MGG-40to80_TuneCP5_13TeV_Pythia8/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15-v2/MINIAODSIM	10912544	241400.0	45.2
/QCD_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	10895375	118100.0	92.3
/DiPhotonJetsBox_M40_80-Sherpa/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	4881986	303.2	16101.5
/DiPhotonJetsBox_MGG-80toInf_13TeV-Sherpa/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM	14857013	84.4	176031.0
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/ RunIIAutmn18MiniAOD-102X_upgrade2018_realistic_v15_ext2-v1/ MINIAODSIM	193215674	5765.4	22711.5

QCD_Pt-30to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8 is not available

5.10 Event selection and reconstruction

5.10.1 L1 trigger and HLT

In 2016 studies with legacy data, two HLT paths were active during the data collecting period and used in the analysis are shown in Table 5-8.

Table 5-8 HLT Trigger paths employed in 2016 data taking.

Trigger Path	Prescale
HLT_Diphoton30EB_18EB_R9Id_OR_IsoCaloId_AND_HE_R9Id_DoublePixelVeto_Mass55_v7	1
HLT_Diphoton30PV_18PV_R9Id_AND_IsoCaloId_AND_HE_R9Id_DoublePixelVeto_Mass55_v7	1

The HLT_Diphoton30EB_18EB_R9Id_OR_IsoCaloId_AND_HE_R9Id_DoublePixelVeto_Mass55 path (called “OR” path) is used for the photons reconstructed in EB. The HLT_Diphoton30PV_18PV_R9Id_AND_IsoCaloId_AND_HE_R9Id_DoublePixelVeto_Mass55 (called “AND” path) was considered for the photons both in EB and EE. It is needed that they pass the requirements on the R_9 and $CaloID + Iso$ filters. The same strategy employed for the trigger efficiency as used in previous published results with 2016 dataset. The HLT paths used during the 2017 and 2018 data taking are listed in Table 5-9. As the instantaneous luminosity was increased, the “OR” path

Table 5-9 HLT Trigger paths employed in 2017 and 2018 data taking.

Year	Trigger Path
2017	HLT_Diphoton30PV_18PV_R9Id_AND_IsoCaloId_AND_HE_R9Id_PixelVeto_Mass55
2018	HLT_Diphoton30_18_R9IdL_AND_HE_AN_IsoCaloId_NoPixelVeto

was removed in order to sustain the HLT bandwidth as well as collecting signal events as much as possible. Only the “AND” path was kept as mentioned above. This path is dedicated for the low-mass analysis and possess some differences w.r.t. the one used for Standard Higgs analysis. Keeping in mind the complexity 2017 HLT path, a new path was proposed for the 2018 data taking period with a different logic as well as the filters reorganization. In this way, the same efficiency was ensured as that of 2017 path and in the meantime, the rate was reduced by a factor of three providing room for cuts relaxation and to extend the search region. As a result, the mass cut and the pixel veto got removed. The HLT paths for 2016, 2017 and 2018 are constructed in two parts. The

”seeded leg part” and the ”unseeded leg part”. A layout of the trigger requirements, the list and various trigger filters can be seen in Figure 5-39, Figure 5-40 and Figure 5-41 for 2016, 2017 and 2018 respectively.

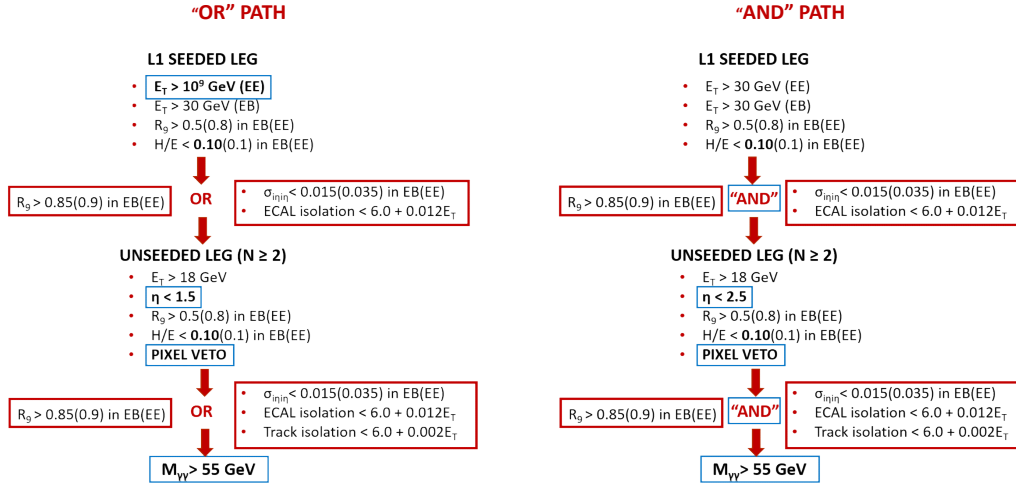


Figure 5-39 Layout of 2016 HLT path showing the list, position and logic of filters.

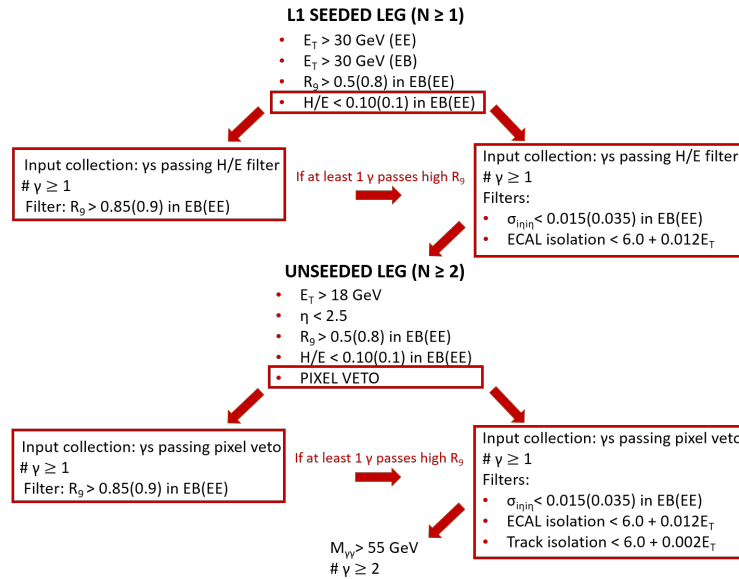


Figure 5-40 Layout of 2017 HLT path showing the list, position and logic of filters.

The **trigger efficiencies** are computed using the tag-and-probe (TnP) technique

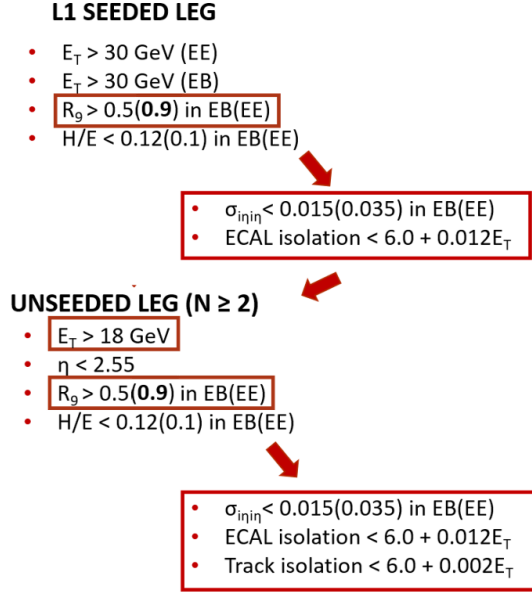


Figure 5-41 Layout of 2018 HLT path showing the list, position and logic of filters.

on $Z \rightarrow ee$ events as done in the standard $H \rightarrow \gamma\gamma$ analysis [106]. By applying the single electron HLT path and with some offline selection, high purity $Z \rightarrow ee$ events can be acquired. In the single electron trigger, the electron from the pair which fires is called "tag" and the second can serve as an unbiased "probe" which is then used to estimate the diphoton path efficiency. All of the tags candidates have to pass the trigger HLT_Ele27_WPTight_Gsf for 2016 and HLT_Ele32_WPTight_Gsf for 2017 and 2018. Additionally, both tags and probes must have to pass the preselection criteria that is used for the analysis (explained later) but with the inverted pixel veto.

The trigger efficiency is defined as the ratio of total number of passing (passing the HLT condition) probes over the number of probes. In 2016 and 2017, cut-and-count method is used while a fitting procedure is employed in 2018. The efficiency is computed for the seeded and unseeded leg separately and then multiplied in order to get a per-event value. In the TnP technique, HLT seeded leg (leading photon candidate) is required to undergo tighter cuts, while for the unseeded leg (sub-leading photon candidate) the selections are relaxed relatively. The trigger efficiencies for 2016 are shown in Figure 5-42 for seeded leg and in Figure 5-43 for unseeded leg.

The 2017 trigger efficiencies are shown in Figure 5-44 for the seeded leg and in Figure 5-45 for the unseeded leg.

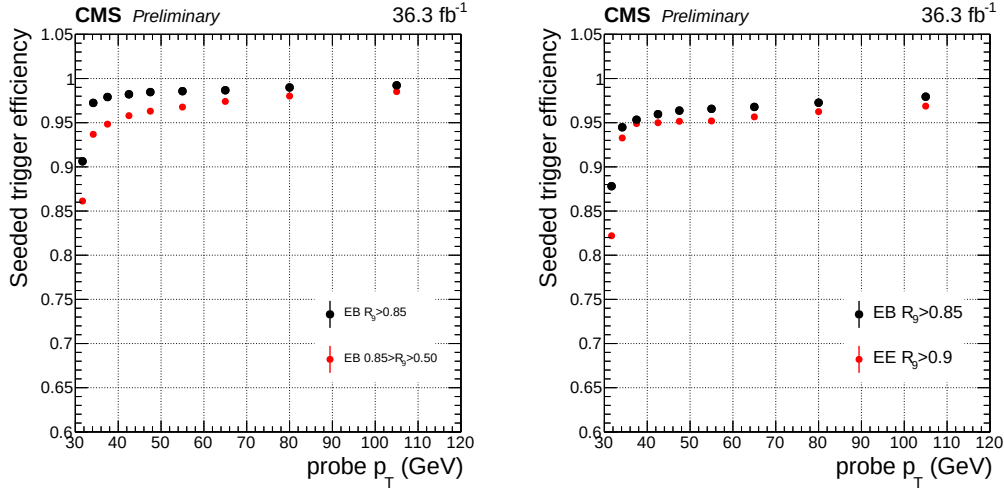


Figure 5-42 (2016) Seeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP technique. Efficiency w.r.t. offline probe p_T with the “OR” path (left) and “AND” path (right).

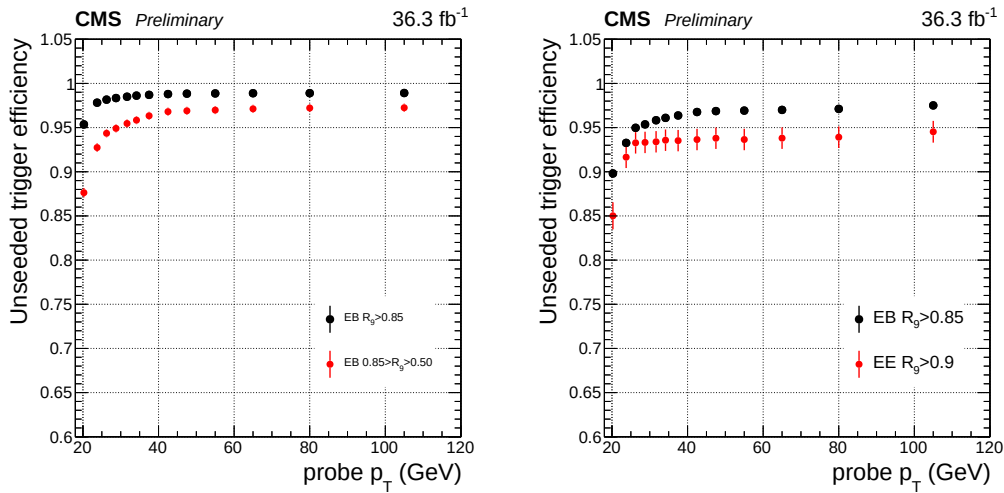


Figure 5-43 (2016) Final trigger efficiency of the “unseeded” leg, measured as the product of the efficiency of the unseeded leg of the standard path up to the mass filter (ϵ_i^{STD}), the ratio between the “OR”/“AND” path efficiency of the unseeded leg up to the pixel veto, the “STD” path efficiency of the unseeded leg just after the H/E filter (SF_i), the pixel-veto efficiency (ϵ_{PV}). The “OR” path results are shown on the left, while the “AND” path shown on right.

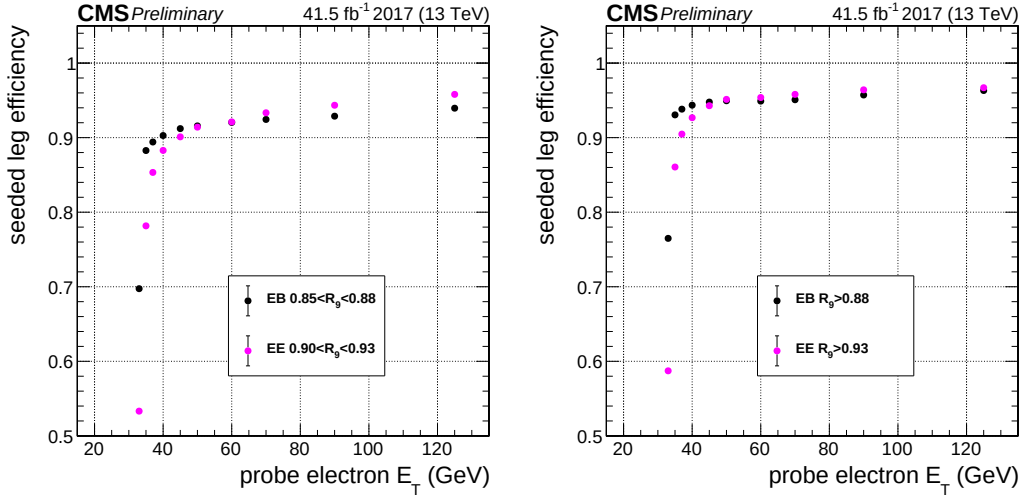


Figure 5-44 (2017) Seeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.

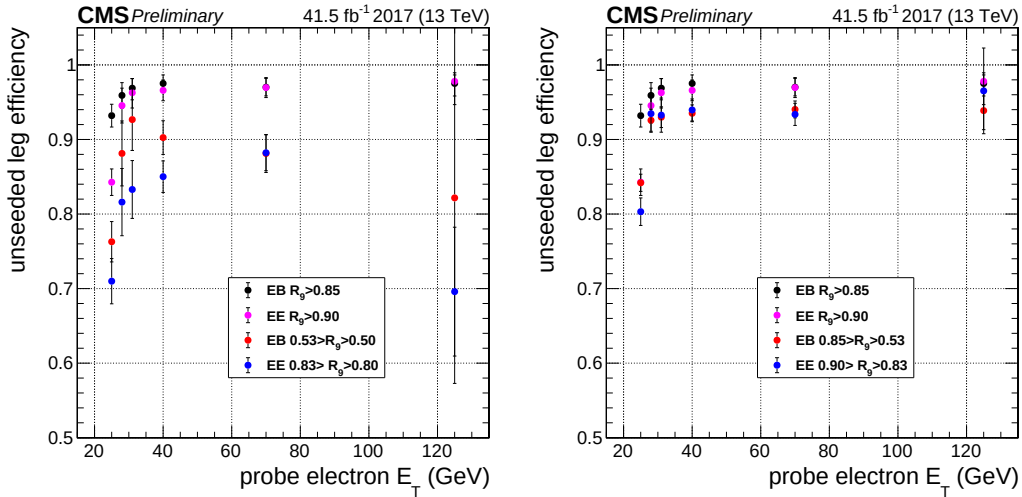


Figure 5-45 (2017) Unseeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.

The trigger efficiencies for 2018 data are shown in Figure 5-46 for the seeded leg and in Figure 5-47 for the unseeded leg.

5.10.2 Photon reconstruction

The photons are reconstructed as a part of global event reconstruction which is described in [107]. They are identified as energy clusters of ECAL. Each crystal energy can also be shared by adjacent clusters considering a transverse profile of EM shower.

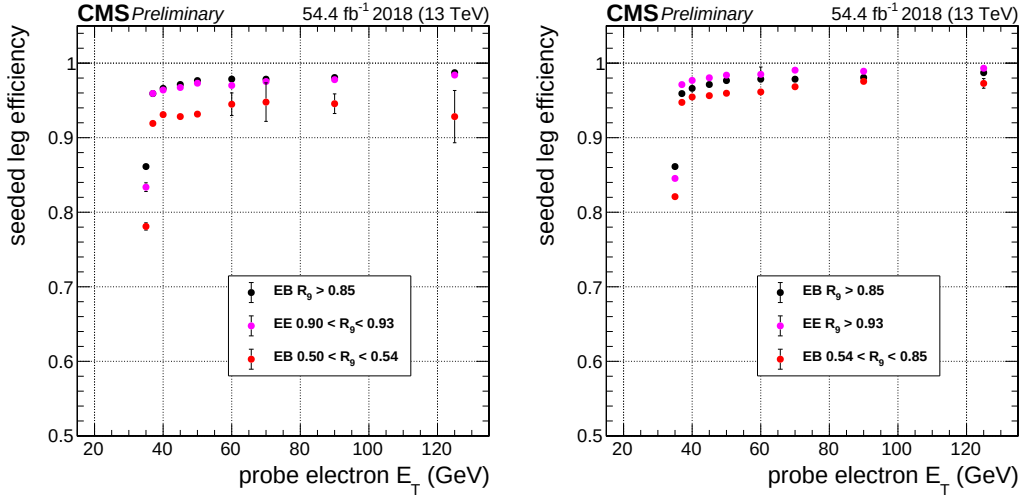


Figure 5-46 (2018) Seeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.

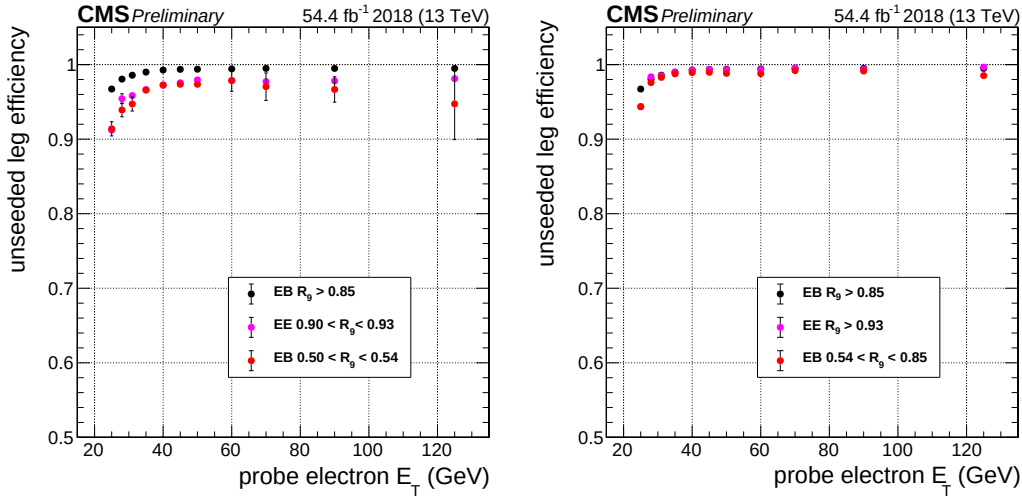


Figure 5-47 (2018) Unseeded leg trigger efficiency measured on data for $Z \rightarrow ee$ events using the TnP method. R_9 turn-on bin on the left and in the R_9 plateau bin on the right.

These clusters are then finally merged in to superclusters (SC's). The photon's energy is then estimated by the sum of the energy of clustered crystals which is calibrated and corrected.

The SC's are then added with the preshower energy in the region where they are installed. In order to optimize the resolution, the energy of the photon is corrected for EM shower containment in the SC's and losses of energy due to the converted photons. The MVA regression method is used to compute the correction for the photon and its uncertainty simultaneously. The inputs are the position variables and the shower shapes

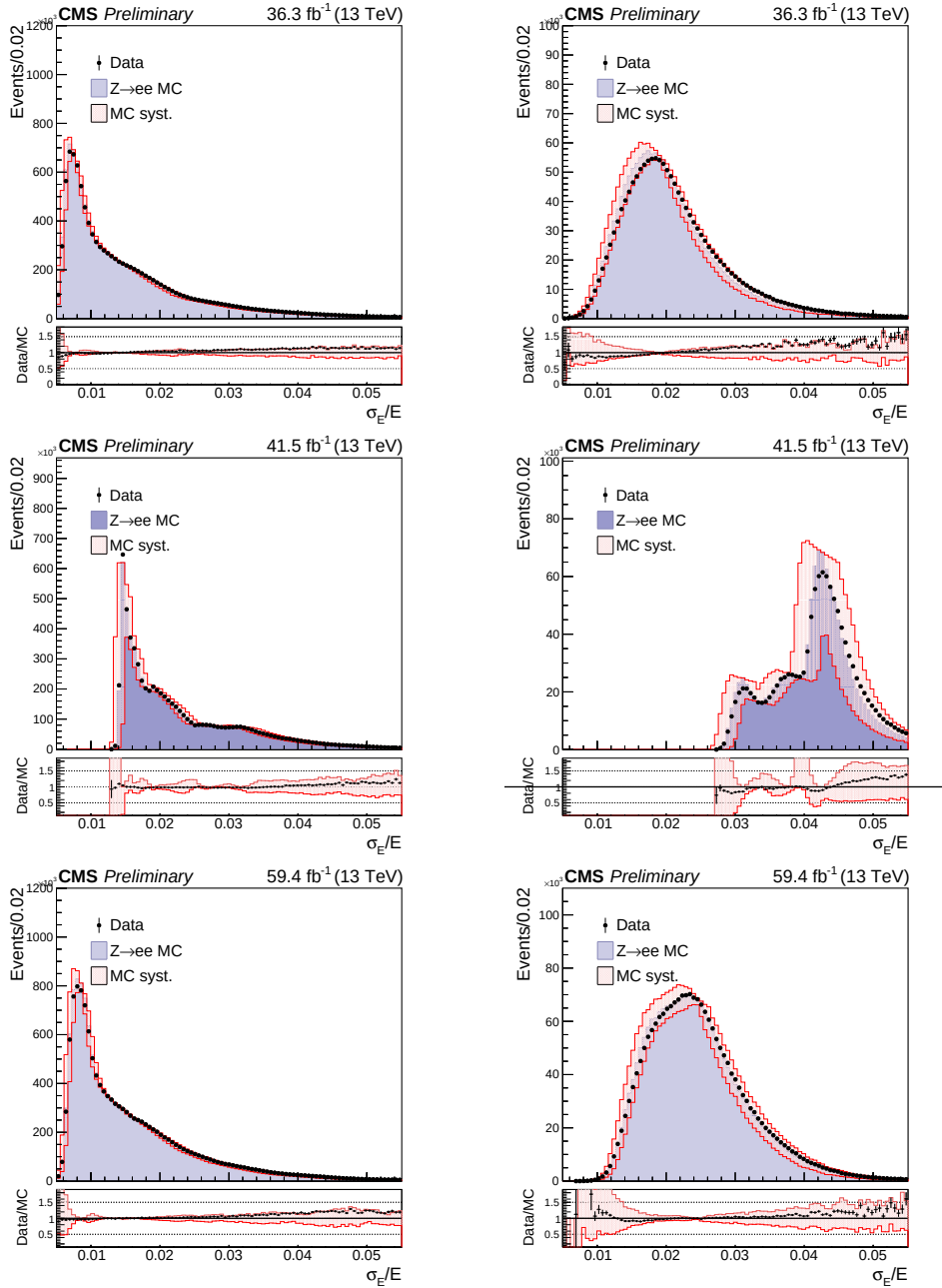


Figure 5-48 Distributions of raw σ_E/E output for 2016 (top), 2017 (middle) and 2018 (bottom) for the EB (left) and EE (right) cases. The electrons from $Z \rightarrow ee$ are reconstructed as photons.

and both of these variables are sensitive to possible un-clustered energy, preshower information and the global event observables that are sensitive to pileup.

The energy resolution (σ_E/E) of photon is validated with electrons originating from the decays of Z boson but reconstructed as photons. Figure 5-48 shows the distributions of σ_E/E for 2016, 2017 and 2018 datasets in barrel and endcaps. The energy resolution

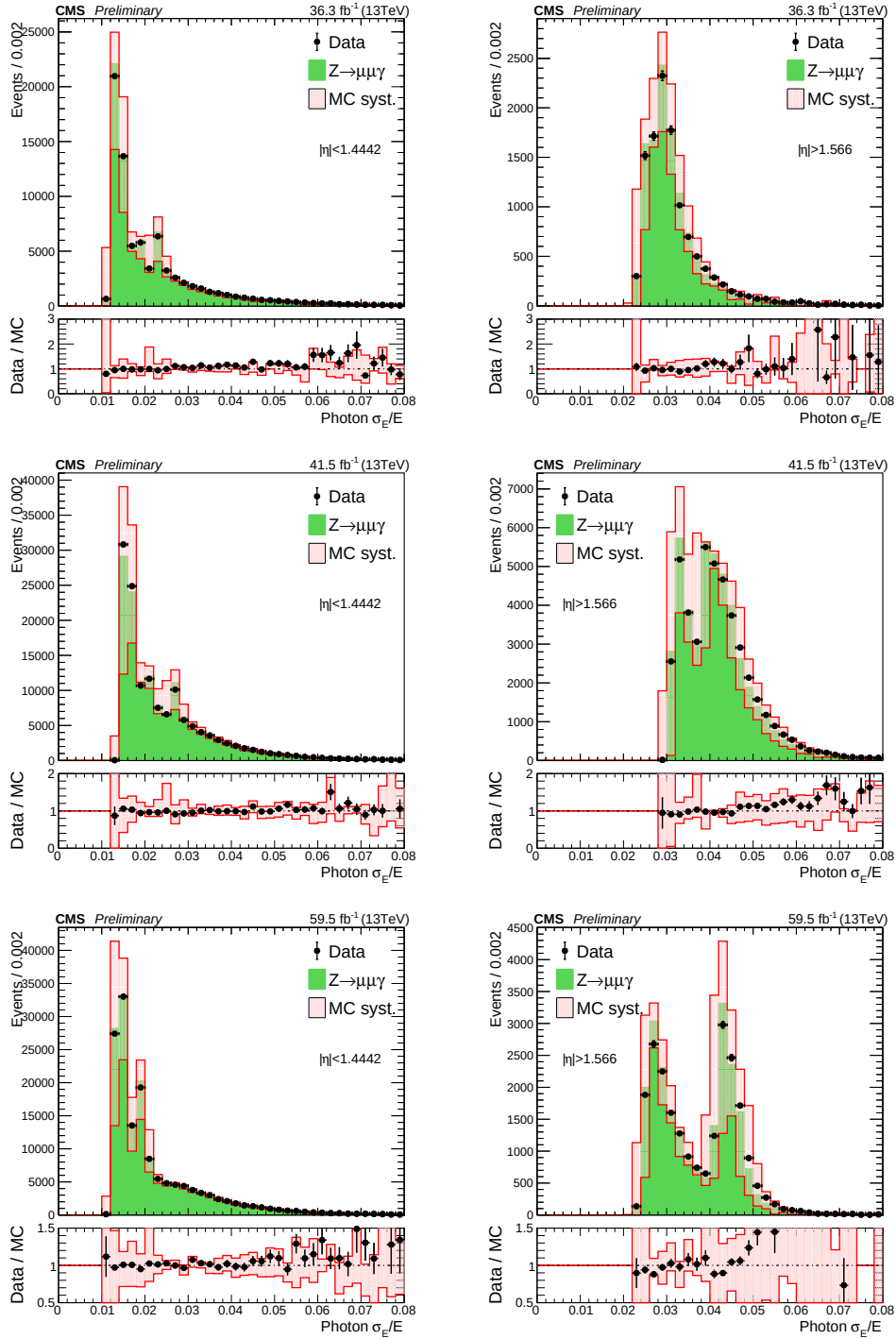


Figure 5-49 Distributions of σ_E/E of photons in top (2016), middle (2017) and bottom (2018) with EB on left and EE on right with photons selected from $Z \rightarrow \mu\mu\gamma$ events. The bottom panel in each figure shows the data to MC ratio together with the systematic uncertainty.

per photon is validated also with the photons selected from $Z \rightarrow \mu\mu\gamma$ events. These validations are shown in Figure 5-49. The systematic uncertainty (red bands in the plots) placed to the obtained shape from MC events relative to a shift of σ_E/E by ± 0.05 .

5.10.3 Photon preselection

All of the diphotons within the geometrical acceptance ($|\eta| < 2.5$ but not in the barrel and endcaps gap corresponding to $1.442 < |\eta| < 1.566$) have to pass some offline preselection criteria which is constructed to reflect the requirements of HLT. This is to keep the coherent phase space in data as well as in simulation (where the HLT is not applied). The preselection requirements are slightly more stringent than the HLT. The variables such as H/E , $\sigma_{i\eta i\eta}$, Photon ISO (ρ corrected), Tracker ISO and R_9 are used in preselection. The leading photon candidate have $p_T > 30$ GeV and the subleading photon must have $p_T > 18$ GeV. Each photon candidate of the diphoton pair has to fulfill the electron veto condition. The photon is removed if it's SC is found to match with the electron track with even one hit in the pixel detector. An MVA based classifier is used for identification which is trained by exploiting isolation, shower shape and kinematic variables. This photon ID MVA aid to discriminate the signal from background photons. A very loose photon ID MVA requirement above -0.9 is applied as a part of further selection. It significantly eliminates a large amount of non-prompt photons and accommodate approximately 100% of prompt photons.

Additionally, a requirement which is designed to minimize the MiniAOD (data format) photon cuts is employed also at the preselection step. Both of the photons have to pass either $R_9 > 0.8$, Charged Hadron Isolation < 20 GeV or (Charged Hadron Isolation) $/p_T < 0.3$ if their $p_T > 14$ GeV and $H/E < 0.15$. The preselection requirements for 2016 analysis are shown in Table 5-10, Table 5-11 for 2017 and Table 5-12 for 2018. In

Table 5-10 2016 preselection requirements.

		R_9	H/E	$\sigma_{i\eta i\eta}$	pfPhoIso	TrackerIso
Both photons in barrel	Barrel	> 0.5	< 0.07	< 0.0105	< 4 GeV	< 6 GeV
At least one in endcap	Barrel	> 0.85	< 0.07	< 0.0105	< 4 GeV	< 6 GeV
At least one in endcap	Endcap	> 0.9	< 0.035	< 0.0275	< 4 GeV	< 6 GeV

2017 analysis, an additional requirement is set, the diphoton pair is selected in a way that

Table 5-11 2017 preselection requirements.

	H/E	$\sigma_{i\eta i\eta}$	R_9	pfPhoIso	TrackerIso
EB; $R_9 > 0.85$	< 0.08	< 0.015	> 0.5	< 4.0	< 6.0
EB; $R_9 \leq 0.85$	< 0.08	< 0.015	> 0.5	< 4.0	< 6.0
EE; $R_9 > 0.90$	< 0.08	< 0.035	> 0.8	< 4.0	< 6.0
EE; $R_9 \leq 0.90$	< 0.08	< 0.035	> 0.8	< 4.0	< 6.0

Table 5-12 2018 preselection requirements.

	H/E	$\sigma_{i\eta i\eta}$	R_9	pfPhoIso	TrackerIso
EB; $R_9 > 0.85$	< 0.08	< 0.015	> 0.5	< 4.0	< 6.0
EB; $R_9 \leq 0.85$	< 0.08	< 0.015	> 0.5	< 4.0	< 6.0
EE; $R_9 > 0.90$	< 0.08	< 0.035	> 0.8	< 4.0	< 6.0

at least one of the photon candidate should be in the higher R_9 region. Finally, for both 2017 and 2018 studies, the diphoton pair is required to satisfy $55 < m_{\gamma\gamma} < 120$ GeV. The efficiencies of preselection and loose photon ID MVA requirement apart from the electron veto, are computed by using $Z \rightarrow ee$ events. In order to estimate the efficiency of electron veto, $Z \rightarrow \mu\mu\gamma$ events are taken into consideration. The efficiencies and scale factors of preselection and loose photon ID MVA requirement are shown in Table 5-13 and Table 5-14 respectively for 2016, Table 5-15 and Table 5-16 respectively for 2017, and Table 5-17 and Table 5-18 respectively for 2018 data. For the pixel-seed electron veto requirement, the efficiencies and scale factors for 2016, 2017 and 2018 are shown in Tabel 5-19, Tabel 5-20 and Tabel 5-21 respectively.

Table 5-13 (2016) Preselection efficiencies measured in the three η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all preselection criteria except electron veto).

	Data		Simulation		Scale	Ratio	
	Eff.	Stat	Eff.	Stat.		Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.8708	0.0001	0.8774	0.0002	0.9925	0.0003	0.0099
Barrel; $R_9 < 0.85$	0.7395	0.0002	0.7371	0.0004	1.0033	0.0006	0.0321
Endcap; $R_9 > 0.90$	0.6219	0.0003	0.6280	0.0006	0.9903	0.0011	0.0222

Table 5-14 (2016) Loose photon ID MVA cut efficiencies measured in the three η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all ID MVA criterion except electron veto).

	Data		Simulation		Scale	Ratio	
	Eff.	Stat	Eff.	Stat.		Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.9996	< 0.0001	0.9996	< 0.0001	1.0000	< 0.0001	0.0001
Barrel; $R_9 < 0.85$	0.9976	< 0.0001	0.9974	0.0001	1.0002	0.0001	0.0004
Endcap; $R_9 > 0.90$	0.9993	< 0.0001	0.9992	< 0.0001	1.0001	< 0.0001	0.0003

Table 5-15 (2017) Preselection efficiencies measured in the four η , $R_9 <$ categories using tag and probe with $Z \rightarrow ee$ events (for all preselection criteria except electron veto).

	Data		Simulation		Scale	Ratio	
	Eff.	Stat	Eff.	Stat.		Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.8699	0.0001	0.8756	0.0002	0.9934	0.0003	0.0139
Barrel; $R_9 < 0.85$	0.5986	0.0002	0.5956	0.0003	1.0050	0.0004	0.0333
Endcap; $R_9 > 0.90$	0.8475	0.0002	0.8494	0.0003	0.9977	0.0004	0.0070
Endcap; $R_9 < 0.90$	0.5739	0.0004	0.5770	0.0006	0.9946	0.0012	0.0155

Table 5-16 (2017) Loose photon ID MVA cut efficiencies measured in the four η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all ID MVA criterion except electron veto).

	Data		Simulation		Ratio		
	Eff.	Stat	Eff.	Stat.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.9943	< 0.0001	0.9963	< 0.0001	0.9979	< 0.0001	0.0017
Barrel; $R_9 < 0.85$	0.9803	0.0001	0.9852	0.0001	0.9950	0.0001	0.0026
Endcap; $R_9 > 0.90$	0.9951	0.0001	0.9950	0.0001	1.0001	0.0001	0.0012
Endcap; $R_9 < 0.90$	0.9730	0.0002	0.9723	0.0003	1.0007	0.0004	0.0022

Table 5-17 (2018) Preselection efficiencies measured in the three η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all preselection criteria except electron veto).

	Data		Simulation		Ratio		
	Eff.	Stat	Eff.	Stat.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.8677	0.0001	0.8631	0.0002	1.0053	0.0003	0.0201
Barrel; $R_9 < 0.85$	0.8160	0.0002	0.8105	0.0003	1.0068	0.0005	0.0588
Endcap; $R_9 > 0.90$	0.7761	0.0002	0.7588	0.0004	1.0228	0.0006	0.0060

Table 5-18 (2018) Loose photon ID MVA cut efficiencies measured in the three η , R_9 categories using tag and probe with $Z \rightarrow ee$ events (for all ID MVA criterion except electron veto).

	Data		Simulation		Ratio		
	Eff.	Stat	Eff.	Stat.	Scale	Stat. Unc.	Syst. Err
Barrel; $R_9 > 0.85$	0.9952	< 0.0001	0.9954	< 0.0001	0.9998	< 0.0001	0.0007
Barrel; $R_9 < 0.85$	0.9841	0.0001	0.9829	0.0001	1.0012	0.0001	0.0029
Endcap; $R_9 > 0.90$	0.9939	< 0.0001	0.9930	0.0001	1.0009	0.0001	0.0006

Table 5-19 (2016) Efficiency of the pixel-seed electron veto, measured in the three photon categories using tag and probe with 2016 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.

	Data		Simulation		Scale Factor	
	Eff.	Stat.	Eff.	Stat.	Eff.	Unc.
Barrel; $R_9 > 0.85$	0.9613	0.0009	0.9775	0.0014	0.9834	0.0017
Barrel; $R_9 < 0.85$	0.8828	0.0026	0.8751	0.0056	1.0088	0.0071
Endcap; $R_9 > 0.90$	0.8329	0.0035	0.8695	0.0064	0.9580	0.0081

Table 5-20 (2017) Efficiency of the pixel-seed electron veto, measured in the four photon categories using tag and probe with 2017 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.

	Data		Simulation		Scale Factor	
	Eff.	Stat.	Eff.	Stat.	Eff.	Unc.
Barrel; $R_9 > 0.85$	0.9382	0.0010	0.9619	0.0013	0.9754	0.0017
Barrel; $R_9 < 0.85$	0.8438	0.0026	0.8781	0.0042	0.9610	0.0054
Endcap; $R_9 > 0.90$	0.6897	0.0034	0.7495	0.0052	0.9202	0.0078
Endcap; $R_9 < 0.90$	0.4561	0.0081	0.4979	0.0133	0.9161	0.0293

Table 5-21 (2018) Efficiency of the pixel-seed electron veto, measured in the three photon categories using tag and probe with 2018 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.

	Data		Simulation		Scale Factor	
	Eff.	Stat.	Eff.	Stat.	Eff.	Unc.
Barrel; $R_9 > 0.85$	0.9343	0.0008	0.9656	0.0013	0.9676	0.0015
Barrel; $R_9 < 0.85$	0.8465	0.0022	0.8742	0.0042	0.9683	0.0053
Endcap; $R_9 > 0.90$	0.6711	0.0028	0.7589	0.0054	0.8844	0.0073

5.10.4 Photon identification

The identification of photon is done by using a photon ID MVA classifier to discriminate prompt photons from non-prompt ones. It include the unconverted photons and mostly the ones converted within tracker sub-detector. Non-prompt are those mainly originating from neutral (high momentum) mesons such as π^0 , which decays into two photons, share same SC (supercluster) and reconstructed as single photon candidate. The prompt photons are the ones coming from quarks, gluons or photons (at generator level). The separation of these two classes of photons, has been done with the BDT classifier with the TMVA 4 package. The BDT score for each of the two selected photons is then used as one of the input variables of diphoton BDT for the diphoton event categorization. The photon ID MVA training has been done by using 2017 $\gamma + jets$ simulated samples. The retraining is also done using 2018 $\gamma + jets$ MC samples and performance was checked with the 2017 photon ID MVA training. The performance of 2018 training was quite similar to 2017 training in a test done on 2018 samples. The 2017 photon ID MVA weights are validated with 2018 $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-\gamma$ data and MC events. In the analysis with 2016 data, the photon ID MVA trained for standard $H \rightarrow \gamma\gamma$ analysis was used.

The BDT is trained by employing 13 TeV MC samples. The signal are the prompt photons and the non-prompt ones are used as background which are taken from the double EM-Enriched $\gamma + jets$ simulated samples. The 2017 simulated samples has been utilized for this task. The samples used are weighted relative to their cross-section, as they do not have similar p_T and mass ranges. The diphoton candidates should pass the preselection criteria which is constructed to have a same phase space for data and simulation. The photon candidate will be considered as signal like prompt photon if matched to a generator level photon with the help of gen matching criteria. The photon candidate is considered as a non-prompt photon if it does not matched to a generator level photon. The prompt and non-prompt photons considered for MVA training have $p_T > 18$ GeV and $m_{\gamma\gamma} > 55$ GeV. The BDT training is made independent of photon kinematic variables, a 2D reweighting which is based on the SC η and the candidate photon p_T of signal photons to background photons is performed. The re-weighted plots are presented in the Figure 5-50.

During the training, BDT classifier differentiate among input variables for prompt

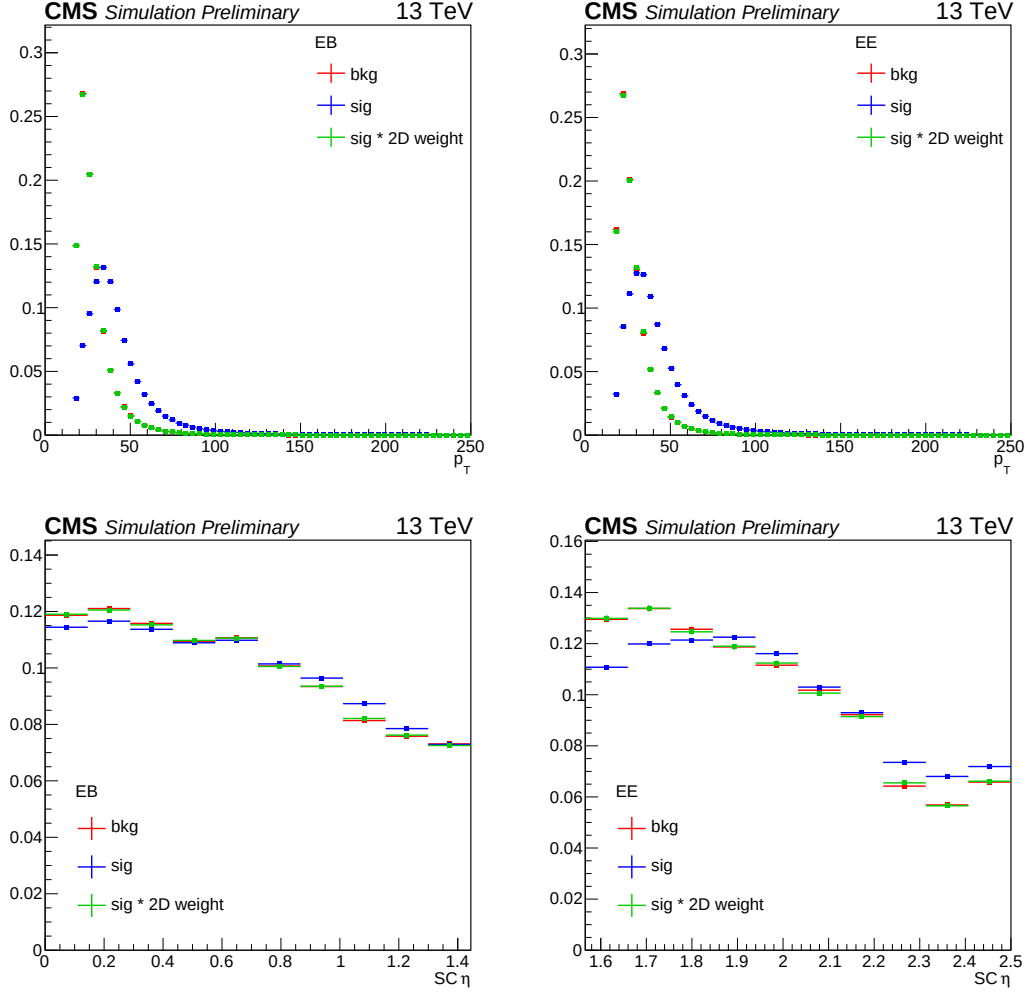


Figure 5-50 (2017) p_T , supercluster η shown for barrel(left), and endcap(right), background photons distributions are shown in red, signal photons are in - blue and the signal photon ones reweighted with 2D $p_T - \eta_{SC}$ weights are shown in green.

and non-prompt photons. For barrel and endcaps, two separate training are performed. The input variables for the training are the photon kinematic, shower shape and isolation variables. ρ , $SC \eta$ are the kinematic variable.. The shower shape variables are $\sigma_{in\eta}$, $Cov_{in\eta\phi}$, $E_{2 \times 2}/E_{5 \times 5}$, R_9 , σ_η and σ_ϕ as defined in section 4.6.1. The isolation variable are the PF Photon ISO, PF Charged ISO (worst vertex) and PF Charged ISO (selected vertex). Supercluster raw energy E_{RAW} is also used for both barrel and endcap photons. For the endcaps, two additional variables are used i.e. preshower spread (σ_{RR}) and E_{ES}/E_{RAW} as defined in section 4.6.1. These two variables help to increase the performance of photon ID in the endcap. The validation of minimum photon ID BDT output score is done by using signal samples and other $H \rightarrow \gamma\gamma$ backgrounds (photon plus jets,

QCD dijet, diPhotons and DYjetsToLL) have been used for 2017 and 2018 data/simulation studies. The photon ID MVA training which is developed for 2017 conditions is applied to diphoton events in data and simulation for 2017 and 2018.

The validation of the photon ID BDT input variables of is done by $Z \rightarrow ee$ events where electrons are considered as reference to photons (i.e. neglecting tracks). The validation process is carried out by exploiting tag and probe technique with the electrons from Z decays in 2016, 2017 and 2018 analysis. The shower shape and isolation corrections are applied to the electrons before making any selections in 2016, 2017 and 2018 simulation samples. The inverted electron veto is then applied to all of the events a part of preselection. Figure 5-51 shows the $Z \rightarrow ee$ validations for two of the variables.

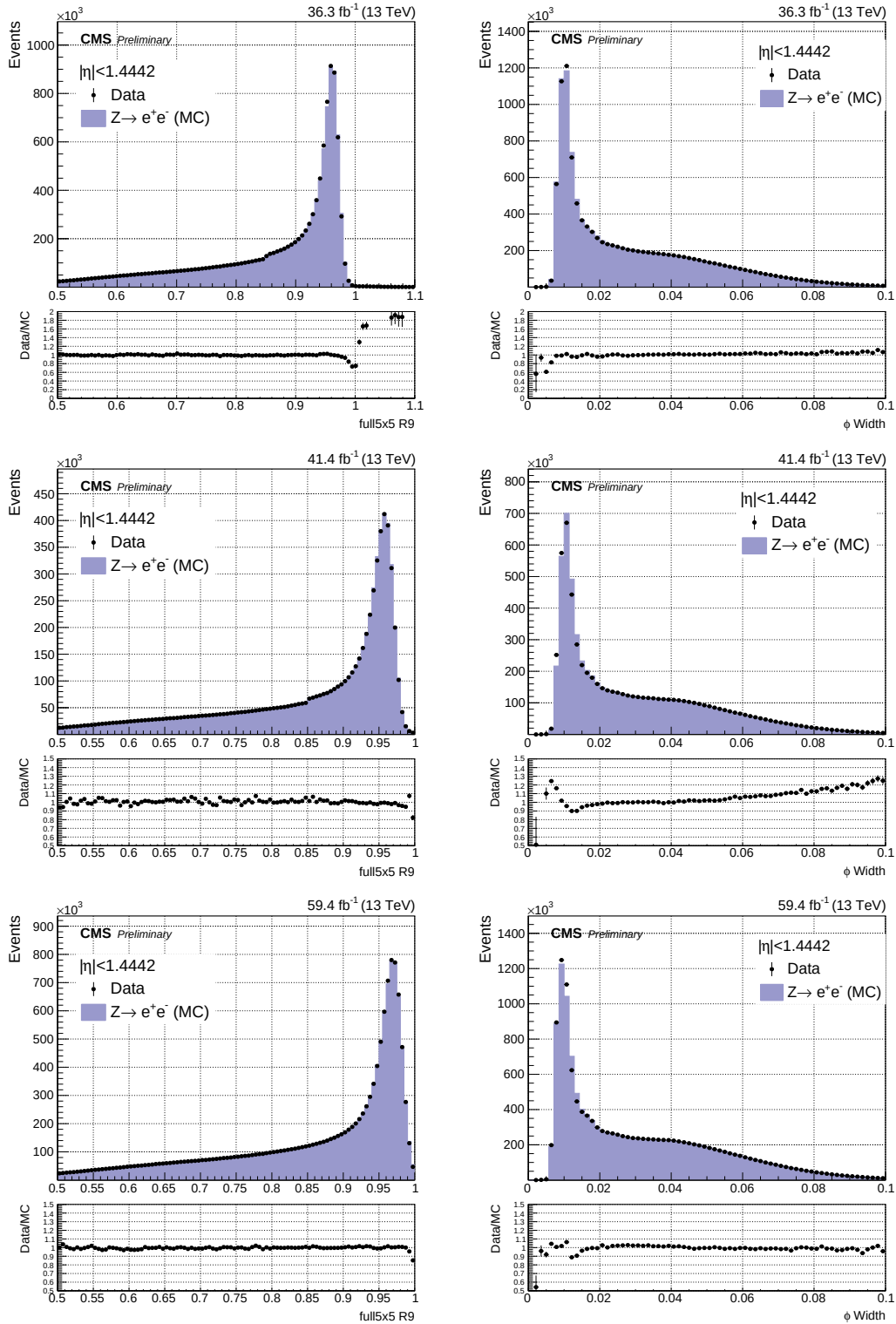


Figure 5-51 Full $5 \times 5 R_9$ (left) and ϕ width (right) for $Z \rightarrow ee$ events in data and simulation in ECAL barrel for probe candidates for 2016 (top), 2017 (middle) and for 2018 (bottom).

The validations are also done for $Z \rightarrow \mu\mu\gamma$ events that are selected in data with high photon purity ($\sim 99.5\%$) after the $Z \rightarrow \mu\mu\gamma$ event selections. The photon candidates has to pass analysis preselections (Section 5.10.3) along with the loose requirement on photon ID with $MVA > -0.9$.

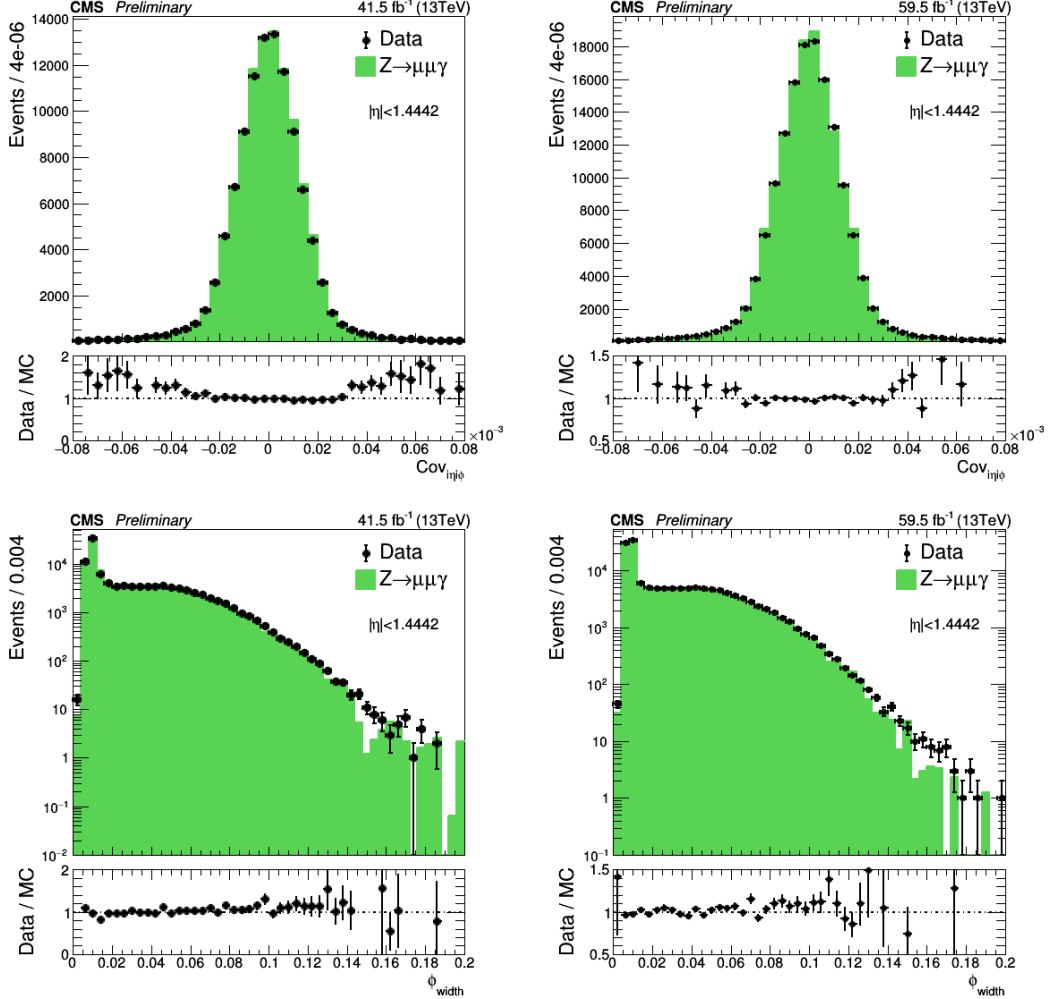


Figure 5-52 2017 distributions of the photon $Cov_{\eta\eta\phi}$ (left) and ϕ width (right) in ECAL barrel with 2017 (top) and for 2018 (bottom). Points are data and the green filled histogram is simulation.

The photon ID MVA score distributions are shown in Figure 5-53 and Figure 5-54 for $Z \rightarrow ee$ and $Z \rightarrow \mu\mu\gamma$ events. The data and simulated distributions agree well within the systematic uncertainty, corresponding to a shift in the value of the BDT score of ± 0.05 along with a linearly increasing term for the tail with low values of the photon ID MVA score in the EB, and a constant ± 0.06 in the EE. In 2016 and 2018, the systematic uncertainty in the MC photon ID BDT score in EB and EE are taken from the

standard $H \rightarrow \gamma\gamma$ with the systematic uncertainty doubled in EE for 2018 in $Z \rightarrow ee$ validations. For the validations with $Z \rightarrow \mu\mu\gamma$ events, the same consideration of the systematic uncertainty as the standard $H \rightarrow \gamma\gamma$ analysis is employed for 2016 and 2018. In 2017, the systematic uncertainty correspond to a shift of ± 0.04 (EB)/ ± 0.06 (EE) in the value of the MVA output along with a linearly increasing term (only EB case) at low values from -0.9 to 0.2.

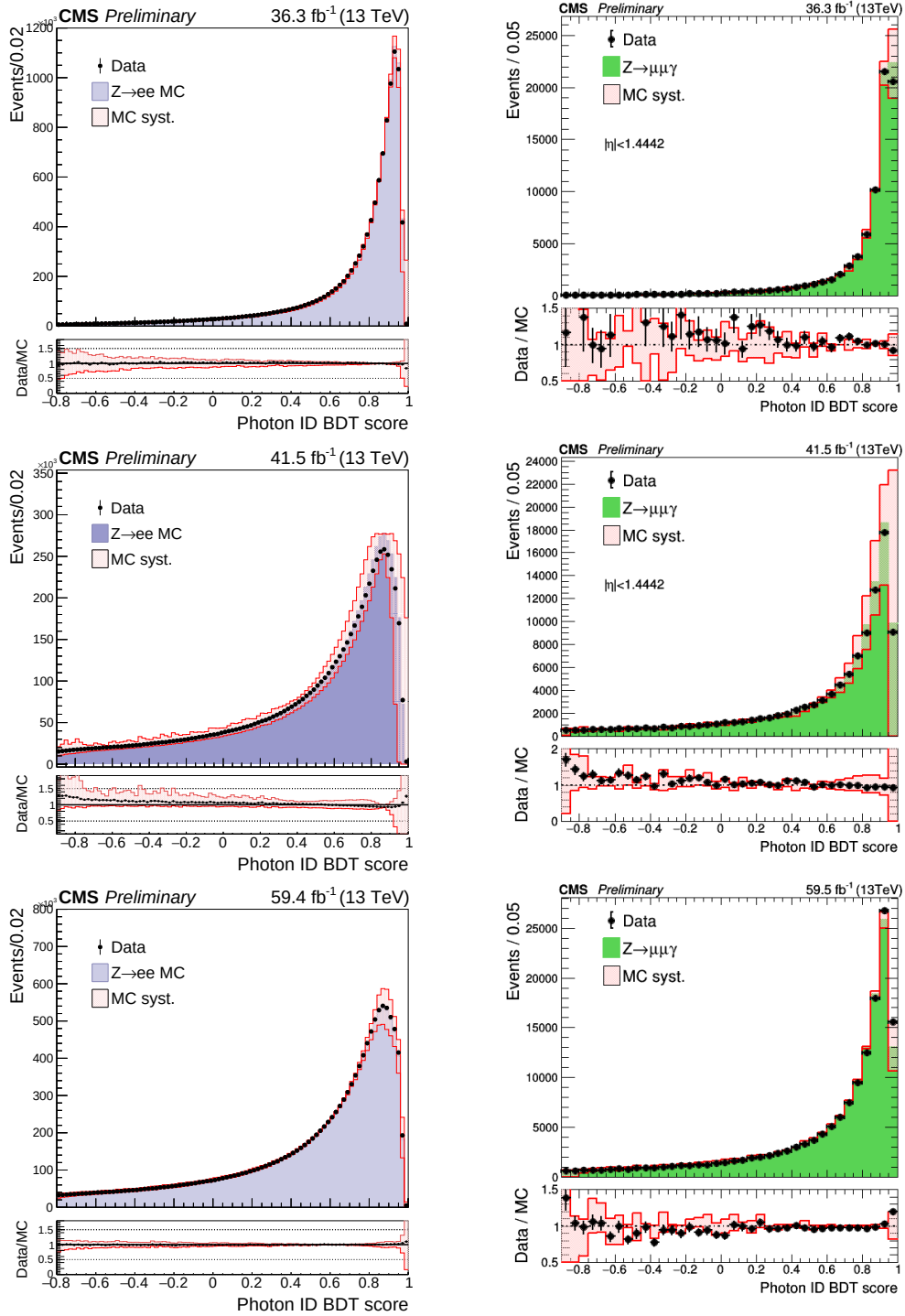


Figure 5-53 Photon ID MVA output distribution for $Z \rightarrow ee$ (left) and $Z \rightarrow \mu\mu\gamma$ (right) in ECAL barrel for 2016 (top), 2017 (middle) and for 2018 (bottom).

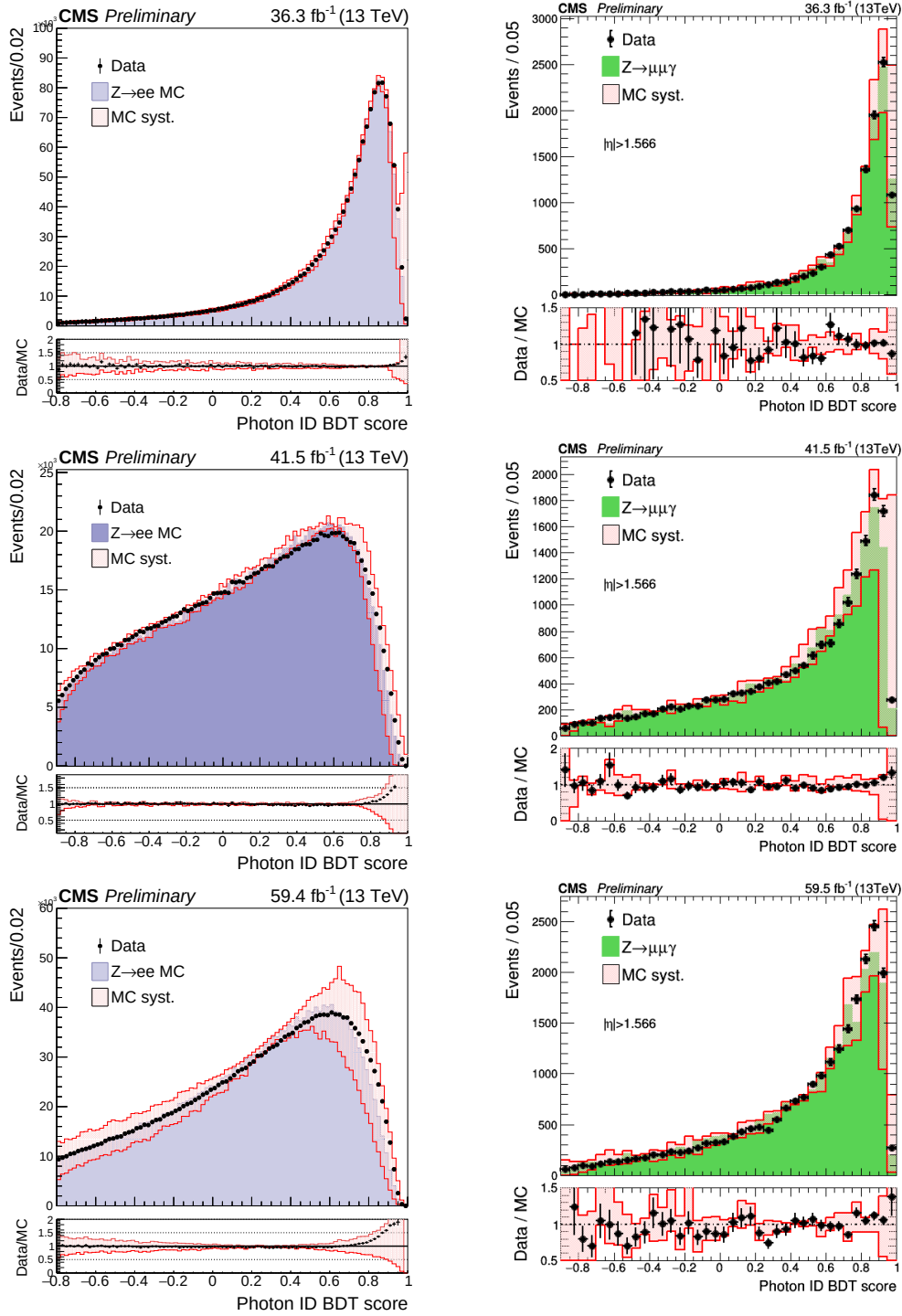


Figure 5-54 Photon ID MVA output distribution for $Z \rightarrow ee$ (left) and $Z \rightarrow \mu\mu\gamma$ (right) in ECAL endcaps for 2016 (top), 2017 (middle) and for 2018 (bottom).

The photon ID validation plot for the lower-scoring photon from diphoton pairs in data and background simulated samples is presented in Figure 5-55. The data and simulated backgrounds events are considered in the final fit range of invariant mass, $M_{\gamma\gamma} \in (65 - 120)$ GeV. The prompt-prompt component of the total background is scaled by a factor of 1.3. The total and individual components of the backgrounds are scaled to data by a global Scale Factor (SF) of 0.869676 for 2017 and 1.08751 for 2018. The MC background samples are used only for the training of BDT, the agreement is more or less consistent with data.

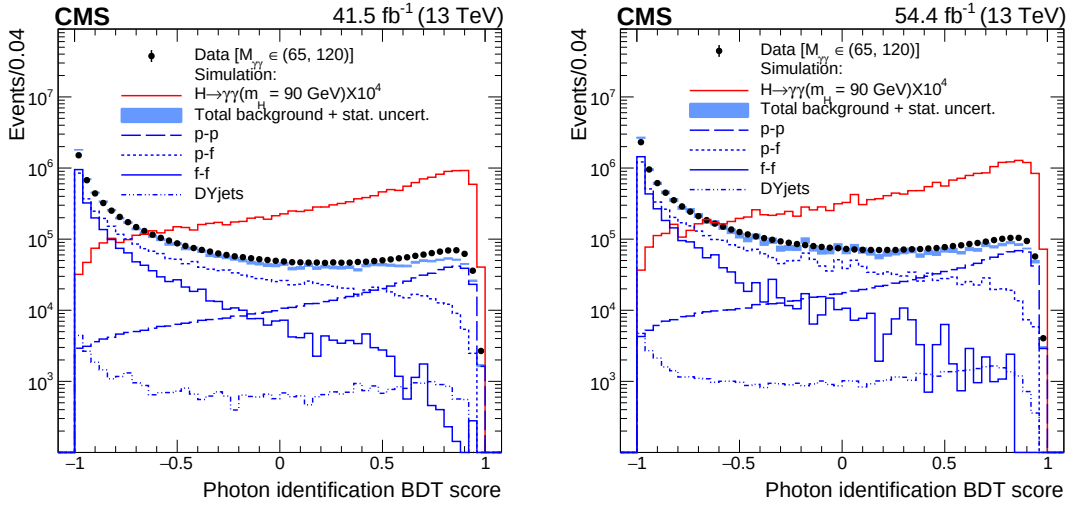


Figure 5-55 Photon identification BDT score of the lower-scoring photon of diphoton pairs in an event in 2017 (left) and 2018 (right).

5.10.5 Additional selections for Drell-Yan suppression

The relic DY events that have a mass peak not very far from observed excess in the previous published results with 2016 dataset [14], are suppressed by introducing two additional cuts as our definitive strategy. The investigation of significantly higher proportion of surviving DY events in 2018 compared to the other years, led us to realize that a stricter DY suppression technique than Pixel-Seed electron veto (PSV) was required. We have spent a considerable amount of time to develop the additional DY suppression strategy. These cuts are defined as follows

- so-called “linear cut”: $\log(\Sigma p_T^2) \leq 0.016 \times p_T^{\gamma\gamma} + 6.0$, where $\log(\Sigma p_T^2)$ is logarithm of the sum of squares of transverse momenta of all the tracks associated with

the chosen vertex as described in next subsection, and $p_T^{\gamma\gamma}$ is the transverse momentum of the selected diphoton candidate. The $\log(\Sigma p_T^2)$ distributions of the relic DY MC events and the signal events after the preselection from 90 GeV ggH production mechanism are shown in Figure 5-56 for 2016, 2017 and 2018 analysis. In 2D plots, white line shows the boundary of this linear cut Figure 5-56. It has the ability to reject almost $\sim 25\%/ \sim 45\%/ \sim 56\%$ of relic DY events as well as keeping $\sim 97\%$ of boosted signal events in 90 GeV ggH sample in for all of the three years as can be seen in Figure 5-57 where red circles shows when events are categorized into 3 untagged event classes.

- “ $N_e^{Matched} = 0$ ”: $N_e^{Matched}$ is the number of electrons that are reconstructed and matching to each of the photon from the diphoton candidate. The p_T of electron should be greater than 9 GeV as required in the μAOD production and used in all of the $H \rightarrow \gamma\gamma$ studies with full Run 2 data. For an electron matches to a photon should satisfy at least one out of the following 4 quantities: $\Delta R(\gamma, electron)$, $\Delta R(\gamma - supercluster, electron)$, $\Delta R(\gamma, electron - supercluster)$ and $\Delta R(\gamma - supercluster, electron - supercluster)$, has to be less than 0.2. As presented in Figure 5-57, the blue circles with this cut, can also efficiently suppress the relic DY background while at the same time, keep the signal efficiency of about 95%. On the top of linear cut, “ $N_e^{Matched} = 0$ ” requirement can aid to mimic an additional $\sim 63\%/ \sim 70\%/ \sim 65\%$ of relic DY events in 2016/2017/2018 and keeping almost 95% efficiency signal events.

As a result, by combining these two requirements, we are able to reject $\sim 55\%/ \sim 60\%/ \sim 70\%$ of relic DY events and keep about $\sim 92\%$ efficiency of the 90 GeV ggH signal events in 2016, 2017 and 2018 analysis, as presented as black points in the Figure 5-57. In the most sensitive class 0, the events with higher diphoton BDT scores, these additional cut help to reject $\sim 63\%/ \sim 70\%/ \sim 65\%$ of relic DY background as well as keeping almost $\sim 90\%$ efficiency of the 90 GeV ggH signal events in 2016, 2017 and 2018 analysis.

The $Z \rightarrow ee$ events are used to measure the efficiencies of the linear cut are for both data and Drell-Yan MC. In order to eliminate the effect because of two electron tracks in original $\log(\Sigma p_T^2)$ with larger values, a modified $\log(\Sigma p_T^2)$ with matched tracks to each of two photon candidates removed, is used to calculate the linear cut efficiencies. The tracks matching criteria to each of leading and subleading photon candidates, is that ΔR between any tracks and each photon candidate or its supercluster should be less

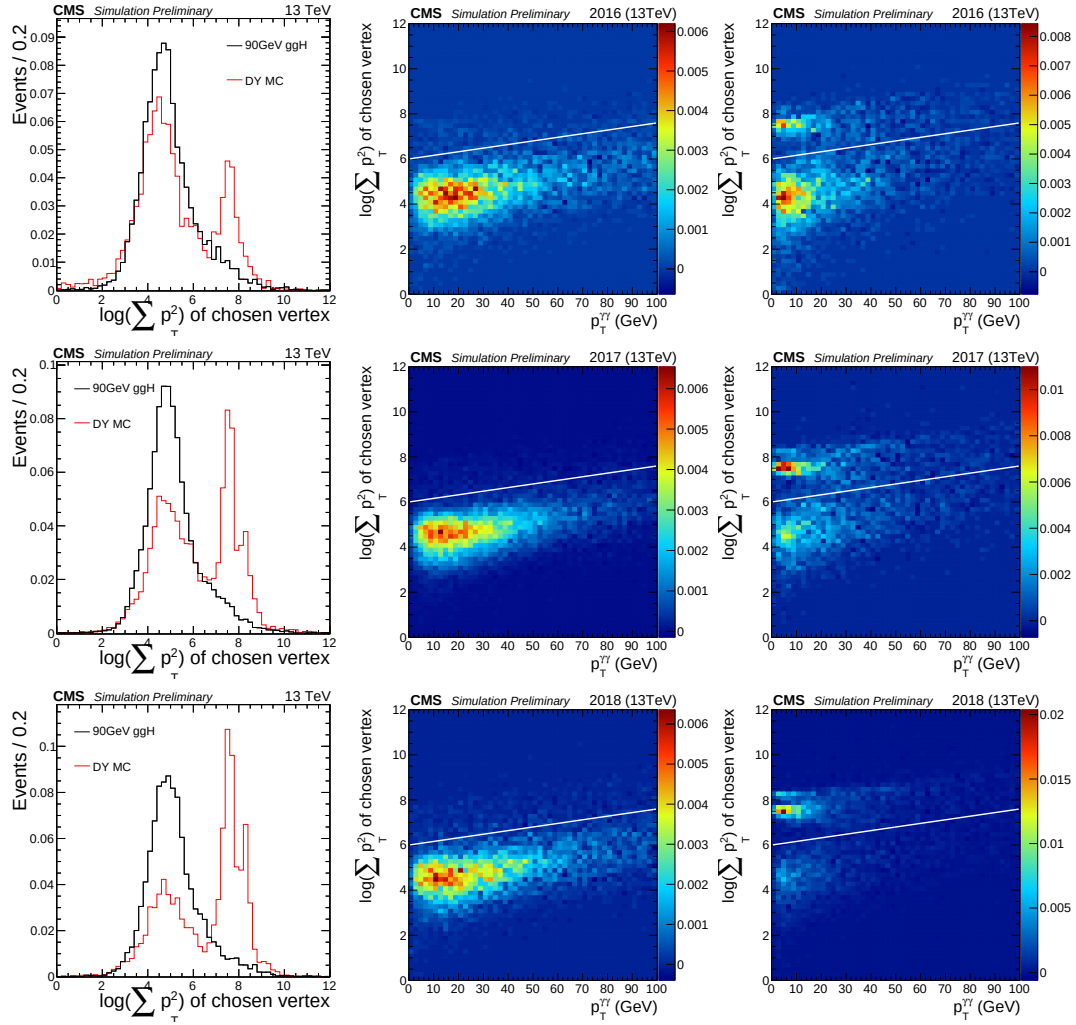


Figure 5-56 Comparisons of the $\log(\sum p_T^2)$ distributions of the relic DY MC events and the signal events from 90 GeV ggH production mode (left), $\log(\sum p_T^2)$ versus $p_T^{\gamma\gamma}$ for the signal events from 90 GeV ggH production mode (middle) and the relic DY MC events (right). The boundary of the linear cut, $\log(\sum p_T^2) \leq 0.016 \times p_T^{\gamma\gamma} + 6.0$, is shown as the white line in each 2D plot. Top plots are for 2016, middle plots are for 2017 and the bottom ones are for 2018.

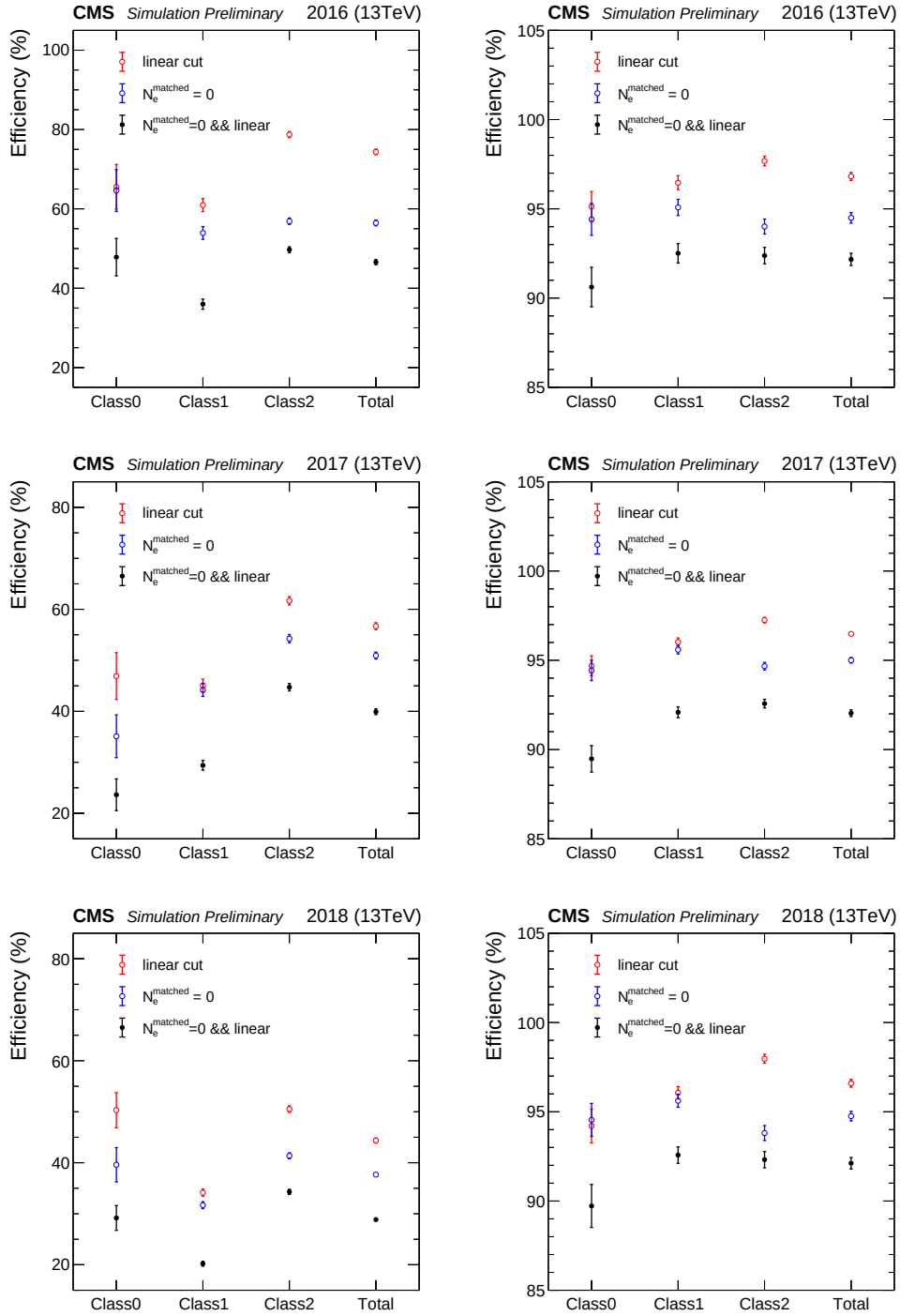


Figure 5-57 Selection efficiencies of the linear cut, “ $N_e^{\text{Matched}} = 0$ ” and the combination of these two requirements on the relic DY events (left) and 90 GeV ggH signal events (right). Top plots are for 2016, middle plots for 2017 and the bottom ones for 2018.

than 0.05. The ratio between the number of $Z \rightarrow ee$ events after the linear cut and the total number of $Z \rightarrow ee$ events after the preselection is referred as efficiency. For the linear cut, efficiencies and scale factors can be seen in Figure 5-58 for 2016, 2017 and 2018. The quadratic sum of the statistical uncertainty (from the nominal fitting) and the systematic uncertainty is taken as total uncertainty.

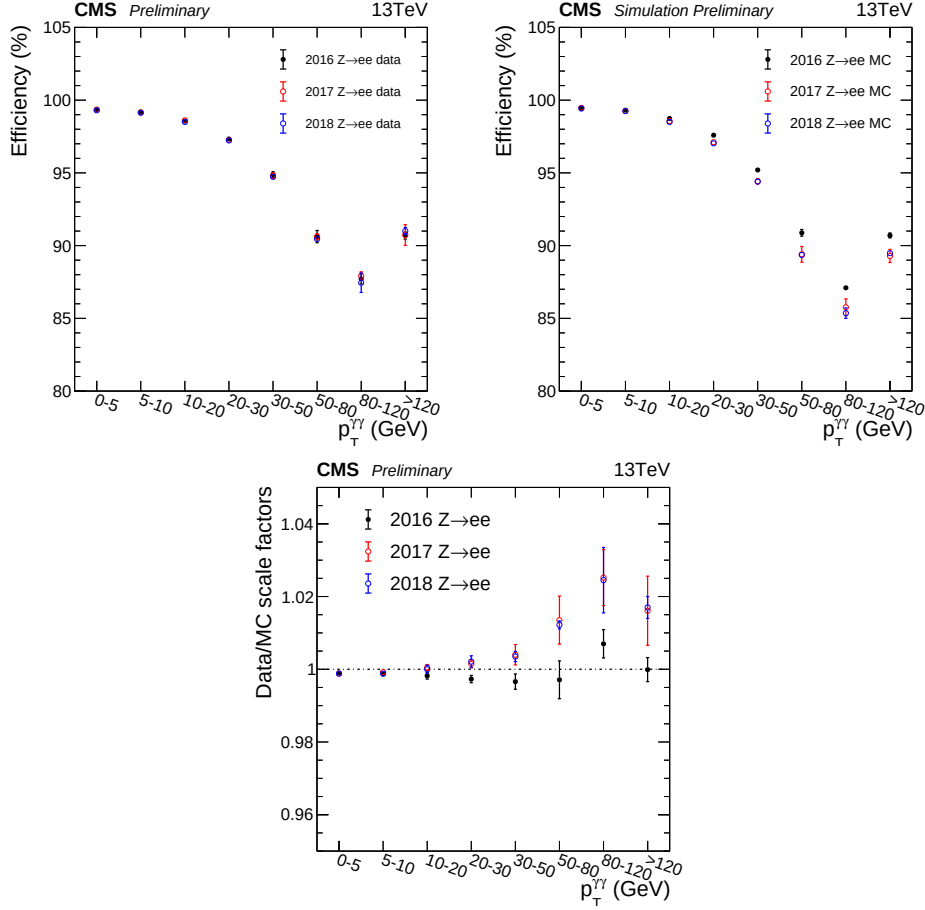


Figure 5-58 Efficiencies measured using $Z \rightarrow ee$ events in data (top left) and MC (top right), and the scale factors (bottom) of the linear cut with 2016, 2017 and 2018 datasets.

The efficiencies of “ $N_e^{Matched} = 0$ ” cut are computed using the photons of $Z \rightarrow \mu\mu\gamma$ events in both Drell-Yan and MC. The e-veto requirement is used for the estimation of efficiencies and scale factors of “ $N_e^{Matched} = 0$ ”. The efficiency is measured as the ratio of photons surviving the “ $N_e^{Matched} = 0$ ” cut divided by the total number of photons. Table 5-22, Table 5-23 and Table 5-24 shows the efficiency results of “ $N_e^{Matched} = 0$ ” cut in data and MC and their ratios, in 2016, 2017 and 2018 datasets.

Table 5-22 (2016) Efficiency of the “ $N_e^{Matched} = 0$ ” requirement, measured in the three photon categories using tag and probe with 2016 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.

Photon categories	Data		Simulation		Scale Factor	
	Eff.	Stat.	Eff.	Stat.	SF	Unc.
Barrel; $R_9 > 0.85$	0.9946	0.0003	0.9947	0.0006	1.0000	0.0007
Barrel; $R_9 < 0.85$	0.9425	0.0018	0.9255	0.0042	1.0184	0.0050
Endcap; $R_9 > 0.90$	0.9618	0.0019	0.9317	0.0045	1.0323	0.0054

Table 5-23 (2017) Efficiency of the “ $N_e^{Matched} = 0$ ” requirement, measured in the four photon categories using tag and probe with 2017 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.

Photon categories	Data		Simulation		Scale Factor	
	Eff.	Stat.	Eff.	Stat.	SF	Unc.
Barrel; $R_9 > 0.85$	0.9928	0.0003	0.9936	0.0005	0.9993	0.0006
Barrel; $R_9 < 0.85$	0.9169	0.0018	0.9232	0.0030	0.9932	0.0038
Endcap; $R_9 > 0.90$	0.9590	0.0017	0.9638	0.0026	0.9951	0.0032
Endcap; $R_9 < 0.90$	0.8035	0.0091	0.8009	0.0140	1.0033	0.0209

Table 5-24 (2018) Efficiency of the “ $N_e^{Matched} = 0$ ” requirement, measured in the three photon categories using tag and probe with 2018 $Z \rightarrow \mu\mu\gamma$ events. The data to simulation ratio is also shown with its uncertainty.

Photon categories	Data		Simulation		Scale Factor	
	Eff.	Stat.	Eff.	Stat.	SF	Unc.
Barrel; $R_9 > 0.85$	0.9928	0.0003	0.9932	0.0005	0.9996	0.0006
Barrel; $R_9 < 0.85$	0.9232	0.0015	0.9253	0.0032	0.9978	0.0039
Endcap; $R_9 > 0.90$	0.9628	0.0013	0.9606	0.0028	1.0023	0.0032

5.10.6 Diphoton vertex identification

For the diphoton vertex identification, the same as described in [105] is used. A BDT classifier is employed to obtain a diphoton vertex among the set of all primary

vertices. The sum of the square of transverse momenta of the charged particle connected with the vertex, two variables which define the vector and scalar balance of p_T of the diphoton system and the charged particle tracks linked with the vertex. Moreover, if the photon is associated with any tracks of charged particle which have been identified as resulting from conversion process, the pull between the longitudinal positions of the primary vertex of the conversion tracks alone as well as from all the associated tracks has been added to the BDT input variable set. The detailed description along with the performance of the vertex ID for all of the years can be accessed in [105].

5.10.7 Event classification

This analysis utilize the events with diphoton pairs and each photon is required that the ratio of its transverse momentum, p_T , to invariant mass of the diphoton system, $m_{\gamma\gamma}$, be $p_T^{\gamma_1}/m_{\gamma\gamma} > 0.47$ ($\approx 30.55/65.$) and $p_T^{\gamma_2}/m_{\gamma\gamma} > 0.28$ ($\approx 18.20/65.$), where $\gamma_{1(2)}$ are the photon candidates with the highest (next-to-highest) value of p_T . The use of p_T thresholds scaled by $m_{\gamma\gamma}$ stop the distortion of the low end of diphoton mass spectrum which results if we use a fixed threshold. Preselection requirements has to be satisfied by each of the photon candidate as well as a minimum requirement on photon ID BDT value (>-0.9). The diphoton system finally have the invariant mass lie within the range $65 < m_{\gamma\gamma} < 120$ GeV.

The analysis sensitivity has been enhanced by categorizing of events using mass resolution, signal-to-background ratio and production modes. The Higgs production mechanisms are identified by selecting the final objects in addition to the diphoton pair except the ggH mode. In the VBF production mode of Higgs boson is accompanied by a pair of jets separated by a large rapidity gap. Dijet events tagging for VBF class has significantly increased the overall sensitivity of the analysis. The dijet objects tagged events considered as exclusive categories and those which are remain untagged are the inclusive categories determined by the ggH mechanism. The inclusive categories are splitted to combine events with identical mass resolution.

The studies have been done with 2017 and 2018 data, so one additional VBF tagged event class was finally introduced in 2017 and 2018 analyses. In 2016 analysis, the events are categorized in to different event classes based on the retrained diphoton BDT. The rest of events in 2017 and 2018 as well as all of the events in 2016 analysis are divided in to classes using the diphoton BDT score. The boundaries are also optimized

for 2016 and 2017. The total number of these event classes and their boundaries are determined, such as to maximize the expected signal significance. The same boundaries are employed for both the 2017 and 2018 analyses. Three untagged event classes are used in 2016, 2017 and 2018 and referred as 0, 1, 2 where class 0 contains the events with greatest expected sensitivity. If there are more than one diphoton candidate within the analysis mass window is present then the candidate with the highest priority tag is selected. In case of multiple diphoton pairs with same tag priority, the diphoton with highest p_T photons is then considered. For the studies presented in this chapter, 3 event classes are considered in 2016 analysis and 4 event classes in 2017 and 2018 analyses.

5.10.8 Diphoton BDT classifier

The sensitivity of analysis is increased by splitting the events in to categories, characterized by a high signal-to-background ratio and better mass resolution. The studies presented here has new updates w.r.t. to the previous published low-mass analysis [14]. The categorization of events is performed based on the output of event classifier which aim to assign events having high scores with two photons such that they fulfill the following criteria:

- signal-like kinematic characteristics
- good diphoton mass resolution
- scores from the photon identification BDT in the ‘photon-like’ region

The MVA classifier does not select the events on the basis of invariant mass of the diphoton pairs such as not be biased by the related mass used at the time of generation of signal component of the training sample. The diphoton events are selected after they fulfill the preselection criteria described in Section 5.10.3 and in addition a loose selection on the ID MVA output (> -0.9) which is 99% efficient for signal events passing the preselection requirements.

The TMVA package is employed for the training of diphoton BDT classifier with the hyperparameters NTrees=2000, MaxDepth=3 and MinNodeSize=0.125 are used. This is done specifically on the low mass signal samples for all of the four production modes (ggH, VBF, VH, and ttH) and $m_H = 90$ GeV. The selection of background events is made from the simulated SM background samples, which includes the prompt (from diphoton + jets sample), prompt-fake (from photon + jet and QCD di-jet sam-

ples) and fake-fake (from QCD di-jet samples) contributions. Due to the presence of high weights in the QCD dijet sample because of not enough simulated events, the events are down-weighted by a factor of 25 to prevent over-training of the input variables. The variables used as input to the classifier are the transverse momenta for both photons, rescaled for the diphoton mass ($p_T^{1,(2)}/m_{\gamma\gamma}$), pseudorapidities of both photons ($\eta^{1,(2)}$), cosine of the angle between the two photons in the transverse plane ($\cos(\Delta\phi)$), identification BDT score for both photons, per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using the correct primary vertex (σ_{rv}), per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using an incorrect primary vertex (σ_{wv}) and per-event probability estimate that the correct primary vertex has been used to reconstruct the mass, based on the event-level vertex selection MVA (p_{vtx}). The shapes of the BDT input variables for 2016 legacy are shown in Figure 5-59. The BDT output variable shapes are shown in Figure 5-60 for 2016 legacy and 2017. The events with the higher BDT score bins shows the properties of signal events (signal kinematics with boosted p_T , more central photon pairs, good mass resolution, “photon-like” values from the photon ID BDT). The discriminating power between signal and background events of the BDT classifier in the form of ROC (Receiver Operating Characteristic) curves are shown in Figure 5-61 for 2016 legacy and 2017. The ROC curves are extracted by using the test sample such as to prevent bias in the results.

As it is already mentioned that 2017 analysis diphoton BDT training is used for 2018 data analysis. Apart from the validations of 2018 analysis diphoton BDT score using $Z \rightarrow ee$ and diphoton data/simulation comparison, we had also checked performance (“ROC”) of 2018 analysis diphoton BDT using 2017 analysis weight file. The “ROC” comparison along with the normalized distributions for signal and background of 2017 and 2018 can be seen in Figure 5-62. The “AUC (Area Under Curve)” value for 2018 analysis with 2017 training, is a little lower (around 2%) than the 2017 one but this difference is acceptable.

The validation is done using electrons reconstructed as photons in $Z \rightarrow ee$ events in data and Drell-Yan simulation with the electron veto inverted applied as a part of pre-selection. The comparison of the signal model shape for $Z \rightarrow ee$ in data and simulation represents a crucial check that the BDT input variables modeling and their correlations

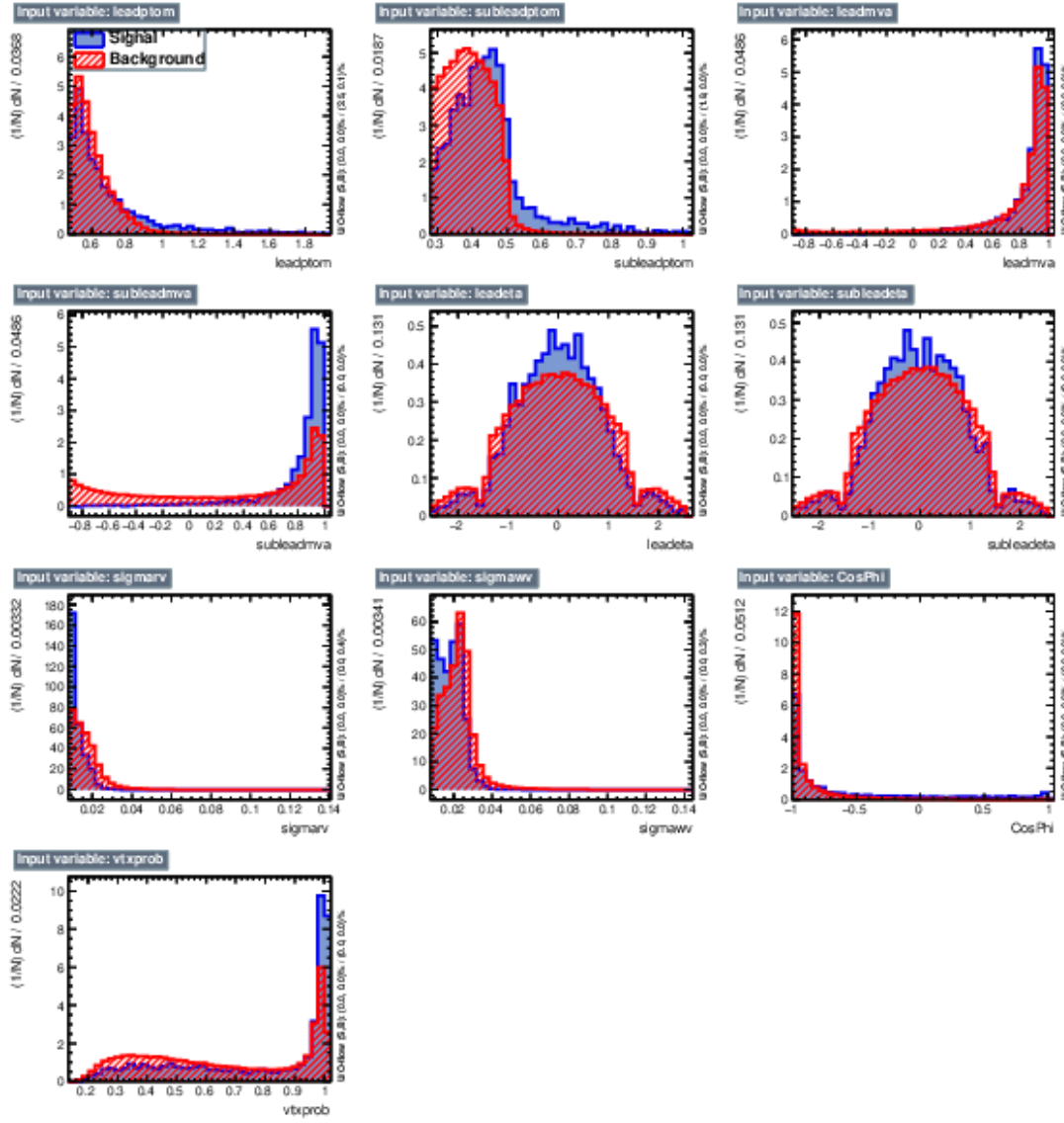


Figure 5-59 Comparisons on the input variables of 2016 legacy diphoton BDT between the signal and background. The blue distribution corresponds to the signal whereas the red one corresponds to the background.

in the simulation are accurate sufficiently. The single electron data samples and the trigger (single electron HLT) path used for 2016 legacy is HLT_Ele27_WPTight_Gsf. For 2017 samples, the trigger path applied to both data and simulation is HLT_Diphoton30PV_18PV_R9Id_AND_IsoCaloId_AND_HE_R9Id_NoPixelVeto_Mass55, prescaled by a factor of 18. For 2018 samples, the default low-mass analysis path is used which is HLT_Diphoton30_18_R9IdL_AND_HE_AND_IsoCaloId_NoPixelVeto_Mass55. Apart from different trigger paths, the construction of signal model was done in similar manner as that for $H \rightarrow \gamma\gamma$ by using the same energy scale and smearing corrections applied in the analysis. The events with mass range between 80 and 100 GeV are used for this

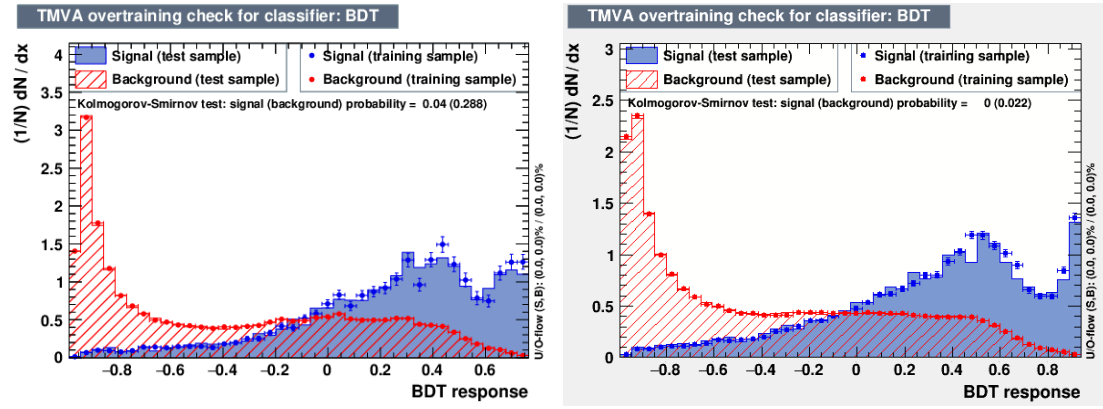


Figure 5-60 2016 (left) and 2017 (right) diphoton BDT outputs of training and testing events in the signal and background samples.

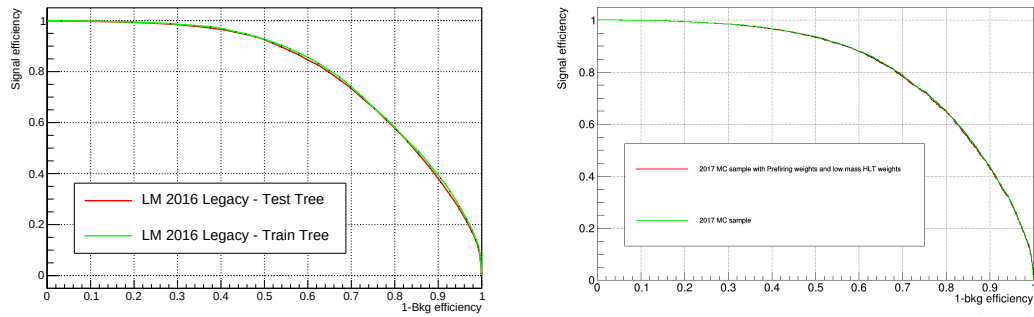


Figure 5-61 2016 (left) 2017 (right) diphoton BDT classifier ROC curve.

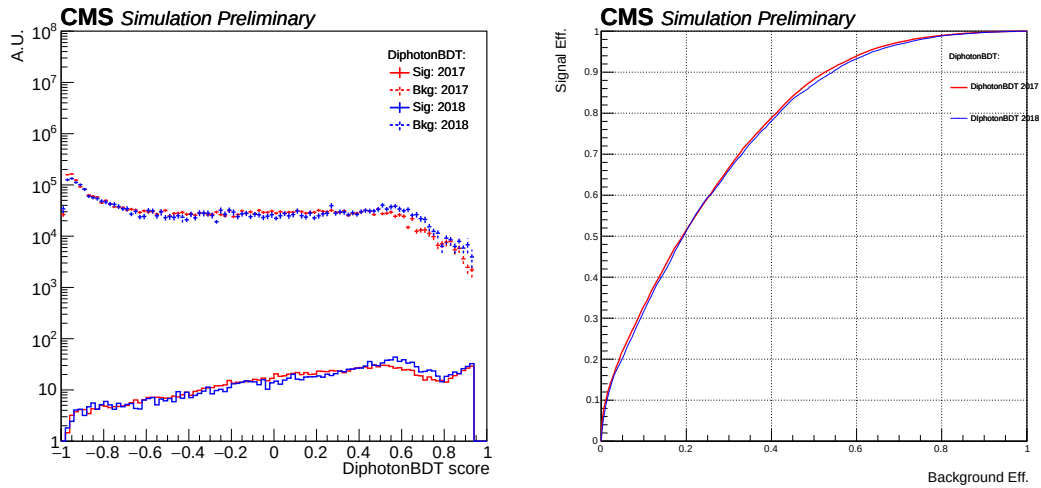


Figure 5-62 Comparison between 2018 analysis diphoton BDT using 2017 analysis weight file and 2017 analysis diphoton BDT for the signal and background distribution with the same normalization across years (left) and for ROC (right).

validation. Two key systematic uncertainties affecting the diphoton BDT and its inputs are included i.e. shift in the photon ID MVA score as well as the uncertainty on the photon energy resolution σ_E/E . A simple re-weighting of the MC sample is applied in $p_{T,Z}$ bin so that simulation sample agrees well with data. This re-weighting also enhance the data/MC agreement for the distributions of diphoton BDT correlated with $p_{T,Z}$ like the vertex probability, the p_T/m and cosine of the angle between two electrons. Some of the comparisons between data and simulation are shown in Figure 5-63 and Figure 5-64. The data/MC agreements are within the uncertainties. It is observed that the diphoton BDT shapes in $Z \rightarrow ee$ events in 2017 and 2018 are a bit different. The reasons is the different mass resolution in 2017 (worse) and 2018 which is inducing the different output shape for 2017 and 2018 as the mass resolutions w.r.t. right vertex and wrong vertex are two of the input variables for the training. The 2017 and 2018 selections are different due to different requirements in HLT paths of 2017 and 2018 is also an additional reason.

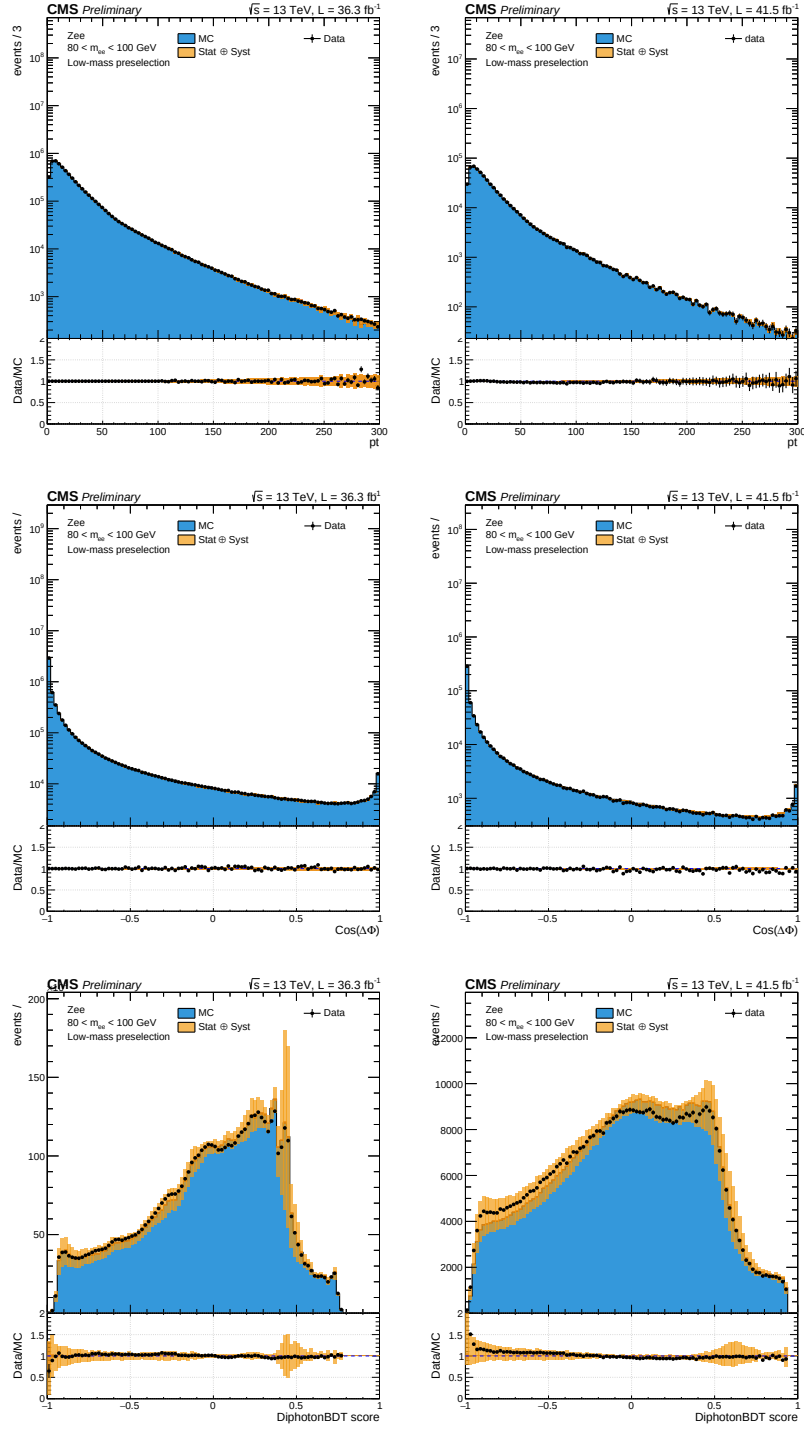


Figure 5-63 Comparisons of data and simulation for the p_T distribution (top), $\cos(\Delta\phi)$ (middle) and diphoton BDT output score distribution (bottom) on electrons reconstructed as photons from Z boson decay for 2016 (left) and 2017 (right).

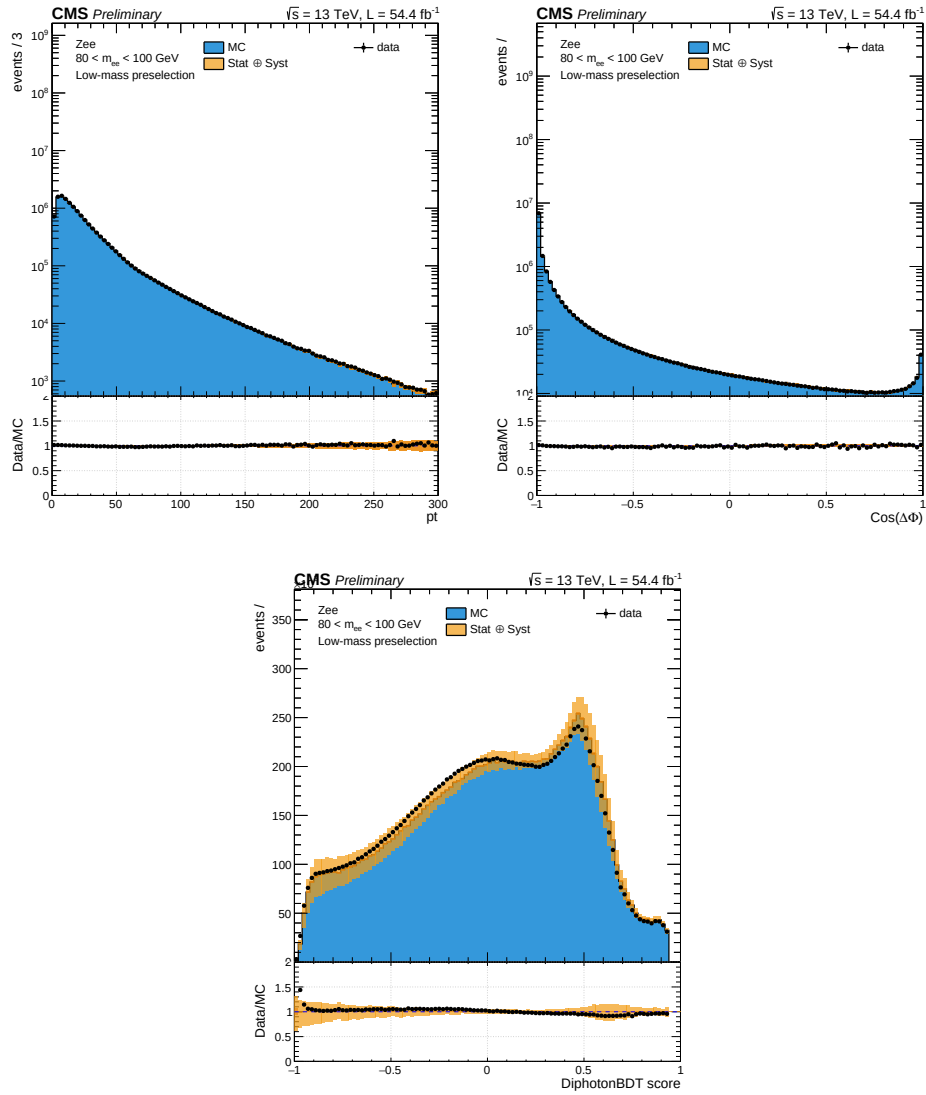


Figure 5-64 Comparisons of data and simulation for the p_T distribution (1a), $\cos(\Delta\phi)$ (b) and diphoton BDT output score distribution (c) on electrons reconstructed as photons from Z boson decay for 2018.

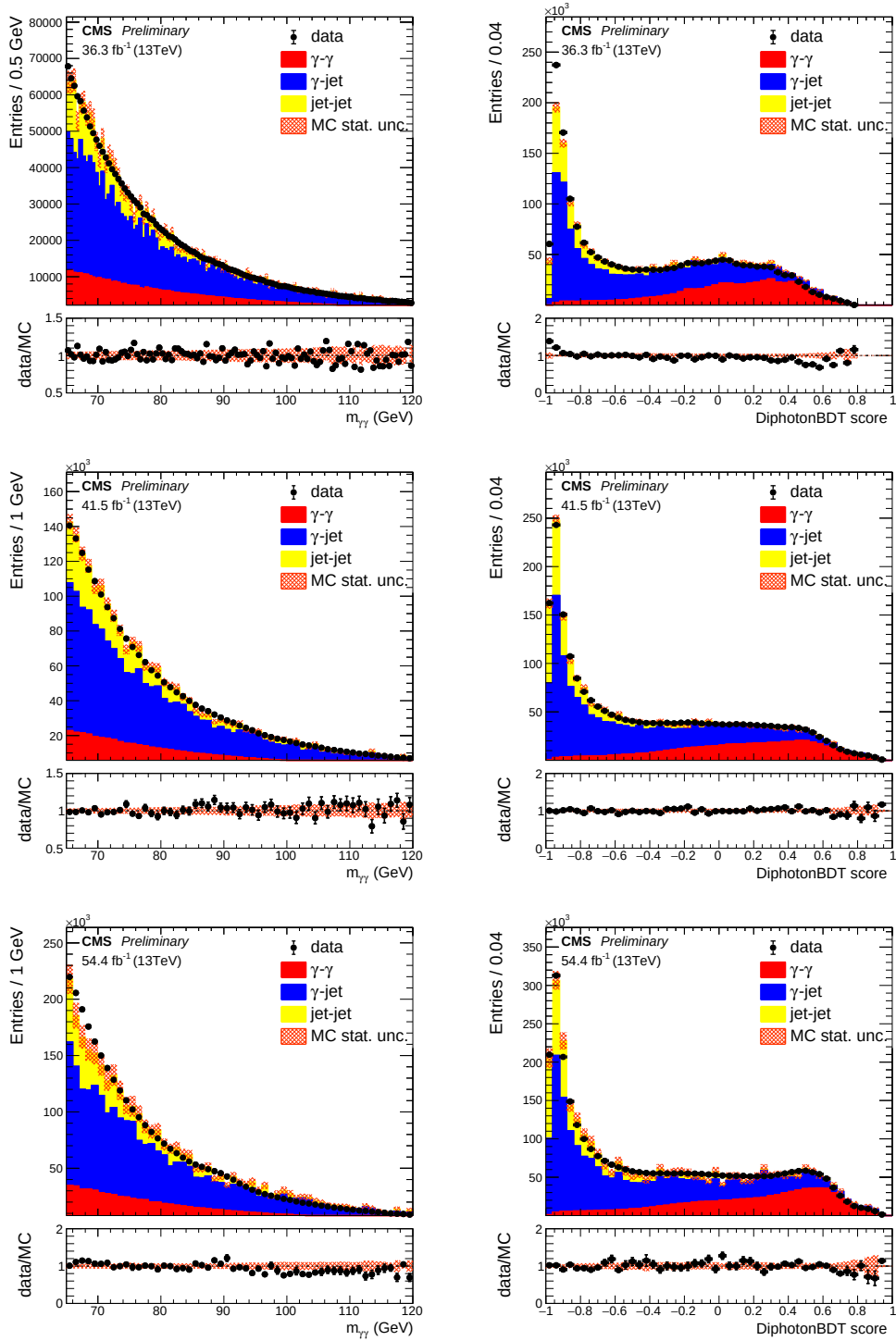


Figure 5-65 Distributions of diphoton invariant mass and diphoton BDT scores for 2016 (top), 2017 (middle) and 2018 (bottom) data sets.

The validation has also been done with selected events in data and MC for the inputs and output of diphoton BDT training. In 2018 analysis, the MC sample “/QCD_Pt-30to40_DoubleEMEnriched_MGG-80toInf_TuneCP5_13TeV_Pythia8” is not available so its contribution in the MC distributions is missing. The background MC is normalized according to the number of events in each 2016, 2017 and 2018 data. In order to validate the inputs and outputs of diphoton BDT training, the events in data are selected from the whole analysis mass range from 65 to 120 GeV. The distribution of some of the input variables of the diphoton BDT training are shown in Figure 5-65. The agreements are observed reasonable for each of the three years.

5.10.9 Boundary optimization

The diphoton BDT (as discussed in the previous section) output score is used to split the analysis in to different classes. The number of classes and position of boundaries are optimized. The categorization of events was done based on their BDT output values. The classes boundaries in the BDT output spectrum are chosen by an optimization method. The ‘n’ boundaries are placed initially at equal distances from each other in signal quantiles such that two fixed boundaries at the edges of the spectrum ± 1 . Using ‘n’ boundaries, $n + 1$ categories are determined and then, only ‘n’ are selected for the studies in order to calculate the expected significance. The events within the lowest BDT score category are not considered. The boundary positions are allowed to float freely and are adjusted by on the basis of combined p-value. The method is same as used in the standard $H \rightarrow \gamma\gamma$ analysis [105]. In each of the category, the simulation is used for the extraction of signal and background models. The signal model shape is extracted by using the sum of two Gaussian functions, one fitting the signal component where the mass resolution is entirely dominated by photon energy resolution (i.e. the distance ‘dz’ between the selected vertex for reconstruction of mass and correct vertex is less than 1 cm along the z-coordinate) and the other Gaussian function is fitting the worse resolution component of signal given by the non-negligible term related to the vertex displacement ($dz > 1$ cm). The p-value (expected significance defined as figure of merit in Figure 5-66) has been extracted from a fit to Asimov toy dataset thrown from the signal plus background model in each category. As the p-value is computed several times during the optimization, so no systematic uncertainty on the signal is included in the procedure and the simplified models for signal and background described

above are considered. The simulated background samples are used to extract the background model through an exponential shape fit to the $m_{\gamma\gamma}$ invariant mass distribution. In order to make the background modeling simple, the DY MC sample is not included in the background samples for the optimization of boundaries. But we make sure that the optimized boundaries have enough DY MC events for the modelling of relic DY background in each of the corresponding event class.

A per class fit will be performed on MC relic $Z \rightarrow ee$ events where both electrons are reconstructed as photons. Therefore, it needs to make sure that enough DY MC events should be present in each class to obtain acceptable fits i.e. the central values of fit parameter has to be greater than their statistical uncertainties as well as the sigma value is not too far away from the real $Z \rightarrow ee$ value. It is necessary to define a minimal width for the size of the classes in terms of diphoton BDT score. The class "0" contains the events which are more signal-like having high BDT diphoton scores. It is the class containing fewest events and is sufficient to find the minimal width of optimal size for this class only. In order to that, relic DY event fits have been performed for various class 0 widths and the optimized value is therefore defined as the minimal width value for which, the parameters are well defined w.r.t. their statistical uncertainties as well as the sigma parameter value is within 1 standard deviation of the true Z sigma.

The width value is then placed as a minimal width value for class 0. By minimizing the p-value for number of classes, a figure of merit as shown in Figure 5-66 for 2016 and 2017 obtained as a function of n. There is no significant change in the figure of merit while going from n=3 to n=4 or n=5. For this reason, it is reasonable to chose n=3 and to proceed with 3 classes having boundaries [1.000, 0.593, 0.246, -0.538] for 2016 and with boundaries [1.000, 0.753, 0.334, -0.364] for 2017 in term of diphoton BDT output value as shown in Figure 5-67 for 2016 and 2017.

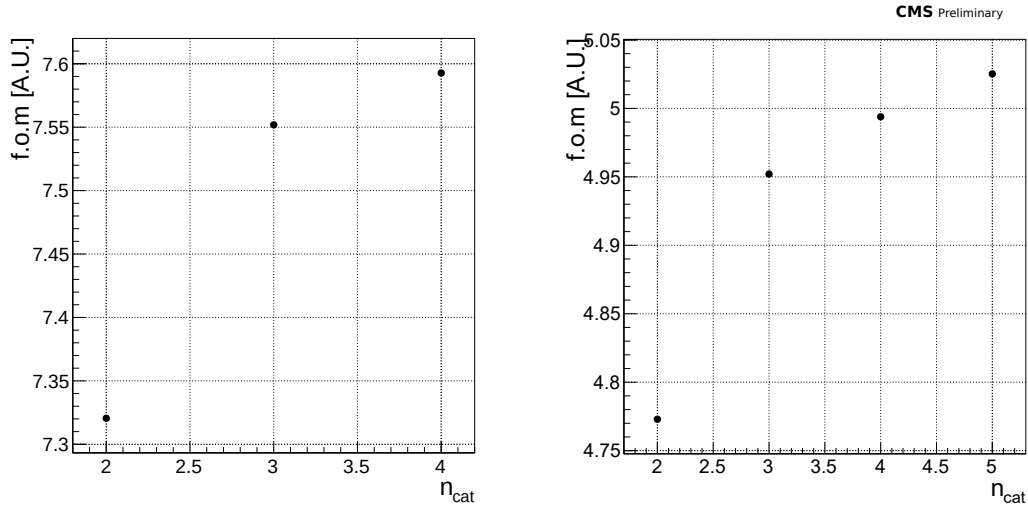


Figure 5-66 Figure of merit based on the p-value as a function of the number of classes n_{cat} for 2016 (left) and 2017 (right).

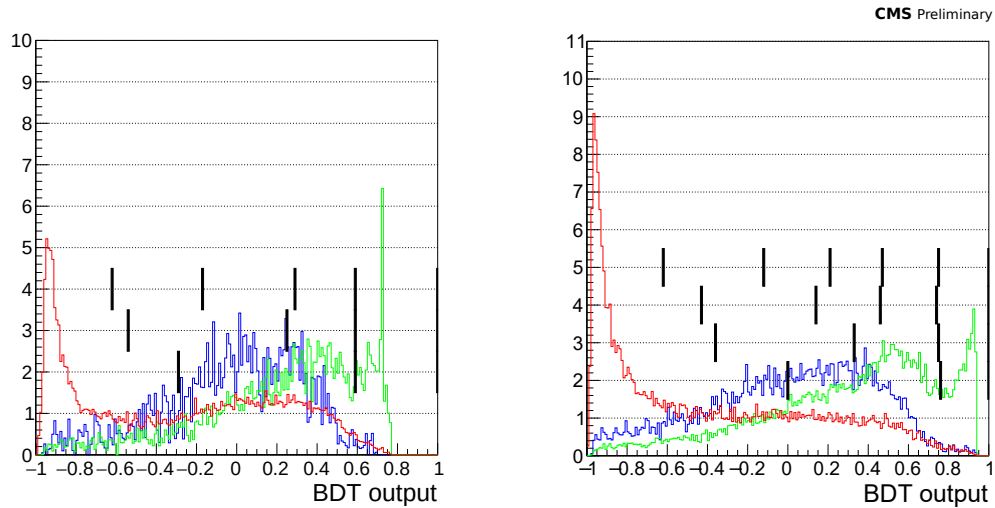


Figure 5-67 Events distribution based on BDT output value. Background MC events (red), right vertex (green) and wrong vertex (blue) signal MC events. Black lines shows the class boundaries. The plots are shown for 2016 (left) and 2017 (right).

5.11 VBF selection strategy

As it is mentioned above that in 2017 and 2018 analyses we introduced VBF class with two jets along with the photon pair in the final state. This is done in order to select events originating via vector boson fusion (VBF) production mechanism. In the studies presented in this section, only the two highest in p_T jets are considered. For this VBF production mode, the SM Higgs cross section is about ten times smaller than that for gluon-gluon fusion. By using multivariate classifier, a category of events can be defined with excellent signal-to-background ratio. The expected signal-to-background ratio of the events selected in VBF class is highest among all the event categories.

This section is dedicated to the dijet and combined MVA training, based on the 2017 data which is also used for 2018 analysis. The strategy of training, the input variables and the training package is inherited from standard $H \rightarrow \gamma\gamma$ analysis [105]. The results of the training with 2017 MC samples are then validated with both 2017 and 2018 $Z \rightarrow ee$ plus 2jets events and also with the diphoton events selected in both data and MC for 2017 and 2018. As it is explained that the $H \rightarrow \gamma\gamma$ events contain two high- E_T isolated photons. The event selections for photons is same as used in the inclusive analysis and described in Sections 5.10.3.

VBF events possess a distinctive topology having two forward jets from the scattered quarks that may be measured in the detector. In order to tag them as VBF like events, they must have at least two jets along with two photons. Based on the combined MVA output score, we add one VBF tagged event class in each of the 2017 and 2018 analyses.

5.11.1 Jet definition

In the VBF tagged events, jets are reconstructed by using the Jet/MET dedicated recipe i.e. the anti- k_T algorithm with radius parameter 0.4 has been applied to Particle Flow (PF) candidates. The charged candidates are excluded which are not associated with the selected vertex of the event i.e. defined by Particle Flow with Charged Hadron Subtraction (PFCHS). Several corrections are applied to the simulated events and data which are described in section 4.2.3.

An algorithm known as Pileup-Jet Identification (PUJID) is used which returns a PUJID jet score to define three kinds of working points i.e. loose, medium and tight.

As we are following the same strategy in [105], so the tight working point is chosen in the studies. It is eventually helpful in order to improve the data/simulation agreement as explained in Ref. [105].

5.11.2 VBF preselection

The contributions of the background processes has been reduced by requiring the the events to pass a selection criteria for photons and the jets. Particularly jet requirements are employed in order to distinguish real VBF signal jets from the ones originating from gluon fusion, the standard model background processes and fake jets (from pileup interactions). The requirements include a high dijet invariant mass m_{jj} , which is a characteristic of VBF event topology. A set of selections referred as VBF preselection requirements are listed below and the events passing these preselections, are possibly considered as VBF candidates.

- Two photons passing the Diphoton Preselection described in Section 5.10.3
- Both photons have a photon ID MVA score > -0.9
- Tight requirements on the photon p_T : $p_T^{\gamma_1}/m_{\gamma\gamma} > 0.47$ and $p_T^{\gamma_2}/m_{\gamma\gamma} > 0.28$
- Two jets passing the PUJID and satisfying the requirements $p_T^{j_1} > 30 \text{ GeV}$, $p_T^{j_2} > 20 \text{ GeV}$, $\eta_{j_1, j_2} < 4.7$, where j_1 (j_2) refers to the jet with the highest (next-highest) p_T value
- $m_{j_1 j_2} > 100 \text{ GeV}$

5.11.3 Dijet MVA and combined MVA categorization

The Dijet MVA used in the study takes the following variables as input from events after passing the VBF preselection:

- The transverse momenta of the leading and subleading photons divided by the invariant mass of the diphoton candidate: $p_T^{\gamma_1}/m_{\gamma\gamma}$ and $p_T^{\gamma_2}/m_{\gamma\gamma}$
 - the transverse momenta of the leading and subleading jets: $p_T^{j_1}$ and $p_T^{j_2}$
 - the dijet invariant mass, $m_{j_1 j_2}$
 - the difference in pseudo-rapidity between the two jets, $\Delta\eta_{j_1 j_2}$
 - the difference in azimuthal angle between the dijet and the diphoton, $\Delta\phi_{(j_1 j_2, \gamma\gamma)}$.
 - centrality variable, defined as $C_{\gamma\gamma} = \exp\left(-\frac{4}{(\eta_1 - \eta_2)^2} \left(\eta_{\gamma\gamma} - \frac{\eta_1 + \eta_2}{2}\right)^2\right)$.
- η_1, η_2 , and $\eta_{\gamma\gamma}$ are the pseudo-rapidities of two jets and the diphoton.
- the difference in azimuthal angle between two leading jets, $\Delta\phi_{jj}$

- the minimum distance between a leading or subleading jet and leading or subleading photon, $\min \Delta R(\gamma, jet)$.

In the training procedure, the invariant mass of diphoton is relaxed to the range 55 GeV to 120 GeV in order to include the simulated 60 GeV signal samples. This is done as a purpose to increase the simulated signal events. The kinematic variables are grouped by BDT using a Gradient Boosting (BDTG) algorithm which is implemented in scikit-learn [108]. This BDTG is trained on the simulated events. The VBF $H \rightarrow \gamma\gamma$ samples with m_H in the range from 60 to 115 GeV with 5 GeV granularity are employed as signal. All of the standard background without Drell-Yan and QCD samples are used. Additionally, ggH samples with m_H varies from 60 to 115 GeV are also used as background in the training to enhance the purity of dijet signal. The training, is done by using 70% of the signal/background events and rest of the events for testing. The distributions of dijet MVA output as well as the ROC curve are shown in Figure 5-68.

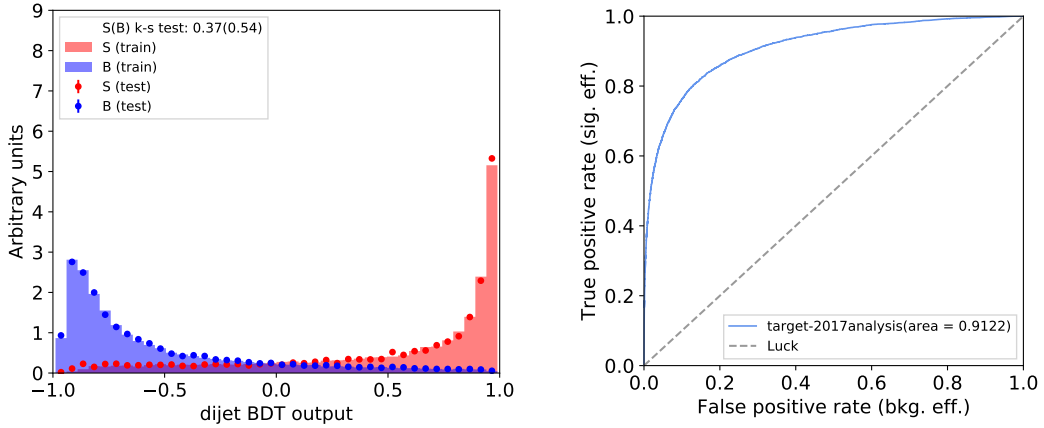


Figure 5-68 Comparison of training and the testing is on left which indicates no obvious sign of over-training and the ROC curve of dijet MVA training is on the right.

The combined (dijet plus diphoton) MVA is a BDT that receives the kinematic dijet MVA score, diphoton MVA score and the $p_T^{\gamma_1\gamma_2}/m_{\gamma\gamma}$ as input variables. This MVA is trained in order to discriminate the dijet signal from background processes by utilizing the information from corresponding objects in the tagged event. The variable $p_T^{\gamma_1\gamma_2}/m_{\gamma\gamma}$ is included due to its significant correlation to the dijet and diphoton MVA scores. The simulated VBF $H \rightarrow \gamma\gamma$ samples with m_H from 60 to 115 GeV (5 GeV step) are used to train this MVA. All of the standard backgrounds are used except QCD samples due to larger statistical fluctuation. The ggH samples are not utilized as a part of background

samples in combined MVA training because of the reason that, if we train with ggH samples, we are penalizing real diphoton events which degrade the performance. The discrimination between ggH and VBF events is done by using the information from the dijet system through dijet MVA. In this training, 50% of signal/background events are used in training and the rest for testing purpose.

As there are not much events left after all applying all the selections, we only have one VBF category. This category has been defined by directly comparing the significance (“p-value” symbolized as Z_0) of different cuts on the combined MVA score. The same method is used as the one employed in the diphoton BDT boundary optimization. Figure 5-69 shows the distributions of combined MVA output as well as the ROC curve.

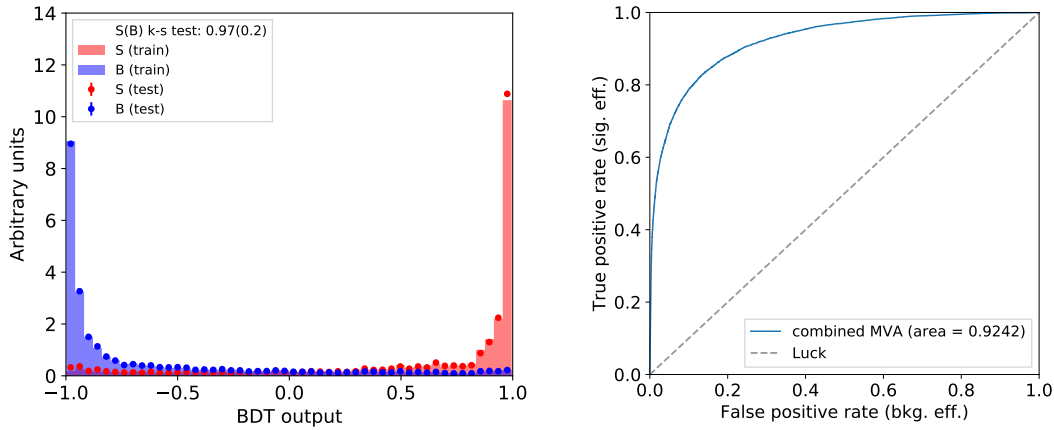


Figure 5-69 Comparison of training and the testing is on left which indicates no obvious sign of over-training and the ROC curve of combined MVA training is on the right.

Figure 5-70 shows significance as a function of combined MVA. The events selected from 90 GeV VBF MC sample are used as ‘signal’ and the ‘background’ events are selected from DiPhotonJetsBox and GJet MC samples. As can be seen from the plot, for highest Z_0 value with a cut at 0.9 (combined MVA > 0.9), will almost reject all of the VBF like events in data as well as simulation. We will be having less ‘single-fake’ events in data as well as DY MC which are not enough for the study of shape uncertainty of the relic DY (described later in this chapter). Its necessary to choose a cut value which give us enough VBF like events along with high significance value. If the cut value is smaller than 0.80, Z_0 value sharply declines. So, the value 0.8 was chosen as the VBF category boundary. It means that the events for the VBF category has to

pass the cut i.e. combined MVA > 0.80 . The events not surviving this cut value will go to the inclusive categories.

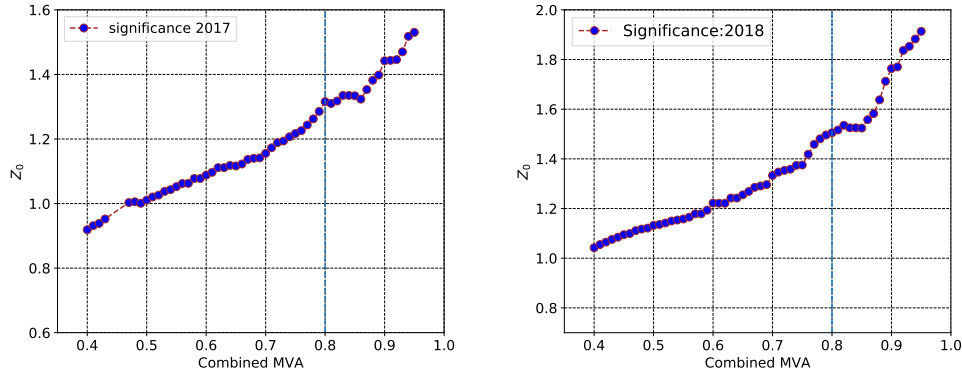


Figure 5-70 The significance, ‘p-value’ symbolized as Z_0 as a function of various cut values on combined MVA score for 2017 (left) and 2018 (right) using the same weight files from 2017 analysis.

As it is already mentioned that, the 2017 weight files for dijet and combined BDT are also employed in 2018 VBF class. Apart from the validations with 2018 $Z \rightarrow ee$ plus 2 jets events and diphoton events in both data and MC, we have also checked the 2018 performance (ROC) of dijet and combined BDT. Figure 5-71 shows the comparisons of ROC curves between 2017 and 2018. The events for these ROC curves are selected from 90 GeV VBF MC as ‘signal’ and the ‘background’ events are selected from DiPhotonJetsBox and GJet MC samples. From the ‘AUC (Area Under the Curve)’ values, there is a bit (2%) lower but still acceptable performance in 2018. For the 0.8 cut value of combined MVA, the signal efficiency is around 54.61% (54.48%) and the background efficiency is about 1.73% (1.81%) in 2017 (2018) from the signal and background samples used in the ‘ROC’ as shown in 5-71. Systematic uncertainties related to jets and to the VBF topology are also considered which are discussed in the dedicated systematics section of this chapter.

5.11.4 Validations

The validation of dijet MVA and combined MVA has been performed using $Z \rightarrow e^+e^-$ events containing two jets. In 2018 studies, we have used the 2017 weight files of dijet and combined BDT after the validations with the 2018 $Z \rightarrow ee$ plus 2 jets events and also the 2018 diphoton events. The events are required to pass the diphoton preselections

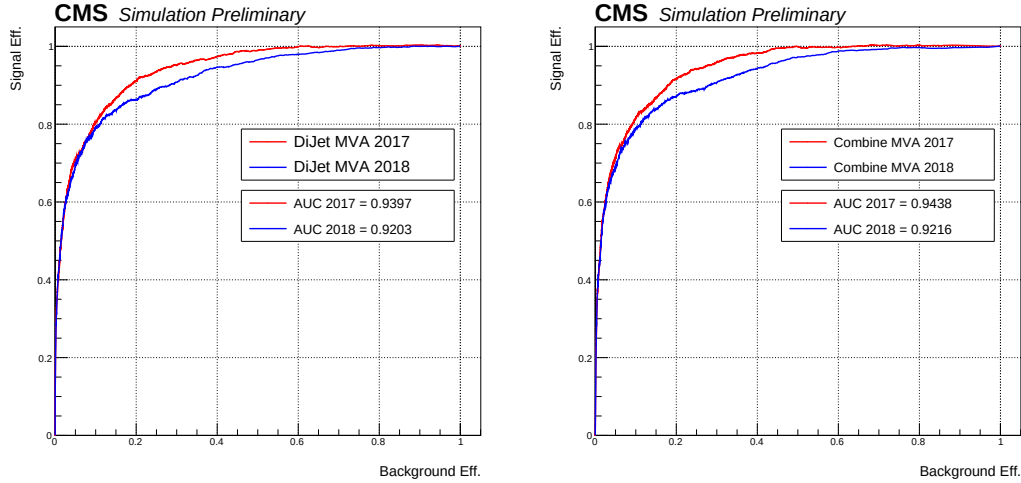


Figure 5-71 ‘ROC’ comparisons between 2017 and 2018 for the dijet MVA (left) and the combined MVA (right) with the same weights files from 2017 used in 2018 analysis.

but with inverted electron veto and a di-electron invariant mass in the range of 80 and 100 GeV. For 2017 data, events should firstly pass the HLT path HLT_Diphoton30PV_18 PV_R9Id_AND_IsoCaloId_AND_HE_R9Id_NoPixelVeto_Mass55, prescaled by a factor of 18, while the HLT path HLT_Diphoton30_18_R9IdL_AND_HE_AND_IsoCaloId_NoPixelVeto as used in the main analysis of 2018 data is applied for 2018 data. In addition, the VBF preselection as are also applied. HLT paths are not applied to MC in both 2017 and 2018.

The main systematics affecting input variables are included in the systematic band of MC. The systematic related to jet includes the uncertainties on the jet energy scale and resolution corrections (JEC and JER respectively) and the PUJID jet selection. The photon related systematics include the uncertainties on the photon ID MVA and the σ_E/E shift. Figure 5-72 and Figure 5-73 shows validations of some of the variables with $Z \rightarrow ee$ plus 2jets events selected from 2017/2018 data and simulation.

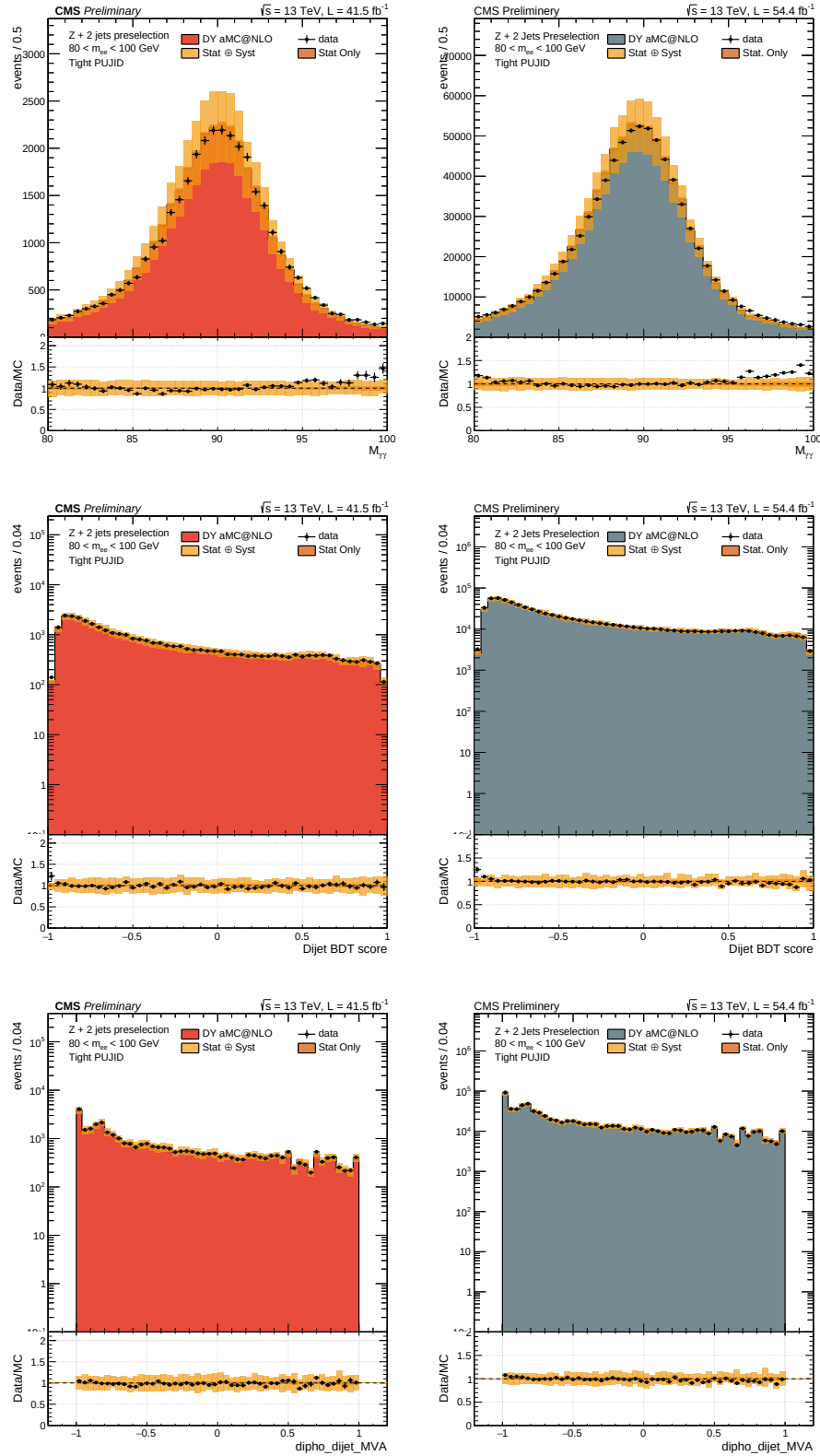


Figure 5-72 Invariant mass of diphoton candidates, dijet BDT and combined BDT in $Z \rightarrow ee$ plus 2jets events selected from 2017 data and MC (left), and in 2018 data and MC (right). MC is normalized according to number of events in data for each year.

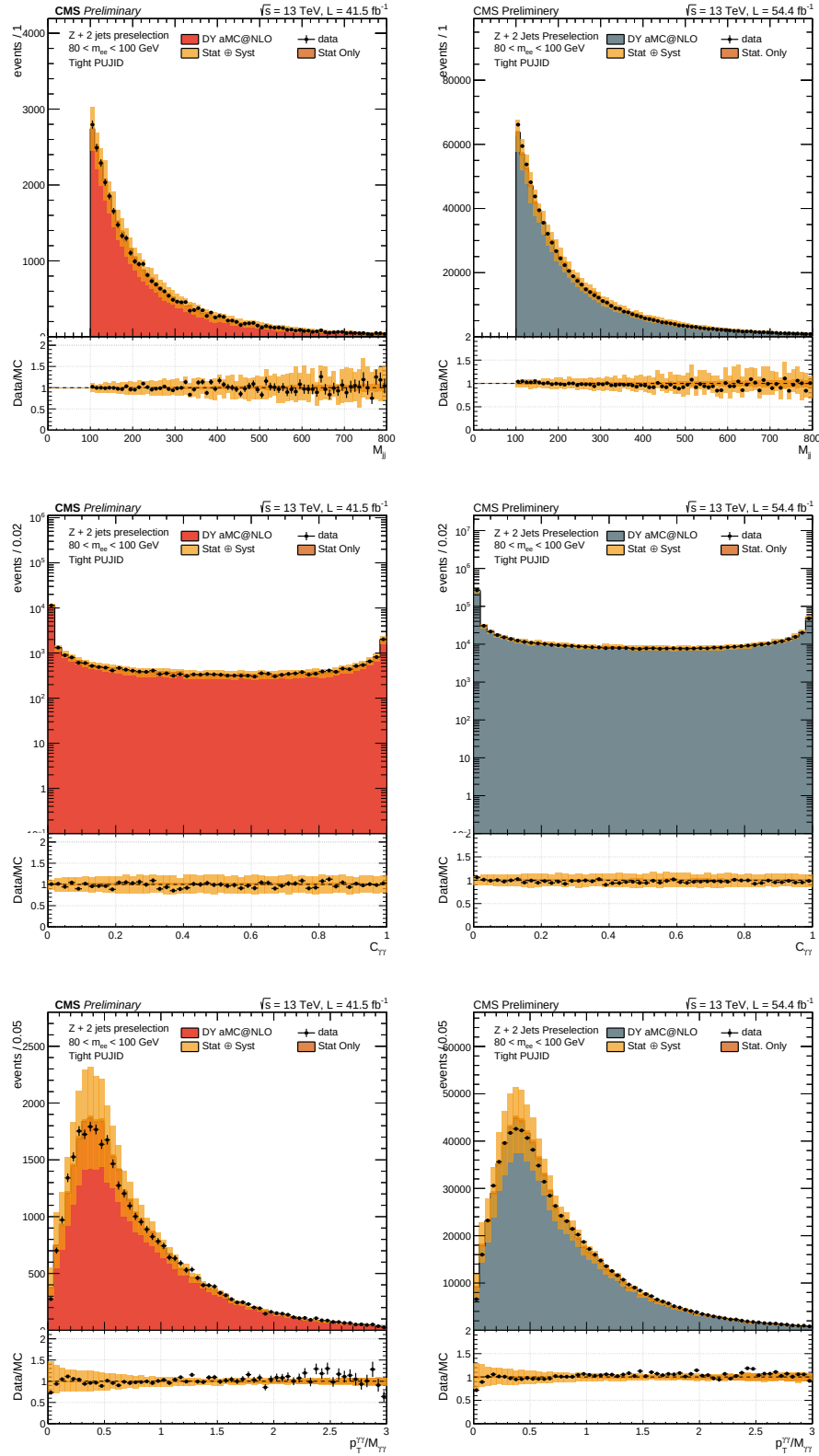


Figure 5-73 m_{jj} , $C_{\gamma\gamma}$ and $p_T^{\gamma\gamma}/m_{\gamma\gamma}$ in $Z \rightarrow ee$ plus 2jets events selected from 2017 data and MC (left), and 2018 data and MC (right). MC is normalized according to number of events in data for each year.

The dijet and combined MVA has also been validated with diphoton data and simulation same as that in diphoton MVA training. The events are selected in mass range $65 \text{ GeV} < m_{\gamma\gamma} < 120 \text{ GeV}$ after the diphoton preselection and the VBF preselection are used. The background MC samples are normalized according to the number of events in each 2017 and 2018 data. Figure 5-74 and Figure 5-75 shows the validations of some variables for 2017 and 2018 data.

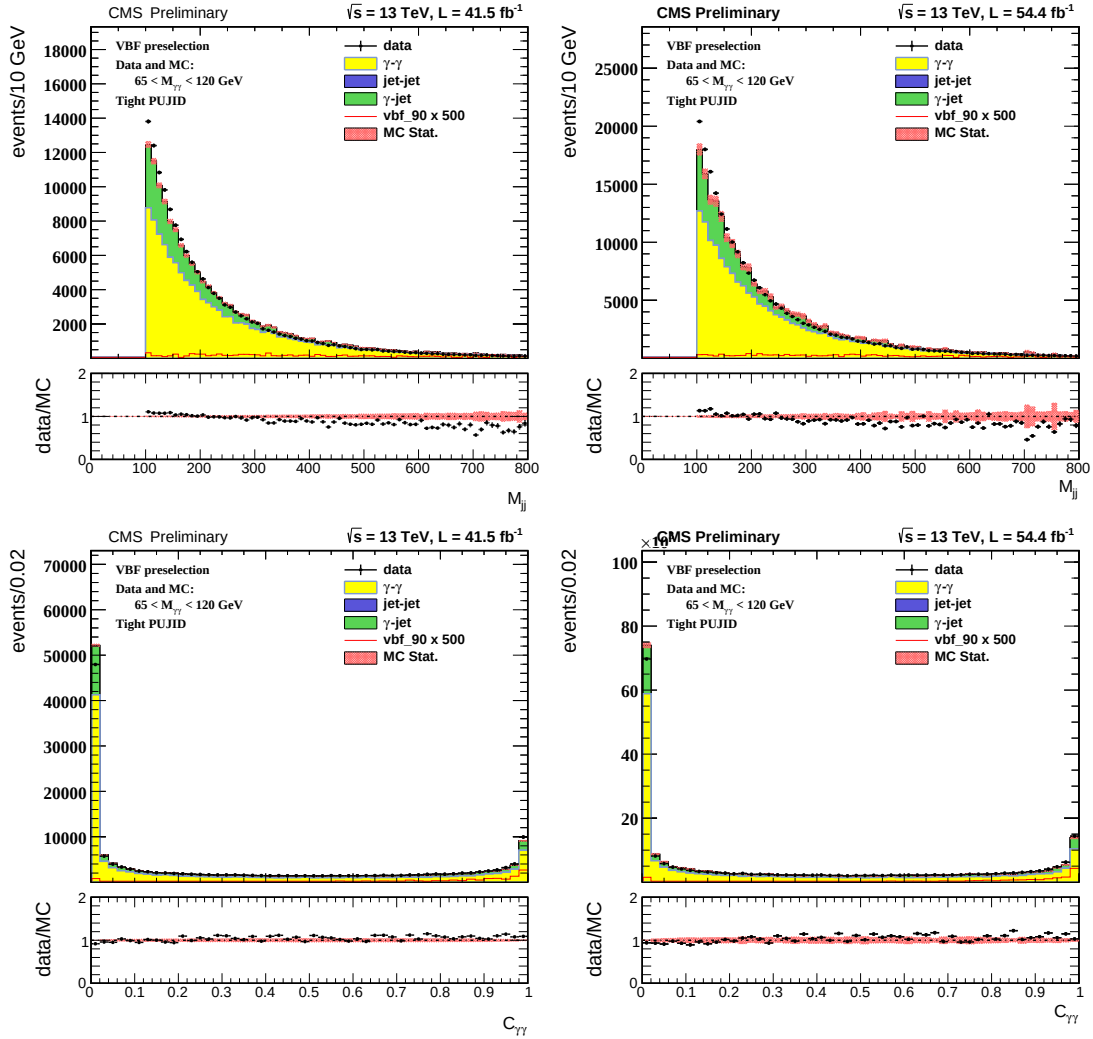


Figure 5-74 The distributions of m_{jj} and $C_{\gamma\gamma}$ in data and MC (without QCD) 2017 (left) and in 2018.

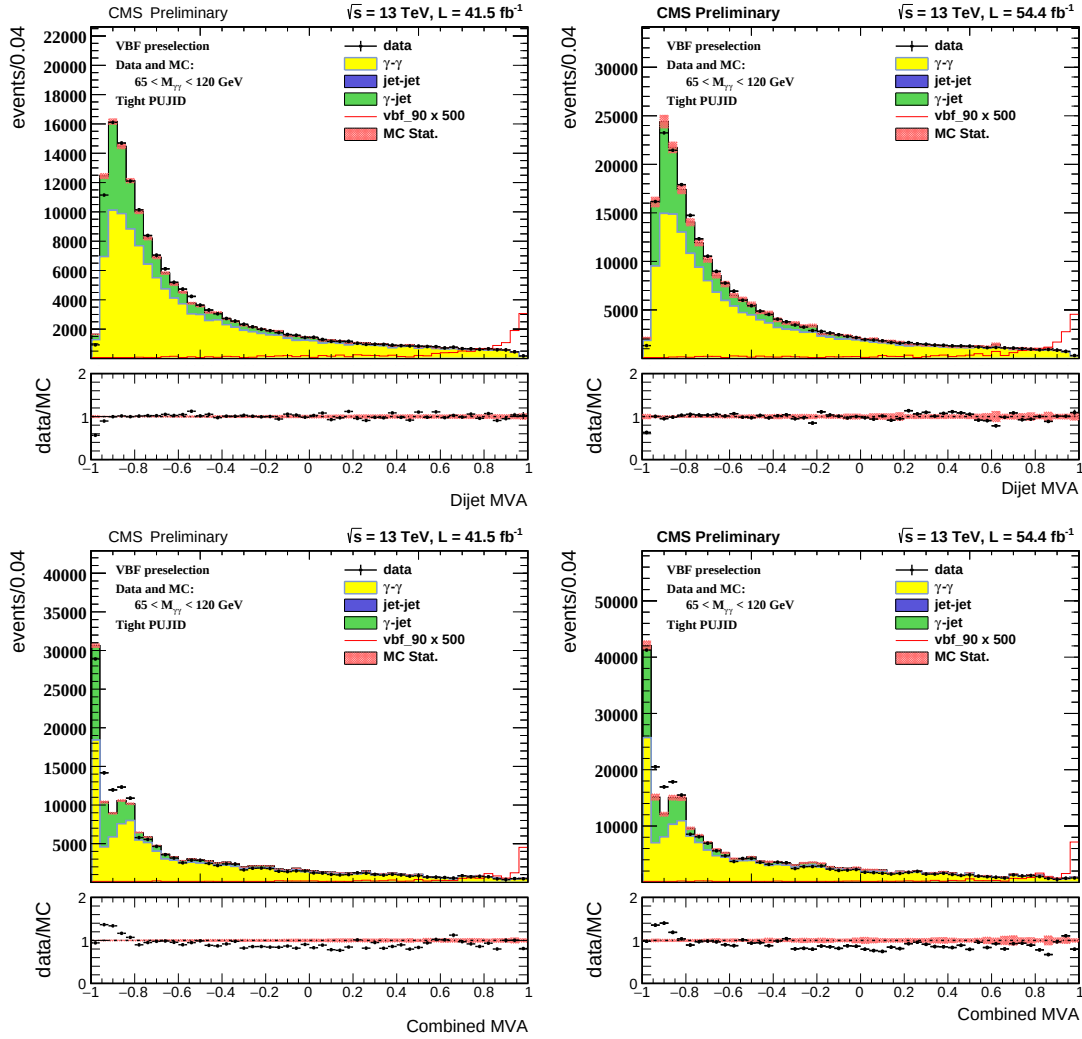


Figure 5-75 The distributions of dijet and combined MVA in data and MC (without QCD) 2017 (left) and in 2018.

5.12 Signal parametrization

The statistical interpretation of observed data is based on the description of signal which specifies the overall efficiency $\times$ acceptance along with the shape of the diphoton mass distribution in each of the event classes. The diphoton mass distribution shape is parameterized separately for each event class and production mode which includes ggH, VBF, $t\bar{t}$ +Higgs associated production, W and Z associated production. The vertex choice affects the diphoton mass distribution shape, so the modeling is done separately for the cases where the right vertex (RV) or wrong vertex (WV) was selected. The right primary vertex is considered if it is associated with the diphoton candidate within 1 cm of the true vertex. The signal distributions of diphoton mass are parameterized using a

sum of up to five Gaussian functions, as described in $H \rightarrow \gamma\gamma$ studies at CMS [105].

The signal shape is influenced by ECAL response and reconstruction, and is modeled empirically by a sum of between two and four (three and four) Gaussian functions for 2017 (2016 and 2018) data, depending on the event class. The signal shape for events having an incorrect primary vertex is smeared significantly by the variation in the z -coordinate of the selected primary vertex w.r.t. the true Higgs boson production vertex. These numbers of Gaussians are determined with 90 GeV signal samples in 2016, 2017 and 2018 separately. The strategy is different from the one used in the previous low-mass analysis with 2016 data set [14]. The simulation is used after the smearing of resolution and applying all the efficiency corrections and scale factors in order to build a parameterized model. The signal is defined continuously for any value of a Higgs boson mass (m_H) between 70 and 110 GeV.

The strategy of signal shape model is to describe the Monte Carlo with an analytic function. The function parameters are determined by a simultaneous fit of all various m_H samples where the individual parameters of functional form are polynomials of m_H . The floating parameters in the fit are the coefficients of these polynomials. The method which is referred as SSF (Simultaneous Signal Fitting), has an advantage of reduction in the total number of parameters used to determine the full model. Moreover, the SSF method is a sensible parametric model and on the other hand, the linear interpolation method can lead to discontinuous or unphysical models. The SSF method has applied separately for each process, category and RV or WV case. The fits for the RV and WV are summed up to make the signal model for each category. Like for the other parameters in the signal model, vertex efficiency is interpolated between mass points while constructing the final parametric model. For the categories having low statistics, the RV fraction is set to 1. Figure 5-76 shows the SSF, interpolation and Figure 5-77 illustrate a full parameterized signal shape for all classes combined case for 2016, 2017 and 2018 datasets. The signal efficiency $\times$ acceptance of the event selection as a function of mass hypothesis is shown in Figure 5-78. The systematic uncertainties related to the scale and smearing of individual photon energies, fraction of events where the right vertex was correctly identified, material corrections and the ECAL crystal light yields are considered into the signal model as additional parameters to be treated as nuisances.

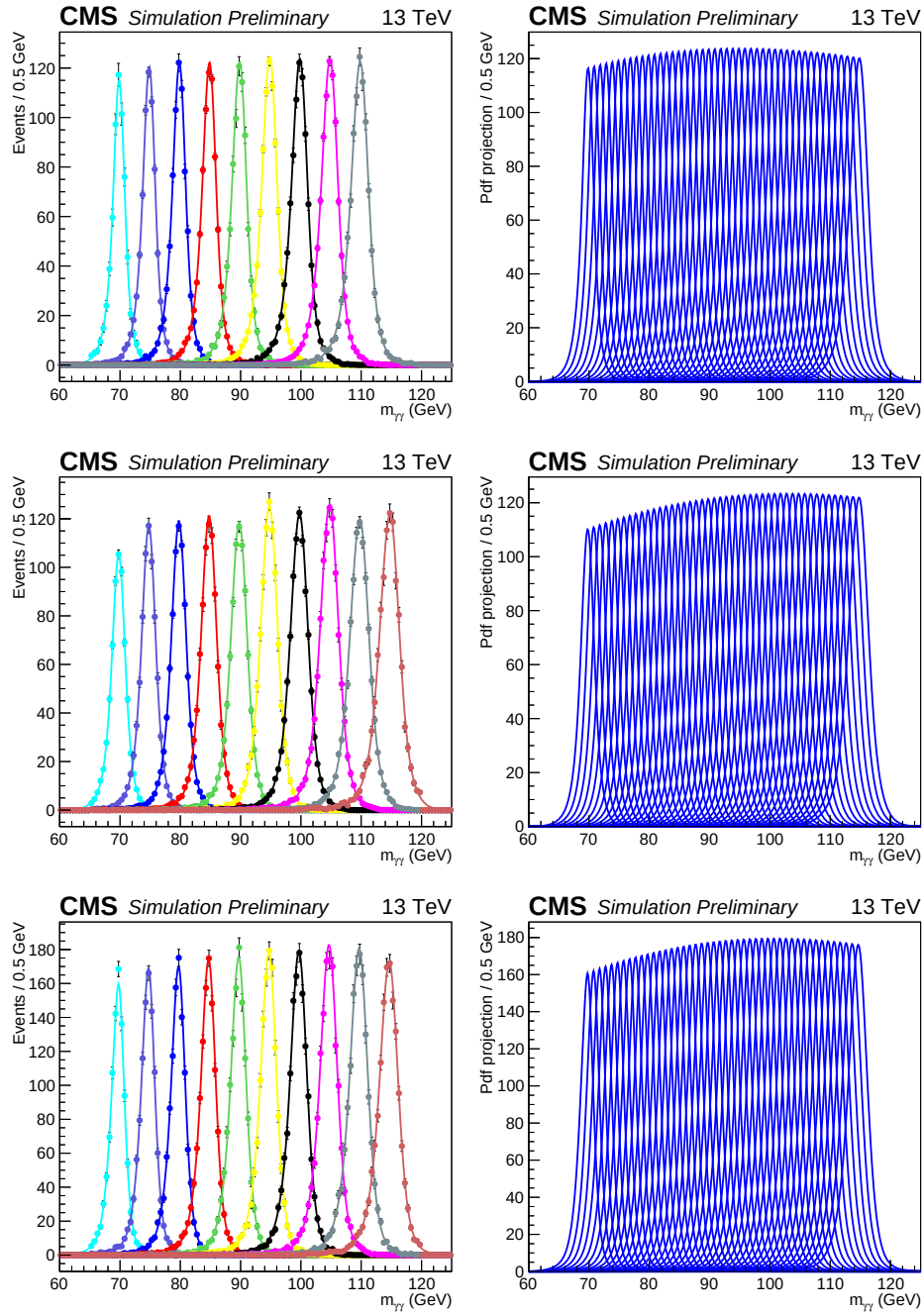


Figure 5-76 (2016 (top), 2017 (middle) and 2018 (bottom)) Simultaneous fit (left) and the interpolation (right) on the simulated signal events for all classes combined. Different production modes are summed according to their relative SM cross sections.

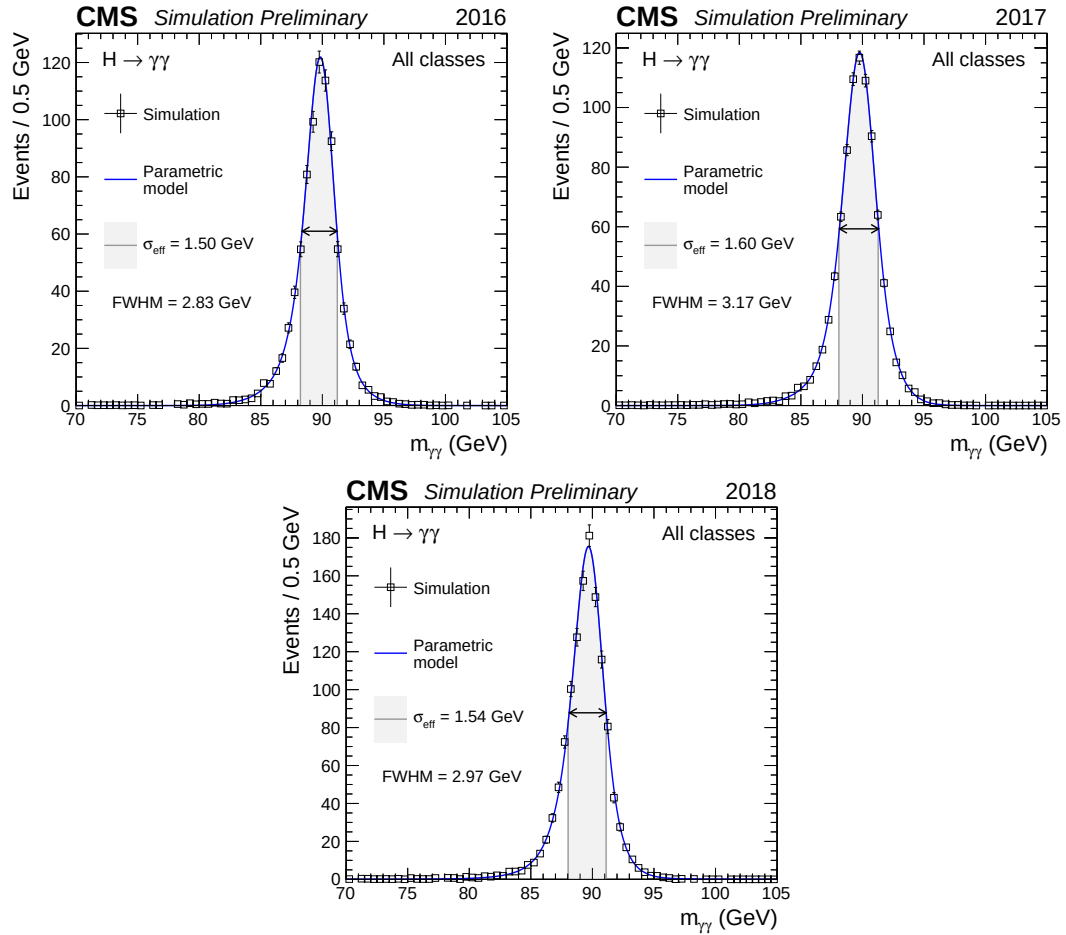


Figure 5-77 Full parameterized signal shape integrated over all event classes in 90 GeV Higgs boson Monte Carlo at $\sqrt{s} = 13$ TeV in 2016 (top left), 2017 (top right) and 2018 (bottom). The black points are the weighted Monte Carlo events and the blue lines are the corresponding models. The effective σ values and corresponding intervals are also shown.

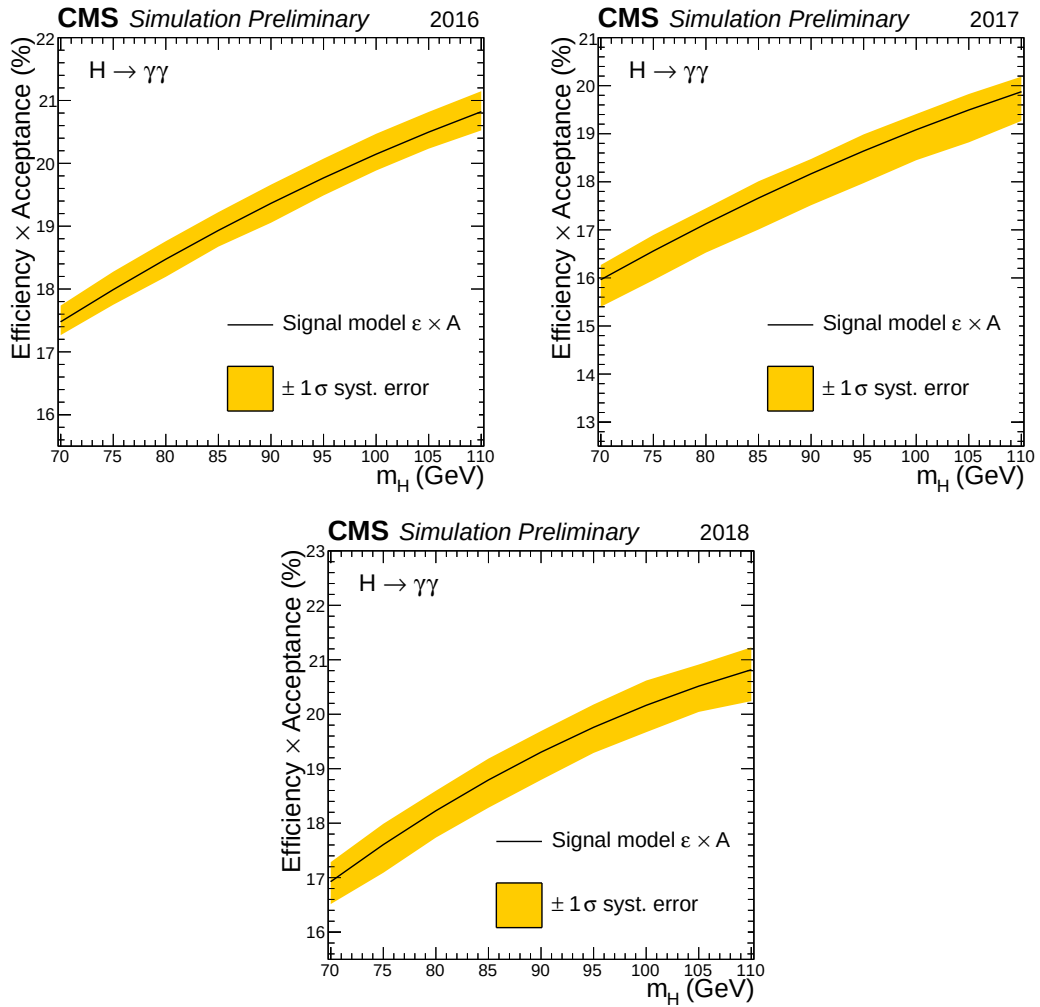


Figure 5-78 Signal efficiency $\times$ acceptance in (top left) 2016, (top right) 2017 and (bottom) 2018 for the analysis as a function of mass hypothesis.

5.13 Background estimation

In the analysis presented here in this chapter, background modelling is done by fitting analytic functions to the observed diphoton mass distributions for each analysis category within the range $65 < m_{\gamma\gamma} < 120$ GeV. The methodology to deal with the continuous background is same as in [109], the ‘envelope’ or ‘discrete profiling’ method [110].

A double-sided Crystal Ball (DCB) function which is defined in Eq. 5-1 is included to treat the background from the Drell-Yan process where, the two electrons are misidentified as photons. The DCB function is identical to a Gaussian distribution but with asymmetric tails. It has six parameters i.e. $\mu, \sigma, \alpha_L, n_L, \alpha_R, n_R$ where μ and σ parameters has same meaning like for a Gaussian function. The parameters $\alpha_L(\alpha_R)$ and $n_L(n_R)$ describe the behavior of left (right) tail. The $\alpha_{L,R}$ is responsible for the tail start and $n_{L,R}$ corresponds to the decreasing power of the tail. An exponential is added to the DCB, as it alone is not sufficient to cover all background events in relic Drell-Yan processes. This exponential has two parameters i.e. exp_p1 and frac_sum . exp_p1 is the coefficient in the additional exponential and frac_sum is the fraction that the exponential function represents in the total Drell-Yan process model. The ‘DCB(x)+exp(x)’ function in the following is in fact ‘ $(1.0-\text{frac_sum}) \times \text{DCB}(x) + \text{frac_sum} \times \text{exp}(x)$ ’ with x being the invariant mass of selected diphoton candidates.

$$DCB(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{n_L}{|\alpha_L|} \right)^{n_L} \exp\left(-\frac{|\alpha_L|^2}{2}\right) \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - \frac{x-\mu}{\sigma} \right)^{-n_L} & , \text{ if } \frac{x-\mu}{\sigma} < -\alpha_L \\ \frac{1}{\sqrt{2\pi\sigma}} \left(\frac{n_R}{|\alpha_R|} \right)^{n_R} \exp\left(-\frac{|\alpha_R|^2}{2}\right) \left(\frac{n_R}{|\alpha_R|} - |\alpha_R| + \frac{x-\mu}{\sigma} \right)^{-n_R} & , \text{ if } \frac{x-\mu}{\sigma} > \alpha_R \\ \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & , \text{ otherwise} \end{cases} \quad (5-1)$$

The set of analytic functions considered in the discrete profiling method are, with N a normalization factor and f the fraction of the continuous background component:

$$\begin{aligned}
 N f \times \text{Exp}(x) + N(1-f)(\text{DCB}(x) + \text{exp}(x)) &= N f \times \sum_{i=0}^M \beta_{2i} e^{\beta_{2i+1}x} + N(1-f)(\text{DCB}(x) + \text{exp}(x)) \\
 N f \times \text{Pow}(x) + N(1-f)(\text{DCB}(x) + \text{exp}(x)) &= N f \times \sum_{i=0}^M \beta_{2i} x^{\beta_{2i+1}} + N(1-f)(\text{DCB}(x) + \text{exp}(x)) \\
 N f \times \text{Ber}(x) + N(1-f)(\text{DCB}(x) + \text{exp}(x)) &= N f \times \sum_{i=0}^M \beta_i b_{i,M} + N(1-f)(\text{DCB}(x) + \text{exp}(x)) \\
 \text{with } b_{i,M} &= \binom{M}{i} x^i (1-x)^{M-i} \\
 N f \times \text{Lau}(x) + N(1-f)(\text{DCB}(x) + \text{exp}(x)) &= N f \times \sum_{i=0}^M \beta_i x^{-\alpha_i} + N(1-f)(\text{DCB}(x) + \text{exp}(x)) \\
 \text{with } \alpha_i &= 4, 5, 3, 6, 2, 7, \dots \quad \text{for } i = 0, 1, 2, 3, 4, 5, \dots
 \end{aligned} \tag{5-2}$$

The DCB+exp parameters are determined by fitting the diphoton invariant mass distributions from relic Drell-Yan events which are known as 'double-fake' events after passing all the selection requirements. As the statistics is low in the Drell-Yan double-fake samples, we have fixed two of the six DCB parameters i.e. α_L and α_R to make the fit stable. The uncertainties on α_L and α_R are considered and discussed in the dedicated systematic section. They are optimized to give minimum or close to minimum χ^2 value of the fit in each 'single fake' event class selected in MC, without QCD events (due to large statistical fluctuation) for 2017, and in data for 2016 and 2018, while also giving reasonable fitted values of the parameters in double-fake relic events in DY MC. The optimized and fixed values of α_L and α_R are different in each event class as shown in Table 5-25 (2016), Table 5-26 (2017) and Table 5-27 (2018). The fits on the mass distributions of 'double-fake' events in DY MC, shown in Figure 5-79 (2016), Figure 5-80 (2017) and Figure 5-81 (2018), are performed in each event class and the parameter values are applied in determining the order for the polynomial component of the functions in Equation 5-2. For relic Drell-Yan events modeling in VBF tagged event class, less double-fake events selected from DY MC samples. To cope up with this issue, the events are selected from DY MC sample passing only the VBF preselection without any cut on combined MVA and due to this, additional systematic has also been considered.

For the required fitting with discrete profiling method, a set of functions within each family is selected. The same strategy as described in [14] is used. Four families of functions are considered i.e. sums of exponential, polynomials in the Bernstein basis, Laurent series and sums of power laws. An F-test is used to select representative functions from each of these four families in order to proceed with the discrete profiling.

Table 5-25 (2016) Fit parameter values for the double-fake Drell-Yan events.

Class	$\mu(GeV)$	$\sigma(GeV)$	n_L	n_R	α_L	α_R	exp_p1	exp_sum1
0	89.66±0.27	1.59±0.27	0.65±0.38	0.88±0.76	1.2	2.0	-0.07±0.14	0.08±0.21
1	89.79±0.20	3.04±0.17	0.47±0.28	1.17±0.28	1.2	2.0	-0.14±0.04	0.24±0.09
2	88.60±0.09	3.27±0.10	0.35±0.22	1.21±0.41	1.4	1.6	-0.08±0.01	0.33±0.04

Table 5-26 (2017) Fit parameter values for the double-fake Drell-Yan events.

Class	$\mu(GeV)$	$\sigma(GeV)$	n_L	n_R	α_L	α_R	exp_p1	exp_sum1
0	91.67±0.84	2.22±1.14	0.22±0.38	1.62±10.02	0.7	2.5	-0.09±0.06	0.22±0.28
1	89.96±0.30	2.26±0.31	0.28±0.16	2.85±5.24	1.0	2.0	-0.08±0.02	0.28±0.12
2	89.53±0.33	2.23±0.39	0.31±0.11	2.26±1.69	0.9	1.8	-0.09±0.01	0.48±0.06
VBF	90.55±0.45	2.33±0.47	0.72±1.46	10.96±10.76	1.1	2.2	-0.12±0.02	0.70±0.16

Table 5-27 (2018) Fit parameter values for the double-fake Drell-Yan events.

Class	$\mu(GeV)$	$\sigma(GeV)$	n_L	n_R	α_L	α_R	exp_p1	exp_sum1
0	88.50±0.29	1.65±0.32	0.60±0.68	1.71±1.46	1.2	1.7	-0.09±0.02	0.51±0.19
1	88.96±0.26	2.14±0.29	0.23±0.09	1.95±0.30	0.7	0.9	-0.12±0.04	0.14±0.08
2	88.91±0.19	2.90±0.30	0.35±0.22	1.21±0.41	0.9	1.7	-0.11±0.03	0.36±0.01
VBF	89.05±0.74	2.34±0.48	8.96±5.37	10.00±1.27	2.3	1.0	-0.12±0.01	0.83±0.03

When fitting the background $m_{\gamma\gamma}$ distribution, the value of twice the negative logarithm of likelihood (-2NLL) is minimized. The 1σ uncertainty is obtained by taking the width of 68% range. Some of the background model fits to the diphoton invariant mass distribution with the no signal hypothesis are shown in Figure 5-82 for 2016, Figure 5-83 2017 and Figure 5-84 for 2018 data. The best-fit function in the event class is identified as the “Best Fit Pdf” in each figure.

Figure 5-85 for 2016, Figure 5-86 2017 and Figure 5-87 for 2018, shows data and the best-fit background function (assuming no signal). The error bands includes the uncertainty only associated with the statistical uncertainties of the fits and are shown for illustration purposes only. The corresponding signal model for $m_H=90$ GeV, multiplied by 10 is also shown for illustration purposes.

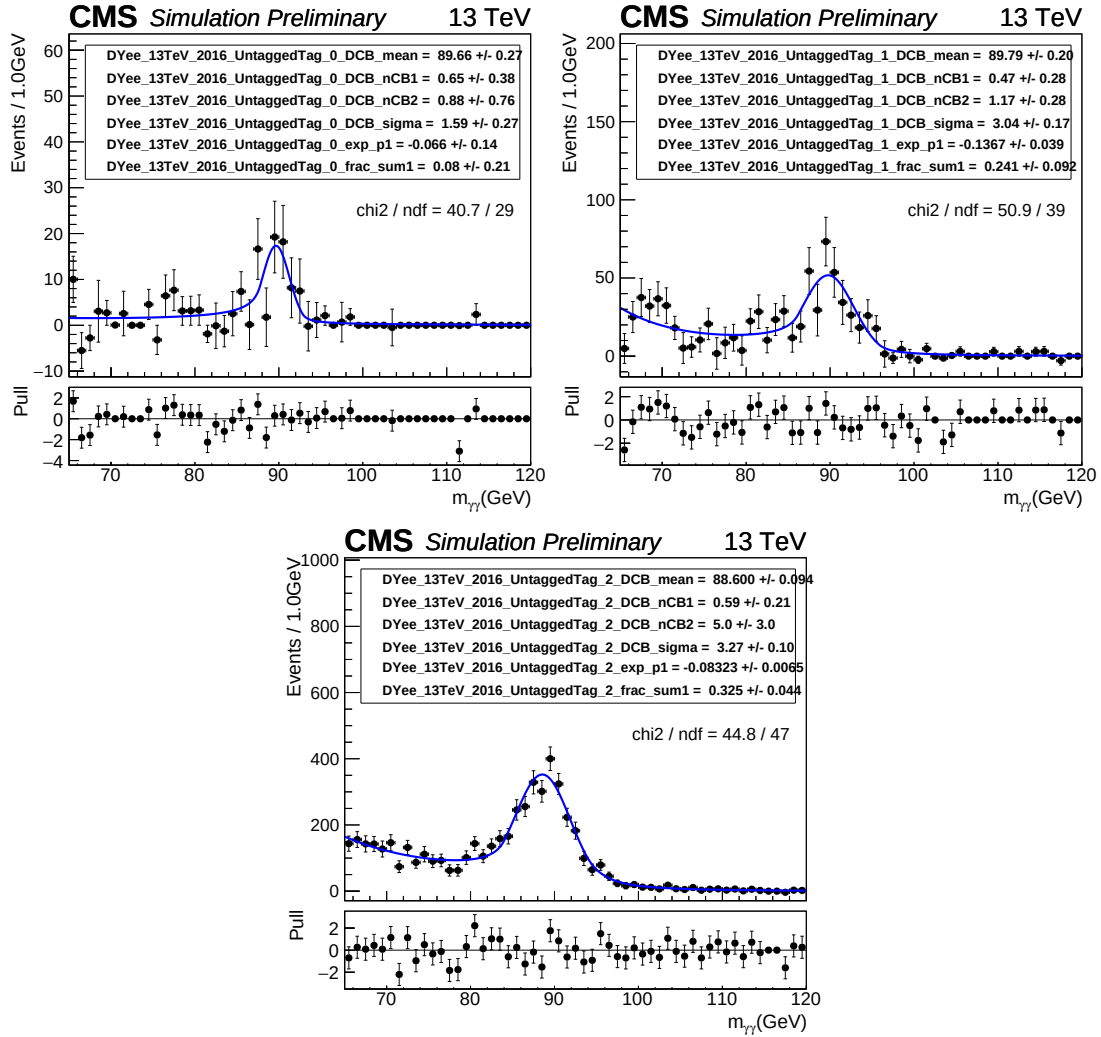


Figure 5-79 Double-sided Crystal Ball function fits to the diphoton invariant mass distributions from 2016 Drell-Yan double-fake events for Class 0 (top left), Class 1 (top right) and Class 2 (bottom).

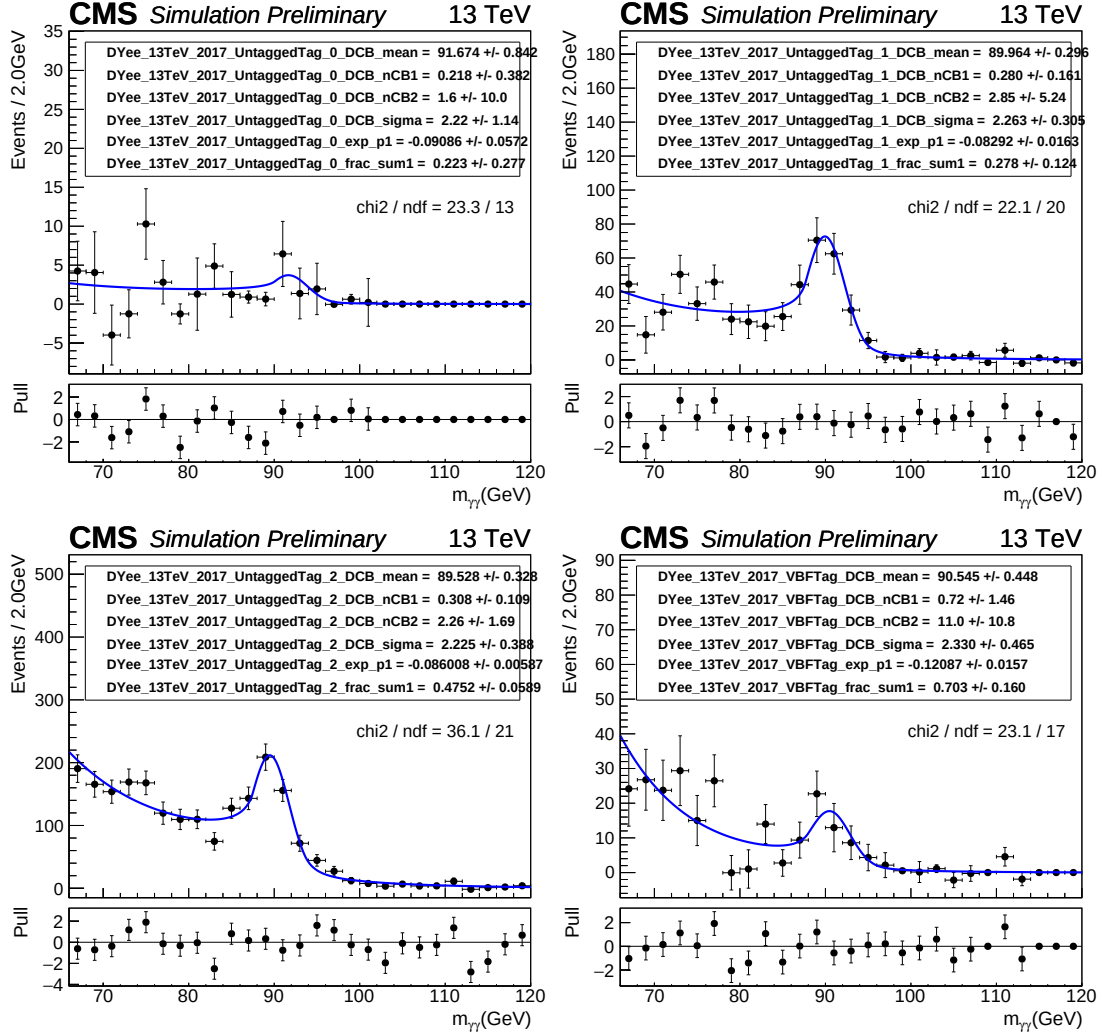


Figure 5-80 Double-sided Crystal Ball function fits to the diphoton invariant mass distributions from 2017 Drell-Yan double-fake events for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right).

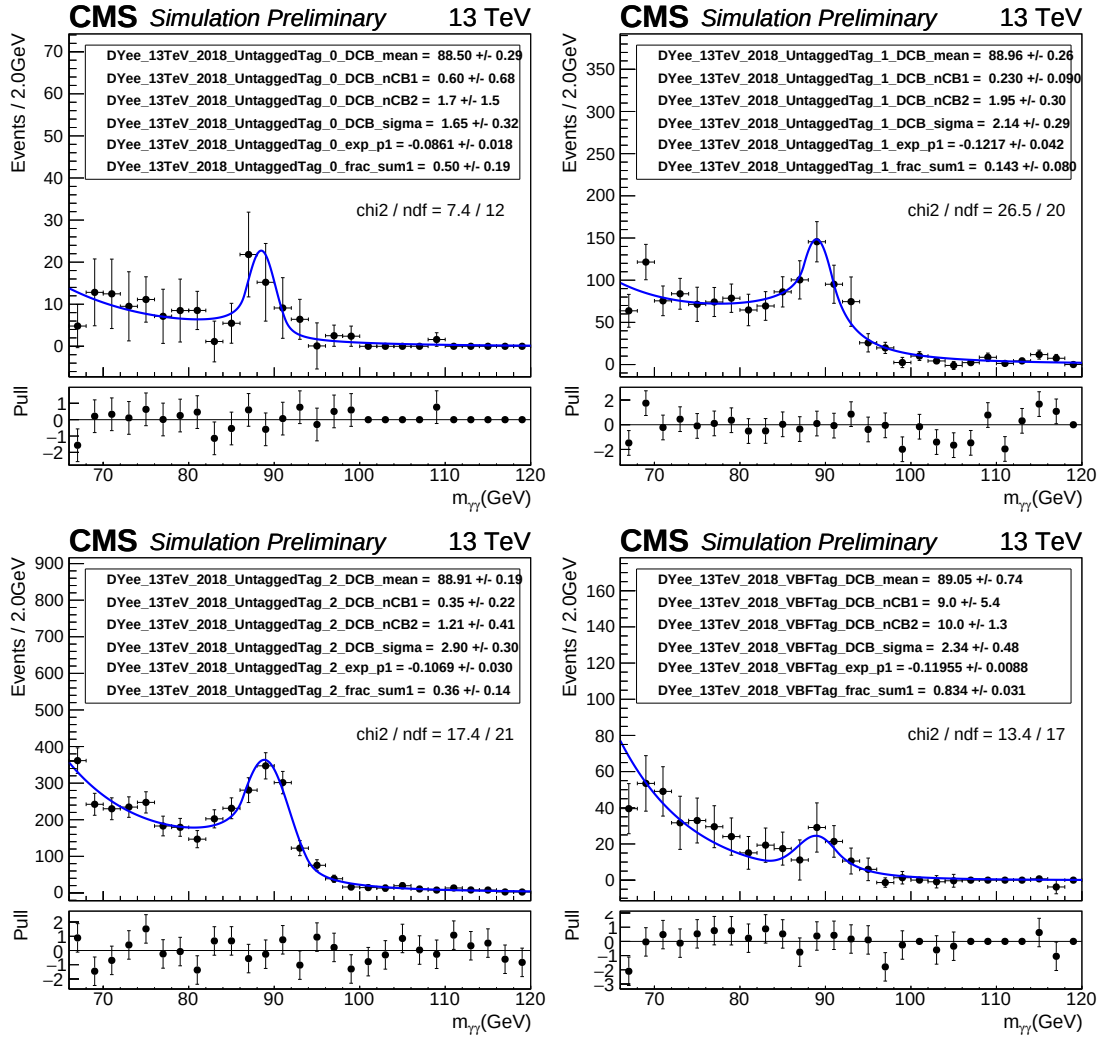


Figure 5-81 Double-sided Crystal Ball function fits to the diphoton invariant mass distributions from 2018 Drell-Yan double-fake events for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right).

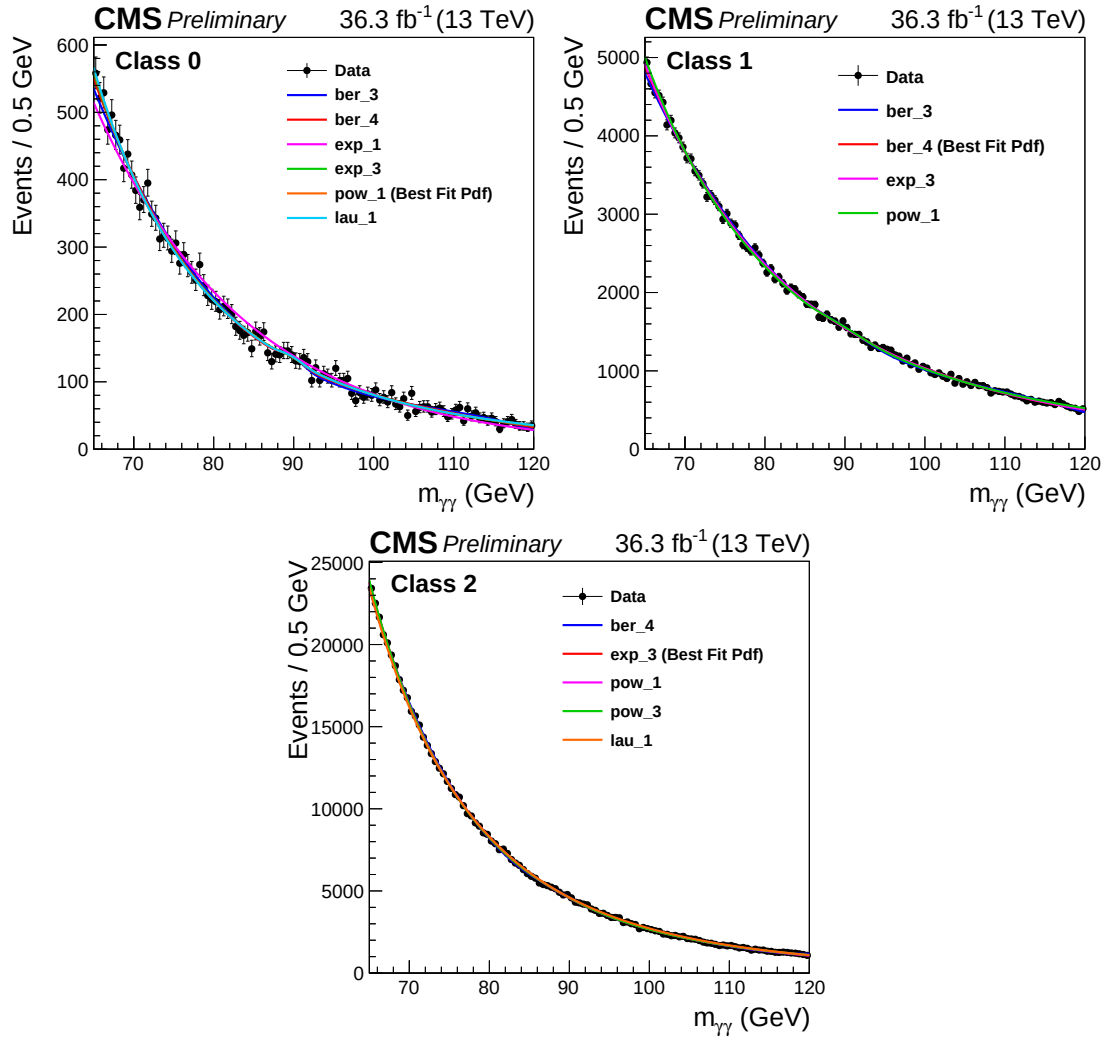


Figure 5-82 A set of functions chosen to fit the background using the discrete profiling method for each of the inclusive classes for 2016. Class 0 (top left), Class 1 (top right) and Class 2 (bottom).

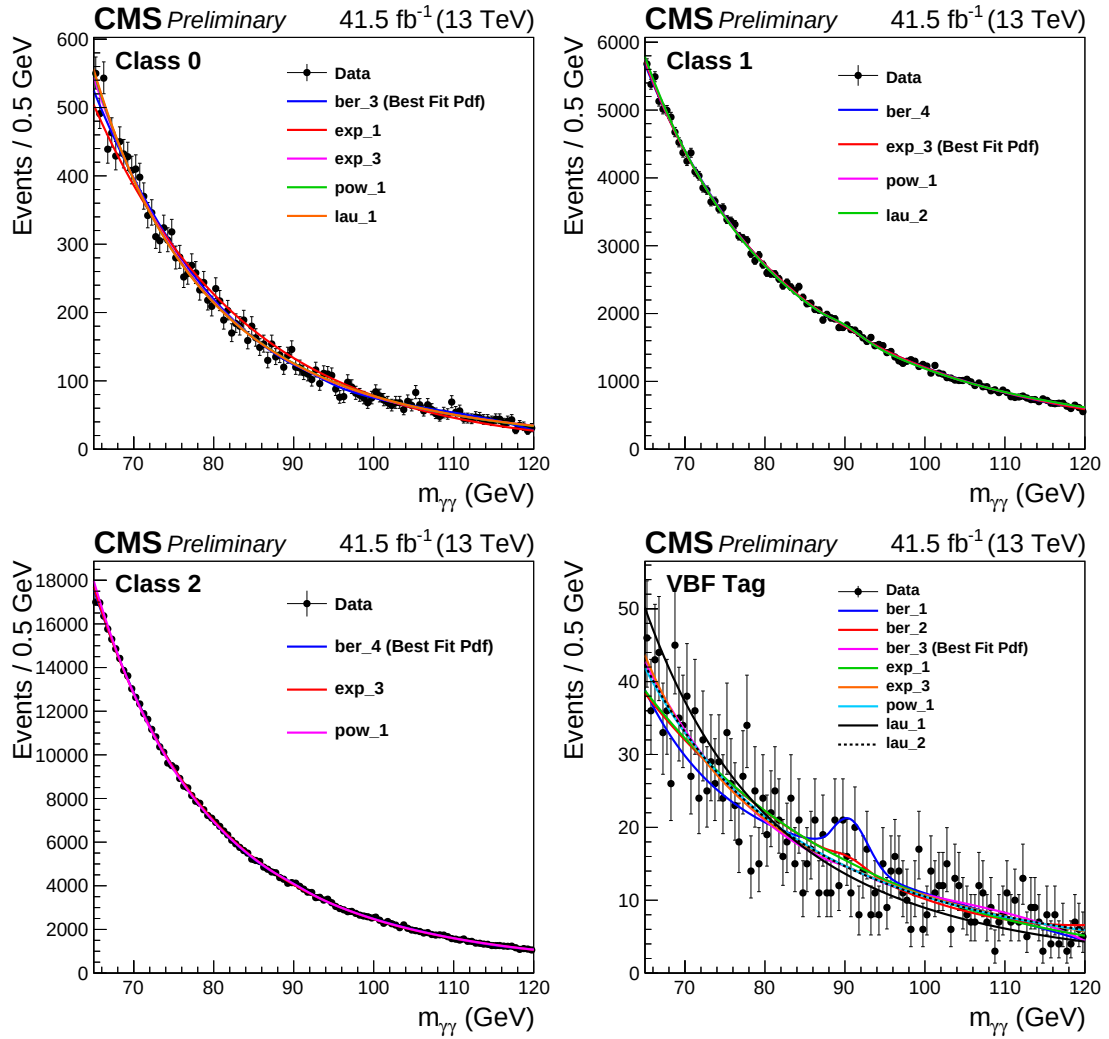


Figure 5-83 A set of functions chosen to fit the background using the discrete profiling method for each of the inclusive and VBF classes for 2017. Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right).

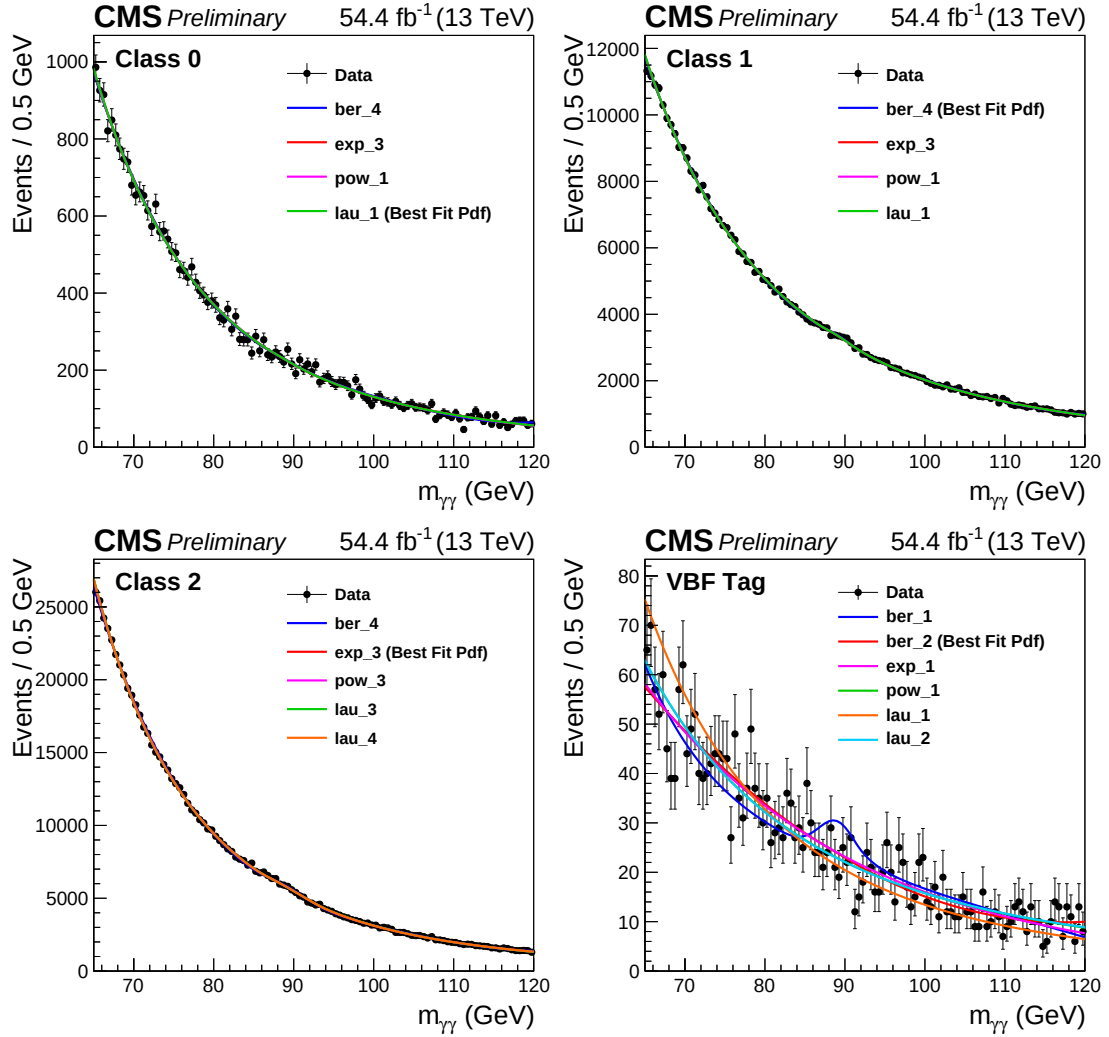


Figure 5-84 A set of functions chosen to fit the background using the discrete profiling method for each of the inclusive and VBF classes for 2018. Class 0 (top left), Class 1 (top right) and Class 2 (bottom left) and VBF Class (bottom right).

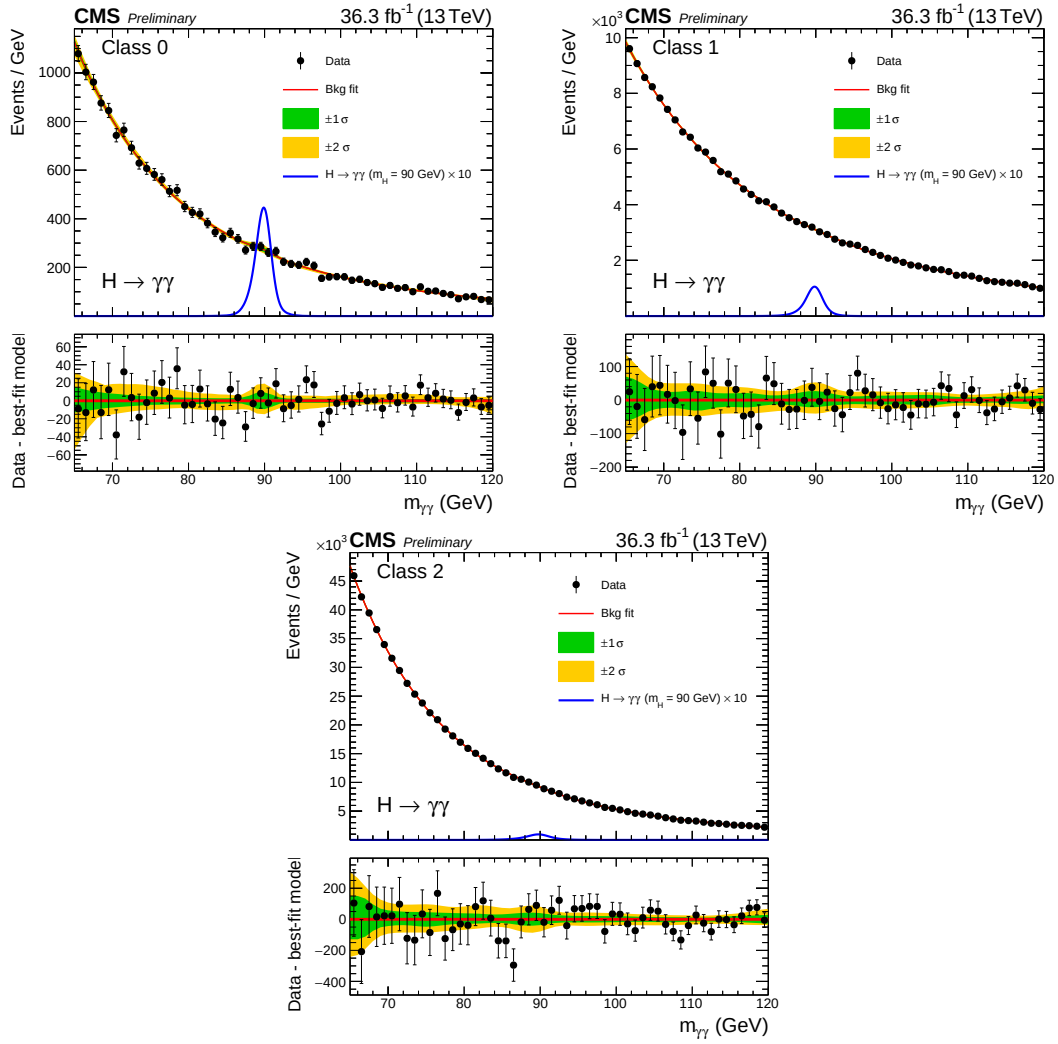


Figure 5-85 (2016) Background model fits using the chosen "best-fit" parametrization to data.

The corresponding signal model for $m_H = 90$ GeV, multiplied by 10, is also shown for Class 0 (top left), Class 1 (top right) and Class 2 (bottom). The difference between the data and the best-fit model is shown in the lower panels.

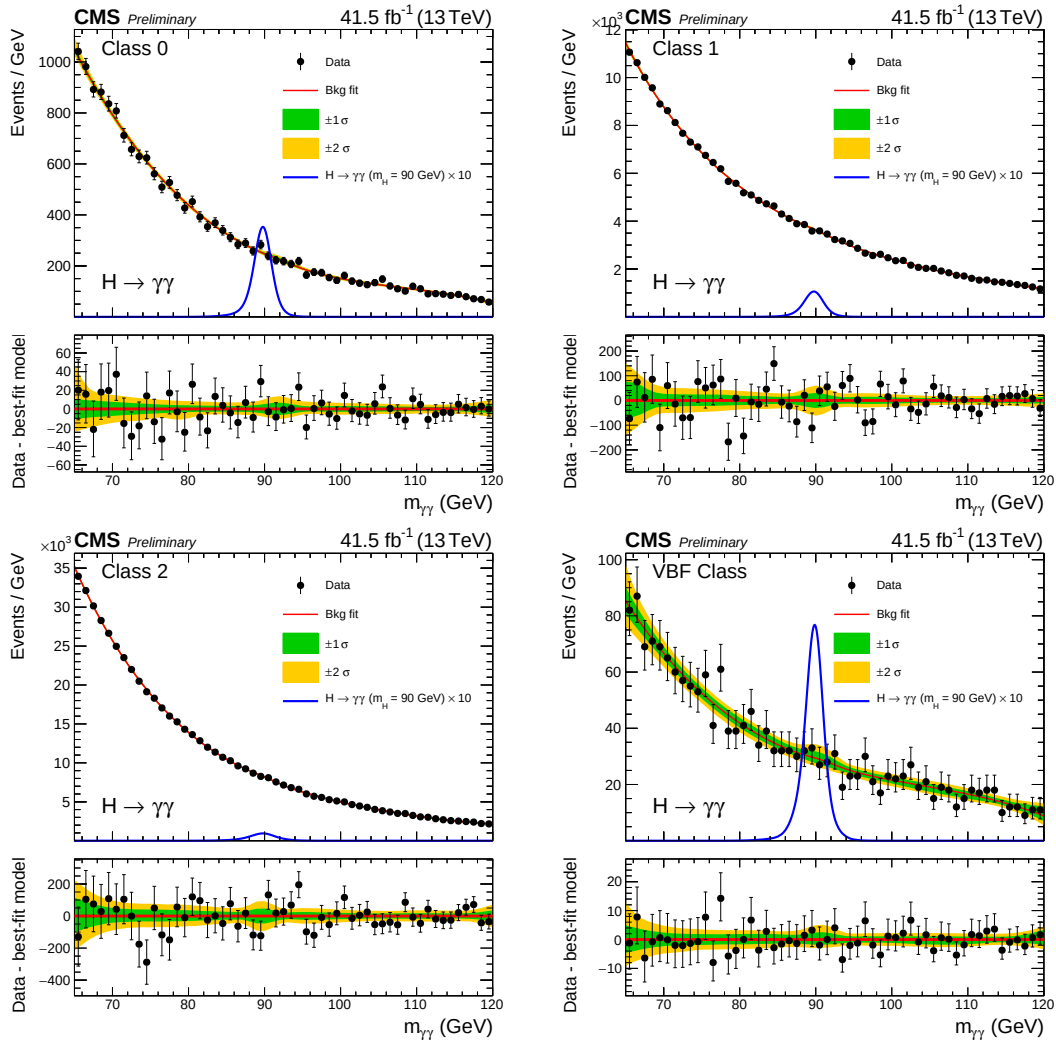


Figure 5-86 (2017) Background model fits using the chosen "best-fit" parametrization to data. The corresponding signal model for $m_H = 90$ GeV, multiplied by 10, is also shown for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). The difference between the data and the best-fit model is shown in the lower panels.

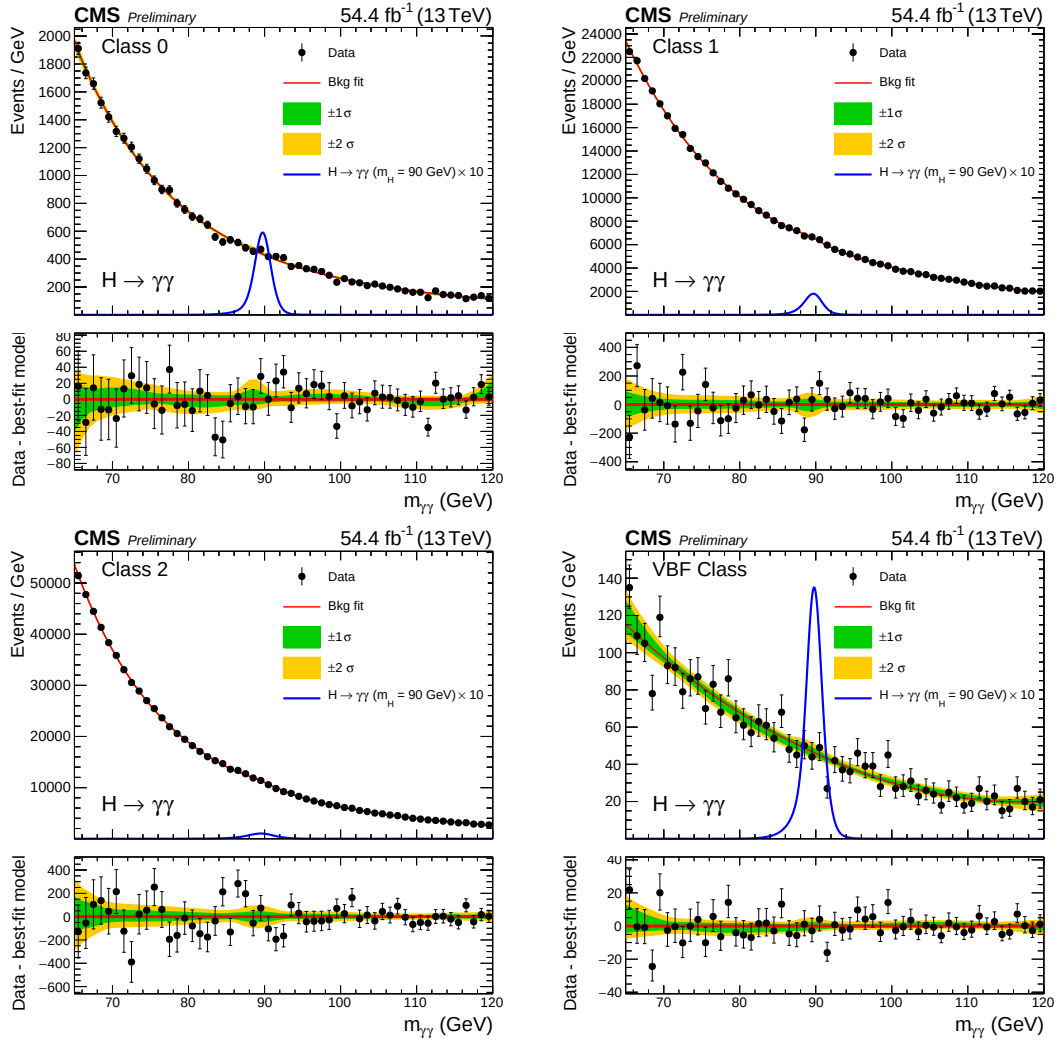


Figure 5-87 (2018) Background model fits using the chosen "best-fit" parametrization to data. The corresponding signal model for $m_H = 90$ GeV, multiplied by 10, is also shown for Class 0 (top left), Class 1 (top right), Class 2 (bottom left) and VBF Class (bottom right). The difference between the data and the best-fit model is shown in the lower panels.

5.14 Systematic uncertainties

Most of the systematic uncertainties relevant to [105] are also considered in this analysis. Those additional systematic uncertainties specific to this analysis are described in more detail below.

Per-photon systematic uncertainties

The systematic uncertainties correspond to the shape of photon ID BDT distribution and the estimate of per-photon energy resolution, as determined in [105], are considered to be sufficient and applicable for our studies. These were calculated using signal sample with $m_H=90$ GeV. For the photon identification BDT distribution shape, the relative efficiency variation of greatest magnitude are 6.9% (2016) occurs for the ttH process in event class 2, 6.4% (2017) occurs for the ggH process in event class 1 or 4.1% (2018) occurs for the VBF process in event class 1 for the other processes they are below 2.2% (2016), 5.3% (2017) or 3.7% (2018). For photon energy resolution, the largest relative variation 18.2%, occurs for the ggH process in 2016 class 1, the largest relative variation 14.6%, also occurs for the ggH process in 2017 class 1 and the largest relative variation 13.7%, occurs for the ttH process in class 2 in 2018. Other relative variations are below 12.9% (2016), 13.7% (2017) or 11.9% (2018).

For 2016, 2017 and 2018 data, uncertainties in the trigger efficiencies give rise to efficiency variations less than 0.8%/1.8%/0.5% respectively. The uncertainties in the scale factors of the preselection give rise to efficiency variations of 3.3%/3.2%/5.2% in 2016, 2017 and 2018 studies. The electron veto and minimum value of the photon identification BDT scale factors uncertainties are important sources of efficiency variations corresponds to less than 0.8%/0.9%/0.8% and 0.1%/0.4%/0.2% for the 2016, 2017 and 2018 data respectively. The scale factors uncertainties of the ' $N_e^{Matched} = 0$ ' requirement, give rise to efficiency variations of less than 0.5%/0.4%/0.4% in 2016/2017/2018 data. The uncertainties arising from material budget modeling, non-uniformity of light collection, non-linearity, imperfect electromagnetic shower simulation and vertex-finding efficiency are the same as those in [105]. The uncertainties in [105] are also propagated to the parametric signal model, where they result in uncertainties in the diphoton efficiency, mass scale, and resolution.

Uncertainties from jets and VBF dijets

The VBF tagged class in 2017 and 2018 relies on the presence of two jets along with

the diphoton. Events in the VBF tagged class are selected by using jet quantities and an MVA classifier which requires quantities related to the jets and their energies. Therefore, systematic uncertainties associated with jet energy scale, smearing corrections and the pileup jet identification are also implemented.

- *Jet energy scale and smearing corrections:* Jet energy scale corrections (JEC) account for maximum of 14.9% (2017) and 15.8% (2018) from VBF category to untagged categories. The jet energy resolution (JER) has an impact on the event migration of less than 1.1% in 2017 and 3.0% in 2018.

- *Pileup jet identification:* This uncertainty is estimated by comparing in data and simulation. The full discrepancy between data and simulation is used to estimate the event migration, which is less than 4.1% in 2017 and 3.2% in 2018.

Per-event uncertainties

Integrated luminosity: The per-event systematic uncertainty on the total integrated luminosity, estimated from data [111–113] which contributes to an uncertainty of 1.2%/2.3%/2.5% on the signal yield in 2016, 2017 and 2018.

Uncertainties from 2016 and 2017 pre-firing: A pre-firing weight for each diphoton object is implemented in MC signal events in order to consider the effect of the trigger pre-firing in 2016 and 2017. The relative variations are 1.3%/1.9%/1.1%/0.2%/1.6% for the ggH/VBF/WH/ZH/ttH production mechanisms in 2017 VBF tagged event class. For the rest of 2017 event classes, the relative variations are below 0.8%. In 2016 data, the relative variations are below 0.5% in all processes and in all of the corresponding event classes.

Underlying event and parton shower uncertainties: As there are no dedicated simulated low-mass signal samples so these uncertainties are calculated using dedicated 2016 125 GeV simulated samples using dedicated low-mass conditions with the only change in mass window i.e. from 65-120 GeV to 100-180 GeV. For the normalisation uncertainties from the modelling of underlying event, the largest variation is 13.7% in 2016 (class 2) in ttH production process. The largest variation is 26.8% in 2017 and 20.4% in 2018 with both occurring in ZH production process in the VBF tagged event class. The normalisation uncertainties from parton shower modelling, largest variation is 9.6% in 2016 (class 2) occurring in ttH production process. 15.9% in 2017 occurring in ZH production process in the VBF tagged event class and 12.0% in 2018 occurring in VBF

production process in the VBF tagged event class. Otherwise, they are typically below 7% in 2016 and 10% in 2017/2018.

Uncertainties from 2018 HEM (HE minus i.e. negative endcap HCAL sectors) issue: During data taking in late June 2018, two neighboring readout boxes (RBXs) in the -z endcap of HCAL, HEM15 and HEM16, lost communication and subsequently event information from the region of the detector covered by these RBXs is lost. In our studies, it has affected mainly the jet reconstruction. Jets near this region of the detector would be reconstructed without their full energy. Similar as the standard $H \rightarrow \gamma\gamma$ analysis [114], we don't drop the jets in this region in 2018 data, but try to quote the uncertainty of jet energy scale with source from this HEM issue. For untagged event classes, the uncertainties are in general less than 0.5% but can be as large as 1.2% for event class 0 in VBF production mode. For VBF tagged event class, the uncertainties range from 1-3% with the largest uncertainty 3.4% occurring in ggH production mode.

Uncertainties from VBF additional jet radiation issue: For the VBF signal sample, there is an issue from the additional jet radiation. With PYTHIA8 version ≥ 8.23 , it is recommended to set "dipoleRecoil" to be "on" in order to cure the un-physical effect due to inaccurate treatment of the initial-final color connections. There is no low-mass VBF signal sample with this issue repaired. We check the effects with two 2018 125 GeV VBF MC samples, one with and the other without this issue repaired with low-mass conditions applied in each 2017 and 2018. The only change is the mass window, from 65-120 GeV to 100-180 GeV. The uncertainties range from about 0.1-3% depending on different event classes, with the largest uncertainty 2.9% in 2017 and 2.6% in 2018 occurring in the VBF tagged event class. The linear cut scale factors uncertainties, give rise to efficiency variations of 0.4%/0.9%/0.5% in the case of the 2016, 2017 and 2018 data.

There are two types of theoretical systematic uncertainties considered in this analysis. Firstly, theoretical uncertainties on the signal acceptance coming from PDF, QCD scale and α_s variations are calculated as in [105] using a signal sample with $m_H=90$ GeV. For the PDF variations, the uncertainties are negligible in each 2016, 2017 and 2018. For the QCD scale variations, largest variation is 11.9% in 2016 (class 0) in ggH production process, 10.0% in 2017 (class 0) in ggH production process and 13.7% in 2018 in ggH production process in the VBF tagged event class. For α_s , the largest variations

is 1.9% in 2016, 3.9% in 2017 and 4.1% in 2018. The overall variations are typically below 0.5%. Secondly, for the calculation of the expected limit on the product of the production cross section times branching ratio into two photons relative to the SM-like BSM expectation and for the expected p-value, the uncertainties on the production cross sections of a Standard Model-like Higgs boson from PDF and scale variations are taken into account, following the recommendations of the LHC Higgs Cross Section working group [59].

Dedicated uncertainties on DCB parameters

The systematic per-event level uncertainty specific to this analysis is the modeling of the Z boson resonance within the background model. The DCB+exp parameters used to model the Z boson resonance are obtained from simulated Drell-Yan events with all of the selections applied including the electron veto requirement (‘double-fake’ events). These parameters might be different in data and simulation. In order to address these differences, we study simulated events from the Drell-Yan, diphoton, photon + jet and jet + jet processes where one photon candidate survives all selection requirements including the electron veto and the ‘ $N_e^{Matched} = 0$ ’ requirement, and the other survives all selection requirements but fails the electron veto and no ‘ $N_e^{Matched} = 0$ ’ requirement on it (‘single-fake’ events). Additionally, similar to $Z \rightarrow ee$ events, the modified $\log(\Sigma p_T^2)$ with the matched tracks to each of two photon candidates removed is used in the linear cut, for the selections of ‘single-fake’ events.

We fit the diphoton invariant mass of single-fake events in data, in Monte Carlo including the sum of all background processes (‘ $MC_{allwoQCD}$ ’) without QCD because of large statistic fluctuation from QCD samples and in DY MC alone (‘ MC_{DY} ’) with a DCB plus exponential component to describe the additional continuum background in the single-fake events. In 2017 analysis, since the prescaled HLT is used for the ‘single-fake’ events in data, resulting not enough events in the VBF tagged event class. So the ‘Replacement’ samples are used for the ‘single-fake’ events in the 2017 data. Consistently the ‘Replacement’ of the ‘single-fake’ events in DY MC and ‘ $MC_{allwoQCD}$ ’ are used in the 2017 analysis. We have also considered an additional systematic on DCB mean and sigma because of ‘Replacement’ in relic DY events modeling. It is done by comparing the fitted values of real VBF selected events (combined MVA > 0.8) with the events passing only VBF preselection cuts (without cut on combined MVA) from

the single-fake DY MC. Some of these fits are shown in Figure 5-88 for 2016 and Figure 5-89 for 2017 and 2018.

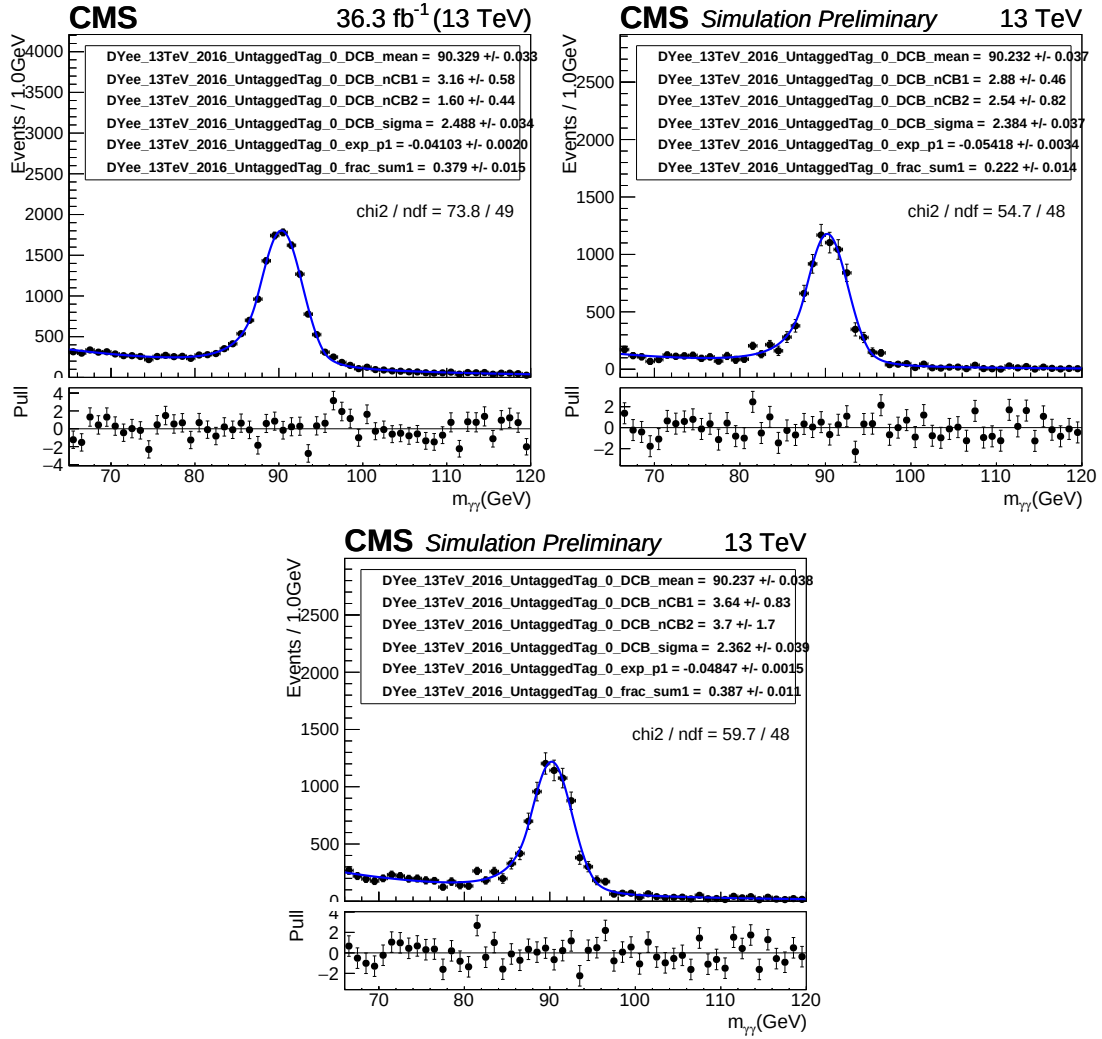


Figure 5-88 (2016: Class 0) Fit of the single-fake events for Data (top left), DY MC (top right) and $MC_{allwoQCD}$ (bottom).

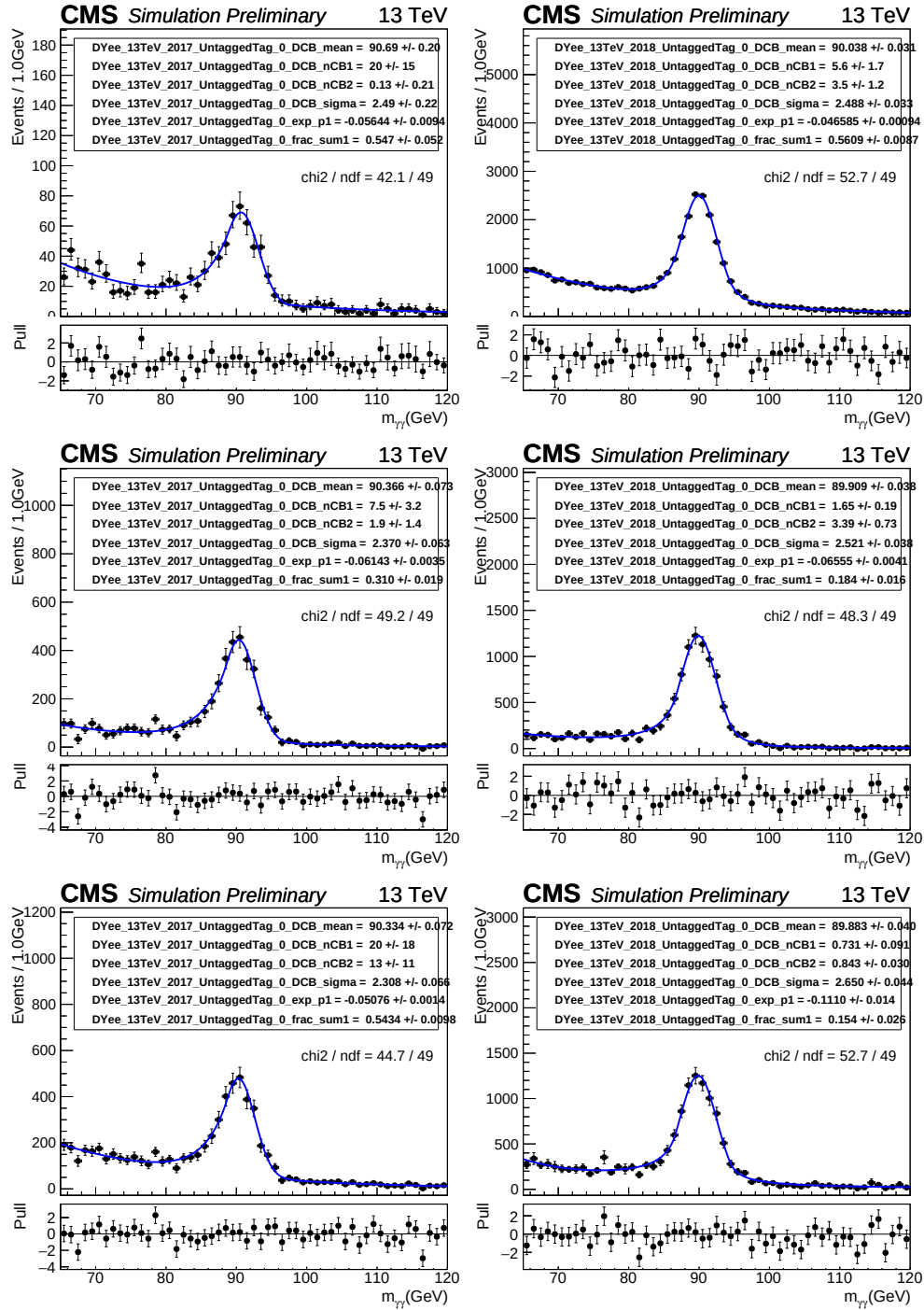


Figure 5-89 (2017 (left) / 2018(right) Class 0): Fit of the single-fake data (top), DY MC (middle) and $MC_{allwoQCD}$ (bottom) events.

We have considered the differences in the DCB μ and σ parameters extracted from these three kinds of fit and those which are statistically significant will enter into the total systematic error on the Z resonance modeling. Tables 5-28 and 5-29 (2016), Tables 5-30 and 5-31 (2017), Tables 5-32 and 5-33 (2018) show the fitted values of μ and σ from the fits of "single-fake" data, $MC_{allwoQCD}$ and MC_{DY} in each event class, together with the statistical errors and the differences between data and $MC_{allwoQCD}$, and between $MC_{allwoQCD}$ and MC_{DY} in units of GeV. The differences marked "Y"("N") will (not) enter into the total systematic error.

Table 5-28 (2016) Fitted values of μ (GeV) from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 3 event classes.

	μ_{data}	$\mu_{MC_{allwoQCD}}$	$\mu_{MC_{DY}}$	$1.0\sigma_{data+MC_{allwoQCD}}$	$ \Delta\mu_{data-MC_{allwoQCD}} $	$1.0\sigma_{MC_{allwoQCD}+MC_{DY}}$	$ \Delta\mu_{MC_{allwoQCD}-MC_{DY}} $
Event class 0	90.33 $\pm$ 0.03	90.24 $\pm$ 0.04	90.23 $\pm$ 0.04	0.05	0.10 (Y)	0.06	0.01 (N)
Event class 1	90.30 $\pm$ 0.01	90.22 $\pm$ 0.01	90.21 $\pm$ 0.01	0.01	0.09 (Y)	0.01	0.01 (N)
Event class 2	89.57 $\pm$ 0.01	89.40 $\pm$ 0.01	89.37 $\pm$ 0.01	0.01	0.20 (Y)	0.01	0.03 (Y)

Table 5-29 (2016) Fitted values of σ from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 3 event classes.

	σ_{data}	$\sigma_{MC_{allwoQCD}}$	$\sigma_{MC_{DY}}$	$1.0\sigma_{data+MC_{allwoQCD}}$	$ \Delta\sigma_{data-MC_{allwoQCD}} $	$1.0\sigma_{MC_{allwoQCD}+MC_{DY}}$	$ \Delta\sigma_{MC_{allwoQCD}-MC_{DY}} $
Event class 0	2.49 $\pm$ 0.03	2.36 $\pm$ 0.04	2.38 $\pm$ 0.04	0.05	0.11 (Y)	0.06	0.02 (N)
Event class 1	2.23 $\pm$ 0.01	2.28 $\pm$ 0.01	2.28 $\pm$ 0.01	0.01	0.08 (Y)	0.01	0.03 (Y)
Event class 2	2.72 $\pm$ 0.01	2.80 $\pm$ 0.01	2.80 $\pm$ 0.01	0.01	0.11 (Y)	0.01	0.03 (Y)

Table 5-30 (2017) Fitted values of μ (GeV) from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.

	μ_{data}	$\mu_{MC_{allwoQCD}}$	$\mu_{MC_{DY}}$	$1.0\sigma_{data+MC_{allwoQCD}}$	$ \Delta\mu_{data-MC_{allwoQCD}} $	$1.0\sigma_{MC_{allwoQCD}+MC_{DY}}$	$ \Delta\mu_{MC_{allwoQCD}-MC_{DY}} $
Event class 0	90.69 $\pm$ 0.20	90.33 $\pm$ 0.07	90.37 $\pm$ 0.07	0.21	0.36 (Y)	0.10	0.04 (N)
Event class 1	90.17 $\pm$ 0.05	89.77 $\pm$ 0.02	89.78 $\pm$ 0.02	0.05	0.40 (Y)	0.03	0.01 (N)
Event class 2	89.68 $\pm$ 0.04	89.34 $\pm$ 0.02	89.34 $\pm$ 0.02	0.04	0.34 (Y)	0.03	0.0 (N)
Event class VBF (replaced)	89.95 $\pm$ 0.18	89.44 $\pm$ 0.06	89.43 $\pm$ 0.06	0.40	0.51 (Y)	0.08	0.01 (N)
	$\mu_{MC_{DY}}$			$1.0\sigma_{MC_{DY}(default)+MC_{DY}(replaced)}$	$ \Delta\mu_{MC_{DY}(default)-MC_{DY}(replaced)} $		
Event class VBF (default)	-	-	90.57 $\pm$ 0.25	0.26	1.14 (Y)		

The central values of the mean μ and σ are taken from double-fake Drell-Yan MC events in each class. The contributions from the single-fake differences has to be applied to double-fake DY MC events to estimate total systematic. Assuming the two fakes are totally correlated with each other, the contribution is doubled of each single-fake difference (for a difference Δ , $\sqrt{\Delta^2 + \Delta^2 + 2 * \Delta * \Delta} = 2\Delta$). The final systematic error in

Table 5-31 (2017) Fitted values of σ from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.

	σ_{data}	$\sigma_{MC_{allwoQCD}}$	$\sigma_{MC_{DY}}$	$1.0\sigma_{data+MC_{allwoQCD}}$	$ \Delta\sigma_{data-MC_{allwoQCD}} $	$1.0\sigma_{MC_{allwoQCD}+MC_{DY}}$	$ \Delta\sigma_{MC_{allwoQCD}-MC_{DY}} $
Event class 0	2.49±0.22	2.31±0.07	2.37±0.06	0.23	0.18 (N)	0.09	0.06 (N)
Event class 1	2.41±0.06	2.43±0.02	2.48±0.02	0.06	0.02 (N)	0.03	0.05 (Y)
Event class 2	2.69±0.05	2.83±0.02	2.89±0.02	0.05	0.14 (Y)	0.03	0.06 (Y)
Event class VBF (replaced)	2.82±0.22	3.07±0.06	3.08±0.06	0.23	0.25 (Y)	0.08	0.01 (N)
				$\sigma_{MC_{DY}}$	$1.0\sigma_{MC_{DY}(default)+MC_{DY}(replaced)}$	$ \Delta\sigma_{MC_{DY}(default)-MC_{DY}(replaced)} $	
Event class VBF (default)	-	-	3.14±0.24	0.25	0.06 (N)		

Table 5-32 (2018) Fitted values of μ (GeV) from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.

	μ_{data}	$\mu_{MC_{allwoQCD}}$	$\mu_{MC_{DY}}$	$1.0\mu_{data+MC_{allwoQCD}}$	$ \Delta\mu_{data-MC_{allwoQCD}} $	$1.0\mu_{MC_{allwoQCD}+MC_{DY}}$	$ \Delta\mu_{MC_{allwoQCD}-MC_{DY}} $
Event class 0	90.04±0.03	89.88±0.04	89.91±0.04	0.05	0.16 (Y)	0.06	0.03 (N)
Event class 1	89.76±0.01	89.77±0.01	89.79±0.01	0.01	0.01 (N)	0.01	0.02 (Y)
Event class 2	89.02±0.01	88.94±0.01	88.95±0.01	0.01	0.08 (Y)	0.01	0.01 (N)
Event class VBF (default)	89.79±0.11	89.81±0.10	89.79±0.09	0.15	0.02 (N)	0.13	0.02 (N)
				$\mu_{MC_{DY}}$	$1.0\mu_{MC_{DY}(default)+MC_{DY}(replaced)}$	$ \Delta\mu_{MC_{DY}(default)-MC_{DY}(replaced)} $	
Event class VBF (replaced)	-	-	88.86±0.03	0.09	0.93 (Y)		

Table 5-33 (2018) Fitted values of σ from single-fake data, single-fake $MC_{allwoQCD}$ and single-fake MC_{DY} in all 4 event classes.

	σ_{data}	$\sigma_{MC_{allwoQCD}}$	$\sigma_{MC_{DY}}$	$1.0\sigma_{data+MC_{allwoQCD}}$	$ \Delta\sigma_{data-MC_{allwoQCD}} $	$1.0\sigma_{MC_{allwoQCD}+MC_{DY}}$	$ \Delta\sigma_{MC_{allwoQCD}-MC_{DY}} $
Event class 0	2.49±0.03	2.65±0.04	2.52±0.04	0.05	0.16 (Y)	0.06	0.13 (N)
Event class 1	1.75±0.01	1.82±0.01	1.82±0.01	0.01	0.07 (Y)	0.01	0.00 (N)
Event class 2	2.79±0.01	2.92±0.02	2.79±0.01	0.02	0.13 (Y)	0.02	0.13 (Y)
Event class VBF (default)	2.44±0.12	2.16±0.07	2.17±0.05	0.14	0.28 (Y)	0.09	0.01 (N)
				$\sigma_{MC_{DY}}$	$1.0\sigma_{MC_{DY}(default)+MC_{DY}(replaced)}$	$ \Delta\sigma_{MC_{DY}(default)-MC_{DY}(replaced)} $	
Event class VBF (replaced)	-	-	2.73±0.02	0.07	0.28 (Y)		

each category for each parameter, μ and σ , is the quadratic sum of all sources including the statistical error from the fit of the double-fake DY MC events. Dedicated systematics due to the replacement of real VBF selected events by the events passing only VBF preselection cuts for the VBF tagged event class is also considered.

The mean values of μ and σ are taken from the double-fake DY MC events and their total uncertainties calculated in each event class has been used in final statistical analysis, are summarized in Table 5-34, 5-35 (2016), Tables 5-36, 5-37 (2017) and Table 5-38, 5-39 (2018).

Table 5-34 (2016) Values of μ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$.

	μ (GeV)	$\Delta\mu_{stat}$ (GeV)	$2.0\Delta\mu_{data-MC}$ (GeV)	$2.0\Delta\mu_{MC-MC_{DY}}$ (GeV)	$\Delta\mu_{tot}$ (GeV)
Event class 0	89.66	0.27	0.20	-	0.34
Event class 1	89.79	0.20	0.18	-	0.27
Event class 2	88.60	0.09	0.40	0.06	0.41

Table 5-35 (2016) Values of σ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class.

	σ	$\Delta\sigma_{stat}$	$2.0\Delta\sigma_{data-MC_{allwoQCD}}$	$2.0\Delta\sigma_{MC_{allwoQCD}-MC_{DY}}$	$\Delta\sigma_{tot}$
Event class 0	1.59	0.27	0.22	-	0.35
Event class 1	3.04	0.17	0.16	0.06	0.24
Event class 2	3.27	0.10	0.22	0.06	0.25

Table 5-36 (2017) Values of μ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$.

	μ (GeV)	$\Delta\mu_{stat}$ (GeV)	$2.0\Delta\mu_{data-MC}$ (GeV)	$2.0\Delta\mu_{MC-MC_{DY}}$ (GeV)	$2.0\Delta\mu_{MC_{DY}(default)-MC_{DY}(replaced)}$ (GeV)	$\Delta\mu_{tot}$ (GeV)
Event class 0	91.67	0.84	0.72	-	-	1.11
Event class 1	89.96	0.30	0.80	-	-	0.85
Event class 2	89.53	0.33	0.64	-	-	0.76
Event class VBF	90.55	0.45	1.02	-	2.28	2.54

Table 5-37 (2017) Values of σ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$.

	σ	$\Delta\sigma_{stat}$	$2.0\Delta\sigma_{data-MC}$	$2.0\Delta\sigma_{MC-MC_{DY}}$	$2.0\Delta\sigma_{MC_{DY}(default)-MC_{DY}(replaced)}$	$\Delta\sigma_{tot}$
Event class 0	2.22	1.14	-	-	-	1.14
Event class 1	2.26	0.31	-	0.10	-	0.33
Event class 2	2.22	0.39	0.28	0.12	-	0.49
Event class VBF	2.33	0.47	0.50	-	-	0.69

Table 5-38 (2018) Values of μ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$.

2018	μ (GeV)	$\Delta\mu_{stat}$ (GeV)	$2.0\Delta\mu_{data-MC}$ (GeV)	$2.0\Delta\mu_{MC-MC_{DY}}$ (GeV)	$2.0\Delta\mu_{MC_{DY}(default)-MC_{DY}(replaced)}$ (GeV)	$\Delta\mu_{tot}$ (GeV)
Event class 0	88.50	0.29	0.32	-	-	0.43
Event class 1	88.96	0.26	-	0.04	-	0.26
Event class 2	88.91	0.19	0.16	-	-	0.25
Event class VBF	89.05	0.74	-	-	1.86	2.00

Table 5-39 (2018) Values of σ , statistical uncertainties from double-fake fits, uncertainties from single-fake fits and the total uncertainty in each event class. Here MC represents the above $MC_{allwoQCD}$.

2018	σ	$\Delta\sigma_{stat}$	$2.0\Delta\sigma_{data-MC_{allwoQCD}}$	$2.0\Delta\sigma_{MC_{allwoQCD}-MC_{DY}}$	$2.0\Delta\sigma_{MC_{DY}(default)-MC_{DY}(replaced)}$	$\Delta\sigma_{tot}$
Event class 0	1.65	0.32	0.32	0.26	-	0.52
Event class 1	2.14	0.29	0.14	-	-	0.32
Event class 2	2.90	0.30	0.26	0.26	-	0.47
Event class VBF	2.34	0.48	0.56	-	1.28	1.48

The DCB parameters α_L and α_R were optimized and fixed during the background modeling. The uncertainties on DCB parameters α_L and α_R are considered. The fits have been performed on the double-fake DY MC events with floating α_L and α_R . For each of the α_L and α_R corresponding to each event category, the central values are optimized and fixed, as summarized in Table 5-25 (2016), Table 5-26 (2017) and Table 5-27 (2018). The fits with α_L and α_R floating are used to obtain the statistical uncertainty. The difference of the fixed value and fitted value from the floating fit has been treated as systematic uncertainty. Finally, the quadratic sum of the statistical and systematic uncertainty is considered as the total uncertainty of fixed alpha. The fitted values of α_L and α_R are summarized in Table 5-40 for 2016, 2017 and 2018. Table 5-41 (2016), Table 5-42 (2017) and Table 5-43 (2018) list the central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R .

Finally, the studies in this chapter takes in account the statistical uncertainties on the values of the DCB parameters n_L and n_R , as well as those from the additional ex-

Table 5-40 Fit values of α_L and α_R while floating in the fitting of the double-fake DY events.

	2016		2017		2018	
class	α_L	α_R	α_L	α_R	α_L	α_R
0	1.04 $\pm$ 0.25	0.85 $\pm$ 0.28	1.27 $\pm$ 0.36	1.49 $\pm$ 0.92	1.67 $\pm$ 0.22	1.09 $\pm$ 0.39
1	0.73 $\pm$ 0.14	3.04 $\pm$ 0.18	1.44 $\pm$ 0.19	2.58 $\pm$ 0.62	0.92 $\pm$ 0.15	1.68 $\pm$ 0.15
2	0.64 $\pm$ 0.11	1.34 $\pm$ 0.16	0.90 $\pm$ 0.18	0.96 $\pm$ 0.11	0.57 $\pm$ 0.11	1.47 $\pm$ 0.21
VBF	-	-	1.78 $\pm$ 0.37	1.83 $\pm$ 0.76	1.04 $\pm$ 0.11	1.75 $\pm$ 0.38

Table 5-41 (2016) Central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R .

	α_L				α_R			
class	Central	Stat.	Syst.	Total Unc.	Central	Stat.	Syst.	Total Unc.
0	1.2	0.3	0.2	0.3	2.0	0.3	1.2	1.2
1	1.2	0.1	0.5	0.5	2.0	0.2	1.0	1.1
2	1.4	0.1	0.8	0.8	1.6	0.2	0.3	0.3

Table 5-42 (2017) Central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R .

	α_L				α_R			
class	Central	Stat.	Syst.	Total Unc.	Central	Stat.	Syst.	Total Unc.
0	0.7	0.4	0.6	0.7	2.5	0.9	1.0	1.4
1	1.0	0.2	0.4	0.5	2.0	0.6	0.6	0.8
2	0.9	0.2	0.0	0.2	1.8	0.1	0.8	0.8
VBF	1.1	0.4	0.7	0.8	2.2	0.8	0.4	0.8

ponential function exp_p1 and frac_sum, obtained from the double-fake fits as listed in Table 5-25 for 2016, Table 5-26 for 2017 and Table 5-27 (2018).

The relic DY events have a similar peak shape as that of the signal. Significant bias observed due to the relic DY, when the signal mass is close to the relic DY fitted peak in 2016 class 0, in 2017 for all classes and in 2018 class 0. The maximum bias, is around 0.6 in 2016 class 0, for all the classes in 2017 and in 2018 class 0. In order to deal with

Table 5-43 (2018) Central values, statistical uncertainties, systematic uncertainties and the total uncertainties of α_L and α_R .

class	α_L				α_R			
	Central	Stat.	Syst.	Total Unc.	Central	Stat.	Syst.	Total Unc.
0	1.2	0.2	0.5	0.5	1.7	0.4	0.6	0.7
1	0.7	0.2	0.2	0.3	0.9	0.2	0.8	0.8
2	0.9	0.1	0.3	0.3	1.7	0.2	0.2	0.3
VBF	2.3	0.1	1.3	1.3	1.0	0.4	0.7	0.8

this, different approaches were tried and the so-called ”inflated “f” uncertainty method” is considered to reduce the bias to acceptable values as well as having a small impact on the expected uncertainty of signal strength. This method inflates the normalization uncertainties of the relic DY background. In 2016 class 0, the inflated “f” uncertainty is 0.02 which is around 4 times the original uncertainty (~ 0.005) for each selected analytic function in this event class. With this method, the maximum bias in 2016 class 0 reduced to less than 0.2. In 2017 data set, the best choice is to employ 4 times the original uncertainty on “f” in event class 0 and event class 1, while to employ 6 times the original uncertainty on “f” in event class 2 and the VBF tagged event class. The original uncertainties are between ~ 0.00001 to 0.0531 in 2017. With this method, maximum bias reduced to around 0.3 to 0.4, depending on the different 2017 event classes. In 2018 class 0, the uncertainty is 4 times the original uncertainty and the maximum bias in 2018 class 0 is reduced to be less than about 0.2. As these fractions “f” with inflated uncertainties are treated as Gaussian constrained nuisance parameters and included as “param” nuisances in the datacards for statistical analysis, they will change the impact ranking of the nuisance parameters on the signal strength. The unblinded impacts for the full Run 2 combination are shown in Figure 5-90 and Figure 5-91.

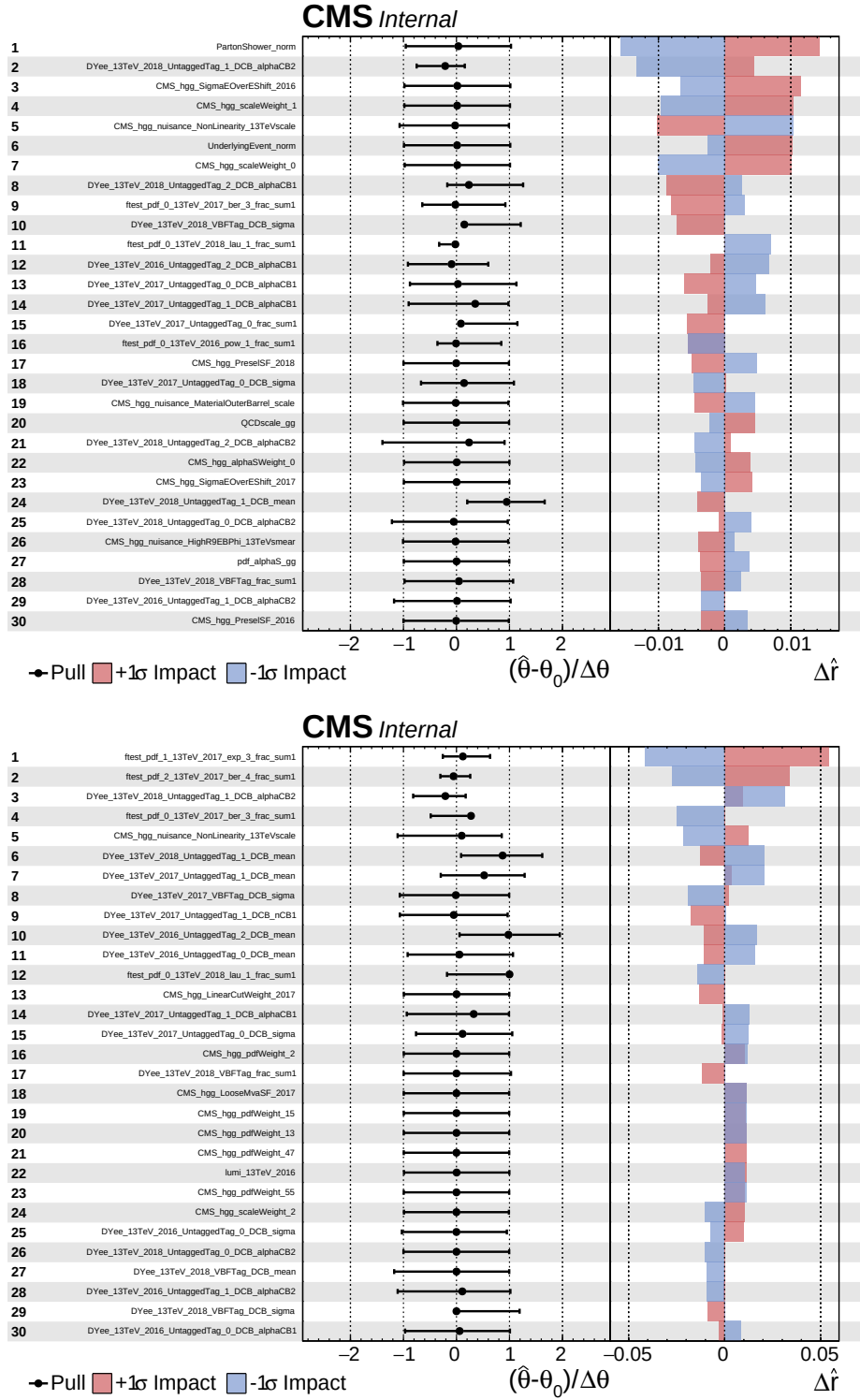


Figure 5-90 Full Run 2 unblinded impact plots of NP on signal strength for 70 GeV (top) and 90 GeV (bottom) signal mass, where θ_0 is the NP before the fit and θ is the NP after the fit. The red band shows the impact on the signal strength when θ reaches its $+1\sigma$ while the blue one shows the impact when θ reaches its -1σ value.

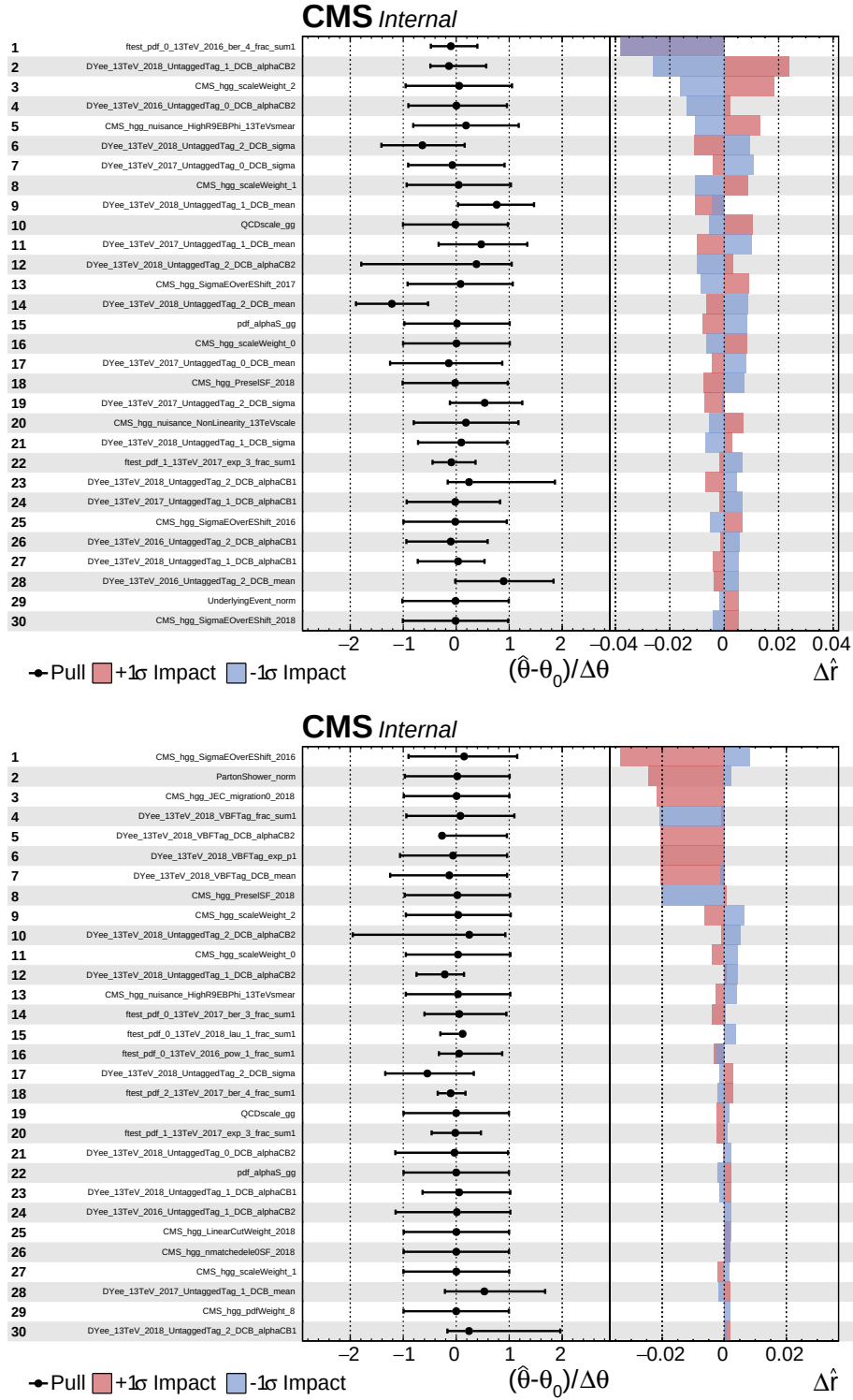


Figure 5-91 Full Run 2 unblinded impact plots of NP on signal strength for 95 GeV (top) and 110 GeV (bottom) signal mass, where θ_0 is the NP before the fit and θ is the NP after the fit. The red band shows the impact on the signal strength when θ reaches its $+1\sigma$ while the blue one shows the impact when θ reaches its -1σ value.

5.15 Results

The final results (limits and p-value) of the study carried out in this chapter are obtained by following the same method as used by SM $H \rightarrow \gamma\gamma$ analysis and are from [115].

The expected number of signal events are listed in Table 5-44 which corresponds to the production of an additional SM-like Higgs boson with $m_H = 90$ GeV from 2016, 2017 and 2018 datasets. The total number is broken down into the contributions from all the production processes in each of the event classes. Also shown the σ_{eff} and σ_{HM} values along with the number of background events per GeV estimated from the background-only fit to the data. The number from the DY process, in the corresponding σ_{eff} window centered on $m_H = 90$ GeV, using the chosen background function is also shown. Figure 5-92 displays the same information visually, with the additional plot of the number of signal events divided by the total number of signal plus background events.

Table 5-44 The expected number of SM-like Higgs boson signal and background events ($m_H = 90$ GeV) per event class and the corresponding percentage breakdown per production process.

Event classes		Expected SM-like Higgs boson signal yield ($m_H = 90$ GeV)								Bkg. (GeV ⁻¹)	DY Bkg. (GeV ⁻¹)
		Total	ggH (%)	VBF (%)	WH (%)	ZH (%)	$t\bar{t}H$ (%)	σ_{eff} (GeV)	σ_{HM} (GeV)		
2016 36.3 fb ⁻¹	0	130	71.9	15.6	6.2	3.6	2.6	1.12	1.00	271	12
	1	304	87.4	6.6	3.6	2.1	0.3	1.25	1.07	3093	33
	2	407	94.7	2.5	1.7	1.0	0.1	1.87	1.51	9190	193
	Total	842	88.5	6.0	3.1	1.8	0.6	1.50	1.20	12 554	239
2017 41.5 fb ⁻¹	0	104	73.4	11.6	7.5	4.3	3.2	1.27	1.13	248	7
	1	347	88.5	5.6	3.5	2.1	0.3	1.40	1.24	3625	83
	2	413	94.4	2.6	1.9	1.1	0.1	1.91	1.64	8169	244
	VBF	26	45.6	51.8	1.0	0.5	1.0	1.33	1.15	29	1
	Total	890	88.2	6.2	3.1	1.8	0.6	1.60	1.35	12 071	338
2018 54.4 fb ⁻¹	0	162	75.1	10.2	7.3	4.3	3.0	1.21	1.05	430	3
	1	585	90.1	4.8	3.1	1.8	0.2	1.34	1.17	6445	378
	2	473	94.4	2.5	1.9	1.2	0.1	2.01	1.73	10 982	720
	VBF	38	45.4	51.9	1.1	0.6	1.0	1.21	1.03	46	1
	Total	1258	88.4	6.1	3.1	1.8	0.6	1.54	1.27	17 902	1104

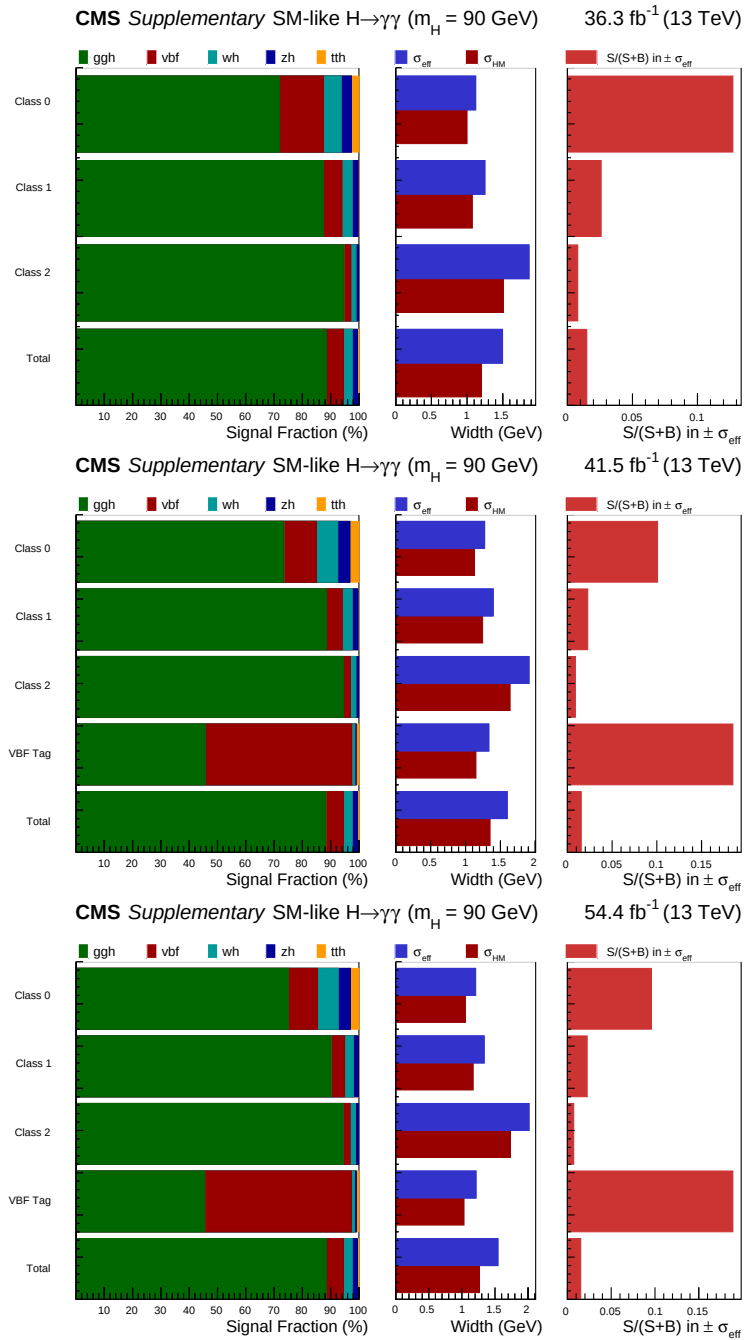


Figure 5-92 The expected number of signal events per event class and the percentage breakdown per production mode in 2016 (top) 2017 (middle) and 2018 (bottom) analysis. The effective σ is also provided as an estimate of the $m_{\gamma\gamma}$ resolution in that category and compared directly to the full-width at half maximum. The ratio of the number of signal events (S) to the number of signal plus background events (S+B) is shown on the right hand side.

The expected and observed 95% confidence level upper limits on the product of the production cross-section times branching ratio into two photons, both in the case relative to the SM-like BSM expectation from the latest LHCHGXSWG [60] and in the unnormalized case, are presented in Figure 5-93 for 2016, Figure 5-94 for 2017 and Figure 5-95 for 2018. A simultaneous binned maximum likelihood fit is performed to the distributions of diphoton invariant mass in all event classes, with a step size of 0.1 GeV within the range $70 < m_{\gamma\gamma} < 110$ GeV by using an asymptotic approach [30, 116, 117]. The related theoretical prediction for an additional SM-like Higgs boson is shown as a solid hatched band line indicating its uncertainty [30]. In 2016 analysis, one excess with approximately 3.3σ local (2.1σ global) significance is observed for an hypothesis mass of 95.7 GeV and no significant excess with respect to the expected limits is observed for 2017 and 2018. The minimum (maximum) observed upper limit on the production cross section times branching ratio is approximately 30 (154) fb corresponding to a mass hypothesis of 87.2 (95.7) GeV in 2016, approximately 34 (159) fb corresponding to a mass hypothesis of 108.8 (70) GeV in 2017 and approximately 24 (94) fb corresponding to a mass hypothesis of 110 (74.6) GeV in 2018.

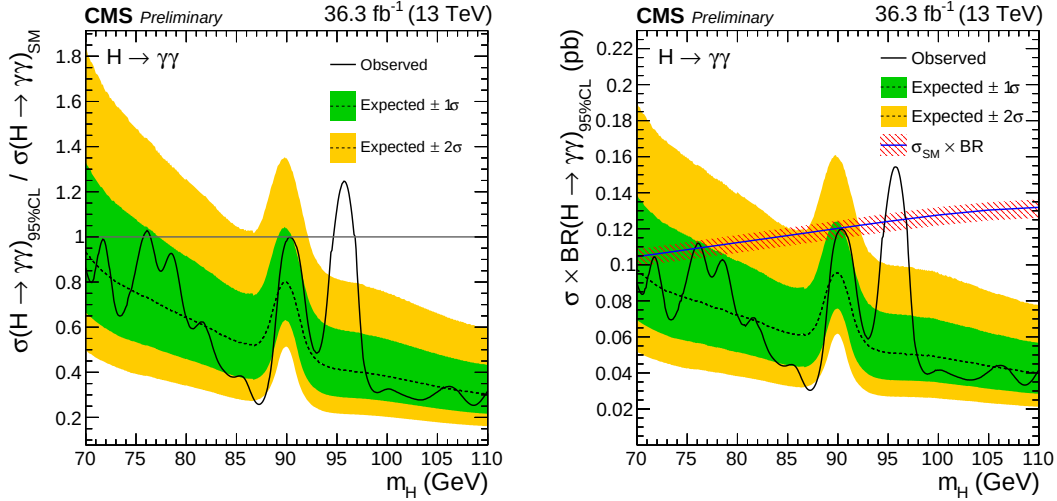


Figure 5-93 (2016) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation.

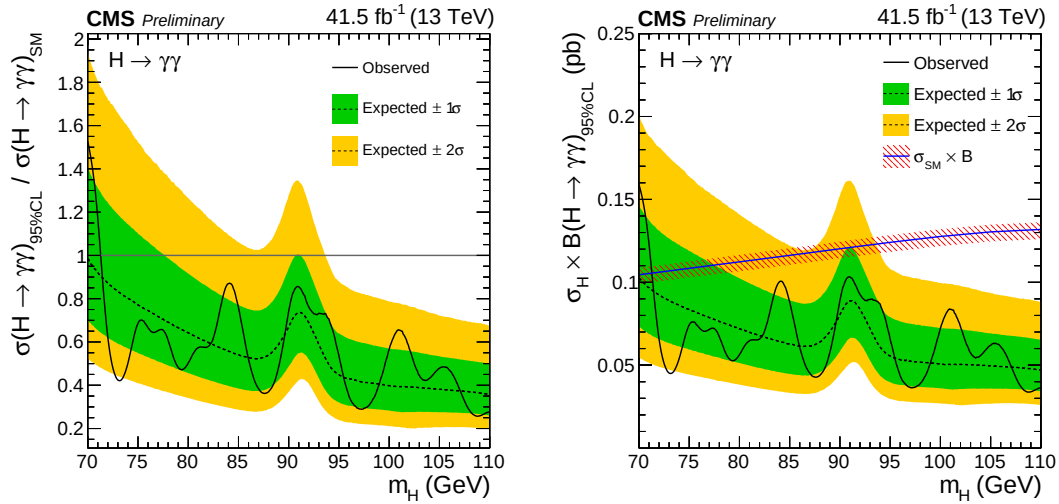


Figure 5-94 (2017) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation.

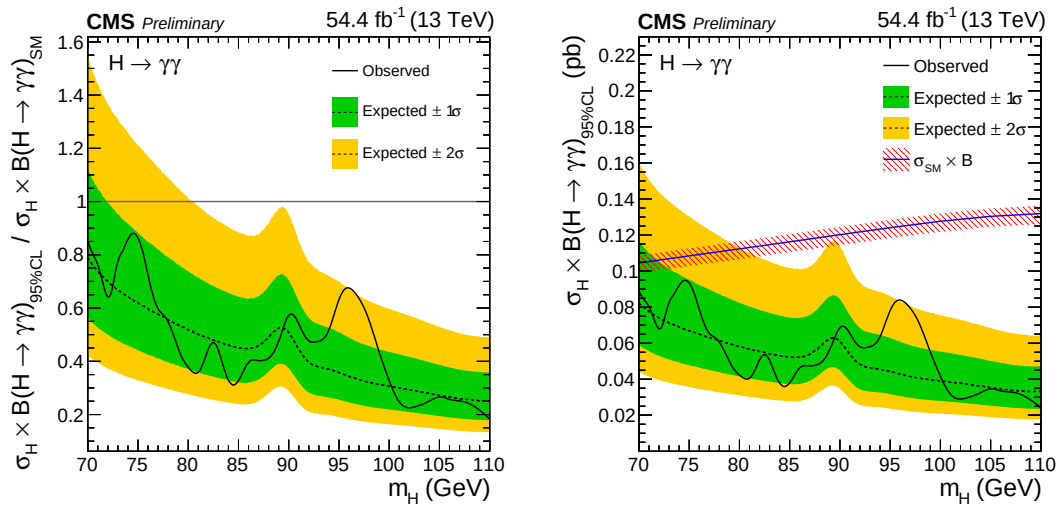


Figure 5-95 (2018) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation.

5.15.1 Full Run 2 combination

The results from 2016, 2017 and 2018 datasets are combined statistically at $\sqrt{s} = 13$ TeV exploiting the same methods which are used to obtain the results from each individual dataset, in the mass range $70 < m_{\gamma\gamma} < 110$ GeV. Figure 5-96 address the expected and observed 95% confidence level upper limits on the production cross-section times branching ratio into two photons in both cases i.e. relative to the SM-like BSM expectation from the latest LHCHGXS WG predictions [30] and in the unnormalized case. The minimum (maximum) observed limit on the production cross section times branching ratio is approximately 15 (73) fb relative to a mass hypothesis of 108.9 (95.4) GeV. Figure 5-97 shows the expected and observed 95% confidence level upper limits on the

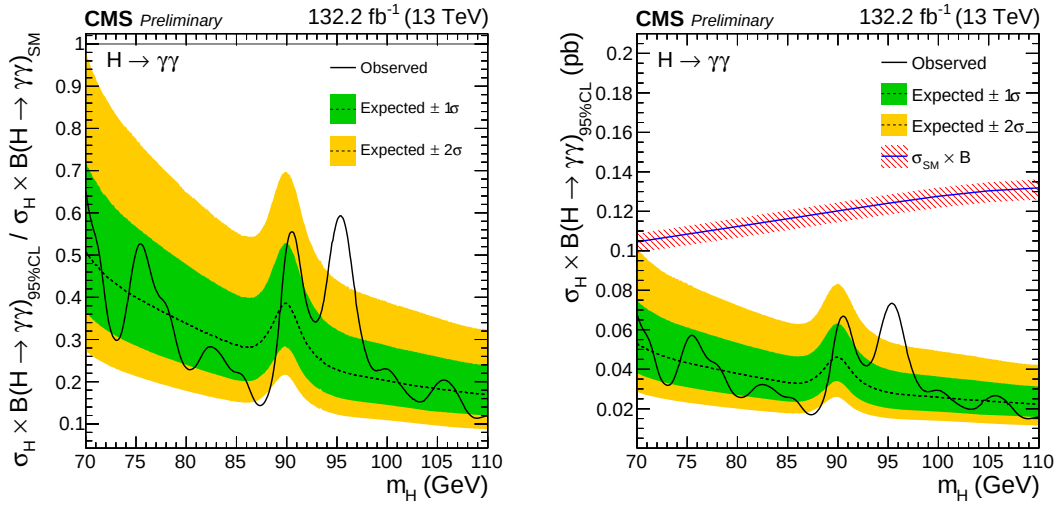


Figure 5-96 (Run 2) Expected and observed exclusion limits (95% CL) on the production cross-section times branching ratio into two photons (left) in the case relative to the expected SM-like BSM cross-section and (right) in the unnormalized case, in the asymptotic CLs approximation.

production cross-section times branching ratio into two photons, for the ggH plus $t\bar{t}H$ and VBF plus VH processes, the VBF only production hypothesis and the VH only production hypothesis from the statistical combination of 2016, 2017 and 2018 datasets.

Figure 5-98 shows the observed local p-values as calculated with respect to the background-only hypothesis using the ratio of a profile likelihood. In the full Run 2 results, one excess approximately 2.9σ local (1.3σ global) significance is observed for

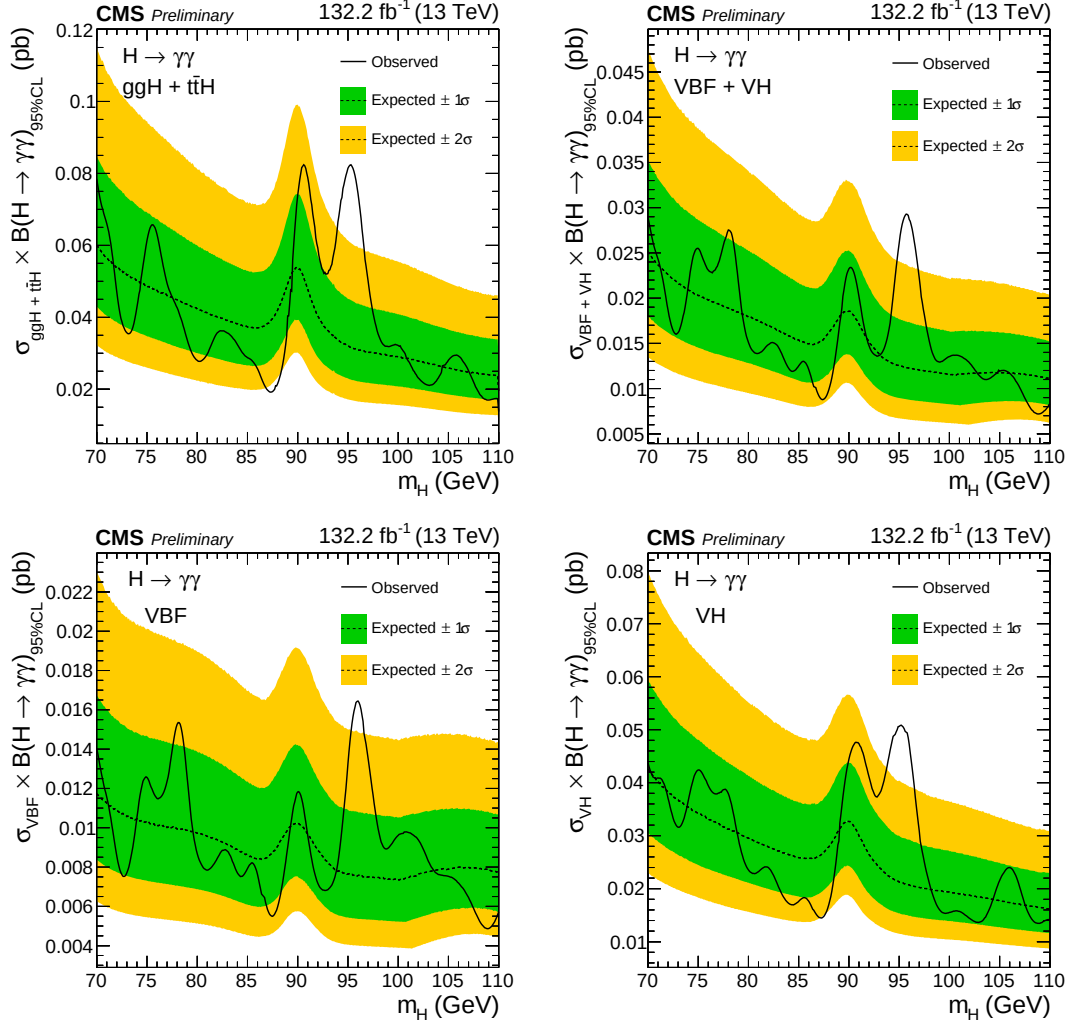


Figure 5-97 (Run 2) Expected and observed exclusion limits (95% CL, in the asymptotic approximation) for the ggH plus $t\bar{t}H$ (top left) and VBF plus VH (top right) processes, 100% VBF production hypothesis (bottom left) and the VH production hypothesis (bottom right). The inner and outer bands indicate the regions containing the distribution of limits located within ± 1 and $\pm 2\sigma$ of the expectation under the background-only hypothesis.

a mass hypothesis of 95.4 GeV where the global significance has been calculated using the method explained in Ref. [118]. In Figure 5-99 and 5-100, there are channel compatibilities for the full Run2 combinations. It has been estimated by fixing the mass value to the maximum significance mass point i.e. 95.4 GeV, corresponds to the full Run2 combination. The horizontal bars indicate 1σ uncertainties in the values, the vertical line and band indicate the value of combined “ μ ” in the fit to data and its uncertainty. Signal strengths at 95.4 GeV are compatible among 2016, 2017, 2018 (with χ^2 probability

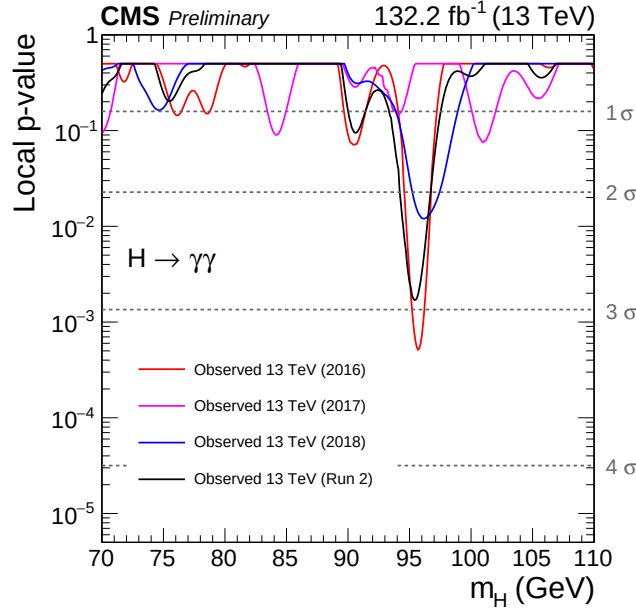


Figure 5-98 (Run 2) Observed local p -values for an additional SM-like Higgs boson as a function of m_H from 2016, 2017, 2018 and the full Run 2 combination.

value of 0.06) and for all the event classes (with χ^2 probability value of 0.68) in 2016, 2017 and 2018 data sets. In the Figure 5-101, $m_{\gamma\gamma}$ distributions are shown for all events

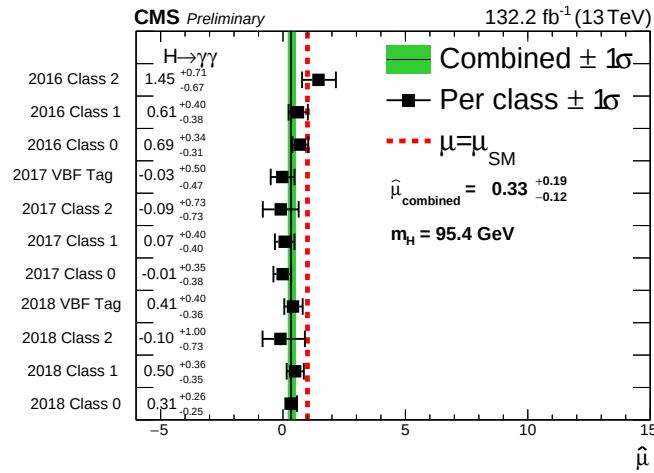


Figure 5-99 Full Run 2 channel compatibility per event class with the categories belong to each corresponding year.

in all classes, in combined 13 TeV data set along with the result of a fit of the signal-plus-background model under a mass hypothesis of 95.4 GeV. In the right plot, each event is weighted according to ratio of the numbers of signal to signal-plus-background

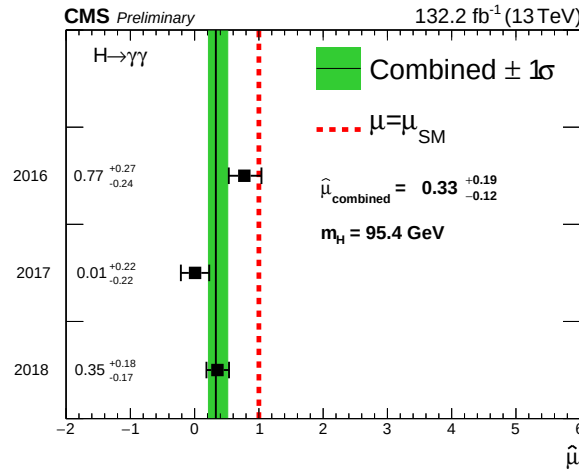


Figure 5-100 Full Run 2 channel compatibility per year.

events ($S/(S+B)$) for its event class, where S and B are the numbers of expected signal and background events.

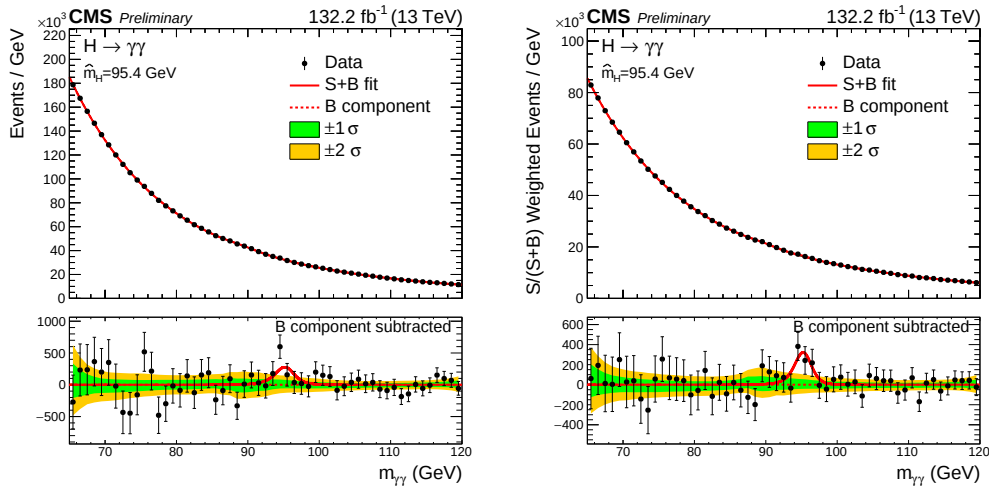


Figure 5-101 Events in all classes of the combined 13 TeV dataset together with the fit of signal-plus-background model, under a mass hypothesis of 95.4 GeV. The left plot is unweighted and in right plot, each event is weighted by the ratio $S/(S+B)$ for its event class, where S and B are the numbers of expected signal and background events.

5.16 Conclusion

The results for the search of a new low-mass resonances in the diphoton decay channel has been reported. The presented study is based on the p-p collisions at $\sqrt{s} = 13 \text{ TeV}$ (center-of-mass energy) collected by the CMS detector at LHC in 2016 (36.3 fb^{-1}),

2017 (41.5 fb^{-1}) and 2018 (54.4 fb^{-1}). The studies are performed in a mass range between 70 and 110 GeV. The expected and observed 95% confidence level upper limits on the product of the cross-section times branching ratio into two photons for both, absolute and normalized to the prediction for an SM-like BSM Higgs boson are shown. The statistical combination of expected and the observed results of upper limit and local p-value from the studies of 2016, 2017 and 2018 datasets are also reported. No significant excess with respect to the expected limits is observed. The maximum observed local significance is approximately 2.9σ for a mass hypothesis of 95.4 GeV, with a global significance of 1.3σ . The observed upper limit on the production cross section times branching ratio into diphoton, ranges from approximately 73 fb at a mass of 95.4 GeV to 15 fb at a mass of 108.9 GeV.

Chapter 6: Higgs properties measurement in the diphoton decay channel

6.17 Introduction

The studies presented in this chapter are mainly about the Higgs boson property measurements which includes, the signal strength modifiers, production cross sections, the Higgs boson's couplings to other particles by using the STXS framework [25]. It also addresses the results of inclusive and differential fiducial cross section measurements [26].

The measurements shown here in this chapter for both of the analyses has been carried out using full Run 2 dataset with the corresponding integrated luminosity of 137 fb^{-1} . The luminosities for 2016, 2017 and 2018 data are 35.9 fb^{-1} , 41.5 fb^{-1} , and 59.4 fb^{-1} respectively. The simulated signal samples are generated considering all of the major Higgs boson production modes. The major background for both of the studies addressed is originating from SM diphoton events and the $\gamma + jets$ or $jet + jet$ events contribute to the minor part of background. Drell-Yan events having final states from leptons are also taken into consideration.

The signal extraction strategy as well as the techniques for the modeling of signal and background are identical for both of analyses. The modeling of signal is done in a parametric way, separately for each observable. The $m_{\gamma\gamma}$ signal shape is simulated at the level of reconstruction which includes the efficiency corrections. A sum of up to four Gaussian pdfs used for fit in each of the bin particle level and reconstruction categories. The right and wrong vertex shapes are fitted separately and combined together in order to have a full signal pdf per bin. The discrete profiling method has been utilized for the background modelling. As the accurate configuration of background is unknown, so various functions are employed to fit the smoothly falling $m_{\gamma\gamma}$ shape and their choice is considered as a nuisance parameter in the fit (maximum likelihood) used to measure the cross sections.

6.18 Signal strengths and STXS with full Run 2 dataset at 13 TeV

6.18.1 The STXS framework

This analysis is mainly based upon the categories to enable measurements within the simplified template cross section (STXS) framework. This framework give an approach to perform measurements of Higgs boson couplings, reducing the theoretical uncertainties and at the same time, permit the use of advanced experimental methods. These techniques were firstly introduced in Ref. [27] followed by subsequent new versions developed according to the experimental results [28]. This framework help to enhance the reinterpretation of the measurement and intends to supersede the traditional measurement with the signal strength modifier.

In STXS framework, the kinematic regions motivated by theory are defined, based upon the Higgs boson production modes. These simplified bins does exist in the changing granularity followed by sequential stages. With the help of increased granularity of the signal bins aid to provide more information for various theory based explanations of the measurements. As a result, the sensitivity is enhanced for the possible BSM signatures. The integrated luminosity of the dataset has been increasing by the LHC resulting an enhanced experimental sensitivity of the measurements could lead to the later stages in the STXS framework.

At the stage 0 of STXS framework, the bins are closely related to the various Higgs boson production modes. In order to reduce the theoretical uncertainty, another requirement on the Higgs boson rapidity ($|y_H| < 2.5$) is placed. Without this, while extrapolation of the measurements to the full phase space, a big part might not be experimentally accessible. With $|y_H| > 2.5$, the experimental acceptance of $H \rightarrow \gamma\gamma$ events is almost negligible but still taken into account. Results with the stage 0 cross sections in the diphoton decay mode were reported in the previous analysis (HIG-16-040) [106].

In the stage 1 of the STXS, bins are further splitted based on various kinematic regions. The stage 1 cross section observations for gluon-gluon fusion (ggH) and vector boson fusion (VBF) were addressed in the prior version of this analysis in HIG-18-029 [29]. Now the updated definitions have become available for stage 1.1 and stage 1.2. The updated stage 1.2 bins is one of the differences between the current studies and the ones reported previously.

In the latest results which corresponds to HIG-19-015 and were published by JHEP

(JHEP07(2021)027) [25], the key differences and enhancements w.r.t. the previous results are:

- The availability of 2018 data increased the integrated luminosity from 77.4 fb^{-1} to 137 fb^{-1} resulting an improvement of about 33% to the most of the measurements which were statistically limited before.
- An enhanced measurement of ggH production is made where a multi-class BDT has exploited to target stage 1 bin in contrary to just using the reconstruction level diphoton p_T and total number of jets. This improved the categorisation effectively and mimic the correlations of ggH parameters in the final statistical results.
- For the VBF production, an optimized approach is made by improving the capability of constraining ggH production with the topology of final state is identical to that of typical VBF events.
- The analysis categories are added in order to target associated production (VH) with the hadronic decays of vector bosons as well as the leptonic decays of vector bosons.
- The tth (top quark-antiquark pair) production categories are also incorporated which were also used before in HIG-19-013 [119].
- A dedicated new category is also added for the first time, for the associated production of single top quark (tH).
- Inclusive as well as per-production mode signal strengths are included along with Higgs boson coupling measurements using κ -framework and additional 2D scans.

This chapter mainly talks about the analysis strategy, categorisation procedure leading towards the final results of Higgs boson properties measurements. The photon identification BDT has been improved significantly as compared to the previous $H \rightarrow \gamma\gamma$ analyses. The chained quantile regression method is used to correct the MC inputs of this BDT in order make the simulation agree with data well. The sequentially corrections are made for the input variables using their cumulative distribution functions to maintain the variables correlations. As a result, the systematic uncertainty belongs to the data/MC differences got reduced for the output score of photon ID BDT. All of the optimizations are done simultaneously three years i.e. 2016, 2017 and 2018 which changes the diphoton mass resolution resulting an enhanced inclusive sensitivity of approximately 5%. The structure of this chapter is as, it briefly describe the analysis strategy, categorisation procedures which covers the ggH, electroweak (EW) qqH,

leptonic VH and ttH production modes followed by the final results and interpretation.

6.18.2 Analysis strategy

In order to correctly define the STXS bins, categorization of the analysis is constructed in a way that each category can be identified separately. This goal has been achieved by reconstructing the detector level similar particle quantities used for the bins definitions. In addition to this, several machine learning (ML) algorithms including boosted decision trees (BDTs) and deep neural networks (DNNs), are used to increase as much as possible the signal to background ratio and eventually increase the analysis sensitivity. By defining the granular categories, a good sensitivity is achieved for more inclusive quantities like the overall signal strength as well as per-production mode signal strengths.

The reconstruction level categorization of events is designed in a way to aim all the bins where a good sensitivity can be achieved in the diphoton decay channel with a prescribed dataset. All of the bins are measured without merging, for the ggH phase space. It is very hard for the high dijet mass (m_{jj}) bins to separate from VBF, even though, it is tried to constrain them collectively. Additionally, very high Higgs boson p_T bins has very low cross section resulting the limited sensitivity. It is rather possible to constrain the main bins for qqH production mode including the VH hadronic ones, the bins with VBF topology for high and low m_{jj} along-with the BSM bin. For the VH events in the leptonic decays of vector boson, the diphoton decay mode has very few events to measure precisely the bins individually. To accomplish this jobs, dedicated studies carried out in order to finely define the analysis categories so that by not effecting the sensitivity to VH leptonic mode. The same studies are performed to determine a precise splitting of Higgs boson p_T for ttH production and a dedicated DNN (Deep Neural Network) referred to as the top DNN is trained to discriminate ttH from tH production.

The analysis categories can be defined separately for each year or by having all of the three years combined together. It is then felt preferred that, if the expected number of events is small, the merging is done for all three years and where it has enough statistics, treated them separately. The merging is then done for in the ttH phase space and at the same time, information of the difference in different years is lost. Its also relevant for the ggH phase space which generally posses high statistics but still, the sensitivity is diphoton mass resolution dependent which is not same among years. Based

on the studies done for both scenarios, it turned out that even for the ggH categories, the sensitivity is higher with the merging of all years. As a result, the MVA's used for the categorization of each intended production mode are trained by combining all of three years. The final categories of the analysis are based on the data from the merging of three years. The MC events are separately generated for each year according to the corresponding detector conditions and they are merged. This is needed in order to include the variation in the detector, event reconstruction and in the LHC beam parameters. In addition to this, the photon energy scale corrections and various procedures related to the event reconstruction are carried out individually for each year. While performing the final categorization of diphoton events, the data and MC events from each year are processed together.

Finally, in order to measure the cross sections along with the uncertainties, a simultaneous fit is performed over all of the analysis. Additionally, signal strengths w.r.t. to the SM expectation are addressed for inclusive and per-production mode. Measurements of various parameters in the κ -framework are also presented. A brief description of different analysis regions, classifiers used and the targeted STXS stage 1.2 bins can be seen in Table 6-45

Table 6-45 Description and the method used for defining the categories of events.

Categorisation region	STXS assignment	Classifiers	Particle-level STXS bin, (units in GeV)	Number of categories
tHq leptonic	-	tHq BDT, top DNN	tHq	1
ttH leptonic	$p_T^{\gamma\gamma}$	ttH lep BDT, top DNN	$p_T^H < 60$	4
			$60 < p_T^H < 120$	2
			$120 < p_T^H < 200$	2
			all $p_T^H > 200$ (2 total)	2
ZH leptonic	-	ZH lep BDT	all qqZH and ggZH bins (10 total)	2
WH leptonic	$p_T^{\gamma\gamma}$	WH lep BDT	< 75	3
			all > 75 (4 total)	3
VH E_T^{Miss}	-	VH MET BDT	all VH leptonic bins (15 total)	2
ttH hadronic	$p_T^{\gamma\gamma}$	ttH had BDT, top DNN	$p_T^H < 60$	4
			$60 < p_T^H < 120$	4
			$120 < p_T^H < 200$	4
			all $p_T^H > 200$ (2 total)	4
VBF	$p_T^{\gamma\gamma}, m_{jj}, p_T^{Hjj}$	dijet BDT, diphoton BDT	$350 < m_{jj} < 700, p_T^{Hjj} < 25$	2
			$350 < m_{jj} < 700, p_T^{Hjj} > 25$	2
			$m_{jj} > 700, p_T^{Hjj} < 25$	2
			$m_{jj} > 700, p_T^{Hjj} > 25$	2
			$p_T^{Hjj} > 200$	2
			all ggH $m_{jj} > 350$ (4 total)	2
VH hadronic	-	VH had BDT, diphoton BDT	$350 < m_{jj} < 700$	2
ggH	ggH BDT, $p_T^{\gamma\gamma}$	diphoton BDT	0J $p_T^H < 10$	3
			0J $p_T^H > 10$	3
			1J $p_T^H < 60$	3
			1J $60 p_T^H < 120$	3
			1J $120 p_T^H < 200$	3
			2J $p_T^H < 60$	3
			2J $60 p_T^H < 120$	3
			2J $120 p_T^H < 200$	3
			$200 < p_T^H < 300$	2
			$300 < p_T^H < 450$	2
			$450 < p_T^H < 650$	1
			$p_T^H > 650$	1
No categories	-	-	qqH 0J, 1J, $m_{jj} < 60$, qqH $120 < m_{jj} < 350$, bbH, tHW	0

6.18.3 Events categorisation

The selection of event in all of the analysis categories needed two leading pre-selected photon candidates to have $p_T^{\gamma 1} > m_{\gamma\gamma}/3$ and $p_T^{\gamma 2} > m_{\gamma\gamma}/4$ respectively with mass range $100 < m_{\gamma\gamma} < 180$ GeV. The scaled photon p_T requirements mimic the distortions at lower end of the $m_{\gamma\gamma}$ spectrum. The categorization of the events for the analysis is made where the signal peak is recognizable from the smoothly falling background of $m_{\gamma\gamma}$ spectrum. The properties diphoton system are used in categorization procedure to optimize the analysis sensitivity. The categorization procedure utilize dedicated criteria and classifiers to select the events consistent with the tH , $t\bar{t}H$, VH , VBF , and ggH production mechanism. As a result, enhanced analysis sensitivity is achieved and measurements are performed for cross sections of individual production mode.

At the stage 1 of STXS framework, bins were splitted using the kinematic properties of events [28]. In this way, the sensitivity of analysis can be enhanced for possible BSM physics signatures. Moreover, the theory dependence of the measurement can be reduced by having increased number of independent bins.

ATLAS collaboration have already been reported measurements at stage 1 of the framework [68–70]. By following these measurements, the enhancements in the framework were made in a way that, the most recent STXS bins definition is stated as STXS stage 1.2. The CMS Collaboration has reported the first measurement of STXS stage 1.2 cross sections [71]. A full set of STXS stage 1.2 bins are illustrated in Figure 6-102.

Categories targeting ggH production

At stage 1.2 of the STXS framework, the gluon fusion process (ggH) is divided into a total of sixteen generator-level bins, as shown in blue region of Figure 6-102. The events are split first by the transverse momentum of the Higgs boson, p_T^H . Events with $p_T^H > 200$ GeV enter a dedicated BSM region, where bins are defined by more granular splits in p_T^H . Four BSM bins are defined by p_T^H boundaries placed at 300, 450, and 650 GeV, the final bin of which extends to arbitrarily high p_T^H . The remaining events are then split initially by the number of jets. At generator-level, jets are defined as anti- k_T (with radius parameter 0.4) jets, built from all stable particles except the products of the Higgs and any associated vector bosons. An additional requirement that $p_T > 30$ GeV is also imposed, to reduce the impact of pileup. There are zero (0J), one (1J), and greater than or equal to two jet (GE2J) bins. In the case where there are no jets present,

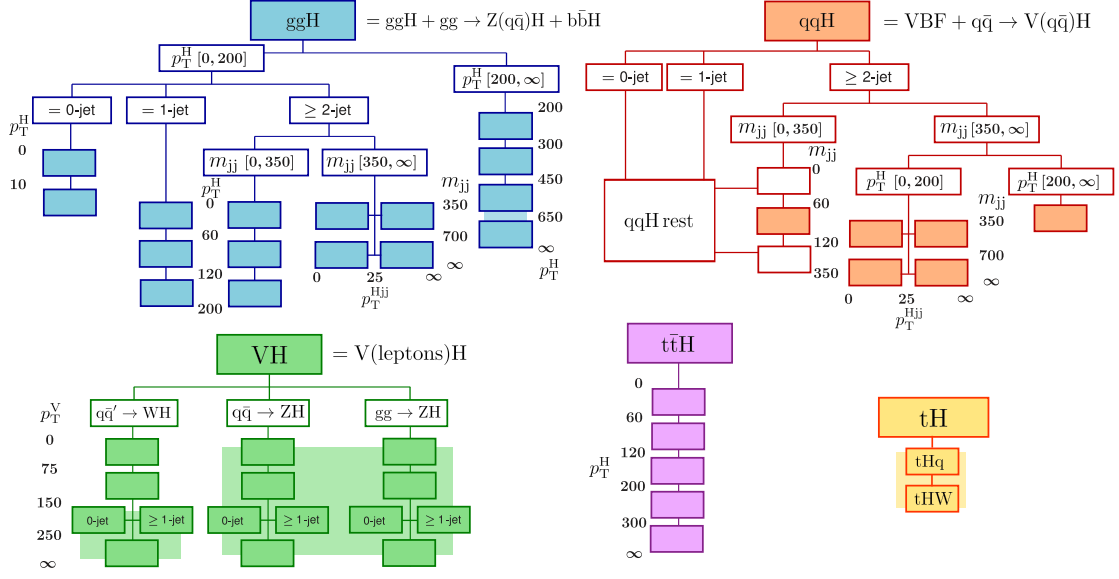


Figure 6-102 Full layout of STXS stage 1.2 bins.

two bins are defined using p_T^H boundaries placed at 10 GeV and 200 GeV. We refer to these bins as 0J low and 0J medium in this note. When one jet is present, boundaries are defined by cuts on p_T^H placed at 60, 120, and 200 GeV, named 1J low, 1J medium and 1J high.

For events with two jets or more, a further split using m_{jj} is performed, with a boundary being placed at 350 GeV. Events falling in the lower m_{jj} bin are split by p_T^H boundaries identical to the one jet case. Events in the upper m_{jj} bin form a separate region dedicated to the VBF-like phase space. This region is split into a two-jet-like ($p_T^H < 25$ GeV) and three-jet-like ($p_T^H > 25$ GeV) bin. Finally these bins are divided into a low and high m_{jj} region, by a boundary placed at 700 GeV.

For the ggH process, dedicated categories targeting each of the stage 1.2 bins. Events are assigned to a category based on the reconstructed equivalent of the generator-level quantities. The number of jets is defined by the number of anti- k_T jets with a measured $p_T > 30$ GeV, and the p_T^H corresponds to the measured p_T of the diphoton system ($p_T^{\gamma\gamma}$). Similarly, the p_T^H has a reconstruction-level analogue in the measured pt of the diphoton plus dijet system. Finally, the m_{jj} corresponds to the invariant mass of reconstructed dijet system.

In general, the $p_T^{\gamma\gamma}$ is well measured entity and the migration between bins is very small (approximately 1-5%). There is however greater migration between the reconstructed jet bins, up to around 30%. Therefore, a sophisticated categorisation techniques

such as classification using multi-class BDT have been studied, in order to make possible improvement in category purity. The analysis sensitivity is further maximized by splitting of events based on the expected signal-to-background ratio (S/B). The diphoton BDT used for this purpose to discriminate the signal from background processes. The uncertainty in the ggH production of events with a VBF-like dijet system is considered by two sources related to the prediction in the two-jet-like and three-jet-like bins. The total magnitude vary from around 5% to 30%, with the events having one or more jets and high values of p_T^H .

Categories targeting qqH production

The categories targeting EW qqH production are splitted into two groups. The first targets the VBF like phase space, using a multi-class dijet BDT to separate true VBF, VBF-like ggH, and non-Higgs backgrounds with dijets. The second targets hadronic VH production, using a VH had BDT trained against both ggH and non-Higgs backgrounds. Both groups additionally use the diphoton BDT to further reject background.

At stage 1.2 of the STXS framework, the EW qqH production is split into six generator-level bins. The number of jets defines three regions: 0-jet, 1-jet and ≥ 2 -jet. The latter region is further divided into two m_{jj} regions of 0 to 350 GeV and ≥ 350 GeV. The ≥ 350 GeV region is further splitted by p_T^H into two bins of 0 to 200 GeV and ≥ 200 GeV respectively, as shown in orange color in Figure 6-102. There is a VH hadronic bin which requires that a dijet is present with $60 < m_{jj} < 120$ GeV. The same approach is used by employing a dedicated BDT to separate VBF from ggH and non-Higgs background. A multi-class BDT is used instead of the as described in Ref. [29, 105]. Categories are constructed using cuts on the multi-class BDT output, together with cuts on the diphoton BDT. In addition, categories are split by p_T^{Hjj} and m_{jj} in order to provide some sensitivity to the different stage 1.2 processes. A data-driven method is used to model background events with non-prompt photons in the multi-class BDT training, replacing the MC. The method is same as used in Ref. [29].

The VH hadronic preselection requires two jets within $|\eta| < 2.4$ and with $p_T > 30$ GeV, and satisfying a pileup jet identification criteria. Category targeting the VH hadronic stage 1.2 bin is defined by particle level bin by $60 < m_{jj} < 120$ GeV and is constructed using an additional dedicated BDT. This is trained with the VH hadronic as signal, and all other Higgs and non-Higgs processes treated as background. Together

with the diphoton BDT and a selection on the dijet mass, this is used to define the final analysis categories. The final categories are defined by placing thresholds on the VH BDT output score and the diphoton BDT. Merging all years, it is found that two categories are required to reach the optimal sensitivity. A data-driven method is used to model $\gamma + jet$ and $jet + jet$, coming mainly from specific types of quark/gluon fragmentation processes and are not easily modelled.

The additional sources are introduced to account for the uncertainty in the modelling of the p_T^H , p_T^{Hjj} and m_{jj} distributions, and the number of jets in the event. A total of six sources are defined to reflect migrations of events across m_{jj} boundaries at 60, 120, 350, 700, 1000 and 1500 GeV. Two additional nuisance parameters account for migrations across the $p_T^H = 200$ GeV and $p_T^{Hjj} = 25$ GeV bin boundaries. A single source is defined to account for a migration between the zero and one. The total magnitude varies between bins is at most 8%. These are considered as uncertainties in the SM expectations while performing measurements for cross section.

Categories targeting leptonic VH production

VH production with leptonic decays (green in Figure 6-102) contains three kinds of decay channels, 2-lepton channel as $Z(\rightarrow ll)H$, 1-lepton channel as $W(\rightarrow l\nu)H$ and 0-lepton channel $Z(\rightarrow \nu\nu)H$. In the VH leptonic studies, the main reco-level categories are also divided to ZHLeptonicTag, WHLeptonicTag and VHMTag. “ZHLeptonicTag” focuses on $Z(\rightarrow ll)H$ with two reconstructed leptons and “WHLeptonicTag” focuses on $W(\rightarrow l\nu)H$ with one reconstructed lepton. “VHMTag” aims $W(\rightarrow l\nu)H$ with one lepton mis-reconstructed or identified and $Z(\rightarrow \nu\nu)H$ without any lepton. In these three channels, a dedicated BDT classifier is employed to separate between the VH signal and background events and each of the BDTs are trained on simulated signal and background events.

In ZH Leptonic Tag (2 reco-level lepton) category, the events with two same-flavour leptons like electrons or muons reconstructed and identified by lepton ID/ISO criteria (LOOSE electron ID and TIGHT muon ID) are considered in this category to extract $Z(\rightarrow ll)H$ from leptonically irrelevant backgrounds like QCD, $\gamma + jets$ and other Higgs production modes like VBF, ggH. However, the irreducible backgrounds for the category including $Z + \gamma$ and Drell-Yan and a few ttH process need to be considered. In previous studies HIG-16-040 [105], cut-based-selection strategy was used in ZH Lep-

tonic Tag category to increase the significance of $Z(\rightarrow ll)H$ signal. Since the current analysis considers to merge three-year data/MC together with larger statistics than previous ones, so MVA strategy is used to enhance the significance of $Z(\rightarrow ll)H$ signal.

In WH Leptonic Tag (1 reco-level lepton) category, the events with one lepton like electron or muon reconstructed and identified by lepton ID/ISO criteria (LOOSE electron ID and TIGHT muon ID) are considered in this category in order to extract $W(\rightarrow l\nu)H$ from leptonic irrelevant backgrounds like QCD, γ + jets and other Higgs production modes like VBF, ggH. However, the irreducible backgrounds for the category including diphoton, $W + \gamma$, Drell-Yan and Top associated production and ZH leptonic decay, ttH process need to be considered. MVA classifier is used to enhance the significance of $W(\rightarrow l\nu)H$ signal. The total magnitude varies between bins but is at most 5% for the dominant WH and ZH leptonic production modes to reflect the migration of events across p_T^V bin boundaries at 75, 150 and 250 GeV.

In VH MET Tag (0 reco-level lepton) category, the events with large MET are considered in this category in order to extract $Z(\rightarrow \nu\nu)H$ and $W(\rightarrow l\nu)H$ with one mis-reconstructed or mis-identified lepton from MET backgrounds like QCD, γ + jets, $\gamma\gamma$ and other Higgs production modes like VBF, ggH, VH hadronic decay. Since final state of VH MET Tag is basically two photons, these MET backgrounds are still dominant in VHMetTag category. However, simulation samples for fake-photon backgrounds like QCD, γ + jets as the main backgrounds are having less statistics, so data-driven QCD/ γ + jets is necessary. A dedicated MVA based BDT is trained to enhance the significance of $Z(\rightarrow \nu\nu)H$ and $W(\rightarrow l\nu)H$ signal.

Categories targeting top quark associated production

The $t\bar{t}H$ categorisation procedure is already described in Ref. [119] and is used as the basis for the ttH categories in this analysis. Two additional considerations are made. The first is to split the categories to target the individual $t\bar{t}H$ STXS stage 1.2 bins, which are divided with boundaries at 60, 120, 200, and 300 GeV in p_T^H . Secondly, an additional effort is made to distinguish $t\bar{t}H$ from tH, since a dedicated category is now included.

Due to the similarities between the final state topologies, there is a significant overlap in the signal regions for $t\bar{t}H$ and tHq events. A dedicated DNN (top DNN) is used to reject possible tHq events from entering the $t\bar{t}H$ tags. This ensures the minimum

possible contamination, and therefore reduces the anti-correlation between the two in the final fit. The inputs to the DNN are similar to those used in the $t\bar{t}H$ background discriminant described in HIG-19-013 [119].

After the application DNN cuts, events are classified according to their $(p_T^{\gamma\gamma})$ values. The stage 1.2 binning for $t\bar{t}H$ and tH are shown in purple and yellow in Figure 6-102 respectively. The categories are constructed to target the two final bins together, using a cut of $p_T^{\gamma\gamma} > 200$ GeV and without an extra boundary at $p_T^H = 300$ GeV. For each targeted region, two sets of categories are constructed using the leptonic and hadronic $t\bar{t}H$ BDTs respectively. For the hadronic decays, it is found that four categories for each target bin is optimal. For leptonic decays, four categories are constructed for the lowest p_T^H bin but only two are needed for the other bins. The tHq leptonic BDT-bkg utilize same input variables to discriminate tHq events from non-Higgs boson SM backgrounds and both are considered from simulation to perform the training. The final background model is modelled from data directly. The final category is constructed by placing a cut on both the score of top DNN and tHq leptonic BDT-bkg. Due to the low expected signal yield of tHq , one analysis category is constructed. For $t\bar{t}H$ production, each nuisance parameter shows migration from one of the boundaries at p_T^H values of 60, 120, 200, and 300 GeV. The uncertainties magnitude have values of up to 9%.

6.18.4 Data and MC samples

The study presented in this analysis is done by utilizing full Run 2 data which was collected by the CMS detector in p-p collisions are 13 TeV. Diphoton trigger is used to select the events photon p_T values of 30 (35) and 22 (25) GeV for 2016 (2017/2018) data. The variable called R_9 is used to define the energy of 3×3 crystals having centred in the most energetic crystal. The R_9 variable is utilized for the photon identification whether undergoes conversion in the upstream of ECAL. The unconverted photons usually exhibit narrow profile of transverse showers, due to which they have higher R_9 values as compared to converted ones. The efficiency of the trigger is validated using $Z \rightarrow ee$ events [120]. The estimated efficiency in data is in the bins of p_T , R_9 , and η is used for the re-weighting of events in MC in order to make it same as that of the observed ones in data.

The MC signal samples that corresponds to the various modes of Higgs boson production are generated using MADGRAPH5_AMC@NLO [121]. The events belong to the

ggH mode are re-weighted as a function of transverse momentum of Higgs boson and the count of jets in order to be coherent with the anticipation from the NNLOPS program [122]. The samples at parton level, are linked with PYTHIA8 for parton showering and hadronization simulation of 2016, 2017 and 2018 data [123]. PDF's (Parton distribution functions) are utilized from the NNPDF 3.0 [124]. Some of the signal samples are also generated with POWHEG 2.0 in perturbative QCD [125].

The major proportion of background events in this study is coming from SM production of diphoton events. A minor part is coming from $\gamma + jets$ or $jet + jet$ events where misidentified photons corresponds to jets. In the final statistical analysis, the estimation of the background has directly made from the distribution of diphoton mass in data. The MC background events from various event generators are utilized only for the training of MVA's. The events for the diphoton background produced by using SHERPA generator [126]. The simulated samples for $\gamma + jet$ and $jet + jet$ backgrounds are made by using PYTHIA8 and after applying a generator level filter to ensure the jets production with large electromagnetic activity. For the validation purpose, Drell-Yan events having final states from leptons are generated with MADGRAPH5_AMC@NLO.

6.18.5 Signal and background modelling

The shape of the signal for the distribution of diphoton invariant mass in each of the category and also for nominal mass of Higgs boson is accomplished by using MC corresponds to each of the production process. As the $m_{\gamma\gamma}$ distribution based on the associated vertex of diphoton candidate that whether the vertex was identified correctly within 1 cm, the scenarios for right and wrong vertex are separately treated for the construction of signal model. For each of the corresponding mechanism and vertex scenario, the diphoton invariant mass distributions are fitted with of five Gaussian functions at most. The sum over the functions belongs to all of the production mechanisms and related vertex scenarios is used to obtain the final fit function for each of the analysis category. Figure 6-103 shows the signal models for each of the individual year and for the three years together. The σ_{eff} belongs to half of the smallest interval which includes the 68.3% of $m_{\gamma\gamma}$ distribution. The weighted MC events are represented by open squares and the signal model is shown by the blue line. The dotted lines shows the contribution of signal model from each year of data. The left plot of Figure 6-103 represent the category corresponding to ggH 0J high p_T^H production and the S/(S+B)

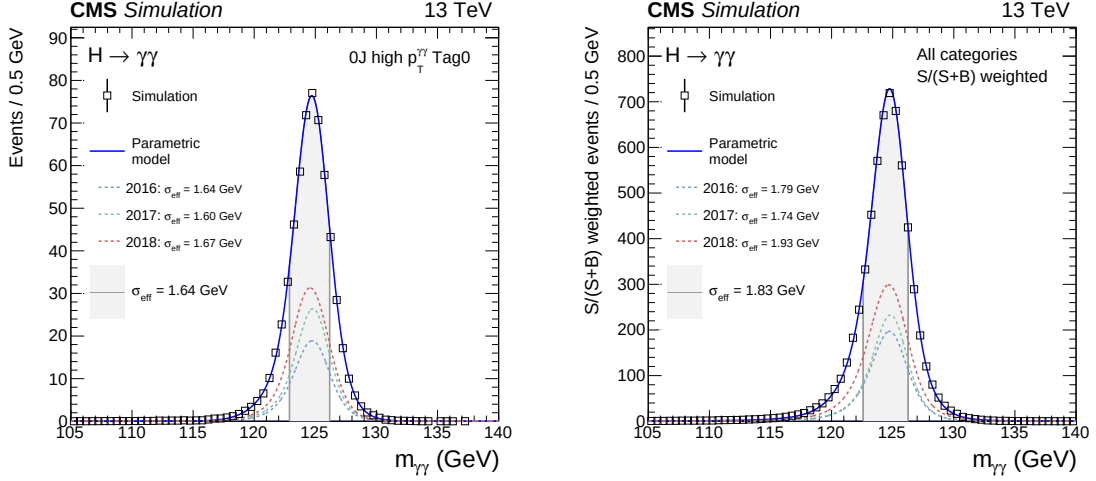


Figure 6-103 Parametric signal shapes for each year of simulation as well as for the sum of all years.

weighted sum for all of the categories is shown on the right of Figure 6-103. S and B are the numbers of expected signal and background events respectively in a $\pm\sigma_{eff} m_{\gamma\gamma}$ window centered on m_H .

The **background model** is constructed from data by utilizing the discrete profiling method [127]. This method evaluate the systematic uncertainty corresponding to the chosen analytic function to fit the background diphoton mass distribution. The background function choice is then treated as a discrete parameter in the fit to data. Many candidate function families are considered such as Bernstein polynomials, exponential functions, power law functions, and Laurent series . An f-test [128] is performed to each family of the functions in order to discover the maximum order, while the minimum order is checked by putting a requirement on goodness-of-fit to data. During the fitting procedure of these functions to the distribution of diphoton, the minimization is done to the value of $-2\Delta NLL$ (twice the negative logarithm of likelihood). When performing an estimation of a given parameter, discrete profiling method reduces the overall $-2\Delta NLL$ by taking in account all of the allowed functions which corresponds to every analysis category.

6.18.6 Systematic uncertainties

The systematic uncertainty linked with the estimation of background is managed by using the discrete profiling method. These uncertainties includes the experimental uncertainties as well as the theoretical sources. The uncertainties which modify the $m_{\gamma\gamma}$

shape distribution are considered into the signal model as nuisance parameters where each Gaussian function's mean and width can be affected. These are typical experimental uncertainties that corresponds to the energy of individual photons. On the other hand, if $m_{\gamma\gamma}$ distribution shape is unaffected, the uncertainty is handled as log-normal variation. For each STXS bin for every analysis category, the uncertainty's impact is individually determined.

The theoretical uncertainties considered in the studies are, the renormalization and factorization scale uncertainties, uncertainties in the ggH , qqH , $t\bar{t}H$ and VH leptonic STXS fractions, uncertainty in the ggH contamination of the top quark associated categories, parton distribution function uncertainties, uncertainty in the strong coupling constant, uncertainty in the $H \rightarrow \gamma\gamma$ branching fraction, underlying-event and parton shower uncertainties.

The experimental uncertainties which effect the signal $m_{\gamma\gamma}$ distribution shape are, photon energy scale and resolution, non-linearity of the photon energy scale, shower shape corrections, longitudinal non-uniformity of light collection, modelling of material in front of the ECAL and the vertex assignment. The overall effect of these shape uncertainties of signal model in the observations of Higgs boson inclusive signal strength modifier is about 2%.

The systematic uncertainties which only alter event yield and have about 4% effect on the strength modifier measurement of inclusive Higgs boson signal. A set of resources are, integrated luminosity, photon identification BDT score, jet energy scale and smearing corrections, per-photon energy resolution estimate, trigger efficiency, photon preselection, missing transverse momentum, pileup jet identification, lepton isolation and identification and b-jet tagging.

6.18.7 Results and conclusion

The composition of expected signal reflecting each of the analysis category in terms of merged STXS bins can be seen in Figure 6-104. It shows the categories of analysis where the common STXS sections are combined together in a way that the compositions of signal for each category are weighted w.r.t. the ratio of S/S+B events. The contribution of final signal yield for every category group originating per process is displayed. The fractional yield corresponds to each of the category (rows) in colour scale

for each STXS process (columns). 100% sum is therefore shown for each row but the values lower than 0.5% are not displayed. In the period of 2016 to 2018 data, MC events for each corresponding years are combined together having appropriate weights w.r.t. the integrated luminosity. The qqH rest labelled in the column indicates contributions related to qqH 0J, qqH 1J, qqH $m_{jj} < 60$ GeV and qqH $120 < m_{\gamma\gamma} < 350$ GeV STXS bins. In Figure 6-105, the sum of categories is shown for the best fit signal and background model. Each of the corresponding category is weighted by S/S+B in order to keep constant absolute signal yield.

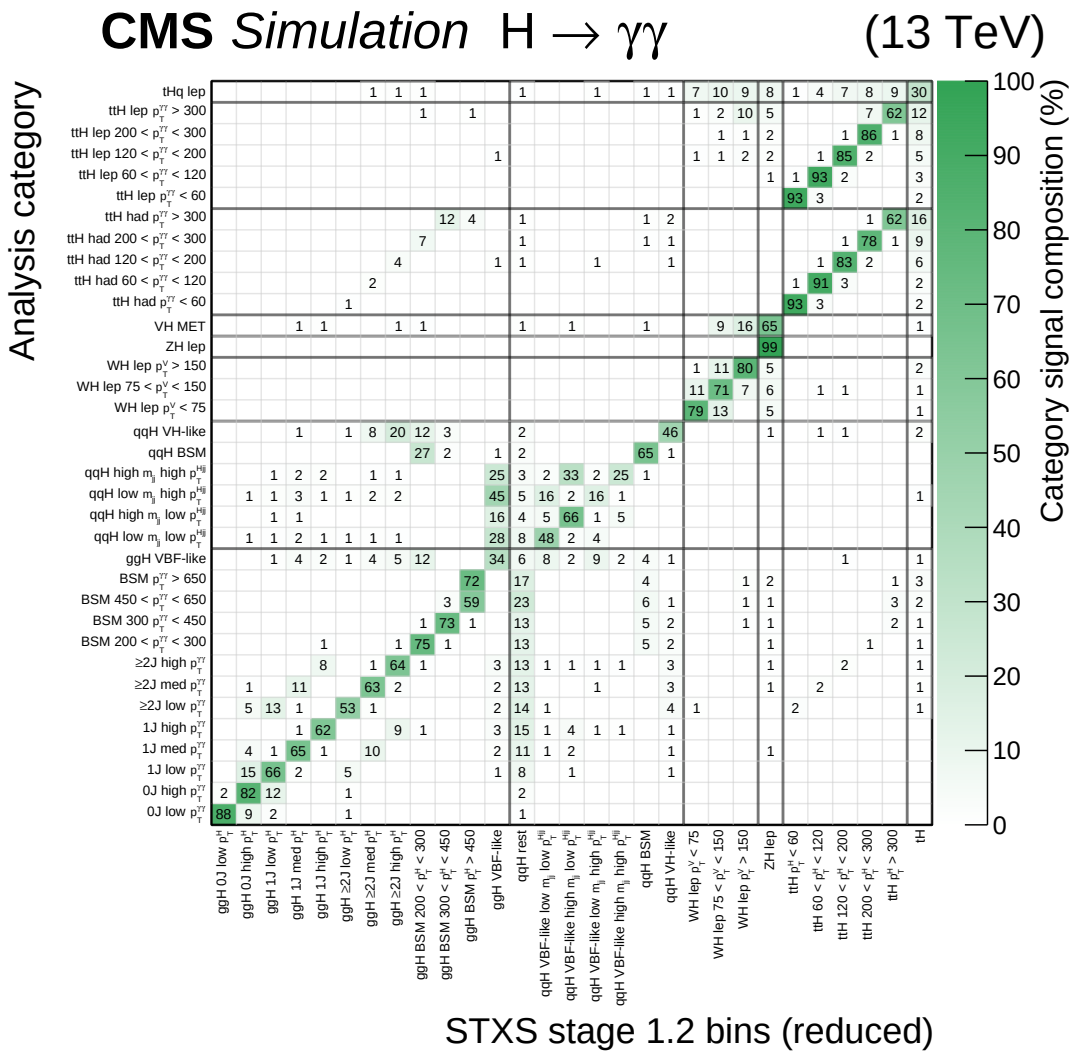


Figure 6-104 Each analysis category composition in terms of a reduced set of STXS bins.

The **signal strength modifier** μ , is known as the ratio of observed product of cross section of Higgs boson and the branching fraction of diphoton relative to the SM expec-

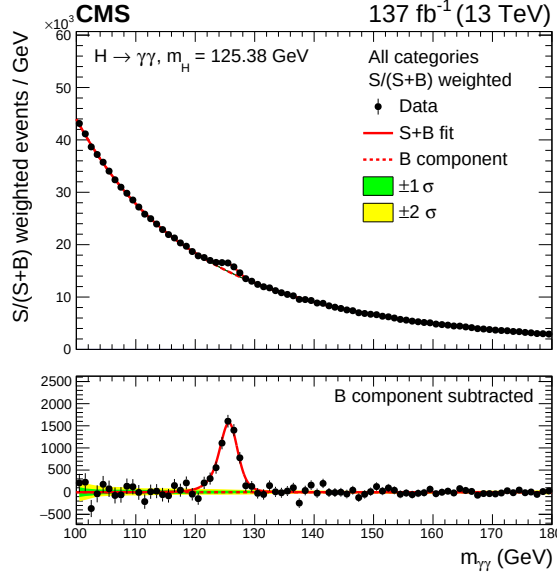


Figure 6-105 Data points in black and fit on the signal plus background model for all categories summed are shown. Each category is weighted by $S/(S + B)$.

tation. It is observed as:

$$\mu = 1.12^{+0.09}_{-0.09} = 1.12^{+0.06}_{-0.06}(\text{theo})^{+0.03}_{-0.03}(\text{syst})^{+0.07}_{-0.06}(\text{stat}).$$

The uncertainty is divided into theoretical, experimental and statistical fragments. The compatibility of the fit relative to SM prediction is approximately 74% which is expressed as a p-value. For each of the Higgs production mechanism, signal strength modifiers are also measured. Contrary to the subsequent STXS fits, the VH hadronic and VH leptonic modes are combined to scale according to μ_{VH} and the VBF production mechanism scales with μ_{VBF} . The μ_{top} parameter is scaled to $t\bar{t}H$, tHq and tHW modes of production equally and μ_{ggH} has been scaled to both ggH and $b\bar{b}H$ production mechanism.

Figure 6-106 shows the values of signal strength modifiers and the corresponding uncertainties per production mode. Comparatively to the previous results, the precision of these measurements has improved significantly in the $H \rightarrow \gamma\gamma$ decay mode. Particularly, the result of μ_{VH} signal strength modifier has been substantially enhanced compared to what shown in reference [105]. Relative to the SM prediction, the p-value of signal strength modifier fit per production mode is approximately 50%.

The **Simplified template cross-sections** are measured by performing the fits within the STXS framework and the corresponding intervals of 68% confidence level. There

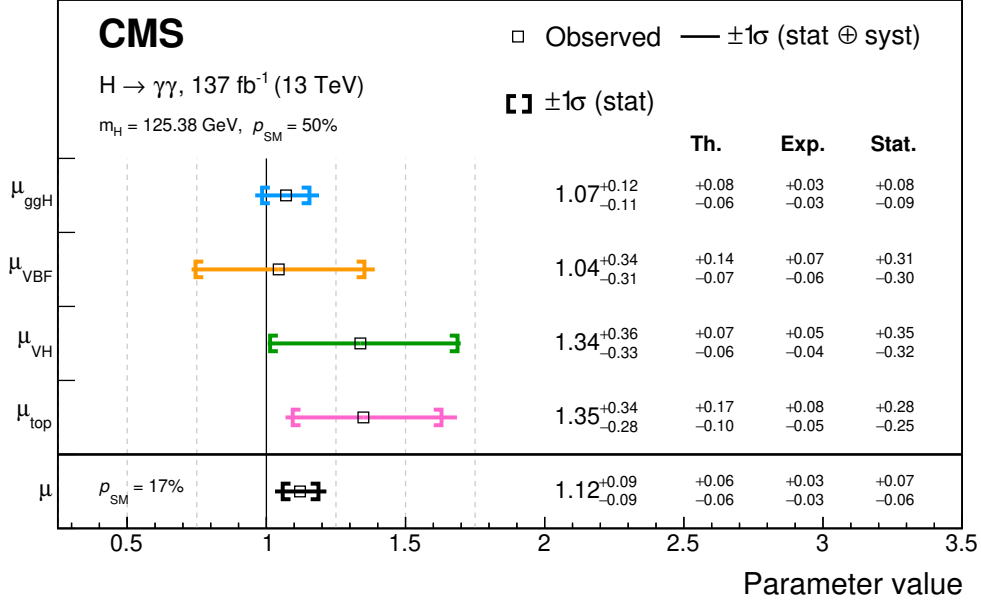


Figure 6-106 Observed results of the fit to four production mode signal strength modifiers.

are two dissimilar parameterizations taken in account based on the varying granular levels of the merging of STXS bins. It was needed to merge bins in order to avoid large uncertainties in some of the parameters as well as parameters correlations. The measurements with less model dependence might have large uncertainties in the cross section results.

The results with two different fits to the cross sections measurement are presented with partially merged STXS bins. The first one is assigned as the "maximal" merging scenario, where the bins are merged as long as their expected uncertainty becomes less than 150% to that of predicted by SM. The "minimal" merging is the second scenario where the fit merges as less bins as possible while ensuring that the parameters should not be much anti-correlated i.e. smaller than approximately 90%.

In the "maximal merging" scenario, there are 17 parameters of interest. The VBF-like sector (≥ 2 -jets, $m_{jj} > 350$ GeV) in the qqH and ggH schemes are merged together in order to define the qqH VBF-like and ggH VBF-like parameters. The four bins having Higgs boson $p_T > 200$ GeV in the gluon-fusion scheme are combined into a single bin referred as ggH BSM bin. The leptonic (WH, ZH) and $t\bar{t}H$ bins are merged fully. The ZH (leptonic) parameter include both ggZH and ZH production.

The second scheme to perform the fit, is the "minimal merging" scenario where even more granular measurement is done with 27 interest parameters. The qqH VBF-

like region has been divided fully in four stage 1.2 STXS bins with boundaries at $m_{jj} = 700$ GeV and $p_T^{H_{jj}} = 25$ GeV. In order to avoid large correlations between fitted parameters, four of the ggH VBF like bins combined with the related bins of the qqH scheme. It also aid to reduce model dependence which is required to fit separate ggH and VBF parameters in an identical phase space. In addition, the splittings are also included in the ggH scheme at $p_T^H = 300$ and 450 GeV and WH leptonic scheme at $p_T^V = 75$ and 150 GeV. The $t\bar{t}h$ region has splitted into five region based upon the boundaries at $p_T^H = 60, 120, 200$ and 300 GeV.

The measured cross sections with 68% confidence level intervals are displayed for the both of the merging schemes in Figure 6-107 and 6-108. In the fits of both, maximal and minimal merging scenarios, the statistical uncertainty component is dominant for all of the observed cross sections. The results corresponding both merging schemes fits are in agreement with what predicted by SM predictions. The p-values w.r.t. the predictions of SM are 31% for the maximal merging scenario and 70% for minimal merging scenario.

In maximal merging scheme (Figure 6-107), the production of ggH having $p_T^H > 200$ GeV, that is sensitive to beyond SM physics particularly, might enters the ggH loop, is observed to a precision which is less than 50% w.r.t. SM prediction. The measured cross section with a value of $0.9_{-0.3}^{+0.4}$ is consistent with the SM prediction. The product of tH cross section times $H \rightarrow \gamma\gamma$ branching fraction is observed to be $1.3_{-0.7}^{+0.8}$ fb belongs to an excess of $6.3_{-3.7}^{+3.4}$ times the expectation of SM. A tH production rate of 14 (8) times the SM prediction is observed (expected) can be excluded at the 95% CL.

The STXS fit for the minimal merging scenario (Figure 6-108) shows the most granular cross section observations done in a single decay channel of Higgs boson. It is sensitive to various regions of phase space belong to the Higgs boson production. The measurement shows the first results of $t\bar{t}H$ production in bins of Higgs boson transverse momentum. In each of the four $t\bar{t}H$ bins, the uncertainty size for $p_T^H < 300$ GeV is less than 100% of what expected in SM. The ggH production cross sections with $p_T^H > 200$ GeV are observed in three different regions and all of the related measurements are in agreement with the respective expectations of SM.

The fitted parameters correlations are shown Figure 6-109 (maximal merging scenario) and Figure 6-110 (minimal merging scenario). The ggH parameters correlations

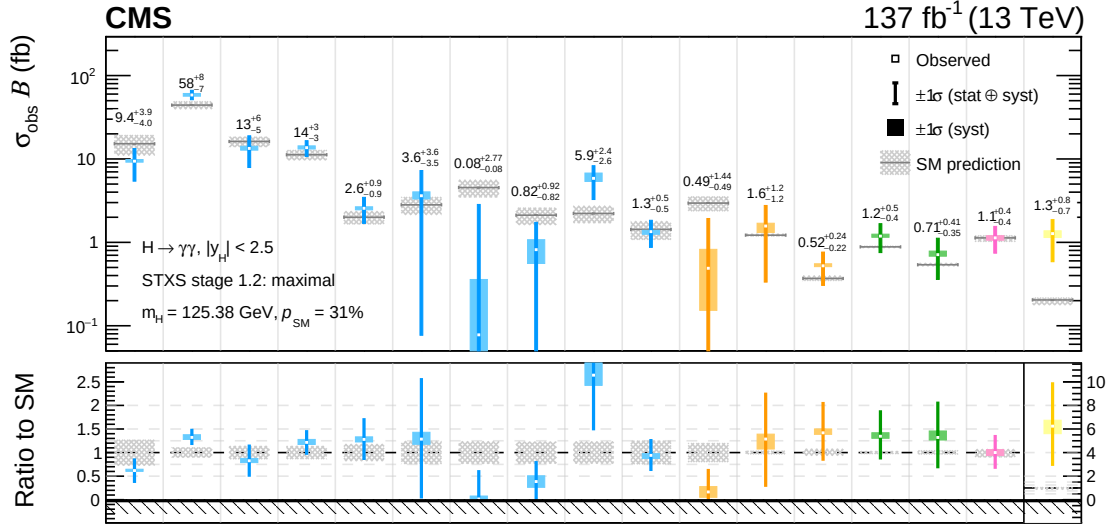


Figure 6-107 Observed results of the maximal merging scheme STXS fit.

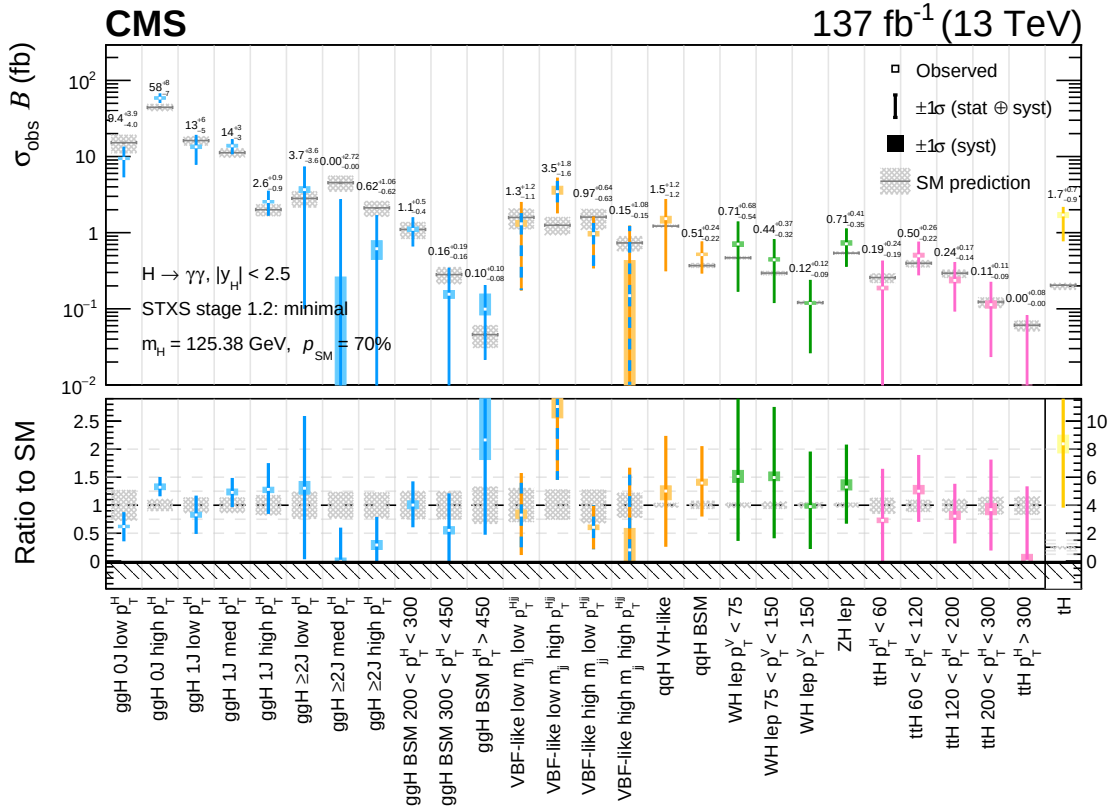


Figure 6-108 Observed results of the minimal merging scheme STXS fit.

are observed smaller between adjacent p_T^H bins and between adjacent jet bins, correlations are larger. The correlations of the parameter are minimized by applying the ggH MVA in the categorization of events. In the maximal merging scheme, the largest correlations does exist between the "ggH VBF like" and "qqH VBF like" cross sections as well as in the $t\bar{t}H$ and tH cross sections having values of -0.76 and -0.59 respectively. These come from the contamination (sizeable) of "ggH VBF like" events in to the qqH

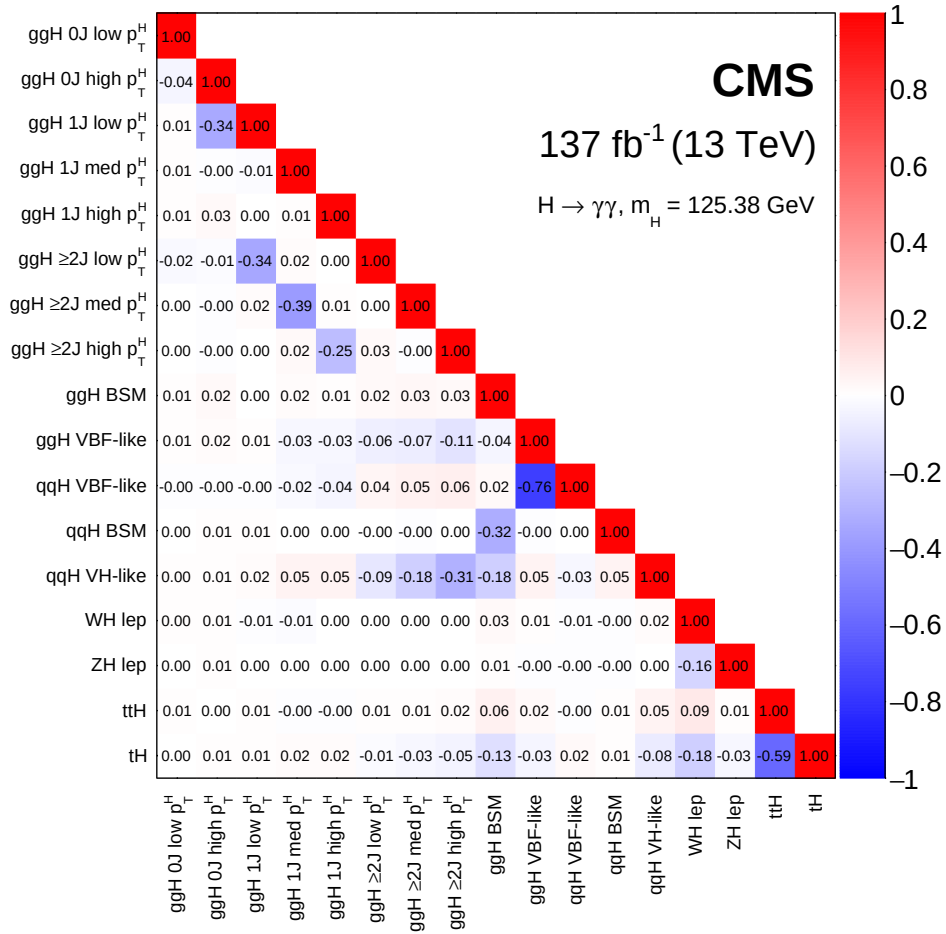


Figure 6-109 Observed correlations in the "maximal" merging fit between the 17 parameters.

categories, also the contamination of $t\bar{t}H$ events in to tH q leptonic category. In the minimal merging scenario, larger correlations in the observations are established because of the splitting of $t\bar{t}H$ production into five individual parameters. The **coupling modifiers** are defined by κ -framework [72] to parameterize directly the deviations from the expectation of SM in the Higgs boson couplings to other particles. In this context, two

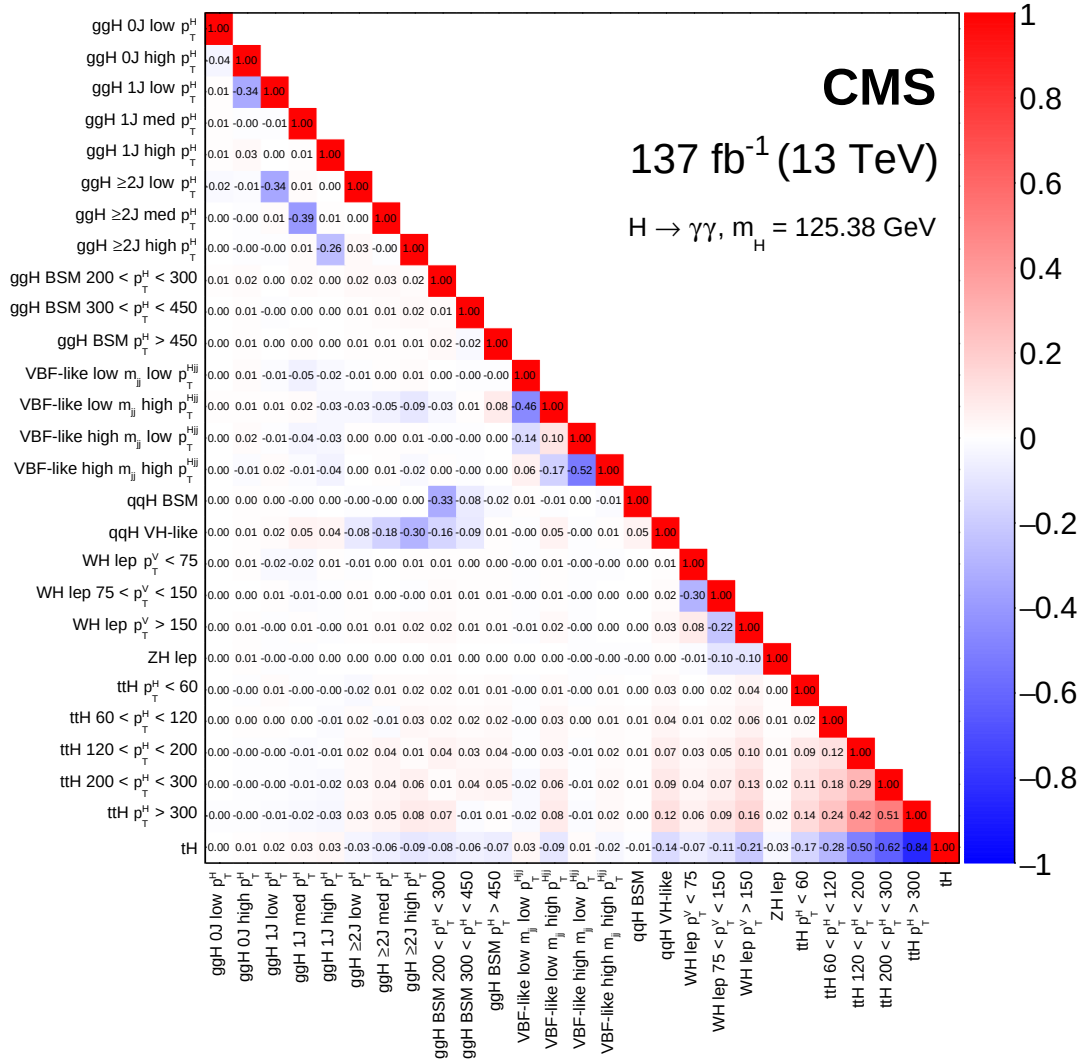


Figure 6-110 Observed correlations in the "minimal" merging fit between the 27 parameters.

2D likelihood scans have been performed.

The resolved κ model is utilized in the first fit. The scaling factors (ggH production and $H \rightarrow \gamma\gamma$ decay) of loops are resolved to their corresponding SM components as a function of other κ parameters. The observation of a 2D scan in κ_f and κ_v with scaling of Higgs boson's coupling to fermions and vector bosons respectively, is shown in Figure 6-111. Other κ parameters are treated as unity in the fit. An interference term is included in tHq and tHW production mechanism proportional to $\kappa_v \kappa_f$, making them sensitive to the relative sign of the t-H and V-H couplings. The study presented in this framework (STXS) exclusively target production of tHq through the tHq leptonic

category. The negative values of κ_f are excluded with significance of 0.5σ (observed) and 2.4σ (expected).

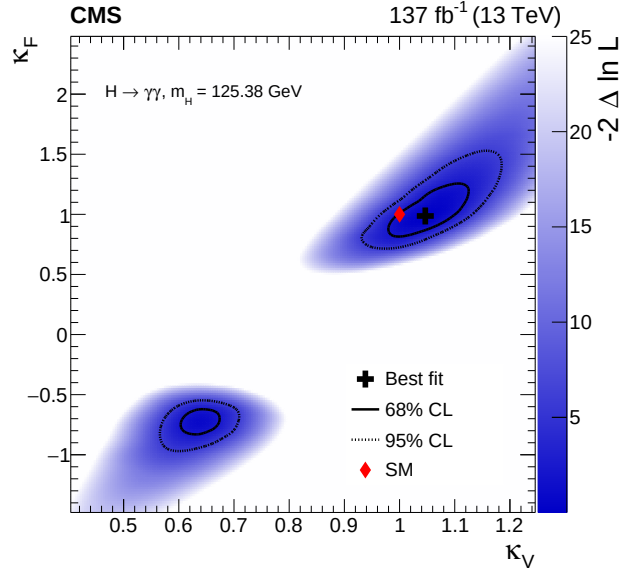


Figure 6-111 Two dimensional likelihood scan (observed) in the κ -framework κ_v -vs- κ_f in the resolved κ model.

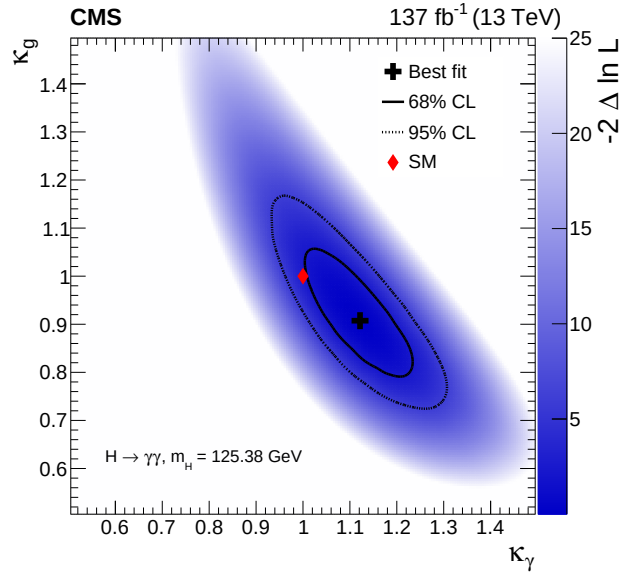


Figure 6-112 Two dimensional likelihood scan (observed) in the κ -framework κ_g -vs- κ_γ in the unresolved κ model.

The second fit is performed by using unresolved κ model. In this model, the $H \rightarrow \gamma\gamma$ and ggH loops are assigned their own scaling factors which are denoted by κ_γ and κ_g respectively. The observed two dimensional results can be seen in Figure 6-112. The

other κ parameters in this model are unity. Higgs boson coupling measurements are in agreement with the expectations of SM.

6.19 Inclusive and differential fiducial cross-sections with full Run 2 dataset at 13 TeV

6.19.1 Introduction

The production cross section of Higgs boson has been explored in different approaches. The STXS approach as described in the previous section 6.18 where Higgs boson production mechanisms separated in to bins in order to mimic the dependence on theory. The study focused in this section of the chapter, is the inclusive and differential observation of fiducial cross sections in order to achieve results with minimum model dependence. The method used for the signal and background modelling is same as discussed in Section 6.17.

The reconstruction of diphoton invariant mass with high precision using $H \rightarrow \gamma\gamma$ decay mode provides the accurate observations of differential production cross sections. In this study, the cross sections are measured w.r.t. many new observables compared to what done in the previous studies. The strategy exploited in this analysis follows the same as that of the two previous measurements at CMS. The methodology is designed as to reduce the dependence on assumption made by the theory for the extraction of results. The fiducial region defined as to lessen the detector acceptance and the categorization of event is done by utilizing the per event mass resolution estimation at reconstruction level.

The diphoton system observables are used for the extraction of fiducial differential cross sections such as the jets transverse momentum produced along with the diphoton, the system of dijet with two leading p_T jets and various event level observables. In the current studies as also described in ref. [26], the measurements of differential cross section are made a function of angular observable ϕ_η^* (in Eq. (6-2)) which aid the measurement of diphoton p_T at low points of its spectrum and τ_C^j (in Eq. (6-3)), the jet p_T which is rapidity weighted. In addition to this, double differential observations are also performed in diphoton p_T and number of jets bins as well as in the bins of diphoton transverse momentum and τ_C^j .

6.19.2 Data and MC samples

The studies for the fiducial and differential fiducial cross sections has been accomplished with full Run 2 data and the diphoton trigger is used to select the events from p-p collision. The asymmetric p_T values of 30 (35) for leading and 22 (25) GeV for sub-leading photon in 2016 (2017/2018) are used respectively with 99% of the efficiency. The diphoton invariant mass should be greater than 90 GeV whereas each of the photon has to pass the calorimetric isolation. The efficiency of the trigger is checked by using tag and probe method from $Z \rightarrow ee$ events. The single electron trigger is used to collect these events. The efficiency in data is measured in p_T and R_9 bins [94] and variable η , is exploit to weight the MC events in order to replicate the observed efficiency of trigger in data.

The MC signal samples for all of the major production mechanism (ggH, VBF, ttH and VH) are produced using MADGRAPH5_AMC@NLO [121]. The ggH events are weighted w.r.t. the Higgs boson p_T and total number of jets in the corresponding event in order to match the NNLOPS (parton showering) prediction [122]. The samples at parton level, are linked with PYTHIA8 for parton showering and hadronization in 2016, 2017 and 2018 data [123]. The recommended cross sections by the LHCXSWG are used as well as the branching fractions.

The dominant background source is diphoton events from QCD. The remaining part is coming from $\gamma + jets$ or $jet + jet$ events where misidentified photons corresponds to jets. The data events are used for estimation of diphoton invariant mass shape and normalization of background. The background events from various event generators are utilized. The training of photon ID MVA is done by using $\gamma + jets$ samples and in order to optimize the categorization of events, $\gamma + jets$ and $\gamma\gamma$ are used. SHERPA version 2.2.4 is considered for the generation of diphoton background events [126]. PYTHIA8 has been utilized to produce $\gamma + jets$ and $jet + jet$ MC backgrounds. The MC events of $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ and $Z \rightarrow \mu\mu\gamma$ are produced with MADGRAPH5_AMC@NLO, used in order to derive the photon energy scales, resolution and correction factors.

6.19.3 Event reconstruction and categorization

The reconstruction of physics objects is done by using the PF algorithm. The purpose of these algorithms is to identify each particle in an event with the information

from different parts of CMS detector.

The construction of a correct primary vertex in diphoton final state is complicated in a way that, photons do not leave any track in the tracker part of CMS detector. For the measurement of kinematic variables of diphoton system, the right vertex assignment is of high importance in order to keep good diphoton mass resolution. A dedicated BDT is used for constructing the correct vertex position of a diphoton event. Simulated samples, using track information are exploited to train the BDT which were produced via ggH mechanism. The performance of this dedicated MVA is validated with data by using $Z \rightarrow \mu\mu$ events where the vertex refitting is done by ignoring muon tracks.

The discrimination of prompt photons w.r.t. those coming from decay of neutral particles such as, from π^0 or η mesons is done by using an MVA classifier known as photon ID MVA, trained with $\gamma + jets$ MC samples. The input variables of this BDT are based on associated SC shapes of candidate photon and isolation sums which are built from EM components of charged and neutral hadrons. [94].

The photon ID BDT is a major mean to reduce $\gamma + jets$ and $jet + jet$ backgrounds. In the MC events, its imperfect modelling is considered as one of the crucial sources of systematic uncertainty. The mis-modelling of MVA input variables is causing the inaccurate modelling and the reason for this, is the changing conditions of ECAL because of radiation damaging. In order to cope up with this, corrections are applied to the MC which are developed with diphoton final states [97, 119]. A dedicated method has been devised to compute a per-photon correction that morphs the input variable distributions in simulation, and their correlations, to match data. Chained quantile regression (CQR) method, is used to estimate corrections at per photon level and applied on MC to replicate data, SC shape and isolation variables. A general layout of CQR method is shown in Figure 6-113 and details can be find here [26].

In the isolation sum of photon, the first part is equalizing of events in the first isolation bin distribution between MC and data by exploiting a stochastic procedure. This is done in order to move the events from peak (zero isolation) to the tail of distribution and vice versa. The random events are picked which has to be moved having a probability estimated from the related events count between peak and the tail. A BDT is trained in order to have zero isolation sum using XGBOOST (Extreme Gradient Boosting). For all of the isolation variables, a new value is assigned for an event to be moved to the tail

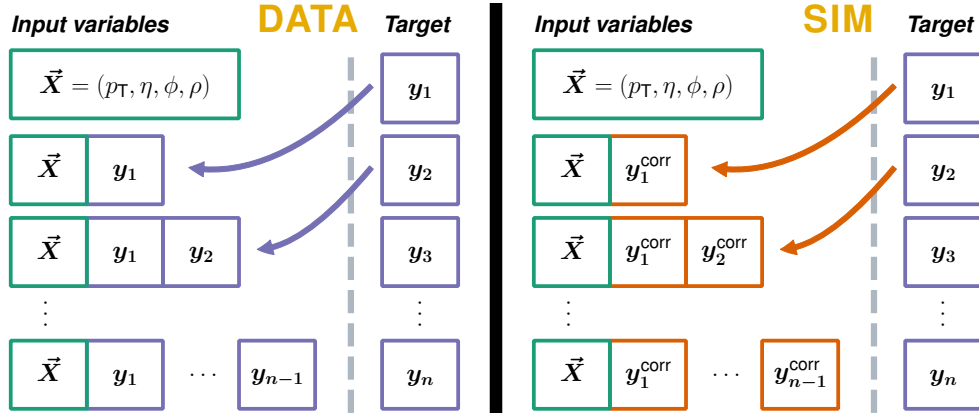


Figure 6-113 A general layout of chained approach for some input variables for quantile BDT's.

The variables to be corrected are also given.

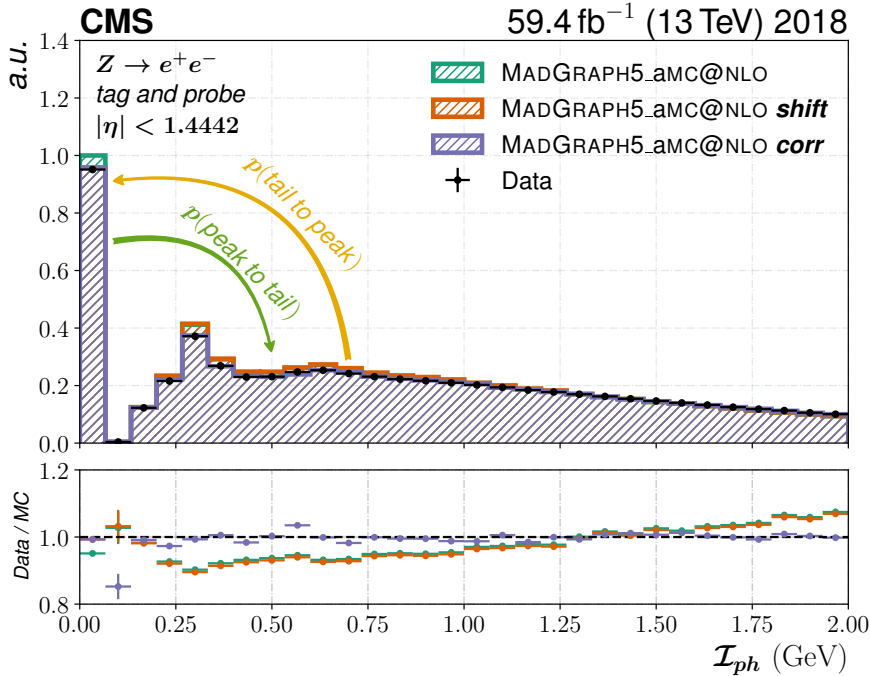


Figure 6-114 The distribution of isolation sum of photon in data (black dots). The histogram in green shows uncorrected distribution and after equalizing number of events having zero isolation in MC with data is in orange colour. The distribution in purple is the one after the application of equalizing step and CQR method to tail part.

part. An example distributions related to the isolation sum of photon for all of the cases as well as equalizing illustration of step is shown in Fig. 6-114.

Charged isolation estimated w.r.t. the vertex with largest value is known as worst charged isolation and is constructed larger as compared to the charged isolation calculated w.r.t. the chosen vertex i.e. the real charged isolation. It may lead to the possibility

of three categories such as, both of the charged isolation sums are zero exactly, the worst charged isolation greater than zero and real charged isolation is zero and the case where both of the isolation larger than zero. An MVA classifier with inputs (p_T , η , ϕ , ρ (average spatial event energy density [105])), is trained to separate them. The mobility of events among three categories is chosen randomly with a probability estimated from the MVA output.

The inputs variables of photon ID MVA are corrected in MC in order to better replicate the distributions to the ones in data with 1% in EB and approximately 5% in EE. It is shown in Fig. 6-115.

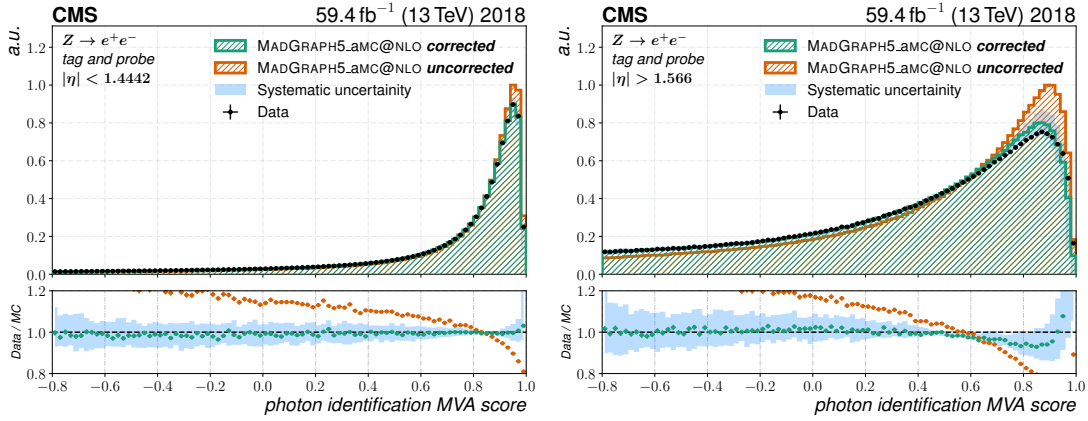


Figure 6-115 Photon ID MVA distribution for probe candidate in tag and probe ($Z \rightarrow ee$) sample for data and the Madgraph5_aMC@NLO simulation with barrel (left) and endcaps (right).

The categorization of diphoton events used in the studies is base upon the decorrelated mass resolution estimator which is explained below as well as the photon ID MVA classifier. This criteria for categorization has been constructed to mimic the model dependence whilst keeping maximum possible sensitivity.

6.19.4 Mass resolution estimator

The photon pairs that enter this analysis are categorized according to the decorrelated mass resolution estimator and the photon identification MVA. These categorization criteria are designed to ensure maximal model independence while yielding the best possible analysis sensitivity.

Mass resolution estimator (σ_m/m) for diphoton pair is quadratic sum of the cor-

responding per photon energy resolution. The energy resolution of ECAL become less with large energies of photons which ultimately reduces the noise as well as the stochastic effects. This causes to a correlation between diphoton invariant mass and (σ_m/m) . This correlation together with selection on σ_m/m can distort the background shape and may ruin the supposition of a smoothly falling background which is required to model it using discrete profiling.

In order to avoid this influence, the decorrelated estimator of mass resolution is considered w.r.t. the invariant mass of diphoton system. This is done by exploiting the quantile morphing technique where σ_m/m distribution is divided in to mass dependent bins. The σ_m^D is known as decorrelated mass resolution estimator and is defined as:

$$\sigma_M/M|_{decorr} = cdf_{M_{ref}}^{-1} (cdf_M(\sigma_M/M)) \quad (6-1)$$

with cdf_M is the cumulative distribution function for mass bin M and M_{ref} is the reference mass for which the corresponding σ_M/M distribution is target distribution. The explanations on this method can be accessed in Ref. [129].

6.19.5 Decorrelated mass resolution categorization

The σ_m^D category boundaries are streamlined at the same time with photon ID MVA classifier selection by utilizing expected significance of signal. MC signal samples at $m_H = 125$ GeV are used for the optimization along with diphoton and $\gamma + jets$ events for background modelling. It is then realized that, for all of the years, the finest number of categories corresponds to σ_m^D is three. The diphoton pair above the σ_m^D highest boundary are not considered. The boundaries of σ_m^D assemble the photon pairs using ECAL regions, where each photon out of the diphoton pair were detected. It results in 75% to 80% efficiency belongs to photon ID criterion for all of the three years data taking. An explanatory breakdown, along with σ_m^D categorization efficiencies are presented in Table 6-46. The events are considered only which satisfies the fiducial selection.

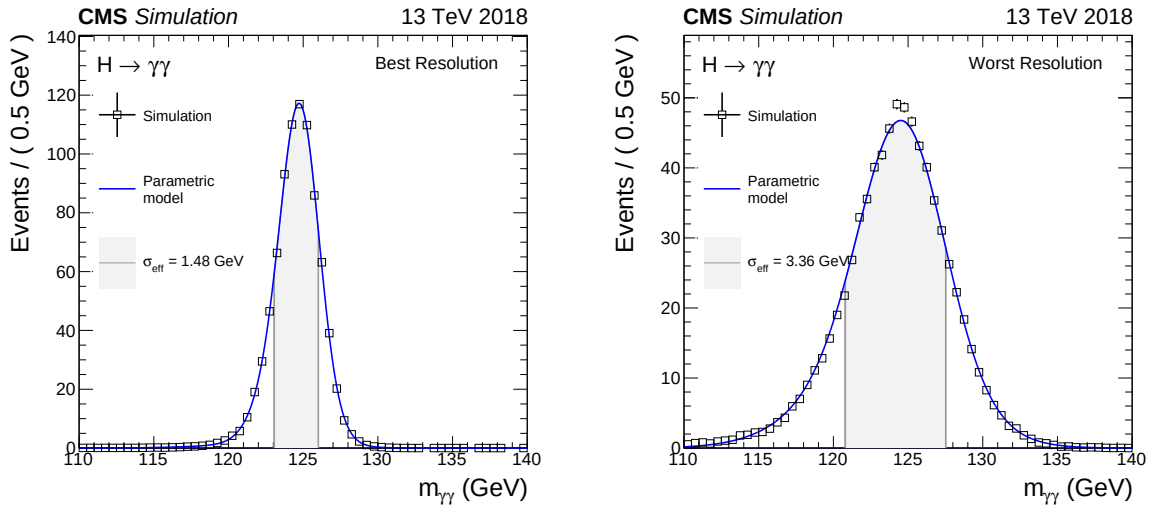
The signal shape in certain categories is shown in Figure 6-116 which includes the signal MC with all of the four main Higgs production modes correspond to a mass of $m_H = 125$ GeV.

Table 6-46 The efficiencies of the photon ID MVA and σ_m^D boundaries using signal sample for 2016, 2017 and 2018 data taking.

2016				
Resolution category	Best	Medium	Worst	Overall
Photon ID MVA efficiency	84.70%	81.71%	70.31%	77.31%
Resolution category efficiency	31.43%	25.39%	41.27%	

2017				
Resolution category	Best	Medium	Worst	Overall
Photon ID MVA efficiency	90.36%	83.92%	73.01%	80.29%
Resolution category efficiency	24.45%	34.04%	38.77%	

2018				
Resolution category	Best	Medium	Worst	Overall
Photon ID MVA efficiency	86.81%	78.76%	63.48%	75.34%
Resolution category efficiency	27.80%	35.91%	35.46%	

**Figure 6-116 Signal models for fiducial cross section measurement for the best and worst resolution categories in 2018.**

6.19.6 Observables and fiducial phase space

In order to mimic the model dependence, the production cross sections (inclusive and differential) measurements carried out in fiducial phase spaces. The identification

and kinematic set of selections are placed at generator level to the diphoton pair as well as single photons are outlined in Table 6-47. The QCD and EWK information are included at the particle level information. $\mathcal{I}_{gen}^\gamma$ is estimated as the hadronic energy sum in a cone of $\Delta R = 0.3$. This requirement is needed to reduce the photon ID MVA classifier selection criteria at level of reconstruction. Along with the main fiducial phase space,

Table 6-47 Fiducial phase space definition where labels 1, 2 refer to the leading and subleading photon.

Observable	Selection
$p_T^{\gamma 1}/m_{\gamma\gamma}$	$> 1/3$
$p_T^{\gamma 2}/m_{\gamma\gamma}$	$> 1/4$
$\mathcal{I}_{gen}^\gamma$	$< 10 \text{ GeV}$
$ \eta^\gamma $	< 2.5

in order to target the specific variables of interests, some other fiducial phase spaces are also defined. The additional jets are considered in the studies and are required to satisfy $p_T^j > 30 \text{ GeV}$. If an additional jet evaluation is required for the measurement of an observable then, 1 jet phase space is utilized which exploit at least one jet should be in the event having leading jet p_T to be $|\eta^j| < 2.5$. The observable which need the involvement of two extra jets, the 2 jet phase space is used with relaxed $|\eta|$ selection ($|\eta^j| < 4.7$) for all of the jets. The 2 jet phase space is constructed to accommodate events coming from the VBF Higgs boson production mode and is referred as VBF enriched phase space. Some more requirements on the events to be in the VBF enriched phase space are: $n_{jets} \geq 2$, $m^{jj} > 200 \text{ GeV}$, and $\Delta\eta^{jj} > 3.5$.

For the system of diphoton, two extra observables are defined. One of them is the cosine of polar angle in the Collins-Soper frame of reference $\cos \theta^*$ [130]. It is sensitive to CP properties and spin of the diphoton resonance. The other one is ϕ_η^* , which is defined to probe the low p_T region while reducing the experimental uncertainties. It is expressed as [131]:

$$\phi_\eta^* = \tan \left(\frac{\phi_{acop}}{2} \right) \sin(\theta_\eta^*) \quad (6-2)$$

where acoplanarity angle is, the angle between photons in the transverse plane as $\phi_{acop} = \pi - \Delta\phi$, and θ_η^* is the scattering angle of photons w.r.t. the proton beam

in a frame of reference where the particles are back to back in (r, θ) plane. The system of diphoton pair and the leading jet, the difference in rapidity and ϕ between jet and the diphoton pair are considered as observables. In order to specifically probe the VBF production mode properties, the τ_C^j observable is taken in to consideration, which is defined as [132]:

$$\tau_C^j = \max_{k \in \text{jets}} \left(\frac{\sqrt{E_k^2 - p_{z,k}^2}}{2 \cosh(y_k - y_H)} \right) \quad (6-3)$$

It is calculated up to six largest p_T jets from which the one with maximum is considered. Transverse momentum of the jet is weighted by a function which depends on the jet's rapidity and mimic the forward jet's contributions. In equation 6-3, the denominator represents the difference y_k (rapidity of the jet) and y_H (rapidity of the Higgs boson) is considered. It does mean that the τ_C^j is constructed in a frame of reference where rapidity of the Higgs boson is zero. The detailed description of the binning as per observable of interest can be seen in the table 3 of Ref. [26].

The categorization of events has been done in bins of observable of interest. In order to unfold the bins at level of reconstruction to the bins at particle level, the signal sample further split into the bins of particle level observable of interest. The signal model is sensitive to events fraction from every particle level bin as being reconstructed in a reconstruction level bin and category, thus maximum likelihood unfolding of the detector response has been performed. The imperfect alignment of the fiducial phase space at the particle level and the reconstruction selection means that there are certain events which enter the fiducial phase space but being generated outside and vice versa. This effect has been considered and is done by deriving a probability density function (pdf) for the signal events coming outside the fiducial phase space at particle level and considering them in maximum likelihood fit as an extra production mechanism. In particular, for the observables measured in restricted phase space regions, the cross section of events which enters the fiducial phase space but not restricted region has been left floating in the fit and treated as additional bin. The product of efficiency with acceptance per particle level bin, the reconstruction level bin and category combination is presented in Figure 6-117 for the year 2018 for p_T and n_{jets} observables.

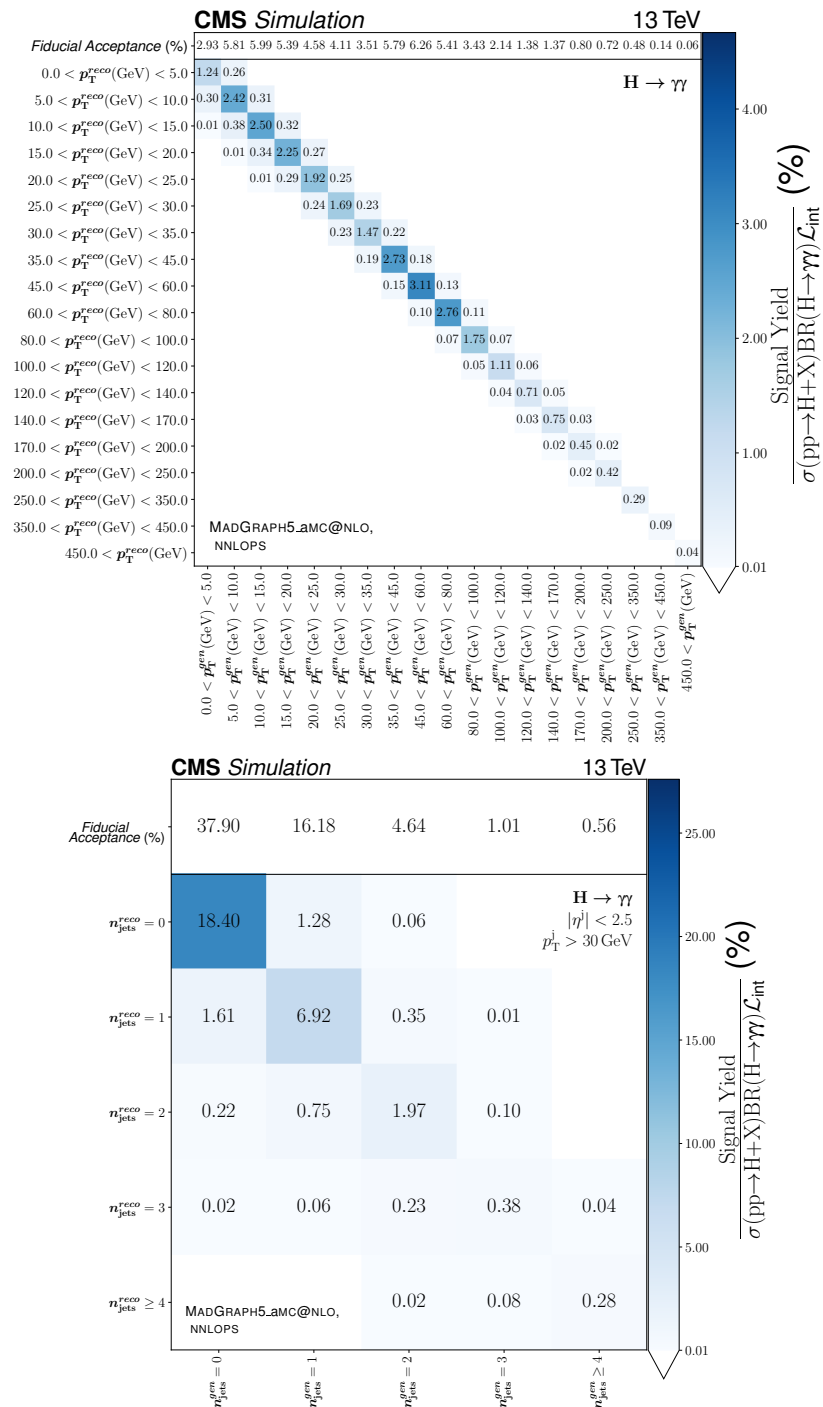


Figure 6-117 The event yields summed across all resolution categories divided by the total $H_{\gamma\gamma}$ cross section multiplied by the integrated luminosity for the bins in particle level and reconstruction level observables for the year 2018.

6.19.7 Systematic uncertainties

This section shows the systematic uncertainties considered in the studies. The theoretical uncertainties taken in account are, the PDF uncertainty, α_s uncertainty and renormalization and factorization scale uncertainties.

The experimental uncertainties are, photon energy scale and resolution, photon energy scale non-linearity, cluster shape uncertainties, non-uniformity of light collection, modelling of material upstream of ECAL and the vertex assignment.

The systematic uncertainties which effect the event yield only includes, integrated luminosity, per photon energy correction, jet energy scale and smearing corrections, per-photon energy resolution estimate, trigger efficiency, photon preselection, photon ID MVA score, missing transverse momentum, pileup jet identification, lepton identification and b tagging efficiency. Table 6-48 shows some of the dominant systematic uncertainties and their impacts.

Table 6-48 Dominant systematic uncertainties and impacts.

Systematic uncertainty	Impact
<i>Per-photon energy resolution</i>	2%
<i>Integrated luminosity</i>	1.5%
<i>Photon identification MVA score</i>	1.5%
<i>Photon energy scale and resolution</i>	1.0%
<i>Photon preselection</i>	0.6%
<i>Theoretical uncertainties</i>	0.5%
<i>Vertex assignment</i>	0.1%

6.19.8 Results and Conclusion

The results of the studies carried out are presented in this section. The measured cross section is as follows for the defined fiducial region:

$$\sigma_{\text{fid}} = 73.40_{-5.3}^{+5.4}(\text{stat})_{-2.2}^{+2.4}(\text{syst})\text{fb} = 73.40_{-5.9}^{+6.1}\text{fb}. \quad (6-4)$$

For the fiducial cross section, likelihood scan and the nominal theoretical predictions are shown in Figure 6-118. 75.4 ± 4.1 fb is the nominal prediction of theory for the fiducial cross section.

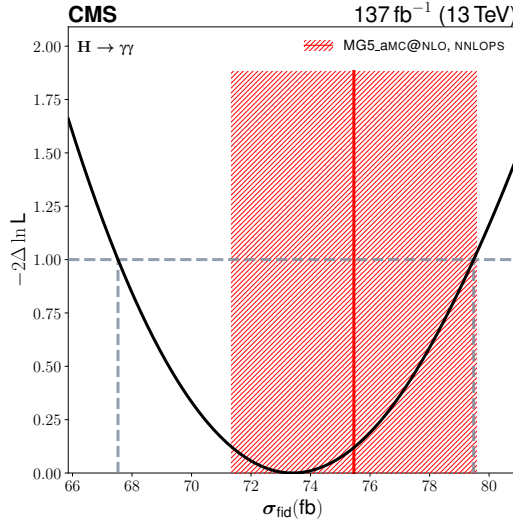


Figure 6-118 Likelihood scan for $H \rightarrow \gamma\gamma$ cross section in the fiducial region.

Figure 6-119 shows the invariant mass of diphoton from the hypotheses for both signal and background. The events are added together from all of the categories and are $S/(S + B)$ ratio weighted for the corresponding category. The best fit result is in agreement with the SM expectation. The statistical uncertainty is about 7% and is the dominated source. The systematic uncertainty component is about 3%.

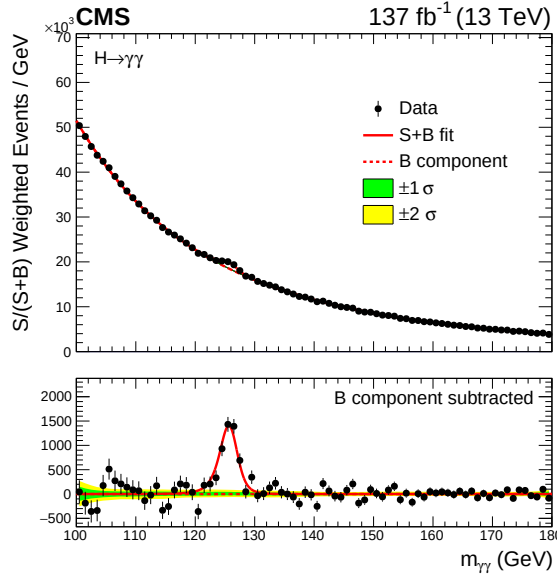


Figure 6-119 Invariant mass distribution of diphoton with all of the categories combined and used for the measurement of inclusive fiducial cross section.

The expected uncertainty in the measurement of inclusive fiducial $H \rightarrow \gamma\gamma$ cross section as visible in form of hatched area in Figure 6-118 and the cross section shown in Figure 6-120 as hatched part for the dedicated regions of phase space are coming from

the combination of various components. The dominating uncertainty is from the cross section of Higgs boson production and second substantial component of the uncertainty is coming from the branching fraction of $H \rightarrow \gamma\gamma$. Both of these uncertainty components are taken from [59]. The expected and the observed cross sections in the dedicated regions of fiducial phase space are displayed in Figure 6-120.

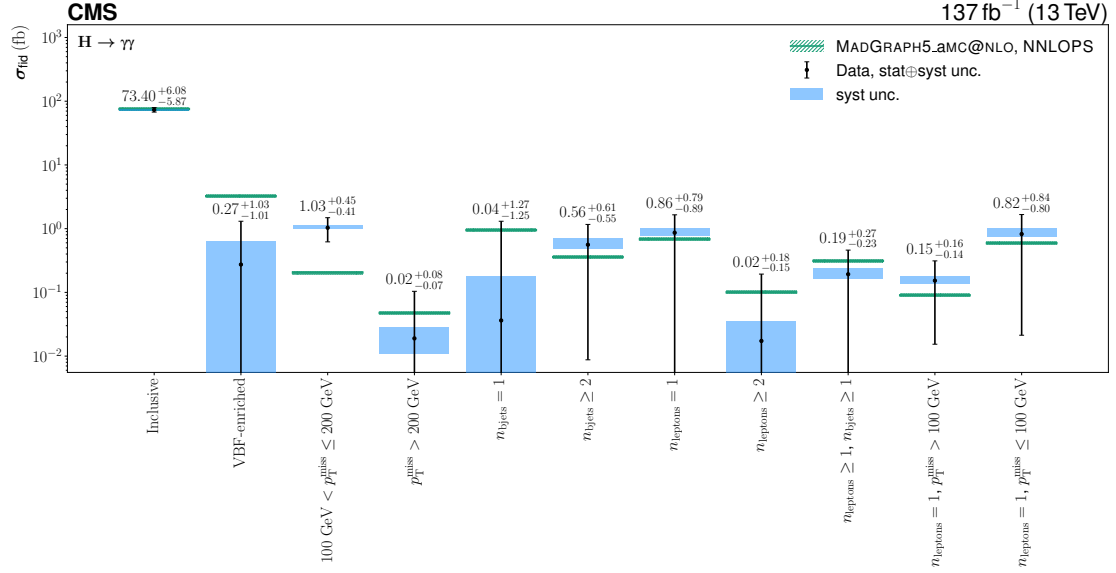


Figure 6-120 Cross sections of $H \rightarrow \gamma\gamma$ in the selected regions of fiducial phase space.

The correlation matrix for the fiducial cross sections per particle level bin of $p_T^{\gamma\gamma}$ are shown in Figure 6-121 and n_{jets} can be seen in Figure 6-122.

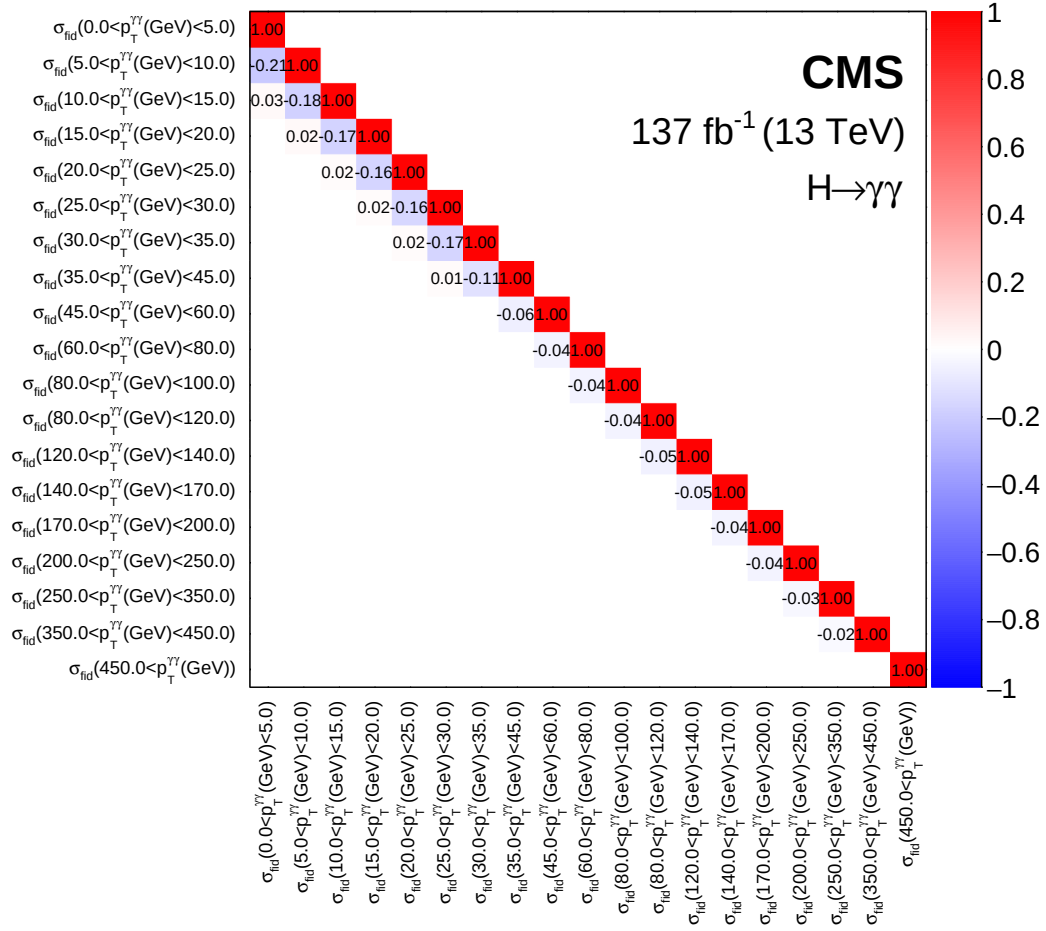


Figure 6-121 Correlation matrix for the cross sections per particle level bin for $p_T^{\gamma\gamma}$, extracted by performing simultaneous maximum likelihood fit for cross sections.

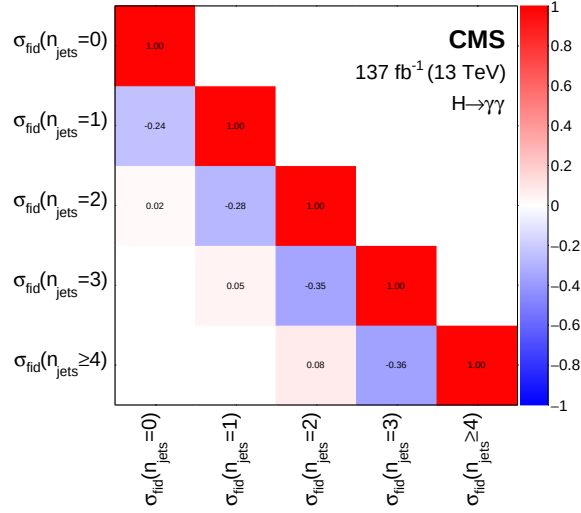


Figure 6-122 Correlation matrix for the cross sections per particle level bin for n_{jets} , extracted by performing simultaneous maximum likelihood fit for cross sections.

The measurement results for differential fiducial cross section w.r.t. some of the observables (results for all of the observables can be accessed in Ref. [26]) i.e. $p_T^{\gamma\gamma}$ (GeV) and n_{jets} are shown in Figure 6-123.

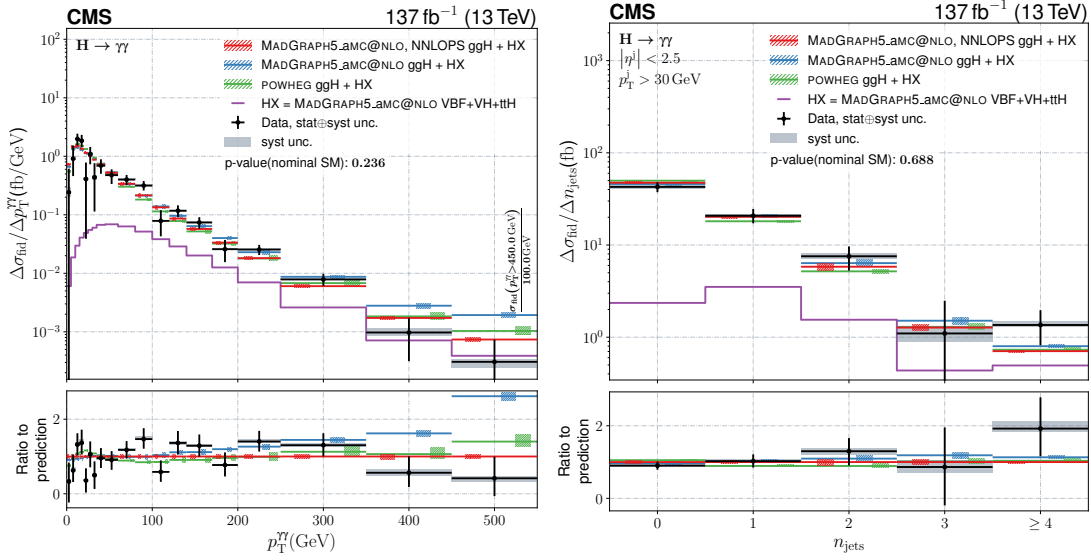


Figure 6-123 Differential fiducial cross sections for $p_T^{\gamma\gamma}$ and n_{jets} . The black dots represent the observed values of differential fiducial cross sections and full uncertainty is shown with vertical bars.

Figure 6-124 shows the differential fiducial cross section results for some of the other observables such as $|y^{\gamma\gamma}|$, $|\cos\theta^*|$, $|\phi_\eta^*|$ and τ_C^j .

The results for the measurement of double-differential cross section w.r.t. p_T and n_{jets}

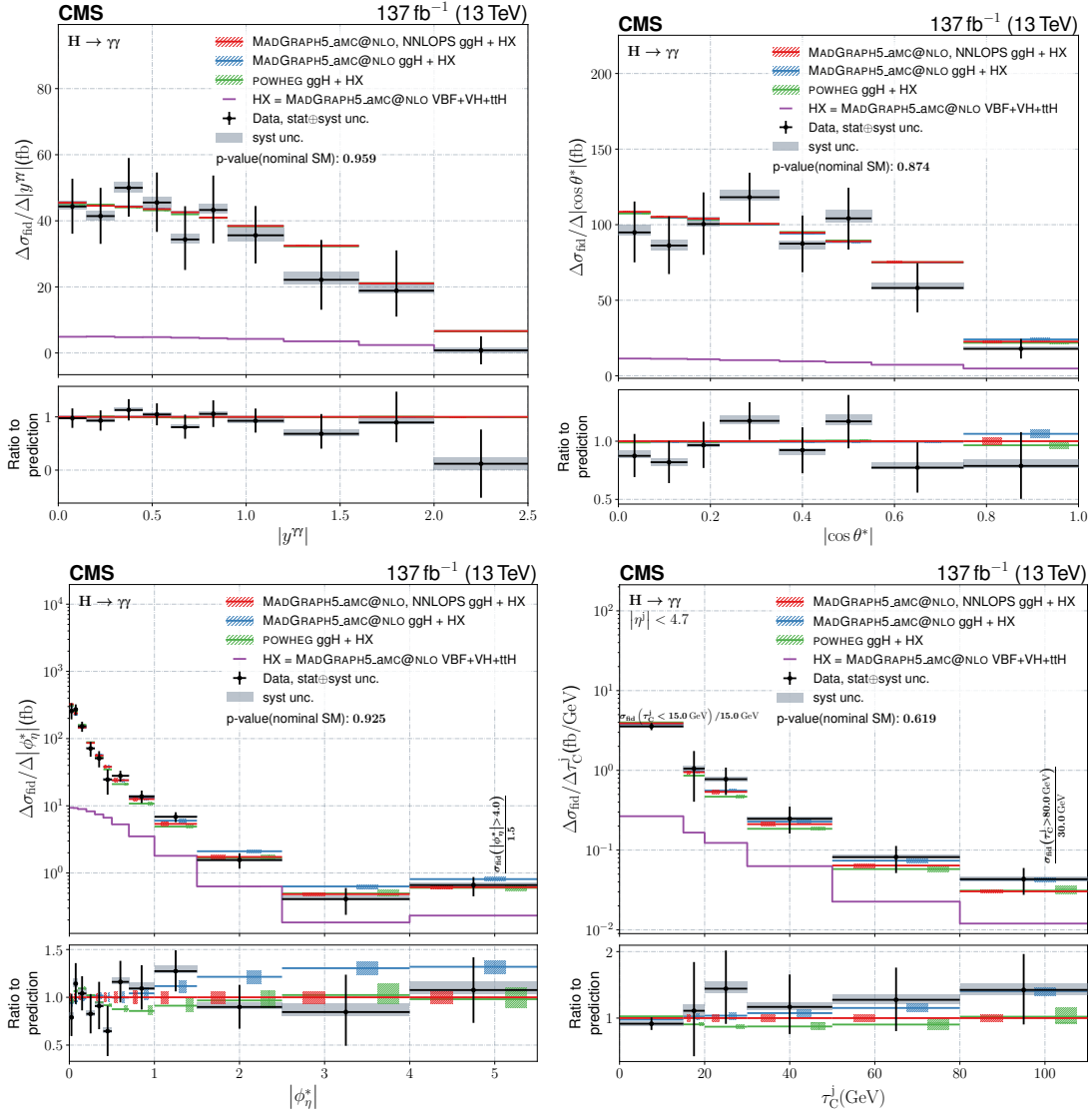


Figure 6-124 Differential fiducial cross sections for $|y^{\gamma\gamma}|$, $|\cos\theta^*|$, $|\phi_\eta^*|$ and τ_C^j observables. The black dots represents the observed values of differential fiducial cross sections and full uncertainty is shown with vertical bars.

as well as p_T and τ_C^j are presented in Figure 6-125 and Figure 6-126 respectively. The comparison of the measurement with the prediction along with various additional predictions are also shown.

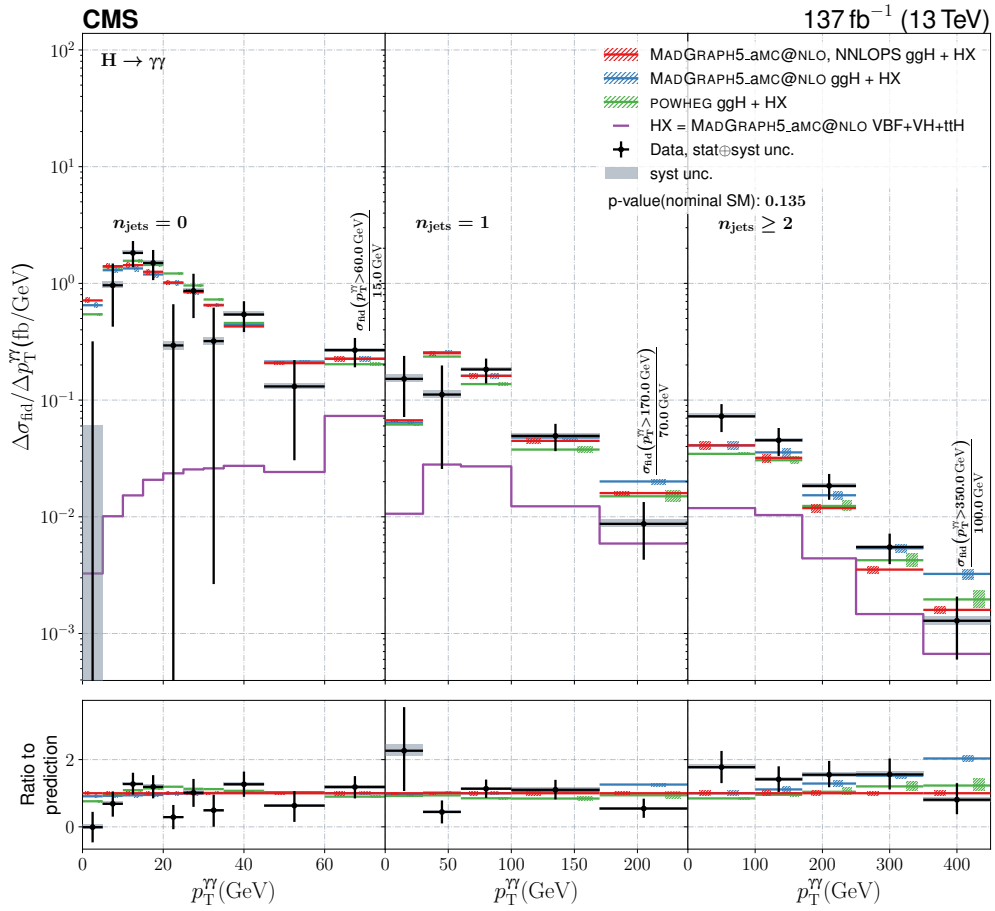


Figure 6-125 Measured double-differential fiducial cross section in bins $p_T^{\gamma\gamma}$ and n_{jets} .

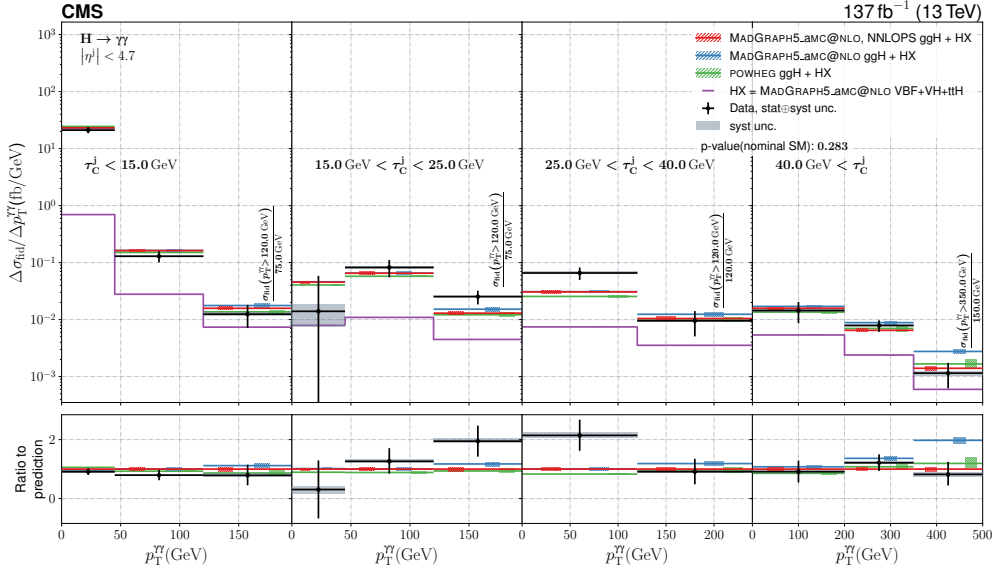


Figure 6-126 Measured double-differential fiducial cross section in bins $p_T^{\gamma\gamma}$ and τ_C^j .

The measured fiducial cross section is $\sigma_{\text{fid}} = 73.40^{+5.4}_{-5.3}(\text{stat})^{+2.4}_{-2.2}(\text{syst}) \text{ fb} = 73.40^{+6.1}_{-5.9} \text{ fb}$. The observed fiducial cross section in the previous measurement done with 2016 (35.9 fb^{-1}) data was $\sigma_{\text{fid}} = 84 \pm 11(\text{stat}) \pm 7(\text{syst}) \text{ fb} = 84 \pm 13 \text{ fb}$ [133]. Overall, the studies carried out for the measurements of differential fiducial cross section in p-p collisions, for the production of Higgs boson, are observed to be in accordance with the expectation of SM within the uncertainties.

Chapter 7: Summary and Outlook

7.20 Summary

The long awaited Higgs boson was finally discovered in 2012 by ATLAS and CMS collaborations at CERN which is the eventual verification of the EWSB (electroweak symmetry breaking) in standard model. After this discovery, many studies were triggered in order to precisely measure the newly discovered particle to check its nature as well to probe for new physics in case of some deviations. So far the measurements made with the available LHC data, the observed Higgs boson is compatible with the predictions of SM.

The Standard Model is a theory which describes all known elementary particles and the forces responsible for their interactions but not fully able to address some fundamental subjects such dark matter, asymmetry of matter and anti-matter and gravity etc. The predictions made by SM are extensively tested to high precision. The Higgs boson is a vital piece of the SM as its presence is needed to describe the origin of masses for bosons and fermions. Therefore, studies related to the precision measurements of the Higgs boson and its properties, is among one of the priorities in modern day particle physics research.

The CMS is a detector with solenoidal magnet producing 3.8T magnetic field. It has greatly efficient tracking system, calorimeter (ECAL and HCAL) along with a powerful muon system which make it capable to precisely reconstruct a variety of final state objects. The CMS trigger system reduce the rate of collisions with efficient processing to a controllable level.

The sub-detectors information has been utilized by the CMS detector in order to reconstruct various kinds of particles with the help of a method known as particle flow. The photon candidates created by particle flow algorithm have to undergo a set of requirements before they enter into the analysis. These certain requirements are established on the shapes of deposited ECAL showers and how they are isolated relative to other particles. The corrected energies are used for the photons which eventually enhance the mass resolution of diphoton events in the studies. A multivariate classifier is trained known as the photon ID BDT to discriminate photons from jets. The jets are

employed to distinguish various Higgs boson production mechanisms.

The BSM physics is also providing a Higgs state that is in agreement with the 125 GeV boson along with the extended space of some BSM models like, Next-to-Minimal Supersymmetric Model (NMSSM), the generalized two-Higgs-doublet models (2HDM) and the Georgi-Machacek (GM) model. These models may lead to the possibility of Higgs bosons and predict some of them may have masses below 125 GeV. In this context, the results were presented for the search of an additional Higgs boson in the mass range 70 to 110 GeV in the diphoton final state with 2016, 2017, 2018 datasets and their combinations. The observed upper limit for the combined data set ranges from 73 fb to 15 fb. One excess with approximately 2.9σ local (1.3σ global) significance is observed for a mass hypothesis of 95.4 GeV. Till date, there are no hints of any new Higgs boson. For this analysis, compared to the previous published results [14], we have updated the strategy to deal with relic DY background and also employed a new method to deal with significant bias around 90 GeV. We have optimized various elements relative to the previous results which includes the photon ID MVA training, diphoton BDT training and the combined MVA training dedicated with low-mass samples. I have contributed significantly in the search of additional standard model-like Higgs boson. I have been doing the diphoton BDT training and related validations ($Z \rightarrow ee$ and diphoton data/MC) followed by the boundary optimization with 2016 legacy data set. As we have used the 2017 combine MVA training in 2018, so I did the corresponding validations ($Z \rightarrow ee$ and diphoton data/MC) and checked the significance. I have also performed the preselection studies, signal and background modelling, systematic studies with 2018 data and performed many iterations of the analysis. Moreover, I have produced all of the statistical results for 2016 (legacy), 2017, 2018, the full Run 2 combination leading to upper-limits and p-values (local and global) and updated corresponding documentation [134]. The expected limit from ATLAS's result is ≈ 185 fb at 70 GeV and ≈ 73 fb at 110 GeV with 80 fb^{-1} data [74]. We get 98/102/82 fb from 2016/2017/2018 separately at 70 GeV and 40/47/33 fb from 2016/2017/2018 separately at 110 GeV. With the 2016 (legacy) and 2017 combination corresponding to 78.1 fb^{-1} which is close to ATLAS's 80 fb^{-1} data, we get the expected upper limit of about 70 fb at 70 GeV and about 31 fb at 110 GeV. From the full Run2 combination with 132.2 fb^{-1} , we get the expected upper limit of 53 fb at 70 GeV and 22 fb at 110 GeV. CMS has much better

sensitivity than ATLAS.

Many of the properties measured for the SM Higgs boson in the diphoton final state are presented in this thesis such as, signal strength, couplings and inclusive and differential fiducial cross section. The observed signal strength (137 fb^{-1}) is $\mu = 1.12^{+0.09}_{-0.09}$. We have reduced uncertainty as compare to previous CMS results with 2016 (35.6 fb^{-1}) data i.e. $\mu = 1.18^{+0.17}_{-0.14}$ [105]. CMS has comparable results with the ones from ATLAS (139 fb^{-1}) i.e. $\mu = 1.04^{+0.10}_{-0.09}$ [135]. The measured fiducial cross section is $\sigma_{\text{fid}} = 73.40^{+5.4}_{-5.3}(\text{stat})^{+2.4}_{-2.2}(\text{syst}) \text{ fb} = 73.40^{+6.1}_{-5.9} \text{ fb}$. The observed fiducial cross section in the previous measurement done with 2016 (35.9 fb^{-1}) data was $\sigma_{\text{fid}} = 84 \pm 11(\text{stat}) \pm 7(\text{syst}) \text{ fb} = 84 \pm 13 \text{ fb}$ [133]. All of these observations are compatible with the expectations of the SM. In the signal strength and coupling studies, an updated and improved strategy for measurements of ggH and VBF production mechanism is employed along with additional categories targeting associated production (VH). It also includes the dedicated analysis category for production in association with a single top quark (tH) for the first time as well as a wide range of measurements and interpretations. In the analysis regarding inclusive and differential fiducial cross section, the strategy is mostly unchanged w.r.t. the previous rounds, however several enhancements have been made. The most important ones are the corrections of the input variables of the photon ID MVA, which reduced the systematic uncertainty coming from the mis-modelling of the photon ID MVA output in simulation by more than 50%. The set of observables with respect to which the $H \rightarrow \gamma\gamma$ cross-section has been extended by two one-dimensional observables, i.e. ϕ_{η}^* and τ_C^j and one double differential set of bins, p_T vs. τ_C^j . The number of bins in each observables has also been increased. These studies were carried out by keeping in mind to minimize the theory dependence. The precision measurements are crucial in a fact that, any deviations may lead to the new physics. In the studies related to Higgs boson properties measurements, I have contributed to common tools for the analyses of Higgs boson in the diphoton decay channel. I performed the studies of electron veto efficiencies and scale factors.

7.21 Outlook

The $H \rightarrow \gamma\gamma$ decay mode has a small ($\approx 0.23\%$) branching ratio but it provides a clean final state topology where a mass peak can be reconstructed precisely. The

measurement of m_H in this channel can be combined with other decay modes to achieve higher precision. There are certain studies that can be carried out and are not presented here for the measurement of Higgs boson properties and with the availability of more data, they can be done and eventually lead to reduction in the systematic uncertainties. In this

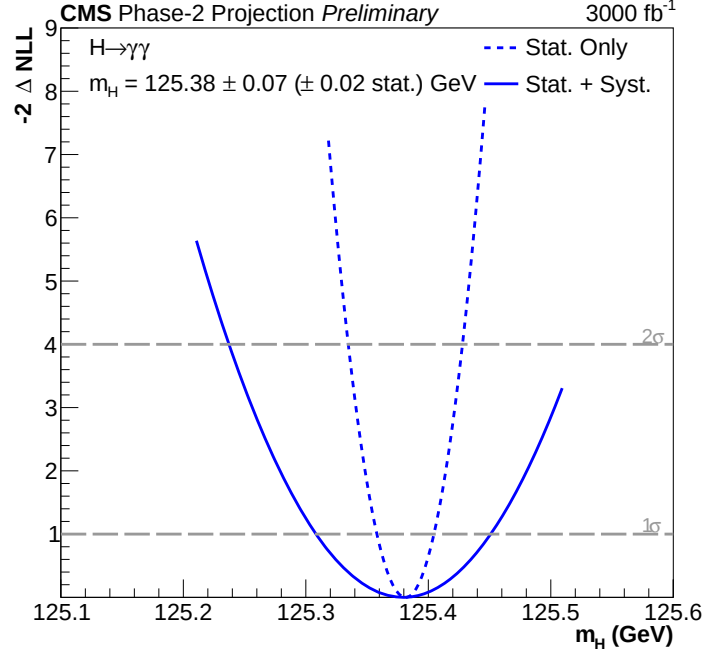


Figure 7-127 (Unweighted) The likelihood scan of the expected Higgs boson mass in diphoton decay mode for an integrated luminosity of 3000 fb^{-1} . The solid lines are for the full likelihood scan including all systematic uncertainties, while the dashed lines shows the same with the statistical uncertainty only [34].

context a study has been done with a projection of the measurement of m_H in diphoton decay mode at the HL-LHC for an integrated luminosity of 3000 fb^{-1} . In order to extract the m_H precisely, a binned maximum likelihood fit is performed to the $m_{\gamma\gamma}$ distributions in all analysis categories simultaneously within the range $100 < m_{\gamma\gamma} < 180 \text{ GeV}$. In the likelihood scan shown in Figure 7-127, systematic uncertainties are included. To extract m_H , in the fit, two independent signal strengths $\mu_{ggH+ttH}$ and μ_{VBF+VH} are free to vary. The best-fit mass of the Higgs boson is expected to be $m_H = 125.38 \pm 0.02 \text{ (stat)} \pm 0.07 \text{ (syst) GeV}$ at an integrated luminosity projected to 3000 fb^{-1} with the improved CMS detector at the HL-LHC. The projected precision of 0.07 GeV on the mass measurement in diphoton decay mode is better by a factor of 3 relative to current measurement with the 2016 dataset.

In future, with more data, the precision measurements may lead to some new endeavours if the deviations from SM expectations would be seen. In the meantime, the searches are ongoing based on the BSM physics based models could also lead us to new observations. In the context of low-mass $H \rightarrow \gamma\gamma$ search, more data (such as Run 3, for which the data taking is on going) might be able to conclude the excess, possible mass extension with boosted events and ZH production model ($Z \rightarrow \mu\mu$ with dimuon HLT), and possible constraints on the BSM models [8, 31, 32].

Bibliography

- [1] Glashow S L. The renormalizability of vector meson interactions [J]. Nuclear Physics, 1959, 10: 107-117.
- [2] Salam A. Weak and electromagnetic interactions [J]. 1994: 244-254.
- [3] Evans L, Bryant P. Lhc machine [J]. Journal of instrumentation, 2008, 3(08): S08001.
- [4] Aad G, Abajyan T, Abbott B, et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc [J]. Physics Letters B, 2012, 716(1): 1-29.
- [5] Chatrchyan S, Khachatryan V, Sirunyan A M, et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc [J]. Physics Letters B, 2012, 716(1): 30-61.
- [6] Franke F. Neutralinos and higgs bosons in the next-to-minimal supersymmetric standard model [J]. International Journal of Modern Physics A, 1997, 12(03): 479-533.
- [7] Jia-Wei F, Jun-Quan T, Yu-Qiao S, et al. Phenomenology of the lightest higgs boson in the next-to-minimal supersymmetric standard model [R]. 2013.
- [8] Tao J Q, Shahzad M A, Zhang S J, et al. Search for a lighter higgs boson in the next-to-minimal supersymmetric standard model [J]. Chinese Physics C, 2018, 42(10): 103107.
- [9] Chang S, Kang S K, Lee J P, et al. Two higgs doublet models for the lhc higgs boson data at 7 and 8 tev [J]. Journal of High Energy Physics, 2014, 2014(9): 1-17.
- [10] Cacciapaglia G, Deandrea A, Gascon-Shotkin S, et al. Search for a lighter higgs boson in two higgs doublet models [J]. Journal of High Energy Physics, 2016, 2016(12): 1-25.
- [11] Georgi H, Machacek M. Doubly charged higgs bosons [J]. Nuclear Physics B, 1985, 262(3): 463-477.
- [12] Logan H E, Rantala V. All the generalized georgi-machacek models [J]. Physical Review D, 2015, 92(7): 075011.
- [13] for Higgs T L W G, Collaboration A, Collaboration D, et al. Search for the standard model higgs boson at lep [J]. Physics Letters B, 2003, 565: 61-75.
- [14] Sirunyan A M, Tumasyan A, Adam W, et al. Search for a standard model-like higgs boson in the mass range between 70 and 110 gev in the diphoton final state in proton-proton collisions at 8 and 13 tev [J]. Physics Letters B, 2019, 793: 320-347.
- [15] Glashow S L. Partial-symmetries of weak interactions [J]. Nuclear physics, 1961, 22(4): 579-588.
- [16] Weinberg S. A model of leptons [J]. Physical review letters, 1967, 19(21): 1264.
- [17] Salam A. Elementary particle theory (p. 367) [J]. Stockholm, Swed.: Almquist & Wiksells, 1969.
- [18] Fermi E. Trends to a theory of beta radiation [J]. Nuovo Cim, 1934, 11(1): 535.

- [19] Englert F, Brout R. Broken symmetry and the mass of gauge vector mesons [J]. Physical review letters, 1964, 13(9): 321.
- [20] Gell-Mann M. A schematic model of baryons and mesons [J]. Phys. Letters, 1964, 8.
- [21] Evans L, Bryant P, et al. Lhc machine. jinst 3: S08001 [Z]. 2008.
- [22] CMS-Collaboration. The cms experiment at the cern lhc [J]. JINST, 2008, 3: S08004.
- [23] Hervé A, Blau B, Brédy P, et al. Status of the construction of the cms magnet [J]. IEEE transactions on applied superconductivity, 2004, 14(2): 542-547.
- [24] Hervé A. The cms detector magnet [J]. IEEE transactions on applied superconductivity, 2000, 10(1): 389-394.
- [25] Sirunyan A M, Tumasyan A, Adam W, et al. Measurements of higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{13}$ tev [J]. Journal of High Energy Physics, 2021, 2021(7): 1-87.
- [26] CMS-Collaboration. Measurement of the higgs boson inclusive and differential fiducial production cross sections in the diphoton decay channel with pp collisions at 13 tev [J]. arXiv preprint arXiv:2208.12279, 2022.
- [27] Group L H C S W, et al. Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector [J]. arXiv preprint arXiv:1610.07922, 2016, 35.
- [28] Berger N, et al. Simplified template cross sections – stage 1.1: LHCHXSWG-2019-003, DESY-19-070 [R]. 2019.
- [29] CMS-Collaboration, CMS-Collaboration, et al. Measurements of higgs boson production via gluon fusion and vector boson fusion in the diphoton decay channel at 13 tev [J]. Physics Analysis Summary CMS-PAS-HIG-18-029, CDS ID, 2019, 2667225.
- [30] de Florian D, Fontes D, Quevillon J, et al. Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector: arXiv: 1610.07922 [M]. CERN, 2016.
- [31] Wang C, Tao J Q, Shahzad M A, et al. Search for a lighter neutral custodial fiveplet scalar in the georgi-machacek model [J]. Chinese Physics C, 2022, 46(8): 083107.
- [32] Cacciapaglia G, Deandrea A, Gascon-Shotkin S, et al. Search for a lighter higgs boson in two higgs doublet models [J]. Journal of High Energy Physics, 2016, 2016(12): 1-25.
- [33] cms-publication-committee-chair@ cern. ch C C. A portrait of the higgs boson by the cms experiment ten years after the discovery [J]. Nature, 2022, 607(7917): 60-68.
- [34] CMS-Collaboration. Higgs to $\gamma\gamma$ mass and width measurements: CMS-PAS-FTR-21-008 [R/OL]. Geneva: CERN, 2022. <https://cds.cern.ch/record/2804042>.
- [35] Clowe D, Bradač M, Gonzalez A H, et al. A direct empirical proof of the existence of dark matter [J]. The Astrophysical Journal, 2006, 648(2): L109.
- [36] Wendell R, Ishihara C, Abe K, et al. Atmospheric neutrino oscillation analysis with subleading effects in super-kamiokande i, ii, and iii [J]. Physical Review D, 2010, 81(9): 092004.
- [37] Martin S P. A supersymmetry primer [M]//Perspectives on supersymmetry. World Scientific, 1998: 1-98.

- [38] Aad G, Abbott B, Abbott D C, et al. Measurements of the higgs boson inclusive and differential fiducial cross-sections in the diphoton decay channel with pp collisions at at 13 tev with the atlas detector [J]. *Journal of High Energy Physics*, 2022, 2022(8): 1-96.
- [39] Khachatryan V, Sirunyan A M, Tumasyan A, et al. Observation of the diphoton decay of the higgs boson and measurement of its properties [J]. *The European Physical Journal C*, 2014, 74(10): 1-49.
- [40] Schneider G, Mooser A, Bohman M, et al. Double-trap measurement of the proton magnetic moment at 0.3 parts per billion precision [J]. *Science*, 2017, 358(6366): 1081-1084.
- [41] Valente M. The standard model of particle physics [J]. *Springer International Publishing*, 2022: 5-25.
- [42] Banner M, Battiston R, Bloch P, et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider [J]. *Physics Letters B*, 1983, 122(5-6): 476-485.
- [43] Arnison G, Astbury A, Aubert B, et al. Experimental observation of lepton pairs of invariant mass around 95 gev/c² at the cern sps collider [J]. *Physics Letters B*, 1983, 126(5): 398-410.
- [44] Bagnaia P, Banner M, Battiston R, et al. Evidence for $z^0 \rightarrow e^+ e^-$ at the cern pp collider [J]. *Physics Letters B*, 1983, 129(1-2): 130-140.
- [45] Thomson M. Modern particle physics [M]. *Cambridge University Press*, 2013.
- [46] Englert F, Brout R. Broken symmetry and the mass of gauge vector mesons [J]. *Physical review letters*, 1964, 13(9): 321.
- [47] Kibble T W. Symmetry breaking in non-abelian gauge theories [J]. *Physical Review*, 1967, 155(5): 1554.
- [48] Guralnik G S, Hagen C R, Kibble T W. Global conservation laws and massless particles [J]. *Physical Review Letters*, 1964, 13(20): 585.
- [49] Higgs P W. Broken symmetries, massless particles and gauge fields [J]. *Phys. Lett.*, 1964, 12: 132-133.
- [50] Higgs P W. Broken symmetries and the masses of gauge bosons [J]. *Physical Review Letters*, 1964, 13(16): 508.
- [51] Higgs P W. Spontaneous symmetry breakdown without massless bosons [J]. *Physical review*, 1966, 145(4): 1156.
- [52] Grojean C, Maltoni F, Mariotti C, et al. Handbook of lhc higgs cross sections: 4. deciphering the nature of the higgs sector [J]. 2017.
- [53] Aad G, Abajyan T, Abbott B, et al. Measurements of higgs boson production and couplings in diboson final states with the atlas detector at the lhc [J]. *Physics Letters B*, 2013, 726(1-3): 88-119.
- [54] Khachatryan V, Sirunyan A, Tumasyan A, et al. Precise determination of the mass of the higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 tev [J]. *European Physical Journal C*, 2015, 75(5).

- [55] collaboration A, et al. Measurements of the higgs boson production and decay rates and constraints on its couplings from a combined atlas and cms analysis of the lh c pp collision data at $\sqrt{s}=7$ and 8 tev [J]. The Journal of High Energy Physics, 2016, 2016(8): 45.
- [56] Wang A Z. Observation of higgs boson production in association with a top quark pair at the lh c with the atlas detector [R]. ATL-COM-PHYS-2018-679, 2018.
- [57] Aaboud M, Aad G, Abbott B, et al. Observation of $h \rightarrow b\bar{b}$ decays and vh production with the atlas detector [J]. Physics Letters B, 2018, 786: 59-86.
- [58] CMS-Collaboration, et al. Observation of higgs boson decay to bottom quarks [J]. Physical Review Letters, 2018.
- [59] Group L H C S W, et al. Handbook of lh c higgs cross sections: 4. deciphering the nature of the higgs sector [J]. arXiv preprint arXiv:1610.07922, 2016, 35.
- [60] Heinemeyer S, et al. The lh c higgs cross section working group collaboration, handbook of lh c higgs cross sections: 3. higgs properties [J]. arXiv preprint arXiv:1307.1347.
- [61] Khachatryan V, et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV [J]. Eur. Phys. J. C, 2015, 75: 212.
- [62] Chatrchyan S, et al. Evidence for the direct decay of the 125 GeV Higgs boson to fermions [J]. Nature Phys., 2014, 10: 557.
- [63] Sirunyan A M, et al. Observation of the Higgs boson decay to a pair of leptons with the CMS detector [J]. Phys. Lett. B, 2018, 779: 283.
- [64] Sirunyan A M, et al. Observation of Higgs boson decay to bottom quarks [J]. Phys. Rev. Lett., 2018, 121: 121801.
- [65] Sirunyan A M, et al. Observation of $t\bar{t}H$ production [J]. Phys. Rev. Lett., 2018, 120: 231801.
- [66] Sirunyan A M, et al. Evidence for Higgs boson decay to a pair of muons [J]. JHEP, 2021, 01: 148.
- [67] Sirunyan A M, et al. A measurement of the Higgs boson mass in the diphoton decay channel [J]. Phys. Lett. B, 2020, 805: 135425.
- [68] Aaboud M, et al. Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector [J]. Phys. Rev. D, 2018, 98: 052005.
- [69] Aad G, et al. Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector [J]. Eur. Phys. J. C, 2020, 80: 957.
- [70] Aad G, et al. Measurements of h and hA production in the $h \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector [J]. Eur. Phys. J. C, 2021, 81: 178.
- [71] CMS-Collaboration, et al. Measurements of production cross sections of the higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13$ [J]. The European Physical Journal C, 2021, 81.

- [72] Ghezzi M, Gomez-Ambrosio R, Passarino G, et al. Nlo higgs effective field theory and κ -framework [J]. Journal of High Energy Physics, 2015, 2015(7): 1-149.
- [73] Palacino G. Direct searches for additional scalar higgs at atlas [R]. ATL-COM-PHYS-2019-516, 2019.
- [74] Search for resonances in the 65 to 110 GeV diphoton invariant mass range using 80 fb^{-1} of pp collisions collected at $\sqrt{s} = 13$ TeV with the ATLAS detector: ATLAS-CONF-2018-025 [R]. Geneva: CERN, 2018.
- [75] CERN L. Design report: Vol. 2 [J]. The LEP Main Ring, 1984.
- [76] Brüning O S, Collier P, Lebrun P, et al. Paul proudlock lhc design report [J]. CERN, Geneva, 2004, 3.
- [77] Acharya B, Alexandre J, Bernabéu J, et al. The physics programme of the moedal experiment at the lhc [J]. International Journal of Modern Physics A, 2014, 29(23): 1430050.
- [78] Anelli G, Antchev G, Aspell P, et al. The totem experiment at the cern large hadron collider [J]. Journal of Instrumentation, 2008, 3(08): S08007.
- [79] Berti E. The lhcf experiment: present status and physics results [J]. arXiv preprint arXiv:1710.03991, 2017.
- [80] Apollinari G, Brüning O, Nakamoto T, et al. High luminosity large hadron collider hl-lhc [J]. arXiv preprint arXiv:1705.08830, 2017.
- [81] CMS-Collaboration, et al. Description and performance of track and primary-vertex reconstruction with the cms tracker [J]. Journal of Instrumentation, 2014, 9(10): P10009.
- [82] Tumasyan A, CMS-Collaboration, et al. The phase-2 upgrade of the cms tracker [R]. Technical Report CERN-LHCC-2017-009, CMS-TDR-014, 2017.
- [83] Orimoto T J. The cms electromagnetic calorimeter: Construction, commissioning and calibration [J]. Proceeding of Science, 2009: 129.
- [84] Canelli M F, Aarrestad T K, Brzhechko D, et al. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at 13 tev [J]. Journal of Instrumentation, 2018, 13(06): P06015.
- [85] CMS-Collaboration, et al. Cms, the compact muon solenoid. muon technical design report [J]. 1997.
- [86] CMS-Collaboration. Motivation of the cms muon system upgrade with triple-gem detectors [J]. Nuclear and particle physics proceedings, 2016, 273: 2506-2508.
- [87] Costantini S. Background in the cms muon detectors: simulation and measurements with pp collision data [J]. Journal of Instrumentation, 2020, 15(02): C02038.
- [88] Bunkowski K, Krolikowski J. Optimization, Synchronization, Calibration and Diagnostic of the RPC PAC Muon Trigger System for the CMS detector: CMS-TS-2010-009 [R/OL]. Geneva: CERN, 2009. <http://cds.cern.ch/record/1308715>.
- [89] Paolucci P. The CMS Muon system: CMS-CR-2006-006 [R/OL]. Geneva: CERN, 2005. <http://cds.cern.ch/record/927394>.

- [90] Fisk I. First Experiences with LHC Grid Computing and Distributed Analysis [C]//21st Hadron Collider Physics Symposium. 2010.
- [91] Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector: CMS-PAS-PFT-10-001 [R]. 2010.
- [92] Sirunyan A M, et al. Particle-flow reconstruction and global event description with the cms detector [J]. JINST, 2017, 12: P10003.
- [93] CMS-Collaboration. Performance of photon reconstruction and identification with the cms detector in proton-proton collisions at $\sqrt{s} = 8$ tev [J]. Journal of Instrumentation, 2015, 10 (8): P08010.
- [94] Sirunyan A M, Tumasyan A, Adam W, et al. Electron and photon reconstruction and identification with the cms experiment at the cern lh [J]. Journal of Instrumentation, 2021, 16(5): P05014.
- [95] Sirunyan A M, CMS-Collaboration, et al. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ tev [J]. Journal of Instrumentation, 2018, 13(06): P06015.
- [96] CMS-Collaboration, et al. Energy calibration and resolution of the cms electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ tev [J]. arXiv preprint arXiv:1306.2016, 2013.
- [97] Sirunyan A M, Tumasyan A, Adam W, et al. Measurements of higgs boson production cross sections and couplings in the diphoton decay channel at 13 tev [J]. Journal of High Energy Physics, 2021, 2021(7): 1-87.
- [98] Daubie E, Beliy N, Caebergs T, et al. Performance of electron reconstruction and selection with the cms detector in proton-proton collisions at 8 tev [J]. Journal of Instrumentation, 2015, 10.
- [99] Canelli M F, Aarrestad T K, Brzhechko D, et al. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at 13 tev [J]. Journal of Instrumentation, 2018, 13(06): P06015.
- [100] Kirschenmann H, CMS-Collaboration, et al. Determination of the jet energy scale in cms [C]//Journal of Physics: Conference Series: volume 404. IOP Publishing, 2012: 012013.
- [101] Cacciari M, Salam G P, Soyez G. The anti-kt jet clustering algorithm [J]. Journal of High Energy Physics, 2008, 2008(04): 063.
- [102] Khachatryan V, Sirunyan A M, Tumasyan A, et al. Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev [J]. Journal of Instrumentation, 2017, 12(02): P02014.
- [103] CMS-Collaboration, et al. First results on higgs to $\gamma\gamma$ at 13 tev [J]. CMS Physics Analysis Summary CMS-PAS-HIG-15-005, 2016.
- [104] Sirunyan A M, Tumasyan A, Adam W, et al. Measurements of higgs boson properties in the diphoton decay channel in proton-proton collisions at 13 tev [J]. Journal of High Energy Physics, 2018, 2018(11): 1-60.

- [105] CMS-Collaboration. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV [J]. JHEP, 2018, 11: 185.
- [106] Sirunyan A M, Tumasyan A, Adam W, et al. Measurements of higgs boson properties in the diphoton decay channel in proton-proton collisions at 13 tev [J]. Journal of High Energy Physics, 2018, 2018(11): 1-60.
- [107] Sirunyan A M, CMS-Collaboration, et al. Particle-flow reconstruction and global event description with the cms detector [J]. JINST, 2017, 12(10): P10003.
- [108] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python [J]. Journal of Machine Learning Research, 2011, 12: 2825-2830.
- [109] Khachatryan V, et al. Observation of the diphoton decay of the Higgs boson and measurement of its properties [J]. Eur. Phys. J., 2014, C74(10): 3076.
- [110] Dauncey P D, Kenzie M, Wardle N, et al. Handling uncertainties in background shapes: the discrete profiling method [J]. JINST, 2015, 10(04): P04015.
- [111] Sirunyan A M, et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS [J]. Eur. Phys. J. C, 2021, 81(9): 800.
- [112] CMS-Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV: CMS-PAS-LUM-17-004 [R/OL]. CERN, 2018. <https://cds.cern.ch/record/2621960>.
- [113] CMS-Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV: CMS-PAS-LUM-18-002 [R/OL]. CERN, 2019. <https://cds.cern.ch/record/2676164>.
- [114] Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s} = 13$ TeV: CMS-PAS-HIG-19-015 [R/OL]. Geneva: CERN, 2020. <https://cds.cern.ch/record/2725142>.
- [115] Cowan G, Cranmer K, Gross E, et al. Asymptotic formulae for likelihood-based tests of new physics [J]. The European Physical Journal C, 2011, 71: 1-19.
- [116] Junk T. Confidence level computation for combining searches with small statistics [J]. Nucl. Instrum. Meth. A, 1999, 434: 435.
- [117] Read A L. Presentation of search results: The CL(s) technique [J]. J. Phys. G, 2002, 28: 2693.
- [118] Demortier L. P values and nuisance parameters [C/OL]//Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007. 2008: 23. <http://cds.cern.ch/record/1099967/files/p23.pdf>.
- [119] Sirunyan A M, Tumasyan A, Adam W, et al. Measurements of $t\bar{t}h$ production and the cp structure of the yukawa interaction between the higgs boson and top quark in the diphoton decay channel [J]. Physical review letters, 2020, 125(6): 061801.
- [120] Chatrchyan S, Khachatryan V, Sirunyan A M, et al. Measurement of the inclusive w and z production cross sections in pp collisions at 7 tev with the cms experiment [J]. Journal of High Energy Physics, 2011, 2011(10): 1-76.
- [121] Alwall J, Frederix R, Frixione S, et al. The automated computation of tree-level and next-to-

- leading order differential cross sections, and their matching to parton shower simulations [J]. Journal of High Energy Physics, 2014, 2014(7): 1-157.
- [122] Hamilton K, Nason P, Re E, et al. Nnlops simulation of higgs boson production [J]. Journal of High Energy Physics, 2013, 2013(10): 1-25.
- [123] Sjöstrand T, Ask S, Christiansen J R, et al. An introduction to pythia 8.2 [J]. Computer physics communications, 2015, 191: 159-177.
- [124] Ball R D, Bertone V, Carrazza S, et al. Parton distributions for the lhc run ii [J]. Journal of High Energy Physics, 2015, 2015(4): 1-148.
- [125] Hartanto H B, Jäger B, Reina L, et al. Higgs boson production in association with top quarks in the powheg box [J]. Physical Review D, 2015, 91(9): 094003.
- [126] Gleisberg T, Höche S, Krauss F, et al. Event generation with sherpa 1.1 [J]. Journal of High Energy Physics, 2009, 2009(02): 007.
- [127] Dauncey P, Kenzie M, Wardle N, et al. Handling uncertainties in background shapes: the discrete profiling method [J]. Journal of Instrumentation, 2015, 10(04): P04015.
- [128] Fisher R A. On the interpretation of χ^2 from contingency tables, and the calculation of p [J]. Journal of the royal statistical society, 1922, 85(1): 87-94.
- [129] Khachatryan V, Sirunyan A M, Tumasyan A, et al. Measurement of differential cross sections for higgs boson production in the diphoton decay channel in pp collisions at 8 tev [J]. The European Physical Journal C, 2016, 76: 1-31.
- [130] Collins J C, Soper D E. Angular distribution of dileptons in high-energy hadron collisions [J]. Physical Review D, 1977, 16(7): 2219.
- [131] Boggia M, Cruz-Martinez J M, Frellesvig H, et al. The higgstools handbook: a beginners guide to decoding the higgs sector [J]. Journal of Physics G: Nuclear and Particle Physics, 2018, 45(6): 065004.
- [132] Gangal S, Stahlhofen M, Tackmann F J. Rapidity-dependent jet vetoes [J]. Physical Review D, 2015, 91(5): 054023.
- [133] CMS-Collaboration. Measurement of inclusive and differential higgs boson production cross sections in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ tev [J]. Journal of High Energy Physics, 2019, 2019(1).
- [134] CMS-Collaboration. Search for low mass resonances in the diphoton final state in pp collisions at 13 TeV with the full Run 2 dataset [J/OL]. Physics Analysis Summary CMS-PAS-HIG-20-002, 2023. <http://cds.cern.ch/record/2852907>.
- [135] collaboration A, et al. Measurement of the properties of higgs boson production at $\sqrt{s} = 13$ tev in the $h \rightarrow \gamma\gamma$ channel using 139fb^{-1} collision data with the atlas experiment [J]. arXiv preprint arXiv:2207.00348, 2022.

Acknowledgements

In the name of ALLAH, the most beneficent and merciful. I am thankful to Almighty ALLAH for giving me strength, ability and opportunity to accomplish this dissertation. I offer my respectful and sincere words to the Holy Prophet Muhammad (PBUH), who is forever a torch of guidance for humanity.

It was not possible to perform the research included in this dissertation without the patience, motivation and enthusiasm of many individuals. I would like to first, express my gratitude to my advisors Prof. Guoming Chen and Prof. Junquan Tao for mentoring me during my graduate studies. Their supervision led me to undertake and significantly contribute to $H \rightarrow \gamma\gamma$ analysis. I am sincerely thankful to them for their continuous help over the course of the analysis and the writing of this thesis. I am also thankful to other faculty members Prof. Mingshui Chen, Prof. Huaqiao Zhang, Prof. Hongbo Liao, Prof. Jin Wang and others for their nice suggestions on my work.

I want to express my sincere gratitude to my parents, brothers, sisters and my wife (Ms. Saadia Abbas) without whose love, support and understanding, it was not possible for me to complete this doctoral degree. Additionally, I would like to thank Dr. Muhammad Ahmad, Dr. Tahir Javaid, Dr. Zhang Sijing (my senior fellow and great friends at the institute) for their support in research.

This research would not have been possible without the assistance and useful discussions from Prof. Dr. Susan Gascon Shotkin, Dr. Antoine Lesauvage, Dr. Camille Camen, and other experts and fellows in the IPNL group. I am really grateful to Prof. Bernard Ille, for all the help he offered as an elder to face all kinds of problems during my stay in IPNL, now IP2I.

At last, I thank the CMS collaboration and all its members for the insightful discussions and opportunities presented to myself.

Author's resume, academic papers and other related academic achievements published during the degree program

Resume of the author:

September, 2004 – July, 2007, B.Sc at University of the Punjab, Lahore, Pakistan.

September, 2007 – August, 2009, M.Sc at University of the Punjab, Lahore, Pakistan.

September, 2010 – August, 2012, MS at International Islamic University, Islamabad, Pakistan.

September, 2017 – June, 2023, pursuing doctoral degree at Institute of High Energy Physics, University of Chinese academy of sciences.

Published (or officially accepted) academic papers:

- (1) CMS Collaboration. “Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV” . In: *Physics Analysis Summary CMS-PAS-HIG-20-002*, CDS ID 2852907 (2023). URL:<http://cds.cern.ch/record/2852907>
- (2) CMS Collaboration. “Measurement of the Higgs boson inclusive and differential fiducial production cross sections in the diphoton decay channel with pp collisions at $\sqrt{s} = 13$ ” . In: *arXiv preprint arXiv:2208.12279 (2022) (accepted by JHEP)*.
- (3) Chu Wang, Jun-Quan Tao, M. Aamir Shahzad, Guo-Ming Chen, S. Gascon-Shotkin, “Search for a lighter neutral custodial fiveplet scalar in the Georgi-Machacek model” , *Chin. Phys. C*, 2022, 46(8): 083107-083107.
- (4) CMS collaboration. “Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ ” . In: *JHEP07 (2021) 027*.
- (5) Jun-Quan Tao, M. Aamir Shahzad, Si-Jing Zhang, Chu Wang, Yu-Qiao Shen, Guo-Ming Chen, He-Sheng Chen, S. Gascon-Shotkin, M. Lethuillier, L. Finco, C. Camen, “Search for a lighter Higgs Boson in the Next-to-Minimal Supersymmetric Standard Model” , *Chin. Phys. C*, 2018, 42(10): 103107

Conference talks:

- (1) Presented a talk in “Physics Plenary session of the CMS week” held at CERN from 30th of January to 3rd of February, 2023, on the topic “Low mass $H \rightarrow \gamma\gamma$ search status with full Run 2 data” .
- (2) Presented a talk in “8th China LHC Physics Conference (CLHCP 2022)” held in Nanjing Normal University, China from 23-27 November, 2022, on the topic “Measurement of Higgs boson inclusive and differential fiducial production cross sections in the di-photon decay channel at CMS.
- (3) Presented a talk in “13th France China Particle Physics Laboratory workshop (FCPPL 2021)” held in France from 13-16 December, 2021, on the topic “Measurements of Higgs boson properties in the di-photon decay channel at CMS” .
- (4) Presented a talk in “7th China LHC Physics Conference (CLHCP 2021)” held in China from 25-28 November, 2021, on the topic “Measurements of Higgs boson production cross-sections and couplings in the di-photon decay channel at 13 TeV” .
- (5) Presented a talk in “Higgs 2021 Conference” held in Stony Brook University, New York, United States from 18th October to 22nd of October on the topic “Measurements of Higgs production and search for an additional SM-like Higgs boson in the di-photon decay channel with the CMS detector” .
- (6) Presented a talk in “CMS China Conference” held in Hangzhou, China from 28 to 29 May, 2021 on the topic “Search for a Standard Model-like Higgs Boson in the mass range 70 and 110 GeV in the di-photon final state at CMS (Compact Muon Solenoid) Experiment” .
- (7) Presented a talk in “Higgs Coupling 2019” held in Oxford University, United kingdom from 30th September to 4th of October on the topic “Search for additional low-mass ($m < 125$ GeV) Higgs Bosons at CMS (Compact Muon Solenoid) Experiment” .
- (8) Presented a talk in “Third China LHC Physics Conference (CLHCP 2017)” held at Nanjing University, China from 22-24 December, 2017, on the topic “Measurement of Higgs Boson in the di-photon decay channel” .