Measurements of Higgs boson properties in the four-lepton channel in pp collisions at centre-of-mass energy of 13 TeV with the CMS detector

Šćulac, Toni

Doctoral thesis / Disertacija

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:354582

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-25**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

NNT : 2018SACLX092

Thèse de doctorat

Measurements of Higgs boson properties in the four-lepton channel in pp collisions at centre-of-mass energy of 13 TeV with the CMS detector

Thèse de doctorat de l'Université Paris-Saclay et de l'Université de Zagreb
préparée à l'École polytechnique

Ecole doctorale n°576 Particules, Hadrons, Énergie, Noyau, Instrumentation, Imagerie, Cosmos et Simulation (PHENIICS)
Spécialité de doctorat : Physique des particules

Thèse présentée et soutenue à Palaiseau, le 05/12/2018, par

ŠĆULAC TONI

Composition du Jury :

Vacavant Laurent Directeur de Recherches, CNRS, Marseille	Président
Kortner Sandra Associate Professeur, Max-Planck-Institut für Physik, München	Rapporteur
De Naurois Mathieu Professeur, LLR, Palaiseau	Examinateur
Kumerički Krešimir Professeur, Faculty of Science, Zagreb	Examinateur
Sirois Yves Directeur de recherche, LLR, Palaiseau	Directeur de thèse
Puljak Ivica Professeur, FESB, Split	Co-directeur de thèse

University of Zagreb

University Paris-Saclay
École Polytechnique

University of Zagreb
Faculty of Science

TONI ŠĆULAC

Measurements of Higgs boson properties in the four-lepton channel in pp collisions at centre-of-mass energy of 13 TeV with the CMS detector

International dual doctorate

President of the Jury	M. LAURENT VACAVANT
Referee	MME. SANDRA KORTNER
Examinators	M. KREŠIMIR KUMERIČKI
	M. MATHIEU DE NAUROIS
Thesis supervisor	M. YVES SIROIS
Thesis co-supervisor	M. IVICA PULJAK

Palaiseau, France, 2018.

University of Zagreb

University Paris-Saclay
École Polytechnique

Sveučilište Zagreb
Prirodoslovno-matematički fakultet

TONI ŠĆULAC

Mjerenja svojstava Higgsovog bozona kroz kanal raspada u četiri leptona u pp sudarima s energijom u centru mase od 13 TeV detektorom CMS

Internacinalni dvojni doktorat

Predsjednik Povjerenstva	M. LAURENT VACAVANT
Ocjenjivač	MME. SANDRA KORTNER
Povjerenstvo	M. KREŠIMIR KUMERIČKI
	M. MATHIEU DE NAUROIS
Mentor	M. YVES SIROIS
Komentor	M. IVICA PULJAK

Palaiseau, Francuska, 2018.

University of Zagreb

Université Paris-Saclay
École Polytechnique

University of Zagreb
Faculty of Science

TONI ŠĆULAC

Mesure des propriétés du boson de Higgs via ses désintégrations en quatre leptons dans les collisions pp à 13 TeV dans le détecteur CMS

International Cotutelle

Président du Jury	M. LAURENT VACAVANT
Rapporteur	MME. SANDRA KORTNER
Exeminateurs	M. KREŠIMIR KUMERIČKI
	M. MATHIEU DE NAUROIS
Directeur de thèse	M. YVES SIROIS
Co-directeur de thèse	M. IVICA PULJAK

Palaiseau, France, 2018.

Acknowledgements

Please forgive me if I missed some names, as there are a lot of people who have helped me in different ways to reach this very important milestone in my scientific journey.

I would like to start with my friends and colleagues from the CMS group at Laboratoire Leprince Ringuet. Thank you **Florian**, **Roberto**, **Jean-Baptiste**, **Olivier**, **Stephanie**, **Arthur**, **Philippe**, and **Giacomo**, you have helped me a lot and made my stay in Paris very enjoyable.
Special thanks goes to **Claude**, you have selflessly invested a lot of your time to teach me about the world of electrons in CMS.
Christophe, thank you a lot for helping me master the world of Higgs and becoming my supervisor from the shadows. While learning a lot from you I have also enjoyed hanging out with you and consider you my friend.
If I have to choose one person from the group I had most fun with, it has to be **Alex**! We had some great times together and I look forward for more to come.
A big thanks goes to my fellow PhD students **Luca**, **Chiara**, **Marina**, **Christina**, **Iurii**, and **Jonas**. I have enjoyed our fruitful discussions and all the fun times we had, you made me feel at home. I have to single out my friend **Philipp** who is a brilliant young mind, I always enjoy learning from. You are a strong gym buddy that inspired me to start working out again. **Simon**, thank you for helping me with every single step of my journey and for being a great friend as well.

Thank you **Duje**, **Darko**, **Damir**, **Dunja**, **Ilja**, **Nikola**, **Suri**, **Bojan**, and **Stipe**, my FESB colleagues and friends. Special thanks goes to **teta Anita**, as it would not be possible for me to enjoy my journey without your enormous help with all the paperwork. **Marko**, you have helped me so much in every possible way and thank you a lot for that, my friend.

I would like to also thank my Johns Hopkins colleagues **Andrei**, **Heshy**, **Ulascan**, and **Meng** for a lot of physics related discussions and collaboration.

Thank you **Sandra Kortner** and **Laurent Vacavant**, my referees, for carefully reading and reviewing my thesis and to other jury members **Krešimir Kumerički** and **Mathieu de Naurois** for your time invested to understand my work and for making my thesis

defense enjoyable.

None of this would be possible without the guidance of my supervisors **Yves Sirois** and **Ivica Puljak**. Thank you for trusting in me and giving me this golden opportunity. I have enjoyed working with you and have learned so much from you. Thank you very much for your unconditional support over the years and for providing me with exceptional working conditions. I consider myself extremely lucky to have you as my supervisors and friends and I hope that I can continue learning from you in many years to come.

I would like to thank all my friends who are always there for me and are a big part of my life. You have helped me relax and take my mind off science, which I find very important.

And last, but certainly not the least, I would like to thank my family for their unconditional love, support and help. In this journey I have traveled to many beautiful places all over the world, but I still find it special to return home to my three ladies **Minja**, **Ana**, and **Rita**.

I would like to dedicate this thesis to my biggest supporter, my grandma. You were the one to recognize my talent from the early age and were always supportive. Following my journey closely, you have always celebrated my success and enjoyed listening to my stories. I will never forget the great times we had together and the unconditional love you have given me.
Hvala ti na svemu bakice!

Summary

This thesis presents the measurements of Higgs (H) boson properties in the four-lepton decay channel in proton-proton collisions at a centre-of-mass energy of 13 TeV with the CMS detector. The discovery of the Higgs boson is considered as the ultimate verification of the Standard Model. It solves the problem of the origin of elementary particles mass and provides unitarization at high energy. After the discovery was announced by the ATLAS and CMS Collaborations further studies of the properties of the Higgs boson were performed exploiting Run I data at collision energies of 7 and 8 TeV. While the discovery relied on a combination of studies in five different decay channels, two of them provided most of the sensitivity and a measurement of the H boson mass: it's decay to a pair of photons and the decay to a pair of Z bosons that both decay into pairs of electrons or muons.

For the Run II at 13 TeV, significant improvements in many aspects of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis are introduced with respect to the previously reported measurements by the CMS Collaboration. Because of its large signal-to-background ratio and complete reconstruction of all final state particles with an excellent resolution, the channel allows for precise measurements of Higgs boson properties. It is often referred to as the "golden channel". My contribution to the analysis is twofold, improving the electron selection efficiency measurements and leading the analysis of data recorded in 2017 as the main author.

For each selection step of the analysis, the efficiency has to be carefully measured in data and in simulation and observed differences need to be accounted for with the application of the scale factors. I have derived scale factors for electrons in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis using the Tag-and-Probe method. While performing my studies I have introduced several improvements and my results were also used by many other analyses in the CMS Collaboration.

There are several important steps in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis. It all starts with the trigger that is designed to have the highest possible efficiency for signal events. Triggered events are studied and the ones with four leptons that pass specially designed selection criteria are used to build ZZ candidates. Selected events are further classified into event categories targeting different H boson production mechanisms. For each

event category and final state, reducible and irreducible backgrounds are estimated and the signal is modelled and parametrized as a function of the H boson mass. Three main observables providing good separation between signal and background are defined and used to build likelihoods used in the statistical analysis: the four-lepton invariant mass, matrix element kinematic discriminants, and per-event mass uncertainties. Finally, each possible source of the systematic uncertainties is carefully studied and its effect is measured and accounted for in the maximum likelihood fits used to extract the results.

Results of the analysis of Run II data are presented for two data sets recorded in 2016 and 2017, corresponding to integrated luminosities of $35.9\ \mathrm{fb}^{-1}$ and $41.5\ \mathrm{fb}^{-1}$, respectively. The signal strength modifiers relative to the Standard Model prediction, the mass and decay width of the boson, differential and inclusive cross sections are measured using 2016 data. All results are found to be in good agreement with the Standard Model expectations for a Higgs boson within the uncertainties. I was the main author of the analysis of 2017 data that yielded a measurement of the signal strength modifiers and simplified template cross sections. Finally, I also performed a measurement of the signal strength modifiers on combined $77.4\ \mathrm{fb}^{-1}$ of data collected in 2016 and 2017 that is again found to be compatible with the Standard Model predictions. A simultaneous fit to all channels, combining 2016 and 2017 data, resulted in a measured global signal strength modifier of

$$\mu = \sigma/\sigma_{\mathrm{SM}} = 1.06 \pm 0.10(\mathrm{stat})^{+0.08}_{-0.06}(\mathrm{exp.syst})^{+0.07}_{-0.05}(\mathrm{th.syst}) = 1.06^{+0.15}_{-0.13}$$

at $m_{\mathrm{H}} = 125.09$ GeV. New frontiers are reached as the statistical and systematical contributions to the total uncertainty on the inclusive signal strength measurement are considerably reduced and of the same size.

Résumé

Cette thèse présente les mesures des propriétés du boson de Higgs avec le détecteur CMS, dans le canal de désintégration à quatre leptons, pour les collisions proton-proton à une énergie de 13 TeV au collectionneur LHC. La découverte du boson de Higgs, considérée comme une clef de voûte du modèle standard, résolve les problèmes d'origine de la masse des particules élémentaires et fournit une unitarisation exacte de la théorie à haute énergie. Après la découverte annoncée par les collaborations ATLAS et CMS, des études supplémentaires sur les propriétés du boson de Higgs ont été effectuées en exploitant les données du Run I. La découverte reposait sur une combinaison d'études dans cinq canaux de désintégration différents, dont deux fournissent l'essentiel de la sensibilité et une mesure de la masse du boson H : sa désintégration en une paire de photons et sa désintégration en une paire de bosons Z qui se désintègrent chacun en paire d'électrons ou de muons.

Pour le Run II à 13 TeV, d'importantes améliorations dans de nombreux aspects du logiciel pour l'analyse du canal $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ sont introduit par rapport aux analyses précédemment publiés par la collaboration CMS. En raison de son rapport signal / bruit important et grâce à la reconstruction complète de toutes les particules d'état final avec une excellente résolution, le canal permet des mesures précises des propriétés du boson de Higgs. Il est souvent qualifié de « canal en or ». Ma contribution à l'analyse est double. Elle concerne d'une part l'amélioration des mesures d'efficacité et de sélection des électrons et d'autre part le pilotage complet de l'analyse des données enregistrées en 2017 en tant qu'auteur principal. Pour chaque étape de sélection de l'analyse, l'efficacité doit être soigneusement mesurée, pour les données réelles ainsi que pour pour la simulation, et les différences observées doivent être prises en compte avec l'application des facteurs d'échelle. J'ai dérivé des facteurs d'échelle pour les électrons dans le canal d'analyse $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ en utilisant la méthode dite de « Tag-and-Probe ». Tout en effectuant ces études, j'ai présenté plusieurs améliorations et mes résultats ont également été utilisés par de nombreuses autres analyses au sein de la collaboration CMS.

Il y a plusieurs étapes importantes dans l'analyse $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$. Tout commence par le déclenchement conçu pour avoir la plus grande efficacité possible pour les événements de signal. Les événements acceptés sont ensuite étudiés et des critères de sélection spé-

cialement adaptés sont utilisés pour identifier ceux possédants quatre leptons. Ils seront utilisés pour construire des candidats ZZ. Les événements sélectionnés sont ensuite classés en catégories d'événements ciblant différents mécanismes de production du boson H. Pour chaque catégorie d'événement et état final, les bruits de fonds réductibles et irréductibles sont estimés, puis le signal est modélisé et paramétrisé en fonction de la masse du boson H. Trois observables principaux fournissant une bonne séparation entre le signal et le bruit de fond sont définis et utilisés pour construire les vraisemblances utilisées dans l'analyse statistique : la masse invariante à quatre leptons, des discriminants cinématiques basés sur des éléments de matrice et les incertitudes sur la mesure de masse par événement. Enfin, chaque source possible d'incertitudes systématiques est soigneusement étudiée puis son effet est mesurée et pris en compte dans l'ajustement par maximum de vraisemblance utilisé pour extraire les résultats.

Les résultats de l'analyse des données du Run II sont présentés pour deux ensembles de données enregistrés en 2016 et 2017, correspondant respectivement à des luminosités intégrées de $35.9\ \mathrm{fb}^{-1}$ et $41.5\ \mathrm{fb}^{-1}$. Les modificateurs d'intensité du signal par rapport à la prédiction du modèle standard, la masse et la largeur de désintégration du boson H, ainsi que les sections efficaces différentielles et inclusives sont mesurées à l'aide des données 2016. Tous les résultats sont en bon accord avec l'attendu pour le boson H du modèle standard, compte tenu des marges d'incertitudes. J'étais l'auteur principal de l'analyse des données de 2017 qui a donné une mesure des modificateurs d'intensité du signal et des sections efficaces de références simplifiées. Enfin, j'ai également effectué une mesure des modificateurs d'intensité combinant $77.4\ \mathrm{fb}^{-1}$ de données collectées en 2016 et 2017 et qui est à nouveau compatible avec les prévisions du modèle standard. Un ajustement simultané de tous les canaux, combinant les données de 2016 et de 2017, a donné une mesure globale pour le modificateur de force associé au signal de

$$\mu = \sigma/\sigma_{\mathrm{SM}} = 1,06 \pm 0,10(\mathrm{stat})^{+0,08}_{-0,06}(\mathrm{exp.syst})^{+0,07}_{-0,05}(\mathrm{th.syst}) = 1,06^{+0,15}_{-0,13}$$

à $m_{\mathrm{H}} = 125,09$ GeV. Une nouvelle frontière est atteinte alors que les contributions statistiques et systématiques sont considérablement réduites et contribuent de façon comparable désormais à l'incertitude totale pour la mesure d'intensité du signal.

Prošireni sažetak

Otkriće Higgsovog bozona se sigurno može svrstati među najvažnije događaje u povijesti fizike i često se smatra kao konačna potvrda Standardnog Modela (SM) elementarnih čestica, teorije koja objašnjava tri od četiri poznate sile. Ovaj rad prezentira istraživanja svojstava novootkrivene čestice kroz njen kanal raspada $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$, $\ell = e, \mu$. Istraživanje je provedeno od 2015. do 2018. godine i analizira $77.4\ \mathrm{fb}^{-1}$ podataka prikupljenih detektorom Compact Muon Solenoid (CMS, Kompaktni Mionski Solenoid).

Uvod u fiziku Higgsovog bozona i Standardni Model

Elementarne čestice u Standardnom Modelu su leptoni, kvarkovi, baždarni bozoni i Higgsov bozon. Elementarnu česticu karakteriziraju svojstva poput mase, naboja, spina i okusa. Leptoni i kvarkovi su grupirani u tri obitelji s po četiri fermiona koji imaju spin $\frac{1}{2}$.

Prenositelji sile u Standardnom Modelu su 3 masivna baždarna bozona Z, W^{+} i W^{-} za slabu silu, foton koji nema masu za elektromagnetsku silu, i osam gluona koji također nemaju masu i nositelji su jake sile. Posljednja otkrivena elementarna čestica je Higgsov bozon, masivna čestica spina nula.

Standardni Model je kvantna teorija polja koja ima lokalnu baždarnu simetriju grupe $SU(3)_C \times SU(2)_L \times U(1)_Y$. Osnovni elementi teorije su kvantna polja, a njihova dinamika je zadana Lagrangianom $\mathcal{L}$. Lagrangiane za jaku i elektroslabu interakciju gradimo tako da zahtijevamo lokalnu baždarnu invarijantnost na grupu simetrije Standardnog Modela. Takva teorija je izrazito uspješna i precizna u predviđanju interakcija elementarnih čestica, ali ima veliku manu da previđa kako niti jedna elementarna čestica nema masu. Grupa od šest znanstvenika: Englert, Brout, Higgs, Guralnik, Hagen i Kibble je u tri nezavisna članka predložila rješenje tog problema kroz primjenu spontanog narušenja simetrije elektroslabog sektora. Takav mehanizam predviđa postojanje novog kvantnog polja u teoriji i nove elementarne čestice Higgsovog bozona.

Postoje razni načini na koje se Higgsov bozon može proizvesti, a najvažniji u hadronskim

sudaračima čestica su: fuzija gluona, fuzija vektorskih bozona, produkcija uz vektorski bozon i produkcija uz gornji ili "vrh" kvark. Također, postoje razni raspadi Higgsa na elementarne čestice, a od bronjih mogućnosti najzanimljiviji za istraživanje su raspadi na par kvarkova, raspadi na par vektorskih bozona ili na par fotona. U ovom radu se razmatra takozvani "zlatni kanal" raspada u dva Z bozona koji se dalje raspadaju na ukupno četiri leptona. To je jedan od glavnih kanala u velikom otkriću Higgsovog bozona koji su 4. srpnja 2012. godine objavili A Toroidal LHC Apparatus (ATLAS) i CMS kolaboracije. Iako je vjerojatnost raspada Higgsovog bozona kroz zlatni kanal jako mala i iznosi samo 0.0124%, ovaj kanal odlikuju neke karakteristike koje ga čine posebnim, a to su: kompletna rekonstrukcija čestica na koje se raspada, vrlo dobra rezolucija kod mjerenja količine gibanja tih čestica i odličan omjer signalnih i pozadinskih događaja.

CMS eksperiment u Velikom Sudaraču Čestica u CERN-u

Europsko vijeće za nuklearna istraživanja (CERN) je međunarodna institucija za nuklearna istraživanja osnovana 1954. kojoj su glavni cilj fundamentalna istraživanja na području fizike elementarnih čestica. CERN je najveći znanstveni laboratorij na svijetu, ima najsloženije znanstvene mjerne instrumente i druge uređaje potrebne za istraživanja. Najvažniji instrumenti u CERN-u su akceleratori i detektori čestica. Jedan od tih akceleratora je veliki hadronski sudarač (LHC) koji leži u tunelu kružnog oblika, opsega 27 km, 175 metara ispod francusko-švicarske granice, nedaleko od Ženeve, Švicarska. LHC je sagrađen dizajniran i izgrađen u suradnji više od 10000 znanstvenika iz preko 100 zemalja kao i stotina sveučilišta i laboratorija. Dva glavna detektora na LHC-u, dizajnirana i izgrađena za potragu za Higgsovim bozonom, su ATLAS i CMS.

CMS je kompaktni mionski solenoid, što znači da su mu glavne karakteristike: male dimenzije s obzirom na masu, napredni sustav za detekciju miona i supravodljivi magnet u obliku zavojnice. Sastoji se od više manjih pod-detektora koji su dizajnirani za obavljanje specifičnih zadataka. Odmah oko točke u kojoj se sudaraju čestice nalazi se sustav za detekciju tragova koji je temeljen na ideji da elementarna čestica kada prođe kroz silicijev detektor ostavi električni signal. Nakon njega slijede elektromagnetski i hadronski kalorimetar kojima je cilj zadržati čestice i izmjeriti njihovu energiju. Supravodljivi magnet služi za zakretanje putanja nabijenih čestica što pomaže kod mjerenja količine gibanja i električnog naboja čestica. Konačno, najveći dio CMS detektora su mionske komore koje služe za detekciju miona. Mioni su elementarne čestice slične elektronima, dolaze iz druge obitelji elementarnih čestica i razlikuju se jedino po masi. Upravo zato što su masivniji od elektrona postoji jako mala vjerojatnost da će interagirati s materijalom iz elektromagnetskog kalorimetra i bez mionskih komora ne bismo ih mogli detektirati. Čestice se sudaraju svako 25 ns, a moguće je sačuvati samo nekoliko stotina Hz podataka, pa je dizajniran posebni sustav za okidanje (trigger) koji u jako kratkom vremenskom roku odlučuje je li neki događaj zanimljiv za pohranu ili ne. Detektori se stalno obnavljaju i unapređuju kako bi mogli s visokom efikasnosti snimati događaje. Upravo zato je na početku 2017. godine unutarnji sustav za detekciju tragova unaprijeđen, što je rezultiralo u boljoj mogućnosti razlučivanja elektrona i miona od pozadinskih događaja.

Događaji koji su pohranjeni idu dalje na rekonstrukciju fizikalnih objekata. Kompliciran

algoritmi su posebno dizajnirani tako da od šume podataka koja se sastoji od pogodaka u sustavu za detekciju tragova i mionskim komorama, te depozita energije u kalorimetrima rekonstruiraju fizikalne objekte i njihova svojstva. Nakon procesa rekonstrukcije imamo elektrone, mione, hadronske mlazove čestica (eng. jet) i nedostajuću transverzalnu energiju (eng. missing transverse energy) koja ukazuje na prisustvo neutrina. Naravno, proces rekonstrukcije nije 100% efikasan, a uz to pozadinski signali mogu biti pogrešno rekonstruirani kao fizikalni objekti. Upravo iz tog razloga analize koje koriste fizikalne objekte dobivene rekonstrukcijom moraju dizajnirati dodatne zahtjeve na takve objekte da bi osigurali što manje pozadine u svojim podatcima.

Odabir elektrona i mjerenja efikasnosti

Glavne čestice koje detektiramo u $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ su elektroni i mioni. Moj rad je doprinio unapređenju mjerenja efikasnosti odabira elektrona na stvarnim i simuliranim podatcima. Rekonstruirane elektrone prvo treba kalibrirati. Naime, detektor prolazi kroz stalne promjene poput gubljenja transparentnosti kristala u elektromagnetskom kalorimetru i kada takve efekte ne bismo uzeli u obzir mjerenje količine gibanja elektrona bilo bi pogrešno. Upravo iz tog razloga se pažljivom analizom podataka kalibrira količina gibanja elektrona u prikupljenim podatcima, ali i u simuliranim podatcima. Važan dio u fizici elementarnih čestica je Monte Carlo (MC) simulacija koja nam omogućava da predviđanja teorije prevedemo u očekivana mjerenja u našem detektoru.

Nakon što su elektroni kalibrirani potrebno je osmisliti uvjete koje ćemo zahtijevati prilikom njihovog odabira za analizu. Cilj takvih uvjeta je što veća efikasnost za prave elektrone, a da je efikasnost za pozadinu što manja. U ovom slučaju pozadina bi bili elektroni iz hadronskih mlazova čestica ili signali iz detektora koji su pogrešno rekonstruirani kao elektroni. Odabir elektrona se odvija u nekoliko koraka:

- Udaljenost od centra sudara, transverzalna količina gibanja i pseudorapiditet: set jednostavnih uvjeta na osnovne elektronske varijable,
- Identifikacija elektrona: analiza na desetke varijabli osjetljivih na različite vrste pozadina koje se kombiniraju uz pomoć tehnika stabala odluke (eng. boosted decision tree) kako bi se postiglo što bolje odvajanje signala od pozadine
- Izolacija elektrona: uvjet da se u neposrednoj blizini elektrona ne nalaze nikakve druge čestice.

Dok se za analizu podataka prikupljenih 2016. godine koristio ovakav pristup, u analizi 2017. je odabir elektrona unaprijeđen korištenjem izolacijskih varijabli u stablima odluke prilikom identifikacije. Ovakav pristup je za istu efikasnost signalnih elektrona znatno smanjio efikasnost pozadine.

Nakon što je napravljen odabir elektrona od velike je važnosti izmjeriti njegovu efikasnost na stvarnim podatcima i u simulaciji i ispraviti sve opažene razlike. Razvijena je standardna metoda "označi i ispitaj" (eng. tag and probe TnP) koja koristi događaje iz

standardne svijeće $\mathrm{Z} \to \ell^+\ell^-$ kako bi izgradila uzorak elektronskih kandidata. Prvo se iz odabranih podataka označe elektroni s vrlo strogim uvjetima koji osiguravaju izrazito veliku vjerojatnost da su to pravi elektroni pa se u tom istom događaju traže elektroni koji u paru s označenim elektronom imaju masu u blizini mase Z bozona. Tako dobivamo kolekciju elektrona na kojima možemo ispitivati efikasnost pojedinog odabira koja je omjer broja elektrona koji zadovoljavaju uvijete odabira i ukupnog broja elektrona za ispitivanje. Na taj način su mjerene efikasnosti odabira $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ analize, a postignuti rezultati su postali službeni rezultati CMS kolaboracije, te su korišteni u raznim drugim analizama u CMS-u. Efikasnost je mjerena na stvarnim i simuliranim podatcima i razlika između njih je ispravljena primjenom faktora za skaliranje (eng. scale factor).

Efikasnost se može mjeriti na razne načine, a jedan od najučestalijih je da se elektroni koji prolaze i ne prolaze zadani odabir prilagode (eng. fit) na neku analitičku funkciju. Prilikom promatranja raznih prilagodbi funkcija uočena je zanimljiva nakupina događaja za elektrone s određenim transverzalnom količinom gibanja. Detaljna ispitivanja na simulacijama su pokazala da su to signalni elektroni koji imaju lošije izmjerenu količinu gibanja zbog izračenog fotona ili zbog toga što su završili u pukotinama između kristala u elektromagnetskom kalorimetru. Kako bi se takvi elektroni bolje tretirali u analizi odlučeno je da se za elektrone iz pukotina posebno mjere faktori za skaliranje, a za elektrone koji su izračili foton je unaprijeđen algoritam za njihov oporavak (eng. FSR recovery algorithm).

Glavni elementi $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ analize

Početna točka analize su prikupljeni i simulirani podatci. Rad u ovoj tezi temelji se na $35.9\ \mathrm{fb}^{-1}$ podataka prikupljenih u 2016. godini i $41.5\ \mathrm{fb}^{-1}$ prikupljenih u 2017. Monte Carlo simulacijom generiraju se signalni i pozadinski procesi koje teorija predviđa. Uz glavne procese simuliraju se hadronizacija kvarkova i gluona, dodatne pozadinske interakcije i naposljetku se simulira prolazak svih tih čestica kroz CMS detektor. Za signal se simuliraju glavni načini proizvodnje predviđenih teorijom za Higgsov bozon mase 125 GeV. Postoje dvije vrste pozadina: nesvodljiva pozadina koja ima iste čestice u konačnom stanju, ali u procesu nije stvoren Higgsov bozon i svodljiva pozadina koja se dobiva zbog nesavršenosti detektora i pogrešne rekonstrukcije čestica. Nesvodljiva pozadina u ovoj analizi dolazi od fuzije gluona i anihilacije para kvark-antikvark koje proizvode par Z bozona koji se dalje raspadaju na ukupno četiri leptona. Svodljivu pozadinu je jako teško simulirati te se iz tog razloga procjenjuje iz podataka koristeći dvije nezavisne metode. Njen dominantan izvor su događaji koji sadrže jedan Z bozon i razne druge čestice koje mogu biti krivo detektirane pa se iz tog razloga grupno označavaju kao $\mathrm{Z} + \mathrm{X}$ pozadina.

Da bismo odabrali događaje za koje mislimo da bi u sebi mogle sadržavati Higgsov bozon koji se raspada u četiri leptona moramo postaviti neke uvjete. Na leptone postavljamo uvjete koji smanjuju pozadinu dok zadržavaju vrlo visoku efikasnost za signal. Leptoni koji prežive taj odabir se kombiniraju u Z bozone tako da sparujemo leptone suprotnog naboja i istog okusa e^+e^-, $\mu^+\mu^-$. U događajima koji imaju dva Z bozona tražimo do-

datne uvjete kako bismo uklonili pozadine koje je teško simulirati i na kraju ostajemo s događajima koji su kandidati za produkciju Higgsovog bozona.

Prikupljeni događaji se kategoriziraju u svrhu razlikovanja različitih načina produkcije Higgsovog bozona. Kategorije su dizajnirane tako da imaju visoku koncentraciju događaja koji su nastali traženom produkcijom Higgsovog bozona dok je doprinos ostalih produkcija što manji moguć. Koristimo nekoliko vrsti varijabli za određivanje kategorije događaja, a to su: varijable koje određuju broj dodatnih objekata poput hadronskih mlazova i dodatnih leptona u događaju i kinematičke diskriminante temeljene na računanju matričnih elemenata. Kinematičke diskriminante se računaju tako da se iz četveroimpulsa leptona i dodatnih objekata u događaju, koristeći Lagrangian Standardnog Modela, računaju amplitude, tj. matrični elementi. Te amplitude se računaju za signalne i pozadinske procese i predstavljaju vjerojatnost pojavljivanja određenog događaja.

Možda i najvažniji element svake analize je odabir varijabli koje omogućuju razlikovanje signala od pozadine i koje se koriste kasnije u statističkoj analizi i oređivanju konačnih rezultata. U $\mathrm{H} \to 4\ell$ analizi tri glavne varijable koje se koriste su: invarijantna masa četiri leptona, kinematička diskriminanta i pogreška izmjerene mase. Ovdje je potrebno naglasiti da postoji kinematička diskriminanta koja za računanje koristi samo varijable osjetljive na raspad Higgsovog bozona. Takva diskriminanta vrlo dobro razlikuje signal produkcije Higgsovog bozona od pozadine ali ne može razlučiti između različitih načina produkcije Higgsovog bozona. Iz tog razloga su kod analize podataka prikupljenih u 2017. godini razvijene dvije nove diskriminante koje za računanje koriste i četveroimpulse dodatnih hadronskih mlazova rekonstruiranih u događaju, te uz pomoć njih uspijevaju razlučiti fuziju vektorskih bozona i produkciju Higgsovog bozona uz vektorski bozon od dominante produkcije fuzijom gluona.

Signalna i pozadinska distribucija invarijantne mase četiri leptona u statističku analizu ulaze kao analitičke funkcije koje se dobivaju prilagodbom na podatke prikupljene iz simulacije. Kako bi analiza bila što fleksibilnija i mogla mjeriti masu Higgsovog bozona signalne funkcije su parametrizirane s obzirom na masu Higgsovog bozona.

Konačno, dodatne analize su potrebne kako bi se detaljno ispitali svi mogući uzroci sistematskih pogrešaka. Dominantni izvori s eksperimentalne strane su mjerenja efikasnosti odabira leptona, kalibracija količine gibanja leptona, mjerenja integriranog luminoziteta prikupljenih podataka, određivanje svodljive $\mathrm{Z} + \mathrm{X}$ pozadine, kalibracija hadronskih mlazova, itd. Uz eksperimentalne postoje i teoretske sistematske pogreške koje dolaze iz nesigurnosti teoretskih proračuna. Neke od njih su odabir partonske distribucijske funkcije, renormalizacijska i faktorizacijska skala, korekcije višeg reda, itd.

Svi elementi analize se ujedinjuju u statističkom modelu koji koristi pristup maksimalne vjerojatnosti (eng. maximum likelihood). U tom pristupu se izgradi funkcija vjerojatnosti (eng. likelihood function) $\mathcal{L}(\mathrm{podatci}|\theta)$, gdje θ označava parametre koje želimo izmjeriti. Parametri θ se određuju tako da maksimiziraju funkciju vjerojatnosti, ili intuitivno biramo one vrijednosti parametara koje čine izmjerene podatke najvjerojatnijim. Funkcije vjerojatnosti mogu biti jedno, dvo ili trodimenzionalne, ovisno o tome koliko varijabli koriste, a grade se od signalnih i pozadinskih funkcija invarijantne mase,

dvodimenzionalnih predložaka (eng. template) kinematičkih diskriminanti i analitičkih funkcija za pogrešku izmjerene mase.

Mjerenje svojstava Higgsovog bozona s podatcima prikupljenim 2016. godine

Podatci prikupljeni u 2016. godini su detaljno analizirani kroz gore opisani proces i izmjerena su svojstva Higgsovog bozona u zlatnom kanalu po prvi puta na novoj energiji od 13 TeV. Prikupljeno je $35.9\ \mathrm{fb}^{-1}$ događaja i oni su uspoređeni sa simulacijom kroz distribucije raznih varijabli i vrlo dobro slaganje je uočeno na punom rasponu invarijantne mase četiri leptona. Također, jako dobro slaganje podataka sa simulacijom je prisutno i kod predviđenog i opaženog broja događaja u pojedinim kategorijama.

Izmjerena je ukupna jakost signala (eng. signal strength), koja se definira kao omjer izmjerenog udarnog presjeka i udarnog presjeka predviđenog Standardnim Modelom, od $\mu = \sigma/\sigma_{\mathrm{SM}} = 1.05^{+0.15}_{-0.14}(\mathrm{stat})^{+0.11}_{-0.09}(\mathrm{syst}) = 1.05^{+0.19}_{-0.17}$ za Higgsov bozon mase $m_{\mathrm{H}} = 125.09$ GeV. Također, izmjereni su jačine signala za glavne načine produkcije Higgsovog bozona, kao i za fermionske doprinose i doprinose vektorskih bozona očekivanom udarnom presjeku. Sve izmjerene vrijednosti su u slaganju s predviđanjima Standardnog Modela unutar nesigurnosti mjerenja.

Nadalje, izmjeren je udarni presjek u volumenu mjerenja (eng. fiducial volume), koji je dizajniran tako da se olakša teorijska reprezentacija rezultata umanjivanjem detektorskih efekata. Izmjeren je udarni presjek od $\sigma_{\mathrm{fid}} = 2.92^{+0.48}_{-0.44}(\mathrm{stat})^{+0.28}_{-0.24}(\mathrm{syst})\,\mathrm{fb}$, što je u slaganju s predviđanjima iz Standardnog Modela $\sigma^{\mathrm{SM}}_{\mathrm{fid}} = 2.76 \pm 0.14\,\mathrm{fb}$. Udarni presjek je izmjeren i kao funkcija energije u centru mase sudara, transverzalne količine gibanja Higgsovog bozona, broja hadronskih mlazova u događaju i transverzalne količine gibanja vodećeg hadronskog mlaza u događaju. Ponovno su sva mjerenja u skladu s očekivanjima Standardnog Modela unutar neodređenosti mjerenja.

Konačno, izmjerena je masa Higgsovog bozona od $m_{\mathrm{H}} = 125.26 \pm 0.20(\mathrm{stat}) \pm 0.08(\mathrm{syst})$ GeV i širina raspada od $\Gamma_{\mathrm{H}} < 1.10$ GeV s razinom pouzdanosti (eng. confidence level) od 95%.

Mjerenje svojstava Higgsovog bozona s podatcima prikupljenim 2017. godine

Zlatni kanal se pokazao kao jedna od najjačih analiza u CMS kolaboraciji koja redovno objavljuje rezultate. Nastavljajući na takvu reputaciju, $41.5\ \mathrm{fb}^{-1}$ podataka prikupljenih 2017. godine je pažljivo analizirano i objavljeni su rezultati za ljetne konferencije u 2018. godini.

Usporedba prikupljenih podataka sa simulacijom je ponovno pokazala dobro slaganje uz mali višak opaženih signalnih događaja u kategoriji koja dominantno sadrži Higgsovo

bozon proizveden fuzijom gluona.

Izmjerena je ukupna jačina signala od $\mu = \sigma/\sigma_{\mathrm{SM}} = 1.10^{+0.14}_{-0.13}(\mathrm{stat})^{+0.11}_{-0.09}(\mathrm{syst}) = 1.10^{+0.19}_{-0.17}$ za Higgsov bozon mase $m_{\mathrm{H}} = 125.09$ GeV. Ponovno su izmjerene jačine signala za glavne načine produkcije Higgsovog bozona, kao i za fermionske kontribucije i kontribucije vektorskih bozona očekivanom udarnom presjeku. Sve izmjerene vrijednosti su profitirale od uvođenja novih kinematičkih varijabli koje koriste podatke o produkciji Higgsovog bozona, te je zaključak ponovno slaganje s predviđanjima Standardnog Modela unutar neodređenosti mjerenja.

Kombinacija rezultata s podatcima iz 2016. i 2017. godine

Kako bi se smanjila neodređenost mjerenja, odlučeno je izvršiti mjerenja jačine signala za ukupnu količinu podataka od $77.4\ \mathrm{fb}^{-1}$ prikupljenu u 2016. i 2017. godini. Analize su kombinirane tako da je izgrađena zajednička funkcija vjerojatnosti koja je korištena za maksimiziranje. Posebna pažnja je pridana tretiranju sistematskih nesigurnosti između dvaju analiza i odabran je konzervativni pristup potpunog koreliranja nesigurnosti.

Konačan rezultat je izmjerena ukupna jačina signala od:

$$\mu = \sigma/\sigma_{\mathrm{SM}} = 1.06 \pm 0.10(\mathrm{stat})^{+0.08}_{-0.06}(\mathrm{exp.syst})^{+0.07}_{-0.05}(\mathrm{th.syst}) = 1.06^{+0.15}_{-0.13} \qquad (1)$$

za Higgsov bozon mase $m_{\mathrm{H}} = 125.09$ GeV. Mjerenje je u slaganju s predviđanjima Standardnog Modela i ukupna pogreška je dostigla razinu kada su statistička i sistematska komponenta slične veličine. To je jako važan događaj za ovaj kanal koji je poznat po tome da nema velik broj događaja. Zahvaljujući konstantom unapređivanju analize to je omogućeno unatoč nikad težim uvjetima sudara s velikom količinom pozadinskih događaja koji znatno otežavaju analizu podataka. Slaganje s predviđanjima Standardnog Modela je uočeno i za izmjerene jačine signala za glavne načine produkcije Higgsovog bozona, kao i za fermionske kontribucije i kontribucije vektorskih bozona očekivanom udarnom presjeku.

Zaključak

Zlatni kanal je i na novoj energiji sudara od 13 TeV pokazao da su sva svojstva Higgsovog bozona u slaganju s predviđanjima Standardnog Modela, unutar neodređenosti mjerenja. Analiza se konstantno unapređuje kako bi donosila sve preciznije rezultate unatoč sve težim uvjetima sudara. Kako se rad oko ove teze bliži kraju tako se i prikupljanje podataka s CMS detektorom završava prije dugog gašenja (eng. long shutdown) koje je planirano pri kraju 2018. godine kako bi se detektor unaprijedio za budućnost. Bit će jako zanimljivo dočekati rezultate analize podataka prikupljenih u 2018. godini, kao i konačne količine podataka previđene do kraja rada LHC-a. Cilj mjerenja svojstava Higgsa je postići što veću preciznost kako bismo bili osjetljivi na moguća odstupanja od predviđanja Standardnog Modela koja bi nam omogućila portal u dosad nepoznatu fiziku.

Contents

Introduction

The discovery of the long sought Higgs boson is often described as the ultimate verification of the Standard Model (SM) of particle physics, a theory describing three of the four known fundamental forces. It was the last unobserved elementary particle, first mentioned in the 1960s in three independent papers from Englert and Brout, Higgs, and Guralnik, Hagen and Kibble. Without the Higgs field, the Standard Model was still a very successful theory yielding many precise theoretical calculations that all withstood numerous experimental checks. However, it had a big shortcoming not being able to explain why any of the elementary particles would have mass without breaking its main cornerstone, the local gauge symmetry. The other big shortcoming was that the theory was becoming non-pertubative at the TeV scale resulting in meaningless results such as probabilities that are larger than 100%. The brilliant idea of these scientists solves these problems in a very elegant way, with the application of the well-known concept of spontaneous symmetry breaking. It introduced a doublet of complex scalar fields to generate masses of elementary particles and was named the Brout–Englert–Higgs (BEH) mechanism. In 2013, the year after the Higgs boson was discovered, the Nobel Prize for physics was awarded to Englert and Higgs for the theoretical description of the BEH mechanism. This may be the best indicator of how important the Higgs boson is for particle physics and our understanding of the universe in general.

It took a lot of time and an immense effort from the scientific community to discover the Higgs boson. Everything started in the mid 1970s when first studies were made, exploring how the Higgs boson signature could be found in particle collision experiments. A big challenge was that the theory had no prediction for the mass of the Higgs boson. First dedicated Higgs searches were conducted at the Large Electron-Positron Collider (LEP) at Conseil Européen pour la Recherche Nucléaire (CERN) and Tevatron at Fermilab. However, all experiments came short of the discovery resulting in experimental exclusions of the possible Higgs boson mass. It is safe to say that at that time the existence of the Higgs boson became one of the inescapable questions and main motivation for the design and construction of the new energy frontier machine, the CERN Large Hadron Collider (LHC). The 27 km in circumference and as deep as 175 metre, LHC was intended to provide proton-proton collisions with a centre-of-mass energy up to 14 TeV, with the Compact Muon Solenoid (CMS) and A Toroidal LHC Apparatus

(ATLAS) detectors placed at its opposite sides as general multipurpose detectors, and a Large Ion Collider Experiment (ALICE) and the Large Hadron Collider beauty (LHCb) as detectors designed for specific measurements. ATLAS and CMS were primarily designed to search for the long awaited boson in a variety of production and decay channels. This huge effort that took more than two decades and thousands of scientists from all around the world working together fully paid off on 4^{th} July 2012, when the ATLAS and CMS Collaborations announced independent observations of a new boson with mass near 125 GeV compatible with the SM Higgs boson.

After the discovery was announced, CMS and ATLAS have compared a number of options for the spin-parity of the newly discovered particle and these all prefer zero spin and even parity, two fundamental criteria of a Higgs boson consistent with the SM. These results, coupled with the measured interactions of the new particle with other particles, strongly indicated that it is a Higgs boson with properties consistent with a minimal scalar sector. While the discovery relied on a combination of studies in five different decay channels, two of them provided most of the sensitivity and a measurement of the mass, the decay to a pair of photons and the decay to a pair of Z bosons that both decay into pairs of electrons or muons. This concluded an extremely fruitful first round of data taking, that lasted from 2009 to 2013 and is referred to as Run I, and allowed for a two year shutdown and upgrade of the LHC and both the CMS and ATLAS experiments in preparation for Run II.

The main goals for Run II, planned to last from 2015 to 2018, were set to be the discovery of all main Higgs production and decay modes and precision studies of the Higgs properties. Any deviations of measurements with respect to the theory prediction could indicate new physics beyond the SM. This thesis presents measurements of Higgs boson properties in the $\text{H} \rightarrow \text{ZZ}^{\star} \rightarrow 4\ell$, $\ell = e, \mu$ decay channel in the CMS detector with Run II data. Because of its large signal-to-background ratio and of the complete reconstruction of all final state particles with excellent resolution, that allow for precise measurements of its properties with only a handful of events, this channel is often referred to as the "golden channel". In order to cope with never before experienced harsh collision conditions, even the golden channel was in need of new analysis developments and several improvements in order to deliver a new set of most precise Higgs properties measurements.

The thesis is structured to follow the conceptual flow of the analysis, but also the time line of my journey as a Ph.D. student. It all starts with a brief theoretical introduction to the SM theory of particle physics and with main emphasis on the Higgs boson in Chapter 1. It is followed up with a Chapter 2 that introduces the reader to the CMS experiment at the LHC, dissecting it to parts in order to understand how it detects particles emerging from collisions. My work as a Ph.D. student started in late 2014, and I have spent almost a full year studying the theory of SM and advanced concepts of experimental high energy physics as part of my exams at the University of Zagreb. This gave me a solid background that allowed me to start my work for the CMS Collaboration in the $\text{H} \rightarrow \text{ZZ}^{\star} \rightarrow 4\ell$ analysis.

In late 2015 I have started a joint Ph.D. program at Ecole Polytechnique and produced a

first set of results for the CMS Collaboration with a study of electron identification efficiency measurements. When trying to disentangle electrons from background, it is very important to understand and correct for the differences that occur between recorded data and simulation prediction. My work was done for the $H \to ZZ^{\star} \to 4\ell$ analysis that uses very challenging low energy electrons, but also for the wide use in the CMS Collaboration, as presented in Chapter 3. Afterwards I have shifted my focus towards the understanding of the building blocks of the $H \to ZZ^{\star} \to 4\ell$ analysis discussed in Chapter 4. Results based on data collected in 2016, 2017, and their combination are presented in Chapter 5. I was becoming more involved in the analysis from late 2016 to late 2017, contributing in various aspects from the studies of data to simulation agreement of input objects used in the analysis to tunning of the analysis selection chain that provides inputs for the properties measurements. My work was recognized by the group and I was awarded the chance to present the pre-approval talk of the analysis that was later published as a first set of legacy Higgs properties measurements in the golden channel with $35.9\ \mathrm{fb}^{-1}$ of Run II data.

Next in line was the analysis of the $41.5\ \mathrm{fb}^{-1}$ of Run II data recorded in 2017. I was chosen for the main author and contact person of the analysis and was heavily involved in all aspects of the analysis. My main contributions come from reducible background estimation and signal normalization parametrization studies, running the analysis chain to produce data to simulation comparison plots, studying systematic uncertainties, and performing the measurement of signal strength modifiers. I was given the opportunity to present the approval talk, and afterwards the analysis was made public in a form of a CMS Public Analysis Summary (PAS) in June 2018. The results also included a combination with the previously published analysis based on 2016 data corresponding to a total of $77.4\ \mathrm{fb}^{-1}$ analyzed Run II data in the golden channel. Thanks to my involvement in the analysis I was awarded an opportunity to present these results to the public on the 39^{th} International Conference on High Energy Physics (ICHEP) that was held in Seoul in July 2018.

Chapter 1

Introduction to the Higgs boson in the Standard Model

We begin our journey with a hard task of summarizing current theoretical knowledge of elementary particles and in particular the Higgs sector, a final piece of this puzzle which was only recently discovered.

The Standard Model of particle physics (SM) is a gauge quantum field theory describing three of the four known fundamental forces in the universe, as well as classifying all known elementary particles. In order for elementary particles in the SM to acquire mass a new scalar field that breaks the symmetry of the theory had to be introduced, predicting a new particle called the Higgs boson. This chapter gives a brief introduction in the theory of the Standard Model of particle physics and the electroweak symmetry breaking, discusses Higgs boson phenomenology at the Large Hadron Collider (LHC) and finally introduces the reader with some distinctive features of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ decay channel ($\ell = e, \mu$).

1.1 Standard Model of particle physics

1.1.1 Elementary particles

All known elementary particles can be divided into two groups, according to their spin: fermions and bosons. Elementary particles in the SM are leptons, quarks, gauge bosons and the Higgs boson. Leptons and quarks can be grouped into 3 families of 4 elementary particles which are all spin $\frac{1}{2}$ fermions. The particles belonging to separate families have different flavour and mass. This classification of fermions can be seen in Table 1.1 and Table 1.2. It is important to mention that every elementary particle listed in these tables has an anti-particle which has opposite quantum numbers and is usually denoted with a bar above the particle symbol. While leptons interact only via electroweak (EWK) inter-

actions and can be found in nature as free particles, quarks on the other hand interact via both strong and EWK interactions and are observed in hadrons, either mesons which are bound states of one quark and one anti-quark or (anti)baryons which are bound states of three (anti)quarks. The only exception is the very massive and short-lived top quark.

Force mediators in the SM are 3 massive gauge bosons (Z, W^+, and W^-) for the weak force, massless photon (γ) for the electromagnetic (EM) force, and 8 massless gluons (g) for the strong force. The last discovered elementary particle is the Higgs boson (H), a massive scalar particle. A short overview of the force mediators in the SM is given in Table 1.3.

Leptons – spin = 1/2				
Name (flavour)	**Label**	**Mass [GeV]**	**Charge** (e)	**Interaction**
electron neutrino	$\boldsymbol{\nu_e}$	$< 1 \cdot 10^{-8}$	0	Weak
electron	$\boldsymbol{e}$	0.000511	-1	Electromagnetic (EM), Weak
muon neutrino	$\boldsymbol{\nu_\mu}$	< 0.0002	0	Weak
muon	$\boldsymbol{\mu}$	0.106	-1	EM, Weak
tau neutrino	$\boldsymbol{\nu_\tau}$	< 0.02	0	Weak
tau	$\boldsymbol{\tau}$	1.777	-1	EM, Weak

Table 1.1 – Classification of leptons in the SM [1].

Quarks – spin = 1/2				
Name (flavour)	**Label**	**Mass [GeV]**	**Charge** (e)	**Interaction**
up	$\boldsymbol{u}$	0.002	$2/3$	All
down	$\boldsymbol{d}$	0.005	$-1/3$	
charm	$\boldsymbol{c}$	1.275	$2/3$	All
strange	$\boldsymbol{s}$	0.095	$-1/3$	
top	$\boldsymbol{t}$	173	$2/3$	All
bottom	$\boldsymbol{b}$	4.18	$-1/3$	

Table 1.2 – Classification of quarks in the Standard Model (SM) [1].

Gauge bosons and Higgs boson					
Name	**Label**	**Charge** (e)	**Spin**	**Mass [GeV]**	**Mediates**
Photon	$\boldsymbol{\gamma}$	0	1	0	EM
W^+ boson	$\boldsymbol{W^+}$	+1	1	80.4	Weak
W^- boson	$\boldsymbol{W^-}$	-1	1	80.4	Weak
Z^0 boson	$\boldsymbol{Z^0}$	0	1	91.2	Weak
gluons	$\boldsymbol{g}$	0	1	0	Strong
Higgs boson	$\boldsymbol{H}$	0	0	125	/

Table 1.3 – Gauge bosons and Higgs boson in the SM [1].

1.1.2 Building the Standard Model with quantum field theory

The Standard Model is a quantum field theory, meaning its fundamental objects are quantum fields which are defined at all points in spacetime. These fields are

- the fermion fields, ψ, which account for matter particles;
- the electroweak boson fields W_1, W_2, W_3, and B;
- the gluon fields, $\mathrm{G_a}$;
- the Higgs field, ϕ.

The dynamics of the quantum state and the fundamental fields are determined by the Lagrangian density $\mathcal{L}$, or usually for short just called the Lagrangian. The Standard Model is furthermore a gauge theory, which means that the Lagrangian is invariant under a local symmetry, in which the transformation parameters can be a function of the space-time position rather than a constant term. The gauge group of the Standard Model is $SU(3)_C \times SU(2)_L \times U(1)_Y$, where the indices stand for color, left-handed, and hypercharge respectively.

Next, the formalism of a local gauge symmetry in the framework of the strong interaction is introduced, which is described by the quantum chromodynamics (QCD), a theory based on the colour group $SU(3)_C$. Then we will apply the same formalism to the electromagnetism and the weak interactions that are unified in the EW theory which relies on the weak isospin and weak hypercharge $SU(2)_L \times U(1)_Y$ group.

Quantum chromodynamics

The (QCD) sector defines the interactions between quarks and gluons, governed by the symmetry group $SU(3)_C$. It is a non-abelian group with 8 generators representable by the 8 Gell-Mann matrices $\frac{\lambda_a}{2}, a = 1, ..., 8$. Quantum number associated to this group is called colour and it can take 3 values: blue, red and green. Since leptons do not interact with gluons, they are not affected by this sector. The Dirac Lagrangian of the quark fields can be written as:

$$\mathcal{L}_{QCD} = \bar{q}_f(i\gamma^\mu \partial_\mu - m)q_f. \tag{1.1}$$

We want this Lagrangian to be invariant under a local symmetry. The quark fields transform linearly under the local gauge symmetry:

$$q_f(x) \to e^{-i\alpha_a(x)\frac{\lambda_a}{2}} q_f(x), \tag{1.2}$$

but the space-time derivatives of the fields do not. Because of this one has to replace ∂_μ with a covariant derivative:

$$D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} G^a_\mu, \tag{1.3}$$

where a coupling constant g_s and 8 gauge vector fields G^a_μ corresponding to the 8 gluons have to be introduced. These have to transform as

$$G^a_\mu \to G^a_\mu + \alpha^b(x) f^{abc} G^c_\mu + \frac{1}{g_s}\partial_\mu \alpha^a(x), \tag{1.4}$$

where the f^{abc} are the structure constants of the group. One can show that there are always as many gauge fields as generators and these represent different force mediators in the SM. Finally, with the above substitutions gauge invariant QCD Lagrangian can be written as:

$$\mathcal{L}_{QCD} = \bar{q}_f(i\gamma^\mu \partial_\mu - m)q_f - g_s \bar{q}_f \gamma^\mu \frac{\lambda_a}{2} q_f G^a_\mu - \frac{1}{4} G^{\mu\nu}_a G^a_{\mu\nu}, \tag{1.5}$$

with summation over all quark fields implied.

First term in this expression is the free Dirac Lagrangian from Eq 1.1, the second term governs the interactions between quark and gluon fields while the last term introduces trilinear and quadrilinear terms that correspond to interactions among the gluon fields.

Electroweak interaction

The electroweak sector defines the interactions between leptons and 4 gauge bosons, governed by the $SU(2)_L \times U(1)_Y$ symmetry group. Quantum numbers associated to this group are weak isospin (I_3) and weak hypercharge (Y), directly linked to the electric charge by the relation $Q = I_3 + \frac{Y}{2}$. One peculiar difference with respect to the QCD sector is that in the EW theory fermions of opposite chirality have different interactions. If we denote $SU(2)_L$ doublets as $L = \begin{pmatrix} \psi_L \\ \psi'_L \end{pmatrix}$, and singlets as ψ_R and ψ'_R, then the free EW Lagrangian can be written as:

$$\mathcal{L}_{EW} = \bar{L} i\gamma^\mu \partial_\mu L + \bar{\psi}^{(\prime)}_R i\gamma^\mu \partial_\mu \psi^{(\prime)}_R. \tag{1.6}$$

In a similar manner as described in the previous section, doublets and singlets transform linearly under the local gauge symmetry:

$$L \to e^{-i\alpha_i(x)\frac{\sigma_i}{2} - i\beta(x)\frac{Y}{2}} L, \quad \psi^{(\prime)}_R \to e^{-i\beta(x)\frac{Y}{2}} \psi^{(\prime)}_R, \tag{1.7}$$

where $\frac{\sigma_i}{2}$ are Pauli matrices that represent 3 generators of $SU(2)_L$. Still following the procedure from the previous section we have to replace partial derivatives with corresponding covariate derivatives:

$$\begin{aligned} L: &\quad D_\mu = \partial_\mu + i g_w \frac{\sigma_i}{2} W^i_\mu + i g \frac{Y}{2} B_\mu, \\ \psi_R, \psi^{(\prime)}_R: &\quad D_\mu = \partial_\mu + i g \frac{Y}{2} B_\mu, \end{aligned} \tag{1.8}$$

where 2 coupling constants g_w and g were introduced as well as 3+1 gauge fields. Gauge fields transfrom as:

$$
\begin{aligned}
W_\mu^i &\to W_\mu^i + \alpha^j(x)\epsilon^{ijk}W_\mu^k + \frac{1}{g_w}\partial_\mu\alpha^i(x), \\
B_\mu &\to B_\mu + \frac{1}{g}\partial_\mu\beta(x),
\end{aligned}
\tag{1.9}
$$

with ϵ^{ijk} being the structure constants of the group. The connection between W_μ^i and B_μ and physical weak boson fields $W_\mu^\pm$ and Z_μ and the photon field A_μ is given with a simple linear combination:

$$
\begin{aligned}
W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\
Z_\mu &= W_\mu^3\cos\theta_w - B_\mu\sin\theta_w, \\
A_\mu &= W_\mu^3\sin\theta_w + B_\mu\cos\theta_w,
\end{aligned}
\tag{1.10}
$$

where θ_w is the weak mixing angle defined as $\sin\theta_w = \frac{g}{\sqrt{g_w^2+g^2}}$.

Substituting partial for covariant derivative we get the expression for the full electroweak Lagrangian:

$$
\begin{aligned}
\mathcal{L}_{EW} \quad = \quad & \bar{L}i\gamma^\mu\partial_\mu L + \bar{\psi}_R^{(\prime)}i\gamma^\mu\partial_\mu\psi_R^{(\prime)} \\
& - g_w\bar{L}\gamma^\mu\frac{\sigma_i}{2}LW_\mu^i - g\bar{L}\gamma^\mu\frac{Y}{2}LB_\mu - g\bar{\psi}_R^{(\prime)}\gamma^\mu\frac{Y}{2}\psi_R^{(\prime)}B_\mu \\
& - \frac{1}{4}W_i^{\mu\nu}W_{\mu\nu}^i - \frac{1}{4}B^{\mu\nu}B_{\mu\nu},
\end{aligned}
\tag{1.11}
$$

with summations over all doublets and singlets implied. First two terms in this expression are the free Dirac Lagrangian from Eq 1.6, the third, fourth and fifth term are the interaction terms that lead to Feynman rules for the interaction between fermions and the γ, $W^\pm$, and Z mediators of the EW interactions. The last two terms describe trilinear and quadrilinear EW interactions among gauge bosons.

Spontaneous symmetry breaking and the Higgs sector

Electroweak and QCD Lagrangians present an elegant but very powerful theory that is able to predict interactions between elementary particles to an unprecedented scale. However, this theory has a couple of fundamental problems. It requires all bosons and fermions to be massless and becomes non-pertubative at the TeV scale. A simple introduction of the mass terms would violate the gauge invariance of the Lagrangian, and this is where the elegant application of the spontaneous symmetry breaking (SSB) solves the problem. It is done with a procedure known as the Brout-Englert-Higgs (BEH) mechanism which was first proposed in 1964 in three independent papers from Englert and Brout [2], Higgs [3], and Guralnik, Hagen and Kibble [4].

The idea of the SSB lies in the fact that the Lagrangian has a global or local symmetry but the quantum theory, instead of having a unique vacuum state which respects this symmetry, has a family of degenerate vacua that transform into each other under the

action of the symmetry group. The BEH mechanism does this by introducing a new field which is symmetric under the gauge transormations of the SM gauge group and that acquires a non-zero expectation value in the vacuum state thus breaking the electroweak symmetry. The simplest field with this properties that can be introduced is a $SU(2)_L$ doublet of complex scalar fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}, \tag{1.12}$$

which can be introduced in the Lagrangian of the SM with the term:

$$\mathcal{L}_{Higgs} = (D^\mu \phi)^\dagger (D_\mu \phi) + V(\phi), \tag{1.13}$$

where D_μ is a covariant derivative from Eq. 1.9 and $V(\phi)$ is the potential of the introduced scalar fields. Again, simplest form for the potential is chosen:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \tag{1.14}$$

with the introduction of constant terms λ and μ^2. In order to have a ground state, the potential has to be bounded from below, i.e. $\lambda > 0$, and for the other constant the $\mu^2 < 0$ is chosen to introduce the spontaneous symmetry breaking. Instead of a trivial $\phi = 0$ ground state one gets a manifold of non-zero values for the ground state given with the expression:

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \tag{1.15}$$

Once we choose a particular ground state the $SU(2)_L \times U(1)_Y$ symmetry gets spontaneously broken to the electromagnetic subgroup $U(1)_{EM}$, which remains a true symmetry of the vacuum.

According to the Goldstone theorem [5] three massless states corresponding to the three broken $SU(2)_L$ generators should appear. We can parametrize the scalar doublet in the general form:

$$\phi = e^{i\frac{\sigma^i}{2}\theta^i(x)} \begin{pmatrix} 0 \\ v + \frac{h(x)}{\sqrt{(2)}} \end{pmatrix}, \tag{1.16}$$

with four scalar fields; three massless $\theta_i(x)$ fields and a massive field $h(x)$.

Like any other field in the SM, the Higgs field also transforms linearly under the local gauge symmetry:

$$\phi(x) \to e^{-i\alpha_i(x)\frac{\sigma_i}{2}} \phi(x), \tag{1.17}$$

with $\alpha_i(X) = 2\theta^i(x)$, which eliminates the θ^i fields turning their three degrees of freedom into the longitudinal degrees of freedom of the now massive $W^\pm$ and Z^0 vector bosons.

Expanding Eq. 1.18 with the expressions for the covariant derivative, scalar fields and

the potential gives us the full Higgs Lagrangian:

$$\begin{aligned}\mathcal{L}_{Higgs} &= \frac{1}{2}\partial_\mu h \partial^\mu h + \mu^2 h^2 \\ &+ \frac{g_w^2 v^2}{4} W_\mu^- W^{+\mu} + \frac{g_w^2 v^2}{8\cos^2\theta_w} Z_\mu Z^\mu \\ &+ \frac{g_w^2 v^2}{2} h W_\mu^- W^{+\mu} + \frac{g_w^2}{4} h^2 W_\mu^- W^{+\mu} + \frac{g_w^2 v}{4\cos^2\theta_w} h Z_\mu Z^\mu + \frac{g_w^2}{8\cos^2\theta_w} h^2 Z_\mu Z^\mu \\ &+ \frac{\mu^2}{v} h^3 + \frac{\mu^2}{4v^2} h^4.\end{aligned} \tag{1.18}$$

The addition of this part to the SM Lagrangian generates mass to the Higgs boson ($m_H = \sqrt{2}|\mu|$) and to the $W^\pm$ ($m_W = \frac{1}{2} g_w v$) and Z^0 ($m_Z = \frac{g_w v}{2\cos\theta_w}$) bosons while the photon remains massless. It also introduces trilinear and quadrilinear couplings of the Higgs boson to weak vector bosons and its self-couplings.

The BEH mechanism is also used to extend the SM Lagrangian with gauge-invariant Yukawa terms that are responsible for generating the fermion masses. After the EW symmetry breaking the additional terms in the Lagrangian are given with:

$$\mathcal{L}_{Yukawa} = \sum_f -m_f \bar{\psi}\psi \left(1 + \frac{h}{v}\right) + \sum_{f'} -m_{f'} \bar{\psi}'\psi' \left(1 + \frac{h}{v}\right), \tag{1.19}$$

where first sum runs over up-type fermions and the second sum over down-type fermions. Together with necessary mass terms for fermions, additional terms imply that the Higgs boson field interacts with fermions with couplings proportional to their masses. With the introduction of the BEH mechanism the overall symmetry of the theory gets reduced since the three generations of matter no longer appear as identical.

The full SM Lagrangian is given with:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}, \tag{1.20}$$

where separate terms are given with Eq 1.5, 1.11, 1.18 and 1.19.

In addition to being able to explain the origin of the particle mass while preserving gauge symmetry of fundamental underlying theory, the BEH mechanism also provides exact unitarisation of the SM. It solves two big shortcomings of the SM and provides a theory that is valid at all scales, at least in principle.

Using this Lagrangian we are able to calculate and predict almost any process, but as with any other theory this has no significant value without an experimental confirmation of our predictions. We will now briefly go trough the phenomenology of the SM Higgs boson at the LHC in order to better understand what are the interesting predictions of the SM theory that are being tested in this impressive worldwide high energy physics effort.

1.2 Higgs boson at the LHC

The discovery of the Higgs boson marked the final triumph for the SM. In order to better understand how the experimental setup was planned and how the search was conducted we will study the main features of the production mechanism and decay modes of the Higgs boson. We will then give a short history of the searches leading to the big discovery announced on 4^{th} July in 2012 and conclude with the latest results and the current status of the searches and properties measurements.

The evolution of SM process cross sections as a function of centre-of-mass energy for hadron colliders is illustrated in Fig. 1.1. It shows the advantage for raising the centre-of-mass energy for the Higgs boson search, motivating the development of new hadron colliders in search for the Higgs boson.

1.2.1 Production mechanisms

According to the SM, Higgs can be produced in many different ways. We will focus on the main production mechanisms contributing to the Higgs boson production at a hadron colliders [6, 7]. The main four production mechanisms are illustrated in the form of Feynman diagrams in Fig. 1.2. The cross sections for main production mechanisms as a function of the centre-of-mass energy are shown in Fig. 1.3.

Gluon fusion

The gluon fusion (ggH) process, where two gluons fuse via an intermediate loop of virtual quarks as illustrated in Fig. 1.2 (a), has the largest cross section dominating other production modes by more than one order of magnitude as one can see in Fig. 1.3. This is because the gluon luminosity is very large in pp collisions at the high centre-of-mass energies provided by the LHC. As the most massive quark, the top quark gives the largest contribution to the loop amounting to ∼90% and the only other seizable contribution arises from the bottom quark accounting for ∼5-10% of the total cross section.

Higher order QCD corrections are very important for this process, for example next-to-leading order (NLO) QCD corrections have been found to increase the cross section by a factor of ∼2. Because of this it is crucial to use state-of-the-art computations. In this thesis work, recent N3LO QCD + NLO EW computations [8, 9] have been used.

Vector boson fusion

Vector boson fusion (VBF) illustrated in Fig. 1.2 (b) is the second production mechanism at the LHC with a cross section roughly an order of magnitude smaller than that of ggH. It occurs when two fermions exchange virtual W or Z bosons, which immediately fuse into the Higgs boson. This production mechanism is very important and interesting because it has a clear experimental signature with two forward and backward energetic jets with high invariant mass. This characteristic topology helps rejecting SM

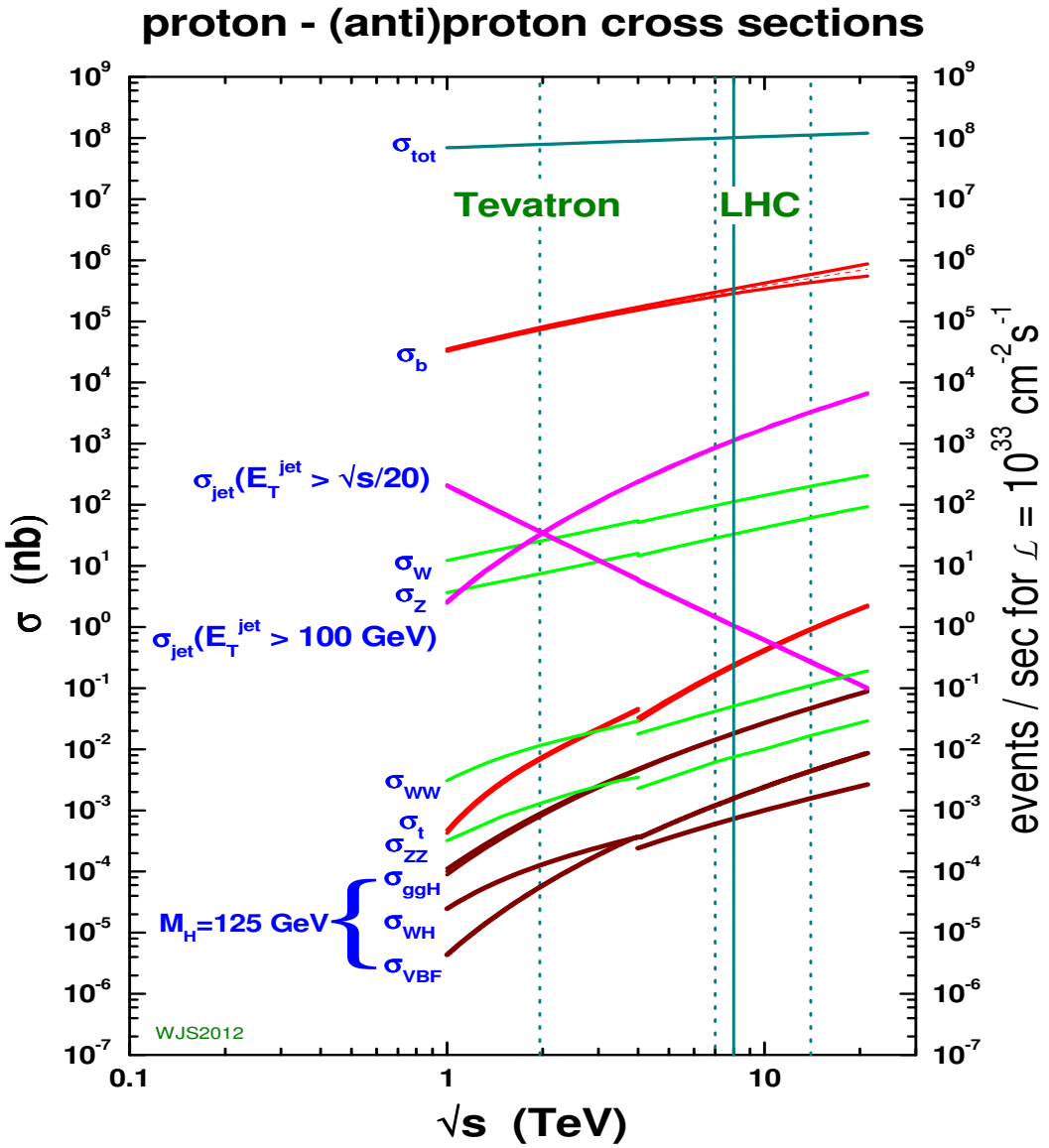

Figure 1.1 – Standard Model process cross sections at hadron colliders as a function of centre-of-mass energy.

backgrounds and ggH production in association with two jets. Both leading order (LO) and NLO cross sections of this process are known [9] with small uncertainties and the higher order QCD corrections are found to be very small.

Associated production with a vector boson

Associated production with a vector boson is the third most prominent production mechanism at the LHC being about twice less frequent than VBF. This process is often called VH associated production or Higgsstrahlung and it is illustrated in Fig. 1.2 (c). In this

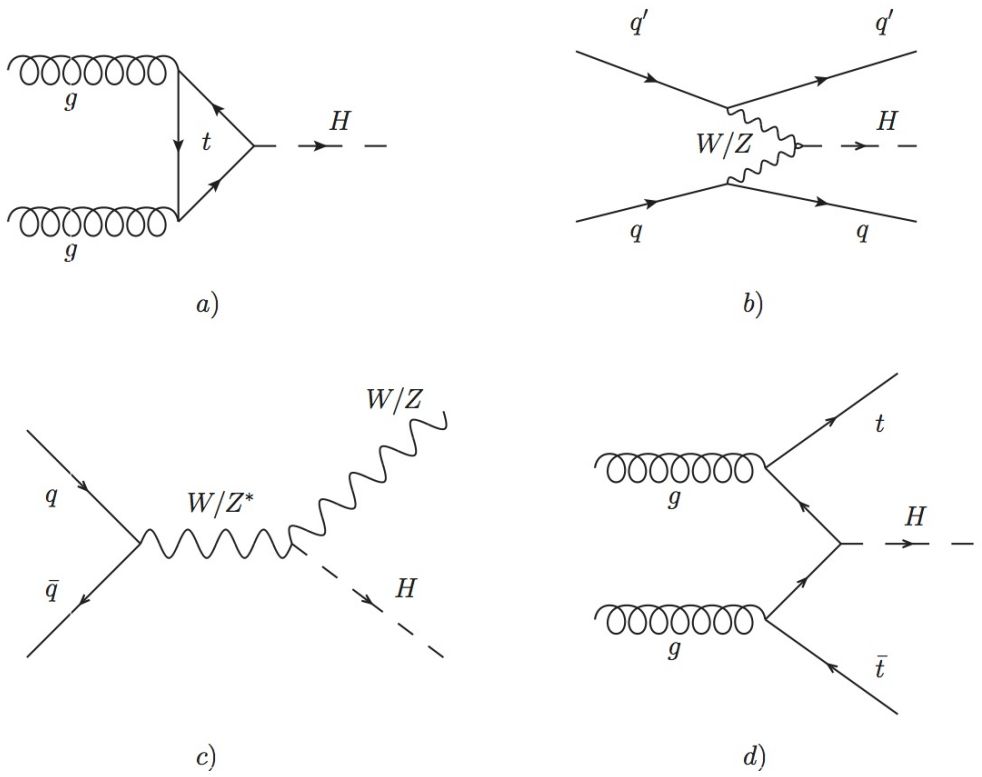

Figure 1.2 – Some of the leading order diagrams for the relevant SM Higgs boson production mechanisms: (a) gluon fusion, (b) vector boson fusion (VBF), (c) W and Z associated production or Higgsstrahlung, and (d) $\mathrm{t\bar{t}H}$ associated production.

process a fermion and anti-fermion collide and produce a W or Z boson which later on radiates a Higgs boson. This leads to experimental signature where Higgs boson is bosted and accompanied by the leptonic or hadronic products of the associated W or Z boson. When the vector boson decays hadronically, a pair of nearby boosted jets with invariant mass close to the nominal mass of the vector boson can be found. On the other hand, leptonic decays either provide one lepton and missing transverse energy for WH, or either a pair of leptons or missing transverse energy for ZH. The higher-order QCD corrections are quite large for this production mechanism and are computed at NNLO QCD and NLO EW accuracy [9].

Associated production with a top quark pair

Associated production with a top quark pair ($\mathrm{t\bar{t}H}$) is almost two orders of magnitude smaller than ggH and several times lower than VBF. It mostly involves two colliding gluons each decaying into a heavy quark-antiquark pair. A quark and antiquark from each pair can then fuse to form a Higgs boson as illustrated in Fig. 1.2 (d). The presence of the $\mathrm{t\bar{t}}$ pair in the final state provides an interesting experimental signature that can be used to study this rare production mechanism. The higher-order QCD corrections are of order of ∼1.2 and are computed at NLO QCD and NLO EW accuracy [9].

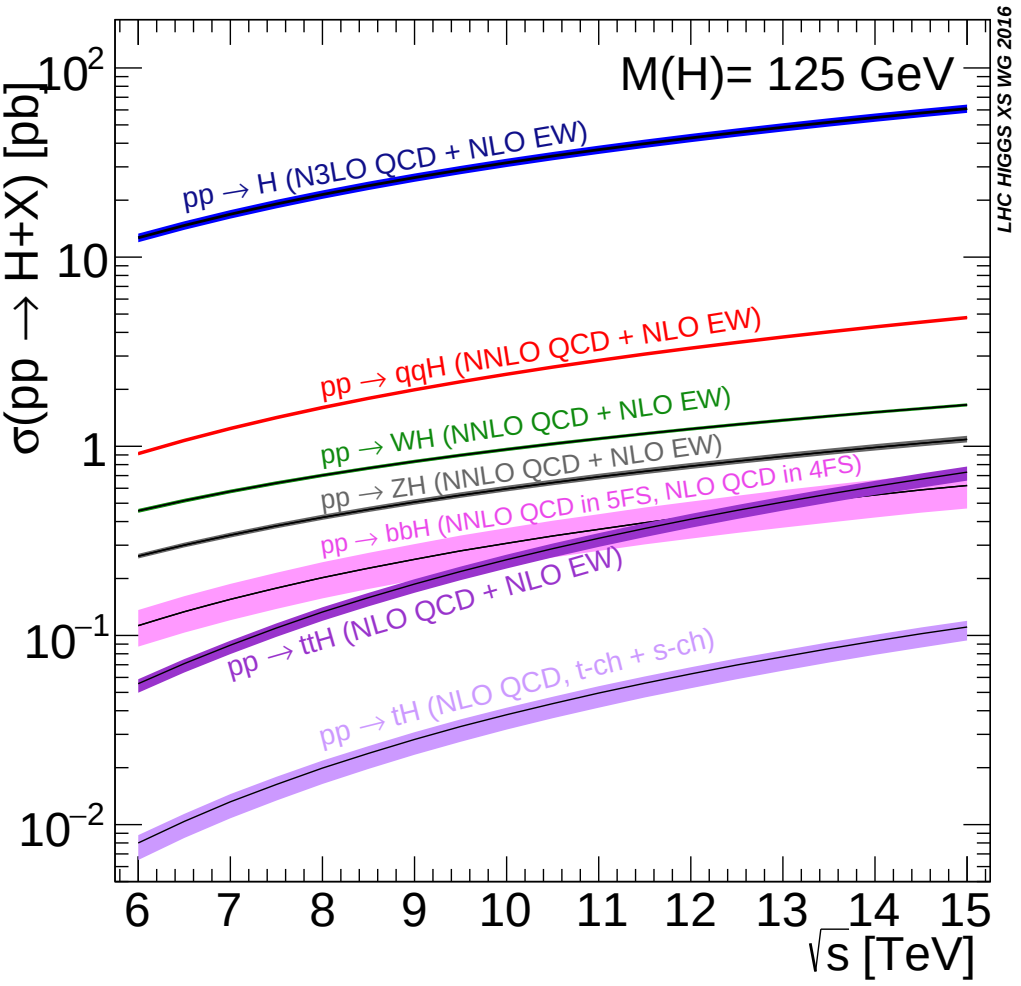

Figure 1.3 – Total production cross section for a 125 GeV SM Higgs boson at the LHC as a function of the centre-of-mass energy.

Other production mechanism

Associated production with a bottom quark pair ($\mathrm{b\bar{b}H}$) and single top production (tqH) are the two other production mechanisms that were considered in the work of this thesis. While the $\mathrm{b\bar{b}H}$ has cross section that is comparable to the one of the $\mathrm{t\bar{t}H}$ process, tqH has a cross section one order of magnitude smaller making it extremely rare. Other Higgs boson production mechanisms have even lower cross sections and therefore are not being considered.

1.2.2 Decay modes

From quantum mechanics we know that if it is possible for a particle to decay into a set of a lighter particles, then it will eventually do so. The Higgs boson is no exception to this. The probability for a decay depends on many factors like the difference in mass,

the strength of interactions, etc. Most of these factors are fixed by the SM, except for the mass of the Higgs boson itself. For a Higgs boson with a mass of 125 GeV the SM predicts a mean lifetime of about $1.6 \cdot 10^{-22}$ s [10]. This mean lifetime is too short for the Higgs boson to reach the detector so we have to search for it through its decay products which are detectable in our experiment.

The SM prediction for the branching ratios of the Higgs boson decay modes depends on the value of its mass and is shown in Fig. 1.4. The Higgs boson has many different processes through which it can decay since it interacts with all the massive elementary particles of the SM and can decay to massless particles as well through an intermediate loop of virtual particles.

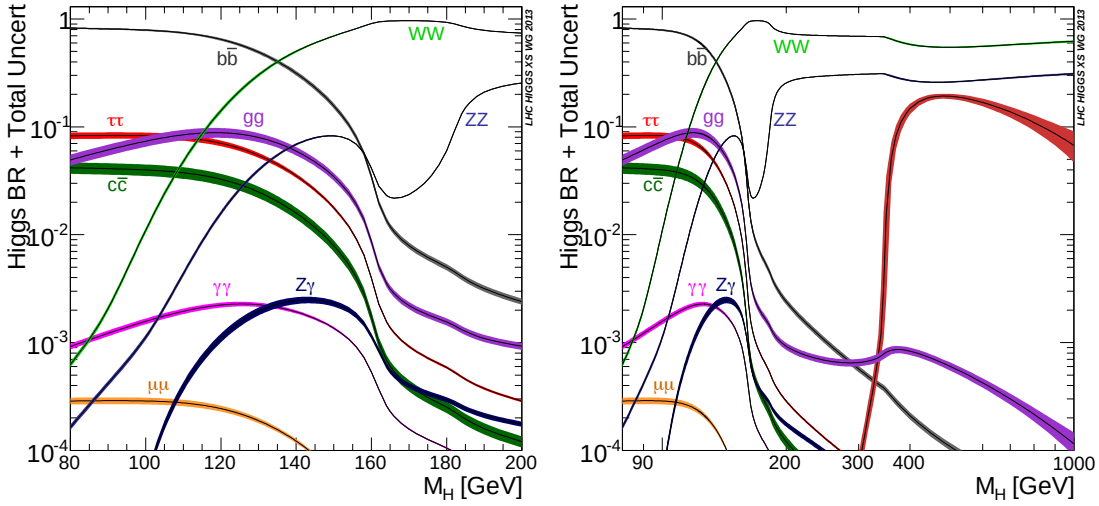

Figure 1.4 – Decay branching ratios with uncertainties from the SM Higgs boson as a function of m_{H}, in a low-mass range (left) and in a range extending to high mass (right).

One possible Higgs boson decay is into a fermion-antifermion pair. The Higgs boson is more likely to decay into heavy than light fermions because strength of its interaction is proportional to the fermion mass. Following this logic the most common decay should be into $\mathrm{t\bar{t}}$ but this decay is possible only if the Higgs boson is heavier than $2m_t$ (∼350 GeV). For a Higgs boson mass of ∼125 GeV the most probable decay is to $\mathrm{b\bar{b}}$ with a branching ratio of ∼58%. Search in this channel is very challenging due to the overwhelming QCD background. Therefore, this channel is mainly exploited in boosted regimes, mainly in the VH production mode with the associated boson decaying leptonically, leaving a signal that is somewhat easier to distinguish from the QCD background. The second most common fermion decay is to a $\tau^+\tau^-$ which happens ∼6% of the time. This channel is probed with a variety of experimental strategies depending on the decay products of the tau leptons. All other possible decays to fermion-antifermion pairs like $\mu^+\mu^-$, $\mathrm{c\bar{c}}$, e^+e^- are still not being exploited, either because their experimental signatures cannot be distinguished from the overwhelming QCD background or because with their very low predicted branching ratio the rate of production is too small for current luminosity of the LHC.

Another possibility is the Higgs boson decaying into a pair of gauge bosons. For a 125 GeV SM Higgs the most probable such decays are to $WW^{(\star)}$, gg, $ZZ^{(\star)}$ and $\gamma\gamma$ respectively. The $WW^{(\star)}$ channel has the largest branching ratio of all with a ~21% probability. It can be studied with different experimental signatures arising from W boson decays into quark-antiquark pairs or into a charged lepton and a neutrino. In the case of quark-antiquark pairs it is challenging to distinguish the signal from the QCD background. Contrarily, thanks to accurate lepton identification and missing transverse energy measurement the $WW \to \ell\nu\ell\nu$ is a sensitive channel for Higgs boson searches around 125 GeV. The only shortcoming is that due to a presence of two neutrinos in the final state it is impossible to reconstruct the invariant mass of the Higgs boson. Next in line is decay to a pair of gluons that happens ~8% of the time. This decay mode is not being studied currently at the LHC due to problems of distinguishing it from the significant QCD dijet background.

Two very interesting bosonic decays of Higgs are the ones to $ZZ^{(\star)}$ and $\gamma\gamma$. Although they suffer from a low branching fractions of only ~2.6% and ~0.2%, they benefit from a complete reconstruction of the final state, with a good experimental invariant mass resolution. With the branching fraction becoming even lower for Z boson decays to pairs of leptons it is essential that in this final state ($H \to ZZ^{\star} \to 4\ell$) a good signal over background ratio is achieved. Because of these privileges this channel is often called "golden channel" for Higgs boson search and properties measurement. The golden channel is the focus of this thesis and its characteristics will be studied in detail from Section 1.3. Even though it does not have a good signal over background ratio like the golden channel, decay to two photons is still an important channel for the search and study of the Higgs boson. Thanks to a good experimental resolution for the diphoton invariant mass, the small Higgs boson signal appears as a peak on top of the backgrounds from QCD production of two photons or jet fragments misidentified as photons.

This concludes a brief overview of most significant Higgs production mechanisms and decay modes together with the signatures they leave in the detector and challenges in distinguishing them from the SM background. We will now shift our focus on the history of the Higgs boson experimental searches and current status of main production mechanisms and decay modes with a separate section on the golden channel being the main topic of this thesis.

1.2.3 Status of the experimental searches for the Higgs boson

The first studies exploring how the Higgs boson may show itself in particle collision experiments were published in the mid 1970s [11] stressing the problem that there was almost no clue about the possible mass of the Higgs boson. Theory left open a very wide range somewhere between 10 GeV [12] and 1000 GeV [13] with no real indication where to look [14]. However, later on the allowed range for Higgs boson mass m_H has been better constrained by theoretical arguments [15, 16]. An upper limit called triviality was obtained by requiring that the running quatric coupling λ remains finite up to the scale Λ. A lower limit comes from vacuum stability, requiring that λ remains positive after including radiative corrections which implies that the minimum of the

potential is absolute.

The first extensive search for the Higgs boson was conducted at the Large Electron–Positron Collider (LEP) at CERN in the 1990s [14]. At the end of its service in 2000, LEP had found no conclusive evidence for the Higgs boson. Direct searches at LEP yielded a lower limit of $m_{\mathrm{H}} > 114.4$ GeV at 95% confidence level (CL) [17].

The search continued at Fermilab in the United States, where the Tevatron had been upgraded for this purpose. There was no guarantee that the Tevatron would be able to find the Higgs boson, but it was the only supercollider that was operational since the LHC was still under construction and the planned Superconducting Super Collider had been cancelled in 1993 and never completed. The Tevatron was only able to exclude further ranges for the Higgs boson mass, and was shut down on 30 September 2011 because it no longer could keep up with the LHC. By the time the LHC got started with high luminosity and high energy collisions, the Tevatron excluded the Higgs boson in the mass range $158 < m_{\mathrm{H}} < 173$ GeV at 95% CL [18].

At that time the LHC experiments had taken over and by the end of 2011 the first direct Higgs boson searches were made public analysing data from pp collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV which amounted to an integrated luminosity of 5.1 fb^{-1}. Using five decay channels CMS Collaboration excluded the Standard Model Higgs boson in the mass range $129 < m_{\mathrm{H}} < 525$ GeV, and the ATLAS Collaboration excluded the $129.2 - 541$ GeV at a 95%CL [19]. At that point it was evident that if a Higgs boson existed it was to be searched for between 120 and 130 GeV.

In 2012 the centre-of-mass energy of LHC was raised to $\sqrt{s} = 8$ TeV and additional data sample of 5.3 fb^{-1} was collected. Finally, on 4^{th} July 2012, the ATLAS and CMS Collaborations announced they had independently made the same discovery, the observation of a new boson with mass near 125 GeV compatible with the SM Higgs boson [20, 21, 22]. In that same time results from the direct searches at the Tevatron were published excluding the Higgs boson in the mass ranges $100 - 103$ GeV and $147 - 180$ GeV at a 95% CL [23]. The results also indicated an excess of events with a global significance of 3.1σ in a mass range consistent [24, 25] with the LHC observations.

The full Run I LHC data sample both at $\sqrt{s} = 7$ TeV and 8 TeV was used to confirm the announced discovery as well as to perform studies of Higgs production mechanisms and decay modes. Mass measurement of a newly discovered boson was obtained using the two high-resolution channels [26, 27, 28, 29] and combining the data from both experiments [30] leading to a value of $m_{\mathrm{H}} = 125.09 \pm 0.21(\mathrm{stat.}) \pm 0.11(\mathrm{syst.})$ GeV. The new boson was shown to have spin-parity $J^P = 0^+$ [31], and its production and decay rates together with its coupling strengths to SM particles turned out to be consistent with the expectations for the SM Higgs boson [32, 29]. Combination of the analyses of ATLAS and CMS Collaboration was again performed and even stronger conclusion was made that all measured properties across different channels are consistent with the SM predictions for Higgs boson [33].

After Run I was concluded both experiments prepared for Run II where focus shifted to

possible discovery of all main production mechanisms and decay modes of the newly discovered Higgs boson. Also, experiments were aiming to study its properties with higher precision in the search for any anomalies with respect to the predictions from the SM indicating possible Beyond the Standard Model (BSM) physics. In the time this thesis was written, Run II was coming to its end with only couple more months of data taking left. Even without the full Run II data sample, both ATLAS and CMS Collaborations have so far announced observations of all four main production mechanisms ggH, VBF [34, 35], VH [36, 37], and $\mathrm{t\bar{t}H}$ [38, 39] and five main decay channels $\mathrm{ZZ}^{(\star)}$, $\mathrm{WW}^{(\star)}$, $\gamma\gamma$, $\mathrm{b\bar{b}}$ [36, 37] and $\tau^+\tau^-$ [40, 41]. Again, both experiments found coupling strengths to SM particles to be consistent with the SM predictions [42, 43] as can be seen in Fig. 1.5 for the CMS measurement.

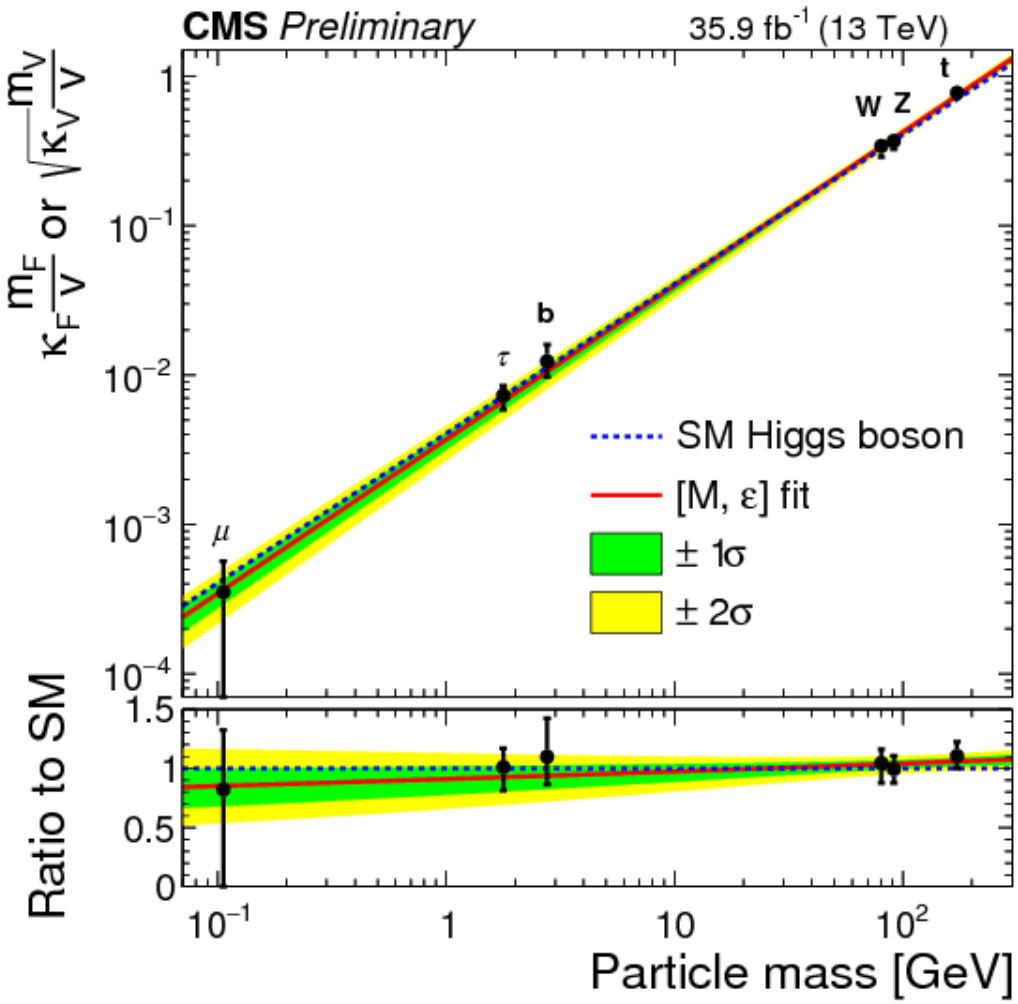

Figure 1.5 – Latest measurements of coupling strengths of the Higgs boson with the gauge bosons and fermions, compared to the SM expectation from the CMS Collaboration.

As already discussed in Section 1.2.2, the $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ is the key channel in Higgs boson properties measurement. We continue by introducing this channel with some of its key features and also mentioning contributions from its searches leading up to the

work performed in the scope of this thesis.

1.3 The golden channel

Work presented in this thesis was performed studying the decay of the Higgs boson to two Z bosons that in turn decay to pairs of electrons or muons, denoted as $H \to ZZ^{\star} \to 4\ell$ or in short $H \to 4\ell$. Of course, finding four leptons that could decay from a Higgs boson in your data does not mean you have found your Higgs boson signal. You have to understand all SM processes that leave identical signature without any presence of the Higgs boson. There are several processes of this kind predicted by the SM. Main are those when a pair of Z bosons or $Z\gamma^{*}$ are produced via gluon fusion or quark-antiquark annihilation. There is also a significant contribution of the so called "Z+X" reducible background, where X stands for Z boson reconstructed by two leptons that are not products of its decay. These leptons can come either from jets or can be some other particles that are misidentified as leptons duo to imperfections in reconstruction.

This channel has three possible final states referred as $4e$, 4μ and $2e2\mu$. Because these final states have different mass resolution and reducible background rate they are analysed separately. It is also possible to search for different Higgs boson production mechanisms mentioned in Section 1.2.1 with this channel. Thanks to a very clean detector signature of only 4 leptons in the final state, different production mechanisms are easily accessible via extra objects in the event like jets, missing transverse energy and extra leptons. Due to its very low branching ratio of only 0.0124% some of the main production mechanisms are still suffering from low statistics, but in the future this channel will for sure filter out as one of the key tools for measuring properties and possible BSM anomalies of main Higgs boson production mechanisms.

It is not a small thing to be proclaimed a golden channel for important discovery such as the one for the final missing piece of the SM puzzle which was the hunt for the Higgs boson. There must be a very good reason for that and here are three which stand out:

- Complete reconstruction of final state objects. Thanks to the ability of LHC experiments to reconstruct leptons even at very low energies it is possible to reverse the decay of the Higgs boson to four leptons. From the kinematic information of the leptons, i.e. their four momentum, it is possible to calculate powerful kinematic discriminants making it much easier to distinguish signal from background,
- Very good momentum resolution. Due to excellent momentum resolution of electrons and muons in the LHC experiments it is possible to measure mass of the Higgs boson with a very high precision,
- Great signal to background ratio. Even though branching fraction of this channel is very low, excellent signal to background ratio of roughly 2 to 1 in a narrow mass range around the Higgs boson peak makes is another key feature of this channel.

The golden channel has defended its name in Run I with a set of impressive results from

the ATLAS and CMS Collaborations:

- Mass measurement [26, 27, 44],
- Width measurement from a direct method [26, 27], from a much more powerfull combined analysis of the on-shell and off-shell regions [45, 46], and finally from the flight distance within the CMS detector [47],
- Signal strength and couplings to fermions and bosons [26, 44],
- Spin-parity [31, 48, 26, 49, 50],
- Fiducial cross section, both integrated and as a function of several parameters [51, 52]
- Anomalous couplings [49, 50, 47],
- Search for additional heavy Higgs bosons [53, 54].

While the full Run II dataset will be analysed in the coming years to improve on all of the measurements mentioned above and to introduce some new properties measurements, work of this thesis is based on the Run II datasets collected between 2015 and 2017. Already, most of these results have been updated with a new centre-of-mass energy of $\sqrt{s} = 13$ TeV with increased statistics and with improved analysis strategy. Main focus of this analysis are mass and width measurement, signal strengths, and fiducial cross section with several significant improvements that will be described in detail.

But before we embark on our journey of hunting for Higgs bosons and studying them carefully, we have to appreciate the huge and long effort that was needed to build such an experiment. For this reason, next chapter is dedicated to give a brief introduction to Conseil Europeen pour la Recherche Nucleaire (CERN) and LHC, and to explain how CMS detector was build and how it detects and reconstructs different types of elementary particles.

Chapter 2

The CMS experiment at the Large Hadron Collider

This thesis was carried out within the CMS experiment, studying data from proton-proton (pp) collisions at a center-of-mass energy of 13 TeV delivered by the Large Hadron Collider. Like for many other thesis in experimental particle physics these days, work was carried out in a big collaboration. It is therefore important to give a clear picture to the reader how such a big project that involves thousands of scientist and technicians from all over the world can be constructed and conducted so efficiently and successfully.

After a short introduction about the history of CERN, this chapter will also provide an introduction to the operation of the LHC and a more detailed description of the CMS experiment.

2.1 A short history of CERN

The convention establishing CERN was ratified on 29 September 1954 by 12 countries in Europe. The acronym CERN originally represented the French words for Conseil Européen pour la Recherche Nucléaire (European Council for Nuclear Research), which was a provisional council for building the laboratory, established by 12 European governments in 1952. The acronym was retained for the new laboratory after the provisional council was dissolved, even though the name changed to the current Organisation Européenne pour la Recherche Nucléaire (European Organization for Nuclear Research) in 1954. The laboratory was originally devoted to the study of atomic nuclei, but was soon applied to higher-energy physics, concerned mainly with the study of interactions between subatomic particles. Therefore, the laboratory operated by CERN is commonly referred to as the European laboratory for particle physics (Laboratoire européen pour la physique des particules), which better describes the research being performed there.

These days CERN is considered to be a world-wide research organization that operates the largest particle physics laboratory in the world. It is based in a northwest suburb of Geneva on the Franco–Swiss border, and has 22 member states. The acronym CERN is also used to refer to the laboratory, which in 2016 had 2500 scientific, technical, and administrative staff members, and hosted about 12000 users. In the same year, CERN generated 49 petabytes of data.

CERN's main function is to provide the particle accelerators and other infrastructure needed for high-energy physics research – as a result, numerous experiments have been constructed at CERN through international collaborations. The main site at Meyrin hosts a large computing facility, which is primarily used to store and analyse data from experiments, as well as simulate events. Researchers need remote access to these facilities, so the lab has historically been a major wide area network hub.

Several important achievements in particle physics have been made through experiments at CERN. They include:

- 1973: The discovery of neutral currents in the Gargamelle bubble chamber,
- 1983: The discovery of W and Z bosons in the UA1 and UA2 experiments,
- 1999: The discovery of direct CP violation in the NA48 experiment,
- 2012: A boson with mass around 125 GeV consistent with the long-sought Higgs boson.

These major scientific discoveries were recognized in the scientific community resulting in several Nobel Prizes. The 1984 Nobel Prize for Physics was awarded to Carlo Rubbia and Simon van der Meer for the developments that resulted in the discoveries of the W and Z bosons. The 1992 Nobel Prize for Physics was awarded to CERN staff researcher Georges Charpak "for his invention and development of particle detectors, in particular the multiwire proportional chamber". The 2013 Nobel Prize for physics was awarded to Englert and Higgs for the theoretical description of the Higgs mechanism in the year after the Higgs boson was found by CERN experiments.

2.2 The Large Hadron Collider

The LEP e^+e^- collider was the largest particle accelerating machine ever constructed at the time. It was built at CERN with circumference of 27 km, in a tunnel roughly 100 meters underground and passing through Switzerland and France. It was used from 1989 until 2000. In 2001 it was dismantled to make way for the LHC, which re-used the LEP tunnel. The LHC took over its place and inherited the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) accelerator systems illustrated in Fig. 2.1.

The LHC is nowadays the world's largest and most powerful particle collider, the most complex experimental facility ever built and the largest single machine in the world. It

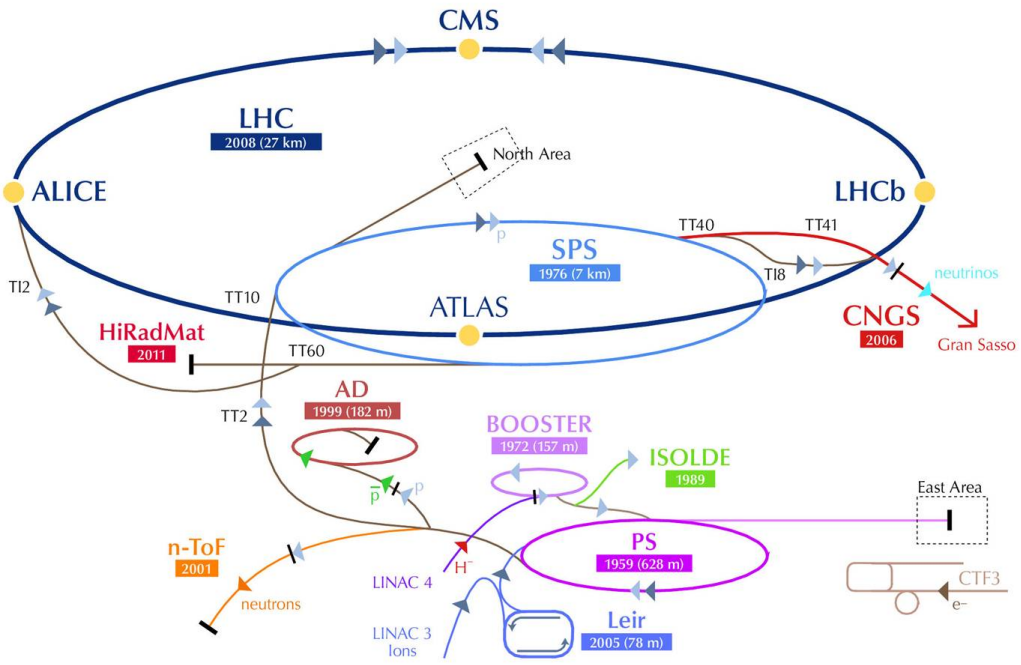

Figure 2.1 – The CERN accelerator complex. Proton acceleration starts from a Linear Accelerator (LINAC) that injects the protons to PS. In the following stage, the SPS accelerates proton beams to a higher energy and subsequently injects the beam into the LHC ring.

was built in collaboration with over 10000 scientists and hundreds of universities and laboratories, as well as more than 100 countries.

2.2.1 Design and performance

The LHC is a two-ring superconducting proton accelerator and collider. It is designed to explore TeV energy scale, probing the SM processes and searching for the Higgs boson, or an alternative new physics at the TeV scale. It is designed to accelerate proton beams up to 7 TeV, resulting in a center-of-mass energy of 14 TeV and instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$.

The collider tunnel contains two adjacent parallel beam pipes that intersect at four points. Each pipe contains a proton or ion beam and the beams travel in the opposite directions around the ring. The beams are kept on their circular path with 1232 dipole magnets, while the additional 392 quadrupole magnets are used to keep the beams focused in order to maximize the chances of interaction at the four crossing points. Approximately 96 tonnes of superfluid helium 4 is needed to keep the magnets at their operating temperature of 1.9 K (-271.25°C). This makes the LHC the largest cryogenic facility in the world at liquid helium temperature.

Before being injected into the LHC, proton beams are prepared by a chain of pre-accelerators that increase the energy in steps. Everything begins by injecting hydrogen

gas into the metal cylinder-Duoplasmatron which is surrounded with an electrical field to break down the gas into its constituent protons and electrons. This process yields protons that are first accelerated to an energy of 50 MeV in the LINAC 2, which feeds the Proton Synchroton Booster (PSB) where they are accelerated to 1.4 GeV. They then reach 26 GeV in the PS, and the SPS further increases their energy to 450 GeV before they are injected into the LHC ring at Point 2 and Point 8. All the parts of this acceleration chain can be seen in the illustration on Fig. 2.1. Rather than continuous beams, the protons are bunched together with about 100 billion protons in each bunch and with more than 2000 bunches in a single beam. The shaping of proton bunches together with the acceleration process in the LHC is done by 16 radiofrequency cavities, where the electromagnetic field oscillates at 400 MHz.

The rate of collisions per unit of time can be factorized into two parts: the cross section σ and the instantaneous luminosity $\mathcal{L}$. In order to maximize the rate of events, one wants large instantaneous luminosities, which can be expressed in terms of the beam parameters as:

$$\mathcal{L} = \gamma \frac{f n_b N^2}{4\pi \epsilon_n \beta^*} R, \tag{2.1}$$

where $\gamma = E/m$ is the relativistic Lorentz factor of the protons, n_b is the number of bunches per beam, N is the number of protons per bunch, f is the revolution frequency, R is a reduction factor due to the beam crossing angle at the interaction point, β^* is the beam beta function at the collision point, and ϵ_n is the normalizied transverse beam emittance [55]. The nominal value of the beta function is $\beta^* = 0.55$ m, and the nominal luminosity is reached with $n_b = 2808$ and $N = 1.15 \cdot 10^{11}$. This choice corresponds to a spacing of 25 ns between bunches, i.e. a bunch crossing rate of 40 MHz. The designed instantaneous luminosity of the LHC is $\mathcal{L} \approx 1 \cdot 10^{34}$ cm^{-2}s^{-1} but in 2017 record breaking instantaneous luminosity of $\mathcal{L} \approx 2.06 \cdot 10^{34}$ cm^{-2}s^{-1} was reached surpassing the nominal design specification by more than a factor of two. For example, colliding protons and anti-protons the Tevatron delivered peak instantaneous luminosities of $\mathcal{L} \approx 1 \cdot 10^{32}$ cm^{-2}s^{-1}, two orders of magnitude smaller than the LHC. This is one of the reasons why for the LHC it was decided not to use anti-protons.

2.2.2 Detectors

Four main detectors have been installed around the LHC ring, located in large caverns excavated at positions of beam intersection points. The two largest are ATLAS and CMS detectors. These are multi-purpose detectors located symmetrically at opposite points of the ring where the collisions are expected, by design, to provide the highest instantaneous luminosities. CMS and ATLAS detectors are designed to cover the largest spectrum of the electroweak and strong physics. The detectors were conceived especially to provide the best possible sensitivity to a Higgs boson or additional new physics such as supersymmetry, dark matter, extra-dimensions, etc. The two other detectors are more specialized. A Large Ion Collider Experiment (ALICE) detector is build to study the quark-gluon plasma that existed shortly after the Big Bang and a wide range of phenomena in heavy ion collisions. The Large Hadron Collider beauty (LHCb) is build to investigate what happened to the "missing" antimatter, since the theory predicts that

equal amounts of matter and antimatter were created at the Big Bang. It is also used to study the physics of B-mesons. Besides the main detectors, three other small experiments are installed as parasites. These are TOTal Elastic and diffractive cross section Measurement (TOTEM), Monopole and Exotics Detector at the LHC (MoEDAL), and Large Hadron Collider forward (LHCf). The TOTEM detector aims at measurement of total cross section, elastic scattering, and diffractive processes. The MoEDAL is performing direct search for the magnetic monopole and other highly ionizable stable massive particles. The LHCf is a special-purpose experiment for cosmic ray astroparticle physics.

2.2.3 Operation

The first beam injections started in September 2008, but due to a faulty resistance of an interconnection between two magnets the LHC stopped for more than one year for repairs and commissioning of further safety measurements. On 20 November 2009, 450 GeV beams circulated in the tunnel for the first time since the incident, and shortly after, on 30 November, the LHC achieved 1.18 TeV per beam to become the world's highest energy particle accelerator. During the 2010 ramp-up of the beam energy continued towards the 3.5 TeV per beam which was achieved on 30 March 2010. It was decided to not immediately aim at the design LHC beam parameters yet, and to only operate the collider with a 50 ns bunch spacing and intermediate centre-of-mass energies until 2012, hoping to discover the long sought Higgs boson in the first data taking era referred to as Run I.

With the first data sample of $47\,\mathrm{pb}^{-1}$ in 2010 at $\sqrt{s} = 7$ TeV, the LHC delivered a high luminosity data set of about 6 fb^{-1} during the 2011 runs, a large fraction of which was collected and studied by ATLAS and CMS allowing the exclusion of the big portion of the allowed range for the Higgs boson mass as discussed in Section 1.2.3. Meanwhile, instantaneous luminosities have been steadily growing as one can see in Fig. 2.2 which shows cumulative and peak luminosities delivered to the CMS experiment in every year of data taking. In early 2012, it was decided to increase the centre-of-mass energy to 4 TeV per beam. By the end of that year, the LHC had delivered a data sample corresponding to an integrated luminosity of 23 fb^{-1} and the Higgs boson discovery was announced based on the analysis of a part of that data set. The further exploitation of the large collected data samples went on during the LS1, providing a quantity of precision measurements in the electroweak and strong sectors.

LS1 brought considerable upgrade and consolidation to the LHC, with the goal of meeting the start-up of Run II in 2015 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, and with the objective to progressively increase the instantaneous luminosities throughout this first year of Run II operations, until surpassing the pp collision rates that had been planed in the original design of the LHC. Although 14 TeV have not been reached, the energy increase to 13 TeV still substantially extends the physics reach of the experiments with respect to Run I. After several months of training of the dipole magnets, the first Run II beam was injected on 5^{th} April 2015, and the first 6.5 TeV beam was obtained on 10^{th} April. The year 2015 was intended for commissioning and started with 50 ns collisions, before moving to the nominal bunch spacing of 25 ns over summer. In total, a data

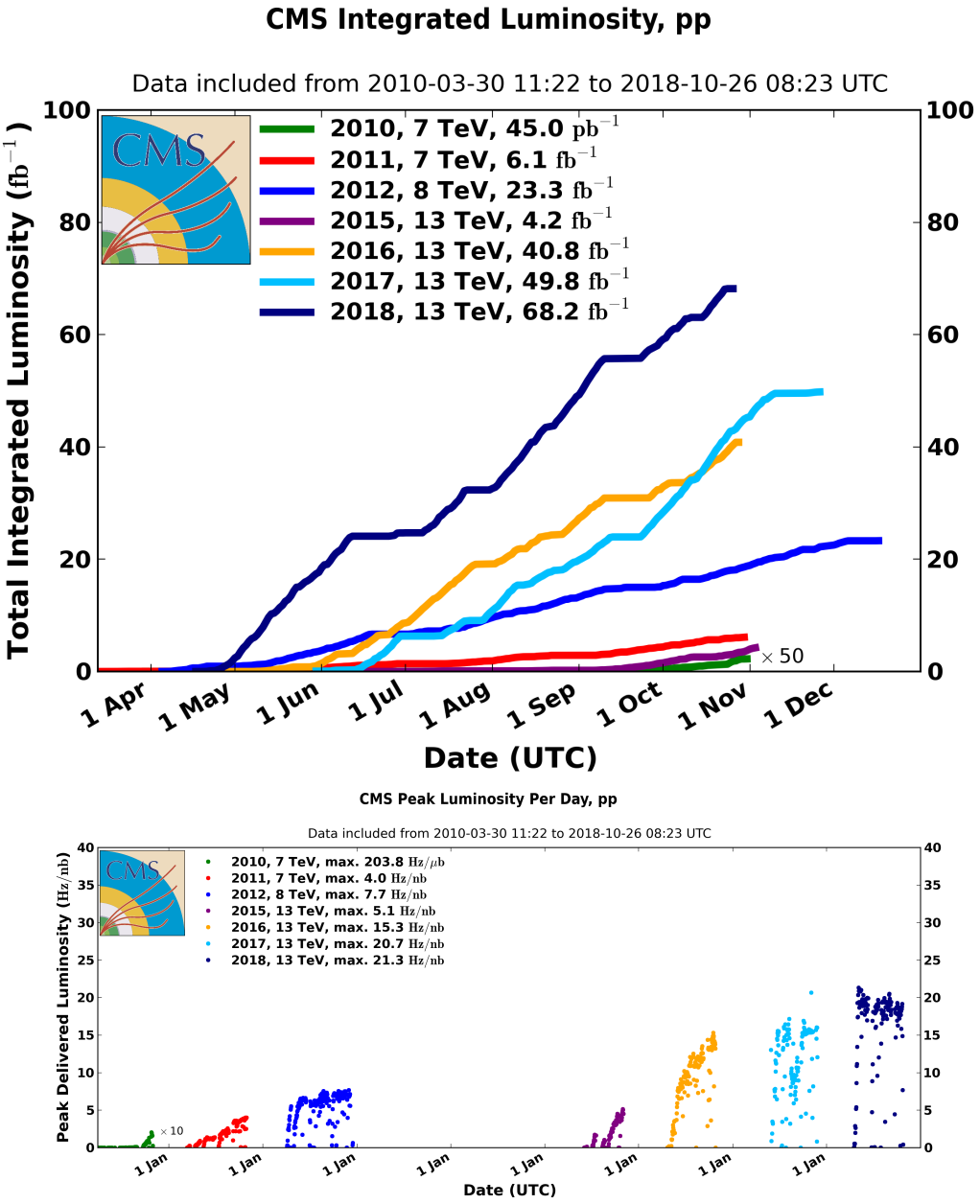

Figure 2.2 – Cumulative integrated luminosity versus day (top) and peak instantaneous luminosity versus day (bottom) delivered to the CMS experiment during stable beams for pp collisions, for the 2010 (green), 2011 (red), 2012 (dark blue), 2015 (purple), 2016 (orange), 2017 (light blue) and 2018 (dark blue) data taking.

sample corresponding to more than 4 fb^{-1} was delivered to ATLAS and CMS.

In 2016, operation immediately started at 13 TeV with a bunch spacing of 25 ns and with a relatively bold set of operational parameters. Even though there were some problems with the SPS beam dump which made it necessary to restrain the number of

bunches to about 2100, the nominal LHC luminosity of 10^{34} cm^{-2}s^{-1} was reached in June 2016. Instantaneous luminosity was increasing steadily during summer reaching more than $1.4 \cdot 10^{34}$ cm^{-2}s^{-1} in the end. The 2016 pp data taking ended in late October with a total delivered luminosity of about 41 fb^{-1}, as illustrated on the left in Fig. 2.3.

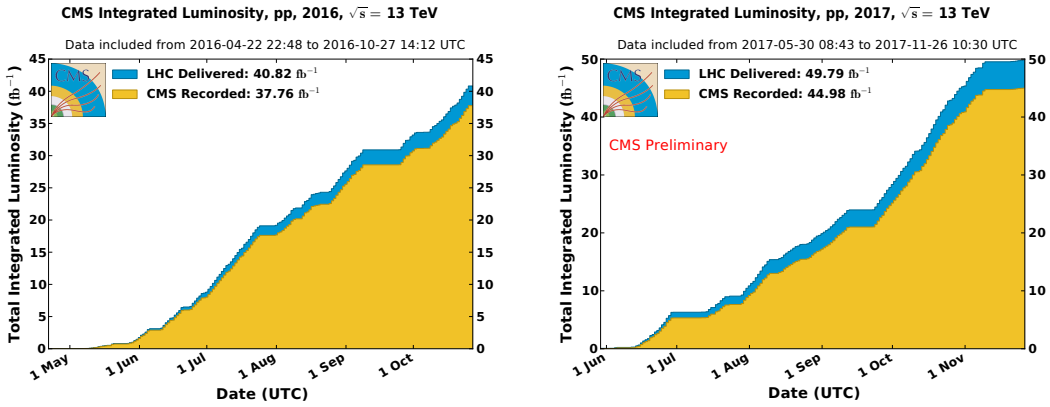

Figure 2.3 – Cumulative offline luminosity versus day delivered to CMS (blue), and recorded by CMS (yellow) during stable beams for pp collisions at centre-of-mass energy of $\sqrt{s} = 13$ TeV, in 2016 (left) and 2017 (right). The delivered luminosity accounts for the luminosity delivered from the start of the stable beams until the LHC requests CMS to turn off the sensitive detectors to allow a beam dump or studies. The luminosity is given as determined from counting rates measured by the luminosity detectors after offline validation.

In 2017 first beams started circulating in the LHC in May. Using a new system put in place, the size of the beams when they meet at the centre of the experiments has been significantly reduced. The more squeezed the beams, the more collisions occur each time they meet. Thanks to these improvements, the instantaneous luminosity record was smashed, reaching $2.06 \cdot 10^{34}$ cm^{-2}s^{-1}, or twice the nominal value and up to 60 collisions were produced at each crossing as can be best seen from Fig. 2.4.

The LHC has far exceeded its target for 2017. It has provided its two major experiments, ATLAS and CMS, with 50 fb^{-1} of data as illustrated on the right in Fig. 2.3. This result is all the more remarkable because the machine experts had to overcome a serious setback. A vacuum problem in the beam pipe of a magnet cell limited the number of bunches that could circulate in the machine. Several teams were brought in to find a solution. Notably, the arrangement of the bunches in the beams was changed. After a few weeks, luminosity started to increase again.

On Saturday, 28 April 2018, the operators of the LHC successfully injected 1200 bunches of protons into the machine and collided them. This formally marks the beginning of the LHC's 2018 physics season. The start of the physics run comes a few days ahead of schedule, continuing the LHC's impressive re-awakening since the end of its annual winter hibernation just over a month ago. The LHC operators will keep ramping up the number of bunches, aiming to hit 2556 bunches in total. This will help them achieve their target of 60 fb^{-1} of proton-proton collisions this year delivered to both ATLAS and CMS. This is the last year with collisions before the LHC enters a period of hibernation

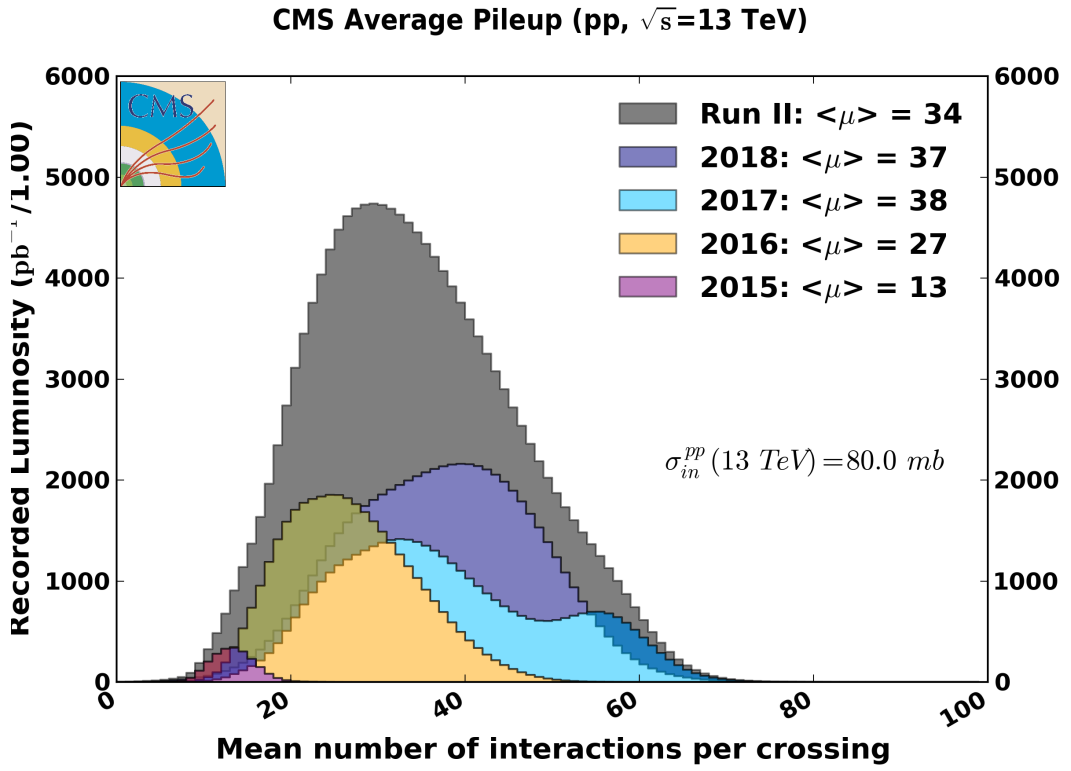

Figure 2.4 – Mean number of interactions per bunch crossing for the proton-proton Run II at 13 TeV.

until spring 2021, known as Long Shutdown 2 (LS2), during which the machine and the experiments will be upgraded. All four experiments will therefore hope to maximise their data-collection efficiency to keep themselves occupied with many analyses and new results over the two-year shutdown, using high-quality data collected this year.

This thesis work exploits both 2016 and 2017 data samples accounting to roughly 77 fb^{-1} of certified data available for Higgs hunting.

2.3 Compact Muon Solenoid

The CMS detector is a general-purpose detector with its key characteristics revealed by its name, Compact Muon Solenoid:

- **Compact** - small dimensions compared to its mass,
- **Muon** - advanced muon detection system,
- **Solenoid** - solenoidal superconducting magnet.

In order to understand how this gigantic yet compact detector works we will first go

through its design where only mentioning its sub-detectors. The coordinate system used, each sub-detector, and the trigger system will then be presented in more details.

2.3.1 Detector design

CMS had to be designed to be able to cope with great expectations of achieving the LHC physics program which includes the search for SM Higgs boson and the investigation of the EWSB, the search for Supersymmetry and other BSM physics at the TeV scale. Since the mass of the Higgs boson is the free parameter in ths SM, it had to be searched for in a large mass window from 100 GeV to about 1 TeV. Due to the strong dependence of the Higgs boson branching ratios on its mass as discussed in Section. 1.2.2, this meant that the detector has to be able to reconstruct and identify the physics objects of the final states best suited for a given mass. Moreover, the unprecedented collision parameters of the LHC set a strong technical constraints on the design which had to be met. Detector had to have fast response of about 20-50 ns in order to cope with LHC collision rates. High granularity was needed to minimize the probability that particles from pileup interactions end up in the same detector element as particles from the main interaction, leading to a high number of electronic channels. Finally, detector had to have high radiation resistance in order to survive the high flux of particles from collisions that damage detector components over the long term.

The CMS detector is located in the underground hall, about 100 meters below the village of Cessy in France. It is 21 m long, 15 m wide and 15 m high. CMS is like a giant filter with onion structure, where each layer is designed to stop, track or measure a different type of particle emerging from pp and heavy ion collisions as illustrated in Fig. 2.5. Determining the energy and momentum of a particle gives clues to its identity and enables physicists to discover new particles and new phenomena with careful studies and thorough analysis.

The detector is built around a huge solenoid magnet which takes the form of a cylindrical coil of superconducting cable that is cooled down to -268.5°C. The coils generate nominal magnetic field of 4 Tesla, about a hundred thousand times that of the Earth's magnetic field. In practice, the working value has been set to 3.8 Tesla for the matters of safety and security during operation time.

Particles emerging from collisions first pass through the tracker system which detects charged particles tracks. Tracks positions are charted as they move through successive silicon layers, allowing us to measure their momentum. Outside the tracker are calorimeters, designed to stop the particles and measure their energy. The Electromagnetic Calorimeter (ECAL) is a sub-detector designed to measure the energy of photons and electrons. Next layer in this onion structure is the Hadronic Calorimener (HCAL), a sub-detector designed to detect hadrons. The size of the magnet allows for the tracker and both calorimeters to be placed inside its coil making it very compact.

As already mentioned, one of the key features of the CMS is to measure muons precisely. This is possible thanks to the four layers of muon detectors that are interleaved with the

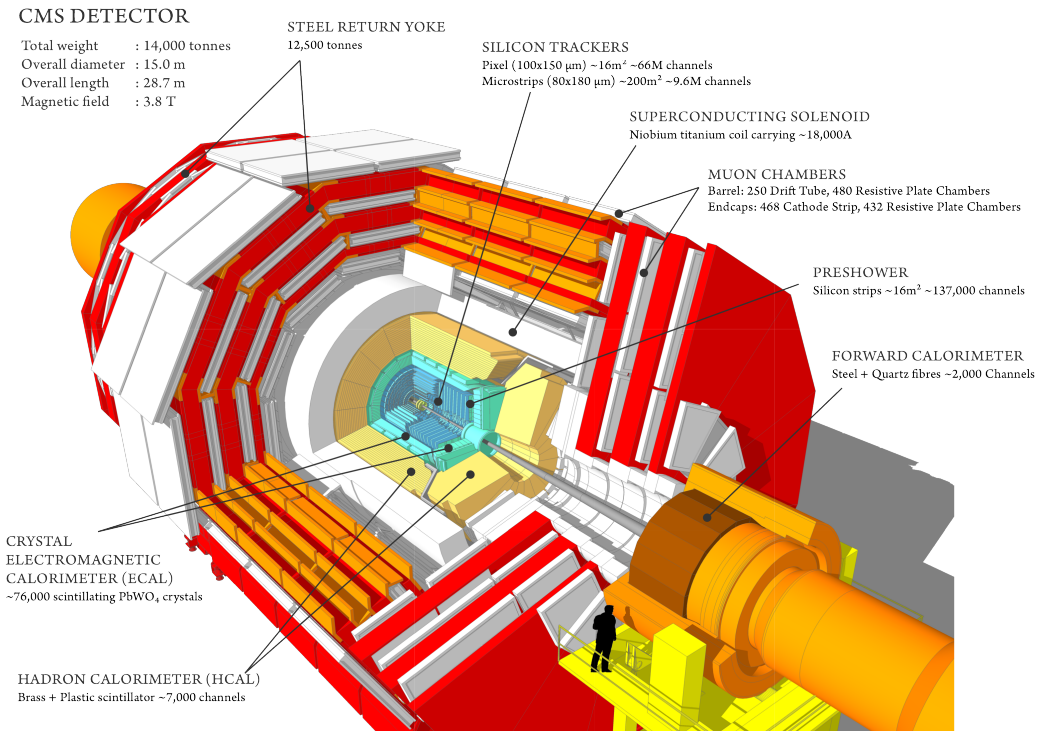

Figure 2.5 – A perspective view of the CMS detector with major sub-detectors indicated.

iron yoke. The only particles from the SM that are not detectable by the CMS detector are neutrinos. Although their presence can not be directly observed it can be inferred from the missing energy in the event.

As we already know, within the LHC bunches of particles collide up to 40 million times per second, so a trigger system that selects only interesting events is essential. It is build in two levels: the so called Level 1 (L1) trigger based on fast hardware and the High Level Trigger (HLT) based on software running on computers. The L1 significantly reduces the event rate using very simple detector information and feeds it to the HLT. The lower rate in the HLT allows time for a more detailed analysis of the event to allow for a better informed decision weather to discard the event or save it on tape for further analysis.

A much more detailed and comprehensive description of the CMS detector can be found in References [56, 57].

2.3.2 Coordinate system

Before we dissect CMS to study its parts in detail we have to deal with the very technical but important task of defining the reference frame for all the measurements discussed in this thesis.

In the CMS detector we have adopted a standardized coordinate system to describe the reconstructed particles. Its origin is at the nominal interaction point at the center of the detector. The x axis points to the center of the LHC ring and the y axis points upwards. The z axis coincides with the proton beam direction and points towards the Jura mountains, giving a right-handed coordinate system. Cylindrical coordinates are commonly used, where the transverse plane is given by the $x-y$ plane. The azimuthal angle ϕ is measured from the x axis in the $x-y$ plane and takes values in $[-\pi, \pi]$. The polar angle θ is measured from the z axis and takes values in $[0, \pi]$.

The particle trajectories are often described in the transverse plane because the activity in this plane is interesting when searching for new phenomena. Momentum and energy measured perpendicular to the beam direction are calculated from the x and y components of energy and momentum (for a massless particle, $E_T = p_T$). The transverse momentum is the projection of any momentum $\mathbf{p}$ onto the $x-y$ plane and often used to denote the magnitude of this projected vector:

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.2}$$

The rapidity of a particle is given by:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{2.3}$$

and approaches the pseudorapidity in the limit where the mass is negligible compared to the energy, $m/E \ll 1$:

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_z}{p - p_z} \right) = -\ln \tan \frac{\theta}{2}, \tag{2.4}$$

which is for example always true for electrons and muons in the scope of this thesis. Keeping in mind the cylindrical shape of the detector, the η coordinate makes the difference between two parts of the sub-detectors: the central part called *barrel*, and the two opposite forward parts called the *endcaps*.

The angular distance between two particles with azimuthal angles ϕ_i and pseudorapidities η_i is commonly expressed as ΔR which is defined as:

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}. \tag{2.5}$$

A particle escaping the detection creates an imbalance in the total transverse energy measurement. Because of this we define the transverse missing momentum p_T^{miss} as the negative momentum sum of all reconstructed particles projected onto the transverse plane, and if the detector is hermetic it is usually interpreted as the total transverse momentum of neutrinos in the event.

2.3.3 Sub-detectors

As previously mentioned, the CMS detector consists of sub-detectors that have different tasks in particle reconstruction puzzle. This can be seen in Fig. 2.6 where it is illustrated

what kind of signatures are left by the passage of different types of particle through the onion-like structure of the CMS detector. We will describe the design and principle of work of all sub-detectors with particular emphasis put on the ECAL and muon system, which play a major role in the $H \rightarrow ZZ^{\star} \rightarrow 4\ell$ analysis.

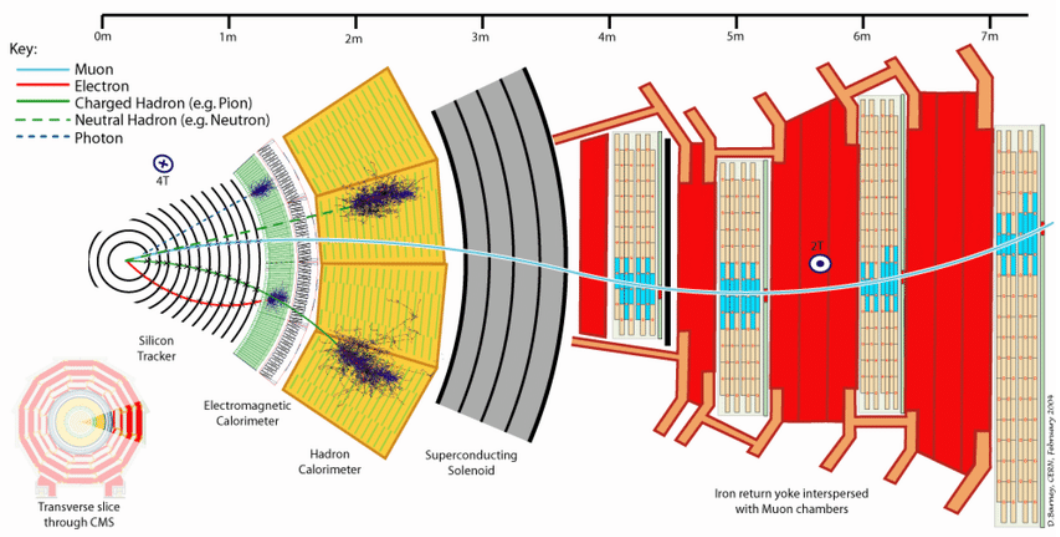

Figure 2.6 – Schematic view of a transverse slice of the CMS detector, illustrating the specific signatures of different types of detected particles.

Tracking system

The closest sub-detector to the particle collisions, and therefore the first detector traversed by the particles coming out of the collisions is the tracking system, or *tracker*. It occupies a cylindrical volume of 5.8 m in length and 2.5 m in diameter, and is immersed in the homogeneous magnetic field provided by the solenoid. The object of the tracker is to reconstruct the trajectories of all charged particles resulting from a collision. By measuring the curvature of the track in the magnetic field, from the Lorentz force the momentum and charge of the particle can be deducted. Considering the large number of interactions per bunch crossing of the LHC, a crucial task of the tracking system is to resolve the large number of pileup interactions and separate them from the interesting hard interaction by reconstructing several vertices of interest in each event. It is also very important to be able to reconstruct secondary vertices indicating late decays of particles such as b-hadrons.

Tracking is achieved by having successive layers of sensitive elements that are capable of registering the passage of a charged particle via its ionization effect. A schematic representation of the overall tracker layout is shown in Fig. 2.7.

The active elements of the CMS tracker are organized in layers and made of thin silicon sensors which are further split in the $x - y$ plane into individual readout cells of rectangular shapes. Two different technologies are used: silicon pixel and silicon strip sensors. A charged particle crossing one of these cells will produce electron-hole pairs which will

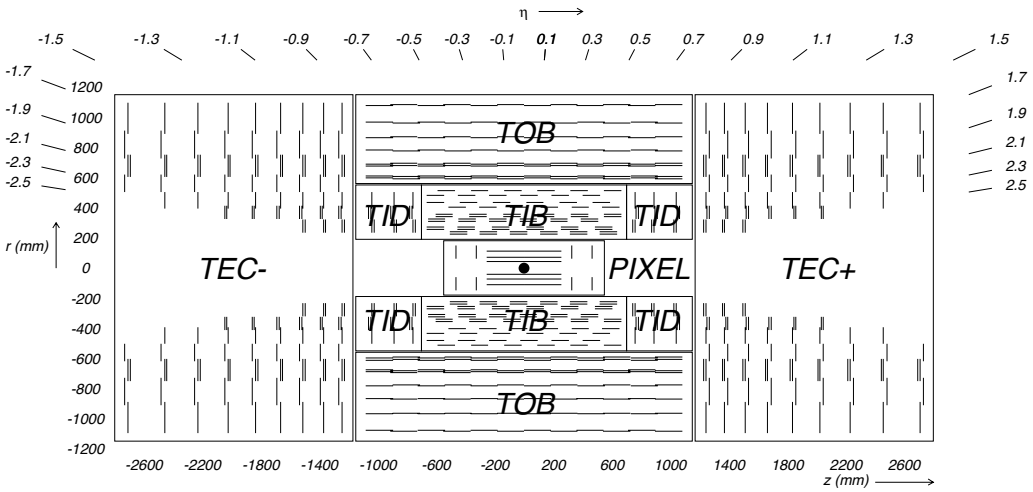

Figure 2.7 – Schematic cross section through the CMS tracker. The pixel detector contains barrel and endcap modules; the silicon strip detector contains two collections of barrel modules: the tracker inner barrel (TIB) and the tracker outer barrel (TOB), and the two collections of endcap modules: the tracker inner disc (TID) and the tracker endcap (TEC).

drift under an applied electric field, giving rise to a current pulse. By having a precise knowledge of the spatial position of the silicon sensors which generate an electric signal, one can reconstruct the particle trajectory.

The innermost part is made of 65 million silicon pixels of size 100×150 $\mu\mathrm{m}^2$, which are arranged in three cylindrical barrel layers at radii between 4.4 cm and 10.2 cm, and two endcap disk layers. This pixel detector achieves a spatial resolution of 10 μm in (x, y) and 20 μm along z, allowing for a three-dimensional vertex reconstruction. The principal reason for the small pixel size, or high granularity, is the need to separate tracks from particles that are near-by and to be capable of identifying secondary vertices in the decay of heavy mesons. However, a cost of having such a large number of readout channels is having more on-chip electronics for the signal readout and high voltage supply which in turn demands a large cooling capacity. All this adds passive or *dead* material to the detector, which increases the chance of particle interactions with the detector with phenomena such as multiple scattering, bremsstrahlung, photon conversion, and nuclear interactions, all of which complicate particle reconstruction.

As discussed in Section 2.2.3 the original performance goal for the LHC was well surpassed reaching instantaneous luminosity twice the nominal value in 2017. It was known that the current pixel system will not sustain the extreme operating conditions expected in 2017. In order to cope with such high rate of collisions and unprecedented values of pileup CMS decided to replace the described pixel detector in the year-end technical stop of 2016-2017 with one that can maintain a high tracking performance at luminosities up to $2 \cdot 10^{34}$ cm^{-2}s^{-1} and average pileup up to and exceeding 50.

Figure 2.8 shows a conceptual layout for the upgraded pixel detector. The 3-layer barrel and 2-disk endcap system is replaced with a 4-layer barrel and 3-disk endcap system for

four hit coverage. Main role of the additional fourth barrel layer at a radius of 16 cm is to provide redundancy in pattern recognition and to reduce fake rates with high pileup.

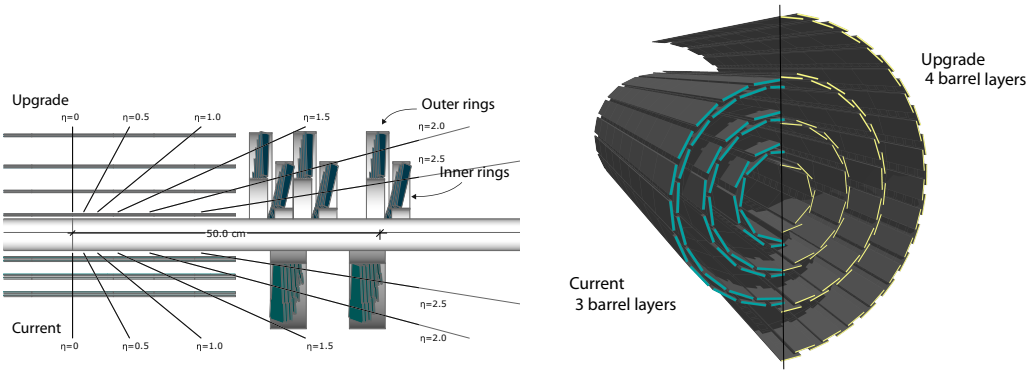

Figure 2.8 – Conceptual layout comparing the different layers and disks in the previous and upgraded pixel detectors (left). Transverse-oblique view comparing the pixel barrel layers in the two detectors (right).

Since the extra pixel layer could easily increase the material of the pixel detector, the upgrade detector, support, and services were redesigned to be lighter than the present system, using an ultra-lightweight support with CO_2 cooling, and by relocating much of the dead material, like the electronic boards and connections, out of the tracking volume. A study of the "material budget" reduction by the upgrade of the pixel detector was done and is shown in Fig.2.9 as a comparison of the radiation length and nuclear interaction length of the pixel detector as a function of η before and after the upgrade.

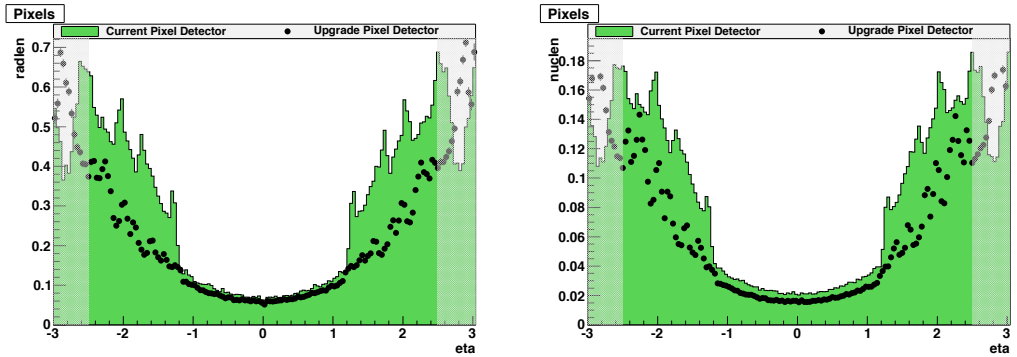

Figure 2.9 – The amount of material in the pixel detector shown in units of radiation length (left), and in units of nuclear interaction length (right) as a function of η; this is given for the pixel detector before the upgrade (green histogram), and the Phase 1 upgraded detector (black points). The shaded region at high $|\eta|$ is outside the region for track reconstruction.

Improvements from the new detector cannot be summed up by one number, but are characterized by higher efficiencies, lower fake rates, lower dead time, lower data loss, and an extended lifetime of the detector. This leads to better muon identification, b-tagging, photon and electron identification, and tau reconstruction, both offline and in the HLT. As it will be shown in the following chapters this upgrade had a significant

impact on the $H \to ZZ^{\star} \to 4\ell$ analysis as well, as best seen from the significant reduction of fake electrons in 2017 with respect to 2016 before the upgrade was employed. A comprehensive and detailed description of the pixel detector upgrade can be found in Reference [58].

The rest of the tracker consists of silicon strips. It covers a total surface area of $65.6\,\mathrm{m}^2$ with 4.6 million channels. They are arranged parallel to the beam axis in the 4 + 6 barrel layers, which extend outwards to $r = 1.1$ m, and radially in the 3 + 9 endcap layers, which provide a pseudorapidity coverage of $|\eta| < 2.5$. The distance between neighboring strips varies from 80 to 205 μm depending on the location in the detector. Silicon strips allow covering large surface areas by reducing the number of readout channels and all the required electronics. This is achieved by increasing the length of a single silicon cell from a hundred μm to several cm, where length is defined as the coordinate that matters the least for the curvature and momentum measurement (the z coordinate in the barrel, and the radial direction in the endcap).

Electromagnetic calorimeter

Next in line is the electromagnetic calorimeter (ECAL). In order to build up a full event picture, CMS must measure the energies of particles precisely. The CMS ECAL provides the energy measurement for electrons and photons, and more so allows triggering on these objects. It is a hermetic, high-resolution, high-granularity homogeneous calorimeter made of lean tungstate ($PbWO_4$) crystals. Its schematic view can be seen in Fig. 2.10 and Fig. 2.11.

The main part of the ECAL calorimeter is cylindrical inner part called barrel (EB) closed at both ends with two endcap discs (EE). The barrel is further divided into two cylinder halves, EB+ and EB-. Each half consists of 18 super-modules each of which has a mass of approximately 1.5 t and contains 1700 crystals arranged in $\eta - \phi$ geometry: 20 crystals in ϕ and 85 in η. Every super-module covers pseudorapidity range $|\eta| < 1.479$ and 20° in ϕ. Super-modules are further divided into 4 modules. The module number 1 contains 20 crystals in ϕ and 25 in η, while all other modules contain 20 crystals in ϕ and 20 in η coordinate. A module is constructed out of 40 or 50 sub-modules which are lightweight glass-fibre boxes in groups of ten. To avoid cracks aligned with particle trajectories, crystals are mounted so that their axes make a small 3° angle with respect to the direction from the nominal interaction point. Still, some gaps which are referred to as *cracks*, remain between modules and complicate the energy reconstruction. Largest cracks present are at $\eta = 0$ and at the EB-EE transition.

Two EE part consists of four *dees*, each containing 3662 crystal complexes in $x - y$ geometry and covering pseudorapidity range $1.479 \leq |\eta| \leq 3$. The crystals are grouped into 5×5 super-crystals. The crystals have front face cross section of $28.62 \times 28.62\,\mathrm{mm}^2$, rear face cross section of $30 \times 30\,\mathrm{mm}^2$, and length of 220 mm.

For extra spatial precision, the ECAL also contains the ECAL preshower (ES) detector, sitting in front of the EE. It allows CMS to distinguish between single high-energy photons

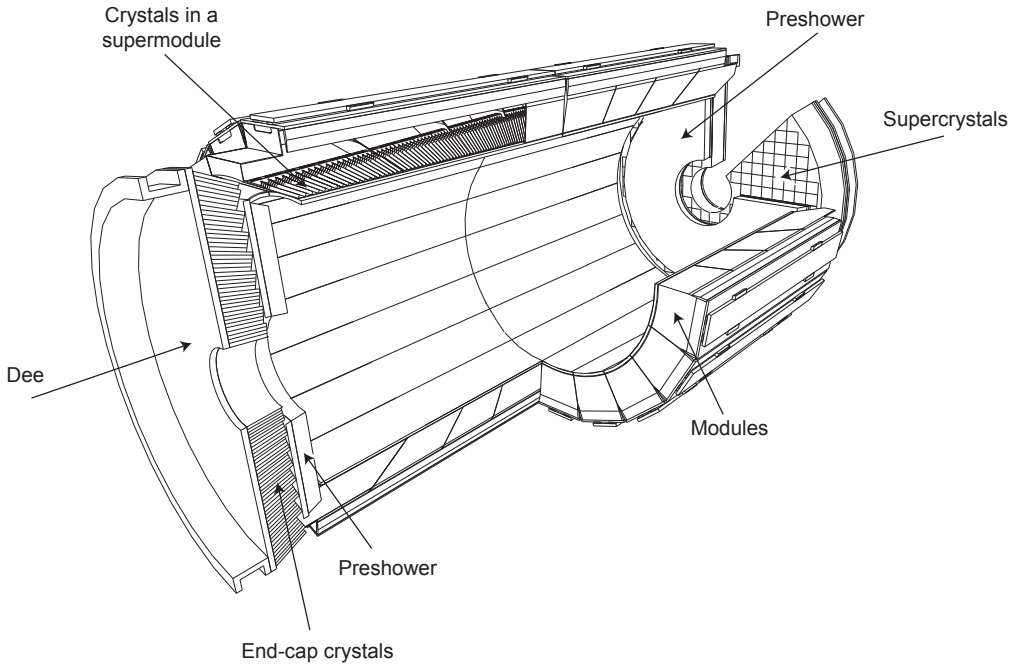

Figure 2.10 – Schematic view of the CMS ECAL and the mechanical structure. In the barrel, crystals are grouped into modules and super-modules. Each endcap consists of two half-disks or *dees*. The pre-shower detector covers most of the endcap surface.

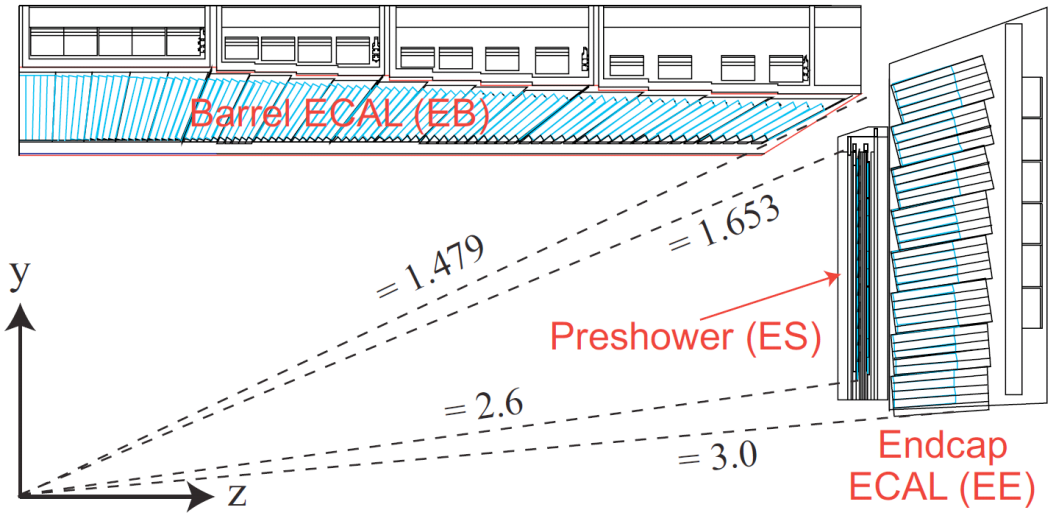

Figure 2.11 – Longitudinal view of a ECAL quadrant. The pseudorapidity coverages of the barrel, encdcap, and preshower systems are indicated.

and the less interesting close pairs of low-energy photons coming from the $\pi^0 \to \gamma\gamma$ decays. Approximately 6% to 8% of the energy in an electromagnetic shower is deposited in the ES. The ES has a much finer granulatiry than the ECAL with the detector strips 2

mm wide, compared to 3 cm wide ECAL crystals. It is made of two lead layers followed by the silicon sensors. Each ES detector consists of 8 m^2 of silicon. Silicon sensors measuring $63 \times 63 \times 0.3\,mm^2$ are divided into 32 strips and arranged in a circle to cover the majority of the EE. For optimal performance, the ES detector silicon sensors are kept at the temperature of -15°C to -10°C. When a photon passes through the lead layer it causes an electromagnetic shower, which the silicon sensors detect and measure. When seemingly high-energy photons are found in the ECAL we can extrapolate their pairs back to the centre of the collision and look in the ES along the way. From the energy deposited there we can deduce if they really were individual high-energy photons or photon pairs.

$PbWO_4$ crystals. The most suitable choice for the construction of ECAL proved to be crystals of lead tungstate shown in Fig. 2.12. Dimensions of these crystals are $2.2 \times 2.2 \times 23\,cm^2$ in the EB and $3 \times 3 \times 22\,cm^2$ in the EE part of the calorimeter. The mass of each crystal is 1.5 kg. Resistance to the high levels of radiation thanks to high density, small radiation length and short relaxation time makes them ideal for the CMS experiment. When an electron or a photon of high energy passes through crystals, it starts an electromagnetic shower. Main processes of the electromagnetic shower are bremsstrahlung radiation of electrons and photon conversion into electron-positron pairs. These processes continue until the energy of the particles is so low that the production of new particles is no longer possible. When this happens the dominant mechanism of energy loss becomes ionization, Compton scattering, and photoelectric effect. Electromagnetic shower can be characterized with the radiation length X_0 and the Moliere raduis R_M. Radiation length is defined as the mean distance over which a high-energy electron loses all but $\frac{1}{e}$ of its energy by bremsstrahlung, and $\frac{7}{9}$ of the mean free path for pair production by a high-energy photon, which in the case of $PbWO_4$ crystal is $X_0 = 8.9$ mm. The Moliere radius is the radius of the base cone within which 90% of the energy of the shower is contained and in the case of $PbWO_4$ equals $R_M = 22$ mm. The particles from the shower excite atoms in the crystal and they emit photons of blue light when returning to the ground state. It is important to note that the intensity of the light emitted is proportional to the energy absorbed by the crystal so in this way we can measure the energy of the incident electron or photon.

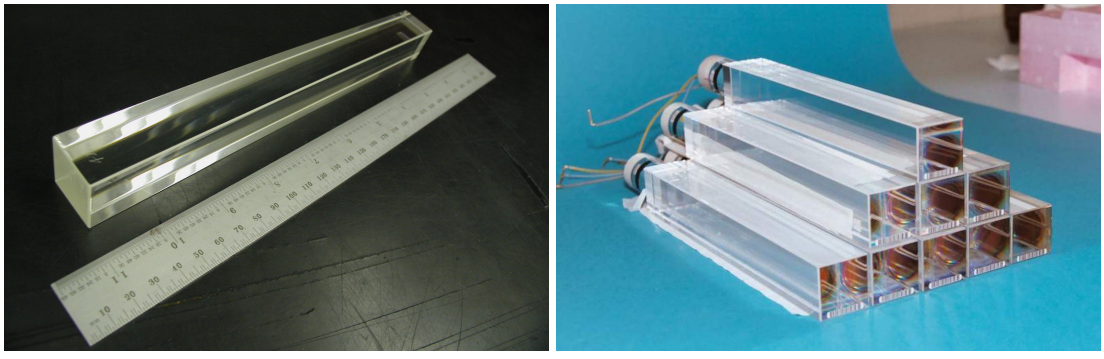

Figure 2.12 – Lead tungstate crystals used in the EB part of the ECAL detector (left) and in the EE part of the ECAL detector (right).

Photodetectors. The photons for the blue light emitted in crystals then reach the photodetector. Photodetectors used in ECAL are avalanche photodiodes (APDs) and vacuum phototriodes (VPTs). APDs are used in the EB part of the detector and are made of semiconducting silicon with a strong electric field applied to them. When a scintillation photon strikes the silicon it knocks an electron out of an atom. The electron is then accelerated by the electric field, and strikes other atoms during his travel through the crystal. These strikes transfer the energy to the electrons and knock them from the atom. As these are also accelerated, this method produces an avalanche of electrons with their numbers increasing exponentially. Thanks to this method APDs are able to produce a very high current in a short time. This is important since the lead tungstate crystals give a low yield of light for each incident particle. The signal is then amplified, digitized and transported by a fiber optic cables. In EE VPTs are used because radiation is too high to use a silicon photodiode. VPTs contain three electrodes within a vacuum. The whole process starts when the light strikes atoms in the first electrode and knocks the electron. The voltage difference between electrodes accelerates the electrons into the second, then third electrode, producing a batch of accelerated electrons. This also creates a large current from the initial tiny amount of scintillation light, which is again turned into a digital signal and sent along the optic fibers.

Laser monitoring. Although the ECAL lead tungstate crystals are resistant to radiation they are not insensitive to it. Their optical transparency decreases by few percent during a run. The loss of transparency has been observed to be strongly correlated with the LHC instantaneous luminosity and with the crystal position in the detector. The main mechanism leading to the transparency loss observed in the detector is the formation of color centers due to ionization radiation. This type of damage is not cumulative and recovers with time when the irradiation stops. Another component of the transparency loss is observed to arise after hadron irradiation due to interactions of energetic hadrons with the crystal lattice. It does not recover at room temperature when irradiation stops and therefore builds up during the data-taking. The laser light injection system has been used to continuously monitor transparency loss during the data-taking [59]. The laser pulses are injected into the crystals via optical fibers. The response is normalized by the laser pulse magnitude measured using silicon photodiodes. The ratio of the crystal response to the photodiode measurement gives the crystal transparency and is shown in Fig. 2.13 for ECAL. To account for the transparency losses, time dependent corrections are applied to measured particle energy.

ECAL calibration. Another technical challenge is calibration, which consists of a global component that gives the absolute energy scale, and of a channel-to-channel relative component, or the so called intercalibration. Intercalibration takes into account the difference in the scintillation light yields between crystals which can be up to ∼15% in the EB crystals and up to ∼25% for EE. Corrections from laboratory measurements and calibration of the crystal light yield and photodetector response reduced the initial channel-to-channel variations. Even better precision on intercalibration constants was further achieved with the use of cosmic rays and electron test beams yielding in a percent-level precision.

Electron test beam measurements on a 3×3 matrix of ECAL crystals show that the

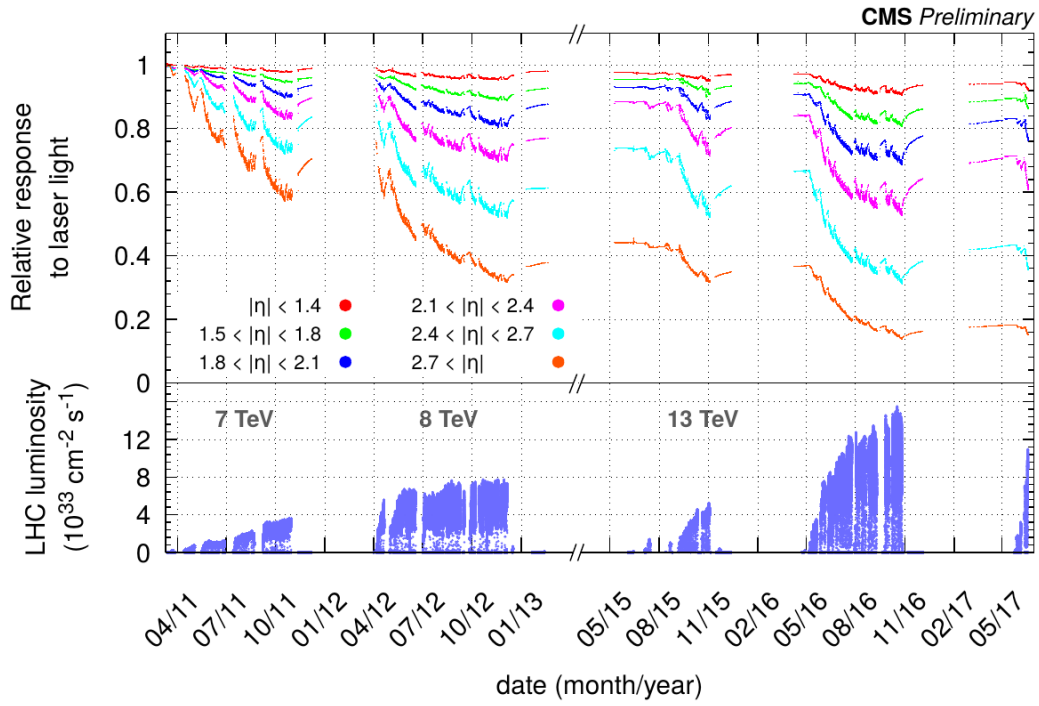

Figure 2.13 – Time evolution of the ECAL response monitoring laser $R(t)$. The reduction in response during data-taking periods, most pronounced for large psudorapidities, is caused by radiation damage to ECAL crystals. The response is monitored and ultimately corrected via the laser monitoring system. Some recovery of the response during periods without collisions is also visible.

measured energy resolution is described by the usual parametrization [60]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 \oplus \left(\frac{12\%}{E}\right)^2 \oplus (0.3\%)^2, \tag{2.6}$$

where the energy E is given in GeV. Noise from the readout electronics, corresponding to the second term, contributes only at very low energies. At intermediate energies the first term called the *stochastic term* contributes the most, accounting in particular for shower-by-shower variations in the scintillation light yield. For electrons above 50 GeV the resolution is mostly determined by the constant term.

Hadronic calorimeter

The hadronic calorimeter (HCAL) serves to measure the energy of long-lived hadrons that traverse the tracker and ECAL. It provides the only energy measurement for neutral hadrons and complements the momentum measurement of the tracker for charged hadrons, thus playing a major role in the reconstruction of jets. Its wide extension in pseudorapidity captures a large fraction of particles emerging from the interaction point, allowing for a reliable measurement of the missing transverse energy which is a

signature of otherwise undetected particles such as neutrinos.

The CMS HCAL is a hermetic and compact sampling calorimeter [57] located within the solenoid. The structure of the CMS HCAL is illustrated in Fig.2.14. It is made of repeating layers of dense absorber and tiles of plastic scintillator. When a hadronic particle hits a plate of absorber, in this case brass or steel, an interaction can occur producing numerous secondary particles. As these secondary particles flow through successive layers of absorber they too can interact and produce a shower of particles. As this shower develops, the particles pass through the alternating layers of active scintillation material causing them to emit blue-violet light. Within each tile tiny optical wavelength-shifting fibers, with a diameter of less than 1mm, absorb this light. These shift the blue-violet light into the green region of the spectrum, and clear optic cables then carry the green light away to readout boxes located at strategic locations within the HCAL volume.

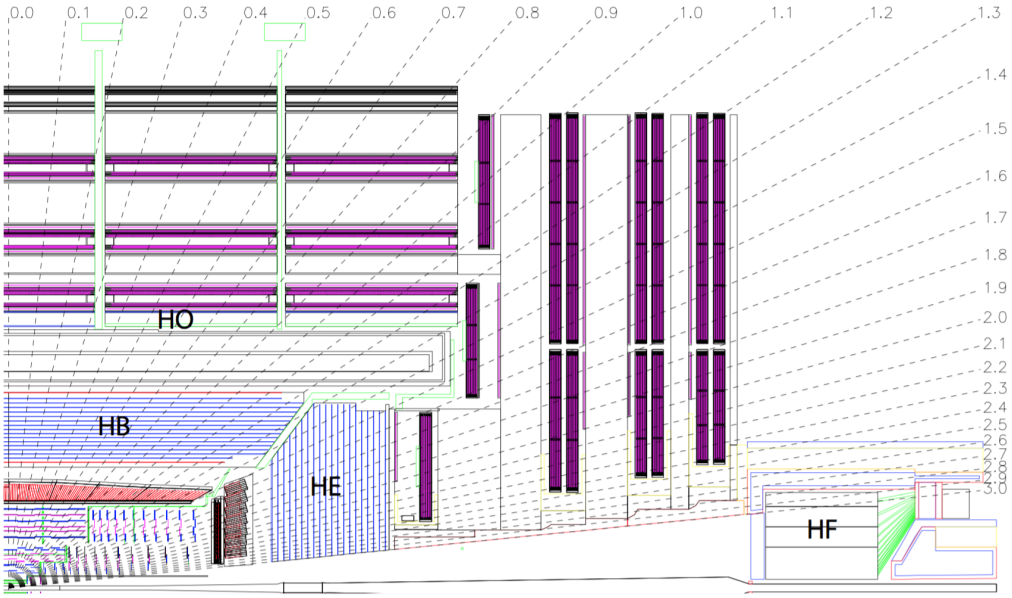

Figure 2.14 – Longitudinal view of the geometry in a HCAL quadrant. The location and psudorapidity coverage of the barrel (HB), outer (HO), endcap (HE), and very forward (HF) hadron calorimeters are illustrated.

The HCAL barrel (HB) is located between the outer extent of the EB and the inner extent of the magnetic coil. It is made of 2304 towers of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ that are read out as a whole. Since it is not enough to fully absorb hadronic showers, an outer hadron calorimeter (HO) is placed outside of the solenoid. It increases the total number of interaction lengths to 10, decreasing the leakage of energetic hadronic jets into the muon system, which helps to reduce non-Gaussian tails in the energy resolution as well lowering the odds of such jets being misidentified as muons. The HB covers the pseudorapidity range $|\eta| < 1.3$, which the HCAL endcaps (HE) extend to $|\eta| < 3.0$. Beyond this, two forward hadron calorimeters (HF) are located 11.2 m from the interaction point and reach $|\eta| = 5.2$, ensuring good hermeticity. Quartz fibers in the HF are used to collect the energy of the showers developing in the iron absorber and to

produce the signal by exploiting the Cherenkov effect.

Superconducting magnet

Next in line is the CMS magnet, the corner stone around which the experiment is built. Its task is to bend the trajectories of charged particles emerging from the interaction point. The more momentum the particle has the less its path is curved by the magnetic field, thus tracing its path enables us to estimate its momentum. When combined with high precision position measurements in the tracker and muon detectors, this allows a very accurate measurement of the momentum even for high energy particles.

The CMS magnet shape is a solenoid made of 2168 coils of wire cooled town to $-268.5°C$. It is also superconducting, allowing nominal current of 19.14 kA to flow without any resistance. It is interesting that, at ordinary temperatures, the strongest possible magnet has only half the strength of the CMS solenoid. The flux is returned through a 10000 t iron yoke which contains 5 wheels and 2 endcaps composed of three disks each. The main role of the yoke is to increase the field homogeneity in the tracker and to reduce the stray field by returning the magnetic flux of the solenoid.

In order to allow accurate reconstruction and Monte Carlo (MC) simulation of events, a detailed map of the CMS magnet magnetic field has been made using the cosmic muons [61]. In the tracker volume the field has been mapped with a precision better than 0.1%. Map of the magnetic field strength in the CMS detector is shown in Fig. 2.15.

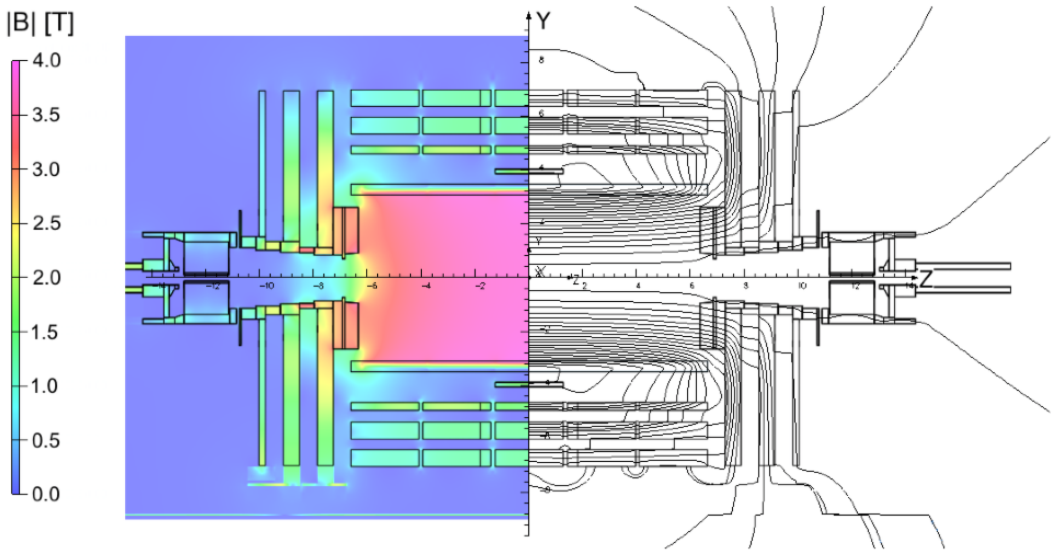

Figure 2.15 – Map of the $|B|$ field (left) and field lines (right) predicted for a longitudinal section of the CMS detector by a magnetic field model at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb.

Muon system

The muon system has been designed to provide an efficient muon trigger and a precise measurement of the muon momentum and charge [62, 63], even without relying on the information from the tracking system. Given the shape of the CMS solenoid magnet, the muon detector is divided into a cylindrical barrel section, and two planar endcap regions with 25000 m^2 of detection planes used. Muon detectors are embedded in the iron return yoke of the CMS magnet, as shown in the Fig. 2.16.

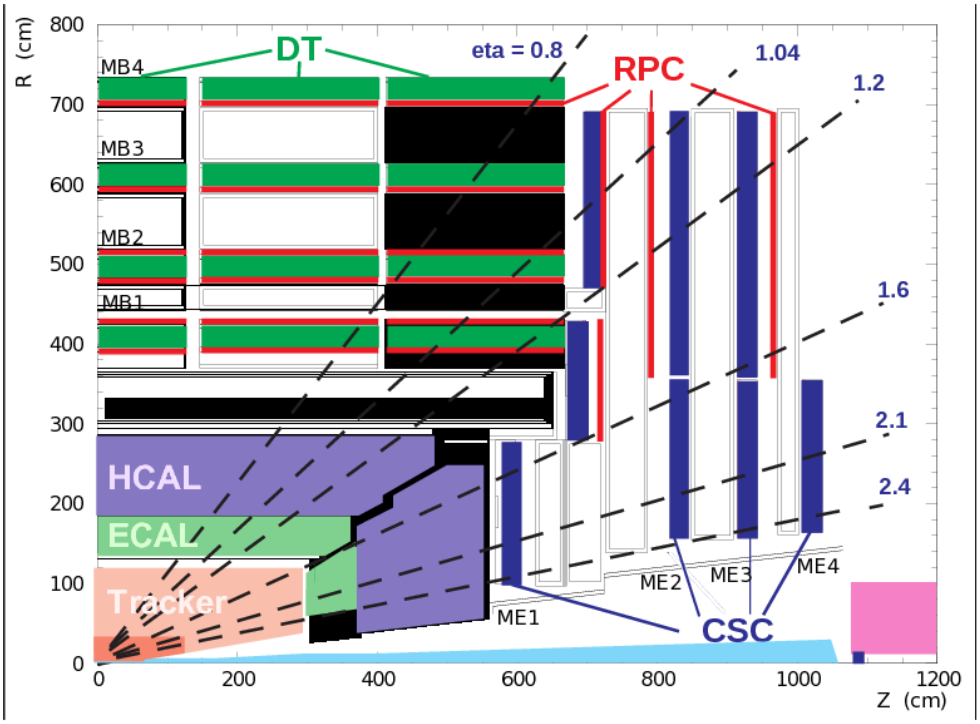

Figure 2.16 – Longitudinal sectional view of a quarter of the CMS detector, showing the four DT stations in the barrel MB1-MB4 (green), the four CSC stations in the endcap ME1-ME4 (blue), and the RPC stations (red).

Drift Tube Chambers. The Drift Tube (DT) system measures muon positions in the barrel part of the detector. Each 4 cm wide tube contains a stretched wire within a gas volume. When a muon, or any charged particle, passes through the volume it knocks the electrons off the atoms in the gas. These follow the electric field ending up at the positively charged wire. By registering where along the wire electrons hit, as well as by calculating the muon original distance away from the wire DTs give two coordinates of the muon position. DT chambers are located in the barrel region $|\eta| < 1.2$, and are divided in five wheels along the z coordinate, each including 12 azimuthal sectors. Each DT chamber is, on average, $2 \times 2.5\,\mathrm{m}$ in size. The basic constituent of a DT chamber is a $42 \times 13\,\mathrm{mm}^2$ cell. A cell is bounded by two parallel aluminum planes and by I-shaped aluminum beams serving as cathodes. The anodes are 50 μm stainless steel wires located in the centre of the cells. The gas mixture within a cell is composed of 85% argon and 15% CO_2. It guarantees good quenching properties and the saturation

of the drift velocity. The efficiency of a single cell is ∼99.8% with a spatial resolution of ∼180 μm.

Cathode Strip Chambers. Cathode Strip Chambers (CSC) have been installed in the endcap region $0.8 < |\eta| < 2.4$, where the residual magnetic field between the plates of the return yoke is intense as seen in Fig. 2.15 and the particle rate is high. They are multi-wire proportional chambers consisting of arrays of positively charged anode wires crossed with negatively charged copper cathode strips within a gas volume. When muons pass through, they knock electrons from the gas atoms, which flock to the anode wires creating an avalanche of electrons. Positive ions move away from the wire and towards the copper cathode, also inducing a charge pulse in the strips. By interpolating among the strips one reaches a very fine spatial resolution of 50 μm. In addition to providing precise space and time information, the closely spaced wires make the CSCs fast detectors suitable for triggering.

Resistive Plate Chambers. Resistive Plate Chambers (RPC) are located both in the barrel and in the endcaps. Their spatial resolution is limited, but their time resolution is very good, about 1 ns. Therefore RPC detectors are used to identify unambiguously a bunch crossing of LHC and to provide prompt trigger decisions. RPCs are made of two parallel plates of bakelite, a high-sensitivity plastic material, with a few mm thick gas gap in between them and a graphite coat outside of them. Aluminum strips, separated from the graphite layers by an insulating Polyethylene Terephtalate (PET) film, read out the signals. The gas mixture filling the gap consists of 95% of $C_2H_2F_4$ and 5% of i-C_4H_{10}. When a muon passes through the chamber, electrons are knocked out of gas atoms. These electrons in turn hit other atoms causing an avalanche of electrons. The electrodes are transparent to the signal, which are instead picked up by an external metallic strips after a small but precise time delay. The pattern of hit strips gives a quick measurement of the muon momentum, which is then used by the trigger to make immediate decisions to keep or discard the event.

Trigger system

When the CMS is performing at its peak, collisions happen every 25 ns so it is impossible to store the detector readout for every collision. In these extreme conditions, an efficient trigger system had to be constructed to perform the event rate reduction from 40 MHz in an optimal way. It should be able to discard low-energy processes while keeping as many interesting high-energy events as possible. This is achieved by splitting the workflow in three steps or levels. The Level 1 (L1) step is totally hardware based, whereas Level 2 (L2) and Level 3 (L3) are sets of software requirements and are usually referred to as High Level Trigger (HLT).

The Level 1 trigger. The L1 trigger consists of mostly custom-designed, programmable hardware capable of bringing down the event rate from the initial 40 MHz to 100 kHz. The full data content is stored in pipelines that can retain and process information from many interactions at the same time. To not confuse particles from two different events, the detectors must have a very good time resolution and the signals from the millions

of electronic channels must be synchronized so that they can all be identified as being from the same event. The maximum allowed latency is about 4 μs, after which if the L1 accepts the event, the data starts being processed by the HLT. Since it would not be possible to read out and analyze the whole information contained in an event, only calorimeters and muon chambers are involved in the L1 step, as illustrated in Fig. 2.17. The calorimeter trigger reconstructs electromagnetic and hadronic clusters forming $e\gamma$ or EG candidates, while the muon trigger is responsible for reconstructing muon candidates. The two trigger flows are then combined for a more sophisticated analysis of the event. The L1 ultimately decides whether to pass the event to the second trigger layer or whether to discard it. Only events that satisfy the requirements of at least one of the L1 *seeds* that form the L1 *trigger menu* are retained. For example, in 2016 around 200 L1 seeds were used out of the 512 allowed by the L1 trigger logic. Each L1 seed specifies a list of requirements that need to be satisfied in order to pass the event to the next trigger level. The readout and electronics of the L1 trigger were significantly improved during the LS1 and following the 2015 run, allowing for more sophisticated algorithms to be run, improving the position and energy resolution for jets and EG candidates in particular.

The High Level trigger. The HLT is a software system running on a farm of commercial processors, designed to reduce the event rate down to the final output of 1 kHz. The basic event building strategy is to reconstruct those parts of each physics object that can be used for selection while minimizing the overall processor usage. In order to pass HLT an event needs to satisfy the requirements of at least one of its *paths*, which are defined in the HLT menu, similar to the L1 trigger. Each trigger path targets a certain event topology, for example the presence of two isolated electrons. The trigger path defines a sequence of modules which are run sequentially. The sequence is organized such that computationally expensive modules are run last in order to speed up the overall execution. The HLT software was upgraded in order to cope with the larger event rates at higher luminosity and increased pileup of the Run II. This was achieved by porting some of the advanced offline reconstruction algorithms to the HLT, in particular the particle flow reconstruction and the associated particle identification and isolation algorithms [64]. If an event satisfies one of the trigger paths it is marked for permanent storage and transferred to CERN Tier0 (T0) in one or more *data streams*. Data streams gather similar trigger paths that are commonly used by the offline analysis.

2.4 Physics objects reconstruction

There is one big step left in the path from particles colliding in the centre of the CMS detector to having information about all particles in the event and their properties like charge, momentum and energy. This step is called reconstruction, and as one can imagine it is an extremely challenging task. We have seen in the previous section how each sub-detector is designed and how it is able to detect different kinds of particles. However, particles only leave some kind of electric signal, so in the end we are left with a lot of *hits* in the tracker and muon chambers, and a large amount of energy deposits coming from many different sectors in ECAL and HCAL. We have seen how some of this infor-

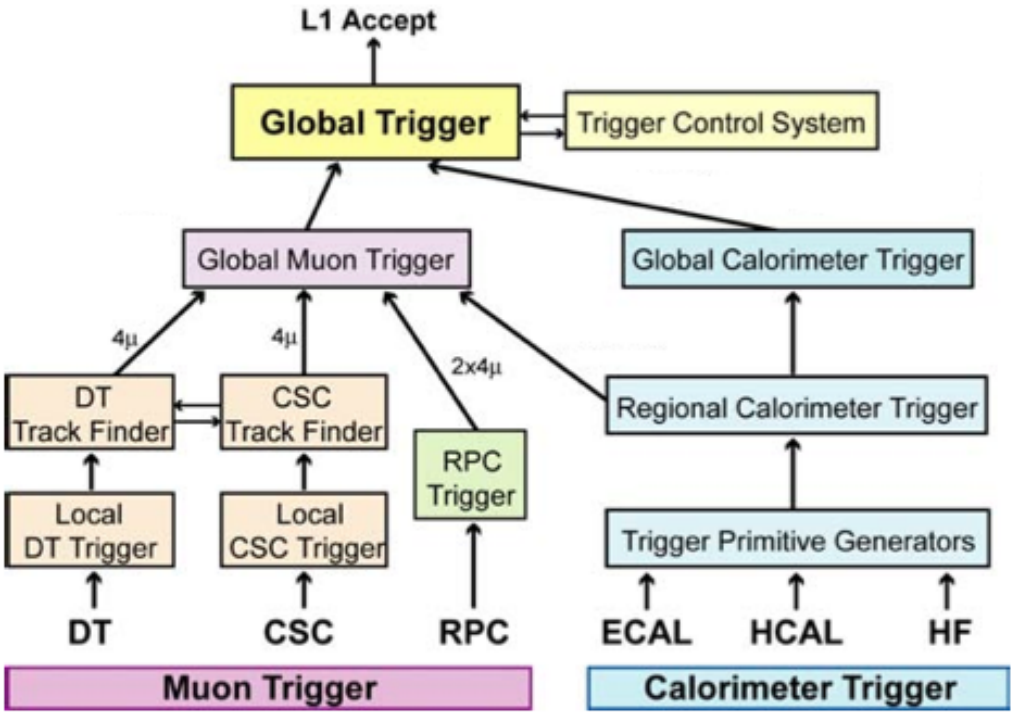

Figure 2.17 – Schematic representation of the CMS L1 trigger system.

mation can be analyzed very fast in order to decide whether the event is worth storing or not. Now, we will describe how this big pile of electronic signals gets reconstructed to physical properties of particles. This step is done offline, which allows more complicated algorithms because of relaxed CPU time limit.

Excellent properties of the CMS detector featuring a highly segmented tracker, a fine-grained ECAL, a hermetic HCAL, a strong magnetic field, and an excellent muon spectrometer are all well suited for particle-flow (PF) reconstruction [65]. For each collision, the comprehensive list of final-state particles identified and reconstructed by the PF algorithm provides a global event description that leads to unprecedented CMS performance for jet and hadronic τ decay reconstruction, missing transverse momentum, and electron and muon identification. This approach also allows particles from pileup interactions to be identified and enables efficient pileup mitigation methods. We will describe advanced algorithms specifically set up for the reconstruction of the basic PF elements. These are in turn used by a *link algorithm* that connects all the PF elements from different subdetectors to produce all stable final state particles as output.

2.4.1 Tracking

Tracking was originally aimed at measuring energetic particles and was limited to well-measured tracks [66]. A combinatorial track finder based on Kalman Filtering (KF) [67] was used to reconstruct these tracks in three stages: initial seed generation with a few

hits compatible with charged-particle trajectory, trajectory building or pattern recognition to gather hits from all tracker layers along this charged-particle trajectory, and final fitting to determine charged-particle properties like origin, transverse momentum, and direction. To be kept for further analysis, the tracks had to be seeded with two hits in consecutive layers, were required to be reconstructed with at least eight hits in total and with at most one missing hit. In addition, all tracks were required to originate from a cylinder of a few mm radius centered around the beam axis and to have p_{T} larger than 0.9 GeV. This yielded in a very low misreconstruction rate but at a cost of reduced track reconstruction efficiency because of the tight selection criteria. It is clear that tracking inefficiency can be substantially reduced by accepting tracks with a smaller p_{T} and with fewer hits. This large improvement, however, comes at the expense of an exponential increase of the combinatorial rate of misreconstructed tracks [68].

Increasing the track reconstruction efficiency while keeping the misreconstructed rate unchanged is therefore critical for PF event reconstruction. To achieve this, the combinatorial track finder was applied in several successive iterations [69], each with moderate efficiency but with as high a purity as possible. At each step, the reduction of the misreconstruction rate is accomplished with quality criteria on the track seeds, on the track fit χ^2, and on the track compatibility with originating from one of the reconstructed primary vertices. The hits associated with the selected tracks are masked, and the remaining hits may be used in the next iteration to form new seeds and tracks with relaxed quality criteria. This process is called *iterative tracking* and is done in ten steps which get more complex and time consuming.

The tracks from the first three iterations are seeded with triplets of pixel hits, with additional criteria on their distance of closest approach to the beam axis. The fourth and fifth iterations aim at recovering tracks with one or two missing hits in the pixel detector. They address mostly detector inefficiencies, but also particle interactions and decays within the pixel detector. The next two iterations are designed to reconstruct very displaced tracks. They also account for nuclear interactions in the tracker material that lead to either a kink in the original hadron trajectory, or to the production of a number of secondary particles where on average two thirds are charged. The eighth iteration addresses specifically the dense core of high-p_{T} jets. In these jets, hits from nearby tracks may merge and be associated with only one or no tracks causing the tracking efficiency to severely decrease. The last two iterations are designed to increase the muon tracking reconstruction efficiency with the use of the muon detector information in the seeding setup.

2.4.2 Clustering

The purpose of the clustering algorithm in the calorimeters is to:

- detect and measure the energy and direction of stable neutral particles,
- separate neutral particles from charged hadron energy deposits,
- reconstruct and identify electrons with all accompanying bremsstrahlung photons,

Table 2.1 – Clustering parameters for ECAL, the HCAL, and the preshower.

	ECAL		HCAL		Preshower
	barrel	endcap	barrel	endcap	
Cell E treshold (MeV)	80	300	800	800	0.06
Seed # closest cells	8	8	4	4	8
Seed E treshold (MeV)	230	600	800	1100	0.12
Seed E_T treshold (MeV)	0	150	0	0	0
Gaussian width (cm)	1.5	1.5	10.0	10.0	0.2

- help the energy measurement of charged hadrons.

This is achieved by a specific clustering algorithm developed for the PF event reconstruction. It performs the clustering independently in ECAL and HCAL, and separately for barrel and endcap while no clustering is done for HF where the energy deposits are directly transformed into clusters. All parameters used in the clustering algorithm were optimized in simulation and their values are reported in Table 2.1.

First, *cluster seeds* are identified as cells with an energy larger than a given seed threshold, and larger than the energy of the neighbouring cells. The cells considered as neighbours are either the four closest cells, which share a side with the seed candidate, or the eight closest cells, including cells that only share a corner with the seed candidate. Second, *topological clusters* are grown from the seeds by aggregating cells with at least a corner in common with a cell already in the cluster and with an energy in excess of a cell threshold set to twice the noise level. In the ECAL endcaps seeds are additionally required to satisfy a threshold requirement on E_T.

An expectation-maximization algorithm based on a Gaussian-mixture model is then used to reconstruct the clusters within a topological cluster. The Gaussian-mixture model postulates that the energy deposits in the M individual cells of the topological cluster arise from N Gaussian energy deposits, where N is the number of seeds. The parameters of the model are the amplitude and the coordinates in the (η, ϕ) plane of each Gaussian while the width σ is fixed to different values for each calorimeter. The fitting of the model is performed via an iterative expectation-maximizing algorithm. At the start of each iteration, the expected energy fraction of a cell in the total energy of a Gaussian is calculated, followed by a maximum-likelihood fit to estimate the model parameters. The initial values for their Gaussian parameters are taken from the cells and the algorithm is repeated until convergence is achieved. The fitted parameters of the model are then used to define *PF clusters*. The PF clusters are then calibrated to compensate the bias in the energy coming from the finite cell thresholds during topological clustering and the energy loss in the dead material between ECAL and HCAL.

One of the goals of the clustering step is to aggregate all energy deposits of a particle. In order to collect the energy of electrons and converted photons an additional clustering step is needed resulting in *superclusters* (SCs). More details on the superclustering step is given in Section. 2.4.4 when electron reconstruction is discussed.

2.4.3 The particle-flow link algorithm

Since a single particle can give rise to several PF elements, the idea is to try and link those elements from various CMS subtedectors that originate from the same particle. This is done by the PF *link algorithm* that is designed to reconstruct particles by connecting PF elements from different subdetectors. In practice this is done by testing the pairs of nearest neighbours in the (η, ϕ) plane. The PF link algorithm then produces *PF blocks* of elements associated either by a direct link or by an indirect link through common elements.

To get a clear picture of how this works in practice we will follow an example illustrated in Fig. 2.18. The track T_1 is linked to the ECAL cluster E_1, and to the HCAL clusters H_1 with a smaller link distance, and H_2 with a larger link distance. On the other hand, the track T_2 is linked only to the HCAL clusters H_2 and H_1. These two tracks form a first PF block with five PF elements: T_1, E_1, and H_1 corresponding to the generated π^-, and T_2 and H_2 corresponding to the generated π^+. The other three ECAL clusters are not linked to any track or cluster and thus form three PF blocks on their own, corresponding to the generated pair of photons from the π^0 decay, and to the neutral kaon.

In each PF block, the identification and reconstruction sequence proceeds in the following order:

1. Muon candidates are identified and reconstructed and corresponding tracks and clusters are removed from the PF block.
2. Electron candidates are identified and reconstructed with the aim of collecting the energy of all bremsstrahlung photons. Energetic and isolated photons are identified in this step as well. Again, corresponding tracks and ECAL or preshower clusters are excluded from further consideration.
3. A track cleaning is performed to reduce the number of fake tracks, removing tracks with large uncertainties.
4. The redundancy of the track and calorimeter measurements are furthermore used to identify muons within jets and fake tracks, both of which can cause the sum of the track momenta to be much smaller than the sum of cluster energies. Muons are selected from the global muon collection with relaxed quality requirements and their tracks removed from the PF block. If the reduced track momentum is still larger than the sum of cluster energies, fake tracks are selected and discarded from the block by ordering all tracks according to their p_{T} and uncertainty σ_{p_T} and removing those with $\sigma_{p_T} > 1$ GeV until the p_{T}-sum of the remaining tracks would be smaller than the energy sum. This cleaning procedure affects only 0.3‰ of tracks in the multijet events.
5. Charged hadron candidates are created for each of the remaining tracks in the PF block, with their momenta set equal to the track momenta. In cases where the sum of track momenta is compatible with the sum of cluster energies within the measured uncertainties, the hadron momenta are redefined to the result of a global fit to the tracks and clusters.

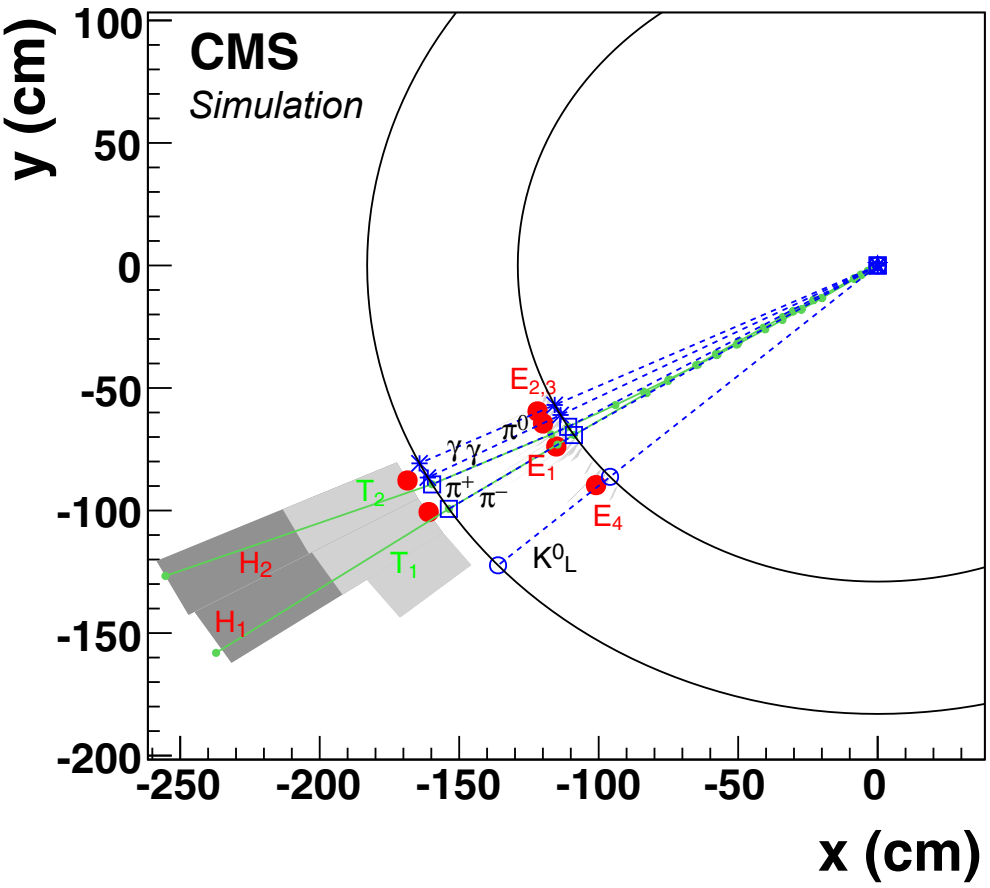

Figure 2.18 – Event display of an illustrative jet made of five particles shown in the (x, y) plane. ECAL and HCAL surfaces are represented as circles centred around the interaction point. The K^0_L, the π^-, and the two photons from the π^0 decay are detected as four well-separated ECAL clusters denoted $E_{1,2,3,4}$. The π^+ does not create a cluster in the ECAL. The two charged pions are reconstructed as charged-particle tracks $T_{1,2}$ appearing as circular arcs. These tracks point towards two HCAL clusters $H_{1,2}$. The cluster positions are represented by dots, the simulated particles by dashed lines, and the positions of their impacts on the calorimeter surfaces by various open markers.

6. If the sum of the cluster energies is larger than the sum of track momenta, the excess is used to create PF photons and neutral hadrons. In cases where the excess is smaller than or equal to the total ECAL energy, the excess is interpreted as a PF photon. In the remaining cases, the ECAL energy is interpreted as a PF photon and the remaining excess as a PF neutral hadron.

7. Clusters that are not linked to tracks are used to create PF photons and neutral hadrons. Within the tracker acceptance $|\eta| < 2.5$, all ECAL and HCAL clusters give rise to photon or neutral hadrons. Outside the tracker acceptance, all ECAL clusters linked to HCAL clusters are interpreted as neutral hadrons, while those not linked to HCAL clusters are interpreted as photons.

The output of the particle-flow link algorithm is a list of mutually exclusive PF candidates, which are then used for further processing like jet reconstruction, particle-flow isolation, or the calculation of missing transverse energy.

Now that we know the global principles of PF event reconstruction in the CMS detector we will describe in detail the process of reconstructing electrons and muons which are essential in the $H \to ZZ^{\star} \to 4\ell$ analysis. Jet reconstruction from PF candidates will also be discussed, since they are used in the $H \to ZZ^{\star} \to 4\ell$ analysis to target specific production mechanisms.

2.4.4 Electrons

With global event reconstruction picture in place we will now discuss how electrons are reconstructed in CMS [70, 71]. The main challenge here is to account for all of the radiated bremsstrahlung photons in the electron trajectory. Because of a significant amount of material budget in the pixel detector as one can see on Fig. 2.9 electrons lose up to 86% of their energy depending on their pseudorapidity. Even though material budget was significantly reduced in the upgrade of the pixel detector in 2017, electrons still lose significant amount of energy this way and accounting for it remains a main challenge of electron reconstruction.

As already mentioned in Section. 2.4.1, the tracks in CMS are reconstructed with a KF approach. This is not suitable for electrons where the dominant effect is the highly non-Gaussian Bremsstrahlung emission. Indeed, in the case of energetic Bremsstrahlung photon emission, causing a kink in the electron trajectory, the pattern recognition is often not able to follow the electron path. In contrast, when the photon energy is moderate, the pattern recognition can succeed in collecting all the hits, but the quality χ^2 of the track fit can be poor. A further challenge arises from bremsstrahlung photons that undergo electron-positron pair production, leading to a complex shower pattern of potentially very short tracks and missing energy.

Indeed, a dedicated tracking algorithm has been developed to cope with these challenges and to improve the efficiency of electron track finding and the accuracy of the parameter estimation.

Tracking for electrons

The Gaussian Sum Filter (GSF) tracking relies on a KF-based pattern recognition with relaxed criteria, allowing the trajectory to be reconstructed often up to the last tracker layer with a specific GSF fit performed in the end. In this approach, the Bremsstrahlung energy loss is modeled by a Gaussian mixture. The large number of components in the GSF track fit allows the sudden curvature radius change, caused by the Bremsstrahlung photon emission, to be properly taken into account. The GSF track thus provides more precise estimates of the inner and outer momentum than the KF. The price to pay for this large number of components is the CPU time consumption, of the order of a few

hundred milliseconds per track. The GSF track reconstruction can thus be run only on a limited number of seeds that represent tracks that are likely to originate from electrons. Two complementary seeding algorithms are used to construct electron seeds:

- **ECAL-driven seeding**. Energetic ECAL superclusters with $E_T > 4$ GeV are used to infer the position of the hits expected in the tracker layers under the assumptions that the cluster is produced by an electron or a positron. Lower energies are not considered in order to save the CPU-time, because tracker-driven seeding proved to be better suited for low-p_{T} electrons. The position and energy of the selected superclusters are used to construct two trajectories which are then propagated from the ECAL surface to the innermost layers of the tracker. Tracker seeds are selected if they are compatible with either trajectory and electron seeds are formed if pairs or triplets of hits are matched. The size of geometric matching window between the extrapolated trajectory and a hit as well as the minimum number of matched hits required to form a seed depends on the tracker subdetector. These parameters are optimized to decrease the rate of fake electrons for chosen high efficiency of finding seeds for true electrons.

- **Tracker-driven seeding**. All tracks found with iterative tracking with $p_{\mathrm{T}} > 2$ GeV are used to search for matching clusters in ECAL. Tracks satisfying pre-identification selection, derived to reduce the rate of fake electrons, are propagated to the ECAL surface and matched to the closest ECAL cluster. The tracker-cluster pair is used to define an electron seed only if the ratio of the cluster energy to the track momentum is compatible with unity and if the extrapolation of the track to the ECAL surface and the cluster position are within a $\Delta\eta$ and $\Delta\phi$ window. Tracks of poor quality pose a challenge and are treated separately. Tracks that have a sufficient number of hits but a large χ^2 are refit using a light version of the GSF fit with a reduced number of components in the energy loss modeling. An example of such a track can come from electrons with small successive energy losses. The final decision to consider the track seed of a refitted track as an electron seed is based on a pre-identification boosted decision tree (BDT). It exploits the track quality parameters of the KF and light GSF fit, together with $\Delta\eta$ and $\Delta\phi$ between the cluster and refitted track.

The seeds obtained with the tracker-driven and ECAL-driven seeding are merged in order to provide a unique collection of seeds, while keeping memory of their origin. For the tracker-driven seeds the reference to the KF track, out of which the seed has been obtained, is kept. Similarly the reference to the standard supercluster is saved for the ECAL-driven seeds. When an ECAL-driven seed, with N hits, and a tracker-driven seed share at least N-1 tracker hits, only the former is kept and the provenance is assigned to both algorithms. The significant increase of seeding efficiency brought by the tracker-driven seeding approach is shown in Fig. 2.19.

This collection of seeds is used to initiate electron track reconstruction. Track building proceeds iteratively from the track parameters provided in each layer, modelling the electron loss with a Bethe-Heitler function [72]. To maintain good efficiency in

the presence of bremsstrahlung, compatibility requirements between the predicted and found position of hits in each layer are loosened up. If several hits are compatible with the predicted one, then several trajectory candidates are created and developed, with a limit of five candidate trajectories for each layer. At most one missing hit is allowed per trajectory. Once all hits are collected, track parameters are estimated by a fit that uses Gaussian Sum Filter. The energy loss in each layer is approximated by a mixture of Gaussian distributions, each of which is attributed a weight that describes its associated probability. The parameters of electron tracks are estimated from the most probable value, the mode, of all track components because it was found that a weighted mean of all components introduced a slight bias in the momentum value yielding a worse resolution. End products are *GSF tracks* which are combined with ECAL superclusters to build electron candidates.

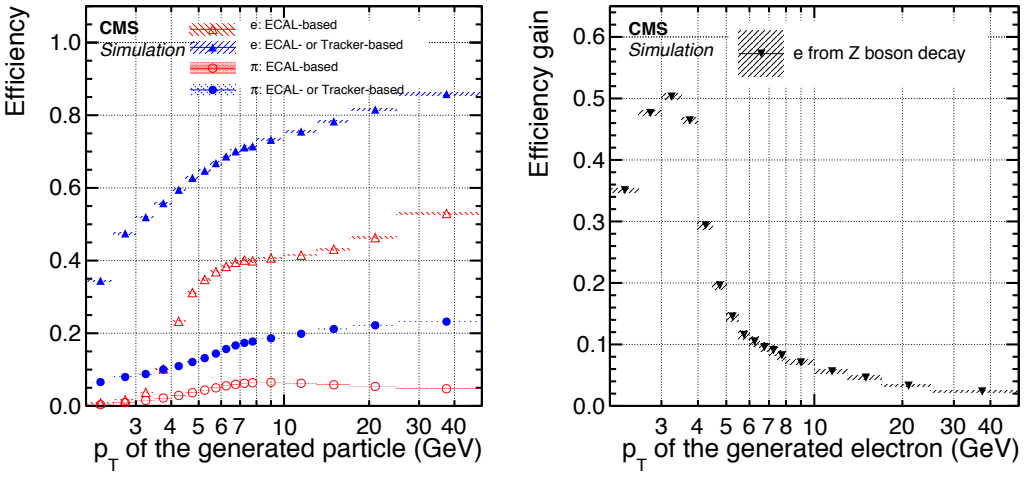

Figure 2.19 – Electron seeding efficiency for electrons (triangles) and pions (circles) as a function of p_{T}, from a simulated event sample enriched in b quark jets with p_{T} between 80 and 170 GeV, and with at least one semileptonic b hadron decay. Both the efficiencies for ECAL-driven seeding only (hollow symbols) and with the tracker-driven seeding added (solid symbols) are displayed (left). Absolute efficiency gain from the tracker-based seeding for electrons from Z boson decays as a function of p_{T} (right). The shaded bands indicate the pT bin size and the statistical uncertainties on the efficiency.

ECAL superclusters

ECAL superclusters are built from ECAL PF clusters described in Section. 2.4.2. In order to collect the energy of electrons and converted photons, which can exhibit a large spread in ϕ due to bremsstrahlung an additional step of clustering algorithm was developed.

In Run I, superclustering was done in a fixed $\Delta\eta - \Delta\phi$-rectangular region around a seed crystal. The rectangular region had to be sufficiently large to capture bremsstrahlung far from the primary electron. A large clustering region, however, poses problems in the presence of energy deposits from pileup interactions close to the electromagnetic cluster,

introducing a bias in the determination of the SC energy and cluster shapes. In order to cope with this for the conditions of increased pileup in Run II, the superclustering was improved to avoid the use of large $\Delta\phi$ region for high-E_T deposits, and also to accommodate for the separation in η for very low-E_T clusters.

Electron reconstruction

Final step in the electron reconstruction is combining GSF tracks and ECAL superclusters. Although the superclustering algorithm is designed to collect the energy deposits from bremsstrahlung it is not perfect in doing so. These inefficiencies arising mainly from situations when primary and bremsstrahlung clusters are very far apart spatially or in case of converted bremsstrahlung photons. An attempt is made to collect these deposits from the unconverted and converted bremsstrahlung photons respectively. For the former case, a tangent to the GSF track is build at each tracker layer, and any ECAL cluster compatible with the tangent is PF-linked to the supercluster as illustrated in Fig. 2.20.

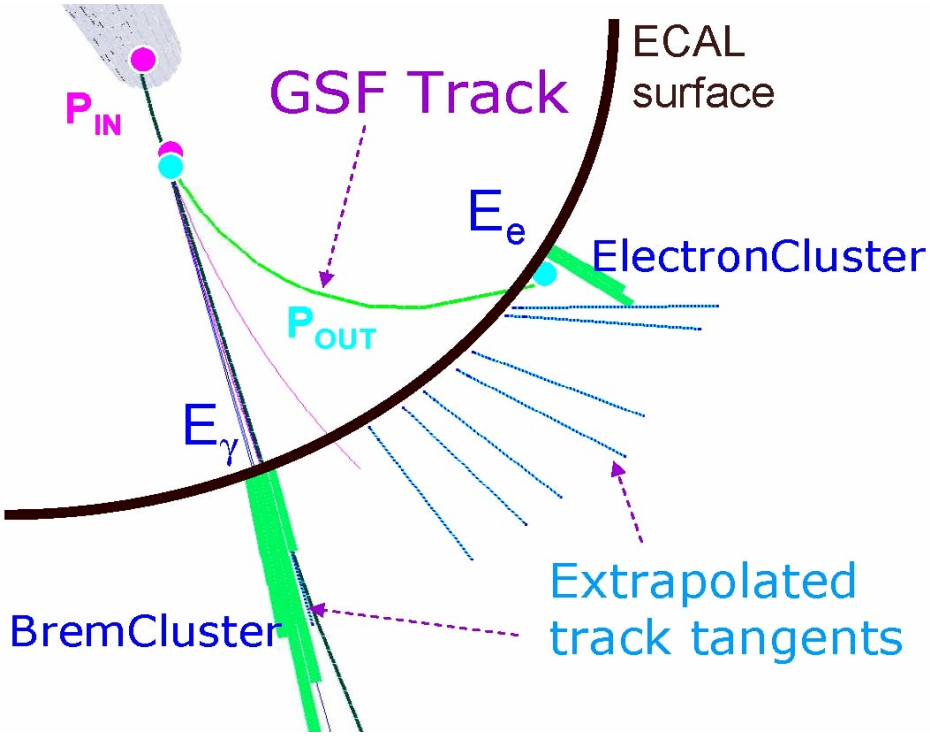

Figure 2.20 – Illustration of an electron undergoing bremsstrahlung emission and the components of the electron reconstruction implemented in the PF algorithm. The initial electron (magenta line) emits a bremsstrahlung photon (gray line), giving rise to two distinctive electromagnetic clusters in the ECAL (green bars). The GSF tracking accommodates the change of the curvature of the electron track (green line) and allows to measure the incoming (p_{in}) and outgoing momenta (p_{out}). Finally, the cluster of the bremsstrahlung photon is linked to the electron cluster via the GSF track tangents, giving the refined supercluster.

The procedure with GSF track tangents is also used to identify cases where the bremsstrahlung photon undergoes conversion in the tracker material. Two algorithms are developed for this purpose, targeting cases where one or both tracks of the conversion electrons are reconstructed by the iterative tracking algorithm. A dedicated conversion finder attempts to find track pairs from conversion vertices. If the direction of the converted photon calculated from the sum of the two conversion tracks is compatible with the one of the GSF tangents, the tracks are PF-linked to the GSF track. The conversion tracks are then used to search for ECAL clusters, which are PF-linked to the supercluster if the ratio of the cluster energy and conversion track momentum is compatible with the electron hypothesis. The single-leg conversion finder targets cases of asymmetric conversions with one very low-p_{T} electrons or cases where one of the electrons from conversion leaves only few hits making it hard to reconstruct its track. In this algorithm the closest KF track in ΔR is identified for each GSF tangent. Tracks passing a preselection on $\Delta\eta$ and $\Delta\phi$ are then submitted to a multivariate discriminant to distinguish conversions from different backgrounds. In addition to the distance of the track to the GSF track, the conversion multivariate discriminant also exploits information about: the radius of the innermost hit on the KF track, the transverse impact parameter with respect to the primary vertex, and the E/p of the KF track ant the ECAL clusters linked to it. Selected KF tracks and their associated clusters are again PF-linked to the GSF track and the supercluster respectively. In this way, a *refined supercluster* is ultimately defined based on the merger of the supercluster the ECAL clusters PF-linked to it.

Next in line is the charge estimation step. One can be tricked into thinking that this is a very trivial step because charge can be easily deducted from the sign of the curvature of the GSF track. However, again due to bremsstrahlung, one can understand how this can become non trivial. Let's imagine for instance a situation where an early bremsstrahlung is followed by a photon conversion. Here, the contribution from conversion can be mistakenly included in track fitting leading in a wrong charge estimate. Because of this two other charge estimates are used to compliment the simple estimation. First one exploits the charge of the KF track associated with the GSF track, if there is any. The second one looks at the difference in ϕ between the vector joining the beam spot to the supercluster position and the vector joining the beam spot to the first hit of the electron GSF track. In the end, the electron charge is chosen as the sign obtained by majority of the three estimates.

The final step is the estimation of the electron momentum. It relies on a combination of the energy of the supercluster and the momentum estimate of the PF-linked GSF track. This is performed in several steps:

- Electrons are first classified into five categories: golden, big brem, showering, crack and bad tracks. These categories are designed to collect electrons with similar properties in order to improve overall momentum resolution. First three categories are associated to different bremsstrahlung patterns. Crack electrons are those that are located near the cracks in the ECAL. Finally, bad track electrons are those with a poorly fitted track.
- The supercluster energy is obtained simply by summing all crystal energies plus the

preshower energies in the endcaps. However, it was shown that effects like energy leakage out of the superclusters, in the ECAL cracks or into the HCAL, energy loss in the tracker material, and energy from pileup severely affect this measurement. In order to correct for these effects a multivariate regression is used exploiting a large number of variables related to all these effects [70]. This was shown to bring an improvement of 20% on average, while also reducing the bias in the peak position. A similar procedure is used to calculate the uncertainty on the corrected energy estimate. Additional calibrations are further applied to correct for various imperfect descriptions of the detector in the simulation and will be discussed in detail in Section 3.1.

- The final estimate of the electron momentum combines the corrected and calibrated supercluster energy E^{SC} with the GSF track momentum p. The latter is expected to be the more precise one below 15 GeV, and for electrons near cracks. The combination again uses a multivariate regression, whereby the main input variables are E^{SC}, p, their relative uncertainties, the electron category, electron position in the barrel or endcaps.

The output of this procedure is a list of electron candidates with their corresponding charge and momentum estimates that are used as inputs to various analysis. Further selection can be made to extract a wanted signal and details of this procedure for the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis will be discussed in Chapter. 3.

2.4.5 Muons

The reconstruction of muons in the CMS detector [63] is much less challenging, as one would expect. Firstly, the probability for a muon to undergo a bremsstrahlung emission is $\sim 10^{10}$ smaller than for electron. Furthermore, other than neutrinos that leave no trace in the CMS detector, muons are the only particles that traverse the inner part of the detector and reach the muon chambers. And finally, as we will see, combining inner tracker information with the hits from the muon system provides high efficiency and good resolution for momentum measurement.

Tracking for muons

Main advantage when tracking muons over other particles is the use of the muon system. Three different types of muon candidates are reconstructed with the PF algorithm:

- **Standalone muons** are reconstructed using only the muon system described in Section. 2.3.3. Hits within each DT or CSC detector are clustered to form track segments. Track segments are then used as seeds for track reconstruction that adds the information from the RPC. The final output are *standalone muon tracks*.
- **Global muons.** If the standalone muon track can be geometrically matched to a track originating from an inner tracker, then the hits from the *inner track* and from the standalone muon track are combined and refitted to form a *global muon track*.

- **Tracker muons.** If at least one muon segment is geometrically matched to an inner track with $p_{\mathrm{T}} > 0.5$ GeV and a total momentum $p > 2$ GeV the inner track is considered a *tracker muon track*.

About 99% of muons produced within the geometrical acceptance of the muon system are reconstructed either as global muon or a tracker muon and very often as both. Global muons improve the momentum resolution with respect to the tracker-only fit, and tracker muons recover efficiency for very low-p_{T} muons that do not always fully traverse the CMS detector. Global muons and tracker muons that share the same inner track are merged into a single candidate. Muons reconstructed only as standalone-muon tracks have worse momentum resolution and a higher admixture of cosmic muons than global and tracker muons.

Muon reconstruction

The charge and momentum of the PF muon are extracted from the fit to the tracker muon track if its p_{T} is smaller than 200 GeV. Above this value, the momentum is chosen according to the smallest χ^2 probability from the different track fits: tracker only, tracker and first muon detector plane, global and global without the muon detector planes featuring a high occupancy.

The PF elements that make up muon candidates are masked against further processing in the corresponding PF block. Similarly to electrons, even though efficiency for reconstructing real muons with this procedure is very high, other particles get misreconstructed as muons and sneak in as a form of background. In order to cope with this kind of background, additional selection can be applied to the muon candidates as well.

2.4.6 Jets

One of the main features of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ channel is having only four well isolated leptons in the final state. However, if one wishes to study different production mechanisms of the Higgs boson jets become as an important final state particles as leptons. Being one of the more complex physics objects, just a brief overview of the jet reconstruction [73, 74] is presented here.

A jet is a narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon, meaning it has a large number of final state particles detected mainly in HCAL. In the view of PF candidates they mostly consist of charged and neutral hadrons, but can also have electrons, muons and photons. The goal of combining all of these particles in a jet is to provide a mean to detect and measure the kinematics of the initial quark or a gluon. The reconstruction of jets is done with jet clustering algorithms [75]. The most common jet reconstruction algorithm in the LHC is the anti-k_T algorithm [76], which is inspired by a reversal of the fragmentation process. This algorithm tends to cluster jets around the hardest particles, resulting in a conic jet shapes. It also infrared and collinear safe, meaning that the result od the algorithm is unaltered

under additional soft gluon emissions or parton splitting. Jet reconstruction in CMS is based on the FASTJET [77] package and uses PF candidates as inputs to the clustering. Several jet collections corresponding to different choice of the cone sizes used in the clustering procedure are produced. In the work of this thesis anti-k_T jets with a cone size of $R = 0.4$ and exploiting the *charged hadron subtraction* (CHS) are used. The goal of CHS is to reduce the pileup dependence by removing all charged hadron PF candidates associated with pileup vertices before the jet clustering. Pileup vertices are defined as all reconstructed vertices, expect the primary vertex.

The momentum of a reconstructed jet is simply computed as the vectorial sum of the momenta of all particles it involves.

Chapter 3

Electron selection and efficiency measurements

In the previous Chapter we have discussed how the CMS experiment was designed and built in order to search for the Higgs boson. We have also described how the CMS detector detects particles and records events, and finally how it reconstructs detector signal to physics object candidates. We are now equipped with a set of events and particles that we can use in our analysis to search for the Higgs boson candidates. However, there is a very important step in between.

As already mentioned, reconstructed particle candidates are contaminated with non-negligible fraction of background. In addition, the momentum measurements are degraded due to many subtle effects that need to be taken into account. Finally, specific analysis will have different requirements for different particles. For example, in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis we require well isolated leptons with low probability of fakes and a very good momentum resolution. We also require high efficiency when selecting very low-p_{T} leptons in order to be able to reconstruct the offshell $\mathrm{Z}^{\star}$. On the other hand, since jets and missing transverse momentum are only used to tag different production mechanisms the requirements on them are not so confined.

In order to be selected in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, leptons must pass a set of identification and isolation requirements specifically designed to extract the signal Higgs boson process while reducing the SM backgrounds as much as possible. Data and MC simulation have to be compared carefully in dedicated control regions in order to understand and correct any differences identified as resulting from detector effects. This is done in several steps. Dedicated time-dependent momentum and energy corrections are derived. On top of that, for each selection step the efficiency is carefully measured in data and in MC simulation, and observed differences are accounted for with the application of scale factors. Failing to account for all of the detector effects could be catastrophic as it could lead to our data showing an excess when there is none, or making it impossible to observe a new effect.

In this chapter a detailed description of this procedure is given for electrons, one of the main particles in the golden channel. Compared to muons, electrons suffer from intrinsically higher background rates, so a more complex selection procedure is needed. My main contributions come from measuring and improving electron selection efficiencies and deriving corresponding scale factors both for the $H \to ZZ^{\star} \to 4\ell$ analysis and for a wider use in the CMS Collaboration. My results were used in the intermediate CMS Public Analysis Summary (PAS) released on the first Run II data collected in 2015 [78].

3.1 Electron momentum calibration

As explained in Section 2.4.4, the determination of the electron momentum relies on a combination from ECAL and tracker. Although the ECAL calibrations and the track reconstruction rely on the best knowledge of the detector conditions, some small discrepancies remain between data and simulation. There are several sources for the effects that harm the electron momentum estimation arising from tracker material modeling, evolution of the crystal transparency, and ECAL pedestal and noise time evolution. These effects are very hard to account for in the simulation, so the observed discrepancies are studied in detail and an additional set of scale and resolution corrections is applied to data and simulation.

The corrections are done by monitoring the measured mass of the Z boson and shifting the electron energy scale such that the corrected mass is equal to the mass in the simulation. This does not imply that the scale is used to shift the Z boson mass to match the experimental value given by the world best average determined by the Particle Data Group (PDG). Instead it is matched to the reconstructed peak position in the simulation which will in general differ from the nominal mass value used to generate the simulation sample due to detector and reconstruction effects.

First, the energy scale of electrons is calibrated in data for different electron categories with the expression:

$$E_{scale} = E_{reco} \cdot scale(Run, p_T^e, |\eta^e|, R_9^e), \tag{3.1}$$

where E_{reco} is the reconstructed mass without any corrections, Run is the run number to account for time evolution of the scale shift, $p_T^e, |\eta^e|$ and R_9^e are electron kinematic observables that proved to be the best minimal choice allowing for a good final calibration. In particular, the R_9 observable is defined as the ratio of the energy in a 3×3 crystal matrix centered around the seed crystal and the supercluster energy, and is shown to be very sensitive to the amount of bremsstrahlung. The scale is derived using the $Z \to e^+e^-$ control sample, where a multidimensional fit in different electron categories is performed in order to minimize the differences between the dielectron mass in data and simulation.

In addition a smearing of the electron energy in the simulation is performed with:

$$E_{smear} = E_{reco} \cdot Gauss(1, \sigma(Run, p_T^e, |\eta^e|, R_9^e)), \tag{3.2}$$

where the electron categories are defined with the same observables as for the scale corrections.

The impact of these corrections in data and simulation can be seen in Fig. 3.1, illustrating a clear improvement in data to simulation comparison after the application of scale and smearing corrections.

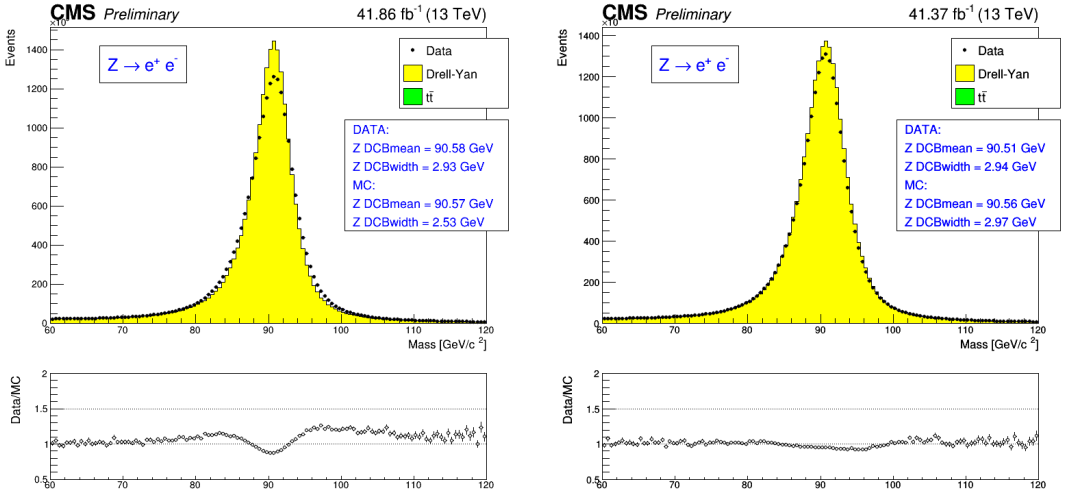

Figure 3.1 – Reconstructed dielectron invariant mass before (left) and after (right) the application of the electron scale and smearing corrections. Stacked histograms represent expected distributions obtained from simulation and black points represent data. The results of the Crystall-ball fit are reported in the figures.

3.2 Electron selection

The efficiency of the electron reconstruction is very high, thus meaning we only miss a very small fraction of real electrons. The main challenge is the presence of a sizable amount of fakes, i.e. other objects that pass the electron reconstruction procedure and are thus considered as electron candidates. In order to deal with these, one has to implement a further set of requirements to reduce the amount of fakes while loosing as few as possible real electrons.

Since in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis at least four electrons have to be selected, the per-lepton efficiency propagates to event selection with the power of four. Therefore, a rather loose lepton selection is chosen in comparison with other analysis using leptons. Together with this main consideration, there are several other important properties in the design of the electron selection. One has to study carefully the dependence of the signal and background efficiencies on the candidate kinematics. As we will see, for example, processes that contribute to the reducible background feature mostly low-p_{T} electrons. Additionally, signal and background efficiency are expected to have no pileup dependence. For every variable used in the selection the quality of the agreement be-

tween data and simulation should be checked with any possible discrepancies understood. Finally, the agreement between the efficiency in data and simulation should be cross checked and validated in control regions with sufficient statistics. Also, modeling and composition of the background at various steps of the selection should be cross checked. With these requirements in mind a selection is performed in three sub-selection steps: kinematic and impact parameter requirements, identification, and isolation.

All selection studies are performed on *signal electrons* and *background* or *fake electrons*. Signal electrons are taken from a gluon fusion $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ sample, and defined as reconstructed electrons that are geometrically matched to generated electrons originating from the Higgs boson decay. On contrary, background or fake electrons are taken from the Drell-Yan sample, and are defined as reconstructed electrons that are geometrically separated from the electrons originating from $\mathrm{Z}^{(*)}$ or $\gamma^{(*)}$ decays. With these two sets of electrons we have exactly the objects that should be selected as candidates in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis and their fakes coming from the largest reducible background contribution to the signal region.

3.2.1 Kinematic and impact parameter requirements

To account for detector acceptance, a cut on the pseudorapidity of the electrons of $|\eta^{e}| < 2.5$ is applied. Furthermore, because of the exponential increase of backgrounds at the very low p_{T} and severe difficulties to reliably reconstruct tracks and measure momentum and efficiencies only electrons with $p_{\mathrm{T}} > 7$ GeV are considered.

The impact parameter selection aims to reduce backgrounds originating from displaced electron candidates. The most important examples are photon conversions and the B meson decay. The main feature that can be used to differentiate these electron candidates from signal electrons is their distance from the primary vertex. This vertex is defined by first associating tracks to reconstructed primary vertices based on vertex fits and proximity criteria, then clustering tracks with a jet algorithm, and finally picking the primary vertex that has the largest sum of squared transverse momenta of the resulting clusters and the per-vertex missing momentum.

The electrons are then required to satisfy the following primary vertex constrains:

$$d_{xy} < 0.5\,\mathrm{cm} \quad \text{and} \quad d_{z} < 1\,\mathrm{cm}, \tag{3.3}$$

where d_{xy} and d_{z} are the absolute values of the electron impact parameter with respect to the primary collision vertex in the transverse plane and in the longitudinal direction, respectively. The *3D impact parameter* $\mathrm{IP}_{\mathrm{3D}}$ between the candidate and the primary vertex is defined as the minimal Euclidean distance between the two. A more robust observable *3D impact parameter significance* is constructed using the tracking uncertainty on the impact parameter $\sigma_{\mathrm{IP}_{\mathrm{3D}}}$, and a selection is made:

$$\mathrm{SIP}_{\mathrm{3D}} \equiv \frac{|\mathrm{IP}_{\mathrm{3D}}|}{\sigma_{\mathrm{IP}_{\mathrm{3D}}}} < 4. \tag{3.4}$$

The distributions of impact parameter variables used in the electron selection are shown

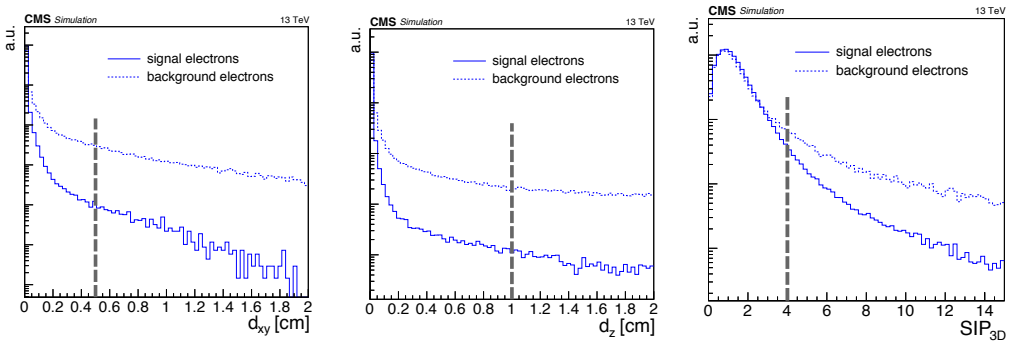

Figure 3.2 – Distribution of the d_{xy} (left), d_z (middle), and 3D impact parameter significance (right) variables for signal and background electrons. The two types of electrons are defined in the text, and the kinematic cuts are applied. The vertical grey dashed lines denote the operating points.

in Fig. 3.2 for signal and background electrons showing a clear separation between them for the chosen cut values.

3.2.2 Identification

With the selection on the impact parameter variables we do get rid of a fraction of background electrons that have significantly displaced vertex. However, we are still left with a sizable contamination of fake electrons, originating mostly from hadronic jets and photon conversions. Hadronic jets can mimic the electron signature with accidental association during reconstruction. For example, the reconstructed track from a charged hadron like a $\pi^{\pm}$ can be in a close vicinity of an electromagnetic cluster of $\pi^0 \to \gamma\gamma$ decays and thus easily mistaken for electron. This is the dominant source of fake electrons in our analysis and *electron identification* (ID) is designed and optimized to distinguish them from the signal electrons.

One could naively think that introducing several cuts on a set of tracking and ECAL related observables should be enough to identify signal electrons. Indeed, this procedure is often referred to as *cut-based identification* and is used in many analysis with electrons in the final state. However, a cut-based ID usually implies a very tight selection that sacrifices signal efficiency in order to reduce the amount of background and it does so mainly in the low-p_{T} region. Because of the need to have a loose electron ID that is able to significantly reduce the amount of fake electrons even in the very low-p_{T} region without sacrificing signal efficiency that enters analysis selection with a power of four, a dedicated electron identification has been developed by the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis. It uses a broad range of observables listed and described in Table. 3.1, combined together into a single multivariate classifier using the boosted decision tree (BDT) techniques.

There are three main classes of electron observables that are used as inputs:

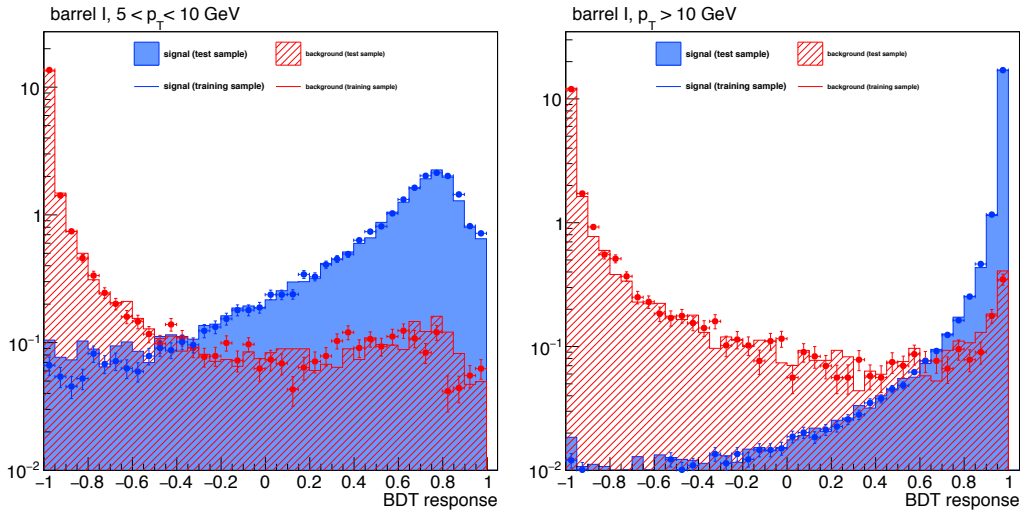

Figure 3.3 – Output of the electron BDT classifier for signal electrons (blue) and background electrons (red) in the two $|\eta| < 0.8$ training bins. The electrons are required to pass kinematic cuts.

- ECAL observables based on the shape of the electromagnetic cluster,
- observables based on tracking information,
- observables that describe the quality of the matching between the supercluster and the track.

The BDT classifier is trained on signal and background electrons both coming from a Drell-Yan sample, with geometrical matching as defined earlier. It was shown that signal electrons from gluon fusion and Drell-Yan samples have almost identical kinematic properties. This is expected since in both cases electrons are coming from $Z \to e^+e^-$ decays. To improve identification performance of low-p_T electrons, the training is done in two p_T bins: $7 < p_T < 10$ GeV and $p_T > 10$ GeV. Since electrons are reconstructed with different subdetectors such as the barrel and endcap parts of the ECAL and tracker, with different material budget distributions, the training is further split into three bins in pseudorapidity: $|\eta| < 0.8$, $0.8 < |\eta| < 1.479$, and $|\eta| > 1.479$. In this way, six training categories are defined, and it was shown that they provide better performance than directly including electron p_T and $|\eta|$ as variables. An example of the BDT classifier output for one training bin can be seen in Fig. 3.3, showing a powerful discriminating power achieved. The final step is to decide on the cut values on the BDT classifier score, or in other words to choose the *working points*. It was decided to go for working points such that the global signal electron efficiency for the $H \to ZZ^\star \to 4\ell$ analysis of ~98% remains the same as in Run I.

Further improvements of the electron ID for the analysis of the 2017 data were done with the introduction of the isolation variables and will be discussed shortly in Section 3.2.4. Results of the electron identification in terms of ROC curves and working points in different categories will be presented in Section 4.2.2.

Table 3.1 – List of input variables with their definition for the electron identification BDT classifier.

observable class	observable	definition
pure cluster shape observables	$\sigma_{i\eta i\eta}$	energy-weighted standard deviation of single crystal η within a 5×5 block of crystals centred on the highest energy crystal of the cluster seed
	$\sigma_{i\phi i\phi}$	similar as $\sigma_{i\eta i\eta}$, but in the ϕ direction
	η width	supercluster width along η
	ϕ width	supercluster width along ϕ
	$1 - E_{5\times1}/E_{5\times5}$	$E_{5\times5}$ is the energy computed in the 5×5 block of crystals centered on the highest energy crystal of the cluster seed, and $E_{1\times5}$ is the energy computed in the strip of crystals containing it
	R_9	energy sum of the 3×3 block of crystals centred on the highest energy crystal, divided by the supercluster energy
	H/E	energy collected by the HCAL towers within a cone of $\Delta R = 0.15$ centred on the supercluster position, divided by the supercluster energy
	E_{PS}/E_{raw}	energy fraction deposited in the preshower subdetector, divided by the untransformed supercluster energy
pure tracking observables	$f_{brem} = 1 - p_{out}/p_{in}$	fractional momentum loss of the GSF track which measures the amount of bremsstrahlung emission
	N_{KF}	number of hits of the KF track
	N_{GSF}	number of hits of the GSF track
	χ^2_{KF}	goodness of fit of the KF track
	χ^2_{GSF}	goodness of fit of the GSF track
	$N_{missinghits}$	number of expected but missing inner hits in the first tracker layers
	$P_{conv.}$	fit probability for a conversion vertex associated with the electron track
track-cluster matching observables	E_{SC}/p_{in}	ratio of the supercluster energy to the track momentum at the innermost track position
	E_e/p_{out}	ratio of the energy of the ECAL cluster closest to the electron track extrapolated to ECAL to the track momentum at the outermost track position
	$\frac{1}{E_{SC}} - \frac{1}{p}$	deviation of the supercluster energy from the electron momentum obtained by combining ECAL and tracker information
	$\Delta\eta_{in} = \lvert\eta_{SC} - \eta_{in}\rvert$	pseudorapidity distance between the energy weighted position of the supercluster and the track position of closest approach to the supercluster, extrapolated from the innermost track position and direction
	$\Delta\phi_{in} = \lvert\phi_{SC} - \phi_{in}\rvert$	similar as $\Delta\eta_{in}$, but in the ϕ direction
	$\Delta\eta_{seed} = \lvert\eta_{seed} - \eta_{out}\rvert$	pseudorapidity distance between the seed cluster position and the electron track extrapolation to the ECAL

Table 3.2 – List of input variables with their definition for the electron isolation. Variables not used in the combination of electron identification and isolation are marked with (*).

observable class	observable	definition
isolation	$\sum_{\text{charged hadrons}} p_T$	p_T sum of all charged hadron reconstructed PF candidates in a cone with $\Delta R = 0.3$ around the electron
	$\sum_{\text{neutral hadrons}} p_T$	p_T sum of all neutral hadron reconstructed PF candidates in a cone with $\Delta R = 0.3$ around the electron
	$\sum_{\text{photons}} p_T$	p_T sum of all photon reconstructed PF candidates in a cone with $\Delta R = 0.3$ around the electron
	ρ	the mean energy density in the event
	A_{eff}(*)	the effective area, defined as the ratio between the slope of the average isolation and that of ρ as a function of the number of reconstructed vertices

3.2.3 Isolation

A very powerful tool in suppressing the dominant fake electron background coming from hadronic jets is *isolation*. The underlying idea is that a signal electron will not be surrounded by other particles coming from the same hard interactions, so finding other PF candidates in a cone around the electron can be used as a sign of a fake electron. Instead of cutting on different isolation related observables, a combination of them is made, and a single cut is applied. A detailed list of isolation variables used is reported in Table 3.2, together with their definitions.

The combined relative isolation is defined as:

$$\mathcal{I}_{\text{PF}} \equiv \frac{1}{p_{\text{T}}} \left(\max \left[0, \sum_{\substack{\text{photons} \\ (\text{exceptFSR})}} p_T + \sum_{\substack{\text{neutral} \\ \text{hadrons}}} p_T - p_T^{\text{PU}} \right] + \sum_{\substack{\text{charged} \\ \text{hadrons}}} p_T \right), \tag{3.5}$$

where photons selected by the $\text{H} \to \text{ZZ}^\star \to 4\ell$ final state radiation (FSR) recovery algorithm which is described in Section are excluded from the sum.

The lack of a reliable vertexing for photons and neutral hadrons makes their contributions highly susceptible to the amount of pileup in the event, so a correction based on the hadronic activity in the event is used to mitigate this pileup dependence. It is based

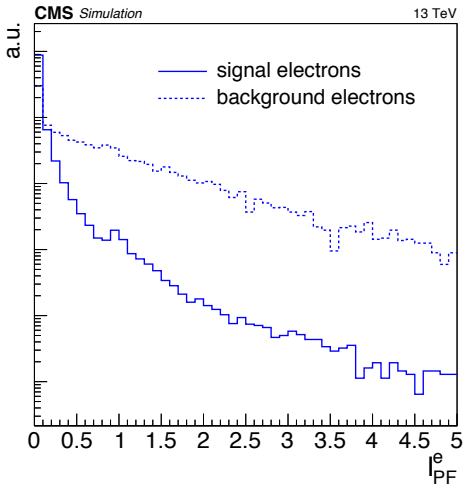

Figure 3.4 – Distribution of the combined relative PF isolation variable $\mathcal{I}_{\mathrm{PF}}$ for signal and background electrons. The electrons are required to be pass kinematic, impact parameter, and identification cuts.

on the effective area technique and is defined as:

$$p_{\mathrm{T}}^{\mathrm{PU}} = \rho \cdot A_{\mathrm{eff}}, \tag{3.6}$$

the effective area is determined in five bins of $|\eta|$. Distribution of the relative isolation for signal and background electrons is shown in Fig. 3.4.

After several careful studies a loose working point was chosen such to have similar high signal efficiency as for the electron identification step. For simplicity, since it provided almost identical signal efficiencies, a single working point of $\mathcal{I}_{\mathrm{PF}} < 0.35$ was chosen for both electrons and muons.

3.2.4 Combining identification and isolation

After Run I was concluded, every step in the electron selection procedure was studied in search for possible improvements. It was still desired to keep a loose selection with high signal efficiency, so a natural place for improvement was trying to reduce the background efficiency, i.e. the number of fake electrons that pass the selection chain.

With this motivation in mind, electron selection was improved for the analysis of the 2016 data. There was not much to gain in the selection on simple impact parameter observables. In the isolation, a shift from a $\Delta R = 0.4$ to $\Delta R = 0.3$ cone when calculating isolation sums brought small improvements. Finally, inclusion of three new tracking related observables an inputs for electron ID BDT proved to be a right direction in further reduction of fake electrons.

For the analysis of the 2017 data, an idea of combining all variables used in the electron selection chain into a single BDT classifier was studied. Impact parameter observables

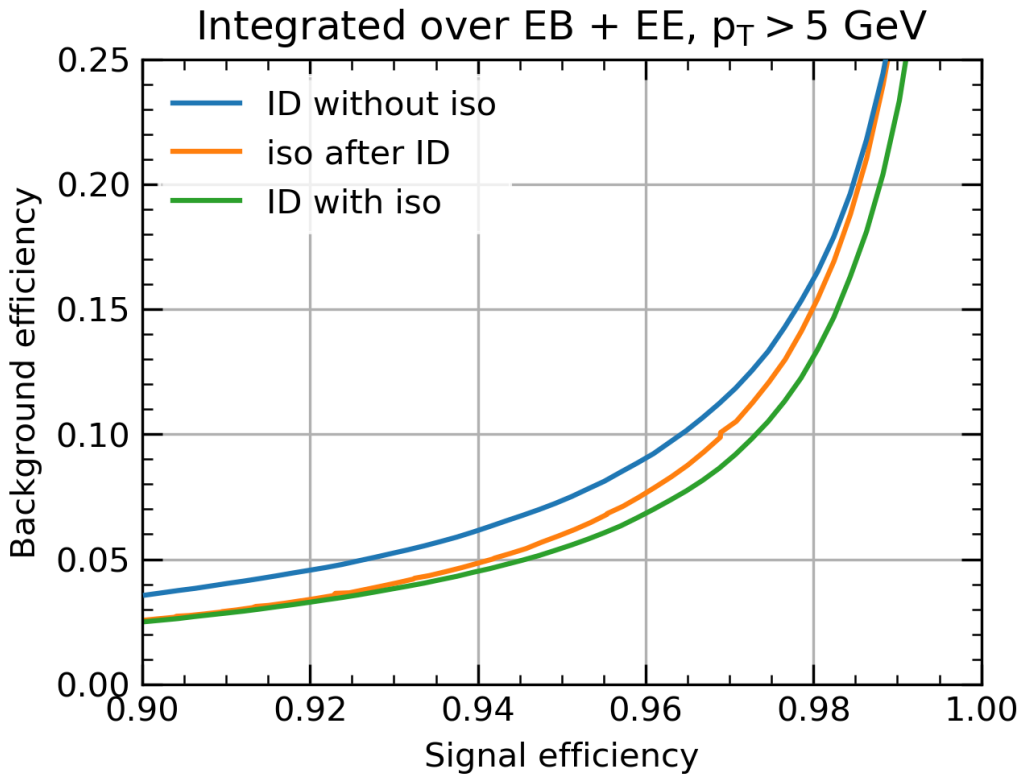

Figure 3.5 – ROC curves of the electron BDT classifier for the 2017 analysis combining identification and isolation (green), electron ID BDT classifier for 2016 analysis combined with a $\mathcal{I}_{\mathrm{PF}} < 0.35$ isolation cut (orange), and electron ID BDT classifier for 2016 analysis without any isolation selection (blue). Signal and background efficiencies are integrated over the full p_{T} and $|\eta|$ range of the analysis and calculated using the Drell-Yan sample.

d_{xy}, d_z, and $\mathrm{SIP}_{3\mathrm{D}}$ were decided to be left out of this combination for a couple of reasons. Firstly, their modeling in simulation is not perfect, thus it is a much safer choice to have a controlled lose cut that does not introduce large data to simulation discrepancies. Furthermore, it is convenient to be able to cut on them separately from further selection steps in order to define control regions that are used for various background estimates and cross checks as it is explained in Section 4.5. However, it was shown that the combination of previously used electron identification observables listed in Table 3.1 with electron isolation variables listed in Table 3.2 can bring significant reduction of fake electrons for the same signal efficiency. It is important to note here that one of the isolation variables, the effective area A_{eff}, is not used in the combination because of its strong correlation with electron pseudorapidity $|\eta|$ that is used to define training bins.

The combined electron identification and isolation BDT was trained in same p_{T} and $|\eta|$ bins and was able to outperform both the identification only BDT and the selection with identification only BDT and the $\mathcal{I}_{\mathrm{PF}} < 0.35$ cut. An example of the performance gain can be seen from the ROC curves shown in Fig. 3.5.

A detailed set of results of the new classifier is presented when discussing lepton selection in Section 4.2.2, and the main gain in the view of significant reduction of reducible background in the four electron final state is presented in Section 4.5.

3.3 Electron efficiency measurements

An ideal set of selection cuts would have identical efficiency in data like in the simulation, and it would efficiently eliminate the background events while keeping the signal rates untouched. Since every real selection has at least slightly different effect in data than in simulation and it affects signal efficiencies also, this effect has to be studied and quantified. These efficiencies can be extracted directly from the simulation, but there is a danger of the bias which can appear because of insufficient event description mostly coming from the lack of understanding of the detector. Mitigating the bias from the MC simulation is achieved by measuring the selection efficiencies directly from the experimental data using the Tag-and-Probe (TnP) technique [79]. In the case of electron efficiency measurements, this method exploits $Z \to e^+e^-$ events to estimate the reconstruction and selection efficiencies and builds on Run I measurements [80, 71, 21] with several improvements introduced.

3.3.1 Tag-and-Probe technique

The TnP method uses the well known $Z \to l^+l^-$ candle to build a collection of unbiased leptons that can be used to study different selection efficiencies. First it requires one lepton candidate from the event, called *tag*, to satisfy a set of very tight selection requirements. This is important because we need a set of very pure leptons with ideally no background present. Because of the enormous number of $Z \to l^+l^-$ events and since we are willing to trade our signal efficiency for excellent purity this is not so hard to achieve. Another lepton from the event, called *probe*, is simply selected as the lepton of same flavour and opposite charge to the tag, and with as loose selection as possible. For example, a very loose mass selection on the electron probes requires the dielectron invariant mass to be in a proximity around the Z boson mass with selection $60 < m_{e^+e^-} < 120$ GeV. In this way, with a very tight selection we are trying to tag one lepton from the Z boson decay and use it to collect its decay pair from the event. On a large enough number of events this ensures an unbiased source of electrons covering the full kinematic phase-space and that can be used to measure efficiency of a given selection cut. In the case of multiple probes per one tag the random choice of probes is made.

The efficiency of the selection criteria is defined as the ratio of the number of *passing* probes N_P to the total number of probes:

$$\epsilon_{sel.} = \frac{N_P}{N_P + N_F}, \tag{3.7}$$

where N_F is the number of the probes *failing* the selection criteria.

The TnP method can, in principle, be employed in two different ways depending on the

purity of the $Z \to l^+l^-$ events before applying the cut under the study. If the purity is very high, a simple Cut-and-Count (CnC) technique gives very good estimate of the efficiency [71]. However, the CnC technique can not be used in the presence of significant background contamination, where the simultaneous fit to the Z boson invariant mass and the background underneath is used instead to estimate the number of signal events. Although the estimated backgrounds mostly come from QCD multi-jet production and usually are at a few percent level, they are larger in low-p_{T} region where a simple CnC would severely bias the measured efficiency.

In practice, probe collection is split in (p_{T},η) bins, making sure that the binning ensures enough statistics while still being fine enough to spot all kinematic dependencies. In each bin, passing and failing probes are fitted simultaneously. Fits can be done using functional forms or templates that are extracted from the fit to the simulation in the same bin. The fits to the simulation are also used to understand the presence and the shape of the background contamination, since there we can use matching to generated leptons from Z boson to ensure only signal electrons as probes. Finally, measurements of the efficiencies in each bin are used to derive *scale factors*, defined as a ratio of the efficiency of a given selection cut in data and in the simulation:

$$SF_l(p^l_{\mathrm{T}}, \eta^l) \equiv \frac{\epsilon_{data}(p^l_{\mathrm{T}}, \eta^l)}{\epsilon_{MC}(p^l_{\mathrm{T}}, \eta^l)}. \tag{3.8}$$

These scale factors, together with their uncertainty are used to scale the simulation to data in the analysis, removing all possible remaining discrepancies left from the imperfect detector modeling in the simulation.

In the $H \to ZZ^\star \to 4\ell$ analysis, scale factors and their corresponding uncertainties enter the final event selection with the power of four. It is therefore crucial to measure them precisely with a good understanding of any deviations from unity. We proceed with the actual measurements of the electron efficiencies and scale factors that were conducted in 2015, at the begining of Run II. As part of my thesis work I have introduced several improvements that have become part of the analysis and are still used when the scale factors are re-derived for new data taking periods.

The overall electron efficiency can be factorized using the efficiencies that are measured independently:

$$\epsilon = \epsilon_{reco.} \cdot \epsilon_{sel.|reco.}, \tag{3.9}$$

where each terms represents the efficiency for the probe to pass a given selection or reconstruction step, given that it passes the criteria for all the previous steps.

3.3.2 Electron reconstruction efficiency

First factor in Eq. 3.9 is the efficiency of reconstructing electrons $\epsilon_{reco.}$. The dominant source of inefficiency in the reconstruction of electrons originates from the reconstruction of the track. The commissioning in Run I showed that the efficiency of the electron cluster reconstruction is very close to 100%. There are additional very loose selections on H/E and track-cluster matching observables as part of the electron seed filtering

to reduce the rate of background electrons, but it was shown that this has negligible impact on signal. The electron reconstruction efficiency $\epsilon_{reco.}$ is thus taken to be the GSF tracking efficiency. It is measured in data by using the superclusters as probes and defining passing probes as those superclusters that are finally reconstructed as electrons. Since every analysis uses a same set of reconstructed electrons, the reconstruction scale factors are derived centrally in the CMS Collaboration and are provided for all analysis with electron final states. Up to some very technical details, the derivation relies on the same TnP technique as used when deriving $H \to ZZ^{\star} \to 4\ell$ selection scale factors as described in Section 3.3.3.

3.3.3 Electron selection efficiency

Second factor in Eq. 3.9 denotes the efficiency of selection cuts on a collection of electrons that already passed the reconstruction step. To measure the electron selection efficiency, data samples that trigger on single electron trigger paths and Drell-Yan simulation samples are used. Events that trigger on double electron trigger paths are not considered because this would include some trigger selection cuts on the probes that would bias our efficiency measurement.

A set of tight requirements on electron tags that ensure high purity imply:

- $p_{\mathrm{T}} > 30$ GeV, $|\eta_{\mathrm{SC}}| < 2.5$, together with the removal of electrons from the EB-EE ECAL crack with $1.4442 < |\eta_{\mathrm{SC}}| < 1.566$,
- tight cut-based ID with working point at 70%,
- geometrical matching of tag to the leg of a single electron trigger object, to ensure that probes do not have any trigger selection cuts,
- in the case of simulation, geometrical matching to the generated electrons coming from Z boson decay.

An example of passing and failing probe distribution can be seen in Fig. 3.6. In the simulation, only electrons matched to generated electrons are used, so there are no background contributions present. Because of this, we are able to choose signal shape model that describes the invariant mass distribution. It is clear that a resonance shape convoluted with some smearing function will describe the obvious Z boson. A Breit-Wigner function convoluted with a double-sided Crystall Ball function was choosen for the signal shape, and other possible choices were studied as sources of systematical uncertainty.

In the data, one can see a clear presence of the background for the failing probes. There are several reasons why the background is only visible for the failing probes. First, because of the high efficiency of the electron ID step only a couple of percent of events end up as failing probes. And as expected, a wast majority of a multi-jet fakes that are present in the probe collection will fail the electron ID thus contaminating a small sample of failing signal electrons with a sizable background contribution. This represents

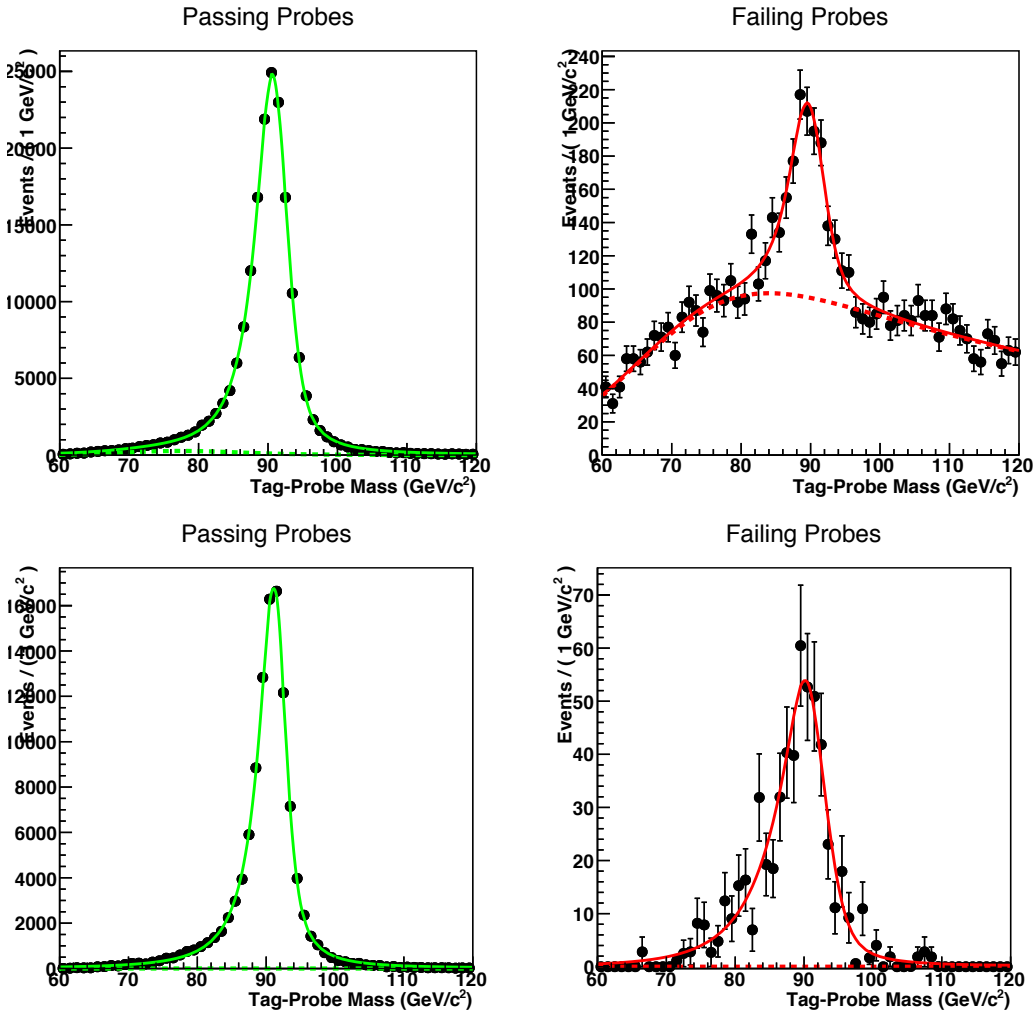

Figure 3.6 – Example of fits to the experimental data (top) and simulated data (bottom) for probe electrons that pass (left) or fail (right) the identification selection. The distributions refer to EB electrons in the $30 < p_{\mathrm{T}} < 40$ GeV window. Black points are experimental or simulated data, dashed lines represent background models and full lines represent background plus signal model.

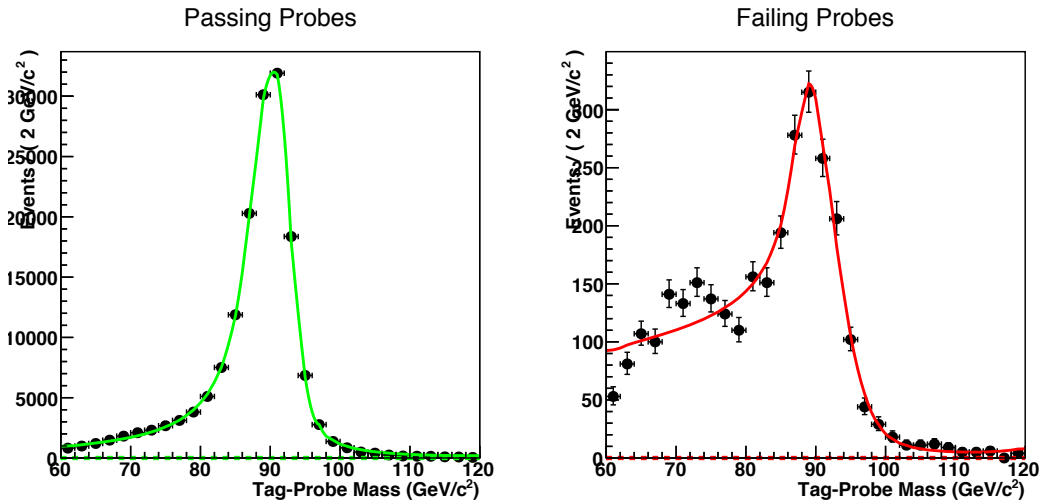

Figure 3.7 – Example of fits to the simulated data for probe electrons that pass (left) or fail (right) the identification selection. The distributions refer to electrons in the $20 < p_{\mathrm{T}} < 30$ GeV and $0.8 < |\eta_{\mathrm{SC}}| < 1.479$ window. Black points are simulated data and full lines represent signal model.

the main challenge of the TnP efficiency measurements, and as we will see in the very low-p_{T} region it is extremely challenging to distinguish any signal amongst the dominant QCD background. For background modeling an error function multiplied with an exponential function was chosen, and again other possible shapes were studied as sources of systematical uncertainties. With a simple function integration one can then obtain the efficiency of ID for data and simulation in the given (p_{T},η_{SC}) bin and calculate the corresponding scale factor using Eq. 3.8.

When studying fits in all defined bins a peculiar feature was observed for failing probe distributions in the $20 < p_{\mathrm{T}} < 30$ GeV window, as can be seen in Fig. 3.7. A fit to the signal shape failed for failing probes since there was a clear accumulation of low-mass events, or a *bump* in the invariant mass spectrum that could not be described with the current signal model. Since these are simulated electrons that are geometrically matched to generated signal electrons, it meant that this contribution was coming from a special population of signal events and that it has to be included in the signal shape in order to measure efficiency properly. Identical effect was observed in Run I and was assigned to kinematic threshold effect arising from the binning in transverse momentum. However, if it was really just a threshold effect, one would expect a similar bump in the passing probes as well. For some reason, these signal events from the bump dominantly failed the electron identification and in order to justify their modeling as signal a detailed investigation was needed.

Spurious events from the bump were studied using MC matching. Various observables were checked, looking for deviations from the expected output. One such deviation was found when supercluster pseudorapity distribution was generated using only events originating from the bump, as shown on Fig. 3.8. A very strong correlation between

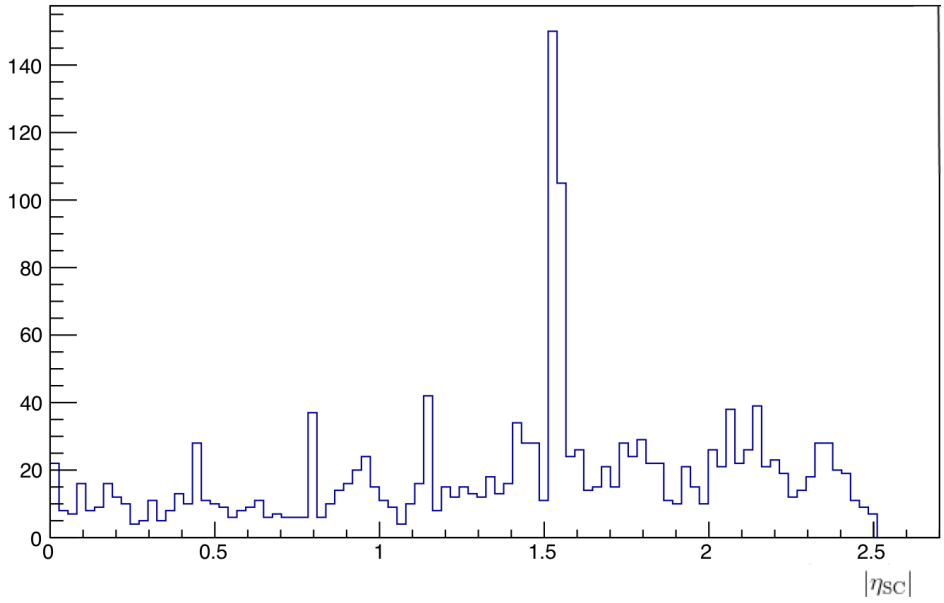

Figure 3.8 – Distribution of the supercluster pseudorapidity for failing probes in the $20 < p_T < 30$ GeV with invariant mass requirement $60 < m_{e^+e^-} < 80$ GeV.

spikes in distribution of the $|\eta_{SC}|$ and the position of cracks between the ECAL modules described in Section 2.3.3 was observed. The largest spike was indicating abundance of events that end up in the big crack in the EB-EE transition region.

Further investigations showed another clear feature, a photon was very often found in the proximity of a tag or a probe for the events from the bump. In the end, it was concluded that the events from bump consist of three main contributions of similar sizes:

- electrons located near cracks between the EB and EE, and between modules of the ECAL barrel,
- electrons that experienced final-state radiation (FSR),
- electrons with significant amount of bremsstrahlung radiation that was not accounted for during reconstruction.

These effects are illustrated in Fig. 3.9 showing passing and failing probe distributions in the low-p_T bin using MC simulation. Clear reduction in the low-mass tail is visible once we remove probes coming from the crack and afterwards probes from events with significant FSR radiation. What is left is a small contribution in the tail coming from events where significant bremsstrahlung radiation was not reconstructed. To summarize, the bump effect comes from poorly measured electrons originating either from cracks or having radiated photons not accounted for. Because of the poor momentum measure-

ment and due to kinematic effect of binning these events accumulated as a bump on invariant mass distributions in certain p_{T} bins.

In order to successfully deal with these effects, three improvements to electron efficiency measurements were proposed:

- measuring crack and non-crack scale factors separately, so that well and poorly measured electrons are not mixed,
- a Gaussian function is added to signal model for low-p_{T} bins to account for this signal contribution,
- FSR recovery algorithm described later in Section 4.2.4 is applied to tags and probes before measuring efficiency.

With the application of the above mentioned steps it was expected that only a small bump will remain for the low-p_{T} crack probes, while for the non-crack probes the effect should completely vanish thanks to the FSR recovery. However, that was not the case. A clear bump was still present and further investigations showed that there was a big inefficiency present in the current version of the FSR recovery algorithm. This algorithm was constructed in Run I, when a different electron reconstruction procedure was used and it was never optimized for the PF reconstruction. One of the steps in the old FSR recovery algorithm was to discard all photon candidates where $\Delta\mathrm{R}_{e,\gamma} < 0.15$. As one can see from Fig. 3.10 a vast majority of photon candidates were rejected with this cut making old FSR recovery algorithm very inefficient for Run II. Indeed, as illustrated in Fig. 3.11 just removing this cut significantly improves the efficiency of the FSR recovery algorithm.

Because of these findings, FSR recovery algorithm was optimized for the analysis of 2016 and 2017 data. A study on the gluon fusion signal sample showed an improvement since the fraction of events with recovered FSR photon increased from 1.5% to 4.3% for the $4e$ events in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis.

Systematic uncertainties

An important aspect of the scale factor determination is the estimate of the corresponding uncertainty. Even more for the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis where the estimated uncertainty enters the final selection with the power of four. It is therefore important not to underestimate the uncertainty, but also not to be too conservative. As we will see when discussing results in Chapter 5, lepton efficiency measurement is the main source of the experimental systematics for many measurements in this analysis.

A total uncertainty for the measurement of the scale factors is a quadratic sum of statistical and systematical uncertainties. Statistical uncertainties are obtained directly from the fits. Systematical uncertainties require a detailed study where one has to be cautious to take all the possible systematics into account but not to double-count. During

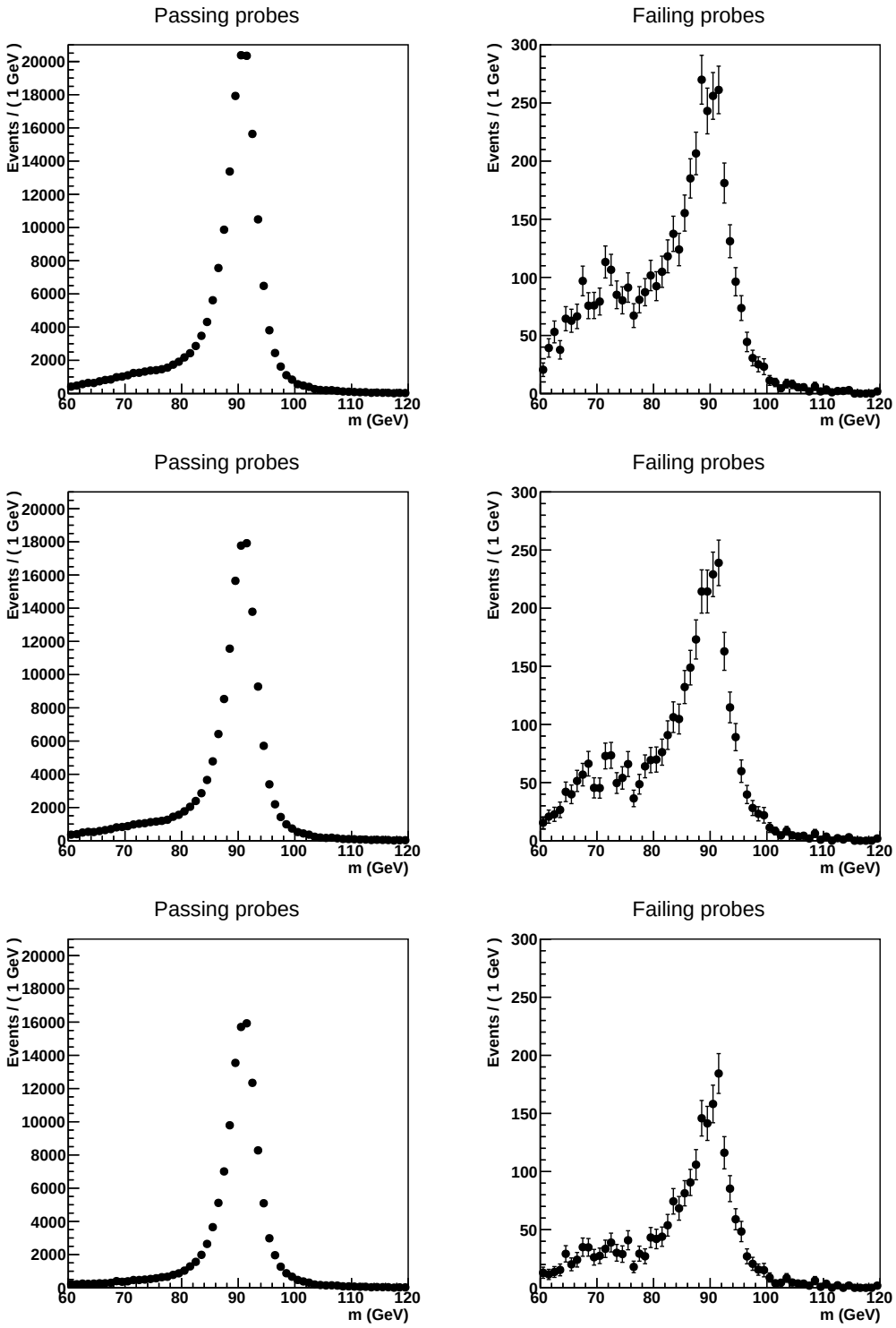

Figure 3.9 – Example of invariant mass distributions for passing (left) and failing probes (right) for simulation sample in $20 < p_{\mathrm{T}} < 30$ GeV bin integrated over $|\eta|$. Identification selection cut is considered on three sets of probes with different signal contributions: all probes (top), all probes that are not coming from ECAL cracks (middle), and all probes other than crack probes and probes that have FSR photon associated to them (bottom).

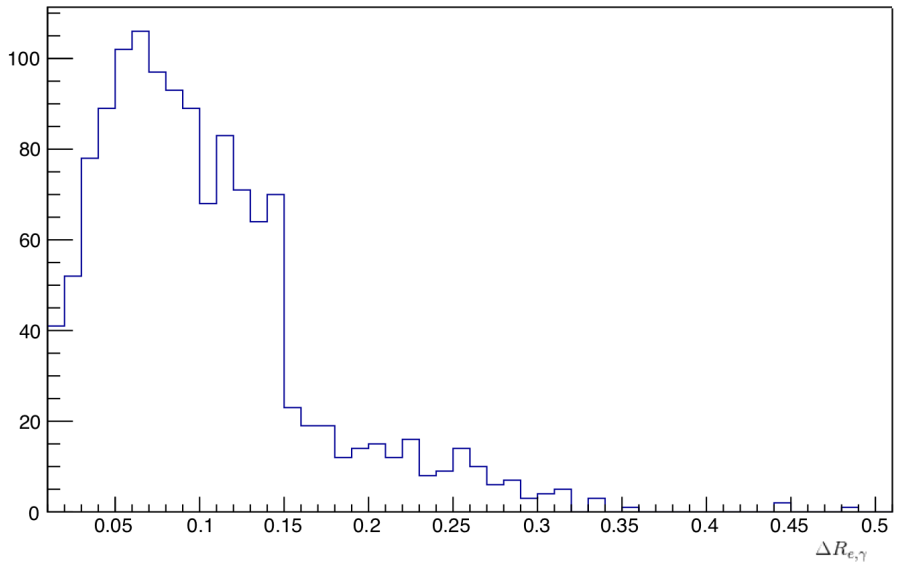

Figure 3.10 – Distribution of the distance between photon candidates and electron probes for failing probes in the $20 < p_{\mathrm{T}} < 30$ GeV with invariant mass requirement $60 < m_{e^+e^-} < 80$ GeV. Only events where FSR photon candidates are found were considered.

the years a list of general systematic studies was developed and is suggested to all TnP studies. It consists of:

- **Uncertainty in the accuracy of the signal model:** Variation of the signal shape from an analytic shape to a simulation-based template, or to a different analytic shape, or from a template to a signal shape depending on the starting choice.
- **Uncertainty in the background modeling:** Variation of the background model to an exponential function, or a high-order polynomial function.
- **Uncertainty in the background coming from the tag selection:** Different selection on the tag.
- **Uncertainty in the overall event description:** Using higher order NLO Drell-Yan simulation instead of the default LO.
- **Uncertainty in the pileup estimation:** Varying the minimum bias cross section used for pileup reweighting for $\pm 5\%$.

With the exception of very forward electrons with $|\eta_{\mathrm{SC}}| > 2.0$ and $p_{\mathrm{T}} > 100$ GeV, the measurement is limited by the systematic uncertainties. The total uncertainty is of the order of 10% for the low-p_{T} bins and around 1-2% otherwise.

To understand the origin of such a high uncertainty in the low-p_{T} region an example fit from the 2015 study is shown in Fig. 3.12. Because of the very high contamination of

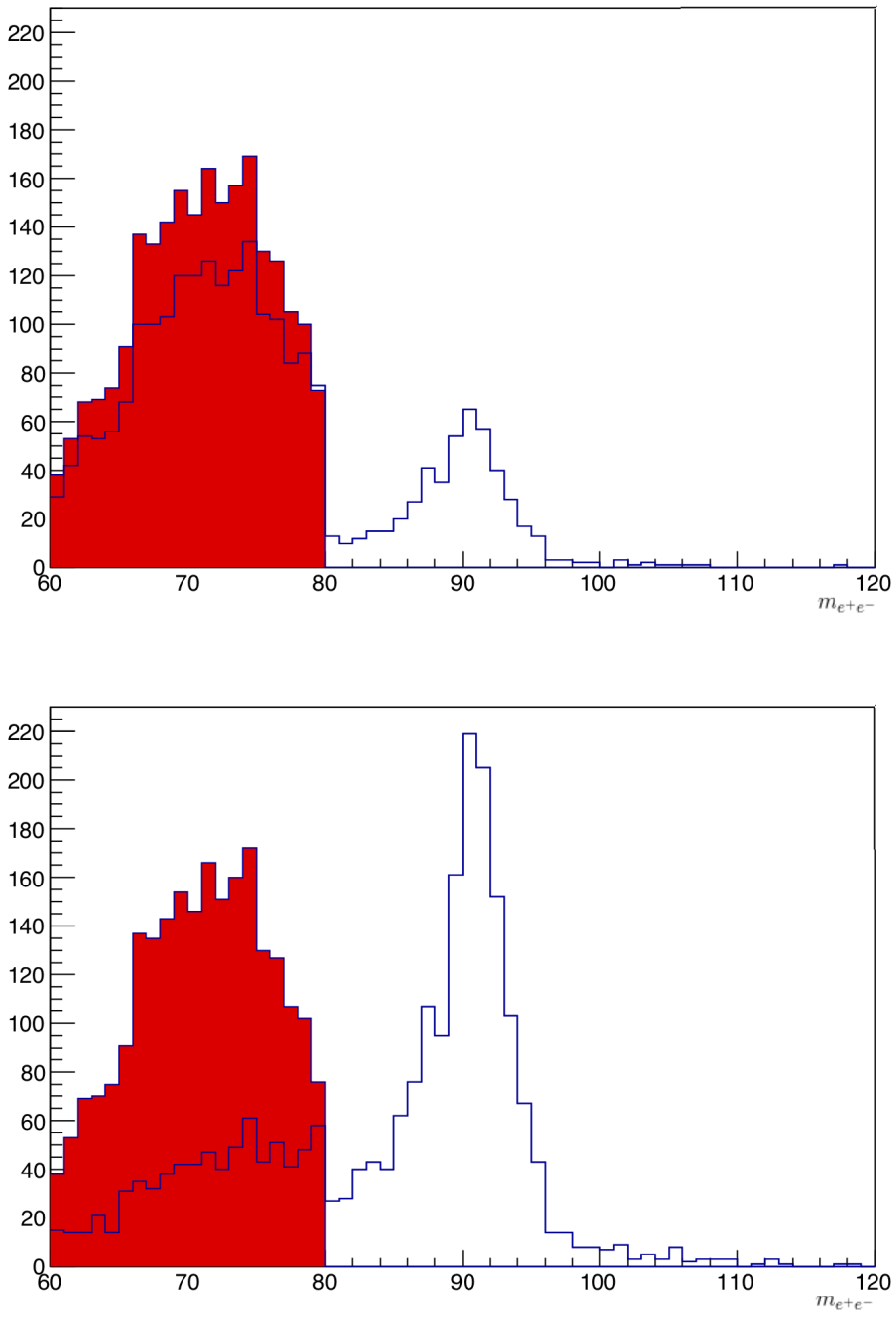

Figure 3.11 – Invariant mass distribution $m_{e^+e^-}$ for the tag and probe pair without consideration of the FSR photons (red filled histogram) and after applying the FSR recovery algorithm (blue empty histogram). Old FSR recovery algorithm from Run I is applied (top), and its version without $\Delta R_{e,\gamma} < 0.15$ cut (bottom). Only probes from the bump mentioned in text are shown, coming from $20 < p_T < 30$ GeV, $0.8 < |\eta_{SC}| < 1.479$, and $60 < m_{e^+e^-} < 80$ GeV window.

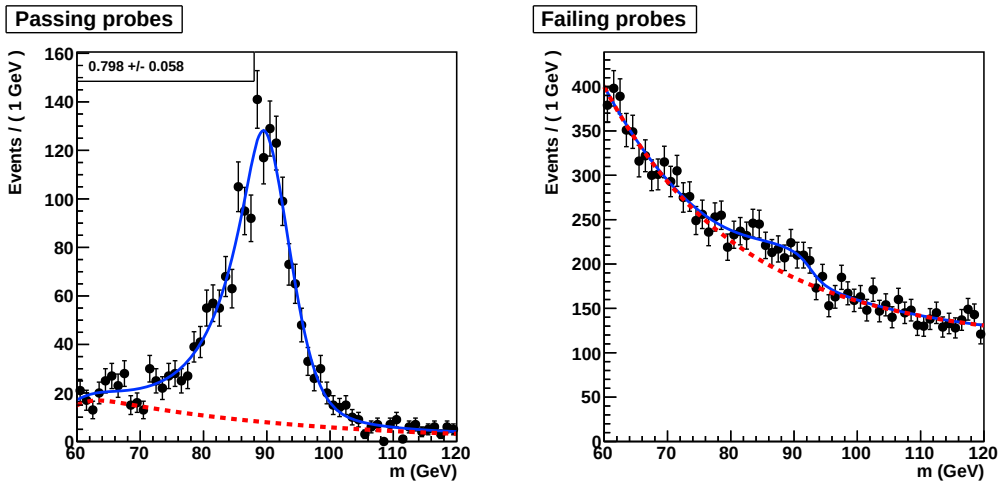

Figure 3.12 – Example of fits to the experimental data for probe electrons in the EE that pass (left) or fail (right) the identification selection. The distributions refer to electrons in the $7 < p_{\mathrm{T}} < 10$ GeV window. Black points are 2015 experimental data, red line is background model and blue line is signal plus background model. Measured efficiency with corresponding statistical uncertainty obtained from the fit is reported.

background in the very low-p_{T} region a Z boson resonance is barely visible in the failing probe invariant distribution. This makes the fit very challenging, and in this case already the statistical uncertainty is ∼6%. When one adds up all the systematic effects described above it is easy to understand how this uncertainty can grow to an order of 10%. This is very problematic for the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, where the need to reconstruct offshell $\mathrm{Z}^{\star}$ requires low-p_{T} leptons in every event. One of the ideas that was investigated was to clean the probe sample from background with some very loose requirements that are not correlated with the selection cuts. If there is no correlation between the selection cuts and cuts on the probe, in principle there would be no bias introduced. However, it was not possible to find any observables that are not contained in at least one of the selection steps of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis.

The only possible solution was to measure the challenging bins as precisely as possible with the given distributions in Run II and to prepare some ideas for the future. Some of the ideas that are being discussed are the use of the $\mathrm{Z} \to l^{+}l^{-}l^{+}l^{-}$ as a standard candle, given there is enough statistics. Another idea is to try and exploit low-mass resonances like $\mathrm{J}/\Psi \to \mathrm{e}^{+}\mathrm{e}^{-}$ and $\Upsilon \to \mathrm{e}^{+}\mathrm{e}^{-}$ for these measurements, as it is done in the muon case. However, it is extremely challenging to trigger on these events for the electron case and new ideas are needed to accomplish this.

My studies on the electron efficiency measurements that were conducted in the analysis of the first 2.8 fb^{-1} of Run II data were presented. Only the work flow and new developments were discussed here, while the results of the efficiency and scale factor measurements will be given for the full 2016 and 2017 datasets in the following chapter.

Chapter 4

Building blocks of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis

In this Chapter a cornerstone of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, were $\ell = e, \mu$, is being discussed. In order to study properties of the SM Higgs boson it is essential to define an optimal selection procedure on all possible events stored by the CMS.

The selection process of this analysis targets signal Higgs boson production while keeping as few background events as possible. A set of observables that are able to distinguish Higgs boson signal from SM background are defined and used for extraction of different results. A crucial part of every analysis is certainly modeling of signal and background estimation. Without a reliable estimate of the signal and background rates and kinematics, using state of the art theoretical knowledge and experimental techniques, we would be insensitive to new physics. There are two different sources of background in every analysis, the irreducible and reducible backgrounds. The irreducible background has a final state identical to the signal but is produced via other SM processes that are not of the interest. For example, in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, main irreducible backgrounds are $\mathrm{gg} \to \mathrm{ZZ}$ and $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ with both Z bosons decaying to a pair of leptons. These processes are simulated, while a more challenging task is modeling of the reducible background that arises from different contributions containing final state objects that are misidentified as leptons coming from Z boson. Dominant sources are events with a Z boson and jets, where for example heavy flavour jets produce secondary leptons or decay of charged hadrons overlapping with π^0 decays get misidentified as leptons. An improvement of the analysis in Run II, which benefits largely from the significant increase in the number of golden Higgs boson events that are being recorded, is the introduction of the categorization scheme that targets all main Higgs boson production mechanisms. One has to also understand all possible sources of systematic uncertainties and estimate their effect. As we will see in the following chapters, this is becoming even more important since some of the interesting results are no longer statistically limited. Finally, all these pieces are glued together in a statistical analysis where a likelihood

approach is used to measure physics parameters of interest.

This Chapter is organized to discuss the building blocks of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis. This is the same way the analysis is actually being performed: optimizing every part of selection, checking data to simulation agreement in control regions, defining measurement techniques and developing them on the simulation before applying everything to recorded data. It is done not to introduce any bias to the analysis, and it is a common practice in CMS, even after the discovery of the Higgs boson. Final results are presented later in Sections 5.1 and 5.2, for 2016 and 2017 respectively. The combined results of two data taking periods are discussed in Section 5.3.

4.1 Data and simulated samples

4.1.1 Recorded data sets

The analysis presented in this thesis benefits from results of two data taking periods, recorded by the CMS detector during 2016 and 2017.

Standard CMS selection of runs and luminosity sections is applied to ensure the highest possible data quality with a good functioning of the different sub-detectors. Thus, similar detector operation conditions and offline data quality checks are imposed to both the 2016 and 2017 datasets for the validation of the data to be used for the analysis. Samples corresponding to $35.9\ \mathrm{fb}^{-1}$ in 2016 and $41.5\ \mathrm{fb}^{-1}$ in 2017 at $\sqrt{s} = 13$ TeV are used.

For computational reasons, the recorded events of each period are split into primary data sets, each of which contains all events collected by a particular subset of all HLT paths that involve similar physics objects. This analysis relies on single and multilepton HLT paths, therefore exploiting the DoubleMuon, DoubleEG, MuonEG, SingleElectron and SingleMuon primary data sets.

The 2016 data set has been *re-reconstructed*, which means that the standard CMS event reconstruction was run again in Winter of 2017 with updated detector conditions and improved calibrations with respect to those used during the data taking. Since the preliminary results using 2017 data were produced during the data taking, they exploit the *prompt reconstruction* workflow, which means that the data underwent the standard CMS event reconstruction as soon as they were recorded, with preliminary cosmics-based detector calibrations. The data is then studied further to better master the detector alignment and fine-tune the calibrations, so that a re-reconstruction can be performed on raw data and used in the final publications, for the benefit of precision measurements such as that of the Higgs boson mass.

The data sets and trigger paths used for 2016 and 2017 analysis summarized in Table 4.1 and Table 4.2.

Table 4.1 – Data sets used in the 2016 and 2017 analysis.

2016	2017
Datasets	
/DoubleMuon/Run2016B-03Feb2017_ver2-v2	/DoubleMuon/Run2017B-17Nov2017-v1
/DoubleEG/Run2016B-03Feb2017_ver2-v2	/DoubleEG/Run2017B-17Nov2017-v1
/MuonEG/Run2016B-03Feb2017_ver2-v2	/MuonEG/Run2017B-17Nov2017-v1
/SingleElectron/Run2016B-03Feb2017_ver2-v2	/SingleElectron/Run2017B-17Nov2017-v1
/SingleMuon/Run2016B-03Feb2017_ver2-v2	/SingleMuon/Run2017B-17Nov2017-v1
/DoubleMuon/Run2016C-03Feb2017-v1	/DoubleMuon/Run2017C-17Nov2017-v1
/DoubleEG/Run2016C-03Feb2017-v1	/DoubleEG/Run2017C-17Nov2017-v1
/MuonEG/Run2016C-03Feb2017-v1	/MuonEG/Run2017C-17Nov2017-v1
/SingleElectron/Run2016C-03Feb2017-v1	/SingleElectron/Run2017C-17Nov2017-v1
/SingleMuon/Run2016C-03Feb2017-v1	/SingleMuon/Run2017C-17Nov2017-v1
/DoubleMuon/Run2016D-03Feb2017-v1	/DoubleMuon/Run2017D-17Nov2017-v1
/DoubleEG/Run2016D-03Feb2017-v1	/DoubleEG/Run2017D-17Nov2017-v1
/MuonEG/Run2016D-03Feb2017-v1	/MuonEG/Run2017D-17Nov2017-v1
/SingleElectron/Run2016D-03Feb2017-v1	/SingleElectron/Run2017D-17Nov2017-v1
/SingleMuon/Run2016D-03Feb2017-v1	/SingleMuon/Run2017D-17Nov2017-v1
/DoubleMuon/Run2016E-03Feb2017-v1	/DoubleMuon/Run2017E-17Nov2017-v1
/DoubleEG/Run2016E-03Feb2017-v1	/DoubleEG/Run2017E-17Nov2017-v1
/MuonEG/Run2016E-03Feb2017-v1	/MuonEG/Run2017E-17Nov2017-v1
/SingleElectron/Run2016E-03Feb2017-v1	/SingleElectron/Run2017E-17Nov2017-v1
/SingleMuon/Run2016E-03Feb2017-v1	/SingleMuon/Run2017E-17Nov2017-v1
/DoubleMuon/Run2016F-03Feb2017-v1	/DoubleMuon/Run2017F-17Nov2017-v1
/DoubleEG/Run2016F-03Feb2017-v1	/DoubleEG/Run2017F-17Nov2017-v1
/MuonEG/Run2016F-03Feb2017-v1	/MuonEG/Run2017F-17Nov2017-v1
/SingleElectron/Run2016F-03Feb2017-v1	/SingleElectron/Run2017F-17Nov2017-v1
/SingleMuon/Run2016F-03Feb2017-v1	/SingleMuon/Run2017F-17Nov2017-v1
/DoubleMuon/Run2016G-03Feb2017-v1	
/DoubleEG/Run2016G-03Feb2017-v1	
/MuonEG/Run2016G-03Feb2017-v1	
/SingleElectron/Run2016G-03Feb2017-v1	
/SingleMuon/Run2016G-03Feb2017-v1	
/DoubleMuon/Run2016H-03Feb2017_ver2-v1	
/DoubleEG/Run2016H-03Feb2017_ver2-v1	
/MuonEG/Run2016H-03Feb2017_ver2-v1	
/SingleElectron/Run2016H-03Feb2017_ver2-v1	
/SingleMuon/Run2016H-03Feb2017_ver2-v1	
Integrated luminosity	
$35.9\ \mathrm{fb}^{-1}$	$41.5\ \mathrm{fb}^{-1}$
Comments	
re-reconstructed	prompt reconstruction

Table 4.2 – HLT paths used in the 2016 and 2017 analysis.

2016	2017
Single muon triggers	
HLT_IsoMu20	HLT_IsoMu27
HLT_IsoTkMu20	
HLT_IsoMu22	
HLT_IsoTkMu22	
HLT_IsoMu24	
HLT_IsoTkMu24	
Single electron triggers	
HLT_Ele25_eta2p1_WPTight_Gsf	HLT_Ele35_WPTight_Gsf
HLT_Ele27_WPTight_Gsf	HLT_Ele38_WPTight_Gsf
HLT_Ele27_eta2p1_WPLoose_Gsf	HLT_Ele40_WPTight_Gsf
HLT_Ele32_eta2p1_WPTight_Gsf	
Double muon triggers	
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	
Double electron triggers	
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	HLT_DoubleEle33_CaloIdL_MW
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL	
Triple muon triggers	
HLT_TripleMu_12_10_5	HLT_TripleMu_10_5_5_DZ
	HLT_TripleMu_12_10_5
Triple electron triggers	
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL
Cross triggers	
HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	HLT_Mu8_DiEle12_CaloIdL_TrackIdL
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	HLT_Mu8_DiEle12_CaloIdL_TrackIdL_DZ
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	
HLT_DiMu9_Ele9_CaloIdL_TrackIdL	

4.1.2 Simulated samples

Monte Carlo (MC) simulation is a cornerstone of every particle physics analysis, being used in various steps from event selection optimization and estimation of background rates to production of final results. Production of simulated samples for CMS analysis can be summarized in five main steps:

1. Hard physics process of interest is generated with a full list of events with all four-momenta of outgoing particles calculated from the SM Lagrangian.
2. Hadronization of quarks and gluons, jet fragmentation and showering, and addition of underlying events is simulated.
3. Overlapping pp (pileup) interactions are included in the simulated sample.

4. Processing through a detailed simulation of the CMS detector, where the output is raw detector information.

5. Emulation of HLT and reconstruction of events with the same algorithms as used for data.

The default parton distribution functions (PDFs) used in all simulations are NNPDF30_nlo_as_0118 [81] for 2016 analysis, and NNPDF31_nlo_hessian_pdfas and NNPDF31_lo_as_0130 [82] for 2017 NLO and LO simulation samples, respectively. All signal and background event generators are interfaced with PYTHIA 8 [83], with CUETP8M1 tune [84], to simulate the multi-parton interaction and hadronization effects. The generated events are processed through a detailed simulation of the CMS detector based on GEANT4 [85, 86] before being reconstructed with the official CMS software.

Signal samples

Information about MC simulated samples for the Higgs boson signal processes is given in Table 4.3. Signal samples are generated at NLO in perturbative quantum chromodynamics (pQCD) with the POWHEG 2.0 [87, 88, 89] generator for the five main production modes [90, 91]: gluon fusion (ggH), vector boson fusion (VBF) and associated production (WH, ZH, and $\mathrm{t\bar{t}H}$). For WH and ZH the MINLO HVJ [92] extension of POWHEG 2.0 is used. Two other production modes, $\mathrm{b\bar{b}H}$ and tqH, are generated using JHUGEN [93, 94, 95, 96]. In all cases the decay of the Higgs boson to four leptons is modeled with JHUGEN [93, 94], which properly accounts for interference effects associated with permutations of identical leptons in the $4e$, 4μ and 4τ final states. Adding $2e2\mu$, $2e2\tau$, and $2\mu2\tau$, six final states are included at generator level in total, even though the analysis does not use reconstructed tau leptons. Nevertheless, owing to the existence of leptonic decays of the taus, a small amount of events involving tau pairs are reconstructed as $4e$ or 4μ events. The cross sections for the various signal processes are taken from ref. [9], and in particular the cross section for the dominant gluon fusion production is taken from ref. [8].

In order to generate an accurate signal model the p_{T} spectrum of the Higgs boson ($p_{\mathrm{T}}(\mathrm{H})$) was tuned in the POWHEG simulation of the dominant gluon fusion production mode to better match predictions from full phase space calculations implemented in the HRES 2.3 generator [97, 98]. To take advantage of the most accurate simulation of gluon fusion available, a reweighting of the gluon fusion sample is defined. Gluon fusion events are separated into 0, 1, 2, and $\geq$3 jet bins, where the jets used for counting are clustered from all stable particles, excluding the decay products of the Higgs boson or associated vector bosons, having $p_{\mathrm{T}} > 30$ GeV. The sum of weights in each sample are first normalized to the inclusive cross section. The ratio of the $p_{\mathrm{T}}(\mathrm{H})$ distribution from the NNLOPS generator [99] to that from the POWHEG generator in each jet bin is applied to the ggH signal samples.

Table 4.3 – List of signal MC samples used in the analysis, for the case where $m_H = 125$ GeV. In this table, l means e, μ, or τ and X denotes the set of any other possible particles in the considered final states. Information on generator software is quoted, along with the production cross section at 13 TeV.

process	generator	σ [pb]
$gg \to H \to ZZ \to 4\ell$	POWHEG 2.0 + JHUGEN V7	48.58
$qq' \to Hqq' \to ZZqq' \to 4\ell qq'$	POWHEG 2.0 + JHUGEN V7	3.782
$q\bar{q} \to W^+H \to W^+ZZ \to 4\ell + X$	POWHEG 2.0 + MINLO HWJ + JHUGEN V7	0.840
$q\bar{q} \to W^-H \to W^-ZZ \to 4\ell + X$	POWHEG 2.0 + MINLO HWJ + JHUGEN V7	0.533
$q\bar{q} \to ZH \to ZZZ \to 4\ell + X$	POWHEG 2.0 + MINLO HZJ + JHUGEN V7	0.884
$gg \to t\bar{t}H \to t\bar{t}ZZ \to 4\ell + X$	POWHEG 2.0 + JHUGEN V7	0.507
$gg \to b\bar{b}H \to b\bar{b}ZZ \to 4\ell + X$	JHUGEN V7	0.488
$qq' \to Htq' \to ZZtq' \to 4\ell tq'$	JHUGEN V7	0.077

In addition to samples simulated for the Higgs boson mass value of $m_H = 125$ GeV, all main production mechanisms are also simulated in other five points: 120, 124, 125, 126, and 130 GeV. These samples are used to parameterize the analysis expectations as a function of the Higgs boson mass, which is discussed in detail in Section 4.6. In 2016 analysis $b\bar{b}H$ and tqH signal contributions were not considered, while in the 2017 analysis $b\bar{b}H$ was generated in all five mass points, and tqH only for $m_H = 125$ GeV scenario.

Additional signal samples are also produced for studies of systematic uncertainties. For example they include different generators, varied pythia parameters for the underlying event tune, varied hadronization scale, and will be mentioned when discussing systematic uncertainties estimation in Section 4.7.

Background samples

Information about MC simulated samples for the SM background processes is given in Table 4.4.

The SM ZZ background contribution from quark-antiquark annihilation is generated at NLO pQCD with POWHEG 2.0, with the same settings as signal samples. The $gg \to ZZ$ process is generated at LO with MCFM [100]. To match the $gg \to H \to ZZ$ transverse momentum spectra predicted by POWHEG at NLO, the showering for MCFM samples is performed with different PYTHIA 8 settings, allowing only emissions up to the parton-level scale. Additional scale factors that account for missing higher-order corrections are applied to both backgrounds, and will be discussed in Section 4.5.1.

Other background samples of WZ, Drell-Yan + jets, and $t\bar{t}$ are generated using either MADGRAPH5_AMC@NLO [101] or POWHEG 2.0. As will be explained in Section 4.5.2, the contribution of these reducible backgrounds is actually estimated from control regions in data, but the simulated samples are used for the optimization and validation of the

methods. Drell-Yan + jets samples are also used in lepton-level optimization studies and efficiency measurements as a source of signal and background leptons, as explained in Sections 3.2 and 3.3.

Table 4.4 – List of MC samples used for the estimation of irreducible backgrounds and for the study of control regions. Information on generator software is quoted, along with the production cross sections at 13 TeV times relevant branching ratios. These cross sections correspond to the phase space of event generation. In this table, l means e, μ, or τ.

process	generator	$\sigma \cdot BR$ [pb]	additional information
$q\bar{q} \to ZZ \to 4\ell$	POWHEG 2.0	1.256	
$gg \to ZZ \to 4e$	MCFM	0.00159	no interference
$gg \to ZZ \to 4\mu$		0.00159	
$gg \to ZZ \to 4tau$		0.00159	
$gg \to ZZ \to 2e2\mu$		0.00319	
$gg \to ZZ \to 2e2\tau$		0.00319	
$gg \to ZZ \to 2\mu2\tau$		0.00319	
$Z/\gamma^* \to \ell\ell +$ jets (Drell-Yan)	MADGGRAPH5_AMC@NLO	5765.4	$m_{ll} > 50$ GeV
$WZ \to 3\ell\nu$	MADGGRAPH5_AMC@NLO	5.290	
$t\bar{t} \to 2\ell2^{\nu}2b$		87.31	

Simulation reweigthing

While pileup interactions do get simulated, it is done with first recorded data and well before the finish of the data taking period. It is thus impossible to predict accurately the exact profile of the pileup and it is needed to perform a reweighting of the simulated samples to match the actual pileup profile in data. The average number of additional pileup interactions is either computed from the number of reconstructed primary vertices or from the measured instantaneous luminosity per bunch crossing, and its distributions in simulation and data are used to compute pileup weights for a particular data taking period. An illustration of the magnitude of these weights is presented in Fig. 4.1 for both data taking periods used in the analysis. This procedure of pileup reweighting is found to have only a sub-percent impact on expected yields in the signal region.

Final step is to scale every simulated event with its corresponding weight, calculated as:

$$w_{\text{event}} = \frac{\mathrm{L_{int.}} \cdot \sigma \cdot \mathrm{BR} \cdot w_{\mathrm{PU}} \cdot \mathrm{SF}_{4\ell} \cdot w_{\text{generator}}}{\sum\limits_{\text{allevents}} w_{\text{generator}}}, \tag{4.1}$$

where $\mathrm{L_{int.}}$ is the integrated luminosity of the data taking period, σ is the best known cross section of the simulated sample, BR is branching ratio, w_{PU} is pileup weight which is obtained as just explained in the text above, $\mathrm{SF}_{4\ell}$ is the scale factor of the event which is obtained by multiplying scale factors of all four leptons as defined in eq. 3.8, and $w_{\text{generator}}$ is generator weight which can be even negative for NLO and higher order samples.

For signal processes, the 125 GeV cross sections used for scaling are listed in Table 4.3, while the branching ratio for Higgs boson decay to four leptons via a pair of Z bosons

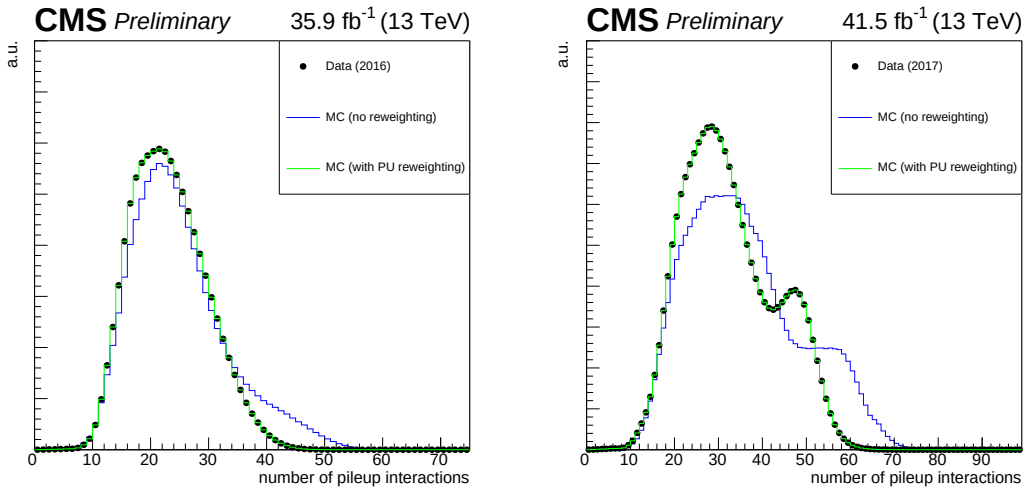

Figure 4.1 – Distribution of the number of pileup interactions in 2016 (letf) and 2017 (right) simulation before and after pileup reweighting, compared to that in data.

is $2.76 \cdot 10^{-4}$. Additional generator filter efficiency is applied to VH and $t\bar{t}H$ to account for the non-resonant signal explained later in Section. 4.6. For background processes, in Table 4.4 a multiplier of cross section times branching fraction is reported.

4.2 Event selection

Event selection in the $H \to ZZ^{\star} \to 4\ell$ analysis can be divided in three main steps:

1. **Trigger selection**: Saving events that potentially have Higgs boson decays for further analysis. Designed to be as close as possible to 100% efficient for the signal.
2. **Lepton selection**: Distinguishing between real and fake leptons by exploiting different observables. Designed to reduce the rate of backgrounds as much as possible without sizable reduction of the signal efficiency.
3. **ZZ candidate selection**: Building Z and ZZ candidates from selected leptons and apply additional cuts on them. Designed to be as close as possible to 100% efficient for signal and to remove peculiar and very hard to simulate backgrounds.

Very efficient event selection, with excellent signal to background ratio in the interesting signal region, was already established in Run I. It was motivated by a very small branching fraction, with very few events per inverse fb of data expected. Main improvements in Run II come from optimization of lepton selection, where background rates have been significantly reduced for same signal efficiencies as we saw on the example of electrons in Chapter 3. Other steps of the analysis proved to be already optimal and underwent only some fine tuning or small changes.

4.2.1 Trigger selection

First, we have to ensure that amongst billions of collisions that are being detected by CMS every second we record all possible Higgs boson candidates. This is achieved with the help of the trigger, and was already described in Section 2.3.3. A dedicated collection of HLT paths has been designed and optimized to cover the phase space of the 4ℓ signal and are listed in Table 4.2 for 2016 and 2017 data taking periods. By definition, the events of the recorded collision data sample are known to have fired at least one path of the HLT menu. This menu is also emulated in MC samples, so that the same trigger requirement can be applied as in data. The trigger requirement is defined as follows: any event where at least one of these 4ℓ-dedicated paths is fired is always accepted, without tying the three possible final states to particular paths. The latter point is motivated by the hunt for production modes: for instance, the triple-electron path is not tied to the $4e$ final state, because a $2e2\mu$ VH event can have additional true electrons coming from the decay of the associated W or Z boson.

Trigger efficiency measurement

Just like any other selection step, the efficiency of the trigger selection needs to be determined in collision data and simulation in order to correct for possible mismatches, either caused by mismodelling effects or by the fact that the menu used in data taking is not well simulated for MC samples. Due to several reasons, a standard TnP approach is not applicable for this measurement. For example, there are no Level-1 seeds with thresholds low enough to measure the efficiencies of the lowest-p_{T} legs of triple lepton triggers. A modified Tag-and-Probe-like method is developed instead, based on 4ℓ events that pass the signal region selection, and using the HLT paths in a more global approach.

One of the four reconstructed leptons considered as a tag, is geometrically matched to a trigger object passing the final filter of one of the single muon or single electron triggers. The other three leptons are used as probes. In each 4ℓ event there are up to 4 possible tag-probe combinations, and all of them are counted in the denominator of the efficiency. If the three probe leptons fire at least one of the 4ℓ-dedicated paths, the set of probes is counted for in the numerator of the efficiency. This method does not have a perfect closure in MC events due to the fact that the presence of a fourth lepton increases the trigger efficiency, and this effect is not accounted for. Also, in the $2e2\mu$ final state, the three probe leptons cannot be combined to form all possible triggers which can collect events with two electrons and two muons. For example, if the tag lepton is an electron, the three remaining leptons can not pass a double electron trigger. Therefore the method is also performed on MC and the difference between data and MC is used to determine the reliability of the simulation. The efficiency plotted as a function of the minimum p_{T} of the three probe leptons in data and MC using this method can be seen in Fig. 4.2. The trigger efficiency in simulation is found to be $> 99\%$ in each final state. From additional studies and checks it was concluded that MC efficiency describes well the data within the statistical uncertainties, and it was decided to not apply any scale factors, but to use a p_{T} dependent systematic uncertainty that covers the small observed discrepancies.

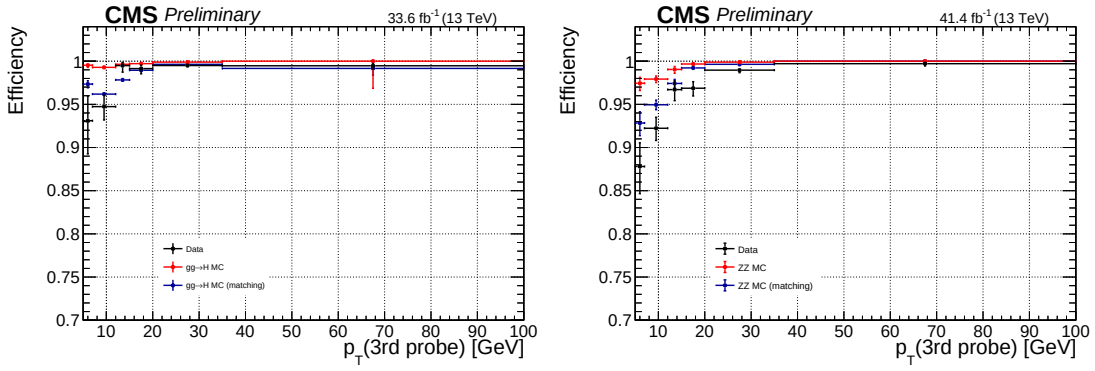

Figure 4.2 – Trigger efficiency measured in data and simulation using 4ℓ events collected by single lepton triggers for the 2016 (left) and 2017 (right) analysis. Efficiency is measured with TnP-like approach described in text in data (black) and simulation (blue) and with direct method in simulation (red).

4.2.2 Lepton selection

In Chapter 3, design and optimization of electron selection was described in detail. In this Section, particularities of muon selection are described before discussing the results of lepton calibration, selection, and efficiency measurements with 2016 and 2017 data.

Kinematic cuts

To be within the detector reconstruction acceptance pseudorapidity cuts of $|\eta^{e}| < 2.5$, and $|\eta^{\mu}| < 2.4$ are applied to electrons and muons, respectively. To be considered for the analysis, reconstructed electrons are required to have a transverse momentum p_{T}^{e} larger than 7 GeV, while low-p_{T} reconstruction of muons is slightly more efficient allowing muons with p_{T}^{μ} larger than 5 GeV.

Impact parameter cuts

Muons and electrons are required to pass identical set of requirements on impact parameter observables: $\mathrm{d_{xy}} < 0.5\mathrm{cm}$, $\mathrm{d_{z}} < 1\mathrm{cm}$, and $\mathrm{SIP_{3D}} < 4$. In the case of muons these cuts mostly aim to suppress fakes that originate from in-flight decays of hadrons and cosmic rays, and for electrons those coming from photon conversions.

Identification

Electron identification for 2016 analysis was discussed in detail in Section 3.2.2. The ROC curves for the 2016 electron BDT classifier are shown in Fig. 4.3 with cuts values listed in Table 4.5.

Electron identification was further optimized with the inclusion of isolation observables

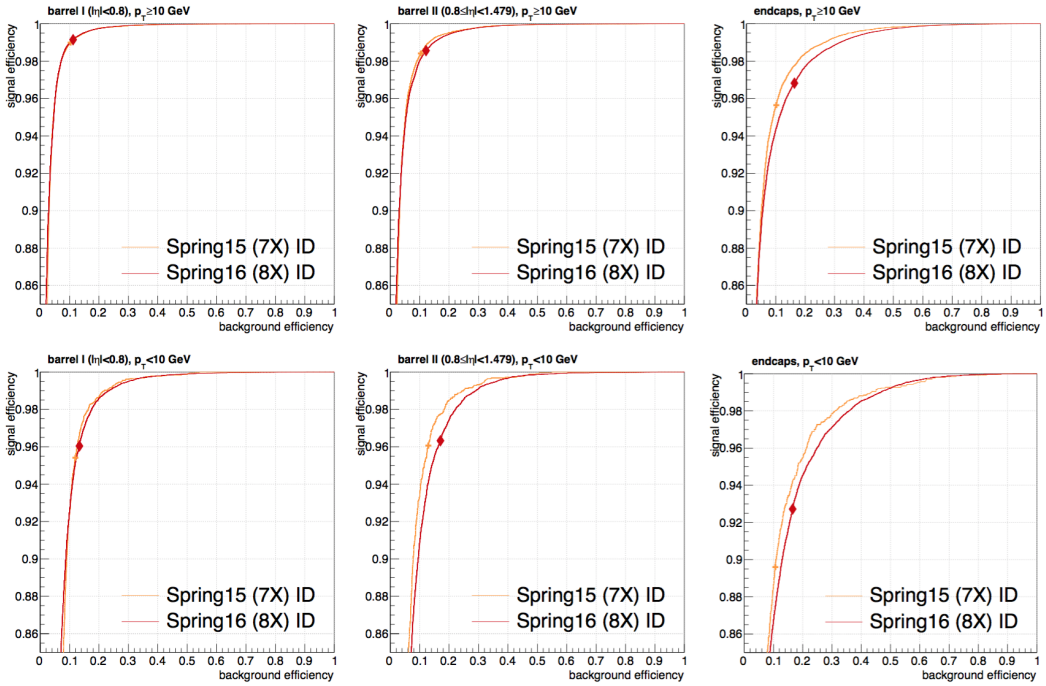

Figure 4.3 – Performance comparison of the electron BDT classifier trained for the 2015 analysis (orange) and the retraining for 2016 conditions (red) in all 6 $(p_{\mathrm{T}}, |\eta|)$ bins. The respective working points are indicated by the markers.

Table 4.5 – Minimum BDT score required for passing the 2016 electron identification.

minimum BDT score	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.479$	$\|\eta\| > 1.479$
$5 < p_T < 10$ GeV	-0.211	-0.396	-0.215
$p_T > 10$ GeV	-0.870	-0.838	-0.763

for 2017 analysis, as discussed in Section 3.2.4. The performance comparison of 2016 and 2017 electron BDT classifier can be seen in Fig. 4.4 for various configurations. It is obvious that for same signal efficiency significant reduction of selected background electrons is achieved. Additional studies showed that the gain in electron discrimination power has similar contributions coming from the improved BDT classifier and from pixel upgrade for 2017 described in Section 2.3.3. For completeness, cut values are listed in Table 4.6. As a reminder, in both cases working points were chosen such to preserve the overall signal efficiency from Run I.

Table 4.6 – Minimum BDT score required for passing the 2017 electron identification with isolation.

minimum BDT score	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.479$	$\|\eta\| > 1.479$
$5 < p_T < 10$ GeV	0.628	0.592	0.637
$p_T > 10$ GeV	0.036	0.043	-0.266

A fact that muons are easier to reconstruct and suffer from much less complicated backgrounds than electrons is maybe best pronounced in the choice of the muon identifi-

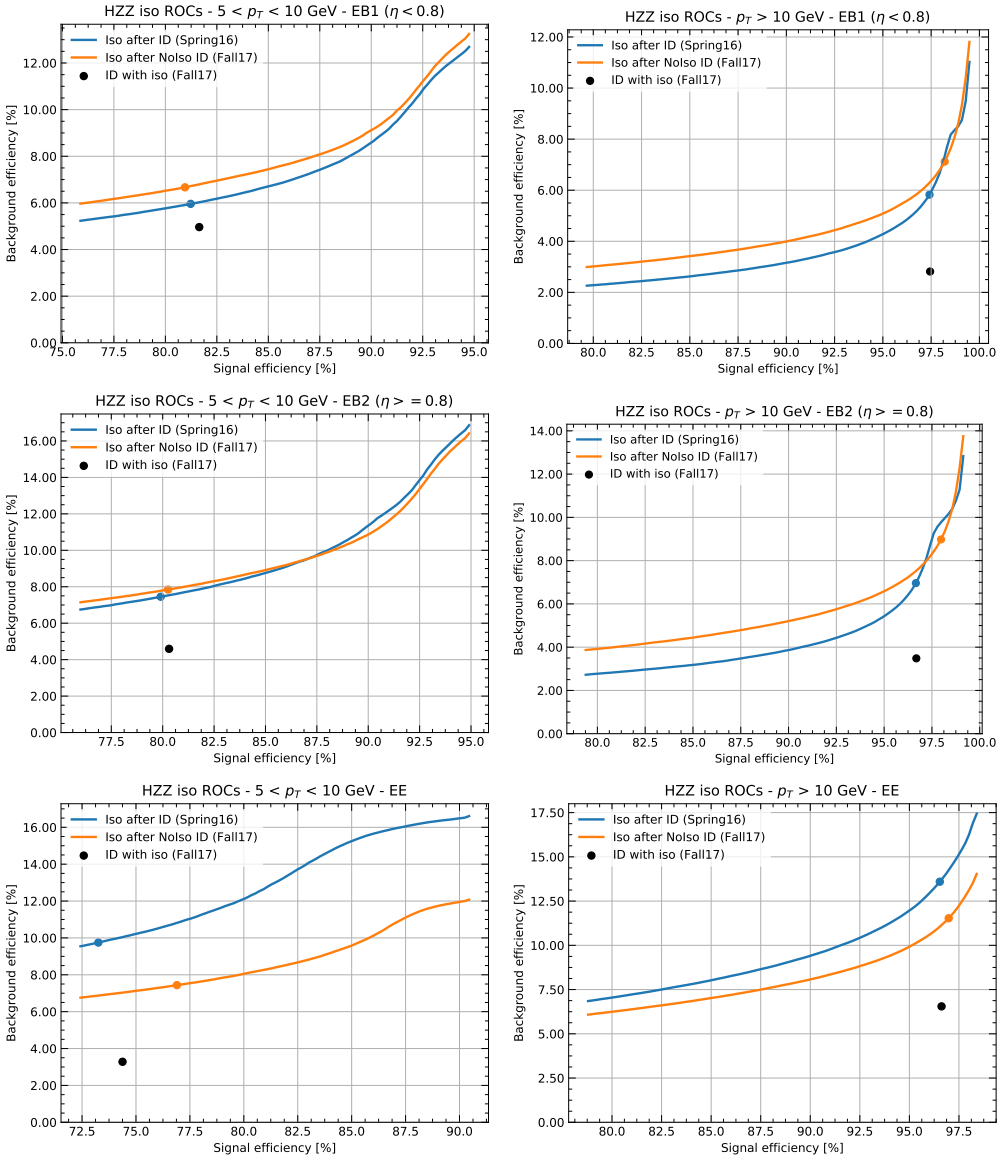

Figure 4.4 – Performance comparison of several electron BDT classifiers: 2016 BDT identification with additional $\mathcal{I}_{\mathrm{PF}} < 0.35$ isolation cut (blue), 2017 BDT trained only with identification observables and with additional $\mathcal{I}_{\mathrm{PF}} < 0.35$ cut (orange), and finally 2017 BD with identification and isolation variables used in training (black). Markers indicate chosen working points. For 2017 BDT with isolation it is impossible to show a ROC curve on this plot since identification variables are used in classifier. Performance are shown for electrons with $5 < p_{\mathrm{T}} < 10$ GeV (left), $p_{\mathrm{T}} > 10$ GeV (right) and $|\eta| < 0.8$ (top), $0.8 < |\eta| < 1.479$ (middle) and $|\eta| > 1.479$ (bottom).

cation criteria. While for electrons, a dedicated BDT classifier had to be developed for the purpose of this analysis, muons profit from simple general PF identification requirements that are provided centrally in CMS and used in most of the analysis with muons in final state.

The PF muons are identified from muons reconstructed by the PF algorithm as described

in Section 2.4.5. Reconstructed muons are considered to be isolated if, in a cone of size $\Delta R = 0.3$ centered on the muon, the sum of the p_{T} of the tracks and of the transverse energy of the calorimeter hits is less than 10% of the muon p_{T}. Such isolated muons have by definition little neighbouring activity, and PF techniques are thus not needed to resolve additional neutral particles in their vicinity. It is important to note here that this detector based isolation is different from the analysis based isolation described next. Maximum efficiency is ensured for the analysis stage by applying only a very loose selection on these isolated muons. They are only required to be global muons, meaning that a valid fit exists between the track from the central tracking system and the track from the muon system. Non-isolated muon candidates can also be selected as PF muons if they both include a minimum number of hits in the muon track and satisfy a matching requirement of the track to hits in the muon stations.

Isolation

As already mentioned, lepton isolation is a powerful tool to suppress backgrounds from misidentified leptons. It is very analysis dependent, so for the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis a detailed study was performed in Run I to chose a working point that ensured satisfactory reduction of fakes for demanded high signal efficiency. This was slightly tuned at the beginning of Run II as described in Section 3.2.3. For 2016 and 2017 analysis, additional studies showed that there is no clear room for improvement so the previous setup remained. Both muons and electrons use the standard relative PF isolation as defined in eq. 3.5 and a cut value of $\mathcal{I}_{\mathrm{PF}} < 0.35$.

The only difference is in the way the PU contribution is subtracted for neutral hadrons and photons. For electrons, correction based on effective areas and mean energy density in the event is used, with expression given in eq. 3.6.

For muons, the so-called $\Delta\beta$ correction, which is defined as:

$$p_{\mathrm{T}}^{\mathrm{PU}} \equiv \frac{1}{2} \cdot \sum_{i} p_{\mathrm{T}}^{\mathrm{PU},i} \tag{4.2}$$

where i runs over the momenta of the charged hadron PF candidates that do not originate from the primary vertex. This formula relies on the assumption that the pileup contribution of neutral particles to the isolation cone is roughly half of that of charged particles associated to pileup.

4.2.3 Lepton momentum calibration and scale factors

To ensure high quality of the data, all important objects and their corresponding observable distributions are being checked and compared with expectations from simulation. It is no different for the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, where all of the observables used for selection on leptons are cross-checked for every data taking period. Data is both being compared with simulation and being checked for any unexpected time dependencies in dedicated control regions. An example of such check on the lepton momentum calibra-

tions is shown in Fig. 4.5 where one can see that the lepton momentum calibrations do a very good job of removing any time dependencies of the reconstructed Z boson mass.

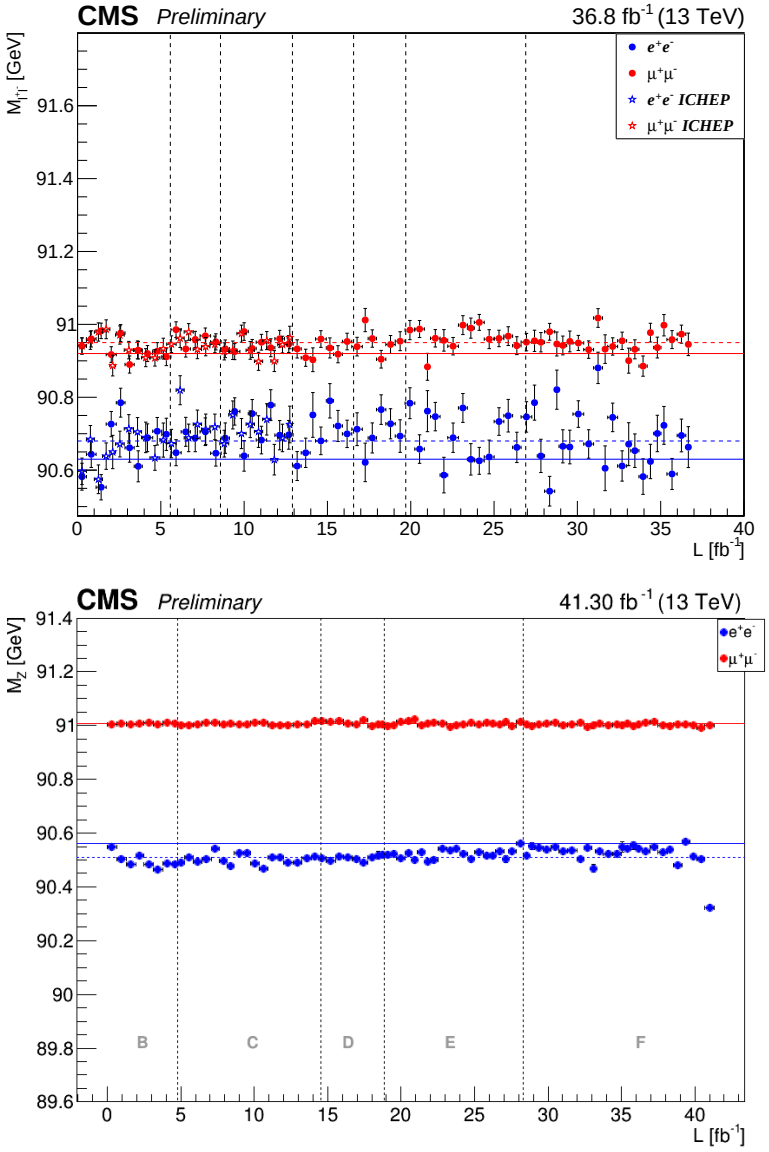

Figure 4.5 – Electron (blue) and muon (red) reconstructed Z boson mass in 500 pb luminosity blocks for 2016 (top) and 2017 (bottom) data taking periods. In each luminosity block, Z boson mass is extracted from the mean value given by the fit of the dilepton invariant mass to a Crystal Ball function. Horizontal lines show the value of the Z boson mass obtained from the fit to full data sample (dashed) and obtained from fit to simulation sample (dashed). Vertical dashed lines indicate different data taking periods. The upper plot includes star markers, indicating values for the first version of $\sim$13 fb^{-1} of prompt reconstructed 2016 data that was used for CMS publications on the 2016 International Conference of High Energy Physics (ICHEP).

Finally, efficiencies for each selection step are evaluated with the Tag-and-Probe (TnP) technique described in Section 3.3.1 and corresponding scale factors are derived to cor-

rect for any possible detector modeling imperfections remaining.

For electrons, this is done in two steps. First, reconstruction scale factors are derived centrally for all analysis and afterwards scale factors for analysis specific selections are derived. Electron reconstruction and selection efficiencies for 2016 analysis are shown in Fig. 4.6. One can immediately notice strong deviations from unity for reconstruction scale factors in endcaps. Deeper investigations showed that in the 2016 dataset the tracking efficiency was reduced due to a lower hit reconstruction efficiency in the silicon strip detector caused by highly ionising particles (HIPs). The large signals generated by HIPs can momentarily saturate the front-end readout chip for the silicon strip tracker, resulting in deadtime in the detector readout system. This example proves the importance of correctly measuring and applying scale factors. For $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ selection scale factors were found to be ∼2% smaller than unity. Total uncertainty on the measurement varies from 1 to 3% in the statistically limited low and high p_{T} regions, and is below 1% for the intermediate region. However, these uncertainties are propagated to the overall event uncertainties with the power of four, and as it will be discussed in Section 4.7 they become one of the leading sources of experimental systematics.

Same procedure is done for muons as well, the only difference being the presence of the $\mathrm{J}/\Psi \to \mu^{+}\mu^{-}$ samples that are helpful for determination of the very low-p_{T} efficiencies. Results of the overall selection scale factors that includes tracking, impact parameter selection, PF identification, and isolation for 2016 analysis is shown in Fig. 4.7. While scale factors for muons with $p_{\mathrm{T}} > 20$ GeV were found to be compatible with unity, low-p_{T} muons showed similar behavior like electrons yielding in scale factors that are 1 to 3% below unity. Total uncertainties were found to be from 1 to 3% in the low-p_{T} region, and well below 1% otherwise.

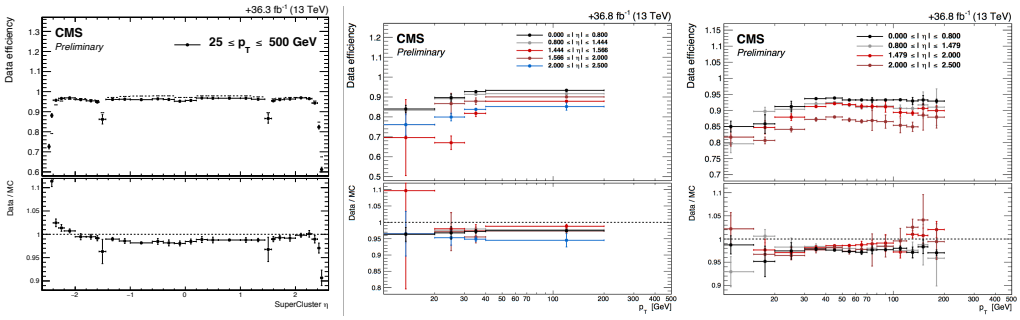

Figure 4.6 – Electron reconstruction (left) efficiency as a function of electron η^{SC}, and $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ selection efficiencies as a function of electron p_{T} for crack electrons (middle) and non-crack electrons (right). Measurements are performed using 2016 data with the Tag-and-Probe technique described in Section 3.3.1. Corresponding scale factors are also shown on the bottom part of the respective figures.

Scale factors have to be rederived for every new data taking period. Electron reconstruction and $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ selection efficiencies and scale factors for 2017 analysis are shown in Fig.4.8. For this data taking period both sets of scale factors were found to be compatible with unity with their corresponding uncertainties. However, due to

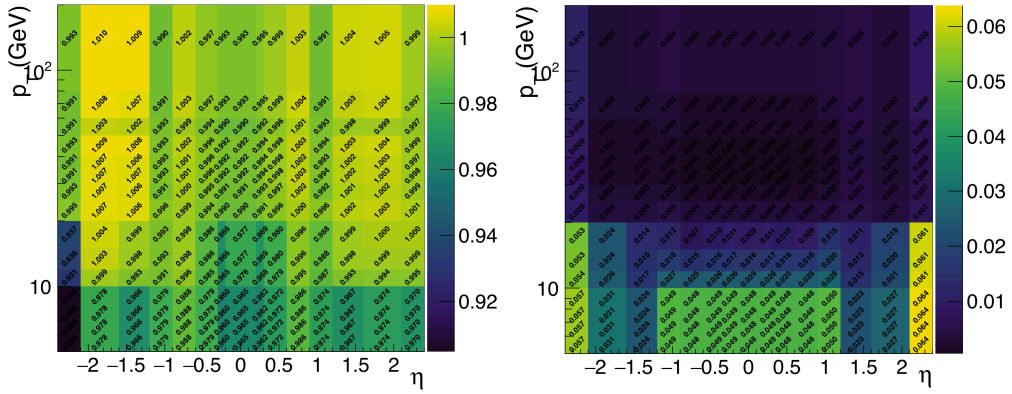

Figure 4.7 – Overall data to simulation scale factors (left) and corresponding uncertainties (right) for muons, as function of p_T and η. Measurements are performed using 2016 data with the Tag-and-Probe technique described in Section 3.3.1 and include muon tracking, impact parameter selection, PF identification, and isolation.

increase of pileup and harsher conditions, measured uncertainties in the challenging low-p_T region increased with respect to the 2016 analysis.

Overall muon scale factors with corresponding uncertainties for the 2017 analysis are shown in Fig. 4.9. They were found to be compatible with unity with their corresponding uncertainties. Thanks to the presence of dedicated $J/\Psi \to \mu^+\mu^-$ uncertainties on the measurement were found to be of same magnitude like in the 2016 analysis.

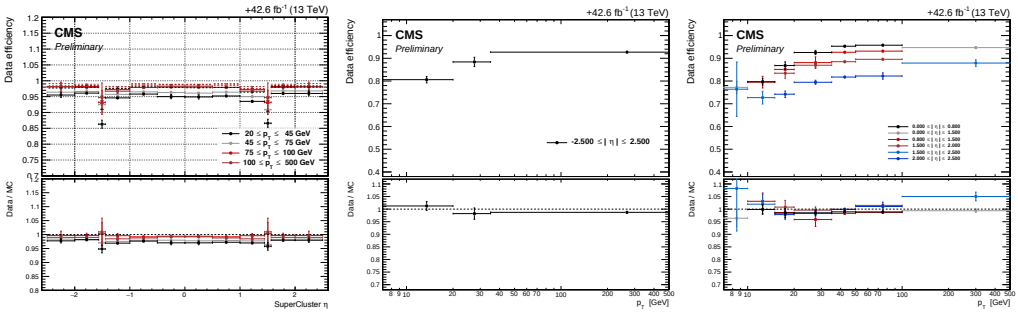

Figure 4.8 – Electron reconstruction (left) efficiency as a function of electron η^{SC}, and $H \to ZZ^{\star} \to 4\ell$ selection efficiencies as a function of electron p_T for crack electrons (middle) and non-crack electrons (right). Measurements are performed using 2017 data with the Tag-and-Probe technique described in Section 3.3.1. Corresponding scale factors are also shown on the bottom part of the respective figures.

4.2.4 FSR photon recovery

Photons are considered in the $H \to ZZ^{\star} \to 4\ell$ analysis only as a possible candidates for final-state radiation (FSR) recovery. Since electrons and muons coming from Z bosons can radiate a high-p_T photon, not considering its four-momentum would significantly

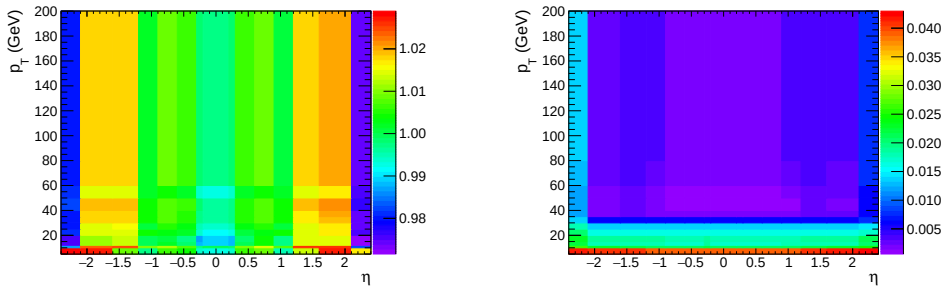

Figure 4.9 – Overall data to simulation scale factors (left) and corresponding uncertainties (right) for muons, as function of p_{T} and η. Measurements are performed using 2016 data with the Tag-and-Probe technique described in Section 3.3.1 and include muon tracking, impact parameter selection, PF identification, and isolation.

degrade the accuracy in the reconstruction of the Z boson kinematics. Motivated by this, a FSR recovery algorithm was developed with a purpose to find FSR photons and to account for them when reconstructing Z boson candidates. Also, these photons are then subtracted from the PF isolation sums defined in Section 3.2.3 recovering some leptons that were otherwise failing the isolation requirement. Starting from the collection of PF photons, provided by the particle-flow algorithm, the selection of photons and their association to a lepton proceeds as follows:

1. The preselection of PF photons is done by requiring $p_{\mathrm{T}}^{\gamma} > 2$ GeV, $|\eta^{\gamma}| < 2.4$, and $\mathcal{I}_{\mathrm{PF}}^{\gamma} < 1.8$. The latter variable is computed using a cone of radius $R = 0.3$, a threshold of 0.2 GeV on charged hadrons with a veto cone of $\Delta R > 0.0001$, and 0.5 GeV on neutral hadrons and photons with a veto cone of $\Delta R > 0.01$, also including the contribution from pileup vertices.

2. Supercluster veto: we remove all PF photons that match with any electron passing both the loose ID and SIP cuts. The matching is peformed by directly associating the two PF candidates. This step was previously performed with a cut on the $\Delta R_{e,\gamma}$ and was very inefficient for PF reconstruction. It was updated thanks to findings from electron efficiency measurements described in Section 3.3.3.

3. Photons are associated to the closest lepton in the event among all those the kinematic and impact parameter cuts.

4. Photons that do not satisfy the cuts $\Delta R(\ell, \gamma) < 0.5$ and $\frac{\Delta R(\ell,\gamma)}{(p_{\mathrm{T}}^{\gamma})^2} < 0.012$ GeV^{-2} are discarded.

5. In the very rare case when more than one photon is associated to the same lepton, photon with the lowest $\frac{\Delta R(\ell,\gamma)}{(p_{\mathrm{T}}^{\gamma})^2}$ is selected.

6. Each FSR photon that was selected is excluded from the isolation sum of all the leptons in the event that pass the kinematic and impact parameter cuts.

Studies on the signal gluon fusion samples showed that the FSR algorithm affects about 5% of all signal events. It has a bigger effect on the muons than electrons, for which the majority of FSR photons are already included in the refined supercluster.

4.2.5 Jet selection

In the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis jets are used to distinguish different Higgs boson production mechanisms. Firstly, reconstructed jets described in Section 2.4.6 are subject to momentum calibrations similar to those applied to leptons. Then, a set of additional cuts that help reduce the background coming from the calorimeter noise and large number of pileup interactions is applied. Finally, special algorithms are developed to distinguish jets originating from bottom quarks.

Jet momentum calibration

The jet momentum determined as a simple vectorial sum of all particle momenta in the jet is found to reproduce the true momentum at the 5 to 10% level over the whole p_{T} spectrum and detector acceptance. The momentum estimate is further improved by applying the *jet energy corrections* (JEC). The JEC are designed as a sequence of corrections, each targeting a specific effect, and implemented as a scaling of the jet momentum based on event and jet observables. The first step removes the dependence of the jet energies on pileup and detector noise via a mean energy density method. The next step attempts to make the jet energy response uniform in η and p_{T}, with corrections derived from simulation and residual corrections obtained from dijet, γ + jet, and Z + jet measurements [102]. Measured JEC are applied to jets in both simulation and data. Besides the correction of the jet energy scale, a smearing of the jet energy resolution is applied in simulation to match that observed in data.

Jet selection cuts

To be selected in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, jets must satisfy $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV and $\left|\eta^{\mathrm{jet}}\right| < 4.7$, and to be separated from all selected lepton and FSR photon candidates by $\Delta R(\ell/\gamma, \mathrm{jet}) > 0.4$. A cut-based jet identification is applied, designed to reject fake, badly reconstructed and noise jets while keeping 98 to 99% of real jets [103]. It takes into consideration the correlations and the discriminating power of all the PF jet observables, as well as the importance of observables for possible HCAL and ECAL noise rejection. For 2016 analysis a loose working point was chosen. Due to improvements of the jet ID in the 2017 analysis a tight working point was chosen with efficiency >99% over the whole η^{jet} range.

For 2017 analysis, the identification of jets from pile-up was adopted. It relies on three types of the jet observables, describing: the compatibility of the jet with the primary interaction vertex, the topology of the jet shape, and the object multiplicity. A cut-based pileup jet ID [103] based on the mentioned variables was chosen with a tight working

point.

Jets that pass all of the above cuts are selected for the further use in the analysis, as main inputs in the event categorization described in Section. 4.3.

Jet b-tagging

Many measurements and searches in CMS rely on the accurate identification of jets originating from the hadronization of bottom quarks. This has motivated the development and optimization of b-jet identification techniques, referred to as b-tagging. These mostly rely on the fact that the B hadrons present in b jets have relatively long lifetime, and thus decay a few millimetres away from the primary interaction vertex. Thanks to the very good position resolution of the pixel tracker, the impact parameters of the resulting displaced tracks are measured, and a secondary vertex may be reconstructed. Similar to electron ID BDT, b-tagging algorithms usually condense the extracted information into one classifier variable, the b tagger, which is computed for every jet. In the present analysis, $\mathrm{t\bar{t}H}$ and ZH with $\mathrm{Z} \to b\bar{b}$ production modes can be targeted by exploiting the b-tagging technique.

In the 2016 analysis, the Combined Secondary Vertex (CSVv2) b-tagger [104] was used. It combines the information on displaced tracks with information related to secondary vertices associated to the jet, using a multivariate regression. The $\mathrm{H} \to 4\ell$ analysis uses a medium 80% efficiency working point, which has a misidentification probability of around 1% for light-flavour jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV.

In the 2017 analysis, a newly developed Deep CSV b-tagger was used with the medium working point with same efficiency but smaller misidentification probability than CSVv2. It exploits same variables as CSVv2 but with the use of deep neural networks to train the classifier.

The imperfect modeling of the distributions of the b tagger causes some discrepancies of the b-jet identification efficiency and misidentification probability between data and simulation. To correct for these differences, scale factors have to be applied to the simulation. Scale factors are applied to data as a function of the jet $p_{\mathrm{T}}^{\mathrm{jet}}$, $\left|\eta^{\mathrm{jet}}\right|$ and flavour.

4.2.6 ZZ candidate selection

Given that most of the object selection is shared between 2016 and 2017 analysis, with several improvements in the electron and jet selection, a clear overview of the object selection for the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis is presented in Table 4.7.

Table 4.7 – Summary of physics object selection for the $H \to ZZ^\star \to 4\ell$ analysis performed on the 2016 and 2017 data.

$H \to ZZ^\star \to 4\ell$ object selection summary	
2016	**2017**
Electrons	
$p_T^e > 7$ GeV $\quad |\eta^e| < 2.5$	
$d_{xy} < 0.5\,cm \quad d_z < 1\,cm$	
$SIP_{3D} < 4$	
BDT ID (cuts from Table 4.5) $\mathcal{I}^e_{PF} < 0.35$	BDT ID with isolation (cuts from Table 4.6)
Muons	
$p_T^\mu > 5$ GeV $\quad |\eta^\mu| < 2.4$	
$d_{xy} < 0.5\,cm \quad d_z < 1\,cm$	
$SIP_{3D} < 4$	
PF muon ID	
$\mathcal{I}^\mu_{PF} < 0.35$	
FSR photons	
$p_T^\gamma > 2$ GeV $\quad |\eta^\gamma| < 2.4$	
$\mathcal{I}^\gamma_{PF} < 1.8$	
$\Delta R(\ell,\gamma) < 0.5 \quad \frac{\Delta R(\ell,\gamma)}{(p_T^\gamma)^2} < 0.012$ GeV^{-2}	
Jets	
$p_T^{jet} > 30$ GeV $\quad |\eta^{jet}| < 4.7$	
$\Delta R(\ell/\gamma, jet) > 0.4$	
Cut-based jet ID (loose WP) CSV b-tagging	Cut-based jet ID (tight WP) Jet pileup ID (tight WP) Deep CSV b-tagging

The event selection of the $H \to ZZ^{\star} \to 4\ell$ analysis is identical for the 2016 and 2017 analysis, and can be summarized in the following steps:

1. The event is required to contain at least one reconstructed primary vertex passing some quality requirements, such as a small radius and collisions restricted along the z axis.

2. *Selected leptons* are defined as the reconstructed leptons that pass kinematic and impact parameter cuts, identification, and FSR-corrected isolation cut, as described in Section 4.2.2.

3. *Z candidates* are defined as all possible pairs of selected leptons of opposite charge and matching flavour, either e^+e^- or $\mu^+\mu^-$, that satisfy $12 < m_{\ell\ell(\gamma)} < 120$ GeV, where the $m_{\ell\ell(\gamma)}$ invariant mass includes the selected FSR photons if any.

4. *ZZ candidates* are defined as all possible pairs of Z candidates that do not share a same lepton. The Z candidate with invariant mass $m_{\ell\ell(\gamma)}$ closest to the nominal Z^0 boson mass given by the PDG is denoted as Z_1, and the other one is denoted as Z_2.

5. ZZ candidates are required to satisfy the following list of requirements:

   - **Ghost removal**: all six pairs that can be built with the four leptons must be separated by $\Delta R > 0.02$.
   - **Lepton p_T**: two of the four selected leptons must pass $p_{T,1} > 20$ GeV and $p_{T,2} > 10$ GeV.
   - **Z_1 invariant mass**: $m_{Z_1} > 40$ GeV.
   - **QCD suppression**: all four opposite-sign pairs that can be built with the four leptons must satisfy $m_{\ell^+\ell^-} > 4$ GeV, regardless of lepton flavour. Here, selected FSR photons are not used in computing $m_{\ell^+\ell^-}$, since a QCD-induced low mass dilepton resonance may have photons nearby.
   - **Alternative pairing check**: defining Z_a and Z_b as the mass-sorted alternative pairing Z candidates, with Z_a being the one closest to the nominal Z^0 boson mass, the ZZ candidate is excluded if $m_{Z_a} < 12$ GeV while m_{Z_a} is closer to the nominal Z_a boson mass than m_{Z_1} is. This cut discards 4μ and $4e$ candidates where the alternative pairing looks like an on-shell Z and a low-mass $\ell^+\ell^-$ resonance.
   - **Four-lepton invariant mass**: $m_{4\ell} > 70$ GeV.

6. The events that contain at least one selected ZZ candidate form the signal region.

7. If more than one ZZ candidate is left, the one with the highest value of $\mathcal{D}^{\text{kin}}_{\text{bkg}}$, a kinematic discriminant which will be defined in Section 4.4.2 is selected as *the best ZZ candidate*. When different ZZ candidates involve the same four leptons, they have identical values of $\mathcal{D}^{\text{kin}}_{\text{bkg}}$, and the candidate with the Z_1 mass closest to the nominal Z^0 boson mass is chosen.

4.2.7 Event selection efficiency

In order to understand how efficient the event workflow is, a study using the signal gluon fusion sample was performed in each of the final states. Results of the study are presented in Fig. 4.10, and show the efficiency of each selection step. About half of the 4ℓ events are found to have all four leptons in the detector acceptance. For electron final states, the least efficient step is the reconstruction, while for muons it is the isolation requirement. Since electrons are more difficult to reconstruct than muons and compete against much more background objects, the efficiency difference between the three final states widens as the selection progresses. While this study was performed on the 2016 simulation with the electron selection that includes identification and isolation as separate steps, an identical efficiency is expected from the 2017 analysis since the combined electron ID with isolation was tuned to exactly match the efficiency of previous analysis. The final efficiency of selecting gluon fusion signal events with the full analysis selection chain is found to be 22%, 40% and 29% for the $4e$, 4μ and $2e2\mu$ final states, respectively.

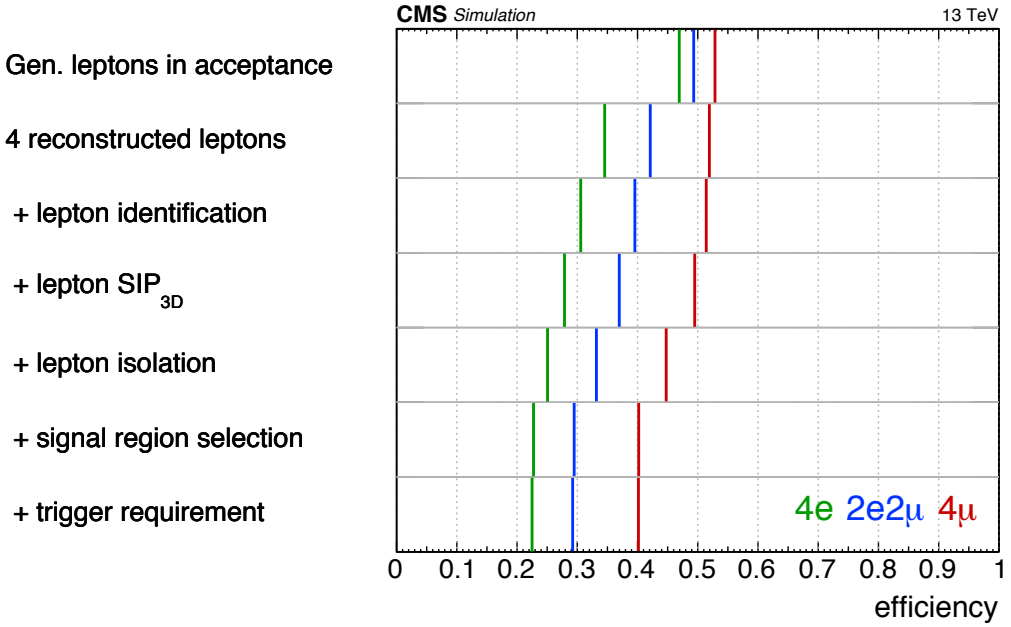

Figure 4.10 – Efficiencies of different steps of the analysis selection, using a gluon fusion simulated sample, comparing three subsets of events corresponding to the generated $4e$, $2e2\mu$ and 4μ final states. The successive steps include the efficiency for the four generated leptons to be in the η and p_{T} acceptance, the efficiency for the event to contain two pairs of opposite-sign same-flavour reconstructed leptons, all lepton and event selection steps, and the efficiency for the event to be triggered.

4.3 Event categorization

In order to study different Higgs boson production mechanisms, selected signal events are distributed in several categories. Idea is to define these categories using observables mainly from additional objects in the event, that are sensitive to different production modes. A first attempt was made already in Run I, with a definition of two categories simply based on the number of jets observed in the event. With such simple definition, it was already possible to perform measurement of Higgs boson couplings to vector bosons and to fermions. Motivated by significant increase in statistics, a more complex categorization has been developed for Run II, targeting all five main production modes. Selected signal events are distributed into several mutually exclusive categories, with each one being design to be enriched in a specific production mode.

In this section, a brief overview of production sensitive observables used in the categorization is given before discussing categorizations for 2016 and 2017 analysis, respectively.

4.3.1 Categorization observables

To successfully extract information about Higgs boson production mechanisms, additional objects in the event, other than the four selected leptons, have to be used. Selected jets defined in Section 4.2.5 and additional leptons defined as selected leptons that are not involved in the best ZZ candidate, are used. There are two different classes of observables used as inputs for event categorization: observables that count number of additional objects in the event and matrix-element based production discriminants that exploit lepton and jet kinematics.

Numbers of additional objects

The simplest observables that can give information about production arise from counting the additional objects. Four such observables were chosen:

- the number of selected jets, targeting production modes like VBF, VH that are expected to have one or two jetst in the event, and $\mathrm{t\bar{t}H}$ with more than three jets expected in the event,
- the number of selected b-tagged jets, to target $\mathrm{t\bar{t}H}$ and ZH with $\mathrm{Z} \to b\bar{b}$,
- the number of additional leptons, targeting leptonic VH and $\mathrm{t\bar{t}H}$ events,
- the number of additional pairs of opposite-sign same-flavour leptons, characteristic for leptonic ZH and fully leptonic $\mathrm{t\bar{t}H}$ events.

The discriminating power of these observables is illustrated in Fig. 4.11.

In addition to the observables that simply count numbers of objects in the event, observable sensitive to presence of neutrinos, missing transverse momentum E_T^{miss}, is used.

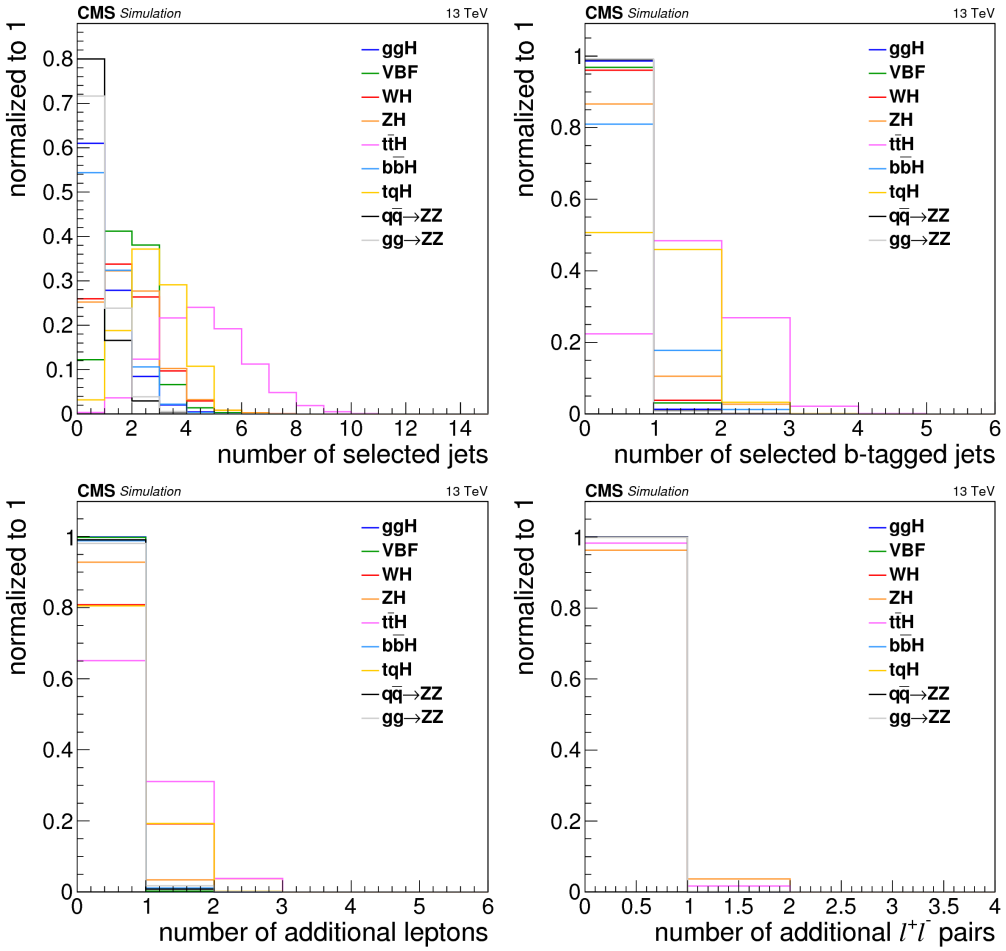

Figure 4.11 – Normalized distributions of the number of selected jets (top left), of selected b-tagged jets (top right) of additional leptons (bottom left), and of additional pairs of opposite-sign additional leptons (bottom right), for the main production mechanisms of the signal and the $gg \to ZZ$ and $q\bar{q} \to ZZ$ background processes, based on 2017 simulation. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

This observable targets VH events with neutrinos present in the decay of vector bosons.

Matrix-element production discriminants

To achieve additional separation power and thanks to the full reconstruction of the final state objects in the $H \to 4\ell$ analysis, kinematic observables such as lepton and jet four momenta and different angles in the event can be used. Instead of forming *kinematic discriminants* by combining sets of observables with the use of multivariate techniques, an approach based on matrix element calculations was developed. It uses kinematic observables from event as inputs, and SM Lagrangian to calculate matrix elements that are directly related to the probability of observing an event. Compared to the commonly used multivariate techniques, this approach has the advantage of extracting the informa-

tion in a transparent way that is directly related to the physics of the targeted process, and does not depend on a prior training, thus avoiding the associated caveats. The matrix-element computations are performed with the Matrix Element Likelihood Approach (MELA) package [22, 93, 94, 95], with JHUGEN is used for signal, and MCFM for background matrix elements. There are four matrix-element production discriminants used for event categorization in the $H \to ZZ^{\star} \to 4\ell$ analysis.

In events containing at least two selected jets, the discriminant sensitive to the VBF signal topology is defined as:

$$\mathcal{D}^{\rm ME}_{\rm VBF-2j} \equiv \left[1 + \frac{\mathcal{P}_{\rm H+JJ}(\mathbf{\Omega}^{\rm H+JJ}|m_{4\ell})}{\mathcal{P}_{\rm VBF}(\mathbf{\Omega}^{\rm H+JJ}|m_{4\ell})}\right]^{-1}, \tag{4.3}$$

where denominator is the probability for VBF Higgs boson production, and the numerator is the probability for ggH + 2 jets production. $\mathcal{P}_{\rm H+JJ}$ and $\mathcal{P}_{\rm VBF}$ probabilities for ggH + 2 jets and VBF are obtained from the JHUGEN matrix-elements [47, 54], and $\mathbf{\Omega}^{\rm H+JJ}$ denotes production related kinematic information associated to VBF and hadronic-VH candidate events that are fully described at leading order by five angles [95] illustrated in Fig. 4.12. If there are more than two selected jets in the event, they are ordered in $p_{\rm T}^{\rm jet}$ and the two leading ones are used in the calculation of the discriminant. Distributions of $\mathcal{D}^{\rm ME}_{\rm VBF-2j}$ are shown for various processes on the left in Fig. 4.13. This discriminant equally well separates VBF from the ggH + 2jets signal and from SM backgrounds with 2 jets in the event, because the latter processes have similar jet kinematics.

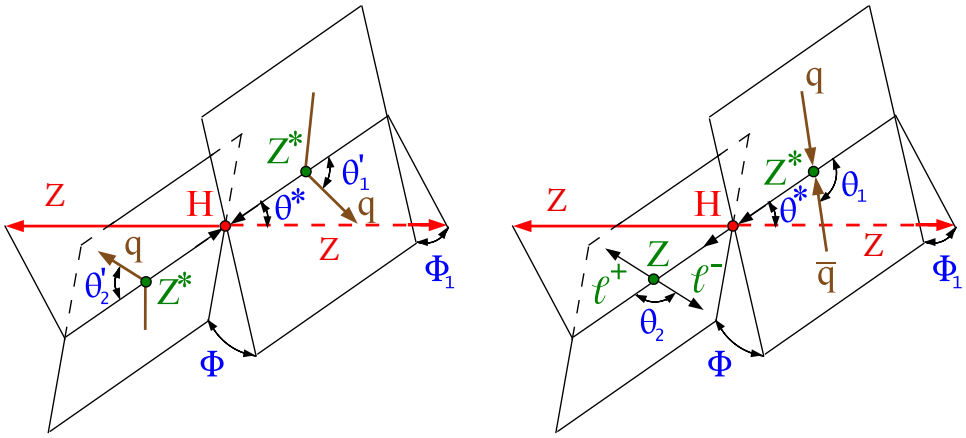

Figure 4.12 – Illustrations of the five angles that fully describe the kinematic configuration of the Higgs boson production chain at leading order in VBF (left) and hadronic-VH (right) processes. The θ^*, Φ_1, and Φ angles are defined in the H rest frame, $\theta_1^{'}$ and $\theta_2^{'}$ are defined in the laboratory frame, and θ_1 and θ_2 are defined in the V* and V rest frames, respectively.

Studies using simulation have shown that ~40% of the VBF events have less than two selected jets in the $H \to 4\ell$ analysis, resulting from a VBF jet that is out of the detector acceptance, or that it is not reconstructed, or fails the selection requirements defined in Section 4.2.5. A VBF signal probability can be constructed in events containing exactly one selected jet by simply integrating $\mathcal{P}_{\rm VBF}$ over the pseudorapidity of the unobserved

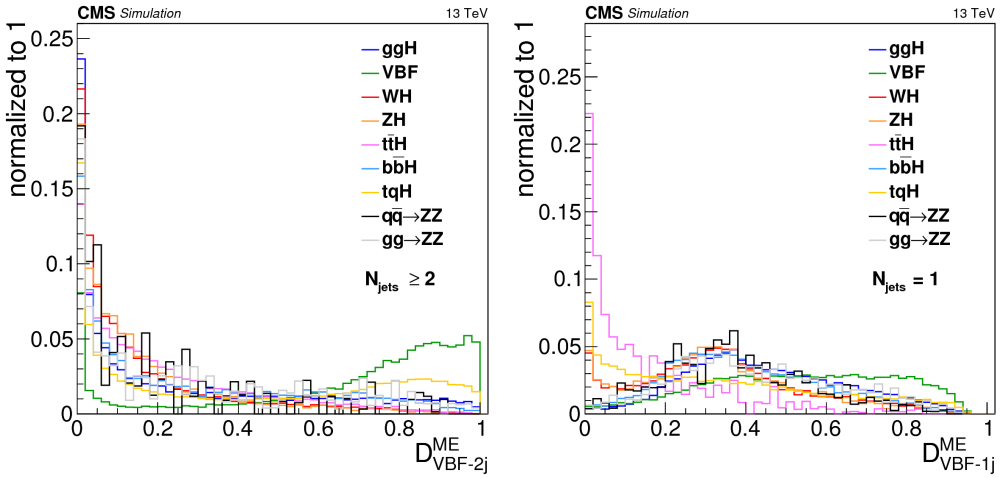

Figure 4.13 – Distributions of the matrix-element based production discriminants sensitive to VBF production $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}}$ (left) and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}}$ (right), in events with at least two or exactly one selected jet, respectively. Distributions are given for the main production mechanisms of the signal and the $\mathrm{gg} \to \mathrm{ZZ}$ and $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background processes, based on 2017 simulation. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

jet while constraining the transverse momentum of the 4ℓ + 2 jets system to be zero, resulting in a discriminant:

$$\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}} \equiv \left[1 + \frac{\mathcal{P}_{\mathrm{H+J}}(\boldsymbol{\Omega}^{\mathrm{H+J}}|m_{4\ell})}{\int \mathrm{d}\eta_{\mathrm{J}} \mathcal{P}_{\mathrm{VBF}}(\boldsymbol{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}\right]^{-1}. \tag{4.4}$$

Distributions of $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}}$ are shown for various processes on the right in Fig. 4.13. While having less separating power with only one jet information, it is still possible to separate the VBF production mechanism and gain information from events that contain only one selected jet.

In a similar manner, two discriminants separating hadronic VH processes from ggH + 2 jets are defined for events with at least two selected jets, as:

$$\mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}} \equiv \left[1 + \frac{\mathcal{P}_{\mathrm{H+JJ}}(\boldsymbol{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}{\mathcal{P}_{\mathrm{ZH-hadr.}}(\boldsymbol{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}\right]^{-1}, \tag{4.5}$$

$$\mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}} \equiv \left[1 + \frac{\mathcal{P}_{\mathrm{H+JJ}}(\boldsymbol{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}{\mathcal{P}_{\mathrm{WH-hadr.}}(\boldsymbol{\Omega}^{\mathrm{H+JJ}}|m_{4\ell})}\right]^{-1}, \tag{4.6}$$

where $\mathcal{P}_{\mathrm{ZH-hadr.}}$ and $\mathcal{P}_{\mathrm{WH-hadr.}}$ are again computed using JHUGEN matrix elements for the ZH and WH process and ggH + 2 jets.

Distributions of $\mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}}$ and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}$ production discriminants are shown for various processes in Fig. 4.14. Since hadronic VH production does not have such a clean detector signature like VBF it is not easy to separate it from the dominant ggH in association with two jets. However, one can see that from the constructed kinematic discriminant moderate discriminating power is achieved. Additional studies using simulation

showed that hadronic VH events with one lost jet exhibit little kinematic information and are therefore not exploited in matrix elements.

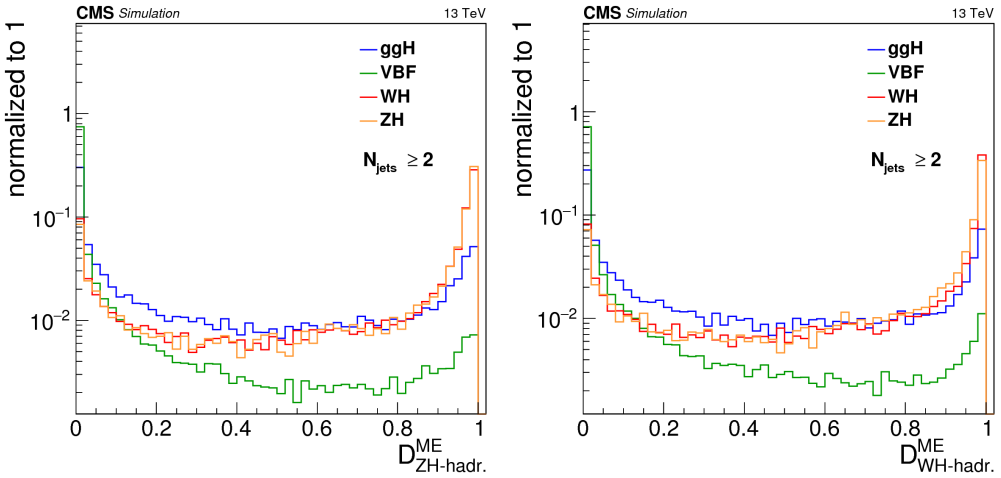

Figure 4.14 – Distributions of the matrix-element based production discriminant $\mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}}$ sensitive to hadronic ZH production (left), and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}$ (right) sensitive to WH hadronic production, in events with at least two selected jets. Distributions are given for the main production mechanisms of the signal, based on 2017 simulation. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

Matrix element probabilities for leptonic VH processes have also been studied, but they are not exploited in the analysis because the sole presence of additional selected leptons was found to provide enough information to extract such processes. In the case of $\mathrm{t\bar{t}H}$ production mode, matrix elements calculates become extremely complicated due to increased combinatorics of additional final state objects. The studies are ongoing to include this kind of discriminant in future iterations of the analysis.

4.3.2 Event categorization for 2016 analysis

This section describes the version of the event categorization that was used in the public results of the 2016 analysis reported in Section 5.1.

In order to design pure event categories, numerous baskets of selected events are defined according to all possible combinations of the values for categorization observables are defined. Afterwards, the composition of every basket is evaluated in terms of relative fractions of production processes. For instance, baskets using a cut on $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}}$ are dominated by VBF, while $\mathrm{t\bar{t}H}$ tends to populate baskets featuring b-tagged jets and/or additional leptons. Finally, the baskets are merged with one another, so as to form only one event category per process of interest.

Using the following criteria applied in this exact order, meaning an event is considered for the subsequent category only if it does not satisfy the requirements of the previous category, seven categories are defined for 2016 analysis:

- The **VBF-2jet-tagged category** requires exactly four leptons. In addition, there must be either two or three jets of which at most one is b-tagged, or four or more jets none of which are b-tagged. Finally, $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}} > 0.5$ is required.

- The **VH-hadronic-tagged category** requires exactly four leptons. In addition, there must be two or three jets, or four or more jets none of which are b-tagged. $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VH-hadr.}} \equiv \max(\mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}}, \mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}) > 0.5$ is required.

- The **VH-leptonic-tagged category** requires no more than three jets and no b-tagged jets in the event, and exactly one additional lepton or one additional pair of OS, same-flavor leptons. This category also includes events with no jets and at least one additional lepton.

- The **$\mathrm{t\bar{t}H}$-tagged category** requires at least four jets of which at least one is b-tagged, or at least one additional lepton.

- The **VH-$\mathrm{E}^{\mathrm{miss}}_{\mathrm{T}}$-tagged category** requires exactly four leptons, no more than one jet and E_T^{miss} greater than 100 GeV.

- The **VBF-1jet-tagged category** requires exactly four leptons, exactly one jet and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}} > 0.5$.

- The **Untagged category** consists of the remaining selected events.

Since VBF extraction relies on two very different discriminants $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}}$ and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}}$ it was decided to have both a VBF-2jet-tagged and a VBF-1jet-tagged category instead of a single VBF-tagged one. Moreover, instead of having a WH-tagged and a ZH-tagged category, a VH-hadronic-tagged one and a VH-leptonic-tagged one are defined. As one can see in Fig. 4.14, $\mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}$ and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}}$ are so similar that they do not really help separating WH from ZH. That is why they are combined in a $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VH-hadr.}}$ discriminant that is used to select hadronic-VH events, and jets and additional leptons are used to select leptonic-VH events. The discriminants were tuned so that the cuts at 0.5 have a good compromise between purity and expected yield in the corresponding categories. The final definitions of the categories were chosen to achieve high signal purity whilst maintaining high efficiency for each of the main Higgs boson production mechanisms. The order of the categories is chosen to maximize the signal purity target in each category.

Figure 4.15 reports the relative contributions of every signal process in the seven event categories together with the expected event yields for 2016 analysis. The VBF-1jet-tagged and VH-hadronic-tagged categories are expected to have substantial contamination from gluon fusion, while the purity of the VBF process in the VBF-2jet-tagged category is expected to be about 49%. Less than one event was expected in categories targeting VH and $\mathrm{t\bar{t}H}$ production, but this categorization scheme is being build for analysis of full Run II data sample which is expected to have around 100 fb^{-1} of data. Figure 4.16 shows how the various signal processes are individually distributed among the seven categories.

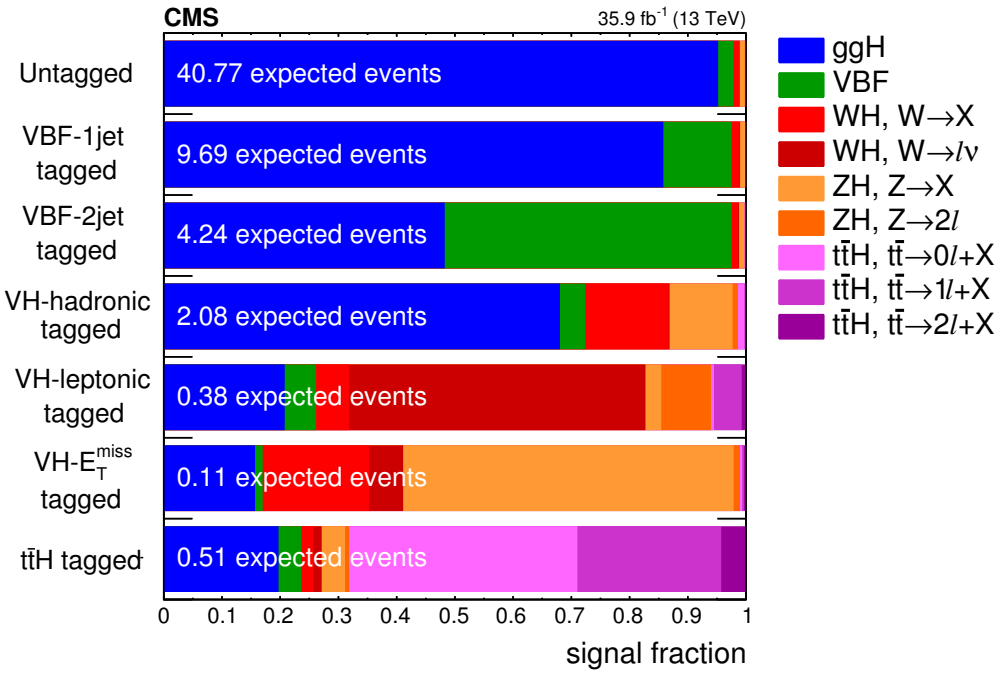

Figure 4.15 – Signal process composition of the seven event categories, for the categorization of the 2016 analysis. The VH and $\mathrm{t\bar{t}H}$ processes are split according to the generated decay of the associated W, Z, or $t\bar{t}$ system, whereby the X symbol stands for a set of any other particles than electrons or muons. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

4.3.3 Event categorization for 2017 analysis

In this section, a version of the event categorization that was used in the preliminary public results of the 2017 analysis reported in Section 5.2 is described.

Because data set of 2017 analysis was only slightly larger than the one for the 2016 analysis it was decided to take the same approach for event categorization. However, there are several changes that need to be mentioned. First improvement is a direct consequence of the improved jet selection, described in Section 4.2.5 and summarized in Table 4.7, coming from more efficient jet identification and b-tagging, and from the introduction of the jet pileup identification. With the improvement of selecting real jets and b-tagged jets over background it translated to slightly higher purities in categories targeting VBF, VH, and $\mathrm{t\bar{t}H}$. Furthermore, a study performed on a simulation showed that by a simple splitting of the $\mathrm{t\bar{t}H}$-tagged category into its hadronic and leptonic counterparts, it is possible to slightly improve sensitivity to the $\mathrm{t\bar{t}H}$ production mode. On top of that, with many more data to be collected in the future, it will be possible to further optimize bosonic and leptonic signatures of this production mode similar to the VH.

Finally, careful checks of all categorization observables showed a big discrepancy between data and simulation for the missing transverse momentum observable. Further

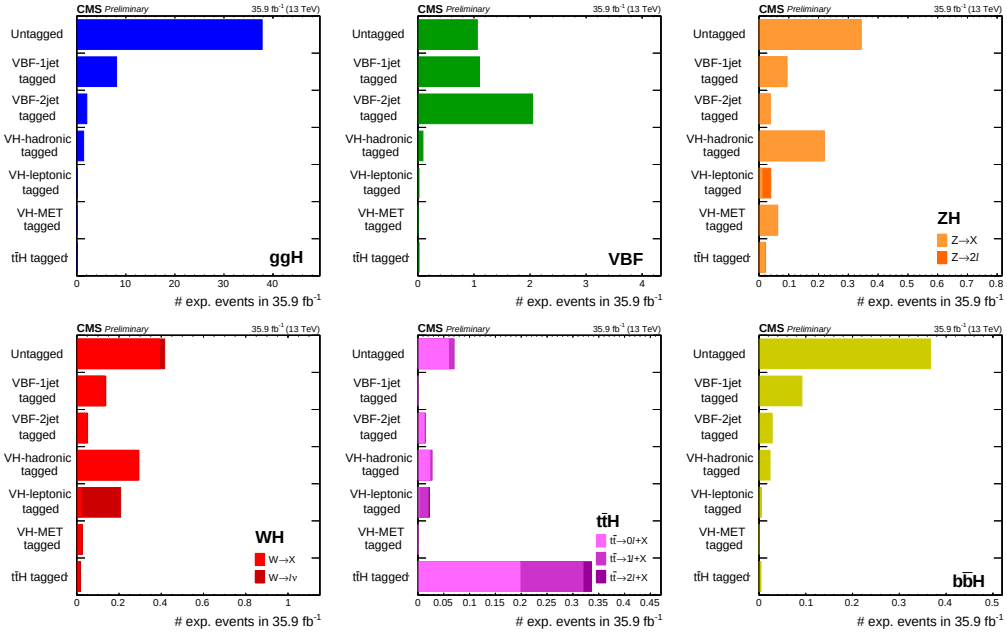

Figure 4.16 – Distributions of ggH (top left), VBF (top middle), ZH (top right), WH (bottom left), $\mathrm{t\bar{t}H}$ (bottom middle), and $\mathrm{b\bar{b}H}$ (bottom right) signal processes in the seven categories, for the categorization of the 2016 analysis. The VH and $\mathrm{t\bar{t}H}$ processes are split according to the generated decay of the associated W, Z, or $t\bar{t}$ system, whereby the X symbol stands for a set of any other particles than electrons or muons. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

studies showed that the origin of the discrepancy is high electronic noise in the very high η regions of the ECAL endcap. This effect is very challenging to correct for and additional work is needed to derive JEC to mitigate it. Since the analysis of 2017 data was planned to be a preliminary result using prompt reconstructed data and definition of a single category is the only place in the whole analysis where E_T^{miss} observable is used, it was decided to drop the VH-$\mathrm{E_T^{miss}}$-tagged category for the 2017 definition of categorization.

Finally, for 2017 analysis, seven categories are defined as follows:

- The **VBF-2jet-tagged category** requires exactly four leptons. In addition, there must be either two or three jets of which at most one is b-tagged, or four or more jets none of which are b-tagged. Finally, $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}} > 0.5$ is required.
- The **VH-hadronic-tagged category** requires exactly four leptons. In addition, there must be two or three jets, or four or more jets none of which are b-tagged. $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VH-hadr.}} \equiv \max(\mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}}, \mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}) > 0.5$ is required.
- The **VH-leptonic-tagged category** requires no more than three jets and no b-tagged jets in the event, and exactly one additional lepton or one additional pair of OS, same-flavor leptons. This category also includes events with no jets and at

least one additional lepton.

- The **$\mathrm{t\bar{t}H}$-hadronic-tagged category** requires at least four jets of which at least one is b-tagged and there are no additional leptons in the event.
- The **$\mathrm{t\bar{t}H}$-leptonic-tagged category** requires at least one additional lepton in the event.
- The **VBF-1jet-tagged category** requires exactly four leptons, exactly one jet and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}} > 0.5$.
- The **Untagged category** consists of the remaining selected events.

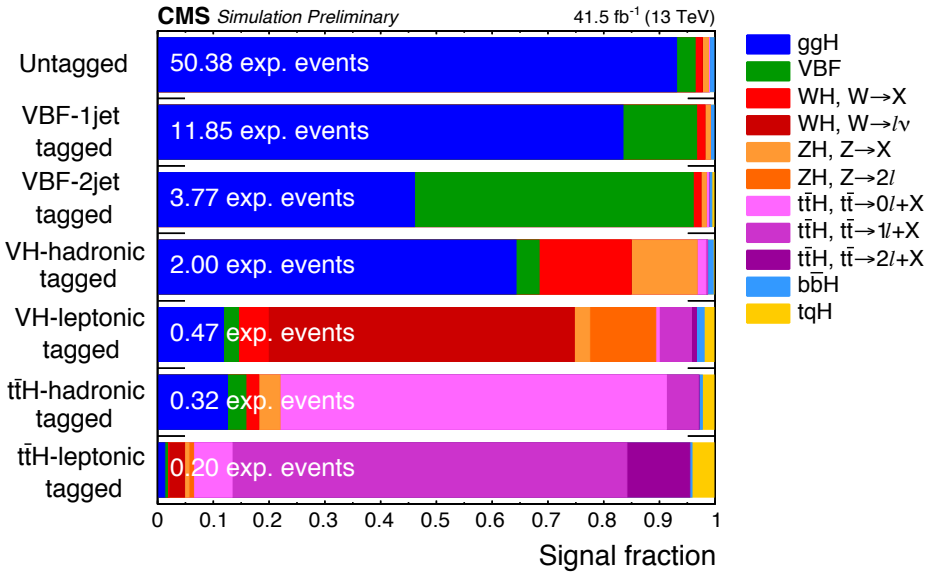

Figure 4.17 – Signal process composition of the seven event categories, for the categorization of the 2017 analysis. The VH and $\mathrm{t\bar{t}H}$ processes are split according to the generated decay of the associated W, Z, or $t\bar{t}$ system, whereby the X symbol stands for a set of any other particles than electrons or muons. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

Figure 4.17 reports the relative contributions of every signal process in the seven event categories together with the expected event yields for 2017 analysis. Due to improvements in the jet selection, events are expected to migrate form categories with higher jet multiplicity to those with lower. For example VBF-2jet-tagged, VH-hadronic-tagged, and $\mathrm{t\bar{t}H}$-tagged categories have lower yields with slightly higher purities in the production modes they are designed to target. The purity of the $\mathrm{t\bar{t}H}$ process in the newly defined $\mathrm{t\bar{t}H}$-leptonic-tagged category is expected to be about 87%. Figure 4.18 shows how the various signal processes are individually distributed among the seven categories.

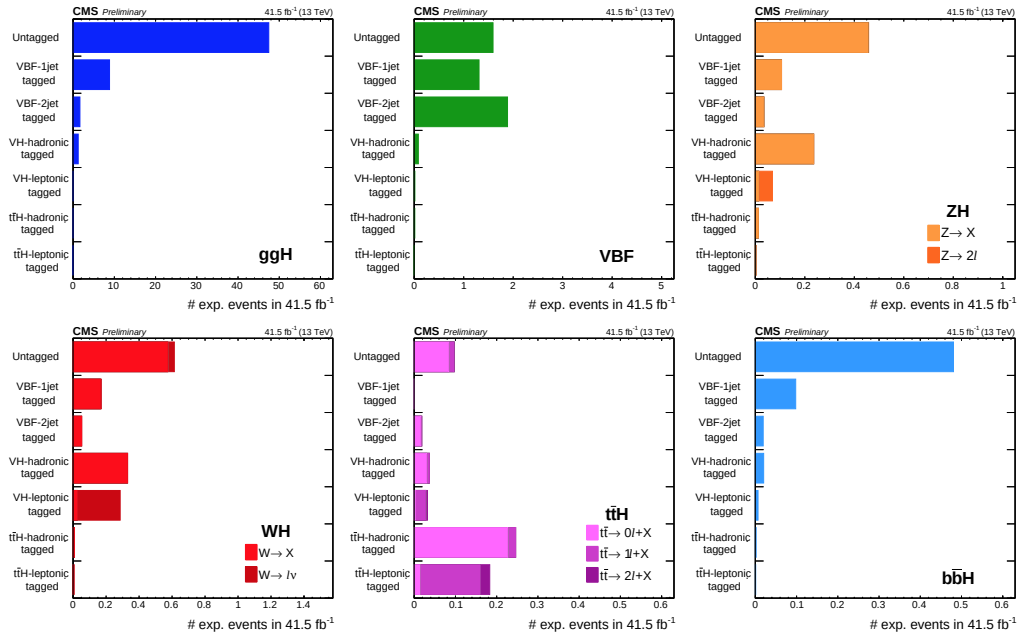

Figure 4.18 – Distributions of ggH (top left), VBF (top middle), ZH (top right), WH (bottom left), $\mathrm{t\bar{t}H}$ (bottom middle), and $\mathrm{b\bar{b}H}$ (bottom right) signal processes in the seven categories, for the categorization of the 2017 analysis. The VH and $\mathrm{t\bar{t}H}$ processes are split according to the generated decay of the associated W, Z, or $t\bar{t}$ system, whereby the X symbol stands for a set of any other particles than electrons or muons. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

4.4 Observables

In order to measure some physics parameters, or to test models of interest, one has to first find observables that are sensitive to it. These observables should be able to discriminate defined signal from backgrounds and be sensitive to the measured parameters. We have already discussed observables that are use to cut on when discussing lepton and event selection, or event categorization. In this section, a different kind of observables that are used to extract results in the statistical analysis is discussed.

4.4.1 Four-lepton invariant mass

A simple, yet powerful observable that is able to well distinguish Higgs boson production over the SM background is the four-lepton invariant mass $m_{4\ell}$. The Higgs boson signal appears as a resonance peak around 125 GeV over the almost flat SM backgrounds. It is used in every measurement in the golden channel, as well as a form of presenting the results with the famous four-lepton invariant mass distribution which will be shown for different data periods in Chapter 5. The $m_{4\ell}$ observable relies strongly on the lepton momentum resolution, and thanks to the impressive performance of the CMS detector and additional fine tuning and offline corrections to data, unprecedented precision on the Higgs boson mass measurement was achieved as will also be discussed in Chapter 5.

4.4.2 Matrix element kinematic discriminants

Similar to the matrix-element based production discriminants for categorization, discussed in Section 4.3.1, one can define similar type of observables for statistical analysis. There are two types of kinematic discriminants that are used for fits in the $\mathrm{H} \to 4\ell$ analysis:

- *Decay only kinematic discriminant* $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$: based on decay only input variables and designed to provide separation between dominant signal gluon fusion production mode and dominant $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ SM background.
- *Production and decay kinematic discriminants* $\mathcal{D}^{\mathrm{VBF+dec}}_{\mathrm{bkg}}$ and $\mathcal{D}^{\mathrm{VH+dec}}_{\mathrm{bkg}}$: based on decay and production input variables and designed to provide additional separation between VH or VBF production mechanisms from the gluon fusion while keeping the separation power from dominant $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ SM background.

Decay only kinematic discriminant

The main building block of the $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$ observable is the fact that the kinematic of the four-lepton decay of the Higgs boson is different from that of the dominant $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background. Or in other words, the four-momenta of the four selected leptons can be used to calculate the probability for event to be signal or background like. It can be shown that eight independent degrees of freedom fully describe $\mathrm{H} \to 4\ell$ configuration in the centre-of-mass frame of the 4ℓ system, up to an arbitrary rotation around the beam axis [93, 94, 95]. A possible set of eight variables is given by:

- $m_{4\ell}$, the four-lepton invariant mass,
- m_{Z_1} and m_{Z_2}, the invariant mass of two Z candidates,
- θ^*, the angle between the beam axis and the Z_1 direction in the 4ℓ rest frame,
- Φ and Φ_1, the two azimuthal angles between the three planes defined by the H boson, Z_1, and Z_2 decay products in the 4ℓ rest frame,
- θ_1 and θ_2, the angles between the Z_1 and Z_2 boost directions and the momenta of their daughter particles in the Z_1 and Z_2 rest frames, respectively.

The five angles mentioned are illustrated on the left in Fig. 4.19. Other seven variables than $m_{4\ell}$ are collectively referred to as $\mathbf{\Omega}^{\mathrm{H}\to 4\ell}$. A decay only kinematic discriminant that is sensitive to Higgs decay is then defined as:

$$\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}} \equiv \left[1 + \frac{\mathcal{P}^{\mathrm{q\bar{q}}}_{\mathrm{bkg}}(\mathbf{\Omega}^{\mathrm{H}\to 4\ell}|m_{4\ell})}{\mathcal{P}^{\mathrm{gg}}_{\mathrm{sig}}(\mathbf{\Omega}^{\mathrm{H}\to 4\ell}|m_{4\ell})}\right]^{-1}, \tag{4.7}$$

where the denominator is the probability for the Higgs boson signal decay to four leptons, and the numerator is the probability for the dominant $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background decay

process, calculated from the LO matrix element squared with JHUGEN and MCFM, respectively [22, 49].

As it is obvious from the right plot in Fig. 4.19, the $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ observable provides great discriminating power between all main Higgs production modes and dominant $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ SM background. However, since it uses decay only information of the four leptons, it provides no discrimination power for different production mechanisms.

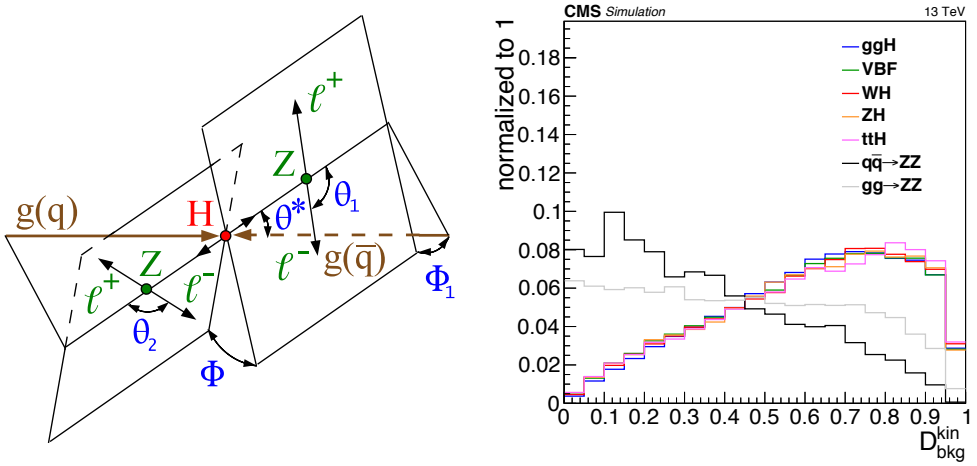

Figure 4.19 – Illustrations of the five angles (left) that characterize the orientation of Higgs boson production and decay as $\mathrm{gg/q\bar{q}} \to \mathrm{H} \to \mathrm{ZZ} \to 4\ell$. The θ^*, Φ_1, and Φ angles are defined in the Higgs rest frame, whereas θ_1 and θ_2 are defined in the Z_1 and Z_2 rest frames, respectively. Distribution of the decay kinematic discriminant $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ (right) for the main production mechanisms of the H(125) signal and for the $\mathrm{gg} \to \mathrm{ZZ}$ and $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background processes. Events are required to pass the full analysis selection and to be in a $118 < m_{4\ell} < 130$ GeV window.

Production and decay kinematic discriminants

Since the $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ observable provides no discriminating power between different production modes, interplay between the expected contributions of different production mechanisms and the number of observed events in each category was the only way to extract information about Higgs boson production modes. This has motivated a development of new kinematic discriminants that combine decay production already used to build $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ and production information to become sensitive to different Higgs boson production mechanisms. A set of twelve variables that fully describe production of Higgs boson and its decay to four leptons is collectively referred to as $\boldsymbol{\Omega}$ and it consists of seven decay variables $\boldsymbol{\Omega}^{\mathrm{H}\to 4\ell}$ and five production angles $\boldsymbol{\Omega}^{\mathrm{H+JJ}}$.

Targeting VBF and VH production modes, two new production and decay kinematic

discriminants are defined as:

$$\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}} \equiv \left[1 + c^{\text{VBF2jet}}(m_{4\ell}) \cdot \frac{\mathcal{P}_{\text{bkg}}^{\text{VBS+VVV}}(\mathbf{\Omega}|m_{4\ell}) + \mathcal{P}_{\text{bkg}}^{\text{QCD+dec}}(\mathbf{\Omega}|m_{4\ell})}{\mathcal{P}_{\text{sig}}^{\text{VBF+VH+dec}}(\mathbf{\Omega}|m_{4\ell})}\right]^{-1}, \quad (4.8)$$

$$\mathcal{D}_{\text{bkg}}^{\text{VH+dec}} \equiv \left[1 + c^{\text{had.-VH}}(m_{4\ell}) \cdot \frac{\mathcal{P}_{\text{bkg}}^{\text{VBS+VVV}}(\mathbf{\Omega}|m_{4\ell}) + \mathcal{P}_{\text{bkg}}^{\text{QCD+dec}}(\mathbf{\Omega}|m_{4\ell})}{\mathcal{P}_{\text{sig}}^{\text{VBF+VH+dec}}(\mathbf{\Omega}|m_{4\ell})}\right]^{-1}, \quad (4.9)$$

where $\mathcal{P}_{\text{sig}}^{\text{VBF+VH+dec}}$ is the probability for VBF and VH signal, $\mathcal{P}_{\text{bkg}}^{\text{VBS+VVV}}$ is the probability for vector boson scattering and tri-boson background, $\mathcal{P}_{\text{bkg}}^{\text{QCD+dec}}$ is the probability for QCD production background, and $c^{\text{VBF2jet}}(m_{4\ell})$ and $c^{\text{had.-VH}}(m_{4\ell})$ are the $m_{4\ell}$-dependent constants used to calibrate the distribution in the VBF-2jet-tagged and VH-hadronic-tagged categories, respectively.

Impressive performance of the new discriminants is illustrated in Fig. 4.20. Top plots show that thanks to combining production with decay information, in their dedicated categories, $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ and $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ outperform the decay only discriminant $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ in separating VBF and VH signals from the dominant $\text{q}\bar{\text{q}} \to \text{ZZ}$ SM background. Furthermore, bottom plots illustrate that while $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ is completely insensitive to VBF or VH production over the ggH, as expected, the new discriminants provide good separation power owing to the use of production related information. Therefore, newly developed production and decay kinematic discriminants were adopted in the 2017 analysis, and their performance and improvement gain will be presented when discussing results in Chapter 5.

4.4.3 Per-event mass uncertainties

Used for the mass measurement, the per-event mass relative uncertainty is defined as:

$$\mathcal{D}_{\text{mass}} \equiv \frac{\sigma_{m_{4\ell}}}{m_{4\ell}}, \quad (4.10)$$

where $\sigma_{m_{4\ell}}$ is the uncertainty on the measurement of the four-lepton invariant mass $m_{4\ell}$. This observable will give higher importance to events with smaller uncertainty on the reconstructed $m_{4\ell}$ and a smaller one to those events with high uncertainty, helping to improve the overall mass measurement and reduce its uncertainty.

For muons, the individual uncertainty on momentum measurement is predicted from the full error matrix that is obtained from the muon track fit. For electrons, the momentum error is estimated from the combination of the ECAL and tracker measurement, neglecting the uncertainty on the track direction from the GSF fit. For FSR photons, momentum error is taken from the PF parametrization. The uncertainty $\sigma_{m_{4\ell}}$ is then calculated for each event, by propagating the uncertainty of the each individual lepton momentum measurements. Two approaches were studied. A simple approach is to calculate the contribution of the transverse momentum of each individual lepton or FSR photon to the overall $m_{4\ell}$ uncertainty separately, and calculate $\sigma_{m_{4\ell}}$ as the sum in quadrature of these contributions. In this approach all directional errors are neglected. The other way

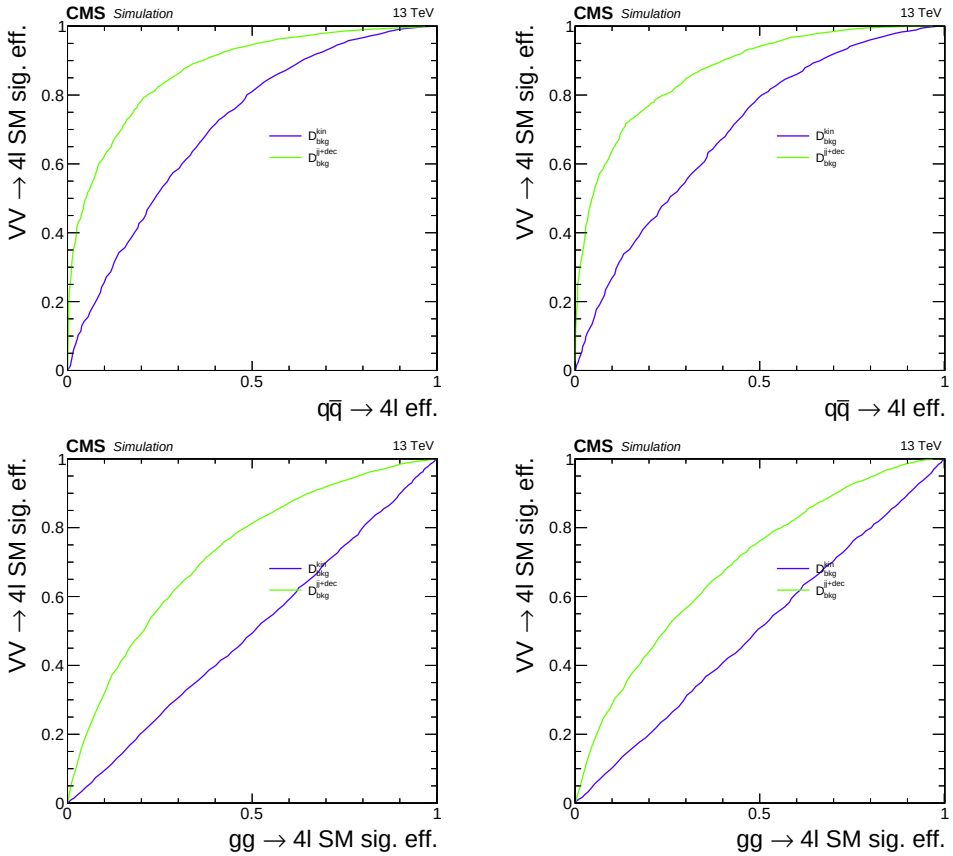

Figure 4.20 – Top left: Comparison of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VBF+dec}}$ and $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ in the VBF-2jet tagged category in terms of efficiency for the VBF signal and $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background. Top right: Comparison of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VH+dec}}$ and $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ in the VH-hadronic tagged category in terms of efficiency for the VH signal and $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background. Bottom left: Comparison of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VBF+dec}}$ and $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ in the VBF-2jet tagged category in terms of efficiency for the VBF and ggH. Top right: Comparison of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VH+dec}}$ and $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ in the VH-hadronic tagged category in terms of efficiency for the VH and ggH. Events are required to pass the full analysis selection and to be in a $105 < m_{4\ell} < 140$ GeV window.

is by using an analytical error propagation including all correlations. These methods have shown to agree at a <1% level.

When studying the distributions of the $\mathcal{D}_{\mathrm{mass}}$ observable in data and simulation, some discrepancies were observed. Even with the application of lepton momentum calibrations described in Section 4.2.3 that provide a good agreement on the transverse momentum, some discrepancies are till left on their uncertainties and additional corrections are needed.

This is done by correcting the lepton-level momentum errors both in data and simulation, using $\mathrm{Z} \to \mathrm{e^+e^-}$ and $\mathrm{Z} \to \mu^+\mu^-$ event samples. Similar like for the lepton efficiency measurements, the lineshape of these resonances is modelled with a a Breit-Wigner function convolved with a double Crystal Ball plus exponential function. Events are divided

into categories based on the p_{T}^{ℓ} and η^{ℓ} of the leptons, and the dilepton mass resolution predicted by the above technique is compared to the resolution extracted from a fit to the dilepton mass in each category, so that a correction to lepton momentum errors can be extracted by an iterative procedure. After the corrections are derived, a closure test of the agreement between the predicted and fitted 4ℓ mass resolution is performed in data and in simulation, in bins of the predicted $m_{4\ell}$ resolution, confirming that the calibration brings it close to the fitted one.

4.5 Background estimation

Without a proper understanding of all possible backgrounds it is impossible to have a reliable way of extracting an interesting signal. One has to ensure that every possible background is considered, and that its prediction is reliable with a correctly estimated uncertainty assigned to it. There are two different kinds of backgrounds that need to be considered in particle physics:

- **Irreducible backgrounds** are all backgrounds coming from physics processes that have exactly the same particles in the final state as the targeted signal.
- **Reducible backgrounds** are all backgrounds coming from physics processes that do not have exactly the same particles in the final state as the targeted signal. However, due to non-perfect detector performance and analysis techniques, they get miss-reconstructed and selected with final state particles matching the targeted signal.

Due to their intrinsic differences they are estimated separately, with different techniques which will be discussed in detail.

4.5.1 Irreducible background estimation

It is impossible to differentiate signal from irreducible background on event basis, meaning we will never be able to say for a single event if it is a Higgs boson production or a SM background process. However, on a larger number of events, and with the use of observables that show different behavior for signal and background it is possible to conclude if there is a signal process occurring amongst the expected background processes. There is a strong implication here that the result will be as reliable as your background estimate.

In the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ there are two main irreducible backgrounds coming from non-resonant gluon fusion $\mathrm{gg} \to \mathrm{ZZ}$, and quark-antiquark annihilation $\mathrm{q\bar{q}} \to \mathrm{ZZ}$. In both cases pairs of Z bosons are produced directly, without an intermediate production of the Higgs boson, and further decay to four leptons. Both irreducible backgrounds are estimated using simulation to obtain their expected yields as well as kinematic distributions.

As already mentioned in Section 4.1.2, there are some missing higher order corrections in simulated samples of these backgrounds. In order to account for those, the *k-factors* are applied as a function of generated mass $\mathrm{m_{ZZ}}$.

The $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background sample is generated at NLO in pQCD, but a full computation of its differential cross section exists at NNLO [105], although it is not yet available in an event generator. Therefore, a NNLO/NLO QCD k-factor is applied, based on cross sections obtained using the same PDF, renormalization, and factorization scales as the NLO sample. This k-factor is applied as a function of $\mathrm{m_{ZZ}}$ and varies between 1.0 and 1.2 as illustrated on left in Fig. 4.21. An additional NLO/LO electroweak correction is applied to the $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background as a function of the initial state quark flavour and kinematics. This correction is applied in the region $\mathrm{m_{ZZ}} > 2m_{\mathrm{Z}^0}$ for which it has been computed, and the average effect is a decrease of the predicted yield by about 2% at 200 GeV, and 14% at 1 TeV, as illustrated on the right in Fig. 4.21. It is not relevant for the work of this thesis, but it is an important piece in the search for additional high mass resonances in the four lepton final state.

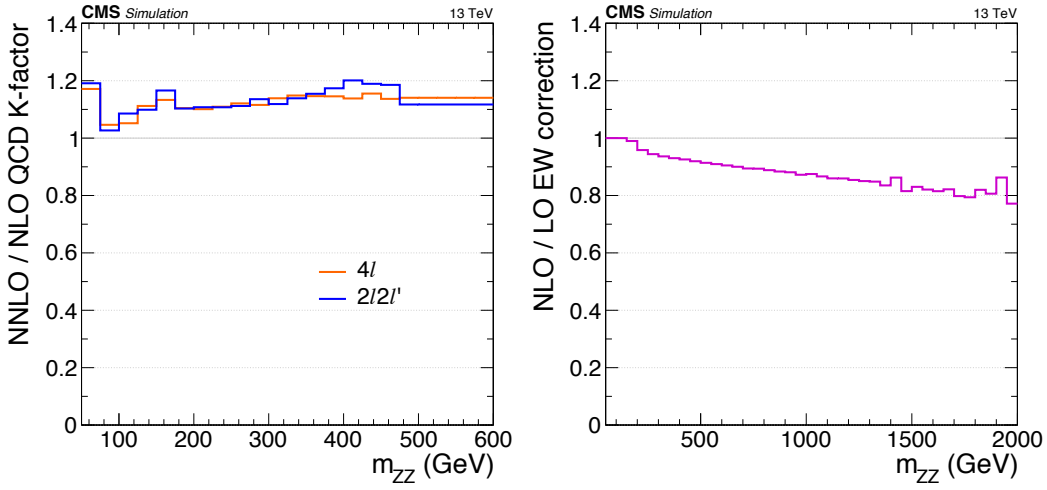

Figure 4.21 – NNLO/NLO QCD k-factor for the $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background (left), as a function of generated $\mathrm{m_{ZZ}}$, for the 4ℓ and $2\ell 2\ell'$ final states. Average impact of the NLO/LO electroweak correction for the $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background on its yields (right), as a function of generator-level $\mathrm{m_{ZZ}}$.

The $\mathrm{gg} \to \mathrm{ZZ}$ background sample is generated at LO in pQCD. Although no exact calculation exists beyond the LO, it has been recently shown [106] that the soft collinear approximation is able to describe the background cross section and the interference term at NNLO. Further calculations also show that the k-factors are very similar at NLO for the signal and background [107] terms, and at NNLO for the signal and interference terms [108], suggesting that the same k-factor can be used for the $\mathrm{gg} \to \mathrm{H} \to \mathrm{ZZ}$ signal and for the $\mathrm{gg} \to \mathrm{ZZ}$ background [109]. This analysis relies on a NNLO/LO k-factor computed for the signal process as a function of $\mathrm{m_{ZZ}}$, using the `HNNLO V2` program [110, 111, 97]. The NNLO and LO $\mathrm{gg} \to \mathrm{H} \to 2\ell 2\ell'$ cross sections, that are illustrated on the left in Fig. 4.22, are computed for a fixed small H boson decay width of

4.07 MeV across the whole m_{ZZ} range (Fig. 5.17, left), and the k-factor is defined as their ratio. It varies from about 2.0 to 2.6 and is 2.27 at $m_{ZZ} = 125$ GeV, as shown on the right plot in Fig. 4.22, stressing the importance of these additional corrections.

The $m_{4\ell}$ histogram templates that are obtained from simulation, and used to present results, are not used directly in the measurements. The analytical shapes are used instead, to smooth out the irregularities due to the finite number of simulated events. An order two Bernstain polynomial was found to describe well both irreducible backgrounds in the signal region defined around 125 GeV. Fits are performed for each final state in each category. For categories with very low expected yields, shapes obtained from the fit in the untagged category are used instead.

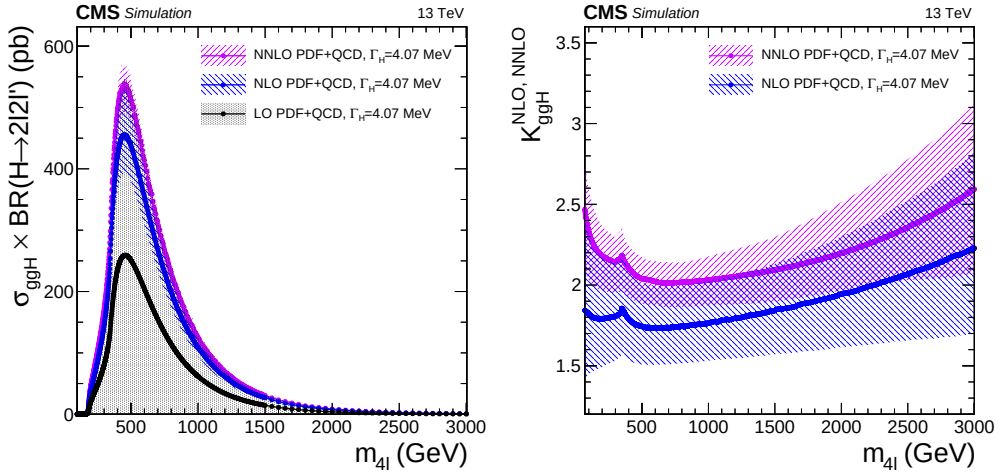

Figure 4.22 – Cross sections for $gg \to H \to 2\ell 2\ell'$ at NNLO, NLO and LO as a function of m_H (left), at the fixed small decay width of 4.07 MeV. Correspondings NNLO/LO and NLO/LO k-factors (right). The former is applied to the $gg \to ZZ$ background, as a function of generator-level m_{ZZ}. Uncertainties on the PDF set and QCD scales are displayed as hatched bands.

4.5.2 Reducible background estimation

With a perfect detector performance, no events coming from reducible background would ever end up as selected ZZ candidates in the $H \to ZZ^\star \to 4\ell$ analysis. A first step in order to mitigate the detector imperfections that cause rise to such events because of misidentification of particles is done with a series of selection cuts. These cuts are designed to reject as much of the reducible background as possible while keeping the signal efficiency high. The selection process was described in detail for electrons in Chapter 3 and a summary for electrons and muons was presented in Section 4.2.2. While this step reduces the amount of reducible background that enters final event selection, there is still a significant contribution left that needs to be accounted for.

Reducible background in the $H \to ZZ^\star \to 4\ell$ analysis originates from events where at least one lepton is not a genuine signal lepton coming from a Z boson. There are three different types of reducible background sources: light flavour hadrons misidentified as

leptons, heavy flavour jets producing secondary leptons through decay of heavy flavour meson, and conversions. Main SM processes that contribute to it are $\mathrm{Z} + \mathrm{jets}$, $\mathrm{t\bar{t}} + \mathrm{jets}$, $\mathrm{Z}\gamma + \mathrm{jets}$, $\mathrm{WZ} + \mathrm{jets}$, and $\mathrm{WW} + \mathrm{jets}$, ordered in importance. All reducible backgrounds in the $\mathrm{H} \to 4\ell$ analysis are collectively denoted as $\mathrm{Z} + \mathrm{X}$, and are calculated using two independent methods based on data rather than simulation. There are no fundamental problems in using the simulation instead, however several physics processes of $\mathrm{Z} + \mathrm{X}$ contribution are not simulated in samples large enough to correctly populate the four-lepton signal region, let alone categories with low expected number of events. Furthermore, simulation also relies on higher order effects, such as the modelling of additional jets, which is not simulated accurately enough yet.

Because of the above mentioned reasons, it is a common practice in many particle physics analysis to estimate the reducible background yields and kinematics from *control regions* defined in data. A control region is a region of the phase space that is constructed to be orthogonal to the signal region and more populated than it.

Both the *opposite sign method* and the *same sign method* define control regions to estimate the *fake rate* which is later applied in a different control region to estimate the $\mathrm{Z} + \mathrm{X}$ background. All parts of this workflow will be discussed and differences between the methods will be described.

Control region definition

There are two main kinds of control regions used in the $\mathrm{Z} + \mathrm{X}$ estimation: $\mathrm{Z} + \mathrm{L}$ and $\mathrm{Z} + \mathrm{LL}$, where in both cases Z denotes a Z candidate, and L denotes a *loose lepton*. A loose lepton is defined as a lepton that passes kinematic and impact parameter cuts as defined in Section 4.2.2. A set of all loose leptons consists of two orthogonal sets: *passing leptons* P, defined as leptons that pass the full lepton selection defined in Section 4.2.2, and *failing leptons* F defined as loose leptons that fail isolation or identification cut, or both.

Both methods use the same $\mathrm{Z} + \mathrm{L}$ control region to estimate the fake rate. It consists of events with a Z candidate, and in addition, there is exactly one loose lepton in the event. Transverse momentum cuts are applied to leptons from the Z candidate asking $p_{\mathrm{T}}(\ell_1) > 20$ GeV and $p_{\mathrm{T}}(\ell_2) > 10$ GeV. The QCD suppression cut $m_{\ell^+\ell^-} > 4$ GeV is also applied using the loose lepton and the opposite sign lepton from the Z_1 candidate.

However, they use a different definition of the $\mathrm{Z} + \mathrm{LL}$ control regions where the fake rate is applied to get the final estimate of the $\mathrm{Z} + \mathrm{X}$ background:

- $\mathrm{2P2L_{SS}}$ control region is used in the same sign (SS) method, where $\mathrm{Z} + \mathrm{LL}$ candidates are required to have two *same sign* loose leptons.
- 2P2F and 3P1F control regions are used in the opposite sign (OS) method, where $\mathrm{Z} + \mathrm{LL}$ candidates are required to have two *opposite sign* leptons where either both leptons are failing leptons, or one lepton is failing lepton and the other one is passing lepton, respectively.

In these control regions, the Z_1 is defined as the normal Z candidate, while the Z_2 is the one with relaxed selection requirement. The FSR recovery is treated the same way as in the signal region. The jet cleaning procedure is extended to also involve the loose and failing leptons of selected $Z + LL$ candidates and their associated photons. The $2P2L_{SS}$ control region uses slight adaptations to kinematic cuts defined in Section 4.2.6: the QCD suppression cut only applies to the three possible opposite sign pairs, while the alternative pairing check only involves the pairing where the Z_a is an opposite sign pair. To avoid overlaps, events selected in the signal region are not used to define control region candidates, and no more than one such candidate is chosen in each event for each method. Nevertheless, one event can still provide candidates for both the 2P2F and 3P1F regions.

As an example, data to simulation comparison for the three different $Z + LL$ control regions for 2017 analysis is shown in Fig. 4.23. It is clear that while for the $4e$ final state simulation can reasonably well describe the data, it completely fails to do so for the 4μ final state, justifying the choice to estimate the $Z + X$ background from data.

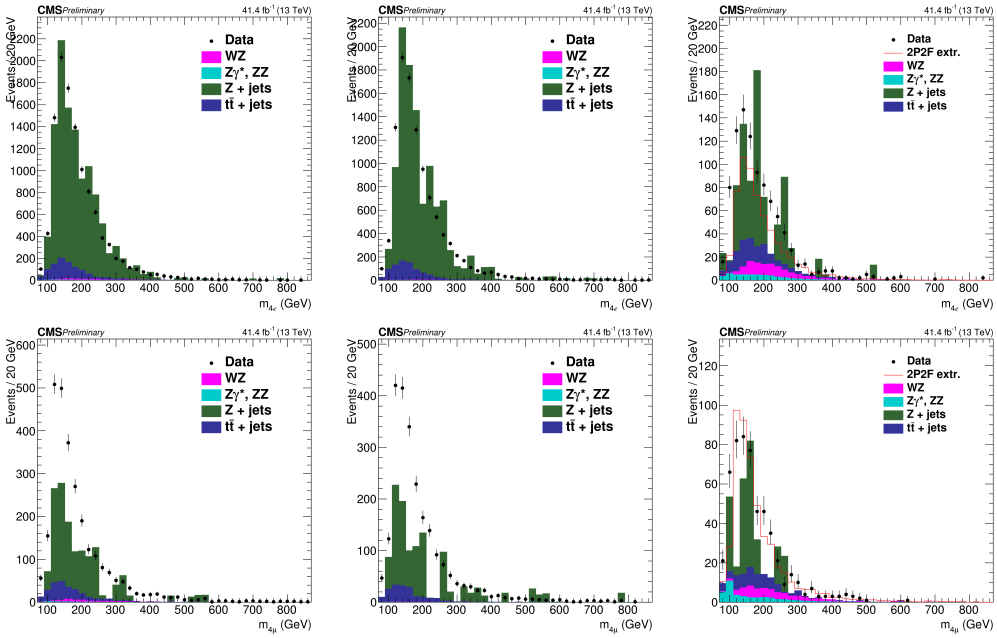

Figure 4.23 – Distribution of the four-lepton invariant mass in the $2P2L_{SS}$ (left), 2P2F (middle), and 3P1F (right) control regions, for the $4e$ (top) and 4μ (bottom) final states. Points with error bars represent the data collected in 2017 and stacked histograms represent expected distributions from simulation, for the main contributing processes. The empty red histogram is the 3P1F distribution extrapolated from 2P2F control region in date as explained later in text.

Fake rate estimation with the OS method

In order to measure the lepton fake ratios f_e and f_μ, defined simply as the fraction of loose leptons that also pass the full selection criteria, the $Z + L$ control region that is

expected to be completely dominated by final states which include a Z boson and a fake lepton is used. Loose leptons from such events are used as probes for the fake rate measurement. The fake rates are measured in bins of the transverse momentum of the loose lepton separately for barrel and endcap regions, using the tight requirement $|m_{Z_1} - m_{Z^0}| < 7$ GeV to reduce the contribution from asymmetric photon conversions. To suppress contamination from WZ and $t\bar{t}$ processes, the missing transverse energy is required to be less than 25 GeV. Finally, remaining non-significant contribution coming from the WZ process is estimated from the simulation and subtracted from the $Z + L$ control region before performing the measurement of the fake rates. An example of fake rates derived with for the OS method in 2017 analysis is shown in Fig. 4.24. Knowing that roughly 80% of fakes have $p_T < 20$ GeV and by comparing the low-p_T fake rate bins for muons and electrons, one can estimate that around 5% electron and 10 to 15% muon fakes pass the $H \to 4\ell$ selection.

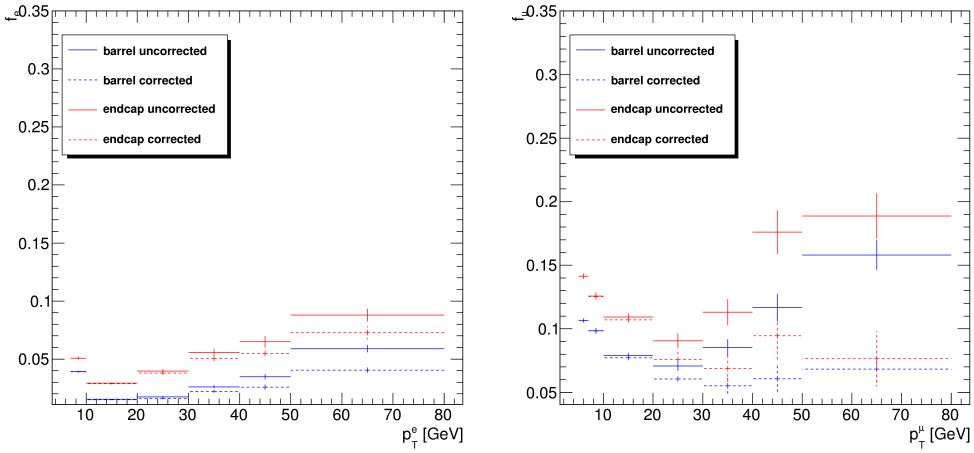

Figure 4.24 – Fake rates as a function of the loose lepton p_T for electrons (left) and muons (right) measured for the OS method in a $Z + L$ control region, on the 2017 data sample. The barrel selection includes electrons with $|\eta^e| < 1.479$, and muons with $|\eta^\mu| < 1.2$. The fake rates are shown before (dashed lines) and after (full line) the removal of the WZ contribution using simulation.

Fake rate application with the OS method

The OS method estimates contribution of the $Z + X$ events coming from 2P2F and 3P1F control regions, while being very careful not to do any double counting:

- The 2P2F control region is used to estimate backgrounds which intrinsically have only two signal leptons, such as $Z + \text{jets}$ and $t\bar{t}$. Their contribution to the signal region is obtained by weighting each event i in the 2P2F region by a factor $[f_3^i/(1-f_3^i)][f_4^i/(1-f_4^i)]$, where f_3^i and f_4^i are the fake rates for the third and fourth lepton, respectively.
- The 3P1F control region is both populated by the previous processes and by back-

grounds that have three genuine signal leptons, such as $\mathrm{WZ}+\mathrm{jets}$ and $\mathrm{Z}\gamma+\mathrm{jets}$ where the photon converts to e^+e^-. Their contribution in the signal region is estimated by weighting each event j of the 3P1F region by a factor $f_4^j/(1-f_4^j)$, where f_4^j is the fake rate for the lepton that does not pass the analysis selection.

- The contribution from ZZ events to the 3P1F region $\mathrm{N}_{\mathrm{3P1F}}^{\mathrm{ZZ}}$, which arises from events where a signal lepton fails identification or isolation requirements, is estimated from simulation and scaled with a factor w_{ZZ} appropriate to the integrated luminosity of the analyzed data set.

- The contamination of 2P2F-type processes in the 3P1F region is estimated as $\sum_i\{[f_3^i/(1-f_3^i)]+[f_4^i/(1-f_4^i)]\}$ and contributes an amount equal to $\sum_i\{2[f_3^i/(1-f_3^i)][f_4^i/(1-f_4^i)]\}$ to the expected yield in the signal region. This amount is subtracted from the total background estimate to avoid double counting.

The total reducible background estimate in the signal region coming from the two categories 2P2F and 3P1F without double counting, $\mathrm{N}_{\mathrm{SR}}^{\mathrm{Z+X}}$, can be written as:

$$\mathrm{N}_{\mathrm{SR}}^{\mathrm{Z+X}} = \sum_j^{\mathrm{N_{3P1F}}} \frac{f_4^j}{1-f_4^j} - w_{\mathrm{ZZ}} \sum_j^{\mathrm{N_{3P1F}^{ZZ}}} \frac{f_4^j}{1-f_4^j} - \sum_i^{\mathrm{N_{2P2F}}} \frac{f_3^i}{1-f_3^i}\frac{f_4^i}{1-f_4^i}, \tag{4.11}$$

where N_{3P1F} and N_{2P2F} are the number of events in the 3P1F and 2P2F regions, respectively. An example of the $\mathrm{Z}+\mathrm{X}$ estimation using the OS method, together with all components used in the calculation is shown in Fig. 4.25 for $4e$ and 4μ final states.

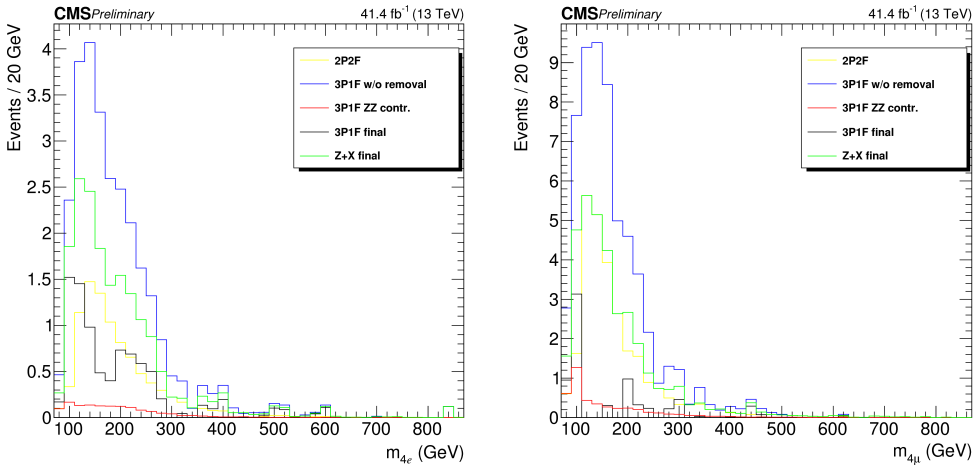

Figure 4.25 – Invariant four-lepton mass distribution for the $\mathrm{Z}+\mathrm{X}$ (green) estimate from the OS method in the $4e$ (left) and 4μ (right) final states using the 2017 data set. All components of the estimate, mentioned in the text, are shown: 2P2F (yellow), 3P1F (blue), 3P1F with subtracted contamination coming from the 2P2F region (black), and contamination of true ZZ events in the 3P1F region (red) that is estimated from the simulation.

Fake rate estimation with the SS method

The lepton fake rates in the SS method are determined in a very similar way as it was described before for the OS method. Same cuts on the transverse momentum of the leptons and the missing transverse energy are used, and WZ subtraction from simulation is performed. The only difference is requirement on the invariant mass of the Z candidate, where a much wider window $40 < m_{Z_1} < 120$ GeV is chosen. Because of this, events where a radiated photon makes an asymmetric conversion contribute significantly to the $Z + e$ control region that is used to measure the electron fake rate. As a result of this enhanced contribution from conversions, the electron fake rates measured with the SS method are larger than the OS fake rates. However, the relative fraction of FSR conversions is not the same in the $Z + L$ control region and in the $2P2L_{SS}$ control region where the fake rates will be applied. A correction accounting for this difference must be applied to the fake rates measured with the SS method. To determine this correction, several fake rate samples of $Z + e$ events are defined by varying the requirements on $|m_{Z_1} - m_{Z^0}|$ and $|m_{Z_1,e} - m_{Z^0}|$, where e is the additional loose electron. These samples are designed to target different amounts of conversions, ranging from minimal to maximal contamination. In each sample, performed in several $(p_T^e, |\eta^e|)$ bins, one determines the fake rate and the average value of the expected missing hits in the inner tracker $< N_{miss\,hits} >$. The $< N_{miss\,hits} >$ is known to be very sensitive to conversions, and is expected to grow linearly with the increase of conversions. Hence, one expects a linear dependence of the fake rate with respect to $< N_{miss\,hits} >$, and indeed one example of such behavior can be seen in Fig. 4.26. In each bin linear fits are performed, which relate the fake rate to $< N_{miss\,hits} >$ and thus indirectly to fraction of conversions in the event.

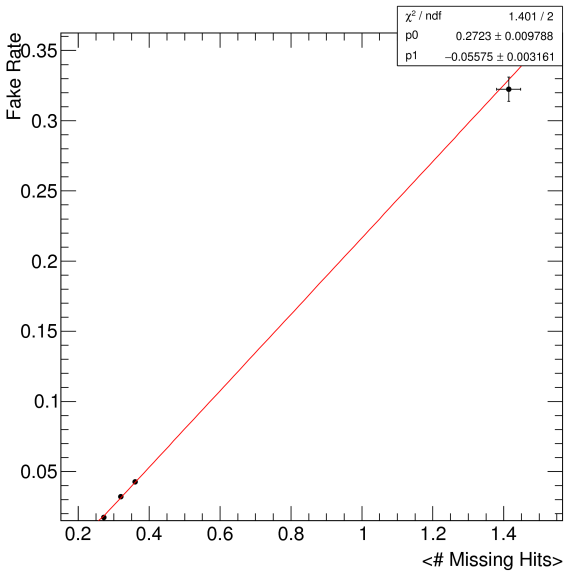

Figure 4.26 – Fake rate as a function of average value of the expected missing hits for four different $Z + e$ samples (black points) with different contributions of conversions for the $20 < p_T^e < 30$ GeV electrons in barrel. Calculation is based on the 2017 data sample. A linear fit (red line) to data is shown, and resulting parameter values are also reported.

Finally, one looks at the loose electrons in the 2P2L$_{\mathrm{SS}}$ control region where the fake rate will be applied, and measures in each $(p_{\mathrm{T}}^{e}, |\eta^{e}|)$ bin the $< N_{miss\,hits} >$ and uses this value together with the linear fit from the same function to calculate the corrected fake rate.

Figure 4.27 shows examples of the resulting corrected electron fake rates, together with the uncorrected fake rates measured with the SS method. It can be seen that the correction for the actual fraction of conversions that is present in the control sample lowers the fake rates considerably. Also, one can see a decrease of the fake rate in the 2017 analysis with respect to 2016, mainly for low-p_{T} electrons in the endcap region that represent most of the fakes. This was expected due to the upgrade of the pixel detector explained in Section 2.3.3 and due to improvements in the electron BDT identification with the addition of isolation variables as explained in Section 3.2.4. This leads to an overall reduction of $\mathrm{Z}+\mathrm{X}$ yield, as will be seen when discussing results of 2016 and 2017 analysis in Chapter 5.

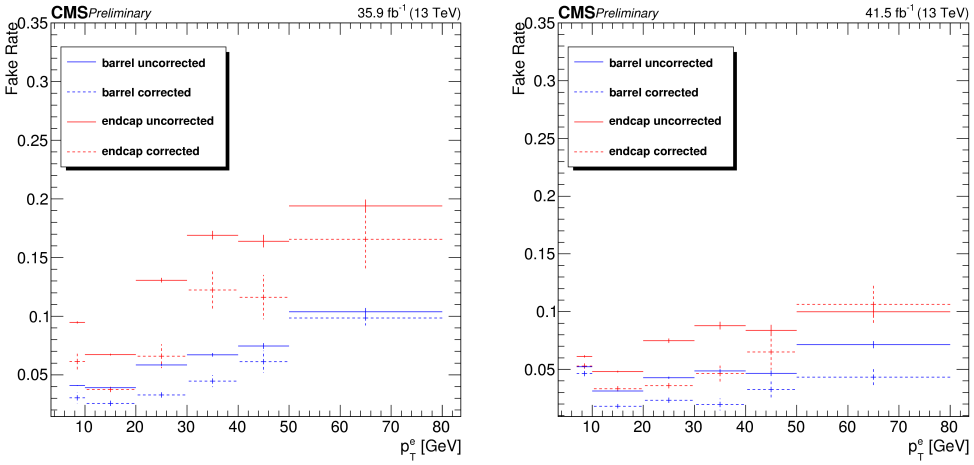

Figure 4.27 – Fake rates as a function of the loose lepton p_{T} for electrons measured for the SS method in a $\mathrm{Z}+\mathrm{L}$ control region, on the 2016 (left) and 2017 (right) data samples. The barrel selection includes electrons with $|\eta^{e}| < 1.479$. The fake rates are shown before (dashed lines) and after (full line) the correction for conversions for electrons as described in text.

Fake rate application with the SS method

The expected number of reducible background events in the signal region can then be written as:

$$\mathrm{N}_{\mathrm{SR}}^{\mathrm{Z+X}} = r_{\mathrm{OS/SS}} \sum_{i}^{\mathrm{N}_{\mathrm{2P2L_{SS}}}} f_3^i f_4^i, \tag{4.12}$$

where $r_{\mathrm{OS/SS}}$ is the ratio between the number of events in the 2P2L$_{\mathrm{OS}}$ and 2P2L$_{\mathrm{SS}}$ control regions, where the 2P2L$_{\mathrm{OS}}$ region is defined analogously to 2P2L$_{\mathrm{SS}}$ but with opposite-sign loose leptons. The correction ratio $r_{\mathrm{OS/SS}}$ is computed for all final states, and was found to be compatible with unity within the uncertainties. An example of the $\mathrm{Z}+\mathrm{X}$ distributions derived for the 2017 analysis, using the SS method are shown

in Fig. 4.28 for $4e$ and 4μ final states. A fit to Landau [112] function is performed to extract the shape of the $m_{4\ell}$ distribution for the reducible background that will be used in the analysis. However, this is not the final estimate since the two methods have to be combined first.

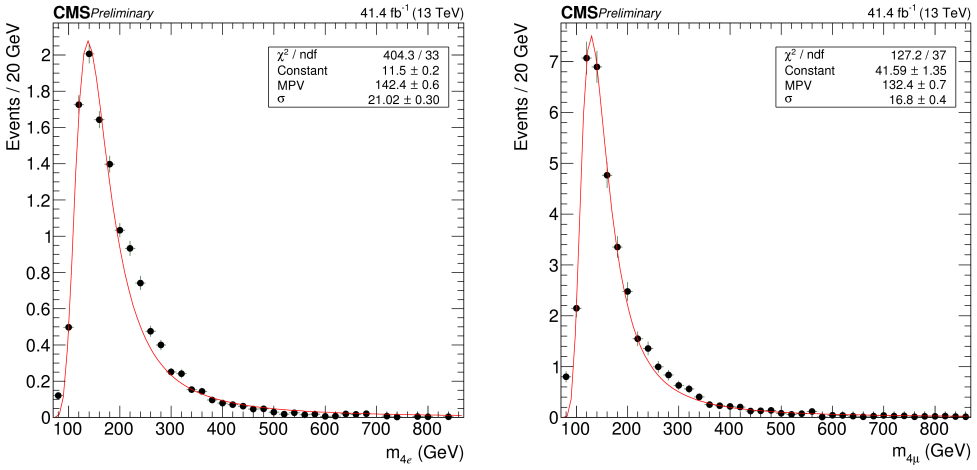

Figure 4.28 – Invariant four-lepton mass distribution for the $Z + X$ estimate from the SS method (black points) in the $4e$ (left) and 4μ (right) final states using the 2017 data set. A fit to Landau function (red line) to data is shown, and resulting parameter values are also reported.

Combination of the OS and SS methods

The predicted yield in the signal region of the reducible background from the two methods was found to be in agreement within their uncertainties, and since they are mutually independent, the results of the two methods are combined. The final estimate is obtained by weighting the individual mean values of both methods according to their corresponding variances. The shape of the $m_{4\ell}$ distribution is then obtained by fitting this combined prediction from the OS and SS methods with empirical functional forms built from Landau and exponential distributions, separately for each final state.

To predict the $Z + X$ yield in every event category, only the SS method is used, and the N_{SR}^{Z+X} are then rescaled accordingly to match the combined estimate when summing all categories, in each of the four final states. Since event categorization relies on associated jets and additional selected leptons, the definitions of categories of the signal region are naturally extended to categories of the $2P2L_{SS}$ region, with the $Z + LL$ candidate being used as a ZZ candidate. In principle, one would like to estimate yields in categories with the OS method as well, and then combine the results like it is done inclusively. However, due to several negative contributions for the OS method in eq. 4.11, and due to very low statistics in most of the categories, the predictions by the OS method are very susceptible to statistical fluctuations. In a similar manner, shapes obtained for inclusive distributions in each final states are then used for every category in the corresponding final states. While slight deviations from the inclusive shape are expected, this is mostly in the very low populated categories where no reliable shape estimate can be done at

the moment.

The numerical results of this entire procedure be presented in Chapter 5 for the 2016 and 2017 analyses, respectively.

Systematic uncertainties

The predicted final yield $\mathrm{N}_{\mathrm{SR}}^{\mathrm{Z+X}}$ is one of the measurements with the highest uncertainties in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis, however with a small impact on the final results. The origin of the large uncertainty comes from several sources:

- A statistical uncertainty arises from the limited size of the data sample in the control regions, and in the regions where fake rates are computed. It is typically in the range of 1 to 10%.
- An important source of systematic uncertainty is the difference in the composition of the $\mathrm{Z}+\mathrm{L}$ control region used to compute the fake rates and the control regions where they are applied. This uncertainty is estimated by first measuring the fake rates for individual background processes in the $\mathrm{Z}+\mathrm{L}$ region in simulation, and combining them to compute a reweighted fake rate according to the exact background composition of the 2P2F region obtained from simulation. The difference between the prediction derived from this reweighed fake rate and the regular one defines the uncertainty, which is of the order or 30–40%, depending on the final state.
- For the SS method, an additional uncertainty arises from the statistical error bars of the measurement of $< N_{miss\,hits} >$ in the $\mathrm{2P2L_{SS}}$ control region, which propagate to the electron fake rates, and finally to $\mathrm{N}_{\mathrm{SR}}^{\mathrm{Z+X}}$.
- Besides the yield predictions, a shape uncertainty defined from the difference between the two methods can be assigned to the final estimate. However, this difference was found to vary slowly with $m_{4\ell}$, so it was decided to only use the uncertainty on predicted yields.

The total uncertainty is then simply defined as a sum in quadrature of all contributions and is found to be around 40% for both 2016 and 2017 analyses.

4.6 Signal modeling

Just like it is done for backgrounds, analytical shapes are used to describe the signal for the measurement purposes. The main difference is, in order to measure m_{H}, or to measure another physics parameter for a certain value of m_{H}, all signal predictions need to be continuously parameterized as a function of m_{H}. This is done using dedicated samples which have been simulated for five different mass points of m_{H}: 120, 124, 125, 126, and 130 GeV, as mentioned in Section 4.1.2. Work presented in this thesis is

focused on the on-shell Higgs with m_{H} ~125 GeV, so a $105 < m_{4\ell} < 140$ GeV window is defined as the signal region where all of the measurements are performed.

To describe the signal probability density function (pdf) for $m_{4\ell}$ with a proper analytical function, one has to take into account both the theory and the experiment. For such low values of m_{H}, the narrow-width resonance hypothesis holds, which means that the theoretical signal lineshape can be described as a relativistic Breit-Wigner function. From the other side, the effects of experimental resolution such as tails due to bremsstrahlung, final state radiation, and energy leakage in the ECAL also need to be taken into account. A double-sided Crystal Ball (dCB) function is usually chosen, which both accounts for the Gaussian resolution of the core of the $m_{4\ell}$ distribution, and for the two asymmetric non-Gaussian tails described by power laws. In practice, given the narrowness of the theoretical lineshape, the signal pdf was found to be well described by a dCB function alone:

$$\mathcal{P}(m_{4\ell}|m_{\mathrm{H}}) = f_{\mathrm{dCB}}(m_{4\ell}|m_{\mathrm{H}}) = N \cdot \begin{cases} A \cdot (B + |\xi|)^{-n_L}, & \text{for } \xi < \alpha_L \\ \exp\left(-\xi^2/2\right), & \text{for } \alpha_L \leq \xi \leq \alpha_R \\ A \cdot (B + |\xi|)^{-n_R}, & \text{for } \xi > \alpha_R \end{cases}, \tag{4.13}$$

where $\xi = (m_{4\ell} - m_{\mathrm{H}} - \Delta m_{\mathrm{H}})/\sigma_m$. This function has six independent parameters: the variance σ_m of the Gaussian core, the systematic mass shift Δm_{H} of the peak, and two parameters for each tail: the powers n_L and n_R control their prominence, while α_L and α_R define the position of the boundary between the core and the tails. The values of A and B ensue from requiring the continuity of the function and its first derivative, while the normalizing constant N is determined separately.

The dependency of each of the six parameters p_i on m_{H} has to be determined for every final state and event category. Here, a linear approximation is used:

$$p_i(m_{\mathrm{H}}) = C_i + D_i \cdot (m_H - 125 \text{ GeV}). \tag{4.14}$$

For every final state, the C_i parameters are first obtained from the shapes in the ggH production mode in the Untagged category for $m_{\mathrm{H}} = 125$ GeV. This is illustrated in Fig. 4.29 for the three final states, showing that resolution is better for muons than for electrons. Then, the D_i parameters are determined from a simultaneous fit to all other m_{H} points. One interesting point is that no significant differences between ggH and VBF signal shapes in any of the well populated categories have been observed. Because of this, statistically limited cases use the parametrization obtained from ggH in the Untagged event category.

In the case of ZH and $\mathrm{t\bar{t}H}$, pairs of signal leptons can also originate from the decay of the associated Z bosons and $\mathrm{t\bar{t}}$ pairs. Therefore, instead of only considering the decay to four leptons, all other decays of the $\mathrm{H} \to \mathrm{ZZ}$ system are allowed at first, and a generator-level filter then requires the presence of two pairs of opposite-sign leptons in the final state, regardless of whether they originate from the H boson or from the associated Z or $\mathrm{t\bar{t}}$. This generator-level filter efficiency is accounted for when reweighting these samples using eq.4.1, as described in Section 4.1.2. In other words, part of the studied ZH and $\mathrm{t\bar{t}H}$ events feature a $\mathrm{H} \to 2\mathrm{l}2\mathrm{X}$ decay where X either stands for a quark or for a neutrino.

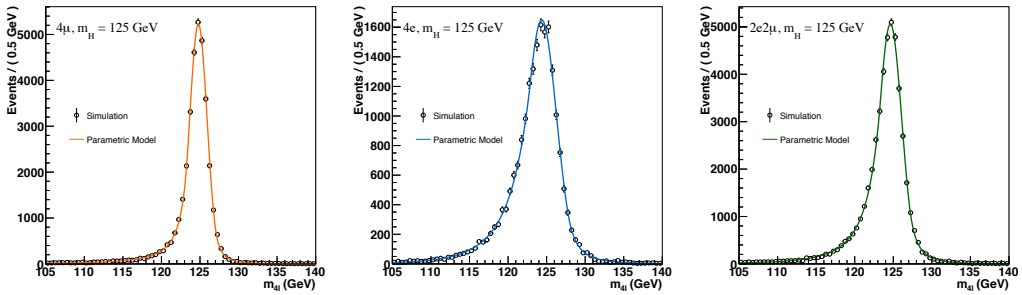

Figure 4.29 – MC simulated $m_{4\ell}$ distributions after full event selection, and fitted pdf $f_{\mathrm{dCB}}(m_{4\ell}|m_{\mathrm{H}})$, for the ggH process generated with $m_{\mathrm{H}} = 125$ GeV, in the Untagged category, for the three final states are compared: 4μ (left), $4e$ (middle), and $2e2\mu$ (right).

These events are often selected by the analysis, forming a non-resonant contribution to the signal region. This contribution is modeled by Landau distribution which is added to the total pdf function to perform the fit in the $m_{\mathrm{H}} = 125$ GeV case, adding two more parameters, which are then fixed for the simultaneous fit. The relative normalization of both components is finally adjusted for every category. An example is given for the ZH process in Fig. 4.30.

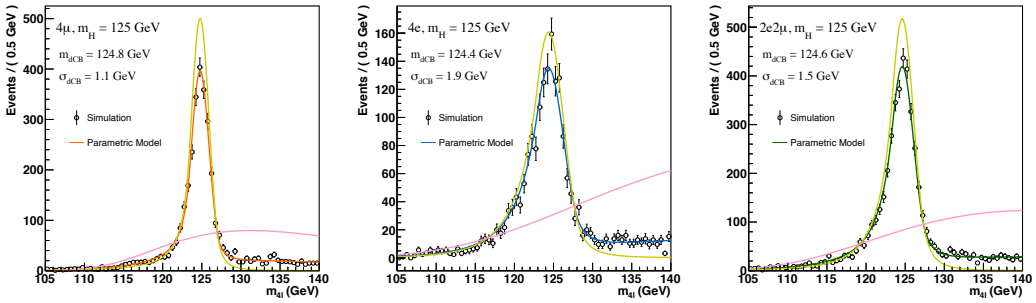

Figure 4.30 – MC simulated $m_{4\ell}$ distributions after full event selection, and fitted pdf $f_{\mathrm{dCB}}(m_{4\ell}|m_{\mathrm{H}})$, for the ZH process generated with $m_{\mathrm{H}} = 125$ GeV, in the Untagged category. The pre-fit pdfs for the resonant and non-resonant component are shown as yellow and pink curves, respectively. The three final states are compared: 4μ (left), $4e$ (middle), and $2e2\mu$ (right).

Final missing part is the parametrization of the normalization constant N, which translates to expected signal yields for a given value of m_{H}. It is performed independently of the shape parametrization, in the same $105 < m_{4\ell} < 140$ GeV window. For every production modes, final state and event category, a fit of expected yields to a second order polynomial is performed using the same five m_{H} points. In cases of very low statistics in simulation, a first-order polynomial is used instead, to get a more robust result. Examples of such fits are presented in Fig. 4.31 for ggH in the Untagged category and ZH in the VBF-2jet-tagged category.

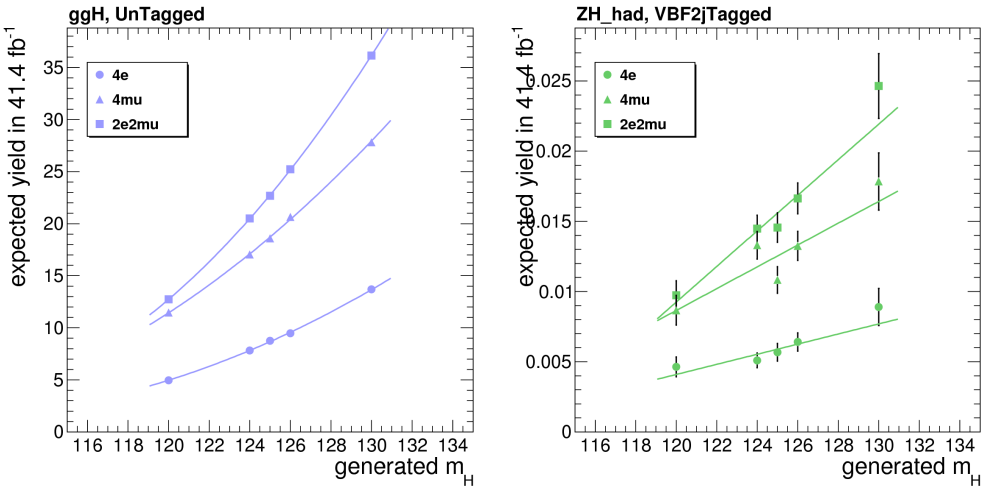

Figure 4.31 – Fits of the m_{H} dependency of the expected signal yields for 41.4 fb^{-1} in the $105 < m_{4\ell} < 140$ GeV window after full event selection, shown for the $118 < m_{4\ell} < 130$ GeV interval. Two examples are shown: ggH in the Untagged category (left) to illustrate a case with a large number of simulated events, and ZH in the VBF-2jet-tagged category (right) for a less populated case.

4.7 Systematic uncertainties

Every serious scientist will never consider a measurement without a corresponding uncertainty. It gives valuable information on the precision of the measurement, and it is impossible to compare two experimental results, or even compare the measurement with theory predictions without it. The overall uncertainty on the measurement has a statistical and systematical contributions. The statistical component mainly depends on the raw amount of data available and motivates us to collect as much data as possible every year, thanks to increased instantaneous luminosity, detector and machine developments, and eventually drives us to design and build new experiments. On the other hand, we have systematic uncertainties that are introduced by an inaccuracy, involving either the observation or measurement process, inherent to the system. When the statistical component completely dominates the measurement, there is no sensitivity for underestimation of systematics and no penalty for being too conservative in the estimate. However, in the regime when the two are of similar sizes, or in the cases when systematic dominates the overall uncertainty it is of utmost importance to get a reliable estimate. Underestimating the uncertainty could have big implications, such as concluding that the measurement is not in agreement with the theoretical expectations, or that you have a new discovery while you do not. On the other hand, being too conservative does not have such a big impact, however it is still important to understand your experimental setup and the measurement, and to produce the result with the best available precision.

The $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis is experiencing an exciting transformation from a discovery analysis to precision properties measurement, where systematical and statistical com-

ponents are starting to have similar contributions. That is why strong efforts are being made to understand all possible sources of systematical uncertainties and estimate them correctly.

In this section an overview of all significant sources of systematical uncertainties on the measurements of Higgs boson properties will be given, providing some information on their estimation and magnitudes.

There are many ways to classify different kinds os systematic uncertainties in the $\mathrm{H} \to 4\ell$ analysis, and in general. For example, they can be divided into *normalization uncertainties* that affect only the yields and *shape uncertainties* that alter the shape of the probability density functions of some of the observables. One could also divide them as *theoretical uncertainties*, coming solely from uncertainties related to theoretical calculations, and *experimental uncertainties* arising from detector and measurement imperfections and biases. Every uncertainty in the $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ analysis can have: *category migration effect*, affecting category composition caused by events migrating between event categories, *inclusive effect* altering only inclusive normalization or shapes with no effect on categories, or both. With so many different classifications available, rather than choosing one, each will be assigned to every uncertainty.

The following sources of uncertainty are taken into account in the $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ analysis. All items of this list are treated as uncorrelated to one another and they are taken into account both in the 2016 and 2017 analysis.

- The measurement of the *integrated luminosity* of the data sample is measured with Van der Meer scans [113], which is defined for all CMS analyses for a given data taking period [114, 115]. It is an experimental uncertainty with inclusive effect on the normalization of about 2%. It is one of the dominant experimental sources in all measurements.

- Experimental uncertainties on the measurements of *efficiencies of trigger and lepton reconstruction and selection* calculated on simulation. First, per-lepton uncertainties on scale factors are obtained by the TnP measurements and explained in Section 3.3.3 and shown in Fig. 4.6, 4.8, 4.7, and 4.9. For each event, assuming that the leptons are uncorrelated, the four-lepton uncertainties are summed in quadrature. An additional uncertainty for trigger efficiency measurement explained in Section 4.2.1, depending on the final state, is also added in quadrature. This is propagated to the four-lepton kinematics using the simulated signal ggH sample and final uncertainty is calculated as the magnitude of the signal yield change. It is applied as a normalization uncertainty and was found to have no category migration effects. Additional studies have also shown that it has a similar effect on the background yields so the same numbers are used. It usually ranges from 5 to 10%, depending on the final state and data taking period, and is the dominant experimental source in most of the measurements.

- A theoretical *QCD uncertainty* is applied to every signal and irreducible background sample in an uncorrelated way. It has two different components. First,

the *renormalization and factorization scales* are varied between 0.5 and 2 times their nominal value, while keeping their ratio between 0.5 and 2. Also, the impact of the *modelling of hadronization and the underlying event* is studied using dedicated generator-level samples with varied PYTHIA parameters for the underlying event tune and hadronization scale. Combined, they were found to have an impact of 3 to 10% on global process normalizations and also cause some anti-correlated migrations of events between different event categories. It affects all measurements.

- The theoretical uncertainty from the choice of a set of *parton distribution functions* (PDFs) is determined following the PDF4LHC recommendations [116]. It is determined independently for different sets of processes grouped by initial state. It was found to have an impact of 3 to 20% on global normalizations and also cause migration effects, affecting all measurements.

- The *electroweak corrections* for the $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background prediction yield in an inclusive normalization and category migration effects of 1 to 8%, depending on the category. It is a theoretical uncertainty that affects all measurements.

- The $\mathrm{gg} \to \mathrm{ZZ}$ background yield is applied an additional theoretical, inclusive normalization uncertainty of 10%, on top of the QCD uncertainty, to account for the fact that its NNLO/LO k-factor was actually computed for a $\mathrm{gg} \to \mathrm{H} \to \mathrm{ZZ}$ signal, as explained in Section 4.5.1.

- A theoretical systematic uncertainty of 2% on the *branching ratio* of $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ is applied to the yields of all signal processes [9].

- Uncertainties on the *reducible background estimation* have been described in Section 4.5.2. The uncertainties are treated as uncorrelated between final states, since each of them can be viewed as a separate measurement. It is an experimental uncertainty that affects only normalization and has an inclusive effect. It is around 40% and affects all measurements, but does not have a significant impact on the total uncertainty.

- The experimental uncertainties on the *lepton energy scale* and the *4ℓ mass resulution* follow from the calibrations described in Section 3.1 and 4.2.3, and are crucial to the measurement of the Higgs boson mass. They are defined as shape uncertainties affecting the mean $m_H + \Delta m_H$ and resolution σ_m of the dCB function defined in Eq. 4.13, respectively. They are determined by considering the $\mathrm{Z} \to \ell^+\ell^-$ mass distributions in data and simulation. Events are separated into categories based on the p_{T} and $|\eta|$ of one of the two leptons, selected randomly, and integrating over the other. A Breit-Wigner parameterization convolved with a double-sided Crystal Ball function is then fit to the dilepton mass distributions. The offsets in the measured peak position, with resolution fixed in the fit, with respect to the nominal Z boson mass in data and simulation are extracted, and the results are shown in Fig. 4.32. In the case of electrons, since the same data set is used to derive and validate the momentum scale corrections, the size of the corrections is taken into account for the final value of the uncertainty. The 4ℓ mass scale uncertainty is

determined to be 0.04%, 0.3%, and 0.1% for the 4μ, $4e$, and $2e2\mu$ final states, respectively. As a cross check, the same measurement was performed by varying the scale of leptons in a simulation with a corresponding scale uncertainty that is calculated when the scale and smearing corrections are derived, as described in Section 4.2.3. These uncertainties are propagated to the $Z \to \ell^+\ell^-$ mass distribution, and scale uncertainty is again extracted from a fit. The results from the two methods are found to be in agreement. In a similar manner, the uncertainties on the smearing procedure, that affects the lepton resolution, can be propagated and a fit with floating resolution and fixed scale can be performed. This procedure yielded a 10% uncertainty for muons and 20% for electrons.

- The imprecise knowledge of the *jet energy scale* affects the number of selected jets in an event, one of the main categorization observables. Varying the jet energy correction factor up and down leads to uncertainties ranging from 2 to 20% depending on the process and category, with largest impacts for the VBF-2jet-tagged and $t\bar{t}H$-tagged categories. It is an experimental uncertainty that only affects the event categorization, having no impact on the overall normalization or shapes. It is fully correlated for different processes, and anti-correlated between categories.

- The scale factors for *b-tagging efficiency* described in Section 4.2.5 are assigned some uncertainties. Their upward and downward variations can modify the number of b-tagged jets per event, altering the expected yields of the $t\bar{t}H$-tagged categories by 2 to 10% depending on the process and data taking period. Just like the jet energy scale uncertainty, it is an experimental uncertainty that only affects the event categorization, having no impact on the overall normalization or shapes. It is fully correlated for different processes, and anti-correlated between categories.

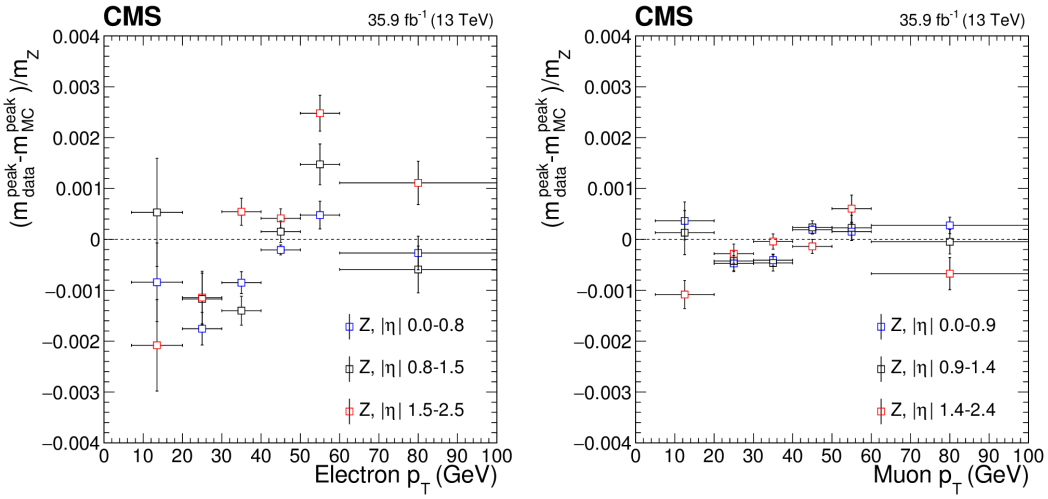

Figure 4.32 – Difference between the $Z \to \ell^+\ell^-$ mass peak positions in data ($m_{\text{data}}^{\text{peak}}$) and simulation ($m_{\text{MC}}^{\text{peak}}$) normalized by the nominal Z boson mass (m_{Z}), as a function of the p_{T} and $|\eta|$ of one of the leptons regardless of the second, for electrons (left) and muons (right).

Many other sources have been considered and studied but showed little or no impact

on the final results. One of the examples is the uncertainty on the $\mathrm{Z}+\mathrm{X}$ shape described in Section 4.5.2, which showed to be almost constant and well absorbed in the normalization uncertainty. Another example features kinematic discriminants described in Section 4.4.2, where many studies were performed to make sure that effects like jet energy scale or QCD scale variations did not alter the shape of the newly designed production and decay discriminants.

While this section focused on different sources and procedures to estimate systematic uncertainties, the way they are taken into account in statistical analysis and propagated to final results will be given in Section 4.8.3. All these numbers are being estimated again for each data taking period, so their values for the 2016 and 2017 analysis will be summarized in Sections 5.1 and 5.2.

4.8 Measurement strategies

With all building blocks of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis falling in place, the missing part is understanding how the final measurements are performed. While in Run I the main focus was on the search for Higgs boson, by looking for significant excess of data with respect to the SM prediction and quantifying it, the focus has shifted to properties measurements. Many different properties are being considered, and this section describes the statistical methods, fit models and treatment of systematic uncertainties that are used to derive the results from inputs defined in the previous sections.

4.8.1 Statistical methods

A well defined common statistical procedure was developed in early Run I by the CMS and ATLAS collaborations in the context of the LHC Higgs Combination Group [117], in order to coordinate searches for the SM Higgs boson.

For measurements of continuous physics parameters, a *maximum likelihood estimation* is used. The method of maximum likelihood is based on the *likelihood function* $\mathcal{L}(\mathrm{data}|\theta)$, where θ denotes a parameter of interest on which the measured data depends. For example, in our case, data can be measured values of the $m_{4\ell}$, while in that case θ parameter would represent a *true* value of the Higgs boson mass m_{H}. The method of maximum likelihood finds the values of the model parameter of interest θ, that maximize the likelihood function $\mathcal{L}(\mathrm{data}|\theta)$. Intuitively, this selects the parameter values that make the observed data most probable.

The most general model used relies on one global *signal strength* parameter μ that multiplies the expected SM Higgs boson cross section such that $\sigma = \mu \cdot \sigma_{\mathrm{SM}}$. All production modes are scaled by this same factor, while branching fractions are assumed to be preserved. Each independent source of systematic uncertainty is assigned a *nuisance parameter* θ_i, the full set of which is denoted as θ. The expected background and SM Higgs boson signal yields are functions of these parameters, and can be written as $b(\theta)$ and $\mu \cdot s(\theta)$, respectively. Systematic uncertainties usually reflect the possible deviations of

a quantity from the input value $\tilde{\theta}_i$ provided by a separate measurement. On that basis, the likelihood can be defined, given the data, either actual data or a pseudo-experiment, and the measurements of $\tilde{\theta}$, as:

$$\mathcal{L}(\text{data}, \tilde{\theta}|\mu, \theta) \equiv \prod_c \mathcal{L}_c(\text{data}|\mu \cdot s(\theta) + b(\theta)) \cdot \prod_i p_i(\tilde{\theta}_i|\theta_i), \tag{4.15}$$

where first product runs over all channels c, and the second one over all possible nuisance parameters i considered in the analysis. In the case of the *binned analysis*, the $\mathcal{L}_c$ functions stand for products of Poisson probabilities for the number of events in every bin of their channel. However, the $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ analysis is an *unbinned analysis*, using simulation to build pdfs for parameters of interest $\mathcal{O}$, as described in Section 4.5 and 4.6 for the $m_{4\ell}$ observable. The likelihood function is expressed as the product over all N observed events of the pdfs of observable $\mathcal{O}$ for signal $f_s(\mathcal{O}|\theta)$, and background events $f_b(\mathcal{O}|\theta)$, weighted by the total expected signal and background rates $S(\theta)$ and $B(\theta)$:

$$\mathcal{L}_c(\text{data}|\mu \cdot s(\theta) + b(\theta)) \equiv \frac{1}{N} \prod_{e=1}^{N} (\mu S(\theta) f_s(\mathcal{O}_e|\theta) + B(\theta) f_b(\mathcal{O}_e|\theta)) \cdot \exp^{-\mu S(\theta) + B(\theta)}. \tag{4.16}$$

In order to measure central values and uncertainty intervals, the strategy then consists in maximizing $\mathcal{L}$ with respect to μ and θ, to obtain the best fit values $\hat{\mu}$ for the parameters of interest, and $\hat{\theta}$ for the nuisance parameters. One defines a *negative log-likelihood function*:

$$-2\Delta \ln \mathrm{L} \equiv -2 \ln \frac{\mathcal{L}(\text{data}, \tilde{\theta}|\mu, \hat{\theta_\mu})}{\mathcal{L}(\text{data}, \tilde{\theta}|\hat{\mu}, \hat{\theta})}, \tag{4.17}$$

and performs a *maximum likelihood fit* (or often just called fit) to data, where $\hat{\theta_\mu}$ maximizes the numerator for a fixed set of values μ of the parameters of interest. The determination of confidence intervals relies on the Wilks theorem [118], which states that for a model with n parameter of interest, the distribution of $-2\Delta \ln \mathrm{L}$ approaches a χ^2 with n degrees of freedom in the limit of a large data sample. For example, when measuring one single parameter μ, the 68% and 95% CL intervals are deduced from the conditions $-2\Delta \ln \mathrm{L} < 1$ and $-2\Delta \ln \mathrm{L} < 3.84$, respectively. In general, the results are both quoted as central values with 68% CL intervals, and displayed graphically as scans of $-2\Delta \ln \mathrm{L}$. Expected results can also be provided for some nominal values of the parameters, which is very useful for estimating the sensitivity of a given measurement. This would in principle require to generate a large number of pseudo-experiments and determine their median outcome, but a very good approximation is provided by the Asimov data set [119], i.e. one single representative data set in which the observed rates and distributions coincide with predictions under the nominal set of nuisance parameters.

This general model can easily be extended to accommodate two or more parameters. One can choose to measure the parameters simultaneously, but also there is a possibility to measure them separately. When measuring one parameter in a model that has more than one, the above method can be applied by either fixing the other parameters to some nominal value, or profiling them, i.e. leaving them floating in the fit.

Table 4.8 – Summary of observables, pdfs and channels used in the physics measurements of the $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ analysis.

Measurement	**pdfs and observables**	**Channels**
Signal strength μ	$f^{\mu}_{\mathrm{2D}}(m_{4\ell}, \mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}})$	7 categories $\times$ 3 final states
Cross sections $\sigma_{\mathrm{fid}}, \mathrm{d}\sigma_{\mathrm{fid}}/\mathrm{dp_T(H)}, \cdots$	$f^{\sigma}_{\mathrm{1D}}(m_{4\ell})$	3 final states
Mass m_{H}	$f^{m}_{\mathrm{1D}}(m_{4\ell})$ $f^{m}_{\mathrm{2D}}(m_{4\ell}, \mathcal{D}_{\mathrm{m}})$ $f^{m}_{\mathrm{3D}}(m_{4\ell}, \mathcal{D}_{\mathrm{m}}, \mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}})$	3 final states
Width Γ_{H}	$f^{\Gamma}_{\mathrm{1D}}(m_{4\ell})$	3 final states

The set of channels c, the choice of observables $\mathcal{O}$, and the construction of the pdfs f_s and f_b for different measurements is summarized in Table 4.8 and discussed in Section 4.8.2, while the choice of the pdfs p_i associated to each nuisance parameter is described in Section 4.8.3.

4.8.2 Multi-dimensional likelihoods

Signal strength measurement

The simplest way of measuring the global signal strength modifier of the Higgs boson signal would be with a one-dimensional fit to the four-lepton invariant mass distribution. The likelihood function then just relies on the 1D mass pdf that has been discussed in Section 4.6 for signal, and in Section 4.5 for background. However, this model can be greatly improved by exploiting the matrix element kinematic discriminant $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$, as a second dimension:

$$f^{\mu}_{\mathrm{2D}}(m_{4\ell}, \mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}) = \mathcal{P}(m_{4\ell}) \times \mathcal{P}(\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}} | m_{4\ell}). \tag{4.18}$$

The second factor controls the shape of $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$ and its correlation with $m_{4\ell}$, and is based on 2D histogram templates, to which a smoothing procedure is applied. The conditional part is implemented by normalizing each bin of the $m_{4\ell}$ dimension to one. As a result, $\mathcal{P}(\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}} | m_{4\ell})$ only provides information on the kinematic discriminant given a certain mass, but no information on the mass itself. The 2D templates are built separately for the 3 final states, from simulation for the signal, $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ and $\mathrm{gg} \to \mathrm{ZZ}$ components and from the control region for $\mathrm{Z} + \mathrm{X}$, using the SS method. Finally, based on the event categories defined in Section 4.3 and on the three final states, the f^{μ} pdfs are built for $7 \times 3 = 21$ channels.

Besides the global signal strength, other similar measurements can be performed using the same framework. Instead of using a global signal strength modifier to vary all channels together, one can define different kinds of groupings resulting in other interesting measurements:

- By trivial definition of different signal strengths for each category μ_{cat} would yield

a signal strength in each event category. This is achieved by replacing the $\mu \cdot s + b$ model with $\sum_{\text{cat}} \mu_{\text{cat}} \cdot s_{\text{cat}} + b$.

- One can define two signal strength modifiers $\mu_{\text{ggH, t}\bar{\text{t}}\text{H}}$ and $\mu_{\text{VBF, VH}}$ as scale factors for the fermion and vector-boson induced contributions to the expected SM cross section. In this model $\mu_{\text{ggH, t}\bar{\text{t}}\text{H}}$ scales the Higgs boson production cross sections in modes related to couplings to fermions, and $\mu_{\text{VBF, VH}}$ scales the cross sections in modes related to couplings to electroweak gauge bosons. This is achieved by replacing the $\mu \cdot s + b$ model with $\mu_{\text{ggH, t}\bar{\text{t}}\text{H}} \cdot (s_{\text{ggH}} + s_{\text{t}\bar{\text{t}}\text{H}}) + \mu_{\text{VBF, VH}} \cdot (s_{\text{VBF}} + s_{\text{VHhad}} + s_{\text{VHlep}}) + b$.
- To fully exploit event categorization defined in Section 4.3, one can introduce signal strentgh modifiers for each of Higgs boson production modes individually. Five signal strength modifiers are defined: μ_{ggH}, μ_{VBF}, μ_{VHhad}, μ_{VHlep}, and $\mu_{\text{t}\bar{\text{t}}\text{H}}$. The WH and ZH processes are merged, and then split based on the decay of the associated vector boson into either hadronic or leptonic decays. This is done because with the current categorization there is not much sensitivity to WH and ZH production modes separately, while different decay modes are used to define event categories. The models therefore assumes that the ratio of HWW to HZZ coupling strengths is as predicted by the SM, an assumption that was already made in considering W boson fusion and Z boson fusion together in the VBF process. This model is achieved by replacing the $\mu \cdot s + b$ model with $\mu_{\text{ggH}} \cdot s_{\text{ggH}} + \mu_{\text{VBF}} \cdot s_{\text{VBF}} + \mu_{\text{VHhad}} \cdot s_{\text{VHhad}} + \mu_{\text{VHlep}} \cdot s_{\text{VHlep}} + \mu_{\text{t}\bar{\text{t}}\text{H}} \cdot s_{\text{t}\bar{\text{t}}\text{H}} + b$.

To further improve sensitivity of the signal strength measurements for different Higgs boson production modes, newly developed production and kinematic discriminants described in Section 4.4.2 were introduced in the 2017 analysis. The procedure is identical to the one just described, the only difference being that instead of the same kinematic observable being used for every event category in Eq. 4.18, different observables are defined for every category:

- $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ is used in the VBF-2jet-tagged category, and $\mathcal{P}(\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}|m_{4\ell})$ 2D templates are built separately for each signal production mode and SM backgrounds, in each final state.
- $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ is used in the VH-hadronic-tagged category, and $\mathcal{P}(\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}|m_{4\ell})$ 2D templates are built separately for each signal production mode and SM backgrounds, in each final state.
- $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ is used in all other event categories, and $\mathcal{P}(\mathcal{D}_{\text{bkg}}^{\text{kin}}|m_{4\ell})$ 2D templates are built separately for each signal production mode and SM backgrounds, in each final state.

A test of the improvement gain from new discriminant was performed, running the full statistical analysis. The likelihood scans for the expected signal strength modifiers corresponding to the five main Higgs boson poduction modes were computed for the nominal

configuration and for the one using new production and decay kinematic discriminant, and an improvement of 10 to 15% was measured on the expected sensitivity on the measurements of μ_{VBF} and μ_{VHhad}, respectively.

Cross section measurement

In theories beyond the SM, the properties of the 125 GeV Higgs boson may not be determined only by a simple scaling of couplings. Instead, the kinematic distributions in the various Higgs boson production and decay channels may be sensitively modified by BSM effects. *Fiducial cross sections*, whether total or differential, for specific states within the phase space defined by experimental selection and acceptance cuts, provide a largely model-independent way to test for such deviations in kinematic distributions. In particular, differential fiducial cross sections are a powerful for scrutinizing the SM Lagrangian structure of the Higgs boson interactions, including tests for new tensorial couplings, non-standard production modes, determination of effective form factors, etc.

There are, however, several questions to address on both, the experimental and the theoretical sides, for making most out of fiducial cross section measurements. On the experimental side these include, for example, the definition of the fiducial volumes and the unfolding procedure, both of which will be described shortly. On the theory side, important issues include the precision of MC simulations, the expected BSM effects, and whether some BSM effects might affect the unfolding procedure.

The definition of the generator-level *fiducial volume*, chosen to match closely the reconstruction level selection, is very similar to the definition used in Ref. [52]. The differences with respect to Ref. [52] are that leptons are defined as *dressed leptons*, as opposed to Born-level leptons, and the lepton isolation criterion is updated to match the reconstruction-level selection. Leptons are *dressed* by adding the four-momenta of photons within $\Delta R < 0.3$ to the bare leptons, and leptons are considered isolated if the scalar sum of transverse momenta of all stable particles, excluding electrons, muons, and neutrinos, within $\Delta R < 0.3$ from the lepton is less than $0.35 p_{\mathrm{T}}$. For the measurement of differential cross sections related to jet observables, only well measured central jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$ are considered in both the fiducial and reconstruction-level selections. To simplify the definition of the fiducial volume, the $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$ discriminant is not used to select the ZZ candidate at the generator level. Instead the Z_1 candidate is chosen to be the one with m_{Z_1} closest to the nominal Z boson mass, and in cases where multiple m_{Z_2} candidates satisfy all criteria, the pair of leptons with the highest sum of the transverse momenta is chosen. The same candidate selection is also used at the reconstruction level for the fiducial cross section measurements to align the reconstruction- and fiducial-level selections as closely as possible. The full definition of the fiducial volume is detailed in Table 4.9 and the acceptance $\mathcal{A}_{\mathrm{fid}}$ for various SM production modes is given in Table 4.10.

A maximum likelihood fit of the signal and background parameterizations to the observed 4l mass distribution, $N_{\mathrm{obs}}(m_{4\ell})$, is performed to extract the integrated fiducial cross section σ_{fid} for $\mathrm{pp} \to \mathrm{H} \to 4\ell$. The fit is done without any event categorization

Table 4.9 – Summary of requirements and selections used in the definition of the fiducial phase space for the $\mathrm{pp} \to \mathrm{H} \to 4\ell$ cross section measurements.

Lepton kinematics and isolation	
Leading lepton p_{T}	$p_{\mathrm{T}} > 20$ GeV
Subleading lepton p_{T}	$p_{\mathrm{T}} > 10$ GeV
Additional electrons (muons) p_{T}	$p_{\mathrm{T}} > 7(5)$ GeV
Pseudorapidity of electrons (muons)	$\lvert\eta\rvert < 2.5(2.4)$
Sum p_{T} of all stable particles within $\Delta R < 0.3$ from lepton	$< 0.35 p_{\mathrm{T}}$
Event topology	
Existence of at least two same-flavor OS lepton pairs, where leptons satisfy criteria above	
Invariant mass of the Z_1 candidate	40 GeV $< m(\mathrm{Z}_1) <$ 120 GeV
Invariant mass of the Z_2 candidate	12 GeV $< m(\mathrm{Z}_2) <$ 120 GeV
Distance between selected four leptons	$\Delta R(\ell_i \ell_j) > 0.02$ for any $i \neq j$
Invariant mass of any opposite sign lepton pair	$m_{\ell^+\ell'^-} > 4$ GeV
Invariant mass of the selected four leptons	$105 < m_{4\ell} < 140$ GeV

Table 4.10 – Summary of the fraction of signal events for different SM signal production modes within the fiducial phase space (acceptance $\mathcal{A}_{\mathrm{fid}}$), reconstruction efficiency (ϵ) for signal events from within the fiducial phase space, and ratio of reconstructed events which are from outside the fiducial phase space to reconstructed events which are from within the fiducial phase space (f_{nonfid}). For all production modes the values given are for $m_{\mathrm{H}} = 125$ GeV. Also shown in the last column is the factor $(1 + f_{\mathrm{nonfid}})\epsilon$ which regulates the signal yield for a given fiducial cross section, as shown in Eq. 4.19. The uncertainties listed are statistical only. The theoretical uncertainty in $\mathcal{A}_{\mathrm{fid}}$ for the SM is less than 1%.

Signal process	$\mathcal{A}_{\mathrm{fid}}$	ϵ	f_{nonfid}	$(1 + f_{\mathrm{nonfid}})\epsilon$
gg→H (POWHEG)	0.398 ± 0.001	0.592 ± 0.001	0.049 ± 0.001	0.621 ± 0.001
VBF (POWHEG)	0.445 ± 0.001	0.601 ± 0.002	0.038 ± 0.001	0.624 ± 0.002
WH (POWHEG+MINLO)	0.314 ± 0.001	0.577 ± 0.002	0.068 ± 0.001	0.616 ± 0.002
ZH (POWHEG+MINLO)	0.342 ± 0.002	0.592 ± 0.003	0.071 ± 0.002	0.634 ± 0.003
ttH (POWHEG)	0.311 ± 0.002	0.572 ± 0.003	0.136 ± 0.003	0.650 ± 0.004

targeting different production modes and does not use the $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$ observable to minimize the model dependence, but does use three different final states.

The number of expected events in each final state f and in each bin i of an observable

considered is expressed as a function of $m_{4\ell}$ as:

$$\begin{aligned} N^{\mathrm{f},i}_{\mathrm{exp}}(m_{4\ell}) &= N^{\mathrm{f},i}_{\mathrm{fid}}(m_{4\ell}) + N^{\mathrm{f},i}_{\mathrm{nonfid}}(m_{4\ell}) + N^{\mathrm{f},i}_{\mathrm{nonres}}(m_{4\ell}) + N^{\mathrm{f},i}_{\mathrm{bkg}}(m_{4\ell}) \\ &= \sum_{j} \epsilon^{\mathrm{f}}_{i,j} \left(1 + f^{\mathrm{f},i}_{\mathrm{nonfid}}\right) \sigma^{\mathrm{f},j}_{\mathrm{fid}} \mathcal{L} \mathcal{P}_{\mathrm{res}}(m_{4\ell}) \\ &\quad + N^{\mathrm{f},i}_{\mathrm{nonres}} \mathcal{P}_{\mathrm{nonres}}(m_{4\ell}) + N^{\mathrm{f},i}_{\mathrm{bkg}} \mathcal{P}_{\mathrm{bkg}}(m_{4\ell}). \end{aligned} \tag{4.19}$$

The shape of the resonant signal contribution $\mathcal{P}_{\mathrm{res}}(m_{4\ell})$ and non-resonant contribution $\mathcal{P}_{\mathrm{nonres}}(m_{4\ell})$ from WH, ZH, and $\mathrm{t\bar{t}H}$ production are modelled by a double-sided Crystal Ball function, and a Landau, respectively, as described in Section 4.6. The normalization of the resonant contribution is proportional to the fiducial cross section, and the non-resonant signal component is treated as background in this measurement.

The $\epsilon^{\mathrm{f}}_{i,j}$ factor represents the detector response matrix that maps the number of expected events in a given observable bin j at the fiducial level to the number of expected events in the bin i at the reconstruction level. This response matrix is measured using signal samples with full detector simulation and corrected for residual differences between data and simulation, and an example for ggH and VBF production modes can be seen in Fig. 4.33. This procedure accounts for the unfolding of detector effects from the observed distributions and is the same as in [52, 120]. In the case of the integrated fiducial cross section measurement the efficiencies reduce to single values, which for different SM production modes are listed in Table 4.10.

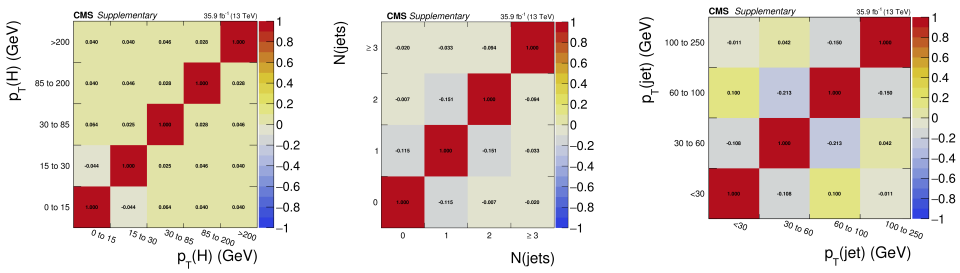

Figure 4.33 – Efficiency matrices for the $p_{\mathrm{T}}(\mathrm{H})$ (left), N(jets) (middle), and $p_{\mathrm{T}}(\mathrm{)jet)}$ (right) observables for the ggH production mode.

An additional resonant contribution arises from events which are reconstructed, but do not originate from the fiducial phase space, N_{nonfid}. These events are due to detector effects that cause differences between the quantities used for the fiducial phase space definition and the analogous quantities at the reconstruction level. This contribution is treated as background and is referred to as the *nonfiducial signal* contribution. The shape of these events is verified using signal samples with full detector simulation to be identical to the shape of the fiducial signal, and its normalization is fixed to be a fraction of the fiducial signal component. The value of this fraction, which we denote as f_{nonfid}, has been determined from signal samples with full detector simulation for each of the signal production modes studied. The value of f_{nonfid} for different signal production modes is reported in Table 4.10.

Mass and width measurements

The measurement of the Higgs boson mass combines the per-event mass uncertainty variable $\mathcal{D}_{\rm m}$, descibed in Section 4.4.3, with $m_{4\ell}$ and $\mathcal{D}^{\rm kin}_{\rm bkg}$ observables. A 1D, 2D and 3D pdfs are thus build as:

$$\begin{aligned} f^{m}_{\rm 1D}(m_{4\ell}) &= \mathcal{P}(m_{4\ell}), \\ f^{m}_{\rm 2D}(m_{4\ell}, \mathcal{D}_{\rm m}) &= \mathcal{P}(m_{4\ell}) \times \mathcal{P}(\mathcal{D}_{\rm m}), \\ f^{m}_{\rm 3D}(m_{4\ell}, \mathcal{D}_{\rm m}, \mathcal{D}^{\rm kin}_{\rm bkg}) &= \mathcal{P}(m_{4\ell}) \times \mathcal{P}(\mathcal{D}_{\rm m}) \times \mathcal{P}(\mathcal{D}^{\rm kin}_{\rm bkg}|m_{4\ell}). \end{aligned}$$

The templates for the $\mathcal{D}_{\rm m}$ variable are built for each final state, from simulation for the signal and irreducible backgrounds, and from the control region for $\rm Z + X$. While they should in principle be conditional upon the other variables, detailed studies showed that $\mathcal{D}_{\rm m}$ is not correlated to $\mathcal{D}^{\rm kin}_{\rm bkg}$, and it is found to have very little dependence on $m_{4\ell}$. Therefore, one dimensional $\mathcal{P}(\mathcal{D}_{\rm m})$ templates are used to extract shapes with fits to the sum of a Landau and a Gaussian function. No event categorization is used for the mass measurement, since the expected contribution from associated production modes, which have slightly different $m_{4\ell}$ lineshape, is found to be very small. There are thus only 3 channels, one for each final state.

To improve the $m_{4\ell}$ resolution further, a *kinematic fit* is performed using a mass constraint on the intermediate Z resonance. Previous studies [26] of the Higgs boson mass show that the selected $\rm Z_1$ has a significant on-shell component, while the invariant mass distribution of the selected $\rm Z_2$ is wider than the detector resolution. Therefore, only the $\rm Z_1$ candidate is considered when performing the kinematic constraint. The likelihood is constructed as:

$$\mathcal{L}(\hat{\mathbf{p}_{\rm T}}^1, \hat{\mathbf{p}_{\rm T}}^2 | \hat{\mathbf{p}_{\rm T}}^1, \sigma_{p^1_{\rm T}}, p^2_{\rm T}, \sigma_{p^2_{\rm T}}) = {\rm Gauss}(p^1_{\rm T}|\hat{\mathbf{p}_{\rm T}}^2, \sigma_{p^1_{\rm T}}){\rm Gauss}(p^2_{\rm T}|\hat{\mathbf{p}_{\rm T}}^2, \sigma_{p^2_{\rm T}})\mathcal{L}(m_{12}|m_{\rm Z}, m_{\rm H}), \tag{4.20}$$

where $p^1_{\rm T}$ and $p^2_{\rm T}$ are the reconstructed transverse momenta of the two leptons forming the $\rm Z_1$ candidate, $\sigma_{p^1_{\rm T}}$ and $\sigma_{p^2_{\rm T}}$ are the corresponding per-lepton resolutions, $\hat{\mathbf{p}_{\rm T}}^1$ and $\hat{\mathbf{p}_{\rm T}}^1$ are the refitted transverse momenta, m_{12} is the invariant mass calculated from the refitted four-momenta, and the term $\mathcal{L}(m_{12}|m_{\rm Z}, m_{\rm H})$ is the mass constraint term. For a Higgs boson mass near 125 GeV, the selected $\rm Z_1$ is not always on-shell, so a Breit-Wigner shape does not perfectly describe the $\rm Z_1$ shape at the generator level. We therefore choose $\mathcal{L}(m_{12}|m_{\rm Z}, m_{\rm H})$ to be the $m(\rm Z_1)$ shape at the generator level from the SM Higgs boson sample with $m_{\rm H} = 125$ GeV, where the same algorithm for selecting the $\rm Z_1$ and $\rm Z_2$ candidates, as described in Section 4.2, is used. For each event, the likelihood is maximized and the refitted transverse momenta are used to recalculate the four-lepton mass and mass uncertainty, which are denoted as $m^{'}_{4\ell}$ and $\mathcal{D}^{'}_{\rm m}$, respectively. The comparison of the reconstructed and refitted four-lepton invariant mass can be seen in Fig. 4.34, where a clear gain in resolution is visible in the refitted case. These refitted distributions are then used to build the pdfs for $m^{'}_{4\ell}$ and $\mathcal{D}^{'}_{\rm m}$ observables, used in the likelihood to extract the Higgs boson mass.

Thanks to the fact that the $m_{4\ell}$ pdfs are parametrized as a function of the Higgs boson

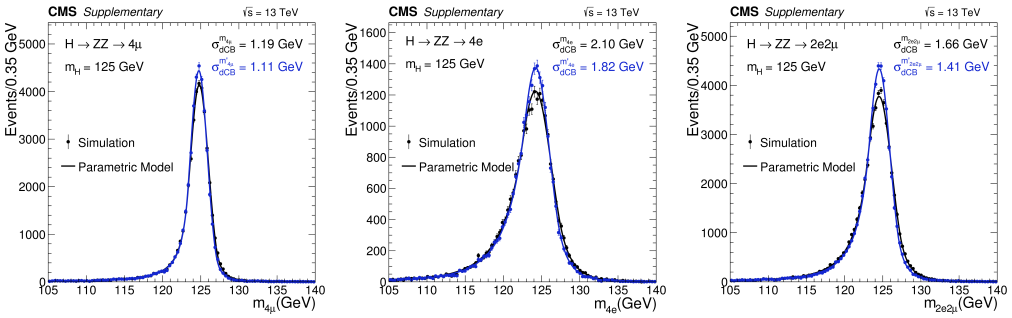

Figure 4.34 – Comparison of the reconstructed $m_{4\ell}$ and refitted $m^{'}_{4\ell}$ in the 4μ (left), $4e$ (middle), and $2e2\mu$ (right) final states of $m_{\mathrm{H}} = 125$GeV signal MC sample. A fit to the double Crystall Ball is shown with resolution parameters extracted from the fit reported on the figure.

width Γ_{H}, it can also be measured with a 1D pdf:

$$f^{\Gamma}_{\mathrm{1D}}(m_{4\ell}) = \mathcal{P}(m_{4\ell}), \tag{4.21}$$

with $f^{\Gamma}_{\mathrm{1D}}(m_{4\ell})$ being identical to $f^{m}_{\mathrm{1D}}(m_{4\ell})$. The only difference is that for the width measurement we profile the Γ_H parameter as well.

4.8.3 Treatment of systematic uncertainties

The factors $p_i(\tilde{\theta}_i|\theta_i)$ in Eq. 4.15 represent pdfs for the measurements associated to each nuisance parameter, i.e. the probabilities of measuring a value $\tilde{\theta}_i$ given the true value θ_i. The different sources of systematic uncertainties are described in detail in Section 4.7. Each independent source of uncertainty is assigned a nuisance parameter θ_i, which is profiled during the minimization. It can affect either one or multiple processes, and in different ways depending on the final state or the event category. As we already saw, the exact way and magnitude have to be determined and quantified from an auxiliary study. All sources of uncertainties are taken to be either 100% *correlated*, either positively or negatively, or *uncorrelated* in which case an independent nuisance parameter is assigned to the uncertainty. There are two ways to treat partially correlated uncertainties. They can be broken down to subcomponents that can be said to be either 100% correlated or uncorrelated. In the case of very small or very large correlations, they can be declared as fully uncorrelated or correlated, whichever is believed to be appropriate or more conservative. This allows us to include all constraints in the likelihood functions in a clean factorized form.

Two classes of uncertainties are considered here. First, normalization uncertainties only affect the expected yields of some processes. Their associated nuisance parameters simply scale the yield in a multiplicative way, and their pdf $p_i(\tilde{\theta}_i|\theta_i)$ follows a log-normal distribution. Second, shape uncertainties alter the shapes of the pdfs of some of the considered observables. This is modelled by considering a family of alternative templates, which is usually governed by one single nuisance parameter that itself follows a Gaussian pdf.

Sometimes, it is useful to understand contributions of statistical and systematical uncertainties. To calculate this, a likelihood scan is performed removing the systematical uncertainties to determine the statistical uncertainty. The systematic uncertainty is then taken as the difference in quadrature between the total uncertainty and the statistical uncertainty.

Chapter 5

Measurement of Higgs boson properties

After a fruitful Run I where the elusive Higgs boson was finally discovered and studied, Run II was supposed to provide enough data to discover its main production and decay modes, as well as to probe its properties to a new precision. In the scope of the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis in the CMS Collaboration, already with the first Run II data with only 2.8 fb^{-1} recorded in 2015, an evidence for Higgs boson was found [78] for the first time at $\sqrt{s} = 13$ TeV. I have contributed to that analysis with the derivation of electron scale factors described in Section 3.3. However, there was not enough data to rediscover the Higgs boson at the new centre-of-mass energy or to measure its properties, so everyone was eagerly awaiting 2016 with a lot more recorded data expected. Already on a partial data sample of 2016, collected between April and July and corresponding to an integrated luminosity of 12.9 fb^{-1}, Higgs boson was rediscovered and a first set of its properties was measured [121]. I was starting to become more involved in other aspects of the analysis, working on the signal normalization parametrization described in Section 4.6 and being involved in running the analysis selection chain to provide the inputs to all results. The analysis evolved using the full 2016 data set corresponding to 35.9 fb^{-1}. Thanks to my contributions to the analysis I was chosen to give the pre-approval talk in the CMS Higgs Physics Analysis Group (PAG). Afterwards, the analysis got approved by the CMS Collaboration and finally published [122] with a set of impressive results that will be presented in Section 5.1.

Continuing to accumulate more data in Run II, in 2017 the CMS detector collected another 41.5 fb^{-1} of data. With the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ being one of the key channels for Higgs boson studies in the CMS Collaboration, it was important to showcase the new improvements being introduced to the analysis and that the newly collected data is well calibrated and understood in new harsh conditions with record breaking pileup. That is why it was decided to perform a preliminary study of the newly collected data, measuring only some basic properties of the Higgs boson.

I have contributed strongly to this analysis, being involved in every aspect from the object selection, background and signal modeling, running the full chain of analysis selection defined in Chapter 4, producing the observed and expected distributions for all main variables and finally performing the measurements. I was chosen as a contact person of the analysis and as a main author, also giving the approval talk of the recently released CMS PAS [123]. It is the first result on Higgs properties from the CMS Collaboration based on data collected in 2017. I have presented the analysis results in the *39th International Conference on High Energy Physics* in Seoul on July 6^{th}, 2018.

5.1 Measurement of Higgs boson properties in 2016 data

5.1.1 Results of event selection

The 2016 analysis was based on re-reconstructed primary data sets listed in Table 4.1 with a total integrated luminosity of $35.9\ \text{fb}^{-1}$, and a full list of trigger paths used to select event candidates is presented in Table 4.2. More details can be found in Section 4.1.1.

After a full event selection is performed, data were compared to MC simulation in the signal region. In Table 5.1, the number of expected background and signal events and the number of events observed in the 2016 data is reported in the full mass range for all three considered final states. Signal and irreducible background distributions are obtained from simulation and reducible background is estimated from data in control regions combining two different methods as explained in Section 4.5.2. Reducible background prediction in all three final states was found to be in agreement between the two methods, within their uncertainties. In total, 1479 events are selected for 1417^{+89}_{-94} expected. For each final state, there is a very good agreement between data and MC simulation, giving confidence in all of the careful studies on corrections and scale factors.

Table 5.1 – The number of expected background and signal events and number of observed candidates after full analysis selection, for each final state, for the full mass range $m_{4\ell} > 70$ GeV and for an integrated luminosity of $35.9\ \text{fb}^{-1}$. Signal and ZZ backgrounds are estimated from Monte Carlo simulation, Z + X is estimated from data. The uncertainties include both statistical and systematic sources.

Channel	$4e$	4μ	$2e2\mu$	4ℓ
$\text{q}\bar{\text{q}} \to \text{ZZ}$	193^{+19}_{-20}	360^{+25}_{-27}	471^{+33}_{-36}	1024^{+69}_{-76}
$\text{gg} \to \text{ZZ}$	$41.2^{+6.3}_{-6.1}$	$69.0^{+9.5}_{-9.0}$	102^{+14}_{-13}	212^{+29}_{-27}
Z + X	$21.1^{+8.5}_{-10.4}$	34^{+14}_{-13}	60^{+27}_{-25}	115^{+32}_{-30}
Sum of backgrounds	255^{+24}_{-25}	463^{+32}_{-34}	633^{+44}_{-46}	1351^{+86}_{-91}
Signal ($m_{\text{H}} = 125$ GeV)	$12.0^{+1.3}_{-1.4}$	23.6 ± 2.1	30.0 ± 2.6	65.7 ± 5.6
Total expected	267^{+25}_{-26}	487^{+33}_{-35}	663^{+46}_{-47}	1417^{+89}_{-94}
Observed	293	505	681	1479

In Table 5.2, the number of observed events for 2016 data is also compared to the number of expected background and signal events, for each category and in a narrow mass range $118 < m_{4\ell} < 130$ GeV around the Higgs peak. Overall very good agreement is observed in all event categories, with a very slight downward fluctuation in the VBF-2jet-tagged category, where 4 events are observed for $5.32^{+0.78}_{-0.65}$ expected. This proved to be very important for the signal strength modifier measurement as will be discussed in Section 5.1.2. No $\mathrm{t\bar{t}H}$ signal candidates are observed for $0.79^{+0.14}_{-0.12}$ events expected in the $\mathrm{t\bar{t}H}$-tagged category.

Table 5.2 – The number of expected background and signal events and number of observed candidates after full analysis selection, for each event category, for the mass range $118 < m_{4\ell} < 130$ GeV and for an integrated luminosity of $35.9\ \mathrm{fb}^{-1}$. The yields are given for the different production modes. Signal and ZZ backgrounds are estimated from Monte Carlo simulation, Z + X is estimated from data. The uncertainties include both statistical and systematic sources.

	Event Category							
	Untagged	VBF-1j	VBF-2j	VH-hadr.	VH-lept.	VH-E_T^{miss}	$\mathrm{t\bar{t}H}$	Inclusive
$\mathrm{q\bar{q}} \to \mathrm{ZZ}$	19.18	2.00	0.25	0.30	0.27	0.01	0.01	22.01
$\mathrm{gg} \to \mathrm{ZZ}$	1.67	0.31	0.05	0.02	0.04	0.01	< 0.01	2.09
Z + X	10.79	0.88	0.78	0.31	0.17	0.30	0.27	13.52
Sum of backgrounds	31.64	3.18	1.08	0.63	0.49	0.32	0.28	37.62
Uncertainties	$^{+4.30}_{-3.42}$	$^{+0.37}_{-0.32}$	$^{+0.29}_{-0.21}$	$^{+0.13}_{-0.09}$	$^{+0.07}_{-0.07}$	$^{+0.14}_{-0.11}$	$^{+0.09}_{-0.07}$	$^{+5.19}_{-4.18}$
ggH	38.78	8.31	2.04	1.41	0.08	0.02	0.10	50.74
VBF	1.08	1.14	2.09	0.09	0.02	< 0.01	0.02	4.44
WH	0.43	0.14	0.05	0.30	0.21	0.03	0.02	1.18
ZH	0.41	0.11	0.04	0.24	0.04	0.07	0.02	0.93
$\mathrm{t\bar{t}H}$	0.08	< 0.01	0.02	0.03	0.02	< 0.01	0.35	0.50
Signal	40.77	9.69	4.24	2.08	0.38	0.11	0.51	57.79
Uncertainties	$^{+3.69}_{-3.62}$	$^{+1.13}_{-1.17}$	$^{+0.55}_{-0.55}$	$^{+0.23}_{-0.23}$	$^{+0.03}_{-0.03}$	$^{+0.01}_{-0.02}$	$^{+0.06}_{-0.06}$	$^{+4.89}_{-4.80}$
Total expected	72.41	12.88	5.32	2.71	0.86	0.43	0.79	95.41
Uncertainties	$^{+7.35}_{-6.27}$	$^{+1.25}_{-1.21}$	$^{+0.78}_{-0.65}$	$^{+0.34}_{-0.28}$	$^{+0.10}_{-0.09}$	$^{+0.15}_{-0.12}$	$^{+0.14}_{-0.12}$	$^{+9.86}_{-8.32}$
Observed	73	13	4	2	1	1	0	94

Data and MC simulation can also be compared visually for a variety of interesting observables that can differentiate signal over the SM background. The reconstructed four-lepton invariant mass distribution is shown in Fig. 5.1 and compared with the SM expectations from signal and background. The error bars on the data points correspond to the so-called Garwood confidence intervals at 68% CL [124]. The observed distribution shows impressive agreement with the SM expectation obtained from simulation, over the whole mass spectrum. In Fig. 5.2, the same can be concluded for the well populated Untagged, VBF-2jet-tagged, VBF-1jet-tagged and VH-hadronic tagged categories, with the rest of the event categories still being statistically limited.

In addition, distributions of other interesting kinematic observables are presented. While for the $m_{4\ell}$ variable, the Z + X normalization and analytical shapes come from the combination of the OS and SS methods described in Section 4.5.2. For all other variables, the

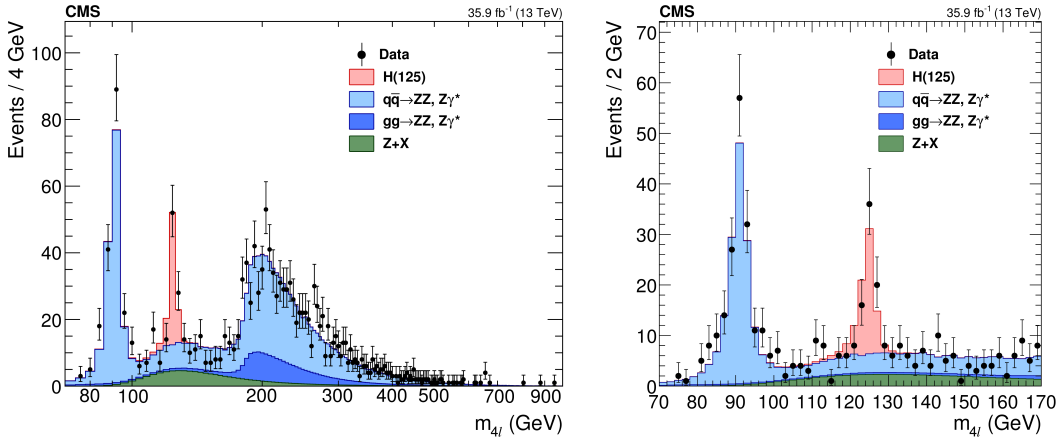

Figure 5.1 – Distribution of the four-lepton reconstructed invariant mass $m_{4\ell}$ in the full mass range (left) and the low-mass range (right). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data. No events are observed with $m_{4\ell} > 1$ TeV.

distributions of $\mathrm{Z} + \mathrm{X}$ are estimated using the SS method, and then renormalized to the combined yield without the use of analytical shapes. The reconstructed invariant masses of lepton pairs selected as Z_1 and Z_2 are shown in Fig. 5.3 in the $118 < m_{4\ell} < 130$ GeV mass range, together with their correlation. Again, a good agreement is seen for the SM prediction that includes the 125 GeV Higgs boson. The correlation between the two variables used in the likelihood fit to extract the results, namely the kinematic discriminant $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$ and the four-lepton invariant mass $m_{4\ell}$ is shown in Fig. 5.4. The gray scale represents the expected contributions of the ZZ background and the Higgs boson signal. The points represent the data and horizontal bars represent the third observable used in the likelihoods, the four-lepton mass uncertainty $\mathcal{D}_{\mathrm{m}}$. This plot is the closest visual representation to the measurements, while keeping in mind that the signal and backgrounds are parametrized as analytical shapes in the likelihoods used to extract results, as explained in Section 4.8. Different marker colors and styles are used to denote the final state and the categorization of the events, respectively. A lot of information can be extracted from this plot, for example one can see that two observed events in the VH-$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$-tagged and $\mathrm{t\bar{t}H}$-tagged categories (empty star and square markers) have low values of $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$, implying that these events are more compatible with the background than the signal hypothesis.

Finally, the distribution of the discriminants used for event categorization and the corresponding working points are shown in Fig. 5.5, with the SM Higgs boson signal being separated into the production mode which is targeted by the specific discriminant and other production modes.

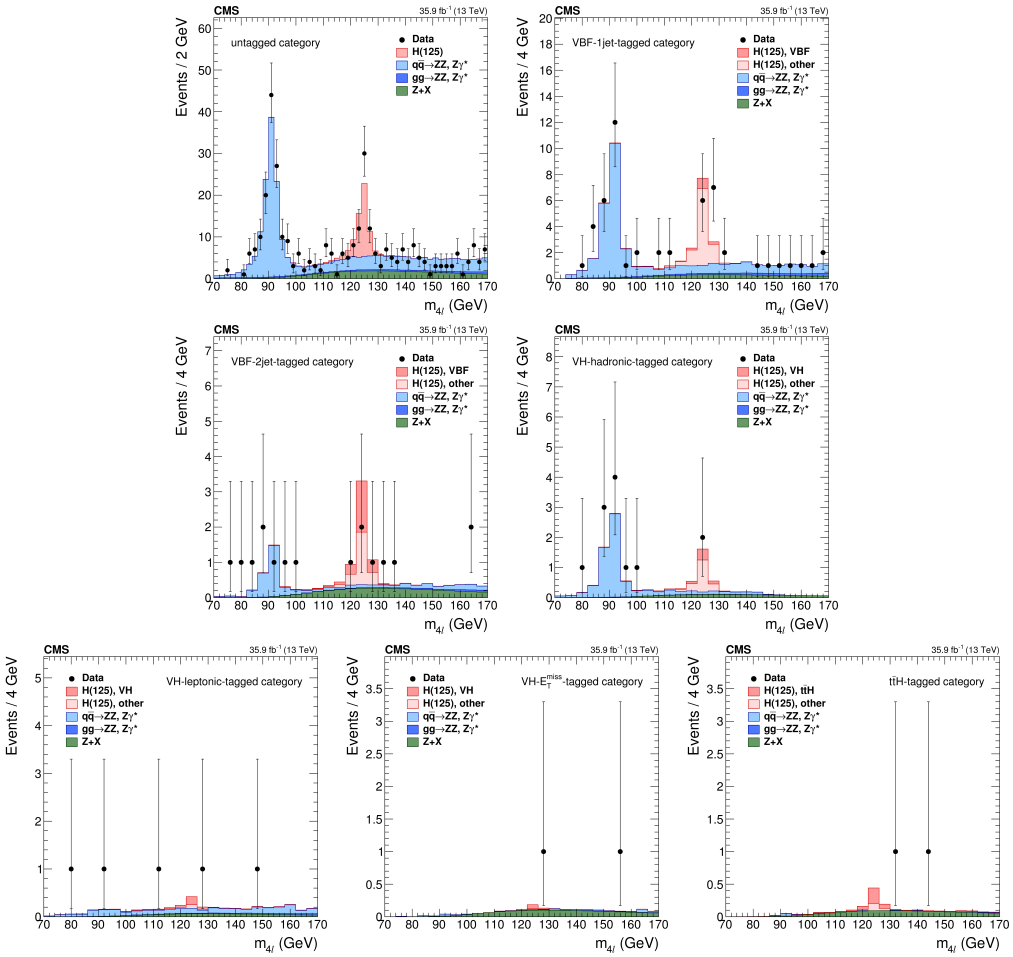

Figure 5.2 – Distribution of the four-lepton reconstructed mass in the seven event categories for the low-mass range in Untagged category (top left,) VBF-1jet-tagged category (top right), VBF-2jet-tagged category (center left), VH-hadronic-tagged category (center right), VH-leptonic-tagged category (bottom left), VH-E_T^{miss}-tagged category (bottom middle), and $t\bar{t}H$-tagged category (bottom right). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_H = 125$ GeV, denoted as $H(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $Z + X$ background to the estimation from data. For the categories other than the untagged category, the SM Higgs boson signal is separated into two components: the production mode that is targeted by the specific category, and other production modes, where the gluon fusion process dominates.

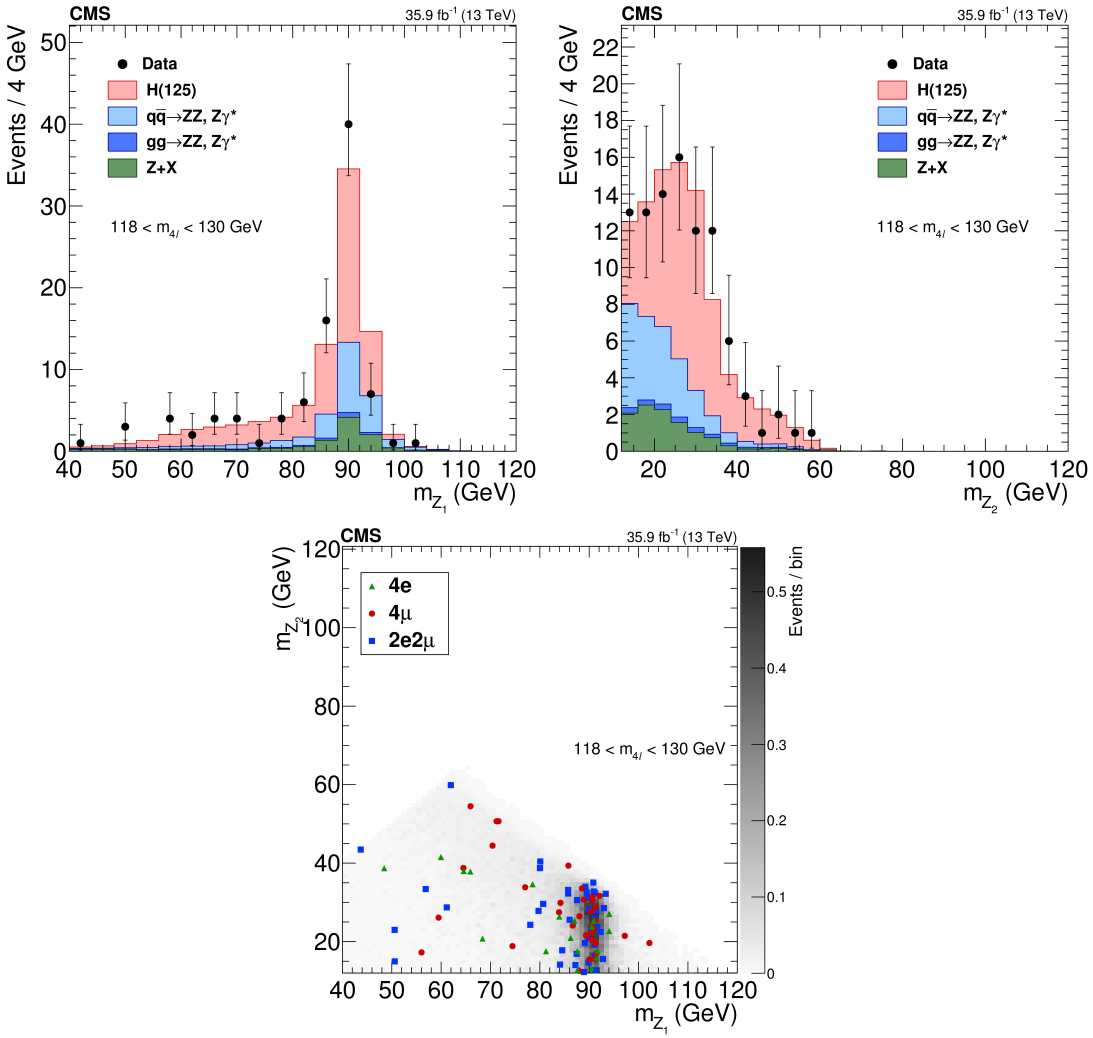

Figure 5.3 – Distribution of the Z_1 (top left) and Z_2 (top right) reconstructed invariant masses and correlation between the two (bottom) in the mass region $118 < m_{4\ell} < 130$ GeV. The stacked histograms and the gray scale represent expected distributions of the signal and background processes, and points represent the data. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data.

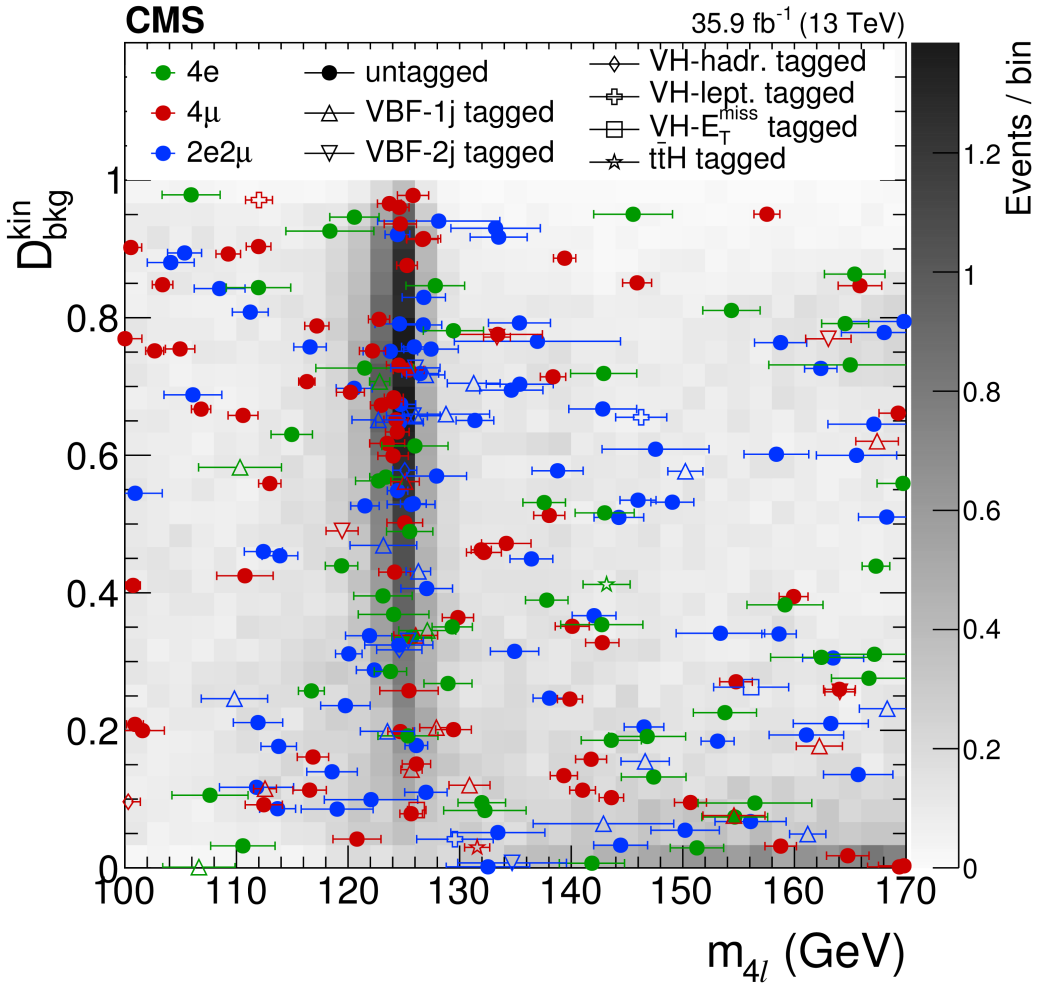

Figure 5.4 – Distribution of $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ versus $m_{4\ell}$ in the mass region $100 < m_{4\ell} < 170$ GeV. The gray scale represents the expected total number of ZZ background and SM Higgs boson signal events for $m_{\text{H}} = 125$ GeV. The points show the data and the horizontal bars represent $\mathcal{D}_{\text{m}}$. Different marker colors and styles are used to denote final state and the categorization of the events, respectively.

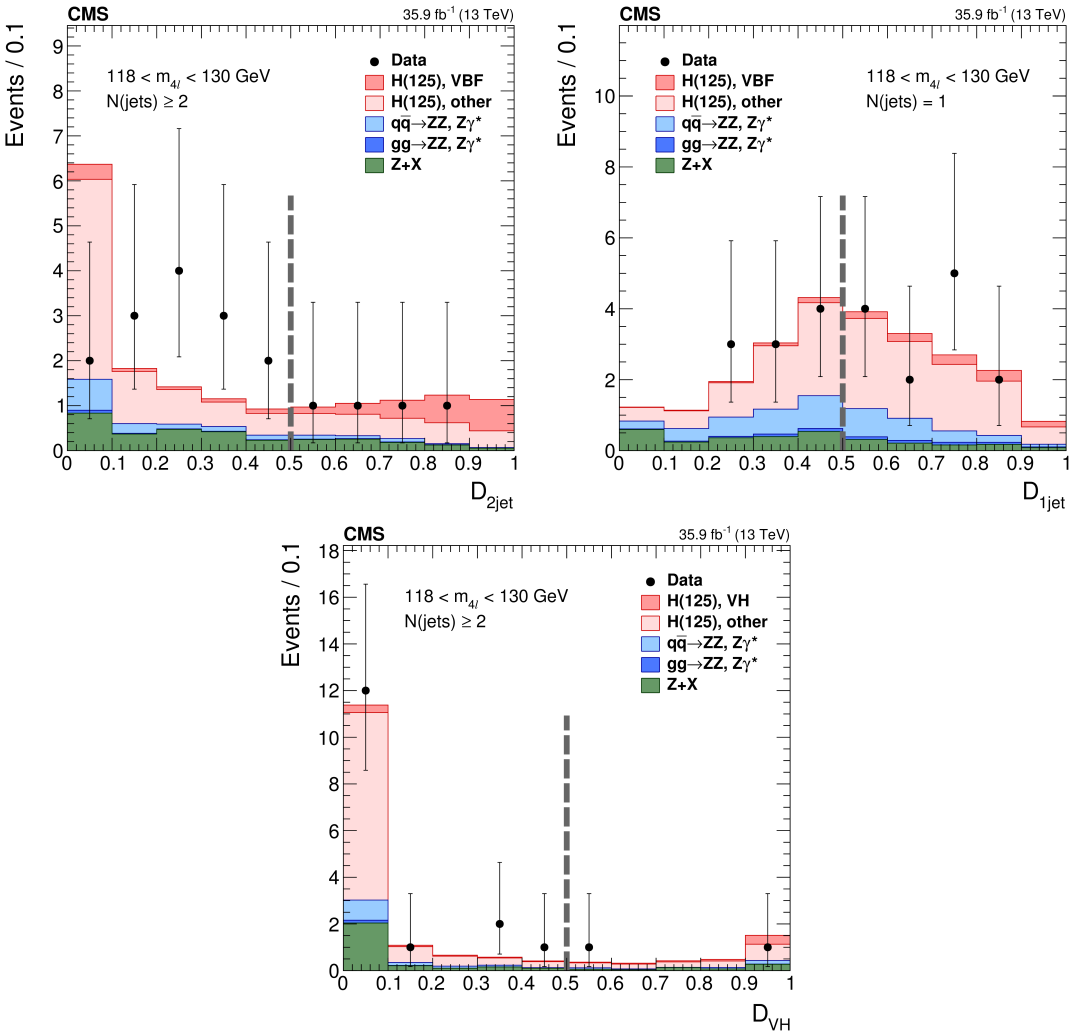

Figure 5.5 – Distribution of categorization discriminants in the mass region $118 < m_{4\ell} < 130$ GeV: $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}}$ (top left), $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}}$ (top right) , and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VH-hadr.}} = \max(\mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}, \mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}})$ (bottom). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data. The vertical gray dashed lines denote the working points used in the event categorization. The SM Higgs boson signal is separated into two components: the production mode which is targeted by the specific discriminant, and other production modes, where the gluon fusion process dominates.

5.1.2 Measurements

In general, a very good agreement is observed between data and MC expectations for the variety of different observables. Before proceeding with the results of several measurements performed, an overview of the values of the systematic uncertainties will be presented. Details on the measurement strategies were given in Section 4.8.2

Systematic uncertainties

The values of the systematic uncertainties used in the measurements of the 2016 analysis are summarized in Table 5.3. Details about the definition and computation of these uncertainties are given in Section 4.7, and how they are treated when extracting results is explained in Section 4.8.3.

Table 5.3 – Summary of the systematic uncertainties affecting measurements of the 2016 $H \to ZZ^{\star} \to 4\ell$ analysis. The uncertainties either affect the shape (S) or only the normalization (N) of predicted distributions, (p) denotes the uncertainties that are process-dependent, and (c) that the uncertainties are category-dependent.

Systematic	$4e$	4μ	$2e2\mu$
Luminosity (N)		2.5%	
Lepton Efficiency (N)	9%	5%	4.6%
QCD scale (N) (p) (c)		2 - 20%	
PDF set (N) (p) (c)		0 - 4%	
Branching ratio (N)		2%	
k-factor for $q\bar{q} \to ZZ$ (N) (c)		1 - 15%	
k-factor for $gg \to ZZ$ (N)		10%	
$Z + X$ estimation (N) (c)	43%	36%	38%
Lepton energy scale (S)	0.3%	0.04%	0.1%
Lepton energy resolution (S)	20%	10%	15%
Jet energy scale (N) (p) (c)		2 - 15%	
b-tagging (N) (p) (c)		0 - 6%	

From the experimental side, dominant sources of uncertainty come from luminosity and lepton efficiency, being 2.5% and 4.6 to 9% depending on the final state, respectively. For the mass measurement, lepton energy scale and resolution are dominant sources, as expected, raging from 0.04 to 0.3% and from 10 to 20%, respectively. Dominant sources from the theoretical side originate from the QCD scale and PDF set uncertainties for dominant gluon fusion production mode.

In the case of the measurements which use event categorization, the main sources of event migrations come from QCD scale and PDF set uncertainties, with combined effect

amounting to 4 to 20% for the signal and 3 to 20% for the background, depending on the category. Additional experimental sources like the imprecise knowledge of the jet energy scale amounts for 2 to 15%, and uncertainty on b-tagging efficiency amounts up to 6% in the $\mathrm{t\bar{t}H}$-tagged category.

Signal strength modifiers

A simultaneous fit to all channels is performed and a global signal strength modifier measured at the ATLAS and CMS Run I combined mass value of $m_{\mathrm{H}} = 125.09$ GeV is $\mu = \sigma/\sigma_{\mathrm{SM}} = 1.05^{+0.15}_{-0.14}(\mathrm{stat})^{+0.11}_{-0.09}(\mathrm{syst}) = 1.05^{+0.19}_{-0.17}$.

Because of several categories with very low expected yields, to keep the fit stable it was chosen to constrain all signal strength parameters to be non-negative. This result is compared to the measurement for each of the seven event categories in Fig. 5.6 (top left). All observed values are consistent with the SM prediction of $\mu = 1$, within the uncertainties.

In order to profit from event catagorization, a fit for five signal strength modifiers μ_{ggH}, μ_{VBF}, μ_{VHhad}, μ_{VHlep}, and $\mu_{\mathrm{t\bar{t}H}}$, that control the contributions of the main SM Higgs boson production modes, is performed. Again, all signal strength modifiers are constrained to positive values. The results are reported in Fig. 5.6 (top right) and compared to the expected signal strength modifiers in Table 5.4. The reason for the very low observed signal strengths for VBF, VH, and $\mathrm{t\bar{t}H}$ processes was studied and understood. Because of the mild excess of events in the untagged category which have higher values of $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$, it leads to a higher than expected signal strength for the gluon fusion that still contaminates significantly other categories with hadronic activity. However, in the categories that are not based on hadronic event activity, events with $m_{4\ell}$ near 125 GeV have low $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$ values favoring the background hypothesis. This drives the fit to recognize majority of events in hadronic categories as gluon fusion and drives all other contributions to zero. This is one of the big shortcomings of using the same $\mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}}$ observable as a second dimension for all production modes, given that it has no sensitivity to distinguish amongst them.

Two signal strength modifiers $\mu_{\mathrm{ggH,\,t\bar{t}H}}$ and $\mu_{\mathrm{VBF,\,VH}}$, representing scale factors to the fermion and vector boson induced contribution to the expected SM cross section are also measured. A two-parameter fit is performed simultaneously to all categories, again assuming a mass $m_{\mathrm{H}} = 125.09$ GeV, leading to measurements of $\mu_{\mathrm{ggH,\,t\bar{t}H}} = 1.19^{+0.21}_{-0.20}$ and $\mu_{\mathrm{VBF,\,VH}} = 0.00^{+0.81}_{-0.00}$. The SM expectation, the measured values, and the 68% and 95% CL contours in the ($\mu_{\mathrm{ggH,\,t\bar{t}H}}$,$\mu_{\mathrm{VBF,\,VH}}$) plane are shown in Fig. 5.6 (bottom). The SM prediction is found to lie within the 68% CL regions of this measurement.

Cross section measurements

In the following, various measurements of the cross section for Higgs boson production are presented. First measurement is very similar to the signal strength modifier mea-

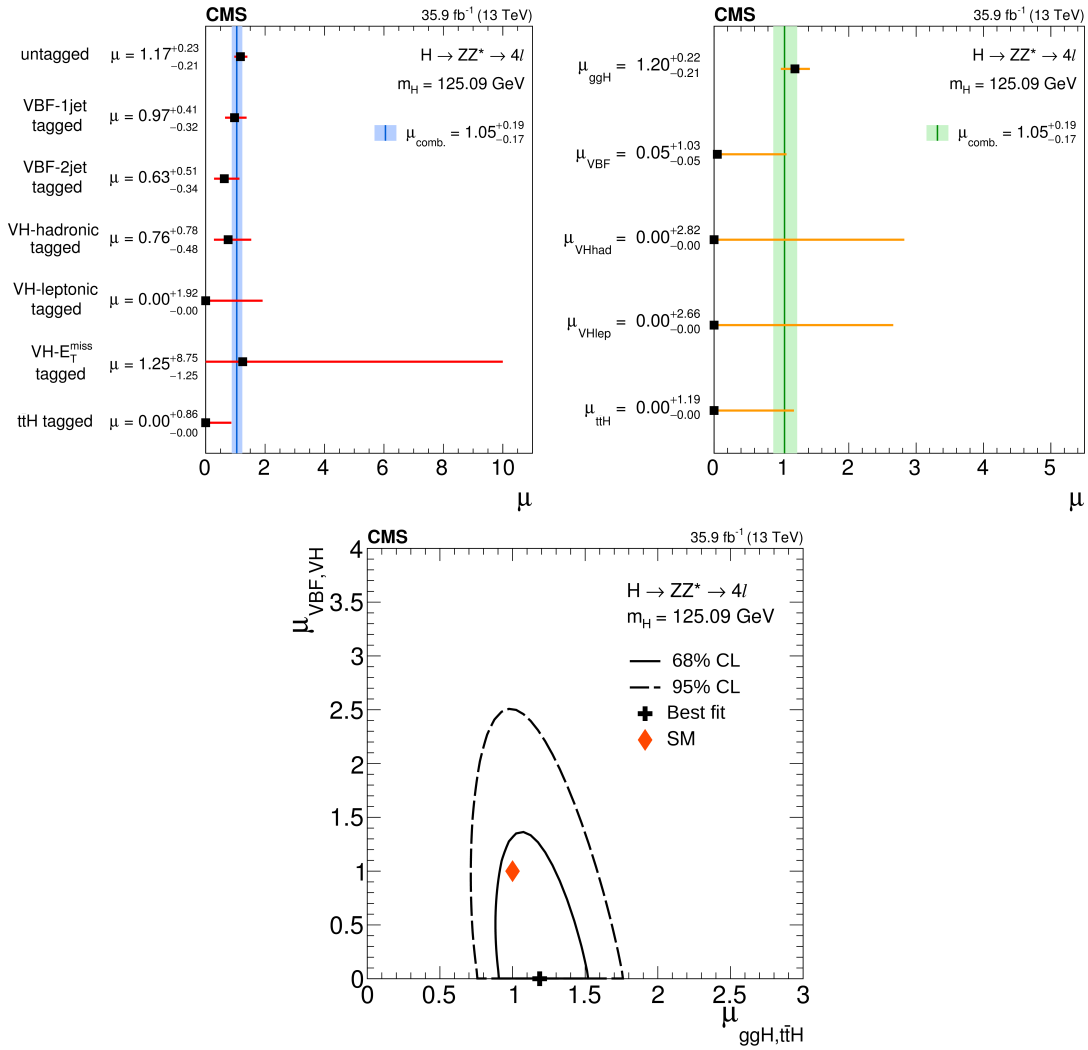

Figure 5.6 – (Top left) Observed values of the signal strength $\mu = \sigma/\sigma_{SM}$ for the seven event categories, compared to the combined μ shown as a vertical line. The horizontal bars and the filled band indicate the $\pm$ 1σ uncertainties. (Top right) Results of likelihood scans for the signal-strength modifiers corresponding to the main SM Higgs boson production modes, compared to the combined μ shown as a vertical line. The horizontal bars and the filled band indicate the $\pm$ 1σ uncertainties. The uncertainties include both statistical and systematic sources. (Bottom) Result of the 2D likelihood scan for the $\mu_{\mathrm{ggH,\,t\bar{t}H}}$ and $\mu_{\mathrm{VBF,\,VH}}$ signal-strength modifiers. The solid and dashed contours show the 68% and 95% CL regions, respectively. The cross indicates the best-fit values, and the diamond represents the expected values for the SM Higgs boson.

surement for different production modes, presented in the previous Section. The idea is to measure cross sections of five SM Higgs boson production processes σ_{ggH}, σ_{VBF}, σ_{VHhad}, σ_{VHlep}, and $\sigma_{\mathrm{t\bar{t}H}}$ in a simplified fiducial volume defined by just requiring selection $|y_{\mathrm{H}}| < 2.5$ on the Higgs boson rapidity. Studies have shown that outside of this volume $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis has a negligible acceptance. The separation of production modes is again achieved thanks to the event categorization. This measurement

Table 5.4 – Expected and observed signal-strength modifiers.

	Inclusive	μ_{ggH}	μ_{VBF}	μ_{VHhad}	μ_{VHlep}	$\mu_{\mathrm{t\bar{t}H}}$
Expected	$1.00^{+0.15}_{-0.14}(\mathrm{stat.})^{+0.10}_{-0.08}(\mathrm{sys.})$	$1.00^{+0.23}_{-0.21}$	$1.00^{+1.25}_{-0.97}$	$1.00^{+3.96}_{-1.00}$	$1.00^{+3.76}_{-1.00}$	$1.00^{+3.23}_{-1.00}$
Observed	$1.05^{+0.15}_{-0.14}(\mathrm{stat.})^{+0.11}_{-0.09}(\mathrm{sys.})$	$1.20^{+0.22}_{-0.21}$	$0.05^{+1.03}_{-0.05}$	$0.00^{+2.83}_{-0.00}$	$0.00^{+2.66}_{-0.00}$	$0.00^{+1.19}_{-0.00}$

corresponds to the *stage-0 simplified template cross section* measurement from Ref. [9], designed to reduce the dependence of the measurements on the theoretical uncertainties in the SM predictions, thus avoiding extrapolation of the measurements to the full phase space which carries nontrivial and sizable theoretical uncertainties. Measurement is performed using the identical procedure as for the signal strength modifiers, the only difference being that the theoretical uncertainties on affecting the signal normalization are removed, while those affecting category migrations and background normalizations are left. The measured cross sections normalized to the SM predictions, taken from [9] and denoted as σ_{theo}, are shown in Fig. 5.7.

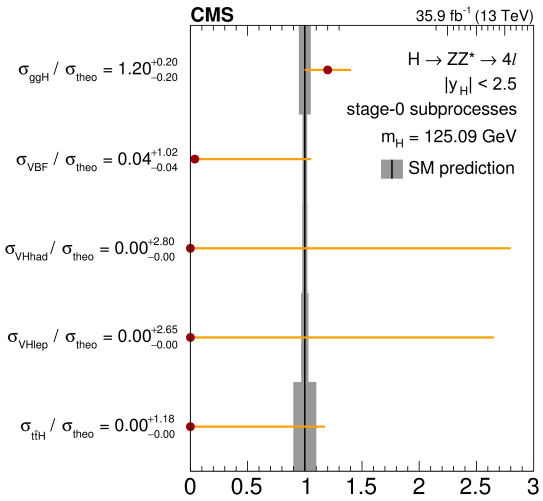

Figure 5.7 – Results of the fit for simplified template cross sections for the 'stage-0 subprocesses', normalized to the SM predictions. The grey bands indicate the theoretical uncertainties in the SM predictions. The orange error bars show the full uncertainty, including experimental uncertainties and theoretical uncertainties causing migration of events between the various categories.

Other cross section measurements are performed in a tight fiducial volume as explained in detail in Section 4.8.2. The results are compared to the predictions obtained from POWHEG and NNLOPS, with NLO and NNLO accuracy in pQCD for inclusive distributions, respectively. In both cases the total cross section for the dominant gluon fusion production mechanisms is taken from [8]. The integrated fiducial cross section is measured to be $\sigma_{\mathrm{fid}} = 2.92^{+0.48}_{-0.44}(\mathrm{stat})^{+0.28}_{-0.24}(\mathrm{syst})\,\mathrm{fb}$, in agreement with the SM prediction $\sigma^{\mathrm{SM}}_{\mathrm{fid}} = 2.76 \pm 0.14\,\mathrm{fb}$, obtained from NNLOPS. The integrated fiducial cross section as a function of $\sqrt{s}$ is shown in Fig. 5.8 (top left) including the Run I measurements and SM predictions. Additional final state compatibility checks were made by comparing

the measured integrated fiducial cross sections in the $4e$, 4μ, and $2e2\mu$ final state with the SM prediction. A likelihood ratio with the three cross sections at their best fit values in the numerator and the three cross sections fixed to the SM predictions in the denominator was used. The compatibility, defined as the asymptotic p-value of the fit, is found to be 88%. Finally, the measured differential cross section results for $p_{\mathrm{T}}(\mathrm{H})$, N(jets), and p_{T}(jet) of the leading associated jet is also shown in Fig. 5.8. In this measurement, systematic uncertainty was found to be dominated by experimental sources. To estimate the model dependence of the measurement, the unfolding procedure was done using different response matrices created by varying the relative fraction of the SM production mode within its experimental constrains. The uncertainty coming from this procedure was also found to be negligible with respect to the experimental systematic uncertainties.

Higgs boson mass measurement

Another interesting property to measure is the mass of the Higgs boson, which is not predicted by the Standard Model. With CMS and ATLAS results from Run I in good agreement and combined it was interesting to see what will first measurements on Run II data bring. As described in Section 4.8.2, besides the three observables $m_{4\ell}$, $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$, and $\mathcal{D}_{\mathrm{m}}$, a kinematic fit using a mass constraint on the selected Z_1 candidate is performed to improve the four-lepton invariant mass resolution. The 1D projections of the likelihood scans vs m_{H} are shown in Fig. 5.9 for the 1D $\mathcal{L}(m_{4\ell}^{'})$, 2D $\mathcal{L}(m_{4\ell}^{'}, \mathcal{D}_{\mathrm{m}}^{'})$, and 3D $\mathcal{L}(m_{4\ell}^{'}, \mathcal{D}_{\mathrm{m}}^{'}, \mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}})$ fits. The $m(\mathrm{Z}_1)$ mass constraint is included in the fit and signal strength modifier μ was profiled in the fit along with all other nuisance parameters described in Section 4.7. Just like for the signal strength measurement, the relative fraction of $4e$, 4μ, and $2e2\mu$ signal events is fixed to the SM prediction. Additional studies showed that this has no significant effect on the final result. The change in the fitted mass value, when allowing the fractions to float, is much smaller than the overall uncertainty on the measurement. One can see a clear decrease in the uncertainty as we include more dimensions.

The best fit masses and expected increase in the uncertainty relative to the most precise 3D fit with the $m(\mathrm{Z}_1)$ constraint for each of the six fits are reported in Table 5.5. The nominal result for the mass measurement is obtained from the 3D fit with the $m(\mathrm{Z}_1)$ constraint, for which the fitted value of m_{H} in the three subchannels is $m_{\mathrm{H}}^{4\mu} = 124.94 \pm 0.25(\mathrm{stat}) \pm 0.08(\mathrm{syst})$ GeV, $m_{\mathrm{H}}^{4e} = 124.37 \pm 0.62(\mathrm{stat}) \pm 0.38(\mathrm{syst})$ GeV, and $m_{\mathrm{H}}^{2e2\mu} = 125.95 \pm 0.32(\mathrm{stat}) \pm 0.14(\mathrm{syst})$ GeV leading to a combined value of:

$$m_{\mathrm{H}} = 125.26 \pm 0.20(\mathrm{stat}) \pm 0.08(\mathrm{syst}) \text{ GeV}. \tag{5.1}$$

This result is well compatible with the Run I LHC combined measurement [30] of the Higgs boson mass $m_{\mathrm{H}} = 125.09 \pm 0.21(\mathrm{stat}) \pm 0.11(\mathrm{syst})$ GeV. What is remarkable is that, thanks to the improvements in the lepton scale and resolution determination as well as in the mass measurement itself, already with first third of Run II data CMS has the most precise measurement of the Higgs mass ever, surpassing the LHC combined measurement of Run I. This is a very exciting prospect for the full Run II measurement and ultimately for a new combination between CMS and ATLAS expected soon afterwards.

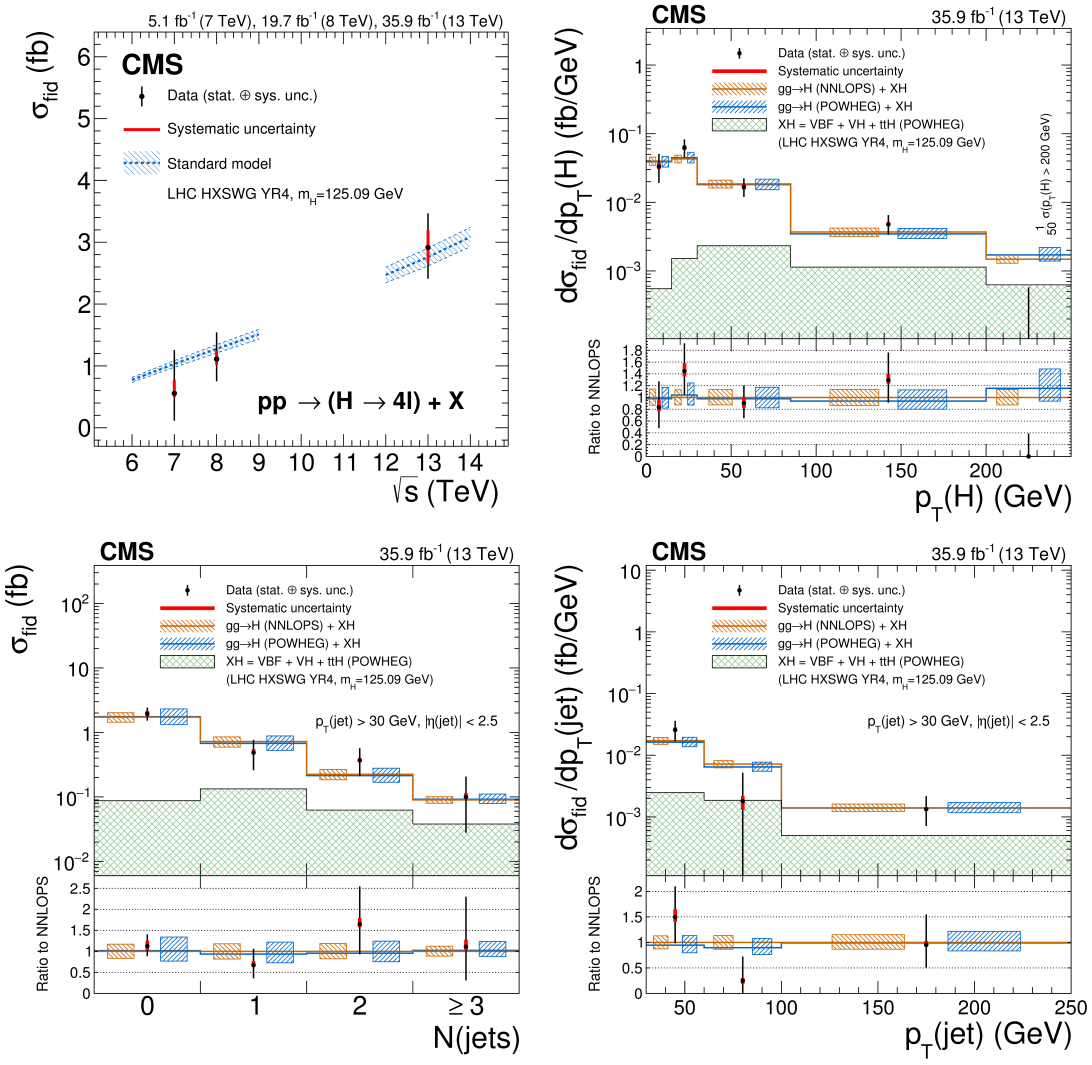

Figure 5.8 – The measured fiducial cross section as a function of $\sqrt{s}$ (top left). The acceptance is calculated using NNLOPS at $\sqrt{s} = 13$ TeV and HRES at $\sqrt{s} = 7$ and 8 TeV and the total cross sections and uncertainties are taken from ref. [9]. The fiducial volume for $\sqrt{s} = 7$ and 8 TeV uses the lepton isolation definition from ref. [52], while for $\sqrt{s} = 13$ TeV the definition described in the text is used. The results of the differential cross section measurements are shown for $p_{\mathrm{T}}(\mathrm{H})$ (top right), N(jets) (bottom left) and p_{T}(jet) of the leading associated jet (bottom right). The acceptance and theoretical uncertainties in the differential bins are calculated using POWHEG and NNLOPS. The subdominant component of the signal (VBF + VH + $\mathrm{t\bar{t}H}$) is denoted as XH. In the differential cross section measurement for $p_{\mathrm{T}}(\mathrm{H})$, the last bin represents the integrated cross section for $p_{\mathrm{T}}(\mathrm{H}) > 200$ GeV and is scaled by 1/50 for presentation purposes. No events are observed with $p_{\mathrm{T}}(\mathrm{H}) > 200$ GeV.

As the mass measurement is completely dominated by the uncertainty in the lepton momentum scale, several studies were performed to determine its value as explained in Section. 4.7. Additional checks were made by estimating the expected uncertainty in the mass measurement on the toy Asimov data sets, before and after the fit is performed.

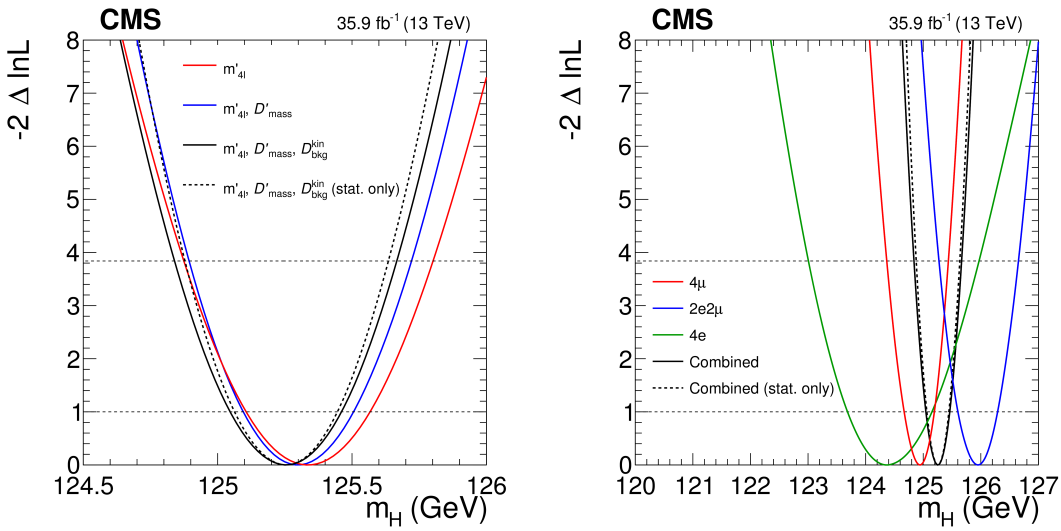

Figure 5.9 – 1D likelihood scans as a function of the Higgs boson mass for the 1D, 2D, and 3D measurement (left). 1D likelihood scans as a function of mass for the different final states and the combination of all final states for the 3D mass measurement (right). The likelihood scans are shown for the mass measurement using the refitted mass distribution with the $m(\mathrm{Z}_1)$ constraint. Solid lines represent scans with all uncertainties included, dashed lines those with only statistical uncertainties.

Table 5.5 – Best fit values for the mass of the Higgs boson measured in the 4ℓ, $\ell = e, \mu$ final states, with 1D, 2D and 3D fit, respectively. All mass values are given in GeV. The uncertainties are the total statistical plus systematic uncertainty. The expected m_{H} uncertainty change shows the change in expected precision on the measurement for the different fit scenarios, relative to 3D $\mathcal{L}(m'_{4l}, \mathcal{D}^{'}_{\mathrm{m}}, \mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}})$.

No $m(\mathrm{Z}_1)$ constraint	3D: $\mathcal{L}(m_{4l}, \mathcal{D}_{\mathrm{m}}, \mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}})$	2D: $\mathcal{L}(m_{4l}, \mathcal{D}_{\mathrm{m}})$	1D: $\mathcal{L}(m_{4l})$
Expected m_{H} uncertainty change	$+8.1\%$	$+11\%$	$+21\%$
Observed m_{H} (GeV)	125.28 ± 0.22	125.36 ± 0.24	125.39 ± 0.25
With $m(\mathrm{Z}_1)$ constraint	3D: $\mathcal{L}(m'_{4l}, \mathcal{D}^{'}_{\mathrm{m}}, \mathcal{D}^{\mathrm{kin}}_{\mathrm{bkg}})$	2D: $\mathcal{L}(m'_{4l}, \mathcal{D}^{'}_{\mathrm{m}})$	1D: $\mathcal{L}(m'_{4l})$
Expected m_{H} uncertainty change	—	$+3.2\%$	$+11\%$
Observed m_{H} (GeV)	125.26 ± 0.21	125.30 ± 0.21	125.34 ± 0.23

First, the *prefit* expected uncertainty is obtained for $m_{\mathrm{H}} = 125$ GeV, $\mu = 1$, and all nuisance parameters fixed to their nominal values reported in Table 5.3 yielding the expected uncertainty of $\pm 0.24(\mathrm{stat}) \pm 0.09(\mathrm{syst})$ GeV. To obtain the *postfit* expected uncertainty, m_{H}, μ, and all nuisance parameters are fixed to their best-fit estimates from the data, yielding a total expected uncertainty of $\pm 0.23(\mathrm{stat}) \pm 0.08(\mathrm{syst})$ GeV. The probability of the prefit uncertainty being less than or equal of the observed value is determined from an ensemble of pseudo-experiments and was found to be about 18%. Furthermore, the mutual compatibility of the m_{H} results in three individual final states is tested using a likelihood ratio with three masses in the numerator and a common

mass in the denominator with the signal strength profiled in both cases. The resulting compatibility, defined as the asymptotic p-value of the fit, is 2.5%. The tension between the three individual channels is driven by the difference in the 4μ and $2e2\mu$ channels, where the compatibility was checked with a simple 1D mass measurement without the $m(\mathrm{Z}_1)$ constraint and was found to be 8%. Extensive studies were made to check for any possible biases like for example the lepton momentum scale or the event-by-event mass uncertainty and everything was found to be consistent.

Higgs boson width measurement

Thanks to the fact that the 1D likelihood used for the mass measurement is also parametrized as a function of the Higgs boson width Γ_H, one can easily extend the measurement by profiling it as well. This yields a model independent measurement in the $105 < m_{4\ell} < 140$ GeV that accounts for the interference between the signal and background production of the 4ℓ final state. However, it is limited by the four-lepton invariant mass resolution and is therefore sensitive to a width of about 1 GeV. With a predicted value for the SM Higgs boson of $\Gamma_\mathrm{H} = 4.07$ MeV it is clear that with the current statistics this measurement will not be sensitive to it in the near future. Because of this, an indirect measurement was developed exploiting the ofshell region as well and profiting from the fact that the SM Higgs boson width is proportional to the ratio of the event yields in the offshell and onshell regions. This measurement was performed only on the first part of the 2016 dataset and a constraint of $\Gamma_\mathrm{H} < 41$ MeV at 95% CL was observed [121].

Coming back to the measurement using the full 2016 data set and the onshell region only, a fit is performed with the strengths of fermion and vector boson induced couplings independent and left unconstrained in the fit. Figure 5.10 shows the joint constrain on the width Γ_H and mass m_H of the Higgs boson (left) and the likelihood as a function of Γ_H with the m_H parameter unconstrained (right). The width is constrained to be $\Gamma_\mathrm{H} < 1.10$ GeV at 95% CL. The observed and expected results are summarized in Table 5.6. As expected, the dominant source when determining the width is the uncertainty on the four-lepton mass resolution.

Table 5.6 – Summary of allowed 68% CL (central values with uncertainties) and 95% CL (ranges in square brackets) intervals on the width Γ_H of the Higgs boson. The expected results are quoted for the SM signal production cross section ($\mu_{\mathrm{VBF,\,VH}} = \mu_{\mathrm{ggH,\,t\bar{t}H}} = 1$) and the values of $m_\mathrm{H} = 125$ GeV. In the observed results $\mu_{\mathrm{VBF,\,VH}}$ and $\mu_{\mathrm{ggH,\,t\bar{t}H}}$ are left unconstrained in the fit.

Parameter	$m_{4\ell}$ range	Expected	Observed
Γ_H (GeV)	$[105, 140]$	$0.00^{+0.75}_{-0.00}$ $[0.00, 1.60]$	$0.00^{+0.41}_{-0.00}$ $[0.00, 1.10]$

5.1.3 Summary

A first set of comprehensive results based on 35.9 fb^{-1} collected in 2016 with the CMS detector is presented. This is a first set of legacy results for the $\mathrm{H} \to \mathrm{ZZ}^\star \to 4\ell$ decay at $\sqrt{s} = 13$ TeV energy. The global signal strength modifier, defined as the ratio of the observed Higgs boson rate to the Standard Model expectation is measured to

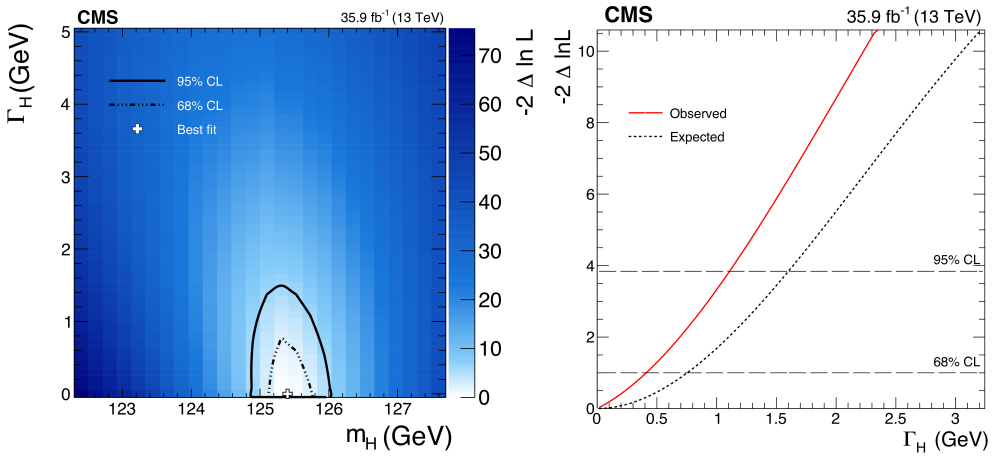

Figure 5.10 – Observed likelihood scan of m_{H} and Γ_{H} using the signal range $105 < m_{4\ell} < 140$ GeV (left). Observed and expected likelihood scan of Γ_{H} using the signal range $105 < m_{4\ell} < 140$ GeV, with m_{H} profiled (right).

be $\mu = 1.05^{+0.19}_{-0.17}$ at $m_{\mathrm{H}} = 125.09$. First measurements of the signal strength modifiers for all main Higgs boson production modes were made using this decay channel. Even though due to fluctuations in data all but the main gluon fusion production modes were measured to be 0 (or very close to 0), they are still found to be compatible with the SM prediction $\mu = 1$ within uncertainties. The fiducial cross section is measured to be $2.92^{+0.48}_{-0.44}(\mathrm{stat})^{+0.28}_{-0.24}(\mathrm{syst})\,\mathrm{fb}$, in agreement with the SM prediction $2.76 \pm 0.14\,\mathrm{fb}$. Differential cross sections as a function of transverse momentum of the Higgs boson, the number of associated jets, and the transverse momentum of the leading associated jet ere all found to show good agreement between observed data and the expectations from the SM. The most precise measurement of the Higgs boson mass is measured to be $m_{\mathrm{H}} = 125.26 \pm 0.20(\mathrm{stat}) \pm 0.08(\mathrm{syst})$ GeV, surpassing the precision of the combined LHC Run I mass measurement. The width is constrained using the onshell invariant mass distribution only, yielding a $\Gamma_{\mathrm{H}} < 1.10$ GeV constraint, at 95% confidence level.

5.2 Measurement of Higgs boson properties in 2017 data

This Section is dedicated to the results based on the 2017 data, while the combination with 2016 data will be presented in Section 5.3.

With respect to the analysis of 2016 data, electron identification is improved with the inclusion of the isolation variables in the BDT resulting in rejection of more fakes, the $\mathrm{t\bar{t}H}$-tagged category is split to two new categories targeting its hadronic and leptonic decay signatures, and finally new matrix based kinematic discriminants are developed and used as observables when performing the measurements.

5.2.1 Results of event selection

The 2017 analysis was based on prompt reconstructed primary data sets listed in Table 4.1 with a total integrated luminosity of $41.5\ \mathrm{fb}^{-1}$, and a full list of trigger paths used to select event candidates is presented in Table 4.2. More details can be found in Section 4.1.1. Because of the data still not being re-reconstructed, precise calibrations of the momentum scale and resolution were not possible, leading to a decision not to perform a full set of properties measurements at this stage.

Observed data compared to the MC simulation predictions for signal and background, after the full event selection is applied, is shown in Table 5.7 for the analysis based on the 2017 data set. Signal and irreducible background distributions are obtained from simulation and reducible background is estimated from data in control regions combining two different methods as explained in Section 4.5.2. One can see that for each final state, there is a very good agreement between the observed number of events and MC predictions, where 307 events in the $4e$ final state are observed for 315^{+41}_{-45} expected, 602 events in 4μ are observed for 589^{+45}_{-49} expected, and 797 events in $2e2\mu$ are observed for 777^{+64}_{-67} expected, yielding a total 1706 events observed for 1681^{+131}_{-140}.

When comparing results to the 2016 analysis shown in Table 5.1 one can see that expected yields for signal and irreducible backgrounds roughly scale with luminosity, as expected. The only thing that does not scale is the reducible background. In the 4μ final state it is only slightly increased, proving that the upgrade of the pixel detector described in Section 2.3.3 helped reduce the number of fake muons that pass the analysis selection cuts. Furthermore, the expected number of $\mathrm{Z+X}$ events in the 2017 analysis in the $4e$ final state is even smaller than for the 2016 analysis, profiting more from the pixel detector upgrade and in addition to the improved BDT identification that now includes isolation variables as explained in Section 3.2.4. These results are even more impressive considering the significant increase in the number of pileup interactions in each collision that is visible in Fig. 4.1.

Table 5.8 shows the number of expected background and signal events and observed events in 2017 data for each category and in a narrow mass range $118 < m_{4\ell} < 130$ GeV around the Higgs boson peak. A slight upwards fluctuation of events is present in the Untagged category, where for $84.76^{+6.52}_{-6.71}$ expected events 103 are observed. Otherwise,

Table 5.7 – The number of expected background and signal events and number of observed candidates after full analysis selection, for each final state, for the full mass range $m_{4\ell} > 70$ GeV and for an integrated luminosity of $41.5\ \mathrm{fb}^{-1}$. Signal and ZZ backgrounds are estimated from Monte Carlo simulation, Z + X is estimated from data. The uncertainties include both statistical and systematic sources.

Channel	$4e$	4μ	$2e2\mu$	4ℓ
$\mathrm{q\bar{q}} \to \mathrm{ZZ}$	235^{+32}_{-36}	443^{+36}_{-40}	572^{+50}_{-54}	1250^{+104}_{-114}
$\mathrm{gg} \to \mathrm{ZZ}$	$49.1^{+8.7}_{-8.8}$	$81.8^{+11.2}_{-10.7}$	$121.5^{+17.1}_{-16.3}$	$252.4^{+35.1}_{-33.5}$
Z + X	$17.1^{+6.4}_{-6.1}$	$35.4^{+12.7}_{-11.4}$	$47.8^{+16.4}_{-15.8}$	$100.3^{+21.3}_{-20.6}$
Sum of backgrounds	301^{+39}_{-43}	560^{+43}_{-47}	741^{+62}_{-65}	1602^{+126}_{-135}
Signal ($m_{\mathrm{H}} = 125$ GeV)	$13.9^{+1.9}_{-2.1}$	$28.9^{+2.5}_{-2.6}$	35.8 ± 3.3	$78.5^{+7.0}_{-7.1}$
Total expected	315^{+41}_{-45}	589^{+45}_{-49}	777^{+64}_{-67}	1681^{+131}_{-140}
Observed	307	602	797	1706

there is a very good agreement in all other categories with still no events observed in the categories targeting $\mathrm{t\bar{t}H}$ production, compatible with $0.24^{+0.03}_{-0.03}$ and $0.49^{+0.11}_{-0.05}$ expected events. Also, one can notice that the $\mathrm{b\bar{b}H}$ and tqH production modes are now considered in the analysis. The $\mathrm{b\bar{b}H}$ production mode has expected yield of similar size of the one from $\mathrm{t\bar{t}H}$ but it is distributed in ends up mostly in the gluon fusion dominated untagged category and is thus scaled with the gluon fusion when signal strength modifiers are measured. The single top tqH production has negligible yield for current statistics, but for consistency it is considered by being scaled together with the $\mathrm{t\bar{t}H}$ process.

The famous reconstructed four-lepton invariant mass distribution for the 2017 analysis is shown in Fig. 5.11 and compared with the SM expectations from signal and background. The observed distribution shows very good agreement with the SM expectation obtained from simulation, over the whole mass spectrum with a slight excess of events in the signal region that end up classified in the Untagged event category. This is best visible in Fig. 5.12, where the four-lepton invariant mass distribution is shown for each of the seven event categories used in the 2017 analysis. As explained in Section 4.3.3, event category definition is very similar to the one used for the 2016 analys, the main differences coming from the splitting of the $\mathrm{t\bar{t}H}$-tagged category to $\mathrm{t\bar{t}H}$-hadronic-tagged and $\mathrm{t\bar{t}H}$-lepton-tagged categories and dropping the VH-$\mathrm{E_T^{miss}}$-tagged category. Other than the slight excess in the untagged event category, a very good agreement between data and MC is observed with one $\mathrm{t\bar{t}H}$ candidate observed in the new $\mathrm{t\bar{t}H}$-leptonic-tagged category, but just outside the analysis signal mass window of $105 < m_{4\ell} < 140$ GeV.

The reconstructed invariant masses of lepton pairs selected as $\mathrm{Z_1}$ and $\mathrm{Z_2}$ are shown in Fig. 5.13 in the $118 < m_{4\ell} < 130$ GeV mass range, together with their correlation. The distribution of the discriminants used for event categorization and the corresponding working points are shown in Fig. 5.14, with the SM Higgs boson signal being separated into the production mode which is targeted by the specific discriminant and other production modes. Already from these distributions we can see promising candidates for the

Table 5.8 – The number of expected background and signal events and number of observed candidates after full analysis selection, for each event category, for the mass range $118 < m_{4\ell} < 130$ GeV and for an integrated luminosity of $41.5\ \mathrm{fb}^{-1}$. The yields are given for the different production modes. Signal and ZZ backgrounds are estimated from Monte Carlo simulation, Z + X is estimated from data. The uncertainties include both statistical and systematic sources.

	Event Category							
	Untagged	VBF-1j	VBF-2j	VH-lept.	VH-hadr.	$t\bar{t}H$-lept.	$t\bar{t}H$-hadr.	Inclusive
$q\bar{q} \to ZZ$	22.72	1.91	0.13	0.23	0.19	0.00	0.01	25.19
$gg \to ZZ$	1.93	0.30	0.03	0.04	0.02	0.00	0.00	2.32
Z + X	9.60	0.80	0.56	0.17	0.56	0.04	0.15	11.87
Sum of backgrounds	34.25	3.00	0.72	0.44	0.77	0.04	0.16	39.38
Uncertainties	$^{+2.79}_{-2.91}$	$^{+0.30}_{-0.29}$	$^{+0.14}_{-0.13}$	$^{+0.04}_{-0.05}$	$^{+0.12}_{-0.12}$	$^{+0.01}_{-0.01}$	$^{+0.10}_{-0.03}$	$^{+3.29}_{-3.39}$
ggH	46.94	9.90	1.74	0.06	1.29	< 0.01	0.04	59.96
VBF	1.68	1.57	1.89	0.01	0.08	< 0.01	0.01	5.24
WH-lep	0.18	0.02	0.01	0.28	0.01	0.01	< 0.01	0.50
WH-had	0.48	0.16	0.05	0.00	0.32	< 0.01	0.01	1.02
ZH-lep	0.29	0.02	0.01	0.07	0.03	< 0.01	< 0.01	0.43
ZH-had	0.32	0.10	0.03	0.00	0.23	< 0.01	0.01	0.69
$t\bar{t}H$	0.11	< 0.01	0.02	0.03	0.04	0.18	0.25	0.65
$b\bar{b}H$	0.48	0.10	0.02	0.01	0.02	< 0.01	< 0.01	0.63
tqH	0.03	< 0.01	0.02	0.01	0.01	0.01	0.01	0.09
Signal	50.51	11.87	3.79	0.47	2.03	0.20	0.33	69.21
Uncertainties	$^{+4.68}_{-4.74}$	$^{+1.41}_{-1.45}$	$^{+0.68}_{-0.59}$	$^{+0.04}_{-0.04}$	$^{+0.28}_{-0.25}$	$^{+0.03}_{-0.02}$	$^{+0.05}_{-0.04}$	$^{+6.13}_{-6.21}$
Total expected	84.76	14.87	4.51	0.91	2.80	0.24	0.49	108.58
Uncertainties	$^{+6.52}_{-6.71}$	$^{+1.59}_{-1.63}$	$^{+0.74}_{-0.64}$	$^{+0.07}_{-0.07}$	$^{+0.32}_{-0.29}$	$^{+0.03}_{-0.03}$	$^{+0.11}_{-0.05}$	$^{+8.21}_{-8.42}$
Observed	103	14	5	2	2	0	0	126

VBF and VH production modes with very high values of their respective categorization discriminants.

Again, the correlation between the two variables used in the likelihood fit to extract the signal strength modifiers is the plot that is closest possible visual representation of the actual measurement performed. One of the biggest improvements introduced in the 2017 analysis was the development of two new kinematic discriminants $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VBF+dec}}$ and $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VH+dec}}$ that combine production with decay information to be able to discriminate between VBF or VH production and the dominant ggH production. The new discriminants are introduced and explained in detail in Section 4.4.2. In Fig. 5.15 one can see the correlation between all three kinematic discriminants $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{kin}}$, $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VBF+dec}}$, and $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VH+dec}}$, and the four-lepton invariant mass $m_{4\ell}$. The distributions are shown for the categories where the discriminants are used in the fit, and the gray scale represents the expected contributions of the reducible and irreducible backgrounds and the Higgs boson signal. One can see for example that from seven observed VBF-2jet-tagged events in the $105 < m_{4\ell} < 140$ GeV window, three have very high values of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VBF+dec}}$ making them more compatible with the signal hypothesis than the background hypothesis. In

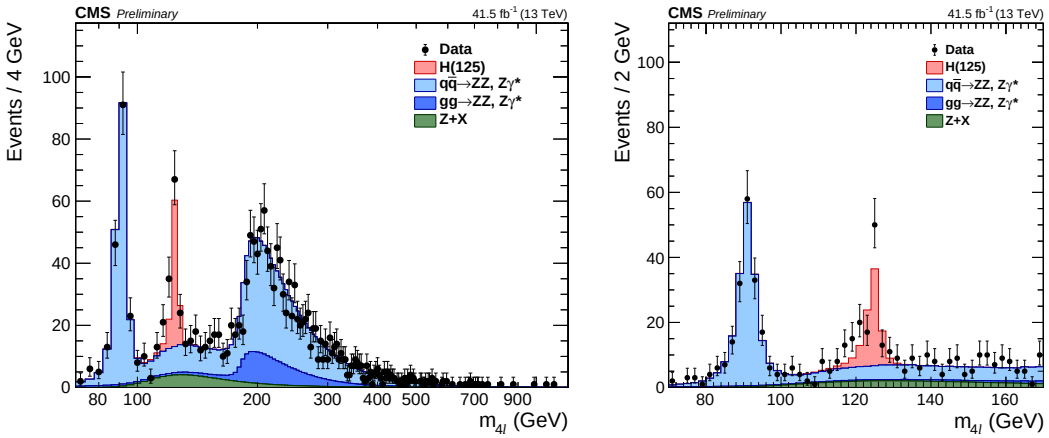

Figure 5.11 – Distribution of the four-lepton reconstructed invariant mass $m_{4\ell}$ in the full mass range (left) and the low-mass range (right). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data. No events are observed with $m_{4\ell} > 1.1$ TeV.

this case, the signal hypothesis means VBF production, and the background hypothesis means ggH or SM background production. We can also see that only one VH-hadronic-tagged event near 125 GeV has a high value of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VH+dec}}$. The 1D distributions of the kinematic discriminants in the $118 < m_{4\ell} < 130$ GeV mass range are shown in Fig. 5.16, where again the SM Higgs boson signal being separated into the production mode which is targeted by the specific discriminant and other production modes dominated by gluon fusion.

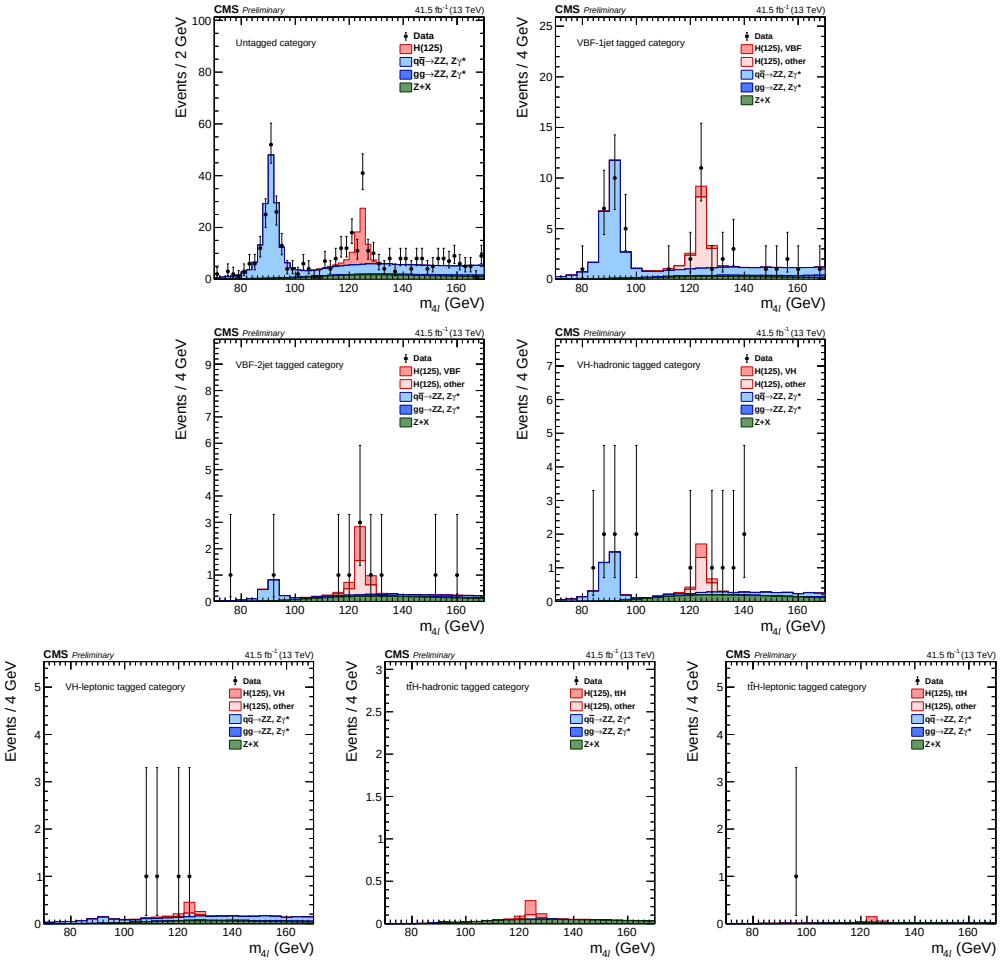

Figure 5.12 – Distribution of the four-lepton reconstructed mass in the seven event categories for the low-mass range in Untagged category (top left,) VBF-1jet-tagged category (top right), VBF-2jet-tagged category (center left), VH-hadronic-tagged category (center right), VH-leptonic-tagged category (bottom left), $\mathrm{t\bar{t}H}$-hadronic-tagged category (bottom middle), and $\mathrm{t\bar{t}H}$-leptonic-tagged category (bottom right). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data. For the categories other than the untagged category, the SM Higgs boson signal is separated into two components: the production mode that is targeted by the specific category, and other production modes, where the gluon fusion process dominates.

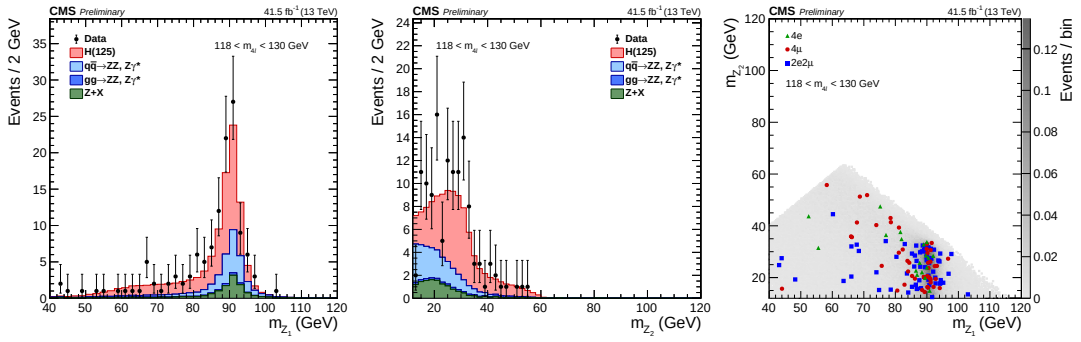

Figure 5.13 – Distribution of the Z_1 (left) and Z_2 (center) reconstructed invariant masses and correlation between the two (right) in the mass region $118 < m_{4\ell} < 130$ GeV. The stacked histograms and the gray scale represent expected distributions of the signal and background processes, and points represent the data. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data.

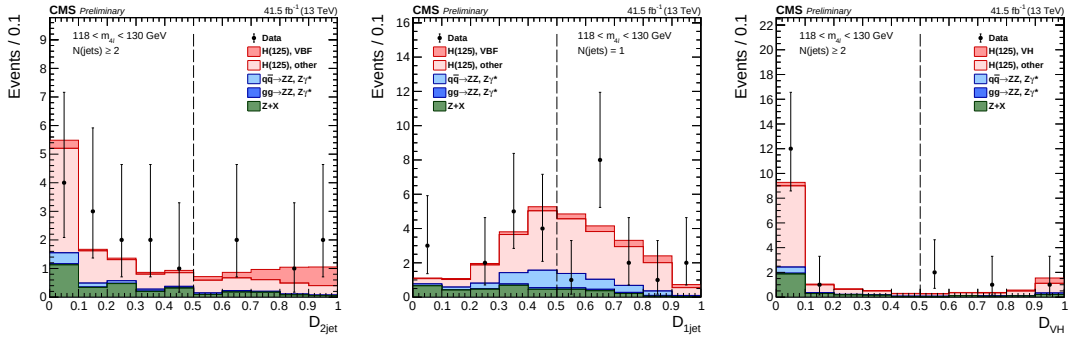

Figure 5.14 – Distribution of categorization discriminants in the mass region $118 < m_{4\ell} < 130$ GeV: $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-2j}}$ (left), $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VBF-1j}}$ (middle) , and $\mathcal{D}^{\mathrm{ME}}_{\mathrm{VH-hadr.}} = \max(\mathcal{D}^{\mathrm{ME}}_{\mathrm{WH-hadr.}}, \mathcal{D}^{\mathrm{ME}}_{\mathrm{ZH-hadr.}})$ (right). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data. The vertical gray dashed lines denote the working points used in the event categorization. The SM Higgs boson signal is separated into two components: the production mode which is targeted by the specific discriminant, and other production modes, where the gluon fusion process dominates.

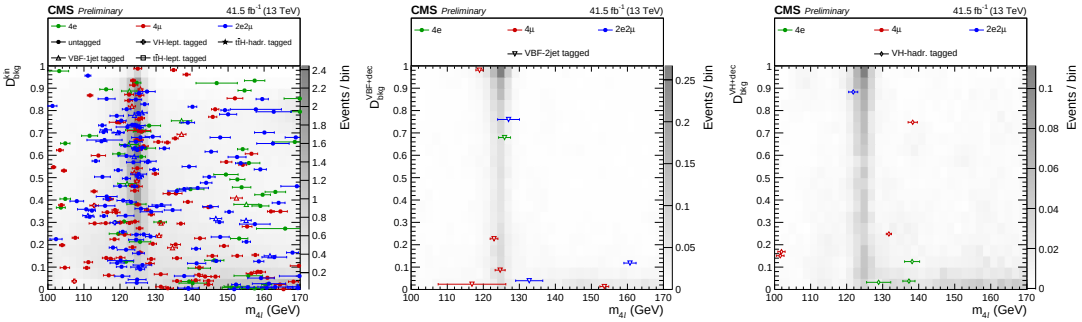

Figure 5.15 – Distribution of three different kinematic discriminants versus $m_{4\ell}$: $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ (left), $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ (middle) and $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ (right) shown in the mass region $100 < m_{4\ell} < 170$ GeV. The gray scale represents the expected total number of ZZ and $\text{Z} + \text{X}$ background and SM Higgs boson signal events for $m_{\text{H}} = 125$ GeV. The points show the data and the horizontal bars represent the measured event-by-event mass uncertainties. Different marker styles are used to denote the categorization of the events.

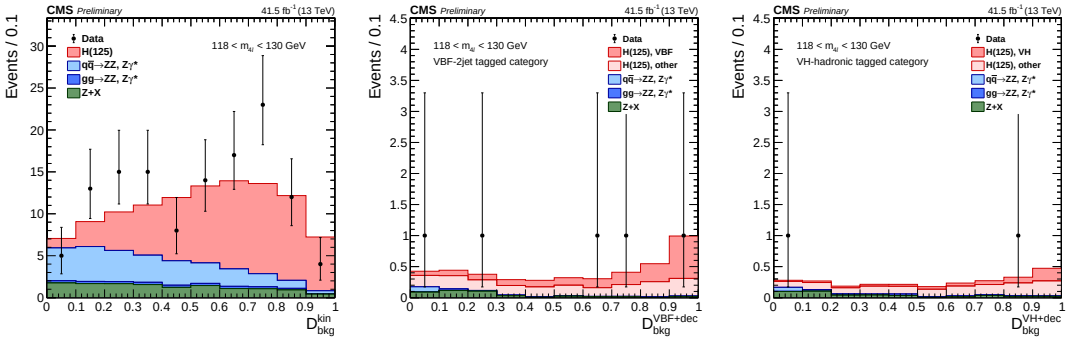

Figure 5.16 – Distribution of kinematic discriminants in the mass region $118 < m_{4\ell} < 130$ GeV: $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ (left), $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ (middle) and $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ (right). Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\text{H}} = 125$ GeV, denoted as $\text{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\text{Z} + \text{X}$ background to the estimation from data. The SM Higgs boson signal is separated into two components: the production mode which is targeted by the specific discriminant, and other production modes, where the gluon fusion process dominates.

5.2.2 Measurements

For the 2017 data signal strength modifiers and stage-0 simplified template cross section are measured. Before presenting the results systematic uncertainties affecting the measurements are discussed.

Systematic uncertainties

The values of the systematic uncertainties used in the measurements of the 2017 analysis are summarized in Table 5.3.

Table 5.9 – Summary of the systematic uncertainties affecting measurements of the 2017 $H \to ZZ^{\star} \to 4\ell$ analysis. The uncertainties either affect the shape (S) or only the normalization (N) of predicted distributions, (p) denotes the uncertainties that are process-dependent, and (c) that the uncertainties are category-dependent..

Systematic	$4e$	4μ	$2e2\mu$
Luminosity (N)		2.3%	
Lepton Efficiency (N)	12.5%	5.6%	6.7%
QCD scale (N) (p) (c)		2 - 40%	
PDF set (N) (p) (c)		0 - 4%	
Branching ratio (N)		2%	
k-factor for $q\bar{q} \to ZZ$ (N) (c)		1 - 15%	
k-factor for $gg \to ZZ$ (N)		10%	
$Z + X$ estimation (N) (c)	42%	44%	33%
Lepton energy scale (S)	0.3%	0.04%	0.1%
Lepton energy resolution (S)	20%	10%	15%
Jet energy scale (N) (p) (c)		2 - 22%	
b-tagging (N) (p) (c)		0 - 11%	

All of the values have been performed again using new data and simulation, while the procedure to derive them was the same as for the 2016 analysis and explained in Section 4.7. The biggest difference was the study of the possible systematical sources affecting the shape of the new kinematic discriminants. All possible sources for this were studied, and 2D templates used to build likelihood were derived and compared to the nominal ones. No significant differences were observed in the shape and it was concluded that it is enough to account for those effects in the form of categorization migration as done before. From the experimental side, dominant sources of uncertainty again come from luminosity and lepton efficiency, being 2.3% and 5.6 to 12.5% depending on the final state, respectively. The increase in the lepton efficiency uncertainty comes mainly from the problems to measure the efficiency of reconstruction for the very low-p_{T} leptons because of harsher pileup conditions. This increased uncertainty

is then propagated to the four-lepton final state resulting roughly in a 40% increase of the uncertainty for the $4e$ final state. Even though mass measurement is not performed, and lepton energy scale and resolution are subdominant sources for the measurements performed, nevertheless they were again measured and showed to be same like for the 2016 analysis. Dominant sources from the theoretical side are again found to be from the QCD scale and PDF set uncertainties for dominant gluon fusion production mode.

In the case of the measurements which use event categorization, the main sources of event migrations remain the same, amounting between 2 to 40% for main production modes, depending on the category. Additional experimental sources like the imprecise knowledge of the jet energy scale amounts for 2 to 22%, and uncertainty on b tagging efficiency amounts up to 11% in the $\mathrm{t\bar{t}H}$-tagged category.

To summarize, almost all uncertainties have increased slightly with respect to the 2016 analysis, mainly due to higher average number of pileup interactions in events.

Signal strength modifiers

A simultaneous fit to all channels, at the ATLAS and CMS Run I combined mass value of $m_{\mathrm{H}} = 125.09$ GeV, resulted in a measured global signal strength modifier of $\mu = \sigma/\sigma_{\mathrm{SM}} = 1.10^{+0.14}_{-0.13}(\mathrm{stat})^{+0.11}_{-0.09}(\mathrm{syst}) = 1.10^{+0.19}_{-0.17}$ for the 2017 data. Again, all signal strength parameters are constrained to be non-negative. This result is compared to the measurement for each of the seven event categories in Fig. 5.17 (top left). All observed values are consistent with the SM prediction of $\mu = 1$, within the uncertainties.

The improvement resulting from the new discriminants is best seen when extracting signal strength modifiers for different production modes. A fit for five signal strength modifiers $\mu_{\mathrm{ggH,b\bar{b}H}}$, μ_{VBF}, μ_{VHhad}, μ_{VHlep}, and $\mu_{\mathrm{t\bar{t}H,tqH}}$, that control the contributions of the main SM Higgs boson production modes, is performed with all signal strength modifiers constrained to positive values. The results are reported in Fig. 5.17 (top right) and compared to the expected signal strength modifiers in Table 5.10. Even though there is a bigger over fluctuation of gluon fusion dominated untagged events than in the 2016 analysis, owing to the introduction of the new discriminant this has not affected the measurements of signal strength for the VBF production and $\mu_{\mathrm{VBF}} = 1.12^{+1.19}_{-0.83}$ was measured. With only a small number of events observed in the VH-hadronic-tagged category and only one of them having a high value of $\mathcal{D}_{\mathrm{bkg}}^{\mathrm{VH+dec}}$, signal strength modifier for hadronic VH production was measured to be zero again, but well compatible with the SM expectation within uncertainties.

Table 5.10 – Expected and observed signal-strength modifiers with 2017 data.

	Inclusive	$\mu_{\mathrm{ggH,b\bar{b}H}}$	μ_{VBF}	μ_{VHhad}	μ_{VHlep}	$\mu_{\mathrm{t\bar{t}H,tqH}}$
Expected	$1.00^{+0.14}_{-0.13}(\mathrm{stat})^{+0.11}_{-0.09}(\mathrm{syst})$	$1.00^{+0.22}_{-0.20}$	$1.00^{+1.19}_{-0.79}$	$1.00^{+3.24}_{-1.00}$	$1.00^{+3.36}_{-1.00}$	$1.00^{+2.47}_{-1.00}$
Observed	$1.10^{+0.14}_{-0.13}(\mathrm{stat})^{+0.13}_{-0.11}(\mathrm{syst})$	$1.14^{+0.23}_{-0.20}$	$1.12^{+1.19}_{-0.83}$	$0.00^{+1.54}_{-0.00}$	$2.23^{+3.95}_{-2.12}$	$0.00^{+0.93}_{-0.00}$

Two signal strength modifiers $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}}$ and $\mu_{\mathrm{VBF,\,VH}}$, representing scale factors to the fermion and vector boson induced contribution to the expected SM cross section are again measured, profiting from the improvements lead by two new discriminants. A two-parameter fit is performed simultaneously to all categories, again assuming a mass $m_{\mathrm{H}} = 125$ GeV, leading to measurements of $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}} = 1.11^{+0.23}_{-0.21}$ and $\mu_{\mathrm{VBF,\,VH}} = 1.00^{+0.96}_{-0.71}$. The SM expectation, the measured values, and the 68% and 95% CL contours in the ($\mu_{\mathrm{ggH,\,t\bar{t}H}}$,$\mu_{\mathrm{VBF,\,VH}}$) plane are shown in Fig. 5.17 (bottom). The SM prediction is found to lie well within the 68% CL regions of this measurement, very close to the observed value.

Cross section measurements

In the 2017 analysis, only the stage-0 simplified template cross section measurement is performed, in a simplified fiducial volume defined as $|y_{\mathrm{H}}| < 2.5$, for the main Higgs boson production mechanisms. The theoretical uncertainties affecting the signal normalization are removed, while those affecting category migrations and background normalizations are kept. The measured cross sections normalized to the SM predictions denoted as σ_{theo}, are shown in Fig. 5.18. Measurement is very similar to the one of the signal strength modifiers for different production modes, with a visible decrease only in the uncertainty on the gluon fusion measurement, being the only one that is not completely statistically dominated.

5.2.3 Summary

Results based on $41.5\ \mathrm{fb}^{-1}$ collected in 2017 with the CMS detector show that the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis can cope with the new harsh collision conditions that lead to 33 pileup interactions in average. To cope with this several improvements were introduced. From the detector side, a new inner layer of the pixel detector helps disregard events coming from pileup interactions, and from the analysis side new developments in the electron identification lead to event better fake rejection than for the 2016 analysis. The global signal strength modifier, measured at $m_{\mathrm{H}} = 125.09$, is found to be $\mu = 1.10^{+0.19}_{-0.17}$, compatible both with the SM prediction and with the 2016 analysis within the uncertainties. The signal strength modifiers for main production mechanisms are also measured, with the improvements coming from new production and decay matrix element discriminants. The measured signal strength modifiers associated with fermions and vector bosons are $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}} = 1.11^{+0.23}_{-0.21}$ and $\mu_{\mathrm{VBF,\,VH}} = 1.00^{+0.96}_{-0.71}$, respectively. The golden channel proved to be leading the way one more time, since this is the first set of measurements of Higgs boson properties based on 2017 data that was made public by the CMS Collaboration.

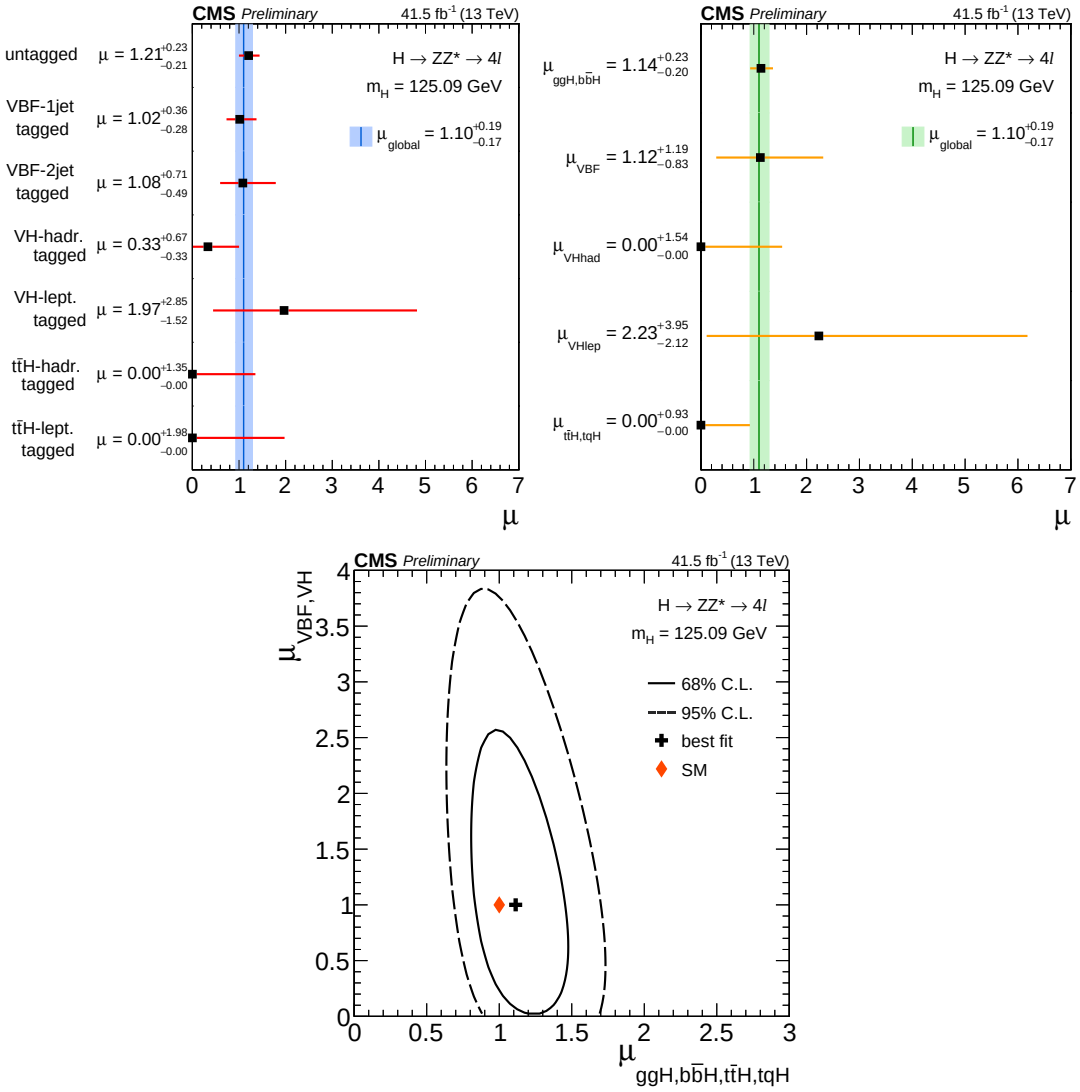

Figure 5.17 – (Top left) Observed values of the signal strength $\mu = \sigma/\sigma_{SM}$ for the seven event categories, compared to the combined μ shown as a vertical line. The horizontal bars and the filled band indicate the $\pm\,1\sigma$ uncertainties. (Top right) Results of likelihood scans for the signal-strength modifiers corresponding to the main SM Higgs boson production modes, compared to the combined μ shown as a vertical line. The horizontal bars and the filled band indicate the $\pm\,1\sigma$ uncertainties. The uncertainties include both statistical and systematic sources. (Bottom) Result of the 2D likelihood scan for the $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}}$ and $\mu_{\mathrm{VBF,\,VH}}$ signal-strength modifiers. The solid and dashed contours show the 68% and 95% CL regions, respectively. The cross indicates the best-fit values, and the diamond represents the expected values for the SM Higgs boson.

5.3 Combined results with data recorded in 2016 and 2017

In addition to measurements performed on $41.5\ \mathrm{fb}^{-1}$ of data collected in 2017, a combination with $35.9\ \mathrm{fb}^{-1}$ data recorded in 2016 by the CMS detector was also performed. I was leading the combination measurement, providing the inputs and building the com-

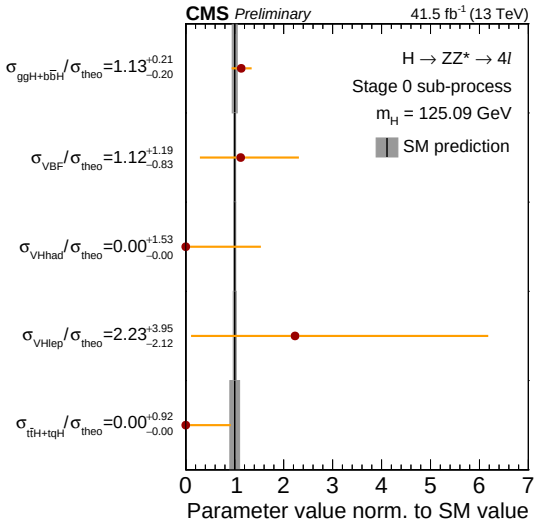

Figure 5.18 – Results of the fit for simplified template cross sections for the 'stage-0 subprocesses', normalized to the SM predictions. The grey bands indicate the theoretical uncertainties in the SM predictions. The orange error bars show the full uncertainty, including experimental uncertainties and theoretical uncertainties causing migration of events between the various categories.

bined likelihood model, that is part of the results in the recently released CMS PAS [123]. First, combination strategy will be discussed, describing how the two datasets are combined, why are systematic uncertainties chosen to be fully correlated and how is the measurement of signal strength modifiers performed. Afterwards, combined results extracted from a total of $77.4\ \mathrm{fb}^{-1}$ of Run II data collected with the CMS detector ar a center-of-mass energy of 13 TeV will be presented.

5.3.1 Combination strategy

There are several ways to extract results from multiple data sets. One can reanalyze the full data set using the latest state of the analysis. In our example, that would mean reanalyzing 2016 data using all of the improvements from 2017 analysis. On the other hand, as it is a usual practice in the CMS Collaboration when combining different channels, one can combine the two analysis at the likelihood level. This section describes how the combined likelihood is built and how the nuisance parameters are treated.

Building the combined likelihood

The likelihood for the combined measurement of the signal strength modifiers has a form of a general likelihood function given in Eq. 4.15. The first product runs over all channels c, and in the case of combination this implies a total of $(3\ \text{final states} \times 7\ \text{event categories})_{2016} + (3\ \text{final states} \times 7\ \text{event categories})_{2017} = 42$ channels. All chan-

nels use the general form of the 2D likelihood for the signal strength measurement given with Eq. 4.18, but with some differences. For the 21 channels of 2016 analysis, all signal and background shapes together with conditional 2D templates are derived from the 2016 MC simulation samples, reflecting detector conditions at that time. Event categories are defined as in Section 4.3.2 and only one discriminant $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ is used in all categories. For the 2017 analysis, signal and background shapes are derived from 2017 MC simulation samples with updated electron selection and detector conditions. The 2D templates are derived using three different kinematic observables $\mathcal{D}_{\text{bkg}}^{\text{kin}}$, $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$, and $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ in their dedicated categories. With all of the ingredients in place one can define three different signal strength modifier measurements:

- A global signal strength modifier can be extracted with a simple $\mu \cdot (s^{2016} + s^{2017}) + b$ model, scaling all signal contributions with a global signal strength μ
- One can define two signal strength modifiers $\mu_{\text{ggH, b}\bar{\text{b}}\text{H, t}\bar{\text{t}}\text{H, tqH}}$ and $\mu_{\text{VBF, VH}}$ to scale the Higgs boson production cross sections in modes related to couplings to fermions and electroweak gauge bosons, respectively. This is achieved by replacing the $\mu \cdot (s^{2016} + s^{2017}) + b$ model with $\mu_{\text{ggH, b}\bar{\text{b}}\text{H, t}\bar{\text{t}}\text{H, tqH}} \cdot (s_{\text{ggH}}^{2016} + s_{\text{t}\bar{\text{t}}\text{H}}^{2016} + s_{\text{ggH}}^{2017} + s_{\text{b}\bar{\text{b}}\text{H}}^{2017} + s_{\text{t}\bar{\text{t}}\text{H}}^{2017} + s_{\text{tqH}}^{2017}) + \mu_{\text{VBF, VH}} \cdot (s_{\text{VBF}}^{2016} + s_{\text{VHhad}}^{2016} + s_{\text{VHlep}}^{2016} + s_{\text{VBF}}^{2017} + s_{\text{VHhad}}^{2017} + s_{\text{VHlep}}^{2017}) + b$.
- To measure five signal strength modifiers of main Higgs production modes $\mu_{\text{ggH, b}\bar{\text{b}}\text{H}}$, μ_{VBF}, μ_{VHhad}, μ_{VHlep}, and $\mu_{\text{t}\bar{\text{t}}\text{H, tqH}}$, one has to replace the global model with $\mu_{\text{ggH, b}\bar{\text{b}}\text{H}} \cdot (s_{\text{ggH}}^{2016} + s_{\text{ggH}}^{2017} + s_{\text{b}\bar{\text{b}}\text{H}}^{2017}) + \mu_{\text{VBF}} \cdot (s_{\text{VBF}}^{2016} + s_{\text{VBF}}^{2017}) + \mu_{\text{VHhad}} \cdot (s_{\text{VHhad}}^{2016} + s_{\text{VHhad}}^{2017}) + \mu_{\text{VHlep}} \cdot (s_{\text{VHlep}}^{2016} + s_{\text{VHlep}}^{2017}) + \mu_{\text{t}\bar{\text{t}}\text{H, tqH}} \cdot (s_{\text{t}\bar{\text{t}}\text{H}}^{2016} + s_{\text{t}\bar{\text{t}}\text{H}}^{2017} + s_{\text{tqH}}^{2017}) + b$.

A small inconsistency is present due to $\text{b}\bar{\text{b}}\text{H}$ and tqH production mechanisms not being considered in the 2016 analysis. However, thanks to the similarities of $\text{b}\bar{\text{b}}\text{H}$ with ggH and tqH with $\text{t}\bar{\text{t}}\text{H}$ in terms of distribution among categories the magnitude of this inconsistency was studied and was found to be negligible.

Correlating systematic uncertainties

Another important task in building the combined likelihood is the second product from Eq. 4.15. As discussed in Section 4.8.3, one has to understand possible correlations between the same nuisance parameters used in the 2016 and 2017 analysis. While for the theoretical sources of uncertainties corresponding to the QCD scale, PDF set, branching ratio, and k-factor uncertainties it is clear that they are fully correlated between the two data taking periods it is not obvious the same is the case for the experimental sources. A study of the possible treatment of experimental uncertainties was performed. Two extremes were considered, measuring global signal strength modifier on Asimov data set with the assumption of fully correlated or fully uncorrelated experimental uncertainties between the two data taking periods. Results of the study are reported in Table 5.11.

As expected, assuming full correlation between the uncertainties is proved to be a more conservative approach. While it is expected for some experimental sources like luminosity or jet energy scale uncertainties to be full correlated, some others like electron

Table 5.11 – Expected signal-strength modifiers for combined 2016 and 2017 data based on two assumptions of fully correlated or uncorrelated experimental uncertainties.

	Fully correlated	Fully uncorrelated
Expected	$1.00 \pm 0.10(\text{stat})^{+0.10}_{-0.08}(\text{syst})$	$1.00 \pm 0.10(\text{stat})^{+0.10}_{-0.06}(\text{syst})$

efficiency and b-tagging are expected to be closer to fully uncorrelated due to significant changes between the analysis. However, a set of dedicated thorough studies would be needed to justify this choice and given the small magnitude of the overall effect and that these are preliminary results based on prompt reconstructed data, in the end it was decided to adopt the conservative approach of considering all experimental sources as correlated between the 2016 and 2017 analysis.

5.3.2 Results

The four-lepton invariant mass distribution

The reconstructed four-lepton invariant mass distribution combining 2016 and 2017 datasets is shown in Fig. 5.19 for the sum of the $4e$, 4μ, and $2e2\mu$ final states, and compared with the expectations from signal and background processes. This distribution is obtained as a sum of the invariant mass distributions for the 2016 analysis shown in Fig. 5.1 and for the 2017 analysis shown in Fig. 5.11. It shows an impressive agreement between data and simulation over the full mass range for $77.4\ \text{fb}^{-1}$ of Run II data collected by the CMS experiment.

Signal strength modifiers

A simultaneous fit to all channels, combining 2016 and 2017 data, resulted in a measured global signal strength modifier of

$$\mu = \sigma/\sigma_{\text{SM}} = 1.06 \pm 0.10(\text{stat})^{+0.08}_{-0.06}(\text{exp.syst})^{+0.07}_{-0.05}(\text{th.syst}) = 1.06^{+0.15}_{-0.13} \tag{5.2}$$

at $m_{\text{H}} = 125.09$ GeV. To obtain different contributions of systematics, a likelihood scan is performed removing the theoretical sources of systematical uncertainties to determine the contribution of the experimental sources of uncertainties. The contribution is calculated as the difference in quadrature between the total uncertainty and the uncertainty obtained from this fit. This result is consistent with the Run I CMS measurement of $\mu = 0.93^{+0.26}_{-0.23}(\text{stat})^{+0.13}_{-0.09}(\text{syst}) = 0.93^{+0.29}_{-0.25}$ [26] obtained for $m_{\text{H}} = 125.6$ GeV, surpassing its sensitivity thanks to increase in statistics while managing to keep the systematics at the same level. Also, already on two thirds of expected Run II data, the measurement of the global signal strength modifier has reached a limit of not being statistically dominated anymore, providing interesting prospects for the upcoming analysis of full Run II statistics.

A fit for five signal strength modifiers $\mu_{\text{ggH,b}\bar{\text{b}}\text{H}}$, μ_{VBF}, μ_{VHhad}, μ_{VHlep}, and $\mu_{\text{t}\bar{\text{t}}\text{H,tqH}}$,

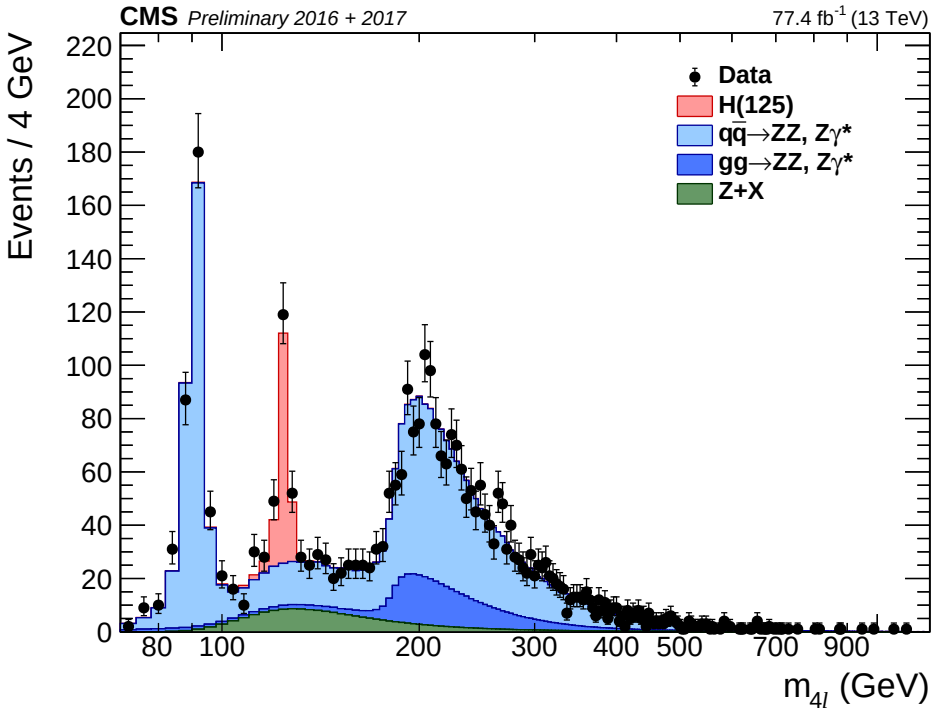

Figure 5.19 – Distribution of the four-lepton reconstructed invariant mass $m_{4\ell}$ in the full mass range combining 2016 and 2017. Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The SM Higgs boson signal with $m_{\mathrm{H}} = 125$ GeV, denoted as $\mathrm{H}(125)$, and the ZZ backgrounds are normalized to the SM expectation, the $\mathrm{Z} + \mathrm{X}$ background to the estimation from data. No events are observed with $m_{4\ell} > 1.1$ TeV.

that control the contributions of the main SM Higgs boson production modes, is performed with all signal strength modifiers constrained to positive values. The results are reported in Fig. 5.20 (left) and compared to the expected signal strength modifiers in Table 5.12. Unfortunately, no signal $\mathrm{t\bar{t}H}$ candidates are observed in the 2016 or 2017 data for roughly 1.5 expected events when combining three categories targeting $\mathrm{t\bar{t}H}$ production. This is still in agreement with the SM expectations well within the 2σ uncertainty. Because of the very small expected number of events and a forced positive constrain on the signal strength, a likelihood function is very steep and 2σ uncertainty is much larger than for cases where significant number of events is observed. Same conclusion is also valid for the measurement of μ_{VHhad}.

Table 5.12 – Expected and observed signal-strength modifiers for combined 2016 and 2017 data.

	Inclusive	$\mu_{\mathrm{ggH,b\bar{b}H}}$	μ_{VBF}	μ_{VHhad}	μ_{VHlep}	$\mu_{\mathrm{t\bar{t}H,tqH}}$
Expected	$1.00 \pm 0.10(\mathrm{stat})^{+0.08}_{-0.06}(\mathrm{exp.\ syst})^{+0.07}_{-0.05}(\mathrm{th.\ syst})$	$1.00^{+0.17}_{-0.16}$	$1.00^{+0.86}_{-0.67}$	$1.00^{+2.39}_{-1.00}$	$1.00^{+2.30}_{-1.00}$	$1.00^{+1.80}_{-1.00}$
Observed	$1.06 \pm 0.10(\mathrm{stat})^{+0.08}_{-0.06}(\mathrm{exp.\ syst})^{+0.07}_{-0.05}(\mathrm{th.\ syst})$	$1.15^{+0.18}_{-0.16}$	$0.69^{+0.75}_{-0.57}$	$0.00^{+1.16}_{-0.00}$	$1.25^{+2.46}_{-1.25}$	$0.00^{+0.53}_{-0.00}$

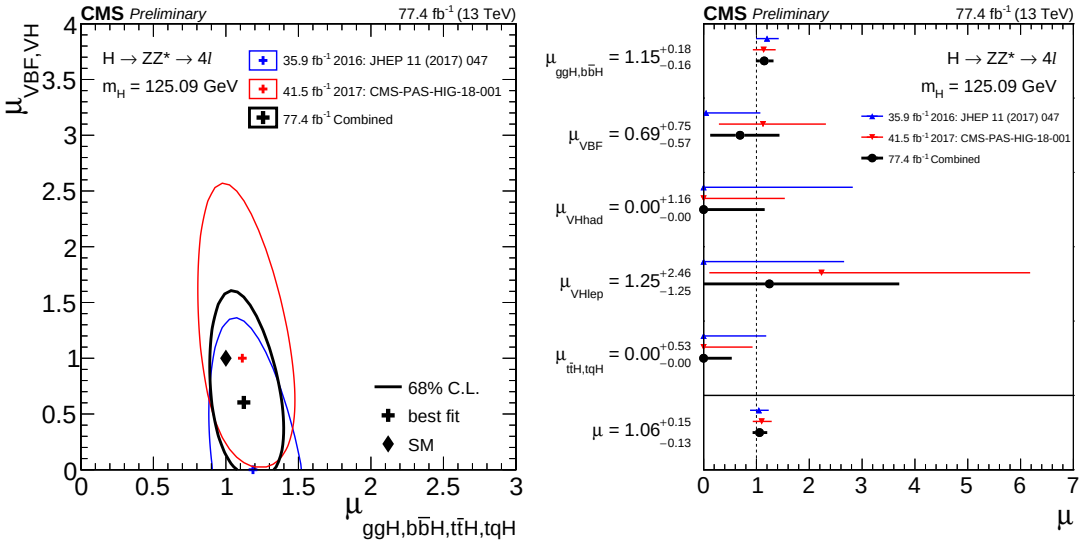

Figure 5.20 – Result of the 2D likelihood scan for the $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}}$ and $\mu_{\mathrm{VBF,\,VH}}$ signal-strength modifiers (left). The solid contours show the 68% CL regions. The cross indicates the best-fit value, and the diamond represents the expected value for the SM Higgs boson. Results of likelihood scans for the signal-strength modifiers corresponding to the main SM Higgs boson production modes, compared to the SM expectation shown as a vertical dashed line (right). The horizontal bars indicate the $\pm$ 1σ uncertainties. The uncertainties include both statistical and systematic sources. The measurements of the global signal strength μ are also shown.

Two signal strength modifiers $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}}$ and $\mu_{\mathrm{VBF,\,VH}}$, representing scale factors to the fermion and vector boson induced contribution to the expected SM cross section are measured to be $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}} = 1.12^{+0.18}_{-0.16}$ and $\mu_{\mathrm{VBF,\,VH}} = 0.60^{+0.62}_{-0.49}$, at $m_{\mathrm{H}} = 125$ GeV. The result is in agreement with Run I CMS measurements of $\mu_{\mathrm{ggH,\,t\bar{t}H}} = 0.80^{+0.46}_{-0.36}$ and $\mu_{\mathrm{VBF,\,VH}} = 1.7^{+2.2}_{-2.1}$, obtained for $m_{\mathrm{H}} = 125.6$ GeV. Significant decrease of uncertainties on this measurement comes from the increase in statistics, introduction of complex event categorization, introduction of new production discriminants used as observables in the fit, and many other improvements like for example electron identification. The SM expectation, the measured values, and the 68% and 95% CL contours in the ($\mu_{\mathrm{ggH,\,t\bar{t}H}}$,$\mu_{\mathrm{VBF,\,VH}}$) plane are shown in Fig. 5.17 (bottom). The SM prediction is found to lie well within the 68% CL regions of this measurement, very close to the observed value.

5.3.3 Summary

The 2016 and 2017 analysis are combined at the likelihood level and signal strength modifiers are measured. The most conservative approach of taking all systematic uncertainties between the two analysis as fully correlated is adopted. The global signal strength is measured to be $\mu = 1.06^{+0.15}_{-0.13}$ at $m_{\mathrm{H}} = 125.09$ GeV. This value is compatible with the SM prediction and with Run I result from the CMS Collaboration. Signal strength modifiers for main Higgs production mechanisms are also extracted and are all

in agreement with the SM prediction. Finally, two signal strength modifiers representing scale factors to the fermion and vector boson induced contributions to the expected SM cross sections are measured to be $\mu_{\mathrm{ggH,\,b\bar{b}H,\,t\bar{t}H,\,tqH}} = 1.12^{+0.18}_{-0.16}$ and $\mu_{\mathrm{VBF,\,VH}} = 0.60^{+0.62}_{-0.49}$, at $m_{\mathrm{H}} = 125$ GeV. The measurement of the global signal strength in the $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ channel has reached a stage when statistical and systematical uncertainties are of the same size. With further improvements in the analysis and with the reduction of experimental systematic uncertainties, full Run II data sample brings exciting possibilities for precision measurements of Higgs boson properties in the golden channel.

5.4 Other measurements

For the completeness of this thesis all other measurements of Higgs boson properties will be mentioned in this Section. These results are based on the same or slightly modified building blocks of the analysis described in Chapter 4.

5.4.1 Anomalous Higgs boson couplings from on-shell production

A detailed description of the analysis techniques is given in [125]. The analysis approach follows previous Run I publication [49] from the CMS Collaboration, expanded in two important ways. Information from the kinematic correlations of quark jets from VBF and VH production is used together with $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ decay information for the first time, in a very similar manner as described in Section 4.4.2. Moreover, data sets corresponding to integrated luminosities of 2.7 and $35.9\ \mathrm{fb}^{-1}$ collected at a center-of-mass energy of 13 TeV in Run II of the LHC during 2015 and 2016, respectively, are combined with the Run I data.

The Higgs boson is assumed to couplee to two gauge bosons VV, and three general tensor structures that are allowed by Lorentz symmetry are tested. It is assumed that all lepton and quark couplings to vector bosons follow the SM predictions. Relaxing this requirement would be equivalent to allowing the contact terms to vary with flavor, which would result in too many unconstrained parameters to be tested with the present amount of data. Anomalous interactions of a spin-zero H boson with two spin-one gauge bosons VV, such as ZZ, Zγ, $\gamma\gamma$, WW, and gg, are parameterized with a scattering amplitude that includes three tensor structures with expansion of coefficients up to (q^2/Λ^2):

$$A(\mathrm{HVV}) \sim \left[a_1^{\mathrm{VV}} + \frac{\kappa_1^{\mathrm{VV}} q_1^2 + \kappa_2^{\mathrm{VV}} q_2^2}{(\Lambda_1^{\mathrm{VV}})^2}\right] m_{\mathrm{V1}}^2 \epsilon_{\mathrm{V1}}^* \epsilon_{\mathrm{V2}}^* + a_2^{\mathrm{VV}} f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + a_3^{\mathrm{VV}} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu}, \tag{5.3}$$

where q_i, $\epsilon_{\mathrm{V}i}$, and $m_{\mathrm{V}i}$ are the four-momentum, polarization vector, and pole mass of a gauge boson, Λ is an energy scale of new physics, $f^{(i)\mu\nu} = \epsilon_{\mathrm{V}i}^{\mu} q_i^{\nu} - \epsilon_{\mathrm{V}i}^{\nu} q_i^{\mu}$, $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} f^{(i),\rho\sigma}$, and a_i^{VV} and $\kappa_i^{\mathrm{VV}}/(\Lambda_1^{\mathrm{VV}})^2$ are parameters to be determined from data.

In Eq.5.3, the only leading tree-level contributions are $a_1^{\mathrm{ZZ}} \neq 0$ and $a_1^{\mathrm{WW}} \neq 0$. The rest of the couplings are considered anomalous contributions. Tiny anomalous terms arise in the SM due to loop effects but are not yet accessible experimentally, and new BSM contributions could make them larger.

It is convenient to measure the effective cross-section ratios (f_{ai}) rather than the anomalous couplings themselves (a_i). First of all, most systematic uncertainties cancel in the ratio. Moreover, the effective fractions are conveniently bounded by 0 and 1 and do not depend on the normalization convention in the definition of the couplings. The effective fractional cross sections f_{ai} and phases ϕ_{ai} are defined as follows:

$$f_{ai} = |a_i|^2 \sigma_i / \sum |a_j|^2 \sigma_j, \text{and } \phi_{ai} = \arg(a_i/a_1). \tag{5.4}$$

The f_{ai} values have a simple interpretation as the fractional size of the BSM contribution

for the $H \to ZZ^\star \to 4\ell$ decay.

The analysis uses same building blocks as described in Chapter 4 with several slight modifications. Only VBF-2jet-tagged, VH-hadronic-tagged, and untagged categories are used to classify events. In addition, new matrix-element based discriminants are used as observables in fits to extract results. These observables are specially constructed to provide separation between SM signal and BSM anomalous coupling contributions, taking into account possible interference effects. An unbinned extended maximum likelihood fit to the events split into the categories according to the lepton flavor and production topology is performed in the on-shell signal region defined as $105 - 140$ GeV. Results are reported in Table 5.13 and illustrated in Fig. 5.21. All results are found to be compatible with the SM predictions, improving the constrains on the anomalous HVV couplings by an order of magnitude with respect to Run I results.

Table 5.13 – Summary of allowed 68% CL (central values with uncertainties) and 95% CL (in square brackets) intervals on anomalous coupling parameters obtained from the combined Run I and Run II data analysis.

Parameter	Observed	Expected
$f_{a3}\cos(\phi_{a3})$	$0.00^{+0.26}_{-0.09}$ $[-0.38, 0.46]$	$0.000^{+0.010}_{-0.010}$ $[-0.25, 0.25]$
$f_{a2}\cos(\phi_{a2})$	$0.01^{+0.12}_{-0.02}$ $[-0.04, 0.43]$	$0.000^{+0.009}_{-0.008}$ $[-0.06, 0.19]$
$f_{\Lambda 1}\cos(\phi_{\Lambda 1})$	$0.02^{+0.08}_{-0.06}$ $[-0.49, 0.18]$	$0.000^{+0.003}_{-0.002}$ $[-0.60, 0.12]$
$f_{\Lambda 1}^{Z\gamma}\cos(\phi_{\Lambda 1}^{Z\gamma})$	$0.26^{+0.30}_{-0.35}$ $[-0.40, 0.79]$	$0.000^{+0.019}_{-0.022}$ $[-0.37, 0.71]$

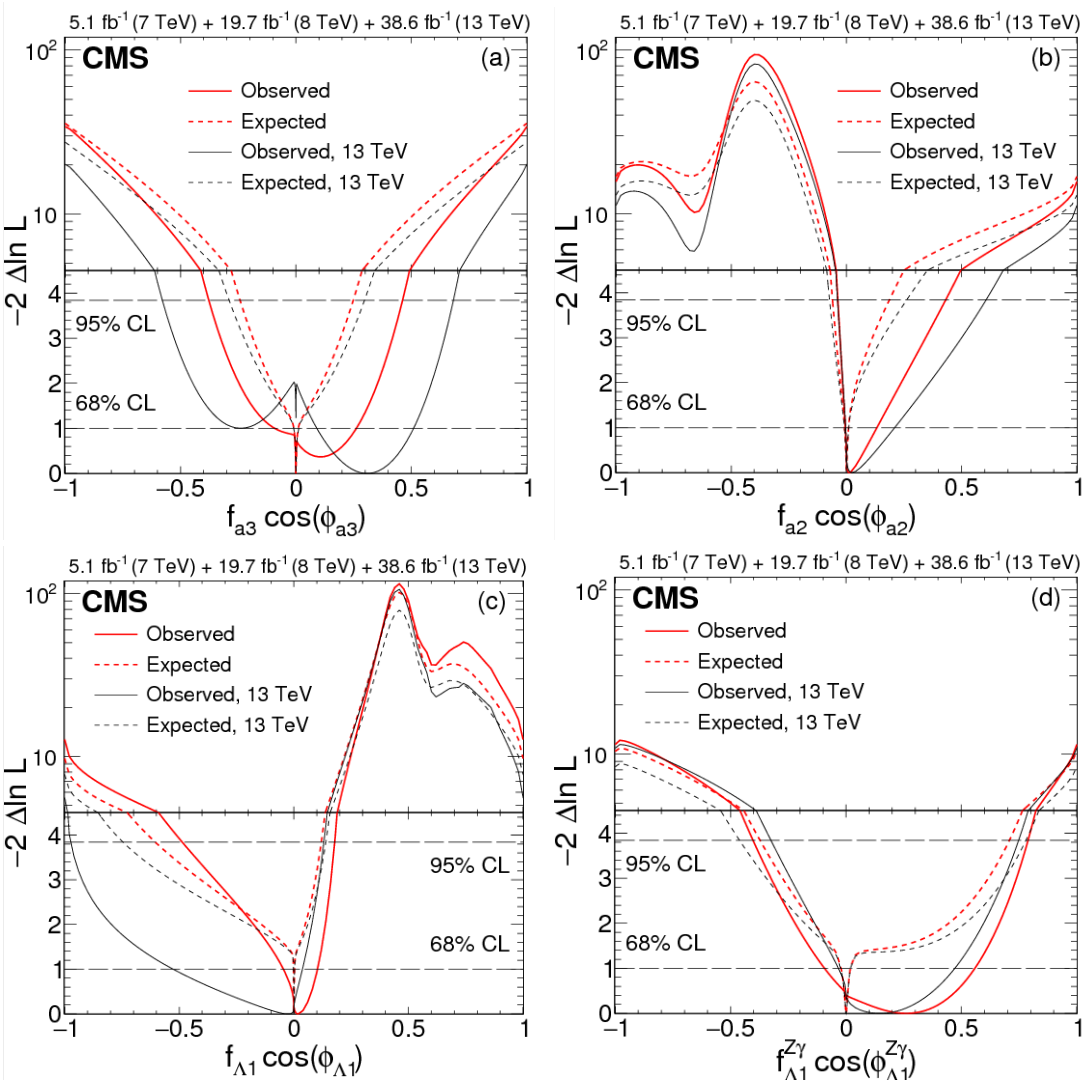

Figure 5.21 – Observed (solid) and expected (dashed) likelihood scans of $f_{a3}\cos(\phi_{a3})$ (top left), $f_{a2}\cos(\phi_{a2})$ (top right), $f_{\Lambda 1}\cos(\phi_{\Lambda 1})$ (bottom left), and $f_{\Lambda 1}^{Z\gamma}\cos(\phi_{\Lambda 1}^{Z\gamma})$ (bottom right). Results of the Run II only and the combined Run I and Run II analyses are shown.

5.4.2 Anomalous Higgs boson couplings and width from on-shell and off-shell production

As already mentioned when discussing Higgs width measurement in Section 5.1.2, the precision on some properties of the Higgs boson can be significantly improved by exploiting the events from the off-shell region. In this section we present results that combine on-shell $105 < m_{4\ell} < 140$ GeV and off-shell $m_{4\ell} > 220$ GeV regions of the $\mathrm{H} \to 4\ell$ analysis to measure Higgs boson width and anomalous HVV couplings with unprecedented precision. The details of the study and analysis techniques are described in detail in [126], and again heavily rely on the building blocks of the main $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis presented in this thesis. The data sample used in this analysis corresponds to integrated luminosities of $35.9\ \mathrm{fb}^{-1}$ collected in 2016 and $41.5\ \mathrm{fb}^{-1}$ collected in 2017 during Run II of the LHC at a center-of-mass energy of 13 TeV. These results are combined with results obtained earlier from the data collected at center-of-mass energies of 7 TeV in 2011, 8 TeV in 2012, and 13 TeV in 2015, corresponding to integrated luminosities of 5.1, 19.7, and $2.7\,\mathrm{fb}^{-1}$, respectively. The increase in both energy and luminosity leads to substantial improvement in the precision of the width measurement using the off-shell techniques, either under the assumption of SM couplings or with BSM effects.

The upper bound on Γ_{H} is set using the off-shell production method, which considers the Higgs boson production relationship between the on-shell and off-shell regions. The on-shell and off-shell yields are related by

$$\sigma^{\mathrm{on-shell}}_{\mathrm{vv}\to\mathrm{H}\to 4\ell} \propto \mu_{\mathrm{vv}\to\mathrm{H}\to 4\ell} \text{ and } \sigma^{\mathrm{off-shell}}_{\mathrm{vv}\to\mathrm{H}\to 4\ell} \propto \mu_{\mathrm{vv}\to\mathrm{H}\to 4\ell} \cdot \Gamma_{\mathrm{H}}, \tag{5.5}$$

where $\mu_{\mathrm{vv}\to\mathrm{H}\to 4\ell}$ is defined as the on-shell signal strength, the ratio of the observed number of on-shell final-state four-lepton events relative to the SM expectation, which is denoted by either μ_F for Higgs production via gluon fusion or in association with a $\mathrm{t\bar{t}}$ or $\mathrm{b\bar{b}}$, or μ_V for Higgs boson production via vector boson fusion or in association with an electroweak vector boson W or Z.

An unbinned extended maximum likelihood fit to the events split into the categories according to the three final states and three event categories. An independent fit is performed for each parameter of interest, the width and four anomalous couplings. A joint fit of the off-shell and on-shell regions is performed in order to determine the width of the H boson under certain assumptions about the couplings or to constrain anomalous couplings under certain assumptions about the width. It is assumed that all other couplings are SM-like when a certain anomalous coupling is tested. Anomalous couplings are also tested using the on-shell region only in order to avoid assumptions about the width, which becomes equivalent to a joint fit without any constraint on the width.

Results on anomalous HVV couplings exploiting only the on-shell region are reported in Table 5.14 and illustrated in Fig. 5.22.

The combination of on-shell and off-shell regions allows the setting of tighter constraints on $f_{ai}\cos(\phi_{ai})$. The on-shell region is analyzed both ways, using 2016 and 2017 data

Table 5.14 – Summary of allowed 68% CL (central values with uncertainties) and 95% CL (in square brackets) intervals on anomalous coupling parameters $f_{ai}\cos(\phi_{ai})$ obtained from the on-shell data analysis of the Run I and Run II combined dataset.

Parameter	Observed	Expected
$f_{a3}\cos(\phi_{a3})$	$-0.0001^{+0.0005}_{-0.0015}$ $[-0.16, 0.09]$	$0.0000^{+0.0019}_{-0.0019}$ $[-0.082, 0.082]$
$f_{a2}\cos(\phi_{a2})$	$0.0004^{+0.0026}_{-0.0007}$ $[-0.006, 0.025]$	$0.0000^{+0.0030}_{-0.0023}$ $[-0.021, 0.035]$
$f_{\Lambda1}\cos(\phi_{\Lambda1})$	$0.0000^{+0.0035}_{-0.0008}$ $[-0.21, 0.09]$	$0.0000^{+0.0012}_{-0.0006}$ $[-0.059, 0.032]$
$f_{\Lambda1}^{Z\gamma}\cos(\phi_{\Lambda1}^{Z\gamma})$	$0.000^{+0.355}_{-0.09}$ $[-0.17, 0.61]$	$0.000^{+0.009}_{-0.010}$ $[-0.10, 0.34]$

only and using the combined Run I and Run II data. The off-shell region is analyzed only using 2016 and 2017 data because no such analysis of the four anomalous couplings has been performed with Run I data. Therefore, the one-parameter likelihood scans of $f_{ai}\cos(\phi_{ai})$ under the assumption $\Gamma_H = \Gamma_H^{SM}$ is shown for two cases in Fig. 5.23: using 2016 and 2017 data only and in addition using earlier data in the on-shell region together with the two-parameter likelihood scans of $f_{ai}\cos(\phi_{ai})$ and Γ_H. The 68% and 95% CLs for the latter configuration are summarized in Table 5.15.

Table 5.15 – Summary of allowed 68% CL (central values with uncertainties) and 95% CL (in square brackets) intervals on anomalous coupling parameters $f_{ai}\cos(\phi_{ai})$ obtained from the on-shell and off-shell data analysis of the Run I and Run II combined dataset.

Parameter	Observed	Expected
$f_{a3}\cos(\phi_{a3})$	$0.0000^{+0.0005}_{-0.0011}$ $[-0.0067, 0.0050]$	$0.0000^{+0.0014}_{-0.0014}$ $[-0.0098, 0.0098]$
$f_{a2}\cos(\phi_{a2})$	$0.0005^{+0.0025}_{-0.0008}$ $[-0.0029, 0.0129]$	$0.0000^{+0.0011}_{-0.0017}$ $[-0.0100, 0.0117]$
$f_{\Lambda1}\cos(\phi_{\Lambda1})$	$0.0001^{+0.0020}_{-0.0010}$ $[-0.0150, 0.0501]$	$0.0000^{+0.0010}_{-0.0010}$ $[-0.0152, 0.0158]$

Limits on Γ_H are set by combining events from the on-shell and off-shell regions. Figure 5.24 (left) shows the results of the likelihood scans of Γ_H for the 2016 and 2017 period of the 13 TeV run and for the full combined dataset from collisions at 7, 8 and 13 TeV under the assumption of SM-like couplings. These combined results are listed in Table 5.16, where the Run I analysis includes both on-shell and off-shell regions in the analysis of the $H \to ZZ^\star \to 4\ell$ decay. This measurement of the Higgs boson width has surpassed significantly the precision from the Run I combined LHC measurement and is the first time ever that a lower bound is set on the Higgs boson width.

The best fitted (μ_V,μ_F) values in these results are $(0.62^{+0.57}_{-0.43}, 1.20^{+0.19}_{-0.16})$ when $\Gamma_H = \Gamma_H^{SM}$, and $(0.65^{+0.61}_{-0.45}, 1.21^{+0.19}_{-0.17})$ when Γ_H is unconstrained. The Γ_H constraints are also obtained with the $f_{a3}\cos(\phi_{a3})$, $f_{a2}\cos(\phi_{a2})$, or $f_{\Lambda1}\cos(\phi_{\Lambda1})$ parameters unconstrained and are shown in Figure 5.24 (right) and in Table 5.16. These results are obtained using combination of Run I and Run II data with a small difference from the above fit with SM-like couplings. There is no analysis of off-shell data with anomalous couplings, therefore only on-shell data are used in anomalous coupling fit with Run I data. On the other hand, 2015 data are included in the on-shell anomalous couplings study only. While

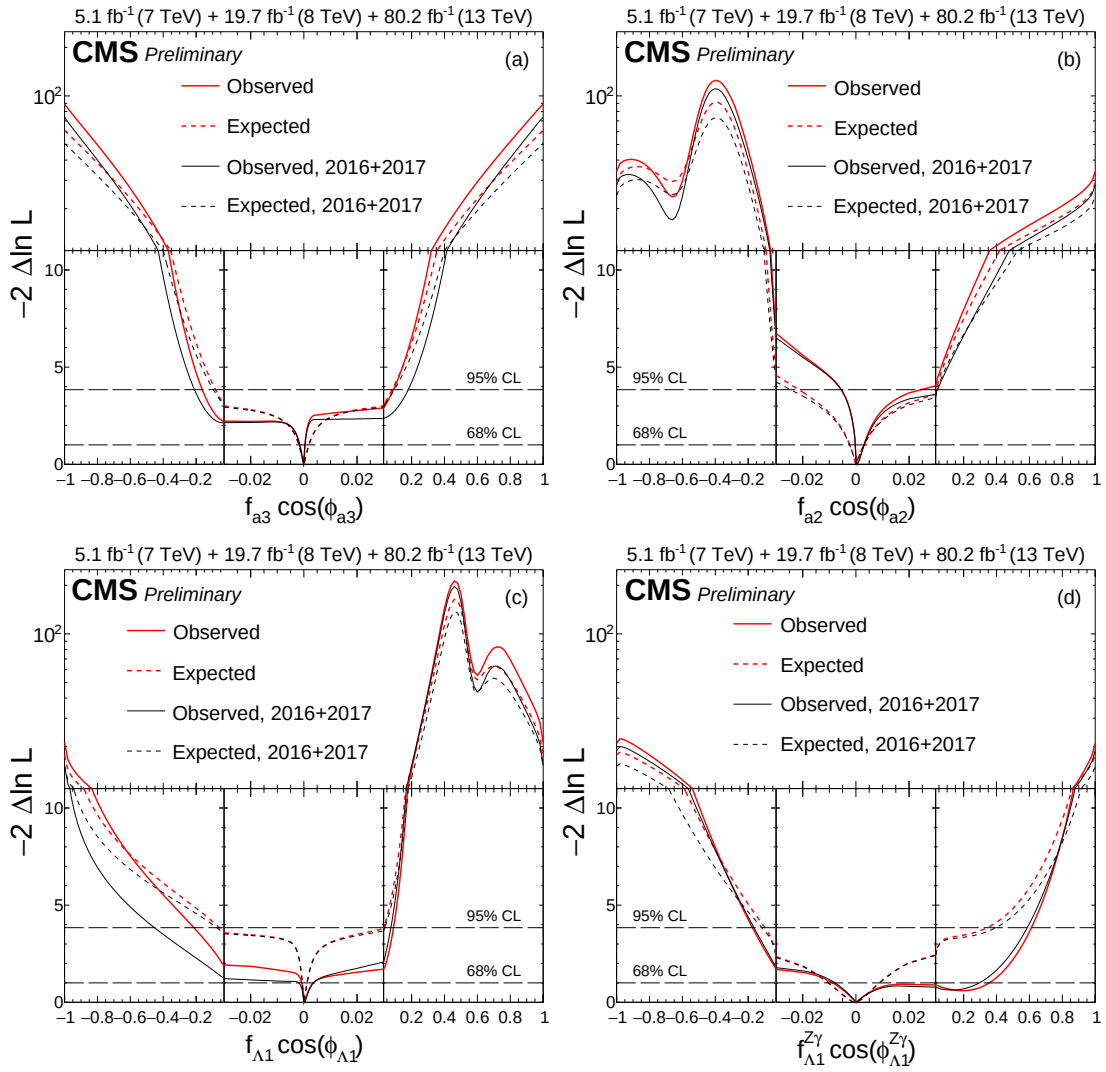

Figure 5.22 – Observed (solid) and expected (dashed) likelihood scans of $f_{a3}\cos(\phi_{a3})$ (top left), $f_{a2}\cos(\phi_{a2})$ (top right), $f_{\Lambda 1}\cos(\phi_{\Lambda 1})$ (bottom left), and $f_{\Lambda 1}^{\mathrm{Z}\gamma}\cos(\phi_{\Lambda 1}^{\mathrm{Z}\gamma})$ (bottom right) using on-shell events only. Results of analysis of the data from 2016 and 2017 only (black) and the combined Run I and Run II analysis (red) are shown. The dashed horizontal lines show the 68% and 95% CL regions.

the expected Γ_{H} constraints are similar but somewhat looser with the unconstrained anomalous coupling parameters, fluctuations in the data lead to somewhat tighter observed constraints.

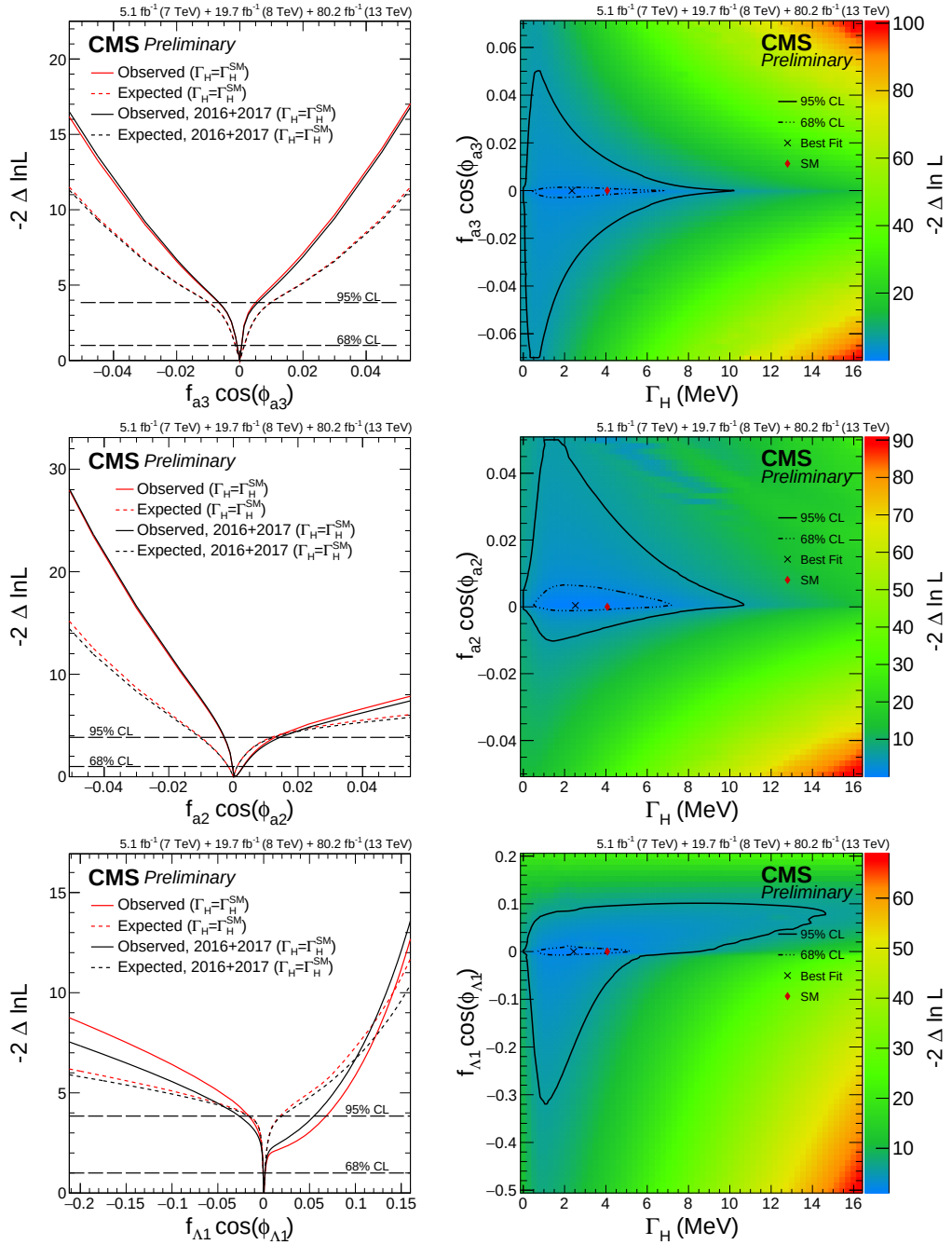

Figure 5.23 – Constraints on $f_{a3}\cos(\phi_{a3})$ (top), $f_{a2}\cos(\phi_{a2})$ (middle), and $f_{\Lambda 1}\cos(\phi_{\Lambda 1})$ (bottom) under the assumption $\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$ (left) and with Γ_{H} unconstrained (right). Left plots: Results of analysis of the data from 2016 and 2017 only (black) and the combined Run II and Run I analysis (red) are shown. The dashed horizontal lines show the 68% and 95% CL regions. Observed (solid) and expected (dashed) likelihood scans are shown. Right plots: Observed 2D likelihood scans are shown for the combined Run II and Run I analysis. The 68% and 95% CL regions are indicated with the dashed lines.

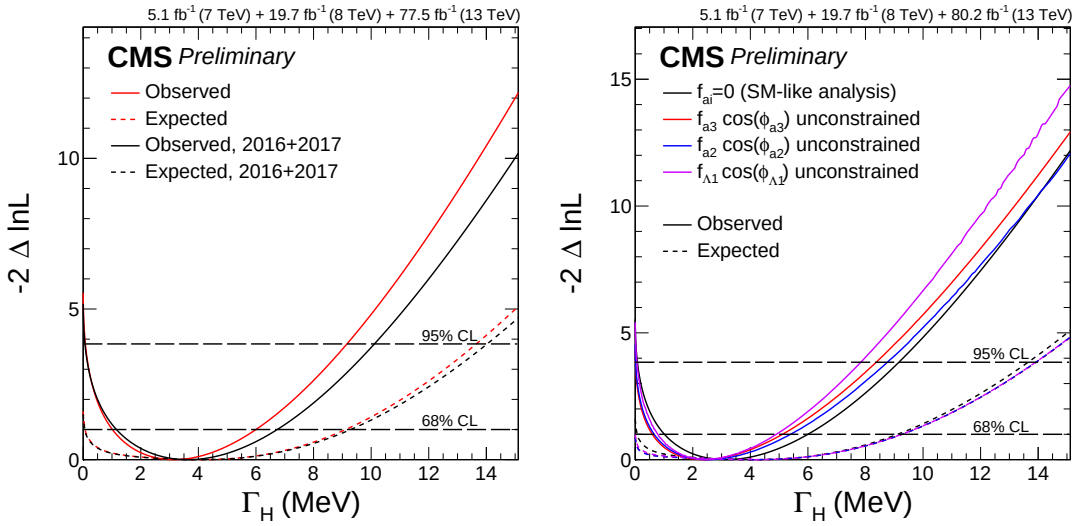

Figure 5.24 – Observed (solid) and expected (dashed) likelihood scans of Γ_{H}. Left plot: Results of analysis of the data from 2016 and 2017 only (black) and the combined Run I and Run II analyses (red) are shown for the SM-like couplings. Right plot: Results of analysis of the data from the combined Run I and Run II analyses for the SM-like couplings and with three anomalous coupling parameters of interest unconstrained: $f_{a3}\cos(\phi_{a3})$ (red), $f_{a2}\cos(\phi_{a2})$ (blue), and $f_{\Lambda 1}\cos(\phi_{\Lambda 1})$ (violet). The dashed horizontal lines show the 68% and 95% CL regions.

Table 5.16 – Summary of the total width Γ_{H} measurement, showing allowed 68% CL (central values with uncertainties) and 95% CL (in square brackets). The limits are reported for the SM-like couplings using the Run I and Run II combination.

Parameter	Observed	Expected
Γ_{H}(MeV)	$3.2^{+2.8}_{-2.2}$ $[0.08, 9.16]$	$4.1^{+5.0}_{-4.0}$ $[0.0, 13.7]$

Table 5.17 – Summary of the total width Γ_{H} measurement, showing allowed 68% CL (central values with uncertainties) and 95% CL (in square brackets). The Γ_{H} limits are reported for the anomalous coupling parameter of interest unconstrained using the Run I and Run II combination.

Parameter	Unconstrained Parameter	Observed	Expected
Γ_{H}(MeV)	$f_{a3}\cos(\phi_{a3})$	$2.4^{+2.7}_{-1.8}$ $[0.02, 8.38]$	$4.1^{+5.2}_{-4.1}$ $[0.0, 13.9]$
Γ_{H}(MeV)	$f_{a2}\cos(\phi_{a2})$	$2.5^{+2.9}_{-1.8}$ $[0.02, 8.76]$	$4.1^{+5.2}_{-4.1}$ $[0.0, 13.9]$
Γ_{H}(MeV)	$f_{\Lambda 1}\cos(\phi_{\Lambda 1})$	$2.4^{+2.5}_{-1.6}$ $[0.06, 7.84]$	$4.1^{+5.2}_{-4.1}$ $[0.0, 13.9]$

Chapter 6

Conclusion and perspectives

While this thesis work is being written, Run II is coming to an end with only a couple of months of data taking left. It is hard to compete with Run I that brought us a discovery of the long sought Higgs boson, but it is safe to say that Run II will also be considered a major success. Already with two thirds of collected data that has been analyzed, both the ATLAS and CMS Collaborations have now observed all main production and decay modes of the Higgs boson. The Higgs boson candle is shining brightly at 13 TeV with hundreds of fully reconstructed Higgs boson candidates in the golden channel, which has once again profiled itself as a leader in Higgs boson precision properties measurements. All of the measurements are found to be in agreement with the Standard Model predictions within their uncertainties, and thanks to the great efforts from the theoretical and experimental side, these uncertainties are still decreasing even though we are experiencing harsher collision conditions. Because of this, we are entering a new era of precision measurements where we will not be statistically limited anymore. This gives us an ability to measure small deviations from the SM expectation that could provide strong hints about possible physics beyond the SM.

The $\mathrm{H} \to \mathrm{ZZ}^{\star} \to 4\ell$ analysis has advanced in many aspects, from improved object selection, developing complex event categorization targeting all main Higgs production modes, to new kinematic observables and techniques that all together improve the precision on the measurements. Besides the significant efforts put to improve the analysis, it is also one of the very few analyses that produces the results regularly, analyzing data sets as they come. Run II data collected from 2015 to 2017 has already been analyzed and many properties measurements have been performed. The impressive list features the CMS Higgs boson mass and width measurements that are the current most precise measurements in the world, surpassing the combined ATLAS and CMS measurements from Run I and a measurement of the signal strength modifiers of main Higgs boson production modes combining $77.4\ \mathrm{fb}^{-1}$ of data collected in 2016 and 2017.

Analysis will continue to improve and will for sure produce a set of impressive results on the full Run II data set. World will be waiting to see if once again everything is in agree-

ment with the SM expectations or we will be lucky to experience another excitement of discovering something new. However, one has to keep in mind that we are just getting started. In the grand scheme of High Luminosity LHC that is planned after Run III, we have only recorded a few percent of the expected data. Even though many projections exist no one can tell for sure what to expect from the huge amount of data that will be recorded and carefully analyzed. Only one thing is for sure, very exciting times are ahead of us and you do not want to miss them!

Bibliography

[1] Particle Data Group. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[2] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321, 1964.

[3] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[4] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585, 1964.

[5] J. Goldstone. Field theories with « superconductor » solutions. *Il Nuovo Cimento (1955-1965)*, 19(1):154–164, Jan 1961.

[6] J. Baglio and A. Djouadi. Higgs production at the lHC. *JHEP*, 03:055, 2011.

[7] J. Baglio and A. Djouadi. Predictions for Higgs production at the Tevatron and the associated uncertainties. *JHEP*, 10:064, 2010.

[8] C. Anastasiou et al. High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP*, 05:058, 2016.

[9] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016.

[10] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. 2012.

[11] M. K. Gaillard J. Ellis and D.V. Nanopoulos. A phenomenological profile of the higgs boson. *Nuclear Physics B*, 106:292 – 340, 1976.

[12] J. E. Weinberg. *Radiative corrections as the origin of spontaneous symmetry breaking*. PhD thesis, Harvard U., 1973.

[13] C. Quigg B. W. Lee and H. B. Thacker. The Strength of Weak Interactions at Very High-Energies and the Higgs Boson Mass. *Phys. Rev. Lett.*, 38:883–885, 1977.

[14] J. Ellis et al. A Historical Profile of the Higgs Boson. 2012.

[15] A. Djouadi. The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys. Rept.*, 457:1–216, 2008.

[16] G. Ridolfi. Search for the Higgs boson: Theoretical perspectives. In *Results and perspectives in particle physics. Proceedings, 15th Rencontres de physique de la Vallee d'Aoste, La Thuile, Italy, March 4-10, 2001*, pages 291–304, 2001.

[17] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565:61–75, 2003.

[18] CDF and D0 Collaborations. Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.2 fb^{-1} of Data. In *Proceedings, 46th Rencontres de Moriond on Electroweak Interactions and Unified Theories: La Thuile, Italy, March 13-20, 2011*, 2011.

[19] ATLAS Collaboration. Combined search for the Standard Model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev.*, D86:032003, 2012.

[20] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[21] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV. *JHEP*, 06:081, 2013.

[22] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[23] Tevatron New Physics Higgs Working Group. Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data. 2012.

[24] CDF and D0 Collaborations. Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron. *Phys. Rev. Lett.*, 109:071804, 2012.

[25] CDF and D0 Collaborations. Higgs Boson Studies at the Tevatron. *Phys. Rev.*, D88(5):052014, 2013.

[26] CMS Collaboration. Measurement of the properties of a Higgs boson in the four-lepton final state. *Phys. Rev.*, D89(9):092007, 2014.

[27] ATLAS Collaboration. Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data. *Phys. Rev.*, D90(5):052004, 2014.

[28] CMS Collaboration. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *Eur. Phys. J.*, C74(10):3076, 2014.

[29] CMS Collaboration. 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. *Eur. Phys. J.*, C75(5):212, 2015.

[30] ATLAS and CMS Collaborations. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[31] CMS Collaboration. Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs. *Phys. Rev. Lett.*, 110(8):081803, 2013.

[32] ATLAS Collaboration. Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment. *Eur. Phys. J.*, C76(1):6, 2016.

[33] ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.

[34] CMS Collaboration. Observation of $\mathrm{t\bar{t}H}$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018.

[35] ATLAS Collaboration. Measurement of the electroweak production of dijets in association with a Z-boson and distributions sensitive to vector boson fusion in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *JHEP*, 04:031, 2014.

[36] ATLAS Collaboration. Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector. 2018.

[37] CMS Collaboration. Observation of Higgs boson decay to bottom quarks. 2018.

[38] CMS Collaboration. Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. *Eur. Phys. J.*, C75(2):66, 2015.

[39] ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. 2018.

[40] CMS Collaboration. Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector. *Phys. Lett.*, B779:283–316, 2018.

[41] ATLAS Collaboration. Cross-section measurements of the Higgs boson decaying to a pair of tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2018.

[42] ATLAS Collaboration. Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. Technical Report ATLAS-CONF-2018-031, CERN, Geneva, Jul 2018.

[43] CMS Collaboration. Combined measurements of the Higgs boson's couplings at $\sqrt{s} = 13$ TeV. 2018.

[44] ATLAS Collaboration. Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *Phys. Rev.*, D91(1):012006, 2015.

[45] CMS Collaboration. Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs. *Phys. Lett.*, B736:64–85, 2014.

[46] ATLAS Collaboration. Constraints on the off-shell Higgs boson signal strength in the high-mass ZZ and WW final states with the ATLAS detector. *Eur. Phys. J.*, C75(7):335, 2015.

[47] CMS Collaboration. Limits on the Higgs boson lifetime and width from its decay to four charged leptons. *Phys. Rev.*, D92(7):072010, 2015.

[48] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett.*, B726:120–144, 2013.

[49] CMS Collaboration. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Phys. Rev.*, D92(1):012004, 2015.

[50] ATLAS Collaboration. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Eur. Phys. J.*, C75(10):476, 2015. [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[51] ATLAS Collaboration. Fiducial and differential cross sections of Higgs boson production measured in the four-lepton decay channel in pp collisions at $\sqrt{s}$=8 TeV with the ATLAS detector. *Phys. Lett.*, B738:234–253, 2014.

[52] CMS Collaboration. Measurement of differential and integrated fiducial cross sections for Higgs boson production in the four-lepton decay channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 04:005, 2016.

[53] ATLAS Collaboration. Search for an additional, heavy Higgs boson in the $H \rightarrow ZZ$ decay channel at $\sqrt{s} = 8$ TeV in pp collision data with the ATLAS detector. *Eur. Phys. J.*, C76(1):45, 2016.

[54] CMS Collaboration. Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons. *JHEP*, 10:144, 2015.

[55] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[56] CMS Collaboration. CMS, the Compact Muon Solenoid: Technical proposal. 1994.

[57] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[58] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Upgrade. 2012.

[59] CMS Collaboration. CMS ECAL Laser monitoring up to 2017, $\pi^0/\eta \to \gamma\gamma$ spectrum and monitoring, ES calibration. Jul 2017.

[60] CMS Collaboration. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013. [JINST8,9009(2013)].

[61] CMS Collaboration. CMS Tracking Performance Results from early LHC Operation. *Eur. Phys. J.*, C70:1165–1192, 2010.

[62] CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC. *JINST*, 8:P11002, 2013.

[63] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[64] S. Regnard. *Measurements of Higgs boson properties in the four-lepton final state at sqrt(s) = 13 TeV with the CMS experiment at the LHC.* PhD thesis, Ecole Polytechnique, 2016.

[65] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[66] CMS Collaboration. CMS Physics. 2006.

[67] P. Billoir. Progressive track recognition with a Kalman like fitting procedure. *Comput. Phys. Commun.*, 57:390–394, 1989.

[68] CMS Collaboration. Track Reconstruction Performance in CMS. *Nucl. Phys. Proc. Suppl.*, 197:275–278, 2009.

[69] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[70] CMS Collaboration. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV. *JINST*, 10(06):P06005, 2015.

[71] S. Baffioni et al. Electron reconstruction in CMS. Technical Report CMS-NOTE-2006-040, CERN, Geneva, Feb 2006.

[72] W. Heitler H. Bethe. On the stopping of fast particles and on the creation of positive electrons. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 146(856):83–112, 1934.

[73] J. Strologas. Performance of Jet reconstruction in CMS at 13 TeV. *PoS*, ICHEP2016:736, 2016.

[74] CMS Collaboration. Determination of the Jet Energy Resolutions and Jet Reconstruction Efficiency at CMS. 2009.

[75] R. Atkin. Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645(1):012008, 2015.

[76] G. P. Salam M. Cacciari and G. Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.

[77] G. P. Salam M. Cacciari and G. Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.

[78] CMS Collaboration. Studies of Higgs boson production in the four-lepton final state at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-15-004, CERN, Geneva, 2016.

[79] CMS Collaboration. Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 10:132, 2011.

[80] M. Kovac. *Electron measurements and constraints on the spin-parity and anomalous HVV couplings in the H to ZZ$^{(*)}$ to 4 leptons decay channel with the CMS detector*. PhD thesis, Ecole Polytechnique, 2015.

[81] R.D. Ball et al. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys.*, B855:153–221, 2012.

[82] R. D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017.

[83] T. Sjöstrand et al. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[84] V. Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016.

[85] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[86] J. Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

[87] S. Alioli et al. NLO vector-boson production matched with shower in POWHEG. *JHEP*, 07:060, 2008.

[88] P. Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[89] S. Frixione et al. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[90] E. Bagnaschi et al. Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. *JHEP*, 02:088, 2012.

[91] P. Nason and C. Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 02:037, 2010.

[92] G. Luisoni et al. $HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.

[93] Y. Gao et al. Spin determination of single-produced resonances at hadron colliders. *Phys. Rev.*, D81:075022, 2010.

[94] S. Bolognesi et al. On the spin and parity of a single-produced resonance at the LHC. *Phys. Rev.*, D86:095031, 2012.

[95] I. Anderson et al. Constraining anomalous HVV interactions at proton and lepton colliders. *Phys. Rev.*, D89(3):035007, 2014.

[96] A. Gritsan et al. Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques. *Phys. Rev.*, D94(5):055023, 2016.

[97] M. Grazzini and H. Sargsyan. Heavy-quark mass effects in Higgs boson production at the LHC. *JHEP*, 09:129, 2013.

[98] D. de Florian et al. Higgs boson production at the LHC: transverse momentum resummation effects in the $\mathrm{H} \to 2\gamma$, $\mathrm{H} \to \mathrm{WW} \to \ell\nu\ell\nu$ and $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ decay modes. *JHEP*, 06:132, 2012.

[99] K. Hamilton et al. NNLOPS simulation of Higgs boson production. *JHEP*, 10:222, 2013.

[100] J. M-Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205-206:10–15, 2010.

[101] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[102] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.

[103] CMS Collaboration. Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[104] CMS Collaboration. Identification of b quark jets at the CMS Experiment in the LHC Run 2. Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[105] M. Grazzini et al. ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD. *Phys. Lett.*, B750:407–410, 2015.

[106] M. Bonvini et al. Signal-background interference effects for $gg \to H \to W^+W^-$ beyond leading order. *Phys. Rev.*, D88(3):034032, 2013.

[107] K. Melnikov and M. Dowling. Production of two Z-bosons in gluon fusion in the heavy top quark approximation. *Phys. Lett.*, B744:43–47, 2015.

[108] C.S. Li et al. Soft gluon resummation in the signal-background interference process of $gg(\to h^*) \to ZZ$. *JHEP*, 08:065, 2015.

[109] G. Passarino. Higgs CAT. *Eur. Phys. J.*, C74:2866, 2014.

[110] S. Catani and M. Grazzini. An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC. *Phys. Rev. Lett.*, 98:222002, 2007.

[111] M. Grazzini. NNLO predictions for the Higgs boson signal in the $H \to WW \to l\nu l\nu$ and $H \to ZZ \to 4l$ decay channels. *JHEP*, 02:043, 2008.

[112] L. Landau. On the energy loss of fast particles by ionization. *J. Phys.(USSR)*, 8:201–205, 1944.

[113] S. van der Meer. Calibration of the effective beam height in the ISR. Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968.

[114] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[115] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018.

[116] J. Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys.*, G43:023001, 2016.

[117] The LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

[118] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1):60–62, 03 1938.

[119] G. Cowan et al. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].

[120] CMS Collaboration. Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C76(1):13, 2016.

[121] CMS Collaboration. Measurements of properties of the Higgs boson and search for an additional resonance in the four-lepton final state at sqrt(s) = 13 TeV. Technical Report CMS-PAS-HIG-16-033, CERN, Geneva, 2016.

[122] CMS Collaboration. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:047, 2017.

[123] CMS Collaboration. Measurements of properties of the Higgs boson in the four-lepton final state at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-18-001, CERN, Geneva, 2018.

[124] F. Garwood. Fiducial limits for the poisson distribution. *Biometrika*, 28(3-4):437–442, 1936.

[125] CMS Collaboration. Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state. *Phys. Lett.*, B775:1–24, 2017.

[126] CMS Collaboration. Measurements of Higgs boson properties from on-shell and off-shell production in the four-lepton final state. Technical Report CMS-PAS-HIG-18-002, CERN, Geneva, 2018.

université PARIS-SACLAY | ÉCOLE DOCTORALE PHENIICS

Titre : Mesure des propriétés du boson de Higgs via ses désintégrations en quatre leptons dans les collisions pp à 13 TeV dans le détecteur CMS

Mots clés : boson de Higgs, expérience CMS, collisioneur LHC, modèle standard, électrons

Résumé : Cette thèse porte sur les mesures des propriétés du boson de Higgs dans le canal à quatre leptons pour des collisions proton-proton à une énergie dans le centre de masse de 13 TeV avec le détecteur CMS. Mes contributions originales sont doubles: une amélioration des mesures d'efficacité de la sélection des électrons et une prise en charge de l'analyse des données enregistrées en 2017 en tant qu'auteur principal. Les résultats de l'analyse des données du ≪ run II ≫ sont présentés pour deux séries de données enregistrées respectivement en 2016 et 2017, correspondant à des luminosités intégrées de 35.9 fb^{-1} et 41.5 fb^{-1}, respectivement. Les modificateurs d'intensité du signal par rapport à la prédiction du modèle standard, la masse et la largeur de décroissance du boson, les sections efficaces différentielles et inclusives sont mesurés, pour le boson de Higgs de 125 GeV. Enfin, une mesure des modificateurs d'intensité du signal sur 77.4 fb^{-1} des données collectées en 2016 et 2017 effectué et est à nouveau trouvé compatible avec les prédictions du modèle. De nouvelles frontières sont atteintes, car les contributions statistiques et systématiques à l'incertitude totale sur la mesure de la force du signal inclusif sont de taille similaire.

Title : Measurements of Higgs boson properties in the four-lepton channel in pp collisions at centre-of-mass energy of 13 TeV with the CMS detector

Keywords : Higgs boson, CMS experiment, LHC collider, standard model, electrons

Abstract : This thesis presents the measurements of Higgs boson properties in the four-lepton decay channel in proton-proton collisions at centre-of-mass energy of 13 TeV with the CMS detector. Significant improvements in many aspects of the analysis are introduced with respect to the previous reported measurements by the CMS Collaboration. My contribution is twofold, improving the electron selection efficiency measurements and leading the analysis of data recorded in 2017 as the main author. Results of the analysis of Run II data are presented for two data sets recorded in 2016 and 2017, corresponding to integrated luminosities of 35.9 fb^{-1} and 41.5 fb^{-1}, respectively. The signal strength modifiers relative to the standard model prediction, the mass and decay width of the boson, differential and inclusive cross sections are measured. All results are in good agreement with standard model expectations for a Higgs boson within the uncertainties. Finally, a measurement of the signal strength modifiers on combined 77.4 fb^{-1} of data collected in 2016 and 2017 is also performed and is again found to be compatible with the SM predictions. New frontiers are reached as the statistical and systematical contributions to the total uncertainty on the inclusive signal strength measurement are considerably reduced and of the same size.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France