

Praktikumsversuch zur Higgs-Boson Suche im 'goldenen Kanal' mit CERN Open Data

(Lab course for the Higgs boson search in the
'golden channel' with CERN Open Data)

Bachelorarbeit
von

Artur Monsch

am Institut für Experimentelle Teilchenphysik

Referent: Prof. Dr. Günter Quast
Korreferent: Priv. Doz. Dr. Roger Wolf

Bearbeitungszeit: 15.08.2019 – 14.02.2020

Inhaltsverzeichnis

Einleitung	1
1. Grundlagen	3
1.1. Invariante Masse	3
1.2. Pseudorapidität	3
1.3. Luminosität	4
1.4. Wirkungsquerschnitt	4
1.5. Higgs-Boson des Standardmodells	5
1.6. Der CMS-Detektor	7
1.7. Invariantes Massenspektrum	8
1.8. Parameterschätzung	9
1.8.1. Methode der kleinsten Quadrate	9
1.8.2. Likelihood-Methode	10
2. Rekonstruktion des Massenspektrums der vier Lepton	13
2.1. Konvertierung von .AOD/.AODSIM nach .csv	13
2.2. Filter	15
2.2.1. Detektorspezifische Filter	16
2.2.2. Kinematische Filter	16
2.3. Rekonstruktion	18
2.4. Doppeltzählung und Skalierung der MC-Simulation	20
3. Statistische Analyse	23
3.1. Untergrundstudie	23
3.2. Signalbreitenabschätzung	25
3.3. Abschätzung lokaler Signifikanz – der lokale p_0 - Wert	27
3.4. Signalstudie	29
3.5. Lokale Signifikanz	31
3.5.1. Massenbestimmung des SM-Higgs-Bosons	32
3.5.2. Bewertung der Methoden	33
4. Entwurf des Praktikumversuchs	35
4.1. Versuchsbeginn	35
4.2. Filter- und Rekonstruktionsschritt	37
4.3. Analyseschritt	37
4.4. Betrugsvermeidung im Rahmen des Praktikums	39
Zusammenfassung	41
Anhang	43
A. Untergrundparametrisierung	43
B. Signalbreitenabschätzung - Signalverteilungsparametrisierung	47
C. p_0 - Wertabschätzung für Histogramme mit wenig Statistik	49

D. Signalparametisierung	50
Literaturverzeichnis	53

Abbildungsverzeichnis

1.1.	Higgs-Boson Produktionskanäle aus pp -Kollisionen am LHC [15]. Die Gluonfusion 1.1a und die Vektorbosonfusion 1.1b sind hierbei die dominanten Produktionskanäle.	5
1.2.	Verzweigungsverhältnisse des Higgs-Bosons mit den geschätzten Unsicherheiten [18]. Der Zerfallskanal in ein Photonenpaar (1.2b) liefert die wichtige Information, dass das ursprüngliche Teilchen einen Spin 0 besitzt. Der Zerfallskanal nach vier Leptonen (1.2a) wird im Rahmen der Arbeit genauer betrachtet.	6
1.3.	Ein Längsschnitt des Detektors, parametrisiert durch die Pseudorapidität, dem Radius des Detektors und der Ausdehnung entlang der Strahlachse [20].	7
1.4.	Vier Leptonen invariantes Massenspektrum in einem Massenbereich zwischen 70 GeV und 181 GeV. Eingezeichnet sind neben den erwarteten Verteilungen für die Untergründe und einer Signalverteilung für ein Standardmodell Higgs-Boson mit einer Masse von 125 GeV und die Messung mit den Unsicherheiten, die am CMS Detektor aufgenommen wurde [3].	9
2.1.	Masseverteilung der rekonstruierten Z_1 - und Z_2 -Bosonen für einen Massenbereich von $106 \text{ GeV} < m_{4\ell} < 151 \text{ GeV}$. Die MC-Simulationen wurden hierfür an die Daten skaliert (Abschnitt 2.4).	18
2.2.	Visualisierung der Anzahl an Ereignissen nach jeweiligem Selektionsschritt, aufgeteilt in die Zerfallskanäle 4μ , $4e$ und $2e2\mu$. Die Prüfung nach der Minimalanzahl an Leptonen und einer existierenden elektrischen Ladungskombination erfolgt nach jeder Filteranwendung bis zum $m_{2\ell}$ -Ausschluss.	19
2.3.	$m_{4\ell}$ -Spektrum in einem Massebereich von 70 GeV und 180 GeV. Abgebildet sind inklusiv die Zerfallskanäle 4μ , $4e$ und $2e2\mu$, wobei die Messung aus dem gesamten Run 2012 stammt. Die Normierung der MC-Simulationen ist nach (2.4) erfolgt.	21
2.4.	$m_{4\ell}$ -Verteilung im Bereich 106 GeV und 151 GeV mit einer Binbreite von 3 GeV. Zusätzlich sind zu der rekonstruierten Untergrundverteilung (blau) und den gemessenen Ereignissen (schwarz) exemplarisch drei mögliche Signal-MCs der Massen 115 GeV, 125 GeV und 140 GeV.	22
3.1.	Histogrammanpassung des an die Daten skalierten Untergrund MC durch das verwendete Modell (3.1) aus einer Linearkombination von Legendrepolynomen bis zum Grad $n = 2$ durch die Softwareumgebung KAFE2 [51]. Das Ratio gibt dabei das Verhältnis zwischen der bestimmten, an die Daten skalierten Wahrscheinlichkeitsdichte an und die für die Anpassung verwendeten Bins. Die Unsicherheiten wurden für die Anpassung mit dem Faktor $\sqrt{\chi_{\min}^2} = 1.015$ skaliert.	24
3.2.	Visualisierung der beschriebenen Skalierung der Unsicherheiten mit dem Faktor $\sqrt{\chi_{\min}^2}$ für unterschiedliche Polynomgrade.	25

3.3.	Anpassung an ein Histogramm des Signal-MC eines SM-Higgs-Bosons der Masse 125 GeV mit der Softwareumgebung <code>KAFE2</code> [51]. Das Signal-MC wurde an die vorhandene Messung skaliert. Das Ratio gibt das Verhältnis zwischen der durch die Anpassung bestimmten und skalierten Wahrscheinlichkeitsdichte und der vorhandenen Messung an. Der Skalierungsfaktor ist $\sqrt{\chi_{\min}^2} = 7,77$	26
3.4.	Lineare Anpassung der unterschiedlichen Breiten σ_G der Gaußverteilungen in Abhängigkeit von $m_{4\ell}$ durch die Verwendung der Softwareumgebung <code>KAFE2</code> [51]. Jeder der Punkte wurde einer Signal-MC der entsprechenden Higgsmasse durch eine Gaußparametrisierung entnommen. Die Anpassung der Signal-MC Verteilungen für weitere Higgsmassen ist im Anhang B hinterlegt. Das Ratio gibt das Verhältnis zwischen der bestimmten Funktion und für die Anpassung verwendeten Messpunkten an.	26
3.5.	Aufgetragene Schätzung von dem lokalen p_0 -Wert über $m_{4\ell}$. Die in Rot eingezeichneten Signifikanzen sind nach Gleichung (3.7) berechnet. Im Bildausschnitt nicht zu sehen ist die 5σ -Grenze für eine Entdeckung die bei einem p_0 Wert von $\simeq 2,87 \cdot 10^{-7}$ liegt.	28
3.6.	Anpassung an ein Histogramm der an die Daten skalierten Signal-MC für eine Higgsmasse von $m_H = 125$ GeV mittels einer DSCB-Verteilung (3.8) durch die Softwareumgebung <code>KAFE2</code> [51]. Das Ratio zeigt das Verhältnis zwischen der angepassten Verteilung, die an die für die Anpassung verwendeten Bins skaliert wurde und den Bins selbst. Der Skalierungsfaktor für die Unsicherheiten ist $\sqrt{\chi_{\min}^2} = 1,292$	30
3.7.	Bestimmung des optimalen Signalanteils α_S mit der Likelihood Methode (1.12) und dessen Unsicherheiten, eingezeichnet in dem gezeigten Ausschnittsfenster. Auf der y-Achse sind nach (3.12) die Signifikanzen eingezeichnet, die $-2 \ln \frac{\mathcal{L}(\alpha_S)}{\mathcal{L}_{\min}}$ für den Punkt $\alpha_S = 0$ annehmen würden. Für die eingezeichnete profile-Likelihood werden die optimalen Parameterwerte θ_{0_S} und θ_{0_U} als festgehaltene Parameterwerte verwendet. Die in die Likelihood einfließenden Verteilungen ist für das Signal die DSCB-Verteilung (3.8) und für den Untergrund die Linearkombination aus Legendrepolyomen bis zum zweiten Grad (3.1).	31
3.8.	Bestimmung des optimalen Signalanteils α_S und der optimalen Masse des SM-Higgs-Bosons als ein zusätzlicher Minimierungsparameter durch die <i>profiled likelihood</i> Methode der extended Likelihood, Gleichung (1.12), zusammen mit ihren Unsicherheiten. Auf der y-Achse sind die Signifikanzen eingezeichnet, welche nach Gleichung (3.12) für den Punkt $\alpha_S = 0$ gelten. Die optimalen Parameterwerte θ_{0_S} (ohne μ) und θ_{0_U} werden festgehalten. Für die Signalverteilung wird die DSCB-Verteilung (3.8) verwendet. Der Untergrund wird durch die Linearkombination aus Legendrepolyomen bis zum zweiten Grad (3.1) beschrieben.	32

4.1.	Darstellung von Zerfällen, deren detektierte Endprodukte vier Myonen bzw. vier Elektronen enthalten, die alle an den Signalprozess gestellten Bedingungen erfüllen. In beiden Abbildungen sind die vom Tracker rekonstruierten Bahnen der nach der Kollision entstandenen Teilchen als gelbe Linien eingezeichnet. In dem 4μ -Zerfall sind die rekonstruierten globalen Myonen in rot eingezeichnet. Zusätzlich sind zum Teil die Driftröhren und Violdrahtkammern angedeutet. Die Komponenten, die die Myonen detektiert haben sind durch eine Umrandung hervorgehoben. Die Energie der Elektronen sind am Rand des schematisch dargestellten ECAL Barrels (blau) als grüne Balken eingezeichnet. Ein Teil des Trackers ist zudem in gelb schematisch im ECAL Barrel dargestellt.	36
4.2.	Beispielhaft ist ein Ausschnitt aus dem Filterprozess dargestellt, der von Versuchsteilnehmern implementiert werden kann. Verwendet werden die Daten aus RunC mit 4μ im Endzustand.	38
A.1.	Anpassungen an ein Histogramm der Untergrund MC durch verschiedene Linearkombinationen aus Legendrepolyomen bis zum Grad n mit der Softwareumgebung <code>KAFE2</code> [51]. Das Ratio gibt das Verhältnis zwischen den für die Anpassung verwendeten Daten und der skalierten Wahrscheinlichkeitsdichte an. Die Unsicherheiten für die Anpassungen wurden mit den Faktoren $\sqrt{\chi_{\min}^2}$ multiplikativ skaliert, derart, das $\chi_{\text{ndf}}^2 = 1$ erfüllt ist. Diese sind in der Tabelle 2.2 zusammengefasst.	45
A.2.	Darstellung der Korrelationen zwischen den Parametern für die Anpassung der Untergrund MC mittels eines nicht zentrierten Polynomen zweiten Grades (A.2b) und einer Linearkombination von zentrierten Legendrepolyomen bis zum zweiten Grad (A.2a).	46
B.3.	Anpassung an ein Histogramm der Signal MC unterschiedlicher Massen des Higgs-Bosons durch eine Gaußverteilung für die Signalbreitenabschätzung. Für die Anpassung wird die Softwareumgebung <code>KAFE2</code> [51] verwendet. Die Signal-MC wurde an die vorhandene Messung skaliert. Das Ratio gibt hierbei das Verhältnis zwischen den skalierten Wahrscheinlichkeitsdichten und der für die Anpassung verwendeten Daten an. Die Unsicherheiten wurden mit dem Faktor $\sqrt{\chi_{\min}^2}$ derart skaliert, dass $\chi_{\text{ndf}}^2 = 1$ erfüllt ist.	48
C.4.	Abschätzung des p_0 -Wertes bei unterschiedlichen $m_{4\ell}$ Werten für Daten die nicht in einem Histogramm zusammengefasst sind (durchgezogene Linie), sowie sie in Kapitel 3 durchgeführt wurde und drei beispielhafte Histogramme mit unterschiedlichen Binbreiten.	49
D.5.	Anpassungen an ein Histogramm der Signal-MC für eine Masse des Higgs-Bosons von 125 GeV. Die Anpassungen erfolgen in der Softwareumgebung <code>KAFE2</code> [51]. Das Ratio gibt das Verhältnis der skalierten Wahrscheinlichkeitsdichten und der verwendeten Daten an. Die Unsicherheiten wurden für die Anpassungen mit dem jeweiligen Faktor $\sqrt{\chi_{\min}^2}$ multiplikativ skaliert und sind in Tabelle 3.4 aufgelistet.	51

Tabellenverzeichnis

2.1.	Auflistung der für die Arbeit verwendeten Daten und MC-Simulationen. Die Tabelle enthält außerdem die Gesamtanzahl der Ereignisse für die jeweiligen MC-Simulationen und Messungen (N_{tot}).	14
2.2.	Größen der Datensätze (initial und final) mit den entsprechenden Reduktionsfaktoren f_{R} und dem Gesamtreduktionsfaktor $f_{\text{R,tot}}$. Für den Mischkanal $2e2\mu$ bei der Messung wird der DoubleMuParked/Run2012[A,B,C,D] Datensatz verwendet. Über die Signal-MC erfolgt das <i>skimming</i> in allen Zerfallskanälen.	15
2.3.	Aufzählung aller Filter und Rekonstruktionsschritte in der im Kapitel erwähnten Reihenfolge. Nach jedem Schritt bis zum $m_{2\ell}$ -Ausschluss wird nach der Minimalanzahl der Leptonen im Ereignis geprüft und eine elektrisch neutrale Ladungskombination gesucht. Der relative Anteil der übrigen Ereignisse bezieht sich immer auf die Ereignisanzahl aus dem vorhergegangenen Filter- bzw. Rekonstruktionsschritt nach der Anwendung des genannten Schrittes.	19
2.4.	Für die Skalierung von MC-Simulationen auf die Messung verwendeten aus der Theorie bestimmten Wirkungsquerschnitte für die unterschiedlichen Zerfallskanäle [5, 57].	20
3.1.	Durch die Anpassung (Abbildung 3.1) bestimmten Parameter für die Untergrundparametrisierung mittels (3.1). Diese Menge an Parametern wird im folgenden als $\theta_{0\text{U}}$ bezeichnet.	24
3.2.	Ergebnis der globalen Skalierung der Unsicherheiten der Bins σ_i durch den Faktor $\sqrt{\chi_{\text{min}}^2}$ derart, dass $\chi_{n_{\text{dof}}}^2 = 1$ gilt. n entspricht dem größten Grad des Legendrepolynoms in der verwendeten Linearkombination.	25
3.3.	Durch die Anpassung (Abbildung 3.4) bestimmten Parameter für die Signalbreitenabschätzung mittels eines linearen Modells . (3.2) für unterschiedliche simulierte Signalverteilungen.	26
3.4.	Skalierungsfaktoren $\sqrt{\chi_{\text{min}}^2}$ der Unsicherheiten, für die Bedingung $\chi_{n_{\text{dof}}}^2 = 1$, welche für die Anpassung eines Signal-MC für eine Higgsmasse von $m_{\text{H}} = 125$ GeV notwendig ist. Die Anpassungen für die aufgeführten Verteilungen sind im Anhang D.5a, D.5b und D.5c hinterlegt.	30
3.5.	Durch die Anpassung (Abbildung 3.6) bestimmten Parameter für die Parametrisierung der Signal-MC mit einer Higgsmasse von $m_{\text{H}} = 125$ GeV mittels der DSCB-Verteilung (3.8). Diese Parametermenge wird im Weiteren als $\theta_{0\text{S}}$ bezeichnet.	30

Einleitung

Die physikalischen Praktika, die im Laufe des Physikstudiums absolviert werden, sollen durch unterschiedliche Experimente einen Einblick in die Bereiche der Physik geben und damit die Wahl eines Themenbereiches für die Spezialisierung erleichtern. Moderne Versuche zur Hochenergiephysik sind zunehmend international aufgestellt. Sie schaffen durch die öffentliche Bereitstellung der aufgezeichneten Daten die einzigartige Möglichkeit im physikalischen Praktikum moderne Fragestellungen zu behandeln. Diese Arbeit hat das Ziel, eines dieser Themen, die Entdeckung des Standardmodell Higgs-Bosons in Form einer Analyse vorzustellen, die in einem physikalischen Praktikum durchgeführt werden kann.

Das von Peter W. Higgs in 1964 postulierte Higgs-Boson [1, 2], als grundlegender Baustein im Standardmodell wurde 2012 am CERN entdeckt [3, 4]. In weiteren Messungen hat sich das anfängliche 'higgsartige' Boson als das Higgs-Boson des Standardmodells der Teilchenphysik bestätigt und führte zur Verleihung des Nobelpreises in der Physik für Peter W. Higgs und François Englert im Jahr 2013. Die Veröffentlichung der für die Entdeckung verwendeten Datensätze auf dem frei zugänglichen CERN Open Data Portal¹ eröffnet die Möglichkeit einer Durchführung einer ähnlichen Analyse, die zur Entdeckung führte. Der Fokus dieser Arbeit wird auf die Analyse des Zerfalles in vier Leptonen gelegt.

Aus den Studien zur grundlegenden Durchführbarkeit [5, 6] wird aufbauend eine Version des Versuches im Kapitel 4 vorgestellt, dessen Anwendung im physikalischen Praktikum oder dem praktischen Teil einer Veranstaltung zur experimentellen Teilchenphysik möglich ist. Bei der Durchführung wird die Verwendung der in der Teilchenphysik üblichen Analysesoftware `ROOT` und der dazugehörigen detektorspezifischen Software vermieden, um auch Versuchsteilnehmern, die bis dahin nur wenige Berührungspunkte mit der Teilchenphysik hatten, die Möglichkeit zu geben den Versuch ohne langwierige Vorbereitung durchzuführen. Hierzu werden die verwendeten Datensätze in das zugängliche CSV-Format konvertiert, welches von den Studenten mithilfe der weit verbreiteten Programmiersprache `PYTHON` analysiert wird.

Die Hauptaufgabe der Versuchsteilnehmer ist die Analyse des Massenspektrums der vier Leptonen, die im Kapitel 3 beschrieben wird und mit der Annahme startet, dass kein Higgs-Boson gemessen wurde. Im Laufe des Kapitels wird diese Annahme widerlegt, auf das Higgs-Boson des Standardmodells geschlossen und seine Masse bestimmt.

Vorbereitend findet im Kapitel 2 eine detaillierte Selektion der Kollisionsereignisse statt, die den gewünschten Zerfall in vier Leptonen enthalten, aus welchen die invariante Masse der vier Leptonen rekonstruiert wird. Die Versuchsteilnehmer lernen, mit den dazu notwendigen Grundlagen, die im Kapitel 1 besprochen werden, den Umgang mit gemessenen und simulierten Daten und deren statistische Analyse.

¹<http://opendata.cern.ch/>

1. Grundlagen

Die Teilchenphysik, besitzt wie andere Teilgebiete der Physik eigene Terminologie und Konventionen, die sich von anderen Gebieten der Physik unterscheiden. In diesem Kapitel wird neben den notwendigen physikalischen Größen und Konzepten auch in Kürze auf den Compact Muon Solenoid Detektor (CMS-Detektor) eingegangen, der zur Datenerfassung in dieser Arbeit benutzt wurde.

Das wichtigste Grundkonzept sind die natürlichen Einheiten, die die physikalischen Größen mithilfe der Naturkonstanten ausdrücken. Aus diesem Grund können physikalischen Konstanten wie die Lichtgeschwindigkeit oder das reduzierte Plancksches Wirkungsquantum auf eins gesetzt werden. Die resultierende Einheit für die Masse ist damit beispielsweise in Elektronenvolt durch das Quadrat der Lichtgeschwindigkeit ($\text{eV}c^{-2}$) gegeben. Das Ziel dieser Konvention ist die Vereinfachung der Darstellung von physikalischen Vorgängen.

1.1. Invariante Masse

Jedes sehr kurzlebige Teilchen kann aus seinen Zerfallsprodukten rekonstruiert werden. Dazu wird mit der Kenntnis der Zerfallsprodukte die invariante Masse (1.1) dieses Teilchen berechnet. Diese ist die quadratische Differenz aus den Energien und den Impulsen der im Zerfallsprozess gebildeten N Teilchen [7, 8].

$$m_{\text{inv}} = \sqrt{\left(\sum_{i=1}^N E_i\right)^2 - \left(\sum_{i=1}^N \vec{p}_i\right)^2}. \quad (1.1)$$

1.2. Pseudorapidität

Die Zerfallsprodukte zweier aufeinander fliegender und kollidierender Teilchen können einen Boost in Flugrichtung der kollidierenden Teilchen aufweisen. Die Geschwindigkeiten $\beta_i = |\frac{v_i}{c}|$ dieser Teilchen lassen sich jedoch nicht additiv kombinieren. Aus diesem Grund wird die Rapidität (1.2) dieser Teilchen betrachtet, welche wiederum additiv ist.

$$y = \frac{1}{2} \ln \left(\frac{1 + \beta}{1 - \beta} \right) = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (1.2)$$

p_z ist dabei die Impulskomponente des Teilchens entlang der Strahlachse. Aus der Rapidität lässt sich unter der Annahme, dass die Ruhemasse eines hochenergetischen Teilchen vernachlässigbar ist, die Pseudorapidität (1.3) bestimmen, die für relativistische Teilchen als ein Winkel zwischen der Flugbahn der kollidierenden Teilchen und einem Zerfallsprodukt interpretiert werden kann [7, 8].

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_z}{p - p_z} \right) \simeq -\ln \left(\tan \frac{\theta}{2} \right), \quad (1.3)$$

1.3. Luminosität

Die Luminosität \mathcal{L} ist eine Größe, die die Anzahl von Teilchenkollisionen pro Zeiteinheit – instantane Luminosität – oder in einem Zeitintervall beschreibt – integrierte Luminosität ($\mathcal{L}_{\text{int}} = \int dt \mathcal{L}$). Für einen Ringbeschleuniger ist diese durch (1.4) definiert.

$$\mathcal{L} = \frac{N_a N_b j f}{A}. \quad (1.4)$$

In die Luminosität fließt die Anzahl j der unterschiedlichen Pakete der Teilchensorten N_a und N_b ein, die sich mit einer Frequenz $f = \frac{v}{U}$ in einem Strahlenquerschnitt A treffen, wobei v die Geschwindigkeit dieser Pakete und U der Umfang des Ringbeschleunigers ist. Die Dimension der Luminosität ist das inverse Produkt aus Fläche und Zeit und besitzt die für die Teilchenphysik typische Einheit $(\text{b} \cdot \text{s})^{-1}$, wobei die Definition von 'barn' $1 \text{ b} = 10 \cdot 10^{-28} \text{ m}^2$ ist [7, 9].

Damit ist die Luminosität eine wichtige Kenngröße zur Beschreibung der Beschleunigungsleistung. Ereignisse, wie der Zerfall eines Higgs-Bosons in vier Leptonen sind äußerst selten, wodurch für die Untersuchung dieser Zerfälle eine hohe Rate an Kollisionen notwendig ist, die in hoher integrierter Luminosität resultieren. Aus diesem Grund findet eine schrittweise Erhöhung der integrierten Luminosität am CERN statt. So betrug diese im Jahr 2018 65 fb^{-1} und war damit um 16 fb^{-1} höher als die Luminosität aus dem Vorjahr [10]. Bei der Entdeckung des Higgs-Bosons betrug die integrierte Luminosität aus dem Jahr 2011 und dem Anfang von 2012 $10,4 \text{ fb}^{-1}$ [3].

1.4. Wirkungsquerschnitt

In der experimentellen Teilchenphysik wird das Wissen über die Eigenschaften, Wechselwirkungen oder Struktur der Teilchen durch Streuexperimente bestimmt. Ein Streuexperiment beschreibt im einfachsten Fall ein Teilchenstrahl, welcher auf ein fixes zu untersuchendes Objekt, das Target, trifft. Für die Anwendung in einem Ringbeschleuniger wird dazu eine Transformation in das Ruhesystem eines der Kollisionsobjekte durchgeführt. Die Wahrscheinlichkeit einer Interaktion zwischen den Stoßpartnern entspricht dem Wirkungsquerschnitt σ , welcher veranschaulicht auch als Trefferfläche angesehen werden kann. Die Wechselwirkungen geschehen jedoch nicht mit den 'Flächen' der untersuchten Objekte, sondern ihren Potenzialen. Die Einheit von σ ist eine Fläche und wird in der Teilchenphysik in 'barn' angegeben [7].

Bei einer Reaktionsrate \dot{N} zwischen dem Fluss der eintreffenden Teilchen (Φ_a) und der Gesamtanzahl an Targetteilchen N_b in einem ausgewählten Querschnitt ist der Wirkungsquerschnitt durch (1.5) beschrieben [7, 11].

$$\sigma = \frac{\dot{N}}{\Phi_a N_b} = \frac{\dot{N}}{\mathcal{L}}. \quad (1.5)$$

Das Produkt $\Phi_a N_b$ kann ähnlich wie in (1.4) als die Luminosität aufgefasst werden. Für eine unvollständige Raumabdeckung muss zudem eine Korrektur durchgeführt werden, die der Integration des differentiellen Wirkungsquerschnittes $\frac{d\sigma}{d\Omega}$ über den vom Detektor abgedeckten Winkelbereich entspricht. Bei einem nicht idealen Detektor muss zudem die Luminosität durch die Akzeptanz A und Effizienz ϵ des Detektors zu $\mathcal{L}' = \mathcal{L} A \epsilon$ korrigiert werden. Damit kann mit dem Wissen über die Gesamtzahl an Ereignissen (N_{tot}) und einem aus der Theorie bekannten Wirkungsquerschnitt σ_{theo} auf die integrierte Luminosität von einer simulierten Messung, einer Monte Carlo Simulation (MC-Simulation), geschlossen werden:

$$\mathcal{L}'_{\text{int}} = \frac{N_{\text{tot}}}{\sigma_{\text{theo}}}. \quad (1.6)$$

Diese Form wird im Kapitel 2 für die Skalierung von MC-Simulationen an die vorhandene Messung verwendet.

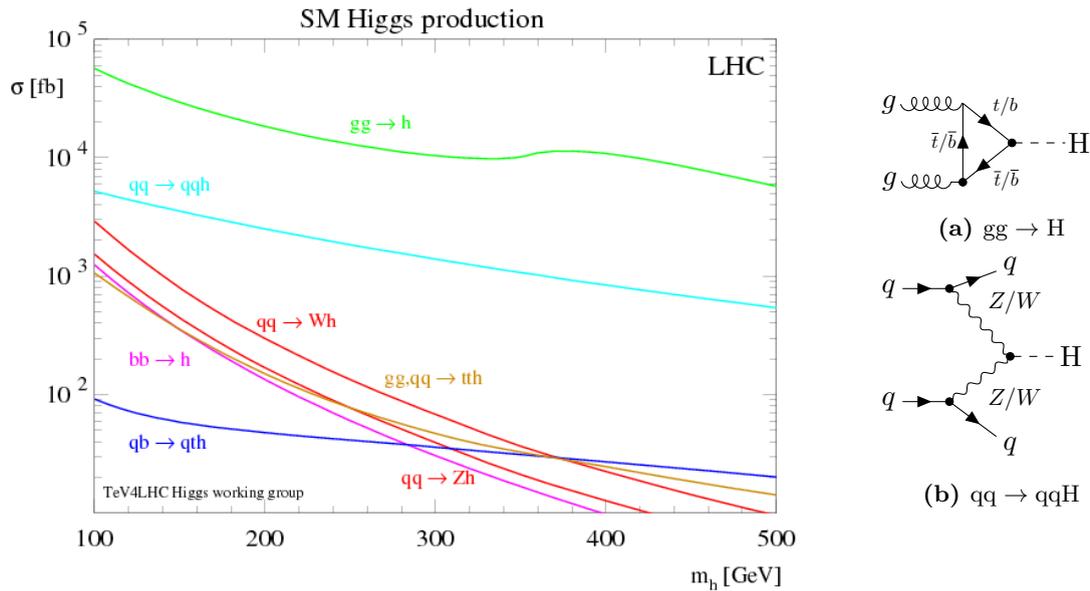


Abbildung 1.1.: Higgs-Boson Produktionskanäle aus pp -Kollisionen am LHC [15]. Die Gluonfusion 1.1a und die Vektorbosonfusion 1.1b sind hierbei die dominanten Produktionskanäle.

1.5. Higgs-Boson des Standardmodells

Nach dem Standardmodell müssen die Eichbosonen, die als ein zusätzliches Feld zu den wechselwirkungsfreien Fermionen als Lösung der Dirac-Gleichung herauskommen, eine verschwindende Ruhemasse besitzen. Für die W - und Z -Bosonen der schwachen Wechselwirkung kann die Ruhemasse jedoch nicht verschwinden, da dadurch die schwache Wechselwirkung eine unendliche Reichweite hätten [7, 11]. Die endliche Masse ist seit der Entdeckung der W - und Z -Bosonen am CERN [12] bestätigt.

Um dieses Problem zu beheben wird ein zusätzliches Feld, das Higgs-Feld, eingeführt. Durch den Higgs-Mechanismus haben auch alle anderen elementaren Teilchen des Standardmodells, bis auf die Gluonen und Photonen, die nicht an das Higgs-Feld koppeln, ihre Masse. Das Higgs-Boson des Standardmodells ist ein elektrisch neutrales Teilchen mit Spin 0 und besitzt aus der Theorie bekannte Verzweigungsverhältnisse, die in Abbildung 1.2 gezeigt werden [1, 2, 13, 14].

Einer Einschränkung der Masse auf kleiner als 1 TeV folgte der experimentelle Nachweis, der von Peter Higgs im Jahr 1964 [1] postulierten Higgs-Bosons, erst über 40 Jahre später. Dies geschah an zwei unabhängigen Experimenten am CERN. Die von dem 'Compact-Muon-Solenoid-Experiment' (CMS) bestimmte Masse betrug $m_H = (125,3 \pm 0,4 \pm 0,5)$ GeV und von 'A Toroidal LHC Apparatus' (ATLAS) $m_H = (126,0 \pm 0,4 \pm 0,4)$ GeV [3, 16]. Weitere Messungen grenzen die Masse des Standardmodell Higgs-Bosons nachfolgend auf $m_H = (125,18 \pm 0,16)$ GeV ein [17].

Die Produktion des Standardmodell Higgs-Boson am LHC ist durch die Nutzung von pp -Kollisionen hauptsächlich durch Gluonfusion und Vektorbosonfusion dominiert (Abbildung 1.1). Das Ziel der Higgs-Produktion ist eine Paarvernichtung von möglichst schweren Elementarteilchen, da mit steigender Masse die Wahrscheinlichkeit der Produktion steigt. Ein Ansatz hierfür ist die indirekte Nutzung von top bzw. bottom Quarks. Sie sind nicht direkt detektierbar, können jedoch bei einer Proton-Proton-Kollision durch die Gluonfusion, gebildet werden. Bei der Gluonfusion (Abbildung 1.1a) geht damit aus einer Schleife von top bzw. bottom Quarks ein Higgs-Boson hervor. Der zweitwichtigste Produktionskanal,

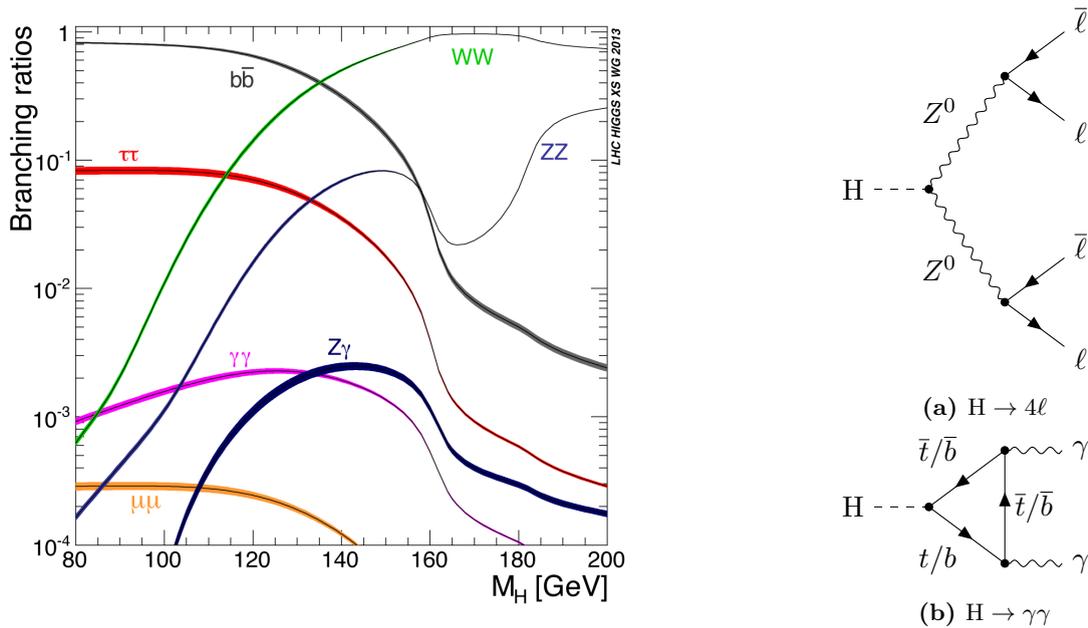


Abbildung 1.2.: Verzweigungsverhältnisse des Higgs-Bosons mit den geschätzten Unsicherheiten [18]. Der Zerfallskanal in ein Photonenpaar (1.2b) liefert die wichtige Information, dass das ursprüngliche Teilchen einen Spin 0 besitzt. Der Zerfallskanal nach vier Leptonen (1.2a) wird im Rahmen der Arbeit genauer betrachtet.

die Vektorbosonfusion (Abbildung 1.1b), erzeugt ein Higgs-Boson aus den von zwei Quarks abgestrahlten W^\pm - oder Z^0 -Boson Paaren. Weitere Kanäle, bis auf die Abstrahlung von einem W - oder Z -Boson, tragen dagegen kaum zu der Higgs-Boson Produktion bei, wie in Abbildung 1.1 erkenntlich ist.

Die möglichen Zerfallskanäle des Standardmodell Higgs-Bosons sind in Abbildung 1.2 in Abhängigkeit ihrer auftretenden Häufigkeit dargestellt. Für diese Arbeit wird der Zerfallskanal in zwei Z -Bosonen und anschließend in vier Leptonen betrachtet (Abbildung 1.2a). Die zu untersuchenden Leptonen sind die e^+e^- - bzw. $\mu^+\mu^-$ -Paare, die aus den elektrisch neutralen Z -Bosonen zerfallen und im Detektor, im Gegensatz zu den kurzlebigen Tau-Leptonen, gut nachzuweisen sind. Die wahrscheinlichen Zerfallskanäle wie der Zerfall in ein bottom und anti bottom Quark oder in ein W -Bosonen Paar zeigen zu dem hohen Untergrundanteil gegenüber dem Signal auch die Problematik, dass die Analyse in diesen Kanälen im Rahmen eines Praktikumsversuches aufwendig ist. Dagegen kann der hohe Untergrundanteil in dem vier Leptonen Kanal ($H \rightarrow 4\ell$), im Praktikum durch eingeführte Selektion derart reduziert werden, dass das Verhältnis aus Signal und Untergrund groß genug für aussagekräftige statistische Analysen ist.

Für eine tatsächliche Bestätigung des Higgs-Bosons des Standardmodells werden zusätzlich notwendige Informationen aus anderen Zerfallskanälen genommen, sowie die Zerfallsverhältnisse in die unterschiedlichen Zerfallskanäle untersucht, da diese aus der Theorie bekannt sind [3]. So kann aus dem Zerfall in ein Photonenpaar (Abbildung 1.2b) auf den Spin des Higgs-Boson geschlossen werden. Der $\gamma\gamma$ -Zerfallskanal ist, trotz des geringen Wirkungsquerschnittes, aufgrund der guten Photonendetektion im Detektor eindeutig nachweisbar. Im Rahmen dieser Arbeit wird dieser Zerfall nicht betrachtet und der Fokus auf den vier Leptonen Zerfall gelegt.

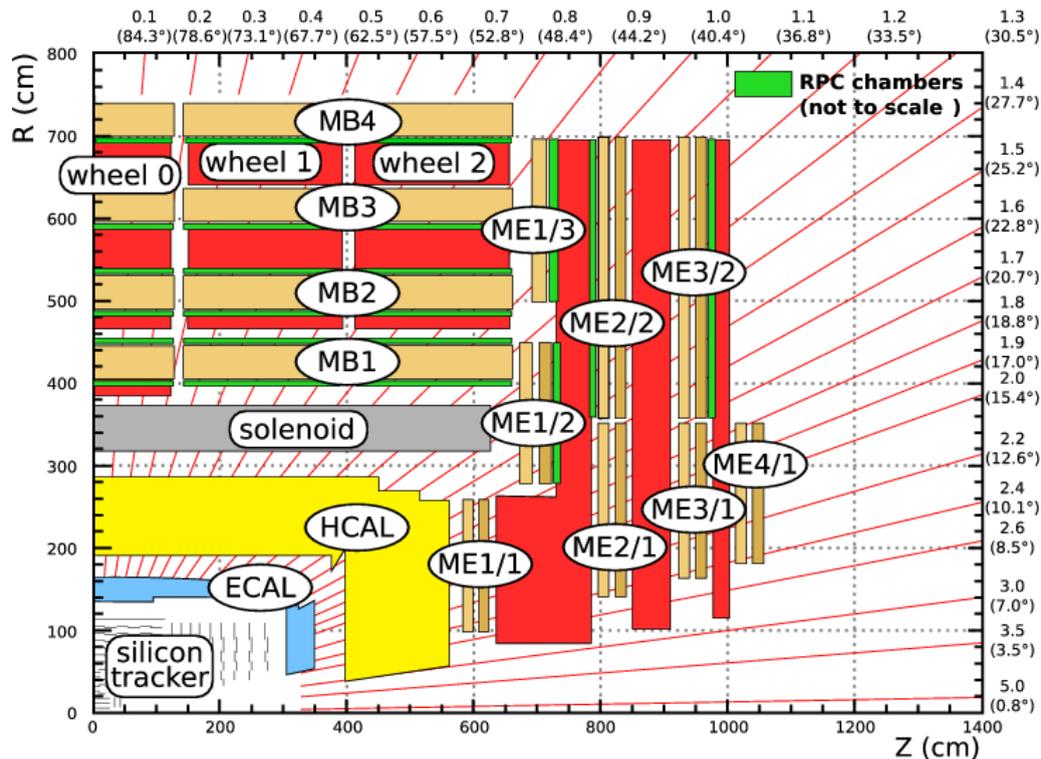


Abbildung 1.3.: Ein Längsschnitt des Detektors, parametrisiert durch die Pseudorapidität, dem Radius des Detektors und der Ausdehnung entlang der Strahlachse [20].

1.6. Der CMS-Detektor

Die Daten für diesen Praktikumsversuch stammen aus dem CMS Detektor (Compact Muon Solenoid), welcher zusammen mit weiteren Detektoren am LHC-Beschleuniger am CERN betrieben wird.

Das für den Detektor gewählte rechtshändige Koordinatensystem hat seinen Ursprung im Mittelpunkt des Detektors, in welchem idealerweise die Kollision stattfindet. Die z-Achse ist entlang der Strahlrichtung gewählt. Die Ausrichtung der x-Achse zeigt radial zum Mittelpunkt des LHC-Beschleunigers und die y-Achse senkrecht nach oben. Es findet eine Parametrisierung durch den Azimutalwinkel ϕ in der x-y-Ebene und einem Polarwinkel θ in der x-z-Ebene statt. Für die Polarwinkelangaben wird die Pseudorapidität (1.3) verwendet. Sie entspricht einem Winkel, den das Teilchen relativ zu der Strahlachse besitzt. Die verschiedenen Detektorschichten lassen sich durch η parametrisieren. Das Wissen der Pseudorapidität kann damit zu Ausschluss von Teilchen genutzt werden, welche beispielsweise an Grenzbereichen der einzelnen Detektorkomponenten, oder fälschlicherweise außerhalb dieser, registriert wurden. Ein Längsschnitt des CMS-Detektors ist in Abbildung 1.3 beispielhaft dargestellt. Genauere Informationen zum datierten Aufbau und der speziell verwendeten Software können dem 'Technical Design Report' des CMS-Detektors [19] entnommen werden.

Der CMS-Detektor ist ein Vielzweckdetektor und ist zwiebelschalenartig aufgebaut. Ausgehend von dem Detektorzentrum durchlaufen die durch Kollision entstandenen Teilchen zunächst den Silizium-Spuredetektor (Silizium-Tracker), welcher in dem Zentrallbereich, dem Barrel, in einen äußeren und inneren Spuredetektor unterteilt wird. Mit den zusätzlichen Spuredetektoren am Anfang und Ende des Barrels, den Endcaps, wird eine Raumwinkelabdeckung von $|\eta| < 2,4$ erreicht. Mithilfe des Spuredetektors kann so der Kollisionspunkt und die anfänglichen Bahnen der Kollisionsprodukte rekonstruiert werden. Um den Spuredetektor befindet sich der elektromagnetischer Kalorimeter (ECAL),

welcher im CMS-Detektor durch szintillierende Bleikristalle als Absorber- und Silizium-Avalanche-Photodiode als Auslesematerial realisiert ist. Der ECAL detektiert Teilchen, die elektromagnetisch wechselwirken und damit ihre Energie im ECAL deponieren. Der Aufbau besteht wie beim Spurendetektor aus einem Barrel und zwei Endcaps. Der maximale effektive Raumwinkel, den der ECAL abdecken kann ist $0 \leq |\eta| < 2.5$. Der Raumwinkel zwischen dem Barrel und den Endcaps ($1,479 \leq |\eta| \leq 1,653$) wird nicht vollständig abgedeckt. Weiter folgt das hadronische Kalorimeter (HCAL), der das ECAL umschließt. Hier werden die teilweise deponierte oder vollständig absorbierte Energie der Hadronen detektiert. Idealerweise werden bis dahin alle Teilchen außer den Myonen und den Neutrinos in dem ECAL und HCAL absorbiert. Für die Krümmung der Teilchenbahnen wird ein supraleitender Solenoid verwendet, der den HCAL einschließt und ein maximales Magnetfeld von 4 T besitzt. Außerhalb des Magnetfeldes befindet sich das Myonen-System des Detektors, welches sich aus Driftröhren im mittleren Bereich und Vieldrahtkammern in den Endbereichen zusammensetzt. Zwischen den Driftröhren und den Vieldrahtkammern befinden sich Widerstandsplattenkammer die eine gute zeitliche Auflösung im Gegensatz zu den Driftröhren und Vieldrahtkammern liefern. Die effektive Raumwinkelabdeckung des Myonen-Systems beträgt $0 \leq |\eta| < 2.6$.

1.7. Invariantes Massenspektrum

Aus der Betrachtung der invarianten Massenspektren kann auf bereits bekannte oder neuartige Teilchen geschlossen werden, welche sich in Form einer Resonanz in einem Spektrum zeigen. Die Suche nach neuen Teilchen erfordert eine genaue Untersuchung des invarianten Massenspektrums des zu betrachtenden Prozesses. Hierzu wird aus einem Kollisionsereignis oder kurz Ereignis durch (1.1) die notwendige invariante Masse bestimmt. Dieser wird im Falle großer Ereignisanzahl in Form eines Histogramms dargestellt, in welchem die Anzahl an beobachteten Ereignissen im bestimmtem Massenbereich festgehalten werden (Abbildung 1.4).

In dem untersuchten invarianten Massenspektrum der vier Leptonen finden sich zu der Signalverteilung eines Standardmodell Higgs-Bosons der Masse $m_H = 126 \text{ GeV}$ (rot) noch weitere Verteilungen von möglichen Untergrundprozessen, die auch ohne die Signalverteilung erwartet werden. Der dominante Teil des Untergrundes ist die Quark Annihilation oder die Gluonfusion. Der entstandene Untergrund kann in einen reduzierbaren (grün) und den nicht reduzierbaren (blau) Untergrund unterteilt werden.

Untergrundprozesse, deren Reduktion möglich ist beinhalten Ereignisse mit nur einem Z -Boson und einer zusätzlichen Komponente und werden mit $Z + X$ zusammengefasst. Diese Prozesse können durch die Betrachtung spezieller Kontrollregionen in den Datensätzen reduziert oder abgeschätzt werden. Im weiter betrachteten Bereich ist der Anteil dieser Untergründe jedoch gegenüber dem nicht reduzierbarem Untergrund klein und wird im Rahmen der Arbeit vernachlässigt [3, 21].

Die nicht reduzierbaren Untergründe beinhalten dagegen die Zerfälle in vier Leptonen aus zwei Z -Bosonen oder einem Z -Boson und einem Photon. Diese stammen nicht aus einem Higgs-Boson, sondern beispielsweise aus der genannten Quark Annihilation. Diese Untergrundprozesse haben eine ähnliche Topologie wie das Signal, sodass keine Trennung möglich ist. Deshalb ist in der Abbildung 1.4 ein ausgeprägter Peak zu sehen, der bei der Masse des Z -Bosons, $91,1876 \text{ GeV}$ [17], liegt. Für die Abschätzung des Untergrundes wird eine MC-Simulation von diesen Zerfällen im Detektor durchgeführt (Untergrund-MC). Zusätzlich findet eine Simulation von Signalverteilungen für unterschiedliche Massen statt. Die Aufgabe besteht darin mit statistischen Werkzeugen eine angenommene Verteilung des Signals durch die Messung zu bestätigen.

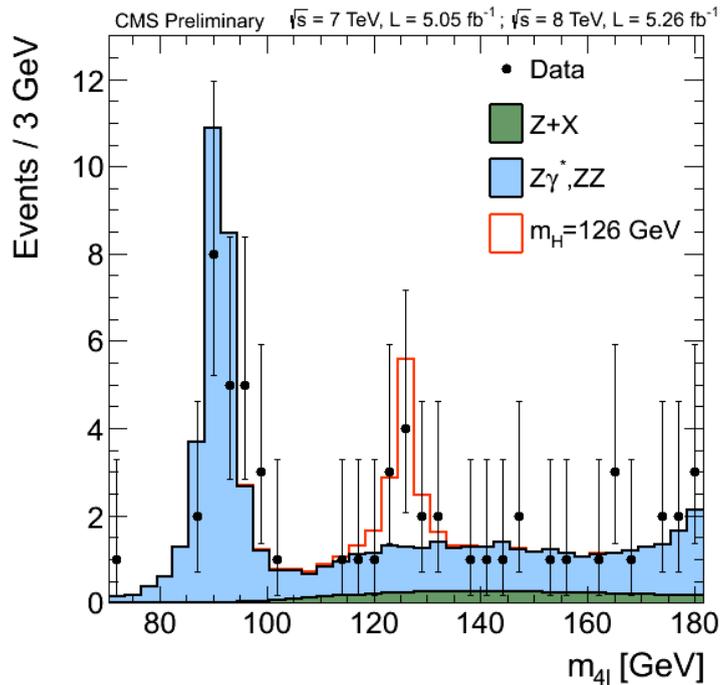


Abbildung 1.4.: Vier Leptonen invariantes Massenspektrum in einem Massenbereich zwischen 70 GeV und 181 GeV. Eingezeichnet sind neben den erwarteten Verteilungen für die Untergründe und einer Signalverteilung für ein Standardmodell Higgs-Boson mit einer Masse von 125 GeV und die Messung mit den Unsicherheiten, die am CMS Detektor aufgenommen wurde [3].

1.8. Parameterschätzung

Die Grundlagen für die statistische Betrachtung im Kapitel 3 ist die Methode der kleinsten Quadrate und die Methode von maximalem Likelihood. Die Anwendung dieser Methoden ist im Rahmen dieser Arbeit auf Parameterschätzung, Bestimmung der lokalen Signifikanz und die Auswahl zwischen möglichen Hypothesen beschränkt, ist aber als Grundkonzept für weitere statistische Betrachtungen unerlässlich.

1.8.1. Methode der kleinsten Quadrate

Diese Methode eignet sich zur Schätzung der Parameter und kann zur Bewertung der Anpassungsgüte eingesetzt werden. Zur Fluktuationsreduktion, ausgelöst hauptsächlich durch die Histogrammierung wird eine Wahrscheinlichkeitsdichtefunktion, eine pdf^1 , bestimmt, die eine gegebene Verteilung, dargestellt durch ein Histogramm, in einem bestimmten Bereich beschreibt. Für die Parametrisierung einer Verteilung ist das Wissen über die Unsicherheit auf die jeweiligen Histogrammeinträge nötig.

Diese sind für große Anzahl an Ereignissen in guter Näherung gaußförmig, können aber auch durch entsprechende Skalierung des Histogramms mit dem gleichen Faktor vergrößert bzw. verkleinert werden. Damit kann aus dem Wissen der bekannten Unsicherheiten σ_i auf jede Binfläche $y_{\text{obs},i}$ durch die Methode der kleinsten Quadrate die optimalen Parameter θ einer Hypothese $f(x|\theta)$ bestimmt werden, die die Messung $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ beschreiben

¹probability density function

soll. n_{tot} ist hierbei die Gesamtzahl an beobachteten Ereignissen. Die Integration in (1.8) erfolgt über den Bereich des betrachteten Bins.

$$\chi^2(\theta) = \sum_{i=1}^N \frac{(y_{\text{obs},i} - \tilde{f}_i(\theta))^2}{\sigma_i^2}, \quad (1.7)$$

$$\tilde{f}_i(\theta) = n_{\text{tot}} \int_{\text{bin}_i} f(x|\theta) dx. \quad (1.8)$$

$\chi^2(\theta)$ ist das Maß für die Übereinstimmung der Hypothese mit den vorhandenen Messdaten. Diese Größe nimmt den minimalen Wert χ_{min}^2 für die optimale Schätzung der Parameter θ_0 an und kann als Aussage über die Anpassungsgüte genutzt werden. Hierzu wird das Verhältnis $\chi_{\text{min}}^2/n_{\text{dof}}$, im Weiteren als $\chi_{n_{\text{dof}}}^2$ bezeichnet betrachtet, wobei n_{dof} der Anzahl an Freiheitsgrade² entspricht. χ_{min}^2 folgt der χ^2 Verteilung mit n_{dof} als Anzahl der Freiheitsgrade und hat damit als Erwartungswert n_{dof} . Damit ist $\chi_{n_{\text{dof}}}^2 \simeq 1$ im Falle einer idealen Anpassung, $\ll 1$ sobald eine Überparametrisierung stattfindet und $\gg 1$ wenn die entsprechende Hypothese nicht der beobachteten Messung entspricht. [22, 23]

1.8.2. Likelihood-Methode

Für eine Betrachtung von einer geringen Anzahl an Ereignissen, wie sie im Weiteren stattfindet, ist die Verwendung einer Abwandlung des Maximum-Likelihood Prinzips für statistische Aussagen nötig, da die Methode der kleinsten Quadrate nur für große Anzahl an Ereignissen gültig ist. Für poissonartige Größen versagt sie, konvergiert aber für große Zahlen gegen den mit dem Faktor zwei skalierte negative Logarithmus der Likelihood. Weil eine Histogrammierung von wenigen Daten diese durch Effekte des Binnings verfälscht werden die Daten ungebinnt betrachtet.

Die Likelihood ist ein Produkt aus bedingten Wahrscheinlichkeiten die Messgröße x_i aus der Stichprobe $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ zu messen unter der Hypothese, dass die Stichprobe der Wahrscheinlichkeitsverteilung f mit den Parametern θ folgt. f sollte hierbei auf dem betrachteten Intervall Ω auf eins normiert sein.

$$\mathcal{L}(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta). \quad (1.9)$$

Aus der Likelihood (1.9) können durch Maximierung die optimalen Parameter θ_0 für die angenommene Wahrscheinlichkeitsverteilung entnommen werden. Die Maximierung der Likelihood entspricht dabei der Annahme, dass die Gesamtwahrscheinlichkeit die gemessene Stichprobe unter der Annahme von f zu erhalten steigt und für optimale Wahl der Parameter maximal wird.

Sollte die Größe der Stichprobe \mathbf{x} selbst eine poissonverteilte Größe sein und damit auch ein Teil des Ergebnisses der Messung darstellen, so kann die Bedingung der Normierung der Wahrscheinlichkeitsverteilung f derart gelockert werden, dass die erwartete Stichprobengröße in Form eines poissonartigen Skalierungsfaktors mitberücksichtigt wird (1.10) [22, 24]. Damit ist im Folgenden bei der Erwähnung der Likelihood immer diese Variante

²degree of freedom

der Likelihood-Methode – die *extended-Likelihood* – gemeint, da die Bedingung für die Verwendung dieser Likelihood-Methode im Weiteren immer erfüllt ist.

$$\mathcal{L}(\mathbf{x}|\theta) = \frac{\nu(\theta)}{n!} e^{-\nu(\theta)} \prod_{i=1}^n f(x_i|\theta), \quad (1.10)$$

$$\nu(\theta) = \int_{\Omega} x f(x|\theta) dx, \quad (1.11)$$

$$-2 \ln \mathcal{L}(\mathbf{x}|\theta) = 2\nu(\theta) - 2 \sum_{i=1}^n \ln(\nu(\theta) f(x_i|\theta)). \quad (1.12)$$

Die Suche nach den optimalen Parametern erfolgt im Weiteren numerisch. Deshalb wird aus Gründen der numerischen Instabilität von Gleichung (1.10) der negative Logarithmus gebildet (Gleichung (1.12)) und alle konstanten Terme fallen gelassen.

Nach dem zentralen Grenzwertsatz kann die Poissonverteilung durch die Normalverteilung beschrieben werden ($\nu \gg 1$). Für den Fall, dass $f(x_i|\theta)$ einer Normalverteilung mit dem Erwartungswert $\nu(\theta)$ und der Standardabweichung $\sigma = \sqrt{\nu(\theta)}$ folgt, wird eine Übereinstimmung zwischen der Methode der kleinsten Quadrate (1.7) und dem negativen Logarithmus der Gleichung (1.9) erreicht, indem ein Faktor zwei eingeführt und in Gleichung (1.12) verwendet wird.

Damit entspricht bei beiden Methoden der Fehler auf den optimalen Parameter den Werten der Parameter, bei denen $\mathcal{L}(\theta)$ und $\chi^2(\theta)$ die Werte $-2 \ln(\mathcal{L}(\theta_0)) + 1$ beziehungsweise $\chi^2(\theta_0) + 1$ annimmt.

2. Rekonstruktion des Massenspektrums der vier Lepton

Das Ziel dieses Kapitels ist es, ausgehend von den CMS veröffentlichten Daten und MC-Simulationen, das invariante Massenspektrum der vier Leptonen ($m_{4\ell}$ -Spektrum) für den Zerfallskanal $H \rightarrow ZZ \rightarrow 4\ell$ zu rekonstruieren. Hierzu wird, nach einer Konvertierung der Daten aus dem CMS spezifischen Datenformat, durch Auswahlkriterien (Cuts) eine Wahl von Leptonen getroffen, aus denen anschließend das invariante Massenspektrum der vier Leptonen ($m_{4\ell}$ -Spektrum) rekonstruiert wird. Die für diese Analyse betrachteten Zerfallskanäle beinhalten die Datensätze für den Zerfall in vier Myonen (4μ - *channel*), vier Elektronen ($4e$ - *channel*), sowie den Zerfall in zwei Myonen und zwei Elektronen ($2e2\mu$ - *channel*).

Für die nachfolgende statistische Analyse wird ein Massenbereich zwischen 106 GeV und 151 GeV gewählt. Eine direkte Suche am LEP Beschleuniger, bis 2003, ergab, dass eine Higgsmasse kleiner als 114,4 GeV mit einem Konfidenzlevel von 95 % ausgeschlossen werden kann [25]. Eine weitere, bis 2010 anhaltende, ebenfalls am LEP Beschleuniger vorgenommene, Messung von elektroschwachen Ereignissen schränkt die gesuchte Higgsmasse mit einem 95 % Konfidenzintervall auf einen Wert kleiner als 152 GeV ein [26]. Als Kontrolle soll zudem die Übereinstimmung von Untergrund MC mit den gemessenen Daten durchgeführt werden. Hierzu wird ein Massenbereich zwischen 70 GeV und 181 GeV gewählt und nach der durchgeführten Kontrolle auf den oben genannten Massenbereich eingeschränkt.

2.1. Konvertierung von .AOD/.AODSIM nach .csv

Die für diese Analyse verwendeten Datensätze, sind seitens CERN auf dem CERN Open Data Portal¹ unter Creative Commons CC0 Lizenz veröffentlicht und in der Tabelle 2.1 aufgelistet. Sie umfassen die im Jahr 2012 von CMS aufgezeichneten Ereignisse, bzw. die für dieses Jahr verwendeten MC-Simulationen und liegen im AOD (Analysis Object Data) und AODSIM (Simulated Analysis Object Data) Format vor. Die Verarbeitung dieser Datensätze geschieht durch die Kombination von ROOT [27] und der CMS eigenen Software-Umgebung CMSSW [28]. Um den roten Faden nicht zu verlieren, werden die Daten, die vom CMS 2011 erhoben wurden nicht berücksichtigt, da die verwendete Schwerpunktsenergie sowie viele der im Weiteren eingeführten Werte für die Cuts sich von dem im Jahr 2012 erhobenem Datensatz unterscheiden.

Die Analyse der Datensätze im AOD bzw. AODSIM Format ist aus Gründen der Komplexität des Datenformats und der dafür notwendigen Software im Rahmen eines Praktikumsversuches nicht praktikabel. Aus diesem Grund werden die Daten in ein zugänglicheres Format, in diesem Fall *comma separated list* (csv), konvertiert. Dadurch können Studenten, die im Laufe ihres Studiums und begleitender Lehrveranstaltung angeeignete Programmiersprache PYTHON, zusammen mit Standardbibliotheken wie NUMPY, SCIPY, PANDAS

¹<http://opendata.cern.ch/>

Tabelle 2.1.: Auflistung der für die Arbeit verwendeten Daten und MC-Simulationen. Die Tabelle enthält außerdem die Gesamtanzahl der Ereignisse für die jeweiligen MC-Simulationen und Messungen (N_{tot}).

Beschreibung	Name der Messung	N_{tot}	
4 μ -Kanal	DoubleMuParked/Run2012[A,B,C,D]	110567953	[29–32]
4e-Kanal	DoubleElectron/Run2012[A,B,C,D]	104906885	[33–36]

Beschreibung	Name der MC-Simulation	N_{tot}	
Untergrund des 4 μ -Kanals	ZZTo4mu_8TeV-powheg-pythia6	1499064	[37]
Untergrund des 4e-Kanals	ZZTo4e_8TeV-powheg-pythia6	1499093	[38]
Untergrund des 2e2 μ -Kanals	ZZTo2e2mu_8TeV-powheg-pythia6	1497445	[39]
Signal für die Higgs-masse $m_{\text{H}} = 125$ GeV	SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUGenV3-pythia6	299973	[40]
Signal weiterer Higgs-massen	SMHiggsToZZTo4L_M-[115, 120, 122, 124, 128, 130, 135, 140, 145, 150]_8TeV-powheg15-JHUGenV3-pythia6	≈ 300000	[41–50]

oder `MATPLOTLIB`, sowie öffentlich zugänglichen Softwareumgebungen wie `KAFE2` [51], zur Lösung des Versuchs nutzen. Das `csv` Format ist für das Speichern von komplexen Datenstrukturen und großen Datenmengen nicht optimal. Daher wird im Rahmen der Konvertierung bereits eine Vorauswahl an Ereignissen getroffen, von welchen anschließend nur Größen gespeichert werden, die im Verlauf in der Analyse verwendet werden sollen. Hierfür wird eine vorhandene Vorlage [5], die den Spezialfall von exakt vier Leptonen behandelt, auf beliebig viele Leptonen erweitert. Für diese Arbeit wurde die maximale Anzahl an Leptonen auf acht festgelegt. Außerdem erfolgt eine absteigende Sortierung der Leptonen nach ihren Transversalimpulsen innerhalb eines Ereignisses um Leptonen, die sehr wahrscheinlich Teil des Signalprozesses sind, durch die begrenzte Anzahl nicht auszuschließen. Nach diesem ersten Selektionsschritt, welcher auch als *skimming* bezeichnet wird, werden nur Ereignisse weiter betrachtet, die bestimmte Bedingungen erfüllen:

- Das Ereignis wird als validiert eingestuft. Damit wird sichergestellt, dass zur Messzeit des Ereignisses der Detektor fehlerfrei lief. Diese Bedingung muss von den MC-Simulationen nicht erfüllt werden. Für die Messung geschieht dies durch Abgleichen mit der im Nachhinein von CMS erstellten Validierungsliste, die im `JSON` Format für jedes Jahr vorliegt [52].
- Ereignisse müssen einen validen Kollisionspunkt (`primvtxHandle`) besitzen. Dies stellt sicher, dass Ereignisse welche mehrere Kollisionspunkte enthalten oder Ereignisse mit einem fehlerhaft rekonstruiertem Kollisionspunkt aussortiert werden.
- Im Ereignis müssen vier oder mehr Leptonen enthalten sein.

Außerdem wird im Rahmen vom *skimming* eine Klassifizierung der Leptonen durchgeführt. Die Klassifizierung der Myonen erfolgt ausgehend davon, ob sie nur im Spuredetektor (**T**) oder noch zusätzlich im Myonen System des Detektors (**G**), also global, detektiert wurden. Ein Myon, dem kein solcher Typ zugewiesen werden kann² erhält die Zuweisung **S**. Die Elektronen werden mit **T** gekennzeichnet, wenn diese im Spuredetektor und in einem der

²Dies erfolgt beispielsweise bei Myonen aus der kosmischen Strahlung.

elektromagnetischen Kalorimeter (ECAL Barrel oder ECAL Endcaps) detektiert werden. Für Elektronen, die dem Kollisionspunkt des Ereignisses nicht zugeordnet werden können, erfolgt ebenfalls wie bei den Myonen die Zuweisung `S`. Abschließend wird versucht eine elektrisch neutrale Kombination aus beliebigen vier der so erhaltenen Leptonen in einem Ereignis zu erhalten. Ist dies möglich, so wird das Ereignis behalten.

In den `.csv` Dateien werden die Zuordnungen zu Myonen bzw. Elektronen durch den entsprechenden Präfix (`myon_` bzw. `electron_`) realisiert. Neben der Laufnummer (`run`), der Eventnummer (`eventId`) und der oben genannten Zuweisung (`type`) werden die Energie (`energy`), alle Impulse (`px`, `py`, `pz`), die Ladung (`charge`) sowie weitere Parameter zur besseren Eingrenzung des Kollisionpunktes (`relPFiso`, `dxy`, `dz`, `SIP3d`) festgehalten. Für Elektronen wird zudem die Anzahl an (`missing_hits`) herausgeschrieben. Diese Variablen werden in der `.csv` Datei durch `'` getrennt. Innerhalb einer Rubrik werden die Parameter der einzelnen Leptonen, sortiert nach ihrem absteigenden Transversalimpuls, nochmals durch `'` separiert.

```
run;eventId;Myon_type;energy;...
193556;33426889;G,G,T,G,T;84.9839,83.2791,5.39887,5.85744,3.55763;...
193556;32666985;G,T,G,G;7.49254,6.97479,6.80645,4.94126;...
:
```

Die dadurch entstandene Form hat die Möglichkeit eine individuelle Teilchenzahl pro Ereignis festzuhalten ohne die Einführung von Platzhaltern. Mit dieser Auswahl an Ereignissen und Variablen wird die belegte Datenmenge um einen Faktor von 9734 reduziert. Die pro Datensatz erreichten Reduktionsfaktoren sind in der Tabelle 2.2 festgehalten.

Nach dem *skimming* ist der Datensatz insgesamt 6,4 GB groß und kann im zeitlichen Rahmen des Praktikums weiterverarbeitet werden. Im Praktikum werden im Gegensatz zur im Kapitel 3 vorgestellte Analyse nur das Signal-MC der Higgsmasse $m_H = 125$ GeV verwendet.

2.2. Filter

Auf die nach dem *skimming* erhaltenen Daten werden in einem weiteren, dynamischen Selektionsschritt spezifische Filter angewendet. Das Ziel der Anwendung der Cuts ist neben der Beschränkung auf Eigenschaften des Signalprozesses in den Ereignissen, wie beispielsweise die relative Isolation der vier Leptonen, die Beseitigung von fehlerhaft detektierten Teilchen in den jeweiligen Ereignissen. Um die Sensitivität zu erhöhen, werden Cuts auf verschiedene kinematische Größen angewendet. Während das *skimming* außerhalb des Praktikumsversuches stattfindet und nur ggf. vom Tutor durchgeführt wird, ist dieser Selektionsschritt der Ergebnisse mithilfe von Filtern und die nachfolgenden Abschnitte in einem variablen Umfang als Teil des Praktikumsversuches angedacht (Kapitel 4).

Tabelle 2.2.: Größen der Datensätze (initial und final) mit den entsprechenden Reduktionsfaktoren f_R und dem Gesamtreduktionsfaktor $f_{R_{tot}}$. Für den Mischkanal $2e2\mu$ bei der Messung wird der DoubleMuParked/Run2012[A,B,C,D] Datensatz verwendet. Über die Signal-MC erfolgt das *skimming* in allen Zerfallskanälen.

	initial in TB	final in GB	f_R	$f_{R_{tot}}$
Messung	60,2	5,96	10100	
Untergrund-MC	1,2	0,38	31507	9734
Signal-MC ($m_H = 125$ GeV)	0,1	0,045	2222	

Dieser und alle nachfolgenden Filter werden für ein Ereignis, also zeilenweise in der .csv Datei angewendet, da durch das Betrachten von mehr als vier Leptonen pro Ereignis, die Gesamtanzahl an vorhandenen Teilchen auch nach einem Entfernen größer als vier sein kann. Entsprechend wird wie für die Überprüfung der Leptonenanzahl auch bis zur Rekonstruktion des Z -Bosonen und des Higgs-Bosons weiterhin eine Möglichkeit gesucht eine elektrisch neutrale Ladungskombination aus den vorhandenen Teilchen im betrachteten Ereignis zu bilden. Grund dafür ist, dass die Bedingung, dass das Higgs-Boson und Z -Boson wie in Kapitel 1 beschrieben selbst ladungsneutral ist. Bei dem Mischkanal wird deshalb außerdem die Bedingung gestellt, dass es sowohl eine elektrisch neutrale Ladungskombination aus Myonen, als auch aus den Elektronen existiert.

Typ Der erste Cut beschränkt sich darauf alle Leptonen entsprechend ihrer im *skimming* durchgeführten Klassifizierung zu filtern. Myonen müssen eine globale Spur haben (G). Die Detektion nur im Spurendetektor (Zuweisung T) allein reicht aufgrund der hohen Fehlidentifikationsrate nicht aus [53]. Die Elektronen werden dagegen nur im Spurendetektor registriert, womit nur Elektronen mit der Klassifizierung (T) behalten werden. Übrige Myonen und Elektronen, die nicht aus dem Hauptkollisionspunkt stammen (S) werden verworfen.

missing hits Der Ausschluss von sekundären Elektronen, die im Spurendetektor durch Paarerzeugung aus Photonen entstehen, geschieht durch die Betrachtung der in .csv Datei gelisteten `missing_hits`. Findet eine Paarerzeugung statt, so wird das Elektron nicht ab dem Kollisionspunkt detektiert, sondern erst zu einer späteren Zeit. Es entsteht eine Lücke bei der Rekonstruktion der Trajektorie des Elektrons in dem Spurendetektor, der einer Anzahl an `missing_hits` entspricht. Des Weiteren kann diese Lücke in der Rekonstruktion auch durch die Detektoreffizienz verursacht werden. Nach [54] darf damit der Wert der `missing_hits` nicht größer als eins sein.

2.2.1. Detektorspezifische Filter

Aus der Funktionsweise des Detektors, zusammen mit seinem Aufbau, können Filter eingeführt werden um detektorspezifische Größen wie die Rekonstruktionseffizienz oder die räumliche Abdeckung zu berücksichtigen.

minimaler Transversalimpuls Die Rekonstruktionseffizienz von Myonen, die einen Transversalimpuls von $p_T < 5$ GeV besitzen ist kleiner als 95 % und die Wahrscheinlichkeit einer Missidentifizierung mit einem Hadron wird größer als 1 % [3, 53]. Aus diesem Grund werden alle Myonen, die diese p_T -Grenze unterschreiten, verworfen. Der minimale Transversalimpuls der Elektronen beträgt 7 GeV [3]. Bei höherem Transversalimpuls ist die Rekonstruktionseffizienz der Elektronen größer als 88 % im ECAL Barrel und über 90 % in den ECAL Endcaps [54].

Pseudorapidität Für die Myonen wird die Pseudorapidität auf den Bereich $0 < |\eta_\mu| < 2,4$ festgelegt. Die gesamte Raumwinkelabdeckung ($|\eta_\mu| < 2,6$) wird aufgrund der auftretenden Randeffekte nicht verwendet. Für die Elektronen ist die Pseudorapidität auf den Bereich $0 < |\eta_e| < 1,479 \wedge 1,653 < |\eta_e| < 2,5$ beschränkt. Grund für die Intervalllücke ist, dass zwischen dem ECAL-Barrel und den ECAL-Endcaps ein Übergangsbereich existiert, der nur teilweise abgedeckt wird. In diesem ist keine sinnvolle Rekonstruktion von Elektronen aus dem Spurendetektor möglich.

2.2.2. Kinematische Filter

Die im Weiteren verwendeten Filter sollen die Kinematik des Prozesses bestmöglich auf den Signalprozess einschränken und die Sensitivität steigern. Leptonen, deren Rekonstruktion ungenügend genau ist oder nicht zum betrachteten Prozess passen, werden so aussortiert.

relative Isolation Die räumliche relative Isolation der Leptonen wird durch den Zusammenhang (2.1) beschrieben.

$$\mathcal{R}_{\text{iso}}^\ell = \frac{\sum_i p_{T_i}}{p_{T_\ell}}. \quad (2.1)$$

Hierzu wird die Summe der Transversalimpulse aller nicht leptonischer Teilchen durch den Transversalimpuls des betrachteten Leptons gebildet [55]. Die relative Isolation entspricht einem Kegel um den Impulsvektor eines Leptons nach der Kollision und besitzt beim Eintritt in den ECAL den Radius ΔR_{ij} . Der Mindestabstand in der $\eta - \phi$ -Ebene, ΔR_{ij} , ist durch (2.2) definiert.

$$\Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 - \Delta\phi_{ij}^2}. \quad (2.2)$$

Mit $\Delta\eta_{ij} = \eta_i - \eta_j$ und $\Delta\phi_{ij} = \phi_i - \phi_j$ der Teilchen i und j . Für die relative Isolation sind somit auch nicht leptonische Teilchen notwendig, die nicht in die csv Dateien herausgeschrieben werden. Die explizite Bestimmung der relativen Isolation für die jeweiligen Leptonen in einem Ereignis findet deshalb im *skimming* statt und ist in den csv Dateien unter der Variable `relIso` aufgelistet. Durch das Filtern nach der relativen Isolation wird ein möglicher Überlapp und damit eine ungenaue Rekonstruktion oder Verwechslung der betrachteten Leptonen mit anderen Teilchen verhindert. Alle Leptonen mit einer relativen Isolation kleiner als 0,35 werden in diesem Filterschritt aussortiert [3].

Stoßparameter Leptonen, die nicht aus derselben Kollision stammen oder einen ungenau bestimmten Entstehungspunkt besitzen werden durch die Eingrenzung des Stoßparameters (2.3) aussortiert.

$$\text{SIP3d} = \frac{\text{IP}}{\sigma_{\text{IP}}}. \quad (2.3)$$

Hierfür muss nach [3, 21] das Verhältnis aus dem dreidimensionalen Stoßparameter IP (*interaction point*), der durch die Koordinaten des Kollisionspunktes bestimmt wird, und seiner Unsicherheit σ_{IP} kleiner als vier sein. Die Unsicherheit auf den Ort des Kollisionspunktes in der x-y-Ebene ($d\mathbf{x}$) muss kleiner als 0,5 und in der z-Richtung ($d\mathbf{z}$) kleiner als eins sein.

eingeschränkter Transversalimpuls Unter den vier Leptonen, die in der späteren Rekonstruktion der Z -Bosonen Paare verwendet werden muss ein Lepton einen Transversalimpuls größer als 20 GeV und ein zweites Lepton einen Transversalimpuls größer als 10 GeV besitzen. Dieser Schwellenwert verbessert die Wahrscheinlichkeit, dass die für die Rekonstruktion verwendeten Leptonen aus dem Hauptarbeitsbereich des Detektors stammen [3, 56]. Ebenso wird dadurch die Wahrscheinlichkeit gesteigert, ein reelles und virtuelles Z -Boson rekonstruieren zu können.

$m_{2\ell}$ -Massenausschluss Des Weiteren werden Ereignisse verworfen, in denen keine invariante Masse aus zwei Leptonen unterschiedlicher elektrischer Ladung ($m_{2\ell}$) gebildet werden kann, die größer als 4 GeV ist. Dadurch werden die unerwünschten hadronischen Resonanzen bei geringeren Massen unterdrückt [3].

$m_{4\ell}$ -Massenausschluss Abschließend werden alle Ereignisse aussortiert in denen keine invariante Masse aus vier Leptonen gefunden werden kann, die in dem am Anfang des Kapitels gewählten Massenintervall von 70 GeV und 180 GeV liegt.

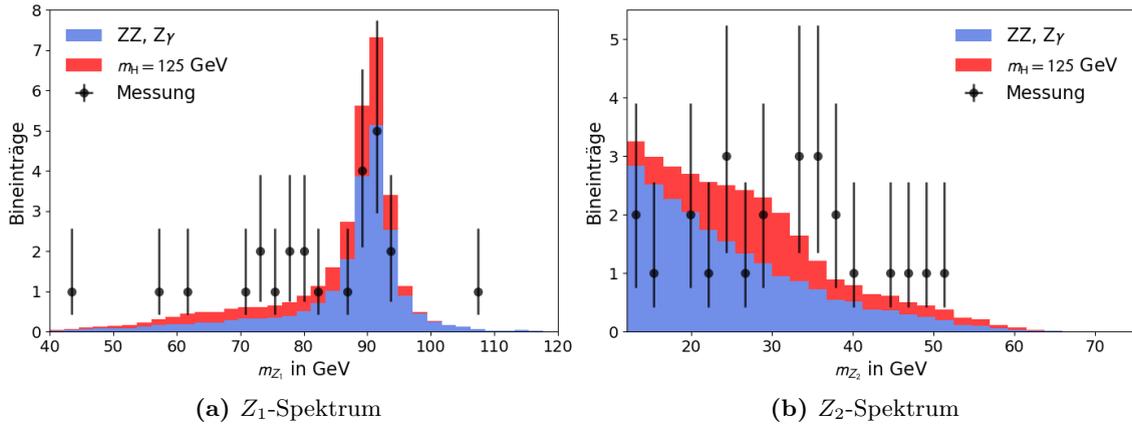


Abbildung 2.1.: Masseverteilung der rekonstruierten Z_1 - und Z_2 -Bosonen für einen Massenbereich von $106 \text{ GeV} < m_{4\ell} < 151 \text{ GeV}$. Die MC-Simulationen wurden hierfür an die Daten skaliert (Abschnitt 2.4).

2.3. Rekonstruktion

Aus den so gefilterten Ereignissen können Z -Bosonen Paare rekonstruiert werden. Jedes in der Rekonstruktion verwendete Leptonenpaar muss elektrisch neutral sein. ΔR_{ij} (2.2) zwischen den zwei Leptonen wird größer als 0,02 gewählt. Für den $2\mu 2e$ -Kanal erfolgt die Rekonstruktion eines Z -Bosons immer nur in einer Leptonenfamilie.

Aus allen möglichen rekonstruierten Z -Bosonen wird der Z_1 -Kandidat ausgewählt, dessen Masse am nächsten zu Z -Boson Masse ist. Für die verbleibenden möglichen Z -Boson Kombinationen wird das Z -Boson mit der größten skalaren p_T Summe als Z_2 gewählt [56]. Abschließend wird verlangt, dass das Z_1 -Boson die Bedingung $40 \text{ GeV} < Z_1 < 120 \text{ GeV}$ und für das Z_2 -Boson die Bedingung $12 \text{ GeV} < Z_2 < 120 \text{ GeV}$ erfüllt. Ziel ist es sicherzustellen, dass das Z_1 -Boson reell und das Z_2 -Boson ein virtuelles Teilchen ist, wodurch gewährleistet wird, dass das Ergebnis der Rekonstruktion von $m_{4\ell}$ im gewählten Massenbereich liegt. Die resultierenden m_{Z_1} - und m_{Z_2} -Spektren sind in Abbildung 2.1 dargestellt. Die Z_2 -Kandidaten sind entsprechend der Forderung hauptsächlich virtuell (Abbildung 2.1b). Die Verteilung der Z_1 -Kandidaten weist einen deutlichen Peak bei der Masse des Z -Bosons und erfüllt die Forderung nach einem reellen Z -Boson. Die Rekonstruktion von $m_{4\ell}$ aus den $Z_1 Z_2$ -Paaren erfolgt bis auf die getroffene Einschränkung auf ein Bereich zwischen 70 GeV und 180 GeV, die vor der Rekonstruktion stattfindet, ohne weitere Cuts.

Alle Filterschritte sind in Tabelle 2.3 mit dem relativen Anteil an Ereignissen gegenüber dem vorherigen Schritt aufgelistet und in Abbildung 2.2 als Absolutwerte aufgetragen. Nach jedem Schritt wird eine Prüfung nach einer möglichen elektrisch neutralen Ladungskombination und der Minimalanzahl der Leptonen im Ereignis durchgeführt. Die Prüfung nach dem Typ verringert wie erwartet hauptsächlich die Anzahl der Ereignisse im 4μ -Kanal. Zu der Entfernungen von Leptonen, die nicht einer Kollision zugeordnet werden können (S) werden alle Myonen entfernt, die, wie die Elektronen, nur im Tracker, aber nicht noch zusätzlich im Myonen Systems detektiert wurden. Auffällig ist, dass die verbleibenden Ereignisse in allen Zerfallskanälen nach der Überprüfung des minimalen Transversalimpulses und der relativen Isolation drastisch sinken. Damit sind diese beiden Filterkriterien am effektivsten bei der Reduktion auf geeignete Ereignisse. Der Cut auf die relative Isolation der Teilchen ist anders als der vom Detektor festgelegte minimale Transversalimpuls ein Filter, zur Sensitivitätssteigerung und damit bis zu einem gewissen Grad ³ frei wählbar. Die Reihenfolge der Filterschritte ist im Rahmen des Versuches frei wählbar.

³Die verwendete untere Grenze für die relative Isolation hat nach [3] den Wert 0,3. Die Wahl einer größeren relativen Isolation ist, mit weiterem Ereignisverlust als Konsequenz, frei.

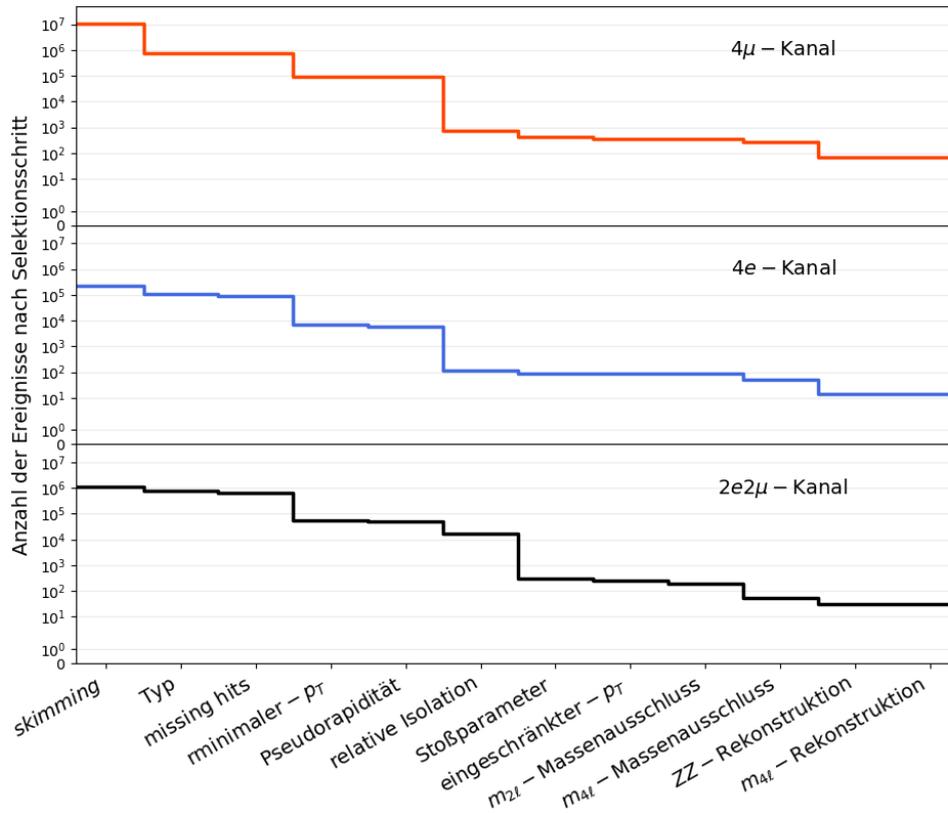


Abbildung 2.2.: Visualisierung der Anzahl an Ereignissen nach jeweiligem Selektionsschritt, aufgeteilt in die Zerfallskanäle 4μ , $4e$ und $2e2\mu$. Die Prüfung nach der Minimalanzahl an Leptonen und einer existierenden elektrischen Ladungskombination erfolgt nach jeder Filteranwendung bis zum $m_{2\ell}$ -Ausschluss.

Tabelle 2.3.: Aufzählung aller Filter und Rekonstruktionsschritte in der im Kapitel erwähnten Reihenfolge. Nach jedem Schritt bis zum $m_{2\ell}$ -Ausschluss wird nach der Minimalanzahl der Leptonen im Ereignis geprüft und eine elektrisch neutrale Ladungskombination gesucht. Der relative Anteil der übrigen Ereignisse bezieht sich immer auf die Ereignisanzahl aus dem vorhergegangenen Filter- bzw. Rekonstruktionsschritt nach der Anwendung des genannten Schrittes.

Filterschritt	Übrige Ereignisse in %		
	4μ	$4e$	$2e2\mu$
<i>skimming</i>	9,21	1,03	0,21
Typ	7,08	49,0	69,41
missing hits	100,00	81,57	85,43
minimaler Transversalimpuls	12,56	7,73	8,01
Pseudorapidität	99,30	83,77	89,19
relative Isolation	0,80	2,08	35,92
Stoßparameter	58,31	73,33	1,81
eingeschränkter Transversalimpuls	84,56	98,86	84,08
$m_{2\ell}$ -Massenausschluss	99,72	100,0	70,08
$m_{4\ell}$ -Massenausschluss	72,96	58,62	28,65
ZZ Rekonstruktion	25,48	27,45	58,49
$m_{4\ell}$ Rekonstruktion	100,00	100,00	100,00

Tabelle 2.4.: Für die Skalierung von MC-Simulationen auf die Messung verwendeten aus der Theorie bestimmten Wirkungsquerschnitte für die unterschiedlichen Zerfallskanäle [5, 57].

Kanal	σ_{theo} in fb
$ZZ \rightarrow 4\mu$	76,91
$ZZ \rightarrow 4e$	76,91
$ZZ \rightarrow 2e2\mu$	176,7
$H \rightarrow ZZ$	6,5

2.4. Doppeltzählung und Skalierung der MC-Simulation

In diesem Abschnitt soll zunächst eine Korrektur durchgeführt werden, die durch Doppeltzählung von Ereignissen entsteht. Anschließend erfolgt nach einer Skalierung der MC-Simulationen an die vorhandene Messung die abschließende Betrachtung des vier Leptonen invarianten Massenspektrums.

Das *skimming* des $2e2\mu$ -Mischkanals geschieht auf den Daten der Messung des 4μ -Kanals. Zudem wird das Signal-MC im *skimming* für alle Zerfallskanäle verwendet. Aus diesem Grund kann es durch die Betrachtung von mehr als vier Leptonen zu der Gefahr kommen, dass Ereignisse doppelt auftauchen. Für die Beseitigung von doppelten Ereignissen werden `run` und `eventId` von den Dateien verglichen. Bei den Messungen werden Dateien verglichen, welche Ereignisse aus denselben Zerfallskanälen haben. Die Ergebnisse der MC-Simulationen werden miteinander verglichen, sobald sie aus derselben Simulationsdatei stammen. Das Ereignis mit der rekonstruierten Z_1 -Masse, die am nächsten zur Z -Boson Masse ist, wird behalten. In dem Signal-MC für $m_H = 125$ GeV und weiteren Massen für das Jahr 2012 finden sich keine doppelten Ereignisse⁴. Die Daten sind aufgrund ihrer starken Reduktion nicht von der Problematik betroffen. Diese Berücksichtigung ist im Hinblick auf die Messung notwendig, da bei einer sehr geringen Anzahl an verbleibenden Ereignissen ein zusätzliches Ereignis im Signalbereich zu einer signifikanten Verfälschung führen kann.

Für die Skalierung der MC-Simulationen an die vorhandene Messung wird der Faktor $f_{\text{MC} \rightarrow \text{D}}$ eingeführt, der einem Verhältnis der Luminositäten aus der Messung (\mathcal{L}_{D}) und der MC-Simulation (\mathcal{L}_{MC}) entspricht. Die Luminosität einer MC-Simulation wird über das Verhältnis aus der Gesamtzahl der Ereignisse in der MC-Simulation (N_{tot}) und dem theoretischen Wirkungsquerschnitt σ_{theo} des betrachteten Zerfallskanal nach Gleichung (1.6) berechnet. Damit findet eine Skalierung nach Gleichung (2.4) statt, in welcher die gestrichene Größe (N'_{ij}) die korrigierte Anzahl an Ereignissen pro Bin i in einem der drei Zerfallskanäle j ist.

$$N'_{ij} = f_{\text{MC}_j \rightarrow \text{D}} \cdot N_{ij} = \frac{\mathcal{L}_{\text{D}} \sigma_{\text{theo}_j} k}{N_{\text{tot}_j}} N_{ij}, \quad j \in \{4\mu, 4e, 2e2\mu\}. \quad (2.4)$$

Der Faktor $k = 1,386$ [5] ist ein für alle Zerfallskanäle globaler Skalierungsfaktor, der eingeführt wird um die MC-Simulation, die nach der führenden Ordnung abrichtet (LO), auf die nächstführende Ordnung zu skalieren (NLO). Die Gesamtanzahl der Ereignisse in den MC-Simulationen sind in der Tabelle 2.1 aufgelistet. Die 2012 erreichte integrierte Luminosität beträgt $19,79 \text{ fb}^{-1}$. Die notwendigen Wirkungsquerschnitte finden sich in der Tabelle 2.4.

Mit dieser Skalierung der MC-Simulationen lässt sich das finale invariante Massenspektrum der vier Leptonen wie in Abbildung 2.3 darstellen und mit der Messung vergleichen. Der

⁴Für die in dieser Studie nicht betrachtete Signal-MC aus dem Jahr 2011 finden sich dagegen 2,93 % doppelt gezählter Ereignisse für eine Higgsmasse von 125 GeV

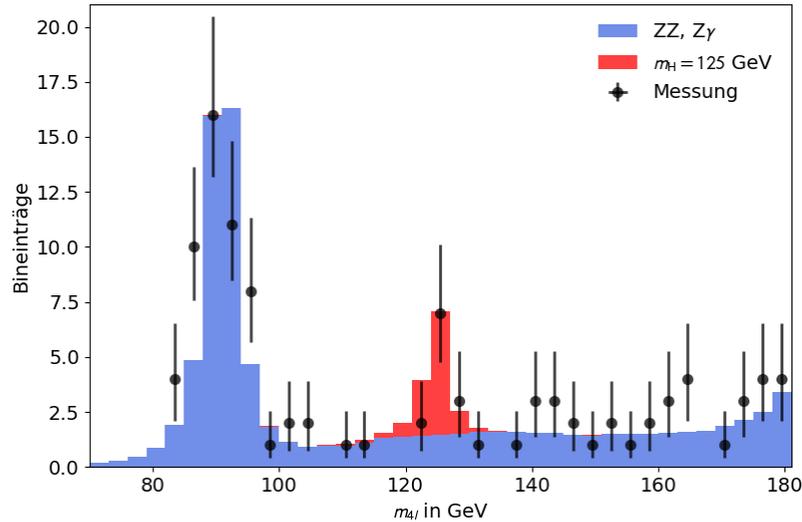


Abbildung 2.3.: $m_{4\ell}$ -Spektrum in einem Massebereich von 70 GeV und 180 GeV. Abgebildet sind inklusiv die Zerfallskanäle 4μ , $4e$ und $2e2\mu$, wobei die Messung aus dem gesamten Run 2012 stammt. Die Normierung der MC-Simulationen ist nach (2.4) erfolgt.

ausgeprägte ZZ -Peak im Massenbereich von 91 GeV deckt sich mit der Simulation des Untergrundes. Die eingezeichneten Unsicherheiten entsprechen dem 68% Intervall einer Poissonverteilung mit der Anzahl der gemessenen Ereignisse pro Bin als Mittelwert. Für kleine Mittelwerte ist die Poissonverteilung stark asymmetrisch und konvergiert erst für große Zahlen gegen die symmetrische Normalverteilung. Die grafische Darstellung soll diese Tatsache den Studenten verdeutlichen, da dies in weiteren Rechnungen implizit verwendet wird. Die in Abbildung 2.3 deutlich erkennbare Resonanz um die Masse des Z -Bosons vom Untergrund-MC deckt ebenso wie die Signal-MC die Messung gut ab.

Im Rahmen des Versuches ist diese Darstellung als solche didaktisch erst nach der Durchführung der im nächsten Kapitel vorgestellten statistischen Analyse sinnvoll, da die Wahl des Signal-MC von $m_{4\ell} = 125$ GeV, trotz des Wissens um die Entdeckung im Jahr 2012, erst getroffen werden kann, nachdem andere MC-Simulationen des Signals (Tabelle 2.1) ausgeschlossen werden. Die Abbildung 2.4 zeigt beispielhaft zwei weitere Verteilungen des Signal-MC, die ohne eine statistischen Betrachtung auch eingezeichnet werden könnten.

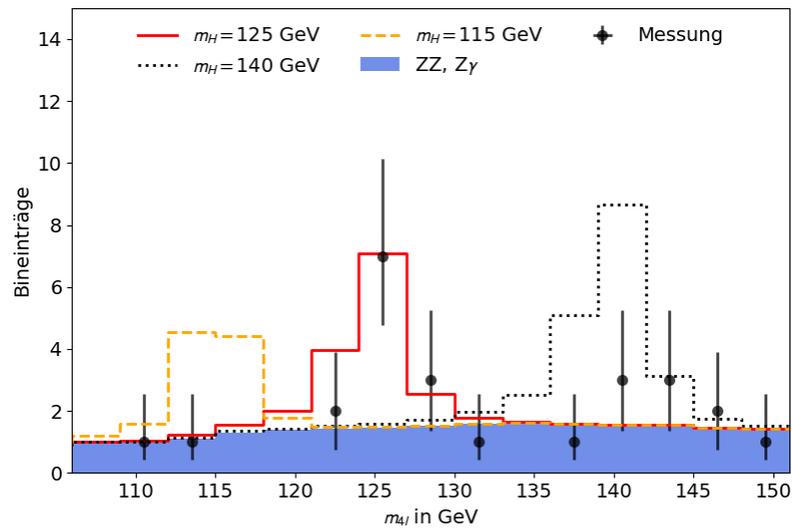


Abbildung 2.4.: $m_{4\ell}$ -Verteilung im Bereich 106 GeV und 151 GeV mit einer Binbreite von 3 GeV. Zusätzlich sind zu der rekonstruierten Untergrundverteilung (blau) und den gemessenen Ereignissen (schwarz) exemplarisch drei mögliche Signal-MCs der Massen 115 GeV, 125 GeV und 140 GeV.

3. Statistische Analyse

In diesem Kapitel sollen die aus dem Kapitel 2 erhaltenen invarianten Massenspektren, mit den im Folgenden eingeführten Methoden statistisch untersucht werden. Das Ziel ist es eine passende Annahme, eine Hypothese, über ein mögliches Signal eines Standardmodell Higgs-Bosons anhand der vorliegenden Messung auszuwählen und dessen lokale Signifikanz zu bestimmen.

Als Ausgangspunkt für die Analyse wird die Annahme getroffen, dass das Untergrund-MC nur Abweichungen von den beobachteten Messungen aufweist, die im Rahmen anderer Studien als vernachlässigbar gering eingestuft wurden [58–60]. Damit soll im Folgenden eine Argumentation aufgebaut werden, die die Hypothese, bei der nur Untergrund angenommen wird, zugunsten einer Hypothese, die neben dem Untergrund auch noch ein Signal annimmt, verwirft, wodurch mit dem Wissen aus anderen Zerfallskanälen auf das Higgs-Boson geschlossen werden kann. Die Betrachtung findet in dem im Kapitel 2 festgelegten Bereich zwischen 106 GeV und 151 GeV statt.

3.1. Untergrundstudie

Die Parametrisierung der Untergrund-MC erfolgt aufgrund der großen Ereignisanzahl und den nicht mehr poissonartigen auf jeden Bin skalierten Unsicherheiten mit der Methode der kleinsten Quadrate. In der Untergrund-MC kommen zu der Tatsache, dass die eingehenden verwendeten Berechnungen aus zeitlichen Gründen nach der führenden Ordnung abbrechen noch zusätzlich die Effekte der Detektorsimulation hinzu. Dadurch kann kein analytisches Modell für die Beschreibung der Simulation gefunden werden. Vor allem entsteht durch den Abbruch der Rechnung nach der führenden Ordnung eine systematische Unterschätzung der Unsicherheiten, wodurch mögliche angenommene Modelle oft $\chi_{n_{\text{dof}}}^2 \geq 1$ aufweisen. Um das zu kompensieren werden für die Wahl eines passenden Modells die Unsicherheiten σ_i in Gleichung (1.7) auf die jeweiligen Bins multiplikativ mit dem Faktor $\sqrt{\chi_{\text{min}}^2}$ derart skaliert, dass $\chi_{n_{\text{dof}}}^2 = 1$ erfüllt ist. Durch eine relative Unsicherheitskalierung sollen Effekte wie die ungenaue Detektorantwort auf eine Ereignisdetektierung, die Akzeptanz und Effizienz vor allem aber die Systematik mitberücksichtigt werden. Die Einführung eines weiteren Faktors für die Korrektur der Unsicherheit geschieht damit ähnlich wie beispielsweise die Betrachtung der Akzeptanz in der MC-Simulation ebenfalls multiplikativ. Die Form des angenommenen Modells bleibt bei einer derartigen Skalierung der Unsicherheiten unverändert und vergrößert nur die Unsicherheiten auf die Parameter des Modells. Das Ziel ist damit ein möglichst einfaches Modell für die Parametrisierung zu wählen, bei welcher der Skalierungsfaktor minimal und größer als eins ist. Damit wird sichergestellt, dass zusätzliche Quellen für die Unsicherheiten die angenommen werden müssen im Vergleich zu den rein statistischen Unsicherheiten klein sind und keine Überparametrisierung stattfindet.

Für die Modellannahme wird eine Linearkombination aus Legendrepolyomen $P_n(x)$ verwendet, die orthogonal zueinander sind, und die x-Koordinate um den Schwerpunkt

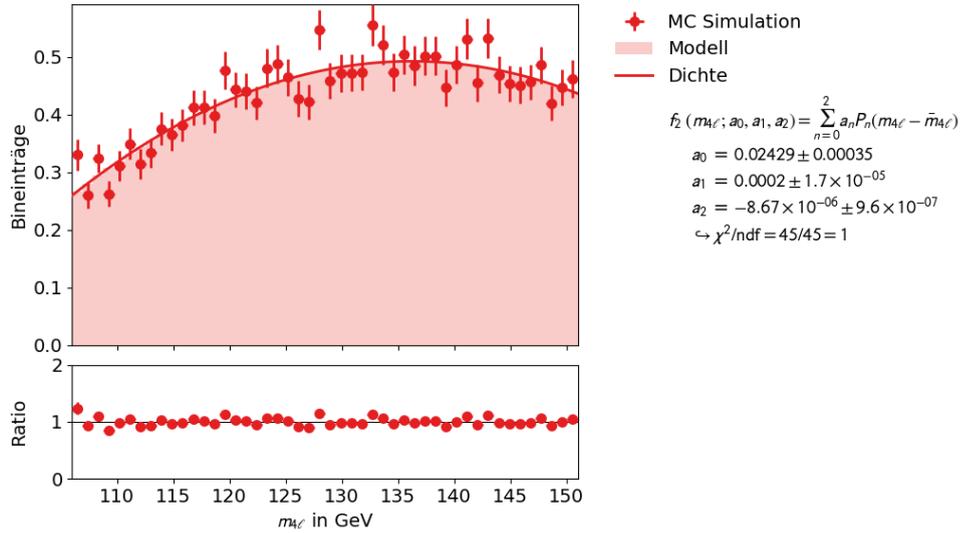


Abbildung 3.1.: Histogrammanpassung des an die Daten skalierten Untergrund MC durch das verwendete Modell (3.1) aus einer Linearkombination von Legendrepolynomen bis zum Grad $n = 2$ durch die Softwareumgebung KAFE2 [51]. Das Ratio gibt dabei das Verhältnis zwischen der bestimmten, an die Daten skalierten Wahrscheinlichkeitsdichte an und die für die Anpassung verwendeten Bins. Die Unsicherheiten wurden für die Anpassung mit dem Faktor $\sqrt{\chi^2_{\min}} = 1.015$ skaliert.

Tabelle 3.1.: Durch die Anpassung (Abbildung 3.1) bestimmten Parameter für die Untergrundparametrisierung mittels (3.1). Diese Menge an Parametern wird im folgenden als θ_{0_U} bezeichnet.

Parameter	Wert
a_0	$0,02429 \pm 3,5 \cdot 10^{-4}$
a_1	$0,0002 \pm 1,7 \cdot 10^{-5}$
a_2	$-8,67 \cdot 10^{-6} \pm 9,6 \cdot 10^{-7}$

der x-Achse ($m_{4\ell} - \bar{m}_{4\ell}$) zentriert. Die einzelnen Parameter werden somit voneinander teilweise entkoppelt, was in einer geringeren Korrelation zwischen diesen resultiert (Anhang A). Diese Wahl ist für die spätere Unsicherheitenbetrachtung vorteilhaft, in welcher eine Variation der Parameter durchgeführt wird. Tabelle 3.2 und Abbildung 3.2 zeigen die Skalierungsfaktoren weiterer Linearkombinationen der Legendrepolynome und verdeutlicht, dass das einfachste Modell zur Beschreibung des Untergrundes eine Linearkombination aus Legendrepolynomen bis zur zweiten Ordnung (3.1) ist. Die Ergebnisse dieser Anpassung (Abbildung 3.1) sind in Tabelle 3.1 aufgelistet. Diese Parameter werden im Weiteren als θ_{0_U} bezeichnet.

$$f_2(m_{4\ell}) = a_0 P_0(m_{4\ell} - \bar{m}_{4\ell}) + a_1 P_1(m_{4\ell} - \bar{m}_{4\ell}) + a_2 P_2(m_{4\ell} - \bar{m}_{4\ell}). \quad (3.1)$$

Linearkombinationen aus Legendrepolynomen höheren Grades neigen zu einer Überparametrisierung, wodurch zunehmend die statistischen Schwankungen der vorhandenen MC-Simulation berücksichtigt werden. Die Anpassungen für Linearkombinationen von Legendrepolynomen der in Tabelle 3.2 aufgeführten weiteren Grade sind im Anhang A hinterlegt. Die Bedingung der Normierung einer Wahrscheinlichkeitsdichte wird in den Parametern der Histogrammanpassung bereits berücksichtigt und findet deshalb explizit nicht statt.

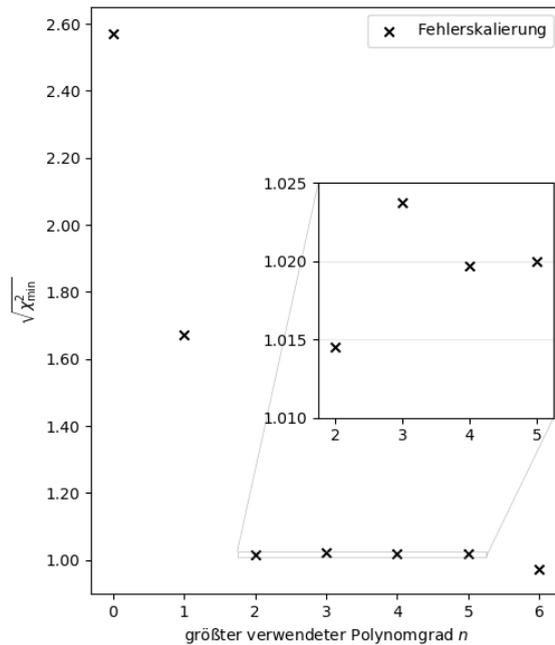


Tabelle 3.2.: Ergebnis der globalen Skalierung der Unsicherheiten der Bins σ_i durch den Faktor $\sqrt{\chi_{\min}^2}$ derart, dass $\chi_{n_{\text{dof}}}^2 = 1$ gilt. n entspricht dem größten Grad des Legendrepolynoms in der verwendeten Linearkombination.

n	$\sqrt{\chi_{\min}^2}$
0	2,572
1	1,672
2	1,015
3	1,024
4	1,020
5	1,020
6	0,974

Abbildung 3.2.: Visualisierung der beschriebenen Skalierung der Unsicherheiten mit dem Faktor $\sqrt{\chi_{\min}^2}$ für unterschiedliche Polynomgrade.

3.2. Signalbreitenabschätzung

Ausgehend von Abbildung 2.4 soll eine passende Signalthypothese ausgewählt werden, welche die vorhandene Messung am besten beschreibt. Die Auswahl an Möglichkeiten wird hierbei im Rahmen der Analyse auf den gewählten Abschnitt beschränkt und enthält simulierte Verteilungen der Detektorergebnisse für Standardmodell (SM) Higgs-MC Simulation (Signal-MC) unterschiedlicher Massen. Die Auflistung der verwendeten Massen ist in Tabelle 2.1 zu finden. Betrachtet wird die Detektorantwort und die daraus resultierende Rekonstruktion des vollständig simulierten Prozesses. Der Vorteil dieser Betrachtung ist, dass in die Simulation des Detektors alle notwendigen Korrekturen bereits einfließen und nicht zusätzlich betrachtet werden müssen. Aus der natürlichen Breite des SM-Higgs-Bosons, die nur wenige MeV groß ist [18], lässt sich schließen, dass die beobachtete Breite in den Signal-MC durch die Transversalimpulsauflösung des Detektors für Myonen und Elektronen dominiert ist, die mehrere Größenordnungen über der natürlichen Breite liegt [61, 62]. Das Ziel ist es eine quantitative Beschreibung der Breite von den Signal-MCs in Abhängigkeit der Masse zu finden, um damit auf die passende Verteilung des Signal-MC zu schließen.

Für die Signalabschätzung wird der einfachste Fall einer gaußförmigen Resonanz mit der Breite $\sigma_G(m_{4\ell})$ und dem Schwerpunkt $\mu_G = m_{4\ell}$ angenommen, da im Rahmen der Abschätzung eine spezifischere Annahme über die Form der Verteilung anhand der vorhandenen Messung nicht sinnvoll ist. Die Parametrisierung mit einer Gaußverteilung für das Signal-MC für eine Higgs-Masse von 125 GeV ist in Abbildung 3.3 beispielhaft dargestellt und soll, vor allem durch den Skalierungsfaktor $\sqrt{\chi_{\min}^2} = 7,77$ verdeutlichen, dass eine Gaußverteilung als finale Parametrisierung nicht ausreichend ist. Der Bereich um den Peak wird, wie der Ratio-Plot zeigt, ausreichend gut durch die Gaußverteilung parametrisiert. Der Effekt der Überschätzung der Breite ist im Rahmen der Abschätzung klein und wird bei der Betrachtung des Ergebnisses diskutiert. Die Anpassung für weitere Massen sind im Anhang B zu finden.

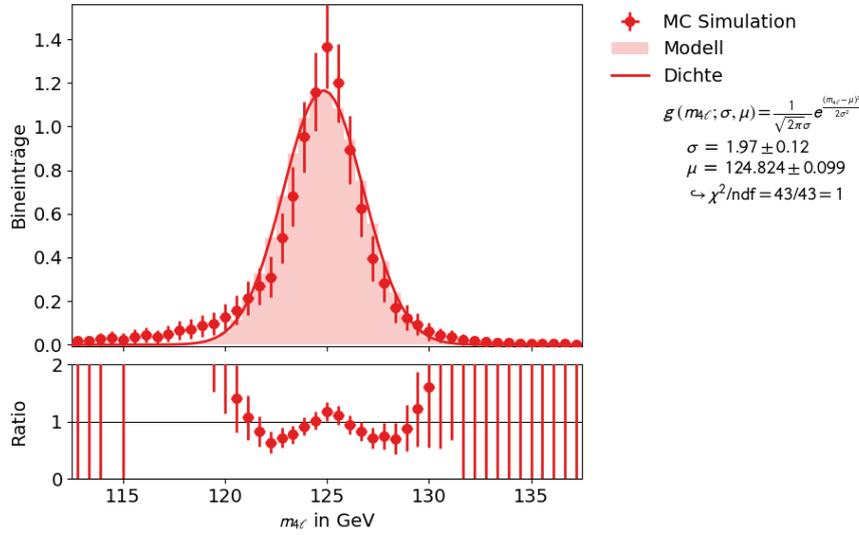


Abbildung 3.3.: Anpassung an ein Histogramm des Signal-MC eines SM-Higgs-Bosons der Masse 125 GeV mit der Softwareumgebung `KAFE2` [51]. Das Signal-MC wurde an die vorhandene Messung skaliert. Das Ratio gibt das Verhältnis zwischen der durch die Anpassung bestimmten und skalierten Wahrscheinlichkeitsdichte und der vorhandenen Messung an. Der Skalierungsfaktor ist $\sqrt{\chi^2_{\min}} = 7,77$.

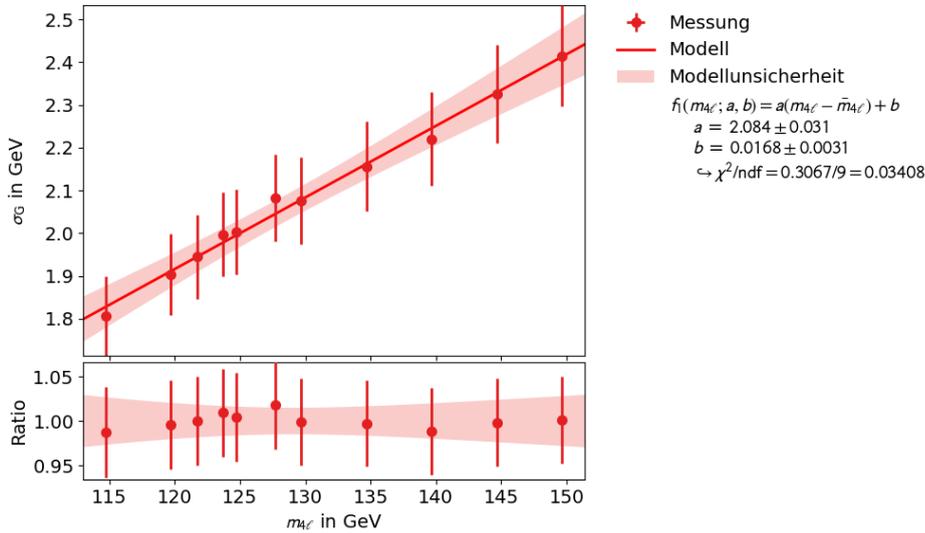


Abbildung 3.4.: Lineare Anpassung der unterschiedlichen Breiten σ_G der Gaußverteilungen in Abhängigkeit von $m_{4\ell}$ durch die Verwendung der Softwareumgebung `KAFE2` [51]. Jeder der Punkte wurde einer Signal-MC der entsprechenden Higgs-Masse durch eine Gaußparametrisierung entnommen. Die Anpassung der Signal-MC Verteilungen für weitere Higgsmassen ist im Anhang B hinterlegt. Das Ratio gibt das Verhältnis zwischen der bestimmten Funktion und für die Anpassung verwendeten Messpunkten an.

Tabelle 3.3.: Durch die Anpassung (Abbildung 3.4) bestimmten Parameter für die Signalbreitenabschätzung mittels eines linearen Modells . (3.2) für unterschiedliche simulierte Signalverteilungen.

Parameter	Wert
a	$2,0538 \pm 0,0037$
b	$0,016\ 15 \pm 0,000\ 35$

Die daraus gewonnene Massenabhängigkeit von $\sigma_G(m_{4\ell})$ lässt sich näherungsweise durch ein lineares Modell

$$\sigma_G(m_{4\ell}) = a(m_{4\ell} - \bar{m}_{4\ell}) + b, \quad (3.2)$$

mit den in Tabelle 3.3 aus der Anpassung (Abbildung 3.4) bestimmten Parametern beschreiben.

3.3. Abschätzung lokaler Signifikanz – der lokale p_0 - Wert

Für die Auswahl eines geeigneten Signal-MC soll ein Hypothesentest für unterschiedliche Higgsmassen durchgeführt werden. Die Hypothese H_0 entspricht dabei der Annahme, dass kein Signal vorhanden ist. Somit kann diese Hypothese mit der pdf des parametrisierten Untergrundes $f_{H_0}(x|\theta_{0_{H_0}}) = f_b(x|\theta_{0_U})$ beschrieben werden. H_1 , die Gegenhypothese, die ein mögliches Signal der Stärke $\hat{\mu}$ annimmt, ist durch $f_{H_1}(x|\theta_{0_{H_1}}) = \hat{\mu}f_s(x|\sigma_G, m_H) + f_b(x|\theta_{0_U})$ parametrisiert. Die Wahrscheinlichkeitsdichte der Signal-MC ist nach der Abschätzung im Abschnitt 3.2 gausförmig. Der optimale Parameter $\hat{\mu}$, auch Schätzer genannt, wird für die jeweilige Higgsmasse separat durch eine Likelihood-Anpassung bestimmt. Störparameter, die multiplikativ mit der bestehenden Likelihood die Gesamtl likelihood bilden werden nicht mitberücksichtigt. Damit soll der Fokus auf der Auswahl eines geeigneten Signal-MC bleiben.

Das Ziel ist es eine bestmögliche Trennung beider Hypothesen zu erreichen um eine Aussage treffen zu können, welche von ihnen verworfen werden kann. Nach Neyman-Pearson-Lemma [63] geschieht dies am besten durch die Teststatistik, die das Verhältnis $\lambda(m_H)$ der Likelihoods beider Hypothesen bildet, Gleichung (3.3). Für die numerische Stabilität wird der negative Logarithmus des Verhältnisses, ähnlich zu der Gleichung (1.12) gebildet. Mit der Einführung des für die asymptotische Form notwendigen Faktors zwei, ist die Teststatistik q_0 gegeben durch:

$$q_0 = q_0(m_H) = -2 \ln \lambda(m_H) = \begin{cases} -2 \ln \frac{\mathcal{L}(x|\theta_{H_1})}{\mathcal{L}(x|\theta_{H_0})} & \hat{\mu} > 0 \\ 0 & \text{sonst.} \end{cases} \quad (3.3)$$

Der Schätzer $\hat{\mu}$ wird auf ein positives reelles Intervall eingeschränkt, da das erwartete Signal sich durch einen Ereignisüberschuss im Massenspektrum zeigt.

Der p_0 -Wert kann als ein kritischer Wert betrachtet werden, der eine Aussage darüber macht, ob die betrachteten Unterschiede zwischen zwei Hypothesen aus reinem Zufall entstanden sind. Er ist mit (3.4) definiert.

$$p_0 = \int_{q_{0_{\text{obs}}}^0}^{\infty} dq_0 f(q_0|\hat{\mu} = 0). \quad (3.4)$$

Die Funktion $f(q_0|\hat{\mu} = 0)$ ist die Wahrscheinlichkeitsdichte von der Teststatistik q_0 unter der Annahme, dass kein Signal ($\hat{\mu} = 0$) erwartet wird. Durch diese Betrachtung wird die Unverträglichkeit von $q_{0_{\text{obs}}}$ mit der Hypothese kein Signal zu messen, $\hat{\mu} = 0$ bzw. $f_{H_0}(x|\theta_{0_{H_0}})$, bestimmt. Für steigende Werte von $\hat{\mu}$ wird diese Hypothese immer weiter ausgeschlossen und die Integration ab dem beobachteten $q_{0_{\text{obs}}}$ bestimmt die Wahrscheinlichkeit kein besseres $q_{0_{\text{obs}}}$ unter einer Experimentwiederholung zu erhalten.

$f(q_0|\hat{\mu} = 0)$ kann durch die Generierung von Pseudo-Daten mittels Monte Carlo Simulation erhalten werden. Jedoch reicht für die nachfolgende Abschätzung die Betrachtung der

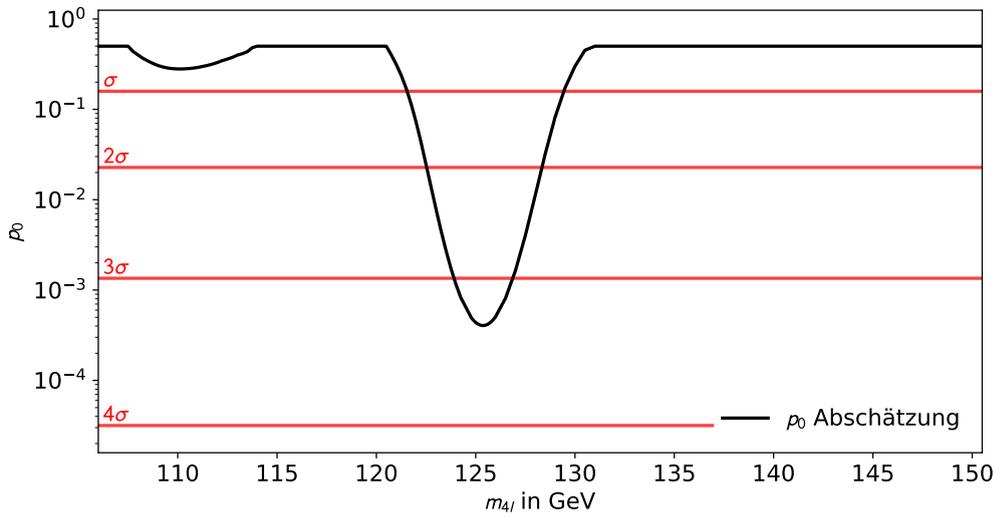


Abbildung 3.5.: Aufgetragene Schätzung von dem lokalen p_0 -Wert über $m_{4\ell}$. Die in Rot eingezeichneten Signifikanzen sind nach Gleichung (3.7) berechnet. Im Bildausschnitt nicht zu sehen ist die 5σ -Grenze für eine Entdeckung die bei einem p_0 Wert von $\simeq 2,87 \cdot 10^{-7}$ liegt.

asymptotischen Form (3.5) [64], welcher $f(q_0|\hat{\mu} = 0)$ bereits für sehr geringe Anzahl an Ereignissen folgt [65]. Diese analytische Form kann zu (3.6) integriert werden.

$$f(q_0|\hat{\mu} = 0) \rightarrow \frac{1}{2} \left(\delta(q_0) + \chi_1^2 \right), \quad (3.5)$$

$$p_0 = \frac{1}{2} \left(1 - \operatorname{erf} \left(\sqrt{\frac{q_{0\text{obs}}}{2}} \right) \right). \quad (3.6)$$

Für die Betrachtung verschiedener Massen kann der p_0 -Wert in Abhängigkeit davon lokal bestimmt werden. Aus diesem lässt sich wiederum die Signifikanz Z berechnen (3.7). Diese ist als die Fläche eines Ausläufers der Gauß-Verteilung definiert, welche $n\sigma$ von dem Mittelwert aus startet und kann durch die Inverse der Gauß'schen Verteilungsfunktion bzw. durch $q_{0\text{obs}}$ beschrieben werden.

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_{0\text{obs}}}. \quad (3.7)$$

Die Abbildung 3.5 zeigt den Verlauf des lokalen p_0 -Wertes für unterschiedliche $m_{4\ell}$ -Massen. Die größte Signifikanz von $\simeq 3,35\sigma$ ist nach dieser groben Abschätzung für eine Higgsmasse von 125,37 GeV gegeben. Die Wahl zwischen den unterschiedlichen Signal-MC aus Tabelle 2.1 fällt auf die zu dem Wert am nächsten liegende Masse von 125 GeV. Damit ist das Ziel dieses Abschnittes nach der Wahl des passenden Signal-MC mit einer statistischen Begründung erreicht. Zusätzlich findet eine erste Abschätzung der lokalen Signifikanz statt, die den Mindestwert für ein Indiz (3σ) überschreitet. Da die hier durchgeführte Studie aus didaktischen Gründen nicht vollkommen blind ist und eine Konzentration auf die Betrachtung eines bestimmten Bereiches stattfindet, besteht die Gefahr einer möglichen rein zufällig entstandenen Beobachtung¹, die jedoch auch bei einer Blindstudie, durch die Wahl eines Parameterraumbereiches nicht vollkommen verschwindet. Aus diesem Grund ist die Grenze für eine Entdeckung mit 5σ , also $p_0 \simeq 2,87 \cdot 10^{-7}$ gewählt, damit entstehende Fluktuationen in der Betrachtung nicht zu einer verfrühten und möglicherweise fehlerhaften Entdeckung führen. Die Grenze von 3σ eignet sich damit als ein guter Ausgangspunkt

¹Dies wird durch den *look-elsewhere*-Effekt beschrieben [66]

für weitere genauere Beobachtungen. Die Abschätzung des p_0 -Wertes für Daten, die in Histogrammen zusammengefasst worden sind und die daraus entstehende Verfälschungen sind beispielhaft im Anhang C hinterlegt.

3.4. Signalstudie

Nach der passenden Wahl des Signal-MC im Abschnitt 3.3, erfolgt in Analogie zum Abschnitt 3.1, die Parametrisierung dieser Verteilung durch eine passende Wahrscheinlichkeitsdichte. Anders als bei der in Abschnitt 3.2 genutzten Abschätzung für die Region des Peaks müssen für eine bessere Parametrisierung auch die Ausläufer mitberücksichtigt werden. Die Parametrisierung durch eine Gaußverteilung (Abbildung 3.3) weicht vor allem in den Bereichen außerhalb des Peaks deutlich von der Simulation ab. Weitere Parametrisierungen durch symmetrische Wahrscheinlichkeitsdichten wie der Cauchyverteilung oder dem Voigt-Profil minimieren den Skalierungsfaktor $\sqrt{\chi_{\min}^2}$ der Unsicherheiten (Anhang D.5a und D.5b), verdeutlichen jedoch die Asymmetrie der Verteilungsausläufer.

Diese Anhäufung von Ereignissen in den Verteilungsausläufern ist durch die Nutzung der Detektorsimulation auf den Detektor und damit verbundenen Detektoreffekten zurückzuführen. Die beobachtete Asymmetrie wird hauptsächlich durch eine größere Anzahl an Ereignissen im linken Ausläufer bestimmt. Grund hierfür ist die unvollständige Zuordnung der abgestrahlten Photonen zu einem detektierten Elektron, welche auf dem Weg zwischen dem Kollisionspunkt der Protonen und dem Eintritt in den ECAL Photonen abstrahlen. Damit ist für die Rekonstruktion der Initialenergie der Elektronen eine eindeutige Zuweisung der abgestrahlten Photonen notwendig. Für manche der abgestrahlten Photonen ist eine derartige Zuweisung nicht möglich. Dadurch fehlt es den Elektronen an Energie, die nach Gleichung (1.1) in einer geringeren Masse resultiert. Der rechte Teil dagegen wird hauptsächlich durch Detektorrauschen verursacht, in welchem den Leptonen mehr Energie zugeordnet wird.

Diese Asymmetrie wird am besten durch die Double-Sided-Crystal-Ball-Verteilung (DSCB-Verteilung) (3.8) beschrieben, welche sich aus einer Gaußverteilung für den mittleren Abschnitt und zwei voneinander unabhängigen Potenzgesetzen als Ausläufer zusammensetzt [67, 68].

$$DSCB = N \cdot \begin{cases} e^{-\frac{\alpha_L^2}{2}} \left(\frac{\alpha_L}{n_L} \left(\frac{n_L}{\alpha_L} - \alpha_L - x \right) \right)^{-n_L}, & x < -\alpha_L \\ e^{-\frac{x^2}{2}}, & -\alpha_L \leq x \leq \alpha_R \\ e^{-\frac{\alpha_R^2}{2}} \left(\frac{\alpha_R}{n_R} \left(\frac{n_R}{\alpha_R} - \alpha_R + x \right) \right)^{-n_R}, & x > \alpha_R \end{cases} \quad (3.8)$$

$$x = \frac{m_{4\ell} - \mu}{\sigma}, \quad (3.9)$$

$$N = \sigma \left(\sqrt{\frac{\pi}{2}} \left(\operatorname{erf} \left(\frac{\alpha_R}{\sqrt{2}} \right) + \operatorname{erf} \left(\frac{\alpha_L}{\sqrt{2}} \right) \right) + \frac{e^{-\frac{\alpha_L^2}{2}} n_L}{\alpha_L (n_L - 1)} + \frac{e^{-\frac{\alpha_R^2}{2}} n_R}{\alpha_R (n_R - 1)} \right)^{-1}. \quad (3.10)$$

Die Skalierungsfaktoren für die unterschiedlichen Verteilungen sind in der Tabelle 3.4 dargestellt. Deutlich zu sehen ist das Versagen der symmetrischen Verteilungen bei der Parametrisierung der Signal-MC. Die Single-Sided-Crystal-Ball-Verteilung (SSCB-Verteilung) ist ähnlich zu der DSCB-Verteilung aus einem Potenzgesetz und einer Gaußverteilung zusammensetzt und zeigt im Vergleich zu der Gaußverteilung eine deutliche Verbesserung des Skalierungsfaktors (Anhang D.5c). Die DSCB-Verteilung verzichtet zusätzlich auch auf den rechten Gauss Ausläufer und parametrisiert die Signal-MC am besten. Die Parameter der Anpassung der DSCB-Verteilung (Abbildung 3.6) sind in Tabelle 3.5 festgehalten und werden im Weiteren als θ_{0s} bezeichnet.

Tabelle 3.4.: Skalierungsfaktoren $\sqrt{\chi_{\min}^2}$ der Unsicherheiten, für die Bedingung $\chi_{n_{\text{dof}}}^2 = 1$, welche für die Anpassung eines Signal-MC für eine Higgsmasse von $m_H = 125 \text{ GeV}$ notwendig ist. Die Anpassungen für die aufgeführten Verteilungen sind im Anhang D.5a, D.5b und D.5c hinterlegt.

Verteilung	$\sqrt{\chi_{\min}^2}$
Gauss	7,768
Cauchy	5,984
Voigt	4,399
SSCB	4,432
DSCB	1,292

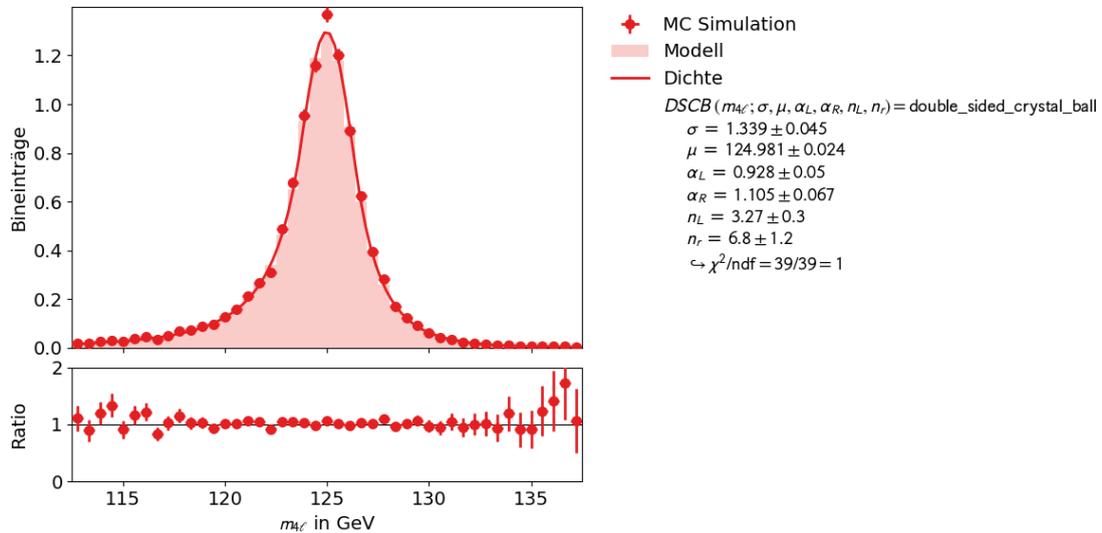


Abbildung 3.6.: Anpassung an ein Histogramm der an die Daten skalierten Signal-MC für eine Higgsmasse von $m_H = 125 \text{ GeV}$ mittels einer DSCB-Verteilung (3.8) durch die Softwareumgebung KAFE2 [51]. Das Ratio zeigt das Verhältnis zwischen der angepassten Verteilung, die an die für die Anpassung verwendeten Bins skaliert wurde und den Bins selbst. Der Skalierungsfaktor für die Unsicherheiten ist $\sqrt{\chi_{\min}^2} = 1,292$.

Tabelle 3.5.: Durch die Anpassung (Abbildung 3.6) bestimmten Parameter für die Parametrisierung der Signal-MC mit einer Higgsmasse von $m_H = 125 \text{ GeV}$ mittels der DSCB-Verteilung (3.8). Diese Parametermenge wird im Weiteren als θ_{0s} bezeichnet.

Parameter	Wert
σ	$1,339 \pm 0,045$
μ	$124,981 \pm 0,024$
α_L	$0,928 \pm 0,05$
α_R	$1,105 \pm 0,067$
n_L	$3,27 \pm 0,3$
n_R	$6,8 \pm 1,2$

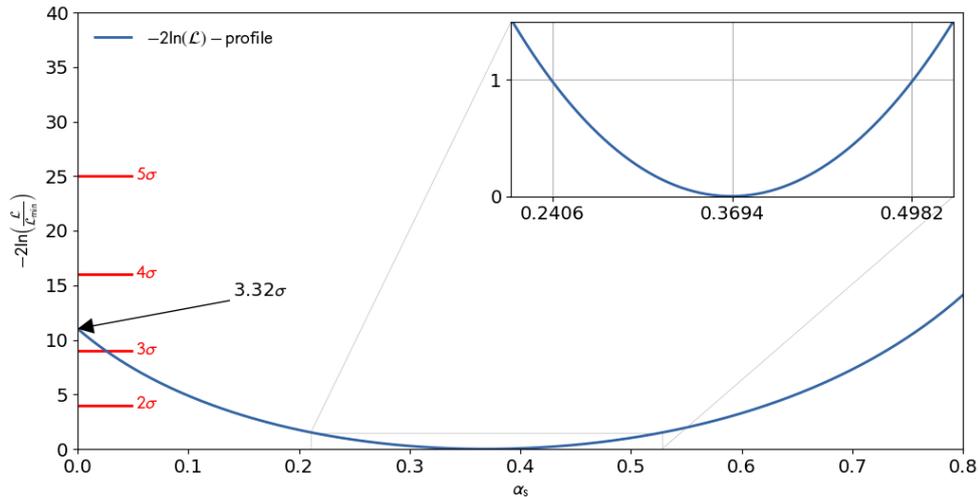


Abbildung 3.7.: Bestimmung des optimalen Signalanteils α_S mit der Likelihood Methode (1.12) und dessen Unsicherheiten, eingezeichnet in dem gezeigten Ausschnittsfenster. Auf der y-Achse sind nach (3.12) die Signifikanzen eingezeichnet, die $-2 \ln \frac{\mathcal{L}(\alpha_S)}{\mathcal{L}_{\min}}$ für den Punkt $\alpha_S = 0$ annehmen würden. Für die eingezeichnete profile-Likelihood werden die optimalen Parameterwerte θ_{0_S} und θ_{0_U} als festgehaltene Parameterwerte verwendet. Die in die Likelihood einfließenden Verteilungen ist für das Signal die DSCB-Verteilung (3.8) und für den Untergrund die Linearkombination aus Legendrepolyomen bis zum zweiten Grad (3.1).

3.5. Lokale Signifikanz

Die lokale Signifikanz beschreibt die Abweichung der beobachteten Verteilung von dem erwarteten Untergrund ohne ein Signal. Für die Bestimmung der lokalen Signifikanz wird die Hypothese (3.11) betrachtet, in welcher α_S den Signalanteil darstellt. Im Folgenden ist nur der Parameter α_S von Interesse. Alle weiteren für die Beschreibung von Signal- und Untergrund-MC verwendeten Parametern, werden bei den zuvor bestimmten Parameterwerte aus den Anpassungen (Abbildung 3.1 und 3.6) festgehalten².

$$f(x|\theta) = \alpha_S f_s(x|\theta_{0_S}) + (1 - \alpha_S) f_b(x|\theta_{0_U}). \quad (3.11)$$

$f_s(x|\theta_{0_S})$ beschreibt die in Abschnitt 3.4 eingeführten DSCB-Verteilung mit den optimalen Parametern θ_{0_S} aus Tabelle 3.5. $f_b(x|\theta_{0_U})$ ist aus der Untergrundstudie bestimmte Linearkombination aus Legendrepolyomen bis zum zweiten Grad mit den optimalen Parametern θ_{0_U} aus Tabelle 3.1.

Durch die Likelihood Anpassung lässt sich der Signalanteil zu $\alpha_S = 0,369_{-0,129}^{+0,129}$ bestimmen. Dieser ist zusammen mit den asymptotischen Signifikanzen in Abbildung 3.7 dargestellt. Es wird, ähnlich zu der im Abschnitt 3.3 durchgeführten Betrachtung, die Abweichung von der Untergrundhypothese ohne einen Signal in dem $-2 \ln \frac{\mathcal{L}(\alpha_S)}{\mathcal{L}_{\min}}$ -Profil untersucht, der dem Punkt entspricht, an der der Signalanteil verschwindet ($\alpha_S = 0$). Daraus kann die Unsicherheit auf den Signalanteil abgeleitet werden, die angenommen werden muss um für die beobachtete Messung noch die Hypothese annehmen zu können, nach der kein Signal vorhanden ist. Im Umkehrschluss, analog zu (3.7) kann in asymptotischer Näherung die Signifikanz des Signals in der vorhandenen Messung mit (3.12) bestimmt werden.

$$Z = \sqrt{\frac{\mathcal{L}(\alpha_S = 0)}{\mathcal{L}_{\min}}}. \quad (3.12)$$

²In der Literatur wird dieses Vorgehen auch als *profiled-likelihood* Methode bezeichnet [69]

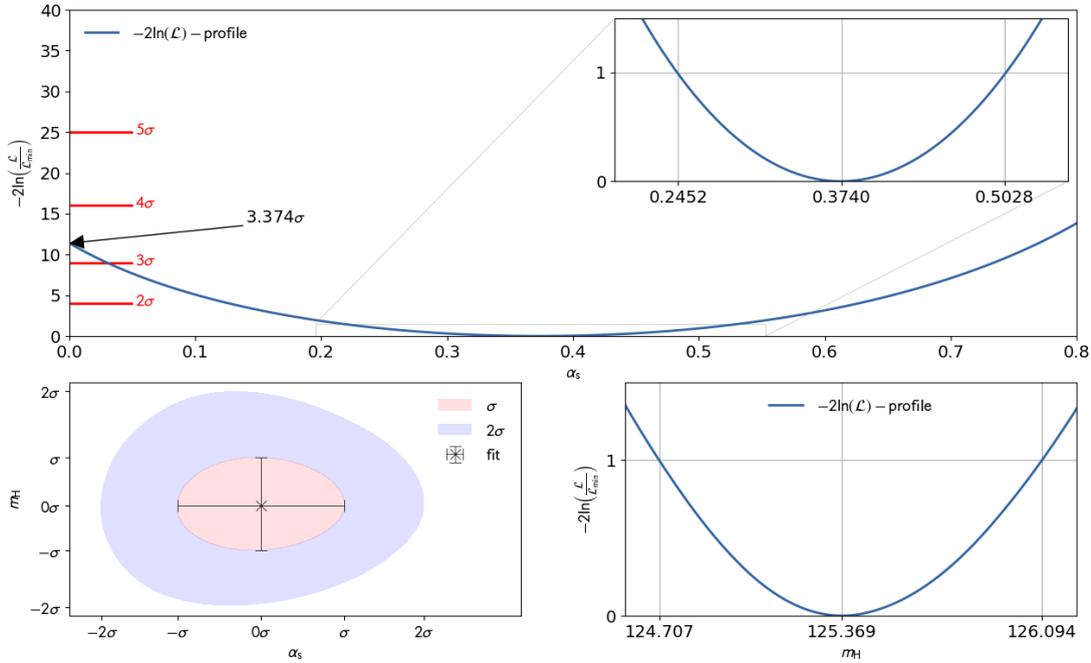


Abbildung 3.8.: Bestimmung des optimalen Signalanteils α_S und der optimalen Masse des SM-Higgs-Bosons als ein zusätzlicher Minimierungsparameter durch die *profiled likelihood* Methode der extended Likelihood, Gleichung (1.12), zusammen mit ihren Unsicherheiten. Auf der y-Achse sind die Signifikanzen eingezeichnet, welche nach Gleichung (3.12) für den Punkt $\alpha_S = 0$ gelten. Die optimalen Parameterwerte θ_{0_S} (ohne μ) und θ_{0_U} werden festgehalten. Für die Signalverteilung wird die DSCB-Verteilung (3.8) verwendet. Der Untergrund wird durch die Linearkombination aus Legendrepoly-nomen bis zum zweiten Grad (3.1) beschrieben.

Die Unsicherheitenbestimmung der Signifikanz geschieht konservativ. Hierzu wird die Signifikanz (3.12) aus allen Kombinationen bestimmt die durch die Parameter θ_{0_S} und θ_{0_U} und ihren Unsicherheiten gebildet werden können. Für die Linearkombination der Signalverteilung (3.8) mit sechs Parametern und der Untergrundverteilung (3.1) mit drei Parametern ergeben sich somit 19683 Kombinationsmöglichkeiten für die Signifikanz. Aus dieser Menge kann eine untere und obere Grenze für die Signifikanz zu

$$Z = 3,320_{-0,040}^{+0,103} \sigma,$$

abgeschätzt werden und überschreitet damit eindeutig die Grenze für ein Indiz (3σ). Diese Herangehensweise ermöglicht es die Überprüfung der Signifikanz einer fest vorgegeben Signal-MC des Standardmodell Higgs-Bosons durchzuführen. Weitere Aussagen über beispielsweise die gemessene Masse können jedoch nicht getroffen werden.

3.5.1. Massenbestimmung des SM-Higgs-Bosons

Für eine Massenbestimmung des SM-Higgs-Bosons muss die Masse als ein zusätzlicher freier Parameter aufgefasst werden. Die Masse wird durch den Parameter μ beschrieben, der den Schwerpunkt der Gaußverteilung angibt, die in die DSCB-Verteilung einfließt. Die Minimierung der so entstandenen zweidimensionalen *profile-likelihood*, ist in Abbildung 3.8 dargestellt. Das Hinzufügen der Masse des Higgs-Bosons als ein zusätzlicher freier Parameter führt zur Erhöhung der Signalstärke zu $\alpha_S = 0,374_{-0,129}^{+0,133}$ und der Signifikanz der Signalverteilung zu

$$Z = 3,374_{-0,048}^{+0,086} \sigma.$$

Das Intervall auf die Unsicherheit wird im Vergleich zu dem Ergebnis aus dem vorherigen Abschnitt damit weiter eingeschränkt. Diese Verbesserung ist ein willkommener Nebeneffekt³, denn das wichtigste Ergebnis dieser Erweiterung ist die Bestimmung der Higgsmasse zu

$$m_{\text{H}} = 125,37_{-0,66}^{+0,73} \text{ GeV}.$$

Diese stimmt im Rahmen ihrer Unsicherheiten mit der offiziellen Messung der Higgsmasse von $m_{\text{H}} = (125,18 \pm 0,16) \text{ GeV}$ [17] aus dem Jahr 2018 und der Messung der Entdeckung³ im Jahr 2012 überein. Damit haben die Studenten im Rahmen des Versuches zwei Methoden zur Signifikanzbestimmung kennengelernt und erhalten zum eine Möglichkeit zur Massenbestimmung des SM-Higgs-Bosons. Nach der in Abschnitt 3.3 thematisierten Einordnung dieses Ergebnisses kann zwar noch nicht von einer Entdeckung gesprochen werden, da die Grenze von 5σ nicht überschritten wird. Jedoch ist die Grenze für ein Indiz (3σ) durch die im Rahmen dieser Studie bestimmte Unsicherheit auf die Signifikanz überschritten, wodurch eine Thematisierung der damit einhergehenden Fragestellungen im Rahmen eines Praktikumsversuches möglich ist.

3.5.2. Bewertung der Methoden

Die systematischen Unsicherheiten wurden im Rahmen dieser Arbeit vernachlässigt. Ihre Betrachtung würde zu einer Korrektur der Ergebnisse führen, jedoch auch einen großen Aufwand im Praktikumversuch darstellen. Aus diesem Grund wird der Einfluss der systematischen Unsicherheiten nicht untersucht.

Die Abschätzung der lokalen Signifikanz aus dem lokalen p_0 -Wert ist ein positiver Zusatz, jedoch im Rahmen dieser Arbeit nicht das Ziel des Verfahrens. Grund hierfür ist, dass, anders als bei der Entdeckung des SM-Higgs-Bosons [3], wo die lokale Signifikanz mithilfe des lokalen p_0 -Wertes bestimmt wurde, im Rahmen dieser Arbeit zu dem Verzicht von Störparametern eine vereinfachte Teststatistik (3.4) verwendet und systematische Effekte vernachlässigt wurden. Die in Abschnitt 3.3 genannte systematische Abweichung der Signalbreite ist ein weiterer Punkt, der den Wert der lokalen Signifikanz verfälscht, da schmalere Signale eine minimal schlechteren p_0 -Wert und somit eine leicht geringere Signifikanz liefern. Der Einfluss der genannten Punkte auf die Auswahl zwischen den unterschiedlichen Signal-MC ist jedoch vernachlässigbar klein, weshalb das Ziel einer Signal-MC Wahl durch das Verfahren sehr gut erreicht wird.

In die *profile-likelihood* Methode fließt ebenfalls im Rahmen dieser Arbeit keine Information über die Systematik ein, jedoch wird die Form des Signals im Vergleich zu der Parametrisierung in der p_0 -Wert Methode, deutlich besser beschrieben. Der nicht vernachlässigbare Unterschied der Skalierungsfaktoren $\sqrt{\chi_{\text{min}}^2}$ auf die Unsicherheiten von 1,29 (DSCB-Verteilung) und 7,77 (Gaußverteilung) führt dazu, dass für die Aussage der lokalen Signifikanz die *profile-likelihood* Methode besser geeignet ist, da in dieser die Signalverteilung deutlich besser beschrieben wird.

³Vergleiche mit der CMS Messung im Kapitel 1: $m_{\text{H}} = (125,3 \pm 0,4 \pm 0,5) \text{ GeV}$

4. Entwurf des Praktikumversuchs

In diesem Kapitel soll ein Konzept für einen Versuch vorgestellt werden, welcher im physikalischen Praktikum zum Einsatz kommen kann. Die Zielgruppe sind Studenten, die im Rahmen ihres Studiums bereits ein grundlegendes Verständnis für die statistischen Techniken zur Datenauswertung haben und einen vertrauten Umgang mit `PYTHON` besitzen. Im Rahmen dieses Versuches werden die Versuchsteilnehmer, außer dem Verständnis des Detektors und den Eigenschaften des Higgs-Bosons im Standard Modell die in der Teilchenphysik verwendete statistische Methoden kennenlernen, die auch in anderen Teilgebieten der Physik Anwendung finden.

Das notwendige Vorwissen für die Durchführung des Versuches wird in Form einer Vorbereitungsmappe zusammengefasst und den Versuchsteilnehmern für die Vorbereitung zur Verfügung gestellt. Die Erstellung dieser Vorbereitungsmappe und den exakten Aufgabenstellungen kann aufbauend auf dieser Arbeit erfolgen. Die Einstufung der Schwierigkeit der Aufgaben und die Umsetzung in dem Versuch muss an interessierten Versuchsteilnehmern getestet und zu einer sinnvollen Kombination zusammengefasst werden.

Die für diesen Versuch verwendete Dateistruktur lässt sich auf folgende sechs Skripte einschränken, die alle notwendigen Implementierungen enthalten.

```
Higgs_Suche
├─ class_collection.py
├─ filter_and_reconstruction.py
├─ reconstruction_plots.py
├─ MC_background_and_signal_fits.py
└─ statistical_analysis.py
```

4.1. Versuchsbeginn

Anknüpfend auf die im Kapitel 1 genannten Grundlagen sollen die Versuchsteilnehmer den CMS-Detektor kennenlernen und seine Funktionsweise nachvollziehen können. Das Verständnis der Funktionsweise des Detektors schafft eine Grundlage, auf welcher der Versuch aufbaut, da die vorhanden Messungen sowie manche der im nächsten Schritt angewendete Cuts maßgeblich von dem verwendeten Detektor abhängen. Die beste Grundlage hierzu bietet das iSpy-WebGL-Interface [70]. In diesem Interface können die wichtigsten Detektorkomponenten einzeln oder in Kombination angezeigt werden, wodurch ihre Rolle für eine Teilchendetektion erklärt werden kann. Zusätzlich können neben dem Detektoraufbau auch gemessene Ereignisse dargestellt werden, die sich im Rahmen des Versuches auf den Zerfall in vier Leptonen beschränkt. In Abbildung 4.1a und 4.1b sind beispielhaft die Zerfälle in vier Myonen bzw. vier Elektronen mit ausgewählten Detektorkomponenten dargestellt.

Damit eignet sich das Interface für ein einleitendes Kolloquium, in dem sichergestellt wird, dass die Versuchsteilnehmer das für das Verständnis notwendige Grundwissen haben.

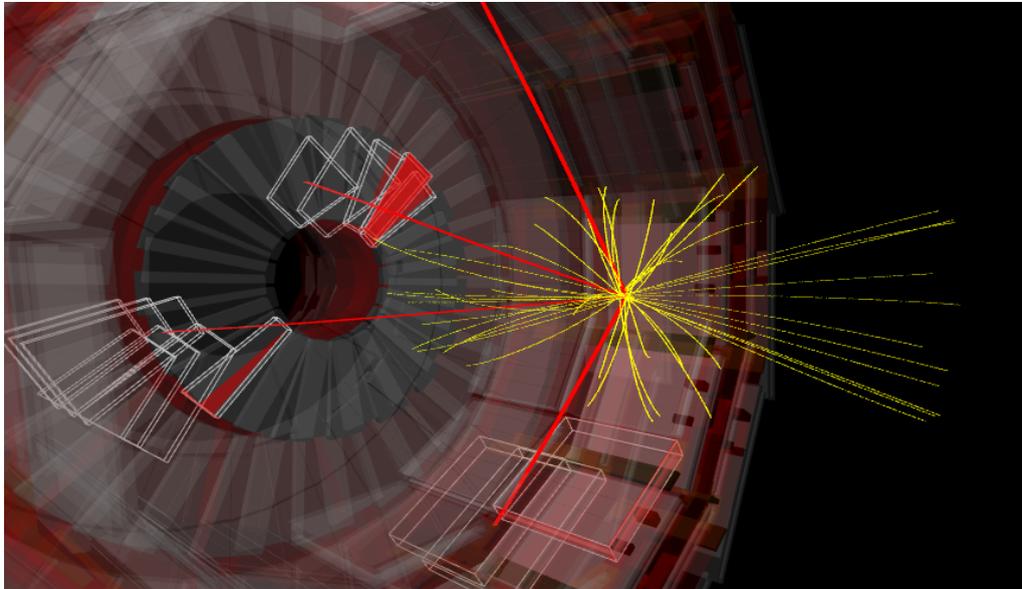
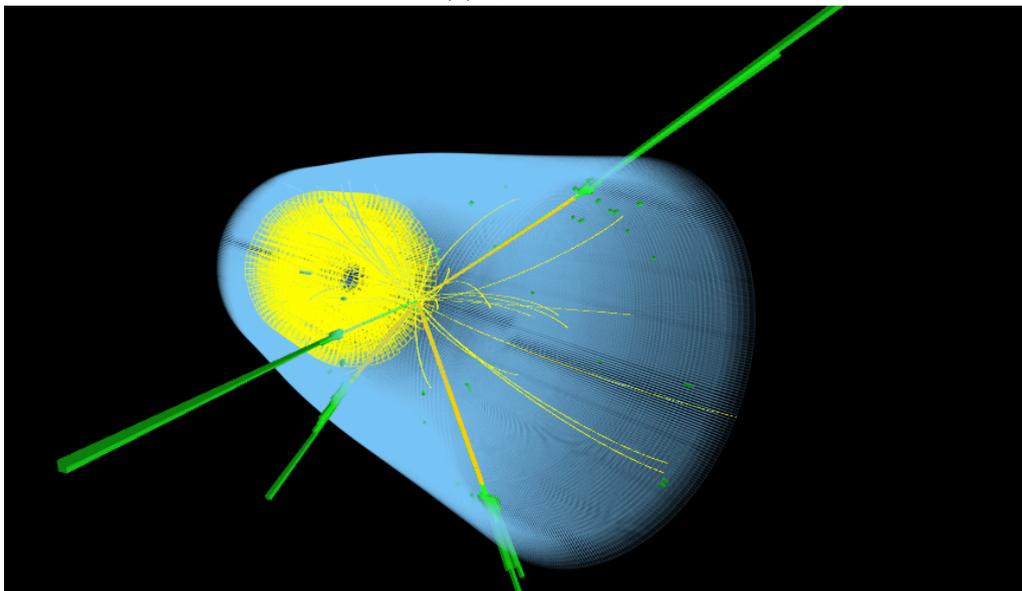
(a) 4μ -Zerfall(b) $4e$ -Zerfall

Abbildung 4.1.: Darstellung von Zerfällen, deren detektierte Endprodukte vier Myonen bzw. vier Elektronen enthalten, die alle an den Signalprozess gestellten Bedingungen erfüllen. In beiden Abbildungen sind die vom Tracker rekonstruierten Bahnen der nach der Kollision entstandenen Teilchen als gelbe Linien eingezeichnet. In dem 4μ -Zerfall sind die rekonstruierten globalen Myonen in rot eingezeichnet. Zusätzlich sind zum Teil die Driftröhren und Violdrahtkammern angedeutet. Die Komponenten, die die Myonen detektiert haben sind durch eine Umrandung hervorgehoben. Die Energie der Elektronen sind am Rand des schematisch dargestellten ECAL Barrels (blau) als grüne Balken eingezeichnet. Ein Teil des Trackers ist zudem in gelb schematisch im ECAL Barrel dargestellt.

Ebenso eignet es sich für die Darstellung der physikalischen Ereignisse aus denen die Versuchsteilnehmer die physikalischen Größen erkennen können.

4.2. Filter- und Rekonstruktionsschritt

In diesem Versuchsteil sollen die Versuchsteilnehmer in den später verwendeten Filtern vorkommenden Größen kennenlernen. Hierzu sollen relevante Größen wie die Invariante Masse, der Transversalimpuls oder die Überprüfung nach einer elektrisch neutralen Ladungskombination der Teilchen im Endzustand von Versuchsteilnehmern eigenständig implementiert werden. Dies geschieht in einem bereits existierenden Programmskelett, das um einzelne Funktionen ergänzt werden soll. Dieser befindet sich aktuell in der Datei `class_collection.py`.

Die in Kapitel 2 vorgeschlagene Reihenfolge der Filter- und Rekonstruktionsschritte kann von den Versuchsteilnehmern eigenständig verändert werden. Daraus entsteht die zusätzliche Aufgabe in `filter_and_reconstruction.py` eine sinnvolle Reihenfolge der möglichen Filter zu finden. In diesem Schritt können die Verteilungen ausgewählter Variablen, die in den Datensätzen enthalten sind, nach jedem Filterschritt dargestellt werden. Damit sollen die Versuchsteilnehmer zu jeder Zeit die Auswirkungen der jeweiligen Filterschritte nachvollziehen und die Funktionalität der eigenen Implementierung überprüfen zu können.

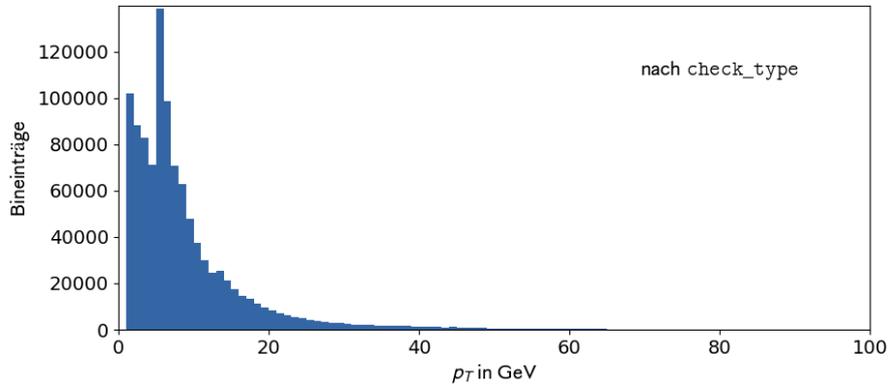
Die in Abbildung 4.2 dargestellte Kombination aus zwei Cuts auf den Transversalimpuls stellen beispielhaft den Ablauf dar, wie vorgegangen werden kann. Ausgehend von globalen Myonen wird die Verteilung der Transversalimpulse nach der Anwendung des Cuts auf den minimalen Transversalimpuls und der weiten Einschränkung des Transversalimpulses dargestellt (Abschnitt 2.2.1 und 2.2.2).

Die Darstellung der Verteilungen aller in den `.csv` Dateien enthaltenen Größen nach der vollständigen Rekonstruktion kann in `reconstruction_plot.py` durchgeführt werden. Nach der Festlegung der Selektionskriterien durch die Studenten sollen aus den verbleibenden Daten das Massenspektrum für die Messung und das Signal-MC, wie in Abbildung 2.4 bzw. Abbildung 2.3 dargestellt, bestimmt werden.

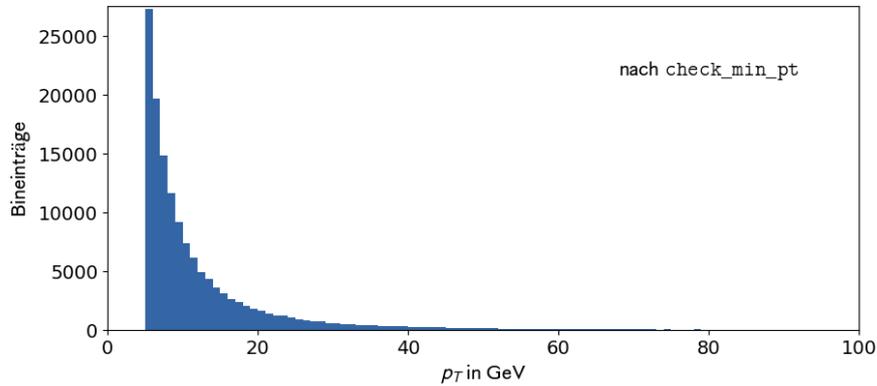
4.3. Analyseschritt

Nach der erfolgreichen Rekonstruktion der Massenspektren folgt dessen statistische Auswertung. Um den roten Faden in dem Analyseschritt nicht zu verlieren, ist die Durchführung der Signalbreitenabschätzung, so wie sie in Abschnitt 3.2 durchgeführt wird nicht sinnvoll. Stattdessen kann direkt die Gleichung (3.2) verwendet werden. Die Implementierung der Untergrund- und Signalstudie, sowie des p_0 -Wertes und der lokalen Signifikanz teilt sich in die für die Analyse notwendigen Klassen in `class_collection.py`, die wieder von den Studenten ergänzt werden können, die durchzuführenden Fits der MC-Simulationen in `MC_background_and_signal_fits.py` und die statistischen Tests in `statistical_analysis.py` auf.

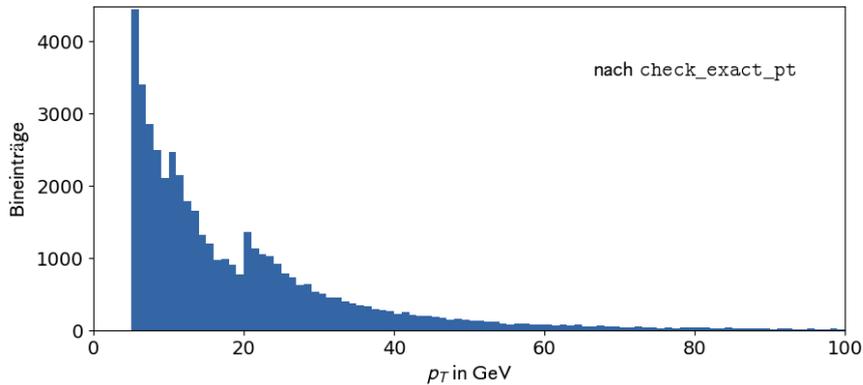
In diesem Abschnitt sollen die Versuchsteilnehmer vor allem das bereits bekannte Konzept von Likelihood Funktionen wiederholen. Mithilfe dieser soll der Likelihood-Quotienten-Test als Hypothesentest kennengelernt und für die Auswahl der richtigen Signal-MC verwendet werden. Damit lernen die Versuchsteilnehmer ein Verfahren kennen, mit dem sie nicht nur zwischen zwei, sondern beliebig vielen Hypothesen auf die am besten passende schließen können. Abschließend sollen die Versuchsteilnehmer mit der ihnen bereits bekannten *profile likelihood* Methode und der im Jahr 2012 von CMS aufgenommenen Daten auf die lokale Signifikanz der zuvor ausgewählten Signal-MC schließen.



(a) Verteilung vom Transversalimpuls im 4μ -Kanal, nach der Entfernung aller Myonen, die nicht als global (G) klassifiziert wurden.



(b) Entstehende Verteilung, nachdem alle Leptonen, die einen kleineren Transversalimpuls als 5 GeV haben entfernt wurden.



(c) Die resultierende Verteilung nach der Anwendung des kinematischen Filters, der sicherstellt, dass in einem Ereignis ein Lepton mit einem Transversalimpuls größer als 20 GeV und ein anderer mit einem Transversalimpuls größer als 10 GeV existiert. Diese Bedingung erklärt die zwei entstandenen Anstiege.

Abbildung 4.2.: Beispielhaft ist ein Ausschnitt aus dem Filterprozess dargestellt, der von Versuchsteilnehmern implementiert werden kann. Verwendet werden die Daten aus RunC mit 4μ im Endzustand.

Die eigenständige Implementierung der Likelihood Funktion kann entweder, wie in dieser Arbeit, im Versuch abgekapselt, oder in der Softwareumgebung `KAFE2` [51] durch die Versuchsteilnehmer implementiert werden¹. Die letztere Variante hat neben der Möglichkeit einer Weiterverwendung der Implementierung außerhalb des Versuches den Vorteil, dass die Versuchsteilnehmer sich mit der Softwareumgebung besser vertraut machen können. Die zusätzlich durchgeführte Unsicherheitenbetrachtung der bestimmten lokalen Signifikanzen kann im einleitenden Gespräch besprochen werden, durch die begrenzte Praktikumszeit aber nur schwer durchgeführt werden.

Im Rahmen dieser Arbeit wurden Messdaten aus dem CMS-Experiment verwendet. Für den abschließenden Teil des Versuches ist eine Kombination mit den Messergebnissen aus dem ATLAS-Experiment möglich. Die erreichte Signifikanz von über 3σ ist für die Durchführung des Versuches ausreichend. Das Einbringen einer zusätzlichen Messung kann aber zur Fragestellung über die Kombination mehrerer Messungen genutzt werden. Für eine Verwendung der Daten aus dem ATLAS-Experiment müssen die Ergebnisse aufgrund der zeitlichen Begrenzung des Praktikums von dem Betreuer außerhalb des Praktikums vorbereitet werden. Damit besteht die Möglichkeit die lokale Signifikanz weiter zu steigern und im besten Fall in den Bereich einer Entdeckung zu bringen.

4.4. Betrugsvermeidung im Rahmen des Praktikums

Für die Betrugsvermeidung im Praktikum muss eine Variation durch unterschiedliche Parameter möglich sein, wodurch die Versuchsteilnehmer individuelle Daten und Zwischen- sowie Endergebnisse erhalten können.

Zunächst kann eine Anpassung der Parameter im *skimming* außerhalb des Praktikums durchgeführt werden. Somit können die Gruppen von vorneherein unterschiedliche Datensätze bekommen, was eine Änderung in den betrachteten Verteilungen in den Zwischenschritten des Filterprozesses erzielt. Andere Parameter in den kinematischen Filtern können für die Variation der MC-Simulationen genutzt werden. Dies kann durch vorgegebene, von dieser Arbeit abweichende Werte, geschehen. Alternativ können Intervalle angegeben werden, aus denen die Versuchsteilnehmer im Rahmen ihrer Einschätzung die Filterparameter selbst bestimmen können. Zu strikte Cuts an dieser Stelle schaffen jedoch die Gefahr einer zu starken Reduktion der benutzten Datensätze, die zu größeren Ungenauigkeit führt, wodurch im Anschluss keine sinnvolle statistische Analyse durchführbar ist. Aus diesem Grund ist es wichtig diese Parameter an freiwilligen Testgruppen vorab zu bestimmen.

¹In `KAFE2` wird der negative Logarithmus der Likelihood Funktion als eine Kostenfunktion implementiert, die in der Anpassung durch die Variation der Parameter minimiert wird.

Zusammenfassung

Im Rahmen dieser Arbeit wurde ein funktionierendes Konzept für den Versuch zur Higgs-Boson Suche im 'goldenen' Kanal vorgestellt, der im physikalischen Praktikum zum Ende des Bachelor- beziehungsweise am Anfang des Masterstudienganges stattfindet oder auch im Rahmen des praktischen Teiles einer Veranstaltung zur experimentellen Teilchenphysik durchgeführt werden kann. Damit wird eine Orientierung bei der Spezialisierung im Physikstudium und der Einstieg in die Teilchenphysik erleichtert.

Die Verwendung der den Versuchsteilnehmern bekannten Programmiersprache PYTHON zusammen mit den Standardbibliotheken wie NUMPY, MATPLOTLIB, PANDAS, SCIPY oder öffentlich zugänglichen Softwareumgebungen wie KAFE2 [51] ermöglicht den Fokus auf die im Versuch vorgestellten und zur Durchführung angesetzter Grundkonzepte, wie sie in einer echten Analyse in der Teilchenphysik angewendet werden.

In diesem Versuch lernen die Teilnehmer den Umgang mit simulierten Daten und deren Verwendung in Kombination mit den gemessenen Daten. Im Laufe des Versuches wird ein grundlegendes Verfahren zur Selektion von gewünschten Daten aus einem größeren Datensatz vorgestellt, das das Ziel einer Steigerung der Sensitivität besitzt. Ohne das Verfahren sind aus einer statistische Analyse keine bedeutungsvollen Aussagen möglich.

Es wird die Anwendung eines Hypothesentests gezeigt und die Grundlagen dazu vorgestellt, mit denen die Versuchsteilnehmern aus unterschiedlichen Annahmen des Standardmodell Higgs-Bosons für verschiedene Massen diejenige mit der passenden Masse wählen. Mit der abschließenden Analyse der statistischen Signifikanz der von den Teilnehmern gewählten Hypothese werden alle Werkzeuge vorgestellt, die für grundlegende moderne Analysen unerlässlich und auch in Gebieten außerhalb der experimentellen Teilchenphysik anwendbar sind.

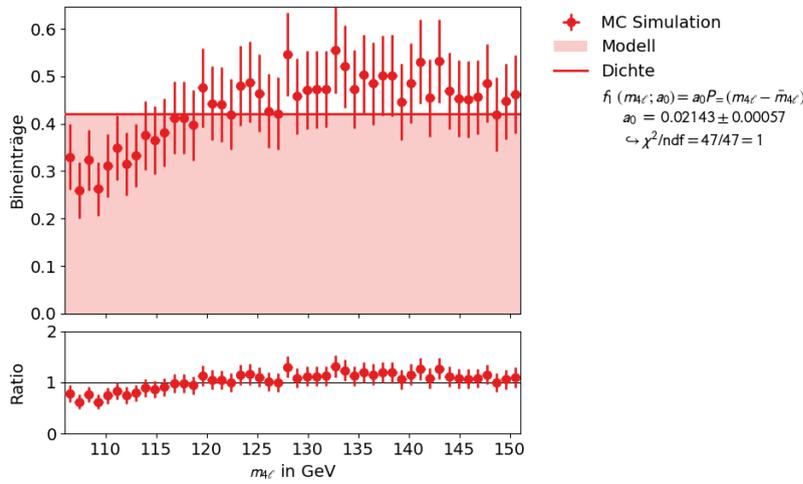
Mit den vorgestellten Methoden wurde basierend auf dem 2012 von CMS aufgezeichneten Datensatz die Existenz des Standardmodell Higgs-Bosons mit einer Signifikanz von $3,374^{+0,086}_{-0,048} \sigma$ nachgewiesen. Die Masse von $125,37^{+0,73}_{-0,66}$ GeV stimmt mit der offiziellen Messung innerhalb der bestimmten Unsicherheiten überein und gibt damit den Teilnehmern des Versuches die Möglichkeit transparent die aktuelle Forschungsergebnisse nachzuvollziehen.

Durch die weiter fortschreitende Veröffentlichung von den am CERN aufgezeichneten Daten kann die Analyse durch ihren modularen Aufbau problemlos angepasst und erweitert werden, sodass die Grenze für eine Entdeckung (5σ) überschritten werden kann. Ebenso steht die Möglichkeit offen, die Anwendung der vorgestellten Methoden auf andere Versuche, wie zum Beispiel die Analyse zum Z-Boson [71], zu erweitern oder Physik jenseits des Standardmodells durch das Einbringen neuer Hypothesen zu untersuchen.

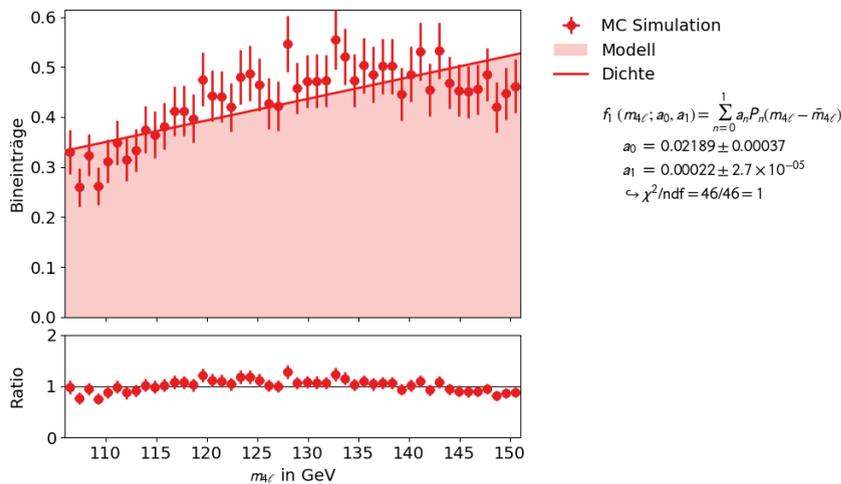
Anhang

A. Untergrundparametrisierung

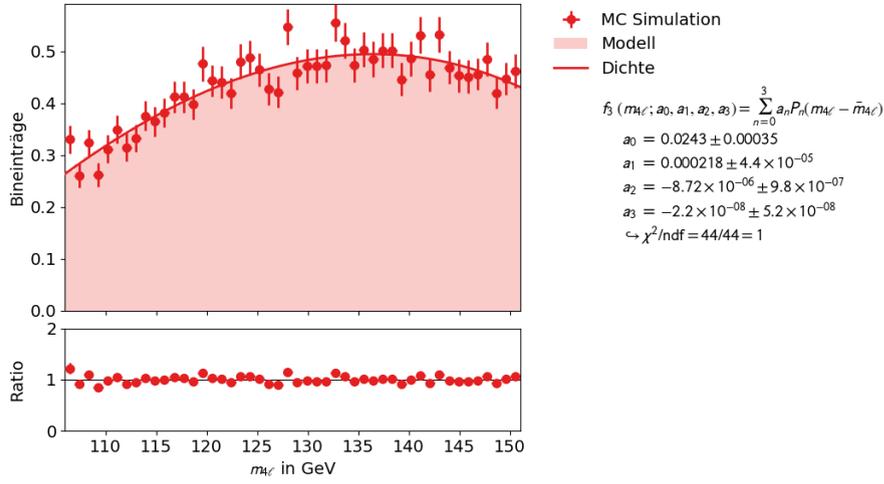
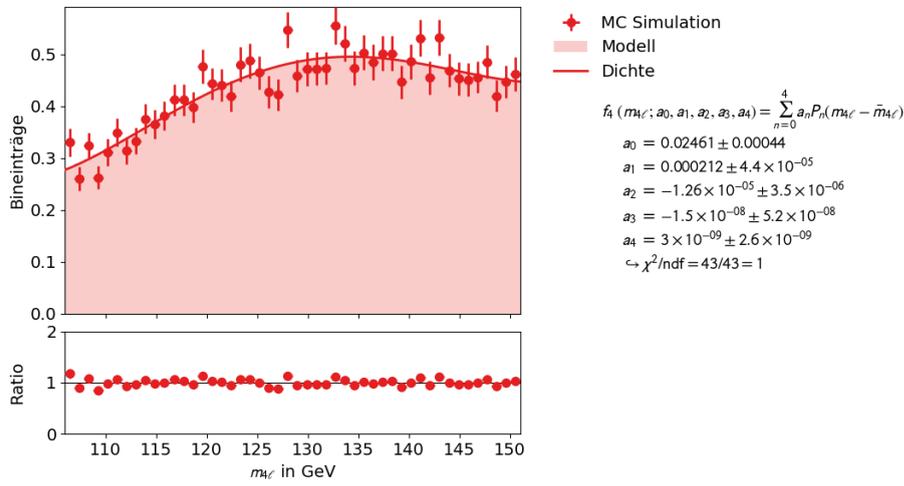
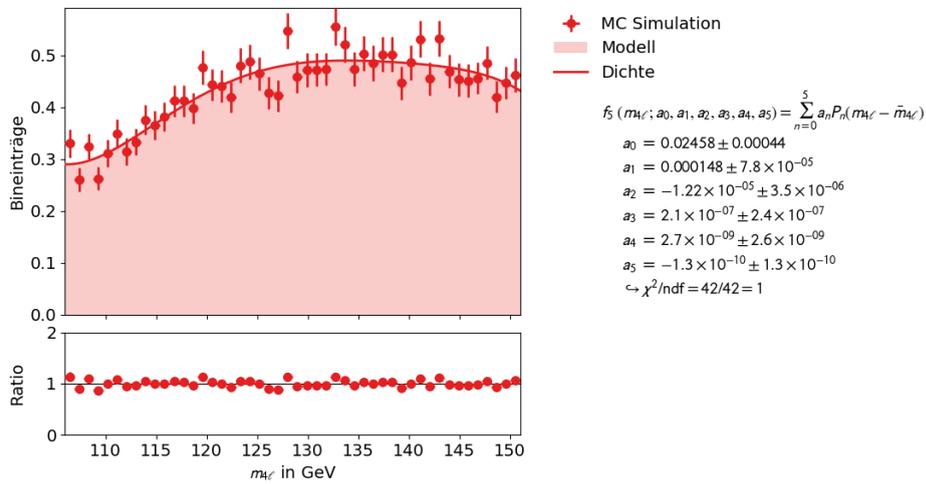
Aufgeführt sind die im Abschnitt 3.1 genannten Anpassungen des Untergrundes mittels Linearkombinationen von Legendrepolyomen unterschiedlichen Grades (Abbildung A.1). In Abbildung A.2a ist die Korrelation zwischen den Parametern dargestellt, die für die Anpassung des Untergrundes für eine Linearkombination von Legendrepolyomen bis zum zweiten Grad verwendet wurde und in Abbildung A.2b sind die Korrelationen zwischen den Parametern dargestellt die für ein einfaches nicht zentriertes Polynom zweiten Grades durch die Anpassung entstehen.

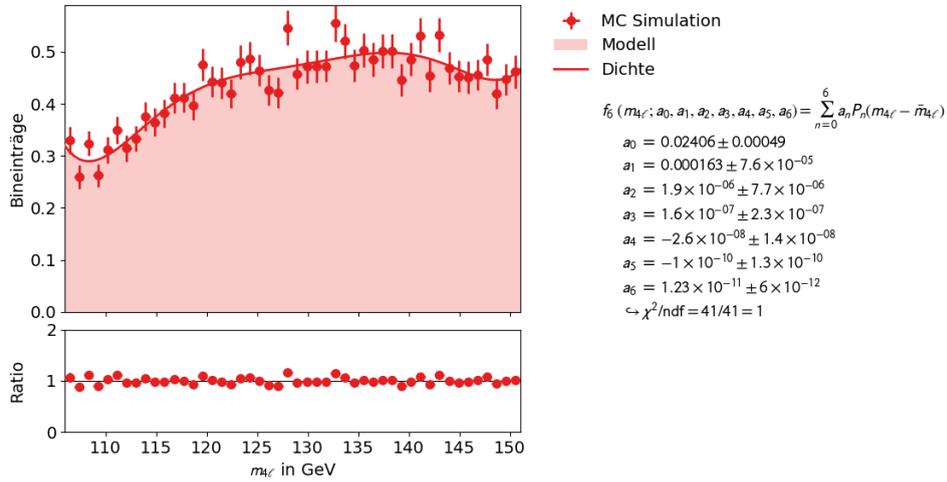


(a) $n = 0$, $\sqrt{\chi_{\min}^2} = 2,572$



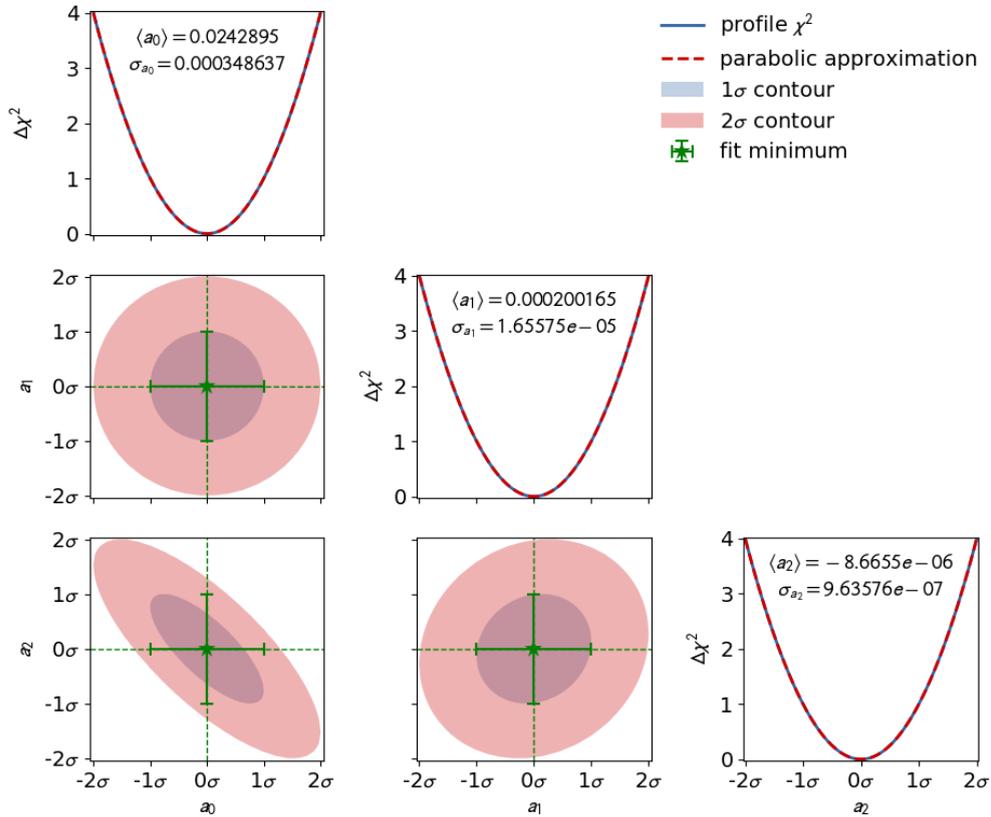
(b) $n = 1$, $\sqrt{\chi_{\min}^2} = 1,672$

(c) $n = 3$, $\sqrt{\chi_{\min}^2} = 1,024$ (d) $n = 4$, $\sqrt{\chi_{\min}^2} = 1,020$ (e) $n = 5$, $\sqrt{\chi_{\min}^2} = 1,020$



(f) $n = 6$, $\sqrt{\chi_{\min}^2} = 0,974$

Abbildung A.1.: Anpassungen an ein Histogramm der Untergrund MC durch verschiedene Linearkombinationen aus Legendrepolyomen bis zum Grad n mit der Softwareumgebung KAFE2 [51]. Das Ratio gibt das Verhältnis zwischen den für die Anpassung verwendeten Daten und der skalierten Wahrscheinlichkeitsdichte an. Die Unsicherheiten für die Anpassungen wurden mit den Faktoren $\sqrt{\chi_{\min}^2}$ multiplikativ skaliert, derart, das $\chi_{\text{ndf}}^2 = 1$ erfüllt ist. Diese sind in der Tabelle 2.2 zusammengefasst.



(a) $\sum_{n=0}^2 a_n P_n(m_{4\ell} - \bar{m}_{4\ell})$

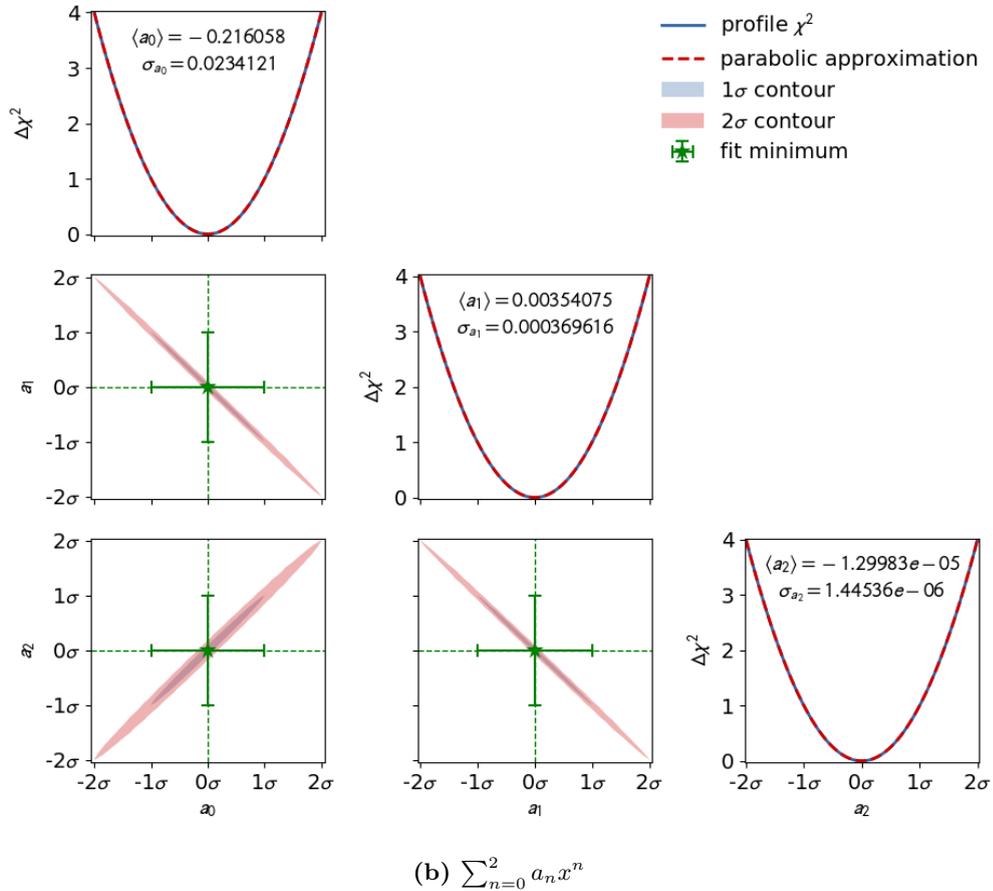
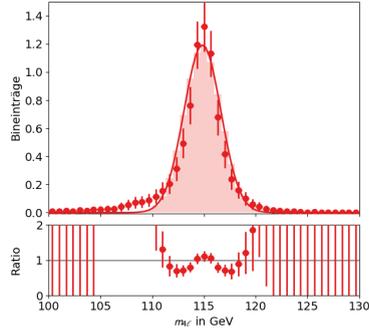
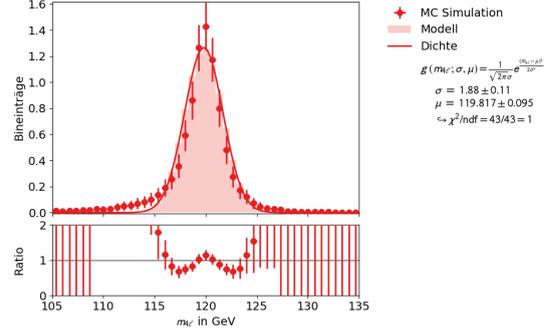
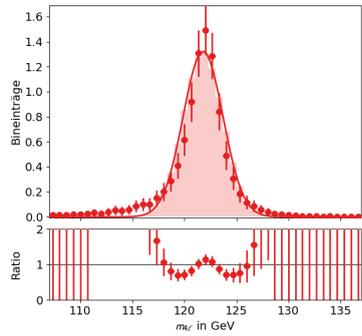
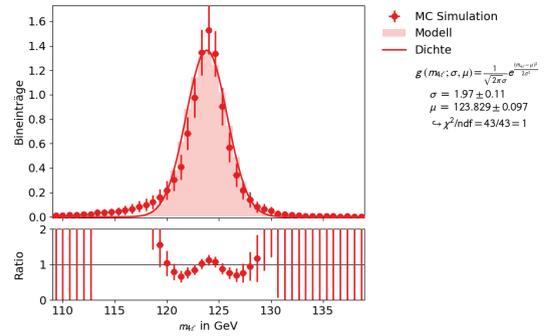
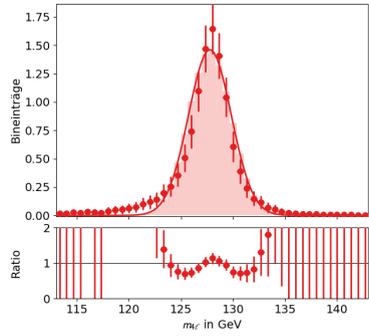
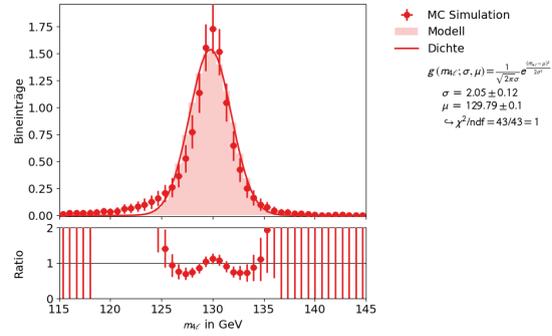


Abbildung A.2.: Darstellung der Korrelationen zwischen den Parametern für die Anpassung der Untergrund MC mittels eines nicht zentrierten Polynoms zweiten Grades (A.2b) und einer Linearkombination von zentrierten Legendrepolynomen bis zum zweiten Grad (A.2a).

B. Signalbreitenabschätzung - Signalverteilungsparametrisierung

In diesem Abschnitt findet sich die Anpassungen an die Signal-MC hinterlegt für weitere Massen des Higgs-Bosons. Es erfolgt nach Abschnitt 3.2 immer die Anpassung durch eine Gaußverteilung.

(a) $m_H = 115$ GeV(b) $m_H = 120$ GeV(c) $m_H = 122$ GeV(d) $m_H = 124$ GeV(e) $m_H = 128$ GeV(f) $m_H = 130$ GeV

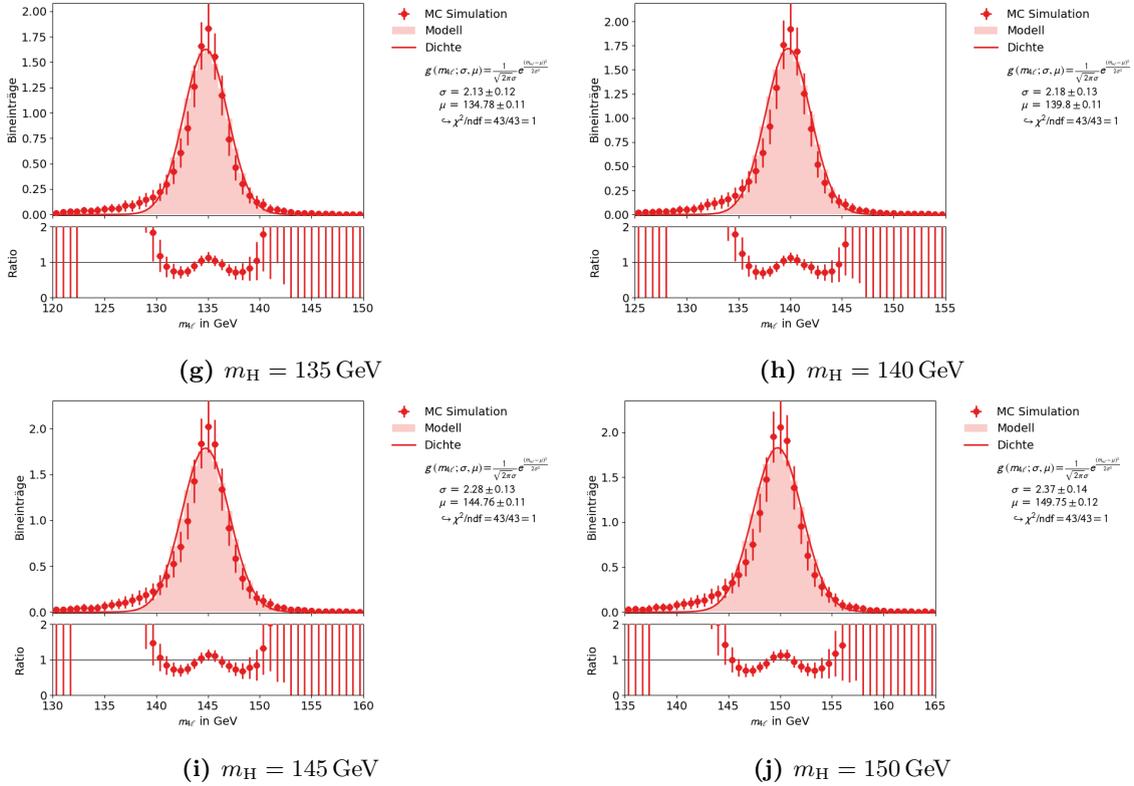


Abbildung B.3.: Anpassung an ein Histogramm der Signal MC unterschiedlicher Massen des Higgs-Bosons durch eine Gaußverteilung für die Signalbreitenabschätzung. Für die Anpassung wird die Softwareumgebung KAFE2 [51] verwendet. Die Signal-MC wurde an die vorhandene Messung skaliert. Das Ratio gibt hierbei das Verhältnis zwischen den skalierten Wahrscheinlichkeitsdichten und der für die Anpassung verwendeten Daten an. Die Unsicherheiten wurden mit dem Faktor $\sqrt{\chi_{\text{min}}^2}$ derart skaliert, dass $\chi_{\text{ndf}}^2 = 1$ erfüllt ist.

C. p_0 - Wertabschätzung für Histogramme mit wenig Statistik

Die im Verlauf der Arbeit verwendete Likelihood Methode für die Abschätzung des p_0 -Wertes und der Signalstärke gilt für Daten, die nicht durch ein Histogramm zusammengefasst wurden. Wird das Vorgehen dagegen mit einem Histogramm durchgeführt, so entsteht die in Abbildung C.4 aufgeführte Problematik: Für geringe Anzahl an Ereignissen verfälscht ein Histogramm die Verteilung, je nach Wahl der Binbreiten. Diese Verfälschung würde ähnliche Auswirkungen auf die Bestimmung des Signalanteils und der lokalen Signifikanz im Abschnitt 3.3 haben. Für große Anzahl an Ereignissen tritt diese Art von Verfälschung ebenfalls auf, ändert jedoch das Ergebnis nicht derart stark wie bei kleiner Anzahl an Ereignissen.

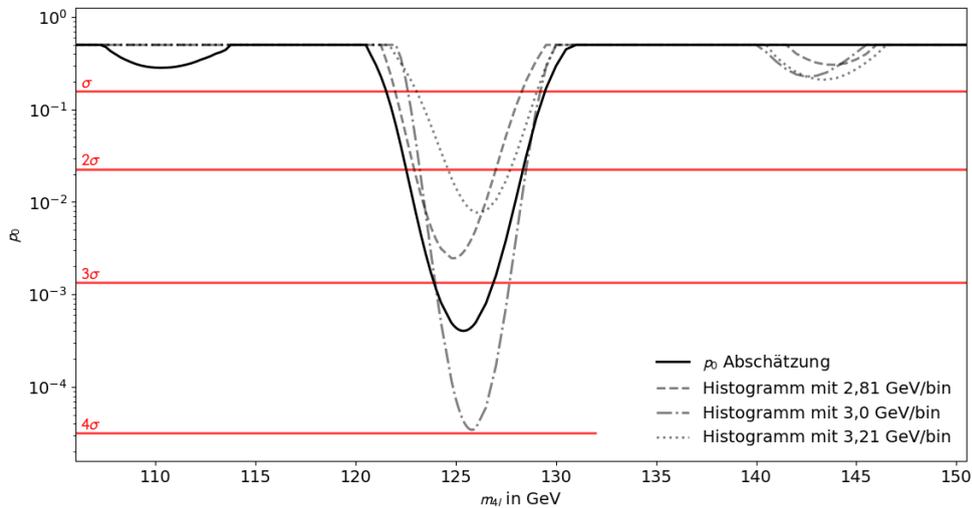
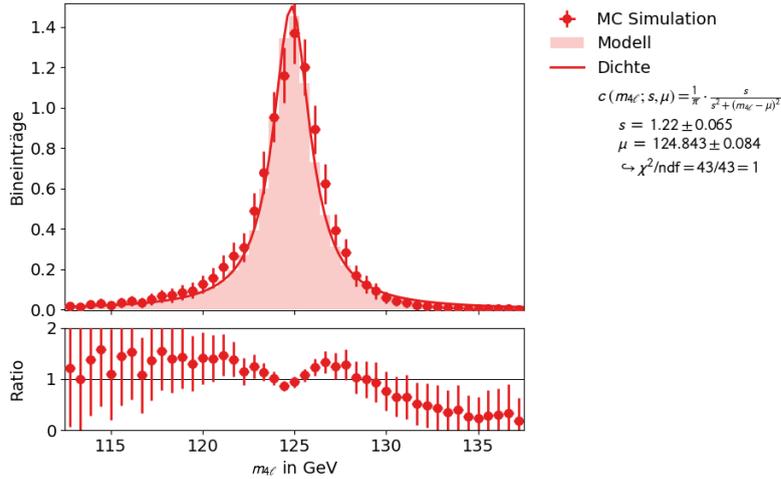


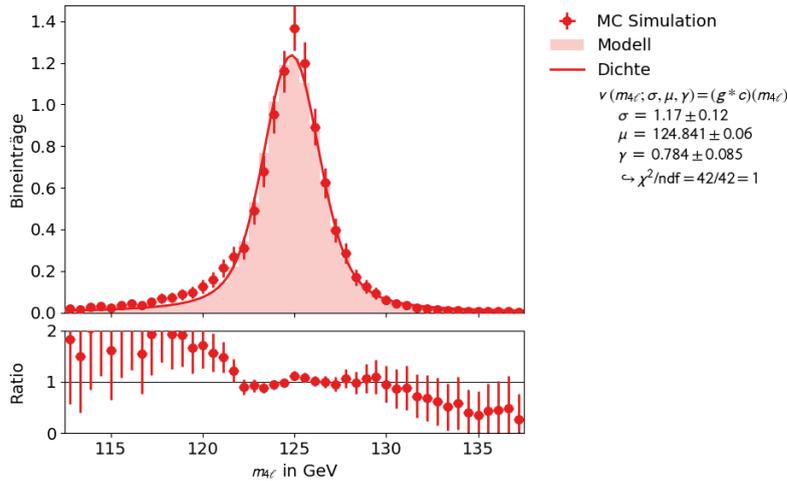
Abbildung C.4.: Abschätzung des p_0 -Wertes bei unterschiedlichen $m_{4\ell}$ Werten für Daten die nicht in einem Histogramm zusammengefasst sind (durchgezogene Linie), sowie sie in Kapitel 3 durchgeführt wurde und drei beispielhafte Histogramme mit unterschiedlichen Binbreiten.

D. Signalparametrisierung

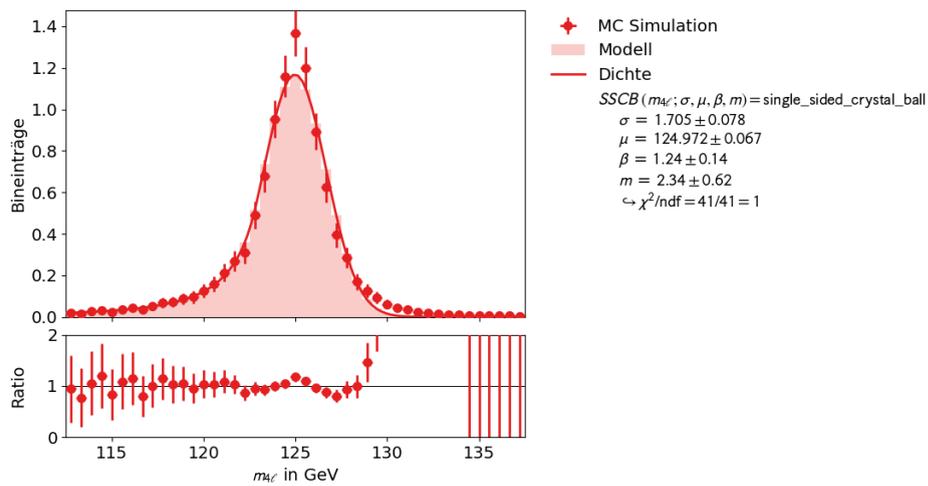
Aufgeführt sind die im Kapitel 3 genannten Anpassungen der Signal-MC für eine Masse des Higgs-Bosons von 125 GeV durch die in Tabelle 3.4 aufgelisteten Verteilungen. Die symmetrischen Verteilungen sind: Cauchy-Verteilung (Abbildung D.5a) und das Voigt-Profil (Abbildung D.5b). Die neben der DSCB untersuchte asymmetrische Verteilung ist die Single-Sided-Crystal-Ball-Verteilung (Abbildung D.5c).



(a) Cauchy-Verteilung, $\sqrt{\chi_{\min}^2} = 5,984$



(b) Voigt-Profil, $\sqrt{\chi_{\min}^2} = 4,399$



(c) SSCB-Verteilung, $\sqrt{\chi_{\min}^2} = 4,432$

Abbildung D.5.: Anpassungen an ein Histogramm der Signal-MC für eine Masse des Higgs-Bosons von 125 GeV. Die Anpassungen erfolgen in der Softwareumgebung KAFE2 [51]. Das Ratio gibt das Verhältnis der skalierten Wahrscheinlichkeitsdichten und der verwendeten Daten an. Die Unsicherheiten wurden für die Anpassungen mit dem jeweiligen Faktor $\sqrt{\chi_{\min}^2}$ multiplikativ skaliert und sind in Tabelle 3.4 aufgelistet.

Literaturverzeichnis

- [1] P.W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Physics Letters* 12.2 (1964), S. 132–133. ISSN: 0031-9163. DOI: [https://doi.org/10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9). URL: <http://www.sciencedirect.com/science/article/pii/0031916364911369>.
- [2] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Okt. 1964), S. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [3] CMS, Serguei Chatrchyan u. a. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett.* B716 (2012), S. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235.
- [4] The ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. 2012. arXiv: 1207.7214.
- [5] Bin Anuar Afiq Aizuddin Jomhari Nur Zulaiha Geiser Achim. *Higgs-to-four-lepton analysis example using 2011-2012 data*. 2017. DOI: 10.7483/OPENDATA.CMS.JKB8.D634. URL: <http://opendata.vern/record/5500>.
- [6] Christopher Zimmer. *Suche nach dem Higgs-Boson als Praktikumsversuch mit CERN open Data*. 6. Nov. 2018.
- [7] Bogdan Povh u. a. *Teilchen und Kerne - Eine Einführung in die physikalischen Konzepte*. 9. Aufl. Berlin Heidelberg New York: Springer-Verlag, 2013. ISBN: 978-3-642-37822-5.
- [8] Raghunath Sahoo. *Relativistic Kinematics*. 2016. arXiv: 1604.02651.
- [9] Werner Herr und B Muratori. “Concept of luminosity”. In: (2006). DOI: 10.5170/CERN-2006-002.361. URL: <https://cds.cern.ch/record/941318>.
- [10] Ramon Cid Manzano Xabier Cid Vidal. *Luminosity*. Aufgerufen am 03.12.2019. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.luminosity.
- [11] Fritz W. Bopp. *Kerne, Hadronen und Elementarteilchen - Eine kurze Einführung*. 2. Aufl. Berlin Heidelberg New York: Springer-Verlag, 2014. ISBN: 978-3-662-43667-7.
- [12] Qing-Hong Cao u. a. *Discovery and Identification of W' and Z' in $SU(2) \times SU(2) \times U(1)$ Models at the LHC*. 2012. arXiv: 1205.3769.
- [13] G. S. Guralnik, C. R. Hagen und T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), S. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [14] Peter W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Phys. Rev.* 145 (4 Mai 1966), S. 1156–1163. DOI: 10.1103/PhysRev.145.1156. URL: <https://link.aps.org/doi/10.1103/PhysRev.145.1156>.

-
- [15] Radja Boughezal. “Theoretical Status of Higgs Production at Hadron Colliders in the Standard Model”. In: *Particles and fields. Proceedings, Meeting of the Division of the American Physical Society, DPF 2009, Detroit, USA, July 26-31, 2009*. 2009. arXiv: 0908.3641.
- [16] The ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. 2012. arXiv: 1207.7214.
- [17] Particle Data Group, M. Tanabashi u. a. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), S. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [18] A. Denner u. a. *Standard Model Higgs-Boson Branching Ratios with Uncertainties*. 2011. arXiv: 1107.5909.
- [19] CMS, G. L. Bayatian u. a. “CMS Physics”. In: (2006).
- [20] S. Costantini u. a. “Uniformity and Stability of the CMS RPC Detector at the LHC”. In: *PoS RPC2012* (2012). [JINST8,P03017(2013)], S. 005. DOI: 10.22323/1.159.0005, 10.1088/1748-0221/8/03/P03017. arXiv: 1209.1989.
- [21] CMS Collaboration. *Measurement of the pp to ZZ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at $\sqrt{s} = 8$ TeV*. 2014. arXiv: 1406.0113.
- [22] Glen Cowan. *Statistical Data Analysis* -. New York: Oxford University Press, 1998. ISBN: 978-0-198-50156-5.
- [23] Roger (Arizona State University) Berger und George (University of Florida) Casella. *Statistical Inference* -. 2 ed. Thomson Learning, 2002. ISBN: 978-0-534-24312-8.
- [24] Roger Barlow. “Extended maximum likelihood”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 297.3 (1990), S. 496–506. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(90\)91334-8](https://doi.org/10.1016/0168-9002(90)91334-8). URL: <http://www.sciencedirect.com/science/article/pii/0168900290913348>.
- [25] “Search for the Standard Model Higgs boson at LEP”. In: *Physics Letters B* 565 (2003), S. 61–75. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(03\)00614-2](https://doi.org/10.1016/S0370-2693(03)00614-2). URL: <http://www.sciencedirect.com/science/article/pii/S0370269303006142>.
- [26] ALEPH Collaboration u. a. *Precision Electroweak Measurements and Constraints on the Standard Model*. 2010. arXiv: 1012.2367.
- [27] R. Brun und F. Rademakers. “ROOT: An object oriented data analysis framework”. In: *Nucl. Instrum. Meth.* A389 (1997), S. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.
- [28] *CMS Offline Software*. URL: <https://github.com/cms-sw/cmssw>.
- [29] The CMS Collaboration. *DoubleMuParked primary dataset in AOD format from Run of 2012 (/DoubleMuParked/Run2012A-22Jan2013-v1/AOD)*. URL: <https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FDoubleMuParked%5C%2FRun2012A-22Jan2013-v1%5C%2FAOD&instance=prod/global>.
- [30] The CMS Collaboration. *DoubleMuParked primary dataset in AOD format from Run of 2012 (/DoubleMuParked/Run2012B-22Jan2013-v1/AOD)*. 2017. DOI: 10.7483/OPENDATA.CMS.YLIC.86ZZ. URL: <http://opendata.cern.ch/record/6004>.
- [31] The CMS Collaboration. *DoubleMuParked primary dataset in AOD format from Run of 2012 (/DoubleMuParked/Run2012C-22Jan2013-v1/AOD)*. 2017. DOI: 10.7483/OPENDATA.CMS.YLIC.86ZZ. URL: <http://opendata.cern.ch/record/6030>.

- [32] The CMS Collaboration. *DoubleMuParked primary dataset in AOD format from Run of 2012 (/DoubleMuParked/Run2012D-22Jan2013-v1/AOD)*. URL: <https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FDoubleMuParked%5C%2FRun2012D-22Jan2013-v1%5C%2FAOD&instance=prod/global>.
- [33] The CMS Collaboration. *DoubleElectron primary dataset in AOD format from Run of 2012 (/DoubleElectron/Run2012A-22Jan2013-v1/AOD)*. URL: <https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FDoubleElectron%5C%2FRun2012A-22Jan2013-v1%5C%2FAOD&instance=prod/global>.
- [34] The CMS Collaboration. *DoubleElectron primary dataset in AOD format from Run of 2012 (/DoubleElectron/Run2012B-22Jan2013-v1/AOD)*. 2017. DOI: 10.7483/OPENDATA.CMS.SOH8.LBD3. URL: <http://opendata.cern.ch/record/6003>.
- [35] The CMS Collaboration. *DoubleElectron primary dataset in AOD format from Run of 2012 (/DoubleElectron/Run2012C-22Jan2013-v1/AOD)*. 2017. DOI: 10.7483/OPENDATA.CMS.SINM.BV86. URL: <http://opendata.cern.ch/record/6029>.
- [36] The CMS Collaboration. *DoubleElectron primary dataset in AOD format from Run of 2012 (/DoubleElectron/Run2012D-22Jan2013-v1/AOD)*. URL: <https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FDoubleElectron%5C%2FRun2012D-22Jan2013-v1%5C%2FAOD&instance=prod/global>.
- [37] The CMS Collaboration. *Simulated dataset ZZTo4mu_8TeV-powheg-pythia6 in AODSIM format for 2012 collision data*. 2017. DOI: 10.7483/OPENDATA.CMS.HJ1F.5U9R. URL: <http://opendata.cern.ch/record/10071>.
- [38] The CMS Collaboration. *Simulated dataset ZZTo4e_8TeV-powheg-pythia6 in AODSIM format for 2012 collision data*. 2017. DOI: 10.7483/OPENDATA.CMS.PEOD.ZZJY. URL: <http://opendata.cern.ch/record/10065>.
- [39] The CMS Collaboration. *Simulated dataset ZZTo2e2mu_8TeV-powheg-pythia6 in AODSIM format for 2012 collision data*. 2017. DOI: 10.7483/OPENDATA.CMS.461G.HELP. URL: <http://opendata.cern.ch/record/10054>.
- [40] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. 2017. DOI: 10.7483/OPENDATA.CMS.G13X.TDSB. URL: <http://opendata.cern.ch/record/9356>.
- [41] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-115_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-115_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [42] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-120_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-120_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [43] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-122_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-122_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.

-
- [44] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-124_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-124_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [45] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-128_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-128_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [46] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-130_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-130_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [47] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-135_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-135_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [48] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-140_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-140_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [49] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-145_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-145_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [50] The CMS Collaboration. *Simulated dataset SMHiggsToZZTo4L_M-150_8TeV-powheg15-JHUGenV3-pythia6 in AODSIM format for 2012 collision data*. URL: https://cmsweb.cern.ch/das/request?input=dataset%5C%3D%5C%2FSMHiggsToZZTo4L_M-150_8TeV-powheg15-JHUGenV3-pythia6%5C%2FSummer12_DR53X-PU_S10_START53_V19-v1%5C%2FAODSIM&instance=prod/global.
- [51] Daniel Sävöiu. *kafe2 - Karlsruhe Fit Environment 2*. <https://github.com/dsavoiu/kafe2>. 2018 - 2019.
- [52] *CMS list of validated runs for primary datasets of 2012 data taking*. 2017. DOI: DOI: 10.7483/OPENDATA.CMS.CO0V.SE32. URL: <http://opendata.cern.ch/record/1002>.
- [53] The CMS Collaboration. *The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC*. 2013. arXiv: 1306.6905.
- [54] CMS Collaboration. *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*. 2015. arXiv: 1502.02701.
- [55] Christopher Brust u. a. *Identifying boosted new physics with non-isolated leptons*. 2014. arXiv: 1410.0362.
- [56] CMS Collaboration. *Measurement of the properties of a Higgs boson in the four-lepton final state*. 2013. arXiv: 1312.5353.

- [57] Ian James Anderson. “A Tale of Two Vertices: Production and Decay of the Higgs VV Vertex at the LHC”. Presented 11 May 2015. Juli 2015. URL: <https://cds.cern.ch/record/2034261>.
- [58] CMS Collaboration. *Measurement of the pp to ZZ production cross section and constraints on anomalous triple gauge couplings in four-lepton final states at sqrt(s) = 8 TeV*. 2014. arXiv: 1406.0113.
- [59] CMS Collaboration. *Measurement of W+W- and ZZ production cross sections in pp collisions at sqrt(s) = 8 TeV*. 2013. arXiv: 1301.4698.
- [60] CMS Collaboration. *Measurement of the ZZ production cross section and search for anomalous couplings in 2l2l' final states in pp collisions at sqrt(s) = 7 TeV*. 2012. arXiv: 1211.4890.
- [61] CMS Collaboration. *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV*. 2015. arXiv: 1502.02701.
- [62] The CMS Collaboration. *The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC*. 2013. arXiv: 1306.6905.
- [63] *Encyclopaedia of Mathematics - Supplement*. Berlin Heidelberg: Springer Science Business Media, 2012. ISBN: 978-9-401-51288-6.
- [64] Glen Cowan u. a. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: <http://dx.doi.org/10.1140/epjc/s10052-011-1554-0>.
- [65] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*. Techn. Ber. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011. URL: <https://cds.cern.ch/record/1379837>.
- [66] Louis Lyons. “Open statistical issues in Particle Physics”. In: *The Annals of Applied Statistics* 2.3 (Sep. 2008), S. 887–915. ISSN: 1932-6157. DOI: 10.1214/08-aos163. URL: <http://dx.doi.org/10.1214/08-AOS163>.
- [67] ATLAS Collaboration. *Search for resonances in diphoton events at sqrt(s)=13 TeV with the ATLAS detector*. 2016. arXiv: 1606.03833.
- [68] Tomasz Skwarnicki. “A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances”. Diss. Cracow, INP, 1986. URL: <http://www-library.desy.de/cgi-bin/showprep.pl?DESY-F31-86-02>.
- [69] D.A. Sprott. *Statistical Inference in Science* -. 2000. Aufl. Berlin Heidelberg: Springer Science Business Media, 2008. ISBN: 978-0-387-22766-5.
- [70] T McCauley. “A browser-based event display for the CMS Experiment at the LHC using WebGL”. In: *Journal of Physics: Conference Series* 898 (Okt. 2017), S. 072030. DOI: 10.1088/1742-6596/898/7/072030. URL: <https://doi.org/10.1088/1742-6596/898/7/072030>.
- [71] Tabea Feßenbecker. *Entwicklung eines Praktikumversuches zur Analyse von Myonen-zuständen in LHC-Daten*. 15. März 2017.