

Besondere Lernleistung

im Fachbereich Physik

Analyse von ATLAS-Experiment-Daten des Higgs-Boson-Zerfalls in zwei Z-Bosonen mittels Peakanalyse und Neuronalen Netzen

Erik Thyzel

geboren am 25.04.2005 in Dresden

Martin-Andersen-Nexö-Gymnasium
in Zusammenarbeit mit
TECHNISCHE UNIVERSITÄT DRESDEN
INSTITUT FÜR KERN- UND TEILCHENPHYSIK

21.12.2022

Eingereicht am 21. Dezember 2022

Externer Betreuer: Max Vincent Stange (TU Dresden)

Betreuender Lehrer: Karen Opitz

Kurzfassung

Diese Arbeit beschäftigt sich mit der Grundlagenforschung im Gebiet der Teilchenphysik, deren Ziel es ist, Eigenschaften über die Elementarteilchen und ihre Wechselwirkungen zu ermitteln. Thema der Arbeit ist dabei die Untersuchung des Zerfalls von Z- und Higgs-Bosonen in Teilchenkollisionen. Aus der Auswertung von seit 2020 frei verfügbaren Daten über Teilchenkollisionen aus dem ATLAS-Detektor am CERN wurden Eigenschaften dieser Wechselwirkungsteilchen ermittelt und schließlich das Higgs-Boson nachgewiesen.

Anhand der Analyse des Z-Bosons wurden am Anfang der Arbeit grundsätzliche Auswertungsmethoden erarbeitet und angewandt. Dabei wurden geeigneten Selektionskriterien festgelegt, um den Anteil der für die Analyse relevanten Ereignisse zu erhöhen. Danach wurde die Peakanalyse des Histogramms über die invariante Masse des Z-Bosons mittels verschiedener Fitting-Methoden und Funktionen durchgeführt. Daraus hat sich ergeben, dass die Voigt-Verteilung die Häufigkeiten der invariante Masse im Experiment am besten beschreibt. Neben der invarianten Masse des Z-Bosons wurde außerdem seine Zerfallsbreite und Lebensdauer mit guter Genauigkeit bestimmt.

Nach der Übertragung und Erweiterung der Konzepte auf die komplexere Analyse des Higgs-Bosons ergaben sich auch hier gute Annäherungen für die Werte der Masse des Z- und Higgs-Bosons. Anschließend wurde anhand dieser Daten das Vorgehen zum Nachweis des Higgs-Bosons verdeutlicht. Außerdem wurde gezeigt, dass die vorhandene Menge an Daten für eine komplexe Analyse und den Nachweis des Higgs-Bosons nicht ausreichend ist und es wurden selbst zusätzliche, zufällige Analysedaten erzeugt.

Um das Higgs-Boson auch ohne Skalierung der Daten nachzuweisen, müssen relevante (Signal-daten) von irrelevanten Daten (Messuntergrunddaten) besser getrennt werden. Eine Variante, um dies zu erreichen, ist die Anwendung von Neuronalen Netzen für die Selektion der Daten. Daher wurde im letzten Teil dieser Arbeit ein Neuronales Netz programmiert, optimiert und schließlich trainiert. Nach zahlreichen Veränderungen konnte so eine statistische Signifikanz von $6,48\sigma$ berechnet werden. Da man in der Teilchenphysik ab 5σ von einer „Entdeckung“ spricht, wurde das Ziel des zweiten Teils der Arbeit, der Nachweis des Higgs-Boson, erreicht.

Abstract

Basic research in the field of particle physics explores what fundamental particles matter consists of and how they interact. In this study data from ATLAS-Detector in the LHC was analysed to determine characteristics of the Z and Higgs boson. Finally, the Higgs boson was confirmed. Analysing the decay of Z bosons, the events measured in the detector were separated by selected criteria. Then the histogram of the invariant mass of the Z boson was analysed with several fitting-functions, which led to the conclusion that the optimal function to describe the measured distribution is the Voigt- distribution. Besides the mass of the Z boson also the decay width and lifetime could be determined with good accuracy. Using similar concepts for the analysis of the Higgs boson, the value for the mass of this particle was also determined with good accuracy. Furthermore, it was proven that the available data was not sufficient for a complex analysis and additional data needed to be generated. With these, the concept of confirming the Higgs boson was explained. While in the first part of this study it was already reached a high accuracy in the determination of characteristics of fundamental particles, the aim of the second part was to increase the statistical significance without creating additional data. For this, an artificial neural network was programmed, optimized and trained. Thus, a statistical significance of $6,48\sigma$ was reached, which in particle physics is considered a proof of the higgs boson.

Inhaltsverzeichnis

1	Einleitung	6
2	Das Standardmodell der Teilchenphysik	7
2.1	Ladungen und Wechselwirkungen	7
2.2	Das Ordnungsschema der Materieteilchen	9
2.3	Das Konzept der Wechselwirkungsteilchen	9
2.4	Feynman-Diagramme und Teilchenumwandlungen	10
2.5	Das Z-Boson - Botenteilchen der schwachen Wechselwirkung	11
2.6	Natürliche Einheiten und das Schwerpunktsystem	11
3	Experimenteller Aufbau	12
3.1	Die Teilchenbeschleuniger am CERN	12
3.2	Der ATLAS-Detektor	13
4	Statistik	14
4.1	Die physikalische Wahrscheinlichkeit	14
4.2	Wahrscheinlichkeitsdichten	14
4.2.1	Definition von Wahrscheinlichkeitsdichten	15
4.2.2	Die Normalverteilung	15
4.2.3	Die Chi-Quadrat-Verteilung	16
4.3	Fitting	17
4.3.1	Kurvenfitting als Auswertungsmethode	17
4.3.2	Minimum-Chiquadrat-Fit	18
4.4	Statistische Signifikanz	19
5	Neuronale Netze	20
5.1	Aufbau Neuronaler Netze	20
5.2	Training Neuronaler Netze	22
6	Z-Boson-Analyse	23
6.1	Signal und Messuntergrund	23
6.2	Der Zerfall des Z-Bosons	23
6.3	Auswahl der Selektionskriterien	24
6.4	Die invariante Masse und die M_{ll} -Verteilung	25
6.5	Fitting der Normalverteilung als Wahrscheinlichkeitsdichtefunktion	26
6.5.1	Auswahl der Fitting-Funktion	27
6.5.2	Schätzung der Fitting-Parameter	28
6.5.3	Chi-Quadrat-Test	28
6.5.4	Anpassung des Fittings	29
6.5.5	Ergebnisse und Auswertung der Analyse mit Normalverteilung	30
6.6	Fitting anderer Wahrscheinlichkeitsdichtefunktionen	32
6.6.1	Die Breit-Wigner-Verteilung	32
6.6.2	Die Voigt-Verteilung	34
6.7	Endergebnisse und Auswertung der kompletten Analyse	36

7	Higgs-Boson-Analyse	37
7.1	Der Zerfall des Higgs-Bosons und die Ereignisanalyse	37
7.2	Analyse der beiden Z-Peaks	38
7.3	Analyse des Histogramms der invarianten 4-Leptonen-Masse	40
7.3.1	Fitting des Z-Boson-Peaks im Histogramm der invarianten 4-Leptonen-Masse	41
7.3.2	Skalierung der Daten	42
7.3.3	Fitting des Higgs-Boson-Peaks	43
7.3.4	Nachweis des Higgs-Bosons	45
7.4	Endergebnisse und Auswertung der kompletten Analyse	46
8	Analyse mittels Neuronaler Netze	47
8.1	Einrichtung des Neuronalen Netzes	47
8.2	Normalisierung	49
8.3	Fitting drei einfacher Neuronaler Netze	49
8.4	Auswertungsmethoden und Histogramme	50
8.5	Ereignisgewichte	53
8.6	Overtraining	55
8.7	Hyperparameter-Optimierung	57
8.7.1	Erster Optimierungsschritt	57
8.7.2	Zweiter Optimierungsschritt	59
8.7.3	Dritter Optimierungsschritt	60
8.7.4	Test des ausgewählten Modells	62
8.8	Anwendung des Neuronalen Netzes und Signifikanzberechnung	63
8.9	Auswertung der Analyse durch Neuronale Netze	65
9	Zusammenfassung und Ausblick	66
	Literaturverzeichnis	68
	Anhang	I

Kapitel 1

Einleitung

Ziel der Naturwissenschaft Physik ist es im Allgemeinen, in der Natur beobachtete Vorgänge auf allgemeine Gesetze zurückzuführen. Dazu werden in der theoretischen Physik Modelle aufgestellt, die dann in der Praxis durch Experimente geprüft werden. Das Teilgebiet der Physik, mit der sich diese Arbeit beschäftigt, ist die Elementarteilchenphysik, die den Aufbau der Materie und ihre Wechselwirkungen im Kleinsten untersucht.

Das Standardmodell der Teilchenphysik bildet das komplexe Grundgerüst der Teilchenphysik und ist in der Lage, Elementarteilchen und ihre Wechselwirkungen zu beschreiben. Experimentell können in der Teilchenphysik aufgestellte Vermutungen und Modelle in Teilchenbeschleunigern und -detektoren überprüft werden, in denen Teilchen mit hoher Geschwindigkeit aufeinandertreffen und die Kollisionsprodukte untersucht werden. Mittels dieser Methode wurden schon viele Elementarteilchen unserer Materie entdeckt. Eine der bedeutendsten Entdeckungen stellt dabei der Nachweis des Higgs-Bosons im Jahr 2012 dar. Schon 1964 wurde von den Physikern Robert Brout, François Englert und Peter Higgs der nach ihnen benannte Brout-Englert-Higgs-Mechanismus vorhergesagt. Dieser erklärte, warum Teilchen, die nach dem Standardmodell der Teilchenphysik masselos sein müssen, sich in der Realität wie Teilchen mit einer Masse verhalten. Die Masse der Teilchen entsteht demnach durch die Wechselwirkung mit einem allgegenwärtigen Feld, dem Brout-Englert-Higgs-Feld. Zu diesem Feld gehört das Higgs-Teilchen, über welches das Feld und der Mechanismus erst im Juli 2012 am internationalen Forschungszentrum CERN nachgewiesen werden konnte. [1]

Trotz vieler Entdeckungen gibt es noch viele offene Fragen, zum Beispiel in Bezug auf Antimaterie, dunkle Materie und die Entstehung unseres Universums, die Gegenstand aktueller Forschung sind. Hier kann das Higgs-Boson ein Schlüssel zum Verständnis dieser Fragen sein, weshalb es auch heute noch untersucht wird. So versucht man beispielsweise seine Eigenschaften immer genauer zu vermessen.

Auch diese Arbeit beschäftigt sich mit der Grundlagenforschung der Elementarteilchenphysik, in der es darum geht, die Eigenschaften elementarer Teilchen zu ermitteln. Dabei wurden frei verfügbare Daten aus dem ATLAS-Detektor, einem Teilchendetektor am CERN ausgewertet. Zuerst wurde die Analyse von Messdaten aus Teilchenkollisionen beispielhaft am Z-Boson, einem Wechselwirkungsteilchen der schwachen Wechselwirkung, durchgeführt und Eigenschaften wie dessen Masse oder Lebensdauer bestimmt. Dazu wurden physikalische Selektionskriterien und mathematische Optimierungsfunktionen verwendet. Durch die Peakanalyse mit verschiedenen Funktionen wurden außerdem statistische Effekte des Experiments berücksichtigt und quantitativ bestimmt. Anschließend wurden die erschlossenen Konzepte auf die komplexere Analyse des Higgs-Bosons übertragen und erweitert. Auch dabei konnten einige seiner Eigenschaften ermittelt werden und abschließend wurde ein Nachweis des Higgs-Bosons anhand der verfügbaren Daten angestrebt.

Da die frei zur Verfügung stehenden Daten nicht ausreichten, um das Higgs-Boson mit genügender Sicherheit nachzuweisen, wurden anschließend zur Selektion der Daten Neuronale Netze angewandt, eine Methode, die heute in der Teilchenphysik sehr beliebt ist. Dazu wurde hier das Neuronale Netz darauf trainiert, Ereignisse, in denen ein Higgs-Boson zerfällt, von anderen uninteressanten Ereignissen zu unterscheiden. Da das Neuronale Netz viel komplexer als klassische Selektionskriterien ist, können so effektiver seltene Ereignisse wie der Zerfall des Higgs-Bosons identifiziert werden.

Kapitel 2

Das Standardmodell der Teilchenphysik

Das Standardmodell der Teilchenphysik bildet das theoretische Fundament für die moderne Forschung der Elementarteilchenphysik. Daher beschäftigt sich das trotz seines Namens sehr komplexe Standardmodell mit den Grundthemen der Teilchenphysik: Aus welchen elementaren Bausteinen ist unsere Materie aufgebaut, welche Eigenschaften besitzen diese und wie können diese Teilchen miteinander in Wechselwirkung treten?

2.1 Ladungen und Wechselwirkungen

Eine entscheidende Eigenschaft eines Teilchens ist dessen Ladung. Jedes Teilchen und jeder Zusammenschluss von Teilchen hat diese Eigenschaft. Die Ladung ist eine Eigenschaft, die ein Teilchen charakterisiert und somit von anderen unterscheidet. Außerdem befähigt die Ladung eines Teilchens es zu einer Wechselwirkung mit anderen Teilchen. Nur wenn ein Teilchen eine Ladung besitzt (also nicht neutral ist) unterliegt es der jeweiligen Wechselwirkung. Es gibt drei verschiedene Arten von Ladungen: die elektrische Ladung, die schwache Ladung und die starke Ladung. Diese generieren jeweils die drei fundamentalen Wechselwirkungen: elektromagnetische, schwache und starke Wechselwirkung. Die vierte Wechselwirkung ist die Gravitation. [2]

Eine Eigenschaft der Ladung ist, dass ihre Werte nicht kontinuierlich verteilt sind, sondern dass die Ladung eines Körpers nur ein Vielfaches einer kleinsten Ladungsmenge annehmen kann. Man sagt dazu auch, Ladungen sind gequantelt. Außerdem gilt in jedem tatsächlich stattfindenden Prozess Ladungserhaltung, was bedeutet, dass alle Ladungen der Ausgangsteilchen sich zu der gleichen Summe addieren wie die Ladungen der entstandenen Teilchen. [2]

Die Gravitation, oder gravitative Wechselwirkung, ist verantwortlich für die Anziehung zwischen Massen und hält so zum Beispiel die Erde auf ihrer Bahn um die Sonne. Die Ladung, die der Gravitation als Wechselwirkung zugeordnet werden kann, ist somit die Masse. Jeder Körper und jedes Teilchen mit einer Masse unterliegt der gravitativen Wechselwirkung. Allerdings ergibt die Einordnung der Gravitation in die Konzepte von Ladungen und Wechselwirkungen auch Probleme und bis heute konnte die Gravitation nicht mit den anderen drei Wechselwirkungen des Standardmodells der Teilchenphysik zu einer alles umfassenden Theorie vereinigt werden. Die Masse ist zum Beispiel im Gegensatz zu den anderen Ladungen nicht gequantelt und kann nicht negativ sein. Damit ist die gravitative Wechselwirkung auch immer anziehend. Die Gravitation spielt bei der Elementarteilchenphysik aber eher eine untergeordnete Rolle, da sie für sehr kleine Massen, wie sie bei Atomen und Elementarteilchen auftreten, vernachlässigbar klein ist und somit praktisch keinen Einfluss mehr ausübt.

Die elektrische Ladung ist die bekannteste der Ladungen des Standardmodells und über die elektromagnetische Wechselwirkung auf atomarer Ebene zum Beispiel dafür verantwortlich, dass Elektronen am Atomkern binden und Atome bilden. Die elektrische Ladung kann positiv oder negativ sein und wird in Vielfachen Z der Elementarladung e angegeben, der Ladung eines Protons. Den Faktor Z nennt man elektrische Ladungszahl. Bisher hat man folgende Ladungszahlen für

die elektrische Ladung gefunden: $-1; -\frac{2}{3}; -\frac{1}{3}; 0; \frac{1}{3}; \frac{2}{3}; 1$. Alle Teilchen, die eine von 0 verschiedene elektrische Ladungszahl besitzen, die also elektrisch geladen sind, unterliegen der elektromagnetischen Wechselwirkung. Diese Wechselwirkung äußert sich in der Anziehung oder Abstoßung von elektrisch geladenen Teilchen. Für die Kraft zwischen den Teilchen gilt in der klassischen Elektrodynamik der Zusammenhang des Coulombschen Gesetzes für Punktladungen: [2]

$$F_C = \frac{e^2}{4 \cdot \pi \cdot \epsilon_0} \cdot \frac{Z_1 \cdot Z_2}{r^2} \quad (2.1)$$

mit der Elementarladung e und der elektrischen Feldkonstante ϵ_0 . In dieser Formel lassen sich typische Eigenschaften der Wechselwirkungen erkennen: die Proportionalität zu den Ladungen der beiden Teilchen Z_1 und Z_2 und die umgekehrte Proportionalität zum Quadrat des Abstandes der beiden Punktladungen r^2 . Außerdem besteht die Formel noch aus einer Konstante, von der die Kraft zusätzlich proportional abhängig ist. Dieser konstante Faktor (hier $\frac{e^2}{4\pi\epsilon_0}$) ist ein Maß für die Stärke der Wechselwirkung. [2]

Die schwache Ladung besitzt analog zur elektrischen Ladung die schwache Ladungszahl I . Es wurden bisher nur Materieteilchen (Fermionen) beobachtet, die die schwache Ladungszahl $I = -\frac{1}{2}$ oder $I = \frac{1}{2}$ besitzen, damit nehmen auch alle Materieteilchen an der schwachen Wechselwirkung teil. Diese ist zum Beispiel für Kernfusionen und -umwandlungen wie die β -Umwandlung verantwortlich. Auch die durch die schwache Wechselwirkung erzeugten Kräfte weisen in sehr kleinen Abständen eine $\frac{1}{r^2}$ -Abhängigkeit auf. Allerdings hat die schwache Wechselwirkung eine sehr viel geringere Reichweite als die anderen Wechselwirkungen, daher auch der Name „schwache“ Wechselwirkung. Das kommt daher, dass die durch die schwache Ladung erzeugte Kraft für größere Abstände der Ladungen sehr schnell exponentiell fällt. [2]

Die dritte Ladung ist die starke Ladung. Sie besteht nicht aus einer Zahl, sondern aus Vektoren, den sogenannten Farbladungsvektoren. Ein Teilchen kann rot, grün oder blau bzw. anti-rot, anti-grün oder anti-blau oder nicht stark geladen sein. Nur Quarks sind stark geladen und nehmen an der starken Wechselwirkung teil. Ein Proton besteht zum Beispiel aus drei so verschieden farbgeladenen Quarks, dass sich die drei Farbladungsvektoren zum Nullvektor addieren und das Proton nicht stark geladen ist. Die starke Wechselwirkung ist beispielsweise die Ursache für die Stabilität des Atomkerns und der Nukleonen, aus denen dieser besteht. Denn nach den anderen Wechselwirkungen müssten sich die Nukleonen abstoßen, und die Gravitation ist bei so enorm kleinen Massen zu klein, um Einfluss auszuüben. Es muss also eine „stärkere“ Kraft geben, die die Nukleonen zusammenhält. Dies ist die starke Wechselwirkung. Die starke Kraft hat für kleine Abstände wieder eine typische $\frac{1}{r^2}$ -Abhängigkeit, für größere Abstände sinkt sie aber nicht weiter, sondern bleibt konstant. Das hat zur Folge, dass Teilchen, die der starken Wechselwirkung unterliegen (Quarks) nicht vollständig voneinander getrennt werden können, da dafür unendlich viel Energie nötig wäre. Damit ist die Reichweite der starken Wechselwirkung in der Realität ebenfalls beschränkt. [2]

Die „Stärken“ der Wechselwirkungen lassen sich anhand des Kopplungsparameters vergleichen, eines Parameters der zu jeder der Wechselwirkungen gehört und in die Berechnung der Kräfte durch die Ladungen einfließt. Dieser ist keine Konstante, sondern ebenfalls mehr oder weniger leicht vom Abstand der Ladungen abhängig. Dennoch kann er gut zum Vergleich der Wechselwirkungsstärken dienen. Damit lässt sich zum Beispiel erklären, dass im Atomkern die starke Wechselwirkung einen höheren Einfluss als alle anderen hat und somit den Kern zusammenhält. [2]

In Tabelle 2.1 sind zur Übersicht die erklärten fundamentalen Wechselwirkungen, deren Ladungen, Reichweiten und Kopplungsparameter zusammenfassend dargestellt (Daten nach [2]).

Ladung	Wechselwirkung	Reichweite	Kopplungsparameter
Masse	gravitativ	unendlich	$\frac{1}{10^{38}} - \frac{1}{10^{45}}$
Elektrische Ladung	elektromagnetisch	unendlich	$\frac{1}{137}$
Schwache Ladung	schwach	$2 \cdot 10^{-18}$ m	$\frac{1}{30}$
Starke Farbladung	stark	$2 \cdot 10^{-15}$ m	$\frac{1}{2} - \frac{1}{10}$

Tabelle 2.1: Übersicht über die Wechselwirkungen, Ladungen und ihre Eigenschaften

2.2 Das Ordnungsschema der Materieteilchen

Die Teilchen, aus denen die Materie aufgebaut ist, nennt man Materieteilchen, in der Fachsprache Fermionen. Diese lassen sich nach bestimmten Eigenschaften ordnen. Die Ladung der Teilchen unterteilt sie in Untergruppen wie Quarks oder Leptonen. Über die Masse lassen sich die Teilchen außerdem in drei Generationen unterteilen. Mit zunehmender Generation einer Teilchenart wird das Teilchen nur schwerer, andere Eigenschaften bleiben gleich. So sind zum Beispiel die schwereren „Versionen“ des Elektrons das Myon und das Tauon, welche die gleiche Ladung wie das Elektron besitzen, aber schwerer sind. [2] Die folgende Tabelle (2.2) stellt das Ordnungsschema der Materieteilchen und wichtige Eigenschaften der Teilchen dar.

	1. Generation	2. Generation	3. Generation	Z	I	\vec{C}
Quarks (elektrisch positiv)	u Up-Quark	c Charm-Quark	t Top-Quark	$+\frac{2}{3}$	$+\frac{1}{2}$	r/ g/ b
Quarks (elektrisch negativ)	d Down-Quark	s Strange-Quark	b Bottom-Quark	$-\frac{1}{3}$	$-\frac{1}{2}$	r/ g/ b
elektrisch geladene Leptonen	e Elektron	μ Myon	τ Tauon	-1	$-\frac{1}{2}$	0
elektrisch neutrale Leptonen	ν_e Elektron-Neutrino	ν_μ Myon-Neutrino	ν_τ Tauon-Neutrino	0	$+\frac{1}{2}$	0

Tabelle 2.2: Das Ordnungsschema der Materieteilchen und ihre zugehörigen Ladungszahlen bzw. Farbladungsvektoren für elektrische- (Z), schwache- (I) und starke Ladung (\vec{C}) (Daten nach [2])

Zu jedem Materieteilchen existiert ein „Gegenstück“, ein sogenanntes Anti-Materieteilchen. Dieses hat die gleiche Masse wie das zugehörige Materieteilchen, aber alle Ladungen haben entgegengesetzte Vorzeichen. Das Antiteilchen des Elektrons ist zum Beispiel das Positron. Man kann also analog zum Ordnungsschema der Materieteilchen (siehe Tabelle 2.2) ein Schema für die Anti-Materieteilchen aufstellen. [2]

2.3 Das Konzept der Wechselwirkungsteilchen

„In der klassischen Physik werden die Fernwirkungen der elektromagnetischen Wechselwirkung und der gravitativen Wechselwirkung jeweils durch Felder [...] beschrieben. In der ersten Hälfte des 20. Jahrhunderts wurde entdeckt, dass die elektromagnetische Wechselwirkung auf quantenphysikalischer Ebene durch Teilchen, die Photonen [...], vermittelt wird. In der zweiten Hälfte des 20. Jahrhunderts wurde zudem herausgefunden, dass auch die starke und schwache Wechselwirkung jeweils durch Teilchen vermittelt werden. Die Teilchen, über die eine Wechselwirkung vermittelt wird, werden Botenteilchen (auch: Austauscheteilchen) [in der Fachsprache „Bosonen“] genannt.“ [2], S. 37

Mithilfe des Feldbegriffs lässt sich die $\frac{1}{r^2}$ -Abstands-Abhängigkeit der Wechselwirkungen mit unendlicher Reichweite geometrisch herleiten (über die Geometrie einer Kugel und deren Oberflächeninhalt). Das Feldmodell erklärt aber nicht die endliche Reichweite der schwachen Wechselwirkung. Daher wurde das Konzept der Bosonen, der Botenteilchen, entwickelt. Dieses besagt, dass Energie im Feld nicht kontinuierlich, sondern in kleinen Portionen, übertragen wird. Den Austausch von Bosonen kann man sich so vorstellen, dass ein Teilchen ein Boson erzeugt, welches ein anderes dann aufnimmt. Es lässt sich eine Gleichung für die Berechnung der mittleren Reichweite eines Botenteilchens λ mit Masse herleiten ($\lambda = \frac{\hbar}{m \cdot c}$) und damit die endliche Reichweite erklären. Aus dieser Formel folgt auch, dass Botenteilchen ohne eine Masse (wie das Photon) und demzufolge auch die dazugehörigen Wechselwirkungen, eine unendliche Reichweite haben müssen. [2]

Das Botenteilchen der elektromagnetischen Wechselwirkung ist das Photon. Das Photon hat

keine Masse und Ladung und unterliegt damit keiner Wechselwirkung. Die Botenteilchen der schwachen Wechselwirkung sind das W^+ , W^- und Z^0 -Teilchen. Deren Massen lassen sich zu etwa 100 GeV aus der mittleren Reichweite bestimmen. Die W-Bosonen sind jeweils elektrisch und schwach einfach positiv bzw. negativ geladen. Das Z-Boson ist nicht elektrisch und schwach geladen. Die starke Wechselwirkung wird über Gluonen vermittelt, ebenfalls masselose Elementarteilchen, die nur eine starke Ladung besitzen und sonst neutral sind. [2]

2.4 Feynman-Diagramme und Teilchenumwandlungen

Durch den Austausch von Wechselwirkungsteilchen können Elementarteilchen in Wechselwirkung treten oder sich ineinander umwandeln. Dies wird in Feynman-Diagrammen einfach und anschaulich dargestellt. Ein Feynman-Diagramm stellt Anfangs und Endzustand einer Teilchenumwandlung und mögliche Zwischenprozesse dar. Anhand eines Feynman-Diagramms lassen sich aber auch die physikalischen Gesetze der Wechselwirkungen aufstellen. Ein Beispiel für ein Feynman-Diagramm ist in Abbildung 2.1 am Beispiel der β^- -Umwandlung dargestellt. In den Diagrammen werden verschiedene Teilchen durch verschiedene Linientypen und optional Beschriftungen an den Linien dargestellt. So sind Materieteilchen als eine durchgezogene Linie mit Pfeil nach rechts, Antimaterieteilchen als eine durchgezogene Linie mit Pfeil nach links, Austauschteilchen der elektromagnetischen und schwachen Wechselwirkung als eine gewellte Linie und Gluonen als eine gekringelte Linie dargestellt. Ein Knotenpunkt (auch: Vertex) ist ein Aufeinandertreffen von mindestens drei Linien und veranschaulicht die Emission oder Absorption eines Botenteilchens. Die vier fundamentalen Vertices (Knotenpunkte) aus denen ein Feynman-Diagramm aufgebaut sein kann sind: Emission oder Absorption eines Botenteilchens, Paarvernichtung und Paarerzeugung. Bei der Paarvernichtung löschen sich ein Teilchen und sein Antiteilchen unter Entstehung eines Botenteilchens aus, während bei der Paarerzeugung aus einem Botenteilchen ein Teilchen-Antiteilchen-Paar erzeugt wird.

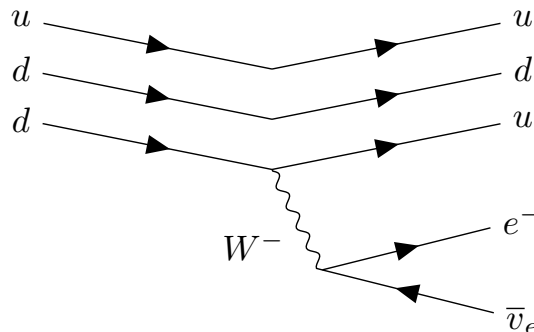


Abbildung 2.1: Feynman-Diagramm für die β^- -Umwandlung, bei dem ein Neutron (links, bestehend aus einem Up- und zwei Down-Quarks) in ein Proton (rechts, bestehend aus einem Down- und zwei Up-Quarks) umgewandelt wird

Für jedes Feynman-Diagramm und jeden Vertex im Feynman-Diagramm gilt die Ladungs-, Energie- und Impulserhaltung. Dies hat zur Folge, dass innerhalb eines Feynman-Diagramms virtuelle Teilchen auftreten können. Das bedeutet, dass diese Teilchen eine fest definierte Energie und einen fest definierten Impuls besitzen (durch die Energie- und Impulserhaltung) aber die relativistische Energie-Impuls-Beziehung ($E = \sqrt{p^2 c^2 + E_0^2}$) verletzen. Es gibt also auch virtuelle Teilchen mit $p > 0$ und $E = 0$ oder mit $p = 0$ und $E > 0$. Die in einem Feynman-Diagramm ein- und auslaufenden Teilchen (Ausgangsteilchen und Endteilchen) sind allerdings immer reelle Teilchen, das heißt, sie erfüllen die relativistische Energie-Impuls-Beziehung und ihnen kann somit auch eine Masse zugeordnet werden.

2.5 Das Z-Boson - Botenteilchen der schwachen Wechselwirkung

Das Z-Boson ist neben den W-Bosonen ein Austauscheteilchen der schwachen Wechselwirkung. Es ist selbst jedoch elektrisch, schwach und stark neutral. Daher unterliegt es selbst nur der gravitativen Wechselwirkung, die durch seine Masse hervorgerufen wird. Auch kann es die Ladungen der Teilchen zwischen denen es wechselwirkt aufgrund der Ladungserhaltung nicht verändern. Somit überträgt das Z-Boson nur Kräfte zwischen Teilchen oder wandelt Neutrinos ineinander um. In folgender Tabelle (2.3) sind wichtige charakterisierende Eigenschaften des Z-Bosons aufgeführt.

Eigenschaft	Wert
elektrische Ladungszahl Z	0
schwache Ladungszahl I	0
starker Farbladungsvektor \vec{C}	$\vec{0}$
Masse m	$91,1876 \pm 0,0021$ GeV
Zerfallsbreite Γ	$2,4952 \pm 0,0023$ GeV
mittlere Lebensdauer	$2,6379 \cdot 10^{-25}$ s

Tabelle 2.3: Überblick über einige wichtige charakterisierende Eigenschaften des Z-Boson, Daten direkt oder berechnet aus [3]

2.6 Natürliche Einheiten und das Schwerpunktsystem

In der Teilchenphysik werden nur selten SI-Einheiten, sondern vielmehr Natürliche Einheiten verwendet, da sie einige Vorteile gegenüber dem in der klassischen Physik verwendeten Einheitensystem haben.

Bei den Natürlichen Einheiten werden für Naturkonstanten Werte festgelegt und darauf ein neues Einheitensystem aufgebaut. In der Teilchenphysik werden zum Beispiel die Lichtgeschwindigkeit und das Plancksche Wirkungsquantum auf 1 festgelegt: $c := 1$; $\hbar := 1$. Damit kann man Geschwindigkeiten immer noch angeben, aber sie sind einheitenlos und werden in Anteilen der Lichtgeschwindigkeit ausgedrückt. Zum Beispiel als $v = 0,5$, was einer Geschwindigkeit von $0,5c$ entspricht. Aus diesem System ergeben sich einige Vorteile. Zum Beispiel haben so Energie, Masse und Impuls die gleiche Einheit, da nach der Äquivalenz von Masse und Energie ($E_0 = mc^2$) bei $c = 1$ gilt: $E_0 = m$.

Als Einheit der Energie wird bei diesem System 1 Elektronenvolt festgelegt:

$$[E] = [m] = [p] = 1 \text{ eV} \quad (2.2)$$

Ein Bezugssystem im Allgemeinen dient dazu, Bewegungen von Körpern zu beschreiben, denn diese sind je nach gewähltem Bezugssystem verschieden. Ein spezielles Bezugssystem ist das Schwerpunktsystem, bei dem der Schwerpunkt der Teilchen im System in Bezug zum System in Ruhe ist. Daraus folgt, dass die Summe aller Impulse 0 ergibt. Die Gesamtenergie des Schwerpunktsystems nennt man dann Ruheenergie E_0 oder invariante Masse M_{inv} . Dieses System hat ebenfalls viele Vorteile. Während die Energie eines Körpers oder eines Systems in allen Bezugssystemen verschieden ist, ist seine invariante Masse immer gleich.

Dieses Bezugssystem kann bei Teilchenkollisionen angewendet werden, da die Impulse der aufeinander treffenden Teilchen gleich groß aber entgegengesetzt gerichtet sind. Die Summe der Impulse ist also 0 und der Schwerpunkt des Systems befindet sich in Ruhe.

Kapitel 3

Experimenteller Aufbau

Teilchenbeschleuniger sind wichtige Instrumente der Grundlagenforschung in der Teilchenphysik. Durch Kollision von hochenergetischen Teilchen und Beobachtung der entstehenden Teilchen können Erkenntnisse über die Grundbausteine der Materie gewonnen werden, Theorien bestätigt oder widerlegt werden und schließlich vielleicht sogar neue Teilchen entdeckt werden. Diese Arbeit beschäftigt sich mit Daten aus dem ATLAS-Detektor im LHC aus dem Jahr 2016. Daher ist im folgenden Kapitel auch der Aufbau des Experiments 2016 beschrieben, da er sich in einigen Bestandteilen vom heutigen Aufbau unterscheidet.

3.1 Die Teilchenbeschleuniger am CERN

Der Large-Hadron-Collider (LHC) am internationalen Forschungszentrum CERN bei Genf ist mit einer Länge von 27 Kilometern der größte Teilchenbeschleuniger der Welt. In ihm werden hochenergetische Protonenstrahlen aufeinander geschossen, die ihre Energie aber auch schon vorher in verschiedenen Beschleunigern portionsweise übertragen bekommen.

Zuerst werden in einem Linearbeschleuniger (Stand 2016: Linac2) Protonen (H^+) auf eine Energie von 50 MeV beschleunigt, was nur etwa 5 % ihrer Ruhemasse ist. [4] Dabei durchfliegt ein Strahl dieser Teilchen eine Röhre, die aus einzelnen Elementen aufgebaut ist, die durch Anlegen einer Wechselspannung elektrisch geladen werden. Durch die Wechselspannung werden die Ionen im elektrischen Feld der Röhrenelemente beschleunigt. Der Linearbeschleuniger eignet sich allerdings nur für die Beschleunigung auf vergleichsweise niedrige Geschwindigkeiten bzw. Energien, da sonst die Länge der benötigten Beschleunigungsstrecke sehr groß wäre. [5]

Deshalb werden die Protonenstrahlen anschließend in drei kreisförmigen Beschleunigern auf eine höhere Energie gebracht. Sie durchlaufen nacheinander Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) und Super Proton Synchrotron (SPS). Dabei durchfliegen die Teilchen immer eine kreisförmige Röhre, an der Magnete angebracht sind. Die durch das Magnetfeld der Ablenkmagnete erzeugte Lorentzkraft hält die Teilchen auf der kreisförmigen Bahn. Die Fokussiermagnete verhindern die seitliche Ausbreitung der Protonenstrahlen. Zusätzlich gibt es auch in Ringbeschleunigern eine Beschleunigungsstrecke, in der die Teilchen auf eine höhere Energie gebracht werden. [5] Nach Verlassen des letzten Vorbeschleunigers, des Super Proton Synchrotrons, haben die Protonen eine Energie von 450 GeV. [6]

Nach den drei Vorbeschleunigern wird der Protonenstrahl in zwei einzelne Strahlen aufgeteilt und diese werden entgegengesetzt in den größten Ringbeschleuniger, den Large Hadron Collider (LHC), eingeleitet und darin bis auf die maximale Energie von 6,5 TeV beschleunigt. An speziellen Punkten im Beschleunigungsring können in Detektoren die nun entgegengesetzt kreisenden Protonenstrahlen zur Kollision gebracht werden. Die Energie, die bei der Kollision letztendlich freigesetzt wird, auch Kollisionenergie genannt, beträgt 13 TeV. [6]

3.2 Der ATLAS-Detektor

Der ATLAS-Detektor ist einer von vier Detektoren im LHC, in dessen Innerem Protonenstrahlen mit einer Energie von jeweils 6,5 TeV aufeinander treffen. Die Funktion des ATLAS-Detektors ist es, die Kollisionsprodukte und deren Eigenschaften zu messen. Der Detektor ist zylinderförmig aufgebaut, hat eine Länge von 46 Metern und einen Durchmesser von 25 Metern. [5] Der Aufbau des ATLAS-Detektors ist in folgender Abbildung (3.1) dargestellt.

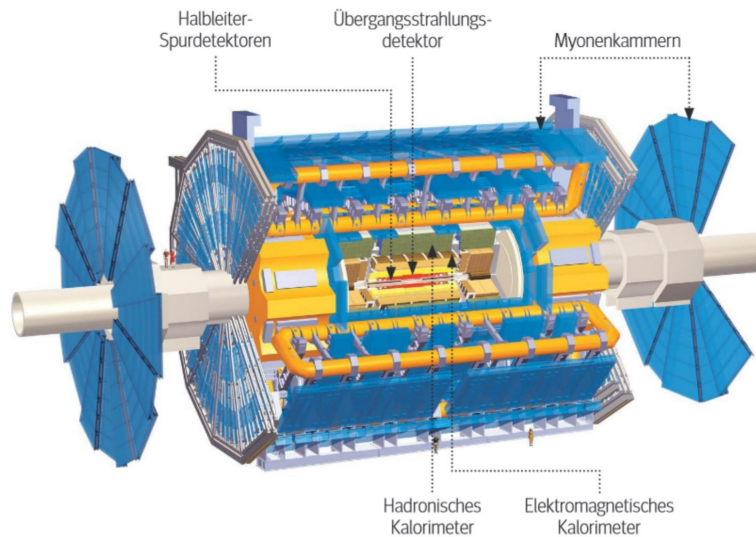


Abbildung 3.1: Aufbau des ATLAS-Detektors, Quelle: [5]

Der ATLAS-Detektor ist in Schichten aufgebaut, die rund um den Kollisionsspunkt angeordnet sind, da sich die entstehenden Teilchen von dort aus nach außen bewegen. In den verschiedenen Schichten des Detektors werden Spuren der Teilchen aufgezeichnet und ihre Energien gemessen. [5]

Die beiden Schichten, die am nächsten am Kollisionsspunkt liegen, sind die Spurdetektoren (Halbleiter-Spurdetektoren und Übergangsstrahlungsdetektoren). Diese zeichnen die Spur eines elektrisch geladenen Teilchens auf, das durch sie geflogen ist. Da in diesen Detektoren ein homogenes Magnetfeld erzeugt wird, wirkt auf die Teilchen außerdem die Lorentzkraft (Kraft auf bewegte Ladungen im Magnetfeld) und die Teilchenspuren beschreiben eine Kreisbahn, anhand derer man das Vorzeichen der Ladung des Teilchens bestimmen kann. Aus dieser Kreisbahn kann nach Identifikation des Teilchens auch sein Impuls bestimmt werden. [5]

Die nächsten beiden Detektorschichten sind Kalorimeter (elektromagnetische und hadronische Kalorimeter). Diese können die Energie mancher Teilchen bestimmen. Elektromagnetische Kalorimeter messen Elektronen, Positronen und Photonen, die in den Kalorimetern Photonen ausstrahlen (sogenannte „Bremsstrahlung“) und somit ihre komplette Energie abgeben. Über die Energie aller ausgestrahlten Photonen und ihrer Zerfallsprodukte kann die Ausgangsenergie des Teilchens bestimmt werden. In den hadronischen Kalorimetern werden die Hadronen (Verbindungen aus Quarks, wie das Proton) gestoppt und ihre Energie auf eine ähnliche Weise ermittelt. [5]

Die äußerste Schicht des ATLAS-Detektors besteht aus Myonenkammern, in denen nur Myonen Spuren hinterlassen und die somit auch nur für den Nachweis von Myonen verantwortlich sind. [5]

Teilchen sind also im ATLAS-Detektor über ihre Spuren, ihre Ladung, ihre Energie, und ihren Impuls nachweisbar. Neutrinos können im ATLAS-Detektor nicht nachgewiesen werden, da sie keine elektrische Ladung besitzen und nicht mit den Komponenten des Detektors in Wechselwirkung treten. Sie können allerdings indirekt nachgewiesen werden. Denn die Teilchenkollision muss immer die Impulserhaltung erfüllen. Daher muss der Gesamtimpuls aller entstandenen Teilchen 0 sein, wie es auch vor der Kollision der Fall ist. Wenn also zur Erfüllung der Impulserhaltung ein Impuls in eine bestimmte Richtung und mit einem bestimmten Betrag fehlt, ist das ein Zeichen dafür, dass in diese Richtung ein Neutrino geflogen sein könnte, das im ATLAS-Detektor nur nicht nachweisbar ist. [5]

Kapitel 4

Statistik

In diesem Kapitel sollen die mathematischen Grundlagen für die verwendeten Funktionen und Methoden in dieser Arbeit erläutert werden. Dabei geht es hauptsächlich um Wahrscheinlichkeiten, Wahrscheinlichkeitsdichten und statistische Datenauswertungsmethoden.

4.1 Die physikalische Wahrscheinlichkeit

Das Ziel der statistischen Physik ist das Beschreiben eines Systems mithilfe von Wahrscheinlichkeitsaussagen. Gerade für Messungen der Quantenmechanik nimmt die Wahrscheinlichkeit eine entscheidende Rolle ein, da es in ihrer Natur liegt, dass das Ergebnis einer Messung nur mithilfe von Wahrscheinlichkeiten vorausgesagt und damit auch beschrieben werden kann.

Bei einem Experiment in der Teilchenphysik werden in jedem Ereignis verschiedene Parameter gemessen, wie zum Beispiel den Impuls eines Teilchens. Diesen einen Parameter bezeichnet man im Konzept der Wahrscheinlichkeiten als Zufallsvariable. Alle möglichen Ausgänge des Experiments (also alle möglichen Werte für diese Zufallsvariable) bilden die Ereignismenge E (wieder bezogen auf diese Variable). Jedem dieser Ausgänge soll eine mathematische Wahrscheinlichkeit zugeordnet werden. Das eigentliche Messergebnis, die absolute Häufigkeit $n(A)$ eines Ereignisses A , trifft aber über dessen Wahrscheinlichkeit $P(A)$ noch keine Aussage. Dazu wird die relative Häufigkeit $h(A)$ verwendet. Diese berechnet sich aus der absoluten Häufigkeit $n(A)$ und der Anzahl der Messungen N : [7]

$$h(A) = \frac{n(A)}{N} \quad (4.1)$$

Wenn die Anzahl der Messungen N sehr hoch ist, wird sich die relative Häufigkeit $h(A)$ der tatsächlichen zu ermittelnden physikalischen Wahrscheinlichkeit $P_{phys}(A)$ des Ereignisses A annähern. Die Physikalische Wahrscheinlichkeit wird also definiert als: [7]

$$P_{phys}(A) = \lim_{N \rightarrow \infty} \frac{n(A)}{N} \quad (4.2)$$

4.2 Wahrscheinlichkeitsdichten

Dieses einfache Prinzip der physikalischen Wahrscheinlichkeit gilt nur für diskrete Zufallsvariablen, also wenn die Anzahl der möglichen Ausgänge eines Experiments endlich ist. Dann kann jedem Ausgang der Ereignismenge E eine Wahrscheinlichkeit zugeordnet werden. Für kontinuierliche Zufallsvariablen ist diese Ereignismenge aber überabzählbar. Daraus ergeben sich Probleme mit der Definition der physikalischen Wahrscheinlichkeit. [7]

Da es in jedem Intervall unendlich viele mögliche Ergebnisse gibt, geht die Wahrscheinlichkeit für jedes bestimmte Ereignis gegen 0. Wenn der Ausgang eines Experiments beispielsweise eine rationale Zahl ist, ist leicht ersichtlich, dass es bei kontinuierlichen Zufallsvariablen unendlich viele mögliche Ausgänge gibt (so viele wie rationale Zahlen). Damit ist die Wahrscheinlichkeit für jeden

einzelnen Wert bei endlich vielen Messungen 0. Es muss also eine neue Definition für die Wahrscheinlichkeit gefunden werden, die auch für kontinuierliche Zufallsvariablen gilt. Daher verwendet man die Wahrscheinlichkeitsdichte, die ebendiese Voraussetzung erfüllt.

4.2.1 Definition von Wahrscheinlichkeitsdichten

Die Wahrscheinlichkeit einer kontinuierlichen Zufallsvariable ξ muss also in Intervallen definiert werden. Man kann auch bei kontinuierlichen Zufallsvariablen eine Aussage darüber treffen, wie hoch die Wahrscheinlichkeit $P(a < \xi < b)$ ist, dass der gemessene Wert in einem bestimmten Intervall $[a, b]$ liegt, wohingegen das für einen bestimmten Wert nicht möglich ist. Wenn diese Intervalle, für die die Wahrscheinlichkeit definiert ist, infinitesimal sind, kann eine Wahrscheinlichkeitsdichte der Funktion

$$f(x) = \frac{dP(x \leq \xi < x + dx)}{dx} \quad (4.3)$$

definiert werden. [7] Das folgt daraus, dass die Wahrscheinlichkeit für ein bestimmtes Integral als Fläche unter der Wahrscheinlichkeitsdichtefunktion definiert ist (siehe Formel 4.4).

Durch diese Definition der Wahrscheinlichkeitsdichte werden benannte Probleme (siehe Kapitel 4.2) umgangen. Die Wahrscheinlichkeit ist nun auch für infinitesimale Intervalle definiert und kann mittels Integration für jedes beliebige Intervall $[x_1, x_2]$ berechnet werden:

$$P(x_1 \leq \xi < x_2) = \int_{x_1}^{x_2} f(x) dx \quad (4.4)$$

Auch diese Dichtefunktion erfüllt noch wichtige Eigenschaften der mathematischen Wahrscheinlichkeit, wie die Gesamtwahrscheinlichkeit, hier die Fläche unter dem Graphen, die 1 beträgt:

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (4.5)$$

Es gibt mit der Wahrscheinlichkeitsdichteverteilung also ein Mittel, um auch für kontinuierliche Zufallsvariablen jedem Ausgang des Experiments (bzw. jedem möglichen Wert) eine Wahrscheinlichkeit zuzuordnen.

4.2.2 Die Normalverteilung

In der statistischen Physik findet besonders häufig die Normalverteilung oder Gauß-Verteilung Anwendung, weil mit ihr gut zufällige Abweichungen von Mittelwerten, wie zum Beispiel Erwartungswerten in Experimenten, beschrieben werden können. Sie ist eine spezielle Dichtefunktion für normalverteilte Zufallsvariablen und ist definiert als

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.6)$$

Dabei gibt μ den Mittelwert der Verteilung an und σ die Standardabweichung von diesem Wert. Das heißt, μ verschiebt die Dichtefunktion entlang der x-Achse, und je größer σ , desto breiter sind die Werte um den Mittelwert gestreut. [8] Das kann man auch in Abbildung 4.1 erkennen, in der Dichtefunktionen der Normalverteilungen für verschiedene Werte der Parameter dargestellt sind.

Ein Spezialfall der Normalverteilung ist die Standardnormalverteilung. Bei dieser Verteilung ist der Erwartungswert $\mu = 0$ und die Standardabweichung $\sigma = 1$. Dadurch ergibt sich eine Normalverteilung, deren Graph um die y-Achse symmetrisch ist. Ihre Dichtefunktion ist nach [8]:

$$s(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \quad (4.7)$$

In der Abbildung 4.2 ist die Dichtefunktion der Standardnormalverteilung dargestellt.

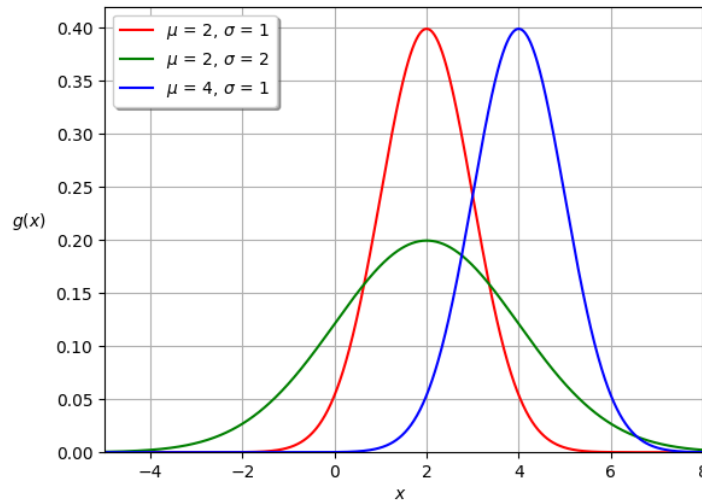


Abbildung 4.1: Dichtefunktionen verschiedener Normalverteilungen mit Mittelwerten μ und Standardabweichungen σ

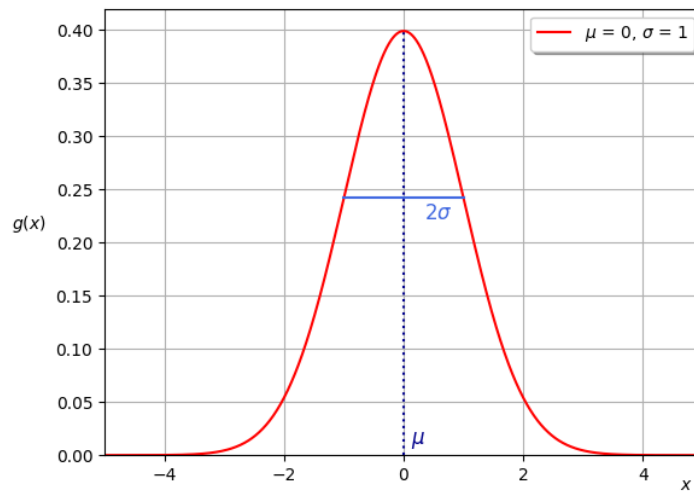


Abbildung 4.2: Dichtefunktion der Standardnormalverteilung mit $\mu = 0$ und $\sigma = 1$

4.2.3 Die Chi-Quadrat-Verteilung

Die Chi-Quadrat-Verteilung bzw. χ^2 -Verteilung ist eine aus der Standardnormalverteilung (siehe Kapitel 4.2.2) abgeleitete Dichtefunktion. Aus einer standardnormalverteilten Zufallsvariable x kann eine neue Variable χ (Chi) berechnet werden, die definiert ist als:

$$\chi := \frac{x - \mu}{\sigma} \quad (4.8)$$

Es ergibt sich auch die neue Variable χ^2 mit $\chi^2 := \frac{(x - \mu)^2}{\sigma^2}$. [7] Diese Variable χ^2 folgt der Dichtefunktion der χ^2 -Verteilung mit einem Freiheitsgrad: [7]

$$f(\chi^2|1) = \frac{1}{\sqrt{2\pi\chi^2}} \cdot e^{-\frac{1}{2}\chi^2} \quad (4.9)$$

Soll der χ^2 -Wert für n standardnormalverteilte Zufallsvariablen berechnet werden, müssen die Einzel- χ^2 -Werte aufsummiert werden. Die Anzahl n der aufsummierten Zufallsvariablen nennt man Freiheitsgrad (engl.: degrees of freedom, abgekürzt: dof). Chi-Quadrat mit n Freiheitsgraden berechnet sich also aus: [7]

$$\chi^2 := \sum_{i=1}^n \chi_i^2 \quad (4.10)$$

Dieses χ^2 folgt der Wahrscheinlichkeitsdichtefunktion $f(\chi^2|n)$ und hat den Mittelwert $\langle \chi \rangle = n$ und die Varianz $\sigma^2 = 2n$. Diese Dichtefunktion ist in Abbildung 4.3 dargestellt.

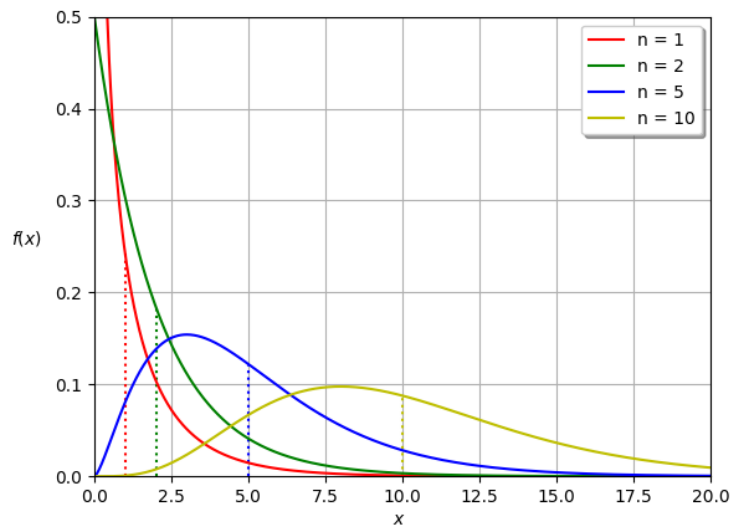


Abbildung 4.3: Dichtefunktion der Chi-Quadrat-Verteilung mit n Freiheitsgraden

Diese Verteilung kann in der Auswertung von Messwerten gut genutzt werden, sodass man sie also zum Beispiel beim Minimum-Chiquadrat-Fit (siehe Kapitel 4.3.2) einsetzt.

4.3 Fitting

4.3.1 Kurvenfitting als Auswertungsmethode

Wie schon in Kapitel 4.2 erklärt, ist die Aufgabe einer Wahrscheinlichkeitsdichte, die Messwerte eines Experiments (bzw. die daraus berechneten Häufigkeiten bzw. Wahrscheinlichkeiten) für eine bestimmte Zufallsvariable darzustellen. Die Dichteverteilung würde also, bezogen auf das Experiment, darstellen, welche Werte für einen bestimmten Parameter wie oft gemessen wurden. Aus einer Normalverteilung etwa könnte man leicht den Erwartungswert μ für die zu ermittelnde Variable ablesen. Die Auswertung der Ergebnisse eines physikalischen Experiments liefert aber keine Wahrscheinlichkeitsdichte, sondern meist ein Histogramm. Denn man kann nicht direkt messen, welche Wahrscheinlichkeiten spezielle Werte haben, sondern nur, wie häufig Werte vorkommen. Zur Erstellung eines Histogramms unterteilt man die möglichen Ausprägungen der Zufallsvariable in Intervalle $[a_i, b_i]$, auch Bins genannt. Im Histogramm wird dargestellt, für wie viele Ereignisse des Experiments die Variable in diesem Intervall lag, also wie oft ein Wert gemessen wurde, der in diesem Intervall lag. Die Intervalllänge nennt man dabei auch Binbreite Δx mit $\Delta x = b_i - a_i$ und den Intervallmittelpunkt Binmitte $x_i = \frac{a_i + b_i}{2}$. [7] Alle erklärten Größen sind in dem Beispielhistogramm in Abbildung 4.4 dargestellt.

Wenn die Binbreite Δx nicht sehr klein ist, beschreibt die Wahrscheinlichkeit für den gesamten Bin aber nicht ausreichend gut die Wahrscheinlichkeit für die Binmitte, da es innerhalb des Intervalls immer noch abweichende Wahrscheinlichkeiten geben kann. Daher sollte die Binaufteilung möglichst fein gewählt werden.

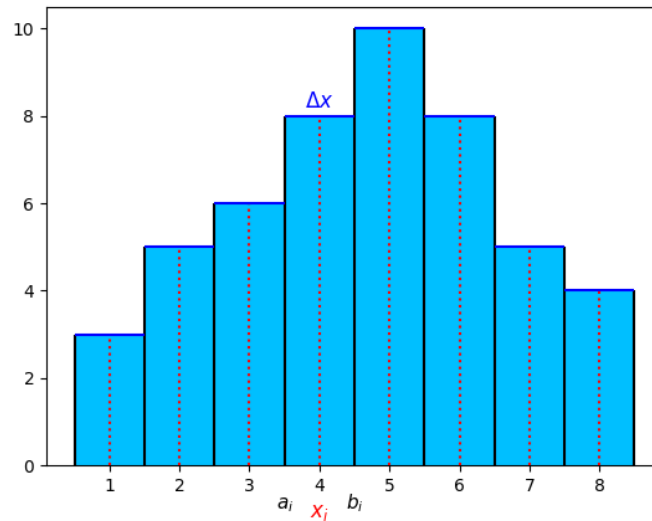


Abbildung 4.4: Beispielhistogramm

Das Ziel des Kurvenfittings ist es, zu diesem Histogramm der Messergebnisse eine Funktion zu finden, die die Werte möglichst gut beschreibt. Es ist hier bewusst von Funktion und nicht Dichtefunktion die Rede, weil diese ganz klar unterschieden werden müssen. Ein Histogramm, wie es aus Experimenten erstellt wird, beschreibt absolute Häufigkeiten, da dargestellt wird, wie oft Werte in bestimmten Intervallen vorkommen. Eine Dichtefunktion beschreibt Wahrscheinlichkeiten, also relative Werte. Wenn man die absoluten Messwerte nun mit einer Funktion (keiner Dichtefunktion!) beschreiben will, müssen entweder die Messwerte zu relativen Häufigkeiten umgerechnet werden (siehe Kapitel 4.1) oder die Dichtefunktion muss mit einem zusätzlichen Parameter skaliert werden, um absolute Werte zu erreichen. Da dann aber das Integral $\int_{-\infty}^{\infty} f(x)dx$ unter der resultierenden Fitting-Funktion $f(x)$ ungleich 1 ist, ist sie mathematisch keine Wahrscheinlichkeitsdichtefunktion mehr.

4.3.2 Minimum-Chiquadrat-Fit

Die eigentliche Frage ist, wie man eine Fitting-Funktion findet, welche die Messwerte möglichst gut beschreibt. Dazu wird die Methode der kleinsten Quadrate, auch Minimum-Chiquadrat-Fit [7] genannt, verwendet. Diese versucht im Allgemeinen die Abweichungen der Fitting-Funktion von den eigentlichen Messwerten zu minimieren. Diese Abstände nennt man Residuen r_i . Für diese gilt der Zusammenhang mit den Messwerten y_i und den Werten der Fitting-Funktion $f(x_i)$:

$$r_i = |y_i - f(x_i)| \quad (4.11)$$

Nun wird nun der χ^2 -Wert berechnet aus den Residuen r_i und den jeweiligen Varianzen σ^2 der Werte in jeder der insgesamt n_B Bins (siehe Formel 4.8):

$$\chi^2 = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{\sigma_i^2} \quad (4.12)$$

Deshalb wird diese Art des Fits auch Minimum-Chiquadrat-Fit genannt. [8]

Man bildet die Residuenquadratsumme $\sum_{i=1}^{n_B} r_i^2$ statt der Residuensumme $\sum_{i=1}^{n_B} r_i$, da größere Residuen (also größere Abweichungen der Werte) einen höheren Einfluss auf die Fitting-Funktion haben sollen. Und dieser Wert wird durch die Teilung durch die Varianz σ^2 gewichtet, damit Werte mit höherer Unsicherheit weniger in das Gesamtergebnis χ^2 eingehen.

Gibt es viele Bins, also viele mögliche Ausprägungen der Zufallsvariable, so ist die Gesamtmenge der Messwerte hoch, aber die Wahrscheinlichkeit für jedes einzelne Ereignis nahezu 0. Daher können

die Bininhalt annähernd durch eine Poisson-Verteilung beschrieben werden (das folgt aus der Definition der Poisson-Verteilung, auf die hier nicht näher eingegangen werden soll). Da für den Erwartungswert $\langle x \rangle$ in der Poisson-Verteilung gilt: $\langle x \rangle = \mu = \sigma^2$, kann auch für die χ^2 -Werte in den Bins angenommen werden, dass die Varianz gleich groß ist wie der Wert der Fitting-Funktion und der Wert des Binmittelpunkts: $\sigma_i^2 = f(x_i) = y_i$. Daraus ergibt sich der Wert für χ^2 :

$$\chi^2 = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{f(x_i)} = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{y_i} \quad (4.13)$$

der sich leichter berechnen lässt, wenn die Varianz des Bins nicht gegeben ist. [8]

Man hat also jetzt einen Wert χ^2 , der für die Messwerte eines Experiments und deren Fitting-Funktion $f(x)$ angibt, wie groß deren allgemeine Abweichung ist. Je größer die Residuen jedes Bins mit der Funktion, desto höher ist χ^2 . Beim Fitting versucht man eine Funktion zu finden, welche die Messwerte optimal beschreibt, also wo der χ^2 -Wert möglichst klein ist. Diese Ermittlung der optimalen Parameter für eine vorgegebene Fitting-Funktion kann numerisch erfolgen. Als Ergebnis erhält man wie gefordert eine Funktion, die optimal an die Messwerte angepasst ist.

4.4 Statistische Signifikanz

Mithilfe von Experimenten sollen Theorien und Annahmen überprüft werden. Ein Maß um zu beurteilen ob die gemessenen Werte die Annahmen und Theorien bestätigen, ist die statistische Signifikanz. Je geringer dieser Wert, desto mehr bestätigen die Messwerte die Annahmen. Hohe Signifikanzen sprechen dagegen dafür, dass das Messergebnis weniger mit der Annahme übereinstimmt.

Die statistische Signifikanz gibt an, wie weit ein bestimmtes Messergebnis y von einem dafür erwarteten Ergebnis y_{exp} abweicht. Diese Abweichung wird aber nicht absolut, sondern als Vielfache der Unsicherheit oder Standardabweichung $\sigma_{y_{exp}}$ des erwarteten Ergebnisses y_{exp} angegeben. Dabei wird angenommen, dass dieser Wert von einer statistischen Unsicherheit behaftet ist, die sich durch die Normalverteilung beschreiben lässt, mit dem Wert y_{exp} als Mittelwert der Verteilung.

$$Z = \frac{|y - y_{exp}|}{\sigma_{y_{exp}}} \quad (4.14)$$

Die Standardabweichung σ des Wertes y_{exp} lässt sich vereinfacht berechnen durch $\sigma_{y_{exp}} = \sqrt{y_{exp}}$. Die Signifikanz kann also, wenn keine Angabe für die Unsicherheit des Werte gegeben ist, vereinfacht berechnet werden durch:

$$Z = \frac{|y - y_{exp}|}{\sqrt{y_{exp}}} \quad (4.15)$$

Mittels der Signifikanz lässt sich dann eine Aussage darüber treffen, wie wahrscheinlich es ist, dass der Wert y zur vorher getroffenen Annahme y_{exp} passt. Die Wahrscheinlichkeiten dafür gehen aus den Integralen der Normalverteilung hervor:

Signifikanz	P(Fluktuation)
1	31,731 %
2	4,550 %
3	0,270 %
4	0,0063 %
5	$5,73 \cdot 10^{-7}$

Tabelle 4.1: Wahrscheinlichkeiten für das Übereinstimmen von Wert und Annahme (P(Fluktuation)) nach Signifikanzen, Quelle für die Daten: [7]

Für Signifikanzen über 3 ist somit die Wahrscheinlichkeit für die Übereinstimmung von Wert und Annahme bereits sehr gering. Signifikanz $Z = 3$ bedeutet einfach ausgedrückt, dass der Wert y drei Standardabweichungen $\sigma_{y_{exp}}$ vom erwarteten Wert y_{exp} entfernt liegt.

Kapitel 5

Neuronale Netze

Neuronale Netze werden auch in der Teilchenphysik immer öfter eingesetzt. Im folgenden Kapitel soll ein kleiner Einblick in das komplizierte Thema geschaffen werden. Im Grunde ist das Ziel Neuroner Netze, eine Struktur zu schaffen, die in der Lage ist, gelernte Inhalte zu übertragen und selbstständig intelligente Entscheidungen zu treffen.

5.1 Aufbau Neuroner Netze

Die Grundbausteine Neuroner Netze sind Neuronen, die miteinander verbunden sind. Da in dieser Arbeit nur einfache vorwärtsgerichtete, vollvernetzte Netze verwendet wurden, wird auch die Erklärung im Folgenden auf diese Modelle beschränkt. Die Neuronen sind in Schichten, sogenannten „layers“ angeordnet. Jedes Neuronale Netz besteht mindestens aus einer Eingabeschicht („input layer“) und einer Ausgabeschicht („output layer“). Dazwischen befinden sich beliebig viele verdeckte Schichten („hidden layers“). [9] In Abbildung 5.1 ist die Struktur eines beispielhaften Neuronalen Netzes mit zwei verdeckten Schichten dargestellt.

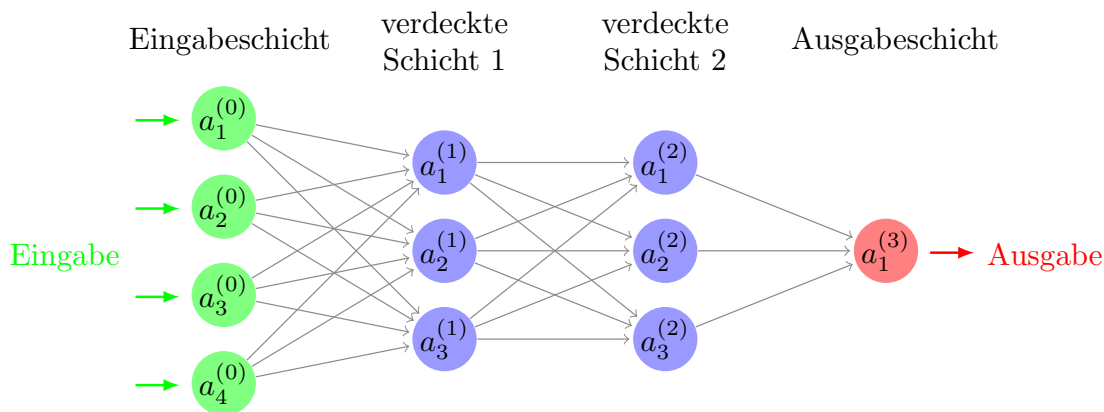


Abbildung 5.1: Aufbau eines beispielhaften Neuronalen Netzes

Jede Schicht besteht aus einer bestimmten Anzahl von Neuronen, die bei vorwärtsgerichteten Neuronalen Netzen in Richtung Ausgabeschicht miteinander verbunden sind. Jedes Neuron besitzt eine Aktivierung $a_n^{(m)}$, wobei m die Schicht ist, in der sich das Neuron befindet, und n die Nummer des Neurons in der Schicht. Die Eingabewerte bestimmen die Aktivierungen der Neuronen in der ersten Schicht. Durch die Verbindungen der Neuronen der ersten und zweiten Schicht müssen nun aus den Aktivierungen der Neuronen der ersten Schicht die Aktivierungen der Neuronen der zweiten Schicht berechnet werden. Das wird dann bis zur Ausgabeschicht wiederholt, die einen Ausgabewert erzeugt. Dazu sind die Neuronen über Gewichte miteinander verbunden, die in der Abbildung 5.1 durch Pfeile veranschaulicht sind. Die Aktivierung eines Neurons wird aus der gewichteten Summe

der Aktivierungen der Neuronen der vorherigen Schicht berechnet. [9] In der Abbildung 5.2 ist die Berechnung der gewichteten Summe für ein Neuron dargestellt.

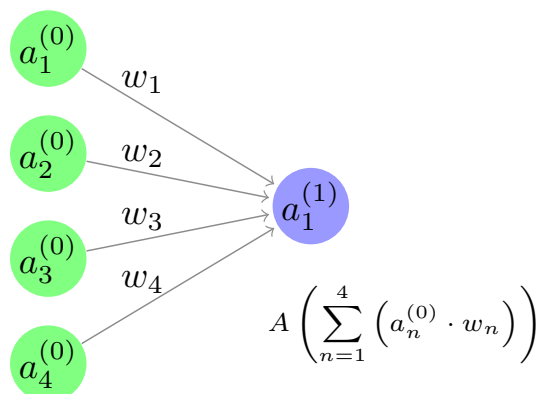


Abbildung 5.2: Berechnung der Aktivierung von Neuronen am Beispiel eines Neurons

Dazu wird für jede Verbindung des Neurons zu einem Neuron der vorherigen Schicht das Produkt aus Aktivierung des Neurons der vorherigen Schicht und Wichtung der Verbindung berechnet ($a_n^0 \cdot w_n$). Diese gewichteten Aktivierungen werden für alle Verbindungen der vorherigen Schicht zu dem Neuron summiert und bestimmen dessen Aktivierung. Oft soll die Aktivierung eines Neurons jedoch nicht ein beliebig hoher Wert sein (wie man es erreicht, wenn einfach nur die gewichtete Summe berechnet wird), sondern in einem bestimmten Bereich liegen (beispielsweise zwischen 0 und 1). Dazu verwendet man Aktivierungsfunktionen, die jeder gewichteten Summe eine bestimmte Aktivierung zuordnen. [9] In den Abbildungen 5.3 und 5.4 sind die beiden Aktivierungsfunktionen Sigmoid und ReLU dargestellt, die auch in dieser Arbeit verwendet werden.

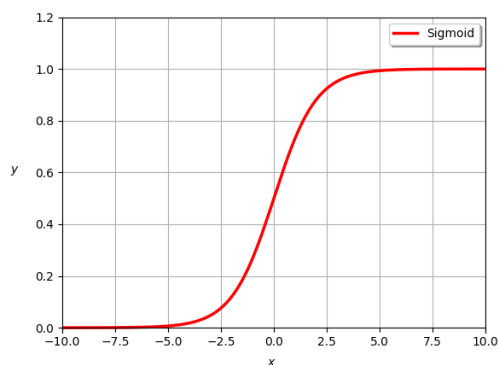


Abbildung 5.3: Sigmoid-Funktion

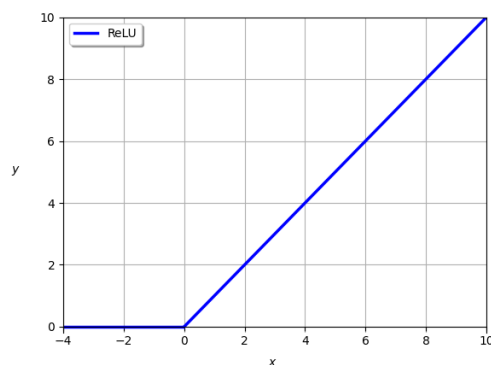


Abbildung 5.4: ReLU-Funktion

Die Aktivierung eines Neurons berechnet sich also aus der Anwendung der Aktivierungsfunktion auf die gewichtete Summe (siehe Abbildung 5.2). Bei der Sigmoid-Funktion wird somit für jedes Neuron eine Aktivierung zwischen 0 und 1 erreicht, bei der Anwendung der ReLU-Funktion wird das Neuron dagegen erst ab einem bestimmten Schwellenwert (hier 0) aktiviert, davor ist es inaktiv.

Ein Neuron ist nun also definiert als eine Funktion, die mit den Aktivierungen der Neuronen der vorherigen Schicht als Eingabewerten und den Gewichten der Verbindungen als Parametern seine Aktivierung berechnet. Auch das gesamte Neuronale Netz kann demzufolge als eine Funktion angesehen werden. Dabei sind die Gewichte der Verbindungen zwischen den Neuronen die Parameter der Funktion und durch die Veränderung dieser Gewichte kann angepasst werden, wie die Aktivierungen der Neuronen sich gegenseitig beeinflussen und so eine Ausgabe erzeugen. Durch die Veränderung der Gewichte wird also das Verhalten des Neuronalen Netzes auf die Eingabewerte verändert, man sagt, das Netz wird trainiert.

5.2 Training Neuronaler Netze

Das Neuronale Netz soll nach dem Training eine bestimmte Aufgabe erfüllen, auf die es trainiert wird. Das Ziel ist es dabei, dass das Neuronale Netz bei bestimmten Eingabewerten bestimmte Ausgabewerte erzeugt. In dieser Arbeit beispielsweise soll ein Neuronales Netz später auf Grundlage von Messergebnissen aus einer Teilchenkollision entscheiden, ob bei dieser Kollision ein Higgs-Boson entstanden ist. Das Neuronale Netz soll also als Ausgabe eine Zahl zwischen 0 und 1 generieren, die angibt, wie wahrscheinlich es ist, dass in diesem Ereignis ein Higgs-Boson entstanden ist. Dazu wird das Netz vorher auf bereits klassifizierte Trainingsdaten angewandt, bei denen also die gewollte (bzw. erwartete) Ausgabe (0 oder 1) bereits bekannt ist. Dabei werden die Parameter (Gewichte) des Neuronalen Netzes so verändert, dass die erwartete und tatsächliche Ausgabe des Netzes möglichst gleich sind, das heißt, dass das Neuronale Netz die Daten möglichst gut klassifiziert. Dazu wird mittels einer Fehlerfunktion berechnet, wie groß die Differenz zwischen erwarteten und generierten Ausgaben ist. Meist verwendet man dazu die Summe der Quadrate der Einzeldifferenzen für die jeweiligen Eingabe-Datenpaare [10]. Diesen Wert nennt man auch „Trainings-Loss“ des Neuronalen Netzes. Das ist vergleichbar mit dem Chi-Quadrat-Wert beim Fitting von Funktionen (siehe Kapitel 4.3.2). Auch der Fehler des Neuronalen Netzes soll während des Trainings minimiert werden, wozu die Parameter des Netzes, also die Gewichte der Verbindungen der Neuronen, verändert werden. Die Fehlerfunktion berechnet aus den Parametern des Neuronalen Netzes den Fehler, das heißt, man sucht während des Trainings das Minimum dieser Fehlerfunktion. Der Unterschied zum Fitting von Funktionen (Kapitel 4.3.2) ist, dass die Fehlerfunktion, deren Minimum beim Training gefunden werden soll, sehr viel komplexer ist als beim Fitting, da sie so viele Parameter wie das Neuronale Netz Gewichte hat (was meist einige hundert bis tausend sind). [10]

Grundsätzlich werden die Gewichte je nach ihrem Beitrag zum Fehler der Ausgabe (bzw. Loss-Wert) verändert. Gewichte, die den Fehler minimieren, werden vergrößert, während Gewichte, die den Fehler vergrößern, verkleinert werden. Und Gewichte, die einen stärkeren Einfluss auf das Ergebnis haben, werden auch stärker verändert. Genauer wird für die Veränderung der Gewichte der Gradient der Fehlerfunktion berechnet und für die Veränderung der Gewichte innerer (verdeckter) Schichten wird sogenanntes „Backpropagation“ genutzt. [10] Dabei werden die Gewichte schichtenweise von der Ausgabeschicht aus verändert, auf die genaue Funktionsweise von Backpropagation soll hier aber nicht näher eingegangen werden.

Das Training Neuronaler Netze findet außerdem in mehreren Epochen statt. Bei einer Epoche wird der komplette vorhandene Datensatz, auf den das Netz trainiert wird, einmal angewandt. Wenn man den gleichen Datensatz nochmals anwendet, kann man meist eine weitere Senkung des Fehlers erreichen, weswegen man ein Neuronales Netz in mehreren Epochen trainiert. [10]

Das Ziel des Trainings eines Neuronalen Netzes ist eigentlich, dass das Netz aus den Trainingsdaten lernt und diese Beispiele abstrahiert und allgemeingültige Gesetze aufstellt. Dieser Prozess heißt Generalisierung und sorgt dafür, dass das Neuronale Netz später auch auf anderen Daten als den Trainingsdaten funktioniert. [10] Vor allem beim Training mit vielen Epochen und einem sehr komplexen Neuronalen Netz (d. h. mit vielen Gewichten) kann aber Overtraining auftreten (wie beim Fitting von Funktionen Overfitting). Dabei generalisiert das Netz nicht, sondern lernt die Trainingsdaten nur „auswendig“. Das führt zwar zu einem sehr kleinen Trainings-Loss, sorgt aber auch dafür, dass das Neuronale Netz auf ihm unbekanntem Daten sehr schlecht funktioniert. Daher muss neben dem Training auch besonderer Wert auf die Auswahl der geeigneten Größe des Neuronalen Netzes gelegt werden. [9]

Kapitel 6

Z-Boson-Analyse

Ziel der Teilchenphysik ist es, die Elementarteilchen, ihre Eigenschaften und Wechselwirkungen zu erforschen. In der Theorie beschreibt dies das Standardmodell der Teilchenphysik. In der Praxis werden Experimente durchgeführt und ausgewertet, um diese Theorien zu bestätigen oder zu widerlegen. Diese Forschung ist auch Teil dieser Arbeit. Zuerst wurde dies am Beispiel des Z-Bosons durchgeführt, eines Botenteilchens der schwachen Wechselwirkung. Das Ziel der Untersuchung war, mittels Aufarbeitung und Analyse der Daten des ATLAS-Experiments, letztendlich die Masse des Z-Bosons zu ermitteln.

Die kompletten in dieser Arbeit zur Analyse verwendeten Daten stammen aus einem Datensatz, der vom ATLAS-Experiment zu Ausbildungszwecken frei zur Verfügung gestellt wurde. Die Daten wurden im Jahr 2016 unter einer Kollisionsenergie von 13 TeV im ATLAS-Detektor bei Protonenkollisionen gemessen. Ebenfalls zur Verfügung gestellt wurden simulierte Daten nach der Theorie des Standardmodells der Teilchenphysik und einige Analyseprogramme, in denen dann eigene Modifikationen und Veränderungen vorgenommen wurden. [11] Auch alle folgenden Histogramme wurden mit den zur Verfügung gestellten Analyseprogrammen erstellt.

6.1 Signal und Messuntergrund

Die Ergebnisse jeder Messung bestehen aus zwei Teilen: Signal und Messuntergrund. Das Signal beinhaltet die Ereignisse, die man bei dem Experiment auch wirklich zu beobachten beabsichtigt. Wenn man zum Beispiel den Zerfall des Z-Bosons messen möchte, besteht das Signal aus Ereignissen, in denen tatsächlich ein Z-Boson zerfallen ist. Diese Ereignisse sind für die Auswertung des Experiments relevant. Anders als der Messuntergrund, Ereignisse die nicht gemessen werden sollen und das Ergebnis verfälschen. Sie sind zur Auswertung daher auch nicht relevant und man versucht ihren Anteil zu minimieren. Ein wichtiger Schritt vor der Auswertung der Messergebnisse ist daher immer die Selektion (siehe Kapitel 6.3).

6.2 Der Zerfall des Z-Bosons

Wenn beim Experiment ein Z-Boson entsteht, zerfällt es wieder nach einer kurzen Zeit von durchschnittlich $2,6 \cdot 10^{-25}$ s. [3] Für den Zerfall gibt es verschiedene Möglichkeiten, die jeweils mit einer bestimmten Wahrscheinlichkeit auftreten. Diese Wahrscheinlichkeiten für die sogenannten „Zerfallskanäle“ des Z-Bosons sind in Tabelle 6.1 aufgeführt.

Die folgende Analyse beschränkt sich auf die Zerfälle des Z-Bosons in Elektron und Positron und in Myon und Antimyon, da alle anderen Zerfälle zur Messung oder Auswertung ungeeignet sind. Zerfälle in Neutrinos sind mittels der ATLAS-Detektoren grundsätzlich nicht nachweisbar, daher können diese auch nicht analysiert werden. Ebenso sind alle Zerfälle, an denen Tauonen beteiligt sind, ungeeignet, da diese eine so enorm kurze durchschnittliche Lebensdauer von $(290,3 \pm 0,5) \cdot 10^{-15}$ s haben [3], dass sie nur über ihre Zerfallsprodukte messbar sind. Zerfälle in Quarks könnten prinzipiell gemessen und in die Analyse mit einbezogen werden, aber bei ihnen ist der Anteil des Messuntergrunds höher als bei den ausgewählten Zerfällen, was die Analyse zusätzlich erschwert.

Zerfallsprodukte	Wahrscheinlichkeit	Zerfallsmöglichkeiten
elektrisch geladene Leptonen	$10,099 \pm 0,0191 \%$	$e^- e^+, \mu^- \mu^+, \tau^- \tau^+$
Neutrinos	$20,000 \pm 0,055 \%$	$\nu_e \bar{\nu}_e, \nu_\mu \bar{\nu}_\mu, \nu_\tau \bar{\nu}_\tau$
Quarks	$69,911 \pm 0,056 \%$	$u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}, b\bar{b}, t\bar{t}$

Tabelle 6.1: Die häufigsten Zerfallskanäle des Z-Bosons und ihre jeweiligen Wahrscheinlichkeiten, Daten nach Particle Data Group [3]

Übrig bleiben also die zwei ausgewählten Zerfälle: In Elektron und Positron und in Myon und Antimyon. Die Feynman-Diagramme für diese Zerfälle sind in den folgenden Abbildungen (6.1 und 6.2) dargestellt.

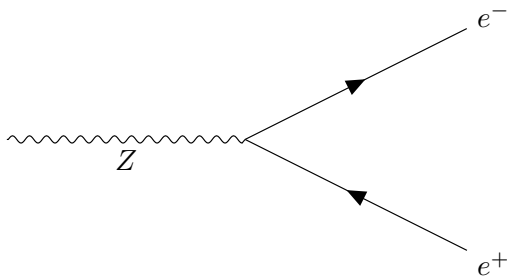


Abbildung 6.1: Feynman-Diagramm des Zerfalls eines Z-Bosons in ein Elektron und ein Positron (Paarerzeugung)

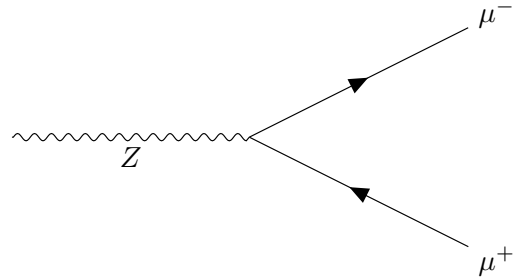


Abbildung 6.2: Feynman-Diagramm des Zerfalls eines Z-Bosons in ein Myon und ein Antimyon (Paarerzeugung)

6.3 Auswahl der Selektionskriterien

Bevor die gezielte Analyse des Z-Bosons anhand dieser beiden ausgewählten Ereignissen durchgeführt werden kann, müssen die Messergebnisse zuerst gefiltert werden, sodass möglichst viel Messuntergrund aussortiert werden kann und im besten Fall nur noch Signal übrig bleibt, was in der Realität aber nicht erreicht wird. Denn ein hoher Messuntergrund-Anteil verfälscht das Ergebnis der Messung. In diesem Fall könnte das zum Beispiel dazu führen, dass Ereignisse mit ausgewertet werden, bei denen gar kein Z-Boson beteiligt war (siehe Kapitel 6.1). Diese Selektion der Messdaten gelingt durch Selektionskriterien. Man muss also geeignete Anforderungen festlegen, nach denen entschieden wird, ob ein gemessenes Ereignis zur Auswertung verwendet werden soll oder nicht. Alle gemessenen Ereignisse werden dann nach und nach auf Erfüllung der festgelegten Kriterien geprüft und dementsprechend aussortiert.

Das Z-Boson zerfällt in jedem Fall in genau zwei Leptonen (siehe Kapitel 6.2). Daher wurde als erstes Kriterium festgelegt, dass bei dem Ereignis genau zwei Zerfallsprodukte gemessen werden.

Im zweiten Kriterium wird überprüft, dass die elektrischen Ladungen der beiden entstehenden Teilchen nicht gleich ist. Das muss auch so sein, da die Ladungserhaltung sonst nicht eingehalten wird. Das Z-Boson ist elektrisch neutral geladen. Daher müssen auch die elektrischen Ladungen aller Zerfallsprodukte sich zu 0 addieren. Dafür gibt es bei zwei Teilchen auch genau zwei Möglichkeiten: Entweder beide sind ebenfalls elektrisch neutral, aber dann wären die Zerfallsprodukte Neutrinos, was nicht möglich ist da Neutrinos im ATLAS-Detektor nicht gemessen werden können (siehe 6.2). Oder die Teilchen haben den gleichen Ladungsbetrag aber unterschiedliche Vorzeichen. Daher muss das zweite Kriterium überprüfen, ob die Beträge der Ladungen der beiden entstandenen Teilchen gleich sind, aber der eine negativ, der andere positiv ist.

Bei den ausgewählten Zerfällen entstehen immer ein Teilchen und das zugehörige Antiteilchen (Elektron und Positron oder Myon und Antimyon). Die Zerfallsprodukte sind also immer vom gleichen Typ (nach dem Particle Data Group „Particle Numbering Scheme“). Daher sollen solche

Kombinationen an Zerfallsprodukten wie Elektron und Antimyon nicht mit analysiert werden. Sie werden aber durch die ersten beiden Kriterien nicht aussortiert. Daher wird ein drittes Kriterium für den Teilchentyp verwendet. Dabei wird geprüft, ob die beiden entstehenden Teilchen vom gleichen Typ sind. Elektron und Positron werden ebenso wie Myon und Antimyon als Teilchen des gleichen Typs angesehen. Somit werden solche unerwünschten Kombinationen aussortiert.

6.4 Die invariante Masse und die M_{ll} -Verteilung

Die invariante Masse ist definiert als die Ruheenergie im Schwerpunktsystem: $M_{inv} := E_0$ und ihre Einheit ist damit auch Elektronenvolt. Wie auch die Ruheenergie ist sie für jedes System unabhängig vom Bezugssystem gleich, da sie die Energie im Schwerpunktsystem ist (siehe Kapitel 2.6).

Die invariante Masse ist außerdem ebenso wie die Ruheenergie eine Erhaltungsgröße. Das heißt, nach und vor der Teilchenkollision ist die invariante Masse des Systems gleich. Die invariante Masse eines Ausgangsteilchens ist also die Gesamtenergie, die in entstehende Teilchen umgesetzt werden kann. Oder andersherum: Mithilfe der Energien und Impulse aller Zerfallsprodukte kann die Masse des Ausgangsteilchens berechnet werden.

Das ist möglich durch die relativistische Beziehung von Energie, Masse und Impuls:

$$E^2 = m^2 + |\vec{p}|^2 \quad (6.1)$$

Es ergibt sich als Formel für die Masse des zerfallenen Teilchens:

$$m_{inv} = \sqrt{E^2 - |\vec{p}|^2} = \sqrt{E^2 - p_x^2 - p_y^2 - p_z^2} \quad (6.2)$$

wobei sich der Impuls p aus den drei Komponenten p_x, p_y, p_z zusammensetzt. Und für die Berechnung der invarianten Masse aus den Zerfallsprodukten (Teilchen 1 und Teilchen 2) müssen ihre Einzelenergien (E_1 und E_2) und -impulse (p_1 und p_2) addiert werden:

$$m_{inv} = \sqrt{(E_1 + E_2)^2 - (p_{x1} + p_{x2})^2 - (p_{y1} + p_{y2})^2 - (p_{z1} + p_{z2})^2} \quad (6.3)$$

Für jedes gemessene Ereignis kann nach dieser Formel die invariante Masse des Systems ermittelt werden. Nach der Selektion (siehe Kapitel 6.3) kann so eine Statistik über die invarianten Massen der Ereignisse erstellt werden. Das geschieht in Form eines Histogramms, bei dem die möglichen Ausprägungen der invarianten Masse in Bins unterteilt werden und die Anzahl der Ereignisse (oder: Events) in jedem Bin gemessen wird. Die resultierende M_{ll} -Verteilung ist in Abbildung 6.3 dargestellt.

Die farbigen Balken in diesem Histogramm zeigen erwartete Werte für verschiedene Zerfallskanäle des Z-Bosons. Diese Werte wurden nach der Theorie des Standardmodells der Teilchenphysik vom Computer berechnet. Bei dieser Simulation werden verschiedene Zerfallsmöglichkeiten des Z-Bosons simuliert, deren Häufigkeiten durch die verschiedenen Farben dargestellt sind. Der blaue Anteil der Balken stellt zum Beispiel den Zerfall in ein Elektron-Positron-Paar, der grüne den Zerfall in ein Myon-Antimyon-Paar dar. Wie im Histogramm ersichtlich ist, dominieren diese beiden Zerfallsmöglichkeiten deutlich. Andere ebenfalls angesprochene Zerfallskanäle, die nur schwer messbar oder auswertbar sind, (siehe Kapitel 6.2) wie der Zerfall in ein Tauon-Antitauon-Paar oder ein Top-Antitop-Quarkpaar machen hier nur einen so geringen Anteil an den simulierten Daten aus, dass sie im Histogramm nicht zu erkennen sind.

Die schwarzen Punkte im Histogramm sind die eigentlich gemessenen Werte aus dem Experiment. Die Kreuze symbolisieren die Binbreite (waagerechte Linien) und die Unsicherheit der Werte in den Bins (senkrechte Linien), für Bins mit mehr Events ist die Unsicherheit erwartungsgemäß höher (relativ zur Eventanzahl ist sie aber dennoch niedriger). Da in den Messdaten der Zerfallskanal nicht eindeutig ersichtlich ist, werden sie nicht wie bei den Simulationsdaten aufgeteilt dargestellt. In einem weiteren Diagramm unter dem Histogramm ist das Verhältnis aus Messdaten zu simulierten Daten dargestellt. Wenn diese genau übereinstimmen würden, lägen diese Werte also alle bei 1. Durch statistische Schwankungen liegen sie hier um 1 verteilt. Auch hier ist die Abweichung vom Optimum (1) im Bereich, in dem es viele Events gibt, geringer, die Daten stimmen

also besser mit den Simulationen überein. Das liegt daran, dass die statistischen Schwankungen bei mehr Ereignissen besser ausgeglichen werden. Für eine erste Übersicht werden hier vereinfacht nur 0,01 % der vorhandene Daten im Histogramm dargestellt.

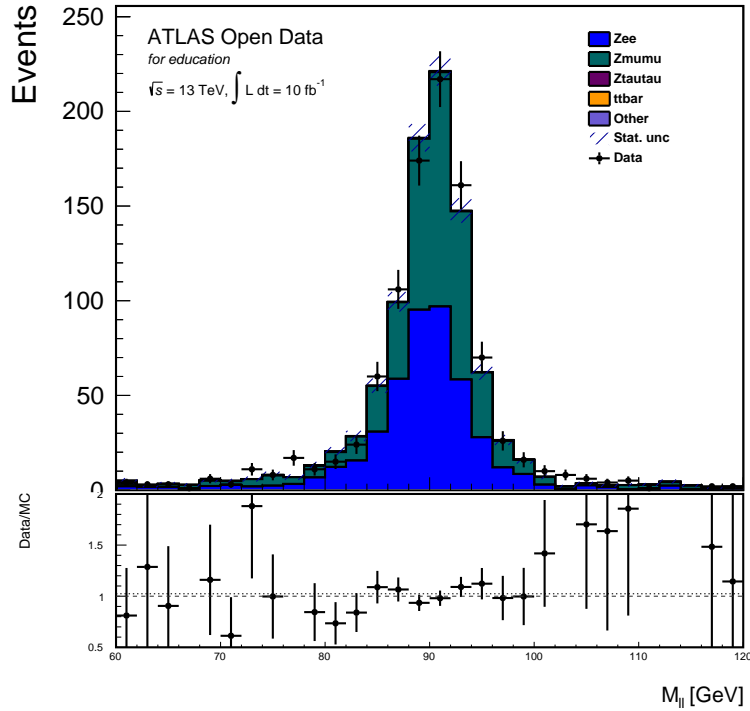


Abbildung 6.3: Histogramm über die invariante Masse des gemessenen Systems mit 0,01 % der Daten, Bin-Breite: 2 GeV

6.5 Fitting der Normalverteilung als Wahrscheinlichkeitsdichtefunktion

Es existieren nun absolute Werte, die eine Aussage über die Wahrscheinlichkeit der möglichen Massen des Z-Bosons treffen. Aus diesem Histogramm kann man gut die mittlere Masse des Z-Bosons abschätzen, die etwa 88 bis 92 GeV beträgt, da dort der Mittelwert der Verteilung liegt. Dieses Ablesen ist aber sehr ungenau, daher wird eine weniger fehlerbehaftete Möglichkeit der Bestimmung gesucht. Dafür eignet sich am besten eine Dichtefunktion, welche die Masse des Z-Bosons beschreibt. Sie hat den Vorteil, dass man aus ihr leicht Werte ablesen kann, wie zum Beispiel auch den Mittelwert, der der ermittelten mittleren Masse des Z-Bosons entspricht. Es muss also erst eine Dichtefunktion festgelegt werden, welche die Masseverteilung beschreibt, und diese muss dann mittels des Minimum-Chiquadrat-Fits an die Werte angeglichen werden (siehe Kapitel 4.3.1).

In allen weiteren Schritten des Fittings wurde die Fitting-Funktion zuerst an die simulierten Daten (siehe Kapitel 6.4) angeglichen, da dies einfacher ist. Erst im letzten Schritt, wenn der komplette Vorgang des Fittings optimiert ist, wurden dazu die eigentlich gemessenen Werte verwendet, um so auch einen Wert für die Z-Boson-Masse aus dem Experiment zu ermitteln.

Des Weiteren wurde die Anzahl der Bins in den Histogrammen erhöht und somit die Binbreite verringert, damit die Werte genauer die wirklichen Wahrscheinlichkeiten der Binmittelpunkte beschreiben (siehe Kapitel 4.3.1). Als Binanzahl wurde 300 gewählt, damit ergibt sich eine Binbreite von $\Delta x = 0,2$ GeV. Auch die Gesamtzahl der Ereignisse, die zur Analyse standen, wurde auf 100 % der verfügbaren Daten erhöht. Es ergibt sich ein neues Histogramm (siehe Abbildung 6.4), an das nun die Fitting-Funktion angeglichen werden soll. Auf der y-Achse sind bei diesem Diagramm die

Anzahl der Ereignisse mit dem Faktor 10^3 angegeben, was an der Vergrößerung der Datenmenge liegt. Auch sind die Daten nur noch als Punkte zu erkennen. Wie man im Diagramm unter dem Histogramm erkennt, liegen die Daten nun auch wie erwartet näher an den simulierten Werten.

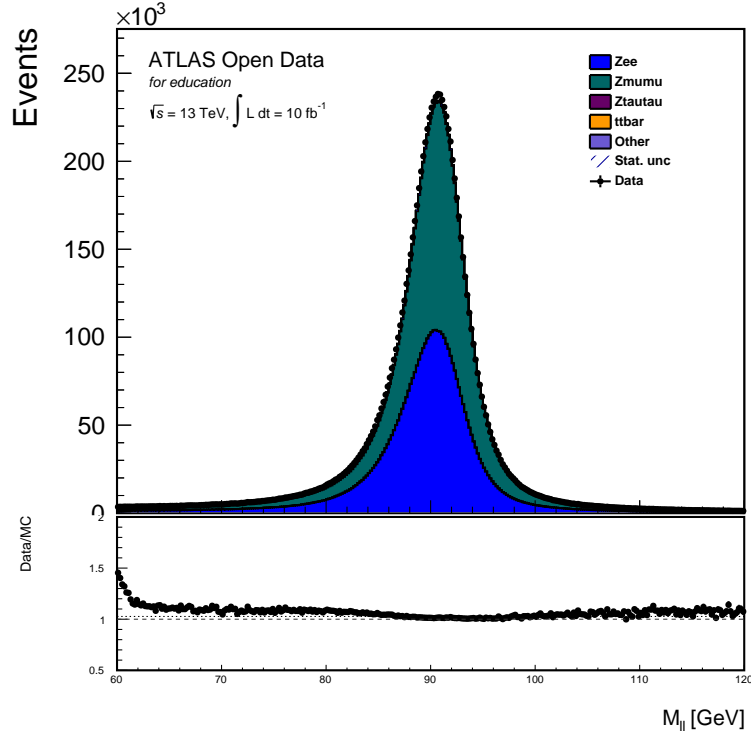


Abbildung 6.4: Histogramm über die invariante Masse des gemessenen Systems mit erhöhter Binanzahl und Datenmenge, an dieses Histogramm soll die Fitting-Funktion angeglichen werden

6.5.1 Auswahl der Fitting-Funktion

Zum Beschreiben von Zufallsexperimenten, deren Ergebnisse in Form eines Peaks um einen Mittelwert gestreut sind, wie es bei der $M_{||}$ -Verteilung der Fall ist, eignet sich besonders gut die Gauß-Funktion oder Normalverteilung. Deren Dichtefunktionsgleichung ist (siehe Formel 4.6):

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6.4)$$

Um mit der Funktion die Werte des Histogramms beschreiben zu können, muss die Funktion mit einem zusätzlichen Vorfaktor skaliert werden. Dabei kann auch gleich der Vorfaktor der Funktion $\frac{1}{\sqrt{2\pi}\sigma}$ durch einen neuen Skalierungsfaktor n ersetzt werden:

$$f(x) = n \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6.5)$$

Da diese Funktion jetzt mathematisch gesehen aber keine Dichtefunktion mehr ist, nennt man sie Fitting-Funktion. Sie soll in den folgenden Schritten an die Werte angeglichen werden.

Bei den Messwerten muss noch der Messuntergrund beachtet werden, der die Ergebnisse verfälscht und möglicherweise nicht mit analysiert werden soll (siehe Kapitel 6.1). Der Messuntergrund, „hebt“ die Werte in allen Bins an, wie man im Histogramm erkennen kann: Auch außerhalb des zu erwartenden Bereichs der Masse von etwa 80 bis 100 GeV sinken die Werte in den Bins nicht auf 0. Damit der Anteil des Messuntergrunds an den Daten nicht mit ausgewertet wird, muss auch die Fitting-Funktion etwas nach oben verschoben werden, sodass der Messuntergrund keine Auswirkung auf

das Fitting der eigentlichen Normalverteilung hat. Diese Verschiebung der Fitting-Funktion in y-Richtung erreicht man durch die Addition einer zusätzlichen Konstante b:

$$f(x) = n \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} + b \quad (6.6)$$

Dies ist die endgültige Funktion, welche an die Werte angeglichen wurde.

6.5.2 Schätzung der Fitting-Parameter

Die Funktion kann nun numerisch an die Werte angeglichen werden. Dazu wird der Minimum-Chiquadrat-Fit angewendet (siehe Kapitel 4.3.2) und die ermittelten Parameter der Fitting-Funktion werden als Ausgabewerte zurückgegeben.

Vorerst müssen aber noch die Parameter geschätzt werden, da zur numerischen Ermittlung der Fitting-Funktion immer Anfangswerte nötig sind. Sonst kann mittels des numerischen Verfahrens kein Optimum für die Parameter gefunden werden. Dieses Abschätzen der Parameter kann leicht anhand des Histogramms erfolgen.

Der Mittelwert der Verteilung beträgt etwa 90 GeV. Auch die Höhe des Peaks ist aus dem Histogramm leicht ablesbar: etwa 250 000. Als Höhenverschiebung wurde erst einmal 0 angenommen, da sie aus dem Histogramm schwer abschätzbar ist und nur knapp über 0 liegt. Der einzige Wert, der nur schwer abgeschätzt werden kann, ist die Streuung σ um den Mittelwert μ . Im Intervall $[\mu - \sigma, \mu + \sigma]$ befinden sich nach der Definition der Normalverteilung etwa 68 % der Ereignisse, das ist im Histogramm etwa bei einem Wert von 4 für σ der Fall. Nach der Schätzung wurde also festgelegt:

$$n = 250\,000; \mu = 90; \sigma = 4; b = 0 \quad (6.7)$$

6.5.3 Chi-Quadrat-Test

Um die Qualität des Fits (die Anpassung der Funktion an die Daten) zu beurteilen, wurde der Chi-Quadrat-Test eingeführt, der genau den Grad der Anpassung bestimmt. Er berechnet den χ^2 -Wert aus den Daten und der Funktion (siehe Kapitel 4.3.2). Das Problem ist, dass man allein durch den quantitativen Wert keine Aussage über die Qualität des Fits treffen kann, dazu benötigt man den p-Wert, den der Chi-Quadrat-Test ebenfalls ausgibt. Dieser Wert gibt an, wie hoch die Wahrscheinlichkeit ist, bei der Fitting-Funktion einen solchen oder noch höheren χ^2 -Wert zu erhalten. Oder anders ausgedrückt wie wahrscheinlich es ist, dass gefittete Funktion und Werte übereinstimmen. Berechnet wird der p-Wert mittels des durch den Chi-Quadrat-Test berechneten χ^2 -Wertes x und der χ^2 -Dichtefunktion für n Freiheitsgrade der Fitting-Funktion $f(\chi^2|n)$:

$$p = \int_x^\infty f(\chi^2|n) \quad (6.8)$$

Im optimalen Fall, wenn die Werte mit der Fitting-Funktion exakt übereinstimmen, ist der χ^2 -Wert 0 und der p-Wert 1. Das ist aber unrealistisch und eher ein Zeichen dafür, dass der Chi-Quadrat-Test seine Aussagekraft verloren hat, zum Beispiel weil die Fitting-Funktion zu speziell gewählt wurde (Overfitting) oder die Anzahl der Bins, an die gefittet wird, zu gering ist. Um das zu vermeiden, sollte man nicht nur auf den p-Wert achten, der möglichst nahe bei 1 liegen soll, sondern auch auf den χ^2 -Wert, den der Chi-Quadrat-Test ausgibt. Der sollte erwartungsgemäß in etwa in der Nähe der Anzahl der Freiheitsgrade (engl.: degrees of freedom) der Fitting-Funktion liegen, da dies genau der Erwartungswert der jeweiligen Chi-Quadrat-Funktion ist. Das erkennt man auch ungefähr bei der χ^2 -Dichtefunktion (siehe Abbildung 4.3). Das heißt, auch der Quotient $\frac{\chi^2}{N_{dof}}$ sollte etwa bei 1 liegen. Die Anzahl der Freiheitsgrade N_{dof} wird bestimmt durch die Anzahl der Datenpunkte N_d und der Anzahl der Parameter N_p der Funktion: $N_{dof} = N_d - N_p$.

Theoretisch wird der χ^2 -Wert berechnet über die Werte in den Bins und der Funktion und der Varianz der Werte in den Bins:

$$\chi^2 = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{\sigma_i^2} \quad (6.9)$$

Das Problem bei diesem Vorgehen ist, dass die χ^2 -Werte beim Fitting auf Simulationsdaten dann unrealistisch groß werden. Das liegt an der Berechnung der Simulationsdatenwerte. Dabei werden viel mehr auf theoretischer Grundlage zufällig erzeugte Daten (Pseudodaten) erzeugt, als tatsächliche Messdaten des Experiments vorhanden sind, um statistische Effekte auszugleichen und eine höhere Genauigkeit zu garantieren. Auf diese große Datenmenge wird die Unsicherheit bestimmt und dann werden Simulationsdatenwerte und deren Unsicherheiten auf die niedrige Menge der Messdaten herunter gerechnet. Dabei entstehen aber Datenwerte für die Simulationsdaten mit enorm kleinen Unsicherheiten, da die berechneten Unsicherheiten durch hohe Divisoren geteilt werden, um die Datenmenge zu reduzieren. Aufgrund der nun sehr kleinen Unsicherheiten auf den Datenwerten ergeben sich sehr große χ^2 -Werte, die damit ihre Aussagekraft verlieren. Wenn man den χ^2 -Wert auswerten will, muss man daher sinnvolle Unsicherheiten für die Simulationsdatenwerte festlegen. Wenn man die üblicherweise verwendete Näherung ansetzt und die Unsicherheit als die Wurzel aus den Werten definiert ($\sigma = \sqrt{y_i}$). Dann erhält man nach der komplexen Berechnung des χ^2 -Wertes

$$\chi^2 = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{\sigma_i^2} = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{\sqrt{y_i}^2} = \sum_{i=1}^{n_B} \frac{(y_i - f(x_i))^2}{y_i}, \quad (6.10)$$

eine vereinfachte Form der Berechnung des Chi-Quadrat-Wertes (wie schon in Kapitel 4.3.2 aufgeführt). Es ergibt also alles in allem Sinn, für die Berechnung des χ^2 -Wertes bei Simulationsdaten trotz vorhandener Unsicherheiten die vereinfachte Berechnung des χ^2 -Wertes zu verwenden. Bei den Messdaten können die vorgegebenen Unsicherheiten aus den Daten verwendet werden, da diese mittels oben erwähnter Näherung ($\sigma = \sqrt{y_i}$) berechnet werden.

6.5.4 Anpassung des Fittings

Nach Schätzung der Parameter und Durchführung des Fittings erhält man eine Dichtefunktion, die an die Werte angeglichen ist. Die Fitting-Funktion ist im Histogramm (siehe Abbildung 6.5) als rote Linie dargestellt.

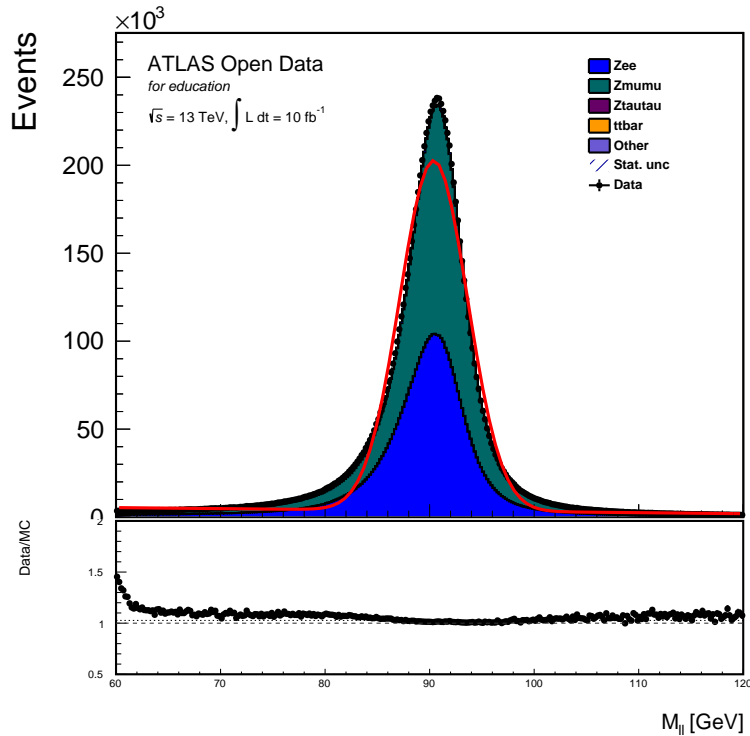


Abbildung 6.5: Histogramm mit angepasster Fitting-Funktion (an die Simulationsdaten)

Wie man aber aus dem Histogramm ablesen kann, beschreibt die Funktion die Werte nicht gut. Vor allem im Bereich von 75 bis 85 GeV gibt es starke Abweichungen. Viel wichtiger für die Ermittlung der mittleren Masse des Z-Bosons ist aber, dass die Funktion den Peak der Verteilung (also die Werte der Bins rund um den Erwartungswert) gut beschreibt. Da auch dies nicht erreicht ist, müssen weitere Anpassungen vorgenommen werden.

Zuerst wurde das Fitting auf ein sinnvolles Intervall begrenzt. Aus dem Histogramm und den Parametern der Fitting-Funktion ist erkennbar, in welchem Bereich die Masse des Z-Bosons etwa liegen wird. Außerhalb dieses relevanten Datenbereichs ist der Messuntergrundanteil in den Bins höher und würde so das Ergebnis verfälschen. Es ergibt also Sinn, nur den relevanten Datenbereich zu betrachten, der auf [85 bis 95 GeV] festgelegt wurde. Die Fitting-Funktion wird im Folgenden nur an diesen Bereich des Histogramms angeglichen. Daher wird in allen folgenden Histogrammen auch der dargestellte Datenbereich verkleinert auf [75 bis 105 GeV]. Wenn die Funktion dann später nach allen Verbesserungen des Fitting-Prozesses auf die eigentlichen Daten gefittet wird, kann dieser Datenbereich sogar noch weiter eingeschränkt werden, zum Beispiel auf [89 bis 92 GeV].

Eine weitere Möglichkeit liegt in der Wichtung der Bins (siehe Kapitel 4.3.2). Wenn die Unsicherheit der Werte in jedem Bin mit in den Prozess des Fittings einfließt, ergibt sich eine besser angepasste Fitting-Funktion. Eine letzte Verbesserung besteht noch darin, die Daten für das Fitting grundsätzlich anders zu ermitteln. Bisher wurden nur die Werte der Binmittelpunkte $y_i = y(x_i)$ als Wert für den gesamten Bin angenommen und die Fitting-Funktion daran gefittet. Das entspricht aber nicht der Realität (siehe Kapitel 4.3.1). Dies wurde nur unter der Voraussetzung einer hohen Binanzahl angenommen, um das Fitting leichter zu machen. Realistisch ist es so, dass die Fläche des Bins ($A_i = \Delta x \cdot y_i$) dem Integral über der Fitting-Funktion $f(x)$ von a_i bis b_i entspricht:

$$\Delta x \cdot y_i = \int_{a_i}^{b_i} f(x) dx \quad (6.11)$$

Daher muss die Fitting-Funktion auch mittels Integralen neu definiert werden. Da bisher an ein Histogramm mit vielen Bins kleiner Binbreite gefittet wurde, ist der Effekt dieser Verbesserung minimal. Dennoch wurde sie hier schon in das Analyseprogramm implementiert, da sie in späteren Analysen noch wichtig wird. Diese drei Veränderungen tragen zur Verbesserung der Anpassung der Funktion bei. Es ergibt sich eine neue Fitting-Funktion, dargestellt in Abbildung 6.6.

Nach allen Anpassungen kann die Fitting-Funktion dann zur abschließenden Analyse auch an die eigentlichen Werte des Experiments (schwarze Punkte im Histogramm) angeglichen werden.

6.5.5 Ergebnisse und Auswertung der Analyse mit Normalverteilung

Zuerst wurde die Fitting-Funktion an die Simulationsdaten angeglichen und dabei der Vorgang des Fittings verbessert. Den größten Einfluss auf die Werte der Funktion hat die Veränderung des Fitting-Bereichs von [60 bis 120 GeV] auf [85 bis 95 GeV]. Bei allen weiteren Änderungen im Fitting-Prozess hingegen verändern sich die Werte der Parameter nur minimal. Das liegt vor allem daran, dass die Gesamtanzahl der Ereignisse und die Anzahl der Bins schon von Anfang an so hoch gewählt wurde, dass die Genauigkeit in den einzelnen Bins schon sehr hoch war. Die letzten Änderungen sind also nur der Korrektheit wegen noch hinzugefügt worden, am Ergebnis verändern sie nicht viel.

Nach der Verbesserung des Fittings wurde die Funktion dann an die eigentlichen Werte angeglichen. Dabei ergaben sich für verschiedene Fitting-Bereiche folgende Ergebnisse für die Werte der Parameter, dargestellt in Tabelle 6.2.

Fitting-Bereich	n	μ	σ	b	χ^2 -Wert	p-Wert
[85 bis 95 GeV]	188193	90,55	2,18	49054	6845	0
[89 bis 92 GeV]	111460	90,71	1,61	126700	5,96	0,74

Tabelle 6.2: Parameter der Fitting-Funktion bei Angleichung an die Messdaten nach verschiedenen Fitting-Bereichen und jeweils die Werte des χ^2 -Tests zur Beurteilung der Anpassung des Fits an die Daten

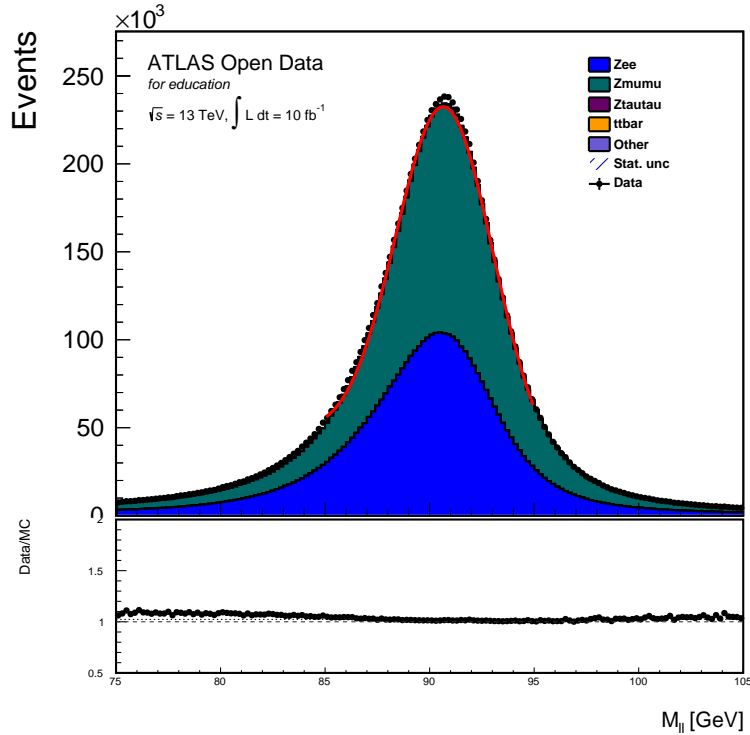


Abbildung 6.6: Histogramm mit angepasster Fitting-Funktion (Fitting nach allen Anpassungen und im Bereich [85 bis 95 GeV], Angleichung an die Simulationsdaten)

Hier lassen sich wieder Abweichungen zwischen den verschiedenen Werten der Parameter bei verschiedenen Fitting-Bereichen gut feststellen. Den gesuchten Wert der Analyse, die Masse des Z-Bosons, kann man nun leicht aus den Daten ablesen: Dies ist der Mittelwert μ der Fitting-Funktion. Auch kann man bei Verkleinerung des Fitting-Bereichs eine starke Verkleinerung des χ^2 -Wertes feststellen, was für eine bessere Anpassung des Fits an die Messdaten spricht. Beim kleinsten Fitting-Bereich ist der Wert so klein, dass sich ein p-Wert von 0,74 ergibt, was sehr gut ist. Zur Definition des p-Wertes siehe Kapitel 6.5.3.

Aus den Werten, vor allem dem Wert für den Parameter b , ergibt sich aber auch eindeutig ein Problem des derzeitigen Fittings. Der Parameter b beschreibt den linearen Untergrund, der der Normalverteilung hinzugefügt wird (siehe Kapitel 6.5.1), der die Funktion also entlang der y-Achse verschiebt. Bei Verkleinerung des Fitting-Bereichs ist vor allem ein starker Anstieg dieses Wertes zu bemerken. Die Funktion wird also zwar besser an die Werte im Fitting-Bereich angepasst, beschreibt aber den Rest der Werte überhaupt nicht, da sie im Fitting unrealistisch hoch in y-Richtung verschoben wird. Denn im Histogramm ist zu erkennen, dass der Messuntergrund nicht über 10 000 ist, daher dürfte auch b nicht größer als 10 000 sein.

Da man bei der Analyse generell an einem Ergebnis interessiert ist, das alle Werte mehr oder weniger stark mit berücksichtigt und nicht nur einen eingeschränkten Bereich der Messung beschreibt, ist das Fitting mit der Normalverteilung physikalisch gesehen wenig brauchbar, auch wenn der ermittelte Wert für die mittlere Masse des Z-Bosons von 90 bis 91 GeV nicht weit vom tatsächlichen Wert $91,1876 \pm 0,0021$ GeV [3] entfernt liegt.

Eine notwendige Maßnahme zur Verringerung dieser Ungenauigkeit des Fittings besteht darin, den Parametern beim Fitting-Prozess Grenzen zu setzen. Man kann also der Optimierungsfunktion vorgeben, dass der Parameter b nur zwischen 0 und 10 000 liegen darf. Das ist nötig, weil im Histogramm zu erkennen ist, dass der Messuntergrund nicht höher liegt. Um die Funktion nicht einfach nur nach oben zu verschieben (was physikalisch keinen Sinn ergeben würde), müssen diesem Parameter Grenzen gesetzt werden. Dies ist aber nicht hinreichend, da dann die Fitting-Funktion die Werte nicht optimal beschreiben kann (was man an einem sehr viel höherem χ^2 -Wert

erkennt). Daraus lässt sich schließen, dass sich die Normalverteilung nicht optimal eignet, um diesen Prozess zu beschreiben. Daher mussten neue Ansätze gesucht werden, vor allem geeignetere Fitting-Funktionen für die Analyse, um die Anpassung an die Werte zu optimieren.

6.6 Fitting anderer Wahrscheinlichkeitsdichtefunktionen

Die Normalverteilung ist offensichtlich nur bedingt geeignet, das Experiment zu beschreiben. Daher wurde der Prozess des Fittings mit allen Verbesserungen beibehalten und nur besser geeignete Funktionen gesucht.

6.6.1 Die Breit-Wigner-Verteilung

Eine geeignete Funktion zur Beschreibung der Ergebnisse ist die Breit-Wigner- oder Lorentz-Verteilung. Ihre Dichtefunktion ist [7]:

$$b(x) = \frac{\Gamma}{2\pi} \frac{1}{(x - \mu)^2 + (\frac{\Gamma}{2})^2} \quad (6.12)$$

mit der Zerfallsbreite Γ , die Auskunft über die Lebensdauer des Z-Bosons gibt und sich in der Funktion in der Breite der Verteilung äußert, wie σ in der Normalverteilung. Die Breit-Wigner-Verteilung, als Vergleich mit Normalverteilung gleicher Parameter, ist dargestellt in Abbildung 6.7.

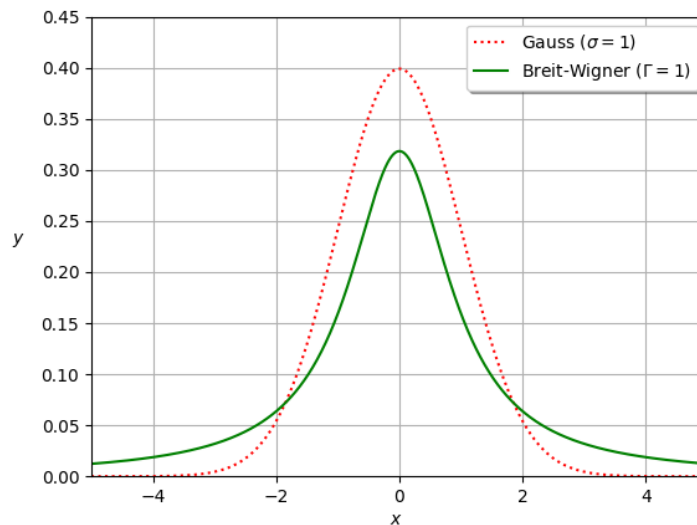


Abbildung 6.7: Normalverteilung und Breit-Wigner-Verteilung für gleiche Parameter

Wie vorher schon bei der Normalverteilung (siehe Kapitel 6.5.1) wurden weitere Parameter hinzugefügt: Die Normierung n und den Untergrund, hier a und b als linearer Untergrund, der eine Asymmetrie der Funktion ermöglicht und sie somit seitlich „kippt“:

$$b(x) = n \cdot \frac{1}{(x - \mu)^2 + (\frac{\Gamma}{2})^2} + ax + b \quad (6.13)$$

Wenn man das Fitting mit dieser Funktion durchführt, erhält man eine andere angeglichene Funktion, dargestellt in Abbildung 6.8.

Auch diese Funktion beschreibt die Werte nicht optimal, wie man leicht im Histogramm (Abbildung 6.8) erkennen kann. Vor allem an den Grenzen und der Mitte des Fitting-Bereichs weicht die Funktion von den Werten ab. Den Grad der Anpassung gibt der Chi-Quadrat-Test an (siehe

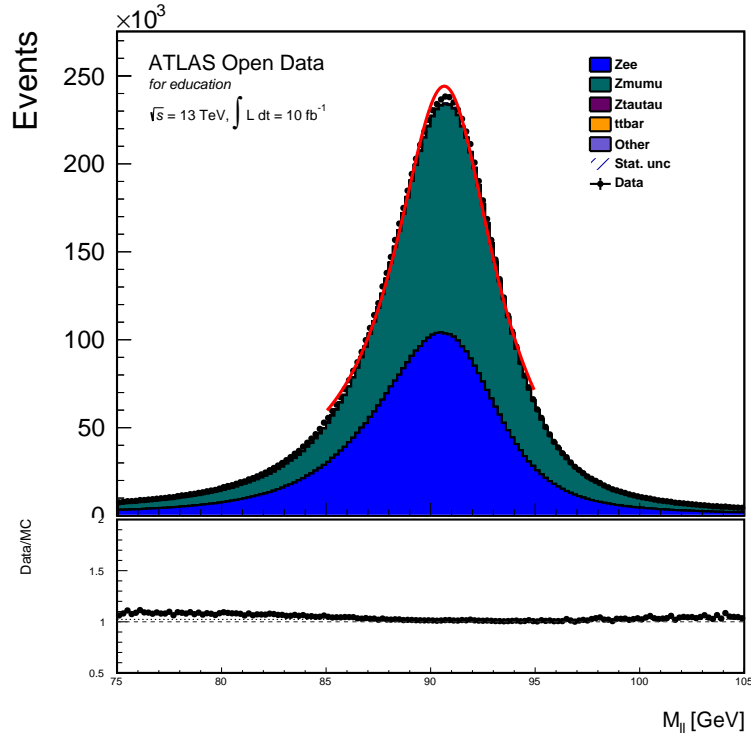


Abbildung 6.8: An die Messdaten angeglichene Breit-Wigner-Verteilung über dem Fitting-Bereich [85 bis 95 GeV]

Kapitel 6.5.3). In der Tabelle 6.3 sind die Parameter der Funktion und χ^2 -, $\frac{\chi^2}{N_{dof}}$ - und p-Wert jeweils für die Normalverteilung und Breit-Wigner Verteilung für Simulationsdaten und Messdaten aufgeführt.

Fitting-Funktion	Daten	n	μ	a	b	σ bzw. Γ	χ^2	$\frac{\chi^2}{N_{dof}}$	p-Wert
Normalverteilung	Simulation	214633	90,51	-	10000	2,73	24254	551,2	0
Normalverteilung	Messdaten	217519	90,48	-	10000	2,73	26779	608,6	0
Breit-Wigner-Verteilung	Simulation	241378	90,67	-581	0	5,85	6371	144,8	0
Breit-Wigner-Verteilung	Messdaten	244794	90,67	-907	0	5,89	6123	139,1	0

Tabelle 6.3: Parameter der Fitting-Funktionen und χ^2 -, $\frac{\chi^2}{N_{dof}}$ - und p-Wert für Normalverteilung und Breit-Wigner-Verteilung im Fitting-Bereich [85 bis 95 GeV]

In dieser Tabelle (6.3) kann man wieder das Problem der Normalverteilung erkennen, wodurch beim Fitting die Funktion auf der y-Achse nach oben verschoben wird. Durch die Begrenzung des Parameters beträgt der Untergrund nur die eingestellten maximalen 10 000 GeV. Wie ebenfalls zu erkennen ist, ist das Problem bei der Breit-Wigner-Verteilung genau umgekehrt: Der Parameter b ist unrealistisch klein und wird nur durch die untere Grenze von 0 GeV begrenzt. Beide Funktionen beschreiben sowohl Simulations- als auch Messdaten nicht gut. Das kann man an den hohen χ^2 -Werten und den verschwindend kleinen p-Werten (die optimalerweise bei 1 liegen und hier so klein sind, dass sie auf 0 gerundet wurden) erkennen. Auch die $\frac{\chi^2}{N_{dof}}$ -Werte sind noch zu groß. Alle diese drei Werte sind zwar besser als bei dem Normalverteilungsfitt aber noch nicht ausreichend. Die Breit-Wigner-Verteilung allein ist also für die Beschreibung des Experiments ebenfalls nicht gut geeignet, deshalb wurde weiter nach einer geeigneteren Funktion gesucht.

6.6.2 Die Voigt-Verteilung

Die Wahrscheinlichkeiten für die Massen des Z-Bosons lassen sich theoretisch mit der Breit-Wigner-Verteilung beschreiben. In der Messung wird dies aber von statistischen Störungen verfälscht, was der Grund dafür ist, dass die Werte dennoch nicht gut durch die Breit-Wigner-Verteilung beschrieben werden könnten. Diese statistische Störungen verändern das Ergebnis der Messung jedes der Ereignisse und schwanken um 0. Als Messergebnis erhält man also den theoretischen Wert plus eine Störgröße, die aus den statistischen Schwankungen bei der Messung von jedem einzelnen Ereignis resultiert. Mathematisch kann man dies mithilfe der Faltung von Funktionen beschreiben. Die theoretischen Werte folgen der Breit-Wigner-Verteilung und die zufälligen Störungen der Normalverteilung. Die Daten beschreibt also die Faltung der Breit-Wigner- mit der Normalverteilung, die auch Voigt-Verteilung genannt wird. [7]

Da sie die Faltung von Breit-Wigner- und Normalverteilung ist, ist sie abhängig von den Parametern μ , σ , Γ , hier erweitert durch die Normierung n und die Untergrundparameter a und b . Auf die spezielle Berechnung der Voigt-Verteilung oder ihre Funktionsgleichung wird aber hier aufgrund der Komplexität nicht eingegangen. In der folgenden Grafik (siehe Abbildung 6.9) ist aber ein Vergleich zwischen den drei verwendeten Fitting-Funktionen abgebildet. Der Parameter σ gibt nun an, wie groß die statistische Schwankung des Messwertes beim Messen jedes einzelnen Ereignisses ist. Ist dieser Parameter 0, ergibt sich wieder eine Breit-Wigner-Verteilung.

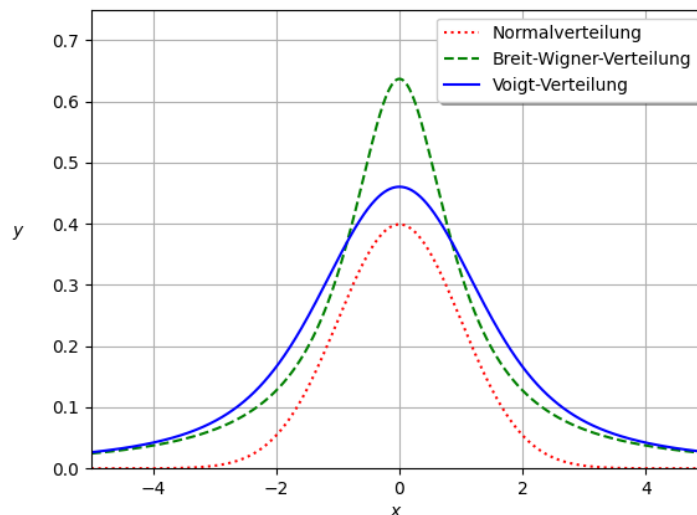


Abbildung 6.9: Normalverteilung, Breit-Wigner-Verteilung und Voigt-Verteilung (für verschiedene Parameter), Breit-Wigner-Verteilung und Voigt-Verteilung sind keine Dichtefunktionen

Nach Festlegung von sinnvollen Startwerten und Grenzen für die einzelnen Parameter der Voigt-Verteilung ergab sich eine angepasste Fitting-Funktion (siehe Abbildung 6.10).

Wie im Histogramm ersichtlich ist, ist diese Fitting-Funktion im Vergleich zu den anderen beiden Funktionen (Normalverteilung und Breit-Wigner-Verteilung) sehr gut an die Daten angepasst. Zum Vergleich sind wieder die Parameter der Fitting-Funktion und die Werte des χ^2 -Tests aufgeführt (siehe Tabelle 6.4).

Aus diesen Werten kann eine Aussage über die Genauigkeit der Anpassung der Fitting-Funktion an die Simulations- und Messdaten getroffen werden. Wie aus den Werten des χ^2 -Tests zu erkennen ist, ist die Anpassung bei der Voigt-Verteilung viel besser als bei den beiden vorherigen Fitting-Funktionen (Normalverteilung und Breit-Wigner-Verteilung), was daran liegt, dass sie das Experiment besser beschreiben kann. Dennoch sieht man auch, dass die χ^2 - und $\frac{\chi^2}{N_{doF}}$ -Werte immer noch so hoch sind, dass der p-Wert bei nahezu 0 liegt. Es kann also von einer besseren, aber nicht guten Anpassung an die Daten gesprochen werden. Dass selbst mit der Voigt-Verteilung nur so vergleichsweise hohe χ^2 -Werte erreicht werden können, liegt daran, dass das Experiment und die

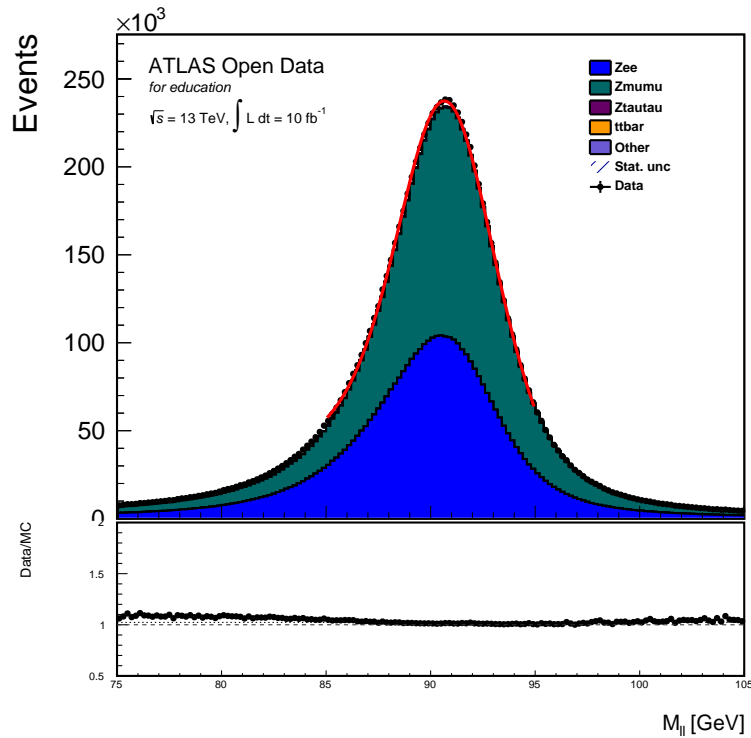


Abbildung 6.10: An die Messdaten angeglichene Voigt-Verteilung über dem Fitting-Bereich [85 bis 95 GeV]

Fitting-Funktion	Daten	n	μ	a	b	σ	Γ	χ^2	$\frac{\chi^2}{N_{dof}}$	p-Wert
Voigt-Verteilung	Simulation	2024773	90,72	-2287	0	1,46	2,08	487	11,3	0
Voigt-Verteilung	Messdaten	2074913	90,72	-2553	0	1,43	2,13	718	16,7	0

Tabelle 6.4: Parameter der Fitting-Funktionen und χ^2 -, $\frac{\chi^2}{N_{dof}}$ - und p-Wert für die Voigt-Verteilung im Fitting-Bereich [85 bis 95 GeV]

daraus erhobenen Daten noch vielen hier nicht beachteten Einflüssen unterliegen, wodurch es die Voigt-Verteilung nicht optimal beschreiben kann.

Eine letzte, schon in Kapitel 6.5.5 angesprochene Möglichkeit besteht noch darin, den Fitting-Bereich, über dem die Funktion an die Daten angeglichen wird, zu verkleinern. Bisher ist dieser Bereich auf [85 bis 95 GeV] festgelegt. Da der Peak offensichtlich um 91 GeV liegt (Mittelwert der Verteilungen), erscheint es sinnvoll, den Fitting-Bereich auf 91 ± 4 GeV festzulegen, was nur eine geringfügige Verkleinerung des Bereiches (von 10 auf 8 GeV) ist. Es ergeben sich neue Fitting-Werte (siehe Tabelle 6.5).

Fitting-Funktion	Daten	n	μ	a	b	σ	Γ	χ^2	$\frac{\chi^2}{N_{dof}}$	p-Wert
Voigt-Verteilung	Simulation	1887416	90,77	-3101	10000	1,37	2,04	36,1	1,10	0,33
Voigt-Verteilung	Messdaten	1948398	90,77	-3476	10000	1,31	2,14	121	3,67	$5,9 \cdot 10^{-12}$

Tabelle 6.5: Parameter der Fitting-Funktionen und χ^2 -, $\frac{\chi^2}{N_{dof}}$ - und p-Wert für die Voigt-Verteilung im Fitting-Bereich [87 bis 95 GeV]

Aus den Werten in dieser Tabelle kann man erkennen, dass schon diese geringfügige Verkleinerung des Fitting-Bereiches die Anpassung deutlich verbessert hat. Die Voigt-Verteilung kann nun

die Simulationsdaten gut beschreiben, wodurch sich ein hoher p-Wert und ein $\frac{\chi^2}{N_{dof}}$ -Wert nahe 1 ergibt. Bei den Messdaten ist die Anpassung weiterhin kleiner aus oben schon genanntem Grund: Im Experiment beeinflussen viele weitere Faktoren die Ergebnisse, die hier nicht mit berücksichtigt wurden. Zur noch genaueren Beschreibung der Messdaten müsste also eine noch komplexere Analyse durchgeführt werden, für diese Arbeit ist die erreichte Genauigkeit aber völlig ausreichend.

6.7 Endergebnisse und Auswertung der kompletten Analyse

Damit wurde also durch besser geeignete Fitting-Funktionen die Genauigkeit der Anpassung von Fitting-Funktionen an die Werte um ein Vielfaches optimiert. Das Problem bei vorhergehenden Fitting-Funktionen war, dass sie die Werte des Experiments nicht gut beschreiben. Sobald der Bereich, für den die Funktion eine Aussage treffen soll, auf eine sinnvolle Größe eingestellt wird, sinkt die Genauigkeit enorm. Durch die Voigt-Verteilung, die statistische Schwankungen der Messergebnisse berücksichtigt, wurde die Genauigkeit des Fits verbessert. Auch unsinnige Verschiebungen der Funktionen in y-Richtung kommen somit nicht mehr vor. Darum kann man jetzt aus den Parametern der Voigt-Funktion Werte wie die invariante Masse des Z-Bosons ablesen, deren Ermittlung Ziel der Analyse war. Dabei äußert sich ein weiterer Vorteil der Voigt-Verteilung im Gegensatz zur Normalverteilung: Man kann direkt aus ihr die Zerfallsbreite Γ des Z-Bosons ablesen, die dem Parameter Γ entspricht. Die Zerfallsbreite ist ebenfalls wie die mittlere Masse eine charakteristische Eigenschaft des Z-Bosons. Aus der Zerfallsbreite kann außerdem mittels folgender Gleichung die mittlere Lebensdauer τ eines Teilchens berechnet werden:

$$\tau = \frac{\hbar}{\Gamma} \quad (6.14)$$

Durch die Voigt-Verteilung wurden die statistischen Effekte, die die Messungen verfälschen, nicht nur berücksichtigt, sondern auch quantitativ bestimmt. Der Parameter σ entspricht den statistischen Schwankungen der Messwerte. Da die Schwankungen durch die Normalverteilung beschrieben werden, kann daher die Standardabweichung für den Messwert jedes Ereignisses direkt aus dem Parameter σ abgelesen werden.

Neben der Berücksichtigung von statistischen Effekten konnten durch das Fitting mit der Voigt-Verteilung also zusätzliche Eigenschaften des Z-Bosons, wie dessen Zerfallsbreite und Lebensdauer annäherungsweise ermittelt werden. Als Ergebnisse dieser Analyse (siehe Tabelle 6.6) wurden die aus den Parametern der Fitting-Funktion durch die Angleichung der Voigt-Verteilung an die Messwerte im Fitting-Bereich [87 bis 95 GeV] ermittelten Werte gewählt.

Parameter	entsprechende physikalische Größe	Analyse-Ergebnis	tatsächlicher Wert [3]
μ	invariante Masse M_{inv}	90,77 GeV	91,1876 GeV
Γ	Zerfallsbreite Γ	2,14 GeV	2,4952 GeV
berechnet aus Γ	mittlere Lebensdauer τ	$3,08 \cdot 10^{-25}$ s	$2,64 \cdot 10^{-25}$ s
σ	statistische Schwankung σ	1,31 GeV	-

Tabelle 6.6: Die Endergebnisse der Analyse und die genaueren, durch die *Particle Data Group* berechneten Werte zum Vergleich

Am besten näherungsweise ermittelt wurde die mittlere invariante Masse des Z-Bosons (mit nur 0,5 % Abweichung vom tatsächlichen Wert), was auch ursprüngliches Hauptziel der Analyse war. Außerdem konnte mit guter Genauigkeit die Zerfallsbreite (14 % Abweichung) bestimmt und daraus die mittlere Lebensdauer (17 % Abweichung) des Z-Bosons bestimmt werden. Alle drei Werte stimmen gut mit den tatsächlichen Werten (nach Particle Data Group) überein. Zusätzlich konnte der Wert der statistischen Schwankung bezogen auf genau dieses Experiment beziehungsweise diese Daten bestimmt werden.

Kapitel 7

Higgs-Boson-Analyse

Die Konzepte der Analyse des Z-Teilchens (siehe Kapitel 6: Z-Boson-Analyse) lassen sich auch auf komplexere Analysen, wie der des Higgs-Bosons, übertragen. Die Analyse ist hier schwerer, da die Gesamtanzahl der Ereignisse, die zur Verfügung stehen, sehr viel geringer ist und die Unterscheidung von Messuntergrund und Daten nicht so einfach ist. Bei der Analyse des Higgs-Bosons wurde wieder wie bei der Z-Boson-Analyse die Ruhemasse des Higgs-Bosons ermittelt. Eine Erweiterung der Konzepte auf andere Eigenschaften des Higgs-Teilchens ist ebenso wie beim Z-Boson möglich. Zusätzlich kann man durch den Vergleich der Messdaten mit den Simulationsdaten, mit und ohne Berücksichtigung des Higgs-Teilchens, den Higgs-Mechanismus nachweisen. Genau das ist auch die Methode, mit der das Higgs-Teilchen von Forschern am CERN nachgewiesen wurde.

7.1 Der Zerfall des Higgs-Bosons und die Ereignisanalyse

Das Higgs-Boson ist ebenso wie das Z-Boson aufgrund der kurzen Lebensdauer nicht direkt, sondern nur über seine Zerfallsprodukte nachweisbar. Der hier ausgewertete ist der Zerfall in zwei Z-Bosonen. Dieser Zerfallskanal tritt allerdings nur in 2 bis 3 % der Fälle auf, wenn ein Higgs-Teilchen zerfällt. [12] Das ist ein Grund dafür, dass die Anzahl der Ereignisse mit einem Higgs-Boson in der Analyse so gering ist. Ein weiterer Grund dafür ist, dass das Higgs-Boson sehr selten erzeugt wird. Andere Zerfallskanäle, wie der in ein Bottom-Antibottom-Quarkpaar, die sehr viel häufiger auftreten, können schlechter identifiziert werden und eignen sich nicht zur Auswertung.

Die zwei entstandenen Z-Bosonen zerfallen dann ebenfalls nach kurzer Zeit wieder und sind auch nur über ihre Zerfallsprodukte nachweisbar. Um die Analyse möglichst einfach zu halten, wurden wieder nur die Zerfälle der Z-Bosonen in jeweils zwei Leptonen ausgewertet. Somit werden bei der Higgs-Boson-Analyse also Ereignisse ausgewertet, bei denen am Ende genau vier Leptonen entstanden sind. Der Zerfallskanal, der ausgewertet wurde, ist im Feynman-Diagramm in Abbildung 7.1 dargestellt.

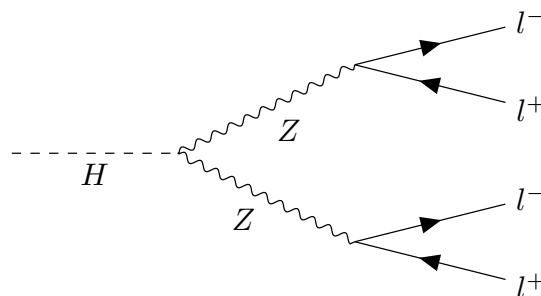


Abbildung 7.1: wichtigster Zerfallskanal des Higgs-Bosons bei der Analyse, die Endprodukte sind vier Leptonen

Diese Kombination aus vier Leptonen kann aber nicht nur durch den Zerfall eines Higgs-Teilchens entstehen. Wenn zum Beispiel ohne Bildung eines Higgs-Bosons zwei Z-Bosonen entstehen, die weiter zerfallen, bleibt am Ende die gleiche Kombination an Teilchen. Eine weitere Möglichkeit ist, dass ein Z-Boson in zwei Leptonen zerfällt, von denen eines ein Photon ausstrahlt. Dieses ausgestrahlte Photon kann dann ein neues Lepton-Antilepton-Paar erzeugen (Paarerzeugung - siehe Kapitel 2.4). So können ebenfalls vier Leptonen entstehen.

Wie auch bei der Z-Boson-Analyse müssen vor der eigentlichen Analyse die gemessenen Ereignisse aussortiert werden. Das gelingt mittels Selektionskriterien und einem Sortierverfahren für die Leptonen der Ereignisse. Sowohl Selektionskriterien als auch Sortierverfahren sind schon in der verwendeten Analysesoftware implementiert. Da diese sinnvoll sind und gut funktionieren, wurden sie nicht verändert, sondern werden im Folgenden nur kurz erklärt. Der erläuterte Algorithmus wird für jedes der gemessenen Ereignisse (bestehend aus gemessenen Teilchen) durchgeführt.

Zuerst werden für das Ereignis alle Teilchen aussortiert, die keine Elektronen oder Myonen (bzw. deren Antiteilchen) sind. Erreicht wird das über den Vergleich der PDGID der Teilchen. Das ist eine durch die Particle Data Group eingeführte Systematisierung der Elementarteilchen, bei der jedes Elementarteilchen (siehe Kapitel 2.2) eine charakterisierende ID besitzt. Jedes im Experiment entstehende Teilchen wird identifiziert und nach diesem Schema eingeordnet. Damit kann überprüft werden, dass nur Ereignisse mit Elektronen und Myonen übrig bleiben. Es sind also nur noch entstandene Leptonen übrig, da Ereignisse untersucht werden sollen, bei denen am Ende Leptonen entstanden sind. Die übrigen Leptonen werden dann in einer Liste absteigend nach ihrem transversalen Impuls sortiert. Der transversale Impuls ist die Komponente des Impulses senkrecht zum Protonenstrahl im Detektor.

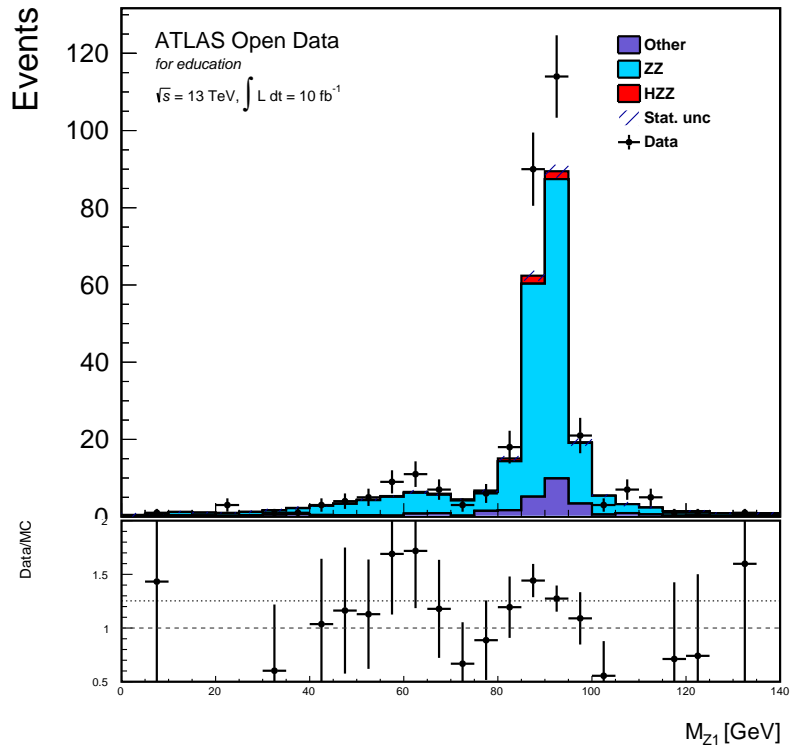
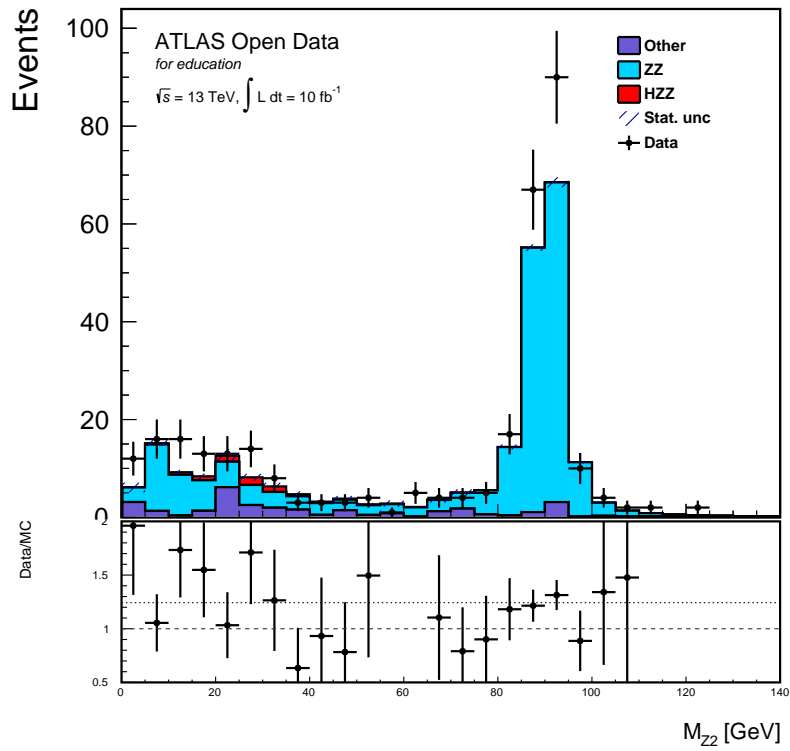
Nach der Sortierung der Leptonen werden auf die Ereignisse jeweils drei Selektionskriterien angewendet. Das erste Kriterium prüft, ob die Anzahl der entstandenen Leptonen genau vier beträgt. Das muss so sein, da nur Zerfälle analysiert werden sollen, in denen genau vier Leptonen entstanden sind. Im zweiten und dritten Selektionskriterium wird überprüft, ob der transversale Impuls vom zweiten und dritten Lepton (die Sortierung erfolgt nach dem transversalen Impuls, siehe oben) bestimmte Werte hat. Es soll gelten: $p_{t(2)} > 15 \text{ GeV}$; $p_{t(3)} > 10 \text{ GeV}$.

Anschließend wird für jedes Ereignis die beste Leptonenpaarung gesucht. Denn jeweils zwei Leptonen sollen nach der Theorie aus dem Zerfall eines Z-Bosons stammen. Um zu rekonstruieren, welche Leptonen zu welchem Z-Boson gehört haben, wird ein Verfahren angewandt, um die beste Paarung von jeweils zwei Leptonen zu ermitteln. Zur Beurteilung wie gut ein Paar passt, dient der Betrag der Differenz zwischen der invarianten Masse des Leptonenpaares und der theoretischen invarianten Masse eines Z-Bosons. Im optimalen Fall beträgt diese Differenz 0, denn die Leptonen sollen ja Zerfallsprodukte des Z-Bosons sein. Nun wird noch die Summe aus den beiden Differenzen (der beiden Leptonenpaare) gebildet und je kleiner diese Summe ist, desto besser passt die Leptonenpaarung. Eine informatische Funktion, die jede mögliche Kombination nacheinander ausprobiert, ermittelt dann für alle möglichen Leptonenpaarungen die beste. Dabei wird außerdem bei jeder Kombination geprüft, ob die Paare vom gleichen Teilchentyp sind und die gleiche Ladung haben (notwendige Eigenschaften). Die am Ende ermittelte beste Kombination wird als die tatsächliche in die Histogramme eingegeben. Falls keine Kombination von Leptonen aus dem Ereignis die Kriterien (Ladung und Teilchentyp) erfüllt, wird das Ereignis verworfen und geht nicht mit in die Analyse ein.

7.2 Analyse der beiden Z-Peaks

In den beiden folgenden Histogrammen (siehe Abbildung 7.2 und 7.3) ist die invariante Masse jeweils eines Leptonenpaar-Systems abgebildet (Ermittlung der Leptonenpaare: siehe Kapitel 7.1).

Wie man erkennen kann, gibt es einen Peak in der invarianten Masse bei beiden Histogrammen um etwa 90 GeV, was der Ruhemasse des Z-Bosons entspricht. Das lässt sich dadurch erklären, dass häufig zwei Z-Bosonen entstehen, die dann in jeweils zwei Leptonen zerfallen. Diese Ereignisse sind bei den Simulationsdaten des Histogramms als hellblauer Anteil dargestellt. Die Position des Peaks bei der Z-Boson-Masse ergibt sich daraus, dass die invariante Masse des Systems aus den zwei Leptonen gerade die invariante Masse des Z-Bosons ist, das in die Leptonen zerfallen ist.

Abbildung 7.2: Häufigkeiten der invarianten Masse M_{inv} des ersten LeptonenpaarsAbbildung 7.3: Häufigkeiten der invarianten Masse M_{inv} des zweiten Leptonenpaars

Nur bei dem roten Anteil der Simulationsdaten im Histogramm zerfällt wirklich ein entstandenes Higgs-Boson in zwei Z-Bosonen, das Ereignis, aus dem die Eigenschaften des Higgs-Teilchens abgeleitet werden sollen, welches aber sehr selten auftritt. Im Histogramm ist also auch im roten Anteil jeweils das Lepton-Antilepton-System aus dem Zerfall eines Z-Bosons entstanden. Diese zwei Z-Bosonen sind im Unterschied zum hellblau dargestellten Anteil aber aus dem Zerfall eines Higgs-Bosons entstanden. Daher müsste der rote Anteil in beiden Histogrammen (beiden Leptonenpaar-Systemen) etwa bei der Masse des Z-Bosons liegen. Beim Histogramm des ersten Z-Bosons ist das auch der Fall, der simulierte Anteil an Higgs-Boson-Zerfällen liegt etwa bei 90 GeV. Beim Histogramm des zweiten Z-Bosons ist der Anteil an Higgs-Zerfällen aber in deutlich niedrigere Energien verschoben, bei etwa 30 bis 35 GeV. Dass die invariante Masse-Verteilung zu einem realen Z-Boson gehört, ist damit ausgeschlossen. Daher muss das zweite Z-Boson, in das das Higgs-Boson zerfällt, ein virtuelles Teilchen sein. Nach der Theorie ist das auch gar nicht anders möglich. Aus einem Higgs-Boson der invarianten Masse 125,25 GeV [3] können nicht zwei reelle Z-Bosonen der invarianten Masse 91,19 GeV [3] entstehen. Invariante Massen sind zwar nicht additiv, das heißt, man kann zur Berechnung der invarianten Masse eines Systems nicht einfach die invarianten Massen der Zerfallsprodukte addieren, aber dennoch lässt sich eine Mindestangabe für die invariante Masse eines Systems berechnen, denn es gilt: $M_{Higgs} \geq M_{Z1} + M_{Z2}$. Daraus folgt für die Masse des zweiten Z-Bosons (eines virtuellen Teilchens): $M_{Z2} = M_{Z*} \leq M_{Higgs} - M_{Z1}$. Bei Einsetzung der Werte ([3]) erhält man also: $M_{Z*} \leq 34,06$ GeV. Diese theoretische Berechnung bestätigt sich auch im Histogramm: Ab einer invarianten Masse von 35 GeV sind keine Ereignisse mehr vorhanden, in denen ein Higgs-Boson in zwei Z-Bosonen zerfallen ist.

7.3 Analyse des Histogramms der invarianten 4-Leptonen-Masse

Im folgendem Histogramm (Abbildung 7.4) ist die invariante Masse des kompletten Systems aus vier Leptonen dargestellt, also die invariante Masse bei allen Ereignissen, bei denen genau vier Leptonen als Endprodukte messbar sind und die nicht durch die Selektionskriterien aussortiert wurden. Wie man im Histogramm erkennen kann, wurde die Binbreite nun deutlich größer eingestellt als bei der Z-Boson-Analyse. Das macht die Analyse zwar deutlich ungenauer, ist aber dennoch zwingend notwendig, da die Menge an Daten aus dem Experiment bei der Higgs-Analyse deutlich geringer als bei der Z-Boson-Analyse ist. Deshalb sind bei dieser Analyse die Ungenauigkeiten auch höher als bei der vorherigen. Trotzdem wurden die Daten ausgewertet und das Fitting nach einigen Modifikationen daran durchgeführt, um Ergebnisse zu erhalten.

Die Simulationsdaten des Histogramms (siehe Abbildung 7.4) lassen sich gut auswerten. Der erste auffällige Peak ist bei etwa 90 GeV. Ereignisse mit dieser invarianten Masse des Systems sind wie in Kapitel 7.2 beschriebene Zerfälle eines Z-Bosons in ein Leptonenpaar unter Emission eines Photons, das ein zweites Leptonenpaar erzeugt. Da alle vier Leptonen aus einem Z-Boson stammen, muss die invariante Masse des 4-Leptonen-Systems etwa der Masse des Z-Bosons entsprechen. Die zweite deutliche Anhäufung von Ereignissen beginnt bei einer invarianten Masse von etwa 180 GeV. Dies sind Ereignisse, bei denen zwei Z-Bosonen (die nicht aus dem Zerfall eines Higgs-Bosons stammen) in jeweils zwei Leptonen zerfallen sind. Da nun ein System aus mehreren Z-Bosonen vorliegt, die nicht aus dem Zerfall eines Teilchens entstanden sein müssen, können diese zusätzlich zu ihrer Ruheenergie noch Bewegungsenergie haben. Diese äußert sich in ihrem nicht gegenseitig zu 0 addierendem Impuls und hat so Einfluss auf die Berechnung der invarianten Masse des Systems, die deutlich über 180 GeV liegen kann. Daher beginnt die invariante Masse dieser Ereignisse auch erst bei 180 GeV, erstreckt sich aber noch weit über 200 GeV, wodurch kein klarer Peak zu beobachten ist, wie bei 90 GeV. Ein etwas kleinerer Peak befindet sich noch bei etwa 120 bis 125 GeV. Dieser besteht zu einem großen Teil aus Ereignissen, bei denen ein Higgs-Boson in zwei Z-Bosonen zerfallen ist (roter Anteil). Da das Higgs-Boson nur in diesem Bereich im Histogramm auftritt, kann man schon aus dem Histogramm ablesen, dass es etwa eine Masse von 120 bis 125 GeV besitzt. Da die beiden entstandenen Z-Bosonen jetzt beide aus dem Higgs-Boson stammen, addieren sich ihre Impulse auch wieder zu 0, und die invariante Masse des Systems entspricht wieder der Masse des Higgs-Teilchens.

Was man ebenfalls im Histogramm erkennen kann, ist, dass die Messdaten in diesem Fall

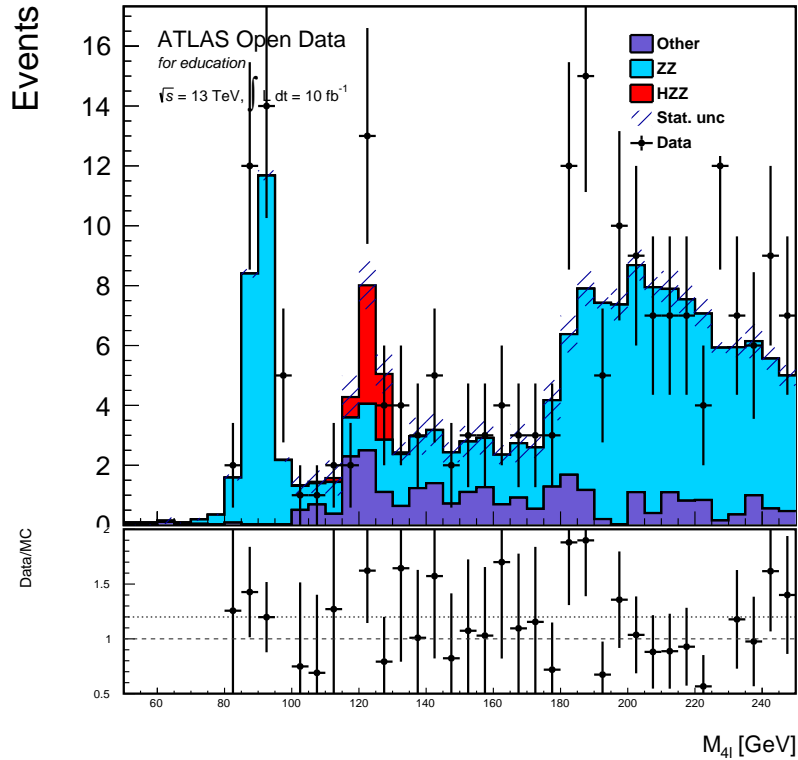


Abbildung 7.4: Häufigkeiten der invarianten Massen des 4-Leptonen-Systems, Darstellungsbereich: 50 bis 250 GeV, Anzahl der Bins: 40, Bin-Breite: 5 GeV

sehr stark von den Simulationsdaten abweichen. Das kann man auch im Streifen darunter sehen. Die Daten stimmen oft auch mit Einbeziehung der Unsicherheit nicht mit den Simulationsdaten überein, das heißt, die statistische Signifikanz ist in manchen Bins über 1 (siehe Kapitel 4.4). Damit bestätigt sich wieder die hohe Unsicherheit der Daten, die eine genaue Auswertung bzw. ein Fitting auf Basis der direkten Messdaten aufgrund der geringen Datenmenge erschwert.

7.3.1 Fitting des Z-Boson-Peaks im Histogramm der invarianten 4-Leptonen-Masse

Der erste Peak, an den eine Funktionsgleichung angeglichen werden kann, ist wieder der Z-Peak bei etwa 90 GeV. Dabei wurde im Bereich [70 bis 110 GeV], also an acht Datenpunkte (Bins) angeglichen. Da bei dieser Auswertung nur so wenige Bins als Datenpunkte für das Fitting verwendet werden können, ist die Normalverteilung als Fitting-Funktion sinnvoll, da sie weniger komplex als die Voigt-Verteilung ist und weniger Parameter hat. Die Ungenauigkeiten sind in diesem Fall so hoch, dass eine Funktion mit vielen Parametern keinen Vorteil bringt. Relevant ist in der Higgs-Boson-Analyse beim Fitting jetzt allerdings, dass die Fitting-Funktion nicht auf die Mittelpunkte der Bins sondern auf Integrale über den Bins angepasst wird (siehe Kapitel 6.5.4), eine Änderung, die bei der Z-Boson-Analyse nur der Vollständigkeit wegen implementiert wurde, die bei der Higgs-Boson-Analyse aber von größerer Wichtigkeit ist. Als Ergebnisse des Fittings erhält man die in Tabelle 7.1 abgebildeten Werte.

Die Anpassung der Fitting-Funktion an die Daten ist relativ hoch (hohe p-Werte) und beim Fitting an die Simulationsdaten mit einem p-Wert von 0,68 sogar sehr hoch. Das liegt vor allem daran, dass die Anzahl der Bins, an die gefittet wird, sehr gering ist und die Funktion somit gut durch diese Werte gelegt werden kann. Bei acht Messwerten und einer Fitting-Funktion mit fünf Parametern ist generell ein hoher p-Wert zu erwarten. Ein weiterer Grund für die gute Anpassung ist die schon erwähnte Änderung bei der Z-Boson-Analyse, nach der die Funktion nicht an die

Daten	n	μ	a	b	σ	χ^2	$\frac{\chi^2}{N_{\text{doF}}}$	p-Wert
Simulation	61,15	90,51	0,16	3,72	3,00	0,78	0,39	0,68
Messdaten	51,94	90,73	0,24	3,78	3,78	3,34	1,67	0,19

Tabelle 7.1: Parameter der Fitting-Funktionen und χ^2 -, $\frac{\chi^2}{N_{\text{doF}}}$ - und p-Wert für die Normalverteilung im Fitting-Bereich [70 bis 110 GeV]

Datenwerte selbst, sondern die Integrale der Funktion an die Bins angeglichen werden. Wie in Kapitel 6.5.4 erklärt, war diese Änderung bei der Z-Boson-Analyse nicht besonders ausschlaggebend, weil die Binbreite klein gewählt wurde. Bei dieser Analyse allerdings ist sie aufgrund der größeren Binbreite zwingend notwendig und trägt zu einem verbesserten Fit bei. Man erhält also auch bei diesem Histogramm realistische Werte für die Masse des Z-Bosons von 90,51 bzw. 90,72 GeV, obwohl die Binbreite 5 GeV beträgt.

7.3.2 Skalierung der Daten

Bei der Analyse des Higgs-Boson-Peaks kann nicht einfach der Minimum-Chiquadrat-Fit angewandt werden, da die Menge an Daten dafür nicht ausreichend ist. Denn eine Voraussetzung für diese Anpassung ist, dass die Werte in den Bins jeweils annähernd normalverteilt um den tatsächlichen Wert sind, was bei dieser geringen Datenmenge nicht gegeben ist. Außerdem ist es durch die teilweise hohen Unsicherheiten schwer, das Higgs-Boson nachzuweisen, was aber Ziel der Analyse ist.

Da diese Arbeit zur Analyse nur die frei zur Verfügung stehenden Daten aus dem ATLAS-Detektor verwendet, dieser Datensatz aber nicht mehr Ereignisse enthält, müssen die vorhandenen Daten hochskaliert bzw. vermehrt (mit einem Faktor multipliziert) werden. Die so selbst erzeugten Daten (Pseudodaten) werden dann wie echte Daten zur Analyse verwendet.

Zuerst muss der Faktor ermittelt werden, mit dem die Daten multipliziert werden sollen. Dazu kann die Signifikanz der Simulationsdaten im Bereich um den Higgs-Peak verwendet werden. Das Ziel ist, dass mittels der Daten das Higgs-Boson nachgewiesen werden kann. Dazu muss der Messwert weit genug über dem angenommenen Untergrund (Theorie ohne Higgs-Boson) liegen, um zu beweisen, dass der Wert nicht zu den Annahmen ohne Higgs-Boson passt. Die Signifikanz muss also groß genug sein. Zur Berechnung der Signifikanzen wurden im Folgenden immer die drei Bins mit Higgs-Boson-Anteil zusammengefasst (115 bis 130 GeV). Die Signifikanz der Simulationsdaten gegenüber dem Untergrund (Simulationsdaten ohne Higgs-Anteil) im Bereich [115 bis 130 GeV] ist damit derzeit:

$$Z = \frac{|y - y_{\text{exp}}|}{\sqrt{y_{\text{exp}}}} \approx \frac{(4 + 8 + 5) - (3,5 + 4 + 3)}{\sqrt{3,5 + 4 + 3}} \approx 2,0 \quad (7.1)$$

Nach Multiplikation der Werte mit dem Faktor a ist die Signifikanz:

$$Z' = \frac{a \cdot |y - y_{\text{exp}}|}{\sqrt{a \cdot y_{\text{exp}}}} = \sqrt{a} \cdot \frac{|y - y_{\text{exp}}|}{\sqrt{y_{\text{exp}}}} = \sqrt{a} \cdot Z \quad (7.2)$$

Von einer ausreichenden Signifikanz für einen Nachweis des Higgs-Bosons spricht man ab dem Wert 5. [13] Mit der Annahme $Z' = 5$ ergibt sich: $a = 6,25$.

Die Simulationsdaten können nun einfach mit diesem Faktor skaliert werden, für die Messdaten muss allerdings eine andere Methode gefunden werden, da das Ziel nicht ist, die Daten einfach zu vermehren, sondern zufällige neue Daten zu erzeugen. Das erreicht man dadurch, dass auch die Messwerte zuerst mit dem festgelegten Faktor multipliziert werden. Dann wird durch die um diesen Wert verteilte Poisson-Verteilung eine Zufallszahl erzeugt, die als Datenwert in das Histogramm eingeht. Es ergibt sich also ein neues Histogramm mit einer höheren Gesamtanzahl an Ereignissen (siehe Abbildung 7.5). Dies entspricht dann einer Messung bei höherer Luminosität von $62,5 \text{ fb}^{-1}$. Die Luminosität ist ein Maß dafür, wie viele Teilchenkollisionen in einem Detektor pro Fläche und Zeit stattfinden.

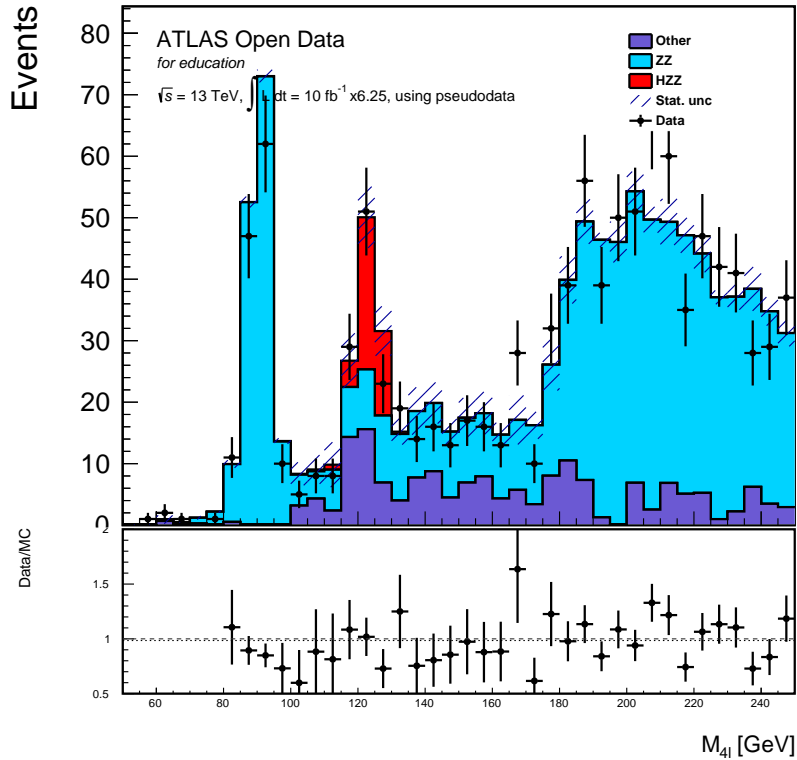


Abbildung 7.5: Histogramm über die invariante Masse des Systems aus vier Leptonen nach der Skalierung mit dem Faktor 6,25, bei den Messdaten zusätzlich Berücksichtigung einer statistischen Schwankung

An diesen Daten kann nun wieder der Minimum-Chiquadrat-Fit durchgeführt werden.

7.3.3 Fitting des Higgs-Boson-Peaks

Bei dem Fitting des Higgs-Boson-Peaks zur Ermittlung der Masse des Higgs-Bosons muss der Messuntergrund zwingend einbezogen werden, denn die meisten Ereignisse, bei denen vier Leptonen entstehen, die also im Histogramm abgebildet sind, sind keine Ereignisse bei denen ein Higgs-Boson vorkommt. Der Messuntergrund ist bei diesem Experiment also größer als der Signalanteil, woraus folgt, dass die Fitting-Funktion nicht einfach an alle Daten angeglichen werden kann. Vielmehr sollte nur an das Signal (die Higgs-Boson-Ereignisse) gefittet werden, also an den roten Anteil im Histogramm. Dieser kann berechnet werden durch: $\text{Signal} = \text{Daten} - \text{Messuntergrund}$. Der Messuntergrund wird dabei aus den Simulationsdaten bestimmt und ist für jeden Bin definiert als die Menge aller Simulationsdatenereignisse, die nicht Higgs-Boson-Ereignisse sind, im Histogramm alle Ereignisse bis auf die roten.

Daher müssen die Fitting-Funktionen neu definiert werden, sodass nicht mehr die Integrale der Fitting-Funktionen an die Bininhalt angeglichen werden, sondern die Integrale der Fitting-Funktionen addiert mit dem jeweiligen Untergrund in den Bins (berechnet aus den Simulationsdatenwerten). Eine weitere Anpassung in den Fitting-Funktionen ist, dass nun die Untergrundparameter a und b , die einen linearen Untergrund in der Fitting-Funktion erzeugen, gestrichen werden, da ein zusätzlicher Untergrund bei der Higgs-Boson-Analyse schon vollständig in den Fitting-Funktionen implementiert ist und somit nicht mehr benötigt und erwünscht ist.

Problematisch beim Fitting ist die starke Abhängigkeit der Ergebnisse von den Startwerten, vor allem dem Startwert für den Mittelwert μ . Optimalerweise sollen so wenige Informationen wie möglich vorgegeben werden und das Programm soll das Fitting dennoch durchführen können. Daher wurde hier eine Methode implementiert, die den besten (bzw. wahrscheinlichsten) Startwert

für den Mittelwert μ ermittelt. Der Hintergrund ist, dass der Mittelwert der Funktion vermutlich an der Stelle liegt, an der die Werte am stärksten vom Untergrund abweichen. Aus statistischer Sicht ist das an der Stelle, an der die statistische Signifikanz am höchsten ist. Der beste Startwert für den Mittelwert wird demzufolge definiert als die Stelle, an der $Z = \frac{\text{Daten-Messuntergrund}}{\text{Unsicherheit}}$ den größten Wert annimmt. Mithilfe dieser Methode wird der richtige Startwert für den Mittelwert bestimmt und direkt an die Fitting-Funktion übergeben. Außerdem kann so der Fitting-Bereich beliebig vergrößert werden, daher wurde er auf den kompletten Darstellungsbereich des Histogramms festgelegt: [50 bis 250 GeV]. Als Fitting-Funktion wurde wieder die Normalverteilung gewählt, da diese das Experiment bei wenig Datenpunkten und Ereignissen am besten beschreibt. Damit ergeben sich für die Fitting-Funktionen die in Tabelle 7.2 aufgeführten Werte. Die dazugehörige Fitting-Funktion ist darunter in Abbildung 7.6 dargestellt.

Daten	n	μ	σ	χ^2	$\frac{\chi^2}{N_{dof}}$	p-Wert
Simulation	29,89	123,7	2,90	0,080	0,002	1,0
Messdaten	30,12	122,3	2,48	44,03	1,22	0,17

Tabelle 7.2: Parameter der Fitting-Funktionen und χ^2 -, $\frac{\chi^2}{N_{dof}}$ - und p-Wert für die Voigt-Verteilung im Fitting-Bereich [50 bis 250 GeV]

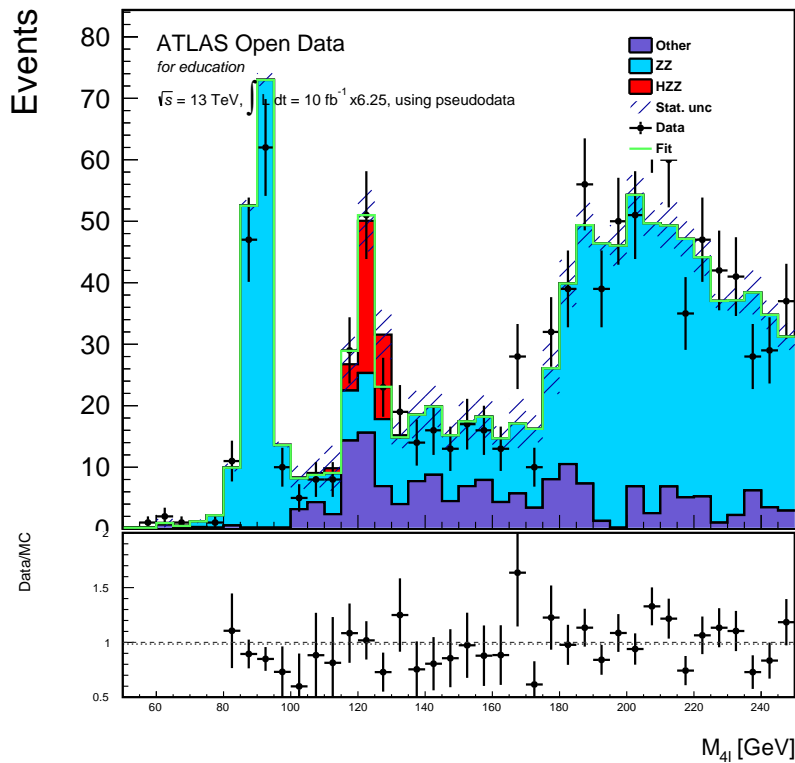


Abbildung 7.6: Histogramm über die invariante Masse des Systems aus vier Leptonen nach der Skalierung mit dem Faktor 6,25, Fitting an die Messdaten, der Fit ist hier nicht als Funktion sondern als Bins dargestellt

Auffällig an den Werten in dieser Tabelle sind die sehr kleinen χ^2 -Werte, wodurch sich sehr hohe p-Werte ergeben. Vor allem beim Fitting an die Simulationsdaten sind die Werte des χ^2 -Tests unrealistisch gut. Das liegt daran, dass die Funktion bei diesem Fitting an das Signal in allen 40 Bins angeglichen wird. Da das Signal aber in nahezu allen Bins 0 ist, wird die Fitting-Funktion

effektiv nur an die drei relevanten Bins mit dem Higgs-Boson-Anteil angeglichen, in allen anderen Bins sind Signal und Wert der Fitting-Funktion fast 0. Bei der Berechnung des χ^2 -Wertes werden aber die Werte von Signal und Funktion in allen 40 Bins verglichen und da die Übereinstimmung dieser Werte in fast allen Bins 100 % ist, ergibt sich ein sehr niedriger χ^2 -Wert. Für das Fitting an die Messdaten hat dieser Effekt einen nicht ganz so großen Einfluss auf die Werte des χ^2 -Tests, da der Untergrund im Experiment durch statistische Schwankungen vom simulierten Untergrund abweicht und sich somit nicht in vielen Bins eine sehr hohe Übereinstimmung ergibt. Dadurch erhält man einen höheren χ^2 -Wert, der allerdings dennoch nur bedingt auswertbar ist bzw. vorerst auf seine Aussagekraft geprüft werden muss.

Eine Möglichkeit, bei der man die Werte des χ^2 -Tests dennoch verwenden kann ist, den χ^2 -Wert für die Übereinstimmung von Messdaten und Simulationsdatenuntergrund als Vergleichswert zu berechnen. Aus genannten Gründen ist die gesamte Übereinstimmung von Fitting-Funktion und Messdaten sehr hoch, da die Übereinstimmung vom Untergrund der Simulationsdaten mit den Messdaten schon hoch ist. Wenn man letztere Übereinstimmung berechnet, hat man einen Vergleichswert für den χ^2 -Wert ohne addierte Fitting-Funktion und kann so beurteilen, um wie viel die Fitting-Funktion die Übereinstimmung mit den Messdaten verbessert. Es ergeben sich folgende Werte für den χ^2 -Test für die Übereinstimmung vom Untergrund der Simulationsdaten (alle Simulationsdaten ohne Higgs-Anteil) und Messdaten:

$$\chi^2 = 73,5; \quad \frac{\chi^2}{N_{\text{dof}}} = 1,93; \quad p\text{-Wert} = 0,0005$$

Wie man an diesen Werten erkennen kann, ist der χ^2 -Wert für die Übereinstimmung von Messdaten und Simulationsdatenuntergrund ohne addierte Fitting-Funktion mit etwa 70 schon gering, wodurch der sehr niedrige χ^2 -Wert beim Fitting an die Messdaten teilweise relativiert werden muss. Allerdings spricht die starke Verringerung des χ^2 -Wertes dennoch für einen sehr guten Fit, dessen Ergebnisse also auch aussagekräftig sind. Im Histogramm (siehe Abbildung 6.6) ist außerdem deutlich erkennbar, dass vor allem in den Bins mit Higgs-Boson-Anteil die Fitting-Funktion sehr gut an die Werte angeglichen ist und die noch verbliebene Unsicherheit demzufolge aus der fehlenden Übereinstimmung von Simulationsdatenuntergrund und Messdaten resultiert, die in einer normalen statistischen Schwankung begründet ist. Eine Anpassung mit einem p-Wert von 17 % ist somit gut.

7.3.4 Nachweis des Higgs-Bosons

Das Higgs-Boson wird darüber nachgewiesen, dass die im Experiment gemessenen Daten verglichen werden mit einer Theorie mit und ohne Higgs-Boson. Statistisch bildet man dazu die Signifikanz von Messwerten und Simulationsdaten ohne Higgs-Boson. Je nachdem, wie hoch diese Signifikanz in den Bins ist, kann man eine Aussage darüber treffen, wie wahrscheinlich die Messdaten zu den Simulationsdaten passen, oder umgekehrt wie wahrscheinlich es ist, dass die Daten (Simulations- und Messdaten) nur unter Einbeziehung des Higgs-Bosons vereinbar sind. Je höher die Signifikanz in einem Bin ist, desto unwahrscheinlicher ist es, dass die Daten zueinander passen und desto wahrscheinlicher ist es, dass das Higgs-Boson bei dieser invarianten Masse auftritt.

Da nur in drei Bins Ereignisse mit Higgs-Bosonen messbar waren, wurden wieder alle drei relevanten Bins zu einem zusammengefasst und von diesem die statistische Signifikanz berechnet:

$$Z_{[120 \text{ bis } 125 \text{ GeV}], \text{ skaliert}} = \frac{|y - y_{\text{exp}}|}{\sqrt{y_{\text{exp}}}} \approx \frac{(29 + 51 + 23) - (23 + 25 + 18)}{\sqrt{23 + 25 + 18}} \approx 4,71 \quad (7.3)$$

Diese hohe Signifikanz konnte nun aber nur durch die Skalierung der Daten um den Faktor 6,25 erreicht werden. Des Weiteren wurde deshalb noch die tatsächliche statistische Signifikanz von Messdaten und Simulation ohne Higgs-Boson vor der Skalierung berechnet:

$$Z_{[120 \text{ bis } 125 \text{ GeV}], \text{ unskaliert}} = \frac{|y - y_{\text{exp}}|}{\sqrt{y_{\text{exp}}}} \approx \frac{(2 + 4 + 13) - (3,5 + 4 + 3)}{\sqrt{3,5 + 4 + 3}} \approx 2,62 \quad (7.4)$$

Wie zu erwarten war, ist diese Signifikanz deutlich geringer als mit Skalierung der Daten, weshalb die Daten ursprünglich auch skaliert wurden (um eine sinnvolle und aussagekräftige Analyse zu ermöglichen).

7.4 Endergebnisse und Auswertung der kompletten Analyse

Die Konzepte der Z-Boson-Analyse konnten erfolgreich auf die Higgs-Boson-Analyse übertragen werden und mit einigen Erweiterungen führte so auch die Higgs-Boson-Analyse zu aussagekräftigen Ergebnissen. So wurde zuerst der Zerfall des Higgs-Bosons analysiert und das Fitting erneut mit gutem Erfolg an einem Z-Boson-Peak durchgeführt, was wieder trotz einer sehr viel geringeren Datenmenge zur genauen Bestimmung der Masse des Z-Bosons geführt hat. Anschließend wurden die Daten hochskaliert, um sie zum Fitting des Higgs-Boson-Peaks zu verwenden. Durch die Skalierung konnte für drei Bins sogar das Higgs-Boson mit sehr hoher Wahrscheinlichkeit nachgewiesen werden. Denn nach der Normalverteilung ergibt sich damit eine Wahrscheinlichkeit für die Notwendigkeit, dass in diesen Bins ein Higgs-Boson vorkommt, von $P(H) = 99,99975\%$. Als Ergebnisse der Analyse des Higgs-Boson-Zerfalls können also folgende Werte (in Tabelle 7.3) aufgeführt werden:

physikalische Größe	Analyse-Ergebnis	tatsächlicher Wert [3]
invariante Masse M_{inv} des Z-Bosons	90,73 GeV	91,1876 GeV
invariante Masse M_{inv} des Higgs-Bosons	122,3 GeV	125,25 GeV

Tabelle 7.3: Die Endergebnisse der vorliegenden Analyse und die genaueren, durch die *Particle Data Group* berechneten Werte zum Vergleich

Auch die mittlere invariante Masse des Higgs-Bosons konnte somit relativ genau bestimmt werden (nur 2,4 % Abweichung vom tatsächlichen Wert). Des Weiteren wurden die folgenden (in Tabelle 7.4) dargestellten statistischen Signifikanzen berechnet werden.

statistische Signifikanz von...	Wert
Simulationsdaten mit und ohne Higgs-Anteil	2,0
Simulationsdaten ohne Higgs-Anteil und Messdaten (vor der Skalierung)	2,62
Simulationsdaten ohne Higgs-Anteil und Messdaten (nach der Skalierung)	4,71

Tabelle 7.4: Die Ergebnisse der Signifikanzberechnungen

Kapitel 8

Analyse mittels Neuronaler Netze

Ziel der Analyse der Daten aus dem ATLAS-Experiment war in den vorherigen Kapiteln vor allem die Ermittlung von Eigenschaften der Teilchen Z- und Higgs-Boson. Dabei wurden mathematische Auswertungsmethoden wie das Fitting von Wahrscheinlichkeitsdichtefunktionen verwendet. In der weiteren Analyse soll es hingegen um den Nachweis des Higgs-Bosons anhand der Daten gehen. Das Grundprinzip dabei besteht immer darin, die erwarteten Ergebnisse vor der Auswertung der im Experiment gemessenen Daten nach dem Standardmodell zu berechnen. Das kann man für zwei Modelle durchführen, mit und ohne Higgs-Boson. Nach der eigentlichen Messung im Experiment vergleicht man dann die Messdaten mit den nach der Theorie berechneten Daten mit und ohne Higgs-Boson. Und wenn die Messergebnisse signifikant nicht zu den Simulationsdaten ohne Higgs-Boson passen, kann die Existenz des Higgs-Bosons damit nachgewiesen werden. Im Grunde wurde dies auch schon am Ende der vorhergehenden Analyse durchgeführt. Das Problem dabei war allerdings, dass die vorhandene Datenmenge im frei zur Verfügung stehenden Datensatz nicht ausreicht, um eine hohe statistische Signifikanz zu erreichen. Daher musste die Datenmenge hochskaliert werden und dabei wurden sogenannte „Pseudodaten“ erzeugt. Das Ziel der Analyse mittels Neuronaler Netze ist nun, das Higgs-Boson auch mit den vorhandenen Daten, also ohne Skalierung, nachzuweisen. Dazu wird ein Neuronales Netz darauf trainiert, die Ereignisse zu klassifizieren. Das heißt, es soll für jedes im Experiment gemessene Ereignis sagen können, ob dabei ein Higgs-Boson zerfallen ist, oder nicht. Damit erhofft man sich keine Erhöhung der vorhandenen Datenmenge, sondern eine Selektion der Daten. Denn wenn die Untergrund-Ereignisse herausgefiltert werden, kann damit auch eine höhere Signifikanz erreicht werden und das Higgs-Boson somit sicherer nachgewiesen werden.

Ein Hauptbestandteil der Arbeit war dabei auch die informatische Implementation der beschriebenen Neuronaler Netze und daran vorgenommenen Anpassungen. Um die Arbeit nicht weiter zu verkomplizieren, wurde aber mit Ausnahme weniger Beispiele weitestgehend auf das Einfügen des während der Arbeit programmierten Codes verzichtet.

8.1 Einrichtung des Neuronalen Netzes

Wie auch bei der vorherigen Analyse des Z- und Higgs-Bosons wird die Erstellung, Konfiguration und das Training des Neuronalen Netzes zuerst an den Simulationsdaten durchgeführt. Dazu wurden bei dieser Analyse viele verschiedene Prozesse genutzt: Einige in denen tatsächlich ein Higgs-Boson zerfällt (Signal-Ereignisse) und einige, in denen nur ähnliche Zerfallsprodukte auftreten, tatsächlich aber kein Higgs-Boson beteiligt war (Untergrund-Ereignisse). Zu jedem dieser Prozesse liegt ein eigener Datensatz mit Ereignissen vor und alle diese Daten wurden in einem Datensatz zusammengefasst und gemeinsam verarbeitet. Dabei wird für jedes Ereignis jedoch weiterhin gespeichert, ob es ein Signal- oder Untergrund-Ereignis ist. Erst im letzten Schritt wird dann für die eigentliche Analyse und Auswertung der Datensatz mit Messdaten genutzt. Das Ziel der Anwendung des Neuronalen Netzes ist es dabei, dass es bei dem Training an den Simulationsdaten darauf trainiert wird, Signal- von Untergrund-Ereignissen zu unterscheiden. Durch die Anwendung des Neuronalen Netzes auf die Messdaten sollen damit dann die hierbei unbekanntes Signal-Ereignisse identifiziert werden.

Auch müssen vor der Anwendung des Neuronalen Netzes auf die Daten diese wieder vorselektiert werden. Dazu wurden im Programmcode wieder Selektionskriterien implementiert, die bestimmte physikalische Kriterien überprüfen (einige davon wurden schon in Kapitel 7.1 erklärt): Erstens sollen alle entstandenen Teilchen Elektronen oder Myonen oder deren Antiteilchen sein, was mit der PDGID überprüft werden kann. Auch sollen insgesamt genau vier Teilchen entstanden sein, die sich zu Paaren mit gleicher Teilchenart (also gleicher PDGID) anordnen lassen. Das lässt sich mit der Summe der PDGID der vier Teilchen überprüfen. Denn Elektronen haben die PDGID 11 und Myonen 13. Wenn die vier Teilchen sich nun zu Paaren aus Elektronen und Myonen zusammensetzen lassen (wobei jedes Paar aus einem Teilchen und seinem zugehörigen Antiteilchen besteht), gibt es die Kombinationsmöglichkeiten $[e^-e^+\mu^-\mu^+]$, $[\mu^-\mu^+\mu^-\mu^+]$ und $[e^-e^+e^-e^+]$. In jedem Fall ist die Summe der PDGIDs 48 oder 52, was im Selektionskriterium überprüft wird. Des Weiteren soll die Summe aller Ladungen der entstandenen Teilchen 0 sein, da auch die Ladung des Ausgangsteilchens (Higgs-Boson) 0 ist. In den Abbildungen 8.1 und 8.2 ist das Histogramm für die invariante Masse der vier Leptonen vor und nach der Selektion dargestellt. Dabei sind die blauen, violetten und grünen Anteile der Balken die Untergrund-Anteile der Simulationsdaten und die gelben, orangenen, braunen und roten Balkenanteile die Signal-Anteile der Simulationsdaten. Die Messdaten (und ihre Fehlerbalken) sind wieder als schwarze Datenpunkte eingezeichnet.

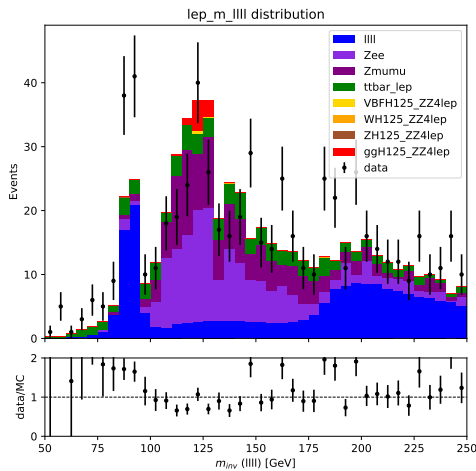


Abbildung 8.1: Histogramm der invarianten Masse des vier-Leptonen-Systems vor der Selektion

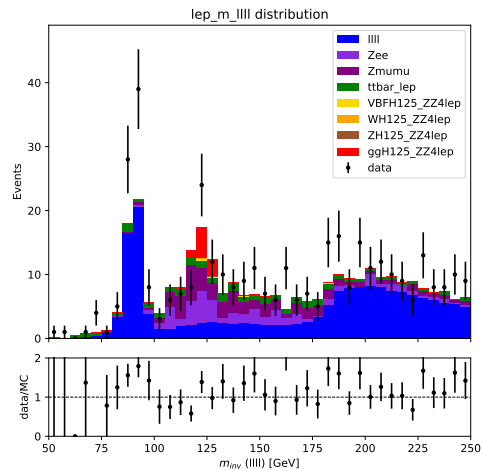


Abbildung 8.2: Histogramm der invarianten Masse des vier-Leptonen-Systems nach der Selektion

Wie man in den Histogrammen erkennen kann, wurden durch die Selektion viele Ereignisse heraus selektiert, in denen ganz offensichtlich kein Higgs-Boson entstanden ist. Das betrifft vor allem Untergrund-Ereignisse, vom roten Signal-Anteil der Simulationsdaten wurden keine Ereignisse selektiert. In den Histogrammen kann man auch sehen, dass auch die Anzahl der Ereignisse in den blauen Balken gleich geblieben ist. Das liegt daran, dass bei diesen Untergrund-Ereignissen auch vier Leptonen entstehen und auf dem ersten Blick keine Unterscheidung von den Higgs-Boson-Zerfällen in vier Leptonen sichtbar ist (siehe Kapitel 7.1).

Nach der Selektion der Daten werden diese aufgeteilt in ein Trainings-, ein Validierungs- und ein Test-Datensatz. Der Trainings-Datensatz wird zum Training des Neuronalen Netzes verwendet. Der Validations-Datensatz wird dazu verwendet zu überprüfen, wie gut das Training funktioniert und um Parameter des Neuronalen Netzes (wie die Anzahl der Schichten) zu optimieren, denn der Validations-Datensatz hat ähnliche, aber nicht die gleichen Daten. Wenn ein Neuronales Netz für den Trainings-Datensatz gut funktioniert, sollte es auch für den Validations-Datensatz gut funktionieren, da dieser aus dem gleichen übergeordneten Datensatz der Simulationsdaten stammt. Anderenfalls ist das ein Zeichen für Overtraining, das heißt, dass das Neuronale Netz zu gut auf die Daten angepasst ist, für andere Daten aber überhaupt nicht gut funktioniert. Zuletzt, nach der kompletten Optimierung des Neuronalen Netzes wird der Test-Datensatz verwendet um wieder zu

überprüfen, ob das Neuronale Netz auch hier gut funktioniert. Häufig ist der Trainings-Datensatz am größten. Auch hier wurde eine Aufteilung des kompletten Simulationsdatensatzes von 80 % für den Trainings-Datensatz und 10 % für je Validations- und Test-Datensatz gewählt.

Vor dem Training des Neuronalen Netzes müssen die Input-Variablen festgelegt werden. Über jedes Ereignis des Datensatzes liegen verschiedene Informationen vor (zum Beispiel die Ladungen der einzelnen Teilchen), was man sich schon bei der Selektion der Daten zunutze macht. Die Input-Variablen sind alle Informationen über die Ereignisse, die zum Training des Neuronalen Netzes verwendet werden. Dabei unterscheidet man zwischen low-level-Variablen (die aus dem Experiment direkt gemessen werden können), wie beispielsweise der Ladung und high-level-Variablen (die aus einzelnen low-level-Variablen berechnet werden müssen), wie etwa die invariante Masse der entstandenen Leptonen. Zum Training verwendet wurden hierbei die low-level-Variablen Anzahl der entstandenen Jets (Spuren in Teilchendetektoren), die Ladungen, PDGIDs, Energien und transversale Impulse der entstandenen Teilchen und die Winkel mit denen sich die Teilchen aus dem Kollisionspunkt entfernen. Die einzige high-level-Variable die verwendet wurde ist die invariante Masse von jeweils zwei Leptonen (also alle sechs Kombinationen aus zwei Leptonen).

8.2 Normalisierung

Die Input-Variablen auf die das neuronale Netz trainiert werden soll müssen vor dem Training normalisiert werden. Das kommt daher, dass die Variablen Werte in sehr unterschiedlichen Größen haben. Während die Ladung der Teilchen beispielsweise nur zwischen 1 und -1 liegt, können invariante Massen sehr viel größere Werte annehmen. Wenn alle Werte unskaliert in das Training eingehen würden, hätten Variablen mit größeren Werten (wie beispielsweise die invariante Masse) einen größeren Einfluss auf das Training und dessen Ausgang als andere Variablen (wie die Ladung). Dies ist aber nicht gewollt, weshalb die Variablen normalisiert werden. Dabei wird für jede Variable der Wert jedes einzelnen Ereignisses so skaliert, dass für die Häufigkeitsverteilung der Variable gilt, dass in den Trainings-Daten ihr Mittelwert 0 und ihre Standardabweichung 1 ist. Dann liegen die Werte aller Variablen in der gleichen Größenordnung und haben etwa den gleichen Einfluss. Dennoch werden die Werte nicht verfälscht, da die Verteilung der Werte nicht verändert, sondern nur skaliert wurde. Diese beschriebene Normalisierung wird normalerweise durch eine „Normalisierungsschicht“ direkt im Neuronalen Netz implementiert. In diesem Fall hat dies aber zu Problemen geführt, weshalb die Normalisierung anders gelöst werden musste, was im folgenden Auszug des Programmcodes dargestellt ist. Eine (informatische) Skalierungsfunktion wird dazu auf den Trainings-Datensatz angepasst (Zeile 3) und mit ihr werden alle drei Datensätze (Trainings-, Validierungs- und Test-Datensatz) dann auf die gewünschte Art normalisiert (Zeilen 4 bis 6) bevor sie dem Netzwerk übergeben werden.

```

1 | from sklearn.preprocessing import StandardScaler
2 | scaler = StandardScaler()
3 | scaler.fit(train_values)
4 | train_values = scaler.transform(train_values)
5 | val_values = scaler.transform(val_values)
6 | test_values = scaler.transform(test_values)

```

8.3 Fitting drei einfacher Neuronaler Netze

Nach der Vorbereitung der Daten durch Selektion, Aufteilung in verschiedene Datensätze und Normalisierung der Input-Variablen können Neuronale Netze definiert und trainiert werden. Dabei wurde das in Python importierbare Framework TensorFlow verwendet, das im Gebiet des maschinellen Lernens oft Anwendung findet. Zur Erstellung eines Neuronalen Netzes müssen verschiedene Schichten mit einer bestimmten Anzahl von Neuronen und einer bestimmten Aktivierungsfunktion definiert werden. Als Ausgabeschicht wurde immer eine Schicht mit einem Neuron und der Sigmoid-Aktivierungsfunktion definiert. Das lässt sich damit begründen, dass es sich hier um ein

klassifizierendes Neuronales Netz handelt. Als Ausgabe wird also nur eine Zahl benötigt, welche das Ereignis als Signal oder als Untergrund klassifiziert. Und die Sigmoid-Funktion liefert als Ergebnis für jeden Eingabewert ein Ergebnis zwischen 0 und 1, was als Klassifikation verwendet werden kann. Um das Neuronale Netz sonst so einfach wie möglich zu halten, wurde für alle weiteren Schichten vor der festen Ausgabeschicht die gleiche Anzahl von Neuronen und die gleiche Aktivierungsfunktion (ReLU) gewählt. Festzulegen sind bei der Definition eines Neuronalen Netzes in dieser Arbeit demzufolge die Anzahl der Schichten vor der Ausgabeschicht und die jeweilige Anzahl der Neuronen. Ein spezielles Neuronales Netz mit bestimmten Hyperparametern wird auch „Modell“ genannt, ein Begriff der im Folgenden für die einzelnen Neuronalen Netze verwendet wird.

Nach der Festlegung dieser sogenannten „Hyperparameter“, wird das Neuronale Netz mit dem Trainings-Datensatz trainiert. Die Trainingsimplementation, die diesen Prozess durchführt, erhält als Eingabewerte neben dem Trainings-Datensatz auch die tatsächliche Klassifikation der Ereignisse, also die Information über jedes der Ereignisse des Datensatzes, ob es ein Signal- oder Untergrund-Ereignis ist.

Zuerst wurden drei einfache Neuronale Netze mit einer verschiedenen Anzahl von Schichten und Neuronen je Schicht in Zehn Epochen trainiert. Berechnet wurden dabei für jedes Modell der Loss (siehe Kapitel 5.2) bei Trainings- und Validierungs-Daten sowie die Fläche unter der ROC-Kurve, der AUC-Wert, (siehe Kapitel 8.4). Die Ergebnisse sind in Tabelle 8.1 dargestellt.

Modell	Schichten	je Neuronen	Trainings-Loss	Validierungs-Loss	ROC-AUC
1	2	3	0,113	0,112	0,987
2	5	20	0,099	0,100	0,991
3	10	100	0,088	0,095	0,992

Tabelle 8.1: Trainings-Ergebnisse drei Neuronaler Netze mit verschiedenen Hyperparametern

Trainings- und Validierungs-Loss sowie die Fläche unter der ROC-Kurve (AUC-Wert) sind einfache Werte, an denen man schnell ungefähr ablesen kann, wie gut das Modell an die Daten angepasst ist (je kleiner die Loss- und je näher der AUC-Wert an 1 liegt, desto besser ist das Modell). Wie man in Tabelle 8.1 sehen kann, wird der Loss von Modell zu Modell kleiner. Ebenso wird der AUC-Wert größer. Für die weitere Analyse wurde vorerst das mittlere Modell 2 verwendet. Eine genaue Analyse, welche Hyperparameter (Anzahl der Schichten und Neuronen) am besten funktionieren, folgt später.

8.4 Auswertungsmethoden und Histogramme

Für das Modell 2 werden im Folgenden zwei komplexere Auswertungsmethoden vorgestellt. Die erste Möglichkeit besteht darin, die ROC-Kurve zu zeichnen. Dies ist ein Diagramm, das darstellt, wie gut ein Neuronales Netz die Trainings-Daten klassifiziert. Nach dem Training wird das Neuronale Netz auf alle Trainings-Daten erneut angewandt und jedes Ereignis wird von ihm als Signal (1) oder Untergrund (0) klassifiziert. Da die Trainings-Daten simuliert sind, weiß man für jedes Ereignis, ob es Signal oder Untergrund ist und kann so über einen Vergleich mit der Klassifikation des Neuronalen Netzes bewerten, wie korrekt die Daten klassifiziert werden. Das Neuronale Netz gibt aber in der letzten Schicht keine eindeutige Zahl (0 oder 1), sondern eine Dezimalzahl zwischen 0 und 1 aus. Je näher die Ausgabe an 1 liegt, desto sicherer ordnet das Netz das Ereignis als Signal ein. Nun muss man festlegen, ab welchem Ausgabewert ein Ereignis als Signal gilt. Der naheliegendste Wert (0,5) ist nicht immer sinnvoll. Je nach Festlegung dieser Grenze treten nämlich unterschiedliche Phänomene auf. Wenn man die Grenze sehr hoch wählt, werden fast alle Ereignisse als Untergrund eingeordnet, dafür kann man mit großer Sicherheit davon ausgehen, dass als Signal eingeordnete Ereignisse auch wirklich Signal-Ereignisse sind. Wenn man die Grenze sehr niedrig setzt, tritt das entgegengesetzte Phänomen auf. Dementsprechend werden die Falsch-Positiv-Rate und die Richtig-Positiv-Rate definiert. Die Falsch-Positiv-Rate gibt an, welcher Anteil der als Signal identifizierter Ereignisse tatsächlich keine Signal-Ereignisse sind, während die Richtig-Positiv-Rate angibt, wel-

cher Anteil der als Signal identifizierter Ereignisse auch wirklich Signal-Ereignisse sind. Im Idealfall ist die Falsch-Positiv-Rate nahe 0 während die Richtig-Positiv-Rate nahe 1 liegt.

Die ROC-Kurve stellt nun dar, für welche Falsch-Positiv-Raten mit dem Neuronalen Netz welche Richtig-Positiv-Raten erreicht werden können oder umgekehrt. Dazu wird das Neuronale Netz auf viele verschiedene Grenzen angewandt und daraus eine Funktion erstellt. Im Programmcode wurde dazu eine Funktion integriert, die automatisch die Werte für eine ROC-Kurve berechnet und diese anschließend zeichnet. Für das Modell 2 erhält man in diesem Fall in Abbildung 8.3 dargestellte ROC-Kurve. Im Prinzip wurde diese Methode auch schon verwendet, nämlich zur Berechnung des AUC-Werts (siehe Tabelle 8.1). Dieser Wert gibt die Fläche unter der ROC-Kurve an, ist also mathematisch das bestimmte Integral unter der ROC-Kurve von 0 bis 1. Diese Fläche ist optimalerweise möglichst nah bei 1, da die Kurve dann eine „eckige“ Form hat, was bedeutet, dass schon bei sehr niedrigen Falsch-Positiv-Raten sehr hohe Richtig-Positiv-Raten erreicht werden.

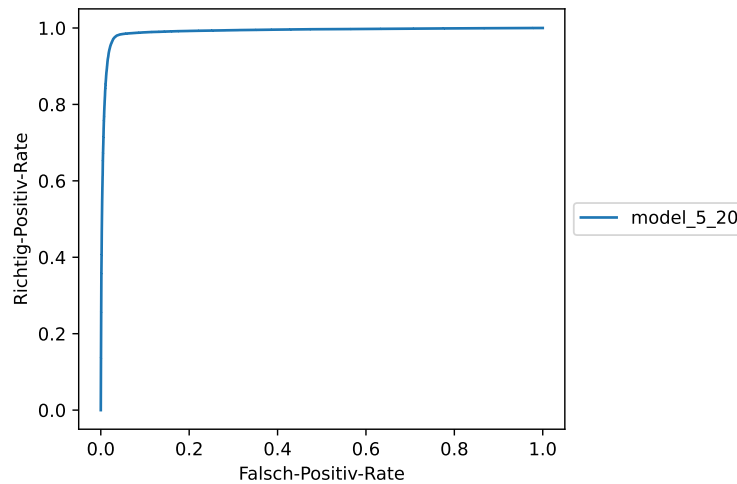


Abbildung 8.3: ROC-Kurve Modell 2

Eine andere Methode zur Beurteilung des Modells und dessen Funktionsfähigkeit ist die Erstellung eines weiteren Histogramms, im Folgenden DNN-Klassifikations-Histogramm genannt (DNN steht für „Deep Neural Network“). Für die Erstellung dieses Histogramms wird das trainierte Neuronale Netz auf alle Daten des Simulations-Datensatzes angewandt, wobei für jedes Ereignis als Ausgabe eine Zahl zwischen 0 und 1 generiert wird. Dieser Wert wird als DNN-Vorhersage bezeichnet. Je näher er bei 1 liegt, desto wahrscheinlicher klassifiziert das Neuronale Netz das Ereignis als Signal. Ein Ereignis mit einer Ausgabe nahe 0 wird dementsprechend als Untergrund-Ereignis eingeordnet. Im DNN-Klassifikations-Histogramm wird nun die Häufigkeit (auf der y-Achse) der verschiedenen DNN-Vorhersagen (auf der x-Achse) dargestellt. In Abbildung 8.4 ist beispielsweise das DNN-Klassifikations-Histogramm des Modells 2 dargestellt.

In diesem Histogramm kann man einfach ablesen, welche Klassifikationen wie häufig vorkommen. Bei Modell 2 sind zum Beispiel Klassifikationen nahe 0 und 1 sehr häufig, wohingegen unklare Klassifikationen dazwischen (0,2 bis 0,8) fast nie vorkommen. Das Neuronale Netz unterscheidet also klar Signal- und Untergrund-Ereignisse. Ob diese Unterscheidung auch korrekt ist, kann man außerdem an der Färbung der Balken erkennen. Die blauen, violetten und grünen Balkenanteile sind Untergrund-Ereignisse und nur die roten Anteile sind Signal-Ereignisse (die gelben, orangenen und braunen Anteile sind ebenfalls Signal-Ereignisse, sie spielen aufgrund ihrer geringen Anzahl aber in der beurteilenden Auswertung per Augenmaß keine Rolle). Wie man also im Histogramm erkennen kann, sind keine Signal-Ereignisse als Untergrund klassifiziert worden (DNN-Vorhersage nahe 0), aber auch einige Untergrund-Ereignisse als Signal (DNN-Vorhersage nahe 1).

Es ergibt Sinn, sich zusätzlich zur ROC-Kurve weitere Histogramme mit der DNN-Vorhersage anzusehen, weil ein kleiner Loss und eine gute ROC-Kurve nicht unbedingt für ein gut funktionierendes Modell stehen. Ein kleiner Loss und ein hoher AUC-Wert entstehen nämlich auch, wenn das Neuronale Netz alle Ereignisse klar als Untergrund klassifiziert, auch die Signal-Ereignisse. Durch

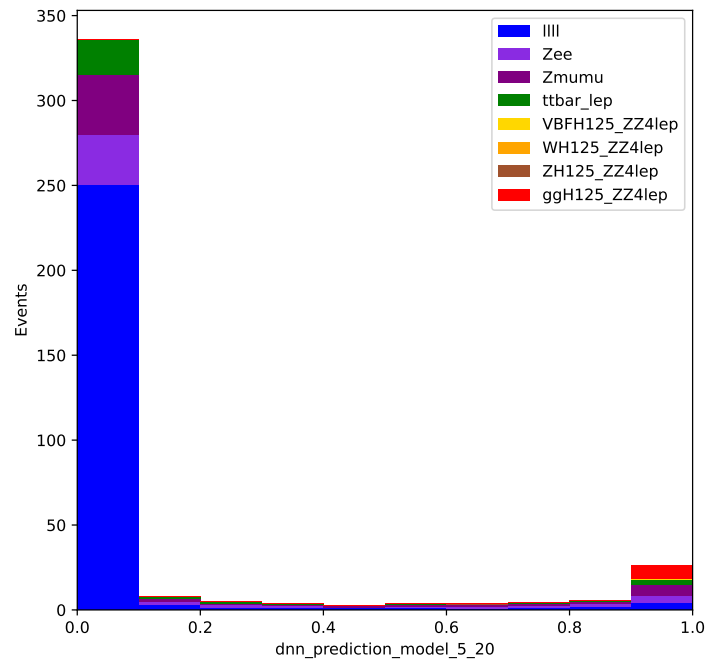


Abbildung 8.4: DNN-Klassifikations-Histogramm Modell 2

die enorme Überzahl der Untergrund- gegenüber den Signal-Ereignissen entsteht so zum Beispiel ein sehr kleiner Loss, der eigentliche Zweck wird aber nicht erfüllt, nämlich dass das Signal klar als Signal klassifiziert wird.

Ein anderes zur Auswertung verwendetes Diagramm stellt die Trainings- und Validierungs-Loss im Verlauf des Trainings dar. Ein Beispiel für solch ein Diagramm ist in Abbildung 8.5 für Modell 2 dargestellt. Daran kann man erkennen, wie sich die Qualität des Neuronalen Netzes während des Trainings verhält. Die gepunkteten Kurven stellen dabei immer die Trainings-Loss und die durchgezogenen Linien die Validierungs-Loss dar.

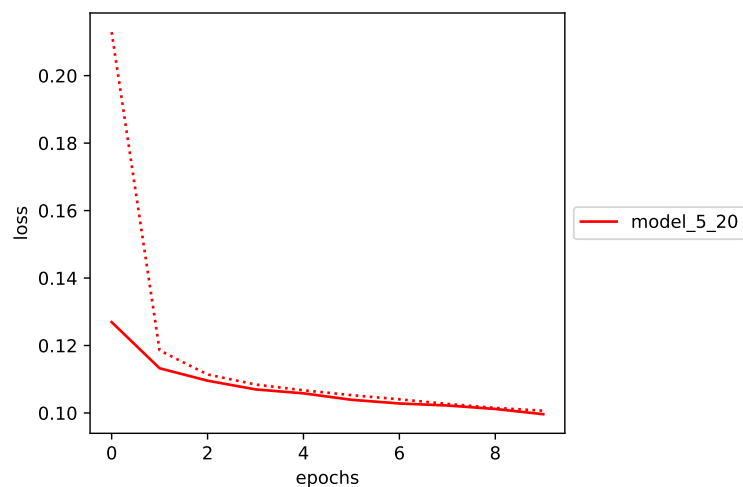


Abbildung 8.5: Loss-Diagramm für Modell 2 (die erste Epoche wird im Programmcode und in den Diagrammen immer als die Epoche 0 definiert)

Die beiden Linien sollten für ein Modell möglichst nah beieinander liegen, aber wie auch hier zu erkennen ist, ist der Validations-Loss in den ersten Epochen immer besser als der Trainings-Loss. Das liegt daran, dass der Validations-Loss nach dem Abschluss der Epoche auf den Validations-Daten berechnet wird, der Trainings-Loss aber während des Trainings. Er ist also ein Mittelwert für die Epoche. Und da gerade bei den ersten Epochen der Trainings-Loss am Anfang der Epoche sehr viel höher ist als am Ende, ist auch der Mittelwert des Trainings-Loss größer als der Validationsloss (und auch größer als der Trainings-Loss, den man erhalten würde, wenn man den Loss am Ende der Epoche auf den Trainings-Daten berechnet).

8.5 Ereignisgewichte

Das Training funktioniert mit diesem Setup zwar schon relativ gut, eine genauere Betrachtung muss man aber noch den Ereignisgewichten schenken. Denn das aktuelle Training der Neuronale Netze ist ohne Ereignisgewichte und somit ungenau, was daran liegt, dass der verwendete Datensatz vom Computer simuliert ist. Dabei werden verschiedene Ereignisse unterschiedlich oft simuliert und dann entsprechend ihrer tatsächlichen Häufigkeit in der Realität gewichtet. So kann es zum Beispiel sein, dass einzelne Ereignisse, die in der Realität selten vorkommen, auf sehr vielen Simulationen beruhen, die dann durch ihre sehr kleine Wichtung auf die tatsächlich niedrigere Anzahl skaliert werden. Oder umgekehrt ist es auch möglich, dass zu in der Realität oft vorkommenden Ereignissen nur wenige Simulations-Rohdaten vorliegen, die dann durch hohe Ereignisgewichte hochskaliert werden. Beachtet man diese Ereignisgewichte beim Training nicht und trainiert das Neuronale Netz ohne Gewichte, führt das dazu, dass tatsächlich seltene Prozesse für das Neuronale Netz „wichtiger“ erscheinen, nur weil dazu mehr Rohdaten vorliegen. Somit „konzentriert“ sich das Training auf seltene Ereignisse und funktioniert im Anschluss schlechter. Um diesen Effekt zu verhindern, werden die Ereignisgewichte im Training als weiterer Eingabewert direkt mit einbezogen.

Dabei trat allerdings ein weiteres Problem auf, was durch die Simulationsdaten verursacht wird. Denn ein Teil der Daten im Datensatz hat negative Ereignisgewichte, das heißt praktisch gesehen werden diese Daten dann als „negative“ Daten angesehen und interpretiert, was in der weiteren Auswertung zu Unsinnigkeiten führt. Durch genauere Analyse wurde klar, dass diese negativen Ereignisgewichte vor allem bei bestimmten Zerfallskanälen auftritt, die vom Generator „SHERPA 2.2“ erzeugt wurden. [14] Dieser Generator ist bekannt dafür, dass das Problem mit den negativen Ereignisgewichten auftritt. Als Lösung wurde für diese Arbeit einfach für jedes Ereignisgewicht der Betrag als neues Ereignisgewicht definiert. Weiterhin müssen nun unabhängig vom Problem der negativen Ereignisgewichte auch in anderen Auswertungsmethoden und -diagrammen ab nun die Ereignisgewichte mit einbezogen werden, wie beispielsweise bei der Berechnung der ROC-Kurve.

Das Problem der negativen Ereignisgewichte war nun gelöst, allerdings trat bei dem Training mit Ereignisgewichten nun ein weiteres Problem auf, das direkt mit den Gewichten zusammenhängt. Zur Veranschaulichung dieses Problems sind in den folgenden Abbildungen die DNN-Vorhersage-Histogramme für das Modell 2 dargestellt, ein mal für das Training ohne Ereignisgewichte (Abbildung 8.6) und einmal für das Training mit Ereignisgewichten (Abbildung 8.7).

Wie man in den Histogrammen sehen kann, funktioniert die Unterscheidung von Signal und Untergrund beim Training ohne Wichtung schon gut, aus oben genannten Gründen ist eine Wichtung aber dennoch nötig. Beim Training mit Wichtung tritt aber offensichtlich ein Problem auf, denn fast alle Ereignisse sind als Untergrund (DNN-Vorhersage nahe 0) und nahezu keine als Signal klassifiziert worden. Außerdem sind die Signalanteile des Histogramms scheinbar wahllos verteilt, sodass auch viele Signal-Ereignisse eine DNN-Vorhersage von nahe 0 haben. Daher kann die ursprüngliche Funktion des Neuronale Netzes, die Klassifikation der Ereignisse als Signal oder Untergrund nicht mehr gewährleistet werden. Das kommt dadurch zustande, dass es ursprünglich viele Signal- und viele Untergrund-Rohdaten gab. Durch die Ereignisgewichte werden die Daten nun auf ihre reale Häufigkeit skaliert. Da Signal-Ereignisse in der Realität sehr selten sind (nur 2,48% der Ereignisse), wird das Modell nun auch auf sehr viele Untergrund-Ereignisse und sehr wenige Signal-Ereignisse trainiert. Die logische Konsequenz ist, dass das Neuronale Netz einfach fast alle Ereignisse als Untergrund klassifiziert und damit in 97 bis 98% der Fälle richtig liegt. Damit wird ein sehr kleiner Loss-Wert erzeugt und im Training wird dieser Effekt verstärkt. Es musste also eine Lösung gefunden werden, sodass das Training nicht nur einseitig auf Untergrund-Ereignissen

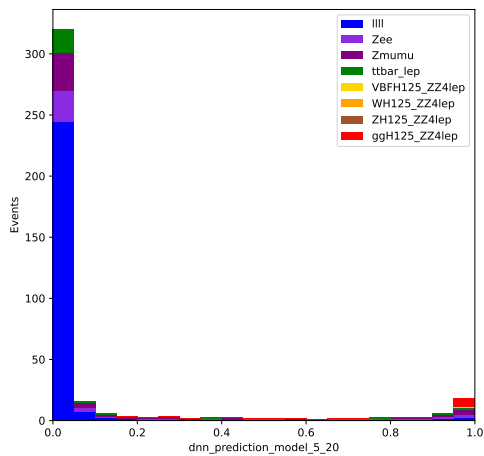


Abbildung 8.6: DNN-Vorhersage-Histogramm bei Training ohne Wichtung

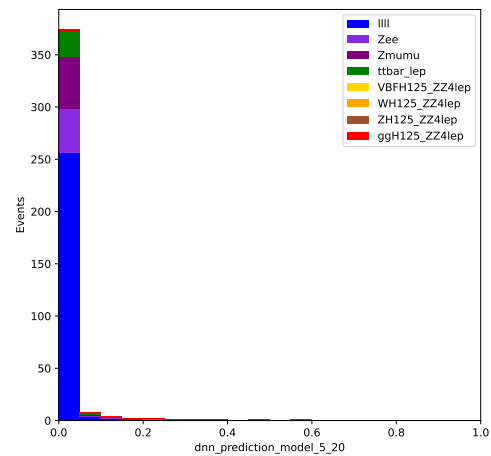


Abbildung 8.7: DNN-Vorhersage-Histogramm bei Training mit Wichtung

stattfindet. Dazu wurden im Folgenden die Ereignisgewichte so skaliert, dass gleich viele Signal- wie Untergrund-Ereignisse existieren. Im folgenden Programmcode ist die Implementation der Lösung dargestellt.

```

1 | def reweight_weights(weights, classification):
2 |     weights_abs = np.abs(weights)
3 |     weights_signal = weights_abs * classification
4 |     weights_background = weights_abs * (1 - classification)
5 |     weights_reweighted = weights_background / sum(weights_background)
6 |     weights_reweighted += weights_signal / sum(weights_signal)
7 |     weights_reweighted /= weights_reweighted.mean()
8 |     return weights_reweighted
9 |
10 | train_weights_normed = reweight_weights(train_weights,
11 |                                         train_classification)
12 | val_weights_normed = reweight_weights(val_weights,
13 |                                       val_classification)
14 | test_weights_normed = reweight_weights(test_weights,
15 |                                       test_classification)

```

Es wurde hier eine Funktion definiert, die die Gewichte entsprechend der erklärten Anforderungen umgewichtet. In dieser Funktion werden zuerst von allen Gewichten die Beträge berechnet, um die negativen Ereignisgewichte zu entfernen (Zeile 2). Danach werden die Ereignisgewichte so verändert, dass die Summe aller Signal-Ereignisse ebenso wie die Summe aller Untergrund-Ereignisse 1 ist (Zeilen 3 bis 6). Dazu wurden jeweils für Signal- und Untergrund-Ereignisse die Ereignisgewichte durch die Summe der Ereignisgewichte geteilt, wodurch die neue Summe der Ereignisgewichte 1 ist. Danach wurden die Gewichte noch normiert (Zeile 7), sodass der Mittelwert aller Gewichte für den Trainings-, Validations- und Test-Datensatz 1 ist (das dient dazu, dass die Loss-Werte vergleichbar sind). Diese „Umwichtungs-Funktion“ wird dann auf die Trainings-, Validations- und Test-Datensätze angewandt (Zeilen 10 bis 12).

Mit der Umwichtung wurde nun erreicht, dass die Ereignisgewichte weiterhin in das Training eingehen, ohne dass die Signal-Ereignisse zu wenig berücksichtigt werden. In der Abbildung 8.8 ist das DNN-Vorhersage-Histogramm für das Modell 2, trainiert mit überarbeiteter Wichtung dargestellt. Wie man darin erkennen kann, tritt das beschriebene Problem für die normale Wichtung

nicht mehr auf, dass sich das Neuronale Netz zu sehr auf die Untergrund-Daten konzentriert. Auch kann man im Vergleich zum DNN-Vorhersage-Histogramm für das Modell 2 ohne Wichtung (Abbildung 8.6) erkennen, dass in den letzten Bins (mit einer DNN-Vorhersage nahe 1) der Untergrund-Anteil in den Simulationsdaten deutlich verringert wurde, was auch für ein gut klassifizierendes Neuronales Netz spricht.

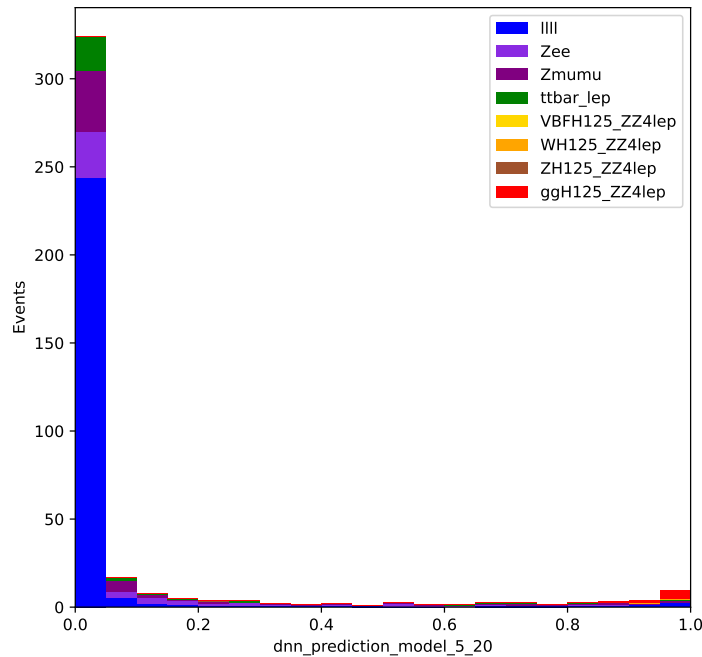


Abbildung 8.8: DNN-Vorhersage-Histogramm bei Training mit neuer Wichtung (Modell 2)

Zu Beachten ist außerdem noch, dass das Training zwar nun immer mit den neuen Ereignisgewichten abläuft, aber die Histogramme immer noch mit den „normalen“ berechnet werden. Daher ist die Gesamtanzahl der Ereignisse auch in allen Histogrammen unabhängig von der Wichtung im Training gleich.

8.6 Overtraining

Ein Problem, das vor allem beim Training eines großen Neuronalen Netzes auftreten kann ist Overtraining. Irgendwann im Training wird das Neuronale Netz so gut angepasst, dass es die Eingabedaten nur noch auswendig lernt, ohne generelle sinnvolle Regeln für die Klassifikation zu finden. Dadurch wird zwar ein kleiner Trainings-Loss erzeugt aber der ursprüngliche Zweck ist nicht mehr erfüllt, da das Neuronale Netz dann nur sehr gut auf den auswendig gelernten Trainings-Daten funktioniert. Man erkennt Overtraining also daran, dass der Trainings-Loss erheblich kleiner als der Validations- und Test-Loss ist. Das Ziel beim Training ist nun, ein Neuronales Netz so lange wie möglich zu trainieren sodass es so gut wie möglich angepasst wird, aber das Training abubrechen, bevor zu viel Overtraining auftritt. Das wird mit einer Funktion realisiert, die Early-Stopping genannt wird. Nach jeder Trainings-Epoche wird überprüft, wie groß der Validations-Loss des Neuronalen Netzes ist. Wird dieser Loss über mehrere Epochen nicht kleiner als der bisher kleinste Wert, wird das Training automatisch abgebrochen. Somit läuft das Training nur so lange, wie das Neuronale Netz auch wirklich bessere Ergebnisse erzielt. Da das Training und damit auch der Loss von Epoche zu Epoche immer statistischen Schwankungen unterliegt, ist es aber wichtig,

dass das Training nicht sofort abgebrochen wird, wenn der Validations-Loss einmal schlechter wird, sondern nur, wenn das mehrere Epochen hintereinander geschieht. Die Anzahl der Epochen, nach denen das Training abgebrochen wird, wenn der Validations-Loss nicht besser wird als der bisher beste Wert wird Toleranz genannt. Für das weitere Training habe ich beispielsweise eine Toleranz von fünf Epochen genutzt. Außerdem ist das Neuronale Netz nach Trainingsabbruch immer schlechter als es vorher schon einmal war (sonst wäre das Training ja nicht abgebrochen worden). Daher muss das Modell immer die Einstellungen (Gewichte) für die bisher beste Epoche, also den bisher besten Zustand des Modells, speichern. Nach Abbruch des Trainings bestimmt dieser dann das resultierende Neuronale Netz, sodass dessen Qualität so hoch wie möglich ist. Dies wird im Programmcode ebenfalls durch die Early-Stopping-Funktion realisiert und „restore best weights“ genannt. Die finale Implementation im Programmcode ist dann wie folgt:

```

1 | import tensorflow
2 | early_stopping = tensorflow.keras.callbacks.EarlyStopping(
   |     monitor='val_loss', patience=5, restore_best_weights=True)

```

Um auch nach dem Training ein Maß für den Overtraining-Grad zu haben, kann man den Quotient aus Validations- und Trainings-Loss verwenden. Je größer dieser Wert, desto mehr Overtraining tritt bei dem Neuronalen Netz auf, denn ein großer Validations-Loss gegenüber einem kleinen Trainings-Loss spricht für Overtraining. Im besten Fall ist dieser Quotient 1, denn das Neuronale Netz sollte auf Validations-Daten etwa genauso gut funktionieren wie auf Trainings-Daten.

Da ein Wert nicht besonders aussagekräftig ist, wurde in der weiteren Auswertung und Hyperparameter-Optimierung außerdem ein weiteres Histogramm zur Auswertung verwendet. In der folgenden Abbildung ist ein solches Trainings-Validations-Histogramm für Modell 2 abgebildet.

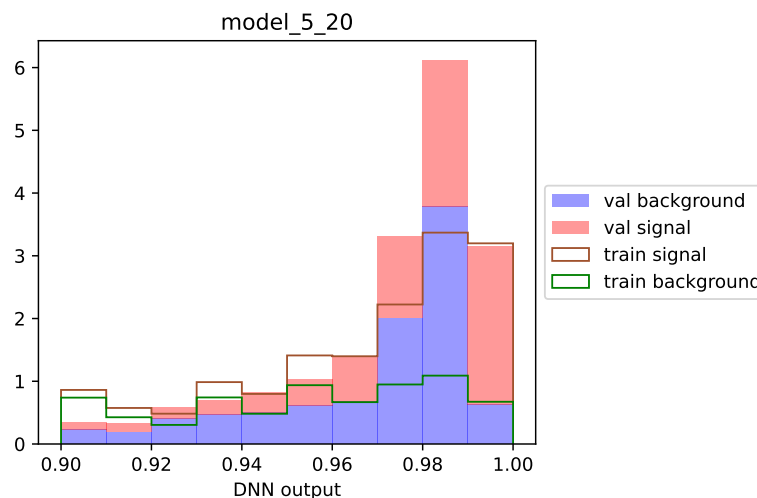


Abbildung 8.9: Trainings-Validations-Histogramm für Modell 2 (Hinweis: in diesen Histogrammen ist nur der interessante Anteil der DNN-Vorhersage (x -Achse) von 0,9 bis 1 dargestellt)

Wie schon beim DNN-Vorhersage-Histogramm ist die Häufigkeit der Ausgabewerte des Neuronalen Netzes dargestellt. Das Problem beim DNN-Vorhersage-Histogramm war aber eben, dass keine Unterscheidung zwischen den Datensätzen (Trainings- und Validations-Daten) möglich ist. Das heißt, es kann sein, dass alle gut eingeordneten Anteile der Daten im Histogramm einfach schon zum Trainings-Datensatz gehören, auf das das Neuronale Netz trainiert wurde. Im neuen Trainings-Validations-Histogramm sind Trainings-Daten und Validations-Daten getrennt dargestellt, Trainings-Daten als Linien und Validations-Daten als ausgefüllte Balken. Anhand dieses Histogramms kann man also einzeln sehen, wie die DNN-Vorhersage-Verteilung bei Trainings- und Validations-Daten ist. Außerdem kann man diese beiden Datensätze direkt vergleichen. Wenn die Linien und die ausgefüllten Balken also nahe beieinander liegen, spricht das für wenig Overtraining.

Weit auseinander liegende Verteilungen für die beiden Datensätze spricht dagegen eher für eine zu gute Anpassung an die Trainings-Daten, also Overtraining. Bei diesem Histogramm (für Modell 2) ist die Übereinstimmung von Trainings- und Validations-Daten beispielsweise sehr gut, vor allem im Signal-Anteil. Der Untergrund-Anteil unterliegt stärkeren Schwankungen, die aber im Rahmen der statistischen Unsicherheit liegen. Also ist Overfitting in diesem Modell nicht problematisch.

8.7 Hyperparameter-Optimierung

Bei der Hyperparameter-Optimierung wird die optimale Größe und Struktur des Neuronalen Netzes ermittelt, um eine größtmögliche Funktionsfähigkeit zu erhalten. Die Funktionsfähigkeit des Neuronalen Netzes bestimmt wie gut das Neuronale Netz die Trainings-, Validations- und Test-Daten kategorisiert und wie gut die Übereinstimmung der Loss ist (Overfitting-Grad). Optimiert wurden in dieser Arbeit die Anzahl der Schichten des Neuronalen Netzwerks und die jeweilige Anzahl der Neuronen in jeder Schicht. Die Optimierung dieser Hyperparameter erfolgt in mehreren Schritten. In jedem Schritt werden drei verschiedene Anzahlen von Schichten und Neuronen ausgewählt und alle neun Kombinationen aus Schichten- und Neuronenanzahl ausprobiert und verglichen. In den ersten beiden Schritten wird der Bereich für die Hyperparameter eingeschränkt, die ein optimales Netz hat. Im dritten Optimierungsschritt wird dann ein konkretes Neuronales Netz bzw. eine konkrete Kombination aus Anzahl der Schichten und Anzahl der Neuronen ausgewählt.

Die Benennung der analysierten Modelle erfolgt im Folgenden immer nach dem Muster „Modell_x.y“, wobei x die Anzahl der Schichten und y die Anzahl der Neuronen je Schicht angibt.

8.7.1 Erster Optimierungsschritt

Im ersten Optimierungsschritt wurden sehr weit auseinander liegende Werte für die Anzahl der Schichten und Neuronen gewählt, um einen möglichst großen Bereich abzudecken. Daher wurden als Modelle alle Kombinationen aus 2, 6 und 10 Schichten und je 10, 50 und 100 Neuronen trainiert. Zum Training verwendet wurden dabei alle erklärten Prinzipien: Normalisierung der Input-Parameter, Eliminierung negativer Ereignisgewichte, Umwichtung der Ereignisgewichte und Early-Stopping. Zur Auswertung wurde ein Loss-Diagramm gezeichnet (das den Verlauf der verschiedenen Loss-Werte für die Modelle im Training darstellt), siehe Abbildung 8.10. Daraus kann man ungefähr ablesen, welche Modelle geringere Loss-Werte erzeugen und damit die Daten besser kategorisieren.

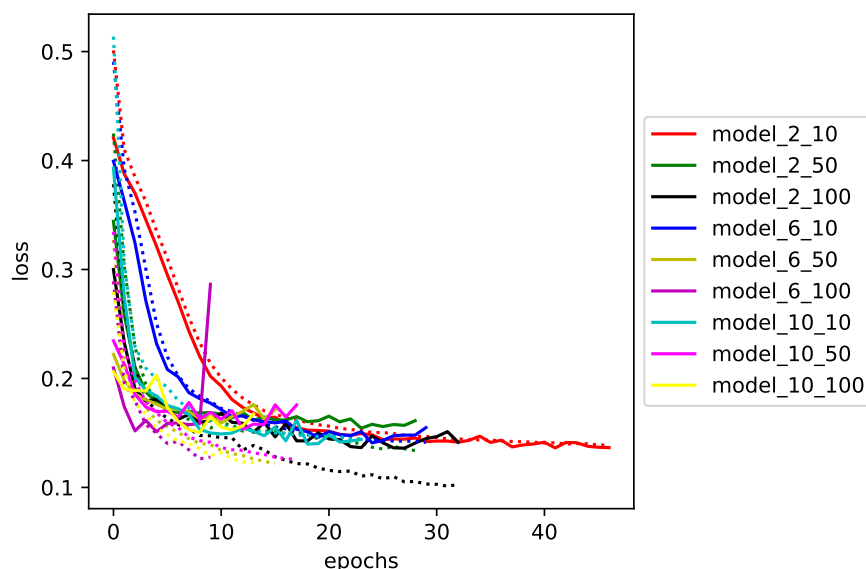


Abbildung 8.10: Loss-Diagramm mit den Loss-Verläufen der Modelle (Validations-Loss sind die durchgezogenen, Trainings-Loss die gepunkteten Linien)

Außerdem wurde eine Heatmap für den Quotient $\frac{\text{Validations-Loss}}{\text{Trainings-Loss}}$ erstellt. Diese stellt anhand von Farbtönen dar, wie groß dieser Quotient für jedes Modell ist.

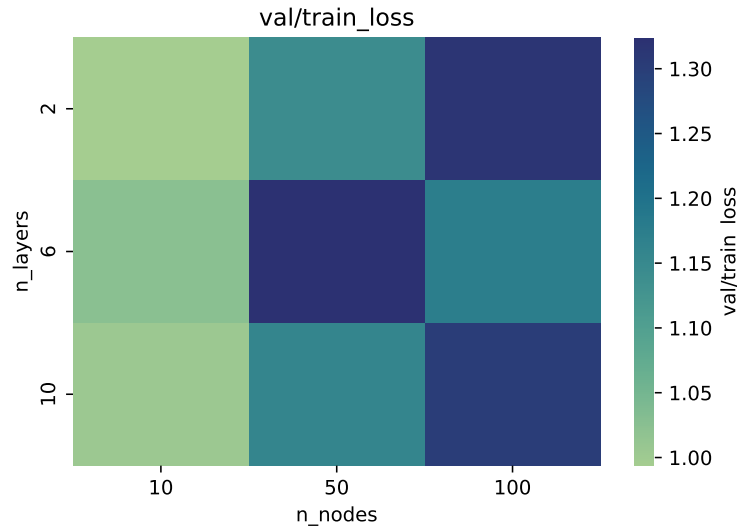


Abbildung 8.11: Heatmap zum Quotient Validation-Loss/ Train-Loss

Als dritte und komplexeste Auswertungsmethode wurde zu jedem Modell ein Trainings-Validations-Histogramm dargestellt. Aus diesem kann man jeweils ablesen, ob und wie viel Overtraining auftritt. Hier sind zwei Beispiele für die Modelle 2_10 und 6_50 dargestellt, alle Trainings-Validations-Histogramme befinden sich im Anhang.

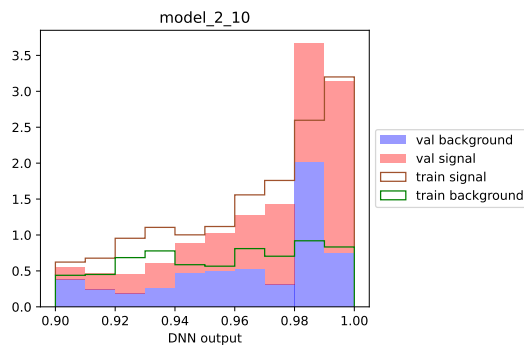


Abbildung 8.12: Trainings-Validations-Histogramm für das Modell 2_10

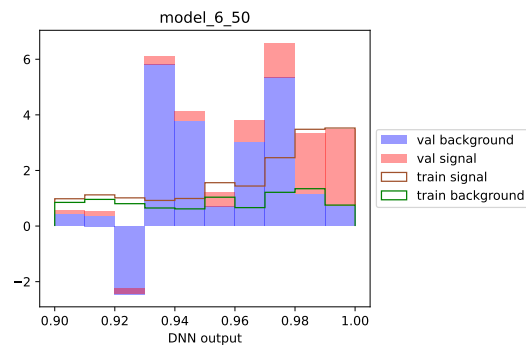


Abbildung 8.13: Trainings-Validations-Histogramm für das Modell 6_50

Schon am Loss-Diagramm (Abbildung 8.10) kann man erkennen, dass die minimalen Validation-Loss-Werte mit geringen Schwankungen für alle Modelle etwa gleich sind. Es kann keine generelle Aussage getroffen werden, ob Modelle mit mehr oder weniger Schichten oder Neuronen erfolgreicher sind. Auffällig ist allerdings in diesem Diagramm, dass vor allem bei den Modellen mit 10 Schichten oder 100 Neuronen die Differenzen zwischen den Trainings- und Validation-Loss groß sind, was auf Overtraining hindeutet. Auch bei der Heatmap (Abbildung 8.11) kann man diesen Eindruck bestätigen: der Quotient aus Validations- und Trainings-Loss ist vor allem bei Modellen mit 100 Neuronen, aber auch schon bei Modellen mit 50 Neuronen groß. Das heißt, Neuronale Netze mit dieser hohen Anzahl von Neuronen laufen auf Overtraining hinaus. Daher müssen in den folgenden Optimierungsschritten vor allem geringere Neuronen-Anzahlen ausprobiert werden. Auch durch die Analyse der Trainings-Validations-Histogramme kann der bereits gewonnene Eindruck bestätigt werden. Die Übereinstimmung von Trainings- und Validations-Daten ist bei einfache-

ren Modellen noch sehr gut (Abbildung 8.12), bei komplexeren Modellen wie Modell 6_50 (mit 6 Schichten und je 50 Neuronen) sind die Abweichungen dagegen schon größer (Abbildung 8.13). Aufgrund dessen wurde im folgenden zweiten Optimierungsschritt als Obergrenze für die Anzahl der Neuronen 40 festgelegt (da schon ab 50 Neuronen teilweise deutliches Overfitting auftrat). Da in keiner der Statistiken ein negativer Einfluss großer Anzahlen von Schichten auffällt (mit steigender Schichten-Anzahl nimmt weder der Validations-Loss zu noch steigt der Quotient aus Validations- und Trainings-Loss erheblich) wurde die obere Grenze für die Anzahl der Schichten nochmals auf 12 erhöht, um ein weiteres Spektrum von Modellen auszuprobieren.

8.7.2 Zweiter Optimierungsschritt

Im zweiten Optimierungsschritt wurden also alle Kombinationen aus 2, 7 und 12 Schichten und je 8, 20 und 40 Neuronen trainiert. In der Abbildung 8.14 ist das Loss-Diagramm für das Training dargestellt.

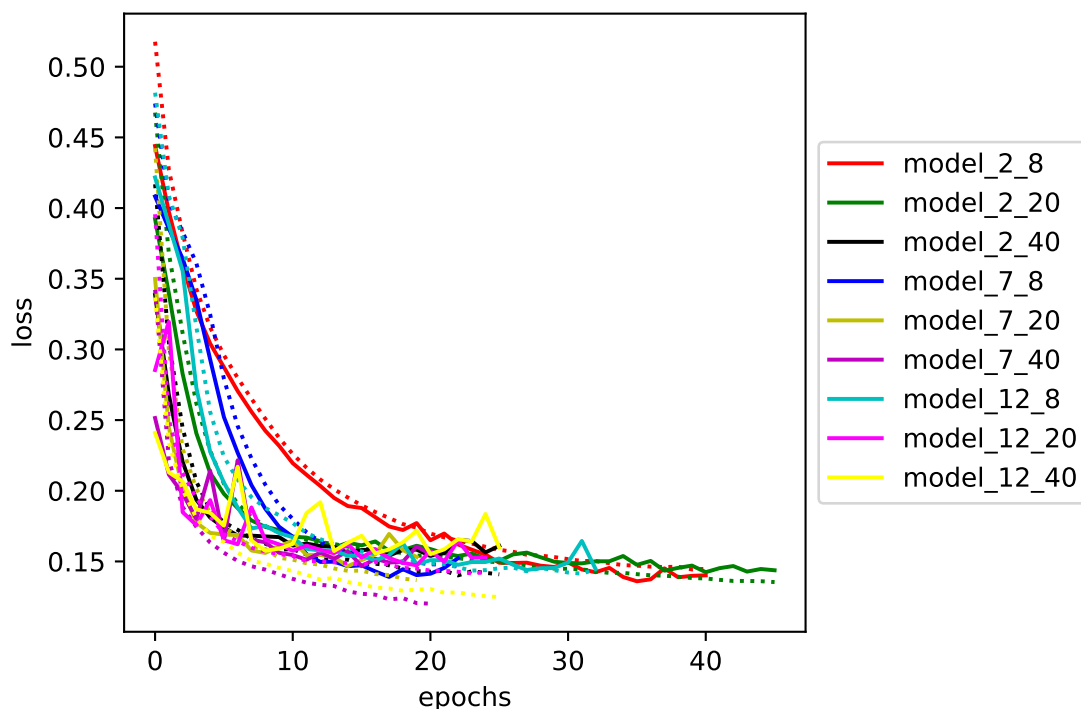


Abbildung 8.14: Loss-Diagramm für den zweiten Optimierungsschritt

Darin kann man keine besonderen Auffälligkeiten erkennen. Deutlich wird nur wieder, dass die Validation-Loss größerer Modelle (zum Beispiel Modell 12_40) im Training größeren Schwankungen unterliegt. Auch wird die Differenz zwischen Trainings- und Validations-Loss mit steigender Epochenanzahl größer, vor allem bei Modellen mit 40 Neuronen je Schicht. An beidem kann man erkennen, dass bei Modellen mit 40 Neuronen je Schicht vermutlich in gewissem Maß Overtraining auftritt. Das muss aber natürlich noch an den anderen Auswertungsmethoden überprüft werden.

In den Abbildungen 8.15 und 8.16 sind die Heatmaps für den Validation-Loss und den Quotienten aus Validation- und Trainings-Loss dargestellt.

Auch darin kann man sehen (siehe Abbildung 8.16), dass bei Modellen mit 40 Neuronen je Schicht (gerade bei größerer Schichten-Anzahl) Overtraining auftritt. Am wenigsten Overtraining tritt erwartungsgemäß bei kleineren Modellen auf, am kleinsten ist der Quotient aus Validations- und Trainings-Loss bei Modellen mit 2 bis 7 Schichten und 8 bis 20 Neuronen je Schicht. Der dritte Optimierungsschritt sollte sich also etwa in diesem Gebiet bewegen. An der anderen Heatmap (Abbildung 8.15) kann man sehen, dass der Validations-Loss auch in diesem Gebiet am kleinsten ist, was für ein gut funktionierendes Modell spricht.

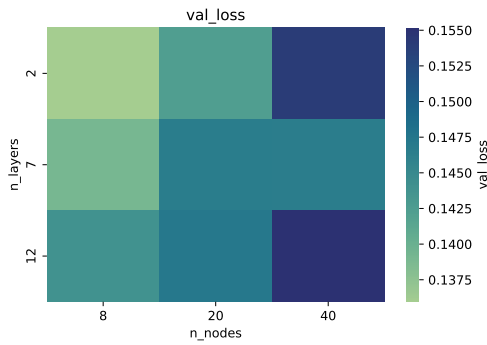


Abbildung 8.15: Heatmap für den Validation-Loss

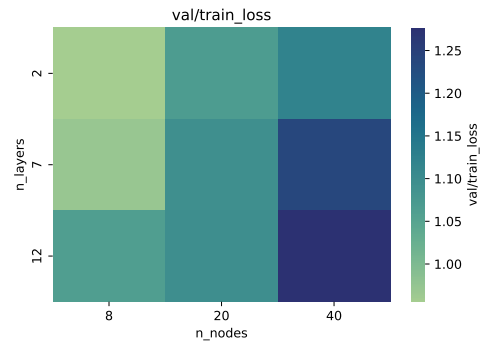


Abbildung 8.16: Heatmap für den Quotienten aus Validation- und Trainings-Loss

Zuletzt müssen noch die Trainings-Validations-Histogramme ausgewertet werden. Dazu hier zwei representantive Histogramme für das kleinste und größte Modell (Abbildung 8.17 und 8.18).

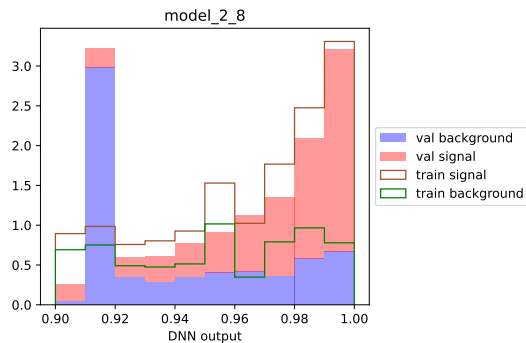


Abbildung 8.17: Trainings-Validations-Histogramm für das Modell 2.8

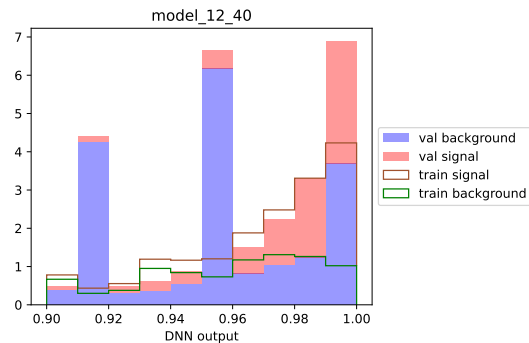


Abbildung 8.18: Trainings-Validations-Histogramm für das Modell 12.40

Wie man darin erkennen kann, ist die Übereinstimmung von Trainings- und Validations-Daten beim einfachen Modell 2.8 sehr gut (nur eine signifikante Abweichung in einem Bin), beim Modell 12.40 ist die Abweichung dagegen schon in mehreren Bins sehr hoch im Vergleich zum erwarteten Wert. Das könnte auch Zufall sein, gerade weil diese große Abweichung in nur zwei Bins auftritt, aber da auch in den anderen Modellen mit 40 Neuronen je Schicht die Übereinstimmung nicht gut ist, kann davon ausgegangen werden, dass hier Overtraining auftritt.

8.7.3 Dritter Optimierungsschritt

Im dritten Optimierungsschritt wurden also Modelle im Bereich von etwa 2 bis 7 Schichten und etwa 8 bis 20 Nueronen je Schicht trainiert. Konkret wurden alle Kombinationen aus 2, 5 und 8 Schichten und je 8, 13 und 18 Neuronen trainiert.

In der Abbildung (8.19) ist das Loss-Diagramm des Trainings dargestellt. Darin kann man fast keine Unterschiede zwischen den Modellen erkennen, deren Loss-Verläufe während des Trainings sich sehr ähneln. Auch sind keine Auffälligkeiten zu bemerken.

Anhand der Heatmap für den Validations-Loss (Abbildung 8.20) kann man aber erkennen, dass trotz der geringen Unterschiede die Modelle 8.8 und 2.18 am besten abgeschnitten haben. Aus der Heatmap für den Quotienten aus Validations- und Trainings-Loss (Abbildung 8.21) lässt sich außerdem ablesen, dass bei allen Modellen sehr wenig Overtraining auftritt (höchster Wert ca. 1,1). Aber vor allem bei Modellen mit 8 Neuronen je Schicht tritt gar kein Overtraining auf (Wert ca. 1). Deshalb wurde als bestes Modell aus dem letzten Optimierungsschritt das Modell 8.8 mit 8 Schichten aus je 8 Neuronen ausgewählt. Bei diesem Modell ist der Validations-Loss am

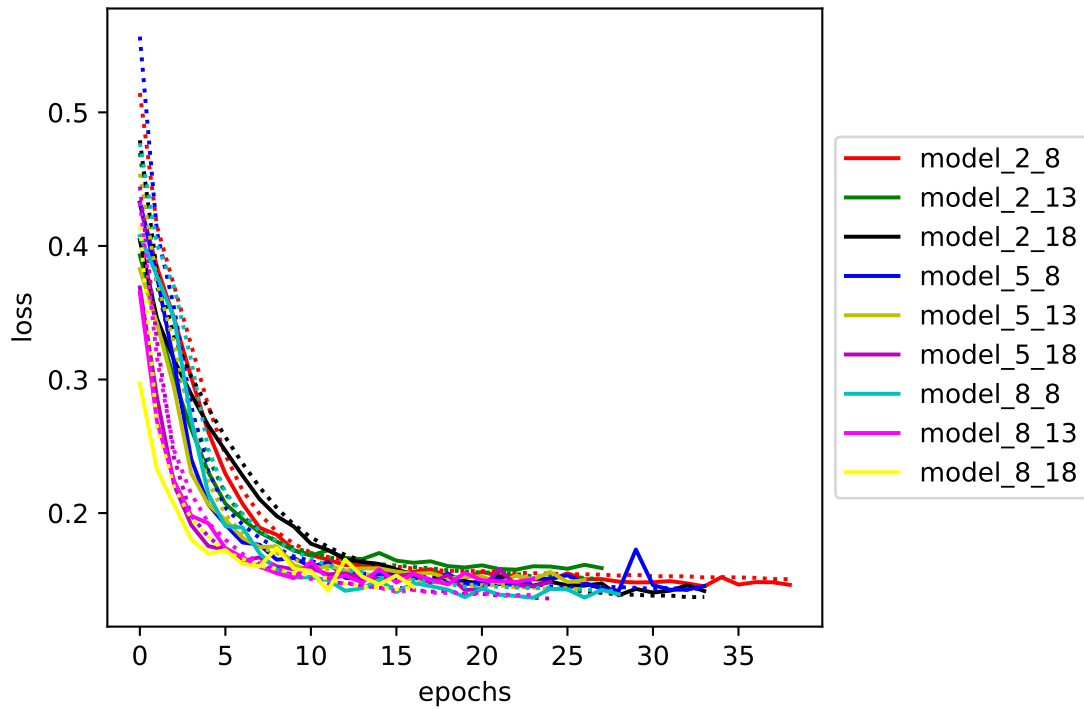


Abbildung 8.19: Loss-Diagramm für den dritten Optimierungsschritt

kleinsten und außerdem tritt kein Overtraining auf. Das kann man auch durch die Auswertung der Trainings-Validations-Histogramme bestätigen (siehe Anhang). An diesen lässt sich erkennen, dass in allen Modellen, in denen scheinbar ein leichtes Overtraining auftritt auch zufällige Aspekte einen erhöhten Quotienten aus Validations- und Trainings-Loss erzeugt haben könnten.



Abbildung 8.20: Heatmap für den Validations-Loss

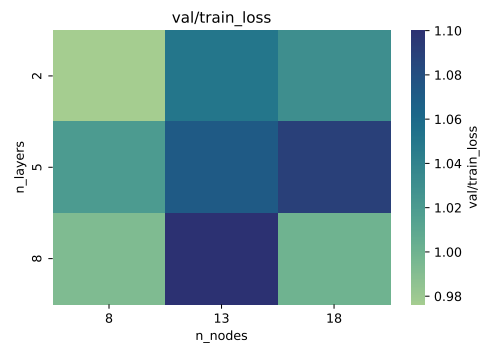


Abbildung 8.21: Heatmap für den Quotienten aus Validations- und Trainings-Loss

Das Histogramm für das ausgeählte Modell 8.8 (Abbildung 8.22) lässt außerdem darauf schließen, dass beim Training dieses Modells fast gar kein Overtraining auftritt, denn in fast allen Bins ist die Übereinstimmung von Trainings- und Validations-Daten sehr hoch.

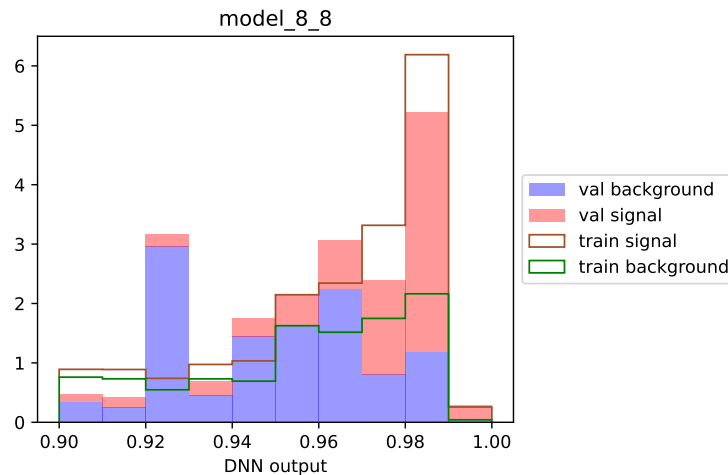


Abbildung 8.22: Trainings-Validations-Histogramm für das Modell 8_8

8.7.4 Test des ausgewählten Modells

Als letzter Schritt der Hyperparameter-Optimierung muss das ausgewählte Modell mit dem Test-Datensatz analysiert werden. Mit dem Trainings-Datensatz wurden alle Modelle trainiert. Um auszuschließen, dass das Training nur bei diesen speziellen Daten gut funktioniert, musste anschließend mit dem Validations-Datensatz überprüft werden, ob das Neuronale Netz auch auf diesen Daten funktioniert. Das wurde für alle Modelle durchgeführt und so das Modell ausgewählt, bei dem wenig Overfitting auftritt. Nun ist es aber möglich, dass auch die Validations-Daten zufällig so zusammengesetzt sind, dass sie besonders gut auf ein bestimmtes Modell passen. Bei so vielen Modellen wie hier ausgewertet wurden, ist das sogar wahrscheinlich. Man würde dann vermuten, dass das Modell (wie hier das Modell 8_8) besonders gut passt, um die Daten zu analysieren. Vielleicht hat man aber auch einfach das Modell ausgesucht, das für die Validations-Daten gerade besonders gut passt. Um dies auszuschließen, hat man den Test-Datensatz. Auf diesen wird das ausgewählte Modell zuletzt angewandt. Wenn das Modell dann die Ereignisse etwa gleich gut klassifiziert wie beim Validations- und Trainings-Datensatz, ist ein Overtraining ziemlich sicher ausgeschlossen. In der folgenden Abbildung (8.23) ist also analog zum Trainings-Validations-Histogramm das Trainings-Test-Histogramm dargestellt.

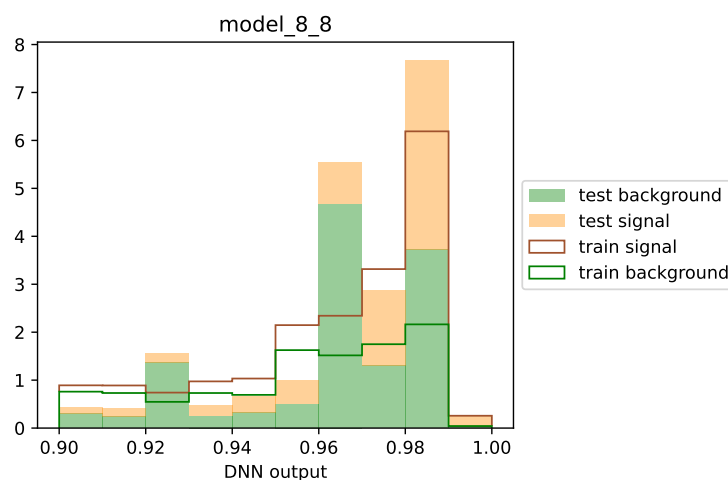


Abbildung 8.23: Trainings-Test-Histogramm für das Modell 8_8

Vergleicht man das mit dem Trainings-Validationshistogramm des gleichen Modells (Abbildung 8.22), erkennt man, dass Validations- und Test-Daten nicht gleich klassifiziert werden, was an statistischen Schwankungen liegt. Im Rahmen dieser Schwankungen ist aber auch die Übereinstimmung von Trainings- und Test-Daten gut. Daher kann das ausgewählte Modell mit 8 Schichten und je 8 Neuronen als Ergebnis der Hyperparameter-Analyse nun für die weitere Analyse der Messdaten verwendet werden.

8.8 Anwendung des Neuronalen Netzes und Signifikanzberechnung

Nach der Optimierung des Setups und der Hyperparameter-Optimierung kann nun das ausgewählte Modell mit 8 Schichten und je 8 Neuronen auf den Mess-Datensatz angewandt werden. Dazu wird das Modell zuerst auf den Trainings-Daten trainiert. (Alle Daten und Diagramme des Training siehe Anhang.) Anschließend wird für jedes Ereignis des Simulations-Datensatzes und Mess-Datensatzes der DNN-Vorhersage-Wert bestimmt. Alle Ereignisse werden dann in ein DNN-Vorhersage-Histogramm eingetragen. In der Abbildung 8.24 ist der für die Auswertung relevante Teil des Histogramms von $[0,9 \text{ bis } 1]$ dargestellt (das vollständige Histogramm befindet sich ebenfalls im Anhang).

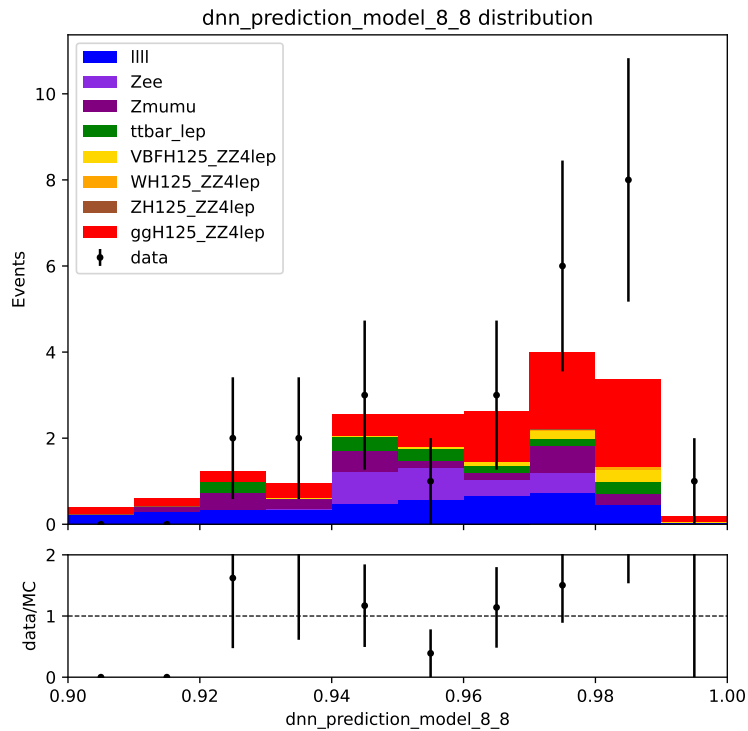


Abbildung 8.24: DNN-Vorhersage-Histogramm für alle vorhandenen Daten (Simulations- und Messdaten) im Bereich $[0,9 \text{ bis } 1]$

Wie man in dem Histogramm schon sehen kann, wurden fast alle Signal-Ereignisse der Simulationsdaten in diesem Bereich klassifiziert und nur vergleichsweise wenige Untergrund-Ereignisse wurden hier eingestuft. Das spricht für ein gut klassifizierendes Neuronales Netz, was auch auf den Messdaten gut funktionieren sollte. Generell kann man auch erkennen, dass in den meisten Bins mehr Messdaten als Simulationsdaten eingeordnet wurden. Das liegt einerseits an dem gut klassifizierenden neuronalen Netz, andererseits kann man auch schon im Histogramm der invari-

anten Masse des vier-Leptonen-Systems nach der Selektion (Abbildung 8.2) erkennen, dass in den relevanten Bins, in denen Higgs-Ereignisse zu vermuten sind (um 125 GeV) mehr Messdaten- als Simulationsdaten-Ereignisse sind. Das lässt vermuten, dass in dem vorhandenen Datensatz zufällig mehr Higgs-Ereignisse gemessen wurden, als es nach der Simulation sein müssten. Das sorgt dann auch nach der Anwendung des Neuronalen Netzes im DNN-Klassifikations-Histogramm (Abbildung 8.24) für einen Überschuss an Messdaten-Ereignissen.

Analog zur Auswertung der Higgs-Boson-Analyse (Kapitel 7.3.4) kann nun im Histogramm eine statistische Signifikanz (siehe Kapitel 4.4) berechnet werden, um das Higgs-Boson nachzuweisen, was das Ziel der Anwendung der Neuronalen Netze war. Dazu habe ich zuerst die statistische Signifikanz der Simulationsdaten berechnet. Diese gibt an, wie weit die kompletten Simulationsdaten vom reinen Untergrund-Anteil der Simulationsdaten abweichen. Damit ist sie ein Maß dafür, wie wahrscheinlich es in Bezug auf die Simulationsdaten ist, dass das Higgs-Boson existiert. In einer umfassenden und vollständig wissenschaftlichen Analyse, wie sie beispielsweise 2012 am CERN zum Nachweis des Higgs-Bosons durchgeführt wurde, wäre die Berechnung der statistischen Signifikanz natürlich nicht ausreichend. Für diese Arbeit soll allerdings eine Abschätzung der Signifikanz genügen, sodass die Berechnung der statistischen Signifikanz auf den Daten als gute Auswertungsmethode verwendet werden kann, um das Higgs-Boson nachzuweisen. Dennoch ist darauf zu achten, dass in dieser Arbeit nicht die Signifikanzen, sondern lediglich statistische Signifikanzen berechnet wurden. Das heißt nicht, dass diese Werte nicht für einen einfachen Nachweis des Higgs-Bosons verwendet werden können. Sie dürfen nur nicht mit den finalen Signifikanzen des ATLAS-Experiments verwechselt werden.

In der Formel für die Berechnung der statistischen Signifikanz

$$Z = \frac{|y - y_{exp}|}{\sqrt{y_{exp}}} \quad (8.1)$$

sind der erwartete Wert y_{exp} der Untergrund-Anteil der Simulationsdaten und der tatsächliche Wert y die kompletten Simulationsdaten. Für die Berechnung der statistischen Signifikanz wurden alle relevanten Bins (also die mit Higgs-Boson-Anteil) in einem Bin zusammengefasst und daraus die Signifikanz berechnet. Die relevanten Bins sind je nach Auswahl die letzten drei, vier, fünf oder sechs Bins. Für alle Möglichkeiten wurde hier die Signifikanz berechnet (Tabelle 8.2).

zusammengefasste Bins	Simulation (Untergrund)	Simulation (komplett)	Signifikanz
letzte 3 Bins	3,0	7,5	2,60
letzte 4 Bins	4,4	10,2	2,77
letzte 5 Bins	6,1	12,7	2,67
letzte 6 Bins	8,2	15,3	2,48

Tabelle 8.2: Ergebnisse der Signifikanzberechnung für Simulationsdaten

Je höher die Abweichung der Simulationsdaten von den Simulationsdaten ohne Higgs-Boson ist, desto geringer ist die Wahrscheinlichkeit, dass diese Fluktuation zufällig auftritt und desto höher ist demzufolge die Signifikanz. Die höchste statistische Signifikanz wird bei der Zusammenfassung der letzten vier Bins erreicht, weswegen nun für diese Kombination auch die Signifikanz für die Messdaten berechnet wird.

Denn das Ziel der Analyse ist der Nachweis des Higgs-Bosons. Das wird wie schon in Kapitel 7.3.4 erläutert durch den Vergleich von Simulationsdaten ohne Higgs-Boson und Messdaten erreicht. Denn wenn es kein Higgs-Boson gäbe, müssten die Messdaten eine Verteilung ergeben, die der Verteilung der Simulationsdaten ohne Higgs-Boson entspricht. Wenn die Verteilungen aber signifikant voneinander abweichen, kann das als ein Nachweis des Higgs-Bosons angesehen werden. Daher wurde hier die Signifikanz für die Messdaten und den Untergrund-Anteil der Simulationsdaten berechnet (für die letzten vier Bins):

$$Z_{Messdaten} = \frac{18 - 4,4}{\sqrt{4,4}} \approx 6,48 \quad (8.2)$$

8.9 Auswertung der Analyse durch Neuronale Netze

Ziel der Analyse der Higgs-Boson-Prozesse durch Neuronale Netze war wie schon bei der mathematischen Analyse der Verteilung der invarianten vier-Leptonen-Masse (Kapitel 7.3.4) der Nachweis des Higgs-Bosons. Das Problem bei der ersten Higgs-Boson-Analyse war die geringe Datenmenge. Durch die Anwendung von Neuronalen Netzen konnte diese Datenmenge nicht erhöht werden. Aber das programmierte, optimierte und trainierte Neuronale Netz konnte die Daten besser als Signal- oder Untergrund-Daten klassifizieren, als die vorher festgelegten starren Selektionskriterien. Daher ist der Anteil der Untergrund-Ereignisse deutlich gesunken, wodurch eine deutliche Erhöhung der statistischen Signifikanz möglich ist. Bemerkenswert ist allerdings, dass die Signifikanz der Messdaten deutlich höher als die Signifikanz der Simulationsdaten ist (2,77 zu 6,48). Das liegt vermutlich daran, dass im Datensatz der Messdaten ein nach der Theorie nicht zu erwartender Überschuss an Higgs-Boson-Ereignisse existiert. Das kann man beispielsweise auch im Histogramm der invarianten vier-Leptonen-Masse sehen (Abbildung 8.1). In den relevanten Bins mit Higgs-Boson-Ereignissen liegen die Messdaten über den Simulationsdaten, was darauf hindeutet, dass in dem analysierten Datensatz mehr Higgs-Bosonen-Ereignisse vorhanden sind, als nach der Wahrscheinlichkeit erwartet werden. Auch das ist wieder darauf zurückzuführen, dass der Datensatz nicht genug Daten für eine gründliche Analyse beinhaltet.

Dennoch konnte mit dem vorhandenen Datensatz eine sehr hohe Signifikanz erreicht werden. Außerdem wurde die Signifikanz durch die Anwendung von Neuronalen Netzen verbessert, wie man in der Tabelle 8.3 sehen kann. Darin sind die Signifikanzen für die Analyse der Higgs-Boson-Ereignisse aus dem Histogramm der invarianten vier-Leptonen-Masse (Kapitel 7.3.4) und mittels Selektion durch Neuronalen Netze (Kapitel 8.8) aufgeführt.

	Analyse mit Selektionskriterien	Analyse mit Neuronalen Netzen
Simulationsdaten $Z_{Simulation}$	2,0	2,77
Messdaten $Z_{Messdaten}$	2,62	6,48

Tabelle 8.3: Ergebnisse aller Berechnungen der statistischen Signifikanz der Analysen für Simulations- und Messdaten

Durch die Anwendung Neuronaler Netze konnte also nicht nur die statistische Signifikanz auf den Simulationsdaten, sondern besonders auch auf den Messdaten erhöht werden. Zu beachten ist außerdem auch noch, dass der hohe Wert für die Signifikanz auf den Messdaten (von 4,71) bei der Analyse des Histogramms der invarianten vier-Leptonen-Masse nur durch die Skalierung der Daten erreicht werden konnte, weshalb in der Tabelle 8.3 für die erste Analyse auch der Wert für die statistische Signifikanz vor der Skalierung dargestellt ist. Und schließlich konnte für die Messdaten sogar die 5σ -Grenze überschritten werden, was in der Teilchenphysik als Entdeckung des Higgs-Bosons gilt. Zu beachten ist bei der Beurteilung dieses Wertes, dass in dieser Arbeit ausschließlich die statistische Signifikanz berechnet und somit systematische Unsicherheiten nicht mit berechnet wurden. Würde man diese mit einbeziehen, ist eine Verringerung des Wertes für die Signifikanz zu erwarten, wobei vermutlich selbst damit für die Analyse durch Neuronale Netze die 5σ -Grenze überschritten werden könnte. Damit wurde das Ziel der Analyse, ein Nachweis des Higgs-Bosons, auch ohne die Skalierung der Daten erreicht.

Kapitel 9

Zusammenfassung und Ausblick

In dieser Arbeit wurden Teilchenkollisionen des LHC am CERN ausgewertet und so ein guter Einblick in einige aktuelle Forschungsmethoden der Teilchenphysik geschaffen. Des Weiteren sind die in dieser Arbeit ausgewerteten Daten erst seit Anfang 2020 frei verfügbar, weshalb diese Arbeit auch eine der Ersten Schülerarbeiten sein dürfte, die sie gründlich ausgewertet hat. Die Analyse von Zerfällen des Z-Bosons eignet sich gut dazu, sowohl physikalische Hintergründe als auch statistische Grundlagen und Methoden kennenzulernen, da sie weniger komplex und damit auch besser verständlich als andere Analysen ist. Es wurden erfolgreich viele Anpassungen vorgenommen, um aus der Analyse so viele Informationen wie möglich zu gewinnen. Dabei wurde vor allem deutlich, dass durch die richtige Wahl der Fitting-Funktion nicht nur viele Werte ermittelt werden können, sondern außerdem statistische Effekte des Experiments berücksichtigt werden. Ermitteln werden konnte so zuerst die Masse des Z-Bosons (90,77 GeV) womit das Ziel der Z-Boson-Analyse erreicht war. Zudem konnte aus der Verteilung noch die Zerfallsbreite des Z-Bosons (2,14 GeV) ablesen werden. Aus dieser wurde zusätzlich die Lebensdauer berechnet und die statistischen Schwankungen des Experiments bestimmt (1,31 GeV), wodurch ein Wert über die Unsicherheit des ATLAS-Detektors abgeleitet werden konnte. Alle ermittelten Werte (mit Ausnahme der Detektorunsicherheit, für die es keinen Vergleichswert gibt) liegen nah bei den theoretischen Werten, was für eine gute Anpassung der Fitting-Funktionen an die Messdaten spricht.

Anschließend wurde die Analyse auf das Higgs-Boson erweitert und so auch die Masse dieses Teilchens (122,3 GeV) ermittelt. Außerdem konnte gezeigt werden, dass die vorhandene Datenmenge nicht ausreichend für eine umfassende Analyse oder den Nachweis des Higgs-Bosons ist, sondern 6,25 mal mehr Daten nötig sind (das entspricht einer Luminosität von $62,5 \text{ fb}^{-1}$). Durch selbst erzeugte Daten konnte das Fitting dennoch durchgeführt und das Higgs-Boson über den Vergleich der Daten mit dem Messuntergrund nachgewiesen werden. Dabei ergab sich eine statistische Signifikanz von $4,71\sigma$, was einer „Entdeckung“ des Higgs-Bosons schon nahe kommt. Allerdings beruht diese Signifikanz auf den skalierten Daten und ist deshalb in ihrer Aussagekraft eingeschränkt. Berechnet man die statistische Signifikanz auf den unskalierten Daten, erhält man nur einen Wert von $2,62\sigma$.

Damit wurden nicht nur die Methoden der Datenauswertungen nachvollzogen und angewandt, die 2012 auch zur Entdeckung des Higgs-Bosons am CERN genutzt wurden. Die meisten in dieser Arbeit beschriebenen Auswertungsmethoden und Veränderungen wurden auch selbst informatisch implementiert.

Ein Problem bei der Analyse des Higgs-Bosons stellte die geringe Datenmenge des frei zur Verfügung stehenden Datensatzes dar, sodass ein Nachweis des Higgs-Bosons mit allein diesen Daten nicht möglich war. Daher wurde im letzten Teil dieser Arbeit ein Neuronales Netz programmiert, optimiert und trainiert, das die Selektion der Daten übernahm. Durch die daraus folgende bessere Unterscheidung von Signal- und Untergrund-Daten wurde ein Nachweis des Higgs-Bosons mit einer statistischen Signifikanz von $6,48\sigma$ erbracht, was in der Teilchenphysik als Entdeckung des Higgs-Bosons zählt. Damit konnte auch das Ziel dieser Analyse erreicht werden.

Alle in dieser Arbeit angewandten und erklärten Methoden können prinzipiell auch für andere Teilchen oder Zerfallsprozesse genutzt werden, wodurch die Arbeit auch eine gewisse Allgemeingültigkeit bekommt. Durch die Anwendung auf neue Daten aus Teilchenbeschleunigern,

beispielsweise mit höheren Kollisionsenergien, könnten somit grundsätzlich auch neue Teilchen entdeckt und ihre Eigenschaften bestimmt werden. Auch hat diese Arbeit mit dem Zerfall des Higgs-Bosons einen in der aktuellen Forschung sehr wichtigen Prozess analysiert. Denn derzeit wird in eben solchen Prozessen aktiv nach Antworten auf bisher ungeklärte Fragen der Teilchenphysik gesucht. Und schließlich konnte in dieser Arbeit auch die Relevanz der Anwendung von Neuronalen Netzen dargelegt werden. Denn aus den enormen Datenmengen, die in teilchenphysikalischen Experimenten entstehen, müssen die wenigen interessanten Ereignisse rausgefiltert werden. Es konnte gezeigt werden, dass ein Neuronales Netz, also eine „Künstliche Intelligenz“ dies nach gründlicher Optimierung deutlich besser kann als eigens festgelegte physikalisch begründete Selektionskriterien. Aufgrund dieser deutlichen Überlegenheit gegenüber klassischen Methoden sind auch Neuronale Netze derzeit wichtige Bestandteile der Auswertung teilchenphysikalischer Experimente.

Literatur

- [1] CERN Collaboration, *The Higgs boson*, URL: <https://home.cern/science/physics/higgs-boson>, accessed: 12.05.2022 (siehe S. 6).
- [2] M. Kobel, U. Bilow, P. Lindenau und B. Schorn, *Teilchenphysik - Ladungen, Wechselwirkungen und Teilchen*, Joachim Herz Stiftung, 2018, URL: <https://www.teilchenwelt.de/material/materialien-fuer-lehrkraefte/unterrichtsmaterial-teilchenphysik/> (siehe S. 7–10).
- [3] P.A. Zyla u. a., *Review of Particle Physics*, PTEP **2020** (2020) 083C01 (siehe S. 11, 23, 24, 31, 36, 40, 46).
- [4] CERN Collaboration, *Linear accelerator 2*, URL: <https://home.cern/science/accelerators/linear-accelerator-2>, accessed: 01.05.2022 (siehe S. 12).
- [5] M. Kobel, U. Bilow, P. Lindenau, G. Bachmann und B. Schorn, *Teilchenphysik - Forschungsmethoden*, Joachim Herz Stiftung, 2021, URL: <https://www.teilchenwelt.de/materialien/vierbaendiges-unterrichtsmaterial/> (siehe S. 12, 13).
- [6] CERN Collaboration, *CERN's accelerator complex*, URL: <https://home.cern/science/accelerators/accelerator-complex>, accessed: 01.05.2022 (siehe S. 12).
- [7] Roland Waldi, *Statistische Datenanalyse - Grundlagen und Methoden für Physiker*, Springer Spektrum, 2015, ISBN: 9783662471449 (siehe S. 14–19, 32, 34).
- [8] Glen Cowan, *Statistical data analysis [with applications from particle physics]*, Repr., Clarendon Press, 2004, ISBN: 9780198501565 (siehe S. 15, 18, 19).
- [9] Uwe Lämmel und Jürgen Cleve, *Lehr- und Übungsbuch Künstliche Intelligenz, 2., korr. Aufl.*, Fachbuchverl. Leipzig im Carl-Hanser-Verl., 2004, ISBN: 9783446225749 (siehe S. 20–22).
- [10] Detlef Nauck, Frank Klawonn und Rudolf Kruse, *Neuronale Netze und Fuzzy-Systeme, 2., überarb. und erw. Aufl.*, Vieweg, 1996, ISBN: 9783528152659 (siehe S. 22).
- [11] CERN Collaboration, *Review of the 13 TeV ATLAS Open Data release*, Techn. Ber., All figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-OREACH-PUB-2020-001>: CERN, 2020, URL: <http://cds.cern.ch/record/2707171> (siehe S. 23).
- [12] CERN Collaboration, *SM Higgs Branching Ratios and Total Decay Widths (update in CERN Report4 2016)*, URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR>, accessed: 19.05.2022 (siehe S. 37).
- [13] CERN Collaboration, *12 steps - From idea to discovery*, URL: <https://home.cern/science/physics/12-steps-idea-discovery>, accessed: 19.05.2022 (siehe S. 42).
- [14] CERN Collaboration, *Overview of the MC samples released in the 13 TeV ATLAS Open Data*, URL: <http://opendata.atlas.cern/release/2020/documentation/datasets/mc.html>, accessed: 28.10.2022 (siehe S. 53).

Anhang

Auswertungsstatistiken erster Optimierungsschritt

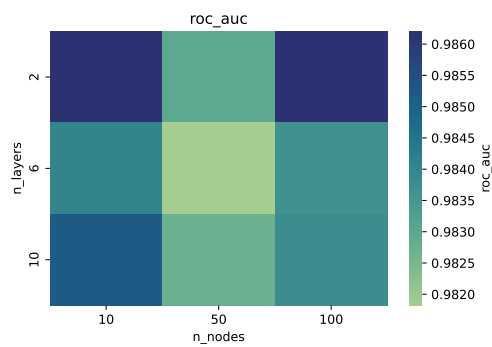


Abbildung 9.1: Heatmap zum ROC-AUC-Wert

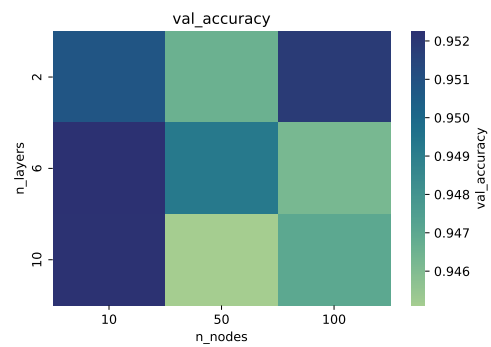


Abbildung 9.2: Heatmap zur Validations-Accuracy

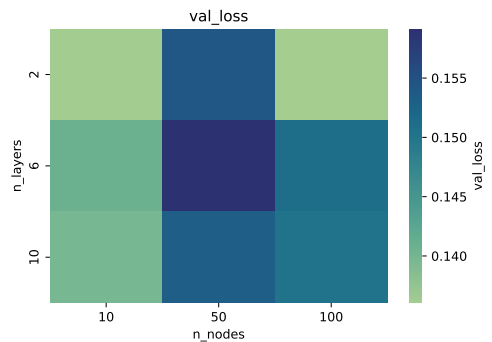


Abbildung 9.3: Heatmap zum Validation-Loss

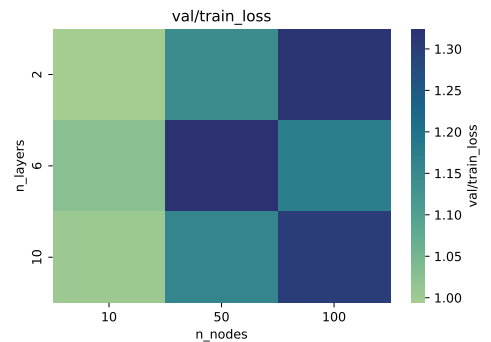
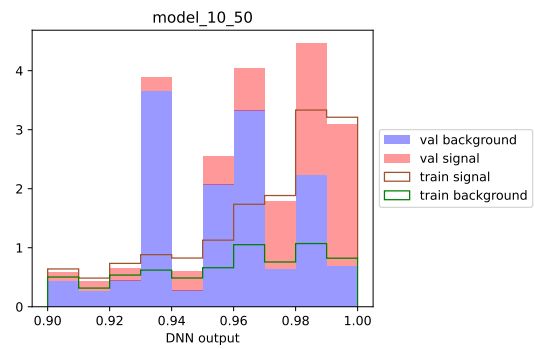
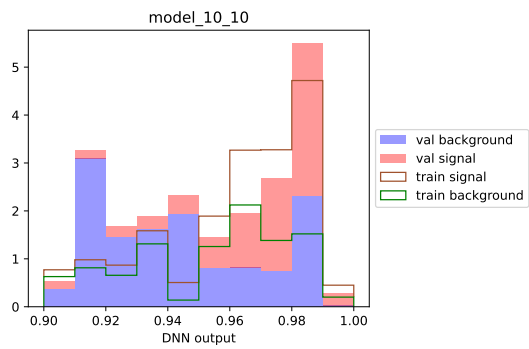
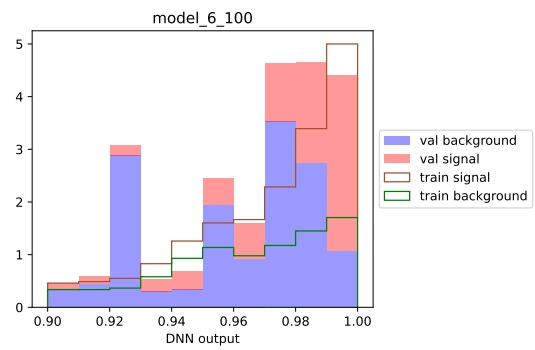
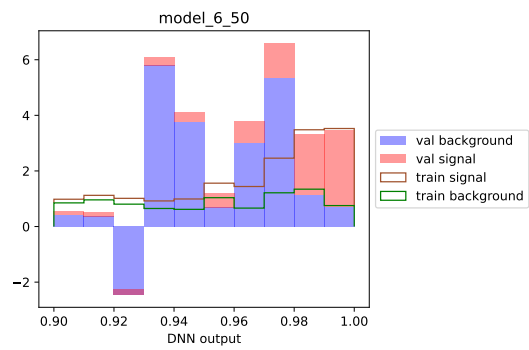
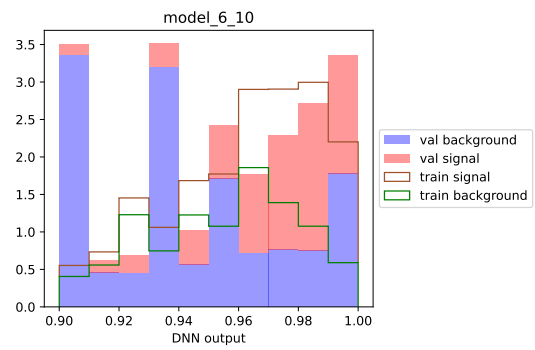
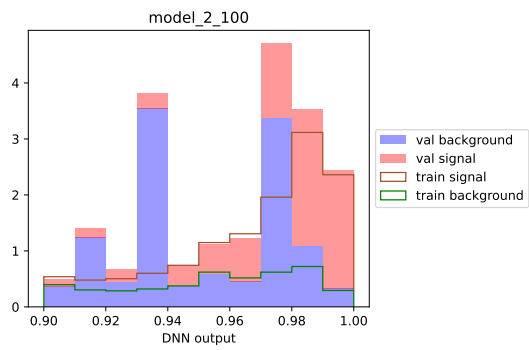
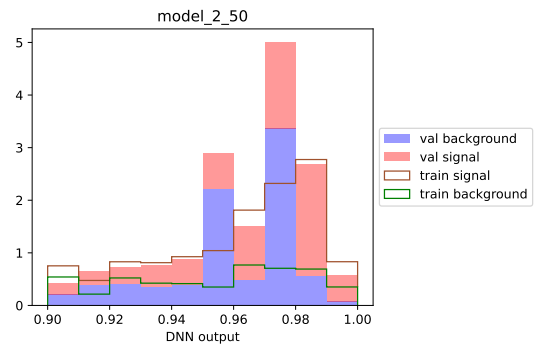
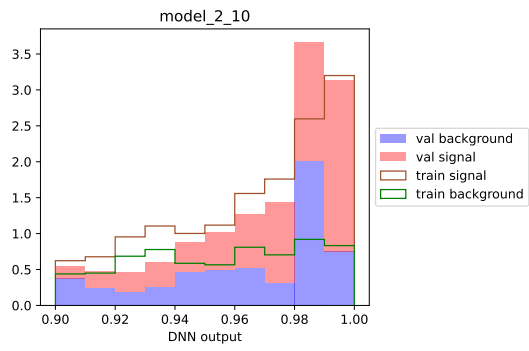
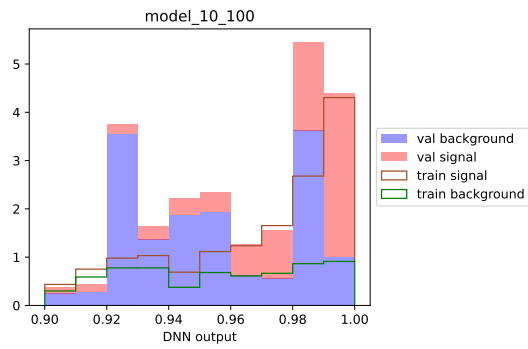


Abbildung 9.4: Heatmap zum Validations-/Training-Loss





Auswertungsstatistiken zweiter Optimierungsschritt

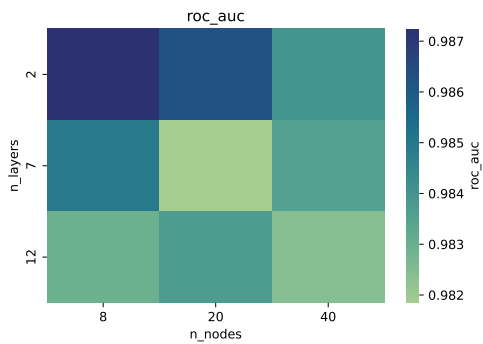


Abbildung 9.5: Heatmap zum ROC-AUC-Wert



Abbildung 9.6: Heatmap zur Validations-Accuracy

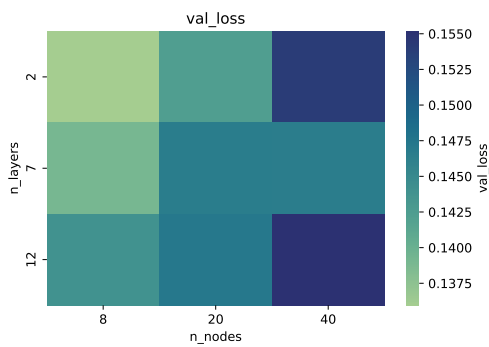


Abbildung 9.7: Heatmap zum Validation-Loss

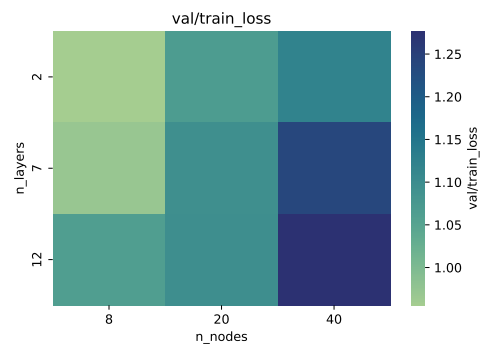
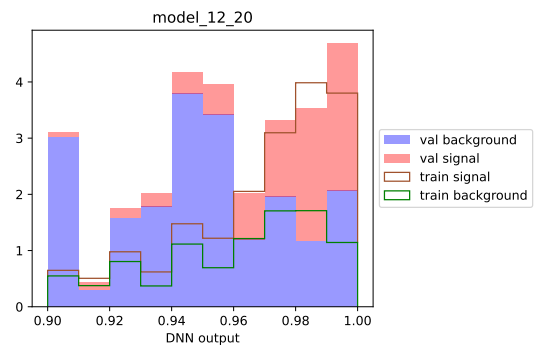
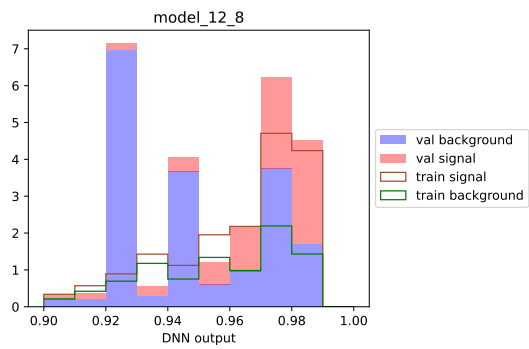
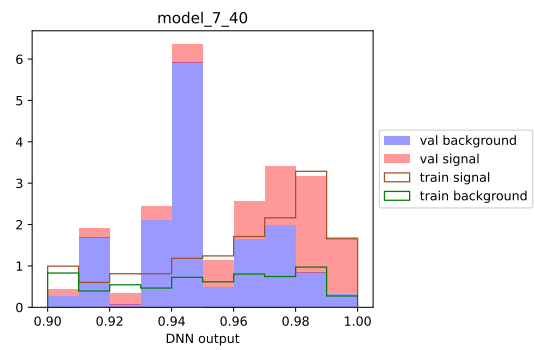
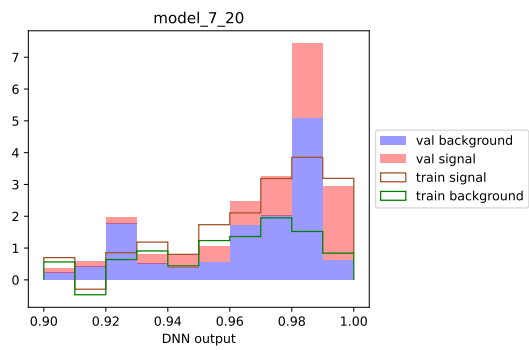
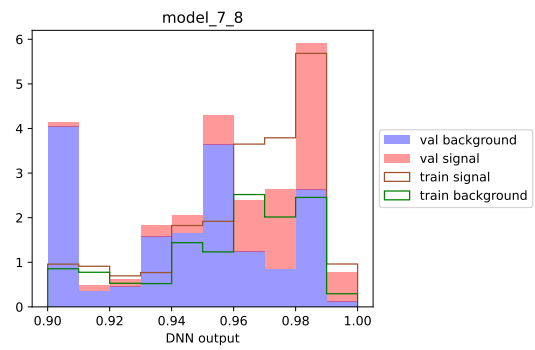
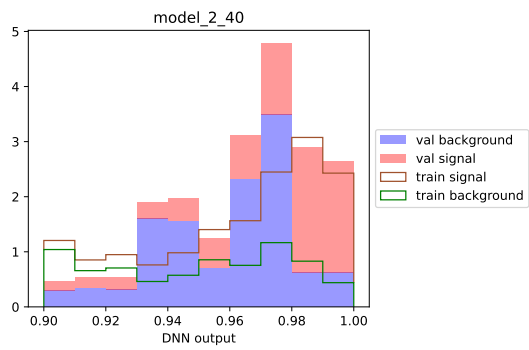
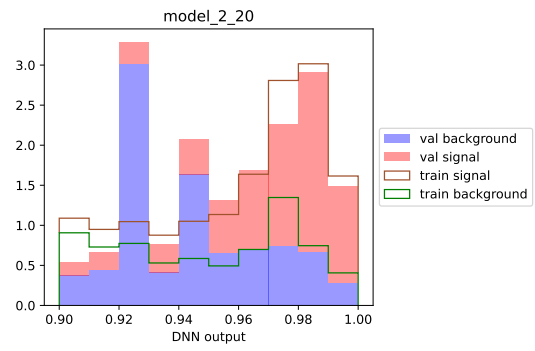
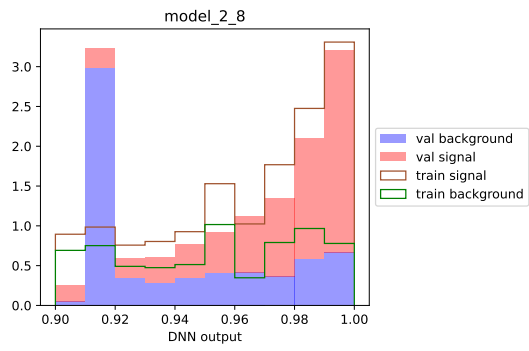
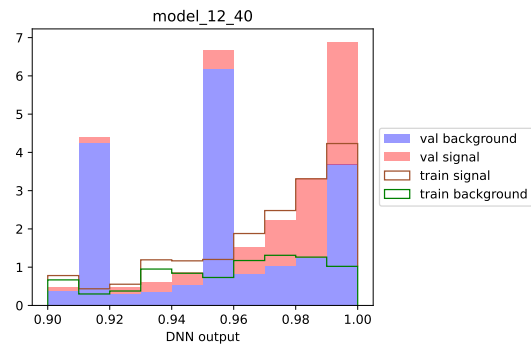


Abbildung 9.8: Heatmap zum Validations-/Training-Loss





Auswertungsstatistiken dritter Optimierungsschritt

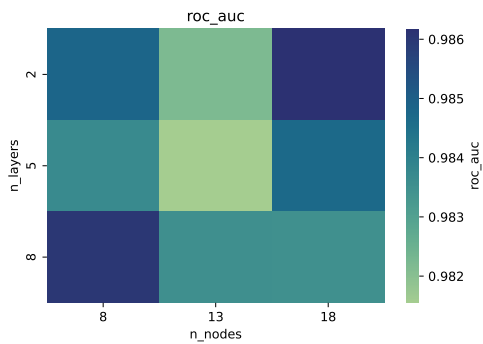


Abbildung 9.9: Heatmap zum ROC-AUC-Wert

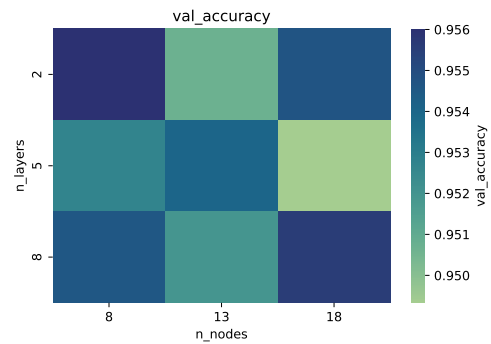


Abbildung 9.10: Heatmap zur Validations-Accuracy

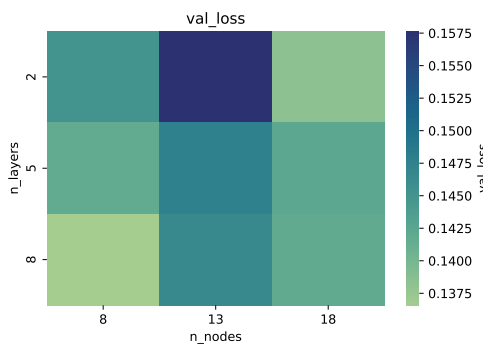


Abbildung 9.11: Heatmap zum Validation-Loss

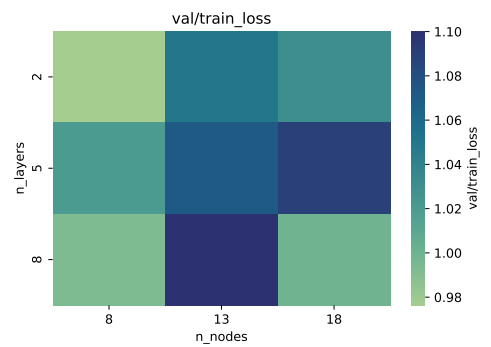
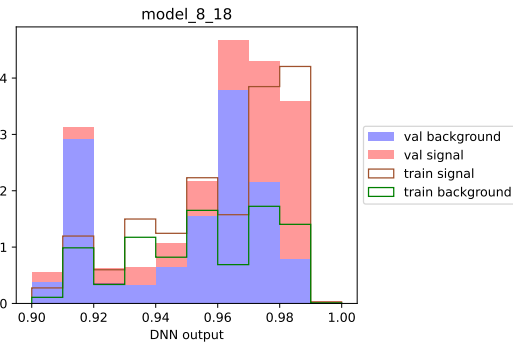


Abbildung 9.12: Heatmap zum Validations-/Training-Loss



Training, Validation und Test des finalen Modells

Epochen	Trainings-Loss	Validations-Loss	Test-Loss	$\frac{Valid.-Loss}{Train.-Loss}$	$\frac{Test-Loss}{Train.-Loss}$	ROC-AUC
22	0,139	0,143	0,138	1,031	0,998	0,984

Tabelle 9.1: Werte für das Training des finalen Neuronalen Netzes (8 Layer mit je 8 Neuronen)

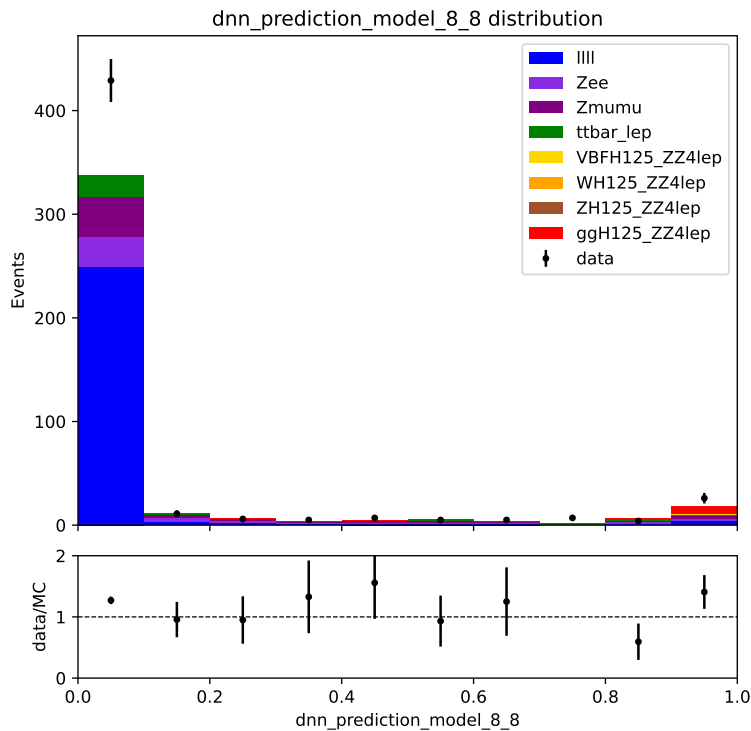


Abbildung 9.13: komplettes DNN-Vorhersage-Histogramm für alle vorhandenen Daten (Simulations- und Messdaten)

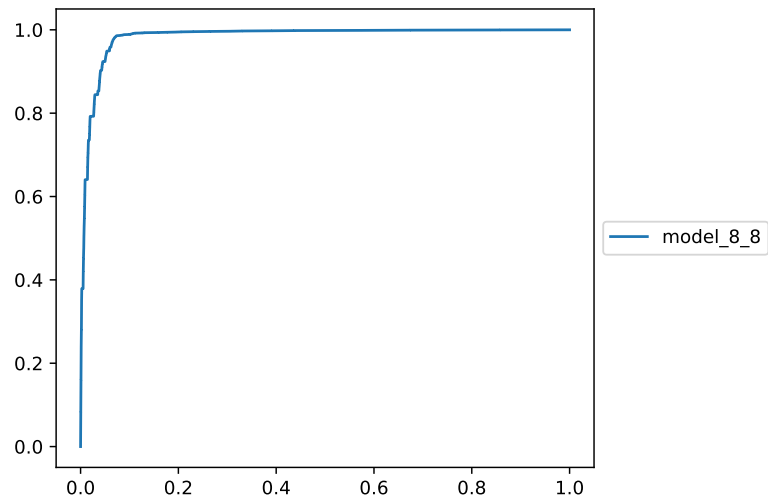


Abbildung 9.14: ROC-Kurve

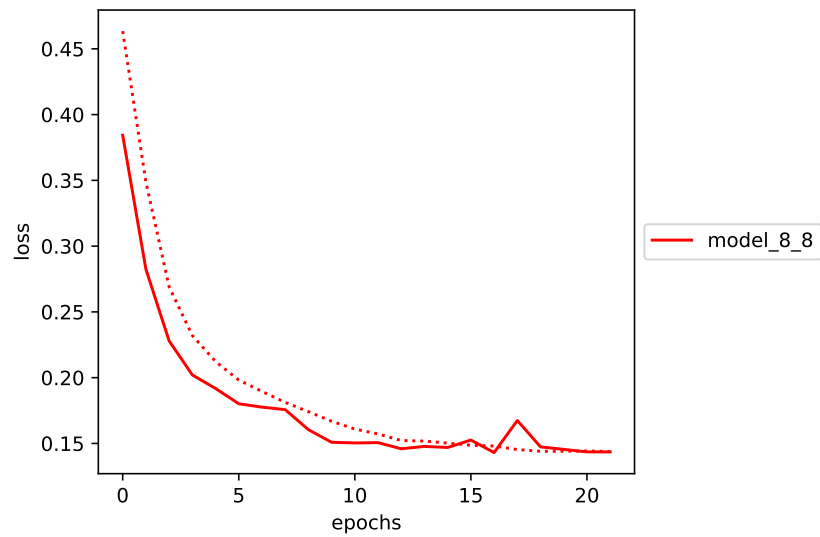


Abbildung 9.15: Loss-Kurven

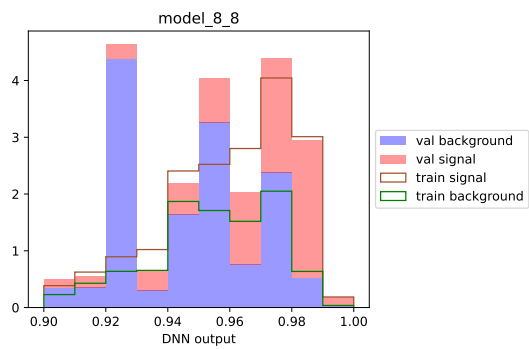


Abbildung 9.16: Trainings-Validations-Histogramm

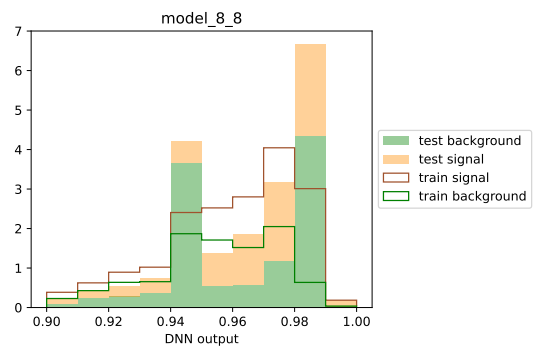


Abbildung 9.17: Trainings-Test-Histogramm

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am *Institut für Kern- und Teilchenphysik* ohne unzulässige Hilfe Dritter verfasst und alle Quellen und Zitate als solche gekennzeichnet habe.

Erik Thyzel
Dresden, 21.12.2022