Pseudo Experiment Based Studies with RooFit of Fits to Events Weighted by the SPlot Technique for Application in the ATLAS B-Physics Data Analysis

Master Thesis
to obtain the academic degree
Master of Science
(M. Sc.)

University of Siegen

Department of Physics

submitted by
B. Sc. Tim-Philip Hücking

December 2020

Overview of the chapters

Contents

1 Preface

Our present knowledge about the elementary constituents of matter and their interactions is described by the Standard Model (SM)[§] of particle physics (Ch. 2). In particle physics, nearly all the measurements are compatible with the SM. Hence, it is a very successful theory. Nevertheless, it has some limitations and therefore many searches for physics beyond the SM are ongoing.

The decays of the B_s^0 and B^0 mesons into two muons are well suited to detect physics beyond the SM. Among others, members of the ATLAS collaboration are measuring the branching ratios for these decays (Ch. 5) with their detector (Ch. 4) at the LHC (Ch. 3). Since these decays are rare, a multivariate classifier is employed in this analysis to suppress the background events. The classifier is partially trained on Monte Carlo (MC) samples. The MC samples are known to have some insufficiencies, e.g. due to a slightly different kinematic regime in the simulation. In order to determine corrections, the simulated signal of a reference channel is compared to the signal in the data of the reference channel. The signal data distribution needs to be extracted from data, which also contains background events. Up to now the sideband subtraction technique is used for this purpose [Lyo86]. This is the technique commonly used in High Energy Physics (HEP) for the purpose of statistically separating signal and background events. It requires a signal region to be defined, as well as sideband regions, where the number of background events dominates the number of signal events. The cuts defined to separate the regions become a source of systematic uncertainty which requires additional studies.

The purpose of this thesis is to explore the sPlot [PD05] technique and unbinned maximum likelihood fits to events weighted by this method. In contrast to the sideband subtraction technique, using this technique does not require the definition of signal and sideband regions. Instead, weights, called sWeights, serve to separate signal and background distributions. Based on a fit to a discriminating variable distribution, a signal and a background sWeight are calculated for each event. Applied to a control variable these weights recover the corresponding signal or background distribution of this variable. No knowledge about the shapes of the signal and background Probability Density Functions (PDFs) of this variable is required. Taking the advantage mentioned above into account, the ATLAS $B_{(s)}^0 \to \mu^+\mu^-$ analysis team considers replacing the sideband subtraction technique in their analysis by the sPlot technique.

[§]A list of all acronyms used can be found in Appendix C.

In many cases it is needed to estimate the parameters of the PDFs, that describe the disentangled signal and background distributions. Such a case is the measurement of the B_s^0-meson lifetime. This lifetime may be estimated by an unbinned maximum likelihood fit to the events weighted by the signal sWeights. Unbinned maximum likelihood fits and all other statistical methods relevant for this thesis are introduced in Chapter 6. However, there is a caveat about unbinned maximum likelihood fits in the presence of sWeights: the uncertainties of the fitted parameters obtained by the standard procedures, as e.g. used by MINUIT [JR75], must be corrected. Three methods to perform this uncertainty correction are studied in this thesis (Ch. 7). The first uncertainty correction method proposed by Eadie et al. [EDJ71] is referred to as the squared weights correction. Two uncertainty correction methods, referred to as asymptotic correction part one and asymptotic correction part two, were proposed recently by Langenbruch [Lan19]. The studies of these methods are based on pseudo experiments. Performing the pseudo experiments, the RooFit [VK03] software toolkit is used as much as possible. The latter is well tested, as it is widely used in HEP. The findings are summarized in Chapter 8.

2 Introduction to elementary particle physics

The SM of particle physics is a combination of relativistic Quantum Field Theorys (QFTs) describing the elementary constituents of matter (Sec. 2.1) and their interactions (Sec. 2.2). Despite its great success, there are limitations of the SM. A probe for physics beyond the SM, which is relevant for this thesis, is presented in Section 2.4.

2.1 The elementary particles

Particles are categorized as fermions and bosons, based on their intrinsic property spin. Fermions have half-integer spin quantum numbers. Their distribution over energy states is described by the Fermi-Dirac statistics [Fer35; Dir26]. Therefore, two fermions can never have identical quantum numbers, which is known as the Pauli exclusion principle [Pau25]. Bosons have integer spin quantum numbers and are distributed over the energy states according to the Bose-Einstein statistics [Bos24; Ein05]. According to the SM, there exist 12 elementary fermions in our universe, grouped in six quarks and six leptons as presented in Figure 2.1. Three of the six quarks have an electric charge of $+2/3\,e$, referred to as up type quarks and three an electric charge of $-1/3\,e$, referred to as down type quarks. The six different types of quarks are denoted as quark flavors. Each of the six quark flavors comes in three different types, which differ in their color charges. Leptons include the electron, the muon and the tau lepton, as well as their neutrino counterparts. In the context of the SM, neutrinos are massless.

In Figure 2.1 the 12 fermions are presented in three columns, referred to as generations. With each generation, the mass of the elementary particles increases. The second and third generation elementary particles are not stable and eventually decay into elementary particles of the first generation. All ordinary matter in our universe consists of the first generation elementary particles only. For each of the fermions listed, an antimatter complement exists. It has the same mass but the charge-like properties have opposite sign. Charge-like properties are e.g. the electric charge and the color charge. The bosons shown on the right-hand side of Figure 2.1 are discussed in the next section.

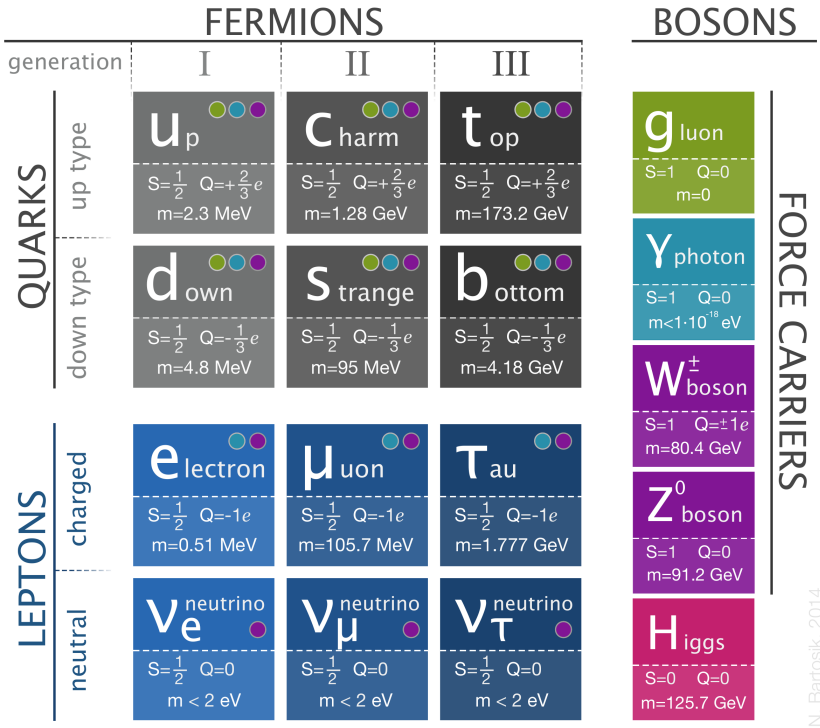


Figure 2.1: Elementary particles of the Standard Model, where S refers to the spin quantum number, Q to the electric charge and m to the mass. The interactions, in which the particles participate, are indicated for each particle by colored circles [Bar15]. Numeric values are taken from the PDG 2014 Particle Listing [Oli+14].

2.2 The fundamental interactions

There are four fundamental interactions, also called forces, for the 12 fermions, namely the electromagnetic, the weak, the strong and the gravitational interaction. The description of the latter is not part of the SM. Compared to the other forces acting on the fermions, the influence of gravity is negligibly small. For high interaction energies, a unified description of the electromagnetic and the weak interaction exists. All interactions between the fermions are realized by the exchange of the mediating bosons, given on the right-hand side of Figure 2.1. Each of the interactions relates to a certain property of the interacting particles, as e.g. the electromagnetic force relates to the electric charge.

Mathematically, the SM is usually formulated in terms of a Lagrangian density, in which the particles are represented as fields. The terms of the free Lagrangian density, describing non-interacting particles, are divided in two categories. The kinetic terms comprise the derivatives of the fields and the mass terms comprise quadratic expressions of the fields. The interaction terms are introduced into the free Lagrangian density by imposing a $SU(3) \times SU(2) \times U(1)_Y$ gauge symmetry on the free Lagrangian density, where Y stands for the hypercharge. To keep the Lagrangian density invariant under these symmetry transformations, additional fields are

introduced. After quantization, they yield the mediating bosons, also known as gauge bosons. The number of bosons mediating an interaction is equal to the number of generators of the corresponding symmetry group.

SU(3) is the symmetry group of Quantum Chromodynamics (QCD). This group has eight generators, corresponding to eight gauge bosons, called gluons. The gluons carry a color charge as well as an anti-color charge and couple to the color charge. Hence, only quarks and gluons participate in the strong interaction, as indicated by the circles colored like the gluons in Figure 2.1. A special property of QCD is color confinement: only particles in a color singlet state are observed. Thus, only bound states of quarks are observed, rather than bare quarks. These composites are referred to as hadrons. They are classified as mesons and baryons, depending on whether they are formed by two or three quarks. Examples for mesons are the B^0 and B_s^0 mesons, which have the quark content $\bar{b}d$ and $\bar{b}s$, respectively. The proton and the neutron are examples for baryons with the quark content uud and udd, respectively.

The electroweak interaction has the symmetry $\mathrm{SU(2)}\times\mathrm{U(1)_Y}$. It is the unified description of the electromagnetic and the weak interaction [Gla59; SW59; Wei67]. The two forces are different manifestations of the unified force if the interaction energy is below the electroweak scale, which is around 246 GeV. Above this energy there are four massless gauge bosons. Below this energy the $\mathrm{SU(2)}\times\mathrm{U(1)_Y}$ symmetry is spontaneously broken to a $\mathrm{U(1)_{EM}}$ symmetry due to the vacuum expectation value of the Higgs field. This results in three mass terms, which, after a transformation, can be identified as the mass terms of the W^+, W^- and the Z bosons. These are the gauge bosons of the weak interaction. This phenomenon is known as Higgs mechanism [Hig64; EB64]. In case of the broken electroweak symmetry, the elementary fermions acquire a mass term proportional to the vacuum expectation value of the Higgs field, by coupling to this field. The Higgs boson, presented at the bottom right of Figure 2.1, is the boson associated with the Higgs field. It was discovered by the ATLAS and CMS collaborations in 2012 [ATL12a; CMS12] and is the last fundamental elementary particle predicted by the SM, which was discovered.

The mediating bosons of the weak interaction, the $W^\pm$ and the Z^0 bosons, have a mass greater than 80 GeV. Therefore, the reach of the weak interaction is of subatomic scale. The weak interaction couples to a property called weak isospin. All 12 fermions can interact via the weak force, indicated by the circles colored like the Z^0 in Figure 2.1. Only by the weak interaction, more precisely by the $W^\pm$ bosons, the quark flavor can change. The amplitudes for such charged weak transitions across the quark generations are proportional to the elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [Cab63; KM73]. The complex phase of the CKM matrix allows for the violation of the CP symmetry in weak interactions. The weak interaction is e.g. responsible for the β-decay, which is the conversion of a neutron into a proton by the transition $d \to u\,W^-$.

The $U(1)_{EM}$ symmetry group is the symmetry of Quantum Electrodynamics (QED), which is the theory describing the electromagnetic interaction. This group has only one generator, which relates to the massless photon. Since the photon is massless and neutral, the electromagnetic force has an infinite reach. It couples to the electric charge of particles. Except for the neutrinos, which have zero electric charge, all fundamental fermions interact via the electromagnetic force, as indicated by the circles colored like the γ in Figure 2.1. It is this force, that permits the existence of atoms, by binding the electron cloud to the nucleus.

2.3 Limitations of the Standard Model

The SM is very successful in predicting and explaining measurement results of particle physics experiments. But there are some puzzles. One of them is the neutrino mass, which is zero in the context of the SM. However, neutrino oscillations [SNO02; Sup98] prove, that each of the neutrino flavors has a different, yet still unknown, mass.

Furthermore, the SM does not appear to explain all matter found in the universe. Astronomical observations suggest that there is additional matter in the universe, which interacts only weakly or even not at all with ordinary matter. Thus it is called dark matter. Such observations are e.g. the velocity of galaxy arms, which is much higher than Keplers second law predicts [CS00], or the differences observed when determining the mass of galaxies by gravitational lensing and X-ray data [Nat+17]. According to present estimates 84 % of the matter in the universe is dark matter [Pla14].

For these reasons – among others –, there are many efforts to develop theories beyond the SM, to check them experimentally, and to discover new phenomena, which are not yet described by a theoretical model. One model predicting additional particles is the Minimal Supersymmetric Standard Model (MSSM) [BT06]. A probe for physics beyond the SM, as e.g. described by the MSSM, is presented in the subsequent section.

2.4 The rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$

A system well suited to probe models such as the MSSM, are the B-mesons, e.g. the decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. These decays belong to a category of processes called Flavor Changing Neutral Currents (FCNCs). The FCNCs are processes that change the flavor content of particles but leave their electric charge unchanged. In the SM, FCNCs are forbidden at tree level. For example, a transition from a b-quark to a s-quark can not happen via exchange of a Z-boson. However, such transitions are possible at one-loop level (Figs. 2.2a and 2.2c).

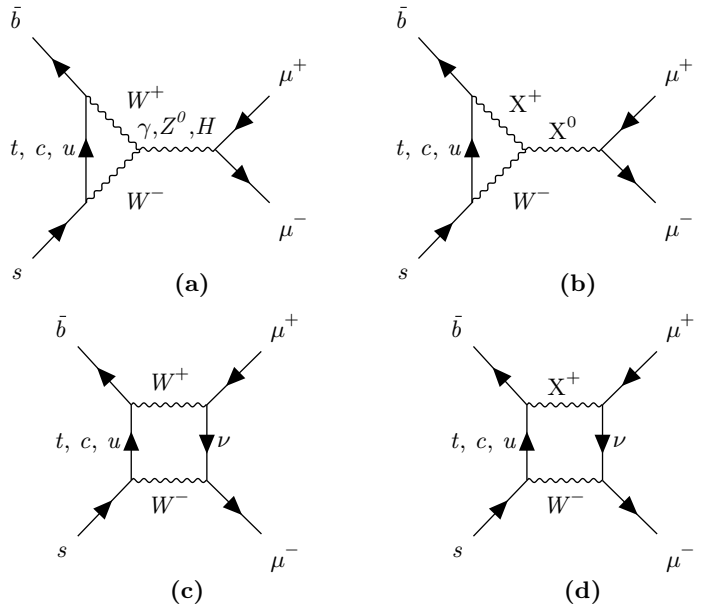


Figure 2.2: (a) and (c) some of the possible $B_s^0 \to \mu^+\mu^-$ Feynman diagrams in the SM. (b) and (d) examples of $B_s^0 \to \mu^+\mu^-$ Feynman diagrams with contributions from generic, non-SM particles, labeled X^0 and X^+.

In such loops also non-SM particles can contribute (Figs. 2.2b and 2.2d), even if they are too heavy to be produced directly. This could lead to an enhanced or suppressed branching fraction for the final state, depending on whether the interference of the different diagrams is constructive or destructive.

The decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are strongly suppressed in the SM, i.e. they are suppressed by the Glashow-Iliopoulos-Maiani (GIM) mechanism [GIM70] and by requirements on the helicity of the final states: the $B_{(s)}^0$ are spin-less and therefore the muons are required to have the same helicity.

The branching ratios of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are measured by the ATLAS collaboration. This analysis is presented in Chapter 5.

3 The Large Hadron Collider

In this chapter the specifications (Sec. 3.1) and the operation in past and future of the Large Hadron Collider (LHC) (Sec. 3.2) are presented.

3.1 Specifications

With a circumference of 26 660 m, the LHC is the largest circular hadron collider built so far [Brü+04]. It is located in the former Large Electron-Positron Collider (LEP) tunnel, 50 m to 175 m beneath the surface. The tunnel crosses the border between Switzerland and France, close to Geneva. The LHC belongs to the European Laboratory for Particle Physics (CERN).

The LHC is used to accelerate and collide mainly protons but also heavy ions. It is not capable of accelerating a proton from rest up to the final energy of 7 TeV [Brü+04]. Hence, the protons are passed through an accelerator chain, prior to injection into the LHC. At CERN this chain is composed of the Linear Accelerator (LINAC), the PS-BOOSTER, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) (Fig. 3.1). The protons are accelerated in bunches, containing up to 1.15×10^{11} protons [Brü+04]. At four interaction points the two antiparallel beams cross and the proton bunches collide. The collisions happen with a frequency of 40 MHz. The four main experiments, located at the four interaction points are:

- ALICE (A Large Ion Collider Experiment) [ALI08], focusing on heavy ion collisions to study e.g. the quark-gluon plasma,
- ATLAS (A Torodial LHC Apparatus) [ATL08], a general purpose experiment (Ch. 4),
- CMS (Compact Muon Solenoid) [CMS08], a general purpose experiment as well and
- LHCb (Large Hadron Collider beauty) [LHC08], designed to study B-mesons and CP violation.

An important property of an accelerator is its instantaneous luminosity. For two opposite beams it is defined by

$$L = f \; n \; \frac{N^2}{A} \; , \tag{3.1}$$

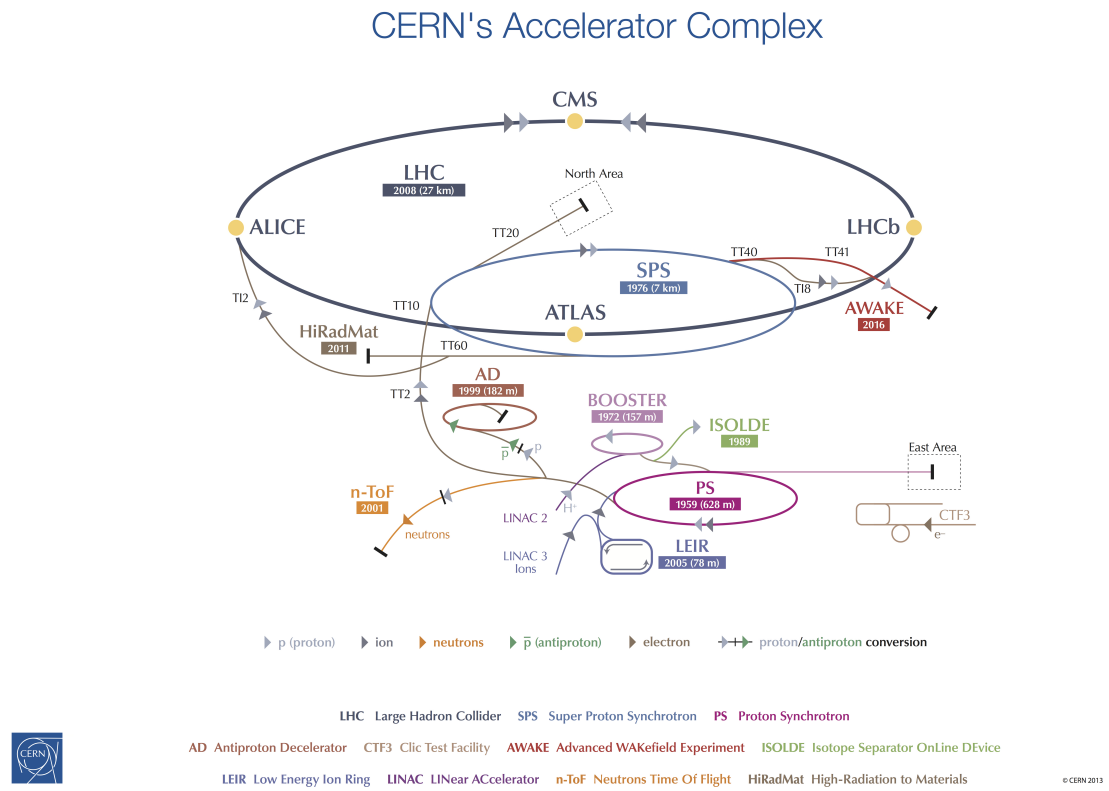


Figure 3.1: Sketch of the CERN accelerator complex, including the LHC and the injecting accelerators. The position of the four major experiments at the four interaction points is indicated as well. [Haf13]

where f is the revolution rate, n the number of bunches, A the cross-section area of the bunches and N the number of particles per bunch. The LHC was designed to deliver an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The luminosity integrated over time is denoted by $\mathcal{L}$. For a given cross section σ_b of a physics process (e.g. $B_s^0 \to \mu^+\mu^-$), the number of events N_b, that are produced on average, is calculated by

$$N_b = \sigma_b \, \mathcal{L} \, . \tag{3.2}$$

Whether a rare process, such as $B^0 \to \mu^+\mu^-$, can be discovered or not depends on the integrated luminosity as long as the energy threshold for the process of interest is reached.

3.2 Overview of LHC operation and future plans

The LHC started operation in 2009 [CER09]. The first period of operation, which ended in 2013, is referred to as Run-I. In Run-I, 75 % of the design instantaneous luminosity was reached (Fig. 3.2). During Run-II even twice the design instantaneous luminosity was reached, which

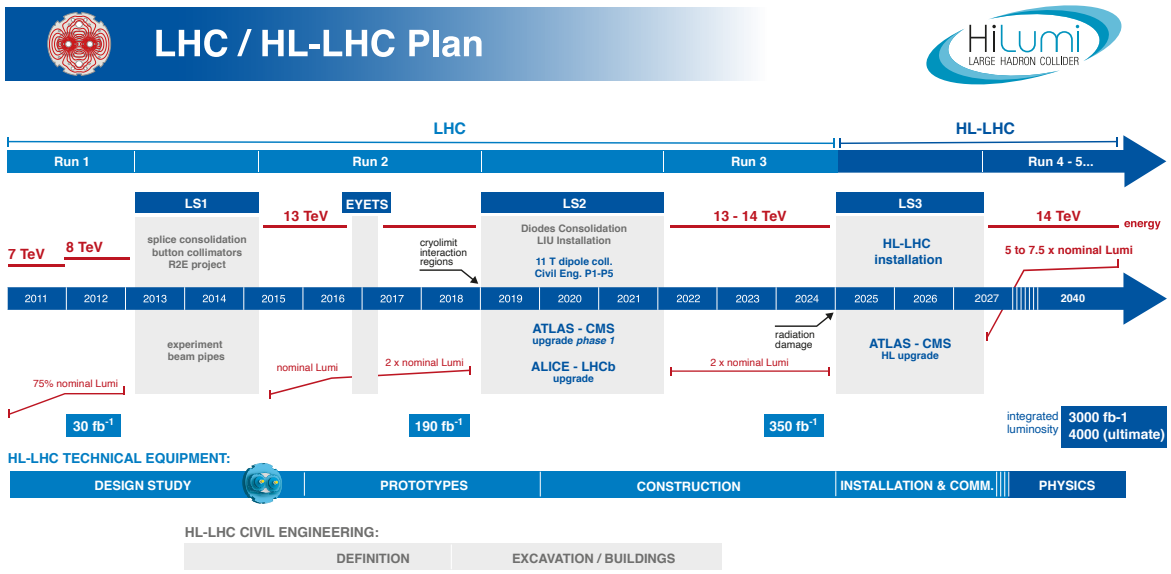


Figure 3.2: Overview of the operation of the LHC from 2009 until 2040 as of August 2020 [HiL20].

led to a total integrated luminosity of $156\,\mathrm{fb}^{-1}$ at the end of Run-II [ATL19b]. The periods of operation are interrupted by long shutdowns (Fig. 3.2), used for upgrades of the LHC and the detectors.

A major upgrade of the LHC is planned for the Long Shutdown 3 (LS3), scheduled for 2025 to 2027 (Fig. 3.2). After that, the LHC will be referred to as High Luminosity LHC (HL-LHC) [G+17]. The HL-LHC is designed to deliver an instantaneous luminosity of $5 \times 10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{2}$, which is five times the design instantaneous luminosity of the LHC. This poses a great technical challenge, e.g. w.r.t. beam stability, beam-beam interaction effects and beam-induced heat load on the cryogenic system [Ard+16]. The design instantaneous luminosity of the HL-LHC corresponds to an integrated luminosity of $3000\,\mathrm{fb}^{-1}$ in 12 years of operation [G+17]. Without the upgrade to the HL-LHC, it would not be beneficial to operate the LHC beyond Run-III due to the slower increase of the discovery potential. It was estimated, that after 2020, the running time required to reduce the statistical uncertainty of certain measurements by a factor two will grow to more than ten years, if the LHC would not be upgraded [G+17].

4 The ATLAS experiment

The ATLAS (A Torodial LHC Apparatus) collaboration, founded in summer 1992, has at the moment about 5500 members from nearly 100 nations, with about 2900 being signing authors. The members of this collaboration developed and built the ATLAS detector (Fig. 4.1) at the LHC (Fig. 3.1). Since 2009 they operate, maintain and upgrade this detector. The ATLAS detector is a general purpose detector to address open questions in particle physics. With a length of 44 m and a diameter of 25 m, it is the largest detector at the LHC. It is composed of several sub-detectors. The Inner Detector (Sec. 4.1), the calorimeters and the Muon Spectrometer (Sec. 4.2) are the main groups of sub-detectors. The calorimeters, namely the Liquid Argon Calorimeter and the Hadronic Tile Calorimeter, are less relevant for the $B^0_{(s)} \to \mu^+\mu^-$ analysis (Ch. 5) and are therefore not presented in the subsequent sections.

The rate of collision events in the ATLAS detector is very high (40 MHz). For reasons of limited computing resources, disk space and read-out bandwidth, only events, that pass the trigger selection (Sec. 4.3), are processed (Sec. 4.4). Future upgrades of the ATLAS detector to cope with the increased luminosities in the HL-LHC era are presented in Section 4.5.

The right-handed ATLAS coordinate system has its origin at the nominal interaction point at the center of the detector. The z-axis is parallel to the beam axis and the x-axis points to the center of the LHC. The angle around the beam axis, in the x-y plane, is denoted by ϕ and the polar angle to the beam axis is denoted by θ. The angle θ is often expressed in terms of the pseudorapidity $\eta = \ln(\tan \theta/2)$.

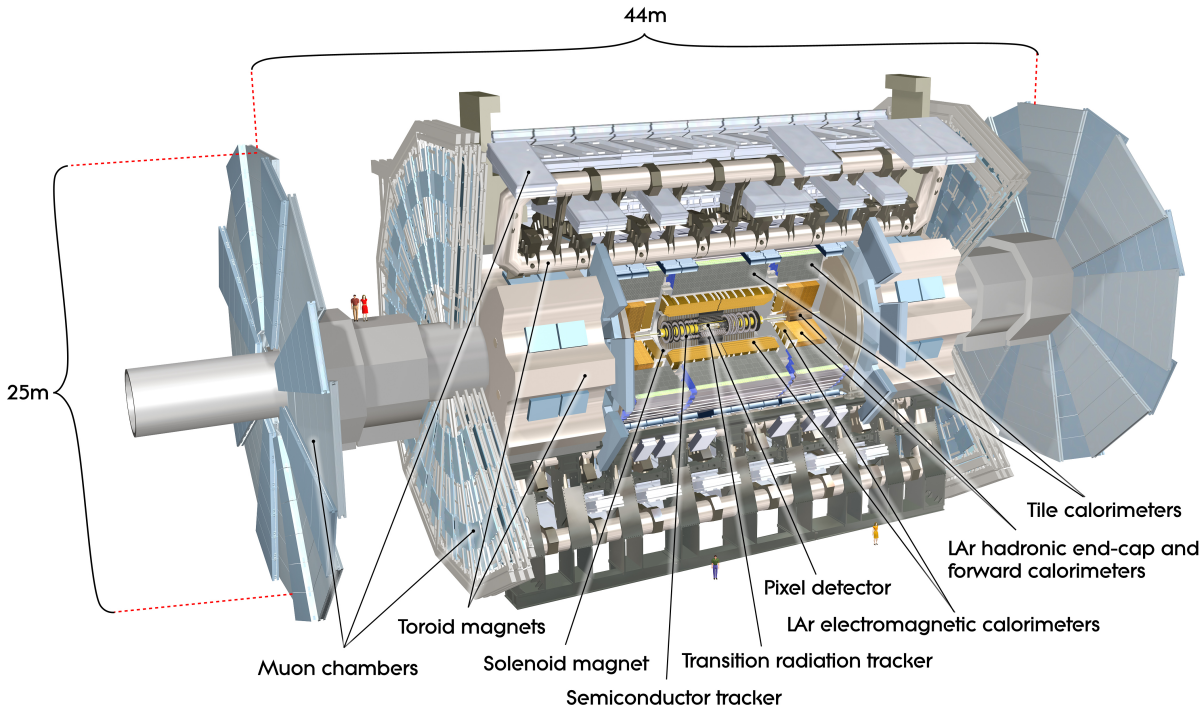


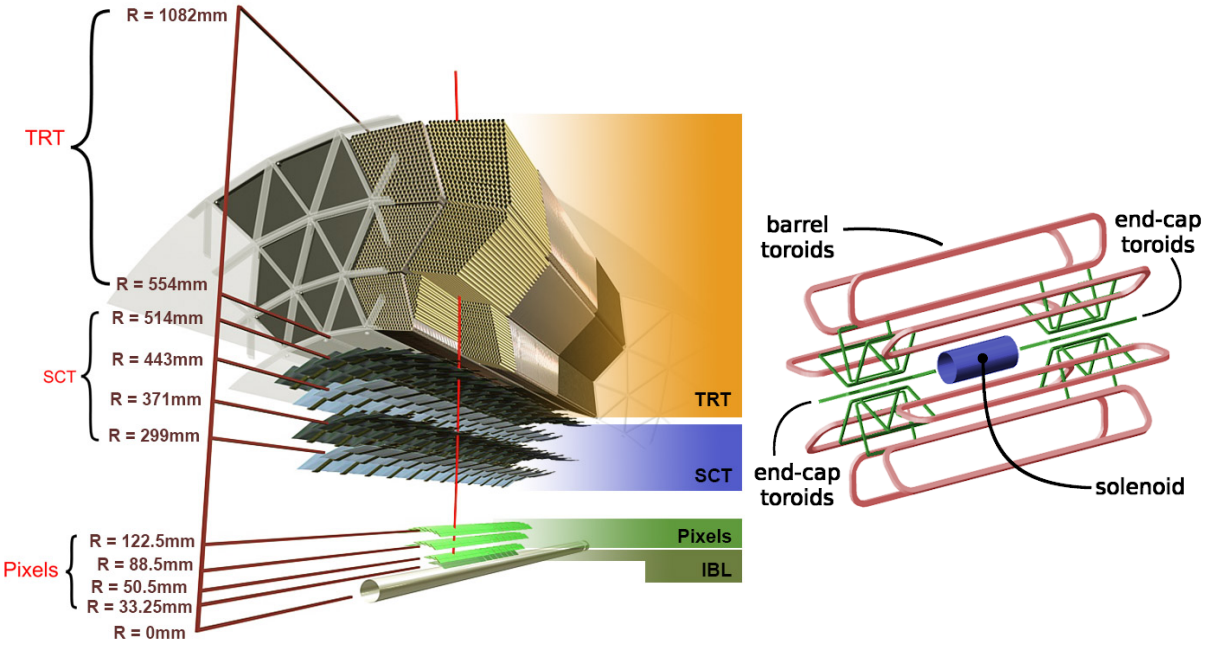
Figure 4.1: Cut-away view of the ATLAS detector system with the individual sub-detectors labeled [Peq08].

4.1 Inner detector

The Inner Detector (ID) is a group of three sub-detectors (Fig. 4.2a): the pixel detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). It is placed inside a 2 T solenoid magnet with a diameter of 2.4 m (Figure 4.2b). The solenoid magnet bends the trajectories of charged particles in the x-y plane, which are tracked by the ID. Based on the particle tracks, the momenta of the particles as well as the potential interaction vertices are determined.

The pixel detector [WH98] is the sub-detector closest to the beam pipe. It consists of four layers in the barrel region, including the Insertable B-Layer (IBL) [Cap+10], and three disks in the endcap regions. The IBL was added during Long Shutdown 1 (LS1) between the original pixel detector and a new beam pipe with reduced diameter. It improves the vertex determination and the identification of hadrons containing a b-quark. Compared to the original pixel detector, the IBL is realized with more advanced technology and has a pixel size reduced to 50 µm × 250 µm instead of 50 µm × 400 µm. The latter results in a spatial resolution of 10 µm × 115 µm for the original pixel detector. Overall there are roughly 92×10^6 read-out channels.

The SCT [ATL97a] covers the region with radii from 30 cm to 55 cm (Fig. 4.2a). It is a silicon detector like the pixel detector. However, to cover an area of 60 m^2, microstrips are employed instead of



(a) (b)

Figure 4.2: (a) Side view of the ID in the barrel region [ATL16b]. (b) Overview of the ATLAS detector magnet system [Pea16]. The ID is located inside the solenoid.

pixels. A typical strip sensor features a microstrip every 80 µm and has a size of 6 cm × 6 cm. Four such strip sensors form a SCT module. For this, two strip sensors are daisy-chained in each case. The sensor chains are mounted on each side of a thermally conducting thermal pyrolytic graphite baseboard. They are rotated against each other by an angle of 40 mrad. This angle allows for spatial resolution along the strips. There are 2112 SCT modules in the barrel region, distributed over four layers. In the endcap regions, 1976 SCT modules are distributed over 9 disks on each side.

The straws of the TRT [ATL97a] fill the outer region of the barrel from 55 cm to 108 cm radius. They are proportional drift tubes with a diameter of 4 mm and are filled with Xenon-rich gas. The space between the straws is filled with a polypropylene transition radiator material. The photons emitted by the radiation material are detected by the gas tubes. The intensity of the emitted photons is used to improve particle identification, especially the identification of electrons. Furthermore, the TRT contributes significantly to a precise momentum measurement.

4.2 Muon Spectrometer

The Muon Spectrometer (MS) [ATL97b] is the outermost group of sub-detectors. It is utilized for momentum measurements of highly energetic muons within a relative uncertainty of a few percent and a time measurement within a precision of a few ns. The momentum measurement

relies on the inhomogenous 4 T magnetic field of the barrel and endcap toroids (Fig. 4.2b), which bends the muon trajectories in the r-z plane.

The ability to trigger on muons is important for many analyses of physics processes, e.g. $H \to 4\ell$ or $B^0_{(s)} \to \mu^+\mu^-$. In order to use the MS in the trigger decision, its timing resolution must be much smaller than the time between two bunch crossings, which is 25 ns. Such a timing precision is achieved with Resistive Plate Chambers (RPCs) in the barrel region ($\eta < 1.05$) and Thin Gap Chamberss (TGCs) in the endcap regions ($1.05 < \eta < 2.7$).

In order to measure the momenta of muons, especially of highly energetic muons with momenta of 1 TeV or higher, with a relative uncertainty of a few percent, a spatial resolution of 0.05 mm must be achieved. Furthermore, the strength of the local magnetic field must be known with a relative uncertainty of at most 1 %. This is achieved by deploying Monitored Drift Tubess (MDTs). The MDTs are aluminum drift tubes with a diameter of 30 mm. They are monitored for deformations by an alignment system. The different layers of MDTs are also monitored by an optical alignment system. Along the MDTs, magnetic sensors measure the strength of the local magnetic field. With these measures the requirements on spatial resolution and magnetic field strength are achieved.

4.3 Trigger system

In many of the collision events happening at the LHC, no physics processes, that are of interest for physicists, occur. Therefore, a trigger system is designed to accept only events of potential interest. On the other hand, the event rate has to be lowered such, that the available computing resources, storage capacities and the read-out bandwidth are sufficient.

Since Run-II, the ATLAS detector has a two level trigger system, as presented in Figure 4.3. The first level (L1) is implemented in hardware. The trigger decision is based on ionisation patterns in the MS and energy deposits in the calorimeters. Out of the 4×10^7 events per second only about 10^5 are accepted. When an event was accepted, the corresponding Region of Interest (RoI) is passed to the High Level Trigger (HLT). The HLT is a software trigger system and reconstructs a partial event (Sec. 4.4) within the RoI. It reduces the event rate by a factor of 100. Given the mentioned constraints, only the resulting reduced amount of events is recorded and stored.

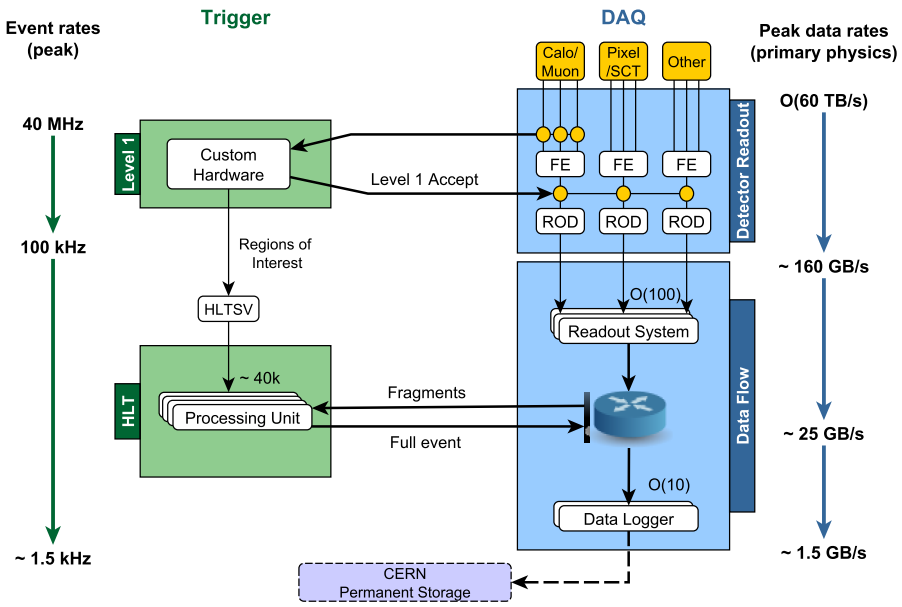


Figure 4.3: Overview of the ATLAS trigger and DAQ system, based on Ref. [Bor+15].

4.4 Event reconstruction, simulation and computing

The aim of the event reconstruction is to identify all the physics objects, such as e.g. muons, missing energy and jets, in the raw data. The offline event reconstruction is performed in two steps. A first reconstruction is performed for each part of the detector. For example, tracks and vertices are identified in the raw data of the ID. These results are fed into a combined reconstruction to identify the physics object. E.g. muons are identified by combining track segments from the MS and the ID. The full offline event reconstruction takes approximately 25 s per event. The online event reconstruction uses the same software as the offline event reconstruction. However, since the online event reconstruction uses faster algorithms, limits the event reconstruction to regions of interest and rejects events early, the time needed to process an event is reduced to 0.5 s.

The event reconstruction is not only applied to data, but also to MC samples. The MC samples are produced by simulating physics processes, passing them through a simulation of the ATLAS detector and finally reconstructing them with the same event reconstruction procedure used for data. MC samples are crucial to understand detector effects and e.g. to study the distributions of background events or to check how signals of yet unknown postulated decays or particles may show up in the detector.

As the data of the three other main experiments at the LHC, the ATLAS data is distributed to and processed in computing centers worldwide, which form the Worldwide LHC Computing Grid (WLCG). The WLCG has a hierarchical structure. The highest instances in this hierarchy

are the Tier-0 and Tier-1 computing centers, which store and process the raw data. The Tier-2 computing centers host the data for the analyzes and provide computing resources to the analyzers.

4.5 Detector upgrades

The upgrade of the LHC to the HL-LHC, as presented in Section 3.2, presents a major challenge for the ATLAS collaboration. At the ultimate instantaneous luminosity of the HL-LHC, a pile-up of about 200 is expected [ATL19a]. This means, that on average 200 proton-proton collisions happen at once for a single bunch crossing. In contrast, the current ATLAS detector was designed for a pile-up of 25 and operated at an average pile-up of about 37 during Run-II. Besides this, the detector has to be much more radiation hard and support higher read-out rates. Therefore, among other upgrades, most of the read-out electronics will be replaced, as well as the data acquisition and trigger system. Two examples of other upgrade projects, the New Small Wheel (NSW) (Sec. 4.5.1) and the Inner Tracker (ITk) (Sec. 4.5.2) are presented subsequently.

4.5.1 New Small Wheel

A replacement of the muon small wheels by the NSWs is ongoing during Long Shutdown 2 (LS2). The muon small wheels and the muon big wheels are used to track muons, that leave the detector in the endcap regions. They also provide the capability to trigger on such muons. Running at instantaneous luminosities higher than the design instantaneous luminosity results in neutrons emitted from activated material. These neutrons may be identified as muons in the muon big wheels. The muon small wheels could be used as a veto, by matching the tracks. However, the spatial resolution of the muon small wheels is not sufficient. This problem is solved by the NSWs.

4.5.2 Inner Tracker

The ID will be replaced by an all-silicon detector, the ITk. The inner part will be a 5 layer pixel detector, surrounded by strip sensors. The pixel size will be reduced to $50\,\mu m \times 50\,\mu m$. It is planned to cut the material budget used for the ITk down by half, compared to the present pixel detector. This reduces the energy losses of particles, when passing through the detector, and in turn improves the accuracy of the particles momentum and energy measurements.

5

Measurement of the $B^0_{(s)} \to \mu^+\mu^-$ branching fractions with the ATLAS experiment

In the present chapter the context of the studies in this thesis is provided, which is the measurement of the branching fractions of $B^0_{(s)} \to \mu^+\mu^-$ [§] based on ATLAS data taken in 2015 and 2016 [ATL18]. As pointed out in Section 2.4, these decays are an important probe for new physics.

The first section is about the measurement principle, followed by Section 5.2 on the data and the MC samples. The MC samples were introduced in Section 4.4. For data the trigger settings and the event selection are presented, whereas for the MC samples the reweighting is explained, which necessitates the introduction of the sideband subtraction technique. This technique could be replaced by the sPlot technique, studied in Chapter 7 of this thesis. A description of the background components in the $B^0_{(s)} \to \mu^+\mu^-$ analysis and the background rejection technique applied are provided in Section 5.3. The background rejection is accomplished using a multivariate classifier, namely a boosted decision tree (BDT). In Section 5.4 the determination of the quantities required to calculate the branching fractions for $B^0_{(s)} \to \mu^+\mu^-$ is discussed. In the last section the final results of the analysis and conclusions are presented. In addition, a combined result from ATLAS, CMS and LHCb is reported.

5.1 Measurement principle

The current predictions for the branching fractions from the SM theory are [BBS19]:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.66 \pm 0.14) \times 10^{-9} ,$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10} .$$

The relative uncertainties for the predictions are below 5 %.

The branching fractions are measured relative to the decay $B^+ \to J/\psi K^+$ ($J/\psi \to \mu^+\mu^-$), which is abundant compared to $B^0_{(s)} \to \mu^+\mu^-$ decays. It can be triggered using a similar dimuon

[§] Throughout this thesis $B^0_{(s)}$ is a shorthand notation for B^0 and B^0_s. Consequently, $B^0_{(s)} \to \mu^+\mu^-$ reads $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$.

trigger and has a well-known branching fraction. Performing a relative measurement leads to a cancellation of some uncertainties common to both channels, e.g. the uncertainty of the absolute luminosity. The absolute branching fractions are then calculated by multiplying the world average values for $\mathcal{B}(B^+ \to J/\psi K^+)$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ to the measured relative branching fractions:

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{N_{d(s)}}{\epsilon_{\mu^+\mu^-}} \times \left[\mathcal{B}\left(B^+ \to J/\psi K^+\right) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)\right] \frac{\epsilon_{J/\psi K^+}}{N_{J/\psi K^+}} \times \frac{f_u}{f_{d(s)}} \ , \quad (5.1)$$

where $N_{d(s)}$ denotes the extracted number of signal events, $N_{J/\psi K^+}$ the number of extracted reference channel events and f_d, f_s, f_u the probability, that a B meson hadronizes into a B^0, B^0_s or B^+. $\epsilon_{\mu^+\mu^-}$ and $\epsilon_{J/\psi K^+}$ are the values of acceptance times efficiency for the final states, accounting for losses due to the phase space fiducial volume chosen and for reconstruction effects.

5.2 Data and MC sample preprocessing

In the next section (5.2.1) the trigger settings and the event selection based on reconstruction conditions are described. This is followed by a short introduction to the sideband subtraction technique (5.2.2), which provides the motivation for the studies of the sPlot technique in this thesis. It is a prerequisite for the third section (5.2.3), in which the preparation of MC samples, that are consistent with data, is presented.

5.2.1 Trigger settings and event selection

Three different data sets are used in the analysis, containing the signal channels $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, the reference channel $B^+ \to J/\psi K^+$ ($J/\psi \to \mu^+\mu^-$) and the control channel $B^0_s \to J/\psi\phi$ ($J/\psi \to \mu^+\mu^-$, $\phi \to K^+K^-$). The signal region in the $B^0_{(s)} \to \mu^+\mu^-$ data sample, defined as the interval from 5166 MeV to 5526 MeV in the invariant dimuon mass ($m_{\mu\mu}$), is blinded. This prevents the analyzer from (unintentionally) optimizing the event selection towards a favored result. The sideband regions are defined as the dimuon mass interval from 4766 MeV to 5966 MeV, excluding the blinded region. An example of the signature for a signal event in the detector is presented in Figure 5.1.

The following trigger settings maximize the covered phase space, given that only a limited amount of data can be recorded. The first-level-dimuon trigger [ATL17b] is passed only by muon pairs with one muon having a transverse momentum p_{T}>4 GeV and the other having p_{T}>6 GeV. The high level trigger confirms the muon p_{T} threshold and selects only muon pairs with an invariant mass in the range 4 GeV to 8.5 GeV.

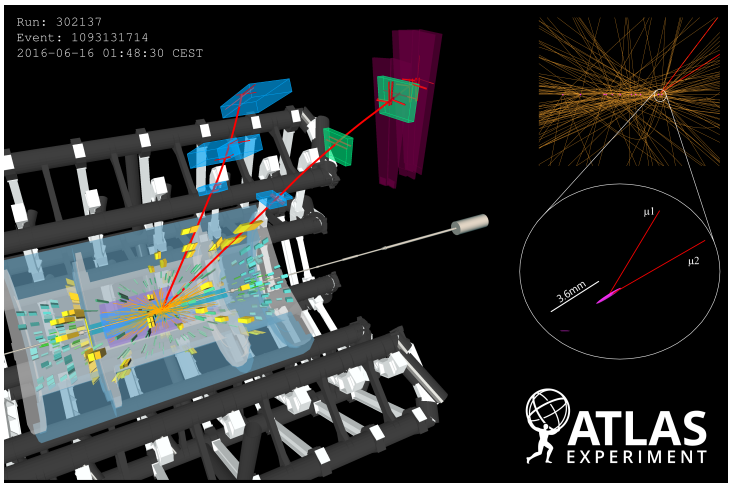


Figure 5.1: Detector signature of an $B^0_{(s)} \to \mu^+\mu^-$ event candidate recorded with the ATLAS detector [ATL20]. The primary vertex and the B decay vertex are shown in the circle at the right side.

After the full reconstruction of the events additional constraints are imposed on the muons as well as on the B candidates. For example, one hit in the pixel, two hits in the semiconductor tracker and reconstruction in the Muon Spectrometer are required for muon tracks. The B candidates are required to have a χ^2 per degree of freedom smaller than 6 for the decay vertex fit. Their p_{T} must be larger than 8 GeV and the absolute value of the η smaller than 2.5. Furthermore, collinearity requirements for the B candidate's momentum $\vec{p}^{\,B}$ are imposed. For example, the angle between $\vec{p}^{\,B}$ and the vector from the primary vertex to the decay vertex must not exceed 1 rad. Additional constraints were applied to J/ψ and ϕ decays in the reference and control channels.

Applying the selections above, a total effective integrated luminosity of $36.2\,\mathrm{fb}^{-1}$ was available for the signal channel and $15.1\,\mathrm{fb}^{-1}$ for each of the other channels. After unblinding about 1×10^6 events were observed in the signal region, that passed this preselection. The final selection is based on the BDT output variable as detailed in Section 5.3.2.

5.2.2 Introduction to the sideband subtraction technique

Prior to explaining the sideband subtraction technique itself, the terminology and the context for the application of this technique is introduced here. An event is defined as a set of variables describing a physical measurement. This could e.g. be the properties of a muon pair such as its invariant mass, the transverse momenta of the muons etc. An event can have different origins, which are denoted as species. In the simplest, non-trivial case there are two species, typically called signal and background. The former could e.g. be the final state particles μ^+ μ^- K^+ from the $B^+ \to J/\psi K^+$ ($J/\psi \to \mu^+\mu^-$) decay and the latter a random combination of tracks, resulting in the same final state μ^+ μ^- K^+. Variables in the events with known PDFs for all species are

called discriminating variables. Control variables are defined as variables with unknown PDFs for the species.

The sideband subtraction technique [Lyo86] is commonly used in HEP to statistically subtract the background contribution from the total control channel distribution. To apply the sideband subtraction technique, a discriminating variable is needed, which must not be correlated to the control variables. In addition, a signal region and sideband regions must be defined for the discriminating variable. In the sideband regions the background contribution should be dominant, although it is possible to account for signal events leaking into the sideband regions [Che19]. The first step in the application of the sideband subtraction technique is to perform a fit of e.g. the signal and background PDFs to the discriminating variable distribution. Based on the resulting PDFs and the fitted number of signal and background events, weights defined as

$$w_i = \begin{cases} -1 \times \frac{\int_{\text{signal region}} f_{\text{bkg}}(x)\ \mathrm{d}x}{\int_{\text{left sideband}} f_{\text{bkg}}(x)\ \mathrm{d}x + \int_{\text{right sideband}} f_{\text{bkg}}(x)\ \mathrm{d}x}\ , & \text{for event } i \text{ in sideband region} \\ 1\ , & \text{for event } i \text{ in signal region} \end{cases} \tag{5.2}$$

are applied to the ith event, where $f_{\text{bkg}}(x)$ represents the background PDF of the discriminating variable x, with parameters extracted by the fit to the discriminating variable. The weights effectively scale the background distribution of the control variable in the sideband regions to the number of background events in the signal region and subtract the resulting distribution from the distribution of the control variable in the signal region. Thus, an approximation of the signal distribution of the control variable remains. This technique is used in the next section to compare MC samples to data. To compare a signal MC sample to data, the background events must be subtracted from data.

The cuts defining the signal region in the discriminating variable become a source of systematic uncertainty, which requires additional studies. The advantage of the sPlot technique is, that it avoids such cuts. In the next chapter the sPlot technique is studied to investigate whether it can be used as a replacement for the sideband subtraction technique in the $B^0_{(s)} \to \mu^+\mu^-$ analysis with the ATLAS experiment.

5.2.3 MC samples, data-MC sample comparisons and reweighting

For each data sample there is a signal-only MC sample. A $b\bar{b} \to \mu^+\mu^- X$ MC sample, also referred to as continuum background MC sample, is used to study background processes. The events are generated with Pythia8 [SMS07] and the detector response is simulated by Geant4 [Ago+03; ATL10]. After that, the events are passed to the ATLAS reconstruction procedure [ATL08].

It needs to be ensured, that the MC samples resemble the data as closely as possible. Therefore the MC samples are compared to data and corrected if necessary. First a pile-up reweighting is

performed to adjust the distribution of the number of primary vertices in the MC sample to the one observed in the recorded data. Next, scale factors to correct for differences in muon trigger and muon offline reconstruction efficiencies are applied, which were obtained by a tag-and-probe method on J/ψ candidates [ATL17a; ATL14].

To account for differences in the kinematic quantities between data and MC samples, corrections for $p_{\mathrm{T}}^{\mathrm{B}}$ and η^{B} are applied. In order to obtain the corresponding event weights the $B^+ \to J/\psi K^+$ MC sample distributions for $p_{\mathrm{T}}^{\mathrm{B}}$ and η^{B} are compared to the $B^+ \to J/\psi K^+$ data after statistically subtracting the background contribution from data. Observed differences are corrected by reweighting the MC samples of the signal channels and the reference channel with weights defined as $w_k = N^{\mathrm{data}}/N^{\mathrm{MC}}$ for the kth $p_{\mathrm{T}}^{\mathrm{B}}$ or η^{B} bin [ATL12b]. These weights agree with weights obtained from $B_s^0 \to J/\psi\phi$ decays, establishing that the reference channel weights are not only applicable to $B^0 \to \mu^+\mu^-$ decays but to the $B_s^0 \to \mu^+\mu^-$ decays too. Only half of the reference channel data is used for this reweighting purpose. The other half is used to extract the B^+ event yield, in order to avoid correlations between event yield and weights.

5.3 Background treatment and final selection

The background events in the data of the $B^0_{(s)} \to \mu^+\mu^-$ analysis are assigned to three categories, which are described in the first section. The next section describes the BDT used for the final event selection and the data-MC sample comparisons for its input variables, which are used to assess systematic uncertainties.

5.3.1 Background components

There are three background categories: First, there is a peaking background, arising from $B^0_{(s)} \to hh'$ decays, where h denotes a hadron, with both hadrons misidentified as muons. The largest contribution arises from $B^0 \to K^\pm\pi^\mp$ decays. The reconstructed mass from such decays is very close to the B mass and lies within the signal region.

Second, the background category of Partially Reconstructed Decays (PRDs) is populated by b-hadron decays with one or more final state particle missed in the reconstruction such that only a muon pair is left, e.g. $B^0 \to K^{*0}\mu^+\mu^-$ decays, with K^{*0} missed in the reconstruction or $b \to c\mu^-\bar{\nu} \to s(d)\mu^+\mu^-\nu\bar{\nu}$, with the muons reconstructed only . Those events accumulate in the lower mass sideband of the $B^0_{(s)} \to \mu^+\mu^-$ data with a tail reaching into the signal region.

Third, the continuum background is composed of events where muons from two uncorrelated decays with an invariant mass falling into the analysis range are reconstructed as a muon pair.

Studies from the signal mass sidebands show, that this background component can be described by the $b\bar{b} \to \mu^+\mu^- X$ MC sample. The efforts to reduce such events are described in the next section (5.3.2).

5.3.2 Background reduction and data-MC sample comparisons on BDT input variables

The final event selection aims primarily at reducing the number of continuum background events. This is achieved by exploiting BDTs. They are trained, validated and evaluated on $B^0_{(s)} \to \mu^+\mu^-$ signal MC samples and data events in the sideband regions following a rotation scheme. Dividing both samples in three subsamples, three BDTs are trained, each on its own subsample. After validating each BDT on another subsample and finding that they provide statistically compatible outputs, each BDT is evaluated on the remaining third subsample not touched by it before. The BDTs have 15 input variables, taken from muon properties, B meson properties and other event properties. The full list is provided in table 1 of [ATL18]. The two variables with the highest separation power are ΔR and $|\alpha_{\mathrm{2D}}|$ (Fig. 5.2a). They are both related to the B meson momentum $\vec{p}^{B}$ and the vector from the primary vertex to the decay vertex $\Delta\vec{x}$. $|\alpha_{\mathrm{2D}}|$ is the absolute value of the angle between $\vec{p}^{B}$ and $\Delta\vec{x}$ in the transverse plane and ΔR is the angular distance between $\vec{p}^{B}$ and $\Delta\vec{x}$, defined as $\Delta R = \sqrt{|\alpha_{\mathrm{2D}}|^2 + (\Delta\eta)^2}$.

After verifying that the BDT output is not correlated with the dimuon mass, the final selection was defined as a BDT output value larger than 0.1439, which corresponds to 72 % signal efficiency and 0.3 % background efficiency. Events passing this selection are split into four BDT bins to account for different background shapes and signal to background ratios. The four BDT bins are indicated in Figure 5.2d.

A first data-MC sample comparison is carried out to check whether the continuum background MC sample describes the sideband data correctly. In case of observed differences, the MC sample could not readily be used to derive the functional forms of the background distributions. However, fair agreement is found as e.g. shown in Figure 5.2a.

Another data-MC sample comparison is conducted to assess the systematic uncertainties of the efficiencies $\epsilon_{\mu^+\mu^-}$ and $\epsilon_{J/\psi K^+}$ introduced by differences in MC sample and data. Since the signal region is blinded, only the MC samples of reference and control channel can be compared to their data counterparts. In order to do this, binned Maximum Likelihood (ML) fits to their invariant mass distributions divided into bins of $p_{\mathrm{T}}^{\mathrm{B}}$ and η^{B} are performed. The binning improves the purity of the signal distribution after background subtraction, as the data is effectively divided in categories of mass resolution and signal to background ratio. After subtracting the background contribution by using the sideband subtraction technique, the differences between MC sample and data are studied and considered as systematic uncertainties, as discussed in Section 5.4.2.

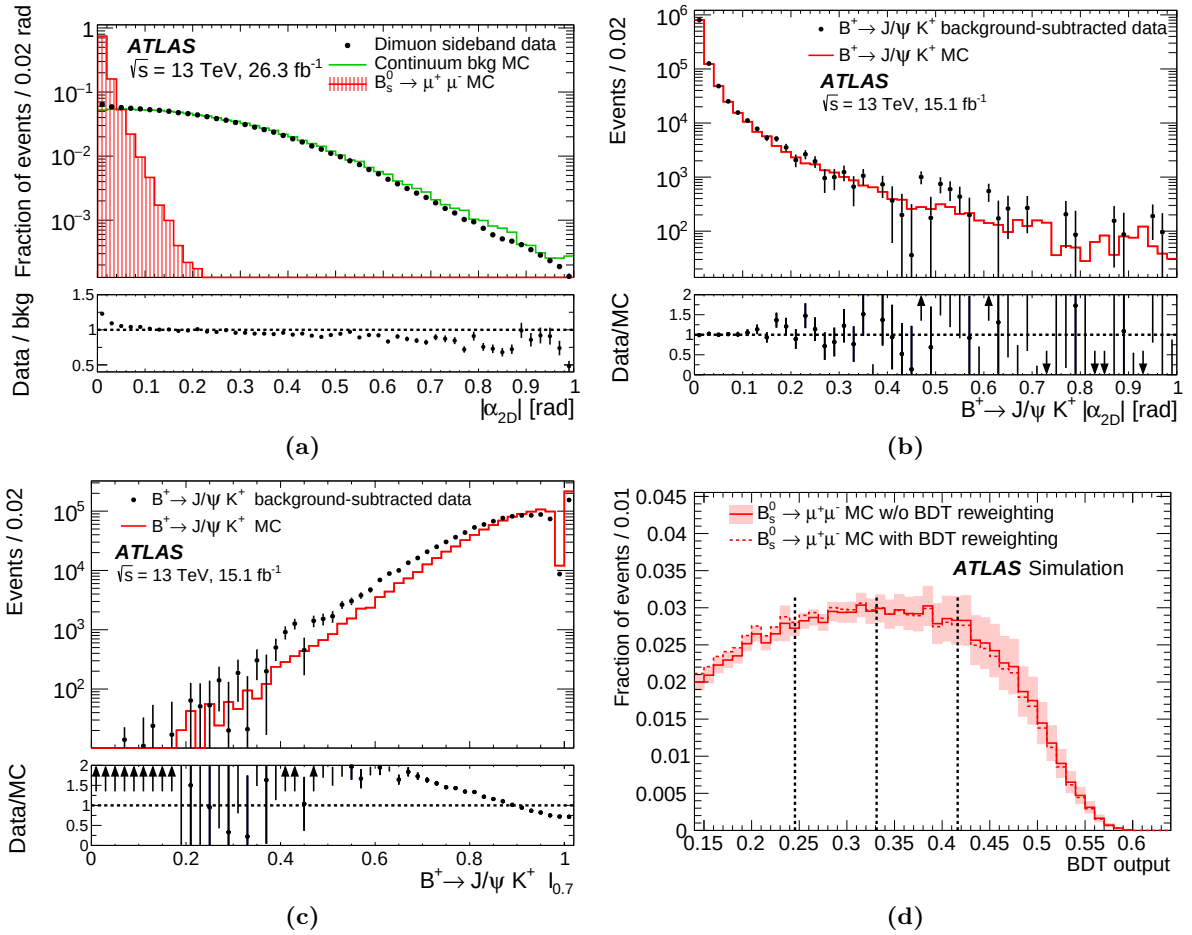


Figure 5.2: In (a) the discriminating power of $|\alpha_{2D}|$ is demonstrated, as well as the fair agreement between sideband data and continuum MC sample, proving that this MC sample can be used to derive the background PDFs for the continuum background contribution. In (b) the data-MC sample comparison for $|\alpha_{2D}|$ is shown, as an example for good agreement between background-subtracted $B^+ \to J/\psi K^+$ data and MC sample. (c) As a contrast the BDT input variable $I_{0.7}$ with the largest discrepancy between MC sample and data is shown. In (d) the BDT output with and without reweighting of the $I_{0.7}$ variable is presented. The red band indicates the systematic uncertainties for the efficiency calculation, derived by the procedure described in Section 5.4.2. The boundaries of the four BDT bins are indicated as vertical, dashed lines. All figures are taken from [ATL18] and its auxiliary material.

Two examples of a data-MC sample comparison for BDT input variables are given in Figures 5.2b and 5.2c, including the B isolation variable $I_{0.7}$ with the largest discrepancies observed and the highly discriminating variable $|\alpha_{\mathrm{2D}}|$.

5.4 Event yield extraction and efficiency determination

Calculating the branching fractions according to Equation (5.1) requires the extraction of the following quantities: $N_{J/\psi K^+}$, the efficiency ratio $\epsilon_{J/\psi K^+}/\epsilon_{\mu^+\mu^-}$ and the signal event yields $N_{d(s)}$, which are presented in this order in the three sections following.

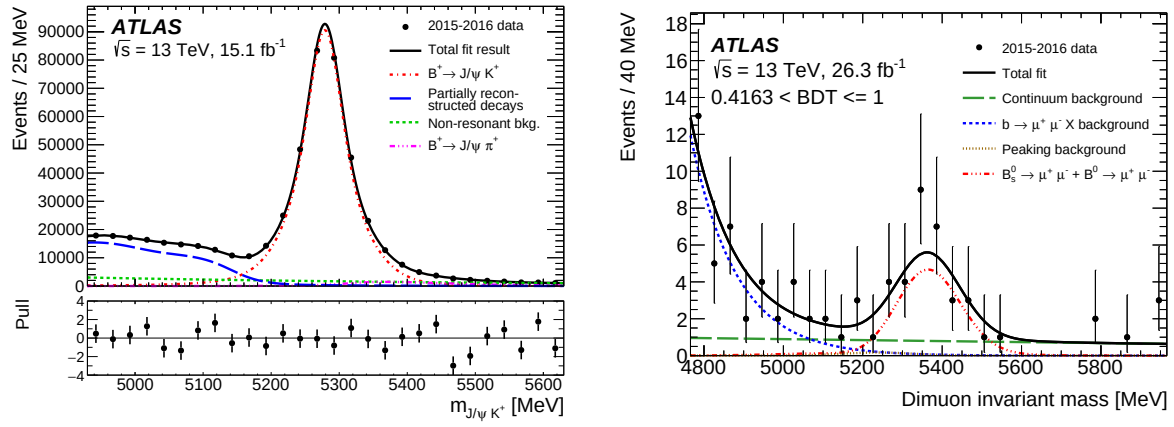
5.4.1 Reference channel event yield extraction

The reference channel event yield is extracted from an unbinned extended ML fit to the invariant mass distribution of $J/\psi\ K^+$. The functional forms for the signal and background PDFs are taken from the MC sample. At first, only the MC sample is fitted to assess the shape parameters. Then the MC sample and the data distribution are fitted simultaneously with shape parameters constrained from the previous fit to the MC sample only, except for the continuum background, whose shape parameters are unconstrained. The event yield parameters are unconstrained, too. Differences between MC sample and data in the mass scale and resolution can arise from imperfections of the detector modeling and are accounted for by adding free parameters for the mass scale and resolution to the fit. The fit result is shown in Figure 5.3a.

5.4.2 Efficiency ratio determination

The efficiencies $\epsilon_{\mu^+\mu^-}$ and $\epsilon_{J/\psi K^+}$ from Equation (5.1) are obtained from MC samples, since the information about the true identity of the event is required. Both, the reference channel and the signal channel events, were analyzed in the fiducial B meson region, defined as $p_{\mathrm{T}}^{\mathrm{B}} > 8\,\mathrm{GeV}$ and $|\eta^{\mathrm{B}}| < 2.5$, which constitutes the acceptance for the B meson. The acceptance for the leading and trailing muon in the final state is $p_T^{\mu_l} > 6\,\mathrm{GeV}$ and $p_T^{\mu_t} > 4\,\mathrm{GeV}$. Furthermore, $\epsilon_{J/\psi K^+}$ and $\epsilon_{\mu^+\mu^-}$ include trigger, reconstruction and selection efficiencies; see [ATL18] for the details.

The largest systematic uncertainty of the efficiency ratio arises from discrepancies between MC sample and data in the BDT input variables, as mentioned in Section 5.2 and shown for few examples in Figures 5.2a to 5.2c. They are investigated by reweighting the MC sample for $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ for each input variable by the ratio of the sideband subtracted $B^+ \to J/\psi K^+$ data and the corresponding MC sample. Only for the isolation variable $I_{0.7}$ the weights for the $B^0_{(s)} \to \mu^+\mu^-$ channels are taken from $B_s^0 \to J/\psi\phi$, since this isolation variable



(a) (b)

Figure 5.3: In (a) the results of the fit to half of the reference channel $B^+ \to J/\psi K^+$data is presented. In (b) the fit to the $B^0_{(s)} \to \mu^+\mu^-$ data in the highest BDT bin is shown. The fit results for $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ are summed in the red dashed-and-dotted line. The PRDs (blue dashed line) are referred to as $b \to \mu^+\mu^- X$. The brown, dotted curve for the peaking background lies very close to the horizontal axis and is hardly visible [ATL18].

is based on charged tracks and differences are expected for B^+ and B^0. The impact of the reweighting on the BDT output is shown in Figure 5.2d.

5.4.3 Signal event yields extraction

The $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ event yields are extracted from an unbinned extended ML fit to the signal invariant dimuon mass. The data is divided in four BDT bins, to account for differences in the signal to background ratio and changes in the background shape. All data in the four BDT bins are fitted simultaneously.

The fit model has four components. The signal PDFs are both described by a double Gaussian PDF. The continuum background contribution is described by a PDF with linear dependence on the invariant dimuon mass. The other components are the peaking background and the PRDs, which are modeled by a double Gaussian PDF and an exponential PDF respectively. For each component, except the continuum background, the shape parameters are constrained from the MC sample. The event yields are not constrained and may become negative. The fit result in the highest BDT bin is shown in Figure 5.3b.

5.5 Results for the branching fractions

The final results for the branching fractions are obtained by a maximum likelihood fit. The yields N_d and N_s in the likelihood are replaced by the corresponding branching fractions by rearranging

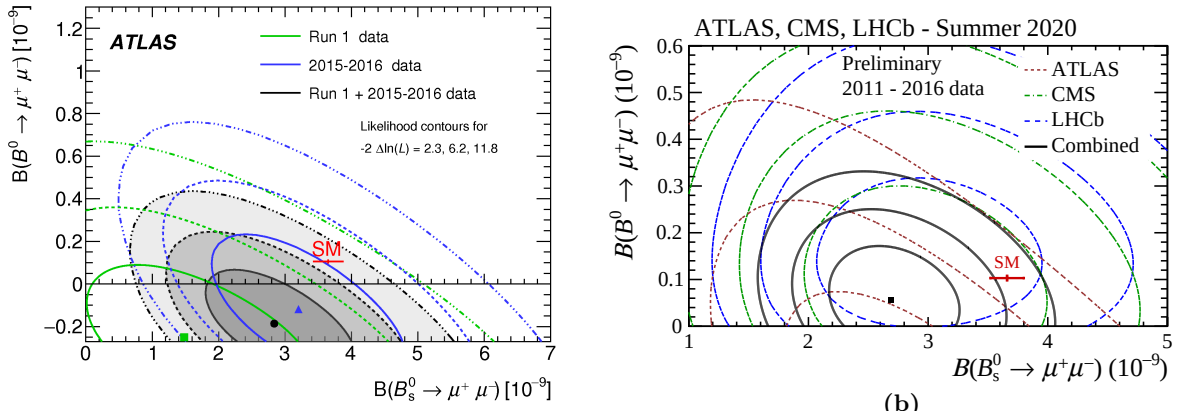


Figure 5.4: In (a) the likelihood contour in the $B^0_s \to \mu^+\mu^- - B^0 \to \mu^+\mu^-$ plane for LHC Run-I, Run-II data taking period and the combination of both is provided [ATL18]. The SM expectation is given in red. In (b) the combined likelihood for ATLAS, CMS and LHCb is given [ACL20].

Equation (5.1). The branching fractions are then fitted simultaneously. The resulting likelihood contours are shown in Figure 5.4a.

The analysis presented here has previously been carried out with data from the LHC Run-I data taking period [ATL16a]. The results are embodied by the green likelihood contour (Fig. 5.4a). The blue line presents the results of the analysis presented in this chapter. Obviously, the likelihood contours are covering areas with negative branching fractions for $B^0 \to \mu^+\mu^-$, which are unphysical. Therefore, a Neyman construction [Ney37] was considered in both analyses.

To shorten the discussion of the results, only the values of the branching fraction obtained from the likelihood combination of the two analyses are presented here. The combined result for the decay of B^0_s is [ATL18]:

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = \left(2.8^{+0.8}_{-0.7}\right) \times 10^{-9} \,.$$

For the decay of B^0 an upper limit of

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.1 \times 10^{-10}$$

at 95 % Confidence Level (CL) was obtained [ATL18].

The combination of the results, as shown in Figure 5.4b, from ATLAS [ATL18], CMS [CMS20] and LHCb [LHC17] for the full LHC Run-I and partial Run-II yields [ACL20]

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = \left(2.7^{+0.4}_{-0.3}\right) \times 10^{-9}$$

and the upper limit for the branching fraction on $B^0_s \to \mu^+\mu^-$ at 95 % CL is

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.6 \times 10^{-10} \, .$$

In summary, an upper limit was established for the $B^0 \to \mu^+\mu^-$ decay. The measurement by ATLAS is compatible with the standard model in the $B^0_s \to \mu^+\mu^- - B^0 \to \mu^+\mu^-$ plane within 2.4 standard deviations and the combination from ATLAS, CMS and LHCb within 2.11 σ. There is no evidence for physics beyond the SM yet.

6 Statistical methods and software toolkits

The methods introduced in this chapter are needed for the studies in Chapter 7, where unbinned maximum likelihood fits to weighted events produced by the sPlot technique are discussed. Special attention is paid to the uncertainties of the fitted parameters.

Therefore, the first method explained (Sec. 6.1) is the sPlot technique, used to calculate event weights, which applied, separate signal and background distributions. The calculation of the weights requires a parameter estimation with the unbinned Maximum Likelihood (ML) method, discussed in Section 6.2. This method can also be used to estimate the parameters of weighted distributions produced by the sPlot technique. Therefore, the discussion includes the treatment of per-event weights in unbinned ML fits, as well as the calculation and correction of the fitted parameter uncertainties. To check the coverage of the uncertainties determined from the maximum likelihood fits, pulls are used. They are discussed in Section 6.3. In the last section an overview of the software toolkits and of numerical techniques used for the studies in Chapter 7 is provided.

6.1 The sPlot technique

The sPlot technique [PD05] is used to statistically separate the distributions of the different species[§] for a control variable by applying per-event weights, which are based on a fit to the discriminating variable distribution. This fit is presented in Section 6.1.1 and the weights, called sWeights are introduced and described in Section 6.1.2. In Section 6.1.3 it is pointed out where care must be taken in the application of the sPlot technique to obtain the correct result. A list of the most important symbols used in this section is provided in Table 6.1.

6.1.1 Discriminating variable fit

Similar to the sideband subtraction technique, introduced in Section 5.2.2, the sPlot technique is based on a fit to a discriminating variable distribution. The discriminating variable PDF is a

[§]The terms species, control and discriminating variable have already been introduced in Section 5.2.2.

Table 6.1: List of the most important symbols used for the description of the sPlot technique in Section 6.1.

Symbol	Definition
y	Discriminating variable
x	Control variable
$f_i(y_e\|\vec{\lambda})$	Component for species i of the discriminating variable PDF evaluated for the discriminating variable value of event e
N	Total number of events in the data set
N_s	Number of species in the data set
$\mathcal{N}_{i,j,k,...}$	Average number of events in species $i, j, k, ...$
V_{ij}^{yields}	Elements of the covariance matrix from the discriminating variable fit with all parameters except the numbers of events (yields) fixed
${}_s\mathcal{P}_i(y_e)$	sWeight of event e for species i

linear combination of the PDFs $f_i(y|\vec{\lambda_i})$ with coefficient a_i for the ith species:

$$f(y) = \sum_{i=1}^{N_s} a_i \; f_i(y|\vec{\lambda_i}) \quad \text{with} \quad \sum_{i=1}^{N_s} a_i = 1 \; \wedge \; a_i > 0 \; . \tag{6.1}$$

For the discriminating variable fit, it is mandatory to use the Extended Maximum Likelihood (EML) method, which will be presented in Section 6.2.4. According to Equation (6.27), the likelihood function of the discriminating variable fit includes a Poisson term for the average total number of events $\mathcal{N}$, written as the sum of the average number of events for all N_s species: $\mathcal{N} = \sum_{k=1}^{N_s} \mathcal{N}_k$. The extended Negative Log Likelihood (NLL) for a data set with N measurements then reads

$$\mathcal{L}(\vec{\lambda}, \vec{\mathcal{N}}) = -\sum_{e=1}^{N} \log \left(\sum_{i=1}^{N_s} \mathcal{N}_i f_i(y_e|\vec{\lambda_i}) \right) + \sum_{i=1}^{N_s} \mathcal{N}_i + \log N! \; . \tag{6.2}$$

6.1.2 Definition and properties of the sWeights

An intuitive guess for the weights to separate the distributions of the species would be the probability that event e belongs to species i:

$$w_i(y_e) = \frac{N_i f_i(y_e|\vec{\lambda})}{\sum_{j=1}^{N_s} N_j f_j(y_e)} \; . \tag{6.3}$$

But this is merely scaling the distributions as e.g. all events in a region, where the signal is dominant, are assigned a high probability for being a signal event, including some background events. Constructing the control variable distribution of e.g. the signal species will therefore include a background contribution, which is not desired. Consequently, negative weights are needed to allow for the statistical subtraction of events pertaining to other species. The sWeights

${}_s\mathcal{P}_i(y_e)$ meet this requirement and are defined for species i as [PD05]

$$ {}_s\mathcal{P}_i(y_e) = \frac{\sum_{j=1}^{N_s} V_{ij}^{\text{yields}} f_j(y_e)}{\sum_{j=1}^{N_s} N_j f_j(y_e)} \,, \tag{6.4} $$

where V_{ij}^{yields} are the covariance matrix elements from the discriminating variable fit with all parameters, except the numbers of events, fixed. The matrix may also be calculated analytically [PD05]:

$$ V_{ij}^{\text{yields}} = \frac{\partial^2 \mathcal{L}(\vec{\lambda})}{\partial N_i \partial N_j} = \sum_{e=1}^{N} \frac{f_i(y_e) f_j(y_e)}{\left(\sum_{j=1}^{N_s} N_j f_j(y_e)\right)^2} \,. \tag{6.5} $$

In case the sWeights for the ith species are applied to all events in the data set, the control variable distributions of this species are obtained. The resulting distributions for species i will be properly normalized to the average number of events in this species, since the sum of all sWeights yields this average number [PD05]:

$$ \sum_{e=1}^{N} {}_s\mathcal{P}_i(y_e) = N_i \,. \tag{6.6} $$

Furthermore, each event is effectively only counted once, as the sum of its sWeights for all species is one:

$$ \sum_{i=1}^{N_s} {}_s\mathcal{P}_i(y_e) = 1 \,. \tag{6.7} $$

6.1.3 Crucial points in the application of the sPlot technique

The sPlot technique is only applicable to a data set if the discriminating variable and the control variables are statistically independent. Thus, the PDFs describing the underlying population, from which the data set was sampled, must factorize for the discriminating and control variables in each species separately.

In case the parameters of the discriminating variable PDF are not known, the PDF must be fitted twice to the discriminating variable distribution. The first time with all parameters allowed to vary freely and the second time with all parameters except the numbers of events fixed to the previously fitted values. Otherwise the covariance matrix results in wrong sWeights and thus the distributions for the species of the control variable distributions are not recovered faithfully.

6.2 Unbinned maximum likelihood parameter estimation

First, in Section 6.2.1, the likelihood function is presented, including some of its properties and its application in parameter estimation. Since the uncertainties of the parameters and their

correlations are crucial for a measurement result, their calculation is explained in Section 6.2.3. The sPlot technique requires the concept of Extended Maximum Likelihood (EML) parameter estimation, which is presented in Section 6.2.4. It also introduces weights, which need to be included in the likelihood function, as explained in Section 6.2.5. However, the inclusion of weights into the likelihood function requires a correction of the resulting covariance matrix, discussed in Section 6.2.6.

6.2.1 Estimation of parameters

The likelihood function is used to estimate the parameters of the PDF, that describes the underlying population of the measurements. Given N measurements $x_1, \ldots, x_N$ of a random variable x and assuming x follows a PDF $f(x|\vec{\lambda})$ with parameters $\vec{\lambda}$, the likelihood function is defined as

$$L(\vec{\lambda}) = \prod_{e=1}^{N} f(x_e|\vec{\lambda}) \; . \tag{6.8}$$

$L(\vec{\lambda})$ describes the probability, that the measurements x_e were sampled from an underlying population described by $f(x|\vec{\lambda})$. By maximizing the likelihood function the most probable values for $\vec{\lambda}$ are found. For convenience usually the negative logarithm of the likelihood function, referred to as NLL, is minimized. It is defined as

$$\mathcal{L}(\vec{\lambda}) = -\log L(\vec{\lambda}) = -\sum_{e=1}^{N} \log f(x_e|\vec{\lambda}) \; . \tag{6.9}$$

The Maximum Likelihood Estimator (MLE) $\widehat{\lambda}_i$ for the ith parameter fulfills the condition

$$\left.\frac{\partial \mathcal{L}(\vec{\lambda})}{\partial \lambda_i}\right|_{\widehat{\lambda}_i} = 0 \; . \tag{6.10}$$

Estimators derived this way have desirable properties. They are consistent, i.e. they converge to the true parameters $\vec{p}$ for large N, which is referred to as the asymptotic limit. In this limit MLEs are also efficient as they reach the smallest possible variance, which is the Cramér-Rao (CR) bound [Cra46; Rao92]:

$$V[\widehat{\lambda}_i] = \frac{\left(1 + \frac{\partial b(\lambda_i)}{\partial \lambda_i}\right)^2}{\mathrm{E}\left[\frac{\partial^2 \mathcal{L}(\vec{\lambda})}{\partial \lambda_i^2}\right]} \; , \tag{6.11}$$

where $b(\lambda_i)$ denotes the bias of the estimator and E[x] the expectation value of x.

6.2.2 Correction of the bias for fitted parameters in the non-asymptotic case

MLEs are guaranteed to be unbiased only in the asymptotic case. In the non-asymptotic case they can be biased. In some cases it is possible to calculate the expected bias and then to correct

the fitted parameter. In general, the bias for an estimator of the parameter λ of a PDF $f(x|\lambda)$, defined for the interval $x \in [a, b]$, is obtained by calculating its expectation value, which can be written as a function $h(\lambda)$:

$$\mathrm{E}\left[\widehat{\lambda}\right] = \int_a^b \widehat{\lambda}\, f(x|\lambda)\, \mathrm{d}x = h(\lambda)\,. \tag{6.12}$$

If $h(\lambda) = \lambda$, the estimator is unbiased. If not, the estimator corrected for the bias is obtained by calculating $h^{-1}(\hat{\lambda})$, if the inverse function $h^{-1}(\hat{\lambda})$ of $h(\lambda)$ exists. Two examples relevant for this thesis are given below, concerning the decay rate λ of an exponential PDF and the standard deviation of a Gaussian PDF.

Correction of the decay constant for an exponential PDF

An exponential PDF for a random variable $x \in [a, b]$ with decay rate $\lambda < 0$ may be defined as

$$e(x) = \frac{\lambda}{\mathrm{e}^{\lambda b} - \mathrm{e}^{\lambda a}}\, \mathrm{e}^{\lambda x}\,. \tag{6.13}$$

In case $a = 0$ and $b = \infty$ the MLE for λ is

$$\widehat{\lambda} = \frac{N}{\sum_{e=1}^{N} x_e}\,, \tag{6.14}$$

with x_e denoting the N data values. The expectation value of $\hat{\lambda}$ and the corresponding corrected estimator $\hat{\lambda}_{\mathrm{corr}}$ are given by [Cow98, p. 73]

$$\mathrm{E}\left[\widehat{\lambda}\right] = \frac{N}{N-1}\,\lambda \quad \Rightarrow \quad \widehat{\lambda}_{\mathrm{corr}} = \frac{N-1}{N}\,\widehat{\lambda}\,. \tag{6.15}$$

Correction of the standard deviation for a Gaussian PDF

A Gaussian PDF for a variable $x \in [a, b]$ with mean μ and standard deviation σ is defined as

$$g(x) = \frac{2}{\mathrm{erf}\left(\frac{b-\mu}{\sqrt{2}\sigma}\right) - \mathrm{erf}\left(\frac{a-\mu}{\sqrt{2}\sigma}\right)} \frac{\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\,. \tag{6.16}$$

In case the distribution is not truncated, i.e. $a = -\infty$ and $b = +\infty$, the MLE for the variance reads

$$\widehat{\sigma^2} = \frac{\sum_{e=1}^{N} (x_e - \widehat{\mu})^2}{N}\,, \tag{6.17}$$

with an expectation value of

$$\mathrm{E}\left[\widehat{\sigma^2}\right] = \frac{N-1}{N}\,\sigma^2\,. \tag{6.18}$$

The correction, known also as Bessel's correction, is given by

$$\widehat{\sigma^2}_{\text{Bes}} = \frac{N}{N-1}\,\widehat{\sigma^2}\,. \tag{6.19}$$

If instead of an unbiased estimation of the variance an unbiased estimation of the standard deviation is of interest, one can not simply take the square root of the corrected MLE of the variance, since the square root is a non-linear function. Holtzman derived the estimator for the unbiased standard deviation to be [Hol50]:

$$\widehat{\sigma}_{\text{corr}} = c_N\,\sqrt{\widehat{\sigma^2}_{\text{Bes}}}\,, \tag{6.20}$$

where c_N is defined by

$$c_N = \sqrt{\frac{N-1}{2}}\,\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}\,. \tag{6.21}$$

The Γ function is the extension of the factorial to real numbers with $\Gamma(n) = (n-1)!$. For n larger than roughly 170 it is not possible to represent the values of $\Gamma(n)$ with double precision floating point numbers. Therefore, in this thesis an approximate formula provided by Brugger [Bru69] based on results of Cureton [Cur68] is used. The approximate unbiased standard deviation is given by

$$\widehat{\sigma}_{\text{corr}} = \sqrt{\frac{N}{N-1.5}}\sqrt{\widehat{\sigma^2}} = \sqrt{\frac{\sum_{e=1}^{N}\left(x_e - \widehat{\mu}\right)^2}{N-1.5}}\,. \tag{6.22}$$

Note on truncated PDFs

In case the conditions for a and b mentioned above are not met, the PDF is called truncated. The MLEs for σ and λ of a truncated exponential or Gaussian PDF can not be determined analytically; Equation (6.10) must be solved numerically. As a consequence neither the expectation value of the estimator nor the correction for a possible bias can be calculated.

6.2.3 Determination of parameter uncertainties

There are several ways to calculate the uncertainties on a ML estimator $\widehat{\lambda_i}$. One option is to use the CR bound (Eq. 6.11), which is only possible for large numbers of events where ML estimators reach the CR bound.

Another option is to exploit the Gaussian shape of the likelihood function in the asymptotic limit [KSO87]. For large N the NLL function will have a parabolic shape and can, considering a single parameter λ, be written as

$$\mathcal{L}(\lambda_i) = \mathcal{L}_{\text{min}} + \frac{\left(\lambda_i - \widehat{\lambda_i}\right)^2}{2\sigma_\lambda^2}\,, \tag{6.23}$$

where $\mathcal{L}_{\min}$ is the minimum of the NLL function. Consequently, the variance of $\widehat{\lambda_i}$ is calculated as

$$\mathrm{V}(\widehat{\lambda_i}) = \left(\frac{\partial^2 \mathcal{L}(\lambda_i)}{\partial \lambda_i^2} \right)^{-1} \tag{6.24}$$

in the one-dimensional case and in the absence of any bias. In the multi-dimensional case the covariance matrix V is given by

$$V = \left[\frac{\partial^2 \mathcal{L}(\vec{\lambda})}{\partial \lambda_i \partial \lambda_j} \right]^{-1} \quad , \tag{6.25}$$

where the square brackets denote the matrix, that is formed by the elements i and j. This convention will be used throughout this thesis.

A third option to determine the uncertainties for the parameter λ_i is to find the values σ_a and σ_b for which

$$\mathcal{L}\left(\lambda_i \, {}^{+\sigma_a}_{-\sigma_b}\right) = \mathcal{L}_{\min} + \frac{1}{2} \, . \tag{6.26}$$

σ_b and σ_a represent the positive and negative uncertainty of the parameter λ_i. In the asymptotic limit this can be easily verified using Equation (6.23). But for finite data sets this is still valid [EDJ71; FST79].

6.2.4 Extended unbinned maximum likelihood estimation

The concept of EML estimation is first mentioned in literature in a note by Orea [Ore58]. It allows the normalization of the PDFs to float and thus accounts for a fluctuating number of events in a data set. This is common in High Energy Physics (HEP), as data is usually taken for a fixed period of time and not until a predefined number of events has been collected. A floating normalization is achieved by multiplying a Poisson term for the average number of events $\mathcal{N}$, also called yield, to the likelihood function build from N measurements:

$$L_{\mathrm{E}}(\lambda) = \frac{\mathrm{e}^{-\mathcal{N}} \mathcal{N}^N}{N!} \, L(\lambda) \, . \tag{6.27}$$

As proved by Barlow, parameters estimated from the likelihood function are equal to those estimated from the extended likelihood function, as long as it is possible to vary the shape of the PDFs independently from the normalization [Bar90]. But the uncertainties of the parameters can be different. Those from the EML function are typically larger since the constraint on the total yield is missing. Whether the ML or EML method is to be used, depends on the situation studied. The former describes the yields in the particular experiment carried out and the latter the yields in an experiment such as the one carried out, that is, if the particular experiment would be repeated [LAP86]. An example for the former is the measurement of the ratio of kaons and pions in cosmic rays and an example for the latter is the determination of the absolute

kaon rate in cosmic rays [LAP86]. In Section 6.1 the EML method was used for the fit to the discriminating variable distribution involved in the sPlot technique.

In case the shape parameters and the yields of the PDFs can not be varied independently even the estimated parameter values from ML and EML method may differ and the parameters estimated from the EML might be more accurate. An example for this is a signal distribution with unknown shape and yield parameters on top of a background distribution with known shape and yield parameters [Bar90].

6.2.5 Unbinned maximum likelihood estimation in the presence of per-event weights

Event weights play an important role in HEP. They are e.g. used to adjust kinematic distributions in Monte Carlo (MC) samples to the corresponding distributions in data to correct for known deficiencies (Sec. 5.2.3) or for performing a statistical background subtraction with the sideband subtraction technique (Sec. 5.2.2), or with the sPlot technique (Ch. 7). They are included by defining the likelihood function as

$$L(\vec{\lambda}) = \prod_{i=1}^{N} f(x_i|\vec{\lambda})^{w_i} \quad \Rightarrow \quad \mathcal{L}(\vec{\lambda}) = -\sum_{i=1}^{N} w_i \log f(x_i|\vec{\lambda}) \; . \tag{6.28}$$

Equation (6.25) for the parameter uncertainties is no longer valid. This is most easily seen by considering a constant weight $w \neq 1$. Then the weight can be written in front of the inverse covariance matrix, resulting in

$$V = \frac{1}{w} \left[\frac{\partial^2 \mathcal{L}(\vec{\lambda})}{\partial \lambda_i \partial \lambda_j} \right]^{-1} \; . \tag{6.29}$$

Consequently, the uncertainties will be underestimated for $w > 1$ and overestimated for $w < 1$. Three methods to correct the uncertainties in the presence of per-event weights are presented in Section 6.2.6.

6.2.6 Corrections of the covariance matrix in the presence of per-event weights

As shown in Equation (6.29) the uncertainties in the presence of per-event weights must be corrected. Corrections proposed by Eadie et al. [EDJ71] and by Langenbruch [Lan19] are described in this section. Langenbruch presents two expressions for the correction of the covariance matrix. The first expression accounts for the presence of the sWeights and the second expression additionally accounts for the uncertainties of the sWeights, arising from the discriminating variable fit. The former is hereafter referred to as asymptotic correction part one and the latter as asymptotic correction part two.

Asymptotic correction part one

Langenbruch derives the covariance matrix V of the fitted parameters $\widehat{\vec{\lambda}}$ of the PDF $f(x|\vec{\lambda})$ estimated with the maximum likelihood method in the presence of weights w_e for event e [Lan19]:

$$\begin{aligned} V = & \left[\sum_{e=1}^{N} w_e \left.\frac{\partial^2 \log f(x_e|\vec{\lambda})}{\partial\lambda_i \partial\lambda_k}\right|_{\widehat{\vec{\lambda}}}\right]^{-1} \times \left[\sum_{e=1}^{N} w_e^2 \left(\left.\frac{\partial \log f(x_e|\vec{\lambda})}{\partial\lambda_i}\right|_{\widehat{\vec{\lambda}}} \left.\frac{\partial \log f(x_e|\vec{\lambda})}{\partial\lambda_k}\right|_{\widehat{\vec{\lambda}}}\right)\right] \\ & \times \left[\sum_{e=1}^{N} w_e \left.\frac{\partial^2 \log f(x_e|\vec{\lambda})}{\partial\lambda_i \partial\lambda_k}\right|_{\widehat{\vec{\lambda}}}\right]^{-1} . \end{aligned} \tag{6.30}$$

This can be identified as $V = \tilde{V} C \tilde{V}$, where $\tilde{V}$ is the uncorrected covariance matrix as in Equation (6.29) and C is the correction matrix including the weights. As mentioned above, the square brackets denote the matrix, that is formed by the elements with the indices found inside, e.g. i and k in the first square brackets in Equation (6.30).

Asymptotic correction part two

The correction of the covariance matrix in Equation (6.30) does not account for the uncertainties of the sWeights, which arise from nuisance parameters in the discriminating variable fit and their potential correlation with the event yields. The nuisance parameters in the sWeight calculation are the shape parameters of the discriminating variable fit model.

Considering the uncertainties of the sWeights leads to the following correction, which needs to be added to Equation (6.30) [Lan19]:

$$\begin{aligned} V = & \left[\sum_{e=1}^{N} w_e \left.\frac{\partial^2 \log f(x_e|\vec{\lambda})}{\partial\lambda_i \partial\lambda_k}\right|_{\widehat{\vec{\lambda}}}\right]^{-1} \\ & \times \left[\sum_{e=1}^{N} \left.\frac{\partial w_e}{\partial p_m}\right|_{\widehat{\vec{p}}} \left.\frac{\partial \log f(x_e|\vec{\lambda})}{\partial\lambda_k}\right|_{\widehat{\vec{\lambda}}}\right] \times \tilde{V}^{\text{discr}} \times \left[\sum_{e=1}^{N} \left.\frac{\partial w_e}{\partial p_n}\right|_{\widehat{\vec{p}}} \left.\frac{\partial \log f(x_e|\vec{\lambda})}{\partial\lambda_l}\right|_{\widehat{\vec{\lambda}}}\right] \\ & \times \left[\sum_{e=1}^{N} w_e \left.\frac{\partial^2 \log f(x_e|\vec{\lambda})}{\partial\lambda_l \partial\lambda_j}\right|_{\widehat{\vec{\lambda}}}\right]^{-1} , \end{aligned} \tag{6.31}$$

where the full set of parameters involved in the discriminating variable fit is denoted by $\vec{p}$, thus including shape and event yield parameters. Their best estimate is denoted by $\widehat{\vec{p}}$. The matrix $\tilde{V}^{\text{discr}}$ is defined as

$$\tilde{V}^{\text{discr}} = V^{\text{discr}} - \left[V_{ij}^{\text{yields}}\right] , \tag{6.32}$$

where the yield matrix $\left[V_{ij}^{\text{yields}}\right]$, defined in Equation (6.5), is filled with zeros to match the dimensions of the full covariance matrix of the discriminating variable fit V^{discr}. Both matrices on the Right Hand Side (RHS) must of course have the same order of parameters in their rows and columns.

Squared weights correction

Eadie et al. [EDJ71] proposed a correction matrix

$$C = \left[\sum_{e=1}^{N} w_e^2 \left. \frac{\partial^2 \log f(x_e|\vec{\lambda})}{\partial\lambda_i \partial\lambda_j} \right|_{\vec{\lambda}} \right] \tag{6.33}$$

to be applied to the uncorrected covariance matrix $\tilde{V}$ in the form of $V = \tilde{V}C\,\tilde{V}$. This is a special case of the asymptotic correction part one, since if [Lan19]

$$\mathrm{E}\left[\frac{w^2}{f(x|\vec{\lambda})} \left. \frac{\partial^2 f(x|\vec{\lambda})}{\partial\lambda_i \partial\lambda_j} \right|_{\vec{\lambda}} \right] = 0 \tag{6.34}$$

both corrections are the same. This can be found by calculating the second derivatives in the C matrix of the squared weights correction under consideration of the chain rule and comparing it to the corresponding terms in the C matrix of the asymptotic correction part one.

The squared weights correction will generally not result in correct parameter uncertainties unless the assumption in Equation (6.34) holds.

6.3 Statistical tests with pull distributions

Pull distributions provide a tool to investigate the consistency of fitted parameter uncertainties with the fluctuation of the fitted parameter values [DL02]. For pseudo experiment studies, where a data set is repeatedly generated and fitted, the contribution to the pull distribution by the ith fit is defined by

$$\theta_i = \frac{\lambda_i^{\mathrm{fit}} - \lambda^{\mathrm{true}}}{\sigma_{\lambda_i}^{\mathrm{fit}}} \,, \tag{6.35}$$

where λ_i^{fit} is the fitted parameter value and λ^{true} is the true value of the parameter used in the generation of the pseudo experiments. Given a set of pseudo experiments, the pull distribution is the distribution of the θ_i. For later reference the difference between the fitted parameter value and its true value is defined by

$$d_i = \lambda_i^{\mathrm{fit}} - \lambda^{\mathrm{true}} \,. \tag{6.36}$$

The distribution of the d_i for a set of pseudo experiments will be referred to as d_i-distribution.

To determine the expected pull distribution, the pull variable is modeled as $z = {}^{x}/_{y}$, where x, y and z are random variables described by the PDFs $p(x, y)$ and $g(z)$. x represents d_i and y the parameter uncertainty $\sigma_{\lambda_i}^{\mathrm{fit}}$. The expected pull distribution is described by the PDF $g(z)$, which is found by solving the integral

$$g(z) = \int_0^\infty y\; p(zy, y)\mathrm{d}y \,. \tag{6.37}$$

The fitted parameter values and their uncertainties are assumed to be Gaussian distributed. Thus, $p(x, y)$ is modeled as a two-dimensional Gaussian distribution with means $\mu_x,\ \mu_y$, standard deviations $\sigma_x,\ \sigma_y$ and correlation coefficient ρ:

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\ \exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right) \tag{6.38}$$

A correlation between the parameter value and its uncertainty occurs e.g. for the MLE of the standard deviation of a Gaussian distribution as well as for the MLE of the decay constant of an exponential distribution. The uncertainties of these estimators are proportional to the estimator itself. For example, the variance of the MLE for the decay constant of an exponential PDF, presented in Equation (6.14), is given by

$$\mathrm{V}\left[\widehat{\lambda}\right] = \frac{\widehat{\lambda}^2}{N}\,. \tag{6.39}$$

This can be calculated using the CR bound from Equation (6.11).

Solving the integral to determine $g(z)$ yields

$$\begin{aligned}
g(z) = \frac{e^{-\frac{(\mu_x-\mu_y z)^2}{2\left({\sigma_x}^2-2\rho\sigma_x\sigma_y z+{\sigma_y}^2 z^2\right)}}}{2\sqrt{2}\pi\sqrt{1-\rho^2}\left(-1+\rho^2\right){\sigma_x}^3{\sigma_y}^3\left(-\frac{{\sigma_x}^2-2\rho\sigma_x\sigma_y z+{\sigma_y}^2 z^2}{\left(-1+\rho^2\right){\sigma_x}^2{\sigma_y}^2}\right)^{3/2}}\Bigg[\\
+\mu_y\sqrt{\pi}\sigma_x(-\sigma_x+\rho\sigma_y z)+\sigma_y\Bigg(\\
\mu_x\sqrt{\pi}(\rho\sigma_x-\sigma_y z) \\
+\sqrt{2}e^{\frac{(\mu_x\sigma_y(-\rho\sigma_x+\sigma_y z)+\mu_y\sigma_x(\sigma_x-\rho\sigma_y z))^2}{2\left(-1+\rho^2\right){\sigma_x}^2{\sigma_y}^2\left({\sigma_x}^2-2\rho\sigma_x\sigma_y z+{\sigma_y}^2 z^2\right)}}\left(-1+\rho^2\right){\sigma_x}^2\sigma_y \\
\times\sqrt{-\frac{{\sigma_x}^2-2\rho\sigma_x\sigma_y z+{\sigma_y}^2 z^2}{\left(-1+\rho^2\right){\sigma_x}^2{\sigma_y}^2}} \\
\Bigg) \\
-\sqrt{\pi}(\mu_x\sigma_y(-\rho\sigma_x+\sigma_y z)+\mu_y\sigma_x(\sigma_x-\rho\sigma_y z)) \\
\times\operatorname{erf}\left(\frac{\mu_x\sigma_y(\rho\sigma_x-\sigma_y z)+\mu_y\sigma_x(-\sigma_x+\rho\sigma_y z)}{\sqrt{2}\left(-1+\rho^2\right){\sigma_x}^2{\sigma_y}^2\sqrt{-\frac{{\sigma_x}^2-2\rho\sigma_x\sigma_y z+{\sigma_y}^2 z^2}{\left(-1+\rho^2\right){\sigma_x}^2{\sigma_y}^2}}}\right) \\
\Bigg]\,,
\end{aligned} \tag{6.40}$$

provided that:

$$\left(\mathrm{Re}\left[\frac{\sigma_x^2 - 2\rho\sigma_x\sigma_y z + \sigma_y^2 z^2}{\left(-1+\rho^2\right)\sigma_x^2\sigma_y^2} \right] \leq 0 \quad \wedge \quad \mathrm{Re}\left[\frac{\mu_x\sigma_y(\rho\sigma_x - \sigma_y z) + \mu_y\sigma_x(-\sigma_x + \rho\sigma_y z)}{\left(-1+\rho^2\right)\sigma_x^2\sigma_y^2} \right] < 0 \right) \tag{6.41}$$

$$\vee \quad \mathrm{Re}\left[\frac{\sigma_x^2 - 2\rho\sigma_x\sigma_y z + \sigma_y^2 z^2}{\left(-1+\rho^2\right)\sigma_x^2\sigma_y^2} \right] < 0 \, . \tag{6.42}$$

The conditions in Equations (6.41) and (6.42) ensure that the pull distribution $g(z)$ is real valued and has no singularity. Their validity has always been checked when using Equation (6.40) in the studies presented in this thesis.

The pull distribution $g(z)$ reduces to a standard Gaussian distribution

$$g(z) = \frac{1}{2\pi} \mathrm{e}^{-\frac{z^2}{2}} \, , \quad \text{if } \rho = 0, \ \mu_x = 0, \ \mu_y = \sigma_x \text{ and } \sigma_y \ll \sigma_x \, . \tag{6.43}$$

These conditions represent an unbiased fitted parameter distribution ($\mu_x = 0$) and correctly estimated uncertainties ($\sigma_x = \mu_y$). The mean value of the distribution of a correctly fitted uncertainty is equal to the standard deviation of the fitted parameter distribution. In most cases the standard deviation of the uncertainty distribution will be much smaller than the standard deviation of the fitted parameter distribution ($\sigma_y \ll \sigma_x$). It might be presumed by intuition, that the expected pull distribution has a standard Gaussian shape. As shown above this is true if $\rho = 0$, but as will be shown below this is generally not true if $\rho \neq 0$.

An example for the function $g(z)$, where x and y have a correlation of $\rho = 0.65$ but all other parameters are chosen such that a standard Gaussian distribution is expected, is given in Figure 6.1. It shows a small bias of 0.009 for the theoretically expected mean value of the pull distribution, denoted hereafter by μ_t:

$$\mu_t = \mathrm{E}\left[g(z)\right] \, . \tag{6.44}$$

During the studies in Chapter 7 the typical values of the biases were smaller than 0.01. The bias is considered to be relevant in a study with N pseudo experiments if the bias is of the same size as the uncertainty of the mean value of the pull distribution, which is given by $1/\sqrt{N}$. Thereby, a standard deviation of one for the pull distribution is assumed. Thus, the effect of a correlation between the parameter value and its uncertainty can be neglected if the number of pseudo experiments is less than $\mathcal{O}(10^4)$. However, for a few pull distributions, with $\rho > 0.97$ and σ_y/μ_y around 5 %, a bias of $\mathcal{O}(0.1)$ was observed. For these pull distributions the effect of the correlation can only be neglected in case the number of pseudo experiments is less than $\mathcal{O}(100)$. In summary, a pull distribution with consistent parameter values and uncertainties may not have an expected mean value statistically compatible with zero.

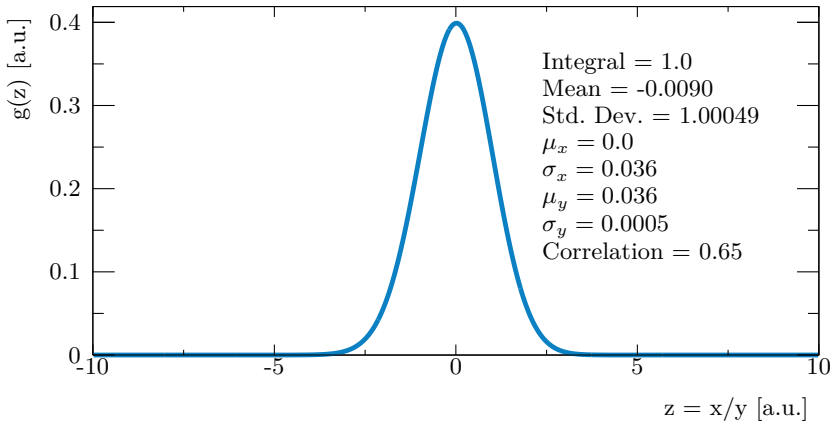


Figure 6.1: Example for the theoretical pull distribution $g(z)$ according to Equation (6.40) with $\mu_x = 0$, $\sigma_x = 0.036$, $\mu_y = 0.036$, $\sigma_y = 0.005$ and correlation coefficient $\rho = 0.65$. From the values next to the curve it can be seen, that the function is properly normalised and more importantly, that a small bias for the mean value is found. The bias of the standard deviation is smaller than the bias of the mean by an order of magnitude.

In Appendix A.1 a validation of Equation (6.40) can be found, as well as a study of the dependence of the mean value and the standard deviation of $g(z)$ on the correlation coefficient.

It needs to be stressed, that in Equation (6.38) an ideal two-dimensional Gaussian distribution is assumed for x and y. For pull distributions of fitted parameters a bias can also arise for other reasons than the correlation, i.e. if the skewness, the third central moment, of the x or y distributions is not compatible with zero (see Section 7.3.1 for an example). If in doubt, whether a bias arises from the correlation between x and y or from some other source, the d_i-distribution needs to be studied. The bias of this distribution is independent of the correlation coefficient.

6.4 Software and Algorithms

In this section the software tools for data analysis and statistical modeling (Sec. 6.4.1) are presented, as well as a technique to minimize rounding errors in computed sums (Sec. 6.4.2).

6.4.1 ROOT, RooFit and RooStats

ROOT [BR97] is an object oriented data analysis framework developed at the European Laboratory for Particle Physics (CERN). It is written in C++ and widely used in HEP. It provides tools for large scale data analysis, such as visualization, histogramming, minimization and fitting. The ROOT file format is used in HEP to store more than 1 EB of physics data [ROO20].

RooFit [VK03] is a C++ software toolkit focused on statistical modeling. It is integrated into ROOT since 2003. Being object oriented, variables and PDFs are implemented as objects. It allows generation of pseudo experiments, fitting of binned and unbinned data sets and provides tools to manage large analyzes. Especially, RooFit takes care of the normalization of the PDFs, performing a numerical integration if an analytical integration is impossible. The version of ROOT used in this thesis is 6.20.04 from April 1st 2020.

RooStats [Mon+10] is a software toolkit built on top of RooFit, providing advanced statistical tools for e.g. the estimation of confidence intervals. In the context of this thesis only the implementation of the sPlot technique by RooStats is used.

6.4.2 Compensated summation

Adding two finite precision floating point numbers, as e.g. specified by the IEEE standard 754 [Ins85], may lead to rounding errors if one number is larger than the other. To add the numbers, they have to be expressed with the same exponent, consequently the smaller number is rounded.

If n numbers x_i are summed up, the rounding errors can accumulate up to a maximum error [Hig93]

$$E_n \leq (n-1)\epsilon \sum_i^n |x_i| + \mathcal{O}\left(\epsilon^2\right) \ , \tag{6.45}$$

where E_n is the difference between the actually computed sum and the sum calculated with infinite precision while ϵ denotes the machine precision, also called unit roundoff or machine epsilon, as defined in Ref. [Gol92].

The Kahan summation [Kah65] is one method to reduce the numerical error due to rounding. This is achieved by storing the rounding error for each addition in a carrier variable and feeding it back into the next addition. For this algorithm the upper bound for E_n is given by [Hig93]

$$E_n \leq \left(2\epsilon + \mathcal{O}\left(n\epsilon^2\right)\right) \sum_i^n |x_i| \ . \tag{6.46}$$

Considering $\mathcal{O}(\epsilon)$ the rounding error is independent of n. This is a significant improvement compared to Equation (6.45). Care must be taken if $\sum_i^n |x_i| >> \sum_i^n x_i$. In these cases the relative error E_n/S_n can be large, where S_n is $|\sum_i^n x_i|$.

Summation occurs in the context of this thesis e.g. for the calculation of mean values or of the asymptotic correction part one (Eq. 6.31). To ensure that the results have a minimal numerical error the Kahan summation is used whenever possible. It will be pointed out where it was used in Chapter 7.

7 Studies of the sPlot technique and of fits to weighted events

The studies conducted in the scope of this thesis are presented in this chapter. They aim to study the sPlot technique and to find the correct approach to obtain the parameter uncertainties of fits to events weighted with sWeights. They are based on pseudo experiments, which are presented in Section 7.1. Some validations are discussed in Section 7.2 prior to presenting the main results in Section 7.3. This is followed by discussions and further investigations of the results (Sec. 7.4). This chapter closes with some conclusions in Section 7.5.

7.1 Description of the studies

First, the statistical model employed by the pseudo experiments is presented in Section 7.1.1, defining the variables and the PDFs. This is followed by a description of the procedure used for a single pseudo experiment, including the sWeight calculation and the various fits performed. Finally, in Section 7.1.3, an explanation, how conclusions are drawn about the three correction methods for the uncertainties in fits to weighted events, is given.

7.1.1 The statistical model

The model used for the studies in this chapter has two species[§], signal and background, and three random variables, called mass, vexex and vgg. The random variable mass, which is the discriminating variable, follows an exponential PDF for background and a Gaussian PDF for signal. The variable vexex follows an exponential PDF for signal as well as for background and the variable vgg follows a Gaussian PDF for both signal and background. The exponential PDF was defined in Equation (6.13) and the Gaussian PDF in Equation (6.16). The parameters of the PDFs in the model are documented in Table 7.1 and typical distributions, found by generating toy data according to the model, are given in Figure 7.1.

The variable mass may represent the mass distribution of a $B_s^0 \to \mu^+\mu^-$ signal on a continuous background in a simplified manner. The variable vexex resembles a simplified model of the

[§]The terms species, control variable and discriminating variable have already been introduced in Section 5.2.2.

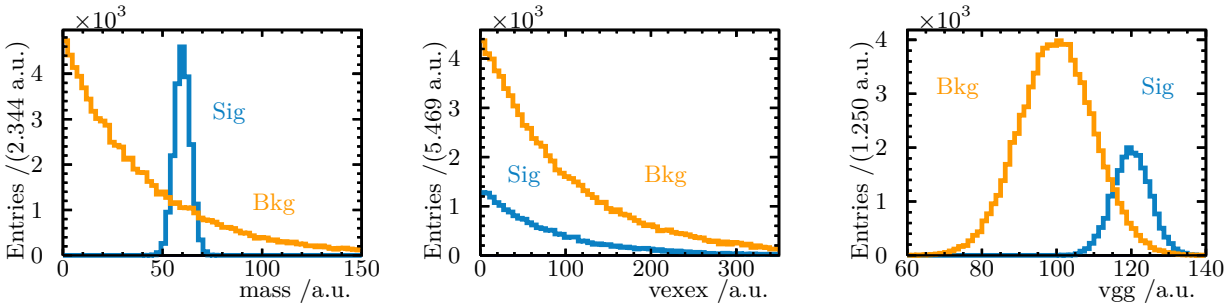


Figure 7.1: Typical distributions for (a) the discriminating variable mass, (b) the control variable vexex and (c) the control variable vgg in a single pseudo experiment. Sig denotes signal and Bkg background.

Table 7.1: PDFs of the model and their parameters. The variable name Mass is abbreviated as m and vexex as vxx to obtain shorter indices.

Variable	Signal PDF	Background PDF	Range / a.u.
mass	Gaussian: $\mu_{\mathrm{m}}^{\mathrm{sig}} = 60\,\mathrm{a.u.}$, $\sigma_{\mathrm{m}}^{\mathrm{sig}} = 4\,\mathrm{a.u.}$	exponential: $\lambda_{\mathrm{m}}^{\mathrm{bkg}} = -0.025\,\frac{1}{\mathrm{a.u.}}$	$[0, 150]$
vexex	exponential: $\lambda_{\mathrm{vxx}}^{\mathrm{sig}} = -0.0125\,\frac{1}{\mathrm{a.u.}}$	exponential: $\lambda_{\mathrm{vxx}}^{\mathrm{bkg}} = -0.01\,\frac{1}{\mathrm{a.u.}}$	$[0, 350]$
vgg	Gaussian: $\mu_{\mathrm{vgg}}^{\mathrm{sig}} = 120\,\mathrm{a.u.}$, $\sigma_{\mathrm{vgg}}^{\mathrm{sig}} = 5\,\mathrm{a.u.}$	Gaussian: $\mu_{\mathrm{vgg}}^{\mathrm{bkg}} = 100\,\mathrm{a.u.}$, $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}} = 10\,\mathrm{a.u.}$	$[60, 140]$

lifetime distribution of the B_s^0. The variable vgg was designed to provide peaking signal and background PDFs in contrast to the vexex variable, inspired by the distribution of one boosted decision tree (BDT) input variable in the $B_{(s)}^0 \to \mu^+\mu^-$ analysis (Ch. 5), namely the χ^2 distance of the muon tracks to any Primary Vertex (PV) [ATL16a, Figure 3b].

7.1.2 Description of a single pseudo experiment

Using the model described above, pseudo experiments are carried out. For each pseudo experiment this involves five major steps:

1. Generating a toy data set,
2. performing the fit to the discriminating variable mass with all parameters floating,
3. performing the fit to the discriminating variable mass with the yield parameters floating and the shape parameters fixed to the values obtained in step 2.,
4. calculating the sWeights based on the discriminating variable fit and
5. performing fits to the control variable signal and background distributions obtained by applying the signal and background sWeights, respectively.

The fit to the discriminating variable with only yield parameters floating (step 3.) will be referred to as yield fit.

The toy data set is generated with a varying number of events. The number of signal and background events is Poisson distributed with a mean of 20 000 events for signal and 80 000 events for background. Details of the event generation and some subtleties are explained in Appendix B.1.

The sWeights are calculated using the SPlot class in RooStats. Care is taken to keep all non-yield parameters of the discriminating PDF constant, which is necessary to obtain correct sWeights (Sec. 6.1.3).

In the last step, the signal and background PDFs of the two control variables are fitted to the corresponding weighted distributions. Each of these fits is performed multiple times with different options for the correction of the uncertainties: squared weights correction, asymptotic correction part one and asymptotic correction part two (Sec. 6.2.6). For the first option the fit as well as the correction is performed using RooFit. For the latter two, the fit is performed in RooFit but the corrections are calculated by a custom implementation of Equation (6.30) and Equation (6.31) due to some limitations of the implementation in RooFit, as detailed in Appendix B.1.

For any fit performed, the standard deviations and the decay constants are corrected for the bias according to Equations (6.15) and (6.22). Any occurrence of a not bias corrected estimator will be marked explicitly. Not applying the bias correction would lead to an expected bias with a significance of $3.35\,\sigma$ for $\sigma_{\mathrm{m}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ in fits to the true distributions. This is derived in Appendix B.2 and shows that the correction for bias is needed.

7.1.3 Procedure to study many pseudo experiments

In order to study fits to events weighted with sWeights, 100 000 pseudo experiments are carried out, referred to as a set of pseudo experiments. The fit results are investigated by means of the parameters of the pull distributions (Sec. 6.3). The focus is on the mean values and the standard deviations of the pull distributions. Higher order moments, e.g. the skewness, are not considered.

To decide whether the estimation of a parameter is unbiased, by looking at the mean value of the corresponding pull distribution, the theoretically expected mean μ_{t} (Eq. 6.44) of the pull distribution needs to be known. As μ_{t} should only account for the influence of the correlation between the fitted parameter and its uncertainty, it is calculated under the assumption, that the fitted parameter distribution is unbiased, i.e. μ_x in Equation (6.40) is set to zero. The parameter σ_x is taken as the standard deviation of the fitted parameter distribution and μ_y (σ_y) as the

mean value (standard deviation) of the fitted parameter uncertainty distribution. Although the d_i-distribution defined by Equation (6.36) offers a more robust method to determine the bias of a fitted parameter distribution, pull distributions are preferred, since they allow to inspect the bias and the correctness of the uncertainties at a glance.

Hereafter a pull distribution is called biased if the absolute value of the difference ($|\Delta_{\mu_t}^{\mu_{\text{pull}}}|$) between the mean value of the fitted parameter pull distribution (μ_{pull}) to μ_t is larger than five times the uncertainty of the fitted parameter pull distribution mean ($\sigma_{\mu_{\text{pull}}}$):

$$|\Delta_{\mu_t}^{\mu_{\text{pull}}}| = |\mu_{\text{pull}} - \mu_t| > 5\,\sigma_{\mu_{\text{pull}}} \; . \tag{7.1}$$

For convenience the signed significance is defined by

$$\mathcal{S}_{\mu_t}^{\mu_{\text{pull}}} = \frac{\Delta_{\mu_t}^{\mu_{\text{pull}}}}{\sigma_{\mu_{\text{pull}}}} \; . \tag{7.2}$$

Any value of $|\mathcal{S}_{\mu_t}^{\mu_{\text{pull}}}|$ below 3 is considered to indicate an unbiased parameter estimation. For values larger than 3, but below 5, no conclusion is drawn. In this case studies with a larger number of pseudo experiments are suggested.

The standard deviation of the fitted parameter pull distribution σ_{pull} is expected to be one, if the uncertainties are estimated correctly. In this case the standard deviation of the fitted parameter distribution is equal to the mean value of the fitted parameter uncertainty distribution. As stated in Section 6.3, a possible correlation of a parameter and its uncertainty will not introduce a significant deviation from one. A fitted parameter uncertainty will be called incorrect if σ_{pull} is not compatible with one, considering a 5σ confidence interval. That is

$$|\mathcal{S}_1^{\sigma_{\text{pull}}}| = |\frac{\Delta_1^{\sigma_{\text{pull}}}}{\sigma_{\sigma_{\text{pull}}}}| = |\frac{\sigma_{\text{pull}} - 1}{\sigma_{\sigma_{\text{pull}}}}| < 5 \; , \tag{7.3}$$

with $\sigma_{\sigma_{\text{pull}}}$ denoting the uncertainty of the parameter σ_{pull}. For values of $\mathcal{S}_1^{\sigma_{\text{pull}}}$ below 5, but larger than 3, no conclusion is drawn. Any value of $\mathcal{S}_1^{\sigma_{\text{pull}}}$ below 3 is considered to indicate correctly estimated uncertainties.

7.2 Validations of the analysis steps

The validations in this section are meant to provide a solid basis for the presentation of the results of the uncertainty correction methods in the next section. First, the event generation and the bias correction is validated, based on fits to the true distributions (Sec. 7.2.1). The parameters of the discriminating variable fit are an input for the sWeights calculation. This fit is validated in Section 7.2.2, followed by a validation of the sWeights calculated (Sec. 7.2.3).

7.2.1 Validations of the event generation and the correctness of the bias correction

The true distributions for each variable and each species are obtained by selecting the events based on their true species recorded at generation time. The fits to the true distributions serve two purposes. First, they should confirm the proper generation of events according to the model described in Table 7.1. Second, they should proof, that the correction of the bias arising for some MLEs in the non-asymptotic case is correct.

Problems with the event generation might be detected by distorted shapes of the fitted parameter distributions, such as reported in Ref. [HHM20a]. No such shape distortions were found in any distribution of a fitted parameter of the model. Two representative pull distributions without shape distortions are shown in Figures 7.2a and 7.2b.

The bias corrections for the fitted parameters given by Equations (6.15) and (6.22) are applied despite the fact, that the PDFs in the model are truncated. In general this might not be correct, as explained in the note on truncated PDFs in Section 6.2.2. But in this model the truncations are of little influence. For example only 3.17×10^{-3} % of the events generated with a non-truncated version of the vgg signal PDF would lie outside the definition interval of the truncated PDF. Ergo it is expected, that these bias corrections are applicable. This is confirmed by the pull distribution mean values in Table 7.2, which are all compatible with the theoretically expected mean. Additionally, two examples of unbiased pull distributions are given in Figure 7.2.

In conclusion, these observations confirm the proper function of the event generation and the correctness of the bias corrections applied to the fitted decay constants and standard deviations.

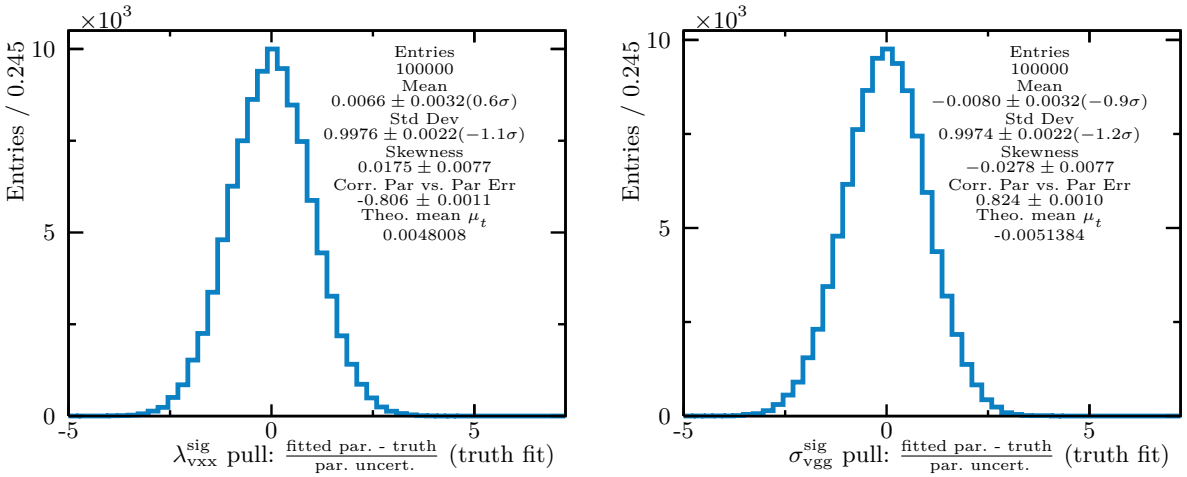


Figure 7.2: Pull distributions for fits to the true distributions, demonstrating that the corrections for bias from (a) Equation (6.15) and (b) from Equation (6.22) yield unbiased parameter estimates. No oddities in the shapes of the distributions are found, indicating proper event generation.
The mean value, the standard deviation and the skewness given in the text field are descriptive statistics values. No fit was used to obtain them. Right after the mean value of the pull distribution the value of $\mathcal{S}_{\mu_t}^{\mu\,\text{pull}}$ is provided in brackets. In the same fashion $\mathcal{S}_1^{\sigma\,\text{pull}}$ is provided. An example how the correlation coefficient between the fitted parameter at its uncertainty is obtained, is found in Appendix B.3. The theoretically expected mean value μ_t, defined by Equation (6.44), is given at the bottom of the text field. It is needed to calculate $\mathcal{S}_{\mu_t}^{\mu\,\text{pull}}$ and describes the bias introduced by correlation between the fitted parameter and its uncertainty. At the end of the axis title the type of fit is given in brackets. In following figures on fits to weighted control variable distributions, the uncertainty correction method applied will be provided as an acronym. All acronyms used are listed in Appendix C.

Table 7.2: Overview of the mean values and the standard deviations of the pull distributions for all the parameters of the model, estimated from fits to the true distributions. For each parameter the pull distribution is unbiased and has a standard deviation compatible with one, confirming proper event generation according to the model. The signed significances $\mathcal{S}_{\mu_t}^{\mu\,\text{pull}}$ and $\mathcal{S}_1^{\sigma\,\text{pull}}$ (Eqs. (7.2) and (7.3)) are provided in brackets.

Parameter	Pull mean	Pull std. dev.
$\lambda_{\text{m}}^{\text{bkg}}$	-0.0005 ± 0.0032 (-0.8σ)	0.9990 ± 0.0022 (-0.5σ)
$\mu_{\text{m}}^{\text{sig}}$	-0.0033 ± 0.0032 (-1.0σ)	0.9986 ± 0.0022 (-0.6σ)
$\sigma_{\text{m}}^{\text{sig}}$	-0.0016 ± 0.0032 $(+1.1\sigma)$	1.0020 ± 0.0022 $(+0.9\sigma)$
$\lambda_{\text{vxx}}^{\text{bkg}}$	$+0.0054 \pm 0.0032$ $(+1.1\sigma)$	0.9992 ± 0.0022 (-0.4σ)
$\lambda_{\text{vxx}}^{\text{sig}}$	$+0.0066 \pm 0.0032$ $(+0.6\sigma)$	0.9976 ± 0.0022 (-1.1σ)
$\mu_{\text{vgg}}^{\text{bkg}}$	-0.0004 ± 0.0032 (-0.1σ)	0.9992 ± 0.0022 (-0.4σ)
$\sigma_{\text{vgg}}^{\text{bkg}}$	$+0.0005 \pm 0.0032$ $(+1.0\sigma)$	0.9971 ± 0.0022 (-1.3σ)
$\mu_{\text{vgg}}^{\text{sig}}$	-0.0056 ± 0.0032 (-1.8σ)	0.9983 ± 0.0022 (-0.8σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0080 ± 0.0032 (-0.9σ)	0.9974 ± 0.0022 (-1.2σ)

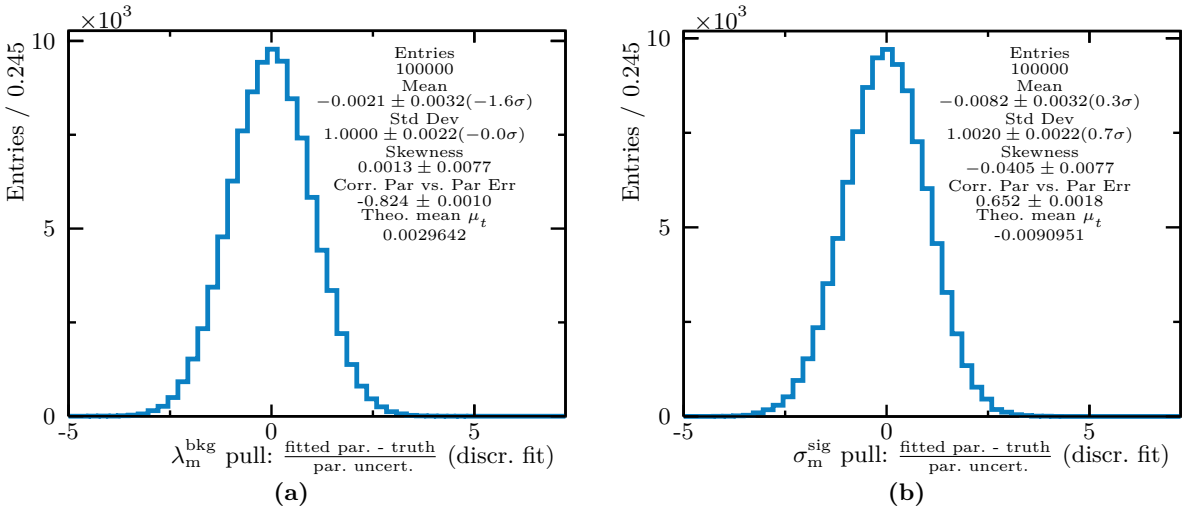


Figure 7.3: Pull distributions for (a) the fitted $\lambda_{\mathrm{m}}^{\mathrm{bkg}}$ and (b) the fitted $\sigma_{\mathrm{m}}^{\mathrm{sig}}$ showing unbiased fit results and correctly estimated parameter uncertainties. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

7.2.2 Validation of the discriminating variable fit

The resulting parameters from the discriminating variable fit are an input for the sWeights calculation. Ergo, it is crucial to ensure that all the shape parameters of the mass PDF are fitted correctly. The pull distributions for these parameters, namely the fitted $\lambda_{\mathrm{m}}^{\mathrm{bkg}}$, $\mu_{\mathrm{m}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{m}}^{\mathrm{sig}}$ were found to have a mean of zero within the uncertainties, proving that there is no bias introduced by the fit. As an example the pull distributions of the fitted $\lambda_{\mathrm{m}}^{\mathrm{bkg}}$ and $\sigma_{\mathrm{m}}^{\mathrm{sig}}$ are given in Figure 7.3.

The asymptotic correction part two depends on the covariance matrix of the discriminating variable fit. The diagonal elements of this matrix are the squared uncertainties of the parameters estimated in the discriminating variable fit. Hence, correct pull distributions for these parameters prove the correctness of the diagonal elements of the covariance matrix. The standard deviations of all pull distributions were compatible with one, showing that the diagonal matrix elements are determined correctly. Two examples are provided in Figure 7.3. The off-diagonal elements have not been checked explicitly.

7.2.3 Validation of the calculated sWeights

Equation (6.6) is exploited to validate the sWeights calculated by RooStats. According to this equation the sum of signal sWeights equals the signal yield determined in the yield fit and the sum of background sWeights equals the background yield determined in the yield fit. For signal and background the difference between the sum of sWeights and the yield from the yield fit is compared. Statistically significant but negligible biases are found. The bias for the

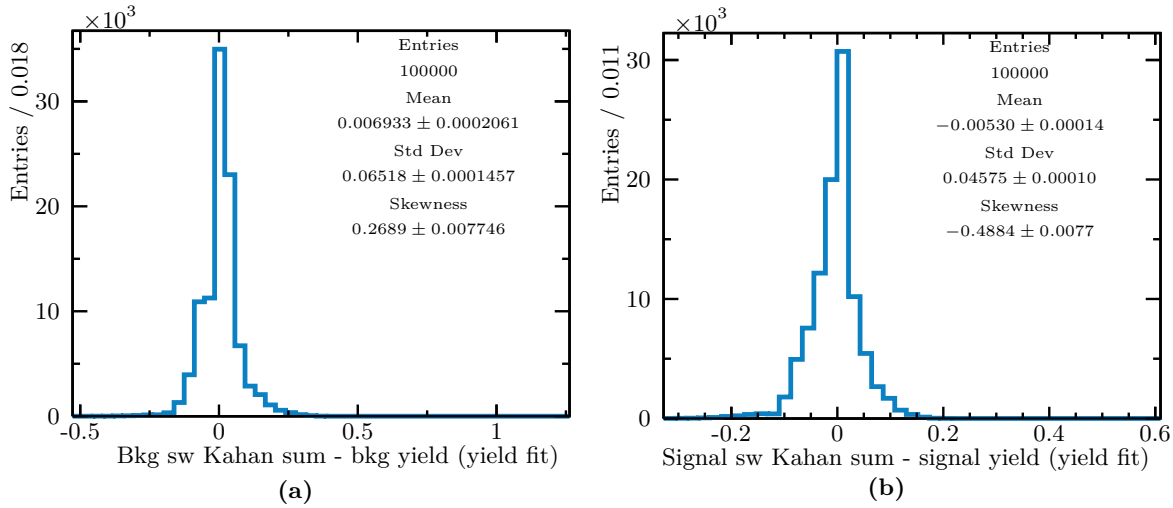


Figure 7.4: Distributions of the difference between the sum of sWeights, and the yield obtained from the yield fit for 100 000 pseudo experiments. Shown (a) for background and (b) for signal. The difference should be zero according to Equation (6.6). Comparing the observed mean values of the differences to the corresponding total event yields, the conclusion is, that the bias is negligible and that the sWeights were calculated properly. The sWeights are calculated by RooStats, and abbreviated as sw in the axis titles. In each pseudo experiment the Kahan summation (Sec. 6.2) was used to calculate the sum of sWeights, in order to reduce numerical instabilities.

sum of background sWeights is 0.0069 ± 0.0002 (Fig. 7.4a). Relating this to the total number of background events, yields a relative bias of 8.6×10^{-5} ‰. For the sum of signal sWeights (Fig. 7.4b) the relative bias is 2.5×10^{-4} ‰. The Kahan summation (Sec. 6.4.2) is used to calculate the sum of the sWeights, excluding numerical instabilities as explanation for the biases observed. Since the relative biases are small, this is not expected to have an influence on the subsequent studies.

Nevertheless, results based on these sWeights, calculated by RooStats, have been cross-checked with the results obtained by using a custom implementation of the sWeights calculation. These custom calculated sWeights do not show a bias for the sum of signal or background sWeights. The standard deviation of the distribution of the difference between the sum of sWeights and the yield from the yield fit is of order 10^{-15} (Figs. B.2a and B.2b). This is much smaller than the values observed for the sWeights calculated by RooStats, which are of order 10^{-2} (Figs. 7.4a and 7.4b). Hence, the custom calculated sWeights fulfill Equation (6.6) more accurately and are thus well suited to be used for cross validations of the sWeights calculated by RooStats. Subsequently the sWeights calculated by RooStats are used unless noted otherwise.

Table 7.3: Comparison of the three uncertainty correction methods used, based on pull distributions. In (a) the mean values of the pull distributions for all control variable parameters are presented and in (b) the standard deviations of the pull distributions. The signed significances $\mathcal{S}_{\mu_t}^{\mu_{\text{pull}}}$ and $\mathcal{S}_1^{\sigma_{\text{pull}}}$ (Eqs. (7.2) and (7.3)) are provided in brackets.

(a)

Parameter	squared weights	asymptotic part one	asymptotic part two
$\lambda_{\text{vxx}}^{\text{bkg}}$	$+0.0065 \pm 0.0032$ $(+1.4\sigma)$	$+0.0065 \pm 0.0032$ $(+1.4\sigma)$	$+0.0065 \pm 0.0032$ $(+1.4\sigma)$
$\lambda_{\text{vxx}}^{\text{sig}}$	$+0.0082 \pm 0.0033$ $(+0.5\sigma)$	$+0.0102 \pm 0.0032$ $(+0.5\sigma)$	$+0.0102 \pm 0.0032$ $(+0.5\sigma)$
$\mu_{\text{vgg}}^{\text{bkg}}$	$+0.0025 \pm 0.0033$ $(+0.8\sigma)$	$+0.0027 \pm 0.0032$ $(+0.8\sigma)$	$+0.0026 \pm 0.0031$ $(+0.8\sigma)$
$\sigma_{\text{vgg}}^{\text{bkg}}$	$+0.0043 \pm 0.0033$ $(+2.0\sigma)$	$+0.0045 \pm 0.0033$ $(+2.0\sigma)$	$+0.0046 \pm 0.0031$ $(+2.0\sigma)$
$\mu_{\text{vgg}}^{\text{sig}}$	$+0.1039 \pm 0.0081$ $(+15.0\sigma)$	-0.0129 ± 0.0032 (-2.4σ)	-0.0125 ± 0.0031 (-2.4σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0791 ± 0.0103 (-7.5σ)	$+0.0344 \pm 0.0032$ (-8.5σ)	$+0.0338 \pm 0.0031$ (-8.6σ)

(b)

Parameter	squared weights	asymptotic part one	asymptotic part two
$\lambda_{\text{vxx}}^{\text{bkg}}$	0.9986 ± 0.0022 (-0.6σ)	0.9996 ± 0.0022 (-0.2σ)	0.9993 ± 0.0022 (-0.3σ)
$\lambda_{\text{vxx}}^{\text{sig}}$	1.0399 ± 0.0023 $(+17.2\sigma)$	1.0005 ± 0.0022 $(+0.2\sigma)$	0.9999 ± 0.0022 (-0.0σ)
$\mu_{\text{vgg}}^{\text{bkg}}$	1.0411 ± 0.0023 $(+17.7\sigma)$	1.0225 ± 0.0023 $(+9.8\sigma)$	0.9915 ± 0.0022 (-3.8σ)
$\sigma_{\text{vgg}}^{\text{bkg}}$	1.0372 ± 0.0023 $(+16.0\sigma)$	1.0291 ± 0.0023 $(+12.6\sigma)$	0.9894 ± 0.0022 (-4.8σ)
$\mu_{\text{vgg}}^{\text{sig}}$	2.5698 ± 0.0057 $(+273.2\sigma)$	1.0239 ± 0.0023 $(+10.4\sigma)$	0.9851 ± 0.0022 (-6.8σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	3.2624 ± 0.0073 $(+310.1\sigma)$	1.0128 ± 0.0023 $(+5.6\sigma)$	0.9824 ± 0.0022 (-8.0σ)

7.3 Results of the corrections of the covariance matrix in presence of weights

In the sections below results for the squared weights correction (Sec. 7.3.1), the asymptotic correction part one (Sec. 7.3.2) and the asymptotic correction part two (Sec. 7.3.3) are shown. These corrections are applied to the uncertainties of the parameters estimated in the control variable fits in the presence of sWeights.

7.3.1 Study of the squared weights correction

In this section the RooFit implementation of the squared weights correction described by Equation (6.33) is studied. It is expected to find incorrect uncertainties since this methods is not in general asymptotically correct. The uncertainties of the fitted parameters are discussed first and then the biases. An overview of the mean values and the standard deviations of the pull distributions for all fitted parameters is given in Tables 7.3a and 7.3b.

In Figure 7.5a the pull distribution obtained for the fitted $\lambda_{\text{vxx}}^{\text{bkg}}$ shows, that for this parameter correct uncertainties were obtained. For the parameters $\lambda_{\text{vxx}}^{\text{sig}}$, $\mu_{\text{vgg}}^{\text{bkg}}$ and $\sigma_{\text{vgg}}^{\text{bkg}}$ (Figs. B.3b to B.3d) the uncertainties are roughly 4 % smaller than the correct value, which is the standard deviation

of the fitted parameter distribution σ_{par}. This was inferred by taking

$$1 - \frac{\mu_{\text{uncert}}}{\sigma_{\text{par}}} \,, \tag{7.4}$$

where μ_{uncert} is the mean value of the parameters uncertainty distribution. This will in most cases be equal to $1 - {}^{1}/_{\sigma_{\text{pull}}}$, where σ_{pull} is the standard deviation of the corresponding pull distribution. For the parameters of the vgg signal Gaussian distribution the deviations are worse. The parameter uncertainties are underestimated by roughly a factor 3.3 for parameter $\sigma_{\text{vgg}}^{\text{sig}}$ (Fig. 7.5b) and a factor 2.6 for parameter $\mu_{\text{vgg}}^{\text{sig}}$ (Fig. 7.5c).

For both $\mu_{\text{vgg}}^{\text{sig}}$ and $\sigma_{\text{vgg}}^{\text{sig}}$ a bias in the pull distribution is observed (Figs. 7.5b and 7.5c). Checking the d_i-distribution of the fitted $\mu_{\text{vgg}}^{\text{sig}}$, which has a mean value of (0.0008 ± 0.0003) a.u., shows that the parameter estimate is unbiased (Fig. 7.5e). The observed bias in the pull distribution arises from the shape of the parameter uncertainty distribution (Fig. 7.5d). This distribution has a strong tail towards smaller values, leading to a skewness of -0.256 ± 0.008. In combination with the negative correlation coefficient of -0.8740 ± 0.0007 between the parameter value and the parameter uncertainty this results in a tail towards higher values in the pull distribution (Fig. 7.5c). This is qualitatively in agreement with the large positive bias of 0.1039 ± 0.0081 observed for the mean value of the pull distribution. Next, the bias of the pull distribution of the fitted $\sigma_{\text{vgg}}^{\text{sig}}$ (Fig. 7.5b) is confirmed by looking at the d_i-distribution (Fig. 7.5f), which has a mean value of (0.0074 ± 0.0009) a.u., confirming the bias observed in the pull distribution with a significance of roughly 8σ. The skewness of the uncertainty distribution for this parameter is (0.055 ± 0.008), thus much smaller than the skewness observed for $\mu_{\text{vgg}}^{\text{sig}}$. The discussion of the bias looking at the d_i-distributions is valid for all correction methods, since the d_i-distributions do only depend on the fitted parameters itself and not on the fitted parameter uncertainties.

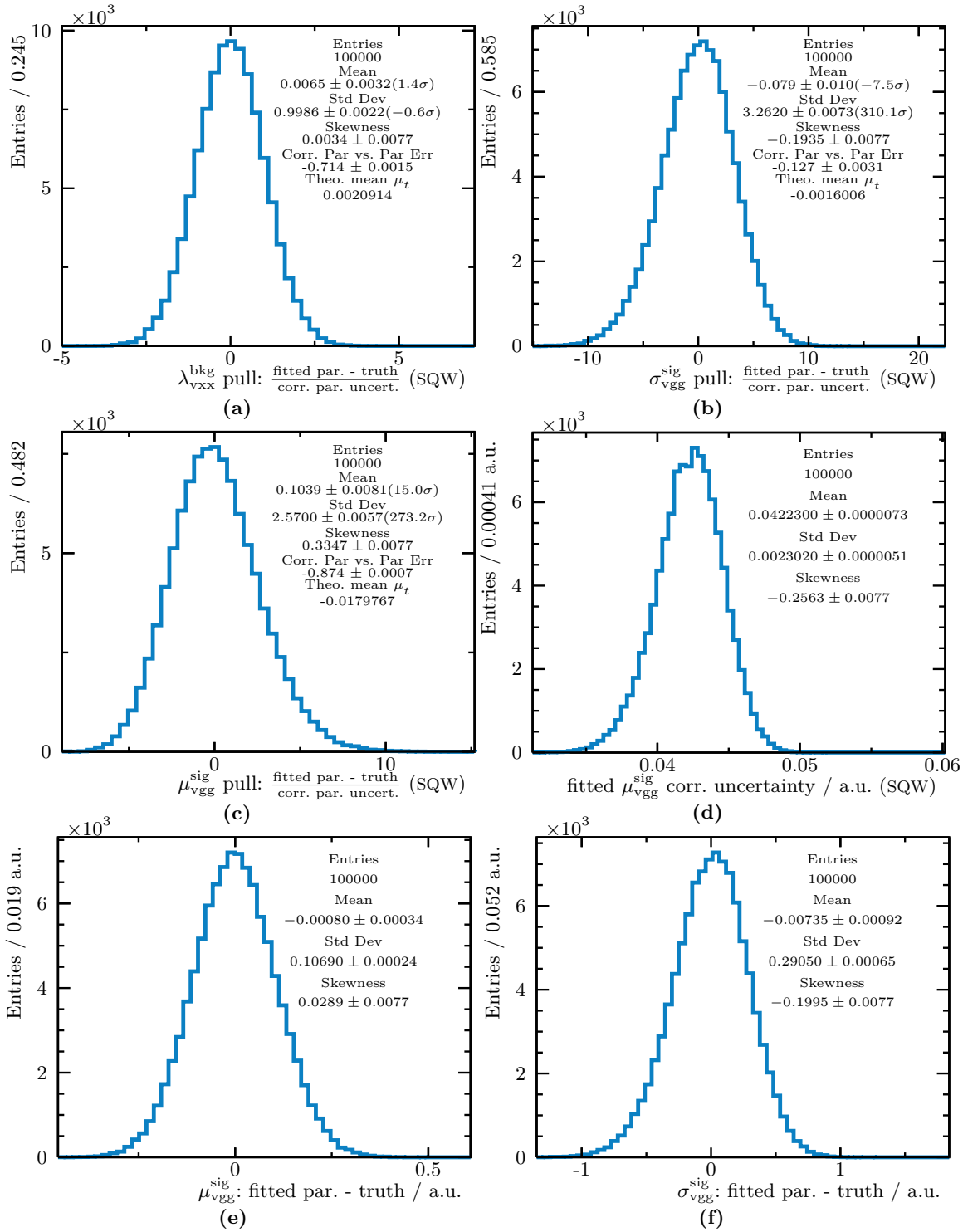


Figure 7.5: Distributions obtained using the squared weights correction (SQW). (a) The only fitted parameter with correct uncertainties and (b) the pull distribution for the fitted parameter with the largest underestimation of the uncertainties. (c) Pull distribution with the largest bias, arising from (d) the shape of the uncertainty distribution for this parameter and not from a real bias in the fitted parameter distribution, which is established by (e) the corresponding d_i-distribution. (f) The d_i-distribution of the fitted parameter σ^{sig}_{vgg} is biased. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

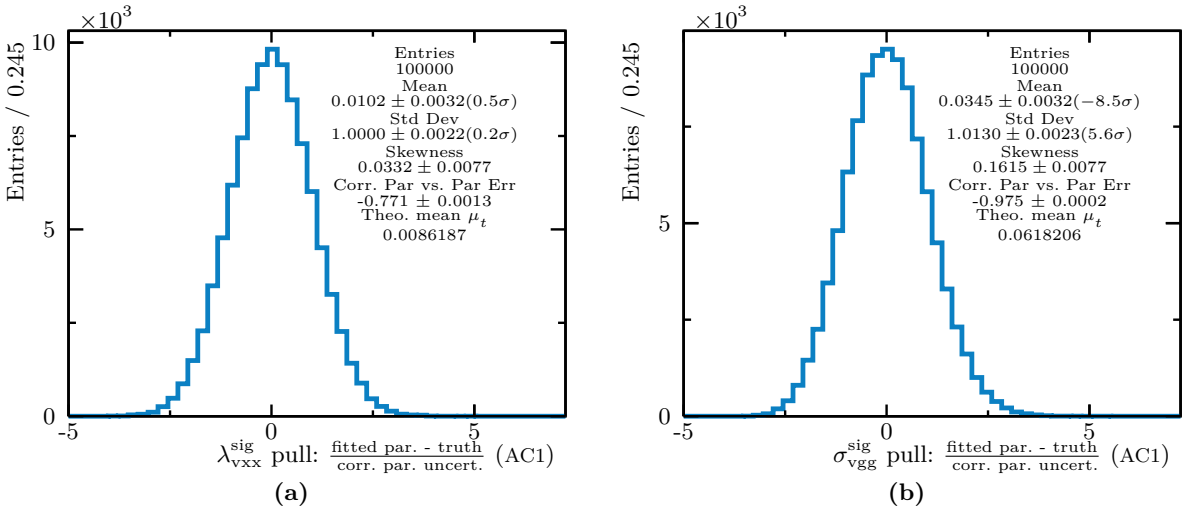


Figure 7.6: Pull distributions obtained using the asymptotic correction part one. (a) Correct uncertainties are obtained for the $\lambda^{\text{sig}}_{\text{vxx}}$. (b) Compared to Figure 7.5b a significant improvement of the uncertainties is found. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

7.3.2 Study of the asymptotic correction part one

In this section a custom implementation of the asymptotic correction part one (Eq. 6.30) is used and not the RooFit implementation, as explained in Appendix B.1. Using the asymptotic correction part one improves the corrected uncertainties significantly, compared to the squared weights correction. An overview of the mean values and the standard deviations of the pull distributions for all fitted parameters is given in Tables 7.3a and 7.3b.

Correct uncertainties are obtained for the decay constants of the vexex signal (Fig. 7.6a) and background (Fig. B.4a) PDF. Using the squared weights correction, this was only the case for the $\lambda^{\text{bkg}}_{\text{vxx}}$ (Fig. 7.5a) but not for the $\lambda^{\text{sig}}_{\text{vxx}}$ (Fig. B.3b). For the parameters of the vgg background PDF the pull distributions have a standard deviation closer to one by a few σ (Table 7.3b). The uncertainties are smaller than the true value by $(2.20 \pm 0.22)\,\%$ and $(2.83 \pm 0.22)\,\%$ for $\mu^{\text{bkg}}_{\text{vgg}}$ and $\sigma^{\text{bkg}}_{\text{vgg}}$, which is calculated employing Formula 7.4.

A significant improvement of the estimated uncertainties is observed for the parameters of the vgg signal distribution. Instead of being underestimated by roughly factor 3, the uncertainties are now underestimated by $(2.33 \pm 0.22)\,\%$ and $(1.26 \pm 0.22)\,\%$ for $\mu^{\text{sig}}_{\text{vgg}}$ and $\sigma^{\text{sig}}_{\text{vgg}}$ (Figs. B.4e and 7.6b). The bias observed for the pull distribution of $\sigma^{\text{sig}}_{\text{vgg}}$ was discussed in the previous section.

7.3.3 Study of the asymptotic correction part two

The asymptotic correction part two is not implemented in RooFit and consequently a custom implementation is used. The results obtained with this uncertainty correction method are compared to the results obtained using the asymptotic correction part one, discussed above. Applying the asymptotic correction part two to the covariance matrices of the fits to the weighted control variable distributions does not alter the correct uncertainties for the parameters of the vexex signal and background PDFs obtained with the asymptotic correction part one but reduces the standard deviation of the pull distributions for the parameters of the vgg signal and background PDFs (Table 7.3b).

The uncertainties of the fitted parameters of the vgg background PDF are improved significantly compared to the asymptotic correction part one. Applying the latter the uncertainties for the fitted $\mu_{\text{vgg}}^{\text{bkg}}$ and $\sigma_{\text{vgg}}^{\text{bkg}}$ were underestimated by $(2.20 \pm 0.22)\,\%$ and $(2.83 \pm 0.22)\,\%$. In contrast, applying the asymptotic correction part two the uncertainties for μ and σ are overestimated by $(0.86 \pm 0.22)\,\%$ and $(1.07 \pm 0.22)\,\%$ (Figs. 7.7a and 7.7b). The standard deviations of the pull distributions for $\mu_{\text{vgg}}^{\text{bkg}}$ and $\sigma_{\text{vgg}}^{\text{bkg}}$ differ from one by $-3.8\,\sigma$ and $-4.8\,\sigma$. Therefore, a study with a higher number of pseudo experiments is needed to decide whether the uncertainties are correct.

For the standard deviations of the pull distributions for the fitted parameters of the vgg signal PDF a significant reduction is observed. This results in significantly overestimated uncertainties instead of underestimated uncertainties, as observed for the asymptotic correction part one. Instead of being underestimated by $(2.33 \pm 0.22)\,\%$ and $(1.26 \pm 0.22)\,\%$ the uncertainties for $\mu_{\text{vgg}}^{\text{sig}}$ and $\sigma_{\text{vgg}}^{\text{sig}}$ are overestimated by $(1.51 \pm 0.23)\,\%$ and $(1.79 \pm 0.24)\,\%$ respectively (Figs. 7.7c and 7.7d), calculated according to Formula 7.4. For these parameters the standard deviations of the pull distributions differ from one by $-6.8\,\sigma$ and $-8.0\,\sigma$, respectively. Thus, the estimated uncertainties are incorrect. The results obtained for the parameters of the vgg signal PDF are the main subject of the subsequent studies.

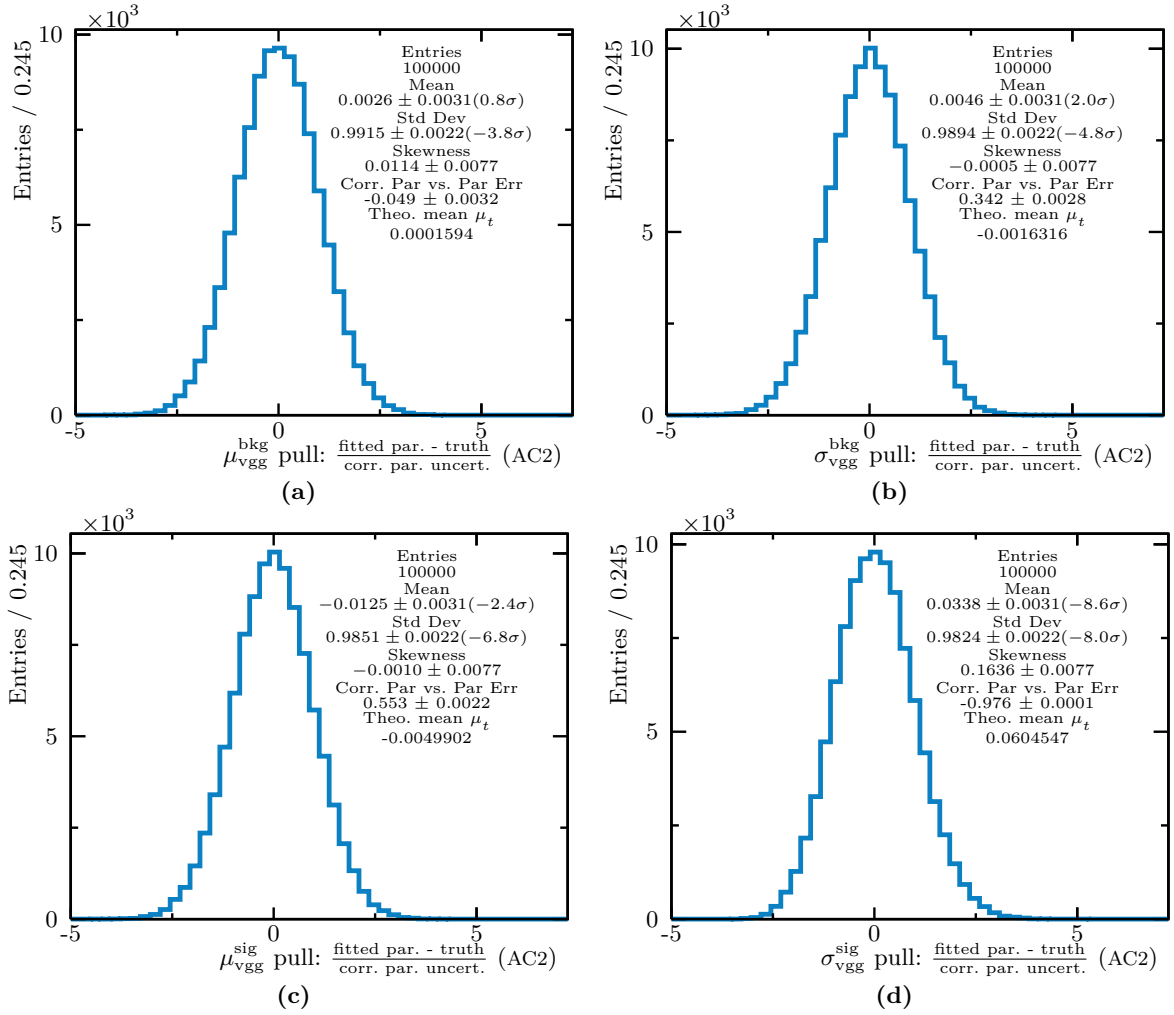


Figure 7.7: Pull distributions obtained using the asymptotic correction part two. (a) For the parameter $\mu_{\text{vgg}}^{\text{bkg}}$ and (b) the parameter $\sigma_{\text{vgg}}^{\text{bkg}}$ no conclusion about the correctness of the uncertainties can be drawn. The pull distributions of (c) the fitted parameter $\mu_{\text{vgg}}^{\text{sig}}$ and (b) the fitted parameter $\sigma_{\text{vgg}}^{\text{sig}}$ show, that the uncertainties of these parameters are significantly overestimated. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

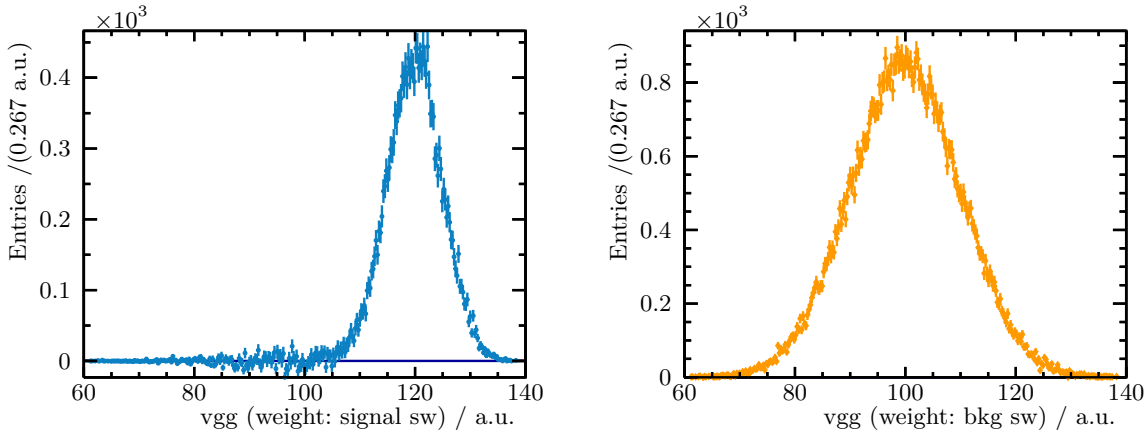


Figure 7.8: Distribution of the generated values of the vgg variable for a single pseudo experiment (a) weighted with signal sWeights and (b) weighted with background sWeights. The pseudo experiment was chosen randomly.

7.4 Discussion of the results

The uncertainties obtained for the fitted parameters $\lambda_{\text{vxx}}^{\text{bkg}}$ and $\lambda_{\text{vxx}}^{\text{sig}}$ using the asymptotic correction part one or asymptotic correction part two were correct and need no further discussion. However, for the fitted parameters of the vgg background PDF, namely $\mu_{\text{vgg}}^{\text{bkg}}$ and $\sigma_{\text{vgg}}^{\text{bkg}}$, it is not clear if their uncertainties obtained by the asymptotic correction part two are correct. To clarify this situation additional studies with a higher number of pseudo experiments are needed. For the fitted $\mu_{\text{vgg}}^{\text{sig}}$ the uncertainties are incorrect and for $\sigma_{\text{vgg}}^{\text{sig}}$ the fit result is biased and the uncertainties are incorrect. Thus, the following discussion mainly concerns the fit results obtained for the parameters $\mu_{\text{vgg}}^{\text{sig}}$ and $\sigma_{\text{vgg}}^{\text{sig}}$.

To guide the discussion, the following hypothesis is proposed here:

Hypothesis 1 (H1): *The bias and the incorrect uncertainties observed for σ_{vgg}^{sig} and the incorrect uncertainties for μ_{vgg}^{sig} are caused by contributions from background events in regions, where the vgg signal PDF reaches values close to zero.*

This hypothesis is motivated by Figure 7.8a. In this figure a histogram with the distribution of the vgg variable values, weighted with signal sWeights for a randomly-chosen single pseudo experiment is presented. On the left side of the signal peak, below roughly 102.5 a.u., fluctuations introduced by the background events are observed. For comparison, the distribution of the vgg variable values weighted with the background sWeights is given in Figure 7.8b. It is important not to draw premature conclusions from the binned distributions shown in Figures 7.8a and 7.8b about the ML fits, since they are unbinned.

7.4.1 Causes of the problems observed that can be excluded

The incorrect uncertainties observed for the fitted $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$, as well as the bias observed for $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ do not arise from the event generation. This was shown in Section 7.2.1. There it was also shown, that the bias correction according to Equation (6.22) is applicable. Since this bias correction is applied for the fitted $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$, the observed bias can not arise from a biased estimator in the non-asymptotic case. Furthermore, the significance of the expected bias in the non-asymptotic case without applying the bias correction is -3.35σ. This is much smaller than the observed bias with a significance of $-8\,\sigma$.

It has been argued in Section 7.2.3, that the biases observed in Figures 7.4a and 7.4b for the difference between the number of events obtained from the yield fit and the sum of sWeights have no influence on the results of the control variable fits. To ascertain this, the control variable fits to weighted events were repeated using the sWeights calculated with a custom implementation of the sWeights calculation as weights. The results obtained for the mean values and the standard deviations of the pull distributions are close to the results obtained using the sWeights calculated by RooStats. The differences observed are much smaller than $1\,\sigma$. An elaborate discussion of this topic is found in Appendix B.4.

On account of some sWeights being negative, it could be suspected, that the minimization of the NLL may have been unsuccessful: e.g. by running into local minima. To clarify this situation the parameters of the vgg signal and bkg PDFs were estimated calculating the minimum of the NLL, defined by Equation (6.10), directly. Based on this equation the estimator of the parameter μ of a non-truncated Gaussian PDF is defined by the mean value of the measurements or by the weighted mean (Eq. B.12) in case weights are present. For the parameter σ of a non-truncated Gaussian PDF the not-bias-corrected estimator is the weighted standard deviation (Eq. B.13). Given a truncated Gaussian PDF, estimating the values of μ and σ is more involved. An equation system for $\hat{\mu}$ and $\hat{\sigma}$ must be solved. This was done numerically for all 100 000 pseudo experiments (Appendix B.7). The distribution of the differences between the fitted parameter values of $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ and the estimated parameter values obtained by direct calculation are given in Figures 7.9a and 7.9b. It is not expected that these differences are always zero, since the minimization performed by Minuit is stopped once the convergence criterion is satisfied. This happens if the Estimated Distance to Minimum (EDM), which is Minuit's estimate of the vertical distance to the minimum of the NLL, is below some defined threshold. The bias of both distributions presented in Figures 7.9a and 7.9b is statistically significant and of order 1×10^{-5} a.u. or smaller. In comparison, the bias observed for the fitted $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ compared to the true value of this parameter is (-0.0074 ± 0.0009) a.u. (Fig. 7.5f) and thus two orders of magnitude larger. Consequently, the bias observed for the fitted $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ compared to its true value does not arise from problems in the minimization process. The biases observed for the distribution of the differences between the estimated parameter values obtained by direct calculation and those

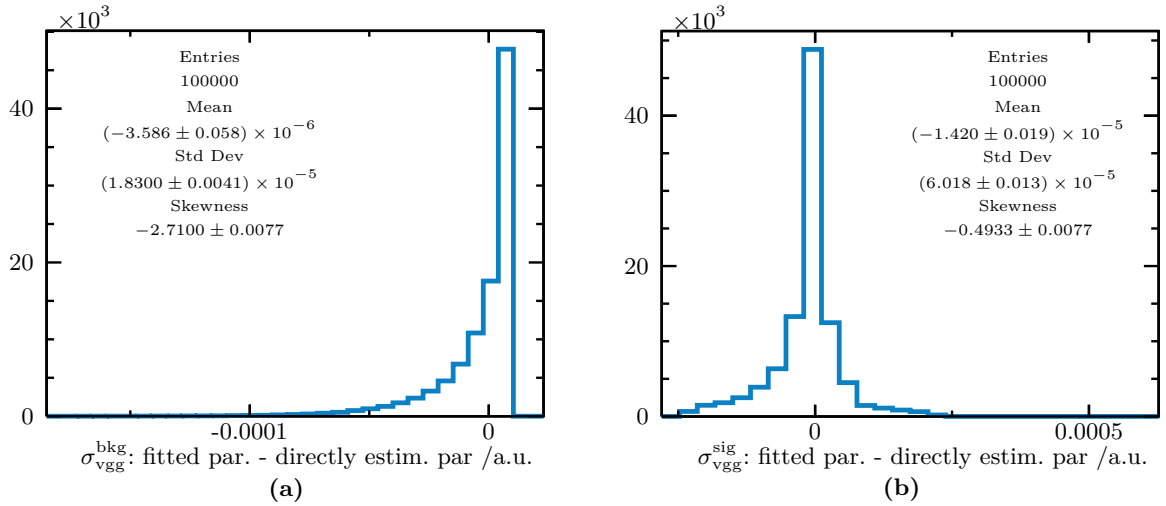


Figure 7.9: Distributions of the differences between fitted parameter values and the directly estimated parameter values obtained by solving the equation system for the estimators of a truncated Gaussian distribution (Eq. B.7) for (a) the fitted parameter $\sigma_{\text{vgg}}^{\text{bkg}}$ and (b) the fitted parameter $\sigma_{\text{vgg}}^{\text{sig}}$. A negligible bias compared to the bias observed for the difference between the fitted value of $\sigma_{\text{vgg}}^{\text{sig}}$ and the true parameter value (Fig. 7.5f) is found.

obtained from the control variable fit for $\mu_{\text{vgg}}^{\text{bkg}}$ and $\mu_{\text{vgg}}^{\text{sig}}$ are even smaller than those observed for $\sigma_{\text{vgg}}^{\text{bkg}}$ and $\sigma_{\text{vgg}}^{\text{sig}}$ (Figs. B.7a and B.7b).

7.4.2 Studies with restricted fit ranges applied

If Hypothesis 1 is true, excluding the regions in the vgg signal fit, in which the values of the vgg signal PDF are close to zero and where the vgg background PDF is dominant, is expected to improve the uncertainties and to reduce the bias of the fitted parameter $\sigma_{\text{vgg}}^{\text{sig}}$. Thus, three different restricted fit ranges were chosen for the vgg signal PDF: [100 a.u., 140 a.u.], [105 a.u., 140 a.u.] and [110 a.u., 140 a.u.]. This corresponds to cuts at 2, 3 and 4 $\sigma_{\text{vgg}}^{\text{sig}}$ on the left-hand side of the signal peak.

Using any of the three restricted fit ranges for the vgg signal PDF leads to an unbiased pull distribution for the fitted $\sigma_{\text{vgg}}^{\text{sig}}$, which is found in any of the Tables 7.4a to 7.4c. However, a bias with a signed significance of −3.3 a.u. is found for the fitted $\mu_{\text{vgg}}^{\text{sig}}$ when using the [110 a.u., 140 a.u.] range. To clarify whether this is a statistical fluctuation or a real bias a higher number of pseudo experiments would be needed.

Using the squared weights correction to correct the uncertainties of the fitted $\mu_{\text{vgg}}^{\text{sig}}$ and $\sigma_{\text{vgg}}^{\text{sig}}$ in combination with any of the three restricted fit ranges for the vgg signal PDF still results in incorrect uncertainties (Table 7.4a).

Applying the asymptotic correction part one or the asymptotic correction part two to correct the uncertainties of the fitted $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$, correct uncertainties are obtained for the [105 a.u., 140 a.u.] fit range and the [110 a.u., 140 a.u.] fit range (Tables 7.4b and 7.4c). For the [100 a.u., 140 a.u.] no decision is made on the correctness of the uncertainties as the $|\mathcal{S}_1^{\sigma_{\mathrm{pull}}}|$ is roughly three. This deviation is observed for both the asymptotic correction part one and the asymptotic correction part two (Tables 7.4b and 7.4c).

To investigate, whether excluding the signal region when fitting the vgg background PDF to the correspondingly weighted data has an influence on the fit results obtained for $\mu_{\mathrm{vgg}}^{\mathrm{bkg}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}}$, two different fit ranges for the vgg background PDF are chosen: [60 a.u., 120 a.u.] and [60 a.u., 130 a.u.]. This corresponds to cuts at 2 and 3 $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}}$ on the right side of the vgg background distribution.

Correct uncertainties are retrieved for the fitted parameters $\mu_{\mathrm{vgg}}^{\mathrm{bkg}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}}$ by applying the [60 a.u., 120 a.u.] fit range in combination with the asymptotic correction part one (Table 7.4b). However, using the asymptotic correction part one in combination with the [60 a.u., 130 a.u.] fit range incorrect uncertainties are found (Table 7.4b). Next, no decision is made on the correctness of the uncertainties obtained when applying the asymptotic correction part two to correct the uncertainties of the fitted parameters and any of the vgg background PDF fit ranges (Table 7.4c)

In summary, using the [105 a.u., 140 a.u.] or the [110 a.u., 140 a.u.] fit range yields unbiased parameter estimates with correct uncertainties for the parameters $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$, when applying either the asymptotic correction part one or the asymptotic correction part two to correct the uncertainties. Furthermore, correct uncertainties are obtained for $\mu_{\mathrm{vgg}}^{\mathrm{bkg}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}}$ using the [60 a.u., 130 a.u.] fit range in combination with the asymptotic correction part one. These findings support Hypothesis 1.

Table 7.4: Mean values and standard deviations of the pull distributions for the fitted parameters of the PDFs of the vgg variable when using restricted fit ranges in the fits to the distributions weighted by the sPlot technique. The uncertainties of the fitted parameters were corrected by the three different methods: (a) squared weights correction, (b) asymptotic correction part one and (c) asymptotic correction part two. The signed significances $\mathcal{S}_{\mu_t}^{\mu_{\text{pull}}}$ and $\mathcal{S}_1^{\sigma_{\text{pull}}}$ (Eqs. (7.2) and (7.3)) are provided in brackets.

(a) Squared weights correction

Fit range / a.u.	Parameter	Pull mean	Pull std. dev.
$[100, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	$+0.0019 \pm 0.0045$ (-1.9σ)	1.4153 ± 0.0032 $(+131.2\sigma)$
$[100, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0131 ± 0.0048 (-2.4σ)	1.5309 ± 0.0034 $(+155.1\sigma)$
$[105, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0054 ± 0.0037 (-2.5σ)	1.1589 ± 0.0026 $(+61.3\sigma)$
$[105, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0089 ± 0.0035 (-0.8σ)	1.1126 ± 0.0025 $(+45.3\sigma)$
$[110, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0051 ± 0.0032 (-3.3σ)	1.0243 ± 0.0023 $(+10.6\sigma)$
$[110, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0105 ± 0.0031 (-0.0σ)	0.9711 ± 0.0022 (-13.3σ)
$[60, 120]$	$\mu_{\text{vgg}}^{\text{bkg}}$	-0.0016 ± 0.0032 $(+0.1\sigma)$	1.0105 ± 0.0023 $(+4.6\sigma)$
$[60, 120]$	$\sigma_{\text{vgg}}^{\text{bkg}}$	-0.0030 ± 0.0032 $(+0.7\sigma)$	0.9963 ± 0.0022 (-1.7σ)
$[60, 130]$	$\mu_{\text{vgg}}^{\text{bkg}}$	$+0.0021 \pm 0.0033$ $(+0.7\sigma)$	1.0421 ± 0.0023 $(+18.1\sigma)$
$[60, 130]$	$\sigma_{\text{vgg}}^{\text{bkg}}$	$+0.0033 \pm 0.0033$ $(+2.0\sigma)$	1.0360 ± 0.0023 $(+15.6\sigma)$

(b) Asymptotic correction part one

Fit range / a.u.	Parameter	Pull mean	Pull std. dev.
$[100, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0068 ± 0.0032 (-2.0σ)	1.0064 ± 0.0023 $(+2.8\sigma)$
$[100, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	$+0.0045 \pm 0.0032$ (-2.4σ)	1.0070 ± 0.0023 $(+3.1\sigma)$
$[105, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0070 ± 0.0032 (-2.5σ)	1.0038 ± 0.0022 $(+1.7\sigma)$
$[105, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0037 ± 0.0032 (-0.8σ)	1.0036 ± 0.0022 $(+1.6\sigma)$
$[110, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0059 ± 0.0032 (-3.3σ)	1.0020 ± 0.0022 $(+0.9\sigma)$
$[110, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0101 ± 0.0032 (-0.0σ)	1.0007 ± 0.0022 $(+0.3\sigma)$
$[60, 120]$	$\mu_{\text{vgg}}^{\text{bkg}}$	-0.0015 ± 0.0032 $(+0.2\sigma)$	1.0034 ± 0.0022 $(+1.5\sigma)$
$[60, 120]$	$\sigma_{\text{vgg}}^{\text{bkg}}$	-0.0029 ± 0.0032 $(+0.7\sigma)$	1.0012 ± 0.0022 $(+0.5\sigma)$
$[60, 130]$	$\mu_{\text{vgg}}^{\text{bkg}}$	$+0.0022 \pm 0.0032$ $(+0.7\sigma)$	1.0233 ± 0.0023 $(+10.2\sigma)$
$[60, 130]$	$\sigma_{\text{vgg}}^{\text{bkg}}$	$+0.0035 \pm 0.0033$ $(+2.0\sigma)$	1.0297 ± 0.0023 $(+12.9\sigma)$

(c) Asymptotic correction part two

Fit range / a.u.	Parameter	Pull mean	Pull std. dev.
$[100, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0068 ± 0.0031 (-2.0σ)	0.9935 ± 0.0022 (-2.9σ)
$[100, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	$+0.0046 \pm 0.0031$ (-2.4σ)	0.9919 ± 0.0022 (-3.6σ)
$[105, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0070 ± 0.0032 (-2.5σ)	0.9994 ± 0.0022 (-0.3σ)
$[105, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0036 ± 0.0032 (-0.8σ)	0.9986 ± 0.0022 (-0.6σ)
$[110, 140]$	$\mu_{\text{vgg}}^{\text{sig}}$	-0.0059 ± 0.0032 (-3.3σ)	1.0012 ± 0.0022 $(+0.5\sigma)$
$[110, 140]$	$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0100 ± 0.0032 (-0.0σ)	1.0002 ± 0.0022 $(+0.1\sigma)$
$[60, 120]$	$\mu_{\text{vgg}}^{\text{bkg}}$	-0.0015 ± 0.0031 $(+0.2\sigma)$	0.9935 ± 0.0022 (-2.9σ)
$[60, 120]$	$\sigma_{\text{vgg}}^{\text{bkg}}$	-0.0028 ± 0.0031 $(+0.7\sigma)$	0.9931 ± 0.0022 (-3.1σ)
$[60, 130]$	$\mu_{\text{vgg}}^{\text{bkg}}$	$+0.0021 \pm 0.0031$ $(+0.7\sigma)$	0.9919 ± 0.0022 (-3.7σ)
$[60, 130]$	$\sigma_{\text{vgg}}^{\text{bkg}}$	$+0.0036 \pm 0.0031$ $(+2.0\sigma)$	0.9901 ± 0.0022 (-4.5σ)

7.4.3 Studies with a modified statistical model

If Hypothesis 1 is true, it should be possible to tune the statistical model defined by Table 7.1 in such a way, that incorrect uncertainties and a potential bias are observed for other parameters than $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$. The modification chosen here is setting $\lambda_{\mathrm{vxx}}^{\mathrm{sig}} = -0.1\,\frac{1}{\mathrm{a.u.}}$ instead of $-0.0125\,\frac{1}{\mathrm{a.u.}}$. This causes the vexex signal PDF to fall quickly and thus a region is artificially created, where the signal PDF has values close to zero and the background PDF is dominant, similar to the case of the vgg signal PDF. Therefore, it is expected to observe incorrect uncertainties for the fitted $\lambda_{\mathrm{vxx}}^{\mathrm{sig}}$ parameter and a potential bias.

Alternatively, it should be possible to obtain unbiased pull distributions for $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ and improved uncertainties for $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$ and $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ by reducing the range in which the values of the vgg signal PDF are close to zero. This is achieved by choosing a wider $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ of 10 a.u. To keep the relative truncation of the vgg signal PDF the same, the right endpoint of the definition interval was increased from 140 to 160 a.u. Thus the vgg signal PDF is truncated at four times the value of $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ above the mean value $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$.

Another set of 100 000 pseudo experiments was generated with these settings. The following discussion of the results obtained based on this set of pseudo experiments primarily focuses on the asymptotic correction part two, since this should be the most accurate correction method for the uncertainties. Additional material for this discussion is presented in Appendix B.6.

The pull distribution for the fitted $\lambda_{\mathrm{vxx}}^{\mathrm{sig}}$ using the asymptotic correction part two based on the set of pseudo experiments with modified parameters is presented in Figure 7.10a. The uncertainties of the fitted $\lambda_{\mathrm{vxx}}^{\mathrm{sig}}$ are underestimated with a signed significance of $-5.2\,\sigma$ (Fig. 7.10a). The bias of the pull distribution has a signed significance of $-14.4\,\sigma$, which is sustained by the corresponding d_i-distribution (Fig. B.6a).

In comparison to Figure 7.5b it is apparent from Figure 7.10b, that the bias of the fitted $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ disappears in case a larger parameter value for $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ is chosen in the statistical model. The absence of this bias is confirmed by the corresponding d_i-distribution (Fig. B.6b). The standard deviations of the pull distribution for the fitted $\mu_{\mathrm{vgg}}^{\mathrm{sig}}$ and the fitted $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$ parameter values (Table B.3) do not allow a conclusion on the correctness of the uncertainties. But compared to the incorrect uncertainties found for the default parameter values of the statistical model (Table 7.3b) an improvement is clearly observed.

These additional observations also support Hypothesis 1.

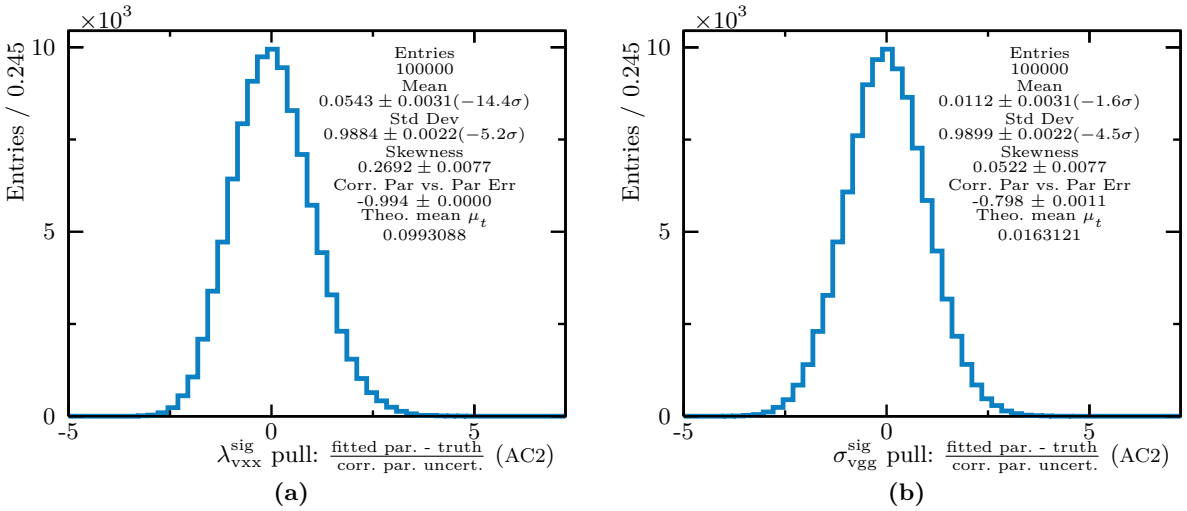


Figure 7.10: Pull distributions for 100 000 pseudo experiments based on the modified statistical model (Table B.2). (a) Pull distribution for the modified parameter $\lambda^{\text{sig}}_{\text{vxx}}$ with a significant bias and incorrect uncertainties. (b) Pull distribution for the modified parameter $\sigma^{\text{sig}}_{\text{vgg}}$ with an unbiased mean value. It is not clear, whether the uncertainties are correct. These figures may be compared to the nominal setup (Figs. B.5b and B.5f). The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

7.5 Conclusions

The aim of the studies presented in this thesis was to see, whether the sPlot technique can be used in the ATLAS $B^0_{(s)} \to \mu^+\mu^-$ analysis. The sideband subtraction technique, used up to now in this analysis (Sec. 5.2.2), requires the specification of a signal region. The sPlot technique does not, which is an advantage. Since no issues with the calculated sWeights have been found, it is suggested to use it in the $B^0_{(s)} \to \mu^+\mu^-$ analysis for data-MC sample comparisons, after performing a direct comparison of the sideband subtraction technique and the sPlot technique for relevant distributions in a future study.

In order to achieve the results presented here, unbinned ML fits to distributions of events weighted by the sPlot technique and the correction of the covariance matrix resulting from these fits had to be investigated. As expected, applying the squared weights correction does not yield correct uncertainties. Correct uncertainties are obtained for the decay constants of the exponential distributions using the asymptotic correction part one or the asymptotic correction part two. This is a proof of principle that either the asymptotic correction part one or the asymptotic correction part two could be used in a measurement of e.g. the B^0_s lifetime, in case an unbinned ML fit is employed. Prior to their application a study with an adequate number events per pseudo experiment and a realistic model of the lifetime distribution in question should be performed. If the asymptotic correction part one and the asymptotic correction part two provide correct uncertainties in this case too, the asymptotic correction part one should be chosen as it is less CPU intensive and less complicated to implement.

For the fitted parameters of the vgg signal PDF and the fitted parameters of the vgg background PDF correct uncertainties were not obtained by using the asymptotic correction part one or the asymptotic correction part two and the full vgg range as fit range. For the distribution of the fitted $\sigma_{\text{vgg}}^{\text{sig}}$ even a bias was observed. All studies presented in Section 7.4 supported Hypothesis 1 and none of them falsified this hypothesis. Therefore, this hypothesis can be generalized:

Hypothesis 2 (H2): *If for a species the PDF of a control variable v reaches values significantly greater than zero only in a small range compared to the full range of that variable, then the events from the other species outside this small range might introduce a bias to the fitted parameter distribution. The corrected uncertainties obtained with the asymptotic correction part one or the asymptotic correction part two might not be correct. This effect is the more pronounced, the smaller the yield of the control variable v is compared to the yields of the other species.*

When performing an unbinned ML fit with unrestricted fit range to events weighted by the sPlot technique, where the PDF fulfills the criteria mentioned in Hypothesis 2, it should be investigated whether a bias or incorrect uncertainties appear.

8 Summary and outlook

In this thesis, the sPlot technique and unbinned maximum likelihood fits to events weighted by this method have been studied (Ch. 7). The presence of the sWeights in the fits necessitates a correction of the covariance matrix as discussed in Section 6.2.6. Besides the correction proposed by Eadie et al. [EDJ71], referred to as squared weights correction, the corrections proposed by Langenbruch [Lan19], referred to as asymptotic correction part one and asymptotic correction part two, have been examined. In contrast to the asymptotic correction part one, the asymptotic correction part two does not only account for the presence of the sWeights but also for their uncertainties, which arise from the discriminating variable fit.

A simple statistical model composed of three variables is employed for the pseudo experiments. The discriminating variable PDF, consisting of an exponential background PDF and a Gaussian signal PDF, is inspired by a typical invariant mass distribution of decaying B-mesons. One control variable, named vexex, is made up of two exponential distributions for signal and background, the other one, named vgg, of two Gaussian distributions. The variable vexex resembles a simplified model of the lifetime distribution of the B_s^0. The variable vgg is designed to provide peaking signal and background PDFs in contrast to the more continuous vexex variable.

Aiming for statistically significant results, 100 000 pseudo experiments have been performed. For each pseudo experiment the numbers of events were drawn from Poisson distributions with a mean of 20 000 events for signal and 80 000 events for background. The main steps in a single pseudo experiment are: generating a data set, performing the discriminating variable fit, performing the yield fit, calculating the sWeights and performing fits to the signal and background distributions, obtained by applying the corresponding sWeights. All fits in this procedure were carried out using RooFit, which internally calls the function minimization tool Minuit [JR75].

Various validations of the analysis steps have been performed (Sec. 7.2). By fits to the true distributions, the correctness of the event generation has been shown. To check the sWeight calculation, the difference between the sum of sWeights and the fitted number of events from the yield fit has been inspected and shown to have a negligible bias.

Most conclusions are drawn from pull plots of the fitted parameters and their uncertainties. An account of the effects of a correlation between the fitted parameter and its uncertainty on the pull distribution was given in Section 6.3. These effects are taken into account in the interpretation of the pull distributions.

The squared weights correction is shown to be insufficient. Only for the decay constant of the vexex background PDF, correct uncertainties are obtained. The uncertainties of all other control variable PDF parameters are underestimated.

In comparison to the squared weights correction, applying the asymptotic correction part one improves the parameter uncertainties significantly. Nevertheless, the uncertainties of the parameters of the Gaussian PDFs are still underestimated. Applying the asymptotic correction part two results in correct uncertainties for the parameters of the vgg background PDF. The uncertainties of the parameters of the vgg signal PDF are still problematic. However, by using restricted instead of unrestricted fit ranges, correct parameter uncertainties are achieved. Furthermore, the bias of the fitted standard deviation of the vgg signal PDF, observed for the unrestricted fit range, vanishes. Supported by further studies, this leads to the conclusion, that the incorrect uncertainties of especially the parameters of the signal PDF were caused by contributions from background events in the regions where the signal PDF reaches values close to zero. In general it is suspected, that, if the PDF of a species reaches values significantly greater than zero only in a small region, the events of the other species outside this region might introduce incorrect uncertainties or a bias, in case an unrestricted fit range is used.

In summary, the studies in this thesis present a step towards the application of the sPlot technique in the ATLAS $B^0_{(s)} \to \mu^+\mu^-$analysis. Furthermore, the lifetime of the B^0_s-meson could be measured in a future analysis utilizing the sPlot technique to extract the signal lifetime, as done by CMS and LHCb [CMS20; LHC17]. In case unbinned maximum likelihood fits are employed, the asymptotic correction part two should be applied.

A

Appendix to Chapter 6

A.1 Studies of the theoretical pull distribution

A pull distribution, as defined in Section 6.3, can be described by the ratio $z = x/y$ of two random variables x and y following a two-dimensional Gaussian PDF. The resulting PDF $g(z)$ is given by Equation (6.40). For some examples Equation (6.40) is validated against a dedicated pseudo experiment, such as the one presented in Figure A.2. Besides this validation, the influence of the correlation coefficient ρ on the mean value of $g(z)$, denoted as μ_t, and on the standard deviation of $g(z)$, denoted as σ_t, is investigated. The parameter values for μ_x, σ_x, μ_y and σ_y are taken from the results of the fit to the vgg signal truth distribution. The mean value and the standard deviation of the distribution of the fitted vgg signal σ are taken as μ_x and σ_x. The values for μ_y and σ_y are taken from the mean value and the standard deviation of the distribution of the uncertainty of the fitted vgg signal σ. The pull distribution for vgg singal σ is given in Figure 7.2b. The correlation coefficient was varied from $-0.999\,99$ to $0.999\,99$ in ten steps.

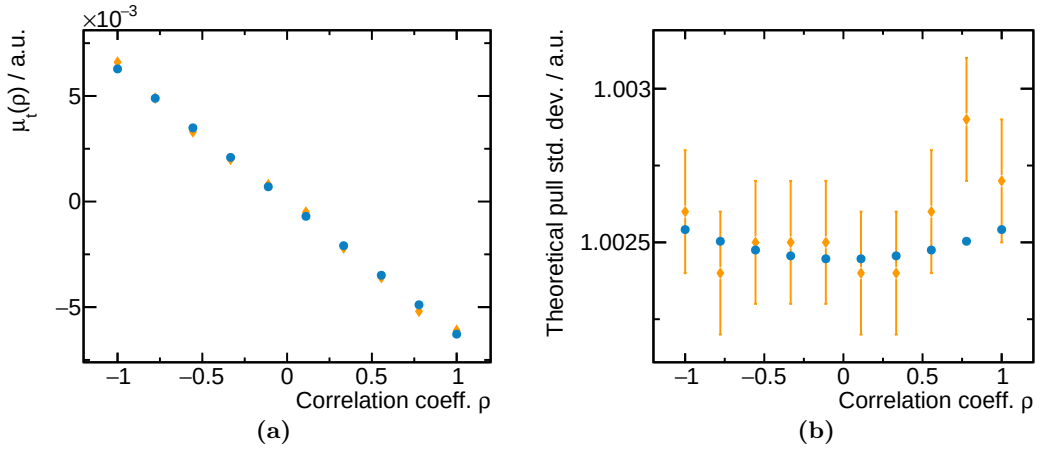


Figure A.1: Mean value (a) and standard deviation (b) of a pull distribution with $\mu_x = 0$, $\sigma_x = 0.02512$, $\mu_y = 0.02506$ and $\sigma_y = 0.0001571$, comparing the values obtained from Equation (6.40) (blue dots) to values obtained by a pseudo experiment (orange diamonds), such as the one presented in Figure A.2. Each of these pseudo experiments contained 2×10^6 events.

For this specific example, a good agreement between the mean values calculated from Equation (6.40) and from pseudo experiments is observed (Fig. A.1a). The same holds for the standard deviation of $g(z)$ (Fig. A.1b). These figures show also a stronger effect of ρ on μ_t than on σ_t. The difference between the minimum and the maximum of μ_t across the entire ρ range is 10^{-2}. For σ_t it is 9×10^{-5}, thus providing an example, where the influence of ρ on σ_t can be neglected. For this example a roughly linear dependence of μ_t on ρ is observed.

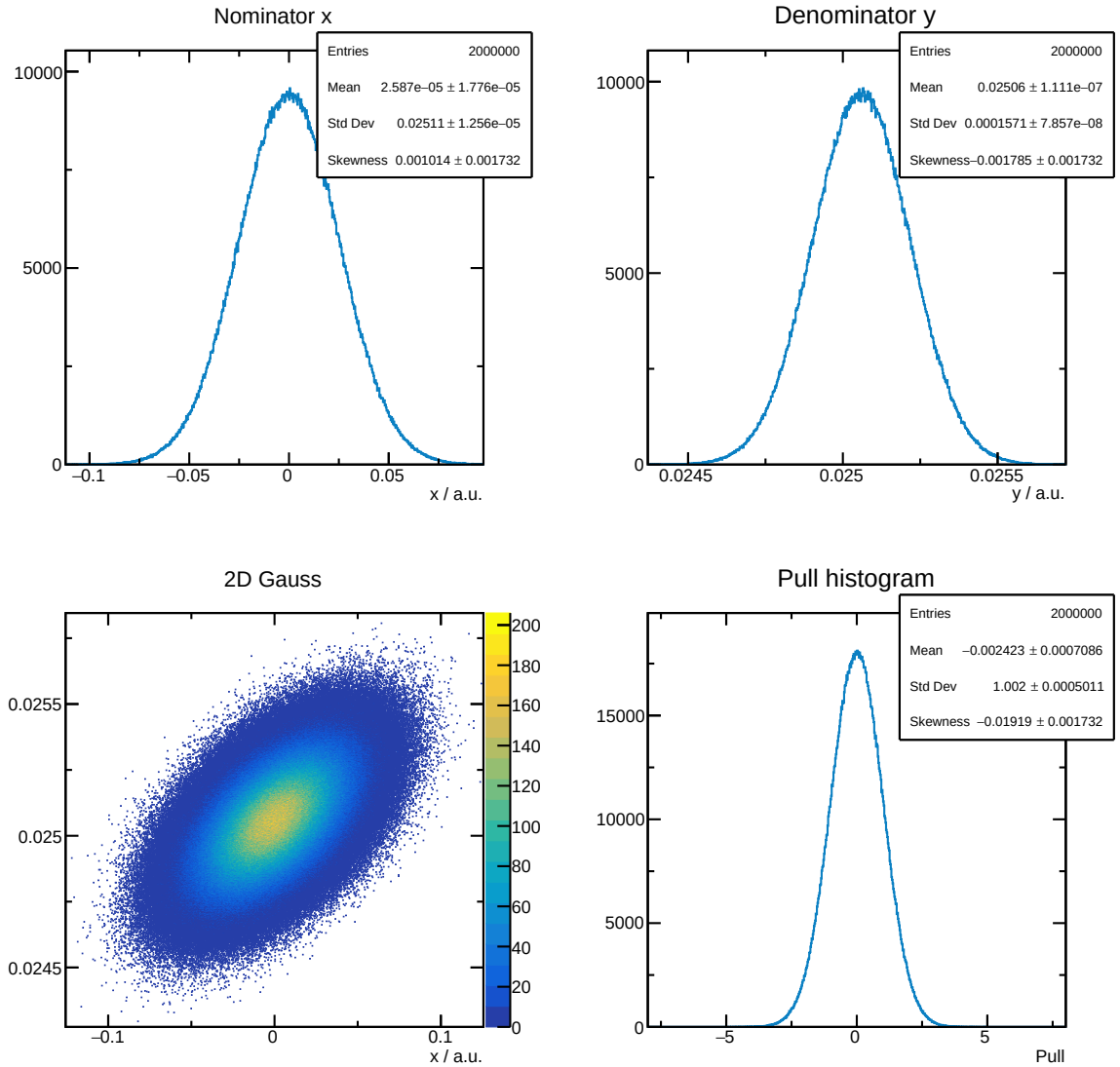


Figure A.2: Example of a pseudo experiment to obtain the pull distribution $g(z)$ on a statistical basis. 2×10^6 values for x as well as for y were generated. The upper left plot shows the distribution of the nominator x, the upper right plot the distribution of the denominator y. The lower left plot shows the two two-dimensional Gaussian distribution of x and y, demonstrating the correlation coefficient. The lower right plot shows the resulting pull distribution. The true parameters are $\mu_x = 0$, $\sigma_x = 0.02512$, $\mu_y = 0.02506$, $\sigma_y = 0.0001571$ and $\rho = 0.55$.

B

Appendix to Chapter 7

B.1 Details of a single pseudo experiment and limitations of RooFit

A rough description of a single pseudo experiment has been given in Section 7.1.2. Below some of the steps are presented in more detail.

Generating a toy data set in RooFit can require a reconfiguration of the internal number generator. Usually, toy data is generated in RooFit by calling RooAbsPdf::generate() on the PDF set up by the user. Generating a toy data set with a Poisson distributed number of events is achieved by passing the flag Extended to this method. Doing so for the model described in Section 7.1.1 set up in RooFit, while using the default number generator settings, leads to generated toy data sets where the phase space spanned by the three variables mass, vexex and vgg is not fully covered [HHM20b]. To avoid this issue the numbers of events are generated by the TRandom3 class of ROOT. The toy data sets for signal and background are then produced by providing the numbers of events to generate to the RooAbsPdf::generate() call. After labeling the events according to their true origin as signal or background the two data sets are merged and shuffled.

All fits are performed with RooFit using the RooAbsPdf::fitTo() method. By default this method uses of the function minimization tool Minuit [JR75]. The function minimizer was not changed for the studies presented here. The version of ROOT used in this thesis is 6.20.04 from April 1st 2020 the DOI in [BR97] is specific for this version. The squared weights correction and the asymptotic correction part one are implemented in the RooAbsPdf::fitTo() method. They are enabled by the flags SumW2Error and AsymptoticError, respectively. The asymptotic correction part one was introduced into ROOT in release 6.20 dated April 2020. Using this method revealed, that it can not be used in conjunction with the Extended flag and that the usage of the AsymptoticError flag requires setting limits on the parameters. These observations were reported to the ROOT-Forum [HHM20b; HM20; HHM20c]. Since parameter limits in Minuit may lead to incorrectly estimated parameter uncertainties [JR75], it is preferred to avoid these. Therefore, the asymptotic correction part one was implemented in the code used by the studies presented in this thesis. Numerical instabilities are avoided by using analytical derivatives only. Furthermore, the asymptotic correction part two was implemented, which is not avail-

able in ROOT as of release 6.22. The derivatives required for the asymptotic correction part two were calculated by the computer algebra software SageMath [The20], using especially the component SymPy [Meu+17]. The latter was also used to generate the corresponding C++ code automatically. The corrections implemented use the fit result obtained by RooAbsPdf::fitTo().

B.2 Calculation of the expected bias of a distribution of an uncorrected standard deviation

In Equation (6.16) the correction for the bias of the standard deviation σ of a Gaussian PDF has been given. To know whether this correction is needed, when looking at many pseudo experiments, the expected bias of the fitted parameter distribution in case of no correction is estimated below.

Given a MLE $\hat{\sigma}$ for a parameter σ_{true} of a Gaussian PDF, the following relation holds approximately (compare Equation (6.22)):

$$\mathrm{E}\left[\hat{\sigma}\right] = \sigma_{\text{true}} \sqrt{\frac{N_{\text{ev}} - 1.5}{N_{\text{ev}}}} \,, \tag{B.1}$$

where N_{ev} is the number of measurements available to calculate $\hat{\sigma}$ and $\mathrm{E}\left[\sigma_{\text{corr}}\right]$ was substituted by σ_{true}. Therefore, the expectation value of the bias b is calculated by

$$\mathrm{E}\left[b\right] = \sigma_{\text{true}} \sqrt{\frac{N_{\text{ev}} - 1.5}{N_{\text{ev}}}} - \sigma_{\text{true}} \,, \tag{B.2}$$

The expected significance of the bias s_b is defined as the expected bias divided by the expected uncertainty of the bias. Given N_{exp} pseudo experiments the bias is calculated as the difference between σ_{true} and the mean value of the distribution of $\hat{\sigma}$. Consequently, the uncertainty of the bias is given by the uncertainty of the mean value of the distribution of $\hat{\sigma}$. This mean value is denoted as $\bar{\hat{\sigma}}$ and its uncertainty is given by

$$\sigma_{\bar{\hat{\sigma}}} = \frac{\sigma_{\text{distr}}}{\sqrt{N_{\text{exp}}}} \,, \tag{B.3}$$

where σ_{distr} is the standard deviation of the distribution of $\hat{\sigma}$. The expectation value of σ_{distr} is equal to the uncertainty of $\hat{\sigma}$. The uncertainty for a standard deviation, such as $\hat{\sigma}$ is in first order given by [AF03]

$$\sigma_{\hat{\sigma}} = \frac{\hat{\sigma}}{2\sqrt{N_{\text{ev}} - 1}} \quad \Rightarrow \quad \mathrm{E}\left[\sigma_{\bar{\hat{\sigma}}}\right] = \frac{\mathrm{E}\left[\hat{\sigma}\right]}{2\sqrt{N_{\text{ev}} - 1}} \,. \tag{B.4}$$

Therefore, the expected significance of the bias is

$$s_b = \frac{\mathrm{E}\left[b\right]}{\mathrm{E}\left[\sigma_{\bar{\hat{\sigma}}}\right]} = 2\sqrt{N_{\text{exp}}}\sqrt{N_{\text{ev}} - 1}\left(1 - \sqrt{\frac{N_{\text{ev}}}{N_{\text{ev}} - 1.5}}\right) \,. \tag{B.5}$$

Considering $N_{\text{ev}} = 20\,000$ for the number of signal events used and $N_{\text{exp}} = 100\,000$ for the number of pseudo experiments conducted, a bias with a significance of -3.35σ is expected. For $N_{\text{ev}} = 80\,000$, the number of background events, a bias with a significance of -1.68σ is expected. This leads to the conclusion, that a bias correction is needed.

B.3 Example of obtaining the correlation coefficient between a fitted parameter and its uncertainty

Considering pseudo experiment studies where a data set is repeatedly generated and fitted, the correlation coefficient between the fitted parameter λ_i^{fit} and its uncertainty $\sigma_{\lambda_i^{\text{fit}}}$ is defined by:

$$\rho = \frac{\sum_{i=0}^{N_{\text{exp}}} \sum_{j=0}^{N_{\text{exp}}} \left(\lambda_i^{\text{fit}} - \overline{\lambda^{\text{fit}}}\right) \left(\sigma_{\lambda_j^{\text{fit}}} - \overline{\sigma_{\lambda^{\text{fit}}}}\right)}{\sum_{i=0}^{N_{\text{exp}}} \left(\lambda_i^{\text{fit}} - \overline{\lambda^{\text{fit}}}\right)^2 \sum_{j=0}^{N_{\text{exp}}} \left(\sigma_{\lambda_j^{\text{fit}}} - \overline{\sigma_{\lambda^{\text{fit}}}}\right)^2}, \tag{B.6}$$

where N_{exp} is the number of pseudo experiments produced and the bar indicates the arithmetic mean. The uncertainty of the correlation coefficient is calculated by [Bow28]

$$\sigma_\rho = \frac{\left(1 - \rho^2\right)}{\sqrt{N_{\text{exp}}}}, \tag{B.7}$$

which might not be correct, if the correlation coefficient is very close to unity or the two-dimensional distribution of the fitted parameter and its uncertainty is not a two-dimensional Gaussian. An example of the correlation between a fitted parameter and its uncertainty is provided in Figure B.1.

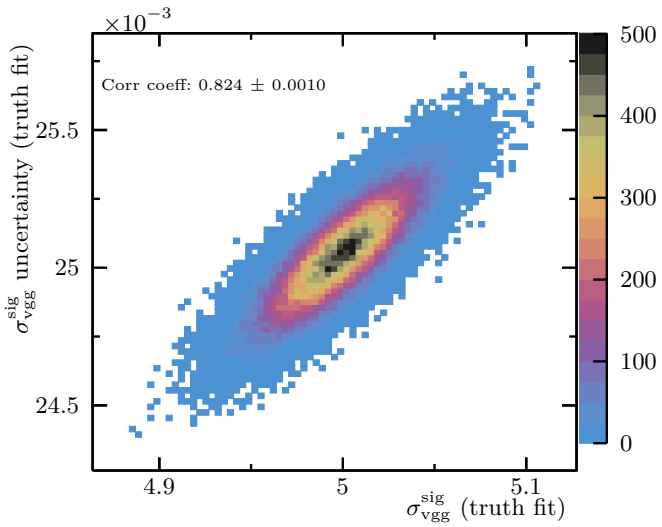


Figure B.1: Two-dimensional histogram of the parameter $\sigma^{\text{sig}}_{\text{vgg}}$ fitted to the true distribution and its uncertainty for 100 000 pseudo experiments, illustrating the correlation between the fitted parameter and its uncertainty. The correlation coefficient and its uncertainty provided in the text field are obtained using Equations (B.6) and (B.7), respectively. The corresponding pull distribution is found in Figure 7.2b.

B.4 sWeights calculated by a custom implementation of the sWeights calculation

The custom implementation of the sWeights calculation according to Equation (6.4) uses the analytical expression for the covariance matrix of the yield fit defined by Equation (6.5). In contrast to the sWeights calculated by RooStats (Figs. 7.4a and 7.4b) the sWeights calculated by the custom implementation show no bias when looking at the difference between the sum of signal or background sWeights and the corresponding yield from the yield fit (Figs. B.2a and B.2b). The fluctuations observed arise from the precision of the double-precision floating point numbers used. According to IEEE 754 there are 53 effective bits of precision for the mantissa [Ins85]. This results in a precision of 15 to 16 digits. Considering that the number of signal and background events is of order 10^4, their precision is of order 10^{-12}. The fluctuations observed in Figures B.2a and B.2b are of this order.

In addition, the control variable fits to weighted events were additionally performed using the sWeights calculated with a custom implementation of the sWeights calculation. The mean values and standard deviations of the pull distributions obtained using either the asymptotic correction part one or the asymptotic correction part two are given in Tables B.1a and B.1b. The results are close to the results obtained using the sWeights calculated by RooStats, which are used by default. The differences observed are much smaller than $1\,\sigma$ (compare Table 7.3 and Table B.1).

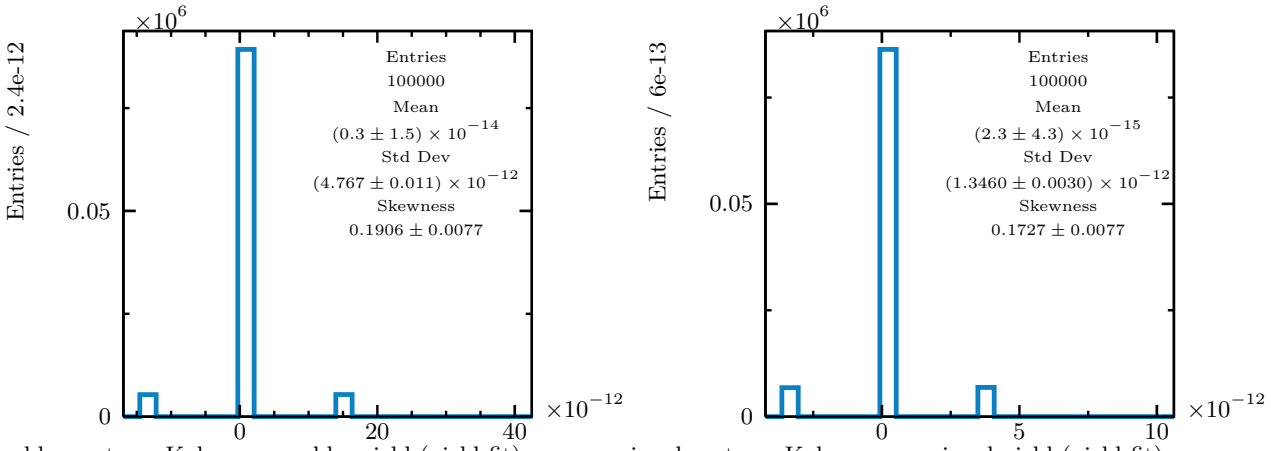


Figure B.2: Distribution of the differences between the sum of sWeights calculated by the custom implementation of the sWeights calculation and the yield obtained from the yield fit for 100 000 pseudo experiments. Shown for (a) background and (b)for signal. The differences are expected to be zero according to Equation (6.6). Good agreement with this expectation is found. The fluctuations observed arise from the finite numerical precision of the double-precision floating point numbers used. The sWeights calculated by the custom implementation of the sWeights calculation are abbreviated as cust. sw in the axis titles. In each pseudo experiment the Kahan summation (Sec. 6.4.2) was used to calculate the sum of sWeights, in order to reduce numerical instabilities.

Table B.1: (a) Mean values and (b) standard deviations of the pull distributions for 100 000 pseudo experiments obtained from the control variable fits to weighted events, using the sWeights calculated by a custom implementation of the sWeights calculation. The uncertainties were corrected using either the asymptotic correction part one or the asymptotic correction part two. The signed significances $\mathcal{S}_{\mu_t}^{\mu_{\text{pull}}}$ and $\mathcal{S}_1^{\sigma_{\text{pull}}}$ (Eqs. (7.2) and (7.3)) are provided in brackets.

(a)

Parameter	asymptotic part one	asymptotic part two
$\lambda_{\text{vxx}}^{\text{bkg}}$	$+0.0065 \pm 0.0032$ $(+1.4\sigma)$	$+0.0065 \pm 0.0032$ $(+1.4\sigma)$
$\lambda_{\text{vxx}}^{\text{sig}}$	$+0.0102 \pm 0.0032$ $(+0.5\sigma)$	$+0.0102 \pm 0.0032$ $(+0.5\sigma)$
$\mu_{\text{vgg}}^{\text{bkg}}$	$+0.0027 \pm 0.0032$ $(+0.8\sigma)$	$+0.0026 \pm 0.0031$ $(+0.8\sigma)$
$\sigma_{\text{vgg}}^{\text{bkg}}$	$+0.0044 \pm 0.0033$ $(+2.0\sigma)$	$+0.0045 \pm 0.0031$ $(+2.0\sigma)$
$\mu_{\text{vgg}}^{\text{sig}}$	-0.0129 ± 0.0032 (-2.4σ)	-0.0125 ± 0.0031 (-2.4σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	$+0.0345 \pm 0.0032$ (-8.5σ)	$+0.0338 \pm 0.0031$ (-8.6σ)

(b)

Parameter	asymptotic part one	asymptotic part two
$\lambda_{\text{vxx}}^{\text{bkg}}$	0.9996 ± 0.0022 (-0.2σ)	0.9993 ± 0.0022 (-0.3σ)
$\lambda_{\text{vxx}}^{\text{sig}}$	1.0005 ± 0.0022 $(+0.2\sigma)$	0.9999 ± 0.0022 (-0.0σ)
$\mu_{\text{vgg}}^{\text{bkg}}$	1.0225 ± 0.0023 $(+9.8\sigma)$	0.9915 ± 0.0022 (-3.8σ)
$\sigma_{\text{vgg}}^{\text{bkg}}$	1.0291 ± 0.0023 $(+12.6\sigma)$	0.9894 ± 0.0022 (-4.8σ)
$\mu_{\text{vgg}}^{\text{sig}}$	1.0239 ± 0.0023 $(+10.4\sigma)$	0.9851 ± 0.0022 (-6.8σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	1.0128 ± 0.0023 $(+5.6\sigma)$	0.9824 ± 0.0022 (-8.0σ)

B.5 Overview of the pull distributions for all three uncertainty correction methods

An overview of the pull distributions for all parameters of the control variable PDFs is shown for the squared weights correction in Figure B.3, for the asymptotic correction part one in Figure B.4 and for the asymptotic correction part two in Figure B.5.

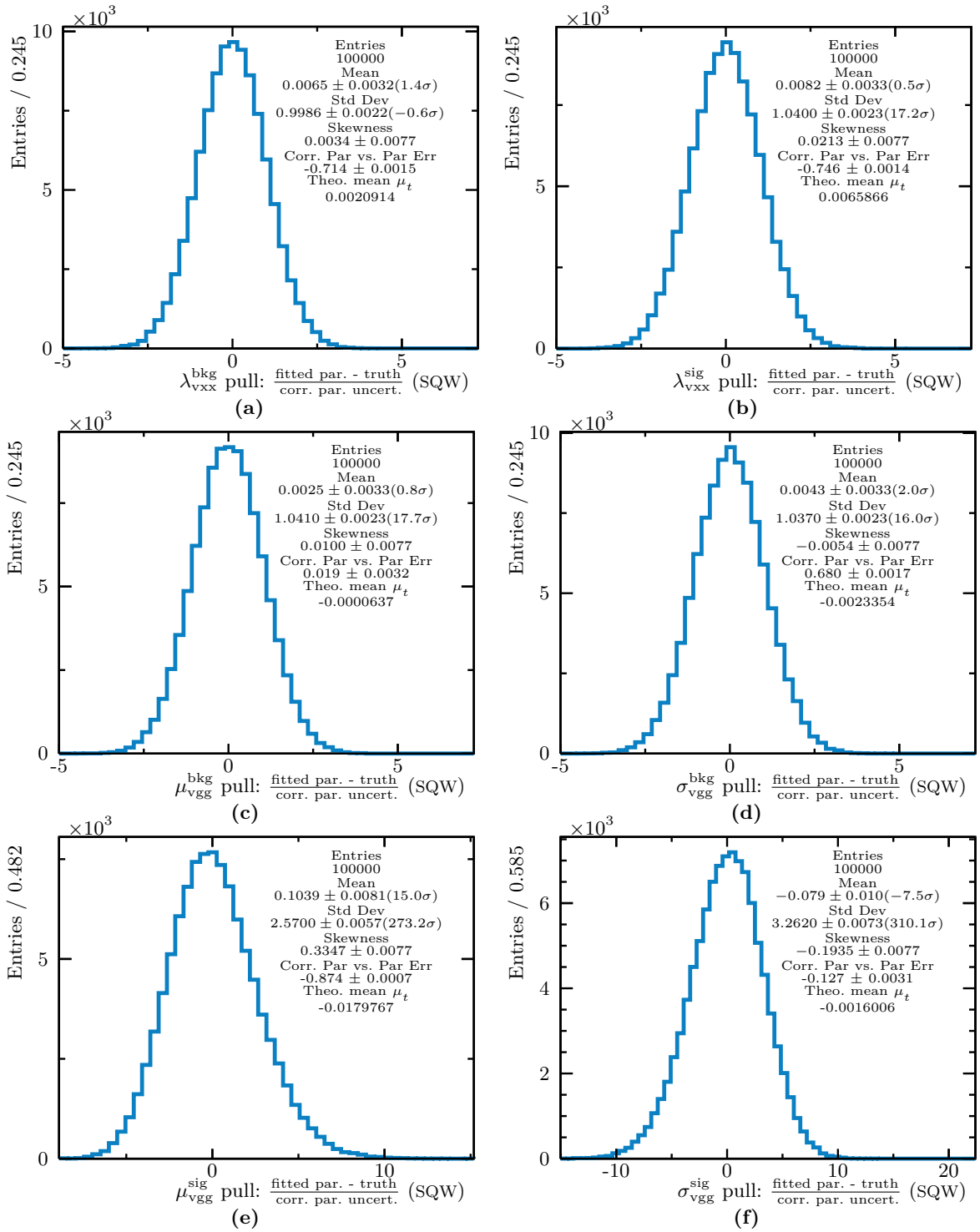


Figure B.3: Overview of the pull distributions for all parameters of the control variable PDFs using the squared weights correction. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

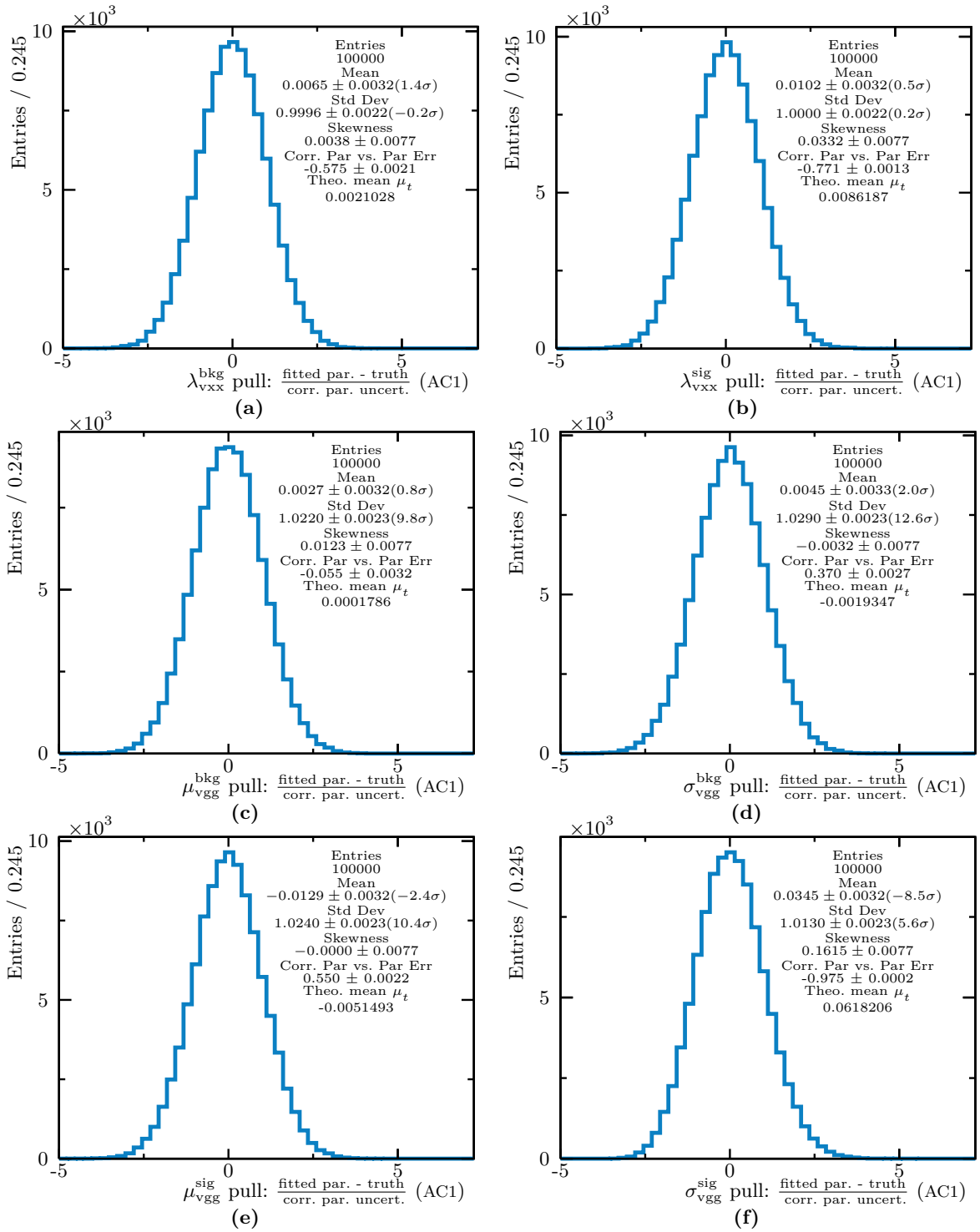


Figure B.4: Overview of the pull distributions for all parameters of the control variable PDFs using the asymptotic correction part one. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

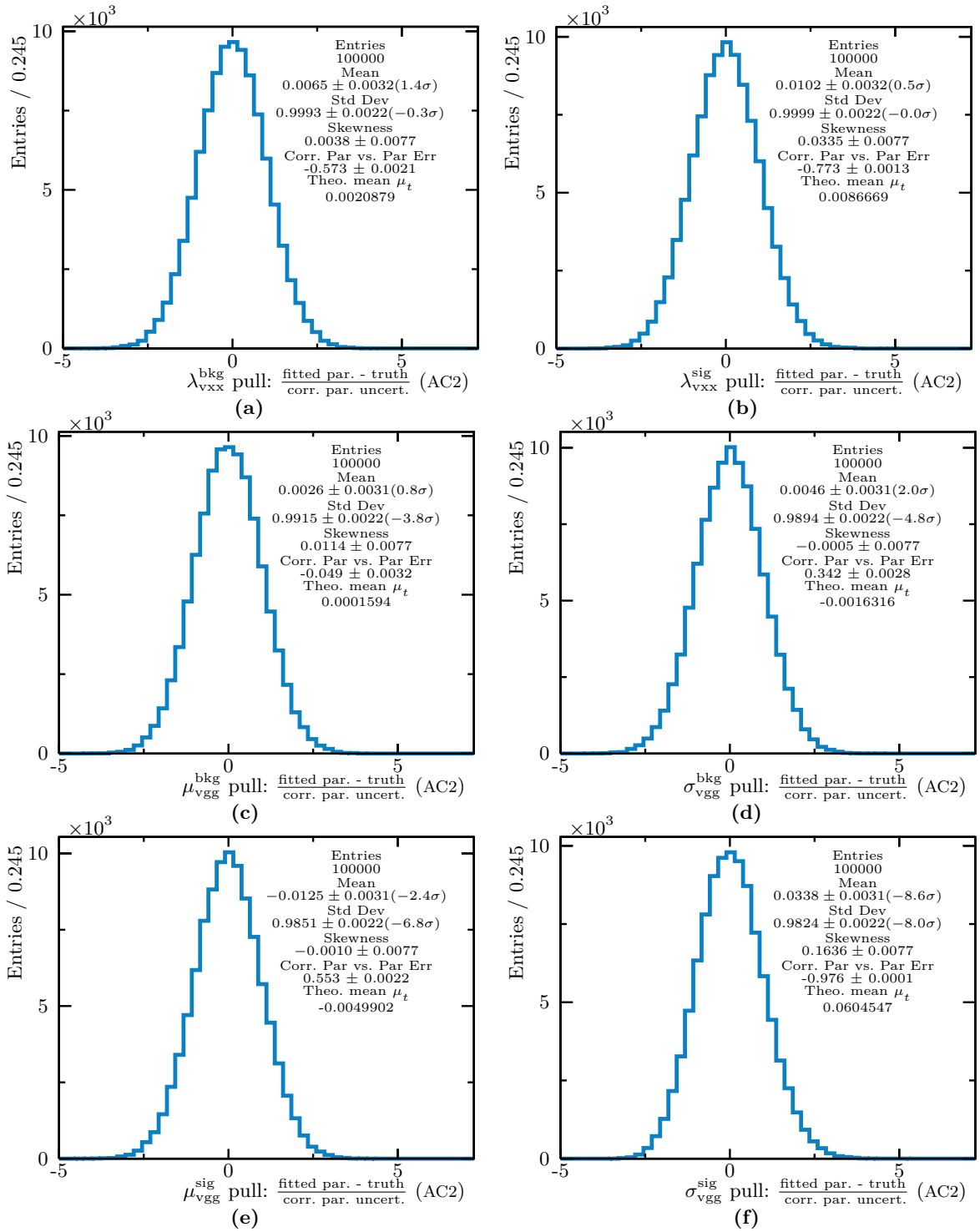


Figure B.5: Overview of the pull distributions for all parameters of the control variable PDFs using the asymptotic correction part two. The meaning of the values in brackets and of μ_t is explained in Figure 7.2.

Table B.2: PDFs of the modified statistical model and their parameters. The variable name Mass is abbreviated as m and vexex as vxx to obtain shorter indices.

Variable	Signal PDF	Background PDF	Range / a.u.
mass	Gaussian: $\mu_{\mathrm{m}}^{\mathrm{sig}} = 60\,\mathrm{a.u.}$, $\sigma_{\mathrm{m}}^{\mathrm{sig}} = 4\,\mathrm{a.u.}$	exponential: $\lambda_{\mathrm{m}}^{\mathrm{bkg}} = -0.025\,\frac{1}{\mathrm{a.u.}}$	$[0, 150]$
vexex	exponential: $\lambda_{\mathrm{vxx}}^{\mathrm{sig}} = -0.1\,\frac{1}{\mathrm{a.u.}}$	exponential: $\lambda_{\mathrm{vxx}}^{\mathrm{bkg}} = -0.01\,\frac{1}{\mathrm{a.u.}}$	$[0, 350]$
vgg	Gaussian: $\mu_{\mathrm{vgg}}^{\mathrm{sig}} = 120\,\mathrm{a.u.}$, $\sigma_{\mathrm{vgg}}^{\mathrm{sig}} = 10\,\mathrm{a.u.}$	Gaussian: $\mu_{\mathrm{vgg}}^{\mathrm{bkg}} = 100\,\mathrm{a.u.}$, $\sigma_{\mathrm{vgg}}^{\mathrm{bkg}} = 10\,\mathrm{a.u.}$	$[60, 160]$

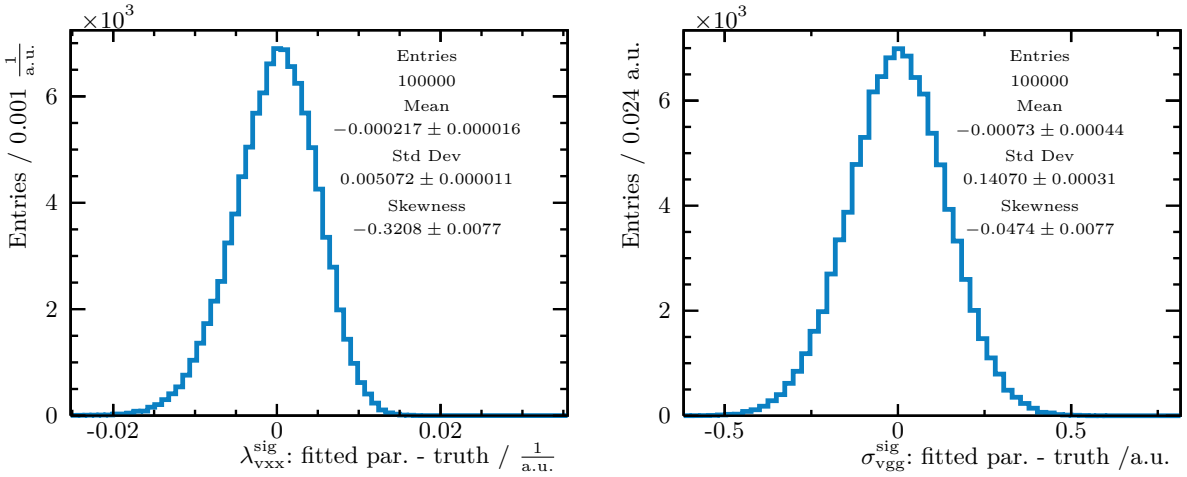


Figure B.6: d_i-distributions for 100 000 pseudo experiments based on the modified statistical model (Table B.2) for (a) the modified parameter $\lambda_{\mathrm{vxx}}^{\mathrm{sig}}$ and (b) the modified parameter $\sigma_{\mathrm{vgg}}^{\mathrm{sig}}$. The bias observed for the pull distribution in Figure 7.10a is confirmed and the observation of no bias in case of the pull distribution in Figure 7.10b is also confirmed.

B.6 Overview of the study with a modified statistical model

In Table B.2 the parameter values of the modified statistical model are presented. The original statistical model is described in Table 7.1. An overview of the mean values and the standard deviations of the pull distributions for all parameters of the control variable PDFs using the three different uncertainty correction methods is presented in Table B.3. The differences between the fitted parameters and the true parameter values presented in Figures B.6a and B.6b verify the observations w.r.t. to the biases in Figures 7.10a and 7.10b.

Table B.3: Overview of (a) the mean values and (b) the standard deviations of the pull distributions for all parameters of the control variable PDFs using the three different uncertainty correction methods for 100 000 pseudo experiments based on the modified statistical model described in Table B.2. The signed significances $\mathcal{S}_{\mu_t}^{\mu_{\text{pull}}}$ and $\mathcal{S}_1^{\sigma_{\text{pull}}}$ (Eqs. (7.2) and (7.3)) are provided in brackets.

(a)

Parameter	squared weights	asymptotic part one	asymptotic part two
$\lambda_{\text{vxx}}^{\text{bkg}}$	-0.0003 ± 0.0032 (-0.7σ)	-0.0003 ± 0.0032 (-0.7σ)	-0.0003 ± 0.0032 (-0.7σ)
$\lambda_{\text{vxx}}^{\text{sig}}$	$+0.0569 \pm 0.0187$ (+8.8σ)	$+0.0556 \pm 0.0032$ (-14.2σ)	$+0.0543 \pm 0.0031$ (-14.4σ)
$\mu_{\text{vgg}}^{\text{bkg}}$	-0.0004 ± 0.0033 (-0.1σ)	-0.0001 ± 0.0032 (-0.1σ)	-0.0001 ± 0.0031 (-0.1σ)
$\sigma_{\text{vgg}}^{\text{bkg}}$	-0.0055 ± 0.0035 (-1.0σ)	-0.0059 ± 0.0033 (-1.2σ)	-0.0056 ± 0.0031 (-1.2σ)
$\mu_{\text{vgg}}^{\text{sig}}$	-0.0018 ± 0.0046 (-2.5σ)	-0.0120 ± 0.0032 (-2.7σ)	-0.0118 ± 0.0031 (-2.7σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	-0.0041 ± 0.0046 (-1.6σ)	$+0.0110 \pm 0.0032$ (-1.6σ)	$+0.0111 \pm 0.0031$ (-1.6σ)

(b)

Parameter	squared weights	asymptotic part one	asymptotic part two
$\lambda_{\text{vxx}}^{\text{bkg}}$	1.0068 ± 0.0023 (+3.0σ)	1.0065 ± 0.0023 (+2.9σ)	0.9981 ± 0.0022 (-0.9σ)
$\lambda_{\text{vxx}}^{\text{sig}}$	5.9182 ± 0.0132 (+371.6σ)	1.0160 ± 0.0023 (+7.0σ)	0.9884 ± 0.0022 (-5.2σ)
$\mu_{\text{vgg}}^{\text{bkg}}$	1.0438 ± 0.0023 (+18.8σ)	1.0212 ± 0.0023 (+9.3σ)	0.9906 ± 0.0022 (-4.2σ)
$\sigma_{\text{vgg}}^{\text{bkg}}$	1.0959 ± 0.0025 (+39.1σ)	1.0422 ± 0.0023 (+18.1σ)	0.9885 ± 0.0022 (-5.2σ)
$\mu_{\text{vgg}}^{\text{sig}}$	1.4700 ± 0.0033 (+143.0σ)	1.0188 ± 0.0023 (+8.3σ)	0.9899 ± 0.0022 (-4.5σ)
$\sigma_{\text{vgg}}^{\text{sig}}$	1.4439 ± 0.0032 (+137.5σ)	1.0126 ± 0.0023 (+5.6σ)	0.9899 ± 0.0022 (-4.5σ)

B.7 Estimators of a truncated Gaussian PDF

For a truncated Gaussian PDF defined in the interval $[a, b]$, where a and b are the given truncation points, the equation system to determine the estimators $\hat{\sigma}$ and $\hat{\mu}$ is defined by [Coh57]

$$\begin{aligned} \hat{\sigma}\left(Z_1 - Z_2 - \xi_1\right) - \bar{x} + a &= 0 \,, \\ \hat{\sigma}^2\left(1 - \xi_1\left(Z_1 - Z_2 - \xi_1\right) - Z_2\frac{b-a}{\hat{\sigma}}\right) - \bar{\sigma}^2 - \left(a - \bar{x}\right)^2 &= 0 \,, \end{aligned} \tag{B.8}$$

with

$$\xi_1 = \frac{a - \hat{\mu}}{\hat{\sigma}} \quad , \quad \xi_2 = \frac{b - \hat{\mu}}{\hat{\sigma}} \tag{B.9}$$

and

$$Z_1 = \frac{\phi(\xi_1)}{I(\xi_1) - I(\xi_2)} \quad , \quad Z_2 = \frac{\phi(\xi_2)}{I(\xi_1) - I(\xi_2)} \,, \tag{B.10}$$

where $I(\xi)$ is defined by

$$I(\xi) = \int_{\xi}^{\infty} \phi(t)\mathrm{d}t \quad \text{with} \quad \phi(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2} \,. \tag{B.11}$$

The weighted mean value of the N measurements x_e used for the estimation is denoted by $\bar{x}$ and defined by

$$\bar{x} = \frac{\sum_{e=1}^{N} w_e x_e}{\sum_{e=1}^{N} w_e} \,. \tag{B.12}$$

The weighted standard deviation is given by

$$\bar{\sigma}^2 = \frac{\sum_{e=1}^{N} w_e \left(x_e - \hat{\mu}\right)^2}{\sum_{e=1}^{N} w_e} \,. \tag{B.13}$$

The value obtained for $\hat{\sigma}$ by solving Equation (B.8) needs to be corrected for the bias in the non-asymptotic case based on Equation (6.16). To solve Equation (B.8) a python program was written, which is provided below. It uses the scientific software toolkit SciPy [Vir+20] and especially its interface to the MINPACK library [MGH80].

Listing B.1: Python code to solve Equation (B.8). To load the class inside ROOT, local variables were avoided and all variables were made class member variables. This results in extensive use of the self keyword.

```
class MLE_Double_Trunc_Gauss():
    """ Calculate estimators for mu and sigma of a doubly truncated Gaussian
    PDF based  Ref. [2]

    [2] Cohen, A. Clifford. "On the Solution of Estimating Equations for
    Truncated and Censored Samples from Normal Populations." Biometrika 44, no.
    1/2 (1957): 225-36. Accessed November 13, 2020. doi:10.2307/2333256.
    """

    def __init__(self, mean, std, a, b, npx=1000):
        """
        @param mean: Mean value of the data set
        @param std: Std. dev. of the data set
        @param a: Left endpoint of the interval of the Gaussian PDF
        @param b: Right endpoint of the interval of the Gaussian PDF
        """
        self.set_mean_set_std(mean=mean, std=std)
        self.a      = a
        self.b      = b
        self.w      = b-a

    def set_mean_set_std(self, mean=None, std=None):
        """
        If no arguments provided just reset the cached mles for mu and sigma
        """
        if mean is not None: self.mean  = mean
        if std is not None:  self.std   = std
        self.cache_mle_mu       = None
        self.cache_mle_sigma    = None

    def phi(self, t):
        """ Standard Gaussian PDF """
        return norm.pdf(t)

    def I(self, xi):
        """ Cumulative function of standard Gaussian """
        return 1.-norm.cdf(xi)

    def Z(self, xi, xi_1, xi_2):
        """ Eq. 6 in [2] """
        return self.phi(xi)/(self.I(xi_1)-self.I(xi_2))
```

```
43      def get_xi_1(self, mu, sigma):
            """ Standardized left endpoint """
45          return (self.a-mu)/sigma

47      def get_xi_2(self, mu, sigma):
            """ Standardized right endpoint """
49          return (self.b-mu)/sigma

51      def eq5(self, mu, sigma, xi_1, Z_1, Z_2):
            """ eq. 5 in [2] involving the sample mean
53          Note for eq. 5: v1 = mean - a
            """
55          return sigma*(Z_1 - Z_2 - xi_1) - (self.mean - self.a)

57      def eq6(self, mu, sigma, xi_1, Z_1, Z_2):
            """ eq. 6 in [2] involving the sample variance
59          Note for eq. 6: v2 = std**2 + (mean - a)**2
            """
61          self.val = sigma**2*(1. - xi_1*(Z_1 - Z_2 - xi_1) - Z_2*self.w / sigma)
            return self.val - (self.std**2 + (self.a-self.mean)**2)
63
        def eq_system(self, args):
65          """ Equation system of eq. 5 and eq. 6 to be solved """
            self.mu, self.sigma = args[0], args[1]
67          self.xi_1 = (self.a - self.mu) / self.sigma
            self.xi_2 = (self.b - self.mu) / self.sigma
69          self.Z_1 = self.Z(xi=self.xi_1, xi_1=self.xi_1, xi_2=self.xi_2)
            self.Z_2 = self.Z(xi=self.xi_2, xi_1=self.xi_1, xi_2=self.xi_2)
71          self.eq5_val = self.eq5(mu=self.mu, sigma=self.sigma, xi_1=self.xi_1,
                                    Z_1=self.Z_1, Z_2=self.Z_2)
73          self.eq6_val = self.eq6(mu=self.mu, sigma=self.sigma,
                    xi_1=self.xi_1, Z_1=self.Z_1, Z_2=self.Z_2)
75          return [self.eq5_val, self.eq6_val]

77      def get_mles(self, meth='hybr'):
            if self.cache_mle_mu is not None and self.cache_mle_sigma is not None:
79              return self.cache_mle_mu, self.cache_mle_sigma
            self.res = root(self.eq_system, x0=[self.mean, self.std], method=meth,
81                      options={'xtol' : 1e-10})
            self.cache_mle_mu, self.cache_mle_sigma = self.res.x[0], self.res.x[1]
83          return self.cache_mle_mu, self.cache_mle_sigma

85      def get_mle_mu(self, meth='hybr'):
            self.get_mles(meth=meth)
87          return self.cache_mle_mu

89      def get_mle_sigma(self, meth='hybr'):
            self.get_mles(meth=meth)
91          return self.cache_mle_sigma
```

The code in Listing B.1 was used to calculate the estimated $\mu_{\text{vgg}}^{\text{bkg}}$, $\sigma_{\text{vgg}}^{\text{bkg}}$, $\mu_{\text{vgg}}^{\text{sig}}$ and $\sigma_{\text{vgg}}^{\text{bkg}}$ values directly for each of the 100 000 pseudo experiments. The differences between this direct calculation and the fitted parameter values obtained from the control variable fit to weighted events are given in Figures B.7a, B.7b, 7.9a and 7.9b. Clear, but non-trivial correlations between the

distributions of the differences mentioned above for $\sigma_{\text{vgg}}^{\text{bkg}}$ and $\sigma_{\text{vgg}}^{\text{sig}}$ and MINUIT's EDM are observed in Figures B.7c and B.7d. These correlations indicate, that the differences observed arise mainly from the convergence criterion of the fit, which is based on the EDM.

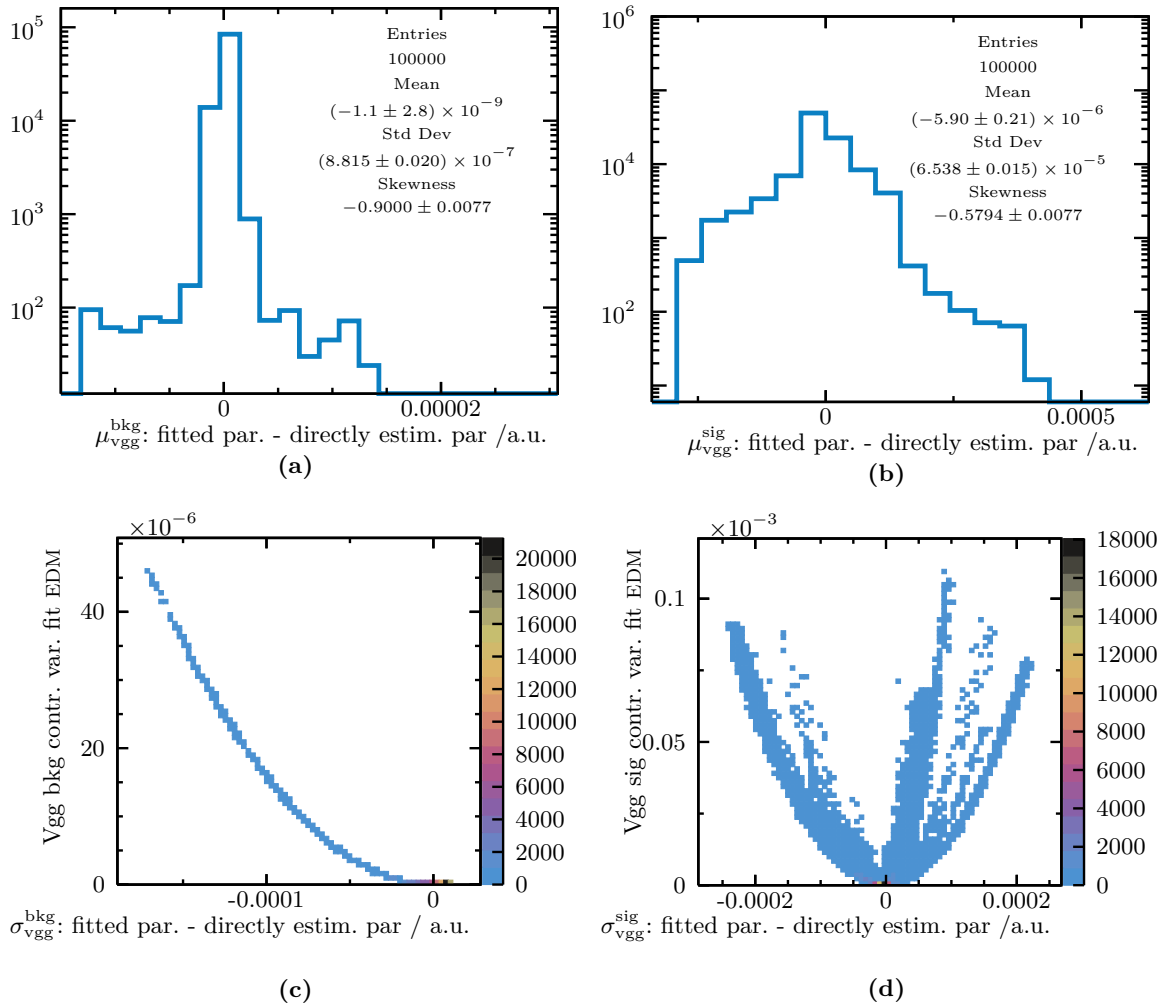


Figure B.7: Distributions of the differences between the estimated parameter values obtained by direct calculation solving Equation (B.8) and those obtained from the control variable fit for (a) the parameter $\mu^{\text{bkg}}_{\text{vgg}}$ and (b) the parameter $\mu^{\text{sig}}_{\text{vgg}}$. The observed bias is small compared to the bias found for the difference between the fitted $\sigma^{\text{sig}}_{\text{vgg}}$ parameter value and its true parameter value (Fig. 7.5f). In (c) and (d) the distributions shown in Figures 7.9a and 7.9b are plotted against the EDM provided by MINUIT after the minimization converged. Clear, but non-trivial correlations are observed, showing that the larger the EDM, the larger is the deviation. This indicates, that the distributions shown in Figures 7.9a and 7.9b might result from the convergence criterion EDM.

C

Acronyms

AC1 asymptotic correction part one

AC2 asymptotic correction part two

ALICE A Large Ion Collider Experiment

ATLAS A Torodial LHC Apparatus

BDT boosted decision tree

CERN European Laboratory for Particle Physics. The acronym originates from Conseil Européen pour la Recherche Nucléaire.

CKM Cabibbo-Kobayashi-Maskawa

CL Confidence Level

CMS Compact Muon Solenoid

CPU Central Processing Unit

CR Cramér-Rao

DAQ Data Acquisition

DOI Digital Object Identifier

EDM Estimated Distance to Minimum

EML Extended Maximum Likelihood

FCNC Flavor Changing Neutral Current

GIM Glashow-Iliopoulos-Maiani

HEP High Energy Physics

HL-LHC High Luminosity LHC

HLT High Level Trigger

IBL Insertable B-Layer

ID Inner Detector

ITk Inner Tracker

LEP Large Electron-Positron Collider

LHCb Large Hadron Collider beauty

LHC Large Hadron Collider

LINAC Linear Accelerator

LS1 Long Shutdown 1

LS2 Long Shutdown 2

LS3 Long Shutdown 3

MC Monte Carlo

MDT Monitored Drift Tubes

MLE Maximum Likelihood Estimator

ML Maximum Likelihood

MSSM Minimal Supersymmetric Standard Model

MS Muon Spectrometer

NLL Negative Log Likelihood

NSW New Small Wheel

PDF Probability Density Function

PDG Particle Data Group

PRD Partially Reconstructed Decay

PS Proton Synchrotron

PV Primary Vertex

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

QFT Quantum Field Theory

RHS Right Hand Side

RPC Resistive Plate Chamber

RoI Region of Interest

SCT Semiconductor Tracker

SM Standard Model

SPS Super Proton Synchrotron

SQW squared weights correction

TGC Thin Gap Chambers

TRT Transition Radiation Tracker

WLCG Worldwide LHC Computing Grid

D

Bibliography

[ACL20] ATLAS Collaboration, CMS Collaboration, and LHCb Collaboration. *Combination of the ATLAS, CMS and LHCb results on the $B^0_{(s)} \to \mu^+\mu^-$ decays*. Tech. rep. CMS-PAS-BPH-20-003. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2727216.

[AF03] S. Ahn and J. A. Fessler. *Standard errors of mean, variance, and standard deviation estimators*. In: EECS Department, The University of Michigan (2003), pp. 1–2.

[Ago+03] S. Agostinelli et al. *GEANT4a simulation toolkit*. In: Nucl Instrum Methods Phys Res , Sect A 506.3 (2003), pp. 250–303.

[ALI08] ALICE Collaboration. *The ALICE experiment at the CERN LHC*. In: Journal of Instrumentation 3.08 (2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[Ard+16] G. Arduini et al. *High Luminosity LHC: Challenges and plans*. In: Journal of Instrumentation 11.12 (2016). DOI: 10.1088/1748-0221/11/12/C12081.

[ATL08] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. In: Journal of Instrumentation 3.08 (2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003.

[ATL10] ATLAS Collaboration. *The ATLAS Simulation Infrastructure*. In: Eur. Phys. J. C (2010). DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568v1 [physics.ins-det].

[ATL12a] ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. In: Phys.Lett. B716 (2012) 1-29 (2012). DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214v2 [hep-ex].

[ATL12b] ATLAS Collaboration. *Search for the decay $B^0_s \to \mu\mu$ with the ATLAS detector*. In: Phys. Lett. B 713 (2012), pp. 387–407. DOI: 10.1016/j.physletb.2012.06.013. arXiv: 1204.0735 [hep-ex].

[ATL14] ATLAS Collaboration. *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*. In: Eur. Phys. J. C 74 (2014). DOI: 10.1140/epjc/s10052-014-3130-x. arXiv: 1407.3935v2 [hep-ex].

[ATL16a] ATLAS Collaboration. *Study of the rare decays of B^0_s and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*. In: Eur. Phys. J. C 76.9 (2016), p. 513. DOI: 10.1140/epjc/s10052-016-4338-8. arXiv: 1604.04263 [hep-ex].

[ATL16b] ATLAS Collaboration. *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*. Tech. rep. ATL-PHYS-PROC-2016-104. 15 pages, EPS-HEP 2015 Proceedings. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2209070.

[ATL17a] ATLAS Collaboration. *Measurement of quarkonium production in proton–lead and proton–proton collisions at 5.02 TeV with the ATLAS detector*. In: Eur. Phys. J. C 78 (2017). DOI: 10.1140/epjc/s10052-018-5624-4. arXiv: 1709.03089v3 [nucl-ex].

[ATL17b] ATLAS Collaboration. *Performance of the ATLAS Trigger System in 2015*. In: Eur. Phys. J. C 77.5 (2017). DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661v2 [hep-ex].

[ATL18] ATLAS Collaboration. *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*. In: JHEP 04 (2019) 098 (2018). DOI: 10.1007/JHEP04(2019)098.

[ATL19a] ATLAS Collaboration. *A 25-year Insider Story Of The Lhc Experiment*. WSPC, 2019. 372 pp. ISBN: 9813271795. URL: https://www.ebook.de/de/product/33731055/the_atlas_collaboration_atlas.html.

[ATL19b] ATLAS collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. ATLAS-CONF-2019-021. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2677054.

[ATL20] ATLAS Collaboration. *$B_s \to \mu^+\mu^-$ candidate event recorded in 2016 (Run 302137, event 1093131714)*. 2020. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Physics#Bs_candidate_event_recorded_in_2.

[ATL97a] ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[ATL97b] ATLAS Collaboration. *ATLAS muon spectrometer: Technical design report*. In: Technical Design Report ATLAS (1997).

[Bar15] N. Bartosik. *Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} = 8$ TeV as test of QCD and background to tt+Higgs production*. Universität Hamburg, Diss., 2015. Dr. Hamburg: Universität Hamburg, 2015, p. 312. DOI: 10.3204/DESY-THESIS-2015-035. URL: https://bib-pubdb1.desy.de/record/222384.

[Bar90] R. Barlow. *Extended maximum likelihood*. In: Nucl. Instrum. Methods Phys. Res., Sect. A 297.3 (1990), pp. 496–506.

[BBS19] M. Beneke, C. Bobeth, and R. Szafron. *Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$*. In: JHEP10(2019)232 (2019). DOI: 10.1007/JHEP10(2019)232. arXiv: 1908.07011v1 [hep-ph].

[Bor+15] A. Borga et al. *The C-RORC PCIe card and its application in the ALICE and ATLAS experiments*. In: Journal of Instrumentation 10.02 (2015), pp. C02022–C02022. DOI: 10.1088/1748-0221/10/02/c02022.

[Bos24] Bose. *Plancks Gesetz und Lichtquantenhypothese*. In: Zeitschrift für Physik 26.1 (1924), pp. 178–181. ISSN: 0044-3328. DOI: 10.1007/BF01327326.

[Bow28] A. L. Bowley. *The Standard Deviation of the Correlation Coefficient*. In: Journal of the American Statistical Association 23.161 (1928), pp. 31–34. ISSN: 0162-1459. URL: http://www.jstor.org/stable/2277400.

[BR97] R. Brun and F. Rademakers. *ROOT: An object oriented data analysis framework*. In: Nucl. Instrum. Meth. A 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.5281/zenodo.3895855.

[Brü+04] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[Bru69] R. M. Brugger. *A note on unbiased estimation of the standard deviation*. In: The American Statistician 23.4 (1969), pp. 32–32.

[BT06] H. Baer and X. Tata. *Weak scale supersymmetry: From superfields to scattering events*. Cambridge University Press, 2006. (Or a similar textbook).

[Cab63] N. Cabibbo. *Unitary Symmetry and Leptonic Decays*. In: Phys. Rev. Lett. 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[Cap+10] M. Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. 2010. URL: https://cds.cern.ch/record/1291633.

[CER09] CERN press release. *Two circulating beams bring first collisions in the LHC*. 2009. URL: https://home.cern/news/press-release/cern/two-circulating-beams-bring-first-collisions-lhc (Accessed: 3 Dec. 2020).

[Che19] S. S. Chelluri. *A study of the sPlot technique for application in the ATLAS B-Physics data analysis*. MA thesis. University of Siegen, 2019. (The treatment of signal events leaking into the sideband regions in the application of the sideband subtraction technique can also be found elsewhere).

[CMS08] CMS Collaboration. *The CMS experiment at the CERN LHC*. In: Journal of Instrumentation 3.08 (2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004.

[CMS12] CMS Collaboration. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. In: Phys. Lett. B 716 (2012) 30 (2012). DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235v2 [hep-ex].

[CMS20] CMS Collaboration. *Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment*. In: JHEP 04 (2020) 188 04.4 (2020), p. 188. DOI: 10.1007/jhep04(2020)188. arXiv: 1910.12127 [hep-ex].

[Coh57] A. C. Cohen. *On the Solution of Estimating Equations for Truncated and Censored Samples from Normal Populations*. In: Biometrika 44.1/2 (1957), pp. 225–236. ISSN: 0006-3444. URL: http://www.jstor.org/stable/2333256.

[Cow98] G. Cowan. *Statistical data analysis*. Oxford university press, 1998.

[Cra46] H. Cramér. *Mathematical methods of statistics*. In: Princeton U. Press, Princeton 500 (1946).

[CS00] E. Corbelli and P. Salucci. *The Extended Rotation Curve and the Dark Matter Halo of M33*. In: Monthly Notices of the Royal Astronomical Society 311.2 (2000), pp. 441–447. DOI: 10.1046/j.1365-8711.2000.03075.x. arXiv: astro-ph/9909252v1 [astro-ph].

[Cur68] E. E. Cureton. *The Teacher's Corner: Unbiased Estimation of the Standard Deviation*. In: The American Statistician 22.1 (1968), pp. 22–22. DOI: 10.1080/00031305.1968.10480435. eprint: https://doi.org/10.1080/00031305.1968.10480435.

[Dir26] P. A. M. Dirac. *On the theory of quantum mechanics*. In: Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 112.762 (1926), pp. 661–677.

[DL02] L. Demortier and L. Lyons. *Everything you always wanted to know about pulls*. In: CDF note 43 (2002). (Or similar literature on this topic).

[EB64] F. Englert and R. Brout. *Broken Symmetry and the Mass of Gauge Vector Mesons*. In: Phys. Rev. Lett. 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[EDJ71] W. T. Eadie, D. Drijard, and F. E. James. *Statistical methods in experimental physics*. 1971.

[Ein05] A. Einstein. *Quantentheorie des einatomigen idealen Gases*. In: *Albert Einstein: AkademieVorträge*. Ed. by D. Simon. John Wiley & Sons, Ltd, 2005, pp. 237–244. ISBN: 9783527608959. DOI: 10.1002/3527608958.ch27. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/3527608958.ch27.

[Fer35] E. Fermi. *On the Quantization of the Monoatomic Ideal Gas*. English. Trans. by A. Zannoni. In: Rend. Accad. Naz. Lincei (1935). arXiv: cond-mat/9912229.

[FST79] A. Frodesen, O. Skjeggestad, and H. Tofte. *Probability and Statistics in Particle Physics*. Bergen, Norway: Universitetsforlaget, 1979. ISBN: 978-82-00-01906-0. (Or a similar textbook).

[G+17] A. G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*. CERN Yellow Reports: Monographs. Geneva: CERN, 2017. DOI: 10.23731/CYRM-2017-004. URL: https://cds.cern.ch/record/2284929.

[GIM70] S. Glashow, J. Iliopoulos, and L. Maiani. *Weak Interactions with Lepton-Hadron Symmetry*. In: Phys. Rev. D 2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[Gla59] S. L. Glashow. *The renormalizability of vector meson interactions*. In: Nuclear Physics 10 (1959), pp. 107–117.

[Gol92] D. Goldberg. *What Every Computer Scientist Should Know About Floating-Point Arithmetic*. In: vol. 23. 1. ACM New York, NY, USA, 1992, pp. 5–48.

[Haf13] J. Haffner. *The CERN accelerator complex*. General Photo. 2013. URL: https://cds.cern.ch/record/1621894.

[HHM20a] S. Hageboeck, T.-P. Hücking, and L. Moneta. *In few cases RooAbsPdf::generate() fails to create correct distributions*. 2020. URL: https://root-forum.cern.ch/t/in-few-cases-rooabspdf-generate-fails-to-create-correct-distributions/40110 (Accessed: 30 Sept. 2020).

[HHM20b] S. Hageboeck, T.-P. Hücking, and L. Moneta. *Is it possible to use the AsymptoticError method in RooAbsPdf::fitTo without parameter limits?* 2020. URL: https://root-forum.cern.ch/t/is-it-possible-to-use-the-asymptoticerror-method-in-rooabspdf-fitto-without-parameter-limits/39985 (Accessed: 30 Sept. 2020).

[HHM20c] S. Hageboeck, T.-P. Hücking, and L. Moneta. *Wrong error on number of events when using AsymptoticError and Extended flag in tutorial rf611_weightedfits.C*. 2020. URL: https://root-forum.cern.ch/t/wrong-error-on-number-of-events-when-using-asymptoticerror-and-extended-flag-in-tutorial-rf611-weightedfits-c/39776 (Accessed: 1 Oct. 2020).

[Hig64] P. W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons*. In: Phys. Rev. Lett. 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[Hig93] N. Higham. *The Accuracy of Floating Point Summation*. In: SIAM J. Sci. Comput. 14 (1993), pp. 783–799.

[HiL20] HiLumi project. *The HL-LHC project timeline as of August 2020*. 2020. URL: https://hilumilhc.web.cern.ch/content/hl-lhc-project (Accessed: 2 Dec. 2020).

[HM20] S. Hageboeck and L. Moneta. *[ROOT-10866] Numerical instabilities when calculating the derivatives for the asymptotically correct erros - SFTJIRA*. 2020. URL: https://sft.its.cern.ch/jira/browse/ROOT-10866 (Accessed: 30 Sept. 2020).

[Hol50] W. H. Holtzman. *The Unbiased Estimate of the Population Variance and Standard Deviation*. In: The American Journal of Psychology 63.4 (1950), pp. 615–617. ISSN: 0002-9556. URL: http://www.jstor.org/stable/1418879.

[Ins85] Institute of Electrical and Electronics Engineers (IEEE). *IEEE Standard for Binary Floating-Point Arithmetic*. In: ANSI/IEEE Std 754-1985 (1985), pp. 1–20.

[JR75] F. James and M. Roos. *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*. In: Comput. Phys. Commun. 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

[Kah65] W. Kahan. *Pracniques: Further Remarks on Reducing Truncation Errors*. In: Commun. ACM 8.1 (1965), p. 40. ISSN: 0001-0782. DOI: 10.1145/363707.363723.

[KM73] M. Kobayashi and T. Maskawa. *CP Violation in the Renormalizable Theory of Weak Interaction*. In: Prog. Theor. Phys. 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[KSO87] M. G. Kendall, A. Stuart, and J. K. Ord. *Kendalls Advanced Theory of Statistics*. USA: Oxford University Press, Inc., 1987. ISBN: 0195205618. (Or a similar textbook).

[Lan19] C. Langenbruch. *Parameter uncertainties in weighted unbinned maximum likelihood fits*. 2019. arXiv: 1911.01303v2 [physics.data-an].

[LAP86] L. Lyons, W. W. Allison, and J. Pañella Comellas. *Maximum likelihood or extended maximum likelihood? An example from high energy physics*. In: Nucl. Instrum. Methods Phys. Res., Sect. A 245.2 (1986), pp. 530–534. ISSN: 0168-9002. DOI: 10.1016/0168-9002(86)91293-3. URL: http://www.sciencedirect.com/science/article/pii/0168900286912933.

[LHC08] LHCb Collaboration. *The LHCb Detector at the LHC*. In: Journal of Instrumentation 3.08 (2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005.

[LHC17] LHCb Collaboration. *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*. In: Phys. Rev. Lett. 118, 191801 (2017) (2017). DOI: 10.1103/PhysRevLett.118.191801. arXiv: 1703.05747v2 [hep-ex].

[Lyo86] L. Lyons. *STATISTICS FOR NUCLEAR AND PARTICLE PHYSICISTS*. 1986, pp. 204+. ISBN: 978-0-521-37934-2. (Or a similar textbook).

[Meu+17] A. Meurer et al. *SymPy: symbolic computing in Python*. In: PeerJ Computer Science 3 (2017), e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103.

[MGH80] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. *User guide for MINPACK-1*. Tech. rep. ANL-80-74. Argonne, IL: Argonne Nat. Lab., 1980. URL: https://cds.cern.ch/record/126569.

[Mon+10] L. Moneta et al. *The RooStats Project*. In: PoS ACAT2010 (2010). arXiv: 1009.1003v2 [physics.data-an].

[Nat+17] P. Natarajan et al. *Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations*. In: Monthly Notices of the Royal Astronomical Society 468.2 (2017), pp. 1962–1980. DOI: 10.1093/mnras/stw3385. arXiv: 1702.04348v2 [astro-ph.GA].

[Ney37] J. Neyman. *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*. In: Phil. Trans. Roy. Soc. Lond. A 236.767 (1937), pp. 333–380. DOI: 10.1098/rsta.1937.0005.

[Oli+14] K. A. Olive et al. *PDG Particle Listings*. In: Chin. Phys. C, 38, 090001 (2014). URL: https://pdg.lbl.gov/2014/listings/contents_listings.html.

[Ore58] J. Orear. *Notes on statistics for physicists*. University of California, Lawrence Radiation Laboratory; printed for the US Atomic Energy Commission, 1958.

[Pau25] W. Pauli. *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*. In: Zeitschrift für Physik 31.1 (1925), pp. 765–783.

[PD05] M. Pivk and F. R. L. Diberder. *sPlot: a statistical tool to unfold data distributions*. In: Nucl. Instrum. Methods Phys. Res., Sect. A 555 (2005). This is a note, pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083v3 [physics.data-an].

[Pea16] B. Pearson. *Searches for a high-mass Higgs-like diboson resonance in the $H \to WW \to \ell\nu qq$ decay channel using pp collisions at both $\sqrt{s} = 8$ and 13 TeV with the ATLAS detector at the LHC*. PhD thesis. Oklahoma U., 2016.

[Peq08] J. Pequenao. *Computer generated image of the whole ATLAS detector*. 2008. URL: https://cds.cern.ch/record/1095924.

[Pla14] Planck Collaboration. *Planck 2013 results. I. Overview of products and scientific results.* In: Astronomy & Astrophysics 571 (2014), A1. DOI: 10.1051/0004-6361/201321529. arXiv: 1303.5062v2 [astro-ph.CO].

[Rao92] C. R. Rao. *Information and the accuracy attainable in the estimation of statistical parameters.* In: *Breakthroughs in statistics.* Springer, 1992, pp. 235–247.

[ROO20] ROOT Team. *ROOT: analyzing petabytes of data, scientifically.* 2020. URL: https://root.cern.ch (Accessed: 5 Aug. 2020).

[SMS07] T. Sjöstrand, S. Mrenna, and P. Skands. *A Brief Introduction to PYTHIA 8.1.* In: Comput.Phys.Commun.178:852-867,2008 (2007). DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820v1 [hep-ph].

[SNO02] SNO Collaboration. *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory.* In: Phys. Rev. Lett. 89 (1 2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301.

[Sup98] Super-Kamiokande Collaboration. *Evidence for Oscillation of Atmospheric Neutrinos.* In: Phys. Rev. Lett. 81 (8 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[SW59] A. Salam and J. Ward. *Weak and electromagnetic interactions.* In: Il Nuovo Cimento (1955-1965) 11 (1959), pp. 568–577.

[The20] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 9.1).* 2020. URL: https://www.sagemath.org.

[Vir+20] P. Virtanen et al. *SciPy 1.0: fundamental algorithms for scientific computing in Python.* In: Nature methods 17.3 (2020), pp. 261–272.

[VK03] W. Verkerke and D. Kirkby. *The RooFit toolkit for data modeling.* In: eConf C0303241 (2003). Ed. by L. Lyons and M. Karagoz, MOLT007. arXiv: physics/0306116v1 [physics.data-an].

[Wei67] S. Weinberg. *A Model of Leptons.* In: Phys. Rev. Lett. 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[WH98] N. Wermes and G. Hallewel. *ATLAS pixel detector: Technical Design Report.* Technical Design Report ATLAS. Geneva: CERN, 1998. URL: https://cds.cern.ch/record/381263.

Acknowledgments

Looking back on more than a year of working on my master thesis and my time as a student at the Siegen University, there are many persons, who deserve my gratitude. Some of them I would like to thank explicitly:

- The living *God*, who has revealed himself in Jesus, is willing to forgive and cares for a master student in Siegen.
- It is really a privilege to have a supervisor like *Prof. Dr. Peter Buchholz* with a wealth of experience. He also shared this experience, as well as his views on e.g. modern developments in particle physics. I will not forget his concerns about younger physicists, who are more after fancy techniques than after doing real physics. I esteem him as a man with profile, something that becomes more and more rare.
- *Prof. Dr. Markus Risse* agreed to be my co-reviewer. Most of my knowledge in particle physics I gained in his lectures. However, not only in his lectures but also at lunch – before the COVID-19 pandemic of course – he would explain some recent results in astroparticle physics, if asked to do so.
- *Dr. Wolfgang Walkowiak* has been an excellent mentor to me. He invested much of his precious time for me. For example, I remember him sitting with me in his office until 11 PM at Friday night to do some calculations. He shaped my thesis with plenty of great ideas and suggestions. I tried to learn a lot from his structured and analytical approach. I hope to develop also such an eagle's eye to spot mistakes at a glance. His proofreading improved my thesis a lot. Furthermore, I enjoyed the conversations on computing, technology, hiking and any other things.
- *Siddardha Chelluri* did the groundwork for my thesis. He investigated the sPlot technique as well as uncertainties of fits to weighted events and introduced me into these topics. He was very interested in my work and we had fruitful discussions. He and his housemates cooked various tasty Indian dishes for me. I always enjoyed it.
- I can not forget the nice collaboration with *Dr. Qader Dorosti* on the Test System, as well as on holding the Physics I exercise classes.

- I received valuable comments from *Sonja Timmermann* on parts of the thesis and enjoyed working together with her on the Test System, before starting the master theses.

- *Nils Kreher* was a nice colleague in the first semesters at University. As a theoretician, he helped me to a better understanding of the Standard Model and supported me with great comments on Chapter 2.

- *Arpan Ghosal and Buddhadeb Mondal* read part of my thesis and contributed to a nice working atmosphere. It was helpful to exchange some knowledge about VIM and other software utilities with Buddhadeb.

- *Jan-Joachim Hahn* provided me with valuable comments on parts of this thesis. He studied with me for the oral examination and explained quite some things to me.

- *Nicolas Seitz* and *Michael Gaida* provided access to Mathematica to calculate Equation (6.40).

- *Dr. Wolfgang Walkowiak and Philip Rühl* are the administrators of the SiMPLE computing cluster, which I used for my studies. They do this job with great commitment.

- All members of the experimental particle and astroparticle physics group. They created a very nice working atmosphere. Furthermore, all people from the Siegen University, the ATLAS collaboration and CERN, who contributed in an indirect way to my thesis, e.g. by cleaning the building, providing computing resources or by developing the ROOT software toolkit.

- I am very glad to have outstanding friends, who transformed this turbulent year to a great experience. They had deep solicitousness for my progress, although they are hardly interested in physics. Together with them, I am looking forward to great experiences and joyful moments to come.

- Last but not least, my parents *Martin and Christiane Hücking* supported me a lot. I think it will take some time until I realize all that I owe to them. Besides my parents, also my sister and her fiancé took great interest in the progress of my thesis and encouraged me.

F

Declaration

Hiermit erkläre ich, dass ich die vorliegende Master-Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate und Ergebnisse Anderer kenntlich gemacht habe.

(Ort) (Datum) (Unterschrift)