

Arbeit zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften

**Tuning of Monte Carlo algorithms with
Bayesian methods for the improvement
of cosmic ray induced air shower
simulations**

Michael Elias Windau
geboren in Unna

2026

AG Kröninger
Fakultät Physik
Technische Universität Dortmund

Diese Dissertation wurde der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegt.

Erstgutachter:	Prof. Dr. Kevin Kröninger
Zweitgutachter:	Prof. Dr. Karl-Heinz Kampert
Vorsitzender der Prüfungskommission:	Prof. Dr. Metin Tolan
Vertreter der wiss. Mitarbeiter:	Dr. Michael Paulus

Datum des Einreichens der Arbeit:	12. Mai 2026
-----------------------------------	--------------

Abstract

This thesis investigates the Bayesian tuning of hadronic interaction models used in Monte Carlo simulations of cosmic-ray-induced air showers. Since hadronic interactions govern the production of secondary particles in the atmosphere, their modeling is essential for describing air-shower observables and for interpreting cosmic-ray measurements.

The hadronic event generator `Pythia 8` is used in this study and constrained with accelerator data from collider and fixed-target experiments. Although the highest-energy cosmic-ray-induced air showers extend beyond the reach of current accelerator measurements, the selected observables constrain aspects of hadronic particle production that are also relevant during air-shower development at lower energy scales. Rather than tuning directly to air-shower data, the aim is to investigate how constraints from accelerator measurements propagate to air-shower simulations.

The analysis focuses on a selected set of sensitive model parameters and observables. The thesis also includes a comparison of surrogate models that parameterize the generator response, thereby enabling the Bayesian tuning procedure. The choice of surrogate model is shown to be crucial both for the quality of the tune and for the reliability of the uncertainty propagation to the simulated observables.

The results show that the selected collider and fixed-target data sets cannot be described optimally by a single common tune, as they favor different regions of parameter space. Separate tunes are therefore constructed and studied. These tunes improve the description of the corresponding accelerator data sets and, in turn, induce measurable changes in air-shower simulations. In particular, the predicted average muon number changes by about 7–9%, demonstrating that accelerator-based tuning can directly impact the modeling of cosmic-ray-induced air showers.

Kurzfassung

Diese Arbeit untersucht das bayessche Tuning hadronischer Wechselwirkungsmodelle, wie sie in Monte-Carlo-Simulationen kosmischer Luftschauer verwendet werden. Da hadronische Wechselwirkungen die Erzeugung sekundärer Teilchen in der Atmosphäre bestimmen, ist ihre Modellierung entscheidend für die Beschreibung von Luftschauer-Observablen und für die Interpretation kosmischer Strahlungsmessungen.

In dieser Arbeit wird der hadronische Ereignisgenerator **Pythia 8** verwendet und mithilfe von Beschleunigerdaten aus Collider- und Fixed-Target-Experimenten eingeschränkt. Obwohl die höchstenergetischen kosmischen Luftschauer Wechselwirkungen jenseits der Reichweite aktueller Beschleunigermessungen umfassen, schränken die ausgewählten Observablen Aspekte der hadronischen Teilchenproduktion ein, die auch bei der Entwicklung von Luftschauern auf niedrigeren Energieskalen relevant sind. Anstatt ein direktes Tuning an Luftschauerdaten durchzuführen, besteht das Ziel darin zu untersuchen, wie sich Einschränkungen aus Beschleunigermessungen auf Luftschauersimulationen übertragen.

Die Analyse ist auf eine ausgewählte Menge sensitiver Modellparameter und Observablen beschränkt. Die Arbeit umfasst außerdem einen Vergleich von Surrogatmodellen, die zur Parametrisierung der Generatorantwort verwendet werden und damit das bayessche Tuning-Verfahren ermöglichen. Es zeigt sich, dass die Wahl des Surrogatmodells sowohl für die Qualität des Tunings als auch für die Verlässlichkeit der Unsicherheitsfortpflanzung vom Tuning auf die simulierten Observablen entscheidend ist.

Die Ergebnisse zeigen, dass die ausgewählten Collider- und Fixed-Target-Datensätze nicht optimal durch ein einziges gemeinsames Tuning beschrieben werden können, da sie unterschiedliche Bereiche des Parameterraums bevorzugen. Daher werden getrennte Tunes konstruiert und untersucht. Diese Tunes verbessern die Beschreibung der jeweiligen Beschleunigerdatensätze und führen zugleich zu messbaren Änderungen in Luftschauersimulationen. Insbesondere ändert sich die vorhergesagte mittlere Myonenzahl um etwa 7–9%, was zeigt, dass ein auf Beschleunigerdaten basierendes Tuning die Modellierung kosmischer Luftschauer direkt beeinflussen kann.

Contents

1	Introduction	1
2	Standard Model of Particle Physics	3
2.1	Fundamental Particles	3
2.2	Hadronization at High-Energy Particle Accelerators	5
3	Cosmic Rays	9
3.1	Cosmic Ray Flux	9
3.2	Extensive Air-Showers	10
3.3	Experimental Detection	12
3.4	The Muon Puzzle	15
3.5	Difference Between Air-Shower and Collider Events	18
4	Event Generation with Monte Carlo Simulations	19
4.1	Hard Process	19
4.2	Parton Showers	20
4.3	Hadronization	21
4.4	Pythia 8 Lund String Model	21
4.5	Rivet	26
4.6	CORSIKA 8	27
5	Bayesian Methods in Particle Physics	29
5.1	Bayes' Theorem	29
5.2	Combining Bayesian Inference and Monte Carlo Methods	31
5.3	Markov Chain Monte Carlo	31
6	Tuning Monte Carlo Event Generators	33
6.1	Adjusting Free Parameters	33
6.2	Tuning Workflow Using Bayesian Inference	34
6.3	Generating Data	35
6.4	Constructing a Parameterization	36
6.5	Fitting to Data	39
6.6	Evaluating the Tuning Result	41
6.7	Uncertainty Propagation	42

7	Feasibility Study of the Tuning of Pythia 8 Simulated Air Showers	45
7.1	Pythia 8 as Hadronic Interaction Model for Air Shower Simulations .	45
7.2	Simulation of Muon Flux with CORSIKA 8	49
8	Exploring the Interplay between Collider and Air-Shower Simulations	55
8.1	Understanding Air Showers with LHC Collider Data	56
8.2	Selection of Pythia 8 Parameters	58
8.3	Selection of Collider Observables	75
8.4	Selection of the Surrogate Model	76
8.5	Analysis of the Posterior Distributions	83
8.6	Impact of Tunes on the Observables	91
8.7	Uncertainty Propagation	96
8.8	Limitations of the Parameterization	98
8.9	Influencing Air-Shower Simulations	106
9	Conclusion	113
10	Outlook	115
A	Appendix	117
	Bibliography	127
	Acknowledgements	135

List of Contributions

This thesis investigates the tuning of hadronic interaction models using Bayesian methods with the aim of improving the simulation of cosmic-ray-induced air showers. Rather than tuning directly to air-shower data, the goal is to study how constraints from accelerator-based measurements propagate to air-shower simulations. This work was carried out in collaboration with other researchers. During this thesis, I contributed to the two publications listed below.

- J. Albrecht *et al.*, *Global tuning of hadronic interaction models with accelerator-based and astroparticle data*, Nat. Rev. Phys. 8 (2025) 98

This article discusses the global tuning of hadronic interaction models using accelerator-based and astroparticle data. My contribution to this publication focused on the section describing the tuning procedure based on Bayesian inference. The tuning procedure is presented in detail in Chapter 6.

- M. Windau, C. Gaudu, K.-H. Kampert, K. Kröninger, *Improving Air Shower Simulations by Tuning Pythia 8/Angantyr with Accelerator Data*, PoS(ICRC2025) (2025) 438

This article presents the tuning of the Pythia 8/Angantyr event generator using accelerator data and studies its impact on air-shower observables. I contributed to all aspects of this work, including the design of the study, the implementation of the tuning workflow, the execution of the tuning and simulations, the evaluation of the results, and the preparation of the manuscript. The published results represent intermediate results of the study presented in Chapter 8.

1 Introduction

Scientific research seeks to understand natural phenomena by combining theoretical ideas with experimental observation. Theories provide descriptions of physical processes and lead to predictions that can be tested against data. Experiments, in turn, provide the information needed to assess how well these descriptions reflect reality. The comparison between model predictions and measurements is therefore a central element of modern physics and relies strongly on statistical and computational methods.

In particle physics, the connection between theory and experiment is often non-trivial, since many relevant processes cannot be calculated analytically in full detail. In particular, the formation of hadronic final states cannot be described entirely from first principles and therefore requires phenomenological modeling, because it involves processes that are not analytically tractable in full detail. Monte Carlo event generators translate theoretical assumptions into predictions for measurable observables by simulating complete collision events. Their predictive power depends crucially on these phenomenological components, whose free parameters are not fixed by theory and must therefore be constrained by experimental data.

Cosmic rays are of interest both as messengers of energetic astrophysical processes and as probes of particle interactions at energies that can exceed those accessible at accelerators. Ultra-high-energy cosmic rays are studied through the extensive air showers they initiate in the atmosphere. The properties of the primary particle must be inferred from measured shower observables by comparison with simulations. Hadronic interaction models play a central role in this procedure, because they govern the first interactions of the shower and influence its subsequent development, thereby affecting key observables measured by air-shower experiments.

Comparisons with cosmic-ray air-shower data reveal persistent discrepancies between simulations and measurements. A well-known example is the *muon puzzle* [1], namely the observation that air-shower simulations tend to underestimate the number of muons with respect to experimental data. This suggests that aspects of hadronic interactions relevant to air-shower development may not be adequately described.

Accelerator measurements provide an essential input for improving and validating hadronic interaction models, because they constrain particle-production processes

that are also relevant for extensive air showers. In accelerator experiments, these processes can be studied under controlled conditions and often with high precision. Although air showers probe hadronic interactions in a kinematic regime that is not fully accessible to accelerator experiments, collider and fixed-target data provide important constraints.

A natural approach to improving hadronic interaction models is to constrain their free parameters by comparison with experimental data. In Monte Carlo event generators, this is achieved through tuning, that is, by determining parameter values for which the simulated observables provide the best possible description of the measurements. In the approach followed in this work, the dependence of simulated observables on the generator parameters is approximated by a surrogate model. Combined with a prior for the parameters, this parameterization is used in Bayesian inference to obtain the posterior distribution of the model parameters. The posterior provides more than a single preferred parameter set. It allows distinct parameter regions compatible with the data to be identified. In addition, it also enables the propagation of tuning uncertainties to simulated observables.

This thesis is structured as follows. Chapter 2 introduces the Standard Model of particle physics and discusses the hadronization of particles at high-energy particle interactions. An overview of cosmic rays, extensive air showers, their experimental detection, and the muon puzzle is given in Chapter 3. Chapter 4 describes Monte Carlo event generation and introduces `Pythia 8` as a hadronic interaction model [2], as well as the `Rivet` [3] and `CORSIKA 8` [4, 5] framework. The Bayesian methods used in this work are presented in Chapter 5. Chapter 6 presents the tuning methodology employed in this work, including the evaluation of the tuning results and the propagation of tuning uncertainties. Chapter 7 investigates the feasibility of tuning `Pythia 8` using simulated air showers and examines the simulation of the muon flux with `CORSIKA 8`. The impact of tuning `Pythia 8` to collider measurements for air-shower predictions is investigated in Chapter 8. Finally, Chapter 9 summarizes the results and Chapter 10 provides an outlook.

2 Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a quantum field theory that describes the interactions between the known fundamental particles in our universe [6–10]. These interactions include three of the four fundamental forces, namely the strong, weak, and electromagnetic forces, with only the gravitational force not described by the SM. Despite being incomplete, the SM successfully explains most experimental observations in particle physics and enables precise predictions for experimental measurements, which in turn continue to test, refine, and extend the theory.

2.1 Fundamental Particles

The fundamental particles are categorized according to their intrinsic properties like spin, mass, and charge. They are grouped into fermions (spin-1/2), vector bosons (spin-1), and a scalar boson (spin-0), as illustrated in Figure 2.1.

Fermions are classified into six quarks and six leptons. For each fermion, there exists a corresponding antiparticle. Antiparticles have the same mass and spin as their corresponding particles, but opposite values of all additive quantum numbers, such as electric charge.

Both quarks and leptons occur in three generations, where each new generation can be regarded as a more massive version of the previous one. For example, the tau lepton is heavier than the muon and the electron, but all three carry the same electric charge of $-1e$. The tau lepton can decay into either of the lighter charged leptons, whereas the muon can decay only into an electron. Each charged lepton has an associated neutrino, which is electrically neutral and massless in the SM.

As in the lepton sector, quarks are arranged in three generations, with the heavier quarks belonging to the second and third generations. The top quark is the most massive known elementary particle. Quarks carry fractional electric charge: the down, strange, and bottom quarks have charge $-1/3e$, while the up, charm, and top quarks have charge $+2/3e$. In addition, quarks carry color charge, which is the charge associated with the strong interaction. Protons and neutrons are composed primarily of first-generation quarks, namely up and down quarks. Consequently,

2 Standard Model of Particle Physics

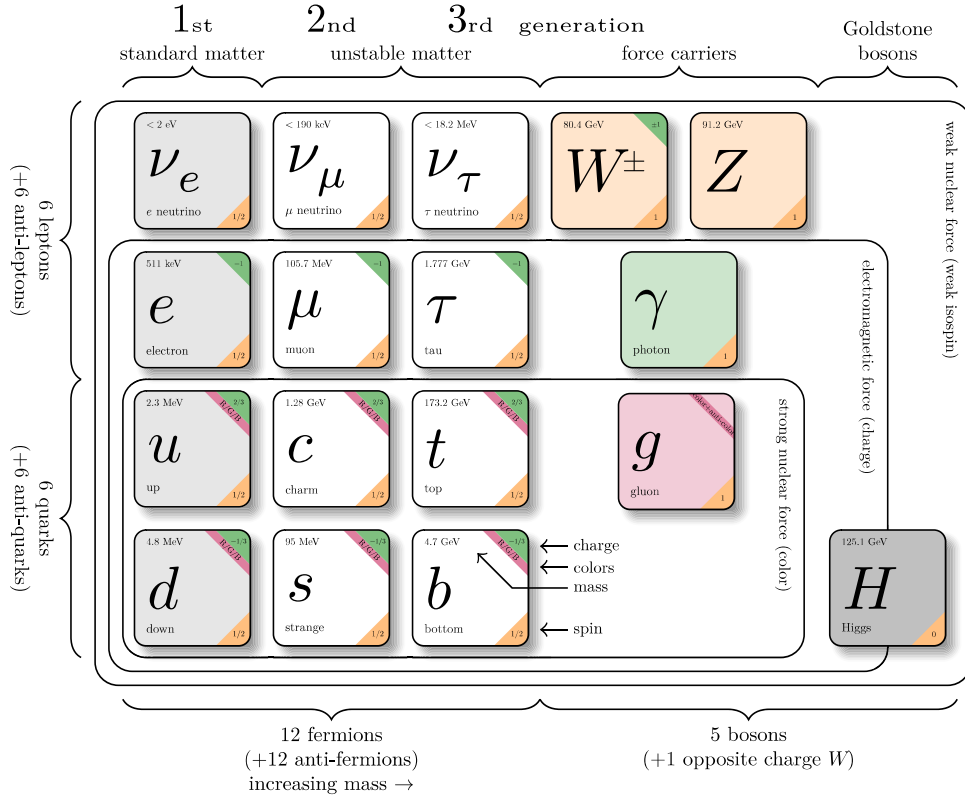


Figure 2.1: Illustration of the Standard Model. Adapted from [11], figure taken from Ref. [12], with particle property values from [13].

ordinary visible matter consists mainly of electrons and atomic nuclei, whose nucleons are built from first-generation fermions.

The interactions between fermions are mediated by vector bosons, often referred to as force carriers. These bosons couple to the corresponding charges of the fermions and, in some cases, also to each other. The photon is the force carrier of the electromagnetic interaction and couples to particles carrying electric charge. This interaction is described in the Standard Model by quantum electrodynamics. Being both massless and neutral, the photon does not interact with itself.

The gluon is the massless vector boson associated with the strong interaction, which is described in the Standard Model by quantum chromodynamics (QCD). In analogy to the photon, the gluon couples to particles carrying the corresponding charge. In QCD, this is color charge rather than electric charge. Unlike photons, however, gluons themselves carry color charge. As a result, they can couple not

only to quarks, but also directly to other gluons. Quarks carry one of three color charges, conventionally denoted red, green, and blue, while antiquarks carry the corresponding anticolors. Gluons carry a combination of color and anticolor, allowing them to change the color of quarks while conserving the total color charge.

The weak interaction is mediated by three massive vector bosons: the W^+ , W^- , and Z bosons. The charged $W^\pm$ bosons mediate transitions between related fermions, such as a charged lepton and its corresponding neutrino or an up-type and a down-type quark. In these processes, electric charge is transferred by the $W^\pm$ boson, so that the total charge remains conserved. The neutral Z boson mediates weak interactions in which the fermion type and electric charge remain unchanged. Unlike the massless photon and gluon, the $W^\pm$ and Z bosons have high mass and rank among the most massive fundamental particles.

The only fundamental scalar boson is the Higgs boson, which was the latest fundamental particle to be discovered in 2012 by the collaborations of the ATLAS and CMS experiments at CERN [14, 15]. Unlike the other bosons, the Higgs boson is not directly associated with a fundamental force. Its existence is a direct consequence of the Higgs mechanism, in which the Higgs field acquires a nonzero vacuum expectation value and causes spontaneous symmetry breaking in the electroweak sector of the Standard Model [16, 17]. This results not only in the appearance of the massive Higgs boson but also in masses for the bosons of the weak interaction. Fermions obtain their masses through Yukawa couplings to the Higgs field.

2.2 Hadronization at High-Energy Particle Accelerators

While the intrinsic properties of the interacting particles govern the types of possible interactions, the strength with which these interactions occur is described by the coupling constants, which differ significantly between interactions. Unlike what the name suggests, these coupling constants are dependent on the energy scale and, therefore, the distance between the interacting particles. The strong coupling constant, which corresponds to the strong force, grows with larger distances (small energies) and declines on small length scales (high energies). As a result, color-charged particles do not occur as isolated particles at small energy scales, a phenomenon known as color confinement. At very high energy scales, however, the strong coupling becomes small, so that quarks and gluons behave approximately as free particles. This property is known as asymptotic freedom. To reach such energy scales to investigate these particles further, high-energy particle accelerators such as the Large Hadron Collider (LHC) at CERN are employed [18]. In these experiments, hadrons, which are composite particles made of quarks, are accelerated

and then collided at designated interaction points in order to produce high-energy quarks and gluons.

Most hadrons are classified as either mesons or baryons. Mesons consist of a valence quark–antiquark pair, while baryons contain three valence quarks. Beyond this valence content, hadrons also contain gluons and sea quark–antiquark pairs generated by the dynamics of the strong interaction. These particles contribute to the internal structure of the hadron and carry fractions of its total momentum. Together, these constituents are called partons, and their respective momentum fraction of the hadron is described by parton density functions (PDF), which give the probability density for finding a parton with a particular momentum fraction at a given energy scale. PDFs depend on the energy scale at which the hadron is probed, typically characterized by the momentum transfer Q^2 in a hard-scattering process. As Q^2 increases, gluons and sea quarks carry an increasingly large fraction of the hadron’s momentum. An example for PDFs of protons, measured in multiple collider experiments, is shown in Figure 2.2.

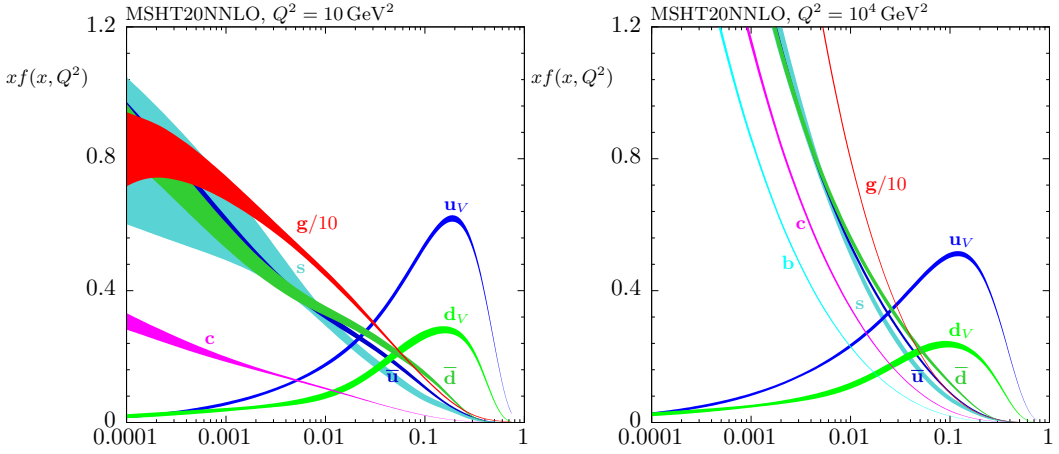


Figure 2.2: Proton PDFs at the energy scales $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ (right), taken from Ref. [19]. The horizontal axis shows the momentum fraction and the v index at the u and d indicates the valence quarks. The curve for the gluon parton is divided by 10 for visibility.

In high-energy collisions with a large momentum transfer, hadrons are resolved into their partonic constituents rather than behaving as indivisible particles. The interacting partons carry fractions of the hadron momentum described by the PDFs. Since the strong coupling becomes small at such scales, quarks and gluons can be treated as approximately free during the short-distance interaction. This enables the study of quark and gluon interactions through high-energy scattering processes.

Due to confinement, colored partons cannot be isolated as free particles. As color-connected partons separate, the energy stored in the confining color field increases. In phenomenological hadronization models, this is often described by the formation of color flux tubes or strings. When the available energy becomes sufficiently large, additional quark-antiquark pairs can be produced, allowing the system to reorganize into hadrons with no net color charge. This non-perturbative transition from colored partons to hadrons is known as hadronization. In high-energy collisions, outgoing quarks and gluons first undergo perturbative radiation, producing a parton shower. After this shower evolution, hadronization converts the colored partons into a collimated spray of final-state hadrons. Experimentally, such sprays are reconstructed as jets using jet algorithms, making jets observable proxies for the underlying quarks and gluons. Accurate modeling of hadronization is therefore essential for high-energy particle physics studies and will be further discussed in Chapter 4.

3 Cosmic Rays

Cosmic rays (CR) are high-energy atomic nuclei that originate from both galactic and extragalactic sources. When these particles enter the Earth's atmosphere, they interact with atmospheric nuclei, producing extensive air showers composed of secondary particles. The study of CRs provides fundamental insights into the nature and location of their sources, the mechanisms responsible for their acceleration to extreme energies, and the processes governing their propagation through magnetized environments within and beyond our galaxy.

3.1 Cosmic Ray Flux

The Earth is continuously bombarded by CRs of vastly different energies, with the highest energy observed above 10^{20} eV [20]. The flux of incoming CRs, shown in Figure 3.1, can be mainly described as a power law that extends over twelve orders of magnitude of the form $E^{-\gamma}$ with E being the CR energy and γ the spectral index. The spectral index takes a value around 3, with corrections occurring in the high-energy regions ($> 10^{15}$ eV). Each of these regions is separated from the rest through a change in the steepening, hinting at possible galactic acceleration limits, extragalactic contributions, and their respective acceleration limits. Due to the extragalactic nature of ultra-high-energy cosmic rays (UHECR), as well as their small deflection angle by magnetic fields, the study of their energy spectrum, mass composition, and arrival direction is crucial. However, the enormous drop in the particle rate from around 10^4 particles per square meter second at 10^9 eV to only 10^{-15} particles at 10^{20} eV, makes direct observation through satellite-borne experiments basically impossible. Consequently, the only way to investigate UHECR is through the indirect measurement of the extensive air shower, created in the interaction with nuclei in the atmosphere. As a result, the mass, energy, and arrival direction are no longer directly measurable, but must be inferred from the development of the air shower.

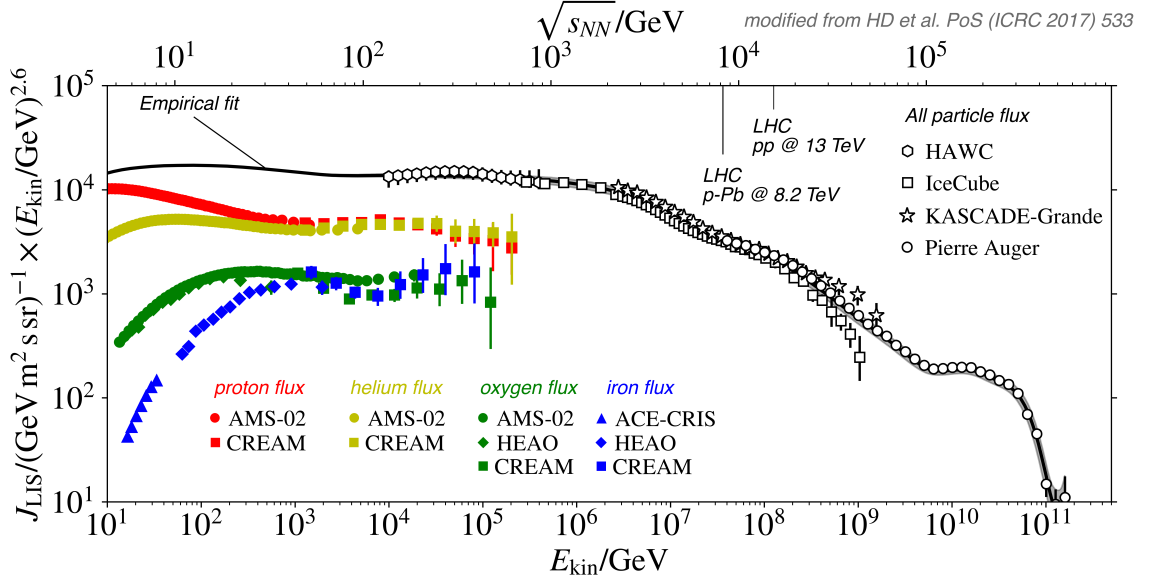


Figure 3.1: Cosmic ray flux combined from multiple measurements. Measurements for the combined flux and the flux for various nuclei are shown. The lower x-axis shows the cosmic ray energy, and the upper x-axis the converted center-of-mass energy of nucleon-nucleon interactions. The spectrum is multiplied by $E_{\text{kin}}^{2.6}$. The figure is taken from Ref. [1].

3.2 Extensive Air-Showers

When UHECRs enter the atmosphere, they interact with atmospheric nuclei and initiate a cascade of secondary particles known as an extensive air shower (EAS), analogous to the production of a jet as described in Chapter 2.2. This process converts the energy of the primary particle into a hadronic, electromagnetic, and muonic component that propagates through the atmosphere from the first interaction point to the ground, as illustrated in Figure 3.2. The resulting secondary particles at ground level can then be detected with large surface detectors, allowing the indirect study of UHECRs.

The hadronic component forms the initiating core of the extensive air shower. When a UHECR collides with a nucleus in the atmosphere, a multitude of secondary hadrons, predominantly pions and kaons, are produced. These secondaries, which still carry a significant fraction of the energy of the primary particle, undergo further inelastic collisions with atmospheric nuclei, giving rise to a hadronic cascade that propagates through the atmosphere.

At each stage of this cascade, part of the hadronic energy is converted into the

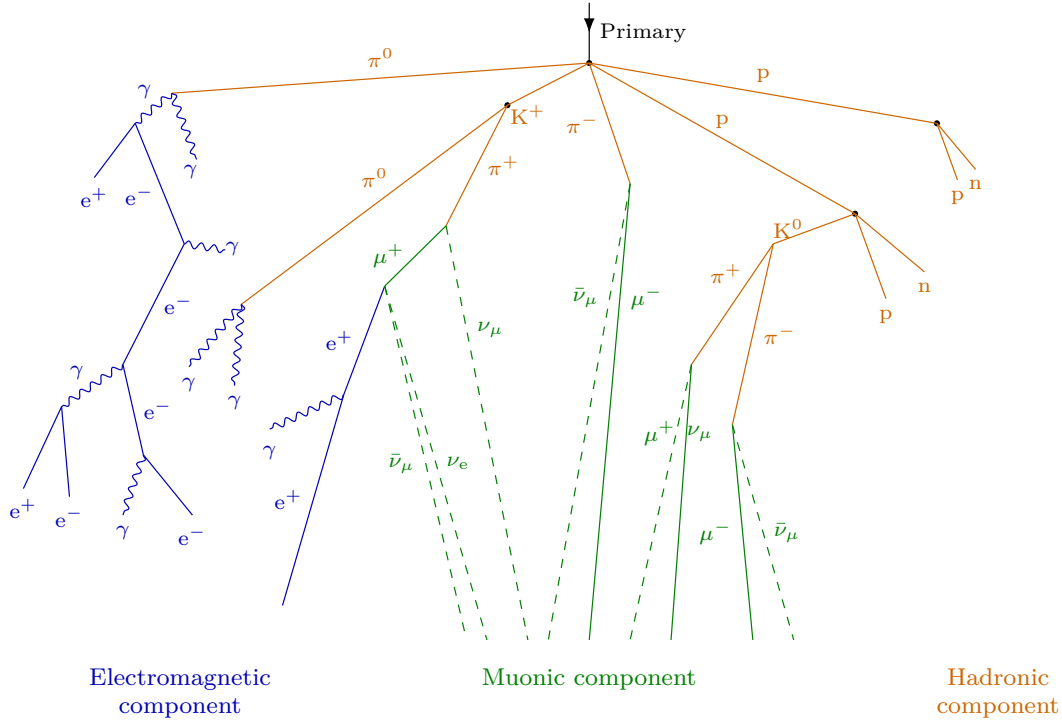


Figure 3.2: Schematic illustration of an extensive air shower, based on Ref. [21] and the original illustration in Ref. [22].

electromagnetic component mainly through the prompt decay of neutral pions into photons, $\pi^0 \rightarrow \gamma\gamma$. In contrast, charged pions and kaons can interact hadronically or decay into muons and neutrinos ($\pi^\pm \rightarrow \mu^\pm \nu_\mu$, $K^\pm \rightarrow \mu^\pm \nu_\mu$), feeding the muonic component of the shower. As the shower penetrates deeper into the atmosphere, the average interaction length of charged pions and kaons for hadronic interactions decreases. At the same time, their energies and thus their Lorentz factors decrease as the shower develops, resulting in shorter decay lengths. Eventually, decay becomes more likely than further hadronic interactions, reducing the probability of further hadron production. At the end of the shower, most hadrons either decay or are absorbed. Through these processes, the hadronic component continuously adds to the other components of the air shower.

Although the number of hadrons reaching the ground is relatively small compared to electrons and photons, the hadronic component carries a substantial fraction of the total energy. It also steers the lateral spread and development of the electromagnetic component, as well as the overall position of the shower maximum. Accurate

modeling of the hadronic interactions is therefore essential for reconstructing the primary energy and composition of UHECRs from air shower observables.

As previously mentioned, most of the electromagnetic component stems from the decay of neutral mesons, like the π^0 , which decay into two photons, initiating electromagnetic sub-showers. The electromagnetic component consists of photons, electrons, and positrons producing particle cascades through pair production ($\gamma \rightarrow e^-e^+$) and bremsstrahlung ($e^\pm \rightarrow e^\pm\gamma$). At high energies, these processes dominate over ionization losses and Compton scattering, leading to an exponential growth in particles as the shower develops. This cascade continues until the average energy of the particles falls below a critical energy, where ionization losses begin to dominate. At this critical energy, the electromagnetic component reaches the maximum number of particles, and since it dominates the number of charged particles in the shower, the distance to this point from the first interaction is called the shower maximum $X_{\max}$. The reconstruction of $X_{\max}$ in measurements is essential for the study of EAS, since its longitudinal development and shape are sensitive to the energy and mass of the primary particle.

The electromagnetic component contains the largest number of particles in the shower and, as the cascade develops, receives an increasing fraction of the primary energy. It eventually carries the dominant share of the primary energy. In addition, it is responsible for most of the observable radiation in the form of fluorescence light and radio emissions, which will be explored in more detail in Section 3.3.

The muonic component originates primarily from the decay of charged pions and kaons produced in the hadronic component. Muons penetrate long distances in the atmosphere with little deflection and radiation losses, resulting in many muons reaching the ground before decaying. The number of muons at ground level N_μ gives an indicator of the total number of charged mesons in the hadronic cascade. Thus, the measurement of N_μ contains essential information about the hadronic interactions and primary particle, as heavier primary particles produce more secondary mesons.

3.3 Experimental Detection

Because UHECRs are exceedingly rare, their properties cannot be measured directly. Instead, their properties are inferred from the extensive air shower produced in the atmosphere. Their detection relies on the measurement and detection of secondary particles and emitted radiation. Experiments employ various complementary methods, with the prominent ones being the use of Cherenkov, Fluorescence, and Radio detectors.

3.3.1 Cherenkov Radiation

Cherenkov radiation is emitted when a charged particle travels through a medium with a velocity v greater than the local phase velocity of light, $v > c/n$, where n is the refractive index of the medium, and c is the speed of light in a vacuum. Through this, charged particles in the atmosphere can emit Cherenkov radiation in the form of a forward cone at an angle θ , given by

$$\cos(\theta) = \frac{1}{\beta n}, \quad (3.1)$$

with $\beta = v/c$.

One way to detect particles with Cherenkov radiation is by using the atmosphere as a detector, measuring the radiation produced by the traversing particle. This technique is used, for example, in imaging atmospheric Cherenkov telescopes (IACT) to measure high-energy gamma rays. However, due to the large spread of air showers induced by UHECR and the small field of view of IACTs, they are not suited to measure UHECR. Consequently, the Pierre Auger Observatory, an observatory designed to detect UHECR, uses water-Cherenkov detectors [23]. These detectors, shown in Figure 3.3, are water tanks with a 3.6 m diameter and contain 12000 liter of ultra-pure water [23]. When a charged particle penetrates the water tank and produces Cherenkov radiation, the radiation is detected through photomultipliers on the inside of the tank. With 1660 detectors placed in a grid of about 3000 km², the Pierre Auger Observatory is able to infer the primary energy through the deposited signal strength and the shower direction through the arrival time of the secondary particles in the tanks. Alternatively to water-Cherenkov detectors, the Telescope Array experiment [24], also designed to observe UHECRs, uses plastic scintillator detectors to measure the energy deposited by charged particles on the surface.

3.3.2 Fluorescence Emission

Charged particles in an air shower can excite nitrogen molecules in the atmosphere, which results in isotropic light emitted in a wavelength range of about 300-400 nm. The emitted fluorescence light is proportional to the local energy deposition, allowing a reconstruction of the longitudinal shower profile and primary energy. Due to this, experiments like the Pierre Auger Observatory and the Telescope Array employ fluorescence telescopes to measure the energy deposit per atmospheric depth dE/dX and shower maximum $X_{\max}$. However, the fluorescence photon yield is low, so observations require dark, cloudless nights to minimize background noise, reducing the duty cycle of fluorescence detectors to roughly 15%.



Figure 3.3: Photo of a Cherenkov detector used in the Pierre Auger Observatory. Picture is taken from Ref. [25].

3.3.3 Radio Emission

The radio emission emitted in air showers consists of different sources. The main contribution is of geomagnetic origin, emitted by a time-varying transverse current arising from the changing number of charged particles during the development of an air shower, together with the constant acceleration of charged particles in the Earth's magnetic field and the deceleration due to atmospheric interactions. The result is a radio emission which is polarized linearly in the direction of the Lorentz force, with the same polarization direction for each observer on the ground, independent of their position relative to the shower axis.

Another smaller contribution to the radio emission of air showers comes from the Askaryan effect [26]. This effect originates from the net negative charge in the shower front, which arises through the annihilation of positrons and the knock-out of electrons from molecules in the atmosphere. In contrast to the first contribution, the polarization is oriented radially in the direction of the shower axis, resulting in a polarization direction that is dependent on the relative position of the observer. These processes generate coherent radio emissions typically observed below the 100 MHz range, with some frequencies reaching the GHz range, due to Cherenkov-like effects in the shower front.

As described in Ref. [27], radio detection has become an established method for

observing UHECR-induced air showers. Not only do radio detectors have a duty cycle of almost 100%, but the radio emission is also detectable through broadband antennas, which are inexpensive compared to other large-scale detector components. Due to the coherent nature of the radio emission, the radiation energy scales approximately quadratically with the primary energy. The shape and slope of the lateral profile are sensitive to X_{max} . Due to the anisotropy of the radio emissions, the radio footprint, which can be as large as several 10 km [27], must be sampled with multiple antennas.

3.4 The Muon Puzzle

A persistent issue in the interpretation of UHECR data is the so-called muon puzzle. As described in Ref. [1], this term refers to the systematic excess of muons observed in extensive air shower measurements compared to predictions from current hadronic event generators. The discrepancy appears in data collected by several experiments, including the Pierre Auger Observatory and the Telescope Array [28, 29]. Despite the use of different detection techniques, reconstruction methods, and energy ranges, all report muon numbers higher than expected from simulations. To compare the measurements between different experiments, without having to take potential biases and mismodelling of the resolution into account, the muon scale parameter z is defined as

$$z = \frac{\ln\langle N \rangle - \ln\langle N \rangle_{\text{p}}}{\ln\langle N \rangle_{\text{Fe}} - \ln\langle N \rangle_{\text{p}}}, \quad (3.2)$$

with $\langle N \rangle$ being the average muon number, and $\langle N \rangle_{\text{p}}$ and $\langle N \rangle_{\text{Fe}}$ being the average muon number from simulations of protons and iron nuclei as primary particles after experiencing the full detector simulation and same analysis as the actual events. As shown in Figure 3.4, for each of the shown hadronic event generators most of the data reach z -values above the allowed band between 0 and 1, being defined by the lightest and heaviest primary particles. The shown gray band depicts the value of z_{mass} , which is the expected value of z for a primary mass composition predicted through independent X_{max} measurements. The dashed line shows z_{mass} values taken from the GSF model of Ref. [30], which is primarily derived from X_{max} measurements.

The majority of muons in an air shower originate from the decay of charged pions and kaons produced in these interactions. Therefore, the muon content is closely linked to the hadronic component of the shower, in particular to the fraction of the primary cosmic-ray energy transferred to hadrons rather than to the electromagnetic component. Hadronic event generators try to describe these non-perturbative

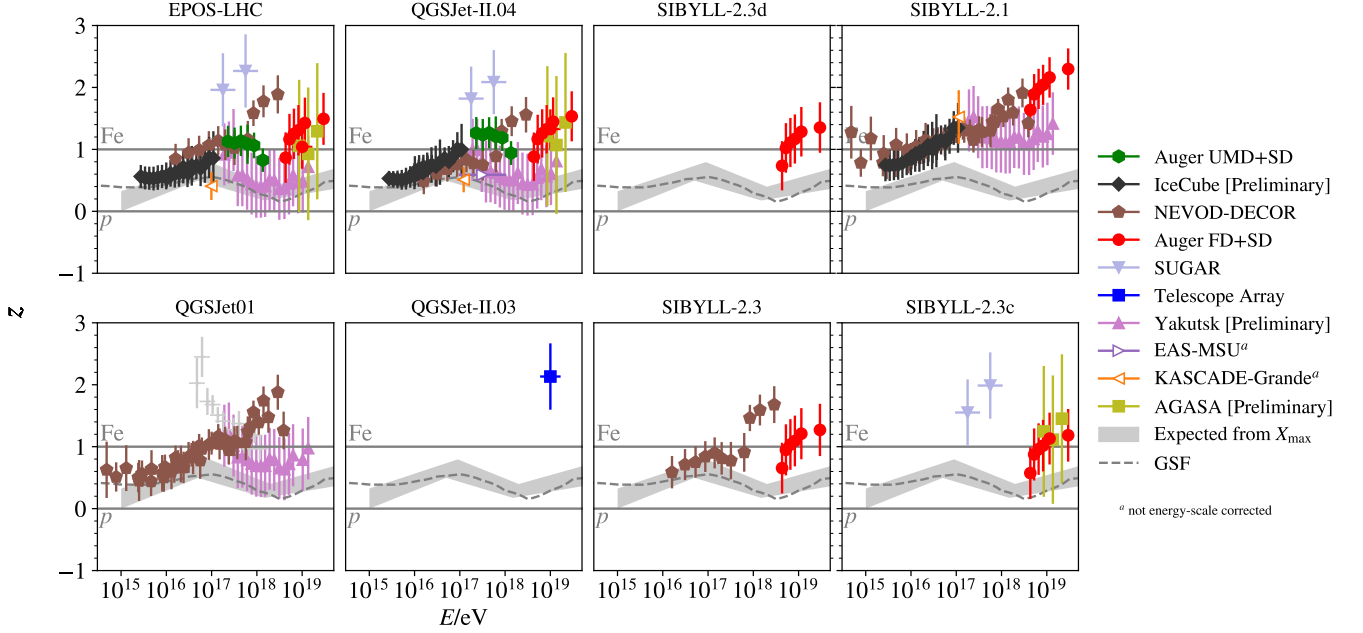


Figure 3.4: Muon measurements converted to the muon scale parameter z scale from different experiments. The results are shown for multiple hadronic event generators. The lines labeled with Fe and p show the range defined by having protons or iron nuclei as primary particles. The gray band and dashed line depict the z_{mass} value prediction based on X_{max} measurements and the GSF model [30], respectively. Figure taken from Ref. [1]

processes using extrapolations from accelerator and air shower data. However, due to the high energy, sparsity, and forward-boosted direction of UHECRs, the data used to tune these models do not cover the necessary phase space. Even small deficiencies in the modeling of multiparticle production can lead to significant deviations in predicted muon yields, resulting in the muon puzzle at high energies.

The difference $\Delta z = z - z_{\text{mass}}$ is a measure for the relative muon deficit and shown in Figure 3.5. An energy-dependent linear increase is visible, which depends only weakly on the choice of hadronic event generator. The start of the deviation around 40 PeV implies that air showers below that energy are described sufficiently, and the deficit in simulations is probably happening in the high-energy early stage of the air shower.

The onset of the muon puzzle at around 40 PeV corresponds to a collider center-of-mass energy of about 8 TeV. Since this energy is reached at the LHC, LHC

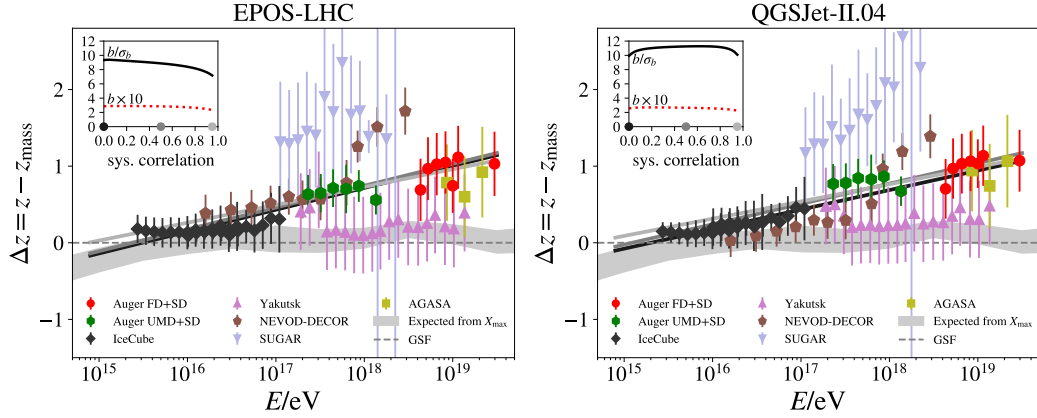


Figure 3.5: The difference $\Delta z = z - z_{\text{mass}}$ in dependence of the primary energy for various muon measurements from different experiments. The results are shown for two different hadronic event generators. The values for z_{mass} are taken from the GSF model ([30]) while the gray band shows the z_{mass} from X_{max} measurements. The slope of the experimental data is fitted for different assumed correlation levels of the experimental uncertainties. Figure taken from Ref. [1]

measurements can provide valuable input for understanding the discrepancy.

3.5 Difference Between Air-Shower and Collider Events

Air-shower interactions differ from collider events in several fundamental ways. At the LHC, proton–proton collisions take place at fixed and precisely known centre-of-mass energies, and the interaction point is confined to a small, well-defined region. In air showers, the first collision happens at a random altitude and with a random geometry, involving a nucleus in the atmosphere. This introduces substantial fluctuations in the starting conditions and in how the cascade develops.

The energy scales also differ. Cosmic rays can reach far higher energies than any accelerator, but the flux at these energies is extremely low, which limits statistics. Collider data, by contrast, are abundant but restricted to much lower energies. As a result, air-shower simulations must often rely on extrapolations from well-measured but less energetic collider regimes into a region where little direct data exist.

A further important distinction is the role of the forward region, illustrated in Figure 3.6. In air showers, a large fraction of the primary energy is carried by hadrons produced at small angles with respect to the incoming particle direction during the first interaction. These forward particles largely determine the subsequent shower development, including the longitudinal profile, the depth of shower maximum, and the muon content at ground level. In contrast, collider measurements and generator tunes are primarily constrained by central observables, where detector acceptance and precision are generally greatest. This difference helps explain why event generators require dedicated tuning for cosmic-ray applications. This issue will be discussed in more detail in Chapter 8.1.



Figure 3.6: Illustrations of an event in a collider configuration with particles of equal energy interacting from opposite directions, resulting in particle production predominantly in the central region (left), and of an event in a fixed-target configuration with the particle production in forward direction (right).

4 Event Generation with Monte Carlo Simulations

Monte Carlo (MC) simulations are an essential tool for physics, offering a statistical framework to model complex processes that cannot be treated analytically. In particle physics, they are used to simulate the probabilistic nature of particle interactions, decays, and hadronization, enabling detailed comparisons between theoretical predictions and experimental data. Event generators implement these methods to produce realistic representations of particle collision events based on current theoretical understanding and effective models. The same approach is used in air shower simulations, where these generators model the high-energy hadronic interactions that occur when cosmic rays strike the atmosphere. By coupling such event generators with dedicated shower propagation codes, one can reproduce the full cascade of secondary particles and compare simulated observables with measurements from experiments.

This chapter first discusses the main components of event generation: the hard scattering process, the parton shower, and the hadronization stage. It then describes the Lund string model as implemented in `Pythia 8` [2]. Finally, the chapter introduces the `Rivet` framework [3], which provides a systematic way to analyze and compare simulated results with experimental data, and `CORSIKA 8` [4, 5], a widely used framework for simulating air showers.

4.1 Hard Process

The hard process represents the fundamental interaction at the core of each simulated event. It involves a short-distance scattering between the elementary constituents of the incoming particles, typically quarks and gluons in the case of hadron collisions. The characteristic momentum transfer in such processes is large compared to the QCD confinement scale, making the interacting particles asymptotically free, which allows the use of perturbative quantum chromodynamics (pQCD) to calculate the interaction probabilities.

At this stage, the event generator evaluates the matrix element for the selected process. These matrix elements are derived from the relevant Feynman diagrams and depend on the type of interaction. The squared matrix element, averaged and

summed over the initial- and final-state quantum numbers, enters the calculation of the differential cross section together with the relevant phase-space factors, flux normalization, and parton distribution functions in the case of hadron collisions.

The result of the hard process stage is a set of outgoing particles with well-defined flavors, color connections, and momenta. These serve as the initial configuration for the subsequent parton shower evolution, where QCD radiation and further branchings are modeled.

4.2 Parton Showers

Parton showers describe the perturbative evolution of the partons produced in the hard interaction as they undergo successive splittings, thereby generating additional quarks and gluons. They offer a probabilistic framework to approximate higher-order QCD radiation, capturing the dominant soft and collinear emissions that shape the structure of hadronic final states [31].

The shower develops as each parton successively radiates new partons, ordered by a decreasing evolution scale such as the emission angle or transverse momentum. This ordering ensures a physically consistent treatment of the radiation and incorporates coherence effects, where interference between different color-connected partons suppresses wide-angle soft emissions. Modern event generators use different evolution variables and shower formalisms. For example, *Pythia 8* employs a transverse-momentum-ordered shower, while other generators may use angular ordering or other approaches.

Parton showers are used both for final-state radiation (FSR) from the outgoing partons produced in the hard scattering and for initial-state radiation (ISR) from the incoming partons before the collision. In the latter case, a backward evolution algorithm is used to remain consistent with the parton distributions of the initial hadrons.

In hadron collisions, modern event generators also simulate multiple parton interactions (MPI), corresponding to additional semi-hard scatterings in the same collision. Together with beam remnants and associated radiation, MPI contribute significantly to the underlying event, increasing the overall particle multiplicity and energy flow in the final state, and are essential for a realistic description of hadronic collisions.

The shower evolution ends at a scale of order 1 GeV, where the strong coupling becomes large and perturbative calculations are no longer valid. The resulting low-energy partons, connected by color flow, provide the starting point for hadronization.

4.3 Hadronization

In this non-perturbative regime, colored partons cannot exist as free particles and must combine into color-neutral hadrons, a process known as hadronization. Since hadronization cannot be calculated from first principles in QCD, event generators rely on phenomenological models tuned to experimental data.

Hadronization models provide an effective description of how the color field between partons converts into hadrons. They aim to preserve global conservation of energy, momentum, and quantum numbers while reproducing measured particle multiplicities and momentum spectra. Two main approaches are commonly used. The string model, employed in **Pythia 8**, treats color-connected partons as endpoints of a color flux tube that breaks through the creation of quark–antiquark pairs from the vacuum, leading to a chain of hadrons. The cluster model, employed in generators such as **Herwig** [32], instead groups nearby color-connected partons into color-singlet clusters that subsequently decay into hadrons.

Although conceptually different, both models implement the same physical principle: the non-perturbative conversion of colored partons into color-neutral hadrons. Their parameters are determined through extensive comparisons with collider data, ensuring that simulated hadronic final states accurately reproduce experimental observations across a wide range of processes and energies.

4.4 Pythia 8 Lund String Model

General-purpose event generators such as **Pythia 8** simulate all stages of a high-energy collision, from the initial hard interaction and parton showers to MPI and hadronization. The non-perturbative transition from colored partons to color-neutral hadrons is modeled through the Lund string model, which provides an effective description of color confinement.

In this model, color-connected partons are joined by a linear color flux tube, or string. As partons separate, the potential energy stored in the string increases linearly with distance until it becomes energetically favorable to produce a new quark–antiquark pair. As illustrated in Figure 4.1, this process breaks the string into smaller segments, each forming a hadron when a new quark or antiquark combines with one of the original string ends. A schematic example of such an event is shown in Figure 4.2, where a proton–proton collision producing a $t\bar{t}$ pair results in a dense network of color strings stretching between the many color-connected partons. The

hadronization of these strings leads to the large number of hadrons observed in the final state.

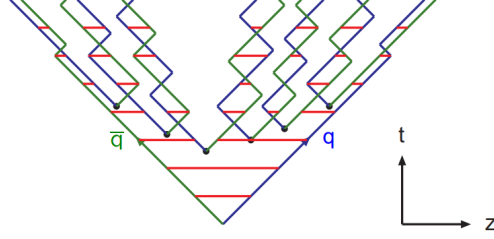


Figure 4.1: Sketch of a connecting string between a quark-antiquark pair in the time-space (t - z) plane. Figure taken from Ref. [33].

4.4.1 String Fragmentation Function

The longitudinal momentum distribution of the produced hadrons is described by the Lund symmetric fragmentation function,

$$f(z) \propto \frac{(1-z)^{a_{\text{Lund}}}}{z} \cdot \exp\left(-\frac{b_{\text{Lund}} m_{\perp}^2}{z}\right), \quad (4.1)$$

where z is the fraction of momentum transferred to the new hadron after string breaking, and $m_{\perp}^2$ is its transverse mass. The parameters a_{Lund} and b_{Lund} control the shape of the fragmentation spectrum, as visualized in Figure 4.3. As described in Ref. [34], a_{Lund} suppresses the hard region $z \rightarrow 1$ for large values, while large b_{Lund} values suppress the soft region $z \rightarrow 0$. Through their adjustment, the fragmentation spectra can be modified. Throughout this thesis, b_{Lund} will be treated as dimensionless, although it formally has the dimension GeV^{-2} .

Because QCD provides no direct prediction for these parameters, they are tuned to experimental data.

4.4.2 Transverse Momentum Generation

At each string breaking, a new quark-antiquark or diquark-antidiquark, which are pairs of two quarks or antiquarks to create baryon-antibaryons during hadronization, is created. The produced particles are assigned transverse momenta equal in magnitude and opposite in direction, ensuring local momentum conservation at

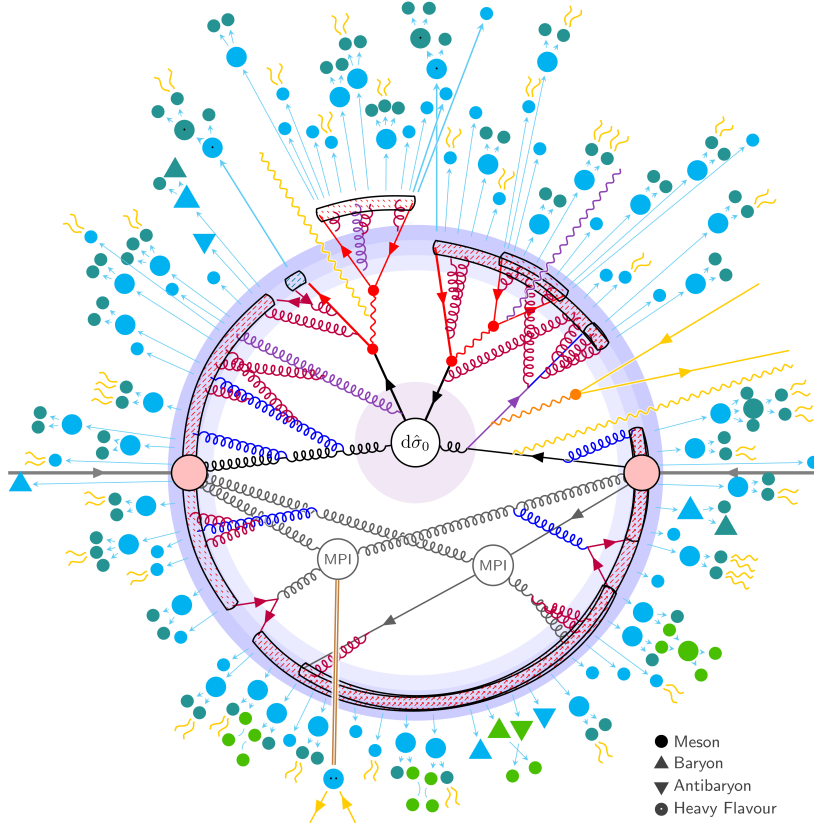


Figure 4.2: Illustration of an event initiated through a proton-proton collision producing a $t\bar{t}$ pair. Figure taken from Ref. [2].

every break. The two transverse components, p_x and p_y , are generated independently according to Gaussian probability distributions. The probability for a quark to acquire a given transverse momentum (p_x, p_y) can therefore be written as

$$dP \propto \exp\left(-\frac{p_x^2 + p_y^2}{2\sigma_q^2}\right) \quad (4.2)$$

where σ_q is the standard deviation of the one-dimensional Gaussian distribution in either transverse direction.

The average squared transverse momentum of a quark is then

$$\langle p_T^2 \rangle = \langle p_x^2 \rangle + \langle p_y^2 \rangle = 2\sigma_q^2 \equiv \sigma^2, \quad (4.3)$$

with the total width σ being an adjustable parameter in Pythia 8.

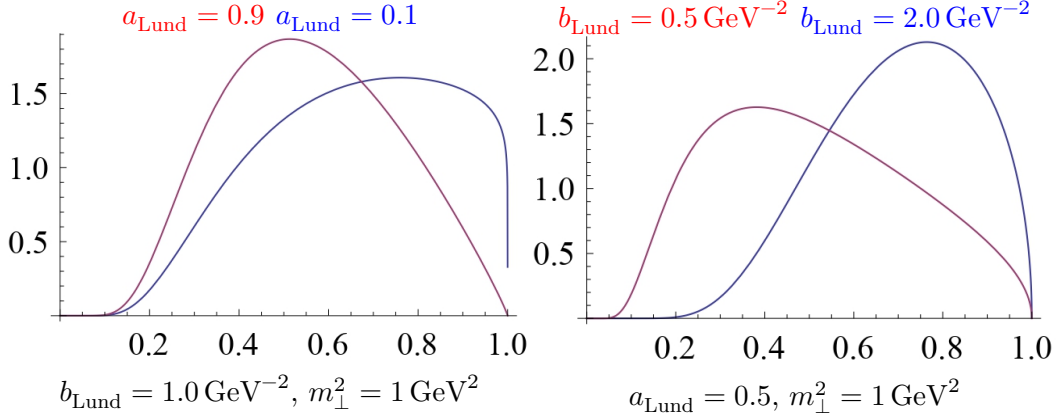


Figure 4.3: Lund fragmentation function for two different values of the a_{Lund} parameter (left) and the b_{Lund} parameter (right). The function is normalized to unity. Adapted from Ref. [34].

Because a hadron typically forms from quarks originating in two neighboring string breakings, it receives transverse momentum contributions from both sides. Consequently, the average squared transverse momentum of a hadron becomes

$$\langle p_{\text{T}}^2 \rangle_{\text{had}} = 2\sigma^2. \quad (4.4)$$

A larger value of σ produces harder transverse momentum spectra, as hadrons on average acquire greater p_{T} . Since the total available energy is fixed, this results in a corresponding reduction in overall particle multiplicity. The σ parameter is therefore an important variable to tune **Pythia 8**, influencing jet widths, hadron multiplicities, and the overall transverse activity in the event.

4.4.3 Multi-Parton Interactions

In high-energy hadron-hadron collisions, several pairs of partons from the incoming hadrons can undergo separate scattering within a single event. These MPIs contribute significantly to the underlying event and are an essential component of modern event generators such as **Pythia 8**.

At leading order, the QCD $2 \rightarrow 2$ parton-parton cross section behaves as

$$\frac{d\hat{\sigma}}{dp_{\text{T}}^2} \propto \frac{\alpha_s^2(p_{\text{T}}^2)}{p_{\text{T}}^4}, \quad (4.5)$$

which diverges for $p_{\text{T}} \rightarrow 0$. This infrared divergence is unphysical because interactions at small momentum transfers are suppressed by color screening between

neighboring partons. To model this effect, **Pythia 8** introduces a regularization scale p_{T0} that softens the divergence:

$$\frac{d\hat{\sigma}}{dp_T^2} \propto \frac{\alpha_s^2(p_T^2 + p_{T0}^2)}{(p_T^2 + p_{T0}^2)^2}. \quad (4.6)$$

This modification ensures a finite total cross section even at small p_T and provides a smooth transition between the perturbative and non-perturbative regimes. Figure 4.4 shows both the divergence of the differential cross section at low p_T , and the dampened form used in **Pythia 8**.

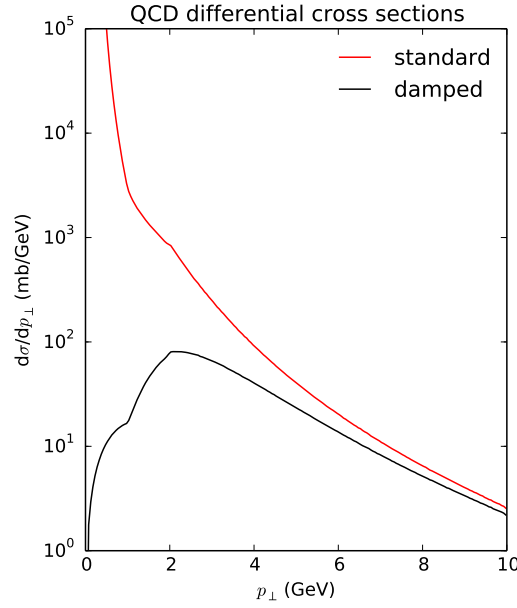


Figure 4.4: Differential cross section for QCD $2 \rightarrow 2$ interactions at 13 TeV, obtained through standard perturbation theory and after multiplication with a damping factor. Figure taken from Ref. [34].

For MPI, the cutoff scale p_{T0} is not fixed but depends on the center-of-mass energy E_{CM} of the collision

$$p_{T0}(E_{CM}) = p_{T0}^{\text{Ref}} \left(\frac{E_{CM}}{E_{CM}^{\text{Ref}}} \right)^{\text{ecmPow}}. \quad (4.7)$$

The energy-scaling power **ecmPow** and especially the reference value p_{T0}^{Ref} are key tuning parameters that determine the relative amount of soft and semi-hard activity in the event. The value E_{CM}^{Ref} is a reference energy scale at which the parameter p_{T0}^{Ref}

is defined. It sets the normalization point for the energy scaling and does not carry direct physical meaning beyond anchoring the power-law behavior. In this work, $E_{\text{CM}}^{\text{Ref}}$ is set to 7 TeV, as it reflects the typical collision energy of LHC data at the time of early tunes of **Pythia 8**. For a center-of-mass energy of $E_{\text{CM}} = 7 \text{ TeV}$, the cutoff scale $p_{\text{T}0}$ is equal to $p_{\text{T}0}^{\text{Ref}}$.

To account for the spatial structure of the proton, **Pythia 8** models MPI within an impact-parameter framework, in which the number of additional parton-parton scatterings depends on the transverse overlap of the colliding hadrons. **Pythia 8** generates MPI together with ISR and FSR within a p_{T} -ordered evolution. This unified treatment ensures that the hardest emission or interaction is generated first and provides a consistent description of radiation and secondary parton-parton scatterings. While MPI play a central role in the description of the underlying event at colliders, they become even more important at the ultra-high energies relevant for cosmic-ray air showers, where smaller parton momentum fractions are probed, leading to higher parton densities. These additional scatterings significantly affect secondary-particle production, multiplicities, and the overall development of the shower.

4.5 Rivet

Rivet is a framework for testing and comparing event generators against published collider measurements in a clean and reproducible way [3]. It provides a large library of analysis modules that encode the exact selection criteria, observable definitions, and binning used by the experiments. Using unfolding techniques, detector-specific reconstruction steps are avoided. This makes comparisons straightforward: once an event sample is produced, **Rivet** fills the relevant histograms and outputs them in a standard format without extra configuration.

A key strength is the effortless comparison between different Monte Carlo models. After generating an event sample, **Rivet** runs the chosen analyses and fills histograms in the YODA format. These files store the binned distributions, statistical uncertainties, and metadata in a consistent structure. Matching the event generator output to the corresponding histograms becomes straightforward, which makes **Rivet** very effective for tuning. Users can compare several event generator outputs to each other and to measurements, which makes the framework ideal for tuning studies.

While **Rivet** excels at validating and tuning event generators against collider and fixed-target data, it is not directly applicable to full air-shower observables (such as shower maximum depth or ground-level muon densities). The main limitation is the **HepMC** event record: it cannot store the detailed particle cascades, large

particle multiplicities, and spatial density profiles needed to describe an extensive air shower. These quantities require specialized air-shower simulation frameworks, which track millions of particles through the atmosphere. Instead, one strategy, as mentioned in Ref. [35], is first to use **Rivet** to ensure that the generator correctly reproduces basic hadronic collisions in the energy and kinematic regime where data exist, and then feed that tuned generator into a dedicated air-shower simulation code. Air-shower observables, therefore, provide an essential test of how the collider-level tuning obtained with **Rivet** translates into the much more complex conditions of atmospheric cascades.

4.6 CORSIKA 8

Simulating an extensive air shower initiated by a high-energy cosmic ray involves several stages, beginning with the first interaction of the primary particle in the atmosphere, followed by secondary interactions, particle propagation through the atmosphere, and the decay of unstable particles. The most widely used framework for this purpose is **CORSIKA 8** [4, 5, 36, 37]. Its predecessor, **CORSIKA**, had been used for more than 30 years [38], but its FORTRAN-based codebase became increasingly difficult to maintain. This motivated the development of **CORSIKA 8** in C++, which began in 2015 [4]. Since then, most of the functionality of the original code has been reimplemented [39], while the new framework also enables additional features, such as the integration of **Pythia 8**, that were not feasible in the older version.

In **CORSIKA 8**, primary and secondary particle interactions are handled by Monte Carlo event generators (MCEG), which are included as hadronic interaction models. The choice of MCEG is made by the user before the shower simulation is run. Current high-energy hadronic models available in **CORSIKA 8** include, among others, **SIBYLL-2.3d** [40], **QGSJet-II.04** [41], **EPOS-LHC** [42], and **Pythia 8**. While the former models are already well established in EAS simulations, the use of **Pythia 8** in this context is comparatively recent and remains under active development. With the release of **Pythia 8.307**, the simulation of EAS improved substantially by enabling event generation at different energies without reinitializing the model for each energy, extending the range of supported primary particles, and introducing further relevant improvements [43]. In the context of tuning a hadronic interaction model, however, the main advantage of **Pythia 8** lies in its convenient access to a large number of free parameters, which can be easily adjusted, making it particularly suitable for tuning. This aspect is discussed in more detail in Chapter 7.1. For these reasons, the present work focuses on the tuning of **Pythia 8**.

Beyond the modeling of individual particle interactions by MCEGs, **CORSIKA 8** simulates the full development of the air shower particle cascade, including its electromagnetic, muonic, and hadronic components, within a realistic model of the atmosphere. The secondary particles produced in the cascade are tracked as they propagate downward, undergoing further interactions, decays, and continuous energy losses. The atmosphere itself is described by realistic density profiles, with additional effects such as the curvature of the Earth taken into account. Shower propagation is followed until particles either reach a predefined observation level or fall below specified energy thresholds. Throughout this process, a wide range of shower properties can be recorded, including the spatial and energy distributions of particles both during the cascade and at ground level. To improve computational efficiency, user-defined energy thresholds can be applied for hadrons, muons, and other particles, below which particles are no longer tracked. In addition, techniques such as thinning are employed, in which only a representative subset of secondary particles is tracked and assigned appropriate weights to account for the large number of particles produced in high-energy EAS.

In contrast to event simulations for the LHC, which are typically performed for a fixed primary particle and a single center-of-mass energy, air-shower simulations often sample a broad range of primary energies and particle types in order to more realistically reflect the complex flux of incoming cosmic rays.

Taken together, **CORSIKA 8** provides a flexible and modern framework for the simulation of extensive air showers, combining realistic shower propagation in the atmosphere with interchangeable Monte Carlo interaction models.

5 Bayesian Methods in Particle Physics

Modern particle-physics simulations depend on numerous parameters that cannot be derived from first principles and must be tuned to data. Since measurements come with statistical fluctuations, detector effects, and modeling uncertainties, a framework is needed that treats these uncertainties consistently and carries them through the tuning procedure. Bayesian methods are well-suited for this because they treat the parameters as uncertain quantities, incorporate prior knowledge, and provide direct access to the posterior distributions of the tuning parameters [44]. The following section introduces Bayes' theorem, which underlies this framework and illustrates its application in Bayesian inference.

5.1 Bayes' Theorem

Probability is often first introduced in its frequentist form, where the probability of an outcome x corresponds to the limiting relative frequency of that outcome in an infinite sequence of repeated trials. It is defined as

$$P(x) = \lim_{n_t \rightarrow \infty} \frac{n_x}{n_t}, \quad (5.1)$$

with n_x denoting the number of occurrences of event x in n_t trials. For example, a fair six-sided die rolled infinitely will show a four in one sixth of all throws. In practical applications, however, experiments cannot be repeated indefinitely, and many questions of interest cannot be interpreted through repeated trials. A typical example is the question of how likely it is to rain on the following day: the answer depends not on infinite repetition, but on expectations drawn from experience, available information, and uncertainty.

This perspective motivates a more subjective interpretation of probability. In Bayesian reasoning, probability quantifies the degree of belief in a proposition, incorporating both prior knowledge and new evidence. The mathematical foundation of this interpretation is Bayes' theorem,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (5.2)$$

where $P(A)$ and $P(B)$ denote the probabilities of the events A and B to happen, while $P(A|B)$ is the conditional probability of event A given that B is true. Equation 5.2 follows directly from equating the probability of both A and B

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A). \quad (5.3)$$

Bayesian Inference

In particle physics and many other fields, one commonly associates B with experimental data and A with a theoretical hypothesis or model. The probability $P(B|A)$ expresses the probability of observing the data assuming the model is correct and is typically expressed as a likelihood. The posterior $P(A|B)$ represents the updated probability of the model in light of the measurement, and the prior $P(A)$ encodes the belief about the model before any data is considered.

Through this concept, Bayesian inference is used to draw conclusions about the physical parameters $\vec{\theta}$ of a model M by updating prior knowledge with the data $\vec{D}$. This generalizes Bayes' theorem to the following form:

$$p(\vec{\theta}|\vec{D}, M) = \frac{p(\vec{D}|\vec{\theta}, M)p(\vec{\theta}|M)}{p(\vec{D}|M)}. \quad (5.4)$$

The theorem yields a posterior probability density that characterizes the model parameters. The denominator $p(\vec{D}|M)$, known as the evidence or marginal likelihood, is given by

$$Z = p(\vec{D}|M) = \int_{\vec{\theta}} p(\vec{D}|\vec{\theta}, M)p(\vec{\theta}|M)d\vec{\theta}. \quad (5.5)$$

While it is essential for Bayesian model comparison, evaluating this integral becomes increasingly demanding for models with many parameters or complex likelihood functions. The calculation of the integral is therefore often avoided. Instead, the posterior is often written as

$$p(\vec{\theta}|\vec{D}, M) \propto p(\vec{D}|\vec{\theta}, M)p(\vec{\theta}|M). \quad (5.6)$$

When dealing with numerous parameters, it is often not feasible to visualize the multidimensional posterior distribution $p(\vec{\theta}|\vec{D}, M)$. In these cases, marginalized posterior distributions are calculated to investigate only the posterior distribution of one parameter by integrating over the other parameters:

$$p(\theta_i|\vec{D}, M) = \int_{\theta_j} p(\vec{\theta}|\vec{D}, M) \prod_{j \neq i} d\theta_j. \quad (5.7)$$

5.2 Combining Bayesian Inference and Monte Carlo Methods

In many applications, the likelihood $p(\vec{D}|\vec{\theta}, M)$ has a complex dependence on the parameters. The resulting posterior distribution can therefore be difficult to characterize analytically, for example when computing marginal distributions, credible intervals, or expectation values. Efficient ways to explore and sample the posterior are therefore essential. A straightforward approach is to evaluate the product of prior and likelihood on a grid, but this quickly becomes infeasible in high dimensions. Several numerical sampling strategies exist to address this, including transform methods, which require invertible cumulative distribution functions and are thus limited in scope, and the more general accept–reject method. In the latter, one introduces an easy-to-sample distribution $g(x)$ and a constant $M \geq 1$ such that $f(x) \leq Mg(x)$, then draws a random point x' from $g(x)$ and a random point u from a uniform distribution between 0 and 1, accepting x' if

$$u \leq \frac{f(x')}{Mg(x')}. \quad (5.8)$$

Repeating this yields samples from the target distribution $f(x)$, though efficiency depends strongly on how well $g(x)$ approximates $f(x)$ and the chosen value of M . More advanced variants of this method exist (see Ref. [45] for more examples), but in practice, the most widely used and scalable approach for sampling posterior distributions is Markov Chain Monte Carlo, which will be discussed in the following.

5.3 Markov Chain Monte Carlo

Rather than generating independent candidates, Markov Chain Monte Carlo (MCMC) methods construct one or more chains, each consisting of a sequence of samples in which every new point depends only on the current one. The transition rule is designed such that, after a sufficient number of steps, the distribution of visited points approaches the target distribution. Since successive samples are generally correlated, however, the amount of effectively independent information is reduced. A central challenge in MCMC is therefore to choose a transition rule that allows the chain to converge reliably and explore the parameter space efficiently. Because the early part of a chain may still reflect the initial starting point, an initial burn-in period is usually discarded. After burn-in, the remaining samples trace the parameter space in proportion to the target probability density, making MCMC particularly

useful for high-dimensional or strongly constrained posterior distributions where direct sampling is impractical.

Metropolis-Hastings Algorithm

The Metropolis–Hastings algorithm formalizes the sampling process shown in the flow chart in Figure 5.1 by iteratively proposing new states x' and deciding whether to accept them based on their relative posterior probability $f(x)$. Starting from an initial point, each step draws a candidate from a proposal distribution $g(x)$ from the allowed parameter space and evaluates it through an acceptance probability

$$A(x', x_t) = \min \left(1, \frac{f(x')g(x_t|x')}{f(x_t)g(x'|x_t)} \right), \quad (5.9)$$

which determines whether the chain moves to the proposal or remains at its current position. This mechanism allows the sampler to favor regions of high probability while still occasionally exploring less likely areas, eventually reproducing the target distribution. The performance of the method is strongly influenced by the proposal distribution, which affects the acceptance rate, autocorrelation between samples, and the ability of the chain to explore the relevant regions of parameter space. The total number of iterations then controls the statistical precision with which posterior quantities can be estimated.

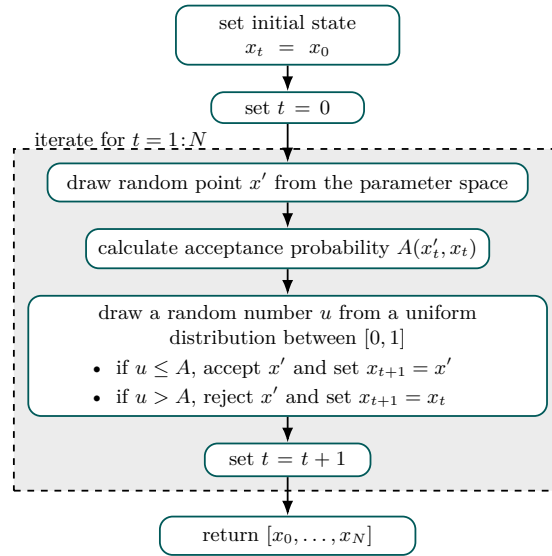


Figure 5.1: Flow chart of the Metropolis-Hastings algorithm.

6 Tuning Monte Carlo Event Generators

This chapter discusses the tuning of Monte Carlo event generators (MCEGs). After introducing various tuning approaches, a Bayesian inference-based method developed in Ref. [46] is presented. Subsequently, the tuning workflow applied in this thesis is described.

6.1 Adjusting Free Parameters

MCEGs used in particle physics typically rely on free parameters that control the generator output and can be adjusted to improve agreement with experimentally measured data. Consequently, tuning an MCEG means identifying a set of free parameters and corresponding values that provide a better description of the experimental observations.

Various tuning strategies exist. The most straightforward approach consists of manually varying parameters, simulating the physical observables of interest, comparing the Monte Carlo (MC) output to data, and repeating this procedure until agreement within an acceptable tolerance is achieved. A more advanced method is provided by the PROFESSOR program [47], which employs a parameterization-based tuning workflow to estimate optimal values for a given set of free parameters, as illustrated in Figure 6.1. Given a set of observables, their measurements from data, and a known set of parameters influencing their simulation, the workflow begins by simulating the observables at multiple randomly sampled points in the parameter space. These points serve as reference samples for a multivariate parameterization, performed bin-wise for each observable as a function of the parameters. The parameterization is then fitted to the data to obtain parameter estimates that minimize the difference to the measurements. While the overall workflow is common to parameterization-based tuning methods, the choice of parameterization and fitting procedure differs. The PROFESSOR approach utilizes polynomials of at least second order, with coefficients determined via the Moore–Penrose pseudoinverse method, as detailed in Ref. [47]. The fit to the data is performed by minimizing a chi-squared function χ^2 of the

form

$$\chi^2(\vec{p}) = \sum_{\mathcal{O}} \omega_{\mathcal{O}} \sum_{b \in \mathcal{O}} \frac{(f_{\mathcal{O},b}(\vec{p}) - N_{\mathcal{O},b})^2}{\sigma_{\mathcal{O},b}^2}, \quad (6.1)$$

where $f_{\mathcal{O},b}(\vec{p})$ denotes the parameterization evaluated at the parameters $\vec{p}$ for the observable $\mathcal{O}$ and the bin b , $N_{\mathcal{O},b}$ is the corresponding reference value, and $\sigma_{\mathcal{O},b}$ is the total uncertainty of that bin. A weight $\omega_{\mathcal{O}}$ can be assigned to each observable to emphasize those considered particularly important for the tuning.

The APPRENTICE tool is a successor to PROFESSOR and introduces several improvements [48]. While the overall workflow remains unchanged, APPRENTICE employs higher-order polynomials and polynomial ratios to more effectively parametrize MCEG responses that are difficult to approximate otherwise. In addition, it provides an automated procedure for weighting observables, improving the robustness of the tuning.

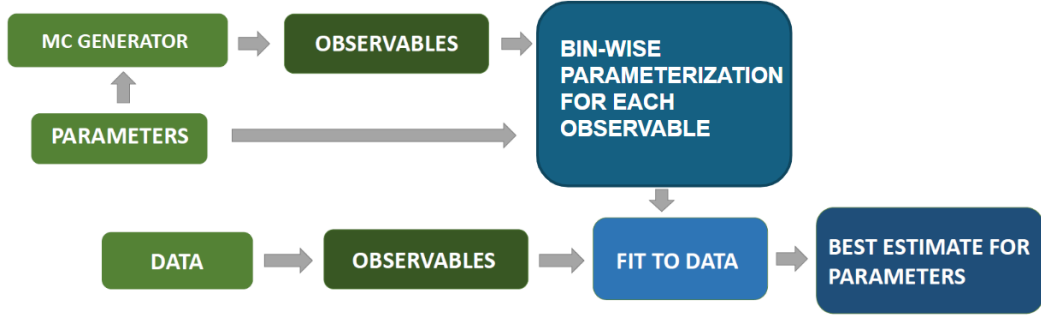


Figure 6.1: Schematic workflow of a parameterization-based tuning of Monte Carlo event generators.

6.2 Tuning Workflow Using Bayesian Inference

Another parameterization-based tuning approach relies on Bayesian inference. Analogous to the PROFESSOR workflow, this method begins by sampling the parameter space of interest to construct a parameterization of the MCEG response as a function of the parameters. The generation of these reference samples is discussed in Chapter 6.3, while the choice of a suitable parameterization strategy is addressed in Chapter 6.4. In contrast to the PROFESSOR approach, a likelihood is constructed from the parameterization and the data, which is combined with a choosable prior for the parameters and explored using an MCMC method to sample the posterior

distribution. The global mode of the posterior distribution is then taken as the best estimate of the parameter values.

Although both tuning approaches yield estimates of optimal parameter values, a key advantage of the Bayesian method is the availability of the full posterior distribution rather than only point estimates with associated uncertainties. This aspect is discussed in more detail in Chapter 6.7.

6.3 Generating Data

To ensure a sufficiently accurate parameterization of the MCEG response, reference points in the parameter space must be sampled and simulated. In this work, all MC samples are generated using the `Pythia 8` event generator. The samples are produced with the `RIVET` toolkit, as described in Section 4.5, which allows the observables of interest to be simulated for arbitrary parameter configurations. A bottleneck in this setup is the output of `Pythia 8` in the `HepMC` format, which requires substantial disk space since `HepMC` events store each particle in human-readable plain text. To avoid this issue, events are passed directly to `RIVET` as they are generated, following the approach described in Ref. [49]. The resulting `YODA` files contain the binned distributions and associated uncertainties for the set of observables required for the tuning workflow while occupying significantly less disk space.

The choice of which parameters to tune and which observables to use is highly problem-dependent. For instance, if the focus lies on the description of the underlying event rather than on rare processes, it is reasonable to tune the parameter $p_{\text{TO}}^{\text{Ref}}$, introduced in Chapter 4.4.3, using underlying-event observables measured by ATLAS and CMS. Otherwise, this parameter can be fixed to its nominal value. The number of observables and, in particular, the number of parameters must be chosen with care, as each additional parameter increases the dimensionality of the tuning problem. Since the allowed parameter phase space must be sampled densely enough to construct a reliable parameterization, adding parameters rapidly increases the number of required reference points. As an illustration, simulating three parameters on a grid with five values each results in $5^3 = 125$ reference points. Using four observables from two different `RIVET` analyses and simulating 10^6 events per reference point leads to a total of $5^3 \cdot 2 \cdot 10^6 = 2.5 \cdot 10^8$ events. Consequently, data generation constitutes the most time-consuming and computationally intensive step of the tuning workflow.

While the reference points in the example above are sampled on a regular grid in parameter space, both the sampling strategy and the parameter ranges are problem-

specific choices. The range is often initially set to the allowed parameter interval and later restricted once a region of interest has been identified. Although grid-based sampling ensures systematic coverage of the parameter space, it becomes inefficient in higher dimensions. For this reason, sampling reference points from probability distributions is also common practice, for example by drawing points from a uniform distribution as done in Ref. [46].

6.4 Constructing a Parameterization

A reliable tuning workflow requires an accurate approximation of the MCEG response. To achieve this, a multivariate parameterization, also referred to as a surrogate model, is constructed from the reference points generated in the previous step. In addition to accurately approximating the MCEG response, the parameterization must be fast to evaluate, as it is queried repeatedly during the posterior sampling process.

To ensure a successful tune, it is important to assess the approximation of the parameterization. While standard goodness-of-fit measures provide a convenient way to evaluate smooth and differentiable models such as polynomials, which come with well-defined statistical uncertainties, other parameterization methods are more difficult to assess. Interpolators defined on rectilinear grids, for example, including methods like linear and spline interpolations, often do not provide uncertainties that can be directly used in a goodness-of-fit function. In such cases, it is usually necessary to visually assess the approximation by inspecting the parameterization for individual bins. Direct visualization is only feasible for up to two parameters, where the parameterization can be represented as a surface, as shown in Figure 6.2. An alternative approach is to visualize one- or two-dimensional slices of the parameterization by fixing the remaining parameters to specific values.

Choosing a Surrogate Model

Since the quality of the parameterization is critical for the tuning, the choice of the surrogate model must be adapted to the problem at hand. Low-order polynomials are an attractive option due to their fast evaluation and analytically accessible uncertainties. However, the MCEG response of an individual observable bin is often complex and requires greater flexibility to be approximated accurately. An example of this limitation is shown in Figure 6.3, where an arbitrary bin is fitted using a cubic polynomial.

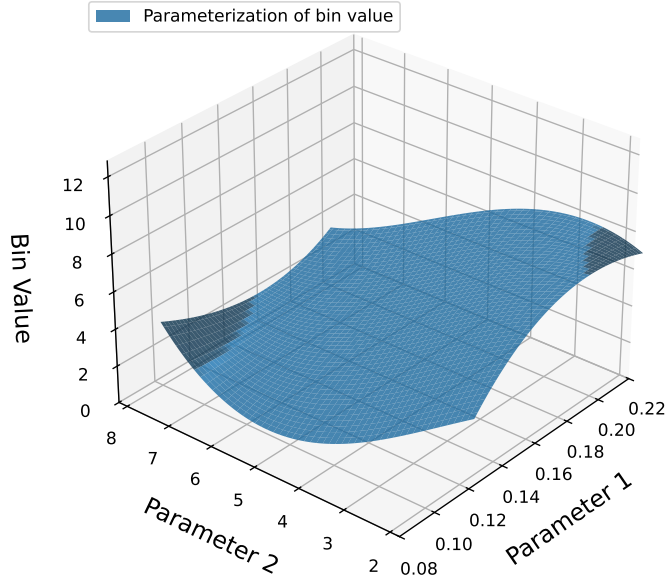


Figure 6.2: Visualization of a parameterization of a bin value of an observable for two parameters.

A natural alternative is to use machine learning models, which can fit parameterizations with arbitrary degrees of freedom. However, training such models typically requires a large number of reference points, which is often impractical given that each reference point may consist of millions of simulated events. Other approaches include ratios of polynomials, as employed by the APPRENTICE tool, and spline interpolations. Spline interpolations construct piecewise polynomials between reference points, providing an exact description at the sampled points while ensuring continuity as well as continuous first and second derivatives. In higher dimensions, however, spline interpolations tend to oversmooth regions with sparse reference points, introducing artificial curvature.

Although the methods discussed above yield differentiable parameterizations, differentiability is not a strict requirement for the tuning workflow, as only an accurate approximation of the MCEG response and fast evaluation are needed. When a sufficient number of reference points is available to cover the parameter space, linear interpolation between neighboring points is often sufficient. Delaunay triangulation provides a non-differentiable approach by partitioning scattered reference points into triangles in two dimensions or simplices in higher dimensions and performing barycentric interpolation within each simplex. While this method naturally handles irregular sampling, its evaluation becomes computationally expensive in higher

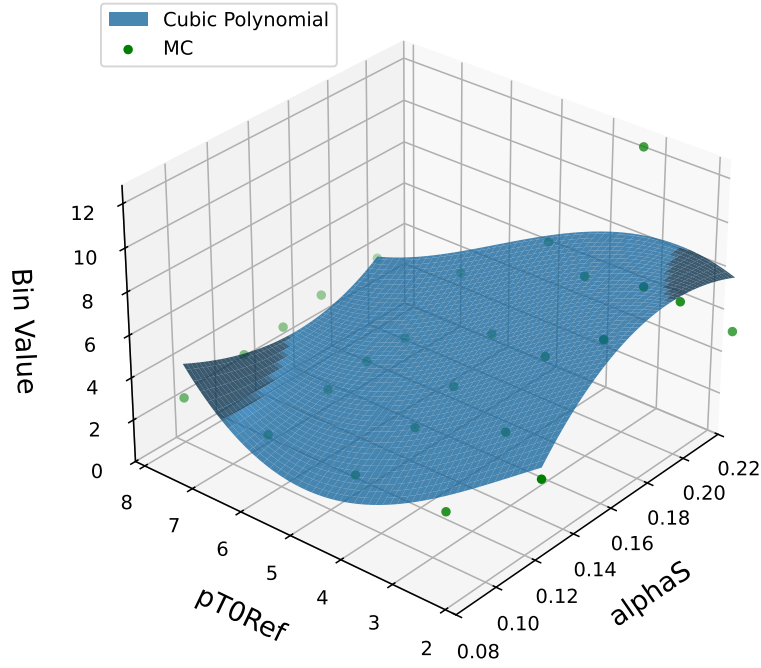


Figure 6.3: Example of parameterizing a bin value of an arbitrary observable as a function of two parameters using a cubic polynomial.

dimensions due to the required geometric searches. To reduce evaluation time, reference points can instead be simulated on a grid-like structure and linear interpolation performed between the corner points of each cell, since locating a point on a grid is computationally trivial. As illustrated in Figure 6.4, grid-based linear interpolation often provides a sufficiently accurate approximation of the MCEG response while remaining computationally efficient. Nevertheless, as discussed in Chapter 6.3, grid-based sampling becomes increasingly expensive as the dimensionality of the parameter space grows.

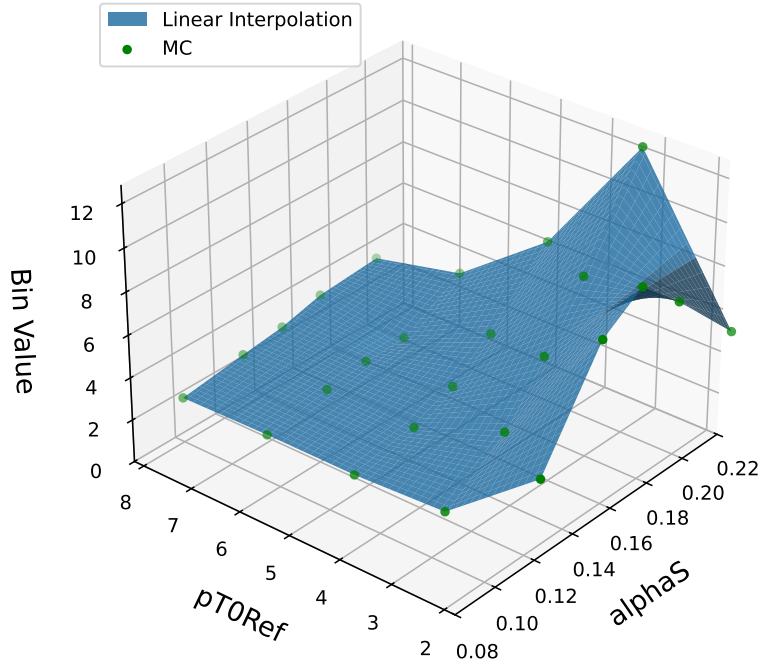


Figure 6.4: Example of parameterizing a bin value of an arbitrary observable as a function of two parameters using linear interpolation on a grid-like structure.

6.5 Fitting to Data

Once a suitable parameterization of the MCEG response has been constructed, it is combined with experimental data to define a statistical model for the tuning. Together with a prior distribution for the parameters, this statistical model is used to sample the posterior distribution of the parameters using MCMC methods. In this work, the sampling is performed using the `BAT.jl` framework [50], which provides a flexible and efficient implementation of Bayesian inference techniques. The posterior samples are then used to determine the best estimates of the tuned parameters, as well as their associated uncertainties.

The prior distribution should reflect the available prior knowledge about the parameters. In most cases, a uniform prior over the allowed parameter range is chosen, representing the absence of additional constraints. However, alternative choices are possible. For example, a normal distribution centered around a parameter value suspected to be close to the optimum can be used to guide the tuning towards a specific region of parameter space. While such informative priors can be beneficial

in certain situations, care must be taken to avoid introducing unintended biases.

The construction of the statistical model itself, that is, the definition of the likelihood function combining parameterization and data, is also a flexible choice. In this thesis, this task is performed using the `EFTfitter.jl` framework [51], which provides a convenient interface for building likelihoods from parametrized model predictions and experimental measurements.

EFTfitter.jl and BAT.jl Framework

The `EFTfitter.jl` framework is written in the Julia programming language and was originally developed for fitting effective field theory models and combining measurements from different experiments. However, the underlying likelihood construction is sufficiently general to be applied to a wide range of fitting problems, including the tuning of Monte Carlo event generators. Julia was chosen because it offers near C-level performance while providing a high-level, expressive syntax similar to Python, making it well-suited for computationally intensive statistical analyses without sacrificing code readability and flexibility.

The likelihood provided by `EFTfitter.jl` is formulated as a multivariate Gaussian distribution and can be written as

$$\ln L(\vec{D}|\vec{\lambda}) = -\frac{1}{2}[\vec{D} - \vec{f}(\vec{\lambda})]^T M^{-1}[\vec{D} - \vec{f}(\vec{\lambda})], \quad (6.2)$$

where $\vec{D}$ denotes the measured data, $\vec{f}(\vec{\lambda})$ represents the parameterization as a function of the parameters $\vec{\lambda}$, and M is the covariance matrix of the data points. In this work, however, correlations are set to zero, since many analyses do not provide full covariance matrices in their `RIVET` implementations. A more detailed discussion of the impact of neglecting correlations on the tuning result can be found in Ref. [12].

Once the likelihood has been constructed from the parameterization, the data, and the covariance matrix, it is combined with the chosen prior distribution to define the posterior distribution of the parameters. The sampling of this posterior is performed by interfacing `EFTfitter.jl` with the `BAT.jl` framework.

The `BAT.jl` framework provides a Julia-based implementation of various Bayesian inference methods, which were introduced in Chapter 5. It includes multiple sampling algorithms, among them the Metropolis–Hastings algorithm, which is used throughout this work. Before starting the sampling, the number of parallel chains and the number of steps per chain must be specified. Each chain begins with a burn-in phase, during which sampler-specific parameters are tuned to efficiently

explore the target distribution. This burn-in procedure consists of several tuning cycles, each consisting of multiple steps. After each cycle, all chains are checked for algorithm-specific requirements and for convergence using a Brooks–Gelman diagnostic [52], which compares the variance within and between multiple Markov chains to assess whether they have converged to the same target distribution. If these criteria are satisfied, the samples generated during burn-in are discarded and the actual sampling phase is started using the predefined number of steps. For the Metropolis–Hastings algorithm, the default proposal distribution in `BAT.jl` is a multivariate Student’s t distribution.

6.6 Evaluating the Tuning Result

After generating reference samples, constructing the surrogate model, and sampling the posterior distribution of the parameters, the best estimates of the tuned parameters are identified. In this work, the best estimates are defined as the global modes of the posterior distribution. An example of a two-dimensional posterior obtained for a tuning involving two parameters is shown in Figure 6.5. On the diagonal, the one-dimensional marginalized distributions for each parameter are displayed, while the off-diagonal panels show the corresponding multidimensional posteriors. Although the example shown appears approximately Gaussian, this behavior is not generic. In many cases, posterior distributions are asymmetric or exhibit multiple local maxima. As a consequence, defining an uncertainty on the best parameter estimate is not straightforward. For approximately Gaussian-like posterior distributions, the uncertainty is often defined by the width of the credible intervals.

To evaluate the quality of the tuning, new Monte Carlo samples are generated using the MCEG with the tuned parameter values set to their best estimates. The resulting distributions are then compared to the experimental data for each observable and compared with the default generator prediction or alternative tuning results. As a quantitative measure of closeness, a reduced chi-squared test is performed for each observable,

$$\chi_{\text{red}}^2 = \frac{\chi^2}{\nu} = \frac{1}{\nu} \sum_b \frac{(M_b - D_b)^2}{\sigma_b^2}. \quad (6.3)$$

where ν denotes the number of degrees of freedom, defined here as the difference between the number of bins and the number of tuned parameters. The quantities M_b and D_b are the bin contents of the MCEG prediction and the data for bin b , respectively, and σ_b is the uncertainty of the corresponding data bin.

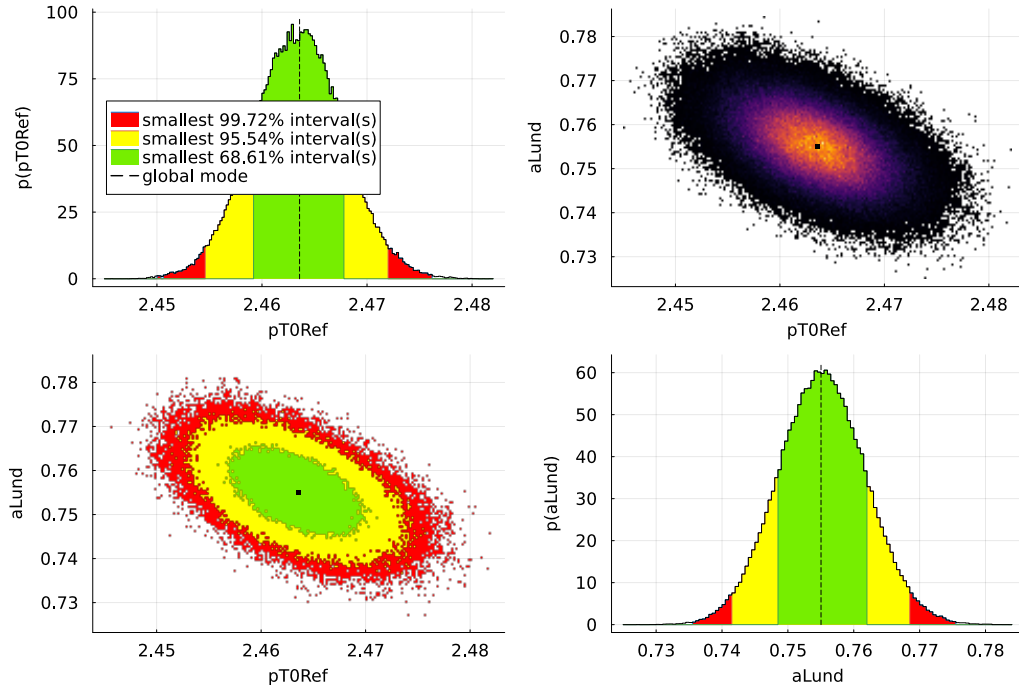


Figure 6.5: Example posterior distribution for two parameters. The diagonal panels show the one-dimensional marginalized posteriors with intervals containing 68.61%, 95.54%, and 99.72% of the posterior probability. The lower off-diagonal panel shows the corresponding two-dimensional credible regions, while the upper off-diagonal panel shows the continuous two-dimensional distribution as a heat map. The dashed lines and dots indicate the global mode.

While the reduced χ^2 provides a convenient metric for comparing different tunes for a given observable, it has important limitations. Since observables differ in shape, binning, and uncertainties, χ^2_{red} values cannot be directly compared across different observables. Moreover, a reduction in χ^2_{red} may arise from improved agreement in one region of a distribution at the expense of worse agreement in another, which is often not desirable.

6.7 Uncertainty Propagation

A major advantage of a Bayesian tuning workflow is the full access to the posterior distribution of the parameters. This not only allows the identification of secondary solutions visible as additional peaks in the marginalized posteriors, but also enables a consistent propagation of the tuning uncertainty to the simulated observables.

Uncertainty propagation is achieved by resampling the parameter posterior and evaluating the surrogate model at the corresponding parameter points. This results in a distribution of predicted values for each bin of each observable, as illustrated in Figure 6.6. The tuning-induced uncertainty on the observable is defined by the one-standard-deviation spread of this distribution around its mean. For this approach

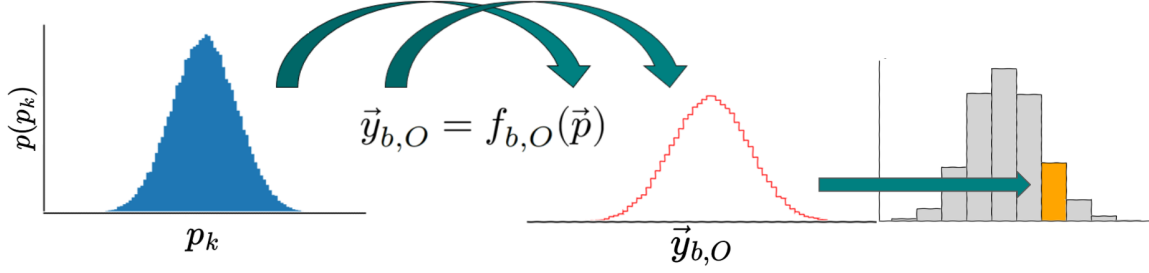


Figure 6.6: Schematic representation of propagating the uncertainty from the parameters posterior to the observables, by evaluating the surrogate model $f_{b,O}$ at resampled points $\vec{p}$ of the posterior distribution. The uncertainty of each bin in each observable is defined by the width of the resulting distribution.

to be meaningful, the surrogate model must approximate the MCEG response close enough, such that the uncertainty introduced by the surrogate model itself is negligible. Otherwise, the propagated uncertainty would be impacted heavily by the distance between the surrogate model and true MCEG response. An alternative approach would be to use the resampled posterior parameter values directly as input to the MCEG, thereby avoiding any bias from the parameterization. While this would provide a better uncertainty estimate, it comes at the cost of a substantially increased computational effort.

7 Feasibility Study of the Tuning of Pythia 8 Simulated Air Showers

This chapter investigates the feasibility of applying the tuning workflow presented in Chapter 6 to simulations of air showers induced by cosmic rays. The suitability of Pythia 8 as a viable model for this purpose is examined in greater detail. As a first step, the tuning of Pythia 8 based on a simplified model for air-shower simulations is presented. Finally, a feasibility study is carried out on tuning the Pythia 8 simulation of the atmospheric flux of high-energy muons using the CORSIKA 8 framework.

7.1 Pythia 8 as Hadronic Interaction Model for Air Shower Simulations

As already noted briefly in Section 4.6, the hadronic interaction models commonly used for simulations of cosmic-ray-induced air showers do not typically include Pythia 8. Only recently has Pythia 8 begun to explore this field in a more direct way [43]. Nevertheless, processes related to dense hadronic environments have long been studied within Pythia, in particular through the heavy-ion model **Angantyr**, which enables the simulation of collisions involving projectiles heavier than protons, relevant for some collider experiments at the LHC. One important limitation for air-shower studies, however, was that a flexible event-to-event variation of the primary energy and mass was not supported until recently in version Pythia 8.310.

This naturally raises the question of why Pythia 8 was chosen as the Monte Carlo event generator for the studies presented in this thesis. The main reason is the comparatively convenient access to its free model parameters, together with the extensive documentation of both the code and the parameter settings. This makes it straightforward to adjust free parameters within the tuning workflow, which is considerably more difficult for most of the other hadronic interaction models used in air-shower simulations. In addition, Pythia 8 is under continuous development due to its widespread use in high-energy collider physics and is regularly improved by a large community of contributors [53].

The extension of `Pythia 8` to a broader range of hadron collisions, as required for air-shower simulations, was implemented and investigated in Ref. [54] using a toy model of the atmosphere. In that work, the propagation of air showers through different simplified atmospheric models is simulated. The corresponding code is distributed with `Pythia 8` in the example program `main183.cc`. This setup provides a useful test environment in which not only the propagation of simplified air showers through basic atmospheric models can be studied, but also the tuning workflow introduced in Chapter 6 can be explored in a controlled toy-model setting.

Tuning Pythia 8 on Air-Showers in a Toy-Model Study

In the following, a toy-model study is performed using the example introduced above. This example is particularly suitable because it provides a computationally inexpensive environment: the propagation of particles through the atmosphere is treated in a simplified way, while several effects that would be relevant in a more realistic setting, such as the electromagnetic component of the shower, the Earth's magnetic field, and the curvature of the Earth, are neglected.

For the toy-model study, an atmospheric model consisting only of nitrogen with an exponentially attenuated density profile is chosen. In addition, the atmosphere has an upper cutoff of at 100 km, and the incoming cosmic-ray protons arrive vertically with a vanishing zenith angle.

Since the example only simulates air showers and does not include any experimental data to which the simulation could be compared, a suitable observable for the tuning procedure first has to be defined. For this purpose, the default configuration of `Pythia 8` is used to simulate 10^4 air showers initiated by protons with a primary energy of 10^7 GeV, to generate a distribution of the number of muons at ground level. Since no real measurement is available, these simulated events are used as pseudo-data, or fake data, for testing the tuning setup. If the workflow performs as intended, the tuned parameter value should reproduce the default setting that was used to generate these fake data.

After generating the fake data, several MC samples are produced for different values of a single `Pythia 8` parameter. In this study, the parameter a_{lund} , implemented in `Pythia 8` as `StringZ:aLund` and introduced in Chapter 4.4.1, is varied over seven values. For each parameter choice, one MC sample of the same observable is simulated, with each sample consisting of $6 \cdot 10^4$ air showers. The resulting distribution, together with the fake data, is shown in Figure 7.1.

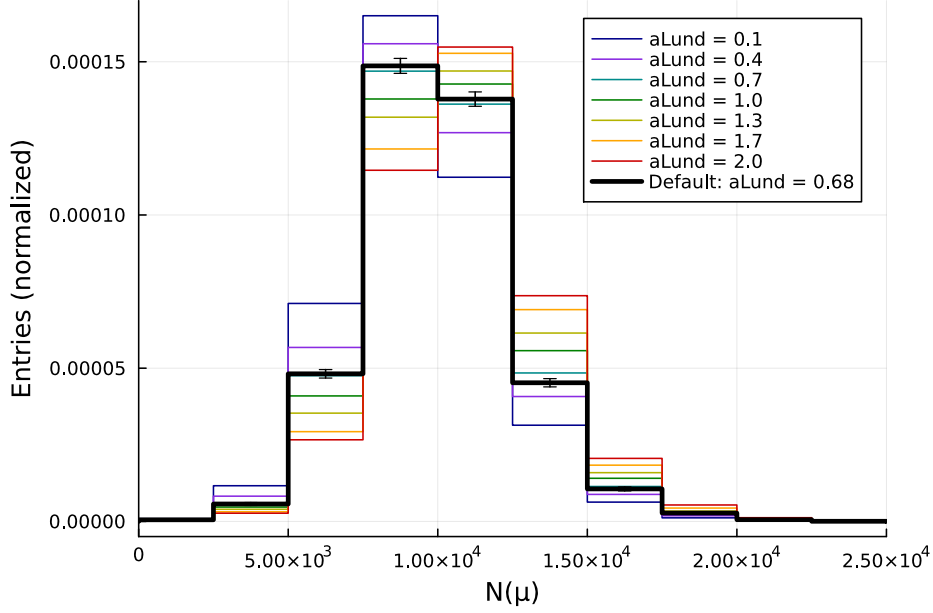


Figure 7.1: The number of muons at ground level simulated using Pythia 8 for different values of `StringZ:aLund`, including the default value.

As expected, varying the parameter away from its default value leads to systematically larger or smaller bin contents across the observable. The individual bins are parameterized independently. An example bin is shown in Figure 7.2.

The behavior shown in the figure is consistent with the expectations. The MC samples define a smooth dependence on the parameter, and the fake data point generated at the default parameter value is compatible with this curve within its uncertainties.

In the next step, the MC response is parameterized by constructing a surrogate model for each bin, describing the bin content as a function of the parameter value. For the toy-model study, a simple second-order polynomial is chosen and fitted to the MC samples. Based on this parameterization, the likelihood is then constructed and the posterior is sampled, following the procedure described in Chapter 6. The resulting posterior distribution is shown in Figure 7.3.

The posterior exhibits an approximately Gaussian shape, with its mean located close to the default value of $a_{\text{Lund}} = 0.68$. This confirms that the tuning workflow performs as intended, since the toy-model study recovers the parameter value that was originally used to generate the fake data.

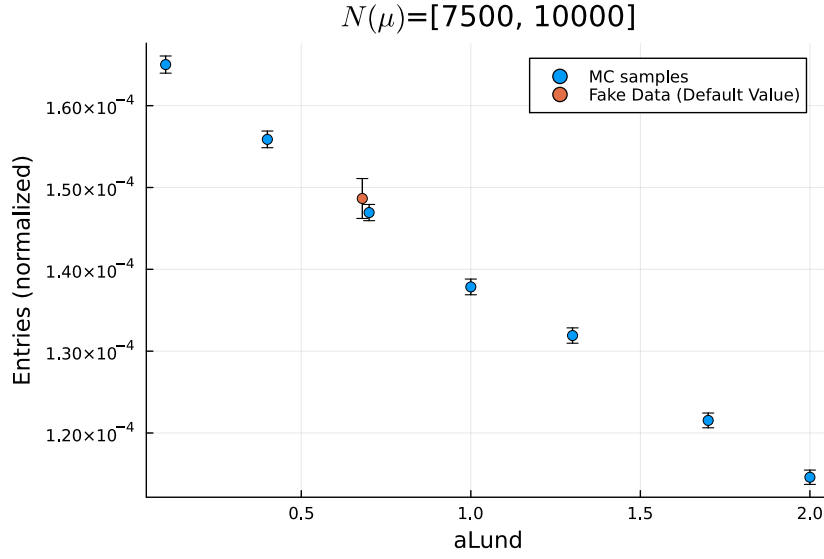


Figure 7.2: Example bin of the muon-number observable at ground level, simulated in the toy-model study as a function of `StringZ:aLund`.

Despite this successful test, several important differences from a realistic tuning of an MCEG should be kept in mind. First, replacing experimental measurements with pseudo-data generated by the same event generator removes any inherent mismodeling from the hadronic interaction model with respect to the data that cannot be corrected by adjusting free parameters alone. Second, the restriction to a single parameter represents a considerable simplification. In a realistic tuning study, each additional parameter increases the dimensionality of the parameter space and thereby also the complexity of the surrogate modeling, as well as the number of MC samples required to cover the parameter space sufficiently.

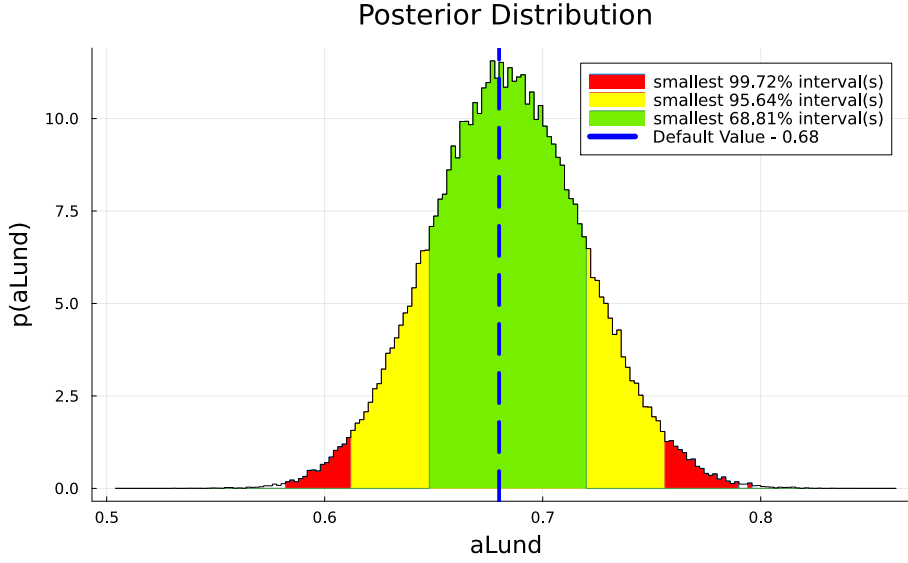


Figure 7.3: The sampled posterior for the toy-model study. The default value at which the MC sample is used as fake data in the toy-model study is marked as a blue dashed line.

7.2 Simulation of Muon Flux with CORSIKA 8

Following the successful toy-model study, a natural next step to further probe the tuning potential of **Pythia 8** in EAS simulations is the atmospheric flux of high-energy muons. This flux is produced in interactions of cosmic rays with the atmosphere and therefore contains information on both the primary cosmic-ray spectrum and the mass composition. Moreover, since atmospheric muons originate predominantly from the decay of secondary hadrons, their flux provides a direct probe of hadronic particle production in air showers. This makes it a suitable observable for investigating the impact of different **Pythia 8** parameter settings.

The data to which the **Pythia 8** simulation is intended to be tuned were measured by the IceCube Collaboration and unfolded in Ref. [55]. The atmospheric muon flux is weighted by the muon energy raised to the power 3.7, to create an approximately horizontal trend in the data. The covered muon-energy range extends from 10^4 to 10^6 GeV.

The aim of this study is to use **CORSIKA 8** with **Pythia 8** as the hadronic interaction model to simulate the atmospheric muon flux and, in a second step, to generate MC samples for different parameter settings in order to tune the event generator to the data. However, as will be discussed below, the status of **Pythia 8** in **CORSIKA 8** at

the time of this study was not yet developed enough to sufficiently simulate and reconstruct the muon flux.

Since `CORSIKA 8` simulates EASs, an appropriate energy range for the primary cosmic rays initiating the showers first has to be identified. As illustrated in Figure 7.4, muons of a given energy are predominantly produced in showers initiated by cosmic rays with energies of roughly $10 \cdot E_\mu$ per nucleon, with a long tail extending to higher primary energies. Based on this, showers with primary energies in the range 10^4 to 10^{10} GeV are simulated.

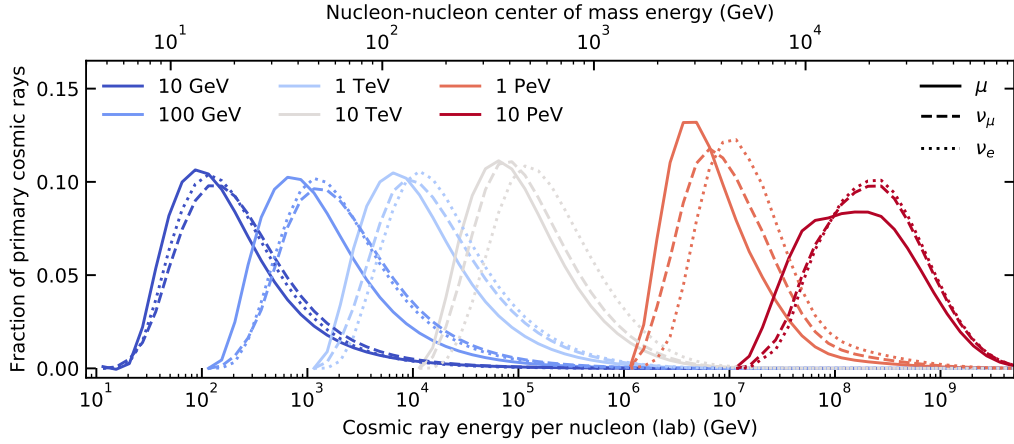


Figure 7.4: Probability density of the primary cosmic-ray energy per nucleon for leptons at several fixed lepton energies. Figure taken from Ref. [40].

At the time of this study, simulations of air showers initiated by primary particles heavier than protons were not possible in `CORSIKA 8` when using `Pythia 8` as the hadronic interaction model. Consequently, only proton-induced air showers are considered here. This is a non-negligible limitation and its full impact would require a dedicated investigation. Nevertheless, an air shower initiated by a heavier nucleus can, as an approximation, be interpreted as the superposition of several nucleon-induced showers, each carrying only a fraction of the total primary energy. This reduces the impact of the limitation to some extent, since most cosmic rays populate the lower-energy end of the spectrum, where heavier primaries tend to produce lower-energy muons than proton primaries of the same total energy.

The primary energy is sampled from a probability density proportional to E^{-1} between 10^4 and 10^{10} GeV. This choice reflects the declining trend of the cosmic-ray spectrum while still ensuring sufficient statistics at high energies. In addition, the zenith angle θ of the shower is sampled according to the isotropic arrival directions

of cosmic rays. Since the differential solid angle is given by $d\Omega = \sin\theta d\theta d\phi$, a uniform sampling in θ would oversample directions close to the poles. Instead, θ is sampled as $\theta = \arccos(u)$ with $u \sim \mathcal{U}(0, 1)$.

After the air showers are simulated with **CORSIKA 8** and **Pythia 8**, all muons reaching detector level are filled into a histogram as a function of the muon energy. Each muon is weighted with the factor

$$\omega_\mu = E_\mu^{3.7} \cdot \Phi(E_{\text{CR}}, \text{M}) / N_{\text{shower}} / P(E_{\text{CR}}), \quad (7.1)$$

where $E_\mu^{3.7}$ corresponds to the same energy weighting applied to the IceCube data, $\Phi(E_{\text{CR}}, \text{M})$ denotes the primary cosmic-ray flux for a given primary energy E_{CR} , primary particle and cosmic-ray flux-composition model M, N_{shower} is the number of simulated showers, and $P(E_{\text{CR}})$ is the earlier mentioned probability density used to sample the primary energy.

The primary cosmic-ray flux determines the occurrence probability of each simulated cosmic ray for a given primary type and energy. Several flux models are available in the literature. In this study, the **H3a** model is used [56]. The division by N_{shower} ensures that the resulting spectrum is independent of the total number of simulated showers, while the division by $P(E_{\text{CR}})$ removes the bias introduced by the chosen sampling distribution.

Because the simulation of each individual shower is computationally expensive, the electromagnetic component is not included. This component constitutes a large fraction of the computational cost, since every photon and electron/positron above a threshold energy has to be propagated and tracked through the atmosphere. Excluding this component is expected to have only a minor effect on the present study, as the electromagnetic component contributes only a small fraction of muons, and variations of hadronic model parameters are not expected to affect these muons strongly.

The resulting simulation of the atmospheric muon flux, using around 2.2×10^4 showers with about 9×10^5 weighted muons, is shown in Figure 7.5.

While the simulation reproduces the overall trend of the data, deviations remain. In particular, the discrepancies in the highest-energy bins are large. Several possible reasons may contribute to this behavior. First, a sample of only about 2.2×10^4 showers is not sufficient to model the flux with high precision.

A second important issue is the long tail in the distribution of event weights defined in Equation 7.1, which is shown in Figure 7.6. While most weights are close to zero, a small number of events receive enormous weights, reaching values of up to about

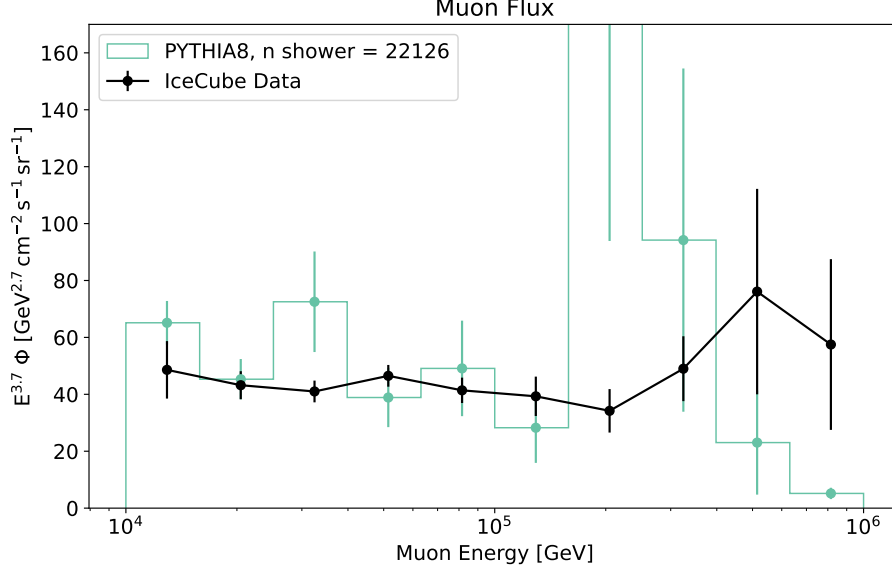


Figure 7.5: Simulated atmospheric muon flux using CORSIKA 8 and Pythia 8 as hadronic interaction model. The data is taken from Ref. [55].

100. As a consequence, individual events can dominate bins of the reconstructed flux.

To illustrate this effect, an arbitrary cut is applied to remove showers with weights above 10. Although this excludes only eight showers, it leads to large changes in some bins, as shown in Figure 7.7. However, without a clear physical understanding of the origin of these large weights, such a cut cannot be justified. Increasing the simulated statistics could help to obtain a more stable approximation of the expected flux. This may reduce the sensitivity of the result to individual showers with unusually large weights.

Although this study is theoretically appealing, since it directly connects the tuning workflow to EAS simulations with Pythia 8, it is not pursued further in this work. The reconstruction and tuning of the atmospheric muon flux based on large numbers of simulated air showers is computationally too demanding, especially since different primary energies, mass compositions, and incoming directions need to be covered sufficiently.

A more suitable approach for future studies may be the use of dedicated tools for the calculation of cosmic-ray-induced particle fluxes in the atmosphere. One example is the MCEq toolkit [57], which models the evolution of particle densities through

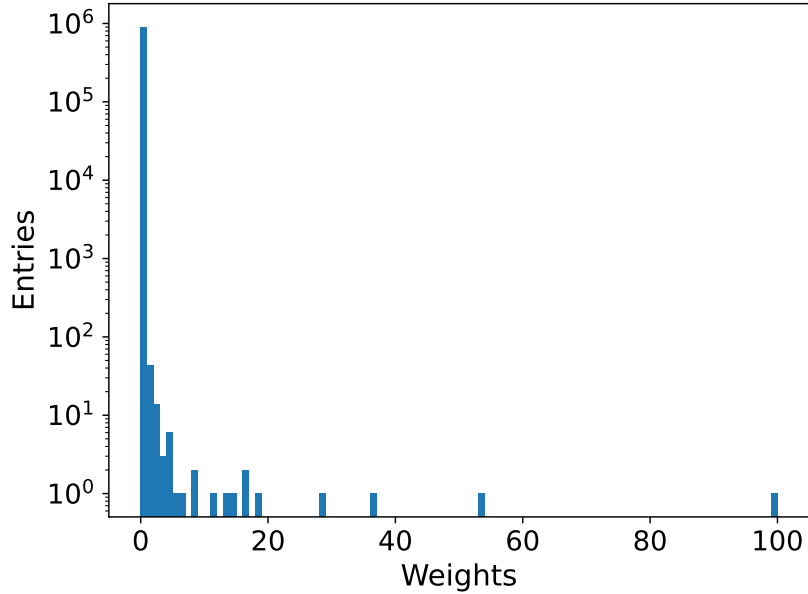


Figure 7.6: Distribution of the weights assigned to the simulated muons.

the atmosphere by solving dedicated cascade equations. In this way, average fluxes can be obtained without shower-to-shower fluctuations, in contrast to *CORSIKA 8*, which requires the simulation of many showers to achieve comparable statistical precision. This results in substantially lower computational cost. However, at the time of writing, using *Pythia 8* with access to the free parameters as MCEG for the code is not possible.

Overall, the studies presented in this chapter indicate that the tuning workflow can in principle be applied to air-shower simulations. At the same time, they also highlight the present practical limitations. Because the use of *Pythia 8* in air-shower physics is still relatively recent, many relevant tools do not yet support *Pythia 8* as a hadronic event generator or do not provide convenient access to its tunable parameters. In addition, the long runtime of simulations of high-energy cosmic-ray-induced air showers makes a full tuning study extremely demanding, since a large number of showers is required to cover the parameter space sufficiently. The integration of *Pythia 8* into air-shower simulation frameworks remains an active area of development. Nevertheless, for the reasons discussed above, a direct tuning of air-shower simulations with *CORSIKA 8* and *Pythia 8* is not pursued further in this thesis. Instead, a more indirect approach is explored in the next chapter.

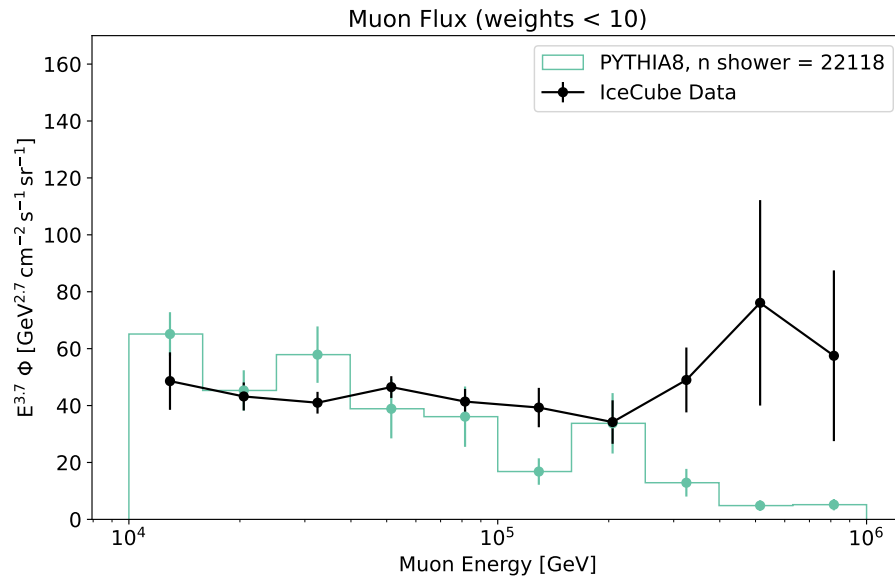


Figure 7.7: Simulated atmospheric muon flux using CORSIKA 8 and Pythia 8 as hadronic interaction model. Weights with values above 10 are removed.

8 Exploring the Interplay between Collider and Air-Shower Simulations

When tuning a general-purpose MCEG such as **Pythia 8**, the choice of observables is a non-trivial task. An improved description of one observable may be accompanied by a deterioration in the description of others. This is acceptable if the analysis is focused on a small set of observables only. However, if the aim is to improve not merely individual observables but the description of air-shower simulations as a whole, while at the same time preserving a reliable description of hadronic interactions at the LHC, the set of observables used for the tune must reflect the relevant physical processes.

Examples of **Pythia 8** tunes include the Monash tune [34], which improved the overall agreement with data, as well as a more recent forward tune [58], which was developed to improve the modeling of particle production in the forward region at the LHC.

Building on these developments, the long-term objective is to improve the modeling of air showers by incorporating data from collider and fixed-target experiments that are sensitive to components of the hadronic interaction model relevant for air-shower simulations. A direct tuning to air-shower observables is currently computationally challenging because of the long runtime of simulation tools such as **CORSIKA 8** and the present level of development of **Pythia 8** in the context of air-shower modeling, as discussed and concluded in Chapter 7. However, tuning on collider and fixed-target data is possible, and the impact of the resulting tune on the description of air-shower simulations can be evaluated.

This chapter presents a study of tuning to collider and fixed-target observables and evaluating the resulting impact on air-shower modeling. First, the role of collider data, in particular from experiments at the LHC, in improving the understanding and modeling of air showers is discussed. Subsequently, the selection of free parameters and observables used in the tune is introduced, followed by an evaluation of the final tune. Afterwards, the parameterization of the detector response is identified as a limiting factor in the tuning procedure and is examined in more detail. Finally, the impact of the tuning on air-shower simulations is evaluated.

8.1 Understanding Air Showers with LHC Collider Data

At the time of writing, the LHC is operating in Run 3 and has already delivered proton-proton collisions at a center-of-mass energy of 13.6 TeV. In addition to the proton-lead and lead-lead collision systems studied in earlier running periods, Run 3 also included, for the first time in 2025, dedicated proton-oxygen, oxygen-oxygen, and neon-neon collisions. Measurements of such systems in the controlled environment of the major LHC experiments provide valuable information on hadronic interactions that are relevant for the development of extensive air showers in the atmosphere. In particular, hadron-oxygen collisions are of special interest, as air showers are initiated through the interaction of cosmic rays with atmospheric nuclei and previously had to be constrained indirectly through interpolation between proton-proton and proton-lead data.

Although the energy of the first interaction in an extensive air shower can exceed the nucleon-nucleon center-of-mass energies available at the LHC by several orders of magnitude, collider data still probe an important part of the relevant hadronic physics. For a fixed-target interaction, as realized when a cosmic ray of energy E strikes an atmospheric nucleus of mass m at rest, the center-of-mass energy is approximately given by

$$\sqrt{s} \simeq \sqrt{2Em}, \quad E \simeq \frac{s}{2m}. \quad (8.1)$$

For example, a collider energy of 14 TeV corresponds to a fixed-target projectile energy of approximately 0.1 EeV for a proton target with $m = 0.938$ GeV. While this still lies below the energy range of ultra-high-energy cosmic rays, such measurements nevertheless provide important constraints on particle production in the cascade development of air showers, as discussed in Ref. [1].

As already discussed in Chapter 3.5, one of the main differences between collider events and air showers lies in the direction of particle production. In collider analyses, the angular distribution of produced particles is commonly expressed in terms of the pseudorapidity η , defined as

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right], \quad (8.2)$$

where θ is the polar angle with respect to the beam axis. The relation between polar angle and pseudorapidity is illustrated in Figure 8.1. The use of pseudorapidity is particularly convenient because, for ultra-relativistic particles, it closely approximates the rapidity. Under Lorentz boosts along the beam axis, rapidity changes only by an additive constant, such that the shape of the rapidity distribution is preserved and only shifted along the rapidity axis.

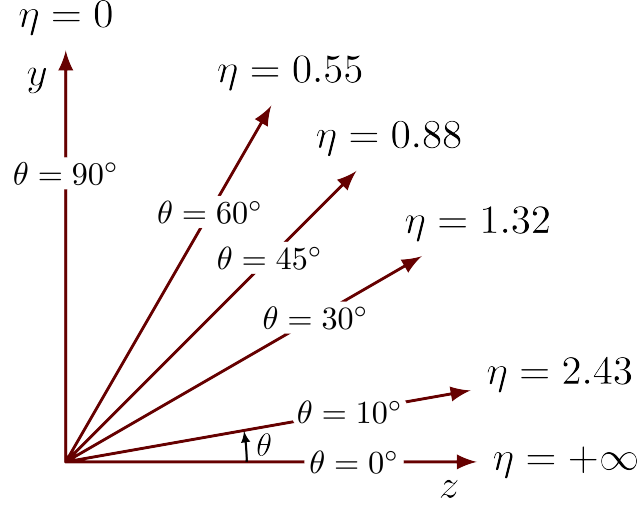


Figure 8.1: Pseudorapidity values corresponding to different polar angles θ . Figure taken from Ref. [59].

In this context, the term *forward region* typically refers to pseudorapidities $\eta > 2$, while the *central region* denotes smaller values. For air showers, the forward region is of particular importance, since a large fraction of the particles in the cascade are produced at high pseudorapidity. Historically, a considerable part of the development of `Pythia 8` has focused on the central region, reflecting the priorities of the main LHC experiments. More recent efforts, such as the forward tune presented in Ref. [58], have extended this program towards the forward phase space. Nevertheless, collider data at very large pseudorapidities remain comparatively limited, even though this region is especially relevant for air-shower observables such as X_{max} and N_μ , both of which are sensitive to charged-particle production and multiplicity in the forward direction. Measurements from dedicated forward experiments, therefore, play an important role in constraining this region.

A complementary approach to improving the understanding of air-shower development is provided by fixed-target experiments. In contrast to collider measurements at the LHC, fixed-target data from experiments such as NA61/SHINE or from the fixed-target program of LHCb probe interaction constraints that are closer to those encountered in the particle cascade of an air shower and are naturally more focused on forward particle production. Their center-of-mass energies are typically well below those of collider data, but they remain highly relevant for the later stages of air-shower development, where the interaction energies have decreased sufficiently. Such measurements therefore provide important input complementary to collider data.

8.2 Selection of Pythia 8 Parameters

In order to identify a suitable set of **Pythia 8** parameters relevant for the physics governing extensive air showers, it is first necessary to consider the interactions that shape the particle cascade. When a high-energy cosmic ray enters the atmosphere and interacts with an oxygen or nitrogen nucleus at an energy well beyond that reached by collider experiments such as the LHC, multiple parton-parton scatterings can occur within the same collision. These multiparton interactions (MPI) play a central role in particle production in collisions between cosmic rays and atmospheric nuclei, substantially enhancing secondary-particle multiplicities and redistributing energy towards the softer components of the shower. As a consequence, MPI directly affect key air-shower observables, including the depth of shower maximum X_{max} , the lateral particle distribution, and the muon content at ground level. An accurate description of MPI is therefore essential for the simulation and interpretation of extensive air showers, in particular for shower reconstruction and mass-composition inference.

Although the highest-energy regime relevant for air showers cannot be probed directly at the LHC, the parameters governing MPI can be constrained using collider data and subsequently extrapolated into more extreme kinematic regimes. A selection of **Pythia 8** parameters affecting MPI was already introduced in Chapter 4.4.3. While collider measurements sensitive to MPI are available in hadronic collisions at the LHC, they are absent in lepton-collider environments such as LEP. Nevertheless, LEP data remain highly relevant, since they provide precise information on hadronization and final-state particle production in a clean environment, without the complications arising from nuclear effects or parton distribution functions in hadrons.

In addition to MPI, other physical processes directly affect hadron multiplicities, the lateral distribution of the shower, and X_{max} . In particular, the energy and momentum distributions of secondary particles are important. Parameters governing these aspects of the hadronic final state are therefore also relevant for the tuning. These include the parameters of the Lund string fragmentation function, as well as parameters controlling the transverse-momentum distribution assigned to hadrons during string breaking, as described in Chapter 4.4.1 and 4.4.2.

The selection of **Pythia 8** parameters for the tune is determined not only by their relevance for air-shower physics, but also by the degree to which they influence the observables included in the tuning dataset. The stronger the dependence of an observable on a given parameter, the larger the potential impact of the tune on that observable. In practice, however, this sensitivity is often difficult to assess a priori, since the part of the hadronic model controlled by a given parameter may play only a subleading role in the observable.

As discussed in Chapter 6, the number of parameters in a tune determines the dimensionality of the posterior distribution. Provided that the chosen parameters are sufficiently sensitive to the observables, increasing their number generally increases the flexibility of the fit, improving the tuning performance. At the same time, however, it substantially increases computational cost due to the larger number of reference points required to be simulated, and results in greater complexity in constructing an accurate surrogate model for the detector response. It is therefore necessary to identify a set of parameters that exhibits strong sensitivity to the observables of interest while keeping the total number of parameters sufficiently small to ensure a manageable number of reference points and an accurate surrogate-model description.

The studies presented in this work focus on exploring the interplay between collider data and their impact on air-shower simulations, as well as on assessing the influence and performance of different surrogate models. Since both objectives require repeated simulations of a large number of reference points, the number of tuned parameters is kept limited. For these reasons, the number of free parameters for the following tune is restricted to three.

Evaluating the Sensitivity

Restricting the tune to three parameters requires a careful assessment of the sensitivity of *Pythia 8* parameters to collider observables. Defining such a sensitivity measure is, however, not straightforward. Moreover, sensitivity is meaningful only in relation to a specific observable. Since each bin of an observable is parameterized independently, any sensitivity measure for a full observable must, in practice, combine the bin-wise sensitivities.

An intuitive starting point is to parameterize the observable with respect to a model parameter and define the partial derivative of the parameterization as sensitivity, since the slope indicates how strongly the bin content changes when the parameter is varied. However, this definition yields a function over the parameter space rather than a single numerical value, such that one must specify the point at which the derivative is to be evaluated. Since the present study builds on established tunes such as the Monash tune, which is currently essentially the default in *Pythia 8*, a plausible choice would be the corresponding default parameter values. The observable sensitivity could then be defined as an average over all bins. This approach was used in Ref. [12], where hadronic models were tuned to LEP data.

For LEP observables, for which *Pythia 8* already provides a well-optimized description, such a definition is reasonable, since the tuned parameter values are not

expected to deviate strongly from the default values. However, for underlying-event observables in the LHC or for fixed-target measurements, `Pythia 8` often fails to model the data sufficiently. In such cases, the optimal parameter value may lie close to the boundary of the allowed range, or the model may be unable to describe the data at all. If `Pythia 8` is unable to model the observable independently of the changes in parameters, a sensitivity measure is obviously irrelevant, since no variation of the free parameters can compensate for this deficiency. But when the optimal parameter value still lies within the allowed range, a purely local measure at the default point can be misleading if the surrogate model is nearly flat there but changes substantially elsewhere in the allowed parameter range.

A possible alternative is therefore to replace the local derivative at the default point by the slope evaluated over a small interval around it,

$$S = \frac{f(x_{\text{def}} + \frac{\Delta x}{2}) - f(x_{\text{def}} - \frac{\Delta x}{2})}{\Delta x}, \quad (8.3)$$

where S denotes the sensitivity, f the surrogate model for a given observable bin, and x the parameter vector. In this case, the interval must be chosen small enough to avoid including local extrema, yet large enough to reflect the behavior over a representative part of the allowed parameter range. One possible choice would be to set the interval dynamically, for instance to 1% of the corresponding parameter range.

To avoid potential bias from mismodeling in the surrogate model, one could simulate dedicated Monte Carlo samples at the shifted parameter values $x \pm \Delta x$ rather than evaluating the surrogate model. For the purpose of identifying suitable parameters and observables, this is in fact necessary, since the parameter space has not yet been sampled to construct the corresponding surrogate model. As an example, the sensitivity of an observable from Ref. [60] to the parameter α_S is shown in Figure 8.2.

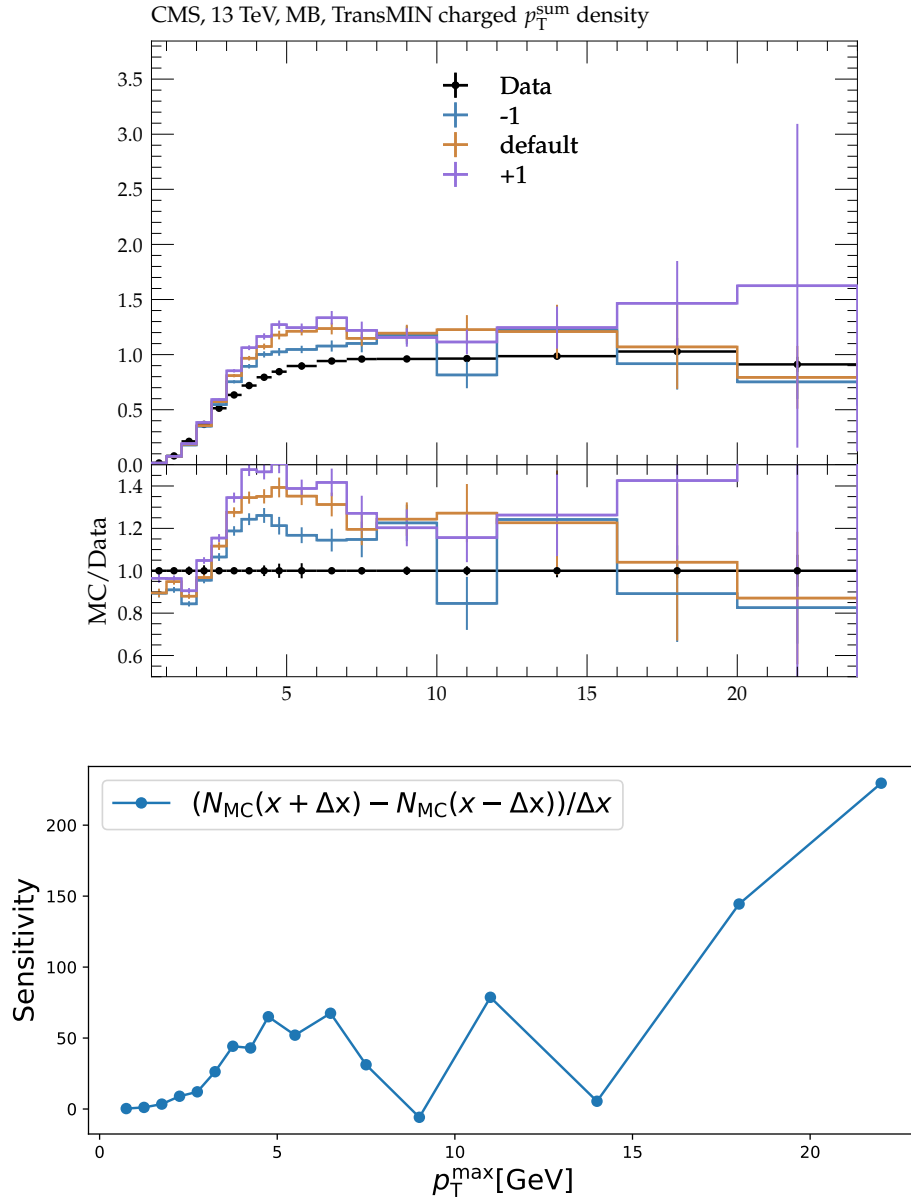


Figure 8.2: Upper panel: p_T^{sum} density in event regions containing only MPI and beam-remnant activity as a function of the transverse momentum of the leading charged particle, measured at the CMS experiment [60], shown for the default value of α_S and for variations of $\pm 1\%$. Lower panel: corresponding sensitivity of the observable to α_S in Pythia 8, obtained from Monte Carlo samples using $\Delta x = 2\%$ of the parameter range.

The main limitation of this definition is its dependence on both the parameter range and the absolute bin value. As a result, the sensitivity cannot be compared directly either between different parameters for the same observable or between different observables. Such comparability is, however, essential when the goal is to identify a restricted set of parameters and observables for the tune.

For this reason, the final selection of parameters is based not on a formal sensitivity metric, but on a broader exploratory study in which a set of observables is simulated for a larger set of candidate parameters. In order to reduce the computational cost, each observable is simulated for only one parameter at a time, using 10^6 events for the default parameter value as well as for the minimum and maximum allowed parameter value. This results in three simulations per observable-parameter pair. The change in the observable between the minimum, default, and maximum parameter value is then compared to the other candidate parameter, to find the parameters with the highest sensitivity.

This approach is not perfect. In particular, the allowed parameter ranges differ substantially from one parameter to another, and the approach assumes that steadily increasing a parameter increases or decreases the observable in a corresponding manner. This does not need to be the case, as the largest deviation of a bin content may occur at an intermediate parameter value rather than at one of the range boundaries.

The set of candidate observables considered in this study includes measurements from LEP, fixed-target experiments, and underlying-event measurements at the major LHC experiments. As discussed above, observables related to particle multiplicities, momenta, and transverse momenta are focused, since these are expected to probe physical properties that are relevant for air-shower modeling. Because only measurements implemented in the **Rivet** framework can be used, the available dataset is significantly constrained. For example, the fixed-target measurements from LHCb are absent in **Rivet** at the time of writing and therefore cannot be included in the tuning workflow. As a consequence, only one fixed-target analysis, from the NA22 experiment [61], is used in this study.

The fixed-target data are expected to be particularly challenging for **Pythia 8**, since the event generator has primarily been developed and optimized for collider data. However, fixed-target measurements are especially useful for constraining air-shower physics, since the fixed-target configuration naturally probes the forward production region that is most relevant for interactions in air showers. A study with a stronger focus on the use of fixed-target measurements to constrain and improve air-shower simulations is presented in Ref. [62]. For the LHC, observables from the analysis in Ref. [63] from the ATLAS experiment are considered. These involve events containing decaying Z bosons and are sensitive to properties of the underlying

event. Underlying-event observables are often difficult to model in hadronic event generators, since they are dominated by soft-QCD effects, where even comparatively small modeling deficiencies can produce visible deviations in the observables. In addition, a CMS measurement from Ref. [60] is included, consisting of the charged-hadron multiplicity as a function of pseudorapidity. In contrast, `Pythia 8` is already highly optimized for LEP data, making LEP observables essential for constraining the parameters associated with hadronization. Two LEP analyses are therefore included, from the L3 [64] and ALEPH [65] experiments.

The initial set of candidate parameters is chosen to include quantities expected to affect either the underlying event directly or the general production of hadrons. The larger candidate set comprises the parameters introduced in Chapter 4, namely `StringZ:aLund`, `StringZ:bLund`, `StringPT:sigma`, `MultipartonInteractions:pT0Ref`, and `MultipartonInteractions:ecmPow`, corresponding to a_{Lund} , b_{Lund} , σ , $p_{\text{T}0}^{\text{Ref}}$, and ecmPow , respectively. In addition, the parameters `StringZ:aExtraSQuark` and `StringZ:aExtraDiQuark`, which control the suppression of strange-quark and diquark production, are included. Further candidate parameters are `BeamRemnants:dampPopcorn` and `BeamRemnants:companionPower`, which affect beam remnants, as well as `StringFlav:popcornRate`, which influences baryon production.

The parameter α_S is not included in the larger candidate set, since it represents the strong coupling at the scale of the Z boson and is treated here as a fixed quantity rather than a freely tunable parameter. This choice is not universal among `Pythia 8` tuning studies, since the event generator only calculates leading-order matrix elements and one may therefore argue that the value of α_S in `Pythia 8` also absorbs effects beyond leading order.

Each analysis is simulated for every selected parameter at the default value and at the lower and upper ends of the chosen range, with reference to the allowed parameter intervals quoted in the current `Pythia 8` manual [66]. In some cases, parameter values close to the formal limits of the allowed range lead to unstable MC generation or to substantially increased runtime. In such situations, values close to, but not exactly at, the formal boundaries are used. In other cases, the default value already lies at the edge of the allowed range. The chosen values, the allowed ranges, and the default values are summarized in Table 8.1.

Combining five analyses, each containing multiple observables, with ten parameters results in a number of distributions that is too large to discuss in detail. In the following, a selection of observable-parameter combinations is presented, with emphasis on multiplicity and momentum distributions. The parameter names are abbreviated as `aLund`, `bLund`, `aExtraSQuark`, `aExtraDiQuark`, `sigma`, `ecmPow`, `pT0Ref`, `dampPopcorn`, `companionPower`, and `popcornRate`. The sensitivity plots for the not shown parameters are collected in Appendix A.1.

Table 8.1: Candidate parameters together with the parameter ranges used in the sensitivity study, the allowed ranges quoted in the *Pythia 8* manual, and the default values.

Parameters	Chosen Range	Allowed Range	Default Value
<code>StringZ:aLund</code>	[0.2, 1.6]	[0.0, 2.0]	0.68
<code>StringZ:bLund</code>	[0.2, 1.6]	[0.2, 2.0]	0.98
<code>StringZ:aExtraSquark</code>	[-1.0, 2.0]	[-1.0, 2.0]	0.0
<code>StringZ:aExtraDiQuark</code>	[0.0, 2.0]	[0.0, 2.0]	0.97
<code>StringPT:sigma</code>	[0.1, 0.8]	[0.0, 1.0]	0.335
<code>MultipartonInteractions:ecmPow</code>	[0.0, 0.5]	[0.0, 0.5]	0.225
<code>MultipartonInteractions:pT0Ref</code>	[1.8, 5.0]	[0.5, 10.0]	2.28
<code>BeamRemnants:dampPopcorn</code>	[0.0, 1.0]	[0.0, 1.0]	1.0
<code>BeamRemnants:companionPower</code>	[0.0, 4.0]	[0.0, 4.0]	4.0
<code>StringFlav:popcornRate</code>	[0.0, 2.0]	[0.0, 2.0]	0.5

Figure 8.3 shows an observable from the ATLAS analysis describing the charged-particle multiplicity as a function of transverse momentum for the parameters **aLund**, **sigma**, **pT0Ref**, and **ecmPow**. While the observable is reproduced reasonably well in much of the distribution, the first bins, which contain the largest event fraction, are not described sufficiently by the default parameter settings and are overestimated up to approximately 1 GeV.

The influence of the parameters on the observable differs a lot. The strongest sensitivity is observed for **pT0Ref**, for which the data lie within the area spanned by the lowest and highest parameter values in most bins. A similar sensitivity is seen for **sigma**, which also contains the data in most bins and exhibits a reversing trend in the first few bins with respect to over- and underestimation. The parameter **aLund** likewise shows significant sensitivity, although it does not fully enclose the data in the first bins. In the intermediate and higher bins, however, its influence remains visible over a broad range. In contrast, **ecmPow** shows almost no sensitivity in this observable, with only minor fluctuations between the two extreme parameter choices. Among the parameters not shown, only **bLund** exhibits an influence comparable to that of **aLund**, while **aExtraSquark** shows a modest sensitivity in the later bins and the remaining parameters have little to no visible effect.

Figure 8.4 shows a second observable from the same ATLAS analysis, the charged-particle multiplicity distribution for events, for the parameters **bLund**, **sigma**, **pT0Ref**, and **aExtraSquark**. Here, the default description is clearly insufficient, illustrating the difficulty of modeling the underlying event. Both the first few bins and the tail of the distribution show sizable deviations from the data.

Compared to the previous observable, the overall parameter sensitivity is reduced.

The strongest effect is again found for **pT0Ref**, large values lead to an overestimation of the first bins, whereas low values underestimate them, with the trend reversing at about $N_{\text{ch}} = 15$. The parameters **bLund** and **sigma** also show noticeable sensitivity, although neither is able to envelope the data in the tail of the distribution. The parameter **aExtraSQuark** exhibits a similar level of sensitivity in the later bins, but has almost no influence in the first bins, where most events are concentrated. The parameter **aLund** performs comparably to **bLund** and **sigma**, whereas the remaining parameters have negligible impact.

For the selected CMS analysis, implemented in **Rivet** as **CMS_2015_I1384119**, only one observable is available. It describes the pseudorapidity density of charged hadrons in proton-proton collisions at $\sqrt{s} = 13$ TeV, measured in an unusual CMS configuration without magnetic field. The sensitivity of this observable to the parameters **aLund**, **sigma**, **pT0Ref**, and **aExtraDiQuark** is shown in Figure 8.5.

The parameters **aLund**, **sigma**, and **pT0Ref** all show strong sensitivity, as the data are enclosed by the parameter variations in each bin. In particular, the lower value of **sigma** at 0.1 yields an almost optimal description of the data. By contrast, **aExtraDiQuark** exhibits no visible sensitivity. Among the parameters not shown, **bLund** and **ecmPow** perform similarly to **aLund**, whereas the remaining parameters show only minor to no influence.

Looking at the LHC observables, **pT0Ref** appears to be the most influential parameter overall, while **aLund**, **bLund**, and **sigma** also show consistently strong sensitivity. The high impact of **pT0Ref** is expected, since MPI are the dominant source of the underlying event and therefore strongly affect both ATLAS observables. Because a lower value of **pT0Ref** corresponds to a larger MPI cross section, which rises steeply at low p_T , the parameter has a large effect on charged-particle multiplicities. This behavior is particularly visible in the CMS observable in Figure 8.5, where changes in the parameter increases the overall bin contents without changing the shape of the distribution.

It is also noteworthy that **aLund** and **bLund** show a very similar level of sensitivity in the LHC observables. Both parameters enter the Lund string fragmentation function and steer the energy fraction assigned at string breaking. Together with **sigma**, which determines the width of the transverse-momentum distribution assigned to hadrons during string breaking, these parameters are therefore expected to exhibit a significant influence on final-state hadron observables.

By contrast, the beam-remnant parameters show only modest impact. This is expected, since they are assumed to contribute mainly in the forward region and at low transverse momenta, and to have minor effects compared to MPI and string-fragmentation parameters.

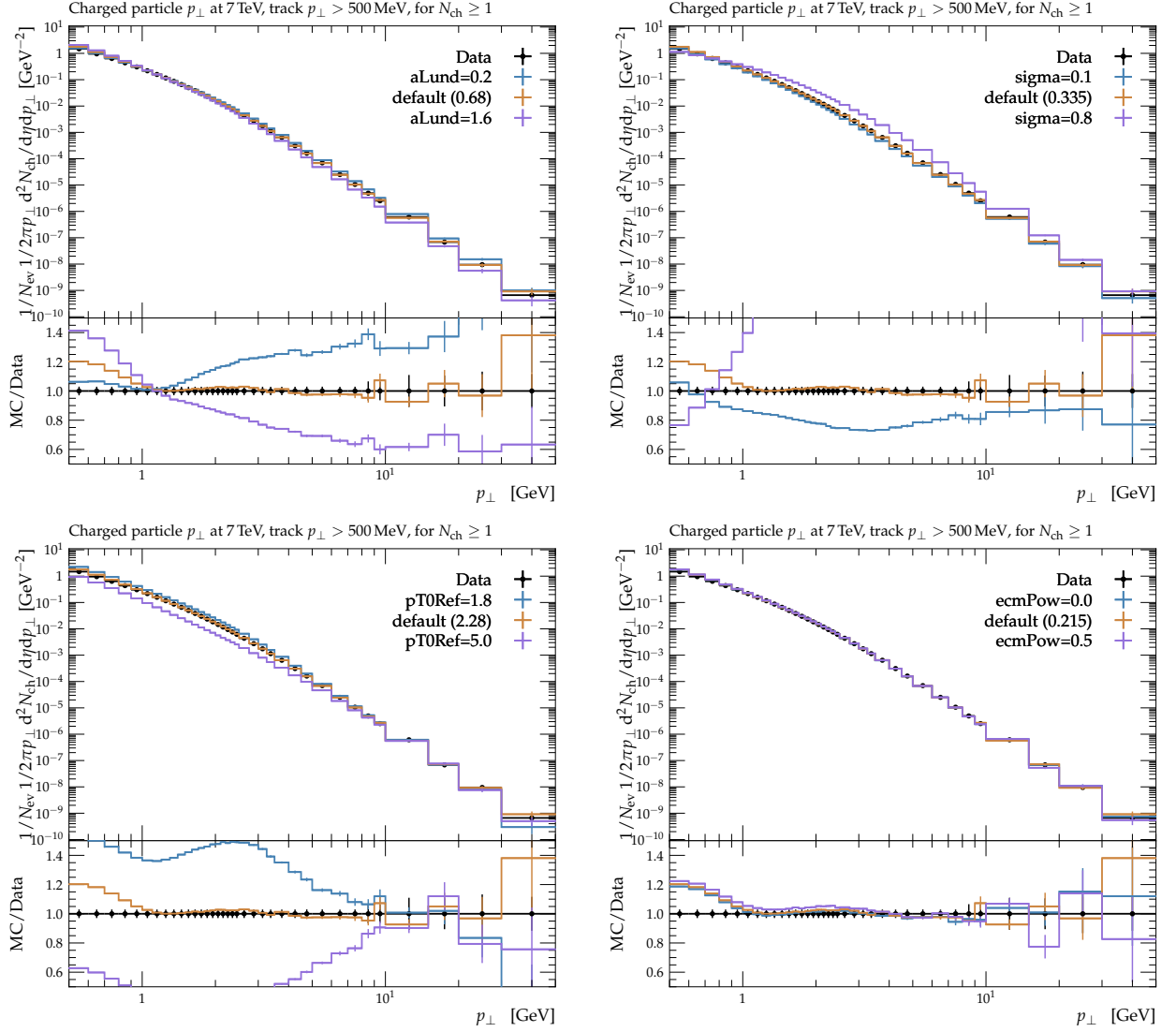


Figure 8.3: Charged-particle multiplicity as a function of transverse momentum for different parameter variations. The observable is taken from the Rivet analysis ATLAS_2010_I882098, histogram d10-x01-y01.

One notable detail is the small but visible sensitivity of $ecmPow$ in the two ATLAS observables. Since the ATLAS analysis was performed at $\sqrt{s} = 7$ TeV, the regularization scale reduces to $p_{T0} = p_{T0,Ref}$, as discussed in Equation 4.7. The parameter $ecmPow$ should therefore not affect these observables through the scaling

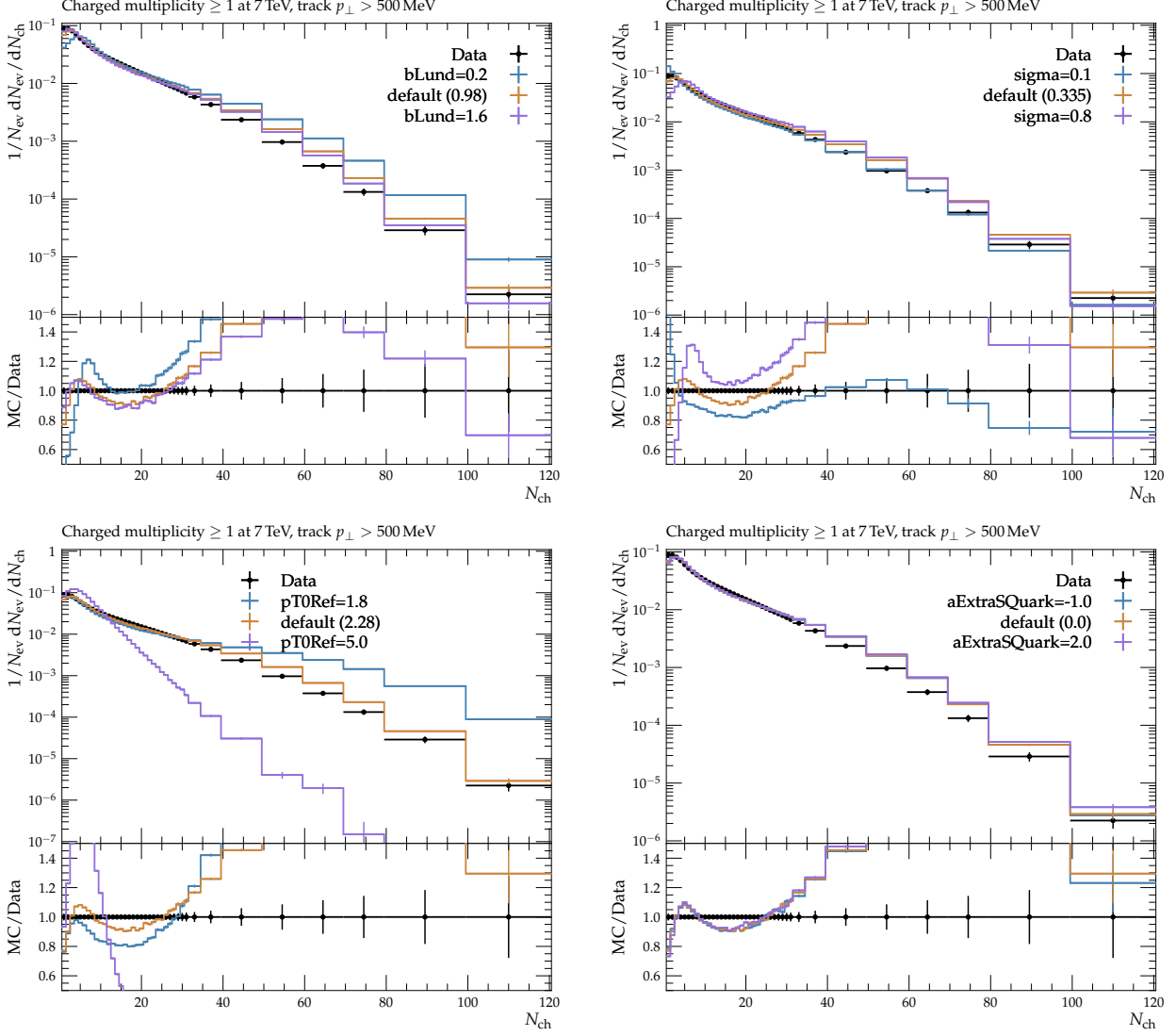


Figure 8.4: Charged-particle multiplicity distribution for different parameter variations. The observable is taken from the Rivet analysis ATLAS_2010_I882098, histogram d17-x01-y01.

factor $\left(\frac{E_{CM}}{E_{CM}^{Ref}}\right)^{ecmPow}$, which is equal to unity in this case. Because the samples for different `ecmPow` values were generated with identical seeds and settings, the resulting observables are expected to appear exactly the same. The small residual differences visible in the plots may therefore indicate that `ecmPow` enters the simula-

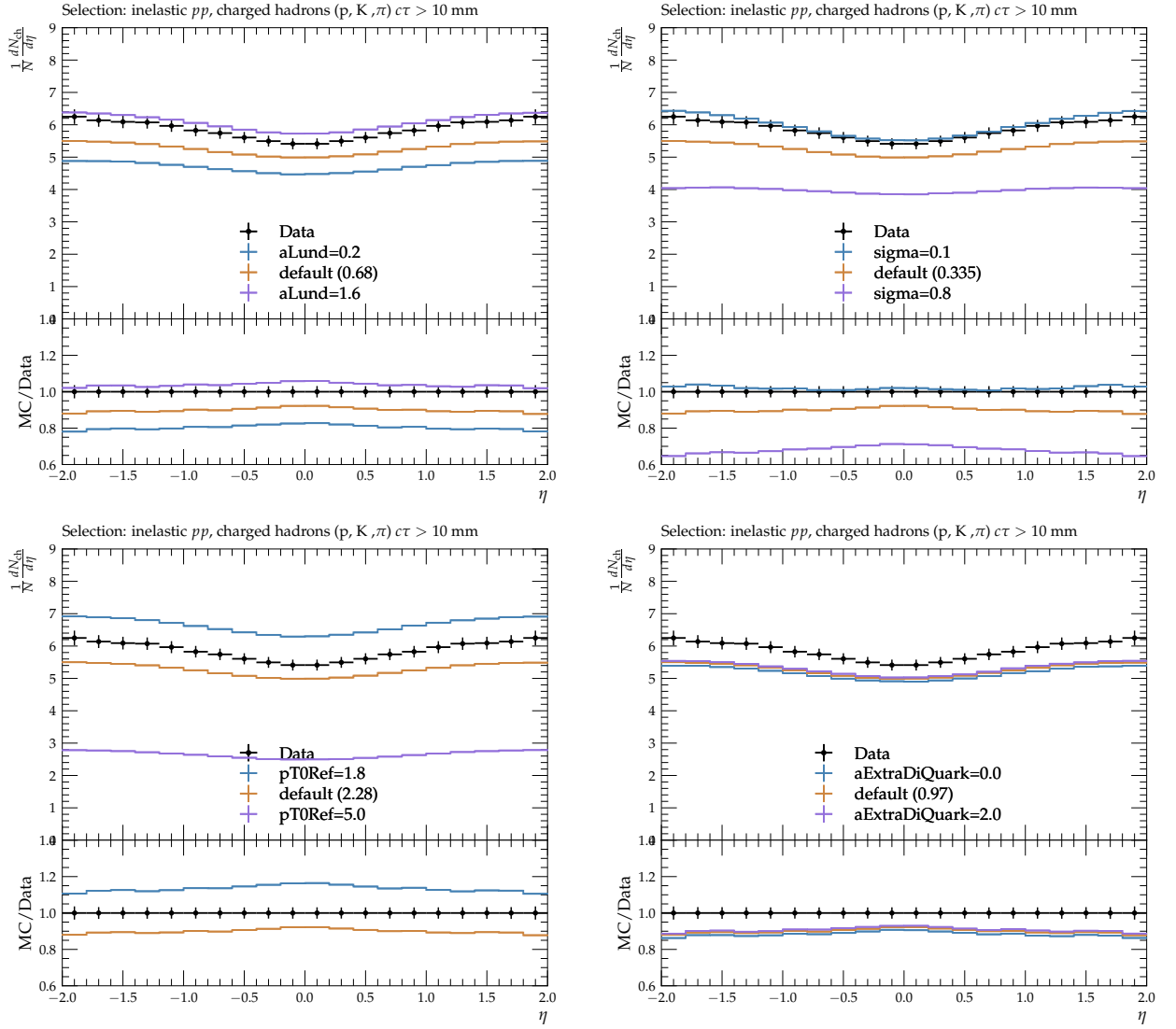


Figure 8.5: Pseudorapidity density of charged hadrons for different parameter variations. The observable is taken from the Rivet analysis CMS_2015_I1384119, histogram d01-x01-y01.

tion elsewhere, or that some internal error or numerical effect is present. Given the minor size of the effect, this issue is not investigated further.

Turning to the LEP observables, the emphasis again lies on multiplicity and momentum distributions. In this case, observables measured at $\sqrt{s} = 91.2 \text{ GeV}$ are

investigated. Figure 8.6 shows the distribution of the logarithm of the scaled momentum, $\xi_p = \log(1/x_p)$ with $x_p = 2|\vec{p}|/E_{\text{cm}}$, for charged particles, for the parameters **aLund**, **bLund**, **aExtraSQuark**, and **pT0Ref**. The distribution is taken from the Rivet analysis ALEPH_1996_I428072.

As expected, both string-fragmentation parameters **aLund** and **bLund** strongly affect most bins and do so with similar strength. In contrast, **pT0Ref** shows no sensitivity at all, which is consistent with the absence of MPI in electron-positron collisions. The parameter **aExtraSQuark** has a noticeable effect in the first few bins, but its influence decreases in the region containing larger event fractions. Among all candidate parameters, the strongest sensitivity is observed for **sigma**, whereas **aExtraDiQuark** and **popcornRate** show only weak effects. The remaining parameters govern MPI or beam remnants and therefore do not contribute in electron-positron collisions.

A similar pattern is observed in the corresponding distribution measured by the L3 experiment, shown in Figure 8.7 for the parameters **bLund**, **sigma**, **aExtraDiQuark**, and **popcornRate**. Once again, **sigma** emerges as the most influential parameter, while **bLund** also shows a strong sensitivity. In contrast, **aExtraDiQuark** and **popcornRate** have only a negligible effect. Overall, the sensitivity pattern is consistent between the ALEPH and L3 measurements. To complement the LEP momentum observable, a charged-particle multiplicity distribution from the L3 experiment is also included. This observable is shown in Figure 8.8 for the parameters **aLund**, **bLund**, **sigma**, and **aExtraSQuark**. As for the other LEP observables, **aLund**, **bLund**, and **sigma** exhibit the largest sensitivity. Among the remaining parameters, only **aExtraSQuark** has a visible effect, while the others show no significant influence. The final part of the sensitivity study concerns the fixed-target analysis implemented in Rivet as EHS_1988_I265504. The EHS/NA22 data include collisions of proton, pion, and kaon beams on a fixed hydrogen target with a beam momentum of 250 GeV/c. One of the observables from this analysis is the transverse-momentum distribution of negatively charged pions produced in collisions involving a pion beam. Figure 8.9 shows the sensitivity of this observable to the parameters **sigma**, **aLund**, **pT0Ref**, and **ecmPow**. As expected for forward data at a much lower center-of-mass energy than the LHC measurements, the observable is described poorly by the default Pythia 8 settings. The first bins, which contain the largest event fraction, are reproduced comparatively well, whereas the later bins deviate substantially from the measured data. The strongest sensitivity is found for **sigma**, **pT0Ref**, and **ecmPow**, which are able to bracket most bins within their chosen ranges. Among these, only **sigma** is capable of enveloping the data in the first bins up to about 1 GeV², while all three parameters only barely reach the data in the later bins, and only near the edges of the chosen parameter ranges. The parameter **aLund** has some influence on the observable, but it does not contain most of the data.

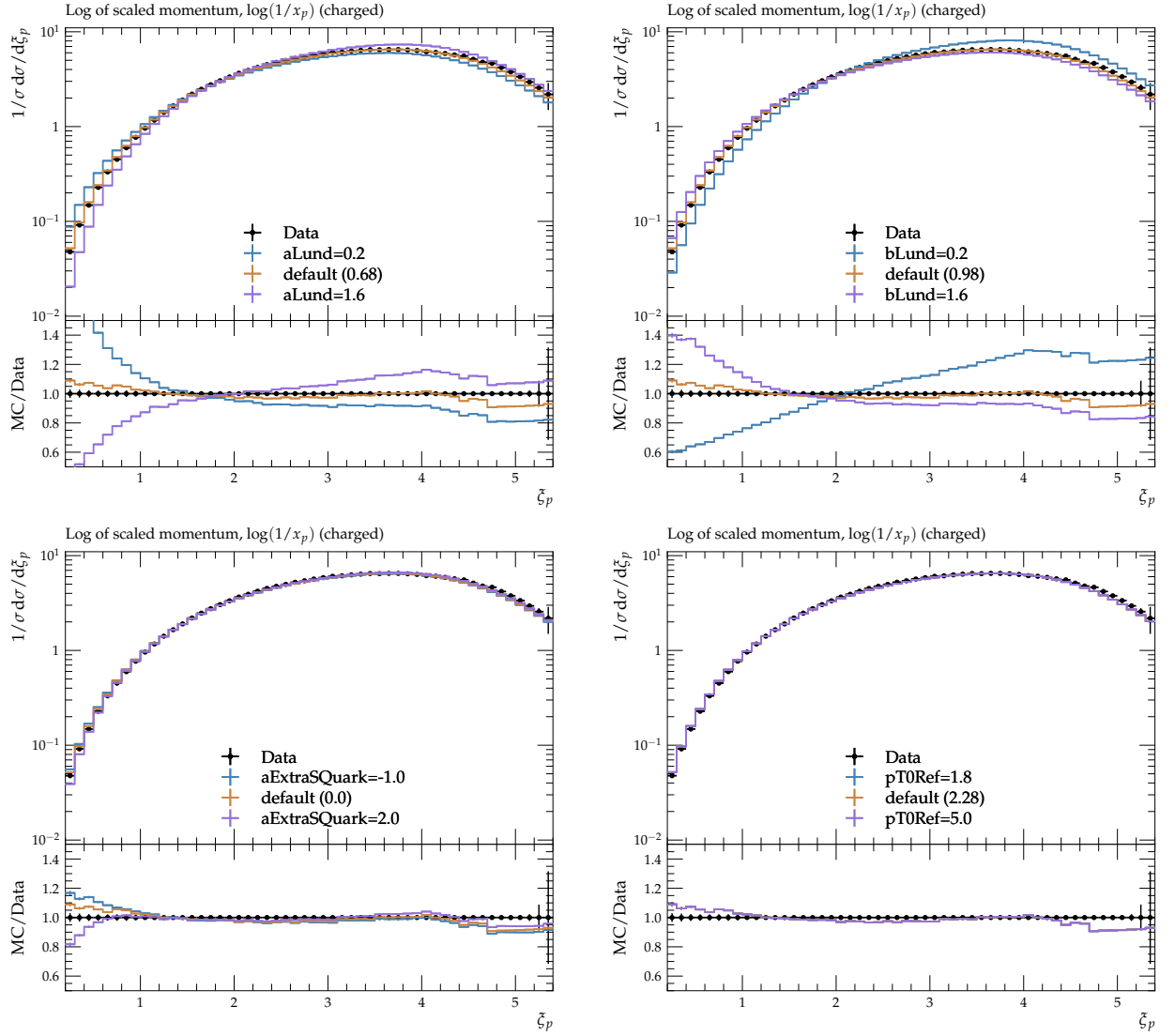


Figure 8.6: Distribution of the logarithm of the scaled momentum, ξ_p , for different parameter variations. The observable is taken from the Rivet analysis ALEPH_1996_I428072, histogram d17-x01-y01.

Among the parameters not shown for this fixed-target observable, only **bLund** exhibits a sensitivity comparable to **aLund**, whereas the remaining parameters change the bin contents only weakly.

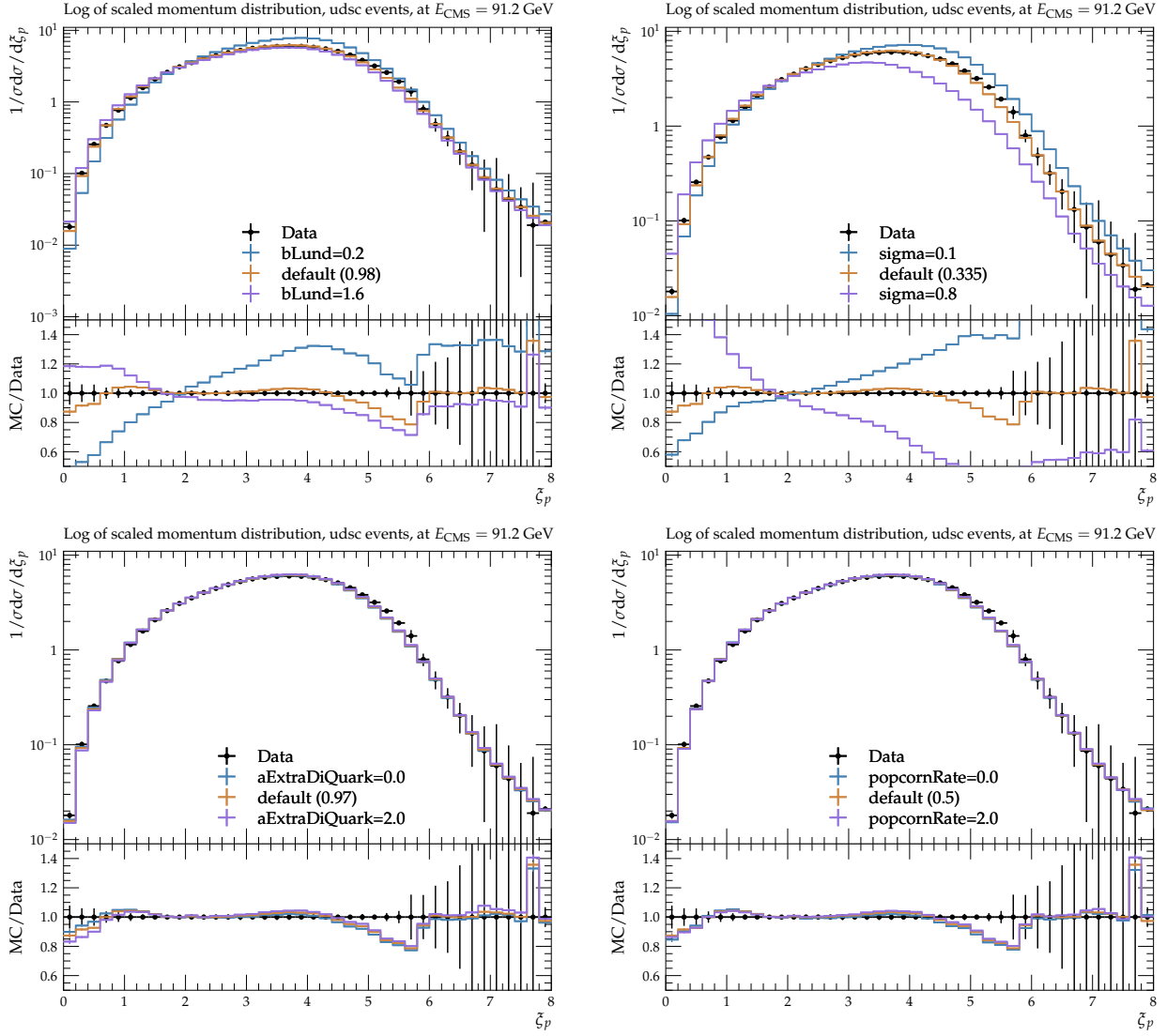


Figure 8.7: Distribution of the logarithm of the scaled momentum, ξ_p , for different parameter variations. The observable is taken from the Rivet analysis L3_2004_I652683, histogram d65-x01-y02.

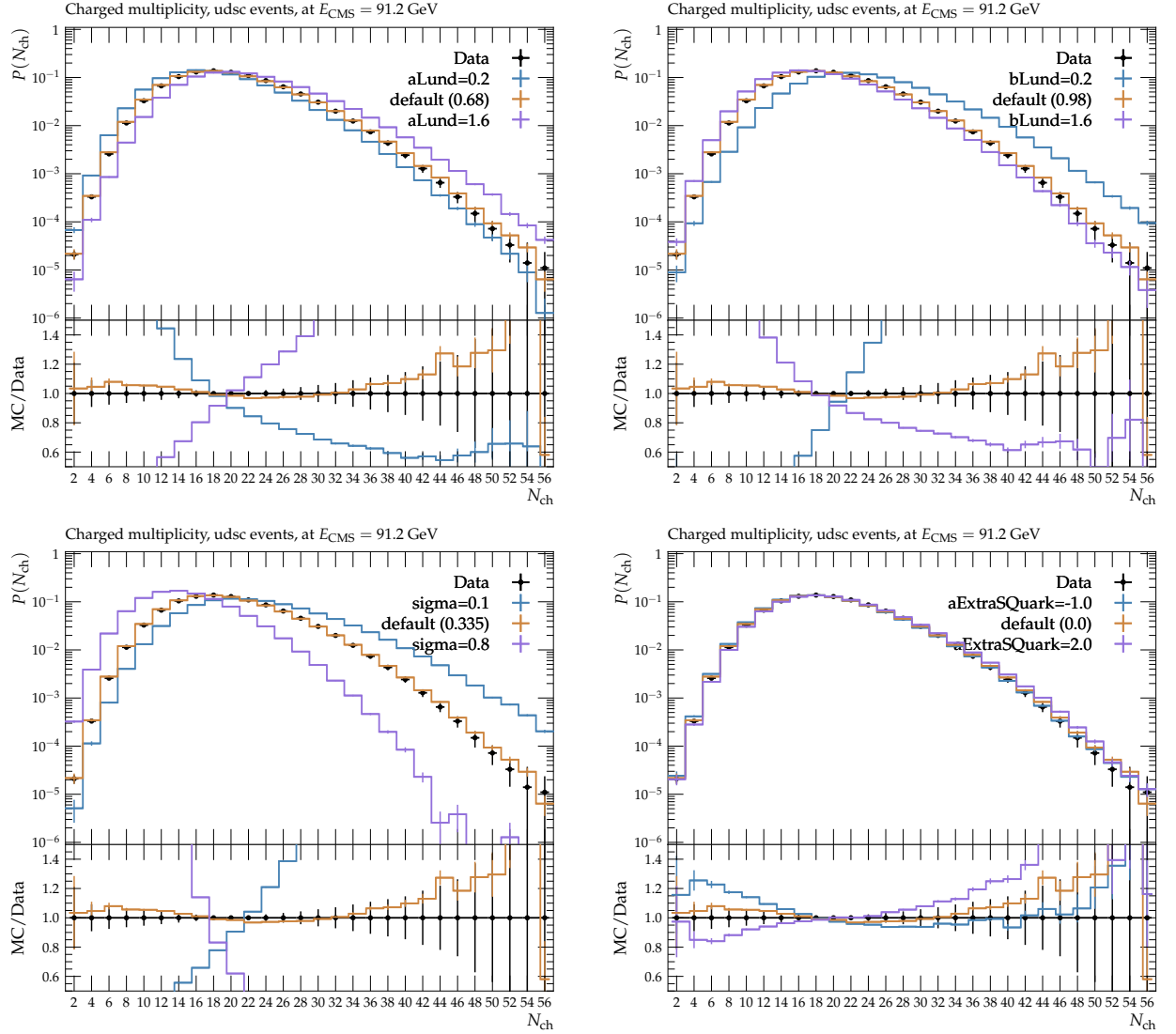


Figure 8.8: Charged-particle multiplicity distribution for different parameter variations. The observable is taken from the Rivet analysis L3_2004_I652683, histogram d59-x01-y02.

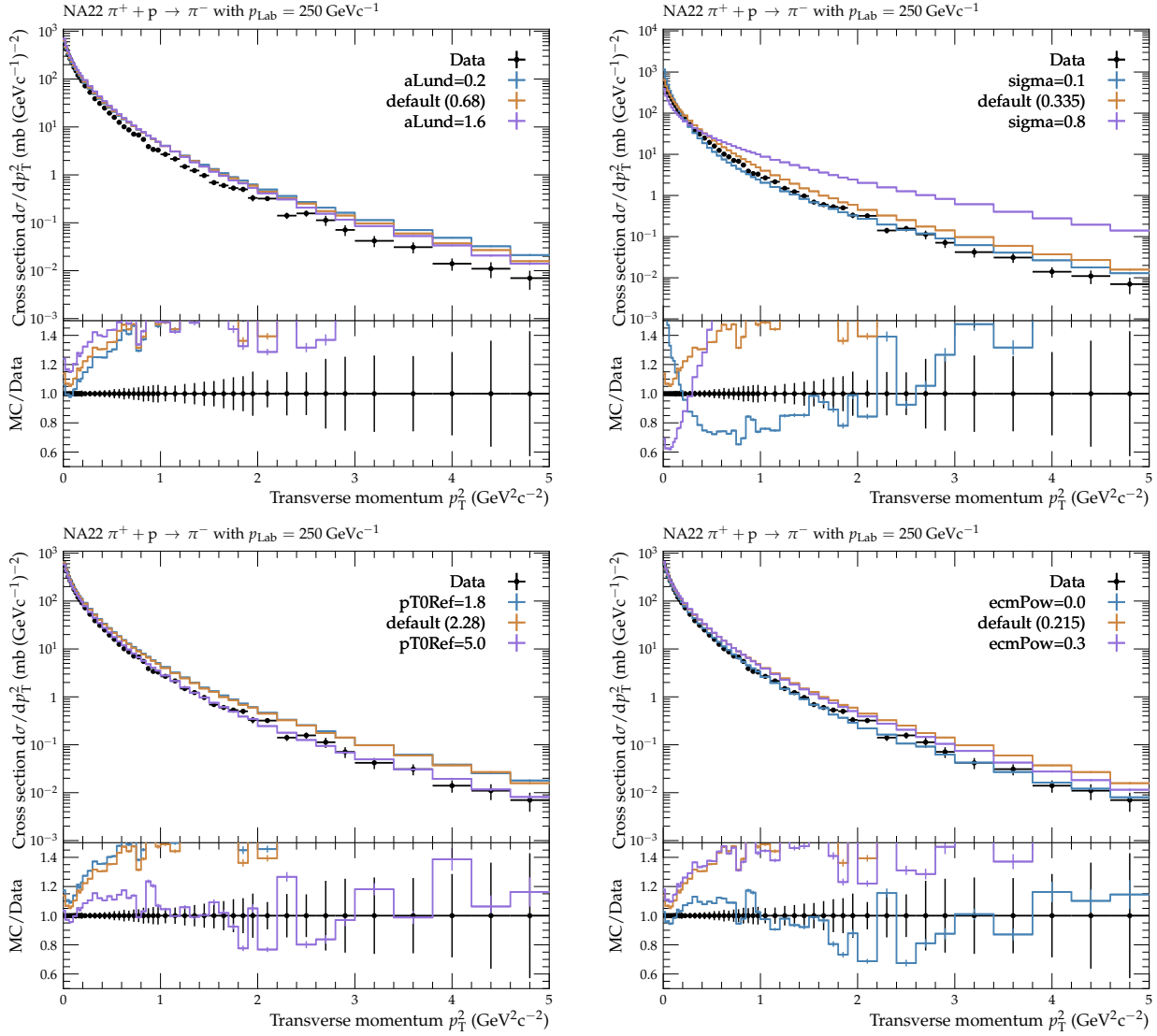


Figure 8.9: Transverse-momentum distribution of negatively charged pions produced in collisions of a pion beam with a fixed hydrogen target, shown for different parameter variations. The observable is taken from the Rivet analysis EHS_1988_I265504, histogram d06-x01-y01.

Choosing three parameters based on the discussed sensitivities, the focus is on parameters essential for simulations with `Pythia 8`, while showing high sensitivity. A binary summary of the sensitivity of each parameter to the three classes of experiments is given in Table 8.2.

In a more comprehensive tune involving a larger parameter set, it would be necessary to tune `pT0Ref` and `ecmPow` simultaneously, since both parameters contribute to the regularization of MPI, with `ecmPow` controlling the energy scaling of `pT0Ref`. In this study, however, the number of free parameters is deliberately kept small. Only `pT0Ref` is therefore retained to control the MPI regularization, assuming that the default value of `ecmPow` provides an adequate energy scaling.

For the Lund fragmentation function, the parameter `aLund` is chosen, as it affects the fraction of energy given to each constituent in a string break, affecting each observable with hadrons in the final state. The second fragmentation parameter, `bLund`, shows a comparable sensitivity and would likewise be a natural candidate in a higher-dimensional tune.

As the third parameter, `sigma` is selected because it shows, on average, the strongest influence across the observables discussed above. Since it determines the width of the transverse-momentum distribution assigned to produced hadrons, it again impacts processes that produce hadrons in the final state.

Table 8.2: Sensitivity matrix summarizing whether the considered observables show a significant sensitivity to each parameter. Entries marked with “X” indicate sensitivity, while “-” denotes no significant sensitivity. The classification is based on a qualitative assessment.

Parameters	LEP	fixed-target	LHC
<code>StringZ:aLund</code>	X	X	X
<code>StringZ:bLund</code>	X	X	X
<code>StringZ:aExtraSQuark</code>	X	-	X
<code>StringZ:aExtraDiQuark</code>	X	-	-
<code>StringPT:sigma</code>	X	X	X
<code>MultipartonInteractions:ecmPow</code>	-	X	X
<code>MultipartonInteractions:pT0Ref</code>	-	X	X
<code>BeamRemnants:dampPopcorn</code>	-	-	-
<code>BeamRemnants:companionPower</code>	-	-	-
<code>StringFlav:popcornRate</code>	-	-	-

8.3 Selection of Collider Observables

In a comprehensive tune of a MCEG, where the objective is not merely to improve the description of a single measurement but to achieve a more general improvement of particle-interaction simulations, it is essential to include a broad set of observables that probe different aspects of the underlying physics. While the inclusion of an additional parameter comes at a substantial computational cost, since it increases the dimensionality of the parameter space to be sampled, the cost of adding further observables scales only linearly with the number of `Rivet` analyses included in the tune. For this reason, it is common practice to include all observables from a given analysis when the data is considered relevant for the intended tuning objective.

In this work, however, the aim is to investigate the interplay between observables from collider and fixed-target experiments, to test different surrogate models for their description, and to study their compatibility within a common tune as well as the impact on the simulation of EAS. For these reasons, only a limited set of observables is selected. The focus is placed on observables related to particle multiplicities and momenta, since these are expected to be most directly relevant for the aspects of hadronic interactions that influence air-shower development.

The selected observables are summarized in Table 8.3. They include all observables discussed in Chapter 8.2, with the exception of the CMS measurement shown in Figure 8.5.

Table 8.3: `Rivet` analyses and observables selected for the tuning studies.

	RIVET plugin	Data	Observable	Ref
LEP	L3_2004_I652683	d59-x01-y02 d65-x01-y02	$P(N_{\text{ch}})$ $d\sigma/d\zeta_p$	[64]
	ALEPH_1996_I428072	d17-x01-y01	$d\sigma/d\zeta_p$	[65]
LHC	ATLAS_2010_I882098	d10-x01-y01 d17-x01-y01	$d^2 N_{\text{ch}}/d\eta dp_{\perp}$ $dN_{\text{ev}}/dN_{\text{ch}}$	[63]
fixed-target	EHS_1988_I265504	d06-x01-y01	$d\sigma/dp_T^2$	[61]

The CMS analysis was not included in the final tuning dataset despite its good modeling and its noticeable sensitivity to the selected parameters. The reason is that the lack of differences in the bins of the pseudorapidity observable does not add much constraint on the chosen parameters, as the bins behave similarly across changes in the parameter values.

8.4 Selection of the Surrogate Model

The importance of choosing an appropriate surrogate model was already discussed in Chapter 6.4. If the MCEG response is insufficiently parameterized, the observables' sensitivity to the model parameters is irrelevant, since the resulting posterior distribution may have a global mode that does not correspond to the true optimal value in the parameter space. One of the objectives of this study is therefore to examine the role of the surrogate-model parameterization by comparing two different approaches.

The first approach is based on piecewise linear interpolation, as already mentioned in Chapter 6.4. In this case, MC samples are generated on a regular grid in parameter space, and the response between neighboring grid points is obtained by linear interpolation. By construction, this surrogate model reproduces the sampled points exactly. However, it cannot capture the smoothness of the true generator response between the sampled points on the grid.

The second approach uses a multivariate polynomial of fifth order,

$$f(p_1, p_2, p_3) = \sum_{i+j+k \leq 5} c_{ijk} p_1^i p_2^j p_3^k, \quad (8.4)$$

for the three parameters p_1 , p_2 , and p_3 . The number of coefficients in a multivariate polynomial of order n_{order} with P parameters is given by

$$N_{\text{coeff}} = 1 + \sum_{i=1}^{n_{\text{order}}} \frac{1}{i!} \prod_{j=0}^{i-1} (P + j). \quad (8.5)$$

For a polynomial of order five in three parameters, this yields $N_{\text{coeff}} = 56$ coefficients. The surrogate model is therefore fully determined by a single coefficient vector $\vec{c}$.

For n simulated MC samples, the construction of the surrogate model can be written as

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\vec{y}} = \underbrace{\begin{pmatrix} 1 & p_{1,1} & \cdots & p_{1,1}p_{2,1}p_{3,1} & \cdots & p_{2,1}^3p_{3,1} & \cdots & p_{3,1}^5 \\ 1 & p_{1,2} & \cdots & p_{1,2}p_{2,2}p_{3,2} & \cdots & p_{2,2}^3p_{3,2} & \cdots & p_{3,2}^5 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & p_{1,n} & \cdots & p_{1,n}p_{2,n}p_{3,n} & \cdots & p_{2,n}^3p_{3,n} & \cdots & p_{3,n}^5 \end{pmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{56} \end{pmatrix}}_{\vec{c}}$$

Here, the vector $\vec{y}$ contains the value of an observable bin for each simulated MC sample, while the matrix $\mathbf{A}$ contains, in each of its n rows, the 56 monomials that define the polynomial basis.

An exact solution of this linear system requires the number of MC samples to be equal to the number of coefficients. In practice, however, n is typically larger than that, such that the system is overdetermined and no exact solution exists in general. The coefficients are therefore obtained by minimizing the residual in the Euclidean norm,

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{A}\mathbf{c} - \vec{y}\|_2 \quad (8.6)$$

In the `Julia` programming language, this is implemented through

```
c = A \ y
```

which, for an overdetermined system of the form $\mathbf{y} = \mathbf{A}*\mathbf{c}$, uses a pivoted QR factorization to determine the best fitting $\mathbf{c}$, corresponding to $\hat{\mathbf{c}}$ in Equation 8.6 [67].

Comparison of the Parameterization

To assess the performance of the two surrogate models, both are constructed from the same set of simulated MC samples. Since the linear interpolation requires samples on a regular grid in parameter space, the MC samples are generated for fixed parameter values along each dimension. Two sample sets are considered, corresponding to a $7 \times 7 \times 7$ grid and a $10 \times 10 \times 10$ grid. At each grid point, 10^7 events are simulated, resulting in two datasets with different sampling densities. The grid points are chosen within the chosen parameter ranges listed in Table 8.1 and are approximately equidistant.

Because each surrogate model describes the dependence of a histogram bin on three parameters, direct visualization is not straightforward. A useful representation is obtained by freezing, i.e. fixing, two parameters and varying only the third, thereby producing one-dimensional slices through the surrogate model. An example is shown in Figure 8.10 for the 10^3 grid. The first bin of the observable `d17-x01-y01` from the ATLAS analysis is shown as a function of `sigma`, while `aLund` and `pT0Ref` are fixed to the values 0.67, 0.82 and 2.16, 2.51, respectively.

For the chosen parameter values, both surrogate models show similar behavior, with only minor differences visible in the ratio panel. The monotonic rise and fall in the ratio can be understood from the placement of the MC samples, which determine the piecewise linear interpolation exactly and are generally also well described by the polynomial. The deviations between the two models increase with distance from the sampled points. As shown in the upper figures, however, the polynomial

does not reproduce every MC sample exactly and can differ by as much as 10% at $\sigma = 0.8$. Nevertheless, both parameterizations provide a sufficient description for the selected bin and parameter values.

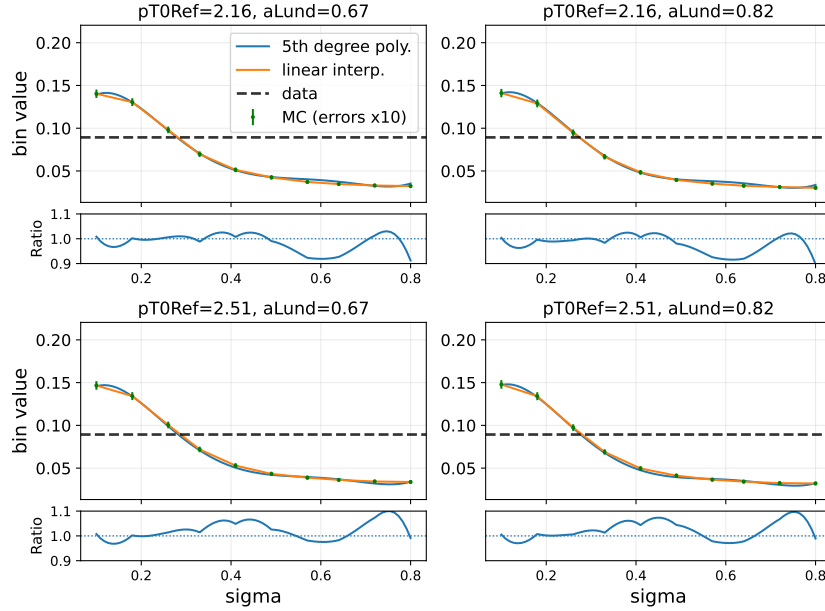


Figure 8.10: Content of the first bin of the observable `d17-x01-y01` from the Rivet analysis `ATLAS_2010_I882098` as a function of `sigma`, by freezing `aLund` and `pT0Ref`. The surrogate models obtained from piecewise linear interpolation and from a fifth-order polynomial fit to the 10^3 grid are shown as blue and orange curves, respectively. The data are shown as a black dashed line and the MC samples as green points. The MC uncertainties are scaled by a factor of 10 for visibility. The lower panel shows the ratio of the linear interpolation to the polynomial parameterization.

The quality of the parameterization depends not only on the chosen surrogate model, but also on the quality of the underlying MC modeling and the statistical population of the bin, since the bin content is assumed to be a sufficiently stable estimate of the observable. It is therefore expected that the first bin of the ATLAS observable, which contains the largest number of entries, is reproduced comparatively well. To compare this with a bin of lower statistical population, the 20th bin of the same observable is shown in Figure 8.11 for the same frozen parameter values. Even for the 20th bin, both models show a satisfying agreement, with only a small increase in the deviations compared to the first bin.

For the LEP observables, the tenth bin of the observable `d65-x01-y02` from the L3 analysis is shown in Figure 8.12 for the 10^3 grid, with `aLund` and `pT0Ref` fixed to

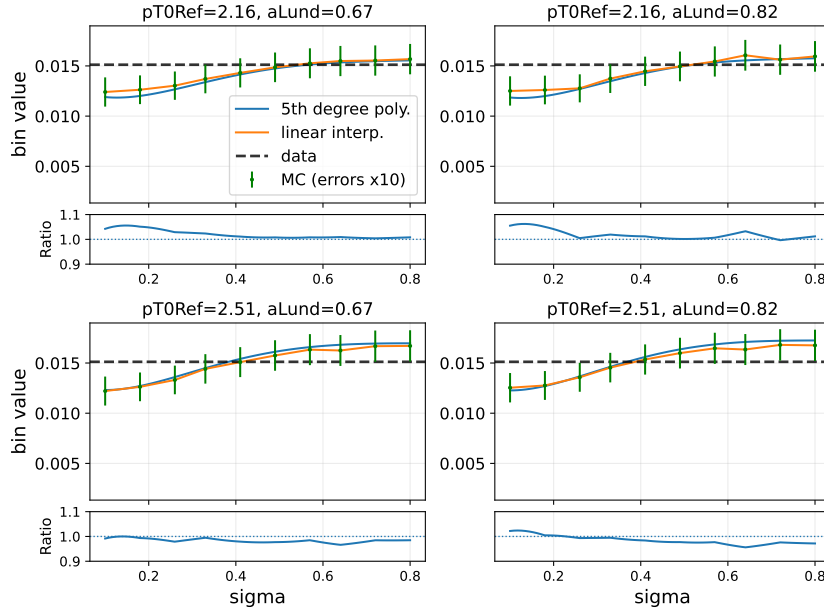


Figure 8.11: Content of the 20th bin of the observable `d17-x01-y01` from the Rivet analysis `ATLAS_2010_I882098` as a function of `sigma`, by freezing `aLund` and `pT0Ref`. The surrogate models obtained from piecewise linear interpolation and from a fifth-order polynomial fit to the 10^3 grid are shown as blue and orange curves, respectively. The data are shown as a black dashed line and the MC samples as green points. The MC uncertainties are scaled by a factor of 10 for visibility. The lower panel shows the ratio of the linear interpolation to the polynomial parameterization.

0.67, 0.9 and 2.33, 2.87, respectively. Although this bin does not correspond to the most populated region of the histogram, both surrogate models again provide a good description of the MC samples. Figure 8.13 shows the 20th bin of the fixed-target observable `d06-x01-y01` from the EHS/NA22 analysis for the 10^3 grid and the same fixed parameter values as before. Again, both surrogate models exhibit a good parameterization of the MC samples, showing only smaller deviations for low `sigma` values around 0.1.

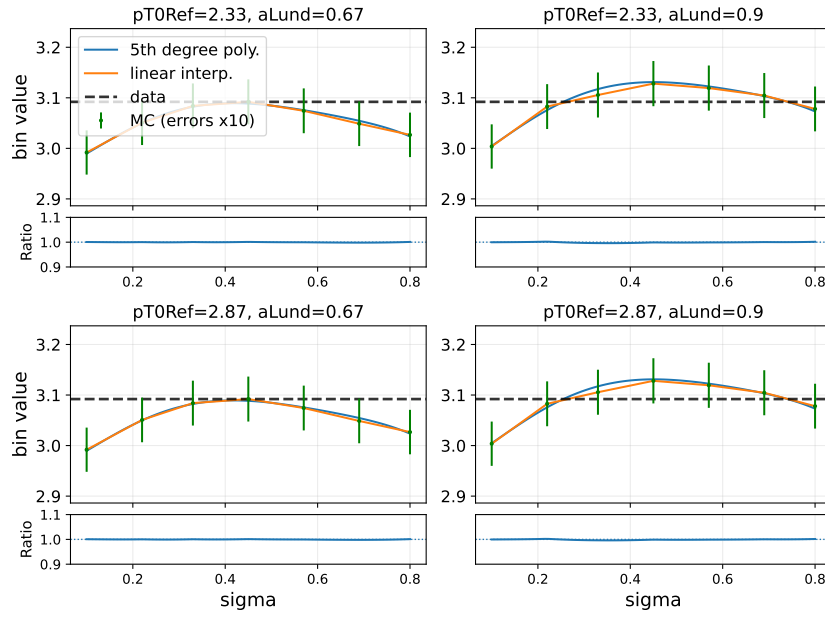


Figure 8.12: Content of the tenth bin of the observable `d65-x01-y02` from the Rivet analysis `L3_2004_I652683` as a function of `sigma`, by freezing `aLund` and `pT0Ref`. The surrogate models obtained from piecewise linear interpolation and from a fifth-order polynomial fit to the 10^3 grid are shown as blue and orange curves, respectively. The data are shown as a black dashed line and the MC samples as green points. The MC uncertainties are scaled by a factor of 10 for visibility. The lower panel shows the ratio of the linear interpolation to the polynomial parameterization.

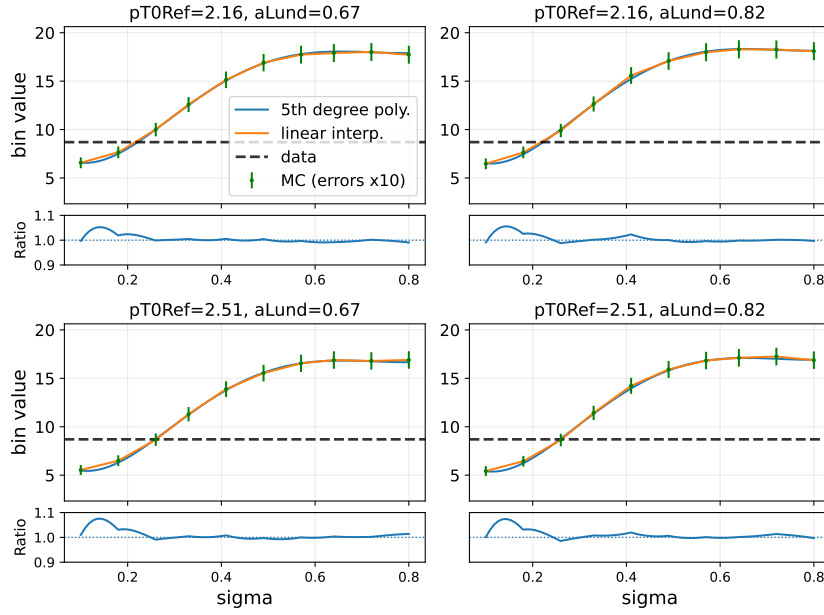


Figure 8.13: Content of the 20th bin of the observable `d06-x01-y01` from the Rivet analysis `EHS_1988_I265504` as a function of `sigma`, by freezing `aLund` and `pT0Ref`. The surrogate models obtained from piecewise linear interpolation and from a fifth-order polynomial fit to the 10^3 grid are shown as blue and orange curves, respectively. The data are shown as a black dashed line and the MC samples as green points. The MC uncertainties are scaled by a factor of 10 for visibility. The lower panel shows the ratio of the linear interpolation to the polynomial parameterization.

Overall, both surrogate models provide a sufficient description of the considered observables on the 10^3 grid. The deviations between the two models remain small, and the fifth-order polynomial reproduces the MC samples with sufficient accuracy in the examples shown.

To evaluate the impact of the sample density on the parameterization, the 10th bin of the ATLAS observable `d17-x01-y01` and EHS/NA22 observable `d06-x01-y01` are shown for the 7^3 grid in Figures 8.14 and 8.15. In both cases, the reduced density of MC samples leads to a visible deterioration of the parameterization quality and to larger deviations between the two surrogate models. Nevertheless, the overall performance on the sparser grid remains acceptable, despite using only $7^3/10^3 = 0.343 \approx 34\%$ of the number of MC samples.

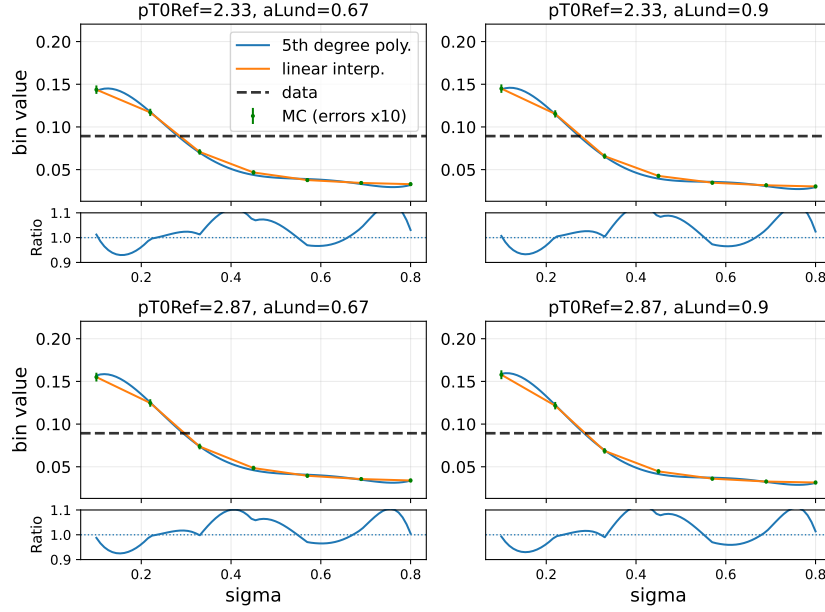


Figure 8.14: Content of the 10th bin of the observable `d17-x01-y01` from the Rivet analysis `ATLAS_2010_I882098` as a function of `sigma`, by freezing `aLund` and `pT0Ref`. The surrogate models obtained from piecewise linear interpolation and from a fifth-order polynomial fit to the 7^3 grid are shown as blue and orange curves, respectively. The data are shown as a black dashed line and the MC samples as green points. The MC uncertainties are scaled by a factor of 10 for visibility. The lower panel shows the ratio of the linear interpolation to the polynomial parameterization.

An important limitation that is not apparent from the figures shown here concerns the unknown performance of the surrogate models in regions of parameter space lying between the sampled values of `aLund` and `pT0Ref`. As illustrated in the figures, both

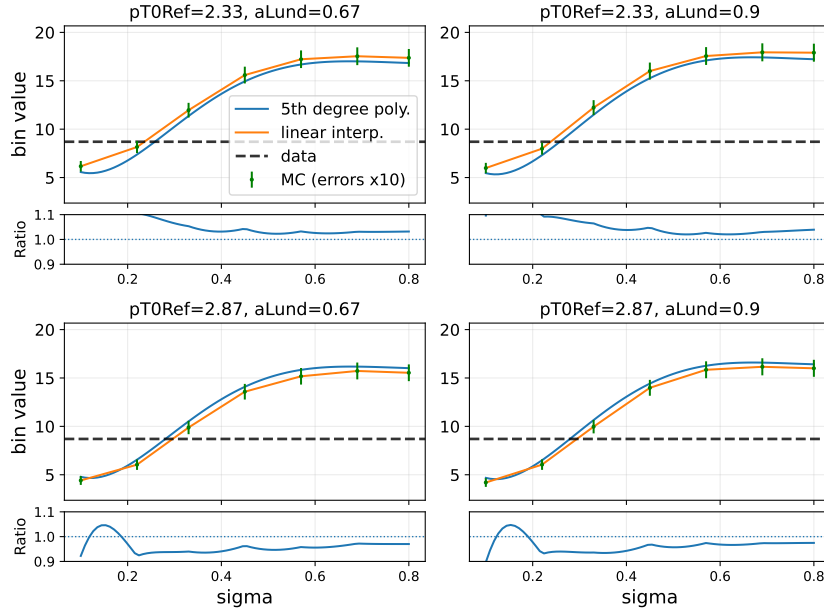


Figure 8.15: Content of the 10th bin of the observable `d06-x01-y01` from the Rivet analysis `EHS_1988_I265504` as a function of `sigma`, by freezing `aLund` and `pT0Ref`. The surrogate models obtained from piecewise linear interpolation and from a fifth-order polynomial fit to the 7^3 grid are shown as blue and orange curves, respectively. The data are shown as a black dashed line and the MC samples as green points. The MC uncertainties are scaled by a factor of 10 for visibility. The lower panel shows the ratio of the linear interpolation to the polynomial parameterization.

models reproduce the sampled points reasonably well, but differ in the unsampled intervals between the chosen `sigma` values. The same issue also applies to the two frozen dimensions. This aspect will become visible in more detail in Chapter 8.7.

Overall, the performance on the 10^3 grid is considered sufficient for both surrogate models in the subsequent studies. In the remainder of this work, only the 10^3 grid is therefore used for the construction of the surrogate models.

8.5 Analysis of the Posterior Distributions

Having access to both the surrogate model and the corresponding data through the Rivet framework, the likelihood can be constructed as described in Chapter 6.5. To this end, a prior for the parameters is required, which in these studies is simply a

uniform distribution over the chosen parameter ranges. After sampling the posterior with the Metropolis-Hastings algorithm, the resulting posterior distribution can be represented through marginalized one- or two-dimensional posterior distributions, obtained by integrating over the remaining parameter dimensions. The smallest 90% credibility regions of the marginalized two-dimensional posterior distributions for the two surrogate models are shown in Figure 8.16.

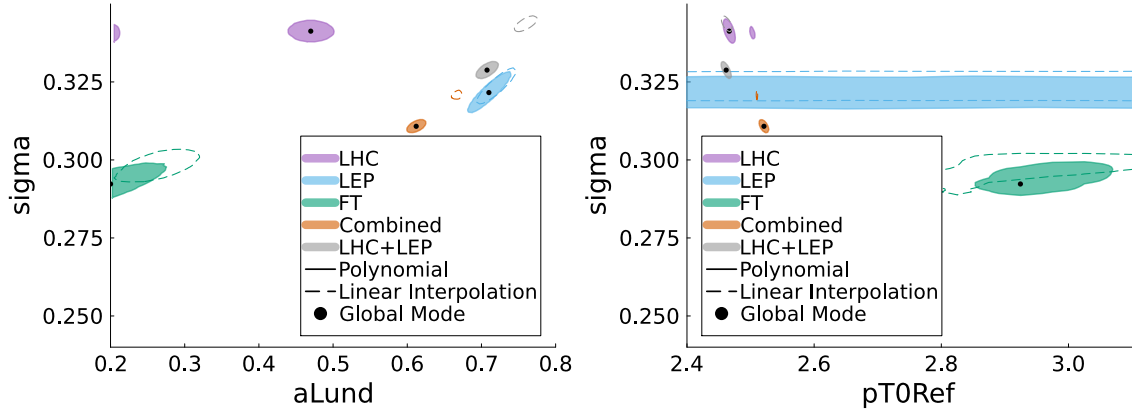


Figure 8.16: Smallest 90% credibility regions of the marginalized two-dimensional posterior distributions for the two surrogate models. The posterior distributions are shown for the full observable set (Combined) and for likelihoods constructed from selected subsets of observables only. The black dots indicate the global modes of the posterior obtained with the fifth-order polynomial surrogate model.

The posterior distributions labeled as "Combined" are obtained from a likelihood that includes every bin of all observables discussed above. In order to study the interplay between different observables and their impact on the posterior, additional likelihoods are constructed using only selected subsets of observables. Figure 8.16 therefore also shows the posterior distributions obtained when the likelihood is built from LHC observables only, LEP observables only, LHC and LEP observables combined (LHC+LEP), and fixed-target (FT) observables only.

Before discussing the position and shape of the posterior distributions in more detail, it is useful to compare the two surrogate models. The differences between the corresponding credibility regions are generally small. The largest deviations are visible for the FT, LHC+LEP, and Combined configurations, whereas the posteriors obtained from LHC-only and LEP-only observables are very similar for the two models. In view of these comparatively minor differences, only the fifth-order polynomial surrogate model is used in the remainder of the study. This choice is primarily motivated by the smoothness of the parameterization between sampled

MC points, which is expected to provide a more realistic approximation of the true generator response, even though the linear interpolation reproduces each sampled point exactly.

The posterior distributions shown in Figure 8.16 reveals that the LEP, LHC, and FT configurations each occupy distinct regions in the marginalized posterior space. Since these regions, or islands, do not overlap within the shown 90% credibility regions, the construction of a single common tune that simultaneously improves the description of all measurements appears difficult. In particular, no single combination of `pT0Ref`, `aLund`, and `sigma` is able to reach, or get close to, all three islands at the same time.

The absence of sensitivity of the LEP observables to `pT0Ref` is reflected in the broad posterior band of the LEP configuration along the `pT0Ref` direction. The LHC posterior, by contrast, shows two distinct solution regions in the `aLund`-`pT0Ref` plane. It is not unusual for a posterior to exhibit a double peak structure, as the combination of multiple parameters often allows for multiple solutions when combining different parameter values.

Another feature worth noting is the small interval of `sigma` values covered by all credibility intervals, especially considering that `sigma` was sampled in the range 0.1 to 0.8. This indicates that the relevant region of parameter space is only sparsely sampled, since the chosen `sigma` interval is considerably larger than necessary. In addition, the chosen lower limit of `aLund` appears to be too high, as both the LHC and the fixed-target posterior islands extend towards lower values of this parameter than those sampled. For this reason, a new set of MC samples is generated on a 10^3 grid covering a reduced parameter range. While `pT0Ref` is still sampled between 1.8 and 5.0, the ranges of `aLund` and `sigma` are reduced to 0.0-1.2 and 0.2-0.4, respectively.

Figure 8.17 shows the impact of reducing the parameter ranges on the marginalized posteriors. Here, "Full Range" refers to the original parameter range, while the filled posterior islands correspond to the posterior obtained from the newly restricted parameter ranges. Although the sampling density is higher in the reduced range, the difference is comparatively minor. It appears that mainly the LHC observables are affected, since the posterior islands obtained using only LEP or only fixed-target observables show little to no shift. In contrast, the posterior island for the LHC observables not only shifts, but the previously observed double-peak structure also merges into one connected island.

The posterior using all observables is still an isolated island, with no overlap with the other posteriors within the 90% credibility interval. The idea of constructing one combined tune for all observables therefore appears unreasonable, since the

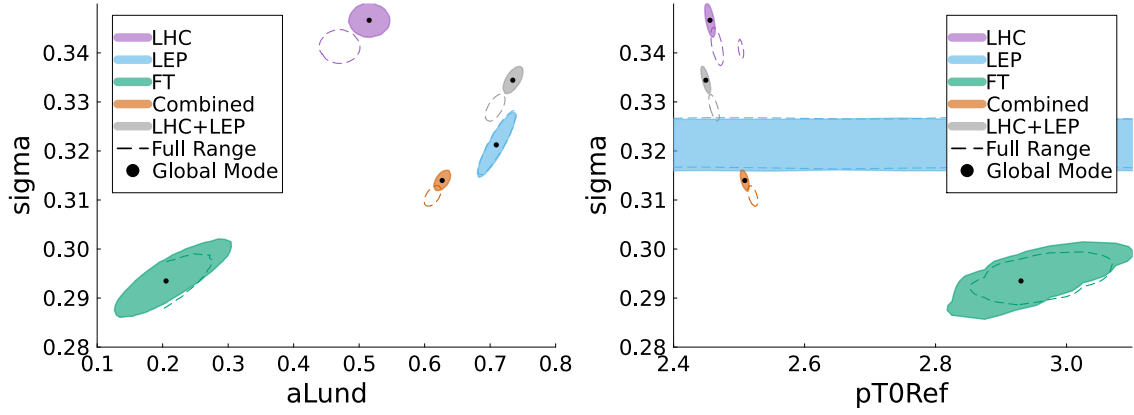


Figure 8.17: Smallest 90% intervals of the marginalized two-dimensional posterior distributions for the fifth-degree polynomial surrogate model. The filled marginals result from using MC samples from the reduced parameter range, while the dashed lines indicate the marginals obtained using the original larger parameter range. The posteriors are shown for all observables combined, and for the case of using only a subset of the observables in the likelihood. The global mode of each posterior using the reduced range is marked as a black dot.

resulting change to the `Pythia 8` model would likely not provide an improvement in the description of the individual measurements. This follows from the fact that the global mode of the combined posterior island lies far from the global modes of the posteriors obtained from the corresponding subsets of observables. In particular, the posterior island using only the fixed-target observable shows a large deviation in `pT0Ref` from the LHC posterior and a moderate deviation in `aLund` from both the LHC and LEP posteriors. By contrast, the posteriors using only the LHC and only the LEP observables are more compatible, showing smaller deviations in `sigma` and `aLund`. For this reason, two tunes are constructed and investigated further instead of one combined tune. One tune is defined by the global mode of the posterior using only the fixed-target observable, while the other is identified as the global mode of the LHC+LEP posterior. In addition, the global modes of the posteriors using only the LHC and only the LEP observables are also further investigated in order to allow a direct comparison between the LHC+LEP tune and the corresponding LHC-only or LEP-only tunes.

It is important to note that the choice of a fixed-target tune and an LHC+LEP tune is based on the posterior study after grouping the observables into LHC, LEP, and fixed-target categories. One could also argue that it would be more appropriate to investigate each individual observable separately instead of clustering them. This is shown in Figure 8.18, which contains the smallest 90% credibility intervals of

the one-dimensional marginalized posterior distributions for each observable and each parameter. As can be seen, the posteriors are spread over a large part of the chosen parameter ranges, such that the relative closeness of the LHC and LEP posteriors in `aLund`, compared to the fixed-target posterior, is less visible than in the previous figures. The larger difference in `pT0Ref` between the LHC and fixed-target observables remains visible, with both LHC observables resulting in posteriors that prefer values close to 2.3, while the fixed-target posterior lies around 2.9.

Nevertheless, as already mentioned, a complete tune of an MCEG such as `Pythia 8` would generally include a large number of measurements covering different physical processes relevant for the event generator. In addition, previous `Pythia 8` tunes [34, 58] focused on LEP and LHC measurements in order to improve the simulation of collider events, while fixed-target data were not included. This is also reflected in the insufficient modeling of the EHS/NA22 observable, as already seen in Figure 8.9. It is therefore considered sufficient for the present study to separate the observables into fixed-target and LHC+LEP categories, while a more detailed investigation of the isolation of the posterior islands under the inclusion of further observables is left to future work.

The resulting tuned values for the `Pythia 8` parameters for the LHC+LEP, fixed-target, LHC, and LEP tunes are listed in Table 8.4. They are identified as the global modes of the posteriors labeled LHC+LEP, FT, LHC, and LEP in Figure 8.17. In order to assign an uncertainty to the tuned parameter values, the half-width of the smallest 68% interval of the one-dimensional marginalized posterior distributions is used, since these distributions are approximately Gaussian for each posterior, except for the LEP observable for `pT0Ref`. This is shown in Figures 8.19 and 8.20 for the fixed-target and LHC+LEP posteriors, respectively. The LHC and LEP posteriors are shown in Appendix A.2. Although this definition is suitable for the present purpose, it should be noted that marginalized posterior distributions do not in general have to be Gaussian. This can already be seen for the fixed-target tune in Figure 8.19, which exhibits a minor secondary solution that becomes visible only in the smallest 99% interval, making the definition of an uncertainty less straightforward. In this case, however, the secondary peak is sufficiently small that the width of the dominant peak is used.

Comparing the resulting tuned parameter values to the default values, it is noticeable that the LHC+LEP tune remains much closer to the default than the fixed-target tune. This is particularly visible for the `sigma` parameter, which showed the highest sensitivity for the observables considered here and whose tuned value in the LHC+LEP configuration is almost identical to the default value. This behavior is expected, since the default description of the collider observables was already reasonably good, whereas the fixed-target observable showed a clearly

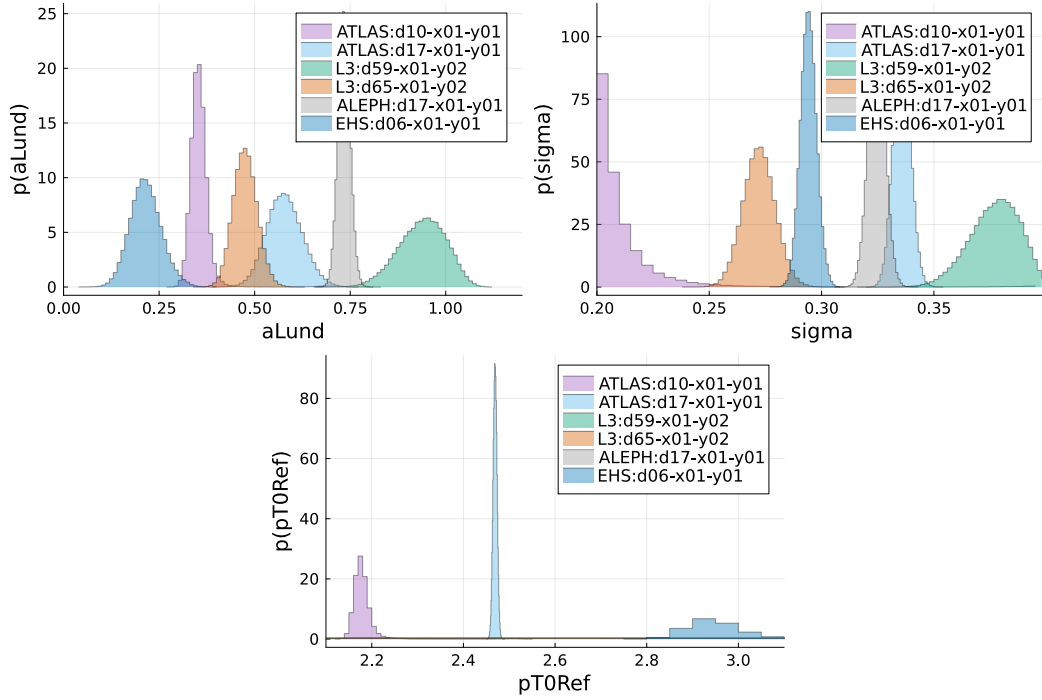


Figure 8.18: Smallest 90% intervals of the marginalized one-dimensional posterior distributions. The posteriors are shown for all observables, labeled by experiment and by the observable code in the corresponding Rivet routine.

insufficient description. It was therefore also expected that the tuned parameter values obtained from the fixed-target posterior would deviate more strongly from the default settings.

Looking at the separate global modes of the LHC and LEP posteriors, one finds that both tunes result in values of a_{Lund} and σ that remain close to the default settings. The p_{T0Ref} parameter does not affect the LEP observables, but remains close to the default value for the LHC posterior.

Table 8.4: Values of the parameters `pT0Ref`, `aLund`, and `sigma` for the default settings of `Pythia 8`, the LHC+LEP tune, the fixed-target tune, the LHC tune, and the LEP tune. The uncertainty on the tuned values is defined as the half-width of the smallest 68% interval of the corresponding marginalized posterior distributions.

	<code>pT0Ref</code>	<code>aLund</code>	<code>sigma</code>
default	2.28	0.68	0.335
LHC+LEP	2.449 ± 0.003	0.735 ± 0.008	0.334 ± 0.001
FT	2.930 ± 0.058	0.205 ± 0.041	0.293 ± 0.004
LHC	2.456 ± 0.004	0.515 ± 0.014	0.347 ± 0.002
LEP	—	0.709 ± 0.014	0.321 ± 0.003

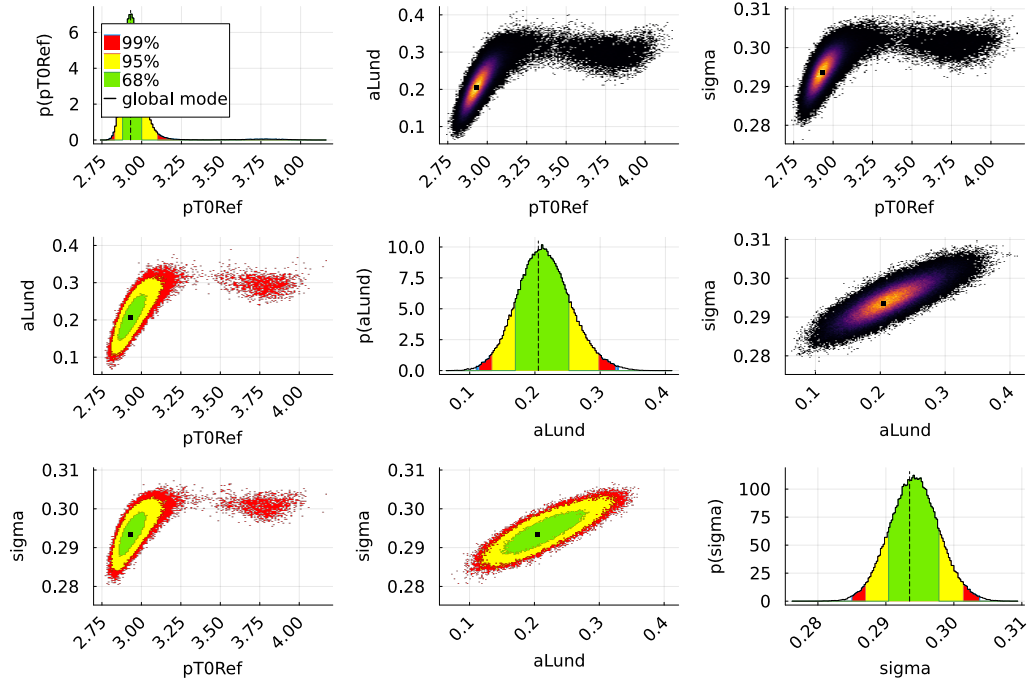


Figure 8.19: One- and two-dimensional marginalized posterior distributions of the parameters `pT0Ref`, `aLund`, and `sigma` for the fixed-target tune. The marginalized distributions are shown for the smallest 68%, 95%, and 99% intervals, while the heat maps of the two-dimensional marginalized posterior distributions are shown in the upper-right triangle.

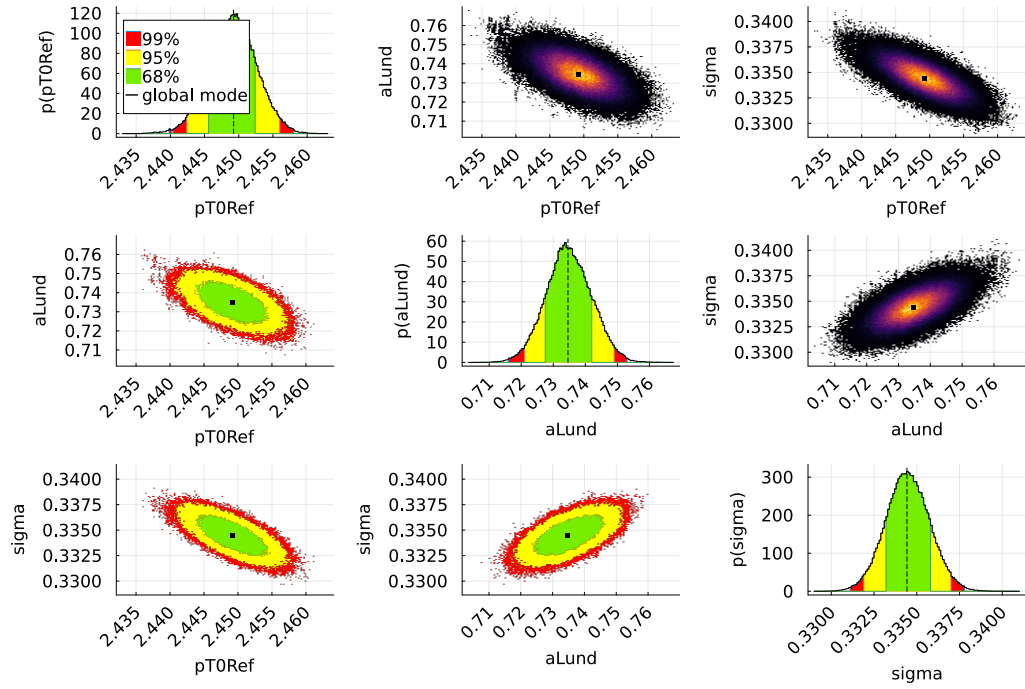


Figure 8.20: One- and two-dimensional marginalized posterior distributions of the parameters $pT0Ref$, $aLund$, and σ for the LHC+LEP tune. The marginalized distributions are shown for the smallest 68%, 95%, and 99% intervals, while the heat maps of the two-dimensional marginalized posterior distributions are shown in the upper-right triangle.

8.6 Impact of Tunes on the Observables

To assess the performance of the tunes, the global modes of the posteriors obtained with the fifth-degree polynomial surrogate model are used as input parameter values in `Pythia 8`. For each of these parameter settings, tuned MC samples of the observables are generated with 10^6 events. For the LEP and LHC observables, this is done using the tunes obtained from LEP-only and LHC-only observables, respectively, as well as the combined LHC+LEP observables. For the fixed-target observable, only the tune based on the fixed-target observable is considered. The quality of the resulting descriptions is quantified by means of the reduced chi-square test, as introduced in Chapter 6.6. The corresponding values are listed in Table 8.5.

Table 8.5: Reduced chi-square values for samples simulated with the default `Pythia 8` settings and the different tunes, evaluated with respect to the corresponding data for each observable.

χ_{red}^2	default	LHC+LEP	FT	LHC	LEP
ATLAS: d10-x01-y01	4.38	10.74	71.68	2.81	4.79
ATLAS: d17-x01-y01	197.55	108.21	466.23	101.06	281.14
L3: d59-x01-y02	0.93	0.89	39.75	13.64	1.38
L3: d65-x01-y02	3.98	3.79	15.26	9.20	3.19
ALEPH: d17-x01-y01	12.50	8.93	128.93	54.66	8.46
EHS/NA22: d06-x01-y01	89.80	61.40	3.83	58.96	84.49

The reduced chi-square values obtained with the tunes reveal a clear pattern across the different observables. Tunes tend to improve the description of the observables used in their respective tuning procedures, but these improvements are often accompanied by a deterioration in the description of the remaining observables relative to the default model. While each tune also yields a small improvement in the fixed-target observables, this effect is minor compared to that obtained with the dedicated fixed-target tune. These results show that the tuning procedure is effective for the observables used as input. At the same time, the anticorrelation between the improvements in different observable groups supports the decision not to construct a single combined tune including all observables.

The two ATLAS observables are shown in Figure 8.21. For the observable displayed in the left panel, both the LHC+LEP and the LHC tune improve the description of the first bins, which contain the largest number of entries. This improvement is, however, accompanied by a deterioration in the description of the bins between 1 and 2 GeV for the LHC tune. For the LHC+LEP tune, this deterioration extends to even larger transverse momenta, reaching up to 100 GeV. For the second observable, the effect of the tunes is less straightforward to interpret. For each tune and the default configuration, *Pythia 8* first underestimates the data in the high-statistics bins and then overestimates them around $N_{\text{ch}} \approx 5$. Both tunes appear to improve the description of the bins at larger values of N_{ch} .

This behavior is also reflected in the reduced chi-square values. For the observable shown in the left panel, the LHC+LEP tune leads to a larger chi-square value than the default configuration, despite the visible improvement in the first bins. By contrast, the LHC tune yields a reduced chi-square value below the default one, hinting at the small modeling differences visible in the ratio as possible reason for the higher chi-square value of the LHC+LEP tune. For the second observable, both tunes result in smaller reduced chi-square values and therefore indicate an improved overall description.

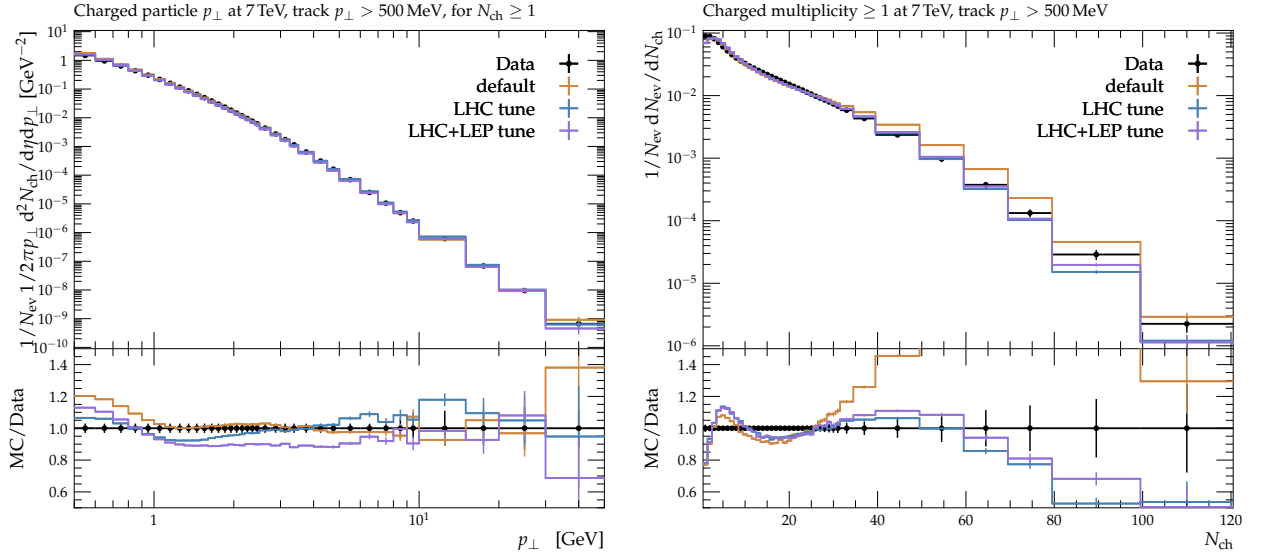


Figure 8.21: The observables d10-x01-y01 (left) and d17-x01-y01 (right) from the selected ATLAS analysis, simulated with the default *Pythia 8* settings and with the tuned parameter values obtained from the LHC+LEP and LHC tunes.

Figure 8.22 shows the observables from the L3 and ALEPH experiments, simulated

with the LHC+LEP and LEP tunes, together with the default `Pythia 8` prediction. For all three observables, the default description is already satisfactory and reproduces most bins well. The most pronounced visible difference appears in the ratio plot of the `d59-x01-y02` observable from the L3 analysis. In this case, the first bins up to approximately $N_{\text{ch}} \approx 18$ are described slightly better by the LHC+LEP and LEP tunes, whereas the bins at larger multiplicities, $N_{\text{ch}} > 32$, are better reproduced by the default settings. The reduced chi-square values confirm that both tunes lead to only modest changes for the LEP observables. In most cases, these changes correspond to small improvements, with the only exception being the LEP tune for the `d59-x01-y02` observable, for which the chi-square value increases slightly with respect to the default.

The remaining observable is `d06-x01-y01` from the EHS/NA22 experiment, shown in Figure 8.23. In this case, the effect of the tune is much more pronounced than for the collider observables. The fixed-target tune yields a substantially improved description compared to the default. Whereas the default `Pythia 8` configuration fails to reproduce the observable satisfactorily, the fixed-target tune provides a reasonable description for most bins. This improvement is also clearly visible in the reduced chi-square values, which decrease strongly relative to the default configuration.

The current default settings of `Pythia 8` is essentially the Monash tune [34], which was constructed using LEP and LHC measurements. It is therefore not surprising that the LHC+LEP, LHC, and LEP tunes lead only to relatively small improvements compared to the default configuration. Nevertheless, the reduced chi-square values show that in most cases some improvement is still possible. At the same time, the tunes considered here are based on only a small set of observables, which makes a direct comparison to a more comprehensive tune difficult. It therefore remains unclear whether the inclusion of a much larger set of observables would lead to a stronger or weaker improvement relative to the default settings. The strong improvement obtained for the fixed-target observable is, by contrast, expected, since the default `Pythia 8` configuration is not optimized for this class of measurements.

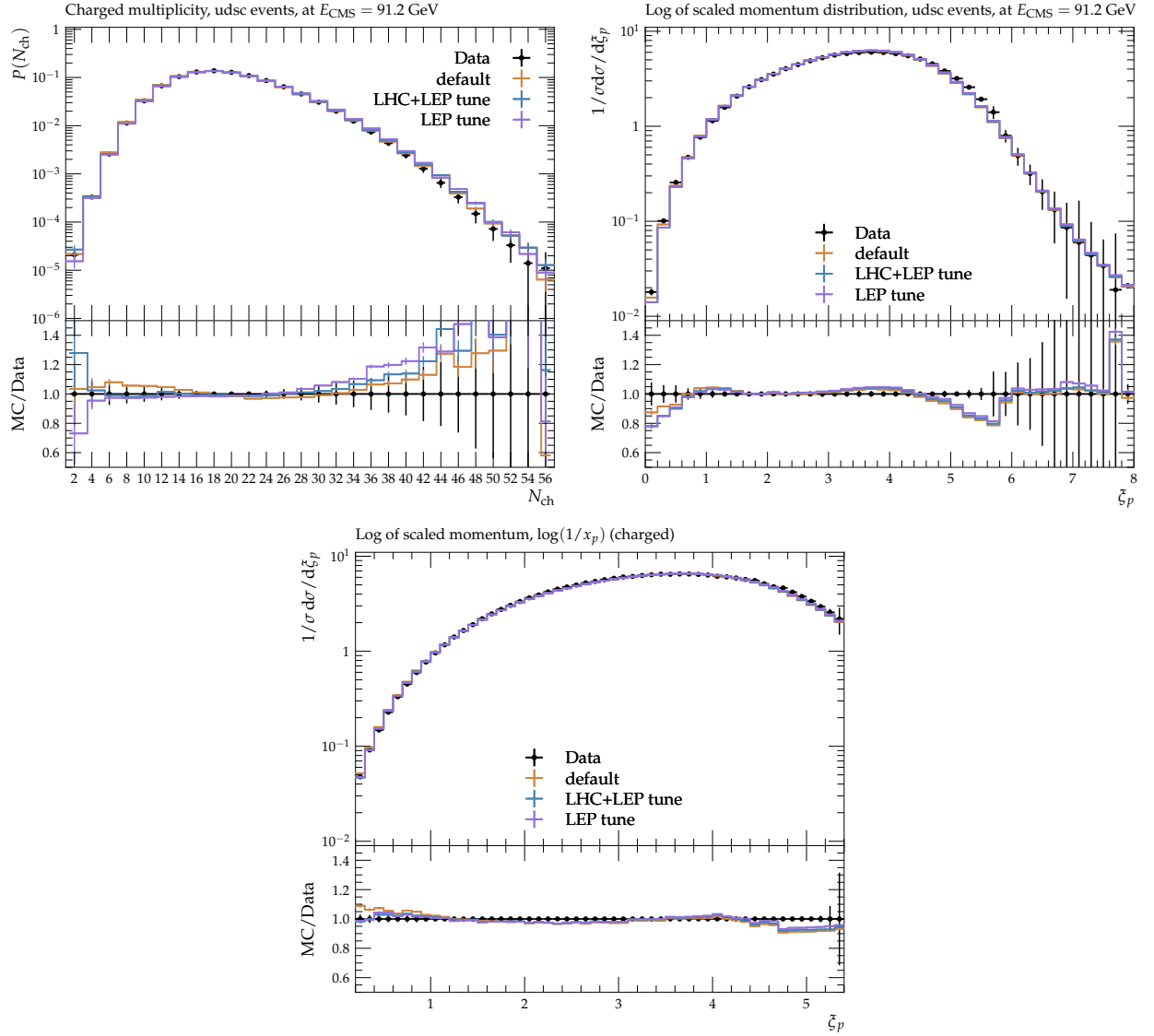


Figure 8.22: The observables d59-x01-y02 (upper left) and d65-x01-y02 (upper right) from the selected L3 analysis, together with the observable d17-x01-y01 (bottom) from the ALEPH analysis, simulated with the default `Pythia 8` settings and with the tuned parameter values obtained from the LHC+LEP and LEP tunes.

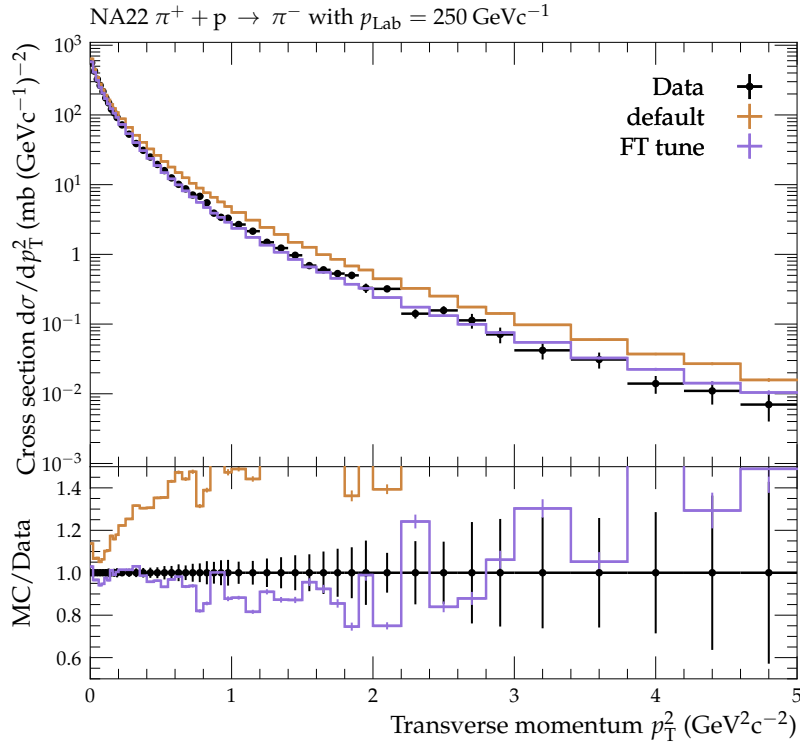


Figure 8.23: The observable d06-x01-y01 from the selected EHS/NA22 analysis, simulated with the default Pythia 8 settings and with the tuned parameter values obtained from the fixed-target tune.

8.7 Uncertainty Propagation

As already discussed in Chapter 6.7, one of the advantages of a tuning workflow based on Bayesian methods is the direct access to the posterior distribution in parameter space. This not only allows for a direct definition of the tuning uncertainty, as used in Chapter 8.6, but also makes it possible to study the shape of the marginalized posteriors and thereby identify potential secondary tuning solutions. A further advantage is the possibility of propagating the uncertainty on the tuned parameters to an uncertainty on the simulated observables. This is particularly relevant because, from the perspective of a user of the tuned MCEG, the impact of the tuning uncertainty on simulated predictions is generally of greater interest than the uncertainty on the parameter values themselves.

The propagated uncertainty is obtained by sampling the posterior and evaluating the surrogate model for each bin at the sampled parameter points. In this way, a distribution of bin values is obtained for every bin of the observable. The width of this distribution is then interpreted as the propagated uncertainty on the observable. This procedure, however, has important limitations, namely that it assumes the surrogate model to be an accurate representation of the MC response. While this assumption also affects the tuning workflow, its impact is more severe on uncertainty propagation. In the tuning procedure, a modest deviation between the surrogate model and the true MC response may result in only a small shift in the global mode relative to the true best-fit parameter value. In the uncertainty propagation, by contrast, the same deviation can lead to a visibly misplaced distribution of bin values. In addition, the width of the posterior distribution is closely linked to the experimental uncertainties of the data, which determine how strongly the likelihood constrains the parameter space. Consequently, the propagated uncertainty also depends strongly on the uncertainties of the data.

Figure 8.24 shows the observable `d17-x01-y01` from the ATLAS analysis and the observable `d17-x01-y01` from the ALEPH analysis, together with the propagated uncertainty obtained from the corresponding LHC and LEP tunes. For the ALEPH observable, the surrogate model provides a good approximation of the new samples simulated with the LEP tune. Apart from a few low-statistics bins at the beginning of the distribution, most bins are reproduced well, with an uncertainty band around the tuned prediction. For most bins, this uncertainty band is very narrow, indicating that the propagated tuning uncertainty on the observable is small. For the ATLAS observable, on the other hand, the parameterization is less accurate, and the surrogate model is unable to reproduce the bins around $N_{\text{ch}} \approx 18$ and above 40 sufficient.

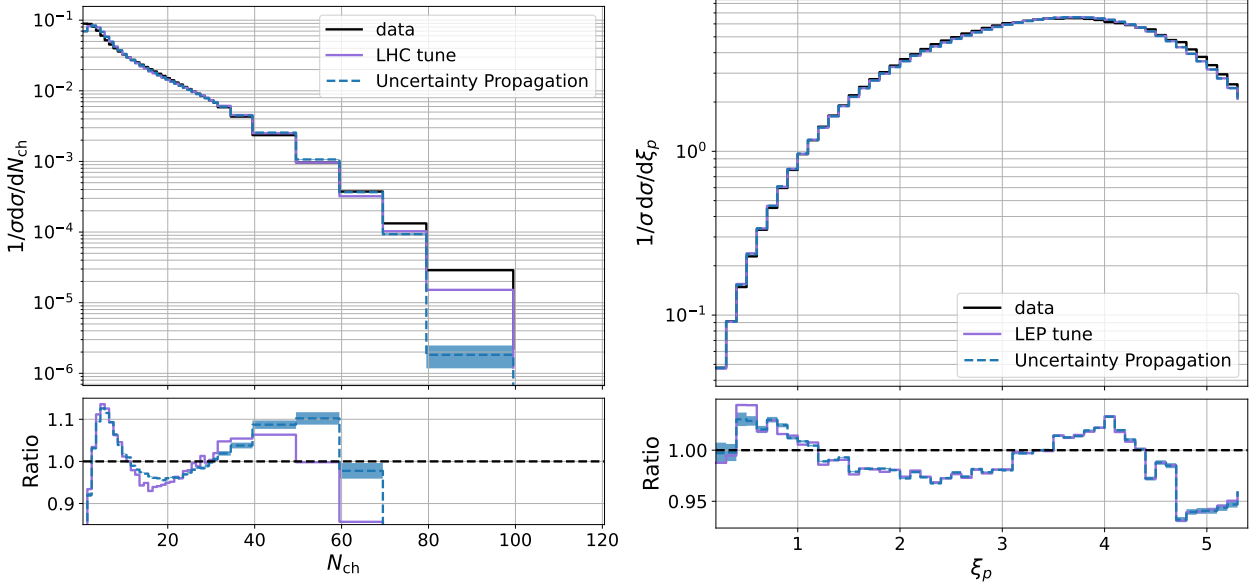


Figure 8.24: The observable `d17-x01-y01` (left) from the ATLAS analysis and the observable `d17-x01-y01` (right) from the ALEPH analysis. The MC samples are simulated using the LHC and LEP tunes, respectively. The uncertainty bands are obtained by propagating the posterior uncertainty through the surrogate model, as described in the text.

The uncertainty propagation for the fixed-target observable `d06-x01-y01` from the EHS/NA22 analysis is shown in Figure 8.25. In this case, the approximation provided by the surrogate model is noticeably worse, as it fails to reproduce most of the bin values of the MC sample simulated with the fixed-target tune. The first high-statistics bins are approximated reasonably well, but the agreement deteriorates strongly for bins above $p_T^2 \approx 0.5 \text{ GeV}^2$.

These results indicate that the uncertainty propagation is not sufficiently reliable in the present study, since the surrogate model does not approximate the MC response accurately enough for the ATLAS and fixed-target observables. An alternative strategy would be to use a subset of points sampled from the posterior directly as input for the MCEG and to generate new MC samples, rather than relying on the surrogate model. The obvious drawback of such an approach is the additional computational cost required for these extra simulations. This cost could be reduced by approximating the posterior with a multivariate Gaussian and again defining the parameter uncertainty through the half-width of the corresponding distribution, as done in Section 8.5. In that case, it would be enough for the uncertainty propagation to simulate the samples defining the width of the posterior.

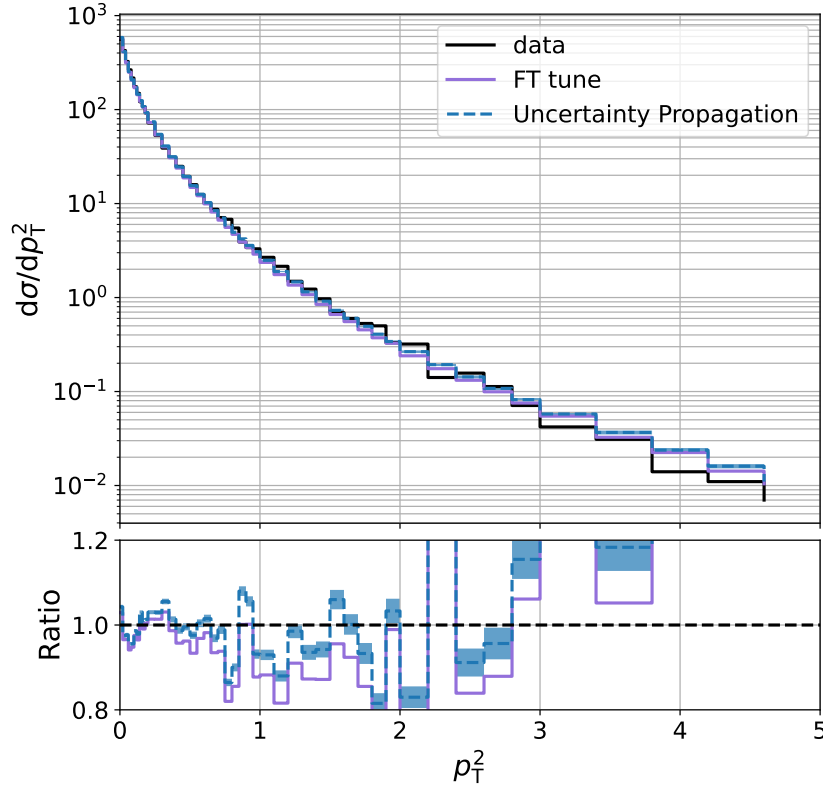


Figure 8.25: The observable d06-x01-y01 from the EHS/NA22 analysis. The MC sample is simulated using the fixed-target tune. The uncertainty band is obtained by propagating the posterior uncertainty through the surrogate model, as described in the text.

8.8 Limitations of the Parameterization

The results presented in this chapter indicate that the surrogate model is the main limiting factor in the tuning of an MCEG such as `Pythia 8`. It was shown that not only the choice of surrogate model, but also the density of the sampled MC points, affects the shape and position of the marginalized posterior distributions. In addition, deviations between the surrogate model and the MC samples lead to an insufficient uncertainty propagation. Furthermore, although Figures 8.10, 8.11, 8.12, and 8.13 suggest a sufficient parameterization of the sampled points, Figure 8.25 reveals a poor description of samples that were not used in the construction of the surrogate model. If the surrogate model does not provide a sufficiently accurate approximation of the MC sample at the global mode, the resulting tuning outcome

is necessarily affected.

It is therefore essential to evaluate the quality of the parameterization in a more systematic way. Figure 8.26 shows the reduced chi-square for each bin of each observable, comparing the surrogate model to the MC samples used in the parameterization. The χ^2 value is computed as the squared difference between the bin content of the MC sample and that of the surrogate model, divided by the uncertainty of the MC sample. The number of degrees of freedom ν , by which χ^2 is divided, is taken to be the difference between the number of MC samples and the number of coefficients in the polynomial, with $n_{\text{coeff}} = 56$.

For the LEP observables, all bins appear to be approximated sufficiently by the surrogate model, with reduced chi-square values close to 1. This is consistent with the behavior already seen for the LEP observable in Figure 8.24, where the surrogate model provides a good approximation of the simulated samples.

The LHC and fixed-target observables, on the other hand, show reduced chi-square values well above 1, confirming the poor approximation visible for the ATLAS observable in Figure 8.24 and for the fixed-target observable in Figure 8.25. The periodic oscillations visible in the chi-square distributions of the LHC and fixed-target observables originate from the changing bin widths of the corresponding histograms.

In order to substantiate these observations, the pull and the relative residuals are also calculated for the individual bins, since the chi-square value produces only one compressed, summarized value for each bin. The pull distribution is defined as the difference between the surrogate model bin value $N_{5\text{th}}$ and the MC sample bin value N_{MC} , divided by the MC uncertainty σ_{MC} . If the approximation is good, the resulting distribution should be centered around 0 and have a width of approximately 1. It therefore allows a direct assessment of both bias and model quality: a shift away from 0 indicates a bias, while a width significantly larger than $[-1, 1]$ indicates a poor approximation. The relative residuals are defined as the difference between $N_{5\text{th}}$ and N_{MC} divided by N_{MC} . They therefore quantify the relative deviation with respect to the bin value itself.

Figure 8.27 shows the pull distribution and the relative residuals for four different bins of the ATLAS observable `d17-x01-y01`. Since the MC samples consist of only 10^3 points, the resulting distributions often appear noisy, such that the width can only be estimated approximately. As expected, the pull distributions are considerably broader than $[-1, 1]$, indicating that the parameterization does not provide a sufficiently accurate approximation. The increasing width of the relative residual distribution towards larger bin numbers further reflects the increasing deviation in bins with lower statistics.

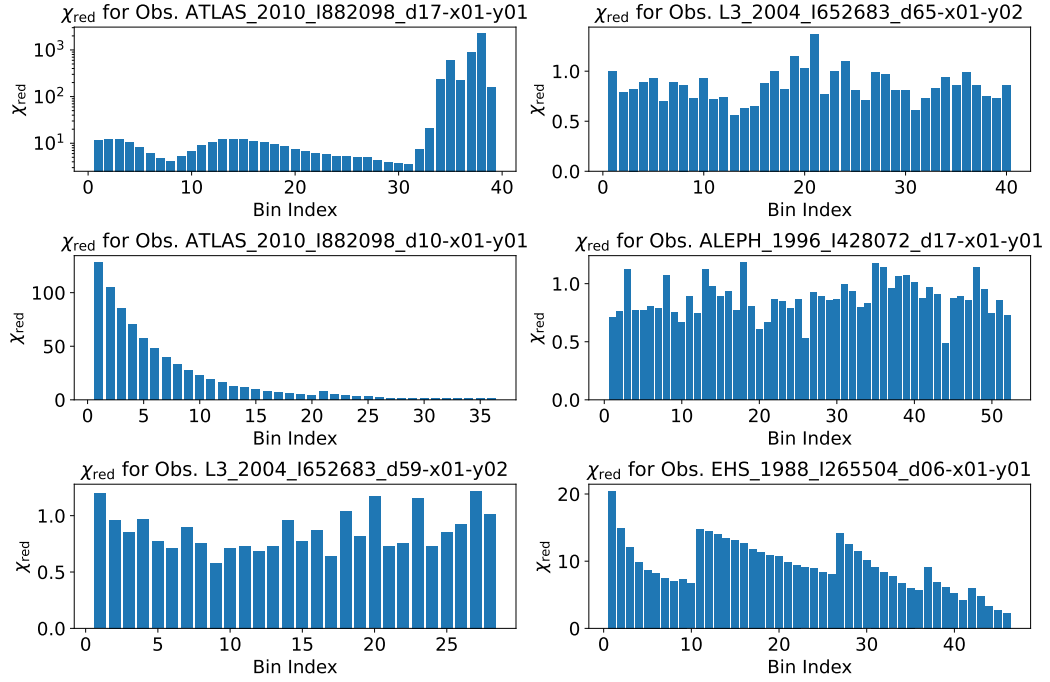


Figure 8.26: Reduced chi-square value for each bin of each observable.

The pull and relative residual distributions for the LEP observable `d65-x01-y02` from the L3 analysis are shown in Figure 8.28. Here, the pull distributions are narrower, with widths close to 1, indicating that the surrogate model provides a good description of the MC response. The same conclusion follows from the relative residuals, which remain narrow in the high-statistics bins.

Figure 8.29 shows the pull and relative residual distributions for the fixed-target observable `d06-x01-y01` from the EHS/NA22 analysis. These distributions are similar to those of the ATLAS observable, exhibiting broad pull distributions and correspondingly large relative residuals.

As a final test, the pull distribution is calculated for the observable `d17-x01-y01` from the ATLAS analysis and for the observable `d65-x01-y02` from the L3 analysis using MC samples generated on a 7^3 grid, in order to study the effect of the sample density. The result is shown in Figure 8.30. Since the distributions now contain only $7^3 = 343$ points, they appear even more noisy than before. Nevertheless, the same trend remains visible: the pull distribution for the ATLAS observable has a width clearly exceeding $[-1, 1]$, while that for the L3 observable remains approximately within this range.

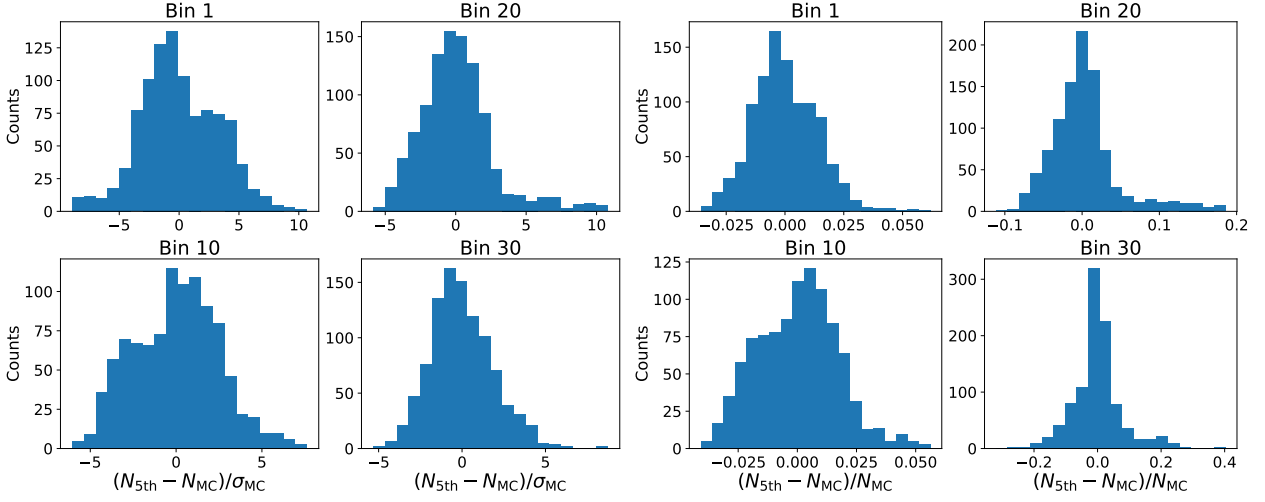


Figure 8.27: Pull distribution (left) and relative residuals (right) for the 1st, 10th, 20th, and 30th bin of the observable `d17-x01-y01` from the ATLAS analysis.

The pull and relative residual distributions therefore confirm the previous observation that the LEP observables are approximated sufficiently well by the surrogate model, whereas the ATLAS and fixed-target observables are described much less accurately. This difference arises from the different mapping of the MCEG parameters to the simulated observables. While the LEP observables are modeled using clean $e^-e^+ \rightarrow \text{hadrons}$ processes, without any beam remnants and MPI, the LHC and fixed-target observables involve considerably more soft-QCD activity. As a result, the LEP observables correspond to much smoother functions in parameter space and are therefore substantially easier to parameterize.

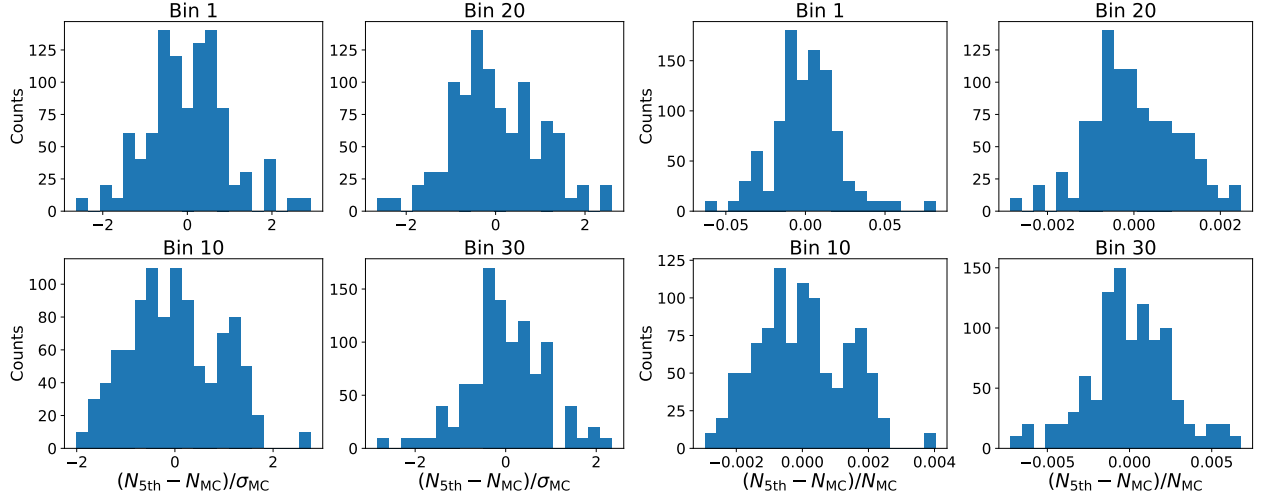


Figure 8.28: Pull distribution (left) and relative residuals (right) for the 1st, 10th, 20th, and 30th bin of the observable d65-x01-y02 from the L3 analysis.

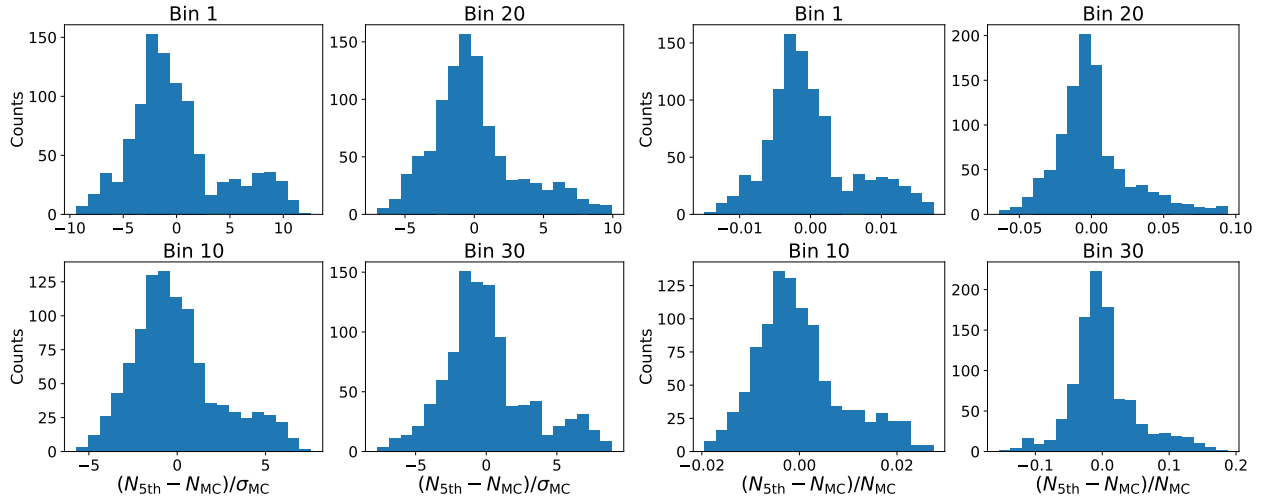


Figure 8.29: Pull distribution (left) and relative residuals (right) for the 1st, 10th, 20th, and 30th bin of the observable d06-x01-y01 from the EHS/NA22 analysis.

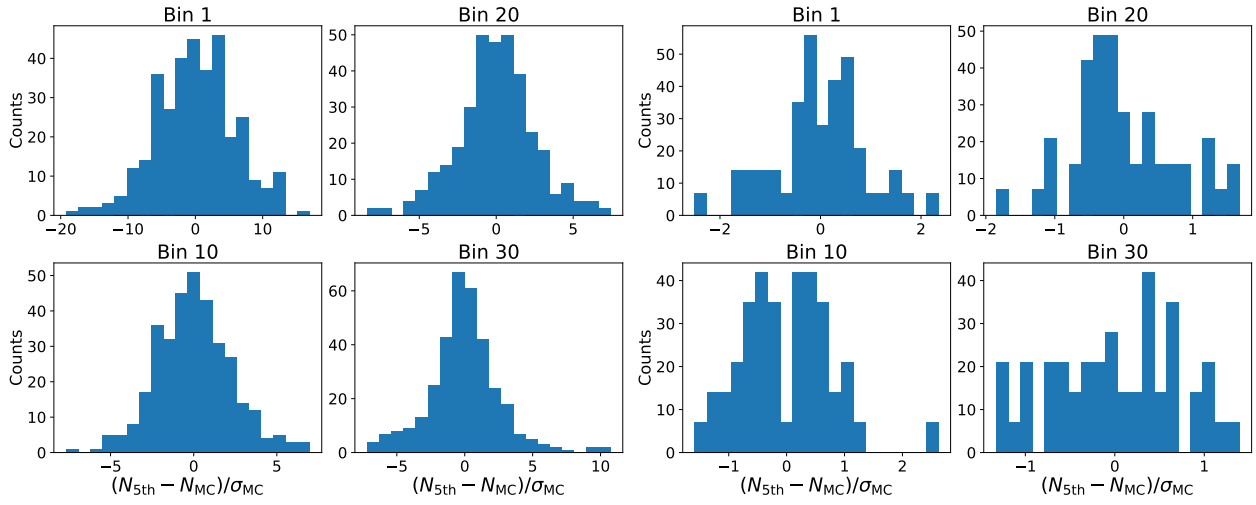


Figure 8.30: Pull distribution for the 1st, 10th, 20th, and 30th bin of the observable `d17-x01-y01` (left) from the ATLAS analysis and the observable `d65-x01-y02` (right) from the L3 analysis, using a surrogate model constructed from MC samples simulated on a 7^3 grid.

Another important aspect in the construction of a surrogate model is the relative size of the MC uncertainty compared to the data uncertainty. It is essential that the MC uncertainty remains significantly smaller, since the surrogate is trained directly on the MC output. Otherwise, the parameterization may not only approximate the physical dependence of the observables on the MCEG parameters, but also statistical fluctuations resulting from limited MC statistics. This can reduce both the smoothness and the accuracy of the surrogate response. Ensuring that the MC uncertainties are subdominant therefore helps the surrogate to represent the true generator behavior and ensures that the comparison to data is driven primarily by the experimental precision rather than by statistical noise in the simulated samples.

To investigate the relative size of MC and data uncertainties, the ratio of the two uncertainties is calculated. For each bin of each observable, the MC sample whose bin value is closest to the data is used, with the comparison performed as a function of the `sigma` parameter, while `aLund` and `pT0Ref` are fixed to selected values. This is shown in Figure 8.31 for the ATLAS observable `d17-x01-y01`, using MC samples simulated on a 7^3 grid with 10^6 and 10^7 events, respectively. Bins in which the data lie outside the band defined by the lowest and highest MC bin value are excluded.

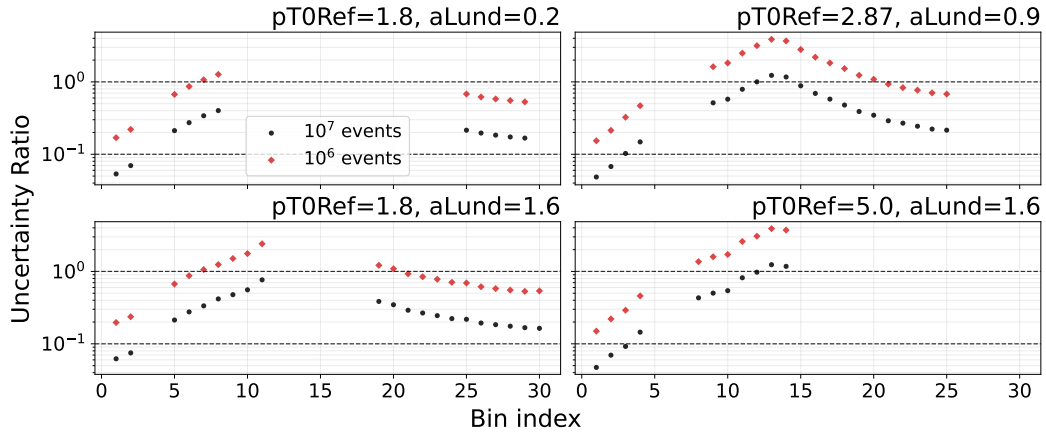


Figure 8.31: Ratio of the MC uncertainty to the data uncertainty for the observable `d17-x01-y01` from the ATLAS analysis, using in each bin the MC sample closest to the data. Bins in which the data lie outside the band defined by the highest and lowest MC bin values are excluded. The MC samples are simulated on a 7^3 grid with 10^7 events (black) and 10^6 events (red). The distribution is shown for four different combinations of `aLund` and `pT0Ref`.

For the ATLAS observable, the data uncertainties are very small, which results in a ratio of MC uncertainty to data uncertainty around 1 for many bins. Only for samples with 10^7 events does the ratio in the first few bins fall below 10^{-1} . As

already seen in the sensitivity studies, several bins are not described sufficiently well by `Pythia 8` for the corresponding data points to lie within the variation induced by the chosen parameter ranges.

The uncertainty ratio for the L3 observable `d59-x01-y02` is shown in Figure 8.32. In contrast to the ATLAS observable, the ratio is below 1 for almost all bins. For the samples consisting of 10^7 events, the ratio is even below 10^{-1} in most bins.

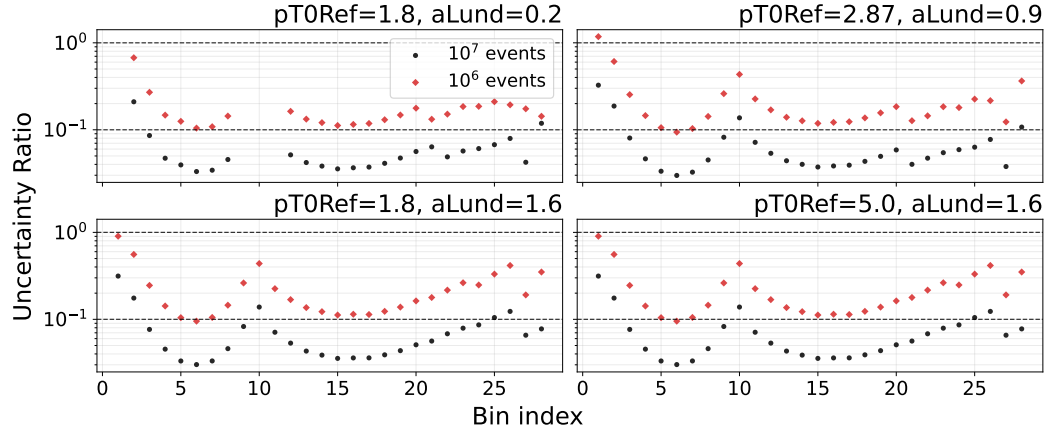


Figure 8.32: Ratio of the MC uncertainty to the data uncertainty for the observable `d59-x01-y02` from the L3 analysis, using in each bin the MC sample closest to the data. Bins in which the data lie outside the band defined by the highest and lowest MC bin values are excluded. The MC samples are simulated on a 7^3 grid with 10^7 events (black) and 10^6 events (red). The distribution is shown for four different combinations of `aLund` and `pT0Ref`.

Figure 8.33 shows the uncertainty ratio for the fixed-target observable. Similar to the LEP observable, the ratio remains small, taking values close to 10^{-1} for the samples with 10^6 events and even smaller values for the samples with higher statistics.

The study of the uncertainty ratios shows that, for the LEP and fixed-target observables, the experimental uncertainties are sufficiently large to justify the use of MC samples with 10^6 events. For the ATLAS observables, however, one may argue that the experimental uncertainties are so small that a larger number of simulated events is required in order to reduce the MC uncertainty further.

Based on the reduced chi-square distributions for the ATLAS observables shown in Figure 8.26, together with the previously discussed pull distributions and relative residuals, the ATLAS observables appear to be among the more challenging cases for the surrogate model. Their description is less accurate than for several other

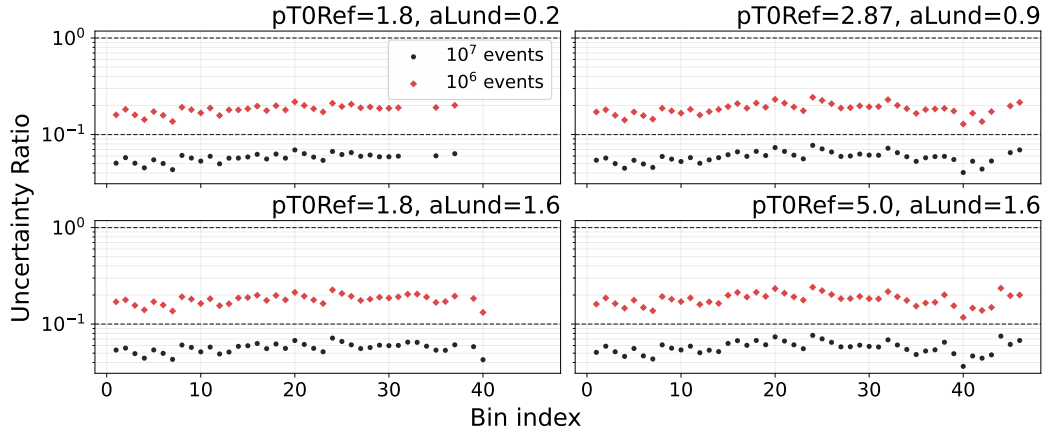


Figure 8.33: Ratio of the MC uncertainty to the data uncertainty for the observable `d06-x01-y01` from the EHS/NA22 analysis, using in each bin the MC sample closest to the data. Bins in which the data lie outside the band defined by the highest and lowest MC bin values are excluded. The MC samples are simulated on a 7^3 grid with 10^7 events (black) and 10^6 events (red). The distribution is shown for four different combinations of `aLund` and `pT0Ref`.

observables, suggesting that higher MC statistics may be required for an efficient tuning procedure. This may partly explain why the LHC and LHC+LEP tunes struggle to improve the description of the ATLAS observables, as shown in Figure 8.21. This limitation is also visible in Table 8.5, where `d17-x01-y01` shows large reduced chi-square values for the default model as well as for the LHC and LHC+LEP tunes. These large values are likely enhanced by the small experimental uncertainties of the data compared to the MC uncertainty

8.9 Influencing Air-Shower Simulations

In the following, the impact of the tuned versions of `Pythia 8`, namely the fixed-target tune and the tune combining the LHC and LEP observables (LHC+LEP), on the simulation of extensive air showers within the `CORSIKA 8` framework is investigated. For this purpose, 10^3 air showers initiated by either a proton or an iron primary are simulated at a primary energy of 10^{17} eV and a zenith angle of 30° . Further used `CORSIKA 8` settings include energy thresholds for particle tracking at 10 MeV for photons, electrons, and positrons, and 300 MeV for muons and hadrons.

8.9.1 Proton-Induced Air Showers

Figure 8.34 shows the average longitudinal profiles of positively and negatively charged muons in proton-induced air showers for the default configuration, the fixed-target tune, and the LHC+LEP tune. The profiles obtained with the default settings and the LHC+LEP tune are nearly identical, whereas the fixed-target tune exhibits a clear deficit over the full shower development.

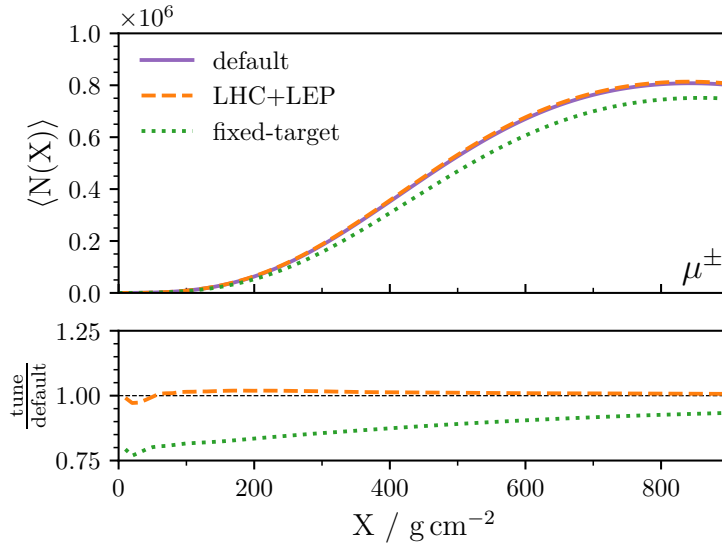


Figure 8.34: Average longitudinal shower profile of muons in proton-induced air showers.

A similar pattern is visible in the median lateral distribution and the median energy spectrum of muons in proton-induced air showers, shown in Figure 8.35. Once again, the LHC+LEP tune yields muon numbers that are very close to the default prediction, whereas the fixed-target tune leads to a visible deficit. In the lateral distribution, this deficit reaches approximately 20% at distances of about $5 \cdot 10^3$ m from the shower axis. Close to the shower core, however, the fixed-target tune produces a small increase in the muon number of about 5% relative to both the default and LHC+LEP configurations. In the median energy spectrum, the fixed-target tune predicts fewer muons than the default and LHC+LEP tunes over almost the full energy range, with a deficit of up to 10%.

As already discussed in Chapter 3, the depth of the shower maximum X_{max} is an important observable for improving the understanding of air-shower development. Figure 8.36 shows the average shower maxima for the all electromagnetic particles, $\langle X_{\text{max}} \rangle$, and for only the muonic component, $\langle X_{\mu, \text{max}} \rangle$. The default and LHC+LEP

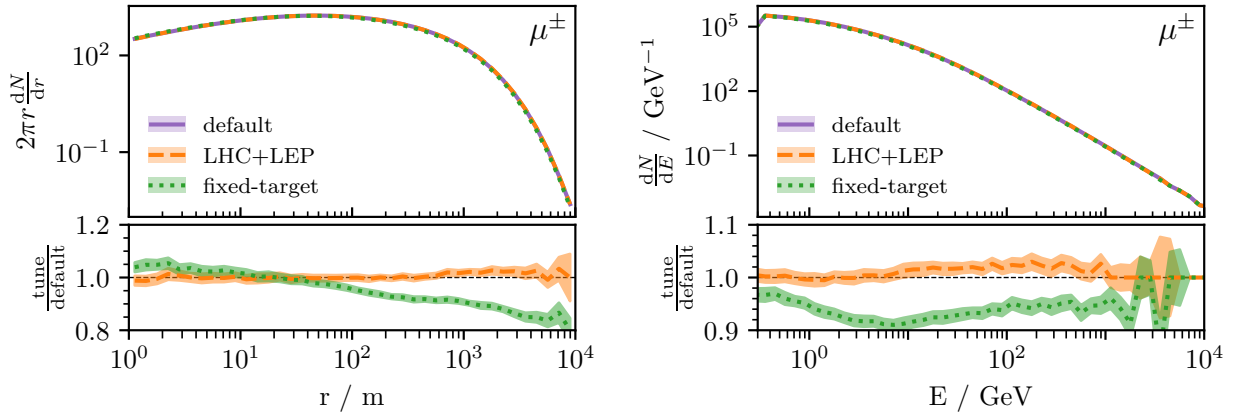


Figure 8.35: Median lateral (left) and median energy (right) distributions of muons produced in proton-induced air showers.

tunes differ only slightly, with the latter showing a decrease of about 1% in $\langle X_{\mu,\max} \rangle$. The fixed-target tune also leads to small deviations, but in this case both shower maxima increase, by about 2% for $\langle X_{\max} \rangle$ and about 3% for $\langle X_{\mu,\max} \rangle$.

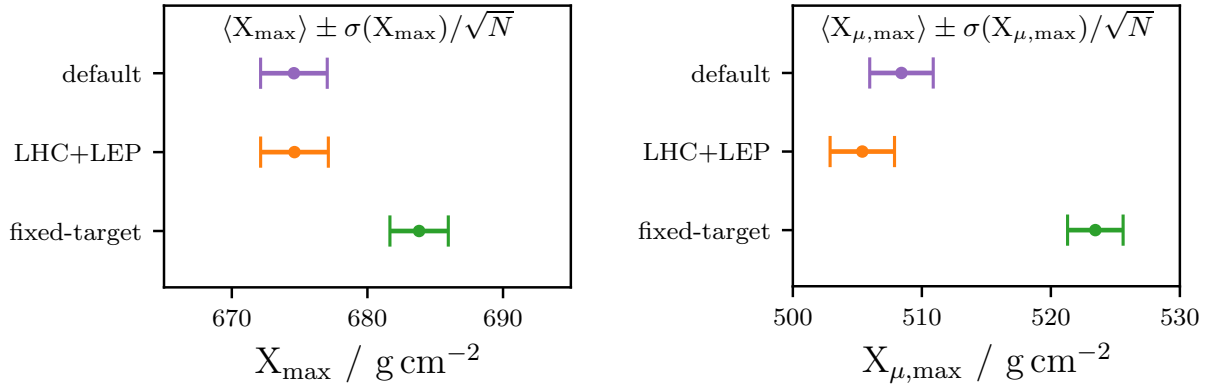


Figure 8.36: Average $X_{\max}$ and $X_{\mu,\max}$ for proton-induced air showers.

8.9.2 Iron-Induced Air-Showers

The same set of observables is considered for air showers induced by an iron primary.

Figure 8.37 shows the average longitudinal profile of positively and negatively charged muons. As in the proton-induced case, the LHC+LEP prediction remains close to the default, whereas the fixed-target tune again leads to a visible deficit.

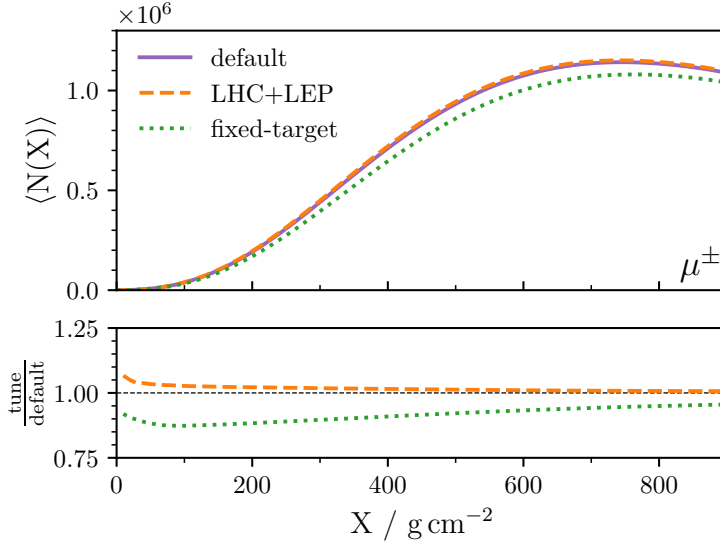


Figure 8.37: Average longitudinal shower profile of muons in iron-induced air showers simulated with CORSIKA 8.

A similar overall behavior, but with a smaller magnitude than in the proton-induced case, is visible in the median lateral and median energy distributions shown in Figure 8.38. While the proton-induced showers show a deficit of up to about 20% in the lateral distribution, the corresponding deficit in the iron-induced case is only about 10%. The same tendency is observed in the energy distributions, where the iron-induced air showers exhibit a similar shape but a noticeably smaller reduction in the muon number.

The depth of the shower maximum for iron-induced air showers is shown in Figure 8.39. As expected, both $\langle X_{\max} \rangle$ and $\langle X_{\mu, \max} \rangle$ assume significantly lower values than in the proton-induced case. This follows from the fact that iron primaries interact earlier in the atmosphere and can, to some approximation, be viewed within the superposition picture as a collection of lower-energy nucleons. Since lower-energy showers reach their maximum earlier, this leads to smaller values of the shower-maximum depth for iron-induced air showers. The same superposition argument also explains the reduced spread in the average shower maximum, as the primary effectively behaves like 56 nucleons with an energy of approximately $E_{\text{primary}}/56$, which reduces fluctuations in the shower development. The impact of the different tunes on the shower maxima is comparable to that observed in the

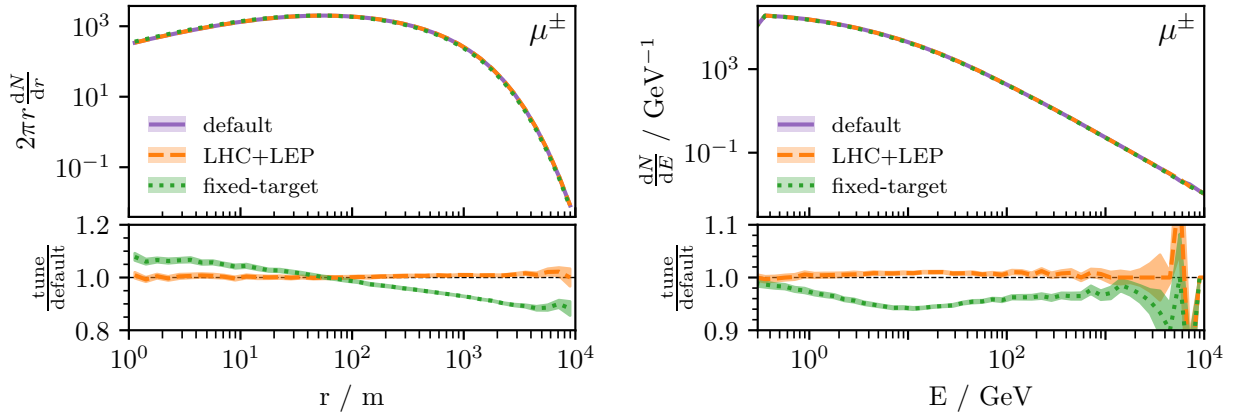


Figure 8.38: Median lateral (left) and median energy (right) distributions of muons produced in iron-induced air showers.

proton-induced case.

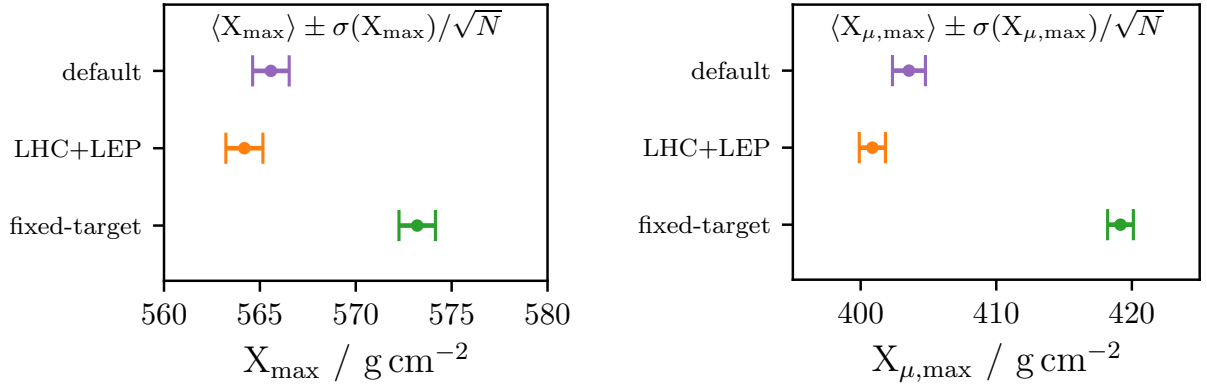


Figure 8.39: Average $X_{\max}$ and $X_{\mu,\max}$ for iron-induced air showers.

8.9.3 Impact on the Number of Muons

As already discussed in Chapter 3.4, the muon puzzle remains an open problem in the simulation of EAS, with current simulations predicting fewer high-energy muons than observed in measurements. According to Ref. [1], one possible route towards resolving this discrepancy is an improved description of hadronic interactions in MCEGs, achievable through constraints by measurements currently available at the

LHC. For this reason, the possibility of modifying the muon yield in EAS simulations through dedicated tunes of MCEGs is of particular interest.

Figure 8.40 shows the average number of muons $\langle N_\mu \rangle$ as a function of the average shower maximum $\langle X_{\max} \rangle$ for both proton- and iron-induced air showers. Since proton and iron represent the lightest and heaviest relevant primary particles, respectively, primaries with intermediate mass are expected to populate approximately the line connecting the two corresponding points.

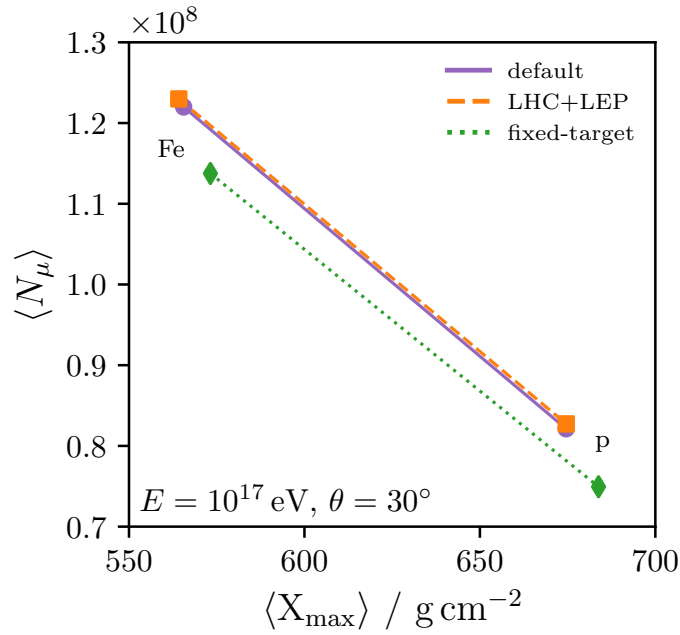


Figure 8.40: Average number of muons $\langle N_\mu \rangle$ as a function of the average shower maximum $\langle X_{\max} \rangle$ for proton- and iron-induced air showers.

The curves corresponding to the default configuration and the LHC+LEP tune yield very similar values of both $\langle N_\mu \rangle$ and $\langle X_{\max} \rangle$, indicating essentially no change in the muon number. By contrast, the fixed-target tune produces a deficit in $\langle N_\mu \rangle$ of about 7-9%, illustrating the potential sensitivity of the muon yield in air-shower simulations to dedicated tunes of the Pythia 8 MCEG.

9 Conclusion

In this thesis, an existing Bayesian tuning strategy for hadronic interaction models was adapted and extended with the aim of improving the description of cosmic-ray-induced air showers with `Pythia 8`. The central goal was to investigate how constraints from accelerator measurements propagate into air-shower simulations and whether such constraints can lead to significant changes in relevant shower observables. The results show that this approach is successful and that the presented Bayesian tuning workflow provides a useful framework for further studies in this area.

A toy-model study with air-shower simulations shows that the tuning workflow works in principle. At the same time, the more realistic application with `CORSIKA 8` revealed the present practical limitations. The high computational cost makes a direct tune to air-shower observables difficult, especially when different primary energies, mass compositions, and directions have to be covered. In addition, `Pythia 8` as a fully tunable hadronic event generator for air-shower applications is still under development. For these reasons, a direct tune to air-shower observables was not pursued further.

An indirect tuning strategy was therefore constructed using selected collider and fixed-target measurements. An important part of the analysis was the comparison of surrogate models used to parameterize the generator response. Two approaches were investigated: piecewise linear interpolation and a fifth-order multivariate polynomial parameterization. Both models described the sampled Monte Carlo points, but the polynomial parameterization was preferred because of its smoother behavior between sampled points. This study also showed that the quality of the surrogate model is determined not only by the interpolation method, but also by the Monte Carlo sample size used for the parameterization. The event statistics must therefore be chosen such that the Monte Carlo uncertainty is smaller than the experimental uncertainty. Otherwise, fluctuations from limited sample size are absorbed by the surrogate model and reduce the robustness of the tuning.

The posterior analysis showed that the selected LEP, LHC, and fixed-target observables favor different regions of parameter space, so that a single common tune is not adequate for the chosen observable set. Separate tunes were therefore constructed. These tunes improve the general description of most observables, with the largest

improvement for the fixed-target observable, which was modeled significantly better after tuning. This demonstrates that the tuning procedure is not only technically functional, but also effective in improving the model.

At the same time, the study showed that the main limitation lies in the parameterization of the generator response. In particular, the propagation of parameter uncertainties to the observables is not yet fully reliable, because the surrogate model does not reproduce the underlying Monte Carlo response with sufficient fidelity for all observables. This does not alter the central conclusion of the thesis: the workflow is established and performs as intended, the tuning leads to a clear improvement in the description of the data.

An important result of this thesis is that the tuned parameter sets lead to visible changes in air-shower simulations. While the LHC+LEP tune remains close to the default configuration, the fixed-target tune results in a deficit in the predicted average muon number of about 7–9%. Although this shift goes in the opposite direction of what would be required to reduce the muon puzzle, it demonstrates that accelerator-based constraints can induce measurable changes in extensive-air-shower predictions and can directly affect observables of central relevance to cosmic-ray physics.

In conclusion, this thesis has shown that Bayesian tuning can be successfully adapted to `Pythia 8` for cosmic-ray applications. The tuning improves the overall description of the selected measurements, especially for the fixed-target observable, reducing the reduced chi-square by more than 95%, from 89.80 for the default parameter set to 3.83 after tuning to the corresponding measurement. In addition, the results show that accelerator-based constraints can lead to meaningful changes in air-shower predictions. At the same time, the results show that the workflow is robust enough to be used in larger studies and provides a solid basis for future work on tuning hadronic interaction models for air-shower simulations.

10 Outlook

Building directly on the studies presented in this work, the improvement of cosmic-ray-induced air-shower simulations can be pursued further by extending the indirect tuning to a larger set of observables and free parameters. The present study has shown that measurements from LHC, LEP, and fixed-target experiments favor different regions of parameter space. A broader analysis could help clarify these differences and improve understanding of the origin of the current modeling tensions. This would not only support future tunes, but could also provide guidance for further development of the underlying physical model.

In this context, it will be especially important to systematically identify the model parameters and observables that have the largest impact on air-shower simulations. A more detailed sensitivity study could determine which accelerator measurements provide the strongest constraints on the parts of the model most relevant for shower development. Such an approach would help to define more targeted tuning strategies and to improve the connection between accelerator data and air-shower observables.

The study also showed that the surrogate model has a strong influence on the tuning result. It will therefore be important to investigate how well the generator response can be approximated in a higher-dimensional parameter space and whether the present parameterization strategies remain adequate in such a setting.

Further progress will also depend on the available analysis tools. Fixed-target measurements from LHC experiments with colliding particles heavier than protons would be highly valuable for these studies, but corresponding `Rivet` routines are not yet available. More generally, the strong reliance on `Rivet` also highlights a central difficulty for direct tuning to air-shower observables, since no comparable framework exists for cosmic-ray-induced air shower data in the same systematic way, as discussed in Ref. [35]. The development of such a tool, together with further improvements to air-shower simulations in `CORSIKA 8` using `Pythia 8`, will be an essential prerequisite for broader direct-tuning studies based on air-shower data.

A Appendix

A.1 Parameter Sensitivities

This section presents the remaining sensitivity plots for the candidate observable–parameter combinations omitted from Chapter 8.

A Appendix

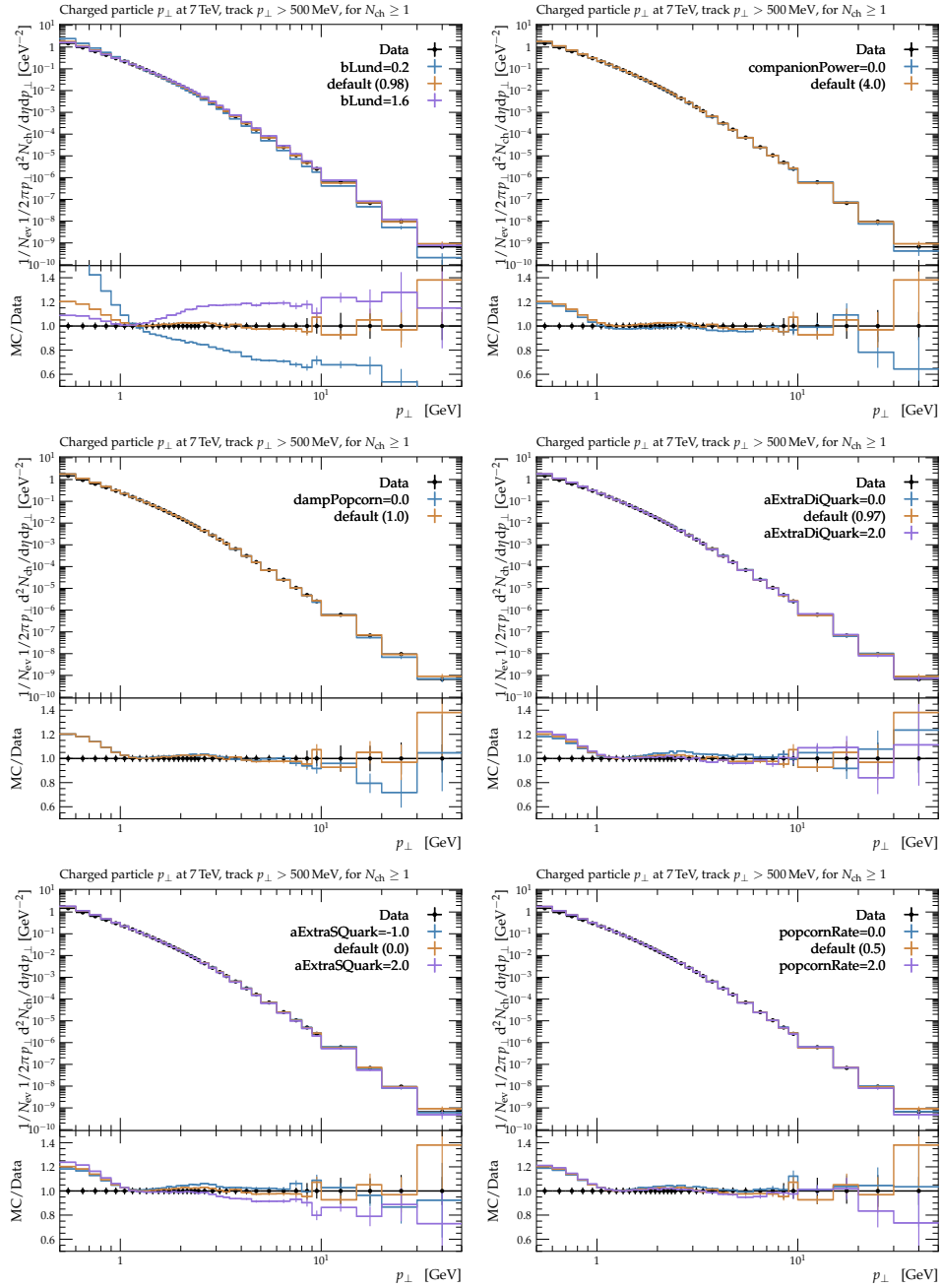


Figure A.1: Charged-particle multiplicity as a function of transverse momentum for different parameter variations. The observable is taken from the Rivet analysis ATLAS_2010_I882098, histogram d10-x01-y01.

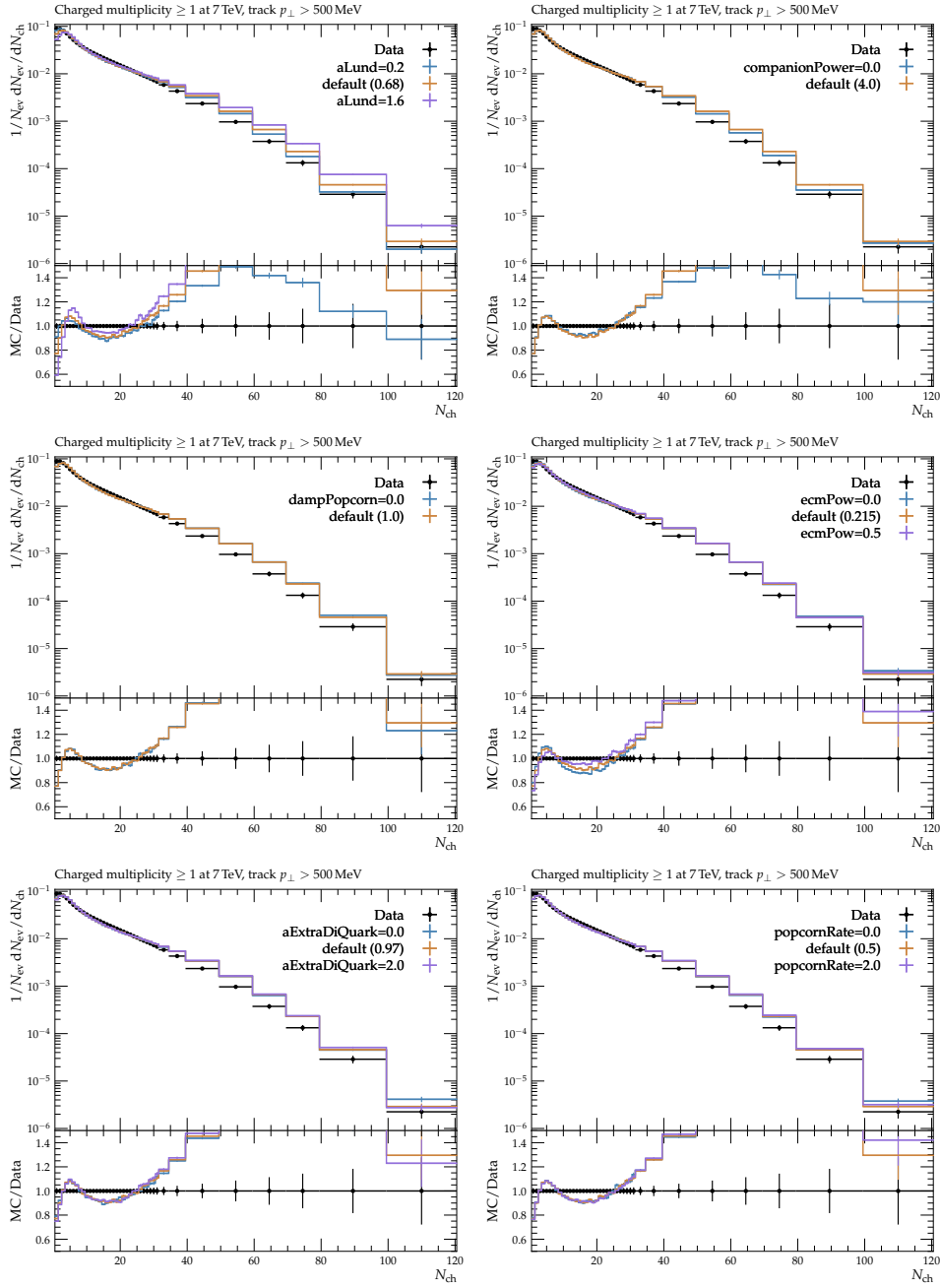


Figure A.2: Charged-particle multiplicity distribution for different parameter variations. The observable is taken from the Rivet analysis ATLAS_2010_I882098, histogram d17-x01-y01.

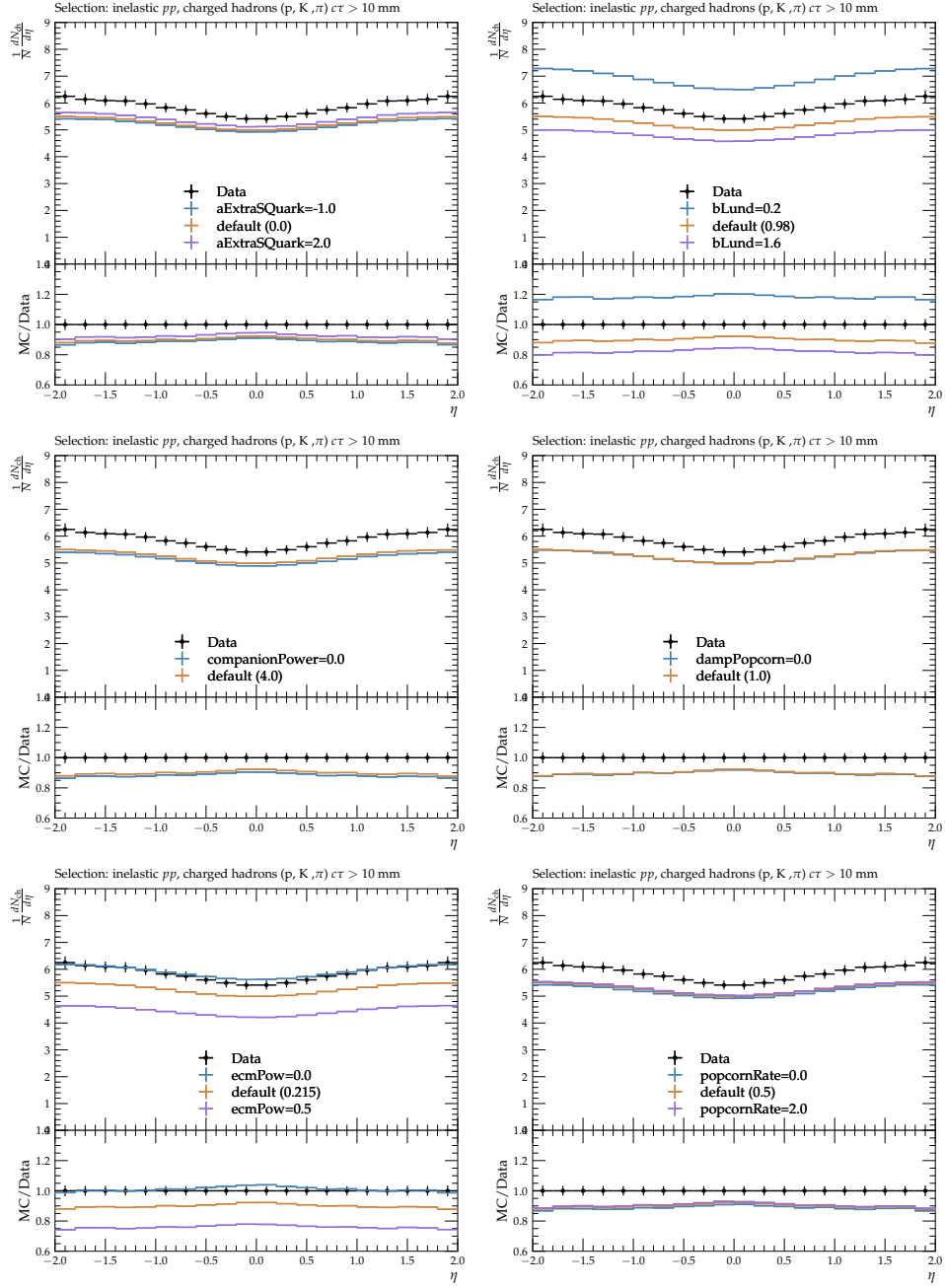


Figure A.3: Pseudorapidity density of charged hadrons for different parameter variations. The observable is taken from the Rivet analysis CMS_2015_I1384119, histogram d01-x01-y01.

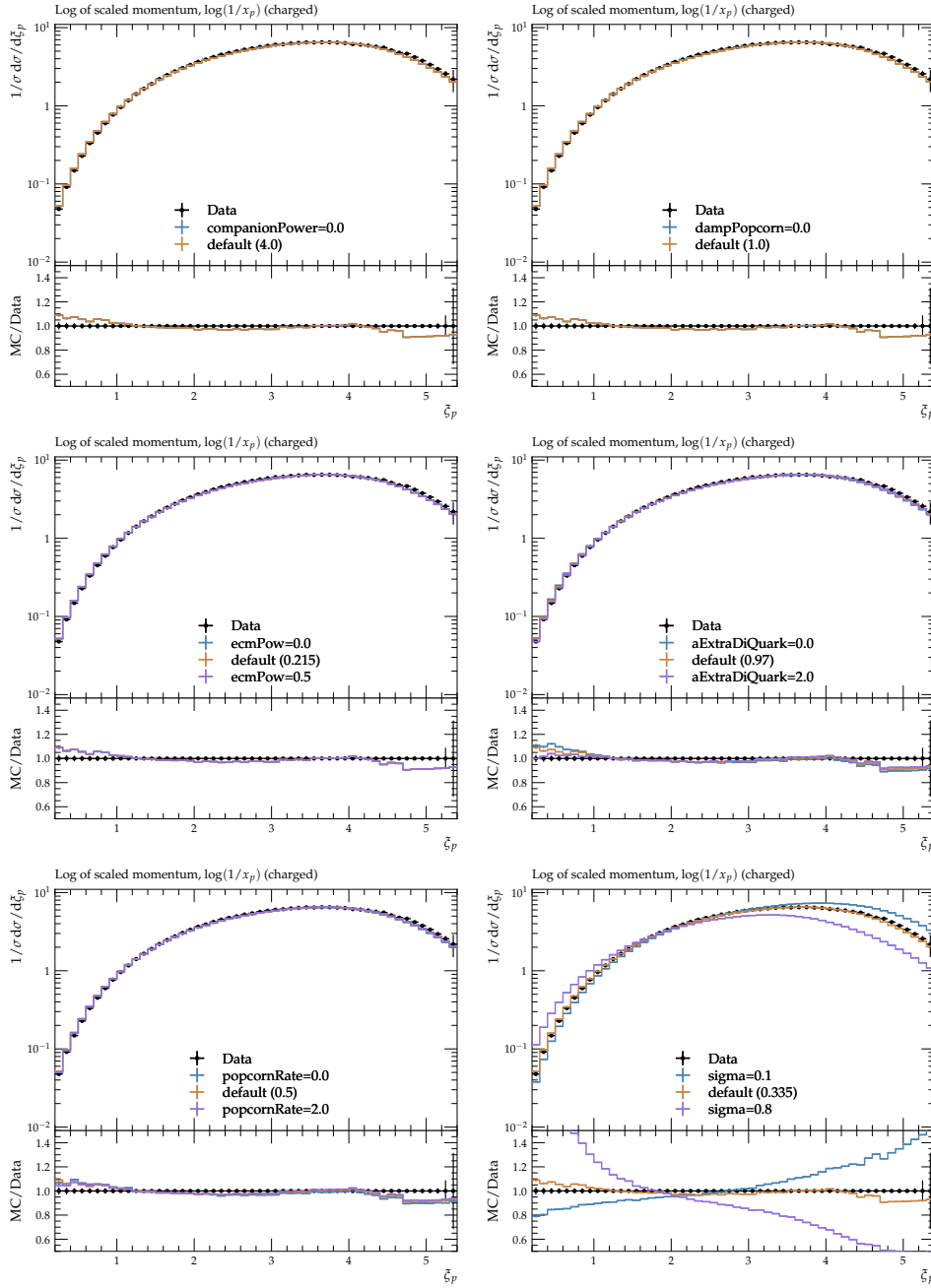


Figure A.4: Distribution of the logarithm of the scaled momentum, ξ_p , for different parameter variations. The observable is taken from the Rivet analysis ALEPH_1996_I428072, histogram d17-x01-y01.

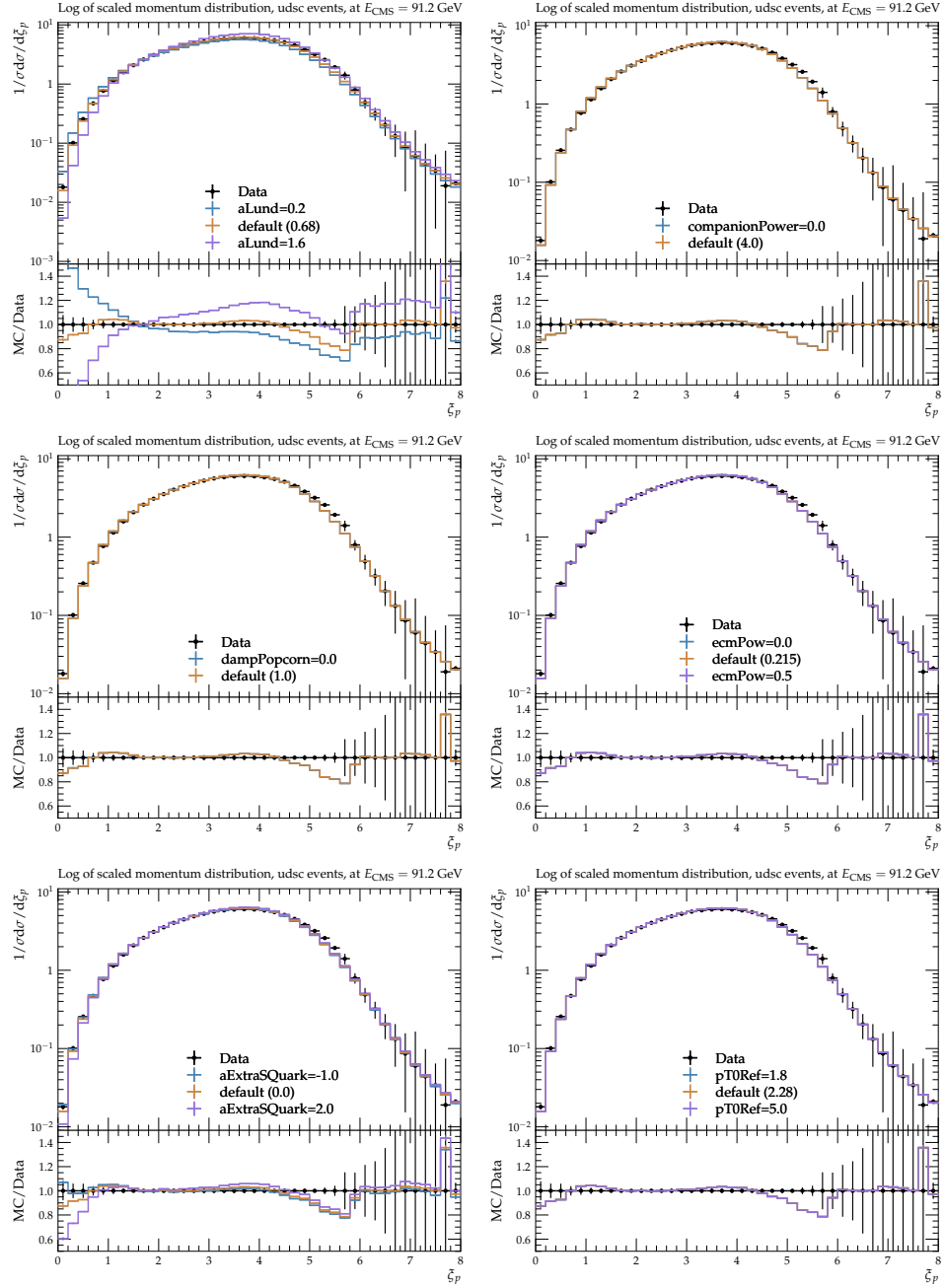


Figure A.5: Distribution of the logarithm of the scaled momentum, ξ_p , for different parameter variations. The observable is taken from the Rivet analysis L3_2004_I652683, histogram d65-x01-y02.

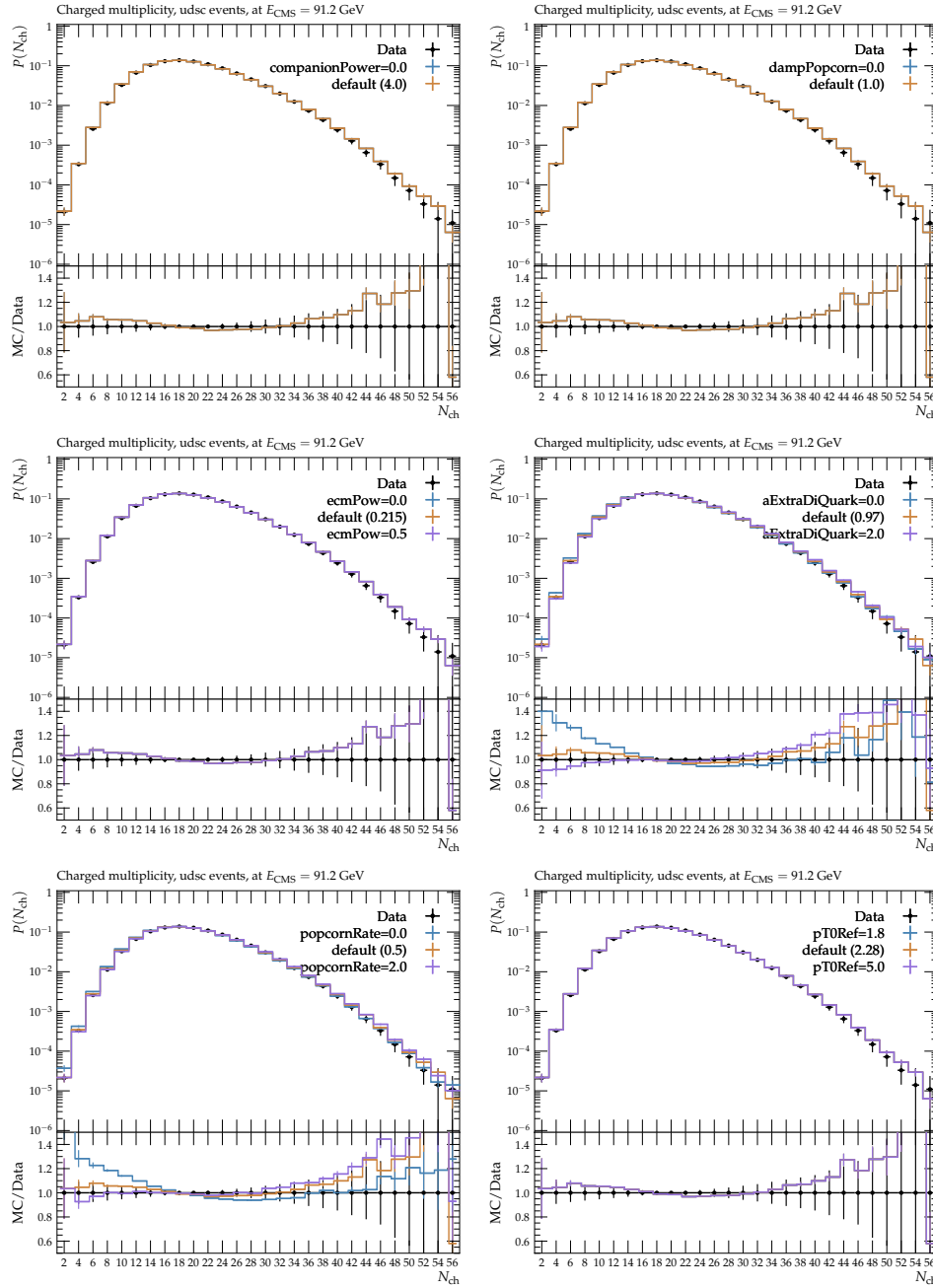


Figure A.6: Charged-particle multiplicity distribution for different parameter variations. The observable is taken from the Rivet analysis L3_2004_I652683, histogram d59-x01-y02.

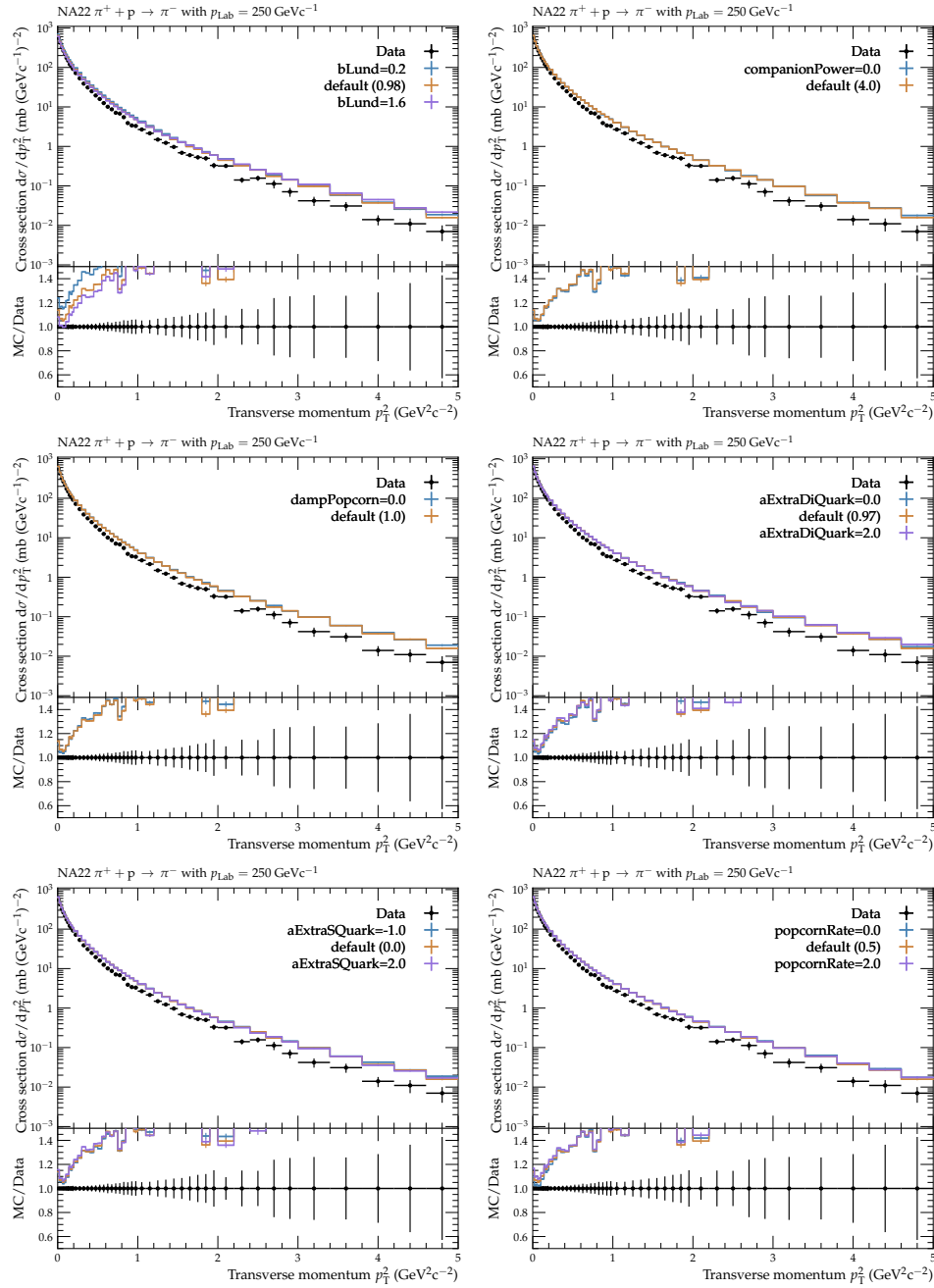


Figure A.7: Transverse-momentum distribution of negatively charged pions produced in collisions of a pion beam with a fixed hydrogen target, shown for different parameter variations. The observable is taken from the Rivet analysis EHS_1988_I265504, histogram d06-x01-y01.

A.2 Posterior Distributions

This section presents the remaining posterior plots for the LHC and LEP tune omitted from Chapter 8.5.

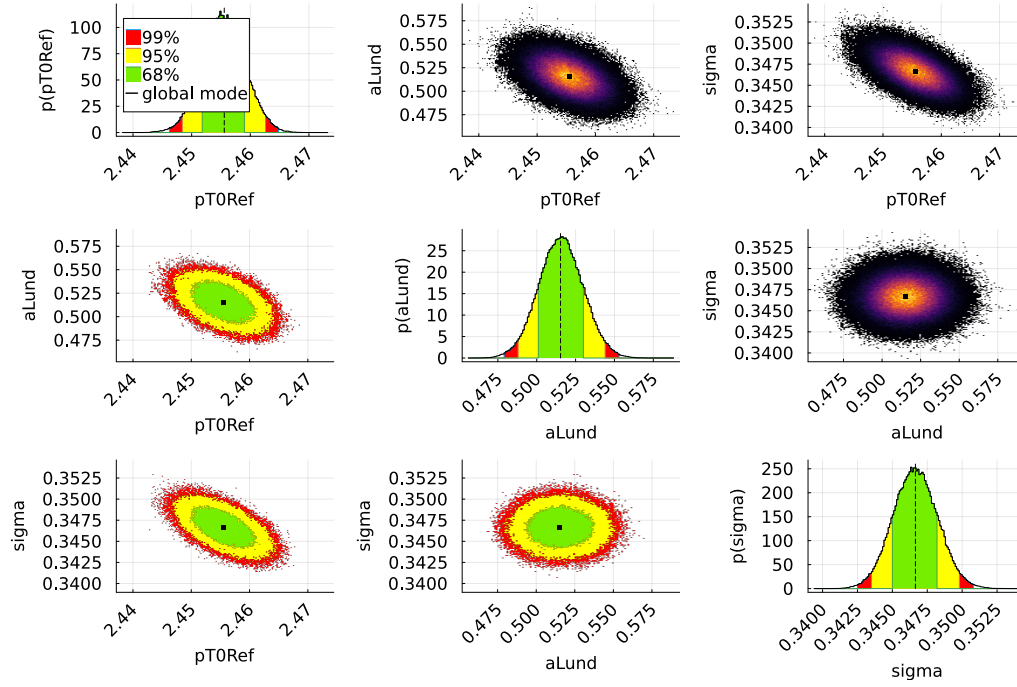


Figure A.8: One- and two-dimensional marginalized posterior distributions of the parameters $pT0Ref$, $aLund$, and σ for the LHC tune. The marginalized distributions are shown for the smallest 68%, 95%, and 99% intervals, while the heat maps of the two-dimensional marginalized posterior distributions are shown in the upper-right triangle.

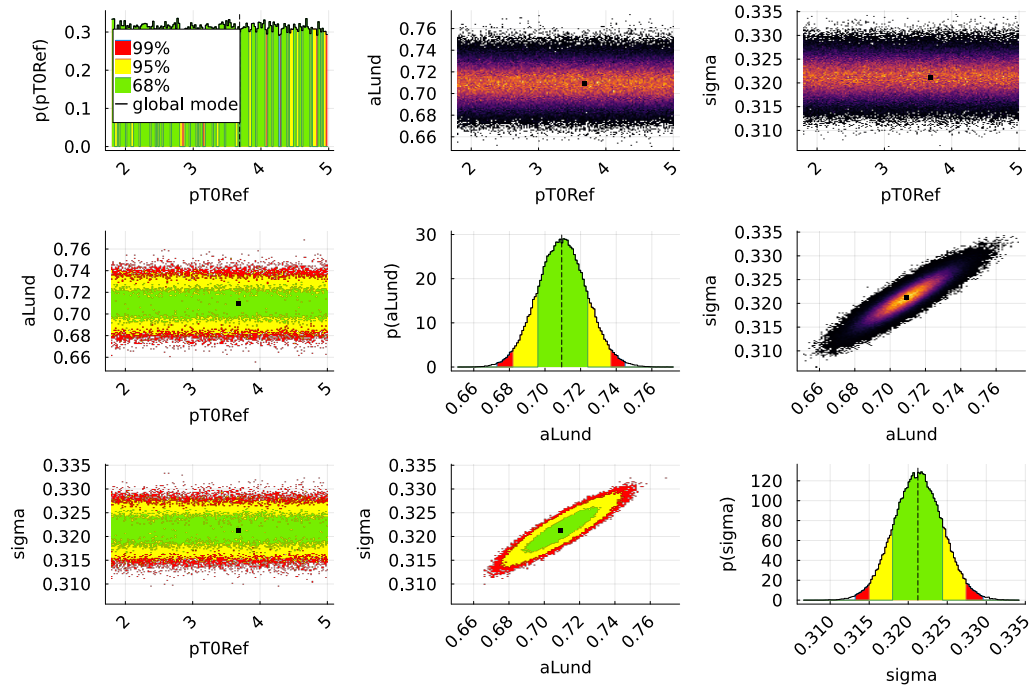


Figure A.9: One- and two-dimensional marginalized posterior distributions of the parameters $pT0Ref$, $aLund$, and σ for the LEP tune. The marginalized distributions are shown for the smallest 68%, 95%, and 99% intervals, while the heat maps of the two-dimensional marginalized posterior distributions are shown in the upper-right triangle.

Bibliography

- [1] J. Albrecht et al., *The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider*, *Astrophys. Space Sci.* **367** (2022), Licensed under CC BY 4.0.
- [2] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codebases* (2022) 8, Licensed under CC BY 4.0.
- [3] C. Bierlich et al., *Robust Independent Validation of Experiment and Theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026.
- [4] R. Engel et al., *Towards A Next Generation of CORSIKA: A Framework for the Simulation of Particle Cascades in Astroparticle Physics*, *Comput. Softw. Big Sci.* **3** (2018).
- [5] J.M. Alameddine et al., *Simulating radio emission from particle cascades with CORSIKA 8*, *Astroparticle Physics* **166** (2025) 103072.
- [6] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264.
- [7] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968) 367.
- [8] S. Weinberg, *Non-Abelian Gauge Theories of the Strong Interactions*, *Phys. Rev. Lett.* **31** (1973) 494.
- [9] S. L. Glashow, *Partial-symmetries of weak interactions*, *Nucl. Phys.* **22** (1961) 579.
- [10] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, *Nucl. Phys. B* **44** (1972) 189.
- [11] D. Galbraith and C. Burgard, *Standard model of physics*, URL: <https://texample.net/model-physics/> (visited on 09/24/2025), Licensed under CC BY-SA 4.0.
- [12] S. La Cagnina, *Development of a tool for Bayesian data analysis and its application in Monte Carlo tuning*, PhD thesis: TU Dortmund, 2023.
- [13] Particle Data Group et al., *Review of Particle Physics*, *Prog. Theor. Exp. Phys.* **2022** (2022) 083C01.

- [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1.
- [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30.
- [16] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.
- [17] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.
- [18] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.
- [19] S. Bailey et al., *Parton distributions from LHC, HERA, Tevatron and fixed target data*, Eur. Phys. J. C. **81** (2021), Licensed under CC BY-SA 4.0.
- [20] D. J. Bird et al., *Detection of a Cosmic Ray with Measured Energy Well beyond the Expected Spectral Cutoff due to Cosmic Microwave Radiation*, Astrophys. J. **441** (1995) 144.
- [21] M. Krause et al., *High energy astroparticle physics for high school students*, PoS(ICRC2015) **236** (2016) 304.
- [22] S. Lafebre, *From cosmic particle to radio pulse*, PhD thesis: Radboud University Nijmegen, 2008.
- [23] The Pierre Auger Collaboration, *The Pierre Auger Cosmic Ray Observatory*, Nucl. Instrum. Methods Phys. Res., Sect. A **798** (2015) 172.
- [24] T. Abu-Zayyad et al., *The surface detector array of the Telescope Array experiment*, Nucl. Instrum. Methods Phys. Res., Sect. A **689** (2012) 87.
- [25] Pierre Auger Observatory, *A Hybrid Detector*, URL: <https://www.auger.org/observatory/auger-hybrid-detector> (visited on 10/15/2025), Licensed under CC BY-SA 4.0.
- [26] G. A. Askar'yan, *Excess Negative Charge of an Electron-Photon Shower and its Coherent Radio Emission*, Zh. Eksp. Teor. Fiz. **41** (1961) 616.
- [27] F. G. Schröder, *Radio detection of ultrahigh-energy cosmic-ray air showers*, Eur. Phys. J. Spec. Top. **234** (2025) 4957.
- [28] R. U. Abbasi et al., *Study of muons from ultrahigh energy cosmic ray air showers measured with the Telescope Array experiment*, Phys. Rev. D **98** (2018) 022002.
- [29] A. Aab et al., *Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory*, Phys. Rev. Lett. **117** (2016) 192001.

-
- [30] H. Dembinski et al., *Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV*, PoS(ICRC2017) (2017) 533.
 - [31] A. Buckley et al., *General-purpose event generators for LHC physics*, Phys. Rep. **504** (2011) 145.
 - [32] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016).
 - [33] R. Pasechnik and M. Šumbera, *Different Faces of Confinement*, Universe **7** (2021) 330, Licensed under CC BY 4.0.
 - [34] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune*, Eur. Phys. J. C **74** (2014), Licensed under CC BY 4.0.
 - [35] J. Albrecht et al., *Global tuning of hadronic interaction models with accelerator-based and astroparticle data*, Nat. Rev. Phys. **8** (2025) 98.
 - [36] J.M. Alameddine et al., *Simulating radio emission from particle cascades with CORSIKA 8*, Astropart. Phys. **166** (2025) 103072.
 - [37] Corsika Collaboration, *CORSIKA 8 Framework for Particle Cascades in Astroparticle Physics*, 2018, URL: <https://gitlab.iap.kit.edu/AirShowerPhysics/corsika> (visited on 01/27/2026).
 - [38] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, Technical Report FZKA-6019 (1998).
 - [39] M. Reininghaus, *Air showers and hadronic interactions with CORSIKA 8*, SciPost Phys. Proc. (2024) 019.
 - [40] A. Fedynitch et al., *Hadronic interaction model SIBYLL 2.3c and inclusive lepton fluxes*, Phys. Rev. D **100** (2019) 103018, Licensed under CC BY 4.0.
 - [41] S. Ostapchenko, *Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model*, Phys. Rev. D **83** (2011) 014018.
 - [42] T. Pierog et al., *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. C **92** (2015) 034906.
 - [43] T. Sjöstrand and M. Uthm, *Hadron interactions for arbitrary energies and species, with applications to cosmic rays*, Eur. Phys. J. C **82** (2022).
 - [44] O. Behnke et al., eds., *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, 1st, WILEY-VCH, 2013.
 - [45] C. Robert and G. Casella, eds., *Monte Carlo Statistical Methods*, Springer New York, 2013.
 - [46] S. La Cagnina et al., *A Bayesian tune of the Herwig Monte Carlo event generator*, JINST **18** (2023) P10033.

- [47] A. Buckley et al., *Systematic event generator tuning for the LHC*, Eur. Phys. J. C **65** (2009) 331.
- [48] M. Krishnamoorthy et al., *Apprentice for Event Generator Tuning*, EPJ Web Conf. **251** (2021) 03060, ed. by C. Biscarat et al.
- [49] Pythia, *RIVET usage*, URL: <https://pythia.org/latest-manual/RIVETusage.html> (visited on 04/10/2026).
- [50] O. Schulz et al., *BAT.jl: A Julia-Based Tool for Bayesian Inference*, SN Comput. Sci. **2** (2021) 210.
- [51] N. Castro et al., *EFTfitter—A tool for interpreting measurements in the context of effective field theories*, Eur. Phys. J. C **76** (2016) 432.
- [52] S. P. Brooks and A. Gelman, *General Methods for Monitoring Convergence of Iterative Simulations*, J. Comput. Graph. Stat. **7** (1998) 434.
- [53] T. Sjöstrand, *The Pythia event generator: Past, present and future*, Comput. Phys. Commun. **246** (2020) 106910.
- [54] T. Sjöstrand and M. Utheim, *Hadron interactions for arbitrary energies and species, with applications to cosmic rays*, Eur. Phys. J. C **82** (2022).
- [55] T. Fuchs, *Charmante Myonen im Eis*, PhD thesis: TU Dortmund, 2016.
- [56] T. K. Gaisser, *Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio*, Astropart. Phys. **35** (2012) 801.
- [57] A. Fedynitch et al., *Calculation of conventional and prompt lepton fluxes at very high energy*, EPJ Web Conf. **99** (2015) 08001, ed. by D. Berge et al.
- [58] M. Fieg et al., *Tuning pythia for forward physics experiments*, Phys. Rev. D **109** (2024) 016010.
- [59] I. Neutelings, *Pseudorapidity*, Licensed under CC BY-SA 4.0, URL: https://tikz.net/axis2d_pseudorapidity/ (visited on 02/17/2026).
- [60] CMS Collaboration, *CMS pythia 8 colour reconnection tunes based on underlying-event data*, Eur. Phys. J. C **83** (2023).
- [61] M. Adamus et al., *Charged Particle Production in K^+p , π^+p and pp Interactions at 250-GeV/c*, Z. Phys. C **39** (1988) 311.
- [62] C. Gaudu, *Implementation of Pythia 8/Angantyr in CORSIKA 8 and Its Tuning for Air Shower Physics*, PhD thesis: Bergische Universität Wuppertal, 2026 (to be published).
- [63] G. Aad et al., *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, New J. Phys. **13** (2011) 053033.
- [64] P. Achard et al., *Studies of hadronic event structure in $e+e^-$ annihilation from 30 to 209 GeV with the L3 detector*, Phys. Rep. **399** (2004) 71.

- [65] R. Barate et al., *Studies of quantum chromodynamics with the ALEPH detector*, Phys. Rept. **294** (1998) 1.
- [66] T. Sjöstrand, *Welcome to PYTHIA - The Lund Monte Carlo!*, 2026, URL: <https://pythia.org/latest-manual/Welcome.html> (visited on 02/18/2026).
- [67] *Linear Algebra*, 2026, URL: <https://docs.julialang.org/en/v1/stdlib/LinearAlgebra/#LinearAlgebra.factorize> (visited on 03/02/2026).
- [68] OpenAI, *ChatGPT*, 2026, URL: <https://chatgpt.com/> (visited on 05/06/2026).
- [69] Grammarly, *AI Writing Assistant*, 2026, URL: <https://www.grammarly.com/ai-writing-assistant> (visited on 05/06/2026).

Declaration on the Use of Generative AI

During the preparation of this thesis, AI-assisted writing tools were used to support language revision, grammar checking, wording, and readability. In particular, ChatGPT [68] and Grammarly [69] were used for these purposes. All suggestions generated by these tools were critically reviewed and edited by the author. The author takes full responsibility for the final content of this thesis.

Acknowledgements

Diese Arbeit wäre nicht ohne die Unterstützung einiger Menschen möglich gewesen, bei denen ich mich im Folgenden bedanken möchte.

Zunächst möchte ich mich bei Prof. Dr. Kevin Kröninger dafür bedanken, dass ich meine Dissertation in dieser Arbeitsgruppe anfertigen durfte. Zudem bin ich zutiefst dankbar für die Betreuung und Zusammenarbeit über all die Jahre hinweg. Mein zweiter Dank gilt meinem Zweitgutachter Prof. Dr. Karl-Heinz Kampert für seine stetige Unterstützung während dieser Arbeit. Außerdem möchte ich mich bei Chloé Gaudu und Felix Riehn für die vielen hilfreichen Diskussionen bedanken.

Als Nächstes möchte ich mich bei der Arbeitsgruppe bedanken. Danke an Andrea Teichmann und Mike Muschak, dass ihr immer Ansprechpartner seid, für jedes Problem eine Lösung habt und den Laden zusammenhaltet. Ein Dank geht auch an die Admins Christopher und Lucas, die mir stets geholfen und die Infrastruktur aufrechterhalten haben. Danke, dass ich unser Cluster so häufig mit Simulationen stopfen durfte.

Für das Korrekturlesen meiner Arbeit möchte ich mich bei Chris, Donna, Felix, Andrea, Cornelius, Lucas und besonders bei Salvatore bedanken, die mir hilfreiche Anmerkungen zur Fertigstellung dieser Dissertation gegeben haben.

Auch wenn meine Zeit in der Arbeitsgruppe endet, werde ich stets Kontakt zu den vielen Freunden halten, die ich in dieser Gruppe kennenlernen durfte. Ich hoffe, wir werden noch viele Jahre lang Padel und Magic spielen, sowie aufregende Abenteuer erleben.

Besonders möchte ich mich bei meiner Lebensgefährtin Paulin bedanken. Danke für die wundervollen Jahre voller Freude und emotionaler Unterstützung während dieser Arbeit. Du hast mir stets dabei geholfen, positiv zu bleiben, egal wie stressig es manchmal wurde. Du hast mich während eines Großteils dieser Arbeit begleitet, und ich kann nicht in Worte fassen, wie sehr du mir in dieser Zeit geholfen hast.

Mein letzter Dank geht an meine Familie, die mich stets auf meinem Weg begleitet und unterstützt hat. Danke, dass ihr mich in jeder Hinsicht unterstützt und mir dadurch diesen Weg ermöglicht habt. Ohne euch wäre dies nie möglich gewesen, und ihr seid genauso mitverantwortlich für die Vollendung dieses Abenteuers.

