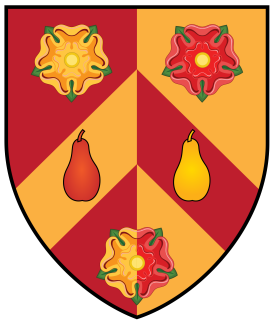
SEARCH FOR QCD INSTANTONS AT THE LHC USING THE ATLAS DETECTOR

Ynyr Theophilus Harris

Wolfson College, University of Oxford



A thesis submitted for the degree of

Doctor of Philosophy

Long Vacation 2023

Abstract

This thesis presents the design of the first search for QCD instantons at the LHC, which is based on $1.693\,\mathrm{nb}^{-1}$ of low-pileup data collected by the ATLAS Experiment in 2015. QCD instantons are semi-classical solutions of the Yang-Mills equations of motion and correspond to tunnelling processes between topologically distinct vacuum sectors. Their very existence reveals the rich vacuum structure of theory of the strong interactions, where the different vacuum sectors are continually being mixed together. Instanton field configurations are deeply embedded in the Standard Model of particle physics, being already essential in explaining the heaviness of the η' meson. Yet they have never been observed. The search presented here uses a Machine Learning technique based on Deep Neural Networks (DNNs) for concomitant signal discrimination and background estimation. Despite the deployment of DNNs, signal discrimination has proved to be difficult, with instanton event kinematics being very similar to those of the minimum-bias background events. Blinded results show that the analysis has an expected exclusion sensitivity to a signal strength value of $\mu^{\mathrm{Exp}} = 412.6^{+151.9}_{-110.3}$.

Acknowledgements

A research degree in particle physics is a complex operation that requires a huge amount of concerted support. For that it pretty much worked out in my case, I am indebted to a correspondingly huge number of people. I use this section to extend my sincere thanks to a few of them.

Starting with the Particle Physics department at the University of Oxford, I would like to thank Sue Geddes and Kim Proudfoot for administrating the logistics of the degree. Thanks to the ATLAS students Gabriel Gallardo, Aaron O'Neill, Santiago Paredes, Jonas Würzinger, Philipp Windischhofer, Iza Velišček, Eimear Conroy, Siyuan Yan, Tim Brückler, Alessandro Ruggiero, Simon Koch, Maggie Chen, Olivier Salin, Radek Grabarczyk, Tom Yildirim, and Tom Dingley, for being so much fun. Claudia Merlassino, *thank you* for dealing with the most difficult questions and edge-cases that I could muster, even when you were not contractually obliged to listen to them. To the other Oxford postdocs, Koichi Nagai, Lydia Beresford, Elisabeth Schopf, Bill Balunas, Chris Pollard, and Holly Pacey, thank you for showing how it's done.

Niamh, Daniel, Jenny, along with the other Mancunians, thanks for coming with me to Oxford and immediately making it feel like home. Your faces bring me joy.

Thank you to Wolfson College for hosting my intellectual life outside of the office at Oxford, for accommodating me (even through a pandemic), and for providing such a community of wonderful people where I have felt completely at home.

I owe my thanks to the UK students and others at CERN who made my eighteen-month attachment there a delight. There are too many of you to name, but you will know who you are. Collectively you changed my priorities and now life is much brighter.

I am privileged to be able to work with an international group of people as part of the ATLAS Collaboration at CERN. I am deeply grateful to the instantons analysis team in particular for standing by me for the last two and a half years. Thank you Tancredi Carli for showing me by example how to do physics properly, and for always taking time to discuss. Thank you Matthias Schott for your flair for physics, efficiency, and for always knowing what to do. Thank you to the Prague contingent, Tom Sykora and Radek Vavřička, for your faithful attendance of our weekly analysis meetings and all the good questions. Thank you all for entrusting me with this analysis, and for your patience while I figured it out. I am obliged to Simone Amoroso for the success of its first phase — that you had to leave ATLAS remains one of the worst things that happened to this thesis.

The results presented here depend critically on work done by Frank Krauss in the SHERPA code. I would like to thank you Frank for implementing the instanton event generation, and for volunteering so much time to explain all of it. Indeed no good deed goes unpunished! For the help with event generators, thanks also to Steffen Schumann and Max Knobbe for useful discussions and for providing tune parameters, as well as to Frank Siegert for advice on running bleeding edge code with ATLAS Collaboration software.

In the broader ATLAS context, many thanks to Lydia and Marek Tasevsky, who convened the group in which the QCD instantons analysis got up-and-running. Thanks in particular to Lydia for long and interesting discussions, and for being a solid friend for the whole duration of my time at CERN. More recently, the analysis was stewarded by Jesse Liu and Karel Cerny, who have taken care of its progression. Thanks in particular to Jesse for his empathy and clarity of instruction as we try

to make a paper. Thanks are owed to Monica Dunford and Oldrich Kepka for licensing the instantons analysis and for many useful conversations over similarly many lunches.

I must thank Susumu Oda for his supervision of my qualification task in the SCT group of ATLAS, and for making a steep learning curve appear surmountable. In this context, thank you to Koichi and Elisabeth again, as well as to Dave Robinson, Taka Kondo, Shigeki Hirose, Hidetoshi Otono, Hannah Borecka-Bielska, Cristiano Sebastiani, and Simran Gurdasani.

Thank you to my supervisor, Alan Barr, for supporting my career and for finding such an interesting project for this thesis. The intellectual freedom that I have been granted throughout my DPhil has allowed me to explore (sometimes in a completely wayward fashion), and has made me a more useful researcher. Thanks in equal measure to Claire Gwenlan, for all the work stuff, but also for all the St Hilda's dinners, for helping us dominate the Oxford pub quiz scene, and for the friendship.

I am grateful to the Science and Technology Facilities Council (STFC) for funding this part of my education, and to the Particle Physics department at the University of Oxford, as well as Wolfson College for supplementing my travel resources.

Diolch i bawb adra, am gefnogi hyn i gyd yn ddiamod.

Preface

The work presented in this thesis was performed in the context of the ATLAS Collaboration at CERN. The ATLAS Experiment is an international collaboration of more than 3000 active members, and necessarily involves sharing work. Here I provide a list of my personal contributions to this programme.

Studies of SCT modelling performance

The modelling of charged-particles as they traverse the Semiconductor Tracker (SCT) of ATLAS is important to understanding data. I performed two studies of the agreement between data and Monte Carlo in SCT-related quantities to better understand some mis-modellings that were known to exist. The first study was performed using Minimum Bias events at low pileup values, while the second was performed using a tag-and-probe method in $Z \to \mu\mu$ events in a high-pileup dataset. I concluded that it is the radiation damage suffered by the SCT over the course of Run 2 that is mis-modelled. This could in principle be corrected by a periodic calibration of the radiation damage parameters used in SCT detector simulation.

Search for QCD instantons

For the QCD instantons search which constitutes the bulk of this thesis, I designed and led the analysis from the offset. All of the results presented in this thesis, including all the figures and tables, were made by me unless clearly stated otherwise

in the captions.

Tuning of the SHERPA 3 Monte Carlo event generator

As a supporting act to the QCD instantons analysis, it was necessary to conduct a comprehensive re-tuning of the phenomenological parameters of the Underlying Event model of the SHERPA 3 event generator. I instigated this effort and single-handedly did all of the work. This project is documented as Appendix A of this thesis, which is cited from Section 5.1 when the signal Monte Carlo samples that were affected are introduced.

Contents

1 Introduction **2**

2 The Standard Model and Beyond **4**

2.1 The SM Lagrangian 6

2.2 Simulation of Scattering Processes 13

2.2.1 MC event generators 14

2.2.2 Detector simulation 19

2.3 Conclusion 19

3 Theory of Instantons **22**

3.1 Generalities 23

3.2 Many Vacua 25

3.3 The Effective Vertex 26

3.4 Predictions for the LHC 27

4 The ATLAS Experiment **31**

4.1 The Large Hadron Collider (LHC) 33

4.1.1 Lattice layout and high luminosity 33

4.1.2 Machine performance 34

4.2 The ATLAS Detector 35

4.2.1 Inner Detector 36

4.2.2 Calorimeters 39

4.2.3 Muon Spectrometers 40

4.2.4 Trigger and Data Acquisition (TDAQ) 40

4.3 Event Reconstruction 41

4.3.1 Tracks . . . 42
4.3.2 Primary Vertices . . . 44
4.3.3 Calorimeter jets . . . 46

5 Search Design **47**

5.1 Data and Monte Carlo Samples . . . 49
5.1.1 Signal modelling . . . 51
5.1.2 Background modelling . . . 54
5.1.3 Event and track preselection . . . 57
5.2 Discriminating Variables . . . 58
5.2.1 Event kinematics . . . 58
5.2.2 Event Shapes . . . 59
5.2.3 Flatenicity . . . 62
5.2.4 Flavour variables . . . 62
5.2.5 Distributions of discriminating variables . . . 63
5.2.6 Correlation matrices of discriminating variables . . . 65
5.3 Region Definitions and Background Estimation . . . 72
5.3.1 Machine Learning approach . . . 72
5.3.2 Implementation of the ML approach . . . 75
5.3.3 Mass binning . . . 78
5.3.4 Network health checks . . . 80
5.3.5 Classifier output distributions . . . 82
5.3.6 What do the networks learn? . . . 85
5.3.7 ABCD background estimation . . . 98
5.3.8 Blinding strategy . . . 100
5.3.9 Background model validation . . . 100

6 Statistical Analysis **107**

6.1 The Profile Likelihood Ratio . . . 108
6.2 Hypothesis Testing . . . 110

6.2.1 CLs . . . 111
6.3 Likelihood Configurations . . . 112
6.3.1 One-bin counting experiments . . . 113
6.4 Systematic Uncertainties . . . 114
6.4.1 Experimental uncertainties . . . 114
6.4.2 Theory modelling . . . 115
6.4.3 Background estimation . . . 123
6.4.4 Prescription . . . 131
6.5 Blinded Results . . . 133

7 Conclusion **139**

A Tuning of Sherpa 3 **141**
A.1 Introduction . . . 141
A.2 Models . . . 142
A.2.1 Multi-Parton Interactions . . . 142
A.2.2 Intrinsic Transverse Momentum . . . 144
A.2.3 Hadronisation . . . 145
A.3 Tuning Strategy . . . 146
A.3.1 The Professor toolkit . . . 148
A.3.2 Underlying Event variables . . . 149
A.3.3 Reference data and MC samples . . . 150
A.4 Results . . . 153
A.4.1 Tune performance and features . . . 154
A.4.2 Conclusion and future work . . . 160

Bibliography **185**

Chapter 1

Introduction

From the point of view of human life on earth, much of Nature is concealed out of sight. This is because human life plays out on length-scales of metres and miles, and timescales of seconds, minutes, and years. In regard of what fabric this life is made of, humans have zoomed-in, using ever more powerful microscopes, to unveil strange interiors.

Contemporary High Energy Physics (HEP), or particle physics, studies the most fundamental building blocks of the universe. Using the most powerful microscope ever built, the Large Hadron Collider (LHC) at CERN, we are able to witness events at length-scales of 1 fm, where quantum mechanics reigns and the main characters are the elementary particles of nature and their interactions. The dynamics of these particles are encoded to the best of our knowledge in the Standard Model (SM) of particle physics. The LHC is a colossal feat of engineering and a triumph of international cooperation, and stands as a monument to human ingenuity, a cathedral to curiosity.

The Standard Model is experimentally-driven, but has held its ground in light of new datasets at higher energies despite exhaustive searches for new physics and stress-tests. With its integrity unbroken, 'SM predictions' are held in high esteem.

Most predictions are made based on perturbation theory, as the equations of the SM can usually not be solved exactly. But this calculational paradigm misses interesting, non-perturbative phenomena. This thesis pursues a phenomenon called the QCD instanton that fits into this category.

QCD instantons are semi-classical solutions of the Yang-Mills equations of motion, and correspond to tunnelling between topologically distinct vacuum sectors. They reveal that the vacuum state of the strong interactions has a rich topological structure consisting of a broiling landscape of instantons. They are a very basic prediction of the SM, and are required for the consistency of the theory, but have never been observed. Their inclusion in a transition amplitude induces a spectacular isotropic decay that should be observable to the ATLAS Experiment at the LHC. In 2020, the cross-section of this process at present LHC energies was predicted and published along with a viable plan for its discovery [1].

Shortly afterwards, the search that became the work of this thesis was instigated within the ATLAS Collaboration with the aim of discovering this qualitatively new effect of the SM. In an early presentation of the plan, a senior physicist struck a cautionary note, referring to QCD instantons as *le serpent de mer* — a recurrent issue that will not go away. In the ensuing analysis, signal discrimination has been a challenge, but the search is now at a mature stage and is undergoing approval by the ATLAS Collaboration. This thesis culminates with blinded results showing the expected sensitivity.

The thesis opens with a review of the theory of particle physics in Chapter 2, before Chapter 3 continues into the theory of QCD instantons specifically. The dataset used for the search comes from the ATLAS Experiment, which is introduced along with the LHC in Chapter 4. Then the stage is set to introduce the design of the search, which is done in Chapter 5. The statistical analysis and the results are presented in Chapter 6, before conclusions are drawn in Chapter 7.

Chapter 2

The Standard Model and Beyond

The Standard Model (SM) of Particle Physics is a triumph of contemporary physics. It was born from the union of quantum mechanics and special relativity as brokered by Dirac, who sought to put time and space coordinates in quantum mechanics on an equal footing. Dirac managed to do this partly by bringing to centre stage the generator of canonical transformations, the action functional [2, 3]. The action is the time-integral of a Lagrangian, which is fit for purpose as it can be expressed in a manifestly Lorentz invariant way. A century later, with the SM in its maturity, everything we know about the elementary particles and their interactions can be summarised in remarkably few equations.

The first equation to note is the Feynman path integral [4], which prescribes how to calculate a quantum mechanical amplitude from Dirac's classical action, S:

$$\langle \Phi(t_f) | \Phi(t_i) \rangle = \int \mathcal{D}[\Phi] \, \mathrm{e}^{iS[\Phi]}, \tag{2.1}$$

where square brackets indicate functional dependence on the field configurations Φ, and the $\mathcal{D}[\Phi]$ notation represents the measure of integration. The integral is over all paths in the space of field configurations between states $|\Phi(t_i)\rangle$ and $|\Phi(t_f)\rangle$ which are defined at times t_i and t_f, respectively. The integrand says that all paths contribute

equally in magnitude, but with a phase given by the classical action (in units of $\hbar$).

The path integral clarifies the connection between the regimes of 'big', where classical physics reigns, and 'small', where quantum mechanics applies. The exponential part of Eq. 2.1 is oscillatory, and has a frequency of $1/\hbar$. In the classical limit where $\hbar \to 0$, the integrand oscillates rapidly as the path changes, and tends to wash out all paths except the one for which the action is stationary. So the integration only sees the path for which the action is stationary, which in classical mechanics is precisely the solution of the equations of motion — a system takes the path of extremal action. Meanwhile in quantum mechanics, $\hbar$ is not small, so every path must be considered — thus a system takes every possible path. Eq. 2.1 graduates us from classical to quantum physics.

In an experimental setting, we are usually concerned with multi-particle scattering processes of the form $2 \to$ many. The formalism for calculating an amplitude for this process considers asymptotic *in* and *out* states of definite momenta p_a, p_b and $k_1, k_2, \ldots, k_n$, respectively. These are defined in the infinite past and the infinite future, respectively. The evolution of the system from the former to the latter is described by a so-called S-matrix, which is defined implicitly in

$$ {}_{\text{out}}\langle \alpha; k_1, k_2, \ldots, k_n | \beta; p_a, p_b \rangle_{\text{in}} = \langle \alpha; k_1, k_2, \ldots, k_n | \hat{S} | \beta; p_a, p_b \rangle. \tag{2.2} $$

On the right-hand side, the states are defined at the same time after the Heisenberg operator $\hat{S}$ has been taken out. In terms of the parameters α and β used to label the states here, the operator form of S can be interpreted as a matrix of elements $S_{\alpha\beta}$. This S-matrix then encodes the full dynamics of the interacting system. It contains a trivial part where no scattering occurs, so it is conventional to define the matrix $\mathcal{M}$ by

$$ S = 1 + i(2\pi)^4 \delta^{(4)} \mathcal{M}, \tag{2.3} $$

where the delta-function that imposes overall four-momentum conservation has been

extracted.

Finally, the total cross-section for the $2 \to$ many process is given by

$$\sigma \sim \int \mathrm{d}\Pi_n \, |\mathcal{M}(p_a, p_b \to k_1, k_2, \ldots, k_n)|^2, \tag{2.4}$$

where the measure of integration is the relativistically-invariant differential phase space of an n-body final state. An experiment will only ever measure a fiducial cross-section, which is a cross-section in a restricted phase space that aligns with the acceptance of the experimental apparatus. A total cross-section can be obtained from a fiducial cross-section by a theory-driven extrapolation to the full phase space.

Equation 2.4 separates the 'dynamics' of the system (parts affected by the details of the Lagrangian), which go into $\mathcal{M}$, from the 'kinematics' of the specific experimental setup, which go into $\mathrm{d}\Pi_n$. Calculating the cross-section from these ingredients is straightforward, so the whole theory of particle physics boils down to evaluating the path integral. To see how this is done, we must specify a form for the action S.

2.1 The SM Lagrangian

In this section, the Lagrangian expression of the SM is introduced, with its content drawn from the textbooks of Refs. [5–8]. The action of Eq. (2.1) is usually written in terms of a Lagrangian density, which is related to the action by $S = \int \mathrm{d}^4x \mathcal{L}$. It is a beautiful fact that if we build this Lagrangian out of all renormalisable terms consistent with Lorentz and local gauge symmetry, we get the Standard Model Lagrangian that is used by Nature:

$$\mathcal{L}_{\mathrm{SM}} = \underbrace{-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{Yang-Mills}} + \underbrace{i\overline{\psi}\not{D}\psi}_{\text{Dirac}} + \underbrace{y\overline{\psi}\phi\psi}_{\text{Yukawa}} + \underbrace{|D_\mu\phi|^2 - V(\phi)}_{\text{Higgs}}\,. \tag{2.5}$$

This expression is schematic, since gauge and flavour indices are suppressed for clarity, as are the chiralities of the spinors ψ and Hermitian conjugate terms. In Eq. 2.5, $\mathcal{L}_{\text{SM}}$ has been parsed into four parts according to interaction type. These interactions are named after their founders.

- ***Dirac.*** Conceived for the relativistic description of the electron in Quantum Electrodynamics (QED), this generalised Dirac part is the kinematic term for fermions, which are written as Dirac spinors ψ. The covariant derivative D is contracted with the Dirac gamma matrices γ, forming $\not{D} = \gamma^\mu D_\mu$, to ensure Lorentz invariance. The covariant derivative arises from the connection of the fermion field at two points in space by the gauge fields.

- ***Yang and Mills.*** This is the kinetic term of the gauge fields. It is built out of the field strength tensors $iF_{\mu\nu} = [D_\mu, D_\nu]$. There is one copy of $F^{\mu\nu}F_{\mu\nu}$ for each generator of gauge transformation. This means that there is one copy for electromagnetism, three for the weak interactions, and eight for the strong interactions.

- ***Yukawa.*** Originally introduced to describe the interactions of pions and nucleons, the Yukawa part describes the interaction of the scalar Higgs field ϕ with the fermion fields ψ, and yields a fermion mass term. The coefficients of this term y contain the fermion masses and the elements of the Cabbibo-Kobayashi-Maskawa (CKM) quark-mixing matrix.

- ***Higgs.*** The Higgs field is a complex scalar doublet whose excitation H is interpreted as a physical Higgs boson. The Higgs term describes the kinematics of this field as well as the shape of its potential V. The form of the potential is taken to be $V = -\mu^2\phi^2 + \lambda\phi^4$. When $\mu, \lambda > 0$, this is the Mexican-hat potential responsible for electroweak symmetry-breaking.

There is a fifth term, $\mathcal{L}_\theta$ called the ***topological term***, that is consistent with symmetry principles, but it is omitted from the canonical expression of $\mathcal{L}_{\text{SM}}$ because

Particle Type	Spin	Field	Representation
Quarks, q^i	$\frac{1}{2}$	$(u \quad d)_L$	$(\mathbf{3}, \mathbf{2}, +1/6)$
		u_R	$(\mathbf{3}, \mathbf{1}, +2/3)$
		d_R	$(\mathbf{3}, \mathbf{1}, -1/3)$
Leptons, l^i	$\frac{1}{2}$	$(\nu_e \quad e)_L$	$(\mathbf{1}, \mathbf{2}, -1/2)$
		e_R	$(\mathbf{1}, \mathbf{1}, -1)$
Gauge, A_μ	1	G^a_μ	$(\mathbf{8}, \mathbf{1}, 0)$
		$W^\pm_\mu, W^0_\mu$	$(\mathbf{1}, \mathbf{3}, 0)$
		B^0_μ	$(\mathbf{1}, \mathbf{1}, 0)$
Higgs, H	0	$(\phi^+ \quad \phi^0)$	$(\mathbf{1}, \mathbf{2}, +1/2)$

Table 2.1: Summary of the particles of the Standard Model grouped according to their representation under $(\mathrm{SU}(3)_c, \mathrm{SU}(2)_L, \mathrm{U}(1)_Y)$ transformations. The superscript i on the quarks and leptons runs over the three generations. The $\mathrm{U}(1)_Y$ charge Y is related to the third-component of isospin T^3 and electric charge Q through $Q = T^3 + Y$.

it is a total divergence and does not appear in perturbative expansions.

SM particles are excitations of the fields participating in $\mathcal{L}_{\mathrm{SM}}$, and they furnish representations of the SM gauge group $\mathrm{SU}(3)_c \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. Gauge bosons go in adjoint representations, while fermions go in fundamental representations as summarised in Table 2.1. The field content of the SM and the representation of the particles is mostly experimentally-driven. Once its physical parameters are determined by measurement, the specification of the SM is complete.

Quantum chromodynamics

Buried in Eq. (2.5) are the Lagrangians of the different interactions incorporated in the SM, which are often regarded separately. The focus of this thesis is on quantum chromodynamics (QCD), so this section will hone in on that part of Eq. (2.5). The QCD Lagrangian can be extracted and written exclusively as

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4} G^{\mu\nu}_a G^a_{\mu\nu} + i\overline{\psi}_i \gamma^\mu D_\mu \psi_j, \tag{2.6}$$

where gauge indices a, and flavour indices i and j, have been written explicitly. The field strength objects have been written as $G^a_{\mu\nu}$ to emphasise that they are gluon field strengths.

QCD is distinguished by the fact that its local gauge transformations respect SU(3) group symmetry. This group spans eight generators t, which are indexed here by $a = 1, ..., 8$. Gauge invariance is maintained in the kinetic term as usual by introducing the covariant derivative, $D_\mu = \partial_\mu - igA^a_\mu t^a$.

The SU(3) group is said to be non-Abelian because its generators do not commute. For commutation relations we have $\left[t^a, t^b\right] = if^{abc}t^c$, where the proportionality symbols f^{abc} are called structure constants. As a result, the gluon field strength tensor has an additional term compared to QED (which is Abelian), becoming

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu. \tag{2.7}$$

Thus the full QCD Lagrangian includes terms that are cubic and quartic in the gauge fields, and gluons can self-interact.

Perturbation theory and renormalisation

With the relevant parts of the path integral of Eq. (2.1) constructed, all that remains is to evaluate it. However, this is difficult for two major reasons [7]. The first subtlety is in the defintion — to evaluate $\langle\Phi(t_f)|\Phi(t_i)\rangle$, we need to know $|\Phi(t_{\{i,f\}})\rangle$ for the boundary conditions, but the definition of these states depends on the solution. To circumvent this, vacuum-to-vacuum transition amplitudes, $\langle\Omega|\Omega\rangle_J$, in the presence of arbitrary source terms J are considered instead. In terms of these amplitudes, the path integral can be used to manufacture correlation functions among fields, which are in turn related to the $\mathcal{M}$-matrix.

The next complication is that the integrations $\int \mathcal{D}[\Phi]$ are intractable. However, the

free theory is solvable, so the problem can be attacked perturbatively. A perturbative expansion works by expanding the exponential part of the path integral as a power series in the coupling constants and keeping the first few orders of terms.

The formalities of this expansion are unwieldy, which is why the series is more commonly assembled pictorially in terms of Feynman diagrams. The resulting Feynman diagrams suffer divergences, which arise loosely because space is big (*infrared* divergences) and because the momenta carried by virtual particles around loops are unbounded from above (*ultraviolet* divergences).

The ultraviolet divergences can be systematically removed by a procedure called renormalisation. This programme consists of adding counter-terms to $\mathcal{L}_{\mathrm{SM}}$ that have the effect of shifting the bare couplings, masses, and fields in such a way that the theory is rendered finite. One of the big moments for the SM's quantum field-theoretic formulation was when 't Hooft and Veltman showed that such redefinitions are always possible for Lagrangian terms of mass dimension up to four [9]. Renormalisation is not mere trickery, since these counter-terms can be interpreted as quantum vertex and vacuum corrections. Thus renormalisability is a physical requirement, which is is why renormalisability was demanded when Eq. 2.5 was written down.

In Euclidean space (related to Minkowski space by a Wick rotation, $t \to -it$), the path integral can alternatively be evaluated in the saddle point approximation to first order in $\hbar$. This proceeds by expanding the action around a particular configuration ϕ_{cl} for which it is stationary at a value S_E^{cl}. Denoting a small perturbation about ϕ_{cl} by $\eta(x)$, then

$$S_E = S_E^{\mathrm{cl}} + \frac{1}{2}\int \mathrm{d}^4x\mathrm{d}^4y\ \eta(x)\frac{\delta^2 S_E}{\delta\phi(x)\delta\phi(y)}\eta(y) + \ldots \tag{2.8}$$

The classical action S_E^{cl} is no longer a function of the fields, and can be taken outside of $\int \mathcal{D}[\Phi]$. What remains is a Gaussian integral that can be performed. This expression provides the interpretation that the full path integral yields quantum corrections to

the classical equations of motion. Physical amplitudes are obtained from here by analytic continuation back to Minkowski space.

Running of the coupling constants

The success of the perturbation theory depends on the coupling constants used as expansion parameters being truly small. Although the coupling constants g are classically scale-independent, their quantum corrections induce a scale-dependence that is of physical consequence. The running of g with energy scale μ is given from the renormalisation group equations by

$$\mu \frac{dg}{d\mu} = \beta(g). \tag{2.9}$$

The character of the running of the coupling therefore depends on the form of the β-function on the right-hand side. In QED, β is slightly positive and evolves slowly with μ, meaning that it is safe to perform a perturbative expansion.

On the other hand, in QCD the β-function is given by

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left[\frac{11}{3}C_2(G) - \frac{4}{3}n_f C(r)\right], \tag{2.10}$$

in terms of the group-theoretic constants $C_2(G) = 3$ and $C(r) = 1/2$, which for small numbers of quark flavours n_f is *negative*. Thus in QCD, the coupling runs in such a way that at small distances and high energies, the interaction strength is small, while at large distances and low energies, the interaction strength is high.

Based on this, QCD is said to have the property of asymptotic freedom. At low energies, the divergent QCD coupling constant gives rise to confinement, which explains why quarks are not seen as free particles at low energies, but rather only as constituents of colour-neutral bound states such as baryons and mesons. This observation was made by Politzer [10], and independently by Gross and Wilczek [11].

The perturbative expansion is only valid where $g^2(\mu) \ll 1$, which enables us to define a dimensionful parameter Λ as a fuzzy boundary between the perturbative and non-perturbative regimes. A reference value for g is usually measured via $\alpha_S = \frac{g^2}{4\pi}$ at the Z-boson mass m_Z, so with $n_f = 5$ we have $\Lambda \sim 1\,\text{GeV}$.

Anomalies

The breaking of the classical scale-invariance of the coupling constants, characterised by the β-function, is an example of an anomaly, which is when an exact symmetry of the classical action is broken by quantum corrections [7].

Another anomaly featured in this thesis is the Adler-Bell-Jackiw (ABJ) chiral anomaly [12, 13]. Looking at $\mathcal{L}_{\text{SM}}$ through classical lenses, QCD does not appear to be a chiral theory, which means that the QCD Lagrangian should be invariant under chiral transformations. However, this symmetry is anomalous as can be seen in the divergence of the axial vector current, j^5_μ, which does not vanish. Instead, it evaluates to

$$\partial^\mu j^5_\mu \propto -iF^{\mu\nu}\widetilde{F}_{\mu\nu}, \tag{2.11}$$

where $\widetilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\sigma\rho}F_{\sigma\rho}$ is the dual of $F_{\mu\nu}$.

The right-hand side of Eq. (2.11) can be seen as arising due to the non-invariance of the measure of integration over fermion fields in the path integral under chiral transformation [14], or from radiative corrections due to "triangle diagrams" [12].

This anomaly in QED contributes to the lifetime of the neutral pion decay $\pi^0 \to \gamma\gamma$, and in QCD contributes to the mass of the η' meson.

2.2 Simulation of Scattering Processes

The partonic-level cross-section obtained through a calculation like Eq. (2.4) is only a part of the picture of what happens during a collision event at the LHC. In proton-proton collisions, the initial state partons are transported inside protons, whose structures are parameterised in terms of parton distribution functions (PDFs). A PDF denoted by $f_i(x, \mu_\mathrm{f})$ can be thought of as the probability of finding a parton of species i carrying a fraction x of the proton's momentum when the proton is probed at a momentum scale μ_f.

The partons participating in a scattering event are sampled from these PDFs, so to calculate a total hadronic cross-section, the partonic-level cross-section must be convolved with the PDFs according to

$$\sigma_X = \int \mathrm{d}x_1 \mathrm{d}x_2 \sum_{i,j} f_i(x_1, \mu_\mathrm{f}) f_j(x_2, \mu_\mathrm{f}) \hat{\sigma}_{ij\to X}(\mu_\mathrm{f}, \mu_\mathrm{r}; x_1 x_2 s), \tag{2.12}$$

where μ_f and μ_r are the factorisation and renormalisation scales, respectively, and $s' = x_1 x_2 s$ is the partonic centre-of-mass energy squared at which the partonic cross-section is evaluated.

Coloured partons cannot exist as stable particles, so when partons exit the scattering event, they shower perturbatively downward in energy until they hadronise non-perturbatively into the stable particles that may be seen in the detector. This downward evolution through energy scales can be written in factorised form as

$$\mathrm{d}\sigma_X(Q) \times \mathrm{PS}(Q \to \mu) \times \mathrm{Had}(\mu \to \Lambda) \times \ldots, \tag{2.13}$$

where $\mathrm{PS}(Q \to \mu)$ describes a parton shower, and $\mathrm{Had}(\mu \to \Lambda)$ describes the hadronisation process.

2.2.1 MC event generators

This factorisation of the whole structure of an event facilitates its simulation using Monte Carlo (MC) techniques. There are a number of codes, called MC event generators, that implement the simulation of scattering events at the LHC according to this picture. The structure of an event as modelled by the PYTHIA 8 event generator is depicted in Figure 2.1 for example, which emphasises the complexity inherent in hadron collisions. MC event generators are important as they are used to separate signal from background processes, as well as for fitting model parameters. In this thesis, three MC event generators are used. These are namely SHERPA [15], PYTHIA 8 [16, 17], and EPOS [18]. The configurations of these samples that were used for this thesis will be described in more detail when they become relevant in Chapter 5. Here, we give a brief overview of the components of these generators based on two recent reviews [19, 20].

Parton showers

One can think of a parton shower as a numerical implementation of the renormalisation group equations that connect the two scales Q and μ [20]. The standard collinear parton shower picture involves a cascade of parton splittings, whereby the partons coming from the hard scatter radiate colour charges in the form of secondary partons until the resultant partons can hadronise.

These parton cascades are implemented probabilistically using a Sudakov form factor, $\Delta_i(Q^2, q^2)$, to encode the probability that no emission occurs between the scales Q^2 and q^2,

$$\Delta_i(Q^2, q^2) = \exp\left\{-\int_{q^2}^{Q^2} \mathrm{d}\Phi(t, z, \phi)\mathcal{K}_{ji}(t, z, \phi)\right\}. \tag{2.14}$$

The phase space for the emission, $\mathrm{d}\Phi(t, z, \phi)$, depends on an ordering parameter, t, the momentum fraction carried by the emitted particle, z, and the azimuth of the emitted particle, ϕ. The splitting kernel $\mathcal{K}_{ji}$ for the splitting between parton

species $j \leftarrow i$ depends on the same phase space parameters, as well as the strong coupling, $\alpha_S(p_\perp)$, evaluated at the current transverse momentum scale. The ordering parameter corresponds to the variable in which the parton shower evolution takes place. For illustrative purposes in Eq. (2.14), the evolution was ordered by parton virtuality, q^2.

The probability kernel $\mathcal{K}_{ji}$ can diverge in two ways: (i) in the limit where the emitted particle becomes soft, with $z \to 0$, and (ii) as the emitted particle becomes collinear with the emitter or equivalently as its transverse momentum vanishes. These soft and collinear divergences are well understood analytically, but correspond physically to emissions that are either indistinguishable or below the resolution of a relevant experimental apparatus. Parton showers generate real quanta by imposing a cut-off for these soft and collinear emissions, for example by demanding a minimal transverse momentum. As a result, parton showers implicitly account for the unresolvable divergences, and thus construct loop corrections from unitarity arguments and the non-emission probability [19].

The virtuality-ordered collinear parton shower evolution shown above does not properly account for interference effects between soft emitted gluons since it does not consider the event's colour structure [19]. Colour interference effects can be simulated when the angle of emission is used as the evolution parameter, leading to angular-ordered (or "colour-improved") parton showers.

A more recent approach to colour interference effects is the colour-dipole parton shower. By following colour flow through a diagram, colour partners are identified and are taken to constitute a colour dipole. These dipoles consist of an emitter and a spectator, and undergo a transverse momentum-ordered evolution. In this formulation of $2 \to 3$ splittings, four-momentum is conserved at every step, and matching the parton shower to the matrix element is straightforward, all while colour coherence is accounted for at leading colour. The colour dipole parton shower formalism is so preferred in practice that some variation of it is used in all the main

MC event generators, including in PYTHIA 8 [21] and SHERPA (in two different implementations [22, 23]). The choice of parton shower realisation has important consequences for the accuracy of the simulated radiation pattern.

Hadronisation

At the end of the parton shower, the surviving colour-charged partons must be converted into observable hadrons. In MC event generator parlance, this process is called hadronisation and is most commonly implemented according to two different schools of thought. These result in namely the Lund string model [24, 25], implemented in PYTHIA [16, 17], and the cluster model [26], implemented in SHERPA [15].

The Lund string model is based on the idea that the colour confinement potential of QCD is linear in the separation of colour charges. This potential energy can be imagined as comprising a colour flux tube between colour charges that stretches and contracts in the course of the yo-yo motion of the charges at the ends. An initial string is drawn between colour partners, before it fragments by a series of string-breaks until only on-mass-shell hadrons remain. Each resulting hadron then corresponds to a small piece of string with a quark on one end and an antiquark on the other.

For generating the quark-antiquark pairs that instigate a string break, the Lund string model introduces quantum mechanical tunnelling whereby a pair is produced at $m = p_\perp = 0$, before being pulled apart by the flux tube in which it is spawned. The amplitude for the tunnelling includes a flavour-independent Gaussian spectrum for the $p_\perp$ of the quark-antiquark pair, and an exponential dependence on quark masses that leads to suppression of heavy-quark production. Charm and heavier quarks are therefore not expected to be produced during soft fragmentation. For baryon production, string breaks into diquark-antidiquark pairs are considered. The probability of spawning a diquark as opposed to a quark is a free parameter, but it

is found that about 10% of string breaks must involve diquarks in order to produce the spectrum of observed hadrons.

The cluster model is similar, but is based on the idea of pre-confinement in parton showers [27]. Cluster hadronisation starts with the non-perturbative splitting of gluons into quark-antiquark (or diquark-antidiquark) pairs. The initial clusters are then formed from colour-connected pairs of partons. From here, clusters fragment by a series of quasi two-body decays until the resultant clusters can be identified by their mass, spin, and flavours, with hadrons. The cluster model in conjunction with an angular-ordered parton shower describes high-energy collision data reasonably well. The Lund string model description is generally slightly better, but accomplishes this at the expense of more free parameters.

A shortcoming of the string and cluster models alike is that strings and clusters fragment independently. If string or cluster densities become so high that they begin to overlap (which is feasible, particularly in the presence of multiple parton interactions), then these models should fail to describe any collective phenomena of the kind seen in heavy-ion collisions.

The EPOS LHC event generator takes a step towards describing this collective behaviour with a model of collective hadronisation. At early proper times, before hadrons have been formed but after string segments have been drawn, EPOS distinguishes between regions of high and low string segment density (in terms of some critical string segment density, ρ_0). Regions of high density are said to belong to a 'core', while regions of low density are said to belong to a 'corona' as described in Ref. [28]. The coronal part is hadronised using a standard string fragmentation model, and dominates the description of peripheral, large rapidity and low multiplicity events, but the core part is evolved collectively.

Clusters of string segments constituting the core are considered to be expanding like a fluid until they have cooled down enough for the string segments, now with modified

kinematics, to freeze out. EPOS evades the microscopic description of this collective motion by providing this simple parameterisation of its particle production instead. Not all particles are completely absorbed by the core — high transverse-momentum particles can punch through. To accommodate this, a phenomenological cut on transverse momentum, $p_{\perp}^{\text{cut}}$, is defined, above which a particle survives the core (at the expense of some momentum loss). Soft particles below $p_{\perp}^{\text{cut}}$ are completely absorbed.

The collective hadronisation model of EPOS LHC is comparable to the Lund string fragmentation model of PYTHIA in the corona, but improves upon the string model in the core with a phenomenological model of collective flow. EPOS LHC manages to describe minimum bias LHC data well [18], and does this with as few as about six free parameters [28]. The success of this hadronisation model suggests that there is indeed a collective flow element to hadronisation in the high-energy hadronic collision environment of the LHC.

Underlying Event

The hard scatter parts of interest, and the methods by which they are usually simulated, have at this point been introduced. However, this activity is accompanied by other effects associated with beam remnants and the composite structure of the colliding hadrons. These additional effects involve the intrinsic transverse momentum of the partons bound in the incoming hadrons, multiple parton interactions (MPI), and the break-up of the colliding hadrons into beam remnants. All of these effects colour-connect across the whole event, modifying final state kinematics. These additional effects are collectively referred to as the underlying event (UE).

The place of the UE in the context of a complete event as simulated in an MC event generator is illustrated in Figure 2.1, where the hard scatter process is labelled $\mathrm{d}\hat{\sigma}_0$. This figure depicts three parton-parton interactions (two of which, being additional

to the hard scatter, are labelled 'MPI'), the colour connections between these scatters, and their subsequent fragmentation.

Each MC event generator implements these effects in a different way. A central part is the attachment of MPI. In Pythia 8 and Sherpa, this is done based on some variant of the Sjöstrand–van Zijl model [29, 30]. The Sherpa implementation is discussed in detail in Appendix A in the context of UE tuning. In Epos, MPI are implicitly included in the ladder picture [31, 32].

2.2.2 Detector simulation

The ellipsis at the end of Eq. (2.13) represents the conversion of the hadronised final state into detector signals which can be compared directly to data. This final step can be decomposed into two parts. In the first part, stable final state particles are propagated through a representation of the detector material and geometry to find the distribution of energy in the detector's active media. Once this is done, the energy deposits are digitised into detector signals of the kind that would be read out for reconstruction by a physical detector. Simulated events are then reconstructed using the same algorithms as are used for the data, and theory predictions can be compared against nature. Detector simulation is performed in the ATLAS simulation infrastructure [33] using the Geant4 toolkit [34].

2.3 Conclusion

Comparison of theoretical predictions with data from the ATLAS Experiment testifies that the SM (as well as the calculational strategies introduced in this chapter) is hugely successful. This success is showcased in Figure 2.2, where predicted total and fiducial cross-sections agree with the data over several orders of magnitude and a variety of production modes and final states.

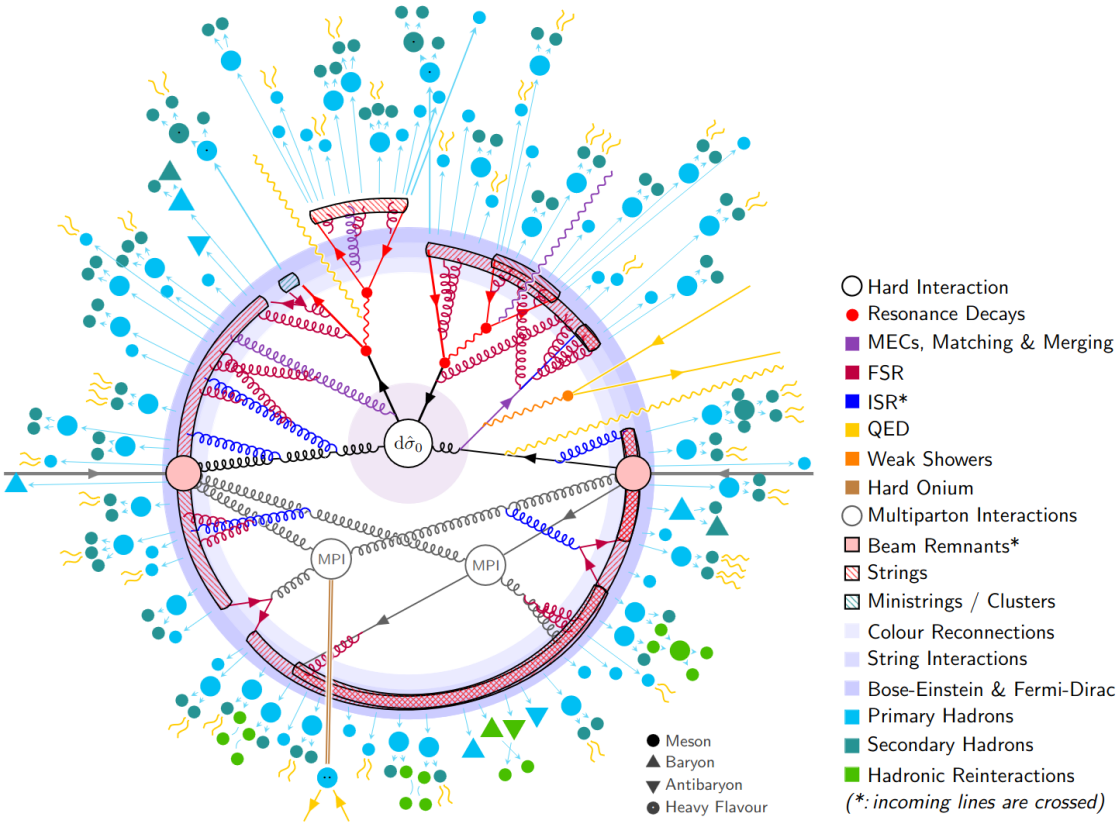


Figure 2.1: Schematic structure of an event in which a top-antitop quark pair is produced, in the language of modelling used by the PYTHIA event generator. This figure is reproduced from Ref. [17], where all of the components shown in the legend are described in more detail. A recent review of general-purpose Monte Carlo event generators for the simulation of scattering processes at the LHC can be found in Ref. [19].

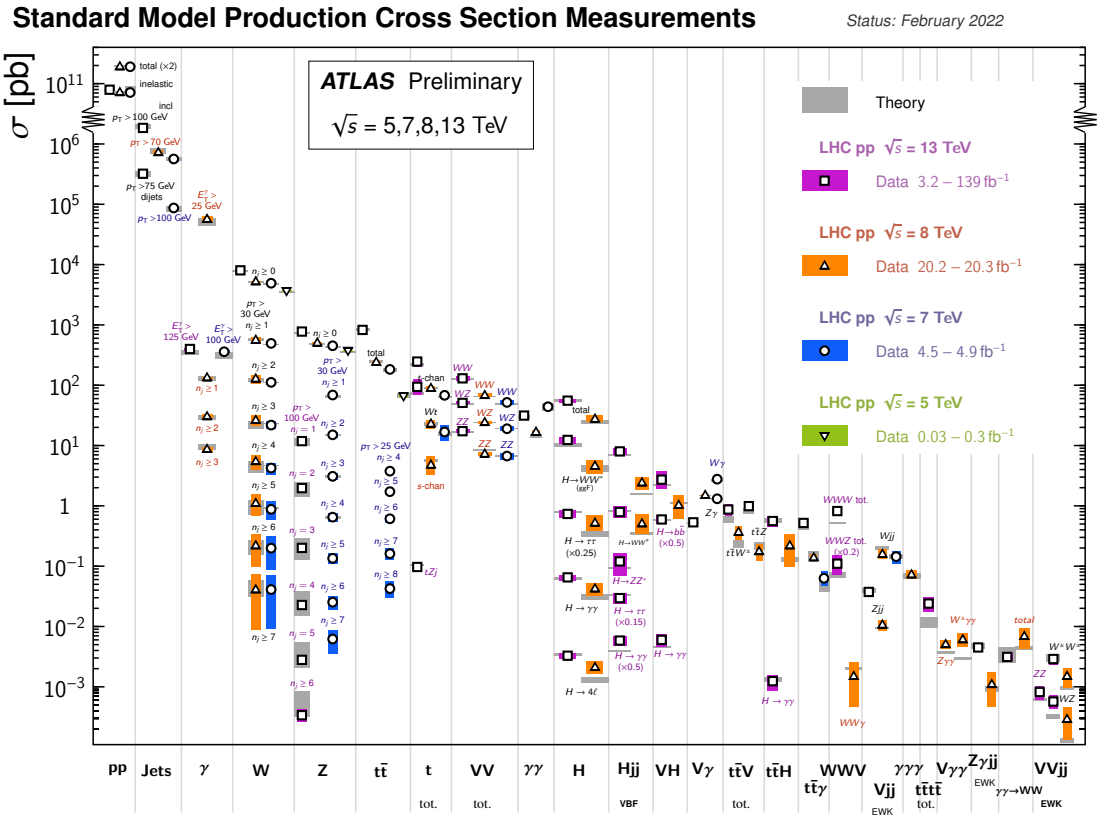


Figure 2.2: Summary of several Standard Model total and fiducial cross-section measurements. The measurements are corrected for branching fractions and compared to theoretical expectations. This figure is reproduced from Ref. [35].

Chapter 3

Theory of Instantons

The hills are shadows, and they flow
From form to form, and nothing stands;
They melt like mist, the solid lands,
Like clouds they shape themselves and go.
— Alfred, Lord Tennyson

Instantons were conceived in 1975 by Belavin *et al.* as semi-classical solutions of a pure Yang-Mills gauge theory [36]. It eventually became clear that these instantons are the cause of the chiral anomaly discussed in Chapter 2, since they are the field configurations that prevent the divergence of the axial vector current of Eq. (2.11) from vanishing [37]. In doing perturbation theory with respect to a background of instanton field configurations, an effective vertex is induced that in the case of three flavours has precisely the quantum numbers of a mass term for the η' meson. As a result, 't Hooft showed how instantons solve the U(1) problem of why the η' meson is so heavy [38], and instantons have been assumed to exist ever since.

Instanton field configurations are a working concept in theory, with a rich but subtle phenomenology. Pictures of the QCD vacuum as an instanton liquid suggest that instantons may have a further role to play in hadronic structure and confinement [39].

This could be the case if the vacuum expectation value in QCD is negative, as then quarks and gluons would be 'expelled' from the vacuum to be confined inside hadrons, with hadrons being bubbles of 'normal' phase in the 'abnormal' vacuum. This allegation is supported by Lattice QCD simulations [40], where instantons emerge in abundance and attest to the consistency of the concept.

The theoretical case for instantons is thus substantial. However, despite attempts at observation by the H1 and ZEUS experiments in electron-proton collisions at the HERA collider in Hamburg, QCD instantons remain unobserved [41–43]. They are a qualitatively different prediction of the SM to those made by perturbation theory, and as such provide a qualitatively different test of the SM's integrity. Furthermore, if the theoretical expectation is correct, then QCD instanton events constitute a large fraction of the collisions that take place at the LHC, and their understanding could improve our understanding and modelling of the LHC data, particularly in the 'Soft QCD' regime where non-perturbative effects are important.

This chapter gives a review of the theoretical background to instantons, following the narrative of the early literature with insight from the reviews of Refs. [6, 8]. The opening gambit is to study field configurations of finite action in pursuit of semi-classical solutions of the Yang-Mills equations[1].

3.1 Generalities

The derivation of the instanton proceeds in Euclidean spacetime in the context of a path integral, where the action of a pure gauge theory is written as a functional $S_E[A]$ of the gauge fields $A_\mu(x)$:

$$S_E[A] = \frac{1}{4}\int \mathrm{d}^4x F_{\mu\nu}F^{\mu\nu}, \tag{3.1}$$

[1] Semi-classical means small $\hbar$, which in a quantum field theory is tantamount to small coupling [8].

where $F_{\mu\nu}$ is the field strength tensor. For this action to be finite, the field strength must vanish faster than $\mathcal{O}(1/x^2)$ as x goes to infinity, which occurs if the gauge field approaches a pure gauge in the same limit. This condition can be expressed as $A_\mu \to g^{-1}(x)\partial_\mu g(x)$, where $g(x)$ are elements of the gauge group G.

From the topological point of view, every field $A_\mu(x)$ taking values in the gauge group defines a mapping from a sphere S^3 into the manifold of the gauge group G. This set of mappings can be partitioned into homotopy classes and has group structure. To each homotopy class, c, we may assign a topological invariant $q(c)$ by way of the Cartan-Maurer from [6]. This permits q to be used to index the homotopy classes and field configurations therein. When G is one of SU(2) or SU(3), this invariant evaluates to $q(c) = 24\pi^2 n$, where n is the number of times the manifold G is covered by S^3. The integer n is variously referred to as the 'winding number', as a topological quantum number, or as a Chern-Simons index.

The desired solutions of the Yang-Mills equations come from extremising the action (3.1) with respect to the gauge fields. This can be done by connecting the Cartan-Maurer form to the action integral, which boils down to

$$S_E = \frac{1}{4}\int \mathrm{d}^4x F_{\mu\nu}F^{\mu\nu} = 8\pi^2|n|. \tag{3.2}$$

Incidental to this result is that semi-classical solutions of the Yang-Mills equation are also solutions of the self-duality equation $F_{\mu\nu} = \pm\widetilde{F}_{\mu\nu}$, where $\widetilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\sigma\rho}F_{\sigma\rho}$ is the dual of $F_{\mu\nu}$ again.

The solution of topological charge 1 found by Belavin *et al.* (called the BPST instanton) is of the form [36]

$$A^{\mathrm{cl},a}_\mu(x) = \frac{2}{g}\frac{\eta^a_{\mu\nu}(x-x_0)^\nu}{(x-x_0)^2+\rho^2}, \tag{3.3}$$

where the coefficients η^a are the 't Hooft eta symbols [44], x_0 is the configuration's translation-invariant centre, and ρ is a free scale parameter. Because this configuration

is instantaneous in time (localised by the ρ parameter) and is soliton-like, it is called an *instanton*[2]. The parameters η^a, x_0^μ and ρ are the collective coordinates of the instanton, with the x_0 interpreted as the instanton's centre, ρ its size, and η^a its orientation in the gauge group.

This action appears in a path integral as a term $\exp(-S/g^2)$ summed over n. Instantons are a purely non-perturbative effect because this term vanishes faster than any power of g in a perturbation series. Instantons of higher winding numbers are exponentially suppressed in term, so the BPST instanton dominates.

3.2 Many Vacua

To obtain the interpretation of these solutions as tunnels between topologically distinct vacua, we must separate the winding number into time and space components according to Ref. [45]. In doing so, the winding number may be decomposed into an expression $n = n_+ - n_-$, the difference between the spatial winding numbers n_+ at $t = +\infty$ and n_- at $t = -\infty$. So the winding number is the difference in topological charge between the initial and final states; and, as the former configuration cannot be continuously deformed into the latter without passing through a configuration of infinite action between the two homotopy classes, the solution obtained must describe a discrete tunnelling process. This interpretation is consistent with the similarity of the $\mathcal{O}(\exp(-S/g^2))$ amplitude with the WKB formula for tunnelling encountered in quantum mechanics [46], which is a similarly non-perturbative effect involving evolution in imaginary time.

This topology unveils the rich vacuum structure. As we may continuously gauge-transform field configurations within homotopy classes down to the vacuum, we see that a Yang-Mills theory nominally possesses a discrete infinity of vacuum states. This situation is complicated by the fact that instanton field configurations, by

[2]Also called an Euclidean gauge soliton [37], or pseudoparticle [36].

tunnelling between vacuum sectors, have the effect of mixing these vacua.

The consequences were investigated in Refs. [45, 47], which show that the true vacuum is a superposition $|\theta\rangle = \sum e^{in\theta}|n\rangle$ of all topologically distinct vacua $|n\rangle$, with θ a free parameter. This mixing induces a new term in the action

$$\mathcal{L}_\theta \sim i\theta F\widetilde{F}, \tag{3.4}$$

which is precisely the so-called topological term that was omitted from the SM Lagrangian in Eq. (2.5). This term describes a different universe for each value of θ and generally violates P and CP symmetries. As it happens, the strong interactions of our universe are approximately CP-conserving, with θ close to zero.

As to why θ may take on a value so arbitrarily close to zero as to explain the approximate CP-conservation of the strong interactions that is observed, the axion particle was invoked [48–50], but has not yet been observed.

3.3 The Effective Vertex

Having found instanton solutions, we see that our usual approach of perturbation theory — of expanding the action about spacetime-independent vacuum expectation values — is incomplete. We must in principle include instantons of all winding numbers in the path integral as zeroth order contributions and consider fluctuations about them as in Eq. (2.8). In evaluating this amplitude, the instanton process is found to induce an effective vertex [44].

The participants in this "vertex" are constrained by strict selection rules on the transition. The transition rules emerge since by comparison of the pure-gauge action, S_E, of Eq. (3.2) with Eq. (2.11), we see that the transition has to produce the ABJ chiral anomaly. Comparison of Eqs. (3.2) and (2.11) shows that if there were no

topologically non-trivial solutions to the field equations, the anomalies would vanish along with the transition amplitude. But to the contrary, the anomalies are enacted by instantons and the amplitude survives. In the SU(2) case, this constraint governs that baryon and lepton numbers are separately violated by one unit each while baryon minus lepton symmetry is conserved. In the SU(3) case, the conservation of chirality is violated. Relating the winding number to the anomalous current shows that an instanton of winding number 1 is associated with a violation of axial charge by $2N_f$ units, where N_f is the number of active flavours [37].

3.4 Predictions for the LHC

Interest in instantons was renewed recently by the release of calculations of instanton event cross-sections at the LHC by Khoze *et al.* in Refs. [1, 51, 52]. They focussed on the amplitude for the reaction

$$g + g \to n_g \times g + \sum_{f=1}^{N_f} \left(q_{Rf} + \overline{q}_{Lf} \right), \tag{3.5}$$

where N_f denotes the number of quark flavours that are light compared to the mass of the instanton, L and R denote left and right-hand chirality, respectively, and n_g is the arbitrary number of gluons that fills the remainder of the phase space once the effective vertex is full of quarks.

Ref. [51] shows how 't Hooft's method of doing perturbation theory in a background of instantons results in an amplitude for the reaction above that looks like

$$\mathcal{A}_{2\to n_g+2N_f} = \int \mathrm{d}^4 x_0 \int_0^\infty \mathrm{d}\rho D(\rho) e^{-S_I} \prod_{i=1}^{n_g+2} A(p_i;\rho) \prod_{j=1}^{2N_f} \psi^{(0)}(p_j;\rho). \tag{3.6}$$

The integrals go over the centres and sizes of the BPST instantons, and are weighted by an instanton density factor $D(\rho)$ that depends on the sizes of the instantons.

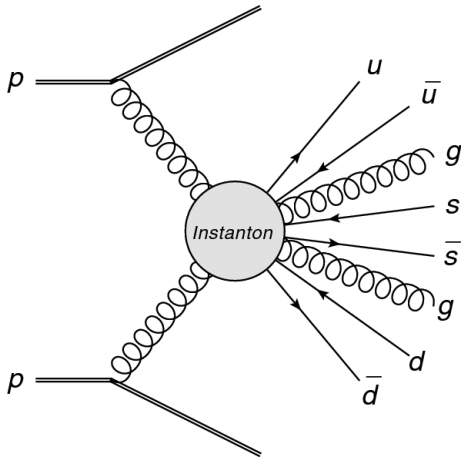


Figure 3.1: Pseudo-Feynman diagram showing the effective vertex of the instanton-induced event considered in this thesis. This is not formally a Feynman diagram since the blob at the vertex corresponding to the background instanton is a fundamentally non-perturbative object. Reproduced from Ref. [53].

The semi-classical action of the instanton enters only in the e^{-S_I} part, which can be regarded as a prefactor. The product of vector boson fields A and fermion zero modes $\psi^{(0)}$ corresponding to the external legs of the amplitude are calculated in the instanton background. Some labels have been dropped compared to Ref. [51] for brevity. This amplitude is evaluated numerically in Ref. [51], without resorting to saddle-point approximations as in previous attempts.

The total partonic cross-section $\hat{\sigma}_{\text{tot}}$ of the instanton-induced process $gg \to X$ is then obtained as a function of the partonic centre-of-mass energy, $\sqrt{\hat{s}}$, by way of the optical theorem.

The effective vertex shown as a reaction in Eq. (3.5) and as a scattering amplitude in Eq. (3.6) is shown as a pseudo-Feynman diagram in Figure 3.1. This is referred to as a *pseudo*-Feynman diagram because the blob at the vertex corresponding to the instanton background is a fundamentally non-perturbative feature, and does not formally appear in Feynman-diagram expansions. This figure shows two gluons coming from the colliding protons and connecting to the instanton field configuration at the effective vertex. The final state coming out of the vertex on the other side contains precisely N_f quark-antiquark pairs (one pair of each accessible flavour), and some number n_g of gluons.

The structure of the amplitude shown in Eq. (3.6) indicates the expected phenomenology of this process. The external legs factorise, which means they are independent of each other at the level of this amplitude, except for overall momentum conservation. As such, the final state should be spherically isotropic in the rest-frame of the instanton. Colour connections among the final state partons and mass effects modify this isotropy, as does the overlay of MPI.

The integral over ρ in Eq. (3.6) is infrared-divergent, which signals the breakdown of the calculation at large instanton sizes where the coupling becomes strong and the semiclassical approximation is invalidated. However, large-size instantons can be cut-off by considering interactions among the initial state partons ('hard–hard interactions'). The hard–hard interactions are provided by Mueller in Ref. [54], who shows that these corrections formally exponentiate. Mueller also explains in Ref. [54] how the extended instanton field configuration, which does not have a hard core, can absorb two high-energy gluons. The gluons can be expressed in terms of many gluons, each carrying a momentum $\sim \rho$, which are easily absorbed by the instanton whose characteristic length-scale is $1/\rho$. This is the argument that motivates the choice of $1/\rho$ for the factorisation and renormalisation scales, μ_{f} and μ_{r}, respectively [51].

Another choice made is to evaluate the strong coupling at the mass of the tau lepton, $\alpha_S(m_\tau)$, below some critical energy where the running prediction becomes unreliable [52]. The resulting values of the total partonic cross-section are shown in Table (3.1) along with some auxiliary information. These values are presumed to carry large and unknown uncertainties, but the experimental signature (3.5) is deemed to be robust [55, 56].

At low centre-of-mass energies where the cross-section for the instanton process is expected to be large, the experimental signature shown in Figure 3.1 is similar to the signature of a minimum-bias event. A minimum bias event can be thought of as an assembly of low-mass $2 \to 2$ scatters (MPI). These scatters are independent (except for overall momentum conservation and colour connections), so minimum-bias events

$\sqrt{\hat{s}}$ [GeV]	$\hat{\sigma}(gg \to X)$ [pb]	$\langle 1/\rho \rangle$ [GeV]	$\alpha_{\mathrm{S}}(1/\rho)$	$\langle N_{\mathrm{gluons}} \rangle$
20	2.01×10^{6}	1.69	0.327	7.81
25	9.49×10^{5}	1.98	0.306	8.58
30	4.64×10^{5}	2.27	0.290	9.07
35	2.32×10^{5}	2.52	0.279	9.61
40	1.25×10^{5}	2.84	0.267	9.67
50	3.89×10^{4}	3.38	0.251	10.56
60	1.38×10^{4}	3.87	0.241	10.89
70	5.45×10^{3}	4.33	0.232	11.38
80	2.36×10^{3}	4.85	0.224	11.67
90	1.08×10^{3}	5.24	0.219	12.31
100	5.44×10^{2}	5.82	0.213	12.10
110	2.79×10^{2}	6.21	0.209	12.62
120	1.53×10^{2}	6.71	0.205	12.77
130	8.56×10^{1}	7.13	0.201	13.04
140	4.99×10^{1}	7.57	0.198	13.25
150	3.01×10^{1}	8.00	0.195	13.45

Table 3.1: The partonic cross-section of the gluon-initiated instanton-induced process $gg \to X$ over a range of partonic centre of mass energies, $\sqrt{\hat{s}}$. Also shown is the calculated average number of gluons, $\langle N_{\mathrm{gluons}} \rangle$. For the renormalisation scale, μ_r, the inverse instanton size $\langle 1/\rho \rangle$ is used, a prescription that gives the values of strong coupling α_S shown in the second-to-last column. These values are all taken from Ref. [52].

will tend to mimic the spherical isotropy of the instanton vertex. However, the instanton vertex is distinguished by the higher particle multiplicity and spherical isotropy of the hardest scatter of the event (namely the instanton vertex).

Besides the distribution of the final state partons, the flavour content of the instanton process is also a distinguishing feature. It is less likely that a minimum-bias event will produce a quark-antiquark pair of each accessible flavour. In principle, these heavy flavours can be tagged and counted in an experimental setting. Failing that, the mass effects of the heavy flavours on the parton shower and hadronisation should leave an imprint on the final state radiation pattern that may be discernible.

Chapter 4

The ATLAS Experiment

The European Organisation for Nuclear Research (CERN) has a broad programme of High Energy Physics in which the Large Hadron Collider (LHC) is the flagship project. The LHC is a two-ring, superconducting synchrotron that accelerates particles to the highest energies attained in the laboratory to date. Along its circumference of 27 km under Swiss and French countryside near Geneva, the LHC collides particle beams at four interaction points (IPs) where experimental apparatuses are situated as shown in Figure 4.1. This thesis uses data from the ATLAS detector, which is situated at LHC Point 1.

We can connect the scattering theory of Chapter 2 to experiment by saying that the LHC prepares the asymptotic *in* state $|\beta; p_a, p_b\rangle$, and that ATLAS detects the *out* state $\langle\alpha; \{k_f\}|$. This lays out the broad requirements for these two systems: to access rare (low cross-section) events, the LHC must provide high-intensity beams, while to make good use of the collision events, ATLAS must be able to detect a variety of $\langle\alpha; \{k_f\}|$ states.

In this chapter, the LHC is briefly introduced in Section 4.1, before the ATLAS detector and its subsystems are introduced in Section 4.2. The reconstruction of the physics objects relevant to this thesis is described in Section 4.3.

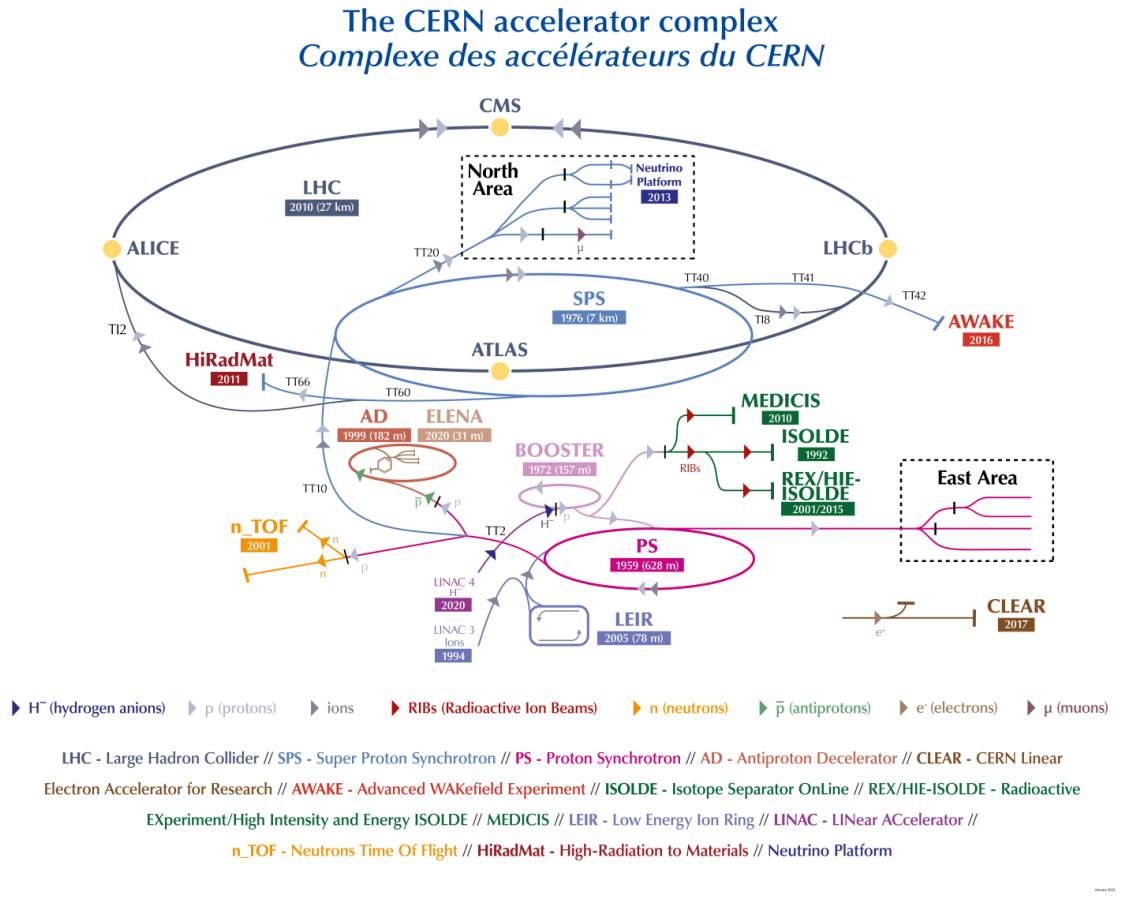


Figure 4.1: Layout of the CERN accelerator complex in 2022, taken from Ref. [57]. The four interaction points along the LHC are shown as yellow dots.

4.1 The Large Hadron Collider (LHC)

The LHC machine is described in Ref. [58]. From the point of view of the experiments, the LHC delivers stable beams at agreed-upon values of instantaneous luminosity, L. This measure of collision rate is defined in relation to the scattering cross-section σ and a corresponding number of events per second, N, through $N = L\,\sigma$.

The luminosity at an IP can be manipulated in terms of LHC beam parameters since these are related by

$$L = \frac{N_b^2 n_b f_\mathrm{r} \gamma}{4\pi \epsilon_n \beta^\star} F, \tag{4.1}$$

where N_b is the number of particles per each of the n_b bunches that constitute a beam. These particles circulate the ring with a revolution frequency, f_r, and relativistic Lorentz factor, γ. The kinematic spread of the beam particles is characterised by the normalised transverse emittance, ϵ_n. Related to particle emittance is the Courant-Snyder β-parameter (denoted $\beta^\star$ at the IP), which determines the transverse size of the beam. As beams collide, particles are lost and the luminosity is reduced in a way characterised by the geometric factor F.

The LHC was designed to provide two high-luminosity experiments, namely ATLAS [59] and CMS [60], with peak proton–proton (pp) luminosity values of $L = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

4.1.1 Lattice layout and high luminosity

The LHC lattice comprises eight straight sections and eight arcs [58]. The synchrotron functionality is accomplished in the arcs using NbTi superconducting dipoles cooled to a temperature of 2 K by superfluid helium. A straight section is about 528 m in length and serves for the insertion of utilities or as the site of an experiment. Beam capture, acceleration and storage is accomplished in one of the straight sections using 400 MHz superconducting radio frequency (RF) cavities. Strong focussing (control of the β-parameter) is carried out by triplets of quadrupoles connected in series. To

support the two counter-rotating same-sign particle beams, the LHC's beam pipes are twin-bore except at the sites of experiments, where the two beams pass through a common beam pipe.

High luminosity collisions at the sites of the ATLAS and CMS experiments are accomplished by configuring the surrounding quadrupole triplets to low-$\beta^\star$ values. Because the beams in these insertion regions occupy a common beam pipe, head-on collisions would cause a number of parasitic collisions along the insertion region. To prevent this, orbit bumps are performed on each side of the experiments, and the beams are kicked such that they cross at an angle.

In this configuration, only one bunch-crossing at a time takes place within each experiment, but an experiment still sees a number of collision events per bunch-crossing because of the particle multiplicity of the bunches.

4.1.2 Machine performance

The lifetime of an LHC fill is finite because the luminosity decays, primarily due to collisions at the interaction points [58]. LHC fills have luminosity lifetimes of around 15 hours, but beams are typically aborted earlier while L is still high to maximise the total integrated luminosity.

Run 2 of the LHC was conducted from 2015 to 2018 with proton beams at 6.5 TeV [61]. In this programme, beam during 2015 was used for re-commissioning after Long Shutdown 1 (LS1) following LHC Run 1. The first physics runs were taken in June with relaxed machine parameters, in particular with 50 ns bunch spacings (as opposed to the design 25 ns spacing), larger IP crossing angles, and $\beta^\star$ set to 80 cm. The LHC then moved to production mode in 2016, delivering 40, 50, and 65 fb^{-1} of pp collisions in 2016, 2017, and 2018, respectively.

The LHC implemented technical improvements throughout Run 2 and exceeded its

design performance, attaining instantaneous luminosities of up to $2 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, and delivering about $160\,\mathrm{fb}^{-1}$ of pp collisions to the ATLAS and CMS experiments.

4.2 The ATLAS Detector

The ATLAS detector (A Toroidal LHC ApparatuS) [59] was designed as a general-purpose detector to exploit the discovery potential of the LHC. The design process considered the search for the Higgs boson to be an important performance benchmark [62]. Without knowing in advance the precise mass of the Higgs boson, ATLAS had to cater sensitivity to a range of production modes and final states. As a result, the ATLAS detector is sensitive to an array of SM and BSM physics scenarios at the TeV scale and can support a diverse and creative research programme. In covering this programme, the ATLAS Collaboration has made over 1000 publications [63].

ATLAS uses a right-hand coordinate system with the origin at the centre of the detector, the x-axis pointing towards the centre of the LHC, the y-axis pointing upwards, and the z-axis running parallel to the beam axis. The $+z$ half of the detector is denoted the A-side, while the $-z$ half is denoted the C-side. Because of the detector's cylindrical symmetry, cylindrical coordinates (r, ϕ) are used in the transverse plane. Angles θ with respect to the $+z$-axis are expressed in terms of pseudorapidity, $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$. Pseudorapidity is useful because differences of pseudorapidity are invariant to Lorentz boosts along the beam axis, which means that particle production on average is uniform in η.

The ATLAS detector is shown schematically in Figure 4.2, where the different detector subsystems are indicated. It comprises a charged-particle tracker, electromagnetic and hadronic calorimeters, and muon spectrometers. In the following, these subsystems are described in detail proportional to their impact on this thesis. A principled review of particle detectors at accelerators can be found in Ref. [64].

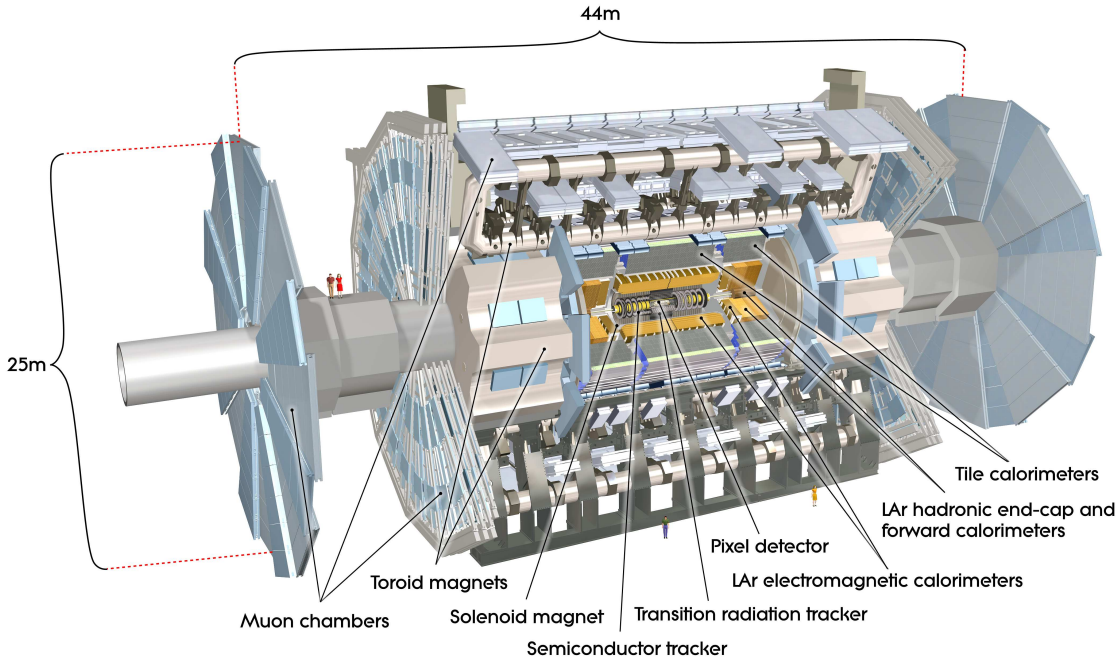


Figure 4.2: Cutaway view of the ATLAS detector. Reproduced from Ref. [59].

4.2.1 Inner Detector

The Inner Detector (ID) [65, 66], occupying the volume between the beam pipe and the calorimeters, is designed to measure charged-particle positions as inputs to charged-particle track reconstruction and vertexing. For these measurements to disentangle particles in the dense tracking environment of the LHC, the innermost radii of the ID are filled with high-granularity silicon pixel and microstrip sensors. The deployment of silicon is limited to these innermost radii because of its high material cost (in units of radiation length, X_0) and financial cost. Precision tracking is therefore continued to the remainder of the ID volume using a cheaper straw tube technology in the form of the transition radiation tracker (TRT). The layout of these ID subsystems is designed to balance the precision of the different detector technologies as shown in Figure 4.3. This whole structure is immersed in a 2 T magnetic field that deflects charged particles in the (r, ϕ)-plane to facilitate momentum determination.

The innermost subsystem of the ID is the Pixel detector, which instruments the vertex region with high-granularity silicon pixel sensors [68]. Modules of these pixel sensors are tiled onto three cylinders in the barrel region, and three disks in each

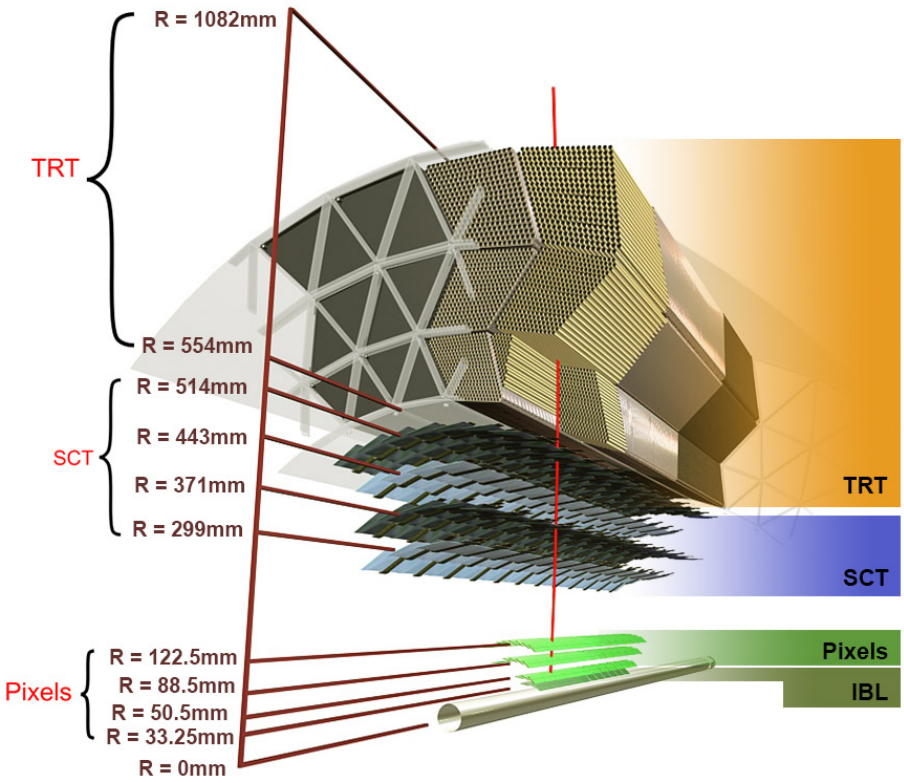


Figure 4.3: Illustration of the Inner Detector with emphasis on the radial dimensions of its subsystems, taken from Ref. [67]. (Ref. [67] reproduces this figure from Ref. [59], but includes the IBL.)

end-cap, which are arranged such that a track passes through at least three modules as it traverses outwards. Pixel sensors are 250 µm thick and have a pixel pitch of $50 \times 400\,\mu\text{m}^2$. The pixels are read-out in arrays of 18×160 rows and columns by chips that negotiate data transfer between the modules and the rest of ATLAS.

With the forecast of higher pile-up values after LHC improvements during LS1, ATLAS installed a new, smaller-radius beam pipe to accommodate an additional pixel layer called the Insertable B-Layer (IBL) closer to the IP [69, 70]. Closer proximity to the IP helps with impact parameter resolution and vertexing, and for mitigating the effects of radiation damage on the original Pixel layers, but brings higher precision requirements and challenges arising from higher particle fluences [71]. Pixels in the IBL are 200 µm-thick (230 µm-thick at high-$|\eta|$ [72]) and have a pitch of $50 \times 250\,\mu\text{m}^2$ [70].

The next layers of silicon constitute the semiconductor tracker (SCT). The SCT

instruments the intermediate radial range with modules of silicon microstrip sensors [73] tiled on four barrels and nine wheels per end-cap. Modules are made of back-to-pack assemblies of silicon wafers, rotated with respect to one another by a stereo-angle of 40 mrad. Two measurements of $r\phi$ on opposite sides of a module can be combined using the stereo-angle to measure the coordinate that runs perpendicular to the strips (z in the barrel and r in the end-caps). All of the modules have a similar structure, except that while barrel modules [74] are rectangular, end-cap modules [75] are trapezoidal in shape. The macro assembly is arranged such that a track passes through eight sensors as it traverses outwards, yielding up to four space-points.

What remains of the ID volume is used for the Transition Radiation Tracker (TRT) [65]. The TRT complements the precise measurements of the silicon detectors with continuous track-following to larger radii by instrumenting its volume with proportional drift tubes ('straws') filled with a Xe-based gas mixture [76]. The straws are close-packed in modules, and the modules are arranged in barrels [77] and end-caps [78] such that a particle traversing outwards passes through a large number (about 30) of straws. In the barrels, straws are oriented longitudinally, while straws are oriented radially in the end-caps. The precision of TRT position measurements is lower than the silicon's, but the TRT intervenes with a smaller material footprint and can furnish a larger volume due to its lower cost.

Additionally, the TRT contributes to electron identification capability by surrounding the straws with radiator material [76]. Highly-ionising particles emit transition radiation (TR) at the boundary with the radiator which can be detected in the straws. Tracking information and TR signals are obtained using low and high-threshold discriminators, respectively.

The design of the ID results in the intrinsic spatial resolutions listed in Table 4.1 and a basis for precision charged-particle track reconstruction.

Sub-system	Technology	Region	Spatial Resolution	Coverage in $\lvert\eta\rvert$
Pixel	Silicon pixels	Barrel	$12 \times 66\,\mu m^2$ [65]	< 2.5
		End-cap	$12 \times 77\,\mu m^2$ [65]	$1.7 - 2.5$
		IBL	$10 \times 67\,\mu m^2$ [79]	< 3.0
SCT	Silicon strips	Barrel	$16 \times 580\,\mu m^2$ [65]	< 1.5
		End-cap	$16 \times 580\,\mu m^2$ [65]	$1.4 - 2.5$
TRT	Straws	Barrel	$130\,\mu m$ [80]	< 1.0
		End-cap	$130\,\mu m$ [80]	$0.7 - 2.5$

Table 4.1: Intrinsic spatial resolutions of the three ID subsystems in the different regions, as obtained from beam tests or LHC collision data. In the spatial resolution column, the resolution is shown in the format $r\phi \times z$ (r) for the barrel (end-caps), except in the case of the TRT, where the spatial resolution $r\phi$ (z) is given per straw. The pseudorapidity acceptance of each subsystem is shown in the last column.

4.2.2 Calorimeters

The calorimeter system serves to provide measurements of electromagnetic activity, jets, and missing transverse energy [81]. It is subdivided into an electromagnetic calorimeter (ECAL) [82] and a hadronic calorimeter (HCAL) [83]. The ECAL is designed to instigate and to capture the whole of an electromagnetic shower, and does so using lead absorber and liquid Argon (LAr) active medium interleaved in an accordion geometry. The HCAL does similarly for hadronic showers using tiles of sandwiched steel absorbers and plastic scintillators. The geometric depth of these systems is determined by the radiation and interaction lengths of the ECAL and HCAL materials, respectively [64]. One hadronic interaction length is greater than an electromagnetic radiation length, so the HCAL is larger and is put behind the ECAL. The precise technology varies with detector region, but all together the calorimeter system has acceptance up to $|\eta| < 4.9$.

The calorimeters are segmented such that measurements are made by 'cells' comprising towers in $\Delta\eta \times \Delta\phi$. The deposits of energy across these cells are used to reconstruct shower shapes and to identify the objects that caused them.

Positioned on the inner faces of the end-cap calorimeter cryostats at $z = \pm 3.56\,m$ are the Minimum Bias Trigger Scintillators (MBTS) [84]. Each of the two MBTS

disks comprises two rings of eight independent azimuthal sectors that together cover the pseudorapidity range $2.09 < |\eta| < 3.84$. The MBTS are essentially hodoscopes and are used primarily for event counting and triggering.

4.2.3 Muon Spectrometers

The muon spectrometers [85] are not used in this thesis. These instrument the $|\eta| < 2.7$ region of the outermost radii with trackers for measuring the minimum-ionising particles that survive the inner detector and the calorimeters. The muon spectrometers are interleaved with the superconducting air-core toroidal magnets that lend their name to ATLAS.

4.2.4 Trigger and Data Acquisition (TDAQ)

At nominal bunch spacings of 25 ns between pp bunch-crossings, collisions occur at up to 40 MHz. This rate exceeds the physical upper-limit on the detector's read-out rate. ATLAS uses a two-level triggering system to select out collision events of interest [86]. The Level 1 (L1) trigger is implemented in custom hardware logic, and is operated to reduce the data rate 40 MHz $\rightarrow$ 100 kHz. The L1 trigger is followed by the High-Level Trigger (HLT) implemented in software which applies a fast reconstruction of an event (thereby defining 'online' physics objects) for selection. The HLT is operated such that the data rate is further reduced 100 kHz $\rightarrow$ 1 kHz (corresponding to an average throughput to permanent storage of about 1.2 GB/s [86]).

The L1 trigger decision is formed by the Central Trigger Processor (CTP), which receives information from all participating detector subsystems, including the MBTS. Upon an L1 accept decision, the read-out of all sub-detectors is triggered and their data are sent to buffers. The HLT running on a computing farm next to the detector requests data fragments from the buffers as required in the course of running feature

extraction and hypothesis algorithms. When the HLT accepts an event, the event is sent to permanent storage and offline reconstruction [87].

ATLAS data are siphoned into *streams* defined by trigger menus and prescales. A trigger menu serves to bias the data collection towards physics processes of interest, typically those residing further to the right along the horizontal axis of Figure 2.2, which are rarer in LHC collisions. The trigger menu in a particular run represents the ATLAS Collaboration's physics interest in that run. Trigger rates are gauged using prescales. A prescale value P attached to trigger i is an instruction for the detector to read out up to $1/P$ of the events firing trigger i. The experimental environment evolves continually, and the trigger prescale menu may be changed by an operator to reflect this. A luminosity block (LB) is defined as a period of about one minute over which the detector conditions are constant.

Over the course of Run 2, the ATLAS detector collected $147\,\text{fb}^{-1}$ of pp data at 13 TeV with a data quality efficiency of 99.7% [88]. The pile-up spectrum of this dataset, parsed into data-taking years, is shown in Figure 4.4. The variability in this spectrum attests to the experimental challenge faced by ATLAS. After imposing data quality requirements, a total of $140.1\,\text{fb}^{-1}$ of this dataset is good for physics [89].

4.3 Event Reconstruction

The ATLAS detector responds to a collision event by producing a large amount of data. For this information to be useful, it is best converted into the particle properties (momentum four-vectors, electric charges, ...) that it represents. This conversion of format is called event reconstruction. In this section, the reconstruction of the relevant parts of an event, namely of charged-particle tracks, primary vertices, and calorimeter jets, is described.

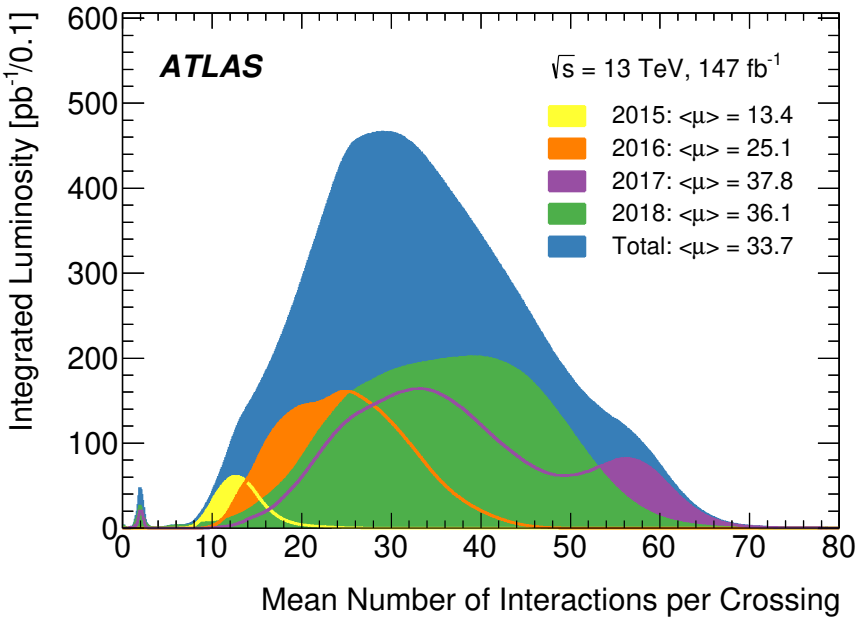


Figure 4.4: Distribution of the mean number of interactions per pp bunch-crossing, μ, over the data-taking years of LHC Run 2. Reproduced from Ref. [88].

4.3.1 Tracks

The basic physics objects used in this thesis are charged-particle tracks. Their reconstruction is a combinatorial joining-of-the-dots problem that takes ID position measurements as input. ATLAS solves this problem using an iterative inside-out method followed by a subsequent outside-in step as described in Ref. [90].

Inside-out tracking

Inside-out track reconstruction starts with the creation of three-dimensional space-point objects from silicon measurements. In this superset of space-points, tracks are seeded using every possible combination of two Pixel detector space-points. This huge number of seeds can be whittled down immediately using a beam-spot constraint supplied by a fast primary vertex search.

These seeds provide enough directional information to instigate the extending of the tracks. The pointing information is used to define 'roads' of encompassing detector

elements, wherein a Kalman fitter-smoother formalism [91] is applied. The Kalman formalism works locally — it looks forwards by predicting the representation of a track on the next surface, and upon finding a space-point in the predicted vicinity, refits the track parameters. Thus it incorporates local pattern recognition and track fitting into one step. Outlier hits are identified along the way as hits making large contributions to the track's goodness of fit χ^2.

Ambiguities arise e.g. due to hits being assigned to multiple tracks ('shared hits'), and fakes (tracks being reconstructed from combinations of silicon hits not due to a single particle). Such ambiguities are resolved by assigning scores to the track candidates, and taking only those tracks whose scores exceed a quality threshold. The scoring depends on the detector geometry and relative precisions of the different sub-detectors. Hits in two geometrically-overlapping sensors are preferred, while tracks containing holes (missing hits) are penalised. A hole in the IBL is strongly penalised.

The resulting track seeds are extended along their pointing direction into the TRT. Tracks of interest, with transverse momenta p_T above about 500 MeV, are seen as approximately straight line segments in the TRT, so the incorporation of TRT hits proceeds by fitting straight lines. This step begins with the silicon part of the track fixed. Only when a set of corresponding TRT hits has been identified is the track refitted as a whole, and its parameters updated.

Outside-in tracking

The inside-out method is liable to miss a track if it is only partially seen in the silicon. This type of track may emerge from highly-displaced secondary vertices, as in the event of photon conversion or K_s decay, or if the underlying particle suffers such substantial energy loss that it does not induce hits in the outer layers of the silicon, thus sending the TRT extension in the wrong direction. The inside-out step

is therefore succeeded by an outside-in step.

This back-tracking works by firstly assembling track segments in the TRT. This is accomplished by performing a Hough transformation [92] of TRT space-points (from r–ϕ coordinates in the barrel and z–ϕ in the end-caps) and extracting the parameters of the straight line-fit from the maximum of the hit distribution in Hough space. These TRT track segments are then extended into the silicon, where they may be combined with existing silicon track segments, or where they can incorporate new, previously unused space-points.

Track selection working points

Once tracks are reconstructed, track selections are applied according to analysis requirements. Selections can be made on numbers of hits and holes in the different sub-detectors, the quality of the track in terms of χ^2, the measured p_T, and track impact parameters with respect to the beam-spot.

ATLAS uses two basic selection working points (WPs) called `Loose` and `TightPrimary`. These selections are defined to satisfy certain efficiency and fake-rate criteria, two important metrics that broadly speaking come at the expense of one another. Their definitions in terms of track parameters, as well as their efficiencies in the central and forward regions, are summarised in Table 4.2. The `Loose` WP represents the track reconstruction requirements, while the `TightPrimary` WP sacrifices some efficiency for significant fake rate reduction by tightening the requirements on the numbers of silicon hits [93].

4.3.2 Primary Vertices

The site of a collision within a bunch-crossing is the source of physics objects seen in the detector. This site is called a primary vertex (PV), and can be reconstructed

Parameter	`Loose`	`TightPrimary`
Maximum η	2.5	2.5
Minimum Silicon Hits	7	9
Maximum Silicon Shared Hits	1	1
Maximum Silicon Holes	2	2
Maximum Pixel Holes	1	0
Minimum η for Strict Minimum Silicon Hits	–	1.65
Minimum Silicon Hits Above η Cut-Off	–	11
Minimum Number Of Both Innermost Layer Hits	0	1
Efficiency in $\lvert\eta\rvert < 0.1$	90%	84%
Efficiency in $\lvert\eta\rvert > 2.3$	71%	62%

Table 4.2: Summary of the `Loose` and `TightPrimary` track selection working points (WPs). The working points consist of an acceptance cut on η, and requirements on the numbers of hits associated to the tracks. The Pixel and SCT sub-systems are jointly referred to as silicon.

from the set of reconstructed tracks. ATLAS uses a *finding-then-fitting* approach to PV reconstruction, as described in Refs. [94] and [95] and briefly summarised here.

The process begins with vertex finding. The first vertex is taken to reside at the mode of the z-coordinate distribution of the reconstructed tracks at their point of closest approach to the beam-line. Based on this vertex seed, an iterative vertex fitting procedure with temperature annealing is performed to minimise the χ^2 of the vertex position while optimising the selection of tracks to associate to it. Tracks incompatible with this first vertex are taken as input to another iteration of the algorithm to find additional PVs.

By default, it is the vertex carrying the greatest sum of squared track transverse momenta, $\sum p_{\mathrm{T}}^2$, that is used for physics. This choice is made because this high-p_{T} vertex is the most likely to correspond to a high-p_{T} physics process of interest. Confusingly, the highest-$\sum p_{\mathrm{T}}^2$ vertex is henceforth called *the* PV following ATLAS convention. The other PVs are regarded as pile-up vertices, nuisances that contaminate the reconstruction of the high-p_{T} process of interest.

4.3.3 Calorimeter jets

The calorimeter showers resulting from the fragmentation of coloured partons are interpreted as jets. This thesis considers particle-flow jets [96]. These are reconstructed from a combination of inner detector and calorimeter information called particle-flow objects, clustered using the anti-k_t algorithm [97] as implemented in the FastJet package [98]. Reconstructed jets may be flavour-tagged [99, 100] using neural networks such as DL1r, which take variables sensitive to the presence of heavy flavour decays (such as secondary vertex coordinates and track impact parameters) as input.

Chapter 5

Search Design

The labour we delight in physics pain.

— Macbeth, Act II, Scene 3, Shakespeare

The search for QCD instantons presented in this thesis is made possible by the calculations of Ref. [51], and the simulation of those instanton events implemented in the SHERPA event generator.

The design of the analysis is challenging for several reasons. The first is the non-resonant character of the instanton. In an ATLAS search for a new physics process, the target physics model often involves some kind of resonance. A search can be formulated in this case as a 'bump-hunt' in some spectrum. The discovery searches for the SM Higgs boson, for example, entailed bump-hunts in four-lepton invariant mass spectra [101, 102]. In contrast, the non-resonance of instanton events means that they tend to spread out in the phase-space of the discriminating variables as will be shown, such that the signal is not localised in any single variable. Thus no distinct bump emerges, and instead instanton events will have the effect of raising entire spectra while changing their shapes.

Secondly, instanton events are kinematically similar to the background minimum

bias events. This aspect combined with the non-resonance motivates us to employ a Machine Learning (ML) technique to exploit higher-dimensional correlations among the discriminating variables than can be inferred from low-dimensional histograms for signal discrimination.

Thirdly, phenomenological models introduced in Section 5.1 are used for the simulation of the minimum bias background. The use of these models potentially introduces a tuning problem. Instead of physical parameters derived from first principles, these models are constructed in terms of free 'phenomenological' parameters that must be tuned to the target data in order to obtain a good description. In principle, the target data include instantons, so it is quite possible that any instantons in the data will have already been attributed to the background by redefinition of the model parameters. It will be emphasised that this tuning-away of instantons ought to be a minimal effect, which is mitigated anyway as part of the analysis design.

The performance of these background models is relatively poor, but they are used for training Deep Neural Networks (DNNs) as part of the ML technique nevertheless. It is therefore important to justify this usage, which is done *a posteriori* in Section 5.3.9. Moreover, the background models are not used directly for the background estimation. Instead, a data-driven 'ABCD'-style background estimation strategy is used. The selected ML technique was chosen with the desire for data-driven background estimation a central consideration.

To avoid complications arising from the overlay of pileup events, which would have the effect of smearing our event-level discriminating variables, we follow the suggestion of Ref. [51] of using a low-pileup dataset. In this dataset, events of interest are triggered using a so-called minimum bias trigger based on the MBTS system. The ATLAS detector is not equipped with a dedicated instanton trigger, so the use of the MBTS is a fallback. The minimum bias triggers are physics-agnostic, and effectively select high cross-section events at low rates.

For the interpretation of the results, the analysis is performed in bins of reconstructed instanton mass. This was a principled choice made to somewhat decouple the analysis from the modelling of the instanton mass spectrum, and to separate instanton events into kinematic regimes which can be different. With some experience of training the DNNs used in the ML part of the analysis, this mass binning became also a practical advantage. In the training dataset, signal events are assigned weights proportional to their cross-section to indicate to the networks that some events are expected to be seen more often than others. However, the signal cross-section falls through almost six orders of magnitude, and training a network over this range of event weights proved difficult. Restricting the training to mass bins eases this issue.

To be precise, three mass bins are defined and are used to separate events into low, medium, and high-mass categories. This enables us to make three separate statements, and to increase our sensitivity to QCD instantons through a statistical combination of the three categories. This statistical treatment is afforded its own chapter, and is presented in Chapter 6.

The analysis design is elaborated in the remainder of this chapter. Section 5.1 begins with the details of the dataset and the MC samples used. In Section 5.2, the discriminating variables used for signal classification and background estimation are presented, before the ML technique is introduced in Section 5.3. The mass binning depends on the details of the ML performance, and is therefore described in Section 5.3.3.

5.1 Data and Monte Carlo Samples

The analysis uses data from six physics runs taken in June 2015 as part of Run 2 commissioning. In these runs, pp collisions were conducted at 13 TeV under low-pileup LHC conditions established with reduced beam currents and a low $\beta^\star$ value. Good Luminosity Blocks (LBs) are selected based on a Good Runs List (GRL),

which specified when the ATLAS detector was fully operational. Event-cleaning is performed by requiring event-wise that no sub-system has reported an error state.

From this dataset, events are collected using the one-armed Minimum Bias trigger `HLT_noalg_mb_L1MBTS_1`. Requiring this trigger corresponds to requiring at least one count above threshold in any of the 32 independent MBTS sectors. The resulting dataset amounts to $1.693\,\mathrm{nb}^{-1}$ of recorded integrated luminosity. A breakdown of how this integrated luminosity was distributed over the constituent runs is shown in Table 5.1.

Since the trigger is prescaled at the average values P shown in Table 5.1, the effective luminosity sampled by the trigger is actually a factor of P higher than the amount recorded. It is thus possible to 'unprescale' the data by multiplying each event by the prescale of the trigger in that event, but this is not done.

The efficiency of `HLT_noalg_mb_L1MBTS_1` in this dataset has been studied in previous analyses using a random control trigger [103, 104]. Starting at 98.8 % in events with only one selected track, this efficiency rapidly rises to 100 % for events with more than five selected tracks.

Another part of the motivation for using `HLT_noalg_mb_L1MBTS_1` is that it is well-understood based on an analysis of charged-particles [103]. Therein a prescription was put forth for quantifying the systematic uncertainty on its efficiency. Two sources are identified: one involves the track selection criteria, while the other involves the difference in efficiency between sides A and C of the Minimum Bias Trigger Scintillator (MBTS) system. Adding these sources in quadrature produces a systematic uncertainty of 0.12 %, which is negligible for this analysis and we do not reproduce it.

Run Number	Del. Lumi. [μb^{-1}]	PS Lumi., L [μb^{-1}]	Average PS, P	Mean $\langle\mu\rangle$
267358	81.1	72.3	1.12	2.4×10^{-3}
267359	240.7	108.9	2.21	6.0×10^{-3}
267360	320.9	76.7	4.18	1.6×10^{-1}
276367	1 538.2	124.9	12.32	3.4×10^{-2}
267385	5 932.2	707.2	8.39	3.0×10^{-2}
267599	5 609.6	602.6	9.31	3.2×10^{-2}
Total	13 722.7	1 692.7	8.11	3.5×10^{-2}

Table 5.1: Summary of the integrated luminosities collected across the runs considered. The 'Delivered Lumi.' column shows the integrated luminosities after live-fraction and LAr Veto corrections. The prescaled (PS) luminosity is the integrated luminosiy collected by the `HLT_noalg_mb_L1MBTS_1` trigger of interest, which was configured at the average prescales shown in the right-most column. The last column shows the average number of pileup interactions per bunch crossing $\langle\mu\rangle$ per run. The mean $\langle\mu\rangle$ shown in the bottom row is calculated as the mean over the runs weighted by the number of good LBs in each run.

5.1.1 Signal modelling

The signal event generation was implemented in SHERPA [15, 105] in support of the phenomenology study of Ref. [1] (a precursor to Ref. [51]). This implementation was done in SHERPA 3 which was at the time in it's alpha stage of pre-release. For our event generation, tag `v3.0.0beta1` was used.

As described in detail in Ref. [1], SHERPA takes a table (in the same format as Table 3.1) of $\hat{\sigma}$ values across a range of s' values as input, and interpolates linearly between those listed values to obtain $\hat{\sigma}$ as a function of s'. The partonic cross-section is convolved with PDFs to produce the hadronic cross-section as a function of instanton mass (which is synonymous with $\sqrt{s'}$). According to this mass spectrum, the final state of partons is then assembled algorithmically. Quark-antiquark pairs are added in increasing order of mass until the mass of the instanton is reached. The number of gluons to produce in association is chosen from a Poisson distribution with mean $N_{\text{gluons}}(s')$. The phase-space for final state partons is generated using the massive-RAMBO algorithm[1] [106], which produces rest frame-isotropic decays. Colours are randomly distributed over the partons participating in the hard scatter before they are showered. For the factorisation scale, μ_f, three choices are considered. These are

[1] *RAndom Momenta Beautifully Organised.*

(i) the $1/\rho$ value interpolated from the data table (the default), (ii) $\sqrt{s'}/N_{\text{gluons}}$, and (iii) $\sqrt{s'}$.[2] Differences among these scale choices, particularly between (i) and (ii), are considered as contributions to the systematic uncertainties in Section 6.4.

A choice is made in relation to the flavour scheme employed in event generation. The run card has three relevant parameters for steering this scheme.

- `INSTANTON_INCLUDE_QUARKS` 5
- `INSTANTON_B_PRODUCTION_THRESHOLD` 100 GeV
- `INSTANTON_C_PRODUCTION_THRESHOLD` 20 GeV

These set the maximum number of quark flavours allowed in the flavour production, and the thresholds in instanton mass above which the decay channels into b and c quarks open, respectively. A wavelength argument is made in Ref. [56] for the default values shown. The argument is based on the observation that the inverse instanton size is significantly smaller than the energy of the system, and the idea that below these thresholds, the instanton in the gluon field itself has too long a wavelength to resolve the heavy flavour fields. This treatment is heuristic, but is a step in the direction of ensuring that the generated flavour production is consistent with the integrals in the calculation of the cross-section in Ref. [51].

Ahead of this latest version of SHERPA, adjustments had been made to the implementation of the Underlying Event (UE) models [15]. A comprehensive re-tuning of the phenomenological parameters of these models was necessitated in order to accommodate these adjustments. The formal re-tuning was conducted by this author using the PROFESSOR toolkit [107], but eventually a best-guess set of tune parameters supplied by SHERPA authors [108, 109] was found to suffice for the analysis. The

[2] Originally, the alternative scale choice $\sqrt{s'}/N_{\text{gluons}}$ was used, but was found to suffer a bug in SHERPA. This bug resulted in a factorisation scale that was significantly higher than the $1/\rho$ choice, causing a hyper-active underlying event, excessively hard final state radiation, and overall our signal distributions to look very minimum bias background-like. This bug has since been fixed, and the $\sqrt{s'}/N_{\text{gluons}}$ choice now looks like a sensible alternative to $1/\rho$.

tuning project is documented in Appendix A.

The number of points provided in Table 3.1 is not sufficient to produce a smooth mass spectrum by linear interpolation. This is a problem since any artificial jaggedness in the mass spectrum may manifest as features in kinematic distributions. For smoothing the spectrum, the data-table is parameterised in terms of arbitrary polynomials as shown in Figure 5.1, from which 100 points are sampled for the event generation's run card.

Our run card describing the instanton-specific parts of the event generation is augmented with ATLAS's standard SHERPA base fragment. This employs the NNPDF3.0 NNLO [110] set of parton distribution functions at $\alpha_S = 0.118$ for the nominal PDF set, and runs a standard set of on-the-fly weight variations discussed in Section 6.4 for studying systematic uncertainties on the theory modelling.

Because the cross-section falls so steeply with increasing $\sqrt{s'}$, signal events are generated in truth-mass slices. We choose to produce events in six mass slices

$$\{(20,\,40),\,(40,\,60),\,(60,\,80),\,(80,\,100),\,(100,\,120),\,(120,\,150)\}\ \text{GeV},$$

which we label `m020`, `m040`, `m060`, `m080`, `m100`, `m120`, respectively. Basic properties of the generated instanton events are validated with a Rivet routine, a picture of which is given in Figure 5.2. It can be seen in the mass distribution that the stiching of the truth-mass slices works as intended, and that the run card containing the interpolated cross-section works. The particle multiplicity and total scalar p_T distributions walk to higher values as the truth mass increases, while the leading particle becomes harder. The structure in the rapidity distribution has been shown to arise from the way that instanton events hit the PDFs.

In this analysis, we only consider the gluon-induced production mode. In principle, one should also consider the quark, anti-quark, quark-gluon, as well as multi-parton production modes. Changing to these production modes simply involves rearranging

the partons connected to the effective vertex shown in Figure 3.1, since the content of this vertex is fixed. As such, if the initial state combination of partons changes, then so does the final state set of partons, and there can be no interference between the final states of the different production modes in the cross-section calculation. Thus the total cross-section comes from adding the individual cross-sections of the different production modes.

Regarding the kinematics of the different modes, it can be seen in the amplitude of Eq. (3.6) that most of the amplitude, particularly its exponent, stays the same. Therefore it is not expected that the final state kinematics will differ drastically between the modes, and the isotropic signature that we target is general enough. In the end, it should suffice to consider only the gluon-initiated mode because the gluon PDF dominates at the relevant values of Björken-x.

As for multiparton initial states, the scenario in which two partons, one from each proton, splits into multiple partons is already accounted for by Mueller's hard–hard interactions term. The scenario in which multiple partons originate from within the protons enter the effective vertex is similar to that concerning the different production modes. This scenario simply involves using multiple PDFs, and will make a manifestly positive contribution to the total cross-section not affecting final state kinematics.

5.1.2 Background modelling

For the nominal background modelling, we choose to use the EPOS LHC event generator [18] based on its competitive performance in Ref. [103] and elsewhere. EPOS[3] is a parton-based Gribov-Regge theory that calculates particle production and cross-sections simultaneously in one consistent formalism [31, 111]. Here multi-parton interactions are implicitly included in the ladder formalism, and diffractive processes

[3] *Energy-conserving quantum mechanical approach, based on Partons, parton ladders, strings, Off-shell remnants, and Splitting of parton ladders.*

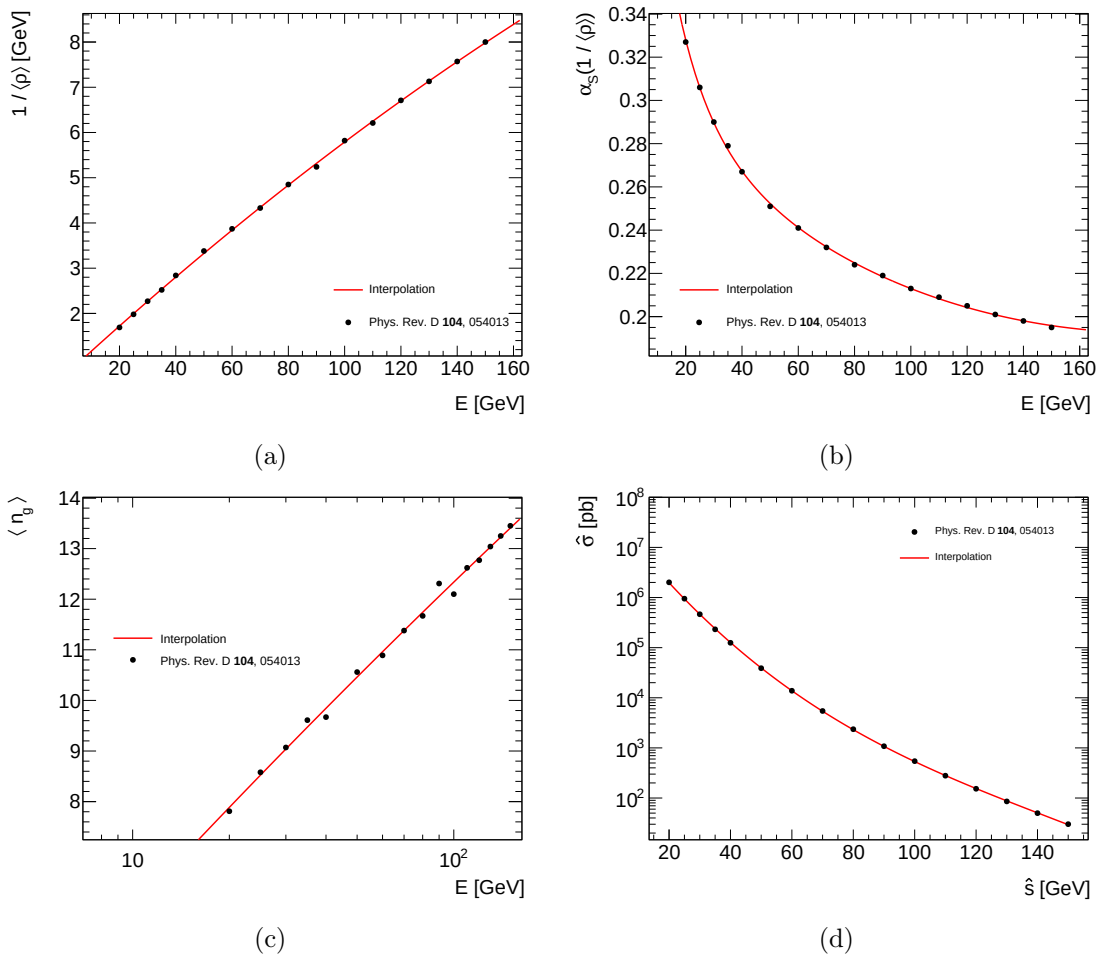


Figure 5.1: Parameterisations of the quantities in the four columns of Table 3.1. These are the parameterisations used to interpolate these quantities as functions of partonic centre-of-mass energy for generating a smoother instanton mass spectrum (as well as for smoothing the ancillary quantities $\langle 1/\rho \rangle$, $\alpha_S(1/\rho)$, and $\langle N_{\text{gluons}} \rangle$).

are realised as closed ladders [31]. EPOS calculates its own PDFs in that same formalism, with external constraints coming through its tuned parameters only.

For the hadronisation part of the event, EPOS LHC uses a model of string fragmentation supplemented by a model of collective hadronisation for regions of higher string segment density, as described in Section 2.2. (The hydrodynamic model of collective hadronisation used to describe the 'core' of an event, for which EPOS is often known, is not needed in EPOS LHC.) Since the EPOS event generator focusses primarily on the description of overall particle production and spectra, heavy flavours are not explicitly treated.

As an alternative background model, we use PYTHIA 8 to generate dijet events [16].

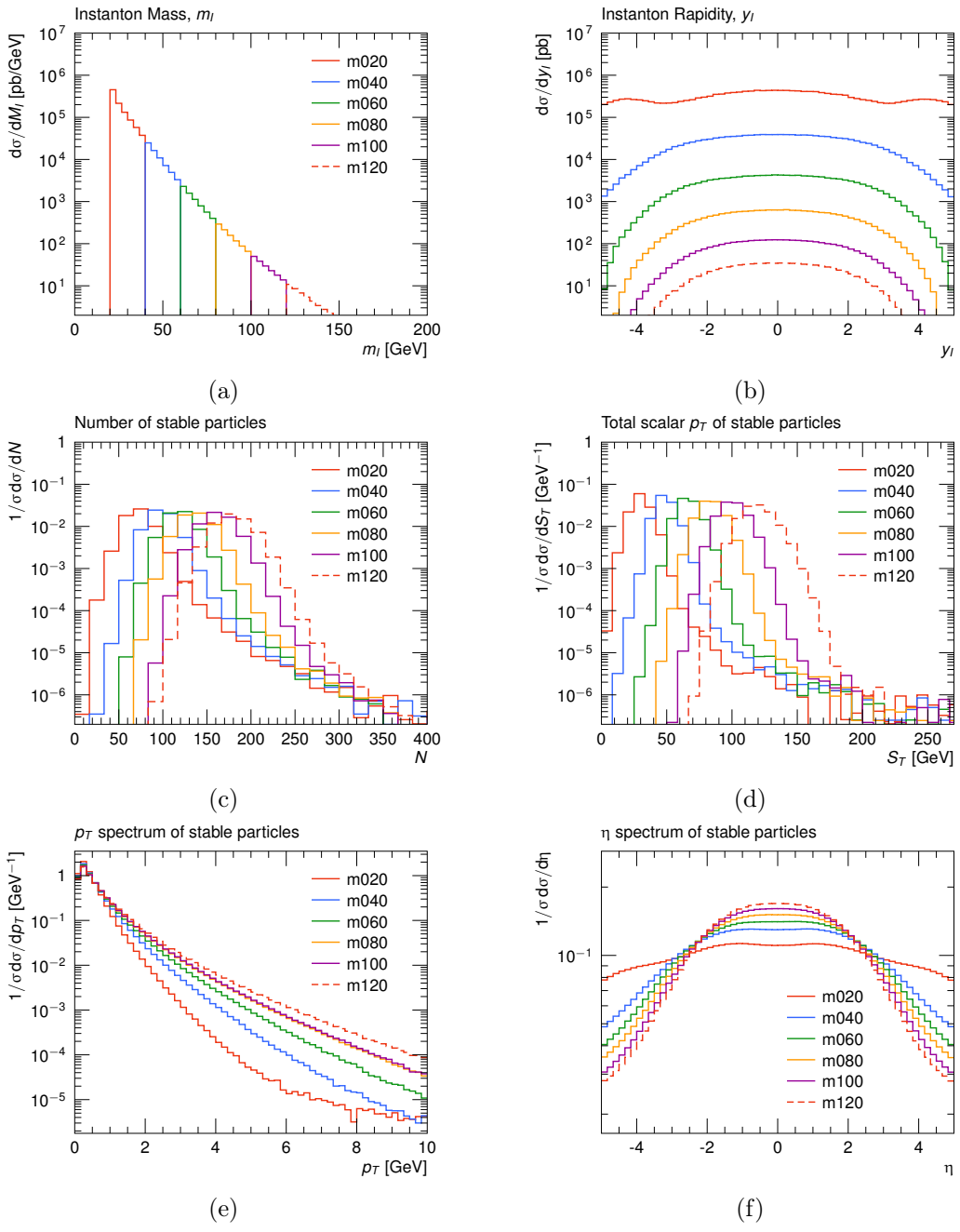


Figure 5.2: Plots from the Rivet routine used to validate the instanton event generation of SHERPA 3. Properties of the instanton itself are shown in the first row, and properties of its stable decay products are shown in the last two rows. Each colour shows events from a different mass slice. The mass distribution of the first plot shows that the interpolations based on Figure 5.1 work as desired, and that the instanton mass slices are produced and combined correctly. The events are found to follow the correct trend of decreasing cross-section with increasing mass.

The configuration for generating these events in ATLAS is described in Ref. [112], and we use the 'JZX' scheme of slicing by leading jet-p_{T}. The dijet part of the event is calculated from matrix element, and is augmented with multi-parton interactions based on the Sjöstrand–van Zijl model [29, 30]. The A14 tune [113] produced by ATLAS is used for the shower and MPI parameters. Of all the JZX samples, we take the lowest three slices 0, 1, and 2, which suffice to describe the data. The JZ0 slice consits of Minimum Bias events only. These ATLAS-common samples ship with all the standard systematic theory variations of scale, PDF, and eigentune, which we consider in Section 6.4.

The phenomenological parts of both EPOS LHC and the PYTHIA 8 are tuned directly to data. As such, it is possible that any instanton signal component present in the data may have been 'tuned away' or absorbed into these background models by redefinitions of the tunable parameters. This is unlikely, since the predicted instanton signal cross-section is about 10^5 times smaller than the total background cross-section. A signal component in the dataset being tuned to is therefore not expected to be visible above the uncertainties.

5.1.3 Event and track preselection

Events are preselected from these samples for compatibility between data and MC, and to avoid pathologies in the discriminating variables that will be exploited. The preselection is taken from Ref. [103]. An event must contain one primary vertex, and is allowed a second PV only if that second pp interaction has fewer than four associated tracks. This latter selection is designed to remove contributions from pileup vertices while keeping events in which the first PV is split by the reconstruction into two separate PVs [103].

With foresight of the event-wise variables that will be calculated from the set of reconstructed tracks in Section 5.2, it is decided that a low fake-rate is the goal

of the track selection. The idea behind this is to preserve the signal properties through the track reconstruction. Fake tracks are an adverse presence in an event because their arbitrary properties will tend to smear distributions and diminish signal discrimination power. Since the `TightPrimary` WP provides reasonably high efficiency while reducing the fake-rate [93], this WP is chosen. Further requiring that all tracks have $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ aids the fake-rate reduction [114] and additionally reduces the systematic uncertainties incurred. At least two tracks with $p_{\mathrm{T}} > 100\,\mathrm{MeV}$ are required for the PV reconstruction, but we additionally require two $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ tracks.

5.2 Discriminating Variables

Since track reconstruction efficiency is so high, track reconstruction produces a lot of information about the stable charged particles produced in a collision event. Assuming that every hadron produced promptly at the IP is a pion ($\pi^{\pm}$ or π^{0}) [64], the set of reconstructed tracks carries about two thirds of the kinematic information about an event, and can be used for signal discrimination. In this section, three classes of discriminating variables are constructed from tracks. These variables all target the isotropy and mass spectrum of instanton events. The analysis that results is entirely tracks-based, which paves the way for a simple analysis that incurs relatively few and small experimental systematic uncertainties.

5.2.1 Event kinematics

To represent event kinematics, the variables listed in Table 5.2 are assembled. These are parsed into two groups. The first group (starting with track multiplicity) is used to target differences of shape between the mass spectra of the signal and background samples. This is reflected in the fact that these variables have dimensions of energy

Variable	Definition
Track multiplicity	$N_{\text{trk}} = \lvert\{p_i\}\rvert$
Scalar sum of p_{T}	$S_T = \sum_i p_{iT}$
Leading track-p_{T}	$p_T^{\text{lead}} = \max p_T$
Number of high-p_{T} tracks	$\lvert\{p_{\text{T}} \mid p_{\text{T}} > 1\,\text{GeV}\}\rvert$
Number of low-p_{T} tracks	$\lvert\{p_{\text{T}} \mid p_{\text{T}} < 1\,\text{GeV}\}\rvert$
Mean p_{T}	$\langle p_{\text{T}} \rangle$
Median p_{T}	med p_{T}
Transverse mass	$m_T = \sqrt{m^2 + p_x^2 + p_y^2}$
Transverse mass per track	m_T / N_{trk}
Mean η	$\langle \eta \rangle$
Standard deviation of ϕ	$\sigma(\phi)$
Standard deviation of η	$\sigma(\eta)$
Standard deviation of $\Delta\phi(i,j)$ of all track pairs	$\sigma(\Delta\phi(i,j))$
Standard deviation of $\Delta\eta(i,j)$ of all track pairs	$\sigma(\Delta\eta(i,j))$
Mean of $\lvert\Delta\eta(i,j)\rvert$ of all track pairs	$\langle\lvert\Delta\eta(i,j)\rvert\rangle$

Table 5.2: List of the discriminating variables constructed to represent event kinematics, along with their rough mathematical definitions.

(except track multiplicity, which is a correlate of mass). The second group (starting with mean η) comprises angular variables assembled to target event isotropy.

5.2.2 Event Shapes

Event Shapes are a category of global collider event observables designed to quantify the geometry of energy flow in the final state [115]. Event Shapes are routinely measured in ATLAS to evaluate modelling and reconstruction performances, of jets in particular [116]. They are used here as a means to quantify the spherical isotropy expected of instanton events, and are discussed in two separate classes: the class based on the thrust axis, and the class based on the sphericity tensor. In what follows, each event is characterised by a set X of four-momentum vectors $x_i = \mathbf{x}_i$, where $i = 1, 2, \ldots, n$ index the n vectors of the event.

Thrust is defined with respect to a thrust axis, e_T, according to

$$T(X) = \max_{e_T} \frac{\sum_i |e_T \cdot x_i|}{\sum_i |x_i|}. \tag{5.1}$$

This axis aligns with the direction of greatest energy flow in X. In the plane perpendicular to the thrust axis, Eq. 5.1 can be applied again to obtain the *thrust major axis* and its value T_{major}. Then orthogonal to both the thrust and thrust major axes is the *thrust minor axis* which carries the thrust minor value, T_{minor}. Assembled into a quantity called oblateness, $O = T_{\text{major}} - T_{\text{minor}}$, these variables were instrumental in the discovery of the gluon in three-jet events [117]. We also consider the ratio of the two most minor thrust values: $R = T_{\text{major}}/T_{\text{minor}}$ in $3d$, and $R = T/T_{\text{major}}$ in $2d$. These ratios may be interesting because they are self-normalising.

At one end of the spectrum, a pencil-like event will score a value of thrust close to 1. At the other end of the spectrum is a perfectly isotropic configuration. To find the thrust value of a perfectly isotropic configuration in 3d, consider the average thrust value over all possible thrust axes [118],

$$\langle T(X, e_T)\rangle_{e_T} = \frac{\sum_i \langle |x_i \cdot e_T|\rangle_{e_T}}{\sum_i |x_i|} = \frac{1}{2}. \tag{5.2}$$

Since $\langle T\rangle = 1/2$, then if $T < 1/2$ for some e_T, then there exists an axis e'_T for which $T > 1/2$, and therefore e_T is not the thrust axis. The only way for T to equal $1/2$ is for $T = 1/2$ for all e_T, which only occurs in perfectly isotropic configurations. Therefore, $T \in [1/2, 1]$. In 2d, a similar argument results in $T \in [2/\pi, 1]$.

For calculating thrust, we use the algorithm described in the PYTHIA 6.4 manual [119].

The sphericity tensor is defined as

$$S^{\alpha\beta}(X; r) = \frac{\sum_i |x_i|^{r-2} x_i^\alpha x_i^\beta}{\sum_i |x_i|^r}, \tag{5.3}$$

where α, β run over the Cartesian coordinates of the p_i, and r is a parameter that may be chosen. The three eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, of this momentum ellipsoid are sums of squares of transverse momenta with respect to the three eigenvector directions. The smallest eigenvalue, λ_3, belongs to the eigenvector along which lies the greatest flow of momentum (the so-called jet axis, especially in a 2-jet event).

Event Shape	Rescaled Definition
$3d$ Thrust, T	$T \to 2T - 1$
$3d$ Thrust major, T_{major}	$T_{\text{major}} \to 1 - T_{\text{major}}$
$3d$ Thrust minor, T_{minor}	$T_{\text{minor}} \to 1 - T_{\text{minor}}$
$3d$ Oblateness, O	$O \to O$
$3d$ Thrust ratio, R	$R \to R$
$3d$ Sphericity, S	$S \to 1 - S$
$3d$ Aplanarity, A	$A \to 1 - A$
$3d$ 3-jet structure, C	$C \to 1 - C$
$3d$ 4-jet structure, D	$D \to 1 - D$
$2d$ Thrust, $T_\perp$	$T_\perp \to \frac{\pi}{\pi-2}\left(T_\perp - \frac{2}{\pi}\right)$
$2d$ Thrust major, $T_{\text{major},\perp}$	$T_{\text{major},\perp} \to 1 - T_{\text{major},\perp}$
$2d$ Oblateness, $O_\perp$	$O_\perp \to O_\perp$
$2d$ Thrust ratio, $R_\perp$	$R_\perp \to R_\perp$
$2d$ Sphericity, $S_\perp$	$S_\perp \to 1 - S_\perp$
$2d$ 3-jet structure, $C_\perp$	$C_\perp \to 1 - C_\perp$

Table 5.3: Table summarising the Event Shape variables considered, and how they are rescaled from the definitions given in the text. The rescaling is applied such that the Event Shapes have a common interpretation, with more isotropic events taking lower values and more pencil-like events taking higher values in the range $[0, 1]$.

The tensor has the desirable properties that (i) for perfect 2-jet events, two of the eigenvalues are zero; (ii) for 3-jet events, one of the eigenvalues is zero; and (iii) for events with a spherical momentum distribution, $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. The normalisation provides that $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

Sphericity event shapes are constructed from the eigenvalues. In 3d, $S = \frac{3}{2}(\lambda_2 + \lambda_3)$. Aplanarity is defined as $A = \frac{3}{2}\lambda_3$, with more planar events having lower values and more isotropic events having higher values in the range $0 \leq A \leq 1/2$. To quantify 3 and 4-jet structures, we define $C = 3(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)$ and $D = 27\lambda_1\lambda_2\lambda_3$, respectively. In 2d, λ_3 vanishes, so there is no notion of aplanarity nor of 4-jet structure. The two remaining sphericities are related, and we choose to use the sphericity value, defined as $S = 2\lambda_2$.

For ease of comparison between the event shapes, all event shapes are rescaled such that isotropic events are mapped to 0, while pencil-like events are mapped to 1. The rescaling formulae are summarised in Table 5.3.

5.2.3 Flatenicity

The isotropy of instanton events means that the instanton is expected to fill the $\eta - \phi$ plane in a characteristic way. As a measure of this distribution, Ref. [120] suggests partitioning the $\eta - \phi$ plane into a grid of cells, and considering the variation of the total cell transverse-momentum, $p_{\mathrm{T}}^{\mathrm{cell}}$. Thus Ref. [120] defines a new variable called 'flatenicity' as

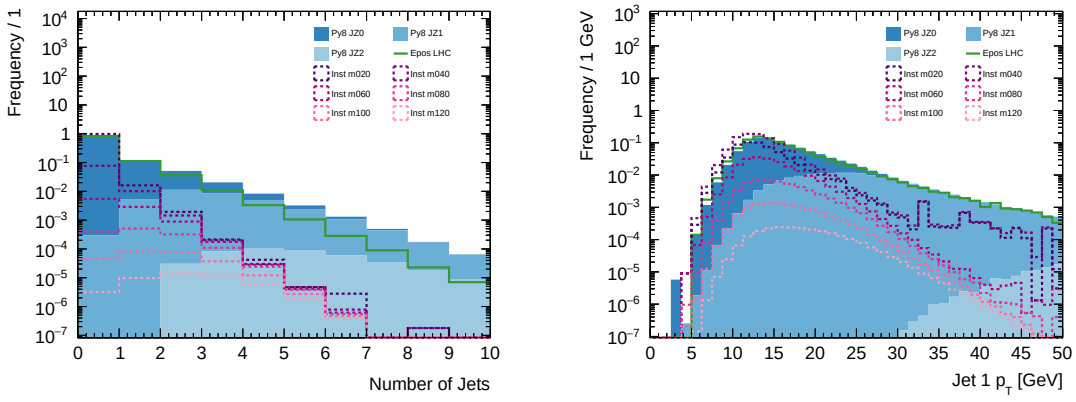
$$\rho = \frac{\sigma_{p_{\mathrm{T}}^{\mathrm{cell}}}}{\langle p_{\mathrm{T}}^{\mathrm{cell}} \rangle}, \tag{5.4}$$

where the numerator is the standard deviation of $p_{\mathrm{T}}^{\mathrm{cell}}$, and the denominator its average. In this analysis, flatenicity is calculated on a uniform grid of 10×10 cells in $\eta - \phi$, and is variously denoted F_{10} and $F(10, 10)$.

5.2.4 Flavour variables

Given the constrained flavour content of the instanton-induced final state, the prospect of flavour-tagging was investigated. The flavour-tagging techniques of ATLAS mostly apply to reconstructed jets [100]. These algorithms and the jets on which they operate are typically calibrated down to about 20 GeV in jet p_{T} [100, 121]. As shown in Figure 5.3, jets are reconstructed in simulated instanton events only about 1% of the time, and when jets are reconstructed, the leading jet of the event has a p_{T} below 20 GeV most of the time (at low mass). At high mass in the `m120` slice, signal events typically produce two jets, but still these jets are below reconstruction and tagging thresholds most of the time.

Consequently, a different flavour-tagging strategy is required. There exists a suite of 'soft' b-tagging algorithms developed for tagging the soft b-hadrons arising in compressed mass-spectrum scenarios of supersymmetry [122]. This suite comprises three methods of soft b-tagging. One is based on track-jets, which are amenable to the usual flavour-tagging algorithms. The other two are based on secondary-vertexing



(a) (b)

Figure 5.3: Distributions of (a) the number of reconstructed jets, and (b) the calibrated p_{T} of the leading jet in events with at least one reconstructed jet. Samples within a process are stacked before being normalised to unit area for comparison of shapes between the total signal and the backgrounds.

algorithms, which reconstruct displaced n-track vertices that are then subject to cuts. These algorithms accomplish tagging efficiencies of up to 20% for the secondary vertex-based algorithms, and up to 60% for the track-jets at about 30 GeV as reported in Ref. [122].

Thus given the low tagging efficiencies in our p_{T}-range of interest, and the anticipation of systematic uncertainty complexities, flavour-tagging is not employed in this analysis.

5.2.5 Distributions of discriminating variables

The distributions of the variables introduced in Sections 5.2.1, 5.2.2 and 5.2.3 are shown here. Figures 5.4 and 5.5 show reconstructed distributions of the event kinematic variables. Figures 5.6 and 5.7 show reconstructed thrust and sphericity distributions, respectively. In Figure 5.8, the reconstructed flatenicity distribution stands alone.

At first glance, it is surprising that the track multiplicity distribution shown in Figure 5.4a is more steeply falling in the case of the signal than the background.

Higher particle multiplicity is supposed to be one of the distinguishing features of the instanton event at particle level. However, this higher multiplicity is not seen at the level of reconstructed objects due to the $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ requirement on reconstructed tracks. Furthermore, the background distribution continues to higher track multiplicities because the background also continues to higher centre-of-mass values, while the signal distribution falls off due to the more steeply falling cross-section. This is reflected in the S_T distribution of Figure 5.4b.

Generally, it appears that the $3d$ Event Shapes are superior to the $2d$ versions at signal discrimination. Although the $2d$ Event Shapes all peak at lower, more isotropic values, signal and background are measured to be similarly isotropic in the transverse plane. The higher mass intervals of the signal are yet more isotropic than the lower mass intervals, which is consistent with the particle multiplicity being higher in the higher mass intervals. The $3d$ Event Shape distributions exhibit a more dijet-like structure as expected from the inclusion of the z-components of the momenta, but the signal appears to have a component that extends to lower vales.

The angular variables in Table 5.2 are inspired by studies of two-particle azimuthal correlations (e.g. Ref. [123]), and serve insights into the Event Shape variables. In Figure 5.5, we see a negative correlation between instanton mass and the $\sigma(\phi, \eta, \Delta\phi, \Delta\eta)$ variables in that lower masses exhibit higher angular variations. In the $\sigma(\Delta\phi, \Delta\eta)$ variables, this is as expected based on the $3d$ Event Shape variables, since more isotropic events should have less variation in the angular separation of pairs of particles. The jaggedness in the $\sigma(\Delta\phi)$ distribution is complicated, but possibly arises from a jettiness that makes certain $\Delta\phi$ values more likely. The flatenicity variable shown in Figure 5.8 has a similar structure.

The $\langle|\Delta\eta(i,j)|\rangle$ variable quite nicely separates signal and background. This is due to the greater spherical isotropy of instanton decays. In terms of pseudorapidity, $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$, a spherical distribution of charged particles looks like a cylindrical band of reconstructed tracks in the detector. A more isotropic event will therefore

have more pronounced band structure, and lower average differences of pseudorapidity between pairs of tracks as captured by the $\langle|\Delta\eta(i,j)|\rangle$ variable.

Differences between the Björken-x values of the incoming gluons result in a translation of this entire band to larger absolute pseudorapidity values. The background consists of MPI, in which approximately independent pairs of partons and their combinations of Björken-x values wash out any isotropy in polar angles to fill the detector more uniformly. The $3d$ Event Shape distributions are almost bimodal, which is possibly an effect of the offsetting of the band structure to produce relatively large momentum-z components. The 'instanton band' idea has been previously discussed in the literature [124, 125].

5.2.6 Correlation matrices of discriminating variables

Correlation matrices made from a subset of the discriminating variables are shown for the signal and EPOS LHC background samples in Figure 5.9. The correlation matrix of the background differs slightly from the correlation matrix of the signal, which ought to lend discrimination power. In the background, the correlation matrix broadly separates into Event Shape and non Event-Shape variables.

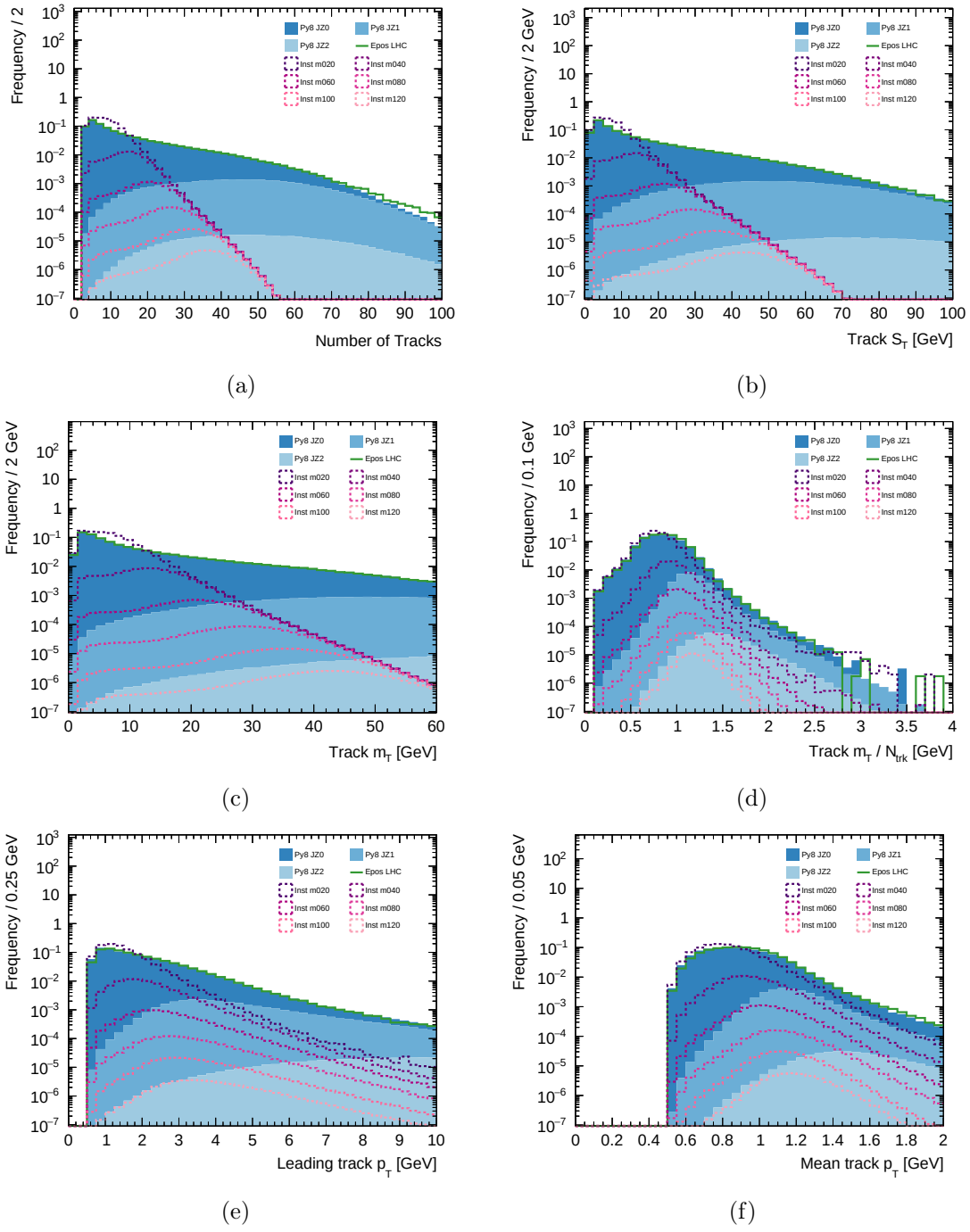


Figure 5.4: Distributions of mass-related discriminating variables. Samples within a process are stacked before being normalised to unit area for comparison of shapes between the total signal and the backgrounds.

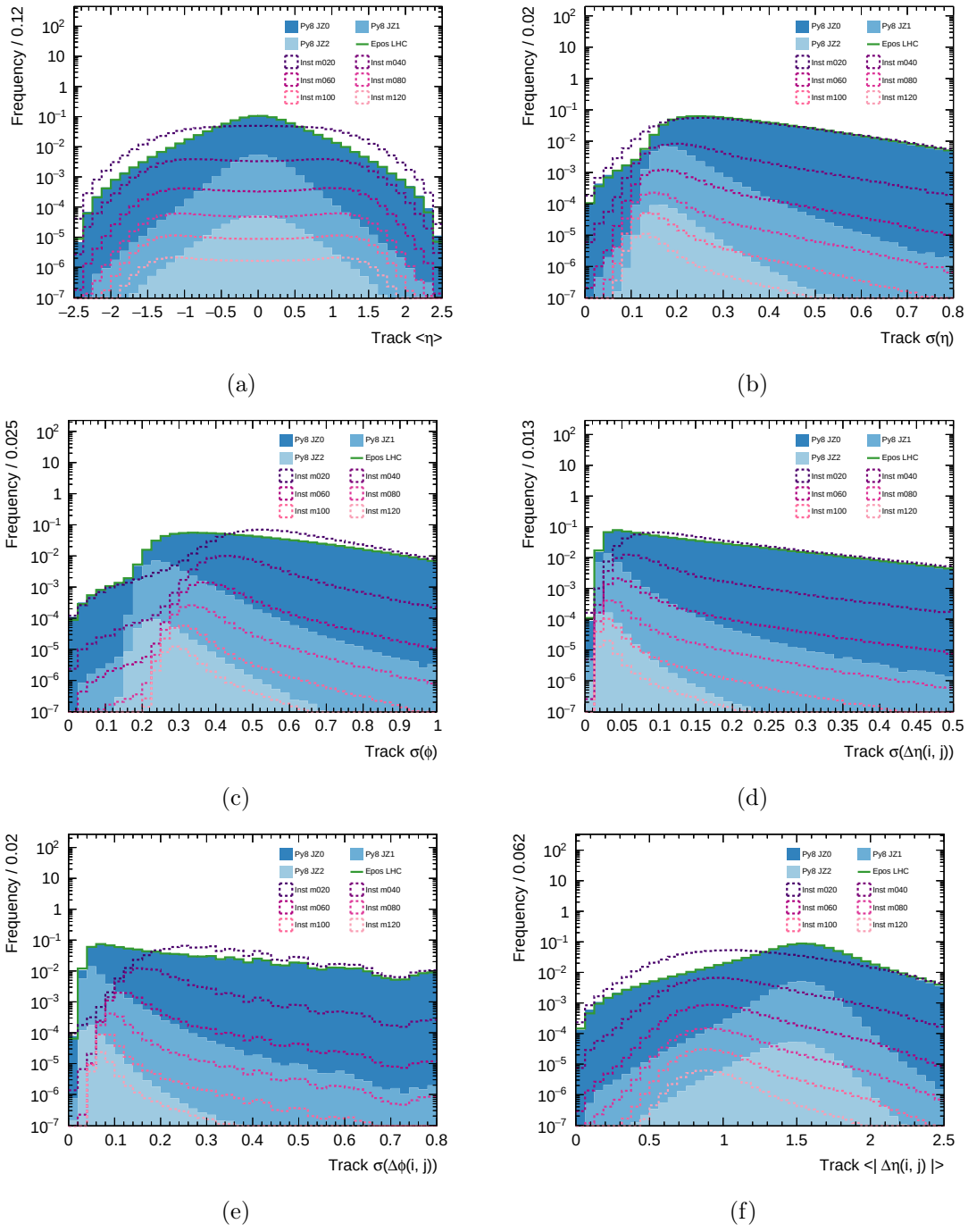


Figure 5.5: Distributions of angular variables. Samples within a process are stacked before being normalised to unit area for comparison of shapes between the total signal and the backgrounds.

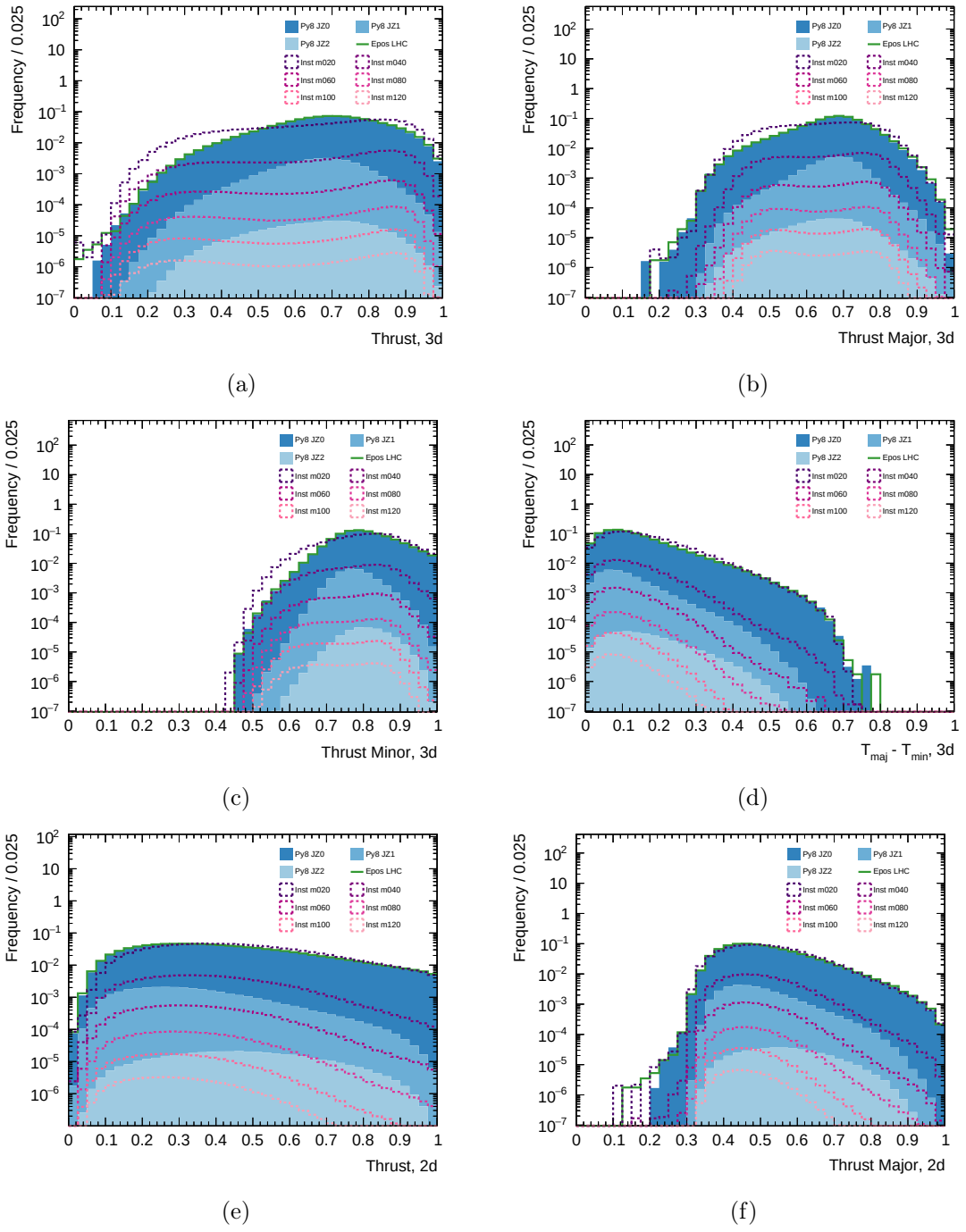


Figure 5.6: Distributions of rescaled thrust variables. Samples within a process are stacked before being normalised to unit area for comparison of shapes between the total signal and the backgrounds.

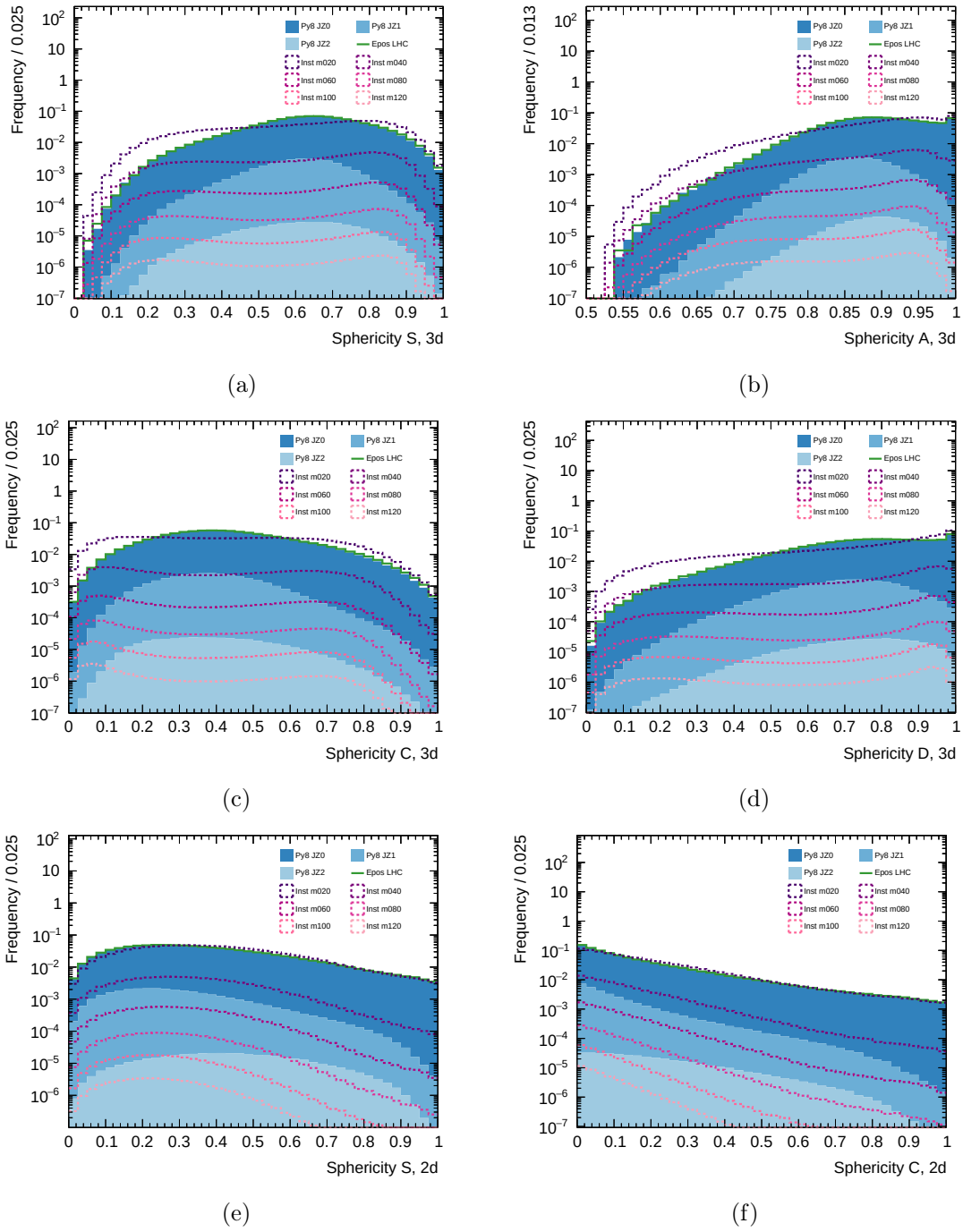
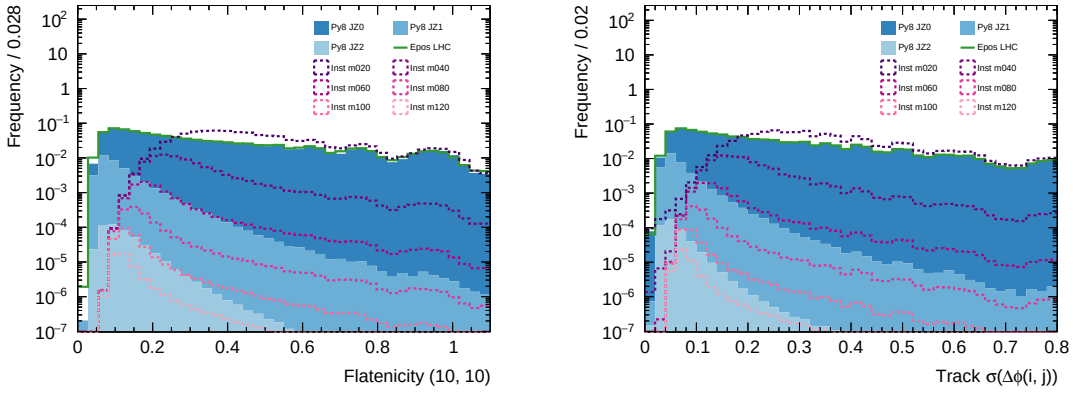
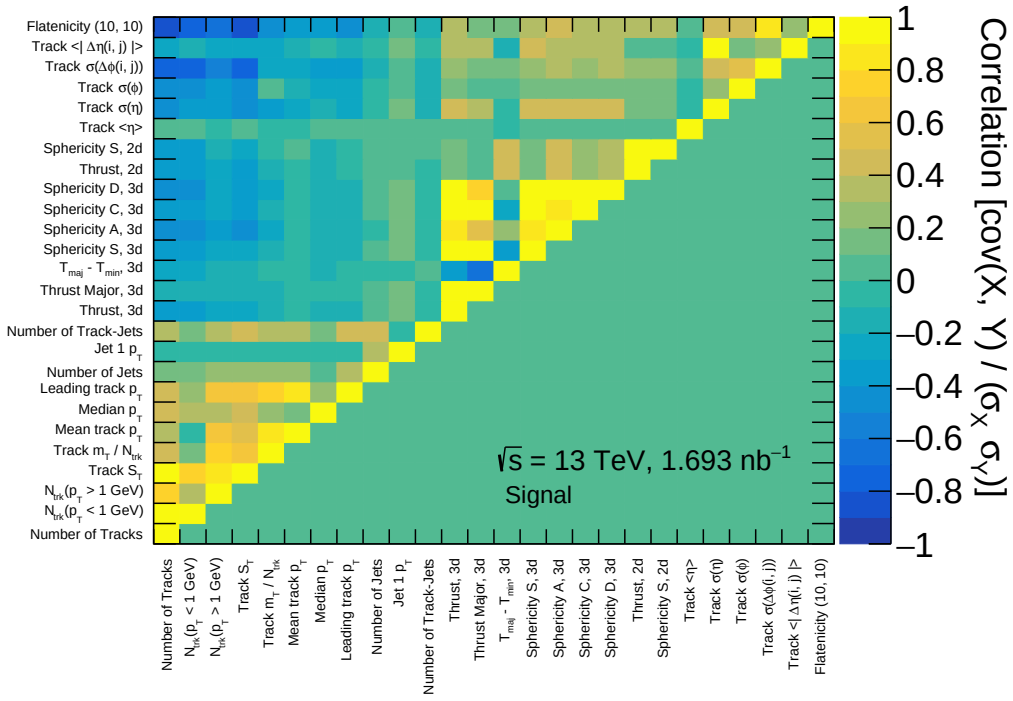


Figure 5.7: Distributions of rescaled sphericity variables. Samples within a process are stacked before being normalised to unit area for comparison of shapes between the total signal and the backgrounds.

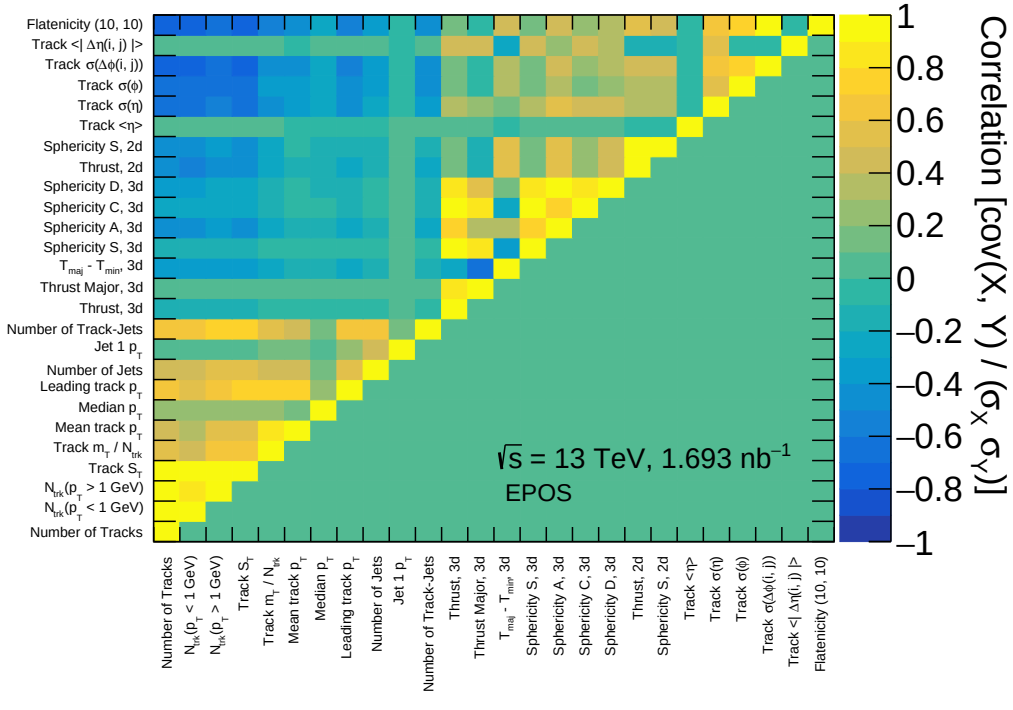


(a) (b)

Figure 5.8: Distribution of events in (a) the flatenicity variable, and (b) the standard deviation of the azimuthal difference between all pairs of tracks (reproduced here for comparison with flatenicity). Samples within a process are stacked before being normalised to unit area for comparison of shapes between the total signal and the backgrounds.



(a) Sherpa 3 signal MC.



(b) Epos background MC.

Figure 5.9: Correlation matrices of a selection of representative reconstructed discriminating variables. The correlation shown is the Pearson correlation as shown in the label on the z-axis.

5.3 Region Definitions and Background Estimation

In this section, the ABCDisCo Machine Learning approach is introduced and implemented as a basis for signal discrimination and background estimation.

5.3.1 Machine Learning approach

The differences between signal and background found in the discriminating variables presented in Section 5.2 are subtle, which makes designing a cuts-based analysis difficult. To exploit differences between signal and background among the correlations of variables, a Machine Learning (ML) approach has been adopted. Furthermore because of the desire for a data-driven estimation of the Soft QCD background, the ABCDisCo method of Ref. [126] was chosen in particular.

The ABCDisCo method works by enforcing the statistical independence of two discriminating variables as a basis for ABCD-style background estimation. The method relies on a Deep Neural Network (DNN) that has an architecture of the form shown schematically in Figure 5.10. This network takes a set $X = \{x_i\}$ of n input variables, and produces outputs at two nodes called A and B. The statistical independence is accomplished by training this network to classify at nodes A and B with a modified loss function that additionally penalises correlation between the outputs at A and B.

The full loss function is

$$\mathcal{L}[f(X), g(X)] = \mathcal{L}_{\text{cl}}[f(X), y] \; + \mathcal{L}_{\text{cl}}[g(X), y] \; + \alpha \mathcal{L}_{\text{dCorr}^2_{y=0}}[f(X), g(X)], \qquad (5.5)$$

which entails the correlation-penalising term, $\mathcal{L}_{\text{dCorr}^2_{y=0}}$, in addition to two classification terms, $\mathcal{L}_{\text{cl}}$. As in Figure 5.10, X denotes the set of input variables, while y represents an array of labels which take the value 0 for the background and 1 for

the signal events in the dataset used for training. The subscript on the correlation-penalising term indicates that this term applies to the background ($y = 0$) only.

For training classifiers A and B, binary cross-entropy terms, $\mathcal{L}_{\text{cl}}$, are used [127]. Binary cross-entropy is a standard choice from the field of information theory. It is given in terms of the training labels y and the predicted labels $\hat{y}$ as

$$\mathcal{L}_{\text{cl}}[\hat{y}, y] = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}). \tag{5.6}$$

In our application, the predicted labels $\hat{y}$ are given by $f(X)$ and $g(X)$ for classifiers A and B, respectively. This is essentially the expectation value of the log-loss, $\log \hat{y}$, calculated over the distribution of true labels, y, and gives a measure of dissimilarity between the distributions $\hat{y}$ and y. Training a DNN to minimise binary cross-entropy therefore corresponds to minimising the difference between the true and predicted distributions.

The correlation-penalising term is simply a statistical measure of non-linear dependence called distance correlation ('DisCo') [128]. DisCo is a function of two random variables U and V and is given in terms of distance covariance

$$\begin{aligned}\mathcal{V}^2(U, V) =& \langle |U_k - U_l| \times |V_k - V_l| \rangle_{kl} \\ &+ \langle |U_k - U_l| \rangle \times \langle |V_k - V_l| \rangle_{kl} \\ &- 2 \langle |U_k - U_l| \times |V_k - V_m| \rangle_{klm},\end{aligned} \tag{5.7}$$

which is calculated over all pairs and triplets of coordinates U and V. Distance correlation is then given as

$$\text{dCorr}^2(U, V) = \frac{\mathcal{V}^2(U, V)}{\sqrt{\mathcal{V}^2(U, U)\mathcal{V}^2(V, V)}}. \tag{5.8}$$

In our application, $\mathcal{L}_{\text{dCorr}^2_{y=0}}(\cdot)$ always simply equals $\text{dCorr}^2(\cdot)$, and the random variables U and V correspond to $f(X)$ and $g(X)$.

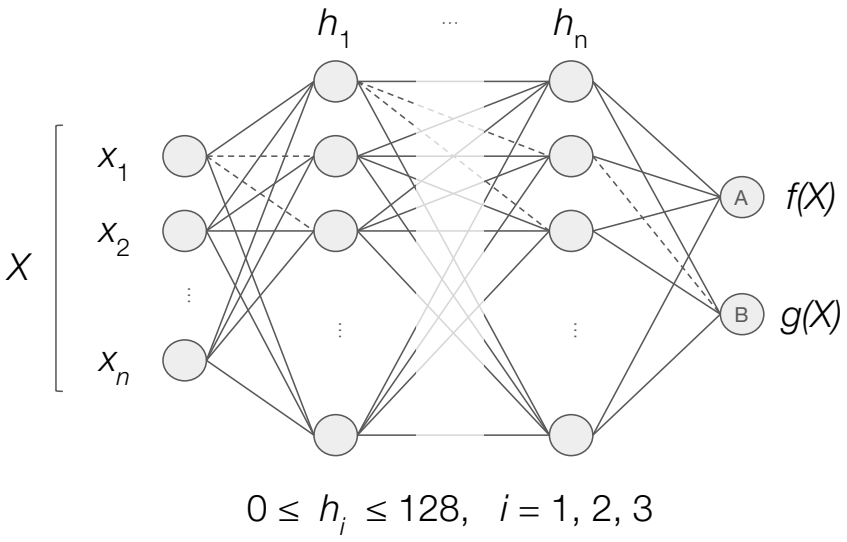


Figure 5.10: Schematic of the template architecture of the DNNs used in the ABCDisCo implementation. The network takes n input variables, and has two output nodes. One output node is interpreted as Classifier A, while the other is interpreted as Classifier B. Classifier A can be thought of as applying a function f to the input variables X, such that it produces predictions $f(X)$. Classifier B similarly applies a function $g(X)$. The hidden layers, indexed by i, are fully-connected and may have up to 128 nodes h_i. Architectures with up to $i = 3$ hidden layers are considered, with the shaded box represeting an optional third layer in the schematic. Drop-out layers with a default dropout probability of 0.25 are shown as dashed lines.

For the purposes of this thesis, it suffices to note that DisCo has the desirable property that it is zero if and only if the variables are statistically independent, and positive otherwise. Thus DisCo is suitable for use as a regularisation term in the loss function of a neural network, as discussed first in Ref. [129] (in the context of decorrelating neural networks more generally). Introducing the correlation-penalising term incorporates one additional hyperparameter: the α parameter is a coefficient that controls the relative importance of classification and decorrelation. This decorrelation coefficient is a hyperparameter of the training whose value is not known in advance.

5.3.2 Implementation of the ML approach

Our implementation of ABCDisCo is based on the Keras framework [127] with the tensorflow backend [130], and makes use of tools from scikit-learn [131].

The signal and background MC samples used are firstly split 80:20 into training and validation datasets, respectively. Then because the raw input variables can differ in magnitude, the ranges of the inputs are normalised to zero mean and unit variance. For each input variable x_i in the set X, the mean and variance are 'learned' from the training dataset, and are 'removed' across the training and validation datasets before they are considered by the network. This transformation is fitted to background events only, with due respect to the null hypothesis that the data consist of background events only. This kind of feature normalisation is important for (i) ensuring that each input feature contributes proportionately to the output, (ii) for the convergence of the optimisation, and (iii) if regularisation is applied, for the regularisation to work similarly across network coefficients.

Training proceeds in epochs and batches. An epoch is a complete pass of the training dataset, which is iterated over in batches. The training works on the training dataset, while the validation dataset is used for monitoring performance metrics during the course of training. Training continues for a pre-defined number of epochs or until some stopping condition is met. We use an early-stopping mechanism, which terminates the training once the loss evaluated on the validation dataset ceases to improve for three successive epochs. If performance decreases over these last epochs, the network is restored to the configuration of highest loss. Because the stopping condition is thus a function of the validation data, the validation dataset influences the loss minimisation, which can result in over-training. Therefore, inference ought in principle to be performed on a third statistically-independent dataset called the *test* dataset. However, our MC sample sizes were deemed too small for this.

To protect against over-training, so-called 'drop-out' regularisation layers can be

included. Drop-out regularisation works by dropping connections between nodes with some probability during training. The drop-out probability used here is set to the default value of 0.25, and drop-out is applied after every hidden layer.

For a DNN to learn the shapes of physical distributions and to pay equal attention to the signal and background classes, training weights are used. Training weights have two components, the first being the total event weight, which is proportional to the MC event generator weight and the process cross-section. The second component, often called a class weight, is a scale-factor for equating the total weights of the signal and background classes.

In practice, it was found that since the classification terms in the loss function drive background events to low classifier output scores, events at low classifier output scores tend to dominate in the calculation of the DisCo term. As a result, the DisCo term tends to fail to account for high classifier output scores and thus fails to decorrelate in that region. To fix this, a second DisCo term is added to Eq. 5.5 that only operates on events above the median scores in the batch of classifier outputs under consideration. This additional term to be added to Eq. 5.5 looks like

$$\mathcal{L}_{\mathrm{dCorr}^2_{y=0}}[f(X), g(X) \mid f(X) > \mathrm{med}\ f(X), g(X) > \mathrm{med}\ f(X)]. \tag{5.9}$$

The same problem was encountered in Ref. [126] and solved using a similar trick.

Hyperparameter optimisation

The architecture of the network is optimised using the random search method of KerasTuner [132]. This configures KerasTuner to randomly sample architectures within the range of configured architectures. The hyperparameters sampled in this manner are namely the number of hidden layers, and the number of nodes in each hidden layer. The hyperparameter search ranges are shown in Table 5.4. This search method tends to conclude that networks of three hidden layers of 128 nodes work

Hyperparameter	Scan Range or Scanned Values
Number of layers	`{1, 2, 3}`
Number of nodes	`[32, 128, 32]`
α	`[0, 30]`

Table 5.4: Table showing the ABCDisCo DNN's hyperparameters and the values scanned by KerasTuner in search of the architecture that minimses the total loss. Curly braces indicate sets, while square brackets indicate ranges, with a third entry indicating the step-size used to step through the range.

well. This network architecture was used everywhere for simplicity.

We scan the decorrelation coefficient α manually. This scan involves repeating the training at different values of α, taken from $\alpha = \{0, 1, 2, ..., 30\}$. Results obtained from training once at each value of α are not easily reproducible due to the stochastic nature of the training. To ameliorate this, a network is trained from random initial weights ten times for each value of α. The best value of α is chosen from this superset of networks by optimising

$$\mathcal{L}_\alpha = (1 - \text{acc}[f(X)])(1 - \text{acc}[g(X)]) + |k|, \tag{5.10}$$

where $\text{acc}[f, g]$ are the accuracies of Classifiers A and B, respectively, at the point of minimal loss during the training, and k is a measure of the distance correlation (called 'non-closure') which will be defined in Section 5.3.7 after the background estimation strategy is explained. Roughly speaking, k is here to ensure that the optimal network produces a background estimation that closes. The addition of these joint inaccuracy and correlation terms reflects the balance that is to be struck between them: when k is smaller, the optimisation should pay more attention to the inaccuracy term, vice versa.

The networks themselves are optimised using the Adam algorithm [133]. The classification terms are evaluated event-wise, and are averaged over a batch for the back-propagation. The DisCo term, however, is calculated across an entire batch. From the point of view of the DisCo term, a large batch size is preferred for accuracy. On the other hand, the efficiency of the classification terms is optimal at intermediate

batch sizes. A batch size of 10 000 was found empirically to produce a stable training featuring good classification and decorrelation.

Network input variables

The discriminating variables given as inputs to the ABCDisCo networks are listed in Table 5.5. Based on the correlation matrices shown in Figure 5.9, there is a physical basis for decorrelation of the two classifiers in that there exist pairs of variables that are uncorrelated or only lightly correlated. In particular, while Event Shapes are highly correlated amongst themselves, they are generally not very correlated with the mass variables. Therefore a range of Event Shape and non Event-Shape variables are inserted into the list of inputs, in addition to some of the obvious discriminating variables identified from Figures 5.4 to 5.8. All of the variables listed in Table 5.5 are given as inputs to both classifiers, as they are part of the same joint network as shown in Figure 5.10.

5.3.3 Mass binning

As laid out in the introduction of this chapter, a part of the search design is to perform the analysis in mass categories. Low, medium, and high-mass categories are defined in terms of the truth instanton mass as the intervals $(20, 40), (60, 80), (100, 150)$ GeV, respectively.

Events are binned in terms of track-S_T, which is now re-interpreted as a proxy for the reconstructed instanton mass. To find the bin boundaries, a search over all possible combinations of four bin edges (allowing gaps between the three categories, with the lowest cut fixed at $\mathtt{x}$ and the highest at ∞) is conducted. The selected edges are

Number of tracks	N_{trk}
Number of tracks above 1 GeV	$N_{\text{trk}}(p_{\text{T}} > 1\,\text{GeV})$
Number of tracks below 1 GeV	$N_{\text{trk}}(p_{\text{T}} < 1\,\text{GeV})$
Scalar sum of track p_{T}	S_T
Transverse mass per track	m_T/N_{trk}
Mean track p_{T}	$\langle p_{\text{T}} \rangle$
Median track p_{T}	med p_{T}
Leading track p_{T}	max p_{T}
$3d$ Thrust	$T, 3d$
$3d$ Thrust major	$T_{\text{major}}, 3d$
$3d$ Oblateness	$T_{\text{major}} - T_{\text{minor}}, 3d$
$3d$ Sphericity	$S, 3d$
$3d$ Aplanarity	$A, 3d$
$2d$ Thrust	$T, 2d$
$2d$ Sphericity	$S, 2d$
Mean pseudorapidity	$\langle \eta \rangle$
Standard deviation of ϕ	$\sigma(\phi)$
Standard deviation of η	$\sigma(\eta)$
Standard deviation of $\Delta\phi(i,j)$ of all track pairs	$\sigma(\Delta\phi(i,j))$
Mean of $\lvert\Delta\eta(i,j)\rvert$ of all track pairs	$\langle\lvert\Delta\eta(i,j)\rvert\rangle$
Flatenicity on a 10×10 grid	$F(10,10)$

Table 5.5: List of variables given as inputs to the ABCDisCo DNNs.

those that score the highest

$$\mathcal{L}_{M_I} = \sum_i \epsilon_i \times (1 - \delta_i), \tag{5.11}$$

where ϵ_i and δ_i are the efficiency and signal contamination in the i^{th} mass category, respectively. The efficiency corresponds to the fraction of events belonging to truth bin i that is captured by reconstructed bin i. Signal contamination corresponds to the fraction of the events captured in reconstructed bin i that do not belong to truth bin i.

Running this search yields the bin edges $(\mathtt{x}, 16), (18, 32), (34, \infty)$ GeV in track-S_T as shown in Figure 5.11. The training does not result in good classification in the case of the low-mass bin when $\mathtt{x} = 0$. This is because at very low masses, the kinematics of signal and background events are indistinguishable as far as a network is

concerned. The confusion is represented as a peak in both the signal and background distributions at about 0.5 in classifier output. The performance of this low-mass network can be recovered through re-interpreting the low-mass network as being parametric in S_T: $f = f(X)(S_T)$. With this re-interpretation, the classifier output distributions are functions of S_T, and a scan of lower cuts on S_T can be performed post-training. This procedure is summarised in Figure 5.12, where $\mathtt{x} = 10\,\text{GeV}$ is found to recover classification performance.

In each mass category, a network is trained according to the procedure laid out above, whence the optimal α parameters are found to be

$$\alpha^\star = \begin{cases} 9 & \text{Low-mass,} \\ 8 & \text{Medium-mass,} \\ 16 & \text{High-mass,} \end{cases} \tag{5.12}$$

with the star on α denoting the optimal value henceforth.

5.3.4 Network health checks

The performance of the training is typically checked by looking at the evolution of the loss function over epochs. 'Network-health' plots showing these evolutions have been made and are shown in the left-hand column of Figure 5.13. Because in some cases the loss appeared to be minimised after the first epoch, beyond which no apparent progress is made, the evolution of the loss over the batches within epochs is also interrogated. These plots are shown in the right-hand column of Figure 5.13. In these plots, it can be seen that the loss decreases within epochs, as well as across epochs, as would be expected in a healthy DNN.

In the specific case of our ABCDisCo networks, it is also interesting to interrogate what happens to the distance correlation in the trained networks as a function of

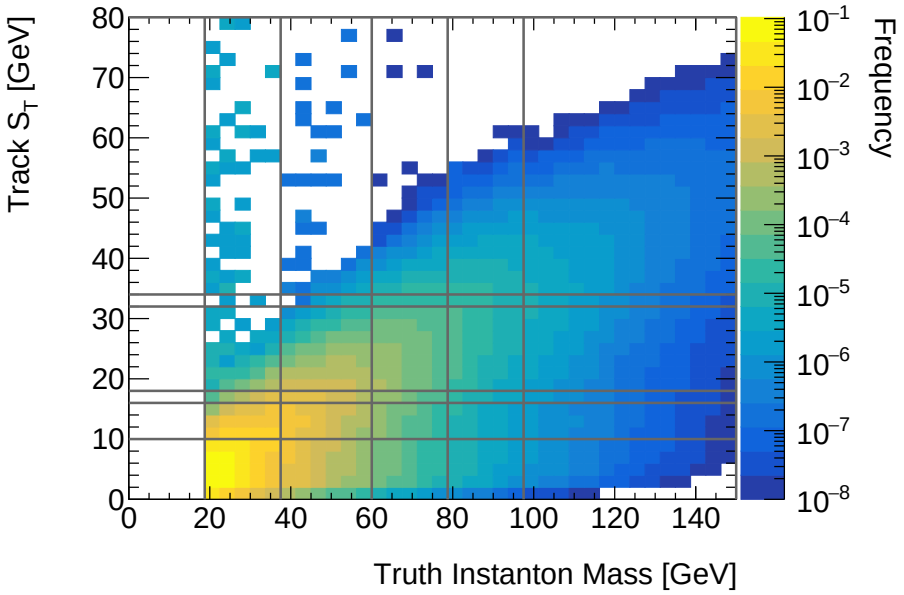


Figure 5.11: Histogram showing the correlation between truth instanton mass and track-S_T, and how the mass category definitions at truth level are translated into reconstructed mass category bins. The vertical grey lines roughly correspond to the generated truth-mass slices described in the text, while the grey horizontal bands correspond to the bin edges separating the reconstructed mass intervals that are also described in the text.

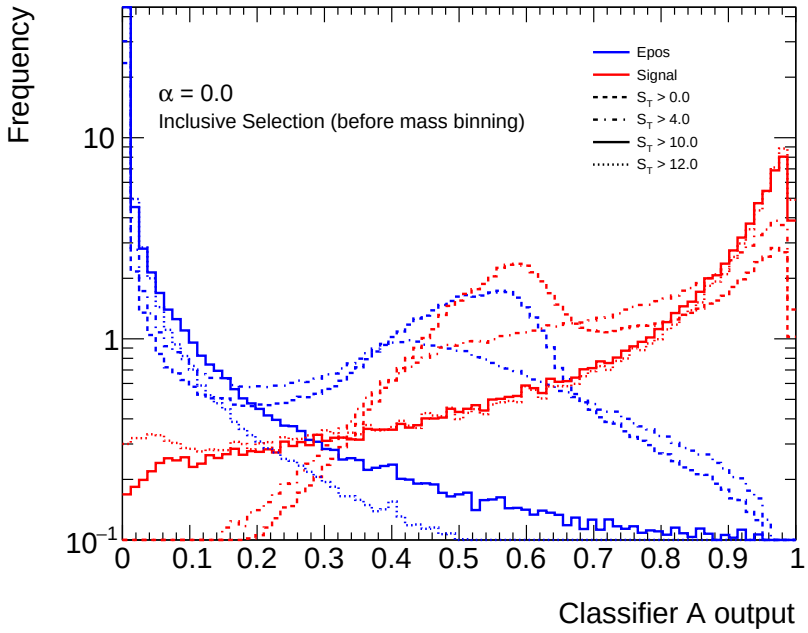


Figure 5.12: Illustration of how increasing the lower cut on track-S_T from 0 to 10 GeV works to recover classification performance in what will become the low-mass bin.

the decorrelation coefficient, α. These plots are shown in Figure 5.14. The spread of the non-closure, characterised by the larger error bars, shows that a given value of decorrelation can result in a range of non-closures. Generally, these plots show that the distance correlation reduces with α until it plateaus. It is possible that this plateau region corresponds to where the distance correlation is saturated at the minimum value attainable in the set of input variables. The optimal α is generally near the onset of this plateau region. This is sensible because here the distance correlation is minimal while α is small, and classification power still correspondingly high, which shows that the optimisation of α based on Eq. (5.10) works.

5.3.5 Classifier output distributions

The distributions of classifier outputs in the three mass categories are shown in Figure 5.15. All of these plots show that some level of classification has been attained, with the background peaking at lower classifier output values and the signal peaking at higher values in every case. The classification performance is shown directly in Figure 5.16 in terms of ROC curves. To quantify discrimination power, Area Under the Curve (AUC) is used as a summary statistic. Each of the ROC curves possesses an AUC value above 0.8, which is deemed to be good classification. The discrimination power is best in the high-mass region, where AUCs get up to 0.98, and is worst in the low-mass region, where the AUCs are about 0.82. This reflects the trade-off that is in play between the signal cross-section and discrimination power as functions of instanton mass. The ROC curves are drawn from the training and validation datasets separately as a check for over-training. That the AUCs for the training and validation datasets are similar indicates that no major over-training has taken place.

Figures 5.15 and 5.16 account for the one-dimensional classification performance of Classifiers A and B separately. In Figure 5.17, the distributions of signal and background events in the two-dimensional planes spanned by Classifiers A and B are

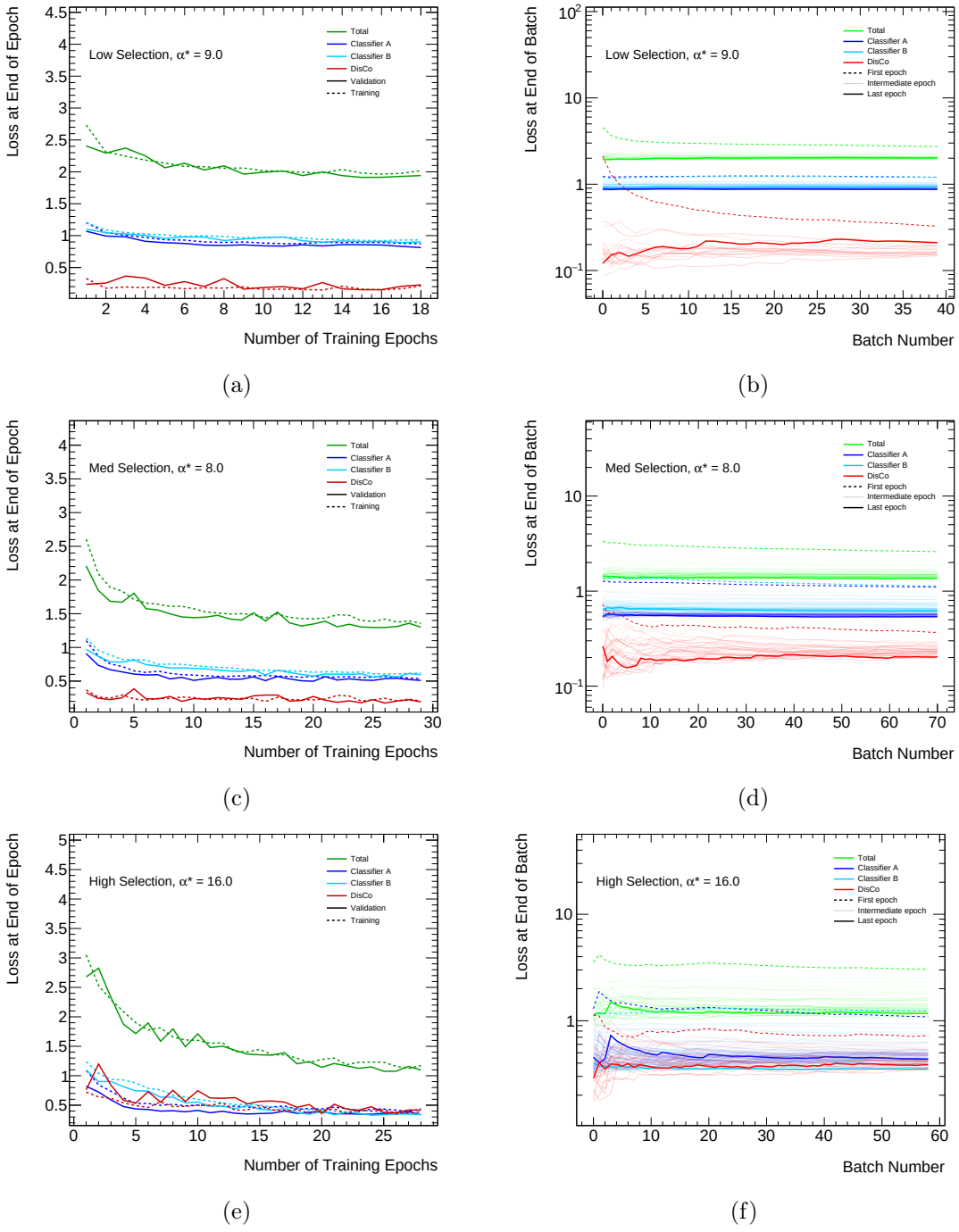


Figure 5.13: 'Network health' plots showing how the value of the loss evolves over epochs (in the left-hand column), and over batches (in the right-hand column). The total loss is separated into its three components to show how the training minimises each of them. The DisCo term shown corresponds to $\alpha\mathcal{L}_{\mathrm{dCorr}^2_{y=0}}$. In the batch plots, the loss through the first epoch is shown as a dashed line. The loss through intermediate epochs is shown as shaded lines, before the loss in the final batch is shown as a solid line. The point of the batch plots is to notice how the loss decreases between the first and the last epochs.

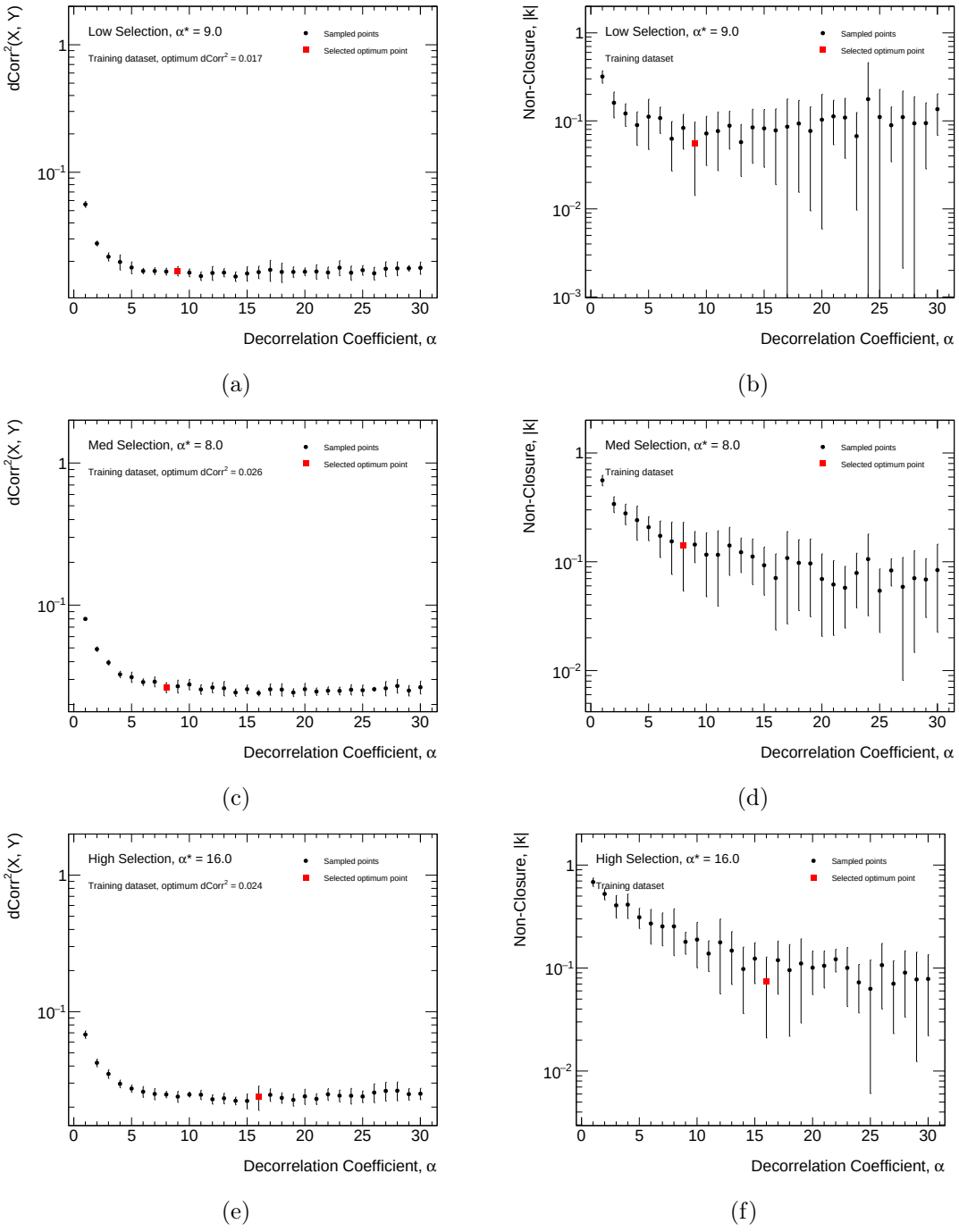


Figure 5.14: Distance correlation (left-hand column) and non-closure (right-hand column) in the training dataset as functions of the decorrelation coefficient, α, for the low, medium, and high-mass categories from top to bottom. The optimal α value is indicated as a red square in each mass category. The 'sampled point' at each value of α corresponds to the mean calculated across the repetitions of the training at that α, while the error bar shows the standard deviation. These data are taken from the training dataset.

shown. Here, it can be seen how the two classifiers work together, moving background events to the bottom left-hand corner and signal events to the upper-right. The decorrelation of the two classifiers is visible.

5.3.6 What do the networks learn?

The functioning of the ABCDisCo networks is non-trivial because of the training demanding concomitant classification and decorrelation. It is important to check that the networks are learning to minimise the loss based on the input features in a physically meaningful way. Moreover, it is interesting from the physics point of view to find out what distinguishes instanton events. In this regard, we have two strategies.

SHAP values

Our first strategy is based on SHAP (SHapley Additive exPlanation) values [134]. The best explanation of a model is the model itself (since it perfectly represents itself), but a DNN it too complex to understand. The SHAP formalism constructs a simplified 'explanation' model, $h(z')$, in terms of simplified inputs z' according to

$$f(x) = h(z') = \phi_0 + \sum_i^N \phi_i z'_i, \tag{5.13}$$

where the ϕ_i coefficient weighs the importance of the i^{th} input variable, and ϕ_0 is the output expected when all inputs are missing. The feature attributions ϕ_i are known from cooperative game theory as Shapley values [135] (nicely reviewed in Ref. [136]).

Shapley values give the marginalised contribution of input feature i, averaged over all feature subsets S, as given by

$$\phi_i = \sum_{S \subseteq X \setminus \{i\}} w_S \left[h(z'_{\setminus i} \cup \{i\}) - h(z'_{\setminus i}) \right]. \tag{5.14}$$

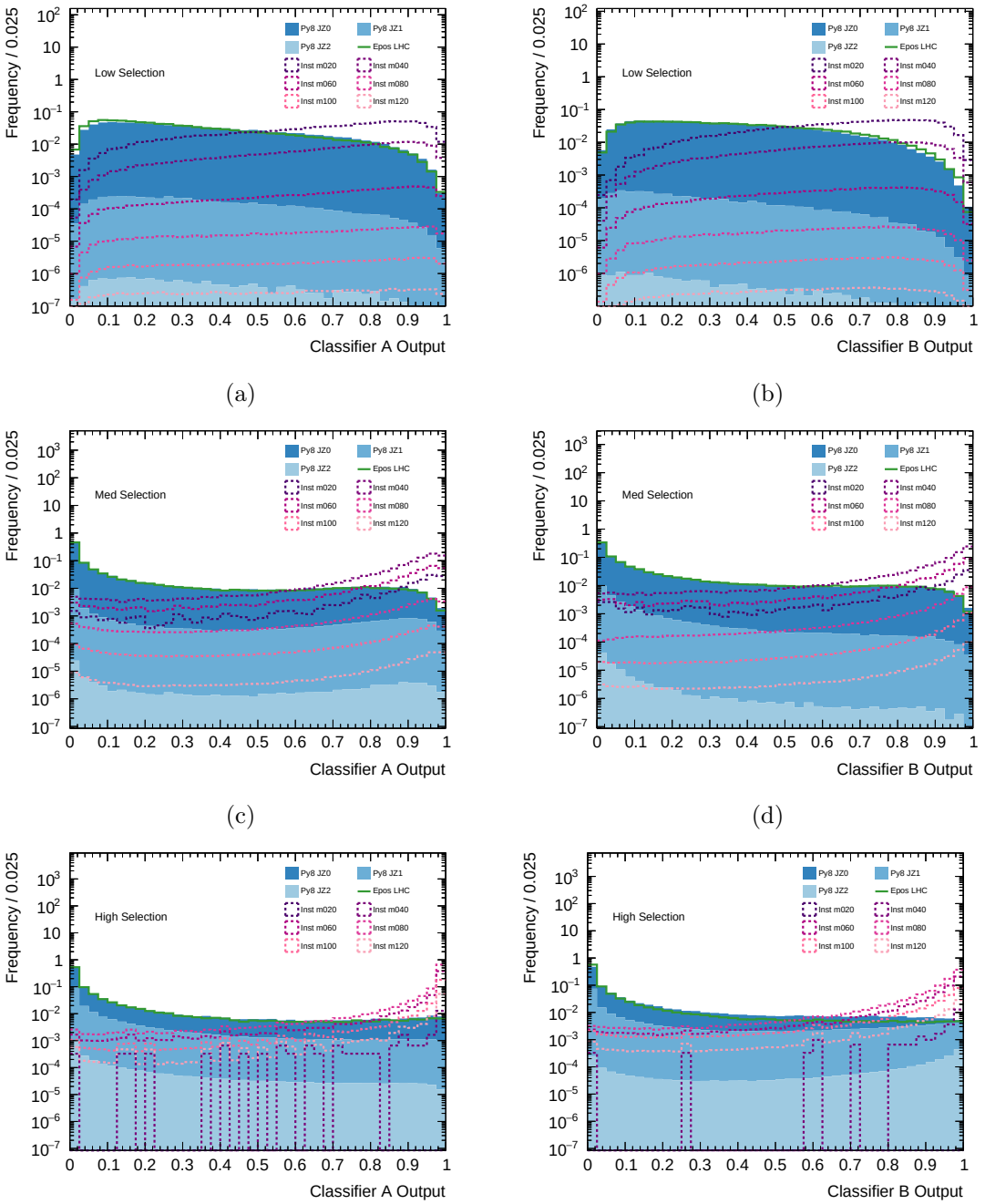
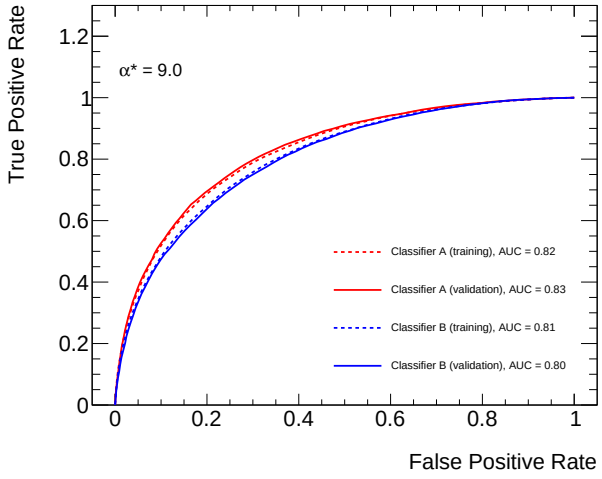
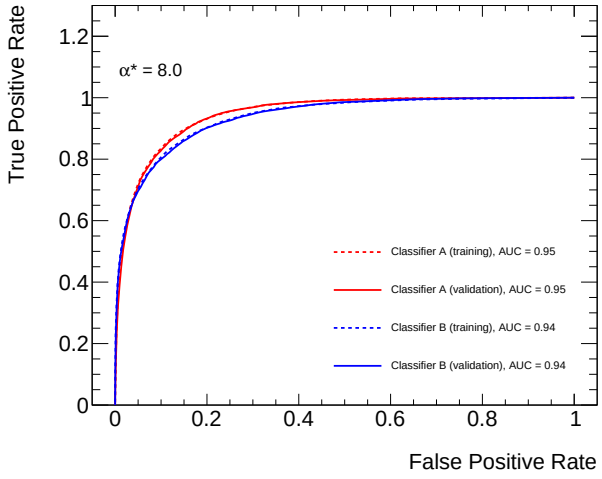


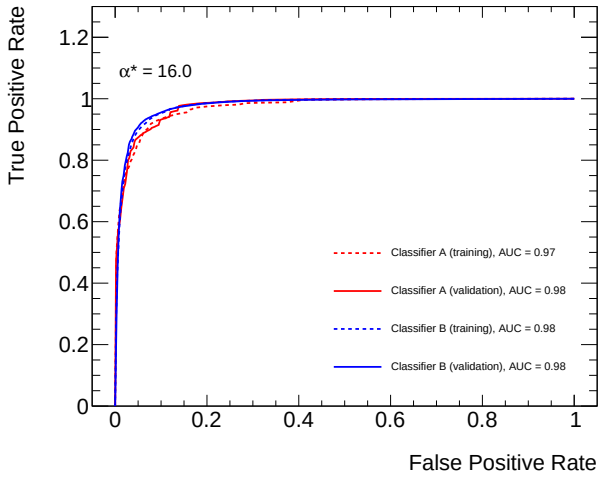
Figure 5.15: Classifier A and B output distributions in the low, medium, and high-mass categories. Distributions are normalised to unit area for comparison of shapes.



(a) Low-mass network.

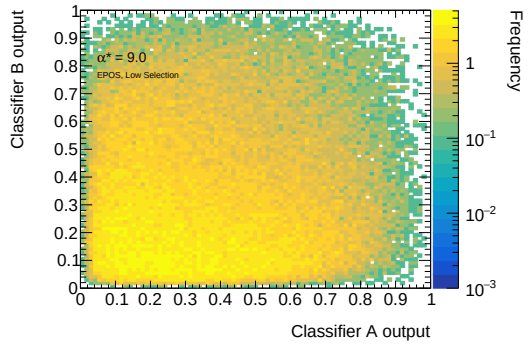


(b) Medium-mass network.

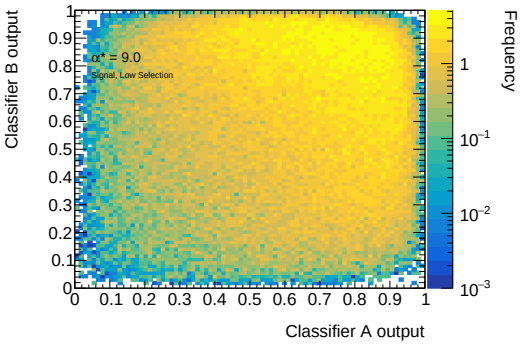


(c) High-mass network.

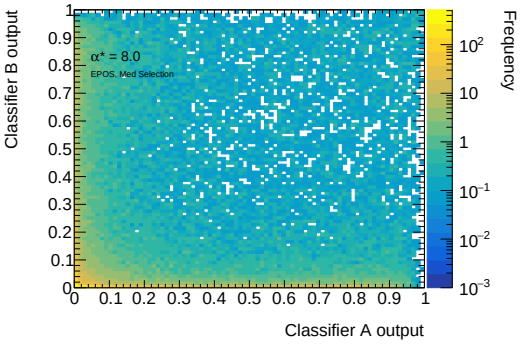
Figure 5.16: Receiver Operating Characteristic 'ROC' curves for the two classifier output variables, f and g. The training (dashed line) and validation (solid line) datasets are shown separately to check for over-training, and are found to agree almost exactly. Each sub-figure shows these ROC curves for the optimised value of α. AUC stands for Area Under the Curve and is used as a summary statistic to quantify the discrimination power of the classifiers.



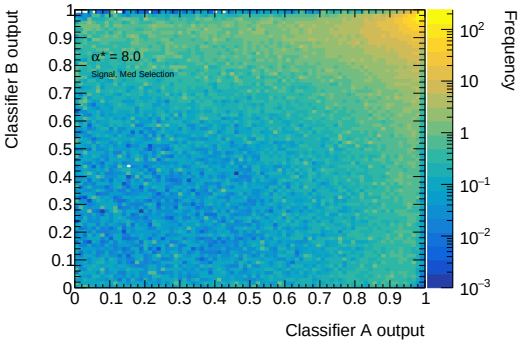
(a) Epos, low-mass region



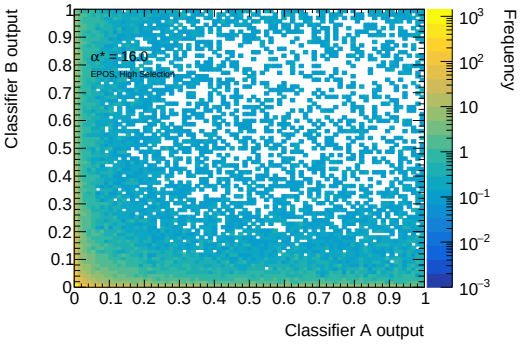
(b) Signal, low mass region



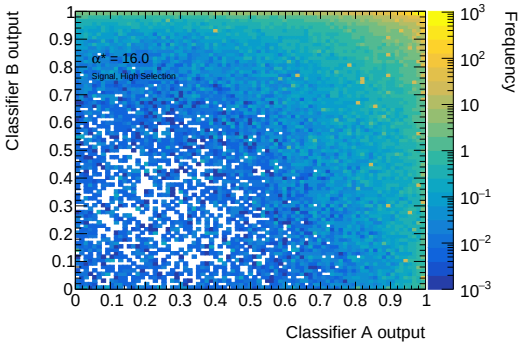
(c) Epos, medium-mass region



(d) Signal, medium-mass region



(e) Epos, high-mass region



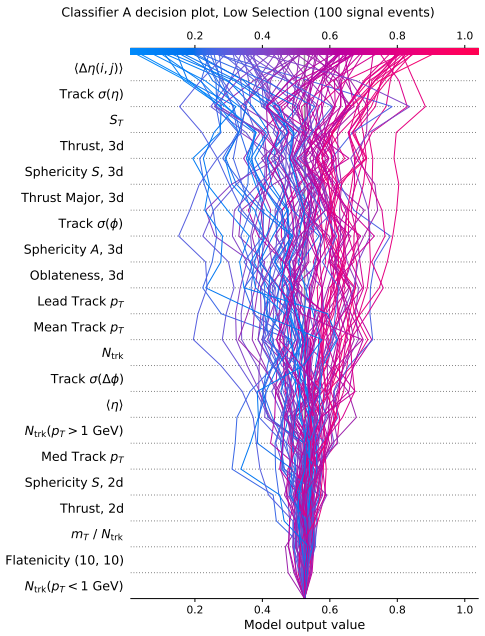
(f) Signal, high-mass region

Figure 5.17: Distributions of Epos background and signal events in the two-dimensional planes spanned by the outputs of Classifiers A and B in the low, medium and high-mass categories, respectively from top to bottom.

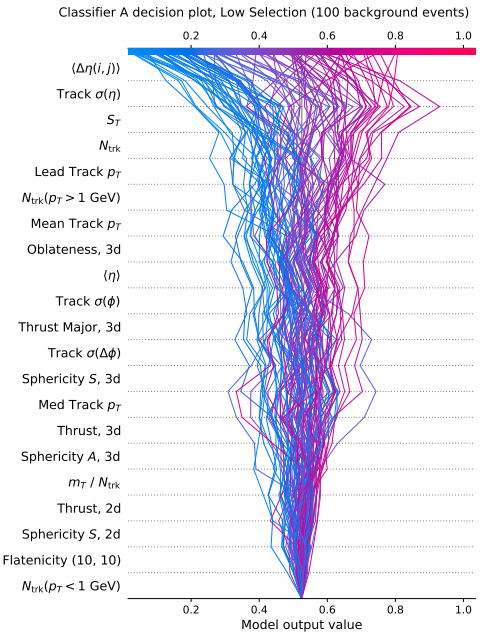
In this equation, $z'_{\setminus i}$ stands for the set of simplified inputs in S, and w_S is the combinatorial weight of subset S. Formally, this evaluation involves retraining the DNN on every subset S of the inputs with the i^{th} feature withheld. This is infeasible, so Ref. [134] assigns $h(z') = \mathrm{E}[f(z) \mid z_S]$, which gives the expectation value of the network's output f over all events, but with the features in the subset S fixed at the current input.

SHAP values are calculated here using the Deep SHAP algorithm [134], which assumes the independence of all input variables. Because of this assumption, the results show the importance of *not removing* the i^{th} variable from the input, as opposed to the importance of the variable for the outcome. A variable providing good discrimination power may score a low SHAP value if it is eclipsed by a more important variable.

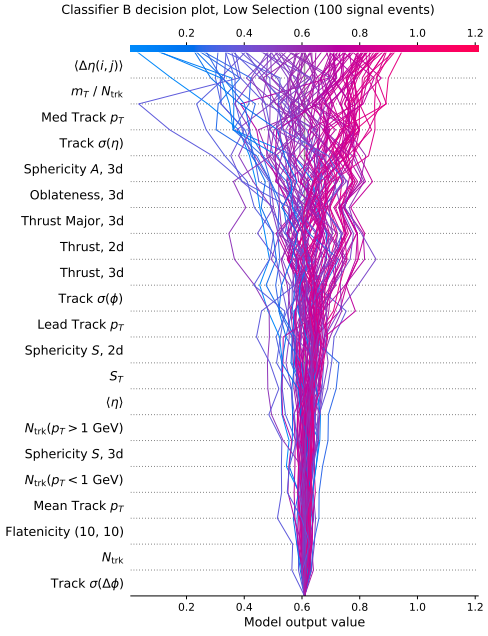
Ranked decision plots based on SHAP values are shown in Figures 5.18, 5.19 and 5.20 for the low, medium, and high-mass categories, respectively. In these plots, each line corresponds to an event, and the input variables along the vertical axis are ranked according to their mean importance in the sample of 100 events shown. The horizontal step taken by a line across a variable on the vertical axis corresponds to the importance of that variable in the current event. On the bottom axis, the starting point corresponds to ϕ_0 in Eq. (5.13). Summing an event's SHAP values over all inputs yields the classifier output score shown at the top with the colour bar. Background events ending above 0.5 and signal events ending below 0.5 in classifier output are said to be misclassified. The ranking of the variables varies with classifier, but generally the variables $\langle|\Delta\eta(i,j)|\rangle$, $\sigma(\eta)$, S_T, and med p_{T} are ranked as important, with $\langle|\Delta\eta(i,j)|\rangle$ clearly being the most important.



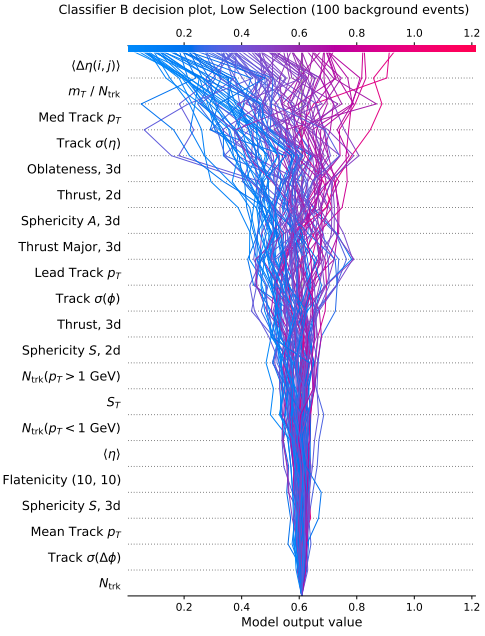
(a) (Classifier A, signal)



(b) (Classifier A, EPOS)

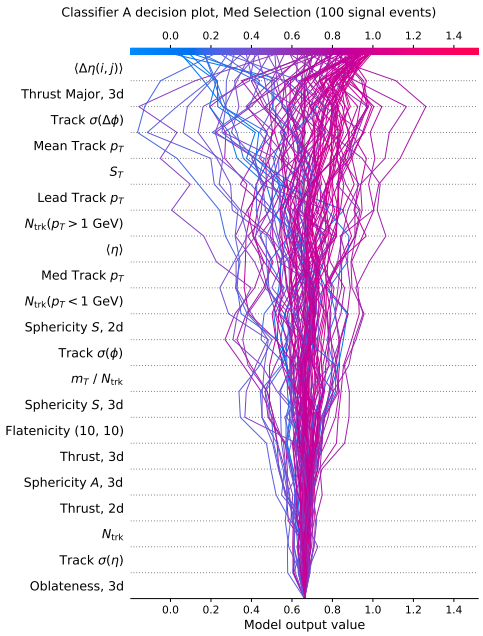


(c) (Classifier B, signal)

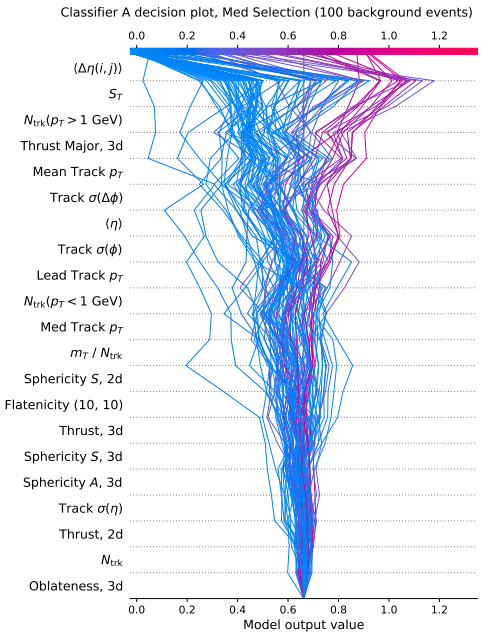


(d) (Classifier B, EPOS)

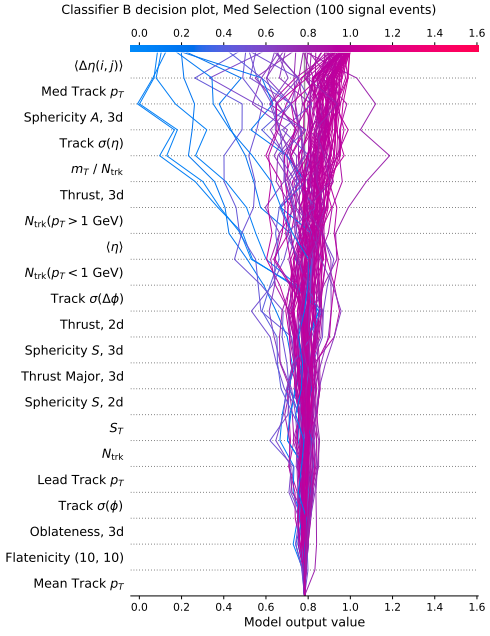
Figure 5.18: SHAP input variable-ranking plots in the low-mass category. Each line corresponds to one event. The Shapley value of each variable (proportional to its importance for the classification result) is shown as a horizontal step from one variable to the next. The most important variables are at the top, and the least important at the bottom.



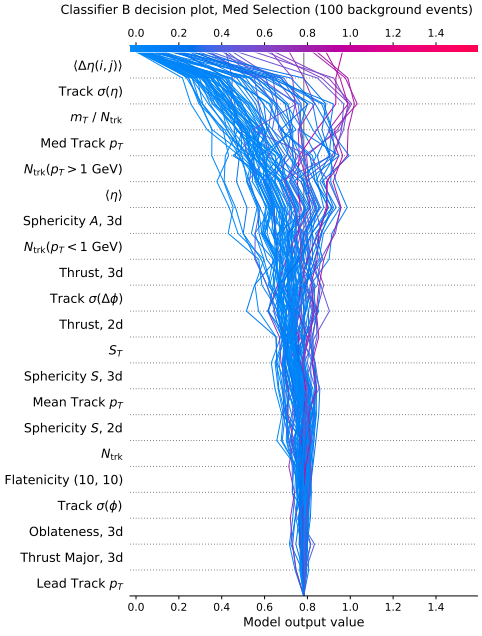
(a) (Classifier A, signal)



(b) (Classifier A, EPOS)

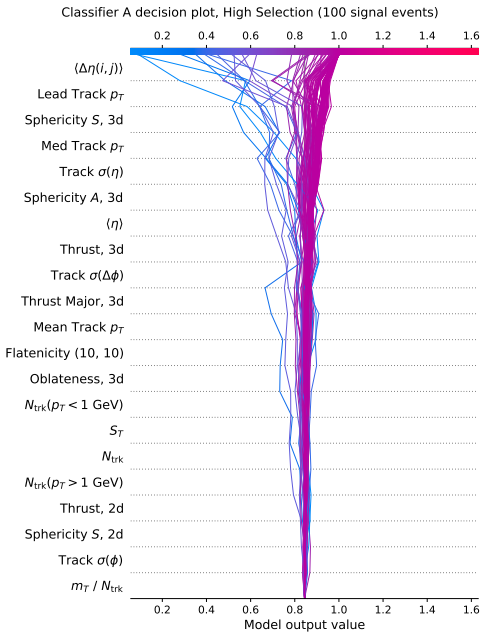


(c) (Classifier B, signal)

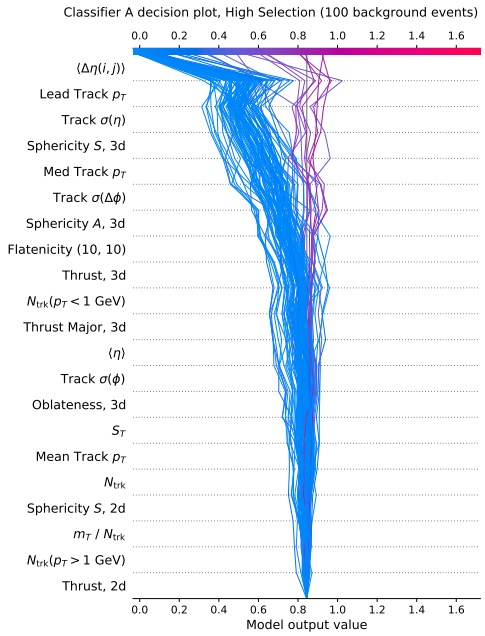


(d) (Classifier B, EPOS)

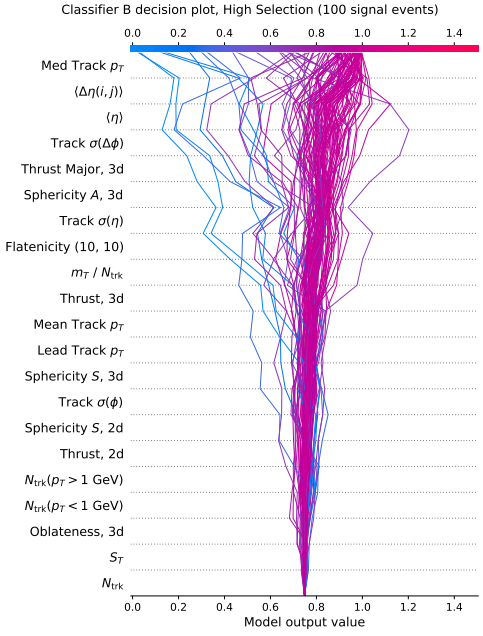
Figure 5.19: SHAP input variable-ranking plots in the medium-mass category. Each line corresponds to one event. The Shapley value of each variable (proportional to its importance for the classification result) is shown as a horizontal step from one variable to the next. The most important variables are at the top, and the least important at the bottom.



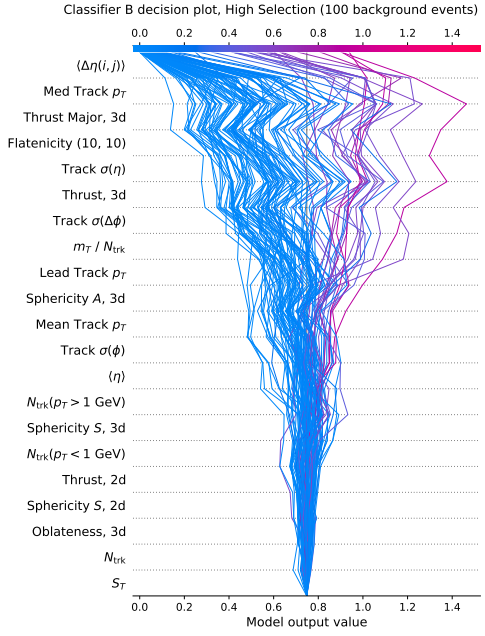
(a) (Classifier A, signal)



(b) (Classifier A, Epos)



(c) (Classifier B, signal)



(d) (Classifier B, Epos)

Figure 5.20: SHAP input variable-ranking plots in the high-mass category. Each line corresponds to one event. The Shapley value of each variable (proportional to its importance for the classification result) is shown as a horizontal step from one variable to the next. The most important variables are at the top, and the least important at the bottom.

Classifier output vs input histograms

The second strategy is complementary, and interrogates how the networks map input variables to classifier output values. This approach consists of making $2d$ histograms of classifier outputs vs inputs to look for functional dependence or the absence thereof. There is a multitude of these plots — one in each mass category ($\times 3$) for each of the signal and background samples ($\times 2$), for each classifier ($\times 2$), for each input variable ($\times \sim 20$), resulting in about 240 histograms, which is too many to provide here. A $2d$ histogram can be reduced by averaging along one of the axes to get a profile histogram. This has been done to produce the profile histograms of mean classifier outputs vs inputs shown in Figures 5.21, 5.22 and 5.23. Profiling enables us to compare the outputs of classifiers A and B between signal and Epos and Pythia 8 backgrounds at the same time.

Where the points of a given sample for a given classifier do not follow a pattern, or have large (or zero-size) error bars, the network is either confused or lacking in training statistics. Where there is zero error on a point that appears to be in the wrong place, it is most likely that that point is calculated as the mean of one single filled bin of classifier output (since the standard error on the mean of one number is zero).

Discussion of network performances

The plots introduced here all together give a comprehensive picture of how the ABCDisCo networks function. Some comments on the main features are made here.

From the profile plots of Figures 5.21 to 5.23, it appears that S_T does not contribute much to the classification since the mean classifier output is only a weak function of it. This therefore makes it surprising that S_T is ranked so highly in the decision plots of Figures 5.18 to 5.20. It is possible that S_T becomes highly-ranked through having an *adverse* effect on the classification. For Classifier A in the low and medium-mass

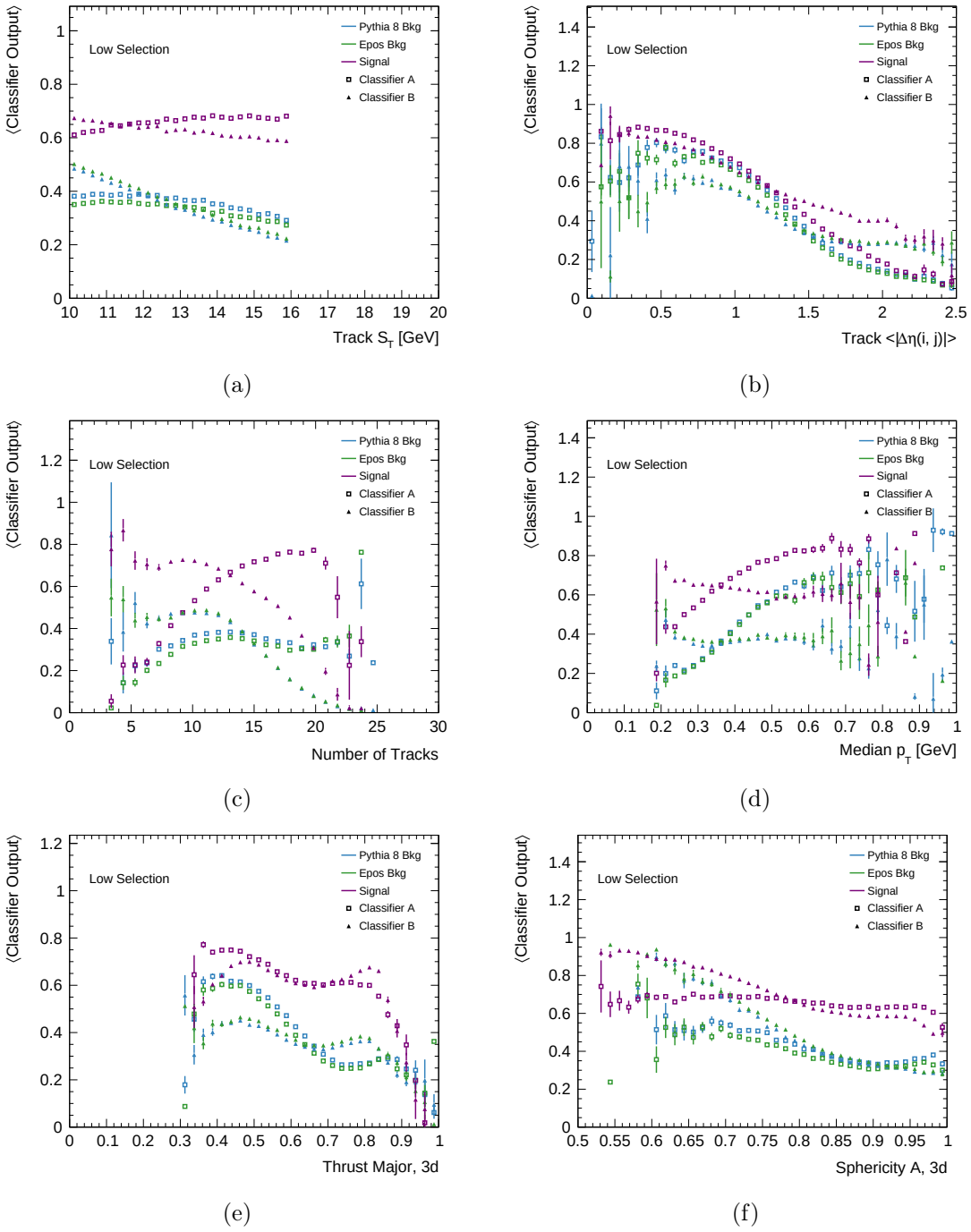


Figure 5.21: Profile plots showing mean classifier outputs as functions of classifier input variables in the low-mass category. Square markers show the distributions of Classifier A outputs, while solid triangle markers show the distributions of Classifier B outputs. The distribution of signal is shown in purple, EPOS background in green, and PYTHIA 8 background in blue. These plots are formed from $2d$ histograms by taking the mean classifier output in each bin of the input variable on the horizontal axis. The error bars show the standard error on the mean.

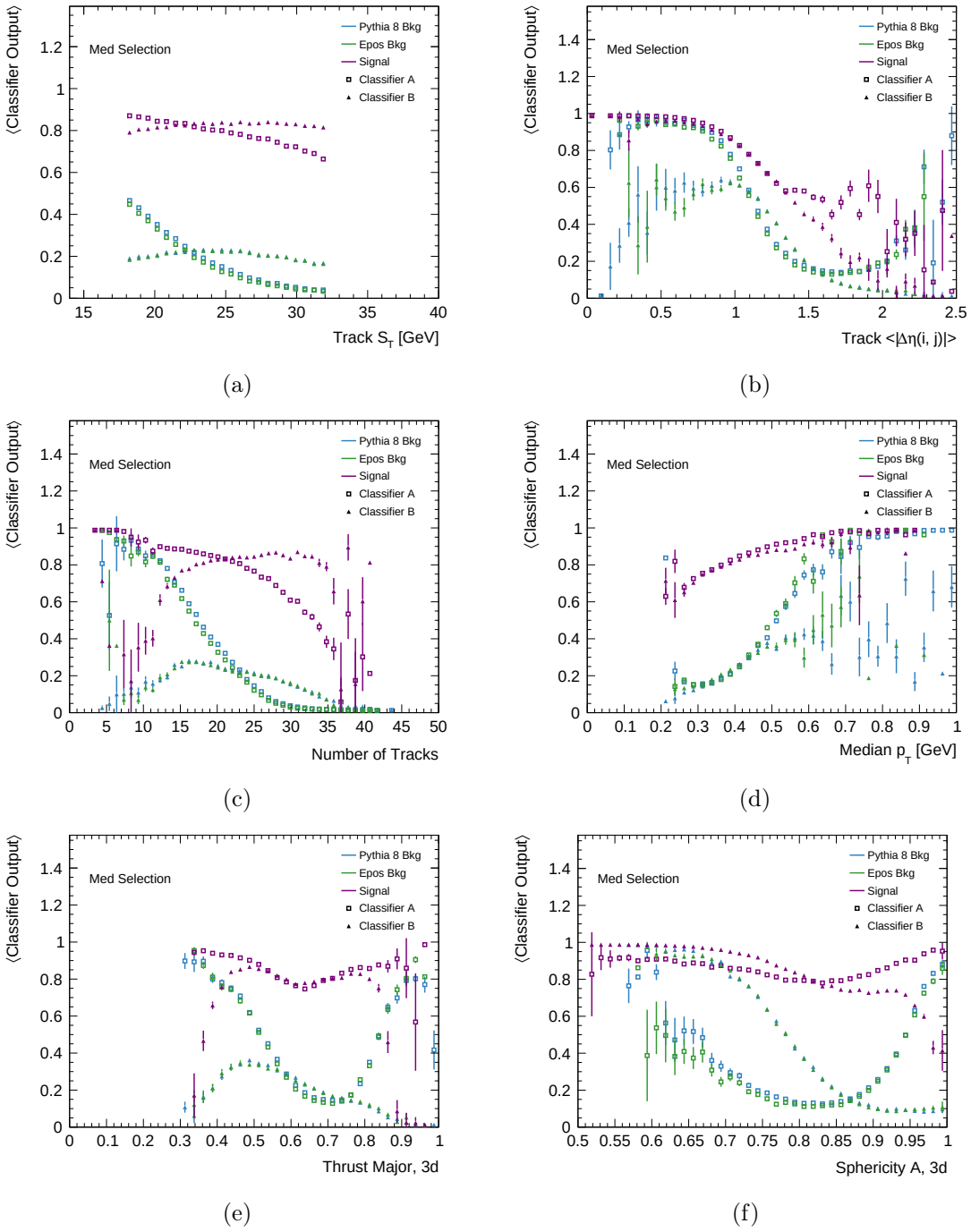


Figure 5.22: Profile plots showing mean classifier outputs as functions of classifier input variables in the medium-mass category. Square markers show the distributions of Classifier A outputs, while solid triangle markers show the distributions of Classifier B outputs. The distribution of signal is shown in purple, EPOS background in green, and PYTHIA 8 background in blue. These plots are formed from $2d$ histograms by taking the mean classifier output in each bin of the input variable on the horizontal axis. The error bars show the standard error on the mean.

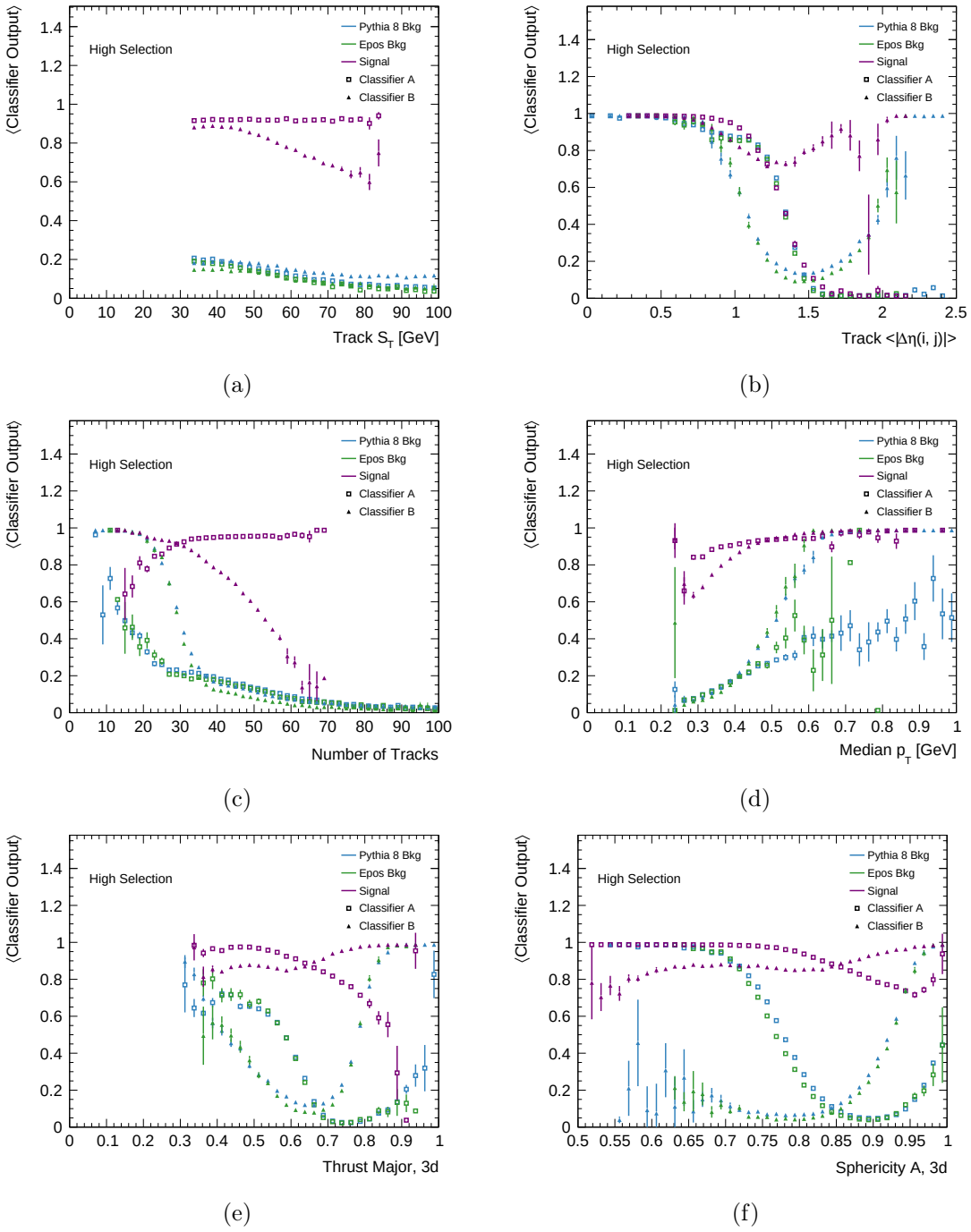


Figure 5.23: Profile plots showing mean classifier outputs as functions of classifier input variables in the high-mass category. Square markers show the distributions of Classifier A outputs, while solid triangle markers show the distributions of Classifier B outputs. The distribution of signal is shown in purple, EPOS background in green, and PYTHIA 8 background in blue. These plots are formed from $2d$ histograms by taking the mean classifier output in each bin of the input variable on the horizontal axis. The error bars show the standard error on the mean.

categories, this conclusion is consistent since S_T often produces the wrong effect in the decision plots, kicking the output in the direction of mis-classification in a large fraction of events. This indicates that it may benefit the classification to remove S_T from the inputs.

The most highly-ranked variable $\langle|\Delta\eta(i,j)|\rangle$ is clearly a functional dependence of the classifier outputs in every category, with lower values generally mapping to higher classifier outputs. The three networks use this variable in a similar way, except that Classifier B in the high-mass category also takes high values to be signal-like. Closer inspection of the medium-mass category shows that the sending of high values to high classifier outputs has already started there. Vertical dispersion in the profile plots occurs due to the classification effects of other input variables. Thus, that so little dispersion takes place in Figure 5.21b shows that $\langle|\Delta\eta(i,j)|\rangle$ is doing most of the classification work, which is the same statement as made by the corresponding decision plots.

The number of tracks is also used in a similar manner in the three categories (although this is concealed by the fact that the roles of Classifiers A and B are swapped in the medium-mass category). In every case, one of Classifiers A and B interprets low-N_{trk} events as signal-like, while the other interprets high-N_{trk} as signal-like. The networks learn that higher values of median p_{T} are more signal-like.

The thrust-major and sphericity-A plots in Figures 5.21 to 5.23 show that the networks have learned that the Event Shape distributions of instantons are broader than those of the background. Thus the network maps central values of the Event Shapes to background classifier values, and values from the extremities to signal. In the low-mass category, the network has even noticed the bimodality of the thrust-major distribution, as reflected in Figure 5.21e.

These features of the networks seem reasonable and agree with the features of the discriminating variables introduced in Section 5.2.

It is possible that the DNNs have learned to discriminate differences between the different parton shower, hadronisation, and UE algorithms employed by EPOS LHC and SHERPA 3. This specialisation cannot be avoided unless the same generator and settings for these components are used for both signal and background simulation. Ultimately, the DNNs' inappropriate use of specific aspects of the modelling (not corresponding to real data) must be captured as systematic uncertainties, as will be discussed in Section 6.4.2.

5.3.7 ABCD background estimation

Here, the ABCDisCo method described above is connected to the traditional ABCD method of background estimation.

The ABCD method is a classic approach to data-driven background estimation and is used extensively at the LHC. It depends on identifying two statistically independent variables f and g. Thresholds on these variables are used to partition the $f-g$ plane into four regions, labelled A, B, C, and D. When A is adjacent to B and C in this 'ABCD' plane, the background in region A can be estimated according to

$$N_{BCD} = \frac{N_B \times N_C}{N_D}, \tag{5.15}$$

where N_i is the yield of data events in region $i \in \{B, C, D\}$, and N_{BCD} is the estimated background in region A.

The ABCD planes constructed by our ABCDisCo networks are spanned by Classifier A and B outputs. Signal and Control Regions are defined as shown in Figure 5.24, where the labels SRA, CRB, CRC, and CRD are defined. To specify the mass category under consideration, these ABCD regions are suffixed with 'Low', 'Med', and 'High' as appropriate. The region definitions have not been significantly optimised, the hope being that the analysis is kept robust against modelling inadequacies involved in

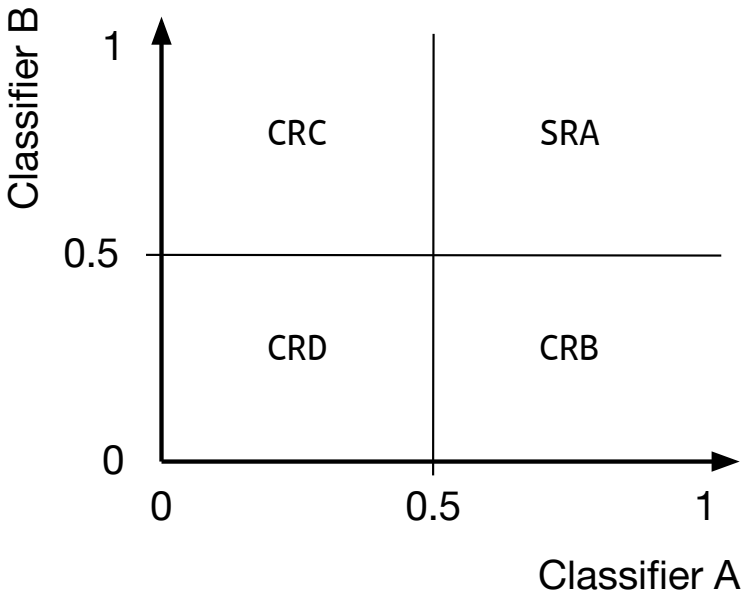


Figure 5.24: Schematic showing the definition of the A, B, C, and D regions in the 'ABCD' plane spanned by Classifier A and Classifier B outputs. These regions are prefixed by SR or CR to indicate their use as Signal or Control regions, respectively. To indicate the mass category in which a region is defined, the suffixes 'Low', 'Med', and 'High' are used.

the training and misconceptions about the details of the signal. Rather, the regions are defined by cuts at 0.5 on classifier output scores. The value 0.5 is privileged by the ML algorithms employed, which use it as the boundary between the signal and background classes during the training. In this sense, the cut is already optimised.

Relation (5.15) holds to the extent that the decorrelation of the Classifier A and B outputs is accomplished. In judging the performance of this background estimation, a central quantity is its closure[4], defined as $k = (N_{BCD} - N_A)/N_A$. This is the quantity that was represented by k in Eq. (5.10) for optimisation of the α-parameter, where the same region definitions as given here were used. Closure will be used again in the treatment of systematic uncertainties in Section 6.4.

[4]Also called 'non-closure', depending on context.

5.3.8 Blinding strategy

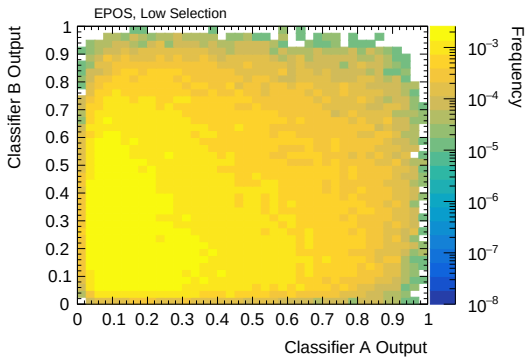
To prevent biasing the analysis design based on expectation or towards statistical fluctuations, it is customary to blind the data in signal regions. This custom is followed in this analysis, and the data in the SRAs are not looked at. The deployment of the blinding can be seen in the $2d$ classifier output distributions of Figure 5.25 as the absence of data events from the top right-hand corners of the ABCD planes. Once the analysis reaches a certain stage of approval by the ATLAS Collaboration, the data may be 'unblinded', and the results can be known.

5.3.9 Background model validation

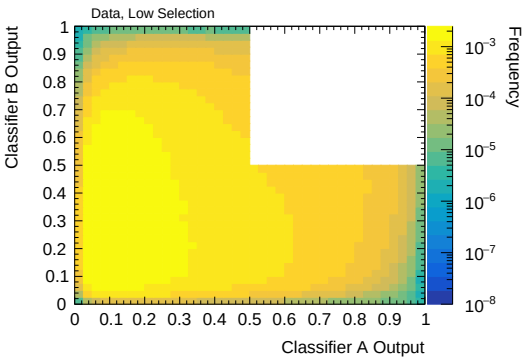
With the blinding strategy established, the agreement between data and background MC can be scrutinised in the CRs .

This opportunity is used to justify *a posteriori* the use of the EPOS MC for training the ABCDisCo networks. For the networks to learn about real background events, certain aspects of the background modelling, in particular the correlations between network input variables, must be good. However, the background modelling does not have to be perfect. The total normalisation does not matter, since the training does not see this, and some mis-modellings of shape can be forgiven so long as they are covered by systematic uncertainties. Ultimately, the background estimation is entirely data-driven, so the EPOS background model enters the analysis only through the definitions of the variables $f(X)$ and $g(X)$.

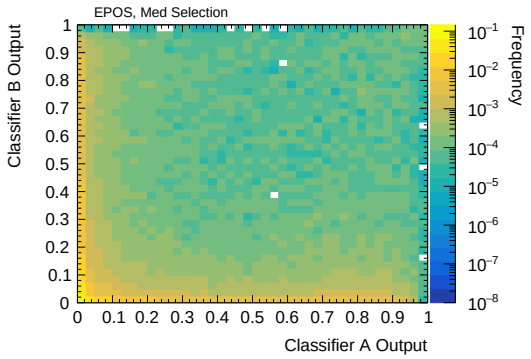
In Figures 5.26 to 5.28, a selection of six variables is used to judge the background modelling in each of the three mass categories. The shape of the track-S_T spectrum is quite well-modelled, with the data–MC ratio being quite flat, except perhaps in the high-mass category where it becomes increasingly over-predicted towards higher values. The ratio is offset to between 0.9 and 0.6, but this is not too important since



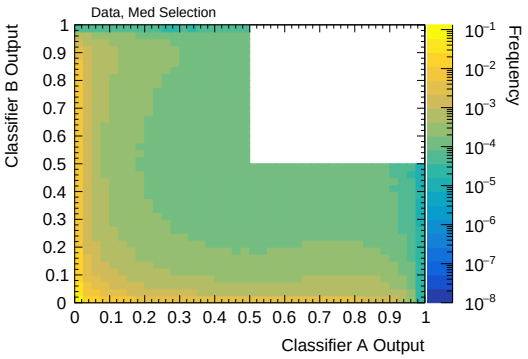
(a) Epos, low-mass category



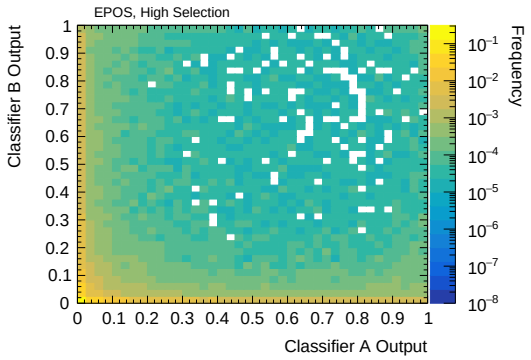
(b) Data, low-mass category



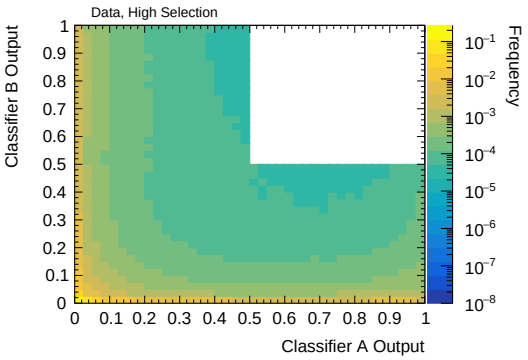
(c) Epos, medium-mass category



(d) Data, medium-mass category



(e) Epos, high-mass category



(f) Data, high-mass category

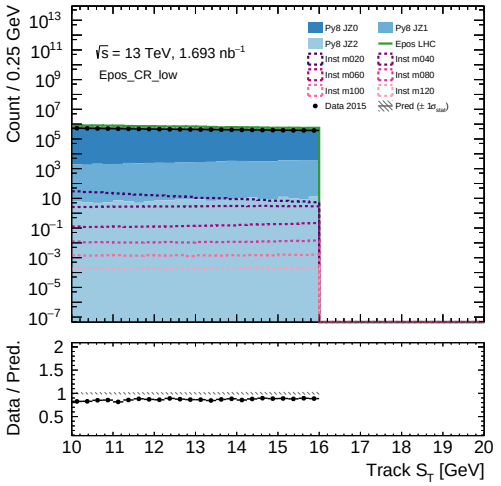
Figure 5.25: ABCD planes in the low, medium, and high mass categories respectively, with the distributions of the Epos model in the left column, and data in the right. The empty regions in the data distributions correspond to the blinded signal regions, SRA.

this overall normalisation discrepancy is corrected as part of the fit performed in Chapter 6.

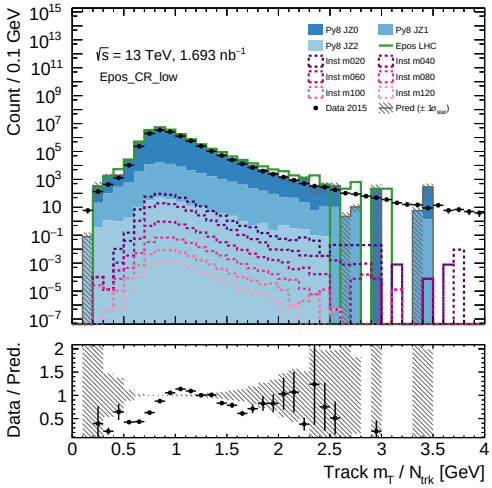
The variables with mass dimensions, m_T/N_{trk} and median-p_{T}, are generally the most poorly modelled. The modelling of median-p_{T} is particularly bad, to the extent that it may have been better not to enlist it as an input to the networks. The angular variables are not too bad at low and medium masses, but the existing pattern of mis-modelling is emphasised at high mass. It is one of the virtues of the Event Shape variables that they are well-modelled [115], and we see that here in the example of the $3d$ thrust distributions.

Mis-modellings of the individual input variables aggregate into mis-modellings of the classifier output distributions, which are shown in Figure 5.29. The shape modelling improves with increasing mass, until the modelling is really quite good in the high-mass category. However, there are unfortunately clear slopes in the ratios in the low and medium-mass categories. These discrepancies ought to manifest as larger systematic uncertainties in these regions.

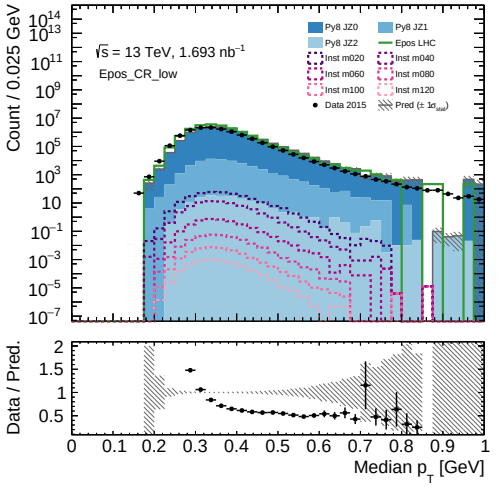
At this stage, the analysis design is established, and we can proceed to make statistical statements about its sensitivity.



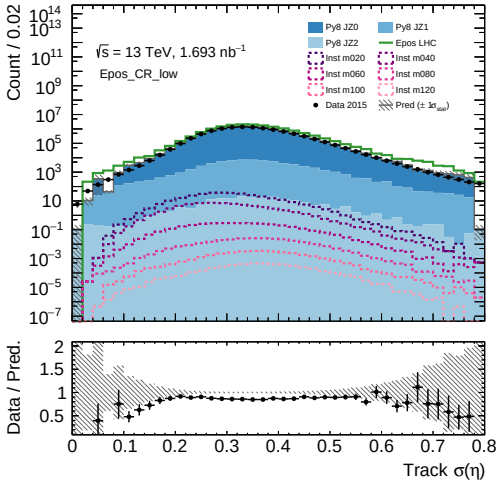
(a)



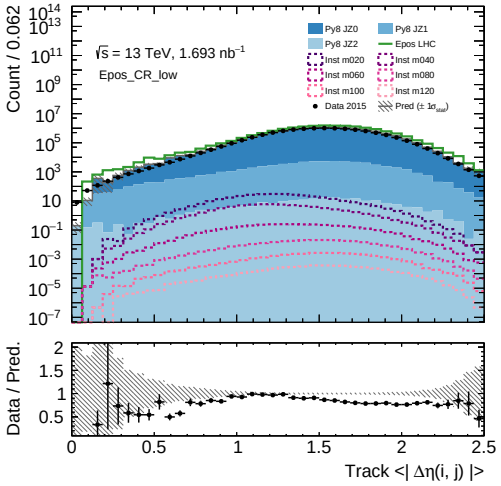
(b)



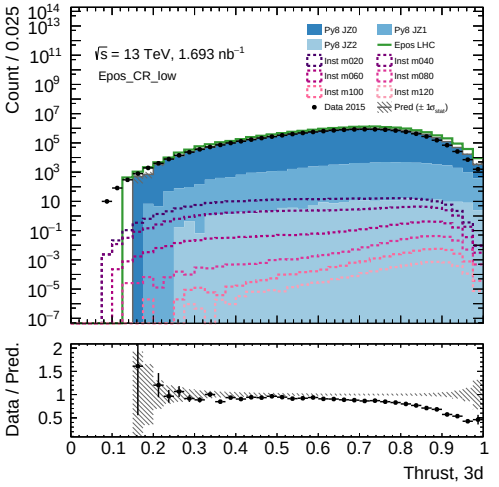
(c)



(d)

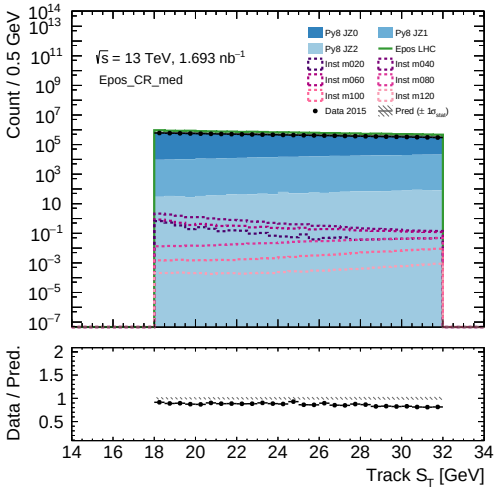


(e)

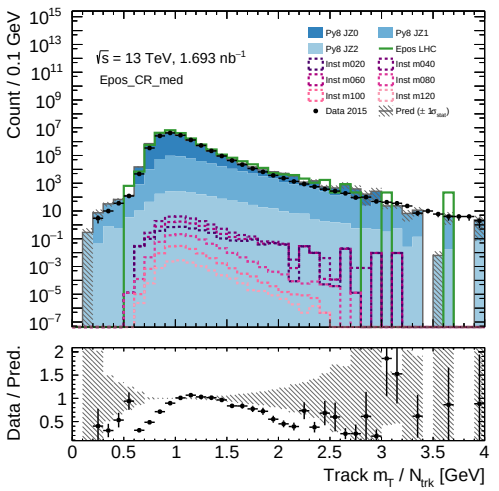


(f)

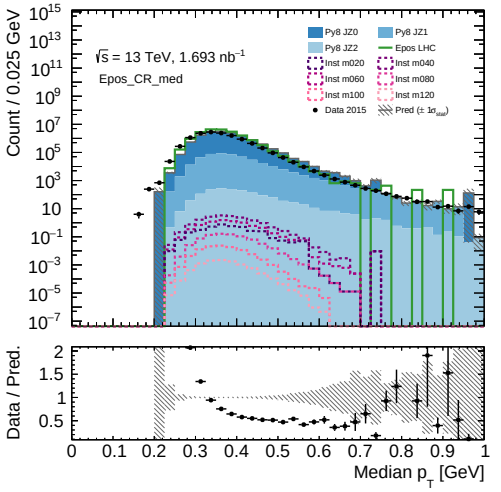
Figure 5.26: Distributions of signal, background, and data events in a selection of variables in the B, C, and D CRs of the low-mass category combined. The PYTHIA 8 model is taken for the denominator of the ratio shown in the bottom panel of each plot. The grey band corresponds to the statistical uncertainty on the PYTHIA 8 model.



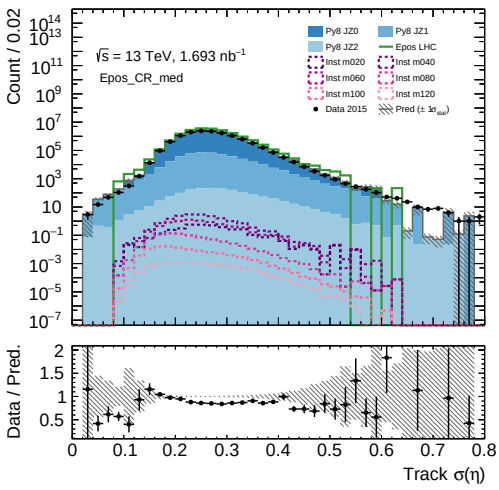
(a)



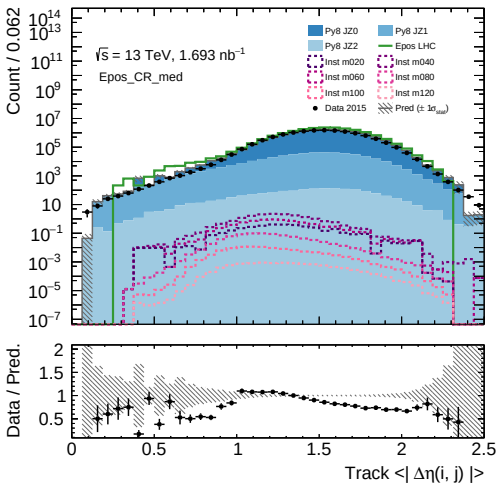
(b)



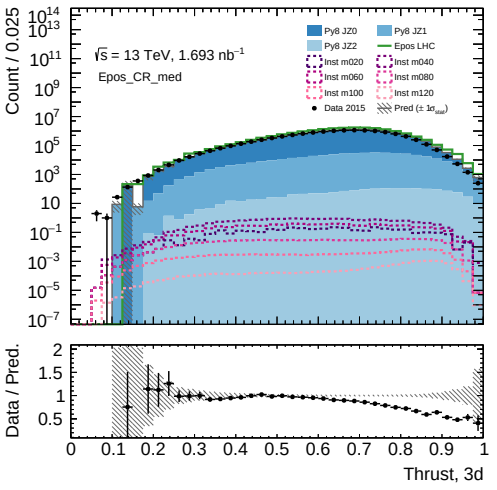
(c)



(d)



(e)



(f)

Figure 5.27: Distributions of signal, background, and data events in a selection of variables in the B, C, and D CRs of the medium-mass category combined. The PYTHIA 8 model is taken for the denominator of the ratio shown in the bottom panel of each plot. The grey band corresponds to the statistical uncertainty on the PYTHIA 8 model.

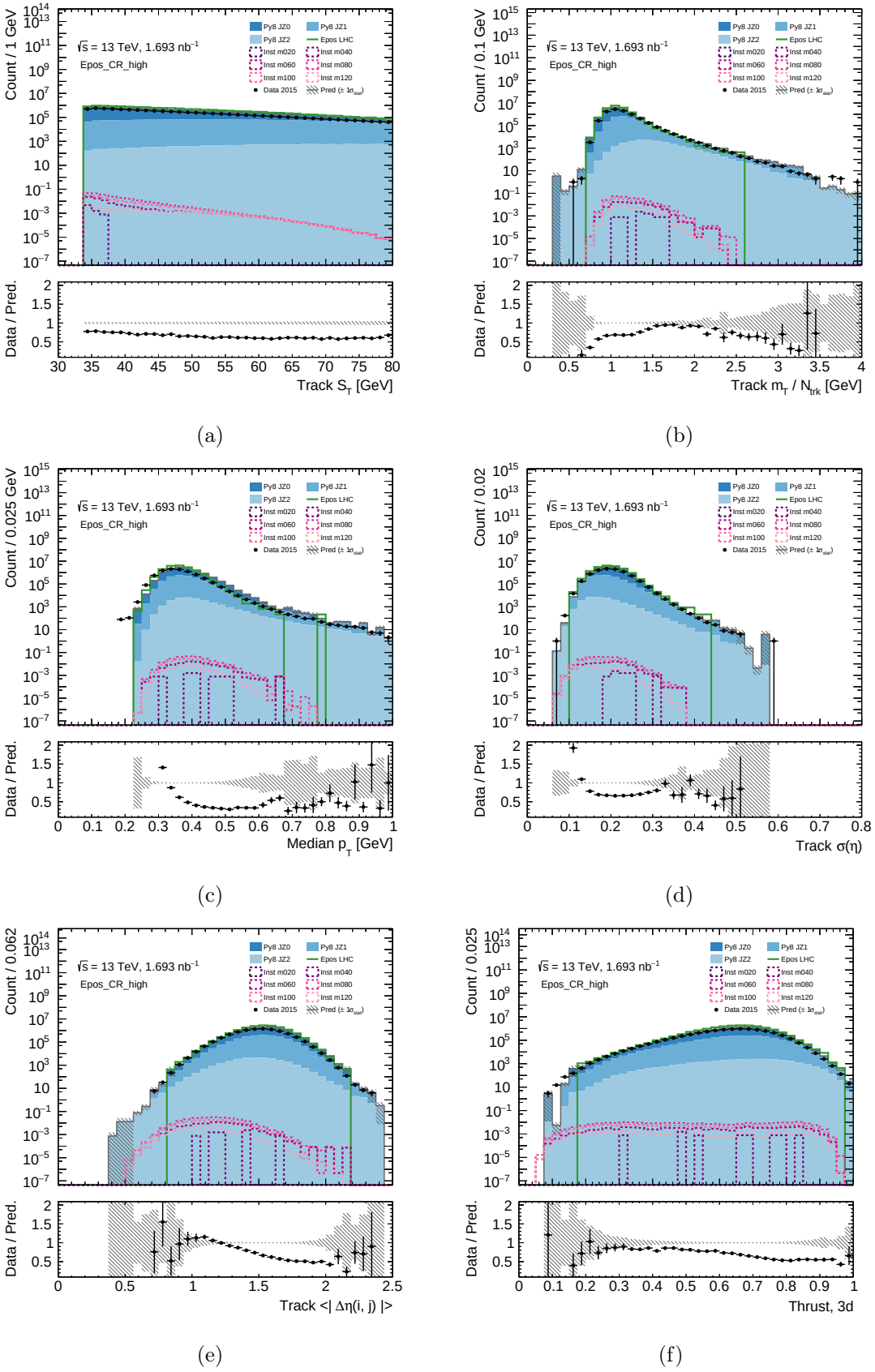


Figure 5.28: Distributions of signal, background, and data events in a selection of variables in the B, C, and D CRs of the high-mass category combined. The PYTHIA 8 model is taken for the denominator of the ratio shown in the bottom panel of each plot. The grey band corresponds to the statistical uncertainty on the PYTHIA 8 model.

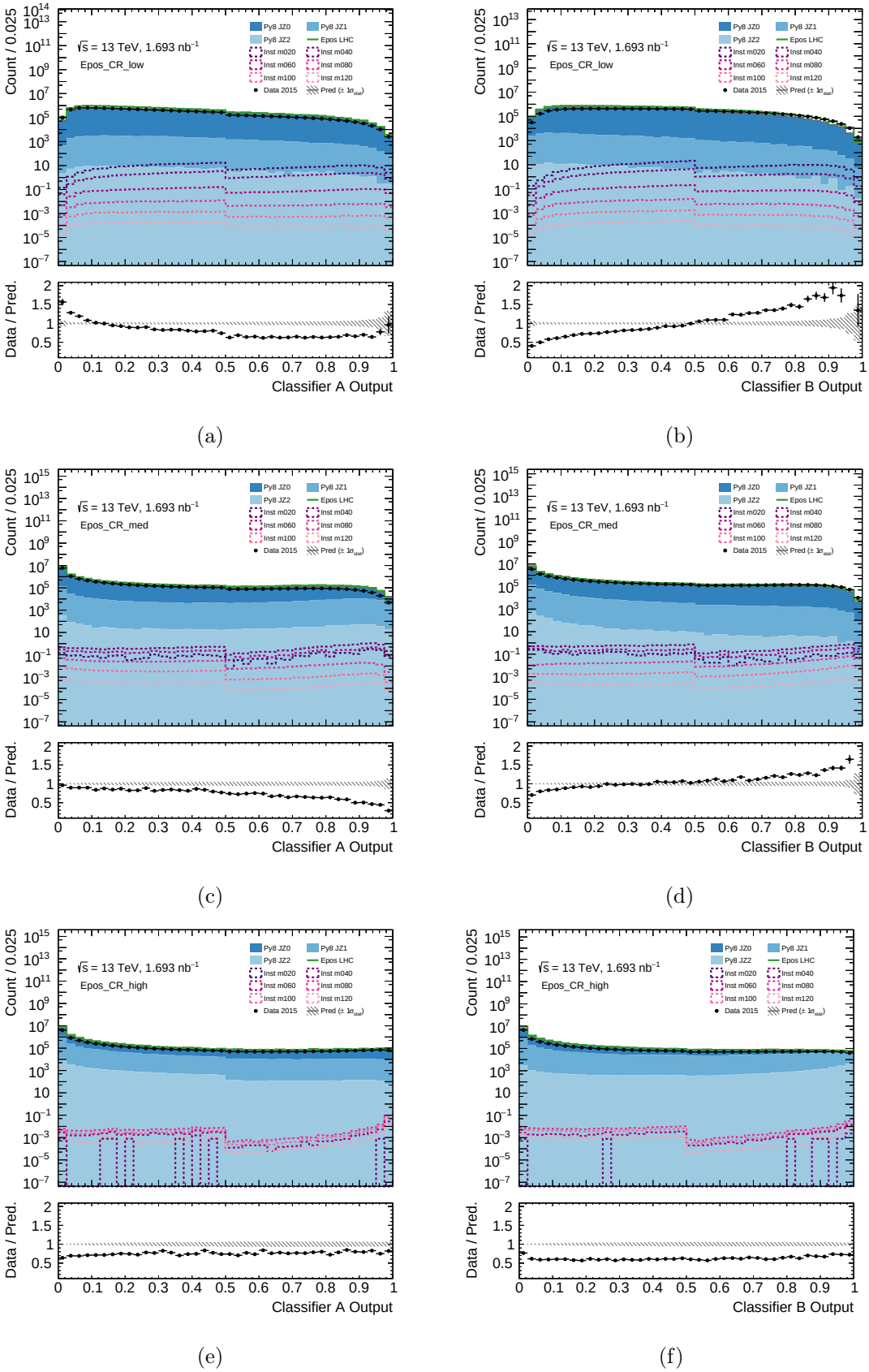


Figure 5.29: Distributions of signal, background, and data events in Classifier A (left) and B (right) outputs in the B, C, and D CRs combined, in the low, medium, and high-mass categories respectively from top to bottom. Pythia 8 is taken for the denominator of the ratio shown in the bottom panel of each plot. The grey band corresponds to the statistical uncertainty on the Pythia 8 model.

Chapter 6

Statistical Analysis

Whereof one cannot speak, thereof one must be silent.

— Wittgenstein

The search for a new physics process can be formulated as a statistical hypothesis test. The statistical approach followed here for hypothesis testing is based on Refs. [137–139]. In Ref. [137], the likelihood ratio at the heart of the formalism is introduced. In Ref. [138], the likelihood ratio is incorporated into a broader analysis strategy comprising signal and control regions. Ref. [138] gives approximations for finding the distributions of test statistics, and introduces the notion of the Asimov dataset used in this thesis prior to data unblinding. In the case of exclusion, a different test statistic called CL_s is prescribed based on Ref. [139]. We use the HistFitter framework [140] for implementing these methods.

This is all part of ATLAS's standard toolkit for searches, and is summarised here in Sections 6.1, 6.2 and 6.3. The systematic uncertainties entering the analysis are discussed in Section 6.4, before blinded results are presented in Section 6.5.

6.1 The Profile Likelihood Ratio

The profile likelihood function is defined as

$$L(\mu, \boldsymbol{\theta}) = P_{\text{SR}} \times P_{\text{CR}} \times C_{\text{syst}} \tag{6.1}$$

$$= \prod_{i=1}^{N} \text{Poiss}(n_i \mid \lambda_i(\mu, \boldsymbol{\theta})) \times \prod_{j=1}^{M} \text{Poiss}(m_j \mid \lambda_j(\mu, \boldsymbol{\theta})) \times C_{\text{syst}}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) \tag{6.2}$$

for a strategy involving N signal regions and M control regions, and where the yields n_i of signal and m_i of background events in the regions are governed by Poisson statistics. The likelihood is written as a function of μ, the signal strength defined as the ratio of the fitted signal cross-section to the nominal, and $\boldsymbol{\theta}$, a vector of nuisance parameters. Nuisance parameters represent systematic uncertainties and other fit parameters, and enter the likelihood through the constraint term $C_{\text{syst}}(\boldsymbol{\theta}^0, \boldsymbol{\theta})$.

The variations allowed by systematics are modelled as Gaussian distributions of unit width where the fitted value of the k^{th} systematic uncertainty, θ_k, is allowed to float around its central value, θ_k^0 which is taken from dedicated auxiliary measurements. When the systematic uncertainties are independent, the constraint term C_{syst} is simply a product of all the individual uncertainties

$$C_{\text{syst}}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \prod_k G(\theta_k^0 - \theta_k).$$

Auxiliary measurements can be provided by the physics working groups responsible for the physics objects of ATLAS, or can come from dedicated studies custom to this analysis such as those detailed in Section 6.4.3 in relation to the background estimation. The presence of nuisance parameters broadens the profile likelihood as a function of μ, reflecting the loss of information about μ as a result of systematic uncertainties. However, the total systematic uncertainty can be reduced since correlations between the uncertainties can be inferred during the fitting procedure, and uncertainties that do not affect (those with $\theta_k^0 - \theta_k \approx 0$) are effectively removed.

The profile likelihood ratio is then defined as

$$Q(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{6.3}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are unconditional maximum-likelihood estimators of the parameters μ and $\boldsymbol{\theta}$, and $\hat{\hat{\boldsymbol{\theta}}}$ is a maximum-likelihood estimator conditional on a fixed μ (and is therefore a function of μ). This is the ordering principle put forth by Feldman and Cousins [137] because it is not the absolute likelihood of a signal strength μ given the data that matters, but rather the likelihood of a hypothesised μ compared to the likelihood of the data given the best-fit μ. The basic test-statistic, sometimes referred to as the negative log-likelihood, is then defined as

$$q_\mu = -2 \ln Q(\mu), \tag{6.4}$$

where the subscript on q refers to the hypothesis being tested. The '$-2 \ln _$' part is inserted due to results of Wilks [141] and Wald [142] stating that in the case of one parameter of interest (POI, in this case μ), the distribution of the test statistic approaches a chi-square distribution. After the transformation of Eq. 6.4, greater values of the test statistic correspond to greater incompatibility between the data and μ.

One is usually interested in positive values of μ only since a signal process cannot subtract from the event yield, but in the fitting of Eq. 6.3, $\hat{\mu}$ is allowed to take on negative values such that it may be modelled as a Gaussian distribution to facilitate approximating the test statistics [138].

For discovery, q_μ is restricted to

$$q_0 = \begin{cases} -2 \ln Q_0 & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \tag{6.5}$$

because best-fit values of μ that are less than zero should not be used to accommodate under-fluctuations of the background or poor background modelling. In other words, the data are considered to show a lack of agreement with the background-only hypothesis only in the case $\hat{\mu} \geq 0$.

For exclusion, the analogous restriction is

$$q_\mu = \begin{cases} -2 \ln Q(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{6.6}$$

because best-fit values of μ that are greater than the posited value of μ do not indicate less compatibility between the data and μ.

6.2 Hypothesis Testing

For testing a hypothesised value of μ in terms of the test statistic q_μ, a p-value is calculated according to

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu) dq_\mu, \tag{6.7}$$

which requires the probability distribution function (pdf) of the test statistic, $f(q_\mu|\mu)$. In $f(q_\mu|\mu)$, the subscript on q still denotes the hypothesis being tested, while the second argument of f is the value of μ assumed in the data. For discovery, the null hypothesis that stands to be rejected is the background-only hypothesis which occurs at $\mu = 0$, so we need $f(q_0|0)$. In the absence of discovery, we resort to setting upper limits which involves testing values of μ greater than 0, for which we need $f(q_\mu|\mu)$. Conventionally, the p-value is converted into a one-sided Gaussian significance, Z.

For characterising the expected sensitivity of an experiment, it is conventional to consider the distribution of q_μ of a given signal strength, μ, assuming data distributed according some other signal strength, μ', specifically $f(q_0|\mu')$ for discovery sensitivity and $f(q_\mu|\mu')$ for exclusion sensitivity. Here one can imagine testing q_0 or q_μ under

the assumption of a dataset in which signal has been 'injected' at a signal strength μ'. For quoting expected significance, it is not the literal expected significance given by inverting the expected p-value using $Z(\langle p \rangle)$ that is used, because this relation is non-linear ($\langle Z(p) \rangle \neq Z(\langle p \rangle)$). Instead, the *median* significance is quoted, since this relation is monotonic (med $Z(p) = Z(\text{med } p)$). The median significance is then regarded as the significance that is expected of the experiment before unblinding. 'Median significance' and 'expected significance' are used interchangably with this meaning.

6.2.1 CLs

Conventionally, a different test statistic called called CL_s [139] is used for setting exclusion upper limits. CL_s is defined as

$$\text{CL}_\text{s} = \frac{\text{CL}_\text{s+b}}{\text{CL}_\text{b}}, \tag{6.8}$$

where

$$\text{CL}_\text{s+b} = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu')dq_\mu, \tag{6.9}$$

$$\text{CL}_\text{b} = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_0|\mu')dq_0. \tag{6.10}$$

Here $f(q_\mu|\mu')$ is the pdf of the signal plus background test statistic, and $f(q_0|\mu')$ is the pdf of the background-only test-statistic, assuming that signal is present in the data at a strength μ'. CL_s is not formally a confidence (it is a ratio of confidences) because it does not give exact frequentist coverage. CL_s can be regarded as a re-normalisation of the p-value obtained from the profile-likelihood ratio, where the re-normalisation is designed to make it less sensitive to fluctuations of the background.

When the data are more compatible with the background-only hypothesis, $\text{CL}_\text{b} \to 0.5$ and $\text{CL}_\text{s+b} \to 0$, such that small values of CL_s result. When the data are more

compatible with the signal plus background hypothesis, $CL_b \to 1$ and $CL_{s+b} \to 0.5$, such that larger values of CL_s result.

6.3 Likelihood Configurations

Using HistFitter, we can perform three types of hypothesis test.

Discovery fit. For discovery, the null hypothesis is that the data consist of background events only. Thus a background-only fit is performed across the regions, with the signal contamination in the CRs assumed to be negligible. Then the null hypothesis is tested by evaluating the p-value resulting from Eq. (6.7) and Eq. (6.5). Injecting $\mu' = 1$ in the SR for the purpose of calculating median significance allows us to see if the experiment is sensitive to discovery if the signal is present at the nominal rate.

Model-dependent upper limits. For setting upper limits on our signal model, we start with a signal model-dependent fit of the signal plus background model across all regions. Assuming a dataset comprising signal plus background, we scan the signal strength parameter μ, evaluating CL_s until it falls below a threshold of 0.05, which corresponds to 95% CL_s. At the point where CL_s drops below 0.05, the upper limit is placed in terms of μ, which can be converted into an absolute value of the signal cross-section. At each point of the scan, the p-value is calculated using Eq. (6.6) in Eq. (6.7).

Model-independent upper limits. For setting model-independent limits on the possible number of BSM events, we use the same fit configuration as for the signal model-dependent results, but use a dummy signal model constructed as a histogram of one entry in the SR. Only a single one-bin SR can be considered in this set-up, since a combination of more bins or SRs requires assumptions about the shape of the signal model.

In the event of discovery, the signal strength can be measured in the signal model-dependent fit configuration. Prior to unblinding, Asimov data defined in Ref. [138] are used instead of the observed data in the SR. Signal can be injected into the Asimov dataset at any desired signal strength μ' to test the expected sensitivity of the experiment. In using this Asimov dataset, the same exact statistical tests are used as would be used after unblinding. The performance of the Asimov data can be cross-checked using pseudo-experiments, but this is not done because the statistics of the dataset used easily satisfy the law-of-large-numbers assumption needed for Asimov dataset generation.

6.3.1 One-bin counting experiments

Based on these possible likelihood configurations, this analysis performs simple one-bin counting experiments. In each mass category, the background is predicted through the ABCD equation in a separate study, and the predicted background yields in the single-bin SRs are exported for the statistical analysis. The one bin counting experiments then involve performing the statistical tests discussed in this chapter on the yields in these single-bin SRs.

The CRs are not included in the likelihood definition. Our background sample consists of one process only, and its normalisation is taken directly from the data through the ABCD equation. The normalisation of the background does not have to be adjusted to accommodate signal contamination, because signal contamination in the CRs is found to be negligible up to large signal strength values.

In the statistical combination, the single common signal strength parameter is fitted across the three mass categories. The fit being performed involves only three bins, one for each of the three mass-category SRs.

6.4 Systematic Uncertainties

Several sources of systematic uncertainty enter the analysis through track reconstruction, background estimation, and theory modelling. Theory modelling is particularly important in the signal model-dependent fit configuration where the signal MC enters. This section discusses the origins of these uncertainties and how they are assessed. An exhaustive list of the affecting systematics is given, but only the dominant ones are considered in producing the main results of this thesis. It will be argued that the neglected systematics will not impact the results.

6.4.1 Experimental uncertainties

All of the experimental systematic uncertainties considered are due to the track reconstruction, except for one additional flat uncertainty of 1.9 % on the integrated luminosity (measured using methods described in Ref. [89]). Track reconstruction systematics originate in (i) our lack of knowledge of the layout of material in the detector, (ii) our lack of knowledge of particle interactions in that material, and (iii) our resolution of the IP. These are assessed by varying the efficiency of track reconstruction, impact parameter resolution, and the alignment of the ID as described here. Systematic uncertainties associated with track reconstruction are typically at the level of 5%, and can be neglected with respect to other sources of uncertainty in our treatment. Tracking uncertainties will nevertheless have to be incorporated eventually, so they are introduced here despite not being used in this thesis.

Track reconstruction efficiency can be regarded as the 'probability of reconstructing a track when there is an underlying charged-particle', and can differ between data and MC. The difference arises principally from our lack of knowledge of the distribution of material in the ID, and from mis-modellings of the physics used in detector simulation.

These uncertainties are assessed by producing alternative MC samples with variations of the material description and with a different physics list [93]. Tracking efficiency is measured in MC simulation as the fraction of reconstructed tracks that can be matched to truth particles, $\epsilon(p_{\mathrm{T}}, \eta)$. Variations are applied by dropping tracks according to the probability that they were *not* reconstructed in their (p_T, η) bin.

Impact paramter resolution. Systematic uncertainties are assigned to the transverse (d_0) and longitudinal (z_0) impact parameters (IPs) to account for the difference in IP resolution between data and MC. IP resolution is difficult to grasp because it is convolved with the resolution of the Primary Vertex. As a first step towards systematic variations, the IP resolution is deconvolved from the PV resolution using an iterative procedure based on the 'split-vertex algorithm' described in Ref. [143]. Deconvolution yields $\sigma(d_0)$ and $\sigma(z_0)$.

To assess the impact of the systematic uncertainty in IP resolution, the IPs of individual tracks are smeared in MC by replacing d_0 (z_0) with a number drawn from a Gaussian distribution with mean d_0 (z_0) and variance $\sigma^2_{\mathrm{smear}}$, where $\sigma^2_{\mathrm{smear}} = \sigma^2_{\mathrm{data}} - \sigma^2_{\mathrm{MC}}$ for d_0 and z_0 as appropriate.

Residual alignment uncertainties. Alignment weak modes (deformations under which track fit-χ^2 is invariant) result in biases on track parameters [144]. Following precision alignment of the ID, which is performed run-by-run during the calibration loop [88], these biases can be corrected during bulk processing of the physics streams. The correction amounts are uncertain, however. To reflect this, systematic variations are produced by biasing the affected track parameters again by the uncertainty in the corrections, thus effectively producing a mis-alignment in the MC.

6.4.2 Theory modelling

The standard prescription for the systematic uncertainty on the theory of a perturbative process is given here. Systematic uncertainties are produced by performing

reasonable variations of the event generation and comparing to the nominal. The recommended variations are either provided by the MC event generator, or must be produced as two-point differences of different configurations.

- **Scale variations.** The variations of the renormalisation and factorisation scales μ_{r} and μ_{f} used in the matrix element by factors of two are applied to provide an estimate of the uncertainty due to missing higher orders in the calculation of the perturbative cross-section. The convention is to perform the seven-point variation

$$\{\mu_{\mathrm{r}}, \mu_{\mathrm{f}}\} \times \{\{0.5, 0.5\}, \{1, 0.5\}, \{0.5, 1\}, \{1, 1\}, \{2, 1\}, \{1, 2\}, \{2, 2\}\},$$

  where $\{1, 1\}$ corresponds to the nominal. The set of variations is combined by taking their envelope, i.e. $\max\{\mathcal{O}(\mu_{\mathrm{r},i}, \mu_{\mathrm{f},i}) - \mathcal{O}(\mu_{\mathrm{r},0}, \mu_{\mathrm{f},0})\}$ in each bin, where i and 0 refer to the i^{th} variation and the nominal, respectively.

- **Parton-shower scale variations** are variations of the renormalisation and factorisation scales used in the parton shower, and are applied in tandem with the scale variations of the matrix element (i.e. the matrix element and parton shower scales are varied upwards together, and downwards together).

- **PDF fit variations** encompass experimental uncertainties that enter through the dataset used in the fit, and the choice of PDF parameterisation. In the case of the NNPDF set used in our signal samples, there is no error PDF set but rather an ensemble of PDF sets. The mean of this ensemble is taken as the best PDF, and the standard deviation of the ensemble as the uncertainty according to

$$\delta\mathrm{PDF} = \sqrt{\frac{1}{N}\sum_{i}(\mathrm{PDF}_i - \mathrm{PDF}_0)^2}, \tag{6.11}$$

  where $N = 100$ is the number of elements in the ensemble.

- **PDF** α_S — the strong coupling, α_S, used in the fitting of the PDF affects

process cross-sections. To account for this, the PDF fit is repeated at upward and downward variations of α_S. The resulting uncertainty on the cross-section σ is defined as

$$\delta\sigma = \frac{\sigma(\alpha_S^{\text{down}}) - \sigma(\alpha_S^{\text{up}})}{2}, \tag{6.12}$$

which must be combined in quadrature with the other PDF uncertainties. The nominal NNPDF set is produced at $\alpha_S = 0.118$, and the $\alpha_S^{\text{down}} = 0.117$ and $\alpha_S^{\text{up}} = 0.119$ variations are used.

- **Algorithmic uncertainties** due to the choice of algorithms for the parton showering and hadronisation are accounted for by taking the difference of two samples produced with different parton shower and hadronisation algorithms but all else the same. This should provide a handle on the possible mis-modellings of these parts of the event. In the case of our signal, we can use SHERPA's interface to the PYTHIA 8 fragmentation model to produce the alternative sample while keeping all the details of the matrix element the same.

Once these variations are obtained and combined as described, they are combined in quadrature to yield the total 'theory' systematic which affects our upper limits on the signal cross-section.

Instanton signal-specific uncertainties

Instanton event phenomenology is not particularly well understood, as emphasised e.g. in Ref. [56], so SHERPA provides the facility to apply some variations of the event generation to capture systematic uncertainties. There is no clear guidance on how to extract an uncertainty from these variations, so a number of possible variations are explored. Variations are steered using the run-card flags listed here.

- `INSTANTON_NGLUONS_MODIFIER`. In the table of cross-section data supplied to the run card, there is a column for the mean of the number of gluons $\langle N_{\text{gluons}} \rangle$

being emitted at a given instanton mass. This number is calculated along with the cross-section [51], and the number of gluons produced in a single event at a given mass is taken to be distributed like a Poissonian. The `INSTANTON_NGLUONS_MODIFIER` parameter changes the number of gluons according to

$$N_{\text{gluons}} = \text{Poiss}(\texttt{INSTANTON_NGLUONS_MODIFIER} \times \langle N_{\text{gluons}} \rangle), \tag{6.13}$$

and is expected to have quite a pronounced effect on event kinematics.

- `INSTANTON_SIGMAHAT_MODIFIER`. SHERPA multiplies the tabulated cross-section by this `MODIFIER` — this moves the cross-section up or down.

- `INSTANTON_ALPHAS_FACTOR`. This modifier multiplies the value of α_S, and is relevant for the cross-section which has a term that looks like

$$(\texttt{INSTANTON_ALPHAS_FACTOR} \times \alpha_S(1/\rho))^{2\beta_0(1/\rho)}\,. \tag{6.14}$$

- `INSTANTON_SCALE_CHOICE`. This is the scale that is being used as the factorisation scale and the parton-shower starting scale. The nominal choice is $1/\rho$, but the $\sqrt{s'}/N_{\text{gluons}}$ and $\sqrt{s'}$ options are also provided. The alternatives will change the amount of ISR and FSR in a drastic way.

Of all these possible sources of uncertainty, we do not consider `INSTANTON_SIGMAHAT_MODIFIER`, since this is a flat variation of precisely the cross-section that this analysis is attempting to constrain.

For understanding the effects of `INSTANTON_NGLUONS_MODIFIER`, `INSTANTON_ALPHAS_FACTOR`, and `INSTANTON_SCALE_CHOICE`, variations are performed and plotted for comparison in Figures 6.1, 6.2 and 6.3, respectively. The variations of $\langle N_{\text{gluons}} \rangle$ by `INSTANTON_NGLUONS_MODIFIER` values of 0.7 – 1.3 produces only slight differences in general. The main differences occur at high particle-p_{T}, where downward-varied events are harder, vice versa. Similar variations of α_{S} by values of 0.7 – 1.3 seem

to affect cross-sections but not shapes. In this sense, the α_S variation is degenerate with the `INSTANTON_NGLUONS_MODIFIER` variation.

Substantial differences are seen between the different scale choices. The factorisation scale ought to have something to do with the momenta of the particles attached to the hard-scatter, so the $\sqrt{s'}$ choice is clearly inappropriate. However, the $1/\rho$ and $\sqrt{s'}/N_{\text{gluons}}$ scales are both reasonable, the former being proportional to the momentum of an incoming gluon by Mueller's argument [54], while the latter is approximately proportional to the momentum of an output gluon. The main effect is again on the cross-section: the ratio in instanton rapidity is flat within the $|\eta| < 2.5$ acceptance of the analysis, and shape differences are largely washed out from the stable particle spectra. The main shape difference appears at high instanton masses (corresponding to high-S_T and high particle multiplicities). This is because N_{gluons} rises slightly slower with $\sqrt{s'}$ than $1/\rho$ does, as can be seen in Table 3.1.

For placing observed limits, it is very important to account for the full systematic uncertainty on the predicted signal cross-section in order to proportionately weaken the upper limits and prevent false exclusion. However, none of these theoretical uncertainties that affect the signal modelling have been implemented yet. In this thesis, pending a more rigorous treatment, a flat uncertainty of 10 % is assigned to the signal yields to cover the shape differences between all of the variations described. In this way, the expected exclusion sensitivities quoted pertain to the sensitivity to 'a process of cross-section σ (up to 10 % shape effects)', which is a stronger statement about the analysis design than if σ were indeed smeared by factors of 100 [56].

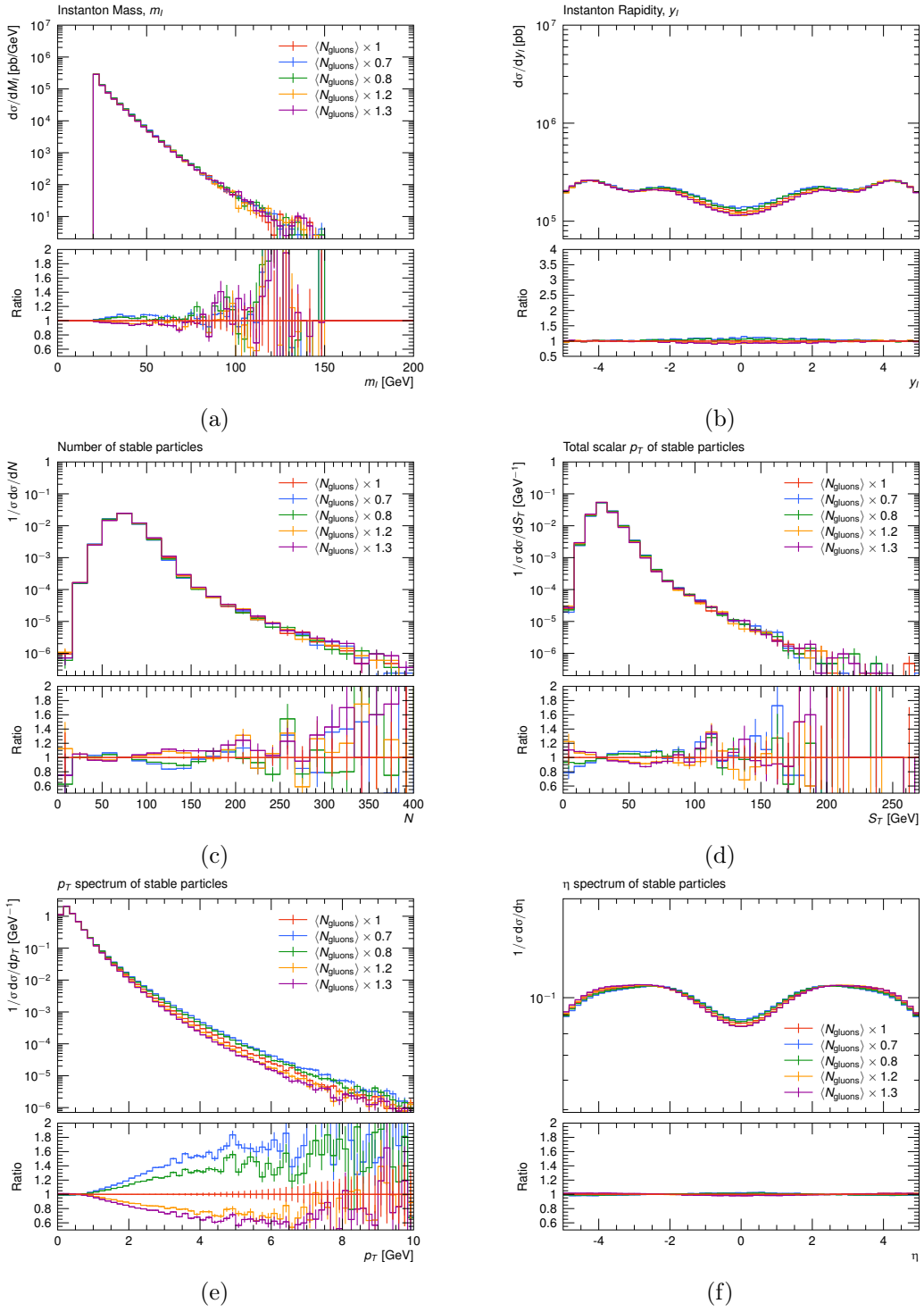


Figure 6.1: Study of the effect on truth-level instanton event distributions of varying the `INSTANTON_NGLUONS_MODIFIER` parameter, which multiplies the mean number of gluons produced at a given instanton mass. Values of {0.7, 0.8, 1.0, 1.2, 1.3} are shown as indicated in the legend, with 1 corresponding to the nominal value.

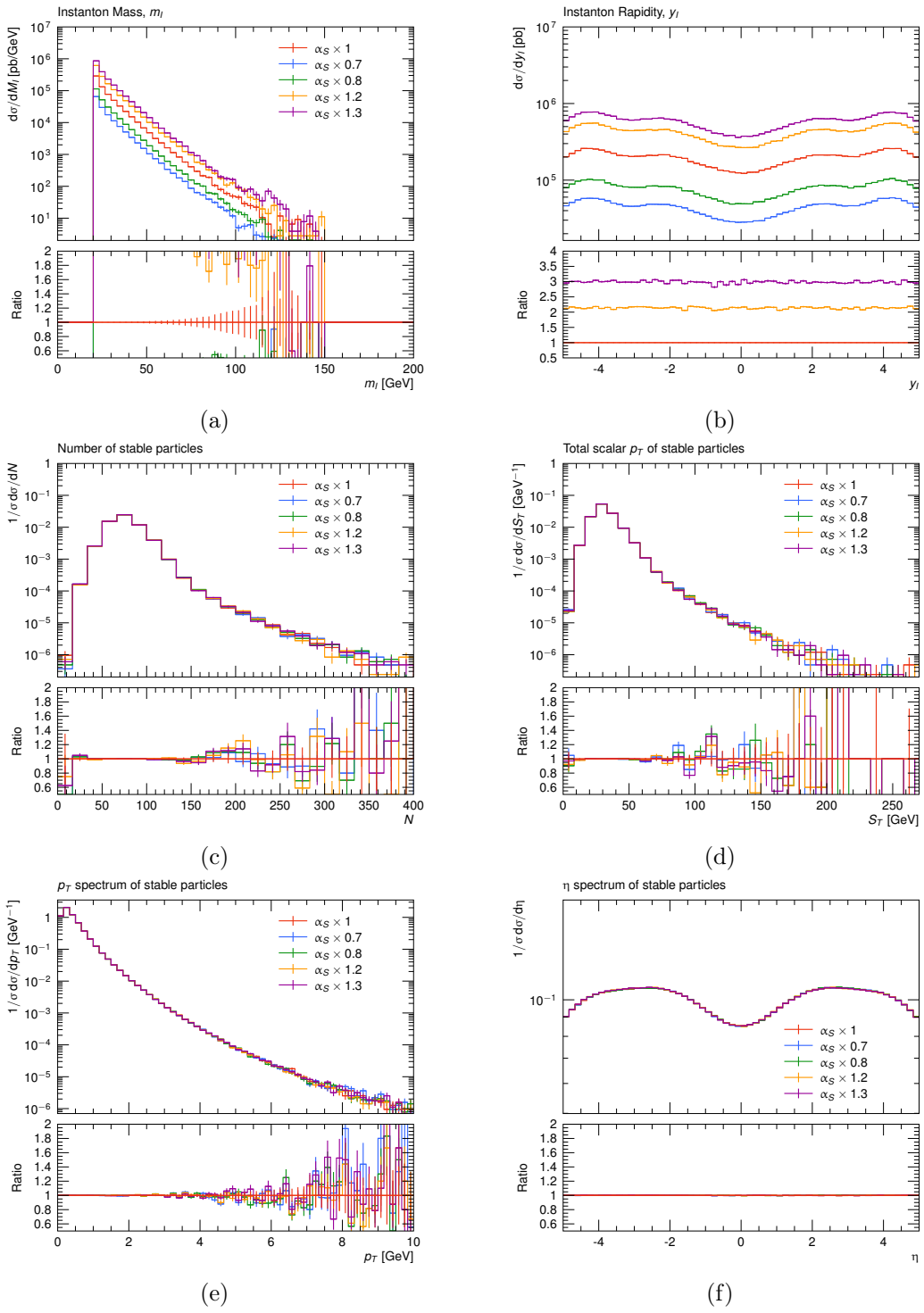


Figure 6.2: Study of the effect on truth-level instanton event distributions of varying the `INSTANTON_ALPHAS_FACTOR` parameter, which multiplies the strong coupling, α_S. Values of $\{0.7, 0.8, 1.0, 1.2, 1.3\}$ are shown as indicated in the legend, with 1 corresponding to the nominal value.

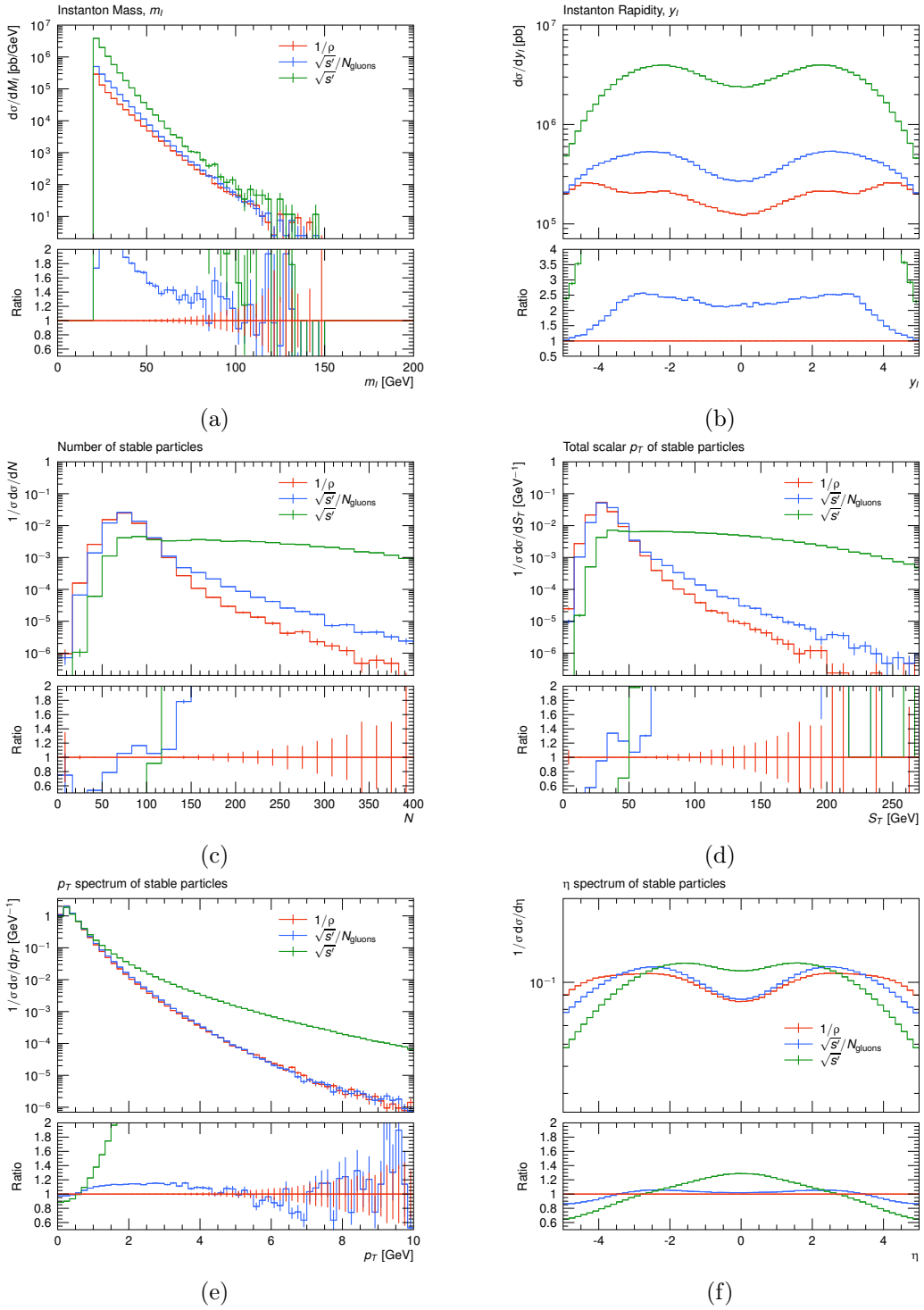


Figure 6.3: Comparison of truth-level instanton event distributions under different μ_f scale choices, as selected using the `INSTANTON_SCALE_CHOICE` parameter of the run card. The nominal choice is one over the instanton size, $1/\rho$. The alternative choices are the partonic centre-of-mass energy divided by the mean number of gluons produced in the instanton decay (labelled $\sqrt{s'}/N$), and the partonic centre-of-mass energy (labelled $\sqrt{s'}$).

6.4.3 Background estimation

Here we discuss some properties of the background estimation with a view to producing a systematic uncertainty on it.

Usually in the deployment of the ABCD background estimation technique in a plane spanned by variables X and Y, there is a third variable Z along which the distribution of background in XY is invariant. The third variable Z is chosen to have discriminating power, such that by inverting a cut on it, a replica ABCD plane is obtained that is depleted of signal and which can be used to validate the method. However, in our ABCD setup, no such variable Z is constructed and the validation of our method depends on cross-checks in XY rather than direct measurement.

These cross-checks consist of performing variations of the background estimation. Five different methods have been considered, two of which are MC-driven, while the others are data-driven. All of these methods should give us an idea of the two underlying systematics: (i) that $\mathrm{dCorr}^2(X, Y) \neq 0$, and (ii) that the background MC does not perfectly describe the data. Both of these imperfections tend to violate the ABCD assumption, thus harming the background estimation.

1. Non-closure calculated from MC (MC-driven)
2. Two-point difference between Epos and Pythia 8 (MC-driven)
3. Variation in ABCD from using subset CRs (Data-driven)
4. Variation in ABCD from eliminating bands around the SR (Data-driven)
5. Closure tests in the data using validation regions (VRs) (Data-driven)

Conceptually, method (1) bears on systematic source (i), method (2) on (ii), and methods (3), (4), and (5) on both (i) and (ii). As context for the numbers presented in this section, a table of the event yields in the ABCD regions is shown in Table 6.1.

Category	Sample	N_{BCD}	N_A	N_B	N_C	N_D
Low	Epos	1 204 258.1 ± 1008.0	1 206 186.2	3 018 921.7	3 787 260.6	9 494 180.4
	Pythia 8	952 710.5 ± 910.9	967 315.2	2 742 033.9	2 452 061.3	7 057 374.9
	Data	933 114.2 ± 964.6	—	1 759 537.0	3 059 306.0	5 768 814.0
	Instantons	389.6 ± 71.6	400.8	138.4	143.6	51.0
Med	Epos	545 848.2 ± 495.4	556 219.5	2 695 774.4	2 656 461.0	13 119 435.3
	Pythia 8	448 505.4 ± 460.4	502 188.9	2 234 515.5	1 980 116.6	9 865 211.6
	Data	395 689.0 ± 441.6	—	1 418 144.0	2 366 395.0	8 481 128.0
	Instantons	27.1 ± 23.5	77.2	9.1	6.2	2.1
High	Epos	167 982.5 ± 201.4	169 349.9	1 588 087.6	1 370 136.0	12 953 109.5
	Pythia 8	230 782.4 ± 278.7	190 548.8	1 398 688.1	1 567 110.9	9 497 690.1
	Data	169 554.5 ± 244.0	—	1 086 672.0	1 004 885.0	6 440 292.0
	Instantons	1.7 ± 19.4	2.1	0.1	0.1	0.0

Table 6.1: Event yields of the signal, background, and data samples in the ABCD regions, and the estimated background, N_{BCD}. Yields are shown for each mass category. The data in region A are blinded. The N_{BCD} column does not apply to the signal, but the 'background estimation' in the signal sample is shown nevertheless. Despite the discrepancy in the cross-section between Epos, Pythia 8, and data, the yields of these processes are not corrected for this table. The uncertainty on N_{BCD} is obtained by propagating the statistical uncertainties on the region yields through the ABCD equation (5.15).

Category	Sample	k [%]	N_{BCD}	Simulated N_A
Low	Epos	-0.2 ± 0.1	$1\,204\,258.1 \pm 1008.0$	1 206 186.2
	Pythia 8	-1.5 ± 0.1	$952\,710.5 \pm 910.9$	967 315.2
Med	Epos	-1.9 ± 0.2	$545\,848.2 \pm 495.4$	556 219.5
	Pythia 8	-10.7 ± 0.2	$448\,505.4 \pm 460.4$	502 188.9
High	Epos	-0.8 ± 0.3	$167\,982.5 \pm 201.4$	169 349.9
	Pythia 8	21.1 ± 0.3	$230\,782.4 \pm 278.7$	190 548.8

Table 6.2: Non-closure of the background estimation calculated from MC in the low, medium, and high mass categories for each background sample. The classifiers in this case were trained with the EPOS background model, which is the nominal choice. The uncertainties shown are statistical, and are obtained from propagating the statistical uncertainties on the CR yields. The non-closure is self-normalising, so no correction for the discrepancy observed between the cross-sections of Epos and Pythia 8 is applied.

For the first method (1), the non-closure of the background estimation is calculated in the Epos and Pythia 8 samples for comparison. The results are shown for each mass category in Table 6.2. The non-closures evaluated in the Epos sample are at the level of a percent, and are much smaller than the non-closures attained in the Pythia 8 sample. This is expected, since the nominal ABCDisCo networks are trained directly on the Epos sample, and the optimisation of the decorrelation parameter explicitly considers closure (through Eq. (5.10)). The non-closure is worst (at 21.1 %) when Pythia 8 events are put through the Epos network of the high-mass category.

Taking the two-point difference between the Epos and Pythia 8 predictions for method (2) produces the numbers shown in Table 6.3. The differences ΔN_i in the four regions reflect mis-modellings of the ABCD plane. Assigning the ΔN_i as uncertainties on the data yields in the CRs and propagating the effects results in the 'two-point' systematics shown in the δN_{BCD} column. The uncertainty is largest in the high-mass category. This may be because flavour is relevant to classifier performance here, and Epos and Pythia 8 are different in this respect.

The variational methods (3) and (4) are based on the constructions shown in Figures 6.4a and 6.4b, respectively, and serve to test the robustness of the method against variations of the CR definitions. The results of applying these variations in

Training Model	Region	N_{BCD}	δN_{BCD}	ΔN_B	ΔN_C	ΔN_D	ΔN_A
Epos	Low	933 114.2 ± 964.6	240 668.4 (25.8 %)	−408 756.0	340 899.9	67 856.2	−52 488.6
	Med	395 689.0 ± 441.6	44 499.4 (11.2 %)	−156 536.0	38 978.7	117 557.4	−68 138.1
	High	169 554.5 ± 244.0	60 529.1 (35.7 %)	−105 916.5	−338 078.2	443 994.8	−39 632.2

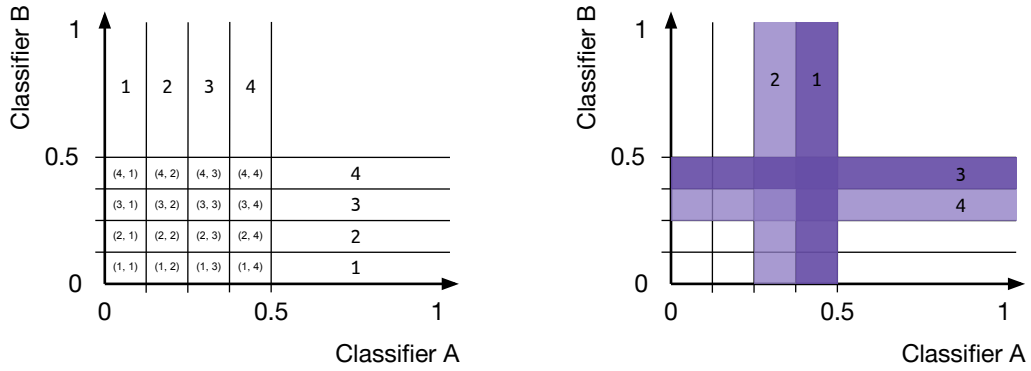
Table 6.3: Two-point difference between the Epos and Pythia 8 predictions in the ABCD regions. The differences in region B, C, and D are propagated as uncertainties through the ABCD equation (Eq. 5.15) to give the values in the δN_{BCD} column. The δN_{BCD} values are shown in the parentheses as a percentage of the nominal estimated background, N_{BCD} (which is calculated from data). The ΔN_i differences are taken in the sense Δ = Epos − Pythia.

the EPOS sample and in data are shown in Figures 6.6 and 6.7, respectively.

The CR subsets method (3) partitions the ABCD plane as shown in Figure 6.4a and produces as many estimations of the background in SRA as there are subregions of CRD (the bottom-left quadrant of the plane). In the tuples labelling the subregions in CRD, the first number refers to the y-bin, while the second number refers to the x-bin. In the summary plots of Figures 6.6 and 6.7, the varied background estimations are labelled 'CR (y, x)', and the subregions along the diagonal of CRD are highlighted for orientation. The band-elimination method (4) partitions the ABCD plane as shown in Figure 6.4b. The varied background estimation in this case involves eliminating each band i in turn (one at a time) so there are four variations. In the summary plots, these variations are labelled 'Band i'.

The pattern of the variations is remarkably similar between data and EPOS, as can be seen by flicking between Figures 6.6 and 6.7. This testifies to rather good modelling of the data. The CR subsets method involves extreme variations of the background estimation, so this method is not considered as a way to produce a systematic uncertainty. The structure seen in the variability of the CR subset estimations can be connected to structures seen in the ABCD planes of Figure 5.25. The band-elimination method involves gentler variations of the background estimation, and we should expect the variations in the estimated background to have less spread. This is indeed found to be the case, and the nominal background estimation is captured within the spread of these variations. The spread of the CR subset variations easily captures both all of the band elimination variations and the nominal estimation.

For the closure tests of method (5), the construction shown schematically in Figure 6.5 is used, which involves defining VRs that are kinematically similar to the SR. These VRs are labelled 1 to 6 with increasing distance away from the SR as indicated. For producing variations of each VR closure test, the same configuration of CR subsets as shown in Figure 6.4a is used, such that multiple background estimations can be performed in each VR. The results of these closure tests are summarised for each



(a) (b)

Figure 6.4: Schematics showing the partitioning of the ABCD plane for (a) the band elimination-based variations, and (b) the CR subsets-based variations of the background estimation. These constructions are used to cross-check the stability of the background estimation.

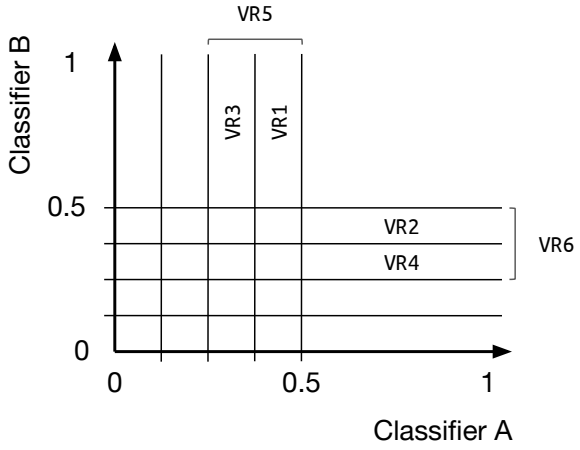
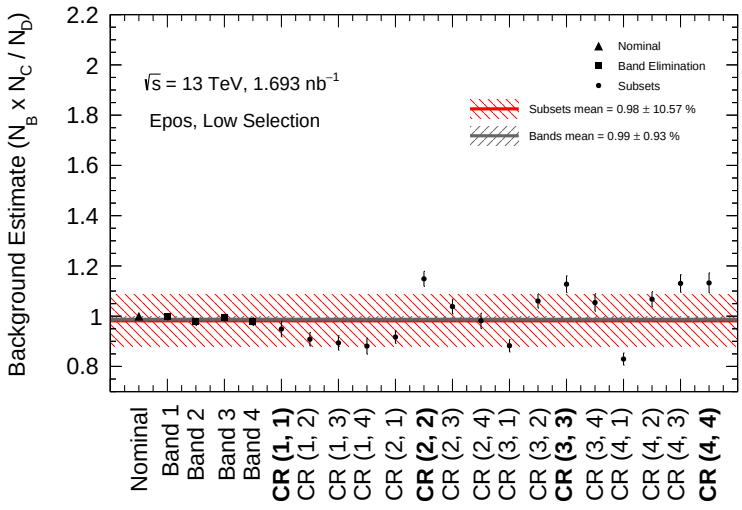
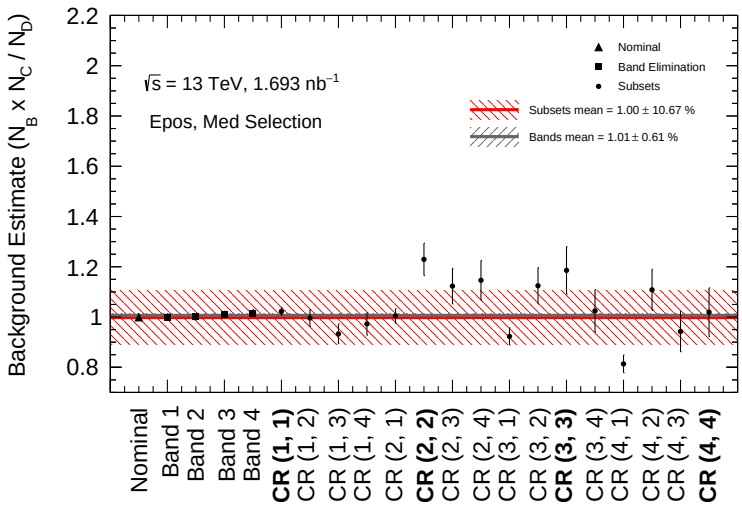


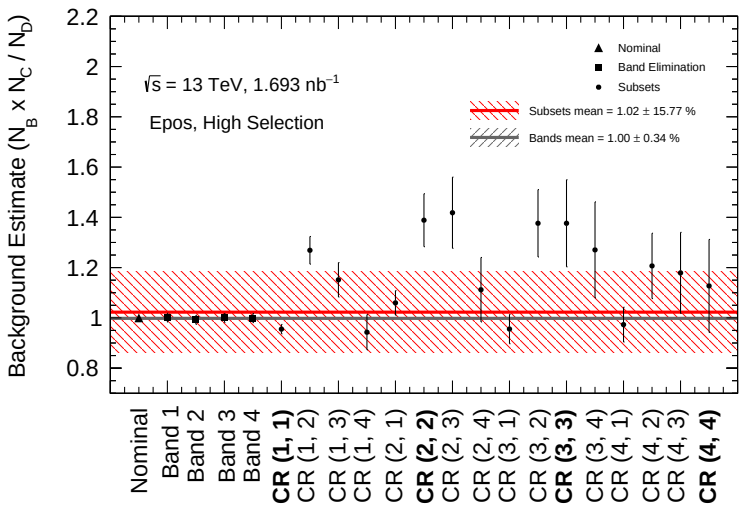
Figure 6.5: Schematic showing the definitions of the validation regions (VRs) surrounding the SR that are used for the closure tests in data.



(a) Epos background, low-mass category.

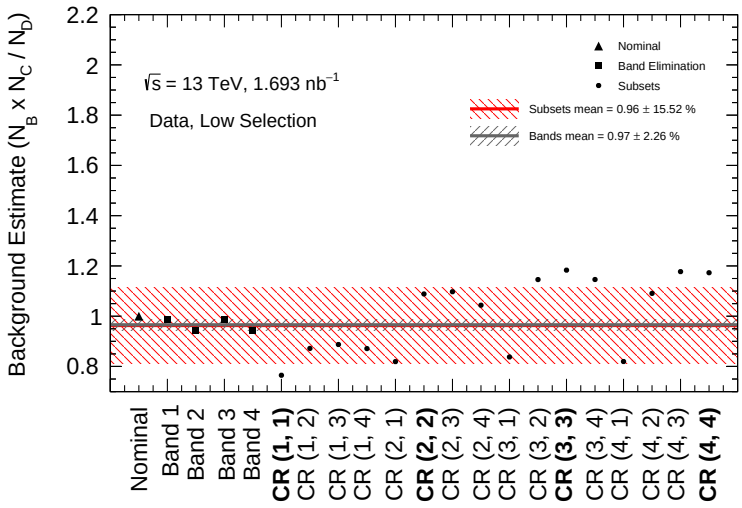


(b) Epos background, medium-mass category.

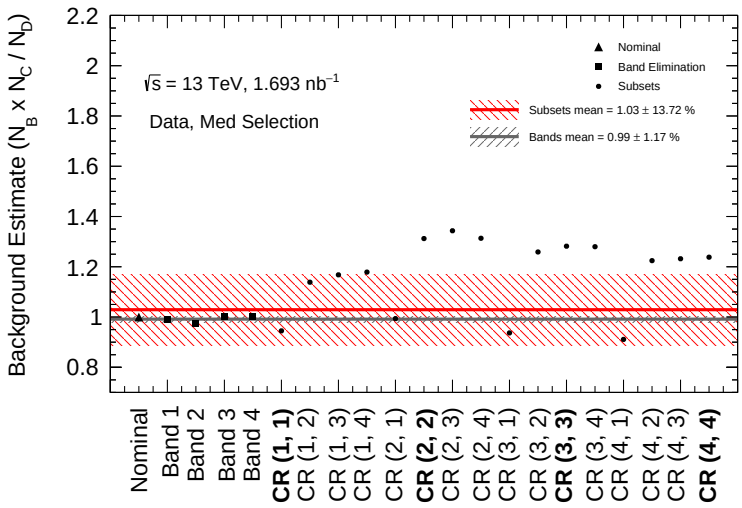


(c) Epos background, high-mass category.

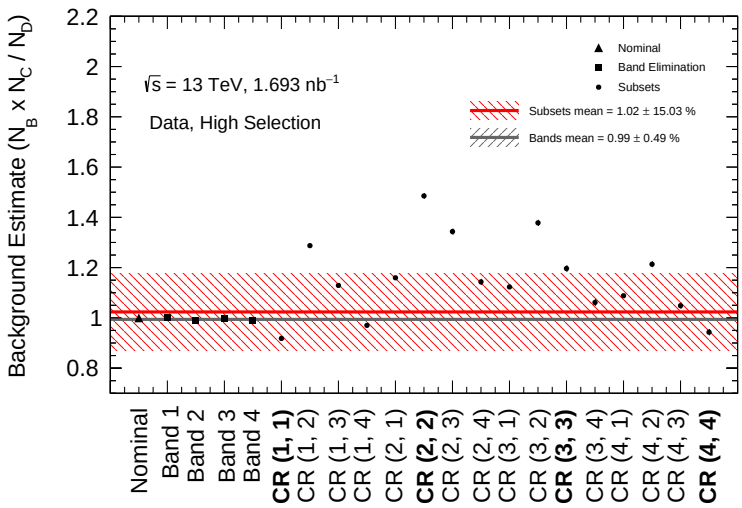
Figure 6.6: Summary of the variations of the ABCD background estimation performed on the EPOS background model. The results of the 'CR subsets' and 'band elimination' methods introduced in the text are shown. Each of these methods entails a different type of variation of the background estimation as represented in Figure 6.4. The band-elimination variations are shown as square markers, and their mean and standard deviation are shown in grey. The CR subsets variations are shown as filled circles, and their mean and standard deviation in red. The variations are all normalised to the nominal background estimation shown in the left-most bin for reference.



(a) Data, low-mass category.



(b) Data, medium-mass category.



(c) Data, high-mass category.

Figure 6.7: Summary of the variations of the ABCD background estimation performed on the data. The results of the 'CR subsets' and 'band elimination' methods introduced in the text are shown. Each of these methods entails a different type of variation of the background estimation as represented in Figure 6.4. The band-elimination variations are shown as square markers, and their mean and standard deviation are shown in grey. The CR subsets variations are shown as filled circles, and their mean and standard deviation in red. The variations are all normalised to the nominal background estimation shown in the left-most bin for reference.

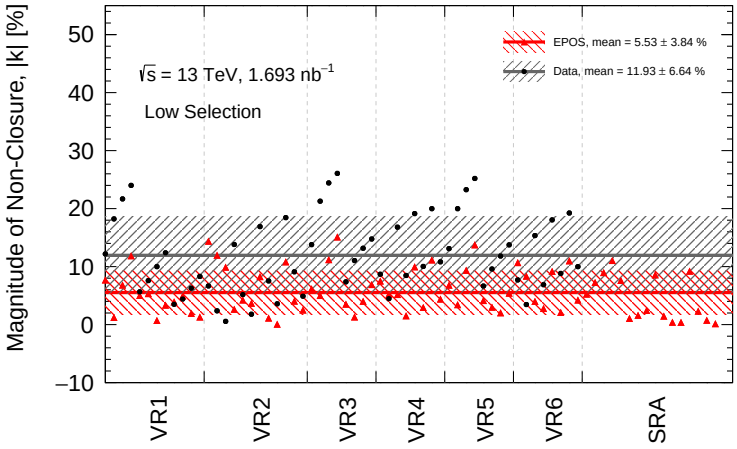
mass category in Figure 6.8.

In the plots of Figure 6.8, the vertical axis shows the magnitude of the non-closure, $|k|$, for each variation of the background estimation in each VR. The magnitude is taken in order to reflect in absolute terms how wrong we may be about the background estimation. The mean $|k|$ is always further from zero in the data than in Epos. This offset is due to the decorrelation accomplished in the Epos sample not transferring perfectly to the data due to mis-modelling. The spread of the non-closures in Epos is smaller, which indicates that the simulated ABCD plane is more uniform than the corresponding one made in data, leading to greater precision of the background estimation in the simulation. The Epos non-closure in the SR is generally very close to zero by construction.

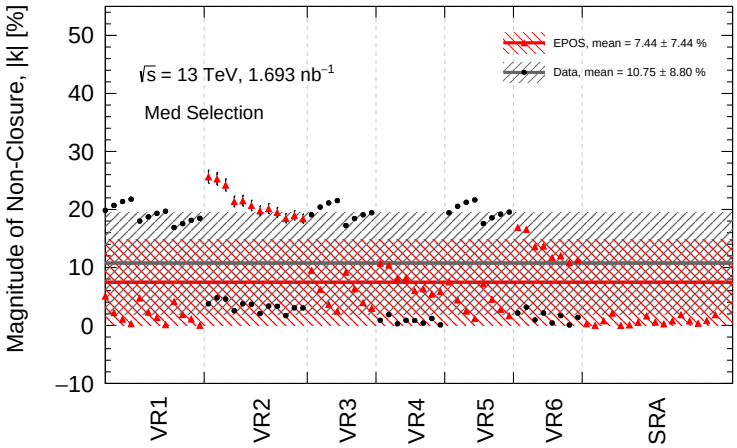
From these plots, the mean $|k|$ in the data is proposed as a sensible systematic uncertainty. The offset of this mean away from the same quantity in Epos contains relevant information about the background mis-modelling affecting the training of the DNNs, while its absolute placement encodes the performance of the decorrelation in the data. As such, the mean $|k|$ in the data constrains both systematic sources (i) and (ii). From this point of view, it is quite irrelevant that Epos and Pythia 8 disagree about the background estimation by the amounts shown in Table 6.3.

6.4.4 Prescription

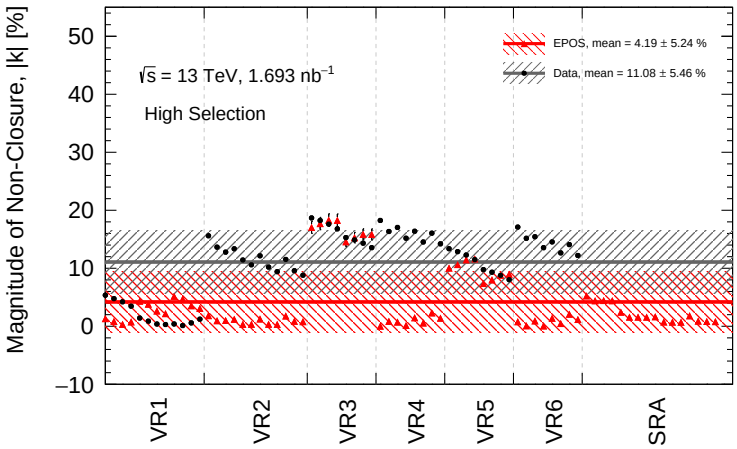
The precise formulation of the systematic uncertainty to use in this analysis is a work in progress in the ATLAS Collaboration. However, a number of possible approaches have been considered in this section. For the purposes of this thesis, the dominant systematic uncertainties are discerned to be due to (i) the background estimation, (ii) the choice of factorisation scale μ_{f} used in the signal modelling, and (iii) the signal cross-section. Uncertainty (i) is taken from the mean non-closure measured directly in data validation regions by the method summarised in Figure 6.8. Uncertainties (ii)



(a) Low-mass category.



(b) Medium-mass category.



(c) High-mass category.

Figure 6.8: Magnitude of the non-closure of the background estimation as measured in validation regions (VRs) in each of the low, medium, and high-mass categories. The definitions of the VRs are shown in Figure 6.5. The mean and standard deviation of the non-closures calculated in Epos and data across all VRs are shown as the red and grey bands, respectively. The means and standard deviations are weighted by the statistical uncertainties in the individual points, such that points derived from lower statistics (which tend to correspond to more extreme variations) contribute appropriately.

and (iii) are loosely constrained by a flat 10 % systematic pending a more rigorous treatment of the theory variations. These are added in quadrature to make a total systematic uncertainty where they affect.

6.5 Blinded Results

In this section, the blinded results are presented. The signal-strength scans used to place expected upper limits on the signal cross-section are shown in Figure 6.9. A summary of the expected upper limits is given in Figure 6.10.

Figures 6.9a, 6.9b and 6.9c show the scans performed in the low, medium, and high-mass categories, respectively, while Figure 6.9d shows the scan performed in the statistical combination of all the mass categories. Across the mass categories, there is a clear pattern in which the sensitivity is greatest in the low-mass category, and worst in the high-mass category. This pattern follows the signal cross-section, and shows that the cross-section's rapid fall through several orders of magnitude is not being counter-acted by the better signal discrimination power at high masses (illustrated in terms of AUC in Figure 5.16). In absolute terms, the expected upper limit at low mass is placed at about 443 times the nominal cross-section. The high mass category is much less sensitive, the expected upper limit there being 16 651 times the nominal cross-section. The medium-mass category is intermediate to these, at 998 times the nominal cross-section.

The analysis comes closest to the nominal cross-section through the statistical combination, which accomplishes $\mu^{\text{Exp}} = 412.6^{+151.9}_{-110.3}$. This statistical combination is essentially a 'shape-fit' across the three mass categories, and as such can exploit differences between the distributions of signal and background events over the three mass categories. That the upper limit is best here shows that the combination successfully enhances sensitivity to the signal in this way. In a combination of the low and medium-mass bins only, the expected upper limit evaluates to $412.7^{+151.9}_{-110.3}$,

which shows that the combined sensitivity comes entirely from the low and medium-mass categories. This, together with the low expected upper limit in the high-mass category, renders the high-mass category redundant.

It should be emphasised that the upper limits placed in the mass categories all apply to the same cross-section. This interpretation contrasts to that in most searches for new physics, where the headline results involve upper limits as functions of hypothetical particle mass [145–148]. In those searches, each different hypothetical particle mass corresponds to a *different signal model*, and an upper limit is placed for each of these different models over the range of targeted particle masses. The instanton, on the other hand, is non-resonant and is produced according to a spectrum. As a result, it is the same signal model that we see from the three mass categories.

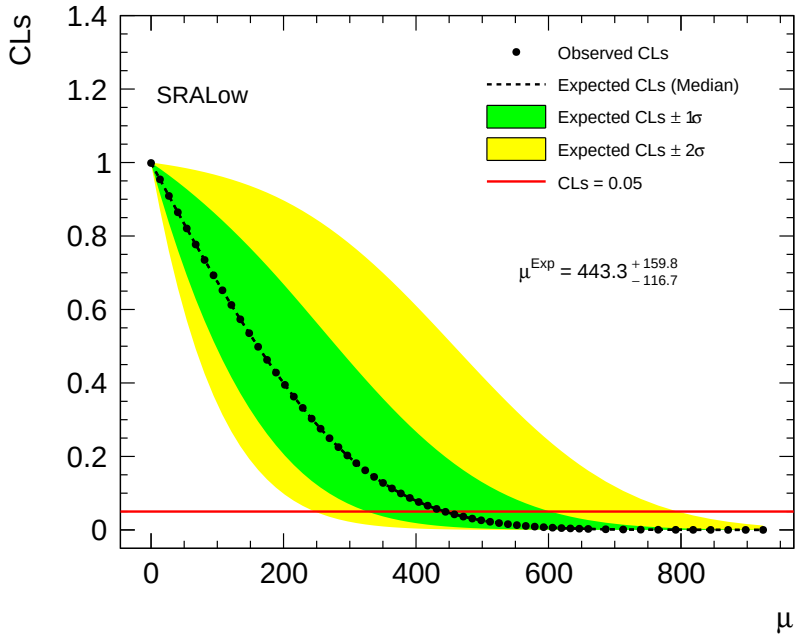
The numbers presented here convey the difficulty of discriminating instanton events from background. Because the expected upper limits show that the analysis is not sensitive to the nominal cross-section, expected discovery sensitivities are not shown, since these summarily convey discovery p-values of zero.

Compared to the upper limits on instanton-induced event rates placed by the H1 [41, 43] and ZEUS [42] Collaborations at HERA, the expected limits shown here are not competitive. The H1 Collaboration went as far as to challenge the theory predictions in Ref. [43]. The analysis contexts are very different between HERA and the LHC, so it is difficult to compare analysis designs directly. However, all of the HERA searches benefitted from shape-fits in the distribution of some multivariate discriminator. This, together with the success of the statistical combination in our analysis, indicate the value of shape differences in enhancing sensivity. The prospect of shape fits in classifier output variables may be considered as part of future work.

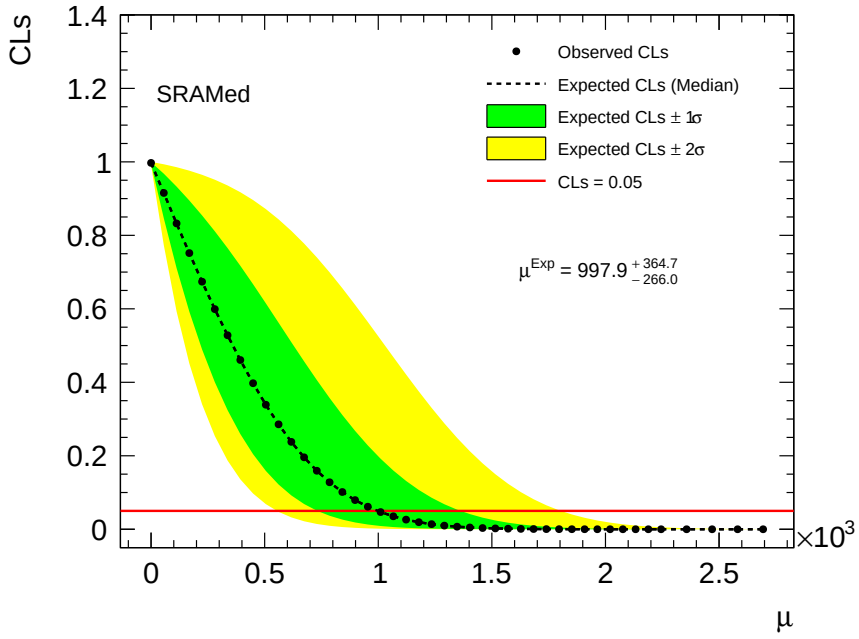
A fundamental problem with the analysis design is the poor background rejection that it accomplishes, which means that despite the good performance of the ABCDisCo networks, the signal regions are still dominated by background events. This follows

from the similarity of instanton-induced events to the minimum bias background, and the resulting difficulty in finding cuts that improve signal significance. An attempt to further reject background will also be made in future work. One possible line of investigation is to only use a subregion of the plane spanned by Classifiers A and B for the 'ABCD' plane in which the background estimation proceeds. By cutting out the most background-like events, and distance-decorrelating only in this sub-ABCD plane, stronger background rejection may be possible, and the decorrelation may even be easier.

The final results of this thesis are summarised in Figure 6.10, where the pattern of exclusion sensitivities over mass categories can be seen. Although the absolute values of these expected upper limits leave much to be desired, this first attempt at a search for QCD instantons has produced lessons for the analysis team and the community. As the analysis undergoes approval in the ATLAS Collaboration, a factor of 400 enhancement in exclusion sensitivity will be sought.

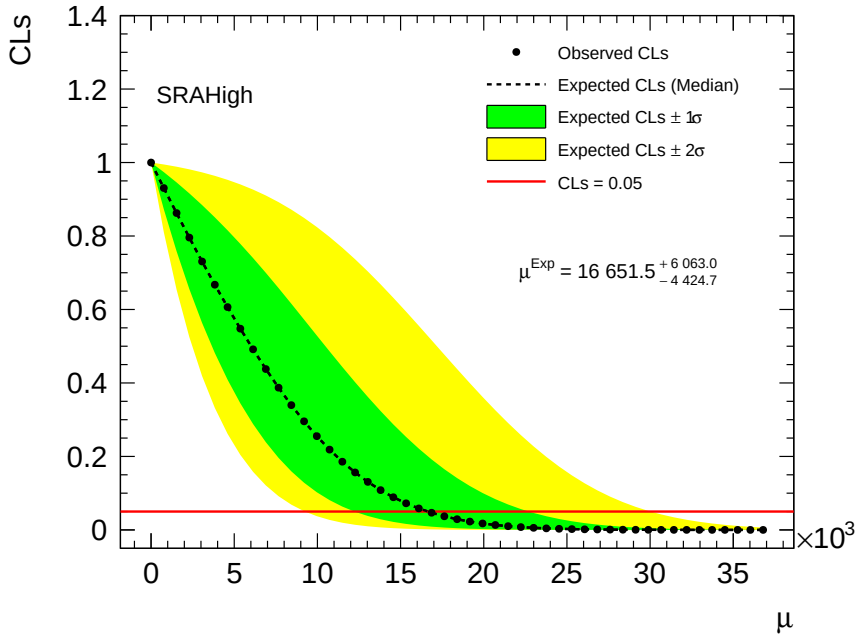


(a) Low-mass category.

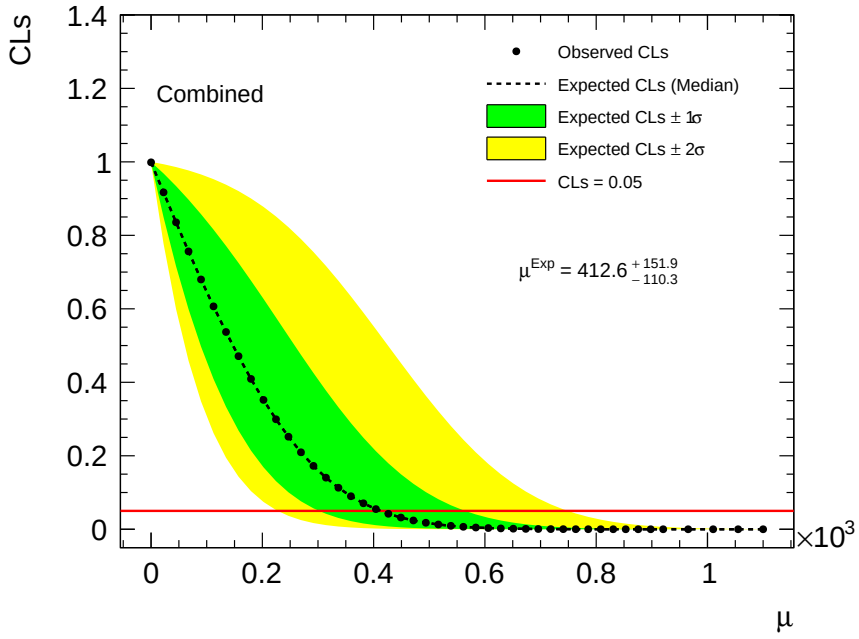


(b) Medium-mass category.

Figure 6.9: Representations of the scans of signal strength, μ, performed for placing exclusion upper limits on the signal model's cross-section. The expected upper limit is placed at the point where the expected CL_s drops below 0.05, which corresponds to 95 % CL_s. The green and yellow bands represent $\pm 1\sigma$ and $\pm 2\sigma$ excursions of CL_s within the total uncertainties, respectively.



(c) High-mass category.



(d) Statistical combination of the mass categories.

Figure 6.9: Representations of the scans of signal strength, μ, performed for placing exclusion upper limits on the signal model's cross-section. The expected upper limit is placed at the point where the expected CL_s drops below 0.05, which corresponds to 95 % CL_s. The green and yellow bands represent $\pm 1\sigma$ and $\pm 2\sigma$ excursions of CL_s within the total uncertainties, respectively.

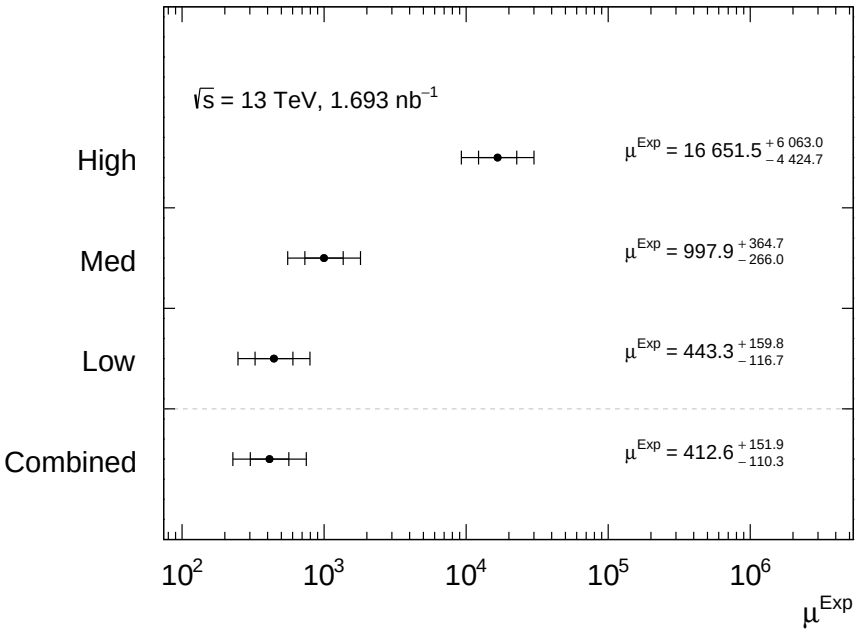


Figure 6.10: Summary of the expected upper limits on the signal strength, μ^{Exp}, calculated based on the signal model-dependent fit configuration using the signal strength-scans shown in Figure 6.9.

Chapter 7

Conclusion

Wisdom is oft-times nearer when we stoop
Than when we soar.
— Wordsworth

This thesis has presented the design of a search for QCD instantons at the LHC using the ATLAS Experiment. This new search was enabled by recent calculations of the instanton-induced event's cross-section at LHC energies [1, 51, 52], which sparked interest with the prospect of discovering a pseudoparticle of nature.

The fact that the predicted signal cross-section falls rapidly through several orders of magnitude with increasing instanton mass motivated performing the analysis in three mass categories, which are enriched in low, medium, and high-mass instanton events, respectively.

The twin cruxes of the analysis that ensues are namely the signal discrimination and background estimation. Due to an absence of significant one-dimensional discrimination power among the selected discriminating variables, a novel ML technique based on DNNs has been employed. These networks perform well, attaining good discrimination powers of between 0.80 and 0.98 in terms of Area Under the Curve,

and distance-correlation scores down to 0.017.

The estimation of the number of background events is produced using the ABCD method in the plane of distance-decorrelated classifier output scores. The background estimation is of low granularity, but that enables good control of the associated systematic uncertainties, which are typically estimated to be about 10 %. The background estimation is ultimately data-driven, due to our lack of confidence in the phenomenological models used to describe the minimum-bias event background.

A one-bin counting experiment is then performed in each mass category to produce expected exclusion sensitivities. The sensitivity is highest in the low-mass category, and rapidly diminishes towards the higher-mass categories as summarised in Figure 6.10. This trend shows that the dominant factor in the sensitivity of the analysis is the signal cross-section, since the DNNs are not capable of counter-acting its fall with correspondingly higher discrimination power. A statistical combination of the three mass categories (which amounts to a shape-fit across the mass spectrum in three bins) enhances sensitivity slightly, and brings the analysis to within about 413 times the predicted signal cross-section.

A number of challenges were encountered along the way, in the sparsity of suitable low-pileup data, the signal cross-section being reduced by an order of magnitude between Refs. [51] and [52], and due to a software bug in our original factorisation scale choice.

But the physics of QCD instantons speaks for itself. They are a deep and qualitatively different prediction of the SM, and a direct search could serve to test its consistency. This thesis has taken a step in the direction of one day discovering QCD instantons. It would be wonderful to eventually find out whether or not nature knows enough about topology to produce these special topological field configurations.

Appendix A

Tuning of SHERPA 3

A.1 Introduction

Underlying Event (UE) effects are not understood from first principles. Instead, they are described by phenomenological models that involve parameters whose values, not being determined *a priori*, must be fitted to data. The process of fitting these parameters is called tuning, and the resulting parameter sets are referred to as tunes. Although not exclusively a part of the UE, the fragmentation of final state partons into colour-neutral hadrons at the transition between perturbative and non-perturbative QCD is usually discussed along with the UE in the context of tuning, since its modelling is similarly phenomenological.

Before the SHERPA 3 event generator could be used for signal modelling in the QCD instantons search of this thesis, a comprehensive re-tuning of SHERPA's UE [15] was needed. This was necessary because technical adjustments that were made to the constituent UE models of SHERPA during the preparation of SHERPA version 3 had to be accommodated. The SHERPA authors had initially provided a best-guess set of tune parameters (based on intuition and rough validation) [108], but it was important for us to formally validate this tune and to potentially improve upon it

more systematically. Indeed, in our original signal samples (which suffered from a bug in the factorisation scale calculation), we found that instanton events were particularly sensitive to UE parameters. This was a false alarm, as the UE is not so prominent in the debugged signal samples used in this thesis. Nevertheless, a formal tuning was performed, which serves to validate the 'best-guess' tune that was used in this thesis.

This appendix gives a brief summary of the tuning effort to supplement the documentation of the signal sample configuration described in Section 5.1. An overview of the UE components and the models used by SHERPA to describe them is given in Section A.2, before the tuning strategy and results are presented in Sections A.3 and A.4, respectively.

A.2 Models

The parameters of the UE that stand to be tuned are summarised in Table A.1 and explained model by model in this section.

A.2.1 Multi-Parton Interactions

Due to hadrons being composite systems of partons, there is the possibility of multiple parton scatterings occurring in each hadron-hadron collision. This idea is simply confirmed by the observation that at low p_{T}, the cross-section for hard parton-parton scattering, $\hat{\sigma}_{2\to2}$, appears to exceed the total proton-proton scattering cross-section, σ_{tot}. This is consistent in the picture of multi-parton interactions (MPI), as then the ratio $\hat{\sigma}_{2\to2}/\sigma_{\mathrm{tot}}$ gives the average number of parton-parton scatters per event.

To simulate MPI, SHERPA implements the Sjöstrand–van Zijl model [149]. This model essentially dresses the primary interaction with secondary $2 \to 2$ scatters

in decreasing order of transverse momentum. In this downwards evolution in p_{T}, additional parton scatters are added using a Sudakov-like factor:

$$\mathcal{P}_{\mathrm{MPI}} = \exp\left(-\frac{1}{\xi\sigma_{\mathrm{ND}}}\int_{p^2_{\mathrm{T,i}}}^{p^2_{\mathrm{T},(i-1)}} \mathrm{d}p^2_{\mathrm{T}}\frac{\mathrm{d}\hat{\sigma}_{2\to2}}{\mathrm{d}p^2_{\mathrm{T}}}\right), \tag{A.1}$$

which encodes the probability of not adding a scatter between $p^2_{\mathrm{T},(i-1)}$ and $p^2_{\mathrm{T},i}$. Here, σ_{ND} is the non-diffractive proton-proton cross-section. The divergence in the differential $\hat{\sigma}_{2\to2}$ due to the t-channel contribution is regulated by the replacement

$$\frac{\mathrm{d}\hat{\sigma}_{2\to2}}{\mathrm{d}p^2_{\mathrm{T}}} \propto \frac{\alpha_S^2(p^2_{\mathrm{T}})}{p^4_{\mathrm{T}}} \to \frac{\alpha_S^2(p^2_{\mathrm{T}}+p^2_{\mathrm{T},0})}{(p^2_{\mathrm{T}}+p^2_{\mathrm{T},0})^2}. \tag{A.2}$$

The dimensionless parameter ξ scales the non-diffractive cross-section to suppress or enhance the probability of a secondary scatter as required. The downwards evolution terminates at a phenomenological value $p_{\mathrm{T,min}}$.

The sharp $p_{\mathrm{T,min}}$ cut-off is included to remove the naive growth in the number of interactions as $p_{\mathrm{T,min}} \to 0$, which occurs in this model due to the increase in the probability of a secondary scatter as $\hat{\sigma}_{2\to2}(p_{\mathrm{T}} \to 0)$ diverges. This would suggest that more and more energy is scattered as $p_{\mathrm{T}} \to 0$, which is clearly not physical.

In principle there must be some mechanism for taming this growth [149]. This should involve the correlation of parton structure functions (PDFs) within the incoming hadrons, but a stronger effect arises from the fact that in the exchange of soft partons, the transverse wavelengths of the probes are too large to resolve the colour-neutral hadrons, and the effective coupling between probe and target is reduced. The latter decoupling mechanism is not understood in perturbative QCD, so instead the $p_{\mathrm{T,min}}$ cut-off, below which no further parton scatters are added, is used to remove it.

The SHERPA implementation further introduces impact parameter dependence in the form of a matter-density profile (again in the same spirit as Sjöstrand and van Zijl). By default, this density profile is taken to be a double Gaussian interpreted as

describing a matter core of radius r_1 containing a fraction $1 - f_{\text{mat}}$ of the hadronic matter, surrounded by a larger sphere of radius r_2 carrying the remainder of the matter:

$$\rho(r) \propto (1 - f_{\text{mat}}) \frac{1}{r_1^3} \exp\left(-\frac{r^2}{r_1^2}\right) + f_{\text{mat}} \frac{1}{r_2^3} \exp\left(-\frac{r^2}{r_2^2}\right). \tag{A.3}$$

The parameters $p_{\text{T,min}}$ and $p_{\text{T,0}}$ apply to a specific collision energy. They are usually fixed along with the other UE parameters at a reference energy $E_{\text{ref}} = 7\,\text{TeV}$, and are evolved to other collision energies E by

$$p_{\text{T,i}}(E) = p_{\text{T,i}}(E_{\text{ref}}) \left(\frac{E}{E_{\text{ref}}}\right)^{\eta^E}, \tag{A.4}$$

where η^E is a tunable parameter.

The variables $p_{\text{T,min}}$, $p_{\text{T,0}}$, ξ, and f_{mat}, r_1, and r_2 are free parameters that must be tuned to reference data. The module in SHERPA 3 that implements MPI is called `AMISIC`.

A.2.2 Intrinsic Transverse Momentum

The partons inside colliding hadrons have transverse momentum, as can be seen in final state transverse momentum distributions. In the case of the Drell-Yan production of two leptons, the transverse momentum of the dilepton system would have a small peak at zero if the partons collided with precisely zero initial transverse momentum. In SHERPA, the partons inside hadrons are therefore endowed with intrinsic transverse momentum independently of Bjorken-x and flavour according to a limited Gaussian distribution of mean $\langle k_{\text{T}} \rangle$ and width σ:

$$\Pr(k_{\text{T}}) \propto \exp\left(-\frac{(k_{\text{T}} - \langle k_{\text{T}} \rangle)^2}{\sigma^2}\right) \times \left[1 - \left(\frac{k_{\text{T}}}{k_{\text{T}}^{\text{max}}}\right)^{\eta^{\text{cut}}}\right], \tag{A.5}$$

for $k_\mathrm{T} < k_\mathrm{T}^\mathrm{max}$. The part in square brackets compresses the distribution polynomially in η.

This model is implemented in SHERPA in the `Primordial_KPerp` module, where $\langle k_\mathrm{T} \rangle$, σ, $k_\mathrm{T}^\mathrm{max}$ and η^cut must be tuned.

A.2.3 Hadronisation

To describe the transition of the quanta of perturbative QCD to observable hadrons, SHERPA employs the cluster fragmentation model introduced in Ref. [150] and adapted in Ref. [151]. Qualitatively, this cluster fragmentation involves clustering quarks and diquarks, then decaying these clusters until they can be identified by their masses and flavours with hadrons which may then decay further. SHERPA's implementation is nicely discussed in Ref. [152] (in the context of a past tuning effort).

This model involves about 50 parameters that must be tuned to reference data. Most of them are in principle universal – applying to every energy and collision system – so they are typically tuned to LEP e^+e^- collision data and exported to the LHC setting. However, three of these parameters pertain to beam fragmentation specifically, and thus depend on beam type. As such, these parameters must be reconciled directly with LHC data.

Beam fragmentation parameters enter through the assignment of cluster decay kinematics. If a cluster $\mathcal{C}$ made of two flavours f_1 and $\bar{f}_2$ is above threshold for decays into hadrons, then it can fission into secondary clusters by "popping" a flavour–anti-flavour pair, $f + \bar{f}$, as in the reaction

$$\mathcal{C}[f_1 \bar{f}_2] \to \mathcal{C}[f_1 \bar{f}] + \mathcal{C}[f \bar{f}_2]. \tag{A.6}$$

Once the popped flavour f is selected (according to "popping" probabilities [152]),

the kinematics of the new clusters must be fixed. The kinematics are determined in a dipole fashion in the rest-frame of the parent cluster, with the momenta p_1 and p_2 of the original cluster constituents f_1 and $\bar{f}_2$ arranged along the positive and negative z-axis. The decay proceeds by sampling a transverse momentum k_{T} from a Gaussian distribution of tunable width $k_{\mathrm{T},0}$. Longitudinal momentum fractions z_1 and z_2 for f_1 and $\bar{f}_2$ are sampled from the probability distribution

$$\mathcal{P}(z) = z^{\alpha}(1-z)^{\beta} \exp\left[-\frac{\gamma}{z} \cdot \frac{k_{\mathrm{T}}^2 + (m_{f_1} + m_{\bar{f}_2})^2}{k_{\mathrm{T},0}^2}\right], \tag{A.7}$$

where α, β, and γ are tunable parameters.

The parameters α, β, and γ depend on whether the flavour being specified is a light quark, heavy quark, diquark, or whether the cluster includes a beam remnant (originating in the break-up of an incoming hadron).

The cluster fragmentation model is implemented in the `AHADIC` module.

A.3 Tuning Strategy

In Section A.2, the SHERPA models which together describe the UE were introduced. The precise parameters of interest for our tuning are summarised in Table A.1. Our tuning strategy depends on the PROFESSOR methodology, which has been used successfully in previous ATLAS tuning campaigns, for example in Refs. [113, 153, 154]. The PROFESSOR methodology is introduced in Section A.3.1, before the UE observables used and their measurements are introduced in Sections A.3.2 and A.3.3, respectively.

Model Parameter	Run-Card Parameter	Description
$p_{\mathrm{T},0}$	`AMISIC::PT_0(ref)`	Transverse momentum regulator
$p_{\mathrm{T,min}}$	`AMISIC::PT_Min(ref)`	Lower cut-off for the further addition of parton scatters
σ_{ND}	`AMISIC::SIGMA_ND_NORM`	Non-diffractive proton-proton cross-section
$1 - f_{\mathrm{mat}}$	`AMISIC::MATTER_FRACTION1`	Fraction of proton matter in the double-Gaussian profile's core
r_1	`AMISIC::RADIUS1`	Radius of the proton core
r_2	`AMISIC::RADIUS2`	Radius of the proton's matter halo
η^E	`AMISIC::Eta`	Exponent for energy extrapolation
$\langle k_{\mathrm{T}} \rangle$	`INTRINSIC_KPERP::MEAN`	Mean of the intrinsic trasnverse-momentum (k_{T}) distribution
σ	`INTRINSIC_KPERP::SIGMA`	Width of the intrinsic-k_{T} distribution
$k_{\mathrm{T}}^{\mathrm{max}}$	`INTRINSIC_KPERP::MAX`	Cut-off of the intrinsic-k_{T} distribution
η^{cut}	`INTRINSIC_KPERP::CUT_EXPO`	Exponent compressing the k_{T} distribution
α_B	`AHADIC::ALPHA_B`	α for fission of beam clusters
β_B	`AHADIC::BETA_B`	β for fission of beam clusters
γ_B	`AHADIC::GAMMA_B`	γ for fission of beam clusters

Table A.1: List of the parameters of interest in SHERPA (corresponding to the parameters introduced in Section A.2), along with their names in the run card used to generate events, and a short description of each parameter's purpose.

A.3.1 The PROFESSOR toolkit

Parameters are fitted to reference data using the PROFESSOR toolkit [107]. PROFESSOR operates on a grid of MC samples, where each point in the grid corresponds to a different set of tune parameters. Using these spans of MC samples, PROFESSOR performs a polynomial interpolation of the MC event generator's response to changes of the tunable parameters in each histogram bin of each observable. With this parameterisation of bin responses, PROFESSOR evaluates a heuristic goodness-of-fit measure, χ^2, which it then minimises over the space of tune parameters with respect to the reference data.

PROFESSOR depends on the Rivet analysis preservation system [155]. The Rivet database stores unfolded data distributions of measured observables, as well as analysis codes that can be executed during MC event generation to produce corresponding simulated distributions. PROFESSOR takes these histograms, of data and of simulated events at the sampled MC points, as inputs.

The χ^2 is calculated over the bins b of the observables $\mathcal{O}$ according to

$$\chi^2(\mathbf{p}) = \sum_{\mathcal{O}} w_{\mathcal{O}} \sum_{b \in \mathcal{O}} \frac{(f^{(b)}(\mathbf{p}) - \mathcal{R}_b)^2}{\Delta_b^2}, \tag{A.8}$$

where $f^{(b)}(\mathbf{p})$ is the parameterisation of a bin's response to changes of a parameter vector $\mathbf{p}$, $\mathcal{R}_b$ is the reference data in bin b, and Δ_b is its associated uncertainty. The weight factors $w_{\mathcal{O}}$ are chosen subjectively to reflect the importance of each observable in the tuning. It is possible to apply weights to specific intervals of the observables, but this facility was not used. The point $\mathbf{p}$ in the space of tune parameters where this goodness-of-fit measure is minimised is the set of parameters that best describes the reference data (as far as the selected configuration can tell).

The observable–weight combinations that we used are listed in Table A.2. Weights were chosen such that the intervals of the selected observables that should physically

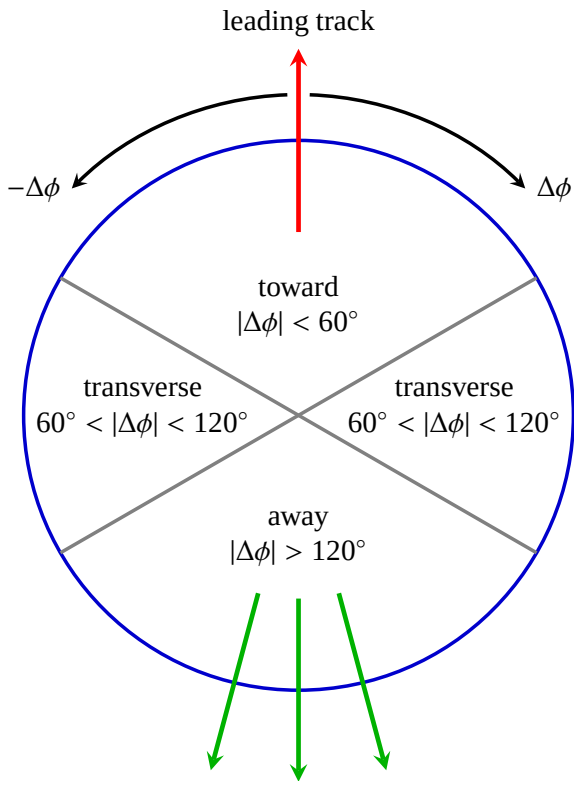


Figure A.1: Schematic showing how the transverse plane is partitioned with respect to a leading object in the event to define the 'transverse' region sensitive to underlying event (UE) activity. This is reproduced from Ref. [156], where the leading object is a track. In the other measurements considered here, the leading object can be a charged-particle jet [157], a calorimeter jet [158, 159], or simply a leading calorimeter deposit [160].

be describable by the models are highly weighted, vice versa. Weights were fine-tuned to a limited extent to minimise the χ^2.

A.3.2 Underlying Event variables

The UE variables used are all based on a partitioning of the transverse plane with respect to an object of leading p_{T}. This partitioning is shown schematically in the case where the leading object is taken to be a track in Figure A.1. The idea behind this partitioning is that the hard scatter component of the event emerges in a back-to-back configuration and produces the highest energy particles. With these highest energy particles, the hard scatter defines the 'toward' and 'away' regions. The region transverse to this hard scatter is relatively enriched in UE activity. Variables defined in the transverse region are thus particularly sensitive to the UE, and are

commonly used as UE tuning targets.

In the transverse region defined in Figure A.1, variables such as the number of charged-particle tracks, N_{ch}, the scalar sum of charged-particle p_{T}, $\sum p_{\text{T}}^{\text{ch}}$, as well as calorimeter sums of transverse energy, $\sum E_{\text{T}}$, may be measured. In the measurement of the UE in events containing a leading jet considered here [157], so-called 'trans-max' and 'trans-min' regions are additionally defined. The trans-max and trans-min regions are defined as the more and less active sides of the transverse region, respectively, as determined in terms of the variable under consideration. The motivation for this is that the trans-max region is more likely to be affected by hard wide-angle emissions and initial state radiation, while the trans-min region is more likely to be populated by MPI and beam remnants [157].

Many of these variables are measured as densities in the $\eta - \phi$ plane, and are profiled in a variable such as the leading object's p_{T} in order to discern any functional dependence thereupon.

A.3.3 Reference data and MC samples

The ATLAS Collaboration has performed many different measurements of the UE. This tuning campaign uses five measurements (separate publications) by ATLAS of observables sensitive to the UE at 7 TeV. The measured observables considered are listed in Table A.2, grouped by publication.

To match these data, 500 000 minimum-bias (MB) events are generated using Sherpa 3 at each sampled point in the space of parameters. There is a subtlety in the comparison of Sherpa MB simulation, which is assembled from MPI with impact-parameter dependence, to MB data. In a physical event, there is a hard scatter process that determines the scale of the first interaction's momentum transfer, but this is absent from the simulated events. This means that consequently the toward and away regions may look different between MB simulation and data.

Nevertheless, the UE component as seen in the transverse region is the same, and the strategy of tuning the UE of simulated MB events to data is valid.

Throughout what follows, p_{T} is synonymous with $p_{\perp}$ (which is often used in Rivet plots to denote 'transverse').

Observable	Weight
Track-based UE [156]	
Transverse N_{ch} density profiles (3 distributions)	3
Transverse $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ density profiles (3 distributions)	3
Std. dev. transverse N_{ch} density profile	3
Std. dev. transverse $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ density profile	3
Transverse $p_{\mathrm{T}}^{\mathrm{ch}}$ profile	3
Calo-based UE [160]	
Profile of transverse N_{ch}	5
Profile of transverse $\sum p_{\mathrm{T}}^{\mathrm{ch}}$	5
Pseudorapidity dependence of $\mathbf{E_T}$ [158]	
$\sum E_{\mathrm{T}}$ distributions in $\lvert\eta\rvert$ intervals (6 distributions)	1
Leading track-jet UE [157]	
Transverse N_{ch} profiles (5 distributions)	1
Mean transverse $p_{\mathrm{T}}^{\mathrm{ch}}$ (5 distributions)	1
Transverse $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ profiles (5 distributions)	1
Transverse N_{ch} profiles in p_{T} intervals (4×5 distributions)	0
Leading jet UE [159]	
Transverse, trans-max, trans-min $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ profiles in incl. jet events	5
Transverse, trans-max, trans-min $\sum p_{\mathrm{T}}^{\mathrm{ch}}$ profiles in dijet jet events	5
Transverse, trans-max, trans-min N_{ch} profiles in incl. jet events	5
Transverse, trans-max, trans-min N_{ch} profiles in dijet jet events	5

Table A.2: List of the observables fitted during the tuning process, along with their weight in the definition of the χ^2 measure used for tune optimisation. Reference data in these observables are taken from the measurements cited in the measurement headings.

A.4 Results

The scan ranges over the parameters are shown in Table A.3. The column showing prior values show the values chosen for the SHERPA author's best-guess tune, and indicate the values that stood to be improved by a systematic tuning. Scan ranges were constructed originally to bracket those prior values shown. After initial tuning, the scan ranges were tightened and parameter points sampled at a higher density therein to better resolve a minimum. The range of `AMISIC::PT_0(ref)` values does not include the prior value. This parameter was found to be correlated with `AMISIC::PT_Min(ref)`, which made it difficult to tune these two parameters together. It was decided to constrain `AMISIC::PT_0(ref)` to the interval 0.1 – 1, which in principle includes more natural values for a transverse momentum regulator.

It was found that the selected UE observables are not particularly sensitive to the beam fragmentation parameters α_B, β_B, and γ_B, so these could not be constrained particularly well. These parameters were instead extracted from deep inelastic ep scattering (DIS) events at HERA by a collaborator [109]. This is why these parameters are not shown being scanned in Table A.3 — they are fixed to the DIS values 15, 0.9, and 15, respectively.

The number of sampled points is an important factor in the accuracy of the interpolations since the density of points decreases exponentially with the number of parameters being scanned. The PROFESSOR documentation lays out a rule for selecting a minimum number of points to scan — there should be at least as many points as there are coefficients in the parameterisation [107]. For a second-order polynomial parameterisation, the number of coefficients to be determined is $1+P+P(P+1)/2 = 78$ for 11 parameters as in this case, and the suggestion is to over-sample by a factor of 2. We generate MC samples at 600 points.

Some dexterity is required in constructing scan ranges and in choosing the sampling density. If the space of points is under-sampled, then the parameterisation may

fail to describe the true MC generator response accurately. Similarly if the scan range in one of the parameters is too large, then structures in the MC generator's response to changes of this parameter, such as multiple local minima, may be grossly simplified in the fitting of the parabolic parameterisation. Physical insight into the expected magnitudes of the tune parameters can be used to avoid such problems, but ultimately the interpolation must be validated. In our case, this is done by comparing the interpolated distributions at the minimum in χ^2 with distributions generated directly at the same point in parameter space.

The final results are shown in Table A.3. The tuned distributions are shown in comparison to the interpolated tune, the best-guess tune, and the reference data in a selection of distributions in Figures A.2 – A.10. With the observable-weight combinations shown in Table A.2, the final tune parameters quoted in Table A.3 result in a $\chi^2/N_{\text{dof}} = 5321/732 = 7.3$, for $N_{\text{dof}} = 732$ degrees of freedom, which counts the weighted number of bins included in the χ^2 definition.

A.4.1 Tune performance and features

Discussion of tune parameters

Comparison of the tuned and prior parameter values in Table A.3 shows some drastic changes. The large relative changes suggest that we have found a different minimum of the tuning problem.

It appears that to compensate for the restriction of the `PT_0(ref)` parameter range, the `PT_Min(ref)` parameter is pulled upwards. In physical terms, `PT_0(ref)` being smaller means that the strong coupling, α_S, is larger, which results in stronger particle production at low p_{T} where $p_{\text{T},0}$ affects. This stronger particle production is compensated by the higher `PT_Min(ref)`, as this terminates the MPI evolution earlier. The normalisation of the non-diffractive cross-section, ξ, meanwhile stays

Parameter	Prior Value	Scan Range	Tuned Value	Relative Change
AMISIC::PT_0(ref)	2.5	0.1 – 1	0.716 ± 0.008	−71.4%
AMISIC::PT_Min(ref)	3	2.5 – 5	4.260 ± 0.009	+42.0%
AMISIC::SIGMA_ND_NORM	0.4	0.4 – 1	0.458 ± 0.003	+14.5%
AMISIC::MATTER_FRACTION1	0.5	0.1 – 0.4	0.187 ± 0.003	−62.6%
AMISIC::RADIUS1	0.4	0.0 – 0.5	0.157 ± 0.002	−60.8%
AMISIC::RADIUS2	1	0.8 – 1.5	0.890 ± 0.005	−11.0%
AMISIC::Eta	0.16	—	0.16	
INTRINSIC_KPERP::MEAN	0	0 – 1.2	1.004 ± 0.009	+∞%
INTRINSIC_KPERP::SIGMA	1.5	0.5 – 3	1.10 ± 0.03	−27%
INTRINSIC_KPERP::MAX	3	2 – 5	2.69 ± 0.03	−10%
INTRINSIC_KPERP::CUT_EXPO	5	3 – 6	5.12 ± 0.04	+2%
AHADIC::ALPHA_B	2.5	—	15	
AHADIC::BETA_B	0.25	—	0.9	
AHADIC::GAMMA_B	0.5	—	15	

Table A.3: List of the parameters in SHERPA that affect the description of the underlying event which needs to be tuned. Noted are the prior values (or 'best-guess' values provided by the SHERPA authors) and the ranges sampled during our tuning. Also shown are the tuned values for comparison. The uncertainties on the tuned values come from the covariance matrix of the fit parameters as reported by the minimiser, and are included to reflect how well the parameters are constrained by the reference data.

close to 0.4. This value of less than 1 decreases the exponent of the MPI probability factor shown in Eq. (A.1), and dampens the probability of not adding a further parton-parton interaction. Because all of our reference measurements are taken at 7 TeV, the energy-extrapolation parameter η^E cannot be constrained. Instead, the prior value of 0.16 is retained.

For the matter profile of the proton, we chose to use a double Gaussian, as this gives the minimisation the freedom to choose between the single and double Gaussian options by control of the `MATTER_FRACTION1` parameter. Most of the matter (about 81.3%) goes into the inner radius, which suggests that a single Gaussian profile may have sufficed.

A large qualitative change is seen in the intrinsic-k_{T} description, as the distribution of partons in the proton acquires a mean transverse momentum of about 1 GeV. The other intrinsic-k_{T} parameters remain similar. The parameter `INTRINSIC_KPERP::MEAN` should be anti-correlated with `PT_0(ref)` and `PT_Min(ref)` to boost particle production, so it is consistent that these three parameters change concurrently. However, the uncertainties on the fitting of some of the intrinsic-k_{T} parameters are relatively large, which suggests that these parameters are correspondingly less constrained by the selected reference data.

Large changes are obtained in the values of the beam cluster fragmentation parameters. Reading the probability distribution for longitudinal momentum, $z_{(1,2)}$ (Eq. (A.7)), the higher values of all parameters suggests that the decays of clusters involving beam remnants are much more asymmetric than was previously thought. In the probability distribution of Eq. (A.7), the α and β parameters determine the skew between small and large values of z. The γ parameter, when positive, shifts the whole distribution towards higher z. If the second fraction in the exponent of Eq. (A.7) (dependent on flavour masses and transverse momentum properties) is not similar to $1/\gamma$ or smaller, then our large value of γ tends to set the whole distribution to zero.

Discussion of distributions

The starting point for the tuning was the measurement of the UE with respect to a leading-p_{T} track [156]. Figures A.2 and A.3 show some of the distributions that were considered. Our tune out-performs the prior tune in every distribution of this measurement, being well within 10% of the data as opposed to 20–30% everywhere. It can be seen in the good agreement between the interpolated and generated tune that the minimum in the χ^2 along PROFESSOR's parameterisation corresponds closely to a real set of parameters. A feature of all the p_{T}-dependent distributions here is the discrepancy at low-p_{T} values showing that the charged-particle multiplicity is over-predicted in events of lower momentum transfers.

The $|\eta|$ distributions of Figures A.3a and A.3b are used to normalise the number of charged-particles per event (since these distributions contain no other shape information). The improved accuracy of our tune in describing these distributions explains how the offset of the prior tune to higher N_{ch} and $\sum p_{\mathrm{T}}$ and lower standard deviations thereof (particularly at low-p_{T}) is fixed.

The distribution of mean charged-particle transverse momentum as a function of leading track transverse momentum, $\langle p_{\mathrm{T}} \rangle$, shown in Figure A.3c, was found to be very easy to tune. The spread in this distribution from sampling different tune parameters is very small and almost entirely contained within the total uncertainties on the data, meaning that this distribution is somehow insensitive to our UE parameters of interest.

The modelling of the UE observables measured with respect to a leading calorimeter cluster [160] is also improved as shown in Figure A.4. This measurement has the advantage that taking a calorimeter deposit for the definition of the 'toward' region makes the measurement independent of the details of the charged-particle tracking as well as of jet reconstruction. Indeed, our tune describes these calo-based UE distributions very well, with the tune going within the data uncertainties everywhere

except the first, lowest-p_{T} bin.

In the pseudorapidity-dependent measurement of the UE [158] shown in Figure A.5, the new tune performs comparably to the prior tune, if not slightly better, at central rapidities and up to about $|\eta| = 2.4$, which is approximately where the ID tracker's acceptance ends. As the selection goes to higher pseudorapidity intervals, our tune becomes increasingly skewed towards lower-$\sum E_{\mathrm{T}}$ values. In this measurement, discrepancies begin to emerge between our interpolated and generated tunes. This may be a matter of statistics, as the selection efficiency decreases towards the higher rapidities and so the bins in these regions may not be faithfully represented in our MC samples. Here, the best-guess tune produces a more pronounced peak where there is a peak in the data at low-$\sum E_{\mathrm{T}}$.

Selected observables from the measurement of the UE with respect to a track-jet [157] are shown in Figures A.6, A.7, and A.8. Here, the charged-particle multiplicity, mean transverse momentum, and total transverse momentum densities are shown, respectively, all as functions of the leading track-jet's transverse momentum, $p_{\mathrm{T}}^{\mathrm{jet}}$, for five different anti-k_t jet radius choices, $R = 0.2, 0.4, 0.6, 0.8, 1.0$. The new and prior tunes perform similarly, except at low values of $p_{\mathrm{T}}^{\mathrm{jet}}$, where the prior tune deviates away from the data less. Qualitatively, these figures show hadronic activity in the transverse region increasing rapidly at low $p_{\mathrm{T}}^{\mathrm{jet}}$, before it plateaus at higher values. Again, among the observables of this measurement, the mean transverse momentum in the transverse region, $\langle p_{\mathrm{T}} \rangle$, is the most well-modelled, coming within approximately 5% of the data everywhere.

In the N_{ch} distributions, and in the $\sum p_{\mathrm{T}}$ distributions at larger jet radius values, there is the appearance of a bump just before the plateau region that is not well-modelled. This same feature is seen and studied in CDF data as discussed in Ref. [161]. In the CDF study, predicted transverse N_{ch} and $\sum p_{\mathrm{T}}$ distributions are sub-divided into two components, the 'hard scattering' part, which includes the hard scatter as well as the associated initial and final-state radiation, and a 'beam-beam

remnant' part, which includes MPI in addition to the the beam remnants. The distribution of the 'beam-beam remnants' part exhibits the same bump as we see in Figures A.6 and A.8, while the distribution of the initial and final-state radiation components increases monotonically with leading particle p_{T}. Such two components can be seen in our case, particularly in the N_{ch} distributions of Figure A.6 where the bump is visible above the rising initial and final-state radiation component.

Assuming that $p_{\mathrm{T}}^{\mathrm{jet}}$ correlates with the scale of the hard scatter, then the placement of the bump along the $p_{\mathrm{T}}^{\mathrm{jet}}$ reflects the point in the hard scattering scale where MPI becomes maximal. The scale used to generate MB events in SHERPA begins at $\sqrt{s}/2$, which ultimately makes the placement of the bump a centre-of-mass energy effect.

As for the apparent R-dependence of the bump's position (it moves from about 9 GeV at $R = 0.2$ to about 20 GeV at $R = 1.0$), the changing definition of $p_{\mathrm{T}}^{\mathrm{jet}}$ as a function of R may be to blame. Larger-R jets typically have a higher reconstructed p_{T} because they capture more particles within their cones, so events of a given hadronic activity effectively migrate up the $p_{\mathrm{T}}^{\mathrm{jet}}$ axis (taking the bump with them). On another note, the efficiency of larger-R jet reconstruction is higher compared to smaller-R jets in events containing less hadronic activity. This could have the effect of reweighting the N_{ch} and $\sum p_{\mathrm{T}}$ distributions towards lower mean values, thus taking away from the bump on the lower side, and changing the shape of the bump structure. All of these features make this bump a great target of tuning (assuming that the initial and final-state radiation components entering the transverse region are already well-modelled).

The observables defined with respect to a leading calorimeter jet [159] are quite well modelled by both our tune and the prior as shown in Figures A.9 and A.10. In each of these figures, distributions are measured in the transverse, trans-max, and trans-min regions in the top, middle, and bottom rows, respectively. Of these three regions, it is the trans-max region (affected more by initial and final-state radiation than the trans-min region) that is most well-modelled, as can be seen in the

$\sum p_{\mathrm{T}}$ distributions in particular. The modelled $\sum p_{\mathrm{T}}$ distribution in the trans-min region agrees with the data well at intermediate $\sum p_{\mathrm{T}}$, but undershoots at high $\sum p_{\mathrm{T}}$ and overshoots in the lowest-$\sum p_{\mathrm{T}}$ bin. This skew survives to a lesser extent in the combined transverse region, and indicates that the simulated UE is not active enough. Perhaps the target of the tuning should have been the trans-min region, as opposed to all transverse regions here, in order to de-emphasise initial and final-state radiation contributions.

Despite being correlated to $\sum p_{\mathrm{T}}$, the N_{ch} distributions shown in Figure A.10 are more well-modelled. For the total transverse momentum distribution to be too soft but the multiplicity distribution to be well-modelled, simulated UE particles must on average be too soft. This conflicts slightly with the good modelling of the $\langle p_{\mathrm{T}} \rangle$ variables (shown in Figures A.3c and A.7).

A.4.2 Conclusion and future work

This appendix has presented an effort to tune the UE parameters of Sherpa version 3. The final tune parameter values are shown in Table A.3, and tuned MC distributions are shown in comparison to reference data in Figures A.2 – A.10. Despite our successful description of this suite of 7 TeV data, there is still work to be done.

First of all, Sherpa 3 in our case is being used for comparison with LHC data at 13 TeV, so the tune must transfer successfully to this higher energy. This extrapolation in energy is controlled by the η^E parameter featured in Eq. (A.4), which must be fine-tuned. This can be done by tuning this single parameter to similar measurements of the UE at 13 TeV, or by allowing some of the UE parameters of interest tuned here to float again in a fit to the 7 and 13 TeV data together. This latter fine-tuning of many parameters may be needed in order to simultaneously describe the UE in 7 and 13 TeV data. Our attempt to do this energy extrapolation not been successful so far.

Secondly, systematic uncertainties on our final tune parameters must be quantified. This can be done using the 'eigentune' method that was used in Ref. [113]. This method involves finding variations of the tune along the principal directions of the covariance matrix at the tune minimum. Since our χ^2 definition does not account for correlations among the bins and observables, it does not follow true χ^2-distribution statistics, so excursions of $\Delta\chi^2 = \pm 1$ about its minimum do not correspond to variations of $\pm 1\sigma$. Instead, a tolerance, T, must be defined heuristically, such that excursions by this tolerance, $\Delta\chi^2 = \pm T$, will produce variations of the tune that bracket the reference data.

Lastly, developments of the Professor methodology have been discussed that should make the tuning procedure more robust, the first pertaining to over-fitting. It often happened in our experience that Professor would over-fit to a high-statistics peak region, sacrificing agreement with the data in less populated distribution tails. To prevent this behaviour, a "sanity limit" could be imposed at say 5%, such that a fit is not rewarded for improving agreement beyond 5%. This limit could be interpreted as representing a theory uncertainty. A second development could address incompatibilities. If a MC model is not able to describe some measurement, or part thereof, then Professor will currently react by compromising everywhere else to blindly minimise the χ^2. In this scenario, Professor could be asked to not penalise the χ^2 beyond a specified maximum amount per bin or observable. While these envisaged improvements do not replace physical insight, they will further automate the tuning procedure.

Once our tune of the Sherpa 3 UE is finalised, it will inform the tune that is provided with the official Sherpa 3 release.

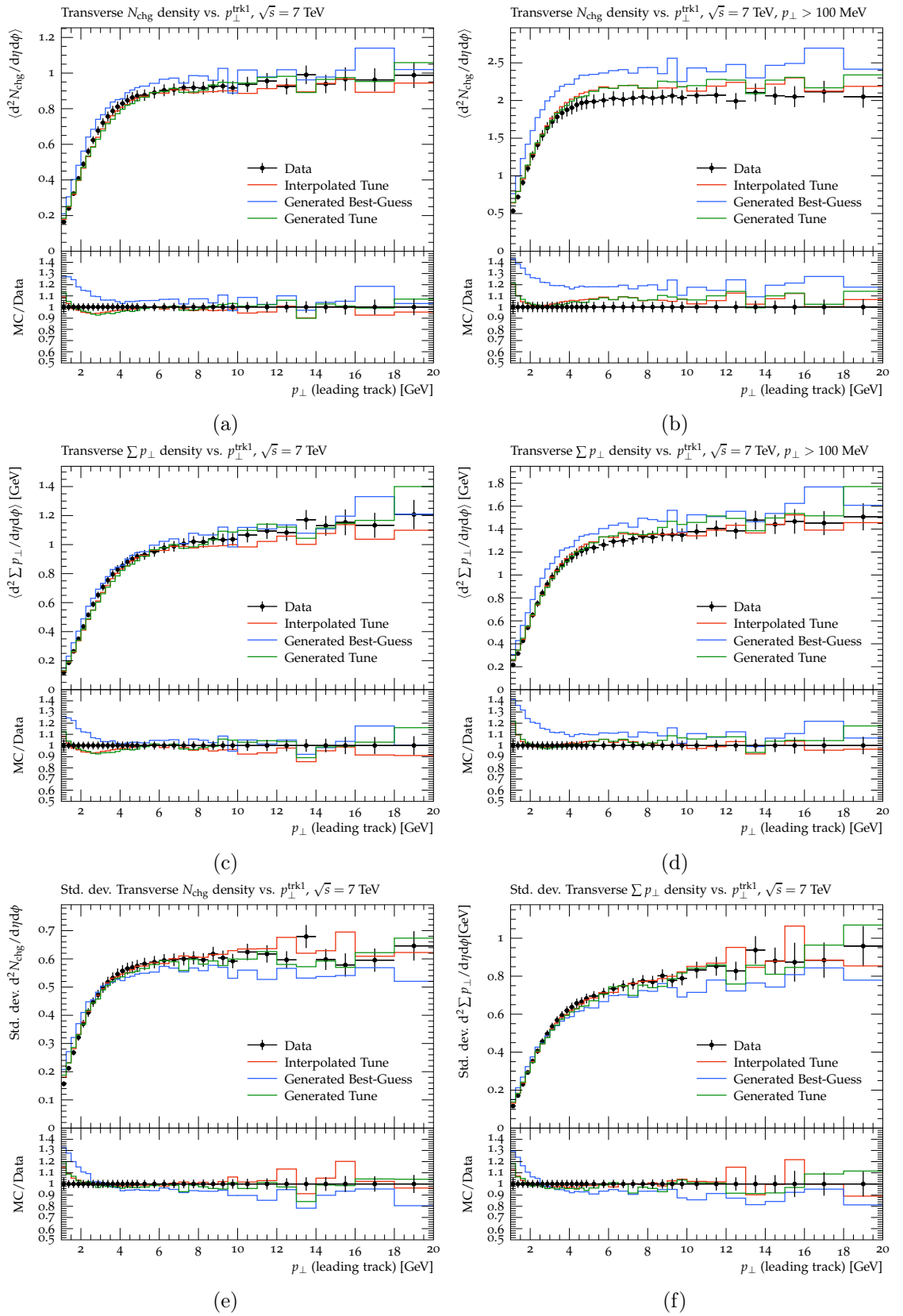


Figure A.2: Charged-particle multiplicity (N_{ch}), total transverse momentum ($\sum p_{\mathrm{T}}$), and standard deviation of N_{ch} and $\sum p_{\mathrm{T}}$, all shown as density profiles in leading track p_{T} for the standard $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ selection (left column) and for the $p_{\mathrm{T}} > 100\,\mathrm{MeV}$ selection (right column). These observables are all measured in the transverse region as part of the leading-track UE measurement [156].

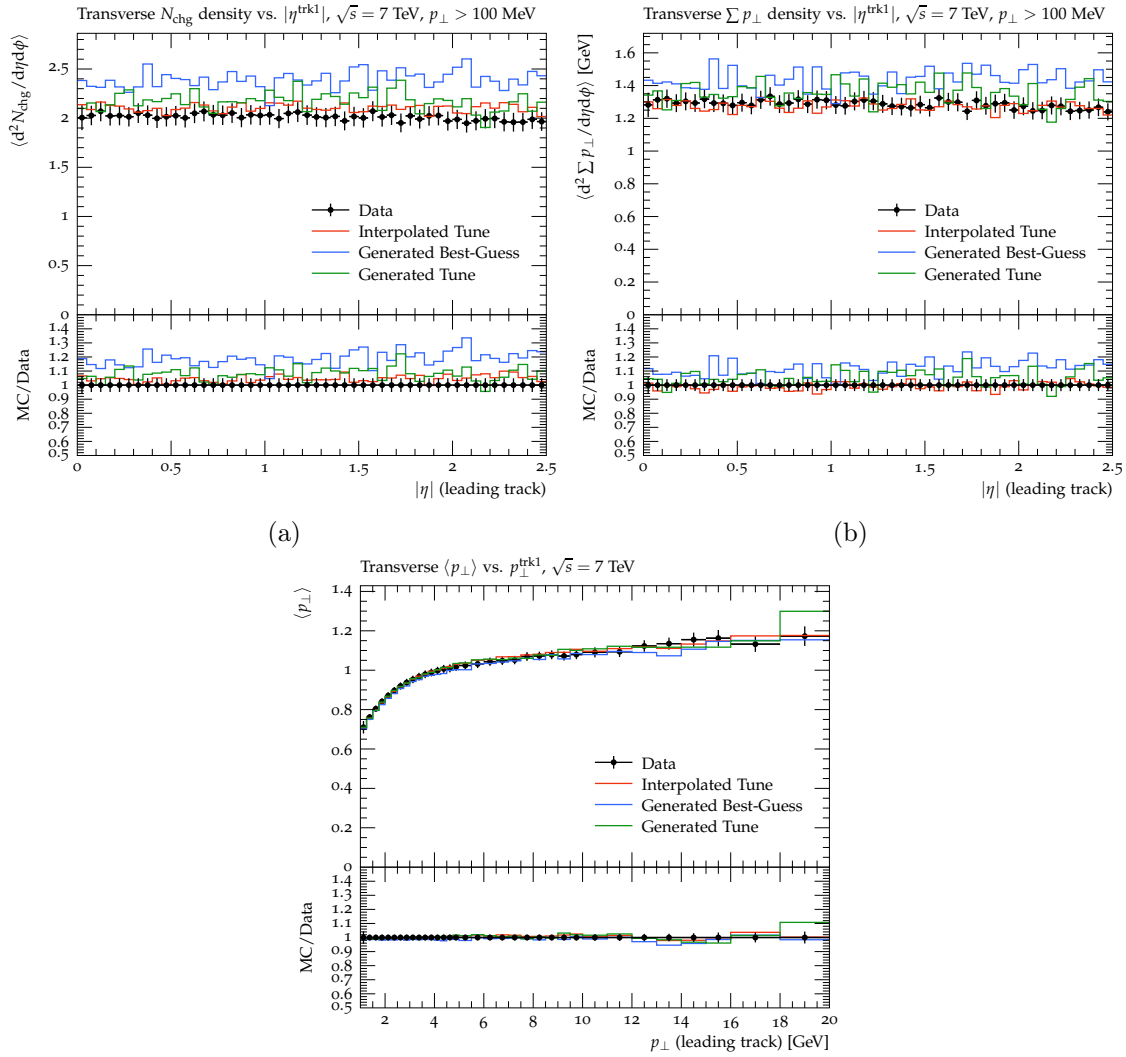
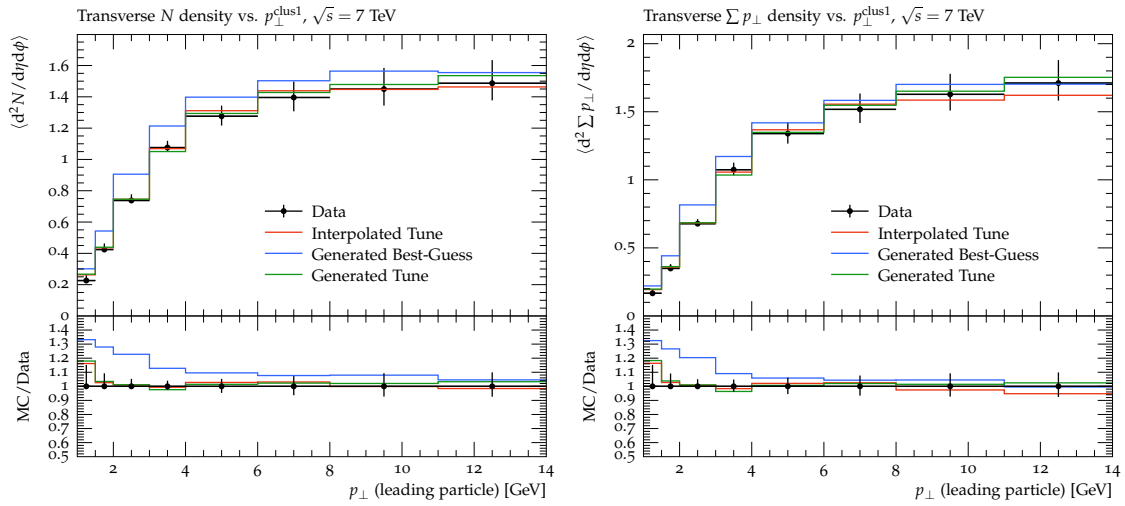


Figure A.3: Charged-particle multiplicity (N_{ch}) and total transverse momentum ($\sum p_{\mathrm{T}}$) density profiles in leading track pseudorapidity, and the mean transverse momentum ($\langle p_{\mathrm{T}} \rangle$) as a profile in leading-track p_{T}, all calculated in the transverse region as part of the leading-track UE measurement [156].



(a) (b)

Figure A.4: The number of charged particles (N), and the total scalar sum of transverse momentum ($\sum p_{\mathrm{T}}$) shown as density profiles in the transverse momentum of the event's leading calorimeter cluster, calculated in the transverse region as part of the calorimeter-based UE measurement [160].

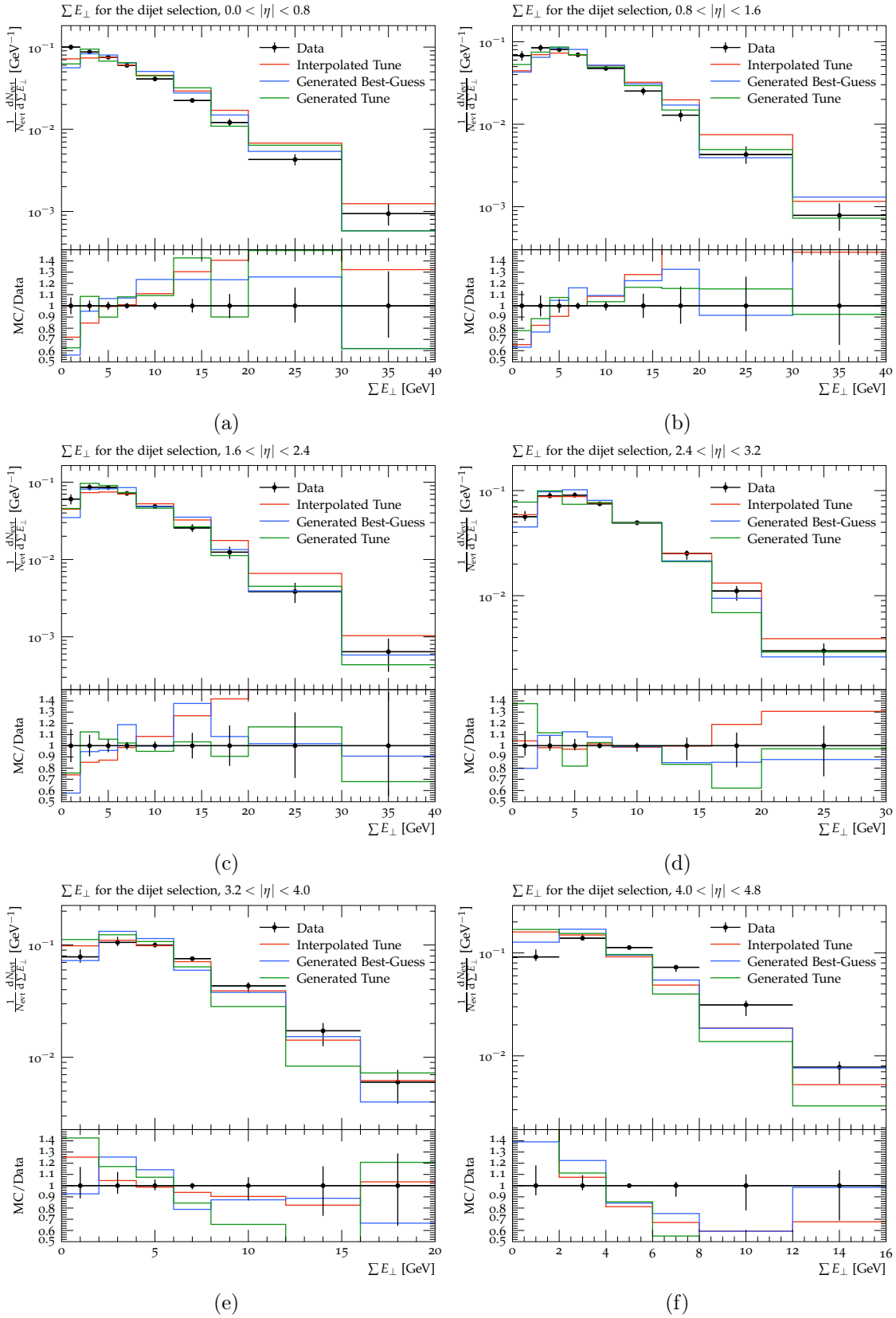


Figure A.5: Number of events, measured differentially in the total scalar sum of transverse energy in the pseudorapidity intervals indicated in the figure titles. These measurements were made to study the pseudorapidity dependence of UE parameters as part of Ref. [158].

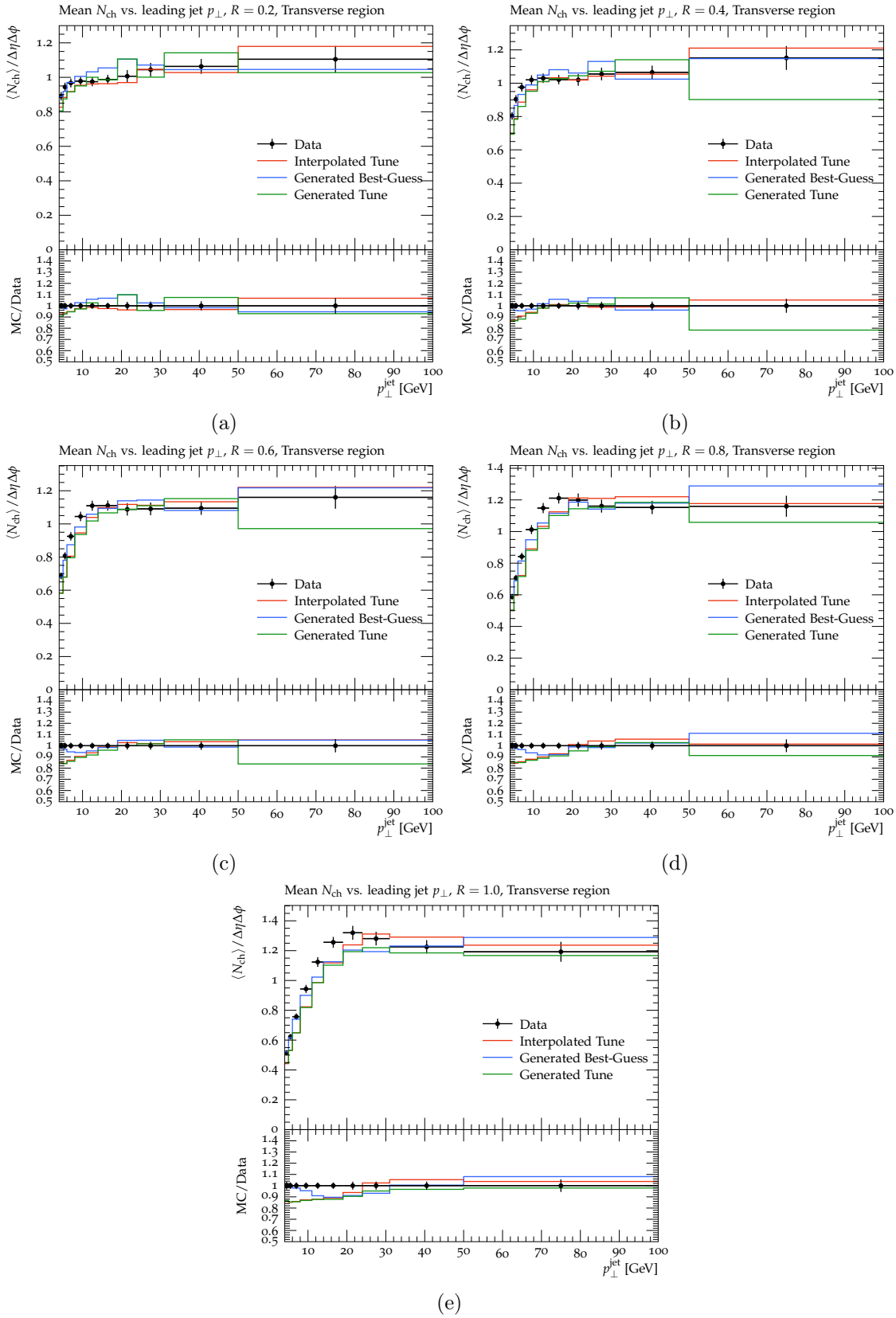


Figure A.6: Charged-particle number density in the transverse region, $N_{ch}/\Delta\eta\Delta\phi$, profiled in the transverse momentum of the leading track-jet in the event, p_T^{jet}, for the different jet-radius values specified in the figure titles. These measurements are detailed in Ref. [157].

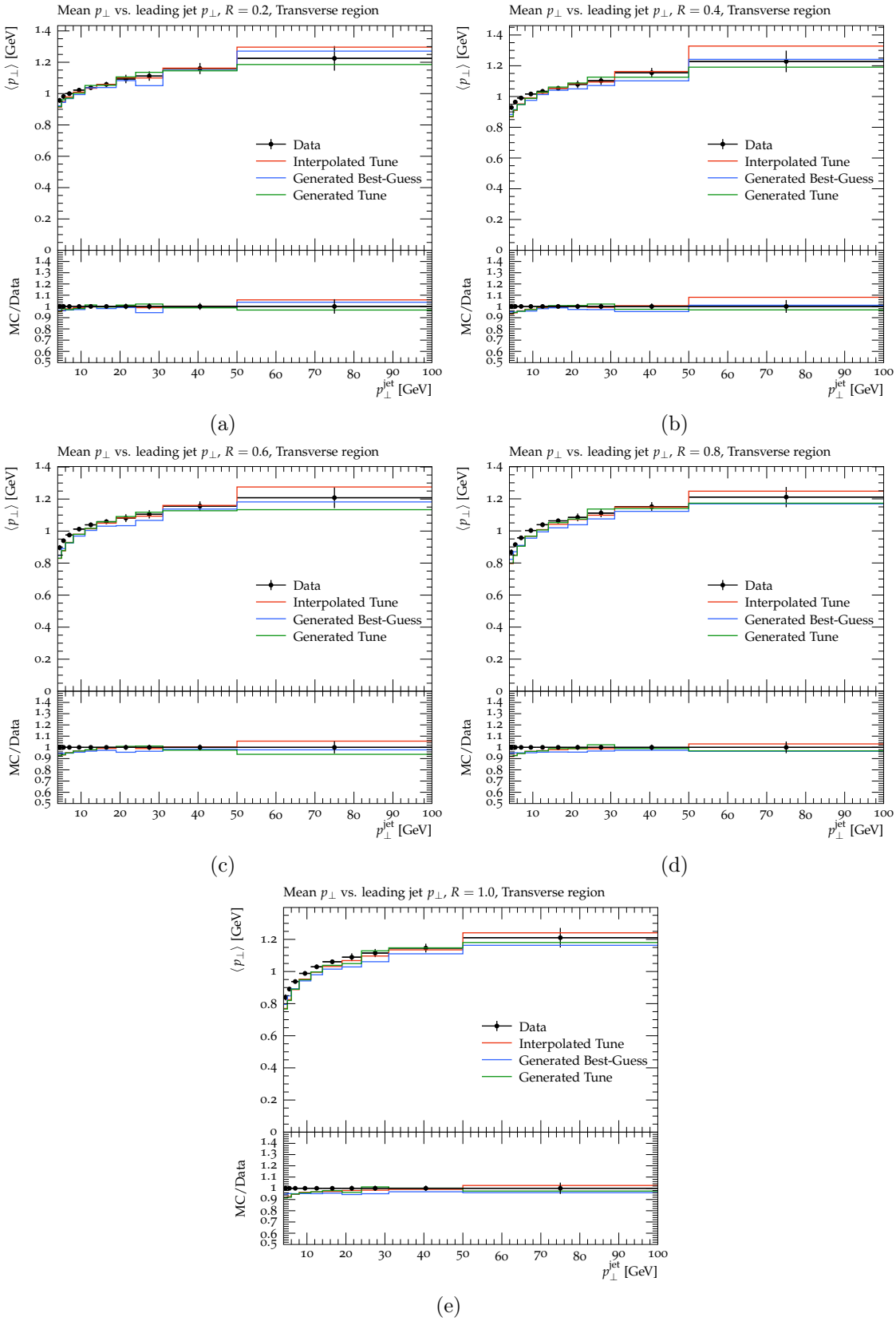


Figure A.7: Mean charged-particle transverse momentum in the transverse region, $\langle p_{\mathrm{T}} \rangle$, as a function of the transverse momentum of the leading track-jet in the event, $p_{\mathrm{T}}^{\mathrm{jet}}$, for the different jet-radius values specified in the figure titles. These measurements are detailed in Ref. [157].

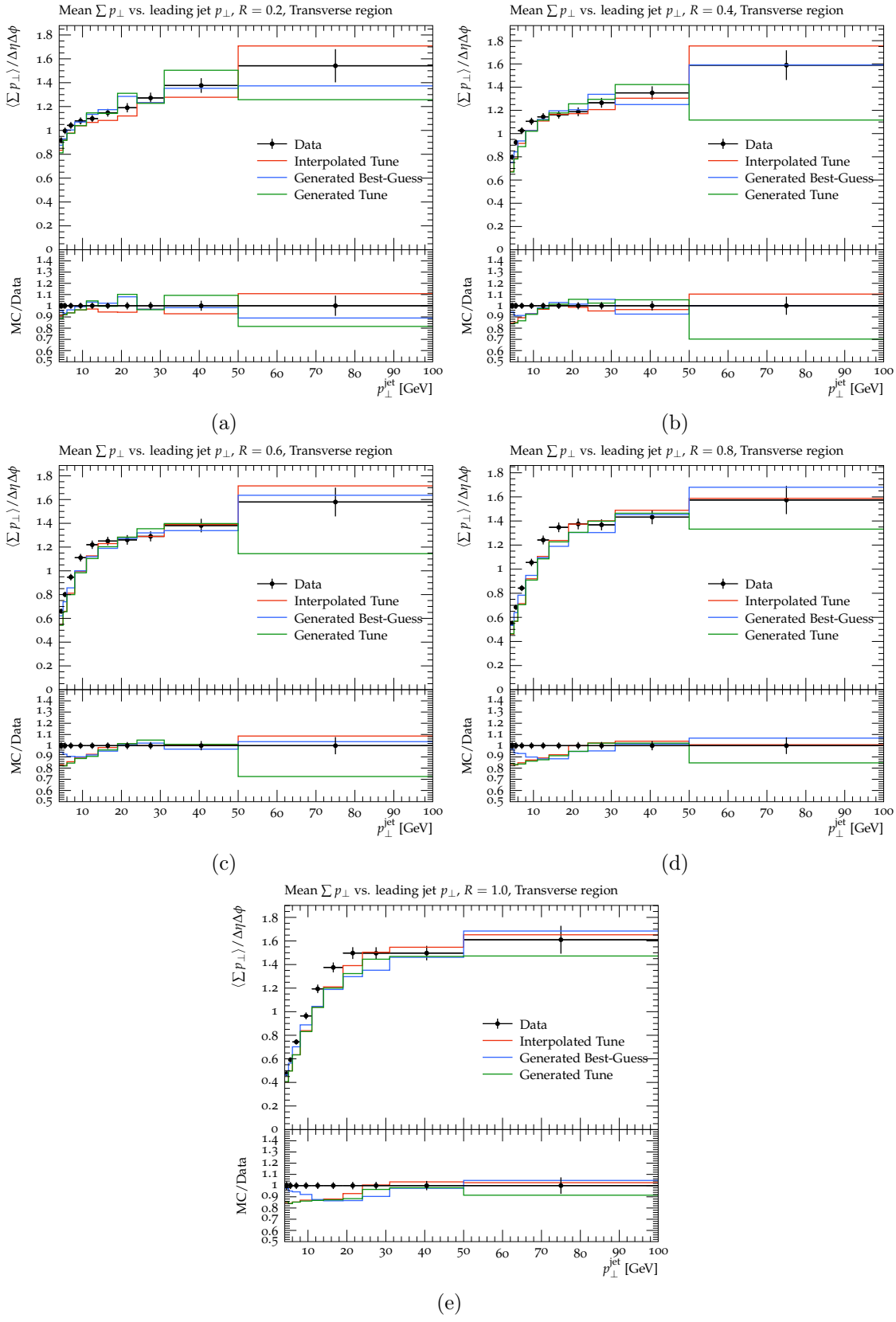


Figure A.8: Mean scalar sum of charged-particle transverse momenta in the transverse region, $\langle \sum p_{\mathrm{T}} \rangle$, profiled in the transverse momentum of the leading track-jet in the event, $p_{\mathrm{T}}^{\mathrm{jet}}$, for the different jet-radius values specified in the figure titles. These measurements are detailed in Ref. [157].

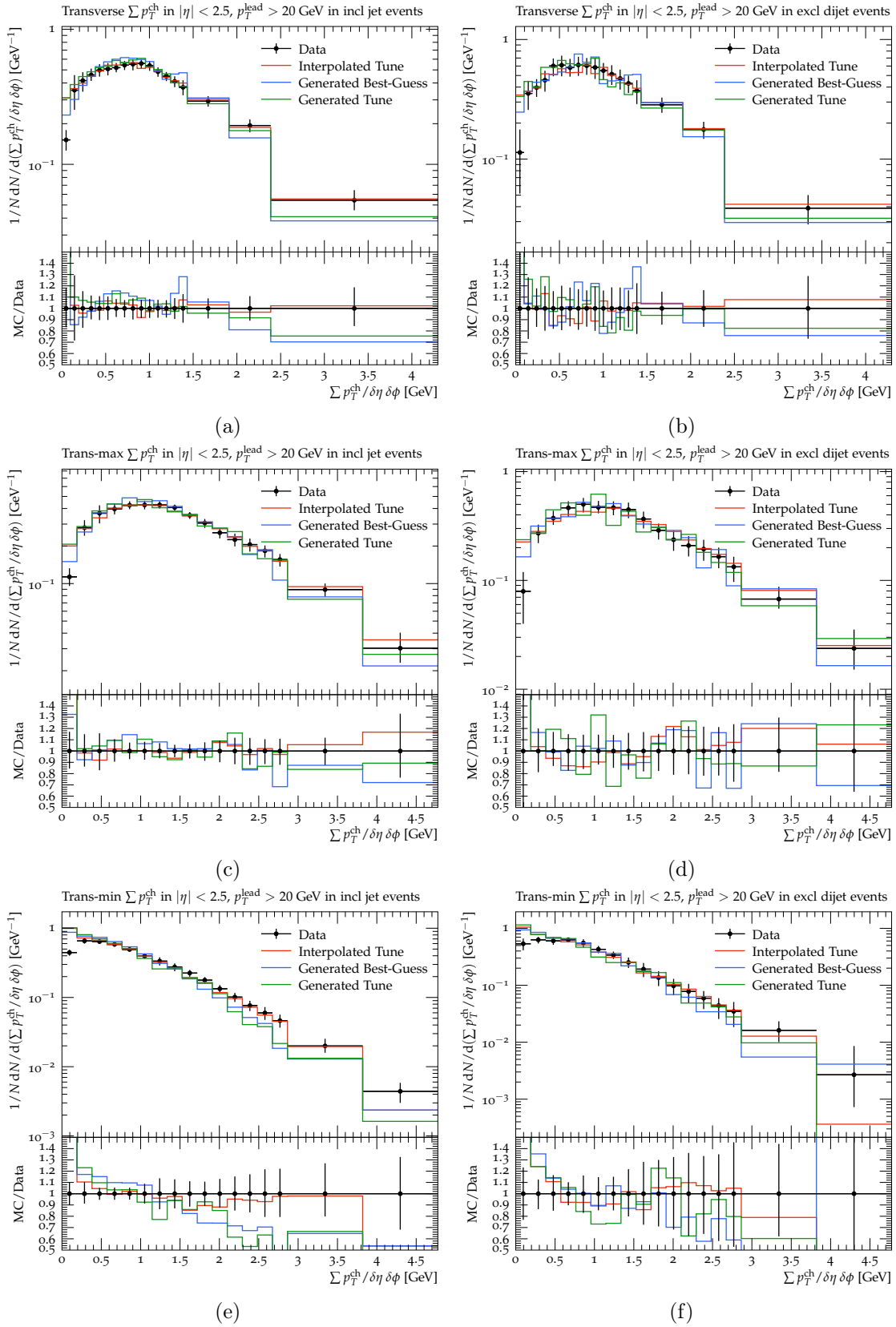


Figure A.9: Number of events, measured differentially in the density of the total scalar sum of charged-particle transverse momenta, $\sum p_{\mathrm{T}}^{\mathrm{ch}}/\delta\eta\delta\phi$. This distribution is measured in the transverse region (top row), trans-max region (middle row), and trans-min region (bottom row), after inclusive jet (left column) and exclusive dijet (right column) selections, respectively [159].

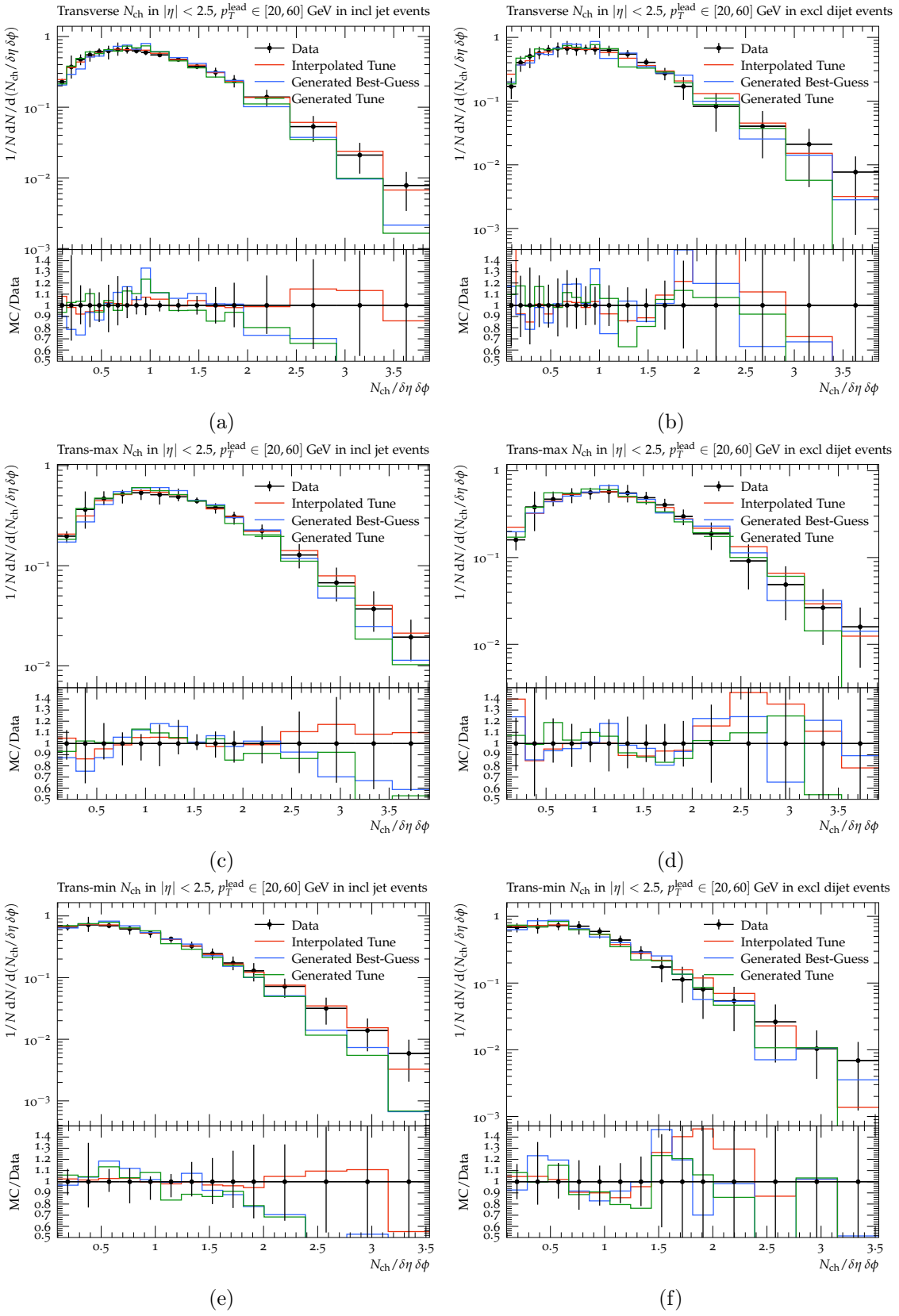


Figure A.10: Number of events, measured differentially in the density of the number of charged-particles in the event, $N_{ch}/\delta\eta\delta\phi$. This distribution is measured in the transverse region (top row), trans-max region (middle row), and trans-min region (bottom row), after inclusive jet (left column) and exclusive dijet (right column) selections, respectively [159].

Bibliography

[1] V. V. Khoze, F. Krauss & M. Schott. *Large effects from small QCD instantons: making soft bombs at hadron colliders*. *JHEP* **2020,** 201. arXiv: 1911.09726 [hep-ph] (Apr. 2020).

[2] P. A. M. Dirac. *On the Analogy Between Classical and Quantum Mechanics*. *Rev. Mod. Phys.* **17,** 195–199 (2-3 Apr. 1945).

[3] P. A. M. Dirac. *The Principles of Quantum Mechanics* 2nd (The Clarendon Press, 1935).

[4] R. P. Feynman. *Space-Time Approach to Non-Relativistic Quantum Mechanics*. *Rev. Mod. Phys.* **20,** 367–387 (2 Apr. 1948).

[5] M. E. Peskin & D. V. Schroeder. *Introduction to Quantum Field Theory*. ISBN: 978-0-201-50397-5 (Addison-Wesley, Reading, USA, 1995).

[6] S. Weinberg. *The Quantum Theory of Fields, Vol. 2: Modern Applications*. ISBN: 0-521-55002-5 (Cambridge University Press, Aug. 1996).

[7] Pierre Ramond. *Field Theory: A Modern Primer* 2nd. ISBN: 0-201-54611-6 (Addison-Wesley, 1990).

[8] S. Coleman. *Aspects of Symmetry*. ISBN: 9780511565045 (Cambridge University Press, 1985).

[9] G. 't Hooft & M. Veltman. *Regularization and renormalization of gauge fields*. *Nuclear Physics B* **44,** 189–213 (1972).

[10] H. D. Politzer. *Reliable Perturbative Results for Strong Interactions? Phys. Rev. Lett.* **30,** 1346–1349 (26 June 1973).

[11] D. J. Gross & F. Wilczek. *Ultraviolet Behavior of Non-Abelian Gauge Theories. Phys. Rev. Lett.* **30,** 1343–1346 (26 June 1973).

[12] S. L. Adler. *Axial-Vector Vertex in Spinor Electrodynamics. Phys. Rev.* **177,** 2426–2438 (5 Jan. 1969).

[13] J. S. Bell & R. W. Jackiw. *A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ-model. Nuovo Cimento* **60,** 47–61 (1969).

[14] K. Fujikawa. *Path-Integral Measure for Gauge-Invariant Fermion Theories. Phys. Rev. Lett.* **42,** 1195–1198 (18 Apr. 1979).

[15] E. Bothmann *et al. Event generation with Sherpa 2.2. SciPost Phys.* **7,** 034. arXiv: 1905.09127 [hep-ph] (2019).

[16] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen & P. Z. Skands. *An introduction to PYTHIA 8.2. Comp. Phys. Comm.* **191,** 159–177. arXiv: 1207.1303 [hep-ph] (June 2015).

[17] C. Bierlich *et al. A comprehensive guide to the physics and usage of PYTHIA 8.3.* 2022. arXiv: 2203.11601 [hep-ph].

[18] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko & K. Werner. *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider. Phys. Rev. C* **92,** 034906. arXiv: 1306.0121 [hep-ph] (Sept. 2015).

[19] A. Buckley *et al. General-purpose event generators for LHC physics. Physics Reports* **504,** 145–233. arXiv: 1101.2599 [hep-ph] (2011).

[20] F. Gross *et al. 50 Years of quantum chromodynamics: Introduction and Review. Eur. Phys. J. C* **83.** arXiv: 2212.11107 [hep-ph] (Dec. 2023).

[21] T. Sjöstrand & P. Z. Skands. *Transverse-momentum-ordered showers and interleaved multiple interactions. Eur. Phys. J. C* **39,** 129–154 (Feb. 2005).

[22] S. Schumann & F. Krauss. *A parton shower algorithm based on Catani-Seymour dipole factorisation. JHEP* **2008,** 038–038 (Mar. 2008).

[23] J.-C. Winter & F. Krauss. *Initial-state showering based on colour dipoles connected to incoming parton lines. JHEP* **2008,** 040–040 (July 2008).

[24] T. Sjöstrand. *The Lund Monte Carlo for jet fragmentation. Comp. Phys. Comm.* **27,** 243–284 (1982).

[25] B. Andersson, G. Gustafson & B. Soderberg. *A General Model for Jet Fragmentation. Z. Phys. C* **20,** 317 (1983).

[26] B. R. Webber. *A QCD Model for Jet Fragmentation Including Soft Gluon Interference. Nucl. Phys. B* **238,** 492–528 (1984).

[27] D. Amati & G. Veneziano. *Preconfinement as a Property of Perturbative QCD. Phys. Lett. B* **83,** 87–92 (1979).

[28] K. Werner. *Core-Corona Separation in Ultrarelativistic Heavy Ion Collisions. Phys. Rev. Lett.* **98,** 152301. arXiv: 0704.1270 [nucl-th] (15 Apr. 2007).

[29] T. Sjöstrand & M. van Zijl. *A multiple-interaction model for the event structure in hadron collisions. Phys. Rev. D* **36,** 2019–2041 (7 Oct. 1987).

[30] T Sjöstrand and P Skands. *Multiple Interactions and the Structure of Beam Remnants. JHEP* **2004,** 053–053 (Mar. 2004).

[31] K. Werner, F.-M. Liu & T. Pierog. *Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider. Phys. Rev. C* **74,** 044902. arXiv: 0506232 [hep-ph] (Oct. 2006).

[32] H. Drescher, M. Hladik, S. Ostapchenko, T. Pierog & K. Werner. *Parton-based Gribov–Regge theory. Phys. Rep.* **350,** 93–289. arXiv: hep-ph/0007198 (2001).

[33] ATLAS Collaboration. *The ATLAS Simulation Infrastructure. Eur. Phys. J. C* **70,** 823. arXiv: 1005.4568 [physics.ins-det] (2010).

[34] J. Allison *et al. Recent developments in* GEANT4. *Nucl. Instrum. Meth. A* **835,** 186–225 (2016).

[35] *Standard Model Summary Plots February 2022.* Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-009 (CERN, Geneva, 2022).

[36] A. Belavin, A. Polyakov, A. Schwartz & Y. Tyupkin. *Pseudoparticle solutions of the Yang-Mills equations. Phys. Lett. B* **59,** 85–87. ISSN: 0370-2693 (1975).

[37] G. 't Hooft. *Symmetry Breaking through Bell-Jackiw Anomalies. Phys. Rev. Lett.* **37,** 8–11 (1 July 1976).

[38] G. 't Hooft. *How instantons solve the U(1) problem. Phys. Rept.* **142,** 357–387. ISSN: 0370-1573 (1986).

[39] E. Shuryak. *The role of instantons in quantum chromodynamics: (II). Hadronic structure. Nucl. Phys. B* **203,** 116–139. ISSN: 0550-3213 (1982).

[40] J. C. Biddle, W. Kamleh & D. B. Leinweber. *Visualization of center vortex structure. Phys. Rev. D* **102,** 034504 (3 Aug. 2020).

[41] H1 Collaboration. *Search for QCD instanton-induced processes in deep-inelastic scattering at HERA. Eur. Phys. J. C* **25,** 495–509 (Nov. 2002).

[42] ZEUS Collaboration. *Search for QCD-instanton induced events in deep inelastic ep scattering at HERA. Eur. Phys. J. C* **34,** 255–265. ISSN: 1434-6052 (May 2004).

[43] H1 Collaboration. *Search for QCD Instanton-Induced Processes at HERA in the High-Q^2 Domain.* 2016. arXiv: 1603.05567 [hep-ex].

[44] G. 't Hooft. *Computation of the quantum effects due to a four-dimensional pseudoparticle. Phys. Rev. D* **14,** 3432–3450 (12 Dec. 1976).

[45] C. Callan, R. Dashen & D. Gross. *The structure of the gauge theory vacuum. Phys. Lett. B* **63,** 334–340 (Aug. 1976).

[46] R. Shankar. *Principles of Quantum Mechanics.* ISBN: 978-0-306-44790-7 (Springer New York, NY, Aug. 1994).

[47] R. Jackiw & C. Rebbi. *Vacuum Periodicity in a Yang-Mills Quantum Theory. Phys. Rev. Lett.* **37,** 172–175 (3 July 1976).

[48] R. D. Peccei & H. R. Quinn. CP *Conservation in the Presence of Pseudoparticles. Phys. Rev. Lett.* **38,** 1440–1443 (25 June 1977).

[49] S. Weinberg. *A New Light Boson? Phys. Rev. Lett.* **40,** 223–226 (4 Jan. 1978).

[50] F. Wilczek. *Problem of Strong P and T Invariance in the Presence of Instantons. Phys. Rev. Lett.* **40,** 279–282 (5 Jan. 1978).

[51] V. V. Khoze, D. L. Milne & M. Spannowsky. *Searching for QCD instantons at hadron colliders. Phys. Rev. D* **103,** 014017 (1 Jan. 2021).

[52] V. A. Khoze, V. V. Khoze, D. L. Milne & M. G. Ryskin. *Hunting for QCD Instantons at the LHC in Events with Large Rapidity Gaps. Phys. Rev. D* **104,** 054013 (5 Sept. 2021).

[53] S. Amoroso, D. Kar & M. Schott. *How to discover QCD Instantons at the LHC. Eur. Phys. J. C* **81,** 624. ISSN: 1434-6052 (July 2021).

[54] A. Mueller. *First quantum corrections to gluon-gluon collisions in the one-instanton sector. Nucl. Phys. B* **348,** 310–326. ISSN: 0550-3213 (1991).

[55] C. Caprini, T. Carli, M. Schott & M. McCullough. *Topological Effects in the Standard Model.* CERN-TH Workshop, https://indico.cern.ch/event/965112. Dec. 2020.

[56] M. Mangano. *Remarks on the numerical impact of potential theoretical systematics in the prediction of QCD instanton cross sections*. in *Topological Effects in the Standard Model: Instantons, Sphalerons and Beyond at LHC* (Jan. 2021). arXiv: 2101.02719 [hep-ph].

[57] E. Lopienska. *The CERN accelerator complex*. CERN-GRAPHICS-2022-001-1. (2023).

[58] L. Evans & P. Bryant. *LHC Machine*. *JINST* **3,** S08001 (Aug. 2008).

[59] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. *JINST* **3,** S08003 (Aug. 2008).

[60] CMS Collaboration. *The CMS experiment at the CERN LHC*. *JINST* **3,** S08004 (Aug. 2008).

[61] B. Salvachua. *Overview of Proton-Proton Physics during Run 2*. in *9th LHC Operations Evian Workshop* (2019), 7–14.

[62] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 2*. https://cds.cern.ch/record/391177 (1999).

[63] K. Anthony. *ATLAS celebrates results of 1000 collision papers*. (2023).

[64] R. L. Workman *et al. Review of Particle Physics*. *PTEP* **2022,** 083C01 (2022).

[65] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report, Volume 1*. ATLAS-TDR-4; CERN-LHCC-97-016 (1997).

[66] ATLAS Collaboration. *The ATLAS Inner Detector commissioning and calibration*. *EPJC* **70,** 787–821. arXiv: 1004.5293 [ins-det] (Aug. 2010).

[67] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*. Accepted by JINST. arXiv: 2305.16623 [physics.ins-det] (2023).

[68] G. Aad *et al. ATLAS pixel detector electronics and sensors*. *JINST* **3,** P07007 (July 2008).

[69] ATLAS Collaboration. *ATLAS Insertable B-Layer: Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013 (2010). Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009 (2012). https://cds.cern.ch/record/1451888.

[70] B. Abbott *et al. Production and integration of the ATLAS Insertable B-Layer*. *JINST* **13,** T05008 (May 2018).

[71] ATLAS Collaboration. *Measurements of sensor radiation damage in the ATLAS inner detector using leakage currents*. *JINST* **16,** P08025. arXiv: 2106.09287 [hep-ex] (2021).

[72] S. Parker, C. Kenney & J. Segal. *3D — A proposed new architecture for solid-state radiation detectors*. *Nucl. Instrum. Meth. A* **395,** 328–343. ISSN: 0168-9002 (1997).

[73] A. Ahmad *et al. The silicon microstrip sensors of the ATLAS semiconductor tracker*. *Nucl. Instrum. Meth. A* **578,** 98–118. ISSN: 0168-9002 (2007).

[74] A. Abdesselam *et al. The barrel modules of the ATLAS semiconductor tracker*. *Nucl. Instrum. Meth. A* **568,** 642–671. ISSN: 0168-9002 (2006).

[75] A. Abdesselam *et al. The ATLAS semiconductor tracker end-cap module*. *Nucl. Instrum. Meth. A* **575,** 353–389. ISSN: 0168-9002 (2007).

[76] The ATLAS TRT Collaboration. *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*. *JINST* **3,** P02013 (Feb. 2008).

[77] The ATLAS TRT Collaboration. *The ATLAS TRT Barrel Detector*. *JINST* **3,** P02014 (Feb. 2008).

[78] The ATLAS TRT Collaboration. *The ATLAS TRT end-cap detectors*. *JINST* **3,** P10003 (Oct. 2008).

[79] ATLAS Collaboration. *Efficiency and Hit Spatial Resolution of ATLAS IBL Sensors in LHC Run 2 Collision Events*. ATL-INDET-PUB-2016-001. 2016.

[80] ATLAS Collaboration. *Basic ATLAS TRT performance studies of Run 1.* ATL-INDET-PUB-2014-001. 2014.

[81] ATLAS Collaboration. *ATLAS Calorimeter Performance: Technical Design Report.* ATLAS-TDR-1; CERN-LHCC-96-040 (1996).

[82] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter: Technical Design Report.* ATLAS-TDR-2; CERN-LHCC-96-041 (1996).

[83] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report.* ATLAS-TDR-3; CERN-LHCC-96-042 (1996).

[84] A. Sidoti. *Minimum Bias Trigger Scintillators in ATLAS Run II. JINST* **9,** C10020 (Oct. 2014).

[85] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report.* ATLAS-TDR-10; CERN-LHCC-97-022 (CERN, 1997).

[86] ATLAS Collaboration. *Operation of the ATLAS trigger system in Run 2. JINST* **15,** P10004. arXiv: 2007.12539 [hep-ex] (2020).

[87] M. Elsing, L. Goossens, A. Nairz & G. Negri. *The ATLAS Tier-0: Overview and operational experience. J. Phys. Conf. Ser.* **219,** 072011 (Apr. 2010).

[88] ATLAS Collaboration. *ATLAS data quality operations and performance for 2015–2018 data-taking. JINST* **15,** P04003. arXiv: 1911.04632 [physics.ins-det] (2020).

[89] ATLAS Collaboration. *Luminosity determination in pp collisions at* $\sqrt{s} =$ 13 *TeV using the ATLAS detector at the LHC.* arXiv: 2212.09379 [hep-ex] (2022).

[90] ATLAS Collaboration. *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT).* ATL-SOFT-PUB-2007-007. 2007.

[91] R. Frühwirth. *Application of Kalman filtering to track and vertex fitting. Nucl. Instrum. Meth. A* **262,** 444–450. ISSN: 0168-9002 (1987).

[92] R. O. Duda & P. E. Hart. *Use of the Hough Transformation to Detect Lines and Curves in Pictures. Commun. ACM* **15,** 11–15. ISSN: 0001-0782 (Jan. 1972).

[93] ATLAS Collaboration. *Early Inner Detector Tracking Performance in the 2015 Data at* $\sqrt{s} = 13$ *TeV.* ATL-PHYS-PUB-2015-051. 2015.

[94] ATLAS Collaboration. *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. Eur. Phys. J. C* **77,** 332. arXiv: 1611.10235 [hep-ex] (2017).

[95] G. Piacquadio, K. Prokofiev & A. Wildauer. *Primary vertex reconstruction in the ATLAS experiment at LHC. J. Phys. Conf. Ser.* **119,** 032033 (July 2008).

[96] ATLAS Collaboration. *Jet reconstruction and performance using particle flow with the ATLAS Detector. Eur. Phys. J. C* **77,** 466. arXiv: 1703.10485 [hep-ex] (2017).

[97] M. Cacciari, G. P. Salam & G. Soyez. *The anti-k_t jet clustering algorithm. JHEP* **04,** 063. arXiv: 0802.1189 [hep-ph] (2008).

[98] M. Cacciari, G. P. Salam & G. Soyez. *FastJet user manual. Eur. Phys. J. C* **72,** 1896. arXiv: 1111.6097 [hep-ph] (2012).

[99] ATLAS Collaboration. *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset.* arXiv: 2211.16345 [physics.data-an] (2022).

[100] ATLAS Collaboration. *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at* $\sqrt{s} = 13$ *TeV. Eur. Phys. J. C* **79,** 970. arXiv: 1907.05120 [hep-ex] (2019).

[101] 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,** 1. arXiv: 1207.7214 [hep-ex] (2012).

[102] CMS Collaboration. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B* **716,** 30. arXiv: 1207.7235 [hep-ex] (2012).

[103] ATLAS Collaboration. *Charged-particle distributions in* $\sqrt{s} = 13$ *TeV pp interactions measured with the ATLAS detector at the LHC. Phys. Lett. B* **758,** 67. arXiv: 1602.01633 [hep-ex] (2016).

[104] T. Martin. *Strange* K_0^S *and* Λ *particle spectra at ATLAS as probes of the Underlying Event in pp interactions at* $\sqrt{s} = 13$ *TeV.* Tech. rep. (CERN, Geneva, 2018).

[105] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert & W. J. *Event generation with SHERPA 1.1. JHEP* **02,** 007. arXiv: 0811.4622 [hep-ph] (2009).

[106] R. Kleiss, W. Stirling & S. Ellis. *A new Monte Carlo treatment of multiparticle phase space at high energies. Comp. Phys. Comm.* **40,** 359–373. ISSN: 0010-4655 (1986).

[107] A. Buckley, H. Hoeth, H. Lacker, H. Schulz & J. E. von Seggern. *Systematic event generator tuning for the LHC. Eur. Phys. J. C* **65,** 331–357. arXiv: 0907.2973 [hep-ph] (Nov. 2009).

[108] F. Krauss. Private communication. July 2023.

[109] S. Schumann & M. Knobbe. Private communication. Nov. 2022.

[110] The NNPDF Collaboration, R. D. Ball, *et al. Parton distributions for the LHC run II. JHEP* **04,** 040. arXiv: 1410.8849 [hep-ph] (2015).

[111] H. Drescher, M. Hladik, S. Ostapchenko, T. Pierog & K. Werner. *Parton-based Gribov–Regge theory. Phys. Rep.* **350,** 93–289 (Sept. 2001).

[112] ATLAS Collaboration. *Multijet simulation for 13 TeV ATLAS Analyses.* ATL-PHYS-PUB-2019-017. Apr. 2019.

[113] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. Nov. 2014.

[114] ATLAS Collaboration. *Low-p_T tracking in nominal pileup*. IDTR-2020-001. 2020.

[115] A. Banfi, G. P. Salam & G. Zanderighi. *Phenomenology of event shapes at hadron colliders*. *JHEP* **2010,** 038. ISSN: 1029-8479. arXiv: 1001.4082 [hep-ph] (June 2010).

[116] ATLAS Collaboration. *Measurement of hadronic event shapes in multijet final states at $\sqrt{s} = 13$ TeV with the ATLAS detector*. ATLAS-CONF-2020-011. 2020.

[117] D. P. Barber *et al. Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA*. *Phys. Rev. Lett.* **43,** 830–833 (12 Sept. 1979).

[118] C. Cesarotti & J. Thaler. *A Robust Measure of Event Isotropy at Colliders*. 2020. arXiv: 2004.06125 [hep-ph].

[119] T. Sjöstrand, S. Mrenna & P. Z. Skands. *PYTHIA 6.4 physics and manual*. *JHEP* **05,** 026. arXiv: hep-ph/0603175 (2006).

[120] A. Ortiz & G. Paic. *A look into the hedgehog events in pp collisions using a new event shape flatenicity*. 2022. arXiv: 2204.13733 [hep-ph].

[121] ATLAS Collaboration. *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. *Eur. Phys. J. C* **81,** 689. arXiv: 2007.02645 [hep-ex] (2020).

[122] ATLAS Collaboration. *Soft b-hadron tagging for compressed SUSY scenarios*. ATLAS-CONF-2019-027. July 2019.

[123] ATLAS Collaboration. *Two-particle azimuthal correlations in photonuclear ultraperipheral Pb+Pb collisions at* 5.02 *TeV with ATLAS*. *Phys. Rev. C* **104,** 014903. arXiv: 2101.10771 [hep-ex] (2021).

[124] T. Carli, J. Gerigk, A. Ringwald & F. Schrempp. *QCD Instanton-induced Processes in Deep-Inelastic Scattering - Search Strategies and Model Dependencies*. 1999. arXiv: hep-ph/9906441 [hep-ph].

[125] S. Knapen, S. P. Griso, M. Papucci & D. J. Robinson. *Triggering soft bombs at the LHC. JHEP* **2017.** arXiv: 1612.00850 [hep-ph] (Aug. 2017).

[126] G. Kasieczka, B. Nachman, M. D. Schwartz & D. Shih. *Automating the ABCD method with machine learning. Phys. Rev. D* **103,** 035021. arXiv: 2007.14400 [hep-ph] (3 Feb. 2021).

[127] F. Chollet *et al. Keras*. https://keras.io. 2015.

[128] G. J. Székely, M. L. Rizzo & N. K. Bakirov. *Measuring and testing dependence by correlation of distances. Ann. Statist.* **35,** 2769–2794 (2007).

[129] G. Kasieczka & D. Shih. *Robust Jet Classifiers through Distance Correlation. Phys. Rev. Lett.* **125,** 122001. arXiv: 2001.05310 [hep-ph] (12 Sept. 2020).

[130] Martín Abadi *et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.

[131] F. Pedregosa *et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research* **12,** 2825–2830 (2011).

[132] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, *et al. KerasTuner*. https://github.com/keras-team/keras-tuner. 2019.

[133] D. P. Kingma & J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[134] S. M. Lundberg & S.-I. Lee. in *Advances in Neural Information Processing Systems 30* 4765–4774 (Curran Associates, Inc., 2017). arXiv: 1705.07874 [cs].

[135] L. S. Shapley. in *Contributions to the Theory of Games (AM-28), Volume II* 307–318 (Princeton University Press, Princeton, 1953). ISBN: 9781400881970.

[136] B. Rozemberczki, L. Watson, P. Bayer, H.-T. Yang, O. Kiss, S. Nilsson & R. Sarkar. *The Shapley Value in Machine Learning*. 2022. arXiv: 2202.05594.

[137] G. J. Feldman & R. D. Cousins. *Unified approach to the classical statistical analysis of small signals*. *Phys. Rev. D* **57,** 3873–3889 (7 Apr. 1998).

[138] G. Cowan, K. Cranmer, E. Gross & O. Vitells. *Asymptotic formulae for likelihood-based tests of new physics*. *Eur. Phys. J. C* **71.** [Erratum: Eur. Phys. J. C **73**, 2501 (2013)], 1554. arXiv: 1007.1727 [physics.data-an] (2011).

[139] A. L. Read. *Presentation of search results: the CLs technique*. *J. Phys. G: Nucl. Part. Phys.* **28,** 2693 (Sept. 2002).

[140] M. Baak, G. J. Besjes, D. Côté, A. Koutsman, J. Lorenz & D. Short. *HistFitter software framework for statistical data analysis*. *Eur. Phys. J. C* **75,** 153 (Apr. 2015).

[141] S. S. Wilks. *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*. *Ann. Math. Stat.* **9,** 60–62 (1938).

[142] A. Wald. *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*. *Trans. Amer. Math. Soc.* **54,** 426–482 (1943).

[143] ATLAS Collaboration. *Tracking Studies for b-tagging with 7 TeV Collision Data with the ATLAS Detector*. ATLAS-CONF-2010-070.

[144] ATLAS Collaboration. *Study of alignment-related systematic effects on the ATLAS Inner Detector track reconstruction*. ATLAS-CONF-2012-141. 2012.

[145] ATLAS Collaboration. *Search for dark matter produced in association with bottom or top quarks in* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*. *Eur. Phys. J. C* **78,** 18. arXiv: 1710.11412 [hep-ex] (2018).

[146] ATLAS Collaboration. *Search for displaced leptons in* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*. *Phys. Rev. Lett.* **127,** 051802. arXiv: 2011.07812 [hep-ex] (2020).

[147] ATLAS Collaboration. *Search for top squarks in events with a Higgs or Z boson using* $139\,fb^{-1}$ *of pp collision data at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector. Eur. Phys. J. C* **80,** 1080. arXiv: 2006.05880 [hep-ex] (2020).

[148] ATLAS Collaboration. *Constraints on spin-0 dark matter mediators and invisible Higgs decays using ATLAS* 13 *TeV pp collision data with two top quarks and missing transverse momentum in the final state. Eur. Phys. J. C* **83,** 503. arXiv: 2211.05426 [hep-ex] (2022).

[149] Sjöstrand, Torbjörn and van Zijl, Maria. *A Multiple Interaction Model for the Event Structure in Hadron Collisions. Phys. Rev. D* **36,** 2019 (1987).

[150] B. Webber. *A QCD model for jet fragmentation including soft gluon interference. Nucl. Phys. B* **238,** 492–528 (1984).

[151] J.-C. Winter, F. Krauss & G. Soff. *A modified cluster-hadronisation model. Eur. Phys. J. C* **36,** 381–395. arXiv: 0311085 [hep-ph] (Aug. 2004).

[152] G. S. Chahal & F. Krauss. *Cluster Hadronisation in Sherpa. SciPost Physics* **13,** 019. arXiv: 2203.11385 [hep-ph] (Aug. 2022).

[153] ATLAS Collaboration. *Example ATLAS tunes of* Pythia8, Pythia6 *and* Powheg *to an observable sensitive to Z boson transverse momentum.* ATL-PHYS-PUB-2013-017. 2013.

[154] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model.* ATL-PHYS-PUB-2016-017. Aug. 2016.

[155] C. Bierlich *et al. Robust Independent Validation of Experiment and Theory: Rivet version 3. SciPost Physics* **8,** 026. arXiv: 1912.05451 [hep-ph] (Feb. 2020).

[156] ATLAS Collaboration. *Measurement of underlying event characteristics using charged particles in pp collisions at* $\sqrt{s} = 900\,GeV$ *and* $7\,TeV$ *with the ATLAS detector. Phys. Rev. D* **83,** 112001. arXiv: 1012.0791 [hep-ex] (2011).

[157] ATLAS Collaboration. *Underlying event characteristics and their dependence on jet size of charged-particle jet events in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector. Phys. Rev. D* **86,** 072004. arXiv: 1208.0563 [hep-ex] (2012).

[158] ATLAS Collaboration. *Measurements of the pseudorapidity dependence of the total transverse energy in proton–proton collisions at* $\sqrt{s} = 7$ *TeV with ATLAS. JHEP* **11,** 033. arXiv: 1208.6256 [hep-ex] (2012).

[159] ATLAS Collaboration. *Measurement of the underlying event in jet events from 7 TeV proton–proton collisions with the ATLAS detector. Eur. Phys. J. C* **74,** 2965. arXiv: 1406.0392 [hep-ex] (2014).

[160] ATLAS Collaboration. *Measurements of underlying-event properties using neutral and charged particles in pp collisions at* $\sqrt{s} = 900$ *GeV and* $\sqrt{s} =$ *7 TeV with the ATLAS detector at the LHC. Eur. Phys. J. C* **71,** 1636. arXiv: 1103.1816 [hep-ex] (2011).

[161] T. Affolder *et al. Charged jet evolution and the underlying event in proton-antiproton collisions at 1.8 TeV. Phys. Rev. D* **65,** 092002 (9 Apr. 2002).