A high-dimensional unbinned measurement of 24 kinematic observables using a machine learning-based analysis of data recorded by the ATLAS detector

by

Laura Miller

A thesis submitted to the

Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy in Physics

Department of Physics

Carleton University

Ottawa-Carleton Institute of Physics

Ottawa, Canada

Abstract

In this thesis, a precision measurement of high momentum $Z \to \mu\mu$ events in proton-proton collision data recorded during 2015–2018 by the ATLAS detector is performed. This measurement is made using a new method, MultiFold, that leverages machine learning to produce a result that is not only corrected for detector effects, but is also high-dimensional and unbinned. This marks the first measurement of its kind based on LHC data and includes full uncertainty covariance. A result in this form greatly increases its utility: it allows the user to construct distributions with any reasonable binning, craft multi-dimensional distributions, and construct new observables that are a function of those included in the result. The measurement is presented in the form of a large event dataset with 24 observables and a variety of weights corresponding to the nominal result and the associated uncertainties. It will be publicly available with an associated user guide to provide examples and give instructions for appropriate use.

The $pp \to Z$+jets cross section is measured as a function of 16 observables that probe the kinematics of the Z boson, the produced muons, and the two highest momentum charged particle jets. A further eight observables probe the internal structure of the two jets. In order to demonstrate the flexibility of the results obtained with the MultiFold method, cross sections are also presented as a function of three additional observables that are calculated from the resulting event sample.

A series of validation tests are performed in order to ensure the measurement is accurate. These tests make use of a set of mock data with known differential cross section distributions, and examine subsets of the dataset and the closure of multidimensional distributions to ensure that agreement within uncertainties is maintained between the measured result and the true values. Based on these tests, a set of recommendations will be provided in the user guide regarding the usage of the results.

Acknowledgments

There have been so many people over the years that have directly and indirectly made this thesis a reality. First, to my supervisor, Dag. You have my deepest appreciation for your guidance and support (and also your patience!). I have learned more about physics over the past five years than I ever thought I would. I couldn't have asked for a better supervisor.

My thanks to everyone I have had the pleasure of working with over the past few years. In particular to my fellow analysis team members: Ben, Dag, Mariel and Adi, thank you for all your hard work and for making this an enjoyable process. Thank you to everyone in the Carleton physics department, and in particular thank you to everyone from the Carleton ATLAS office over the years: Stephen, Alex L., Ben, Alex B., Dylan, Ian, Bryce, Zeke, Callan and Matt you've all been amazing coworkers and drinking buddies. A shoutout also to everyone in the JetEtMiss group, you've all been wonderfully supportive and really helped me find my feet in ATLAS.

To my friends: Liam, Matt C., Matt S., Brody and Ian. While I certainly cannot say you've kept me sane, you have provided some much needed levity and invaluable support throughout this endeavour. Thank you, truly. I would also like to extend my deepest thanks to my family for continuing to support me, even if I'm sometimes a bit terrible at answering your emails.

Finally, I want to thank my Mom and Dad. There are no words to express how grateful I am for the constant love and support you've given me over the years. Without you there's no way I would be here. Thank you, and I love you.

Statement of Originality

The ATLAS Collaboration consists of thousands of scientists, engineers, and technicians from all over the world and as such the data used in this thesis is the combined work of many. The operation of the ATLAS detector and the LHC, the processing and calibration of the data, and the production of simulated samples is all handled centrally. I have contributed to low-level basic reconstruction software efforts in the past. While that work is not detailed in this thesis, it involved implementing software links between reconstructed objects as part of an ongoing effort to improve the event reconstruction within the ATLAS experiment using a new technique known as Global Particle Flow.

The figures in this thesis are created by me unless cited otherwise, and all text is my own. I was one of the main contributors to this analysis and so my work began with designing and implementing the event selection, applying the calibration and all data-driven corrections to the simulation, processing the data and MC samples, and the evaluation of all systematics at the event level. The additional post-processing required for the samples to be used in neural networks and the pseudodata production was a joint effort between my supervisor, Dag Gillberg, and I. I was also responsible for the implementation of the iterative Bayesian unfolding procedure and the corresponding results.

The implementation of the MultiFold algorithm was developed by Ben Nachman, undergraduate student Adi Suresh and postdoctoral researcher Mariel Pettee (all affiliated with the Lawrence Berkeley National Laboratory). While my initial involvement in the MultiFold part of the analysis was low, it grew with time as the analysis progressed. Initially, I helped evaluate the output produced by the Berkeley team and helped to identify issues. In the past year, my involvement in the neural

network training has increased substantially, in particular in regards to the systematic uncertainty propagation. The results for the MultiFold method were produced by the LBNL team, but the tests of the method are performed by me.

At the time of writing this thesis the result is undergoing internal review within the ATLAS Collaboration. A pre-print of the paper will be released shortly, although the final journal is still being decided upon. Along with the publication, we will publish the data as $\mathcal{O}$(GB)-sized event data samples that constitute the measurement. These will likely be hosted on the cloud storage site Zenodo. Along with the samples, a Jupyter notebook will also be published to demonstrate the usage of the results. This will include the production of the 24 differential cross sections presented in this thesis, including the full uncertainty covariance.

Contents

List of Tables

List of Figures

Conventions and acronyms

These are the conventions and acronyms used in this thesis. The acronyms are listed in the order in which they appear.

Conventions

Due to the magnitudes that are dealt with in particle physics measurements, the conventional choice of unit for energy is the electron volt (eV) as opposed to the SI unit Joules (J). Furthermore, this thesis makes use of natural units throughout. This constitutes the convention of setting the speed of light and the reduced Planck constant to one, $c = \hbar = 1$. This leads to the momentum (eV/c), energy (eV) and mass (eV/c^2) all being in units of energy (eV). It is also conventional to express the electric charge in units of the elementary charge.

This thesis makes use of data collected from proton-proton collisions in a circular collider. Thus, the collision energy is expressed as the centre of mass energy (E_{CM}) and is frequently written in terms of the Mandelstam variable s,

$$s = (p_1 + p_2)^2$$

where p_1 and p_2 are the four vectors of the two incoming particles, thus in a circular collider in the high energy limit $\sqrt{s} = E_{\mathrm{CM}}$.

Acronyms

SM - Standard Model

QED - Quantum electrodynamics

QCD - Quantum chromodynamics

EW - Electroweak

MC - Monte Carlo

COM - Centre of mass

ME - Matrix element

LO - Leading order

NLO - Next-to-leading order

PDF - Parton distribution function

IR - Infrared

UV - Ultraviolet

pQCD - Perturbative QCD

UE - Underlying event

MPI - Multiparton interactions

LHC - Large Hadron Collider

CERN - European Organization for Nuclear Research

PS - Proton Synchrotron

SPS - Super Proton Synchrotron

ATLAS - A Toroidal LHC ApparatuS

CMS - Compact Muon Solenoid

ALICE - A Large Ion Collider Experiment

LHCb - Large Hadron Collider beauty

IP - Interaction point

ID - Inner detector

IBL - Insertable B-layer

SCT - Silicon microstrip tracker

TRT - Transition radiation tracker

LAr - Liquid argon

ECAL - Electromagnetic calorimeter

HCAL - Hadronic calorimeter

FCal - Forward calorimeter

EM - Electromagnetic

EMEC - Electromagnetic end-cap calorimeters

HEC - Hadronic end-cap calorimeters

MS - Muon spectrometer

MDT - Monitored drift tubes

CSC - Cathode strip chambers

RPC - Resistive plate chambers

TGC - Thin gap chambers

L1 - Level 1 (trigger)

RoI - Region of interest

HLT - High-level trigger

NN - Neural network

CB - Combined (muons)

IO - Inside-out combined (muons)

ME - Muon spectrometer extrapolated (muons)

ST - Segment-tagged (muons)

CT - Calorimeter-tagged (muons)

IBU - Iterative Bayesian unfolding

pdf - Probability density function

ML - Machine learning

DNN - Deep neural network

GRL - Good Runs List

TTVA - Track-to-vertex association

RMS - Root mean square

Chapter 1

Introduction

The Standard Model of particle physics has proven to be a worthy companion since its full realization in the 1970s. It gives remarkably accurate predictions of many natural phenomena, and is extremely well scrutinized and well studied. Despite its successes, it is known that the Standard Model is not a perfect theory. It lacks an explanation for gravity, and does not explain the neutrino mass or the presence of dark matter and dark energy. These open questions have led not only to many precision measurements being made in an effort to observe deviations from Standard Model predictions, but also many searches for new phenomena beyond the Standard Model. With the advent of the Large Hadron Collider, physicists have been able to probe the Standard Model up to higher and higher energies. Notably this led to the discovery of the final unmeasured particle predicted by the Standard Model, the Higgs boson, in 2012 [1,2] by the ATLAS [3] and CMS [4] Collaborations. However, as of the writing of this thesis, there has been no proof for any specific beyond the Standard Model theory.

Unlike in the days of its discovery in the UA2 experiment [5] where Z bosons were a rare occurrence, they are abundant at the Large Hadron Collider. This has not

only allowed for precision measurements of the Z boson properties, but also for the measurement of their production in association with high energy quarks and gluons. These high energy quarks and gluons produce collimated showers of particles within a particle detector and are known as a jet. Examining the properties of these jets is an important test of perturbative QCD, and this process as a whole marks an important test of both theoretical predictions and the accuracy of simulation software. Due to the aforementioned abundance of Z+jets events in proton-proton collisions at the Large Hadron Collider, producing accurate simulations of this process is vital as it is a major background for many beyond the Standard Model searches [6] and precision Higgs boson measurements [7]. With the energy accessible at the Large Hadron Collider, Z+jets processes can also be used to probe the characteristics of the triple gauge couplings via higher order electroweak processes [8] and in the case where the produced jet is the result of a b quark, be used to study heavy-flavour production and its modelling [9].

Although the detectors located at the Large Hadron Collider are remarkable machines, the measurements they make suffer from various limitations and blurring effects due to the detector. To probe the underlying process the measurement needs to be corrected for such detector effects, which is done through a technique known as *unfolding*. Traditional unfolding methods face a variety of drawbacks. For one, the measured data must be binned into histograms, resulting in potential feature loss, and they cannot be rebinned after the unfolding is complete. Secondly, typical methods are limited to unfolding only one or low dimensional distributions. Lastly, the unfolding procedure only takes into account information from the included features and thus changes in detector response due to auxiliary features may not be correctly accounted for.

Various new unfolding methods have been proposed to address these challenges, with a subset aiming to do so by incorporating elements of machine learning into the unfolding procedure. One such novel method is the recently developed MultiFold method (and its parent method OmniFold [10]), which leverages neural networks to perform a measurement that is highly dimensional and unbinned. In this thesis, it is used to perform a measurement of this kind using ATLAS data for the first time. Unlike other proposed methods [11–13], OmniFold employs a strategy of using neural networks to iteratively reweight the existing Monte Carlo sample. Thus, the neural networks need only learn small incremental corrections as opposed to generative models which require a full estimate of the underlying probability density of the data. Access to a measurement of this nature would be extremely beneficial, for the reasons outlined above, but also from a practical perspective. Due to the ability to rebin the results, examine them in multiple dimensions, and derive new observables from them, the longevity and convenience of results of this kind cannot be understated.

This thesis presents a measurement of high-p_{T} Z+jets events based on the full LHC Run 2 proton-proton dataset collected by the ATLAS detector. The data is recorded with centre of mass energy equal to 13 TeV and has a total integrated luminosity of $139.0\,\mathrm{fb}^{-1}$. Only events where the Z boson decays to a muon and anti-muon pair, $\mu^+\mu^-$, with $p_{\mathrm{T}}^{\mu\mu} > 190\,\mathrm{GeV}$ are considered. The data is unfolded using the MultiFold method with 24 observables as input, thus the measurement will be 24 dimensional and unbinned. The 24 observables are presented differentially, and are compared with a traditional unfolding method, iterative Bayesian unfolding, as well as theoretical predictions. Some additional results unique to the MultiFold method are also explored, including differential measurements of observables derived from the unfolded result.

The theoretical background of the Standard Model, proton-proton collision modelling and Z+jets production are first presented in Chapter 2, before introducing the Large Hadron Collider and the ATLAS Experiment in Chapter 3. Chapter 4 introduces the samples used in this analysis and Chapter 5 discusses the methods used to reconstruct objects in the ATLAS detector. Chapter 6 then discusses both the methods used to make the measurement: the traditional method (iterative Bayesian unfolding) and the MultiFold method. Chapter 7 then outlines the definition of the fiducial volume for this analysis, as well as discussing the modifications made to the samples used as input to the MultiFold algorithm and the implementation of both unfolding algorithms. Finally, Chapter 8 describes the results, first for mock data and then for actual data before concluding remarks are made in Chapter 9.

Chapter 2

Theoretical background

This chapter begins with a short review of the Standard Model of particle physics in Section 2.1. This is followed by a description of the modelling of proton-proton collisions in Section 2.2. Finally, Section 2.3 describes the production mechanisms of Z+jets final states in a proton-proton collider, which are the subject of this thesis.

Unless otherwise specified, Section 2.1 draws on Refs. [14–20]. More information regarding the Lagrangian formalism and the development of quantum field theories are also covered extensively in these references.

2.1 The Standard Model

Introduced in its totality in the 1970s, the Standard Model of particle physics (SM) describes the fundamental particles and how they interact via three of the four fundamental forces of nature: the strong force, the weak force and the electromagnetic force. The fourth fundamental force, gravity, is not included within the SM framework. For our purposes, this force can be safely ignored given that any gravitational effects will be incredibly weak within the realm of particle physics. The SM has

been rigorously studied from both a theoretical and experimental perspective, and it explains nearly all natural phenomena studied to date[1].

The particle content of the SM is summarized in Figure 2.1. The fundamental spin-1/2 fermions that compose matter interact via the exchange of force-carrying spin-1 gauge bosons. The SM is a quantum field theory, and thus these particles are represented by individual fields. Furthermore, the SM is governed by the gauged symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \ . \tag{2.1}$$

This gauge group describes symmetry transformations under which the SM Lagrangian must be invariant, and is used in combination with the individual particle fields to determine the form of that Lagrangian. By Noether's theorem, this leads to several conserved quantities within the SM.

The $SU(3)_C$ group describes the strong interaction, where C represents the colour charge. The colour charge can take on one of three possible quantities: red, blue, or green, and it must be conserved in any interaction involving the strong force. The $SU(2)_L \otimes U(1)_Y$ group describes the electroweak interactions. The conserved quantity in this case is the weak hypercharge, $Y = 2(Q - I_3)$. The electric charge, Q, is the conserved quantity of the $U(1)_{\text{EM}}$ group, and I_3 is the third component of the weak isospin, which is the conserved quantity associated with the $SU(2)_L$ group. The L refers to the fact that only fermions with left-handed chirality may participate

[1]Notable phenomena not currently explained within the framework of the SM are: a theory of gravity, the small observed mass of neutrinos, the asymmetry of matter and antimatter in the universe, and the existence of dark matter and dark energy. Refs. [17] and [19] in particular contain discussions of these phenomena. Efforts to provide explanations for these observations make up a very active field of study known as beyond the Standard Model (BSM) physics. It has led to numerous theories, such as supersymmetry [21], but no experimental confirmation for any specific theory has been observed thus far.

in weak interactions. Similarly, only fermions with electric charge may participate in electromagnetic interactions and only those with colour charge may participate in strong interactions.

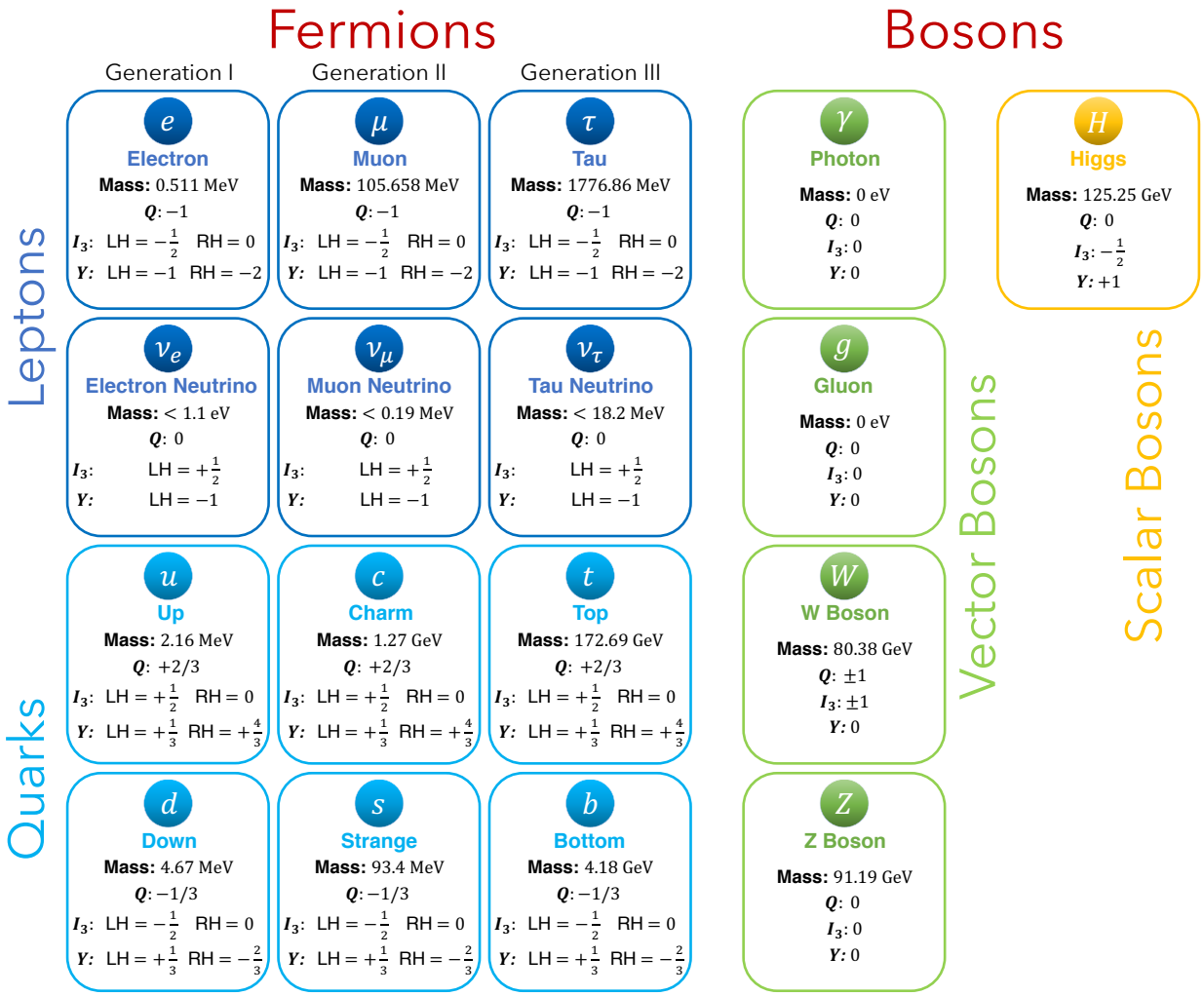


Figure 2.1: A summary of the particle content of the Standard Model. Numerical quantities are taken from Ref. [22]. Q represents the electric charge of each particle in units of elementary charge, I_3 the third component of the weak isospin, and Y is the weak hypercharge.

2.1.1 Particle content of the Standard Model

2.1.1.1 Fermions

There exist two broad categories of fermions within the SM: leptons and quarks. There are three generations of leptons: the first generation contains the electron (e) and the electron neutrino (ν_e), the second the muon (μ) and the muon neutrino (ν_μ), and the third the tau (τ) and the tau neutrino (ν_τ). The electron, muon and tau all possess electric charge $Q = -1$ and thus interact via the electromagnetic force, while the neutrinos do not as they have charge $Q = 0$. Leptons do not possess a colour charge and thus do not participate in the strong interaction. Charged leptons can be divided into left-handed and right-handed chirality, while neutrinos may only be left-handed. The left-handed charged leptons and corresponding neutrinos have a weak isospin equal to $\pm 1/2$ and form weak isospin doublets,

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L , \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L , \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L , \tag{2.2}$$

while right-handed charged leptons form weak isospin singlets, ℓ_R, and have a weak isospin equal to 0. Thus, only the left-handed leptons may participate in weak interactions.

It is also important to note that while neutrinos are determined to be massless within the framework of the SM, they are known to possess a small mass [22]. This was confirmed in 2002 when evidence of neutrino oscillations was discovered by the SNO [23,24] and Super-Kamiokande [25] experiments. The origin of massive neutrinos represents an ongoing area of study within particle physics.

Similar to leptons, there are also three generations of quarks: the first generation

contains the up (u) and down (d) quark, the second the charm (c) and strange (s) quark and the third the top (t) and bottom quark (b). All quarks possess electric charge ($Q = +\frac{2}{3}$ for the u, c and t and $Q = -\frac{1}{3}$ for the d, s and b) and so interact via the electromagnetic force. Furthermore, unlike the leptons, quarks do possess a colour charge, and so participate in strong interactions. The two quarks in each generation form a weak isospin doublet, for example

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \tag{2.3}$$

and right handed singlets, u_R and d_R, where the first generation has been used as an example. Thus, the quarks also participate in weak interactions.

Quarks only exist in colourless bound states and so cannot be directly observed experimentally. As an example, the proton is composed of a bound state of two u quarks and one d quark.

2.1.1.2 Gauge bosons

The gauge bosons serve to mediate interactions between the fermions. They are spin-1 particles that are associated with the generators of the gauge theory defined in Eqn. 2.1.

The $SU(3)_C$ group has eight generators, resulting in eight massless gauge bosons G^a_μ ($a = 1, ..., 8$). They are known as gluons and are the mediators of the strong interaction. A further three gauge bosons, W^a_μ ($a = 1, 2, 3$), are associated with the $SU(2)_L$ group generators and a final gauge boson, B_μ, is associated with the generator for the $U(1)_Y$ group. These final four gauge bosons are mediators of the electroweak interactions and are related to the physical $W^\pm$, Z^0 and the photon γ. Of these, only

the photon is massless while the other three gauge bosons possess a mass generated by the spontaneously broken symmetry of the $SU(2)_L \otimes U(1)_Y$ gauge group. This is outlined in more detail in Section 2.1.4.

2.1.1.3 Higgs boson

The Higgs boson is a massive spin-0 particle that is a consequence of the aforementioned spontaneous electroweak symmetry breaking and is necessary to explain the masses of particles within the Standard Model.

2.1.2 Quantum electrodynamics

In the modern context, the particles and interactions within the Standard Model are described using quantum field theories and use the Lagrangian formalism. This discussion of quantum field theories will begin with a description of quantum electrodynamics (QED), which describes the interactions of charged fermions mediated by a photon.

The Lagrangian for a free fermion ψ with mass m can be written as

$$\mathcal{L}_{\text{free}} = \bar{\psi}(i\not{\partial} - m)\psi \,. \tag{2.4}$$

Here, Feynman slash notation is used as shorthand to represent $\not{\partial} = \gamma^\mu \partial_\mu$, where γ^μ are the gamma (also known as Dirac) matrices. This Lagrangian must be invariant under global phase transformations, $\psi \to \psi e^{-i\alpha}$. Furthermore, it must possess local gauge invariance, and thus must also be invariant under the transformation $\alpha \to \alpha(x)$ for any position x. Combining these conditions, the Lagrangian must be invariant

under the transformation

$$\psi \to \psi e^{-i\alpha(x)}. \tag{2.5}$$

However, if the transformation in Eqn. 2.5 is substituted into Eqn. 2.4, it leads to the following change to the derivative term,

$$\bar{\psi}\partial_\mu\psi \to \bar{\psi}\partial_\mu\psi - i\bar{\psi}\partial_\mu\alpha(x)\psi. \tag{2.6}$$

Thus, the Lagrangian is not invariant. In order to mitigate this, the derivative must be replaced with the covariant derivative,

$$\partial_\mu \to D_\mu = \partial_\mu + ieA_\mu, \tag{2.7}$$

where e is the coupling constant of QED and A_μ is a vector field that transforms under Eqn. 2.5 as

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\alpha. \tag{2.8}$$

Making the replacement described in Eqn. 2.7 results in a Lagrangian that is invariant under local gauge transformation as required. The final missing piece is a kinetic term for the new field A_μ that is similarly invariant under local gauge transformations, this results in

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}(\partial_\nu A_\mu - \partial_\mu A_\nu)(\partial^\nu A^\mu - \partial^\mu A^\nu) = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.9}$$

Thus, the total Lagrangian for the theory of QED is

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi. \tag{2.10}$$

$\mathcal{L}_{\text{QED}}$ gives the kinetic terms for the fermion fields, as well as for the A_μ field (the photon). It also provides the interaction term necessary to construct the QED vertex,

$$\mathcal{L}_{\text{int}} = -e\bar{\psi}\gamma^\mu A_\mu \psi. \tag{2.11}$$

This fundamental vertex is shown in Figure 2.2.

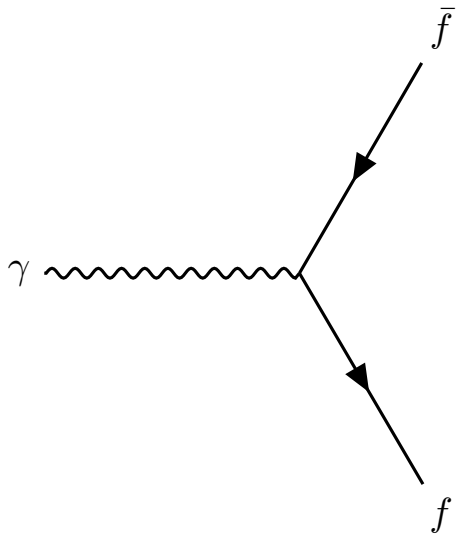


Figure 2.2: The fundamental vertex described by the theory of QED.

2.1.3 Quantum chromodynamics

The theory of quantum chromodynamics (QCD) describes the interactions of quarks and gluons that hold a colour charge. It is a manifestation of the unbroken $SU(3)$ symmetry group. This results in a derivation of the Lagrangian that is very similar,

albeit slightly more complex, to the determination of the QED Lagrangian.

The free-quark Lagrangian may be written analogously to Eqn. 2.4, where the quarks are represented in QCD as a colour triplet,

$$\psi = \begin{pmatrix} \psi_r \\ \psi_b \\ \psi_c \end{pmatrix}. \tag{2.12}$$

Like with QED, the Lagrangian must be invariant under local $SU(3)$ gauge transformations,

$$\psi \rightarrow \psi e^{ig_s \boldsymbol{\alpha}(x)\cdot\hat{\mathbf{T}}}, \tag{2.13}$$

where g_s represents the coupling constant and $\hat{\mathbf{T}}$ represents the generators of the $SU(3)$ group. These eight generators may also be written as

$$T^a = \frac{1}{2}\lambda^a, \tag{2.14}$$

where λ^a represents the Gell-Mann matrices. In a similar procedure to QED, the Lagrangian can be made invariant under Eqn. 2.13 by replacing the derivative with the covariant derivative

$$D_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} G^a_\mu, \tag{2.15}$$

and introducing eight new gauge boson fields G^a_μ that transform as

$$G^a_\mu \rightarrow G^a_\mu - \frac{1}{g_s}\partial_\mu \alpha_a - f_{abc}\alpha^b G^c_\mu. \tag{2.16}$$

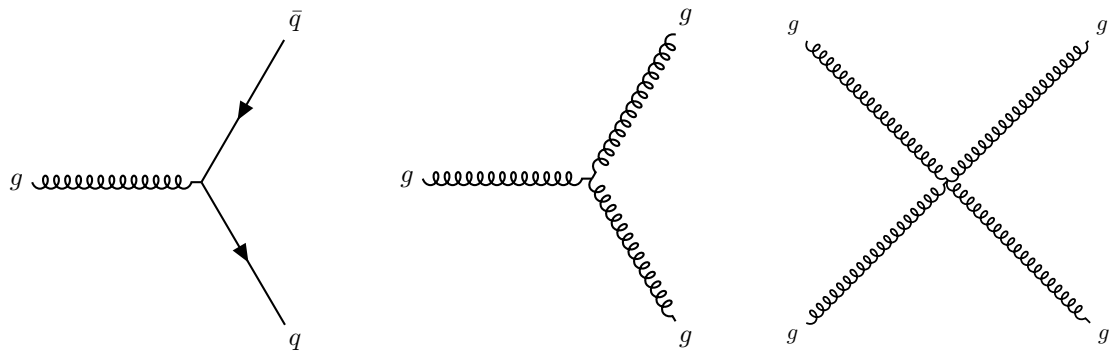


Figure 2.3: The fundamental vertices of QCD.

Here, the structure constants f_{abc} have been introduced, which are related to the Gell-Mann matrices by $[\lambda_b, \lambda_c] = 2if_{abc}\lambda_a$.

The kinetic term for the gluon fields may then be written as,

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}G^{a\mu\nu}G^a_{\mu\nu}, \tag{2.17}$$

where $G^{a\mu\nu}$ is the field strength tensor,

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu. \tag{2.18}$$

The extra term in the field strength tensor gives an additional feature of QCD: gluon self-couplings, which are not present in QED. The form of the resulting QCD Lagrangian is,

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G^{a\mu\nu}G^a_{\mu\nu} + \bar{\psi}(i\not{D} - m)\psi. \tag{2.19}$$

The interactions associated with QCD are summarized in Figure 2.3.

The theory of QCD leads to some interesting physical consequences. The gluon self-couplings result in the theory being asymptotically free, i.e. the coupling constant

increases at low energy or large length scales and may be expressed as

$$\alpha_s(Q^2) = \frac{g_s^2}{4\pi} \simeq \frac{12\pi}{(33 - 2n_f)\ln\left(\frac{Q^2}{\Lambda^2}\right)}, \tag{2.20}$$

where Q^2 represents the energy transfer, Λ is the QCD scale and n_f represents the number of quark flavours. QCD also exhibits what is known as colour confinement, which means that quarks cannot be observed in isolation in nature and only exist in colourless bound states.

2.1.4 Electroweak unification

The concept of electroweak (EW) unification was introduced by Glashow [26], Weinberg [27] and Salam [28]. This theory is necessary to describe the weak interactions, particularly at high energy. There are some properties of the electroweak interactions that contribute to differences between this theory and those of QED and QCD. For one, the weak interaction is mediated by massive bosons, $W^{\pm}$ and Z^0, as opposed to the massless photon and gluons of QED and QCD. Furthermore, for QCD in particular, it may be observed that the coupling constant is consistent for all eight gluons. However in the EW interactions, it can be observed experimentally that the W and Z bosons do not have the same couplings. A unified theory for the electroweak interactions must include these elements, as well as maintaining the expected properties of the photon that were developed in the theory of QED. The basic theory is thus constructed by maintaining invariance under the $SU(2)_L \otimes U(1)_Y$ group transformations, as well as spontaneously breaking this gauge symmetry via the Higgs mechanism.

2.1.4.1 Gauge theory

The first part of the electroweak Lagrangian may be derived by examining the transformations under the $SU(2)_L \otimes U(1)_Y$ symmetry group. This proceeds very similarly to QED and QCD. The weak interactions are associated with the $SU(2)$ group, and thus must be invariant under the transformation,

$$\phi \rightarrow e^{ig\boldsymbol{\alpha}(x)\cdot\frac{\boldsymbol{\sigma}}{2}}\phi, \tag{2.21}$$

where σ^a $(a = 1, 2, 3)$ represent the Pauli matrices and g is the coupling constant.

There is an additional symmetry to consider due to the $U(1)_Y$ group and thus there must also be invariance under the phase transformation,

$$\phi \rightarrow e^{i\frac{g'}{2}Y\alpha(x)}\phi, \tag{2.22}$$

where Y is the hypercharge and g' is the coupling constant.

These combined conditions lead to the requirement that the covariant derivative be

$$D_\mu = \partial_\mu - \frac{1}{2}igW_\mu^a\sigma^a - \frac{1}{2}iYg'B_\mu. \tag{2.23}$$

Four massless gauge bosons have been introduced, W_μ^a $(a = 1, 2, 3)$ for the three associated with the $SU(2)$ group and an additional one associated with the $U(1)$ group, B_μ. This leads to the following kinetic terms for the gauge fields,

$$\mathcal{L} = -\frac{1}{4}W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{2.24}$$

with the field strength tensors defined as

$$W^a_{\mu\nu} = \partial_\nu W^a_\mu - \partial_\mu W^a_\nu - g\epsilon_{abc} W^b_\mu W^c_\nu \tag{2.25}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.26}$$

where ϵ_{abc} is the Levi-Cevita tensor. Finally, the interactions with the fermions may be added to obtain the final expression for this part of the electroweak Lagrangian. Keeping in mind that the weak force interacts differently with left-handed and right-handed fermions, we may write

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi}_L i \not{D} \psi_L + \bar{\psi}_R i \not{D} \psi_R. \tag{2.27}$$

2.1.4.2 Spontaneous electroweak symmetry breaking

The mechanism by which these massless gauge bosons will acquire mass, spontaneous symmetry breaking, can now be discussed. In order to introduce the symmetry breaking, a potential must be added to the Lagrangian that will result in a shift in the vacuum expectation value. To this end, a complex scalar doublet known as the Higgs field is introduced,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{2.28}$$

ϕ is symmetric under $SU(2) \otimes U(1)$. This introduces the following terms to the electroweak Lagrangian

$$\mathcal{L}_{\text{Higgs}} = |D_\mu \phi|^2 + V(\phi^\dagger \phi) - \bar{\psi}_R \mathcal{M} \psi_L - \bar{\psi}_L \mathcal{M}^\dagger \psi_R, \tag{2.29}$$

where the covariant derivative is as defined in Eqn. 2.23. The final two terms represent the interaction between the Higgs boson and the fermions, and provide the mechanism by which the fermions acquire mass. The form of $\mathcal{M}$ depends on the type of fermion.

The form of the potential term $V(\phi^\dagger\phi)$ must be such that the theory is renormalizable, thus it is restricted to having a maximum of quartic terms. Furthermore, it must be rotationally invariant and so we are limited to terms with powers of $\phi^\dagger\phi$. Thus, the general form is

$$V(\phi^\dagger\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \tag{2.30}$$

where in order to preserve stability, λ is chosen to be a positive real constant and μ^2 is positive such that the vacuum expectation value is not invariant.

The vacuum expectation value is then denoted to be $v/\sqrt{2}$, where

$$v = \sqrt{\frac{\mu^2}{\lambda}}\ . \tag{2.31}$$

Furthermore, without loss of generality the Higgs field can be defined in terms of the physical Higgs field, h:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix}. \tag{2.32}$$

By substituting this into Eqn. 2.29, the terms in the electroweak Lagrangian that

involve the Higgs boson can be determined,

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} =& \frac{1}{2}\partial_\mu h \partial^\mu h + \frac{(v+h)^2}{8}\left(g^2(W_\mu^1 + iW_\mu^2)(W^{\mu 1} - iW^{\mu 2}) + (g'B_\mu - gW_\mu^3)^2\right) \\ &+ \lambda v^2 h^2 + \lambda v h^3 + \frac{1}{4}\lambda h^4 - \bar{\psi}_R \mathcal{M} \psi_L - \bar{\psi}_L \mathcal{M}^\dagger \psi_R.\end{aligned} \tag{2.33}$$

The physical states for the $W^\pm$ bosons can then be written as

$$W^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right), \tag{2.34}$$

and in order to diagonalize the masses, the physical states for the Z boson and the photon can be written as:

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \tag{2.35}$$

$$A_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu, \tag{2.36}$$

where θ_W is the weak mixing angle and is defined as $\tan\theta_W = g'/g$. Substituting this back into Eqn. 2.33 gives,

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} =& \frac{1}{2}\partial^\mu h \partial_\mu h + \frac{v^2 g^2}{4} W_\mu^- W^{\mu +} + \frac{v^2 g^2}{8\cos^2\theta_W} Z_\mu Z^\mu + \frac{v g^2}{2} W_\mu^- W^{\mu +} h + \frac{v g^2}{4\cos^2\theta_W} Z_\mu Z^\mu h \\ &+ \frac{g^2}{4} W_\mu^- W^{\mu +} h^2 + \frac{g^2}{8\cos^2\theta_W} Z_\mu Z^\mu h^2 + \lambda v^2 h^2 + \lambda v h^3 + \frac{1}{4}\lambda h^4 \\ &- \bar{\psi}_R \mathcal{M} \psi_L - \bar{\psi}_L \mathcal{M}^\dagger \psi_R.\end{aligned} \tag{2.37}$$

This gives the masses for the W ($m_W = \frac{gv}{2}$), Z ($m_Z = \frac{m_W}{\cos\theta_W}$), and Higgs boson ($m_h = \sqrt{2\lambda}v$), as well as the interaction terms which are summarized in Figure 2.5. The W and Z self-couplings and their couplings to the fermions can be derived by

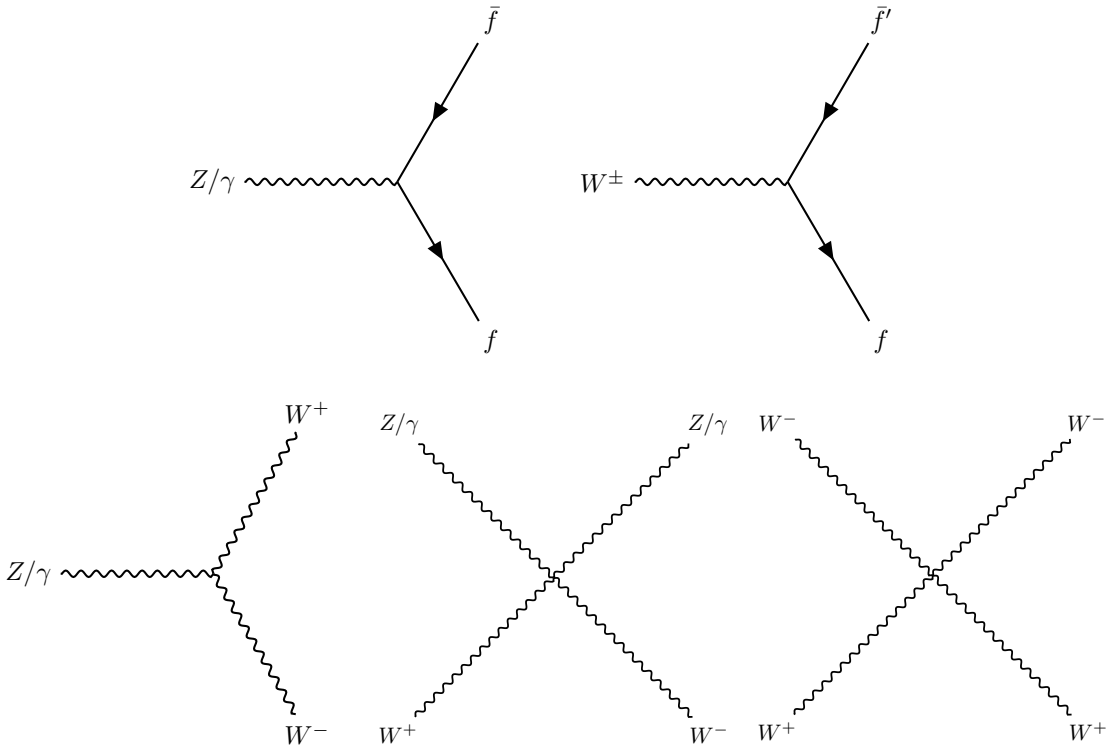


Figure 2.4: The fundamental vertices of the electroweak theory involving electroweak gauge boson self-interactions, as well as their interactions with fermions.

expanding Eqn. 2.24. These interactions are summarized in Figure 2.4.

2.2 Modelling *pp* collisions

Due to the complexity of many modern experimental measurements, it is crucial to have theoretical predictions to compare with. This need for suitable theoretical predictions has led to the development of Monte Carlo (MC) simulation software to model processes.

As this thesis pertains to data produced by colliding protons at a very high centre of mass (COM) energy ($\sqrt{s} = 13$ TeV), this section will focus on describing the simulation of proton–proton collisions. First, the calculation of the matrix element

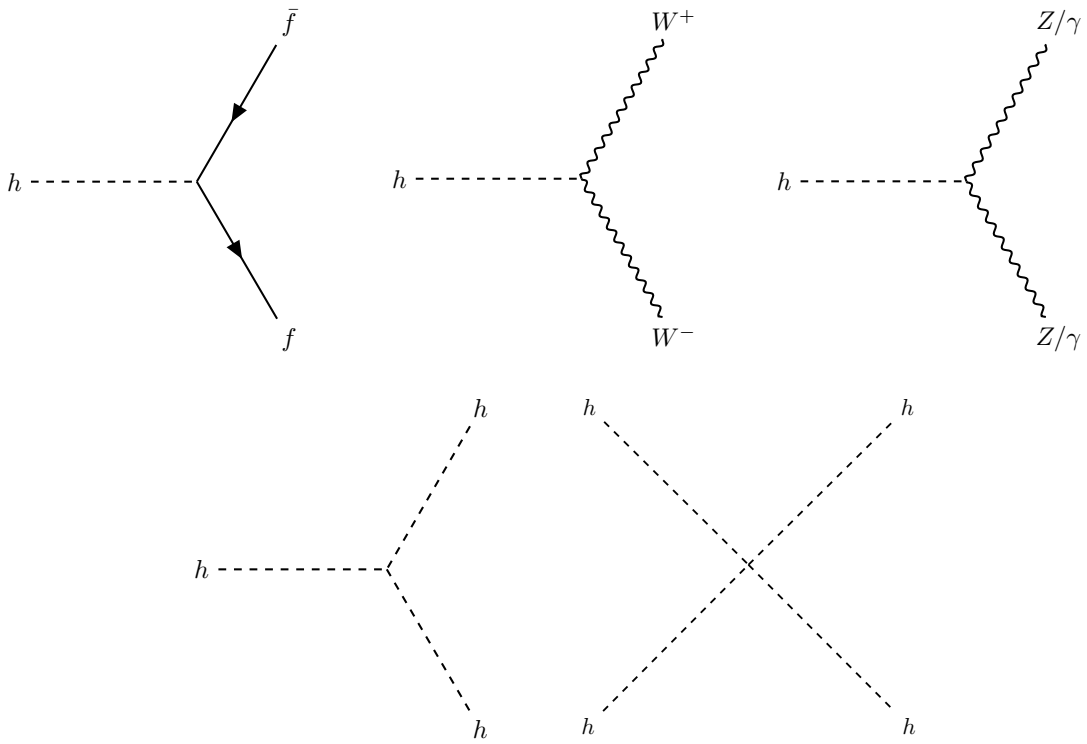


Figure 2.5: Interactions in the electroweak theory involving the Higgs boson.

will be discussed, followed by a description of parton showering algorithms and finally a brief discussion of hadronization.

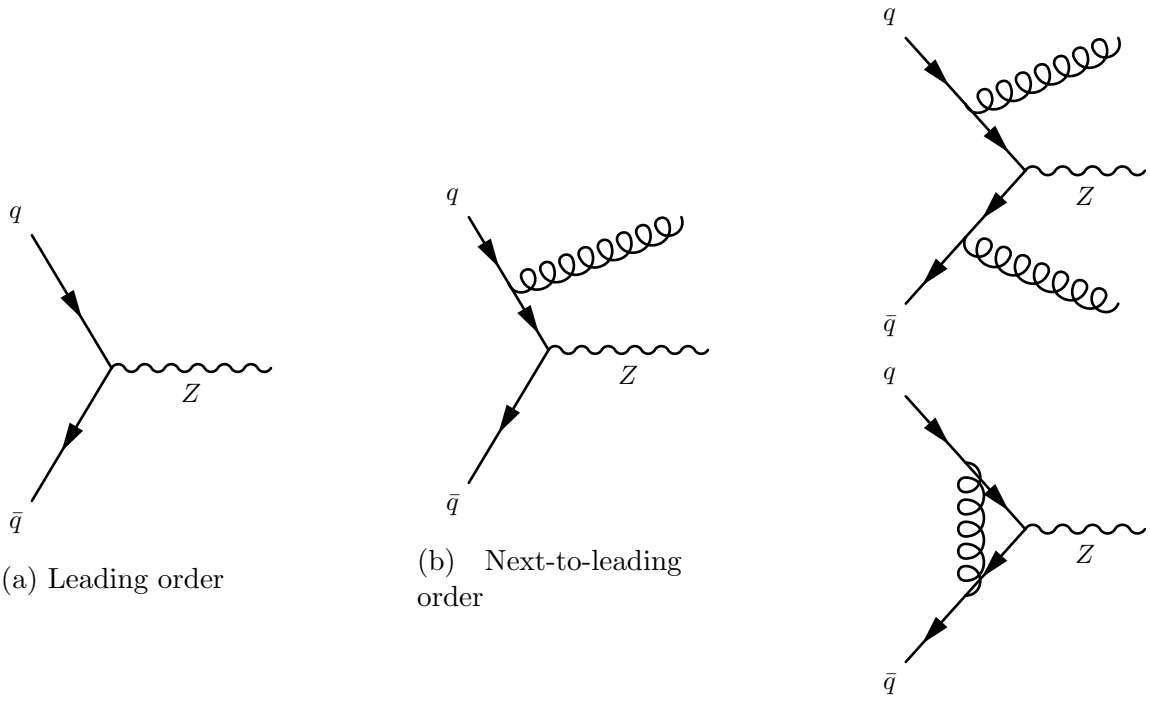
2.2.1 Matrix element and cross section calculations

The first step to the simulation of a particular process is the calculation of the matrix element (ME), $\mathcal{M}$. The ME is calculated by summing contributions from all the possible Feynman diagrams describing a specific process using the principles developed in the previous section. For a specific example, we may consider the annihilation of a quark anti-quark pair to produce a Z boson as outlined in Figure 2.6. This figure also shows diagrams of different orders: leading order (LO) diagrams are defined as the diagrams with the least vertices, next-to-leading order (NLO) are then the diagrams with one additional vertex, and so on. In practice, while calculating diagrams up to a higher order leads to more precision in the ME, it also becomes more computationally expensive. Thus, MC simulations calculate diagrams only up to a fixed order. In this analysis, the majority of the MC generators calculate diagrams up to at least NLO accuracy.

In this thesis, the quantity of interest is the rate at which a specific process occurs. This is generally expressed as a cross section, σ. For example, if the process being considered is two incoming partons (quarks or gluons) a and b with a high COM energy that produce some final state X with an arbitrary number of outgoing particles, then the partonic production cross section for the $ab \to X$ process may be expressed as [19]

$$\hat{\sigma}_{ab\to X} \propto \frac{1}{2\hat{s}} \int d\Phi_n |\mathcal{M}|^2. \tag{2.38}$$

The amplitude $|\mathcal{M}|^2$ is as described above, $\hat{s}$ represents the Mandelstam variable in



(a) Leading order

(b) Next-to-leading order

(c) Next-to-next-to-leading order

Figure 2.6: Representative Feynman diagrams for a quark and an antiquark annihilating to produce a Z boson, (a) shows the leading order process with the least number of vertices, (b) shows NLO with one additional vertex and (c) shows NNLO with two additional vertices. Note that the loop diagram in (c) will also contribute to the leading order diagram via interference.

relation to the incoming partons, and the remaining term containing Φ_n describes the final state phase space available for the process in question. The proportionality is simply an overall factor related to the particle characteristics. This gives the basis for calculating the cross sections of processes in proton-proton collisions. There are, however, other considerations that must be taken into account.

The first additional consideration is the structure of the incoming protons. Equation 2.38 describes the production cross section in the case of two incoming partons, however, the proton is not a fundamental particle and is instead a bound state of two u quarks and one d quark. The quarks that are part of the bound state are

known as the valence quarks. These valence quarks will interact with each other via gluons which themselves produce quark-antiquark pairs, producing what are known as sea quarks. The partons will each carry a fraction of the proton's total momentum, which must be taken into account when making predictions for a final state. The partons within the incoming protons can undergo deep inelastic scattering when the momentum transferred in the proton-proton collision is high enough to cause the proton to break apart. This deep inelastic scattering process can be parameterized by two quantities, Q^2 and x. Q^2 represents the squared momentum transferred in the collision and x represents the fraction of the proton's momentum carried by the parton involved in the collision.

It is possible to construct a probability distribution for the momentum of the parton involved in the collision, known as the parton distribution function (PDF), $f(x, Q^2)$. The PDF serves to give the probability that a parton within the proton has a momentum fraction between x and $x + dx$ for a certain momentum transfer Q^2. The PDFs must be determined experimentally, but they are universal and can therefore be derived from a combination of a wide array of experiments. An example of a PDF set can be seen in Figure 2.7.

Thus, in order to obtain the cross section for the $ab \to X$ process where a and b are part of two colliding protons p_1 and p_2, we must include the PDFs,

$$\sigma_{p_1 p_2 \to X} = \sum_{a,b} \int_0^1 dx_1 f_a(x_1, Q^2) \int_0^1 dx_2 f_b(x_2, Q^2)\, \hat{\sigma}_{ab \to X}, \tag{2.39}$$

where x_1 and x_2 is the momentum fraction of p_1 and p_2 carried by their constituent partons, denoted a and b respectively, and $\hat{\sigma}_{ab \to X}$ is the partonic cross section presented in Eqn. 2.38. Note that $\hat{s}$ in that equation may now be expressed as $x_1 x_2 s$, where s represents the COM energy of the incoming protons.

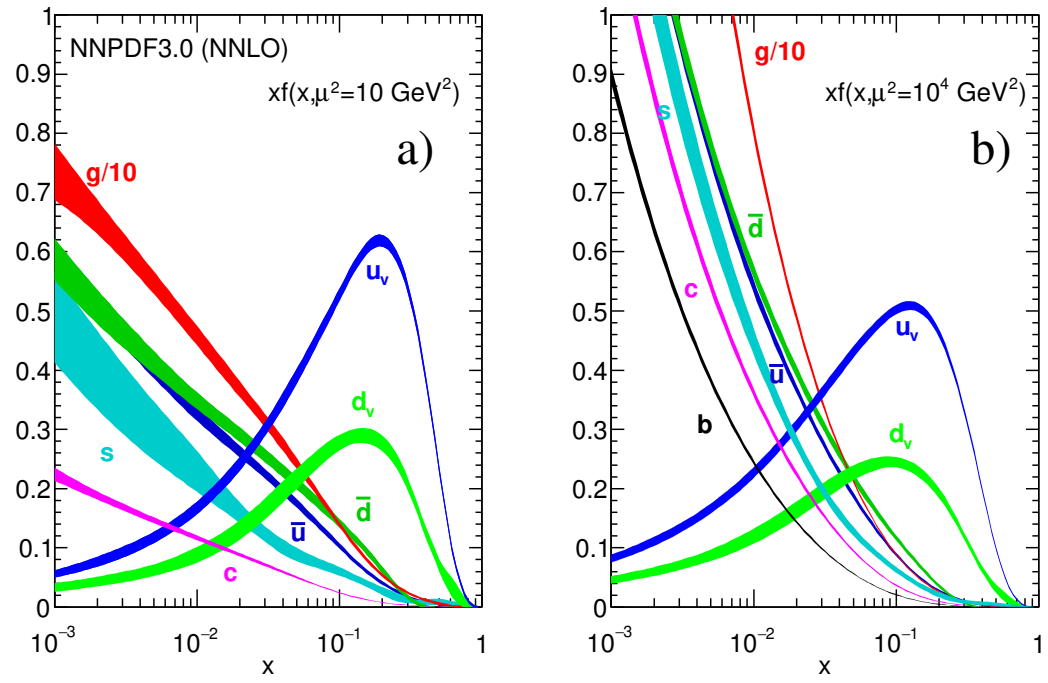


Figure 2.7: PDFs from the NNPDF3.0 PDF set shown with uncertainty bands for (a) $Q^2 = 10$ GeV^2 and (b) $Q^2 = 10^4$ GeV^2 [29].

The effects of higher order diagrams on the cross section measurement must now be considered. There are two main items to highlight here, the first is the existence of infrared (IR) divergences caused by the emission of a soft (low energy) gluon from the initial incoming particles (see Figure 2.6b in the case where the gluon is soft, for example). These divergences are absorbed by the PDFs below a certain scale defined as the factorization scale μ_F. For higher scales, where the gluon emission is harder (higher energy), they are included in the partonic cross section calculation. The second item to highlight is the ultraviolet (UV) divergences caused by loop corrections at high energy (see for example the loop diagram in Figure 2.6c). These are absorbed using the $\overline{\text{MS}}$ scheme and result in the strong coupling constant α_s

becoming a function of the renormalization scale μ_R. This gives [30]:

$$\sigma_{p_1 p_2 \to X} = \sum_{a,b} \int_0^1 dx_1 f_a(x_1, \mu_F) \int_0^1 dx_2 f_b(x_2, \mu_F)\, \hat{\sigma}_{ab\to X}(x_1 x_2 s, \alpha_s(\mu_R), \mu_R, \mu_F). \tag{2.40}$$

For a process such as the one described above at a high COM energy, the calculations remain firmly within the region of perturbative QCD (pQCD), where the coupling constant α_s is much less than 1. This means that the partonic cross section can be freely expressed as,

$$\sigma_{ab\to X} = \sigma^{\mathrm{LO}} + \alpha_s\, \sigma^{\mathrm{NLO}} + \alpha_s^2\, \sigma^{\mathrm{NNLO}} + ... \tag{2.41}$$

2.2.2 Parton showers

The outgoing partons from the ME calculation will continue to radiate additional partons until the individual parton energy is low enough for hadronization to occur (typically around $\mathcal{O}$(GeV)). These subsequent partons are modelled using a parton showering algorithm.

The basic principle of the parton showering algorithm is to take the final state partons from the ME calculation and determine the probability that the system will produce another parton and what energy the resulting parton will have. This is determined using a Sudakov form factor that is a function of the splitting functions and the energy of the system [31].

Parton showering algorithms generally deal well with soft and collinear emissions, but do not provide as adequate a prediction for the case where the emission is harder or at a wide angle. In this case, the emissions are better modelled using ME calculations. This leads to a natural desire to combine these elements in order to

construct the best possible model of the hard scatter event. Several schemes are in existence to accomplish this, two examples are the CKKW-L [32, 33] and FxFx [34] algorithms.

Another consideration when modelling the parton showering in a hard scatter event is the effects from the underlying event (UE). This includes all activity that cannot be directly contributed to the initial hard scatter. As an example, this can include multiparton interactions (MPI), which result from the interactions of additional partons from the protons involved in the hard scatter, or from interactions stemming from the beam remnants. These additional interactions have a tendency to be relatively soft. They will, however, contribute additional energy and influence the colour flow in an event leading to a higher hadron multiplicity.

2.2.3 Hadronization

Once the parton showering algorithm has evolved the partons down to a sufficiently low energy, the hadrons, which are the observable final state, must be modelled. There are two primary models to perform the hadronization, the Lund string model [35, 36] and the cluster model [37]. Both of these models require input from experimental data in order to determine necessary parameters. If the initially produced hadrons are unstable, the decays to stable particles are appropriately simulated by the MC generator.

The hadrons produced by these models mark the measurable state for partons within a particle detector. If the parton has a sufficiently high transverse momentum, it will create a collimated shower of particles known as a jet, which are part of the measurement that is the subject of this thesis.

Figure 2.8 gives a representation of an event within a detector.

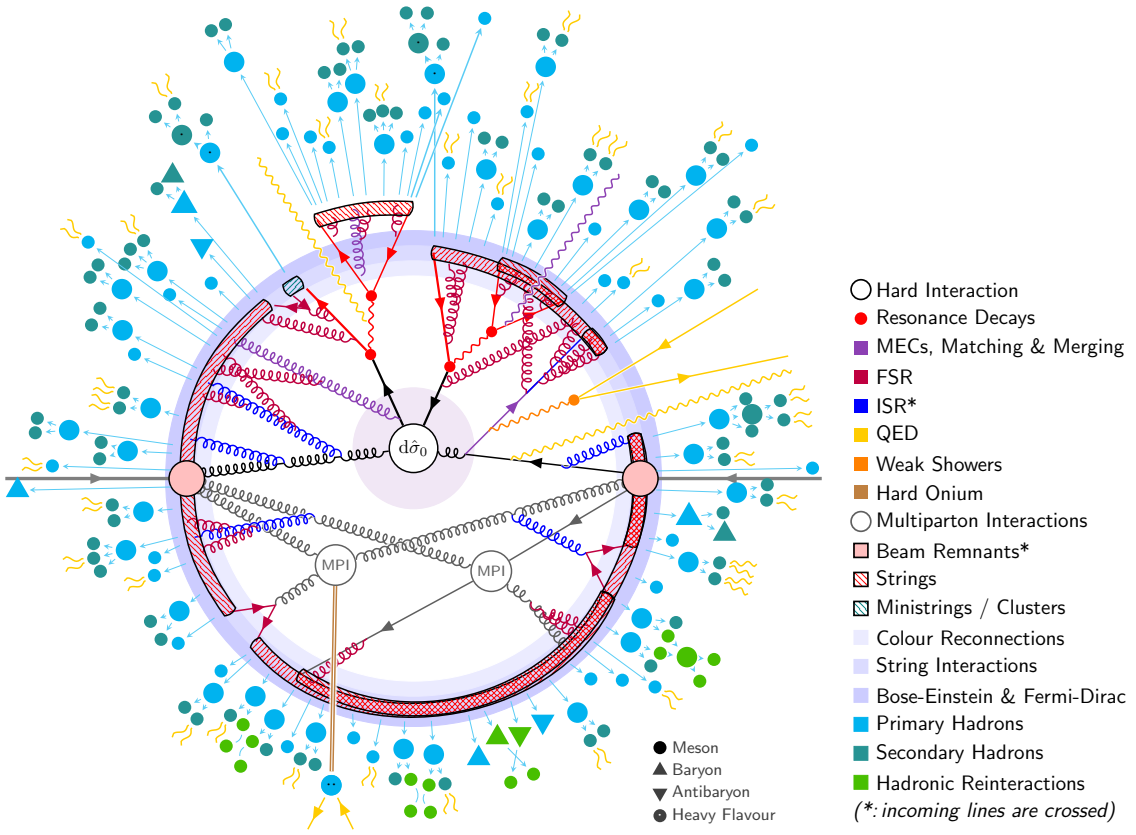


Figure 2.8: Figure taken from Ref. [38]. An illustration of a $pp \to \bar{t}t$ event giving an overview of the main parts of the event as simulated by the PYTHIA MC generator. It shows the results of the hard scatter event, the parton showering and the hadronization with the final state hadrons represented in green and blue.

2.3 Z+jets production in pp colliders

Measurements of Z production via $pp \to Z + X$ are very useful for numerous applications within the particle physics field. It is a relatively low background and frequent process at the LHC and the Z boson is comparatively easy to identify, meaning that it is a process that can be measured with a good deal of precision. This makes it very useful for performing precision measurements of the Z boson properties, and constraining PDFs [39], for example.

When the Z boson is produced in association with a jet, measurements provide

a means of testing perturbative QCD and can be used to test theoretical predictions at higher orders and in extreme regions of phase space. Particularly in the case of producing accurate simulations, this is vital and Z+jets measurements are frequently used for this purpose [40], as well as for fine tuning parton showering algorithms [41]. It is an important process to have a good understanding of in all regions of phase space, as it constitutes an important background to many Higgs measurements [7], as well as beyond the Standard Model searches [6]. It is also used to derive calibration procedures for measured objects [42–44].

With the energy accessible at the LHC, high-p_{T} Z+jets production provides a probe into higher order electroweak processes and can be used to probe the characteristics of the triple gauge couplings [8], as one example. It can also be used to study heavy-flavour production and its modelling [9], in the case where the jet produced is the result of a b quark.

There are several processes that will give a Z+jets final state that are examined in the following sections. The total measured cross section for these processes is also outlined in Figure 2.9 in comparison to other SM processes. The dominant process is examined first, followed by the diboson processes and finally the electroweak processes.

In this analysis, only the case where the Z boson decays to a muon and an anti-muon is considered. The Z boson decays to all fermion anti-fermion pairs within the SM except for top anti-top, which is forbidden by conservation of energy. The branching ratio for $Z \to \mu^+\mu^-$ is 3.36% [22]. For the remainder of this thesis, this will be written as $Z \to \mu\mu$ for simplicity.

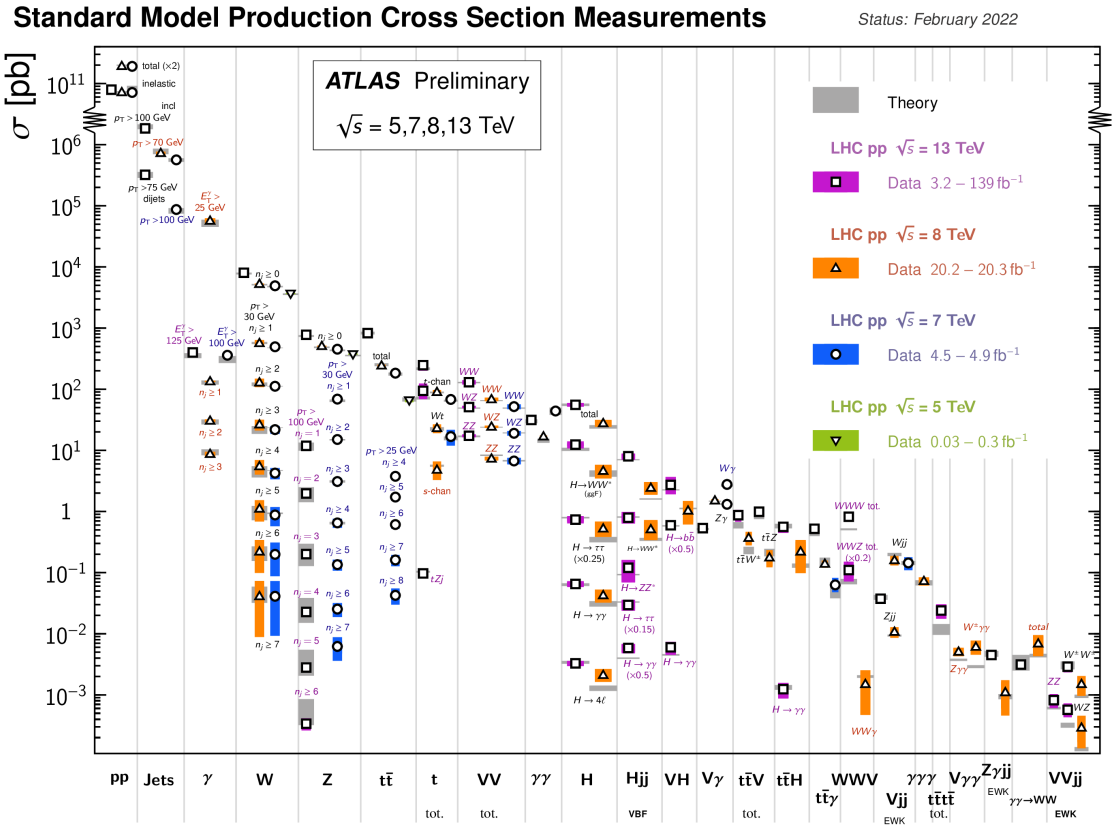


Figure 2.9: A summary of ATLAS cross section measurements for various SM processes [45]. Of particular interest for this analysis: the Z+jets measurements in the fifth column are dominated by the strong Z+jets (Drell-Yan) process described in Section 2.3.1; the ZZ and WZ (see the VV column) represent the diboson processes discussed in Section 2.3.2; and, Zjj EWK (see the Vjj EWK column) represents the electroweak processes in Section 2.3.3. These measurements agree with theoretical predictions, providing good substantiation for the SM.

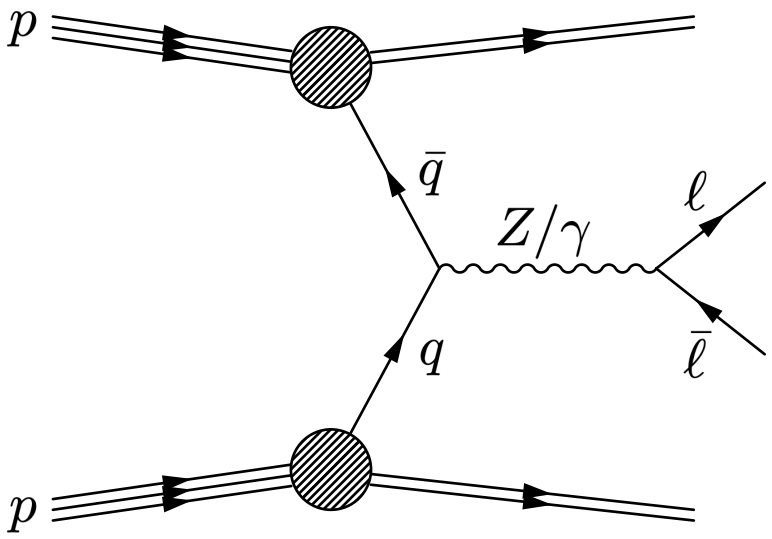


Figure 2.10: The LO Feynman diagram of the Drell-Yan process. In this analysis, only the case where the Z decays to $\mu^+\mu^-$ is considered. Note that at higher orders (NLO and above), additional partons can be produced via initial state radiation.

2.3.1 Strong Z+jets

In this analysis, the main process of interest is known as the Drell-Yan process, where the Z boson is produced by a quark anti-quark annihilation and subsequently decays to a lepton and an anti-lepton. The leading order production mechanism of this process is shown in Figure 2.10.

The number of jets associated with the Drell-Yan process varies. Here, the focus is the case where there is at least one jet produced in association with the Z boson. The simplest LO diagram is shown in Figure 2.11, which can result in one jet being produced. The subsequent Figure 2.12 shows the possible production mechanisms for the case where two jets can be produced, and the production of more jets is possible with higher order diagrams.

As this process is initiated by a strong interaction, it will be referred to as strong Z+jets throughout this thesis.

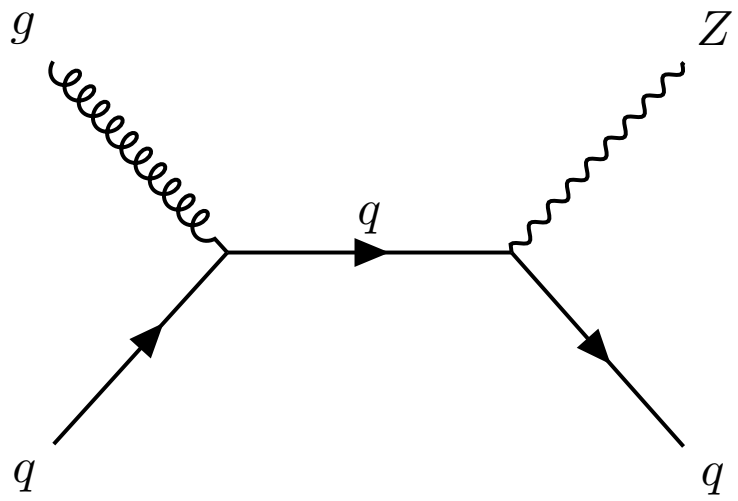


Figure 2.11: Representative Feynman diagram for the strong $Z+1$ jet process. This is one of the NLO diagrams of the Drell-Yan process as seen in Figure 2.10.

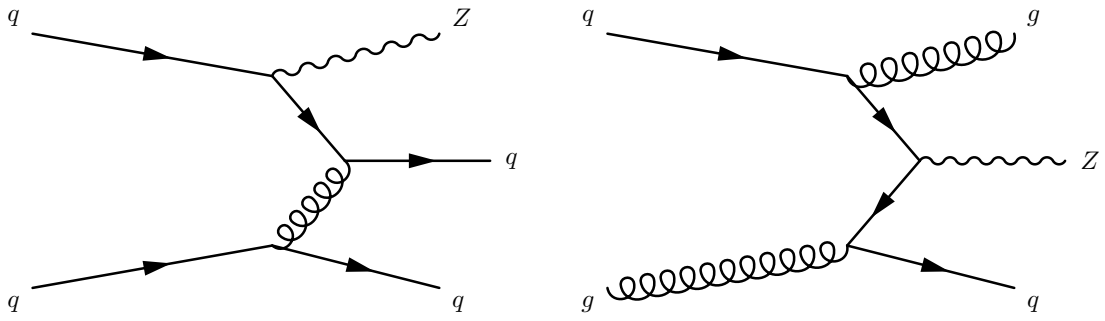


Figure 2.12: Representative Feynman diagrams for the strong $Z + 2$ jets process.

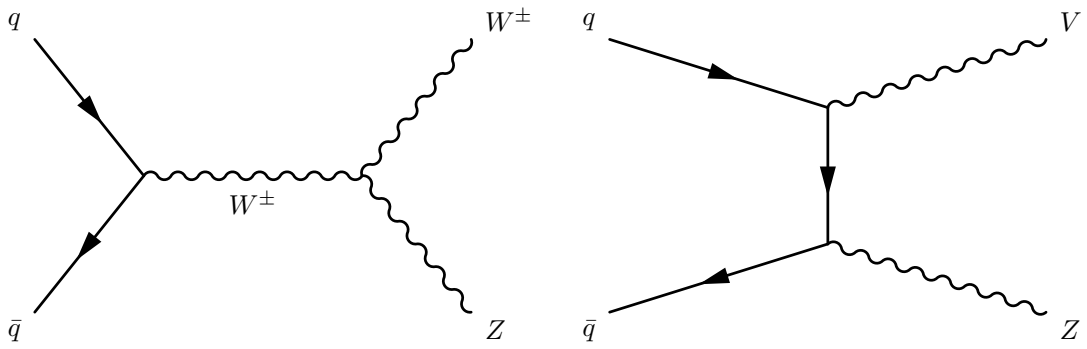


Figure 2.13: Representative Feynman diagrams of diboson processes that produce Z+jets final states. In the case illustrated in the diagram on the left, the W boson decays to pair of quarks. For the diagram on the right, the V boson may be either a W or a Z boson which decays to two quarks.

2.3.2 Diboson Z+jets

A Z+jets final state may also be produced as a result of diboson production. This occurs via the $qq \to ZV$ process, where V denotes either a W or Z boson. The leading order diagrams for diboson Z+jets production are summarized in Figure 2.13. While diboson Z+jets production is overall much less frequent than strong Z+jets production ($\sigma_{\text{strong}}/\sigma_{\text{diboson}} \approx \mathcal{O}(1000)$), the cross section is enhanced in the high p_{T}^{Z} region making it slightly more significant (around 2%, as can be seen in Section 7.5) in the phase space of this analysis.

2.3.3 Electroweak Z+jets

The final way to produce the Z+jets final state is via the exchange of an additional Z or W boson, which will be referred to as EW Z+jets production throughout this thesis. Some examples of this production mechanism at leading order are shown in Figure 2.14. As with diboson Z+jets production, this process occurs much less often than strong Z+jets production. However, like in the diboson case, this pro-

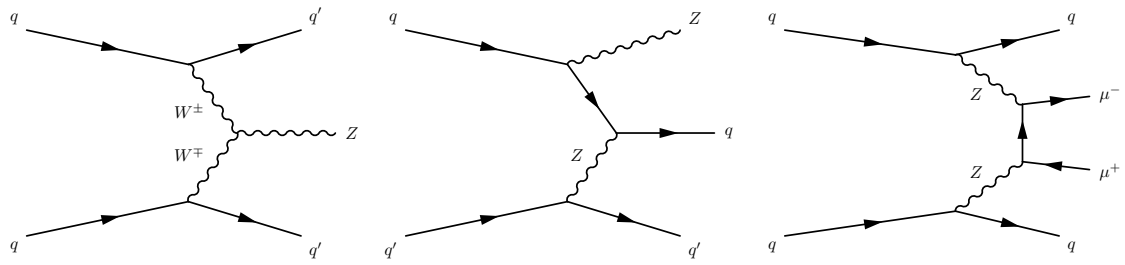


Figure 2.14: Representative leading order diagrams for electroweak Z+jets production. All diagrams of the electroweak Z+jets process have two additional electroweak vertices compared to the strong Z+jets process.

cess becomes more important in the higher p_{T}^{Z} region and so makes a slightly larger contribution in the phase space of this analysis.

Chapter 3

The ATLAS Experiment

This chapter introduces the experimental setup used to collect the data analyzed in this thesis. The ATLAS Collaboration is one of the largest scientific collaborations in the world, with over 3000 collaborators representing institutions from 42 countries [46]. The CERN accelerator complex, described in Section 3.1, is the home of the ATLAS detector, which is described in detail in Section 3.2.

3.1 The Large Hadron Collider and CERN

Located near Geneva, Switzerland, the Large Hadron Collider (LHC) [47] is a part of the largest particle accelerator complex in the world, the European Organization for Nuclear Reseach (CERN). The LHC is a circular accelerator that lies at an average depth of 100 m underground and has a total circumference of 27 km. The full extent of the LHC with respect to its surroundings can be seen in Figure 3.1. Its primary function is to collide bunches of protons at high energy, producing data that is then analyzed in order to examine outstanding questions in particle physics. The LHC also functions secondarily as a heavy ion collider.

The LHC is the largest ring in the CERN accelerator complex, a schematic of which can be seen in Figure 3.2. The protons are initially produced by ionizing hydrogen gas, and undergo four preliminary acceleration steps before they are injected into the main LHC beam line. The protons are first taken to an energy of 50 MeV by Linac2, a linear accelerator [48]. This is followed by an acceleration to 1.4 GeV in the PS Booster before the protons are fed to the Proton Synchrotron (PS), which accelerates them to 25 GeV. Finally, they are accelerated to an energy of 450 GeV in the Super Proton Synchrotron (SPS) before being transferred to the LHC, where they are circulated until they reach the desired energy.

The LHC was designed to attain a COM collision energy of $\sqrt{s} = 14\,\text{TeV}$. Work is ongoing to achieve this, primarily via training of the magnets [49]. The LHC data-taking periods are divided into so-called runs, which have had increasing COM energy since the LHC began taking physics data in 2009. A summary of past and future planned LHC runs can be seen in Figure 3.3. This analysis is concerned with data taken during LHC Run 2, which achieved a COM energy of $\sqrt{s} = 13\,\text{TeV}$.

The protons are injected into the LHC in bunches with around 10^{11} protons per bunch, which collide up to every 25 ns at various interaction points located around the LHC. During Run 2, this translated to a peak instantaneous luminosity of $2.1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, twice the LHC design luminosity of $1.0 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$. Four of the interaction points are outfitted with one of the major experiments located around the LHC ring: ATLAS [3], CMS [4], ALICE [50] and LHCb [51]. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are both general purpose detectors designed to measure the full information from a *pp* collision, while ALICE (A Large Ion Collider Experiment) focuses on the measurement of heavy-ion collisions and LHCb (Large Hadron Collider beauty) is primarily designed to perform

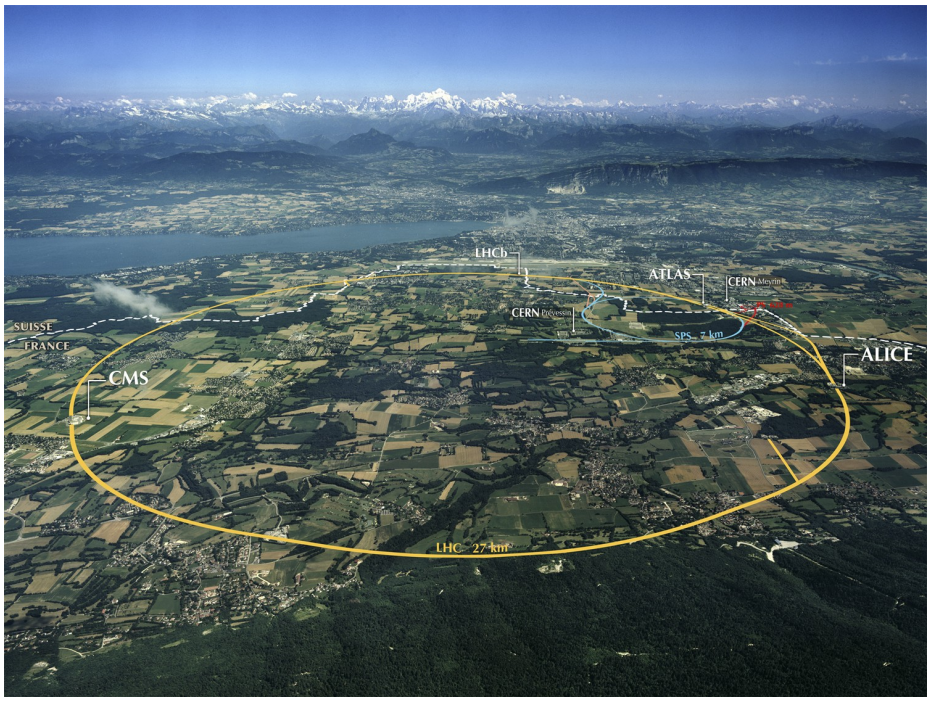


Figure 3.1: An aerial view of the LHC, including the positioning of each of the four major experiments around the LHC ring. The Swiss Alps and Lake Geneva can be seen in the background [52].

measurements related to hadrons containing b quarks.

The protons are kept on the correct trajectory in the beam line using a series of bending magnets. The most important of these are the superconducting dipole magnets, which produce a magnetic field strength of up to 8.3 T with proper training. In order to maintain optimal performance they must be cooled to 1.9 K, which is achieved using a liquid helium cryogenic system. The beam line at the LHC must also be kept at ultrahigh vacuum in order to eliminate collisions with gaseous particles.

3.2 The ATLAS detector

The ATLAS detector is a general purpose detector located at point 1 on the LHC beam line, as seen in Figures 3.1 and 3.2. It is built in a cylindrical configuration

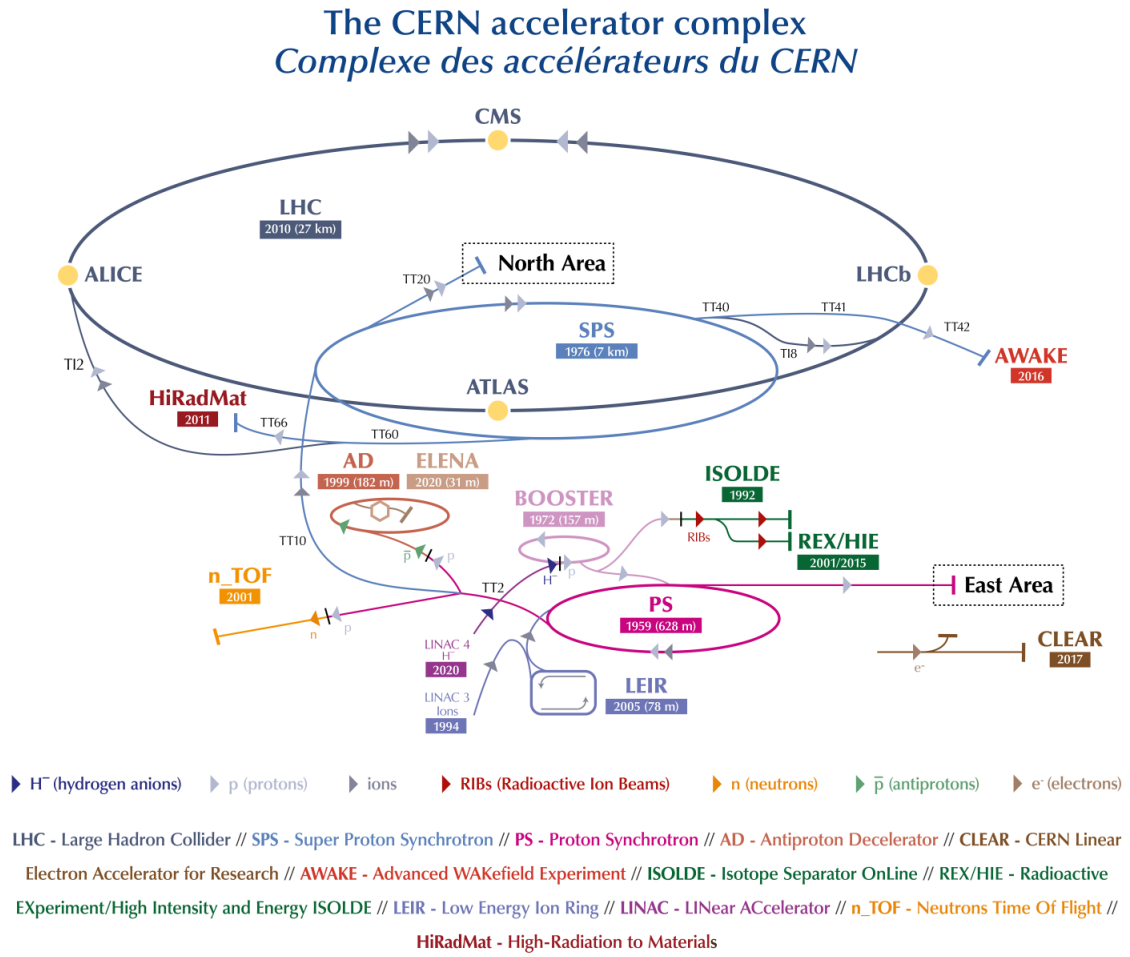


Figure 3.2: A schematic description of the CERN accelerator complex [53]. Protons are accelerated using Linac2 (replaced with Linac4 in 2020, as shown in the figure), followed by the PS booster, the PS and finally the SPS before arriving at the main LHC ring.

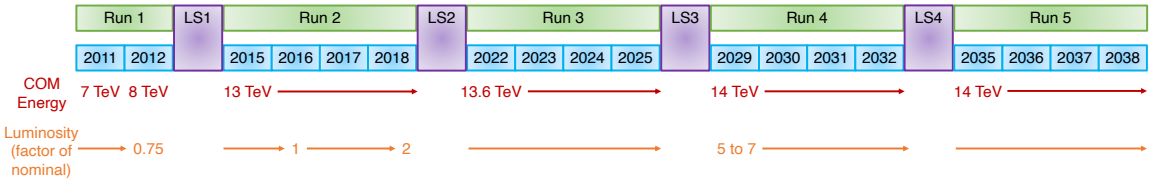


Figure 3.3: A timeline of past and future planned Runs at the LHC. The COM energy and instantaneous luminosity for each Run is also shown for reference. This thesis is concerned with data taken during Run 2.

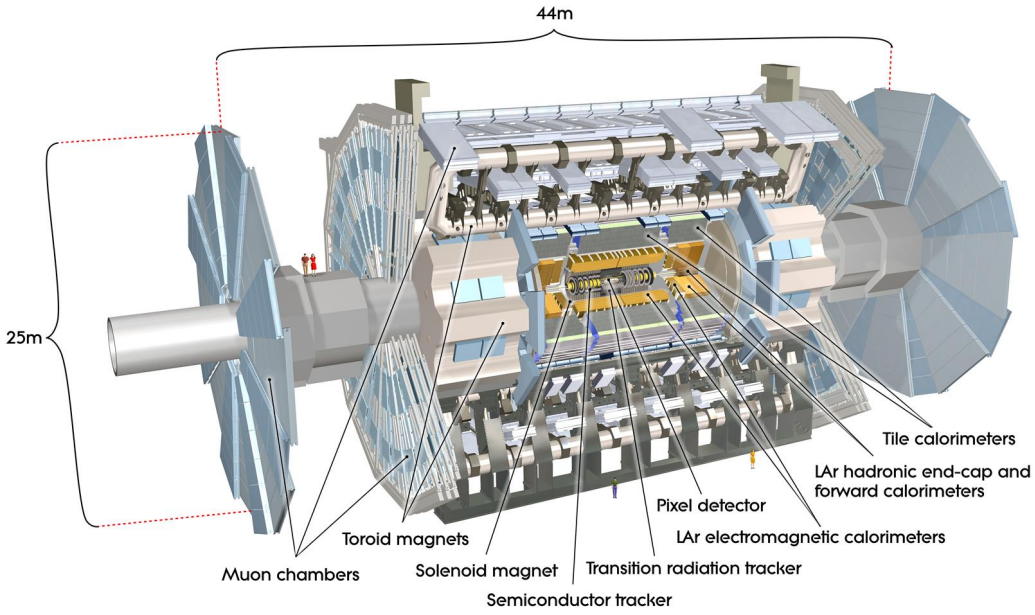


Figure 3.4: A cut away view of the ATLAS detector with the various detector components labelled [54].

around the nominal interaction point (IP), providing nearly complete 4π solid angle coverage. The ATLAS detector is divided into a variety of subdectectors, with the inner detector lying closest to the beam line, followed by the electromagnetic and hadronic calorimeters and finally the muon spectrometer. These various subdetectors are detailed in the following sections, and Figure 3.4 shows a diagram of the complete ATLAS detector with these various subdetectors labelled.

ATLAS uses a right-handed coordinate system with the IP defined as the origin. The positive x axis is directed towards the centre of the LHC ring, with the beam line specifying the z direction and the positive y axis pointing upwards. The polar and azimuthal angles, θ and ϕ, are defined with respect to the z and x axes, respectively. Angular quantities within the ATLAS detector are frequently expressed using ϕ and

the rapidity,

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{3.1}$$

which uses the same symbol as the y coordinate. For the remainder of this thesis, y will refer to the rapidity unless otherwise specified. The distance between objects can then be defined as,

$$\Delta R = \sqrt{\Delta y^2 + \Delta \phi^2}. \tag{3.2}$$

As opposed to using θ directly, the pseudorapidity may also be used,

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right). \tag{3.3}$$

It is an approximation of y, and is equivalent when the particle in question is massless. It is useful for defining detector boundaries as it does not depend on the particle's momentum.

Transverse quantities are also often used within ATLAS, and are defined in the xy cartesian plane. For example, the transverse momentum (p_{T}) is defined as

$$p_{\mathrm{T}} = \sqrt{p_{\mathrm{x}}^2 + p_{\mathrm{y}}^2}. \tag{3.4}$$

Figure 3.5 shows a schematic description of the coordinate system.

3.2.1 Inner detector

The inner detector (ID) makes use of a combination of detector technologies to measure charged particles within the $|\eta| < 2.5$ region. This section will outline the various

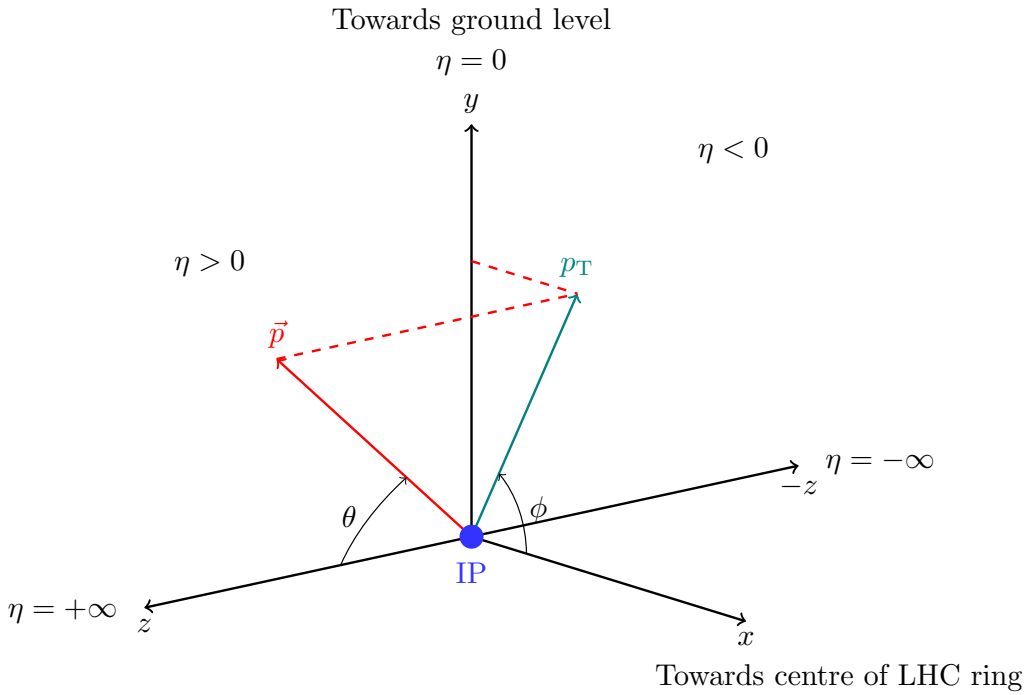


Figure 3.5: A graphical representation of the coordinate system used in the ATLAS detector with relevant quantities labelled. The momentum of a hypothetical particle is shown in red, while the transverse momentum for the particle is shown projected onto the xy cartesian plane.

detector systems, which are also summarized in Figure 3.6. The ID is immersed in a 2 T solenoidal magnetic field, giving charged particles a curved trajectory. The trajectory of a charged particle is known as a track, and is used to reconstruct the particle's momentum.

In addition to measuring the charged particle momentum, the ID is also the primary subdetector used for performing vertex identification. A vertex marks the location within the beam line where a proton-proton collision occurs. Due to the high beam intensity at the LHC, there are typically several primary and secondary vertices per bunch crossing, which can be identified by determining the origin of the tracks as measured by the ID. The procedure for track and vertex reconstruction is described in more detail in Section 5.1.

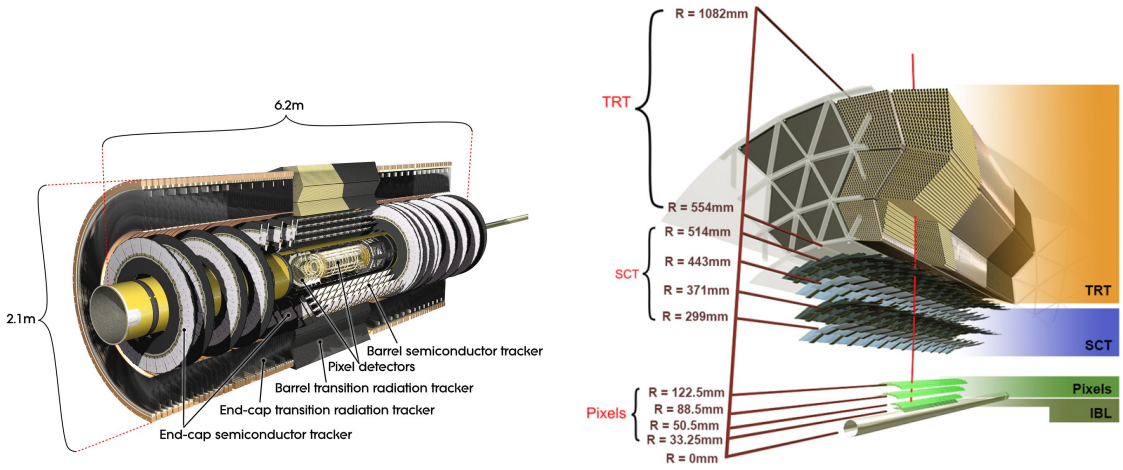


Figure 3.6: Two views of the ATLAS inner detector. The left shows a cut away view with each detector component shown. The right shows a more detailed view of the various detector components as a function of the radial distance R from the beam line [55, 56].

The high-granularity silicon pixel detectors [57] are placed closest to the beam line. In the barrel region, the pixel layers are arranged around the beam line in concentric cylinders, while each end-cap region is outfitted with three disks centered around the beam line. The barrel region was initially comprised of three layers, however an additional layer of pixel detectors known as the insertable B-layer (IBL) [58] was added closest to the beam line after the completion of Run 1. The IBL was needed to improve the robustness and precision of tracking measurements in light of energy and luminosity increases going into Run 2 and beyond. Each pixel in the original layers is $50 \times 400\,\mu\text{m}^2$ in size, while the pixels in the IBL have a slightly smaller size of $50 \times 250\,\mu\text{m}^2$. This results in a resolution of $10\,\mu\text{m} \times 115\,\mu\text{m}$ for the original pixels, and $10\,\mu\text{m} \times 60\,\mu\text{m}$ for the IBL.

The silicon microstrip tracker (SCT) [59] consists of four layers of silicon strip detectors in the barrel region, and nine disks in each of the end-cap regions. Each layer consists of two strip sensors with a strip pitch of 80 µm and an angle of 40 mrad

between the two sensors in order to have access to all spatial coordinates. This results a SCT resolution of 17 µm × 580 µm.

The transition radiation tracker (TRT) [60] is the outermost layer of the ID and is comprised of a collection of 4 mm diameter proportional counter drift tubes known as straws. The TRT covers the $|\eta| < 2.0$ region, with 144 cm long tubes arranged parallel to the beam line in the barrel region and disks in the end-cap region with 37 cm long straws arranged radially. The straws contain a gas mixture, which is ionized as charged particles pass through the straw. The resulting free electrons are amplified and collected at the anode. Additionally, the straws are interspersed with a radiator material which results in transition radiation photons being emitted when charged particles cross from the radiator to the straws. These transition radiation photons can also be absorbed by the gas mixture and result in a larger signal amplitude. This allows for the TRT to be used for electron identification as the electrons produce a larger amount of transition radiation than a minimum ionizing particle. The TRT has an overall resolution of 130 µm which is significantly worse than the silicon based detectors, however it compensates by providing many more hits.

3.2.2 Calorimeters

Next closest to the beam line after the ID is the ATLAS calorimeter system, which is outlined in Figure 3.7. The combined system is arranged cylindrically around the beam line and provides coverage in the $|\eta| < 4.9$ region. It is designed to contain all activity from electrons, photons, and hadrons. The individual systems consist of a variety of sampling calorimeters, which are comprised of alternating passive absorber layers and active layers that provide the measurements. There are two main subsystems in the ATLAS calorimeter system: first is the liquid argon (LAr) elec-

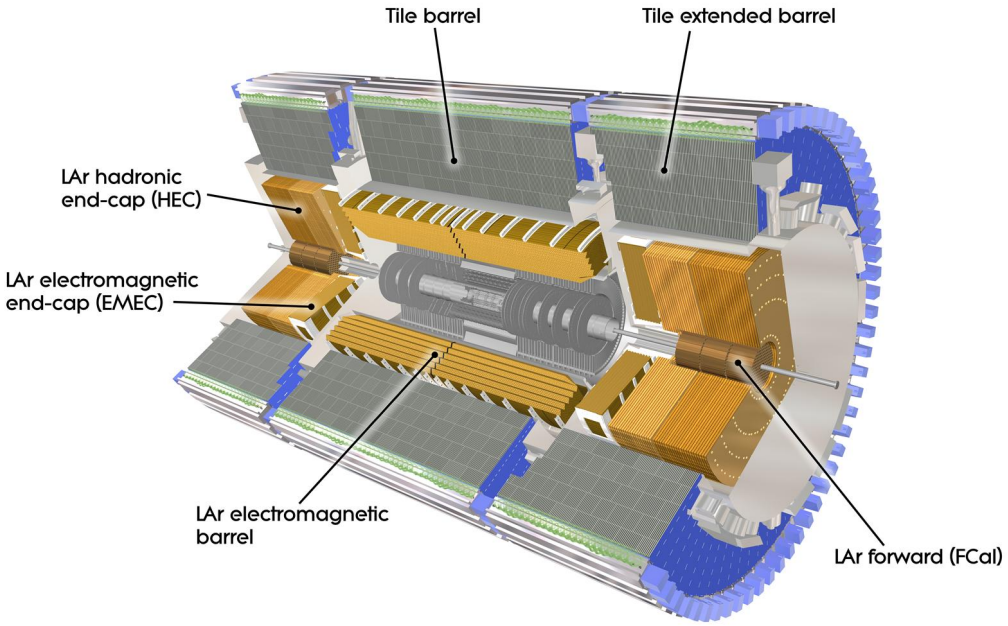


Figure 3.7: A cut away view of the ATLAS calorimeter system [61].

tromagnetic calorimeter (ECAL), and second the hadronic calorimeter (HCAL). The ECAL is designed to contain all activity originating from electrons and photons, while the HCAL is designed to contain all remaining hadronic activity. These subsystems are outlined in the sections below.

3.2.2.1 LAr electromagnetic calorimeter

In the barrel region, which covers the $|\eta| < 1.475$ region, and the end-cap region, which covers the $1.375 < |\eta| < 3.2$ region, the ECAL is an accordion shaped lead-LAr detector [62], with lead being the absorber and LAr being the active material. In the forward region, $3.1 < |\eta| < 4.9$, the first layer of the forward calorimeter (FCal) is designed to contain electromagnetic (EM) showers. It also uses LAr as the active material, but copper is used as an absorber instead of lead. The thickness of the lead absorber plates is optimized as a function of $|\eta|$, resulting in a thickness larger than

22 radiation lengths (X_0) in the central region, and larger than $24X_0$ in the end-cap and forward regions.

In the barrel region, the EM calorimeter contains three layers of varying granularity as well as a presampler layer, which is designed to correct for energy lost by electrons and photons before reaching the calorimeter. The first layer is finely segmented in η in order to obtain an accurate position measurement. The second layer is the largest and is designed to contain the majority of the EM shower, and the third layer has coarser granularity and is designed to contain high energy electrons and photons.

The electromagnetic end-cap calorimeters (EMEC) consist of two coaxial wheels in each end-cap region, with the outer wheel providing coverage in the $1.375 < |\eta| < 2.5$ region and the inner wheel covering the $2.5 < |\eta| < 3.2$ region. Like the ECAL in the barrel region, the outer wheel possesses three layers, while the inner wheel has two layers with a slightly coarser granularity.

The FCal, which was partially constructed at Carleton University, provides coverage in the high flux $3.1 < |\eta| < 4.9$ region. Due to this high flux, the FCal has a very dense configuration and thus has smaller LAr gaps than the ECAL. Only the first of three layers, each 45 cm thick, is dedicated to the containment of EM showers.

3.2.2.2 Hadronic calorimeters

The HCAL is a combination of calorimeter technologies. The tile calorimeter [63], which uses steel as an absorber material and scintillating tiles as the active material, covers the $|\eta| < 1.0$ barrel region and also has an extended barrel portion that provides coverage in the $0.8 < |\eta| < 1.7$ region. The tile calorimeter is divided into three layers, totaling a thickness of around 10 interaction lengths (λ).

Covering the $1.5 < |\eta| < 3.2$ region, the hadronic end-cap calorimeters (HEC)

are located directly behind the EMEC. The HEC calorimeters utilize LAr as an active material as in the ECAL, however the absorber material is flat copper plates as opposed to lead. The HEC is composed of two wheels with two layers each, for a total of four layers.

Finally, the outer two layers of the FCal are dedicated to containment of hadronic showers. Like the first layer, these two layers are 45 cm thick, but use tungsten as opposed to copper as the absorber material.

3.2.3 Muon spectrometer

The muon spectrometer (MS) [64] lies outside the hadronic calorimeters and is the detector subsystem that is the farthest from the beam line. Due to their nature as minimum ionizing particles, muons generally pass through the calorimeters without depositing very much energy. Therefore, a dedicated system is required to identify and perform precise measurements of muons.

The muon spectrometer is shown in Figure 3.8 and operates based on the deflection of muon tracks within a magnetic field produced by toroid magnets. The $|\eta| < 1.4$ region is immersed in a magnetic field with an approximate magnitude of 0.5 T produced by a large barrel toroid. The $1.6 < |\eta| < 2.7$ region is immersed in an approximately 1.0 T magnetic field produced by two smaller end-cap toroids. A diagram of the barrel toroid, as well as the two end-cap toroids is shown in Figure 3.9. The muon spectrometer provides measurements in the $|\eta| < 2.7$ region and is also equipped with a trigger system that provides coverage in the $|\eta| < 2.4$ region.

The muon spectrometer is instrumented with various detector systems depending on the η region in question. The precision tracking measurements are performed by the monitored drift tubes (MDT) and the cathode strip chambers (CSC). The

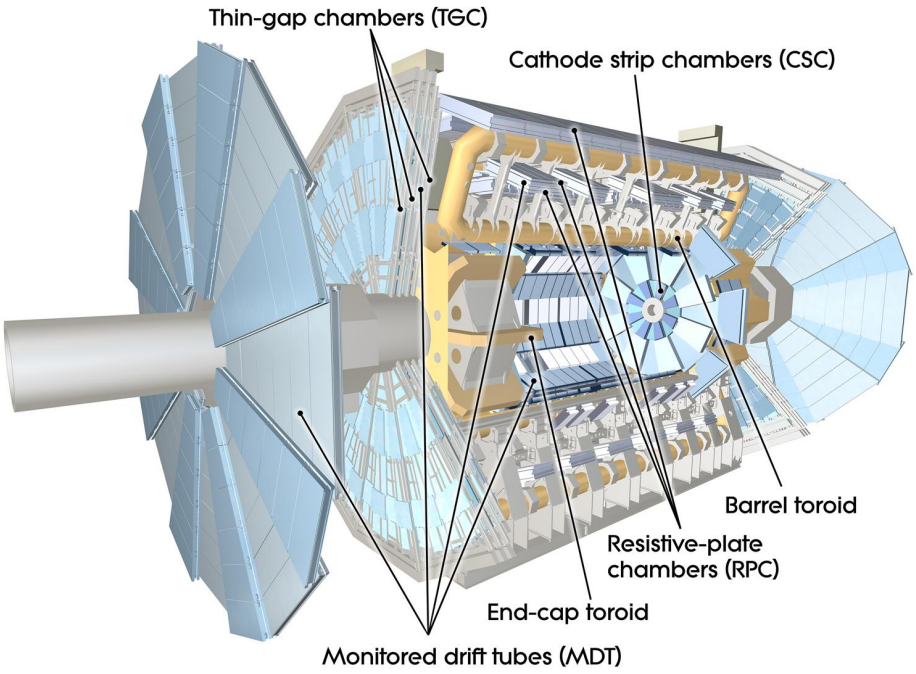


Figure 3.8: A cut away view of the muon spectrometer in the ATLAS detector [65].

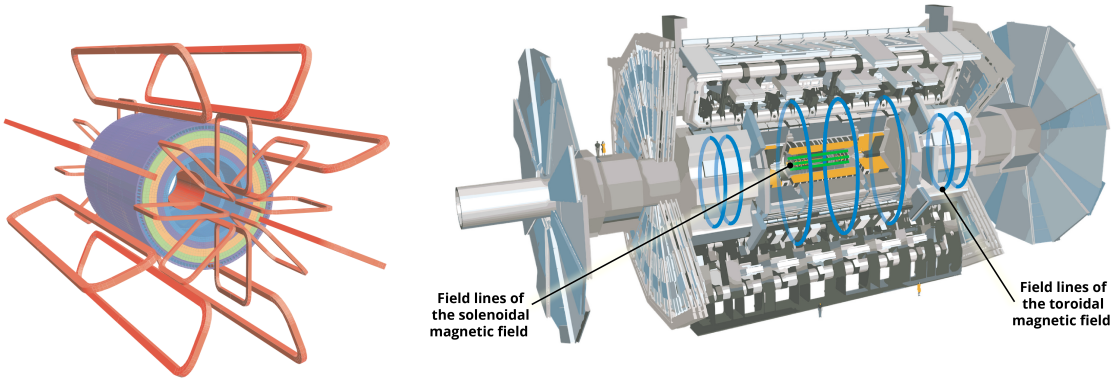


Figure 3.9: The diagram on the left, taken from Ref. [3], shows the eight individual coils for the air-core barrel toroid arranged outside the hadronic calorimeter (the cylindrical shell in the centre of the image) as well as the coils for the two end-cap toroids. The right image [66] shows the direction of the resulting magnetic field in the muon spectrometer due to the toroids, as well as the direction of the magnetic field produced by the solenoid in the inner detector.

MDTs provide measurements in the $|\eta| < 2.7$ region and are complemented by the CSCs over $2.0 < |\eta| < 2.7$ due to the higher event rate in that region and the need for better time resolution.

Triggering is performed using resistive plate chambers (RPC) in the $|\eta| < 1.05$ barrel region and using thin gap chambers (TGC) in the $1.05 < |\eta| < 2.4$ end-cap region. The MDTs only provide information in the bending plane, thus the RPCs and TGCs are used as secondary trackers to give the muon tracks a coordinate in ϕ once the hits from the MDT and the trigger are matched.

3.2.4 Trigger system

The goal of the trigger system [67] is to identify events of interest in order to reduce the amount of information that is saved from the detector. In normal operation, the event rate in the ATLAS detector is 40 MHz and saving every event is not feasible.

The ATLAS trigger system is divided into two main stages. The level 1 (L1) trigger is a hardware-based system and makes use of reduced granularity information from the calorimeters and the muon spectrometer in order to identify particle candidates, as well as large amounts of missing transverse energy. It also identifies regions of interest (RoI) in η and ϕ to be investigated further. The L1 trigger reduces the event rate to 100 kHz and has a latency of 2.5 µs.

The high-level trigger (HLT) is the software based second stage of the trigger system. It examines events that are passed from the L1 trigger and more closely examines the L1 trigger RoIs with progressively more complex algorithms to determine whether the event is useful for physics. If the event passes the HLT then it is saved to permanent storage. The HLT reduces the event rate to 1.2 kHz, resulting in an average 1.2 GB/s of data being saved to permanent storage.

3.2.5 Luminosity measurement

The LHC delivers a very high instantaneous luminosity, L, to the ATLAS detector. The luminosity is determined by the beam properties but can be conceptualized as a measure of the number of particles per unit area per unit time. In the LHC case, this quantity can be expressed as

$$L = \frac{f N^2 n_b}{4\pi \sigma_x \sigma_y}, \tag{3.5}$$

where f is the revolution frequency of the LHC, N is the number of protons in each colliding bunch, n_b is the number of bunches and σ_x and σ_y are the width and height of the beam. In practice, the exact beam parameters are not known. Furthermore, the beam parameters may flucuate as the bunches circulate in the beam pipe and thus the quantity reported by the LHC may not match with the luminosity being received by the ATLAS detector. It is thus simpler and more accurate to simply measure the luminosity directly. At ATLAS, this task is primarily performed by the LUCID-2 (LUminosity Cherenkov Integrating Detector) detector [68].

At the LHC, data is taken over a long period of time. Thus, the total integrated luminosity over a period of data-taking can be defined as

$$\mathcal{L} = \int L \, dt. \tag{3.6}$$

This quantity can be thought of as the size of the dataset. As an example, Figure 3.10 shows the measured total integrated luminosity over the course of Run 2, with the final total considered suitable for physics measurements for this period of data-taking being $139\,\text{fb}^{-1}$.

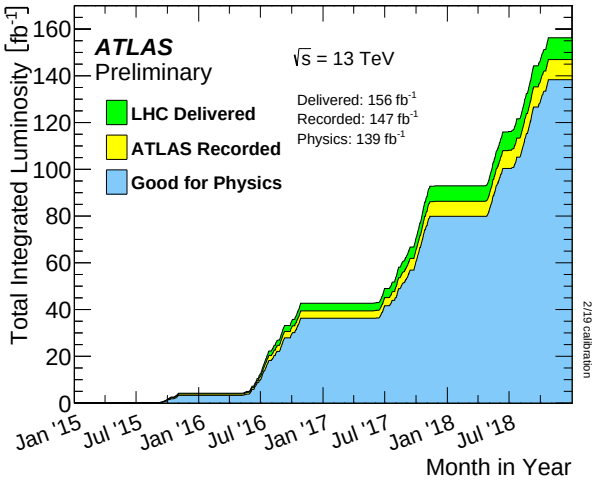


Figure 3.10: The total integrated luminosity measured over the Run 2 data-taking period [69]. The green represents the total integrated luminosity delivered by the LHC, the yellow represents the total integrated luminosity recorded by the ATLAS detector, and the blue represents the total integrated luminosity that is suitable for making physics measurements.

Chapter 4

Data and simulation samples

This chapter describes the recorded and simulated datasets that are analyzed for this thesis. Section 4.1 describes the data collected using the ATLAS detector and Section 4.2 discusses the various Monte Carlo simulated samples that provide predictions of the data.

4.1 Data

This analysis is performed using data from pp collisions with $\sqrt{s} = 13\,\text{TeV}$ produced by the LHC over the full Run 2 data-taking period from 2015–2018, and measured by the ATLAS detector. Of the data collected by the trigger during Run 2, a total integrated luminosity of $139.0\,\text{fb}^{-1}$ with an uncertainty of 1.7% is considered to be suitable for use in physics measurements [70]. Due to the high beam intensity of the LHC more than one inelastic pp interaction per bunch crossing is expected. In Run 2, the average number of interactions per bunch crossing, $\langle\mu\rangle$, was 33.7 as shown in Figure 4.1. However, this quantity varies widely from year to year depending on the luminosity conditions within the LHC. As such, both the total integrated luminosity

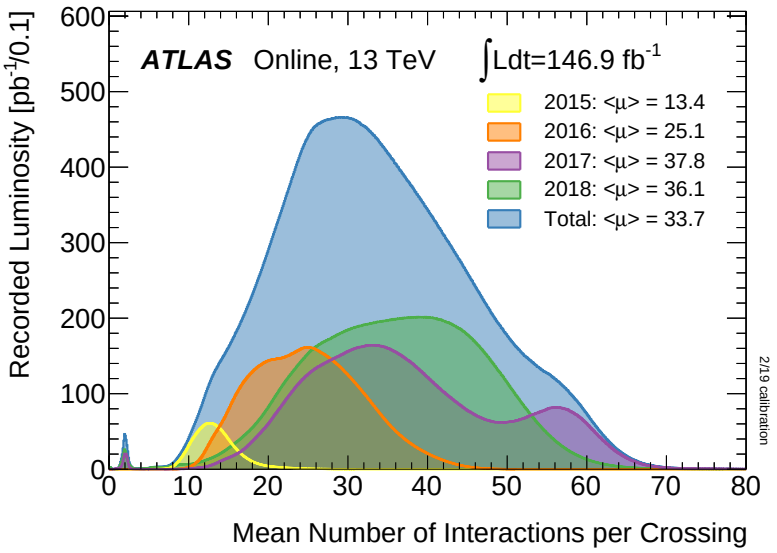


Figure 4.1: Luminosity weighted distribution of the measured μ in the ATLAS Run 2 dataset [69].

Year	$\mathcal{L}_{\text{int}}$ (fb^{-1})	$\langle\mu\rangle$
2015	3.21	13.4
2016	32.99	25.1
2017	44.31	37.8
2018	58.45	36.1
Total	139.0	33.7

Table 4.1: The integrated luminosity, $\mathcal{L}_{\text{int}}$, and the average number of interactions per bunch crossing, $\langle\mu\rangle$, of the data for each year of data-taking in the full Run 2 dataset [69].

suitable for use in physics measurements and the average number of interactions per bunch crossing for each year are summarized in Table 4.1.

4.2 Simulation

The production of MC samples that can be compared with the data collected by the ATLAS detector is a multi-step process. The first step is generation, whose basic methodology was introduced in Section 2.2. This section will describe the various

generators used in this analysis in more specific terms. After the generation step has produced the relevant particles, the sample will undergo two subsequent steps known as simulation and digitization, discussed in Section 4.2.5. These steps will determine how the particles produced in the generation step will interact with the ATLAS detector material, as well as the raw signals that the detector will produce as a result.

One of the main benefits of utilizing MC samples is that they allow for access to the true particles in an event (known as the truth or particle level MC) as well as the reconstructed versions of the particles (reconstructed or dectector level MC). The relationship between the reconstructed and truth level MC can thus be used to correct for the effects of the ATLAS detector such that the data can be recovered at truth level. This process is known as unfolding and is discussed in greater detail in Chapter 6. MC samples are also used to estimate the number of events in the data that originate from background processes.

A dedicated MC sample is produced that is specific to each individual process (e.g. strong $Z(\rightarrow \mu\mu)$+jets, electroweak $Z(\rightarrow \mu\mu)$+jets, etc.) and thus there are several MC samples present in this analysis. These are outlined in detail in the sections below, and can be found summarized in Table 4.2. Due to differing pileup and detector conditions, MC samples are divided into different campaigns depending on the years of data-taking they correspond to: MC16a for 2015–2016, MC16d for 2017 and MC16e for 2018. The analysis is performed independently on these three campaigns with a procedure identical to the one used for the corresponding data-taking years.

4.2.1 Strong Z+jets samples

In total, there are five strong $Z(\rightarrow \mu\mu)$+jets samples considered in this analysis, with some being used to a greater extent than others.

The MadGraph+Pythia8 FxFx sample [71] uses MadGraph5_aMC@NLO version 2.6.5 [72] with the NNPDF3.1 NNLO PDF set [73] to produce up to three additional final state partons at NLO accuracy. The parton showering and hadronization is then performed using Pythia 8.240 [74] with the A14 tune [41] and the NNPDF2.3 LO PDF set [75]. The different jet multiplicities are merged using the FxFx NLO matrix element and parton shower merging prescription [34]. EvtGen 1.7.0 [76] is used to simulate the decays of bottom and charm hadrons and Pythia8 is used to simulate final state QED radiation.

Two datasets are created using the Sherpa MC generator. The first, produced using Sherpa 2.2.1 [77, 78], uses the Comix [79] and OpenLoops [80] libraries to calculate matrix elements for up to two additional final state partons at NLO accuracy and up to four additional partons at LO accuracy. The default Sherpa showering and hadronization methods are used with a tune developed by the Sherpa authors. The matrix elements and parton shower are matched using the MC@NLO algorithm [81], and different jet multiplicities are merged using an improved CKKW matching procedure [82, 83] which is extended to NLO using the MEPS@NLO method [84]. The NNPDF3.0 NNLO PDF set [39] is used for the Sherpa 2.2.1 sample.

Sherpa 2.2.11 [71, 78] utilizes a similar configuration to version 2.2.1, however it includes LO accurate matrix elements for up to five additional final state partons instead of four and the Hessian NNPDF3.0 NNLO PDF set [39, 85] is used. Sherpa 2.2.11 implements an improved splitting procedure in the parton showering algorithm which results in improved modelling. It also includes an improved match-

ing procedure facilitated by an updated treatment regarding unordered parton shower histories[1], which results in a suppression of the cross section for high p_{T}^{Z} compared to Sherpa 2.2.1.

The MadGraph+Pythia8 CKKW-L sample uses MadGraph5_aMC@NLO version 2.2.3 [72] with the NNPDF3.0 NLO PDF set [39] to produce LO accurate matrix elements for up to four final state partons. The parton showering and hadronization are then performed using Pythia 8.210 [74] with the A14 tune [41] and the NNPDF2.3 LO PDF set [75]. Overlaps between the matrix element and the parton shower are removed using the CKKW-L merging procedure [32, 33] and EvtGen 1.2.0 [76] is used to simulate the decays of bottom and charm hadrons.

Lastly, the Powheg+Pythia8 sample uses Powheg Box v1 [86–88] to simulate the hard scatter process at NLO accuracy in QCD using the CT10nlo PDF set [89]. Parton showering, hadronization, and the underlying event are modelled using Pythia 8.186 [74] with parameters set according to the AZNLO [90] tune and using the CTEQ6L1 PDF set [89]. EvtGen 1.2.0 [76] is used to simulate the decays of bottom and charm hadrons and PHOTOS++ [91,92] is used to simulate final state QED radiation.

The MadGraph+Pythia8 FxFx sample is the primary sample used for this analysis, with Sherpa 2.2.11 used as an alternative. Powheg+Pythia8 is used to perform certain tests, and Sherpa 2.2.1 and MadGraph+Pythia8 CKKW-L are

[1] A part of the matching procedure within the Sherpa generator involves identifying the start of a parton shower. This is done by performing the showering backwards using a k_{T}-like algorithm (see Section 5.3). α_s is evaluated at each node such that it can be used to calculate a Sudakov form factor that is then applied to each propagator. This generally leads to a sequence of nodes with an increasing k_{T} scale (related to the transverse momentum), and small Sudakov form factors suppressing contributions from high α_s values. However, in extreme regions of phase space this strict sequence of increasing k_{T} may be violated. This can lead to the absence of the suppressing Sudakov factors, resulting in artificially inflated values of the cross section. Sherpa 2.2.11 better handles these cases compared to previous versions of the generator by using the largest previously identified k_{T} scale if the order of the sequence is violated.

examined only for comparison and as a cross-check.

4.2.2 Diboson samples

Various diboson samples with either fully leptonic or semileptonic final states are also considered. These samples are simulated using either Sherpa 2.2.1 or Sherpa 2.2.2 [78] depending on the process in question and use the NNPDF3.0 NNLO PDF set [39]. Matrix elements are generated at NLO accuracy for up to one parton and at LO accuracy for up to three additional partons. The parton showers are produced and matched to the matrix element calculations with Catani–Seymour dipole factorization [93] using the MEPS@NLO method [84]. Virtual QCD corrections were provided by the OpenLoops library [80] and a tuned set of parton shower parameters developed by the Sherpa authors were used.

4.2.3 Electroweak Z+jets samples

Electroweak Zjj production is simulated using the vector-boson fusion (VBF) approximation[2] [94, 95] with Herwig 7.2.0 [96, 97]. The events are produced at NLO accuracy in the strong coupling using the VBFNLO v3.0.0 program [98] to provide the loop amplitudes. The MMHT2014LO PDF set [99] is used, and the parton showering, hadronization and underlying event uses the default set of tuned parameters provided by the generator.

4.2.4 Top samples

Various processes involving top quarks are considered due to their potential to appear as background events within the analysis.

[2]This approximation neglects colour exchange between the incoming fermions as well as certain interference effects between the t and u channel diagrams.

$t\bar{t}$ events with non-all-hadronic decays are considered, and are produced at NLO accuracy using POWHEG BOX v2 [86–88, 100]. The NNPDF3.0 NLO PDF set [39] is used and the h_{damp} parameter, which is a resummation damping factor and one of the parameters that controls the matching of matrix elements to parton showering, is set to $1.5m_{\text{top}}$ [101]. The parton level events are then interfaced with PYTHIA 8.230 [74] which models the parton showering, hadronization and underlying event using the NNPDF2.3 LO PDF set [75] and parameters set according to the A14 tune [41]. EVTGEN 1.6.0 [76] is used to simulate decays of bottom and charm hadrons.

The associated production of top quarks with W bosons (tW) is modelled with POWHEG BOX v2 at NLO accuracy using the five-flavour scheme and the NNPDF3.0 NLO PDF set. The diagram removal scheme [102] was used to remove interference and overlap with the $t\bar{t}$ sample. The events are then passed to PYTHIA 8.230, which performs the showering, hadronization and underlying event using the A14 tune and the NNPDF2.3 LO PDF set.

Single top production was considered in both the s and t channel. Both channels were modelled with POWHEG BOX v2 at NLO accuracy using the four-flavour scheme and the NNPDF3.0 NLO PDF set. The events are then passed to PYTHIA 8.230, which performs the showering, hadronization and underlying event using the A14 tune and the NNPDF2.3 LO PDF set.

4.2.5 Simulating ATLAS detector conditions and response

The detector response is determined by inputting the simulated events into the GEANT4 simulation [103] for the ATLAS detector [104]. The events are then reconstructed using the same procedure that is used for data (see, for example, the procedures outlined in Chapter 5).

Process	Generator	ME order	Generator PDF	Parton showering (Generator, PDF, tune)
Strong Z+jets				
Strong Z+jets	MadGraph5_aMC@NLO	0-3j@NLO	NNPDF3.1 NNLO	Pythia 8.210, NNPDF2.3 LO, A14
	Sherpa 2.2.11	0-2j@NLO+3-5j@LO	Hessian NNPDF3.0 NNLO	Sherpa 2.2.11, defaults
	Sherpa 2.2.1	0-2j@NLO+3-4j@LO	NNPDF3.0 NNLO	Sherpa 2.2.1, defaults
	MadGraph5_aMC@NLO	0-4j@LO	NNPDF3.0 NLO	Pythia 8.210, NNPDF2.3 LO, A14
	Powheg Box v1	NLO	CT10nlo	Pythia 8.186, CTEQ6L1, AZNLO
Electroweak Z+jets				
EW Z+jets	Herwig 7.2.0 + VBFNLO	NLO	MMHT2014LO	Herwig 7.2.0, defaults
Diboson				
Fully-leptonic	Sherpa 2.2.2	0-1j@NLO+2-3j@LO	NNPDF3.0 NNLO	Sherpa 2.2.2, defaults
Semi-leptonic	Sherpa 2.2.1	0-1j@NLO+2-3j@LO	NNPDF3.0 NNLO	Sherpa 2.2.1, defaults
Top				
$t\bar{t}$	Powheg Box v2	NLO	NNPDF3.0 NLO	Pythia 8.230, NNPDF2.3 LO, A14
tW	Powheg Box v2	NLO	NNPDF3.0 NLO	Pythia 8.230, NNPDF2.3 LO, A14
Single t	Powheg Box v2	NLO	NNPDF3.0 NLO	Pythia 8.230, NNPDF2.3 LO, A14

Table 4.2: A summary of the MC samples used in this analysis, as discussed in detail in section 4.2. The generator used to produce the process is shown in the second column and the third column shows the matrix element order produced by the generator, where j represents final state partons (jets). The PDF used by the generator, as well as the details related to the parton showering are also shown.

The effects due to multiple pp collisions that occur within the same or a neighbouring bunch crossing (pileup) are simulated [105] using inelastic pp collisions produced using the Pythia8 event generator with the A3 tune [106] and using the NNPDF2.3 LO PDF set. The resulting events are then overlaid with the existing sample at a rate that is consistent with the measured value from data.

4.2.6 Monte Carlo event weights

A key difference between the data collected by the ATLAS detector and the events produced by an MC generator is that the MC generator produces events by sampling the scattering amplitude for a specific process over the full partonic phase space [107]. As some regions of phase space will have an associated cross section that is larger than others, produced MC events are assigned a weight to reflect this.

Negative weights are also present in the MC samples described in this section. These are largely due to corrections at NLO, which can result in negative contributions to the cross section from both the matrix element calculation and the parton showering algorithm [108]. Having a high fraction of negative weights is detrimental

to the statistical power of the sample, an effect which is discussed in more detail in Chapter 7.7.

In order to make use of the MC samples in an analysis, they must be normalized to the predicted rate: $\sigma_{\mathrm{MC}} \times \mathcal{L}$, where σ_{MC} is the cross section provided by MC generator. In the case where there exists a cross section prediction at higher order than the MC generator, then σ_{MC} may be scaled by a so-called k-factor in order to bring it into agreement with the improved prediction. This gives the following MC event weight for each event i,

$$w_i^{\mathrm{MC}} = w_i^{\mathrm{gen}} \frac{k\mathcal{L}\sigma_{\mathrm{MC}}}{\sum_i w_i^{\mathrm{gen}}}, \tag{4.1}$$

where w_i^{gen} is the weight directly supplied by the generator. This event weight will also undergo further corrections as detailed in Section 7.4 in order to adjust the simulation to better agree with what is observed in the data. In order to illustrate the distributions of the MC event weights, Figure 4.2 shows the normalized MC event weights in the strong Z+jets samples after the selection and corrections discussed in Chapter 7.

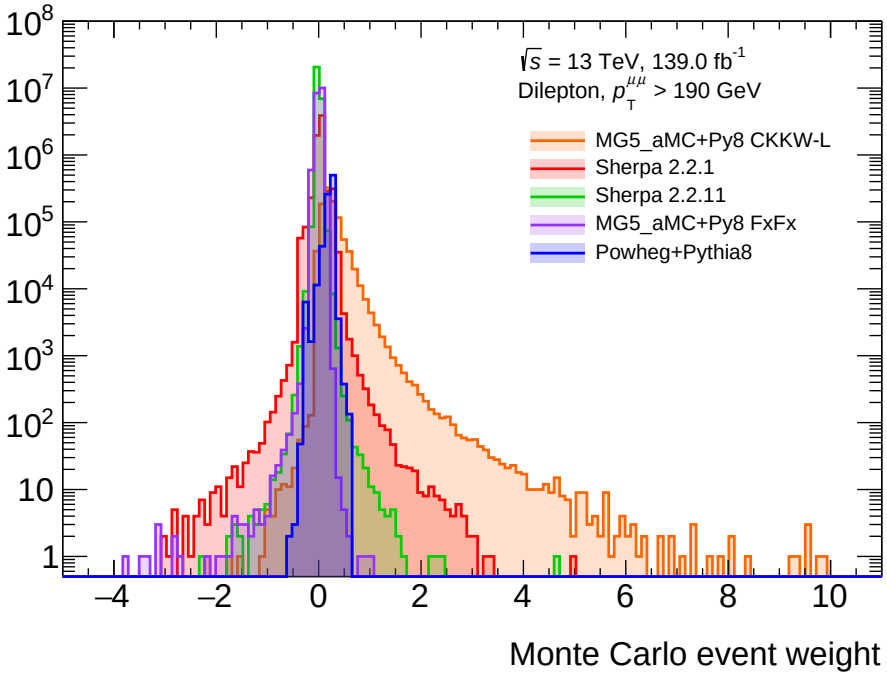


Figure 4.2: The distributions of MC event weights for each of the strong Z+jets samples present in this analysis. The normalization in Eqn. 4.1 as well as the selection criteria and corrections described in Chapter 7 are applied.

Chapter 5

Object reconstruction

In order to be useful for physics measurements, the raw information obtained from the ATLAS detector systems must be processed to reconstruct the particles that created the detector signals. In general, there are two main inputs to these reconstruction schemes: the measured trajectories of charged particles, known as tracks, and energy deposits in the calorimeter. Tracks are discussed in Section 5.1, while additional information on calorimeter energy clustering can be found in Ref. [109]. As calorimeter energy clusters are not used in this analysis, they will not be described in detail here.

Electrons and photons, muons, taus and jets all have their own dedicated reconstruction algorithms that use tracks and calorimeter clusters as input. In the analysis performed for this thesis, only muons and jets built from tracks are considered and thus their reconstruction algorithms are described below in Sections 5.2 and 5.3.

5.1 Tracks and vertices

As was discussed in Section 3.2.1, tracks are measured using information collected from the inner detector. As charged particles pass through the layers in the ID they

deposit energy, creating what are known as hits in the detector layers. These hits form the starting place for the track reconstruction algorithm [110, 111].

The first step of the track reconstruction algorithm is known as clusterization, which groups together hits in adjacent pixels or strips in the silicon layers of the ID. These clusters are then used to construct a three-dimensional coordinate, known as a space-point, where a charged particle crossed the detector layer. Once all possible space-points have been determined, track seeds are formed from sets of three compatible space-points.

The so-called impact parameters are then estimated for each track seed. A reconstructed track is described by five parameters, $(d_0, z_0, \phi, \theta, q/p)$, defined with respect to the track's closest approach to a reference point, as shown in Figure 5.1. d_0 and z_0 describe the transverse and longitudinal distance from the point of closest approach to the reference point and are known as the transverse and longitudinal impact parameters, respectively. ϕ and θ are as defined in the previous chapter, and q/p is the charge of the track divided by the track momentum. In the case of the track seeds, the reference point is chosen to be the centre of the beam spot.

In order to reduce the contribution from fake track seeds, basic selection criteria are applied based on the estimated impact parameters and track seed momentum. Furthermore, one additional space-point is required to be compatible with the track seed. Finally, the purity of the track seed depends on the part of the detector it is measured in, thus the track seeds are considered first if they are measured in the SCT, followed by those measured in the pixel detectors, and lastly the mixed detector seeds. A combinatorial Kalman filter [113] is then used to construct track candidates by extrapolating the track seed to incorporate additional compatible space-points.

There may be more than one track candidate associated with a single space-

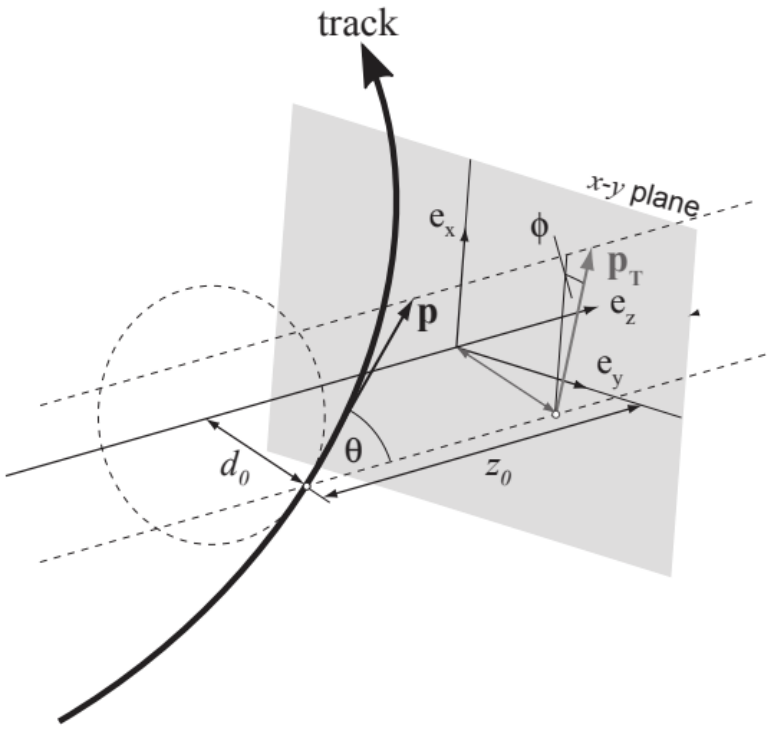


Figure 5.1: A representation of a track with momentum p and charge q in the ATLAS detector. The θ and ϕ angles have been labelled, in addition to the transverse and longitudinal impact parameters d_0 and z_0. The origin represents the centre of the beam spot [112].

point and thus an ambiguity resolving step is required. Each track candidate is given a score, which depends on the number of clusters assigned to the track candidate, the number of holes (layers where a hit is expected but none is observed), the χ^2 of the track fit and the momentum of the track candidate. Ambiguities are then resolved based on this track score and the determination of whether a specific cluster is likely to be the result of multiple tracks based on the result of a neural network (NN) classifier.

A high resolution global fit is performed for the surviving track candidates and an extension to the TRT is attempted before the track candidates are added to the final track collection. The extension to the TRT once again makes use of a combina-

torial Kalman filter, and the compatible TRT hits are added to the track candidate. If the extension is successful, then the fit is performed once again including the additional TRT hits before the track candidate is added to the final track collection.

The above description is known as inside-out track building, as the process begins by examining information from the pixel detector and the SCT before considering the information from the TRT. There is a secondary method known as outside-in tracking, which is used to provide better reconstruction efficiency for charged particles produced further from the beam line (e.g. a photon converting to an electron-positron pair, secondary decays) or in cases where a track is removed due to a large amount of ambiguous hits. In order to perform outside-in tracking, track segments in the TRT are constructed by means of a Hough transform [114] using hits that have not been previously assigned to a track using the inside-out track reconstruction algorithm. Close by to such a TRT segment, track seeds in the silicon detectors composed of two space-points are constructed. These track seeds are then extended in the same manner as the inside-out method and added to the final track collection.

Primary vertex reconstruction [115] is performed using the final track collection after the application of cuts to reduce the number of fake tracks. Due to the large number of protons per bunch crossing, there are typically several reconstructed primary vertices in each one. In order to determine the position of a vertex, a seed position must first be chosen. The transverse seed position is defined as the centre of the beam spot, which indicates the position where pp collisions occur and extends 42 mm in the z-direction. The mode of the z_0 of all the input tracks is used as the z-coordinate. An iterative χ^2 minimization fitting procedure is then applied. For each iteration, tracks that are less compatible with the vertex position are down-weighted and the vertex position is recomputed. Once the vertex position is determined, the

tracks that are incompatible with it are used to determine the position of additional vertices.

After all the primary vertices have been reconstructed, the hard scatter vertex is identified as the primary vertex from which the highest $\sum p_{\mathrm{T}}^2$ originates. The remainder of the primary vertices are classified as pileup vertices. Pileup refers to any activity that cannot be attributed to the hard scatter. It can result from additional proton-proton collisions within the same bunch crossing, from proton-proton collisions within adjacent bunch crossings, or other background events that result in a signal in the detector (e.g. cavern background, proton collisions with gaseous particles in the beam pipe). The activity related to pileup tends to be of lower energy than the activity from the hard scatter vertex, and is an important background to understand and minimize.

5.2 Muons

Muon reconstruction [116] is primarily based on information collected from the MS and the ID, with some information from the calorimeter also being incorporated. Tracks in the ID are reconstructed as discussed in the previous section, while tracks in the MS have their own dedicated reconstruction scheme.

Tracks in the MS are reconstructed by first using a Hough transform to identify short segments of hits within an individual detector system in the MS. The segments that are found to be loosely compatible with each other are combined into preliminary track candidates. In order to create three-dimensional track candidates, the precision measurements are used to establish the coordinate in the bending plane, while the secondary coordinate is obtained using the trigger detectors. The muon trajectory is found using a global χ^2 fit, and any outlier hits are removed while additional

compatible hits are added. If the hits in the track candidate are updated, then the fit is performed once again. The track candidates then undergo an ambiguity resolving step, and a final track fit is performed and extrapolated back to the beam line taking into account the position of the interaction point and the estimated energy losses in the calorimeter.

The use of the full detector information for global muon reconstruction results in five distinct types of reconstructed muons:

- Combined (CB): an MS track can be matched to an ID track. A track fit is then performed utilizing the hits from both the ID and the MS, as well as the estimated energy loss in the calorimeter.

- Inside-out combined (IO): ID tracks are extrapolated to the MS and if at least three loosely aligned MS hits are found a combined track fit is performed using the ID track, the estimated energy loss in the calorimeter and the MS hits. As this method does not require an MS track, it serves to recover some efficiency in regions with poor MS coverage or in the case of low p_{T} muons.

- Muon spectrometer extrapolated (ME): in the case where an MS track cannot be matched to an ID track, the parameters of the MS track are extrapolated back to the beam line. This exploits the larger η coverage of the MS compared to the ID.

- Segment-tagged (ST): if an ID track extrapolated to the MS matches with at least one MS segment, then it produces a reconstructed muon with its parameters taken from the ID track fit. Similar to IO muons, ST muons recover efficiency in regions where the MS has lower coverage or in the case of low p_{T} muons.

- Calorimeter-tagged (CT): if an ID track extrapolated to the calorimeter is compatible with energy deposits that are consistent with a minimum ionizing particle, then the ID track is tagged as a muon. CT muons are designed to recover efficiency in regions where the MS is only partially instrumented ($|\eta| < 0.1$).

5.3 Jets

As was previously discussed in Section 2.2, jets are collimated showers of final state hadrons resulting from energetic parton production originating from the initial pp collision. In order to reconstruct jets within the ATLAS detector, low-level objects (e.g. tracks, clustered energy deposits in the calorimeter, or a combination thereof) are used as input to a jet finding algorithm. Truth particles from the MC simulation can separately be used as inputs. The particles that are used as inputs are known as jet constituents in the context of discussions regarding jets.

A common class of jet finding algorithms are known as sequential recombination jet algorithms, which proceed as follows for a list of input particles:

1. Two distance metrics are defined, the distance between each pair of particles i and j:

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2}, \tag{5.1}$$

where ΔR_{ij} is as defined in Eqn. 3.2, and the distance between each particle i and the beam (B):

$$d_{iB} = p_{\mathrm{T},i}^{2p}. \tag{5.2}$$

There are two model parameters present in these calculations, p and R. These

are chosen depending on the desired jet properties and will be discussed in detail below.

2. The smallest distance in the $\{d_{ij}, d_{iB}\}$ set is then determined. If d_{ij} is the smaller, then particles i and j are combined into one and remain in the particle list. If d_{iB} is the smaller, i is placed in the final jet collection and removed from the list of particles.

3. If particles remain in the list, return to step 1.

The R parameter is a radius parameter and describes the size of the jets in y-ϕ space. Meanwhile, the p parameter describes how the input objects are combined. Setting $p = 1$ gives what is known as the k_t algorithm [117], which combines smaller p_{T} constituents first. If p is instead set to 0, as it is in the Cambridge-Aachen algorithm [118,119], then only the angular separation between constituents is relevant to the recombination and close-by constituents are the first to be combined. The anti-k_t algorithm [120] sets $p = -1$, and is the most commonly used jet clustering algorithm in the ATLAS experiment. The anti-k_t algorithm combines higher p_{T} constituents first, resulting in very circular jets in y-ϕ with radius R.

Each algorithm has its own benefits and downsides [121,122], largely related to its ability to resolve so-called jet substructure, which studies the internal structure of the jet, and its susceptibility to pileup. The k_t algorithm, while it does a good job of resolving jet substructure, is extremely susceptible to pileup and other detector noise due to it clustering soft particles first. In general, this makes it unideal for use within ATLAS's high pileup environment. In addition, due to the clustering method the area of the jet in y-ϕ space varies widely. In contrast, the anti-k_t algorithm is minimally susceptible to pileup and noise and the jet area is very consistent due to its preference

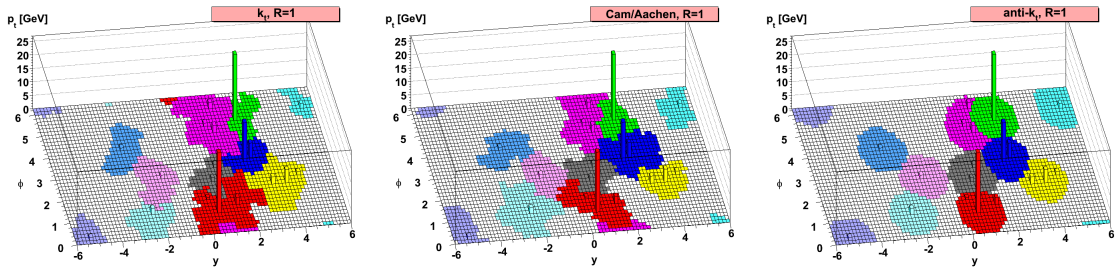


Figure 5.2: The same set of particles used as input to the various sequential recombination jet clustering algorithms [120]. The results of the k_t algorithm are shown on the left, Cambridge-Aachen in the centre and anti-k_t on the right. Particles clustered into the same jet are shown in the same colour.

to cluster hard particles first. However, it is poor for studying jet substructure. The Cambridge-Aachen algorithm is somewhat susceptible to pileup and area fluctuations due to only considering angular information, but it is well-suited for performing jet substructure studies. Figure 5.2 shows the same set of particles clustered using the three different algorithms.

As susceptibility to pileup is a chief concern, the anti-k_t algorithm is preferable for performing jet clustering within ATLAS. Furthermore, it provides improved correspondence between reconstructed and truth jets compared to the other algorithms. In order to perform dedicated substructure studies, a secondary algorithm (typically Cambridge-Aachen) is often used to recluster the jet constituents from an anti-k_t jet. For example, this procedure is applied in the jet substructure studies found in Refs. [123, 124].

Chapter 6

Measuring the differential cross section

The measurement made in this thesis will ultimately be expressed as a set of differential cross sections, which is a common way of expressing precision measurements in particle physics. As such, this chapter will first outline the definition of the differential cross section in Section 6.1, before discussing the way in which this measurement is traditionally made in Section 6.2. The traditional methods described will produce a measurement that is binned and will only probe one observable at a time, i.e. is one dimensional. The new method being explored in this thesis to perform the measurement, MultiFold, will then be described in Section 6.3. Unlike in the traditional case, this method aims to produce a measurement that is both unbinned and highly dimensional (24 observables are probed simultaneously).

6.1 Differential cross section definition

The method by which the production cross section is theoretically calculated was described in Chapter 2, and the process by which it can be determined from experimental data can now be discussed. The cross section is directly proportional to the rate at which a process occurs, and can be expressed as

$$\frac{d\nu}{dt} = \sigma L, \tag{6.1}$$

where $d\nu/dt$ represents the event rate and L is the instantaneous luminosity. Integrating with respect to time gives,

$$\sigma = \frac{\nu}{\mathcal{L}}\,. \tag{6.2}$$

In the case of the production cross section, ν would represent the total number of expected events occurring in nature over a certain period of data-taking.

This analysis is interested in the $pp \to Z(\to \mu\mu) + X$ process, where the $pp \to Z + X$ process has production cross section σ_Z, and the Z boson decays to two muons with branching ratio $\mathcal{B}_{Z\to\mu\mu}$. Thus, the production cross section for the $pp \to Z(\to \mu\mu) + X$ process will be,

$$\sigma_{Z\to\mu\mu} = \mathcal{B}_{Z\to\mu\mu}\,\sigma_Z = \frac{\nu_{Z\to\mu\mu}}{\mathcal{L}}, \tag{6.3}$$

where $\nu_{Z\to\mu\mu}$ represents the total number of expected $pp \to Z(\to \mu\mu) + X$ events produced in nature over the data-taking period.

As discussed in Section 3.2.5, the integrated luminosity is measured with a good deal of precision and thus the only unknown quantity in Eqn. 6.3 that is required to

determine the cross section is $\nu_{Z\to\mu\mu}$. If the ATLAS detector were able to identify and measure each $Z \to \mu\mu$ event with perfect accuracy, then the number of observed data events, n^{data}, would on average be ν. Unfortunately, this is not the case in reality and the effects of the ATLAS detector must be considered in order to properly determine the cross section. This process is generally known as unfolding within the particle physics community, but is synonymous with deconvolution. In general, an unfolding procedure must take into account the following issues.

1. Detector bias and resolution effects: the value of an observable measured by the detector is subject to bias and resolution effects, and thus the measured quantity will not be equal to its true value.

2. Efficiency and acceptance effects: an event in the detector may not be measured.

3. Fakes and out of fiducial events: a measured event may be the result of detector noise, or a truth event that does not fall within the fiducial volume defined by the analysis selection.

4. Background contributions: there may be events that originate from a non-signal process.

Put simply, the unfolding procedure will provide the best estimate of the number of $Z \to \mu\mu$ events that occur in nature based upon the observed data counts.

Firstly, not all $Z \to \mu\mu$ events that occur in nature will produce muons that can be detected, either because they fall outside of spatial acceptance, or because their p_{T} is too low. The probability that a $Z \to \mu\mu$ event satisfies these kinematic conditions is known as the acceptance, α. As the detector can only measure events that are within acceptance, one way to remedy this is to define a so-called fiducial phase space, Ω, that defines the region of phase space in which the detector is capable

of performing high quality measurements. This fiducial phase space will define the kinematic region of the final measurement, and the so-called fiducial cross section,

$$\sigma_{\text{fid}} = \alpha_{\text{fid}}\, \sigma_{Z\to\mu\mu}, \tag{6.4}$$

which is the cross section as measured in the fiducial phase space, will be the result as opposed to the production cross section. From this point forward, the subscript will thus be dropped, and following this section the fiducial cross section will simply be referred to as the cross section.

The definition of the fiducial cross section partially deals with effect (2) as listed above. The efficiency, ε (see also Eqn. 6.20), can then be defined as the probability that an event that enters into the fiducial phase space will also be detected. The $Z \to \mu\mu$ events that are detected enter into the so-called analysis phase space, A, which mimics the fiducial phase space at detector level. Similarly, effect (3) can be quantified by the introduction of the fiducial factor, f^{fid} (see also Eqn. 6.24), which dictates the probability that an event that enters into the analysis phase space also enters into the fiducial phase space. Finally, to address effect (4), if there exists contributions to n^{data} that originate from a background process with an estimated number of events equal to $\hat{\beta}$, then the best estimate of the fiducial cross section is,

$$\hat{\sigma} = \frac{\hat{f}^{\text{fid}}\,(n^{\text{data}} - \hat{\beta})}{\hat{\varepsilon}\,\mathcal{L}}. \tag{6.5}$$

The fiducial cross section can also be considered on a per-bin basis as a function of a certain observable of interest, x. In this case, the migration effects of point (1) also need to be considered. To study an observable, both the fiducial and analysis phase spaces are typically partitioned into N_{bins} analogous bins with respect to x,

where the bins in the analysis phase space are indexed with i and the bins in the fiducial phase space are indexed with j. Events that enter into bin i in A may not enter into the same bin in Ω. How this particular effect is incorporated depends largely on the individual unfolding procedure, and to simplify notation the best estimate of the fiducial cross section in a bin j in the distribution of a certain observable can be written as,

$$\hat{\sigma}_j = \frac{\mathcal{U}(\vec{n}^{\,\text{data}})_j}{\mathcal{L}}, \tag{6.6}$$

where $\vec{n}^{\,\text{data}}$ represents the observed data counts in each bin i and $\mathcal{U}(\vec{n}^{\,\text{data}})_j$ represents the number of events in bin j after the unfolding procedure is applied to the data events. This represents the best estimate of the number of $Z \to \mu\mu$ events occurring in nature in the jth bin of the distribution of observable x.

The differential cross section is then defined as the fiducial cross section expressed differentially with respect to observable x,

$$\frac{d\hat{\sigma}_j}{dx} = \frac{\mathcal{U}(\vec{n}^{\text{data}})_j}{\Delta x_j \mathcal{L}}, \tag{6.7}$$

where Δx_j represents the bin width of the jth bin. This is the form in which the results in this thesis will be presented in later chapters.

The major piece left to determine in Eqn. 6.7 is the unfolding procedure itself, which will be described in the following sections.

6.2 Traditional unfolding methods

This section serves to outline the traditional methods used to perform the unfolding procedure. This section will begin by outlining the basic mathematical background of

the unfolding procedure in Section 6.2.1. The practical quantities required from the MC samples to perform the unfolding will then be defined in Section 6.2.2. Finally, Section 6.2.3 will outline the methodology for iterative Bayesian unfolding (IBU). As IBU is the most common unfolding method used within the ATLAS Collaboration, it will be used as a comparison to the MultiFold method in the following chapters.

6.2.1 Mathematical formulation of unfolding

This section will focus on developing the general mathematical formulation of the unfolding procedure [125]. Consider an observable whose detector (or reconstructed) level value is x^{reco} and whose particle (or truth) level value is x^{truth}. If the detector measuring the data were able to perform a perfect measurement, then the value of x^{reco} would be equivalent to x^{truth}. Unfortunately, due to the effects discussed in the previous section this is not the case in practice and thus additional steps must be taken in order to obtain a measurement at the particle level.

The unfolding problem may be framed as determining the probability distribution function (pdf) for the true values, x^{truth}, given that x^{reco} has been measured and that x^{truth} and x^{reco} are related by some function. Thus, the relation between the pdfs for x^{reco}, $p_{\text{r}}(x^{\text{reco}})$, and x^{truth}, $p_{\text{t}}(x^{\text{truth}})$, may be written as

$$p_{\text{r}}(x^{\text{reco}}) = \int R(x^{\text{reco}}|x^{\text{truth}})\, p_{\text{t}}(x^{\text{truth}}), \tag{6.8}$$

where $R(x^{\text{reco}}|x^{\text{truth}})$ represents the response function. Given the effects of the detector, it describes the probability to observe x^{reco} given that the truth value was x^{truth}.

The standard way to approach finding a solution to this problem is to construct histograms for x^{reco} and x^{truth}, where the number of expected events in a bin i of the

reconstructed level distribution can be written as

$$\nu_i^r = \nu_{\text{tot}}^r \int_{\text{bin } i} p_{\text{r}}(x^{\text{reco}})\, dx^{\text{reco}}, \tag{6.9}$$

where ν_{tot}^r is the total number of detected events. Similarly, the number of events in a bin j of the truth level distribution can be written as,

$$\nu_j^t = \nu_{\text{tot}}^t \int_{\text{bin } j} p_{\text{t}}(x^{\text{truth}})\, dx^{\text{truth}}, \tag{6.10}$$

where ν_{tot}^t is the total number of truth events. For the remainder of this discussion, it will be assumed that each distribution is identically binned into a total of N_{bins} bins (although this need not necessarily be the case in general). Equation 6.8 may then be expressed as,

$$f_i^{\text{fid}}(\nu_i^r - \beta_i) = \sum_{j=1}^{N_{\text{bins}}} R_{ij}\, \nu_j^t, \quad \text{where } i = 1, ..., N_{\text{bins}}. \tag{6.11}$$

Here, R represents the response matrix, which is a discretized version of the response function and dictates the probability that an event is detected in bin i given that the event is in bin j at truth level, and f_i^{fid} and β_i are the fiducial factor and estimated background contribution in bin i. For the sake of brevity, f^{fid} will be ignored in the following discussion.

Equation 6.11 can be generalized to

$$\boldsymbol{\nu}^r = R\boldsymbol{\nu}^t + \boldsymbol{\beta}, \tag{6.12}$$

where $\boldsymbol{\nu}^r$ and $\boldsymbol{\nu}^t$ are simply the number of detector and truth level events in each bin represented as a column vector. Similarly, $\boldsymbol{\beta}$ is a column vector of the number

of background events in each bin. Typically, the quantities in the above equation are obtained using MC samples designed to precisely model the detector level and truth level distributions.

The ultimate goal of the unfolding procedure is to determine the estimators for $\boldsymbol{\nu}^t$, $\hat{\boldsymbol{\nu}}^{\boldsymbol{t}}$, given that a certain number of data events were observed. In theory, the most straightforward way one could envision accomplishing this task is to simply invert Eqn. 6.12,

$$\boldsymbol{\nu}^t = R^{-1}(\boldsymbol{\nu}^r - \boldsymbol{\beta}). \tag{6.13}$$

Determining the estimators in this manner corresponds to deriving them using maximum likelihood estimation.

When a measurement is performed a set of values is obtained for the number of observed events, $\boldsymbol{n} = (n_1, ..., n_N)^{\mathrm{T}}$. In general, the n_i values may be considered random Poisson variables with a mean (or expectation value) of ν_i^r. Thus, writing out the log-likelihood function gives,

$$\log L(\boldsymbol{\nu}^t) = \sum_{i=1}^{N_{\text{bins}}} \log(\text{Poisson}(n_i; \nu_i^r)) = \sum_{i=1}^{N_{\text{bins}}} \log\left(\frac{(\nu_i^r)^{n_i} e^{-\nu_i^r}}{n_i!}\right), \tag{6.14}$$

where log is the natural logarithm. Taking the derivative with respect to $\boldsymbol{\nu}^t$ and setting it equal to zero results in the estimators for $\boldsymbol{\nu}^r$ and $\boldsymbol{\nu}^t$ being,

$$\hat{\boldsymbol{\nu}}^r = \boldsymbol{n}, \tag{6.15}$$

and

$$\hat{\boldsymbol{\nu}}^t = R^{-1}(\boldsymbol{n} - \hat{\boldsymbol{\beta}}). \tag{6.16}$$

Unfortunately, this method is not generally viable. The response matrix R, which is estimated from the MC sample, has in many cases substantial off-diagonal elements, and performing the unfolding by matrix inversion can thus result in large statistical fluctuations. This is due to the fact that any small differences between $\boldsymbol{n}$ and $\boldsymbol{\nu}^r$, even if they are only statistical in nature, will be interpreted as a real effect due to the detector rather than simple statistical variations, and they will be propagated to the final result for $\hat{\boldsymbol{\nu}}^t$ [126]. Due to this limitation, this method is not very commonly used except in very rare cases where the off-diagonal elements of the response matrix are minimal enough that the variance falls within an acceptable range [127]. Thus, other analytic methods have been explored in order to unfold experimental data that reduce localized statistical fluctuations using a process known as regularization.

The basic principle behind regularization involves introducing an element of smoothness to the estimators for $\boldsymbol{\nu}^t$. Imposing this smoothness serves to reduce the variance on the estimators, while ideally leaving the underlying data distribution only minimally affected. Different approaches to regularized unfolding provide various ways of imposing this smoothness, and it is usually quantified by a so-called regularization parameter that is unique to that unfolding procedure. One possible and very common method will be discussed in Section 6.2.3, but examples of other methods include singular value decomposition (SVD) [128] using Tikhonov regularization [129, 130], the iterative dynamically stabilized method (IDS) [131, 132], and full Bayesian unfolding (FBU) [133].

6.2.2 Relevant MC quantities

Before explaining any further unfolding methods in detail, some relevant quantities from the MC samples must be defined. Throughout this section, ν^t and ν^r will represent the number of truth and reconstructed events from the MC sample, respectively, and can be written as,

$$\nu^t_{\text{tot}} = \sum_{k=1}^{N_{\text{truth}}} w_k^{\text{truth}}, \quad \text{and} \tag{6.17}$$

$$\nu^r_{\text{tot}} = \sum_{m=1}^{N_{\text{reco}}} w_m^{\text{reco}}, \tag{6.18}$$

for the total number of each, respectively, where $w^{\text{truth(reco)}}$ represents the MC event weights for the truth (reconstructed) level events as described in Section 4.2.6.

Now, recalling from the previous section that the response matrix dictates the probability that a quantity, x^{reco}, is detected in bin i, given that the truth level quantity, x^{truth}, is in bin j, it may be calculated practically using

$$R_{ij} = P(x_i^{\text{reco}}|x_j^{\text{truth}}) = \frac{P(x_i^{\text{reco}} \cap x_j^{\text{truth}})}{P(x_j^{\text{truth}})} = \frac{\nu_{ij}^{t,\text{truth}\wedge\text{reco}}}{\nu_j^t}, \tag{6.19}$$

where $\nu_{ij}^{t,\text{truth}\wedge\text{reco}}$ represents the number of truth events in bin j reconstructed in bin i and ν_j^t represents the total number of truth events in bin j. Here, x_i^{reco} and x_j^{truth} have been used as shorthand notation to represent $x^{\text{reco}} \in A_i$ and $x^{\text{truth}} \in \Omega_j$, respectively, where A_i represents the analysis phase space in bin i and Ω_j represents the fiducial phase space in bin j.

Another important quantity that can be determined from the response matrix is the efficiency, which was previously discussed in Section 6.1. It represents the probability that an event with a truth level quantity x_j^{truth} enters into the analysis

phase space in any bin,

$$\varepsilon_j = \sum_i R_{ij} = \sum_i P(x_i^{\text{reco}}|x_j^{\text{truth}}) = \frac{\nu_j^{t,\text{truth}\wedge\text{reco}}}{\nu_j^t}, \tag{6.20}$$

where $\nu_j^{t,\text{truth}\wedge\text{reco}}$ represents the number of truth events in bin j that are reconstructed in any bin. The efficiency can be used to relate another object known as the migration matrix to the response matrix. The migration matrix may be defined as,

$$M_{ij} = \frac{\nu_{ij}^{t,\text{truth}\wedge\text{reco}}}{\nu_j^{t,\text{truth}\wedge\text{reco}}}, \tag{6.21}$$

such that,

$$\sum_{i=1}^{N} M_{ij} = 1. \tag{6.22}$$

It is thus related to the response matrix by

$$M_{ij} = \frac{R_{ij}}{\varepsilon_j}. \tag{6.23}$$

There are two additional quantities related to the detector level events, the first of which is the fiducial factor (also discussed in brief in Section 6.1). The fiducial factor represents the probability that an event with $x_{\text{i}}^{\text{reco}}$ enters into the fiducial phase space in any bin. In other words, the fraction of reconstructed events that are not the result of either fakes or truth events that are outside of the fiducial volume defined by the analysis. It can be written as

$$f_i^{\text{fid}} = \frac{\nu_i^{r,\text{truth}\wedge\text{reco}}}{\nu_i^r}, \tag{6.24}$$

where $\nu_i^{r,\text{truth}\wedge\text{reco}}$ is the number of detector level events in bin i that appear at truth level and ν_i^r represents the total number of detector level events in bin i. Finally, the purity is the probability that an event with x_i^{reco} appears in the same bin i in the fiducial phase space (i.e. $j = i$),

$$\mathcal{P}_i = \frac{\nu_{ii}^{r,\text{truth}\wedge\text{reco}}}{\nu_i^r}. \tag{6.25}$$

6.2.3 Iterative Bayesian unfolding

Iterative Bayesian unfolding (IBU) [134, 135] is the most commonly used unfolding method within the ATLAS experiment. In this analysis, it will serve as the "standard" result to which the main MultiFold result will be compared. This section will thus discuss the methodology behind IBU.

IBU works by incorporating Bayesian statistics into the unfolding. Recall that the response matrix for a specific observable can be written as the conditional probability,

$$R_{ij} = P(x_i^{\text{reco}}|x_j^{\text{truth}}), \tag{6.26}$$

where x_i^{reco} and x_j^{truth} are as previously defined. The number of unfolded events can then be estimated using the conditional probability $P(x_j^{\text{truth}}|x_i^{\text{reco}})$. Analogously to Eqn. 6.11, one can write,

$$\nu_j^t = \sum_i P(x_j^{\text{truth}}|x_i^{\text{reco}})\nu_i^r, \tag{6.27}$$

where efficiency effects, contributions from fakes and background contributions are

temporarily neglected. Using Bayes' theorem, $P(x_j^{\text{truth}}|x_i^{\text{reco}})$ may be written as

$$P(x_j^{\text{truth}}|x_i^{\text{reco}}) = \frac{P(x_i^{\text{reco}}|x_j^{\text{truth}})P(x_j^{\text{truth}})}{P(x_i^{\text{reco}})} = \frac{P(x_i^{\text{reco}}|x_j^{\text{truth}})P(x_j^{\text{truth}})}{\sum_j P(x_i^{\text{reco}}|x_j^{\text{truth}})P(x_j^{\text{truth}})}, \tag{6.28}$$

and so Eqn. 6.27 becomes

$$\nu_j^t = \sum_i \frac{P(x_i^{\text{reco}}|x_j^{\text{truth}})P(x_j^{\text{truth}})}{\sum_j P(x_i^{\text{reco}}|x_i^{\text{truth}})P(x_j^{\text{truth}})}\nu_i^r \tag{6.29}$$

$$\nu_j^t = \sum_i \frac{R_{ij}P(x_j^{\text{truth}})}{\sum_j R_{ij}P(x_j^{\text{truth}})}\nu_i^r. \tag{6.30}$$

The iterative part of this process comes from the estimation of the prior, $P(x_j^{\text{truth}})$. On the first iteration, $P(x_j^{\text{truth}})$ is defined using the MC truth distribution, i.e. $P(x_j^{\text{truth}}) = \nu_j^t/\nu_{\text{tot}}^t$. It is subsequently updated using the result for ν_j^t from the current iteration, thus

$$\nu_j^{t,(k+1)} = \sum_i \frac{R_{ij}P^{(k)}(x_j^{\text{truth}})}{\sum_j R_{ij}P^{(k)}(x_j^{\text{truth}})}x_i^{\text{reco}}, \tag{6.31}$$

where k represents the current iteration number.

In order to apply this method to a measurement using data, which will again be represented as $\boldsymbol{n}$, the effects that were neglected eariler must also be included. Incorporating them gives,

$$\mathcal{U}(\boldsymbol{n})_j = \frac{1}{\hat{\varepsilon}_j}\sum_i \frac{\hat{R}_{ij}P^{(k)}(x_j^{\text{truth}})}{\sum_j \hat{R}_{ij}P^{(k)}(x_j^{\text{truth}})}\hat{f}_i^{\text{fid}}(n_i - \hat{\beta}_i). \tag{6.32}$$

$\mathcal{U}(\boldsymbol{n})_j$ represents the unfolded result in bin j after k iterations, which gives the best estimate of the particle level data in bin j. Notably, this is equivalent to the value specified in Eqn. 6.7 in relation to the differential cross section measurement.

The number of iterations serves as the regularization parameter for IBU. Increasing the number of iterations increases the variance of the result while decreasing the bias. In fact, after a very large number of iterations IBU will approach the matrix inversion solution and thus become unregularized. The number of iterations must be optimized for each analysis, but in most cases very few iterations are needed to obtain a sufficient bias reduction.

6.3 The MultiFold method

The traditional methods for measuring the differential cross section discussed in the previous section have some key drawbacks, as listed below.

1. The data must be binned into histograms and the exact list of observables determined prior to unfolding.

2. Only a small number of observables may be unfolded simultaneously (in the last chapter, the focus was on unfolding observables one at a time but this could potentially be done in two or three dimensions).

3. The response matrix is only parameterized as a function of the observables included in the unfolding, and thus may neglect effects from so-called hidden observables that are not included.

The MultiFold method [10] aims to address these drawbacks by performing an unbinned, highly dimensional unfolding that is made feasible by the utilization of neural networks. By obtaining a result that is highly dimensional and unbinned, its flexibility is greatly increased. In addition to the ability to freely bin the results, new observables may be calculated from the results or phase space reductions may be applied. When using the traditional methods, adding a new observable or changing

a phase space definition would require the full analysis to be rerun. As typical measurements occur on the timescale of years, it is cumbersome to incorporate such a change at a later date.

This section will first outline some general background on neural networks in Section 6.3.1 before explaining the methodology of the MultiFold procedure in Section 6.3.2.

6.3.1 Neural networks

This section aims to provide a basic introduction to deep neural networks, which are the type of machine learning algorithm used in this thesis. This will, as such, be a very narrow introduction to a very rich field of study. The following discussion draws largely from Refs. [136–138], which provide a much more in-depth overview of machine learning topics. Put very broadly, a machine learning (ML) algorithm is one that is able to learn from a set of input data. The ML algorithm produces a function, f, that is designed to predict outputs $\boldsymbol{y}$ from inputs $\boldsymbol{x}$. In the cases discussed in this thesis, this is done by examining a set of labelled training data[1]. Two very common tasks that are often performed are classification and regression tasks. In classification tasks, the algorithm predicts which class an input belongs to and the output y is a value mapping the input $\boldsymbol{x}$ to a specific category. Regression tasks, meanwhile, predict a numerical value given the inputs $\boldsymbol{x}$.

ML algorithms are especially useful for approximating solutions to problems that are too complex to be reasonably solved using a man-made program. They are also, in theory, generalizable and once trained can be applied to new data without degradation of performance. Over the past several years as the technical implemen-

[1]This is known as supervised learning. Input data can also be provided without labels and the ML algorithm asked to learn the features of the dataset. This is known as unsupervised learning.

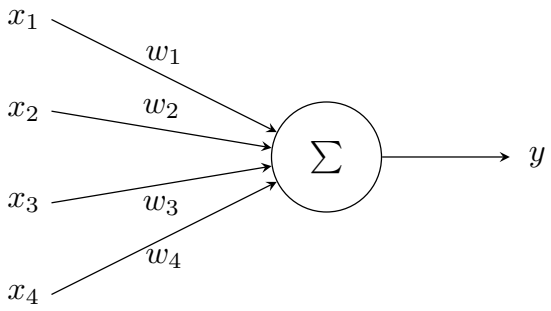


Figure 6.1: A schematic diagram of a single neuron, where the input signal vector $\boldsymbol{x}$ has four components.

tation of NNs has become more accessible, the usefulness of NNs within the particle physics community has been explored with some depth and has found applications in a great number of problems (an ever-growing list is maintained in Refs. [139, 140]). As a few specific examples, ML has seen use in classification tasks in order to differentiate between signal and background events, as well as in object reconstruction where regression tasks can provide improved predictions of the object properties.

The most basic structure in a neural network is the eponymous neuron. The basic principle behind a neuron is that given a vector of input signals, $\boldsymbol{x}$, an output signal y may be calculated by multiplying the inputs by a corresponding vector of weights, $\boldsymbol{w}$, and introducing a possible bias term b. In equation form,

$$y = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b. \tag{6.33}$$

A basic schematic of an individual neuron is also shown in Figure 6.1. These neurons are often grouped into so-called layers, where an input vector of signals $\boldsymbol{x}$ is now multiplied by a weight matrix $\boldsymbol{W}$ to get an output vector $\boldsymbol{y}$. Or,

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}. \tag{6.34}$$

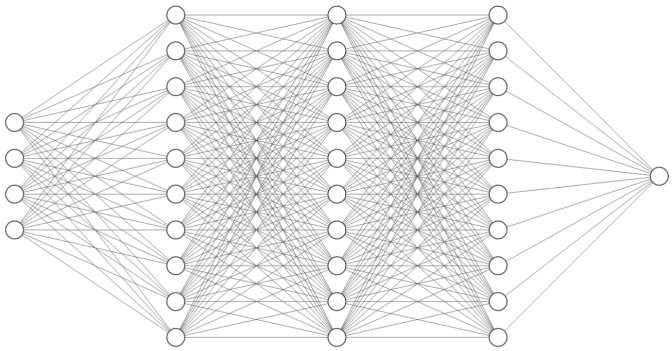


Figure 6.2: A schematic of a DNN composed of an input layer with four features, three hidden layers with ten nodes (neurons) each and an output layer.

More specifically, this kind of layer is known as a linear layer. Stacking a set of linear layers would be equivalent to simply multiplying together the weight matrices from each layer. To extract more information and introduce an element of non-linearity to the function determination, a non-linear activation function may be applied. These involve applying a transformation in the form of a non-linear function, ϕ, to each layer,

$$\boldsymbol{y} = \phi(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}). \tag{6.35}$$

This combination of a linear layer and an activation function is what is known as a hidden layer. If the network architecture includes an input layer (simply the inputs $\boldsymbol{x}$), one hidden layer and an output layer (the outputs $\boldsymbol{y}$) then this is what is known as an artificial neural network. The basis for a deep neural network (DNN) is combining a series of fully-connected hidden layers, where fully-connected simply means that the output of each neuron (also known as a node) in the layer is fed into all the nodes of the next layer. A visualization of this kind of architecture is shown in Figure 6.2, which shows a schematic of a DNN with one input layer, three hidden layers, and one output layer.

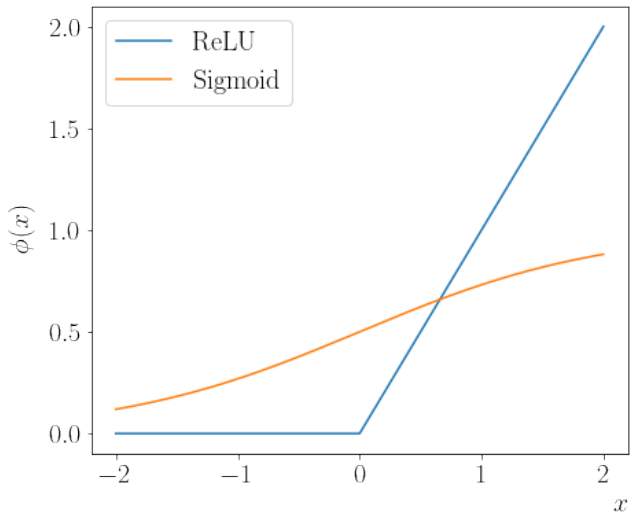


Figure 6.3: The two activation functions that are used in this thesis. The ReLU function is shown in blue, and the sigmoid in orange.

The two activation functions present in this thesis are the Rectified Linear Unit (ReLU) and the sigmoid (σ), which are written as

$$\phi(x) = \text{ReLU}(x) = \max\{0, x\}, \text{ and} \tag{6.36}$$

$$\phi(x) = \sigma(x) = \frac{1}{1 + e^{-x}}, \tag{6.37}$$

respectively. The range of the ReLU function is $[0, +\infty)$, while the range for the sigmoid function is $(0, 1)$. This makes the sigmoid function useful for classifier tasks. Meanwhile, the ReLU function is useful for the optimization of the network, and so is frequently used in layers before the output.

The above outlines how a function f is produced using a DNN, however this output must now be evaluated to determine its performance. The performance of a network is quantified using a so-called loss function, $\mathcal{L}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents the parameters of the model (the weights and the bias in the context of the above

discusssion). The basic principle behind a loss function is that if the network output is far from the target, the loss function will be larger. Thus, the best estimate for the parameters can be obtained using,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{\theta}). \tag{6.38}$$

The loss function can have a variety of forms, however one common choice is based on the principle of maximum likelihood estimation. Typically, the model determined from NN training can be represented by a pdf $p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$. Thus, the principle of maximum likelihood dictates that the best estimate for the parameters $\boldsymbol{\theta}$ is given by,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta}), \tag{6.39}$$

where N represents the number of examples in the input training data. This is equivalent to calculating the minimum of the negative log likelihood,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left(\frac{1}{N} \sum_{i=1}^{N} [-\log p(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta})] \right), \tag{6.40}$$

where without loss of generality this has been scaled by $1/N$. More compactly, this can be expressed in terms of the expectation value,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, E[-\log p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})]. \tag{6.41}$$

Thus, the loss function that corresponds to the maximum likelihood estimate is,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} [-\log p(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\theta})]. \tag{6.42}$$

The exact form of this loss function depends largely on the task the network is being assigned to perform. As a specific example, consider the case where a binary classifier is trained with one output labelled $y = 1$ and the other $y = 0$. For an input random variable $\boldsymbol{X}$, the probability that it will belong to the category labelled $y = 1$ is q and the probability that it belongs to the $y = 0$ category is $1 - q$ (i.e. the Bernoulli distribution). If the classifier is equipped with a sigmoid activation function on the output layer, then $f(\boldsymbol{X}) = q$. This leads to a probability mass function equal to

$$\text{pmf}(y; q) = \begin{cases} q & \text{if } y = 1 \\ 1 - q & \text{if } y = 0 \end{cases} = q^y (1 - q)^{1-y}. \tag{6.43}$$

Taking the log results in,

$$\log(q^y (1 - q)^{1-y}) = y \log(q) + (1 - y) \log(1 - q), \tag{6.44}$$

and inserting this into Eqn. 6.42 gives the following loss function,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} [-y_i \log f(\boldsymbol{x}_i; \boldsymbol{\theta}) - (1 - y_i) \log(1 - f(\boldsymbol{x}_i; \boldsymbol{\theta}))]. \tag{6.45}$$

This is known as the binary cross entropy loss function.

The discussion up until now has described the DNN architecture and how it uses a series of inputs $\boldsymbol{x}$ to determine a model $f(\boldsymbol{x}; \boldsymbol{\theta})$ that gives an output $\boldsymbol{y}$, where the accuracy of the model is quantified by a loss function that examines the agreement between the outputs from the model and the labels assigned to each example in the training data. Now, the way in which the network learns must be discussed. This process is known as optimization, the main idea behind it being to tune the parameters such that the loss function moves towards its minimum value. The basic

technique used to determine how the parameters should be changed is known as gradient descent. This involves evaluating the gradient of the loss function with respect to the parameters, $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})$, using a procedure known as backpropagation, where the gradient is determined by allowing information from the loss function to flow back through the network. The gradient will give the direction of fastest increase of the loss function, thus by taking the negative one can find the direction of steepest descent. Taking a step in this direction in parameter space results in the loss function getting closer to its minimum value. Thus, the parameters are updated using

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - \gamma \nabla_{\boldsymbol{\theta}}\mathcal{L}, \tag{6.46}$$

where γ is the learning rate or the step size. This is optimized depending on the network.

There is one major drawback to the gradient descent method as it is written above, that being that the calculation of the gradient for each event in the training set is computationally expensive. In order to mitigate this, a modified version of this method known as stochastic gradient descent is frequently employed. Since the loss function is an expectation value, its gradient is itself an expectation value. Thus, the idea behind stochastic gradient descent is that the gradient may be approximated by calculating it for only a small subset of samples known as a batch. The parameters are then updated according to Eqn. 6.46 using the gradient calculated using the batch.

The number of times the parameters are updated by passing through the training set is known as the number of epochs. In order to prevent overfitting, where the model becomes very good at predicting the outputs of the training set but does not generalize well, early stopping is one method that is employed. The resulting function, f, determined using the training set is evaluated on a separate validation

set. If the loss function after applying the model to the validation set is reduced from the previous epoch, then the model parameters are saved. If the validation set error does not improve for a specified number of epochs, then the algorithm is terminated and the best parameters are saved.

6.3.2 Methodology

The shortcomings of traditional unfolding methods outlined in the introduction of this section can be remedied through the usage of neural networks. The MultiFold algorithm is underpinned by the principle of using NNs to perform a smooth, highly dimensional reweighting of input MC samples. This reweighting procedure will first be outlined before discussing how it is used in the MultiFold algorithm.

Neural networks provide a means by which to reweight samples in a highly-dimensional and unbinned manner via a process known as density ratio estimation [141]. This forms the basic principle behind the MultiFold method. To frame this reweighting problem, consider two datasets, A and B, that describe the same phase space Ω. The goal will be to reweight dataset A such that it better matches dataset B. If the datasets are described by the probability density functions $p_{\mathrm{A}}(\boldsymbol{x})$ and $p_{\mathrm{B}}(\boldsymbol{x})$, where $\boldsymbol{x} \in \Omega$ represents an event, then the per-event weight to reweight sample A to sample B could be written as

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{B}}(\boldsymbol{x})}{p_{\mathrm{A}}(\boldsymbol{x})}, \tag{6.47}$$

which is also known as the likelihood ratio, $p(\boldsymbol{x}|\mathrm{B})/p(\boldsymbol{x}|\mathrm{A})$. This provides a means by which to indirectly estimate $p_{\mathrm{B}}(\boldsymbol{x})$. The events in each sample may in general be weighted, but for simplicity this will be neglected for the remainder of this discussion. Looking at this equation, the natural thought would perhaps be to simply determine

$p_{\mathrm{B}}(\boldsymbol{x})$ and $p_{\mathrm{A}}(\boldsymbol{x})$ separately and then take the ratio. Unfortunately, this is does not necessarily represent the best approach. For one, estimating the probability density function directly is a much more difficult problem than estimating the ratio. It relies to a larger degree on the training data being representative of the full sample, and constitutes a more complicated network architecture than in the case of estimating the ratio. This is also the reason for not directly estimating $p_{\mathrm{B}}(\boldsymbol{x})$. Furthermore, due to the division being performed, any errors in the estimation of the pdfs may be artificially inflated. This is particularly problematic for highly-dimensional problems like the one in this thesis, where the density values are often very low. Furthermore, estimating the individual pdfs without considering the fact that their ratio will be taken may result in a suboptimal model for the ratio. The NN models used to estimate the pdf would be optimized specifically for that pdf, but this will not generally translate to an optimized model for their ratio. It is, therefore, simpler and often more accurate to estimate the ratio directly.

One possible method to estimate the density ratio involves the use of Bayesian statistics to rewrite the ratio as,

$$r(\boldsymbol{x}) = \frac{p_{\mathrm{B}}(\boldsymbol{x})}{p_{\mathrm{A}}(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\mathrm{B})}{p(\boldsymbol{x}|\mathrm{A})} = \frac{p(\mathrm{B}|\boldsymbol{x})P(\boldsymbol{x})}{P(\mathrm{B})} \frac{P(\mathrm{A})}{p(\mathrm{A}|\boldsymbol{x})P(\boldsymbol{x})} \tag{6.48}$$

$$r(\boldsymbol{x}) = \frac{P(\mathrm{A})}{P(\mathrm{B})} \frac{p(\mathrm{B}|\boldsymbol{x})}{p(\mathrm{A}|\boldsymbol{x})}. \tag{6.49}$$

The $P(\mathrm{A})/P(\mathrm{B})$ term is simply proportional to the number of events in each sample, and may be approximated as $P(\mathrm{A})/P(\mathrm{B}) \approx N_{\mathrm{A}}/N_{\mathrm{B}}$ in the case where all events have a weight of one. It will be neglected for the remainder of the discussion, as it will be incorporated in the normalization factors at a later stage. The estimator for the

density ratio may then be expressed as,

$$\hat{r}(\boldsymbol{x}) = \frac{\hat{p}(\mathrm{B}|\boldsymbol{x})}{\hat{p}(\mathrm{A}|\boldsymbol{x})}. \tag{6.50}$$

In order to calculate this practically, a classifier, $f(\boldsymbol{x})$, can be trained to differentiate between the two datasets [138, 142–144]. Let $\boldsymbol{X}$ and $Y \in \{0, 1\}$ be a random set of input features and a random variable output labelling one of the two datasets, respectively. As described in Section 6.3.1, the optimal form of $f(\boldsymbol{x})$ can be determined by minimizing the loss function,

$$f(\boldsymbol{x}) = \underset{f'}{\operatorname{argmin}}\, E_{\boldsymbol{X},Y}[\mathcal{L}(f'(\boldsymbol{X}), Y)], \tag{6.51}$$

where the expectation value is taken over the joint probability distribution and f' represents a specific trained classifier. Without loss of generality, this can instead be written as,

$$f(\boldsymbol{x}) = \underset{f'}{\operatorname{argmin}}\, E_{\boldsymbol{X}} \left(E_{Y|\boldsymbol{X}}[\mathcal{L}(f'(\boldsymbol{X}), Y)|\boldsymbol{X}] \right), \tag{6.52}$$

and thus it is sufficient to minimize,

$$f(\boldsymbol{x}) = \underset{f'}{\operatorname{argmin}}\, E_{Y|\boldsymbol{X}=\boldsymbol{x}}[\mathcal{L}(f'(\boldsymbol{x}), Y)|\boldsymbol{X} = \boldsymbol{x}], \tag{6.53}$$

for all $\boldsymbol{x}$. Returning to the binary cross entropy loss function defined in Eqn. 6.45, and rewriting it with the quantities defined in this discussion,

$$\mathcal{L}(f'(\boldsymbol{x}), Y) = -Y \log\left(f'(\boldsymbol{x})\right) - (1 - Y) \log(1 - f'(\boldsymbol{x})). \tag{6.54}$$

Substituting this into Eqn. 6.53 gives the following to be maximized,

$$f(\boldsymbol{x}) = \underset{f'}{\operatorname{argmax}}\, E[Y \log(f'(\boldsymbol{x})) + (1-Y)\log(1-f'(\boldsymbol{x}))|\boldsymbol{X}=\boldsymbol{x}] \tag{6.55}$$

$$= \underset{f'}{\operatorname{argmax}}\, E[Y|\boldsymbol{X}=\boldsymbol{x}]\log(f'(\boldsymbol{x})) + (1-E[Y|\boldsymbol{X}=\boldsymbol{x}])\log(1-f'(\boldsymbol{x})). \tag{6.56}$$

The maximum is achieved by setting $f(\boldsymbol{x}) = E[Y|\boldsymbol{X}=\boldsymbol{x}]$, which can be determined by taking the derivative with respect to $f'(\boldsymbol{x})$ and setting it equal to 0.

If the output layer of the classifier is equipped with an activation function with the appropriate asymptotic behaviour and an output between 0 and 1, like the sigmoid function that was discussed in the previous section, then $f(\boldsymbol{x})$ will approach $E[Y|\boldsymbol{X}=\boldsymbol{x}]$ and $1-f(\boldsymbol{x})$ will approach $E[1-Y|\boldsymbol{X}=\boldsymbol{x}]$. Thus,

$$\frac{f(\boldsymbol{x})}{1-f(\boldsymbol{x})} \approx \frac{E[Y|\boldsymbol{X}=\boldsymbol{x}]}{E[1-Y|\boldsymbol{X}=\boldsymbol{x}]} \tag{6.57}$$

$$= \frac{p(Y=1|\boldsymbol{X}=\boldsymbol{x})}{p(Y=0|\boldsymbol{X}=\boldsymbol{x})} \tag{6.58}$$

$$= \frac{p(\boldsymbol{X}|Y=1)}{p(\boldsymbol{X}|Y=0)}\frac{P(Y=1)}{P(Y=0)} \tag{6.59}$$

$$\propto \frac{p(\boldsymbol{X}|Y=1)}{p(\boldsymbol{X}|Y=0)}, \tag{6.60}$$

where once again on the last line the overall factor related to the sample size has been neglected as it was in Eqn. 6.48. Therefore, by expressing the result of the classifier in this manner it will provide an estimator of the density ratio, $\hat{r}(\boldsymbol{x})$. In the context of dataset A and B, this gives,

$$\hat{r}(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{1-f(\boldsymbol{x})} \approx \frac{p(\boldsymbol{x}|\mathrm{B})}{p(\boldsymbol{x}|\mathrm{A})} = \frac{p_{\mathrm{B}}(\boldsymbol{x})}{p_{\mathrm{A}}(\boldsymbol{x})}. \tag{6.61}$$

Thus, dataset A can be reweighted such that it matches dataset B by applying a

weight to each event in A in the form of the above equation. In order to obtain better statistics and a better estimate of $p_{\mathrm{B}}(\boldsymbol{x})$, one would simply need to generate more events in dataset A.

A frequent task in experimental particle physics is to examine the differences between the data sample and the reconstructed MC sample (as in Section 7.5 for example). To improve the MC prediction, one may wish to reweight the reconstructed MC sample to match the data. In order to do this, a classifier would be trained using a set of features $\boldsymbol{x}$, with a label of $y = 0$ corresponding to the reconstructed MC, and a label of $y = 1$ corresponding to the data. The events in the reconstructed MC sample would then be reweighted such that the new event weight for an event i in the sample would be,

$$w_i^{\mathrm{rew}} = \frac{f(\boldsymbol{x}_i)}{1 - f(\boldsymbol{x}_i)} w_i, \tag{6.62}$$

where w_i is the original event weight. This would result in MC events in a region of phase space more densely populated by data events receiving a higher weight, and correspondingly MC events in a region of phase space less densely populated by data events being down-weighted.

The main idea behind the MultiFold algorithm is to iteratively perform reweightings in this fashion to produce the unfolded result. The inputs to the NN in the case in the last paragraph will be a set of features (the observables) and the corresponding event weights from both the data sample, with $(w_i^{\mathrm{data}} = 1, \vec{x}_i^{\,\mathrm{data}})$, and the detector level MC sample, with $(w_j^{\mathrm{reco}}, \vec{x}_j^{\,\mathrm{reco}})$. In order to obtain the unfolded data at truth level, the weight necessary to reweight each MC event at detector level is applied at the truth level, where the MC event j has truth level weight and features $(w_j^{\mathrm{truth}}, \vec{x}_j^{\,\mathrm{truth}})$. A new reweighting function is then determined that reweights the

truth level MC sample with the original MC weights applied to the truth level MC sample with the detector level reweighting applied. The final unfolded result will thus be entirely determined by applying this new truth level reweighting function to the original truth level MC sample. Ideally, the result of this reweighting will begin to approach the best estimate of the data at truth level. This will be discussed in more detail below.

The MultiFold algorithm is a part of a larger suite of algorithms, each with slightly different inputs and outputs, which are collectively known as the OmniFold algorithm. The OmniFold algorithm does not specifically require the use of neural networks to perform the reweighting, however as NNs are very well-suited to the task of determining reweighting functions they are an obvious choice. The three flavours of the OmniFold algorithm are listed below.

- UniFold: essentially an unbinned version of IBU. One observable with the corresponding event weight is used as input to the algorithm. The output is then a set of MC events with the truth level observable plus a modified event weight.

- MultiFold: takes a list of observables as input along with the corresponding event weight. The output is then the list of observables plus a modified event weight.

- OmniFold: takes as input all of the particle four vectors in an event and the corresponding event weight. The output is then a variable-length list of four vectors plus a modified event weight.

This analysis will focus specifically on the MultiFold version of the algorithm with 24 input observables. These observables are discussed in more specific detail in Section 7.1.

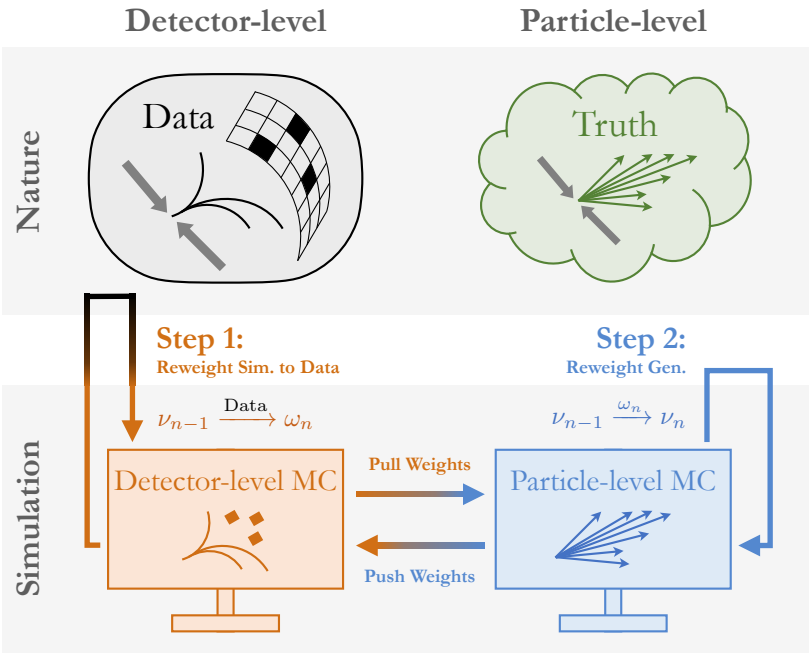


Figure 6.4: A schematic representation of the OmniFold algorithm.

The MultiFold algorithm proceeds as shown schematically in Figure 6.4. The starting inputs for the algorithm are the data sample, with

$$\mathcal{D} = ((w_1 = 1, \vec{x}_1), (w_2 = 1, \vec{x}_2), ..., (w_{N_{\text{data}}} = 1, \vec{x}_{N_{\text{data}}})), \tag{6.63}$$

where N_{data} is the total number of data events and $\vec{x}_i$ represents the 24 observables for event i, and the MC sample at both the detector level and truth level, with

$$\mathcal{MC}_{\text{reco}} = ((w_1^{\text{reco}}, \vec{x}_1^{\text{reco}}),\ (w_2^{\text{reco}}, \vec{x}_2^{\text{reco}}),\quad \text{none} \quad , ..., (w_{N_{\text{MC}}}^{\text{reco}}, \vec{x}_{N_{\text{MC}}}^{\text{reco}})) \tag{6.64}$$

$$\mathcal{MC}_{\text{truth}} = ((w_1^{\text{truth}}, \vec{x}_1^{\text{truth}}),\quad \text{none} \quad , (w_3^{\text{truth}}, \vec{x}_3^{\text{truth}}), ..., (w_{N_{\text{MC}}}^{\text{truth}}, \vec{x}_{N_{\text{MC}}}^{\text{truth}})) \tag{6.65}$$

where N_{MC} represents the total number of MC events. $w_j^{\text{reco(truth)}}$ and $\vec{x}_j^{\text{reco(truth)}}$ represent the MC event weight at the detector (truth) level and the 24 observables in event j at the detector (truth) level. As depicted above, an MC event j need

not necessarily exist at both the detector level and the truth level. Having outlined the intial inputs, the basic steps for a single iteration, k, of the method are then as follows.

Step 1 Using the reweighting method described above, the function $\omega^{(k)}(\vec{x}^{\,\mathrm{reco}})$ is determined that reweights the detector level MC from the previous iteration to the data, $\mathcal{D}$:

$$\mathcal{MC}^{(k)}_{\mathrm{reco}} = (w_j^{\mathrm{reco}\,(k)} = w_{\mathrm{push},j}^{\mathrm{reco}\,(k-1)} \times \omega^{(k)}(\vec{x}_j^{\,\mathrm{reco}}), \vec{x}_j^{\,\mathrm{reco}}) \;\; \forall\; j \text{ where } \exists\; \vec{x}^{\,\mathrm{reco}}. \tag{6.66}$$

Here, $w_{\mathrm{push},j}^{\mathrm{reco}\,(k-1)}$ is the detector level MC weight for event j after the completion of the previous iteration. For the first iteration, it is simply the initial w_j^{reco}.

"Pull" weights The weights are propagated ("pulled") in the form of multiplicative factors to the particle level MC, this results in a reweighted truth level MC sample that is denoted by $\mathcal{MC}^{\mathrm{pull}\,(k)}_{\mathrm{truth}}$:

$$\mathcal{MC}^{\mathrm{pull}\,(k)}_{\mathrm{truth}} = (w_{\mathrm{pull},j}^{\mathrm{truth}\,(k)} = w_j^{\mathrm{truth}\,(k-1)} \times \omega^{(k)}(\vec{x}_j^{\,\mathrm{reco}}), \vec{x}_j^{\,\mathrm{truth}}) \;\; \forall\; j \text{ where } \exists\; \vec{x}^{\,\mathrm{reco}}. \tag{6.67}$$

For particle level MC events that are not reconstructed, the pulled weights are estimated by determining a reweighting function $\rho^{(k)}(\vec{x}^{\,\mathrm{truth}})$ that reweights $\mathcal{MC}^{(k-1)}_{\mathrm{truth}}$ to match $\mathcal{MC}^{\mathrm{pull}\,(k)}_{\mathrm{truth}}$ $\forall\; j$ where $\exists\; \vec{x}^{\,\mathrm{reco}}$:

$$\mathcal{MC}^{\mathrm{pull}\,(k)}_{\mathrm{truth}} = (w_{\mathrm{pull},j}^{\mathrm{truth}\,(k)} = w_j^{\mathrm{truth}\,(k-1)} \times \rho^{(k)}(\vec{x}_j^{\,\mathrm{truth}}), \vec{x}_j^{\,\mathrm{truth}}) \;\; \forall\; j \text{ where } \nexists\; \vec{x}^{\,\mathrm{reco}}. \tag{6.68}$$

Step 2 A new reweighting function, $\nu^{(k)}(\vec{x}^{\,\mathrm{truth}})$, is determined that replicates the reweighting performed using the weights from Step 1 using the particle level quanti-

ties, $\vec{x}^{\,\text{truth}}$. This is done by reweighting the MC truth sample, $\mathcal{MC}^{(k-1)}_{\text{truth}}$, to the MC truth sample with the pull weights applied, $\mathcal{MC}^{\text{pull}\,(k)}_{\text{truth}}$:

$$\mathcal{MC}^{(k)}_{\text{truth}} = (w_j^{\text{truth}\,(k)} = w_j^{\text{truth}\,(k-1)} \times \nu^{(k)}(\vec{x}^{\,\text{truth}}), \vec{x}^{\,\text{truth}}) \;\; \forall\; j \text{ where } \exists\; \vec{x}^{\,\text{truth}} \quad (6.69)$$

If the desired number of iterations has been reached, this is when the algorithm stops.

"Push" weights The weights are next propagated ("pushed") to the reconstructed level MC in the form of multiplicative factors. This results in a reweighted reconstructed level MC sample that is denoted by $\mathcal{MC}^{\text{push}\,(k)}_{\text{reco}}$:

$$\mathcal{MC}^{\text{push}\,(k)}_{\text{reco}} = (w_{\text{push},j}^{\text{reco}\,(k)} = w_j^{\text{reco}\,(k-1)} \times \nu^{(k)}(\vec{x}^{\,\text{truth}}), \vec{x}^{\,\text{reco}}) \;\; \forall\; j \text{ where } \exists\; \vec{x}^{\,\text{truth}}. \quad (6.70)$$

For reconstructed events that do not exist at truth level, the pushed weights are estimated by determining a reweighting function $\xi(\vec{x}^{\,\text{reco}})$ that reweights $\mathcal{MC}^{(k-1)}_{\text{reco}}$ to $\mathcal{MC}^{\text{push}\,(k)}_{\text{reco}}$ $\forall\; j$ where $\exists\; \vec{x}^{\,\text{truth}}$:

$$\mathcal{MC}^{\text{push}\,(k)}_{\text{reco}} = (w_{\text{push},j}^{\text{reco}\,(k)} = w_j^{\text{reco}\,(k-1)} \times \xi^{(k)}(\vec{x}^{\,\text{reco}}), \vec{x}^{\,\text{reco}}) \;\; \forall\; j \text{ where } \nexists\; \vec{x}^{\,\text{truth}}. \quad (6.71)$$

This process is iterated for a chosen number of iterations. The final result is a reweighting function $\nu(\vec{x}^{\,\text{truth}})$ that reweights the particle level MC to become the "unfolded data". In other words, the final output will be the set of events from the truth level MC sample with a set of modified event weights: $(w_i^{\text{truth}} \times w_i^{\text{MF}}, \vec{x}_i^{\,\text{truth}})$, where $w_i^{\text{MF}} = \nu(\vec{x}_i^{\,\text{truth}}) \prod_{k=1}^{N_{\text{iter}}} \nu_k(\vec{x}_i^{\,\text{truth}})$ is the weight given by the MultiFold algorithm for an event i. These unfolded events may then be used to construct distributions of the observables included in the algorithm for any choice of binning with reasonable statistics. Furthermore, distributions for additional observables may be constructed

provided they may be derived from the input observables.

Chapter 7

Analysis

The ultimate goal of this analysis is to perform a precision measurement of the $Z(\rightarrow \mu\mu)$+jets process using the method described in Section 6.3.2, MultiFold. Performing the measurement in this manner will not only make it highly dimesional and unbinned, but also greatly increase its long term utility. In order to perform the measurement, events that are a part of the $Z \rightarrow \mu\mu$ signal process must be selected. These events will include contributions from the strong, electroweak and diboson Z+jets processes that contain a $Z \rightarrow \mu\mu$ decay.

The relevant observables for this analysis are first discussed in Section 7.1. The process of selecting events such that background contributions are minimized is then described in Section 7.2 for the analysis phase space, and Section 7.3 for the fiducial phase space. Following this, Section 7.4 describes the corrections applied to the MC samples before the data and reconstructed level MC distributions for each of the observables being measured are shown in Section 7.5. Section 7.6 describes the various uncertainties that must be considered for each of the events passing the analysis selection. Section 7.7 then describes some modifications made to the MC samples in order to better facilitate their use in the unfolding procedure, as well as production

of a set of mock data to be used for validation tests. Finally, Section 7.8 details the implementation of both the traditional method for performing the measurement, IBU, and the new method, MultiFold.

7.1 Observables

This analysis focuses on the measurement of 24 observables within the $Z(\rightarrow \mu\mu)$+jets events:

- the transverse momentum and rapidity of the dimuon system: $p_{\mathrm{T}}^{\mu\mu}$ and $y_{\mu\mu}$
- the transverse momentum, pseudorapidity and azimuthal angle for the leading (highest p_{T}) and subleading (second highest p_{T}) muon: $p_{\mathrm{T}}^{\mu 1}$, $\eta_{\mu 1}$, $\phi_{\mu 1}$, $p_{\mathrm{T}}^{\mu 2}$, $\eta_{\mu 2}$ and $\phi_{\mu 2}$
- the transverse momentum (p_{T}), rapidity (y) and azimuthal angle (ϕ) for the leading and subleading charged particle jets
- the mass (m), charged hadron multiplicity (n_{ch}), and N-subjettiness (τ_1, τ_2 and τ_3) for the leading and subleading charged particle jets

The N-subjettiness is an observable designed to probe the jet substructure, and is defined as [145]

$$\tau_N = \frac{1}{\sum_{i\in\mathrm{jet}} p_{\mathrm{T},i} R_0} \sum_{i\in\mathrm{jet}} p_{\mathrm{T},i} \min_{j\in\{1,\ldots,N\}} \{\Delta R_{j,i}\}, \tag{7.1}$$

where i runs over the constituents of a given jet, j runs over the subjet axes and R_0 is the radius parameter used in the jet building algorithm. τ_N is designed to quantify the degree to which a jet can be classified as N subjets.

7.2 Analysis phase space definition

Requirements are placed on events in this analysis primarily to limit contributions from processes other than the $Z \to \mu\mu$ signal process. These requirements include ensuring the events are of good quality, pass criteria related to the muon quality and satisfy various kinematic cuts. The selection performed in this analysis is very typical for any ATLAS analysis selecting $Z \to \mu\mu$ events.

Firstly, the event must have been measured during a period of time where the LHC and ATLAS were operating as expected. This means that the event must be part of the ATLAS Good Runs List (GRL) [146], which defines periods of data-taking that are suitable for use in analyses. Secondly, there is a step which removes events that were measured during times where the detector was experiencing an issue due to: an error in the LAr or tile calorimeter system, an error in the SCT, or a restart of the trigger, timing and control system that causes the event to be incomplete. These two criteria are the cause of the discrepancy between "recorded" and "good for physics" in Figure 3.10.

The data were collected using the ATLAS trigger menu, which selects events containing a wide range of characteristics of interest. In this analysis, only events that satisfy a single muon trigger are selected [147]. The single muon trigger requires the event to have at least one muon with a p_{T} higher than a threshold value (50 GeV), or exceed a lower p_{T} threshold but meet basic isolation requirements. Due to the increasing luminosity and pileup during 2016–2018, the latter condition has a slightly higher p_{T} requirement ($p_{\mathrm{T}} > 26\,\mathrm{GeV}$) as well as more stringent isolation criteria than for the condition applied during 2015 ($p_{\mathrm{T}} > 20\,\mathrm{GeV}$). The event is also required to have at least one reconstructed primary vertex.

The selection criteria are designed to provide a pure and high statistics sample of

$Z \to \mu\mu$ events, which requires that the muons be measured with high efficiency and a low fake rate. The remaining selection criteria thus require defining the reconstructed particles within each event in order to fully define the analysis phase space. Physics objects within an event in the ATLAS detector are reconstructed using the algorithms discussed in Chapter 5, and have a selection applied as outlined below. This selection is also summarized in Table 7.2.

The muons must pass so-called "Medium" identification requirements [116]. In order to satisfy this requirement, the muon must:

- be either a CB or IO muon (see Section 5.2);
- have a q/p significance $= \frac{|(q/p)_{\mathrm{ID}} - (q/p)_{\mathrm{MS}}|}{\sqrt{\sigma((q/p)_{\mathrm{ID}})^2 + \sigma((q/p)_{\mathrm{MS}})^2}} < 7$, where $\sigma(q/p)$ is the uncertainty on the q/p quantity in both the ID and the MS. This serves as a loose compatibility check between the ID and MS momentum measurements in order to minimize the number of hadrons misidentified as muons;
- have more than one hit in the MDT, or one hit and less than two holes within the $|\eta| < 0.1$ region.

Utilizing the "Medium" requirement results in an identification efficiency of 96.1% for muons with $20\,\mathrm{GeV} < p_{\mathrm{T}} < 100\,\mathrm{GeV}$, measured using simulated $\bar{t}t$ events containing a W boson decay to a prompt muon [43].

Additionally, the muons must satisfy track-to-vertex association (TTVA) requirements, which are designed to minimize the contributions from muons that do not originate from the hard scatter vertex. The TTVA requirements are:

- $|d_0|/\sigma(d_0) < 3$;
- $|z_0 \sin\theta| < 0.5\,\mathrm{mm}$.

The muons are also subject to an isolation requirement in order to reduce the number that originate from hadronic sources as opposed to from a Z boson. The isolation criteria requires that the p_{T} from other sources within a certain ΔR of the muon does not surpass a threshold. Lastly, the individual muons must have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and be within the $|\eta| < 2.4$ region.

The remaining physics objects that must be properly defined are the tracks and the track jets. Tracks must first fulfill a "Loose" selection criterion, which requires the track to have:

- $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ and $|\eta| < 2.5$;
- at least seven hits in the silicon layers of the ID;
- no more than two holes in the silicon layers and no more than one hole in the pixel layers;
- no more than one shared hit.

They must also pass a "Tight" TTVA selection in order to limit the number of tracks that do not originate from the hard scatter vertex. This requires that the tracks have $|d_0| < 0.5\,\mathrm{mm}$ and $|z_0 \sin\theta| < 3\,\mathrm{mm}$. Additionally, any tracks that are part of a reconstructed muon are excluded from the track collection.

The general prescription for jet building was discussed in Section 5.3. For this analysis, all tracks passing the criteria discussed in the previous section are used as inputs to the jet building algorithm. The anti-k_t algorithm is implemented using FastJet [148] with the radius parameter set to $R = 0.4$ to produce the final track jet collection. There is no p_{T} requirement for the track jets, and thus it is possible (although very rare, particularly for the leading track jet) for the track jet p_{T} to be only $500\,\mathrm{MeV}$ if composed of only one low p_{T} track.

Event selection	Description
Basic data quality	All ATLAS detector systems must be operational No LAr calorimeter, tile calorimeter, or tracker errors Event is complete
Trigger	Pass a single muon trigger `HLT_mu20_iloose_L1MU15_OR_HLT_mu50` (2015) `HLT_mu26_ivarmedium_OR_HLT_mu50` (2016-18)
Primary vertex	Have at least one primary vertex
Muons	$N_{\text{muons}} \geq 2$ (as defined in Table 7.2) Opposite charges Pass TTVA recommendations for muons $81\,\text{GeV} < m_{\mu\mu} < 101\,\text{GeV}$ $p_{\text{T}}^{\mu\mu} > 190\,\text{GeV}$

Table 7.1: A summary of the selection criteria applied for this analysis.

Object	Definition
Muons	$p_{\text{T}} > 25\,\text{GeV}$, $\lvert\eta\rvert < 2.4$ Medium quality, pass isolation
Tracks	$p_{\text{T}} > 500\,\text{MeV}$ Loose quality Pass tight TTVA criteria Do not originate from a muon
Track jets	Constructed using anti-k_t algorithm with $R = 0.4$ Input tracks must fulfill the above requirements

Table 7.2: A summary of the selection criteria applied to each reconstructed physics object in the analysis.

There are several requirements placed on the reconstructed dimuon system. The event is required to have at least two muons with opposite charge and the reconstructed dimuon mass, $m_{\mu\mu}$, must be between 81 GeV and 101 GeV such that it falls within a 10 GeV range of the Z boson mass. The dimuon system is also required to have $p_{\text{T}}^{\mu\mu} > 190\,\text{GeV}$. This has primarily been chosen to ensure that the balancing jet(s) have a high quantity of charged particles and to reduce the overall dataset size.

A summary of the event selection criteria can be found in Table 7.1.

7.3 Fiducial phase space definition

The previous section discussed the selection criteria applied to the events and objects at the reconstructed level. This is mimicked at the particle (truth) level in order to define the so-called fiducial volume, which defines the measurement and is required to compare with theoretical predictions (e.g. the particle level MC samples, or predictions from theorists).

At the event level, the fiducial phase space is defined using an identical kinematic selection to that applied at the reconstructed level. Thus, there must be at least two truth muons with opposite charge that have a dimuon mass in the $81\,\text{GeV} < m_{\mu\mu} < 101\,\text{GeV}$ range, and $p_{\text{T}}^{\mu\mu}$ must be larger than 190 GeV.

The fiducial volume is defined at the particle level, which means that all particles considered must be stable at the detector scale. This is ensured by imposing a requirement that $c\tau > 10\,\text{mm}$ for each particle, which coincides with the standard definition used for LHC experiments. The particle level muons are evaluated after being "dressed". The dressing procedure accounts for final state radiation photons emitted from the muon, and adds the four momenta of any photons within $\Delta R < 0.1$ of the muon to the muon four momentum. The individual muons must have a p_{T} larger than 25 GeV and be within $|\eta| < 2.4$, identical to the kinematic selection applied at the reconstructed level. Stable charged particles, excluding the muons, are used to define the tracks at particle level. They must have $p_{\text{T}} > 500\,\text{MeV}$ and $|\eta| < 2.5$, as at the reconstructed level. The track jets are then built using the same method as at the reconstructed level but using the charged particles as input to the jet building algorithm.

Object	Selection criteria
Dressed muons	$p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$
Charged hadrons	$p_{\mathrm{T}} > 500\,\mathrm{MeV}$, $\lvert\eta\rvert < 2.5$
Fiducial volume	At least two oppositely charged muons $81\,\mathrm{GeV} < m_{\mu\mu} < 101\,\mathrm{GeV}$ $p_{\mathrm{T}}^{\mu\mu} > 190\,\mathrm{GeV}$

Table 7.3: A summary of the fiducial selection criteria applied for this analysis.

7.4 Corrections to Monte Carlo samples

Multiple corrections designed to account for known modelling differences between simulation and data are applied to the MC samples in order to improve their agreement with the observed data. The first such correction is known as pileup reweighting [149], which adjusts the MC sample such that the pileup conditions are more in line with those observed in data. This is done by first calibrating the μ value, the number of interactions per bunch crossing, and then correcting the distribution of μ such that it better matches between MC and data. While the μ is a measured quantity in data, in the MC samples it is an assigned parameter that determines the number of pileup interactions to be overlaid (see Section 4.2.5). To derive this correction, the first step is to apply a calibration factor to the measured μ value in data. This is done since the MC simulation of an event with a given μ is known to be too hard, i.e. it better models data events with a higher observed μ value. Thus, the reconstructed μ value in data is scaled by a factor of $1/1.03$ before deriving the correction required to bring the data and MC distributions of μ into agreement. This second correction is applied to the MC event weight in the form of a multiplicative factor, and is, in general, a small correction with a mean of 1 and a standard deviation of around 0.2.

There are also various corrections applied due to observed differences between the data and MC samples related to the muons. These are applied in the form of

correction factors to the MC events, and are designed to correct for inaccuracies in the muon efficiency, i.e. the probability for a real muon to be reconstructed, to fulfill isolation criteria, and to fulfill TTVA criteria. They depend on the properties of the muon(s) within the event, and are derived by measuring the various efficiencies using the tag-and-probe method with J/ψ and Z events in the central region [43] separately for MC and data. An appropriate correction is determined for each muon in the event and then multiplied together before being applied to the MC event weight. These correction factors are all on average very close to one, with standard deviations of less than 0.01.

An additional correction is applied to the simulated muon p_{T}, as differences are observed between its value in simulation versus its value in data. This muon momentum calibration depends on the η and ϕ position of the muon within the detector, and the correction is derived using the same tag-and-probe method that was used to determine the muon efficiency correction factors [43].

Lastly, there are differing trigger efficiencies between data and simulated events. This correction is also applied as a correction factor, and once again depends on the positioning of the triggering muon within the detector as well as its p_{T}. The correction factors are determined using the tag-and-probe method that was discussed previously [147].

7.5 Observed and predicted event yields

Following the corrections to the MC samples described above, and the application of the selection criteria to both the data and MC samples in order to obtain a comparable set of signal events to be analyzed from both, the predicted event yield from the MC samples and the associated distributions for each observable may be compared with

Sample	Selection criteria						
	None	DQ	Trigger	Dimuon	$p_{\mathrm{T}}^{\mu\mu}$ $> 165\,\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu\mu}$ $> 190\,\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu\mu}$ $> 200\,\mathrm{GeV}$
Data	1115M	1082M	274.1M	76.10M	391.0k	246.7k	207.3k
Strong $Z(\to \mu\mu)$+jets							
MG5_aMC+Py8 FxFx	315.2M	313.3M	178.8M	84.73M	388.5k	245.9k	206.7k
Sherpa 2.2.11	308.9M	307.1M	173.8M	81.51M	368.3k	231.6k	194.2k
Sherpa 2.2.1	290.1M	288.4M	163.4M	76.91M	371.1k	235.9k	198.6k
MG5_aMC+Py8 CKKW-L	292.6M	290.9M	168.7M	80.28M	436.9k	281.0k	240.8k
Powheg+Pythia8	271.1M	269.5M	159.3M	77.04M	304.1k	188.9k	157.8k
EW $Z(\to \mu\mu)jj$	86.16k	85.63k	68.88k	36.64k	5.355k	3.928k	3.484k
$Z(\to \mu\mu)V$	3.330M	3.312M	903.7k	143.8k	10.06k	7.079k	6.183k
Other VV	10.53M	10.47M	1.464M	14	0	0	0
Top	83.08M	82.71M	17.44M	97.62k	1.358k	649	498

Table 7.4: Observed and predicted event yields for data and the MC samples after the data quality (DQ), trigger, and muon selection criteria are applied as discussed in Section 7.2. Differences in yield between MC and data are expected up to the dimuon selection as the data collected by the full trigger menu will contain substantial QCD and W+jets background while the MC samples will not. These contributions are negligible after dimuon selection is applied.

data. Table 7.4 presents the observed event yield in data and the predicted event yields from the MC samples at reconstructed level at various stages of the event selection, and Table 7.5 shows the relative reduction between the stages. Figures 7.1 to 7.6 show the data and reconstructed MC distributions with the $p_{\mathrm{T}}^{\mu\mu} > 190\,\mathrm{GeV}$ selection criterion applied for a subset 24 observables of interest for this analysis. The remaining distributions can be found in Appendix A.

The agreement between the reconstructed MC and data varies depending on the MC sample. In general, the MC generators that contain additional higher order parton emissions (i.e. Sherpa 2.2.1, Sherpa 2.2.11 and MadGraph+Pythia8 FxFx) agree better with data. In particular, MadGraph+Pythia8 FxFx approaches the event yield of data very closely and is thus used as the primary sample for this analysis. Sherpa 2.2.11 also performs well and is used as the alternative MC sample.

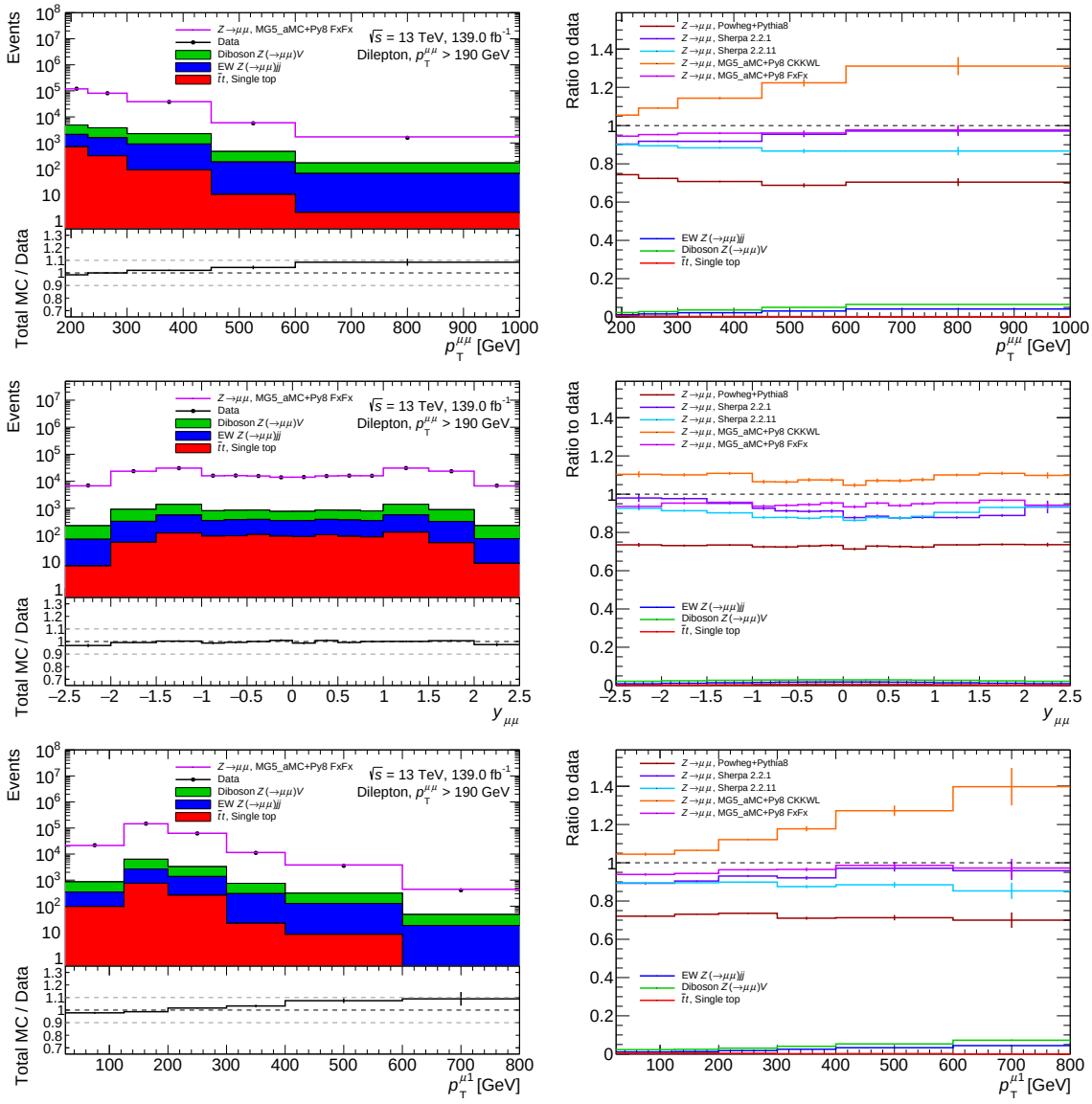


Figure 7.1: MC predicted distributions for $p_{\mathrm{T}}^{\mu\mu}$, $y_{\mu\mu}$ and the p_{T} of the leading muon compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

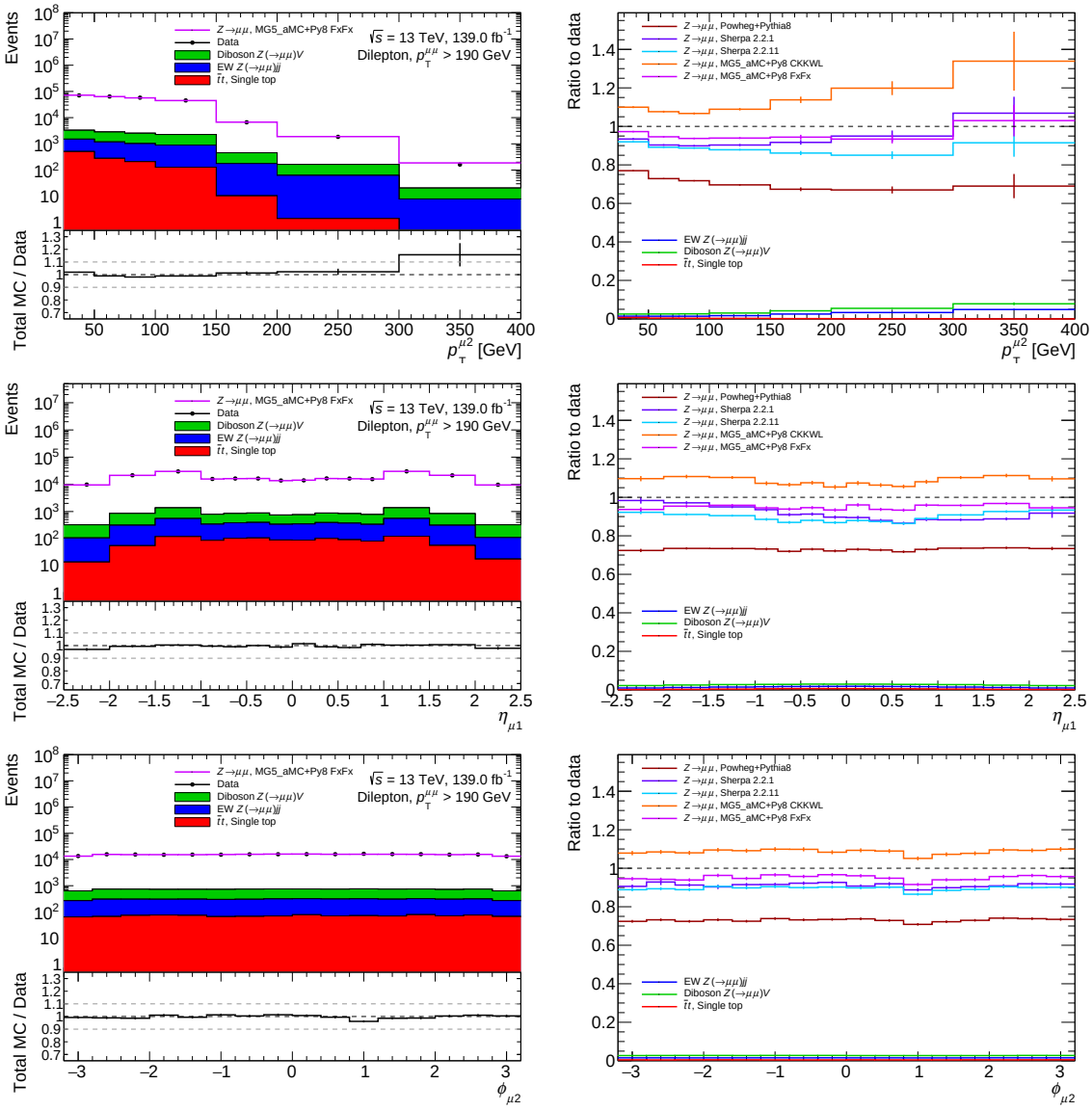


Figure 7.2: MC predicted distributions for the subleading muon p_{T}, η of the leading muon and ϕ of the subleading muon compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

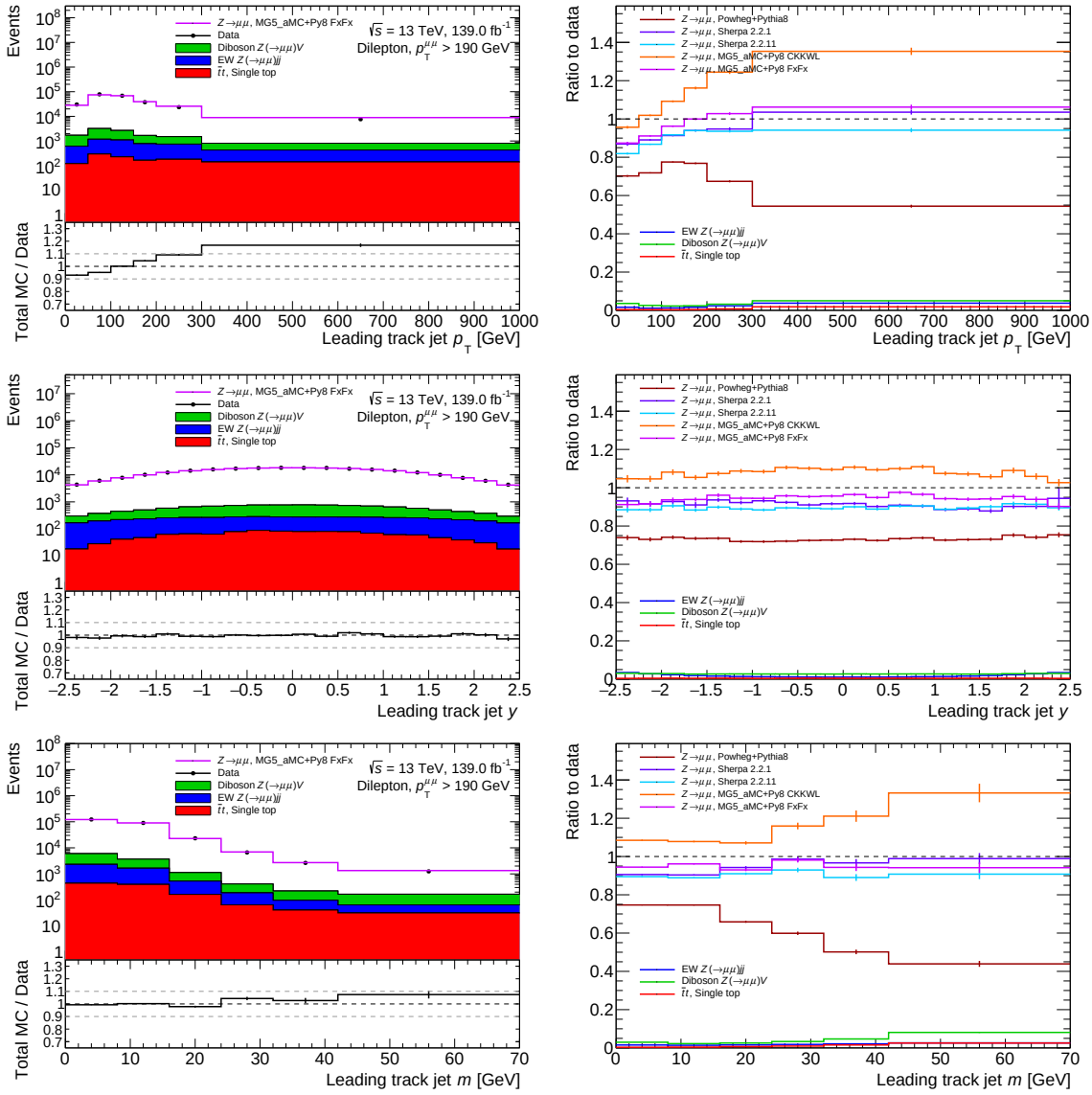


Figure 7.3: MC predicted distributions for the p_{T}, y and m of the leading track jet compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

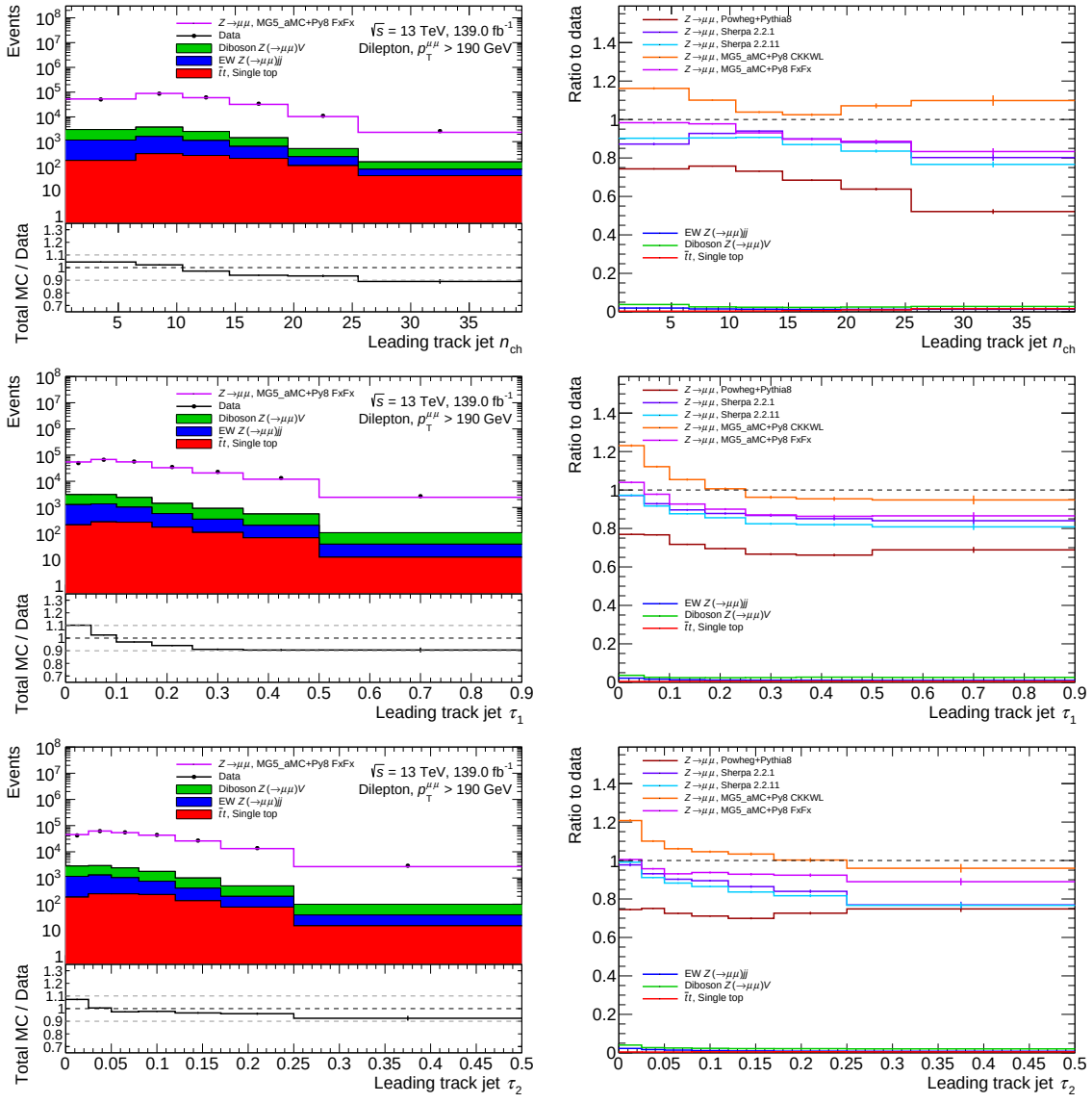


Figure 7.4: MC predicted distributions for the number of charged constituents, τ_1 and τ_2 for the leading track jet compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

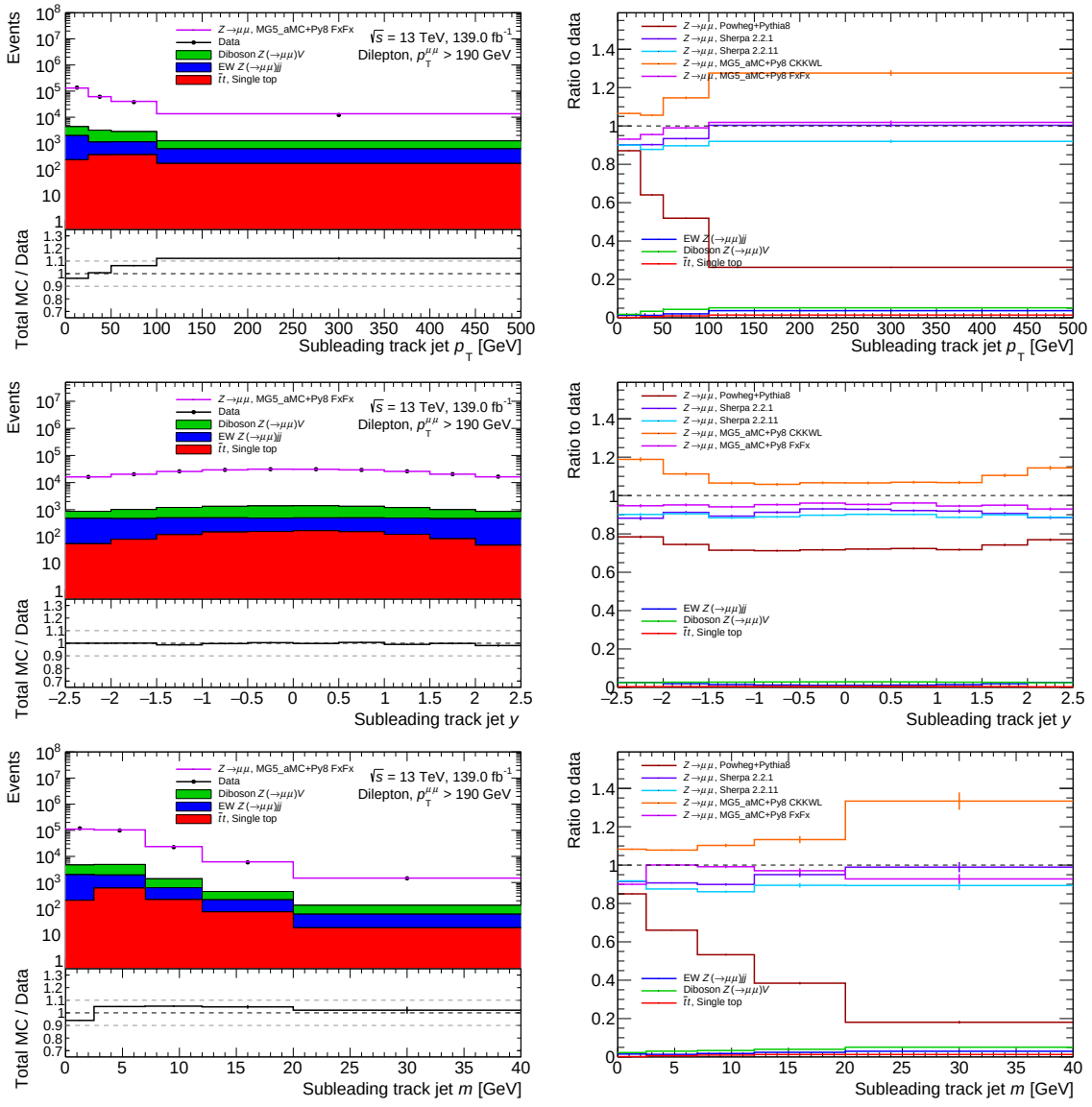


Figure 7.5: MC predicted distributions for the p_{T}, y and m of the subleading track jet compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

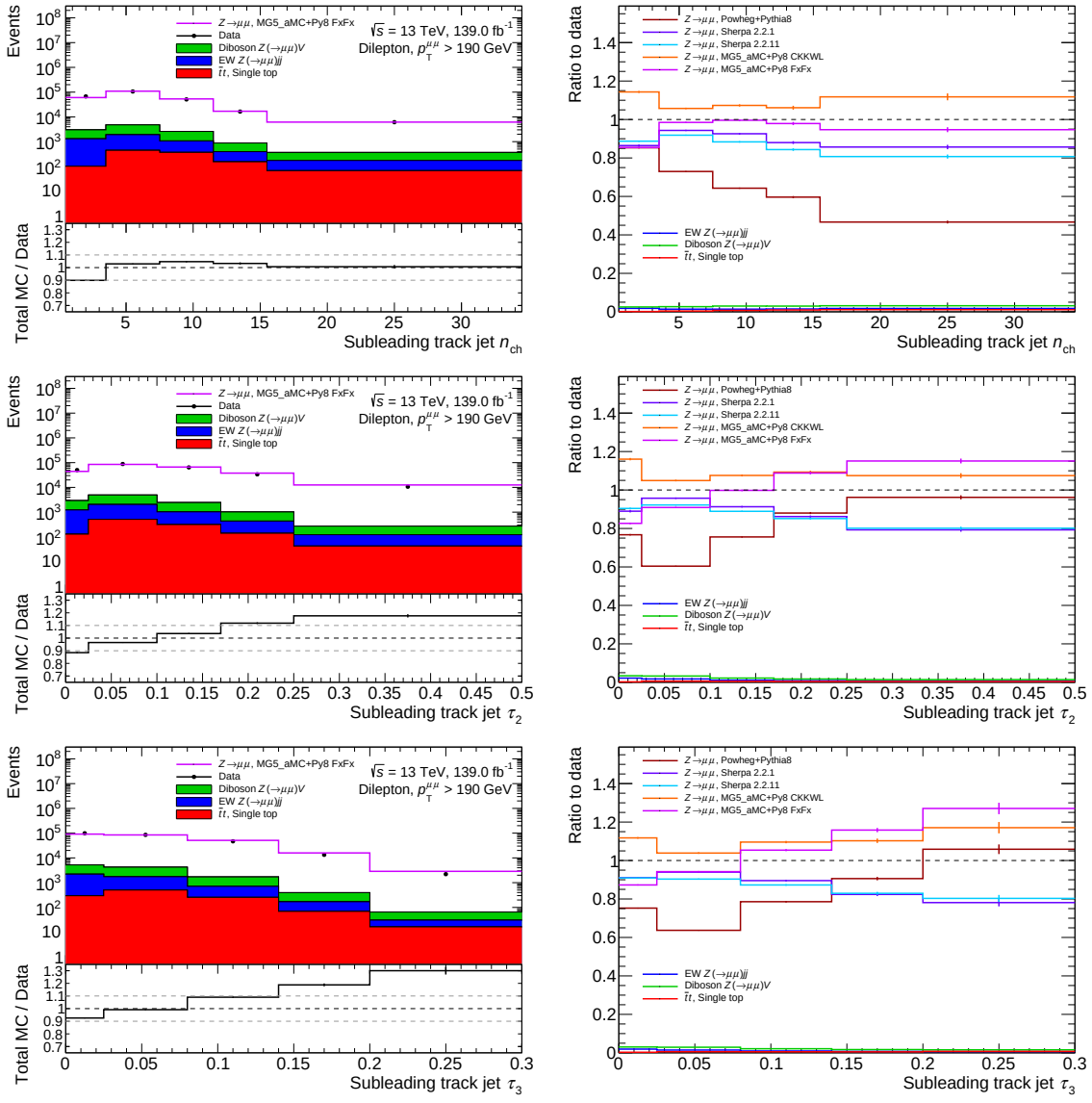


Figure 7.6: MC predicted distributions for the number of charged constituents, τ_2 and τ_3 for the subleading track jet compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

Sample	Selection criteria					
	DQ	Trigger	Dimuon	$p_{\mathrm{T}}^{\mu\mu}$ > 165 GeV	$p_{\mathrm{T}}^{\mu\mu}$ > 190 GeV	$p_{\mathrm{T}}^{\mu\mu}$ > 200 GeV
Data	-3.02%	-74.66%	-72.24%	-99.49%	-36.91%	-15.95%
Strong $Z(\rightarrow \mu\mu)$+jets						
MG5_aMC+Py8 FxFx	-0.58%	-42.93%	-52.62%	-99.54%	-36.69%	-15.96%
Sherpa 2.2.11	-0.59%	-43.42%	-53.09%	-99.55%	-37.11%	-16.15%
Sherpa 2.2.1	-0.59%	-43.36%	-52.92%	-99.52%	-36.43%	-15.81%
MG5_aMC+Py8 CKKW-L	-0.58%	-42.00%	-52.42%	-99.46%	-35.69%	-14.29%
Powheg+Pythia8	-0.58%	-40.88%	-51.65%	-99.61%	-37.88%	-16.46%
EW $Z(\rightarrow \mu\mu)jj$	-0.61%	-19.57%	-46.80%	-85.38%	-26.65%	-11.32%
$Z(\rightarrow \mu\mu)V$	-0.54%	-72.71%	-84.09%	-93.00%	-29.64%	-12.65%
Other VV	-0.54%	-86.02%	-99.99%	-99.90%	-25.96%	-0.00%
Top	-0.45%	-78.91%	-99.44%	-98.61%	-52.23%	-23.22%

Table 7.5: The percentage of events rejected after the application of each set of selection criteria relative to the previous set.

7.6 Uncertainties

A complete set of systematic uncertainty sources is evaluated for this analysis, with the goal of providing a 68% confidence interval to the true underlying physics. They can be divided into three main categories: experimental systematic uncertainties, which are largely related to aspects of the event and object reconstruction and are discussed in detail in Section 7.6.1; theoretical uncertainties, which are associated with aspects of the theoretical modelling within the MC generators and are discussed in detail in Section 7.6.2; and, statistical uncertainties, which are discussed in Section 7.6.3. Uncertainties associated with the unfolding procedure are also present and will be discussed in more detail in association with each unfolding method in Sections 7.8.1.2 and 7.8.2.2. The uncertainties are evaluated using two techniques: the "Hessian" method, where the $\pm 1\sigma$ effect is estimated for each uncertainty source and each is considered uncorrelated to other uncertainties; and replica variation. Each variation is independently propagated through the full analysis in order to determine the impact on the final results.

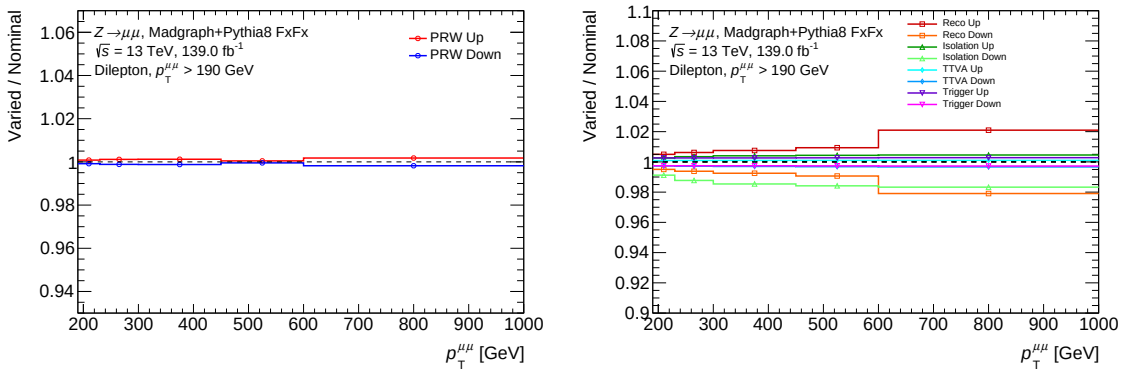


Figure 7.7: The impact of the up and down variations of the pileup reweighting (left) and the up and down variations of the muon efficiency correction factors (right) on the reconstructed dimuon distribution in the MadGraph+Pythia8 FxFx sample.

7.6.1 Experimental systematic uncertainties

As has been previously discussed, there is a 1.7% uncertainty associated with the measured luminosity. The majority of the remaining experimental uncertainties are associated with the MC corrections discussed in Section 7.4 and affect quantities related to the reconstructed MC.

The pileup reweighting procedure involves applying a correction factor of 1/1.03 to the measurement of μ before reweighting the MC μ distribution. Up and down systematic variations associated with the pileup reweighting procedure are determined by setting this correction to 1/1.072 and 1/0.988, respectively, followed by a dedicated reweighting. Next, there are up and down variations for the correction factors associated with the trigger, reconstruction, isolation and TTVA efficiencies in the muon reconstruction. These systematic variations affect the reconstructed MC event weights and are dependent on the muon kinematics, and thus result in a shift across all observables. The effect of these variations on the reconstructed $p_{\mathrm{T}}^{\mu\mu}$ distribution can be seen in Figure 7.7.

There are systematic uncertainties related to the muon momentum calibration

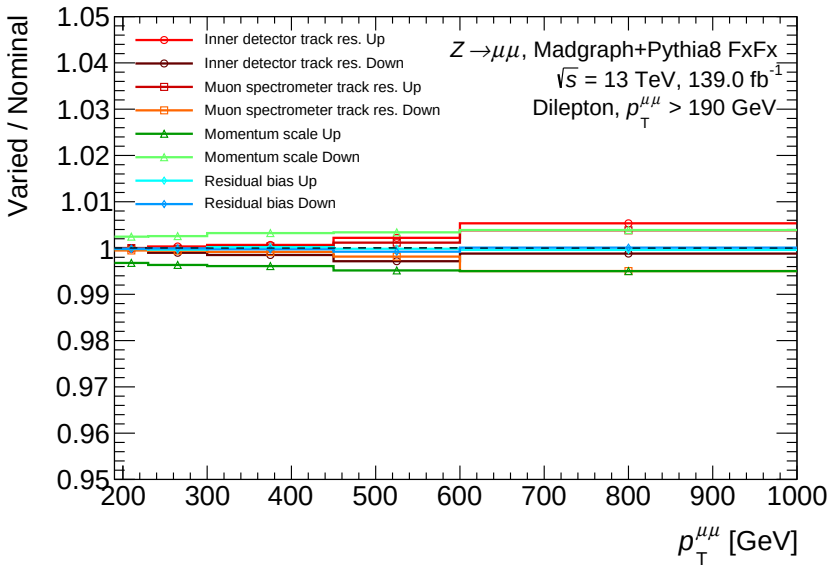


Figure 7.8: The impact of the up and down systematic variations to the muon calibration on the distrbution of the reconstructed dimuon p_{T} for the MadGraph+Pythia8 FxFx sample.

procedure. Four uncertainties, each with up and down variations, are considered: one for momentum variations due to measurements made in the inner detector, one for momentum variations due to measurements made in the muon spectrometer, one for the momentum scale and one related to the measured sagitta[1] value. These uncertainties affect the muon p_{T}, and depend on the muon p_{T}, y and ϕ. As an example, the effect of these systematic uncertainties on the reconstructed $p_{\mathrm{T}}^{\mu\mu}$ distribution can be seen in Figure 7.8.

There are four systematic uncertainties associated with the tracks: one to account for imperfect knowledge of the residual alignment of the inner detector, one to account for variations in the track reconstruction efficiency, one to account for variations in the track reconstruction efficiency when the track is inside a jet, and another to account for variations in the rate of fake tracks. The first of these is a variation

[1]The distance between the centre of circular arc to the centre of its base. The arc in this case is defined by the path the charged particle takes through the ID.

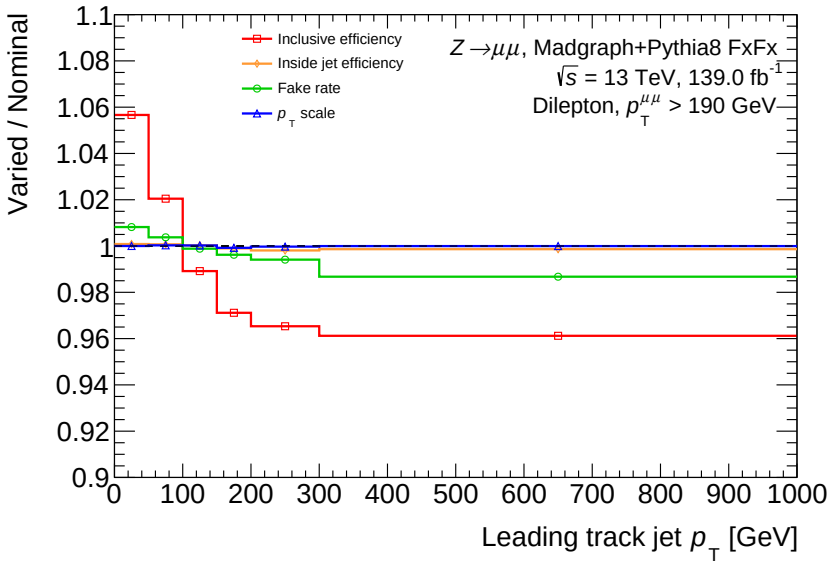


Figure 7.9: The impact of the systematic uncertainties associated with the tracks on the reconstructed distribution of the leading track jet p_{T} in the MadGraph+Pythia8 FxFx sample.

of the track p_{T}, while the remaining three uncertainties involve removing a subset of tracks from the MC event. Unlike the majority of the systematic uncertainties discussed in this section, the track uncertainties are propagated using only the down variation for each uncertainty before being symmetrized in the final result. In order to determine the impact of these uncertainties on the track jet observables, the track jets are rebuilt using the modified track collection for each variation. The effect of these variations on the reconstructed distribution for the leading track jet p_{T} is shown in Figure 7.9.

7.6.2 Theoretical systematic uncertainties

The theoretical uncertainties are associated with the modelling parameters chosen by the MC generators. They thus impact the MC event weights and result in variations to both the reconstructed and truth level MC distributions. They can be broken down

into four categories: the perturbative uncertainties (QCD scale variations), the PDF uncertainties, variations to the chosen value of α_s, and the uncertainties associated with the parton showering algorithm.

The variations of the QCD renormalization (μ_r) and factorization (μ_f) scales are designed to account for the lack of higher order corrections. These variations are determined by the MC generator by recalculating the MC event weights after doubling or halving the individual scales. This results in six uncertainties corresponding to various possible combinations. For example, if both scales are multiplied by two it corresponds to the up-up ("uu") variation, whereas if μ_r is kept at the nominal value and μ_f is halved it corresponds to the nominal-down ("nd") variation. Increasing the scales results in a smaller cross section, however varying the factorization scale results in a much smaller effect than varying the renormalization scale. The impact of each variation on the truth level $p_{\mathrm{T}}^{\mu\mu}$ distribution is shown in Figure 7.10. The largest variation, "dd", is propagated through the full analysis and is taken as the associated uncertainty.

Next, the variability of the underlying PDF is determined. The nominal PDF used by the generator is the central value of a distribution of PDFs. The generator provides a set of 100 PDF replicas sampled from this distribution in order to determine the uncertainty [39]. In a typical analysis, each of the 100 variations is propagated through the full analysis and the final uncertainty is taken to be the root mean square (RMS) of the 100 variations in each bin of the observable distributions.

As a specific example, the changes to the nominal truth level $p_{\mathrm{T}}^{\mu\mu}$ distribution using the 100 PDF replicas as well as the corresponding RMS values are shown in Figure 7.11. Due to the nature of the MultiFold method, the procedure applied for this analysis is slightly modified. The algorithm in Section 7.7.2 is used to produce

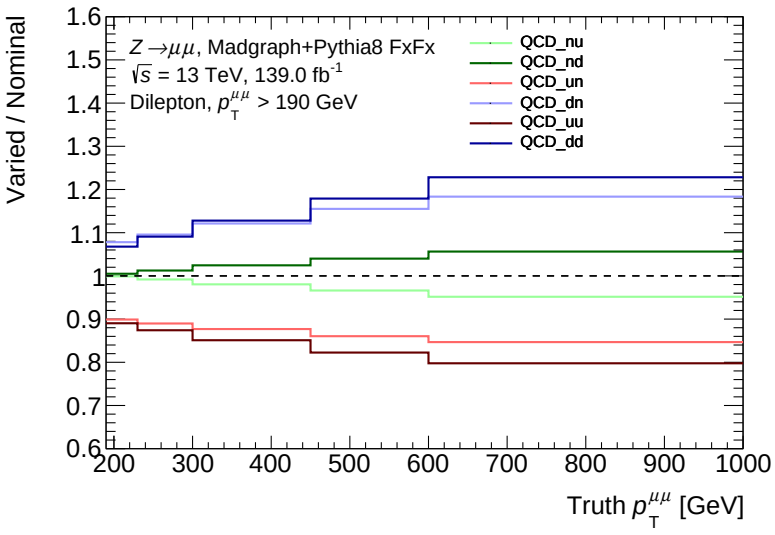


Figure 7.10: The impact of the QCD scale variations in the MadGraph+Pythia8 FxFx sample on the truth level dimuon p_{T} distribution.

a reweighting function to reweight the nominal sample to the sample varied by the RMS of the PDF variations and this varied sample is used to propagate the error through the full analysis.

Upwards and downwards variations of the strong coupling constant α_s to 0.119 and 0.117 from its nominal value of 0.118 are also included as an uncertainty. This affects both the α_s value in the PDF, as well as the α_s value used by the MC generator (see for example Eqn. 2.41). These variations are derived using the Sherpa 2.2.11 sample, as they are not available in the MadGraph+Pythia8 FxFx sample due to a technical issue. They can be seen applied to the truth $p_{\mathrm{T}}^{\mu\mu}$ distribution of the Sherpa 2.2.11 sample in Figure 7.12. In order to determine the variation for the MadGraph+Pythia8 FxFx sample a function is derived to reweight the nominal Sherpa 2.2.11 sample to the varied sample using the algorithm discussed in Section 7.7.2. The reweighting function is then applied to the MadGraph+Pythia8 FxFx sample to determine the variation, which is then propagated through the un-

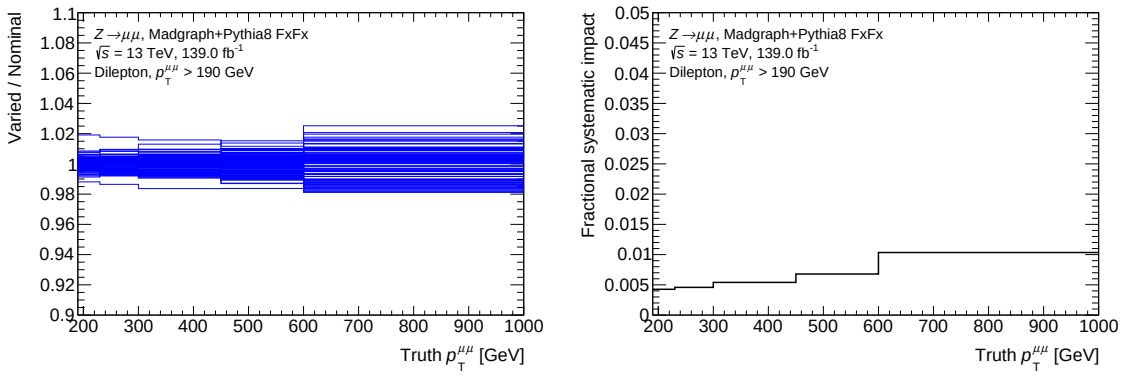


Figure 7.11: The impact of the 100 replicas of the NNPDF3.0 PDF set on the truth level dimuon p_{T} distribution is shown on the left. The distribution on the right shows the fractional systematic impact based on the RMS of these 100 replicas.

folding procedure.

Lastly, there are four uncertainties associated with the parton showering, each with an up and down variation. Two sets of tune variations are considered, one that mainly accounts for underlying event effects (Var1) and another that mainly accounts for effects related to jet shape and substructure (Var2). Renormalization scale variations (Ren) and variations related to multiple parton interactions (MPI) are also considered. These uncertainties are derived with a dedicated Powheg+Pythia8 sample using the AZNLO tune. An example of the impact of these systematics on the number of constituents in the leading track jet can be seen in Figure 7.13. The renormalization scale variation causes the largest shift to the distribution, while the remaining variations are much smaller. In an identical procedure to the one used for the α_s uncertainties, in order to apply these variations to the MadGraph+Pythia8 FxFx sample a reweighting function is derived using the method described in Section 7.7.2 which reweights the nominal Powheg+Pythia8 sample to match the varied sample. This reweighting function is then applied to the MadGraph+Pythia8 FxFx sample to determine the systematic variations to be propagated through the

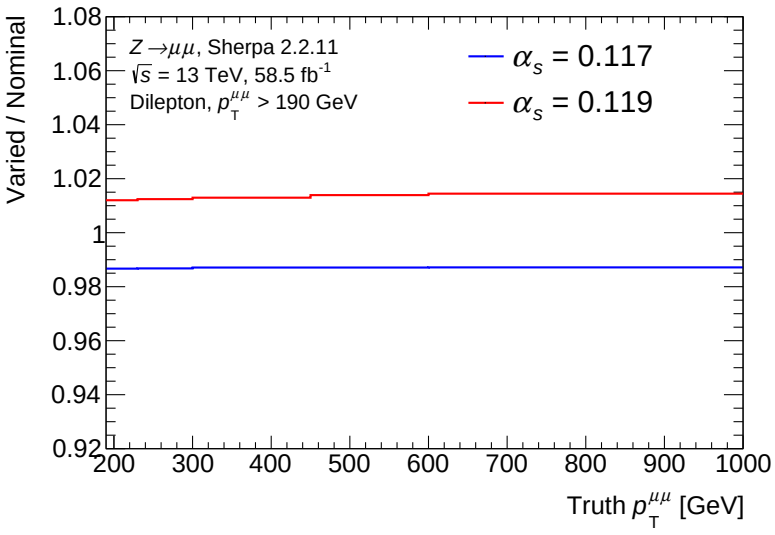


Figure 7.12: The impact of the α_s variations in the SHERPA 2.2.11 sample on the truth level dimuon p_{T} distribution.

unfolding procedure.

7.6.3 Statistical uncertainties

The statistical uncertainties for both data and the MC samples are determined using bootstrapping [150], a procedure which involves producing 100 replica datasets for each sample. In order to produce a replica dataset, each event i is assigned a modified event weight equal to $w_{i,\mathrm{BS}} \times w_i$, where w_i is the original event weight for the ith event and $w_{i,\mathrm{BS}}$ is a randomly sampled value from a Poisson distribution with a mean of one. The individual replica datasets are individually propagated through the full analysis to obtain an associated measurement and the final statistical uncertainty is determined using the spread of these results.

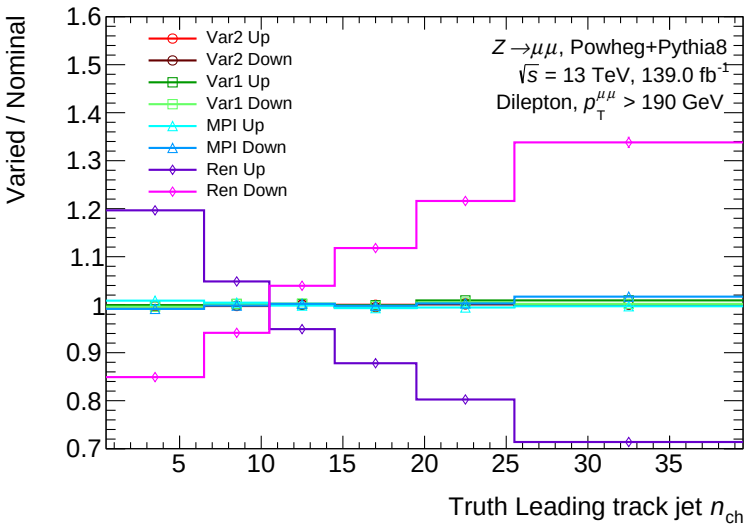


Figure 7.13: The impact of the uncertainties related to the parton showering algorithm on the truth distribution for the number of constituents in the leading track jet. These are derived using a dedicated set of Powheg+Pythia8 samples.

7.7 Sample processing

As discussed in Chapter 6, the MultiFold algorithm takes as input a data sample with $(w_i = 1, \vec{x}_i)$ and an MC sample with $(w_j^{\text{reco}}, \vec{x}_j^{\text{reco}})$ and $(w_j^{\text{truth}}, \vec{x}_j^{\text{truth}})$ for the reconstructed and truth level events, respectively. As stated in Section 7.5, the MadGraph+Pythia8 FxFx sample is used as the nominal in this analysis. One potential issue that could arise if the entire dataset is used to train the model is that of overtraining. This occurs when the output model does a suitable job of modelling the training data, but does not generalize well. In order to mitigate this, the MadGraph+Pythia8 FxFx sample is divided into a subset to be used for training the model, and a subset to which that model is applied. These are known as the train and test splits, respectively, and their derivation is outlined in Section 7.7.1.

While the MadGraph+Pythia8 FxFx sample models the data relatively well, it is also the strong Z+jets sample with the largest fraction of negative MC

event weights. This creates two issues: it inflates the input file size, and negative weights are known to create difficulties for the NN. A way must thus be developed to eliminate these negative weights from the sample. Another point that has yet to be discussed in detail in this thesis is the mock data or pseudodata that is mentioned in Chapter 1 in reference to performing tests of the MultiFold method. In order to perform these tests it is required that the pseudodata emulate the data as closely as possible, but access to a target truth distribution is also required. Therefore, the pseudodataset must be produced using MC samples, but must be adjusted such that it agrees with the data at both the detector level and the truth level.

Both of the points raised in the previous paragraph imply some kind of reweighting. However, this reweighting must be disconnected from the main MultiFold algorithm. For one, as was already stated, negative weights can create issues for NNs. Thus, using NN reweighting to produce a positive weight MadGraph+Pythia8 FxFx sample by inputting the sample with the large percentage of negative weights is undesirable. In the case of the reweighting to data required for the pseudodataset, this should ideally be performed with a separate algorithm in order to avoid biasing the unfolding. An alternative reweighting algorithm, OmniSequential, was thus developed for this analysis in order to accomplish these two tasks. It is discussed in detail in Section 7.7.2. Sections 7.7.3 and 7.7.4 then outline the specifics of its use in the elimination of negative weights and the pseudodataset production, respectively.

7.7.1 Test and train splits

As referenced above, the NNs required for the MultiFold method are trained using a subset of the MC sample (the train set). After the model has been trained, it is applied to an independent subset of the sample (the test set) in order to obtain the

unfolded result. The number of effective events is the metric used to determine how the samples should be split. It is defined as,

$$N_{\text{eff}} = \frac{\left(\sum_{i=1}^{N_{\text{evts}}} w_i\right)^2}{\left(\sum_{i=1}^{N_{\text{evts}}} w_i^2\right)}, \tag{7.2}$$

where w_i represents the event weight for the ith event, and N_{evts} represents the total number of events in the sample. For data, where $w_i = 1$ for all events, N_{eff} will always be equal to the total number of data events. For weighted MC events this will not be the case. As was briefly discussed in Section 4.2.6, this is also further complicated by negative weight events, which in general require that more events be produced in order to obtain the same statistical power as a sample with only positive weights.

The MadGraph+Pythia8 FxFx sample, which is used as the nominal MC sample for this analysis due to its good agreement with the data (see Section 7.5) contains roughly four times the effective statistics of data. Thus, a total of four times the effective statistics of data are kept for each MC sample to remain consistent with the main one. These events are then split with one quarter being included in the test file, and the remaining three quarters being included in the train file. It is advantageous to make the training dataset as large as possible, however, if the test dataset is too small the statistical errors will greatly increase. Therefore, this split is chosen in order to balance these two considerations. The event weights are also scaled according to the fraction kept as well as the test/train split percentage such that the overall yield is unaffected. The remaining events, if any, are placed in a "rest" file and are used for the pseudodata production.

Table 7.6 summarizes the number of effective events for each sample as well as the fraction of the sample that is kept to be used for the test and train samples. The

Sample	N_{evts}	N_{eff}	Predicted yield	N_{eff}/Pred.	Percent kept	Negative weight fraction
Data	246677	246677	–	–	–	–
MG5_aMC+Py8 FxFx	19260277	963297	234400	4.11	100%	35.86%
Sherpa 2.2.11	27618109	3155432	219595	14.37	25%	14.17%
Sherpa 2.2.1	6836375	709826	225405	3.15	100%	16.30%
MG5_aMC+Py8 CKKW-L	1004321	557025	267875	2.08	100%	0.06%
Powheg+Pythia8	822091	743068	180067	4.13	100%	1.02%
EW $Z(\to \mu\mu)jj$	571378	196736	3728	52.77	6%	6.77%
$Z(\to \mu\mu)V(\to jj)$	991378	149367	6738	22.17	15%	7.61%
Top	4849	4498	612	7.35	50%	0.25%

Table 7.6: The number of effective events (N_{eff}) and other related quantities for the samples used in this analysis. The percent kept represents the percentage of each sample kept for the test/train samples, the remaining events are saved in a rest sample that is used in the pseudodata production.

fraction of negative weights in each sample is also summarized.

7.7.2 OmniSequential

Referring back to the discussion in the introduction of this section, there are two instances in this analysis where a sample reweighting must occur that is unrelated to the unfolding procedure: reweighting the MC samples such that the negative weights are eliminated, and in the construction of the pseudodata. In order to perform these reweightings, a new method was developed for this analysis, known as OmniSequential.

OmniSequential takes inspiration from the principle behind the reweighting used in the MultiFold method, but employs more standard data analysis techniques. While density ratio estimation (see Section 6.3.2) uses neural networks to perform a high-dimensional smooth reweighting, OmniSequential instead applies a series of reweightings to each event based on one-dimensional and finely binned projections of the samples to obtain a smooth reweighting function. The full procedure will be examined in detail below. As OmniSequential does not examine the complete phase space, it is not necessarily expected to perform well in all regions. However, for included observables and observables related to them, it should provide a good

approximation of the target distribution.

OmniSequential takes two samples as input: a target sample and one sample to be reweighted to match this target sample. The first step of the OmniSequential algorithm is to determine a sensible binning for each observable. A range is manually selected such that it includes at least 99.5% of the data, with the possibility for overflow or underflow bins included according to the nature of the observable in question. The bin width is chosen based on Scott's rule [151], and if the observable has a skewness ($\mu_3 = E[(x-\mu)^3]$) that is larger than two then log binning is used. This choice was found to perform well, and applies to observables with a long tail (e.g. the p_{T} distributions). This allows for good statistical precision in the tail regions that is similar to that of the main body of the distribution. One iteration of the OmniSequential algorithm then proceeds as follows.

- Histograms with the fixed binning outlined above are created for each observable being used as input to the algorithm. One is created from the target sample, and another from the input sample being reweighted.

- The observable with the worst agreement between the input sample and the target is determined using the chi-squared pull:

$$z_{\chi^2} = \frac{\chi^2 - E[\chi^2]}{\sigma[\chi^2]} = \frac{\chi^2 - n_{\mathrm{dof}}}{\sqrt{2n_{\mathrm{dof}}}} \tag{7.3}$$

- A fit, g, (either a polynomial fit or using a Gaussian Kernel) is then applied to the ratio between the input and target histograms for the worst observable.

- The events in the input sample are then reweighted based on the result of the

fit. For each event i, the new weight in the current iteration k, $w_i^{(k)}$, will be:

$$w_i^{(k)} = g(x_i^{(k)}) \times w_i^{(k-1)}, \tag{7.4}$$

where $g(x_i^{(k)})$ is the fit evaluated at the value of the observable with the worst agreement in the current iteration and $w_i^{(k-1)}$ is the weight from the previous iteration.

The iterations continue until a maximum allowed number is reached or closure, where all observables have a $z_{\chi^2} < 1$, is achieved.

The end result of the OmniSequential algorithm will be a weight to be applied to each event i in the input sample,

$$w_i^{\mathrm{OS}}(\vec{x}_i) = \prod_k^{N_{\mathrm{iter}}} g^{(k)}(x_i^{(k)}), \tag{7.5}$$

where $\vec{x}_i$ represents the set of input observables for event i, $g^{(k)}$ represents the fit to the ratio between the input sample and the target for the observable with the worst agreement in iteration k, and $g^{(k)}(x_i^{(k)})$ is the fit evaluated at the value of the worst observable in iteration k for event i.

7.7.3 Positive reweighting

As discussed in the introduction to this section, the first application of the OmniSequential algorithm is to eliminate the negative weights from the MC samples that are to be inputs to the MultiFold algorithm. There are two considerations when reweighting these MC samples to contain only positive weights. Firstly, the distributions of the observables must be statistically compatible before and after the reweighting, which is achieved by appropriately modifying the event weights. Secondly, it is vi-

tal that the overall statistical error is preserved in the reweighting. In a weighted sample, the statistical error in a certain part of phase space containing N events can be expressed as $\sqrt{\sum_{i=1}^{N} w_i^2}$. Naively, then, one could attempt to preserve the statistical uncertainties by ensuring that the distributions of the observables weighted by the usual weights squared are also statistically compatible. Unfortunately, this is not sufficient as the weight landscape will change based on the reweighting applied in order to ensure the distributions of the observables are statistically compatible. Some events are thus dropped from the final sample in order to preserve the variance for localized regions of phase space. This is done in accordance with the procedure outlined in Ref. [152] and while the main procedure is summarized here, additional details and derivations may be found therein.

In order to reweight the weights, w, the following is done.

- The target distributions for the OmniSequential algorithm are determined using the original MC sample at truth level. The input sample to the algorithm has the same events, but with the weights all set to the average from the original sample in order to eliminate the negative weights.

- The average weight sample is then reweighted at truth level to match the target using the OmniSequential algorithm.

In order to preserve the statistical uncertainties, the following steps are undertaken.

- The reweighting for w^2 subsequently proceeds in much the same way as the reweighting for w, except the events in the input and target distributions are weighted by w^2 instead of w.

- For each event i in the reweighted sample, calculate

$$p_i = \frac{(w_{i,\text{rew}})^2}{(w_i^2)_{\text{rew}}}, \tag{7.6}$$

  where $w_{i,\text{rew}}$ is the reweighted weight for event i, and $(w_i^2)_{\text{rew}}$ is the reweighted squared weight for event i.

- Keep the event with probability p_i.

- If the event is kept, set the MC event weight w_i to $w_{i,\text{rew}} \times p_i$.

- Scale the weight for the corresponding reconstructed MC event weight by $\frac{w_i}{w_{i,\text{orig}}}$.

The above procedure is applied in full to the MadGraph+Pythia8 FxFx sample, as it is used as the input to the MultiFold algorithm. The results of reweighting the MadGraph+Pythia8 FxFx sample are shown in Appendix B. Parts of this procedure are also required for the construction of the pseudodata sample, which is discussed in the following section.

7.7.4 Pseudodata production

In order to perform tests of the MultiFold method, access to a pseudodataset is required. The pseudodata is designed to match the data as closely as possible, thus the observable distributions must be statistically compatible between it and the data sample. Furthermore, in order to ensure that the MultiFold result agrees with an expectation "truth data" distribution, the truth level information for the pseudodata must be accessible. The pseudodata must also have event weights that are equal to 1 for all events, like in the actual data.

The MC samples primarily used to construct the pseudodata are the "rest" splits of the Sherpa 2.2.11 sample, the electroweak Z+jets sample and the diboson Z+jets

sample. Before any reweighting to the data is performed, the negative weights in these samples must first be eliminated. This is done using the procedure outlined in the previous section, however in this case the preservation of the statistical uncertainties is not relevant, so only the first reweighting with OmniSequential described in the previous section is performed.

Once the SHERPA 2.2.11, the electroweak and the diboson rest samples have been reweighted such that they contain only positive weights, the three individual samples are combined into one total sample. The OmniSequential algorithm is then used to determine a reweighting function that reweights the combined MC sample to the data with the top background subtracted as the target. As the input sample is being reweighted to data, the reweighting cannot be done at truth level. This creates complications as the truth distributions must be recoverable after reweighting. The sample is thus reweighted using the reconstructed quantities where available, but w_{MC} is reweighted as opposed to the reconstructed weight, w (similar to the MultiFold algorithm). In general, the events in the sample can be divided into three categories:

1. Events that pass both the truth and reconstructed level selection criteria;
2. Events where all reconstructed quantities are defined, but the event does not pass the selection at the reconstructed level;
3. Events where one or both reconstructed muons are not defined.

In the case of category (1), w_{MC} is reweighted based on the reconstructed observables and the corresponding reconstructed weight is adjusted by a factor of $w_{\mathrm{MC,rew}}/w_{\mathrm{MC,orig}}$.

The vast majority of events in category (2) are simply those which fail the final $p_{\mathrm{T}}^{\mu\mu}$ selection, but pass the selection criteria preceeding it. In this case, all the reconstructed quantities are saved and can thus be used to reweight the event using

the reweighting function determined by OmniSequential. The only additional consideration that must be made in this case is if $p_{\mathrm{T}}^{\mu\mu}$ is one of the variables appearing in the reweighting function. In this case, the function is extrapolated to the appropriate value of $p_{\mathrm{T}}^{\mu\mu}$. The reconstructed MC event weight is then adjusted in the same manner as in (1).

The events in category (3) are those in which one or both of the muons are not reconstructed. In this case, the track jet observables are all present at the reconstructed level and are used as usual. For the muon variables, there are two possible cases:

1. The leading muon is defined and the subleading muon is not. This is largely due to cases where the subleading muon fails the $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ criterion due to resolution effects and thus the following is done:

    - the reconstructed quantities are used as usual for the leading muon;
    - the p_{T} of the subleading muon is substituted to be $25\,\mathrm{GeV}$;
    - $p_{\mathrm{T}}^{\mu\mu}$ is approximated to be $p_{\mathrm{T}}^{\mu 1} + 25\,\mathrm{GeV}$;
    - the truth level quantities are used for the subleading muon angular variables.

2. Neither muon is defined. In this case, the truth level quantities are used for all muon observables.

Once again, the reconstructed MC event weight is then adjusted in the same manner as in category (1).

The final step in the pseudodata production is to unweight the events in the reweighted sample. For each event i with weight w_i, a new weight is assigned that is

equal to:

$$w_{\text{pseudodata}} = \text{Poisson}(w_i). \tag{7.7}$$

If $w_{\text{pseudodata}} = N > 1$, then the event is added to the final pseudodata sample N times with weight 1. At this point, the contribution from the top background is also added to the pseudodata sample with the same unweighting procedure.

The final result of this procedure is a pseudodata sample that, while constructed using MC samples, has observables very similar to the data distributions and an equivalent weight distribution where each event has a weight of 1. Furthermore, the truth distributions can be recovered, which gives the ability to check agreement between them and any results obtained. The comparison of the pseudodata sample with the full Run 2 data distributions for the 24 observables are shown in Appendix B.

7.8 Implementation of unfolding methods

Having now defined the fiducial and analysis phase spaces, as well as the samples that are used as inputs to the unfolding algorithms, discussion can now turn to the specifics of the implementation of the unfolding procedures required to obtain the differential cross section measurements as detailed in Chapter 6. This section first presents the procedure implemented to obtain the IBU measurement (introduced in Section 6.2.3), followed by a section detailing the procedure to obtain the measurement using the MultiFold method (introduced in Section 6.3.2).

As a reminder, the goal is to obtain a high dimensional and unbinned measurement of each of the 24 observables outlined in Section 7.1 using the MultiFold method. The role of IBU in this analysis is to perform the same measurement (i.e.

the same 24 observables will be measured using the same samples as input to the algorithm), but in a traditional way to serve as a comparison and validation for the MultiFold measurement.

7.8.1 Iterative Bayesian unfolding implementation

This section will first outline the technical details of the IBU implementation, before showing some relevant quantities from the MC sample that are required for the IBU procedure. Following this, the results of a technical closure test are shown, before finally presenting a discussion of how the uncertainties discussed in Section 7.6 are propagated through the IBU procedure.

The first step of the IBU measurement is to determine an appropriate binning for each of the 24 observables. There is a trade off to consider when determining an appropriate bin width. Having a larger number of bins means that more information about the overall shape of the observable is preserved, however increasing the number of bins also increases the statistical fluctuations. The bin width is thus chosen to be roughly the detector resolution, which translates to a purity (see Eqn. 6.25) of around 60–70% being targeted in each bin.

The MC samples used to construct the response matrices for the IBU method are the combined train splits of the strong Z+jets MadGraph+Pythia8 FxFx sample, the EW Z+jets sample, and the diboson Z+jets sample after they have undergone the positive reweighting detailed in Section 7.7.3. The chosen binning is shown in Table 7.7 and Figures 7.14 and 7.15 show the efficiency (Eqn. 6.20), purity (Eqn. 6.25) and fiducial factor (Eqn. 6.24) in each bin for each observable. The migration matrices (Eqn. 6.21) for each observable are then shown in Figures 7.16 to 7.19. As discussed, the purity is chosen to be between roughly 60–70% in each bin of the distributions as

can be seen in Figures 7.14 and 7.15. The efficiency is driven largely by the proper reconstruction of both muons, as evidenced by the drop in efficiency in the central η region due to the lack of instrumentation in that area of the MS (see Section 5.2), and the dependence on the muon p_{T}. Overall, the total per-event efficiency is typically within the 70–85% range. The fiducial factors quantify the per-bin probability that a reconstructed event is not the result of a fake or an out-of-fiducial truth level event. This is most prominently demonstrated by the reduction in the fiducial factor in the first bin of the $p_{\mathrm{T}}^{\mu\mu}$ distribution, where events have a truth level value that falls outside of the fiducial volume, but due to resolution effects the reconstructed value will enter the analysis phase space. Overall, the per-event fiducial factor is typically within the 90–99% range.

The method is implemented using the RooUnfold [153] package. After examining the increase in uncertainties versus the decrease in systematic bias over different numbers of iterations, the choice was made to run IBU for two iterations. The contribution from the top background is neglected for IBU, as the impact on the unfolded result due to its inclusion is smaller than the statistical errors and is negligible in most regions of the phase space.

7.8.1.1 Closure tests

To ensure that the implementation of the IBU method is functioning as intended, a technical closure test is performed. This is done by unfolding the MadGraph+Pythia8 FxFx plus non-strong MC sample using the response matrix derived from the same sample. If the unfolding is working as intended, the unfolded result will exactly match the truth MC distributions. Perfect closure is observed for all observables, a subset of which are shown in Figures 7.20 and 7.21.

Observable	**Bin Edges**
$p_{\mathrm{T}}^{\mu\mu}$ [GeV]	190 230 300 450 600 1000
$y_{\mu\mu}$	-2.5 -2 -1.5 -1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1 1.5 2 2.5
$p_{\mathrm{T}}^{\mu 1}$ [GeV]	25 125 200 300 400 600 800
$p_{\mathrm{T}}^{\mu 2}$ [GeV]	25 50 75 100 150 200 300 400
$\eta_{\mu 1}$	-2.5 -2 -1.5 -1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1 1.5 2 2.5
$\eta_{\mu 2}$	-2.5 -2 -1.5 -1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1 1.5 2 2.5
$\phi_{\mu 1}$	-3.2 -2.8 -2.4 -2.0 -1.6 -1.2 -0.8 -0.4 0 0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2
$\phi_{\mu 2}$	-3.2 -2.8 -2.4 -2.0 -1.6 -1.2 -0.8 -0.4 0 0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2
Leading track jet p_{T} [GeV]	0 50 100 150 200 300 1000
Subleading track jet p_{T} [GeV]	0 25 50 100 500
Leading track jet y	-2.5 -2.25 -2.0 -1.75 -1.5 -1.25 -1.0 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1.0 1.25 1.5 1.75 2.0 2.25 2.5
Subleading track jet y	-2.5 -2 -1.5 -1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1 1.5 2 2.5
Leading track jet ϕ	-3.2 -2.8 -2.4 -2.0 -1.6 -1.2 -0.8 -0.4 0 0.4 0.8 1.2 1.6 2.0 2.4 2.8 3.2
Subleading track jet ϕ	-3.2 -2.8 -2.4 -2.0 -1.6 -1.2 -0.8 -0.4 0 0.4 0.8 1.2 1.6 2.0 2.4 2.6 2.8 3.0 3.2
Leading track jet m [GeV]	0 8 16 24 32 42 70
Subleading track jet m [GeV]	0 2.5 7 12 20 40
Leading track jet n_{ch}	1 7 11 15 20 30 40
Subleading track jet n_{ch}	1 4 7 11 15 20 30 55
Leading track jet τ_1	0 0.05 0.1 0.17 0.25 0.35 0.5 0.9
Subleading track jet τ_1	0 0.1 0.2 0.35 0.5 0.9
Leading track jet τ_2	0 0.025 0.05 0.08 0.12 0.17 0.25 0.5
Subleading track jet τ_2	0 0.025 0.1 0.17 0.25 0.5
Leading track jet τ_3	0 0.025 0.05 0.1 0.2 0.3
Subleading track jet τ_3	0 0.025 0.08 0.14 0.2 0.3

Table 7.7: The chosen bin edges for each of the 24 observables examined in this analysis.

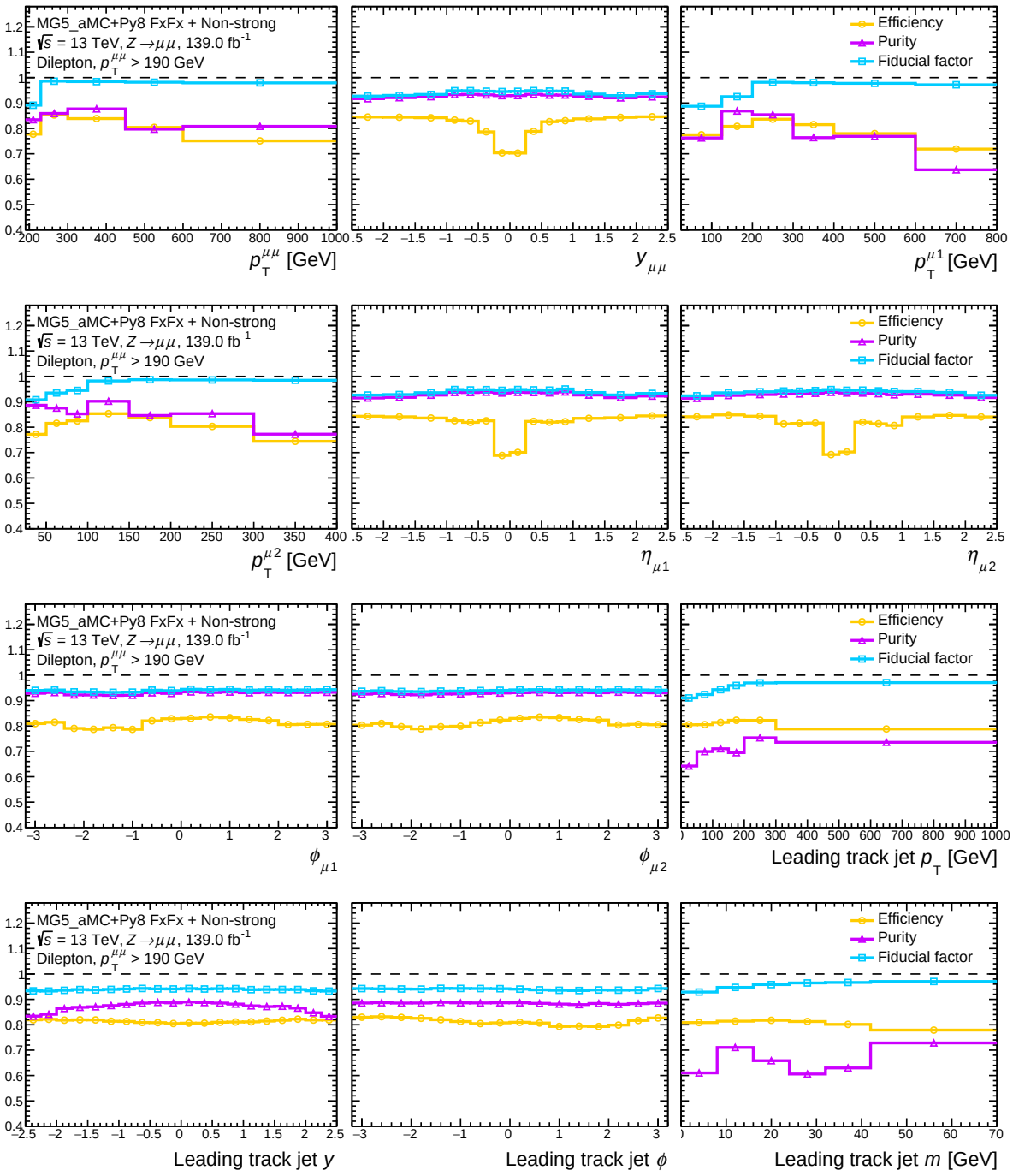


Figure 7.14: The efficiency, purity and fiducial factor as defined in Section 6.2.2 for the dimuon p_{T} and y, the p_{T}, η and ϕ for the leading and subleading muon, and the p_{T}, y, ϕ and m of the leading track jet. The binning reported in Table 7.7 is used.

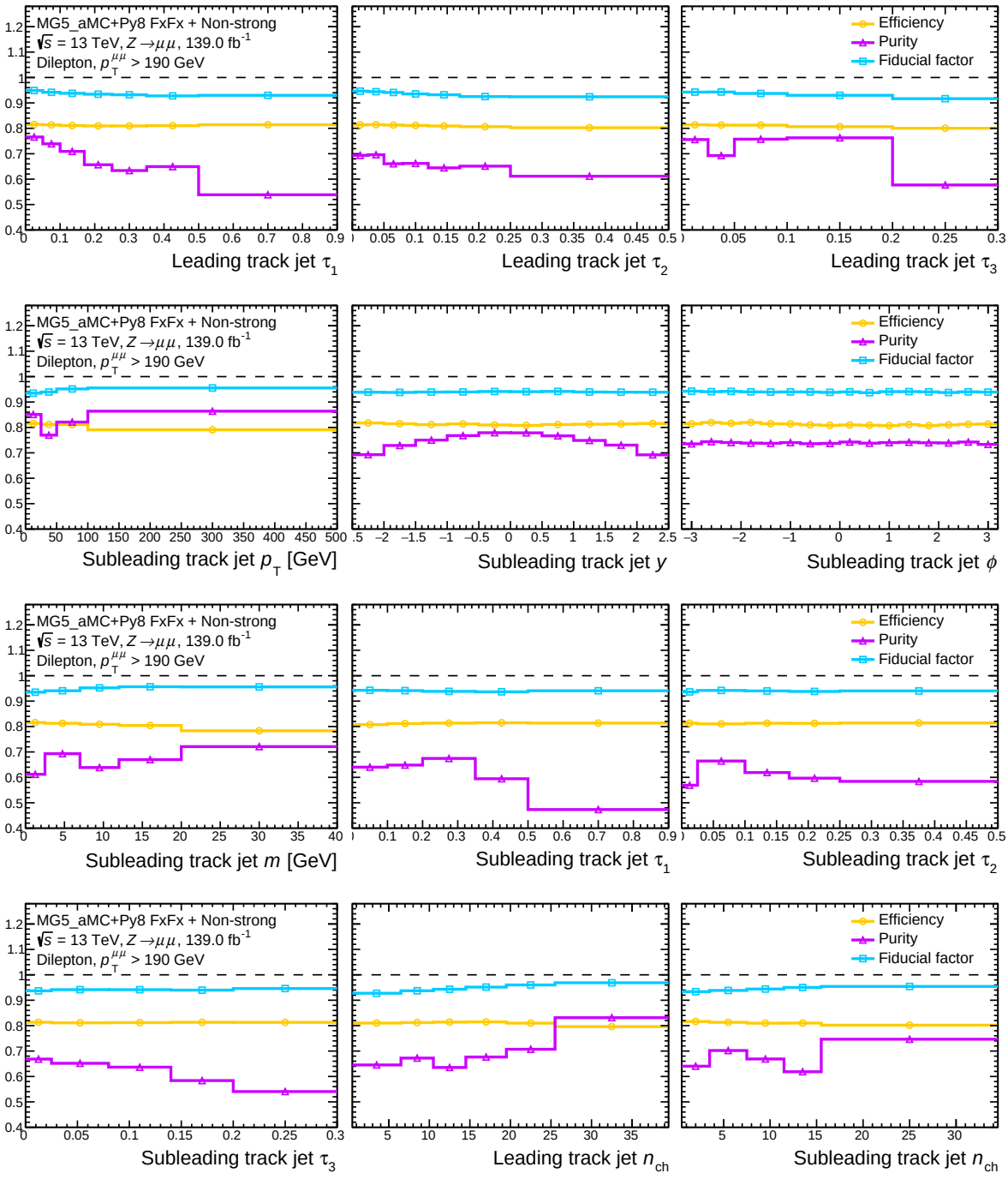


Figure 7.15: The efficiency, purity and fiducial factor as defined in in Section 6.2.2 for the subleading track jet p_{T}, y, ϕ and m, and τ_1, τ_2, τ_3 and the number of constituents for the leading and subleading track jet. The binning reported in Table 7.7 is used.

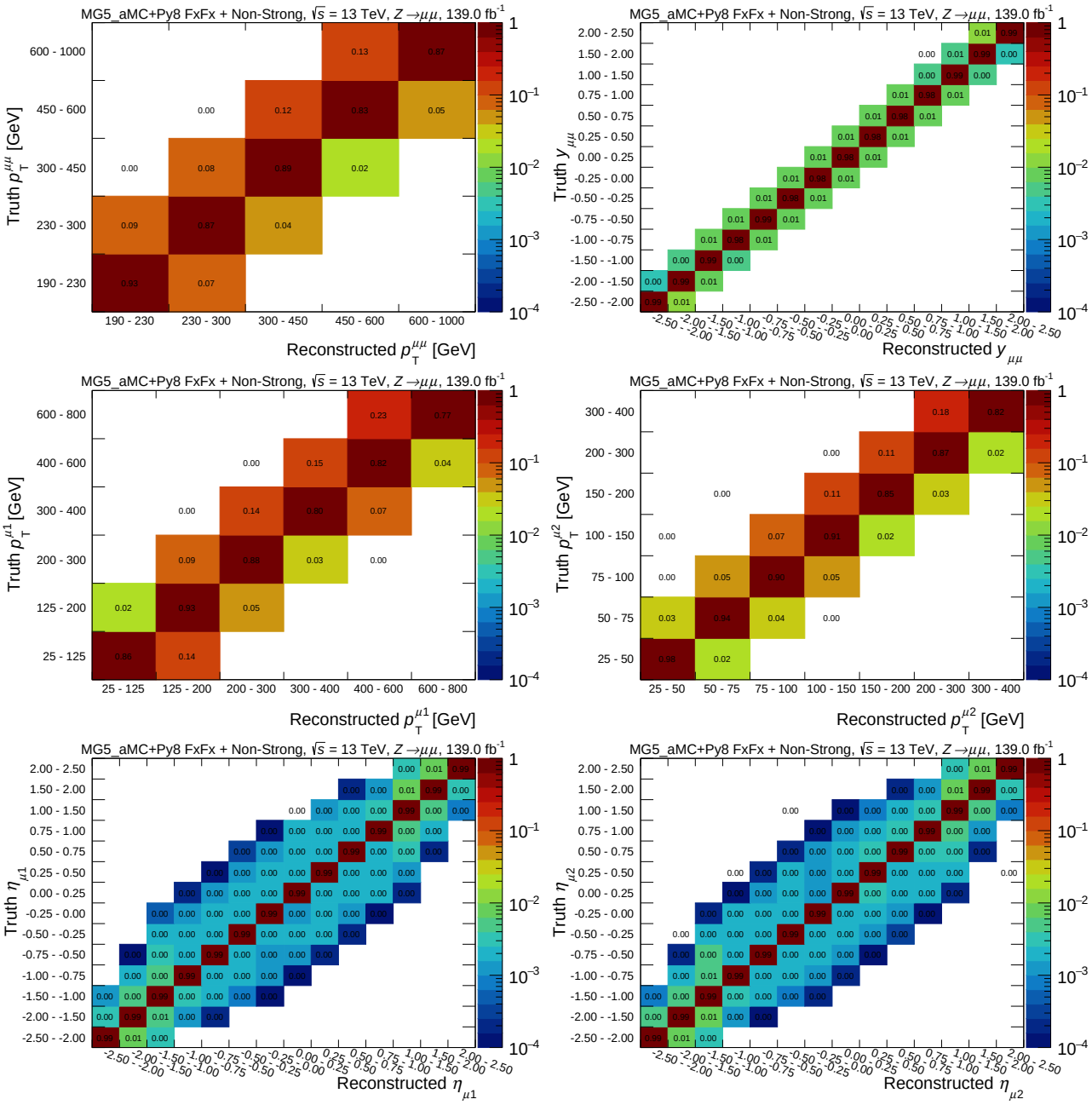


Figure 7.16: The migration matrices as defined in Section 6.2.2 for the dimuon p_{T} and y, and the p_{T} and η of the leading and subleading muon. One can observe that each of the rows sum to one as expected.

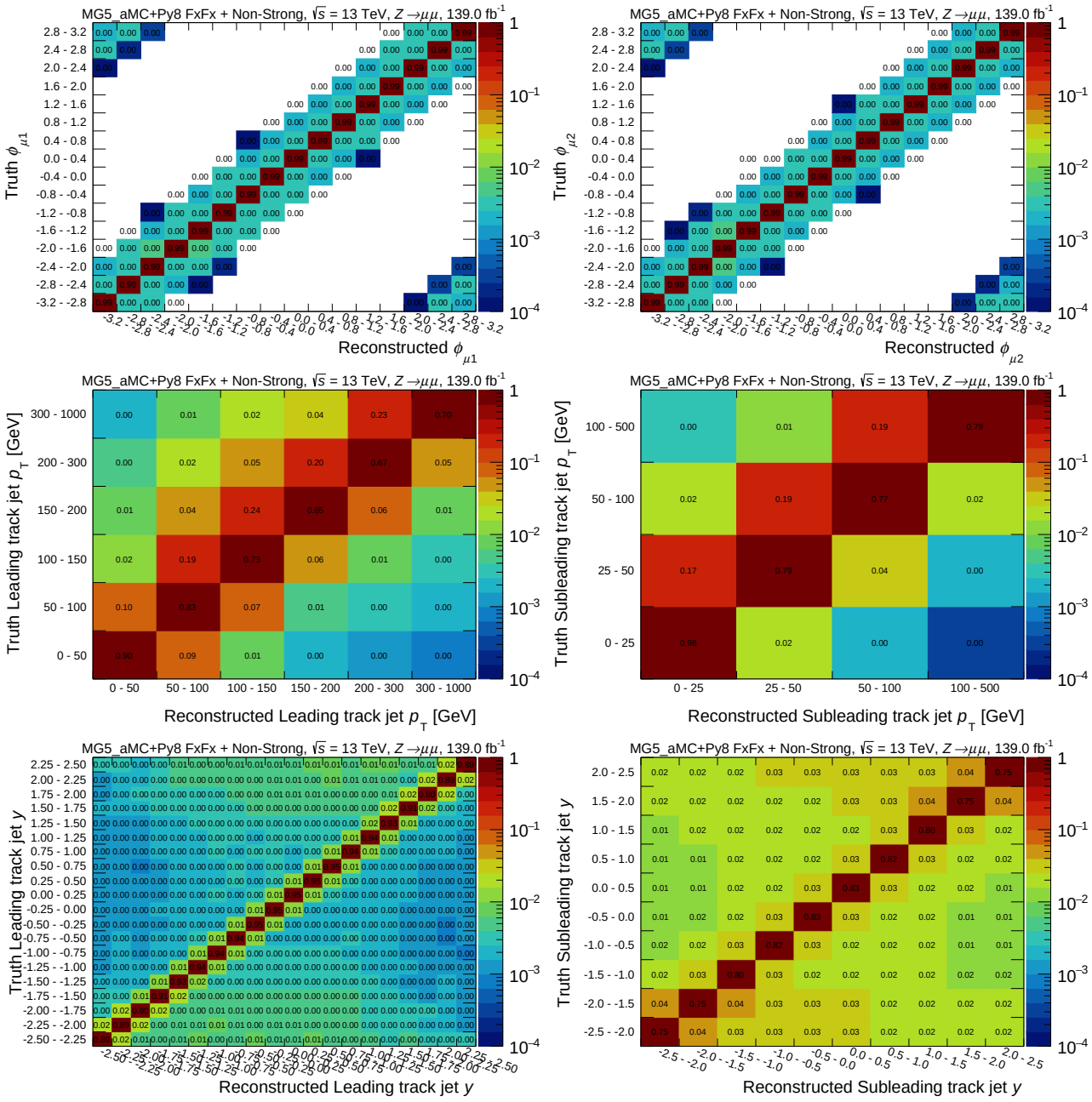


Figure 7.17: The migration matrices as defined in Section 6.2.2 for the ϕ of the leading and subleading muon, and the p_{T} and y of the leading and subleading track jet.

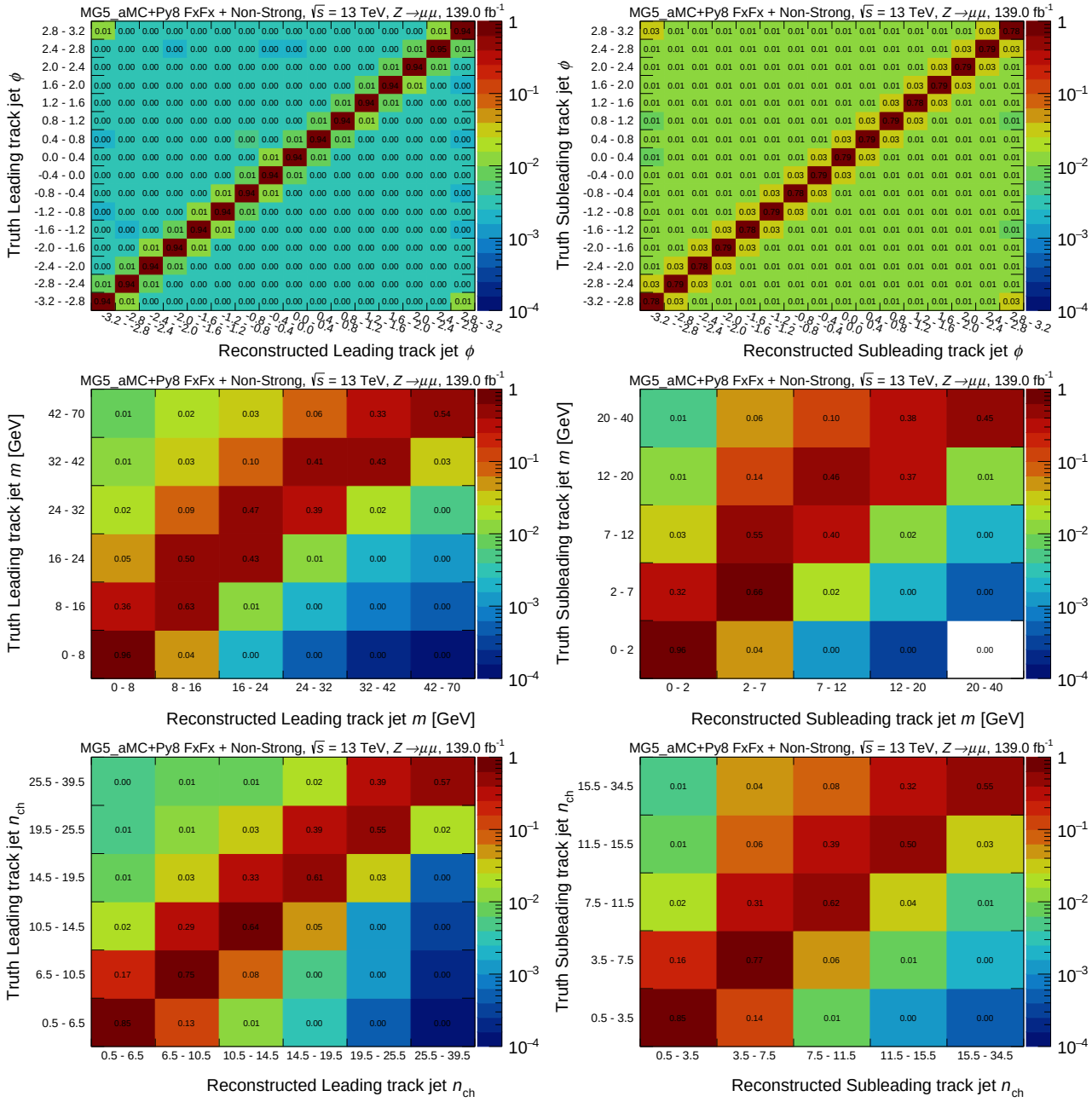


Figure 7.18: The migration matrices as defined in Section 6.2.2 for the ϕ, m and number of constituents for the leading and subleading track jet.

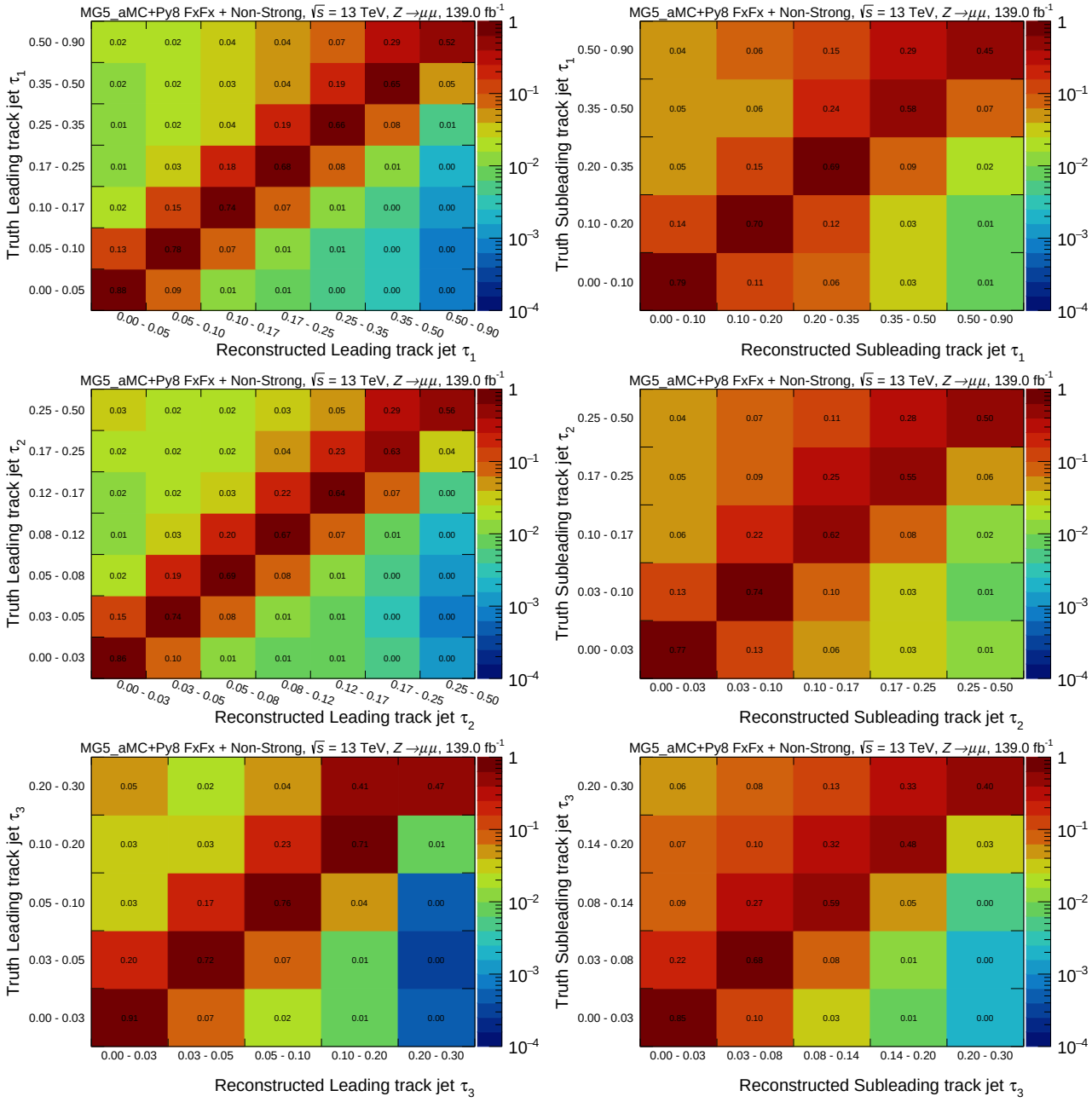


Figure 7.19: The migration matrices as defined in Section 6.2.2 for the τ_1, τ_2 and τ_3 for the leading and subleading track jet.

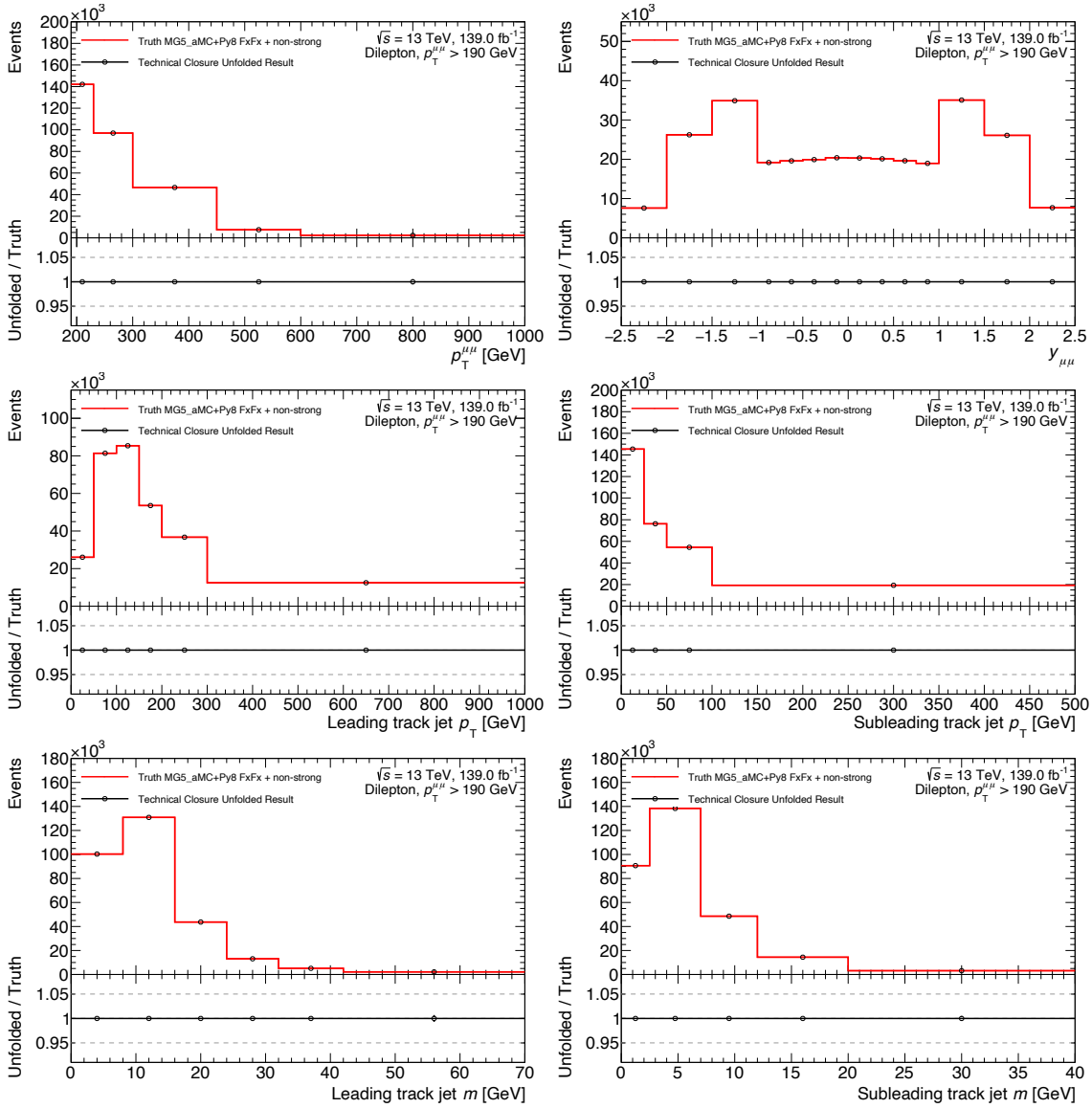


Figure 7.20: The results of performing a technical closure test for the IBU method. The dimuon p_{T} and y, and the p_{T} and m of the leading and subleading track jet are unfolded. The MadGraph+Pythia8 FxFx plus non-strong MC sample is used to construct the response matrix and the detector level MC is used as the input to the unfolding. If the method is performing as expected, the unfolded result (black) will match the truth MC distribution (red). As expected, perfect agreement is observed.

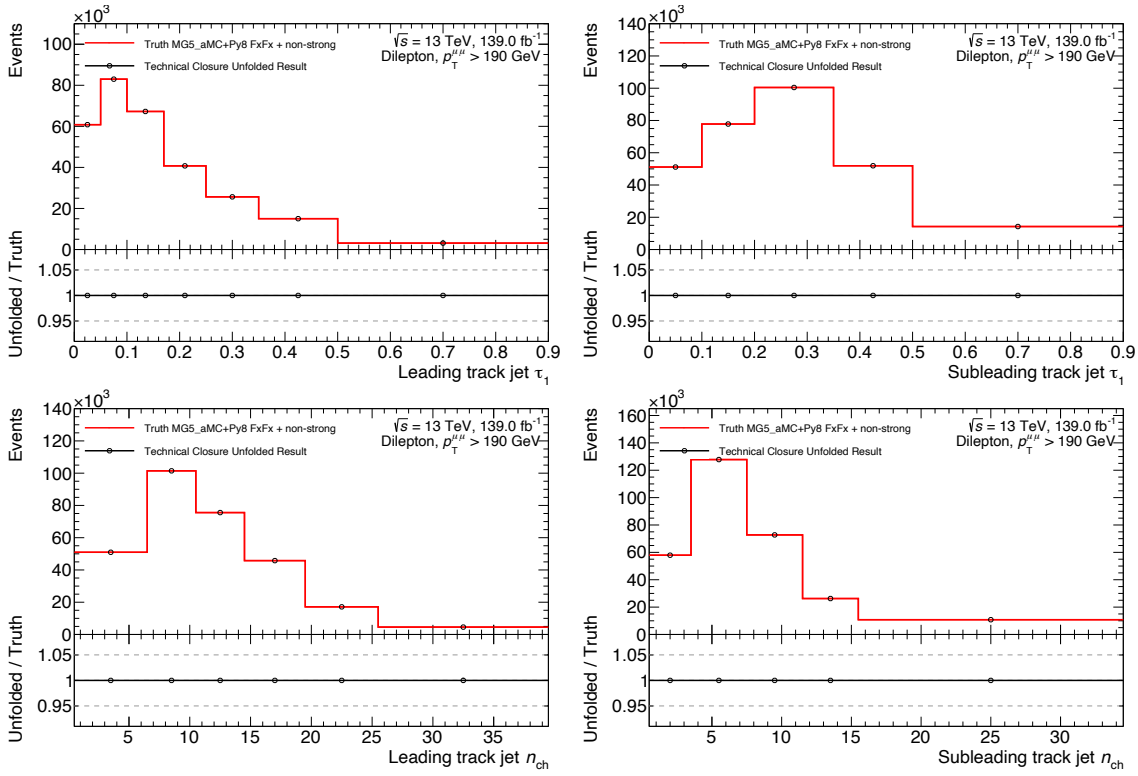


Figure 7.21: The results of performing a technical closure test for the IBU method. τ_1 and the number of constituents in the leading and subleading track jet are unfolded. The MadGraph+Pythia8 FxFx plus non-strong MC sample is used to construct the response matrix and the detector level MC is used as the input to the unfolding. If the method is performing as expected, the unfolded result (black) will match the truth MC distribution (red). As expected, perfect agreement is observed.

7.8.1.2 Uncertainties

The uncertainties are evaluated as described in Section 7.6, and are individually propagated through the unfolding procedure to produce a varied result. The uncertainty on the measurement due to that specific source of uncertainty is then determined by taking the difference between the varied unfolded result and the nominal unfolded result.

Firstly, the experimental systematic uncertainties (introduced in Section 7.6.1) and the theoretical systematic uncertainties (introduced in Section 7.6.2) result in

changes to the MC sample. Therefore, they impact the response matrix, as well as the efficiencies and the fiducial factors, and thus produce a varied result when propagated through the unfolding procedure. It is worth noting that since the theoretical uncertainties modify the MC event weight they affect the reconstructed and truth level MC distributions in similar ways, and so the resulting response matrix is typically very similar to the nominal case.

The statistical uncertainty (as discussed in Section 7.6.3) on the MC sample results in a change to the MC event weights, and thus changes to the response matrix, the efficiencies, and fiducial factors. As this analysis uses 100 replica datasets, this results in 100 variations to the MC inputs and a corresponding 100 variations to the unfolded result. The total uncertainty due to the MC statistics is then determined by taking the RMS of the 100 variations in each bin of the unfolded distribution. The replica datasets for the data, meanwhile, will result in a change to the initial (pseudo)data distribution that is to be unfolded. Thus, in order to determine the uncertainty due to the data statistics, each replica dataset is unfolded using the nominal response matrix from the MC. The standard deviation of the 100 variations in each bin of the unfolded distribution is then taken to be the uncertainty due to the data statistics.

Lastly, there are uncertainties associated with the unfolding procedure that have yet to be discussed. The standard procedure used by the ATLAS Collaboration considers two uncertainties [126]: the basic unfolding uncertainty (also known as the data-driven closure test) and the hidden variable uncertainty. The data-driven closure test is designed to examine the sensitivity of the unfolding method to shape differences between the data and the MC sample. In order to accomplish this, the truth distribution for each observable is reweighted such that the reconstructed MC distri-

bution better matches the (pseudo)data distribution. The reweighted reconstructed level MC is then unfolded using the nominal response matrix. Ideally, the unfolded result will match the reweighted truth MC distribution. The difference between the unfolded result and the reweighted truth distribution is taken to be the uncertainty.

The hidden variable uncertainty is designed to check for any dependence on variables that effect the detector response but are not included in the unfolding procedure. In order evaluate this, the alternate Sherpa 2.2.11 sample is reweighted such that the truth distribution matches that of the MadGraph+Pythia8 FxFx sample. The reweighted Sherpa 2.2.11 reconstructed distribution is then unfolded using the nominal response matrix. The uncertainty is then taken to be the difference between this result and the reweighted Sherpa 2.2.11 truth distribution.

7.8.2 MultiFold implementation

This section presents the procedure implemented to obtain the measurement using the MultiFold method (introduced in Section 6.3), which constitutes the main result for this analysis. The MultiFold algorithm was discussed in detail in Section 6.3. Here, its implementation and its differences from IBU will be highlighted for clarity before showing the technical closure test results and describing the uncertainties.

The neural networks implemented for MultiFold have three hidden layers with 200 nodes each. The ReLu activation function is used for the internal layers and the sigmoid activation function is used for the final layer. Training proceeds for 200 epochs with early stopping using a patience of 10 epochs, i.e. if the validation loss does not improve for 10 epochs then the training ends and the network weights with the lowest validation loss are saved. The input training data is randomly divided into two parts, with 75% to be used in training the network and 25% to be used for

validation. The input features are standardized by subtracting the mean and dividing by the standard deviation. In order to mitigate difficulties encountered due to the neural network struggling to learn the wrap around effect, $(\phi = 0 = 2\pi)$[2], the choice was made to decompose the ϕ observables in this analysis into $\phi \rightarrow (\sin\phi, \cos\phi)$. The neural networks are implemented with Keras [154] and Tensorflow [155] and optimization is performed with Adam [156]. As previously discussed, the chosen loss function is the binary cross entropy, which is necessary for the density ratio estimation to function.

MultiFold is trained for five iterations of the algorithm, using the portion of the MadGraph+Pythia8 FxFx sample set aside for training (the train split with the positive reweighting applied, as discussed in Sections 7.7.1 and 7.7.3) as input. The model from the final iteration is then saved and applied to the part of the MadGraph+Pythia8 FxFx sample set aside for the test split. The results using the test split constitute the unfolded results for the MultiFold method. This is done as an additional measure to eliminate any potential bias from overtraining, and to provide similar statistics to the data sample. The vast majority of the output MultiFold weights are small. However, in some cases these weights may be large or in rare cases even infinite in regions where phase space coverage differs between samples. The MultiFold weights are thus capped at a value of 100. This cap only affects around 0.003% of events.

One major difficulty encountered with regard to the utilization of neural networks is that they are randomly initialized. This means that the weights in Eqn. 6.34 are typically initialized to small positive random numbers before training begins. This

[2]Due to the nature of the ATLAS coordinate system, the ϕ coordinate varies from 0 to 2π. This has the effect that two particles that are in fact very close together in ϕ, but are situated at $\phi_1 = 0.1$ and $\phi_2 = 6.0$ for example, will appear to the neural network as being very far away from each other. This was found to result in biases in the training. By inputting these observables into the NN training in terms of $\sin\phi$ and $\cos\phi$ the NN is able to interpret this effect more correctly.

randomness will ultimately result in a different unfolded result each time the network is trained with the same input information, which is undesirable. In order to mitigate this effect, 100 randomly initialized models are trained to create an ensemble, and the nominal result is taken to be the median weight from the final ensemble of random initializations. An additional uncertainty is also assigned in order to quantify the spread of these networks. It is discussed in more detail below, however it should be noted that this additional uncertainty does result in a total uncertainty that is slightly larger than what is seen on the IBU result. Computing these additional models also considerably increases the computing time and resources required to complete an analysis such as this, particularly since this ensembling technique must also be used for each of the systematic uncertainty variations, including the MC and data bootstraps.

A final point on the implementation of the MultiFold method is that of the normalization. As the input samples are standardized such that the mean of the weights in each sample is equal to unity, the output will not be normalized to the cross section. This is thus achieved by utilizing Eqn. 6.5 to obtain a normalization factor,

$$w_i^{\text{norm}} = \frac{w_i^{\text{MF}}}{\sum_i w_i^{\text{MF}}} \frac{f^{\text{fid}}}{\varepsilon} n^{\text{data}}, \tag{7.8}$$

where w_i^{MF} is the weight produced by the MultiFold algorithm for an event i, and ε and f^{fid} are as previously defined in Eqns. 6.20 and 6.24 and are calculated using the MC samples used as input to the MultiFold algorithm. After the normalization is applied the sum of the normalized weights will give the fiducial cross section if inserted into Eqn. 6.5, as desired. It should be noted that as part of the MultiFold procedure, this normalization factor will be affected by the uncertainties detailed in

the dedicated section below.

Unlike in the IBU case, the nominal input MC sample used for MultiFold is the strong Z+jets MadGraph+Pythia8 FxFx sample. The inclusion of the EW and diboson Z+jets events were found to produce a result that was in general agreement within statistical uncertainties with the result obtained with the strong-only unfolding, and no clear improvement to the agreement with the target when used to unfold the psuedodata sample.

For emphasis, although the MultiFold results are displayed in one-dimensional histograms, the actual output from the MultiFold algorithm is a set of weights to be applied to the MadGraph+Pythia8 FxFx test sample. This reweighted MC sample is the unfolded result, and thus can be freely rebinned or used to calculate additional observables from the 24 included in the unfolding.

7.8.2.1 Closure tests

As was done with IBU, the implementation of the MultiFold method is tested by performing a technical closure test. In this case, the train split of the MadGraph+Pythia8 FxFx sample is used to train an ensemble of 25 models to be used for the unfolding with the same sample used as the input data. The model is then used to unfold the test split of the MadGraph+Pythia8 FxFx sample and the result of the unfolding is compared with the truth level MadGraph+Pythia8 FxFx test sample. If the method is working as intended, the unfolded result and the truth-level distribution should agree. The results of the technical closure test for each of the 24 observables are shown in Figure 7.22, and as expected closure is achieved for each observable.

Figure 7.22: The results of performing a technical closure test for the MultiFold method. The model is trained using the train split of the MadGraph+Pythia8 FxFx sample, and the result is obtained by applying the model to the test split of the same sample. The results of the unfolding are shown in black, while the truth level test split is shown in green. If the method is performing as expected, then these should match, which is observed.

7.8.2.2 Uncertainties

The vast majority of the systematic and statistical uncertainties are propagated in an identical manner to IBU, just with varied input samples instead of a varied response matrix or data histogram. As with the nominal result, each systematic variation is determined using an ensemble of 100 random initializations of the network. There are some subtle differences in the derivation of the unfolding uncertainties between the two methods due to the multidimensional nature of the unfolding, as well as some additional uncertainties unique to the MultiFold method. These will be discussed further in the following.

The uncertainty associated with the data-driven closure test is designed to account for the sensitivity of the unfolding method to differences in shape between the MC and data samples. In the case of the MultiFold algorithm, this uncertainty is determined by running the MultiFold algorithm for an additional 5 iterations for a total of 10. The result obtained after 10 iterations is then compared with the nominal unfolded result and the difference is taken as the uncertainty. The effect of performing the test in this way should be equivalent to the method used in the IBU case, where the reconstructed level MC distributions are reweighted at truth level such that they better match the (pseudo)data before unfolding them with the nominal response matrix. By performing the initial 5 iterations, the reconstructed MC has been made to match the (pseudo)data and performing the additional 5 iterations will give the unfolded result based upon the reweighted starting dataset.

The hidden variable uncertainty is designed to quantify the dependence of the result on any variables that effect the detector response but are not included in the unfolding procedure. It is evaluated in similar fashion to IBU, just with the reweighting applied to the "response matrix" as opposed to the input sample. This produces an

equivalent result, but performing the test in this manner for the MultiFold method is easier to implement. The alternate SHERPA 2.2.11 sample is reweighted at truth level using a neural network such that it better agrees with the MADGRAPH+PYTHIA8 FXFX sample. The (pseudo)data is then unfolded with the reweighted SHERPA 2.2.11 sample and the uncertainty is obtained by comparing this unfolded result with the nominal one.

The unfolding uncertainties are frequently the dominant uncertainties for a given observable. As they are evaluated by switching out one MC sample for another, they are known as two-point systematics, and can be interpreted as a random sampling of the underlying uncertainty pdf. While this is expected to give the correct magnitude, no information regarding the internal correlations of the uncertainty are accessible. This can create challenges as the default assumption is that uncertainties are fully correlated between bins. This is discussed more specifically in Section 8.1.

As discussed at the beginning of this section, the EW $Z(\rightarrow \mu\mu)$+jets sample and the diboson $Z(\rightarrow \mu\mu)$+jets are not included in the MC sample used in the MultiFold algorithm. In order to be conservative, an uncertainty is assigned to account for the effects of these contributions. In order to derive it, the train splits of the MADGRAPH+PYTHIA8 FXFX sample, the EW sample and the diboson sample are combined and used in the NN training to unfold the (pseudo)data. The resulting model is applied to the combined test set and the difference between this result and the nominal one is used as the uncertainty. In a similar fashion, in order to quantify the small error due to the top background the algorithm is run on pseudodata with the top contribution not included. The difference between this and the nominal result is then taken as an uncertainty. The size of this uncertainty gives an estimate of the magnitude of the differential cross section of the top processes entering the analysis

phase space.

Due to the random initialization of the NNs there is some variation in the results obtained from MultiFold. In order to determine the uncertainty due to this, the algorithm is run with 100 random initializations and each set of MultiFold weights is saved. The final uncertainty is calculated after the unfolded result has been binned. It is taken to be the standard error on the mean in each bin, i.e. $\frac{\sigma}{\sqrt{N}}$, where N is the number of random initializations in the ensemble.

Finally, there is an additional statistical uncertainty inherent to the MultiFold method. Since the final result is computed using the test split of the MadGraph+Pythia8 FxFx sample (see Sections 7.7.1 and 7.7.3), an additional uncertainty must be included due to the limited statistics of this sample. This is evaluated according to the usual formula for the error on the number of counts in a weighted sample,

$$\sigma_i = \sqrt{\sum_{j \in i} w_j^2}, \tag{7.9}$$

where i indicates the bin and j represents the individual events.

Chapter 8

Results

This chapter shows the differential cross section results and their associated uncertainties for both the pseudodata described in the previous chapter and the ATLAS Run 2 dataset. The main results are obtained using the MultiFold method (see Sections 6.3.2 and 7.8.2), while the results obtained with the traditional method, IBU (see Sections 6.2.3 and 7.8.1), are also shown for the sake of comparison and validation.

Section 8.1 shows the results using pseudodata. The differential cross section distributions for the 24 observables used as input to the unfolding are presented for both the MultiFold method and IBU. This serves as an important test of the MultiFold method as it allows for a comparison with a known target "truth pseudodata" sample, as well as with a well-established method. This is followed by the results of a series of tests performed using the results obtained from the MultiFold method, including constructing new observables from the results, examining them over kinematic subregions, and constructing a series of multidimensional distributions. The results of these tests are also compared with the target distributions. Finally, the differential cross section results for the ATLAS Run 2 dataset are shown in Section 8.2 for the 24 observables used as input to the unfolding procedure, as well as for three

additional observables that are calculated from the MultiFold result. The theoretical predictions from the MadGraph+Pythia8 FxFx sample and the Sherpa 2.2.11 sample are compared with the differential cross section measurements.

8.1 Results with pseudodata

8.1.1 Results comparison

This section will show the unfolded results for each of the 24 observables using the pseudodata sample as the input to the IBU procedure and the MultiFold algorithm. The results are presented as differential cross sections and will be compared with the target truth pseudodata. In order to quantify the agreement of the result with the target, the χ^2 is calculated using

$$\chi^2 = \boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{D}^T. \tag{8.1}$$

Here, $\boldsymbol{D}$ is a row vector containing the bin-wise difference in the number of events between the unfolded distribution and the target, i.e. the known underlying truth distribution used to generate the pseudodata. $\boldsymbol{C}$ represents the total uncertainty covariance matrix, and is determined by summing the uncertainty covariance matrices from each source of uncertainty.

The results, along with their corresponding p-values, are shown in Figures 8.1 to 8.5. The p-values are also shown tabulated in Table 8.1. The corresponding relative uncertainties as well as the correlation matrices are shown for a subset of the 24 observables in Figures 8.6 to 8.8 and Figures 8.9 to 8.11 for IBU and MultiFold, respectively. The remainder can be found in Appendix C. The p-values are generally larger than 0.5, as opposed to a random uniform distribution between 0 and 1, indi-

cating that the uncertainties are overestimated for both methods. With the exception of poor agreement for the subleading track jet mass and number of constituents in the leading track jet for MultiFold, the unfolded differential cross section distributions are found to be in agreement with the target truth pseudodata distributions. Observing this level of agreement is not unexpected given the amount of observables being examined. Furthermore, the track jet mass is one of the observables with the most pronounced shape difference observed between data and MC (see Figures 7.3 and 7.5).

In many cases, including for the two observables that exhibit poor agreement, the unfolding uncertainty is the dominant one. As discussed in Section 7.8.2.2, the internal correlations inherent to these systematics are not known. As they are taken to be fully correlated between bins when calculating the p-value, this could be unfairly affecting it. If the unfolding uncertainties are considered uncorrelated between bins, then the p-values for the disagreeing subleading track jet mass and number of constituents in the leading track jet would be brought into agreement with 0.0534 and 0.239, respectively. As such, in the case where these uncertainties dominate it may be prudent to explore alternative correlation models when comparing to a target distribution.

These results show that MultiFold gives comparable results to those obtained with the traditional unfolding method of IBU, but with all the additional benefits of the result being high dimensional and unbinned. Some of these additional benefits will be explored in the following section. In most bins, the MultiFold method does result in a slightly higher uncertainty compared to IBU (see Figure 8.12), which is to be expected due to the additional uncertainty sources inherent to the MultiFold algorithm. In regards to the NN initialization uncertainty, it can likely be reduced

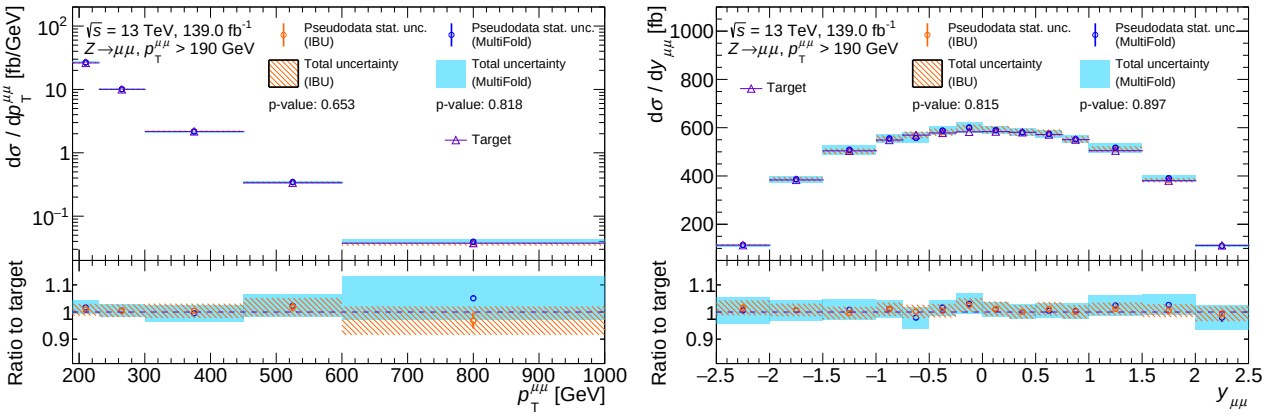


Figure 8.1: The differential cross section as a function of the dimuon p_{T} (left) and y (right). The pseudodata unfolded with IBU is shown in orange with the total uncertainty shown in the hatched band, and the MultiFolded pseudodata is similarly shown in blue. The target truth pseudodata is shown in purple, and the p-value quantifying the agreement between the unfolded result and the target is displayed on the figure for each method.

with the introduction of additional networks in the ensemble. Other ML methods could also be explored to perform the density ratio estimation that may reduce this uncertainty. Another contributor to this increased uncertainty is due to the multi-dimensionality of the result. If one considers the IBU case, the systematics related to the muon calibration procedure will not in principle effect the track jet observables beyond some very small acceptance effects. Similarly, the systematics related to tracking will not impact the muon observables. In MultiFold, this is not the case due to the multidimensionality of the reweighting, and the application of these systematic uncertainties will also result in changes to observables that are not directly linked to the object(s) affected by the systematic in question.

Observable	**MultiFold *p*-value**	**IBU *p*-value**
$p_{\mathrm{T}}^{\mu\mu}$	0.817	0.653
$y_{\mu\mu}$	0.897	0.815
$p_{\mathrm{T}}^{\mu 1}$	0.686	0.549
$p_{\mathrm{T}}^{\mu 2}$	0.733	0.626
$\eta_{\mu 1}$	0.907	0.780
$\eta_{\mu 2}$	0.906	0.931
$\phi_{\mu 1}$	0.999	0.918
$\phi_{\mu 2}$	0.971	0.958
Leading track jet p_{T}	0.976	0.793
Subleading track jet p_{T}	0.805	0.628
Leading track jet y	0.999	0.991
Subleading track jet y	0.955	0.958
Leading track jet ϕ	0.991	0.909
Subleading track jet ϕ	0.984	0.896
Leading track jet m	0.921	0.0506
Subleading track jet m	0.0173 (0.0534)	0.301
Leading track jet n_{ch}	0.00845 (0.239)	0.274
Subleading track jet n_{ch}	0.901	0.567
Leading track jet τ_1	0.893	0.875
Subleading track jet τ_1	0.914	0.603
Leading track jet τ_2	0.919	0.579
Subleading track jet τ_2	0.922	0.953
Leading track jet τ_3	0.965	0.922
Subleading track jet τ_3	0.912	0.986

Table 8.1: The p-values that result from performing a χ^2 test to test the agreement between the unfolded pseudodata results obtained with MultiFold or IBU and the target truth pseudodata distribution. Agreement is achieved within 2σ for all observables except for the subleading track jet mass and the number of constituents in the leading track jet in the MultiFold case. The values in parentheses indicate the p-value when the unfolding uncertainties are decorrelated, as discussed in the main text.

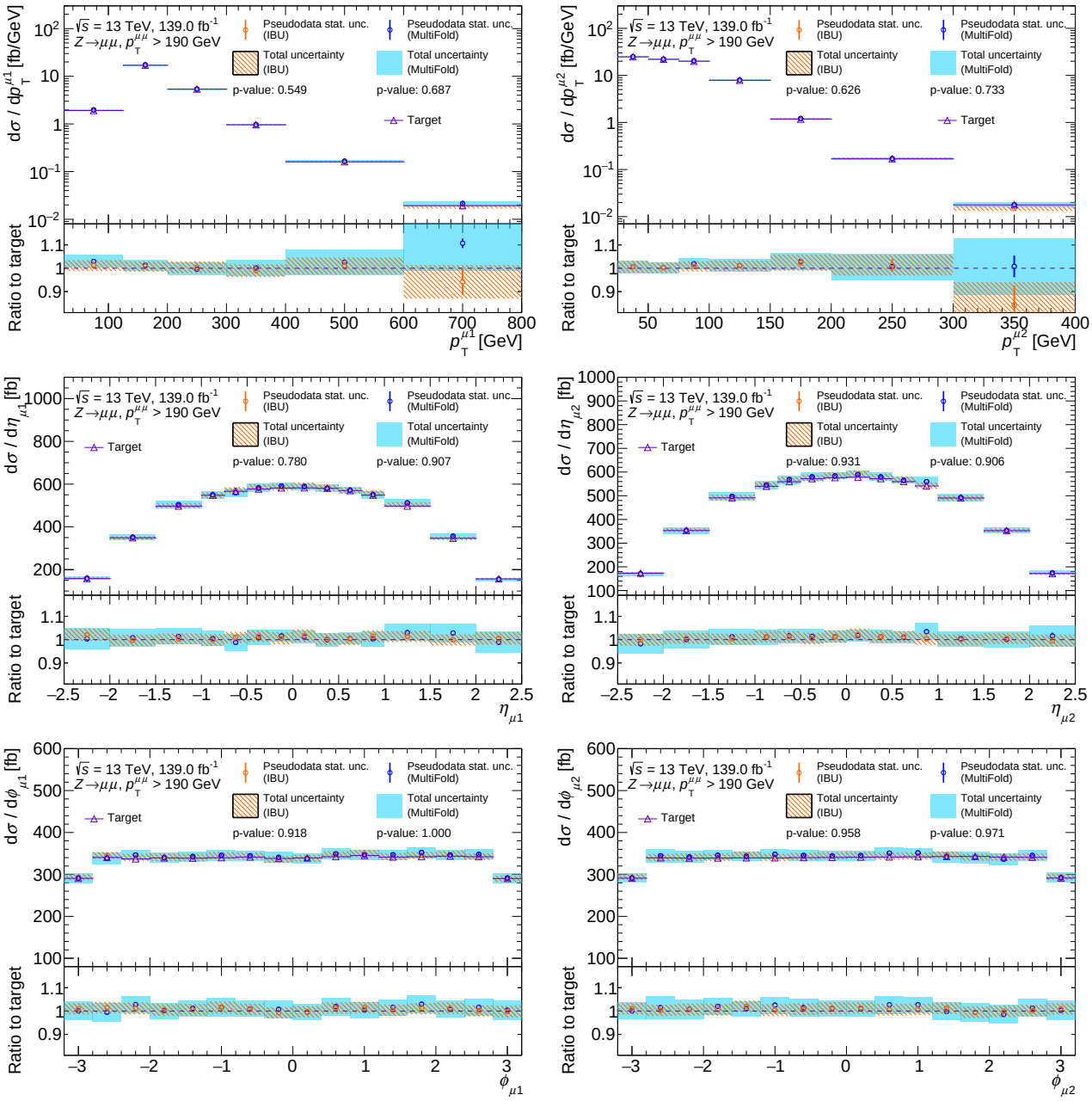


Figure 8.2: The differential cross section as a function of the leading and subleading muon p_{T}, η and ϕ. The pseudodata unfolded with IBU is shown in orange with the total uncertainty shown in the hatched band, and the MultiFolded pseudodata is similarly shown in blue. The target truth pseudodata is shown in purple, and the p-value quantifying the agreement between the unfolded result and the target is displayed on the figure for each method.

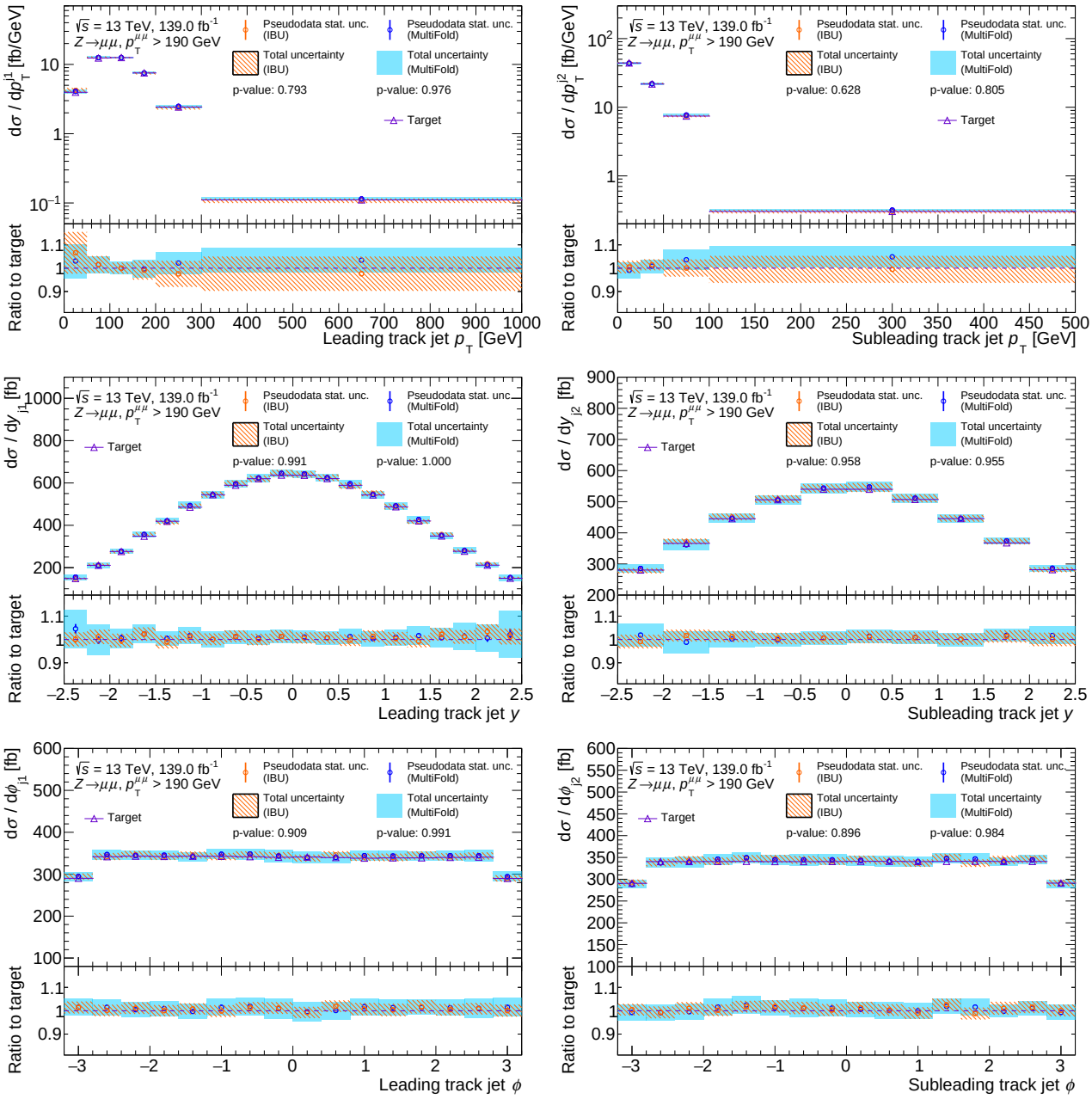


Figure 8.3: The differential cross section as a function of the leading and subleading track jet p_{T}, y and ϕ. The pseudodata unfolded with IBU is shown in orange with the total uncertainty shown in the hatched band, and the MultiFolded pseudodata is similarly shown in blue. The target truth pseudodata is shown in purple, and the p-value quantifying the agreement between the unfolded result and the target is displayed on the figure for each method.

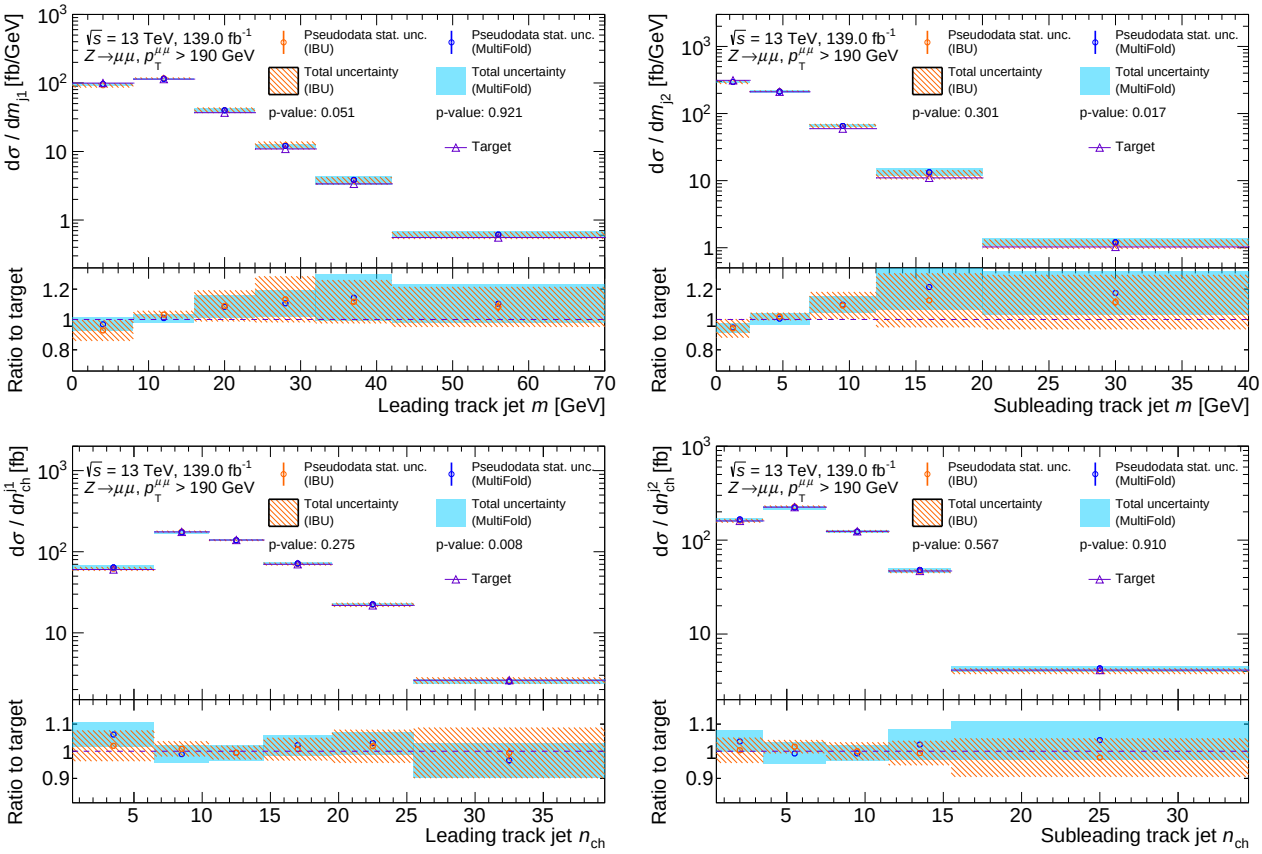


Figure 8.4: The differential cross section as a function of the leading and subleading track jet m and number of constituents. The pseudodata unfolded with IBU is shown in orange with the total uncertainty shown in the hatched band, and the MultiFolded pseudodata is similarly shown in blue. The target truth pseudodata is shown in purple, and the p-value quantifying the agreement between the unfolded result and the target is displayed on the figure for each method.

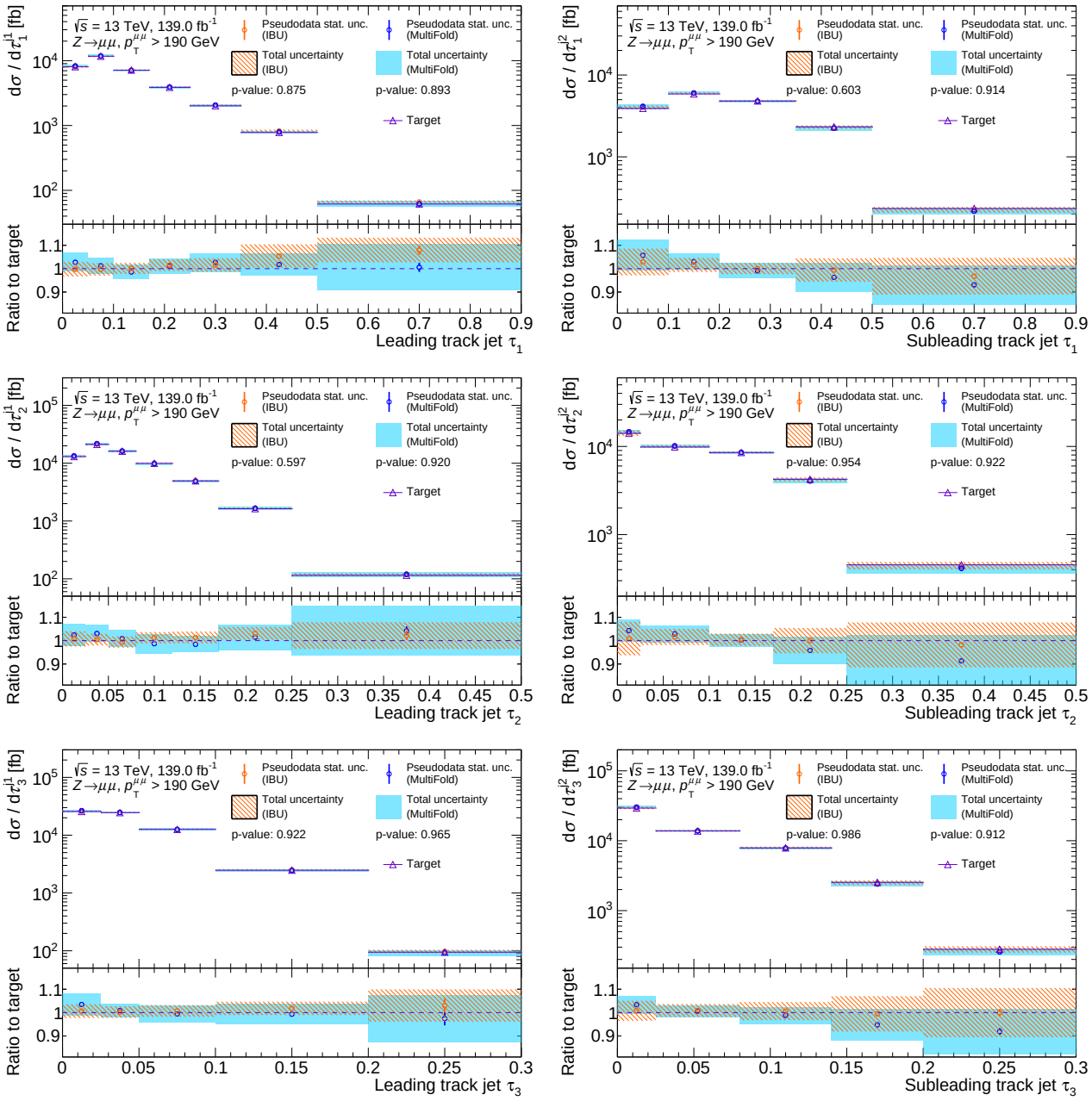


Figure 8.5: The differential cross section as a function of the leading and subleading track jet τ_1, τ_2 and τ_3. The pseudodata unfolded with IBU is shown in orange with the total uncertainty shown in the hatched band, and the MultiFolded pseudodata is similarly shown in blue. The target truth pseudodata is shown in purple, and the p-value quantifying the agreement between the unfolded result and the target is displayed on the figure for each method.

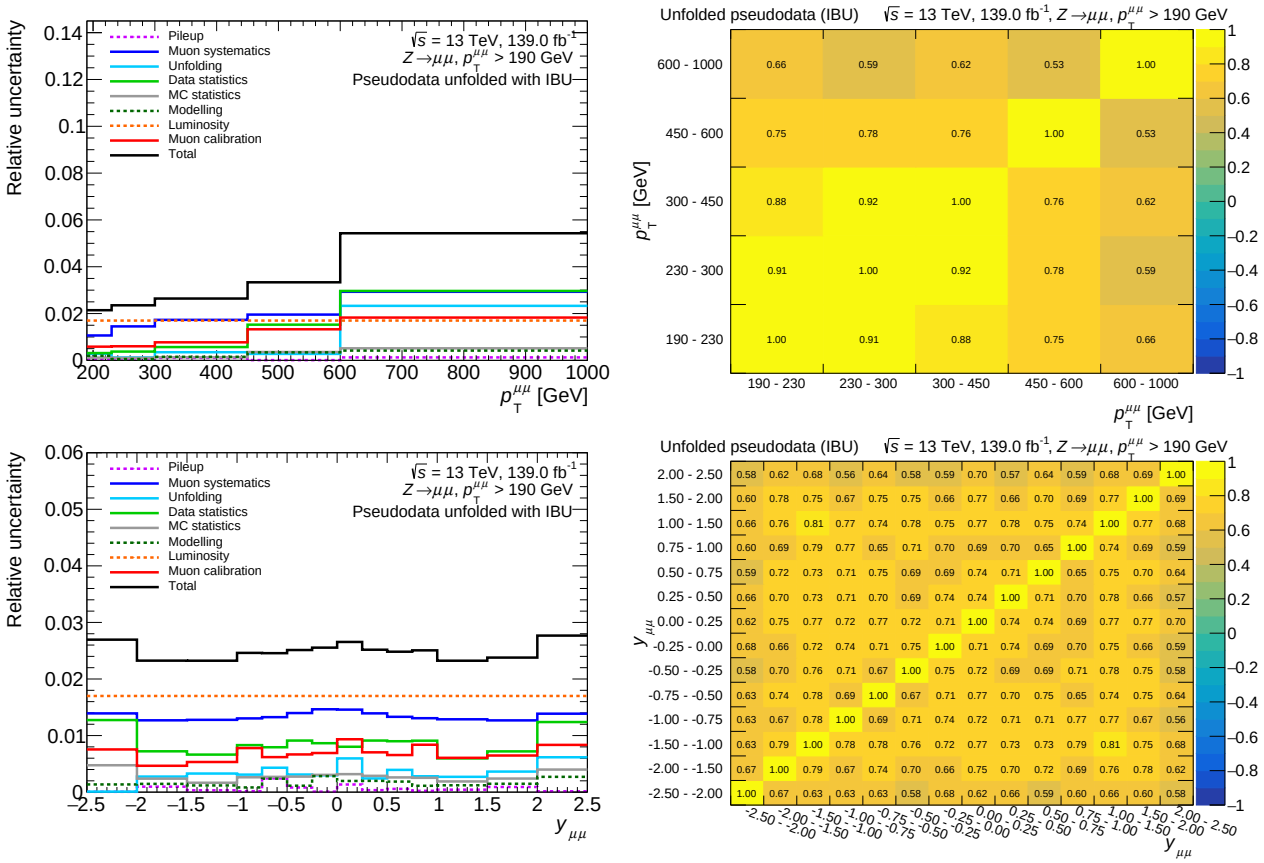


Figure 8.6: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the dimuon p_{T} and y broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

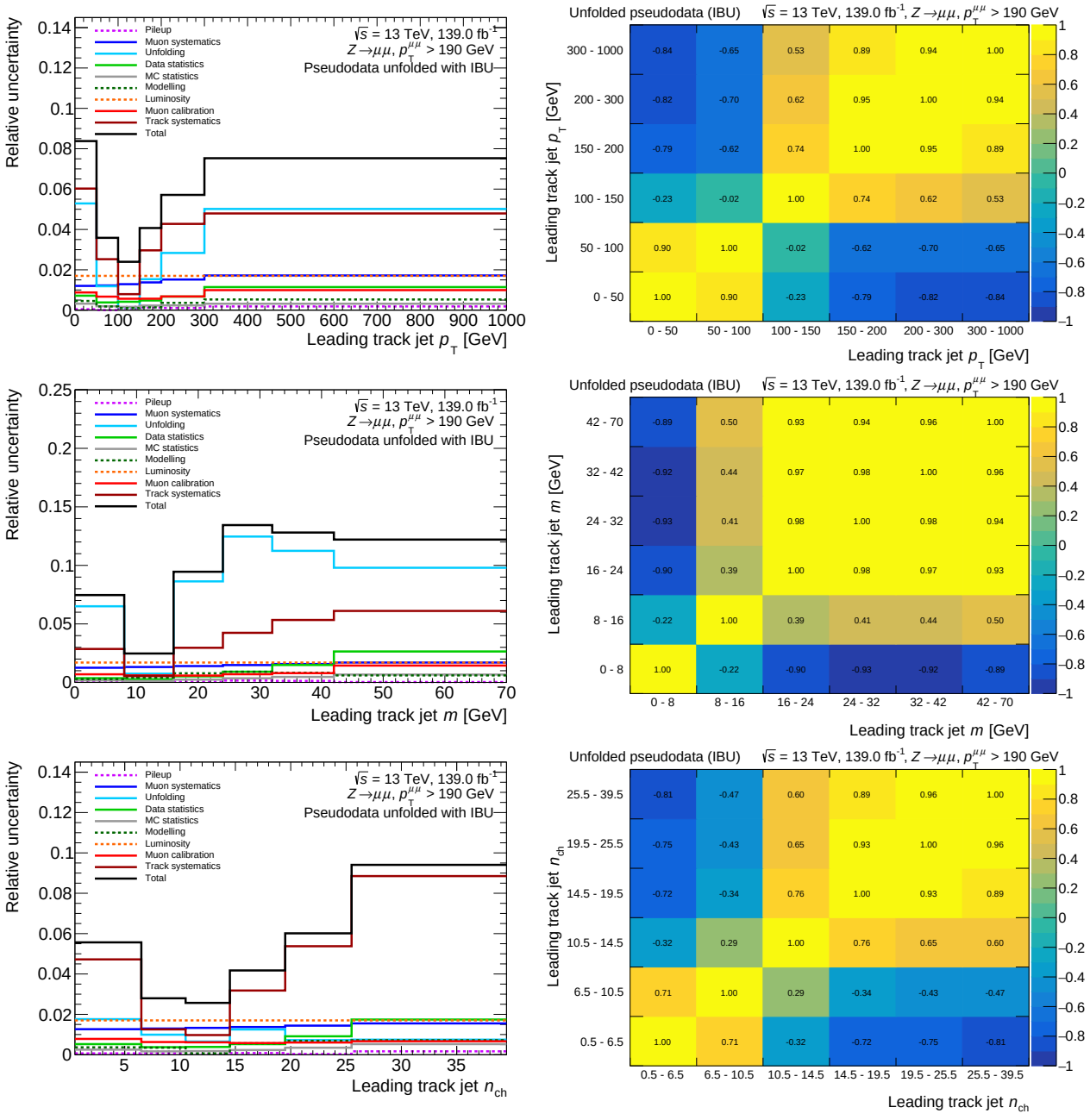


Figure 8.7: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading track jet p_{T}, m and number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

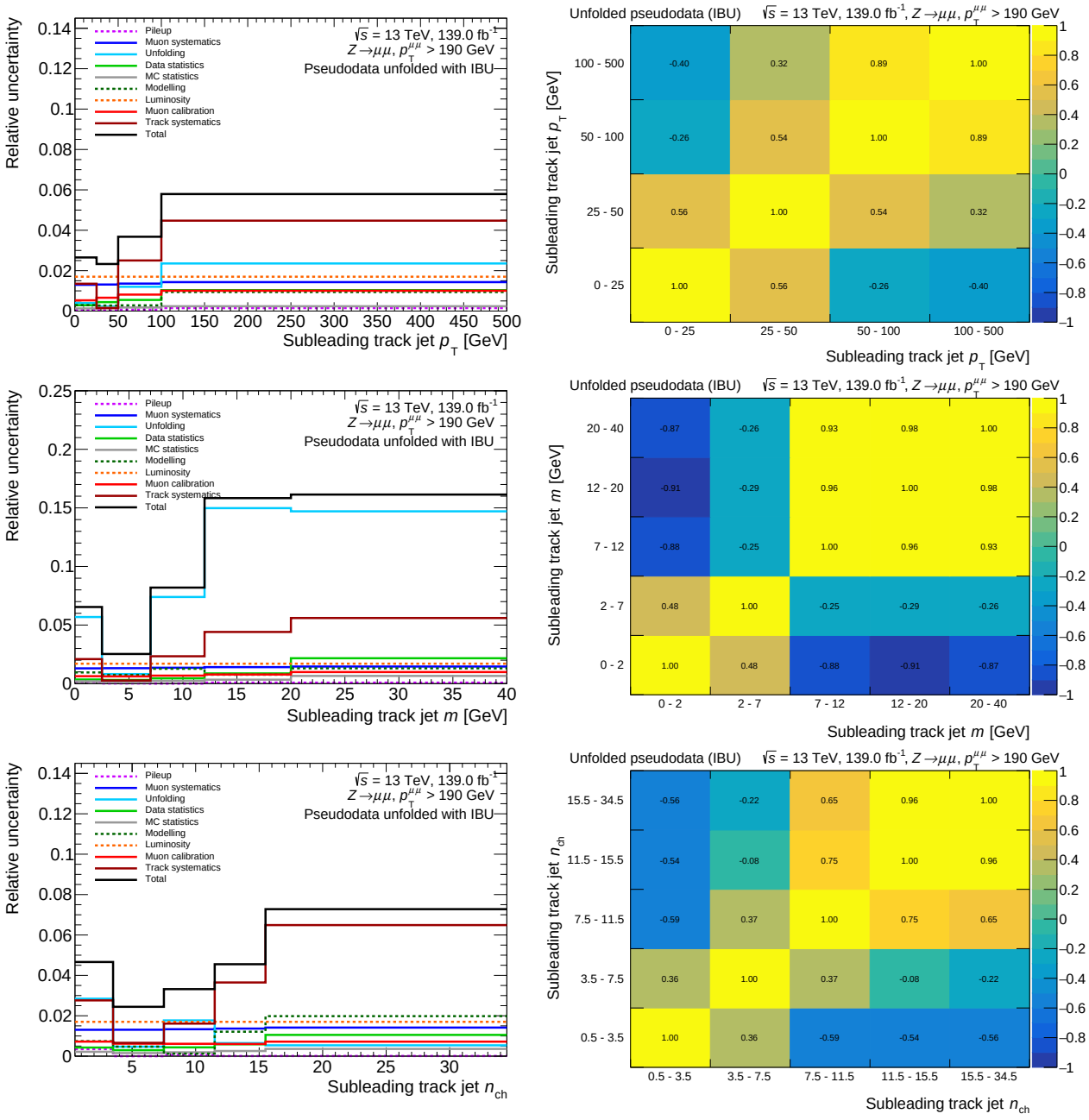


Figure 8.8: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading track jet p_{T}, m and number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

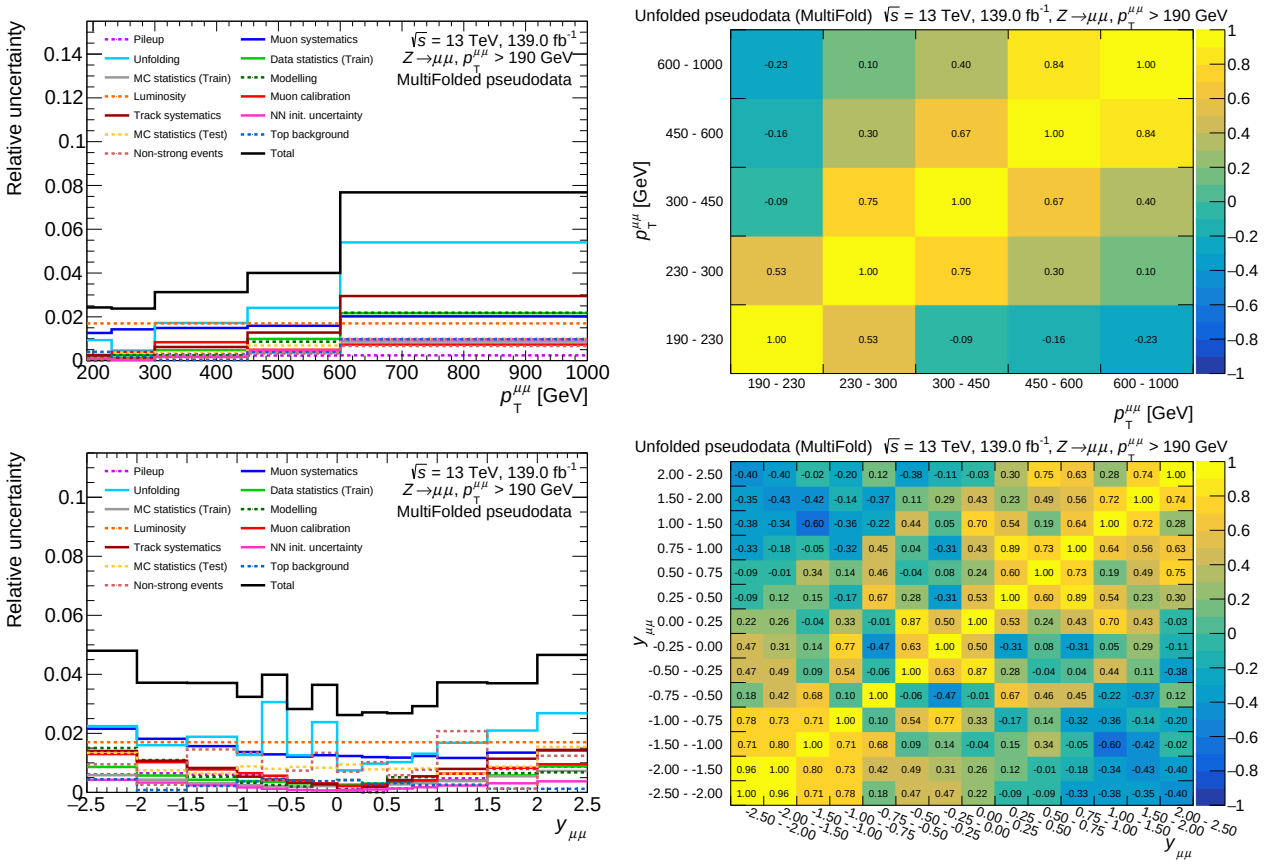


Figure 8.9: The plots on the left show the relative uncertainty on the Multifolded result for the dimuon p_{T} and y broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

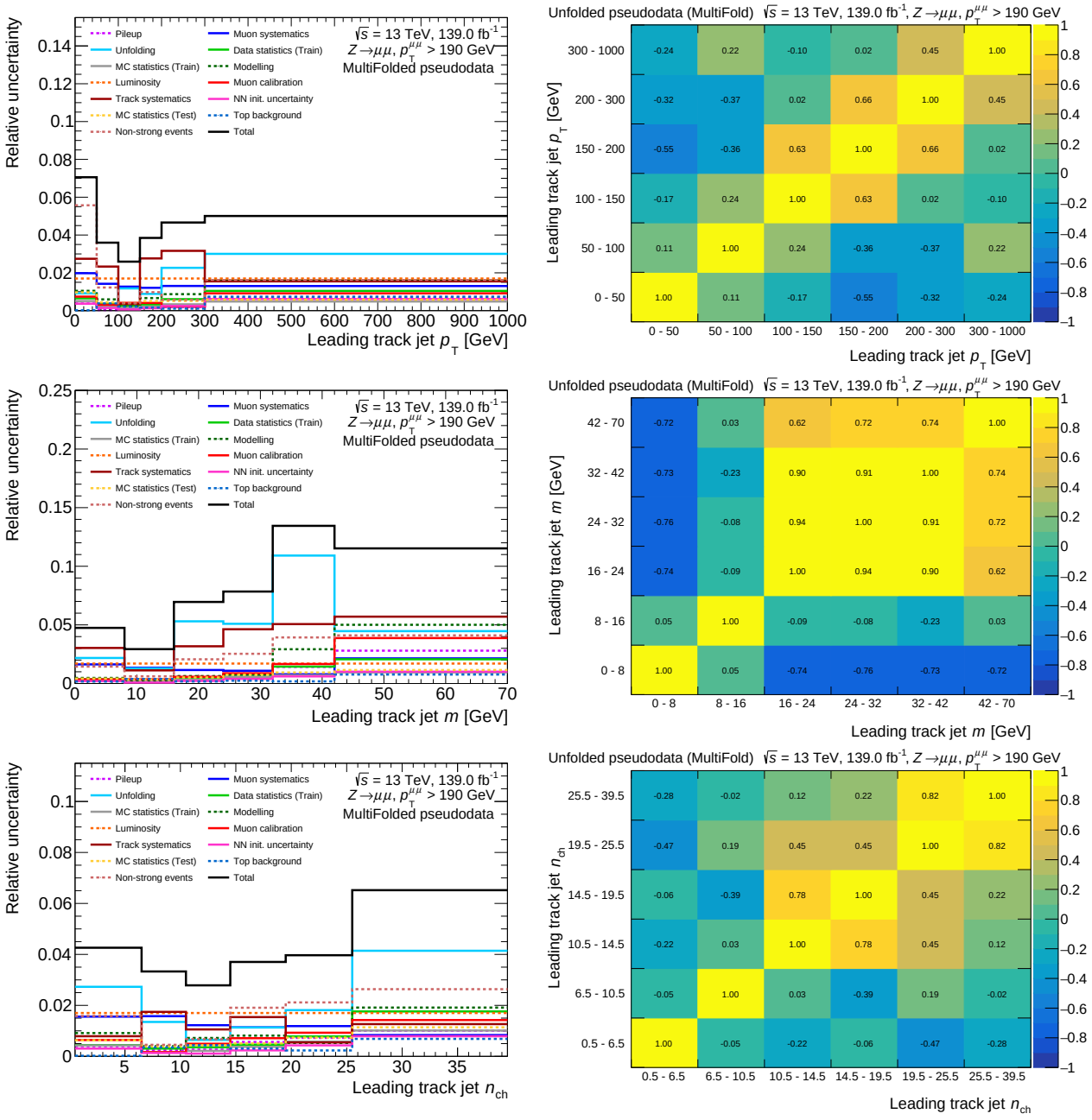


Figure 8.10: The plots on the left show the relative uncertainty on the MultiFolded result for the leading track jet p_{T}, m and number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

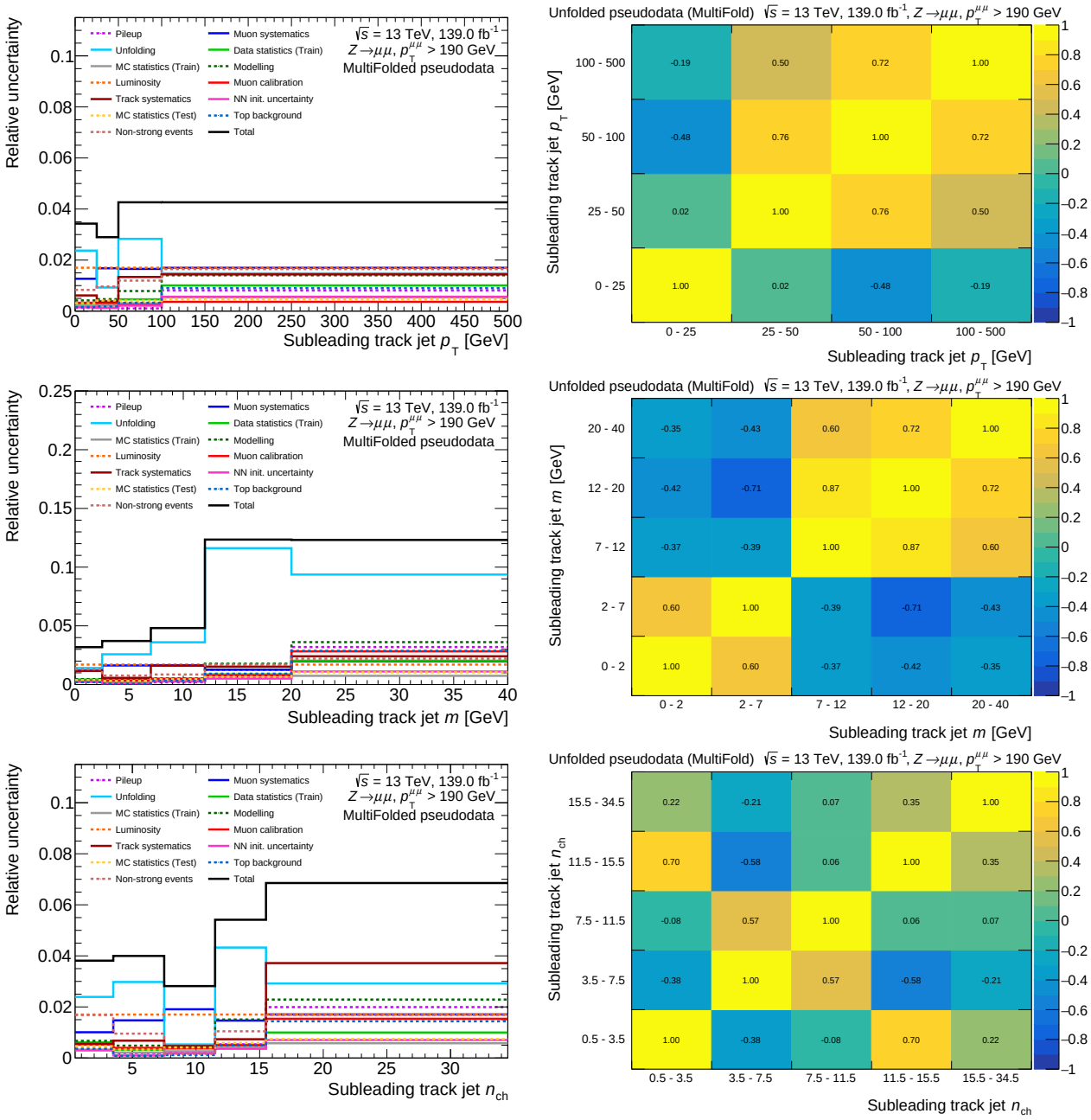


Figure 8.11: The plots on the left show the relative uncertainty on the MultiFolded result for the subleading track jet p_{T}, m and number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

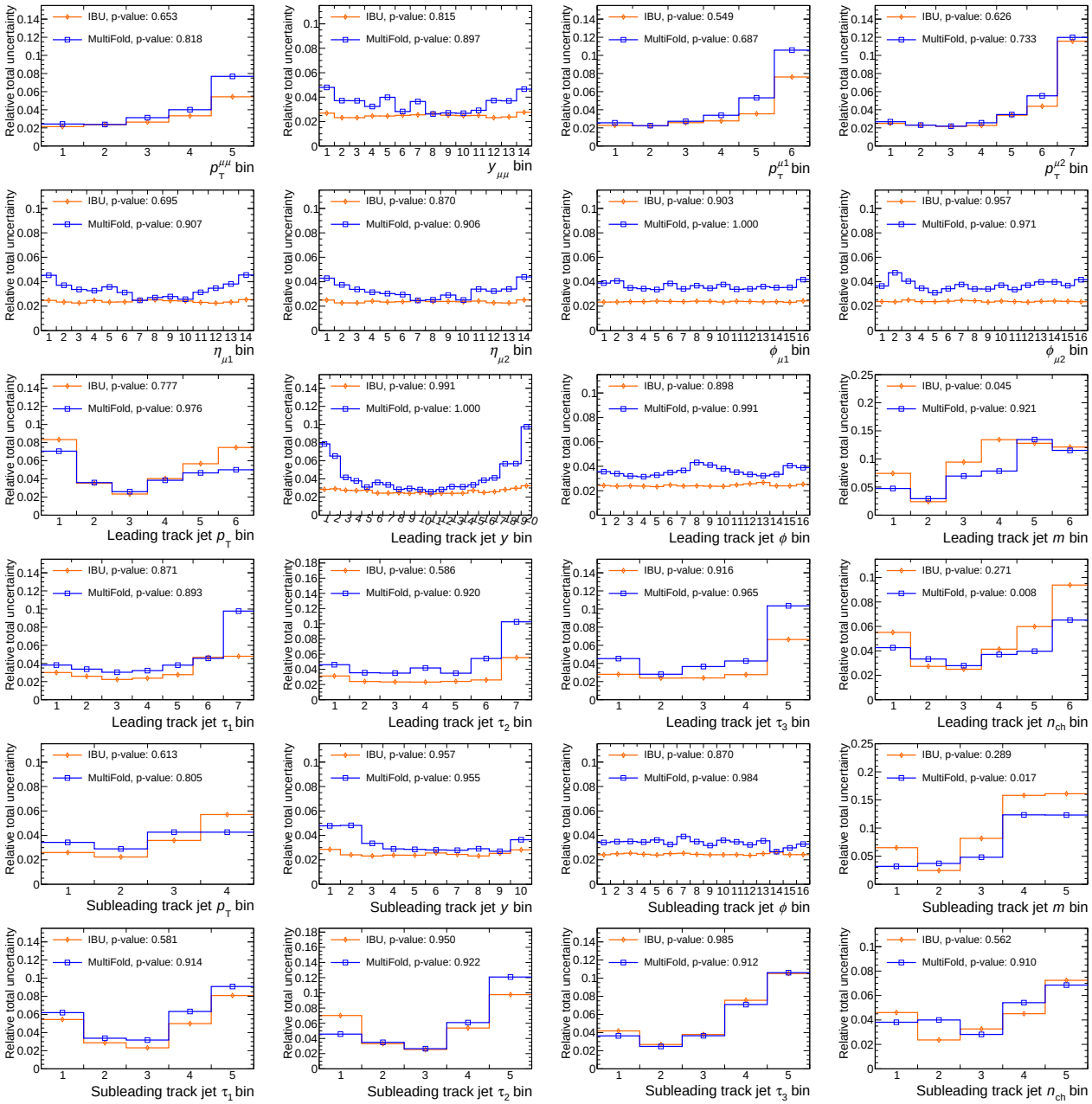


Figure 8.12: The relative total uncertainty in each bin (corresponding to the bins in Table 7.7) for the unfolded results from IBU (orange) and MultiFold (blue). For most bins, MultiFold is observed to have a slightly higher uncertainty, as discussed in the text.

8.1.2 Tests of the MultiFold result

Having shown that good agreement is attained between the MultiFold result and the target truth pseudodata for the 24 observables used as input, some additional results can be explored that demonstrate some of the features of the MultiFold method. Firstly, some observables will be derived from a subset of the 24 using the Multi-Folded result. Secondly, the differential cross section results for the 24 observables will be reexamined in a series of kinematic subregions. Lastly, some two dimensional distributions will be constructed. All of these results will once again be compared to the target distribution and the agreement quantified by the p-value.

8.1.2.1 Derived observables

One of the notable benefits of the MultiFold algorithm is that observables can be calculated after the unfolding has been performed, provided the necessary quantities to calculate the new observable are included in the inputs. This increases the utility of the results by eliminating the need for the analysis to be redone if a new observable is desired. As such, this section explores some of the observables that may be derived from the 24 inputs. They are:

- τ_{21}^{j1}: the ratio of the leading track jet τ_2 and τ_1, i.e. $\tau_{21}^{j1} \equiv \tau_2^{j1}/\tau_1^{j1}$
- $\Delta R_{\mu\mu,j1}$: the angular distance between the Z boson and the leading track jet. This involves performing four vector addition for the leading and subleading muons, and calculating ΔR between the dimuon system and the leading track jet. It is thus a function of 8 out of the 24 input observables.
- Track m_{jj}: the invariant mass of the leading and subleading track jet. This involves the four vector addition of the two track jets, and is thus a function of

Observable	MultiFold p-value	IBU p-value
Leading track jet τ_{21}	0.913	0.763
$\Delta R_{\mu\mu,j1}$	0.953	0.0723
Track m_{jj}	0.451	0.560

Table 8.2: The p-values that result from performing a χ^2 test to test the agreement between the unfolded pseudodata and the target. The observables listed are unfolded directly using IBU, but are calculated after unfolding in the MultiFold case.

8 input observables.

These three observables are also unfolded directly using IBU to serve as a comparison. Figure 8.13 shows the differential cross section results obtained by calculating these observables from the MultiFolded results compared with the IBU result and the target. The resulting p-values for both methods are displayed alongside the results and also in Table 8.2. All three observables exhibit good agreement with the target. The relative uncertainties and the corresponding uncertainty correlation matrices can be found in Appendix C.

8.1.2.2 Results in kinematic subregions

In order to further validate the MultiFold results presented in this section, the differential cross section distributions for the 24 observables are examined in a series of kinematic subregions and their agreement with the target is evaluated once more. The kinematic subregions presented in this section can be broken down into three main categories: a higher p_{T}^{Z} region, a region with a higher EW Z+jets contribution, and a region with an enhanced diboson contribution.

Higher p_{T}^{Zjj} region

The first criterion applied in this case is that the subleading track jet p_{T} be larger than 50 GeV, which enforces a topology where both jets have high p_{T}. The binning is kept consistent where possible with Table 7.7, however the following modifications

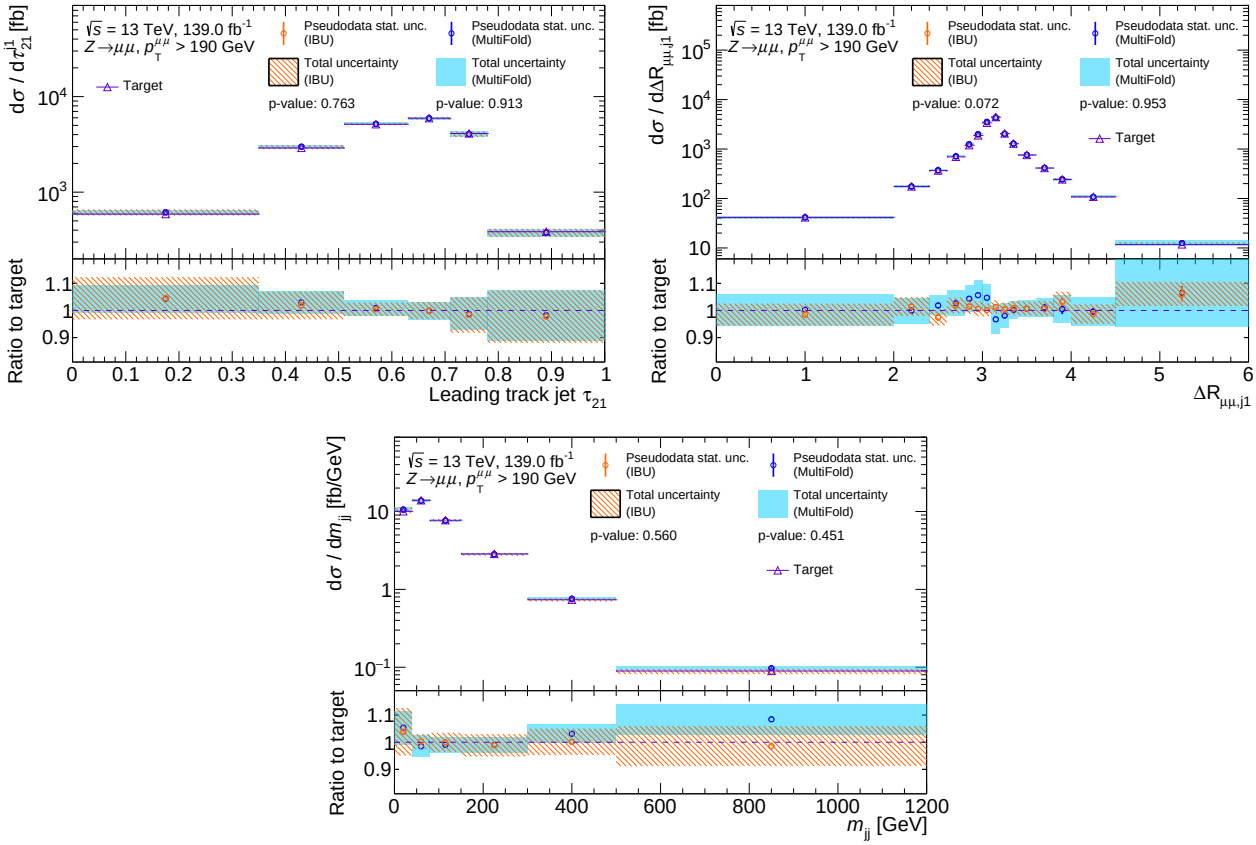


Figure 8.13: The differential cross section as a function of the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and the track dijet mass. The pseudodata unfolded with IBU is shown in orange with the total uncertainty shown in the hatched band. The MultiFolded pseudodata for these three observables is calculated after the unfolding is performed, and is shown similarly in blue. The target truth pseudodata is shown in purple, and the p-value quantifying the agreement between the unfolded result and the target is displayed on the figure for each method.

are made in the case of empty or low statistics bins:

- Subleading track jet p_{T}: (0, 25, 50, 100, 500) to (50, 75, 100, 150, 500);
- Leading track jet p_{T}: (0, 50, 100, 150, 200, 300, 1000) to (50, 100, 150, 200, 300, 1000).

The results for each of the 24 observables are shown in Figure 8.14 and the corresponding p-values are summarized along with the other results in this section in Table 8.3.

The second criterion applied is simply that the dimuon p_{T} be larger than 350 GeV. The binning modifications in this case are:

- $p_{\mathrm{T}}^{\mu\mu}$: (190, 230, 300, 450, 600, 1000) to (350, 400, 450, 600, 1000);
- $p_{\mathrm{T}}^{\mu 1}$: (25, 125, 200, 300, 400, 600, 800) to (125, 200, 300, 400, 600, 800).

Results are shown in Figure 8.15 and p-values are summarized in Table 8.3.

For both of these kinematic subregions, agreement with the target is seen for all obervables with the exception of the subleading track jet mass. If the unfolding uncertainty is considered decorrelated, then the p-value for the subleading track jet mass in the $p_{\mathrm{T}}^{j2} > 50$ GeV region improves slightly (from 0.0212 to 0.0216) but still exhibits tension on the order of 2–3σ with the target. Conversely, in the $p_{\mathrm{T}}^{\mu\mu} > 350$ GeV region, the p-value becomes 0.289, bringing the subleading track jet mass into agreement with the target.

Electroweak enhanced region

To examine a region of phase space with an enhanced electroweak Z+jets contribution, two separate criteria are placed on derived observables to target the so-called VBF topology. The first criterion is that the track m_{jj} be larger than 200 GeV. The

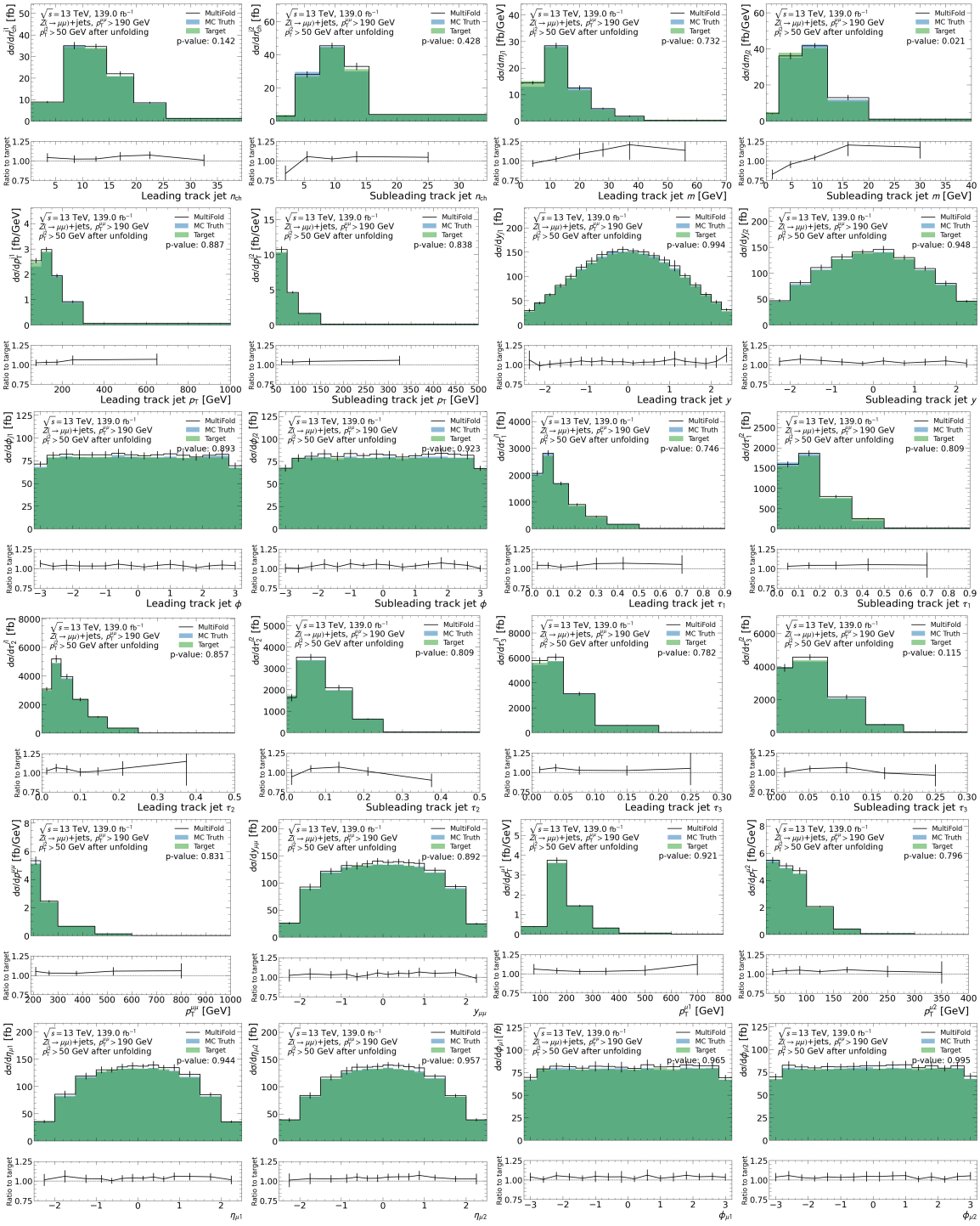


Figure 8.14: The differential cross section measurement for each of the 24 observables with a selection criterion of $p_{\mathrm{T}}^{j2} > 50\,\mathrm{GeV}$ applied after the MultiFold result has been obtained. The MultiFold result is in black, the original MC truth (the starting place for the MultiFold algorithm) is in blue, and the target truth pseudodata is in green.

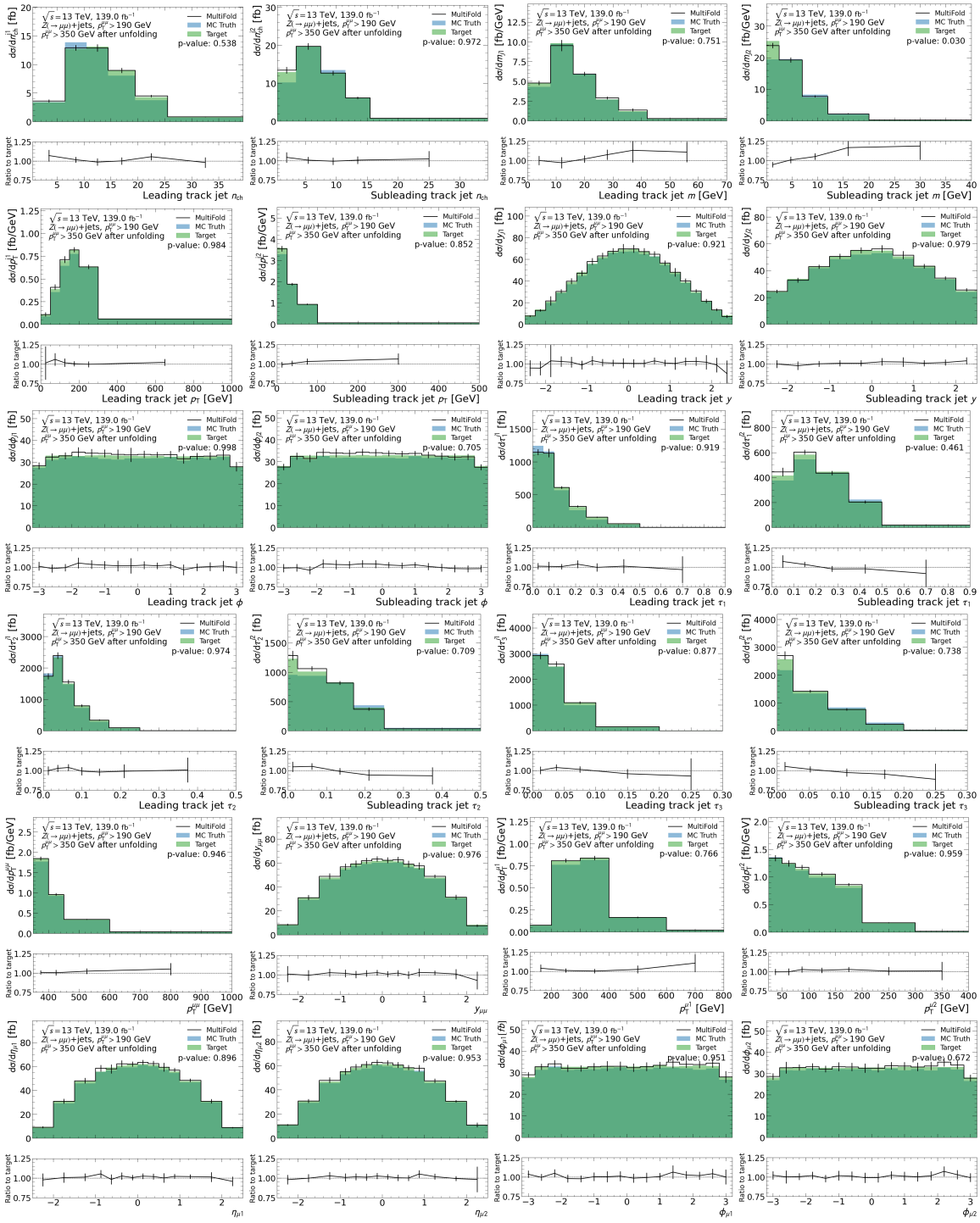


Figure 8.15: The differential cross section measurement for each of the 24 observables with a selection criterion of $p_{\mathrm{T}}^{\mu\mu} > 350\,\mathrm{GeV}$ applied after the MultiFold result has been obtained. The MultiFold result is in black, the original MC truth (the starting place for the MultiFold algorithm) is in blue, and the target truth pseudodata is in green.

differential cross section results for the 24 observables and their agreement with the target is shown in Figure 8.16.

The second derived observable considered is the track dijet rapidity separation: $|\Delta y_{jj}| = |y_{j1} - y_{j2}|$. The criterion in this case is that $|\Delta y_{jj}| > 2$, and the results of its application are shown for each of the 24 observables in Figure 8.17.

As before, the p-values for each of the 24 observables after these selection criteria have been applied can be found in Table 8.3. For the track $m_{jj} > 200\,\text{GeV}$ criterion agreement is observed for all 24 observables, and for the $|\Delta y_{jj}| > 2$ criterion the only observable to exhibit poor agreement is the subleading track jet mass (if the unfolding uncertainties are decorrelated, then agreement is achieved with a p-value of 0.0720).

Diboson enhanced region

In order to examine how the MultiFolded results perform in a region of phase space where the diboson contribution is enhanced, a selection is applied such that the leading track jet mass must be larger than 32 GeV. The changes to the binning are:

- Leading track jet m: (0, 8, 16, 24, 32, 42, 70) to (32, 36, 42, 70);
- Leading track jet p_{T}: (0, 50, 100, 150, 200, 300, 1000) to (100, 150, 200, 300, 1000);
- Leading track jet n_{ch}: (0.5, 6.5, 10.5, 14.5, 19.5, 25.5, 39.5) to (0.5, 10.5, 14.5, 19.5, 25.5, 39.5).

Note that this criterion reduces the dataset size by approximately 98% and so the statistical uncertainties are substantial. The differential cross section results for each of the 24 observables after the application of this selection criterion is shown in Figure 8.18. The p-values are also once again tabulated alongside the other results from this section in Table 8.3. Agreement is seen for all observables in this case.

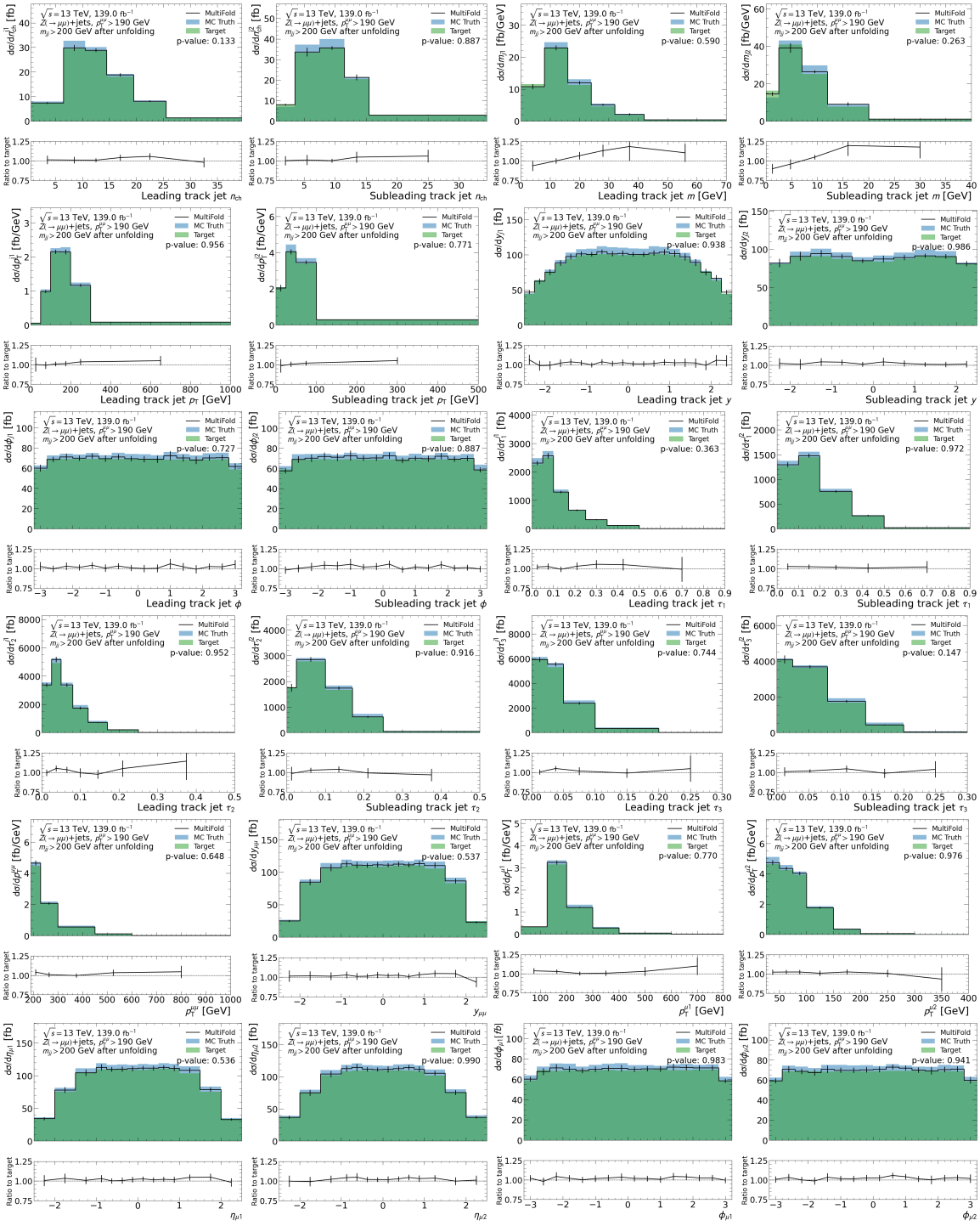


Figure 8.16: The differential cross section measurement for each of the 24 observables with a selection criterion of track $m_{jj} > 200\,\text{GeV}$ applied after the MultiFold result has been obtained. The MultiFold result is in black, the original MC truth (the starting place for the MultiFold algorithm) is in blue, and the target truth pseudodata is in green.

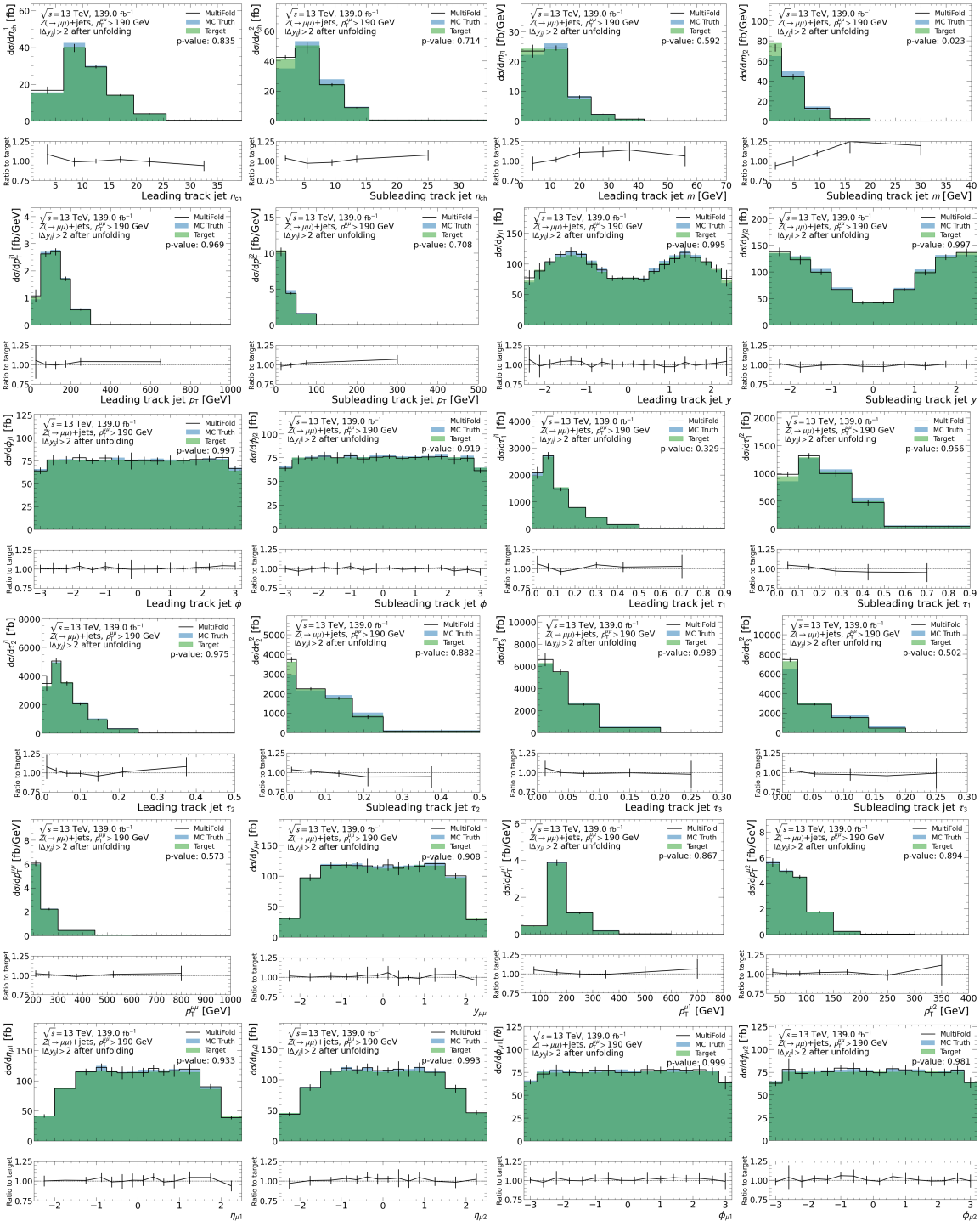


Figure 8.17: The differential cross section measurement for each of the 24 observables with a selection criterion of track $|\Delta y_{jj}| > 2$ applied after the MultiFold result has been obtained. The MultiFold result is in black, the original MC truth (the starting place for the MultiFold algorithm) is in blue, and the target truth pseudodata is in green.

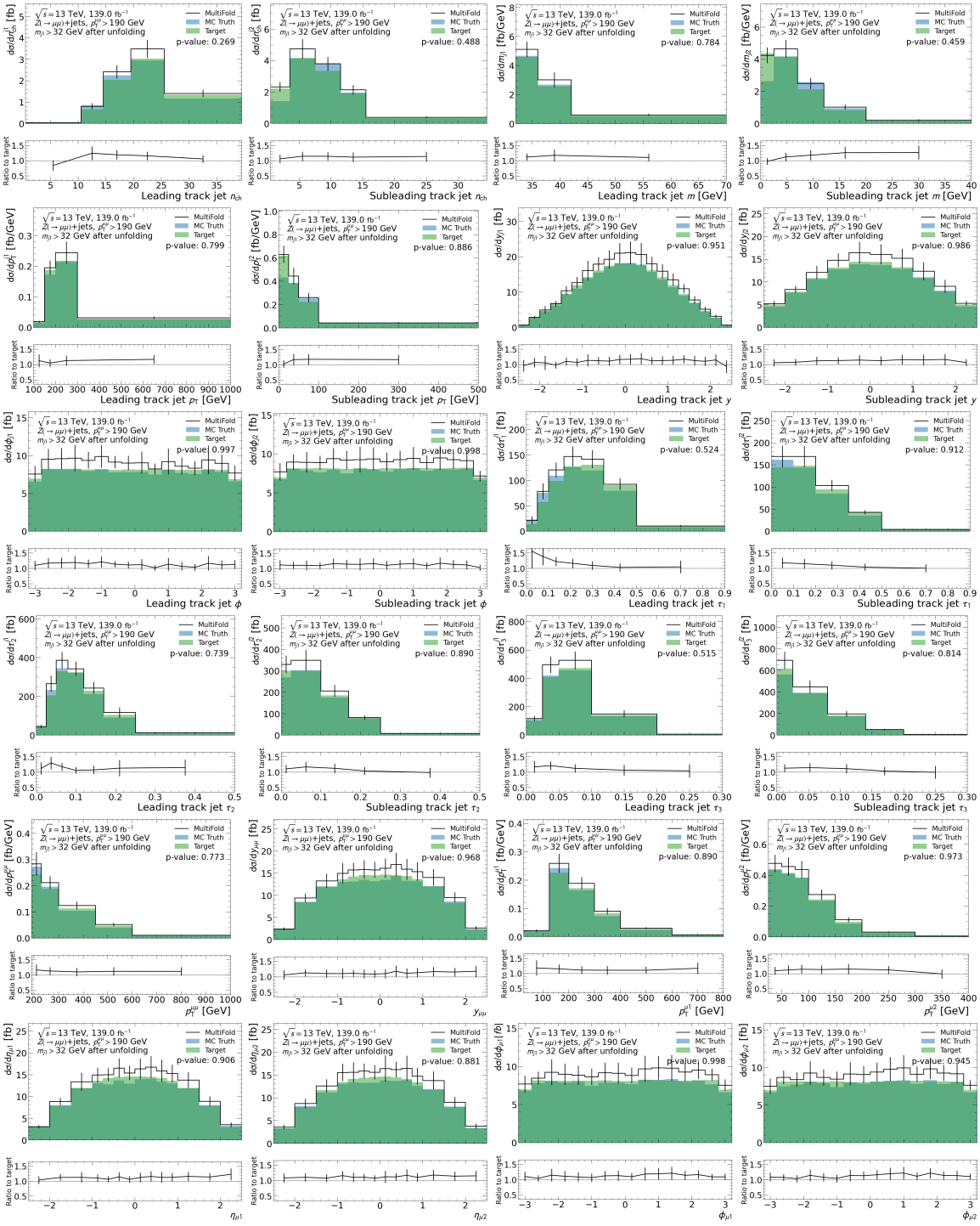


Figure 8.18: The differential cross section measurement for each of the 24 observables with a selection criterion of $m_{j1} > 32\,\text{GeV}$ applied after the MultiFold result has been obtained. The MultiFold result is in black, the original MC truth (the starting place for the MultiFold algorithm) is in blue, and the target truth pseudodata is in green.

Observable	***p*-values**				
	$p_{\mathrm{T}}^{j2} > 50\,\mathrm{GeV}$	$p_{\mathrm{T}}^{\mu\mu} > 350\,\mathrm{GeV}$	$m_{jj} > 200\,\mathrm{GeV}$	$\lvert\Delta y_{jj}\rvert > 2$	$m_{j1} > 32\,\mathrm{GeV}$
$p_{\mathrm{T}}^{\mu\mu}$	0.831	0.946	0.648	0.573	0.772
$y_{\mu\mu}$	0.892	0.976	0.537	0.908	0.968
$p_{\mathrm{T}}^{\mu 1}$	0.921	0.766	0.770	0.867	0.890
$p_{\mathrm{T}}^{\mu 2}$	0.796	0.959	0.976	0.894	0.973
$\eta_{\mu 1}$	0.944	0.896	0.536	0.933	0.901
$\eta_{\mu 2}$	0.957	0.953	0.990	0.993	0.881
$\phi_{\mu 1}$	0.965	0.951	0.983	0.999	0.998
$\phi_{\mu 2}$	0.995	0.672	0.941	0.981	0.945
p_{T}^{j1}	0.887	0.984	0.956	0.969	0.799
p_{T}^{j2}	0.838	0.852	0.771	0.708	0.886
y_{j1}	0.994	0.921	0.938	0.995	0.951
y_{j2}	0.948	0.979	0.986	0.997	0.986
ϕ_{j1}	0.893	0.998	0.727	0.997	0.997
ϕ_{j2}	0.923	0.705	0.887	0.919	0.998
m_{j1}	0.732	0.751	0.590	0.592	0.784
m_{j2}	0.0212	0.0295	0.263	0.0228	0.459
n_{ch}^{j1}	0.142	0.538	0.133	0.835	0.269
n_{ch}^{j2}	0.428	0.972	0.887	0.714	0.488
τ_1^{j1}	0.746	0.919	0.363	0.329	0.524
τ_1^{j2}	0.809	0.461	0.972	0.956	0.912
τ_2^{j1}	0.857	0.974	0.952	0.975	0.739
τ_2^{j2}	0.809	0.709	0.916	0.882	0.890
τ_3^{j1}	0.782	0.877	0.744	0.989	0.515
τ_3^{j2}	0.115	0.738	0.147	0.502	0.814

Table 8.3: The p-values quantifying the agreement between the MultiFolded pseudodata result and the target truth pseudodata where a series of selection criteria have been applied after the unfolding has been performed.

Obs. 1	Obs. 1 bin edges	Obs. 2	Obs. 2 bin edges	p-value
$p_{\mathrm{T}}^{\mu\mu}$ [GeV]	190, 230, 300, 450, 1000	p_{T}^{j1} [GeV]	0, 50, 100, 200, 500, 1000	0.959
p_{T}^{j1} [GeV]	0, 50, 100, 200, 500, 1000	p_{T}^{j2} [GeV]	0, 25, 75, 200, 500	0.486
m_{j1} [GeV]	0, 8, 16, 24, 32, 70	τ_1^{j1}	0, 0.05, 0.1, 0.17, 0.25, 0.35, 0.5, 0.9	0.761
m_{j2} [GeV]	0, 2.5, 7, 12, 20, 40	τ_2^{j2}	0, 0.1, 0.2, 0.35, 0.5, 0.9	4.53×10^{-6}
$p_{\mathrm{T}}^{\mu\mu}$ [GeV]	190, 230, 300, 450, 1000	$y_{\mu\mu}$	-2.5, -1.5, -1, -0.5, -0.25, 0, 0.25, 0.5, 1, 1.5, 2.5	0.952
$p_{\mathrm{T}}^{\mu\mu}$ [GeV]	190, 230, 300, 450, 1000	m_{jj} [GeV]	0, 40, 80, 150, 300, 500, 1000	0.291

Table 8.4: The chosen observables (obs. 1 and 2) and their corresponding binning for the two dimensional distributions constructed using the MultiFolded pseudodata results. The corresponding p-value is also shown, which quantities the agreement between the result and the target truth pseudodata.

8.1.2.3 Two-dimensional closure tests

The final feature of the MultiFold method that will be explored in this chapter is the ability to make multidimensional distributions after obtaining the unfolded result. Six such distributions will be examined here and their agreement with the target truth pseudodata will once again be checked. The observables and their chosen binning, as well as the resulting p-values are shown in Table 8.4. Meanwhile, Figures 8.19 and 8.20 show the ratio to the target distribution in each bin for each two-dimensional distribution.

Good agreement is attained for all two-dimensional distributions with the exception of the one involving the subleading track jet mass. If the unfolding uncertainty is once again decorrelated, then agreement is observed with a p-value of 0.562.

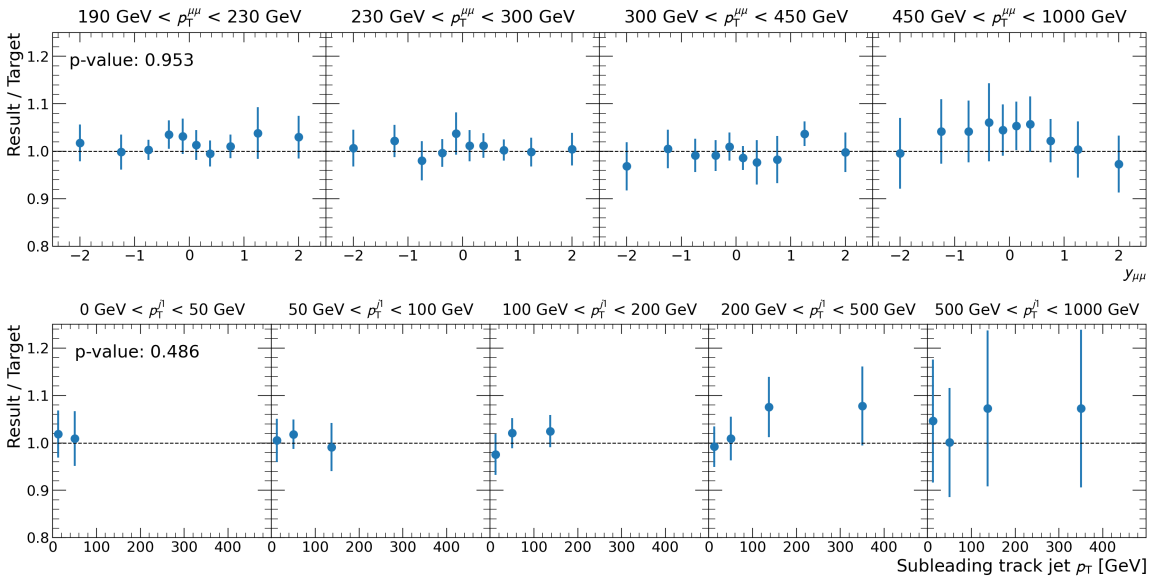


Figure 8.19: Two dimensional distributions for the dimuon p_{T} and rapidity, and the leading and subleading track jet p_{T} created using the MultiFolded pseudodata result. The ratio of the results to the target truth pseudodata are displayed as a function of one observable per bin of the other observable. The corresponding p-values are also shown.

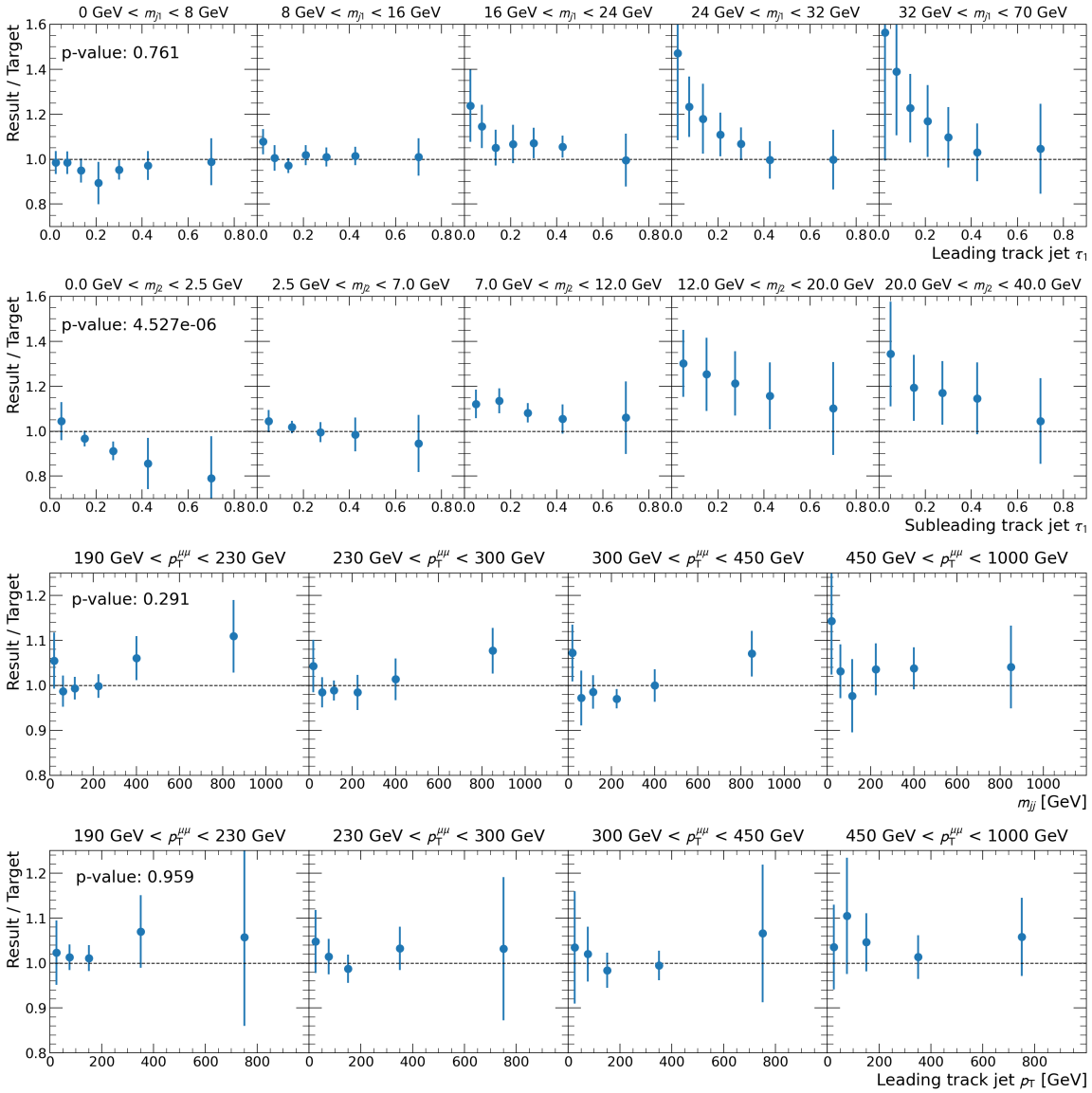


Figure 8.20: Two dimensional distributions for the leading track jet m and τ_1, the subleading track jet m and τ_1, the dimuon p_{T} and the track dijet mass, and the dimuon p_{T} and the leading track jet p_{T} created using the MultiFolded pseudodata result. The ratio of the results to the target truth pseudodata are displayed as a function of one observable per bin of the other observable. The corresponding p-values are also shown.

8.2 Results with data

This section presents the final results of this analysis. The differential cross section distributions for each of the 24 observables using the Run 2 ATLAS data unfolded with the MultiFold method are shown in Figures 8.21 to 8.25. Additionally, the differential cross section distributions are shown for the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and m_{jj} in Figure 8.26. The results obtained using the MultiFold algorithm are compared with the results obtained using IBU, as well as with the theoretical predictions from the MadGraph+Pythia8 FxFx sample and the Sherpa 2.2.11 sample with the additional EW and diboson truth level events included.

The uncertainties on the results for both IBU and MultiFold are as described in Sections 7.8.1.2 and 7.8.2.2, and the relative uncertainties and corresponding correlation matrices can be found in Figures 8.27 to 8.29 for IBU and in Figures 8.30 to 8.32 for MultiFold for a selection of the 24 observables. The remainder are shown in Appendix C. As was the case with the pseudodata, the total uncertainties on the MultiFold result tend to be larger than those of IBU for the same reasons discussed in the previous section. The uncertainty on the theoretical predictions include the PDF and α_s uncertainty, and the QCD perturbative uncertainties. The increased error in the Sherpa 2.2.11 sample compared to the MadGraph+Pythia8 FxFx sample is due to the QCD perturbative uncertainty, which is larger in the Sherpa 2.2.11 sample. This has also been observed in previous analyses utilizing these generators, for example in Ref. [40].

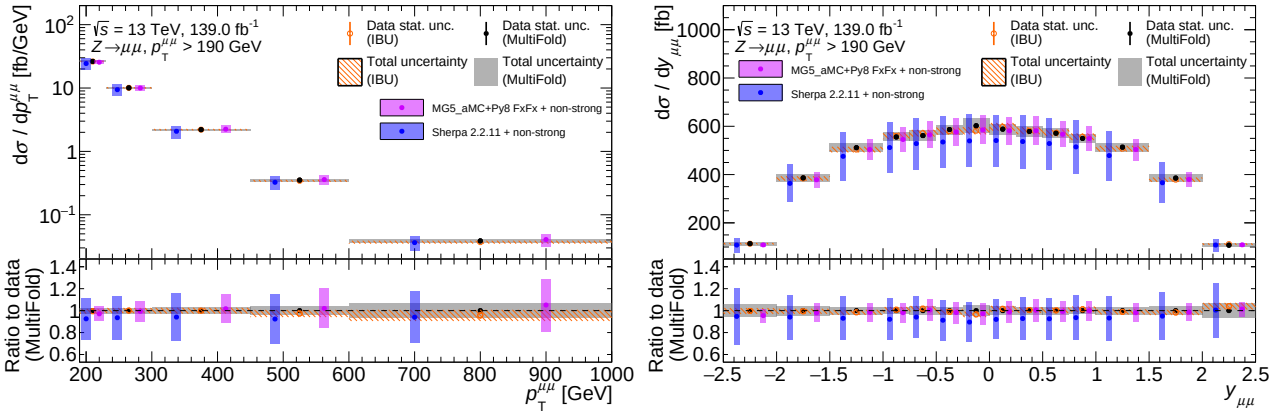


Figure 8.21: The differential cross section as a function of the dimuon p_{T} (left) and y (right). The MultiFold measurement is shown in black with the grey error bands, and is used as the reference distribution for the ratio plot. A separate measurement based on IBU is shown in orange with the total uncertainty shown in the hatched band, and the theoretical predictions from MadGraph+Pythia8 FxFx and Sherpa 2.2.11 are shown in pink and blue, respectively.

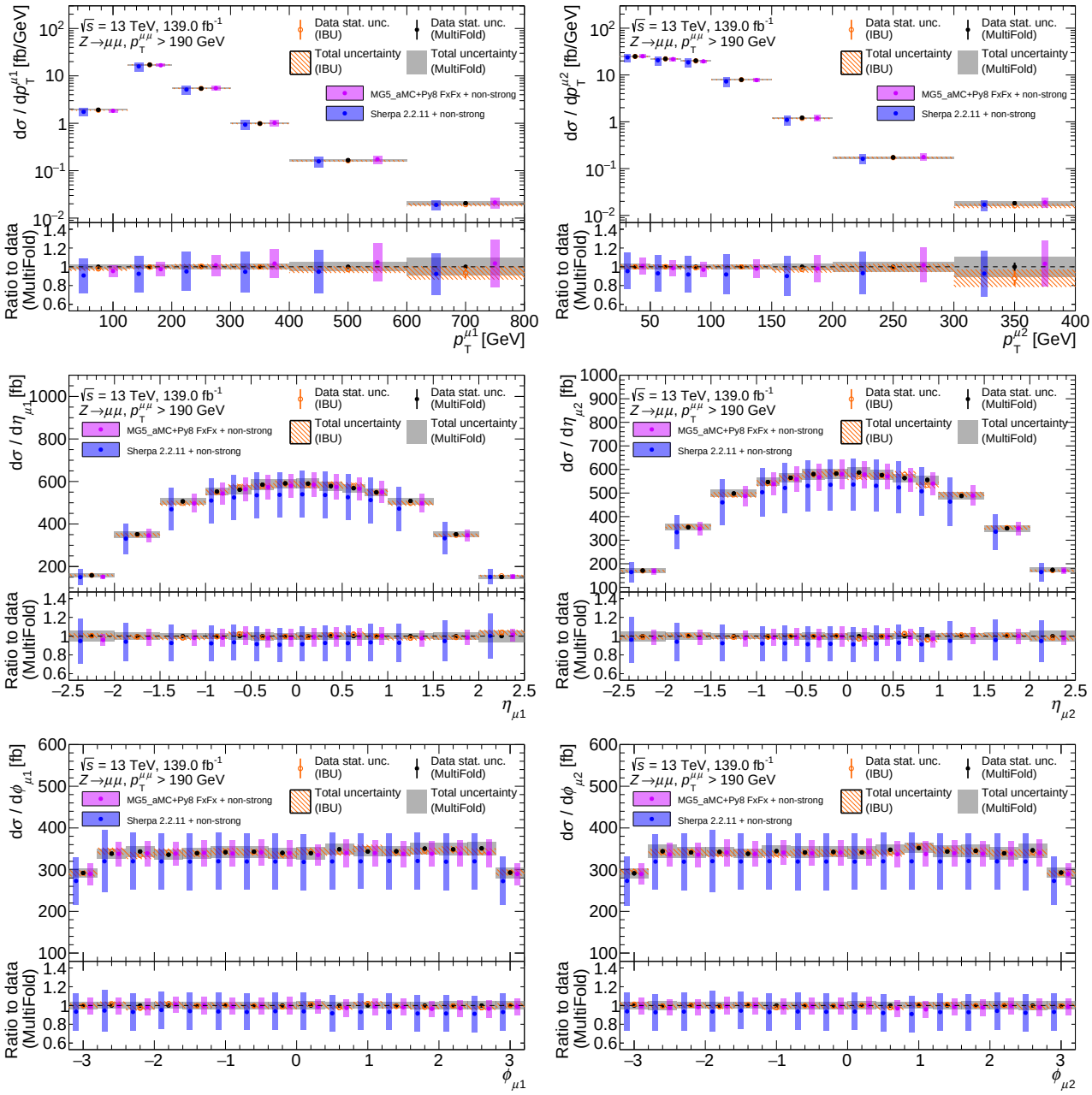


Figure 8.22: The differential cross section as a function of the leading and subleading muon p_{T}, η and ϕ. The MultiFold measurement is shown in black with the grey error bands, and is used as the reference distribution for the ratio plot. A separate measurement based on IBU is shown in orange with the total uncertainty shown in the hatched band, and the theoretical predictions from MadGraph+Pythia8 FxFx and Sherpa 2.2.11 are shown in pink and blue, respectively.

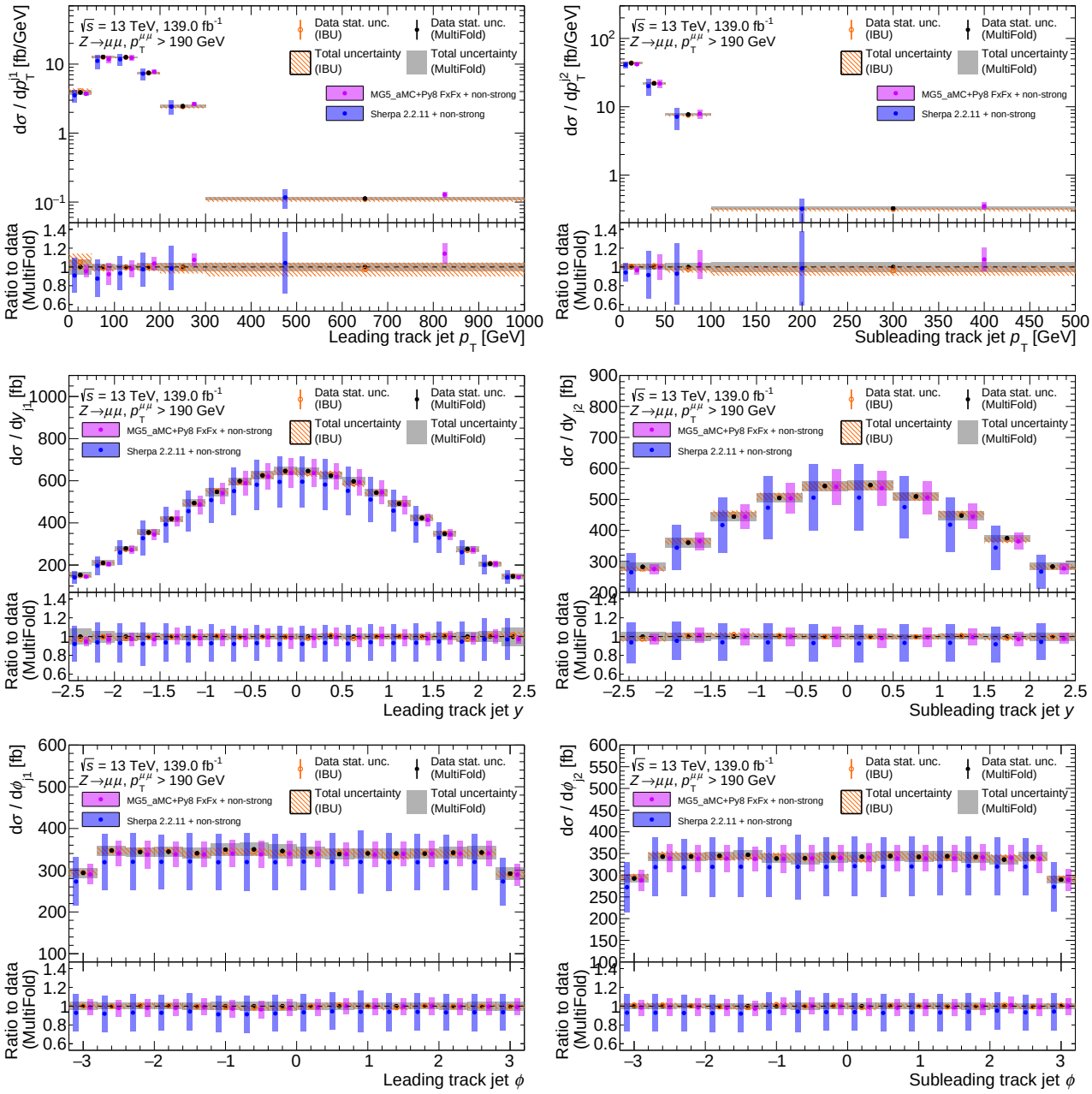


Figure 8.23: The differential cross section as a function of the leading and subleading track jet p_{T}, η and ϕ. The MultiFold measurement is shown in black with the grey error bands, and is used as the reference distribution for the ratio plot. A separate measurement based on IBU is shown in orange with the total uncertainty shown in the hatched band, and the theoretical predictions from MadGraph+Pythia8 FxFx and Sherpa 2.2.11 are shown in pink and blue, respectively.

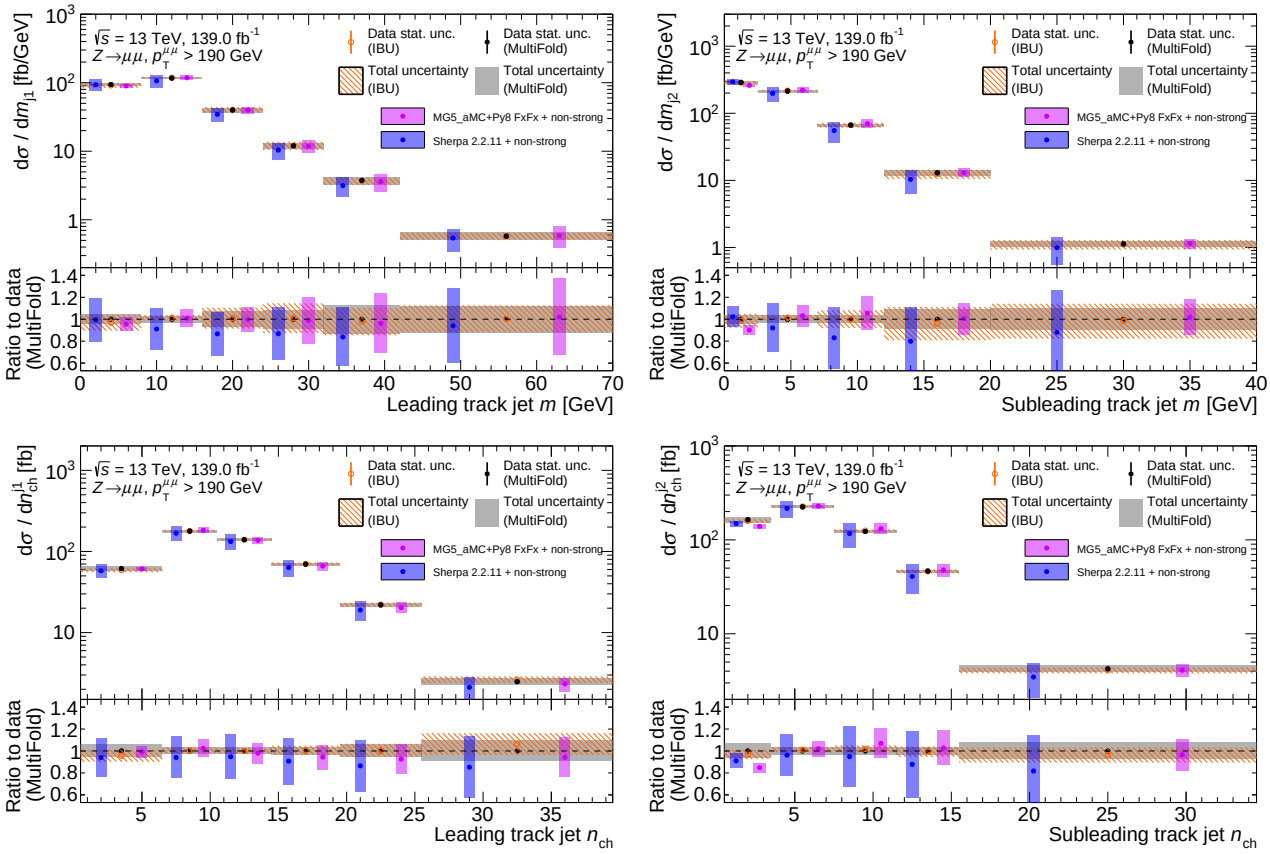


Figure 8.24: The differential cross section as a function of the leading and subleading track jet mass and number of constituents. The MultiFold measurement is shown in black with the grey error bands, and is used as the reference distribution for the ratio plot. A separate measurement based on IBU is shown in orange with the total uncertainty shown in the hatched band, and the theoretical predictions from MadGraph+Pythia8 FxFx and Sherpa 2.2.11 are shown in pink and blue, respectively.

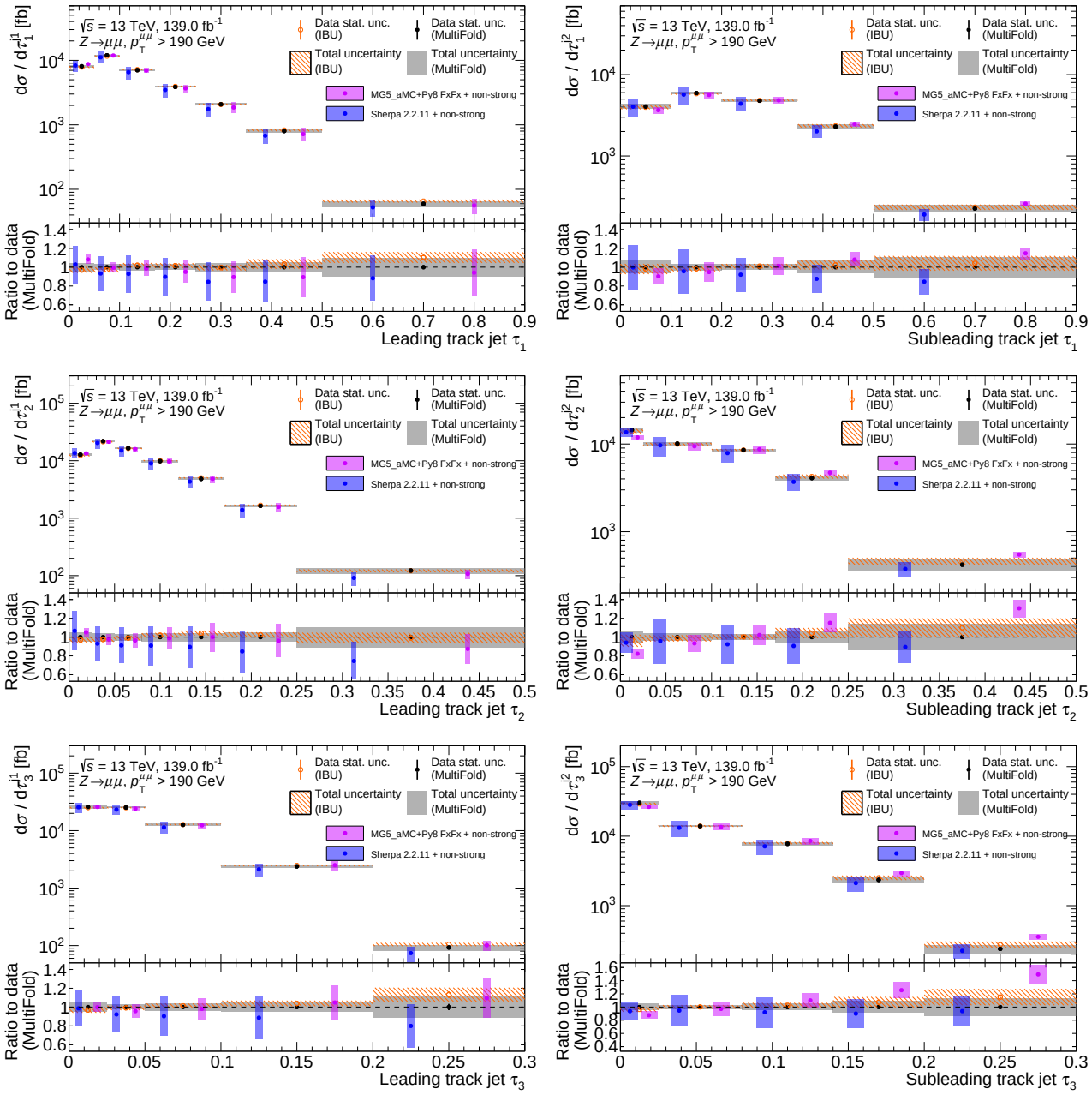


Figure 8.25: The differential cross section as a function of the leading and subleading track jet τ_1, τ_2 and τ_3. The MultiFold measurement is shown in black with the grey error bands, and is used as the reference distribution for the ratio plot. A separate measurement based on IBU is shown in orange with the total uncertainty shown in the hatched band, and the theoretical predictions from MadGraph+Pythia8 FxFx and Sherpa 2.2.11 are shown in pink and blue, respectively.

In general, good agreement within uncertainties is observed between the measurements and the theoretical predictions, the exception being some tension between the measurement and the MadGraph+Pythia8 FxFx sample for the subleading track jet τ_2 and τ_3, in the first bin of the subleading track jet mass and number of constituents, and in the final two bins of the leading track jet τ_{21} distribution. The observed results using the MultiFold method are also found to be in agreement with those obtained using IBU. As was observed in Section 7.5, the central value of the Sherpa 2.2.11 predictions is seen to slightly underestimate the differential cross section in the majority of the distributions. However, as the uncertainties are on the order of 20% in most bins this underestimation is accounted for within the uncertainties.

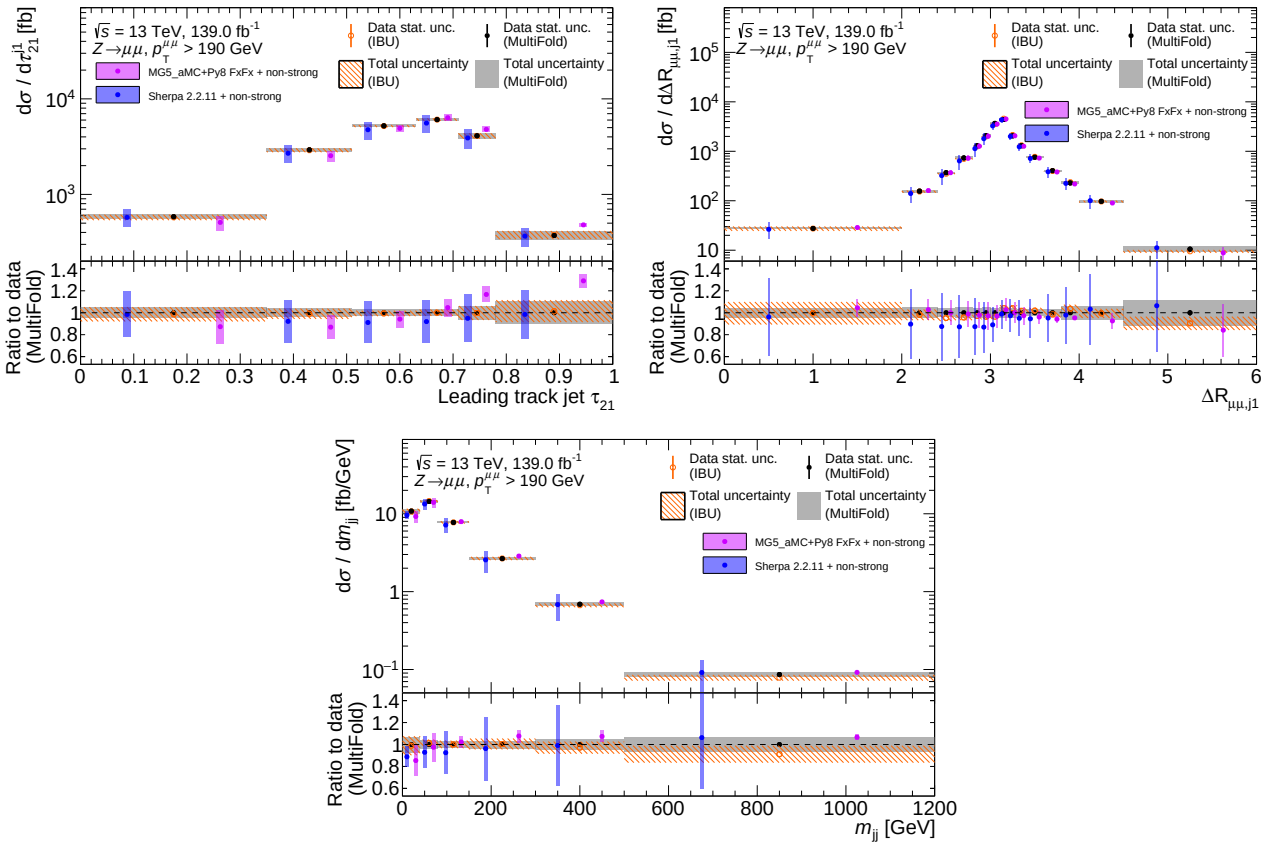


Figure 8.26: The differential cross section as a function of the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and m_{jj}. The MultiFold result is shown in black with the grey error bands, and is used as the reference distribution for the ratio plot. IBU is shown in orange with the total uncertainty shown in the hatched band, and the theoretical predictions from MadGraph+Pythia8 FxFx and Sherpa 2.2.11 are shown in pink and blue, respectively.

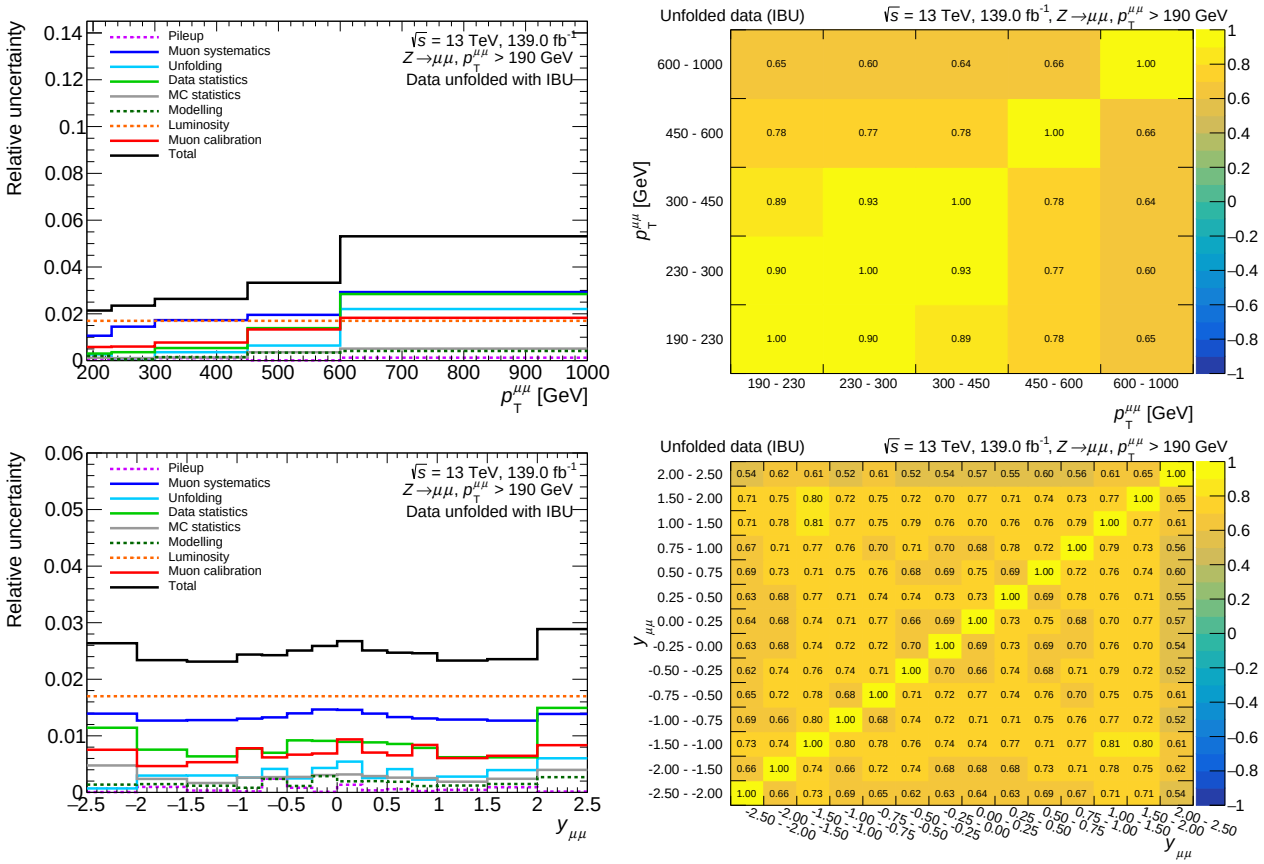


Figure 8.27: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the dimuon p_{T} and y broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

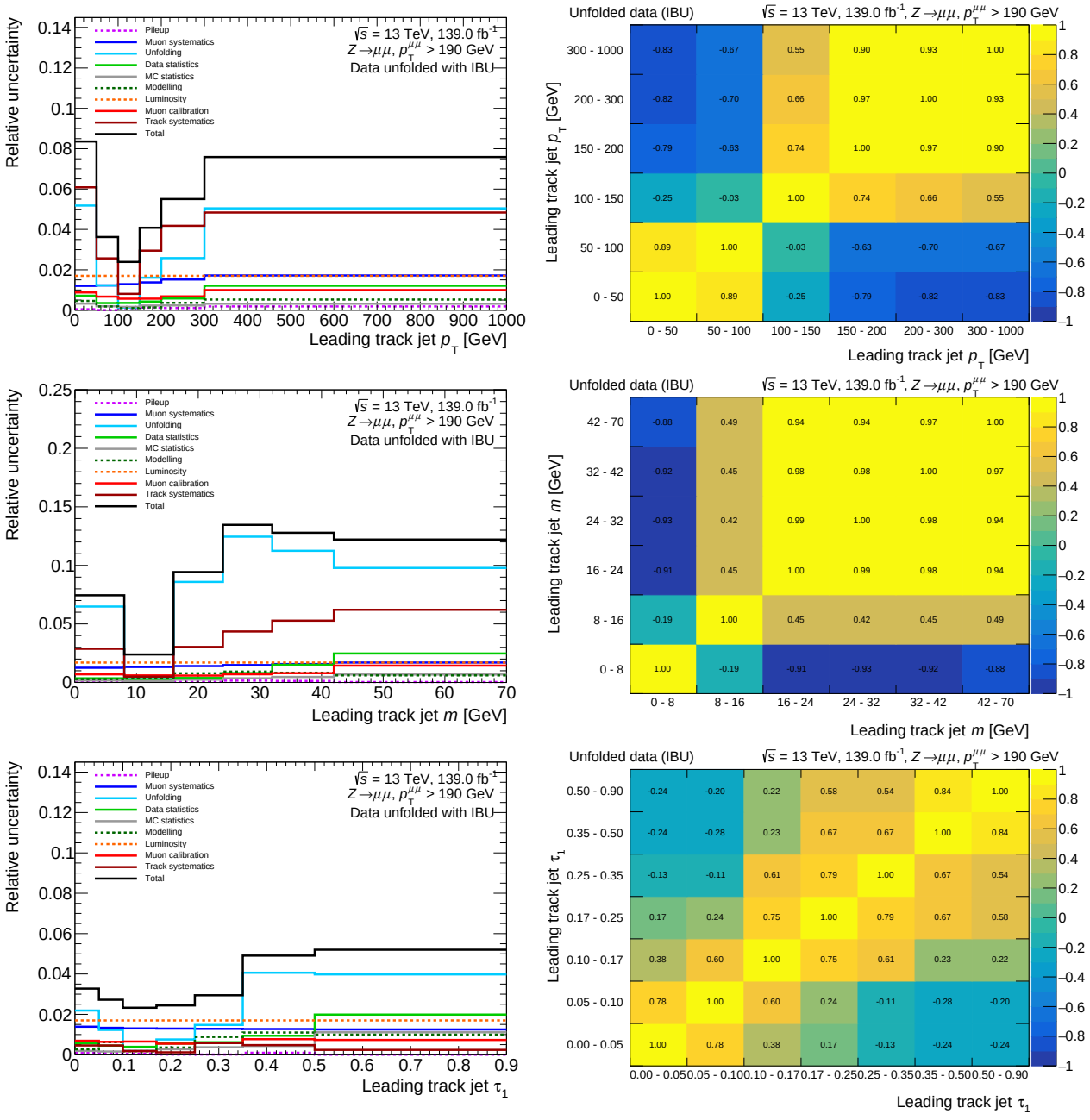


Figure 8.28: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading track jet p_{T}, m and τ_1 broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

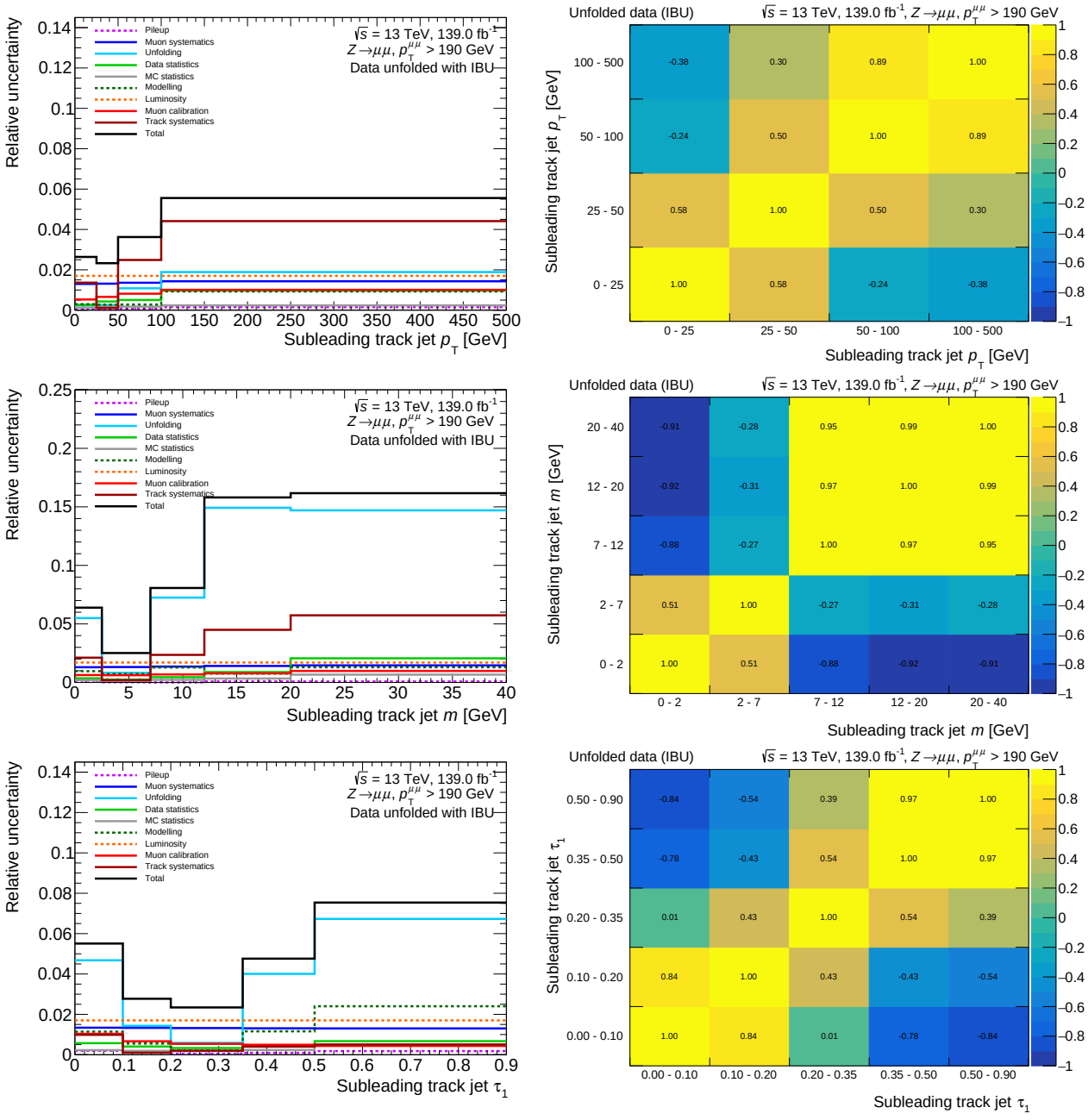


Figure 8.29: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading track jet p_{T}, m and τ_1 broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

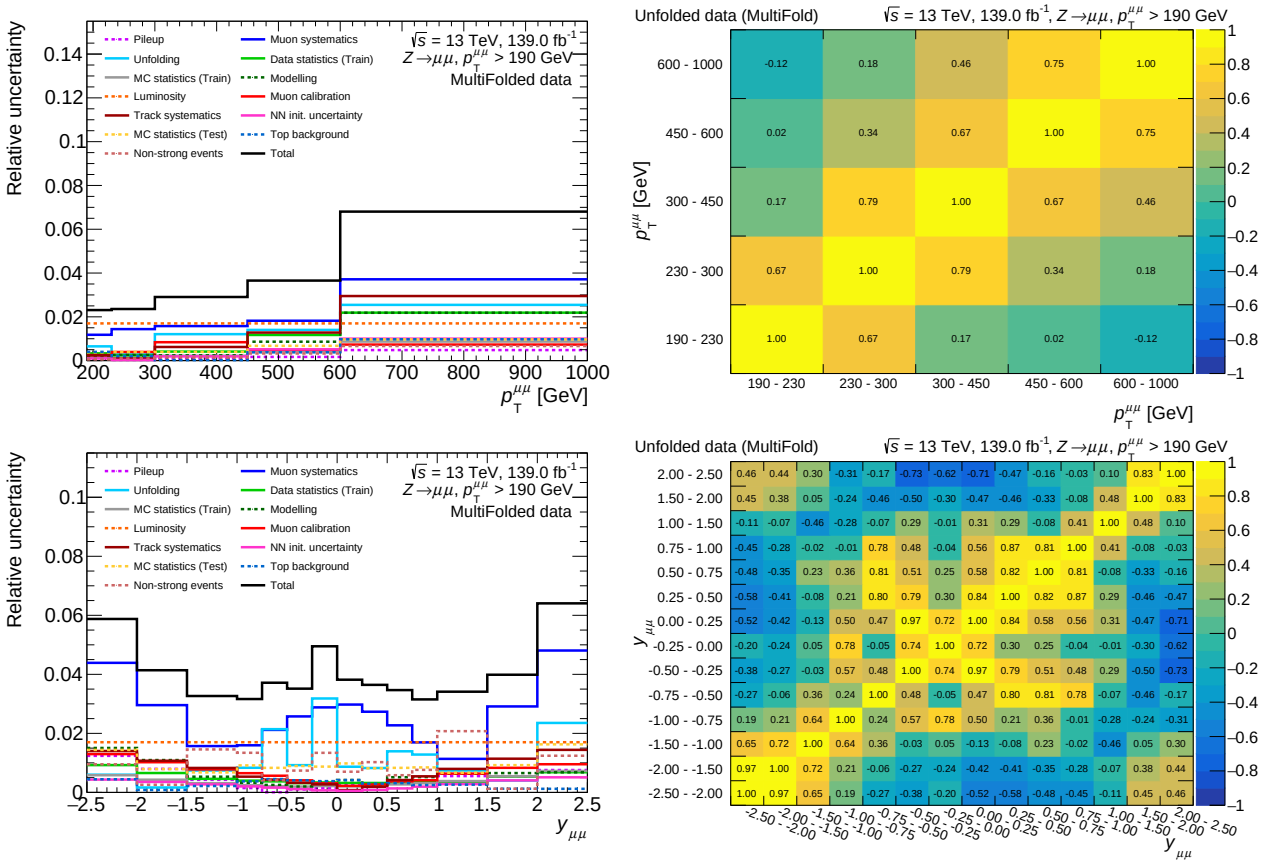


Figure 8.30: The plots on the left show the relative uncertainty on the Multifolded result for the dimuon p_{T} and y broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

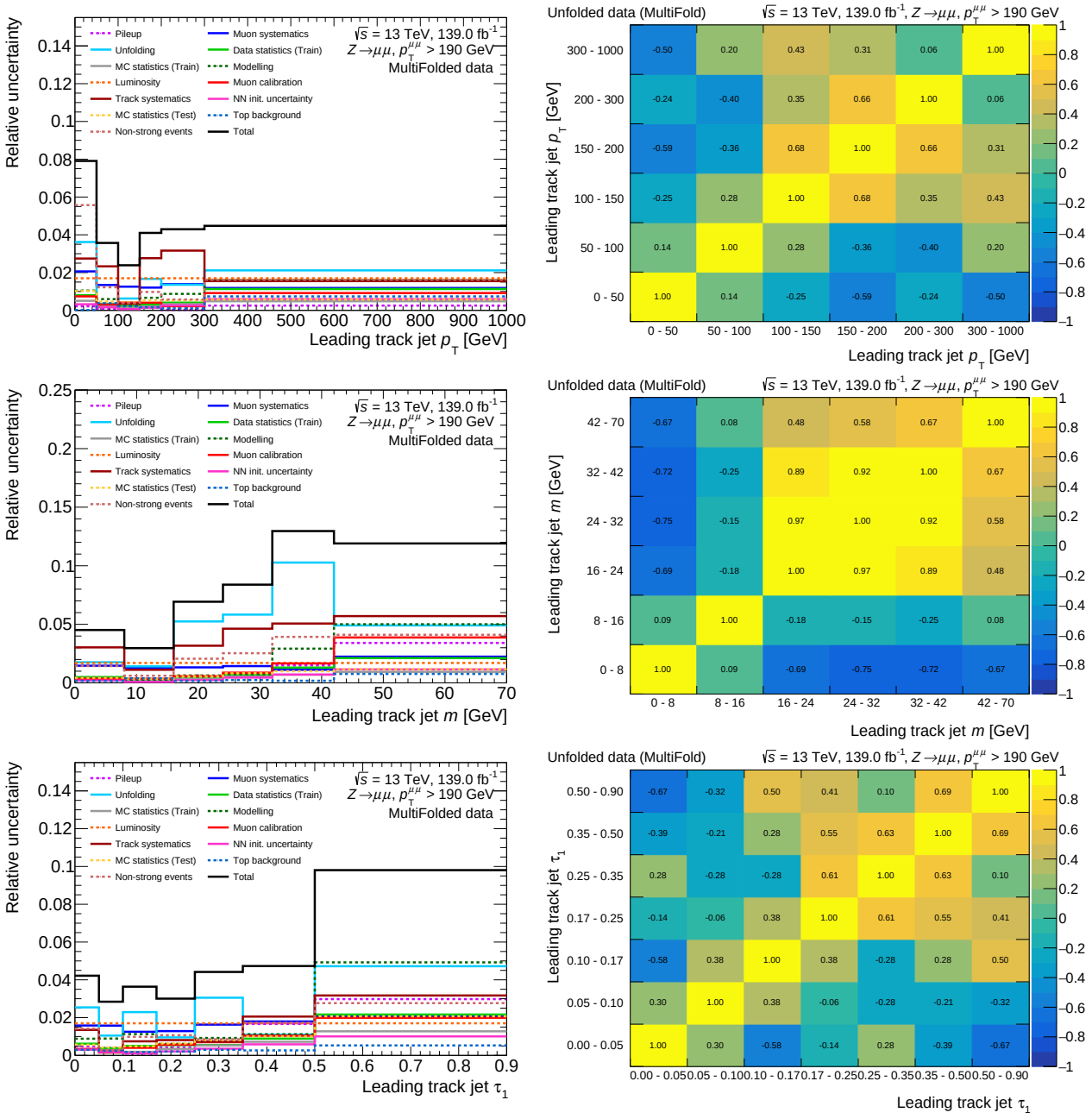


Figure 8.31: The plots on the left show the relative uncertainty on the Multifolded result for the leading track jet p_{T}, m and τ_1 broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

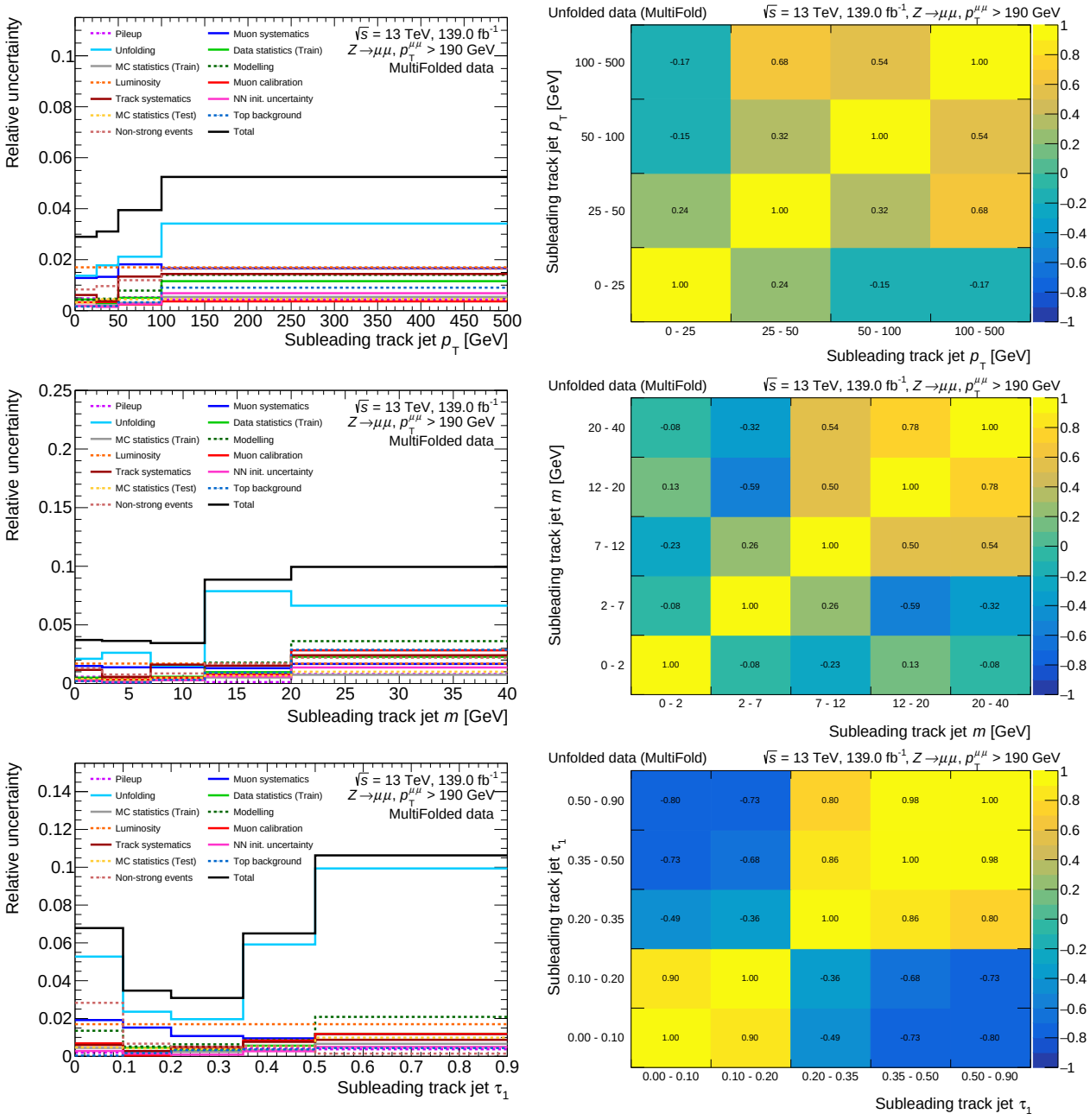


Figure 8.32: The plots on the left show the relative uncertainty on the Multifolded result for the subleading track jet p_{T}, m and τ_1 broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

Chapter 9

Conclusion

This thesis demonstrates the first use of the MultiFold method in an ATLAS analysis. The high-dimensional and unbinned measurement is used to construct differential cross section distributions as a function of 24 observables in Z+jets events in the full ATLAS Run 2 dataset. In order to further demonstrate the benefits of the method, three additional differential cross section distributions are presented as a function of observables calculated from the MultiFold result.

The theoretical predictions provided by the MadGraph+Pythia8 FxFx plus non-strong and Sherpa 2.2.11 plus non-strong generators show generally good agreement with the differential cross section measurements for all 27 observables, with the exception of some tension between the data and the MadGraph+Pythia8 FxFx prediction for the subleading track jet τ_2 and τ_3, and the leading track jet τ_{21}. The first bins of the subleading track jet m and number of constituents also show some small deviations. The Sherpa 2.2.11 sample is found to agree with the data within uncertainties, however the predicted central value underestimates the data by approximately 5–10% in the vast majority of bins.

In comparing the MultiFold results to those obtained using the traditional

method, IBU, they are found to be in agreement with each other within uncertainties. Performing the unfolding with the MultiFold method results in slightly larger total uncertainties than when using IBU. This is due, in part, to the NN initialization uncertainty, which arose due to the stochastic nature of the neural networks. It could potentially be reduced by producing additional models for inclusion in the ensemble. It may also be worth exploring the use of boosted decision trees, for example, to perform the reweighting necessary for the MultiFold algorithm. As they are not of a stochastic nature, they will produce a consistent result if the same training data is used. Such a study is beyond the scope of this thesis, but if tested the performance of the reweighting would need to be carefully evaluated to ensure it performs as well as the neural network reweighting. The other sources of increased uncertainty in the MultiFold result are caused by its intrinsic multidimensionality, and the additional statistical uncertainty from the dataset to which the final model is applied. The first is largely a consequence of the uncertainties associated with the tracks. In IBU, or other one dimensional unfolding methods, the uncertainties on the tracks used to build the jets will not affect the muon observables, but this will not be the case in a high dimensional unfolding. Both of these effects are inherent to the MultiFold method.

The MultiFold method is also used to perform a measurement using realistic pseudodata. This serves as the primary validation of the method as it allows for a comparison with a known truth target (from which the pseudodata was sampled). The agreement between the MultiFold result and the target is quantified by performing a χ^2 test and reporting the resulting p-value. Agreement within 2σ is seen for 22 of the 24 observables, the exceptions being the subleading track jet mass and the number of constituents in the leading track jet. This level of agreement can be explained by

the nature of the unfolding uncertainties. The structure of the internal correlations in both the basic unfolding uncertainty and the hidden variable uncertainty are not known. In the χ^2 test the uncertainties are assumed to be fully correlated between bins, but this is not necessarily the case for the unfolding uncertainties. Indeed, if both of these uncertainties are considered uncorrelated between bins, then both disagreeing observables come into agreement with the target. This indicates that in the case where a user is comparing results with a known target distribution it may be prudent to explore alternative correlation schemes for these two uncertainties.

The pseudodata results are further used to perform a series of tests of the MultiFold result to ensure that agreement with the target is maintained under extensions of the method. Firstly, three additional observables are calculated from the MultiFold result dataset. These are all found to be in agreement with the target. Secondly, a series of kinematic subregions are examined in order to highlight regions with: an increased p_{T}^{Zjj}, an enhanced electroweak Z+jets contribution, and an enhanced diboson Z+jets contribution. Once again, excellent agreement is maintained with the target with the exception of the subleading track jet mass in three of the five subregions, where it exhibits tension at the 2–3σ level. If the unfolding uncertainties are decorrelated, then this observable comes into agreement in two out of the three subregions, with the third continuing to exhibit tension in the 2–3σ range. Lastly, a series of two-dimensional distributions were constructed from the MultiFold result and agreement with the target was once again observed in all distributions except for the one involving the subleading track jet mass. Once more, agreement for this distribution is observed if the unfolding uncertainties are decorrelated.

In conclusion, the MultiFold method has been used successfully to perform the first unbinned 24 dimensional measurement. Furthermore, extensions of the method

that leverage these features have been demonstrated to perform well. These results will be published in the near future in the form of a dataset with a series of events containing the value of each of the 24 observables, and a series of weights corresponding to the nominal result and each of the systematic variations. A user guide will also be established to aid in instructing users on usage of the samples, including how to choose an appropriate binning, and considerations regarding the correlation schemes as discussed in relation to the unfolding uncertainties. By enabling observables to be calculated directly from the final results, allowing for the construction of multidimensional distributions, and giving flexible binning options, results in this format are of great utility to the particle physics community, and have the potential to set a new standard in the way measurements are made and shared going forward.

Appendix A

Additional data to Monte Carlo comparisons

This appendix presents the figures comparing the reconstructed MC and data distributions that were not presented in Section 7.5.

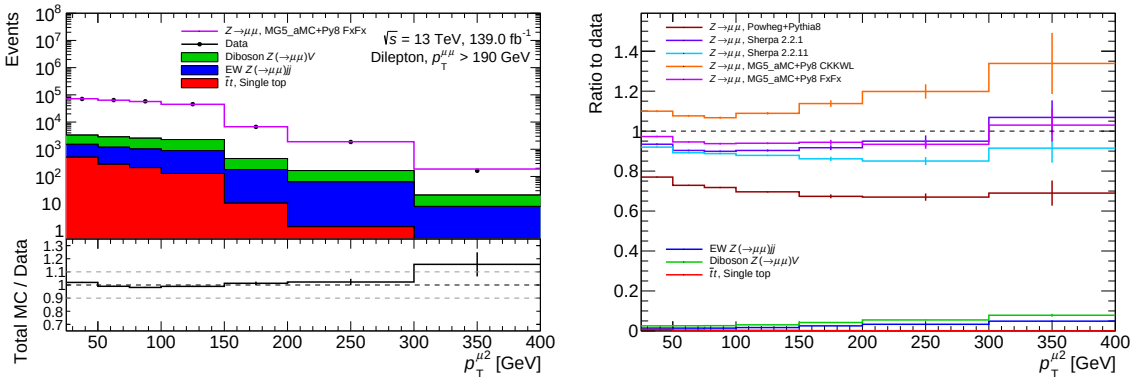


Figure A.1: MC predicted distribution for the subleading muon p_{T} compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

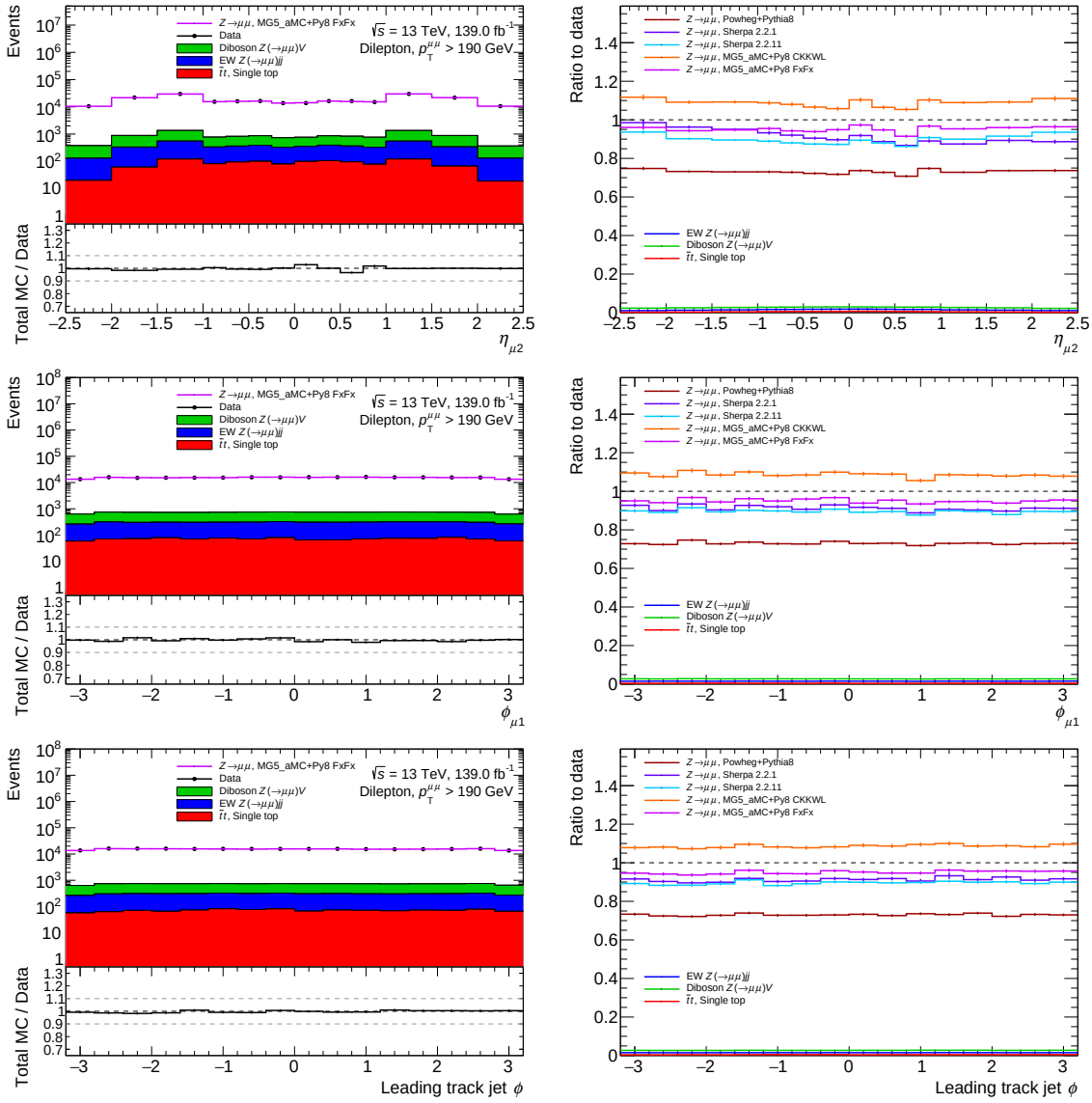


Figure A.2: MC predicted distributions for the subleading muon η, and the leading muon and leading track jet ϕ compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

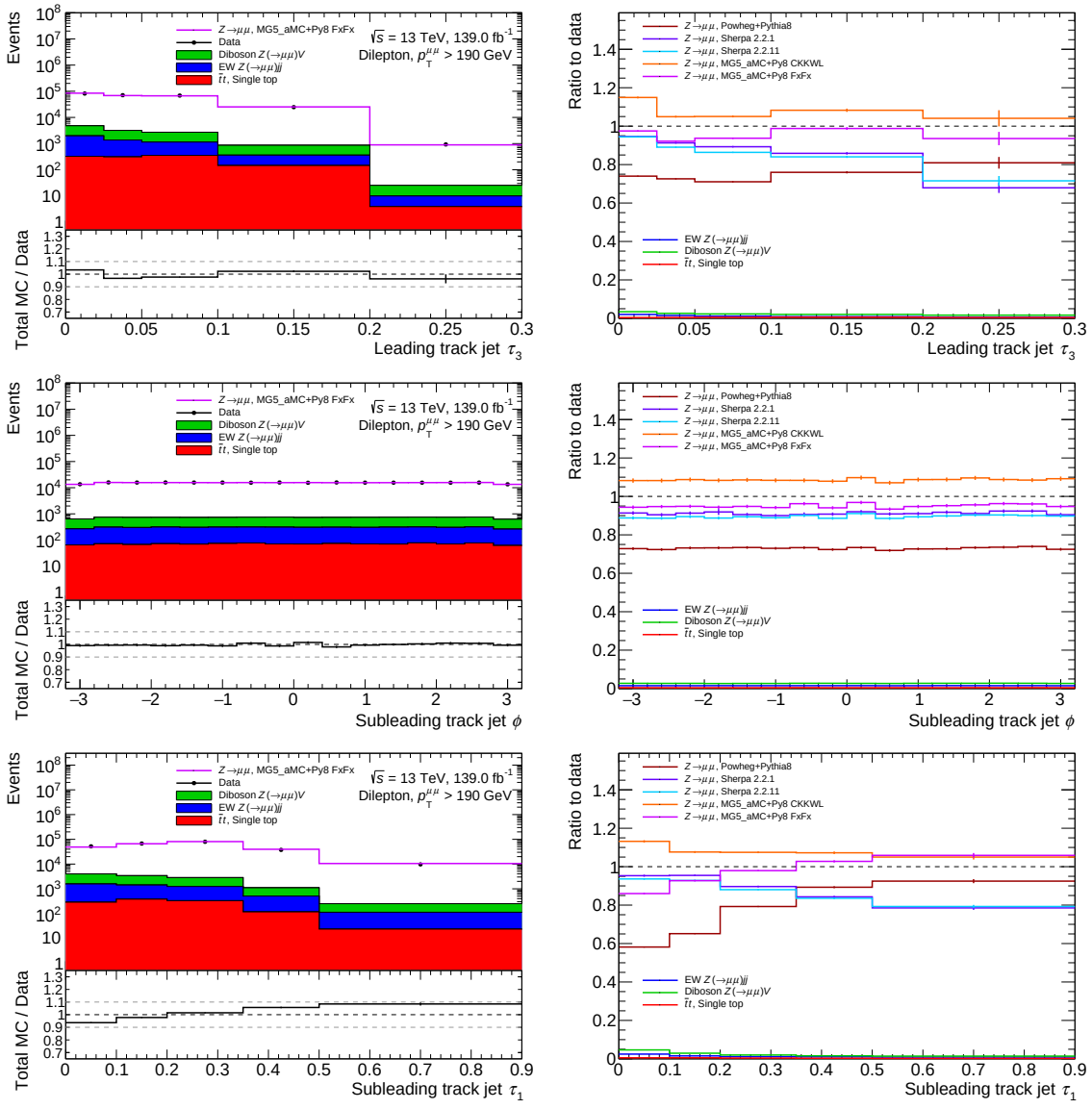


Figure A.3: MC predicted distributions for the leading track jet τ_3, and the subleading track jet ϕ and τ_1 compared with data are shown on the left using the MadGraph+Pythia8 FxFx sample as the strong Z+jets MC. The ratio to data for all of the MC samples for the same observables are shown on the right.

Appendix B

Sample reweighting validation plots

This appendix shows the results of the different sample reweightings performed in Section 7.7. First, the results of reweighting the MadGraph+Pythia8 FxFx sample such that the negative weights are eliminated is shown. Figures B.1 and B.2 show the results of the reweighting at the reconstructed level, while Figures B.3 and B.4 show the results at the truth level. Figures B.5 and B.6 show the results of reweighting the combined Sherpa 2.2.11, EW, and diboson "rest" samples such that they match the data.

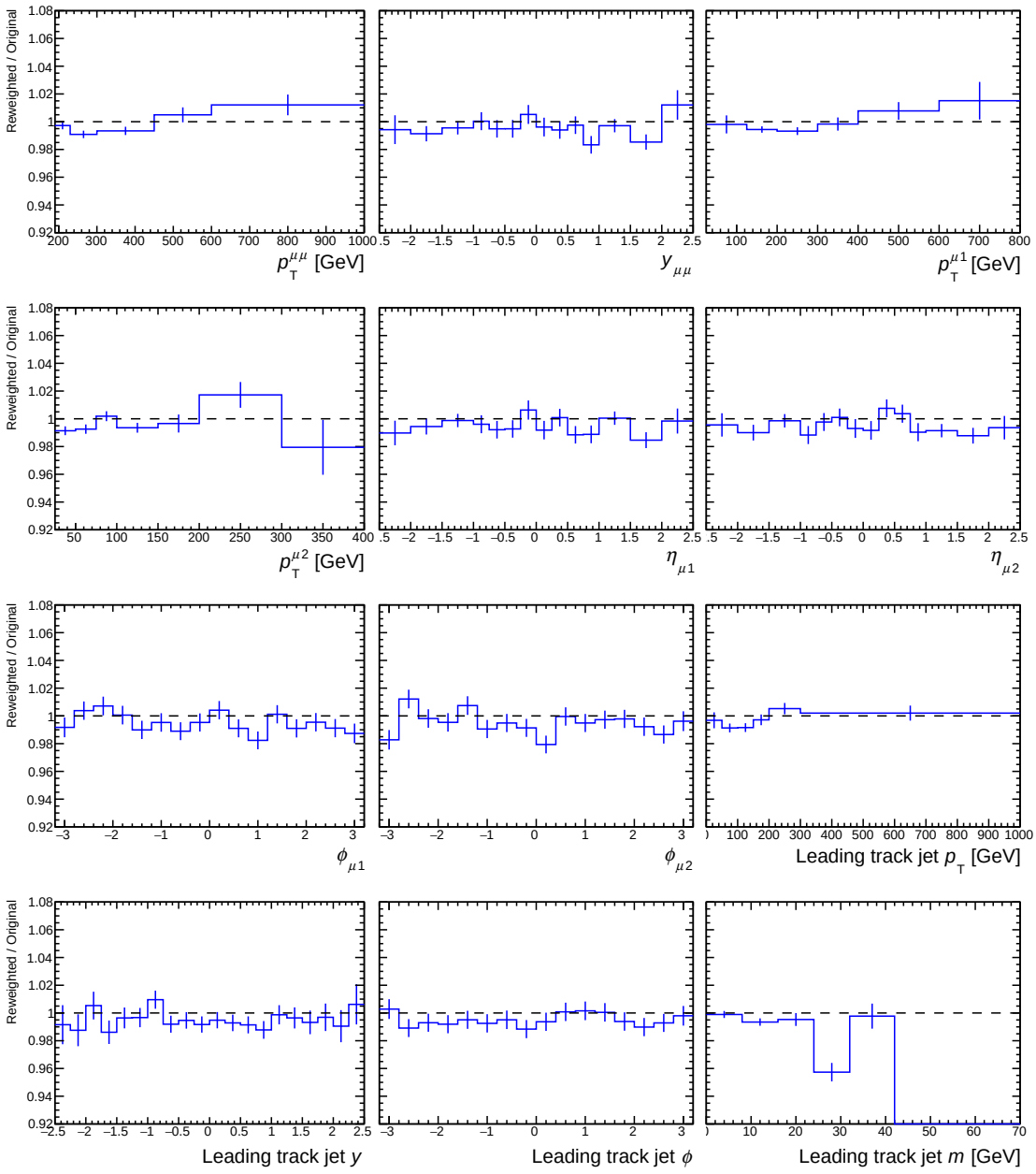


Figure B.1: The ratio between the reweighted and original MadGraph+Pythia8 FxFx distributions for the reconstructed dimuon p_{T} and rapidity, leading and subleading muon p_{T}, η and ϕ, and the leading track jet p_{T}, rapidity, ϕ and mass.

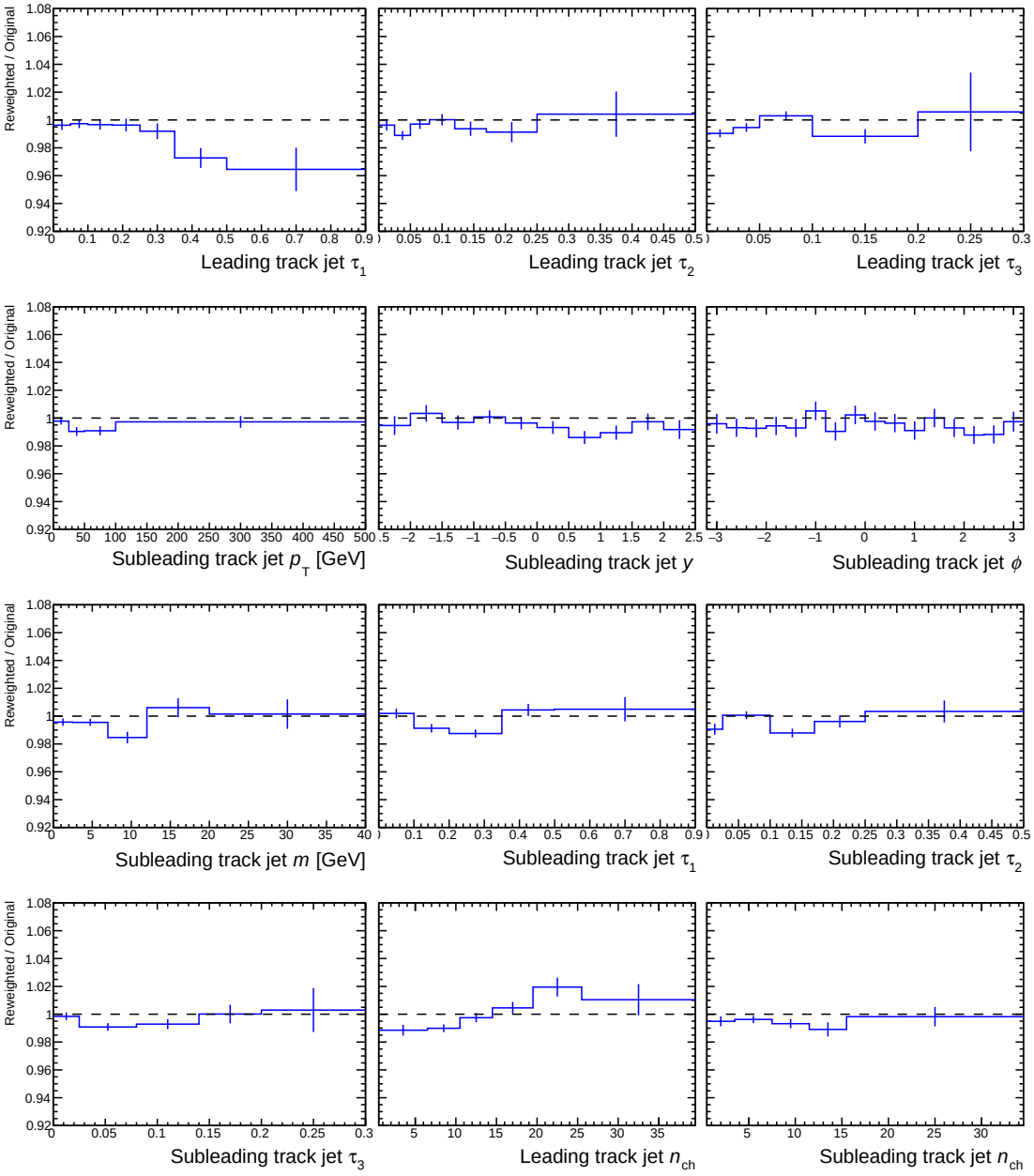


Figure B.2: The ratio between the reweighted and original MadGraph+Pythia8 FxFx distributions for the reconstructed leading track jet τ_1, τ_2 and τ_3, the subleading track jet p_{T}, rapidity, ϕ, mass, τ_1, τ_2 and τ_3, and the number of constituents in the leading and subleading track jet.

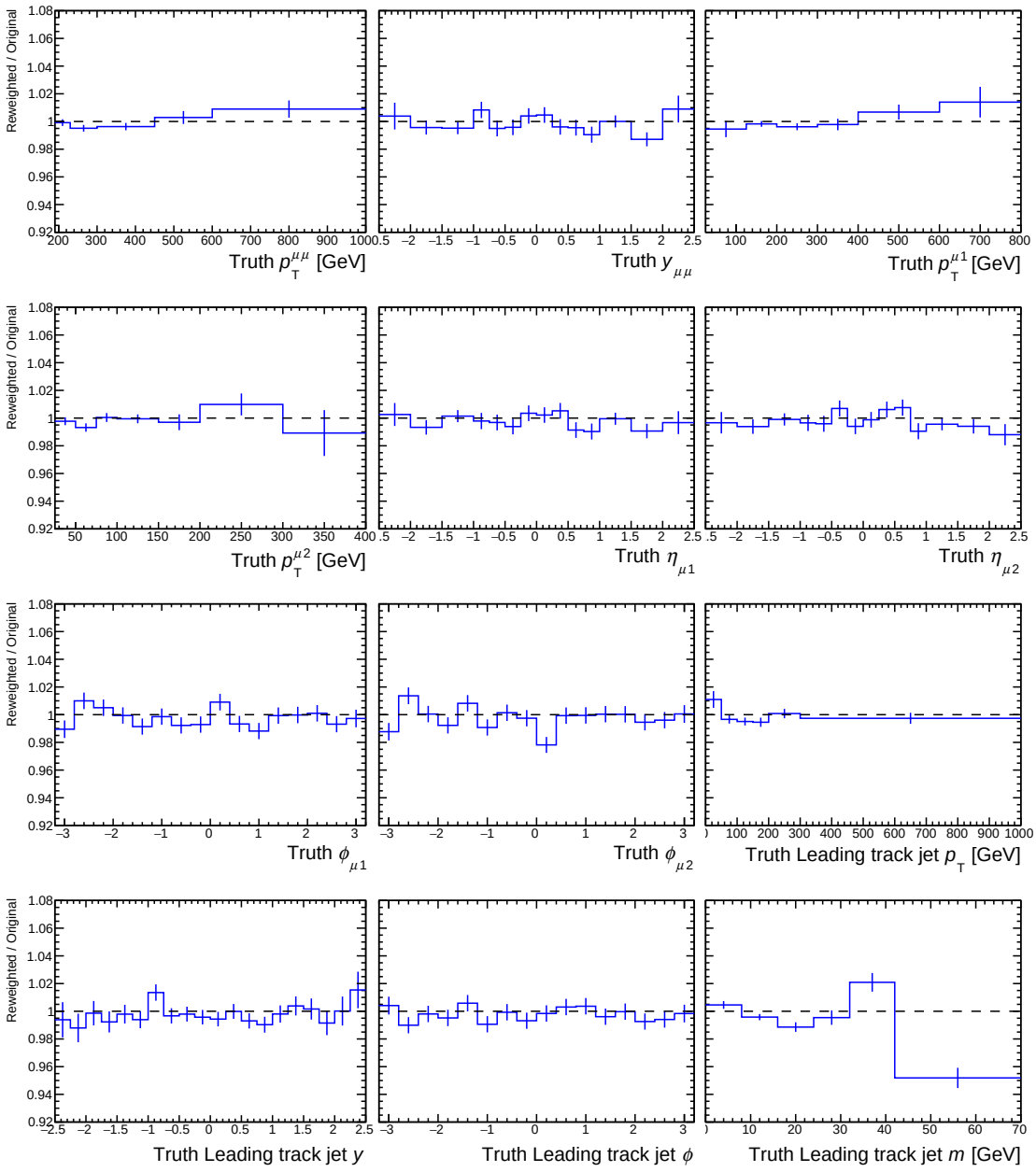


Figure B.3: The ratio between the reweighted and original MadGraph+Pythia8 FxFx distributions for the truth level dimuon p_{T} and rapidity, leading and subleading muon p_{T}, η and ϕ, and the leading track jet p_{T}, rapidity, ϕ and mass.

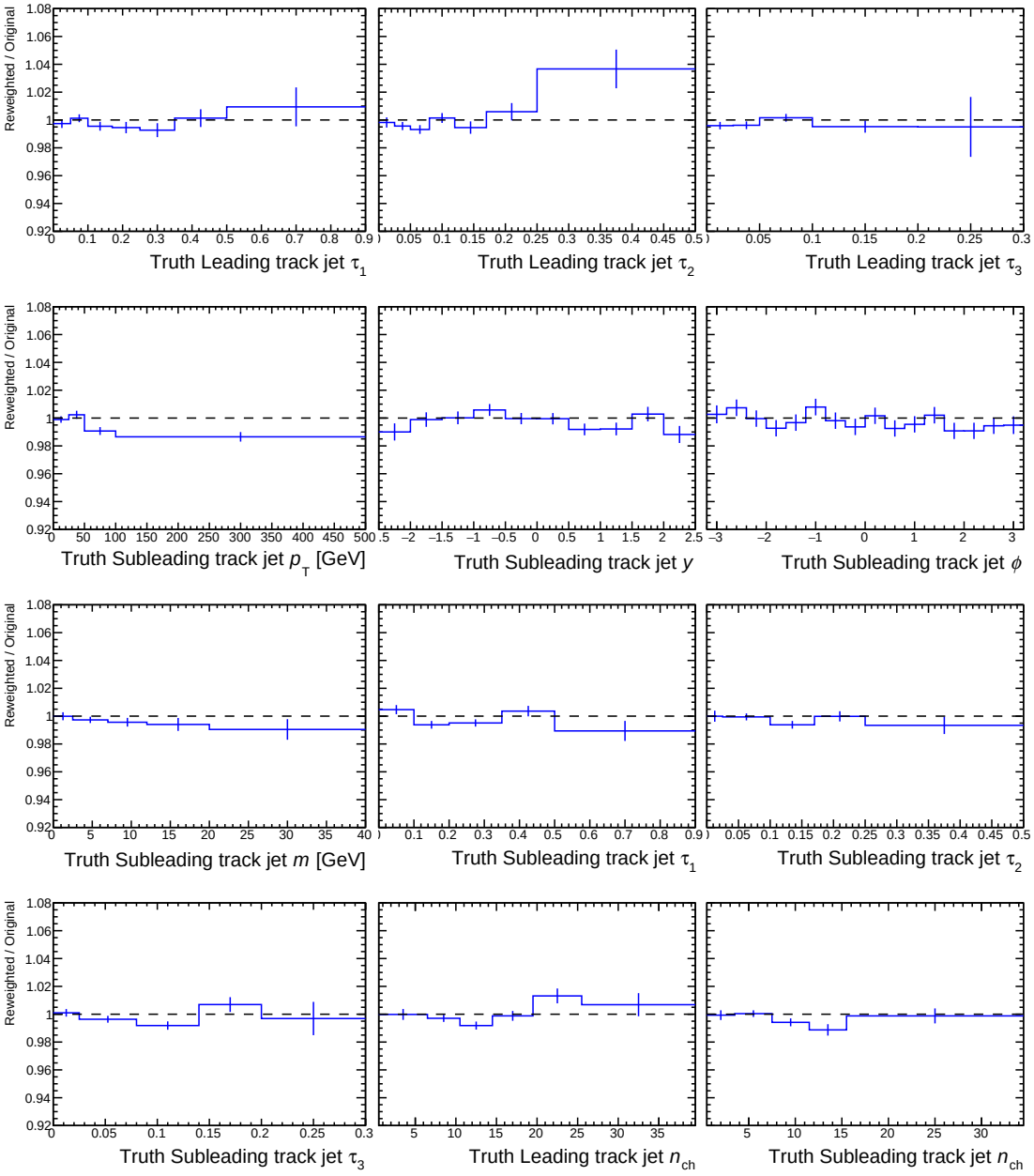


Figure B.4: The ratio between the reweighted and original MadGraph+Pythia8 FxFx distributions for the truth level leading track jet τ_1, τ_2 and τ_3, the subleading track jet p_T, rapidity, ϕ, mass, τ_1, τ_2 and τ_3, and the number of constituents in the leading and subleading track jet.

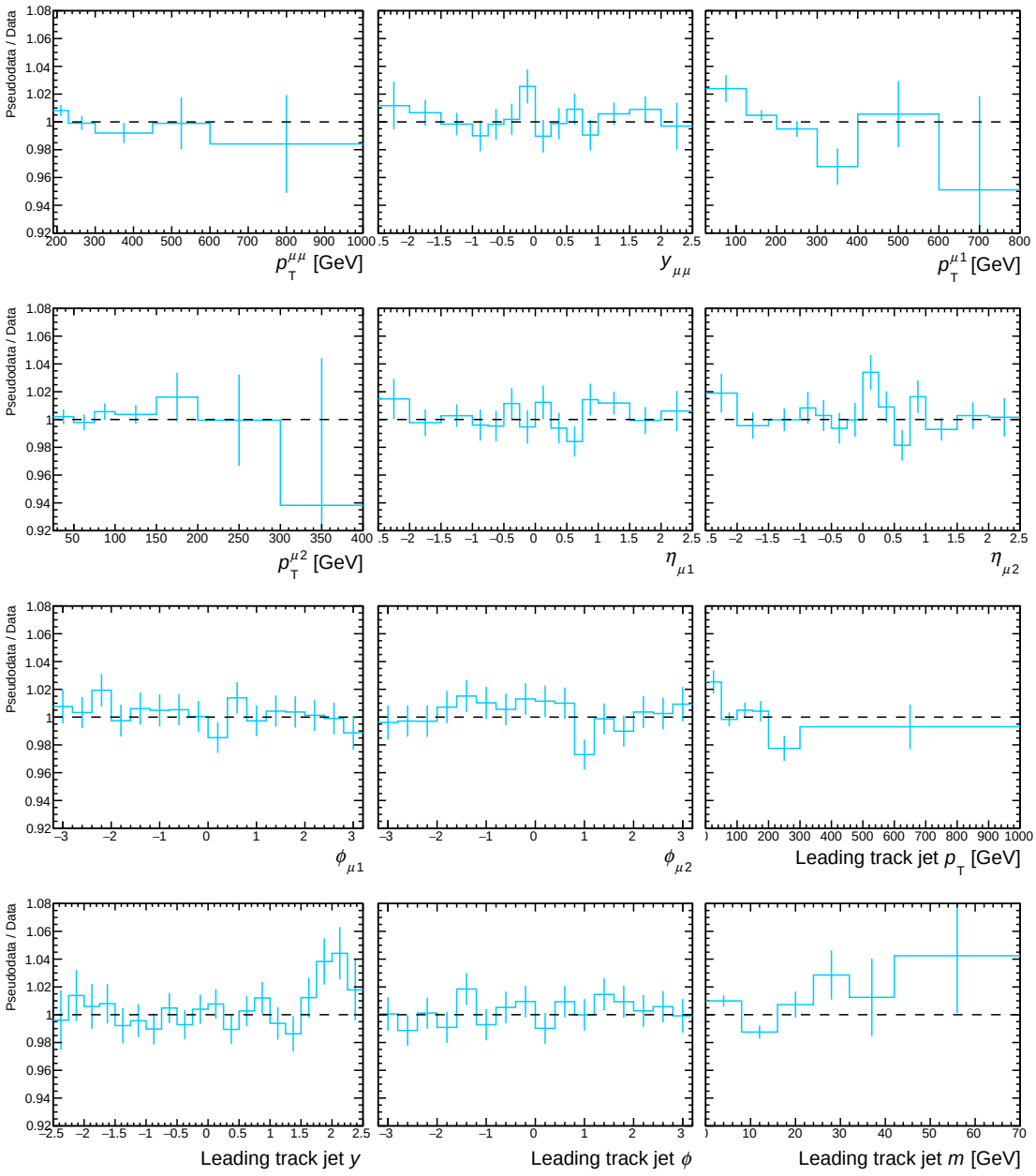


Figure B.5: The comparison between the pseudodata and Run 2 data distributions for the dimuon p_{T} and rapidity, the p_{T}, η and ϕ of the leading and subleading muon, and the p_{T}, y, ϕ and m of the leading track jet.

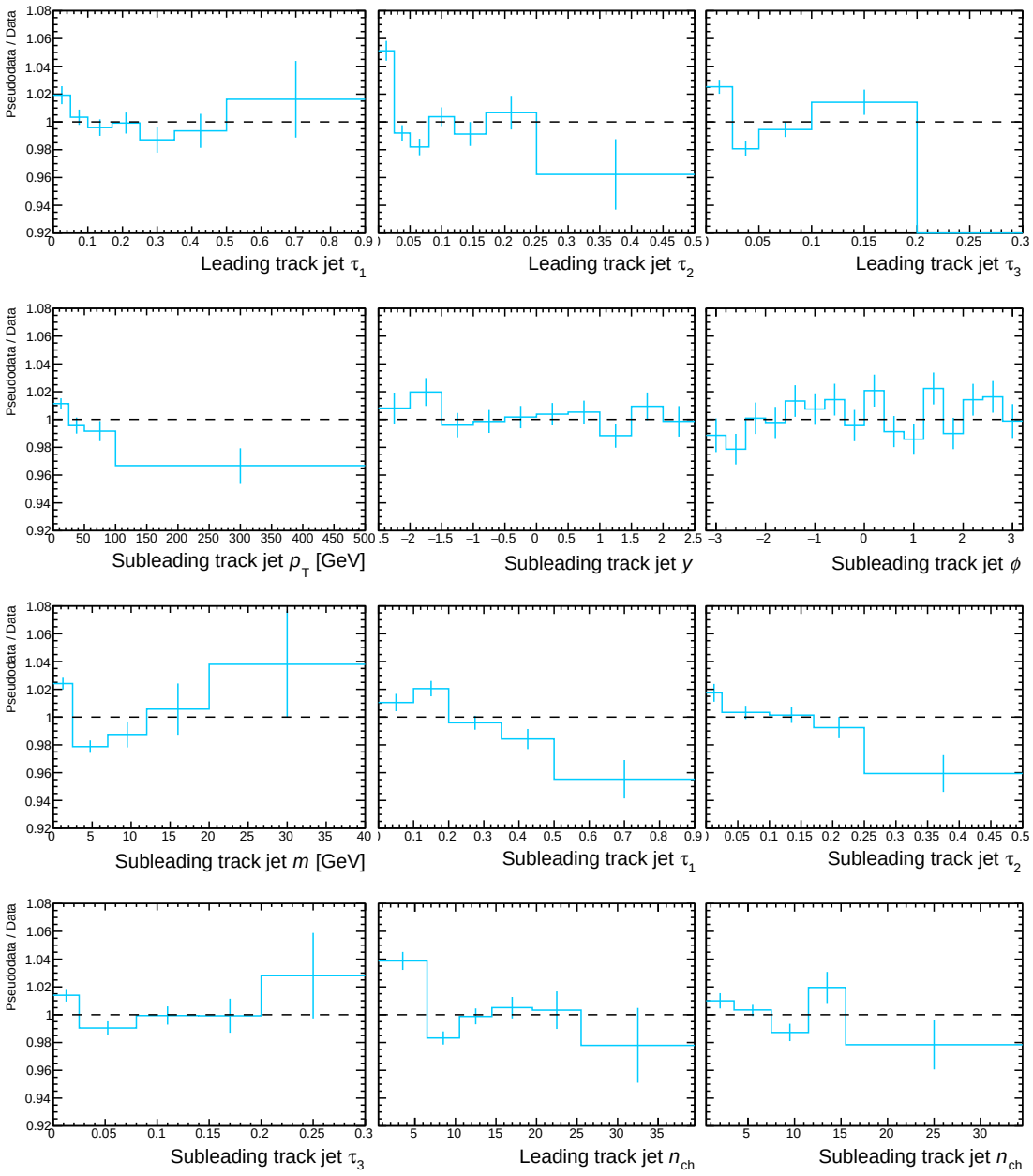


Figure B.6: The comparison between the pseudodata and Run 2 data distributions for the leading track jet τ_1, τ_2 and τ_3, the p_{T}, y, ϕ, m, τ_1, τ_2 and τ_3 of the subleading track jet, and the number of constituents of the leading and subleading track jet.

Appendix C

Additional uncertainty results

This appendix shows the remaining uncertainty plots that were not shown in Chapter 8.

C.1 Pseudodata results

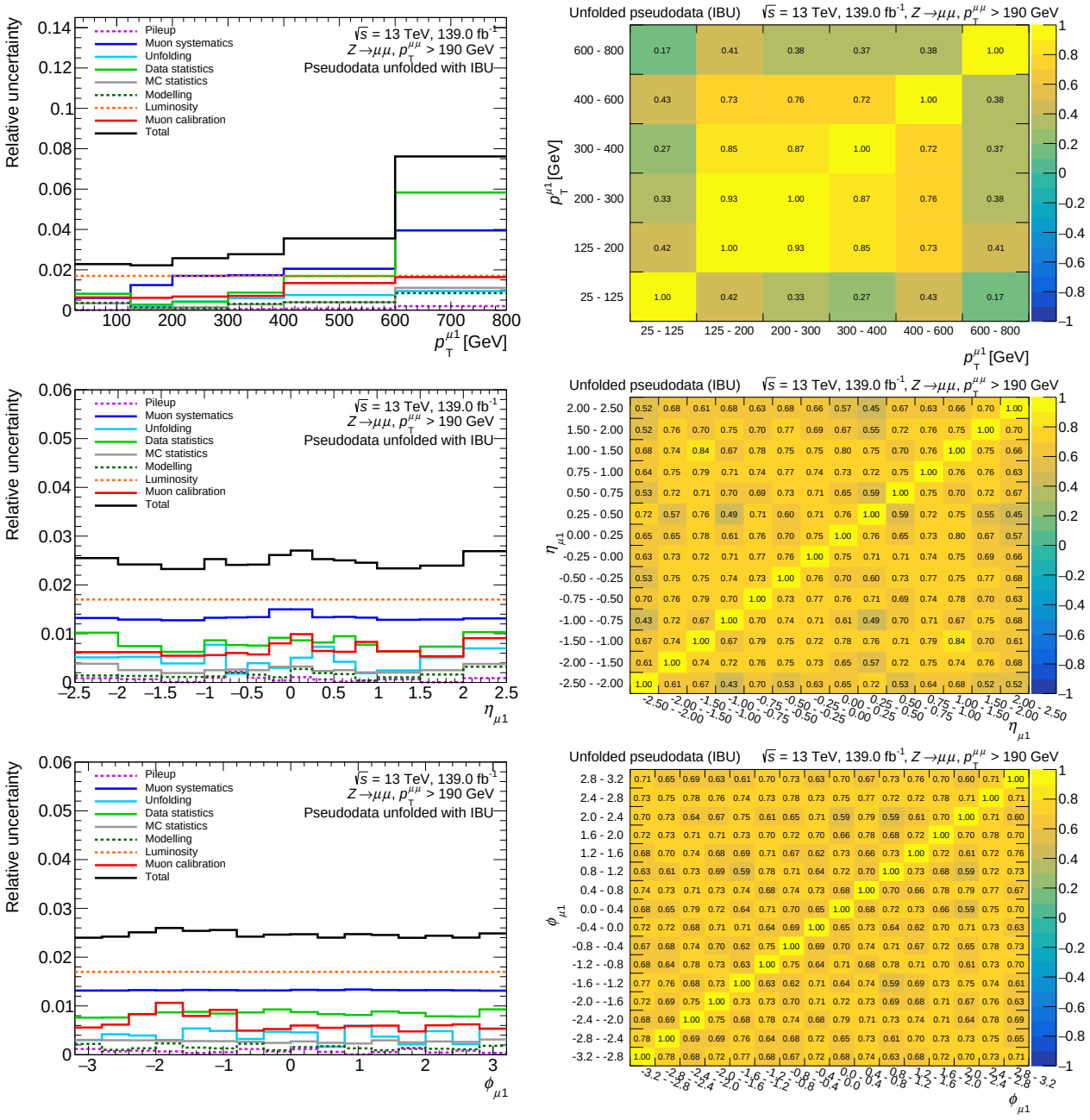


Figure C.1: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

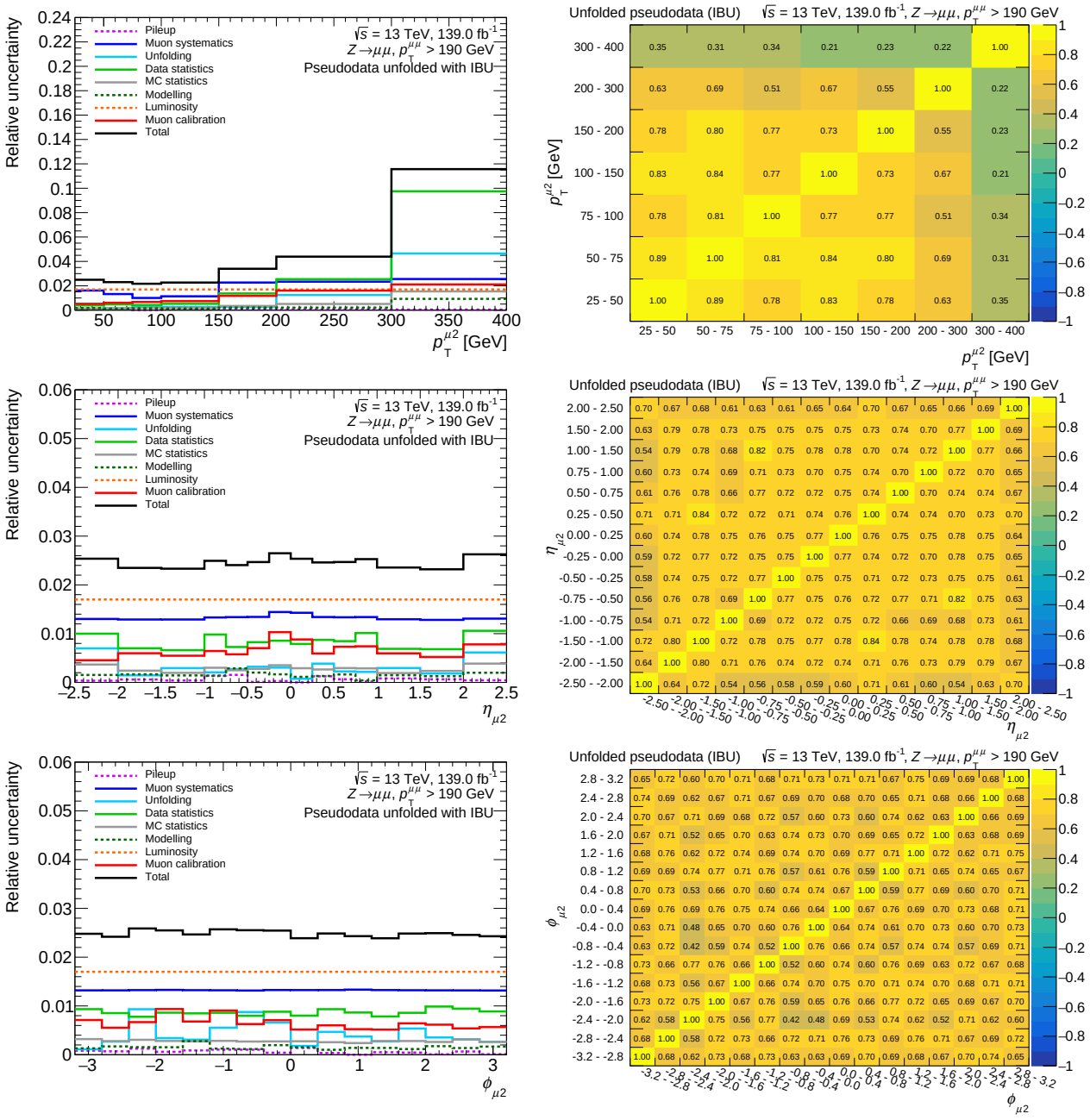


Figure C.2: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

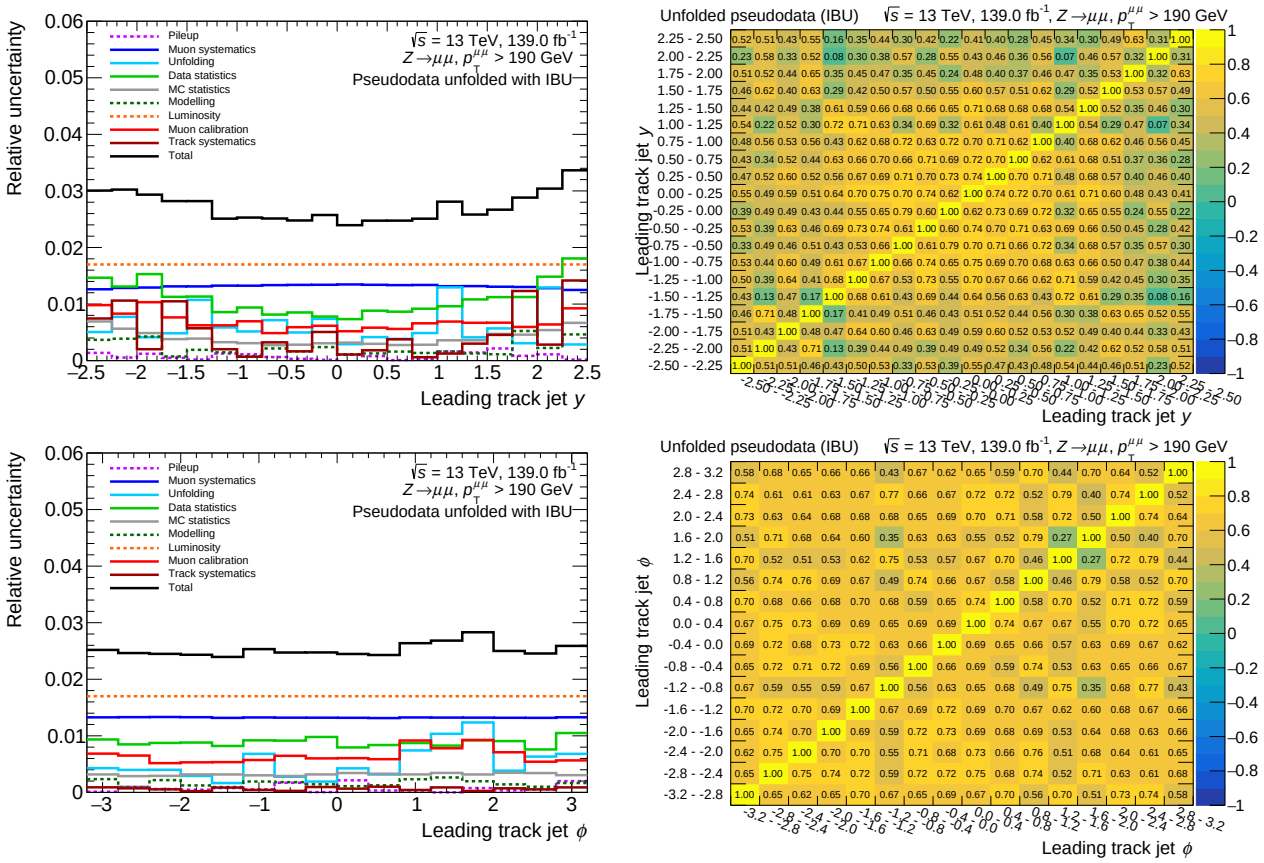


Figure C.3: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

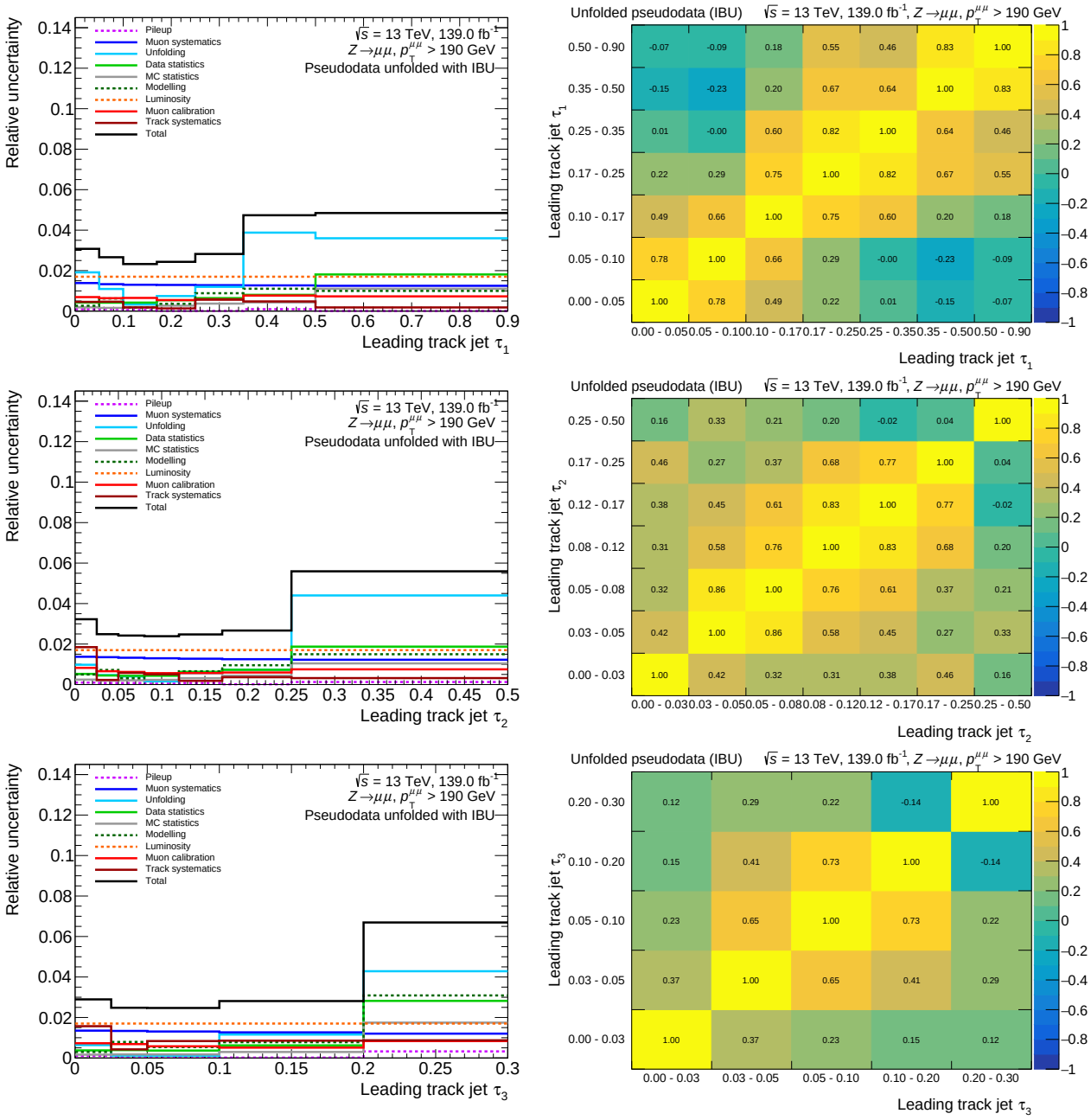


Figure C.4: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading track jet τ_1, τ_2 and τ_3 broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

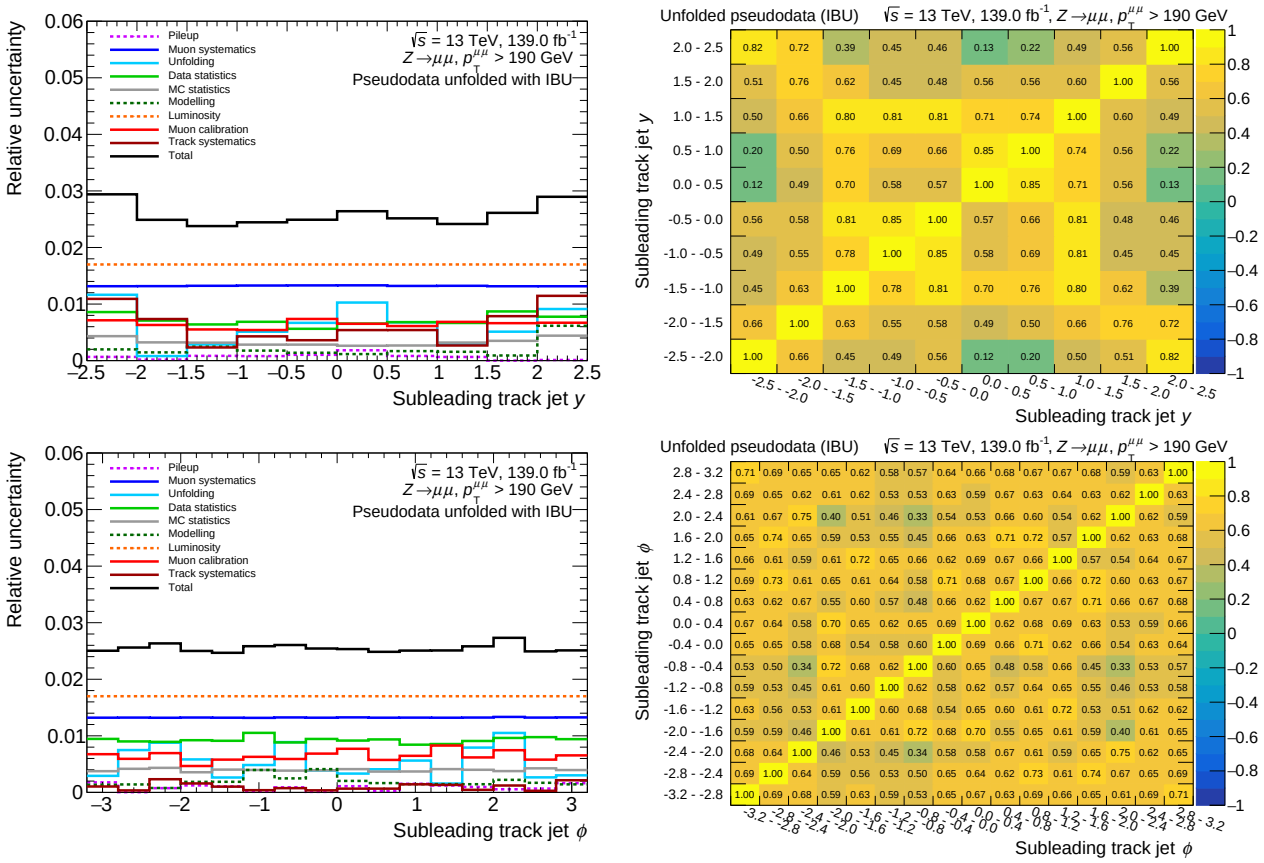


Figure C.5: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

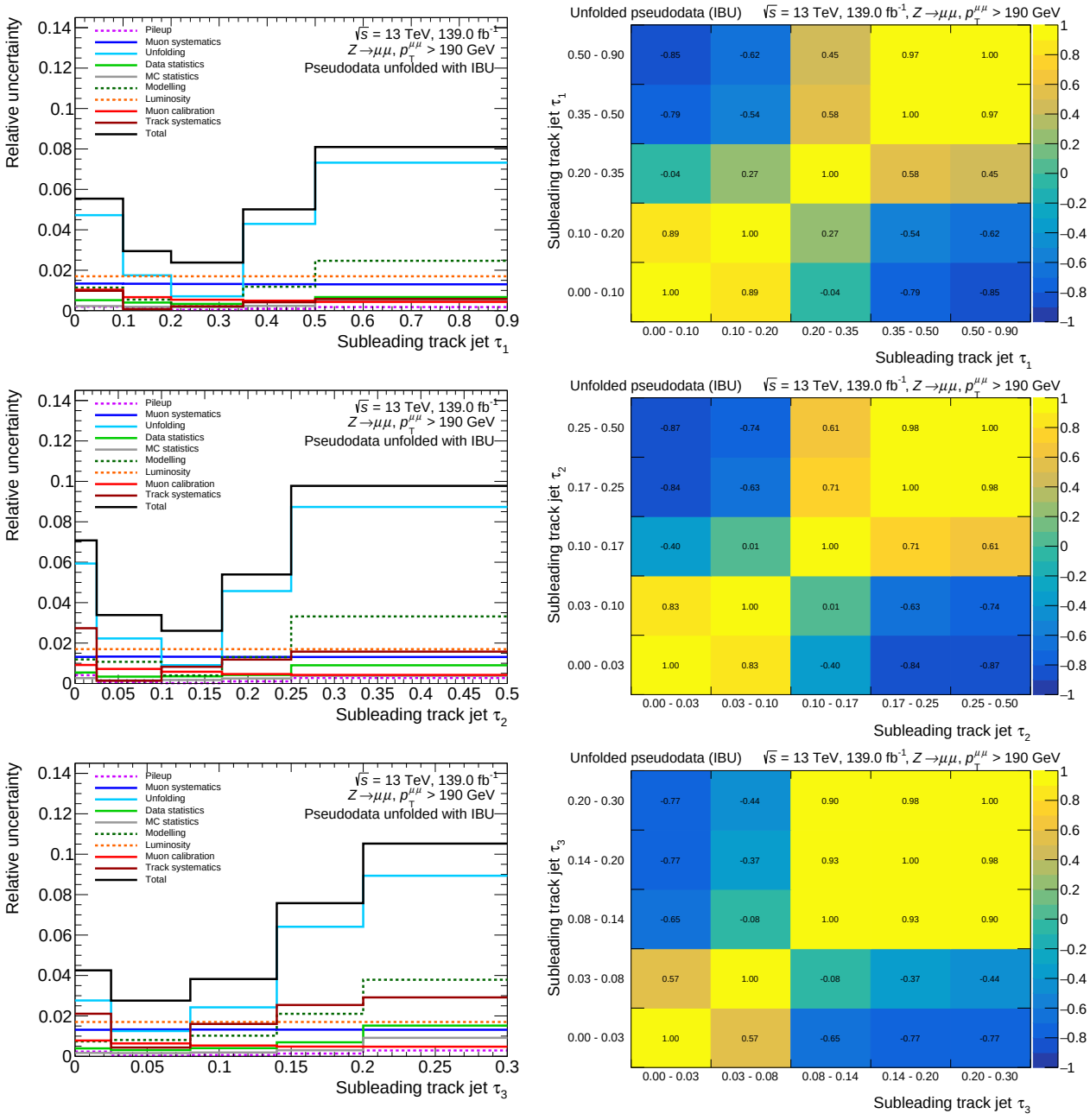


Figure C.6: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading track jet τ_1, τ_2 and τ_3 broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

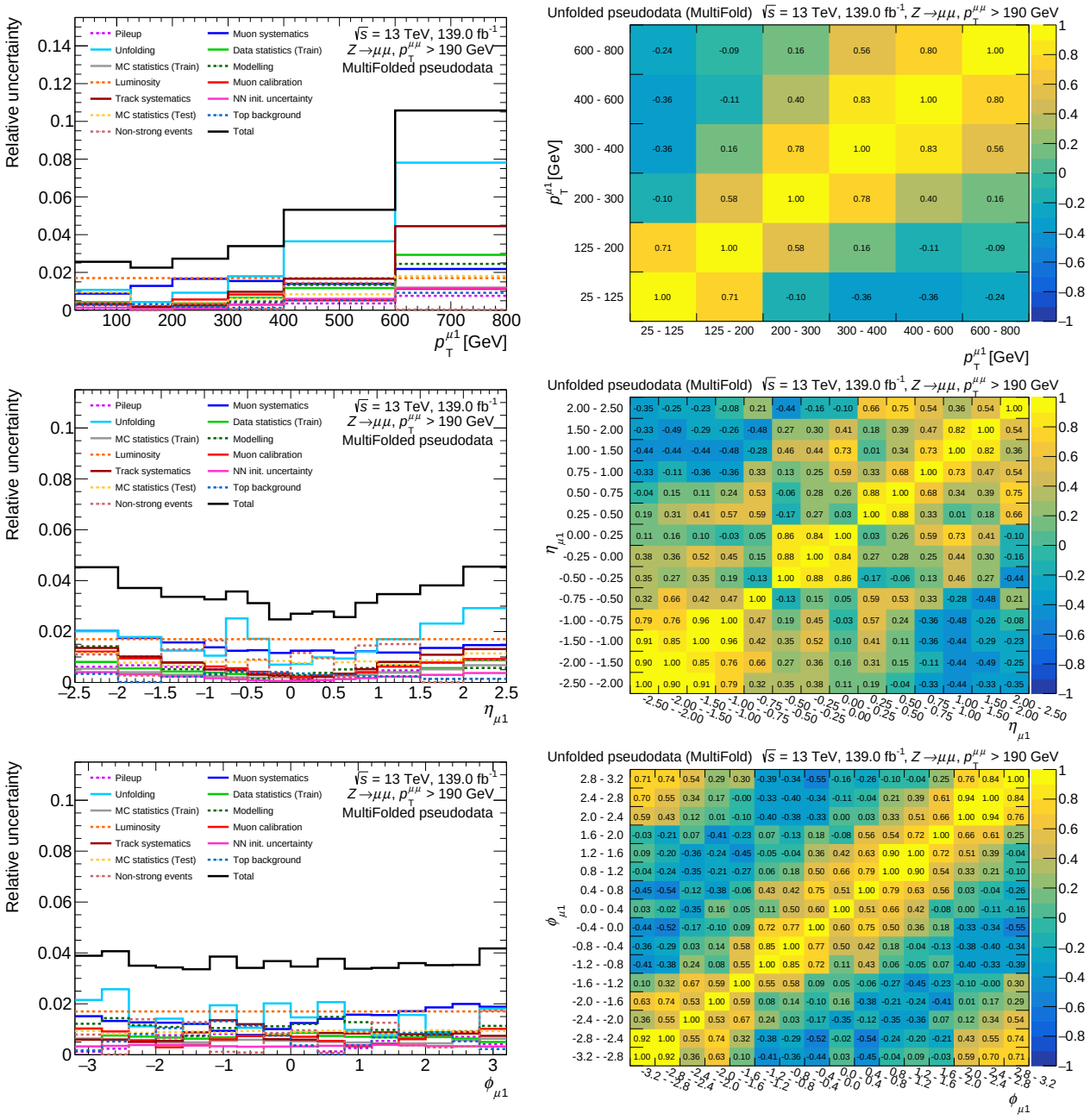


Figure C.7: The plots on the left show the relative uncertainty on the MultiFolded result for the leading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

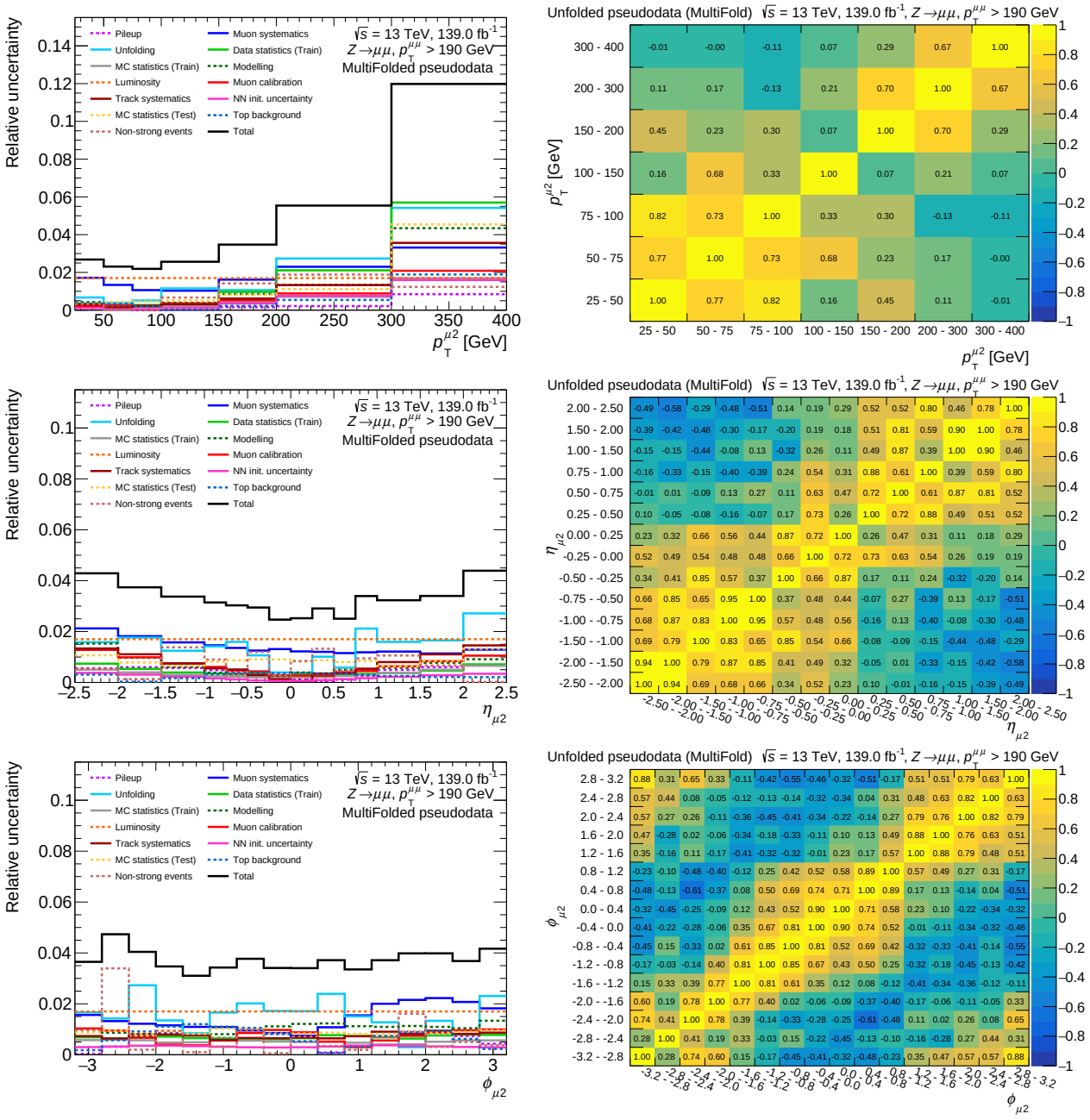


Figure C.8: The plots on the left show the relative uncertainty on the MultiFolded result for the subleading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

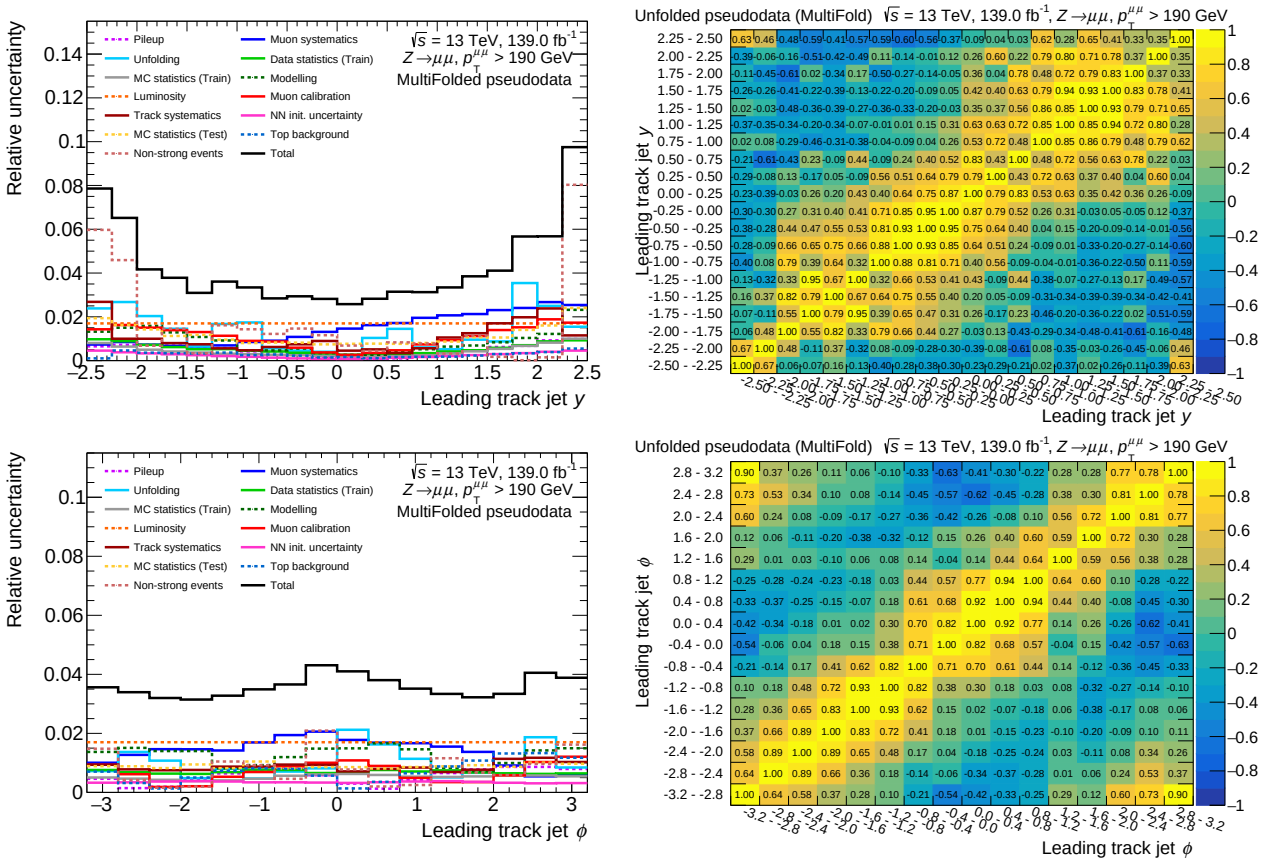


Figure C.9: The plots on the left show the relative uncertainty on the MultiFolded result for the leading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

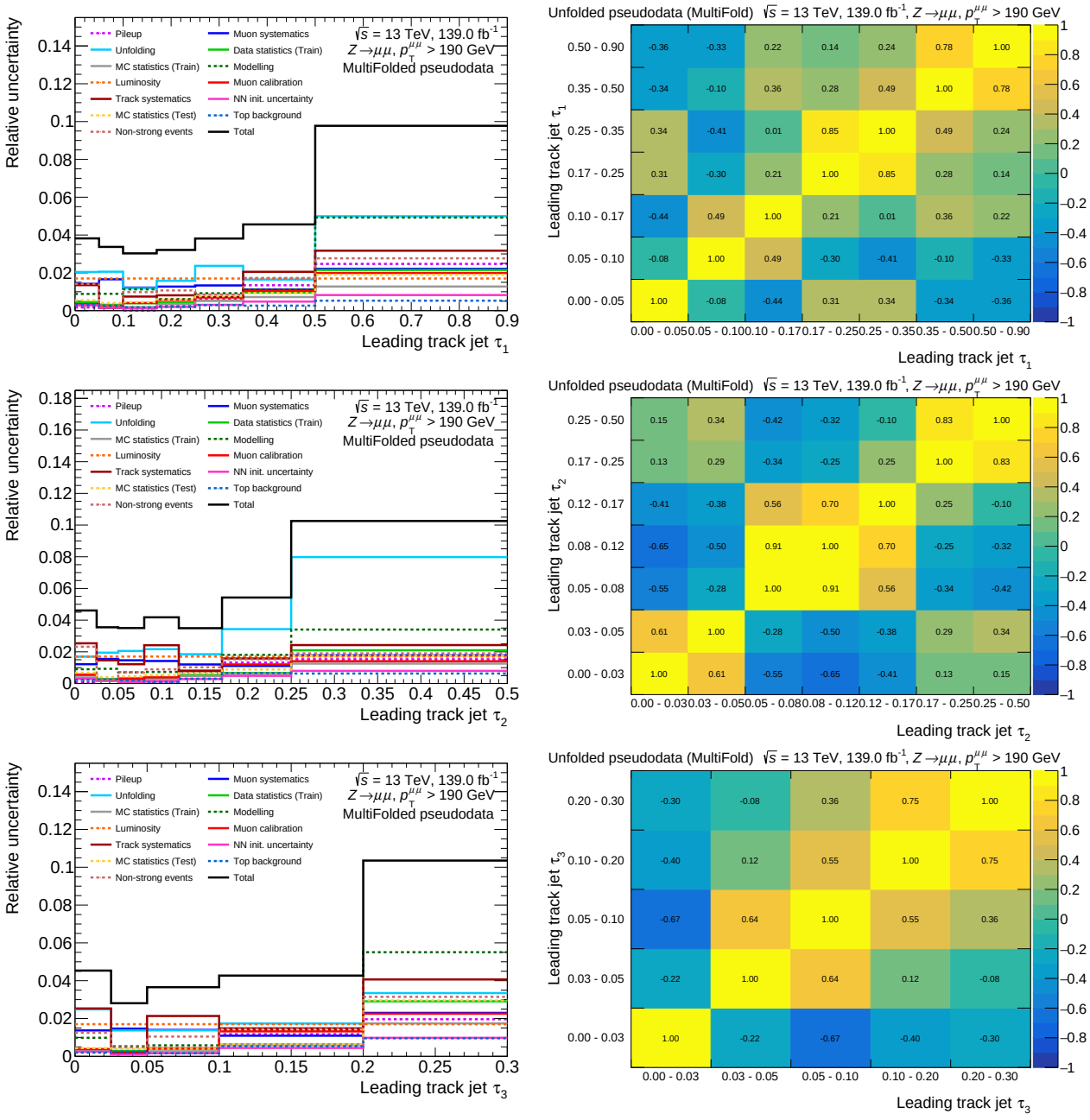


Figure C.10: The plots on the left show the relative uncertainty on the MultiFolded result for the leading track jet τ_1, τ_2 and τ_3 broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

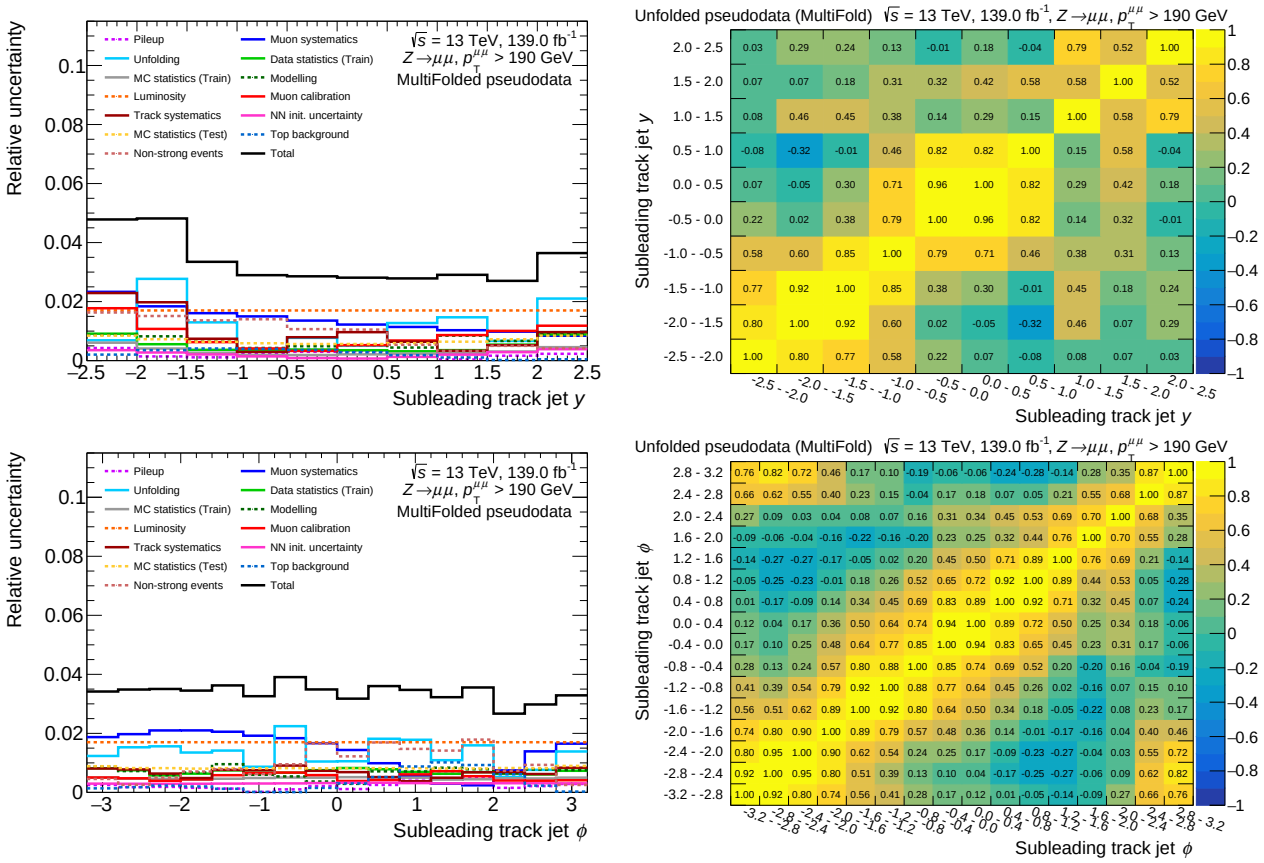


Figure C.11: The plots on the left show the relative uncertainty on the MultiFolded result for the subleading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

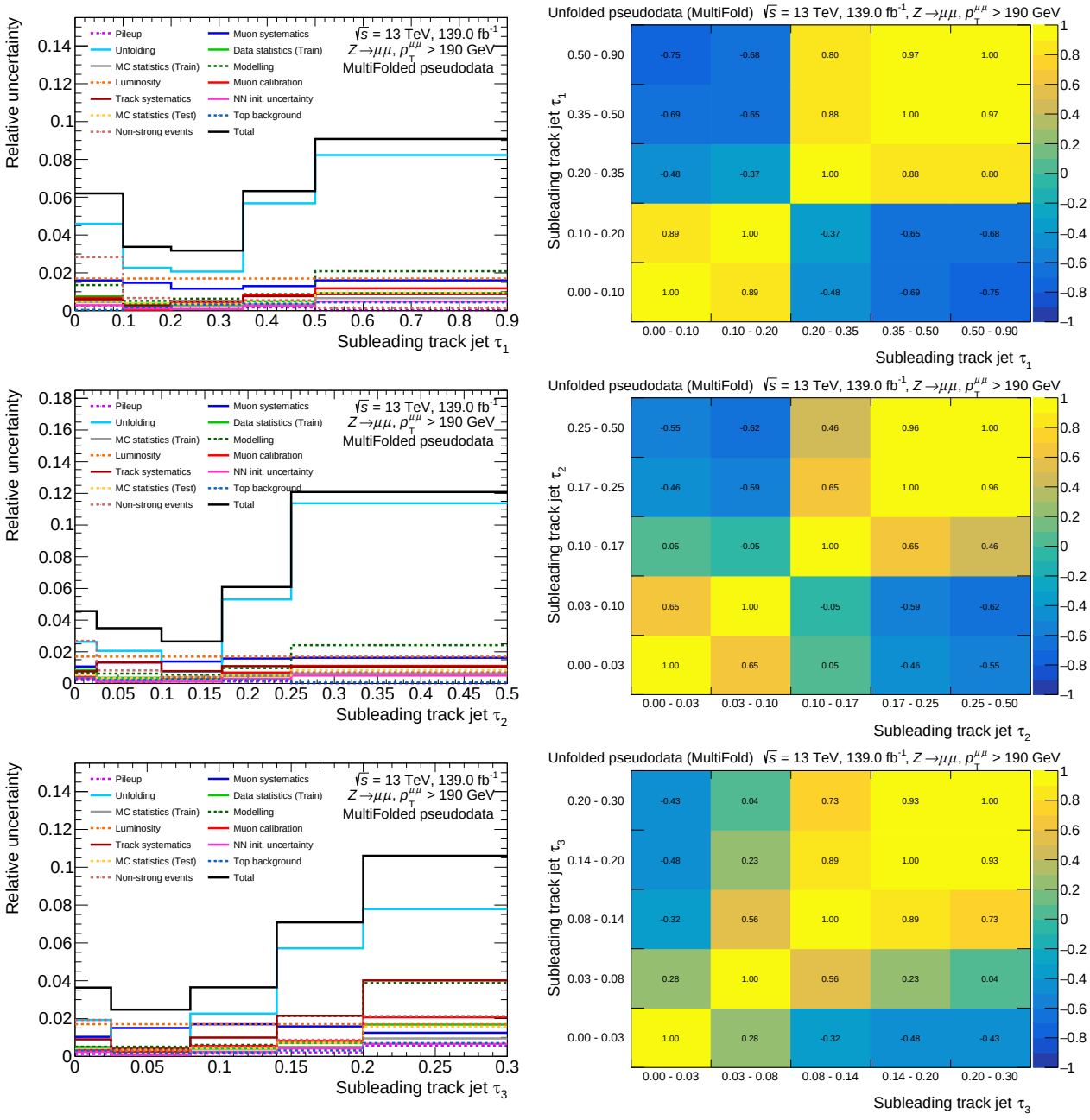


Figure C.12: The plots on the left show the relative uncertainty on the MultiFolded result for the subleading track jet τ_1, τ_2 and τ_3 broken down by source with the total uncertainty shown as the solid black line. The right shows the corresponding uncertainty correlation matrices for the same observables.

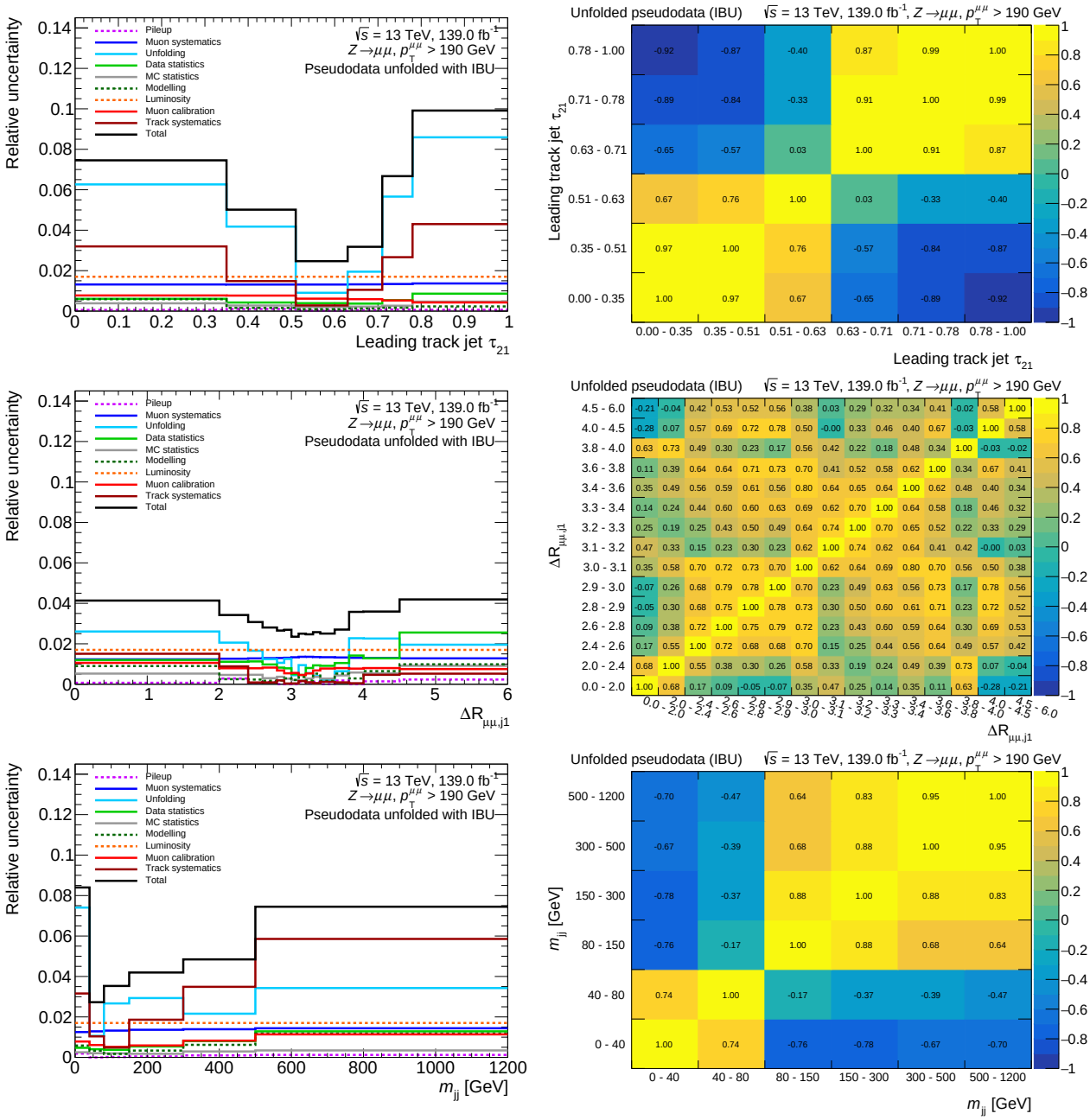


Figure C.13: The relative uncertainty on the unfolded result obtained using IBU broken down by source (left) and the corresponding uncertainty correlation matrix (right) for the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and the track dijet mass.

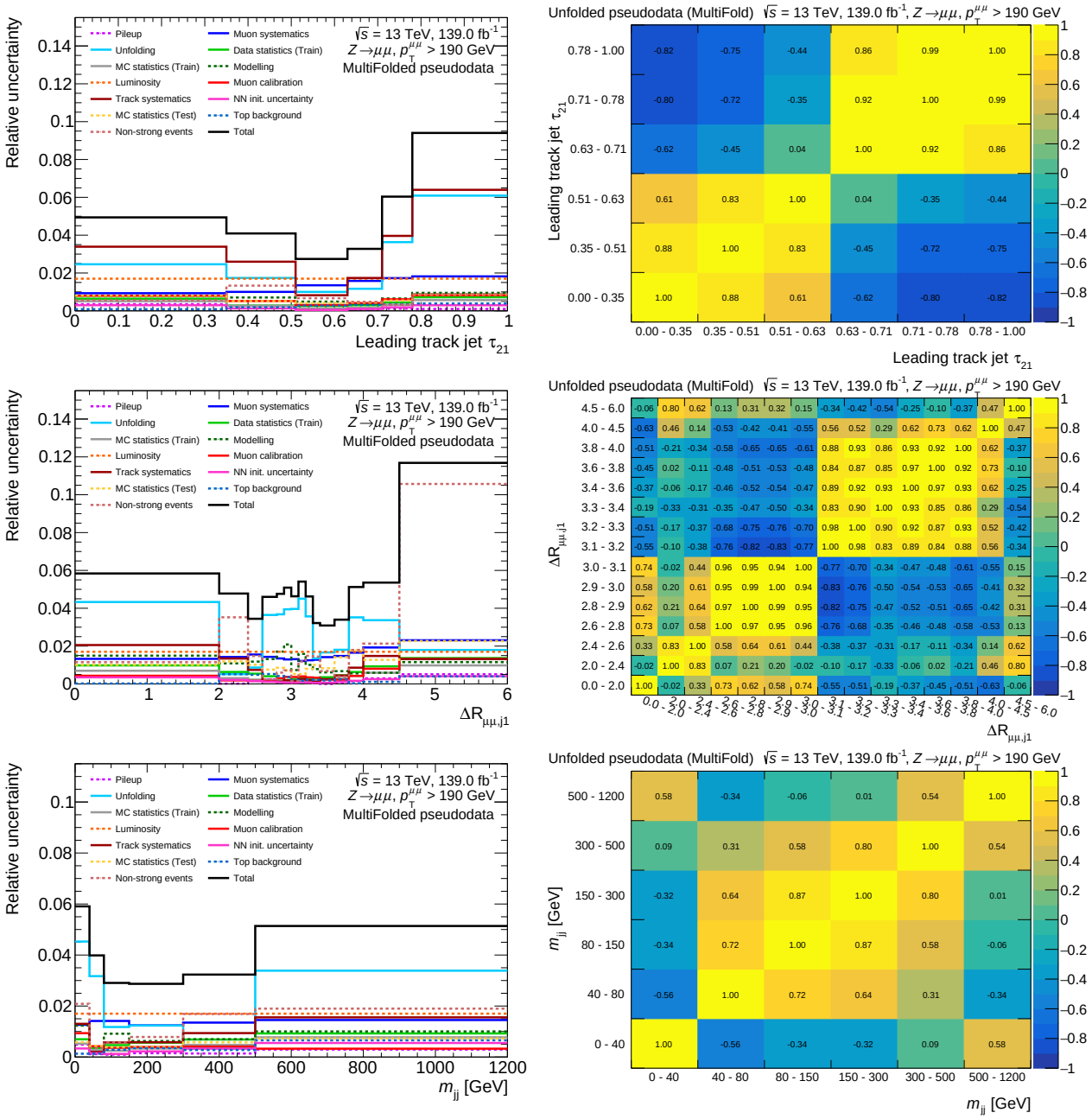


Figure C.14: The relative uncertainty determined for the unfolded result obtained by calculating the observable after performing the unfolding using the MultiFold method (left) and the corresponding uncertainty correlation matrix (right) for the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and the track dijet mass.

C.2 Data results

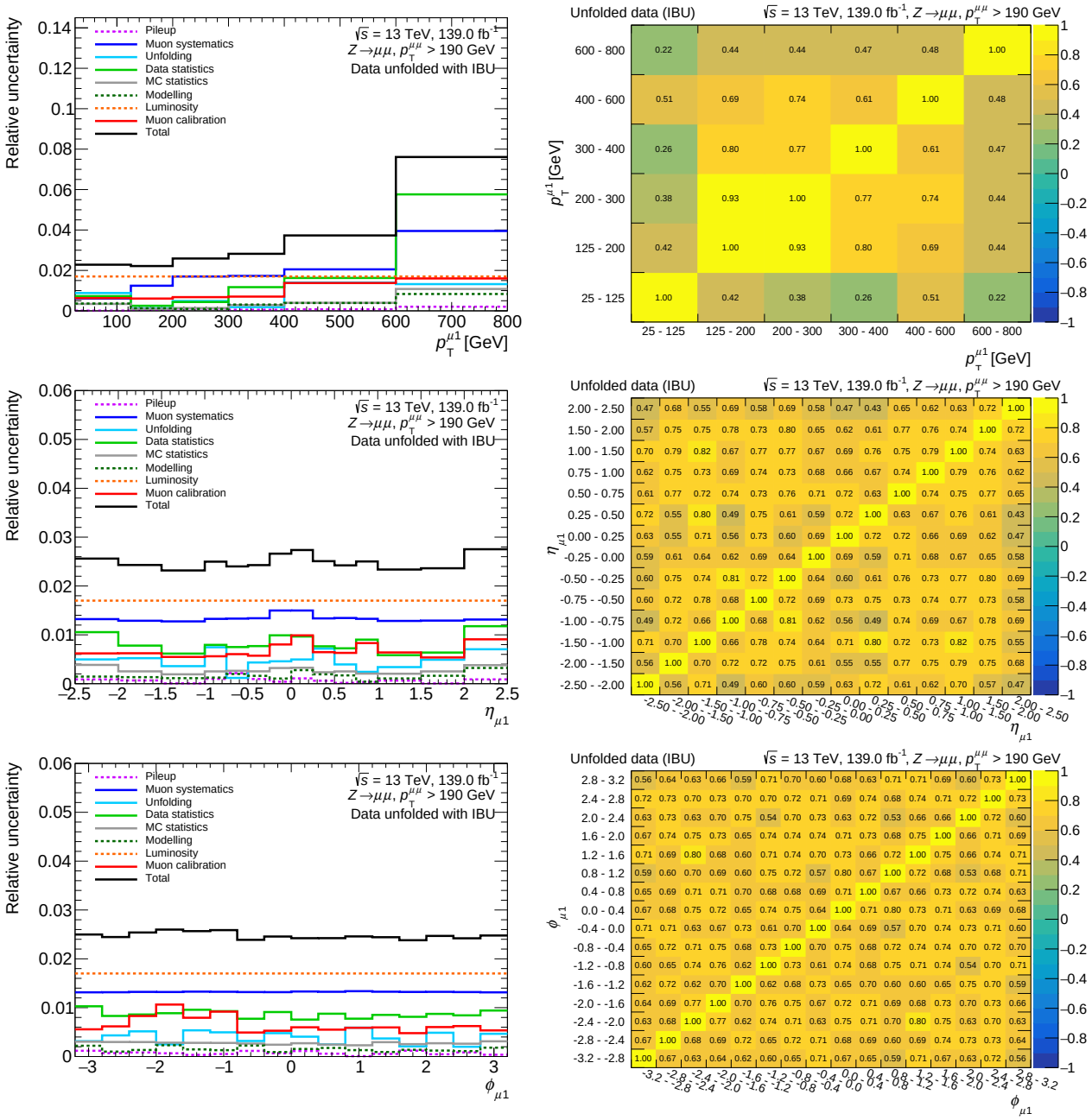


Figure C.15: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

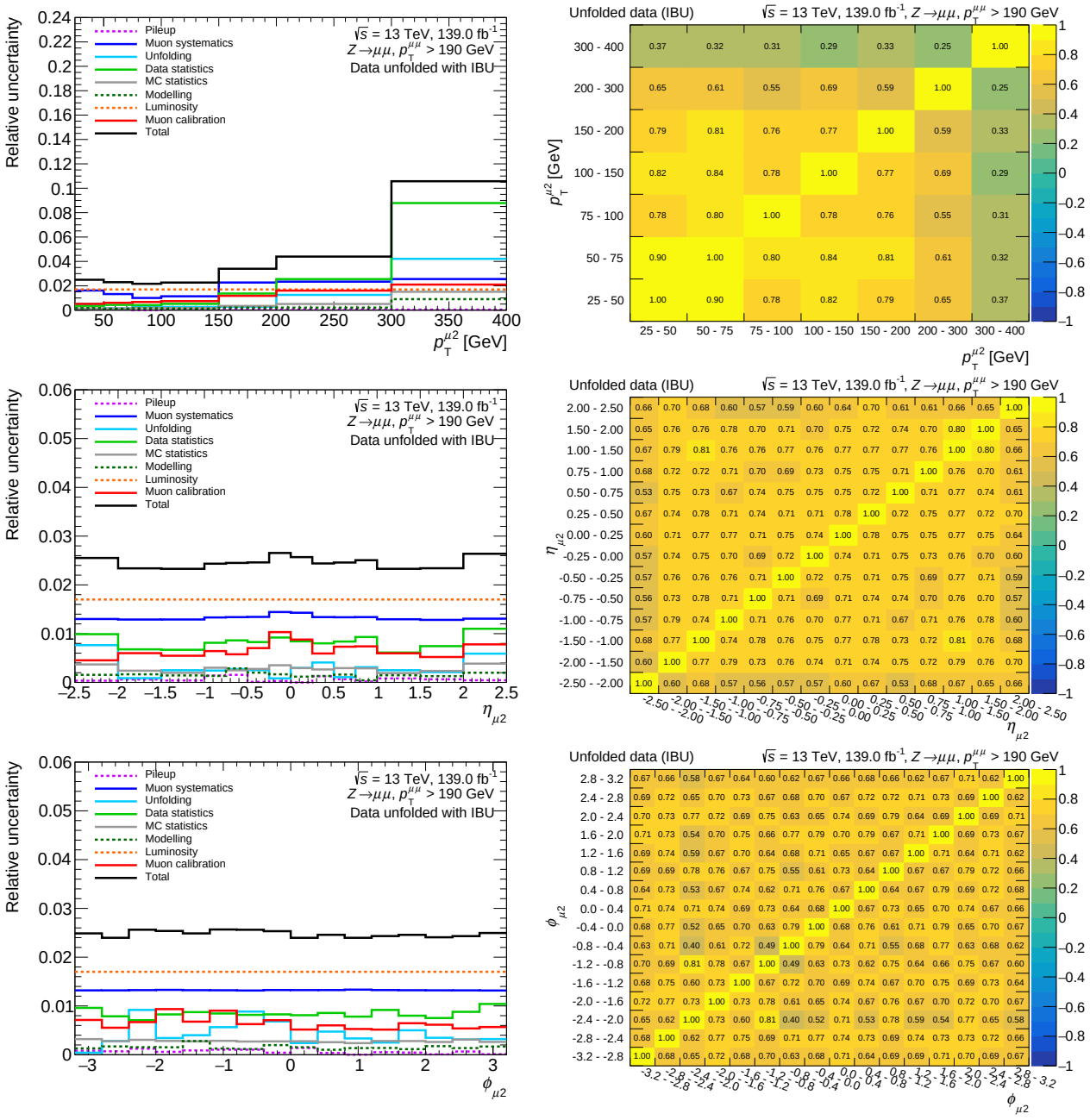


Figure C.16: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

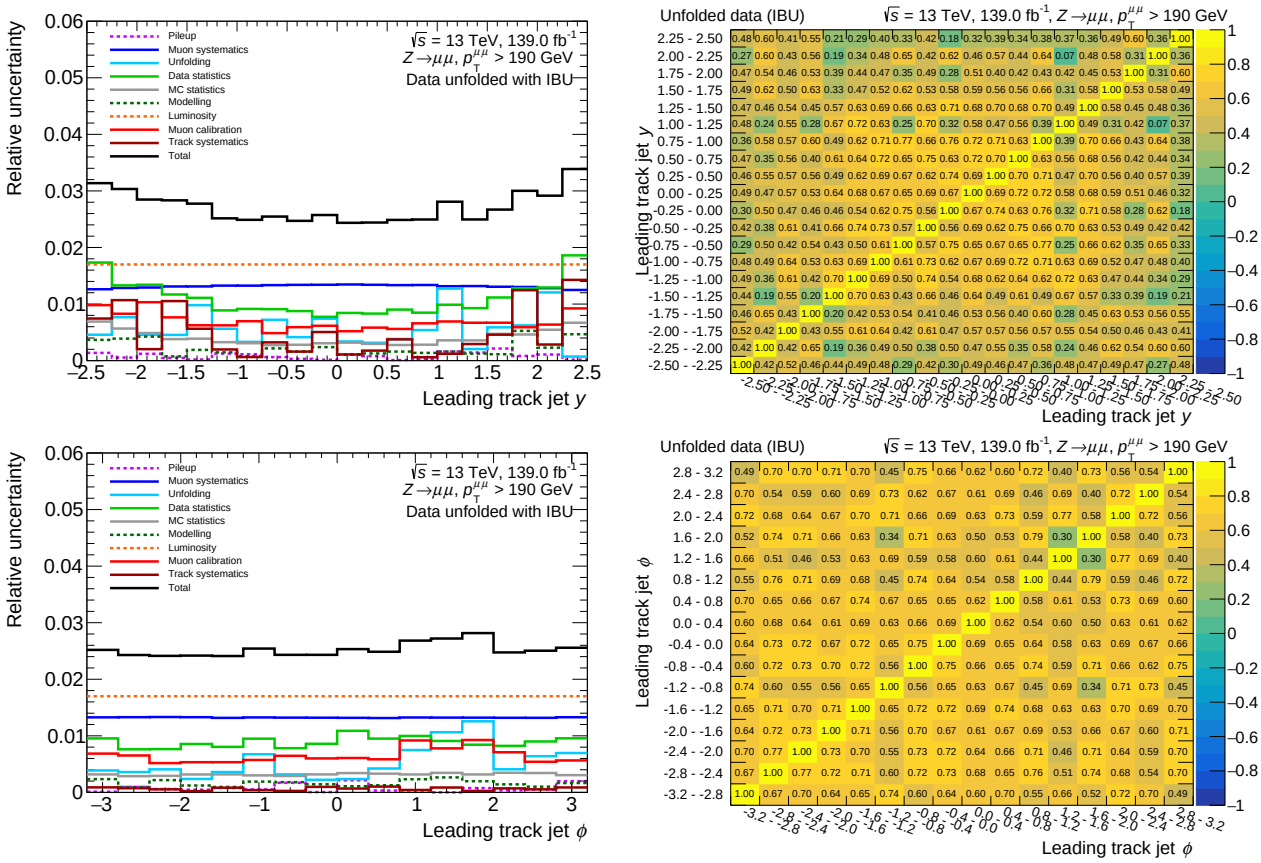


Figure C.17: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

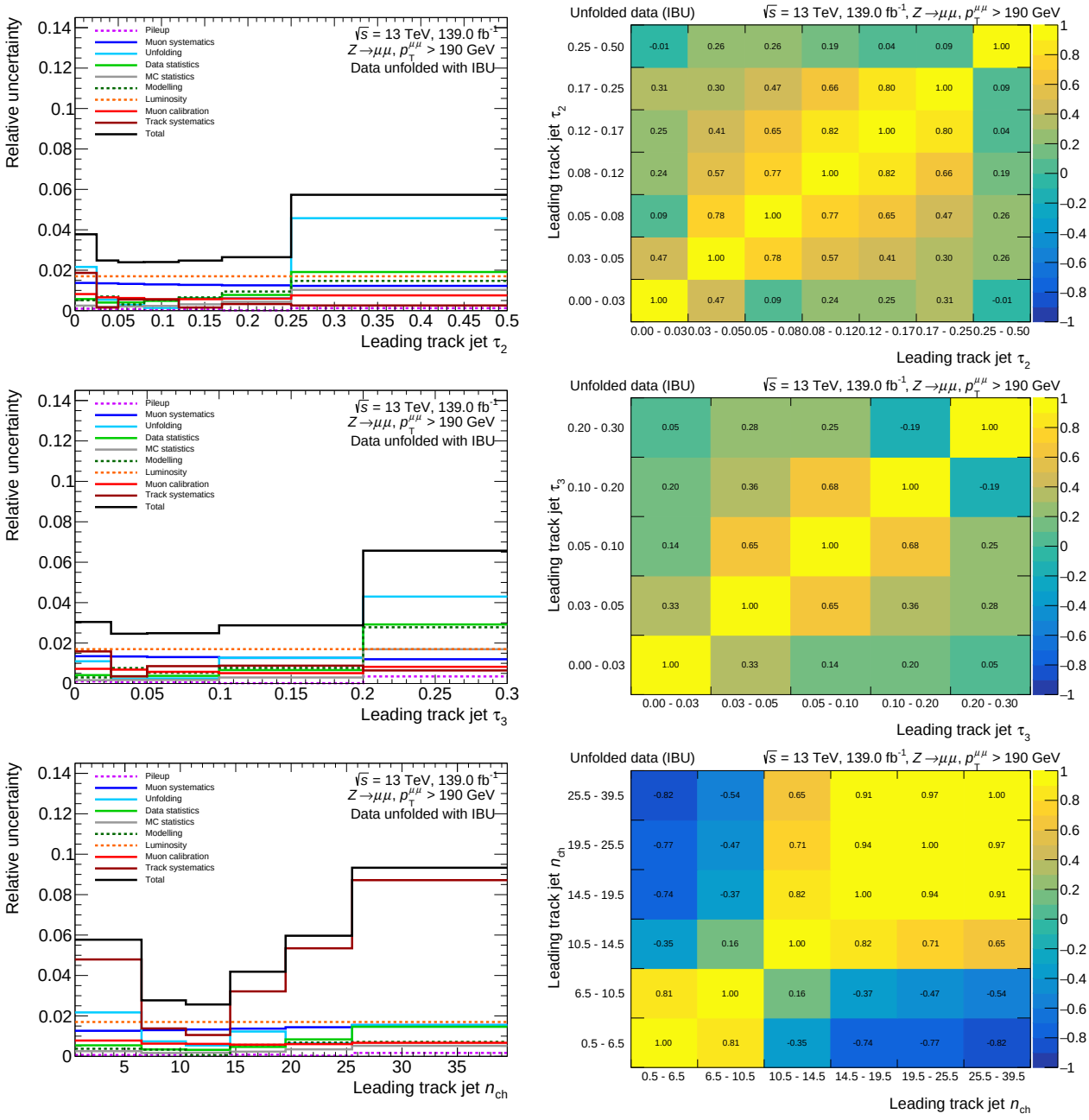


Figure C.18: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the leading track jet τ_2, τ_3, and the number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

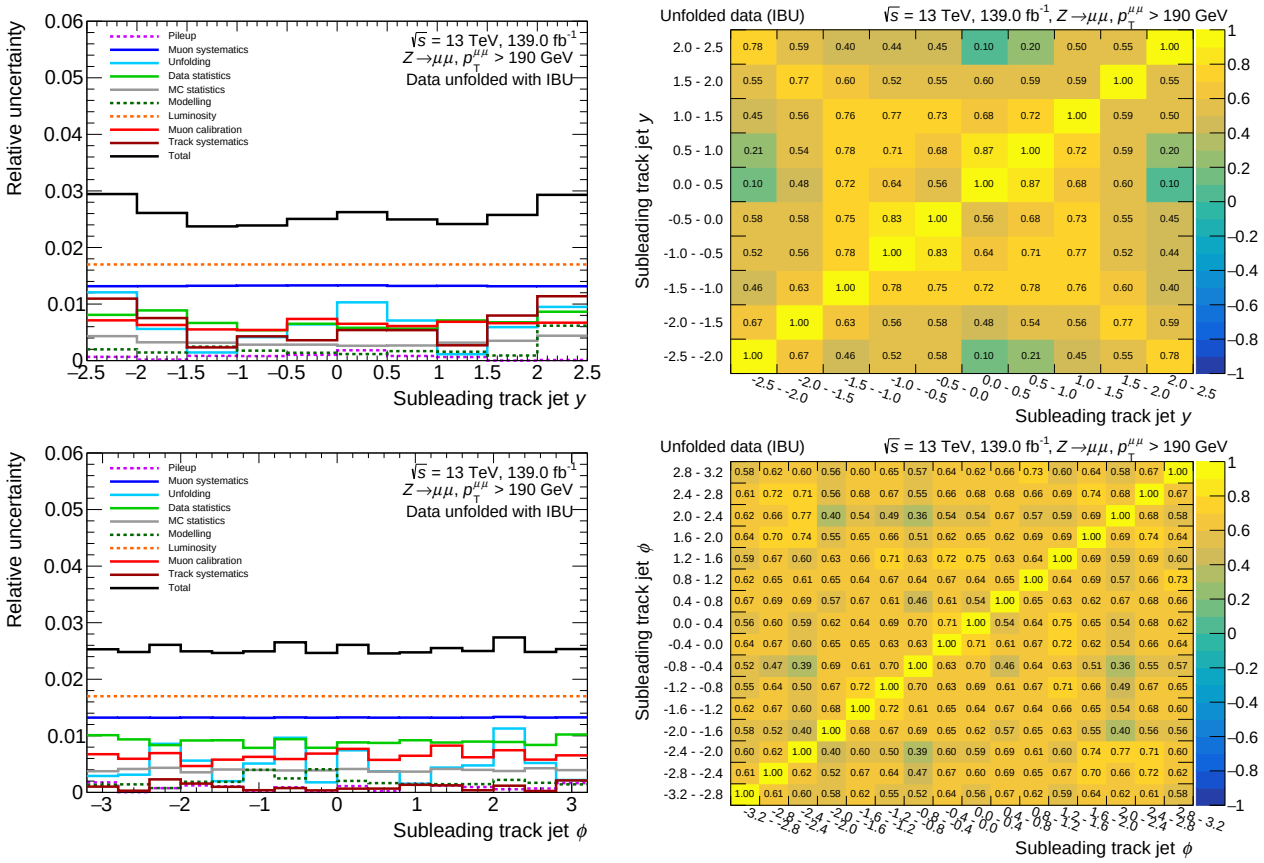


Figure C.19: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

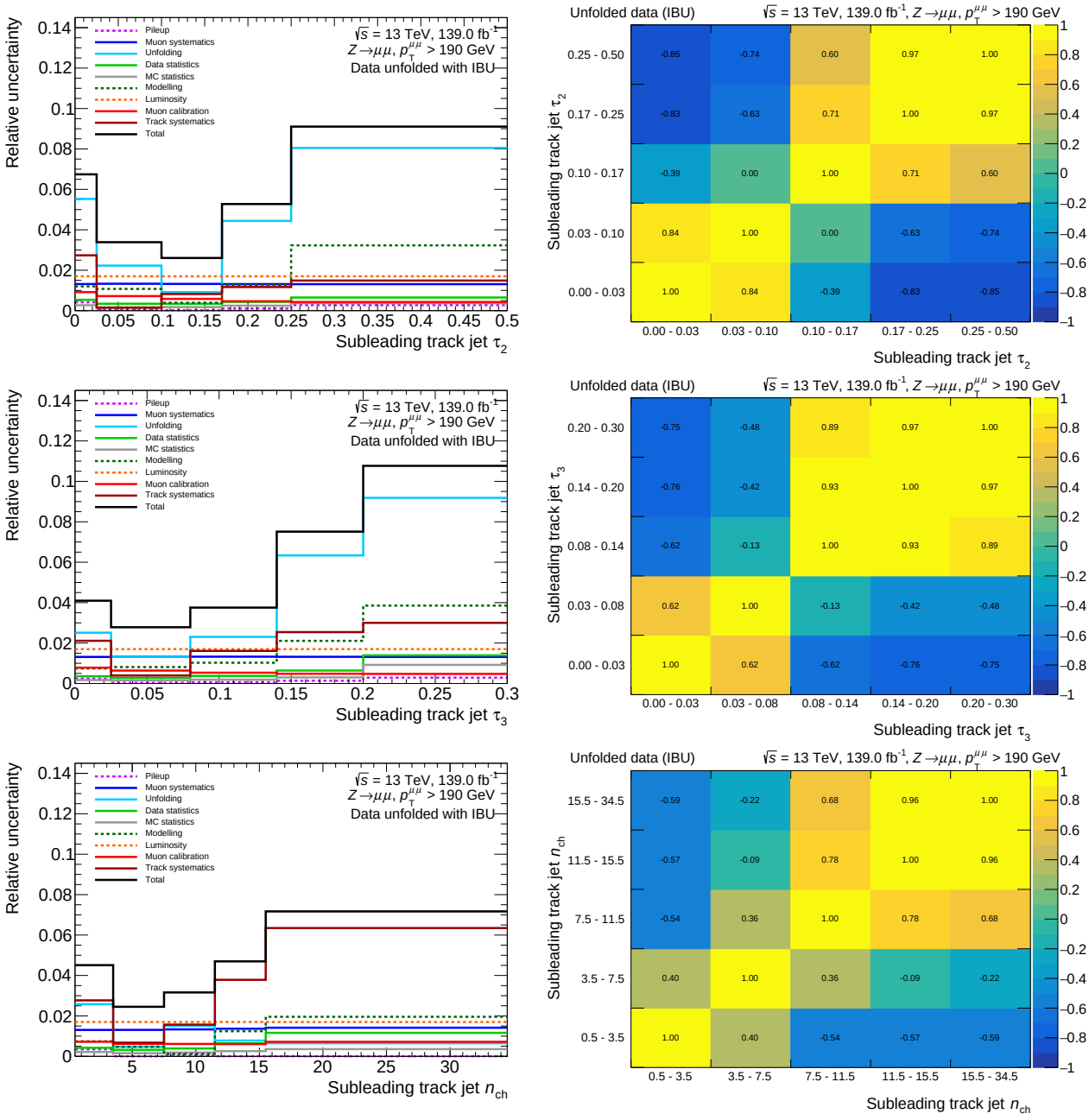


Figure C.20: The plots on the left show the relative uncertainty on the unfolded result obtained using IBU for the subleading track jet τ_2, τ_3, and the number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

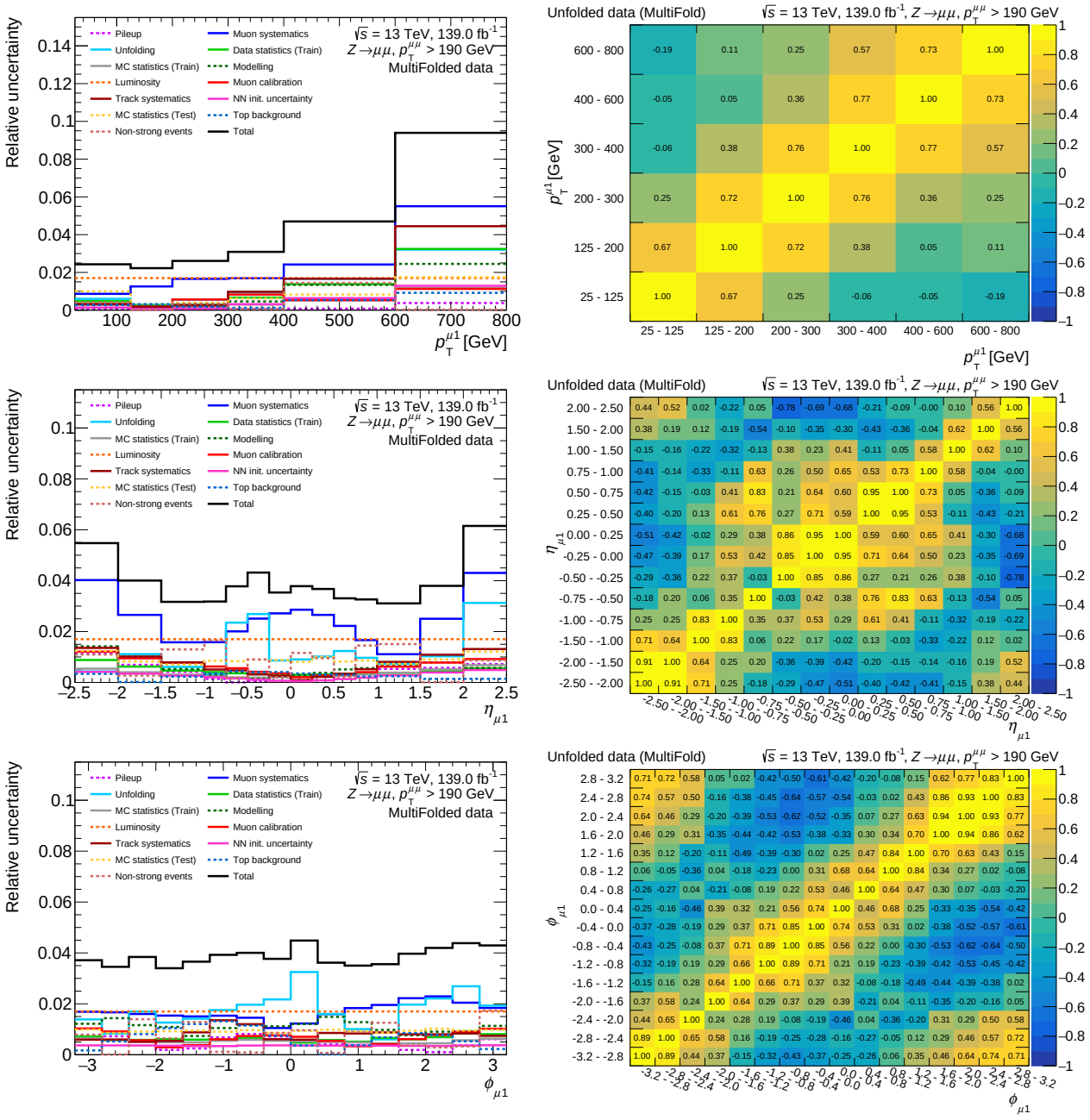


Figure C.21: The plots on the left show the relative uncertainty on the Multifolded result for the leading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

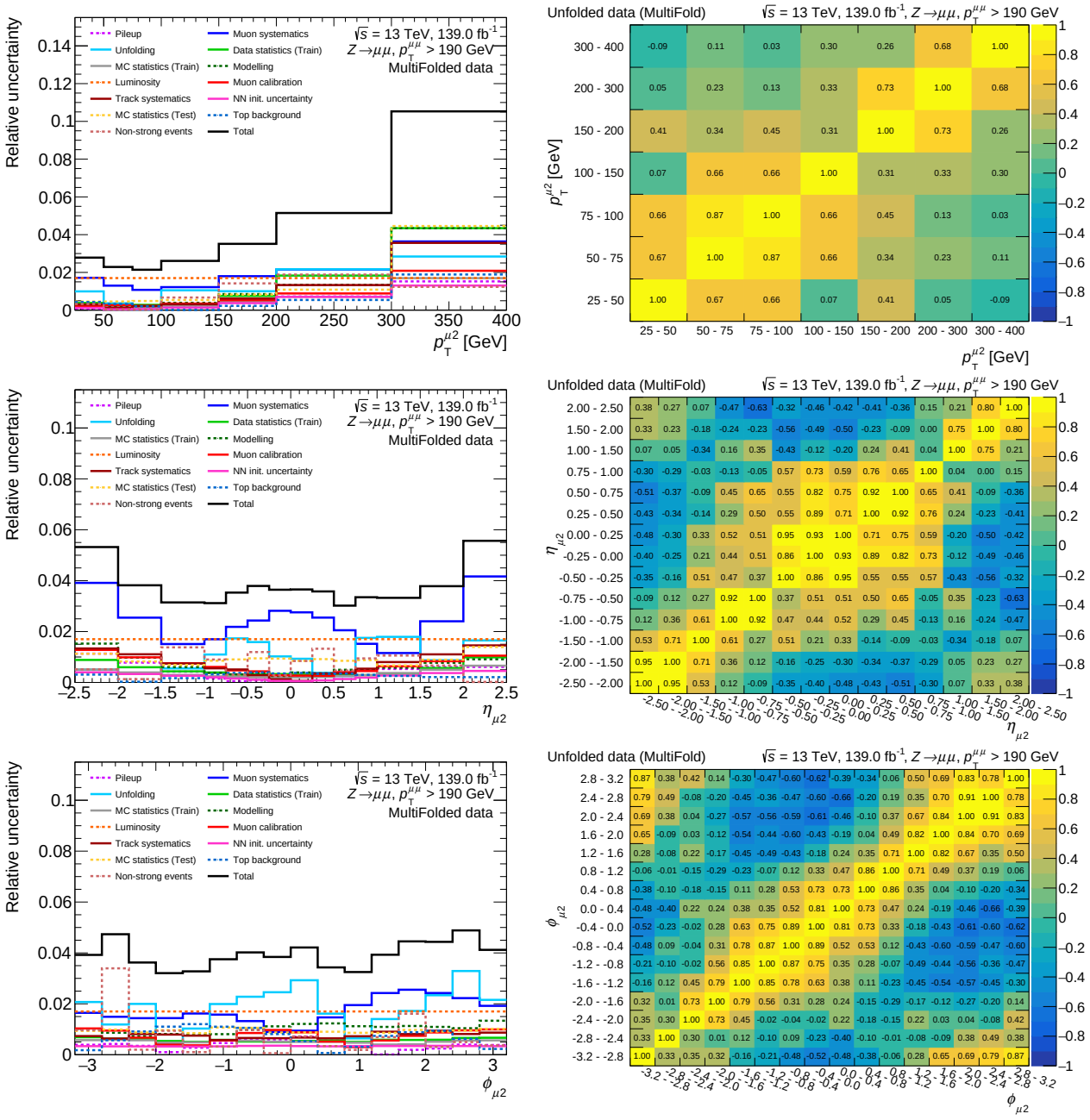


Figure C.22: The plots on the left show the relative uncertainty on the Multifolded result for the subleading muon p_{T}, η and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

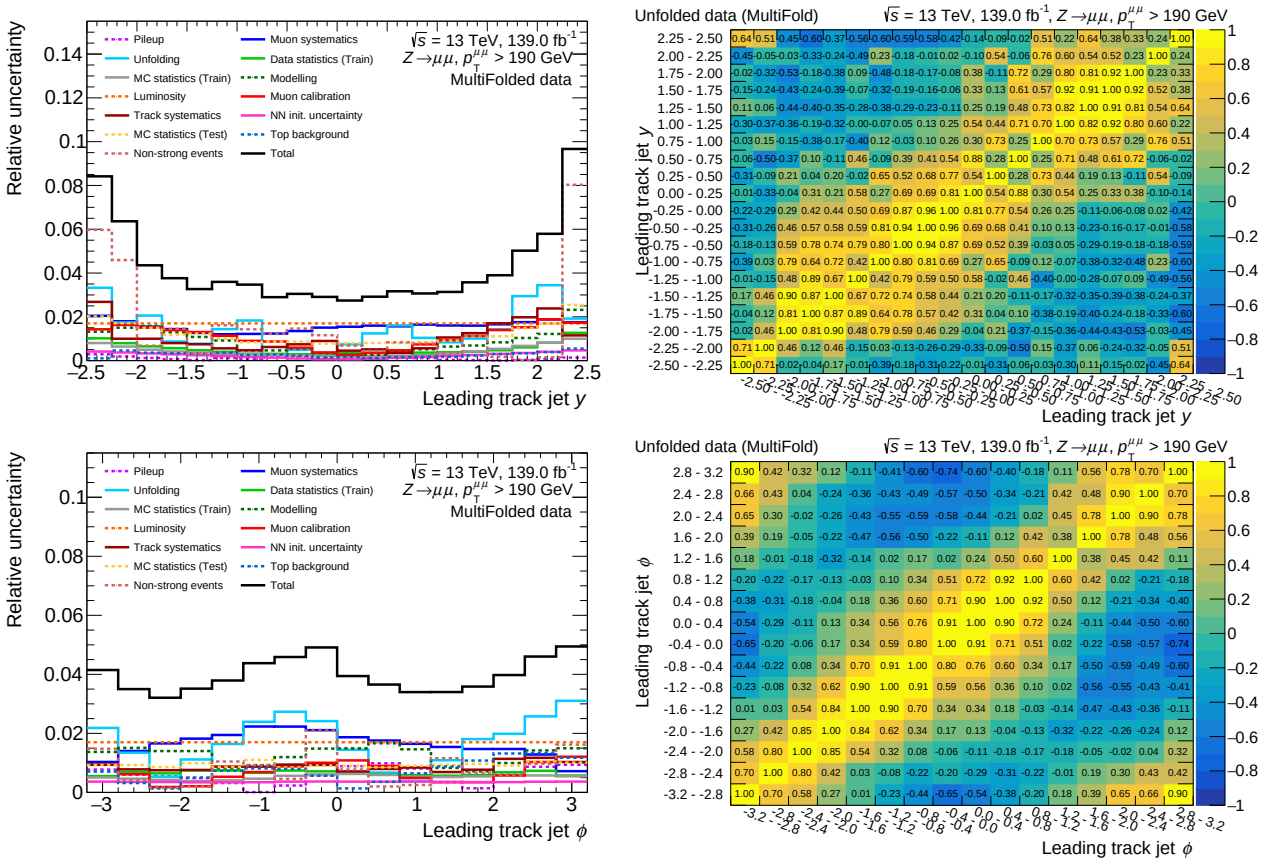


Figure C.23: The plots on the left show the relative uncertainty on the Multifolded result for the leading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

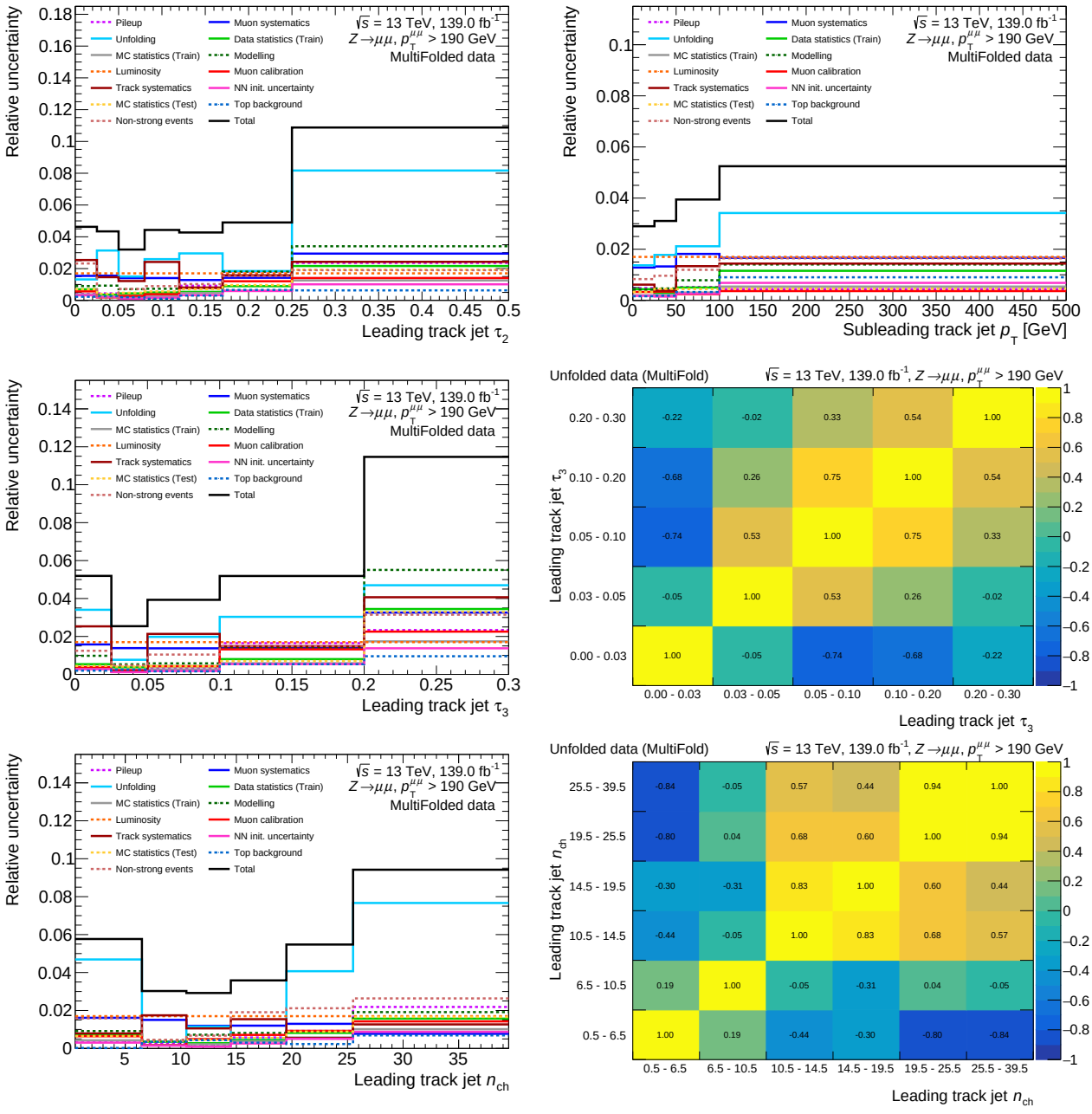


Figure C.24: The plots on the left show the relative uncertainty on the Multifolded result for the leading track jet τ_2, τ_3 and the number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

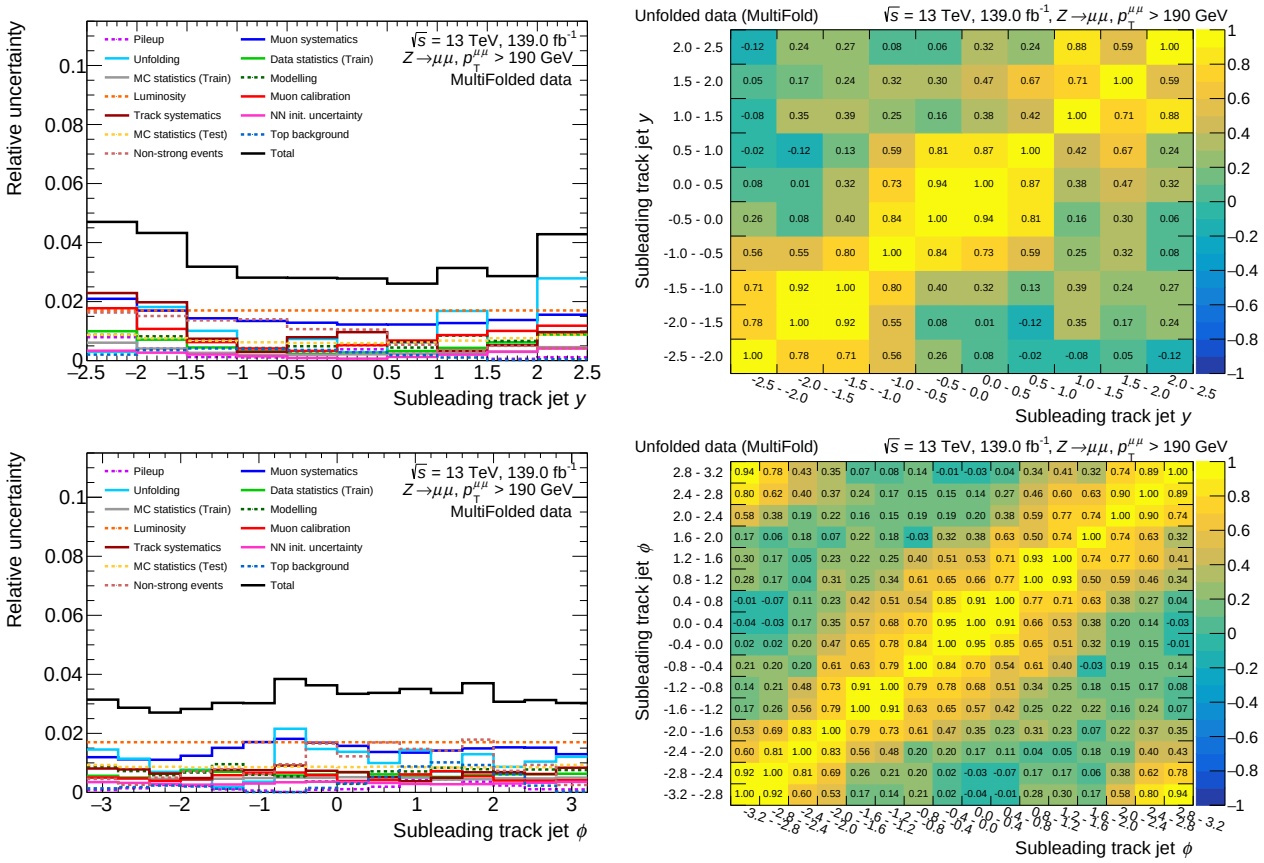


Figure C.25: The plots on the left show the relative uncertainty on the Multifolded result for the subleading track jet y and ϕ broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

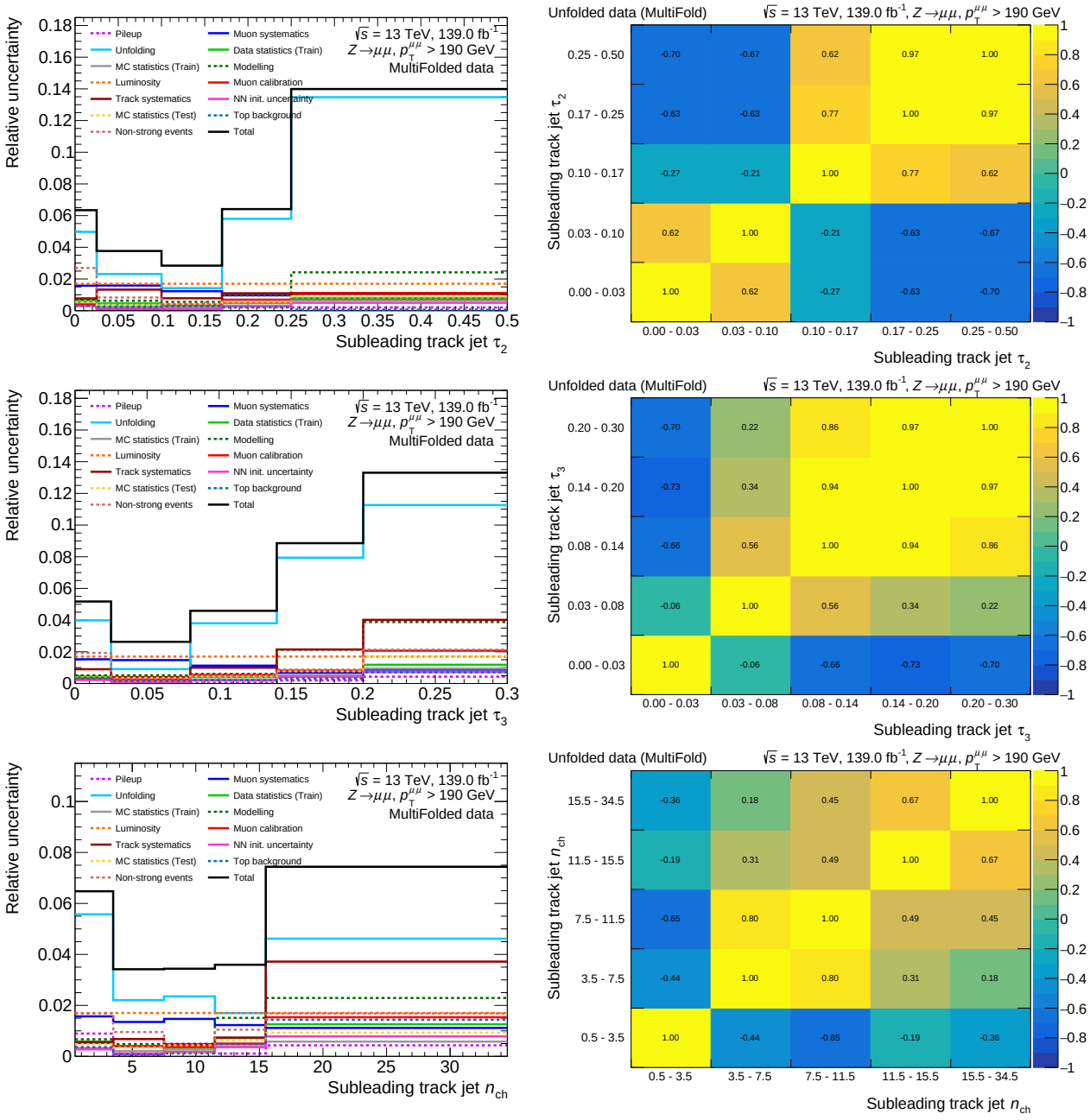


Figure C.26: The plots on the left show the relative uncertainty on the Multifolded result for the subleading track jet τ_2, τ_3 and the number of constituents broken down by source with the total uncertainty shown as the solid black line. The right shows the correlation matrices for the same observables.

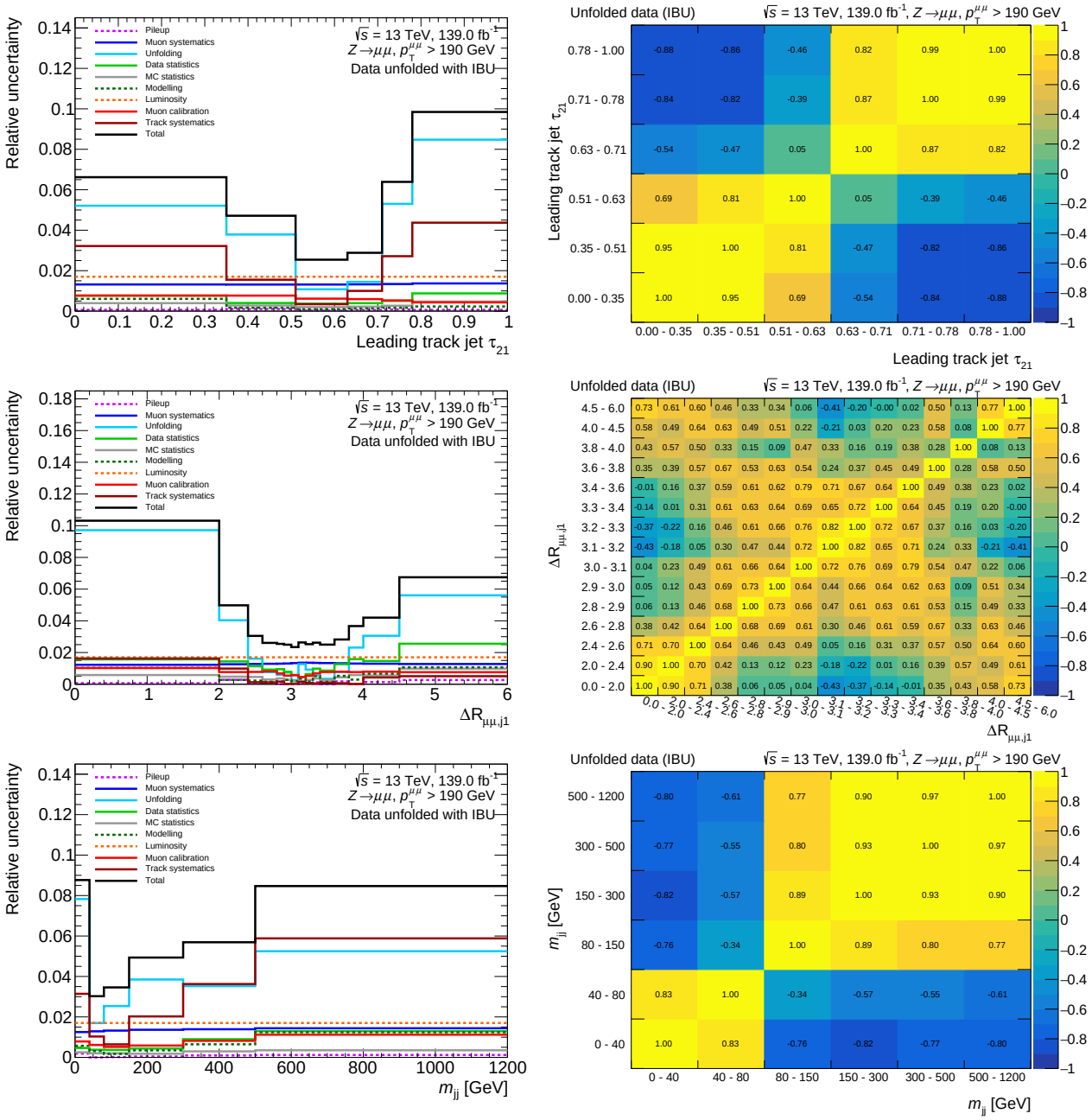


Figure C.27: The relative uncertainty on the unfolded data result obtained using IBU broken down by source (left) and the corresponding uncertainty correlation matrix (right) for the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and the track dijet mass.

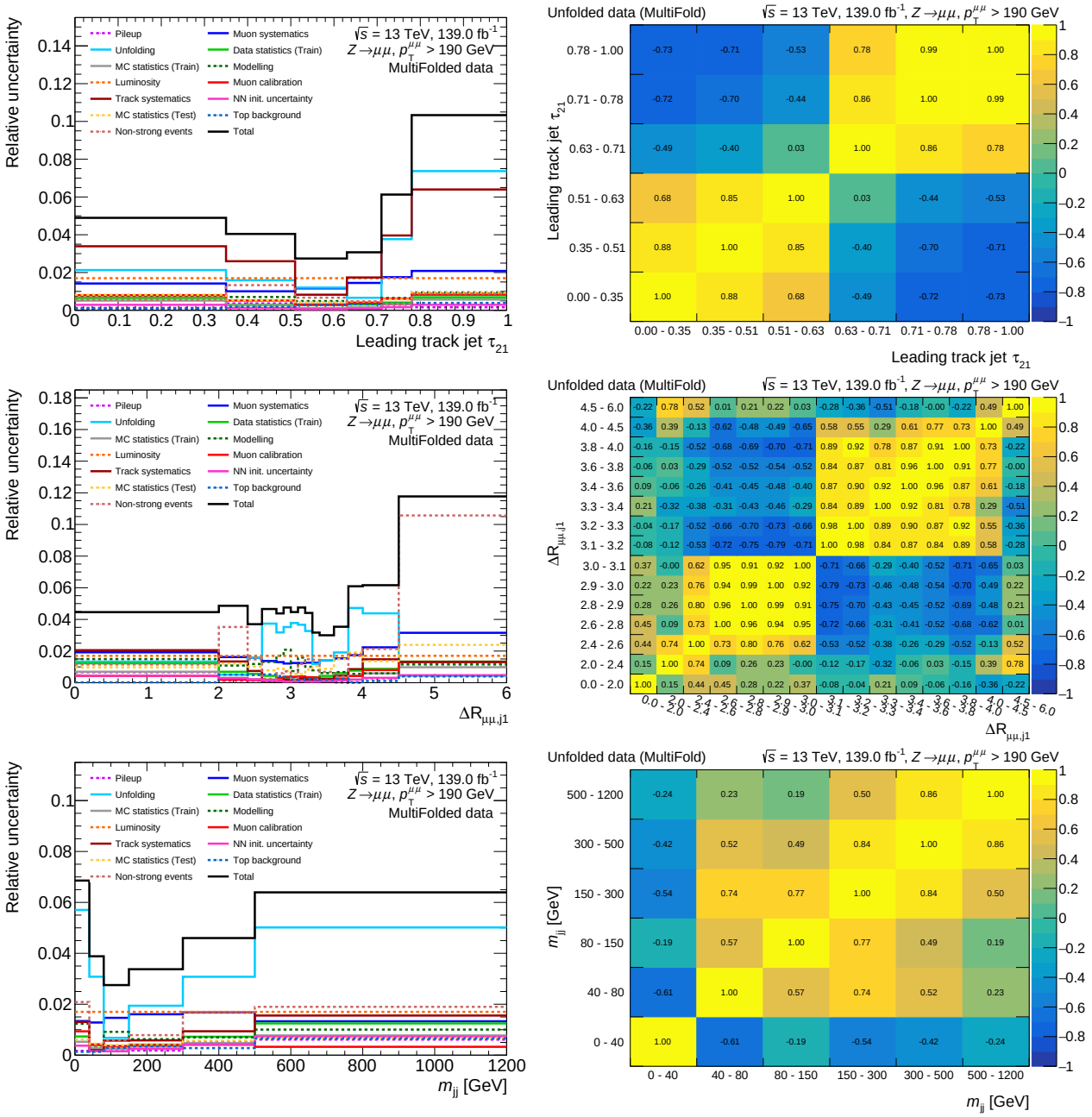


Figure C.28: The relative uncertainty determined for the unfolded data result obtained by calculating the observable after performing the unfolding using the MultiFold method (left) and the corresponding uncertainty correlation matrix (right) for the leading track jet τ_{21}, $\Delta R_{\mu\mu,j1}$ and the track dijet mass.

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1–29, [arXiv:1207.7214].

[2] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30–61, [arXiv:1207.7235].

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

[4] CMS Collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[5] UA2 Collaboration, *Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider*, *Phys. Lett. B* **129** (1983) 130–140.

[6] ATLAS Collaboration, *Searches for new phenomena in events with two leptons, jets, and missing transverse momentum in 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, arXiv:2204.13072.

[7] ATLAS Collaboration, *Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **798** (2019) 134949, [arXiv:1903.10052].

[8] ATLAS Collaboration, *Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton–proton collisions at ATLAS*, *Eur. Phys. J. C* **81** (2021) 163, [arXiv:2006.15458].

[9] ATLAS Collaboration, *Measurement of cross-sections for production of a Z boson in association with a flavor-inclusive or doubly b-tagged large-radius jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, arXiv:2204.12355.

[10] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, *OmniFold: A Method to Simultaneously Unfold All Observables*, *Phys. Rev. Lett.* **124** (2020) 182001, [arXiv:1911.09107].

[11] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn, and R. Winterhalder, *How to GAN away Detector Effects*, *SciPost Phys.* **8** (2020) 070, [arXiv:1912.00477].

[12] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn, A. Rousselot, R. Winterhalder, L. Ardizzone, and U. Köthe, *Invertible Networks or Partons to Detector and Back Again*, *SciPost Phys.* **9** (2020) 074, [arXiv:2006.06685].

[13] M. Backes, A. Butter, M. Dunford, and B. Malaescu, *An unfolding method based on conditional Invertible Neural Networks (cINN) using iterative training*, arXiv:2212.08674.

[14] C. P. Burgess and G. D. Moore, *The Standard Model: A Primer*. Cambridge University Press, 12, 2006.

[15] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions: Second Edition*. Princeton University Press, USA, 9, 2013.

[16] G. Altarelli, *Collider Physics within the Standard Model: A Primer*. Springer, 2017.

[17] M. Thomson, *Modern Particle Physics*. Cambridge University Press, New York, 2013.

[18] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[19] D. Griffiths, *Introduction to Elementary Particles*. Wiley, 2008.

[20] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Addison-Wesley, Reading, USA, 1995.

[21] S. P. Martin, *A Supersymmetry primer*, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1–98, [hep-ph/9709356].

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

[23] SNO Collaboration, *Measurement of the rate of* $\nu_e + d \to p + p + e^-$ *interactions produced by* ^8B *solar neutrinos at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **87** (2001) 071301, [nucl-ex/0106015].

[24] SNO Collaboration, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **89** (2002) 011301, [nucl-ex/0204008].

[25] Super-Kamiokande collaboration, *Solar B-8 and hep neutrino measurements from 1258 days of Super-Kamiokande data*, *Phys. Rev. Lett.* **86** (2001) 5651–5655, [hep-ex/0103032].

[26] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579–588.

[27] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[28] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968) 367–377.

[29] P. A. Zyla et al. (Particle Data Group), *Review of Particle Physics*, *PTEP* **2020** (2020), no. 8 083C01.

[30] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145–233, [arXiv:1101.2599].

[31] F. Gross et al., *50 Years of Quantum Chromodynamics*, arXiv:2212.11107.

[32] L. Lonnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046, [hep-ph/0112284].

[33] L. Lonnblad and S. Prestel, *Matching Tree-Level Matrix Elements with Interleaved Showers*, *JHEP* **03** (2012) 019, [arXiv:1109.4829].

[34] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061, [arXiv:1209.6215].

[35] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31–145.

[36] B. Andersson, *The Lund model*, vol. 7. Cambridge University Press, 7, 2005.

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

[38] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codebases* (2022) 8, [arXiv:2203.11601].

[39] NNPDF Collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849].

[40] ATLAS Collaboration, *Cross-section measurements for the production of a Z boson in association with high-transverse-momentum jets in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, arXiv:2205.02597.

[41] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014. URL: https://cds.cern.ch/record/1966419.

[42] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton-proton collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 1104, [arXiv:1910.04482].

[43] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ *=13 TeV*, *Eur. Phys. J. C* **76** (2016) 292, [arXiv:1603.05598].

[44] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *JINST* **14** (2019) P12006, [arXiv:1908.00005].

[45] ATLAS Collaboration, *Standard Model Summary Plots February 2022*, ATL-PHYS-PUB-2022-009, 2022. URL: http://cds.cern.ch/record/2804061.

[46] ATLAS Collaboration, "ATLAS Collaboration Map (February 2023)." URL: https://cds.cern.ch/record/2654108, 2023.

[47] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[48] A. Lopes and M. L. Perrey, "FAQ-LHC The guide." URL: https://cds.cern.ch/record/2809109, 2022.

[49] O. Aberle et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2020.

[50] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[51] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[52] M. Brice, "Aerial View of the CERN taken in 2008." URL: https://cds.cern.ch/record/1295244, 2008.

[53] E. Mobs, "The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019." URL: https://cds.cern.ch/record/2684277, 2019.

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

[55] J. Pequenao, "Computer generated image of the ATLAS inner detector." URL: https://cds.cern.ch/record/1095926, 2008.

[56] ATLAS Collaboration, "Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment." URL: https://cds.cern.ch/record/2723878, 2020.

[57] ATLAS Collaboration, *ATLAS pixel detector: Technical Design Report*. CERN, Geneva, 1998.

[58] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19, 2010.

[59] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, *JINST* **9** (2014) P08009.

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

[61] J. Pequenao, "Computer Generated image of the ATLAS calorimeter." URL: https://cds.cern.ch/record/1095927, 2008.

[62] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. CERN, Geneva, 1996.

[63] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. CERN, Geneva, 1996.

[64] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. CERN, Geneva, 1997.

[65] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem." URL: https://cds.cern.ch/record/1095929, 2008.

[66] A. M. Rodriguez Vera and J. Antunes Pequenao, "ATLAS Detector Magnet System." URL: https://cds.cern.ch/record/2770604, 2021.

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

[68] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018) P07017.

[69] ATLAS Collaboration, "Luminosity Public Results Run 2." URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2, 2019.

[70] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, arXiv:2212.09379.

[71] ATLAS Collaboration, *Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment*, *JHEP* **08** (2022) 089, [arXiv:2112.09588].

[72] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [arXiv:1405.0301].

[73] R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663, [arXiv:1706.00428].

[74] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820].

[75] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244–289, [arXiv:1207.1303].

[76] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152–155.

[77] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [arXiv:0811.4622].

[78] E. Bothmann et al., *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, [arXiv:1905.09127].

[79] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, [arXiv:0808.3674].

[80] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, [arXiv:1111.5206].

[81] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, [arXiv:1111.1220].

[82] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063, [hep-ph/0109231].

[83] S. Hoeche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, [arXiv:0903.1219].

[84] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, *JHEP* **04** (2013) 027, [arXiv:1207.5030].

[85] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, and J. Rojo, *An Unbiased Hessian Representation for Monte Carlo PDFs*, *Eur. Phys. J. C* **75** (2015) 369, [arXiv:1505.06736].

[86] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[87] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092].

[88] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [arXiv:1002.2581].

[89] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024, [arXiv:1007.2241].

[90] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2014) 145, [arXiv:1406.3660].

[91] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, *Eur. Phys. J. C* **45** (2006) 97–107, [hep-ph/0506026].

[92] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and Physics Documentation*, *Comput. Phys. Commun.* **199** (2016) 86–101, [arXiv:1011.0937].

[93] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038, [arXiv:0709.1027].

[94] T. Figy, V. Hankele, and D. Zeppenfeld, *Next-to-leading order QCD corrections to Higgs plus three jet production in vector-boson fusion*, *JHEP* **02** (2008) 076, [arXiv:0710.5621].

[95] B. Jager, S. Schneider, and G. Zanderighi, *Next-to-leading order QCD corrections to electroweak Zjj production in the POWHEG BOX*, *JHEP* **09** (2012) 083, [arXiv:1207.2626].

[96] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639–707, [arXiv:0803.0883].

[97] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196, [arXiv:1512.01178].

[98] J. Baglio et al., *VBFNLO: A Parton Level Monte Carlo for Processes with Electroweak Bosons – Manual for Version 2.7.0*, arXiv:1107.4038.

[99] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) 204, [arXiv:1412.3989].

[100] S. Frixione, P. Nason, and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, [arXiv:0707.3088].

[101] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016.

[102] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, *Single-top hadroproduction in association with a W boson*, *JHEP* **07** (2008) 029, [arXiv:0805.3067].

[103] S. Agostinelli et al., *Geant4—a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250–303.

[104] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823–874, [arXiv:1005.4568].

[105] Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, *J. Phys. Conf. Ser.* **513** (2014) 022024.

[106] 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, 2016. URL: https://cds.cern.ch/record/2206965.

[107] A. Buckley, *Computational challenges for MC event generation*, *J. Phys. Conf. Ser.* **1525** (2020) 012023.

[108] K. Danziger, S. Höche, and F. Siegert, *Reducing negative weights in Monte Carlo event generation with Sherpa*, arXiv:2110.15211.

[109] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490, [arXiv:1603.02934].

[110] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger, *The new ATLAS track reconstruction (NEWT)*, *J. Phys. Conf. Ser.* **119** (2008) 032014.

[111] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) 673, [arXiv:1704.07983].

[112] ATLAS Collaboration, *ATLAS Track Software Tutorial*, 2021. URL: https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoverview/.

[113] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444–450.

[114] J. Illingworth and J. Kittler, *A survey of the Hough transform*, *Computer Vision, Graphics, and Image Processing* **44** (1988) 87–116.

[115] S. Boutle et al., *Primary vertex reconstruction at the ATLAS experiment*, *J. Phys. Conf. Ser.* **898** (2017) 042056.

[116] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **81** (2021) 578, [arXiv:2012.00578].

[117] S. Catani, Y. Dokshitzer, and B. Webber, *The k_t-clustering algorithm for jets in deep inelastic scattering and hadron collisions*, *Phys. Lett. B* **285** (1992) 291–299.

[118] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[119] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, pp. 270–279, 1998. hep-ph/9907280.

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

[121] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, vol. 958 of *Lecture Notes in Physics*. Springer, 2019.

[122] R. Atkin, *Review of jet reconstruction algorithms*, *J. Phys. Conf. Ser.* **645** (2015) 012008.

[123] ATLAS Collaboration, *Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton-Proton Collisions with the ATLAS Detector*, *Phys. Rev. Lett.* **124** (2020) 222002, [arXiv:2004.03540].

[124] ATLAS Collaboration, *Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at* $\sqrt{s}$ *=13 TeV*, *Phys. Rev. D* **101** (2020) 052007, [arXiv:1912.09837].

[125] G. Cowan, *Statistical data analysis*. Clarendon Press, Oxford, 1998.

[126] A. Armbruster, K. Kroeninger, B. Malaescu, and F. Spano, *Practical considerations for unfolding*, ATL-COM-PHYS-2014-277, CERN, Geneva, 2014.

[127] S. Biondi, *Experience with using unfolding procedures in ATLAS*, *EPJ Web Conf.* **137** (2017) 11002.

[128] A. Hocker and V. Kartvelishvili, *SVD approach to data unfolding*, *Nucl. Instrum. Meth. A* **372** (1996) 469–481, [hep-ph/9509307].

[129] A. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*. Wiley, 1977.

[130] A. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and Y. A. G., *Numerical methods for the solution of ill-posed problems*. Springer, 1995.

[131] B. Malaescu, *An Iterative, dynamically stabilized method of data unfolding*, arXiv:0907.3791.

[132] B. Malaescu, *An Iterative, Dynamically Stabilized (IDS) Method of Data Unfolding*, in *PHYSTAT 2011*, pp. 271–275, 2011. arXiv:1106.3107.

[133] G. Choudalakis, *Fully Bayesian Unfolding*, arXiv:1201.4612.

[134] G. D'Agostini, *A multidimensional unfolding method based on bayes' theorem*, *Nucl. Instrum. Meth. A* **362** (1995) 487–498.

[135] G. D'Agostini, *Improved iterative Bayesian unfolding*, arXiv:1010.0632.

[136] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[137] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2022.

[138] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023.

[139] M. Feickert and B. Nachman, *A Living Review of Machine Learning for Particle Physics*, arXiv:2102.02770.

[140] HEP ML Community, "A Living Review of Machine Learning for Particle Physics." URL: https://iml-wg.github.io/HEPML-LivingReview/.

[141] M. Sugiyama, T. Suzuki, and T. Kanamori, *Density Ratio Estimation in Machine Learning*. Cambridge University Press, 2012.

[142] A. Andreassen and B. Nachman, *Neural Networks for Full Phase-space Reweighting and Parameter Tuning*, *Phys. Rev. D* **101** (2020) 091901, [arXiv:1907.08209].

[143] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, arXiv:1506.02169.

[144] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[145] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [arXiv:1011.2268].

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

[147] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, *JINST* **15** (2020) P09015, [arXiv:2004.13447].

[148] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, [arXiv:1111.6097].

[149] W. Buttinger, *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*, ATL-COM-SOFT-2015-119, 2015. URL: https://cds.cern.ch/record/2014726.

[150] ATLAS Collaboration, *Evaluating statistical uncertainties and correlations using the bootstrap method*, ATL-PHYS-PUB-2021-011, 2021. URL: https://cds.cern.ch/record/2759945.

[151] D. W. Scott, *On optimal and data-based histograms*, *Biometrika* **66** (12, 1979) 605–610.

[152] B. Nachman and J. Thaler, *Neural resampler for Monte Carlo reweighting with preserved uncertainties*, *Phys. Rev. D* **102** (2020) 076004, [arXiv:2007.11586].

[153] L. Brenner, R. Balasubramanian, C. Burgard, W. Verkerke, G. Cowan, P. Verschuuren, and V. Croft, *Comparison of unfolding methods using RooFitUnfold*, *Int. J. Mod. Phys. A* **35** (2020) 2050145, [arXiv:1910.14654].

[154] F. Chollet et al., "Keras." URL: https://keras.io, 2015.

[155] M. Abadi et al., *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*, arXiv:1603.04467.

[156] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv:1412.6980.