TECHNISCHE UNIVERSITÄT DRESDEN

Unfolding algorithms and their application in the first diphoton measurement at 13 TeV

Master-Arbeit
zur Erlangung des Hochschulgrades
Master of Science
im Master-Studiengang Physik

vorgelegt von

Ken Kreul
geboren am 07.02.1995 in Oelsnitz/Vogtl.

Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften
Technische Universität Dresden

Eingereicht am 26. November 2018

1. Gutachter: Dr. Frank Siegert
2. Gutachter: Prof. Dr. Michael Kobel

Zusammenfassung
Deutsch:

Eines der Hauptziele des Large Hadron Collider (LHC) war die Entdeckung des Higgs Bosons. Mehrere Experimente, wie zum Beispiel ATLAS oder CMS, wurde am LHC gebaut. Diese fanden das Higgs Boson 2012 [1] als der LHC bei einer Schwerpunktsenergie von $\sqrt{s} = 8$ TeV betrieben wurde. Einer der Zerfallskanäle, in dem die Masse des Higgs bestimmt wurde, ist der Zerfall des Higgs Boson in zwei Photonen. In diesem Kanal sind alle anderen Prozesse, die zwei Photonen (sogenannter "Diphoton" Prozess) produzieren können, der Untergrund. Deshalb war die Messung der Eigenschaften des Higgs Bosons nur durch das genaue Vermessen des Diphoton Prozesses möglich. Für den ersten Teil des Run 2 (2015-2016) wird die Analyse wiederholt, um das Verständnis des Prozesses zu erhöhen. Dies ist für weitere Suche und Messung von Teilchen entscheidend.

Die Analyse von ATLAS Daten wird in mehreren Schritten durchgeführt. Zum Beispiel müssen alle Prozesse, die nicht von Interesse sind, herausgefiltert werden. Der letzte Schritt der meisten Analysen ist das sogenannte "Unfolding". Hier werden Effekte vom Detektor, wie zum Beispiel die endliche Auflösung, in der Messung korrigiert. Dies ist unter anderem wichtig, um die Ergebnisse von anderen Detektoren und Experimenten zu vergleichen. In dieser Arbeit werden einige Unfolding Methoden verglichen und die Anwendung dieser in der ersten 13 TeV Diphoton ATLAS Analyse betrachtet.

Abstract
English:

One of the main goals of the Large Hadron Collider (LHC) was the measurements of the Higgs boson. Several experiments like ATLAS and CMS were set up and they found the Higgs boson 2012 [1] when the LHC was running at the center of mass energy $\sqrt{s} = 8$ TeV. One of the decay channel, where the mass of the Higgs boson is measured, is the Higgs to two photons. In this decay channel all the other Diphoton production processes are the background. Therefore the measurement of the Higgs boson was only possible because the background of the Diphoton process was measured and well known. For the data taken in a part of Run 2 (2015-2016) the analysis is repeated to improve the knowledge of the process for further particle searches and measurements.

Analysis of ATLAS data are done in several steps. For example, all the processes that are not of interest in the analysis have to be eliminated in the background subtraction. The last step of the analysis is usually the so called unfolding. Here the effects of the detector, like for example the finite resolution, have to be corrected to compare data of different experiments

and to make a theory comparison easy. In this work, several unfolding methods are discussed and the unfolding of the first 13 TeV Diphoton ATLAS analysis is presented.

Contents

1 Introduction

Particle physicists in the 20^{th} century built and validated the Standard Model of particle physics, which describes the interactions of the fundamental components of matter. The theory had success in predicting most of the experiments, but one fact caused trouble. Large Electron Positron collider (LEP) and other experiments detected a mass for the weak interaction bosons. The Standard Model had no theoretical foundation for this mass and in fact mass terms for the bosons caused several conceptual issues . One idea to solve this contradiction was invented by P.W. Higgs [2] and others [3]. In this theory, the interaction of the particles with a new particle, later called the Higgs boson, is the reason for their mass. In order to find the Higgs boson, particle interactions at higher energy than achieved in the experiments of the 20^{th} century were needed. This was one of the main reasons for the construction of the LHC.

In the LHC protons are brought to a collision at a center of mass energy of 13 TeV. To investigate the interactions, several experiments are located at the points of the collisions. In this thesis data from the ATLAS experiment is used. It is a multipurpose detector that is measuring and distinguishing several different particles to provide a picture of the particle in the final state. This cannot be done perfectly because of for example the finite resolution of the detector or misidentification of particles. All these imperfections of the detector change the measured/reconstructed values (reco) compared to the actual properties of the particle (truth). For theorist it is more convenient to compare their physical calculations on truth level i.e. without detector effects with the data of the experiment. Otherwise a CPU-expensive detector simulation must be used on the theory prediction to compare it to uncorrected data. Hence, the measured data has to be corrected and this step is called unfolding. In other words the data has to be transformed from reco level to truth level.

The unfolding is usually the last step of an analysis. Therefore all the uncertainties estimated in earlier steps, like for example the background subtraction, have to be propagated through the unfolding procedure. Since the unfolding is using Monte-Carlo simulations (MC), uncertainties of the theory predictions and the detector simulation have to be estimated.

In this work, the unfolding step of the first 13 TeV ATLAS diphoton analysis is presented. Since the Higgs can also decay to two photons, the diphoton process is the main background in this decay channel. Also most of the theories beyond the Standard Model (BSM) predict particles that can be visible in the diphoton final states. Therefore precise measurements are needed to improve the background subtraction in Higgs and BSM measurements. Furthermore

because the electrodynamic theory is well known, conclusions for example on the structure of the protons can be made. This can improve the understanding of the strong interaction.
In this work, in section 2 the experimental setup is presented and in section 3 several approaches to the problem of unfolding are introduced and discussed. In section 4 the application of these methods in the diphoton analysis is presented, before concluding and giving an outlook on future developments.

2 Basic Foundation

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [4] is a particle accelerator made for high energy particle interactions. It was constructed from 2000 to 2008 by the European Organization for Nuclear Research (CERN) at the France-Switzerland border near Geneva. There the tunnel of the Large Electron-Positron Collider (LEP) built between 1984 and 1989 was used for the LHC to reduce the costs of the project. This tunnel is between 45 and 170 m under the surface.
In the LHC two rings with counter-rotating beams with protons or heavy ions are accelerated by a complex system of electromagnetic fields. The beams consist of up to 2808 bunches each with about $2 \cdot 10^{11}$ particles. Since these beams are both positively charged particle beams, the magnetic flux has to circulate in the opposite sense in the two beam pipes shown in figure 2.1. This is achieved by around 1230 dipole magnets and some 400 quadrupole magnets. The magnets are put into a superconducting state at a temperature of 2 K and they produce a magnetic field of up to 8 T.

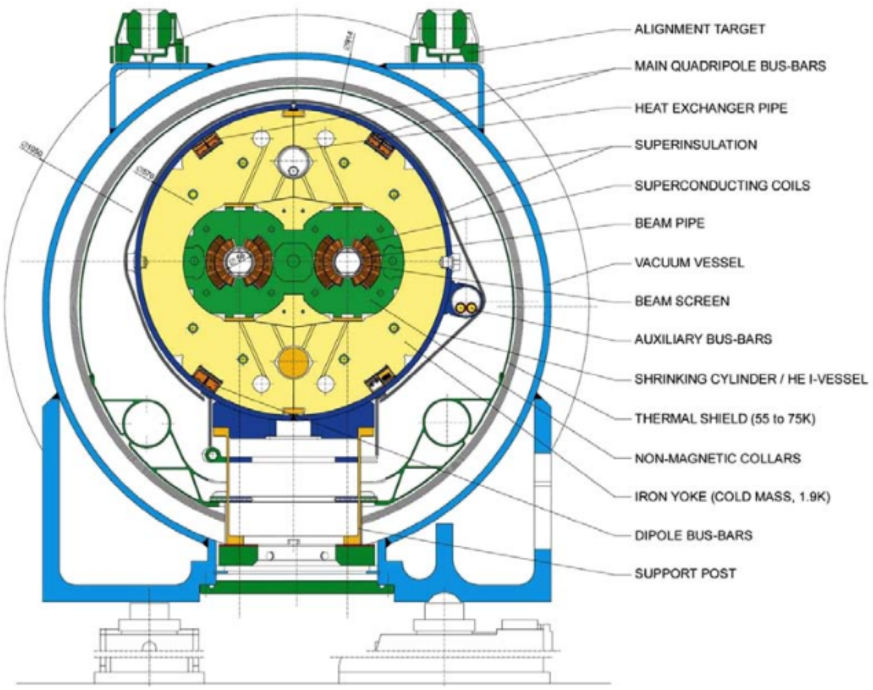


Figure 2.1: Beampipes of the LHC [4]

There are four crossing point of the two beams with detectors in the LHC ring. The LHCb detector [5] and ALICE experiment [6] focus on specific physical processes whereas the CMS detector [7] and ATLAS detector [8] are multipurpose detectors.
The LHC was designed for a center of mass energy of $\sqrt{s} = 14$ TeV but it started in 2010 with Run 1 at $\sqrt{s} = 8$ TeV. During this Run 1 the Higgs boson of the Standard Model was found from CMS and ATLAS [1]. In February 2013 the LHC was shut down to make adjustments for the Run 2 at $\sqrt{s} = 13$ TeV. Run 2 started in August 2015 and will end in 2018.
After the shutdown in 2019 the LHC will run at the designated energy of $\sqrt{s} = 14$ TeV.

2.2 The ATLAS Experiment

2.2.1 The Detector

As you can see in figure 2.4, the ATLAS detector is a 44 m wide and 25 m high cylinder. It is built at one of the four crossing points of the LHC. Protons enter the detector from both sides and collide in the middle of the detector. Most of the particles that are produced in these collisions decay before they reach the detector. Particles that are stable enough to fly (atleast) few cm to the detector systems are the protons p, neutrons n, electrons e^-, muons μ^-, photons γ, neutrinos ν and pions. The detector has to be able to distinguish these particles and measure their energy and momentum.
The closest part of the detector to the point of interaction is the tracking system. This part measures the trajectory of charged particles. Since the tracker is in a strong magnetic field, the charge and the momentum of the particle is measurable through the curvature of the track. The second part of the detector system is the electromagnetic calorimeter. Here light charged particles and photons are stopped and the energy is measured. For the photons the tracking system and the electromagnetic calorimeter are the main systems to measure their properties. Therefore these two elements of the detector are further explained in the sections 2.2.1.1 and 2.2.1.2. Heavier charged particle like the proton and pions are not stopped in the electromagnetic calorimeter. These particles are measured in the hadronic calorimeter. Stable particles that can pass all these layers are muons and neutrinos. Neutrinos only interact by the weak force. Therefore they cannot be detected and they are only reconstructed by the missing energy in the energy and momentum conservation. Muons are too heavy to be stopped in the electromagnetic calorimeter and they have no strong charge to produce a shower in the hadronic calorimeter. Therefore dedicated muon chambers are used as the last layer of the detector to measure the energy of the muons.
In the ATLAS detector the cartesian coordinate system has the origin at the point of interaction of the particles. The x-axis points to the middle of the LHC ring, the y-axis upwards and the z-axis is in beam direction. To express the direction of the particles angular coordinates are commonly used. In ATLAS, the polar angle in the x-y plane is ϕ and the azimuthal angle is θ.

This azimuthal angle is often expressed as pseudorapidity $\eta = -\ln(\tan(\frac{\theta}{2}))$. η of 0 corresponds to perpendicular to the beam and $\eta = \infty$ is in beam direction.

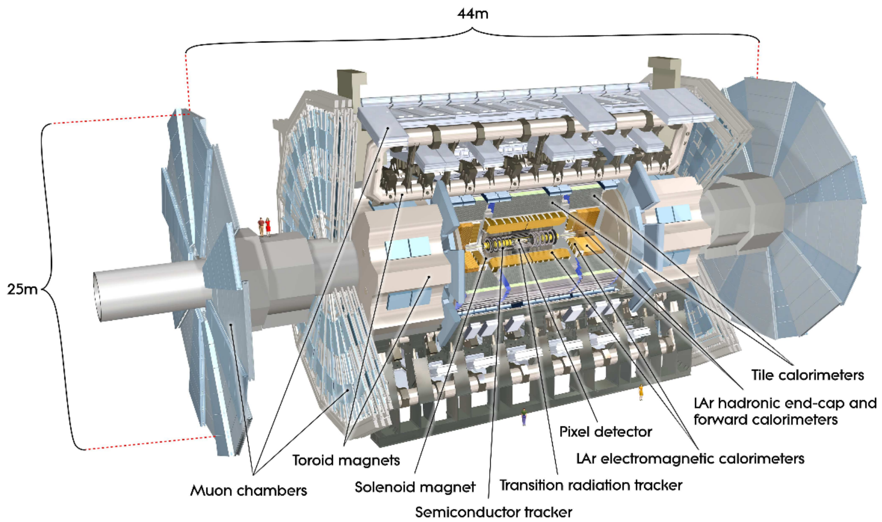


Figure 2.2: The ATLAS detector

2.2.1.1 Tracking System

During Run 2 ~ 40 interactions are expected. These interactions result in approximately 1000 particles in the tracking system every 25 ns [8]. Therefore every component has to be able to withstand the immense particle flux.

The inner detector shown in figure 2.3 is designed to track charged particles with high resolution. At innermost layer the high resolution is achieved by discrete space-points from silicon pixel layers [8]. These pixels are 250 μm thick n-type wafers with readout. There are 1744 pixel sensors with 47232 pixels with a size of $50 \cdot 400\ \mu\text{m}^2$ in the ATLAS detector. The complete Pixel detector has to be cooled to withstand the amount of incoming particles. The next detector system consist of pairs of silicon microstrip (SCT) layers. These are 15912 strips of 285μm thick silicon detectors operating at voltages of 250 to 350 V. The last part of the Tracking system is the transition radiation tracker (TRT). There are Polyimide drift (straw) tubes of 4 mm diameter filled with a mixture of 70% Xe, 27% CO_2 and 3% O_2. They measure the ionization of the gas from the charged particles. With an average of 36 hits per track in this system, the trajectory can be reconstructed. A charged particle produces around 36 hits which provides continuous tracking, good momentum resolution and high efficiency.

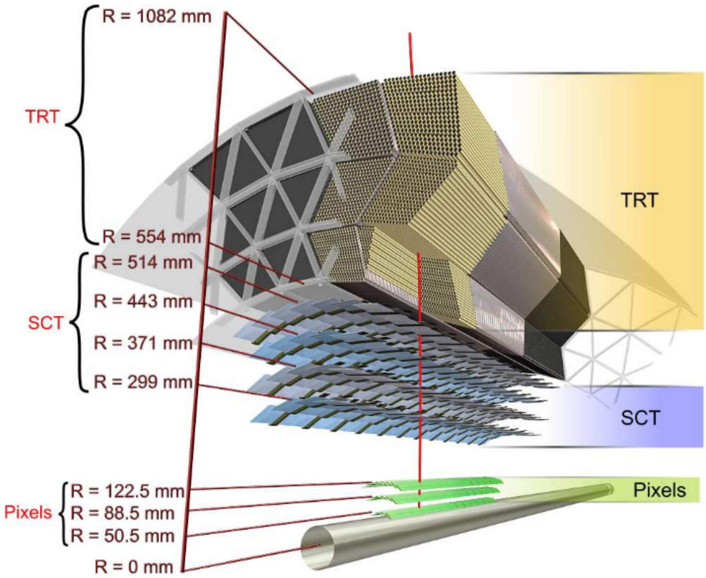


Figure 2.3: The inner detector. [8]

2.2.1.2 Electromagnetic Calorimeter

The electromagnetic calorimeter consist of two parts. The end-cap calorimeter at $1.375 < |\eta| < 3.2$ and the barrel calorimeter $|\eta| < 1.475$. The overlap of these parts are usually excluded in the analysis. The calorimeter consists of active material and dense passive material. The passive material is usually Lead bent in accordion shape to provide full ϕ symmetry [9]. The space between the passive absorber material is filled with the active material which is liquid Argon (LAr) and the readout electrodes, which consist of three conductive Copper layers separated by insulating polyimide sheets [8]. The two outer Copper layers are at the high voltage potential and the inner layer is used to read out the signal.
The incoming electron or photons interact with the Lead and produce secondary charged particles. These particles ionize the argon and produce electrons which induce the signal in the read out electrodes. The secondary particles from the first reaction have to much energy to be measured by the first layer of readout electrodes. Therefore these electrons reach the next absorber layer and produce secondary particles themselves. This multiplication of a single incoming electron or photon into several secondary particles is called shower.
The calorimeter is split up into fine cells with independent readout. These cells are used to locate the incoming particles.

2.2.2 Data Selection and Trigger

The ATLAS detector is measuring proton bunch collisions every 25 ns (40 MHz). The interactions produce several particles which results in too much data for any storage available.

Therefore the interesting events have to be identified. This is done by the Trigger system. The level 1 Trigger is filtering the events very fast with hardware on the detector. It is using only very limited and fast accessible information of the particle measurement. The level 1 Trigger also defines *Region-of-Interest* (RoI) where interesting features are measured. The level 2 Trigger is using all the information the detector is providing in these RoI's to further reduce the amount of data. In the end the event filter is processing all the data to make the decision which event is kept. After these selections only one in 40.000 events is stored and can be used for further analysis.

2.2.3 Photon Reconstruction

Photons produce a shower in the electromagnetic calorimeter. This shower is depositing energy in multiple cells of the calorimeter. The energy of around 3·5 cells are added for the total photon energy. The same is done for electrons because they produce similar shower. Therefore these two particle have to be distinguished with the help of the tracking system. Electrons produce a track because of their electromagnetic charge. Photons themselves cannot produce tracks but they can interact with the material of the detector to produce a electron positron pair. These photons are called *converted* photons and photons that do not produce electron positron pairs before the electromagnetic calorimeter are called *unconverted* photons. Therefore unconverted photons are identified as energy deposits in the electromagnetic calorimeter without a track in the tracking system. The detector is identifying converted photons as two deposits of energy in the electromagnetic calorimeter with the constraint that their tracks have to be oppositely charged and have to originate from one area in the tracking system, the so called conversion vertex. Electrons are the particles that produce a track in tracking system and deposit their energy in the calorimeter.

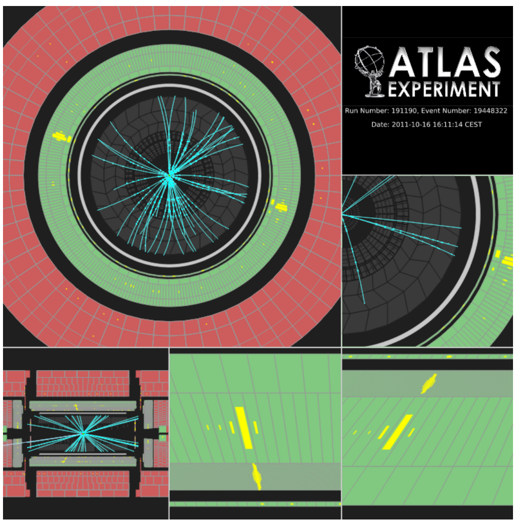


Figure 2.4: Event display of the ATLAS detector. In the top left illustration, a unconverted photon with no track is hitting the green electromagnetic calorimeter on the left and on the right a converted photon is measured. [10]

This procedure is not perfect. If for example a track of an electron is not measured, the electron can be reconstructed as a photon. These things have to be considered in the analysis.

2.3 Physics in the ATLAS Experiment

In the beginning of the 20th century several experiments for example Rutherford scattering [11] and the beta decay provided hints for the interaction and structure at very small scales. From these results the first theories for quantum electrodynamics (QED), the weak interaction and the strong interaction (QCD) were built independently. Theoreticians had many ideas to improve these parts and merge them into one model of particle interactions. One main idea that led to the now widely accepted Standard Model of particle physics was the concept of gauge symmetry. For every fundamental interaction (except for gravity) a gauge symmetry group was found and the interactions were built upon this insight. For further explanation on the history of the Standard Model refer to [12].

The Standard Model consist of the three fundamental interactions which are the electromagnetic, weak and strong interaction. Each interaction has one or more bosons. The weak interaction is mediated by the weak gauge bosons W^+, W^- and the Z. As the name suggests, this interaction is by several orders of magnitude weaker than the other interactions but it is the reason for e.g. the β-decay. In the order of strength, the next fundamental force is the electromagnetic interaction. The boson for the electromagnetic interaction is the photon (γ). The strong interaction the reason for the stability of protons and neutrons which build the matter we know. The bosons for this interaction are the gluons g.

The Standard Model consists of the leptons electron e^-, muon μ^- and tau τ^- with the associated neutrino which are the electron-neutrino ν_e, muon-neutrino ν_μ and the tau-neutrino ν_τ. Neutrinos only interact with the weak gauge bosons. The other leptons can, in addition to this weak force, interact with the electromagnetic force. The last group of particles are the quarks q. The lightest quarks are the up and down quarks. They build the proton and neutron which are the components of the stable matter. The two heavier brothers of the up quark are the charm and top quark. The strange and bottom quarks are the brothers of the down quark with more mass. These particles are summarized in figure 2.5. The main problem of this model was the mass of the W and Z bosons which was an experimental fact but mass terms were causing problems in the theory. The idea to solve this problem was spontaneous broken symmetry. This theory results in a particle that was observed in 2012 [1].

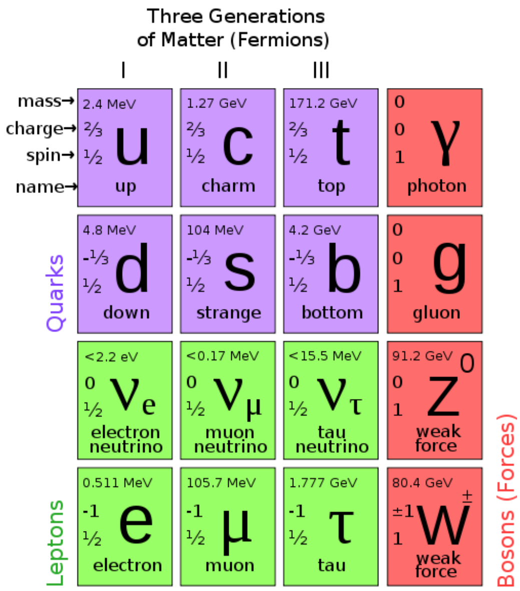


Figure 2.5: The Standard Model of particle physics [13]

The Standard Model predictions are accurate for most experiments but some phenomena (for example dark matter) in nature hint to new physics. Several ideas of theories try to improve the Standard Model. To search for evidence of different ideas, the Standard Model processes have to be measured with high accuracy because they serve as benchmarks and background for these theories.

The here presented analysis focuses on the diphoton process. Final states, that contain two photons, are investigated to decrease uncertainties and provide a comparison for theory simulations. Because the Higgs boson can decay into two photons, the knowledge of this process as a background was important for the discovery of the Higgs in this decay channel. Hence discoveries of new physic can only be done when the already known interactions are measured with high accuracy.

The leading-order (LO) Feynman diagrams (graphical representation of the interactions) of diphoton production are shown in figure 2.6. The quarks in the proton are annihilated with their anti particle to produce a photon. Gluons and quarks in the final state produce jets of many particles in the detector. These interactions can be modeled almost exclusively with the well known QED interaction. To calculate higher order diagrams like in figure 2.7 QCD interactions have to be simulated.

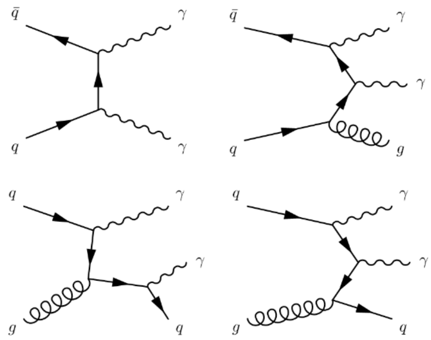


Figure 2.6: LO processes for the diphoton final state [14]

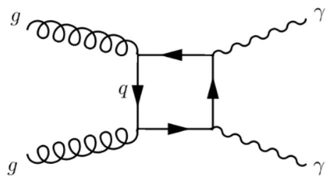


Figure 2.7: Example of a NLO process for the diphoton final state [14]

3 Unfolding Methods

Suppose, one is in the situation that one has a sample of data and wants to find the underlying probability distribution. Normally one would just fit the data by a known parametric function to obtain estimators for the parameters. If the parametric form is not known one can fill the data in a histogram and divide by the total number of events to get a estimator of the probability in every bin (cf. [15]). The problem of unfolding is more complicated. One wants to know the distribution of events of the underlying physical process but one has a sample of data containing statistical fluctuations and effects of the detector such as measurement errors, finite resolution and so on. To fit the data to theory predictions, the effects of the detector have to be corrected. This step in the analysis is called unfolding.

Mathematically, detector effects are treated by folding (convolution) of the truth distribution t with the detector response function R.

$$\int_{-\infty}^{\infty} R(s,x)t(x)dx = r(s) \tag{3.1}$$

The task of unfolding is now the inversion of this equation namely correcting the data distribution (on reco level) to get the truth distribution. But this folding integral is not analytically invertible. Therefore the problem is discretized. The response matrix $\boldsymbol{R}$ is the discrete version of the detector function R. The truth distribution t is converted into the truth vector $\boldsymbol{t}$ and the data distribution r becomes the data vector $\boldsymbol{r}$.

$$\boldsymbol{R}\boldsymbol{t} = \boldsymbol{r} \tag{3.2}$$

The inversion of this equation may look easy: invert the matrix and use it on the measured vector $\boldsymbol{r}$. But in practice the exact response matrix corresponding to our detector and our measurement is not known because it cannot be calculated analytically from the known detector effects. But a best guess is estimated for the response matrix $\boldsymbol{R}^{\mathrm{MC}}$ by using a MC simulation for the process to get the MC truth vector $\boldsymbol{t}^{\mathrm{MC}}$ and use a detector simulation to produce the MC events on reco level $\boldsymbol{r}^{\mathrm{MC}}$ of these truth events.

In the next sections four basic unfolding concepts are discussed.

3.1 Matrix Inversion

The inversion of the equation for the MC distributions

$$\boldsymbol{R}^{\mathrm{MC}}\boldsymbol{t}^{\mathrm{MC}} = \boldsymbol{r}^{\mathrm{MC}} \tag{3.3}$$

is solved by inverting the matrix to

$$\boldsymbol{R}^{\mathrm{MC}-\mathbf{1}}\boldsymbol{r}^{\mathrm{MC}} = \boldsymbol{t}^{\mathrm{MC}}. \tag{3.4}$$

Since this is the exact mathematical inversion, the exact MC truth vector is reproduced.

But a measured data vector $\boldsymbol{r}$, which differs atleast by statistical fluctuations, has to be unfolded. Sometimes, the MC simulation is not perfect and the data has features that are not in the simulated MC reco vector $\boldsymbol{r}^{\mathrm{MC}}$. Since the exact response matrix for the measurement $\boldsymbol{R}$ is not known, the response matrix obtained from the MC simulation $\boldsymbol{R}^{\mathrm{MC}}$ is used.

$$\boldsymbol{R}^{\mathrm{MC}-\mathbf{1}}\boldsymbol{r} = \boldsymbol{t}. \tag{3.5}$$

Because only the response matrix of the MC simulation is known and can be used, the superscript MC is not used anymore for the response matrix.

The procedure is investigated in a 2 bin toy example (cf. [16]). Assume the response matrix is

$$\boldsymbol{R} = \frac{1}{2}\begin{bmatrix} 1+\epsilon & 1-\epsilon \\ 1-\epsilon & 1+\epsilon \end{bmatrix} \quad \text{with} \quad 0 \leq \epsilon \leq 1. \tag{3.6}$$

ϵ parameterizes the quality of the detector. $\epsilon = 1$ is the perfect detector and $\epsilon = 0$ corresponds to a detector which cannot distinguish between the two bins. In order to get some insights into the unfolding process a Singular Value Decomposition is used on $\boldsymbol{R}$. The Singular Value Decomposition of the real matrix $\boldsymbol{R}$ is the factorization in three matrices.

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{3.7}$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices and $\boldsymbol{S}$ is a diagonal matrix. The Eigenvalues of $\boldsymbol{S}$ are called singular values. These singular values are non-negative and they carry important information of the matrix $\boldsymbol{R}$. For example if $\boldsymbol{A}$ is degenerated at least one singular value is zero. In this two bin case this results in

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \quad \text{with} \quad S = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix}. \tag{3.8}$$

To unfold, the response matrix is inverted and used on the measured data vector $\boldsymbol{r}$.

$$\boldsymbol{t} = \boldsymbol{R}^{-1}\boldsymbol{r} = \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{r} = \frac{r_1 - r_2}{2\epsilon}\begin{pmatrix}1\\-1\end{pmatrix} + \frac{r_1 + r_2}{2}\begin{pmatrix}1\\1\end{pmatrix} \tag{3.9}$$

Assume that bin 1 and bin 2 have almost the same number of events. Then the difference $r_1 - r_2$ is a random number because it is dominated by statistical fluctuations. When ϵ is small the random value of the difference of the bins is amplified over the well behaved sum of the bins. Therefore the result is mainly influenced by the not significant and statistically fluctuating term $r_1 - r_2$.

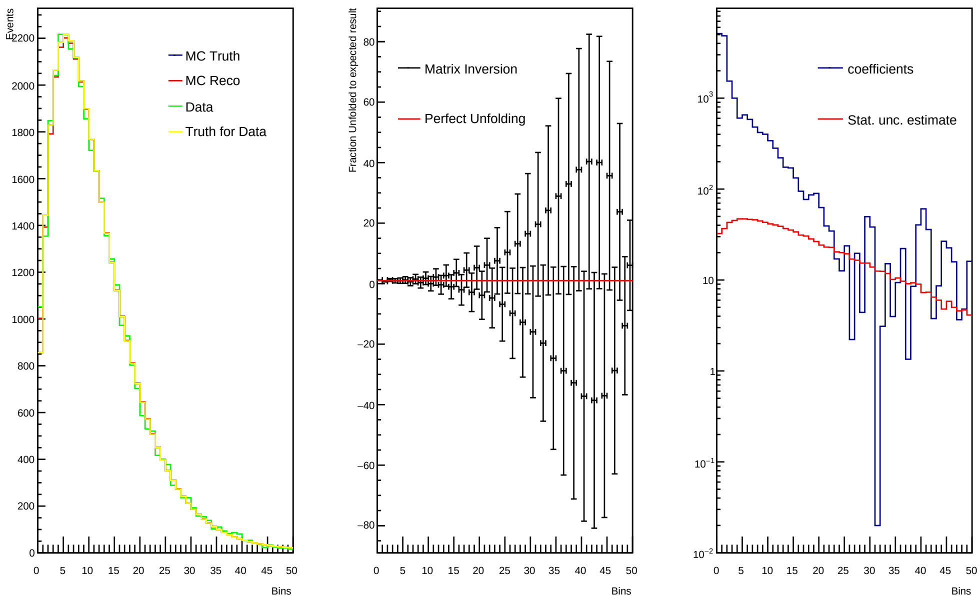


Figure 3.1: Left: Comparison of MC truth and reco distribution with the data. Middle: ratio of the unfolded data with matrix inversion and the expected result. (ratio of 1 $\widehat{=}$ perfect unfolding.) Right: Coefficients $\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{r}$ and the estimate of statistical uncertainty. Beyond bin 20 coefficients are dominated by statistical uncertainty and cause large fluctuations in the result of the unfolding.

To further study this behavior a 50 bins example is shown in figure 3.1. A toy response matrix is built with the same structure as the two bins example in equation 3.8. The migration parameter ϵ is $0,5$. In the toy model shown in figure 3.1 the MC truth distribution is assumed to have the form $x\exp(-ax)$. The MC reco distribution is produced by folding the MC truth with the response matrix. The data, that is unfolded, is a statistical fluctuation of the MC distribution. In the middle plot the ratio of the unfolded data with matrix inversion and the (only in toy models known) expected result is shown. In the first few bins the matrix inversion is working well until large oscillations with large uncertainties start. This is the result of the effect described in the two bin example above. When the coefficients $\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{r}$ (generalization of the terms like $r_1 - r_2$ of the two bin example) are smaller than the statistical uncertainty, they become random and cause the oscillating pattern of the result. This phenomenon is common for the unfolding with matrix inversion. It is hard to compare theory predictions with these oscillating unfolded results, but the result of the matrix inversion is unbiased. Other unfolding algorithms were invented that produce smooth results at the cost of a bias. The effect of the bias has to be estimated and results in a uncertainty of the method.

3.2 Bin-by-Bin Unfolding

The easiest approach to the unfolding problem is Bin-by-Bin unfolding. The ratio of the MC truth to reco distribution is multiplied with the data.

$$t_i = r_i \frac{t_i^{MC}}{r_i^{MC}} \tag{3.10}$$

This very easy unfolding method solves the problem of the oscillating results of the unfolding with matrix inversion. In figure 3.2 the exact same setup as in figure 3.1 is shown. The result of the unfolding with the Bin-by-Bin method is not fluctuating as much as the result of the matrix inversion.

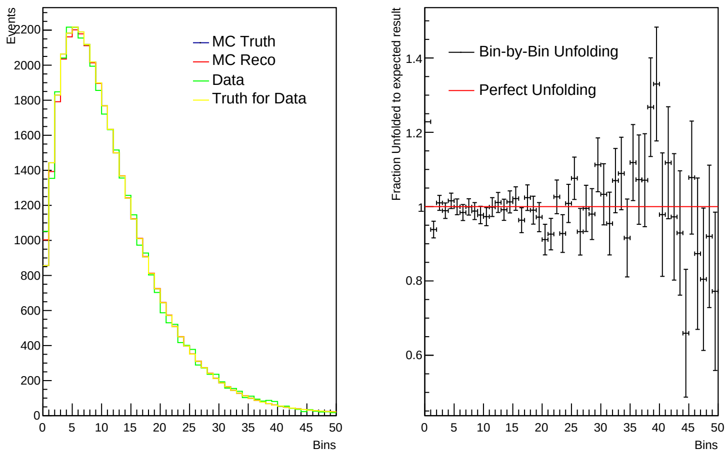


Figure 3.2: Bin-by-Bin unfolding in the same setup as in figure 3.1.

In this example the MC simulation matches the data very well. In common use cases the MC description of the data is not as accurate as in this toy example. To show how the Bin-by-Bin unfolding is performing for use cases like these, another toy example is created. In this toy, plotted in figure 3.3, the data has an additional peak which is not in the MC simulation (for example a new particle that is not simulated in the MC). The detector is smearing and biasing the distributions. In the ratio of the result to the expected result the additional peak of the data is clearly visible and the uncertainties are small. The Bin-by-Bin method is adding a strong bias when the MC simulation is not matching the data well. Because the Bin-by-Bin method is only applying one correction factor for each bin (in equation 3.10 truth bin i is only affected by reco bin i), the method cannot account for the off diagonal elements in the response matrix (migrations) which describe a shift of events from one truth bin to another reco bin. The bias B of this method can be calculated as (τ is the true value of the truth distribution)

$$\begin{aligned} B_i &= E[t_i] - \tau_i \\ &= E\left[r_i \frac{t_i^{MC}}{r_i^{MC}}\right] - \tau_i. \end{aligned} \tag{3.11}$$

Using the expectation of the data $E[r_i] = \beta_i$ the bias is (cf. [15])

$$B_i = \left(\frac{t_i^{\mathrm{MC}}}{r_i^{\mathrm{MC}}} - \frac{\tau_i}{\beta_i}\right)\beta_i \tag{3.12}$$

When the the ratio of the MC truth and reco is not matching the ratio of the data at truth and reco level, the Bin-by-Bin method has a nonzero bias. In figure 3.3 on the right side the factor $\frac{t_i^{\mathrm{MC}}}{r_i^{\mathrm{MC}}} - \frac{\tau_i}{\beta_i}$ is shown. Around bin 35 the bias is increasing and the Bin-by-Bin method is failing to correctly unfold and without the knowledge of the truth for the data (as in real applications of unfolding) this could lead to false conclusions.

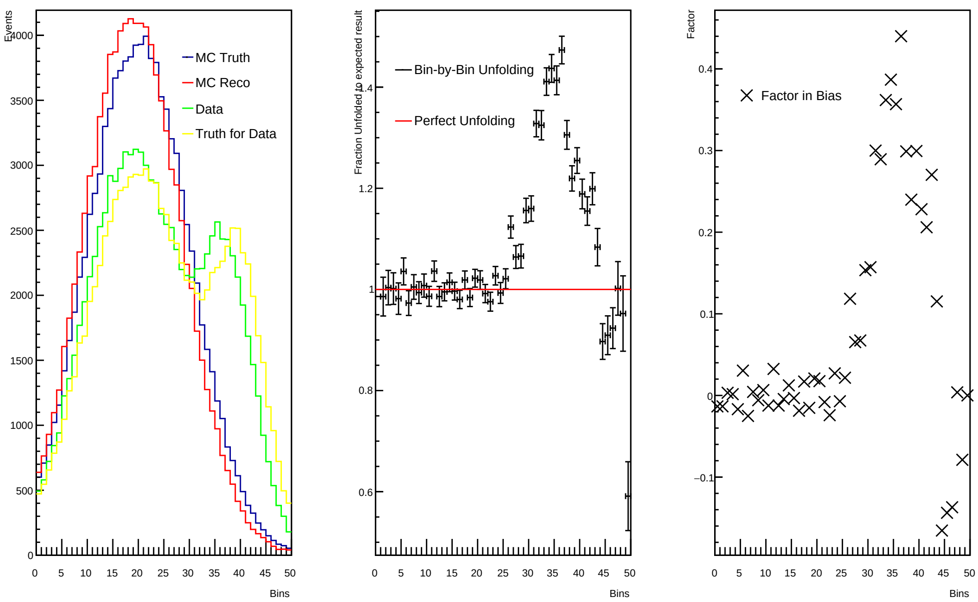


Figure 3.3: Bin-by-Bin unfolding in a setup where the data has features that are not in the MC simulation. Left side: MC truth and reco as well as data and truth for data are shown. Middle: Ratio of the unfolded with the expected result (truth for data). Right side: Factor in the bias calculation $\frac{t_i^{\mathrm{MC}}}{r_i^{\mathrm{MC}}} - \frac{\tau_i}{\beta_i}$ is plotted

3.3 Regularized Unfolding

As shown in section 3.1 the small singular values in equation 3.9 amplify statistical fluctuations. To solve this problem consider the equivalent least squares problem [16] (n_r: number of reco bins, n_t: number of truth bins)

$$\sum_{i=0}^{n_r} \left(\sum_{j=0}^{n_t} R_{ij} t_j - r_i \right)^2 = \min. \tag{3.13}$$

Since the measurement uncertainties vary from bin to bin, the equations have different significance. Therefore the following equations should be considered which weight each equation according to their uncertainty.

$$\sum_{i=0}^{n_r} \left(\frac{\sum_{j=0}^{n_t} R_{ij} t_j - r_i}{\Delta r_i} \right)^2 = \min. \tag{3.14}$$

The general case of 3.14 reads

$$(\boldsymbol{Rt} - \boldsymbol{r})^T \boldsymbol{B}^{-1} (\boldsymbol{Rt} - \boldsymbol{r}) = \min \tag{3.15}$$

$\boldsymbol{B}$ is the covariance matrix of the data vector $\boldsymbol{r}$. Since this formulation of the problem is completely equivalent to the problem in section 3.1, the result is fluctuating in the same way. The oscillations have to be suppressed by adding a regularization term

$$(\boldsymbol{Rt} - \boldsymbol{r})^T \boldsymbol{B}^{-1} (\boldsymbol{Rt} - \boldsymbol{r}) + \tau (\boldsymbol{Xt})^T (\boldsymbol{Xt}) = \min. \tag{3.16}$$

$\boldsymbol{X}$ defines some knowledge of the result. Since physical results are usually smooth functions, a matrix $\boldsymbol{X}$ is used that penalizes not smooth behavior. This problem can be solved with Singular Value Decomposition in a similar way as the two bin example in section 3.1. In essence the problem causing factor in equation 3.9 is transformed.

$$\frac{1}{\epsilon} \to \frac{\epsilon}{\epsilon^2 + \tau} \tag{3.17}$$

τ regularizes singularities due to small singular values ϵ. This prevents the amplification of statistical fluctuations and improves therefore the unfolding. The strength of the regularization τ can be chosen arbitrarily.

In figure 3.4 the same toy example is shown as in the Bin-by-Bin study in figure 3.3. The unfolded result is relatively smooth and is almost everywhere matching the expected result in the uncertainties.

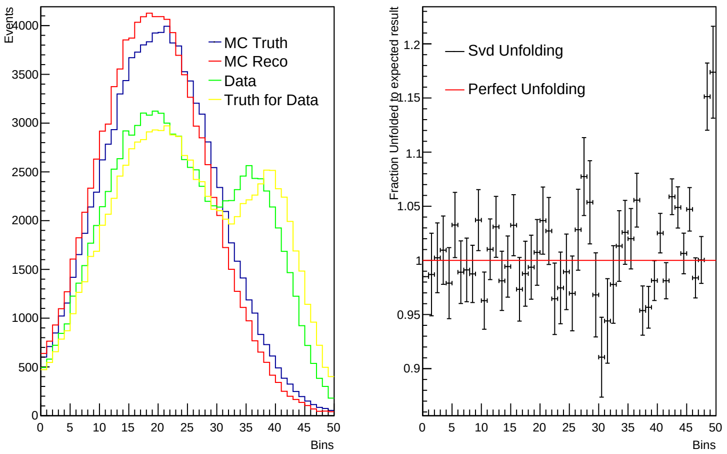


Figure 3.4: unfolding with regularized unfolding method. Left: Toy model. Right: Ratio of unfolded result with expected result.

3.4 Bayesian Unfolding

Bayesian unfolding (also called iterative unfolding) is based on the Bayes Theorem. It was invented by D'Agostini [17].
Unfolding aims to get the number of events in a bin on particle level (truth bin) out of a given number of events on detector level (reco bin). To get the number of events for the truth bin $n(t_i)$ the conditional probability [1] $P(t_i|r_j)$ is multiplied with the number of events of reco bin $n(r_j)$.

$$n(t_i) = \sum_{j=1}^{n_r} P(t_i|r_j)n(r_j) \tag{3.18}$$

n_r is the number of reco bins and n_t the number of truth bins. The conditional probability $P(t_i|r_j)$ is calculated using the Bayes Theorem.

$$P(t_i|r_j) = \frac{P(r_j|t_i)P(t_i)}{\sum_{l=1}^{n_t} P(r_j|t_l)P(t_l)} \tag{3.19}$$

$P(r_j|t_i)$ is extracted from the detector simulation by normalizing the response matrix in every truth bin (called folding matrix). The probability of the truth $P(t_i)$ is not known a priori. One has to assume this so called prior. The standard assumption is the MC truth distribution. To reduce the effect of the assumption of the prior the method is done for multiple iterations. Therefore the number of iterations is the regularization parameter of the Bayesian unfolding. Few iterations correspond to a highly regularized result and with higher number of iterations the statistical uncertainties increases. In every application of this method, the number of iterations has to be chosen independently.

[1] In [17] and the other literature truth bins are called causes and reco bins are the effects.

In figure 3.5 the Bayesian unfolding is shown for the same toy model as in the previous sections. The Bayesian unfolding is producing stable and smooth results even when the MC model is not simulating the data well.

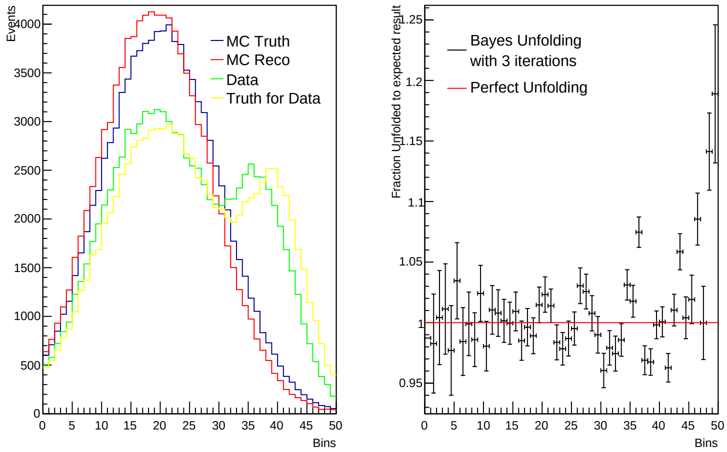


Figure 3.5: Bayesian unfolding, Left: Toy model, Right: Ratio of unfolded result with expected result.

4 Diphoton Analysis

4.1 Introduction

The measurement of the diphoton final state is of big interest to understand the strong interactions in the Standard Model. Photons have good coupling to quarks and a clean experimental signature, therefore they can be used as a probe to study the strong interactions.
As shown in section 4.3.8 leading order simulations of the process cannot explain the observation. Therefore precise measurements of the diphoton system offer valuable benchmarks for higher order corrections. Furthermore, understanding the diphoton process is important for new physics models because it is a background for new processes predicting two photons in the final state. This state is measured in 8 observables. There are the transverse momenta of the individual photons E_{T,γ_1} and E_{T,γ_2}, the invariant mass of the diphoton system $m_{\gamma\gamma}$, the transverse momentum of the diphoton system $p_{T,\gamma\gamma}$ and it's transverse component with respect to the thrust axis $\hat{t} = (\vec{p}_T^1 - \vec{p}_T^2)/|\vec{p}_T^1 - \vec{p}_T^2|$ called $a_{T,\gamma\gamma}$, the angle $\Delta\phi_{\gamma\gamma}$ between the photons in the azimuthal plane, the absolute value of the cosine of the scattering angle with respect to the direction of the proton beams expressed as a function of the difference in pseudorapidity between the photons $|\cos\Theta_\eta^*| = \tanh(\frac{|\Delta\eta_{\gamma\gamma}|}{2})$ and the ϕ_η^* variable defined as $\phi_\eta^* = \tan(\frac{\pi - \Delta\phi_{\gamma\gamma}}{2})\sin(\Theta_\eta^*)$ ([18]). These observables are measured with different resolution and the binning is chosen accordingly.

4.2 Summary of the Analysis on Reco Level

This analysis is done using the full 2015 and 2016 data of the ATLAS detector. It contains ~ 36.1 fb^{-1} $\pm 3.2\%$. Only events acquired in stable beam conditions and passing detector and data-quality requirements are considered. During this period the Trigger used to record diphoton candidates is g35_loose_g25_loose. This Trigger requires for the leading photon (photon with higher transverse momentum) $E_{T,\gamma_1} \geq 35$ GeV and for the subleading photon $E_{T,\gamma_2} \geq 25$ GeV. Both photons have to pass the loose identification. In the analysis, the reconstructed photons are required to have 5 GeV higher transverse momentum than in the trigger, because the trigger efficiency is not stable at 35 and 25 GeV. Photons reconstructed in the region where the end-cap and the barrel of the electromagnetic calorimeter is jointed together ($1.37 < |\eta_\gamma| < 1.56$) are excluded, as they are poorly reconstructed. Photons are also required to be in the region $|\eta_\gamma| < 2.37$, which is the region covered by the tracker and where the electromagnetic calorimeter have good granularity, allowing good reconstruction of photons. These selection criteria are called pre-selection.

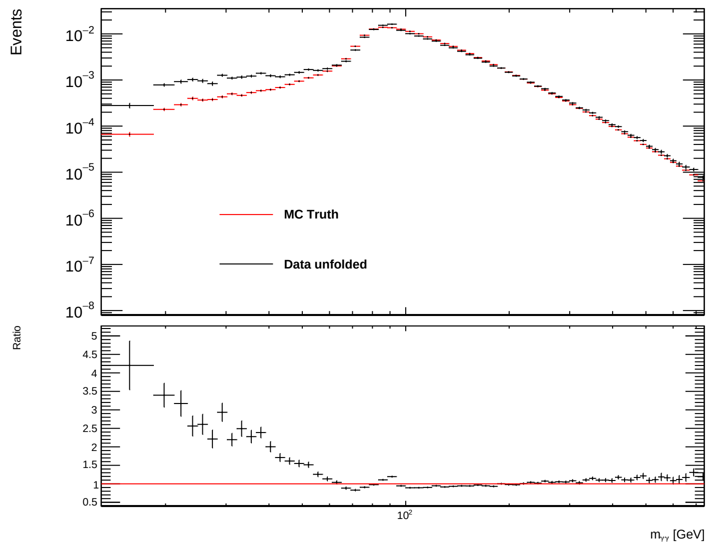


Figure 4.1: Normalized MC truth and unfolded data distribution with statistical uncertainty only. In this data the Drell-Yan($Z \to ee$) background is not subtracted. This results in a small peak at the mass of the Z at ~ 91 GeV.

In order to reject background of for example jets, more strict criteria are applied in the analysis. Since the interactions and shapes of the shower produced by a photons are well known, the photon identification (photon id.) is using the fine granularity of the electromagnetic calorimeter to reject background. Here variables, that reflect the profile of the energy deposits in the calorimeter, are built. Cuts are applied on these variables to decrease the fraction of background in the sample. Another selection of events is done using the isolation of the photons. A clustering algorithm reconstructs particle hits in the calorimeter. Then the energy of all the clusters in a cone around the photon impact is summed up and a rectangle of a few cells centered at the photon candidate is removed. A cut is applied on this energy to analyze only isolated photons. This achieves further background rejection because jets usually consist of many particles that produce several impacts in an area in the calorimeter. For these criteria information of the MC simulation of the process and the detector simulation is used. The uncertainties in these simulation have to be taken into account and result in uncertainties of these analysis steps. The photon identification and isolation accomplishes very good efficiency of identifying real photons but the cross-section of jet production is by several orders of magnitude higher than the diphoton production. Therefore despite the very good identification, around 29% of the sample is background caused by jet misidentification. Other background processes are $\gamma + e$ and ee with misidentified electrons. The background is estimated extrapolating from control regions rich on background to the signal region. The extrapolation is performed assuming that the photon identification and the isolation criteria are independent of each other. This is almost the case, but small remnants of correlations between the two criteria has to be taken into account by assigning a systematic uncertainty. The background of two misidentified electrons is 2%. Since the electron cross section is peaked around the Z boson mass ($\sim$ 91 GeV), the distribution of invariant mass $m_{\gamma\gamma}$ is peaking at this value when the ee background is not subtracted. This is shown in figure 4.1.

4.3 Unfolding

In this section the unfolding of the 13 TeV analysis is presented. The Bayesian unfolding method is used. The uncertainties from previous analysis steps are propagated in section 4.3.5 and the uncertainties of the unfolding are estimated in section 4.3.6. The selection criteria discussed in section 4.2 are applied to the MC events and result in the corrections in section 4.3.1. In the main text only the results for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ are shown. The same plots for all observables are in the appendix 6.

4.3.1 Corrections: Fakes and Misses

According to the selection criteria discussed in section 4.2 the MC events can be classified in three groups. Events that pass both the reco and the truth selection are filled in the response matrix. These events are the input for the actual unfolding step. Events that pass reco selection but don't pass truth selection are called fakes. Fakes are filled in a separate histogram in RooUnfold. The subtraction of the fake events is the first step in the unfolding. Events that fail reco selection but pass truth selection are misses. They result in a multiplicative correction at the end of the unfolding. To reduce statistical fluctuations both misses and fakes are smoothed with the class TGraphSmooth by ROOT [19].

In figure 4.2 the fractions of the events that pass truth and reco selection relative to all truth events called efficiency and the fraction of events that pass truth and reco cuts compared to all reco events, called purity, are shown. The efficiency and purity are the complement to the miss and fake rates. Note that in contrast to this definition, purity is in the context of unfolding sometimes defined as the fraction of events in the diagonal elements of the response matrix.

$$\text{efficiency } \mathcal{E} = \frac{n^{pass\ reco\ and\ truth}}{N^{pass\ reco}} \qquad \text{purity } P = \frac{n^{pass\ reco\ and\ truth}}{N^{pass\ truth}} \tag{4.1}$$

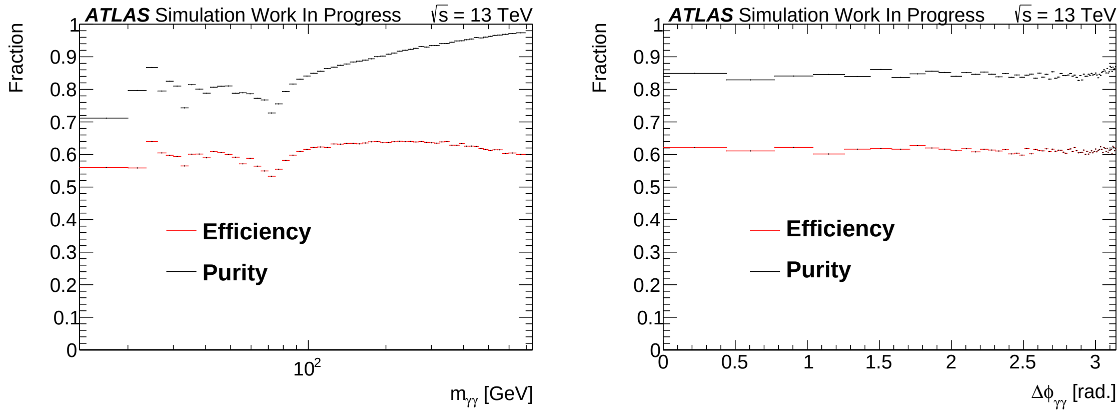


Figure 4.2: Efficiency and Purity for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

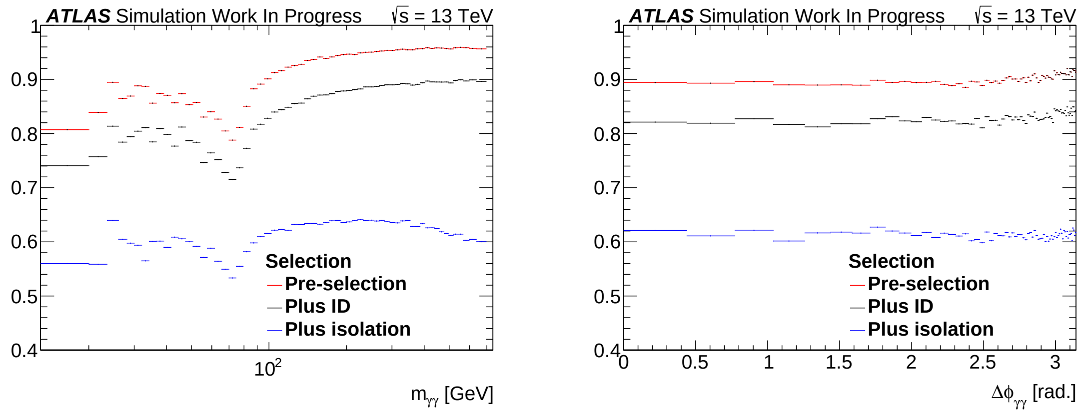


Figure 4.3: Efficiency for different selections for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

4.3.2 Method: Bayesian Unfolding

In the diphoton analysis the Bayesian unfolding method, described in section 3.4, from the RooUnfold package [20] is used because the results of this method are stable and smooth. The number of iteration is 3 for all observables. This choice is justified in section 4.3.6.4. To check the method a sanity check is performed. The MC reco distribution is unfolded and compared to the MC truth in figure 4.4. Since the prior is the MC truth the unfolding is perfectly restoring the MC truth.

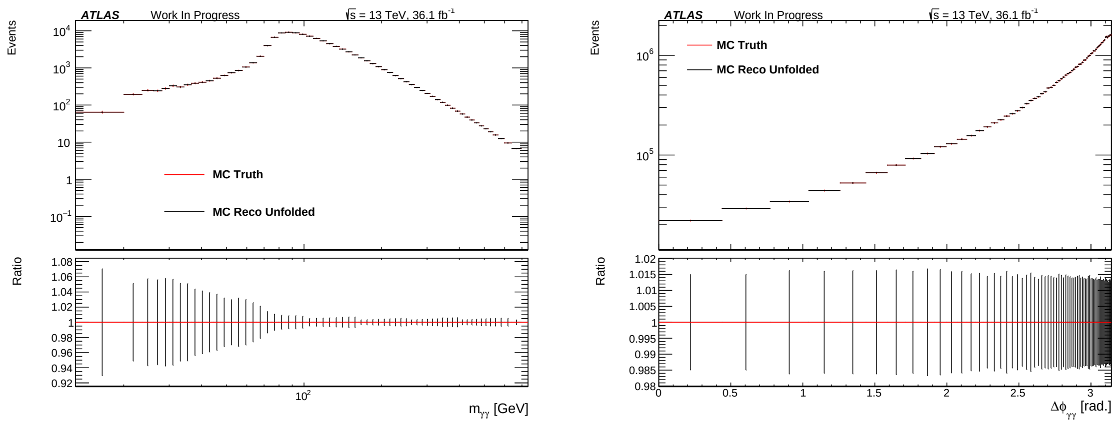


Figure 4.4: Sanity Check: The MC reco distribution used to fill the response matrix is unfolded. Here the sanity check for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only are presented. All observables are in appendix 6

4.3.3 Detector Response Matrix

As shown in section 3 the response matrix is modeling the detector effects. It is filled with events that pass truth and reco selection.

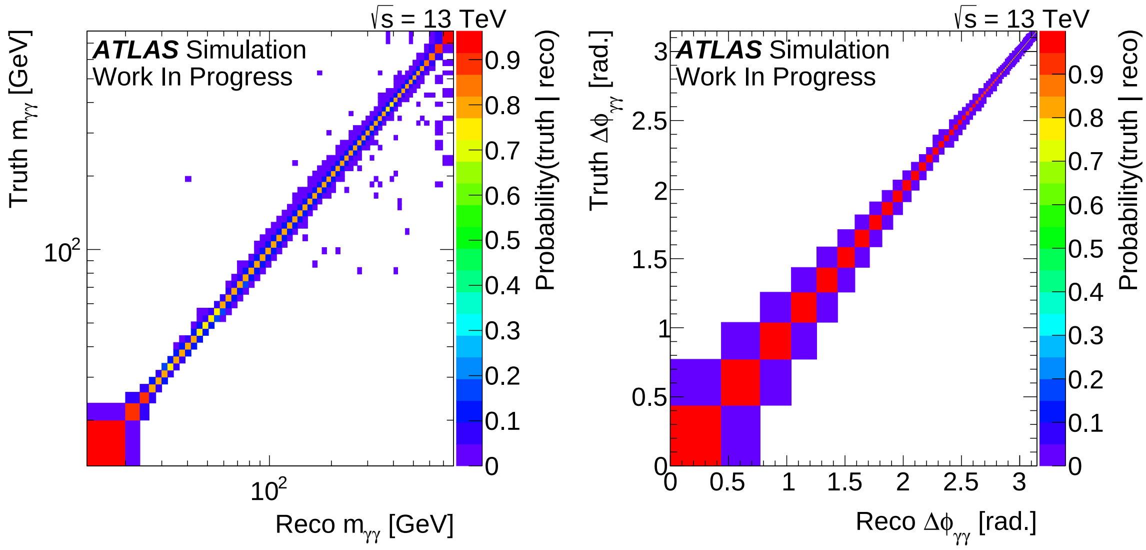


Figure 4.5: Response matrices normalized in every reco bin for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

The binning is chosen taking into account the detector resolution; the bin size is set to be four times the resolution, except for the regions with low statistics, where larger bins are used. In this way, 70 to 80 % of the events are in the diagonal bins, and only a small fraction of events fall in the overflow or underflow bins. The response matrices normalized in every reco bin are shown in figure 4.5. The matrices are nicely diagonal but for example in $m_{\gamma\gamma}$ a few events are very far away from the diagonal elements. This is probably a result of truth reco mismatching of a few events. This needs further investigations but has a negligible impact on the unfolding. The ATLAS detector has a fine resolution in the angular observables which results in the fine binning of these variables.

4.3.4 Results

In figure 4.6 the effects of the steps discussed in section 4.3.1 and 4.3.2 are shown separately for $m_{\gamma\gamma}$. The peaks in the miss and fakes at 70 GeV corresponds to the p_T cuts at 30 and 40 GeV of the photons. Due to the almost diagonal response matrix the iterative unfolding step has only a small effect on the result of the complete unfolding.

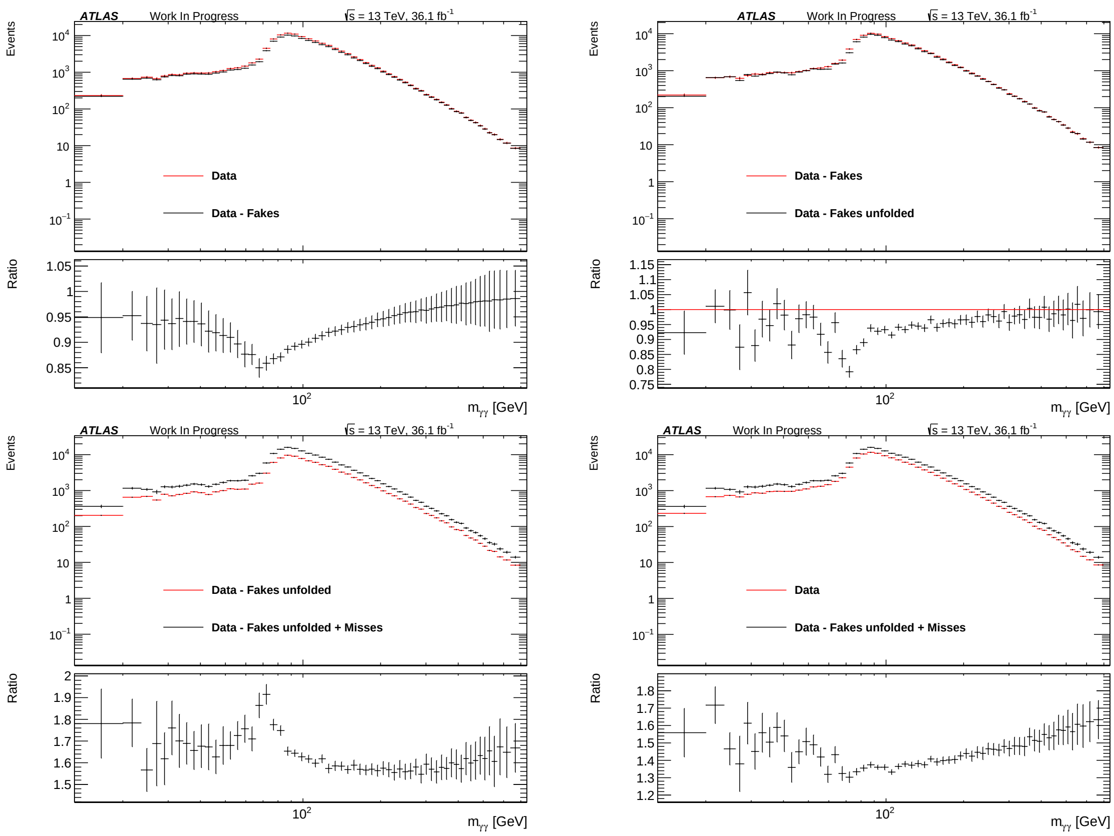


Figure 4.6: Plots showing every step of the unfolding for $m_{\gamma\gamma}$. Top left: Effect of only the subtraction of fakes. Top right: Effect of the iterative unfolding method. Bottom left: Effect of missed events. Bottom right: Full unfolding containing fakes and misses. Uncertainty bars correspond to data statistical uncertainty only.

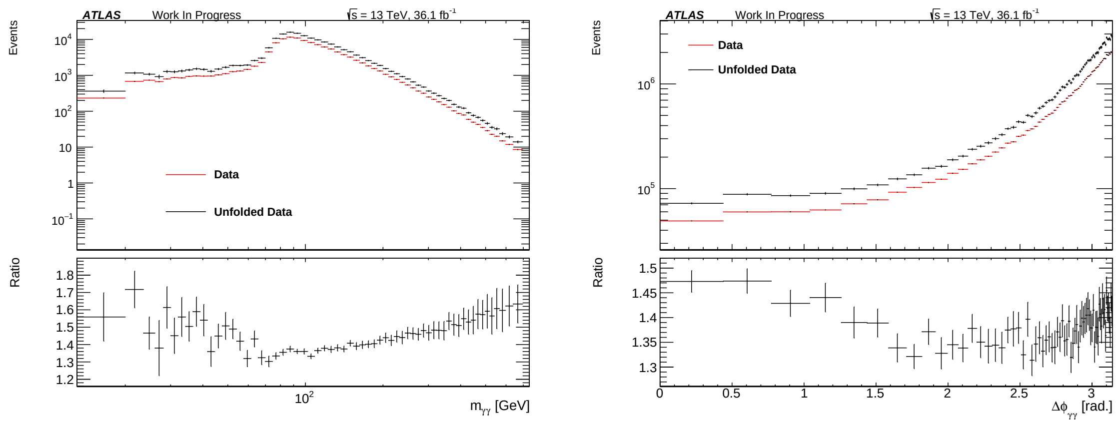


Figure 4.7: Data and unfolded data with data statistical uncertainties only are shown for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

4.3.5 Propagation of Background Estimation Uncertainties

As presented in section 4.2, Uncertainties originating in the background estimation and other steps of the analysis have to be propagated through the unfolding procedure. There are two kinds of these uncertainties. There is the point to point uncorrelated data statistical uncertainty and the point to point correlated systematic uncertainties. Since we are using a iterative unfolding method the uncertainties cannot be propagated analytically but toys (pseudo-experiments) have to be generated.

For the statistical uncertainty on the data toys are generated by picking a random number according to a normal distribution. The statistical uncertainty is multiplied with this random number and the nominal value of that bin is added. The resulting value is the nominal value of the toy. This is done for every bin with an independent random number. The toy is unfolded with the nominal response matrix and the final uncertainty is estimated as the root mean square (RMS) of 100 different toys.

The systematic uncertainties of the background estimation have to be treated differently because of their correlation. A variation is generated by shifting the nominal value of all bins up by one sigma. This variation is unfolded using the nominal procedure and the uncertainty is the difference of this variation to the nominal unfolding result. In the same way, the one sigma down uncertainties are propagated to the unfolded result.

Photon identification and isolation are affecting the background estimation as well as the unfolding itself because for example the miss and fake rates change. Therefore the uncertainty from identification and isolation have to be estimated in the unfolding procedure. These two parts are added in a correlated way which is described in section 4.3.6.1.

The systematic and statistical uncertainties from the background estimation are shown in figure 4.8. For the photon identification and isolation only the propagated uncertainty from the background estimation is shown.

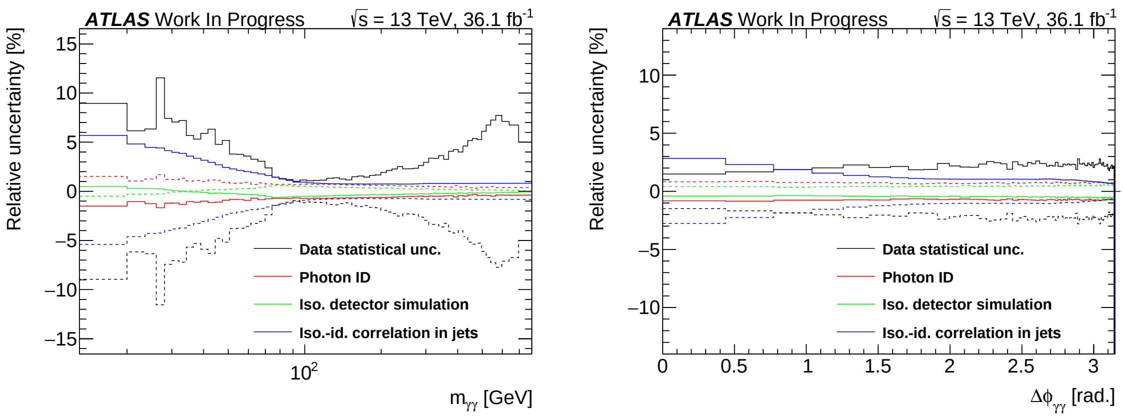


Figure 4.8: Propagation of signal yields uncertainties through the unfolding code for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

4.3.6 Unfolding Uncertainties

Uncertainties discussed in this section are mostly coming from imperfect modeling of the detector. These uncertainties result in changes in the response matrix and/or the miss and fake rates. An exception is the method uncertainty discussed in section 4.3.6.4. This uncertainty is the effect of the regularization in the unfolding procedure.

4.3.6.1 Detector Simulation Uncertainties

Uncertainties considered in this section originate in the simulation of the detector. To estimate the effect of for example the calibration, tools provided by the $e\gamma$-group are used. These tools change the energy scale, energy resolution, photon identification and isolation in the simulation. These variations affect the response matrix as well as the miss and fake rates. Therefore new response matrices, miss and fake rates are filled with these changed events. The data is unfolded and the difference between the data unfolded with the nominal setup and the result of the varied response matrix is taken as the uncertainty. The uncertainty is then smoothed using TGraphSmooth from ROOT [19].

The simulation predictions for the efficiencies of the photon identification and isolation requirements affect both, the background estimation as well as the unfolding inputs. These two aspects of the uncertainty have to be combined in a correlated way by using the up(down)-variation of the data described in section 4.3.5 and the up(down)-variation of the response matrix in the unfolding. On the other hand, in this analysis, the energy scale and resolution simulation uncertainties only affect the unfolding inputs, and have negligible impact on the background decomposition.

The uncertainties are shown in figure 4.9. For the photon identification only the impact in the unfolding part is plotted. The component of the isolation efficiency is not yet estimated.

The energy scale uncertainty is dominating the other uncertainties in every observable. These uncertainties are of the order of 1%.

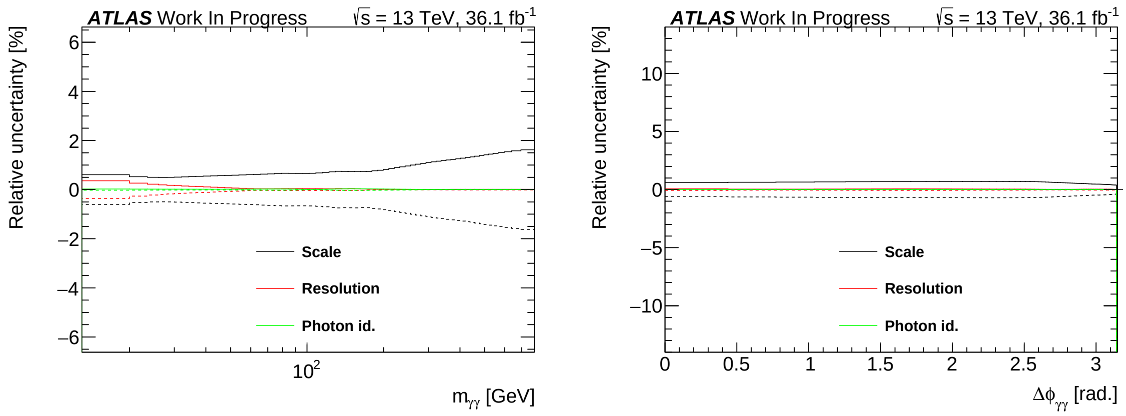


Figure 4.9: Systematic Uncertainties on the detector simulation for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

4.3.6.2 Physics Modeling

For the response matrix and the fake and miss rates a SHERPA [21] sample is used. To probe the impact of the modeling of the physical process, the unfolding with the SHERPA input will be compared to the unfolding with PYTHIA (other MC event generator) [22] input. This is not yet done because the Pythia samples were not produced at the time of writing.

4.3.6.3 MC Statistical Uncertainty

The lack of infinite statistics of the MC sample induces uncertainties in the response matrix, miss and fake rates. To estimate this uncertainty a Bootstrap package is used. For every event a random number is produced according to a Poisson distribution with mean 1. This random number is multiplied with the original weight of the event. The event is filled with the new weight into a replica (toy) of the response matrix, the fake and miss rates. The procedure is repeated for 100 replicas. After these replicas are filled the fakes and misses are smoothed. The unfolding is done using these replicas and the uncertainty of the sample is calculated in every bin. In principle the MC statistical uncertainties are uncorrelated bin by bin, but the unfolding and the smoothing can add some correlation between the bins. This is shown in figure 4.10 for $m_{\gamma\gamma}$. Bins close to the diagonal are correlated with each other. The bins far off the diagonal show only small correlation. This correlation is propagated to the full correlation matrix shown in figure 4.15.

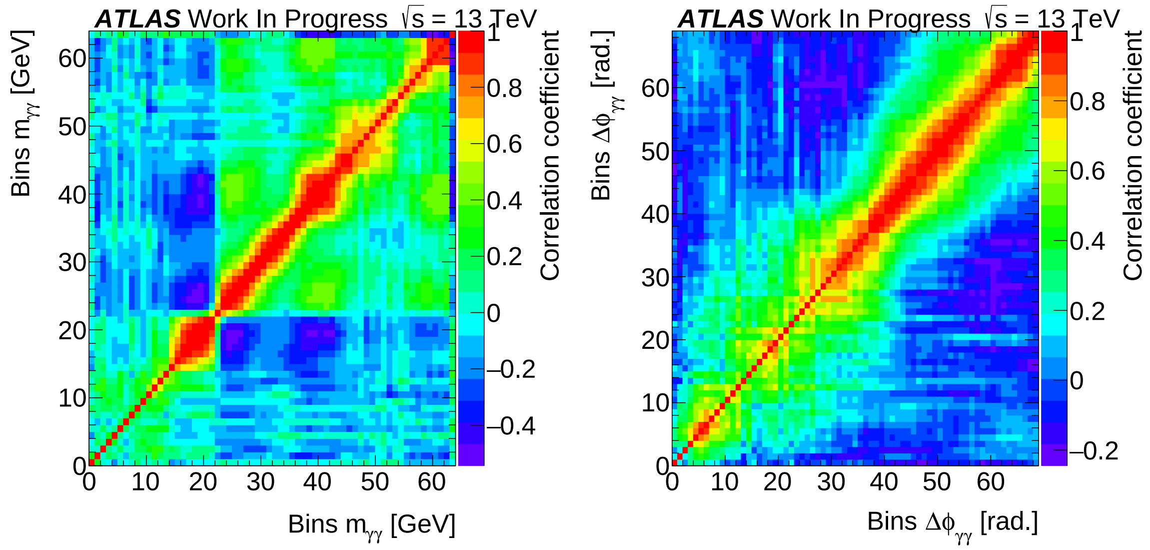


Figure 4.10: Correlation matrix for the MC statistical uncertainty only

The uncertainty is shown in figure 4.12.

4.3.6.4 Unfolding Method Uncertainty

In the previous sections we described the uncertainty on the unfolding with the origin in the MC simulation. In this section we estimate the uncertainty of the unfolding procedure itself. By definition unfolding with matrix inversion has no uncertainty because in this method there is no regularization and therefore there is no information added which can be a source of an uncertainty. Since we are using a unfolding method with regularization the uncertainty of this step has to be taken into account. This is done with a data driven method. The main idea of this method is to reweight the MC truth such that the corresponding MC reco matches the data (better). This reweighted reco distribution is unfolded and compared to the MC truth to have an estimate on how good the unfolding is performing. The test is done in several steps:

1. Unfold the data
2. Reweight the MC truth distribution with a smooth function to match the unfolded data
3. Fold the reweighted truth distribution with the response matrix to get the reweighted MC reco
4. Unfold the reweighted MC reco
5. Compare this unfolded reweighted MC reco to the MC truth to estimate the uncertainty
6. Smooth the uncertainty with the class TGraphSmooth of ROOT

The uncertainty is shown in figure 4.12.

To motivate the choice of 3 iterations, we present in figure 4.11 the data statistical uncertainty and the method uncertainty for $m_{\gamma\gamma}$ with different numbers of iterations used in the unfolding. The data statistical uncertainty is increasing when the number of iterations is rising. In contrast, the method uncertainty is decreasing when the number of iterations is increased. The improvement in the method uncertainty going from 2 to 3 iterations is significant, while not increasing the data statistical uncertainty by a large amount. 4 iterations are not improving the method uncertainty anymore, therefore the number of iterations used in the analysis is 3 for all observables.

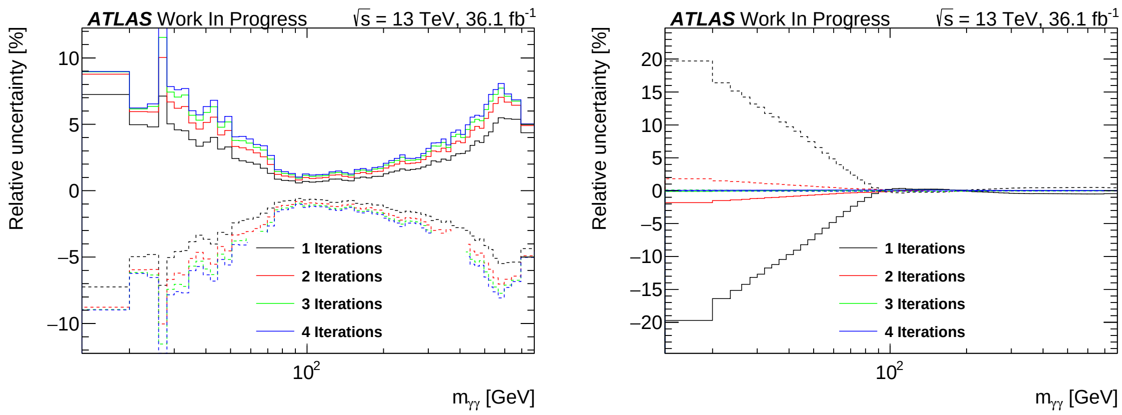


Figure 4.11: Uncertainties for different number of iterations. Data statistical uncertainty is shown on the left side and method uncertainty is shown on the right side.

4.3.6.5 Summary of Unfolding Uncertainties

To summarize this section we show in figure 4.12 all the previously discussed uncertainties. The MC statistical uncertainty and the detector simulations uncertainties are in all observables of the order of a few percent.

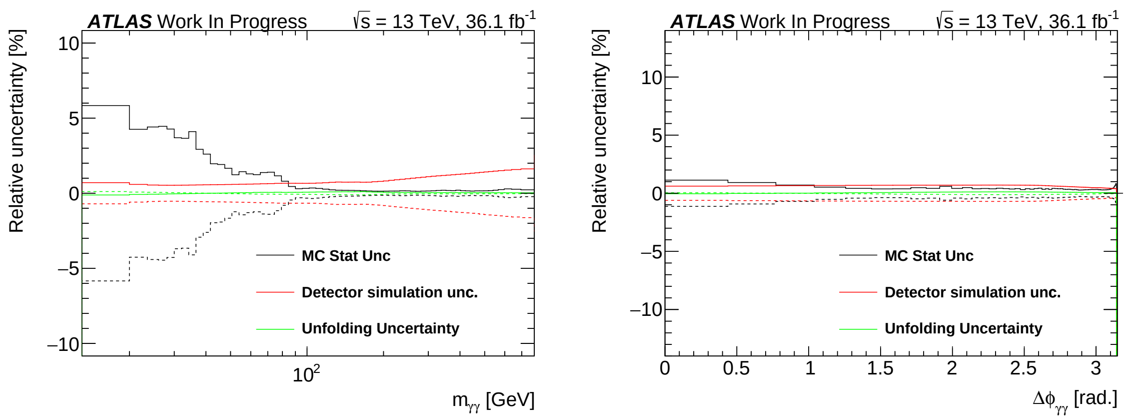


Figure 4.12: Systematic Uncertainties on the detector simulation for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

4.3.7 Uncertainty Summary

In figure 4.13 all previously discussed uncertainties are presented. The efficiency uncertainties are the photon identification and isolation. The uncertainties on the energy measurement are the energy scale and energy resolution combined. All systematic uncertainties combined and the statistical uncertainty are shown in figure 4.14. The uncertainties on these results have different correlation configuration, some of them are correlated bin by bin, with different correlation patterns, while some others are uncorrelated between the bins. In order to properly take into account the uncertainty correlation in future uses of this measurement, for instance in comparison with different theory predictions, we are extracting and providing the bins correlation matrix in figure 4.15 for each observable.

The statistical uncertainty is in almost all observables the dominating uncertainty. Overall the uncertainties are in the order of 5% except for low $m_{\gamma\gamma}$.

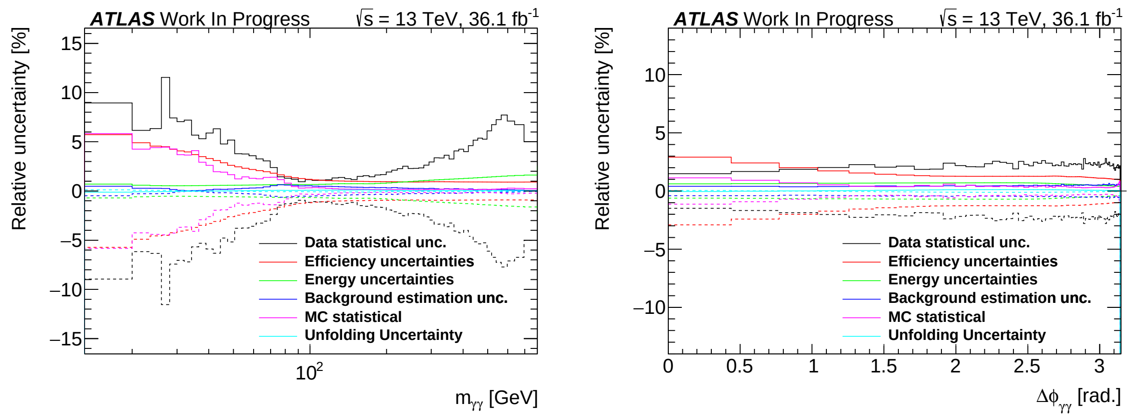


Figure 4.13: Summary of all uncertainties for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

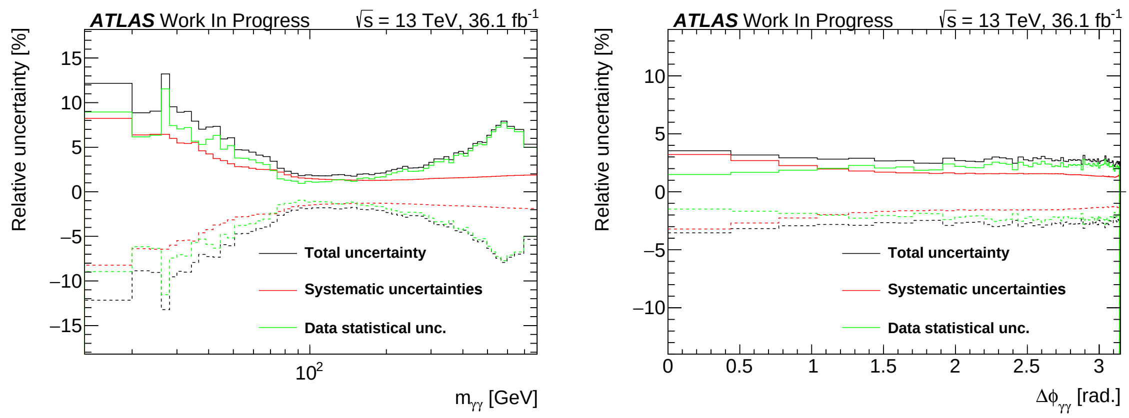


Figure 4.14: Systematic and statistical uncertainties combined for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

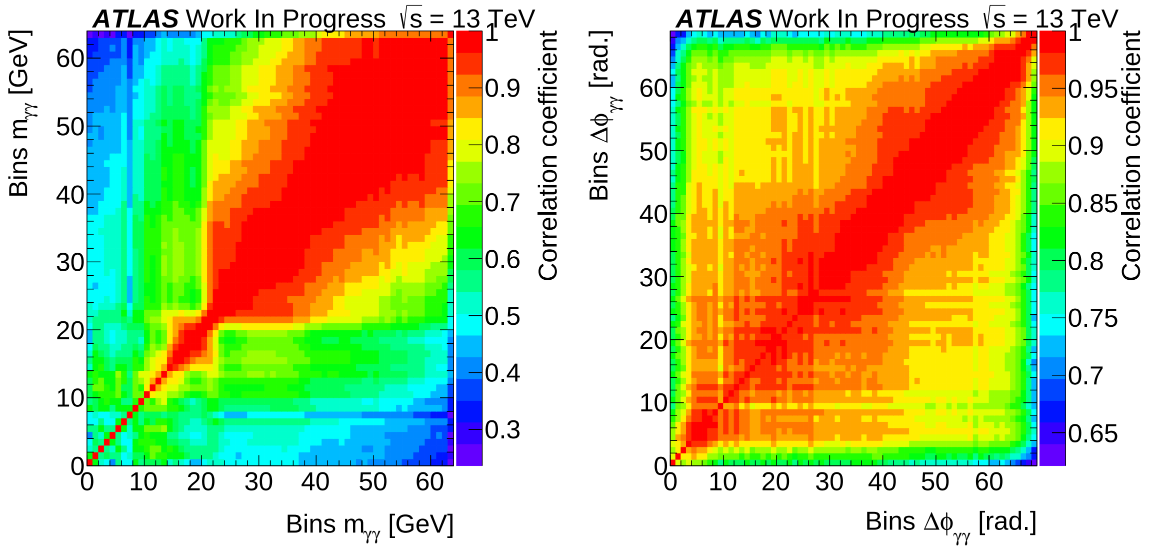


Figure 4.15: Correlation Matrices for $m_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$ only. All observables are presented in the appendix 6

.

The results of this 13 TeV analysis can be compared to the 8 TeV analysis [18]. In the momentum based observables ($m_{\gamma\gamma}$,$p_{T,\gamma\gamma}$ and $a_{T,\gamma\gamma}$) the statistical uncertainty decreased at high values of the observables. In the 8 TeV analysis the uncertainties of some observables were dominated by the systematic uncertainties but here in the 13 TeV analysis systematic uncertainties are lower than the statistical uncertainty. Therefore the accuracy of the measurement can be significantly improved by the full Run 2 data which contains approximately 4 times higher luminosity.

4.3.8 Comparison with LO Predictions

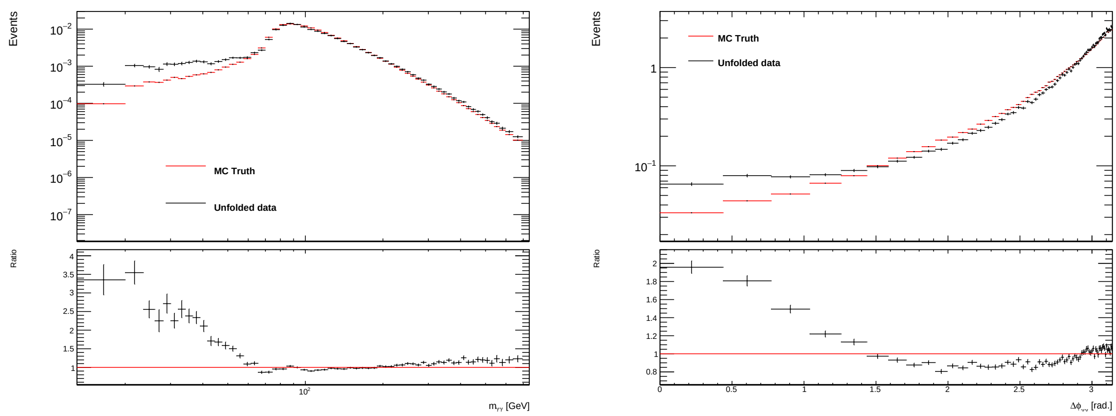


Figure 4.16: Comparison of LO predictions with unfolded data with all uncertainties

In figure 4.16 the unfolded results an the LO predictions are shown. LO predictions cannot describe the observed distributions for all observables. In both $E_{T,\gamma 1}$ and $E_{T,\gamma 2}$ the predictions are not matching the unfolded data. In the low $m_{\gamma\gamma}$ region the simulation predict up to a factor of 3 less events than observed. Similar behavior is observed at $|\cos\Theta^*_n|$ close to 1, low $\Delta\Phi_{\gamma\gamma}$ and high Φ^*_η. These regions have to be improved by for example NLO calculations.

5 Summary and Conclusion

In this study the basic unfolding methods were presented and discussed. The unfolding with matrix inversion is unbiased but results in large fluctuation which make a comparison with theory by eye difficult. Other unfolding methods accept a bias but produce smoother result. The bin by bin method is the easiest method, but as was pointed out in section 3.2, it cannot account for migrations, if the MC and data do not match well. An other unfolding method accounting for these migrations is for example the regularized unfolding. Here the small singular values that cause the oscillation in the unfolding with matrix inversion are regularized to decrease their impact on the result of the unfolding. The method that is used in the presented diphoton analysis is the Bayesian unfolding. Bayes' Theorem is applied and the result of the unfolding is improved with several iterations.
The application of the Bayesian unfolding in the 13 TeV analysis was presented in section 4. Uncertainties were estimated and propagated to the final results. The comparison with LO predictions showed significant differences in several observables. Higher order simulations have to be compared to the unfolded data to check the theory models. To decrease the statistical uncertainties, the analysis will be done for the full Run 2 data.

6 Appendix

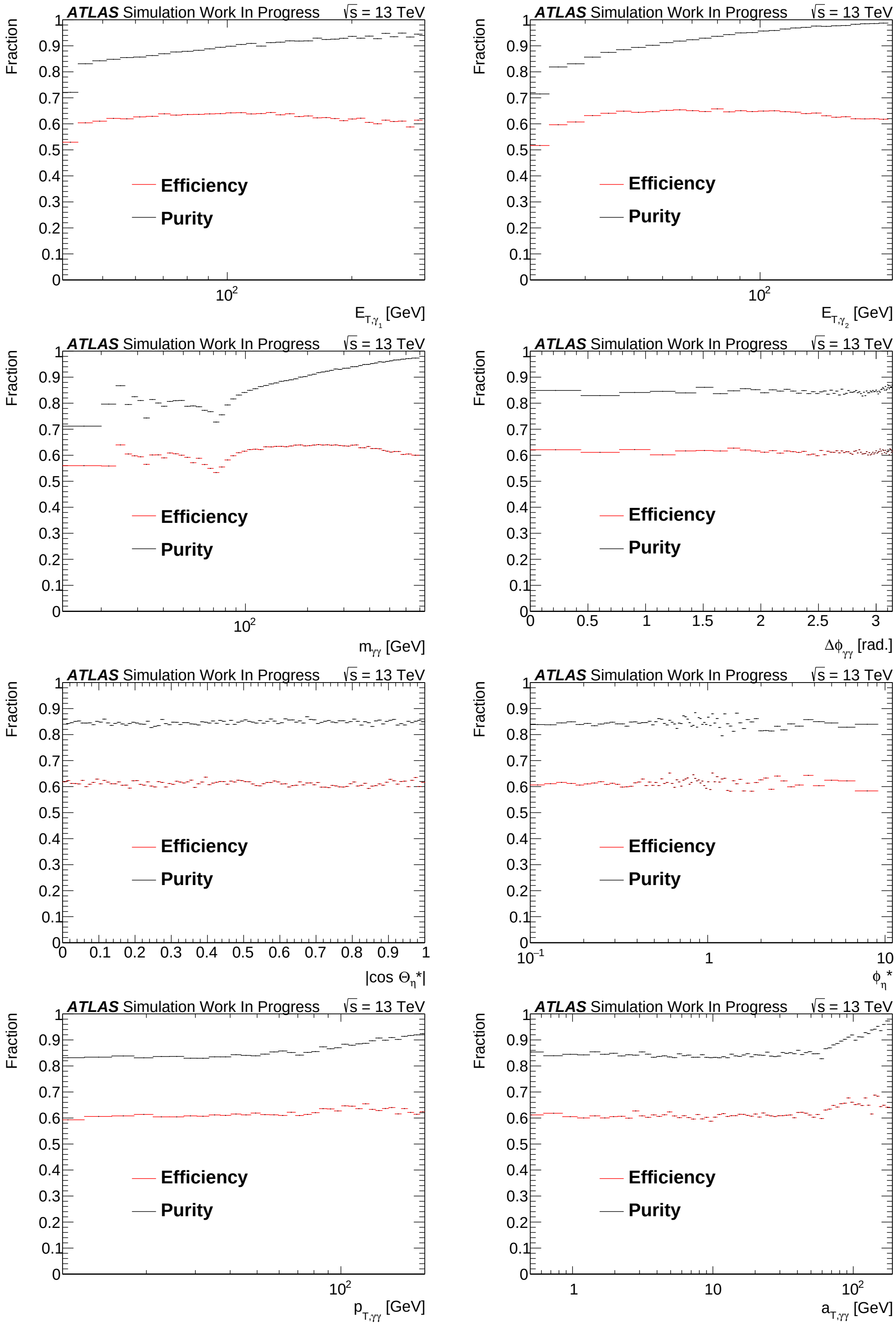


Figure 6.1: Efficiency and fake rate

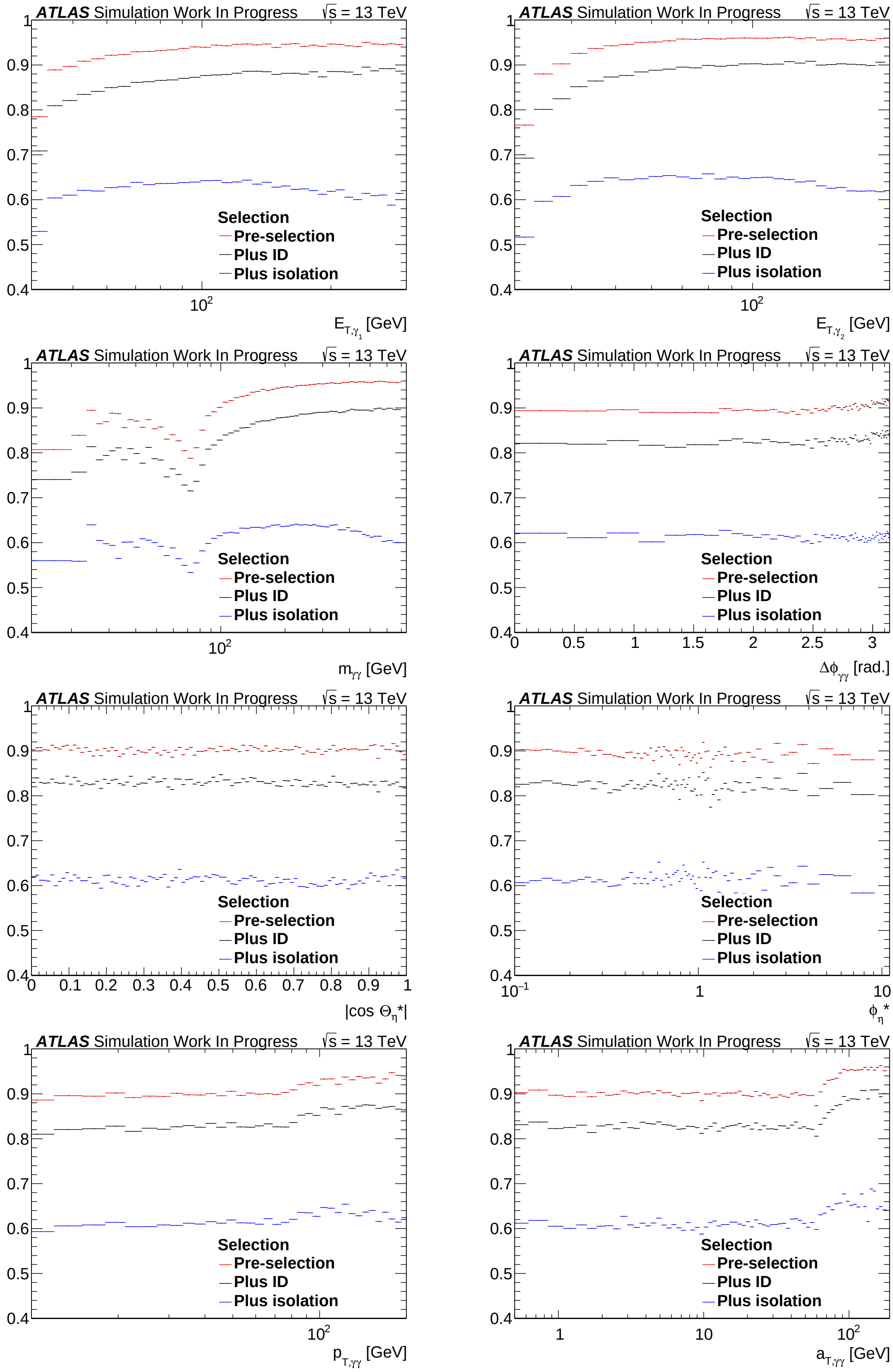


Figure 6.2: Efficiency for different selections

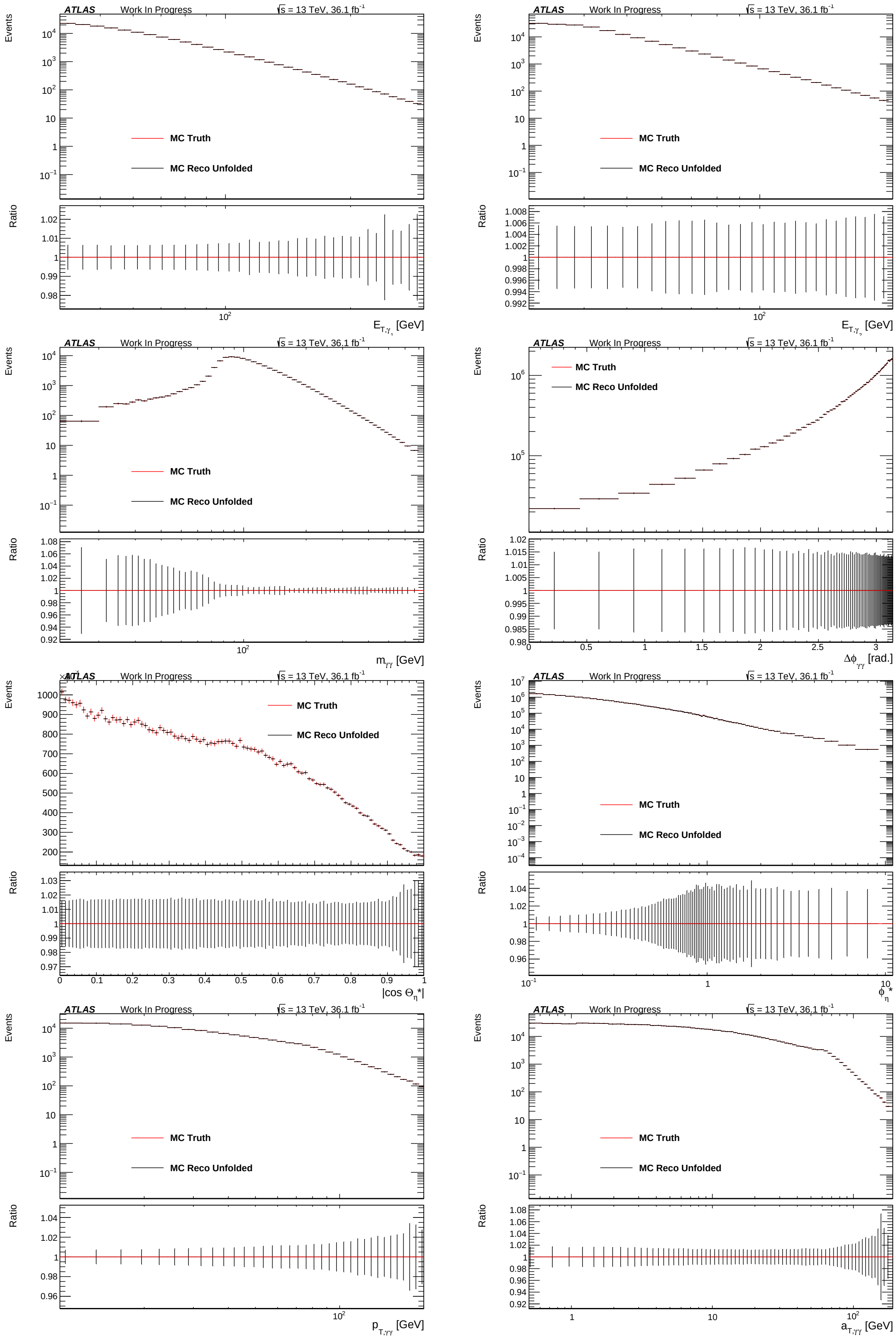
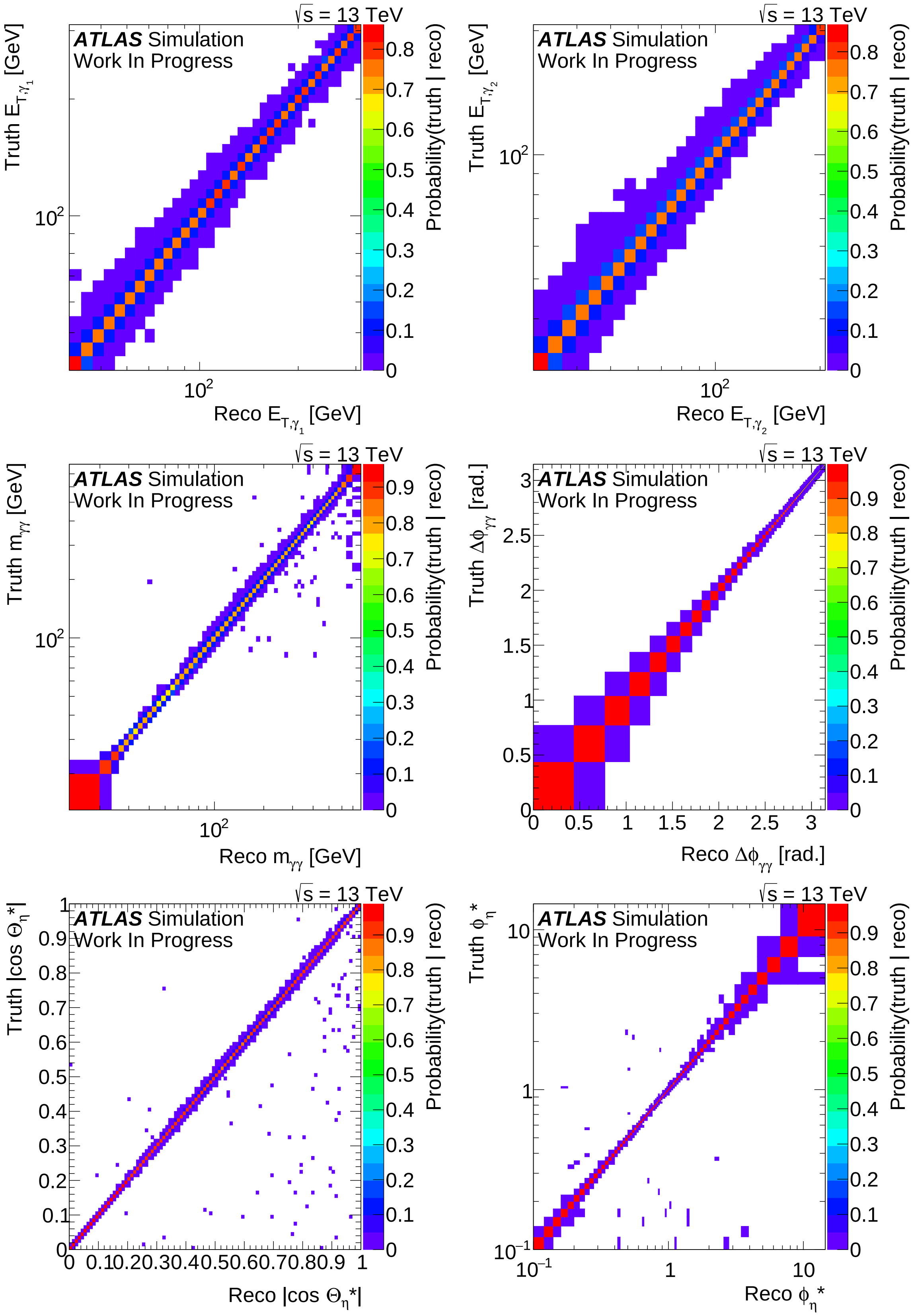


Figure 6.3: Sanity Check: The MC reco distribution used to fill the response matrix is unfolded.


√s = 13 TeV
ATLAS Simulation
Work In Progress
Truth E_{T,γ1} [GeV]
Reco E_{T,γ1} [GeV]
Probability(truth | reco)
√s = 13 TeV
ATLAS Simulation
Work In Progress
Truth E_{T,γ2} [GeV]
Reco E_{T,γ2} [GeV]
Probability(truth | reco)
√s = 13 TeV
ATLAS Simulation
Work In Progress
Truth m_{γγ} [GeV]
Reco m_{γγ} [GeV]
Probability(truth | reco)
√s = 13 TeV
ATLAS Simulation
Work In Progress
Truth Δφ_{γγ} [rad.]
Reco Δφ_{γγ} [rad.]
Probability(truth | reco)
√s = 13 TeV
ATLAS Simulation
Work In Progress
Truth |cos Θ_η*|
Reco |cos Θ_η*|
Probability(truth | reco)
√s = 13 TeV
ATLAS Simulation
Work In Progress
Truth φ_η*
Reco φ_η*
Probability(truth | reco)

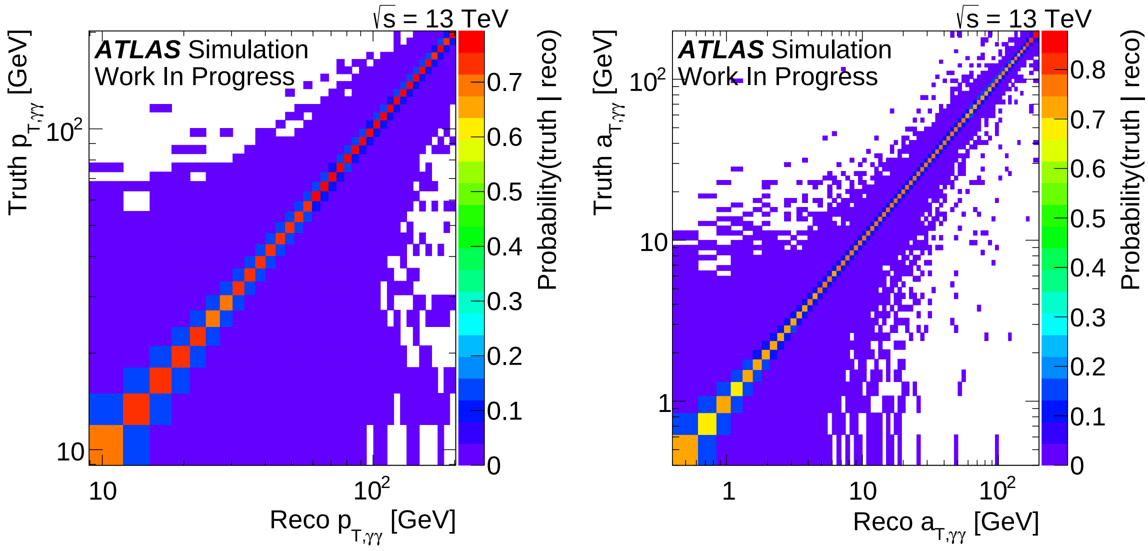


Figure 6.4: Response matrices for the observables normalized in every reco bin

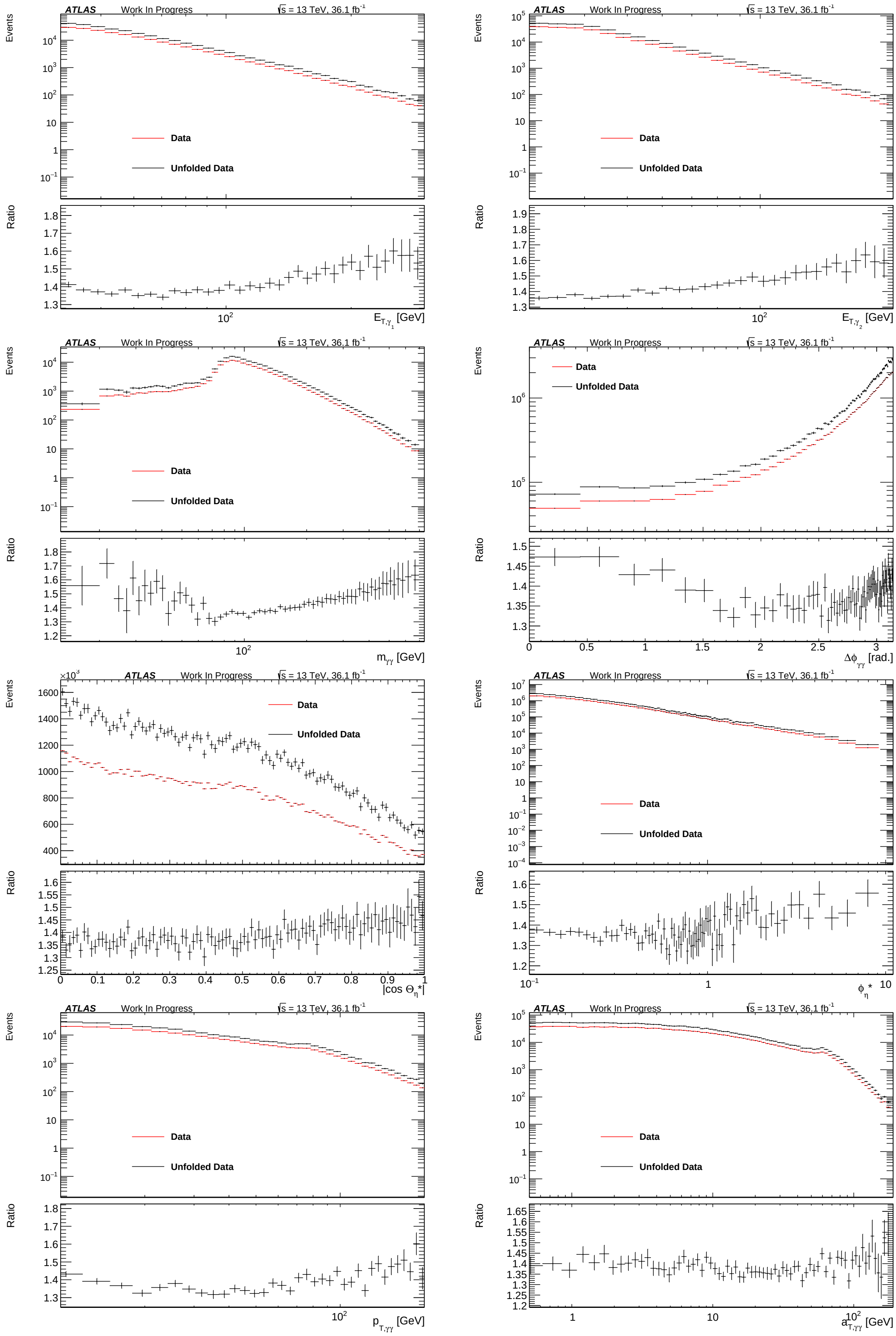


Figure 6.5: Data and unfolded data with data statistical uncertainty only are shown

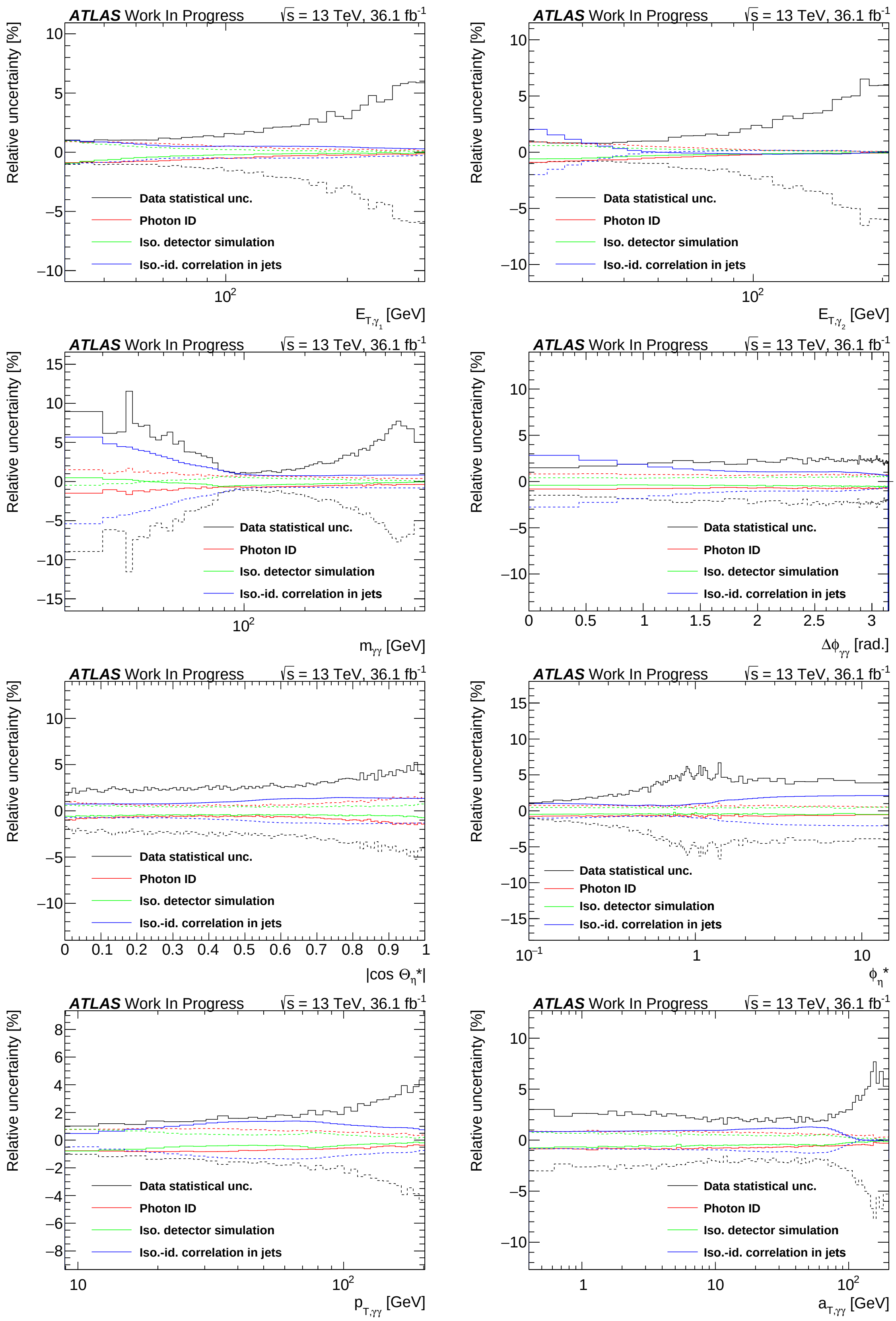


Figure 6.6: Propagation of signal yields uncertainties through the unfolding code

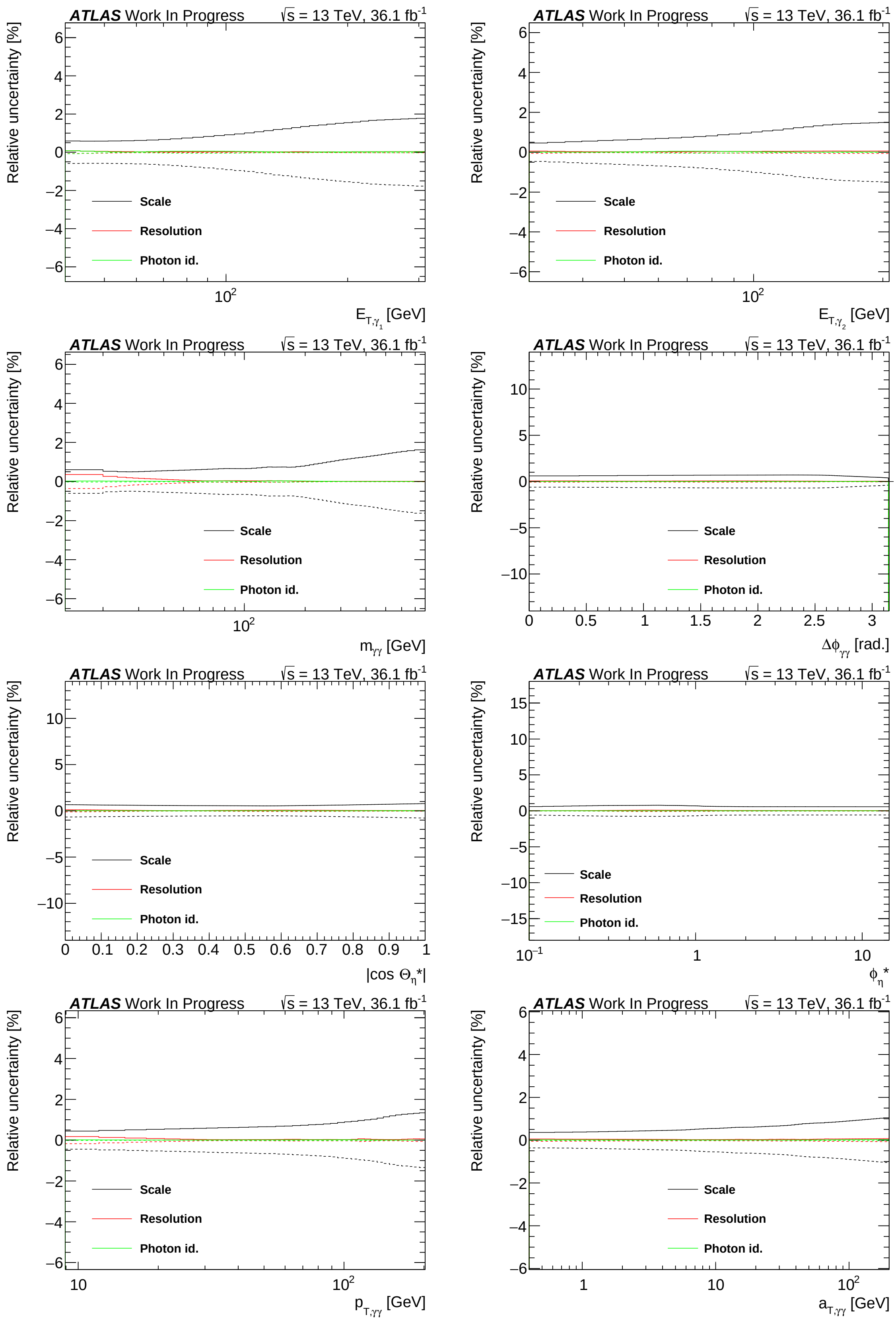


Figure 6.7: Systematic Uncertainties on the detector simulation

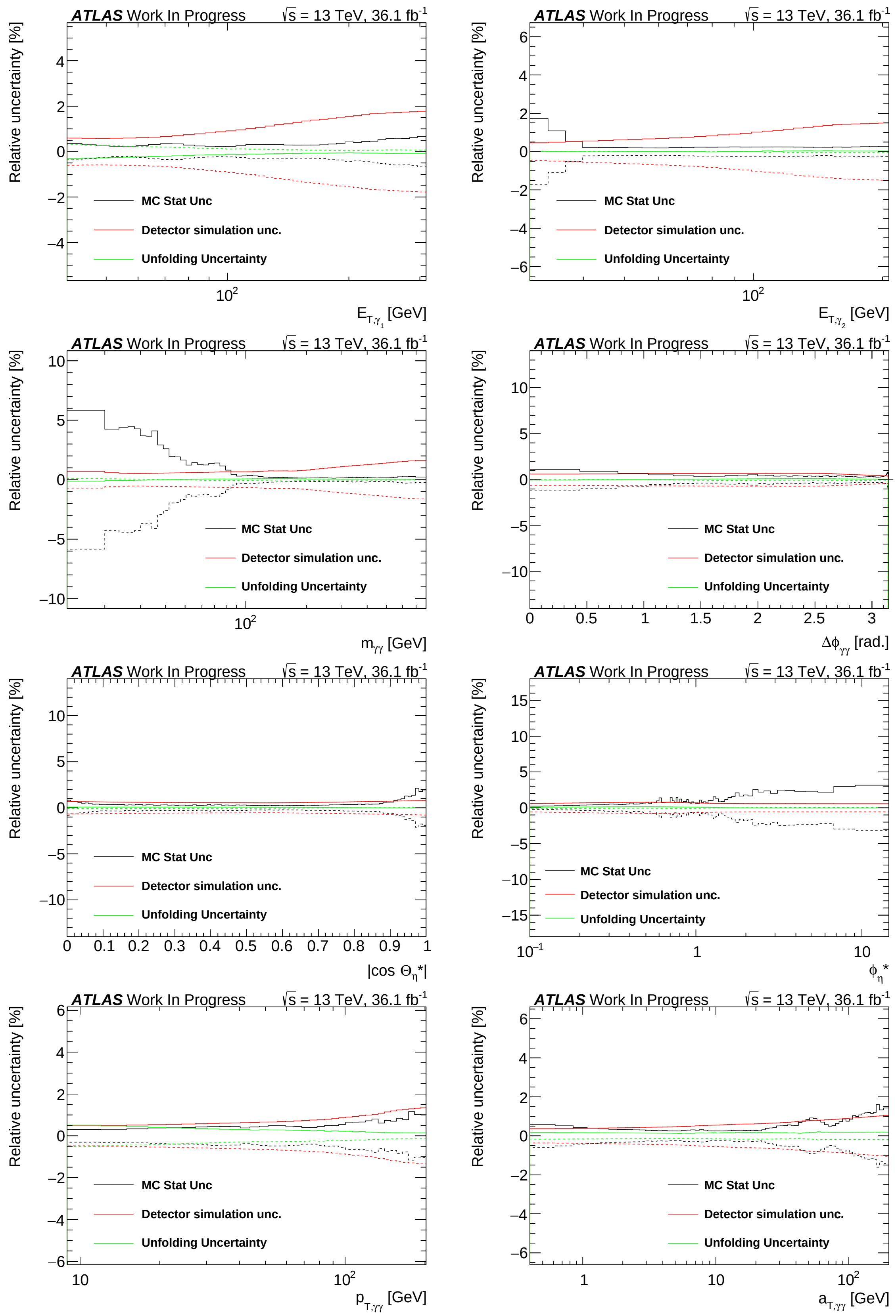


Figure 6.8: Systematic Uncertainties on the detector simulation

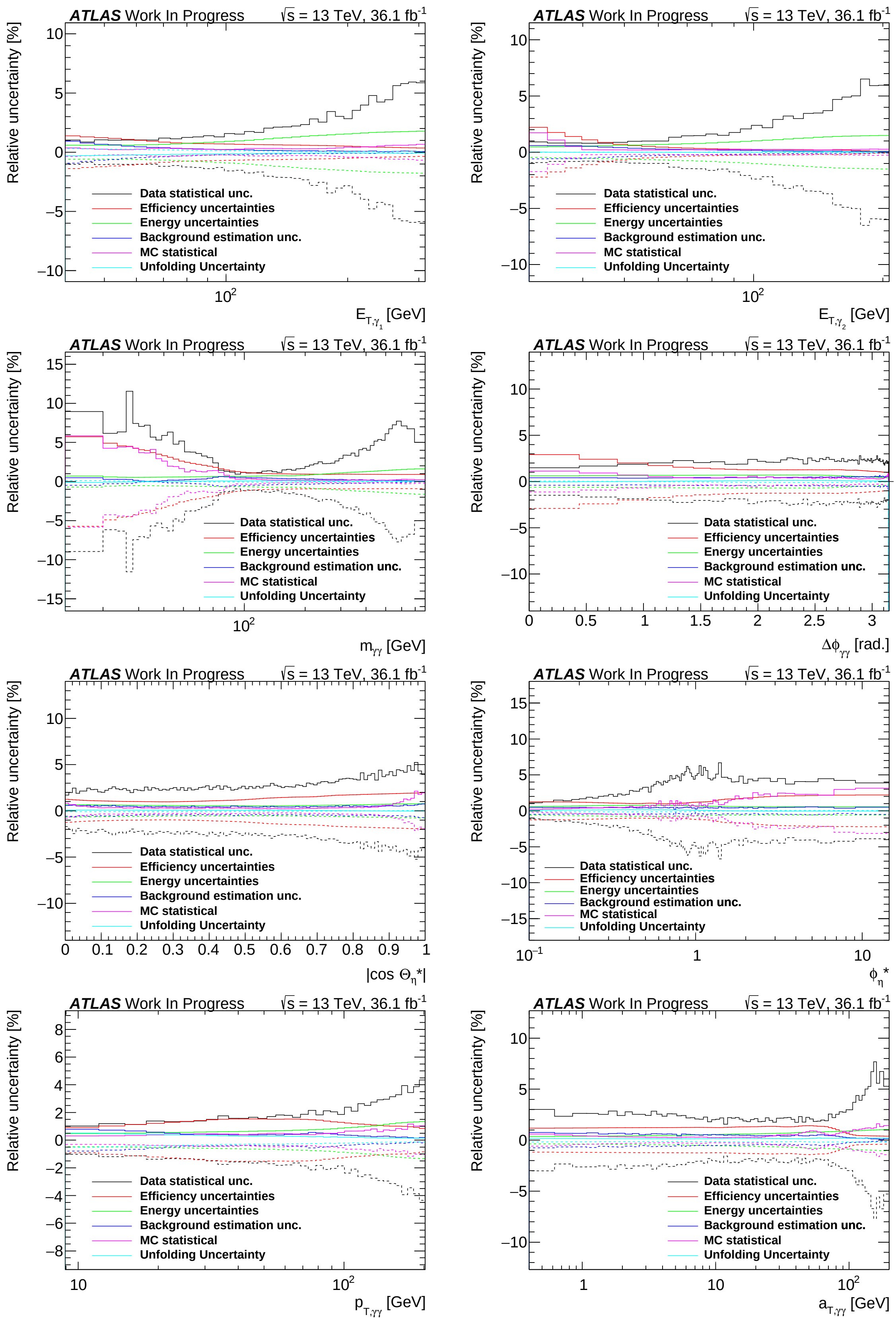


Figure 6.9: Summary of all uncertainties

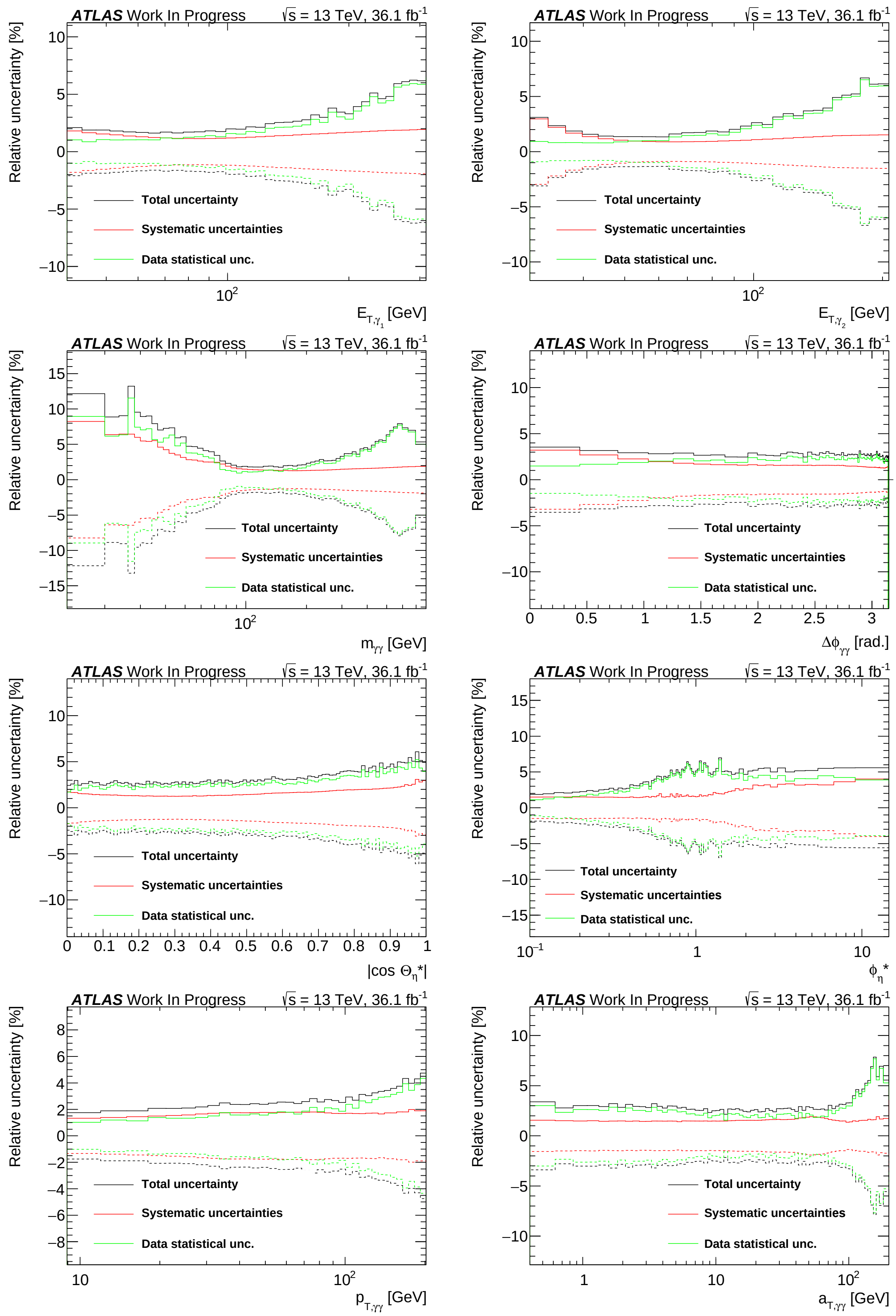


Figure 6.10: Systematic and statistical uncertainties combined

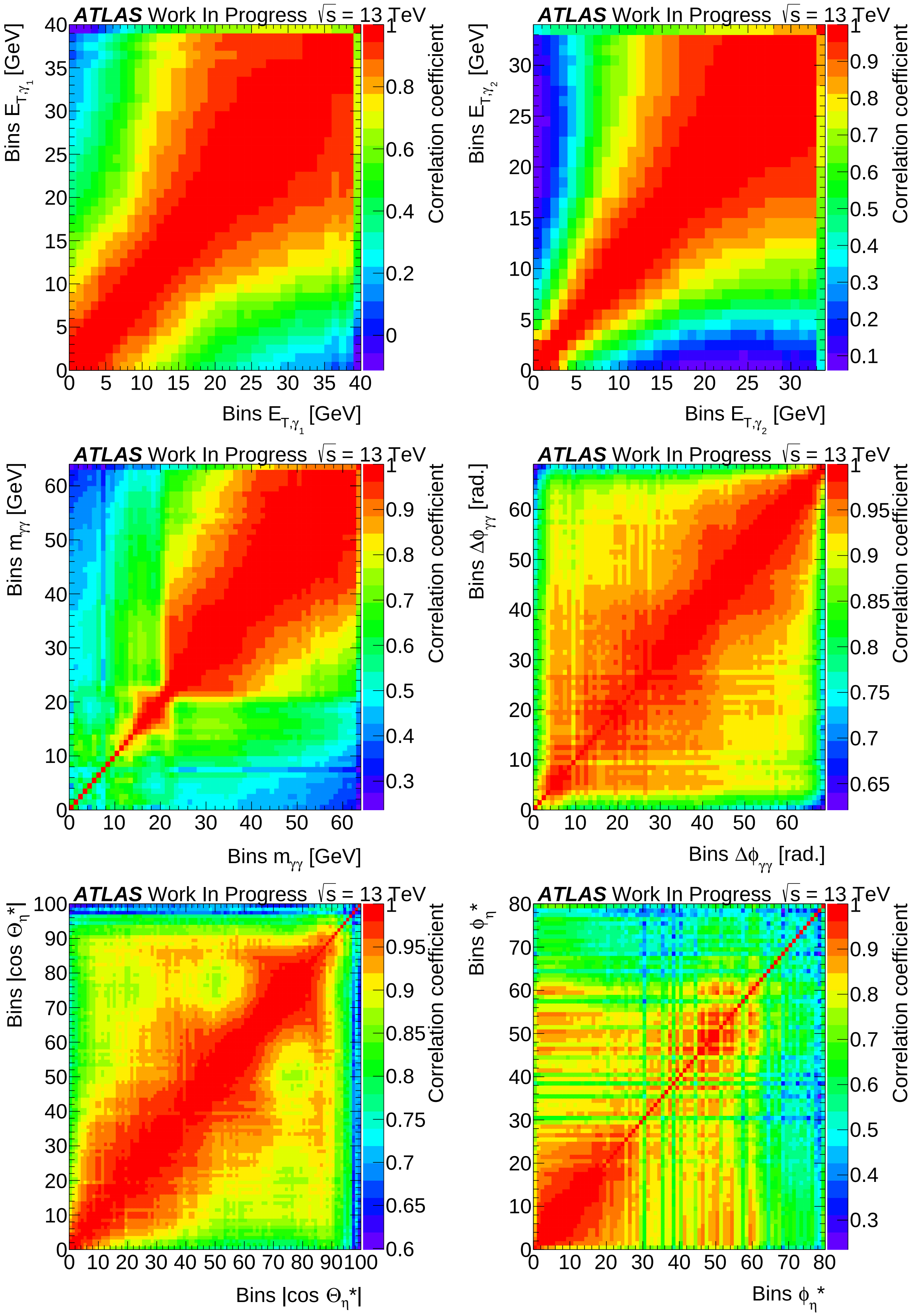


Figure 6.11: Correlation matrix of the total uncertainty for all observables

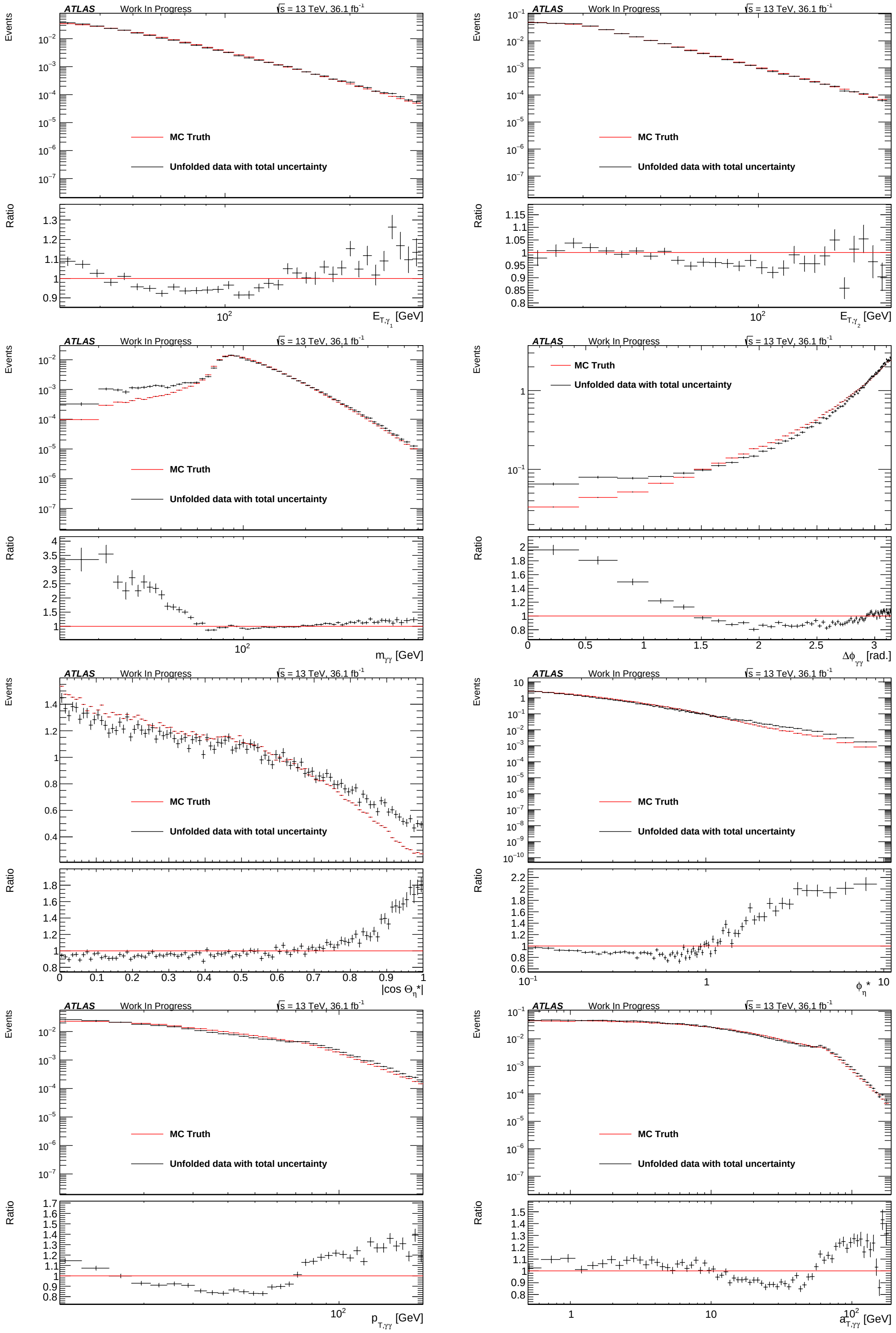


Figure 6.12: Comparison of LO predictions with unfolded data with all uncertainties

List of Figures

Bibliography

[1] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[2] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[3] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[4] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[5] The LHCb Collaboration. The lhcb detector at the lhc. *Journal of Instrumentation*, 3(08):S08005, 2008.

[6] The ALICE Collaboration. The alice experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08002, 2008.

[7] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, 2008.

[8] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[9] W. Bonivento. Overview of the atlas electromagnetic calorimeter. *Nuclear Physics B - Proceedings Supplements*, 78(1):176 – 181, 1999. Advanced Technology and Particle Physics.

[10] Search for the Standard Model Higgs boson in the diphoton decay channel with 4.9 fb-1 of ATLAS data at sqrt(s)=7TeV. Technical Report ATLAS-CONF-2011-161, CERN, Geneva, Dec 2011.

[11] Professor E. Rutherford F.R.S. Lxxix. the scattering of α and β particles by matter and the structure of the atom. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(125):669–688, 1911.

[12] Steven Weinberg. The Making of the Standard Model. *Eur. Phys. J. C*, 34(hep-ph/0401010):5–13. 21 p. ; streaming video, 2003.

[13] Francesco De Lorenzi and R Mc Nulty. Parton distribution function studies and a measurement of drell-yan produced muon pairs at lhcb. 01 2011.

[14] Martin Bessner. *Measurement of differential di-photon plus jet cross sections using the ATLAS detector*. Dissertation, Universität Hamburg, Hamburg, 2017. Dissertation, Universität Hamburg, 2017.

[15] G. Cowan. A survey of unfolding methods for particle physics. *Conf. Proc.*, C0203181:248–257, 2002. [,248(2002)].

[16] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nucl. Instrum. Meth.*, A372:469–481, 1996.

[17] G D'Agostini. Improved iterative bayesian unfolding. 10 2010.

[18] The ATLAS Collaboration. Measurements of integrated and differential cross sections for isolated photon pair production in pp collisions at $\sqrt{s} = 8$ TeV with the atlas detector. *Phys. Rev. D*, 95:112005, Jun 2017.

[19] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[20] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, pages 313–318, Geneva, 2011. CERN, CERN.

[21] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[22] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

Erklärung

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

Ken Kreul
Dresden, November 2018