FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Bachelor's thesis

Study of light-by-light scattering with the ATLAS Forward Proton (AFP) Detector at CERN

Tomáš Chobola

Department of Applied Mathematics
Supervisor: doc. Dr. André Sopczak

June 1, 2020

Acknowledgements

I would like to thank my supervisor doc. Dr. André Sopczak for the opportunity to collaborate on such interesting project, and for his patience and kind approach during the development of this thesis. Also I would like to thank my family and friends for supporting me during my studies.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as school work under the provisions of Article 60(1) of the Act.

In Prague on June 1, 2020 .....

Czech Technical University in Prague
Faculty of Information Technology

Citation of this thesis

Chobola, Tomáš. *Study of light-by-light scattering with the ATLAS Forward Proton (AFP) Detector at CERN*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2020.

Abstrakt

Tato práce zkoumá problém rozptylu světla světlem ve standardním modelu částicové fyziky a v jeho rozšíření, které předpověděli R. Peccei a H. Quinn. Pomocí Monte Carlo simulací porovnává fotony způsobenou produkci párů leptonů s produkcí hypotetické částice podobné axionu, která by rozptyl světla světlem mohla způsobovat. Dále využívá neparametrický Bayesovský přístup, konkrétně Gaussovský proces, k modelování pozadí založeného na hmotnosti roztříštěných fotonů z ultraperiferálních kolizí pozorované při experimentu ALTAS. Model byl následně použit na zakomponování simulovaného signálu reprezentující produkci částice podobné axionu, pro kterou byla dále určena pravděpodobnost její detekce detektorem AFP.

Klíčová slova částicová fyzika, simulace, analýza dat, vizualizace dat, regresní analýza, Gaussovský proces

Abstract

This thesis deals with the study of light-by-light scattering in the Standard Model of particle physics and in its extension predicted by Peccei–Quinn theory. It compares photon induced lepton pair productions with the hypothetical axion-like particle production, which could mediate light-by-light scattering, based on Monte Carlo simulations. Furthermore, it applies non-parametric Bayesian approach, the Gaussian Process, to model smooth background based on the invariant mass of scattered diphotons observed in ultraperipheral collisions with the ATLAS experiment. The estimated background is then used with simulated signal injection representing production of axion-like particle to determine the detection probability.

Keywords particle physics, simulation, data analysis, data visualisation, regression analysis, Gaussian Process

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the most significant technological accomplishments in recent history is the construction of Large Hadron Collider at CERN, that is capable of accelerating particles almost up to the speed of light, and subsequently cause them to collide, releasing immense amounts of energy enabling scientists to observe interactions that can help us understand the deep structure of space and time.

When two cars on the road collide, their kinetic energy needs to go somewhere, which leads to chassis deformation, shattering of windows and worrying sounds of metals bending. The same applies on particle level, but instead of shattering windows, the protons that are being collided scatter into quarks and gluons from which they are made of. This high energy process sometimes leads to creation of a new particle that only exists for a very brief period of time before it decays into other particles. This poses a problem of how to observe such particle. One of the options, which was for example applied in search for the Higgs boson, is to detect the particles into which the new particle decays. For this purpose, the ATLAS detector was installed at CERN that gathers properties of particles produced in proton collisions. Because the protons during the experiment collide approximately 600 million times per second, the amount of obtained data is very extensive. This information is then thoroughly analysed, and events, that meet a certain requirement, could indicate an observation of new physics. One of the approaches is to detect high energy photon pairs, which are emitted in every interaction at the LHC. Even though a photon itself is massless, the emitted pair can obtain invariant mass from the newly formed particle that was created in the proton collision and later decayed into the said photons, meaning that the invariant mass of the diphoton is the same as the mass of the produced particle. The invariant mass distribution of the photons often follows a smooth falling line. However, if conditions necessary for creating a new particle are met, a deviation from the smooth line would appear signifying increased number of events at particular mass caused by a particle production.

In order to be able to recognise these “bumps”, this smooth falling line needs to be modelled as precisely as possible. If it is not modelled correctly it can lead to false assumption of new physics discovery or lessen the importance of a deviation.

1.1 Purpose and structure of the thesis

In 2017 the ATLAS Forward Proton detector took large scale data from proton-proton collisions, including attributes of scattered diphotons. This thesis deals with the problem of estimating diphoton invariant mass distribution by using regression analysis methods, more precisely the widely used non-linear curve fitting algorithms and the Gaussian Process, powerful machine learning tool that is becoming a frequent method in particle physics. The purpose of this modelling is to then inject a signal signifying a hypothetical axion-like particle production that could be mediated by emitted photons in proton collisions and determine its observational probability and detection sensitivity. The results can then be applied in conducting an experiment that could lead to an axion discovery and therefore answer open questions in particle physics.

Chapter 2 introduces concepts and terms of particle physics that are required for understanding the work of this thesis. It is followed by Chapter 3 that describes the applied regression methods, the Levenberg–Marquardt algorithm and the Gaussian Process. Chapter 4 explains conducted simulator runs of event generator, that was used for determining the differences between photon induced production of axions and lepton pairs, which is closely related process that has already been observed. The results were then provided for higher precision external simulations and its outcomes were applied in regression analysis of the diphoton invariant mass distribution measured by the 2017 ATLAS experiment in Chapter 5.

Chapter 2

Preliminaries

In this chapter, the basic physics principles required for understanding the topic of this thesis are laid out. It starts with a brief introduction of the particles from the Standard Model of particle physics that are present in the following analysis and a set of processes in which they occur. Furthermore, it introduces the hypothetical axion-like particle and methods that could lead to its discovery.

2.1 Particles

A system defining the fundamental forces, particles and their interactions is encapsulated into the Standard Model of particle physics, which describes the building blocks of our universe. For the observation purposes of this thesis, it is necessary to introduce five particles [1] from the model.

Proton	Particle with one unit of electric charge and with mass of approximately one dalton.
Electron	Particle with one unit of negative electric charge and much lower mass than proton.
Heavy-ion	Particle heavier than proton with one or more units of electric charge.
Photon	Quantum of electromagnetic radiation, and elementary particle with no electric charge and zero mass.
Muon	Particle similar to the electron with one unit of negative electric charge, but with greater mass.

2.2 Ultraperipheral collisions

One of the specific kinds of interactions that particle physics explores, is an interaction where particles do not collide head-on, but get so close to each other (into the so-called interaction range) they interact through their electromagnetic fields made out of a cloud of virtual photons [2]. When these photons have high energy, they can decay into single particle or particle-antiparticle pair during the interaction. However, the newly produced particles can then quickly annihilate and produce two new photons, which is a very rare process called light-by-light scattering. This process is forbidden according to the classical theory of electrodynamics, and its evidence was difficult to find for decades, yet it was attained with the ATLAS detector at the Large Hydron Collider (LHC) in heavy-ion collisions. The ATLAS Collaboration reported [3] in Nature Physics in 2017 thirteen event candidates found in data recorded in 2015. At the Rencontres de Moriond conference in 2019, the ATLAS Collaboration reported [4] observation of 3.6 times more events in the 2018 heavy-ion run of the LHC compared to the previous run[1].

The analysis in this thesis explores the light-by-light scattering (also denoted as $\gamma\gamma \to \gamma\gamma$ where $\gamma\gamma$ is the photon pair) by comparing it to closely related particle-antiparticle pair productions, in which they do not decay into photons. More specifically, an electron-positron pair production $\gamma\gamma \to e^+e^-$ and a muon-antimuon pair production $\gamma\gamma \to \mu^+\mu^-$. Moreover, it focuses on predicting a hypothetical axion-like particle production, that is introduced in the following Section 2.3, which might be formed in this photon interaction and cause light-by-light as well.

The interest of recently proposed search methods is to observe the photon interactions in proton ultraperipheral collisions at CERN, where the photon pair is detected in the central detector and the intact protons, that are bent out of the beam, are tagged with the forward proton detectors at the LHC [5], which are placed symmetrically with the distance of 205 and 217 meters with respect to the interaction point. The tagging is based on observing protons that lost a fraction of their original momentum in the photon exchange, as the detectors are able to capture protons with relative energy loss between 2% and 10%. The energy loss of the protons is defined as $\xi_{\mathrm{AFP}} = \Delta E/E$, where ΔE is the amount of energy lost in the interaction, and E is the initial energy of the beam proton. Two types of tagging are considered, single tag and double tag. The single tag means that at least one of the two beam protons was detected with relative energy loss in the specific range and similarly, the double tag denotes detecting both beam protons.

Before experiments can be conducted, the probability of a process needs to be calculated to outline the observation prediction. This probability is defined as cross section $\hat{\sigma}$ and is dependent on various conditions, such as the energy

[1] Evidence for light-by-light was also reported before in experiments at lower energy.

of the colliding particles or the angle under which they interact. The cross section multiplied by the luminosity $\mathcal{L}$, the quantity that measures the ability of a particle accelerator to produce the required number of interactions, and the detector efficiency ε, gives a numerical estimate of an interaction rate $\mathcal{N}$ as

$$\mathcal{N} = \hat{\sigma} \cdot \mathcal{L} \cdot \varepsilon. \tag{2.1}$$

Apart from observing the event itself, various particle properties need to be considered. First of all, the particle energy E, calculated from its three-dimensional momentum (p_x, p_y, p_z) and mass m, where speed of light is set to $c = 1$, as

$$p = \sqrt{p_x^2 + p_y^2 + p_z^2}, \tag{2.2}$$

$$E = \sqrt{m^2 + p^2}. \tag{2.3}$$

When calculating the invariant mass m_p of the pair produced in photon interaction, whether it is the photon-photon pair, the electron-positron pair or the muon-antimuon pair, the energies of individual particles E_1, E_2 and their respective momentum p_1, p_2 are required as

$$m_p = \sqrt{(E_1 + E_2)^2 + (p_1 + p_2)^2}. \tag{2.4}$$

Even though photon is a massless particle, the invariant mass of the diphoton pair can reach high values depending on energy and momentum of its components.

The angle property of a particle relative to the beam axis denoted by θ is defined with particle energy E and z component of its three-momentum as

$$\theta = \arccos\left(\frac{p_z}{E}\right). \tag{2.5}$$

In particle physics, pseudorapidity denoted by η is used instead of θ

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.6}$$

The rapidity y is calculated with particle energy E and z component of its three-momentum as

$$y = \frac{1}{2}\ln\left(\frac{E \mp p_z}{E \pm p_z}\right). \tag{2.7}$$

Same as for the beam protons, the relative energy loss ξ can be calculated for the produced particles. The formula is based on the collision energy $\sqrt{s}$, particles mass m and pseudorapidity η. Specifically for produced photon-photon pair the formula can be altered by replacing the pseudorapidity η with the rapidity y, because photon is a massless particle

$$\xi = \frac{m}{\sqrt{s}} \cdot e^{\pm\eta}. \tag{2.8}$$

Moving particles can be traced with a Cartesian coordinate system for a three-dimensional space where X-axis points upwards while Z-axis follows the beam trajectory. The point of origin of the reference system is the interaction point. The side with positive values Z is defined as A-side, whereas the side with negative values of Z is defined as C-side. Therefore two values of relative energy loss ξ for the pair are calculated, one for the particle on the A-side and one for the particle on the C-side.

When the particles of a produced pair are scattered, their momentum perpendicular to the beam line can be measured. This transverse momentum p_T is defined by x and y components of their three-momentum as

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.9}$$

Furthermore, the measure of degree to which the paths of the scattered particles deviate from being coplanar, i.e. contained in single geometric plane, is defined as acoplanarity. It is denoted with x and y components of the three-momentum of the scattered particles together with their momentum p as

$$\mathcal{A} = 1 - \frac{|\phi|}{\pi}, \tag{2.10}$$

where

$$\phi = \arccos\left(\frac{p_{x_1}p_{x_2} + p_{y_1}p_{y_2}}{p_1 p_2}\right). \tag{2.11}$$

The physics analysis requirement on acoplanarity for the $\gamma\gamma \to \gamma\gamma$ process is to be less than 0.01.

2.3 Axion-like particle

An axion is a hypothetical particle, which has not been observed, yet it is a good candidate for explaining dark matter and could resolve current problems in quantum chromodynamics. The proposal of an axion particle is based on the idea by Peccei and Quinn [6] to explain conservation laws in physics. It

is estimated that the particle is produced in ultraperipheral collisions during a photon-photon interaction the same way as the electron-positron pair e^+e^- and the muon-antimuon pair $\mu^+\mu^-$, and then decays into a new pair of photons inducing the light-by-light scattering. The search for the particle can be conducted at the LHC, where it is possible to observe collisions with heavy-ions and protons. The heavy-ions can constrain ALP mass only to the range from 10 GeV to about 100 GeV [7]. To probe higher masses up to 2 TeV, which can increase the detection probability, the proton-proton collisions need to be explored. This was claimed as result of detailed study of possibilities regarding the axion-like particle search [8] presented by Baldenegroa, Fichetb et al., which determined that the search for an axion produced via light-by-light scattering at the LHC can be competitive in mass range between 0.5 TeV and 2 TeV in proton collisions, as it offers the highest observational sensitivity.

2.3.1 Searching for axion

One of the most important aspects of a search for a new particle is to know where to look. As particles produced in photon interaction have short lifetime, observing them is very challenging. However, the ATLAS detector can perceive photons that are produced in every proton collision at the LHC and as the photons, into which a newly created particle decays, inherit its mass, the properties of the particles can then be reconstructed upon their detection.

Usually, the invariant mass distribution of observed photons follows a smooth falling line called background, as the photons with higher invariant mass are more difficult to observe. This helps physicists to predict the behaviour of the photons excluding the production of a new particle. If the requirements necessary for creating such particle, for example an axion, are met, the number of events in which the photons inherited its mass will create a peak called signal in otherwise smooth distribution. This peak deviating from the background could signify an observation of new physics.

Observed data from an experiment is often not so smooth and it slightly varies from the predicted background. In order to claim a discovery the deviation needs to be significant enough, which reduces the probability that the divergence is just a statistical accident. The significance of the deviations is measured in standard deviation σ and if it exceeds 5σ in an experiment, a discovery could be claimed. The 5σ significance means that the probability of it being just an accident is 1 in 3.5 million.

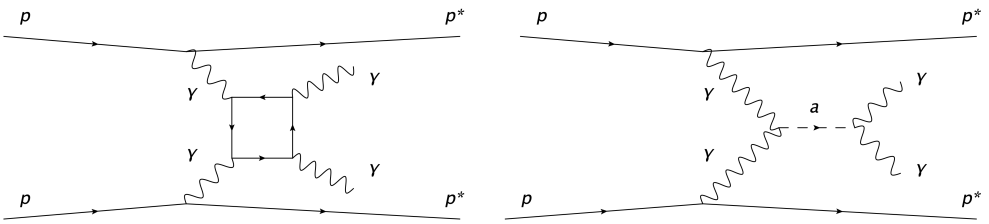

Figure 2.1: Left: Feynman diagram of light-by-light scattering in ultraperipheral proton collision. Right: Feynman diagram of ALP mediated light-by-light scattering in ultraperipheral proton collision.

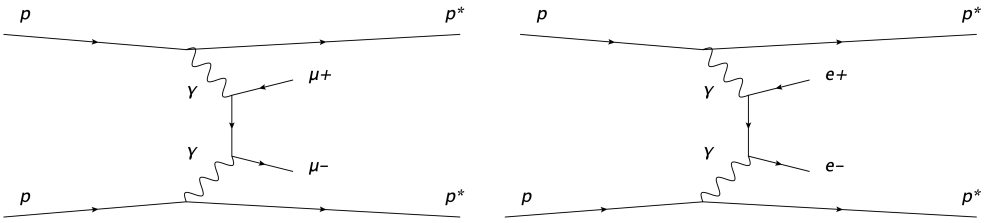

Figure 2.2: Left: Feynman diagram of muon-antimuon pair production from emitted photons in ultraperipheral proton collision. Right: Feynman diagram of electron-positron pair production from emitted photons in ultraperipheral proton collision.

Chapter 3

Regression analysis

Machine learning considers three different approaches to training: reinforcement learning, supervised learning and unsupervised learning. The reinforcement learning focuses on predicting certain outputs based on an environment in which a learning agent is set. Its guesses then receive a feedback based on how good they are, and therefore it forms its abilities by trial and error. Contrarily the task of supervised learning is described as learning by an example as data feeded into a model consists of input-output pairs, from which the abilities of a model are inferred. The unsupervised learning is characterised by its objective which is a search for patterns and connections between data points without any previous labels. Two main problems in supervised learning are classification and regression. Classification focuses on outputting discrete values, whereas in regression the main interest is in predicting continuous values. The latter problem corresponds to the present experiment analysis and smooth distribution modelling in the following Chapter 5, as it poses a problem of a search for relationship between dependent and independent variables, i.e. number of events and diphoton invariant mass.

This chapter serves as a guide and series of explanations of the methods used in the modelling. It starts with a description of the curve fitting problem in Section 3.1, which is one of the most frequent approaches to regression analysis. Then it proceeds with the exposition of methods that solve such problem, namely the gradient descent, the Gauss-Newton algorithm and its extension, the Levenberg–Marquardt algorithm, in Subsections 3.1.1, 3.1.2 and 3.1.3, respectively.

Furthermore, it introduces the Bayesian inference in Section 3.2 and its use in obtaining predictions conditioned on observed data. On this foundation, the non-parametric Bayesian method used in regression analysis called Gaussian Process is explored in Section 3.3. This framework is gaining in popularity and becoming a frequent modelling method in particle physics and was used as the main methodology for diphoton invariant mass background modelling described later in this thesis.

3.1 Curve fitting

Regression analysis is a statistical expertise that approximates a connection between dependent and independent variables. Suppose an observed point y at location $[x_1, x_2, \ldots, x_n]^\top$ and a task of finding values $\boldsymbol{w} = [w_0, w_1, \ldots, w_n]^\top$, so the value y can be estimated by a linear function as

$$y(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{x}, \tag{3.1}$$

where $\boldsymbol{x} = [1, x_1, x_2, \ldots, x_n]^\top$. The parameters $\boldsymbol{w}$, also known as weights, define the relationship between the depended and independent variables. To increase the flexibility of the model a different function, such as polynomial, can be considered as well.

To fit the function to locations $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m]^\top$ and observed values $\boldsymbol{Y} = [y_1, y_2, \ldots, y_m]^\top$, the parameters need to be optimised using a learning method based on minimising or maximising a certain mathematical criterion. It is highly probable that the estimated function will not duplicate the measured observations, as the function returns only an approximation of y_i. One of the frequently used optimisation methods focusing on minimising the residuals denoted as

$$r_i(\boldsymbol{w}) = y_i - y(\boldsymbol{x}_i, \boldsymbol{w}) \tag{3.2}$$

is the least squares method [9]. Its objective is to find such parameters $\boldsymbol{w}$ that the scaled sum of squared residuals is minimised over the whole set of observed points

$$\phi(\boldsymbol{w}) = \frac{1}{2} \sum_{i=1}^{m} (r_i(\boldsymbol{w}))^2. \tag{3.3}$$

The minimum of $\phi(\boldsymbol{w})$ can then be found by using the following optimisation methods.

3.1.1 Gradient descent

The gradient descent is an optimisation algorithm for finding a local minimum of a differentiable function. It takes the advantage of a gradient, which is a vector-valued function $\nabla f : \mathbb{R}^m \to \mathbb{R}^m$. The elements of the gradient are

partial derivatives of a function f denoted as

$$\nabla f(\boldsymbol{w}) = \begin{bmatrix} \frac{\partial f}{\partial w_0} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{bmatrix}. \tag{3.4}$$

The gradient represents a direction and a rate of fastest increase and it is a tangent vector in respect to the plane that goes through the point for which the function f is evaluated. If a function has an extreme in a certain point, then its gradient at that point is equal to zero. In case of the least squares function 3.3 the gradient is denoted as [10]

$$\nabla\phi(\boldsymbol{w}) = \frac{\partial\phi(\boldsymbol{w})}{\partial\boldsymbol{w}} = \frac{\partial}{\partial\boldsymbol{w}}\frac{1}{2}\sum_{i=1}^{m}(r_i(\boldsymbol{w}))^2 = \sum_{i=1}^{m} r_i(\boldsymbol{w})\nabla r_i(\boldsymbol{w}) = (\boldsymbol{J}(\boldsymbol{w}))^\top \boldsymbol{r}(\boldsymbol{w}), \tag{3.5}$$

where $\boldsymbol{r}(\boldsymbol{w}) = [r_1(\boldsymbol{w}), \ldots, r_m(\boldsymbol{w})]^\top$ and

$$\boldsymbol{J}(\boldsymbol{w}) = \left[\frac{\partial r_i}{\partial w_j}\right]_{i=1,\ldots,m;\ j=0,\ldots,n} = \begin{bmatrix} (\nabla r_1(\boldsymbol{w}))^\top \\ (\nabla r_2(\boldsymbol{w}))^\top \\ \vdots \\ (\nabla r_m(\boldsymbol{w}))^\top \end{bmatrix}, \tag{3.6}$$

the optimum can be found solving the normal equation

$$\nabla\phi(\boldsymbol{w}) = (\boldsymbol{J}(\boldsymbol{w}))^\top \boldsymbol{r}(\boldsymbol{w}) = \boldsymbol{0}, \tag{3.7}$$

and therefore in case of a linear function f, it poses a problem of solving a series of linear equations. In case of finding a minimum of a non-linear function an iterative approach needs to be considered [11], where the parameters $\boldsymbol{w}$ are updated in each iteration by making steps in the opposite direction of the gradient as

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - \alpha(\boldsymbol{J}(\boldsymbol{w}^{(k)}))^\top \boldsymbol{r}(\boldsymbol{w}^{(k)}), \tag{3.8}$$

where α is a learning rate that can depend on k, which is the iteration index. The iteration stops when a termination condition is satisfied. Such condition can be maximum number of iterations or when the improvement drops under a certain threshold.

Gradient descent is a simple and intuitive way of searching for a minimum,

but suffers from convergence problems [12] as it is not resistant to gentle gradients, where the algorithm takes only small steps and becomes extremely slow. Assuming a surface in form of a valley (such as the Rosenbrock function [13]), the algorithm will quickly descent from the walls of the valley, but then will do exceptionally small steps in the direction of the valley.

3.1.2 The Gauss-Newton algorithm

To solve the issue of the gradient descent method an additional information about the function is needed. This is achieved by using the Newton's method for optimisation, that iteratively approximates the curvature [14] of the surface by fitting a paraboloid to an analysed point, and then proceeds to the minimum of that paraboloid. The recursive formula for Newton's method to find a minimum [15] of function f based on parameters $\boldsymbol{w}$ is denoted as

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - (\nabla^2 f(\boldsymbol{w}^{(k)}))^{-1} \nabla f(\boldsymbol{w}^{(k)}), \tag{3.9}$$

where $\nabla^2 f(\boldsymbol{w}^{(k)})$ is the Hessian matrix with second partial derviatives of f denoted as

$$\nabla^2 f(\boldsymbol{w}) = \boldsymbol{H}_f(\boldsymbol{w}) = \begin{bmatrix} \frac{\partial^2 f}{\partial w_0^2} & \frac{\partial^2 f}{\partial w_0 \partial w_1} & \cdots & \frac{\partial^2 f}{\partial w_0 \partial w_n} \\ \frac{\partial^2 f}{\partial w_1 \partial w_0} & \frac{\partial^2 f}{\partial p_1^2} & \cdots & \frac{\partial^2 f}{\partial w_1 \partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial w_n \partial w_0} & \frac{\partial^2 f}{\partial w_n \partial w_1} & \cdots & \frac{\partial^2 f}{\partial w_n^2} \end{bmatrix}. \tag{3.10}$$

The Hessian matrix of the scaled sum of squared residuals 3.3 is then denoted as [16]

$$\boldsymbol{H}_\phi(\boldsymbol{w}) = \sum_{i=1}^{m} \nabla r_i(\boldsymbol{w}) (\nabla r_i(\boldsymbol{w}))^\top + \sum_{i=1}^{m} r_i(\boldsymbol{w}) (\nabla^2 r_i(\boldsymbol{w}))^\top \tag{3.11}$$

$$= (\boldsymbol{J}(\boldsymbol{w}))^\top \boldsymbol{J}(\boldsymbol{w}) + \sum_{i=1}^{m} r_i(\boldsymbol{w}) (\nabla^2 r_i(\boldsymbol{w}))^\top. \tag{3.12}$$

The recursive formula 3.9 can be applied to the least squares problem using the already obtained gradient 3.7 as

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - (\boldsymbol{H}_\phi(\boldsymbol{w}^{(k)}))^{-1} (\boldsymbol{J}(\boldsymbol{w}^{(k)}))^\top \boldsymbol{r}(\boldsymbol{w}^{(k)}). \tag{3.13}$$

The Gauss-Newton optimisation method, a specific version of the Newton approach for minimising the least squares function 3.3, approximates the Hessian matrix as $\boldsymbol{H}_f(\boldsymbol{w}) \approx (\boldsymbol{J}(\boldsymbol{w}))^\top \boldsymbol{J}(\boldsymbol{w})$ by eliminating its second term [17]. This choice is appropriate near the solution, because the residuals are becoming negligible [18], and therefore it influences the Hessian matrix in an insignificant way. It also means that the computationally complex process of finding the second derivatives is not required. The recursive formula then becomes

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - ((\boldsymbol{J}(\boldsymbol{w}^{(k)}))^\top \boldsymbol{J}(\boldsymbol{w}^{(k)}))^{-1} (\boldsymbol{J}(\boldsymbol{w}^{(k)}))^\top \boldsymbol{r}(\boldsymbol{w}^{(k)}). \tag{3.14}$$

The Gauss-Newton optimisation brings good results and converges faster then the gradient descent if the initial guess is near an extreme, but otherwise behaves rather wildly [19].

3.1.3 The Levenberg–Marquardt algorithm

A complementary combination of the gradient descent and the Gauss-Newton methods is offered by the Levenberg–Marquardt algorithm [20, 21]. It incorporates the process of updating the parameters by making steps in the steepest-descent direction from the gradient descent, and also the approximation of a paraboloid near an optimum to increase the convergence from the Gauss-Newton algorithm [22]. The iterative algorithm introduces a non-negative damping parameter λ that controls which method will be used in a following iteration as

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - ((\boldsymbol{J}(\boldsymbol{w}^{(k)}))^\top \boldsymbol{J}(\boldsymbol{w}^{(k)}) + \lambda \boldsymbol{I})^{-1} (\boldsymbol{J}(\boldsymbol{w}^{(k)}))^\top \boldsymbol{r}(\boldsymbol{w}^{(k)}), \tag{3.15}$$

where $\boldsymbol{I}$ is the identity matrix.

High values of λ outweigh the term $(\boldsymbol{J}(\boldsymbol{w}))^\top \boldsymbol{J}(\boldsymbol{w})$, and therefore the algorithm behaves as the gradient descent. On the opposite, low values of λ substantially increase the significance of the term and push the Gauss-Newton approach in the forefront [23]. The selection of the damping parameter λ significantly influences the behaviour of the algorithm, and therefore it is crucial to set certain rules that can be followed and lead to optimal results. Starting with higher values is a robust strategy that ensures that $\phi(\boldsymbol{w}^{(k+1)}) < \phi(\boldsymbol{w}^{(k)})$ (especially for sufficiently large values) causing the descent in the steepest direction [24]. Depending on the size of $\Delta\boldsymbol{w}^{(k)} = \boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)}$, the parameter λ can be altered to increase the importance of the Gauss-Newton step in the search for an optimum. Once the set of parameters starts changing slowly, λ is reduced by a factor of $1/\nu$, where $\nu > 1$ is the Marquardt parameter. If the reduction results in an unsuccessful step, λ is increased by ν [25].

The Levenberg–Marquardt algorithm implemented in the SciPy package [26] was used for purposes of this thesis.

3.2 Bayesian learning

Approaches described in the previous Section 3.1 deal only with scalar values, they do not take any subjective belief in the initial values into account and also lack the probability in the modelled results. These requirements are met by the Bayesian inference, a learning method that updates previous beliefs based on observed data.

This statistical concept deals with a prior expert knowledge about certain quantities that are determined before any data on which a model can learn has been seen. With this prior knowledge defined as probability distributions a likelihood of how well the quantities define the observed data can be obtained. Using the Bayesian inference, the prior knowledge can be updated based on the seen data and turned into posterior probability distributions. The Bayesian inference [27] is defined as

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{marginal likelihood}}. \tag{3.16}$$

Assuming observed data $\mathcal{D}$ and a parameter ω of some parametric model, with prior probability distribution $p(\omega)$, the likelihood of the data given the parameter is denoted as $p(\mathcal{D}|\omega)$. Furthermore, the marginal likelihood evaluates the model evidence and guarantees that the posterior is a valid probability, and is denoted as

$$p(\mathcal{D}) = \int p(\mathcal{D}|\omega)p(\omega)\ d\omega. \tag{3.17}$$

Therefore the posterior distribution of parameter ω is defined as

$$p(\omega|\mathcal{D}) = \frac{p(\omega)p(\mathcal{D}|\omega)}{p(\mathcal{D})}. \tag{3.18}$$

Since the posterior is a probability distribution, the result is not a single value, but rather a confidence in the value ω. Once new data is observed, posterior becomes prior and using the Bayesian inference it is once more updated to become posterior again. The posterior can be used to predict unseen data $\hat{\mathcal{D}}$ as well by averaging over all the possible values of ω as

$$p(\hat{\mathcal{D}}|\mathcal{D}) = \int p(\hat{\mathcal{D}}|\omega)p(\omega|\mathcal{D})\ d\omega. \tag{3.19}$$

When compared to the approaches described in Section 3.1 it shows a strong contrast as instead of returning a fixed set of predictions the Bayesian approach returns a set of probability distributions expressing the confidence in each prediction. Even though the formula of Bayesian inference seems rather

simple the computational process can be very demanding for more complex models [28].

3.3 Gaussian Process

The methods described so far are build on a priori defined functional form and optimisation of its parameters. The choice of a functional form is crucial for describing the data and often ad-hoc solutions have no connection to the area in which they are being used. Assuming the problem is mapped to a completely unknown function and there is little or no information about the functional form, the fitting depends on assumptions. The parametric functions also lack the pliability needed for observations in particle physics, and therefore a finer and more flexible way of describing the data needs to be considered.

The Gaussian Process [29] does not use a fixed parametric function to describe the data, but instead it lets the data to pick the functional form itself, and therefore it is a powerful tool in terms of machine learning. The process is formally defined as an infinite collection of random functions, where any finite subset of which creates a joint Gaussian distribution.

The Gaussian distribution of a set of n random functions indexed by a set of indices $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]^\top$ is specified by a mean vector $\boldsymbol{\mu}$ and a covariance matrix Σ as

$$\boldsymbol{f} = [f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n)]^\top \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \tag{3.20}$$

whereas the Gaussian Process is completely defined by a mean function

$$m(X) = \mathbb{E}[f(X)] \tag{3.21}$$

and a covariance function

$$k(X, X') = \mathbb{E}[(f(X) - m(X))(f(X' - m(X'))] \tag{3.22}$$

as

$$\boldsymbol{f} \sim \mathcal{GP}(m(X), k(X, X')). \tag{3.23}$$

Therefore the Gaussian Process can be described as a Gaussian distribution with an infinitely long mean vector and an infinite by infinite covariance matrix. The mean function describes the mathematical expectation of a function f at an input $\boldsymbol{x}$ and usually, for model simplicity, is set to zero $m(\boldsymbol{x}) = 0$. The result of the Gaussian Process depends heavily on the covariance func-

tion [30], but in regions with missing data the predictions will fall towards the mean, therefore the choice of a particular mean function is justified when there is an estimation or observation of how the distribution should behave. For purposes of this thesis, the function used for predicting photon-related backgrounds [31] was selected as the mean function

$$m(x) = p_0 \times \psi(x)^{p_1 + p_2 \log \psi(x)} \times \left(1 - \frac{1}{1 + e^{\frac{\psi(x) - p_3}{p_4}}}\right), \tag{3.24}$$

where $(p_0, p_1, p_2, p_3, p_4)$ are free parameters defined during the fit of the data (for example with the Levenberg–Marquardt method as described in Subsection 3.1.3 above) and $\psi(x) = \frac{x}{\sqrt{s}}$ is based on the centre-of-mass collision energy $\sqrt{s}$.

The covariance function evaluates each pair of arguments $\boldsymbol{x}$ and $\boldsymbol{x}'$ to $\mathbb{R}$ and creates a covariance matrix describing the influence of points on each other. It is important to note that the covariance function is only specified by the input locations and not by the observed values. For purposes of this thesis the kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ has to represent the physics behind the experiment in an intuitive way, where the mass resolution is not constant and is captured by a linear dependence. This varying length scale defined as $l(x) = bx + c$ is described by the Gibbs kernel [32] as

$$k(x, x') = A\sqrt{\frac{2l(x)l(x')}{l(x)^2 + l(x')^2}} e^{\frac{-(x-x')^2}{l(x)^2 + l(x')^2}}, \tag{3.25}$$

where only scalar inputs are considered for simplicity of the exposition. The initial hyperparameters (A, b, c) are defined during the fit of the data as described below in Subsection 3.3.2. This approach was successfully applied in smooth background modelling and search for dijet resonances at the Large Hadron Collider [33], therefore it creates a strong foundation for purposes of this thesis.

Based on the defined mean and covariance functions, the Gaussian Process model can sample prior function values. To obtain the posterior values, the model needs to conditioned on training data using the Bayesian inference.

3.3.1 Regression with Gaussian Process

Assuming the Gaussian Process is used in regression and there is a function f that given an input $\boldsymbol{x}$ returns a scalar value. Furthermore, there are noisy observations $Y = [y_1, y_2, ..., y_n]^\top$ at input locations $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n]^\top$. The

prior functional values at inputs X are denoted as

$$p(\boldsymbol{f}|X) = \mathcal{GP}(m(X), k(X, X')) \tag{3.26}$$

and the objective is to find a posterior distribution $p(\boldsymbol{f}|X, Y)$ of the functional values given the observed data. When working with real data, the observed values are often corrupted by a noise and therefore this needs to be quantified. Assuming the noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ follows a Gaussian distribution, set of noisy targets Y is then described as

$$Y = f(X) + \epsilon. \tag{3.27}$$

As the targets Y and function values $\boldsymbol{f}$ come from the same model, they create a joint Gaussian distribution denoted as

$$\begin{bmatrix} Y \\ \boldsymbol{f} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma^2 \boldsymbol{I} & K(X,X) \\ K(X,X) & K(X,X) \end{bmatrix} \right), \tag{3.28}$$

where $K(X,X)_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{I}$ is the identity matrix. By conditioning the multivariate Gaussian distribution, the posterior distribution of the functional values is obtained as [29]

$$p(\boldsymbol{f}|X, Y) = \mathcal{N}\left(\mathbb{E}[p(\boldsymbol{f}|X, Y)], \mathbb{V}[p(\boldsymbol{f}|X, Y)]\right), \tag{3.29}$$

where

$$\mathbb{E}[p(\boldsymbol{f}|X, Y)] = m(X) - K(K + \sigma^2 \boldsymbol{I})^{-1}(Y - m(X)), \tag{3.30}$$

$$\mathbb{V}[p(\boldsymbol{f}|X, Y)] = K - K(K + \sigma^2 \boldsymbol{I})^{-1} K \tag{3.31}$$

and $K = K(X, X)$.

Regression problems are not based only on input locations X of the observed data, but include predictions of functional values at new input locations X^* as well. This is achieved by the same approach of relying on Gaussian conditioning, because the values that need to be obtained are from the same model as the observed ones

$$\begin{bmatrix} Y \\ \boldsymbol{f}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X^*) \end{bmatrix}, \begin{bmatrix} K(X,X) + \sigma^2 \boldsymbol{I} & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \right). \tag{3.32}$$

The posterior distribution $p(\boldsymbol{f}^*|X, Y, X^*)$ of the functional values on new lo-

cations X^* is obtained as

$$p(\boldsymbol{f}^*|X,Y,X^*) = \mathcal{N}\left(\mathbb{E}[p(\boldsymbol{f}^*|X,Y,X^*)], \mathbb{V}[p(\boldsymbol{f}^*|X,Y,X^*)]\right), \tag{3.33}$$

where

$$\mathbb{E}[p(\boldsymbol{f}^*|X,Y,X^*)] = m(X^*) - K(X^*,X)(K(X,X)+\sigma^2\boldsymbol{I})^{-1}(Y-m(X)), \tag{3.34}$$

$$\mathbb{V}[p(\boldsymbol{f}^*|X,Y,X^*)] = K(X^*,X^*) - K(X^*,X)(K(X,X)+\sigma^2\boldsymbol{I})^{-1}K(X,X^*), \tag{3.35}$$

and the covariance matrices are defined as

$$K(X^*,X)_{i,j} = k(\boldsymbol{x}_i^*,\boldsymbol{x}_j), \quad K(X,X^*)_{i,j} = k(\boldsymbol{x}_i,\boldsymbol{x}_j^*),$$

$$K(X^*,X^*)_{i,j} = k(\boldsymbol{x}_i^*,\boldsymbol{x}_j^*).$$

Notice that the predicted distributions are noise free as they are related only to functional values $\boldsymbol{f}^*$. To make predictions for Y^* at locations X^*, the covariance matrix $K(X^*,X^*)$ needs to be altered to include the Gaussian representation of the noise as $K(X^*,X^*)+\sigma^2\boldsymbol{I}$.

3.3.2 Learning with Gaussian Process

The previous section describes the process of fitting the data into a model that is already optimised. Recalling the mean function 3.24 and the covariance function 3.25, there are sets of hyperparameters that need to be found in order for the Gaussian Process to work. To make the Gaussian Process fully Bayesian a prior on the hyperparameters $\boldsymbol{\omega}$ can be set and using the inference the posterior distribution of the functional values can be obtained while averaging over the set of hyperparameters $\boldsymbol{\omega}$ as

$$p(\boldsymbol{f}|X,Y) = \int p(\boldsymbol{f}|X,Y,\boldsymbol{\omega})p(\boldsymbol{\omega})\, d\boldsymbol{\omega}. \tag{3.36}$$

However, because of the fact that mean and covariance functions are often complex, solving the integral can be very computationally demanding. To solve this, a non-Bayesian approach for parameter optimisation is considered by maximising the likelihood $p(Y|X,\boldsymbol{\omega})$ of the observed data given the model. This likelihood is obtained by averaging over the distribution of functional

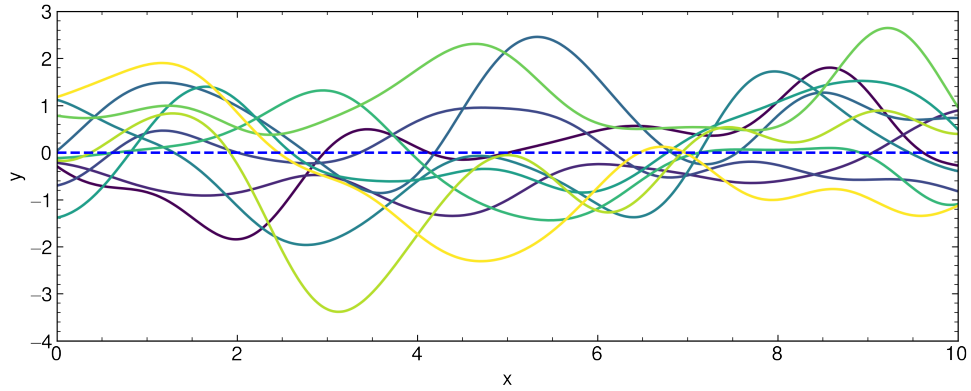

Figure 3.1: An example of randomly sampled functional values from a Gaussian Process. For exposition purposes, this particular model uses simple mean function $m(\boldsymbol{x}) = 0$ and radial basis function as kernel $k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{x}')^2}{2l^2}\right)$, where l is a free parameter. The functional forms are not conditioned on any data.

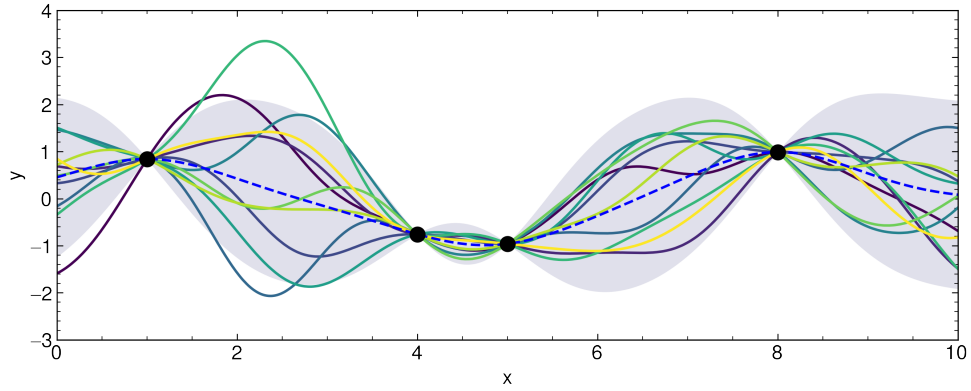

Figure 3.2: Assuming four observed values at input locations $\boldsymbol{x}^* = [1, 4, 5, 8]^\top$ the Gaussian Process from Figure 3.1 can then be conditioned and create posterior predictions. Those observations are represented by the black dots, the blue line is the altered mean and the grey area represents the 2σ uncertainty. The sampled functional forms are then based on the newly observed data points.

values $\boldsymbol{f}$ as

$$p(Y|X,\boldsymbol{\omega}) = \int p(Y|X,\boldsymbol{\omega},\boldsymbol{f})p(\boldsymbol{f}|X,\boldsymbol{\omega})\, d\boldsymbol{f}, \tag{3.37}$$

where $p(Y|X,\boldsymbol{\omega},\boldsymbol{f})$ is the likelihood of the data given the functional values $\boldsymbol{f}$ and $p(\boldsymbol{f}|X,\boldsymbol{\omega})$ is the prior distribution on the functional values $\boldsymbol{f}$. The explicit form of the integral can be obtained since the distributions from $\boldsymbol{f}$ are Gaussian, as well as the assumed noise. The log marginal likelihood is then defined as

$$\log p(Y|X,\boldsymbol{\omega}) = -\frac{1}{2}Y_\bullet^\top K^{-1}Y_\bullet - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi, \tag{3.38}$$

where $K = K(X,X) + \sigma^2\boldsymbol{I}$, $Y_\bullet = Y - m(x)$,

$-\frac{1}{2}Y_\bullet^\top K^{-1}Y_\bullet$	is the measure of how well the current model explains the dependent variable,
$-\frac{1}{2}\log\|K\|$	is the model complexity penalization term,
$-\frac{n}{2}\log 2\pi$	is a normalization constant [34].

The objective is then to maximise it using an optimisation method such as the gradient ascent, which requires an initial set of hyperparameters. A common practice in Gaussian Process research is to randomly sample initial hyperparameters from an uniform distribution using a prior knowledge [35, 36], proceed with the optimisation, and then compare their respective likelihoods. The best result is then selected [37].

Optimising the values by maximising the likelihood often leads to a local extreme, because finding the global extreme is rather difficult as the likelihood surface can be highly multimodal. Every extreme that is found corresponds to one potential interpretation of the model, and therefore multiple optimisation runs are recommended to get the best possible estimate.

3.3.3 Implementation of Gaussian Process

The information from preceding sections gives the basics needed for the model creation. As stated in the previous sections the hyperparameters $\boldsymbol{\omega}_m$ of the mean function and the hyperparameters $\boldsymbol{\omega}_c$ of the covariance function need to be optimised to maximize the log marginal likelihood. For purposes of this thesis the Minuit interface [38] was used to minimise the negative log marginal likelihood 3.38. The interface searches for a minimum in an user-defined func-

tion, hence the minimised function needs to be negative log marginal likelihood. As described in Subsection 3.3.2, multiple runs of the hyperparameter search were initiated. The initial values were randomly generated in each run from a predefined uniform distribution, and then optimised to find a local minimum. The most promising result was then selected and used in the modelling in Chapter 5. The model implementation shown below uses Cholesky decomposition of covariance matrix $K(X,X) + \sigma_\epsilon^2 \boldsymbol{I}$ which is the most computationally demanding process of the modelling with $\mathcal{O}(n^3)$ operations and storage complexity $\mathcal{O}(n^2)$, where n is the number of elements in X.

1: **input**: X (input), Y (observed values), k (covariance function),
m (mean function), σ^2 (variance), X^* (input for prediction)
2: $L := \text{cholesky}(K(X,X) + \sigma^2 \boldsymbol{I})$
3: $\boldsymbol{\alpha} := L^\top \setminus (L \setminus (Y - m(X)))$
4: $Y^* := m(X^*) + K(X,X^*)^\top \boldsymbol{\alpha}$
5: $\boldsymbol{v} := L \setminus K(X,X^*)$
6: $\mathbb{V}[Y^*] := K(X^*,X^*) - \boldsymbol{v}^\top \boldsymbol{v}$
7: $\log p(Y|X) := -\frac{1}{2}(Y - m(X))^\top \boldsymbol{\alpha} - \Sigma_i \log L_{ii} - \frac{n}{2} \log 2\pi$
8: **return**: $\boldsymbol{f}^*$ (mean), $\mathbb{V}[\boldsymbol{f}^*]$ (variance), $\log p(Y|X)$ (LML)

Algorithm 5.1: The following implementation [29] returns the predicted distributions based on observed values Y at input locations X and the log marginal likelihood. It works with already optimised hyperparameters for mean and covariance functions. The required matrix inversion is computed using the Cholesky factorization, as it is fast, numerically extremely stable and able to solve linear systems with symmetric positive definite coefficient matrix.

Chapter 4

Simulation

This chapter introduces the Monte Carlo event generator SuperChic-3 that is used for simulating production processes in proton and heavy-ion collisions and was applied for purposes of this thesis, as the productions of photon-photon, electron-positron, muon-antimuon pairs and axion-like particle were simulated. The first section introduces the generator together with an overview of its control file that defines the required initial parameters of the simulation. It also describes the output file that contains information about simulated events.

The second section shows analysis of the simulated events. It begins with measurements of proton tagging probabilities for each simulated production process, as defined in Section 2.2, and their corresponding cross sections. Furthermore, the properties of the generated particles are explored and the differences between lepton pair productions and axion production are analysed. A validation webpage summarising the observations from the simulations was published on the central scientific results page of the ATLAS Collaboration in the internal section for the members of the collaboration.

4.1 SuperChic-3

Monte Carlo event generators are integration tools that are essential for experimental analysis as they provide estimations of the experiment outcome by simulating the final states of high-energy collisions [39]. The generators use Monte Carlo method that relies on a class of computational algorithms which are based on repeated random sampling from a set of probability distributions. The generators are able to compute the kinematics of a process and calculate the properties of the present particles. The main objective of the generator is to produce detailed events as if they were observed by a perfect detector. All aspects of the events are simulated which provides high comparability to the real data. Fortran based Monte Carlo event generator SuperChic-3 [40] was used for the purposes of this thesis, as it provides numerical results of

Table 4.1: Overview of the PDG particle codes used for analysis of the output file.

PDG code	Particle
2012	proton
22	photon
13	muon
11	electron
93	parton system in independent fragmentation
90	intermediate pseudoparticle

the processes based on the Standard Model of particle physics and the model predicting the existence of an axion–like particle as well.

The program is operated through input file `input.DAT` found in the SuperChic's directory and consists of a set of adjustable parameters defining the simulated process. By running `init` executable found in the `bin` directory, files controlling the simulation based on the parameters from the input file are created. They are then used in the subsequent execution of file `superchic` responsible for the simulation, generating events related to the proir settings.

The information about the generated events outputted by SuperChic-3 is stored in Les Houches Event File (LHEF) format, which consists of two blocks. The first one covers initialisation information, most importantly beam identities (i.e. colliding particles) and its energies, and the probability density function sets on which the simulated events are based. The second block contains information about each generated event. The most significant information for the following analysis is the number of particles present in the event, their three-momentum, energy and mass. For each event the output file labels present particles with the Particle Data Group (PDG) particle codes [41] and for purposes of this thesis the knowledge of six PDG codes, that are shown in Table 4.1, labeling photons, electrons, muons and axions is required. (In the present analysis the parton system in independent fragmentation labeled with code 93 represents generated diphoton, muon-antimuon and electron-positron pair, and the intermediate pseudoparticle with code 90 labels the axion–like particle.)

4.1.1 Events simulation

The initial focus of the work with the generator was directed on cross section calculations of the $\gamma\gamma \to \gamma\gamma$ process in the Standard Model of particle physics. For this purpose, SuperChic's input file set to light-by-light scattering in heavy-ion collisions at the CMS collision energy $\sqrt{s} = 5.02$ TeV was imported and validated. As a first cross check of the SuperChic implementation, the results from the generator were compared with the results from Ref. [4], which describes the observation of the light-by-light scattering process in heavy-ion collisions. The cross section calculated by SuperChic is 49.98 nb,

Table 4.2: Summary of used set of SuperChic parameters with their descriptions and assigned values.

Setting	Description	Value
`nev`	The number of events.	10000
`rts`	The CMS collision energy in GeV.	13000
`beam`	The type of the beam.	'prot'
`ymin`	Cut on the minimal central system rapidity.	-2.4
`ymax`	Cut on the maximal central system rapidity.	2.4
`mmin`	Cut on the minimal mass of the object in GeV.	6
`mmax`	Cut on the maximal mass of the object in GeV.	500
`ptamin`	Cut on the minimal transverse momentum of produced particle *a*.	3.0
`ptbmin`	Cut on the minimal transverse momentum of produced particle *b*.	3.0
`etaamin`	Cut on the minimal θ of the produced particle *a*.	-2.4
`etabmin`	Cut on the minimal θ of the produced particle *b*.	-2.4
`etaamax`	Cut on the maximal θ of the produced particle *a*.	2.4
`etabmax`	Cut on the maximal θ of the produced particle *b*.	2.4
`PDFname`	'MMHT2014lo68cl'	

Table 4.3: Overview of the simulated productions and their respective process numbers defined by SuperChic's `proc` parameter.

`proc`	Process description
56	Two-photon induced electron-positron pair e^+e^- production.
57	Two-photon induced muon-antimuon pair $\mu^+\mu^-$ production.
59	The production of photon pairs $\gamma\gamma$ in the $\gamma\gamma \to \gamma\gamma$ process.
68	Axion-like particle production.

and the reference states cross section of 50 ± 5 nb, where the uncertainty comes from limited knowledge of the nuclear form-factors, i.e. the scattering angle and the related initial photon fluxes (the number of photons per second per unit area) [42]. The high comparability of the results proved correct setting of the simulator.

With the confirmed integrity of SuperChic-3 implementation, the following objective was oriented on two-photon induced productions of $\gamma\gamma$, e^+e^-, $\mu^+\mu^-$ pairs and the axion-like particle in proton collisions with collision energy $\sqrt{s} =$ 13 TeV. The simulator setting that was used for every simulation run is shown in Table 4.2 and the used `proc` parameters defining the generated processes are shown in Table 4.3. For the axion–like particle production two additional parameters are present, the ALP mass `mapl` defined in GeV and the ALP coupling parameter `gax` defined in GeV^{-1}. The parameters were set to 1000 GeV and 0.001 GeV^{-1}, respectively.

4.2 Simulation analysis

Following the generator runs, the data from output files were extracted using a simple text file parser, whose functionality is based on LHEF file structure described in Section 4.1. For each production process the proton tagging probabilities were calculated and are shown in Subsection 4.2.1. Moreover, the analysis of properties of produced particles, that were defined in Section 2.2, is shown in Subsection 4.2.2.

4.2.1 Observation probability

When two beam protons interact, they lose a fraction of their energy. If they do not dissociate into a low mass state and their lost energy is within the range 2% to 10%, they can be tagged by the ATLAS Forward Proton (AFP) detector. Based on the simulation results, the predicted relative energy loss of the protons was calculated for each process together with the tagging probabilities as number of events meeting the tagging requirement divided by number of total simulated events. Based on the results, the tagging expectation for productions of photon pairs and lepton pairs in Standard Model is very low, with the highest estimation being 0.34%. For the axion–like particle production, the probability of tagging at least one of the protons is nearly 92%, significantly surpassing the observation probabilities of the lepton pair productions. To explore how proton tagging possibility changes with increasing transverse momentum, a different cut was set requiring it to be 10 GeV instead of 3 GeV, as initially defined in Table 4.2. Slight rise in the probabilities came at the cost of significant cross section decrease for the lepton pair productions. For the axion-like particle model the cut did not show any deviation from the previous values. The tagging probabilities for each production process based on the cut on transverse momentum are shown in Tables 4.4 and 4.5, and the cross sections of the processes are shown in Table 4.6. The overview of the relative energy loss of the beam protons is shown in Figure 4.1.

4.2.2 Data exploration

Apart from the tagging probabilities, multiple properties of particles present in the events such as energy, invariant mass and relative energy loss were explored. Analysing the energy distributions, which are shown in Figure 4.2, it is clear that produced photons, electrons and muons from the Standard Model follow a very similar pattern, where in majority of events the pairs obtain energy of approximately 10 GeV. The distributions then keep falling up to 100 GeV, which is the highest energy the pairs obtained in the simulation. On the contrary, the produced axion-like particle obtains much higher energy up to 2.5 TeV.

Table 4.4: Overview of the ATLAS Forward Proton detector tagging probabilities of the two beam protons given in per cent (%). The relative energy losses of the two respective protons are denoted as ξ_1 and ξ_2. If the energy loss is in range between 2% and 10%, it is considered in and therefore tagged by the detector. The values correspond to 3 TeV transverse momentum cut on produced particles and the photons, into which axion dissociates.

Prod.	ξ_1 in $\wedge$ ξ_2 out	ξ_1 out $\wedge$ ξ_2 in	ξ_1 in $\vee$ ξ_2 in	ξ_1 in $\wedge$ ξ_2 in
$\gamma\gamma$	0.17 ± 0.04	0.16 ± 0.04	0.34 ± 0.06	0.01 ± 0.01
ALP	34.25 ± 0.47	33.25 ± 0.47	91.82 ± 0.27	24.36 ± 0.43
e^+e^-	0.24 ± 0.07	0.08 ± 0.04	0.16 ± 0.06	0 ± 0
$\mu^+\mu^-$	0.15 ± 0.04	0.12 ± 0.04	0.27 ± 0.06	0 ± 0

Table 4.5: Overview of the ATLAS Forward Proton detector tagging probabilities of the two beam protons given in per cent (%). The relative energy losses of the two respective protons are denoted as ξ_1 and ξ_2. If the energy loss is in range between 2% and 10%, it is considered in and therefore tagged by the detector. The values correspond to 10 TeV transverse momentum cut on produced particles and the photons, into which axion dissociates.

Prod.	ξ_1 in $\wedge$ ξ_2 out	ξ_1 out $\wedge$ ξ_2 in	ξ_1 in $\vee$ ξ_2 in	ξ_1 in $\wedge$ ξ_2 in
$\gamma\gamma$	1.18 ± 0.11	1.48 ± 0.12	2.75 ± 0.16	0.09 ± 0.03
ALP	34.25 ± 0.47	33.25 ± 0.47	91.82 ± 0.27	24.36 ± 0.43
e^+e^-	2.55 ± 0.16	2.61 ± 0.16	5.24 ± 0.22	0.08 ± 0.03
$\mu^+\mu^-$	1.29 ± 0.11	1.46 ± 0.12	2.77 ± 0.16	0.02 ± 0.01

Table 4.6: Overview of cross sections calculated by the SuperChic generator with the dependency on transverse momentum p_T cut applied on produced particles.

Production	3 GeV p_T cut	10 GeV p_T cut
$\gamma\gamma$	7.2327 ± 0.0276 fb	0.4643 ± 0.0019 fb
ALP	6.0821 ± 0.0276 fb	6.0821 ± 0.0276 fb
e^+e^-	18.9427 ± 0.0908 pb	1.0236 ± 0.0042 pb
$\mu^+\mu^-$	18.9250 ± 0.0907 pb	1.0240 ± 0.0045 pb

The same applies for the invariant mass distributions shown in Figure 4.3, where the photon, electron and muon pairs from the Standard Model show a very similar behaviour. Because the mass of simulated axion was set to 1 TeV, the distribution peaks at the corresponding mass, resembling the normal distribution, and therefore strongly differs from the former projections.

Lastly, from the analysis of the relative energy loss shown in Figure 4.4 it is clear that the two-photon induced photons and leptons from the Standard Model lose significantly less energy than the photons into which the axion-like particle decays. It was observed that the energy loss of photons in axion production spans over tenths of the original energy, while on the contrary in the Standard Model the loss is lower by a factor of ten. It is important to note that the relative energy loss of beam protons shown in Figure 4.1 correlates with the relative energy loss of the produced particles.

4.2.3 Axion production variations

Further analysis was focused specifically on the axion-like particle production. It was oriented on determining the influence of axion mass on the acoplanarity of photons into which it decays and on the proton tagging probabilities. Moreover the impact of the coupling constant on the shape of diphoton invariant mass distribution was explored.

For the first objective, axion productions at masses ranging from 200 GeV to 1000 GeV with 100 GeV increments were simulated, where 5000 events were produced in each run. As expected, higher ALP masses result in lower acoplanarity values as shown in Figure 4.5. The highest observed acoplanarity was 0.008 for events in which an axion with mass of 200 GeV was produced. Therefore the conclusion of the analysis is that all generated events met the light-by-light physics analysis requirement for photon pair acoplanarity to be less than 0.01 regardless the ALP mass in range from 200 GeV to 1000 GeV and that by probing higher masses the requirement will still be met.

For determining the proton tagging probabilities, the set of simulated events from acoplanarity exploration was extended by simulated events for axion masses in range from 1000 GeV to 2000 GeV with 200 GeV increments. The results are shown in Figure 4.6 and indicate that probability of tagging at least one of the beam protons increases for productions of axions with masses up to 1200 GeV and then it starts to decrease. For tagging both protons in the same event, the probability reaches maximum at 600 GeV and then falls to zero at 1400 GeV.

Furthermore, the dependency of axion mass distribution width on coupling constant was explored. For this reason three different coupling values were selected, 0.001 GeV^{-1}, 0.0005 GeV^{-1} and 0.00025 GeV^{-1}. The results are shown in Figure 4.5 and indicate relation between the axion mass distribution width and the coupling constant, as with lower coupling, the width gets narrower and peaks more sharply.

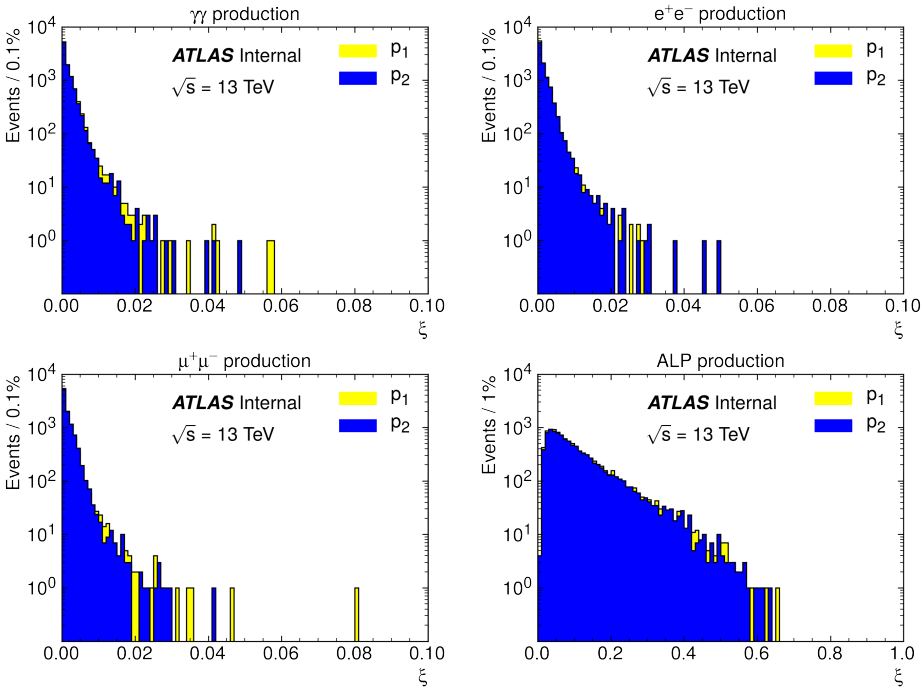

Figure 4.1: Distributions of the relative energy loss of the beam protons for corresponding production processes.

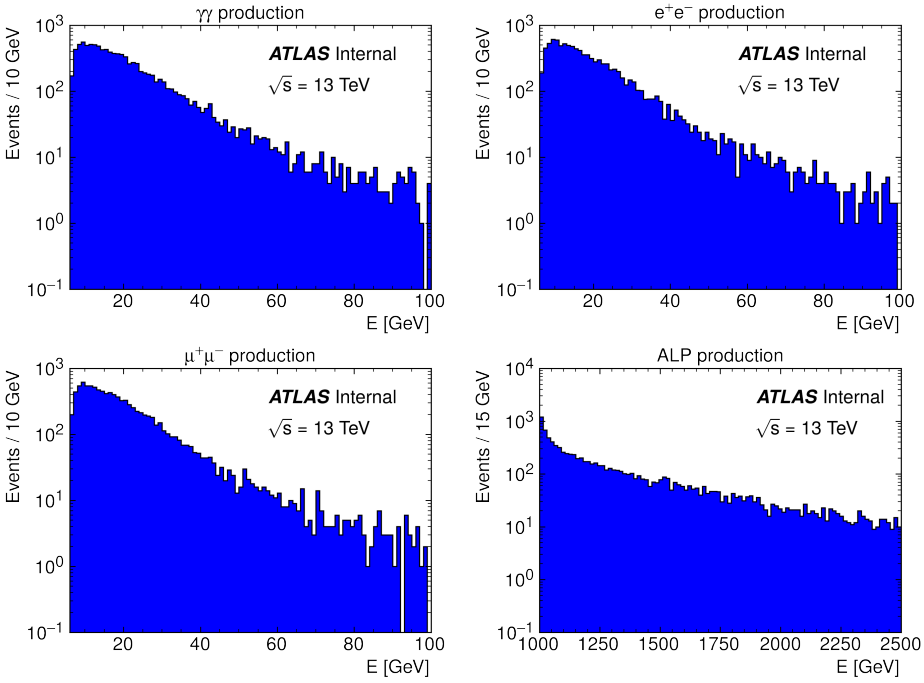

Figure 4.2: Energy distributions of the generated photon-photon pairs, electron-positron pairs, muon-antimuon pairs and the axion-like particle.

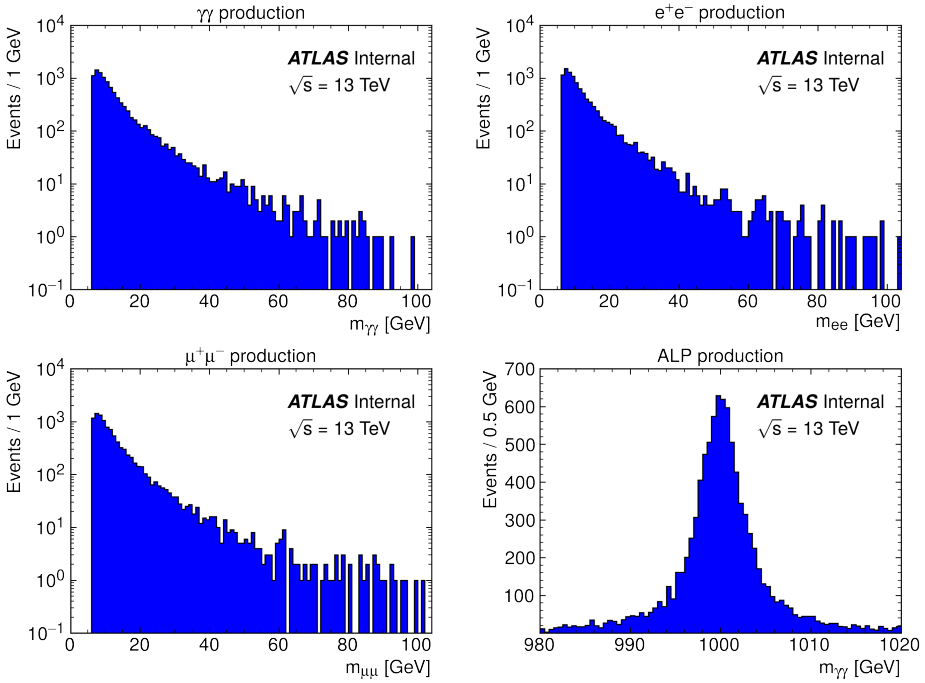

Figure 4.3: Invariant mass distributions of the generated photon-photon pairs, electron-positron pairs, muon-antimuon pairs from the Standard Model and the photon-photon pairs into which axion-like particle decays.

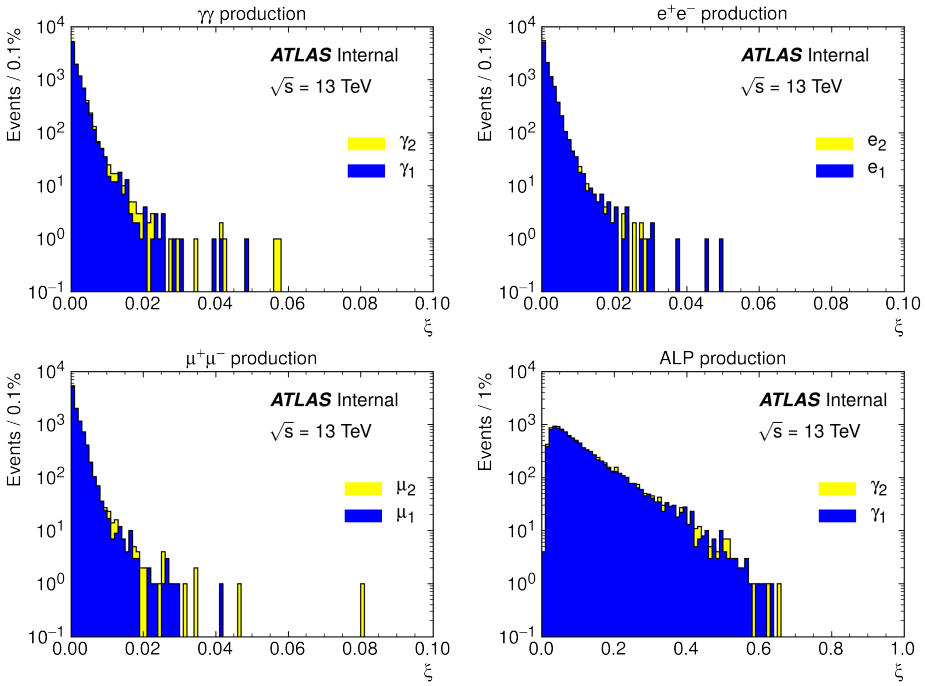

Figure 4.4: Relative energy loss distributions of the generated photon-photon pairs, electron-positron pairs, muon-antimuon pairs from the Standard Model and the photon-photon pairs into which axion-like particle decays.

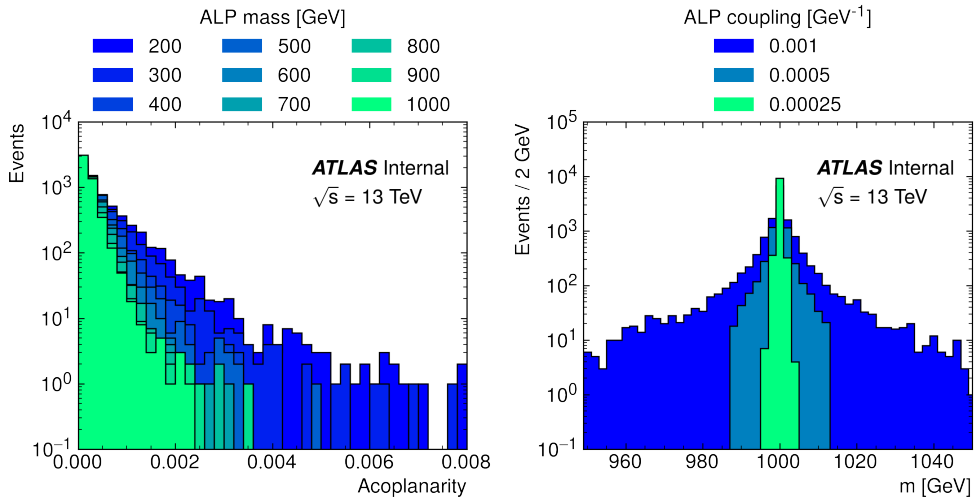

Figure 4.5: Left: acoplanarity distributions based on the axion–like particle mass. They show that regardless of the ALP mass, the acoplanarity satisfies the requirement of being lower than 0.01 as stated in Section 2.3. Right: peak of the axion–like particle mass distribution becomes sharper with smaller coupling parameter.

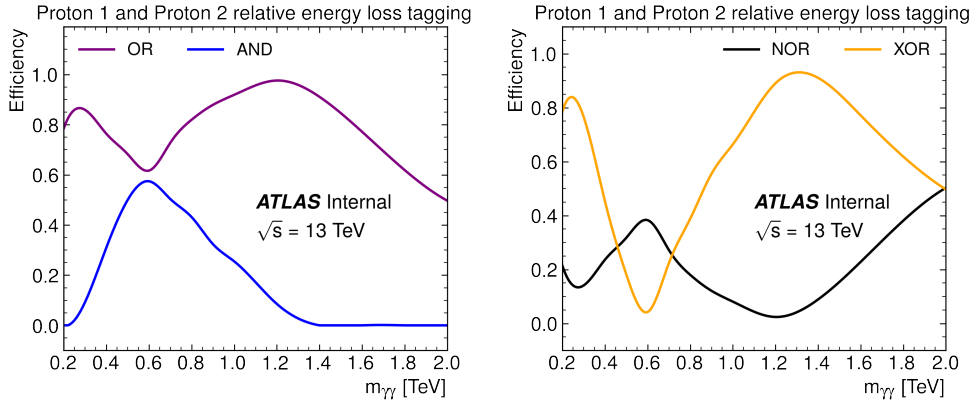

Figure 4.6: Proton tagging probabilities in axion–like particle production process based on scatted diphoton invariant mass.

Chapter 5

Smooth background modelling

This chapter explores the application of regression methods, that were introduced in preceding Chapter 3, on measurements obtained during the 2017 ATLAS experiment. It describes the process of modelling the diphoton invariant mass background using the Levenberg-Marquardt curve fitting method and the Gaussian Process for purposes of framework determination required for an axion detection, and compares the performances of both approaches.

In the second part of the chapter estimation of an axion production at 1 TeV is injected into the model in form of simulated data that was externally produced based on the results and parameters presented in Chapter 4. It serves as a projection of possible signal that could be observed in conducted experiment. However, as the search for an axion can be competitive in mass range between 0.5 TeV and 2 TeV, as described in Section 2.3, and not just at one constant mass, the whole spectrum is explored together with different detection strategies as well.

The end of the chapter discusses the differences between the two modelling approaches and the obstacles that were recognized during the process.

5.1 Analysed experiment

For purposes of this thesis dataset with measurements taken from conducted ATLAS experiment in 2017 containing information about events in which scattered diphotons with invariant mass ranging from 0 TeV to 4 TeV were detected in the central detector at the LHC was imported. The requirement on diphoton acoplanarity, i.e. "blinding criteria", $\mathcal{A} > 0.01$ was applied on all imported events for purpose of excluding potential events in which the axion with mass higher than 0.2 TeV could be produced, as shown in Subsection 4.2.3.

The data was further partitioned based the observations from Subsection 4.2.2 that relative energy losses of beam protons and photon pairs are correlated in photon mediated productions. The detectors at the LHC can tag

Table 5.1: The number of imported events from the experiment dataset and the simulation runs. For the measured data, it only includes events satisfying the acoplanarity $\mathcal{L} > 0.01$ cut and the corresponding matching strategies requirements.

	Experiment			Simulation	
	No AFP matching	A or C matching	A and C matching	Base level	Generator level
Events	4,487,204	80,392	66	502,614	7,377

outgoing particles in both directions relative to the interaction point, i.e. A-side and C-side. If the energy losses of scattered photon and deflected proton are matched by the A-side, respectively C-side, detector, meaning that the difference between them is less than 10%, it is considered as A-matching, respectively C-matching. Therefore the derived detection strategies are A or C matching and A and C matching, which selects events where the energy losses are matched by at least one detector and events in which the energy losses are matched by both detectors, respectively. Based on these matching strategies, three subsets were extracted from the 2017 ATLAS data. The first subset does not consider any matching (no AFP information used), the second includes events satisfying the A or C requirement and the third one corresponds to the A and C matching. The latter two requirements significantly reduce the number of observable events, and therefore increase the background modelling difficulty, as the insufficient data only partially describes the observed behaviour. The invariant mass distributions of each subset are shown in Figure 5.1 and the numerical values are noted in Table 5.1.

Apart from the observed events externally simulated axion-like particle production at mass of 1 TeV and with coupling parameter 1 TeV^{-1} was imported. The simulated data consists of two parts. The first one includes "real observations", meaning it represents the process as if it would be observed by a perfect detector. The second part was obtained by running a successive simulation replicating the function of the AFP detector on the data from the first part using the AFP toolbox, which takes the reconstruction of the photons in the ATLAS central detector into account. It also includes the uncertainties on the reconstructed photons and therefore represents an estimation how the process could be detected.

Diphoton mass distributions from the two simulated sets are shown in Figure 5.2.

5.2 Background modelling

Using the extracted sets from the experiment dataset, three separate falling backgrounds corresponding to each matching strategy were modelled using the standard regression model based on parametric function 3.24 optimised using the Levenberg-Marquardt method, and the Gaussian Process defined with

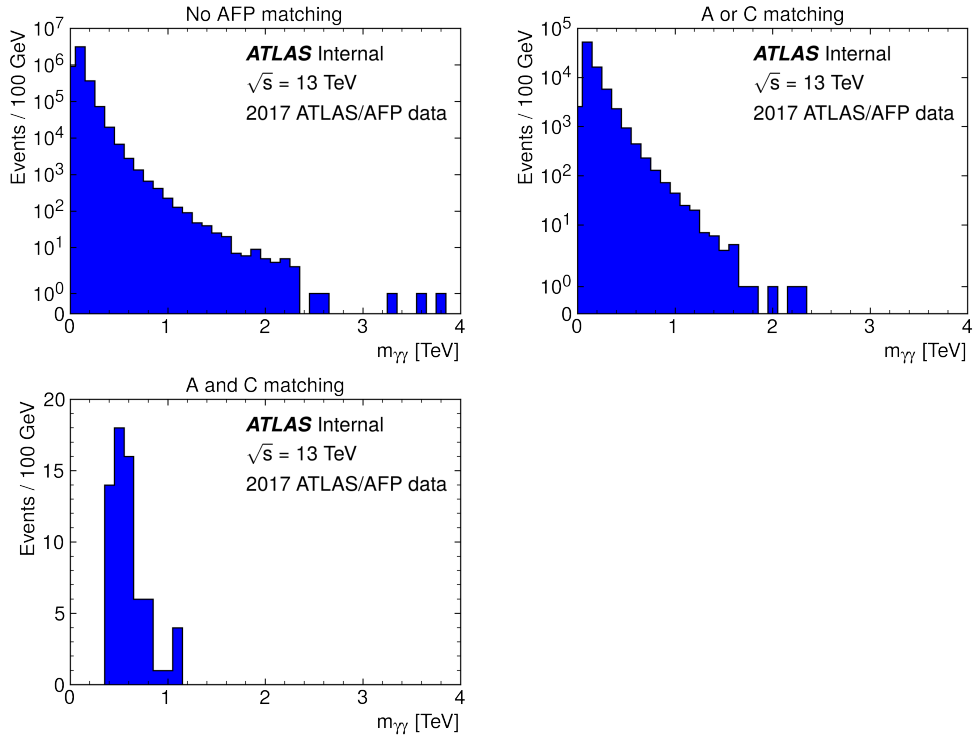

Figure 5.1: Invariant mass distributions of the observed diphotons from 2017 ATLAS experiment based on applied matching strategies together with the blinding criteria. With increasing matching restriction, the number of events that can be analysed decreases significantly.

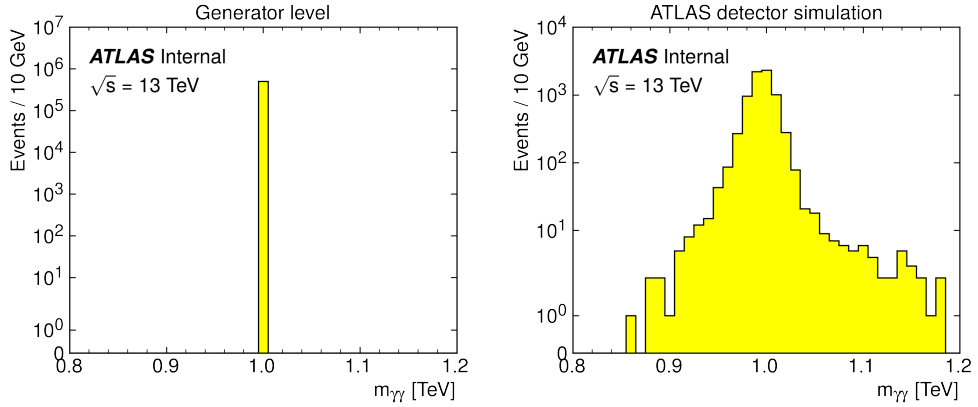

Figure 5.2: Left: invariant mass distribution of externally generated diphotons into which an axion dissociated, as it would be observed by a perfect detector. Right: invariant mass distribution of diphotons as it would be seen by the AFP detector if the externally simulated events would occur.

Table 5.2: The number of events with diphoton invariant mass in range between 0.5 TeV and 2 TeV from the experiment dataset satisfying the acoplanarity $\mathcal{L} > 0.01$ cut and the corresponding matching strategies requirements.

	No AFP matching	A OR C matching	A AND C matching
Events	8,790	1,419	43

mean function 3.24 and covariance function 3.25. The former model represents the most frequent approach to background modelling, and therefore it was used for determining the precision of the GP model. Both frameworks are build on the same function, hence the GP model can be think of as an extension of the former model. As discussed in Section 2.3, the most interesting mass range in which the search for an axion–like particle can take place is between 0.5 TeV and 2 TeV. That being so, the backgrounds were not modelled on the whole observed range of 4 TeV, but only on the bounded spectrum as the analysis of lower and higher masses is not required. This reduced the number of analysed events even further. The amounts of events in range between 0.5 TeV and 2 TeV with applied matching restrictions are shown in Table 5.2.

The precisions of modelled backgrounds were then compared. As shown in Table 5.3 the metrics determining the goodness of fit were mean squared error, mean absolute error, coefficient of determination and maximum residual error. The biggest difference between the models is in the background for events not restricted by any matching requirement. It also offers the highest number of events which leads to the most accurate representation of the diphoton behaviour. Comparing the scores of both fits, the Gaussian Process achieves better performance than the simple fit in every metric, apart from mean absolute error for background corresponding to A and C matching. From the visual representation of the backgrounds shown in Figure 5.3 it can be seen that the simple function fit is much smoother and does not vary. On the contrary the GP is much more sensitive to the variability of the data and therefore takes the physics behind the experiment into account, which offers an important advantage for calculations that are based on the background model. This flexibility is reduced in modelled backgrounds for A or C and A and C matching sets as the statistics is much lower than in the former set. The hyperparameters of the covariance function of the model could be set in a way that the background would follow the data more closely, but that would result in complex, and more importantly, overfitted model that would introduce nonphysical structures. Because of the positive outcome of the regression analysis based on Gaussian Process, its results were used in the following sections as the main background estimations.

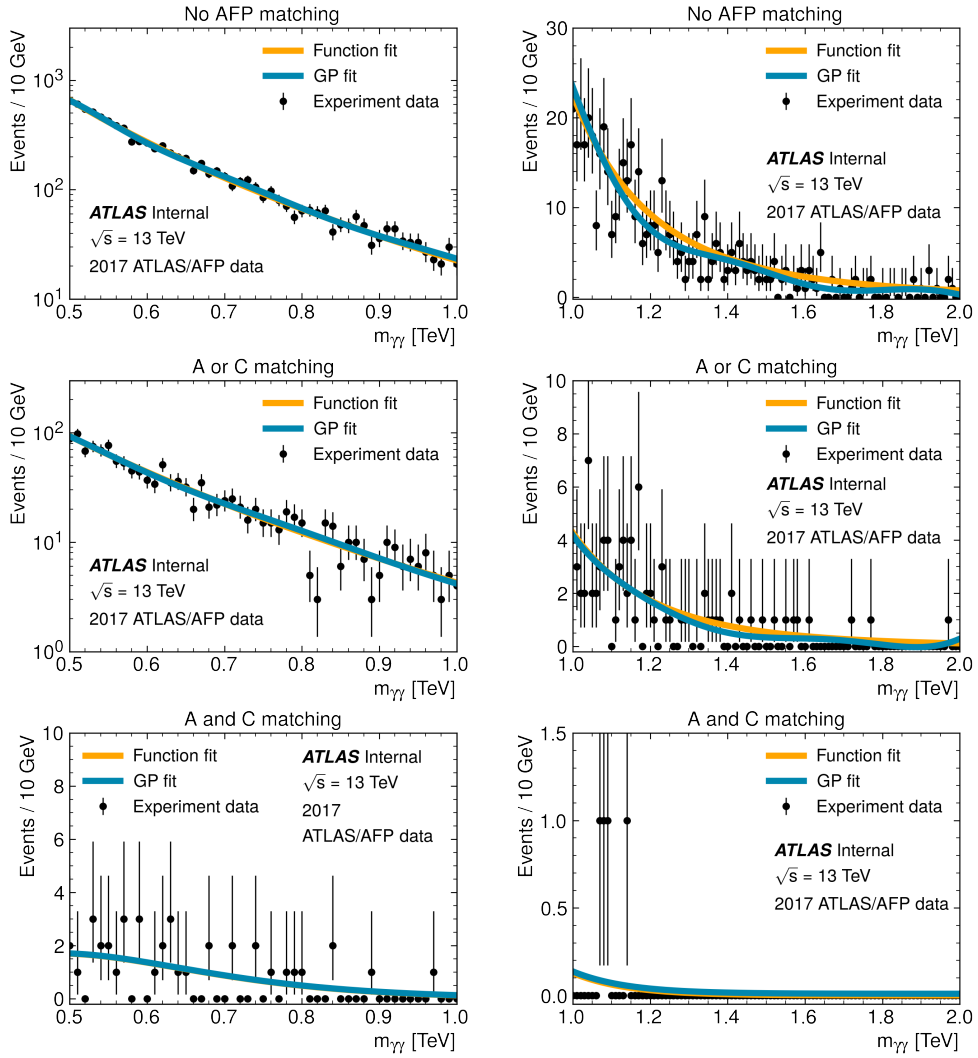

Figure 5.3: Comparison between the Gaussian Process fit and the simple function fit on diphoton mass range from 500 GeV to 2000 GeV. On each row are distributions for different data sets extracted from 2017 ATLAS experiment measurements. The sets corresponding to no AFP matching, A or C matching, and A and C matching are ordered from top to bottom. The plots on the left show the low mass distributions on logarithmic scale (with the exception for the A and C matching plot) and the plots on the right show high mass distributions on linear scale. The error bars represent 68% confidence level in corresponding observation.

Table 5.3: Overview of the goodness of fit metrics for the Gaussian Process fit and the mean function fit for each matching subset.

No AFP matching		
Metric	Gaussian Process fit	Standard curve fit
MAE	**3.5747**	3.8682
MSE	**42.4946**	47.5192
R^2	**0.9971**	0.9968
Max. residual error	**42.8861**	47.7941

A or C matching		
Metric	Gaussian Process fit	Standard curve fit
MAE	**1.6344**	1.6785
MSE	**9.0584**	9.1406
R^2	**0.9733**	0.9731
Max. residual error	**13.7326**	13.9640

A and C matching		
Metric	Gaussian Process fit	Standard curve fit
MAE	0.2661	**0.2595**
MSE	**0.2509**	0.2510
R^2	**0.4645**	0.4643
Max. residual error	**1.7756**	1.7849

5.3 Signal injection

After the regression analysis, the externally simulated signal indicating diphoton invariant mass peak caused by hypothetical axion-like particle production was injected into the no AFP matching model by placing it on top of the background. Because the number of generated signal events exceeds the data from the ATLAS experiment, the signal had to be scaled in order for the peak to match the expected 5σ significance. Since the simulation was based on axion production with mass of 1 TeV, the number of events needed to be observed on top of the background at that particular bin to reach 5σ significance was calculated using the Poisson statistics formula for continuous values as

$$h = F^{-1}(1 - \alpha/2; \lambda + 1), \tag{5.1}$$

where F^{-1} is the inverse quantile function of a gamma distribution with scale parameter 1, α is the upper bound of confidence level equal to 5σ, which is approximately 99.99994%, and λ is the estimated number of events in corresponding bin based on the background. The signal distributions were then scaled by parameter $\mathfrak{p} = h/n_{\mathbb{1}}$, where $n_{\mathbb{1}}$ is the number of simulated events

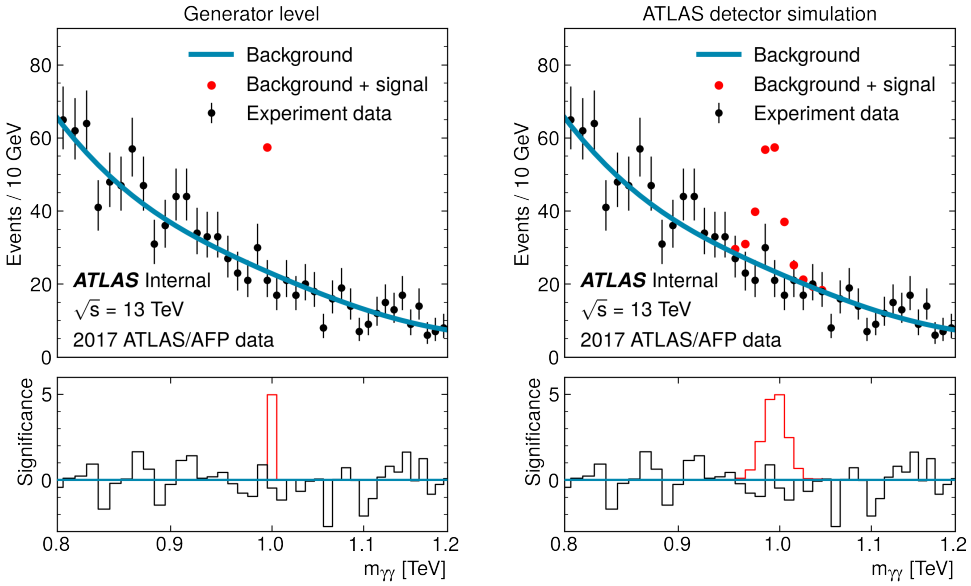

Figure 5.4: Signal injections into the modelled no AFP matching background in range 0.8 TeV to 1.2 TeV based on observational model. The significance of the peak caused by simulated axion-like particle production is shown under the background projection together with deviations of the measured data from the background.

inside the bin at 1 TeV. The scaled signals placed on top of the background are shown in Figure 5.4.

Furthermore, regression analysis of the scaled signal based on ATLAS detector simulation was performed using Gaussian Process defined with simple mean function $m(\boldsymbol{x}) = 0$ and radial basis function

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathfrak{m} \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{x}')^2}{2l^2}\right) \tag{5.2}$$

as covariance function, where $\mathfrak{m}$ is free parameter and l defines fixed length scale. Since the resulting fit that is shown in Figure 5.5 strongly resembles normal distribution, its corresponding standard deviation was calculated using full width at half maximum (FWHM), that determines the width of values higher than half of the maximum and is related to standard deviation σ as

$$\mathrm{FWHM} = 2\sqrt{2 \ln 2}\sigma. \tag{5.3}$$

The obtained result was 10.6 GeV which supports the presented binned analysis based on bin width of 10 GeV.

5.4 Observation sensitivity

So far the projections were made with an assumption of an axion production at 1 TeV. However, it is possible that the process could be occurring at different mass, and therefore it is necessary to explore these options as well. Staying at a sensitive range of 0.5 TeV to 2 TeV, it is required to calculate how many events are needed to be observed in each bin anywhere on the distribution to reach significant deviation on all three backgrounds. For this purpose two significances were considered, 5σ as in Section 5.3 and 2σ corresponding to upper bound of 95.44% confidence level. Therefore six sets of values come from this stage, as for the three backgrounds two different sets of events representing signal at each bin are calculated using equation 5.1.

Next, the cross sections of assumed signal events were calculated using equation 2.1 with luminosity set to 32 fb^{-1} and selection efficiency to 100%. For the results relative to background for A or C matching upper and lower 68% and 95% confidence limit bands were determined using equations 5.1 and 5.4, respectively, and projected in Figure 5.6.

$$h' = F^{-1}(\alpha/2; \lambda) \tag{5.4}$$

Determining the observational sensitivity is also based on the strength of ALP coupling to photons. As observed from the SuperChic-3 simulations from Subsection 4.2.3, with lower coupling the invariant mass distribution of the diphotons, into which created axions decay, becomes narrower and therefore for the same production rate creates higher peak with increased significance in relation to the background. This leads to escalated sensitivity, i.e. probability of detection, because it means that lower production rate is required in order to observe significant deviation. Therefore determining the coupling, i.e. the axion production rate, that would cause observation of the assumed signal

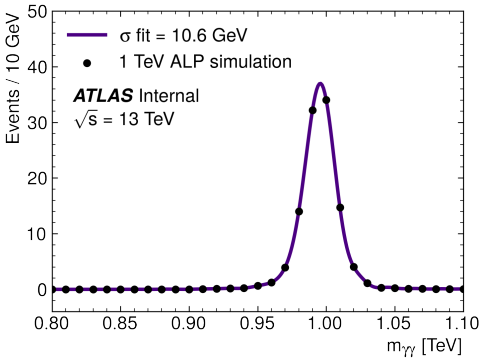

Figure 5.5: Signal fit based on ATLAS detector simulation modelled with Gaussian Process.

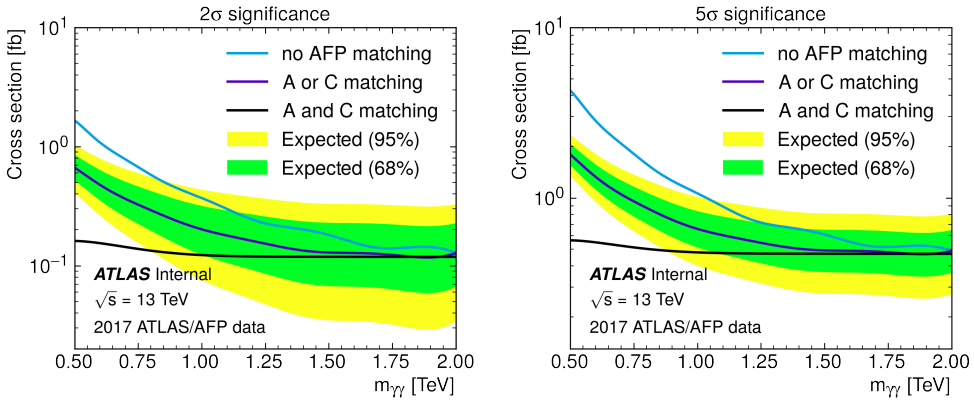

Figure 5.6: Cross section projections of events needed to reach deviation from the background with 2σ and 5σ significances over the whole analysed range and with applied matching restrictions. For signal events based on background for A or C matching 68% and 95% confidence limit bands are projected.

with the efficiencies of the detectors taken into account is an important part of the analysis. For assumed signal events corresponding to each 10 GeV bin in range 0.5 TeV to 2 TeV on all three backgrounds, the coupling f^{-1} was calculated using the proton tagging efficiencies from Subsection 4.2.3 and photon detection efficiency that was externally determined by the ATLAS central detector simulation as 0.7377. The f^{-1} values were then obtained as

$$\mathfrak{g} = \sqrt{\frac{\hat{\sigma}_{\mathrm{lim}}}{\hat{\sigma}_{\mathrm{sim}}}}, \tag{5.5}$$

$$f^{-1} = \frac{\mathfrak{g}}{4}, \tag{5.6}$$

where $\hat{\sigma}_{\mathrm{lim}}$ is the cross section for assumed signal events calculated with detector efficiencies denoted above and luminosity set to 32 fb^{-1}, and $\hat{\sigma}_{\mathrm{sim}}$ is the cross section calculated by SuperChic-3 for an axion production at respective mass with 1 TeV^{-1} coupling. The results are projected in Figure 5.7 and signify that different matching strategies need to be applied based on axion mass and are the following: A and C matching for low masses up to 800 GeV, A or C matching for higher masses up to approximately 1.6 TeV and no AFP matching for the highest masses.

5.4.1 Alternative sensitivity

The analysis was repeated with different three region of logical functions. The matching strategies correspond to logical NOR, XOR and AND. These regions unlike the ones described before do not have any common events, and therefore they can be combined into one smooth sensitivity curve. The

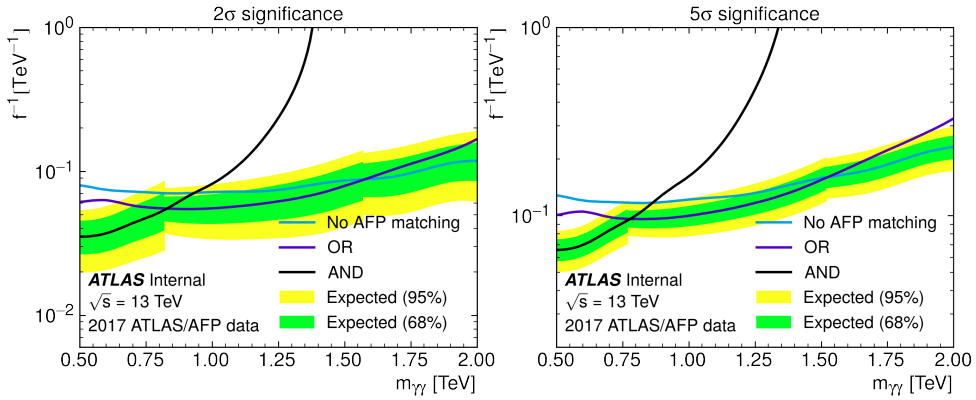

Figure 5.7: Sensitivity projection on the ALP–photon coupling f^{-1} and the diphoton mass $m_{\gamma\gamma}$ plane based on applied matching restrictions.

Table 5.4: The number of events with diphoton invariant mass in range between 0.5 TeV and 2 TeV from the experiment dataset satisfying the acoplanarity $\mathcal{L} > 0.01$ cut and the corresponding alternative matching strategies requirements.

	A NOR C matching	A XOR C matching	A AND C matching
Events	7,371	1,376	43

number of events in range between 0.5 TeV and 2 TeV with applied matching restrictions is shown in Table 5.4.

To obtain the smooth sensitivity curve an iterative approach was utilised. Its purpose was to find such f^{-1} values based on diphoton mass that would correspond to production of events for which the combined probability of being only statistical error above NOR, XOR and AND backgrounds would be 5%. The combination is defined as

$$(1 - \mathrm{CL}) \approx (1 - \mathrm{CL}_{\mathrm{NOR}}) \cdot (1 - \mathrm{CL}_{\mathrm{XOR}}) \cdot (1 - \mathrm{CL}_{\mathrm{AND}}), \tag{5.7}$$

where CL is the confidence level equal to 95% and the remaining terms $(1 - \mathrm{CL}_{\mathrm{NOR}})$, $(1 - \mathrm{CL}_{\mathrm{XOR}})$ and $(1 - \mathrm{CL}_{\mathrm{AND}})$ match the probabilities that the number of events produced with production rate f^{-1} above corresponding background happened randomly as statistical error. Therefore for obtaining the smooth curve the process from Section 5.4 was reversed and instead of calculating the number of events that would reach a certain significance, the combined confidence of possibly produced events based on production rate was determined. This approach was iteratively repeated with increasing f^{-1} value until the requirement 5.7 was met for each analysed diphoton mass.

Apart from the smooth combined curve the sensitivity curves for each alternative matching were obtained as well using the same approach as in Section 5.4. The results are shown in Figure 5.8.

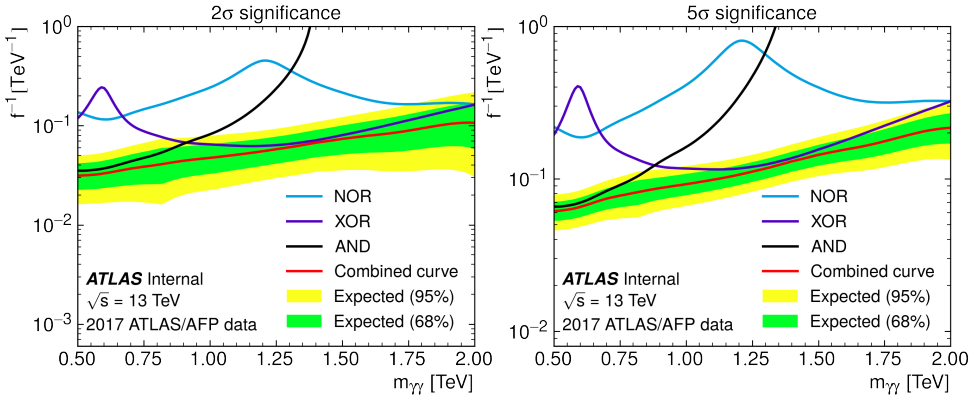

Figure 5.8: Sensitivity projection on the ALP–photon coupling f^{-1} and the diphoton mass $m_{\gamma\gamma}$ plane based on applied alternative matching restrictions.

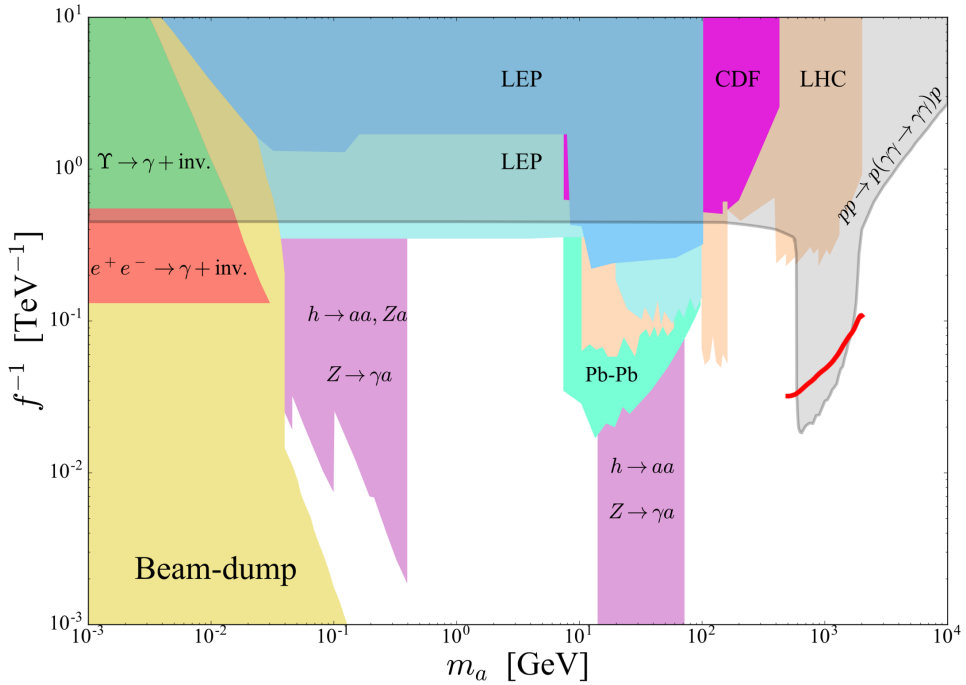

Figure 5.9: Red combined curve for 2σ significance peaks imported into the ALP–photon coupling f^{-1} and the diphoton mass $m_{\gamma\gamma}$ plane from Ref. [8]. The plot shows how the sensitivity curve matches the expected sensitive region for axion–like particle detection at the LHC.

5.5 Discussion

One of the main problems in modelling using Gaussian Process is the estimation of the prior set of hyperparameters, respectively determining the bounds of uniform distributions from which they are sampled and later optimised. The model had to be tested multiple times on different sets of parameter bounds, because the optimisation using Minuit would often run into the bounding value, which indicates that a more likely model could be found with hyperparameters excluded by the bounds. However, since the background modelling deals only with very limited dataset, the model is very prone to overfitting. This problem was mainly occurring at high masses where the statistics is very low and the prediction would include nonphysical structure in form of "waves", that would capture bins with few events in otherwise empty space. This issue would be even more escalated in backgrounds that take matching strategies into account. The problem can be solved by increasing length scale $l(x)$ of the Gibbs kernel. Although this would smooth out the predictions at low masses, it could suppress the influence of the mean function and the model would not follow the exponential form of the background, but fall into negative values. For this reason multiple optimisation runs were performed to find the optimal hyperparameter bounds.

Apart from the Gibbs kernel, two different covariance functions were inspected. The radial basis function 5.2, that was used for signal modelling, was not able to precisely model the background at the whole mass range, because of its fixed length scale causing nonphysical structures. Another tested kernel was augmented Gibbs kernel, that introduces two new free parameters a and d and multiplies the standard version by term $\exp\left(\frac{d-(\boldsymbol{x}+\boldsymbol{x}')}{2a}\right)$. The resulting background did not follow a smooth line and included nonphysical fluctuations as well and because of the increased number of hyperparameters, it was even more prone to overfitting. Measuring its goodness of fit by the same metrics as in Table 5.3, it achieved lower scores than the standard function fit, and therefore was considered ill-suited for the task of background modelling. The advantage of modelling with Gaussian Process becomes more apparent by increasing the modelled mass range. This is shown in Figure 5.10, where the background is modelled for a set of events unrestricted by matching strategies at range 100 GeV to 4000 GeV. It is visibly apparent how especially at low masses the function fit is much less suited for capturing the variances of the background, as the prediction does not go through the observations, but stays under them. This would cause low accuracy in deducing the number of events needed for potential discovery of new physics. To achieve higher precision, the background would have to be modelled in smaller steps. (This is the reason why the function fit performs relatively well for mass range 0.5 TeV to 2 TeV.) On the contrary, it shows the power of Gaussian Process as it manages to fix the issue and even includes the "bump" around 3.5 TeV. Although it might be a statistical error, it could also signify a production that would be otherwise

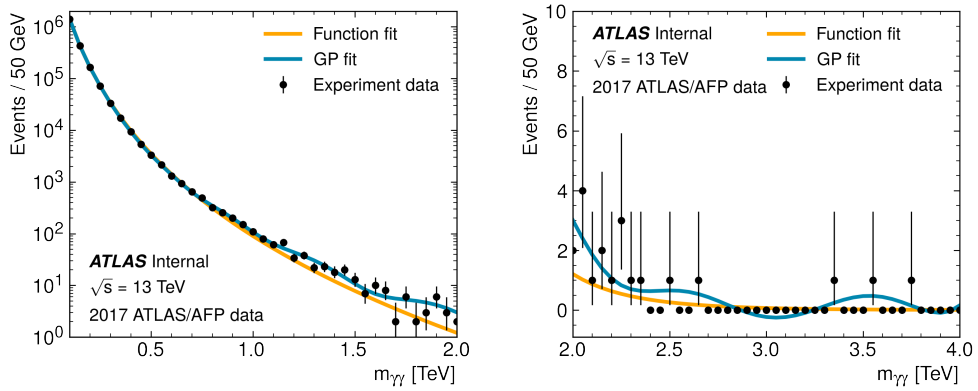

Figure 5.10: Comparison of background modelling results on wide mass range from 100 GeV to 4000 GeV.

left undetected. However, it also creates a dip at 3 TeV that is certainly invalid, because the number of observed events cannot be negative. This issue would need to addressed in an analysis based in this particular mass range.

Conclusion

The objective of this thesis was to apply an existing analysis framework involved in detecting lepton production induced by photon interactions in proton ultraperipheral collisions at CERN on studying light-by-light scattering caused by a hypothetical axion-like particle production and extend it by using a machine learning algorithm that would increase its detection efficiency.

The first part of the task was completed by initiating the SuperChic-3 event generator runs with focus on photon induced lepton pairs production and production of axions with mass of 1 TeV. The results were then compared observing much higher relative energy losses in beam protons, and photons into which the axion decays, than in the lepton productions. Furthermore, the simulations show correlated relationship between relative energy loss of protons and relative energy loss of produced particles induced by emitted photons in ultraperipheral collisions. Moreover, a relation between ALP coupling strength and the width of its mass distribution was observed, as the lower coupling causes the distribution to be narrower. The physics requirement for light-by-light observation on diphoton acoplanarity was tested for axion masses in range from 200 GeV to 1000 GeV for all of which the requirement was met. From analysis of the relative energy loss of simulated beam protons, AFP detector tagging probabilities were estimated based on production processes. For the axion production, the probabilities strongly exceeded those for lepton productions. The simulation analysis was published on the central ATLAS Collaboration webpage and it provided a foundation for external ATLAS detector simulations.

In the second part the problem of estimating the invariant mass distribution of scattered diphotons was addressed. Machine learning approach, the Gaussian Process, was applied on data obtained from the 2017 ATLAS experiment and its results were compared with the standard curve fitting method. Three subsets were extracted from the dataset corresponding to three different proton and photon relative energy loss matching strategies. The first one did not take any AFP information into account, the second one included

events where the losses are matched on at least one side relative to the interaction point, and the last subset applied the requirement for both sides. Based on goodness of fit metrics, the Gaussian Process achieved higher scores and therefore provided a better estimation of the backgrounds. It also included the physics behind the diphoton behaviour which could not be achieved by the standard curve fitting method.

Moreover, a simulated signal representing the axion-like particle production at mass 1 TeV was injected into the model to project the hypothesis. This showed a predicted shape and distribution of the events as if they would be observed by the AFP detector.

The backgrounds were then used to estimate number of events needed for an axion production observation on mass spectrum between 0.5 TeV and 2 TeV, which lead to determining the observational sensitivity for different matching strategies and showed that the most promising results could be obtained by combining the proton and photon relative energy loss matching strategies. For low masses up to 800 GeV the A-side and C-side matching should be applied, for higher masses up to approximately 1.6 TeV the A-side or C-side should be applied and for the highest masses no AFP matching could be potentially the most promising approach.

Bibliography

[1] Basu D. K. *Dictionary of Material Science and High Energy Physics*, CRC Press, January 2001, ISBN 0-8493-2889-6.

[2] Baltz A.; Baur G.; d'Enterria D.; et al. The physics of ultraperipheral collisions at the LHC. *Physics Reports*, vol. 458, n. 1-3, March 2008, ISSN 0370-1573, doi: 10.1016/j.physrep.2007.12.001.

[3] Aaboud, M.; Aad G.; Abbott B. et al. Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC. *Nature Physics*, Springer Science and Business Media LLC, vol. 13, n. 9, p. 852—858, September 2017, ISSN 1745-2481, doi: 10.1038/nphys4208.

[4] ATLAS Collaboration. Observation of Light-by-Light Scattering in Ultraperipheral Pb + Pb Collisions with the ATLAS Detector. *Physical Review Letters*, American Physical Society (APS), vol. 123, n. 5, July 2019, ISSN 1079-7114, doi: 10.1103/physrevlett.123.052001.

[5] Baldenegro C.; Hassani S.; Royon C.; Schoeffel L. Extending the constraint for axion-like particles as resonances at the LHC and laser beam experiments. *Physics Letters B*, Elsevier BV, vol. 795, p. 339-345, August 2019, ISSN 0370-2693, doi: 10.1016/j.physletb.2019.06.029.

[6] Peccei R. D.; Quinn H. R. CP Conservation in the Presence of Pseudoparticles. *Physical Review Letters*, American Physical Society, vol. 38, n. 25, p. 1440-1443, June 1977, doi: 10.1103/PhysRevLett.38.1440.

[7] Knapen S.; Lin T.; Melia T. et al. Searching for axion-like particles with ultra-peripheral heavy-ion collisions. *Physical Review Letters*, American Physical Society, vol. 118, n. 17, April 2017, doi: 10.1103/physrevlett.118.171801.

[8] C. Baldenegro, S. Fichet et al. Searching for axion-like particles with proton tagging at the LHC. *Journal of High Energy Physics*, Springer Science

and Business Media LLC, vol. 2018, n. 6, June 2018, ISSN 1029-8479, doi: 10.1007/jhep06(2018)131.

[9] Rawlings J.; Pantula S.; Dickey D. *Applied Regression Analysis: A Research Tool*. Springer-Verlag New York, vol. 2, p. 3-6, 1998, ISBN 0-387-98454-2.

[10] Oymak S.; Soltanolkotabi M. *Overparameterized Nonlinear Learning: Gradient Descent Takes the Shortest Path?* December 2018, arXiv:1812.10004.

[11] Mehta P.; Wang C. et al. *A high-bias, low-variance introduction to Machine Learning for physicists*. May 2019, arXiv:1803.08823.

[12] *Advances in Human Factors in Cybersecurity: Proceedings of the AHFE 2016 International Conference on Human Factors in Cybersecurity, July 27-31, 2016, Walt Disney World®, Florida, USA*. Springer, 2016, p. 418, ISBN 978-3-319-41932-9.

[13] Zisserman A. Lecture *B1 Optimization*, Department of Engineering Science, University of Oxford, 2012, available at `http://www.robots.ox.ac.uk/~az/lectures/b1/lect2.pdf`.

[14] J. Martens. Deep learning via Hessian-free optimization. *Proceedings of the 27th International Conference on Machine Learning*, Omnipress, p. 735—742, 2010, ISBN 978-1-60558-907-7.

[15] Agarwal N.; Bullins B.; Hazan E. *Second-Order Stochastic Optimization for Machine Learning in Linear Time*. November 2017, arXiv:1602.03943.

[16] Dennis J.; Gay D.; Welsch R. An Adaptive Nonlinear Least-Squares Algorithm. *ACM Trans. Math. Softw.*, Association for Computing Machinery, vol. 7, n. 3, p. 348—368, September 1981, ISSN 0098-3500, doi: 10.1145/355958.355965.

[17] Gratton S.; Lawless A.; Nichols N. Approximate Gauss-Newton methods for nonlinear least squares problems. *SIAM Journal on Optimization*, SIAM Publications, vol. 18, n. 1, p. 106–132, 2007, doi: 10.1137/050624935.

[18] Furmston T.; Lever G.; Barber D. Approximate Newton Methods for Policy Search in Markov Decision Processes. *J. Mach. Learn. Res.*, JMLR.org, vol. 17, n. 1, p. 8055—8105, January 2016, ISSN 1532-4435, doi: 10.5555/2946645.3053508.

[19] Rebentrost P.; Schuld M.; Wossnig L.; Petruccione F.; Lloyd S. *Quantum gradient descent and Newton's method for constrained polynomial optimization*. August 2019, arXiv:1612.01789.

[20] Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, American Mathematical Society, vol. 2, p. 164-168, July 1944, ISSN 1552-4485, doi: 10.1090/qam/10666.

[21] Marquardt D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, Society for Industrial and Applied Mathematics, vol. 11, n. 2, p. 431—441, June 1963, ISSN 0368-4245, doi: 10.1137/0111030.

[22] Ahookhosh M.; Artacho F. J.; Fleming R. M.; Vuong P. T. *Local convergence of the Levenberg–Marquardt method under Hölder metric subregularity*. February 2019, arXiv:1703.07461.

[23] Gavin H. *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. Department of Civil and Environmental Engineering, Duke University, August 2019, available at `http://people.duke.edu/~hpgavin/ce281/lm.pdf`.

[24] Transtrum M. K.; Sethna J. P. *Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization*. January 2012, arXiv:1201.5885.

[25] Finsterle S.; Kowalsky M. B. A Truncated Levenberg-Marquardt Algorithm for the Calibration of Highly Parameterized Nonlinear Models. *Computers & Geosciences*, Elsevier, vol. 37, n. 6, p. 731-738, June 2011, ISSN 0098-3004, doi: 10.1016/j.cageo.2010.11.005.

[26] Virtanen P.; Gommers R.; Oliphant T. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, Springer Science and Business Media LLC, vol. 17, n. 3, p. 261–272, February 2020, ISSN 1548-7105, doi: 10.1038/s41592-019-0686-2.

[27] Tipping M. E. Bayesian Inference: An Introduction to Principles and Practice in Machine Learning. *Advanced Lectures on Machine Learning*, Springer, p. 46-52, June 2006, ISBN 978-3540231226.

[28] Robert C. P. *Bayesian Computational Tools*. June 2013, arXiv:1304.2048.

[29] Rasmussen C. E.; Williams C. K. *Gaussian Processes for Machine Learning*. MIT Press, 2006, ISBN 978-0-262-18253-9.

[30] Lalchand V.; Rasmussen C. E. *Approximate Inference for Fully Bayesian Gaussian Process Regression*. December 2019, arXiv:1912.13440.

[31] ATLAS Collaboration. Search for resonances in diphoton events at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Journal of High Energy Physics*,

Springer Science and Business Media LLC, vol. 2016, September 2016, ISSN 1029-8479, doi: 10.1007/jhep09(2016)001.

[32] M. N. Gibbs. *Bayesian Gaussian Processes for Regression and Classification*. Ph.D. thesis, University of Cambridge, 1998.

[33] Frate M.; Cranmer K. et al. *Modeling Smooth Backgrounds and Generic Localized Signals with Gaussian Processes*. September 2017, arXiv:1709.05681.

[34] Schulz E.; Speekenbrink M.; Krause A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, Elsevier, vol. 85, p. 1-16, August 2018, doi: 10.1016/j.jmp.2018.03.001.

[35] Chen Z.; Wang B. *How priors of initial hyperparameters affect Gaussian process regression models*. October 2017, arXiv:1605.07906.

[36] Swain P. S.; Stevenson K.; Leary A. et al. *Inferring time-derivatives, including cell growth rates, using Gaussian processes*. Cold Spring Harbor Laboratory, 2016, bioRxiv:055483, doi: 10.1101/055483.

[37] A. G. Wilson. *Covariance Kernels for Fast Automatic Pattern Discovery and Extrapolation with Gaussian Processes*. Ph.D. thesis, University of Cambridge, 2014.

[38] James F.; Roos M. Minuit - A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Computer Physics Communications*, Elsevier, vol. 10, n. 6, p. 343–367, December 1975, doi: 10.1016/0010-4655(75)90039-9.

[39] Seymour M. H.; Marx M. *Monte Carlo Event Generators*. April 2013, arXiv:1304.6677.

[40] Harland-Lang L. A.; Khoze V. A.; Ryskin M. G. Exclusive LHC physics with heavy ions: SuperChic 3. *The European Physical Journal C*, Springer Science and Business Media LLC, vol. 79, n. 1, January 2019, ISSN 1434-6052, doi: 10.1140/epjc/s10052-018-6530-5.

[41] Particle Data Group. Review of Particle Physics. *Phys. Rev. D*, American Physical Society, vol. 98, n. 3, p. 562, August 2018, doi: 10.1103/PhysRevD.98.030001.

[42] ATLAS Collaboration. Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC. *Nature Physics*, Springer Science and Business Media LLC, vol. 13, n. 9, p. 852–858, August 2017, ISSN 1745-2481, doi: 10.1038/nphys4208.

Appendix A

Acronyms

CERN Conseil Européen pour la recherche nucléaire

ATLAS A Toroidal LHC Apparatus

ALP Axion-like particle

AFP ATLAS Forward Proton

Appendix B

Contents of enclosed SD card

- `readme.txt` ..... the file with SD card contents description
- `thesis.pdf` ..... the thesis text in PDF format
- `thesis` ..... the directory of LaTeX source codes of the thesis
- `implementation` ..... the directory of the implementation
  - `AFP` ..... the directory with tools needed for the experiment analysis
    - `notebooks` ..... the directory with Jupyter notebooks
    - `src` ..... the directory with source codes
    - `root` ..... the directory with .ROOT files
  - `simulation` ..... the directory with results from the simulations
    - `LHEF` ..... the directory with simlation results in LHEF format
    - `notebooks` ..... the directory with Jupyter notebooks
    - `src` ..... the directory with source codes