



Improved machine learning models for local optics corrections in the LHC triplet magnets

Master thesis
presented by

Elene Kravishvili

Internship carried out at
CERN

Under the supervision of
Félix Simon Carlier

Master degree of Physics
Erasmus Mundus LASCALA

Abstract

The thesis aims to develop a beam optimization process in LHC, particularly the correction of the triplet magnet at each interaction point. This study will construct a machine learning algorithm to predict the quadrupole field errors around IPs, making the correction process more precise and efficient. The models include two types of features: phase advances and β -beating around IPs. The impact that each feature has on precise predictions will be explored. The thesis also explores models' robustness while incorporating the various hyperparameters, compares their global and local performances, and assesses their ability to make local corrections. Models are also trained to predict different error sources, specifically energy offsets for both beams of LHC. Studies have shown that the triplet field errors are predicted with the high precision, especially while including β -beating around IPs into the features. Prediction of energy offset showed that it can be predicted by the model very well, and moreover, can develop the field error precisions since these errors are presented in the real-time beam. Studies have shown that incorporating β -beating makes models very robust in the global case, but this is no longer the case locally. The beam's performance response to the different hyperparameters does not match in the global and local cases, which motivates further studies and suggests incorporating a local approach into the training process.

Contents

1	Introduction	2
2	Large Hadron Collider	5
3	Theory	8
I	Introduction of beam dynamics	8
I.1	Courant-Snyder parameterization	11
I.2	Phase space and beam emittance	12
I.3	Luminosity	13
II	Perturbation	13
4	Machine Learning concepts	16
5	Measurement techniques, analysis and corrections	20
I	Data acquisition	20
II	Optics analysis	21
III	Correction techniques	24
6	General framework and methodology	26
7	Results	29
I	Triplet prediction and β -beating	31
II	Energy offset prediction	35
III	Hyperparameters tuning	38
IV	Validation with segment-by-segment simulations	42
8	Conclusion	50

Chapter 1

Introduction

“What we require is an apparatus to give us a potential of 10 million volts which can be safely accommodated in a reasonably sized room and operated at a few kilowatts of power. We require too an evacuated tube capable of withstanding this voltage... I see no reason why such requirements cannot be made practical” (Rutherford, 1930) [1, p. 444].

This statement represents one of the earliest ideas for particle accelerators. After that, particle accelerators developed gratefully. Particle accelerators are complex machines which accelerates particles to very high velocities, to achieve energies necessary for a variety of scientific and practical applications. The Livingston plot, shown in Fig.1.1 illustrates the progression of particle accelerators over the years, along with their development through increasing center of mass energy.

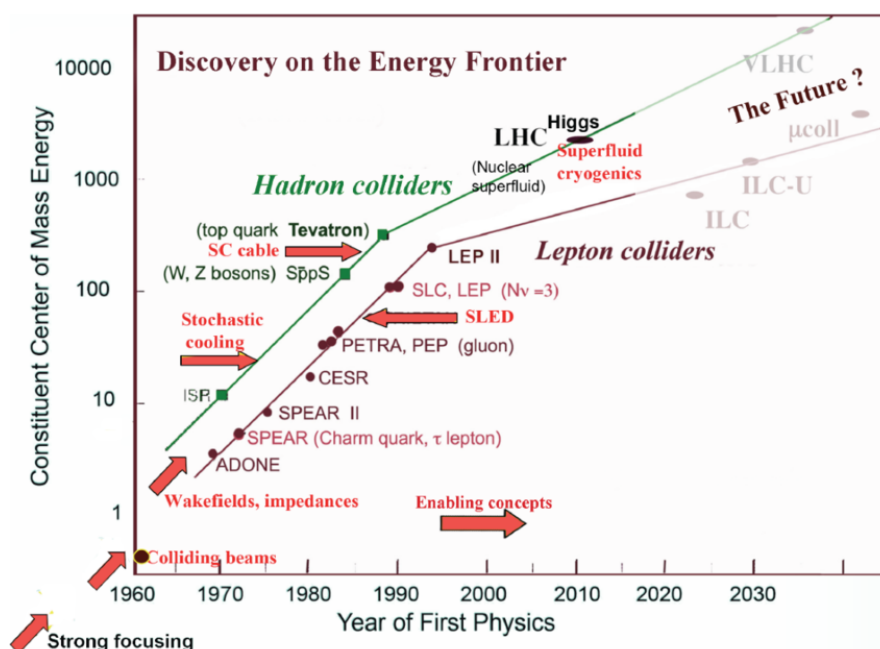


Figure 1.1: The Livingston plot illustrates the historical progression of particle accelerators by charting the center-of-mass energy against the physics results achieved that year. This figure is adapted from [2].

Their applications cover topics from high-energy physics to medical science, materials science, and industrial processes. Accelerators are essential in high energy physics research for exploring fundamental particles and validating the standard model, but they also serve as powerful light source for synchrotron radiation facilities, which are very important to advance imaging techniques in material science and biology. Additionally, they play a huge role in medical science for cancer treatment

through radiation therapy.

There are different types of accelerators, such as linear accelerators (linacs), cyclotrons, synchrotrons, and colliders. Colliders accelerate particles and collide them, resulting in particle interactions, which offers deep insights into the fundamental forces. The most advanced accelerator today is the Large Hadron Collider (LHC)[3] at CERN. Its most celebrated achievement, the discovery of the Higgs boson in 2012 [4] [5].

The LHC is a highly complex machine which needs precise control over its optics and it needs the significant preparation before collisions. It is crucial to ensure that the operation and performance parameters are achieved and maintained as they were designed. Unexpected events, which can easily occur due to the thousands of elements in the machine can damage it, even if they are minor, and must be avoided. It is also essential to ensure that each element among many is accurately calibrated. LHC must reproduce the same experimental results in repeated trials and provide the desired luminosity for collisions.

One of the most significant challenges in operating the LHC is maintaining a stable beam, which is crucial for the successful experiment. The first step in the LHC operation cycle is making linear optics corrections, which are adjustments to the magnetic fields and alignment of components, as well as keeping correct focus and shape of the beam [6] [7] [8] [9] [10]. This process should ensure that the desired parameters are maintained, such as position, size, and intensity, throughout the entire accelerator.

Numerous studies and research efforts focus on the measurement and correction of the beam dynamics and in particular linear optics. One common approach of measurement involves extracting data from Beam Position Monitors (BPMs) and analyzing it using Fourier Transform (FT), referred to as turn-by-turn techniques. Another is K-modulation, which means varying the quadrupole strength to calculate the β -function [11] [12] [13] [14]. Another approach is segment-by-segment (SBS) technique [15] [16] [17], relevant for the interaction points of LHC. These regions are highly sensitive to errors and SBS techniques are used to localize measurements and corrections.

While these methods are effective, there is always a need for improvement. As we look toward the future, enhancing these correction techniques becomes very important. The upcoming High Luminosity LHC (HL-LHC) upgrade and the potential construction of the Future Circular Collider (FCC) represent the next stage in accelerator technology. Developing the optimizing techniques for the beam will be crucial for these upcoming projects.

One of the most innovative and promising approaches to beam optimization today is the use of machine learning (ML) techniques. Machine learning is a branch of artificial intelligence (AI) with which computers are developed to make predictions depending on the learning from the data. Instead of following fixed instructions, ML algorithms analyze data and its patterns and improve over time. This technology has already proven to be incredibly useful in many different areas and shows great potential for advancing scientific research. It can handle complex tasks that are difficult or even impossible for humans to do manually. Various ML techniques, like linear regression, neural networks, and deep learning, have been developed and are showing great promise for a wide range of applications in science and engineering.

Recent studies have shown the potential of machine learning to greatly improve the way we manage and optimize particle accelerators. For example, ML models have been successfully used to predict errors in magnets based on measured optics parameters, with promising results [18].

However, further study of these techniques is essential to realize their full potential and make sure their effectiveness. This thesis has two main goals. The first is to develop machine learning techniques that focus on the local performance of the beam by correcting magnetic field errors. Here, "local" refers to the interaction region sections where each detectors are located. These sections require precise luminosity control, and the beam is highly sensitive to errors in these areas, particularly because the β -functions are very large before the collision. This research aims to enhance the robustness of the model, refine its feature configuration, tune the hyperparameters and focus on the beam's local performance in the desired sections while maintaining overall good results.

The second objective is to extend the ML model to predict and correct other types of errors

beyond magnetic field deviations, specifically energy offsets. These offsets have been a critical part of the correction process and have traditionally been corrected based on empirical methods. Therefore, it is crucial to develop an approach that improves the techniques for correcting these error sources. These two studies will be presented in this thesis.

This thesis is organized as follows. It begins with an overview of the LHC, including a detailed explanation of its role and operational cycle. The next chapter provides a comprehensive analysis of the beam dynamics within the context of accelerators and colliders, followed by a detailed explanation of the measurement and correction techniques currently used in the LHC. Subsequently, an overview of machine learning techniques and an in-depth explanation of relevant concepts are provided. The thesis then proceeds with a chapter that outlines the step-by-step approach to the problem, including the study process and the proposed solutions, followed by results that highlight the effectiveness of the models.

Chapter 2

Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest particle accelerator, serving as the final machine in CERN's accelerator complex as shown in Fig. 2.1. Initially, negative hydrogen ions (H^-) are accelerated in Linac4 to an energy of 160 MeV. These ions are then stripped of their electrons in the Proton Synchrotron Booster (PSB) and accelerated to 2 GeV. The protons then continue to the Proton Synchrotron (PS), reaching 26 GeV, and to the Super Proton Synchrotron (SPS), where they are boosted to 450 GeV. Finally, in the LHC, the protons are accelerated to a maximum energy of 6.8 TeV per beam.

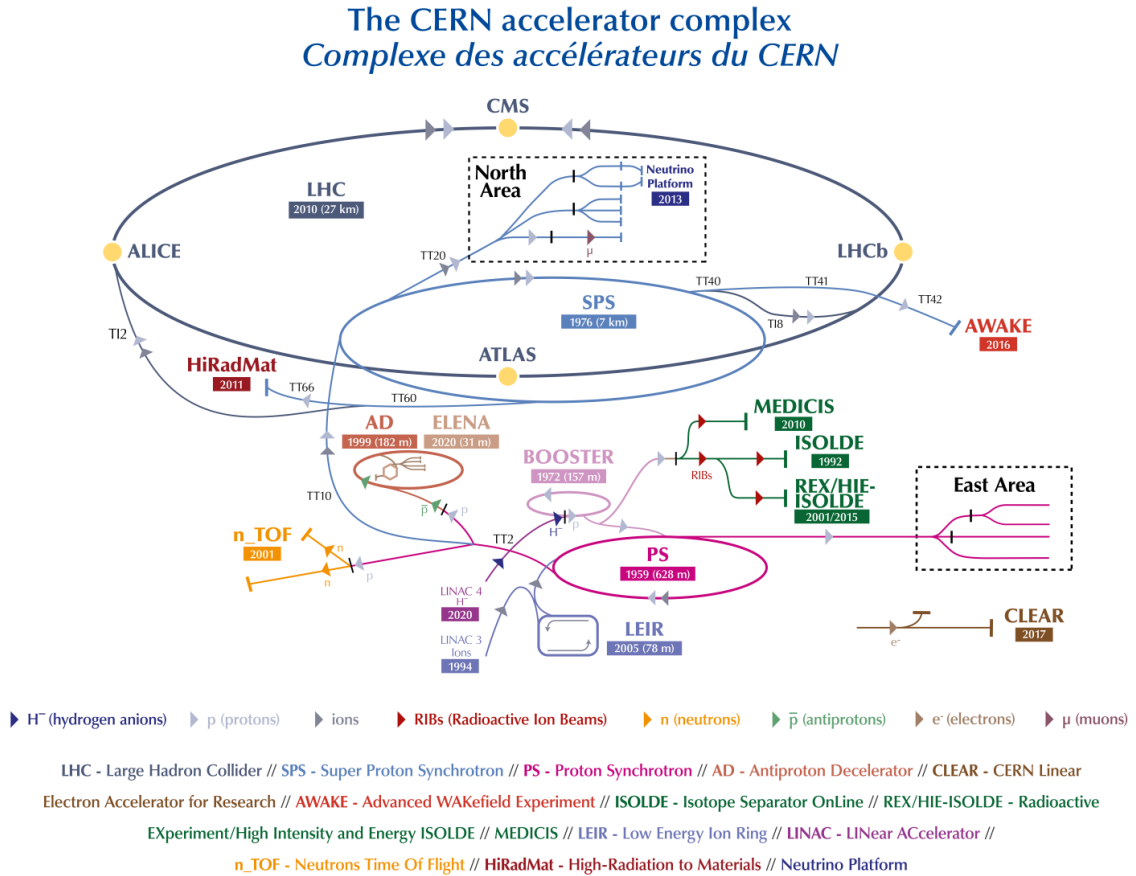


Figure 2.1: The CERN accelerator complex. [19]

In the LHC, two beams are circulated in opposite directions, which are directed to collide at four interaction points (IPs). Each proton beam in LHC consists around 2400 bunches, with each bunch initially containing around 1.6×10^{11} protons. The primary role of the LHC is to collide these two

beams with high precision, yielding high luminosity for the particle detectors, located at each IP: ATLAS, CMS, ALICE, and LHCb as shown in Fig. 2.2. ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are versatile detectors designed to explore a wide range of physics phenomena, including the search for the Higgs boson and potential new particles. ALICE (A Large Ion Collider Experiment) is dedicated to studying the quark-gluon plasma, a state of matter that existed shortly after the Big Bang. LHCb (Large Hadron Collider beauty) focuses on understanding the differences between matter and antimatter by investigating b-quarks (bottom quarks).

The LHC ring's circumference is 27 kilometers, and it includes thousands of accelerator components. These are dipoles for bending the beam, quadrupoles for focusing it, and RF cavities for acceleration. Other crucial elements are collimators, vacuum systems, and cryogenics; they also play vital roles. Each component ensures precise control of the beam, keeping it on course and preventing deviations that could potentially damage the machine.

Operation

The commissioning of the LHC begins with the pre-cycle phase, during which the magnetic states of the machine are re-established prior to the filling process. Following the pre-cycle, a pilot bunch, consisting of one of few (approximately ten) bunches, is introduced to verify the proper setup of the machine. Once verification is complete, the optics cycle begins with the injection phase, where two LHC beams are progressively filled with bunches from Super Proton Synchrotron (SPS). After the completion of the filling phase, the next stage, known as the ramp, starts. During the ramp, the bunches are accelerated to the target energy of 6.8 TeV, reaching a state of flat top. Subsequently, the process moves to the squeeze phase, during which the beams are brought into collision. To ensure the beams provide optimal results for the particle detectors by maximizing the number of collisions, it is crucial to minimize the beam size [20]. For that, at each interaction point, the beam size is initially increased significantly and then squeezed tightly for collision. This process is facilitated by two sets of triplet quadrupoles, with three magnets positioned around each side of the IPs. The current squeeze optics achieve a β^* (beta function at the collision point) of 30cm . Lastly, the beams are extracted through IR6 of the LHC, which is the beam dumping system (see Fig. 2.2). That directs the beams to an external absorber while ensuring the effective beam dispersion [21].

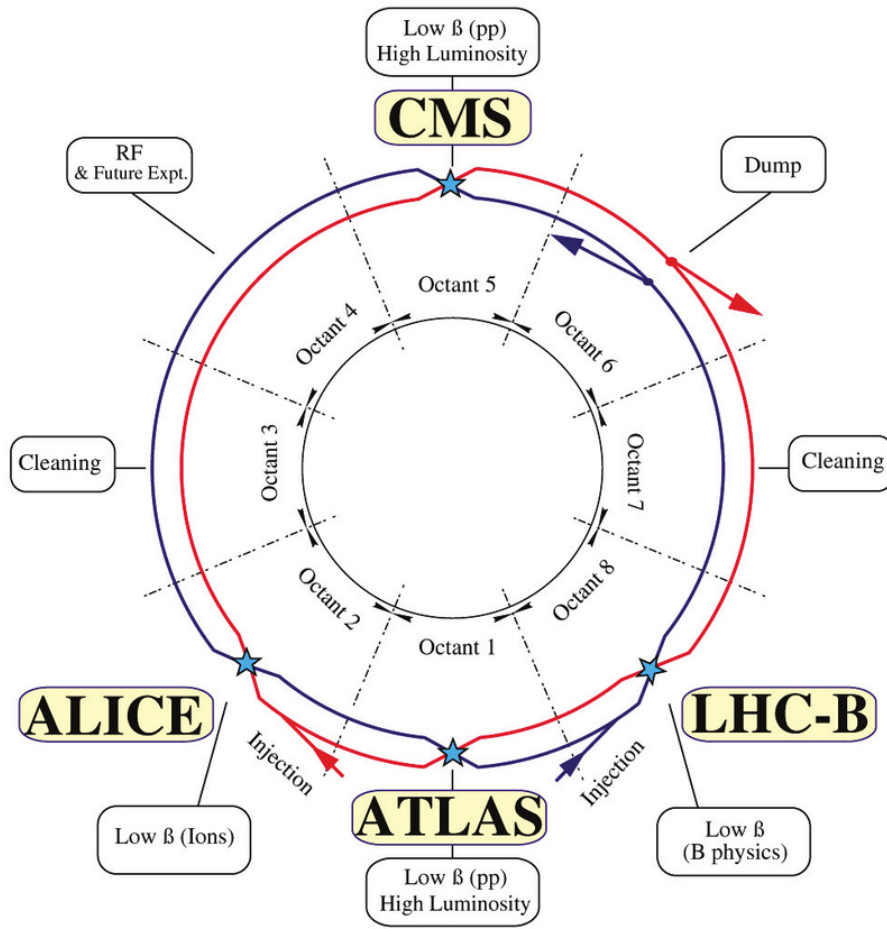


Figure 2.2: Layout of the LHC, highlighting the four particle detectors: ATLAS, CMS, ALICE, and LHCb. Image from[22]

Chapter 3

Theory

In this chapter, the theoretical aspects which are essential to this research presented are explored. The primary theoretical insights depends on the books by S.Y. Lee [23], A. Wolski [24], and H. Wiedemann [25]. The discussion begins with an overview of linear beam dynamics. Following this, the effects of magnetic field perturbations in quadrupole magnets are analyzed and the particular attention to the impact of misalignments. The next chapter introduces measurement and correction techniques used in the Large Hadron Collider. Following that, a dedicated chapter will explore the machine learning techniques relevant to the research work presented in this thesis.

I Introduction of beam dynamics

The motion of particles in accelerators is governed by the Lorentz force (Eq. 3.1). This force is due to the interaction of charged particles with magnetic and electric fields.

$$F = q \left(\vec{E} + \vec{V} \times \vec{B} \right) \quad (3.1)$$

In high energy particle accelerators particles move at velocities almost equal to the speed of light. At these relativistic speeds, the magnetic field's influence dominates, therefore the force from the magnetic field is higher. Magnetic fields are more effective for steering the beam, while electric fields are there to accelerate the particles. The momentum of particles in a magnetic field can be expressed as 3.2.

$$p = qB\rho \quad (3.2)$$

Here, ρ denotes the bending radius, and the term $B\rho$, known as magnetic rigidity, is a key characteristic of particle dynamics in accelerators.

In circular accelerators, the very important designed parameter is reference orbit. In the case of synchrotrons this reference orbit is closed orbit and it is defined with the dipole magnets. Closed orbit is the reference path thorough which the particles are moving. The dynamics of the particle are analyzed around to this orbit. For this analysis reference frame, the Frenet-Serret coordinate system is employed, shown in the Fig. 3.1. Which is particularly useful as it moves along with the particles. In this coordinate system, s denotes the propagation direction of the beam, while x and y correspond to the horizontal and vertical planes, respectively.

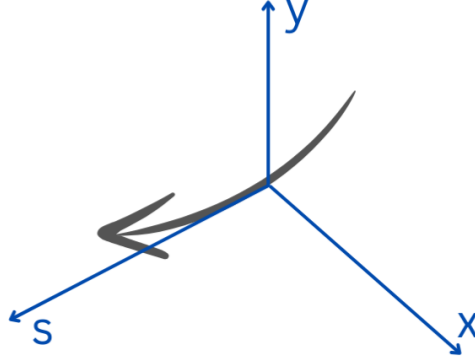


Figure 3.1: The Frenet-Serret coordinate system used in particle accelerators, consisting of the tangent vector along the particle's path and two perpendicular transverse vectors, denoted as s , x , and y , respectively.

The dynamics of particle motion in this coordinate system are described by Hill's equations 3.3.

$$\begin{aligned} x''(s) + K_x(s)x(s) &= 0 & K_x(s) &= \frac{1}{\rho^2} - K_1(s), \\ y''(s) + K_1(s)y(s) &= 0 & K_y(s) &= K_1(s) \end{aligned} \quad (3.3)$$

Here, $K_1 = \frac{1}{\rho^2}$ represents the focusing strength of the magnet, and ρ is the bending radius. The solution to Hill's equation can be expressed using matrix formalism. As particles travel through different magnetic elements in the accelerator, the effect of each element can be determined with this formalism.

$$\mathbf{z}(s) = M(s|s_0) \mathbf{z}(s_0), \quad (3.4)$$

$$\text{Where } \mathbf{z}(s) = \begin{pmatrix} z(s) \\ z'(s) \end{pmatrix} \quad (3.5)$$

Where $M(s|s_0)$ is a transfer matrix that relates the initial and final transverse position and the angle from position s_0 to s . This matrix formalism is applied to various elements within the accelerator.

In particle accelerators, the linear dynamics in the transverse plane result from the combined effects of drift spaces, dipole magnets, and quadrupole magnets. A drift space is a section where no magnetic field is present. Dipole magnets are components used to guide the particles through the circular accelerator. These magnets create a uniform magnetic field with parallel field lines that bend the particles along its path. Quadrupole magnets, in contrast, are responsible for focusing the beam. Their magnetic field lines have a hyperbolic pattern, with a focusing effect in one plane and a defocusing effect in the orthogonal plane, as shown in Fig. 3.2.

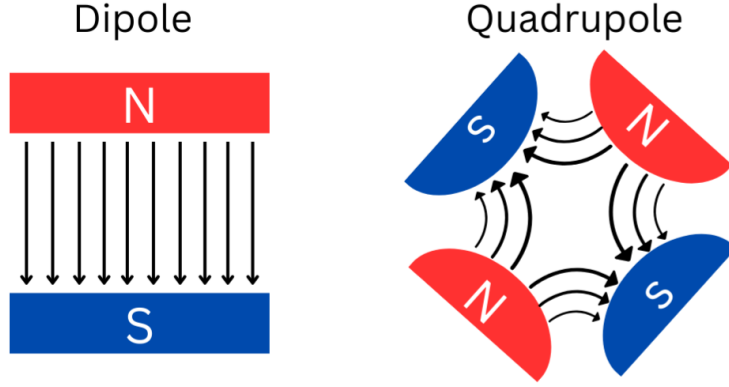


Figure 3.2: The figure illustrates the configurations of dipole and quadrupole magnets, showing the distribution of magnetic field lines in each.

Drift space

In the drift space there is no magnetic field, that implies that the magnetic field coefficient $K(s)$ is zero. Consequently, the transfer matrices takes the following form:

$$M(s|s_0) = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \quad (3.6)$$

Where the element length is defined as $l = s - s_0$.

Dipole

Specifically, for a dipole magnet where $K_1(s) = 0$, the transfer matrix is given by:

$$M(s|s_0) = \begin{pmatrix} \cos \theta & \rho \sin \theta \\ -\frac{1}{\rho} \sin \theta & \cos \theta \end{pmatrix}, \quad (3.7)$$

Where $\theta = \frac{l}{\rho}$.

Quadrupole

In the case of a quadrupole magnet, where $\frac{1}{\rho^2} = 0$ and $K_1(s)_x = -K_1(s)_y$, the transfer matrices are represented as:

$$M(s|s_0) = \begin{pmatrix} \cos \sqrt{K}l & \frac{1}{\sqrt{K}} \sin \sqrt{K}l \\ -\sqrt{K} \sin \sqrt{K}l & \cos \sqrt{K}l \end{pmatrix} \quad (3.8)$$

And for the other defocusing plane with $K_1 = -|K_1|$ the transfer matrix is:

$$M(s|s_0) = \begin{pmatrix} \cosh \sqrt{|K|}l & \frac{1}{\sqrt{|K|}} \sinh \sqrt{|K|}l \\ \sqrt{|K|} \sinh \sqrt{|K|}l & \cosh \sqrt{|K|}l \end{pmatrix} \quad (3.9)$$

The transfer matrix for a beam segment can be expressed as the multiplication of each matrix corresponding to the elements in the segment. For the complete turn around the accelerator, all the effect from each element should be considered and can be represented by the matrix, which is usually called one turn map, denoted as M_{OTM} as in Eq. 3.10.

$$\mathbf{z}(s_0 + C) = M_N \cdot M_{N-1} \cdot \dots \cdot M_2 \cdot M_1 \mathbf{z}(s_0) = M(s_0 + C | s_0) \mathbf{z}(s_0) \quad (3.10)$$

The most common beamline segment in accelerators is the FODO lattice. FODO refers to focusing (F) and defocusing (D) quadrupoles separated by drift spaces (O), which comprises alternating focusing quadrupoles along with drift spaces, often filled with the dipoles. This structure ensures overall stability by compensating for the focusing and defocusing effects. Figure 3.3 illustrates the β -functions on the x and y directions. There also is shown the quadrupole locations and their field strengths. As it is seen field strength are changing the sign, that indicates that the magnets are changing their focusing and defocusing properties. The envelope of the beam exhibit the same periodicity.

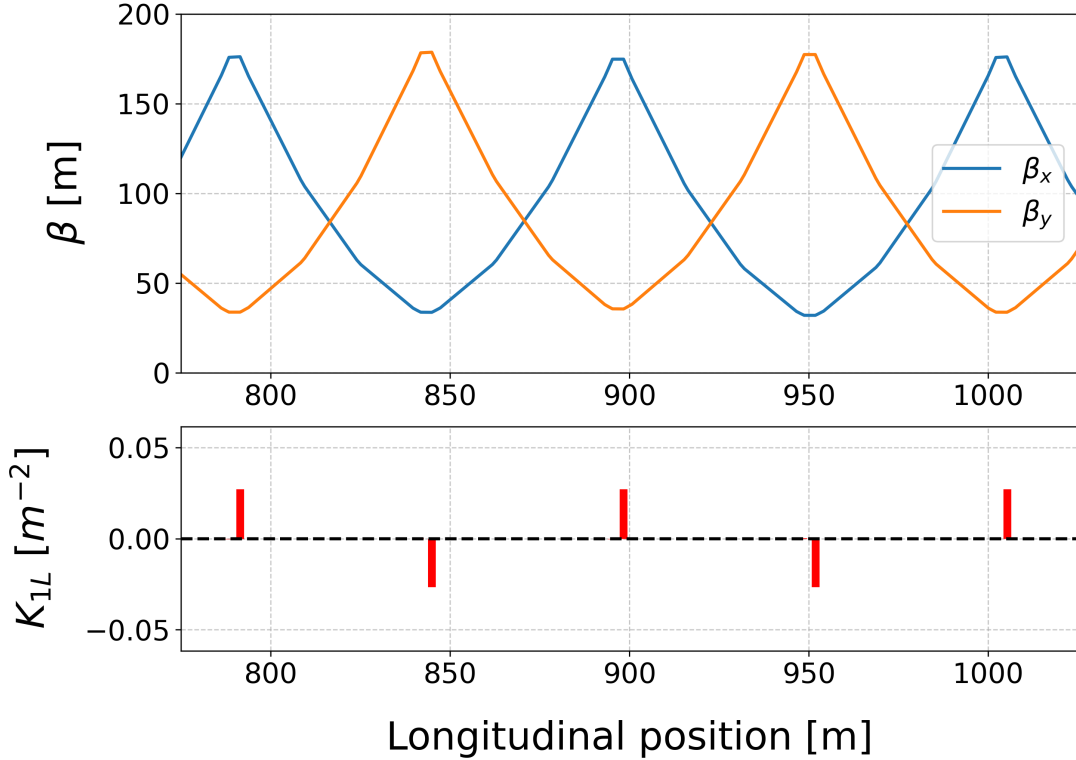


Figure 3.3: The upper plot is representing β -functions of the beam on the horizontal (in blue) and vertical (in yellow) planes, while the lower plot illustrates the corresponding quadrupole field strength distribution in the LHC FODO lattice.

I.1 Courant-Snyder parameterization

When considering the periodic structure of the accelerator lattice, it is often assumed that the function $K(s)$ exhibits the same periodicity as the lattice period L . This can be expressed as $K(s) = K(s + L)$ [26]. A more general solution of Hill's equation can be represented by introducing the Courant-Snyder parameterization. The transverse displacement $x(s)$ can be expressed as Eq. 3.11.

$$x(s) = A(s) \cos(\phi(s) + \phi_0), \quad (3.11)$$

Where $A(s)$ is the amplitude function, $\phi(s)$ is the phase function, and ϕ_0 is the initial phase.

The Courant-Snyder parameters are defined as in 3.12.

$$\begin{aligned}\beta(s) &= \frac{A(s)^2}{\epsilon}, \\ \alpha(s) &= -\frac{1}{2} \frac{d\beta(s)}{ds}, \\ \gamma(s) &= \frac{1 + \alpha(s)^2}{\beta(s)}.\end{aligned}\tag{3.12}$$

Where ϵ is the beam emittance, which is the area occupied by a beam of particles in phase space, describing the beam's spread. $\beta(s)$ characterizes the transverse size of the beam, and the physical transverse beam size is given by $\sqrt{\epsilon\beta(s)}$.

Using these parameters, the transfer matrix $M(s_0, s)$ can be expressed as as 3.13.

$$\begin{pmatrix} \sqrt{\frac{\beta(s)}{\beta(s_0)}} \cos \Delta\phi + \alpha(s_0) \sqrt{\frac{\beta(s)}{\beta(s_0)}} \sin \Delta\phi & \sqrt{\beta(s)\beta(s_0)} \sin \Delta\phi \\ -\frac{1}{\sqrt{\beta(s)\beta(s_0)}} [(\alpha(s_0) - \alpha(s)) \cos \Delta\phi - (1 + \alpha(s)\alpha(s_0)) \sin \Delta\phi] & \sqrt{\frac{\beta(s_0)}{\beta(s)}} \cos \Delta\phi - \alpha(s) \sqrt{\frac{\beta(s_0)}{\beta(s)}} \sin \Delta\phi \end{pmatrix}\tag{3.13}$$

where $\Delta\phi = \phi(s) - \phi(s_0)$ is the phase advance of the betatron oscillation from s_0 to s . The phase at position s is defined as Eq.3.14.

$$\phi(s') = \int_0^{s'} \frac{1}{\beta(s)} ds.\tag{3.14}$$

The phase advance over one revolution is defined as the tune , shown in Eq. 3.15.

$$Q = \frac{1}{2\pi} \oint \frac{1}{\beta(s)} ds.\tag{3.15}$$

I.2 Phase space and beam emittance

In linear dynamics, particle distribution when it is represented in the phase space, in (x, x') space, is the ellipse. The parameters of this ellipse are very well described by the Courant-Snyder parameters as it is visualized in Fig.3.4. Here the α , β and γ are used to characterise the dimensions and configuration of the ellipse. When there is no influence of conservative forces, the area of the ellipse stays constant, and is related to the beam emittance with the relation 3.16.

$$A = \pi\epsilon_u\tag{3.16}$$

The beam emittance ϵ is defined as in Eq. 3.17.

$$\epsilon = \gamma(s)x(s)^2 + 2\alpha(s)u(s)x'(s) + \beta(s)x'(s)^2.\tag{3.17}$$

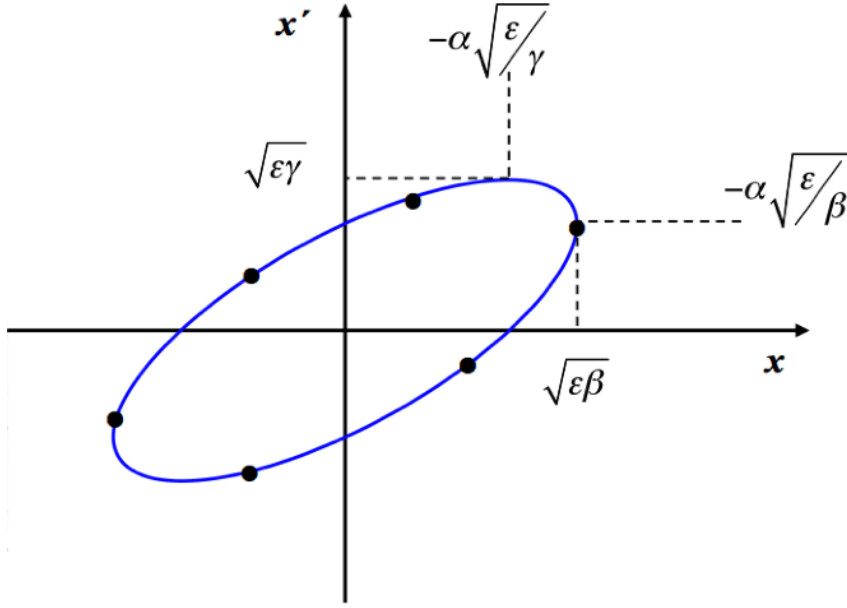


Figure 3.4: Phase space ellipse in the x - x' plane, with axes defined by $\sqrt{\epsilon\beta}$ and $\sqrt{\epsilon\gamma}$, and tilt determined by α . Figure taken from [27].

I.3 Luminosity

The luminosity L of a particle collider is given by Eq.3.18. It is the one of the main characterisation important for the detectors in the LHC.

$$L = \frac{N_1 N_2 \cdot R \cdot \omega_{fr}}{N_b \cdot 4\pi\sigma_x\sigma_y} \quad (3.18)$$

Where N_1 and N_2 are the intensities of beam 1 and beam 2, respectively. N_b is the number of bunches in each beam. R is the reduction factor, ω_{fr} is the revolution frequency of the beams, σ_x and σ_y are the horizontal and vertical beam sizes at interaction points.

II Perturbation

Linear magnet imperfections, resulting from both dipole and quadrupole field errors, can induce two significant effects in particle accelerators [25]. These effects are the distortion of the closed orbit and the perturbation of the betatron amplitude. Both phenomena will be examined in the subsequent sections.

Closed orbit distortion

The first effect is closed orbit distortion, which is mainly an effect in the dipole magnets. When there is a deviation in the magnetic strength of the dipole field, particles are deflected from the reference trajectory as they pass through the dipole, causing the closed orbit to no longer coincide with the reference trajectory. This error can also arise from misaligned quadrupoles. Particles not centered in the orbit experience a feed-down effect from the magnet. In the quadrupole, they encounter a dipole field, affecting particles like dipole field errors.

Focusing errors

The second effect is the perturbation of the betatron amplitude function, which is chiefly caused by quadrupole field errors, can also be referred to as focusing errors. They occur when the magnetic field gradients of quadrupole magnets deviate from their nominal values, leading to these focusing errors. One can show that the change in the β -function due to a quadrupole error is proportional to the error strength Δk , as shown in Eq.3.19, the β -beating, relative deviation of β -function from the nominal value is shown in the Fig.3.5.

$$\frac{\Delta\beta(s)}{\beta(s)} = -\beta(s_0)\Delta k \sin 2\phi + \beta(s_0)^2 \Delta k^2 \sin^2 \phi \quad (3.19)$$

Where $\Delta\beta(s) = \beta_{measured}(s) - \beta(s)$, $\beta(s)$ is nominal beta function.

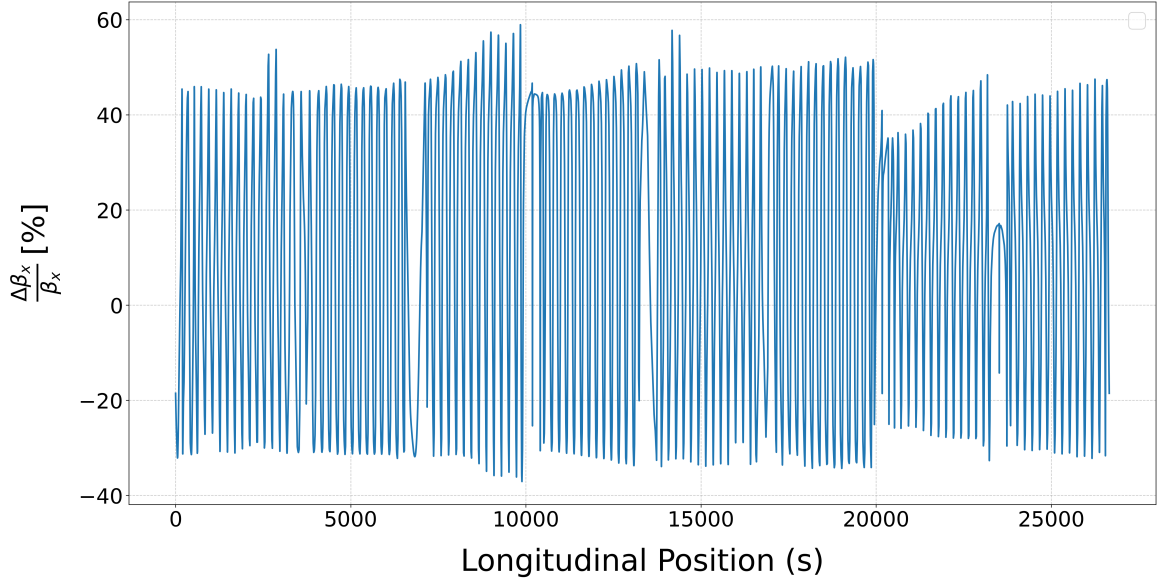


Figure 3.5: The figure illustrates the resulting β -beating, $\frac{\Delta\beta}{\beta}$, when the beam is not optimized and includes errors in various elements. The x-axis represents the longitudinal position along the LHC beam, while the y-axis shows the percentage deviation of the β -function.

Energy offset

An energy offset refers to the deviation of the beam's energy from its nominal or design value. An energy offset in the accelerator can also cause β -beating. This phenomenon can be easily confused with errors coming from quadrupole magnets, as both result in similar distortions in the beam's transverse position. Distinguishing between β -beating due to energy offset and that due to quadrupole errors can be challenging. β -function beating caused by this error source is shown in the Fig.3.6.

Misalignment

Misalignments errors are primarily due to the physical misalignment of the magnets themselves. Longitudinal misalignment is when the quadrupole displaces along the beam axis, badly effects the synchronization between the magnet's field and the particle beam. This misalignment can also lead to improper focusing or defocusing. Transverse misalignment is when the quadrupole is displaced

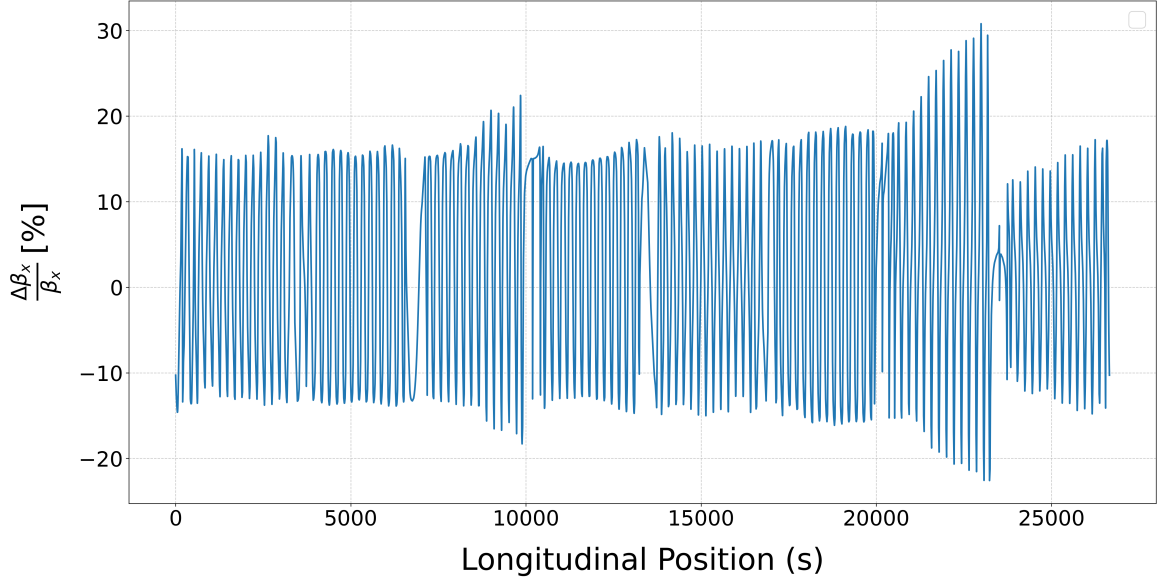


Figure 3.6: This figure illustrates the β -beating in the beam induced by an energy shift, assuming no other sources of error are present. The plot demonstrates that the relative change in the β -function, $\frac{\Delta\beta}{\beta}$, caused by this type of error exhibits behavior similar to that observed from magnet errors.

perpendicular to the beam axis. This effect causes lower-order multipole components such as dipole fields, often called feed down effect. These can create unwanted magnetic forces on the beam, which will deviate it from its intended path. Rotational misalignment of a quadrupole magnet, where the magnet rotates around its axis, introduces a rotated quadrupole component to the magnetic field. β -beating cause by this effect can be expressed with Eq.3.20. [28]

$$\frac{\Delta\beta_{x,y}(s)}{\beta_{x,y}(s)} = B_{x,y} \cos[2\pi Q_{x,y} \mp 2\mu_{x,y}(s) \pm \varphi_{x,y}] \quad (3.20)$$

Where,

$$B_{x,y} = \frac{K\Delta s}{2\sin(2\pi Q_{x,y})} \sqrt{\beta_{x_1,y_1}^2 + \beta_{x_2,y_2}^2 - 2\beta_{x_1,y_1}\beta_{x_2,y_2} \cos(2\Delta\mu_{x,y})} \quad (3.21)$$

Chapter 4

Machine Learning concepts

Machine learning is a branch of artificial intelligence. Depending on it, computers learn from data and make predictions and decisions. The machine learning process involves training models on data, which define patterns and can apply this generalization to new, unseen data. Machine learning models include two main categories: supervised learning and unsupervised learning.

Supervised learning is a machine learning model in which the data used during training includes input data with the corresponding output data. Commonly called features and targets, creating the sample. The model studies and defines the correlation between the input and output pair and can predict the relevant targets for new data. Linear models, such as linear regression and logistic regression, are common in supervised learning. They predict the output and classify the data. Linear regression is used to predict the values by finding the linear relationship between input features and the target variable. Logistic regression is used for binary classification problems. It estimates the probability that a specific input belongs to a particular category. Techniques like Ridge and Lasso regression add regularization to linear models, which helps prevent overfitting by including a penalty on very complex models.

Unsupervised learning, in contrast to a supervised model, is trained with unlabelled data. It identifies patterns within the data and structures them. These learning models include K-modulation, clustering, hierarchical clustering, anomaly detection, and many others. In this thesis, a supervised learning algorithm will be used.

The following sections will provide a detailed explanation of linear regression and Ridge regression.

General Linear Model

A linear model is a supervised learning algorithm that learns from the data to define the relationship and dependence between dependent and independent variables. The learning process includes defining the coefficients of dependency, and the model with this coefficient predicts the variable from the new data. This is shown in Eq.4.1.

$$\hat{y} = w_0 + \sum_{i=1}^j w_i x_i, \quad (4.1)$$

Where $\hat{y}$ is the predicted dependent variable, w_i are the coefficients for each independent variable, w_0 is called the bias term, x_i are the independent variables, commonly called features, and j is the number of independent variables.

$\hat{y}$ can be a scalar as well as a vector. In cases where $\hat{y}$ is a vector of size k , the weights w_i will be a matrix of size $j \times k$.

During the learning process, the best-fitting line is found by minimizing the sum of squared

differences between the predicted values and actual values. With this best-fitted line, model will be able to provide accurate predictions with the new, unseen data. This process includes updating the coefficients by calculating the residual sum of squares (RSS), as shown in Eq.4.2.

$$RSS = \sum_{m=1}^N (y_m - \hat{y}_m)^2, \quad (4.2)$$

Where N is the number of observations, y_m is the actual value of the dependent variable for the m -th observation, $\hat{y}_m$ is the predicted value of the dependent variable for the m -th observation.

Linear model looks for the best-fitted coefficients by minimizing the residual sum of squares. The optimal values of coefficient will be able to best describe the relationship between independent and dependent variables. To minimize the RSS, gradient descent iteratively updates the coefficients. The gradient (partial derivative) of the RSS with respect to each coefficient w is given by Eq. 4.3.

$$\frac{\partial RSS}{\partial w_i} = -2 \sum_{m=1}^N (y_m - \hat{y}_m) x_{m,i} \quad (4.3)$$

Using these gradients, the coefficients are updated in the direction that reduces the RSS, as shown in Eq. 4.4.

$$w_i \leftarrow w_i - \eta \frac{\partial RSS}{\partial w_i} \quad (4.4)$$

Here, η is the learning rate, a very important hyperparameter that determines the step size of each update. The size of the learning rate is very important for training process, large learning rate could cause the algorithm to overshoot the minimum RSS, thus failing to converge. On the opposite side, if learning rate is too small the algorithm could require a very large number of iterations to converge, making it slow.

In the process of gradient descent the model repeatedly adjusts the coefficients to minimize the RSS. Iterations stop when the difference in RSS between consecutive iteration is below the desired threshold or when the maximum number of predefined iterations is reached. This iterative process ensures that the model approaches the optimal set of coefficients.

Loss Function for Regression

The loss function indicates the discrepancy between predicted dependent variable and actual dependent variable, therefore it is good indicator of how good the model is. For linear regression, the most commonly used loss function is the Mean Squared Error (MSE), which is an average of the squared differences between the predicted values and the actual target values. MSE is defined by Eq. 4.5.

$$MSE = \frac{1}{N} \sum_{m=1}^N (y_m - \hat{y}_m)^2 \quad (4.5)$$

Where N is the number of observations, y_m is the actual value of the dependent variable for the m -th observation, and $\hat{y}_m$ is the predicted value of the dependent variable for the m -th observation.

The main objective in regression is to minimize this loss function, thereby improving the model's accuracy. The smaller the MSE, the closer the model's predictions are to the actual values, indicating better model performance.

The reason MSE is preferred is because it penalizes large errors more than the smaller ones,

making the model less sensitive to outliers. However, depending on specific goals of analysis other loss functions like Huber loss is used.

Loss functions play an important role in the training process of regression models. The MSE is often used for its smoothness and simplicity, but alternative loss functions may be more appropriate depending on the characteristics of the data and the specific objectives of the regression task.

Overfitting

Generalization describes a models ability to make accurate predictions on unseen data, meaning the model could be applied to new relevant cases. To ensure effective generalization during the training process, the available data is typically divided into training and test datasets. However supervised learning is susceptible to overfitting. This happens when the model learns random noise and fluctuations as well as the general patterns in the training data. When overfitting occurs the model exhibits excessively high accuracy on the training data by fitting each data point with great precision, but performs poorly on test or unseen data. Thus indicating poor generalisation skills, despite excellent fitting on the training data. Such case is illustrated in the Fig. 4.1.

Several factors can contribute to overfitting. A high number of features relative to the number of observations can make the model overly complex, leading it to fit the noise in the training data and interpret random fluctuations as genuine patterns, resulting the overfitting. Moreover, conducting gradient descent for an excessive number of iterations can worsen overfitting, as the model continues to adjust to even minor variations in the training data.

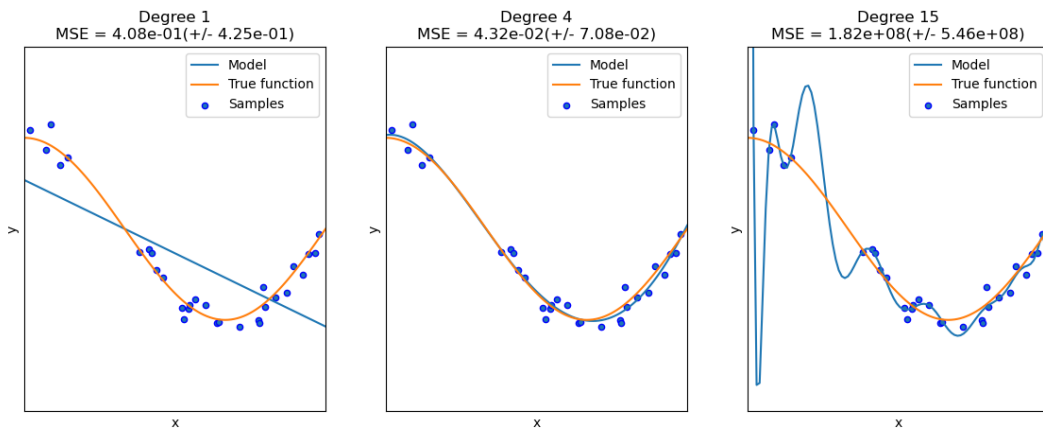


Figure 4.1: Illustration of machine learning model performance. The underfitting model (left), the good fit model (center) and the overfitting model (right). Figure taken from [29]

Regularized regression

A commonly used approach to solve the problem of overfitting is regularization. This works by imposing a penalty on the update of model parameters during the supervised training phase. By doing so regularization helps control these adjustments, making the model more stable against data variations. There are multiple regularisation techniques, one such technique is Ridge regression, a linear regression method that includes a regularization term to prevent overfitting by penalizing large coefficients. Ridge regression minimizes the RSS between the true target values in the training data and the values predicted by the model.

The regularisation coefficient (often denoted as α or λ) is a key component of Ridge regression. It effects the coefficients of the linear model and prevents them to have negative impact on training.

Small changes in the model or data can cause inappropriate updates to the coefficients during training. As the value of α increases, the "shrinkage" of the coefficient values also increases. The regression task is then formulated as minimizing the loss function, which now includes this regularisation term as shown in Eq. 4.6.

$$\min_{\mathbf{w}} \|\mathbf{w}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{w}\|_2^2 \quad (4.6)$$

Where $\mathbf{w}$ is the matrix containing the weights of the regression model, $\mathbf{x}$ is the input data vector, and $\mathbf{y}$ is the target vector. The last term helps to shrink the coefficients and stabilize the model against the outliers, especially when dealing with multicollinearity and high-dimensional data.

Tuning additional model parameters, such as the Ridge penalty (α) is typically done using cross-validation. In cross-validation, the training data is randomly divided into several equal parts, for example, ten subsets. The regression algorithm is then trained on nine-tenths of the data and the loss is computed on the remaining one-tenth.

The regularisation parameter will play an important role in the studies conducted in this thesis. Its impact on the specific case will be investigated to explore how the parameter influences models' performance and stability and its role in model's generalization ability.

Chapter 5

Measurement techniques, analysis and corrections

Effective measurement and correction techniques are essential for achieving precise control of a beam, ensuring optimal performance and the desired luminosity without deviations. Therefore, accurate optics measurements and corrections are crucial. These concepts will be elaborated upon in the following chapter.

I Data acquisition

Optics measurements are conducted on low intensity 'pilot' bunches ($\approx 10^{10}$ protons per bunch), to maintain machine safety. The primary instrument used for data acquisition in these measurements is the Beam Position Monitor (BPM), located at approximately 550 positions in the LHC.

Beam position monitors are looking to the transverse position of a particle beam at their locations in the accelerator beamline. They monitor particles location at each turn around the ring and with that information they are providing the data, which usually is called turn-by-turn data. Subsequently this data is analyzed to derive optical functions. Typically, BPMs are constructed with either two or four conductor plates, with two plates in opposite directions of each other in the horizontal position measurements and the remaining two for vertical position measurements. At the Large Hadron Collider dual-plate button-type BPMs are utilized. As the beam propagates within this devices, an electric charge is induced on each electrode and results a voltage difference between opposing electrodes. By measuring this voltage difference at each turn, the transverse position of the beam can be accurately determined, as described by Eq. 5.1.

$$y \approx \frac{w}{2} \frac{U_+ - U_-}{U_+ + U_-}, \quad (5.1)$$

Where U_+ and U_- are voltage signals from the opposite electrodes (right and left, or up and down) and $\frac{w}{2}$ is the effective width of the monitor. However, Eq. 5.1 could require further nonlinear calibration depending on the geometry of the pickup [30].

Before the measurement and taking the data from BPMs, the beam needs to be excited and create the betatron oscillation, which is the oscillation around the beam orbit. It is important that the amplitude of these oscillations is sufficient to ensure that the signal is larger than the BPM noise. Additionally, the oscillation needs to last long, so the good spectral analysis to be provided, at the same time measurement process should remain non-destructive. This beam excitation can be

achieved through two methods: with ac dipoles ([31], [32], [33], [34]) or using ac dipole-like excitation in the transverse damper [35].

An AC dipole is a dipole magnet with an oscillating current. It operates at a frequency close to the natural tune of the beam in the accelerator. To create the betatron oscillation, the field strength of the AC dipole is increasing, until the oscillation has the sufficient amplitude. Then it is at this level for the data acquisition and then it is ramped down. It is worth to note, that the change of the amplitude happens adiabatically, that ensures that transverse emittance is not effected while creating the coherent oscillation.

The second method to initiate the betatron oscillation is to use LHC transverse damper (ADT). This transverse damper can excite the beam for unlimited time, that was not the case for AC dipole. That makes us able to conduct the data acquisition for the unrestricted period of time. However the amplitude is not as high as in the case of AC dipole. It is because the limited excitation strength of the ADT.

On the data taken from the BPM the harmonic analysis should be done. But before that, the important step is the data to be cleared and denoised. This usually is achieved with singular value decomposition (SVD) based technique. With this method uncorrelated signals are removed from the data. After that the spectral analysis is done, for that Fourier Transform (FT) approach is used. The FT converts the time-domain signal into its frequency-domain representation, enabling the identification of dominant frequencies that correspond to the resonant tunes of the beam, as it is shown in the Fig. 5.1.

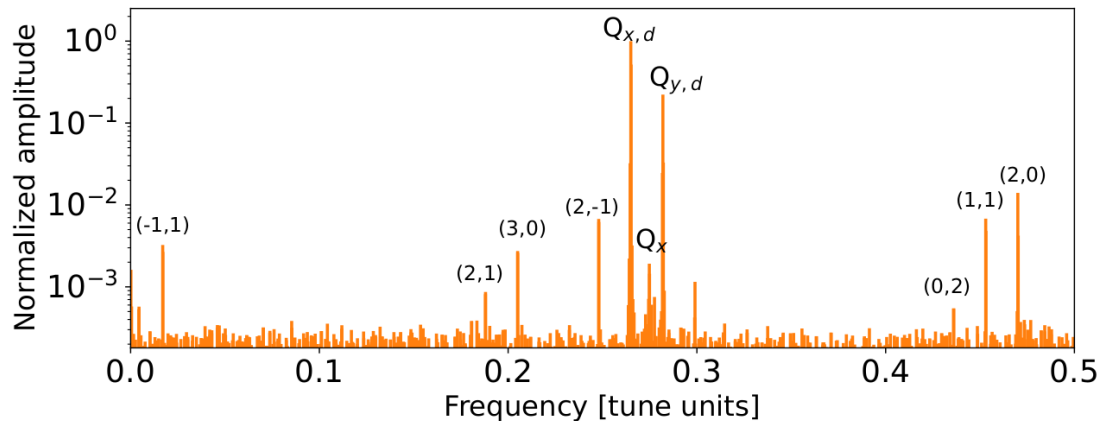


Figure 5.1: Spectral analysis of turn-by-turn data.

II Optics analysis

Calculation of the β function

From the data acquisition, the phase difference of the main mode between consecutive BPMs is obtained. It makes us able to calculate β -function at each corresponding BPMs. The value of this optical function can be derived in three ways. The first one is called 3-BPM method, second is the N-BPM method. And the last slightly different method is calculating β from the amplitude of oscillation.

In the 3-BPM method, three consecutive BPMs are used. The β -function is calculated using the phase advances between them, as shown in Eq.5.2.

$$\beta_i = \beta_{mod} \frac{\cot(\phi_{i,j}) - \cot(\phi_{i,k})}{\cot(\phi_{i,j})^{mod} - \cot(\phi_{i,k})^{mod}} \quad (5.2)$$

Here, the notation "mod" indicates the model values. i, j and k indicates three consecutive BPMs and ϕ - phase advance between them.

In Figure 5.2 the 3-BPM scheme is shown, the red triangle indicates the probed BPM, where the β -function is calculated using the depicted phase advances.

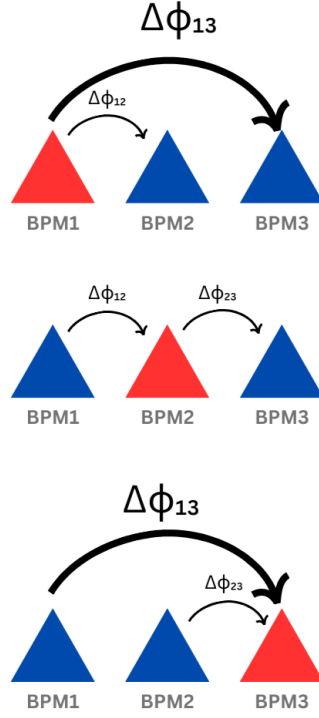


Figure 5.2: 3 BPM scheme

At the LHC, the phase advance between two consecutive BPMs is $\frac{\pi}{4}$. When the phase advance between the BPMs is close to 0, $\frac{\pi}{2}$, or π , the error in the beta function becomes significant. To avoid these unsuitable phase advances and improve measurement precision, the N-BPM method was developed. This method allows for skipping BPMs around the probed one, and the beta function is derived as shown in Eq. 5.3.

$$\beta_l(s_i) = \beta(s_i)_{mod} \frac{\cot(\phi_{i,j,l}) - \cot(\phi_{i,k,l})}{\cot(\phi_{i,j,l})^{mod} - \cot(\phi_{i,k,l})^{mod}} \quad (5.3)$$

$$\beta(s_i) = \sum_l \beta_l(s_i) g_l$$

Where g_l is the weight, l is the index of the probed BPM and i, j, k are the indexes of non-consecutive BPMs.

In the Figure 5.3 the N-BPM scheme is shown. Here the red triangle is the probed BPM, and gray is the one which is skipped. For the calculation the indicated phase advances are used.

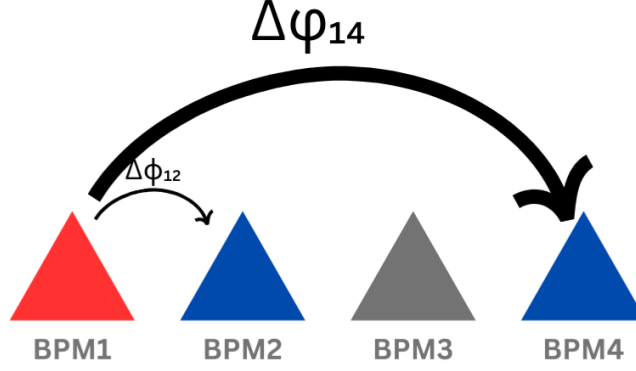


Figure 5.3: N-BPM scheme

β -function from amplitude

The orbit due to the betatron oscillation can be written as:

$$x(N, s) = \sqrt{A} \cos(2\pi Q_x N + \phi_x(s)) + \Delta x(s) \quad (5.4)$$

Where A is the amplitude of oscillation and it is defined as Eq.5.5.

$$A(s) = \sqrt{2J\beta(s)} \quad (5.5)$$

Here, β -function can be defined from that as shown in Eq.5.6.

$$\beta(s) = \frac{A(s)^2}{2J} \quad (5.6)$$

K-mod and β^* calculation

The previously described methods are used to calculate the β function at the locations of BPMs. However, at certain points in the LHC, such as interaction points and their surrounding areas, there are only a few BPMs, with the phase advances of approximately 0° in triplets and 180° between those next to IPs. This configuration is unfavorable for traditional measurement methods. To measure the β -function at the interaction points, a different technique called K modulation (K-mod) is required. In this technique it is used that the varying the gradient of quadrupolar fields causes a change in tune. With these variations, the average β function at the quadrupole's location can be calculated using Eq. 5.7 [36].

$$\bar{\beta}_{x,y} = \pm (\cot(2\pi Q_{x,y}) (1 - \cos(2\pi \Delta Q_{x,y})) + \sin(2\pi \Delta Q_{x,y})) \frac{2}{\Delta K L} \quad (5.7)$$

In the approximation where ΔQ is small the equation reduces to,

$$\bar{\beta}_{x,y} \approx \pm 4\pi \frac{\Delta Q_{x,y}}{\Delta K L}, \quad (5.8)$$

Where the $\pm$ sign corresponds to the horizontal and vertical planes, ΔQ is the variation of the tune and ΔK is the variation of the quadrupole strength, and L is the length of the quadrupole. β -functions at interaction points, can be obtained with the measured average β -function in the quadrupoles left and right of the IPs.

III Correction techniques

During the commissioning of the LHC one of the very first steps is the correction processes. In the LHC two correction approach is conducted. The first is local correction and after that the global corrections. The global correction technique involves adjusting quadrupole strengths in the arcs of the LHC and the local correction process focuses on specific segment of the beam. In the case of the LHC these specific segments are interaction point. Both techniques are essential for optimizing the performance of the beam. They will be discussed in detail in the following sections.

Global correction

Global correction techniques are adjusting different optical functions, including phase advance, β -function, dispersion, and coupling. These corrections are computed using a response matrix, often referred to as the R-matrix. the R-matrix establishes the relationship between changes in quadrupole gradients and deviations in the optical parameters using the simulation done in MADX [37]. The relationship is shown in the Eq.5.9[38].

$$\left(\Delta\vec{\phi}_x, \Delta\vec{\phi}_y, \frac{\Delta\vec{\beta}_x}{\beta_x}, \frac{\Delta\vec{\beta}_y}{\beta_y}, \frac{\Delta\vec{D}_x}{\sqrt{\beta}}, \Delta Q_x, \Delta Q_y \right) = \mathbf{R} \Delta\vec{k} \quad (5.9)$$

The required quadrupole strength adjustments Δk for the corrections are determined by inverting the response matrix, as demonstrated in Eq. 5.10.

$$\Delta\vec{k} = -\mathbf{R}^{-1} \left(w_1 \Delta\vec{\phi}_x, w_2 \Delta\vec{\phi}_y, w_3 \frac{\Delta\vec{\beta}_x}{\beta_x}, w_4 \frac{\Delta\vec{\beta}_y}{\beta_y}, w_5 \frac{\Delta\vec{D}_x}{\sqrt{\beta}}, w_6 \Delta Q_x, w_6 \Delta Q_y \right) \quad (5.10)$$

$w_1, w_2, w_3, w_4, w_5, w_6$ are weighting factors that can be applied to correct the individual optical functions. Optic variables here are now from the measurements and the process is to match measurements with the response matrix.

Local correction

Segment by segment approach is different method which identifies local sources of the errors.[39], [40]. This method examines optics parameters locally, in the particular segment. In the LHC, these segments are mainly the IRs. Local errors in the IRs are most important because of the large β functions. To observe the propagation of the beam within each segment, the measured β and α functions at the start of each segment are used as initial conditions. The propagation inside the segment is done by using the optics modeling tool MAD-X [37]. If there is no gradient error in the specific section, the propagation of optics parameters should show no divergence from the ideal model. Identifying any deviation indicated the local error source. This approach reduces the dimensionality, as this is a single pass approach, there are no other contributions from the rest of the ring and the errors are isolated.

The key parameter used during SBS approach is the deviation of the phase advances between simulated and measured beams. This optics function is a better indicator of the presence of the error

in the segment, as they are more directly observable than for instance the β function. As shown in Fig. 5.4, there is a phase advance discrepancy from the nominal beam. Initially, there is no departure, as expected, because they have the same initial conditions. However, once the gradient error is introduced, the shift becomes evident. Correction process is to match the measured variables to the simulated ones.

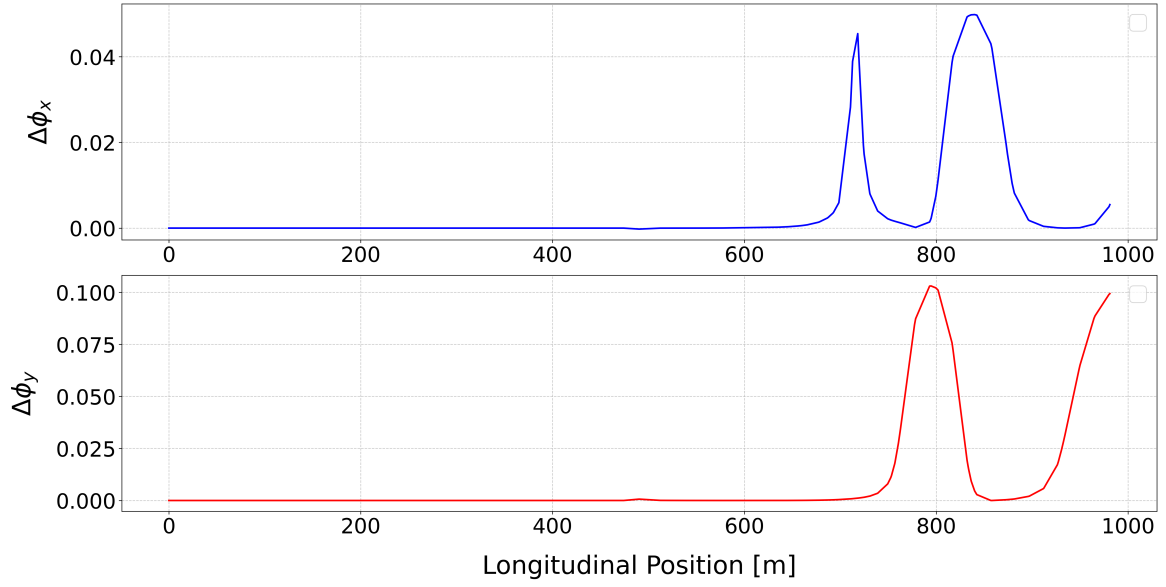


Figure 5.4: Deviation of the phase advance in the segment of IP1 (ATLAS) at LHC. Shift indicates that there are local gradient errors in this segment.

Chapter 6

General framework and methodology

The primary motivation of this study is to improve correction processes at the LHC by refining their accuracy and efficiency. Since machine learning has not been traditionally involved in these processes and has shown promising results, extensive studies are necessary to incorporate them into the correction processes. The aim of this thesis is to investigate and develop a machine learning model for this purpose. To achieve this the research is focused on the local correction of the LHC, specifically addressing the corrections of the magnetic field errors around the interaction points. There are four such interaction points in the LHC, each housing a detector. Ensuring that these detectors receive the desired luminosity while maintaining machine safety is crucial, which is why there is a strong motivation for a more advanced beam optimization techniques.

The model that will be trained is ridge regression, chosen based on previous studies [18]. As explained in the previous chapter, ridge regression is a supervised learning model that is trained using features and targets to find correlations between them. The target of interest in this study is the quadrupole field errors around the IPs, comprising a total of 32 elements, often referred to as triplets. One of the key aspects of this work will be selecting the relevant features and hyperparameters, including the noise level in the test set and the regularization parameter.

The features for the primary model are the simulated phase advance deviations from the design values between each pair of consecutive BPMs, with approximately 2000 BPMs. A second model, which includes additional features beyond $\Delta\phi$, such as β -beating at the closest BPMs on the left and right of each IP, totaling 32 additional features, will be compared to the primary one.

Testing the hyperparameters and finding the optimal values will be done for both overall beam performance and local performance, as the main focus is on local corrections.

To conduct these studies, numerous simulations are required to generate the data needed to train the model. Simulations of the LHC beam are performed using MADX [37], one of the main tools in accelerator beam physics for simulating various scenarios. In these simulations, errors on the triplets and in the arcs are generated randomly. Additionally, misalignment of the triplets is also included in the simulation, but with the different proportions of the size on each of them, specifically 10:2:1:1 at IP1, IP5 IP8 and IP2 respectively. The presence of other type of errors rather than triplets in the simulation is crucial because, as mentioned earlier, local corrections are performed first during the correction process. This means that while correcting the triplets, errors are also present in the arcs, so the model should be trained with this information. With these errors, simulations of both LHC beams 1 and 2 are conducted, and the optical functions data is extracted. The error sources causing β -beating are illustrated in Fig.6.1. The figure shows the beating caused by triplet errors separately and by arc errors separately for beam 1 and beam 2.

With this data, along with the corresponding quadrupole field errors, a sample for ML training is created. For this study, nearly 40,000 simulations were conducted, generating 40,000 samples. The same simulation results were used for both models, with the difference that one model included

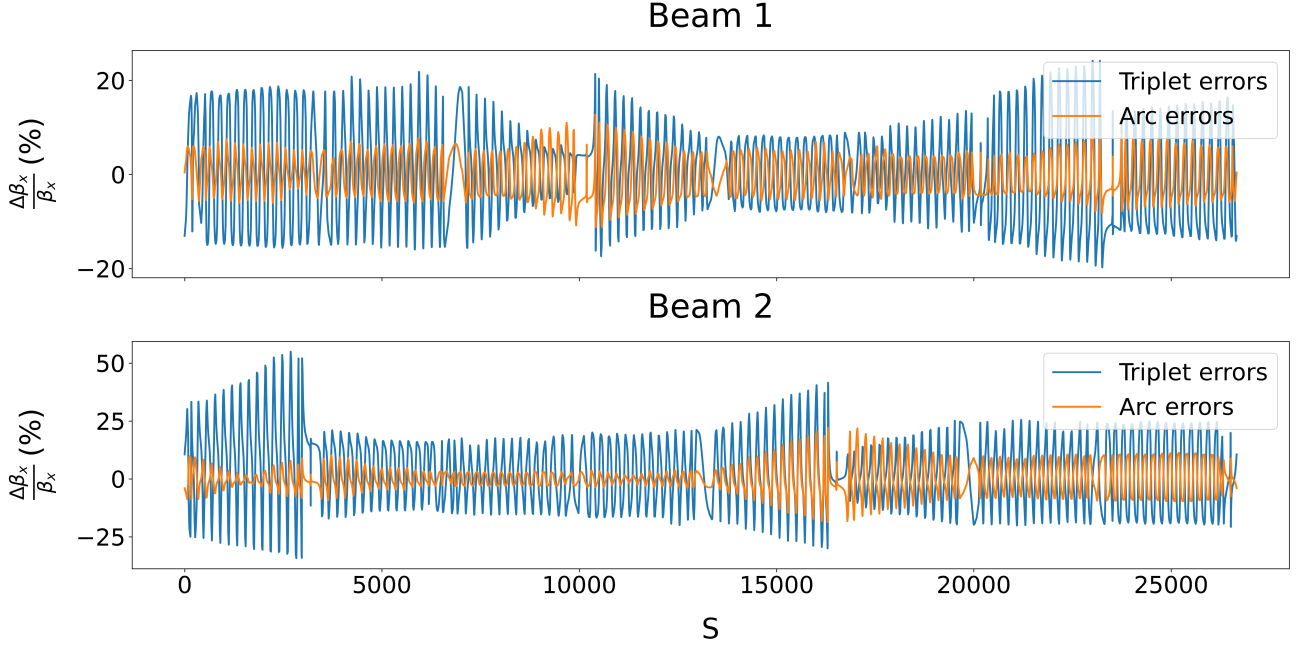


Figure 6.1: Horizontal β -beating as a function of the longitudinal position S (in meters), resulting from arc errors (shown in yellow) and triplet errors (shown in blue). The top plot represents the results for Beam 1, while the bottom plot corresponds to Beam 2. These results are derived from a single seed of the simulation.

additional features in the sample.

After the model is trained, it is tested with various methods, which will be explained later. However, the most important and final step is its local performance, which will be tested using the segment-by-segment approach. The outline of this process is illustrated in Fig.6.2. The final step involves validation with SBS to identify the optimal model.

Alongside these studies, another major focus of this thesis is the creation of a model that can predict the energy offset, which is also a very important and challenging error source for corrections. For this, in the ridge regression sample with features of $\Delta\phi$ and $\frac{\Delta\beta}{\beta}$, additional parameters are included in the target energy offset for beam 1 and beam 2 separately, resulting in 34 targets for the models. For this study, a new set of simulations was conducted, using MADX. In addition to assigning errors in the arcs and triplets, energy shifts were also assigned to both beams within a range of 10^{-4} using a random normal distribution. The same training process was then conducted, followed by performance testing.

The overall configuration of the ML model samples is illustrated in Fig.6.3. In this thesis, three different models will be investigated, focusing on their hyperparameters and local behavior.

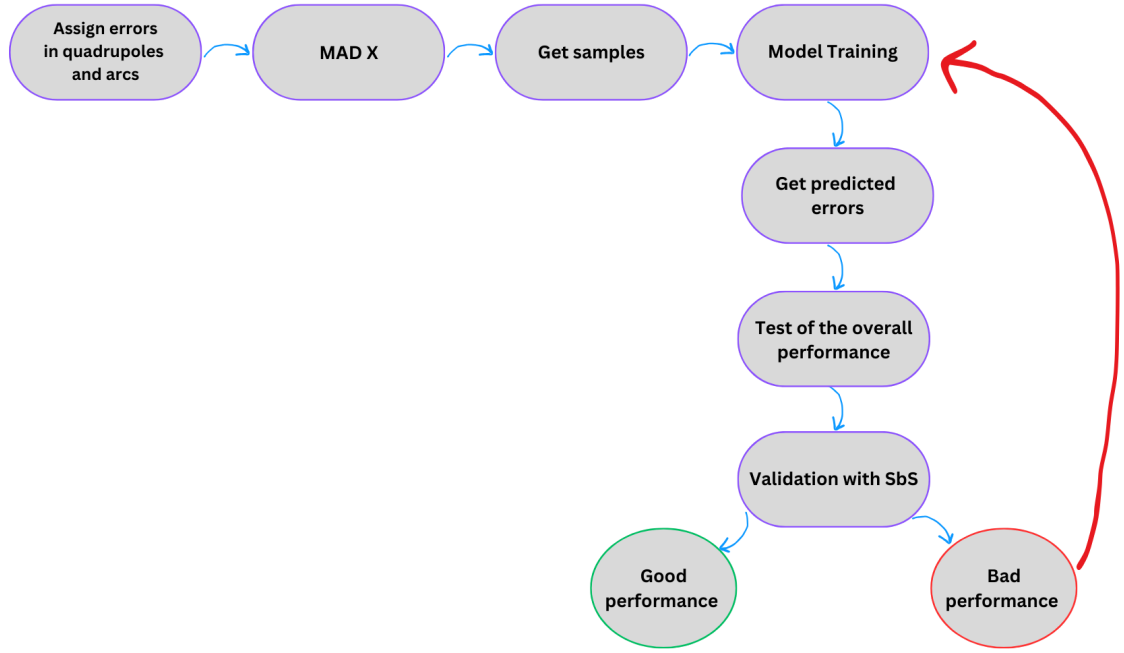


Figure 6.2: Flowchart of the analysis process. The process begins with the assignment of errors, followed by sample generation and model training. The model's performance is then validated, with outcomes leading to a classification of either good or poor performance. In the case of poor performance, the process loops back for re-training.

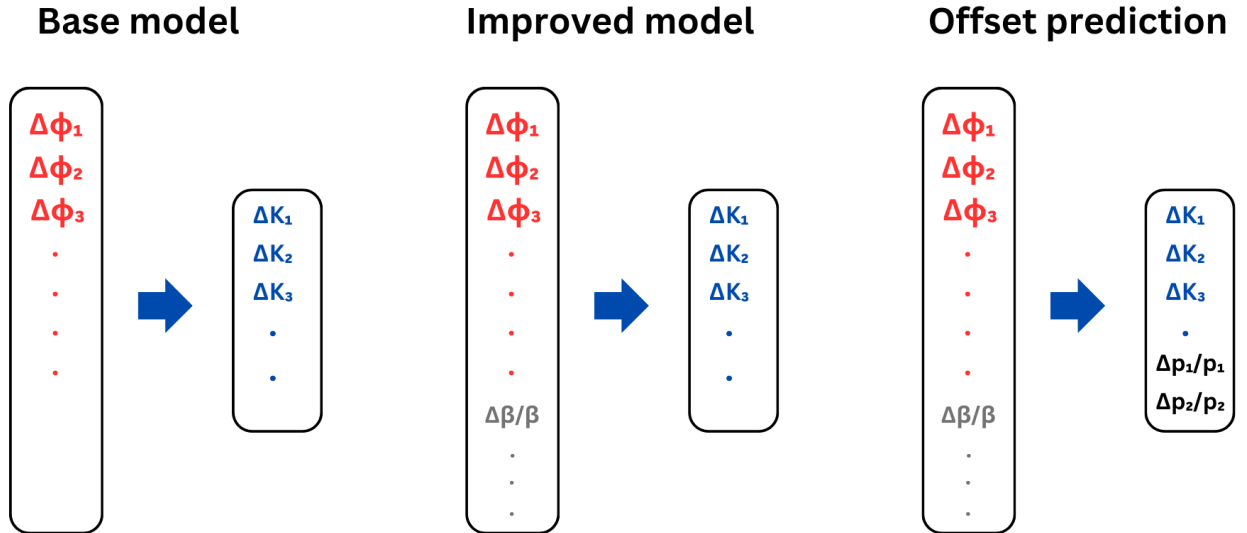


Figure 6.3: Three models on which studies have been conducted. Diagram illustrating the progression from a base model to an improved model, culminating in offset prediction. The base model includes a set of parameters $\Delta\phi$ in the features, which are enhanced in the improved model by incorporating additional parameters $\frac{\Delta\beta}{\beta}$, both targeting field errors ΔK . The final step involves the additional prediction of offset.

Chapter 7

Results

The following chapter will present the results. Key findings will be detailed and analyzed.

To evaluate model performance, the dataset is split into two parts: 80% is used for training and the remaining 20% is reserved for testing. Two key values are used to determine the model's accuracy: mean absolute error (MAE) and the coefficient of determination (R^2). MAE is the average of the absolute value of the prediction error, where prediction error is the difference between actual value and predicted value. Specifically measuring the deviation in quadrupole errors within the triplets. On the other hand, R^2 in regression analysis, indicating the proportion of variance in the dependent variable that is explained by the independent variables. An R^2 value close to 1 signifies a strong model fit, while a lower MAE indicates more accurate predictions. The formulas for these metrics are provided in Eq. 7.1 and Eq. 7.2.

$$R^2(\vec{y}, \hat{\vec{y}}) = 1 - \frac{\text{Var}\{\vec{y} - \hat{\vec{y}}\}}{\text{Var}\{\vec{y}\}} \quad (7.1)$$

$$MAE(\vec{y}, \hat{\vec{y}}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7.2)$$

Where $\hat{y}_i$ are the predicted magnet errors, $\hat{y}_i$ is the predicted value of the i -th sample, $\vec{y}$ are corresponding true values for all N samples, and Var is the square of the standard deviation.

Figures 7.1, 7.2, and 7.3 illustrate the convergence trends of MAE and R^2 values for each model analyzed in this thesis. The base model's MAE stabilizes at 1.4×10^{-5} , with an R^2 value of 0.75. In contrast, the improved model achieves a lower MAE of 1.3×10^{-5} and a higher R^2 value of 0.8, showing the better results for the improved model. However, when an additional target, specifically the energy offset, is introduced, the MAE increases slightly to approximately 1.4×10^{-5} , and the R^2 value decreases to 0.75. This indicates that the inclusion of energy offsets challenges the model slightly but the overall performance still remains strong. Additionally, the lack of significant performance improvement beyond 40,000 samples suggests that this number of samples is most favorable for the models to achieve their optimal predictive capability.

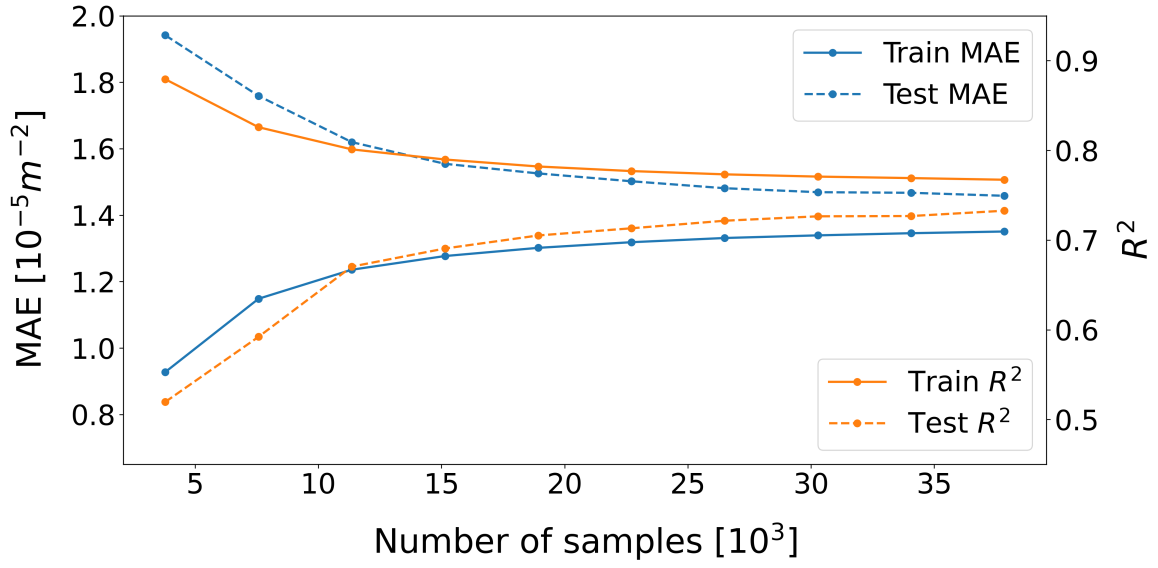


Figure 7.1: Cross-validation of the base model, highlighting the convergence of the mean absolute error (MAE) and the R^2 coefficient. This model analyzes the deviation between phase advances as features and the quadrupole strength errors around the interaction points (IPs) as targets.

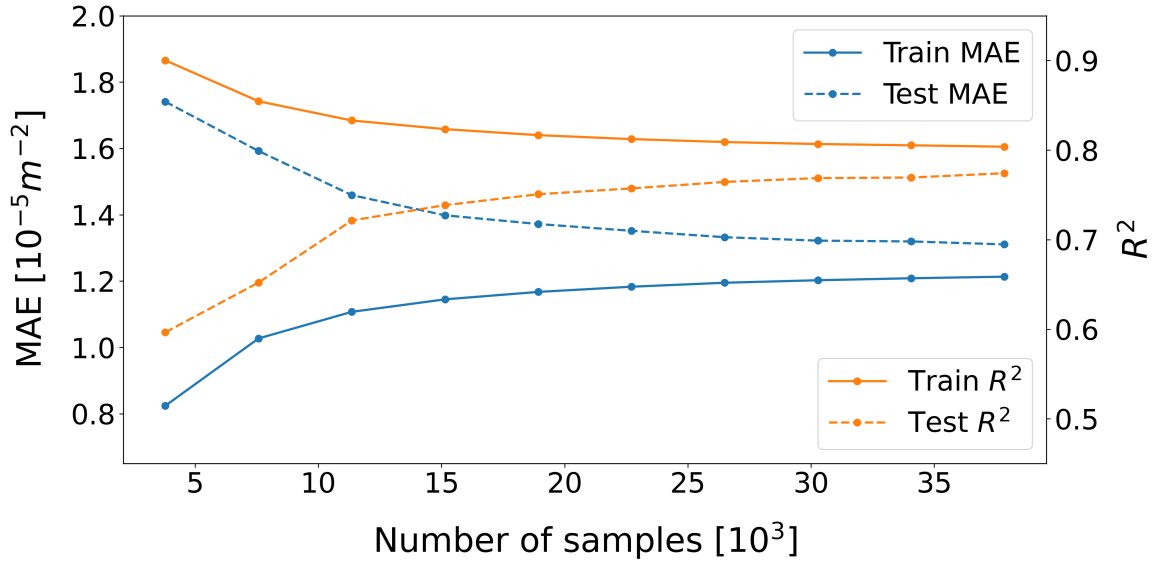


Figure 7.2: Cross-validation of the improved model, demonstrating the convergence of the mean absolute error (MAE) and the R^2 coefficient. This model includes the deviation between phase advances and β^* s as features and the quadrupole strength errors around the interaction points as targets.

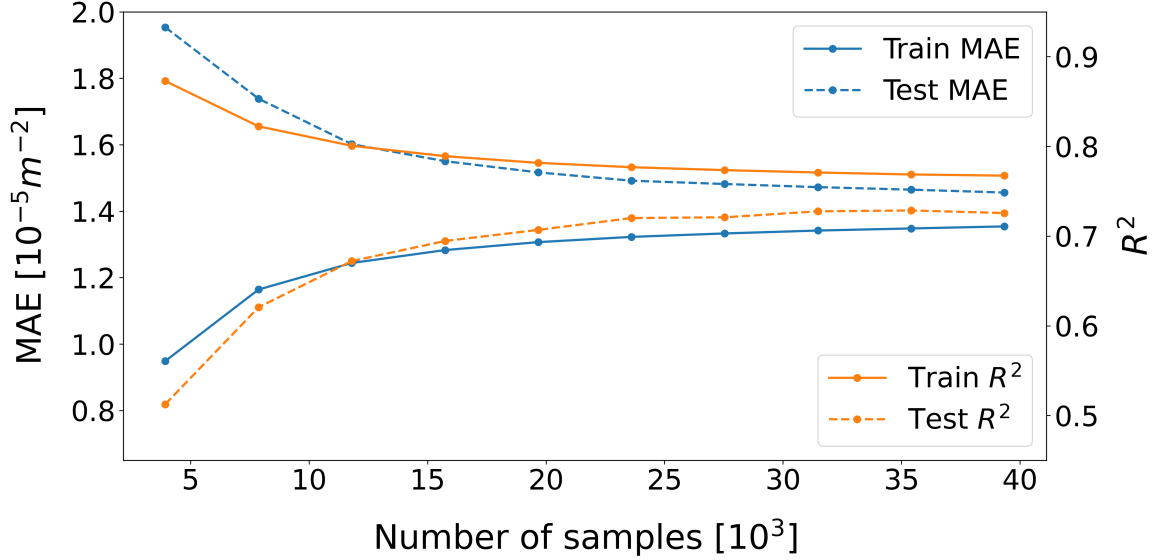


Figure 7.3: Cross-validation of the model, which predicts energy offsets as well as quadrupole strength errors around the interaction points. The features include deviations of phase advances and β^* values, while the targets are quadrupole errors and energy offsets for each beam separately. The convergence of the mean absolute error and the R^2 coefficient indicates successful model training.

I Triplet prediction and β -beating

To further analyze the models performance, additional tests and examinations are necessary, besides the satisfactory values of MAE and R^2 scores. Since the model is intended to predict the quadrupole field errors around interaction points, with a total of 32 values (4 IPs, each with 8 quadrupoles), it is beneficial to compare the errors set during the beam simulation with the predicted values derived from the optical functions extracted from the same simulations. All tests were conducted on the test data. Figure 7.4 displays two plots illustrating the same metrics for different models. The first plot pertains to the regression model utilizing only phase advances as features, while the second plot corresponds to the model incorporating additional features, specifically $\frac{\Delta\beta}{\beta}$ around the IPs, totaling 32 values, with 2 at each of the 4 IPs for both beams in each direction. These plots show the distributions of the quadrupole field errors for all the test data, including the real errors used during the simulations and the predicted errors, as well as the difference between them, denoted as residuals for better understanding. For both models, the predictions align closely with the actual errors, which is validated by the distribution of the residual errors, concentrated near zero.

Figure 7.5 illustrates the residual field error distributions for both models on the same plot for a more clear visual comparison. These values are identical to those in Figure 7.4. It is evident that the model incorporating $\frac{\Delta\beta}{\beta}$ around the IPs as additional features demonstrates greater performance, as indicated by a smaller standard deviation. Specifically, the base model has a standard deviation of 2.93×10^{-4} , while the improved model has a standard deviation of 2.59×10^{-4} . This shows that the distribution of residual field errors is more concentrated around zero. This observation serves as a clear validation of the model's improvement with the inclusion of extra features, underscoring their importance in the model's predictive performance.

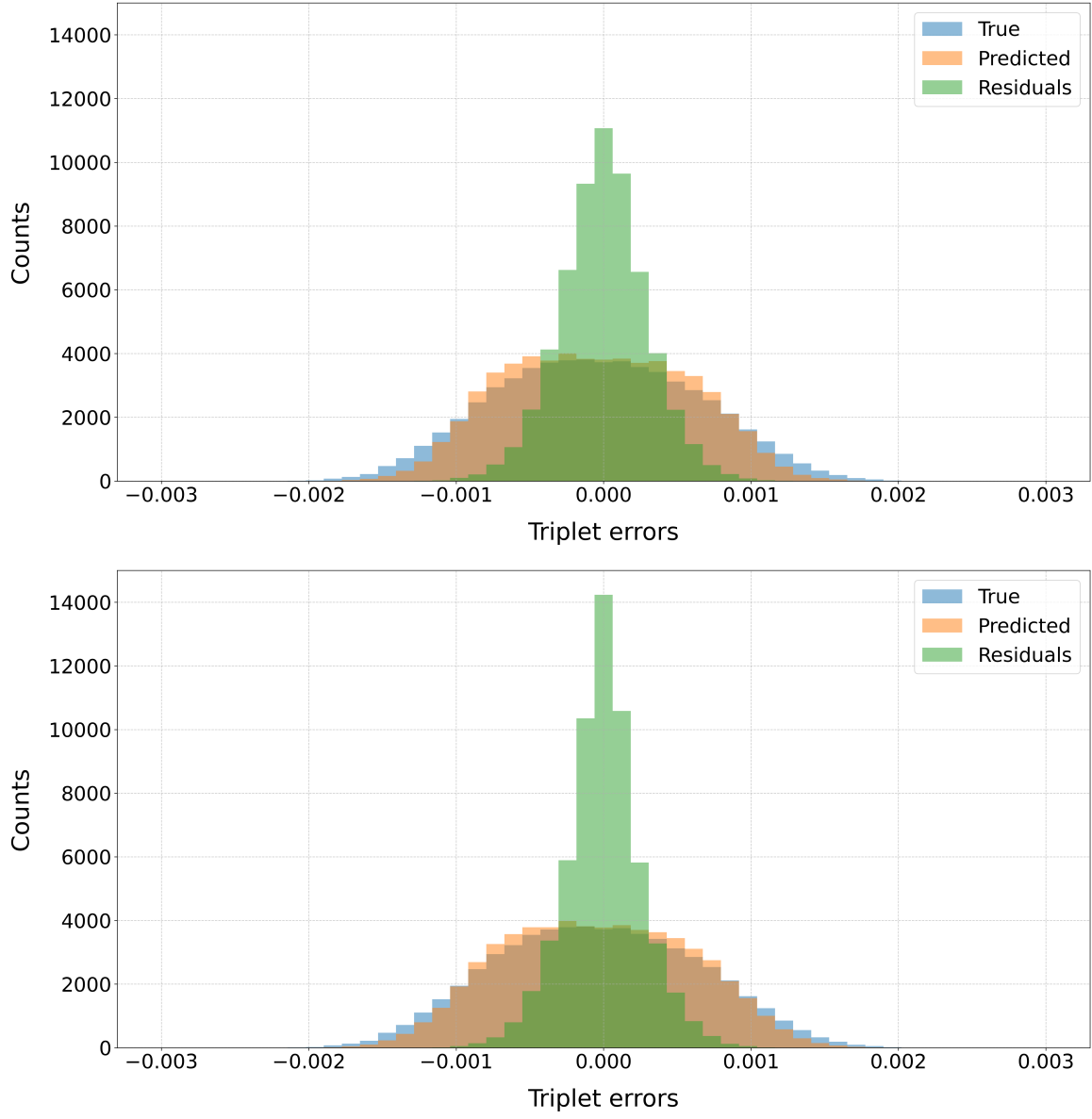


Figure 7.4: The upper plot shows ridge regression model with $\Delta\phi$ as features and quadrupole field errors as targets. The bottom plot shows the model with $\Delta\phi$ and $\frac{\Delta\beta}{\beta}$ around IPs as features and quadrupole field errors as targets. On the plots there are true quadrupole field error values from the test set in blue, predicted values in yellow, and the residuals (difference between true and predicted values) in green. In both cases the predicted values are very close to true values.

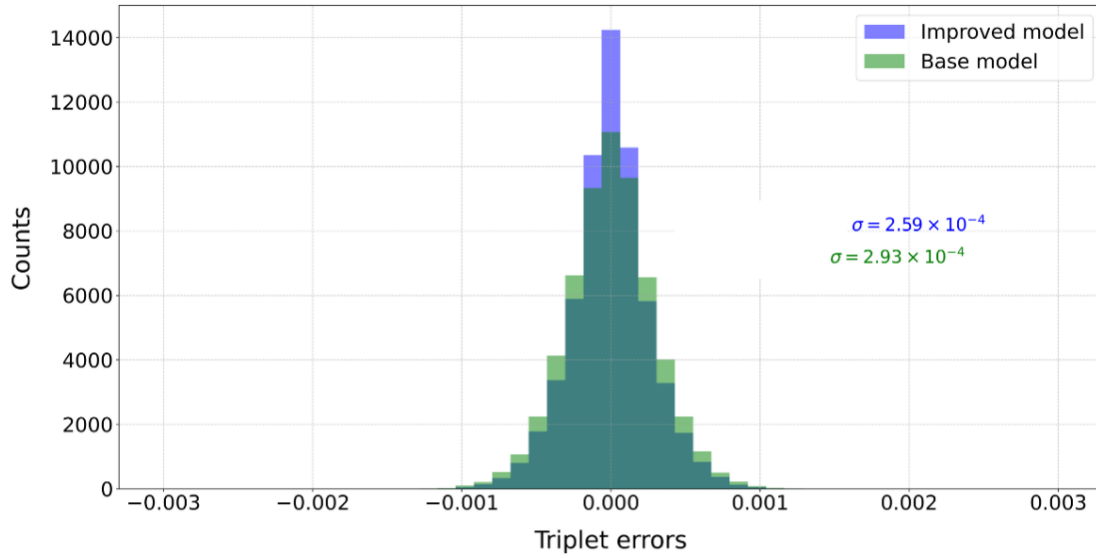


Figure 7.5: Comparison of residual triplet errors corrected using values predicted by the base model (green) and the improved model (blue). The visual improvement is evident, and the numerical data also indicate that the mean residuals from the improved model are closer to zero, with less deviation compared to the base model.

Another crucial aspect for examining model robustness is the observation of β -beating. This optical function is often analyzed during the measurement and correction of the beam, as it provides a better understanding of the beam's performance. A high β -beating indicates greater beam deviation, and the objective is to maintain these values as low as possible, ideally under 10%. Therefore, the models should be evaluated based on their β -beating behavior. Figure 7.6 presents the RMS values of β -beating from the test set, with one plot corresponding to the base model and the other to the improved model. RMS values are calculated from the β -beating of the simulated beam, once with only true errors assigned in the quadrupoles and once with residual errors (the difference between true and predicted errors) assigned, essentially correcting the beam using the model. Both models demonstrate a significant improvement in beam performance and optimization, as the initial β -beating was quite large with the real errors but was greatly reduced after correction. The values are maintained under 10%, which is the desired target for optimization.

To further compare the performance of each model, Figure 7.7 shows the same RMS values of β -beating for corrected beams but with different models, the base model, which includes phase advances as features, and the improved model, which incorporates extra parameters. Although the improvement is not substantial, it is still evident. The base model yields a mean value of 5.73 and a standard deviation of 3.73 in the horizontal direction, compared to the improved model's mean value of 5.53 and standard deviation of 3.43. In the vertical direction, the base model has a mean value of 5.83 and standard deviation of 3.92, while the improved model shows a mean value of 5.56 and standard deviation of 3.66. These numbers indicate some improvement, even though it was more pronounced in previous tests.

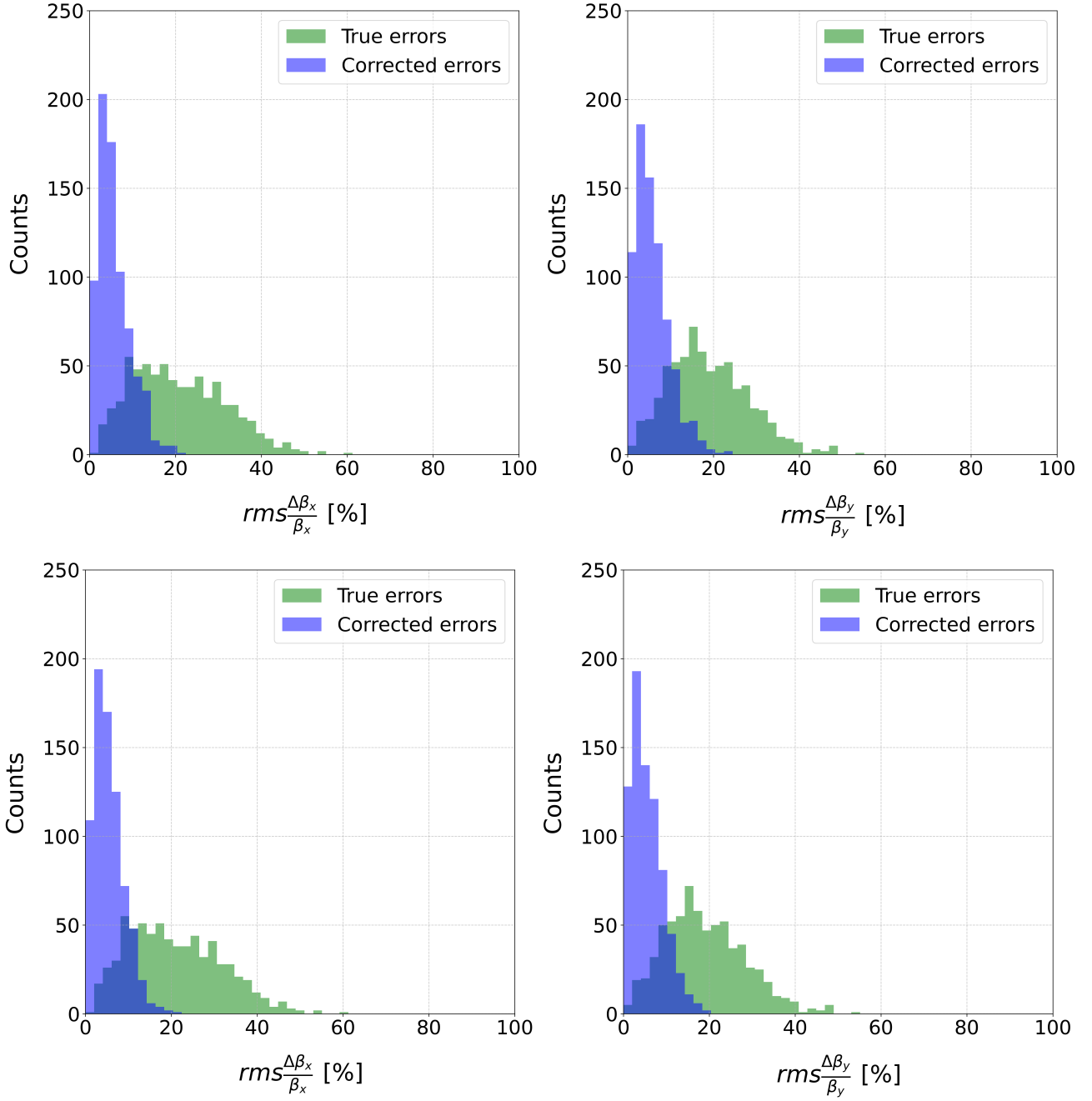


Figure 7.6: β -beating of beam 1, left plots correspond to x direction, while right correspond to y direction. Green represents the reconstruction of the optics with true errors generated during simulation, while blue represents the reconstruction using the same true errors corrected with predicted values. The top plot shows predictions made by the base model, which includes only $\Delta\phi$ as features, while the bottom plot illustrates the performance of the improved model, which incorporates $\frac{\Delta\beta}{\beta}$ as additional features. The analysis is based on approximately 700 simulations.

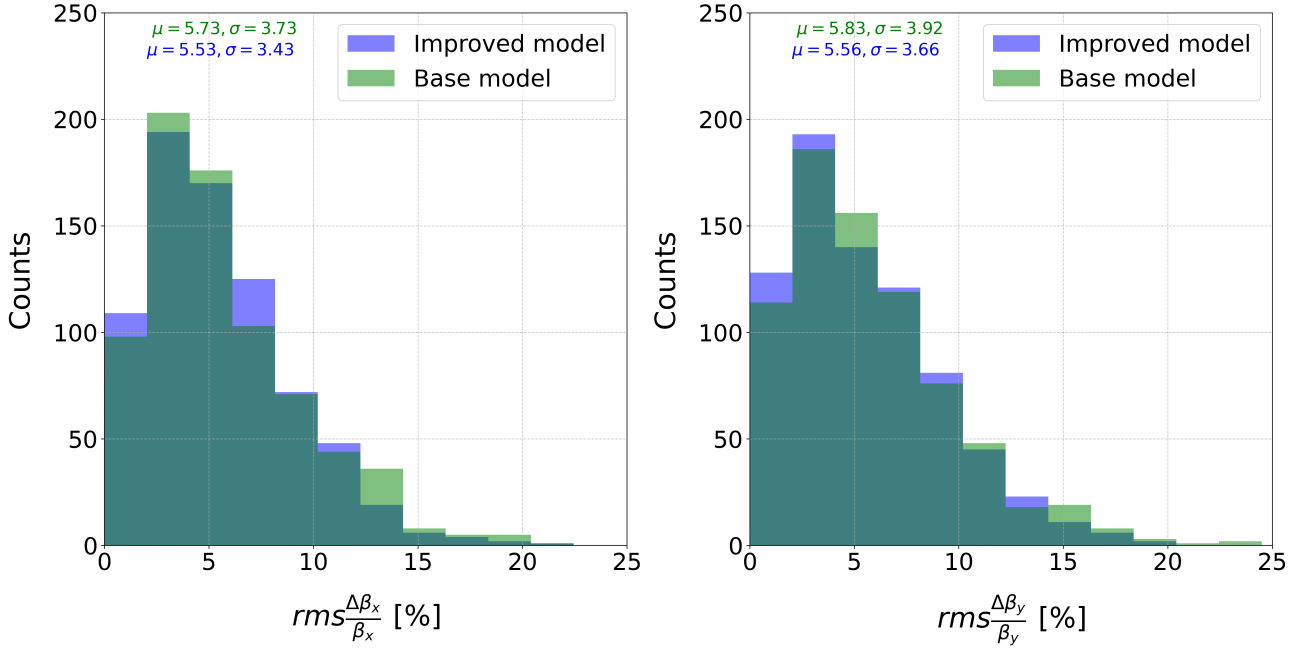


Figure 7.7: Comparison of the RMS values of the beating of the β function. The green bars represent residual beating after correcting the triplets using the base model, while the blue bars show the same correction using the improved model. For both values mean RMS and σ the improvement is clearly evident.

II Energy offset prediction

In accelerators, numerous error sources can be found. Previous studies have mostly been focused on identifying errors originating from the magnets, specifically quadrupoles. However, errors can also arise from relative energy shifts induced by an average kick of the corrector magnets [36]. Such energy shifts cause β -beating similar to that caused by quadrupole magnet errors. This overlap presents a challenge in accurately identifying and distinguishing between these error sources.

The complexity is further heightened in the Large Hadron Collider, which operates with two beams that share a common aperture in the triplet quadrupoles at interaction points. Magnetic errors from these shared quadrupoles affect both beams simultaneously, when the energy offsets are unique for each beam. Consequently, correcting errors that differ for each beam with the shared sources may prevent achieving optimal performance.

Given these challenges, developing effective methods for accurately identifying these error sources is crucial for improving correction efficiency. In this study, a machine learning model was developed to address this specific challenge. The model builds upon a previously used ridge regression approach designed to predict triplet errors by targeting quadrupole field errors in the interaction regions (IRs) using the features $\Delta\Phi$ and $\frac{\delta\beta}{\beta}$. The model was modified to include data simulated with energy shifts for both beam 1 and beam 2, in addition to the arc and triplet errors. A new dataset comprising 40,000 samples was generated, where energy offsets $\frac{\Delta p}{p}$ were also targeted alongside the field errors. These energy shift values were randomized for each simulation, creating a normal distribution within the range of 10^{-4} , as empirical evidence suggests that errors typically fall within this range. Errors for the triplets and arcs were assigned similarly to previous models.

To determine if including an additional parameter in the features has any detrimental effect on the accuracy of the previous results or not, it is crucial to first analyze the distribution of the triplet predictions. The distribution of errors is illustrated in Fig. 7.8, which, similar to prior studies, displays the true errors assigned during simulation, the predicted triplet errors, and the difference between them, denoted as residuals. All tests were conducted on a test dataset. The results indicate that including energy offsets as prediction targets does not degrade the prediction accuracy of triplet magnetic errors. In fact, the results demonstrate that the model can effectively handle the added

complexity.

Histograms, now focus on the energy offset values, are analyzed to further investigate the accuracy of the energy offset predictions. The results for these predictions are presented in Fig. 7.9. In this figure, the true energy shift values for each beam, taken from the test dataset, are shown in blue. The predicted values for the same dataset are depicted in yellow, and for easier comparison, the residuals (the difference between true and predicted values) are shown in green. As seen from the plot, the predictions are accurate, with RMS distribution of a very small standard deviation. These low RMS values indicate that the model predicts energy offsets with high precision, indicating the high robustness of the model.

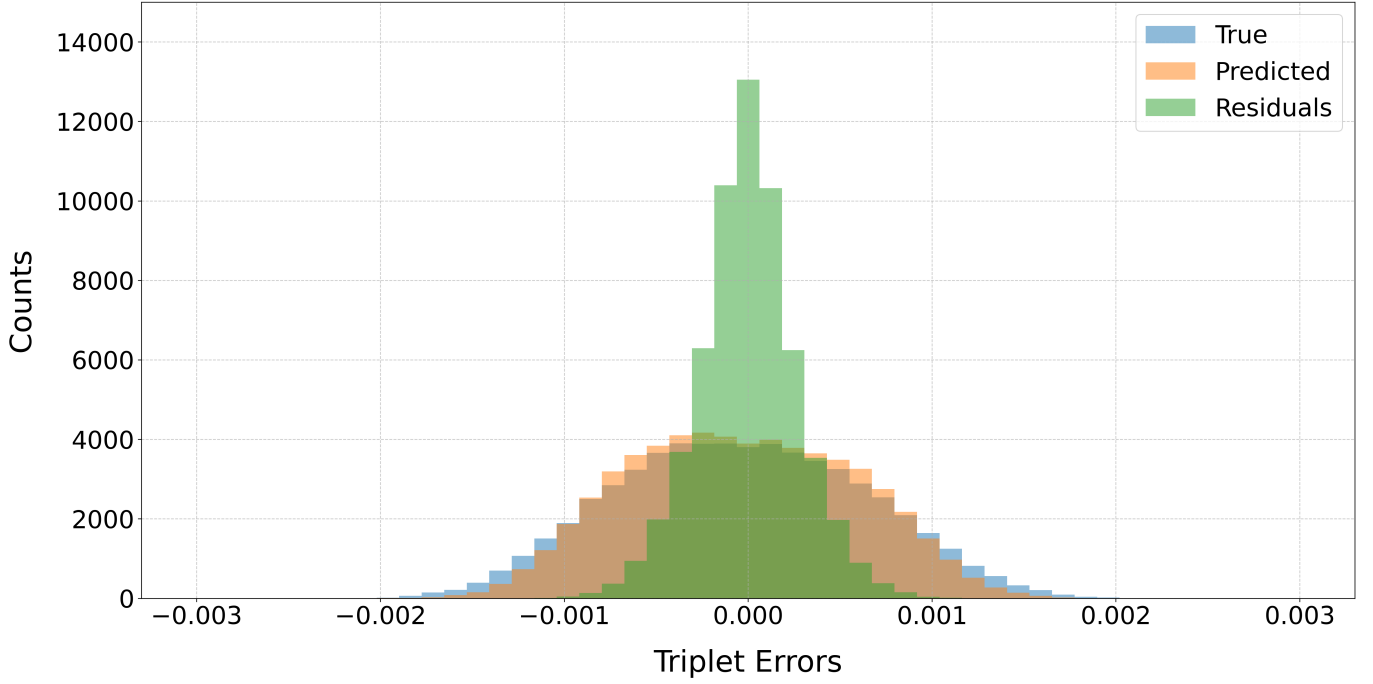


Figure 7.8: The figure represents the true triplet values from the test dataset, alongside the predicted values generated by the model, which includes both energy offsets and field errors in its targets. The residual errors, depicted in green, represent the difference between the true and predicted values. Notably, this study utilized a different dataset from that employed in prior research, therefore direct comparisons of RMS values between this study and previous predictions are less relevant for evaluating model performance.

A more critical evaluation of the model's performance involves assessing the β -function's beating, which is a key indicator of beam performance. Figure 7.10 depicts the β -beating resulting from triplet field errors and energy offset errors, alongside the results after correcting for both. The figure demonstrates that the β -beating is maintained below 10 %, signifying good performance.

To emphasize the impact of energy shifts on beam performance, Figure 7.11 compares two scenarios: one where β -beating is simulated with both triplet and energy offset errors corrected, and another where β -beating is simulated with only triplet errors corrected. As expected, the presence of uncorrected energy offsets degrades beam performance, with the mean RMS values being approximately 1% larger without correcting for it. This represents a significant deviation caused by just one source of error, especially considering that the goal is to achieve the lowest possible β -beating.

In real-time beam energy offset errors are often present and can negatively impact the precision of triplet predictions from models that only target field errors. When addressing this issue, the model incorporating energy offsets in its targets is expected to be more robust. To evaluate this hypothe-

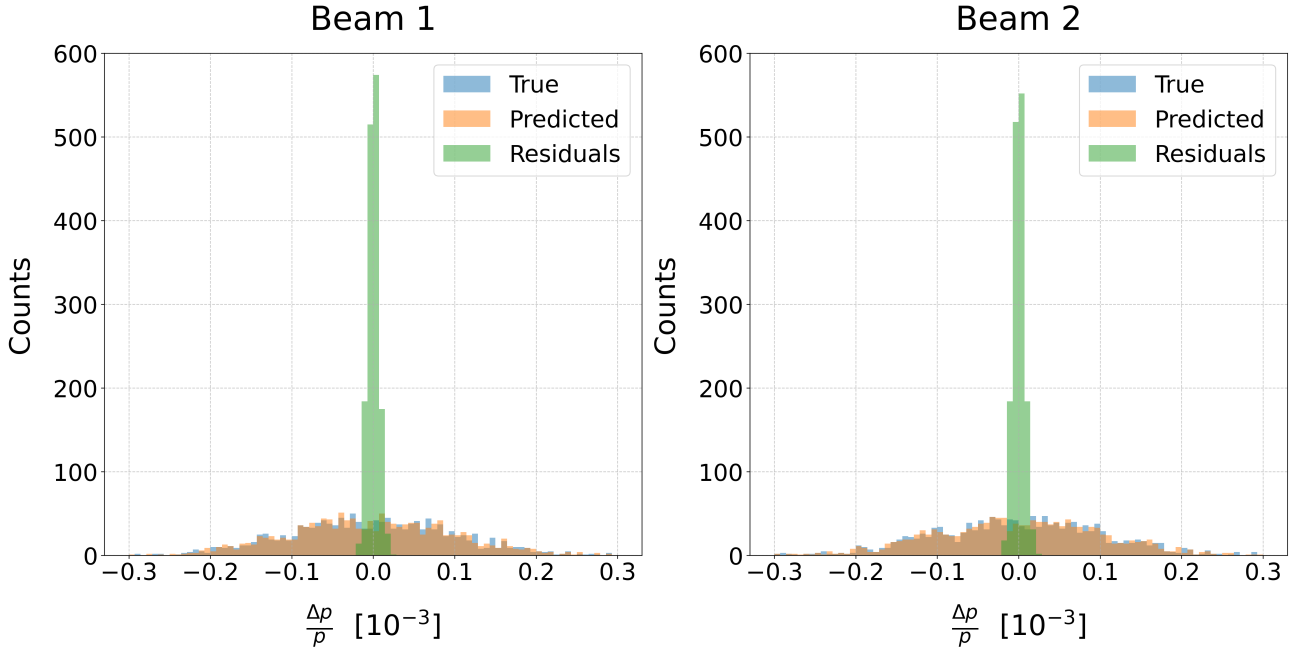


Figure 7.9: The figure shows the distribution of energy shifts which were used in the simulation's test set for both beam 1 and beam 2, as indicated. It also includes the corresponding model predictions. The green distribution represents the residual values, calculated as the difference between the true and predicted values. The close alignment between the true and predicted distributions, along with the narrow distribution of residuals, demonstrates the high precision of the model's predictions.

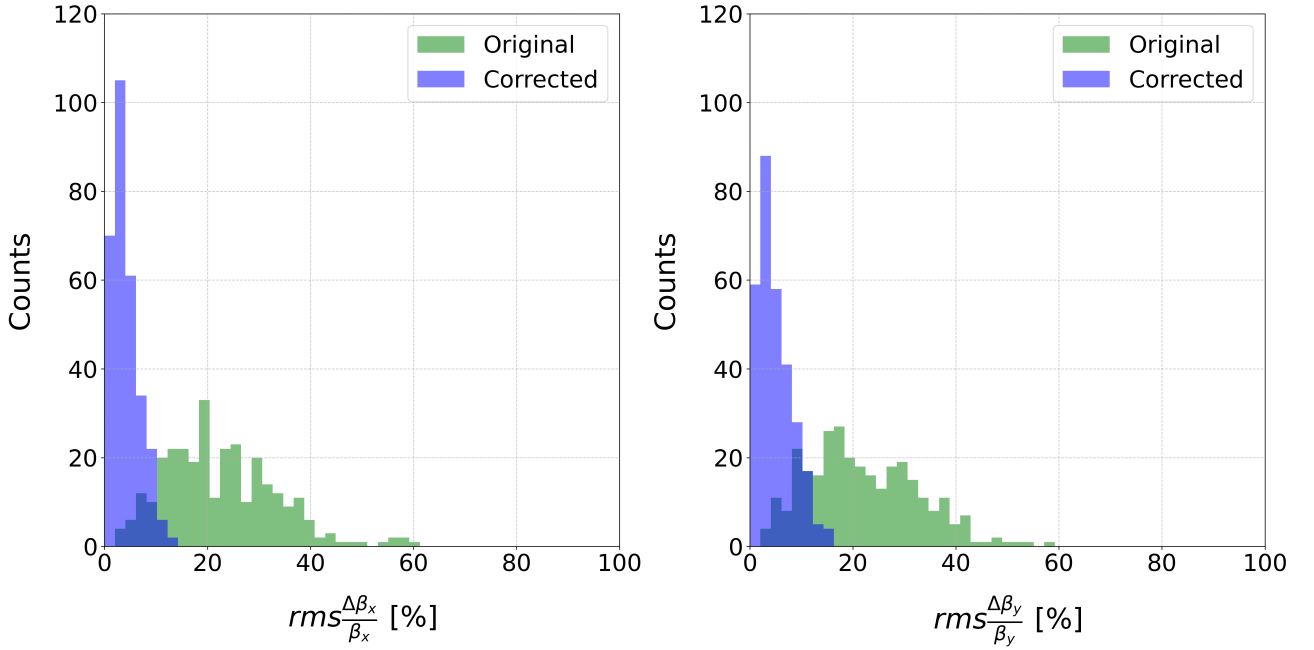


Figure 7.10: The β -function for Beam 1 in the x and y direction is depicted, where the green histogram shows the optics reconstruction based on true errors and energy offsets generated during the simulation. The blue histogram illustrates the reconstruction with these errors and offsets corrected using predicted values.

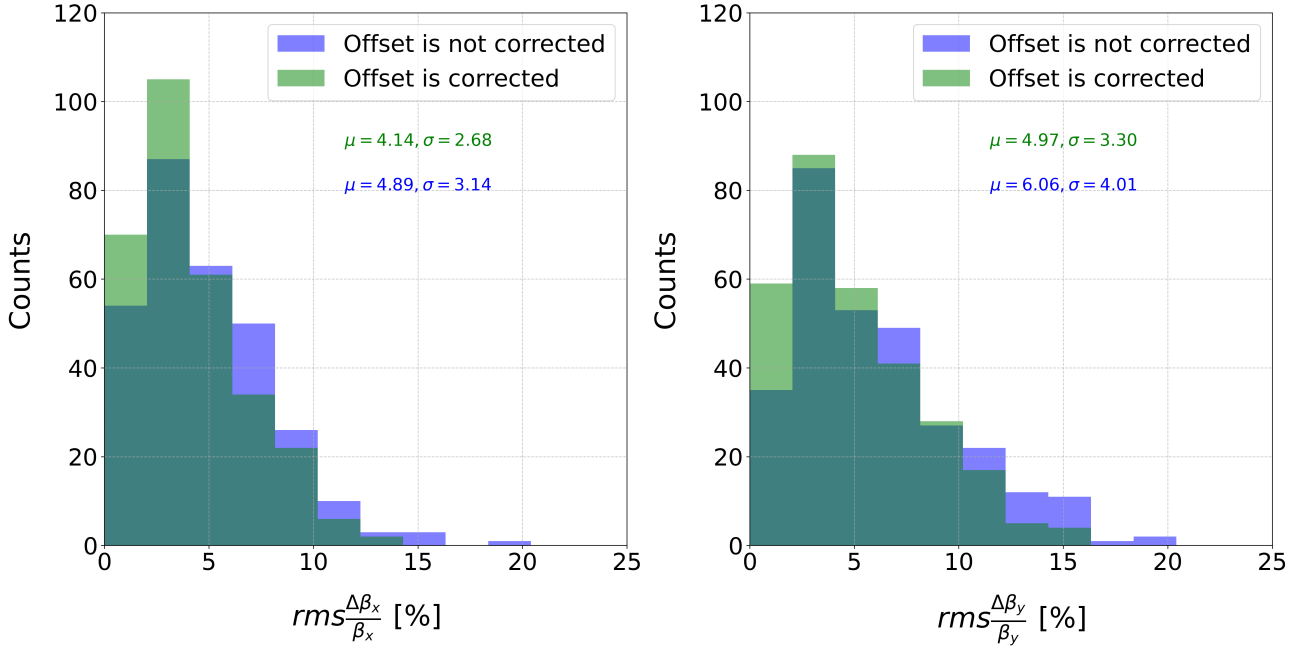


Figure 7.11: RMS distribution of β -beating for the test set, comparing the cases where both field errors and energy offsets are corrected and where only field errors are corrected. The results highlight that uncorrected energy offsets contribute to an increase in β -beating by nearly 1%, highlighting the significance of correcting both error sources.

sis, tests were conducted using data generated with energy offsets, comparing predictions from two distinct models: one trained with the offset and one trained without it. Figure 7.12 illustrates the β -beating of beams simulated with the residual triplet errors corrected by both models. The results indicate that the model which predicts the energy offset demonstrates higher robustness in accurately predicting triplet errors. Specifically, the mean RMS values are 4.62 and 4.81 for the horizontal and vertical directions, respectively, with a sigma of 2.98 and 3.20. In contrast, the model excluding the offset yields mean RMS values of 5.24 and 5.16, with sigma values of 3.26 and 3.18 for the horizontal and vertical directions, respectively, displaying clear advantage of the model including energy offsets.

The analysis shows that incorporating energy offsets as prediction targets does not reduce the model's accuracy for triplet error predictions, and the model predicts energy offsets with high precision.

III Hyperparameters tuning

Achieving better generalization in a model requires incorporate noise during training. This practice helps the model become robust to variations and inaccuracies found in real-world measurement data. It will result more reliable predictions in practical cases. Additionally, the added noise enhances regularization, preventing the model from fitting very well to the training data and reducing the chance of overfitting. Training the model with noisy data, make it able to identify general patterns rather than relying on specific features, which will promote better generalization to new, unseen data.

In this study, the model under consideration is the ridge regression model, which incorporates $\Delta\phi$ and β -beating around IPs, with the quadrupole field errors at IPs as target parameters. The noise level for $\Delta\phi$ will range from 10^{-4} to one digit, representing the noise levels relative to the true values. While varying the noise on $\Delta\phi$, the noise on the other parameter, $\frac{\Delta\beta}{\beta}$, will follow a normal distribution with the mean values at 3% of the actual value, truncated at 3σ . Notably, the

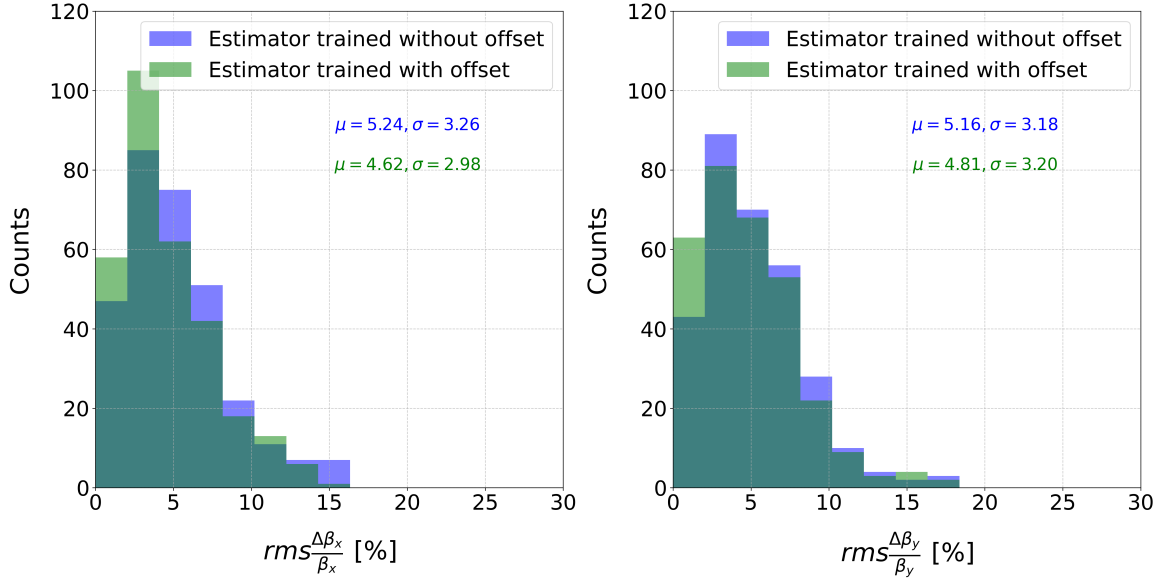


Figure 7.12: This plot illustrates the comparison between predictions made by two distinct estimators on the same dataset, which was simulated with both triplet error sources and energy shifts. The first estimator, depicted in green, was trained on data that included the offset and targeting offset as well, while the second estimator, shown in blue, was trained on data that excluded the offset. The figure represents the resulting β -beating residuals after triplet correction in both scenarios. The lower RMS observed for the estimator with the offset is observed.

noise levels of phase advances and β -beating around the IP are uncorrelated because they originate from two distinct measurement techniques, turn-by-turn data and k-modulation. This independence allows for the separate treatment of these noise levels.

In addition to investigating noise levels, the regularization coefficient for ridge regression, α , is a critical parameter affecting model performance. This study will explore the impact of α within a range from 10^{-4} to 10^{-1} , in conjunction with varying noise levels.

To closely approximate real measured data, the noise level on the test set of $\Delta\phi$ should first be set to 10^{-3} relative to their values. This level has been defined with the empirical approach. Additionally, the same noise level applied to $\frac{\Delta\beta}{\beta}$ during training should be maintained for the test set.

Incorporating noise into the data prior to prediction challenges the model's robustness and motivates for a deeper investigation into hyperparameters. To evaluate this against noisy data and identify the optimal noise parameters for training, it is essential to compare the triplet field error predictions with actual values across various scenarios. The noise applied to the test set remains constant at the most nominal level, closely reflecting real measured data. As depicted in Fig. 7.13, three scenarios are considered, where the test set noise remains unchanged, but the training noise levels on $\Delta\phi$ vary, specifically at 0.0001, 0.001, and 0.01, while the regularisation coefficient was fixed to the value of 0.001. The figure presents the distributions of true errors, predicted errors, and their differences, denoted as residuals. The models' performance varies visibly across different training noise levels, as confirmed by the RMS values of the distributions. The model trained with a noise level of 0.0001 exhibits the highest RMS residual value at 5.52×10^{-4} , while the model trained with a noise level of 0.001 shows a lower RMS value of 3.23×10^{-4} . This shows once again that this noise level parameter is very important.

As it is evident that the tuning of parameters significantly influences prediction accuracy. Therefore looking at a wider range of each parameters should be interesting. In Fig. 7.14, the RMS values of residual triplet error, which are the differences between true values and predicted values, are shown

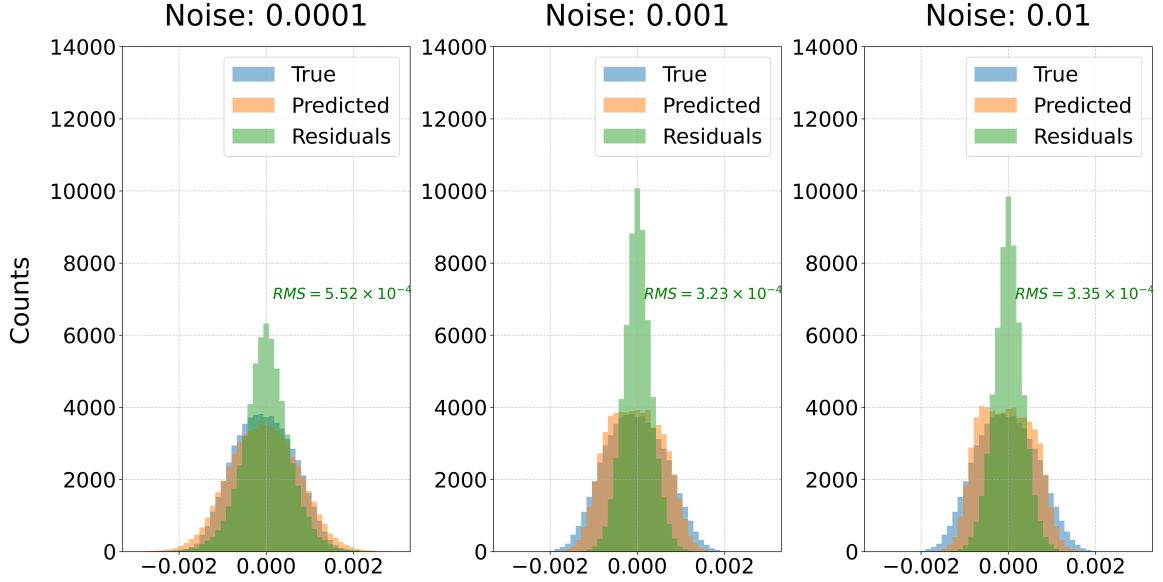


Figure 7.13: The figure presents the predictions made by three machine learning models, demonstrating the impact of different noise levels in the training set, with a regularization parameter of 0.001. The one in blue represent the true error values assigned during simulation, the yellow show the errors predicted by the machine learning models, while the one in green illustrate the residuals, which are the differences between the true and predicted errors. It is evident that the prediction precision is affected by the noise levels in the training set. The highest precision is observed with a noise level of 0.001, while the lowest precision occurs with a noise level of 0.0001

for models trained with corresponding α and noise levels as indicated in the plot. It is worth noting that the RMS of the true error values is 6.8×10^{-4} .

As clearly seen, models trained with lower noise and α exhibit poor results. Performance improves with increasing noise and α , after which there is not much difference in the results, even when the noise level on $\Delta\phi$ is as high as 1 or 5, rendering these features essentially just noise. This phenomenon can be explained by the fact that lower noise and α place excessive importance on phase features relative to β -beating, potentially leading to overfitting. The stability of the model, even with very noisy $\Delta\phi$ data, can be attributed to the ridge regression's ability to handle noise. In high-dimensional spaces, noisy inputs can still contain useful information as the noise does not completely obscure the signal. The model might still detect patterns or correlations despite the added noise, with $\frac{\Delta\beta}{\beta}$.

To further investigate this assumption, a similar study was conducted using a model that only includes $\Delta\phi$ and not β -beating as features and field errors as targets. Fig. 7.15 shows the RMS values of residual field errors for models trained with the corresponding α and noise levels. It can be clearly seen that in this case high noise degrades performance significantly. In this scenario, the presence of noise in the features leads to the anticipated outcome, a significant reduction in the prediction precision of the errors. This finding showed once again that the inclusion of β -beating is making the model more robust and confirmed the fact that β -beating is a very important parameter for field predictions.

The results showcasing the important role that $\frac{\Delta\beta}{\beta}$ holds in the training process provide a motivation to study the model that would only include $\frac{\Delta\beta}{\beta}$ around IPs as features, and incorporates both quadrupole field errors and energy offsets in the target parameters. By including energy offsets, the investigation aims to determine if the robustness observed in predicting field errors can be extended to predicting offsets as well.

In Fig. 7.16, the model's predictions of the actual values of the triplet field errors are displayed. The model performs well, demonstrating satisfactory accuracy in predicting these errors. Conversely, Fig. 7.17 shows the predictions for energy offsets using the same model configuration. It is evident

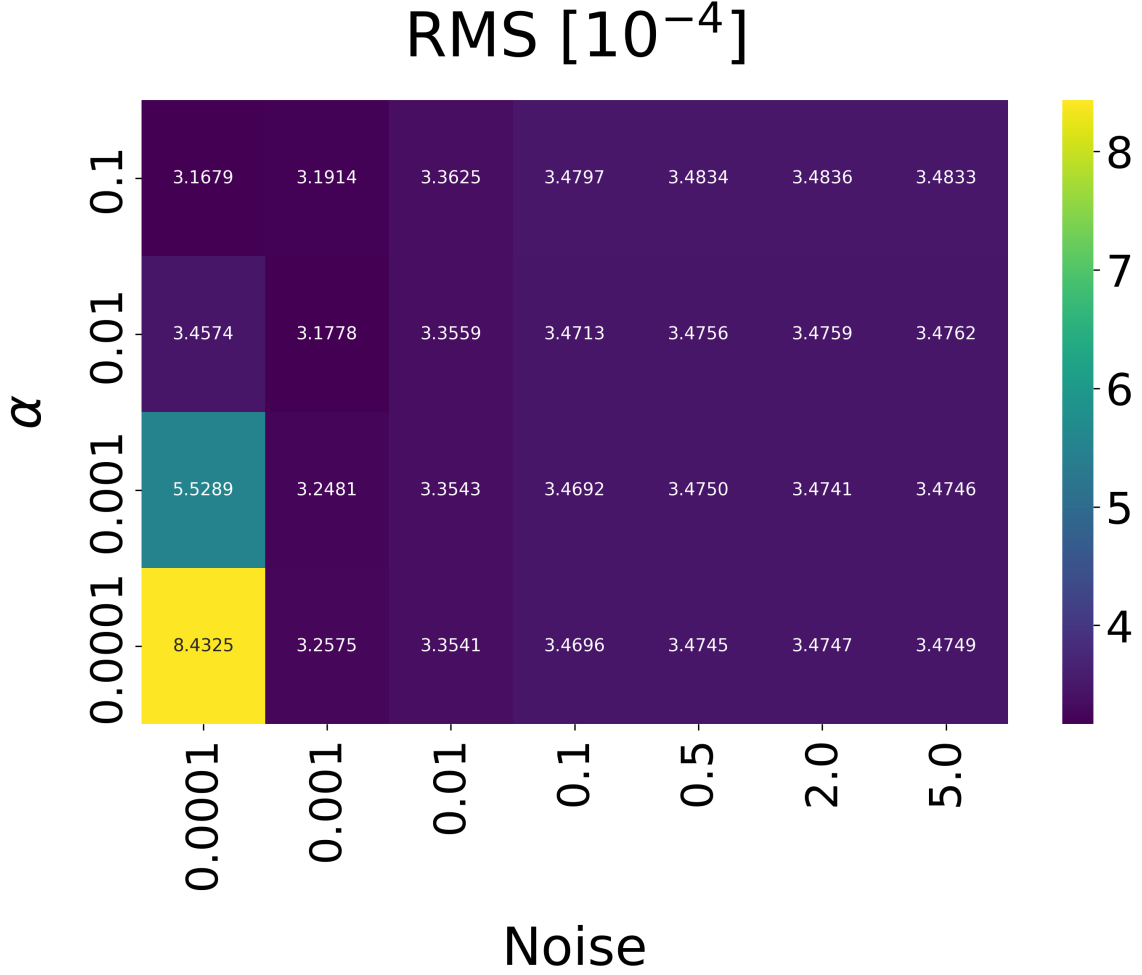


Figure 7.14: The plot shows the RMS values of the residual triplet field errors, which are the differences between the true and predicted values, obtained from different ML models. All models use $\Delta\phi$ and $\frac{\Delta\beta}{\beta}$ as features, and quadrupole field errors as the target. The models are trained with the parameters shown on the graph. The regularization parameter varies as indicated, and the noise levels shown on the plot correspond to the noise values on $\Delta\phi$, while the noise on $\frac{\Delta\beta}{\beta}$ is kept constant, following a normal distribution with a mean of 3% of the real value. The plot demonstrates that low noise levels degrade performance, while performance remains stable even with very high noise levels.

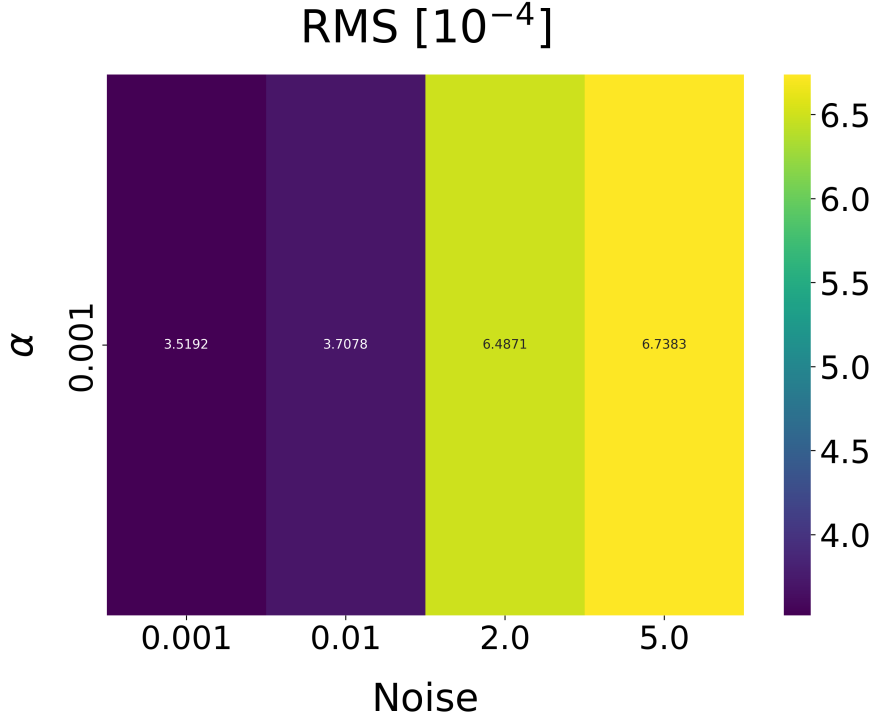


Figure 7.15: The figure shows the RMS values of the residual triplet errors. The model includes only $\Delta\phi$ excluding previously used $\frac{\Delta\beta}{\beta}$ in the features and using field errors as the target. The noise levels in the training set vary as shown, with the regularization parameter kept constant at 0.001. The heatmap demonstrates that in this case higher noise levels can significantly deteriorate the model's precision.

that the model does not predict energy offsets accurately.

This indicates that while incorporating β -beating in the features provides robustness in predicting quadrupole field errors, this robustness does not extend to energy offset predictions. However, phase advances play a substantial role in maintaining the precision of energy offset predictions, suggesting that both β -beating and phase advances are necessary for models targeting energy offsets.

IV Validation with segment-by-segment simulations

The primary objective of this study is to address local beam optics deviations within the triplets of the interaction regions. To achieve this, a focused investigation of the model's performance in these specific regions is essential. Unlike previous evaluations, which primarily assessed the overall predictive accuracy of the models and their ability to reduce global β -beating, a segment-by-segment approach is more suitable for localized analysis. This method involves simulating the beam with measured initial conditions at the start of the interaction region and identifying deviations from the expected nominal behavior. If there is found any deviations it indicates that there is a local error source. This technique can be applied in two ways. First, by validating the hyperparameters of the model or by intergating this methodology into the learning process for model improvement.

Figure 7.18 illustrates the horizontal and vertical β -functions alongside the quadrupole field strength for a nominal beam without errors. The figure highlights the significant increase in the β -function within the triplet quadrupoles of the interaction region. This confirms that errors in interaction regions have a far greater impact on perturbed beam optics than errors occurring in the LHC arcs.

A local, segment-by-segment simulation was performed to investigate the beam's behavior and its response to corrections provided by a machine learning model. This simulation was conducted

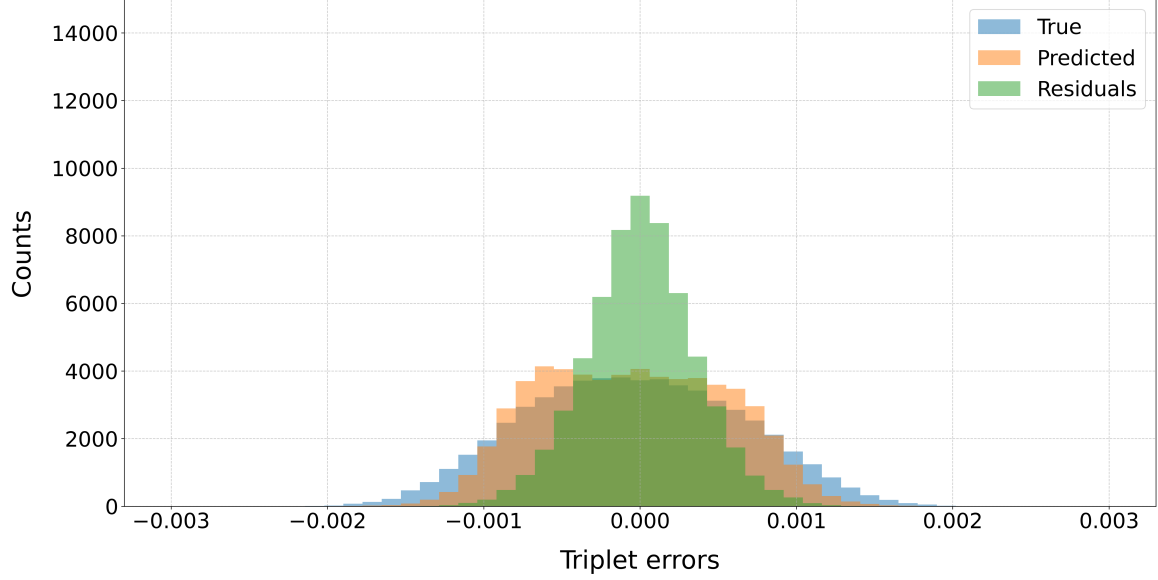


Figure 7.16: The plot illustrates the models' ability to predict triplet values. The machine learning models use only $\frac{\Delta\beta}{\beta}$ as features and include triplet field errors and energy offsets for both beams separately in the targets. The plot demonstrates that the triplet values can be predicted fairly well, as the predictions are reasonably close to the true values.

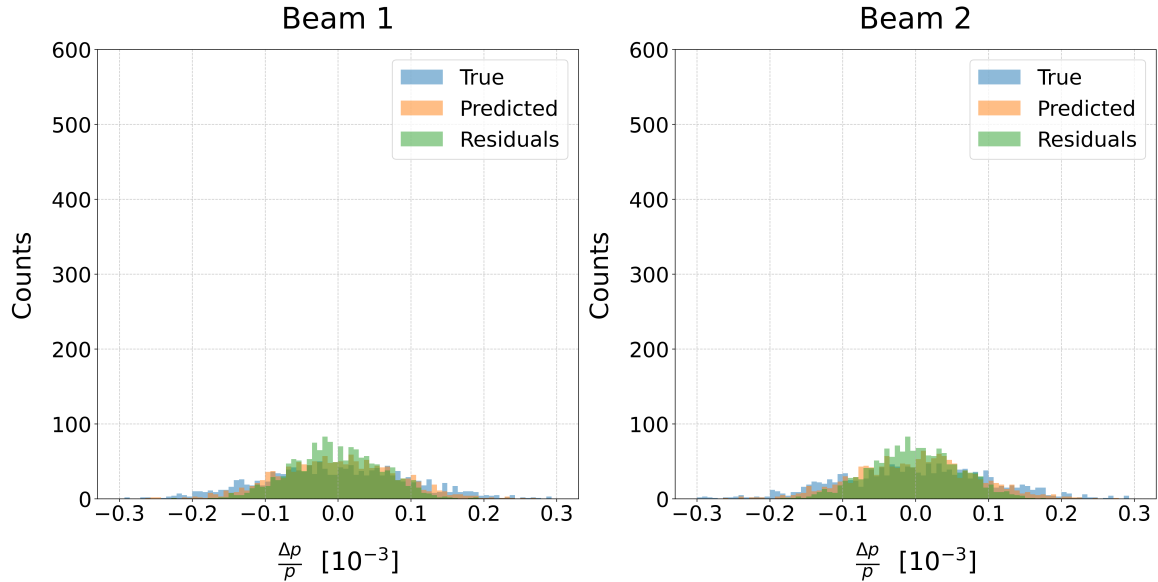


Figure 7.17: The plot illustrates the energy offset predictions for Beam 1 and Beam 2. The machine learning models use only $\frac{\Delta\beta}{\beta}$ as features and include triplet field errors and energy offsets for both beams separately in the targets. In the figure, the residuals (the differences between the true energy offsets and the predicted values) are shown in green. The broad distribution of the residuals indicates that the predictions are quite poor.

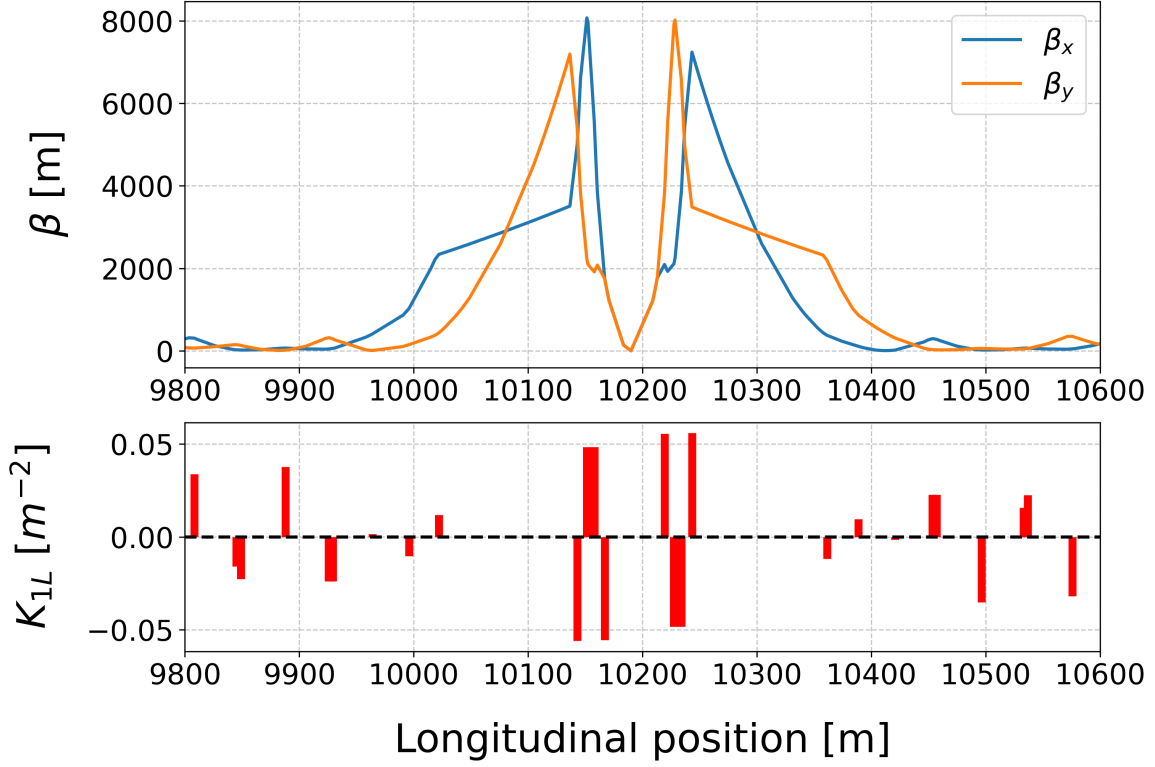


Figure 7.18: The top figure shows the horizontal (blue) and vertical (yellow) β -functions of the nominal beam at IP1. The bottom figure displays the quadrupole field strength in the segment, with the triplet closest to the left and right of the IP point exhibiting greater values compared to the others.

at each interaction point, from the 12th BPM to the left of the IP to the 12th BPM to the right, treating this segment as a line. The initial conditions used in the simulation were the β and α optical functions of a nominal beam, followed by the introduction of segment-specific errors. The goal was to achieve minimal deviations from the nominal beam, as this indicates more effective correction by the model. This approach provides a better assessment of beam correction accuracy locally. To ensure the local results are consistent with the overall performance, a ridge regression model was evaluated with features based on the phase advances across the entire ring and the $(\frac{\Delta\beta}{\beta})$ at the BPMs near interaction points, targeting quadrupole field errors. Different analysis confirmed that this model is significantly more robust than the baseline model, which does not incorporate $\frac{\Delta\beta}{\beta}$ around the IPs.

As practice has shown, the most effective way to investigate beam behavior locally is through the study of the phase advance difference ($\Delta\phi$) in the segment-by-segment method. This difference represents the deviation between the simulated or measured phase advance and the nominal beam phase advance at each location within the interaction region. Figure 7.19 presents the behavior of $\Delta\phi$ under various conditions to better understand the correction process and how this variable responds to it. The x-axis represents the location along the interaction region, starting from the beginning to the end. The y-axis depicts the $\Delta\phi$ variable. The plot shows various scenarios: the results from the simulated beam when the true errors were assigned to the quadrupoles in the region, and scenarios where instead predicted errors were assigned, these predictions are based on the same data from which the true errors were derived. The plot illustrates the correction process, showing the deviation of $\Delta\phi$ when partial errors are assigned to the triplet, for example, the deviation when 50% or 20% of the true errors are applied. It clearly demonstrates that as the errors decrease, $\Delta\phi$ approaches the x-axis, indicating that when the errors are zero in the segment, $\Delta\phi$ will lie on the x-axis. Additionally, the plot shows the difference between the true errors and the predicted errors, illustrating the prediction precision. In this example, the model demonstrates an 80% correction precision for the beam, as indicated by the RMS values of approximately 0.0022 for the 20% deviation and 0.0024 for the corrected errors. Notably, the RMS values for the true and predicted errors are an order of

magnitude larger than those for the corrected errors.

An interesting aspect to explore is the behavior of optical functions in conjunction with the

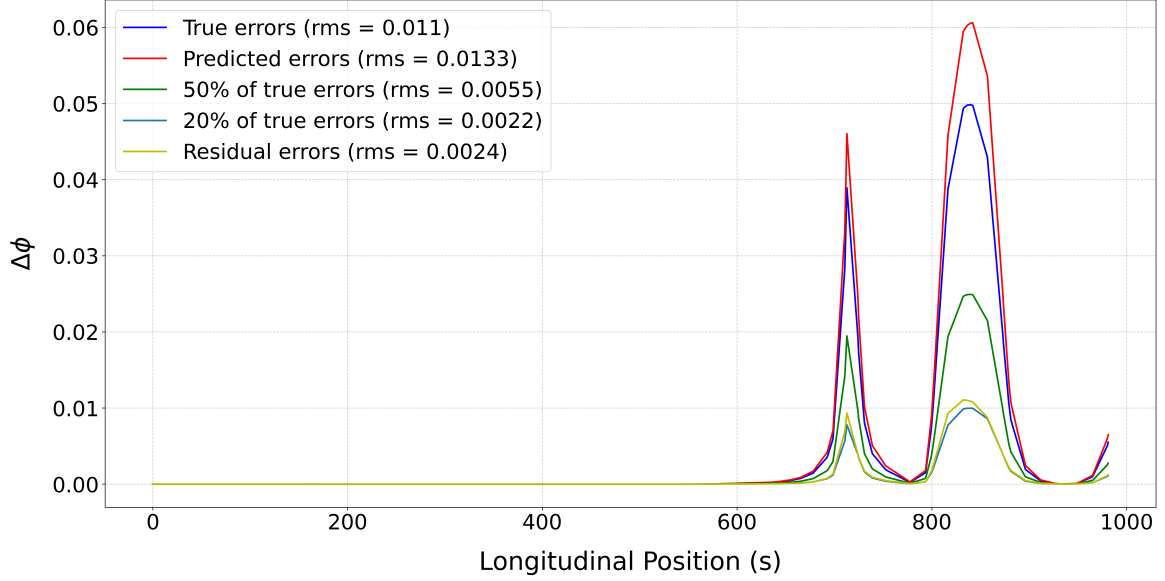


Figure 7.19: Phase advance deviation between simulated and nominal values is plotted against the longitudinal position in the segment containing IP1. Several cases are shown, when the beam is simulated with true errors, when the beam is simulated with predicted errors and the results from the simulated beam with residual errors, which represent the difference between true and predicted errors. The partial deviation serves as a measure of how accurately the predicted errors can correct the beam. Notably, the deviation due to residual errors is approximately 20% of the deviation observed when true errors are applied.

precision of field error corrections. Figures 7.20 and 7.21 illustrate the phase advance and β -beating of the corrected beam, along with the a true, predicted, and residual field error values at the locations of quadrupoles. The beam was simulated with the corrected, residual errors, which are indicated in green on the plot. In this specific example, the phase advance corrected to 80% accuracy, as previously mentioned, and β -beating maintained below 2-3%. These simulations can offer a better figure of merit on the validity of the models compared to the R^2 and MAE, since they better reflect the optics and degeneracy in the machine.

To further explore the impact of hyperparameters on model performance and assess their alignment with previous studies, a series of simulations were conducted, that involved segment corrections with errors rectified by models trained under varying hyperparameters, noise levels on features, and regularization parameters. A total of 50 simulations were performed for each parameter pair at three interaction points (IP5, IP8, and IP1). It is important to note that misalignment errors were excluded from the test data to prevent the model from mistaking these errors for field errors, which could lead to incorrect compensation attempts. Additionally, the ratio of misalignment errors while generating the training data at IP5, IP8, and IP1 was set at 2:1:10, respectively. RMS of the phase advance difference ($\Delta\phi$) in the corrected beam was utilized as a key performance metric. Figures 7.22 and 7.23 illustrate the results for the segments of IP5 and IP8, respectively. The plots depict the mean RMS values of $\Delta\phi$ for each simulation, corresponding to each pair of hyperparameters used to train the model on each beam in both horizontal and vertical directions. The heatmaps indicate increasing RMS values, with a gradient ranging from dark purple to yellow.

The behavior observed in IP5 and IP8 is consistent in terms of their response to hyperparameters. Notably, the regularization parameter does not significantly impact performance in these cases, similar to its effect on triplet predictions (see Fig. 7.14). When introducing noise, specifically a 0.1 noise level on the phase advances in the feature, optimization performance deteriorates, which contrasts with the response observed in error prediction. However, incorporating the $\frac{\Delta\beta}{\beta}$ as an additional feature demonstrated the model's robustness, despite the substantial noise in the phase advance, the

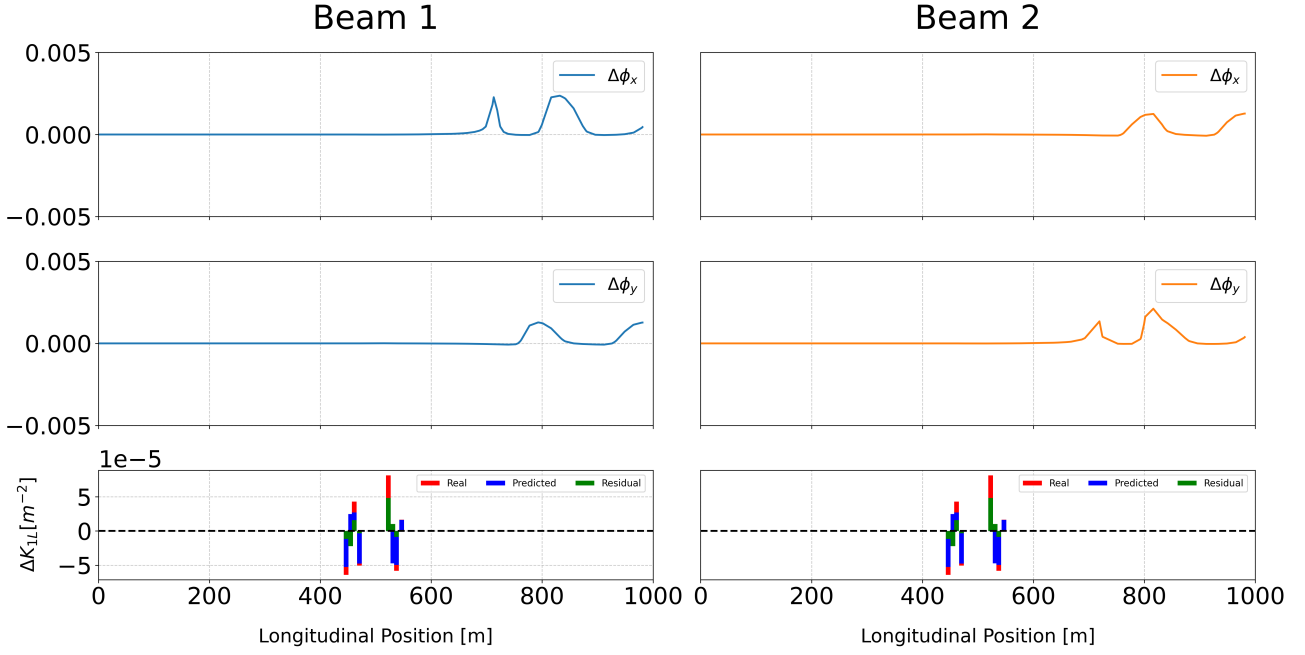


Figure 7.20: Simulated segment of IP1 illustrating the $\Delta\phi$ deviation between simulated and nominal phase advances at each point for both beams in the horizontal and vertical directions. Additionally, the true quadrupole field errors, predicted errors, and the differences between them are depicted in green. The simulation incorporates residual field errors, representing the corrected discrepancies.

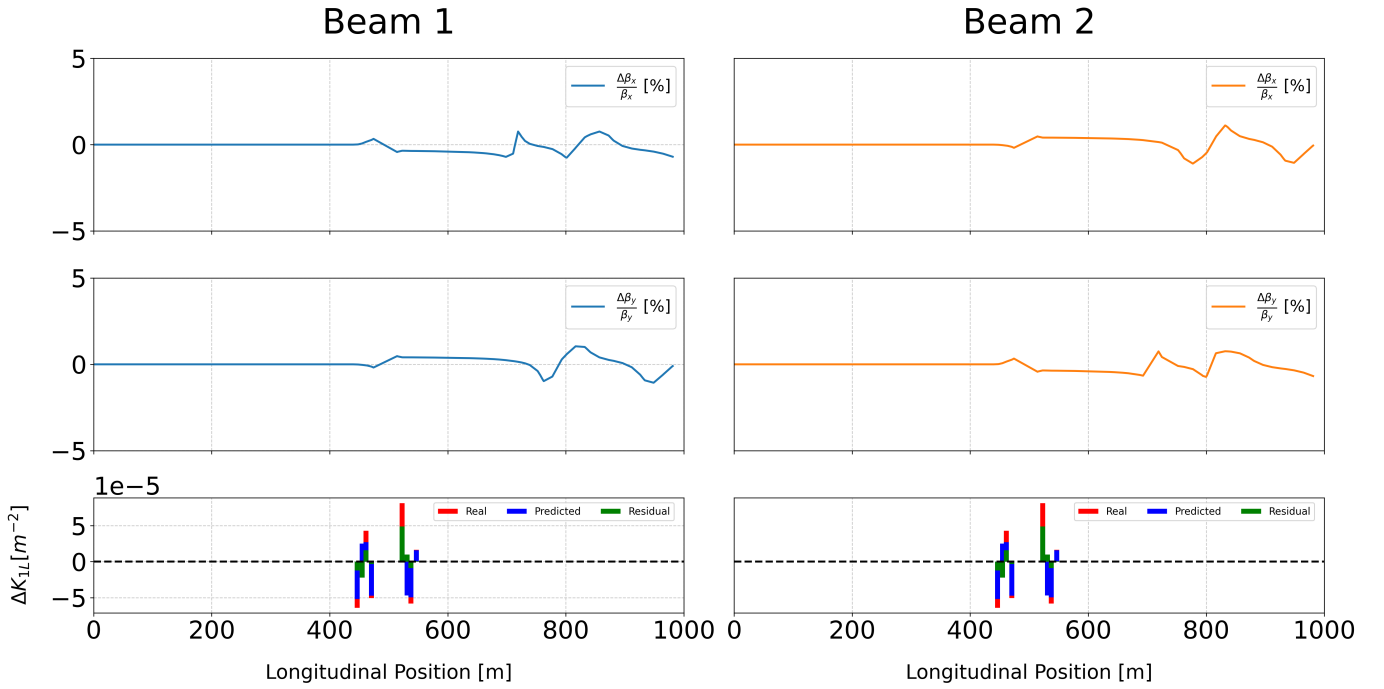


Figure 7.21: Simulated segment of IP1 illustrating β -beating for both beams in the horizontal and vertical directions. Additionally, the true quadrupole field errors, predicted errors, and the differences between them are depicted in green. The simulation incorporates residual field errors, representing the corrected discrepancies.

model still achieved reliable predictions, conversely in the local context, this robustness is no longer observed, and the noise level has a more pronounced negative effect on performance.

It is also worth noting that the RMS values of $\Delta\phi$ for the corrected beam are approximately five times lower at IP8 than IP5, indicating better optimization at IP8. This can be explained by the fact that the β -function at IP5 is larger than at IP8.

In Fig.7.24, the same metric is shown as in Fig.7.22 and Fig.7.23. The figure presents the

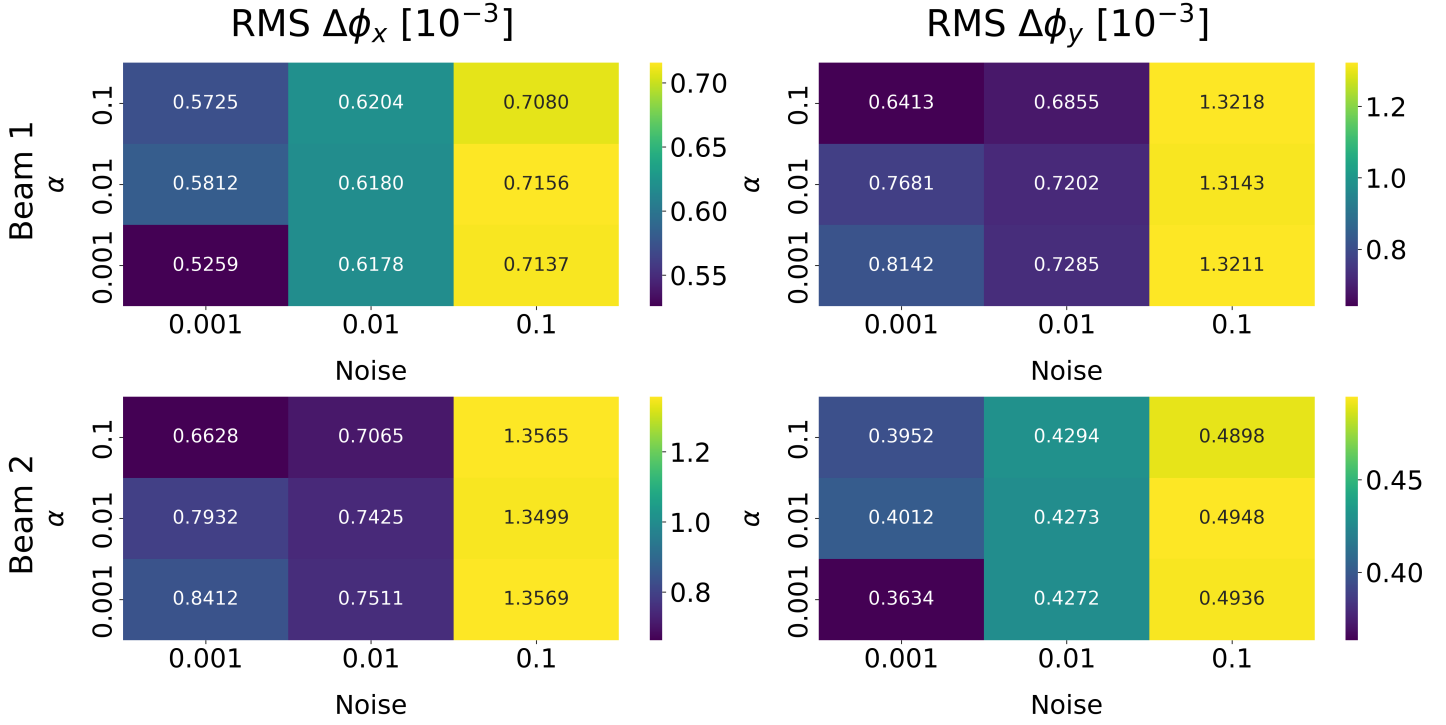


Figure 7.22: The figure shows a heatmap of the mean RMS values of $\Delta\phi$ obtained from 50 simulations at IP5, where the CMS detector is located. The x-axis represents different noise levels applied to the phase advance features, while the y-axis corresponds to the regularization parameters used in the ridge regression model. Higher RMS values are depicted in yellow, indicating greater deviation, while lower RMS values are shown in dark purple, indicating better optimisation. The heatmap illustrates how the model's performance varies with changes in both noise levels and regularization parameters.

mean RMS of $\Delta\phi$ from 50 simulations at IP1 for each pair of hyperparameters in each beam and in each direction. Here, the beam's behavior in response to noise differs notably from the other IPs, with RMS values approximately five times higher than at IP5. However, since the β -functions at IP5 and IP1 are similar in value, they should exhibit the same precision and behavior. The observed difference arises from misalignment of the quadrupoles. As previously mentioned, the ratio of these errors was 1:5 in the data used to train the model. Although these errors were excluded from the test data, they still impact the optimization. This is not evident at IP5 and IP8 because the values were smaller.

In all three plots for IP1, IP5 and IP8, a clear cross-correlation between the beams was observed. The behavior of beam 1 in the x direction mimics the behavior of beam 2 in the y direction. For example, significant deviations in beam 1 along the x axis are consistently accompanied by corresponding deviations in beam 2 along the y axis. This cross-correlation can be attributed to the shared aperture of the triplet magnets, along with the differential behavior of the quadrupoles in different planes, wherein the focusing effect in the x direction corresponds to a defocusing effect in the y direction. This correlation could prove the underlying physics in interaction region.

As the results have demonstrated, the overall beam optimization does not align with the tendencies of local optimizations as shown in the previous discussion. This discrepancy motivates incorporating the segment-by-segment approach into the learning process. By performing both optimizations, the model can adapt to the local behavior of the beam while maintaining effective overall optimization.

One method with which segment by segment approach can contribute the learning process is to include it in the loss function and create the costume loss function. More specifically, the variable that can be used into the costume loss function can be the deviation of phase advances between

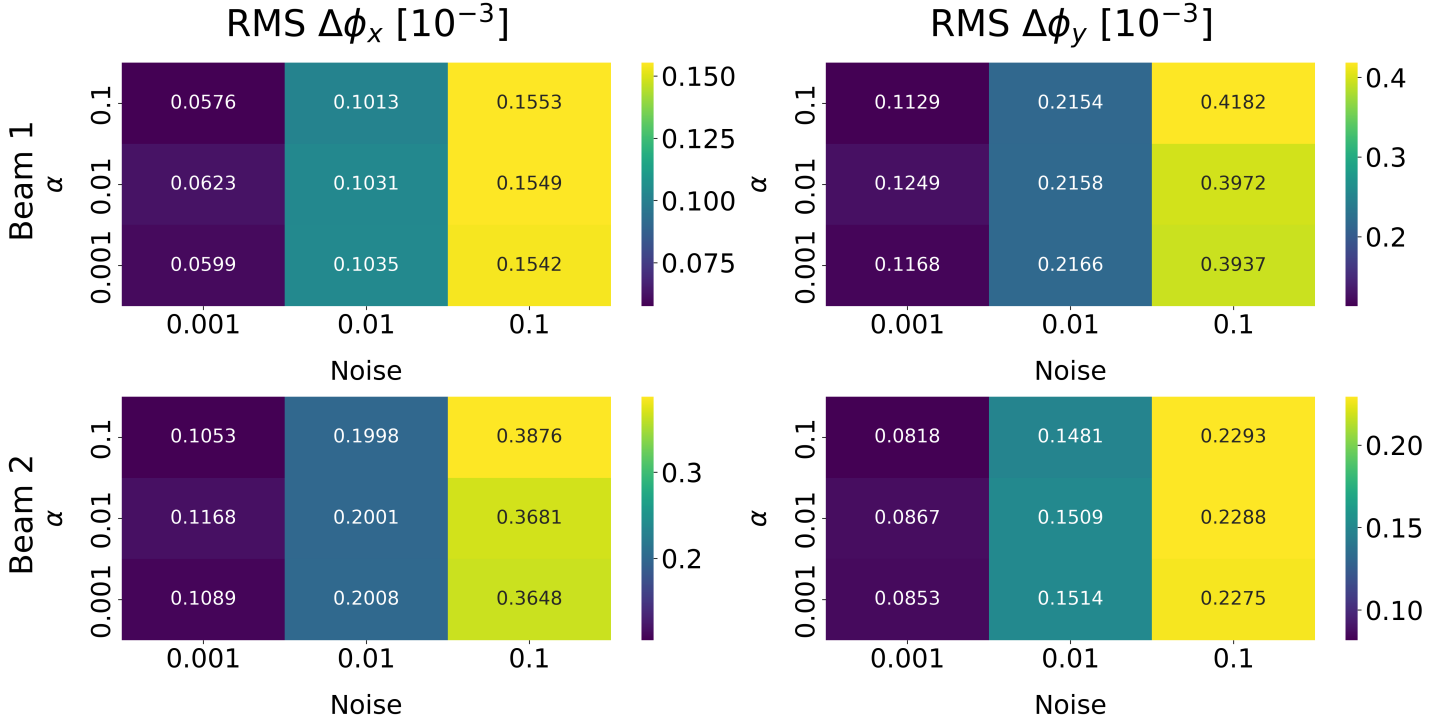


Figure 7.23: Heatmap illustrating the mean RMS of $\Delta\phi$ from a simulation consisting of 50 segments at IP8, the location of the LHCb detector. The x-axis represents the noise level in the phase advance, while the y-axis corresponds to the regularization parameter used during model training. The color gradient from dark purple to yellow indicates the range of values, from lowest to highest.

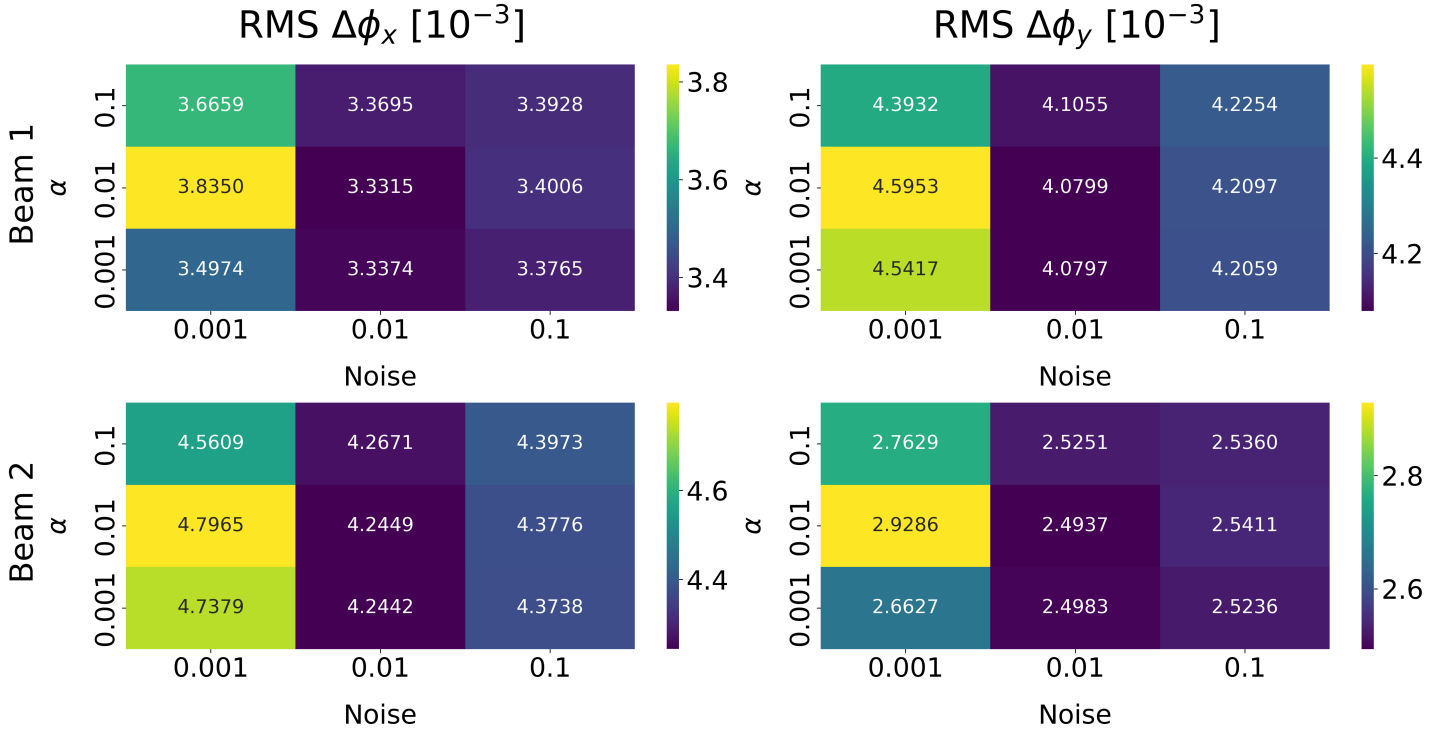


Figure 7.24: The figure illustrates the results for IP1, the location of the ATLAS detector. The plot displays the mean RMS values of $\Delta\phi$ derived from 50 simulations, reflecting the noise level on the phase advance feature and the ridge regression regularization parameter used during training. The values range from dark purple to yellow, indicating increasing magnitudes.

simulated and nominal beam, as it was already mentioned that this variable is the better figure of merit when it comes to analyze the local behaviour of the beam. The new costume loss function

should be built with the previous loss function, the sum of the deviation of phase advance terms and L_2 regularisation, as it is shown in Eq.7.3.

$$L_{costume} = L + \sum \Delta\phi + L_2 \quad (7.3)$$

Chapter 8

Conclusion

This research aimed to investigate the ability of machine learning models to optimize the beam in the Large Hadron Collider (LHC), with a primary focus on local corrections at the interaction points. These IPs, where each detector is located, are the most sensitive to errors. Therefore, optimising the beam optics at IPs is crucial for achieving the desired luminosity and maintaining machine safety.

Several machine learning models were tested. Based on previous studies, ridge regression was identified as the best-performing model. This research further explored the feature configurations relevant to the problem. The study examined two configurations: the first using the feature $\Delta\phi$ —the difference in phase advance between consecutive Beam Position Monitors for simulated and nominal beams and the second configuration adding an extra feature, β -beating at the nearest BPMs left and right of the IPs, with a total of 32 features (two around each of the four IPs, in x and y direction in each beams). The ML models specifically target quadrupole field errors around the IPs, which are the optics errors with the largest impact on beam optics. Several tests were conducted to evaluate model performance, including precision in predictions and overall beam β -beating. In both cases, the model that included the additional feature of $\frac{\Delta\beta}{\beta}$ demonstrated improved results, as the standard deviation of the residual triplet error distribution was smaller for the improved model (2.59×10^{-4}) compared to the base model (2.93×10^{-4}). For the $\frac{\Delta\beta}{\beta}$ distribution in the improved model, the mean (μ) and standard deviation (σ) in the x direction were $\mu = 5.53$ and $\sigma = 3.43$, and in the y direction, $\mu = 5.56$ and $\sigma = 3.66$. In contrast, for the base model, the x direction had $\mu = 5.73$ and $\sigma = 3.73$, while the y direction had $\mu = 5.83$ and $\sigma = 3.93$. These results suggest that including $\frac{\Delta\beta}{\beta}$ as a feature improves the model.

Another significant aspect of the thesis was training the model to predict energy offsets of the beam. Previous experience has shown that energy offsets can complicate optics measurement and correction because the error they introduce mimics the beam's behavior resulting from quadrupole field errors. This problem is particularly pronounced in the LHC, which has two beams, with energy offsets acting as separate error sources for Beam 1 and Beam 2, while the quadrupoles at the IPs are shared between both beams. Attempts have been made to correct these separate errors using shared sources, but this can create a feedback loop that is difficult to optimize. Therefore, a crucial development was creating a model capable of identifying these errors. The study trained the ML model, using ridge regression with the features $\Delta\phi$ and $\frac{\Delta\beta}{\beta}$, to predict energy offsets for both beams separately, in addition to quadrupole errors. The results demonstrated good prediction precision, and further studies investigated how well the model distinguished between quadrupole error sources and energy offsets, with good precision.

In order to make the test set more realistic, a noise level of 10^{-3} was applied to the $\Delta\phi$ values, and a noise level of approximately 3% of the actual value was assigned to $\frac{\Delta\beta}{\beta}$. Incorporating these noise levels into the prediction process made the model's hyperparameters important, particularly the noise in the training set and the regularization parameter used during training. It was found that determining the appropriate values for these parameters is crucial for the model's robustness. The prediction precision varied with different noise levels in the training set. For example, with a noise

level of 0.0001 in the training set, the RMS value of residual field errors (i.e., the difference between true and predicted values) was 5.52×10^{-4} . In contrast, for a noise level of 0.001, the RMS of residuals decreased to 3.23×10^{-4} . This indicates that difference in the hyperparameter can significantly affect prediction precision. To investigate this issue further, several machine learning models were trained with different combinations of hyperparameters, and their prediction precision was evaluated. The studies revealed that for low noise levels, the predictions were not precise. For instance, with a noise level of 0.0001 and regularization parameters of 0.001 and 0.0001, the RMS of residual field errors was approximately three times larger compared to the optimal case. Additionally, the studies demonstrated that the model maintained good prediction precision even with substantially higher noise levels in the $\Delta\phi$ features.

This result suggests that incorporating $\frac{\Delta\beta}{\beta}$ as an additional feature significantly enhanced the model's robustness. To verify this, an ML model was trained using only the $\frac{\Delta\beta}{\beta}$ feature, and its prediction performance was evaluated. The results showed that the precision was fairly good for the field errors. However it was also found that the model struggled to accurately predict energy offsets, indicating that $\Delta\phi$ is important parameter for precise offset predictions.

The primary focus of this study was the local correction of optics errors in the interaction regions, achieved through a segment-by-segment approach. This approach involves simulating specific regions of the machine, treated as linear segments. The simulations were conducted at IP1, IP5, and IP8, corresponding to the locations of the ATLAS, CMS, and LHCb detectors, respectively. Evaluating the optics function within each segment provided a more accurate approach of the model's validity. The results demonstrated that under favorable conditions, the model could locally correct the beam optics parameters with an 80% precision rate. Because the impact of hyperparameters on the model's robustness was significant, extensive testing was conducted to determine which configurations minimized the RMS of the deviation of phase advances between simulated and nominal beams, both horizontally and vertically in both beam 1 and beam 2. Notably, quadrupole misalignments were excluded from the test set to prevent the model from confusing them with field error sources. The results showed that corrections at IP5 and IP8 were more effective than those at IP1. This can be attributed to the larger misalignment errors at IP1 compared to the other IRs, with the ratio of misalignment values at IP1, IP5, and IP8 being 10:2:1. Although these misalignments were not present in the test set, they still influenced the correction process, as the model had been trained with these errors. Additionally, the correction at IP8 was superior to that at IP5, likely due to the lower β -function at IP8. Most interestingly, the study also found that the responses at IP5 and IP8 differed from previous studies in terms of the response to hyperparameters. While the model previously exhibited strong robustness when incorporating the β -beating, this was no longer the case. These findings suggest that further investigation is needed, particularly by integrating the segment-by-segment approach into the training process. The simulation results from this approach could be incorporated into a custom loss function for regression. Moreover, the results demonstrated a clear cross-correlation between the beams, the behavior of beam 1 in the x direction mirrored that of beam 2 in the y direction. This correlation can be explained by the shared aperture of the triplet magnets and the different behavior of the quadrupoles in different planes.

In conclusion, the inclusion of the $\frac{\Delta\beta}{\beta}$ features yielded promising results by enhancing the model's ability to predict beam optics errors. The model also demonstrated excellent performance in predicting energy offsets. However, the segment-by-segment approach demonstrated that this approach should be included into the learning process.

Bibliography

- [1] D. Wilson. *Rutherford: Simple genius*. MIT Press, Cambridge, MA, 1983.
- [2] W. A. Barletta, M. Bai, M. Battaglia, O. Bruning, J. Byrd, R. Ent, J. Flanagan, W. Gai, J. Galambos, G. Hoffstaetter, et al. Planning the future of us particle physics (snowmass 2013): Chapter 6: Accelerator capabilities. *arXiv preprint arXiv:1401.6114*, 2014.
- [3] M Benedikt, P Collier, V Mertens, J Poole, and K Schindl. Lhc design report, volume iii, the lhc injector chain. *CERN, Geneva*, page 35, 2004.
- [4] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, R. Aben, B. Abi, M. Abolins, O. S. AbouZeid, et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.
- [5] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.
- [6] T. Persson, F. Carlier, J. M. Coello de Portugal, A. Garcia-Tabares Valdivieso, A. Langner, E. H. Maclean, L. Malina, P. Skowronski, B. Salvant, R. Tomás, et al. Lhc optics commissioning: A journey towards 1% optics control. *Physical Review Accelerators and Beams*, 20(6):061002, 2017.
- [7] E. H. Maclean, R. Tomás, F. S. Carlier, M. Solfaroli Camillocci, J. W. Dilly, J. M. Coello de Portugal, E. Fol, K. Fuchsberger, A. Garcia-Tabares Valdivieso, M. Giovannozzi, et al. New approach to lhc optics commissioning for the nonlinear era. *Physical Review Accelerators and Beams*, 22(6):061004, 2019.
- [8] R. Tomás, T. Bach, R. Calaga, A. Langner, Y. I. Levinsen, E. H. Maclean, T. H. B. Persson, P. K. Skowronski, M. Strzelczyk, G. Vanbavinckhove, et al. Record low β beating in the lhc. *Physical Review Special Topics—Accelerators and Beams*, 15(9):091001, 2012.
- [9] A. Langner, G. Benedetti, M. Carla, U. Iriso, Z. Martí, J. M. Coello de Portugal, and R. Tomás. Utilizing the n beam position monitor method for turn-by-turn optics measurements. *Physical Review Accelerators and Beams*, 19(9):092803, 2016.
- [10] A. Wegscheider, A. Langner, R. Tomás, and A. Franchi. Analytical n beam position monitor method. *Physical Review Accelerators and Beams*, 20(11):111002, 2017.
- [11] R. Calaga, R. Tomás, R. Miyamoto, and G. Vanbavinckhove. Beta* measurement in the lhc based on k-modulation. Technical Report CERN-ATS-2011-149, CERN, 2011.
- [12] M. Kuhn, V. Kain, G. Trad, R. Tomás, R. Steinhagen, and B. Dehning. New tools for k-modulation in the lhc. Technical report, 2014.
- [13] F. Carlier and R. Tomás. Accuracy and feasibility of the β^* measurement for lhc and high luminosity lhc using k modulation. *Physical Review Accelerators and Beams*, 20(1):011005, 2017.

- [14] M. Hofer and R. Tomás. Effect of local linear coupling on linear and nonlinear observables in circular accelerators. *Physical Review Accelerators and Beams*, 23(9):094001, 2020.
- [15] A. Langner, J. M. Coello de Portugal, P. Skowroński, and R. Tomás. Developments of the segment-by-segment technique for optics corrections in the lhc. Technical Report CERN-ACC-2015-331, CERN, 2015. MOPJE054.
- [16] H. García Morales, R. Tomás García, and J. Cardona. Jacow: Comparison of segment-by-segment and action-phase-jump techniques in the calculation of ir local corrections in lhc. *JA-CoW IPAC*, 2021:632–635, 2021.
- [17] J. Coello de Portugal, R. Tomás, and M. Hofer. New local optics measurements and correction techniques for the lhc and its luminosity upgrade. *Physical Review Accelerators and Beams*, 23(4):041001, 2020.
- [18] E. Fol, R. Tomás, and G. Franchetti. Supervised learning-based reconstruction of magnet errors in circular accelerators. *The European Physical Journal Plus*, 136(4):365, 2021.
- [19] CERN. The cern accelerator complex. <https://www.home.cern/science/accelerators/accelerator-complex>. Accessed: 2024-06-06.
- [20] O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. Lhc design report (volume i, the lhc main ring). *Reports- CERN*, 2004.
- [21] G. Apollinari, I. Béjar Alonso, O. Brüning, M. Lamont, and L. Rossi. High-luminosity large hadron collider (hl-lhc): Preliminary design report. Technical report, Fermi National Accelerator Lab. (FNAL), Batavia, IL, United States, 2015.
- [22] L. Stoel, W. Herr, T. Kramer, A. Milanese, G. Rumolo, B. Goddard, W. Bartmann, E. Shaposhnikova, F. Burkart, and M. Barnes. High energy booster options for a future circular collider at cern. Report CERN-ACC-2016-291, CERN, 2016.
- [23] S.-Y. Lee. *Accelerator physics*. World Scientific Publishing Company, 2018.
- [24] A. Wolski. *Beam dynamics in high energy particle accelerators*. World Scientific, 2014.
- [25] H. Wiedemann. *Particle accelerator physics*. Springer Nature, 2015.
- [26] E. D. Courant and H. S. Snyder. Theory of the alternating-gradient synchrotron. *Annals of Physics*, 3(1):1–48, 1958.
- [27] V. Kain. Beam dynamics and beam losses-circular machines. *arXiv preprint arXiv:1608.02449*, 2016.
- [28] S. D. Fartoukh and M. Giovannozzi. Target values for the azimuthal misalignment of the lhc magnets. Technical report, 2005.
- [29] scikit learn. Underfitting vs. overfitting. https://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.html. Accessed: 2024-07-28.
- [30] A. Nosych, U. Iriso, A. Olmos, and M. Wendt. Overview of the geometrical non-linear effects of button bpms and methodology for their efficient suppression. In *Proceedings of the 3rd International Beam Instrumentation Conference*, pages 298–302, 2014.
- [31] M. Bai, S. Y. Lee, J. W. Glenn, H. Huang, L. Ratner, T. Roser, M. J. Syphers, and W. Van Asselt. Experimental test of coherent betatron resonance excitations. *Physical Review E*, 56(5):6002, 1997.
- [32] M. Bai, L. Ahrens, J. Alessi, K. Brown, G. Bunce, P. Cameron, C. M. Chu, J. W. Glenn, H. Huang, A. E. Kponou, et al. Overcoming intrinsic spin resonances with an rf dipole. *Physical Review Letters*, 80(21):4673, 1998.

- [33] R. Tomás. Adiabaticity of the ramping process of an ac dipole. *Physical Review Special Topics—Accelerators and Beams*, 8(2):024401, 2005.
- [34] R. Miyamoto, S. E. Kopp, A. Jansson, and M. J. Syphers. Parametrization of the driven betatron oscillation. *Physical Review Special Topics—Accelerators and Beams*, 11(8):084002, 2008.
- [35] T. Nissinen, A. Wegscheider, M. Le Garrec, F. Carlier, R. Tomás García, E. Maclean, and T. Persson. Jacow: Lhc optics measurements from transverse damper for the high intensity frontier. *JACoW HB*, 2023:479–482, 2024.
- [36] F. Carlier, T. Levens, J. Keintzel, T. H. B. Persson, E. Fol, V. Riesen-Haupt, M. Le Garrec, A. Costa Ojeda, F. Soubelet, A. Wegscheider, et al. Jacow: Challenges of k-modulation measurements in the lhc run 3. *JACoW IPAC*, 2023:MOPL014, 2023.
- [37] CERN. MAD-X — Methodical Accelerator Design. <https://madx.web.cern.ch/>. Accessed: 2024-08-06.
- [38] R. Tomás, M. Aiba, A. Franchi, and U. Iriso. Review of linear optics measurement and correction for charged particle accelerators. *Physical Review Accelerators and Beams*, 20(5):054801, 2017.
- [39] M. Aiba, S. Fartoukh, A. Franchi, M. Giovannozzi, V. Kain, M. Lamont, R. Tomás, G. Vanbavinckhove, J. Wenninger, F. Zimmermann, et al. First β -beating measurement and optics analysis for the cern large hadron collider. *Physical Review Special Topics—Accelerators and Beams*, 12(8):081002, 2009.
- [40] R. Tomás, O. Brüning, M. Giovannozzi, P. Hagen, M. Lamont, F. Schmidt, G. Vanbavinckhove, M. Aiba, R. Calaga, and R. Miyamoto. Cern large hadron collider optics model, measurements, and corrections. *Physical Review Special Topics—Accelerators and Beams*, 13(12):121004, 2010.