
Application of Machine Learning in Beam Optics Measurements and Corrections

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Physik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Elena Fol
aus Nowosibirsk

Frankfurt am Main 2021
(D 30)



ABSTRACT

The present research in high energy physics as well as in the nuclear physics requires the use of more powerful and complex particle accelerators to provide high luminosity, high intensity, and high brightness beams to experiments. With the increased technological complexity of accelerators, meeting the demand of experimenters necessitates a blend of accelerator physics with technology. The problem becomes severe when optimization of beam quality has to be provided in accelerator systems with thousands of free parameters including strengths of quadrupoles, sextupoles, RF voltages, etc. Machine learning methods and concepts of artificial intelligence are considered in various industry and scientific branches, and recently, these methods are used in high energy physics mainly for experiments data analysis.

In Accelerator Physics the machine learning approach has not found a wide application yet, and in general the use of these methods is carried out without a deep understanding on their effectiveness with respect to more traditional schemes or other alternative approaches. The purpose of this PhD research is to investigate the methods of machine learning applied to accelerator optimization, accelerator control and in particular on optics measurements and corrections. Optics correction, maximization of acceptance, and simultaneous control of various accelerator components such as focusing magnets is a typical accelerator scenario. The effectiveness of machine learning methods in a complex system such as the Large Hadron Collider, which beam dynamics exhibits nonlinear response to machine settings is the core of the study. This work presents successful application of several machine learning techniques such as clustering, decision trees, linear multivariate models and neural networks to beam optics measurements and corrections at the LHC, providing the guidelines for incorporation of machine learning techniques into accelerator operation and discussing future opportunities and potential work in this field.

ZUSAMMENFASSUNG

Das Feld der künstlichen Intelligenz wird durch das Ziel getrieben, Maschinen mit menschlicher Intelligenz zu versorgen. Die Methoden des maschinellen Lernens (ML) und Konzepte aus dem Gebiet der Künstlichen Intelligenz sind in vielen industriellen und wissenschaftlichen Branchen vertreten. In der Hochenergie- und Teilchenphysik werden diese Methoden zum größten Teil für die Analyse von den Experimentaldaten, Identifikation von neuen Ereignissen und zum Lösen der Regressionsprobleme verbreitet angewendet. In der Beschleunigerphysik dagegen hat der ML Ansatz noch keine breite Anwendung gefunden und soll daher erforscht werden.

Fragestellungen im Bereich der Beschleunigerphysik befassen sich oft mit komplexen Optimierungsproblemen, wobei tausende von freien Parametern und zeitlich variierenden Strahleigenschaften und Beschleunigereinstellungen, die eine nicht-lineare Beziehung aufweisen berücksichtigt werden müssen. Diese Besonderheiten entsprechen den Einschränkungen der traditionellen Optimierungsmethoden und machen dieses Problem zu einem perfekten Kandidaten für die Anwendung von ML-basierten Techniken.

Traditionelle Optimierungsansätze demonstrieren eine hervorragende Leistung, wenn sie auf lineare Optimierungsaufgaben und Probleme mit einer begrenzten Anzahl von Optimierungszielen angewendet werden. Größere Herausforderungen ergeben sich, wenn die Diagnose eines komplexen nichtlinearen Verhaltens und mehrerer Variablen erforderlich ist. Vor allem, kann der für die Anwendung traditioneller Methoden benötigte Rechen- und Zeitaufwand für zukünftige Beschleunigeranlagen, deren Komplexität mit der Zeit kontinuierlich ansteigt, unakzeptabel werden. Eine weitere Einschränkung traditioneller Optimierungsmethoden besteht darin, dass die Zielfunktion oder spezifische Regeln und numerische Ausgangsparameter bekannt sein müssen. ML-Methoden hingegen sind in der Lage, beliebige Funktionen zu approximieren, ohne dass explizite Regeln erforderlich sind. Dies zeigt einen signifikanten Vorteil im Vergleich zu traditionellen Ansätzen zur Datenanalyse und Optimierung.

Das zentrale Ziel dieser Arbeit ist es, die Methoden aus dem Bereich des ML zu untersuchen, die effizient integriert werden können, um den Prozess der Strahloptikmessungen und -korrektur am LHC zu optimieren, indem bestimmte Schritte der Analyse automatisiert und im Bezug auf Genauigkeit der Ergebnisse, und erforderlichen Zeitaufwand verbessert werden. Dies soll dazu dienen, um eine methodologische Vorgehensweise der Auswahl, Implementierung, Evaluation und Integration der ML Anwendungen in Beschleunigeroptimierung und -operation zu demonstrieren.

Somit, wird der Fokus nicht nur auf die praktische Umsetzung der ML Konzepte zur Verbesserung der Optikmessungen gelegt, sondern vielmehr auf die Entwicklung einer allgemeiner Methode, um die ML-Algorithmen im Bereich der Beschleunigerphysik erfolgreich zu integrieren.

Der Large Hadron Collider (LHC) ist ein Teilchenbeschleuniger, der von der Europäischen Organisation für Kernforschung (CERN) betrieben wird und für die Erforschung bekannter und noch unbekannter Elementarteilchen und Materiezustände eingesetzt wird. In Bezug auf Energie und Häufigkeit der Teilchenkollisionen ist der LHC der leistungsstärkste Teilchenbeschleuniger der Welt. Da die Maschine sehr enge Toleranzen hat, ist eine gute Kontrolle des Teilchenstrahls entscheidend für den Maschinenschutz und die Leistung. Darüber hinaus ist die Kontrolle der Strahloptik ein wichtiger Faktor für die Erzielung der hohen Luminosität an den Hauptexperimenten CMS und ATLAS.

Im folgenden, wird der Aufbau der Arbeit und die konzeptionelle Vorgehensweise präsentiert. Zunächst müssen die Schwächen der aktuell verfügbaren traditionellen Methoden ermittelt werden, um potenzielle ML-Aufgaben zu definieren. Eine beschränkte Auswahl anwendbarer Algorithmen für jedes der identifizierten Probleme sollte analysiert werden, um die am besten geeignete Technik und mögliche Alternativen zu bestimmen. Die technische Implementierung, die Integration in der vorhandenen Software-Infrastruktur und die Anpassung der verfügbaren ML-Algorithmen an den gegebenen Problemkontext sind ebenfalls ein wesentlicher Bestandteil dieser Forschung. Um die Bewertung der entwickelten ML-basierten Lösungen durchführen zu können, müssen umfangreiche Simulationsstudien und Tests durchgeführt werden, gefolgt von der Analyse der erzielten Ergebnisse und der Bestimmung zukünftiger Pläne und potenzieller Verbesserungen. Der Schwerpunkt liegt nicht nur auf der Entwicklung praktischer Lösungen, sondern auch auf der Demonstration am Beispiel der Einführung von ML-Techniken in optische Messungen und Korrekturen am LHC, wie geeignete ML-Methoden für bestimmte Aufgabentypen ausgewählt und validiert werden sollen, um die Leistung eines Beschleunigers zu verbessern und das detaillierte Verständnis angewandter ML Lösungen zu vermitteln.

Um einen kurzen, aber für das Verständnis der präsentierten Konzepte ausreichenden Hintergrund zu verschaffen, werden in dem ersten Kapitel die grundlegenden Prinzipien des maschinellen Lernens, als auch der Strahloptik und Strahldynamik vorgestellt. Allerdings, umfasst diese Einführung ausschliesslich die Beschreibung der ML-Algorithmen, die in den demonstrierten neu entwickelten Methoden angesetzt wurden. Die Einführung in die Strahloptik enthält besondere Konzepte, wie z.B. die Methoden zur Optikkorrektur, die speziell für Speicherringe entwickelt wurden. Ein wichtiger Teil der Einführung ist die Diskussion über die aktuellen Fortschritte im Bereich der Anwendung von ML in Beschleunigerphysik. Dies soll dem Leser einen Einblick in den Stand der Forschung erschaffen, um die in der vorliegenden Dissertation

erarbeitete Methode einwandfrei in dem Themengebiet einordnen zu dürfen. Die nächsten drei Kapiteln stellen die bestimmten Herausforderungen, die im Laufe dieser Arbeit gelöst wurden vor. An dieser Stelle wird eine kurze Zusammenfassung der entwickelten Methoden gegeben.

Automatische Erkennung von Instrumentenfehlern am Beispiel von anomalen Signalen von fehlerhaften Strahlpositionsmonitoren.

Fehlerhafte Beam Position Monitors (BPMs) erzeugen ein fehlerhaftes Signal, das eine unzuverlässige Berechnung der Optikfunktionen verursacht. Daher ist die Erkennung fehlerhafter BPMs vor der Optikberechnung für eine präzise Optikanalyse von entscheidender Bedeutung. Die meisten Fehler können durch Anwendung traditioneller Methoden der Signalverarbeitung, wie Singular Value Decomposition (SVD) behoben werden. Optikfunktionen, die aus den gereinigten Turn-by-Turn-Daten rekonstruiert wurden, weisen jedoch regelmässig einige unphysikalische Werte auf, die auf die Präsenz verbleibender fehlerhafter BPMs hinweisen. Es wurde eine neue Methode entwickelt, die auf dem Isolation Forest Algorithmus aus dem Bereich des unüberwachten Lernens basiert und die erforderliche Leistung im Laufe der Messungen demonstriert und zu Verbesserung der Qualität der Optikanalyse führt.

Diese Technik wurde erfolgreich in die Optikanalyse integriert, um fehlerhafte Signale vor der Berechnung der Optikfunktion automatisch zu entfernen, wobei eine manuelle Datenbereinigung und anschliessende Wiederholung des kompletten Analysevorgang vermieden werden kann. Zusammen mit der Integration der neuen Methode wurden die Parameterwerte von SVD-basierten Methode und die numerischen Grenzwerte für das Filtern der fehlerhaften Signale optimiert, um die Anzahl der verbleibenden fehlerhaften BPMs zu minimieren, und gleichzeitig sicherzustellen, dass relevante Signalartefakte erhalten bleiben, die zur Berechnung von wichtigen Optikfunktionen erforderlich sind. Diese Optimierung sowie die Identifizierung optimaler Einstellungen des Isolation Forest Algorithmus waren möglich, da realistische Simulationen fehlerhafter BPM-Signale auf der Grundlage statistischer Analysen früherer Messungen und bekannter BPM-Fehler entwickelt wurden. Insgesamt ermöglichte die Anwendung der neuen Methode zusammen mit der Optimierung der Einstellungen der zuvor vorhandenen Werkzeuge, die Anzahl der unphysikalischen Ausreisser in den optischen Funktionen um den Faktor 2 bis weniger als 1% der Gesamtzahl der verfügbaren BPMs zu reduzieren. Dies wurde durch die realistischen Simulationen demonstriert. Ausserdem, wurde der ausgewählte Algorithmus mit anderen Clustering Methoden verglichen. Darüber hinaus wurde an experimentellen LHC-Daten gezeigt, dass das neue Verfahren BPM-Fehler identifizieren kann, die a priori unsichtbare Fehlermuster aufweisen. Nach der vollständigen Umsetzung wurde die neue Methode als Standardbestandteil der Optikanalyse verwendet, um die Qualität der Messdaten zu verbessern. Die Ergebnisse der letzten LHC-Messungen zeigen, dass mit Isolation Forest das Auftreten von 80% der fehlerhaften Signale verhindert werden kann, die sonst in den Daten verbleiben und die Optikanalyse und aufeinander folgende Korrekturen negativ beeinflussen. Das verbesserte Verständnis der in dieser Arbeit erhaltenen Kombination von SVD-Cleaning und IF-Algorithmen

wird in Zukunft die LHC-Messungen weiter unterstützen, insbesondere angesichts der für den LHC-Lauf III implementierten Upgrades des BPM-Systems.

Gleichzeitige Rekonstruktion von mehr als tausend quadrupolaren Gradientenfeldfehlern von Quadrupolmagneten basierend auf den gemessenen Strahloptikfunktionen.

Regressionsmodelle, die auf dem Konzept des Unüberwachten Lernens basieren, ermöglichen es, die physikalischen Korrelationen zwischen einer grossen Anzahl von unabhängigen Variablen (3304 Werte des Phasenvosprungs gemessen an allen BPMs in LHC) und 1256 Magnetfehler als Ausgangsvariablen zu lernen. Damit, kann die Korrektur der Optik anhand von ermittelten Magnetfehlern erfolgen. Der besondere Vorteil dieser Methode besteht darin, dass die Korrekturen für beide Strahlen an allen Positionen im Ring, inklusive Interaction Points, wo die Experimente sich befinden, gleichzeitig berechnet werden können.

Die Ergebnisse von der Evaluierung der Methode an den Simulationsdaten mit einer anspruchsvollen Optikskonfiguration ($\beta^* = 40$ cm) weist einen relativen Prädiktionsfehler von 30% bzw. 16% für die Triplett- bzw. Arcs-Magnete. Die Anwendung eines kombinierten Modells basierend auf der sogenannten "Bagging" Technik auf experimentelle LHC Daten zeigte, dass durch die Simulation des β -beatings mit den vorhergesagten Magnetfehlern zu einer genauen Rekonstruktion der ursprünglich gemessenen β -Beating führt, mit einer Differenz von 7% in den Arc-Regionen und unter 3% in den beiden Experimenten. Diese Differenz entspricht der potentiellen Abweichungen der gemessenen Optikfunktionen von Design nach der Implementierung der Korrektur, die anhand von mithilfe des ML-Modells berechneten Magnetfehlern erstellt ist. Der neu entwickelte Ansatz zur Rekonstruktion von Quadrupolarfehlern wurde erfolgreich an Simulationen und historischen LHC-Daten demonstriert. Der wichtigste zukünftige Schritt ist eine experimentelle Demonstration der entwickelten ML-basierten Optikkorrekturen im LHC Run III sowie die Integration der Methode in die Softwareinfrastruktur der online Optikanalyse.

Reduzierung des Störsignals und Rekonstruktion des Phasenvosprungs an den Positionen der fehlerhaften BPMs, wo keine Messung möglich ist.

Die in Simulationen gezeigte Reduzierung des Störsignale wurde durch Anwendung von einer speziellen Art der neuronalen Netze, nämlich Autoencoder. Der erzielte Vorteil liegt nicht nur darin, dass die Messungen der Synchrotronphase mithilfe des Autoencoders entauscht werden können, sondern auch in der Möglichkeit, gleichzeitig das fehlende Signal zu rekonstruieren. Durch das Trainieren des Neuralen Netzes an zahlreichen Simulationsdaten, die aus einem vollständigen Satz von Phasenvorsprungswerten an jedem BPM bestehen, ist es möglich, eine zuverlässige Rekonstruktion des Phasenvorsprungs an den Positionen fehlerhafter BPMs zu erhalten. Das Turn-by-Turn Signal von diesen BPMs wird bei den vorherigen Schritten der Optikanalyse aus den Messdaten entfernt und somit, ist die Berechnung durch die traditionellen Methoden nicht

möglich.

Dieser Ansatz wurde anhand historischer LHC-Messdaten verifiziert, was den Ergebnissen von Simulationen entspricht und eine hohe Übereinstimmung zwischen gemessenen und vorhergesagten Werten demonstriert. Die genaue Rekonstruktion fehlender Phasenversprungswerte zusammen mit der Reduzierung des Rauschens verspricht Verbesserungen in der quadrupolaren Fehlerrekonstruktion, da die Qualität von Regressionsmodellen stark von den Signalstörungen in den Eingangsvariablen abhängt. Im Allgemeinen öffnet die Reduzierung des Rauschpegels bei den Phasenmessungen, die die Grundlage der Optikanalyse und Korrekturberechnung sind, Türen für weitere Verbesserungen der Genauigkeit und Qualität der gesamten Optikanalyse.

Virtuelle Diagnostik zur Berechnung der optischen Funktionen, ohne ohne Durchführung von direkten Strahlmessungen.

Die Anwendung von linearen Regressionsmodellen ermöglicht es, Schätzwerte bestimmter optischen Funktionen zu erhalten, deren Berechnung ansonsten zusätzliche Strahlmanipulationen erfordert. Dies ist z.B. der Fall für die Messung der horizontalen Dispersion oder die Durchführung komplexer zeitaufwändiger Methoden wie k-Modulation, um β^* zu messen. In dieser Studie werden überwachte Regressionsmodelle verwendet, um eine schnelle und präzise Vorhersage von zwei wichtigen Observablen der linearen Optik zu ermöglichen: horizontale normalisierte Dispersion in dem gesamten LHC und die β -Funktion an den BPMs links und rechts von Injection Points. In beiden Fällen verwendet das Regressionsmodell die Phasenmesswerte als Inputvariablen, die bei jeder Messdatenerfassung durch harmonische Analyse von Turn-by-Turn-Daten verfügbar sind. Der Ansatz wurde an einer Vielzahl von LHC-Simulationen verifiziert, die eine Genauigkeit von 7% für die Rekonstruktion der Dispersionsfunktion und 1% für die Rekonstruktion von β -Funktionswerten aufweisen.

Die neuen Methoden wurden speziell für die Anwendung am LHC implementiert, teilweise sind sie bereits in die Softwareinfrastruktur der online Optik-Analyse integriert. Dies erfolgte nach einer umfassender Verifizierung an realistischen Simulationen sowie an historischen LHC-Daten aus vergangenen Optikmessungen. Die erzielten Ergebnisse demonstrieren die Fähigkeit von ML, grosse und hochkomplexe Systeme wie den LHC mit mehreren Tausend unabhängigen Variablen zu modellieren und zu optimieren. Neben den praktischen Vorteilen, wie der Reduzierung des zeitlichen Aufwands und der Erhöhung des Automatisierungsgrades der optischen Messungen und Korrekturen, wurde in der vorgestellten Studie auch eine methodologische Vorgehensweise für die Integration der ML-Algorithmen in das neue Anwendungsgebiet vorgestellt. Ausserdem, wurden zukünftige Schritte zur Erweiterung der Anwendung von ML in der Strahloptik erarbeitet, was potenziell neue Forschungsfragen und -aufgaben eröffnen kann.

ACKNOWLEDGMENTS

What brought me to the point in time when I find myself writing these sentences for my doctoral dissertation in physics? Surely, the support of my colleagues, friends and family in the past three years made a great contribution to this achievement. However, it is important to realize that not only the time at CERN allowed me to arrive at the end of this exciting PhD journey, but also many years before and a number of important people I met along my way. For me, it all started with receiving a birthday present from my father - a children's book, where a cat called Fortran and sparrow Basic explain how these smart tiny (as tiny as they could be in 1996) machines called computers work.

Long time after this, I met people who I would like to thank in first place - my supervisors at CERN and at the Goethe University Frankfurt, Rogelio Tomás García and Giuliano Franchetti for believing in my ideas, being patient teaching a computer scientist about physics and always having time for fruitful discussions. During the past few years at CERN, I had the pleasure to work with many great colleagues from whom I have learned a lot not only about accelerator physics, but also about professional life in academia. A special thanks goes to the members of Optics Measurements and Corrections Team, who always have been an excellent company, regardless if it was for sleepless nights in the control room, lunch breaks, hiking outings or Friday afternoons in Geneva. Who would have thought, that during four years at CERN one could gain not only a large research expertise, but also obtain such new skills as cycling? I've been very lucky to had the opportunity to join this team and to be able to stay connected even after the end of my PhD time.

I would also like to thank the colleagues from SLAC, DESY and many other laboratories, who are working on establishing machine learning techniques in accelerator physics for many interesting discussions during workshops and conferences. Seeing the interest in this new topic, knowing about success and future plans of other scientists working in the same domain always inspired me and provided additional motivation to progress on my studies.

It is also with great pleasure that I thank my friends in Geneva, Karlsruhe, St. Petersburg and many other places spread all over the world. Thank you for your support, patience and interest in my (sometimes) endless monologues about machine learning and accelerator physics even while being on vacations. In the end, I would like to express my gratitude to my parents and sister for always being sympathetic towards all my passions in work and life and supporting my endless search for changes and new starts. Surely, more of those are to come in the future!

TABLE OF CONTENTS

	Page
1 Introduction	1
1.1 Motivation	1
1.2 Targets and Problem Definition	2
1.3 Relevant Machine Learning Concepts	2
1.3.1 How to define a Machine Learning Task?	2
1.3.2 Supervised and Unsupervised Learning	3
1.3.3 General linear model	4
1.3.4 Generalization and over-fitting	5
1.3.5 Neural Networks	7
1.3.6 Decision Trees and Ensemble Methods	9
1.3.7 Unsupervised techniques: Clustering and Autoencoder	11
1.4 Introduction to Optics Measurements and Corrections at the LHC	13
1.4.1 Fundamentals of Beam Optics	13
1.4.2 Effect of magnetic errors	17
1.4.3 Optics measurements	19
1.4.4 Correction methods	22
1.4.5 Limitations of existing methods	23
1.5 Recent Advances on Machine Learning in Accelerator Physics	24
2 Detection of faulty Beam Position Monitors	29
2.1 General Concept	29
2.2 Traditional techniques	32
2.3 Applying Isolation Forest in BPMs Data Analysis	32
2.4 Results on Operational Data in 2018	33
2.5 Simulating faulty BPM signal identification	36
2.5.1 Model of faulty BPMs	36
2.5.2 Isolation Forest results	37
2.5.3 Exploring SVD settings	38
2.5.4 Comparison to clustering	40

TABLE OF CONTENTS

2.6	Faulty BPMs detection in the presence of local coupling	41
2.7	Detecting unknown failure mode in experimental data	43
2.8	Summary	45
3	Reconstruction of quadrupolar errors from linear optics observables	47
3.1	Introduction	47
3.2	General Concept	48
3.2.1	Preliminary studies	49
3.3	Data generation for prediction of individual magnetic errors	51
3.3.1	Magnet errors as target variables	52
3.3.2	Optics functions as input features	55
3.4	Model selection and training	55
3.5	Results on simulations	57
3.5.1	Optics with $\beta^* = 40$ cm	58
3.5.2	Ballistic optics	61
3.6	Experimental Data	64
3.6.1	Uncorrected optics: LHC commissioning, $\beta^* = 40$ cm	64
3.6.2	Ballistic optics	65
3.7	Summary	67
4	Denoising and reconstruction of Optics Measurements	71
4.1	Noise and magnet errors prediction	72
4.2	Denoising and reconstruction of phase advance measurements using autoencoder	74
4.3	Estimating optics observables using linear regression	76
4.3.1	Reconstruction of normalized dispersion	77
4.3.2	Reconstruction of β -function in Interaction Regions	78
4.4	Summary	79
5	Conclusion	81
5.1	Achieved Results	81
5.2	Discussion	84
5.3	Outlook	86
	Bibliography	89

INTRODUCTION

1.1 Motivation

Machine Learning (ML) as a research field incorporates a broad variety of techniques ranging from traditional statistical methods to deep neural networks capable to surpass human performance in control and optimization tasks. Accelerator problems related to optimization and diagnostics have to deal with non-linear, multi-objective functions which depend on thousands of time-varying machine components and settings. These properties meet the limitations of traditional optimization methods and make this problem a perfect candidate for application of ML-based techniques.

Traditional optimization tools demonstrate successful performance in applications on linear optics corrections and problems with limited amount of optimization targets [1–6]. Bigger challenges emerge when diagnostics of complex non-linear behavior is required and several variables have to be taken into account as final objective. The amount of time and computational power required by traditional methods might become unacceptable for future accelerator facilities. A further limitation of traditional optimization methods is that the objective function or specific rules and thresholds have to be known. ML methods on the other hand are capable of approximating arbitrary functions without requiring explicit rules demonstrating significant benefits compared to traditional data analysis and optimization approaches.

ML is well known for surpassing human performance in some specific tasks such fraud detection, forecasting of market trends and risks, online recommendations, recognition of voice and images and generally in discovering correlations in large scale data sets. Most of these tasks can find analogies in beam control and diagnostics as demonstrated in recent advances of applying machine learning to various accelerator tasks.

1.2 Targets and Problem Definition

The central objective to be addressed in this work is to investigate on ML techniques which can be efficiently incorporated to mitigate existing challenges in optics measurements and corrections at the LHC, build and evaluate them. In the first place, the limitations of current state-of-the-art approach have to be identified in order to define potential ML tasks. A group of applicable algorithms for each of the identified problems should be analyzed, determining the most appropriate technique and possible alternatives. The technical implementation, deployment into existing software infrastructure and adapting available ML algorithms according to the given problem context are also one of the essential parts of this research. In order to conduct the evaluation of developed ML-based solutions, extensive simulation studies and operational tests have to be performed, followed by analysis of obtained results and determining future plans and potential improvements.

The focus not only lies on developing practical solutions, but also on the demonstration on the example of introducing ML techniques into optics measurements and corrections at the LHC, how to choose appropriate ML methods for particular types of tasks and validate them, improving the performance in terms of problems in the domain of accelerator physics and providing understanding of applied solutions using ML. For that purpose, a brief summary on recommendations for suitability between accelerator related problems and corresponding ML methods will be provided along with an outlook into future potential developments.

1.3 Relevant Machine Learning Concepts

Machine Learning can be seen as a natural development of the intersection between computer science and statistics. *"ML techniques aim to build computer programs and algorithms that automatically improve with experience by learning from examples with respect to some class of task and performance measure, without being explicitly programmed"* - this definition by T. Mitchell illustrates the key characteristics of ML, which distinguish this concept from traditional programming, namely the ability to automatically achieve performance improvement based in provided data. The following brief introduction to ML paradigms and methods is mainly based on [7] and [8] where detailed discussion on the subject of ML and statistical learning can be found. This chapter focuses exclusively on particular algorithms applied to the new methods developed in this thesis.

1.3.1 How to define a Machine Learning Task?

In the context of intelligent techniques, *learning* is referred to as a process of performance improvement. In general, a ML model can improve its performance by adjusting the free parameters of the chosen approximation function with respect to the given data. To be noted is

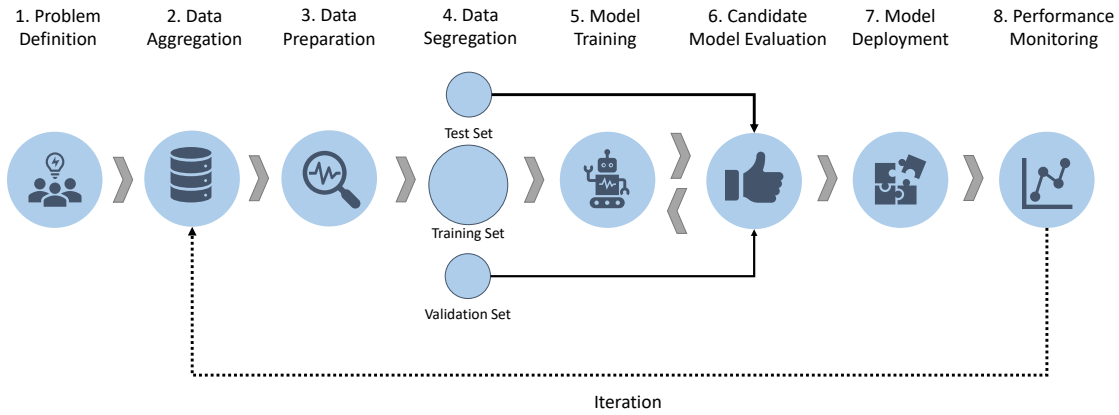


Figure 1.1: Illustration of the development process of a ML based solution including the fundamental steps from the problem definition to the final monitoring of operational model.

that the learned model will always be an approximation, which nevertheless should be sufficiently reliable for a specific task. The requirements on the model performance in terms of accuracy, precision and reliability depend on a particular task and application domain. The performance metric, the type of experience to be obtained via learning and the specific ML approach and algorithm are determined by the task with respect to which the learning is conducted. These three parameters (task, measure of performance and source of experience) are the fundamentals of a well-defined ML problem. In practice, it is reflected in the selection of appropriate ML method and function approximation algorithm, definition of the loss function and its acceptable values, as well as the preparation of data. The latter usually appears to be not trivial and can require dedicated techniques and algorithms e.g. for feature extraction or correlation analysis. Figure 1.1 provides an overview on the basic technical aspects of developing a ML-based solution and incorporating it into existing infrastructure.

1.3.2 Supervised and Unsupervised Learning

Depending on the problem definition and availability and structure of learning examples, different approaches are preferred. *Unsupervised learning* algorithms solve tasks where only input data is available, aiming to find new patterns or to extract relevant information from the given unstructured data. Unsupervised learning approach is suitable to solve the problems such anomaly detection, signal denoising, dimensionality reduction and feature extraction.

ML algorithms that learn from the sample pairs consisting of input variables and corresponding target outputs are called *supervised learning* algorithms. Supervision is provided in the form of determined outputs for each given example. During the training, predictions are computed

as a function of the incoming input and are then compared to the actual corresponding output. The target is to minimize the difference between predictions made during the training and the true output by updating adjustable parameters of a chosen model (approximation function). The difference is defined as *loss function* (e.g. mean absolute error) which is minimized using an optimization algorithms suitable for a particular problem to be solved with the trained ML model.

The complexity of the model depends strongly on the specific task and can range from univariate linear regression to deep neural networks with a complex customized architecture.

Depending on the desired prediction target - category or a quantity, the problem can be defined either as *classification* or *regression* task.

1.3.3 General linear model

Linear models compute the estimates of the output values using a linear function of the input features [9]. For regression, the general estimation for the predicted output $\hat{y}$ of a linear model is formulated as follows:

$$(1.1) \quad \hat{y} = w_0 + \sum_{i=1}^j w_i x_i$$

where $\vec{x}$ denotes the input vector containing j features, $\vec{w}$ is the vector of learned coefficients of the model with w_0 being the bias. For a simplified illustration of the linear regression approach, a single output is modeled, so $\hat{y}$ is a scalar. In general, $\hat{y}$ can be a K -vector, in which case the parameters w are represented by a $j \times K$ matrix.

The training of linear regression model is usually performed using *Least Squares* method. Given a set of training data with N samples and y being the true output of a single sample, the parameters of the linear model are updated with each incoming sample aiming to minimize the residual sum of squares

$$(1.2) \quad RSS(\vec{w}) = \sum_{m=1}^N (y_m - \hat{y}_m)^2$$

Ideally, after performing the training, the resulting coefficients are general enough to produce sufficiently accurate predictions for all provided data samples.

1.3.3.1 Polynomial Regression Tasks

Some learning tasks require nonlinear predictors, such a fitting of one dimensional polynomial function of degree n , that is,

$$(1.3) \quad p(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

where $(a_0 \dots a_n)$ is a vector of coefficients of size $n + 1$. One way to train the model for prediction of such problems is to reduce the problem to linear regression [10]. To define a polynomial regression

task as linear regression problem, a mapping such that $\psi(x) = (1, x, x^2, \dots, x^n)$ is introduced and the problem reduces to

$$(1.4) \quad p(\psi(x)) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n = \langle \vec{a}, \psi(x) \rangle$$

so that the optimal vector $\vec{a}$ can be found by minimizing RSS as in the case of linear regression or another linear optimization function.

1.3.4 Generalization and over-fitting

In supervised learning, the target is to build a model which is able to make accurate predictions on new, unseen data based on the experience provided by the training data with known target outputs. The ability to predict on unexplored data is called *generalization*, which is the most important property of the model to be evaluated [11, 12]. The evaluation of the predictive power of the selected model is usually performed on an independent test data set. In most cases, training and test data sets have enough common properties and relations in the data, such that the model can predict accurately also on the test set. However, in case of high-complex models, very high accuracy can be achieved on the training data, although a possible consequence is a very poor prediction on the test data. Too complex models (e.g. neural networks with several hidden layers) tend to fit to the individual data points of the given training data too closely and hence perform insufficiently on new data. Building a model that is too complex for the patterns to be decoded from the given data is called *over-fitting*. Figure 1.2 demonstrates the importance of indicating the appropriate complexity of the model with respect to the data, which is representative for the case of future application on unseen input.

One of the natural ways to control the over-fitting problem is to provide a set of validation data to the algorithm in addition to the training data. The training set is used to drive the learning procedure. After the training, the error is then evaluated with respect to the validation set. The performance of the model measured on the training set should be nearly equal to the one measured on the validation set. If the model performance on the training set is significantly higher, the model is over-fitting and further adjustments on the model complexity are needed.

On the other hand, if the model is too simple, it will not be able to produce accurate prediction on the training data and to learn hidden information which explain the input-output correlations and its variability. The complexity of the model corresponds directly to the variation of sample pairs contained in the training data. Hence, it is crucial to control the size and the variation of the data in order to build an appropriate model.

1.3.4.1 Regularized Regression

A possible approach to resolve the over-fitting problem is to apply regularization, i.e. imposing a penalty on the update of free parameters of the model during supervised training. The model parameters are updated during the training attempting to fit every incoming input-output

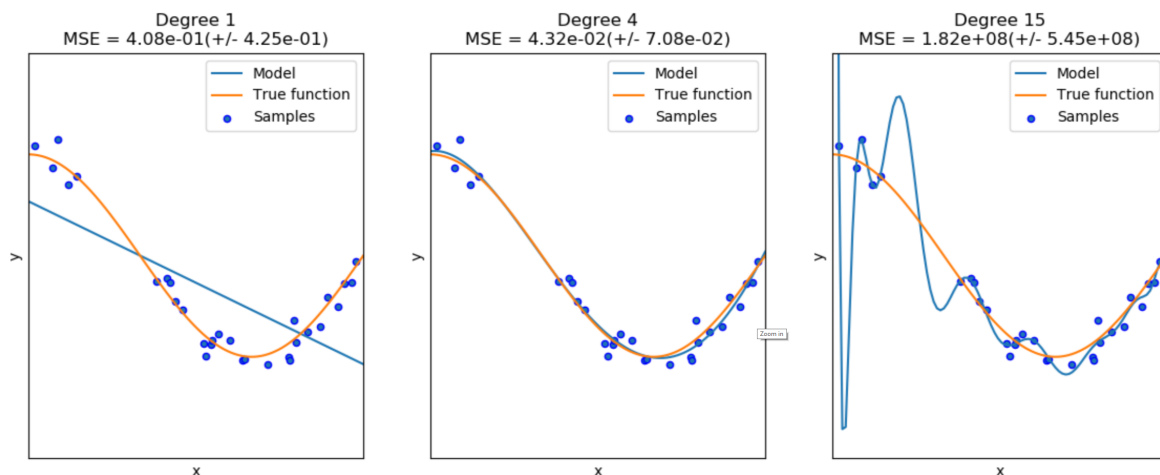


Figure 1.2: Complexity of the model directly affects the prediction accuracy and generalization ability of the model. While simple models tend to deviate from data too much, introducing high bias and low variance, complex models tend to over-fit becoming disadvantageously sensitive to the variations in the data. *Bias-Variance trade-off* needs to be identified in order to create a reliable model.

pair. The penalty controls the change of these parameters during the training such that the model becomes more stable against variations in the data. This technique is known as *L2 regularization* or in the context of supervised regression - *Ridge regression* [13]. Ridge regression model minimizes the residual sum of squares between the true targets in the training data, and the targets predicted by the linear approximation. The tuning parameter α is responsible for the coefficients of the linear model and aims to mitigate the negative impact of collinearity, i.e. linear dependencies between input variables in a regression model. In presence of such dependencies, the update of coefficients during the training may change inappropriately in response to small changes in the model or the data. With growing value of α the "shrinkage" of weights values increases and this the coefficients become more robust to collinearity in the provided training data. The regression task of minimizing the loss between model output, computed using the coefficients and given input, and true target values is then formulated as

$$(1.5) \quad \min_w ||w\vec{x} - \vec{y}||_2^2 + \alpha ||w||_2^2$$

where w is the matrix containing the weights of regression model, $\vec{x}$ is the input data vector and $\vec{y}$ the vector of targets to be predicted by the model.

Alternatively, over-fitting can be mitigated by reducing the number of input features keeping only uncorrelated features for the function approximation. *Lasso regression* [14] uses an *L1* norm in order to push some of the regression coefficients to zero, which not only improves the model in terms of the variance but also allows automatic feature selection.

The tuning of additional model parameters such as the Ridge penalty or the number of input features to be kept as in Lasso regression is based on *cross-validation*. Generally speaking,

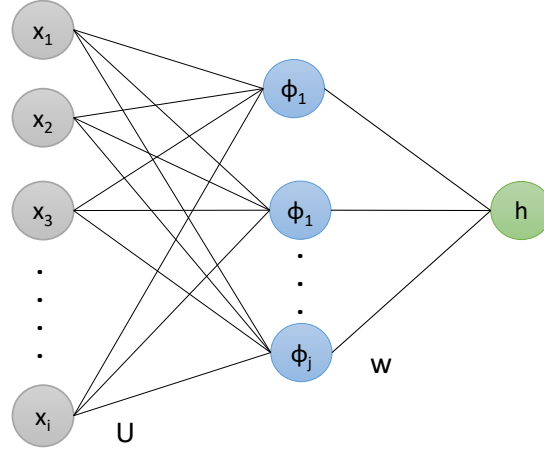


Figure 1.3: Simplified sketch of a Neural Network with hidden layer composed of neurons with a nonlinear activation function ϕ .

cross-validation works by dividing the training data randomly into a several equal parts, e.g. ten subsets. The regression algorithm is then fit to nine-tenth of the data samples, and the loss is computed on the remaining one-tenth. This procedure is repeated for each one-tenth subset of the data and all ten estimates of prediction error are averaged. Since the tuning of regularization parameters is part of training process, cross-validation is applied to the training set, meaning that the data has been previously separated into training and test sets. Test set is exclusively used to judge the performance of the selected model on new unseen data.

1.3.5 Neural Networks

Artificial Neural Networks (NN) are well suited for learning tasks, where data is represented by noisy, complex signals and the target output function may consist of several parameters. A basic NN consists of a single processing unit (*neuron*), that takes a linear combination of the inputs (*activation*), where the weights in linear combination are adaptive parameters, and an additional *activation function* to introduce the nonlinearity.

Considering a simple one-layer NN illustrated in Figure 1.3, the activation function ϕ_i for each neuron in the hidden layer is formulated as

$$(1.6) \quad \phi_j = \sigma(a_j)$$

with σ being an arbitrary differentiable nonlinear activation function such as logistic sigmoid defined as

$$(1.7) \quad \sigma(a_j) = \frac{1}{1 + \exp(-a_j)}$$

and a being the activation linearly combining M input features $\vec{x}$ and the corresponding element u of the adjustable parameter matrix

$$(1.8) \quad a_j = \sum_{i=1}^M x_i u_{ij}$$

The full model of the single output neural network becomes

$$(1.9) \quad h(\vec{x}\mathbf{U}\vec{w}) = \sum_{j=1}^N w_j \phi_j \left(\sum_{i=1}^M x_i u_{ij} \right)$$

with $\mathbf{U}$ being the matrix of free model parameter between the input and hidden layer and $\vec{w}$ the weights connecting the hidden layer to the output h .

For more complex practical problems, NNs are composed of several interconnected *hidden layers* with multiple neurons stacked. NNs can be used for both regression and classification problems. In case of classification the output can be either a class label or a probability of an item belonging to a class. The learning of NN is performed using *backpropagation* algorithm [12] on a set of examples. For each example the training algorithm computes the derivatives of the output function of the network. The obtained gradients with respect to all weights are then used to adjust the weights in order to achieve a better fit to the target output, which is then updated with every incoming input-output pair such that sufficient level of generalization can be achieved.

The optimization technique called *gradient descent* [15] founds on searching for the global minimum of a function $f(x)$ by moving x in small steps with the opposite sign of the derivative. Optimization algorithm may fail to find a global minimum in presence of multiple local minima or plateaus. The problem arises especially in case of multidimensional functions. However, learning algorithms often accept local solutions in case they correspond to significantly low values of loss function, even if they are not globally optimal.

In order to minimize $f(x)$, the gradient descent finds the direction in which f decreases fastest using directional derivative. The function $f(x)$ is then decreasing by moving in the direction of the negative gradient. Gradient descent proposes a new point

$$(1.10) \quad x_{i+1} = x_i - \epsilon \nabla_x f(x)$$

where ϵ is the *learning rate* determining the size of the step and $\nabla_x f(x)$ is the vector with partial derivatives of $f(x)$. In context of learning, the function $f(x)$ is the loss function that can be expressed by the MSE as described above. During the training on a data set, gradient descent computes the average per complete training set in each step, which makes the solution computational expensive and time demanding.

In *stochastic gradient descent* algorithm, instead of averaging the gradient of the loss function over the complete training set, the loss function is computed for particular training sample. Thus, stochastic gradient descent approximates the solution and reduces the required computational resources and time significantly. Stochastic gradient descent is widely used in modern learning

techniques. Several extensions of the algorithm are introduced during the past decades such *momentum* method [16], *AdaGrad* [17] and *Adam* [18].

Neural Networks with many hidden layers called *deep neural networks* are able to use fewer neurons per layer and have a better generalization ability, however the optimization of the structure and training of these networks is not trivial. There are no strict rules for building NN architecture (number of neurons, layers, initial weights) as it usually heavily depends on a particular problem. However, techniques to adjust the architecture parameters exist. A detailed overview on various NN architectures and training methods and their suitability for different applications can be found in [19].

1.3.6 Decision Trees and Ensemble Methods

Decision tree learning is a method for approximating discrete-valued target functions, which are represented by decision trees. Considering the case of classification, decision trees sort down the input instances from the root to leaf nodes. Usually, the splitting is based on one of the input parameters or a specified set of splitting criteria. Each leaf corresponds to one class representing the most appropriate class label.

For regression problems the leaf nodes correspond to an approximation of target values [20]. The decision function fitting the training data is constructed by performing the splits on a selected feature, dividing the data space into regions. By separating the multi-dimensional data space into regions and observing the average value of the target variable in these regions, one can relate the values of the independent input variables to the estimate of the dependent output variable. As in other regression algorithms, the splits can be optimized with respect to the least square minimization of the error between estimated and true output values. The crucial parameter of a decision tree based regression model is the maximal depth of the tree which is directly related to the complexity of the model as shown in Fig. 1.4. The deeper the tree, the more complex the decision rules and the better the fit to the training data and hence, the depth needs to be controlled to prevent possible over-fitting.

Compared to NNs, decision trees are simpler to interpret and to understand its way of obtaining the final results and the underlying process, e.g through the feature importance analysis. Feature importance analysis helps to understand the contribution of each input parameter to the decision during the training process by computing the decrease in the *impurity function* after performing a data split based on a chosen input parameter. While constructing a decision tree, impurity function plays the same role like loss function in other supervised learning techniques and therefore, can be also defined e.g. by MSE for regression problems or cross entropy for classification tasks. This ability of decision trees to evaluate the importance of input parameter is a significant advantage of these algorithms. Knowing the importance of the features allows to reduce the model complexity and to simplify the data preprocessing steps without significant accuracy loss.

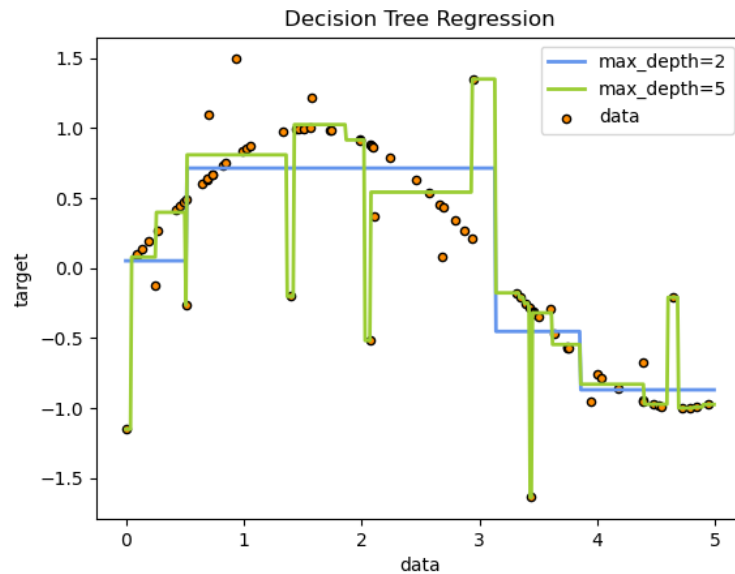


Figure 1.4: Illustration of decision tree based regression using different parameters for the depth of the tree [21]. The resulting decision function and the model performance on unseen data highly depend on the depth of the tree, i.e. number of splits performed on the data during the training.

Using a single tree, a model might not be able to generalize and perform poorly on unexplored sample. One possible solution to overcome this problem is to build ensembles of trees [22]. By training several slightly different models and taking the average prediction, the variance of the model can be reduced. This approach can be applied to an arbitrary supervised learning technique to improve the predictive performance following the general principle of constructing linear combination of some function approximation technique instead of using a single estimator.

Bagging standing for bootstrap aggregating is an ensemble method which aims to reduce the variance of the model prediction [20]. It is based on constructing several independent estimators trained on so called bootstrap samples obtained by randomly drawing with replacement a subsets of available data. After training individual learners, their predictions are averaged in order to obtain the final output. A widely used algorithm based on this approach is e.g. Random Forest [23] which relies on training multiple decision trees on bootstrap samples. The result of the introduced randomness is a minor increase of the ensemble bias (with respect to the bias of a single non-randomized tree). On the other hand, due to the averaging the total ensemble of trees also experiences a decrease of variance, which is greater than a value needed just to compensate the increase in introduced bias. Therefore, a better overall model can be achieved.

The main parameters to adjust when using Random Forest are number of trees and maximum number of features to consider when looking for the best split. The number of trees affects the quality of the model, but also the computation time. The results will not get significantly better beyond a critical number. Empirical good values for maximum number of features is the number

of input parameters in dataset for regression tasks. For classification, the maximum number of features should be the square root of input size. Most of modern implementations of this algorithm allow parallel construction of the trees and parallel computation of the prediction [21].

Boosting algorithms [24] also incorporate the idea of combining individual estimators in order to improve the overall prediction. However, unlike bagging, boosting is based on sequential training of weak learners: each model in the sequence is trained giving more importance to observations in the trained data that were poorly predicted by the previous models in the sequence.

1.3.7 Unsupervised techniques: Clustering and Autoencoder

The application of above mentioned learning techniques based on neural networks and decision trees can be designed in supervised manner as well as unsupervised learning based solutions. In the following, some additional unsupervised learning methods will be presented.

Clustering

Cluster analysis includes methods of grouping or separating data objects into clusters, such that dissimilarity between the objects within each cluster is smaller than between the objects assigned to different clusters [25, 26]. Data clusters can be considered as a summarized representation of the data, such that group labels can describe patterns or similarities and differences in the data. Moreover, clustering can be used for prediction, such that classification of unseen data is performed based on knowledge about the properties of present data and by evaluating their similarity to the incoming data sample. The significant benefit of cluster analysis is the possibility to learn hidden patterns in the data without providing true output values.

The simplest and the most commonly used clustering algorithm is k-means [27], which is based on centroid search. Another kind of clustering algorithms are the density-based algorithms such DBSCAN [28], that views clusters as areas of high density separated by areas of low density, instead of looking for the centroids. Density based approaches are widely applied to the problems where distance based similarity measures are not able to adequately separate the data, e.g. due to the selected features and large variance of their values such that a generally valid threshold for a distance measure cannot be defined. Local Outlier Factor [29] as well as Expectation Maximization [30] are further density-based methods which are appropriate for the application to this type of problems.

Decision tree based methods also can be applied for cluster analysis using the data splits based on different features. Most of cluster analysis techniques allow to build clusters in a multidimensional space.

Apart from classification and pattern recognition, cluster analysis can be used as denoising and cleaning method looking for abnormalities in the signal. Moreover, building clusters combin-

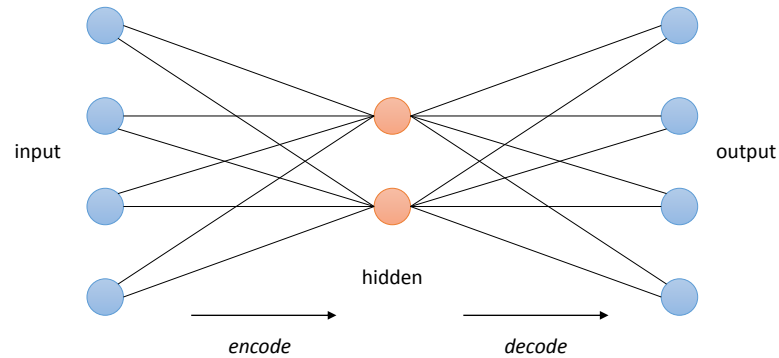


Figure 1.5: Schematic illustration of autoencoder neural network. For simplification, the multiple hidden layers are represented by a single layer of two neurons. The hidden layers allow to map the original input information in a lower-dimensional space by capturing important input features.

ing a large set of different observables can simplify the data visualization and manual analysis, such elimination of outliers in the measurements and detection of anomalies.

Autoencoder Neural Networks

Autoencoder is a specific type of a NN that in its original application attempts to reproduce its input as its output. The network consists of two parts: an encoder function $h = f(x)$ describing a hidden layer h and a decoder that produces a reconstruction $r = g(h)$ as illustrated in Figure 1.5. Autoencoders are usually restricted to copy only input that approximates the training data. Therefore, the model is forced to prioritize which aspects of the input should be reproduced and to learn useful properties of the data in this way [19].

In modern understanding autoencoders generalize the idea of encoder and decoder beyond deterministic functions to stochastic mappings $p_{\text{encoder}}(h|x)$ and $p_{\text{decoder}}(x|h)$. Initially, autoencoder concept was introduced in the tasks of dimensionality reduction and feature learning. Recently, autoencoders are also widely used in generative modeling. As autoencoder is considered as a special case of a feedforward network, they can use the same techniques for training, which are usually gradient descent methods for gradients computed by backpropagation. Apart from typical techniques, autoencoders may learn using *recirculation* [31]. Recirculation is a technique based on comparison between activations of the network on the original input and the activations in the reconstructed input.

Regarding the problem of obtaining useful features from data, an autoencoder built with hidden layer of smaller dimension than the input can be used. This type of autoencoders is called *undercomplete autoencoders*. Since the representation is undercomplete, the autoencoder

is forced to extract most important features of the training data. The learning process in this case is described as minimizing a loss function

$$(1.11) \quad L(x, g(f(x)))$$

penalizing $g(f(x))$ for being dissimilar from x .

Denoising autoencoder is an autoencoder that receives a corrupted data point as input and is trained to predict the original, uncorrupted data point as its output [19]. Instead of adding a penalty to the loss function an autoencoder can be trained by changing the reconstruction error term. A denoising encoder minimizes the function $L(x, g(f(\tilde{x})))$, where $\tilde{x}$ is a copy of x that has been corrupted by some kind of noise. Denoising encoder eliminates this corruption instead of copying the original input to the output.

1.4 Introduction to Optics Measurements and Corrections at the LHC

The Large Hadron Collider (LHC) is CERN's largest facility and the most powerful accelerator in the world with the design energy of proton beams collision up to 14 TeV. The beams are pre-accelerated by the LHC injector chain which consists of several accelerators as shown in Figure 1.7. The beam sizes at the Interaction Point (IP) have to be squeezed in order to provide maximum number of collisions to the experiments. After the squeeze the beams are colliding at 4 different IPs. Bunches cross in average about 30 million times per second, so the LHC generates about 1 billion particle collisions per second.

The 4 main detectors of the LHC are indicated in Figure 1.7: ATLAS [32] and CMS [33] aim to explore new heavy particles, LHC beauty (LHCb) studies [34] focus on the antisymmetries between matter and anti-matter and ALICE [35] is specialized in heavy-ions collisions to study the properties of quark-gluon plasma. The different experiments require different rates of collisions, therefore the beam is squeezed more for the general purpose experiments (ATLAS and CMS) and less for LHCb and ALICE. The requested beam settings also rely upon different modes of operation. This flexibility given by wide selection of operation modes extends the physics potential of the LHC but results in higher complexity. This complexity implies the requirement of various optics settings which need to be tested and corrected to ensure reliable beam control.

1.4.1 Fundamentals of Beam Optics

In this section the basic principles and main aspects of transverse beam dynamics are discussed. For a broader and more detailed introduction into accelerator physics the references [36–38] can be addressed, on which the following brief overview on fundamental beam optics concepts is based on.

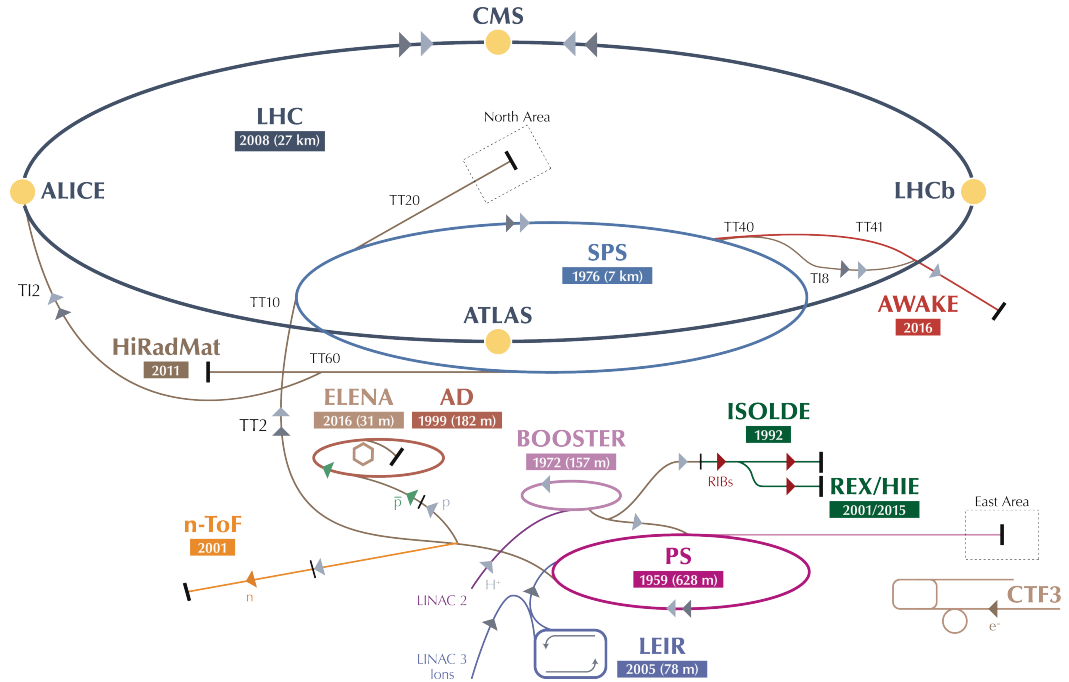


Figure 1.6: The schematic of CERN accelerator complex presenting the acceleration chain and facilities dedicated to different experiments and studies.

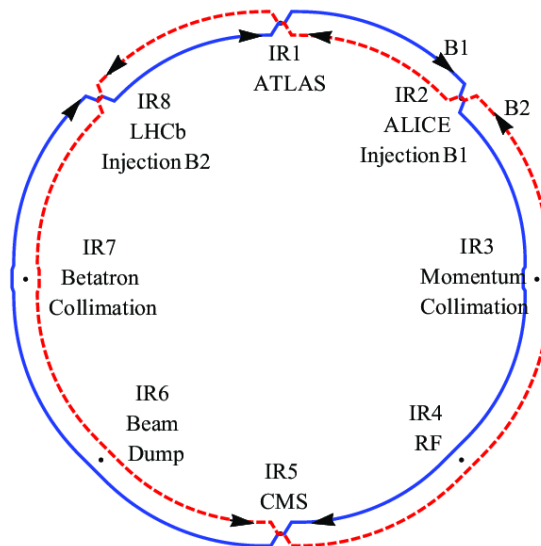


Figure 1.7: Layout of the LHC illustrating the Interaction Regions (IR) and their specific purposes, including the experiments with the collisions of beam 1 (blue) and beam 2 (red).

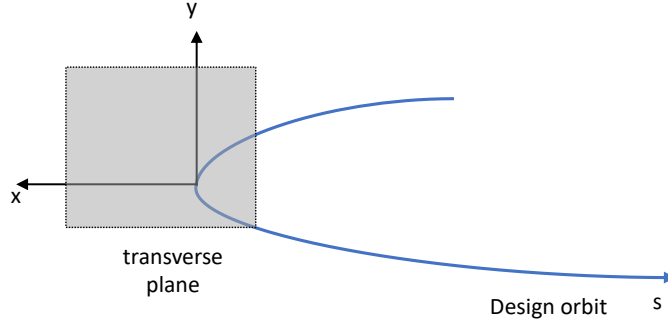


Figure 1.8: Coordinate system used to describe the particle motion in accelerators.

The control of the beam by the means of magnetic fields along the storage ring is the central aspect of beam optics. While dipole magnets can be used to bend the beams direction, focusing is done by quadrupole magnets. Since quadrupoles cannot focus in both planes at the same time, a special lattice design has to be used in order to focus and control the beam as required by the operation. Usually, the quadrupoles of alternating polarity are used in order to focus the beam transversely. The horizontal motion of a particle experiencing a quadrupolar field is then described as

$$(1.12) \quad \frac{d^2x}{ds^2} = -k(s)x$$

where x is the horizontal coordinate of the particle with respect to the design orbit s as defined by the coordinate system shown in Figure 1.8 and $k(s)$ is the focusing strength of a magnetic field. Magnetic field B generated by a quadrupole increases linearly proportional to the distance from the magnet center

$$(1.13) \quad \begin{aligned} B_y &= -gx \\ B_x &= -gy \end{aligned}$$

where x and y refer to the horizontal and vertical coordinates and g being the gradient of the magnetic field. Normalizing the quadrupolar field to the magnetic rigidity $B\rho$, the strength k is defined as

$$(1.14) \quad k = \frac{g}{(B\rho)}$$

The simplest possible strong focusing lattice is the *FODO cell* which represents a unit cell consisting of a pair of quadrupoles each followed by a drift space. The focusing is achieved by combining the focusing and defocussing quadrupoles which can be considered as lenses compared to light optics [36]. A beam line built on repeated FODO cells is shown in Figure 1.9. Replacing the drift space with a dipole magnet to bend the beam, one obtains a FBDB cell. Another possible lattice design is a triplet, which contains three quadrupoles with the polarity of the center

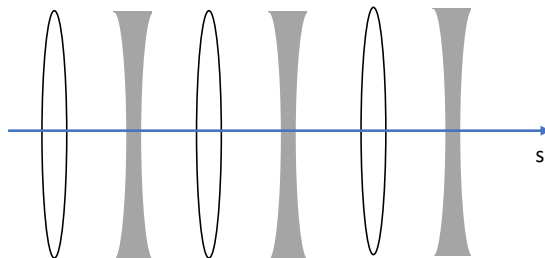


Figure 1.9: Beam line in a FODO cell consisting of focusing and defocussing magnets with drift spaces between the elements.

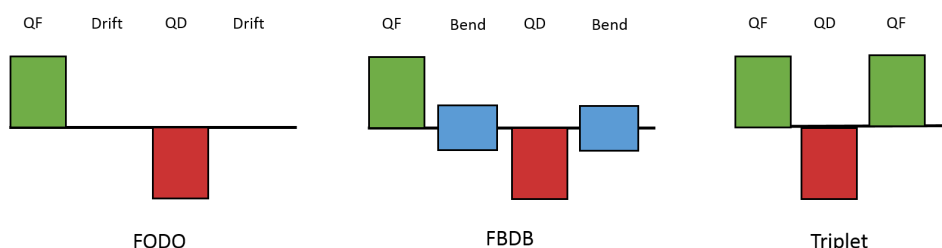


Figure 1.10: Schematics of different magnet configurations of a lattice design in particle accelerators.

quadrupole opposite to the two quadrupoles on the outside of a cell. Figure 1.10 shows the described lattice configurations. In a high energy circular collider, FBDB cells are frequently used for transporting charged particle beams in the arc sections, while triplets are usually installed in the interaction region to strongly focus the charged particle beam into small beam sizes to facilitate collisions as it is done in the LHC.

Assuming the periodicity in a circular accelerator, the particle motion equation 1.12 can be written as so called *Hill's equation* using the amplitude A and phase advance ϕ varying with s

$$(1.15) \quad x(s) = A(s) \cos(\phi(s) + \phi_0)$$

The maximum amplitude of the particle oscillation around the design orbit is described by the emittance ϵ and β -function

$$(1.16) \quad A(s) = \sqrt{\beta(s)\epsilon}$$

The Betatron function β is one of the main parameters to define the beam optics. It defines the beam dimension together with the beam emittance which measures the envelope of all particle trajectories at a given position s along the nominal orbit. The β -function is influenced by the all focusing fields and reflects the periodicity of the lattice structure. In case of large β the beam is less focused, hence occupying a larger transverse space compared to locations with small β .

The betatron tune Q defines the number the of betatron oscillations in horizontal and vertical plane performed by the particle beam in one revolution and is related to the β -function as follows:

$$(1.17) \quad Q_{x,y} = \frac{1}{2\pi} \oint \frac{ds}{\beta_{x,y}(s)} .$$

The tune value affects the dynamics of the beam motion and at certain values it can lead to beam instabilities. As in any oscillating system, the resonance conditions have to be avoided by keeping the frequency of the transverse motion not equal to (or an integer multiple of) the revolution frequency. It is desired to keep the tune away from fractional values with small denominators such as $\frac{1}{2}$, $\frac{1}{3}$.

Dispersion is caused by the fact that the particles with different energies are bent differently. A particle with higher energy will be bent less compared to the one with lower energy. The dispersion is described as the deviation from the reference orbit $\Delta x(s)$ with respect to relative momentum deviation $\Delta p/p$:

$$(1.18) \quad D(s) = \frac{\Delta x(s)}{\Delta p/p}$$

The deflection of a particle by a quadrupole magnet in the beam line depends on the trajectory of the particle as it enters the quadrupole, the length of the quadrupole, the magnetic field in the quadrupole, and on the energy of the particle. The dependence of the particle motion on the energy of the particle will lead in particular to the variation of the tune with energy [36]. This variation is known as chromaticity and is illustrated in Figure 1.11, described as

$$(1.19) \quad Q' = \frac{\Delta Q}{\Delta p/p}$$

where ΔQ is the tune change and $\Delta p/p$ is the relative momentum deviation. The natural chromaticity, i.e. the chromaticity arising from the linear magnet lattice is always negative, which is to be expected since focusing is less effective for higher energy particles ($\Delta p/p > 0$).

1.4.2 Effect of magnetic errors

Imperfections in the real accelerator appear in case of uncertainties or errors in magnets, magnets misalignment, power converter regulations, beam momentum, or Radio Frequency (RF) fluctuations. The field quality and alignment precision of the magnets are crucial targets to be achieved in order to meet the performance specifications of an accelerator, making the identification and corrections of significant error sources to a major task during machine commissioning and operation. In the following a brief overview about the imperfections caused by the magnetic errors is given.

Closed orbit distortion

A magnetic dipole field causes the change in particle momentum and the closed orbit is no longer the reference trajectory. The horizontal dipole field located at a single point s in the storage ring

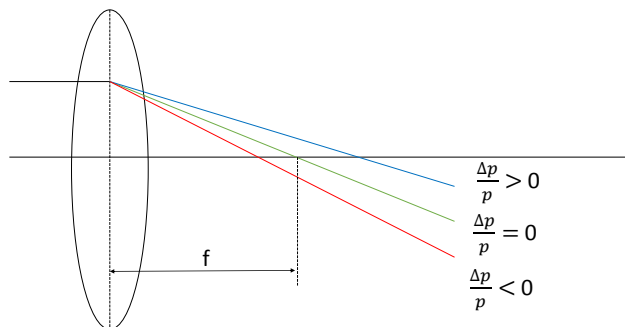


Figure 1.11: Higher energy particles are focused less than a particle with nominal energy and lower energy particles are over-focused.

causes a change Δp_y in the particle momentum p_y :

$$(1.20) \quad \Delta p_y = \frac{q}{P_0} \int B_x ds,$$

where q is the charge of the particle passing the location of the dipole, P_0 is the reference momentum and $\int B_x ds$ is the integrated strength of the horizontal dipole field along the reference trajectory. To be noticed is that the closed orbit at any point on a storage ring depends on the β -function at that point, as well as on the β -function at the location of the dipole field error: if the β -function is large, then a small deflection can lead to a large closed orbit distortion [36]. The following equation describes the closed orbit resulting from distributed dipole perturbations:

$$(1.21) \quad CO(s) = \frac{\sqrt{\beta(s)}}{2 \sin \pi Q} \sum_i \sqrt{\beta_i} \theta_i \cos(\pi Q - |\phi(s) - \phi_i|)$$

where Q is the tune, $\beta(s)$ the β -function at the location s , θ_i is the dipolar kick angle and $|\phi(s) - \phi_i|$ phase advance between the point s and the location of the dipole i .

Quadrupole alignment errors are another significant source of closed orbit distortion. In the case of a vertically misaligned quadrupole, a particle moving along the reference trajectory sees a horizontal magnetic field causing a vertical deflection to the particle.

Focusing errors

Focusing errors in a storage ring can easily arise from variations in the current flowing through the coils in the quadrupoles. Since the β -function is proportional to the magnitude of perturbation and the amplitude of the β -function itself at the location of perturbation, a quadrupole strength change from its nominal value Δk_i at the i th error source leads to the following changes in the β -function up to first order on Δk_i :

$$(1.22) \quad \frac{\Delta \beta(s)}{\beta} \approx \sum_i \frac{\Delta k_i \beta_i}{2 \sin(2\pi Q)} \cos(2\pi Q - 2|\phi(s) - \phi_i|)$$

The beating of the betatron function (β -beating) describes the relative deviation of the measured betatron function with respect to the nominal design function,

$$(1.23) \quad \frac{\Delta\beta}{\beta} = \frac{\beta_{meas} - \beta_{model}}{\beta_{model}}$$

The tune shift in horizontal (ΔQ_x) and vertical (ΔQ_y) planes is as well proportional to the gradient error and to the β -function at its location. Up to first order, this relation is expressed as

$$(1.24) \quad \Delta Q_x \approx \sum_i \frac{1}{4\pi} \beta_{ix} \Delta k_i \quad \Delta Q_y \approx \sum_i -\frac{1}{4\pi} \beta_{iy} \Delta k_i$$

Thus, the change in betatron tune depends on the change in focusing strength and on the value on the β -function at the location on the error.

Coupling

Skew quadrupole field errors and detector solenoids generate betatron coupling between the horizontal and vertical planes of motion. A particle passing through a skew quadrupole experiences a horizontal deflection proportional to its vertical position and vice-versa. It is important to control the coupling in an otherwise uncoupled storage ring such as the LHC, since its appearance can push tunes into resonances which has to be avoided to ensure safe beam operation [39, 40].

The good control of the optics in the LHC was a large contribution to the success in exploration of a new energy scale and important discoveries. In the next section the methods to measure and correct the optics of the LHC are discussed.

1.4.3 Optics measurements

Successful optics control requires appropriate techniques to obtain precise and reliable measurements of optics observables. In this section, the traditional procedure of optics measurements and corrections at the LHC is presented [41–43], followed by the discussion on the limitations of current methods which are to be mitigated by the introduction of ML concepts developed as part of this thesis.

Data acquisition

In order to obtain the optics observables needed to compute the corrections of the magnetic errors, an excitation of the beam is required. In the LHC the beams are excited with the help of so-called kicker magnets, usually an alternating current (AC) dipole is used to initiate the optics measurements process. The AC-dipole is a fast oscillating magnet, which can be adiabatically turned on and off. In this way it creates coherent oscillations of the beam particles without affecting the transverse emittance [44].

During the kick a distortion of the transverse beam center position is introduced, which allows to measure the betatron oscillations. The amplitude of the kick needs to be sufficiently high in

order to produce the oscillations which can be recorded by Beam Position Monitors (BPMs), of which the LHC is equipped with about 500 per plane and per beam [45]. The BPMs record the beam position every turn, which is used to reconstruct the optical functions by performing a spectral analysis on turn-by-turn data [46, 47]. Prior to harmonic analysis, the turn-by-turn data is processed by a special singular value decomposition (SVD)-based technique for noise reduction by filtering uncorrelated signals [48]. This technique along with other traditional cleaning tools are discussed in more details as part of the introduction to the presentation of newly developed ML-based cleaning technique for the BPM data in Chapter 2.

Optics analysis

The phase of measured betatron oscillation can be inferred from a harmonic analysis of the turn-by-turn data, i.e. measurements of the position of transverse beam center for several consecutive turns, recorded using the BPMs system. The measurement of the beam position at turn N and position s is related with the amplitude A and phase ϕ of the oscillations introduced by the AC-dipole as follows:

$$(1.25) \quad x(N, s) = \sqrt{A} \cos(2\pi Q_x N + \phi_x(s)) + \Delta x(s)$$

where $\Delta x(s)$ is the closed orbit offset at this position. Using the three BPM method [49] through the phase advance between 3 consecutive BPMs the β in the position of the BPM i can be reconstructed by using

$$(1.26) \quad \beta(s_i) = \frac{\cot\phi_{ij} - \cot\phi_{ik}}{\cot\phi_{ij}^m - \cot\phi_{ik}^m} \beta_m(s_i)$$

where the ϕ_{ij} and ϕ_{ik} are the measured phase advances between BPM_i and BPM_j and BPM_i and BPM_k respectively; the terms with the superscript m refer to the model values of corresponding observables. The potential issue of applying this method arises from the fact that the accuracy depends not only on the knowledge of the optics model and the precision of the measured phase but also on the value of the phase advance between the BPMs.

The N-BPM method [50] allows to use more BPM combinations from a larger range of BPMs to increase the amount of information that is used in the measurement of β -function and hence, improving the reliability of the β -function calculation. A number of N BPMs is chosen close to the probed BPM as shown in Figure 1.12. The best estimate of the measured β -function out of M combinations of three BPMs is to be found by performing a least squares minimization of the function

$$(1.27) \quad S(\beta) = \sum_{i=1}^m \sum_{j=1}^m (\beta_i - \beta) V_{ij}^{-1} (\beta_j - \beta),$$

where β_i are the β -functions obtained from different BPM combinations and V_{ij} are the elements of the covariance matrix for the different β_i . The measured β -function at the probed BPM

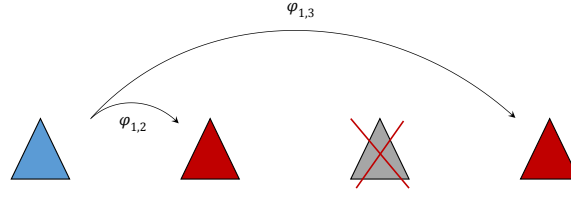


Figure 1.12: The blue BPM is probed, the gray BPM should be skipped and the two red BPMs are included into the β -function computation. In this way, the N-BPM method allows to use a bigger amount of data to increase statistics and to avoid unsuitable phase advances.

position is then a weighted average of the m β -functions. Recently, an analytical version of this method has been developed improving the measurement accuracy, being applicable to beam optics exhibiting large perturbations from various error sources [51].

Alternatively, the amplitude of the transverse motion of the beam excited by a kicker magnet can be utilized in order to obtain the β -function. The amplitude A of the recorded transverse beam motion is inferred by the harmonic analysis of turn-by-turn data and is proportionally related to $\sqrt{\beta}$ and the $\sqrt{2J}$:

$$(1.28) \quad A = \sqrt{2J\beta}$$

where $2J$ is the action being the following average computed using the model β -function at the location of every BPM:

$$(1.29) \quad 2J = \left\langle \frac{A^2}{\beta^m} \right\rangle$$

which allows to reconstruct the real β function from amplitude which might deviate from the model as follows

$$(1.30) \quad \beta = \frac{A^2}{2J}.$$

The detailed analysis of uncertainties of this technique are described in [52]. Each of these two methods for the computation of the β -function has its specific limitations and hence, depending on the given optics settings and operation mode, one of the methods is preferred.

Special measurements are needed in order obtain the β^* , i.e. the β -function at the location of IPs [53–55]. β^* is not measured directly at its location as in the case of BPMs, but is reconstructed from the β -function at the position of quadrupoles magnets surrounding an IP. The method is called *k-modulation* as it requires a variation of the gradient of quadrupolar fields. Based on the tune change induced by the a current modulation in a quadrupole the β at its location can be derived through the relation

$$(1.31) \quad \beta = \pm \frac{2}{l\Delta k} \left[\cot(2\pi Q) - \frac{\cos(2\pi(Q + \Delta Q))}{\sin(2\pi Q)} \right]$$

between the strength change Δk and the corresponding tune change ΔQ , with l being the length of the quadrupole and Q the nominal tune. This solution is general for the horizontal and vertical plane, differentiating the planes by the introduced $\pm$ sign. K-modulation technique is of great importance for the optics measurements at the LHC, since it is used to compute the β^* by modulating individually powered quadrupoles of the final focusing triplet.

Linear coupling is another observable which can be inferred from the harmonic properties of BPM signal. One possibility to measure linear coupling in circular machine is the use of resonance driving terms (RDTs) f_{jklm} obtained from the spectral lines. The concept of RDTs is introduced and presented in detail in [56, 57]. Each term corresponds to a mode in the transverse beam motion and hence contributes to a certain frequency of the motion. While the main frequency in the spectra refers to the betatron tune, the secondary line indicates the coupling harmonics. The relation between the amplitude of this secondary line and the main tune line is incorporated into the computation of the coupling RDTs $|f_{1001}|$, $|f_{1010}|$ and is described as

$$(1.32) \quad \begin{aligned} H(1-j+k, m-l) &= 2j|f_{jklm}|(2J_x)^{\frac{j+k-1}{2}}(2J_y)^{\frac{l+m}{2}} \\ V(k-j, 1-l+m) &= 2l|f_{jklm}|(2J_x)^{\frac{j+k}{2}}(2J_y)^{\frac{l+m-2}{2}} \end{aligned}$$

where J_x and J_y are the action in horizontal and vertical plane, respectively. $|f_{1001}|$ and $|f_{1010}|$ observables are given as complex numbers and their amplitude is constant in sections free of multipoles but increases at the location of these coupling sources. Coupling RDTs can be measured via the following equations derived in [56, 58]

$$(1.33) \quad |f_{1001}| = \frac{1}{2} \sqrt{\frac{H(0,1)V(1,0)}{H(1,0)V(0,1)}} \quad |f_{1010}| = \frac{1}{2} \sqrt{\frac{H(0,-1)V(-1,0)}{H(1,0)V(0,1)}}$$

where $H(1,0)$ and $V(0,1)$ are the amplitudes of the betatron tune lines in the horizontal and vertical planes, and $H(0, \pm 1)$ and $V(\pm 1, 0)$ correspond to the coupling harmonics. In the presence of coupling, f_{1001} will give rise to a peak in the horizontal spectrum at the frequency of the vertical tune [58].

1.4.4 Correction methods

Optics corrections aim to minimize the measured optics errors caused by magnets imperfections as described above. According to the type of error sources to be mitigated, the techniques can be divided into global and local corrections. Global corrections are used to correct the optics errors globally along the lattice in each beam separately, while local corrections aim to identify error sources which cause optics perturbations in a particular machine segments by using individual quadrupoles, which affect both beams at the same time.

The global correction incorporates the *response matrix* approach, modeling a matrix R that relates the optics functions as response to changes in the corrector magnets strengths $\Delta \vec{k}$ [59].

Correctors in the response matrix represent either a set of quadrupoles powered in series (circuit) or an individually powered corrector quadrupole. The response matrix R is computed using the ideal model of the machine providing the relation between phase-beating $\Delta\phi_{x,y}$, beta-beating $\Delta\beta_{x,y}$, normalized dispersion-beating ΔND_x and tune-beating $\Delta Q_{x,y}$ and changes in the quadrupole strength as follow:

$$(1.34) \quad R\Delta\vec{k} = (\Delta\vec{\phi}_{x,y}, \Delta\vec{\beta}_{x,y}, \Delta\vec{ND}_x, \Delta\vec{Q}_{x,y})$$

Furthermore the method allows the specification of quantity specific weights w . Thus the strengths of the correctors needed to perform the corrections can be computed using the measured values of optics observables and the inverse R^{-1} of the response matrix as

$$(1.35) \quad \Delta\vec{k} = -R^{-1}(w_\phi\Delta\vec{\phi}_{x,y}, w_\beta\Delta\vec{\beta}_{x,y}, w_{ND}\Delta\vec{ND}_x, w_Q\Delta\vec{Q}_{x,y})$$

Recently, an analytical approach for the response matrix has been introduced into the optics corrections at the LHC [60]. This allows to improve the speed and robustness of the global corrections computation compared to evaluating the response of the observables from the models computed via MAD-X simulation tool [61]. Response matrix approach is utilized also in the correction of the coupling present in the machine. The required skew quadrupoles strengths settings are computed by constructing the response matrix which relates the sum and difference resonance deviations with respect to the model to the strength changes in sextupoles.

The identification of local error sources in the LHC is based on the Segment-by-Segment (SbS) technique [62–64] and is used mostly around the IPs. The optics parameters at the location of the segment are obtained from the measurement and propagated with the ideal model. By comparing the propagated model with the measured optical parameters in the segment, local deviations can be observed. The phase deviation between the propagated model and measured optics are matched using the least square minimization method on

$$(1.36) \quad \Delta\phi_{n+1} = \phi_{n+1}^{meas} - \phi_n^{meas} - (\phi_{n+1}^{mod} - \phi_n^{mod})$$

representing the deviations in the phase advance ϕ between each BPM in the region of interest. After the phase deviation is reproduced with target accuracy, the identified magnet strength are incorporated as correction setting in order to eliminate this deviation. Another approach to identify local error sources is based on Action phase jump analysis [65]. It incorporates the idea of using the action and phase of the arcs outside the IR to estimate the distortion produced by a magnetic error in the IR.

1.4.5 Limitations of existing methods

Considering the process of acquiring the BPM signal, analysing the optical functions and computing required correction settings, specific problems appear at these individual stages. In the following, the limitations to be mitigated by the means of ML are defined and are further

discussed in the dedicated chapters where ML solutions developed within the frame of this work are presented.

Since the computation of the optical functions is mainly based on turn-by-turn BPMs readings, the entire optics analysis relies heavily on the quality of the BPM data provided to the analysis tools. The main challenge in identifying faulty BPMs is the fact that some of the faults can be observed only after the full optics analysis is completed. The unphysical outliers in the computed optical functions indicate the presence of faulty BPMs, however due to the application of N BPM method, determining which of the BPMs used to compute a specific data point caused the erroneous result is not trivial. In case such outliers are observed, the corresponding BPM reading needs to be manually removed from the harmonic analysis data before the optics computation can be repeated. This time consuming procedure should be replaced by applying unsupervised learning, improving the quality of the turn-by-turn data and reducing the manual effort.

The discussed above corrections computation at the LHC is designed to be performed in several steps, conducting different methods for the corrections depending on the location of the error sources, individually per beam in case of global corrections. In order to develop a general correction approach, capable to provide the corrections of all quadrupoles for both beams simultaneously, supervised learning techniques are investigated.

Due to the presence of faulty BPMs and the need to remove such signals, the optics functions used to compute the corrections settings are corrupted by absence of data points at the locations of identified faulty BPMs. Furthermore, there are certain noise artefacts present in the optics data. These issues can potentially lead to imprecise correction computation and hence, will be also addressed in the search for appropriate ML solutions for the reconstitution of missing data and denoising.

1.5 Recent Advances on Machine Learning in Accelerator Physics

ML is being applied to a large number of problems across different industrial and scientific domains, including High Energy Physics where different ML algorithms became a standard for data analysis boosting the search for new discoveries [66]. However, in accelerator physics, the interest in applying ML has grown only recently. Noticeable advances have been achieved especially in the domain of diagnostics, control and automation. The target of this section is to provide a brief overview on recently developed ML applications across different accelerator facilities which efficiently complement or even surpass traditional methods or human performance. For more detailed overview on opportunities in ML for particle accelerators see [67, 68] and [69] specifically for incorporating neural networks into accelerator controls.

Virtual Diagnostics

Various instruments and diagnostics techniques are required in order to monitor the beam itself and variables which affect its parameters. Virtual diagnostics can assist in case a direct measurement would have a destructive impact on the operation or in the locations where no physical instrumentation can be placed. ML can provide techniques to build such virtual beam diagnostics instruments. Simulation studies and experimental demonstrations have been carried out on Facility for Advanced Accelerator Experimental Tests (FACET-II) and Linac Coherent Light Source (LCLS) to study ML-based longitudinal phase space (LPS) prediction. Training data for a feed-forward NN has been acquired from a large number of simulations that represent changes in LPS distribution as response to the change of various accelerator parameters, as well as from the existing measurements at LCLS. ML models built on such datasets demonstrate a good agreement between the prediction and simulated or measured LSP images allowing to acquire fast and precise LPS reconstruction [70]. Another example is the estimation of oscillation amplitude and synchrotron damping time based on LPS measurements at Shanghai Synchrotron Radiation Facility (SSRF) [71]. Here, Gradient descent algorithm is used to estimate the fitting parameters which are then used as target variables in a supervised model. A neural network is trained to predict these values from longitudinal phase measurements obtained from the Beam Position Monitors (BPM). Another example from SSRF is a study on correlations between the beam size and the images from multi-slit imaging system, aiming to improve the accuracy of BPMs using neural networks [72].

A special kind of neural networks, *convolutional neural networks* (CNN) [73] have been applied at Fermilab Accelerator Science and Technology (FAST) facility on image based diagnostic during beam operation [74]. A combination of a CNN and a feed-forward neural network yields promising results for the prediction of beam parameters on simulated data sets. The model uses simulated cathode images, solenoid strengths and the gun phase as inputs and produces a prediction for various downstream beam parameters. Application of neural networks can be found also in correction of distorted beam profiles measured at ionization profile monitors (IPM) [75]. A neural networks-based model has been trained on IPM simulations in order to establish the mapping between measured profiles together with bunch length and bunch intensity to the original beam profile.

Automatic control and optimization

ML methods can be especially useful to solve problems in non-linear and time-varying systems with large parameter spaces. Operation of a complex system such as an accelerator, whose beam dynamics exhibits nonlinear response to machine settings can be considered as a typical ML task. Due to the constant increase of machine design complexity and development of new interacting systems, traditional techniques might become insufficient. Reinforcement learning, a special concept of training ML-models based on the interaction with task-specific environment,

demonstrates a great ability to solve complex control tasks in robotics and automation [76]. Recently, its application has been studied on control tasks in the domain of accelerators, e.g. for the control of the micro-bunching instability at the KIT storage ring Karlsruhe Research Accelerator (KARA) [77]. Reinforcement learning is being investigated at CERN at various facilities as an alternative to traditional optimization problems applied to autonomous accelerator tuning. Several model-free approaches have been successfully tested promising improvements in performance stabilization of several facilities across CERN accelerator complex. Application of supervised learning using NNs has been demonstrated as efficient approach to model the beam dynamics in large accelerators [78]. By representing the Taylor maps with adjustable parameters of the NN structure, the dynamical system can be approximated with required accuracy enabling the application of this approach to simulations and to the tuning of beam optics models with experimental data. Apart from neural networks, it is also possible to apply other kinds of regressors or classifiers in accelerator control such Gradient Boosting classifier as it was shown in beam loss pattern classification for LHC [79]. The beam loss maps performed in controlled conditions are used in order to train a model to classify the type of losses during the LHC machine cycle. Another example of ML-based automation is the automatic alignment of collimators in SPS and LHC [80]. The method computes optimum angular settings for the collimators without human intervention. Fully automated alignment was achieved after the introduction of the ML-based detection of alignment spikes in the beam loss records which indicate the position of the collimator with respect to the beam.

Another ML-based approach to boost the accelerator performance is the application of gaussian process and bayesian optimization as a control technique, e.g. to maximize the average pulse energy in FELs. The methods developed based on this approach allow to tune a large amount of machine components simultaneously based on noisy measurements [81]. Extremum Seeking (ES) technique [82, 83] for adaptive feedback in combination with ML is being applied to accelerator tuning and optimization providing promising results. ANN is used for fast tuning in order to obtain a close approximation for the target settings, while ES is applied to track the actual time-varying optimal parameter settings. The advantages of this method are the model-independent approach, the ability to perform the tuning on many coupled parameters and handle time-varying noisy data.

The application of ML-based solutions in accelerators domain clearly demonstrates its great potential and advantages against manual or poorly automated operation. Especially, for the scenarios where large amount of free parameters has to be taken into account in order to meet the requirements and find optimal machine settings under given limitations. Only few parameters can be processed by a human at once and it is not feasible to produce forecasts taking into account all possible factors and correlations. Moreover, humans can perform differently on the tasks where the decisions can be subjective, which might lead to significantly different results and hence, unstable machine operation. Furthermore, the automation of routine tasks could bring

significant improvements into daily operation, such that the focus of operation can be transferred to complex tasks and rare events that require expert knowledge.

DETECTION OF FAULTY BEAM POSITION MONITORS

Faulty BPMs produce erroneous signal causing unreliable computation of optics functions and therefore, detection of faulty BPMs prior to optics computation is crucial for adequate optics analysis. Most of the faults can be removed by applying traditional cleaning techniques. However, optics functions reconstructed from the cleaned turn-by-turn data regularly exhibit a few non-physical values which indicate the presence of remaining faulty BPMs. A novel method based on Isolation Forest algorithm has been developed demonstrating the required performance in operation and improving the quality of the optics measurements. This chapter presents this concept and summarizes the results obtained during LHC beam commissioning and Machine Development sessions (MDs) in 2018, discusses the evaluation of the developed method under simulated failure conditions together with the optimization of the preexisting cleaning technique. The advantages of the chosen algorithm compared to some other unsupervised learning techniques are also discussed.

2.1 General Concept

Data cleaning is an important part of optics analysis consisting of several steps along the analysis process. Cleaning of optics measurements data refers to reduction of the noise of the obtained BPM signal, as well as removing anomalous signal in turn-by-turn data, including detection of faulty signal artefact in later stages of data processing such as harmonic analysis. Several numerical thresholds as well as a cleaning technique based on Singular Value Decomposition (SVD) are used in order to remove faulty signals [48, 84]. Since non-physical values still can be observed in the optics functions computed from the cleaned data, it is not uncommon that additional manual cleaning of harmonic analysis data is required, followed by repeating the optics computation. Therefore, the preexisting traditional techniques appear to be insufficient and

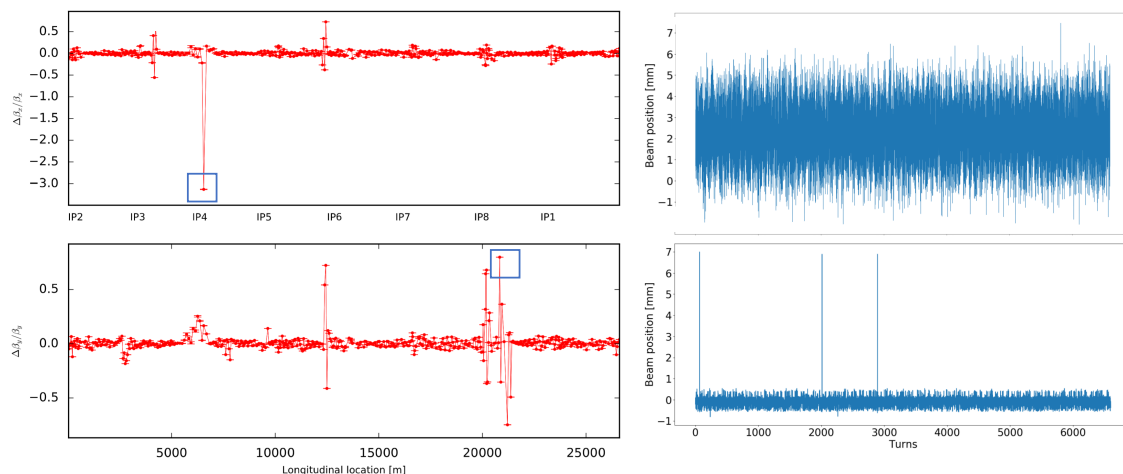


Figure 2.1: Including the turn-by-turn signal of two BPMs shown on the right leads to unphysical outliers in the computed β -function (on the left). The outliers at the location of the demonstrated BPMs are indicated by the blue squares.

alternatives for automatic identification of faulty BPMs are required to avoid human intervention and ensure that faulty signal does not corrupt the optics analysis.

Training a supervised classification model on the data labeled by cleaning results from the past is not possible, since the new method aims to improve the quality of turn-by-turn data instead of reproducing the results of existing tools. Since the reasons of BPM failures are partially unknown, it is unfeasible to define rules which would indicate faulty BPMs that actually cause the erroneous optics computation. To illustrate a challenging case of identification of faulty BPM signal by the means of static thresholds and manual cleaning, Fig. 2.1 presents an example of turn-by-turn data obtained from two different faulty BPMs and the outliers in the reconstructed β -function caused by these BPM faults. While some of faulty BPMs clearly exhibit anomalous patterns such as amplitude spikes as shown in the illustrated example, other faulty BPMs provide signal which appear "normal" however can be related to erroneous data points in the optical functions.

It has to be noted that there is not necessarily a direct relation between the location where the error is observed and the actual bad BPMs. Due to the way how the optics is calculated [51, 85], a single faulty BPM may cause erroneous optics calculation at multiple locations, i.e. produced errors might appear not directly at the position of the bad BPM, but propagate to the locations of adjacent BPMs. The new method should automatically recognize bad signals in the online provided turn-by-turn data prior to optics computation, without requiring predefined rules or thresholds. In this work a demonstration that unsupervised learning successfully meets these requirements is provided.

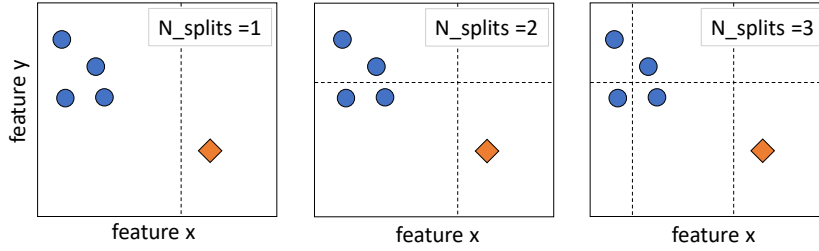


Figure 2.2: A conceptual illustration of IF algorithm. An anomalous point is more likely to be isolated by just one random split on a randomly selected input feature, compared to three splits required to isolate each of the core data points.

Application of Unsupervised Learning

Unsupervised learning deals with tasks where only input data is available and the target is to find patterns in the given data or to extract new information. The ability of unsupervised techniques to find hidden patterns in the data discussed in Sec. 1.3.7 is a powerful tool for faulty BPMs identification. The signal produced by a faulty BPM has different properties compared to normal functioning BPMs. Unsupervised methods can automatically find relevant data properties and the differences which indicate anomalous points in the data provided to the algorithm.

Most common clustering techniques based on centroid search such as K-means [27] appear not to be suitable for faulty signal detection, since the appearance of outliers (data points that significantly differ from other observations) affects the computation of the mean of parameters. A possible solution is to apply density-based algorithms which is discussed in Sec. 2.5.

Several unsupervised learning algorithms have been explored to improve the quality of the optics measurements at LHC, among others the Isolation Forest (IF) algorithm [86]. It detects data anomalies using an ensemble of randomized decision trees. As discussed in 1.3.6, building an estimator consisting of a single tree can lead to over-fitting, and hence using an ensemble of decision trees is preferred in favor of bias-variance trade-off. A decision tree represents a sequence of random splits which are performed until each single data point is “isolated”. The principle is illustrated in Figure 2.2. The split is selected randomly between maximum and minimum values of randomly selected data feature. On average fewer random splits will be needed to isolate an anomalous data point. The number of splits represents the path length from the root to the leaf of a decision tree. Using an ensemble of decision trees, the path lengths are averaged over a number of trees (forest). Based on the path lengths, “anomaly score” of each isolated point is calculated. To distinguish between anomalies and normal data points, one needs to specify a threshold, which defines how many data points will be determined as anomalies. Such threshold can be expressed as expected proportion of outliers, so-called *contamination rate*. The anomalies are then characterized as a fraction of data points with the highest anomaly scores, i.e. with the shorter path lengths and smallest numbers of isolating splits. Significant advantage

of the Scikit-Learn [21] implementation of the algorithm used in this work is that it requires only the number of trees and contamination rate as tuning parameters. IF does not require any data normalization or re-scaling and can be applied directly on harmonic analysis of BPM signals before computing the optics functions.

2.2 Traditional techniques

Traditional techniques for the cleaning of turn-by-turn data include thresholds to define anomalously large and low values of the measured beam position, identification of zero values which replace unphysical BPM readings, as well as manual cleaning. A special list of systematically failing BPMs is used in the dedicated analysis software to exclude these BPMs from the actual data processing. An approach to identify these BPMs dynamically is discussed in Section 2.7.

Most of the noise and faulty signals can be removed using these methods, as well as through applying advanced signal improvements techniques based on SVD which allows to identify faulty BPMs and to reduce the noise in recorded signal. The singular vectors of a turn-by-turn data matrix containing the signal of several BPMs correspond to temporal and spatial modes variations describing the beam motion [48]. SVD modes with localized spikes in their spatial vectors indicate faulty BPMs using so-called *SVD cut* as threshold to find such spikes. To globally reduce the noise on all BPM readings, only a predefined number of strongest singular modes, defined by the *SVD mode* setting, remain in the turn-by-turn data. While the SVD cut value has a direct influence on the number of BPMs identified as faulty, the SVD mode setting affects the overall noise level in turn-by-turn signal. Section 2.5 demonstrates the importance of choosing appropriate SVD settings and the interplay between IF which is the main subject of this work, and the SVD technique with respect to the overall cleaning performance.

The default settings used in LHC operation in 2018 were originally defined from RHIC turn-by-turn data statistical analysis [48]. Table 2.1 summarizes the SVD settings and numerical thresholds used as defaults in 2018. Peak thresholds choose the minimum peak-to-peak signal and maximum value of the signal to be considered physical. Recording exact 0 in one or several turns indicates a bad BPM since this value is used to replace unphysical BPM readings. The betatron tune obtained as the main frequency of the BPM signal can serve as another indicator for faulty BPMs. The threshold for the tune deviation from the average value over the entire set of BPMs defines the limit of tune variability for correct BPM signals. The operational results of the preexisting cleaning tools presented in the following section are obtained using default settings.

2.3 Applying Isolation Forest in BPMs Data Analysis

In the LHC there are 523 BPMs per plane and per beam. The analysis of the results obtained by traditional cleaning tools from the measurements before 2018 has shown that around 10% of

Table 2.1: Default SVD settings and signal cuts used in 2018.

SVD settings		Signal-based cuts	
SVD cut	0.925	min peak-to-peak	10^{-2} mm
SVD modes	12	max peak value	20 mm
		max tune deviation	10^{-5}
		unphysical signal	exact 0

all BPMs are identified as faulty using these tools [87]. Due to experience with the repeating observation of few non-physical outliers in the reconstructed optics functions, it is assumed that only a small fraction of bad BPMs is remaining in the data, so most of the bad BPMs are eliminated by existing techniques.

As discussed above, IF requires the contamination (expected fraction of anomalous samples) as input of the algorithm. During beam commissioning and MDs in 2018 the contamination was set to 1% in the arc sections and 2.5% in the IRs. A higher contamination rate in the IRs is assumed based on the analysis of previous measurements in [87]. The separation into IRs and arcs BPMs is needed due to the fact that BPM hardware installed in IRs is different from the arcs sections [88], therefore they have to be treated separately.

The target is to apply the new cleaning method on the properties of BPM readings computed by a special harmonic analysis [46, 89]. The parameters that are considered as significant for bad BPMs identification are the betatron tune, the amplitude obtained from the FFT scaled by factor 2 with respect to the oscillation amplitude, A , and the noise to amplitude ratio, σ_{scaled} . The latter is defined as

$$(2.1) \quad \sigma_{scaled} = \frac{std(x_{raw} - x_{clean})}{A\sqrt{N}},$$

where x_{raw} and x_{clean} stand for the BPM data before and after applying the SVD cleaning, respectively. The standard deviation of the difference of these two signals, $std(x_{raw} - x_{clean})$, represents the BPM noise. N is the number of turns. Figure 2.3 illustrates detection of faulty signals by IF, using the described features as input data.

2.4 Results on Operational Data in 2018

During commissioning and MDs in 2018 IF algorithm was used complementary to the preexisting cleaning tools. During the evaluation phase of the new introduced technique, the optics was first computed using the data cleaned with traditional tools only and then repeated with turn-by-turned data additionally cleaned with IF. This allowed to observe the positive impact of the new method on the quality of optics analysis. Figure 2.4 shows an exemplary comparison between the β -beating reconstructed from the measurements cleaned with the traditional tools only and the measurements additionally cleaned with IF. It demonstrates that most of the observed outliers in

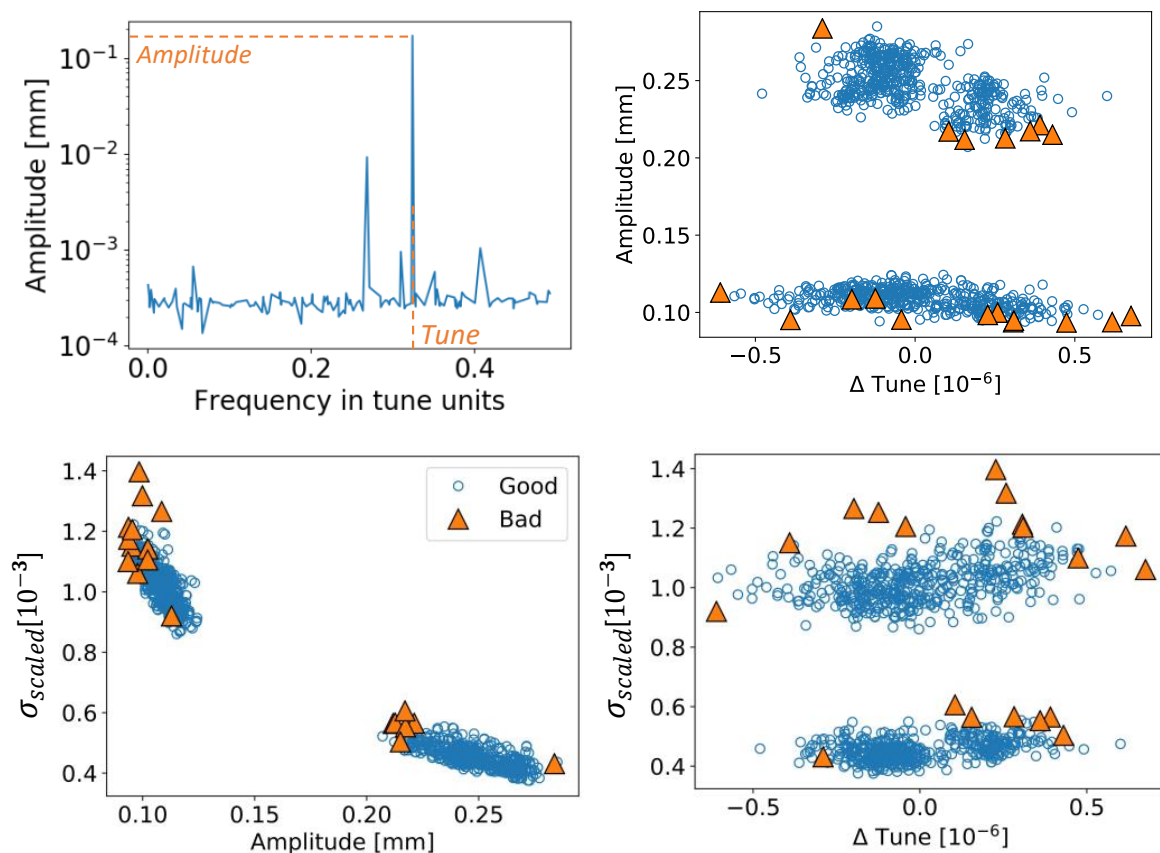


Figure 2.3: The upper left plot represents, for a single BPM, the tune and amplitude from harmonic analysis that are used together with noise to amplitude ratio σ_{scaled} as input features for the IF algorithm. 2D-projections of the feature space illustrate the detection of bad BPMs.

the optics obtained from SVD-cleaned data, can be prevented by using IF. The removal of BPMs at the locations where no spike has been observed, did not cause significant data loss.

In order to conclude on the effectiveness of the new method for online optics analysis, a collected summary of the measurements where IF has been used. Figure 2.5 shows the summary of cleaning results on several measurements in 2018 that are listed in Table 2.2. The BPMs at the locations of outliers usually need to be manually analysed and removed from harmonic analysis of turn-by-turn data before recomputing the optics. This procedure requires additional human intervention and still does not guarantee that bad BPMs are correctly manually detected and erroneous measurements do not appear in the recomputed optics. The statistics on operations data shows that most of the spikes remaining after SVD and thresholds-based cleaning could be successfully removed by IF, without requiring any manual cleaning.

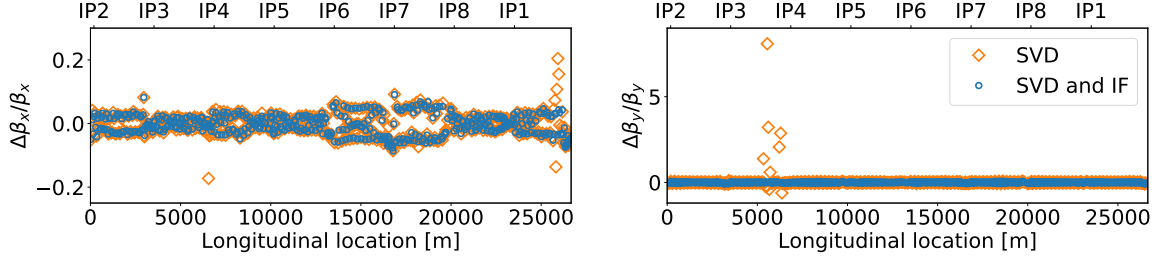


Figure 2.4: Comparison between β -beating measured from SVD-cleaned data and the data additionally cleaned with IF. The data is obtained during ion commissioning in 2018. Without applying IF nonphysical spikes appear in β -beating reconstructed from the BPM signal.

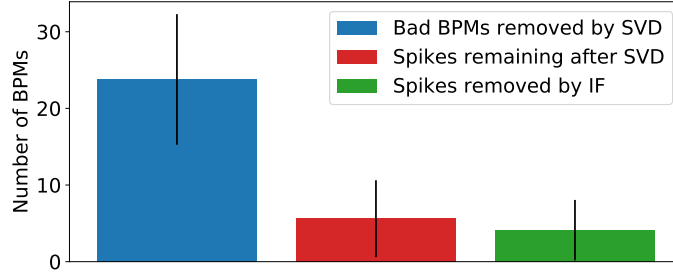


Figure 2.5: Summary on cleaning results based on the number of outliers (spikes) per plane per beam appearing in computed β -beating and phase advance averaged over 10 measurements listed in Table 2.2.

Table 2.2: List of measurements where IF has been used during online data analysis in 2018. Beam 1 appears more often in the statistics just due to the human intervention in the optics analysis and subjective decision about the need for additional cleaning with IF.

Date	Optics, β^*	Beam	Corrected
29.05.2018	commissioning, 90m	1	no
29.05.2018	commissioning, 90m	1	yes
29.05.2018	commissioning, 90m	2	no
12.06.2018	flat, 60 / 15cm	1	no
12.06.2018	flat, 60 / 15cm	1	yes
22.08.2018	ion commissioning, 1m	1	yes
08.10.2018	ion commissioning, 50cm	1	no
08.10.2018	ion commissioning, 50cm	1	yes
08.10.2018	ion commissioning, 50cm	2	no
08.10.2018	ion commissioning, 50cm	2	yes

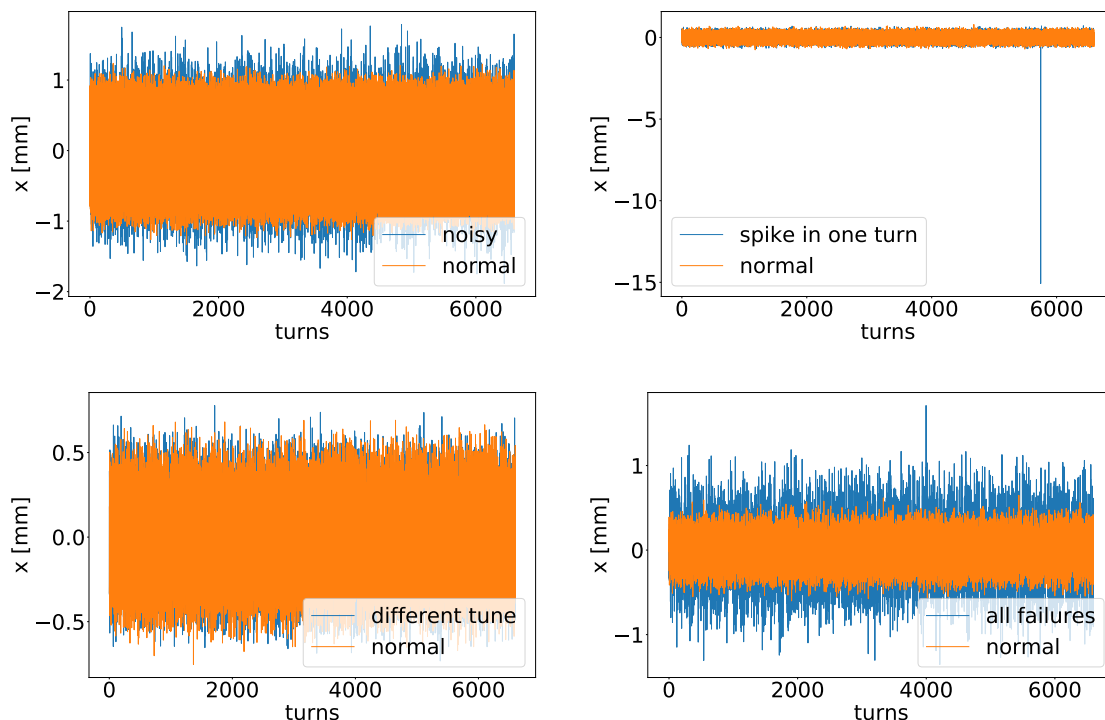


Figure 2.6: Simulated turn-by-turn data of the bad BPMs compared to normal unperturbed signal over 6600 turns.

2.5 Simulating faulty BPM signal identification

Since the definition of erroneous values is exclusively based on the observation of computed optics functions, it is not possible to conclude about the exact amount of actual faulty BPMs which are removed and the number of good BPMs which are wrongly recognized as faulty. The knowledge about actual malfunctioning BPMs currently present in the machine is not available during the operation. Therefore, it has to be studied on simulated measurements with artificially introduced BPM faults, such that cleaning results can be verified against the ground truth. This avoids subjective counting of outliers in the computed optics functions as efficiency measure of applied cleaning methods.

2.5.1 Model of faulty BPMs

First, turn-by-turn BPM signals are generated for 6600 turns using ion optics model with $\beta^* = 50$ cm in IP1, 2 and 5 using MAD-X and a dedicated Python script. Every BPM is given 0.1 mm Gaussian noise. In the second step, the signal of some randomly chosen BPMs is artificially perturbed - these BPMs have to be identified as bad. In real measurements, the reasons for the appearance of faulty signal are unknown, but there are specific artifacts which

are known to be related to faulty BPMs. Therefore, these known properties are used to model a presence of faulty BPMs in the LHC. It has to be noted that these artifacts do not describe all possible failures, however in order to verify the effectiveness of cleaning methods, it should be sufficient to assign the known failures to a realistic number of BPMs in the simulations according to the number of observed anomalies in the measurements data.

Following failure modes are used to introduce BPM faults:

- Gaussian noise added to the signal is 0.3 mm, 3 times higher compared to good BPMs.
- Signal in 1 turn is replaced by a random value in range $[-20, 20]$ mm, such that produced local spike is smaller than the threshold for the maximal absolute peak value used for simple cuts.
- Tune computed from the signal deviates by 10^{-5} from the rest of BPMs.
- All described failures are present.

The examples of perturbed turn-by-turn signal produced by the introduction of the listed failure modes are shown in Figure 2.6. In addition to the described failure modes, BPMs with flat zero signal in all turns and zero signal in 10 randomly chosen turns are also introduced. These 2 failure types are trivial to detect and hence not relevant for the method verification. Nevertheless, they have to be included to produce more realistic turn-by-turn data simulations. Considering the number of bad BPMs found by traditional tools and remaining spikes before applying IF (≈ 30 per plane per beam as shown in Fig. 2.5), the ratio of bad BPMs over the total number of BPMs per plane per beam in 2018 was $\approx 5.5\%$. Hence, 5.5% of BPMs in original simulated turn-by-turn data are perturbed. All failures, apart from flat zero signal are equally distributed over generated BPM signal with 5 occurrences and flat zero signal is simulated at 2 BPMs, producing 27 bad BPMs per plane in total.

2.5.2 Isolation Forest results

First, harmonic analysis is executed on the generated turn-by-turn data using the traditional cleaning techniques without changing the default settings. Knowing the actual bad BPMs and their faults, unsupervised method can be evaluated in the combination with traditional tools, applying IF on harmonic analysis of the SVD cleaned BPMs data.

Statistical analysis of operational data presented in Sec. 2.2 shows, that around 15 bad BPMs remain after using traditional cleaning tools and 12 bad BPMs are removed, the contamination should be set to $15/(524-12) \approx 0.029$. To study the influence of the contamination parameter on the optics computation, simulations are used to run IF multiple times increasing the contamination number from 0 to 0.15 step-wise. Figure 2.7 illustrates the trade-off between eliminating bad BPMs and removing good BPMs as side effect. In order to find an optimal contamination factor, a definition of an acceptable maximum number of missing good BPMs which does not cause negative

Table 2.3: Number of faulty BPMs detected by SVD together with combination of thresholds compared to number of BPMs detected after additionally applying IF on cleaned data. The results are obtained using SVD default cut value 0.925 and IF contamination factor set to 0.02. The numbers are averaged over 20 simulated measurements. The basic zero cut failures are not included into comparison.

Failure	SVD	IF	Not detected
Higher Noise	0	3±1	1±1
Random values	1±1	2±0	1±1
Tune deviation	0	2±1	2±1
All failures present	1±1	3±1	0

Table 2.4: Demonstration of the change in the cleaning results, if the SVD cut is set to the optimal value 0.3 and IF contamination factor is accordingly changed to 0.01.

Failure	SVD	IF	Not detected
Higher Noise	5	0	0
Random values	2±1	1±1	2±1
Tune deviation	0	3±1	2±1
All failures present	5	0	0

effects on optics analysis is required. In Figure 2.7 a sharp increase of the number of detected bad BPMs and slow increase of removed good BPMs can be observed. After contamination factor reaches 0.02 the trend changes and the rise of the number of removed good BPMs becomes steeper than the increase of detected bad BPMs. After contamination number reaches 0.04 there is no significant increase in the number of removed bad BPMs anymore. Based on the obtained results, the conclusion is that the optimal contamination factor lies between 0.02 and 0.04 as expected, if the data is previously cleaned by SVD using the default parameter settings. Considering the different failure modes, Table 2.3 shows that under the default settings, IF complements the traditional techniques exactly in the cases where they are insufficient.

2.5.3 Exploring SVD settings

The simulated data with artificially introduced bad BPMs has been also used to explore if the change of SVD cut threshold can improve the detection of faulty signals. The change in cleaning results with respect to SVD cut values presented in Figure 2.8 shows that the optimal SVD cut threshold range is [0.3, 0.6], noting equal results for 0.4 and 0.3 values. Bigger values lead to an increase in the number of the remaining bad BPMs, which has been demonstrated not only in simulations, but also in LHC measurements in relation to the new failure mode described in 2.7. Lowering the threshold to values smaller than 0.3 leads to an increase in the number of good BPMs wrongly identified as bad.

The change of SVD cut from its default value 0.925 requires also the change of the IF contamination factor. Since less faulty BPMs will appear in the data, lower contamination factor

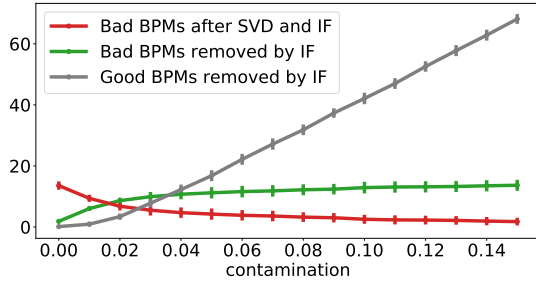


Figure 2.7: Adjustment of contamination factor of IF algorithm. The target is to keep the number of bad BPM remaining after applying SVD and IF (indicated by the red line) low, while avoiding a significant number of good BPMs to be wrongly identified as bad.

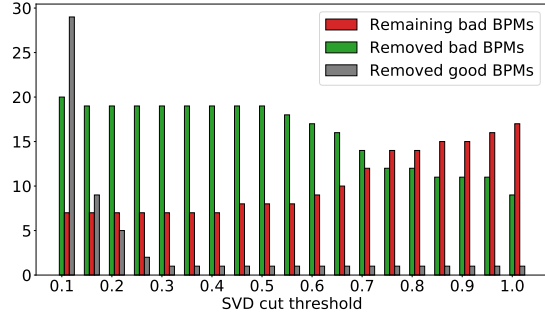


Figure 2.8: Number of removed and remaining bad BPMs after applying SVD depending on SVD cut value.

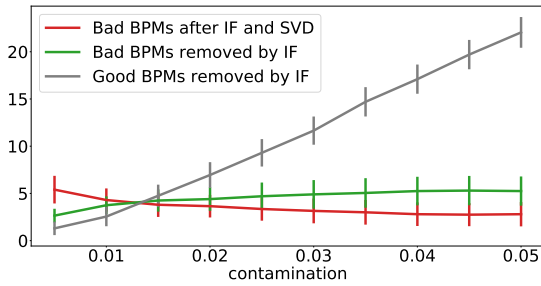


Figure 2.9: Adjustment of contamination factor of IF algorithm after changing the SVD cut to its lowest optimal value 0.3.

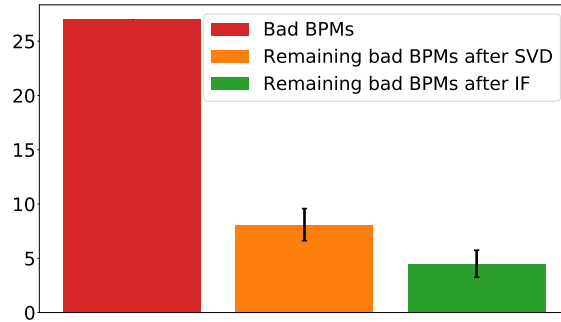


Figure 2.10: Averaged results of faulty BPMs detection on 20 simulated measurements using IF contamination factor 0.01 and SVD cut 0.3.

should be needed. So, the scan of IF contamination factor is repeated on the data cleaned with SVD using the lowest optimal cut value 0.3 in order to compare it with IF performance under the default SVD settings. Figure 2.9 demonstrates how the cleaning result changes with the increase of contamination factor. In case the SVD cut is set to 0.3, the IF contamination factor should be within the range [0.01, 0.015] since further increase results in a bigger number of good BPMs wrongly removed by the algorithm. The study on the detection of particular failures is then repeated using the obtained optimal settings for both, SVD and IF, namely SVD cut of 0.3 and contamination of 0.01. The averaged results for each of the introduced failure modes are summarized in Table 2.4. The final results on data cleaning applying both, previous cleaning techniques using the optimal SVD cut value and IF with accordingly adjusted contamination rate, are summarized in Fig. 2.10.

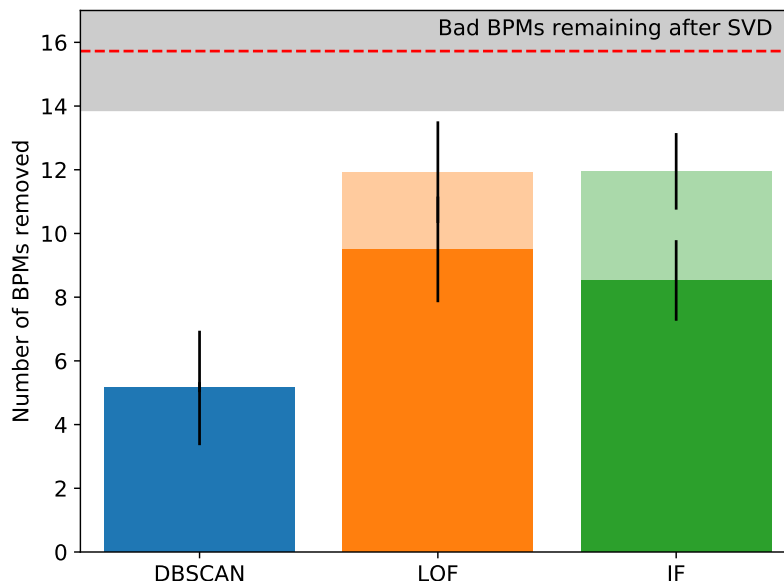


Figure 2.11: The comparison is carried out on 20 simulations for each plane, the results are averaged. The simulated turn-by-turn data has been cleaned using default numerical thresholds and SVD settings described in Table 2.1. Each bar represents the number of BPMs removed by the indicated method. Dark color corresponds to the number of removed BPMs that are actually bad.

2.5.4 Comparison to clustering

Prior to the integration of IF algorithm into optics measurements, several clustering techniques have been considered as possible solutions to improve the cleaning results. These techniques have been tested together with IF on the harmonic analysis of LHC turn-by-turn measurements obtained in the past. One of the investigated approaches is a density-based DBSCAN algorithm [28] which views clusters as areas of high density separated by areas of low density. The method finds core points which build a cluster center, assigns neighboring points to this cluster and considers as anomalies the points which do not belong to any cluster. A core point is defined by the minimum number of points within a distance. Hence, there are following parameters to be defined: minimum number of samples in the neighborhood of a core point, the distance to the neighbors and the metric to be used to compute the distance. The results of applying DBSCAN to bad BPMs detection demonstrated improvements on data cleaning [90], however a significant amount of outliers remained present in the measured optics functions. Another technique which has been applied to the LHC turn-by-turn data is Local Outlier Factor (LOF). This factor indicates the local deviation of density of a given sample with respect to its nearest neighbors [29]. LOF measures how isolated an object is with respect to the surrounding neighborhood, which is very similar to the IF algorithm. However, apart from the contamination rate, LOF requires the number of neighboring data points and the definition of a metric to compute the distance between

the points in order to find the nearest neighbors. Since the structure of the measurements data can vary significantly depending on the BPM location and machine settings, a general valid definition of the distance and number of neighbors to build a cluster becomes problematic.

In order to examine the performance and suitability of each method for faulty BPMS detection, the previously described simulation procedure is carried out. The parameters of clustering techniques has been defined empirically during the application on LHC data. The distance between the points to define a cluster in DBSCAN is set to 0.7, LOF contamination is 0.05. Both methods use euclidean metric to compute the distance and the minimum of 70 neighbors to build a cluster. SVD and thresholds based cleaning have been applied prior to clustering using the default values described in Sec. 2.2. Figure 2.11 summarizes the obtained result. The comparison shows nearly identical performance of LOF and IF algorithms on simulated BPM signal. Applying different methods on LHC measurements, one observes that the optics reconstructed from the data cleaned with IF contains fewer unphysical outliers compared to the other two methods. Moreover, due to a smaller number of settings, IF allows simpler tuning of the cleaning algorithm and more general application. Therefore, it is preferred as an alternative cleaning tool for faulty BPMS detection at the LHC.

2.6 Faulty BPMS detection in the presence of local coupling

Betatron coupling drives the appearance of oscillations in the horizontal plane with the vertical tune and vice-versa. Strong local coupling sources must be corrected to avoid luminosity loss [91] or its propagation to the rest of the machine, therefore it has to be ensured that cleaning tools do not have a negative impact on the computation of local coupling. The coupling is measured in terms of its Resonance Driving Terms (RDTs), f_{1001} and f_{1010} as described in 1.4.3. Since these RDTs are calculated from the spectrum of cleaned BPM signal, the settings of the cleaning tools affect the coupling computation. As the local coupling information is contained in secondary lines of only few BPMS spectra, using a small number of modes in SVD cleaning causes coupling information to be discarded as noise. Considering IF cleaning, the BPMS whose signal contains the information about local coupling might be removed completely based on the difference from the rest of BPMS. In the following the ability of IF to distinguish local coupling related signal from faulty BPMS is presented, along with the optimization of SVD mode number with respect to local coupling.

To simulate faulty BPMS detection in the presence of local coupling, first MAD-X tracking using ion optics from 2018 with $\beta^* = 50$ cm is performed. In order to introduce a local coupling bump, the integrated field strength of the skew quadrupoles around IP2 are changed by ± 0.001 m⁻¹. Also, a small global β -beating (1%) is introduced in order to get more realistic optics from the simulated signal. Produced tracking simulations are then converted into turn-by-turn measurements in order to introduce BPM faults as described in 2.5. In this case it is

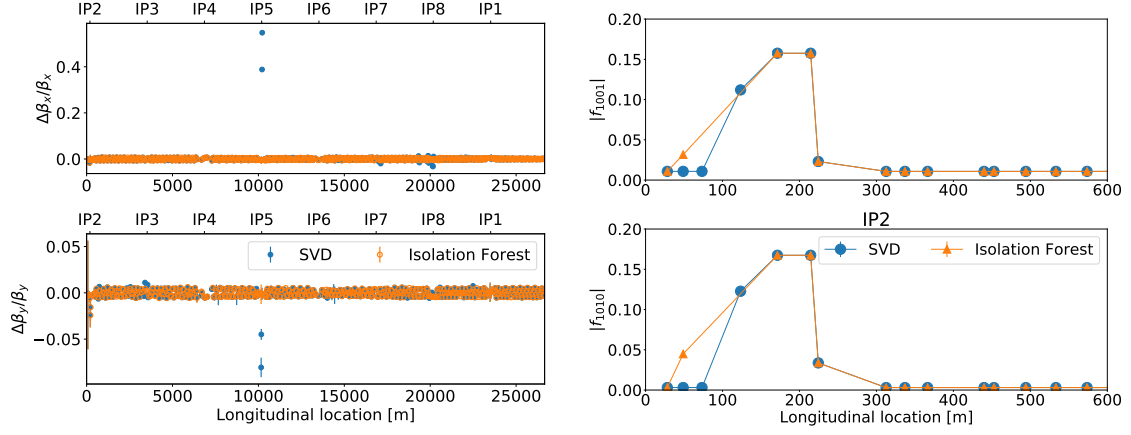


Figure 2.12: Application of IF on SVD cleaned data reduced the number of outliers in beta-beating (left) so the simulated bad BPMs are detected correctly. At the same time the introduced coupling bump in IP2 is still present (right), showing that the signal at the location of the coupling bump (BPMSW.1R2.B1, BPMSW.1L2.B1) where f_{1001} and f_{1010} reach their maximum values is not classified as anomaly as expected.

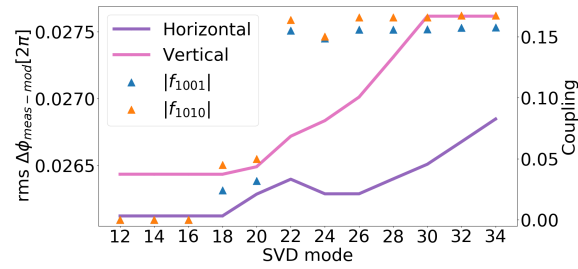


Figure 2.13: Trade-off between correct computation of the expected coupling at the location of introduced skew quadrupole field error and increase of the noise in the measured phase advance errors $\Delta\phi_{\text{meas-mod}}$ defined as rms between measured and model phase advances. Right scale shows the computed values of f_{1001} and f_{1010} depending on the number of SVD modes used in SVD-cleaning.

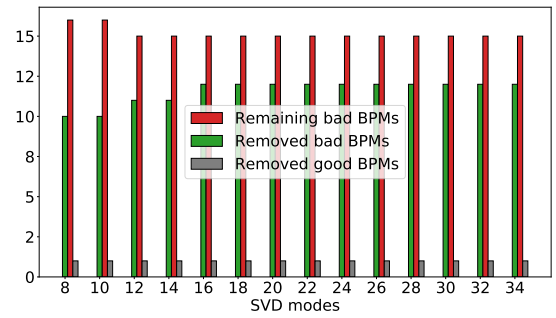


Figure 2.14: The relation between BPM cleaning results and the change of the number of SVD modes used in analysis.

ensured that no fault is assigned to the BPMs at the location of coupling bump (BPMSW.1R2.B1, BPMSW.1L2.B1). In the following, the influence of SVD mode setting on coupling computation is addressed, keeping SVD cut and IF at default values previously used in operation. The default SVD mode number used for the LHC in 2018 is 12. However, when reconstructing f_{1001} and f_{1010} from SVD-cleaned data, the presence of local coupling can be observed only by increasing SVD mode to 18 as shown in Figure 2.13. To reproduce the actual expected value of simulated coupling around IP2 at least 22 SVD modes are needed. Further increase of SVD modes still produces correct f_{1001} and f_{1010} values, however it affects the phase computation negatively since the measurement becomes more noisy. Therefore, increasing the default value of SVD modes up to 22 could help to gain more information about local coupling without significant increase of noise in computed phase. The scan of this parameter with respect to the number of removed BPMs performed on simulations is shown in Figure 2.14. Increasing the number of modes to 16 and further gives a slight improvement in bad BPMs identification while keeping the number of good wrongly removed BPMs constant. Conclusively, the number of SVD modes used in turn-by-turn data analysis should be increased in order to achieve more reliable coupling computation without causing negative effects on signal noise cleaning and faulty BPMs detection. Figure 2.12 demonstrates the result of additional cleaning with IF compared to the optics obtained from data cleaned with traditional techniques only. The local coupling can be observed after application of IF since BPMSW.1R2.B1, BPMSW.1L2.B1 are correctly identified as non-anomalies. At the same time, the outliers in β -beating produced by the simulated faulty BPM signal are eliminated.

2.7 Detecting unknown failure mode in experimental data

It has been observed in operational measurements that some BPMs repetitively caused erroneous optics calculations. These BPMs (23L6.B1, 16R3.B1, 22R8.B1, 15R8.B1) have been manually analysed in regards of betatron tune domain and phase shifts aiming to identify a signal properties which can serve as indicator for fault detection [92, 93]. In the past, such indicators could not be found, and these BPMs were simply removed from the data prior to data processing. Recently, these four BPMs together with three further BPMs (25L5.B1, 10R6.B1, YB.4R8.B1) [94], are found to exhibit identical pattern in the spectra with side bands around the tune frequency line [95]. Figure 2.15 shows examples of the signal related to the described pattern. The reason of the recently discovered failure mode still remain to be understood. We demonstrate that BPMs related to this failure mode can be identified using the combination of preexisting cleaning tools and IF, without observing the BPMs signal spectra a-priori. The seven listed BPMs can be detected using either specific settings of SVD-clean or by applying IF in addition to SVD-clean with arbitrary settings. This is crucial for successful automatic data cleaning since even a careful manual inspection did not guarantee full elimination of faulty BPMs in the computed optics functions.

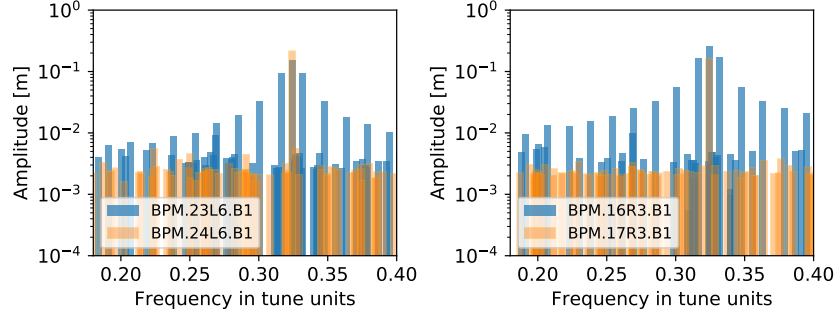


Figure 2.15: Examples of the faulty BPMs signal spectra in the vertical plane (blue) compared to the spectra of normal functioning BPMs (orange). The measurement is performed on Beam 1, $\beta^*=65$ cm.

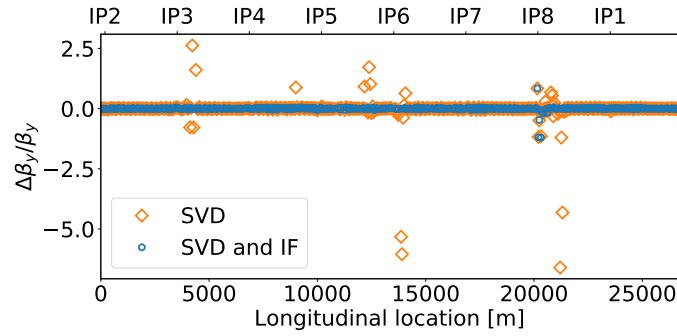


Figure 2.16: In case the optics is computed from the data additionally cleaned by IF, the appearance of nearly all unphysical outliers is prevented.

In total there are seven BPMs found to have the described spectra pattern, however the presence of each of these BPMs causes multiple outliers in the computed optics as shown in Fig. 2.16. SVD is incapable of identifying this failure, unless reducing SVD cut to 0.6 which cleans 20% of affected BPMs. Removing all seven BPMs is possible only by using SVD cut 0.3 which agrees with the optimal range for this setting obtained from simulations described in the previous section. Independently of SVD settings, applying IF helps to clean 90% of these BPMs. The presented showcase demonstrates the advantage of combining traditional cleaning tools with unsupervised learning and the ability of presented techniques, given the optimal settings obtained from extensive simulation studies, to identify BPM faults without providing manually identified rules and to detect even a-priori unseen failure pattern.

2.8 Summary

The newly developed method based on the Isolation Forest (IF) algorithm successfully detected faulty BPMs during LHC operation and MDs in 2018 where this was previously done via tedious human intervention. Extensive studies and simulations presented here show that indeed the previously existing techniques cannot perform as efficiently as in combination with IF. Moreover, the presented result shows that the simulations qualitative reproduce experiment observations, such that an understanding of unsupervised learning as a method for faulty BPMs identification could be obtained.

Further investigations have been carried out on the interplay between previously available cleaning techniques and the introduced IF algorithm, demonstrating the advantage of applying IF prior to optics computation, instead of using SVD only to identify BPM faults. The number of bad BPMs remaining in the data can be reduced by factor 2 to less than 1% performing anomaly detection with IF on SVD-cleaned data. Both methods, SVD and IF can be used complementary noting the importance of adapting the thresholds accordingly. In this regards, it has been demonstrated that measurement of important optics observables such as local coupling require increasing the number of SVD modes above the default value of 12. IF does not influence negatively the local coupling measurement.

Considering a more general application, the developed cleaning approach can be potentially used in other accelerators, performing IF on a given set of signal properties suitable to a particular accelerator type. The choice of contamination rate has to be carefully estimated according to the available BPM infrastructure, by the means of statistical analysis of historical data or simulating BPM failures as demonstrated in this work. The improved understanding of the combination of SVD cleaning and IF algorithm obtained in the presented study will further benefit future measurements during Run III of the LHC.

RECONSTRUCTION OF QUADROPOLAR ERRORS FROM LINEAR OPTICS OBSERVABLES

Magnetic field errors and misalignments cause optics perturbations, which can lead to machine safety issues and performance degradation. The correlation between magnetic errors and deviations of the measured optics functions from design can be used in order to build supervised learning models able to predict error sources directly from a selection of measured optics observables. Extending the knowledge of real magnetic errors offers potential improvements of beam control by including this information into optics models and corrections computation. In the following, the methodology of utilizing the supervised learning approach and linear regression for quadrupole field errors estimation based on measured optics perturbations is presented. The objective of this chapter is to study and to demonstrate the usefulness of supervised ML algorithms for magnet errors reconstruction in a circular accelerator such as the LHC, for which the presented method has been applied in simulated environment, as well as on experimental data.

3.1 Introduction

Optics corrections are crucial for safe machine operation and reliably high performance in terms of luminosity balance between experiments. Currently, LHC optics corrections are performed in two steps, i.e. local corrections based on Segment-by-Segment technique [63, 64] or Action Phase Jumps [65], and global corrections using Response Matrix approach [62]. Local corrections are applied in interaction regions where the strongest error sources are usually observed, while global corrections are performed to correct the error sources globally, including the optics errors in the arcs. The corrections to be applied in the LHC are computed as a set of magnetic field

strength changes - either in the so-called circuits (quadrupoles powered in series) or individual quadrupoles that can be trimmed independently. More details on the correction methods applied to the LHC are given in the dedicated section 1.4.4. At the LHC, several steps of acquiring data for optics analysis, computing and applying local and global corrections for each beam separately are usually needed. These methods allow achieving unprecedentedly low β -beating [42, 96] by computing the strengths change needed to compensate the measured optics perturbations.

While this approach aims to minimize the measured optics errors, it does not provide the information about actual errors in individual magnets which caused these perturbations. Selection of appropriate variables together with optimization of corresponding weighting factors in the response matrix are fundamental to produce efficient global corrections. Although these optimizations can be achieved based on the existing optics model, in practice the variables and weights are defined empirically. Using supervised regression this procedure can be improved by learning the variable importance and weights automatically from the training data. Building a technique able to include the optics perturbations from both beams and to consider the whole set of individual error sources along the lattice promises to reduce the time needed for optics correction and allows to extend the knowledge about the present magnet errors. The study presented in this chapter demonstrates the ability of supervised regression models trained on a large number of simulations to predict the individual quadrupole errors given the measured optics perturbations caused by these errors in one step for both beams simultaneously. For the LHC, more than a thousand of quadrupolar gradient field errors are predicted from several thousands of optics observables, including local optics in IRs, as well as the arcs sections.

First, the concept of the method is presented and a summary of the preliminary studies which demonstrated the promise of supervised learning based approach is provided. In Sec. 3.3 the details on data set generation and the setup of simulation environment are given, followed by the discussion on supervised training using the generated data and model selection with regards to generalization and accuracy presented in Sec. 3.4. Finally, Sec. 3.5 presents the results of model prediction, obtained on realistic simulations considering individual magnet classes, followed by the validation of introduced ML methods on the experimental LHC data is presented in Sec. 3.6.

3.2 General Concept

Magnetic field and misalignment errors distort optics parameters of accelerators. The general idea of applying supervised learning to optics corrections is to create a regression model which will automatically learn the correlation between the field and alignment errors of the magnets and the resulting optics perturbations, from the provided simulations data. Given a set of features $X = x_1, x_2, \dots, x_m$ and targets $Y = y_1, y_2, \dots, y_n$, an estimator can learn a regression model or a non-linear approximation. Such estimator should use the measured optics functions deviations from design as the set of input features and predict the error sources as output targets. Supervised

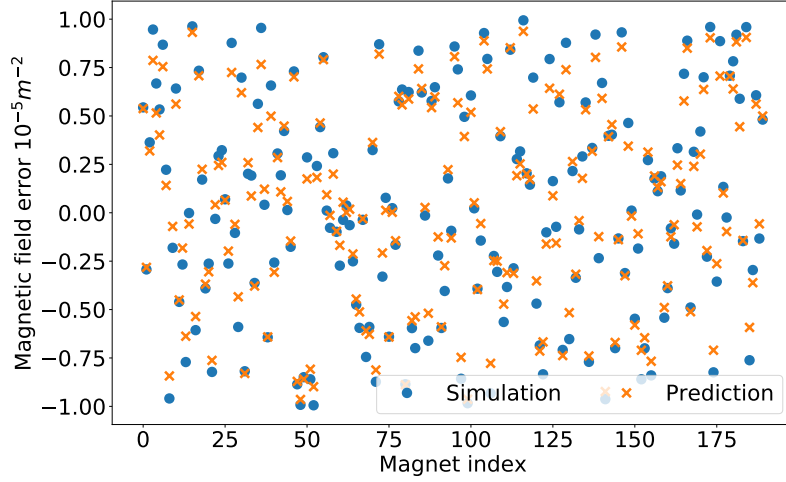


Figure 3.1: Random Forest prediction result on a random sample from the test set giving mean absolute error $0.02 \times 10^{-5} [m^{-2}]$.

training requires a large set of training data in order to be able to generalize and produce reliable results on unseen data. The correction results obtained with traditional methods in the past years are not suitable as a source to acquire the training data, since corrections are only performed few times per year in the LHC and hence, not enough data is available. Moreover, the corrections applied in the LHC utilize the strength of “circuits”, sets of quadrupoles powered in series, and hence the corrections do not correspond to the errors of the individual elements.

3.2.1 Preliminary studies

As a first step towards building a supervised learning based correction approach, simplified proof-of-concept studies have been performed [97, 98]. In these studies, MAD-X variables that represent correction circuits in the arcs sections are used as error sources perturbing the optics functions, also the noise in the input variables is neglected. In the light of assumed simplifications, a training set of a few thousand of samples is sufficient in order to provide acceptable performance of a predictor and hence, several regression models based on different algorithms have been trained on 7000 training samples. The lowest error of prediction has been achieved by Random Forest algorithm [23] based on applying an ensemble of decision trees to regression or classification tasks as described in Sec. 1.3.6. Figure 3.1 demonstrates the comparison between true simulated data and Random Forest prediction on an arbitrary chosen sample from the test set. Expected optics corrections resulting from the Random Forest prediction of circuits errors are compared to Response Matrix global corrections, in the absence of triplet errors and local corrections. The comparison is shown in Fig. 3.2 demonstrating promising performance of ML-based correction approach.

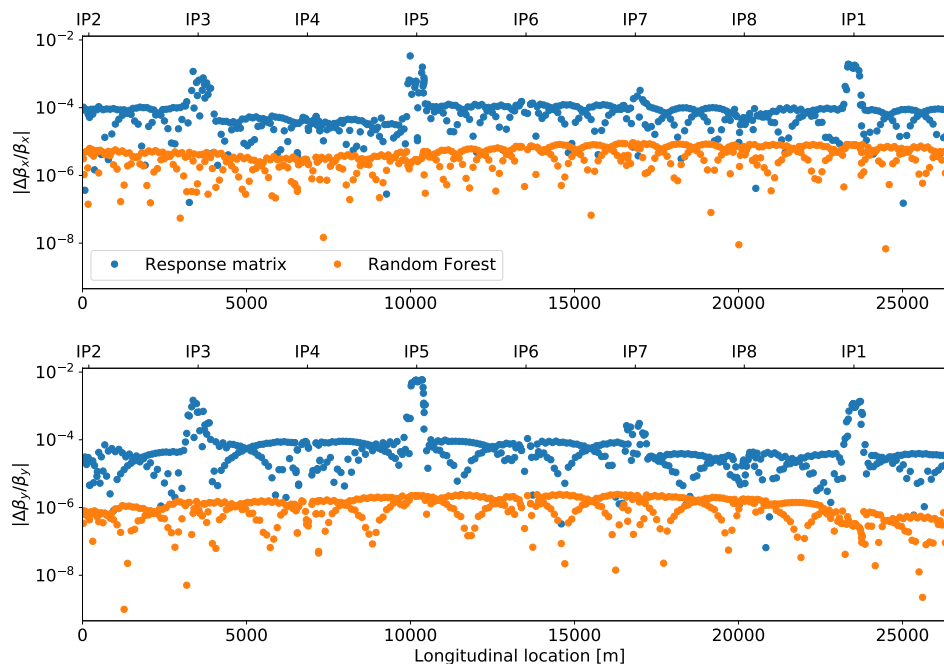


Figure 3.2: Expected β -beating after applying corrections computed with linear response matrix and Random Forest regressor on simulated measurement.

Moving towards the evaluation of the concept on a more realistic use case, the regression models trained on the simulations of the optics perturbed by the errors in the circuits variables have been tested on simulations generated using magnet errors in individual quadrupoles, neglecting the errors in the triplets, but adding Gaussian noise to the input variables [99]. The target variables are represented by the circuits strength changes and the resulting corrections of the optics errors after applying the settings to the circuits, can be directly compared to Response Matrix corrections results. Figure 3.3 shows an example of simulated optics and its corrections using Response Matrix and a Convolutional Neural Network (CNN) [73]. The main advantage of this special type of neural networks is the ability to capture the spatial dependencies through the application of learned filters. In other words, the network can be trained to understand and extract spatially correlated features. This fact makes the CNN specially appealing for the application on correction prediction, as the determination of many optical parameters depends on the relationship between neighboring BPMs as described in 1.4.3.

Different ML algorithms for regression have been compared demonstrating similar performance, producing corrections results on the level of the traditional Response Matrix approach as shown in Table 3.1. However, the realistic errors of every single magnet have a different effect on the optics compared to the strength change in the circuits. In the following, the approach of simulating the optics perturbation in the training data using individual magnet errors will be discussed. The target is to obtain the information on the the actual error sources, i.e. errors of

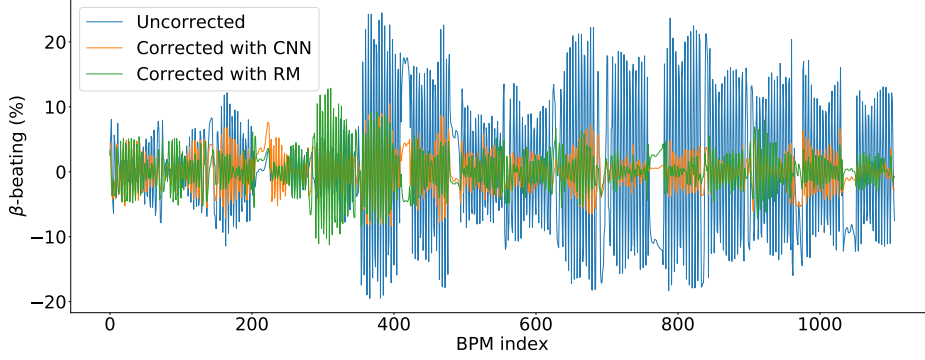


Figure 3.3: Simulated β -beating after applying corrections computed with linear RM and CNN trained on errors in the circuits, reducing rms β -beating from 9.5% to 3.2% and 3.0%, respectively. The measurement is simulated using the optics with $\beta^* = 40$ cm and with the introduction of phase advance noise.

Table 3.1: β -beating after applying the predicted circuits errors as correction settings, the results are averaged over 100 simulations of perturbed collision optics with $\beta^* = 40$ cm.

β -beating %	peak	rms
Uncorrected	32 $\pm$ 10	11 $\pm$ 3
Response Matrix	11 $\pm$ 5	3 $\pm$ 2
Orthogonal Matching	11 $\pm$ 2	3.5 $\pm$ 0.8
Convolutional Neural Network	11 $\pm$ 2	3.2 $\pm$ 0.5
Ridge regression	10 $\pm$ 2	2.9 $\pm$ 0.8
Linear regression	9 $\pm$ 2	2.6 $\pm$ 1.7

individual magnets, instead of predicting the strength change of correctors circuits.

3.3 Data generation for prediction of individual magnetic errors

In order to build a data set for supervised training, a uniform distributed random field errors are generated and assigned to all quadrupoles available in the LHC for both beams. Additionally, a fix set of quadrupolar errors is also applied to the dipoles in order to obtain more realistic simulations. The generated errors are then added to the nominal model of collision optics with $\beta^* = 40$ cm, producing differences between the design and simulated optical functions. These deviations from ideal optics together with a set of associated magnet errors in one simulation build a corresponding training sample, of which several thousands are created. The summary of this concept is presented in Fig. 3.4.

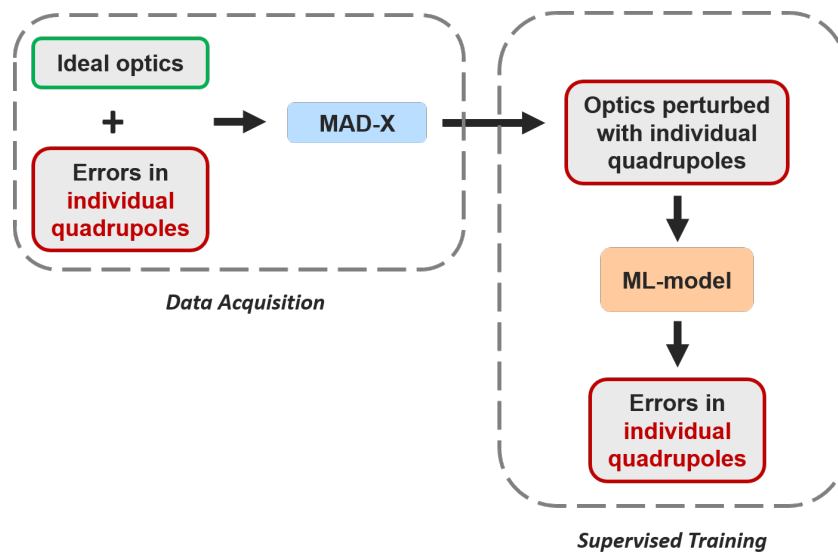


Figure 3.4: Conceptual representation of data generation and supervised model creation for the estimation of individual magnet errors from optics deviations.

Table 3.2: RMS values of integrated relative field errors σ_K/K_1 assigned to different quadrupole (MQ) families and misalignments of quadrupoles and sextupoles (MS), with σ_s and σ_x being the RMS of longitudinal and transverse misalignment respectively, used for training data generation.

Magnet	$\sigma_K/K_1[10^{-4}]$	$\sigma_s[\text{mm}]$	$\sigma_x[\text{mm}]$
MQ	19	...	...
MQX	4	6	...
MQY	11	...	...
MQM	12	...	...
MQW	15	...	...
MQT	75	...	...
MS	...	...	0.3

3.3.1 Magnet errors as target variables

In this study, following error sources are considered: quadrupolar integrated field errors, longitudinal misalignments of quadrupoles and sextupole transverse misalignments. The errors are simulated according to the magnet measurements [100] using MAD-X simulation framework. All quadrupoles in the lattice are assigned a random relative gradient error obtained from uniform distribution with the same rms error σ per magnets family. In case of the triplet magnets, also the systematic error (the mean of the errors in a magnet family in a single LHC simulation) is defined as uniform distribution in a range $[-10^{-3}, 10^{-3}]$. The distribution of relative field errors are defined in Table 3.2 along with alignment errors of the sextupoles and triplet magnets.

Transverse sextupole misalignments induce a quadrupolar field error. A dedicated study has been performed in order to explore replacing misalignment errors by modified field errors in quadrupoles while keeping the optics deviations produced by included error sources realistic.

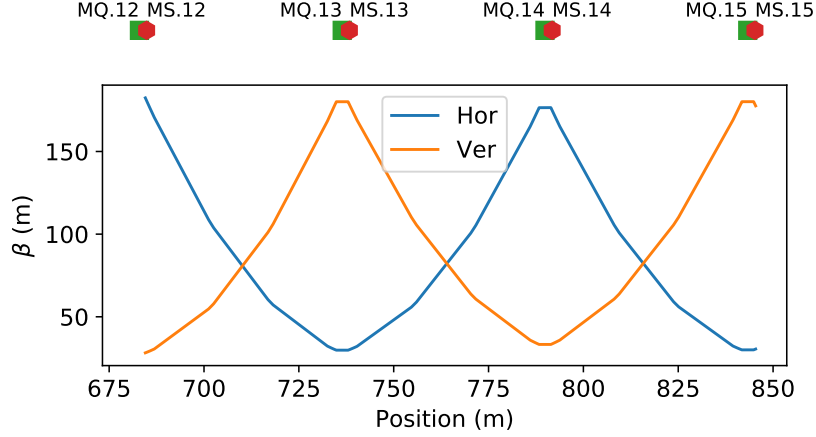


Figure 3.5: A fragmentary representation of LHC arc elements and β -function showing the adjacent placement of arc quadrupoles (MQ) and sextupoles (MS).

The target is to mitigate the degeneracy caused by the close proximity of the sextupoles to the quadrupoles in the LHC arc sections illustrated in Fig. 3.5. The degeneracy derives from the fact that the offset in the position of a sextupole and quadrupolar field error can be assumed to be the same source of β -beating.

For the verification, the approach is to match the β -beating produced by a single sextupole with 1 mm offset with the field change in the neighboring quadrupole. In case the residual β -beating is negligible, the offsets can be reflected by increasing quadrupolar field error with the approximation

$$(3.1) \quad \sigma_K/K_{1,new} = \sqrt{\sigma_K^2/K_{1,original}^2 + \sigma_x^2(K_2L)^2/K_{1,original}^2}$$

where K_2L is the integrated sextupolar strength and $\sigma_K/K_{1,original}$ and $\sigma_K/K_{1,new}$ being original rms of simulated arc quadrupole gradient field errors and the new rms which reflects the presence of sextupole transversal offset. Figure 3.6 shows the reconstruction of the β -beating induced by the displacement of a sextupole in the arc on the right of IP2 (MS.13R2) using the strength change of the adjacent quadrupole. This result confirms that the optics deviation produced by the expected transversal offset of sextupoles is reproducible by increasing the rms of the quadrupolar field errors distribution. Since the sextupoles are also used to correct chromaticity and the β -function in the arcs changes depending on the optics scheme, Eq. (3.1) is optics dependent.

This approach has been explored also to study the longitudinal offset of triplet magnets. According to matching results, a longitudinal error of one magnet cannot be fully replaced by a gradient field error. It has been shown for the optics with $\beta^* = 20$ cm, that β -beating introduced by longitudinal misalignments of the triplets can be reproduced with corresponding field errors [101]. Figure 3.7 shows an example for the optics with $\beta^* = 40$ cm, where the additional contribution

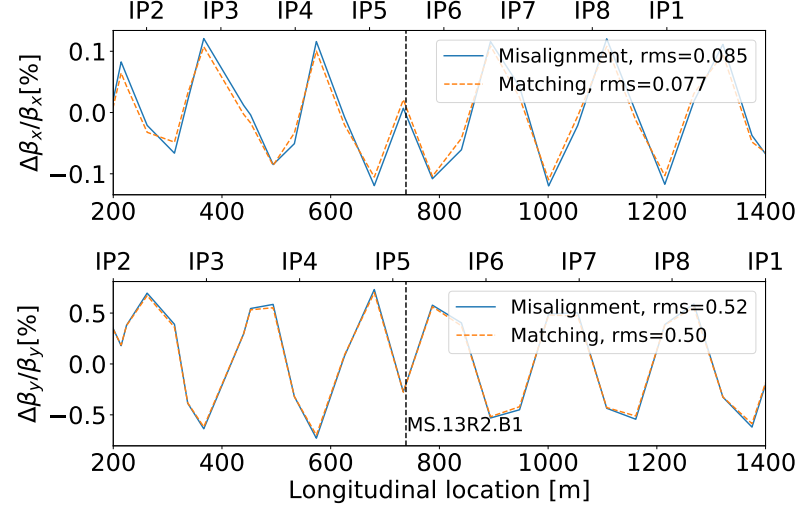


Figure 3.6: The optics perturbation produced by the transversal misalignment of a sextupole can be matched using the adjacent quadrupole field change, with an excellent agreement between the resulting rms β -beating values indicated in the legend.

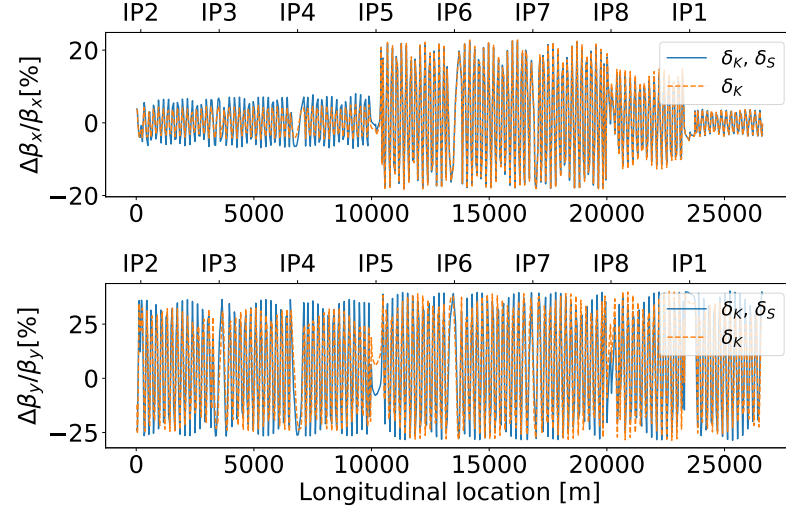


Figure 3.7: The comparison between β -beating produced by quadrupolar field errors in the triplet magnets (δ_K) distributed according to Tab. 3.2, compared to β -beating caused by additionally adding longitudinal misalignments (δ_S) of 6mm.

of the misalignment to β -beating is small compared to the perturbations caused by gradient field errors. Hence, the longitudinal misalignments of the triplet magnets are neglected for the purpose of this study, however the effect of the longitudinal offsets should be addressed in the future related work.

In the following, the reconstruction of effective quadrupolar field errors will be addressed, by training ML-model to predict the total number of 1256 quadrupolar field error in the entire LHC lattice optics observables simulated at all BPMs for both beams.

3.3.2 Optics functions as input features

Feature engineering, i.e. the process of finding the most relevant properties in the input data for predicting the output targets incorporates techniques to measure the feature importance and to build new input sets in order to simplify and improve the model training. Collecting appropriate data is a crucial milestone of the ML-pipeline since the model selection and achieved performance strongly depend on the amount and quality of the given data.

The correction of β -beating can be ensured with the phase beating correction [41]. Hence, the simulated phase advance deviation from the design between each pair of consecutive BPMs is considered as one input feature. The β -functions next to the interaction points (IPs) indicate the local optics errors which are dominated by the errors in the triplet magnets. Thus, the horizontal and vertical β at the BPMs left and right of main interaction points (IP1, IP2, IP5, IP8) of both beams are included as input features into the regression model, in order to improve the prediction of the triplet errors. The deviations of measured dispersion prior to optics corrections are related to quadrupolar errors. The measurement of normalized horizontal dispersion $D_x/\sqrt{\beta_x}$ is independent of BPM calibration errors [102] and hence, its deviation from the nominal values is also added to the set of important input parameters.

In total, 3346 features are extracted from the simulation data. Realistic noise estimated from the measurements data is added to phase advance and dispersion features. Phase advance noise is assumed to be $10^{-3} \times 2\pi$ in a BPM with $\beta=171$ m and it is scaled with the $1/\sqrt{\beta}$ at the rest of locations. The normalized dispersion noise is estimated from the recent optics measurements with $\beta^* = 25$ cm and is assumed to be a chi-square distribution with non-centrality parameter 4 and scaled with $10^{-3}\sqrt{m}$ as demonstrated in Fig. 3.8. The random noise added to the original simulated deviations in normalized dispersion is generated accordingly.

3.4 Model selection and training

As it was shown in [99], applying complex models such as convolution neural network does not result in significantly better corrections, so the model based on the least-squares linear regression with weights regularization [9, 13], so called *Ridge Regression* is used as baseline model for presented studies. Another reason for the choice of linear model is that the error sources to be

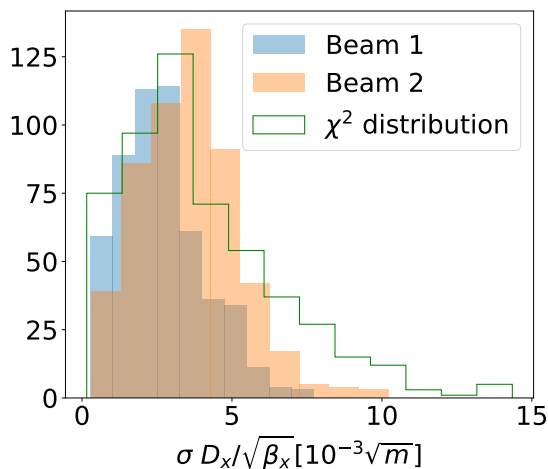


Figure 3.8: The noise distribution in the dispersion measurements of beam 1 and beam 2 compared to the simulated noise generated as a chi-square distribution with non-centrality parameter 4.

predicted are known to introduce linear optics perturbations. This regression technique has been introduced in Sec. 1.3.4.1.

The trained model needs to be evaluated using adequate quality metrics considering the data representation, desired result and specific task to be solved. In order to conclude on the learning performance, the data set is separated into training (80%) and test (20%) sets. The typical figures of merit for regression tasks are the mean absolute error (MAE) to compare the difference between true target values and the output of the model and the coefficient of explained variance R^2 defined as

$$(3.2) \quad R^2(\vec{y}, \hat{\vec{y}}) = 1 - \frac{Var\{\vec{y} - \hat{\vec{y}}\}}{Var\{\vec{y}\}}$$

$$(3.3) \quad MAE(\vec{y}, \hat{\vec{y}}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where $\hat{\vec{y}}$ are the true simulated magnet errors with $\hat{y}_i$ being the predicted value of the i -th sample, $\vec{y}$ the corresponding true values for N total input-output pairs (samples), and Var is the variance, the square of the standard deviation.

Another technique to improve the prediction quality on unseen data is *bagging*. Bagging is based on the idea of generating multiple versions of an estimator and using the averages of targets over the versions for final predictions [20]. After extensive validation, a model with 10 Ridge regression - estimators, each using 80% of the training data, regularized by $\alpha = 1 \times 10^{-3}$ has been found as the optimal setup.

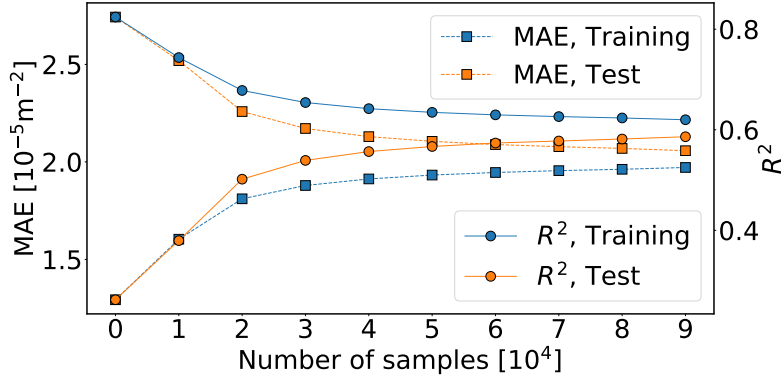


Figure 3.9: Model cross validation based on the loss (MAE) and R^2 coefficient depending on the number of available samples. The loss is constantly decreasing with the growing number of samples, while R^2 is increasing. This trend indicates a reasonable learning behaviour, however using datasets larger than ca. 80 000 samples does not improve the scores significantly.

Increasing the number of training samples does not necessarily result in a large increase of predictive power of the model. Considering the amount of time and storage needed to handle the training simulation data, especially for the future online application, we need to determine the optimal training set size. The change of the model scores with respect to the number of samples (*learning curve*) also indicates the ability of the model to learn from the given data and the data set size required to achieve the optimal model performance as shown in Fig. 3.9. The model scores MAE and R^2 are given as total scores computed from all predicted 1256 magnets from the simulated data samples, without separating the magnets into specific classes described in Tab. 3.2. The R^2 training and test scores converge after the total number of samples reaches 80000, with $R^2 \approx 0.6$ and $\text{MAE} \approx 2 \times 10^{-5}$, compared to the average of absolute magnet errors in the simulated data samples is 3.2×10^{-5} . The next section presents the results from the regression model with described parameters using 80000 samples in total for training and test.

3.5 Results on simulations

Identification of local error sources such as quadrupolar field errors is known to be a degenerate problem with multiple solutions. However, the estimates found by ML-model can be validated in two different ways - comparing the true simulated magnet errors to the ML-model prediction or by comparing the β -beating simulated by true errors to the β -beating produced by predicted errors. This allows to test how reliable prediction is on unseen data and to investigate the ability of the model to learn the physical correlations between the linear magnetic field errors and optics perturbations. The presented approach is tested on two different optics settings with the magnet errors generated according to the distribution listed in Table 3.1. First, the results obtained from

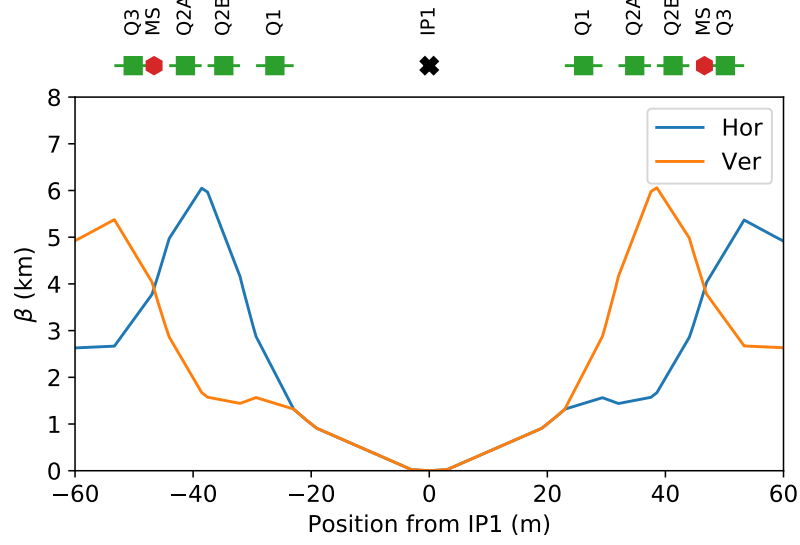


Figure 3.10: Schematic of the magnets layout (triplet quadrupoles Q1, Q2, Q3 and sextupoles MS), with error bars indicating the lengths of elements and β -function around IP1 for the optics settings $\beta^* = 40\text{cm}$ as used in training data generation. Please note that the value of the beta is given at the end of each element.

LHC simulations using the collision optics settings with $\beta^* = 40\text{ cm}$ are presented. Another special optics design, used to validate the ML-models is the so-called ballistic optics characterized by switching off the triplets in IR1 and IR5 [103].

3.5.1 Optics with $\beta^* = 40\text{ cm}$

As discussed above, the regression model is trained to predict the entire set of quadrupolar field errors consisting of 1256 variables. The field errors in the 32 triplet magnets installed in IP1, 2, 5 and 8 affect both beams simultaneously and produce the largest contribution to the optics perturbations. Hence, their prediction is evaluated separately from the rest of the magnets - 612 quadrupoles for each beam 1 and beam 2. This validation approach aims to ensure the ability of the presented method to reconstruct the most significant error source. Figure 3.10 demonstrates the large values of β function in the triplet region and illustrates the layout of single magnets (Q1, Q2, Q3) in one triplet circuit.

Figure 3.11 summarizes the triplet errors prediction results. In order to compare these results to the original simulated error distribution presented in Sec. 3.3, the model prediction is given as relative integrated field errors. To be noted that in the simulated datasets for training and test, the absolute values of quadrupolar field errors are used as provided by MAD-X simulations. To inspect the predictive power of the trained regression model, we compare the true simulated magnet error to corresponding residuals, computed as the difference between true and predicted values. The correlation between the individual simulated magnetic error and residuals

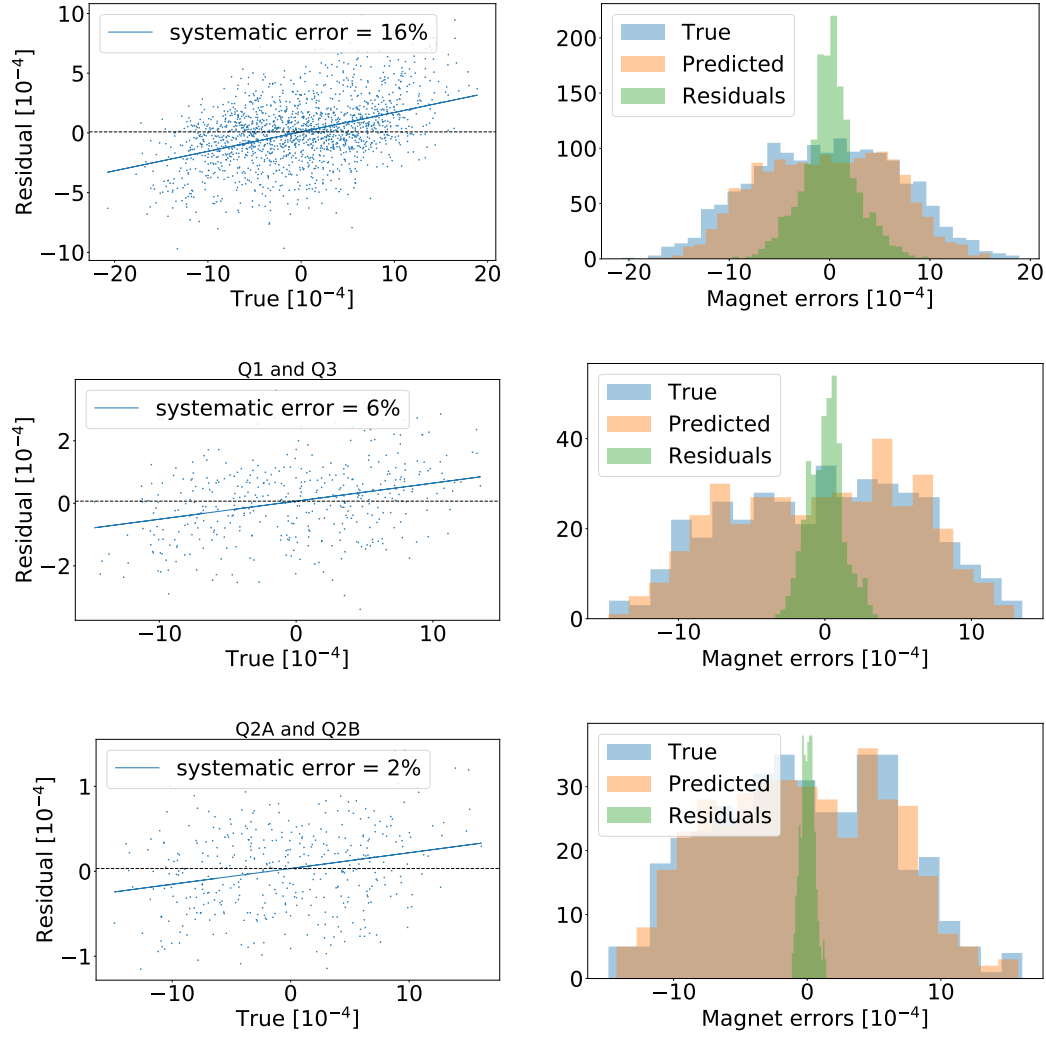


Figure 3.11: Results of predicting the quadrupolar field errors in the triplet for 100 LHC simulations. The upper row illustrates the prediction results obtained for the the field errors in individual triplet quadrupoles, while the middle and bottom rows correspond to the combination of Q1 and Q3 magnets and Q2A and Q2B magnets pieces respectively. Residual error is computed as the difference between true simulated errors and ML-model reconstruction. The slope in the plots on the left illustrates the systematic bias of regression model computed as correlation between residual error and corresponding true values.

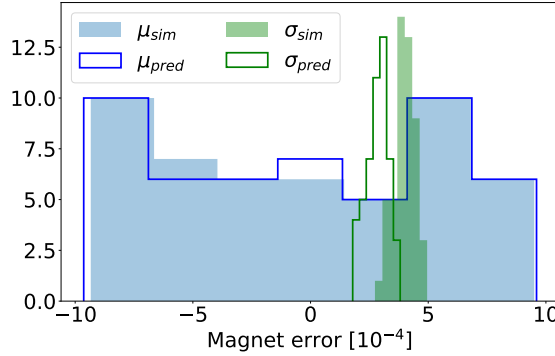


Figure 3.12: The systematic and random components of the true simulated quadrupolar field errors, computed as mean μ and rms σ of the errors distribution, compared to their predicted values.

demonstrates a systematic bias of prediction of 16%. The predicted errors are systematically smaller than corresponding true values in simulations. Since the spread of residual errors is large, it is not possible to correct this systematic obtained as statistical result from 100 LHC simulations.

The residual error reduces by incorporating the knowledge about the relations between single magnets in the triplet circuits. Q1 and Q3 magnets are powered together, while Q2 magnets have two pieces (Q2A and Q2B), which strengths can be balanced. Averaging the values of Q1 and Q3 magnets, the correlation between residual error and true simulated gradient field errors is reduced by a factor of 2 to 6%, while the combination of Q2A and Q2B pieces results in more significant reduction to 2%. The error of prediction for the arc magnets is 30%, being as expected larger compared to the triplet magnets, since the reconstruction of the error sources located in the arcs is affected by degeneracy stronger compared to the triplet magnets installed around the IPs. Inspecting the reconstruction of particular components of the simulated errors (systematic and random components of the expected magnet errors as described in Sec. 3.3), Fig. 3.12 demonstrates a very good agreement between the systematic components of the simulated and predicted triplet errors, whereas the random part appears to be the biggest contribution to the overall prediction error of the trained estimator. Combining individual triplet magnets also leads to a better agreement between rms values of simulated and predicted field errors. Table 3.3 provides a summary on the metrics used to evaluate the reconstruction of the quadrupolar field errors for the triplet and arc magnets separately, after predicting the entire set of errors by ML-model. In total, the resulting accuracy computed as relative absolute error of prediction of all 1256 quadrupolar gradient field errors, averaged over 100 simulations is 30%, while total R_2 score is 0.68.

We also perform the evaluation of model prediction in terms of reproducing the original optics errors. For each validation sample, we compare the β -beating in the original optics simulation

Table 3.3: Explained variance R^2 , residual error of prediction relative to the corresponding simulated true values and comparison between predicted and simulated RMS (σ_{pred} , σ_{sim}) values of the relative integrated field errors.

	Triplets			Arcs
	Individual	$\emptyset$ Q1,Q3	$\emptyset$ Q2A,Q2B	
R^2	0.84	0.94	0.98	0.6
$\frac{res}{true}[\%]$	16	6	2	30
$\sigma_{pred}/K_1[10^{-4}]$	2.8	5.8	6.1	15
$\sigma_{sim}/K_1[10^{-4}]$	4	6	6.2	20

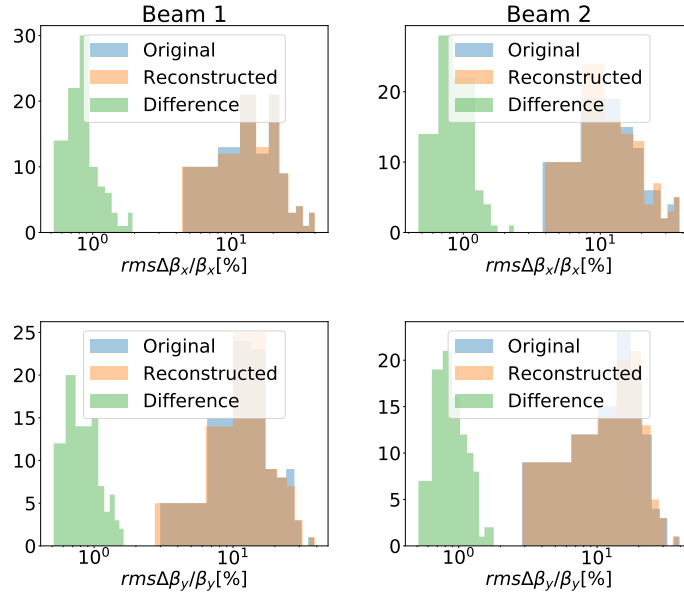


Figure 3.13: The original simulated β -beating compared to the reconstructed β -beating using the magnet errors predicted by Ridge regression model and their differences. The histograms represent rms values of 100 simulations.

where the sample's input has been extracted from, to the β -beating induced by the magnet errors predicted from the given input. Figure 3.13 summarizes the results for 100 validation simulations, per beam and per plane individually. The results demonstrate a very good agreement between the original and reconstructed optics errors, with the average rms difference of 1%.

3.5.2 Ballistic optics

Ballistic optics design implies deactivation of the triplet magnets in the IR1 and IR5. The initial purpose of this optics configuration was the alignment of the triplet quadrupoles Q1, Q2 and Q3. Later, this optics has been extended by switching off also Q4 magnets in order to perform BPM calibration studies [104, 105]. Switching off the triplet magnets which are responsible for the final focusing of the beam before IP makes the beam control especially challenging. Main

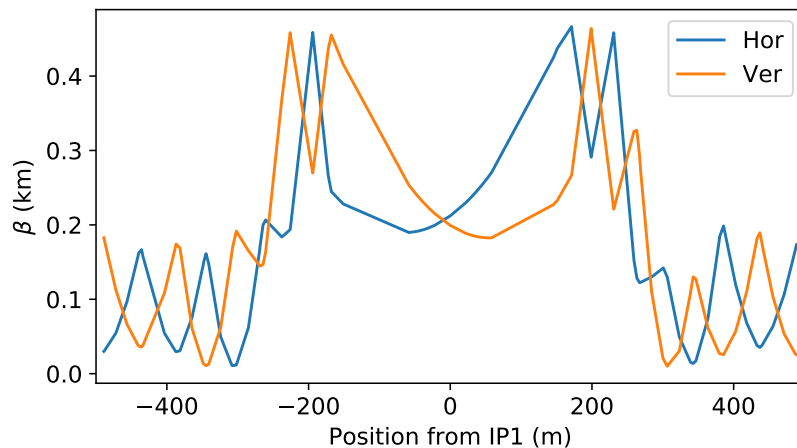


Figure 3.14: β -function around IP1 for the ballistic optics settings used in training data generation. IP5 has equivalent optics configuration.

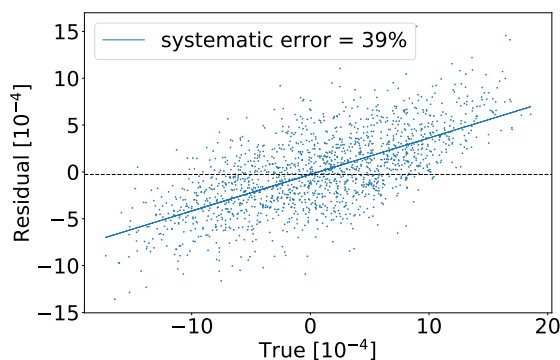


Figure 3.15: Correlation between the true simulated errors in the triplet magnets in IP2 and IP8 with the residual error between the true and predicted values, indicating the systematic bias of the model.

limitation induced by ballistic optics design is the large β -function values caused by the drift space between the quadrupoles, as shown in Fig. 3.14.

Given the importance of this special optics configuration used for alignment and BPM calibration studies, training and validation of ML-models for quadrupolar errors prediction has been performed. The data simulation, model training and testing are executed following the same procedure as for the $\beta^*=40$ cm optics, described above and using the magnet errors distribution listed in Table 3.2. The only difference between the training data used for difference optics settings is that in the case of ballistic optics, the input features do not include the β -function values at the BPMs close to IP1 and IP5 since the triplet magnets in these regions are deactivated.

In order to test the model performance, the residual error of prediction as well as mean and rms values of predicted and true quadrupolar errors in the arcs and the triplets located in IR2

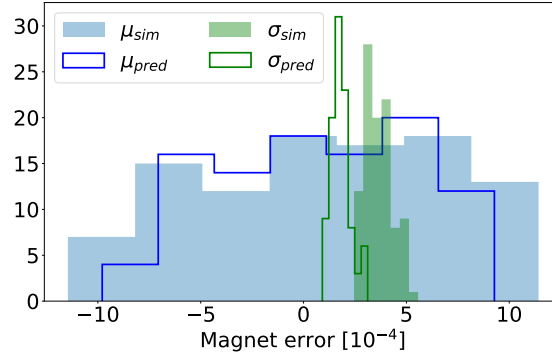


Figure 3.16: Mean and rms values of the full set of relative gradient field errors of all quadrupoles in each of 100 test simulations, compared to the corresponding values predicted from simulated errors in the optics functions.

Table 3.4: Testing ML model on ballistic optics simulation: explained variance R^2 , residual error of prediction relative to the corresponding simulated true values and comparison between predicted and simulated RMS (σ_{pred} , σ_{sim}) values of the relative integrated field errors.

	Triplets	Arcs
R^2	0.65	0.61
$\frac{res}{true}[\%]$	39	32
$\sigma_{pred}/K_1[10^{-4}]$	2.2	15
$\sigma_{sim}/K_1[10^{-4}]$	4	20

and IR8 have been analysed. Figure 3.15 demonstrates the systematic bias of the prediction of the triplet errors in IR2 and IR8. Compared to the results obtained for the optics with $\beta^*=40$ cm, the regression model demonstrates higher residual errors and higher systematic error of prediction - 38% against 16% in case of predicting the individual errors of Q1, Q2 and Q3 magnets. Also, the agreement between the mean and rms values of simulated and predicted triplet errors shown in Fig. 3.16 is lower. Table 3.4 summarizes the results of testing the regression model on ballistic optics data.

While the quality of prediction of triplet errors is significantly reduced compared to the results obtained with $\beta^*=40$ cm optics, the performance of the prediction of arc magnets errors is not affected by change of the optics design. A possible explanation for this observation can be related to the fact that the global effect of the triplet errors on the optics functions is smaller, since the focusing is missing in IR1 and 5. Hence, less importance is given to the triplet errors and the update of the regression coefficients during the training is less sensitive in response to the triplet errors compared to the training on the simulation data with $\beta^*=40$ cm where the triplet errors have the biggest contribution to optics deviations from design.

The results of optics reconstruction using the predicted quadrupole errors for beam 1 and beam 2 are presented in Fig. 3.17, demonstrating high agreement between original simulated β -

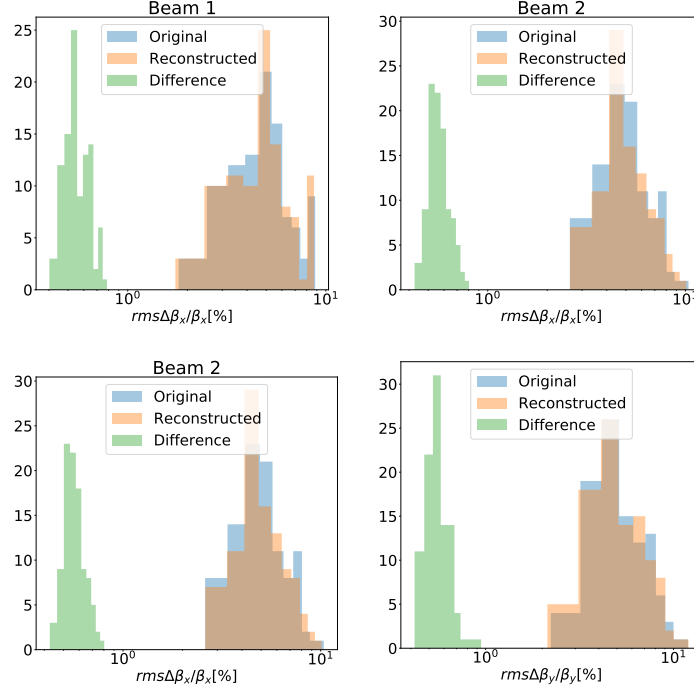


Figure 3.17: The original simulated β -beating using the ballistic optics settings and generated magnet errors compared to the β -beating induced by predicted field errors.

beating and the reconstruction. Despite high error for the prediction of triplet magnets discussed above, the predicted full sets of quadrupole errors in the arcs and IR2 and IR8 triplets reproduce the simulated β -beating within the rms difference of 0.5%.

3.6 Experimental Data

The previously discussed results are obtained from LHC simulations. In this section, the trained regression model is verified against the application on historical LHC measurements data.

3.6.1 Uncorrected optics: LHC commissioning, $\beta^* = 40$ cm

The target is to estimate magnet errors from LHC commissioning data, measured with $\beta^* = 40$ cm, before applying any corrections to the machine. Since β -function at the BPMs next to IPs and normalized dispersion measurement are not present in the corresponding historical data, the model to be applied on the measured data is trained on the phase advance deviations only as input features.

Performing the magnets errors prediction on simulation data, the full set of input features providing the phase advances at every BPM can be used. Due to the cleaning of faulty BPMs prior to the computation of phase advance, missing data points appear in the input features

extracted from experimental LHC data. In the next section, the approach for the reconstruction of missing measurements data, as well as for the denoising of phase advance will be presented in detail. This approach, based on autoencoder neural networks allows to overcome the issue of missing data at the location of faulty BPMs. The output of the autoencoder, i.e. the full set of phase advance deviations is then provided to the regression model to predict the field errors in individual magnets.

The actual magnet errors that generate the measured optics perturbations in the uncorrected machine are unknown such that a comparison between model prediction and corresponding known true values as in the case of simulations is not possible. Instead, we use the predicted errors to simulate optics perturbations and compare this simulation to the actual measurement. The residual difference indicates how well the estimated magnetic errors reflect the actual error sources present in the uncorrected machine. The optics is reconstructed using two different sets of magnet errors: (i) errors predicted from the phase advance deviations replacing the missing values with 0, (ii) errors estimated from the output of autoencoder performing reconstruction of missing values along with denoising. Processing the original phase advance data with autoencoder results in reducing the rms of residual β -beating by 3% and the peak β -beating by 8% with respect to utilizing the original measured phase advances as predictor's input.

The final results on comparing the optics simulated with predicted magnetic errors to the actual measurement in uncorrected machine are demonstrated in Fig. 3.18. The residual β -beating also indicates a potential optics correction that can be achieved by computing correction settings based on predicted errors. The agreement between the measured and reconstructed optics shows that rms β -beating in Beam 1 can be potentially reduced from 12% to 3% and from 54% to 9% in horizontal and vertical planes, respectively. For Beam 2, the rms β -beating decreases from 49% to 15% in the horizontal and from 12% to 2% in the vertical plane. Even though the β -beating for the virgin machine was above 100%, which is significantly larger than the β -beating introduced in the training simulations, the optics reconstruction using the predicted magnet errors results in 7% rms β -beating around the ring and below 3% at the two main experiments CMS and ATLAS.

3.6.2 Ballistic optics

To verify the application of ML-based estimation of quadrupolar errors on the ballistic optics measurements, equivalent approach as for the $\beta^*=40$ cm is followed. The estimation of the magnet errors and the reconstruction of the measured optics deviation from design is carried out for the ballistic optics measurement performed in 2016 on the flattop energy, after implementing coupling corrections, but before applying local and global corrections of β -beating. Apart from having the triplets off in IR1 and 5, another special characteristic of this optics is the reduced total phase advance, such that lowering of the integer part of the tune values to be matched has been required.

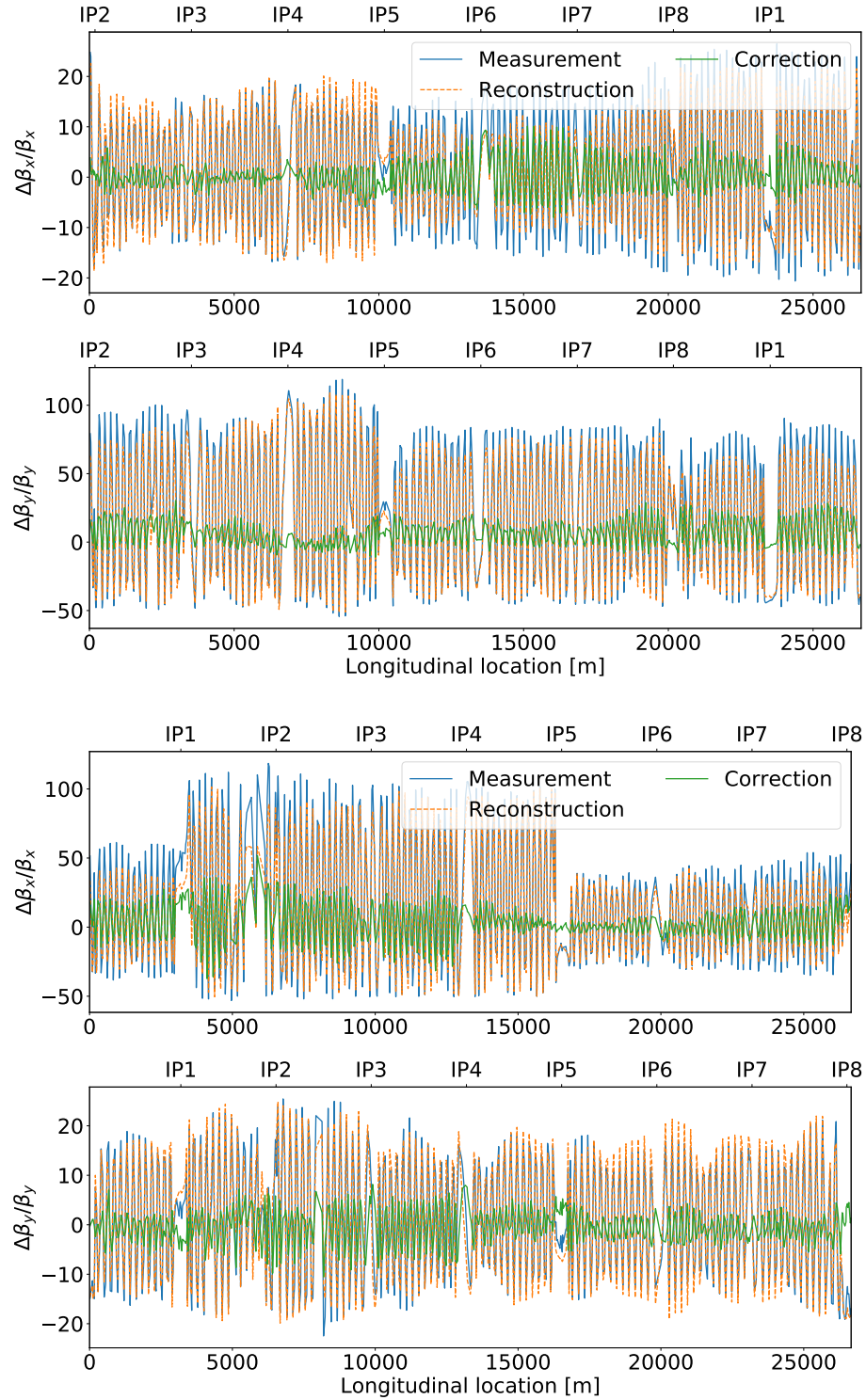


Figure 3.18: Comparison between original β -beating measured in the virgin machine and reconstructed optics produced by predicted quadrupole errors. The difference reflects the potential correction, reducing the average rms and peak β -beating in beam 1 (left) and beam 2 (right) from 32% to 7% and from 72% to 25% respectively.

As in the case of $\beta^*=40$ cm measurements data, also in the ballistic data the measurements of β function next to IR2 and 8 and the normalized dispersion are now available. Hence, a model trained with phase advance only as input features is used. The autoencoder has been re-trained on the ballistic data in order to perform the reconstruction of the model's input variables at the location of faulty BPMs. The comparison between the original measured β -beating and the β -beating obtained from the optics simulation using the predicted quadrupole errors is presented in Fig. 3.19. To be noted is that the measured β -beating is lower compared to the measurements with $\beta^*=40$ cm. In beam 1, the rms of measured β -beating is 5% and 3% in horizontal and vertical planes respectively, the corresponding values obtained from the reconstruction are 5% and 2% demonstrating a great agreement. However, the peak β -beating in the measurement and reconstruction agrees significantly less, namely 18% and 15% measured in both planes for beam 1 against reconstructed peaks of 10% and 8%. A similar comparison is observed in beam 2. Nevertheless, this result shows that the reconstructed quadrupole errors match to a significant level with the optics deviations from design caused by the actual magnet errors present in the machine during ballistic optics measurements.

3.7 Summary

The traditional approaches used for optics corrections in circular accelerators aim to compute correction settings needed to compensate the measured optics deviations from design. Instead, the presented ML-based approach allows determining the actual individual quadrupole field errors currently present in the machine. Building supervised regression models from simulations provides an opportunity to obtain the estimation of the large number of all quadrupolar gradient field errors simultaneously, namely the errors in 32 triplet magnets shared between beam 1 and beam 2, as well as 612 arc magnets for each of the beams. Tests on a large number of simulations demonstrated relative prediction error of 30% and 16% for the triplet and arcs magnet classes respectively, for the optics design with $\beta^* = 40$ cm. For the ballistic optics design, which will be the first machine setting to use in the Run 3 of the LHC, the total relative error of prediction for all magnets is 32%. Moreover, the application of the trained ML-models on the historical measurements data also demonstrated successful verification of the proposed ML-based approach to magnet errors estimation and its potential to reduce the β -beating in an uncorrected machine by providing the knowledge about the local errors which is in particular of a great importance for the triplets.

The advantage of regression models is the ability of extracting an average linear response over the training population instead of only using the unperturbed model and the response of a single observable to a strength change in a single corrector, as it is done in the response matrix approach. The regression models extract the relation between 3304 optics observables and 1256 magnet field errors by updating the regression coefficients based on 80000 training simulations

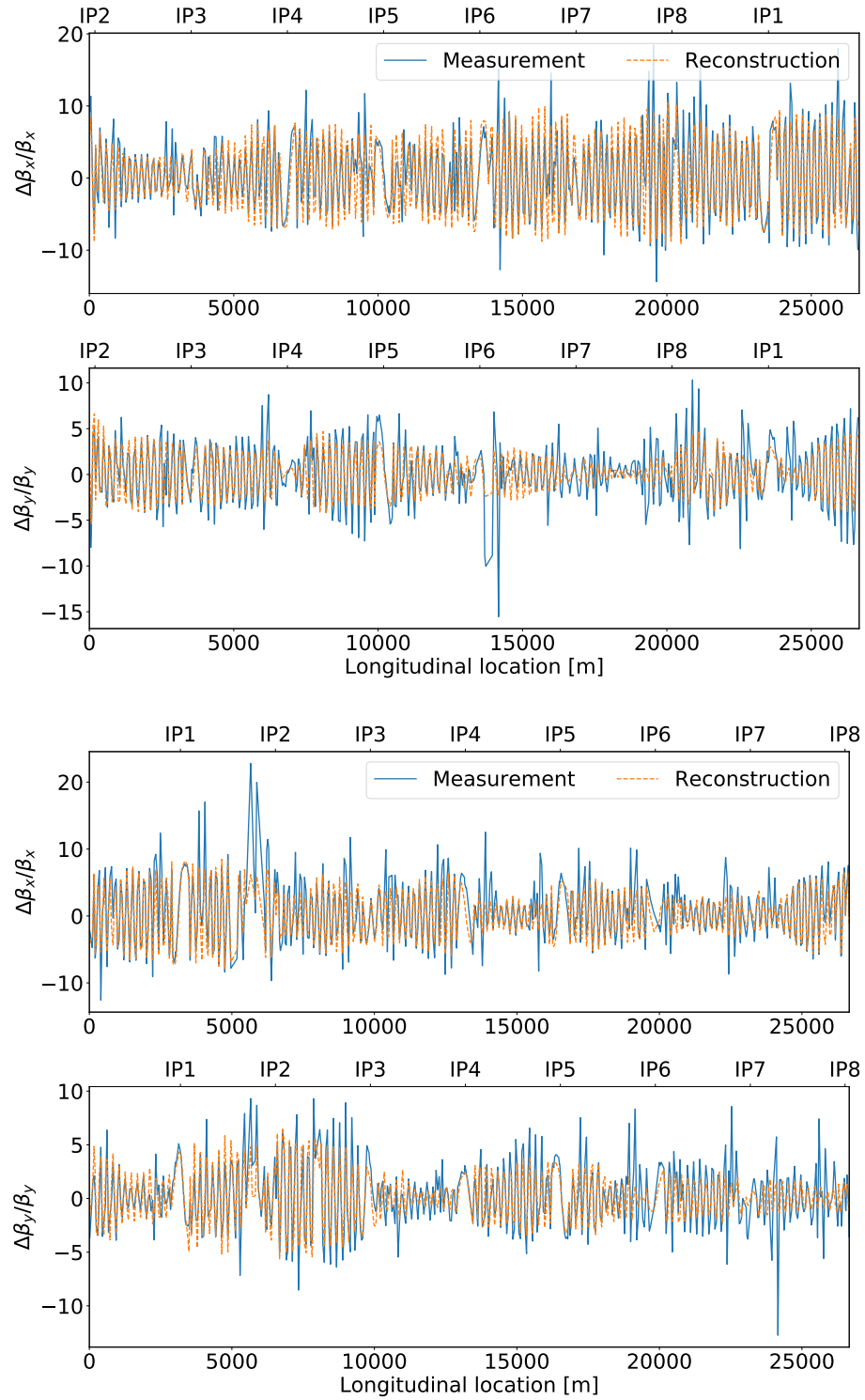


Figure 3.19: Comparison between original β -beating measured with the ballistic optics settings in 2016 and reconstructed optics simulated using the magnet errors predicted with the regression model.

samples. Hence, the weighting factors of specific observables are found automatically instead of defining them manually as needed for the application of currently available correction tools for the LHC.

One of the limitations of supervised models based on linear regression is low prediction accuracy considering the quadrupoles located in the arcs, which can be explained with the fact that the error sources in the arcs exhibit a weaker effect on the optics compared to the triplet magnets located around the IPs. Therefore, it is more difficult to extract the relation patterns between provided model input and corresponding output during the training and produce accurate predictions. Potentially, additional input features which are relevant for arc magnets predictions have to be identified and included into the regression model.

Currently, individual regression models have to be generated for each specific optics design, as it has been presented for the optics with $\beta^* = 40$ cm and ballistic optics settings, since the response of the perturbations introduced in the input features by the magnet errors depends on particular optics. However, it should be possible to allow more general application of created models in the future, by performing the training on the data simulated using different optics settings, reducing the effort for the model preparation in case of optics and machine settings modifications during operation.

The most important future step towards applying the presented approach in LHC operation, will be the computation of actual correction settings to be implemented in the LHC after predicting the individual magnet errors from measured optics. The significant decrease of prediction error after combining individual triplet magnets according to the correlations between these magnets, promises improvements of local corrections around the IPs by using the predicted field errors. Future advances in instrumentation and optics analysis techniques, along with the application of denoising autoencoder presented in the next chapter, will allow significant gain in the accuracy of magnet errors reconstruction and consequently lead to better correction results and optics control.

DENOISING AND RECONSTRUCTION OF OPTICS MEASUREMENTS

Optics measurements data are normally affected by the noise artefacts due to instrumentation imperfections or uncertainties in the applied analysis methods. A special type of semi-supervised neural networks, autoencoders, are widely applied to denoising tasks in image and signal processing as well as in generative modeling. In the following, an approach for denoising of the phase advance measurements obtained from harmonic analysis of turn-by-turn data will be presented. In the previous chapter, the problem of missing data features due the cleaning of faulty BPMs has been addressed with regard to ML-models trained on simulations being applied to the measurements data. In addition to denoising, autoencoder also offers a possibility to reconstruct these missing input features from the original incomplete set of measured phase advance deviations from the nominal model.

In the LHC, most of the optical functions can be obtained from turn-by-turn data recorded at a single beam excitation by a kicker magnet as discussed in [1.4.3](#). However, the analysis of such observables as β -function around the IPs and dispersion function require special dedicated techniques and additional operational time. This chapter presents an alternative approach to estimate these observables using supervised learning and linear regression models, in case direct measurements are not available. Besides providing additional features to the models for the reconstruction of quadrupolar errors, this approach can be advantageous also for the analysis of historical data where the measurements of such observables have not been performed.

The study is presented as following: first, the importance of reducing the noise in the measurements data is demonstrated, followed by the concept of denoising and reconstruction of phase measurements using an autoencoder neural network. The results of applying this approach are demonstrated on simulations and historical experimental data. Further, this chapter presents the reconstruction of normalized dispersion and β -function around the IPs based on the phase advances, by using supervised linear regression models. The simulation results, as well as the

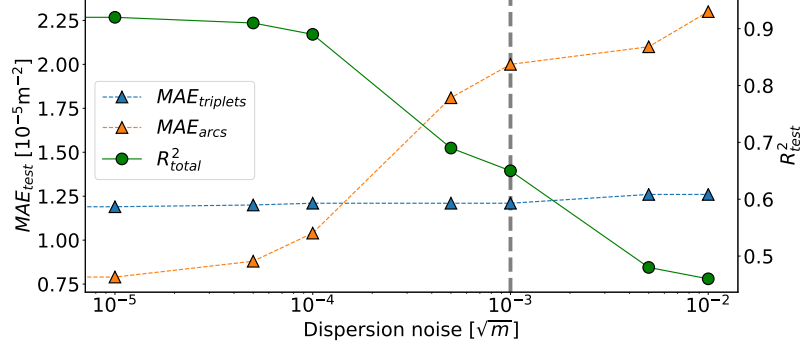


Figure 4.1: The effect of the noise added to normalised dispersion function ($\sigma \frac{D_x}{\sqrt{\beta_x}}$) used in training data simulation on the predictive power of a regression model. Mean absolute error (MAE) of prediction is given in the units of absolute quadrupole errors. R^2 defines the explained variance. The grey line marks the expected noise in the LHC measurements.

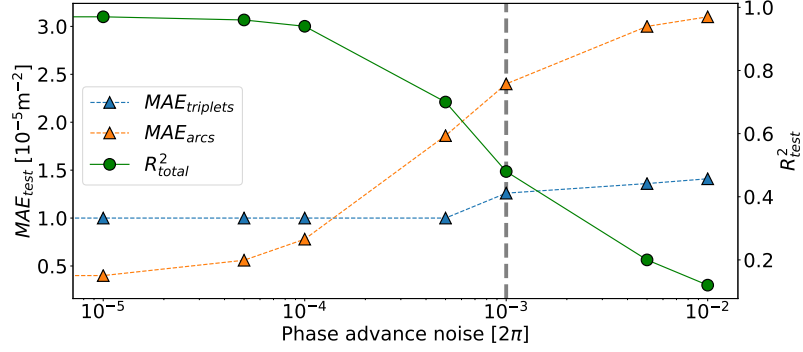


Figure 4.2: The change of model scores trained on phase advance features only, depending on the noise added to simulated phase advance deviations as described in Sec. 3.3.

results obtained from experimental data are given on the example of $\beta^* = 40$ cm optics.

4.1 Noise and magnet errors prediction

In order to investigate the effect of the noise on the predictive power of the regression model for the quadrupole errors reconstruction, the MAE and R^2 scores of regression models trained on the input data given different noise levels are compared. The effect of the noise added to different features groups included in the model input are treated separately - dispersion noise and phase advance noise. Figure 4.1 shows the change in model performance depending on the noise added to the simulated normalized dispersion deviations, while keeping the phase advance noise unchanged. Figure 4.2 presents the results obtained from regression models trained using the noisy phase advance features only, without including normalized dispersion and β -function

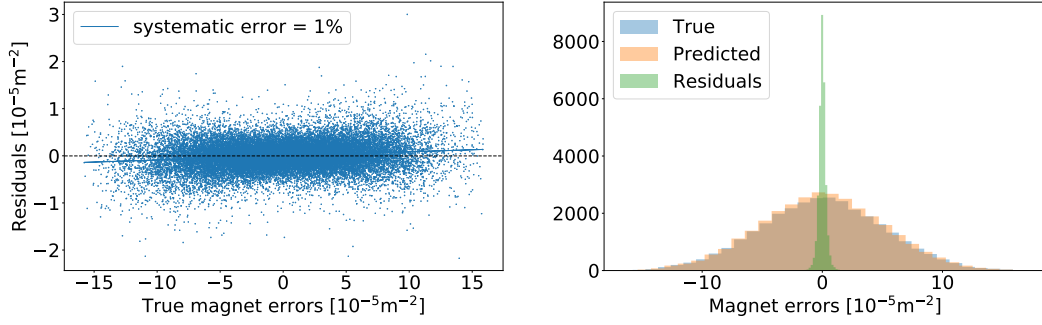


Figure 4.3: Result of the prediction of absolute quadrupolar field errors in the arcs for 100 LHC simulations in the absence of noise. The systematic correlation between the residual error of prediction and actual true simulated values is indicated by the slope on the left. This low systematic regression bias leads to an excellent agreement between simulation and prediction as demonstrated in the histogram on the right.

values as input of the model.

In addition, the noise-free data is also used to verify the accuracy of the triplet and arc magnets error prediction as presented in Sec. 3.5, including the optics reconstruction by applying the predicted errors to ideal lattice. 80000 training and test data samples are generated following simulation set-up described in Sec. 3.3, using additional 100 LHC simulations to present validation results. The absence of the noise in training, as well as in validation input data, leads to the reduction of systematic error of prediction for the arc magnets from 30% to 1%, while for the triplet magnets the systematic error reduces less significantly, from 16% to 12%. Figure 4.3 illustrates the summary of validating the prediction of quadrupole errors in the arcs. This result agrees with the change of ML-model scores in response to increase of the noise, presented in Fig. 4.1 and 4.2, where the quality of arc magnets prediction is observed to be more strongly affected by increase of the noise compared to triplet errors prediction.

Considering the β -beating reconstruction using the predicted quadrupole errors shown in Fig. 4.4, the rms difference between the original simulated and reconstructed β -beating reduces by a factor of 2, compared to the results obtained from errors prediction based on the noisy input data. This result provides another indication for the importance of lowering the noise in the optical functions computed from turn-by-turn data.

To be noted is that the change of the phase advance noise has a bigger impact on the model performance than the noise in the dispersion function. This observation can be explained due to stronger correlation between the magnetic errors and the deviations in the phase advances compared to the correlation with the dispersion perturbation, as the dispersion is suppressed in the quadrupoles located in the straight sections.

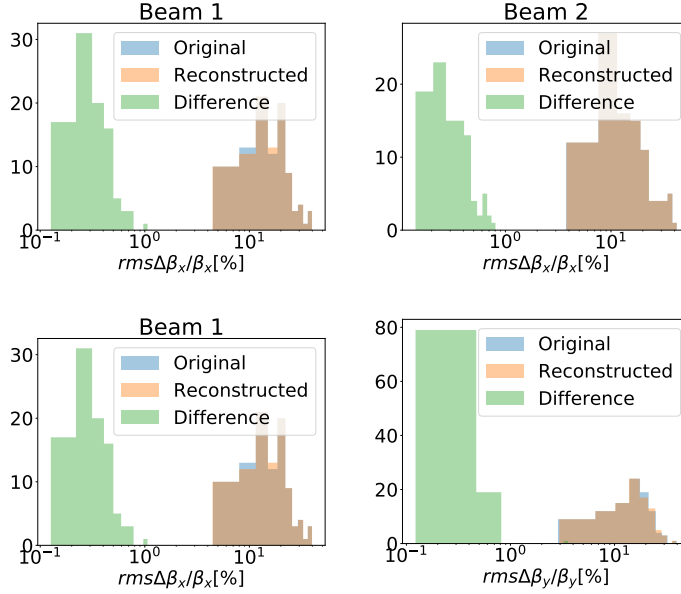


Figure 4.4: The original simulated β -beating using $\beta^* = 40$ cm optics settings and generated magnet errors compared to the β -beating reconstructed with the field errors predicted in the absence of noise.

4.2 Denoising and reconstruction of phase advance measurements using autoencoder

As shown above, keeping the measurements noise as low as possible, as well as including all relevant variables into the training data is crucial for the accuracy of the magnetic error reconstruction. The presence of the noise enforces acquisition of several turn-by-turn measurements for each beam in order to obtain statistically significant computation of the optics functions and uncertainties. Moreover, the signal of identified faulty BPMs is excluded from the optics analysis, and hence the optics observables used as input features of regression model but also for the traditional global corrections based on response matrix, are missing at the locations of faulty BPMs. This section focuses specifically on these two problems and their possible solution.

An ML-based approach to mitigate the effect of the noise and to reconstruct missing data points is the application of autoencoder. This specific type of a neural network is described in detail in 1.3.7. To perform denoising and reconstruction of the original input data, the encoder extracts relevant information from the input by lowering its dimension and filtering the noise. The original input is then reconstructed from the compressed encoded representation of the original data by the decoder. During the training, the encoder learns to recognize the noise patterns in the input and to keep only the signal relevant for the reconstruction performed by the decoder.

In order to apply this approach to denoising and reconstruction of phase advances, the

4.2. DENOISING AND RECONSTRUCTION OF PHASE ADVANCE MEASUREMENTS USING AUTOENCODER



Figure 4.5: Example of reconstructing the phase advance deviations at the locations of simulated missing BPMs in horizontal plane, beam 1 using autoencoder neural network.

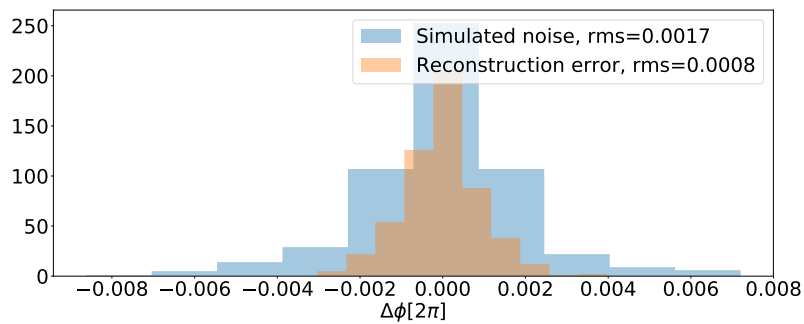


Figure 4.6: Comparison between the noise added to original simulation of phase advance deviations in beam 1, horizontal plane to the reconstruction error from autoencoder prediction, performed on 100 LHC simulations.

autoencoder network is trained on the phase advance deviations produced in the simulations used for the magnet errors reconstruction. The simulated differences between the model and perturbed phase advances are given the noise as described in 3.3 and are utilized as training input of autoencoder. To simulate missing data points in validation data, 10% of input values are omitted in each of the validation samples, according to statistical analysis of BPM discarded by cleaning tools [87]. The original full set of simulated phase advance deviations without noise are the output targets. The results presented in the following are obtained with an autoencoder built using *Keras* library [106], consisting of 4 hidden layers, using MAE as loss function and *Adam* [18] as optimization algorithm.

Exemplary comparison between autoencoder prediction at the location of discarded data points and corresponding original true simulated values is shown in Fig. 4.5 indicating a very good agreement. Processing the phase advance data with autoencoder also allows to reduce the noise by a factor of 2 as demonstrated in Fig. 4.6. The rms error of 100 validation samples, for both beams, horizontal and vertical planes is $0.8 \times 10^{-3}[2\pi]$ which is below the used uncertainty estimate for the phase advance measurements in the LHC. The advantage of applying autoencoder neural network is the possibility to combine two different objectives using one ML technique, namely the reconstruction of missing data and measurements denoising.

The approach for measurements reconstruction using the autoencoder is also verified against the application on measurements data from LHC commissioning in 2016, $\beta^* = 40$ cm. Unlike in simulations, here the true values of missing data points are unknown. Hence, the full set of computed phase advance deviations from ideal optics, instead of the values at the locations of faulty BPMs, is compared to the autoencoder output in order to assert the reliability of prediction. Missing values in the measurements data are replaced by zeros. The high agreement between available measured phase advance deviations and autoencoder-based reconstruction demonstrated in Fig. 4.7, together with the statistical results from simulations, confirm that this approach can produce reliable reconstruction also of the values at the location of faulty BPMs.

4.3 Estimating optics observables using linear regression

Considering the reconstruction of specific optical functions, also linear regression appears as appropriate technique while enabling faster training and simpler model parameter tuning compared to autoencoder.

This section demonstrates the application of Ridge regression to reconstruct the β -function at the BPMs next to IPs and the full set of horizontal normalized dispersion deviations. In both cases, the target output is predicted from the noisy phase advance deviations produced in simulations described in 3.3.

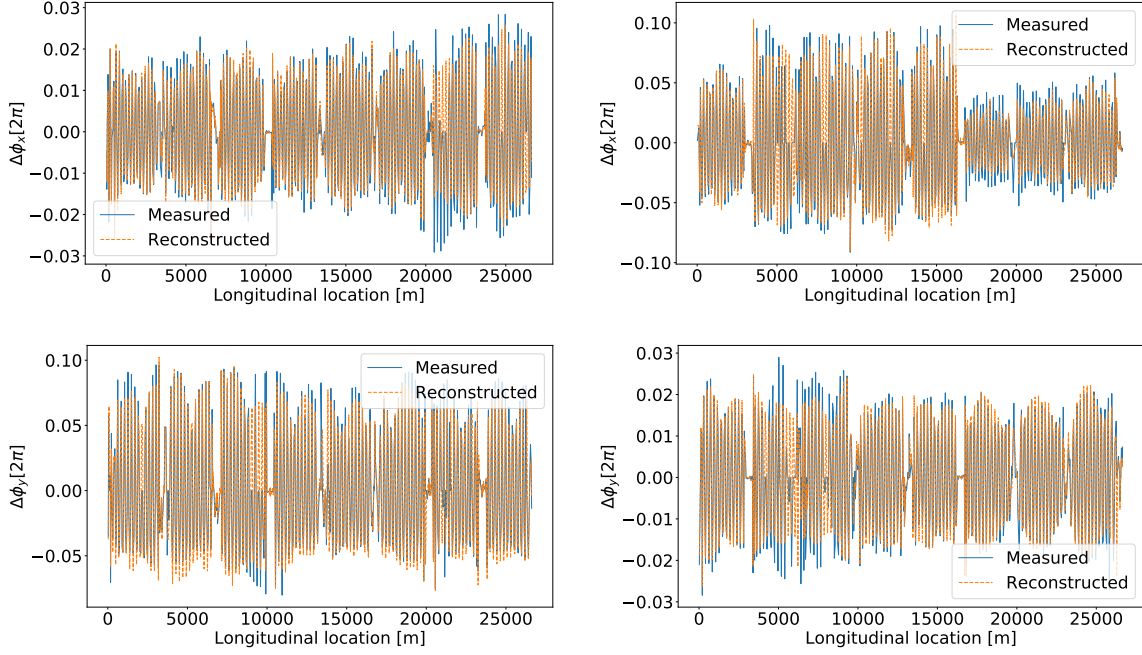


Figure 4.7: Comparison between LHC measurements and reconstruction of phase advance using autoencoder in horizontal and vertical plane for beam 1 (right) and beam 2 (left). The optics analysis provides phase advances at the locations of well-functioning BPMS only. Zeroes in the measurements indicate the locations of faulty BPMS where the measurements are not available and are to be potentially replaced by autoencoder-reconstruction.

4.3.1 Reconstruction of normalized dispersion

In order to obtain the normalized dispersion function, several beam excitations on and off momentum are required. The normalized dispersion $\frac{D_x}{\sqrt{\beta_x}}$ is used as an observable which does not depend of calibration, as it is calculated as a ratio between two calibration-dependent quantities, canceling the effect of calibration factor.

Reconstructing the normalized dispersion measurements from phase advance deviations in one step can save operational time by avoiding numerous measurements acquisitions on and off momentum. Ridge regression model is trained on 8000 samples generated with the set-up described in Sec. 3.3. Here, individual regression models are built for each beam separately and hence, the input consists of noisy phase advance deviations in horizontal and vertical planes of one beam only. Figure 4.8 shows the comparison between normalized dispersion function predicted by regression model and corresponding actual simulated function. The relative rms prediction error of 1000 simulations is 7%.

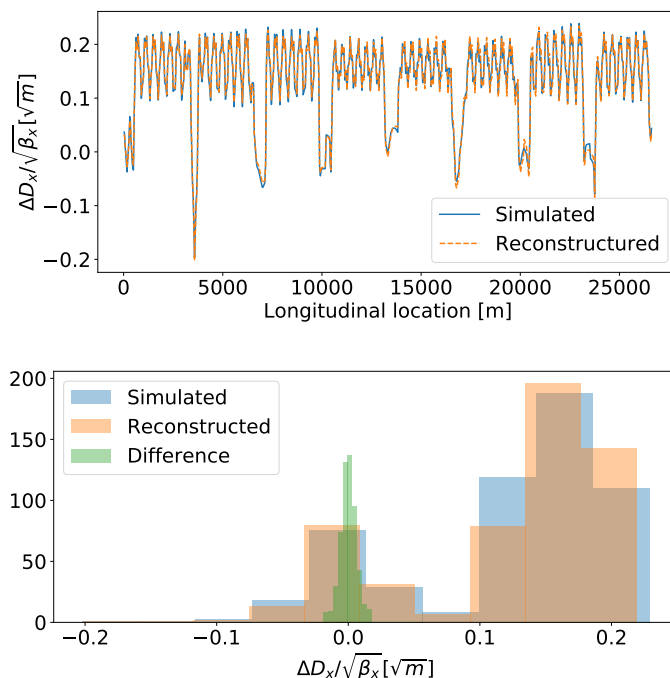


Figure 4.8: Reconstruction of horizontal normalized dispersion deviations in beam 1 from noisy phase advance data. The upper plot illustrates the agreement between simulated and reconstructed values in one simulation, while the histogram on the bottom shows statistics obtained from 1000 simulations.

4.3.2 Reconstruction of β -function in Interaction Regions

The β -function directly at the IP is typically computed in the LHC using the k -modulation technique [53], which also produces accurate β measurements at the BPMs around the four main experiments that are included in the input of regression models for magnetic errors estimation. These observables have been found to be important for the estimation of quadrupolar field errors, however they are not always available in the measurements data.

Results on simulations

The result of reconstructing the β -beating at the BPMs next to the IPs is presented in Fig. 4.9. Ridge regression model is trained on 8000 training samples consisting of simulated noisy phase advance deviations for both beams and planes as the input and β deviations in horizontal and vertical planes simulated at the BPMs left and right next to the IPs 1, 2, 5 and 8 being the output. The validation is then performed on a set of 1000 LHC simulations. The error of reconstruction defined as the difference between simulated and reconstructed values relative to simulated β is 1% which is comparable to the uncertainty of k -modulation technique for the β^* measurements with $\beta^*=40$ cm [107].

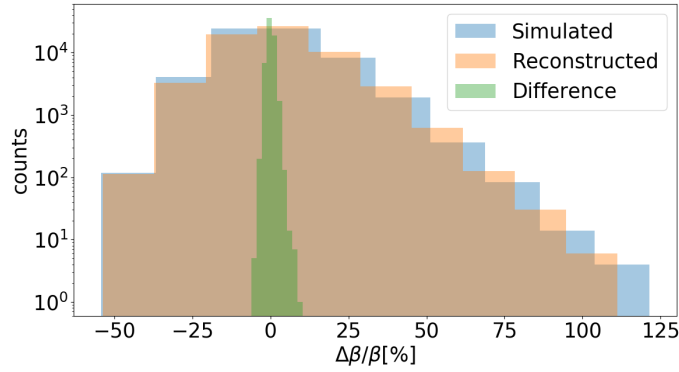


Figure 4.9: Comparison between simulated and reconstructed β -beating in both planes and for both beams at the BPMs next to the main IPs for 1000 simulations, demonstrating the relative rms difference of 1%.

Reconstruction of LHC commissioning data

In order to verify the ability of linear regression model to predicted β functions in IRs from phase advances on realistic LHC data, the measurements taken in uncorrected machine in 2016 with $\beta^* = 40$ cm are used. As in the case of predicting the quadrupolar errors using the measurements data as input features, also here the missing phase advance needed to be included into the input of the model trained on simulations are replaced by the phase advance deviations predicted by autoencoder.

Since k-modulation measurements are not available in this data, the prediction of ML-model is compared to the values computed by the N-BPM method, where the β -function is computed from the phase advances between different BPMs combinations. Both, regression model and N-BPM method use the phase advances to reconstruct the β -function. However, the calculation of β from phase using N-BPM method is known to be less reliable in the IRs. Fig. 4.10 shows the β -function values at the BPMs next to the four IPs, left and right, computed for both beams in horizontal and vertical planes (32 values in total). The comparison between the measured values computed using N-BPM method and the reconstruction predicted by regression model trained on noisy phase advance simulations shows that ML-model is capable to make prediction on unseen operational data.

4.4 Summary

Reducing the noise in the measurements of optics observables used as input for magnetic errors estimation can potentially improve the prediction accuracy. Especially in terms of arc magnets errors, the noise becomes the biggest limitation for the presented ML-based estimation of individual quadrupolar field errors. These results can be potentially used as future requirements

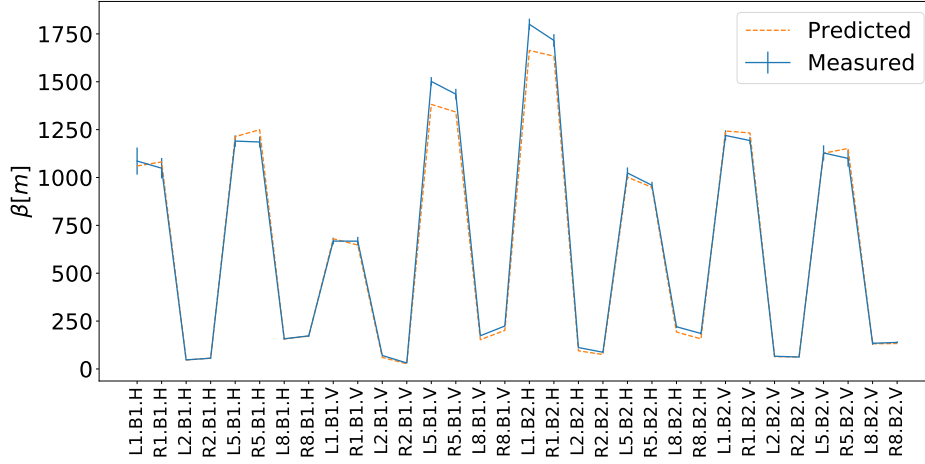


Figure 4.10: Comparison between measured and predicted β -function at BPMs left and right from IP1, 2, 5 and 8, including the values for both beams, horizontal and vertical planes.

on the instrumentation and analysis techniques used to obtain optics functions from the turn-by-turn measurements data.

The application of autoencoder neural network has been found to be a promising solution for denoising and the reconstruction of phase advance measurements at the location of faulty BPMs, improving the quality of the input data provided to magnet error prediction models. As shown on simulations, the noise can be reduced by a factor of 2, while the resulting relative rms error of reconstruction the phase advance function is 20% lower than the phase advance noise estimated as uncertainty in the measurements at the LHC. The comparison between measured β from phase at the BPMs next to IPs and ML-model reconstruction, together with the results obtained on simulations demonstrating the accuracy of 1%, which is equivalent to the accuracy of k-modulation technique. This result provides a strong evidence that the presented supervised learning based approach can be used for the estimation of β -function when direct measurements are not available. Although the reconstruction of normalized dispersion is verified on simulations only, the obtained accuracy of the reconstruction obtained by the prediction of supervised trained regression model is 7%, promising a successful application on experimental data.

The possibility to reconstruct normalized dispersion in one step instead of handling several beam excitation currently needed for the dispersion function computation and to obtain β around the IPs faster than by performing k-modulation, demonstrates how optics analysis can benefit from including supervised learning to the analysis tools. It allows not only to include these observables into corrections computation even if dedicated measurements needed to obtain these optical functions have not been performed, but also opens the perspective to extend offline optics analysis and to speed up optics measurements saving the costly operational time or by providing additional data for offline optics analysis.

CONCLUSION

5.1 Achieved Results

How can we benefit from Machine Learning in the field of Accelerator Physics? The answer to this question can be found in numerous attempts to apply ML ranging from surrogate modeling and simulations to online feedback systems and operation automation conducted across different laboratories. Although quantitative goals such as speed up or full automation of particular operational tasks has been achieved, a systematic methodology for the integration of ML algorithms into the domain of accelerator physics, describing how to approach a typical accelerator related problem as a ML task was not present. The result of this thesis is not only the demonstration of new methods for beam optics measurements and corrections based on supervised and unsupervised learning, but most importantly, this work presents a general methodology of integrating ML techniques into a new field in the domain of accelerator physics. This methodology can be potentially used in order to investigate on further potential ML-based methods for particle accelerators, since it provides a path to identify appropriate techniques from the big variety of existing algorithms, to evaluate the designed solutions and to implement and integrate them enabling the most efficient employment of ML concepts and their advantages.

The described methodology allowed to mitigate some limitations of traditional techniques and to develop new ML-based methods for the following tasks in the field of beam optics measurements and corrections in storage rings:

- I. **Automatic identification of instrumentation faults such as anomalous signal from faulty Beam Position Monitors, prior to optics analysis.**

The newly developed unsupervised learning method based on Isolation Forest (IF) algorithm for anomalies detection has been successfully incorporated into the optics analysis to

automatically remove faulty signal prior to the computation of the optics function avoiding manual data cleaning and repeating the analysis procedure. Overall, the application of the new method together with optimizing the settings of the previously existing cleaning tools allowed to reduce the amount of outliers in the optical functions by a factor of 2 to less than 1% of the total number of available BPMs as shown in simulations.

Moreover, it has been demonstrated on experimental LHC data that the new method is capable of identifying BPM faults exhibiting a-priori unseen failure patterns. After becoming fully operational, the new IF-based method has been used as a standard part of optics analysis to improve the quality of measurements data. The averaged results obtained from the latest LHC measurements show that IF allows to prevent the appearance of 80% of faulty signals otherwise remaining in the data and contaminating the optics analysis and consecutive corrections. The improved understanding of the combination of SVD cleaning and IF algorithms obtained in this work will further benefit LHC measurements in the future, especially in the light of instrumentation upgrades implemented for the LHC Run III.

II. Simultaneous reconstruction of more than a thousand of quadrupolar gradient field errors based on the measured deviations of beam optics observables from design.

Regression models based on supervised learning approach allow to include a large amount of input variables (3304 optics observables) and 1256 quadrupolar gradient field errors being output targets, in order to obtain optics corrections for both beams and all machine regions simultaneously by reconstructing the individual quadrupolar errors present in the machine. Verification on the optics with $\beta^* = 40$ cm demonstrated relative error of regression model's prediction of 30% and 16% for the triplet and arc magnets respectively. The application of a combined model based on several Ridge Regressors to experimental data showed that adding the predicted magnet errors to the ideal optics design leads to an accurate simulated reconstruction of original measured β -beating, with an rms difference of 7% in total and below 3% at the two main experiments. These results indicates also the potential residual β -beating after implementing the predicted magnet errors as correctors settings.

The newly developed approach for quadrupolar errors reconstruction has been successfully demonstrated on simulations and historical LHC data using two different optics settings, namely $\beta^* = 40$ cm and ballistic optics. The most important future step will be an experimental demonstration of the developed ML-based optics corrections in the LHC Run III, as well as the integration of the method into the operational optics analysis software infrastructure.

III. Denoising and reconstruction of synchrotron phase advance measurements obtained from the FFT-analysis of Beam Position Monitors signal.

The reduction of the noise by a factor 2 as shown on simulations has been achieved by processing the phase advance data with an autoencoder neural network. In this application, the optics analysis benefits not only of the ability of autoencoders to denoise the data provided as input, but also from the possibility to simultaneously reconstruct the missing signal. By performing supervised training on simulations consisting of a full set of phase advance values at every BPM location, it is possible to obtain a reliable reconstruction of phase advance at the location of faulty BPMs removed by cleaning tools from the real measurements data. Moreover, it will allow a direct application of regression models on the measurements data to obtain the individual quadrupolar errors, instead of training a new ML-model for each measurements set using input variables only from currently available BPMs.

This approach has been verified on historical LHC measurements data, proving the results obtained on simulations and demonstrating high agreement between measured and predicted values. The accurate reconstruction of missing phase advance values along with the reduction of the noise promises improvements on the quadrupolar error reconstruction since the performance of regression models highly depends on the noise in the input variables. Generally, reducing the noise level in the phase advance measurements which are the basis of the advanced optics analysis and correction computation will open doors for more improvements on analysis precision and quality.

IV. Estimation of the optical functions values when direct measurements of specific observables are not available.

The application of linear regression models allows to obtain estimates of the quantities which computation otherwise requires additional beam excitation as in the case of dispersion function or performing advanced optics analysis techniques as k-modulation in order to measure β^* . In this study, separate supervised regression models are used in order to enable fast and precise estimation of two different sets of important optics observables: horizontal normalized dispersion in the entire machine, producing 1024 output values and 32 values of β at the BPMs close to experiments. In both cases, a regression model estimates its output from the phase advances which are available in any measurement set through harmonic analysis of turn-by-turn data. The approach is verified on a large number of LHC simulations, providing the accuracy of 7% for the reconstruction of dispersion function and 1% for the reconstruction of β -function values, which is comparable to the performance of k-modulation technique currently applied at the LHC.

Apart from incorporating these observables as input variables into the regression models for the estimation of individual quadrupole errors, also the traditional corrections methods as response matrix can potentially benefit from an increased number of variables to be included into the computation of correctors settings.

The new methods have been implemented specifically for the application on the LHC and they are partially integrated into dedicated optics analysis tools, after extensive verification on realistic simulations as well as on historical LHC data available from the past optics measurements. The achieved results demonstrate the capability of ML to cope with large and highly complex systems as the LHC involving several thousands of independent variables and prediction targets. Apart from the practical benefits, such as reduction of manual effort and potential speed up of optics corrections, the presented study also opened up new research questions which are discussed in Sec. 5.3.

5.2 Discussion

The existing traditional techniques for common tasks in accelerators research and technology such as machine performance optimization and design and modeling of the beam dynamics have demonstrated successful applications, however they exhibit certain limitations. Firstly, the application of traditional tools can become problematic, when the given problem deals with an a priori unknown behaviour or in the case of modeling large systems with thousands of free parameters and optimization targets. Even harder problems arise when the objective functions are undetermined and analytical solutions are not available. Machine Learning appears as a natural support to mitigate such limitations, due to the ability to learn and to construct arbitrary models directly from given examples, without requiring predefined function and explicit rules.

As a matter of course, there are certain types of techniques which are more or less suitable for a particular accelerator related task based on their design or traditional fields of application in other research or industry areas. At the same time, there is a need to identify which problems and limitations in the field of accelerator physics can be solved by using ML efficiently, trading off the efforts of implementing new solutions and rewarding advantages. Table 5.2 presents a brief overview on the potential or already solved accelerator problems and corresponding ML methods, their benefits and related critical considerations. Several different ML techniques can be considered as applicable to one specific problem. The presented overview offers a suggestion on how a choice of appropriate technique can be made, however in practice this decision mostly depends on available expertise, technical infrastructure and data resources. At the same time, a particular algorithm can be useful to solve several tasks even across different accelerator facilities, which makes a standardized approach to develop ML solutions highly important in light of future developments.

Considering the technical challenges of incorporating ML into accelerators, data management becomes critical especially for the supervised learning algorithms since the success of offline trained models in actual online operation strongly depends on the amount and quality of the training data. In the light of this challenge, unsupervised learning appears to have more potential for the problems which allow its application, since direct processing of the current

Problem	ML methods	Benefits	To be considered
Automation of machine operation [81, 108–110]	Reinforcement Learning, Bayesian optimization, Gaussian Process, Adaptive Feedback	Simultaneous optimization targeting several beam properties, allows faster tuning offering more user time.	Ensuring that all important properties are included as optimization targets.
Detection of anomalies (see Sec. 2)	Unsupervised Learning, Classification	Preventing faults before they appear without defining data-specific thresholds, no training is needed and can be directly applied online.	In unsupervised methods, no ground truth is available, so methods should be verified on simulations.
Computationally heavy, slow simulations [78, 111]	Supervised Learning (linear models or NN for non-linear problems)	Learning underlying physics directly from the data, faster execution	Performance depends on realistic simulations for the training data.
Reconstruction of non-measurable quantities (see Sec. 4.3, [70, 112])	Supervised Learning and Regression	Learning underlying physics directly from the data, faster execution	Measurements or high-fidelity simulations needed to evaluate results.
Identification of relevant observables or machine parameters [113, 114]	Clustering, Feature Importance Analysis using Decision trees	Speed up of available methods, streamlined models	Parameter selection and combination (feature engineering) can have significant impact on ML methods performance.
Measurements noise reduction (see Sec. 4.2)	Autoencoder Neural Networks	Improved data quality	Significant information should not be removed from the signal.

Table 5.1: Examples of possible tasks which can be efficiently solved by applying ML and corresponding suggestion of suitable techniques, potential benefits and critical considerations.

data is possible, without requiring historical measurements or simulations. However, also in the case of unsupervised learning a well-structured representation of the data is crucial for robust solutions.

The complexity of the given task and underlying physical processes is another important consideration when building ML based solutions. While non-linear systems require the incorporation of e.g. deep neural networks with custom architecture, many processes in accelerator design and operation can be streamlined to simpler models requiring less resources and being easier to interpret and control compared to complex models. Hence, it is advantageous to build ML solutions for particular machine components or processes, rather than attempting to design an integrated precise model of an entire system.

Machine learning as a toolbox of various concepts and algorithms offers certain benefits,

however it is crucial for its successful application to choose the most appropriate technique and to integrate it into the solution of a well-defined problem. Domain knowledge is the central requirement of approach to ML-based solutions, since it is essential for appropriate selection and validation of ML models, and more importantly, it is also the central component of understanding why particular methods shows specific performance, leading to correct interpretation of the results obtained from ML algorithms. Thus, ML algorithms are not suitable to be applied "out of the box", a special systematic approach has to be developed in every new domain of application, which unavoidably requires domain knowledge.

5.3 Outlook

The new techniques for optics measurements and corrections presented in this thesis have potential for further investigations in terms of possible improvements or alternative ML concepts to be applied. Regarding BPM faults detection, one potential research question is to study the actual reasons of the faults, which currently remain unclear e.g by extracting common hidden patterns in the signal from the BPMs frequently appearing as faulty. This can be performed by applying clustering methods or feature importance analysis by decision trees. The reconstruction of magnet errors using supervised learning highly depends on how well the training data correspond to the real machine settings and measurements. One way to partially overcome this limitation is to experimentally perform a dedicated study by changing the strengths of individually powered quadrupoles and measuring the corresponding change in the optics functions, such that these measurements can be included into the training data set along with simulations, allowing to create more realistic regression models and hence, more accurate prediction of quadrupolar errors. One of the future objectives is to translate the predicted errors of individual magnets into correction circuits settings. Alternative way is to implement reinforcement-learning based corrections computation, where the model will directly provide the required strength change in the correctors circuits, which will not offer the possibility to extract the actual magnet imperfections present in the machine at the time point of the measurements, however this approach will potentially allow a fast computation of the corrections for the entire machine and both beams simultaneously in just one step.

Considering the application of ML in accelerators from the global perspective, more potentially useful but not yet widely used methods and concepts exist, such as transfer learning or generative modelling. Transfer learning is a technique which allows to train a model on one problem domain and apply it on another task after re-training on significantly smaller amount than used in the initial training, offering the possibility for real-time application in changing systems, e.g. by transferring the data recorded online in operation into previously trained model on different machine settings, taking advantage from both, previous efforts and currently available data. Such methods as Generative Adversarial Networks and Variational Autoencoders which are

widely used in the areas of image, music and text generation could be also of a great interest for simulation studies. In light of recent technical advances in hardware implementation of neural networks on special processing units such as TPUs (Tensor Processing Unit) and FPGAs (Field Programmable Gate Array), these concepts have a potential to allow faster training of ML predictors and modelling of complex processes and systems which are the subjects of most of the modern accelerator physics problems.

BIBLIOGRAPHY

- [1] N. Walker, J. Irwin, and M. Woodley, *Global tuning knobs for the SLC final focus* (1993).
- [2] J. Flanagan, K. Oide, N. Akasaka, A. Enomoto, K. Furukawa, T. Kamitani, H. Koiso, Y. Ogawa, S. Ohsawa, and T. Suwada, *A simple real-time beam tuning program for the kekb injector linac* (1999).
- [3] Y. Funakoshi et al., *Performance of KEKB with crab cavities* (2008), pp. 1893–1897.
- [4] W. Fischer, J. Beebe-Wang, Y. Luo, S. Nemesure, and L. Rajulapati, *RHIC proton beam lifetime increase with 10-and 12-pole correctors* (2020).
- [5] M. Aiba, M. Boege, N. Milas, and A. Streun, *Ultra low vertical emittance at SLS through systematic and random optimization*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **694** (2012), p. 133–139.
doi:10.1016/j.nima.2012.08.012.
- [6] X. HUANG, J. Corbett, J. Safraneck, and J. Wu, *An algorithm for online optimization of accelerators*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **726** (2013).
doi:10.1016/j.nima.2013.05.046.
- [7] T. M. Mitchell, *Machine Learning*, McGraw-Hill, Inc., New York, NY, USA, 1 ed., 1997.
- [8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, USA, 2001.
- [9] T. Lai, H. Robbins, and C. Wei, *Strong consistency of least squares estimates in multiple regression II*, Journal of Multivariate Analysis **9** (1979), pp. 343 – 361.
doi:https://doi.org/10.1016/0047-259X(79)90093-9.
- [10] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, New York, NY, USA, 2014.
- [11] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review (1958).

- [12] C. Bishop, *Neural networks for pattern recognition*, Oxford University Press, USA, 1995.
- [13] R. M. Rifkin and R. A. Lippert, *Notes on regularized least-squares*, tech. rep., 2007.
- [14] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological) **58** (1996), pp. 267–288.
url: <http://www.jstor.org/stable/2346178>.
- [15] A. Cauchy, *Méthode générale pour la résolution des systemes d'équations simultanées*, Comp. Rend. Sci. Paris **25** (1847).
- [16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Neurocomputing: Foundations of research*, MIT Press, Cambridge, MA, USA, 1988.
- [17] J. Duchi, E. Hazan, and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, J. Mach. Learn. Res. **12** (2011).
- [18] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, CoRR **1412.6980** (2014).
- [19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, 2016.
Book in preparation for MIT Press.
url: <http://www.deeplearningbook.org>.
- [20] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Statistics/Probability Series, Wadsworth Publishing Company, Belmont, California, U.S.A., 1984.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python*, Journal of Machine Learning Research **12** (2011).
- [22] T. Dietterich, *Ensemble methods in machine learning*, 01 2000, pp. 1–15.
- [23] L. Breiman, *Random forests*, Mach. Learn. **45** (2001).
- [24] R. Schapire, *The strength of weak learnability*, vol. 5, 01 1990, pp. 28–33.
doi:10.1109/SFCS.1989.63451.
- [25] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering: a review*, ACM computing surveys (CSUR) **31** (1999).
- [26] B. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster analysis*, Wiley, 5th ed., 2011.
- [27] S. Lloyd, *Least squares quantization in pcm*, TIT **28** (2006), pp. 129–137.

- [28] M. Ester, H. Kriegel, and X. Sander, J. and Xu, *A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, AAAI Press, 1996.
- [29] M. Breunig, H.-P. Kriegel, R. Ng, and J. Sander, *LOF: Identifying density-based local outliers.*, vol. 29, 06 2000, pp. 93–104.
doi:10.1145/342009.335388.
- [30] X. Jin and J. Han, *Expectation Maximization Clustering*, Springer US, Boston, MA, 2010, pp. 382–383.
url: https://doi.org/10.1007/978-0-387-30164-8_289, doi:10.1007/978-0-387-30164-8_289.
- [31] G. E. Hinton and J. L. McClelland, *Learning representations by recirculation*, in Neural Information Processing Systems, American Institute of Physics, 1988.
- [32] *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008), p. S08003. 437 p.
Also published by CERN Geneva in 2010.
url: <http://cds.cern.ch/record/1129811>.
- [33] *The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment*, JINST **3** (2008), p. S08004. 361 p.
Also published by CERN Geneva in 2010.
url: <http://cds.cern.ch/record/1129810>.
- [34] *The LHCb Detector at the LHC*, JINST **3** (2008), p. S08005.
Also published by CERN Geneva in 2010.
url: <https://cds.cern.ch/record/1129809>, doi:10.1088/1748-0221/3/08/S08005.
- [35] *The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment*, JINST **3** (2008), p. S08002. 259 p.
Also published by CERN Geneva in 2010.
url: <https://cds.cern.ch/record/1129812>.
- [36] A. Wolski, *Beam Dynamics in High Energy Particle Accelerators*, Imperial College Press, London, 2014.
url: <http://www-spires.fnal.gov/spires/find/books/www?cl=QC787.P3::2014>.
- [37] S. Y. Lee, *Accelerator Physics*, WORLD SCIENTIFIC, 3rd ed., 2011.
url: <https://www.worldscientific.com/doi/abs/10.1142/8335>, arXiv:<https://www.worldscientific.com/doi/pdf/10.1142/8335>, doi:10.1142/8335.

- [38] M. G. Minty and F. Zimmermann, *Measurement and control of charged particle beams*, Particle acceleration and detection, Berlin, Springer, 2003.
- [39] T. Persson and R. Tomás, *Improved control of the betatron coupling in the Large Hadron Collider.*, Phys. Rev. Accel. Beams **17** (2014), p. 051004.
- [40] E. H. Maclean, F. Carlier, S. Fartoukh, T. Persson, P. Skowronski, R. Tomás, and D. Wierichs, *Demonstration of coupling correction below the per-mil limit in the LHC* (2016).
url: <https://cds.cern.ch/record/2210530>.
- [41] R. Tomás, M. Aiba, A. Franchi, and U. Iriso, *Review of linear optics measurement and correction for charged particle accelerators*, Phys. Rev. Accel. Beams **20** (2017).
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.054801>.
- [42] T. Persson, F. Carlier, J. Coello de Portugal, A. Garcia-Tabares Valdivieso, A. Langner, E. H. Maclean, L. Malina, P. Skowronski, B. Salvant, R. Tomás, A. C. García Bonilla, *LHC optics commissioning: A journey towards 1% optics control*, Phys. Rev. Accel. Beams **20** (2017).
url: <https://cds.cern.ch/record/2272377>.
- [43] F. S. Carlier, J. Coello De Portugal, S. Fartoukh, E. Fol, D. Gamba, A. Garcia-Tabares, M. Giovannozzi, M. Hofer, A. S. Langner, E. H. Maclean, L. Malina, L. E. Medina Medrano, T. H. B. Persson, P. K. Skowronski, R. Tomás Garcia, F. Van Der Veken, and A. Wegscheider, *Optics Measurements and Correction Challenges for the HL-LHC* (2017).
url: <https://cds.cern.ch/record/2290899>.
- [44] X. Shen, S. Y. Lee, M. Bai, S. White, G. Robert-Demolaize, Y. Luo, A. Marusic, and R. Tomás, *Application of independent component analysis to ac dipole based optics measurement and correction at the relativistic heavy ion collider*, Phys. Rev. ST Accel. Beams **16** (2013), p. 111001.
url: <https://link.aps.org/doi/10.1103/PhysRevSTAB.16.111001>, doi:10.1103/PhysRevSTAB.16.111001.
- [45] C. Bovet, J. Papis, H. Schmickler, and L. Vos, *LHC BPM design* (1997).
- [46] R. Bartolini and F. Schmidt, *A Computer Code for Frequency Analysis of Non-Linear Betatron Motion*, Tech. Rep. SL-Note-98-017-AP, CERN, Geneva, Feb 1998.
url: <https://cds.cern.ch/record/702438>.
- [47] L. Malina, J. M. Coello de Portugal, J. Dilly, P. Skowronski, R. Tomás, and M. Toplis, *Performance Optimisation of Turn-by-Turn Beam Position Monitor Data Harmonic Analysis*, in 9th International Particle Accelerator Conference, 2018, p. THPAF045.

doi:10.18429/JACoW-IPAC2018-THPAF045.

- [48] R. Calaga and R. Tomás, *Statistical analysis of RHIC beam position monitors performance*, Phys. Rev. ST Accel. Beams **7** (2004).
doi:10.1103/PhysRevSTAB.7.042801.
- [49] P. Castro, *Luminosity and beta function measurement at the electron-positron collider ring LEP*, PhD thesis, CERN-SL-96-070-BI, 1996.
Presented on 25 Nov 1996.
url: <https://cds.cern.ch/record/316609>.
- [50] A. Langner, G. Benedetti, M. Carlà, U. Iriso, Z. Martí, J. C. de Portugal, and R. Tomás, *Utilizing the n beam position monitor method for turn-by-turn optics measurements*, Phys. Rev. Accel. Beams **19** (2016), p. 092803.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.19.092803>, doi:10.1103/PhysRevAccelBeams.19.092803.
- [51] A. Wegscheider, A. Langner, R. Tomás, and A. Franchi, *Analytical N beam position monitor method*, Phys. Rev. Accel. Beams **20** (2017), p. 111002.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.111002>, doi:10.1103/PhysRevAccelBeams.20.111002.
- [52] A. García-Tabarés Valdivieso and R. Tomás, *Optics-measurement-based beam position monitor calibrations in the LHC insertion regions*, Phys. Rev. Accel. Beams **23** (2020), p. 042801.
- [53] M. Kuhn, V. Kain, A. Langner, and R. Tomás, *First K -Modulation Measurements in the LHC During Run 2*, in Proceedings, 4th International Beam Instrumentation Conference, IBIC2015, 2016.
url: <http://inspirehep.net/record/1480645/files/mopb046.pdf>, doi:10.18429/JACoW-IBIC2015-MOPB046.
- [54] L. van Riesen-Haupt, J. Coello de Portugal, E. Fol, A. Seryi, R. Tomás, et al., *K -modulation developments via simultaneous beam based alignment in the LHC*, in 8th Int. Particle Accelerator Conf.(IPAC'17), Copenhagen, Denmark, 14 -19 May, 2017, JACOW, Geneva, Switzerland, 2017.
- [55] F. Carlier and R. Tomás, *Accuracy and feasibility of the β^* measurement for LHC and High Luminosity LHC using k modulation*, Phys. Rev. Accel. Beams **20** (2017).
- [56] R. Tomás García, *Direct Measurement of Resonance Driving Terms in the Super Proton Synchrotron (SPS) of CERN using Beam Position Monitors*, 2003.
Presented on 30 Mar 2003.

- url: <https://cds.cern.ch/record/615164>.
- [57] R. Tomás, M. Bai, R. Calaga, W. Fischer, A. Franchi, and G. Rumolo, *Measurement of global and local resonance terms*, Phys. Rev. ST Accel. Beams **8** (2005), p. 024001.
url: <https://link.aps.org/doi/10.1103/PhysRevSTAB.8.024001>, doi:10.1103/PhysRevSTAB.8.024001.
- [58] T. Persson, *Beam-Based Error Identification and Correction Methods for Particle Accelerators*, PhD thesis, Chalmers University of Technology, 2014.
CERN-THESIS-2014-324.
- [59] Y. Chung, G. Decker, and K. Evans, *Measurement of beta-function and phase using the response matrix*, in Proceedings of International Conference on Particle Accelerators, May 1993, pp. 188–190 vol.1.
doi:10.1109/PAC.1993.308986.
- [60] J. W. Dilly, L. Malina, and R. Tomas Garcia, *An Updated Global Optics Correction Scheme*, CERN-ACC-NOTE-2018-0056 (2018).
url: <https://cds.cern.ch/record/2632945>.
- [61] H. Grote and F. Schmidt, *MAD-X: An Upgrade from MAD8* (2003), p. 4 p.
url: <https://cds.cern.ch/record/618496>.
- [62] M. Aiba et al., *First beta-beating measurement and optics analysis for the CERN Large Hadron Collider*, Phys. Rev. ST Accel. Beams **12** (2009).
doi:10.1103/PhysRevSTAB.12.081002.
- [63] A. Langner, J. Coello de Portugal, P. Skowronski, and R. Tomás, *Developments of the segment-by-segment technique for optics corrections in the LHC* (CERN-ACC-2015-331, 2015).
url: <https://cds.cern.ch/record/2141783>.
- [64] J. Coello de Portugal, R. Tomás, and M. Hofer, *New local optics measurements and correction techniques for the LHC and its luminosity upgrade*, Phys. Rev. Accel. Beams **23** (2020), p. 041001.
- [65] J. F. Cardona, A. C. García Bonilla, and R. Tomás García, *Local correction of quadrupole errors at LHC interaction regions using action and phase jump analysis on turn-by-turn beam position data*, Phys. Rev. Accel. Beams **20** (2017), p. 111004.
- [66] K. Albertsson, P. Altoe, D. Anderson, J. Anderson, M. Andrews, J. P. A. Espinosa, A. Aurisano, L. Basara, A. Bevan, W. Bhimji, D. Bonacorsi, B. Burkle, P. Calafiura, and M. Campanelli, *Machine learning in high energy physics community white paper*, 2019.
arXiv:1807.02876.

-
- [67] A. Scheinker, C. Emma, A. L. Edelen, and S. Gessner, *Advanced control methods for particle accelerators (acm4pa) 2019 workshop report*, 2020.
arXiv:2001.05461.
- [68] S. Biedron, *Adding Data Science and More Intelligence to Our Accelerator Toolbox*, in 10th International Particle Accelerator Conference, 2019, p. TUZPLM1.
doi:10.18429/JACoW-IPAC2019-TUZPLM1.
- [69] A. Edelen, S. Biedron, B. Chase, E. Jr, S. Milton, and P. Stabile, *Neural networks for modeling and control of particle accelerators*, IEEE Transactions on Nuclear Science **63** (2016), pp. 878–897.
doi:10.1109/TNS.2016.2543203.
- [70] C. Emma, A. Edelen, M. Hogan, B. O’Shea, G. White, and V. Yakimenko, *Machine learning-based longitudinal phase space prediction of particle accelerators*, Phys. Rev. Accel. Beams **21** (2018), p. 112802.
doi:10.1103/PhysRevAccelBeams.21.112802.
- [71] X. Xu, Y. Leng, and Y. Zhou, *Machine Learning Image Processing Technology Application in Bunch Longitudinal Phase Data Information Extraction*, in 8th International Beam Instrumentation Conference, 2019, p. WEPP021.
doi:10.18429/JACoW-IBIC2019-WEPP021.
- [72] B. Gao, Y. Leng, and X. Xu, *Deep Learning Applied for Multi-Slit Imaging Based Beam Size Monitor*, in 10th International Particle Accelerator Conference, 2019, p. WEPGW049.
doi:10.18429/JACoW-IPAC2019-WEPGW049.
- [73] Y. Lecun, K. Kavukcuoglu, and C. Farabet, *Convolutional networks and applications in vision*, 05 2010, pp. 253–256.
doi:10.1109/ISCAS.2010.5537907.
- [74] A. Edelen, S. Biedron, J. Edelen, and S. Milton, *First Steps Toward Incorporating Image Based Diagnostics into Particle Accelerator Control Systems Using Convolutional Neural Networks*, in 2nd North American Particle Accelerator Conference, 2017, p. TUPOA51.
arXiv:1612.05662, doi:10.18429/JACoW-NAPAC2016-TUPOA51.
- [75] D. Vilsmeier, M. Sapinski, and R. Singh, *Space-charge distortion of transverse profiles measured by electron-based Ionization Profile Monitors and correction methods*, Phys. Rev. Accel. Beams **22** (2019), p. 052801.
arXiv:1811.00371, doi:10.1103/PhysRevAccelBeams.22.052801.
- [76] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, *Continuous control with deep reinforcement learning*, 2019.

- arXiv:1509.02971.
- [77] T. Boltz, T. Asfour, M. Brosi, E. Bründermann, B. Härer, P. Kaiser, A.-S. Müller, C. Pohl, P. Schreiber, and M. Yan, *Feedback Design for Control of the Micro-Bunching Instability based on Reinforcement Learning*, in 10th International Particle Accelerator Conference, 2019, p. MOPGW017.
doi:10.18429/JACoW-IPAC2019-MOPGW017.
- [78] A. Ivanov and I. Agapov, *Physics-based deep neural networks for beam dynamics in charged particle accelerators*, Phys. Rev. Accel. Beams **23** (2020), p. 074601.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.23.074601>, doi:10.1103/PhysRevAccelBeams.23.074601.
- [79] G. Valentino and B. Salvachua, *Machine Learning Applied at the LHC for Beam Loss Pattern Classification*, J. Phys. Conf. Ser. **1067** (2018), p. 072036.
doi:10.18429/JACoW-IPAC2018-WEPAF078.
- [80] G. Azzopardi, B. Salvachua, G. Valentino, S. Redaelli, and A. Muscat, *Operational results on the fully automatic LHC collimator alignment*, Phys. Rev. Accel. Beams **22** (2019), p. 093001.
doi:10.1103/PhysRevAccelBeams.22.093001.
- [81] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner, *Bayesian optimization of a free-electron laser*, Phys. Rev. Lett. **124** (2020), p. 124801.
url: <https://link.aps.org/doi/10.1103/PhysRevLett.124.124801>, doi:10.1103/PhysRevLett.124.124801.
- [82] A. Scheinker and D. Scheinker, *Extremum seeking optimal controls of unknown systems*, 2018.
arXiv:1808.05181.
- [83] A. Scheinker, S. Hirlaende, F. M. Velotti, S. Gessner, G. Z. D. Port, V. Kain, and R. Ramjiawan, *Online multi-objective particle accelerator optimization of the awake electron beam line for simultaneous emittance and orbit control*, 2020.
arXiv:2003.11155.
- [84] R. Tomás et al., *LHC optics measurement and correction software progress and plans*, in 10th International Particle Accelerator Conference, 2019, p. WEPGW116.
doi:10.18429/JACoW-IPAC2019-WEPGW116.
- [85] A. Langner and R. Tomás, *Optics measurement algorithms and error analysis for the proton energy frontier*, Phys. Rev. ST Accel. Beams **18** (2015).

- doi:10.1103/PhysRevSTAB.18.031002.
- [86] F. T. Liu, K. M. Ting, and Z.-H. Zhou, *Isolation forest*, in Proc. 8th IEEE Int. Conf. on Data Mining (ICDM'08), IEEE Computer Society, 2008, pp. 413–422.
doi:10.1109/ICDM.2008.17.
- [87] E. Fol, *Evaluation of machine learning methods for LHC optics measurements and corrections software*, Master's thesis, University of Applied Sciences, Karlsruhe, 2017.
- [88] M. Gasior, J. Olexa, and R. Steinhagen, *A high-resolution diode-based orbit measurement system - prototype results from the LHC*, Tech. Rep. CERN-ACC-NOTE-2017-0020, CERN, 2011.
- [89] L. Malina, J. M. Coello de Portugal, J. Dilly, P. Skowroński, R. Tomás, and M. Toplis, *Performance Optimisation of Turn-by-Turn Beam Position Monitor Data Harmonic Analysis*, in 9th International Particle Accelerator Conference, 2018, p. THPAF045.
doi:10.18429/JACoW-IPAC2018-THPAF045.
- [90] E. Fol, *Detection of faulty beam position monitors*.
presented at ICFA Beam Dynamics Mini-Workshop: Machine Learning Applications for Particle Accelerators, 2018.
- [91] R. Tomás et al., *LHC run 2 optics commissioning experience in view of HL-LHC*, in Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019, 2019, pp. 508–511.
- [92] A. Garcia-Tabares, *Beta from amplitude - ballistic and 2.5 km review*.
OMC Team meeting, <https://indico.cern.ch/event/549409/>, 8. July 2016.
- [93] A. Garcia-Tabares, *Bad BPMs*.
OMC Team meeting, <https://indico.cern.ch/event/579332/>, 21. October 2016.
- [94] L. Malina, *Noise and stabilities*.
OMC Team meeting, <https://indico.cern.ch/event/859128/>, 30. October 2019.
- [95] D. W. Wolf, *Analysis of tune modulations in the LHC*, bachelor's thesis, Karlsruhe Institute of Technology (KIT), 2018.
- [96] R. Tomás et al., *Record low beta beating in the LHC*, Phys. Rev. ST Accel. Beams **15** (2012).
doi:10.1103/PhysRevSTAB.15.091001.
- [97] E. Fol, J. M. Coello de Portugal, and R. Tomás, *Application of Machine Learning to Beam Diagnostics*, in 7th International Beam Instrumentation Conference, JACoW, 2019, p. TUOA02.
doi:10.18429/JACoW-IBIC2018-TUOA02.

- [98] E. Fol, F. Carlier, J. M. Coello de Portugal, A. Garcia-Tabares, and R. Tomás, *Machine Learning Methods for Optics Measurements and Corrections at LHC*, in 9th International Particle Accelerator Conference, 2018, p. WEPAF062.
doi:10.18429/JACoW-IPAC2018-WEPAF062.
- [99] E. Fol, J. M. Coello de Portugal, G. Franchetti, and R. Tomás, *Optics corrections using machine learning in the LHC*, in 10th Int. Particle Accelerator Conf. (IPAC'19), 2019, pp. 3990–3993.
doi:10.18429/JACoW-IPAC2019-THPRB077.
- [100] P. Hagen, M. Giovannozzi, J.-P. Koutchouk, T. Risselada, F. Schmidt, E. Todesco, and E. Wildner, *WISE: A Simulation of the LHC Optics including Magnet Geometrical Data*, LHC-Project-Report-1123 (2008).
url: <https://cds.cern.ch/record/1123714>.
- [101] J. Coello de Portugal, *Results form operation and MDs and implications for HL-LHC: Linear corrections*.
113th HiLumi WP2 Meeting, <https://indico.cern.ch/event/685264>, 19. December 2017.
- [102] R. Calaga, R. Tomás, and F. Zimmermann, *BPM calibration independent lhc optics correction*, in 2007 IEEE Particle Accelerator Conference (PAC), 2007, pp. 3693–3695.
- [103] A. Verdier, *Alignment optics for LHC*, Tech. Rep. LHC-PROJECT-NOTE-325, CERN, Geneva, Jul 2003.
url: <http://cds.cern.ch/record/691914>.
- [104] L. van Riesen-Haupt, *Advanced accelerator interaction region optics for LHC operation and future hadron colliders*, PhD thesis, 2019.
- [105] A. G.-T. Valdivieso and R. Tomás, *Optics-measurement-based beam position monitor calibrations in the lhc insertion regions*, Phys. Rev. Accel. Beams **23** (2020), p. 042801.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.23.042801>, doi:10.1103/PhysRevAccelBeams.23.042801.
- [106] F. Chollet et al., *Keras*.
<https://github.com/fchollet/keras>, 2015.
- [107] F. Carlier and R. Tomás, *Accuracy and feasibility of the β^* measurement for lhc and high luminosity lhc using k modulation*, Phys. Rev. Accel. Beams **20** (2017), p. 011005.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.20.011005>, doi:10.1103/PhysRevAccelBeams.20.011005.

- [108] S. C. Leemann, S. Liu, A. Hexemer, M. A. Marcus, C. N. Melton, H. Nishimura, and C. Sun, *Demonstration of machine learning-based model-independent stabilization of source properties in synchrotron light sources*, Phys. Rev. Lett. **123** (2019), p. 194801.
url: <https://link.aps.org/doi/10.1103/PhysRevLett.123.194801>, doi:10.1103/PhysRevLett.123.194801.
- [109] A. Scheinker, D. Bohler, S. Tomin, R. Kammering, I. Zagorodnov, H. Schlarb, M. Scholz, B. Beutner, and W. Decking, *Model-independent tuning for maximizing free electron laser pulse energy*, Phys. Rev. Accel. Beams **22** (2019), p. 082802.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.22.082802>, doi:10.1103/PhysRevAccelBeams.22.082802.
- [110] V. Kain, S. Hirlander, B. Goddard, F. M. Velotti, G. Zevi Della Porta, N. Bruchon, and G. Valentino, *Sample-efficient reinforcement learning for CERN accelerator control*, Phys. Rev. Accel. Beams **23** (2020), p. 124801.
doi:10.1103/PhysRevAccelBeams.23.124801.
- [111] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelman, *Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems*, Phys. Rev. Accel. Beams **23** (2020), p. 044601.
url: <https://link.aps.org/doi/10.1103/PhysRevAccelBeams.23.044601>, doi:10.1103/PhysRevAccelBeams.23.044601.
- [112] A. Hanuka, C. Emma, T. Maxwell, A. S. Fisher, B. Jacobson, M. J. Hogan, and Z. Huang, *Accurate and confident prediction of electron beam longitudinal properties using spectral virtual diagnostics*, Scientific Reports **11** (2021).
url: <http://dx.doi.org/10.1038/s41598-021-82473-0>, doi:10.1038/s41598-021-82473-0.
- [113] L. T. D. Coyle, *Machine learning applications for Hadron Colliders: LHC lifetime optimization*, 2018.
Presented 2018.
url: <http://cds.cern.ch/record/2719933>.
- [114] G. Valentino and B. Salvachua, *Machine learning applied at the LHC for beam loss pattern classification*, Journal of Physics: Conference Series **1067** (2018), p. 072036.
url: <https://doi.org/10.1088/1742-6596/1067/7/072036>, doi:10.1088/1742-6596/1067/7/072036.

Curriculum Vitae

Elena Fol

Rue de Contamines 18
1206 Geneva, Switzerland

Email: elena.fol@cern.ch

Born: April 7, 1991 in Novosibirsk, Russian Federation

Nationality: German

Current position

Senior Fellow, European Organization for Nuclear Research (CERN), Meyrin, Switzerland

Research & Professional Experience

2018 - 2021	Doctoral Student, CERN, Goethe University Frankfurt
2016 - 2017	Technical Student, CERN, Meyrin
2013 - 2016	Test Automation and Software Development, Thomson Reuters Germany, Karlsruhe

Education

2015-2017	MSc in Computer Science, Karlsruhe University of Applied Science
2011-2015	BSc in Business Informatics, Karlsruhe University of Applied Science

Publications & talks

JOURNAL ARTICLES & CONFERENCE PROCEEDINGS

- E. Fol, R. Tomas, G. Franchetti, “Estimation of linear optics error sources in circular colliders using Supervised Learning”, The European Physical Journal Plus volume 136, Article number: 365 (2021).
- E. Fol, J. M. Coello de Portugal, R. Tomas, and G. Franchetti, “Detection of Faulty Beam Position Monitors using Unsupervised Learning”, Phys. Rev. Accel. Beams 23, 102805 – Published 27 October 2020.
- P. Arpaia et al., “Machine learning for beam dynamics studies at the CERN Large Hadron Collider”, NIM A, 2020.
- E. Fol, J. M. Coello de Portugal, R. Tomas, and G. Franchetti, “Optics Corrections Using Machine Learning in the LHC”, 10th Int. Particle Accelerator Conf. (IPAC’19), Melbourne, Australia, May 2019, paper THPRB077
- E. Fol, J. M. Coello de Portugal, and R. Tomas, “Unsupervised Machine Learning for Detection of Faulty Beam Position Monitors”, 10th Int. Particle Accelerator Conf. (IPAC’19), Melbourne, Australia, May 2019, paper WEPGW081

- E. Fol, "Evaluation of Machine Learning Methods for LHC Optics Measurements and Corrections Software", Master thesis, CERN-THESIS-2017-336

INVITED TALKS

- "Machine Learning for Optics Measurements and Corrections at the LHC", German Physical Society: Machine Learning Seminar, September 2020.
- "Machine Learning and Colliders", Center for Bright Beams Annual Symposium, Cornell University, July 2020.
- "What did we learn from Machine Learning?", ARIES-APEC workshop "Mitigation Approaches for Hadron Storage Rings and Synchrotrons", June 2020.
- "Machine Learning Techniques for Optics Measurements at the LHC", 11th Int. Particle Accelerator Conf. (IPAC'20), May 2020.
- "Application of Machine Learning to Beam Optics", Free Electron Laser Conf. (FEL'19), Hamburg, Germany, August 2019, paper WEBo3.
- "Application of Machine Learning to Beam Diagnostics", 7th International Beam Instrumentation Conference (IBIC'18), Shanghai, China, Sep. 2018, paper TUOAo2.
- "Detection of faulty beam position monitors", ICFA Mini-Workshop on Machine Learning for Charged Particle Accelerators, February 2018, SLAC.