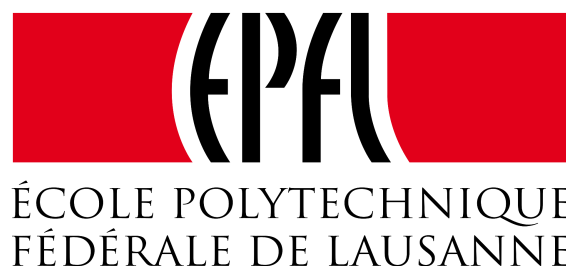


GÉNIE ÉNERGÉTIQUE ET NUCLÉAIRE
2018

Machine learning applications for Hadron Colliders: LHC lifetime optimization



Supervisors:

Dr. Belen SALVACHUA
belen.salvachua@cern.ch

Dr. Tatiana PIELONI
tatiana.pieloni@epfl.ch

Prof. Leonid RIVKIN
leonid.rivkin@epfl.ch

Phelma Tutor:

Dr. Fabrice NARAGHI
naraghi@lpsc.in2p3.fr

Author: Loïc Coyle

From 19/02/2018 to 19/08/2018

LPAP

EPFL-SB-IPHYS-LPAP
Cubotron (CUB/BSP)
CH-1015 Lausanne



Acknowledgements

This work would not have been possible without the support of many people that I would like to gratefully acknowledge. I would like to thank my supervisors Tatiana PIELONI, Belen SALVACHUA and Leonid RIVKIN for giving me this outstanding opportunity and for their guidance throughout my stay, my colleagues Claudia TAMBASCO, Lota METHER and Javier BARRANCO for their help and co-operation. Last but not the least, my partner Zineb ALY for her moral support and encouragement.

Contents

1	Introduction	5
1.1	EPFL : Ecole Polytechnique Fédérale de Lausanne	5
1.2	LPAP : Particle Accelerator Physics Laboratory	5
1.3	CERN : European Organization for Nuclear Research	5
1.3.1	CERN accelerator complex	6
2	The Large Hadron Collider	6
2.1	Overview	7
3	Accelerator physics : refresher	8
3.1	Multipolar Magnets	8
3.2	Tune	11
3.3	Beam Chromaticity	11
3.4	Landau Octupole Magnets	14
3.5	Machine performance	15
4	LHC Operation with beam	16
4.1	Machine filling	16
4.2	LHC beam modes	16
4.3	Beam Instrumentation	17
4.4	Data logging and acquisition	17
5	Characterization of losses during the energy ramp	18
5.1	Optics match points	18
5.2	Peak finder	18
6	LHC and Machine Learning	20
6.1	The 2017 data	20
6.2	Preprocessing	23
6.3	Choosing the model - Model comparison	25
6.4	Gradient Boosted Decision Tree	27
6.4.1	Decision trees	27
6.4.2	Gradient Boosting	28
6.5	Training process	29
6.5.1	Hyperparameter tuning	29
6.5.2	Training of the model	30
6.5.3	Performance of the model	30
6.5.4	Feature importance	31

6.6	Model trained on bare 2017 data	32
6.7	Application to 2018 data	36
6.8	Optimization proof of concept	37
6.9	Corroboration through simulations	38
6.10	Time series prediction	39
6.10.1	Feature engineering and preliminary results	40
7	Conclusion	41
A	Visualisation 2017	46
B	Visualisation 2018	48
C	Training model on 2018 data	50
D	Tune Diagrams	53
E	Gantt Chart	54
F	Summaries	54
F.1	English	54
F.2	Français	55

Abstract

The Large Hadron Collider is an indescribably complicated system with numerous intertwined systems, each impacts in its own way the dynamics and stability of the protons. As such, building a model of the particle losses occurring within the LHC is an extremely daunting task, but it would offer valuable insight into the inner workings of the machine, and could potentially be used to further optimize the working points of the systems. This project aims to characterize the beam losses during the ramp of energy and use Machine Learning to build models of the LHC capable of predicting the intensity lifetimes of the beams at injection energy given a set of input parameters, such as tunes, chromaticities, intensities, octupole currents and a number of other features. The main goal is to determine the optimal settings of these parameters which would help operators in the decision making process, when striving to improve the performance and physics of such colliders. Two input feature sets are defined, one encompasses as many lifetime relevant measurements as possible and allows us to train a model as information complete as possible. The other, more practical feature set, only covers parameters which the operators can control, the knobs of the machine. The models are trained on experimental data using a Gradient Boosted Decision Tree supervised learning algorithm. From the models and covariance calculations we are able to extract the most relevant features in terms of contribution to the beam lifetimes. These results would be useful in understanding the machine on a fill by fill basis and also shed light on the physics mechanisms behind the lifetime variations.

To go further, a similar method will be used on simulation data. In parallel a numerical model has been developed to physically explain the results gleaned from the machine learning analysis. A similar method could be applied to train on a dataset produced by simulations. The aim is to then use this model instead of the extremely computationally expensive SixTrack simulations, thus allowing the gained CPU time to be spent elsewhere, leading to more thorough studies.

1 Introduction

1.1 EPFL : Ecole Polytechnique Fédérale de Lausanne

EPFL is a public institution founded in 1969 [1], mainly funded by the Swiss Confederation, considered to be in the top 20 scientific institutes worldwide, it is without a doubt at the forefront of science and knowledge. It offers a plethora of bachelors and masters in a variety of fields, ranging from Architecture to Basic Sciences. EPFL welcomes roughly 14 000 students, staff or researchers from all over the world, creating a truly international learning environment where curiosity is key. Emphasis is put on encouraging innovation by providing both the infrastructure, such as high performance computer systems, and the support required to nourish new ideas. In fact, it is home to over 350 laboratories and is the birthplace of 245 start-ups established between 2000 and 2017, such as Nestle, Logitech, Credit Suisse, Constellium, Cisco and Siemens.

1.2 LPAP : Particle Accelerator Physics Laboratory

EPFL is home to the Particle Accelerator Physics Laboratory, LPAP[2]. The LPAP, headed by Prof. Rivkin, teaches courses in accelerator physics to EPFL students and is advancing the understanding of accelerator physics in many fields and applications. In fact, accelerators are not only the engines of discovery used to probe the fundamental composition of matter, as is done in colliders worldwide. They are also essential for biomedical and materials research, for diagnosing and treating illnesses, and for a growing number of manufacturing tasks, in energy technology, and in homeland security. Incredibly, 25 Nobel prizes in physics have had contributions from particle accelerators, and 20 more in physics, medicine and chemistry used X-ray sources. The LPAP works on projects in major accelerator institutions, such as the Paul Scherrer Institute (PSI) where they are pioneering the development of the X-Ray Free Electron Laser project, SwissFEL[3]. They are involved in developing accelerator applications for life sciences which include hadron therapy of tumours. Moreover, partnered with CERN, they contribute to the continued optimization of the Large Hadron Collider (LHC) and its upgrades[4], and to the design of the Future Circular Collider, LHC's proposed successor.

1.3 CERN : European Organization for Nuclear Research

CERN[5] is a research laboratory which makes use of the world's largest accelerator complex to answer some of humanities toughest questions. What is the universe made of? How did it begin? Founded in 1954, CERN has consistently been the world leader in the field of high energy colliders and is home to the world's biggest and most powerful complex of accelerators Figure 1. Physicists at CERN use these accelerators and detectors to test the predictions and limits of the Standard Model. The Standard Model has, over the years, explained many experimental results and precisely predicted a range of phenomena, most recently the discovery of the Higgs Boson by the ATLAS and CMS experiments of the Large Hadron Collider has widened its acceptance. However, CERN's contribution to the scientific community are not only in the field particle physics. CERN's infrastructure is the perfect test environment for new accelerator technologies such as novel super conducting magnet designs, cryogenic systems and data storage technologies. In fact, CERN is the birth place of the World Wide Web, now essential to everyday life.

1.3.1 CERN accelerator complex

The CERN accelerator complex Figure 1 is a network of interconnected accelerators of increasing energy. The accelerators successively increase the energy of the beam of charged particles and feed it to the next one. The complex is capable of accelerating particles up to the current record of 6.5 TeV, at which point the beams are collided. Many of the accelerators have their own lower energy experiments, for example the Antiproton Decelerator (AD)[6] feeds a variety of experiments aiming to shed light on the properties of antimatter. The complex is mostly used to accelerate protons, however, it is capable of also accelerating Pb^{82} and other ion species to high energies. The typical proton beam begins from a pressurised bottle of hydrogen, the hydrogen is stripped of its electron, producing a protons. They are then accelerated in the LINAC2[7] to 50 MeV, which feeds the Proton Synchrotron Booster[8] in which they are accelerated to 1.4 GeV. The next step is the Proton Synchrotron (PS)[9], which gets to beam up to 25 GeV and sends the beam to the Super Proton Synchrotron (SPS)[10] which accelerates it to 450 GeV. Finally the beam is transferred to the LHC[11], where two beams circulate both clockwise and anticlockwise, and are accelerated to close to the speed of light and remain stable for many hours.

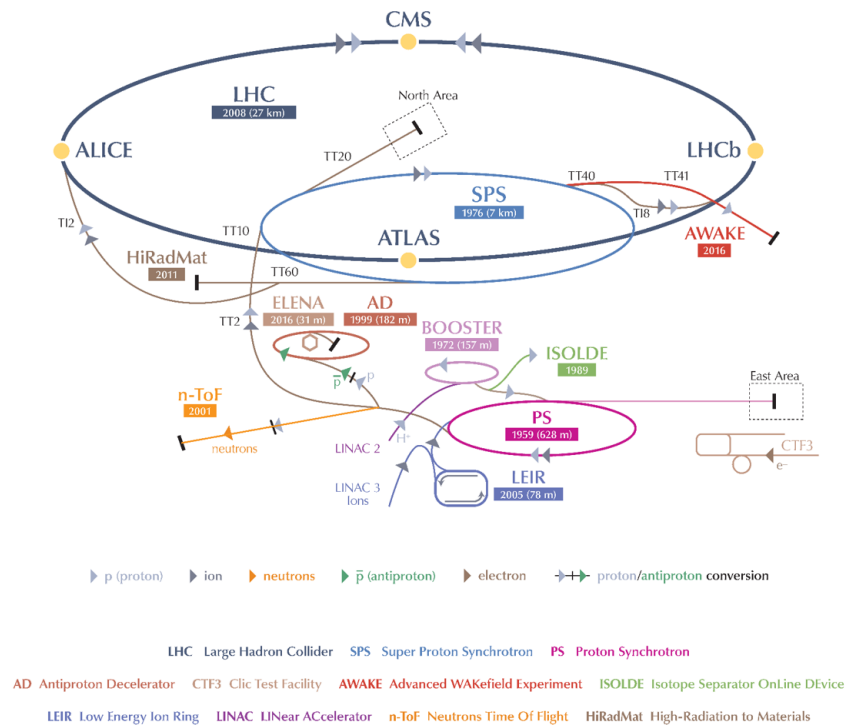


Figure 1: CERN accelerator complex[12]

2 The Large Hadron Collider

CERN's crown jewel is the Large Hadron Collider, spanning 27 km and presently reaching energies of 6.5 TeV, it is the world's biggest and most powerful particle collider. Built in the old Large Electron-Positron (LEP)[13] collider tunnel, the LHC is filled with 2 counter rotating beams which collide with a centre of mass energy of 13 TeV, offering physicist a window from which to peer into the fundamental constituents of matter. The LHC is without a doubt one of the greatest feats of engineering to date and sheds light on scientists' greatest questions.

2.1 Overview

A schematic view of the LHC is given by Figure 2. The ring is split into 8 arcs and 8 straight sections called Interaction Regions (IRs). In order to keep the particles circulating with the desired orbits, a variety of magnets are used. In the arcs, 1232 dipole super conducting magnets that provide a magnetic field of 8.33T when operated at 1.9K (-271°C) are used to keep the particles on the desired orbit. Moreover, 3136 quadrupole, 688 sextupole and 2688 octupole magnets allow the operators to set and control the beam's properties to maximise efficiency and stability.

The Radio Frequency (RF)[14] cavities, located in IR 4, transfer energy to the beams through finely tuned electromagnetic waves. Powered by Klystrons, these RF cavities operate at 400 MHz and provide the necessary energy to accelerate the beams from 450 GeV to the top energy of 6.5 TeV.

The Dump system located in IR 6, is used to protect the magnets from dangerous particle irradiation which could quench or severely damage the machine's magnets. This system is able to extract, in a single turn, both LHC beams in a safe and controlled way.

Passive protection of the LHC magnets and equipment is done by a complete collimation system located in IR 7 for the betatron cleaning and in IR 3 for the off momentum cleaning. The aim of the system is to catch beam halo particles that can be deviated from the irregular orbit of diffused to larger amplitudes.

The beams collide at four Interaction Points (IPs), IP 1,2,8 and 5, where are located the LHC's four main experiments, A Toroidal LHC ApparatuS (ATLAS)[15], A Large Ion Collider Experiment (ALICE)[16], Compact Muon Solenoid (CMS)[17] and the Large Hadron Collider beauty (LHCb)[18] experiment.

In addition to the collimation system in IR7 and IR3, local triplet magnet protection is done on each side of the colliding IPs. Injections and dump regions are also provided with dedicated collimation protection.

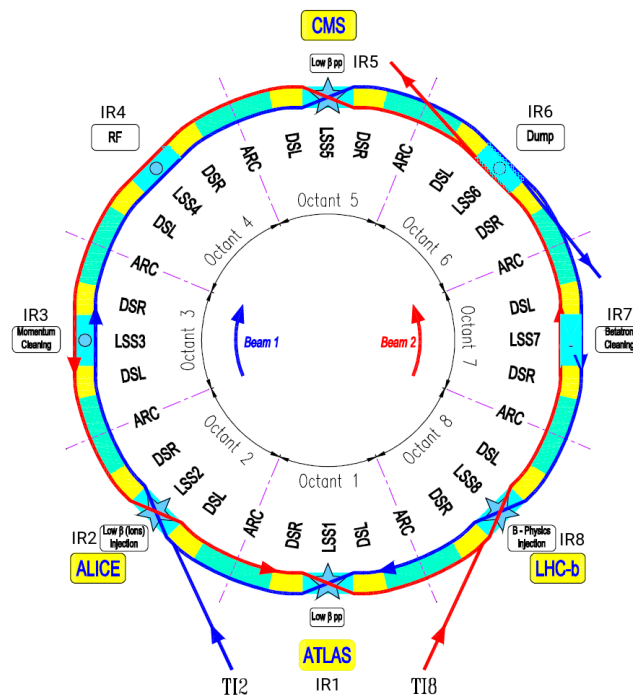


Figure 2: Schematic layout of the LHC [4]

3 Accelerator physics : refresher

This section will serve as an introduction to the key Accelerator Physics terms used in this work, to hopefully guarantee a good comprehension of the physical phenomena.

3.1 Multipolar Magnets

The motion of a charged particle in an magnetic field is obtained with the Lorentz force, equation 1. with q the charge of the particle, $\vec{E}$ the electric field, $\vec{v}$ the velocity of the particle and $\vec{B}$ the magnetic field

$$\vec{F} = m\vec{a} = q(\vec{E} + \vec{v} \times \vec{B}) \quad (1)$$

The LHC magnets utilise different polar configuration to achieve the desired effect on the beams. The bending magnets used in the LHC are dipole magnets. The field lines created by the dipole magnets are as shown in Figure 3. In these constant field lines, the protons are bent around them. The gyroradius is given by the magnetic rigidity Equation 2, where p is the momentum, B the magnetic field and ρ the gyroradius. $B\rho$ is called the magnetic rigidity.

$$\frac{1}{\rho} = \frac{eB}{p} \leftrightarrow B\rho = \frac{p}{e} \quad (2)$$

Technological limitations with regards to the strength of these magnets determine the maximum energy of the particles. The current superconducting magnets used in the LHC produce a magnetic field strength of 8.33 T.

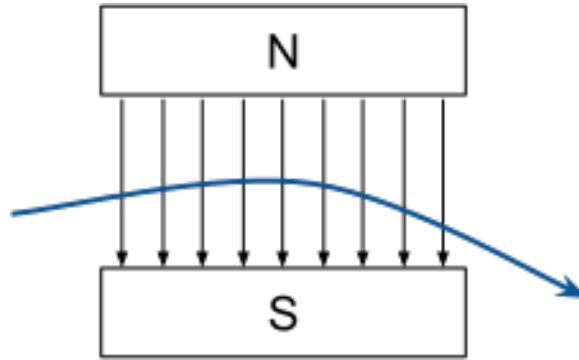


Figure 3: Particle bent in dipole field

These bending magnets are able to control the path of the protons through the machine. In the case of particles offset from the ideal trajectory, they also provide some weak geometric focusing of the beam in the horizontal plane.

Nonetheless, additional focusing is needed to keep the particle close to the ideal orbit and this is provided by quadrupole magnets, similar to optical lenses. However, unlike optical lenses, these elements focus in one plane and defocus in the other, as shown in Figure 4. Thus, the quadrupoles are placed along the ring, with alternating orientations creating, for a given plane, a succession of focusing and defocusing lenses. The field produced by these magnets scales linearly with the deviation from the ideal trajectory Equation 3, with g the field gradient.

$$B_z = -gx \quad B_x = -gz \quad (3)$$

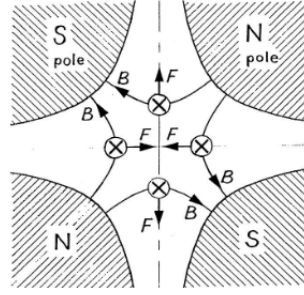


Figure 4: Quadrupole field lines and resulting forces

These two elements, dipole magnets and quadrupole magnets can be considered linear elements. Let us study the case of a particle subject to both the bending and focusing magnets. The bending magnets produce a field along $\vec{u}_z$, $B_{dip} = B_{0z}$, the quadrupole field is described by Equation 3. We first recall the acceleration in cylindrical coordinates, Figure 5, and solve Newton's second law with the Lorentz force Equation 1 with $\vec{E} = 0$:

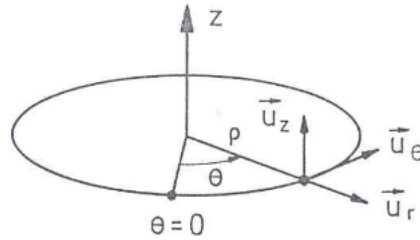


Figure 5: Cylindrical coordinates [19]

$$\begin{aligned}\ddot{\vec{R}} &= (\ddot{r} - r\dot{\theta}^2)\vec{u}_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})\vec{u}_\theta + \ddot{z}\vec{u}_z \\ m\ddot{\vec{R}} &= q\vec{v} \times \vec{B} \\ m\ddot{\vec{R}} &= q[(r\dot{\theta}B_z - \dot{z}B_\theta)\vec{u}_r + (\dot{z}B_r - \dot{r}B_z)\vec{u}_\theta + (\dot{r}B_\theta - r\dot{\theta}B_r)\vec{u}_z]\end{aligned}\tag{4}$$

In our case, $B_\theta = 0$ as neither the dipole nor the quadrupole magnets induce a longitudinal B field. We get :

$$\begin{aligned}m\ddot{R}_{/\vec{u}_r} &= m(\ddot{r} - r\dot{\theta}^2) = qr\dot{\theta}B_z \\ m\ddot{R}_{/\vec{u}_z} &= m\ddot{z} = -qr\dot{\theta}B_r\end{aligned}\tag{5}$$

Along $\vec{u}_z$ we have both the dipole and the quadrupole magnets, thus with Equation 3 we have : $B_z = B_{0z} - gx$ and along $\vec{u}_r$ only the quadrupole magnets are to be taken into account, $B_r = -gz$. Let us also introduce a displacement from ideal orbit by means of $r = \rho + x$ with $\rho = cst$ and x being the displacement, injected into Equation 5.

$$\begin{aligned}m(\ddot{x} - r\dot{\theta}^2) &= qr\dot{\theta}(B_{0z} - gx) \\ m\ddot{z} &= qr\dot{\theta}gz\end{aligned}\tag{6}$$

Seeing as the velocity of the particle is mainly in the azimuthal direction u_θ , $v \approx v_\theta = r\dot{\theta}$, and time can be substituted with the position along the desired orbit, $s = vt$.

$$\ddot{x} = \frac{d^2x}{dt^2} = \frac{d^2x}{ds^2} \frac{ds}{dt} \approx \frac{d^2x}{ds^2} \left(\frac{ds}{dt}\right)^2 = x'' v^2 \quad (7)$$

Which now gives us :

$$\begin{aligned} x'' &= \frac{1}{r} + \frac{qB_{0z}}{mv} - \frac{qg}{mv}x \\ z'' &= \frac{qg}{mv}z \end{aligned} \quad (8)$$

By recalling the magnetic rigidity equation Equation 2 and by now introducing a deviation from ideal momentum through : $mv = p = p_0(1 + \frac{\Delta p}{p_0})$. With a 1st order expansion when $x \ll \rho$ of $\frac{1}{r} = \frac{1}{\rho+x} = \frac{1}{\rho}(\frac{1}{1+x/\rho}) \approx \frac{1}{\rho}(1 - \frac{x}{\rho})$. We get for small $\frac{\Delta p}{p_0}$:

$$\begin{aligned} x'' + (k + \frac{1}{\rho^2})x &= \frac{1}{\rho} \frac{\Delta p}{p_0} \\ z'' - kx &= 0 \end{aligned} \quad (9)$$

With $k = qg/p_0$ which describes the strong focusing force of the the quadrupole magnets and $\frac{1}{\rho^2}x$ is the weak geometric focusing of the bending magnets. In the case of big accelerators such as the LHC, the geometric focusing is negligible compared to the strong focusing. Finally we can introduce $K(s) = k + \frac{1}{\rho^2}$ in the horizontal plane and $K(s) = k$ for the vertical plane. Moreover, if we make the assumption of $K = cst$ in each element, this allows us to deal with a simpler form, that of a harmonic oscillator. By considering only on momentum particle $\Delta p = 0$ and by using the periodicity of a circular machine, i.e. $K(s)$ is periodic $K(s) = K(s + C)$ with C the circumference of the machine. We get to the Hill equation in both planes :

$$u'' + Ku = 0 \quad (10)$$

One solution to Equation 10 is found using a pseudo-harmonic oscillator :

$$u(s) = a\sqrt{\beta(s)}\cos(\phi(s) - \phi_0) \quad (11)$$

The amplitude of the oscillations are modulated by the periodic $a\sqrt{\beta(s)}$ term, $\beta(s)$ is know as the beta function and gives valuable information on the beam's behaviour and is only defined by the optical properties of the lattice. Figure 6, the beta function for both planes along the LHC ring are in opposition of phase, as the effect of the quadrupole magnets is different for both planes. Moreover, the beta functions are minimised at the collision IP positions, at 0 m, 3332 m, 13329 m and 23315 m, to reduce the beam size and maximise performance. A comprehensive example of the impact of the beta function of the trajectory of the particles can be seen in Figure 7, from [20]. It can be shown, that $a = \sqrt{\epsilon}$ with ϵ the emittance of the beam. The emittance is the size of the beam in phase space (x, x') , it is constant under optical transformations. Contrary to the beta function which is a property of the optical lattice, the emittance is a property of the beam itself. The oscillation envelope is thus given by Equation 12, this defines the actual size of the beam.

$$\sigma = \sqrt{\epsilon}\sqrt{\beta} \quad (12)$$

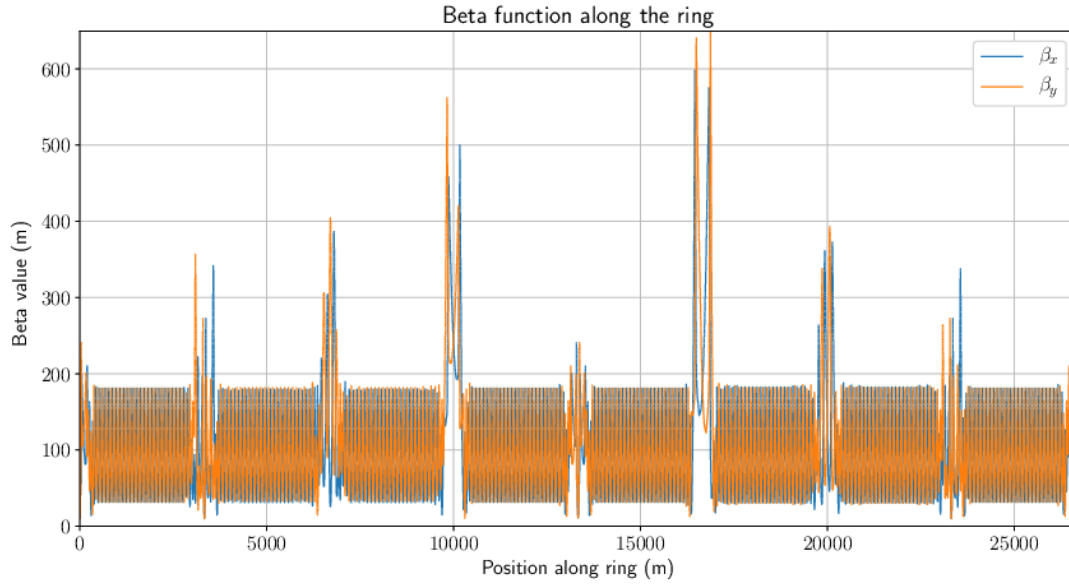


Figure 6: Beta function along the LHC ring [19]

3.2 Tune

From the beta function, one can define the tune of a particle as being the number of betatron oscillations occurring in one turn. Formally, it is given by :

$$Q = \frac{1}{2\pi} \int \frac{ds}{\beta(s)} \quad (13)$$

The tune is one of the most important parameters for beam stability. A badly set tune can lead to self-amplified instabilities, beam growth and ultimately particle losses. For example, let us consider a integer tune value, for a given point along the ring, the particle will have the same phase every turn. This will cause any infinitesimal magnetic kicks, due to momentum particles, slight magnet misalignments or imperfect currents, to accumulate every turn leading to beam instabilities. It is the fractional part of the tune which determines the particle's proximity to these resonances.

For the LHC, the nominal working point is different for both injection and collision, Figure 8. At injection the tunes are around $Q_x = 62.27$ and $Q_y = 60.29$. At collision they are set to $Q_x = 62.31$ and $Q_y = 60.32$.

To measure the value of the tunes in the machine, a few techniques are used. Firstly, it is possible to measure the beam's response to a short excitation, a so-called chirp, by using the array of BPMs. A passive alternative is to use the diode-detection based Base-Band-Q (BBQ) techniques described in [21], this method has a resolution of 10^{-4} to 10^{-6} . In the LHC however, the tight constraints on beam size do not often allow for the use of chirps as the excitation can lead to large particle losses. The beam feedback system, uses a post-processed value of this BBQ signal. Practically, the BBQ measures the frequency (tune) distribution of the particles then a peak finding algorithm determines the tune. The tune feedback system acts on a sliding time window average of the BBQ value.

3.3 Beam Chromaticity

When we take into account a spread in momentum, $\Delta p \neq 0$, particles will interact differently with the quadrupole magnets, similarly to different wavelength photons in a prism. These off mo-

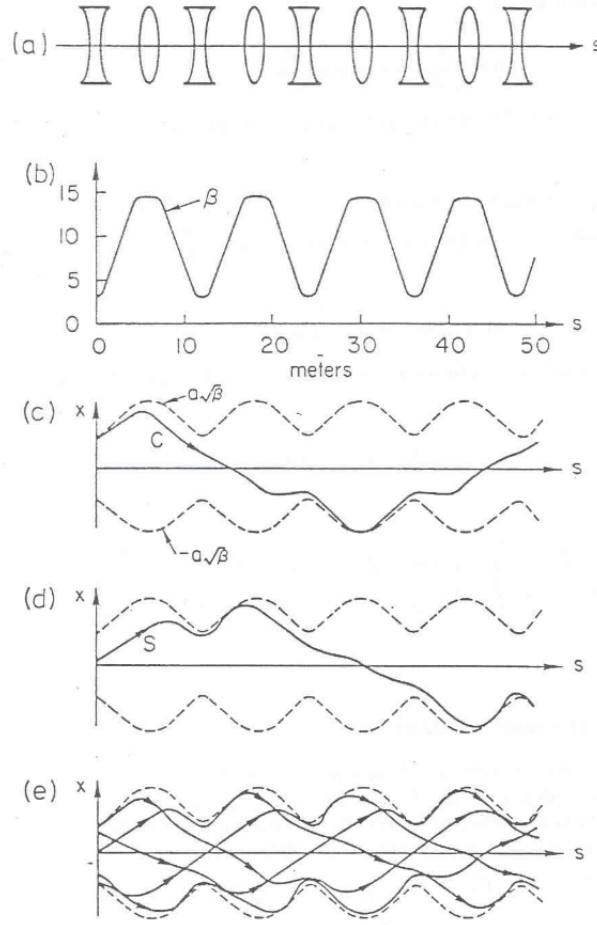


Figure 7: a) A regular lattice of focusing and defocusing lenses b) Beta function. c) Cosine-like trajectory for $s = 0$. d) Sine-like trajectory for $s = 0$. e) One trajectory on several successive revolutions. [19]

momentum particles will circulate around the machine on stable offset orbits as shown in Figure 9. The offset momentum can be considered as a field gradient error in the quadrupoles and thus can be solved using :

$$\Delta Q = \frac{1}{4\pi} \oint \beta(s) \Delta K(s) ds \quad (14)$$

Considering only the strong focusing term $K = \frac{eg}{p_0}$, with off momentum particles we get : $\Delta K = \frac{dK}{dp_0} \Delta p = -\frac{eg}{p_0} \frac{\Delta p}{p_0} = -K \frac{\Delta p}{p_0}$. We then introduce this into Equation 14 to get :

$$\Delta Q = \frac{1}{4\pi} \oint \beta(s) \left(-K(s) \frac{\Delta p}{p_0} \right) ds = \xi \frac{\Delta p}{p_0} \quad (15)$$

With $\xi = -\frac{1}{4\pi} \oint \beta(s) K(s) ds$, the chromaticity. Chromaticity, is a property of the lattice, regions with high beta values and high strength, such as in the quadrupole magnets, contribute most to the machine's chromaticity. Comprehensively, it determines how sensitive the tune is to the momentum spread. In the prism analogy, it would relate to how much well the prism separates the different wavelengths. From Equation 15, the tune shift induced by the momentum spread scales linearly with the chromaticity. This tune shift is to be taken into account when determining the optimal working point as it could lead to unwanted proximity to resonances and can trigger 'head-tail' instabilities.

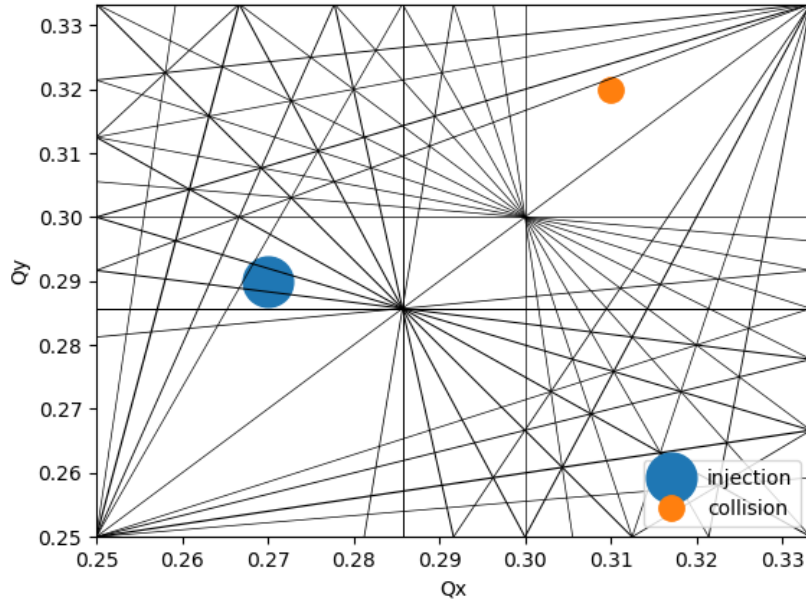


Figure 8: Tune resonance lines up to order 10, blue dot is measured tunes at injection and orange at top energy

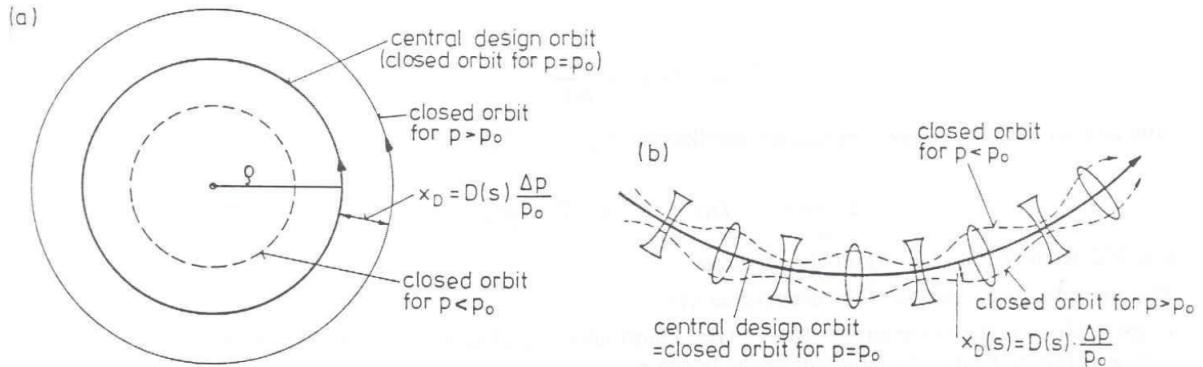


Figure 9: Closed orbit for particles with momentum $p > p_0$ in a weakly (a) and strongly (b) focusing circular accelerator[19].

Chromaticity is controlled using sextupole magnets, see Figure 10. It is at the centre of the focusing quadrupole magnets where the displacement from ideal orbit is most pronounced, thus the sextupole magnets are generally placed after the focusing quadrupoles. The field of these magnets is described with the following :

$$B_z = \frac{1}{2}g'(x+z)^2 \quad B_x = g'xz \quad (16)$$

Let us consider a off momentum particle, circulating on an larger stable closed orbit, with nonetheless some small deviations δx and δz :

$$x = x_D + \delta x \quad z = 0 + \delta z \quad (17)$$

By introducing the dispersion function defined with : $x_D(s, \frac{\Delta p}{p_0}) = D(s) \frac{\Delta p}{p_0}$, and injecting Equation 17 into Equation 16 we get :

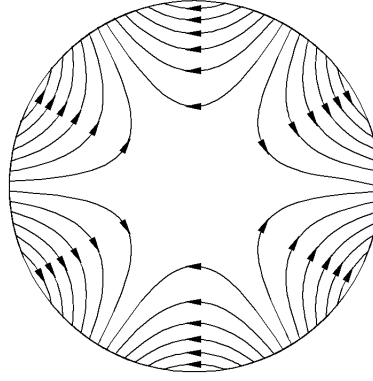


Figure 10: Sextupole magnet field lines[22]

$$\begin{aligned} B_z &= \frac{1}{2}g'(x_D + \delta x + \delta z)^2, & B_x &= g'(x_D + \delta x)\delta z \\ \leftrightarrow B_z &\approx \frac{1}{2}g'x_D^2 + g'D(s)\frac{\Delta p}{p_0}(\delta x + \delta z), & B_x &\approx g'D(s)\frac{\Delta p}{p_0}\delta x \end{aligned} \quad (18)$$

From Equation 18 the first term of B_z deflects the offset closed orbit, whereas the other terms of B_z and B_x act like quadrupole magnets with a strength which scales with the momentum offset. By placing these magnets along the ring, it is possible to control the chromaticity. However, they introduce some non-linear effects and a coupling between the horizontal and vertical planes.

The measure of chromaticity is a destructive process, which is why, in the LHC the chromaticity is controlled using a feed-forward system in which a pilot bunch of protons is sent through the machine at injection, its chromaticity is measured and corrected to 15 units in most cases. The machine is then considered calibrated for the subsequent beam. A novel, non destructive method has been developed using a Schottky measurement, described in [23].

3.4 Landau Octupole Magnets

Despite accurate control over the tune and chromaticity of the beam, coherent instabilities caused by the machine impedance, or other sources, may arise if they are not properly damped. One of the main oscillation damping mechanism used in the LHC is Landau damping [24, 25]. Landau damping is the collisionless damping of collective modes of oscillation. The complete mathematical derivation is beyond the scope of this report but a basic intuitive understanding can be gotten by considering a particle of frequency/tune ω , subject to an external driving force $f(t)$ of frequency Ω . The equation of motion is that of a harmonic oscillator :

$$\ddot{x} + \omega^2 x = A \cos(\Omega t) \quad (19)$$

For $x(0) = 0$ and $\dot{x} = 0$, the solution is :

$$x = -\frac{A}{(\omega^2 - \Omega^2)}(\cos(\Omega t) - \cos(\omega t)) \quad (20)$$

A particle beam is of course made up of many particles, with a given tune distribution $\rho(\omega)$ and spread $\Delta\omega$, the tune ω_x of the beam is the peak of the distribution. The average beam response is given by $\langle x(t) \rangle = \int_{-\infty}^{+\infty} x(t)\rho(\omega)d\omega$. Replacing with Equation 20, and considering a small tune spread and an excitation ($\Omega \approx \omega_x$) we get :

$$\langle x(t) \rangle = -\frac{A}{2\omega_x} \int_{-\infty}^{+\infty} \frac{A}{(\omega - \Omega)} (\cos(\Omega t) - \cos(\omega t)) \rho(\omega) d\omega \quad (21)$$

If we consider a complex $\Omega = \Omega_r + i\Omega_i$, it can be shown that, see [26] for a more detailed calculation:

$$\langle x(t) \rangle = \frac{A}{2\omega_x} \left[\pi \rho(\Omega) \sin(\Omega t) + \cos(\Omega t) P.V. \int_{-\infty}^{+\infty} d\omega \frac{\rho(\omega)}{\omega - \Omega} \right] \quad (22)$$

This first term is in phase with the excitation absorbing energy, it damps the excitation. The second term is the reaction of the beam to the excitation. Landau damping does not allow the particles to organise into collective instabilities through phase mixing, the collective response is damped to zero. With further calculations, it can be shown that increasing the tune spread of the particles $\Delta\omega$ or changing the tune distribution increases the stability region.

In order to have the necessary tune spread to benefit from Landau damping, 144 octupole magnets are used throughout the machine. These octupole magnets actively induce tune spread in the beam.

$$B_x = O_3(x^3 - 3xy^2) \quad B_y = O_3(3x^2y - y^3) \quad (23)$$

With O_3 the octupolar strength. It can be shown [27] that the induced tune spread from the octupoles is given by :

$$\begin{aligned} \Delta Q_x &= \left[\frac{3}{8\pi} \int \beta_x^2 \frac{O_3}{B\rho} ds \right] J_x - \left[\frac{3}{8\pi} \int 2\beta_x\beta_y \frac{O_3}{B\rho} ds \right] J_y \\ \Delta Q_y &= \left[\frac{3}{8\pi} \int \beta_y^2 \frac{O_3}{B\rho} ds \right] J_y - \left[\frac{3}{8\pi} \int 2\beta_x\beta_y \frac{O_3}{B\rho} ds \right] J_x \end{aligned} \quad (24)$$

Where J_x and J_y are the betatron action variables, a microscopic metric of emittance.

Using octupole magnets does allow for larger stable regions, but however there are some drawbacks, the introduction of more non linear field leads to more sensitivity to resonances. Moreover, it can induce unwanted changes to optical properties. Perhaps most importantly, using octupole magnets reduces the maximum amplitude at which particles are stable ,i.e. the dynamic aperture, in the machine. This in turn reduces the lifetime of the beam.

3.5 Machine performance

A particle collider's designed goal is to produce physics events. The metric for performance colliders is the collision energy and also the luminosity of the beam. Luminosity is the ratio of events per times unit and per interaction cross-section. It can directly be linked to the number of physics event $R = \frac{dN}{dt}$ through the cross-section σ of said event. Luminosity is generally defined as [4]:

$$\begin{aligned} \mathcal{L} &= \frac{R}{\sigma} \\ \mathcal{L} &= \frac{N_1 N_2 f_{rev} N_b \gamma}{4\pi \epsilon_n \beta^*} F \end{aligned} \quad (25)$$

With N_1 and N_2 the number of protons in both beams, f_{rev} the frequency of revolution, N_b the number of bunches, ϵ_n the normalised emittance and β^* the beta function at the collision point. F is

a geometric luminosity reduction factor due to the crossing angle at the IP. The unit of luminosity is $\text{cm}^{-2}\text{s}^{-1}$. To maximise the luminosity of the machine, one must aim to minimise the beam's size while having as many particles and bunches as possible.

In addition to the luminosity, the intensity lifetime of the beam determines how long two beams can collide producing physics events before they must be dumped and replenished. When the beams collide only a fraction of the particles actually interact, these particles burn-off in the physics producing interactions. The beam's intensity lifetime is defined as the time for the total intensity to reduce by a factor $1/e$.

$$\tau = \frac{-I}{dI/dt} \quad (26)$$

In colliders it is desirable that the only contribution to the beam's lifetimes are from the collisions. However, there are unavoidable beam losses due, for example, to beam-gas interaction, halo cleaning, orbit drifts, etc. These losses need to be monitored and minimised over the whole operational cycle of the LHC.

4 LHC Operation with beam

4.1 Machine filling

In the LHC and throughout the CERN complex, the beam is not a continuous stream of particles. It is instead a series of concentrated bunches of particles. A nominal LHC beam is made up of up to 2556 bunches each separated by 25ns gap. These bunches are themselves made up of roughly 1.15×10^{11} protons. The size of these bunches vary from a few $10\mu\text{m}$ to a few mm, as they circulate through the machine, in the collision points they are at their smallest so as to maximise the number of collisions.

4.2 LHC beam modes

After filling the LHC ring at injection energy 450 GeV from the the SPS the beams will need to be accelerated to 6.5 TeV, squeezed in the colliding IPs to reduce the β^* (the β at the collision IPs) and increase the instantaneous luminosity, and collapsed to provide collisions. This is done in the LHC cycle and each of these steps is segmented in so-called LHC beam modes. In Figure 11, a schematic view of the LHC's optimal cycle showing the beam mode status. It shows the evolution of the dipole current through the beam mode cycle. In order to keep the beam confined in the LHC ring through the LHC cycle, the bending radius must remain constant, therefore of the strength of the bending magnets B must scale with momentum p , as described in Equation 2.

For each fill, the LHC operators must keep the beams as stable as possible while gradually preparing them for collision. Firstly, the beams are injected into the machine. During INJECTION, the beam's properties are measured and magnet strengths are adjusted, after which the machine is ready to proceed. Then the operators enter PRERAMP mode where the injection collimators are retracted and the system is being prepared to start a synchronous ramp up of energy. In the RAMP mode, the beams are accelerated and all the magnets, collimators and RF cavities follow predefined functions to preserve the quality of the beam during the ramp. Once the beams reach the maximum energy of 6.5 TeV, the FLATTOP takes care of levelling off the energy smoothly and does pre-SQUEEZE checks, collision tunes are set. During the SQUEEZE, the beams sizes are locally reduced near the experimental regions to maximise the interaction probability at the IRs. The ADJUST mode, collapses the separation between the two beams. Finally, in STABLE mode collisions occur and physics events are produced, the LHC is able to stay STABLE for more than 15 hours.

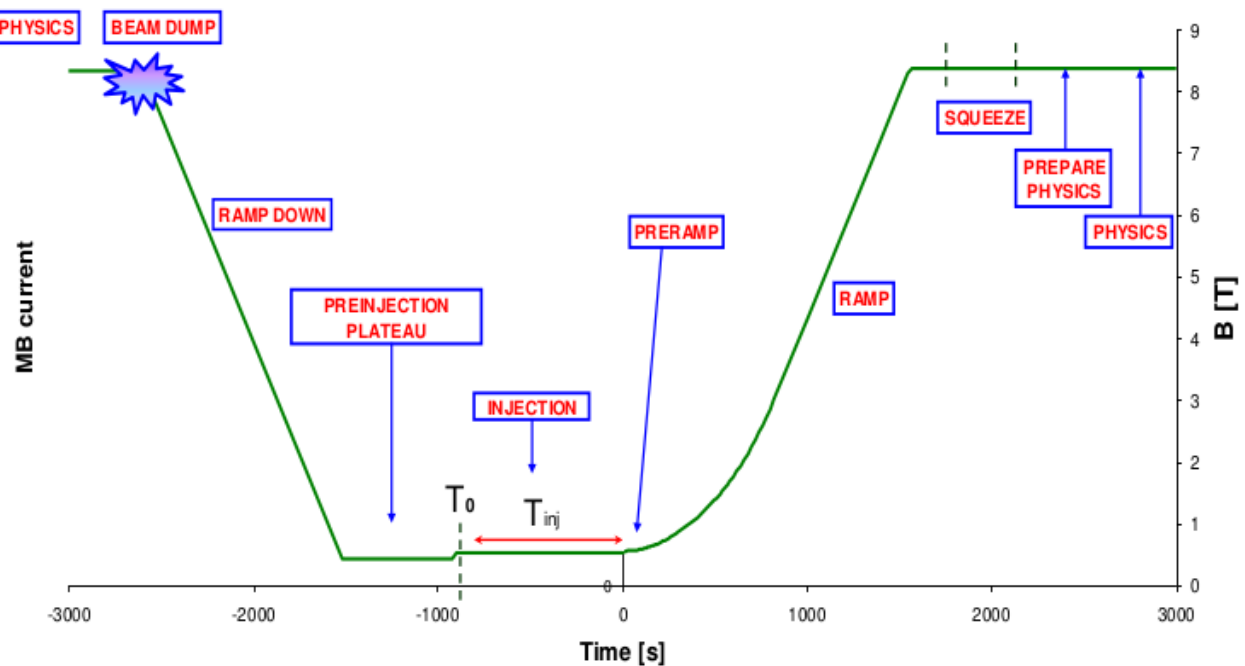


Figure 11: The evolution of the bending magnet current for a nominal LHC beam mode sequence

4.3 Beam Instrumentation

Many beam instrumentation devices are placed in and around the beam pipes to monitor the beams' properties as they circulate through the machine. In order to determine the orbit of the beam and adjust when needed, 1056 Beam Position Monitors (BPMs) are placed throughout the machine, see Figure 12. These devices are installed around both beam pipes, in both the horizontal and vertical planes. Using the electric field produced by the passing beams, the array of BPMs is able to determine the position of the beams throughout the machine to within 10 μm . The knowledge of the beams' positions is essential for the Beam Orbit Feedback Service Unit (BOFSU) which acts on the BPM measurements to constantly readjust the dedicated orbit corrector magnets, correcting deviations in the beams' orbits.

Beam Loss Monitors (BLMs) are ionisation chambers placed in key spots around the machine, typically after aperture constraints such as collimators, as shown in Figure 13. They measure the particle showers produced by stray particles. These BLMs give the operators insight into the expected lifetime of the beams but are also part of the active machine protection system, indeed each of the close to 4000 BLMs along the LHC can trigger a beam dump within 3 turns. It is of utmost importance to continuously monitor the irradiation of the magnets, as with any superconducting magnet, if they are subject to high amounts of radiation they may quench. The quench protection system safely controls the release of 11 GJ of energy stored in these magnets.

These two are the main systems to measure beam position and beam losses. However, other devices are also installed to measure specific beam properties, like the transverse beam profile, longitudinal beam profiles, tunes, etc.

4.4 Data logging and acquisition

The data used throughout this project was acquired through the CERN Accelerator Logging Service, the backbone of CERN's data storage infrastructure allowing of physicists from around the world real time access to CERN data.

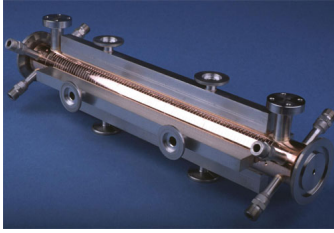


Figure 12: Beam Position Monitor [28]



Figure 13: Beam Loss Monitors on LHC [29]

To extract data through the CERN Accelerator Logging Service, physicists typically use a custom built Java API through the Graphical User Interface called Timber. In order to better automate the data extraction, in this work, PyTimber¹ was used. PyTimber is a python based alternative to Timber which facilitates automatic and complex data analysis.

5 Characterization of losses during the energy ramp

5.1 Optics match points

During the energy ramp of the beams, the protons go from the injection energy of 450 GeV to the collision energy of 6.5 TeV. At the same time part of the squeeze process is happening. Beams are squeezed from β^* of 10 m down to 1 m for ATLAS and CMS. During this process, the machine's optics change accordingly to the new β^* . At different points in the ramp, i.e. for different energies (or times, as the ramp is synchronous), optic matching points have been defined. At these points, optimal optical settings have been determined through simulations and matched. Between these points, predefined 'trim' functions are used to smooth out the evolution, in other words, the optics values in between matching points are interpolated. The trim functions provide the necessary smoothing in order to minimise losses. However, despite this, operators have observed systematic particle losses occurring at specific points along the ramp. Figure 14 shows an example of the evolution of intensity during the ramp for both beams along with the increase in energy. For beam 1 the losses occur sharply at the start of the ramp then the intensity gradually decreases from energies onward of 3 TeV, ending with a sharp drop off at the end of the ramp. For beam 2, there is a gradual loss at the start of the ramp with the same slow decrease from 3 TeV onward, however the sharp drop off at the end of the ramp is absent.

Figure 15, shows the distribution of the relative loss in intensity during the ramp for both beams, for a selection of machine fills. It quantifies the the amount of intensity lost through the ramp. On average, beam 1 is subject to more loses than beam 2, roughly 0.56% for beam 1 compared to around 0.34% for beam 2.

Figure 16, shows an example of the evolution of the beam lifetime for the the fill 6778. It shows the presence of the oscillating structures for both beams, which emerge roughly 600 seconds into the ramp.

5.2 Peak finder

To better understand these losses, the signals from the BLMs closest to the primary collimators were used. It is at the primary collimators where the aperture of the machine is smallest and thus

¹<https://github.com/rdemaria/pytimber>

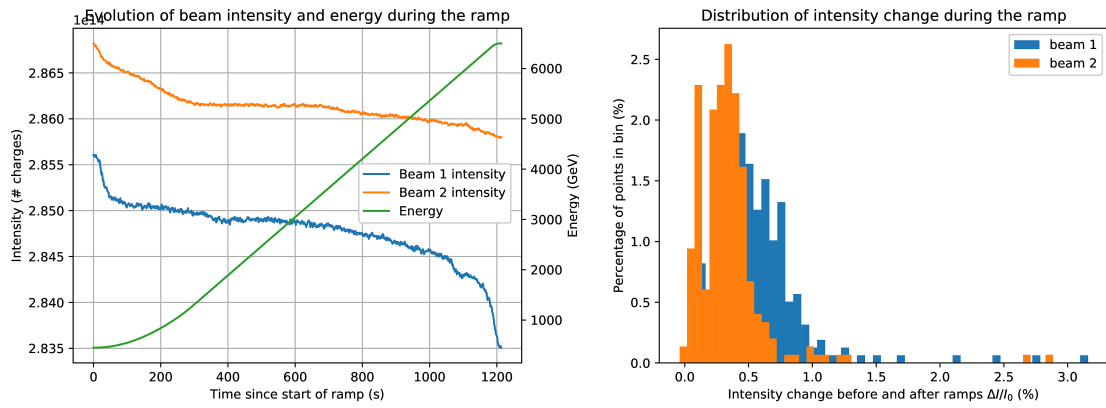


Figure 14: Evolution of energy and beam intensities during the ramp for fill 6778

Figure 15: Distribution of intensity change during ramp for a selection of fills

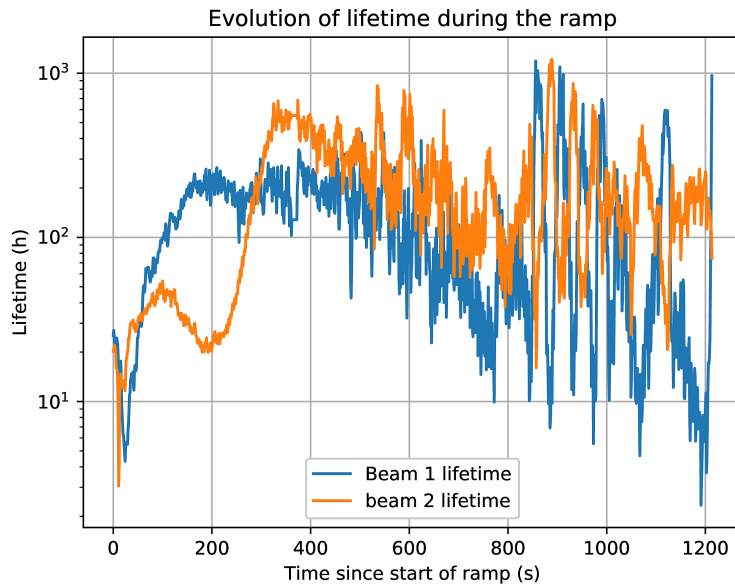


Figure 16: Typical evolution of lifetime during the ramp for fill 6778

where the losses are greatest. These collimators scrape away the eccentric particles in both momentum and betatron oscillation. As shown earlier in Figure 2, the primary collimators are situated in IRs 3 and 7. For this analysis 2 BLMs per beam were used, one at each primary collimator, Figure 17 shows the position of the monitors within the LHC ring, along side a typical measured loss signal around the ring, for a typical fill during the ramp. In this case, it is fill 6778, 1067 seconds into the ramp process, which correspond to a lifetime of 4.6 hours for beam 1. It clearly shows that most of the beam losses occur at the chosen collimators.

In Figure 18, the BLM signals from both BLMs (IR3 and IR7) are summed and the result is plotted for both beams and for a selection of fills. Each line in the plot represent the losses for a given fill. The signals are normalised to the same amplitude and a peak finding algorithm is used to extract the times of the loss peaks, the red dots. It confirms the presence of the systematic losses occurring in the ramp. All of the fills in 2017 and 2018 were analysed and losses appear systematically at the same times. Figure 19 shows the timings of the peaks found by the peak finder for different fills. The beam 1 peaks are shown on the left and the beam 2 on the right. The horizontal lines correspond to the matching points. It is observed that for beam 1 most of the losses occur exactly at the matching points, whereas for beam 2 they occur in between. These peaks detected correspond to the dips

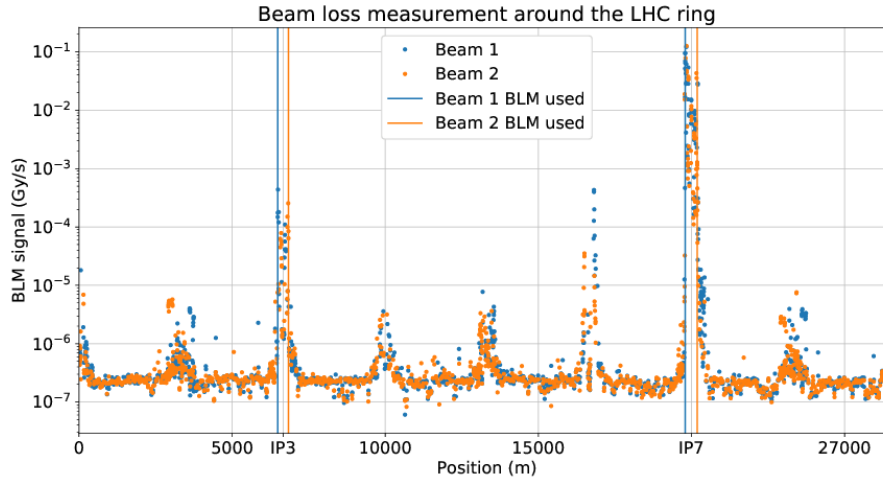


Figure 17: Distribution of BLM signal around the LHC ring for fill 6778 at $t=1067$ s.

observed the signals of lifetime in Figure 16.

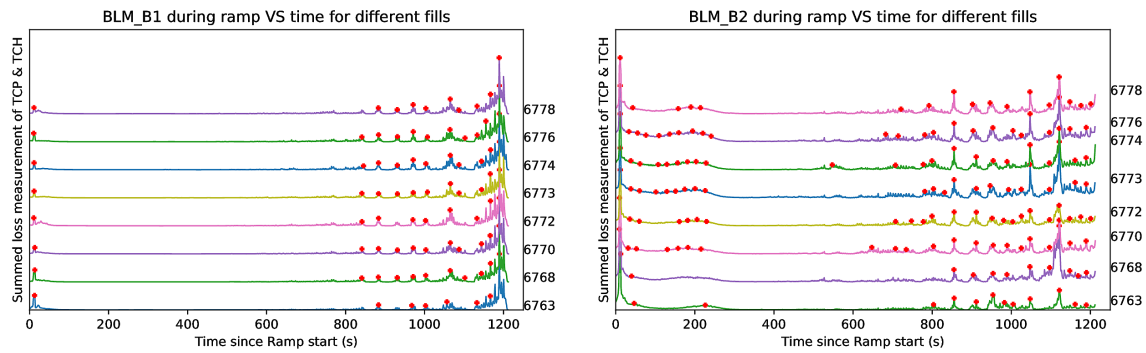


Figure 18: Time evolution of summed collimator BLM signal for different fills.

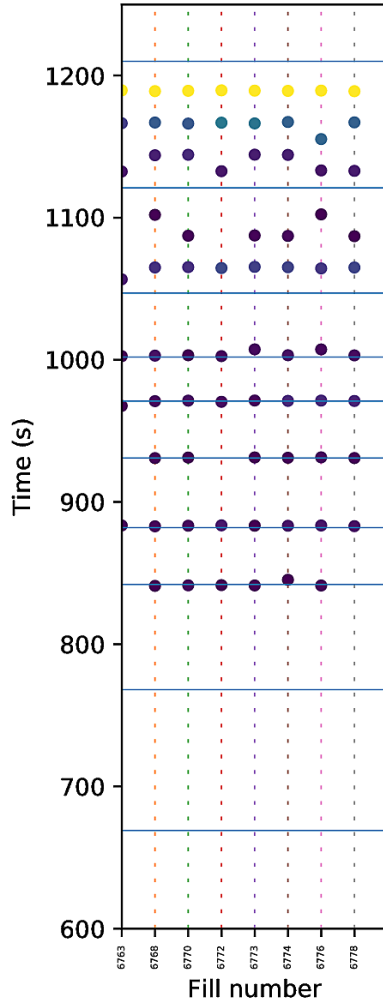
The orbit at the primary collimators was also studied. The same peak finder process was applied to the interpolated position of beam at the primary collimator. More specifically, the value shown is the interpolated absolute offset from reference orbit of the beams, in both the horizontal and vertical planes, at the primary collimator of IP7. This signal is gotten by linearly interpolating the offset orbit from the two closest BPMs before and after the IP7 collimator. In the same way as previously, the signals are scaled to the same range and a peak finding pass is done, Figure 20. In the horizontal plane, top row, we observe similar systematic peaks in the position offset. However, the timing offset between the two beams is much less pronounced, the only significant timing discrepancy is for the peaks around 1000s which means that contrary to the beam losses the orbit drifts are small and occur in the same way for Beam 1 and Beam 2. For the vertical plane, bottom row, the values seem random which indicates no significant orbit drifts are measured. For Beam 2 the values in the logging were not physical. This is due to a faulty BPM corrupting the Beam 2 interpolation. However, we can infer from the Beam 1 signal, that a vertical plane offset is not the culprit of these losses.

6 LHC and Machine Learning

6.1 The 2017 data

The aim is to use Machine Learning to train models of the LHC capable of predicting the lifetimes of the beams given a set of input parameters. The models are trained with a supervised

BLM_B1 Times of peaks Vs fill number



BLM_B2 Times of peaks Vs fill number

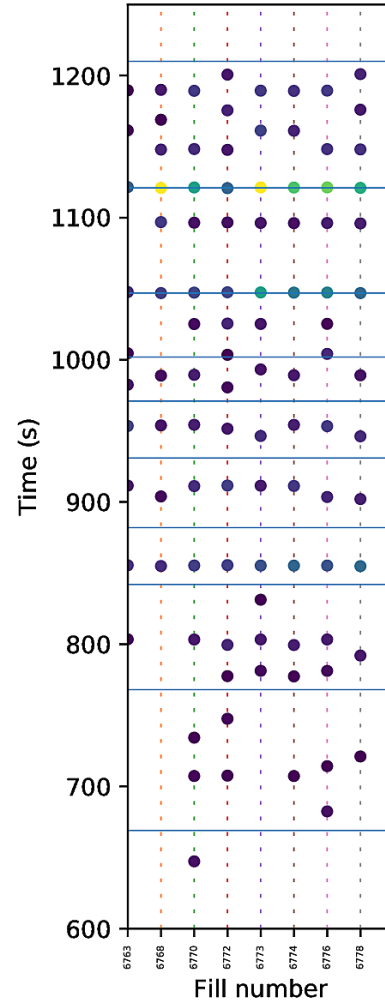


Figure 19: BLM signal peaks along with matching points for different fills, the colour of the points indicates the amplitude of the peak

learning method on LHC experimental data, more specifically, on data produced from 2017-09-04 and until the end of 2017. With hopes of applying said model on this year's data. The data is comprised of 183 machine fills which gives us around 40000 data points. When training a model, a portion of the data is kept from the training process to validate the model. This is done to avoid the model simply memorising the input/output couples and overfitting, instead of determining the relationship between them. Accordingly, the data is split randomly, 60% for the training set and 40% for the test set. Moreover, to keep the input/output relationship as simple as possible, the data in this study is taken from the injection PRERAMP beam mode of the machine. To avoid any problems caused by the different refresh rate of the measuring devices resulting in different number of points for each feature, all the features are synchronised to slowest refreshing feature, roughly 1 measurement per second.

The first, tentative, approach was to give as much information as possible to the model. Thus, an information complete dataset was created. This dataset covers the time period previously described, the input features are as follows :

- Tunes H/V/B1/B2
- Chromaticities H/V/B1/B2

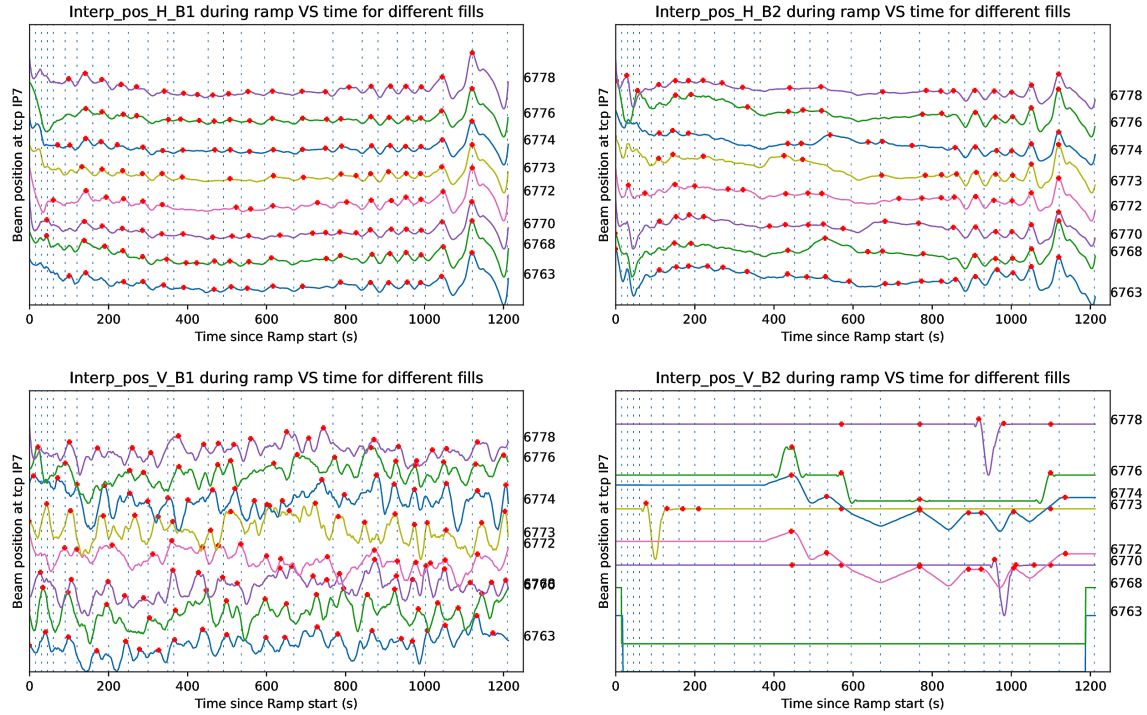


Figure 20: Time evolution of the beam position at IR 7 for different fills, top : horizontal position, bottom : vertical position, left : beam 1, right : beam 2, vertical dashed lines are the optical matching point timings

- Beam intensity B1/B2
- Beam position B1/B2
- Number of bunches
- Heat load²
- Octupole current B1/B2
- Beam losses B1/B2
- Emittances H/V/B1/B2
- Time elapsed since start of PRERAMP

The beam loss measurements are gotten with the same BLMs as used previously, two BLMs for each beam, one at each primary collimator.

The values of average emittances are gotten through the Beam Synchrotron Radiation Telescope (BSRT)[30] which takes periodic measurements of the emittance of a rolling subset of bunches in the beams. Thus, for our purposes, we extrapolate the gaps for each bunch and then average over all the bunches. This gives us an idea of the beam emittance.

The beam position data is calculated by averaging the offset between the beam and the ideal orbit throughout the machine by using the signal from the BPMs positioned throughout the machine.

To reiterate, the values we want the model to predict are the beam lifetimes.

²Heat load is a measure of the heat extracted through the cryogenics. It is a qualitative metric of the presence of electron clouds in the beam pipe. Electron clouds create instabilities and ultimately particle losses.

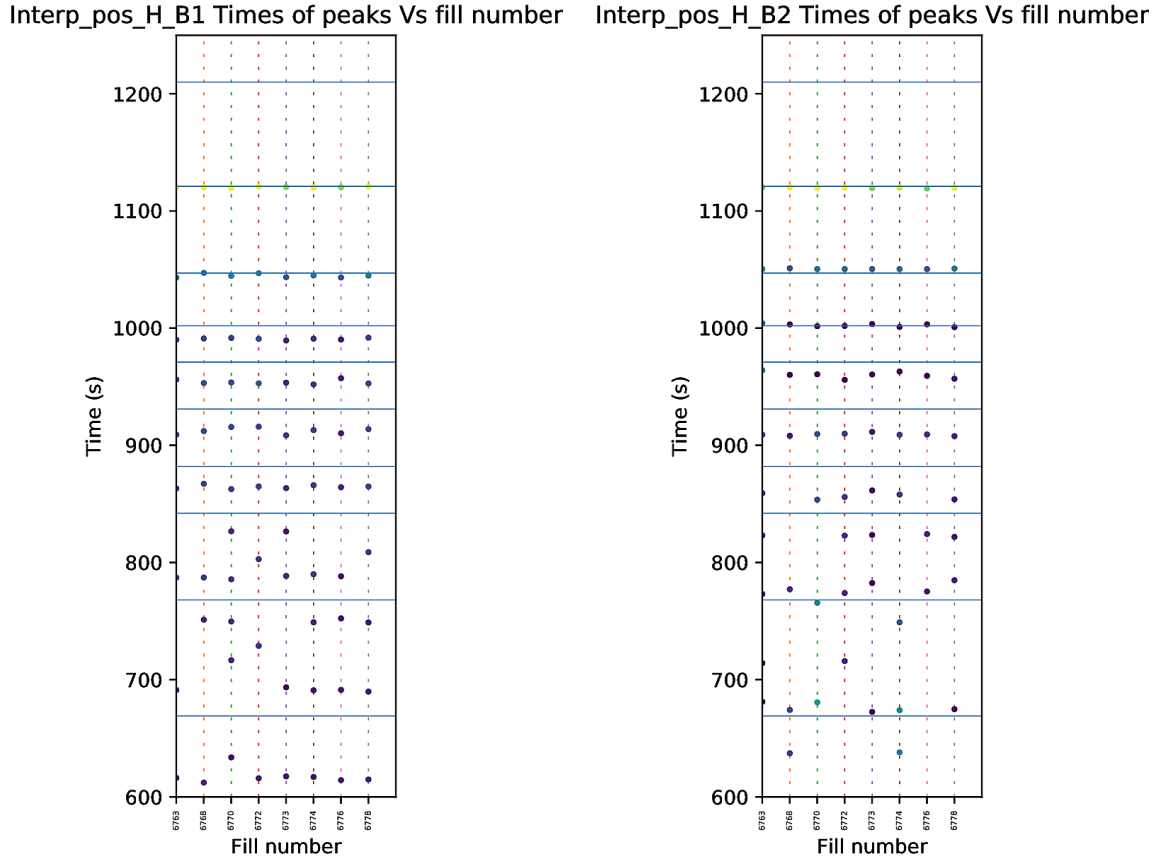


Figure 21: BLM signal peaks along with matching points for different fills, the colour of the points signifies the amplitude of the peak

With this dataset it is expected that the models should predict the lifetime with good accuracy since the information of all lifetime components is included in the model meaning that it should be able to infer the relationship described in Equation 26. The main components are the total beam intensity and the beam losses. We expect the beam losses to be the most important features for lifetime optimization, therefore minimising the beam losses will optimize the machine performance.

Figures 41 and 42, in Appendix A show a simple visualisation of the selected dataset for Beam 1 and Beam 2. The beam lifetime of each beam is plotted against each of the initially selected features, both axis are log scaled. This provides an general overview of the dataset trends. In addition to the lifetime dependence with the beam losses (by definition) one can observe also a dependence with the beam horizontal tune, especially for Beam 1, it seems that higher tunes lead to better lifetimes. The second parameter that has a direct correlation with the beam lifetime are the beam emittances. The dataset shows two major groupings, emittances of about $1.3 \mu\text{m}$ and $2 \mu\text{m}$. The smaller emittances lead to larger lifetimes. This is expected as lower emittances lead to smaller beams thus less losses at the small aperture of the collimators. For the rest, it is not obvious to infer any significant correlation with this qualitative visualisation method. In particular because not all parameters have been scanned.

6.2 Preprocessing

When dealing with widely varying feature magnitudes, it is essential to equalise the importance of each of the parameters to the model. Without this, features with higher order of magnitude values will have an over estimated impact on the lifetime. In the studied case, the beam intensities

are of the order of 10^{14} protons and the fraction beam tunes of 10^{-3} , certain preprocessing must occur to account for these. The elimination of this bias is done by scaling all the features to the same range, between 0 and 1.

Accordingly, for each feature, the following transformation is applied.

$$\hat{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (27)$$

The importance of this step is illustrated by Principal Component Analysis[31] (PCA). PCA is a statistical method for determining new orthogonal components, from linear combinations of the variables, which maximise the variance of the data along these components. By looking at the composition of these new components, it is possible to determine which of the input features contribute most to the variance of the dataset. The PCA is done both before and after the scaling of the data, the results of which are shown in Figure 22.

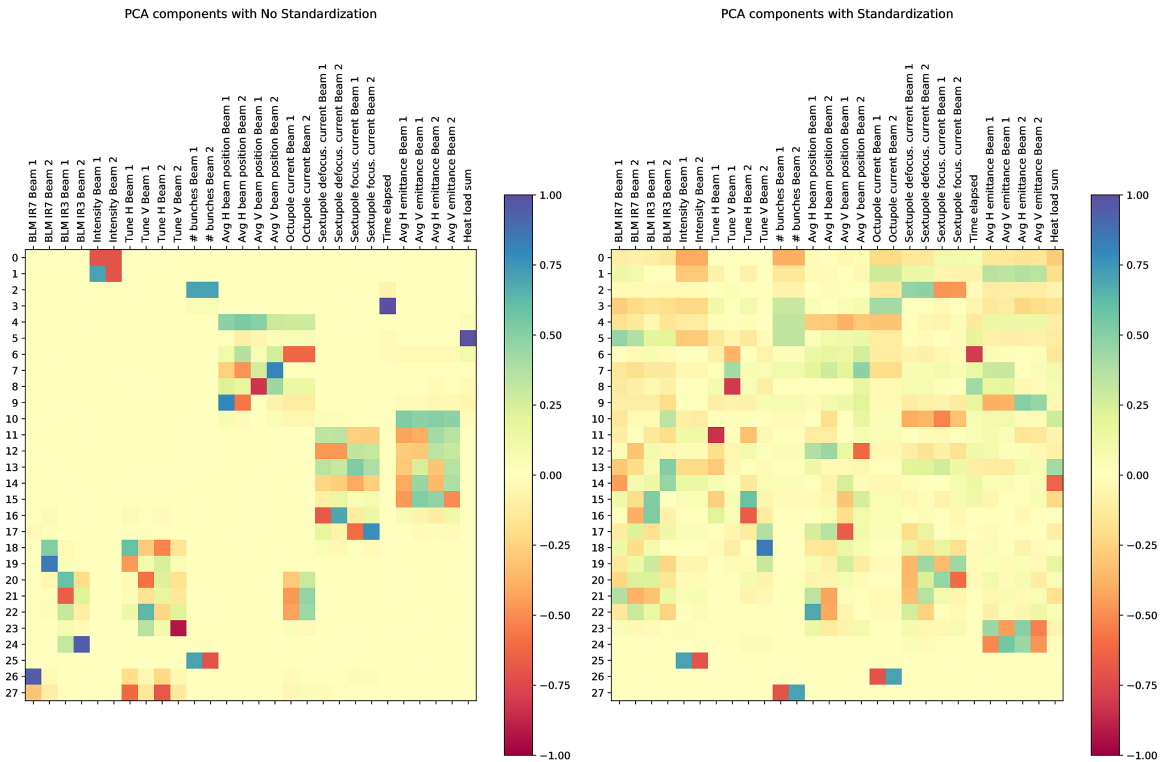


Figure 22: Result of PCA analysis

Figure 22 shows the result of the PCA analysis without preprocessing, on the left, and with preprocessing, on the right. These matrices show the composition of the components of the PCA ordered with respect to their contribution to the variance of the dataset. Without preprocessing, the first two components of the analysis, are only composed of the beam intensities. Whereas after preprocessing, these components are made up of many of the dataset's variables.

Figure 23 illustrates the contribution of each of the components in explaining the variance of the dataset. Without preprocessing, on the left, the entirety of the variance of the dataset can be explained with only the first PCA component, which is itself composed of solely the beam intensities. This indicates that the large magnitude of the beam intensity severely biases the analysis. After rescaling the features, on the right, we require roughly 25 PCA components to explain the entirety of the variance, with the components themselves composed of more of the variables. In addition to providing insight into the importance of scaling, PCA is also a rigorous method to reduce the

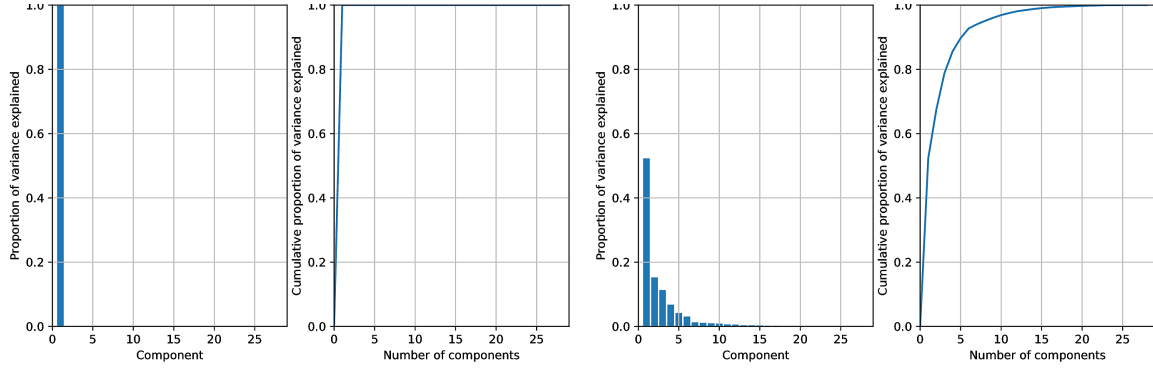


Figure 23: Impact of preprocessing on PCA

dimensionality of the problem. This is done by substituting the input features with a subset of the PCA components, however this was not deemed necessary in this study.

Before training any machine learning models on the now scaling bias free dataset, a statistical metric, the Spearman coefficient is used to better understand the interplay between the features. The Spearman coefficient assesses how strongly two features are related through a monotonic function. In the correlation matrix, Figure 24, a positive correlation indicates that the two variables can be described with a increasing monotonic function, whereas a negative value indicates a decreasing monotonic function.

From the correlation matrix, looking at the lifetime columns, we again see the expected negative correlation with BLM signals at IR3 and IR7. Secondly, we observe a positive correlation with the horizontal tunes, and not with the vertical ones, this is perhaps due to the range of tunes observed in the horizontal plane compared to the vertical, as shown in Figures 41,42 in appendix A. The number of bunches in the beams are negatively correlated with the lifetime, perhaps due to . Lastly, the negative correlation with emittance, confirms the trend previously observed.

6.3 Choosing the model - Model comparison

When it comes to machine learning there is an extensive choice of models. However, not all machine learning algorithms are suitable for a given dataset. In order to find the model best suited to our data, a selection of different models were trained and their performance on the test set compared using the Mean Squared Error (MSE) and the R^2 coefficient. The MSE is a good metric for prediction quality, the lower the MSE the better the prediction quality of the model. The R^2 coefficient is a rescaled version of MSE, independent of the variance of the dataset, it is bound between 0 and 1, the higher the value the better. For a set of values y_i and predictions $\hat{y}_i$ with $i \in [1, n]$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, MSE and R^2 are defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2 \quad R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (28)$$

Most of the models used are from the scikit-learn [32] python library, which provides a robust and open-source API for machine learning models.

The models considered are the following :

- Generalised linear models :

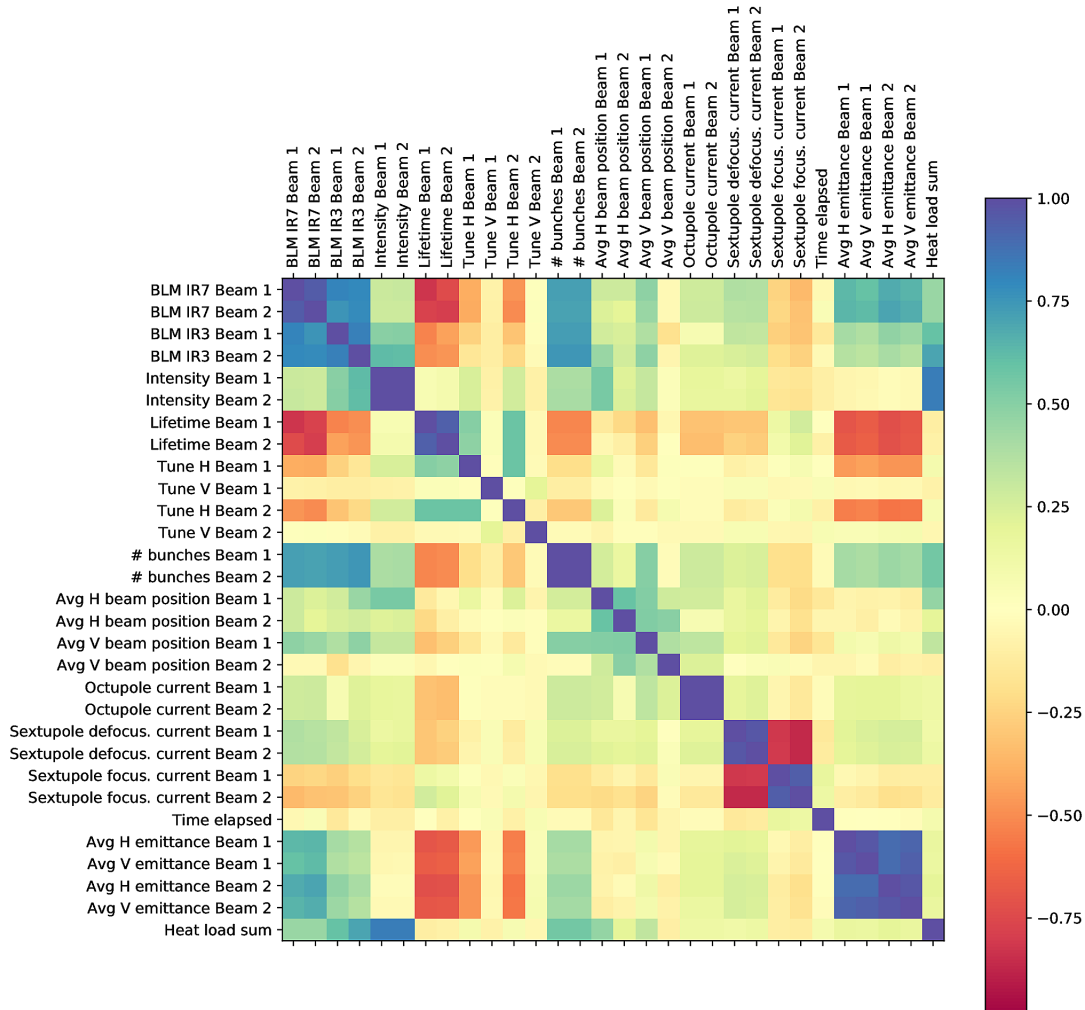


Figure 24: Spearman Correlation matrix for 2017 data

- Ordinary Least Squares[33] : is the simplest of the models, it determines the coefficients $\omega = (\omega_1, \dots, \omega_n)$ which minimise the l-2 norm $\min_w \|X\omega - y\|_2^2$. It can be seen as linear regression generalised to n dimensions.
- Ridge[34] : is similar to the Ordinary Least Squares method but introduces a complexity penalty on the ω coefficients, it solves : $\min_w \|X\omega - y\|_2^2 + \alpha \|\omega\|_2$ with α the penalty parameter. By introducing this complexity penalty it favours smaller values of the ω coefficients, improving the stability of the predictions.
- Lasso[35] : very much like Ridge but using the l-1 norm for the complexity parameter term, as opposed to the l-2, it solves : $\min_w \frac{1}{2n_{\text{samples}}} \|X\omega - y\|_2^2 + \alpha \|\omega\|_1$. The use of the l-1 norm results in a tendency to favour fewer coefficients thus effectively reducing the dimensionality of the data.
- Elastic Net[36] : is a combination of both Ridge and Lasso : $\min_w \frac{1}{2n_{\text{samples}}} \|X\omega - y\|_2^2 + \alpha \rho \|\omega\|_1 + \frac{\alpha(1-\rho)}{2} \|\omega\|_2^2$ with ρ the ratio of l-1/l-2 regularisation. When properly tuned this model should be able to utilise the advantages of both Ridge and Lasso.
- Neural network :
 - Multilayer Perceptron[37] : a fully connected neural network capable of learning non linear relations. Requires considerable tuning as the topology of the network is of utmost importance.

- Tree based models :
 - Random Forest[38] : an ensemble of decision trees each trained on a bootstrap sample of the training dataset. Their predictions are then averaged.
 - Gradient Boosted Decision Trees (GBDT)[39] : see section 6.4.

Model	Beam 1		Beam 2	
	MSE	R^2	MSE	R^2
Ordinary Least Squares	9482.435	0.666	12107.742	0.604
Ridge	9521.182	0.664	12129.580	0.603
Lasso	9773.588	0.655	12302.786	0.597
Elastic Net	24376.165	0.141	25742.933	0.158
Multilayer Perceptron	4856.423	0.833	7571.265	0.759
Random Forest	101.971	0.996	58.612	0.998
Gradient Boosted Decision Trees	76.577	0.997	54.411	0.998

Table 1: Model performances

From this comparison, Table 1, the best performing model is the Gradient Boosted Decision Tree. This model is widely used in high performance machine learning models, it has been a staple of machine learning competitions since its emergence. The GBDT algorithm used is from Microsoft's LightGBM python package[40].

6.4 Gradient Boosted Decision Tree

6.4.1 Decision trees

Decision trees are a machine learning method revolving around the creation of a series of conditions relating the input features to the outputs. To better understand the algorithm behind these trees, let us consider the following example. We want to predict the time spent outside.

outlook	temp.	humidity	windy	time
rainy	hot	high	false	25
rainy	hot	high	true	30
overcast	hot	high	false	46
sunny	mild	high	false	45
sunny	cool	normal	false	52
sunny	cool	normal	true	23
overcast	cool	normal	true	43
rainy	mild	high	false	35
rainy	cool	normal	false	38
sunny	mild	normal	false	46
rainy	mild	normal	true	48
overcast	mild	high	true	52
overcast	hot	normal	false	44
sunny	mild	high	true	30

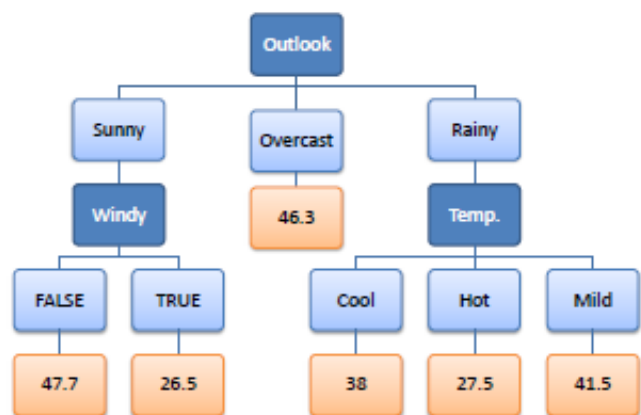


Figure 25: Example dataset and fully grown tree

In our case we are training a decision tree for regression, i.e with continuous outputs, however, the decision tree model can also be applied to classification problems. Once trained, we obtain a decision

tree composed of decision nodes, where the decisions take place (outlook, windy and humidity) and leaf nodes which are the decisions. The building of decision trees relies on efficient algorithms, the most widely used method is with the ID3[41] algorithm. This algorithm determines the optimal features on which to split the tree. This is done by using the standard deviation of the target (Y), in this case, the standard deviation of the time, is calculated as shown in Equation 29.

Time	25	30	46	45	52	23	43	35	38	46	48	52	44	30
------	----	----	----	----	----	----	----	----	----	----	----	----	----	----

$$S(Y) = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x})^2}{n}} \quad (29)$$

$$S(Y) = 9.321$$

In addition to the standard deviation of the target, ID3 uses the standard deviation of the target split along a feature, the time/outlook calculation is shown in Equation 30.

		STD(Time)	Count
Outlook	sunny	10.87	5
	overcast	3.49	4
	rainy	7.78	5

$$S(Y, X) = \sum_{c \in X} P(c) S(c)$$

$$S(Time, Outl.) = P(sunny)10.87 + P(overcast)3.49 + P(rainy)7.78$$

$$= \frac{5}{14}10.87 + \frac{4}{14}3.49 + \frac{5}{14}7.78 = 7.658 \quad (30)$$

The reduction in standard deviation of a split on feature X is defined as $SDR(T, X) = S(Y) - S(Y, X)$. At each step the algorithm determines the feature with maximal SDR and splits the tree with respect to said feature. The algorithm proceeds recursively on each branch. When a branch's coefficient of variation $CV = \frac{S}{\bar{x}}$ is below a threshold, it no longer needs splitting and becomes a leaf node, the value of which is the average of its subset.

6.4.2 Gradient Boosting

Gradient boosting is an algorithm which uses a combination of weak learners, models only slightly correlated with the true target, to build an high performing model, a so called strong learner. This is done by successively training models on the residual of the previous one. For example, let us consider the same example dataset as above, Figure 25, with the resulting fully grown tree, tree 1 in Table 2.

From this initial tree, the residual is used to build a second decision tree, tree 2, using the same process as previously described, but with the residual of the first as the target. This process is iterated as many times as needed, each model training on the residual of the previous model. These weak learners are then combined, building a strong model. This explanation is a simplified view of gradient boosting, in reality, the successive weak learners are not fit to the residuals of the predictions but to the components of the gradient of a loss function with respect to the prediction. Moreover, this gradient boosting method is not limited to decision trees, it can be applied to any weak learner. In practice, the growth of the tree is controlled by imposing constraints on the trees, this is done to limit overfitting and improve the model's generalisation to independent datasets.

sample #	time	Tree 1		Tree 2	combined model
		prediction	residual	predicted residual	prediction
1	25	27.5	-2.5	-2	25.5
2	30	27.5	2.5	3	30.5
3	46	46.3	-0.3	-0.5	45.8
4	45	47.7	-2.7	-2	45.7
5	52	47.7	4.3	5	52.7
6	23	26.5	-3.5	-3.5	23
7	43	46.3	-3.3	-3.5	42.8
8	35	41.5	-6.5	-6.5	35
9	38	38	0	-0.5	37.5
10	46	47.7	-1.7	-2	45.7
11	48	41.5	6.5	6	47.5
12	52	46.3	5.7	6	52.3
13	44	46.3	-2.3	-2	44.3
14	30	26.5	3.5	3	29.5

Table 2: Gradient Boosting example

6.5 Training process

6.5.1 Hyperparameter tuning

In order for a model to perform well on a dataset, its hyperparameters must be adequately tuned. Hyperparameters are variables specific to each model, which greatly impact the building of the models, there is no global optimal set of hyperparameters and no analytical method of determining the best set. Hyperparameters are tuned on a case by case basis by empirically determining the best settings. A tuned model must not be overfit to the training sample, and show high predictive power on the test set. These parameters are specific to a given model and can vary from one implementation to another. In the case of the LightGBM's gradient boosted decision tree, the list of considered parameters for tuning are the following :

- `max_depth` : limits the maximum depth of the tree.
- `num_leaves` : the maximum number of leaves on a tree.
- `min_data_in_leaf` : the minimum amount of data samples to form a leaf.
- `learning_rate` : the learning rate of the model.
- `bagging_freq` : the bagging frequency, bagging is the creation of an alternative dataset by randomly picking from the original data with replacement.
- `bagging_fraction` : determines how much of the original data is to be resampled.
- `feature_fraction` : randomly select a fraction of the features at each iteration, is similar to bagging fraction but for the features.

The tuning was done using random search to obtain a broad set of parameters which is refined through grid searching. When tuning the hyperparameters, the performance of each model is evaluated through 3 fold cross-validation over the entire dataset.

Cross validation is a robust method for evaluating a model's generalised performance. In k-fold cross validation the entire dataset is split into k random and equal size subsets. One of these k subsets is chosen as the validation set and the (k-1) others are the training set. This process is iterated k times, thus each fold is used for validation exactly once. The scores are average over the iterations.

The best hyperparameters are found to be :

- max_depth : 5
- num_leaves : 31
- min_data_in_leaf : 10
- learning_rate : 0.01
- bagging_freq : 1
- bagging_fraction : 0.7
- feature_fraction : 0.7

6.5.2 Training of the model

The GBDT models are trained until their performance does not improve for 1000 iterations. The model is expected to perform better on the training set than the test set. However, if the disparity is 'large' this would be a sign of overfitting, and if the prediction performance is hindered, the hyperparameters would have to be adjusted accordingly.

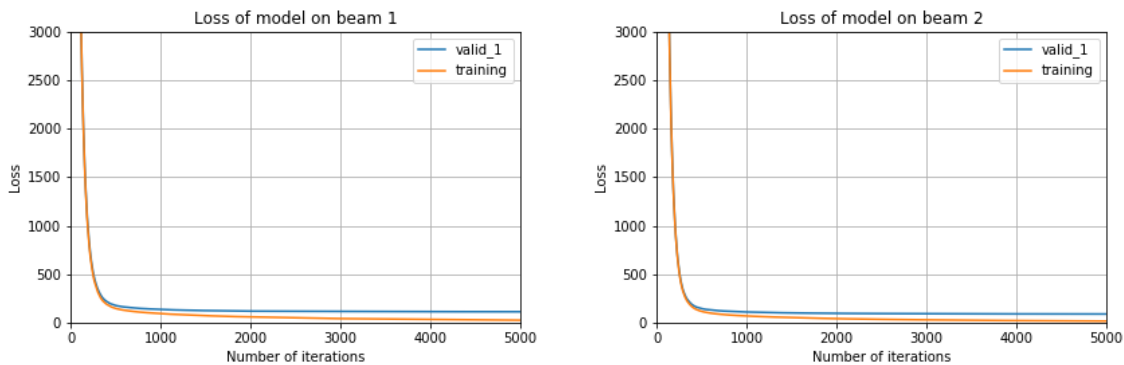


Figure 26: Loss function evolution beam 1 & 2, complete model on 2017 data

For both beams, we observe only slight disparities between the performance on both the training and test sets, overfitting does not seem to be an issue. Moreover, both models have converged successfully.

6.5.3 Performance of the model

As shown previously in Table 1, the GBDT model performs very well on this dataset, as expected given its completeness. In Figure 27 are plotted the models' predictions along with the true values of lifetimes. The lifetimes are sorted and plotted against their sample number. The predicted lifetimes agree with the measured beam lifetimes for both beam 1 and beam 2, with occasional outliers. The model seems to reliably predict the beam lifetimes.

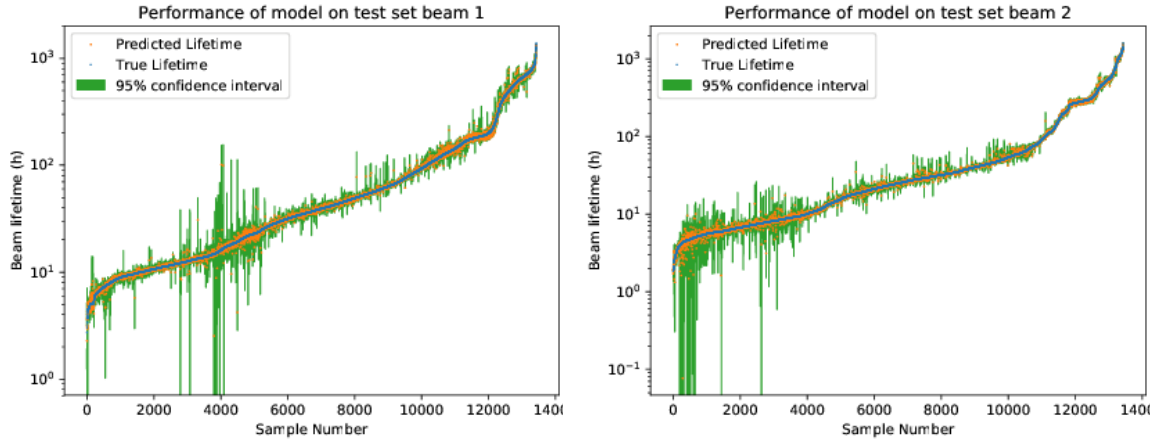


Figure 27: Model predictions beam 1 & 2, complete model on 2017 data

An estimation of the confidence interval is gotten through bootstrapping[42]. Bootstrapping is a statistical technique to estimate a statistic of a population by measuring the same statistic on the bootstrap sample of said population. Bootstrapping consists in resampling a dataset with replacement. In our case, the training set was resampled 200 times, a model is trained on each of these bootstrapped samples. The predictions of the test set lifetimes of each of these trained models are used to broadly estimate the 95% confidence interval by using the 2.5% and 97.5% percentile values of the set of predictions of each sample. The main drawback of this method is the computationally expensive training of models for every bootstrap sample. This imposes a practical limit on the number of bootstrap samples, generally a few hundred samples is considered respectable. The 95% confidence interval of the predictions seem reasonable in most cases, however, there seems to be some divergence for lower lifetime values. The cause of this divergence is unknown.

6.5.4 Feature importance

An additional advantage of using a tree based method is that it does not fall victim to the fundamental flaw of many machine learning techniques, especially neural networks, where the model becomes an impenetrable black box. With tree based models we are able to extract, from the topology of the trees, the importance of each feature in the decision making process. This can give valuable insight into the process being modelled. The importance of a feature is determined by simply counting the number of splits occurring on said feature, the results are shown in Figure 28. For both beams, the front runners in terms of importance are the BLM signals at IR7 of the respective beam, followed by the tunes, horizontal and vertical, of both beams. This observed dependence on the other beam's tune is perhaps an artefact the fact that, both beam often undergo the same tune changes at the same times, which the model could be wrongfully interpreting as significant. These findings tend to agree with the Spearman correlation matrix, however, when looking at the emitances, their significance is drastically lower with this metric than with the Spearman matrix, the same can be said for the number of bunches.

With the Spearman matrix, the correlation of two features do not depend on the correlation of the others, meaning the presence of a high correlation with some features does not diminish the correlation elsewhere. This is not the case with the feature importance as seen by the model. If a feature has a high importance then it makes up many of the split in the trees leaving less for the other features. A high importance feature can drown out the relevance of others. This could be the cause for the discrepancies between the methods.

The trends observed with these models do comply with the physics of the beams and is an encour-

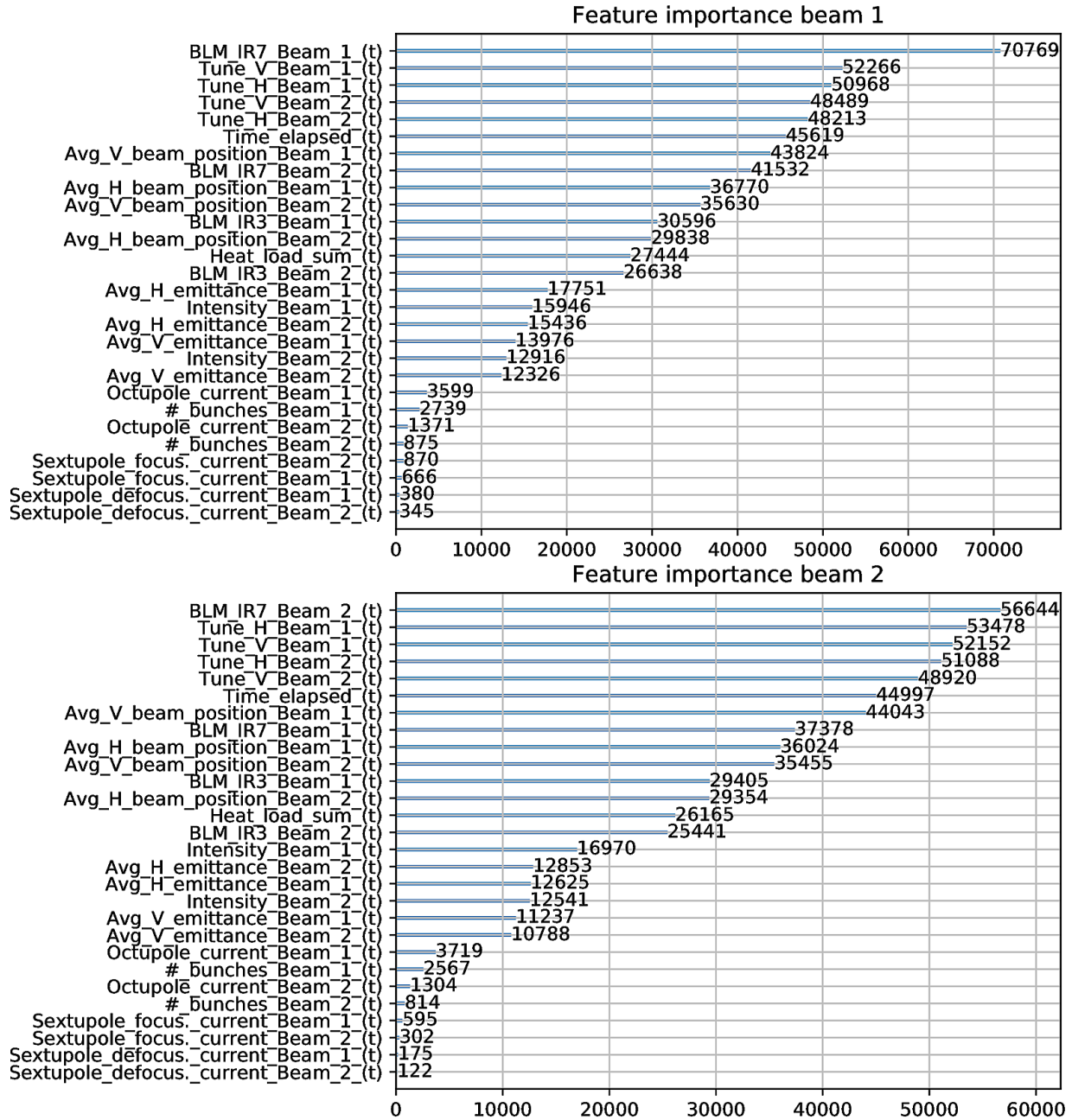


Figure 28: Feature importance for beam 1 & 2, complete model on 2017 data

aging first step. However, these models are not of great utility as they depend heavily on the BLM measurements which can be used to analytically calculate the lifetime. Moreover, as it requires many purely measured quantities as inputs it cannot easily be used to determine novel working points. In practice the operator only controls a handful of beam/machine properties, referred to as knobs. These knobs are the tunes, chromaticities, octupole current and the filling scheme, i.e. number of bunches. To address this, an alternative dataset was used.

6.6 Model trained on bare 2017 data

Contrary to the first dataset, this "bare" dataset only uses the knobs as input features, in hopes of eliminating the dominance of the BLM measurements in the feature importance, allowing for more interesting trends to be extracted between the lifetimes and the knobs themselves. It covers the same time span as the previous dataset, from 2017-09-04 until the end of 2017.

The input features are :

- Tunes H/V/B1/B2
- Chromaticities H/V/B1/B2
- Number of bunches
- Octupole current B1/B2
- Time elapsed since start of PRERAMP

The same procedure is used to prepare the data and train the model. The Spearman correlation matrix for this dataset is recalled below. As expected, the trends observed are identical to the complete feature set previously shown. With this metric, the main features impacting the lifetimes are the horizontal tunes of the beams and the number of bunches.

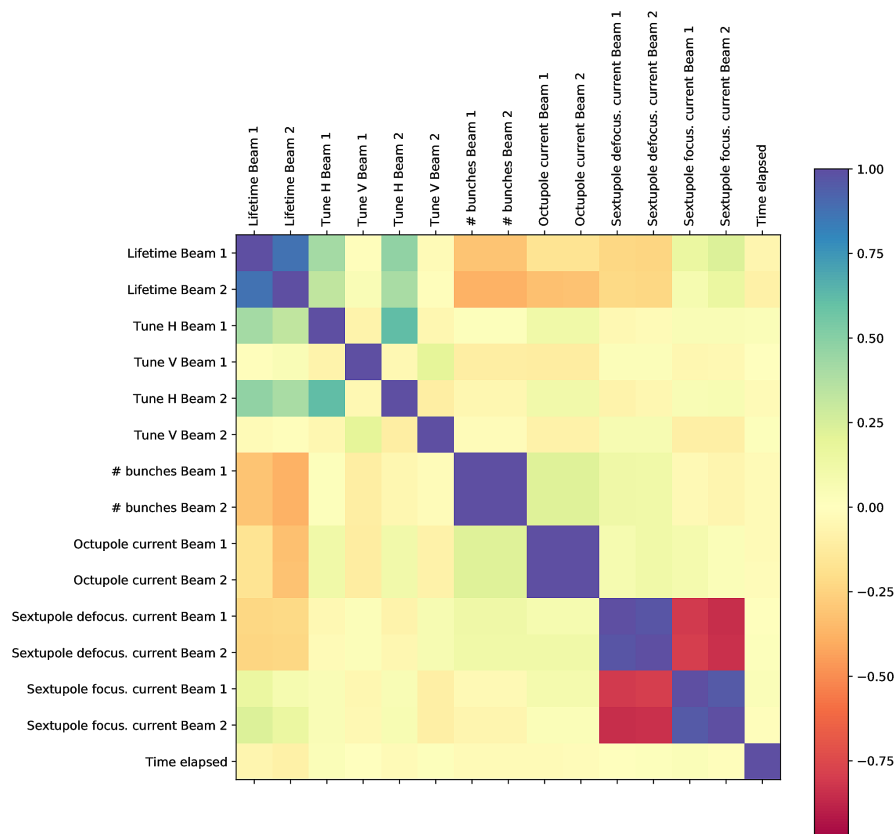


Figure 29: Spearman Correlation matrix for 2017 data

After tuning the hyperparameters using random and grid searching, the best set was determined to be:

- max_depth : -1 (no limit)
- num_leaves : 200
- min_data_in_leaf : 17
- learning_rate : 0.01
- bagging_freq : 1

- bagging_fraction : 0.9
- feature_fraction : 0.9

The value of the l-2 loss function during the training process is shown in Figure 30. The evolution is similar to Figure 26, however, there is a bigger disparity between the training and the test set. This indicates presence of overfitting. Moreover, the disparity is much stronger in the beam 2 model, this would indicate that the dataset is less applicable to beam 2 and perhaps is an indicator of dissimilarities between the two beams. Nonetheless, the performance on both beams are similar, Figure 31. The predictions correctly follow the trends of the data set. However, there is a drastically bigger spread in values, as shown by the larger confidence interval. Moreover, the number of outliers has significantly increased. This reduction in accuracy is expected as the model is given less information compared to previously, resulting in worse, but still respectable, MSE and R-2 coefficients, table 3.

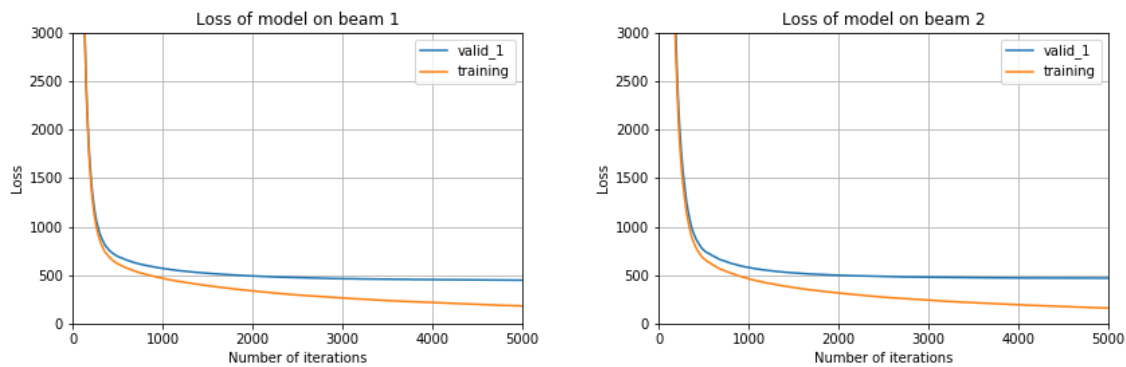


Figure 30: Loss function evolution beam 1 & 2

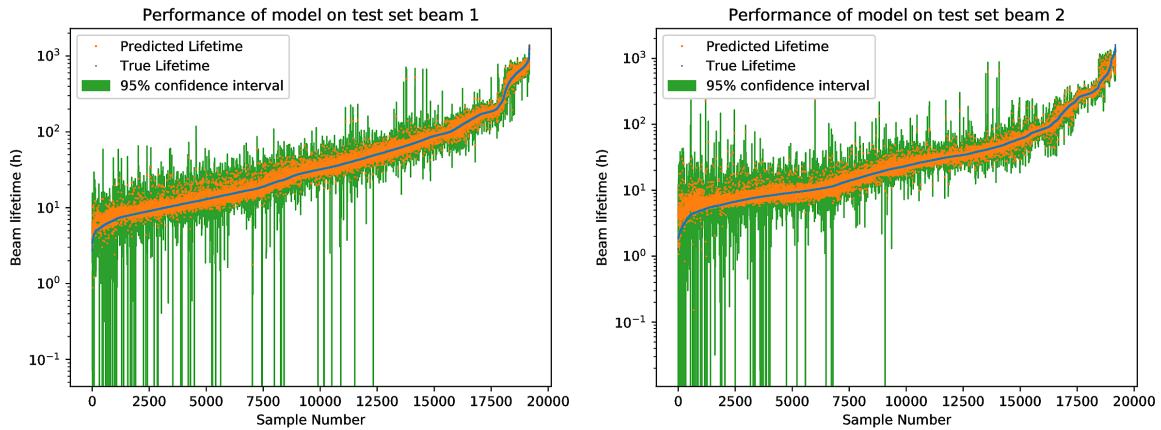


Figure 31: Model predictions beam 1 & 2, bare model on 2017 data

Model	Beam 1		Beam 2	
	MSE	R^2	MSE	R^2
Bare dataset 2017	393.973	0.98	1040.9	0.95

Table 3: Performance of model, bare model on 2017 data

From this model, we can once again get the feature importances, Figure 32, comparing this against the importances in the complete model, Figure 28, the time elapsed feature, which is the time since the start of the PRERAMP beam mode, has significantly grown in importance, it has overtaken the tunes. The dominance of the time elapsed is surprising and difficult to interpret. In the PRERAMP

mode, the LHC operators manually trigger the steps in the sequence which change in machine configuration, thus there is no consistency in the time evolution of the steps which could explain this dependence. This is not the case for other beam modes, such as the ramp where the machine undergoes a series of preprogrammed changes, consistent timing wise, for all fills. The time elapsed feature does not give any obvious information to the model but, nonetheless, is an essential part in the building of a decently performing model. Further, investigation must be done to understand this phenomenon.

Aside for the time elapsed, for beam 1, the tunes have remained in the upper ranks of importance, followed by the sextupole, which, is a drastic increase compared to Figure 28 where the sextupole currents are of least importance. This is even more pronounced for beam 2 where some sextupole currents overtake the vertical tunes. Removing the BLM measurements, and using this reduced feature set, has allowed the model to learn more subtle trends within the data.

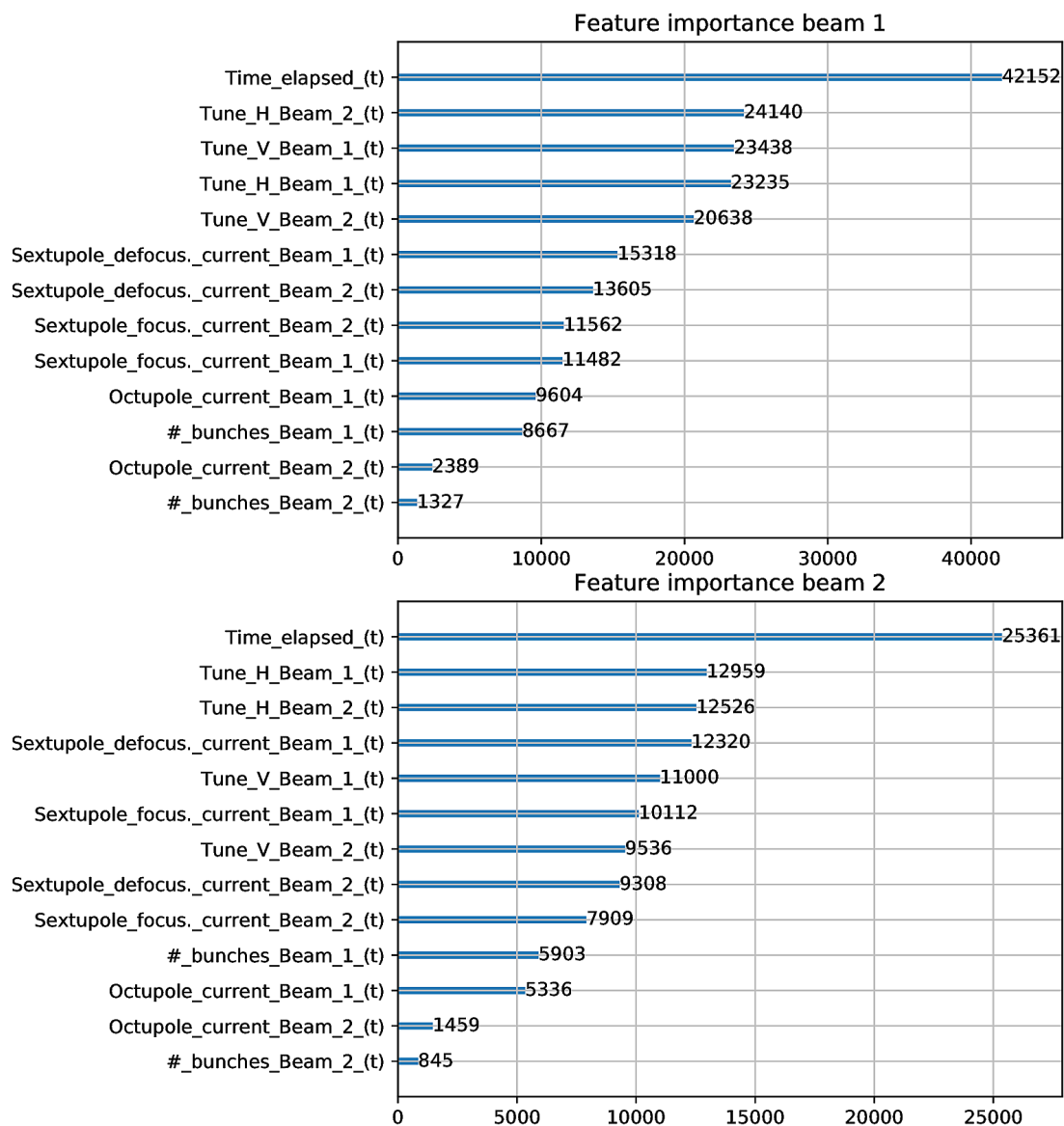


Figure 32: Feature importance for beam 1 & 2, bare model on 2017 data

Interestingly, when the model is trained on the tune values given by the BBQ system, as opposed to the BOFSU feedback system, the correlation with the lifetimes disappears. This is unexpected and merited some investigation as it could indicate some unwanted phenomena in either the BBQ or the feedback system, which could hinder the performance of the machine. The tune signals from both systems are plotted for the PRERAMP beam mode of fill 6762 in Figure 33. It shows, that the BBQ

signal is erratically oscillating between the BBQ value and a lower tune value. Figure 34, shows the distribution of the difference between the two signals, it indicates a consistent offset of roughly -0.0075 Hz between both signals. The lower values of the BBQ system are thought to be due to the peak finding algorithm used to extract the tune from the beam's frequency distribution choosing the wrong peak. These incorrect value are eliminated, through filtering, in the post-processing of the BBQ signal which explains why they are not observed in the BOFSU values. The tune diagrams of the the entire dataset for both signals are shown in Figures 50 in appendix D, the colour indicates the lifetime of the beam. The BBQ signal is significantly noisier and there are diagonal structures which are most likely due to the inconsistent peak finding discussed above. The lack of correlation when using the BBQ tunes is explained by both these factors. This illustrates the sensitivity of the training process with regards to the quality of the data provided. Noisy or nonsensical data will severely impact the learning process and drastically reduce the performance of the models.

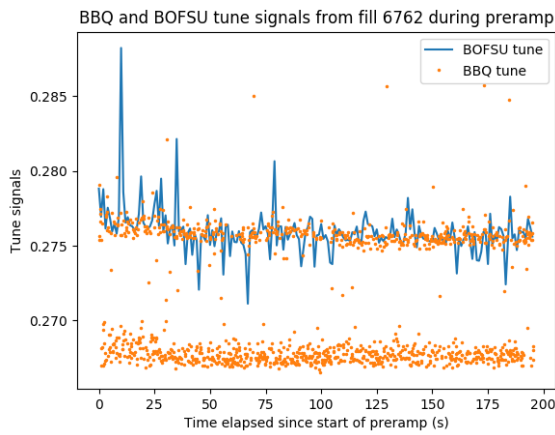


Figure 33: BBQ and BOFSU signals for PRE-RAMP mode of fill 6762

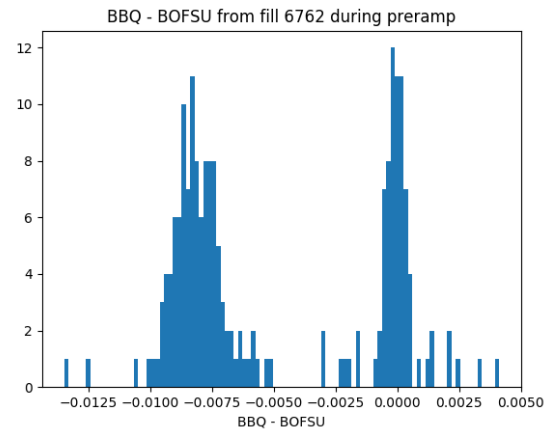


Figure 34: Distribution of signal difference between both systems for PRERAMP of fill 6762

6.7 Application to 2018 data

Every year, during the winter months, as the cost of electricity is higher, the LHC shuts down to undergo upgrades and fine-tuning. Changes on beam optics and machine parameters are also prepared and commissioned after these shutdowns in order to push for better performance. Thus, every year the behaviour of the beams could be slightly different. As such, the applicability of these 2017 models on 2018 data is non trivial. The 2017 model must be benchmarked on this year's data. Unfortunately, as shown in Figure 35, the trained 2017 models do not perform well on 2018 data.

This is a big limitation of this year per year approach, each year's model is not necessarily applicable to the next. Thus, a new model must be trained every year, which requires a large amount of data. Consequently, with this method no model can be used early on in the year as the dataset would most likely be too small to achieve good predictive performance. This issue could perhaps be bypassed by training on a dataset which encompasses many years worth of data to improve the cross year generalisation of the models, this avenue has not yet been explored in this study.

Nonetheless, a model has been successfully trained on 2018 data following the same method as previously described, as shown in appendix C.

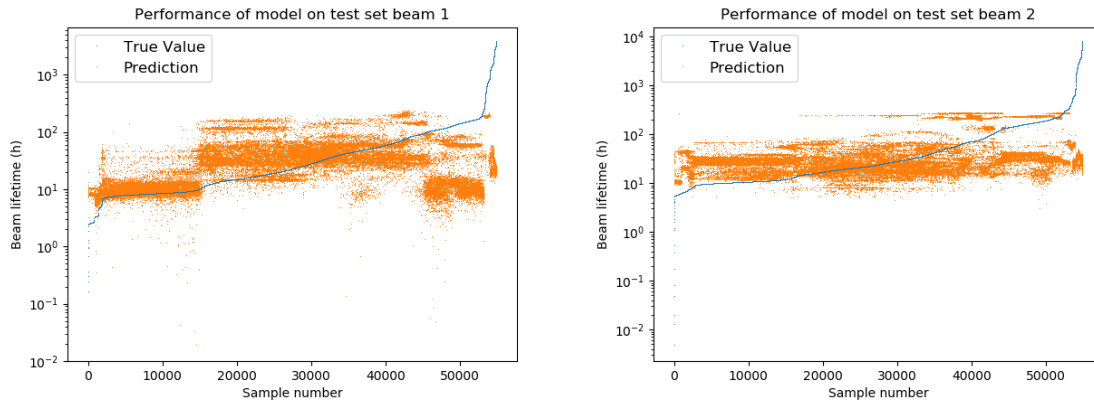


Figure 35: Bare 2017 model predicting 2018 lifetimes

6.8 Optimization proof of concept

Models such as these can potentially be applied to further the optimization of the machine. In this work, a proof of concept is presented. The goal is to use the bare model trained on 2018 data to determine the optimal knob settings which maximise the lifetime of the beam during the PRERAMP beam mode for an arbitrary machine fill, number 7056. To achieve this, the trained model is coupled with a Nelder-Mead[43] optimization algorithm. The optimization algorithm is configured to maximise the lifetime, by exploring the input feature space of the model. The starting point of the Nelder-Mead algorithm is the measured working point of the fill, this optimization is done for each time step of the fill. Once completed, see Figure 36, we have the time evolution of the knobs which would, as determined by the model, maximise the beam lifetimes. With this technique, it is also possible to impose constraints on individual features, as is done for the number of bunches. The number of bunches of both beams is constant for a given fill, thus, to stop the model from naively tuning this feature, they are imposed constant. Moreover, the feature evolutions are extremely erratic, which in practice would induce significant particle losses. Indeed, the model has no concept of smoothness, it is possible to apply a smoothness algorithm in postprocessing but this approach would not be feasible in online control of the machine and would also reduce the accuracy of the predicted lifetime. Moreover, as we have two models which are trained on the same dataset, they are both requesting different settings of the same knob. This conflict can be bypassed by minimising both models' lifetimes at the same time by configuring the optimization algorithm to maximise the average of both lifetimes for example. Keeping the optimization separate does have the advantage of allowing us to examine the model's recommendations for both beams independently. We can glean some information from this proof of concept. Firstly, the model predicts a significant improvement could be made at the start of the PRERAMP beam mode, and the lifetime gain would taper off in time. Secondly, it seems the model encourages higher horizontal tune values. The optimized beam 1 octupole currents seem to agree well with the measured values, however, the erratic nature of the optimization renders any further trends difficult to interpret. Lastly, the sextupole currents are of questionable utility as the absolute value is of little interest due to the feed forward system discussed earlier, and the erratic prediction renders any interpretation inconclusive. The fact that the model was able to extract the trends observed with the horizontal tunes observed previously is encouraging, however this optimization setup is rudimentary and requires additional work to fully understand the implications.

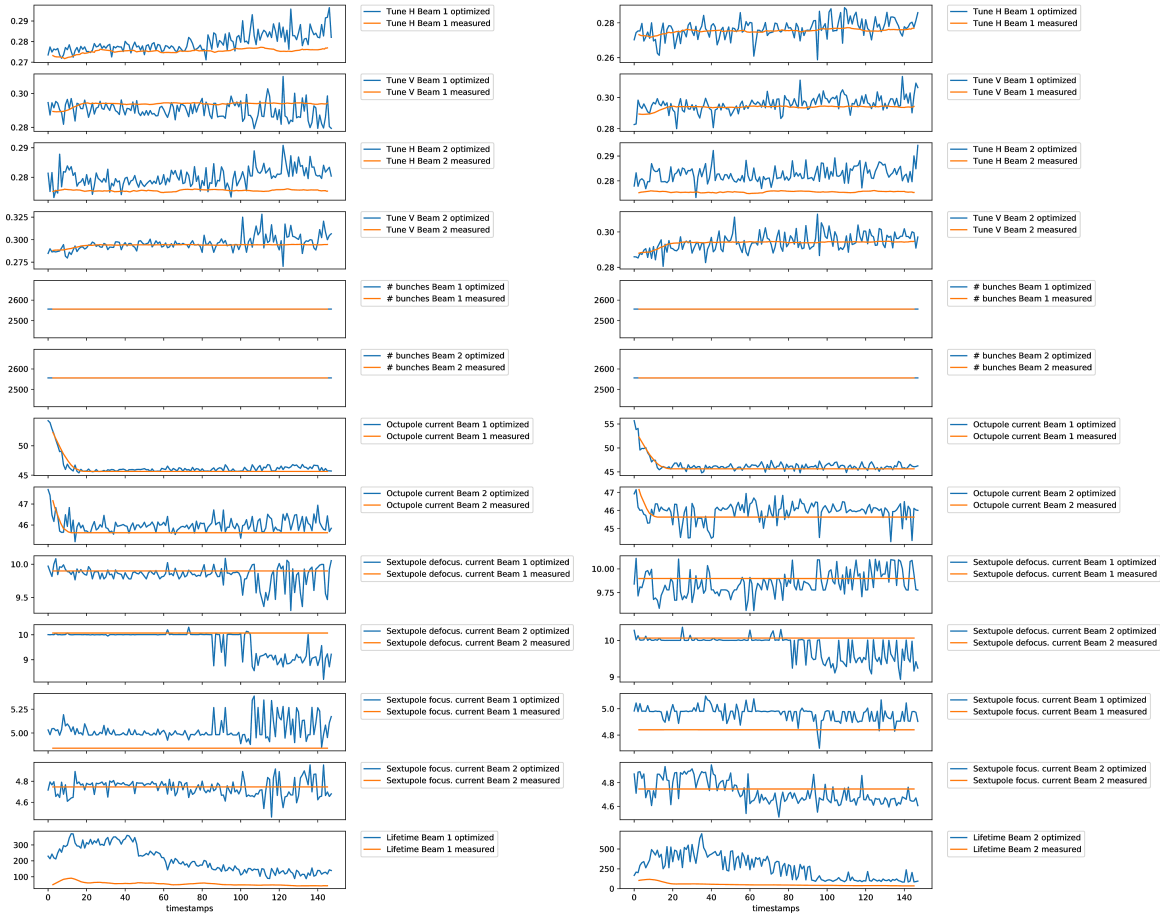


Figure 36: optimization using the trained bare 2018 models on fill 7056, beam 1 model on the right and beam 2 model on the left

6.9 Corroboration through simulations

In order to shed light on the physical reasons behind the models recommendations, some progress has been made in using long term particle tracking simulations to corroborate and elucidate the conclusions of the model. The tracking is done by using Mad-X[44] and Sixtrack[45]. Mad-X is used is provided the injection energy optics files of the machine and creates the necessary inputs required by Sixtrack. Sixtrack then performs the tracking of the particles over 10^6 turns. In our case only 4D tracking was used, as some as yet unexplained behaviour occurred when taking into account the momentum spread and RF cavity of the machine. From the Sixtrack simulations, it is possible to extract the Dynamic Aperture for the given parameters which can then be related to beam lifetime. In order to corroborate the trend observed, in which higher horizontal tunes result in better lifetimes, the tracking was done for different tune working points and octupole currents, based on the values observed in the machine learning dataset. Sixtrack is given an uniform particle distribution which occupies the available space between the collimator jaws at injection Figure 37 (left). The distribution is tracked through the machine and every 10^4 turns the resulting particles are Gaussian weighted to more accurately model the actual beam distribution Figure 37 (right). From this, we can estimate the losses by counting the number of particles outside of the collimator aperture. Using this in combination with the tune diagrams of the tracking one can hopefully explain the origin of the losses. However, as of yet, no pertinent results have been produced and more work must be done to fully understand the phenomena.

In addition to explaining the origin of the losses, simulation campaigns could be used to create additional data to train machine learning models. This would allow data to be generated while the

machine is offline. Moreover, a simulation based dataset could potentially be used to train a model which could serve as a rough substitute to computationally expensive dynamic aperture studies. Thus, severely cutting the simulation time to near instantaneous while providing physicists with decently accurate results. This in particular would be of great use in designing the Future Circular Collider, where the gained efficiency would be used to increase the resolution elsewhere.

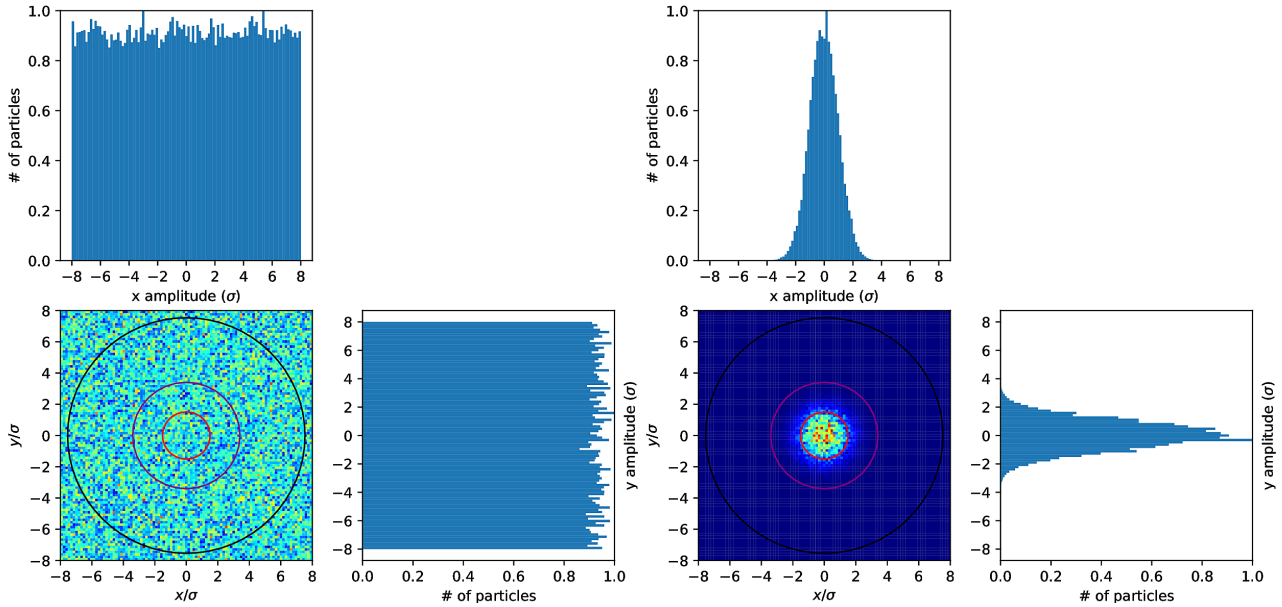


Figure 37: Initial uniform particle distribution (left), Gaussian weighted particle distribution (right), the red circle indicates the beam σ , the purple 3σ and the black the collimator aperture at roughly 7.5σ

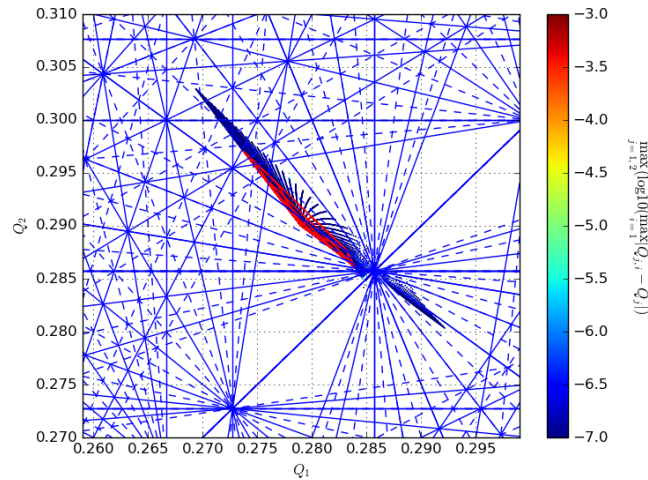


Figure 38: Tune diagram, Mad-X tracking in red, Sixtrack tracking blue points, for octupole current 12A

6.10 Time series prediction

An alternative technique for lifetime prediction has also been explored. This second method revolves around time series prediction. Previously, the model was built to predict the lifetime given as set of input features. However, in this section the model is given the lifetime and a set of input features and is built to predict the lifetime however many seconds in the future. If done properly, this

could be a valuable tool for LHC operators, as one might be able to predict the lifetime at collision from measurements at injection, allowing operator to fine tune the working point at injection to achieve the optimal collision configuration. In order to achieve a high quality time series predicting model, two approaches can be taken. Firstly, Long Short Term Memory (LSTM) unit, pioneered by Hochreiter et al. [46], in recurrent neural networks can be used as they have the ability of utilising previously seen data samples to influence their output. These networks are the standard choice for modelling natural language recognition, music composition and other applications where the output should depend on the previously seen sequence. Secondly, it is possible to reframe a supervised learning problem, as we have in our case, into a time series paradigm through feature engineering. The LSTM approach has been attempted unsuccessfully, as a result, the second method is presented.

6.10.1 Feature engineering and preliminary results

In order to achieve the desired training for our model the input data must be accordingly transformed. To better understand the process, let us consider the following mock dataset, with the target for prediction being feature 3, 2 seconds later.

time	feature 2	feature 3
1	10	0.251
2	15	0.256
3	11	0.245
4	9	0.236
5	14	0.234
6	16	0.235
7	53	0.247
8	40	0.234
9	23	0.248
10	12	0.230

time	feature 2	feature 3	feature 3(t+2)
-1	NaN	NaN	10
0	NaN	NaN	15
1	0.251	10	11
2	0.256	15	9
3	0.245	11	14
4	0.236	9	16
5	0.234	14	53
6	0.235	16	40
7	0.247	53	23
8	0.234	40	12
9	0.248	23	NaN
10	0.230	12	NaN

Table 4: Dataset pre (left) and post (right) transformation

The target feature is simply copied, and shifted the required number of time steps and concatenated to the dataset. This procedure, introduces numerical errors (NaNs) at the start and end of the dataset, the rows affected are removed. We sacrifice 2 datapoint for each time step but this is usually negligible compared to the number of datapoints in the set. In our case, as we are dealing with multiple machine fills, which are to be considered as independent times series, this transformation must be applied on each fill separately, we thus would lose $2 \times \text{number of fills}$ datapoints, which remains nonetheless negligible.

The first tentative application of this method, was to use the same 2018 dataset during the PRERAMP beam mode, the dataset is appropriately transformed so as to train a model capable of predicting the lifetime at $t+50s$. The results of which is shown in Figure 39. This encouraging first attempt yields a slightly stronger model, Table 5, but requires the initial measurement of beam lifetime.

	Beam 1		Beam 2	
Model	MSE	R^2	MSE	R^2
Bare dataset 2018 $t+50s$	761.70	0.993	1512.01	0.994

Table 5: Performance of model, bare model on 2018 data, predicting lifetime at $t+50s$

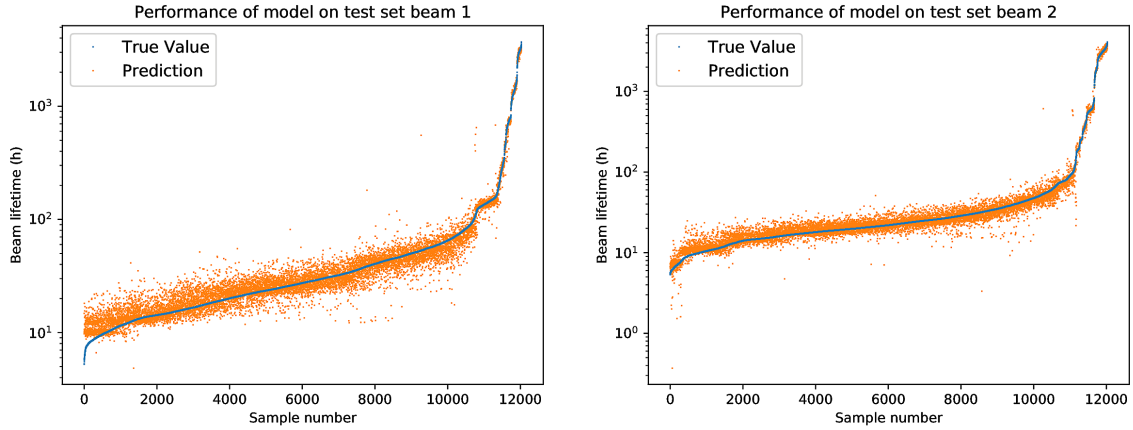


Figure 39: Model predictions beam 1 & 2, bare model on 2018 data at t+50s

However, models such as these, are of limited utility as they are only valid in the PRERAMP beam mode. Preliminary attempts have been made to train a model capable of crossing the beam modes, in the hope of being able to predict the lifetime at collision energies from inputs at injection energies. The dataset covers the same time period as the previously used 2018 dataset, however it is no longer limited to solely the PRERAMP beam mode, instead the data is acquired for the PRERAMP, RAMP, FLATTOP, SQUEEZE, ADJUST and STABLE beam modes, this dataset is comprised of 2762628 samples. The information of the beam mode is given to the model by one hot encoding the beam mode. The same transformation as described above is used to train a model capable of predicting the lifetime post RAMP from measurements at PRERAMP, by drastically increasing the time step offset to 3600s. The predictions of the model are of dubious relevance as it does seem to be able to reliably determine the trends across beam modes which impact the intensity lifetimes of the beams, Figure 40.

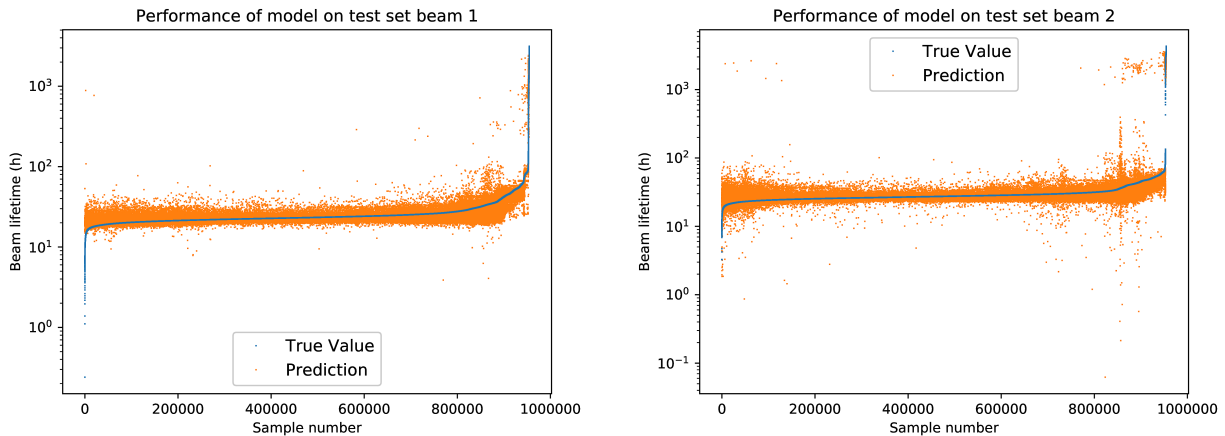


Figure 40: Model predictions beam 1 & 2, bare model on 2018 data at t+50s

7 Conclusion

This work has studied the losses occurring during the ramp up of energy and pre-squeeze at the LHC. It was found that the maximum peak of loss occurs for Beam 1 exactly at the optics match points whereas for Beam 2 it occurs in between match points. The orbit measurements show that no significant orbit drifts occur for both beams. Only small drifts appear similarly for Beam 1 and Beam 2 in the horizontal plane but are not large enough to explain the observed losses.

This study has shown that machine learning techniques are capable of extracting the trends necessary to predict with decent accuracy the lifetimes of both beams using a limited set of input parameters. It has served to elucidate the strengths and weaknesses of these models when applied to LHC data. Notably, the year per year hermiticity of these models which is a significant hindrance and merits to be addressed. These models can then be used for a variety of applications. In this work, we have explored the strength of combining such models with an off the shelf optimization algorithm. It demonstrates the potential of these models for optimization purposes. Some limited headway has been made in reframing the problem as a time series, to hopefully in the future train a model capable of predicting the collision lifetimes from injection inputs. Nonetheless, more work must be done to fully understand the implications of these techniques. Such as, combining these methods with simulation campaigns which would go a long way in demystifying their implications. Moreover, the same process could be applied to a simulation based dataset which would significantly help in the design process of the Future Circular Collider, which requires extremely computationally expensive simulation campaigns. However, it remains a solid first step in using machine learning models for LHC optimization.

References

- [1] *EPFL at a glance*. URL: <https://information.epfl.ch/glance> (visited on 08/21/2010).
- [2] *LPAP*. URL: <https://lpap.epfl.ch/> (visited on 08/21/2010).
- [3] *SwissFEL*. URL: <https://www.psi.ch/swissfel/> (visited on 08/21/2010).
- [4] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. URL: <https://cds.cern.ch/record/782076>.
- [5] *CERN*. URL: <https://home.cern/> (visited on 08/21/2010).
- [6] *Antiproton Decelerator*. URL: <https://home.cern/about/accelerators/antiproton-decelerator> (visited on 08/21/2010).
- [7] *LINAC 2*. URL: <https://home.cern/about/accelerators/linear-accelerator-2> (visited on 08/21/2010).
- [8] *Proton Synchrotron Booster*. URL: <https://home.cern/about/accelerators/proton-synchrotron-booster> (visited on 08/21/2010).
- [9] *Proton Synchrotron*. URL: <https://home.cern/about/accelerators/proton-synchrotron> (visited on 08/21/2010).
- [10] *Super Proton Synchrotron*. URL: <https://home.cern/about/accelerators/super-proton-synchrotron> (visited on 08/21/2010).
- [11] *Large Hadron Collider*. URL: <https://home.cern/topics/large-hadron-collider> (visited on 08/21/2010).
- [12] Esma Mobs. “The CERN accelerator complex. Complexe des accélérateurs du CERN”. In: (July 2016). General Photo. URL: <https://cds.cern.ch/record/2197559>.
- [13] *LArge Electron Posotron-Collider*. URL: <https://home.cern/about/accelerators/large-electron-positron-collider> (visited on 08/21/2010).
- [14] *Radio Frequency cavities*. URL: <https://home.cern/about/engineering/radiofrequency-cavities> (visited on 08/21/2010).
- [15] *ATLAS*. URL: <https://atlas.cern/> (visited on 08/21/2010).
- [16] *ALICE*. URL: <http://aliceinfo.cern.ch/Public/Welcome.html> (visited on 08/21/2010).
- [17] *CMS*. URL: <https://home.cern/about/experiments/cms> (visited on 08/21/2010).
- [18] *LHCb Experiment*. URL: <https://home.cern/about/experiments/lhcb> (visited on 08/21/2010).
- [19] P. Schmüser J. Rossbach. *Basic course on accelerator optics*. URL: <http://inspirehep.net/record/347558/files/Cern.pdf> (visited on 08/21/2010).
- [20] Matthew Sands. “The Physics of Electron Storage Rings: An Introduction”. In: *Conf. Proc. C6906161* (1969), pp. 257–411.
- [21] R. Jones M. Gasior. “High Sensitivity Tune Measurement By Direct Diode Detection”. In: *Conf. Proc. C6906161* (2005).
- [22] Andrzej Wolski. “Maxwell’s Equations for Magnets”. In: *Proceedings, 2009 CAS-CERN Accelerator School: Specialised course on Magnets: Bruges, Belgium, June 16 - 25, 2009*. 2011, pp. 1–38. arXiv: 1103.0713 [physics.acc-ph]. URL: <http://inspirehep.net/record/891330/files/arXiv:1103.0713.pdf>.

- [23] M. Betz et al. “Bunched-beam Schottky monitoring in the LHC”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 874 (2017), pp. 113–126. ISSN: 0168-9002. doi: <https://doi.org/10.1016/j.nima.2017.08.045>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900217309452>.
- [24] Jacques Gareyte, Jean-Pierre Koutchouk, and F Ruggiero. *Landau damping dynamic aperture and octupole in LHC*. Tech. rep. LHC-Project-Report-91. CERN-LHC-Project-Report-91. revised version number 1 submitted on 2003-08-21 14:12:02. Geneva: CERN, Feb. 1997. URL: <https://cds.cern.ch/record/321824>.
- [25] J. Scott Berg and Francesco Ruggiero. “Landau damping with two-dimensional betatron tune spread”. In: (1996).
- [26] W Herr. “Introduction to Landau Damping”. In: arXiv:1601.05227 (2014). Comments: 27 pages, contribution to the CAS - CERN Accelerator School: Advanced Accelerator Physics Course, Trondheim, Norway, 18-29 Aug 2013, 27 p. URL: <https://cds.cern.ch/record/1982428>.
- [27] D Möhl. “On Landau damping of dipole modes by non-linear space charge and octupoles”. In: *Part. Accel.* 50.CERN-PS-95-08-DI (1995), pp. 177–187. URL: <https://cds.cern.ch/record/280706>.
- [28] *Beam Position Monitor (BPM)*. URL: <http://www.lhc-facts.ch/index.php?page=bpm> (visited on 08/21/2018).
- [29] *BLM photo*. URL: https://ab-div-bdi-bl-blm.web.cern.ch/ab-div-bdi-bl-blm/Beam_loss_detectors/blm_photo/P1000814.JPG (visited on 08/21/2010).
- [30] A Goldblatt et al. “Design and Performance of the Upgraded LHC Synchrotron Light Monitor”. In: CERN-ACC-2013-0300 (Sept. 2013), 4 p. URL: <https://cds.cern.ch/record/1637774>.
- [31] Svante Wold, Kim Esbensen, and Paul Geladi. “Principal component analysis”. In: *Chemometrics and Intelligent Laboratory Systems* 2.1 (1987). Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists, pp. 37–52. ISSN: 0169-7439. doi: [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9). URL: <http://www.sciencedirect.com/science/article/pii/0169743987800849>.
- [32] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [33] George Seber and Alan J Lee. *Linear Regression Analysis*. Vol. 1. Jan. 2012.
- [34] Donald W. Marquardt and Ronald D. Snee. “Ridge Regression in Practice”. In: *The American Statistician* 29.1 (1975), pp. 3–20. ISSN: 00031305. URL: <http://www.jstor.org/stable/2683673>.
- [35] Tibshirani Robert. “Regression shrinkage and selection via the lasso: a retrospective”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.3 (), pp. 273–282. DOI: 10.1111/j.1467-9868.2011.00771.x. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2011.00771.x>. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00771.x>.

- [36] Zou Hui and Hastie Trevor. “Regularization and variable selection via the elastic net”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (), pp. 301–320. DOI: 10 . 1111 / j . 1467 - 9868 . 2005 . 00503 . x. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2005.00503.x>. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x>.
- [37] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan, New York, 1962.
- [38] Leo Breiman. “Random Forests”. In: *Machine Learning* 45.1 (Sept. 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10 . 1023 / A : 1010933404324. URL: <https://doi.org/10.1023/A:1010933404324>.
- [39] Leo Breiman. “Arcing The Edge”. In: (Aug. 1997).
- [40] Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 3146–3154. URL: <http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf>.
- [41] J. R. Quinlan. “Induction of Decision Trees”. In: *Mach. Learn.* 1.1 (Mar. 1986), pp. 81–106. ISSN: 0885-6125. DOI: 10 . 1023 / A : 1022643204877. URL: <http://dx.doi.org/10.1023/A:1022643204877>.
- [42] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. In: *Ann. Statist.* 7.1 (Jan. 1979), pp. 1–26. DOI: 10 . 1214 / aos / 1176344552. URL: <https://doi.org/10.1214/aos/1176344552>.
- [43] J. A. Nelder and R. Mead. “A Simplex Method for Function Minimization”. In: *The Computer Journal* 7.4 (1965), pp. 308–313. DOI: 10 . 1093 / comjnl / 7 . 4 . 308. eprint: /oup/backfile/content_public/journal/comjnl/7/4/10.1093/comjnl/7.4.308/2/7-4-308.pdf. URL: <http://dx.doi.org/10.1093/comjnl/7.4.308>.
- [44] *madx*. URL: <http://madx.web.cern.ch/madx/> (visited on 08/21/2010).
- [45] *sixtrack*. URL: <http://sixtrack.web.cern.ch/SixTrack/> (visited on 08/21/2010).
- [46] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-term Memory”. In: 9 (Dec. 1997), pp. 1735–80.

Appendix A Visualisation 2017

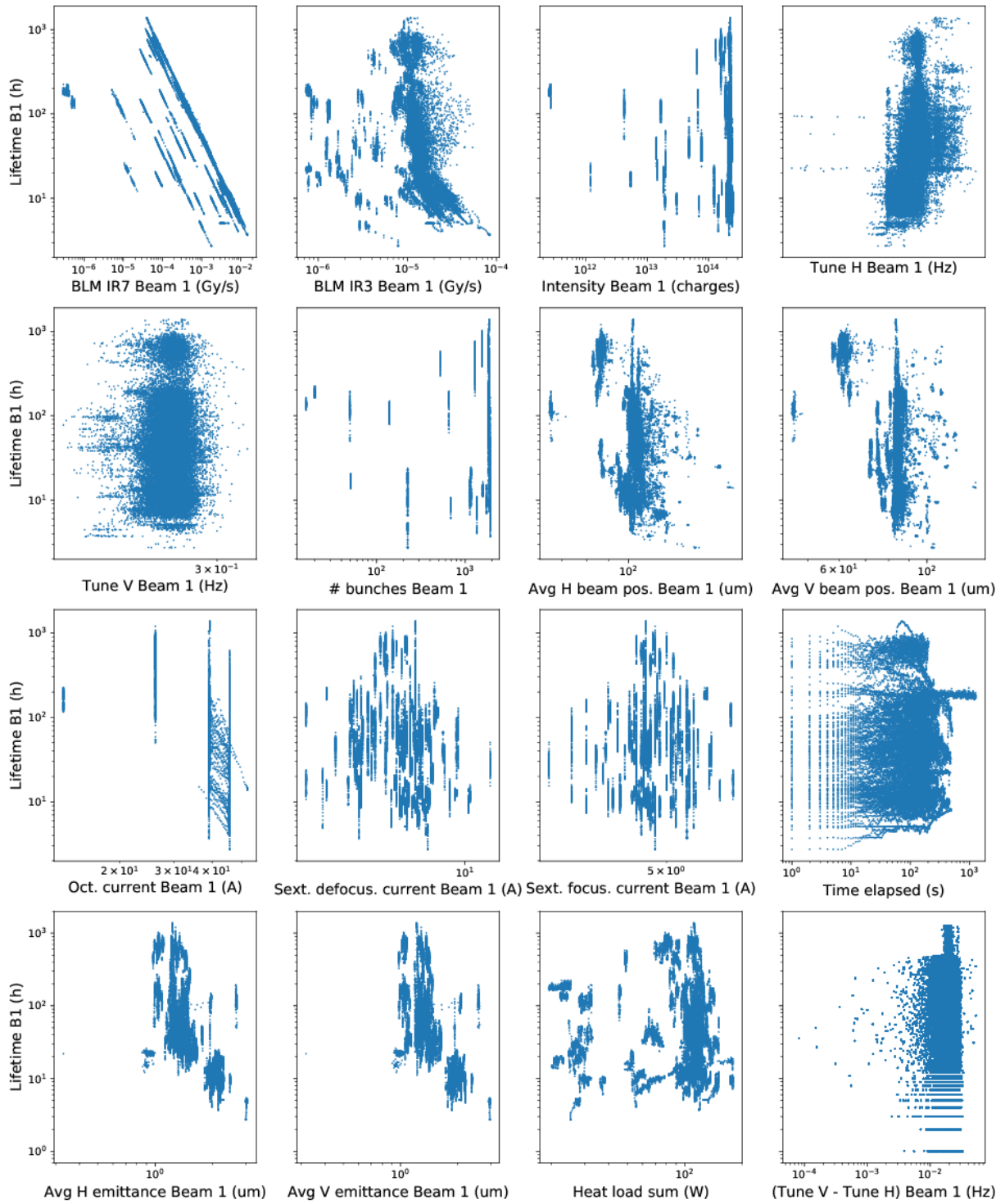


Figure 41: Visualisation of dataset for Beam 1, both axis are log scaled

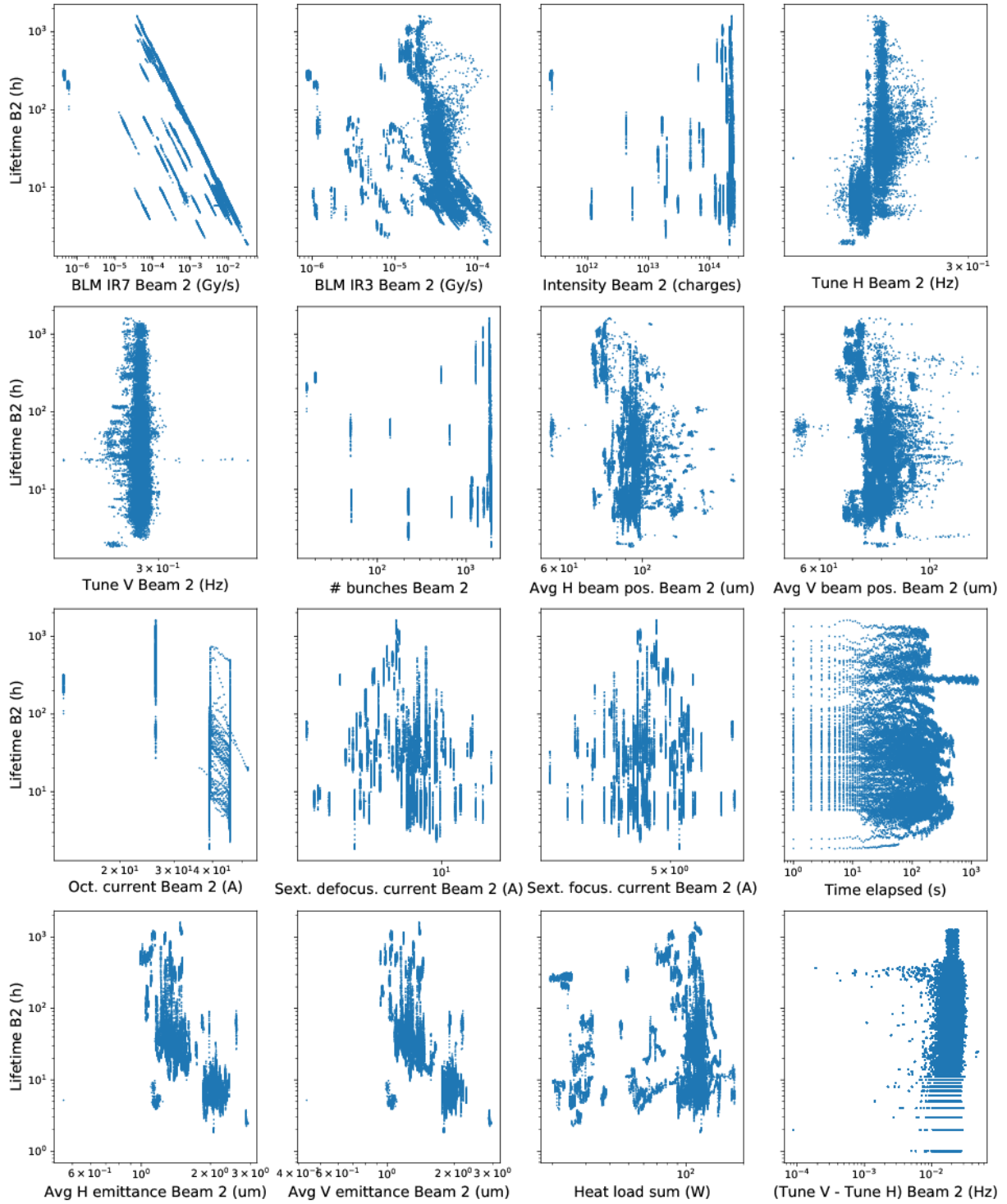


Figure 42: Visualisation of dataset for Beam 2, both axis are log scaled

Appendix B Visualisation 2018

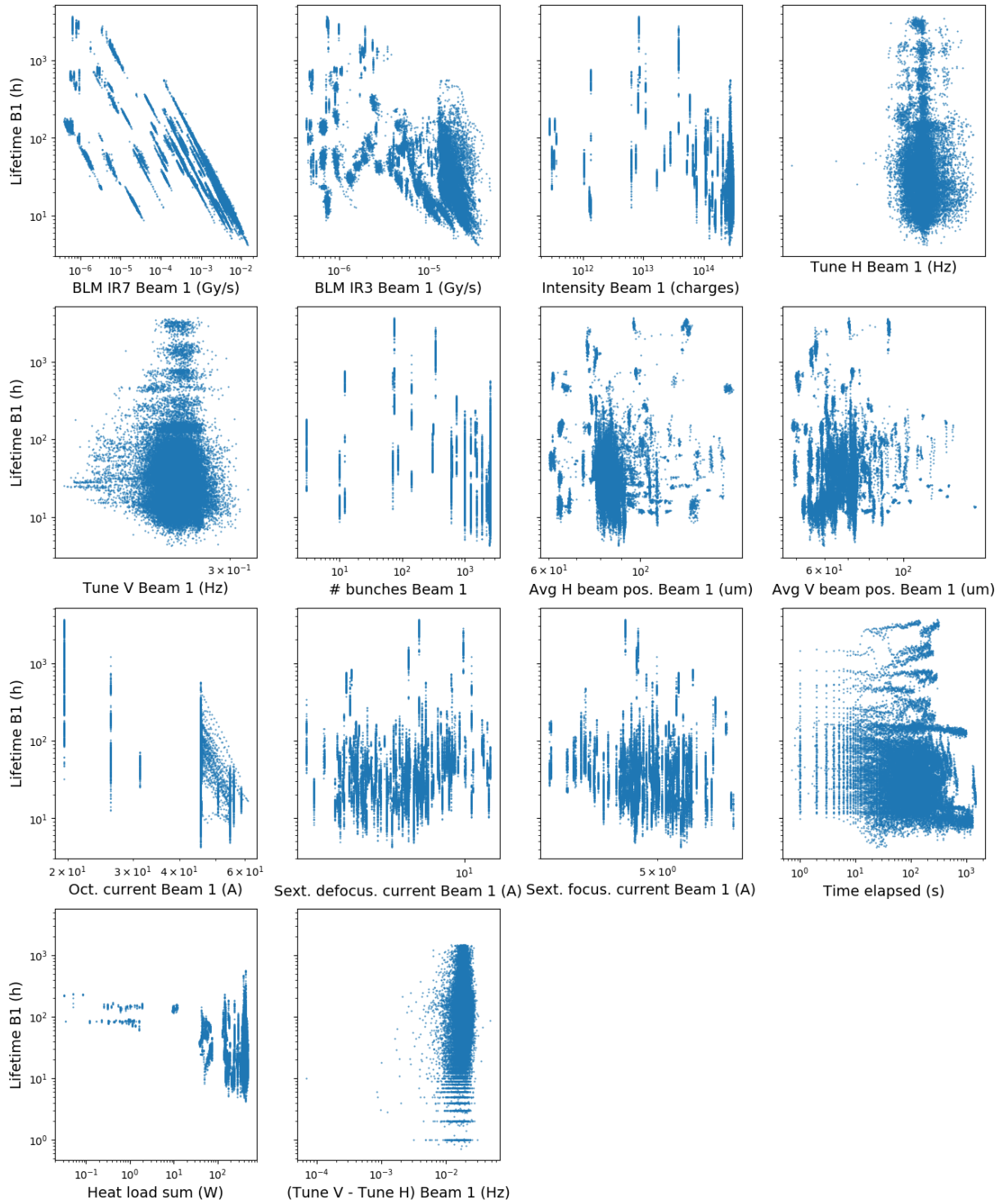


Figure 43: Visualisation of dataset for Beam 1, both axis are log scaled

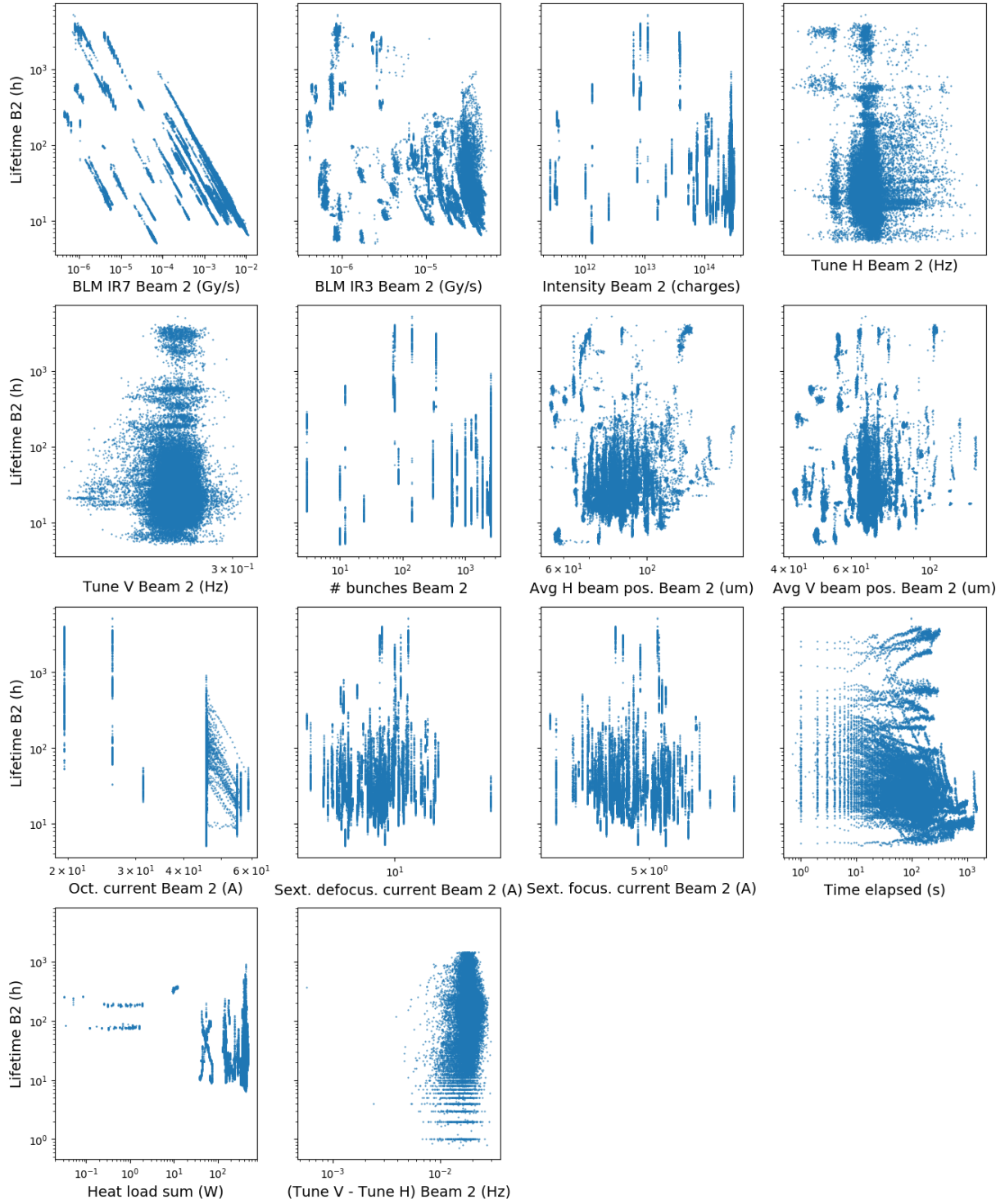


Figure 44: Visualisation of dataset for Beam 2, both axis are log scaled

Appendix C Training model on 2018 data

The training process for the 2018 model is identical to the aforementioned models. The data in this dataset covers all the fills from the start of 2018 onward. The dataset is visualised in the same way as previously, Figure 43,44 appendix B, all the input features are plotted against their respective beam's lifetime. There are fewer obvious links between the features and the lifetimes compared to the 2017 dataset. Once again, as expected, the losses at both IPs are very strongly linked to the lifetimes. The slight visual trend for higher lifetime at higher horizontal tunes previously observed do not seem to be present here. Moreover, we can identify that lower heat loads do seem to correlate to higher lifetimes which is the expected behaviour, as the heat load indicates the presence of electron clouds which cause particle losses. Unfortunately the BSRT system used previously for the emittance values was not adequately calibrated and the values could not be used, thus the emittances are absent.

The Spearman correlation coefficient matrix, Figure 45, shows the expected link between the BLM signals and the lifetimes. However, in this dataset, the correlations with the other features have drastically changed compared to the 2017 dataset. Overall the correlations are much less pronounced, and indeed, the correlations with the horizontal tunes is no longer present. The octupole current also seem to impact the lifetimes similarly to the 2017 dataset, this negative correlation is expected, as discussed in section 3.4.

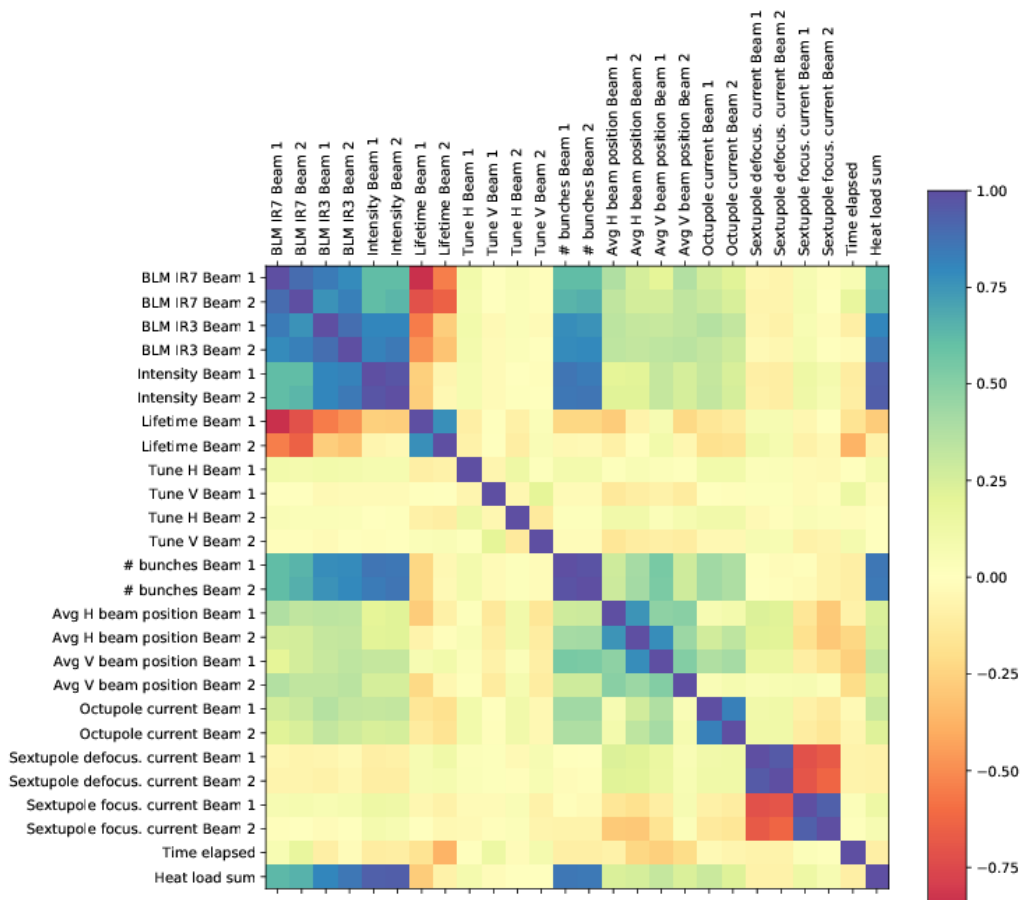


Figure 45: Spearman Correlation matrix for 2018 data

The tuning of the model gives us the following parameter set:

- max_depth : -1 (no limit)
- num_leaves : 58

- min_data_in_leaf : 3
- learning_rate : 0.01
- bagging_freq : 7
- bagging_fraction : 0.98
- feature_fraction : 0.98

The training of the model is shown in Figure 46. There is a significant offset between the loss on the training set and the loss on the test set. However, it does not seem to significantly impact the performance of the model, the R^2 coefficient is still reasonable and the actual predictions match well with the true lifetime values. The performance metrics of the model are shown in Table 6. The models have successfully converged and their predictions of the test set is shown in Figure 47, the predictions of the models are on par with the previous models.

Model	Beam 1		Beam 2	
	MSE	R^2	MSE	R^2
Bare dataset 2018	829.2	0.992	3129.2	0.985

Table 6: Performance of model trained on bare dataset 2018

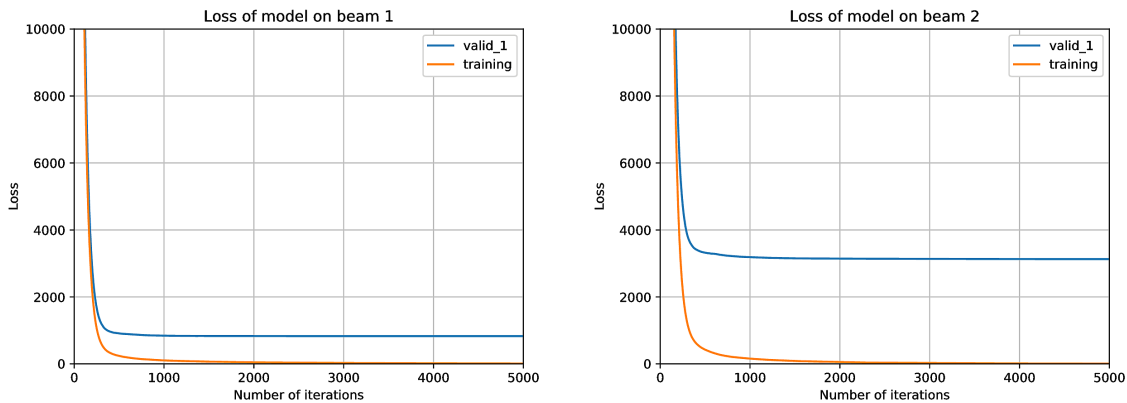


Figure 46: Loss function evolution beam 1 & 2, bare model on 2018 data

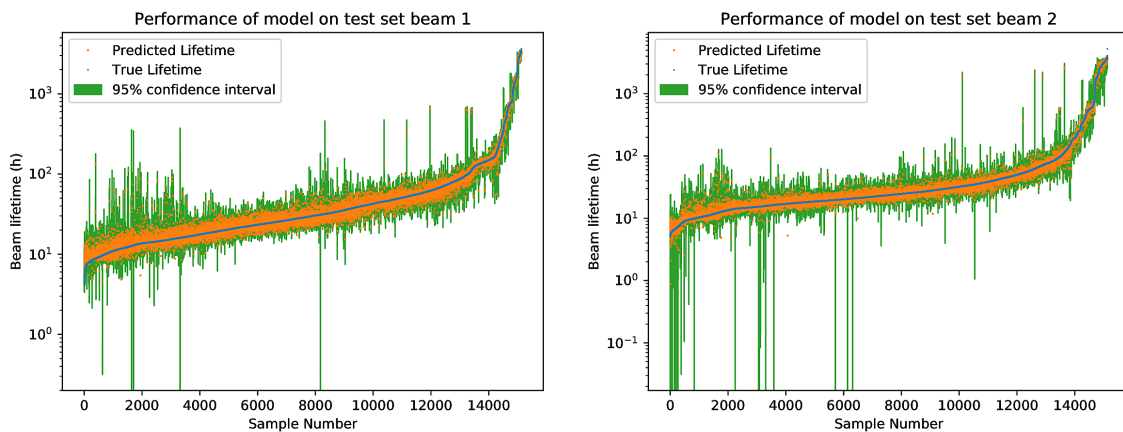


Figure 47: Model predictions beam 1 & 2, bare model on 2018 data

For the feature importances, we observe similar trends as in Figure 32, the main contributor is the time elapsed since the start of the fill followed by the tunes for both beams, Figure 48.

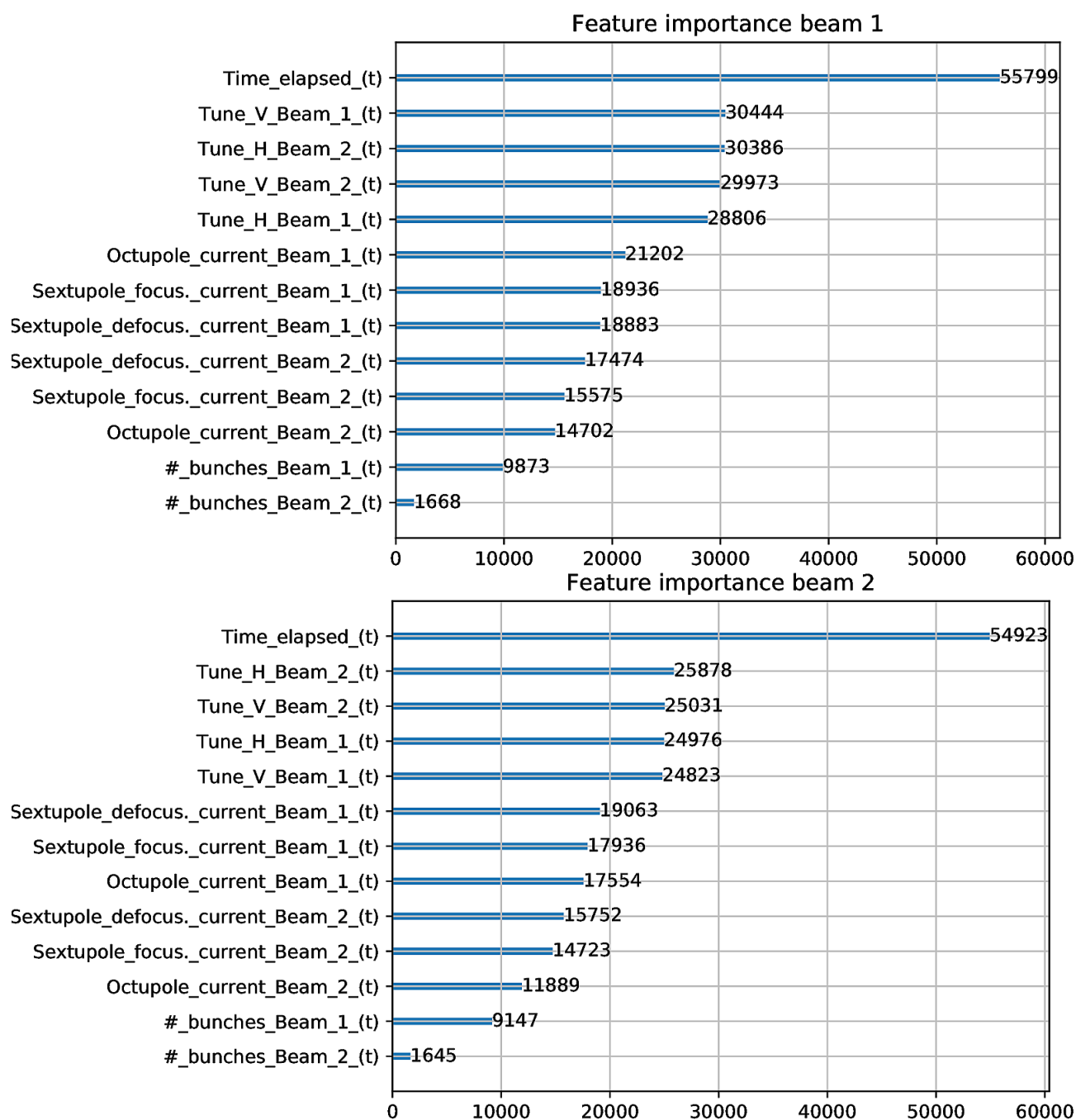


Figure 48: Feature importance for beam 1 & 2, bare model on 2018 data

Appendix D Tune Diagrams

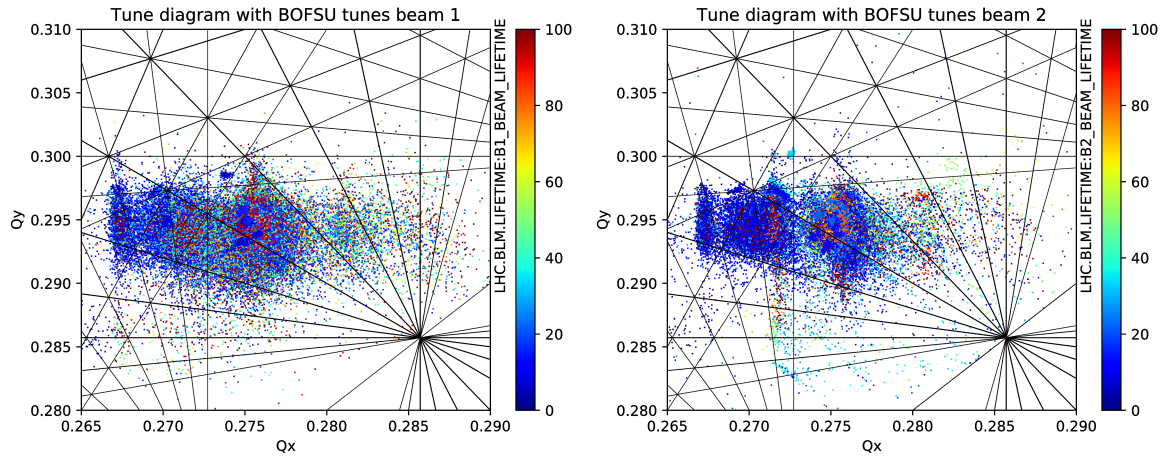


Figure 49: BOFSU tune diagram for 2017 data, colour indicates the beam lifetime

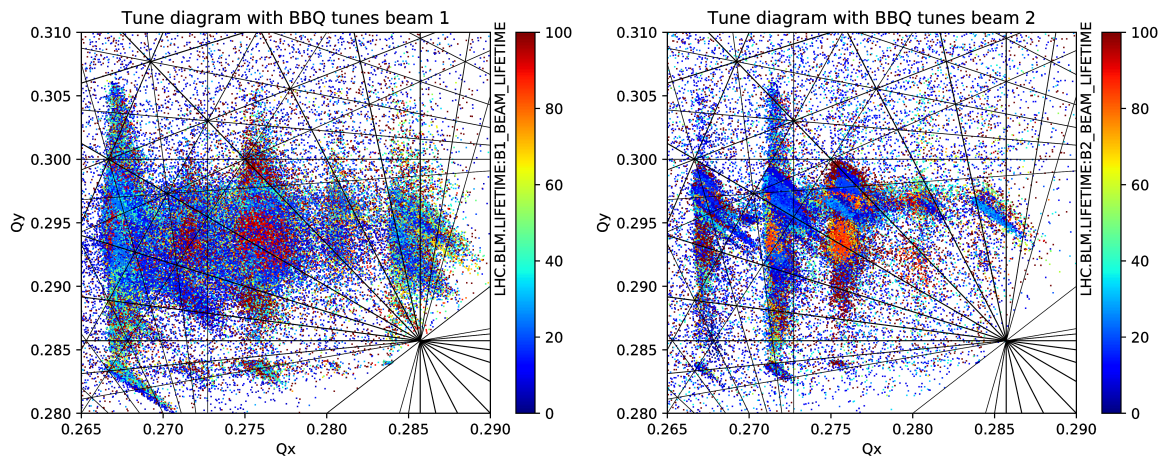


Figure 50: BBQ tune diagram for 2017 data, colour indicates the beam lifetime

Appendix E Gantt Chart

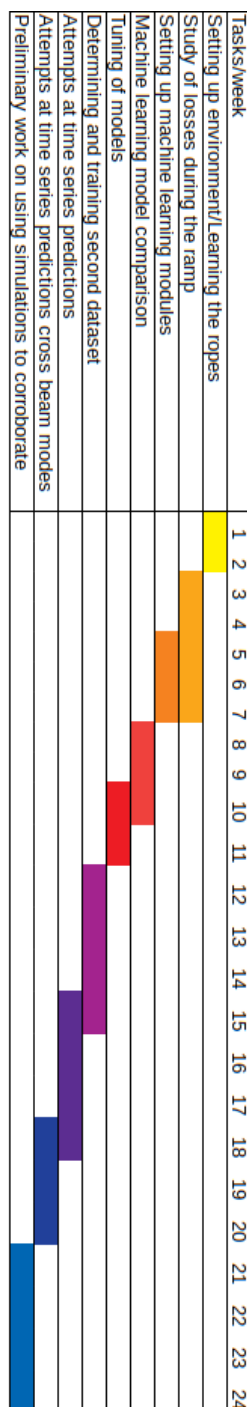


Figure 51: Gantt chart

Appendix F Summaries

F.1 English

This project set out to help in the optimization of the LHC by studying the losses occurring during the ramp and applying machine learning techniques to optimizing the intensity lifetimes of the machine.

The losses occurring in the ramp, over many machine fills, have been characterized and their timings indicate a link with the optical matching points of the machine. Discrepancies between the two beams were also demonstrated. By looking at the interpolated position of the beam at the primary collimators, there seems to be a correlation with the horizontal drift of the beams, however this drift does not seem to be the cause of the losses. More work must be done to fully elucidate their origin.

A first attempt at applying machine learning to the optimization of the LHC was done. A variety of models have been successfully trained on LHC data with decent performances. These models shed some light on the influence of some parameters on the beam lifetimes. They also seem to give counter intuitive results notably the importance of the time elapsed on the models' decision making process. More work must be done to fully understand this. The models were paired with an optimization algorithm which serves as a proof of concept for online machine optimization. Nonetheless, it is just a first step towards successfully applying machine learning to the optimization of the LHC.

F.2 Français

Le but de ce projet était d'aider à optimiser le LHC en étudiant les pertes qui ont lieu pendant la phase d'accélération des protons et d'utiliser des techniques de machine learning pour optimiser la durée de vie en intensité des faisceaux.

Les pertes pendant la phase d'accélération ont été caractérisées sur plusieurs remplissages du LHC, le timing de ces pertes indique un lien avec les points de raccordement des paramètres optiques de la machine. De plus, il semble avoir une différence entre les deux faisceaux. En regardant la position interpolée des faisceaux au niveau des collimateurs primaires, il apparaît une corrélation entre les écarts des faisceaux de leurs orbites idéales. Néanmoins, ces écarts ne semblent pas être la cause des pertes et plus de études doivent être réalisées pour en extraire l'origine.

Une première tentative dans l'application du machine learning à l'optimisation du LHC a été faite. Une variété de modèles, ayant une bonne performance, ont été entraînés avec succès sur les données du LHC. Ces modèles ont permis de cerner d'autant plus l'influence de paramètres sur la durée de vie des faisceaux. Ils ont aussi donné des résultats inattendus comme l'importance du temps écoulé dans les décisions du modèle. Une étude plus poussée est nécessaire pour comprendre ces résultats. De plus, les modèles ont été combinés avec un algorithme d'optimisation pour illustrer une première approche à l'optimisation en temps réel du LHC. Néanmoins, ceci reste un premier pas dans l'application du machine learning à l'optimisation du LHC.