CERN

Machine learning in accelerator physics: Applications at the CERN Large Hadron Collider

AISIS 2019, Ciudad de México

Friday October 25, 2019

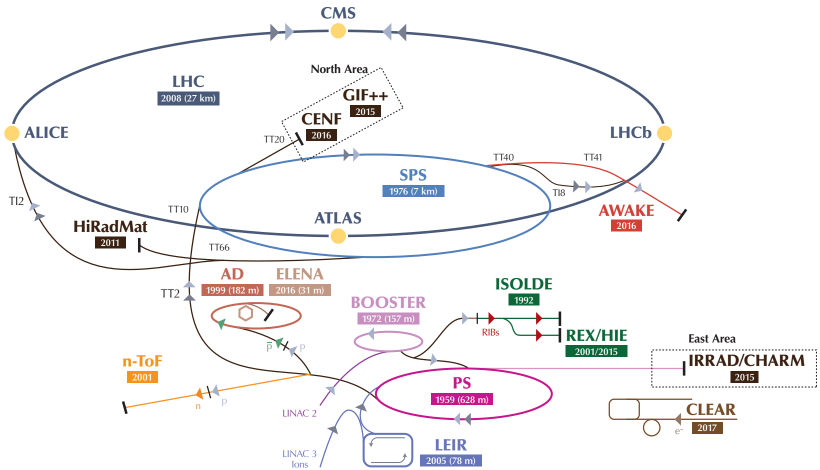
F.F. Van der Veken[1,2],

G. Azzopardi[1,2], F.H. Blanc[4], L.T.D. Coyle[1,4], E. Fol[1,3], M. Giovannozzi[1], T. Pieloni[1,4], S. Redaelli[1], B.M. Salvachua Ferrando[1], M. Schenk[1,4], R. Tomas Garcia[1], G. Valentino[1,2],

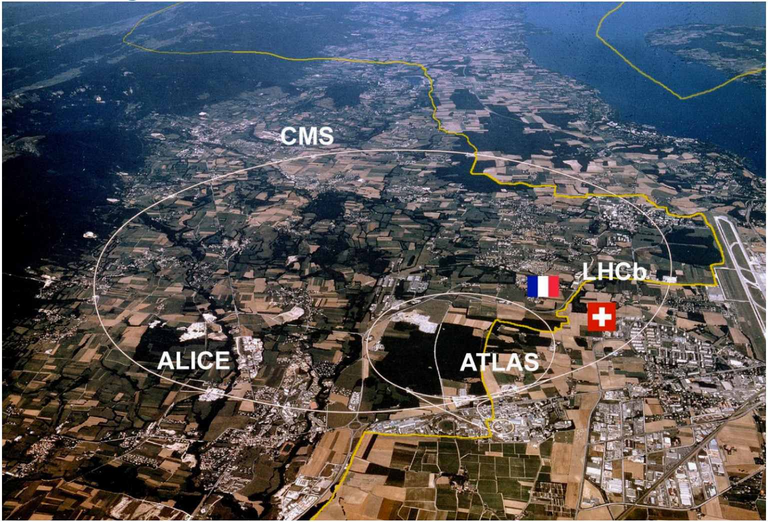
[1] *CERN,* [2] *University of Malta,* [3] *Johann-Wolfgang-Goethe University,* [4] *Ecole Polytechnique Federale Lausanne*

Outline

CERN Accelerator Complex



The Large Hadron Collider



(figure shamelessly stolen from Maciej's talk on Monday)

The Large Hadron Collider

The LHC is a huge and extremely complex machine:

- 26.659km tunnel, 50m-150m underground
- Accelerates particles up to 7TeV
- 9593 magnets, most cooled to 1.9K
- $\sim$ 1.5 billion collisions per second
- Total cost about 6.5 billion CHF
- Total energy usage about 230MW
- Extremely sensitive: moon tides, day/night tariff, tgv, ...

$\Rightarrow$ ¡ Operating it is far from trivial !

Operating the LHC

Operating the LHC

Machine Learning at the LHC

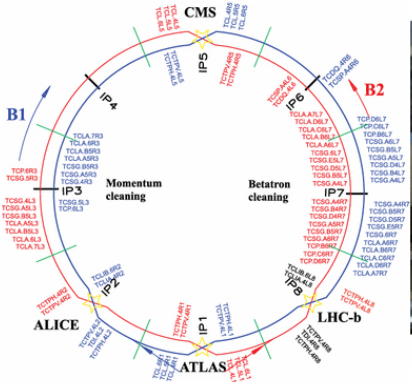
Different Collaborations

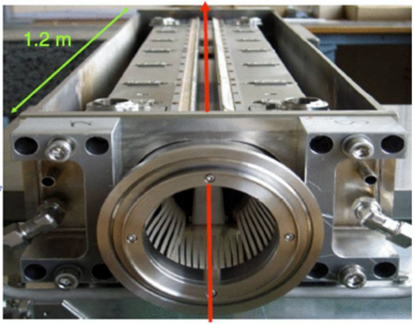
- Concerning **machine operation**:
  - Collimator alignment
  - Recognition of faulty monitors
  - Correction of beam optics variables
- Concerning **analysis** of measurements and simulations:
  - Anomaly detection in tracking simulations
  - Extrapolation of tracking simulations
  - Modelling beam lifetime by operational settings

Outline

L-Università ta' Malta

Context





362 MJ proton beam

Context

LHC Protection System

- The LHC uses a system of **100 collimators** for protection
- These must be **aligned** around the two beams with a precision better than 50μm
- Alignments are performed yearly before start of operation

Applications with Machine Learning

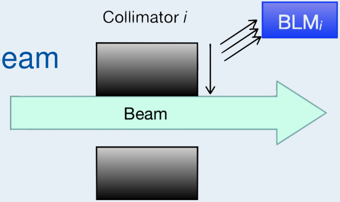
- Alignment is tedious, time-consuming, and repetitive
- Ideal situation for machine learning

$\Rightarrow$ *supervised learning*

Setup

Beam Loss Monitors (BLMs)

- record **losses** as they touch the beam
- experts monitor these losses to deduce collimator **alignment**

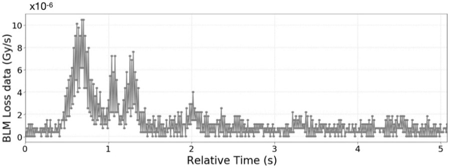


Setup

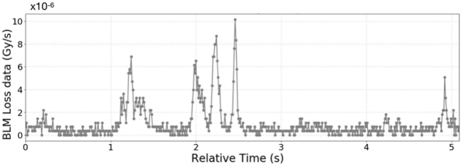
Spikes

- data sample taken when collimator stops moving
- spike when threshold in BLM is passed
- goal is to distinguish **real** spikes (beam is hit) from noise

Alignment Spike



Non-Alignment Spike

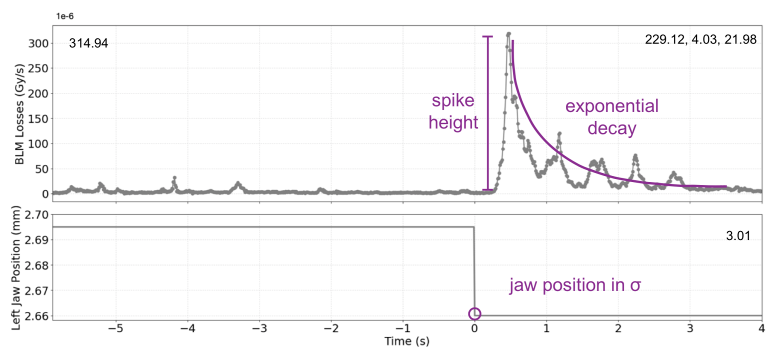


Spike Classification

Spike Parameterisation

5 parameters:

jaw position (1), spike height (1), and decay fit (3)



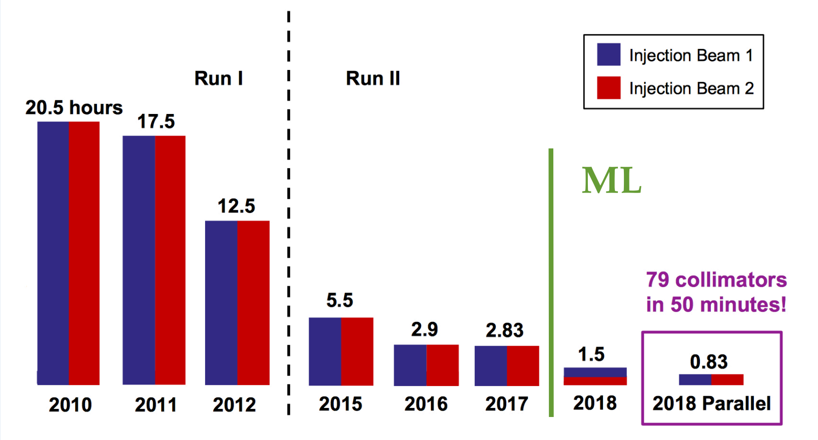
Spike Classification

Approach

- Six ML models for spike classification were compared

  Logistic Regression, Neural Network, SVM, Decision Tree, Random Forest, Gradient Boost
- data split into: 6446 samples for training, 1778 for testing
- enforce: **no false positives**
- false negatives are OK (because alignment will continue)
- **no retraining needed** unless hardware changes
- Analysis of **beam crosstalk** allows parallel alignments

Results



Run I
Run II
Injection Beam 1
Injection Beam 2
20.5 hours
17.5
12.5
5.5
2.9
2.83
ML
1.5
79 collimators
in 50 minutes!
0.83
2010
2011
2012
2015
2016
2017
2018
2018 Parallel

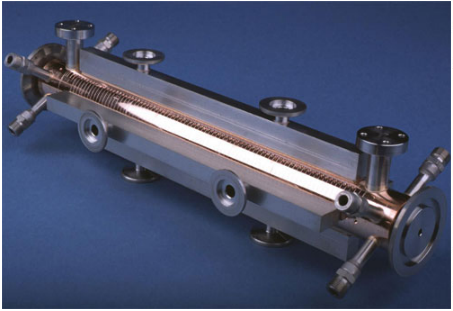
Results

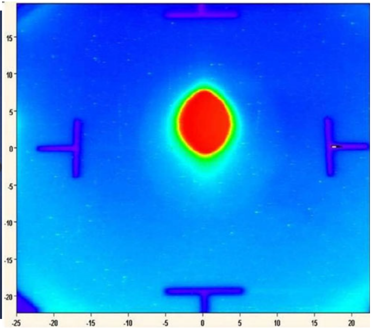
Results

- ML can replace human operators for alignments
- More than three times faster!
- ML-based alignment will be **default** from now on

Outline

Context





Context

Analysis of Beam Optics

- Beam Position Monitors (**BPM**s) measure excited beam
  - Harmonic analysis of BPM signal gives **optics functions**
  - These typically differ from design optics
- Unphysical values in optics stem from **faulty BPMs**

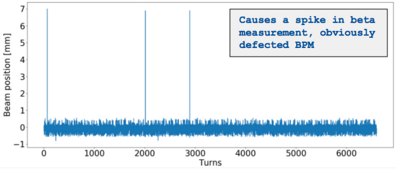
Applications with Machine Learning

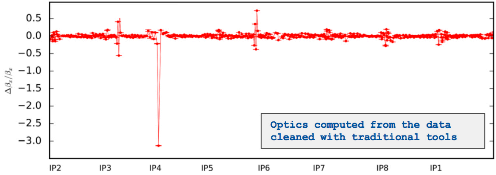
- **Identify** and remove faulty BPMs from data

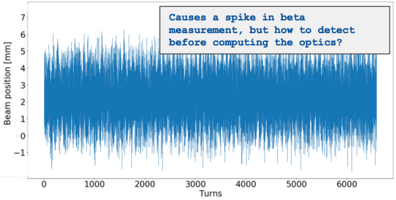
  $\Rightarrow$ anomaly detection by *unsupervised learning*
- Calculate **optimal machine settings** that minimise difference between measured and design optics

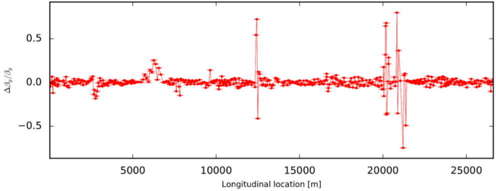
  $\Rightarrow$ *supervised learning*

Anomaly Detection









Anomaly Detection

Approach

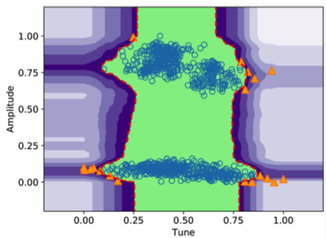
- Past measurements show that $\sim 10\%$ of BPMs are faulty
- Non-physical spikes in optics are artefact of bad BPMs
  - $\Rightarrow$ Use ML to identify faulty BPMs from harmonic analysis
  - $\Rightarrow$ To avoid spikes in optics functions
- enforce: **no false negatives** (don't keep a bad BPM)
- false positives are OK (we have $>$1000 BPMs...)

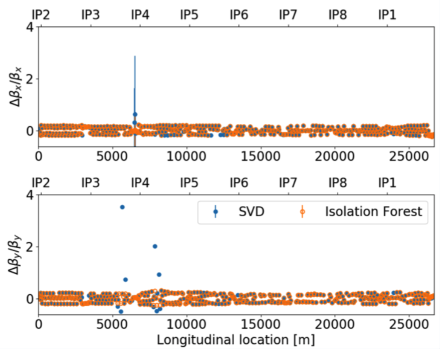
- Four ML algorithms are compared:

  K-means, DBSCAN, Local Outlier Factor, ***Isolation Forest***

Anomaly Detection

Results





Anomaly Detection

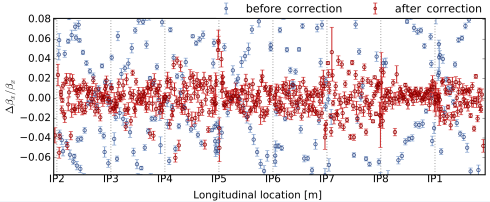
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Results

- Very good recognition of faulty BPMs
- Now integrated by **default** into optics measurements at LHC
- Successfully used during commissioning and machine developments

Correction of Beta-Beating

- β-function calculated from harmonic analysis of BPMs
- β-beating is ratio of measured over designed β-function
- Corrections in the LHC are based on **response matrix**



Machine Learning *(work in progress)*

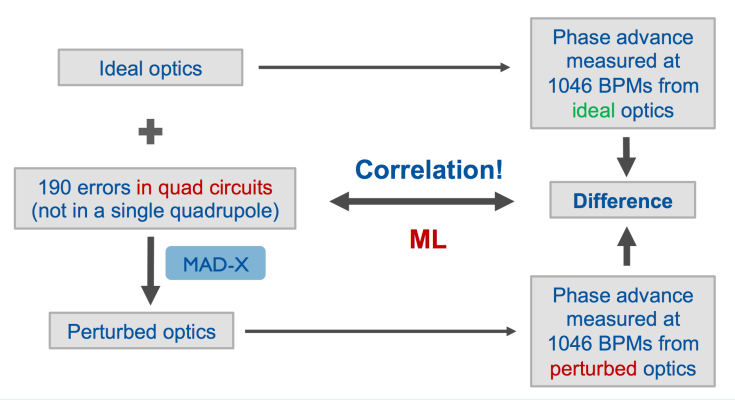
ML to reconstruct magnet errors everywhere at once

$\Rightarrow$ *supervised learning*

Correction of Beta-Beating

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Approach



Correction of Beta-Beating

Approach

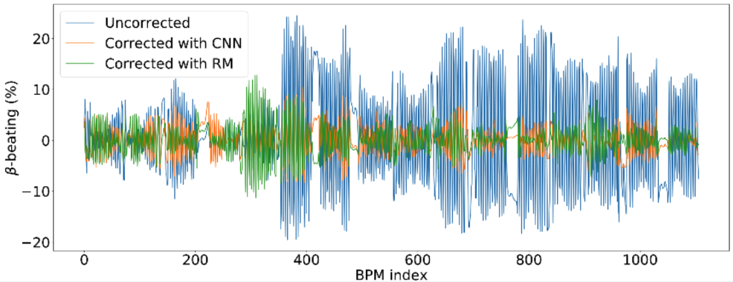
- Three ML algorithms are compared:

  Convolutional Neural Network, Linear Regression, Ridge
- CNN (Keras with TensorFlow backend):
  - Used for image processing
  - Spatially dependent features: phase advance between neighbouring BPMs
  - Different deep layers look for different features
- Very simple model is applied: no parameter tuning, no optimisation

  $\Rightarrow$ Lots of improvements are possible

Correction of Beta-Beating

Results



Correction of Beta-Beating

Results

β-beating %	peak	rms
Uncorrected	32±10	11±3
Response Matrix	11±5	3±2
CNN	11±2	3.2±0.5
Ridge regression	10±2	2.9±0.8
Linear regression	9±2	2.6±1.7

- All methods demonstrate **similar performance**
- Linear Regression ML achieves best correction

Outline

Context

Dynamic Aperture

- A tool to estimate **beam quality**
- It is the volume of the smallest connected region in phase space that remains **stable** for a certain amount of time
- Its **evolution** over time can be estimated with scaling laws
- DA can describe **beam losses** and **luminosity evolution**

Applications with Machine Learning

- Anomaly detection $\Rightarrow$ *unsupervised learning*
- DA extrapolation $\Rightarrow$ *supervised learning*

Anomaly Detection

Setup

- 60 random realisations ('seeds') in LHC simulations
- Sometimes one seed gives **very bad DA** for one angle (because close to resonance, internal cancellations, . . .)

Machine Learning

- Use ML to flag these outliers

  $\rightarrow$ let human decide whether or not to remove
- Investigate anomaly dependence on angles or seeds

Anomaly Detection

L-Università ta' Malta

Approach

Points are sometimes clustered in several groups

$\Rightarrow$ **DBSCAN** to recognise clusters

(scaled over population, min 3 points in a cluster)

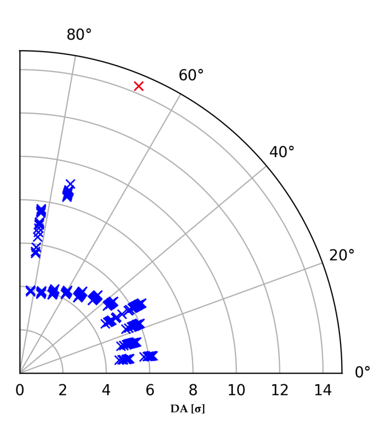
points not in cluster are possible outliers

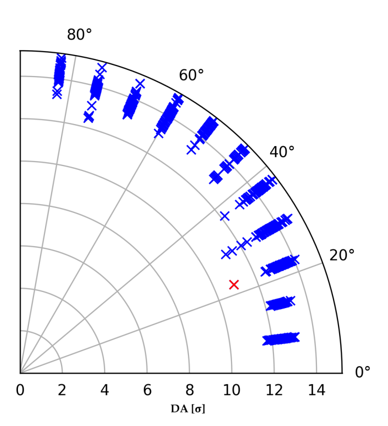
$\Rightarrow$ **LOF** to quantify outlier strength

$\Rightarrow$ Cut off at **minimum threshold**, and outliers can only exist as minima or maxima (not in between)

Anomaly Detection

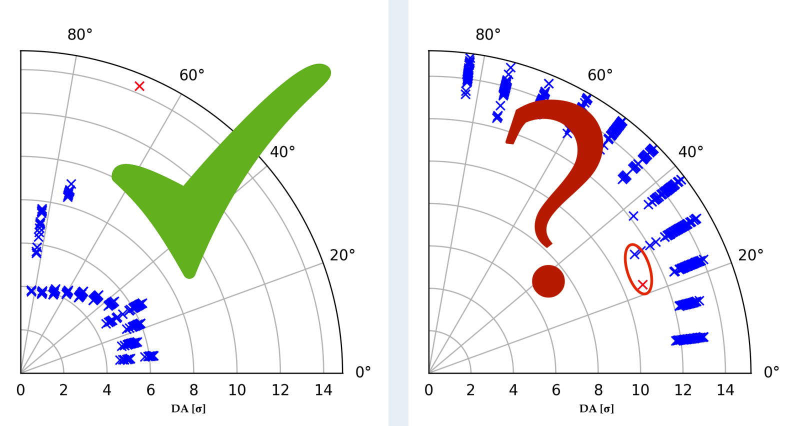
Results





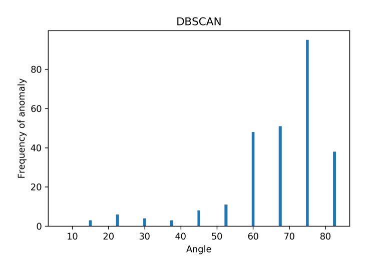
Anomaly Detection

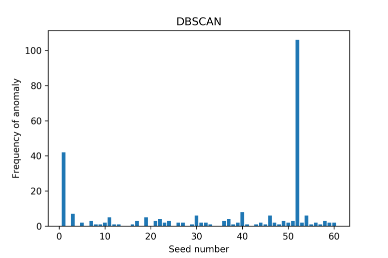
Results



Anomaly Detection

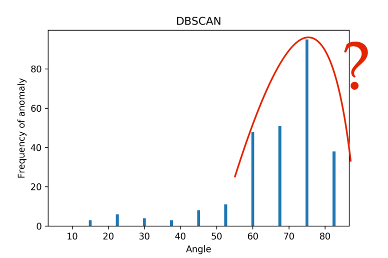
Results

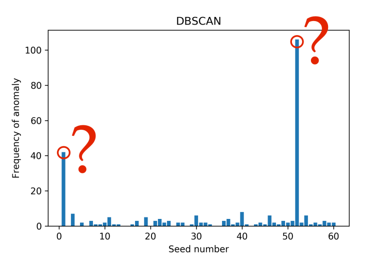




Anomaly Detection

Results





Anomaly Detection

Results

- Outlier detection per angle works as expected

  But human verification is indeed needed!

  $\rightarrow$ to decide whether or not to remove a particular seed

  (depending on behaviour of nearby angles)

- $\approx 10\times$ more outliers at large angles and seeds 1 and 52

  $\Rightarrow$ further investigation needed

Curve Fitting and Extrapolation

Setup

- DA simulations are very CPU-intensive
  - $\Rightarrow$ only $10^5 - 10^6$ turns (~ 1 minute) are achievable
- Realistic timescales are much larger (~ 10 hours)
  - $\Rightarrow$ simulations need to be **extrapolated**
- Scaling laws exist to describe **evolution** over time

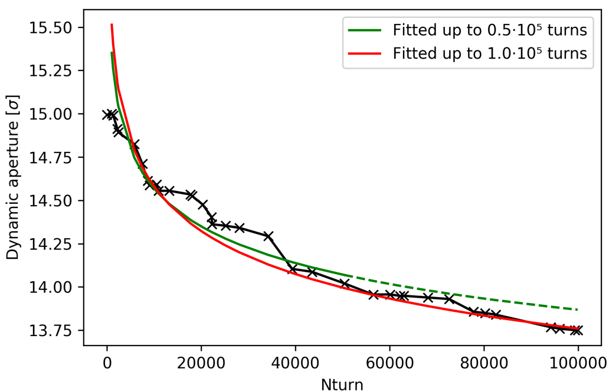
Machine Learning *(work in progress)*

- Use ML to improve fitting to scaling laws
- Recurrent Neural Network to make prediction estimates

Curve Fitting and Extrapolation

Approach

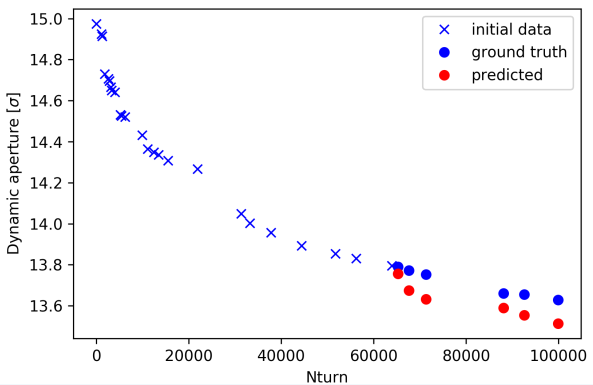
- Existing scaling laws work well to **describe** the data
- But not that much to **predict** (sensitivity of fit parameters)



Curve Fitting and Extrapolation

Trying with a Neural Network

- Brute-force approach: not including any info from scaling
- Time series analysis (LSTM with Keras)



Curve Fitting and Extrapolation

Trying with a Neural Network

- Results aren't very impressive; deeper investigation is needed
- Alternative: use a Neural Network to find optimal **weights** to fit to existing scaling laws

Outline

Context

Beam Lifetime

- Is the time τ such that intensity $I(\tau) = \frac{1}{\mathrm{e}} I_0$
- **Real-life** counterpart of DA, describing **beam quality**
- Strongly influenced by **operational settings**
- Extraction from simulation is difficult (coherent instabilities)

Applications with Machine Learning

- Avoid time- and CPU-consuming tracking simulations
- Model that directly relates **lifetime** to **machine settings**
- Ample data available, focus on **2017** and **2018**

  $\Rightarrow$ *supervised learning*

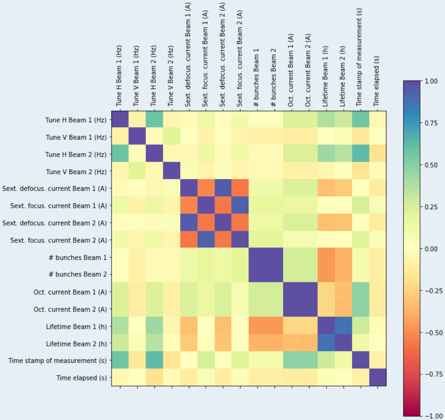
Beam Lifetime Model

Approach

- Input:
  - tunes (H/V, B1/B2)
  - sextupole strengths (B1/B2)
  - elapsed time
  - number of bunches (B1/B2)
  - emittances (H/V, B1/B2)
  - octupole strength (B1/B2)
  - timestamps
  - ...
- Output:
  - beam lifetimes (B1/B2, from slope of BCTs)
- Data from Run 2

Internal Correlations

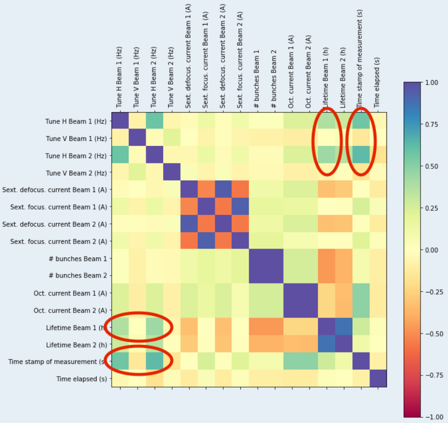
Correlations



Spearman correlation coefficient

Internal Correlations

Correlations



Lifetimes depend on tunes from **both** beams!

$\Rightarrow$ Need to de-correlate before continuing

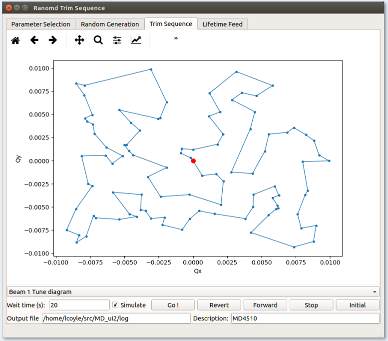
Robustness of Model

Machine Development

- Use dedicated MD run:
  - to decorrelate tunes between two beams
  - to extend tune range further than only current operational settings
- This allows us to test robustness of model:
  - does the tunes correlation matter?
  - behaviour of other beam parameters when lifetime is large?

Robustness of Model

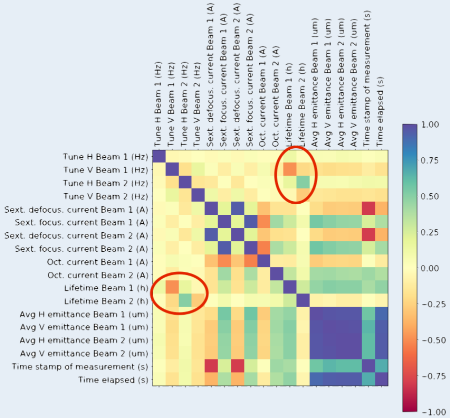
Machine Development



- random walk over tunes
- different random walk for beam 2 at the same moment
- do this for different operational settings

Robustness of Model

Machine Development

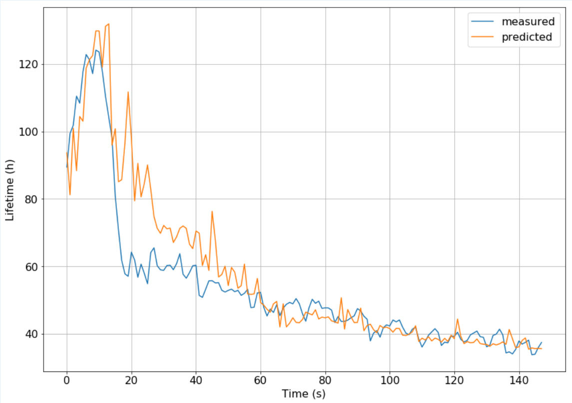


No more correlations between beams!

However, emittance becomes important

Beam Lifetime Model

Result: Prediction of Lifetime (with LightGBM algorithm)

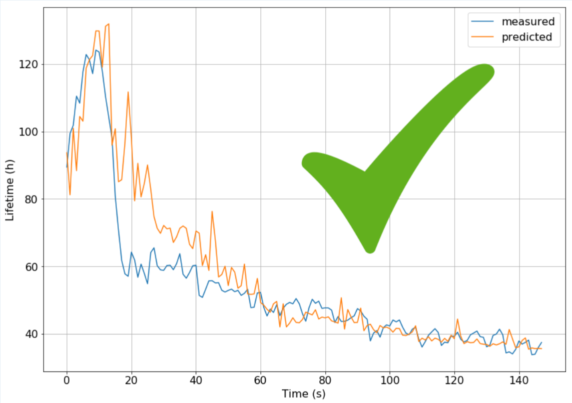


Fill 7056

Beam 2

Beam Lifetime Model

Result: Prediction of Lifetime (with LightGBM algorithm)



Fill 7056

Beam 2

Good agreement!

Multi-Parameter Optimisation

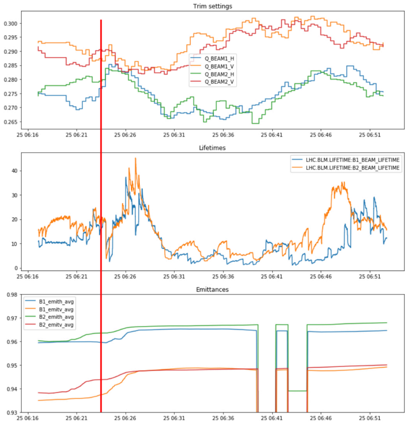
Optimal Settings

- Close to resonances: highest lifetime
- However this also gives emittance blow-up
- Latter is unwanted as it decreases luminosity

$\Rightarrow$ Multi-objective optimisation problem

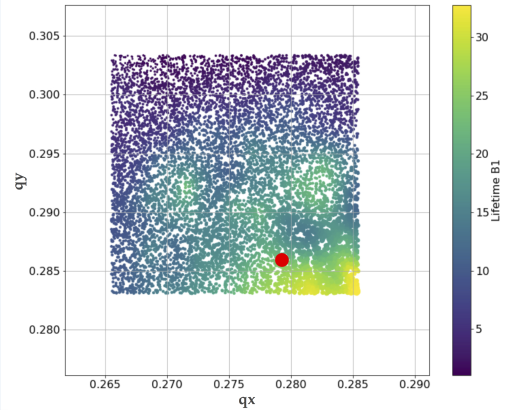
Multi-Parameter Optimisation

Optimal Settings



Multi-Parameter Optimisation

Optimal Settings



Beam 1 recommended settings:

$q_x = 0.279$

$q_y = 0.286$

Beam 2 similar

Outline

Conclusions

- **Collimator Alignment:**
  - ML is now the standard tool for collimator alignments
- **Optics Measurements and Correction:**
  - ML is now the standard tool to find faulty BPMs
  - Reconstruct true errors of single quadrupoles instead of corrector circuits: better results than using Response Matrix
  - Linear Regression is sufficient to correct linear optics errors
- **Dynamic Aperture:**
  - Anomaly detection is very efficient
- **Beam Lifetime:**
  - First steps are made towards a model that predicts lifetime in function of the operational parameters

Outlook

- **Collimator Alignment:**
  - Advanced crosstalk analysis $\rightarrow$ more alignments in parallel
- **Optics Correction:**
  - Larger dataset $\rightarrow$ more general models
  - Increase complexity of optics $\rightarrow$ more complex models
    - Add more sources of errors and non-linearities
  - Reinforcement Learning
- **Dynamic Aperture:**
  - Anomaly detection by centralised supervised learning
  - Improve prediction algorithms using high-precision data
  - Use supervised learning on fitting weights
- **Beam Lifetime:**
  - Larger dataset and more operational parameters $\rightarrow$ more general model

www.cern.ch