Machine Learning at the LHC and for Muon Collider Design Studies

Elena Fol, CERN BE-ABP

R. Tomas, G. Franchetti, D. Schulte, C. Rogers

iFAST Strategy & Brainstorming Workshop, Valencia, 29 March - 1 April 2022

Overview

1. **ML application at the LHC**
    - Instrumentation faults detection
    - Magnets sorting
    - Local optics corrections
    - Denoising and reconstruction of optics observables
    - Detection of coping sources

2. **ML applied to Muon Collider design studies**:
    - Automatic optimisation of Final Cooling system
    - Speeding up simulations using supervised learning

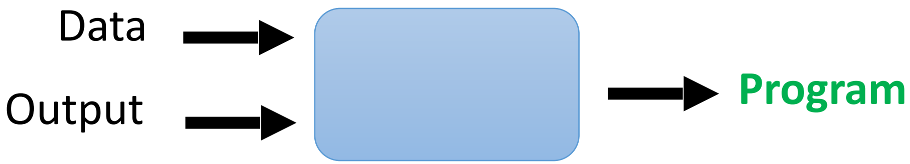
3. **General Considerations and Conclusions**

Teaching machines to learn from experience

- Traditional programming

creating **manually a set of commands** and rules

- Machine Learning approach



learn from data automatically

Why applying ML to accelerators?

Accelerators

- Operation
- Diagnostics
- Beam Dynamics Modeling

Which limitations can be solved by ML with reasonable effort?

➢ large amount of optimization targets
➢ computationally expensive simulations
➢ direct measurements are not possible
➢ previously unobserved behaviour
➢ non-linear interacting sub-systems, rapidly changing environment.

Why applying ML to accelerators?

Accelerators

- Operation
- Diagnostics
- Beam Dynamics Modeling

Which limitations can be solved by ML with reasonable effort?

Machine Learning:

✓ ***Learn arbitrary models***

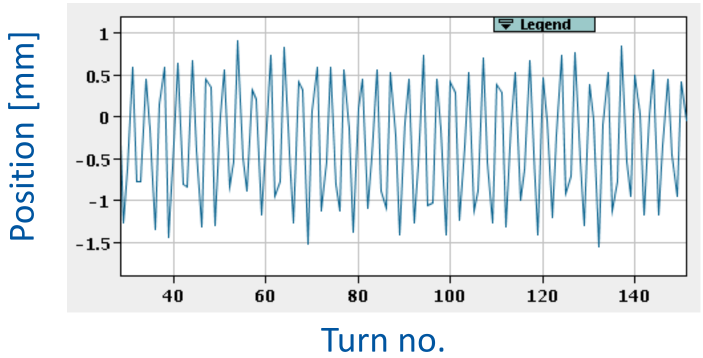
✓ ***Directly from provided data***

➢ large amount of optimization targets
➢ computationally expensive simulations
➢ direct measurements are not possible
➢ previously unobserved behaviour
➢ non-linear interacting sub-systems, rapidly changing environment.

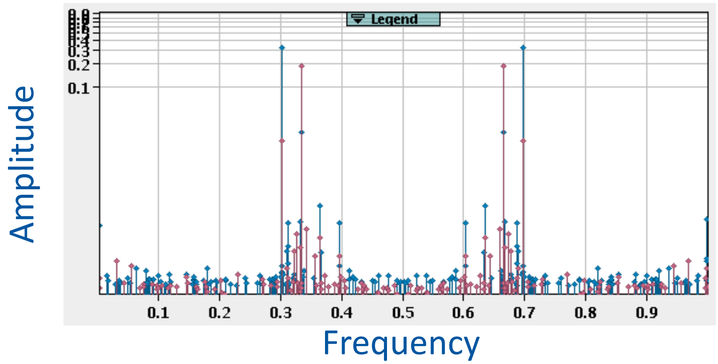
ML for LHC: Unsupervised Learning

How faulty BPMs affect the optics measurements?

Turn-by-turn beam position



Spectrum



- Excite the beam to perform transverse oscillations.

→ **Beam Position Monitors (BPMs) to measure the beam centroid turn-by-turn**

- Harmonic analysis using Fast Fourier Transform (FFT)

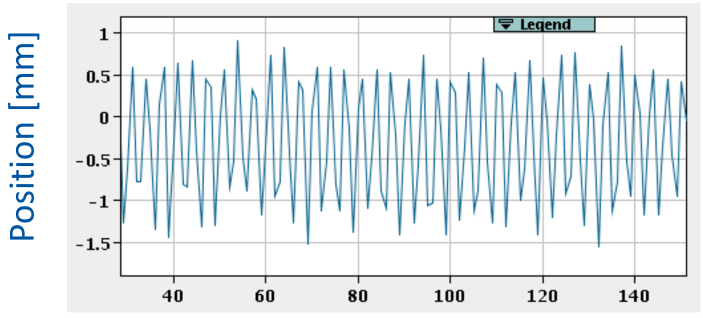
Denoising (SVD)
Signal cuts

Semi-automatic and manual cleaning of outliers

What are the limitations of traditional techniques?

How faulty BPMs affect the optics measurements?

Turn-by-turn beam position

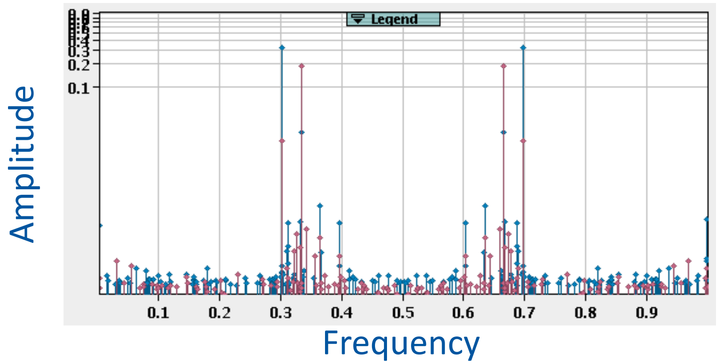


- Excite the beam to perform transverse oscillations.

→ **Beam Position Monitors (BPMs) to measure the beam centroid turn-by-turn**

Denoising (SVD)
Signal cuts

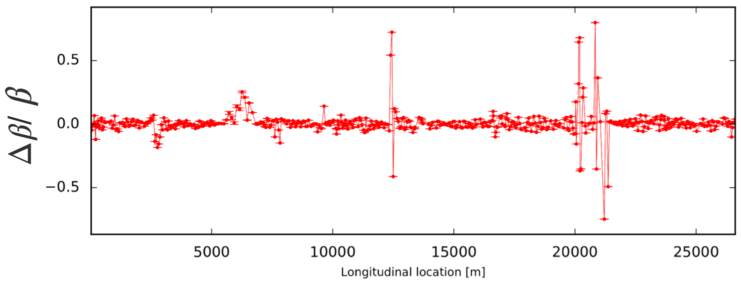
Spectrum



- Harmonic analysis using Fast Fourier Transform (FFT)

Semi-automatic and manual cleaning of outliers

Optics

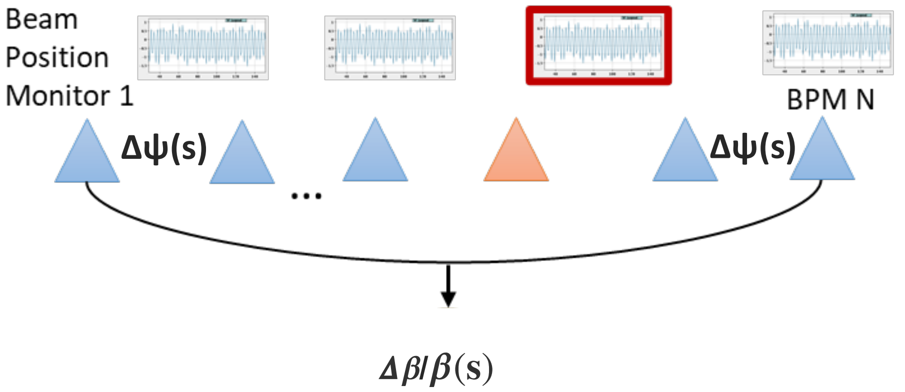


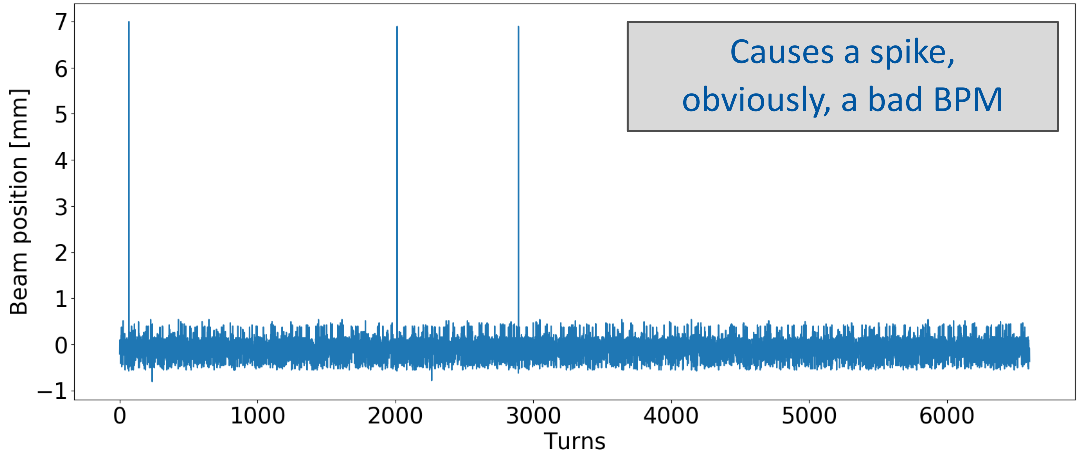
- Compute beta-beating and other optics functions

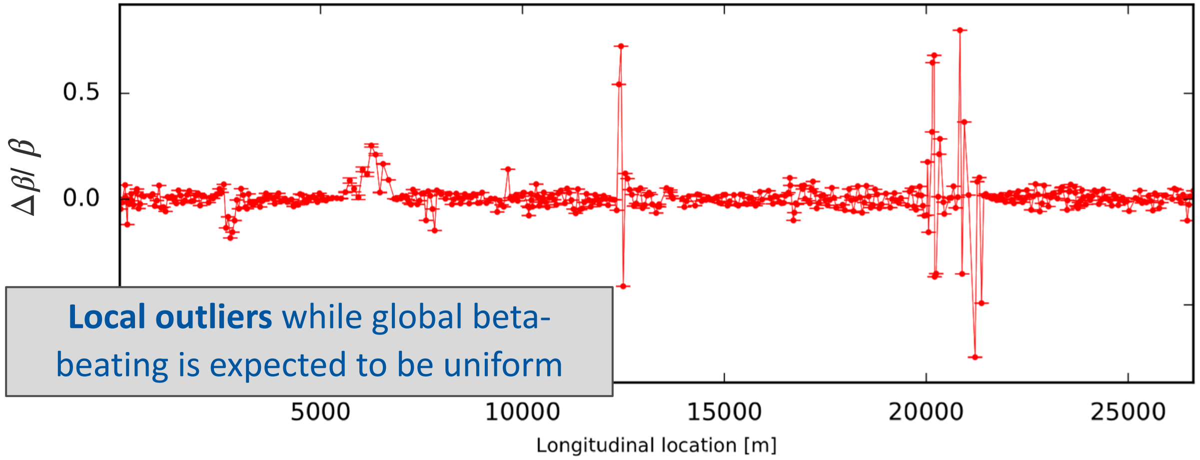
Unphysical values still can be observed

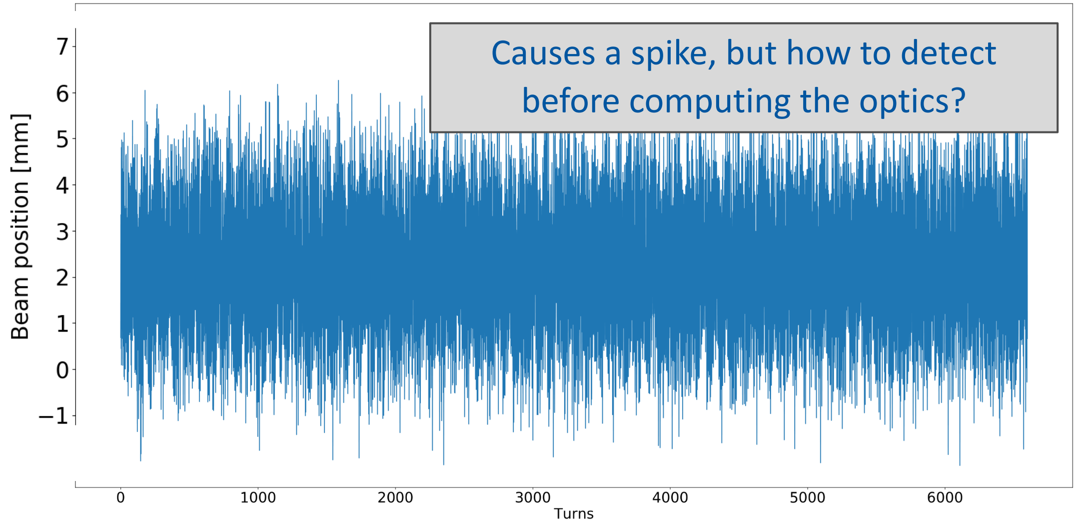
What are the limitations of traditional techniques?

How faulty BPMs affect the optics measurements?









What are the limitations of traditional techniques?

Detection of faulty Beam Position Monitors

Problem: Faulty BPMs **are a-priori unknown**:

—> cause erroneous computation of optics functions

—> manual cleaning is required

—> repeating optics analysis after manual cleaning

Detection of faulty Beam Position Monitors

Problem: Faulty BPMs **are a-priori unknown**:

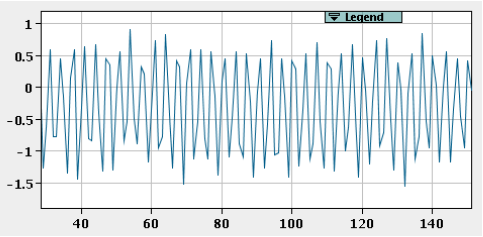
—> cause erroneous computation of optics functions

—> manual cleaning is required

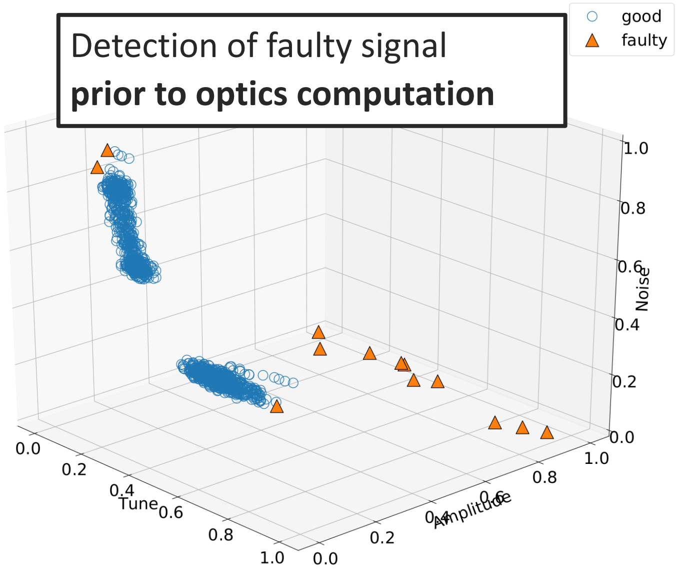
—> repeating optics analysis after manual cleaning

Anomaly detection using Unsupervised Learning

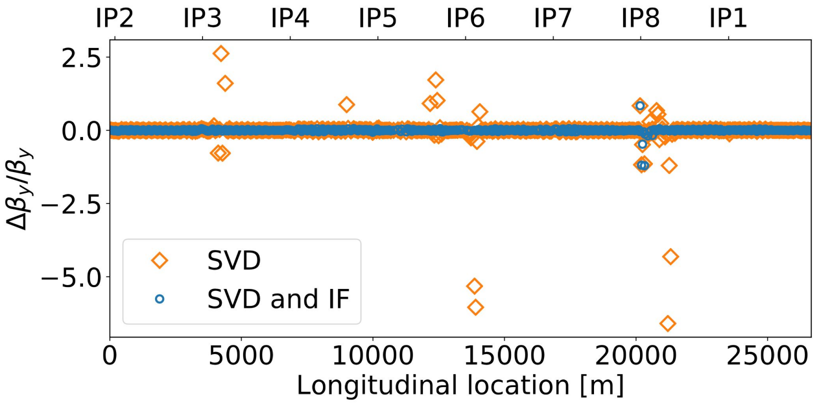
Harmonic analysis of all BPMs



Detection of faulty signal
prior to optics computation



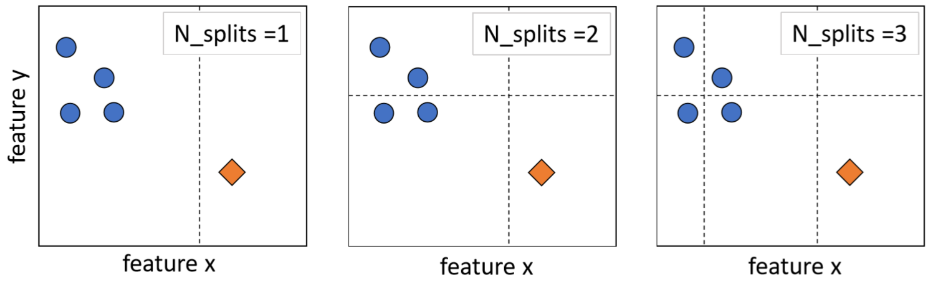
Avoid the appearance of erroneous optics computation



- Outlier detection based on combination of several signal properties
- Immediate results

Isolation Forest Algorithm

- Forest consists of several **decision trees**
- **Random splits aiming to “isolate” each point**
- The less splits are needed, the more “anomalous”
- **Contamination factor**: fraction of anomalies to be expected in the given data

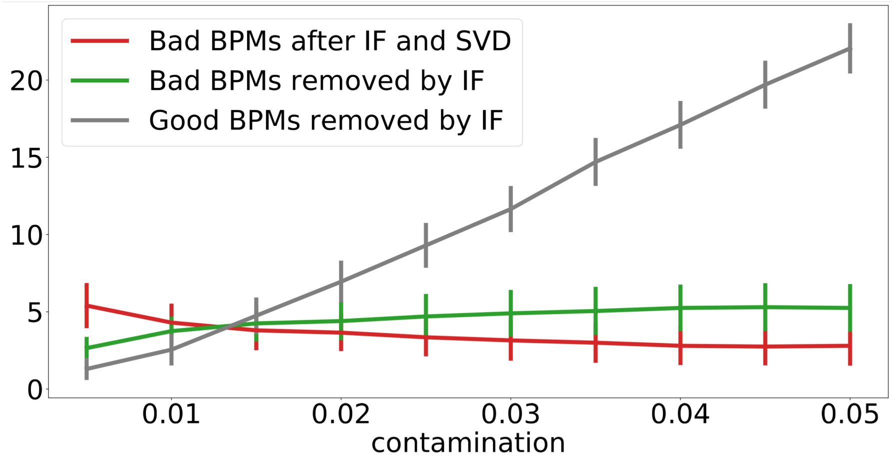


Conceptual illustration of Isolation Forest algorithm

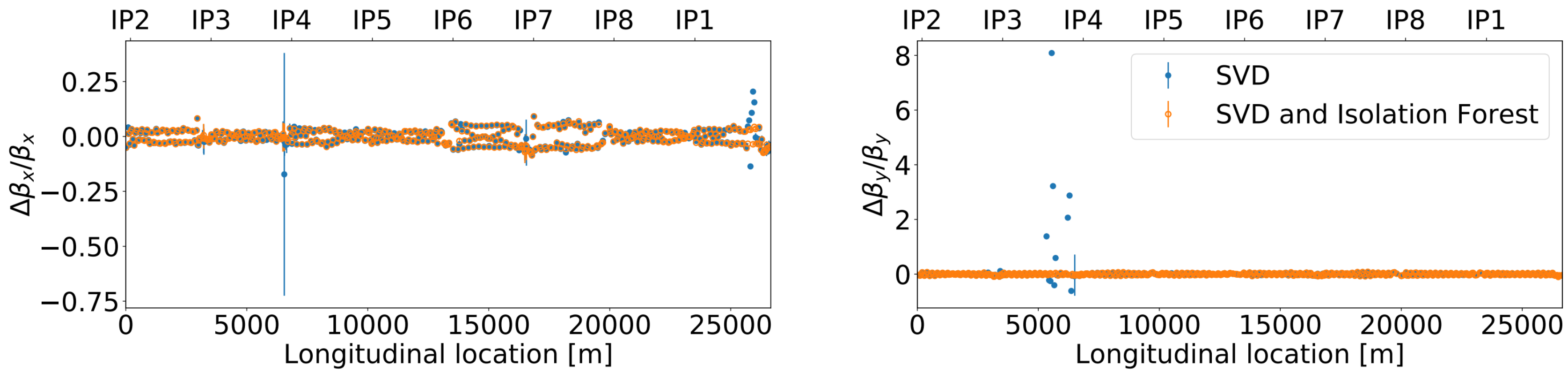
Tuning of IF-algorithm:

→ Trade-off between eliminating bad BPMs and removing good BPMs as side effect by setting the expected contamination rate

→ Optimising in combination with other available cleaning tools (SVD), finding new thresholds



Operational results at the LHC



- Instant faults detection instead of offline diagnostics.

- Full optics analysis is possible directly during dedicated measurements session instead of iterative procedure of cleaning and analysis.

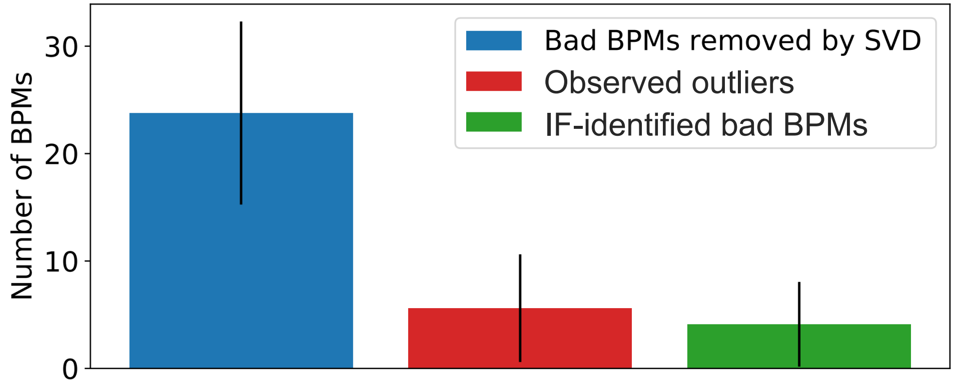
✓ **Fully integrated** into optics measurements at LHC
✓ **Successfully used in operation** under different optics settings.

"Detection of faulty beam position monitors using unsupervised learning", Phys. Rev. Accel. Beams 23, 102805.

Are the BPMs really faulty?

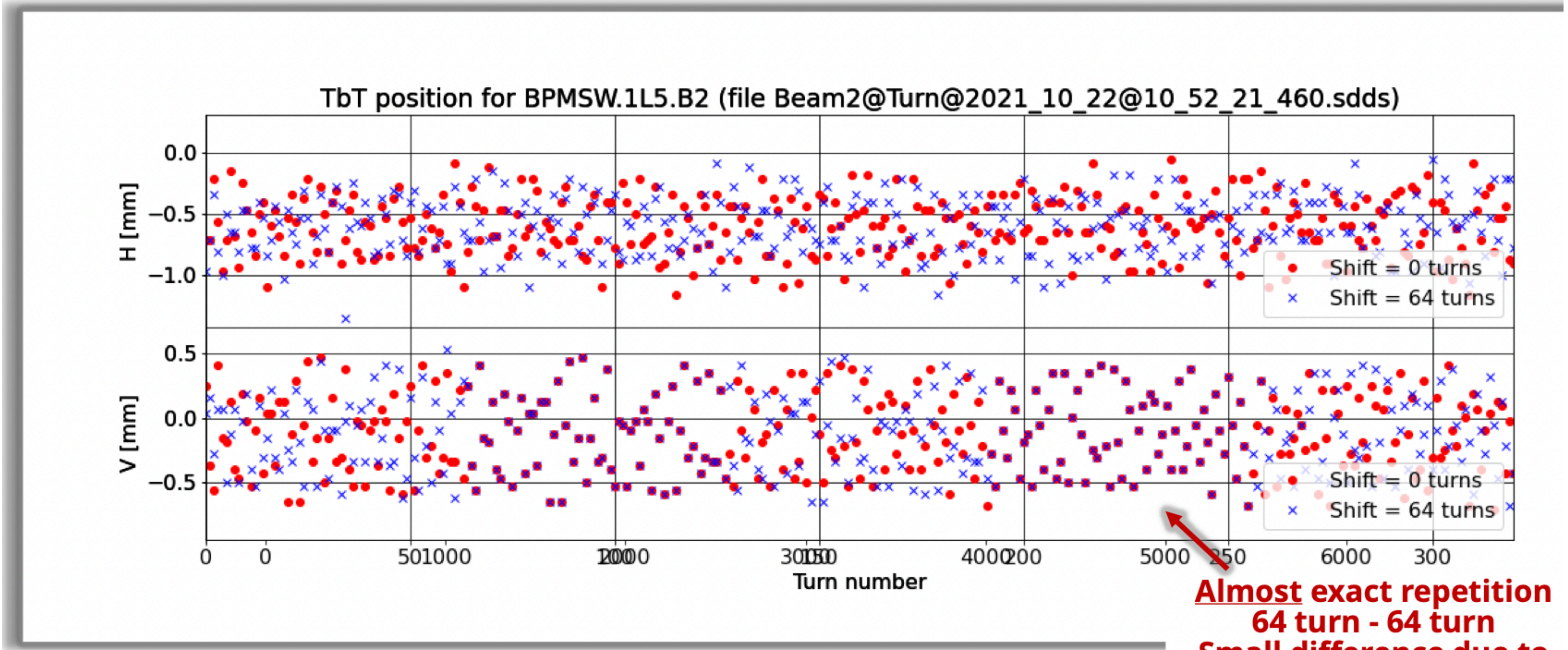
- Collecting cleaning results from different years of LHC operation
- Fault types based on pre-defined thresholds and Isolation Forest input features
- **Extensive analysis and tests done by BI experts**

Reduction of non-physical outliers in beta-beating: Averaged cleaning results, optics measurements in 2018.



Beam Instrumentation checks

BPMSW.1L5.B2



Almost exact repetition
64 turn - 64 turn
Small difference due to cross-term polynomial

BI comment: “repeated samples”, **CRITICAL FOR ABP**

Beam Instrumentation checks

Executive summary of BI analysis

- Out of the 132 BPM flagged by ABP as suspicious
  - **62 BPMs** – look fine to BI (including critical BPMS.2L1.B1)
  - **29 BPMs** – “exact zero” problem, investigations ongoing, beam measurements needed (including crical BPMSW.1L1.B2 and BPMSW.1R5.B2)
  - **27 BPMs** – memory problems, will be solved for Run 3 (including critical BPMSW.1L5.B2)
  - **6 BPMs** – phased incorrectly (only zeros), will be rephased in Run 3
  - **5 BPMs** – disconnected from electronics, already fixed
  - **2 BPMs** – a huge offset (~50 mm) normal due to installation on the dump lines

Beam Instrumentation checks

Executive summary of BI analysis

- Out of the 132 BPM flagged by ABP as suspicious
  - **62 BPMs** – look fine to BI (including critical BPMS.2L1.B1)
  - **29 BPMs** – "exact zero" problem, investigations ongoing, beam measurements needed (including crical BPMSW.1L1.B2 and BPMSW.1R5.B2)
  - **27 BPMs** – memory problems, will be solved for Run 3 (including critical BPMSW.1L5.B2)
  - **6 BPMs** – phased incorrectly (only zeros), will be rephased in Run 3
  - **5 BPMs** – disconnected from electronics, already fixed
  - **2 BPMs** – a huge offset (~50 mm) normal due to installation on the dump lines

Thanks to ML: → **Detection of otherwise unexplored hardware and electronics problems in BPMs**

- 50% of BPMs reported as faulty by cleaning algorithm are actually "broken"
- Verifying false positive BPMs: keeping them in the data does not cause outliers in optics functions, removed as trade-off for detecting actual faults.

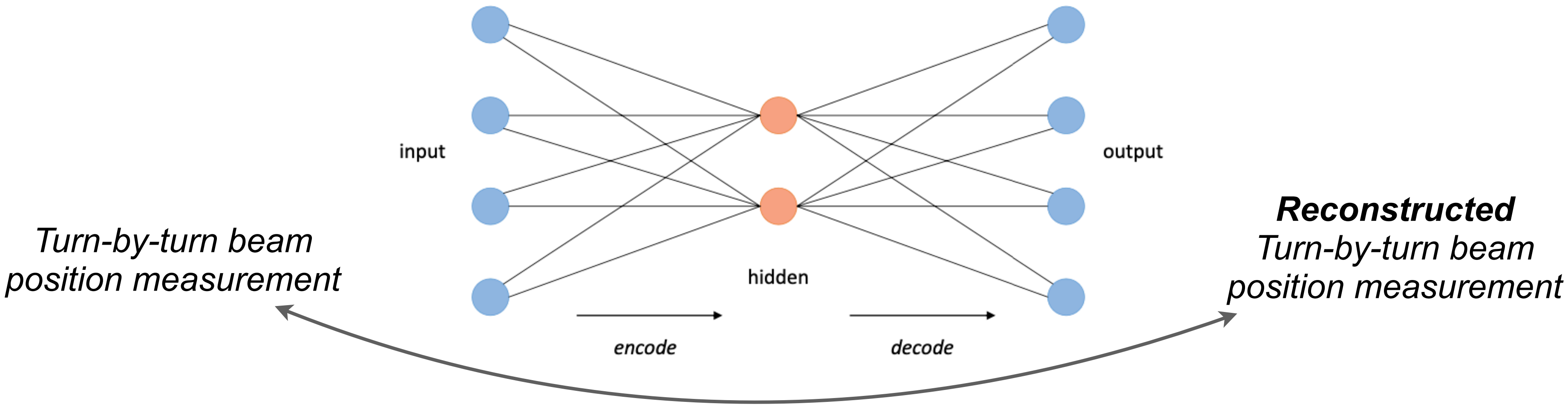
Alternative approach: Autoencoder NN

- Autoencoder can be trained to reproduce the input data in the output layer

Anomaly detection:

1. Training on “clean” data
2. Verify that cleaned signal can be reconstructed with desired low prediction error
3. Reconstruct anomalous signal: prediction error will be higher

—> Need to define a threshold for prediction error to define anomalies



Difference is low for good signal
high for faulty signal

In collaboration with University di Napoli Federico II, A. Apicella, A. Gilardi

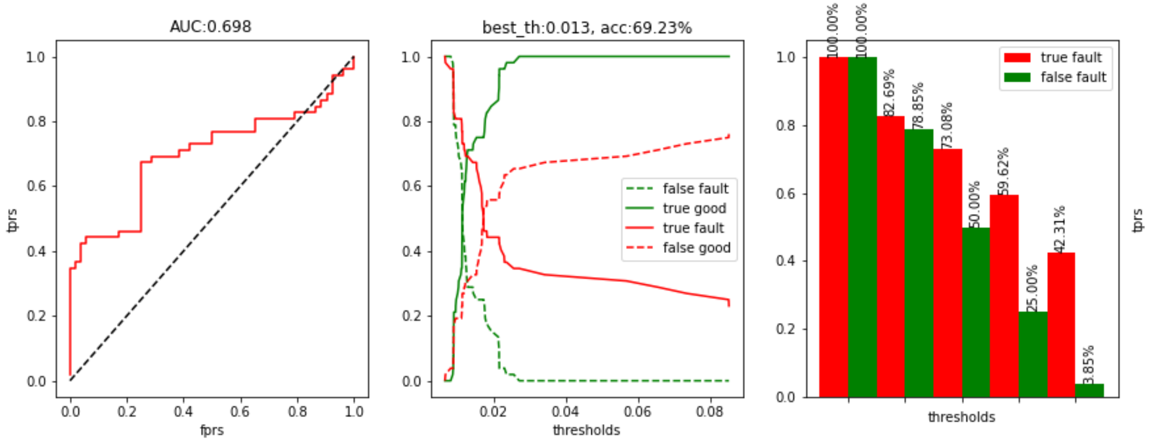
Alternative approach: Autoencoder NN

Advantages compared to existing cleaning tools for the LHC optics measurements:

- Cleaning of different signal artefacts can be done in one step
- Applied directly on raw turn-by-turn data
- Noise reduction can be done at the same time (dimensionality reduction in hidden layers)

Promising results from preliminary studies:

- ✓ Simulated data for training and test
- ✓ Different NN architectures: feed-forward, CNN (with CNN giving accurate reconstruction even for faulty BPMs)



Note: *important to identify as many true faults as possible, on the cost of some false faults*

- ✓ Find optimal threshold for prediction error by analysing ROC-curves

In collaboration with University di Napoli Federico II, A. Apicella, A. Gilardi

Faulty BPMs detection: summary

Instrumentation faults → ❌ Unreliable optics measurements
✅ Detection of faults on early stages of signal processing for optics analysis

Important considerations:

- Unsupervised Learning in this context still requires data (historical, simulations) **to verify the method and to tune the algorithm**
- Decision trees as good alternative to NN: **easier to interpret**

Next steps:

- Applying **Isolation Forest in LHC commissioning after fixes provided by BI** experts and updating cleaning thresholds
- Comparison of **Autoencoder applied to raw tbt-data vs. current cleaning procedure**

Open questions:

- Possibility to **extract fault patterns** from the reduced data representation in hidden layers of Autoencoder

More Unsupervised Learning: Betatron tune measurement

Online tune measurements from LHC BBQ —> used for optics analysis, e.g. K-modulation

Problem: outliers due to wrong data acquisition

❌ Tune uncertainty —> imprecise measurements of derived quantities

❌ Manual fitting and cleaning of outlier measurements

➡ **Unsupervised Learning to automatically detect outliers in tune measurements**

Approach:

- treating tune measurements as time series: how to deal with the changes in working point?

➡ Qx,Qy- space

➡ Clustering to distinguish noise from signal and classify different working points segments

UNSUPERVISED LEARNING TECHNIQUES FOR TUNE CLEANING MEASUREMENT, H. GARCIA-MORALES, E.FOL, R. TOMAS,IPAC’21 doi:10.18429/JACoW-IPAC2021-MOPAB184

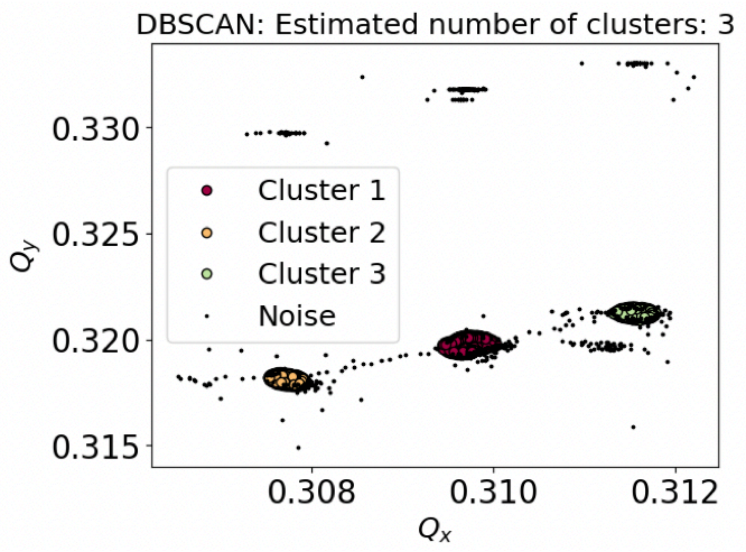
Betatron tune measurement

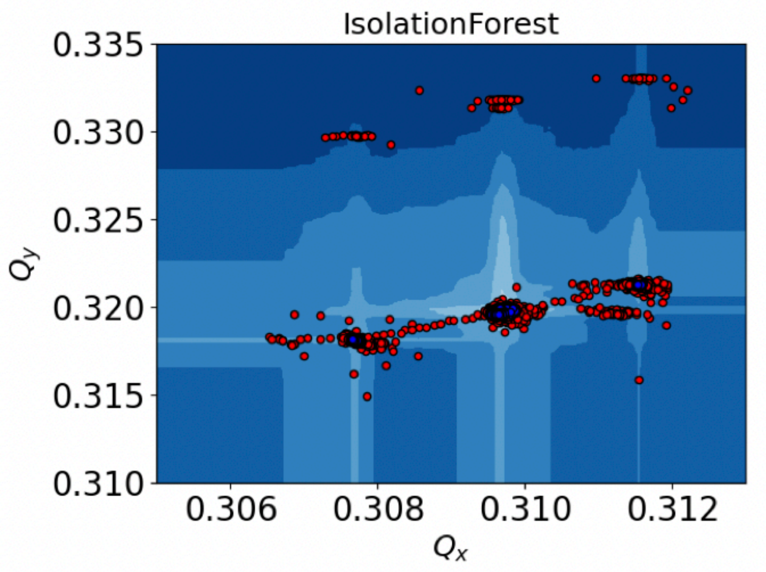
Clustering algorithms: k-means, DBSCAN, Local outlier factor, Isolation forest, ….

How to choose the most appropriate one?

—> Number of parameters to tune, ability to deal with noise?

DBSCAN vs. Isolation Forest





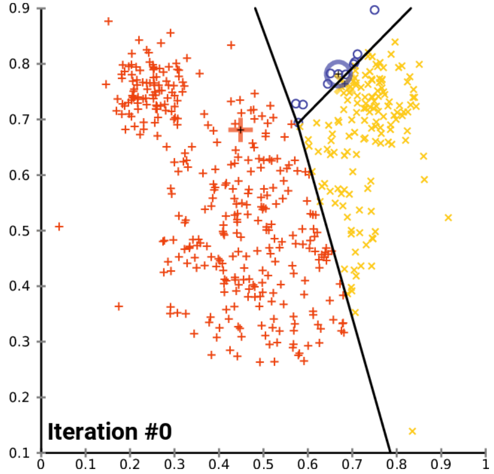
- Both algorithms correctly identify the outliers
- DBSCAN can automatically detect the clusters corresponding to different working points

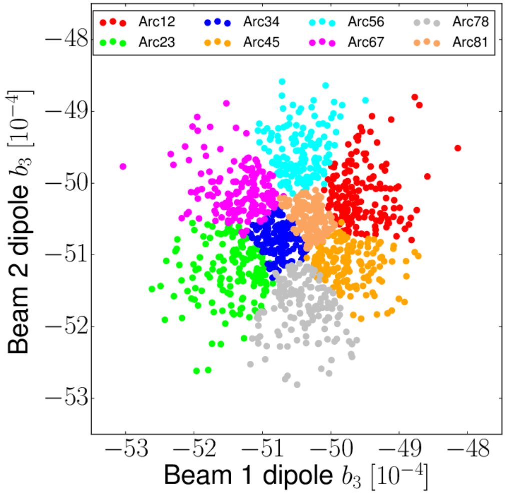
Clustering algorithm for magnet sorting in HE-LHC

- Twin aperture dipoles in (HE-)LHC
  - Only the average over a sector is corrected
  - A priori, magnet field errors between apertures not correlated

- Grouping dipoles with similar error in one aperture may spoil other aperture
  - Use of clustering to optimize on both apertures
  - Additional constraint: Same size clusters

- Use of modified K-means clustering algorithm
  - Groups dipoles with similar error in both apertures together
  - Shown to increase dynamic aperture
  - Option to include more components

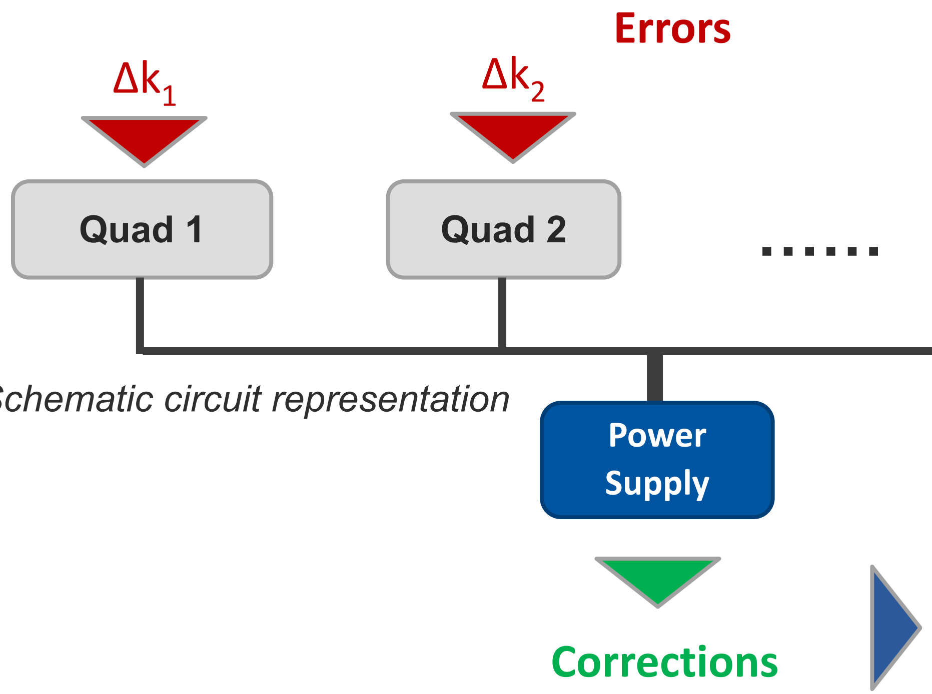




IPAC19 MOPMP023

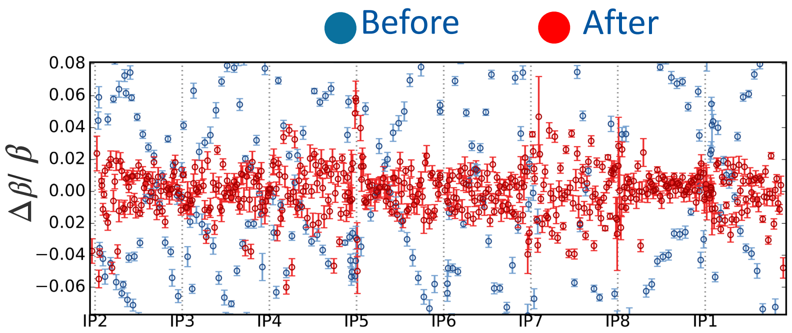
Supervised Learning
for Optics Measurements and Corrections

Correcting the optics



Schematic circuit representation

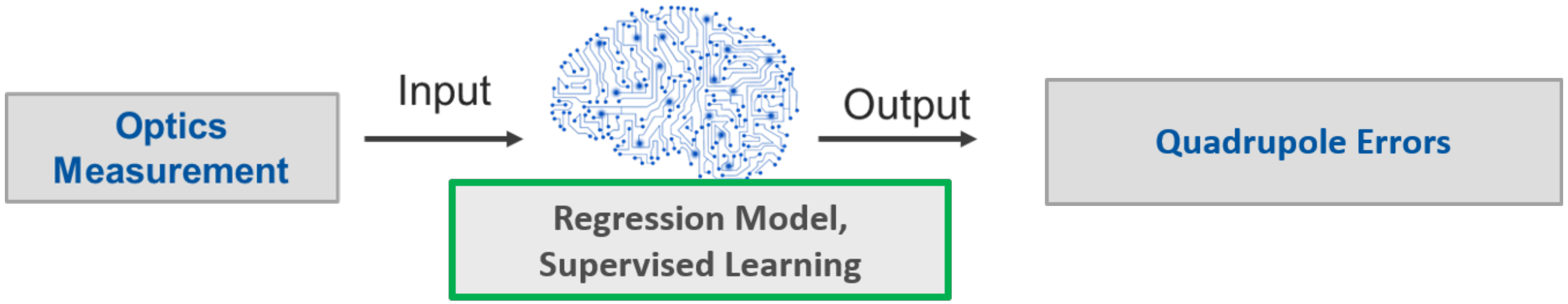
- Corrections are implemented by changing the strength of **circuits**
- Optics perturbations are caused by all **individual magnets.**



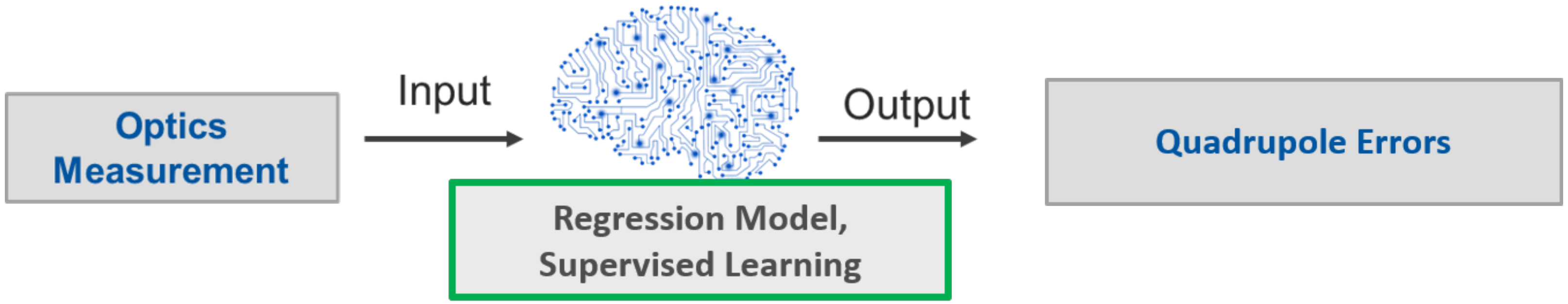
➢ What are the **actual errors of individual quadrupoles?**

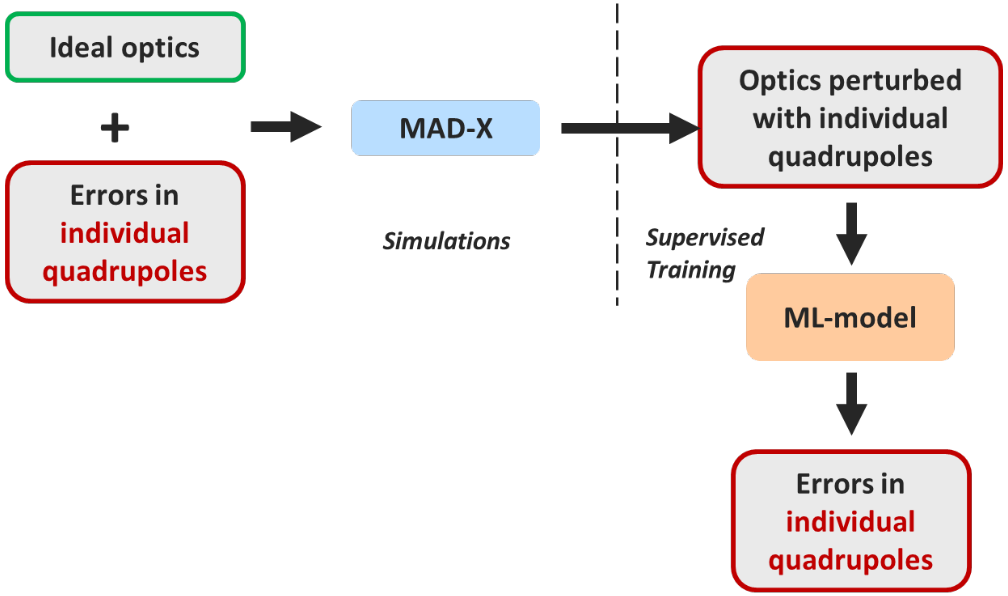
➢ How to obtain the **full set of errors in one step**?

Estimation of quadrupole errors



Estimation of quadrupole errors





<u>Training ML- regression model:</u>

- **1256 target** variables: randomly assigned field errors in quadrupoles + other error sources (dipole errors, sextupoles misalignments)

- **3304 input** variables: optics functions (phase advances, β-function in IRs, normalised horizontal dispersion)

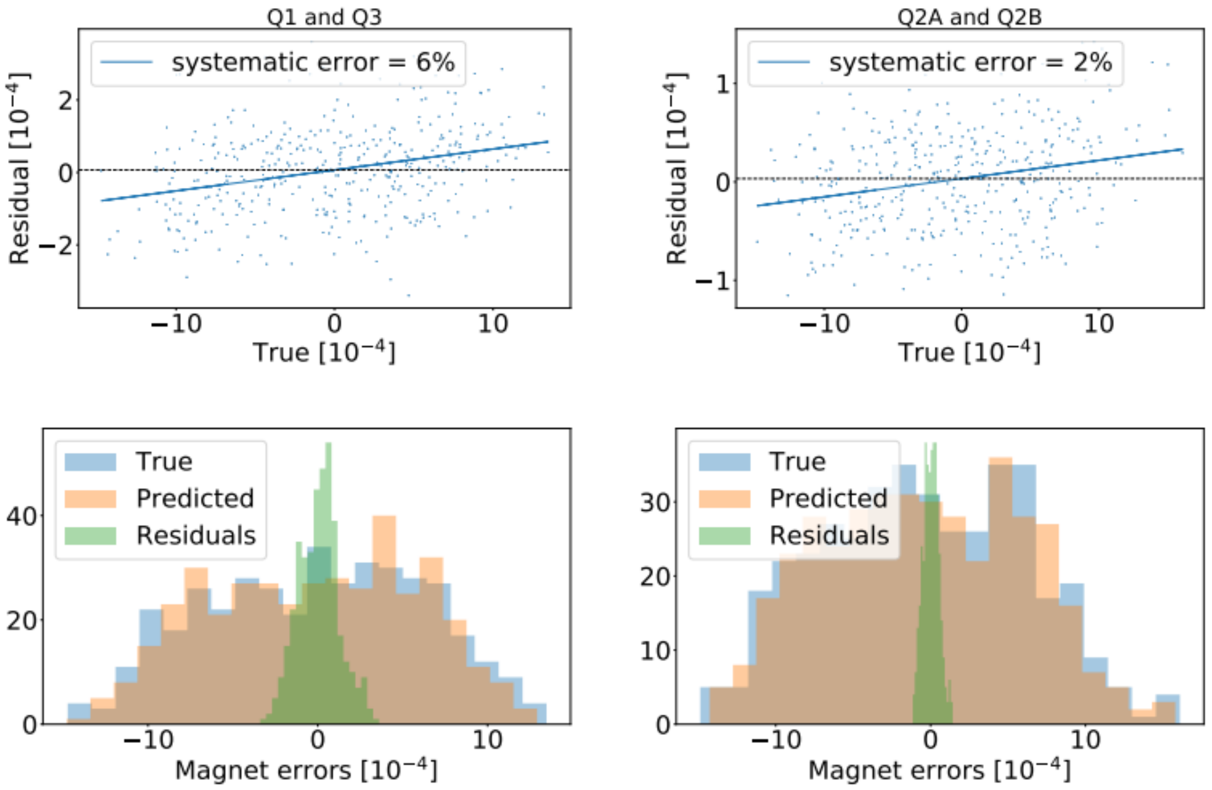
- Using **Ridge Linear Regression as baseline model**

$$\min_{w} \|Xw - y\|_2^2 + \alpha \|w\|_2^2$$

Verifying ML approach: simulations

Simulations: **true magnet errors are known**

→ directly compare prediction to simulated data → **residual error**



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

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

How well can we correct the optics with predicted errors?

Estimation of quadrupole errors: measurements

Measurements: **true magnet errors are unknown**

→ **Control beta-beating**

Measurement → **ML-model**

Difference?

predicted errors

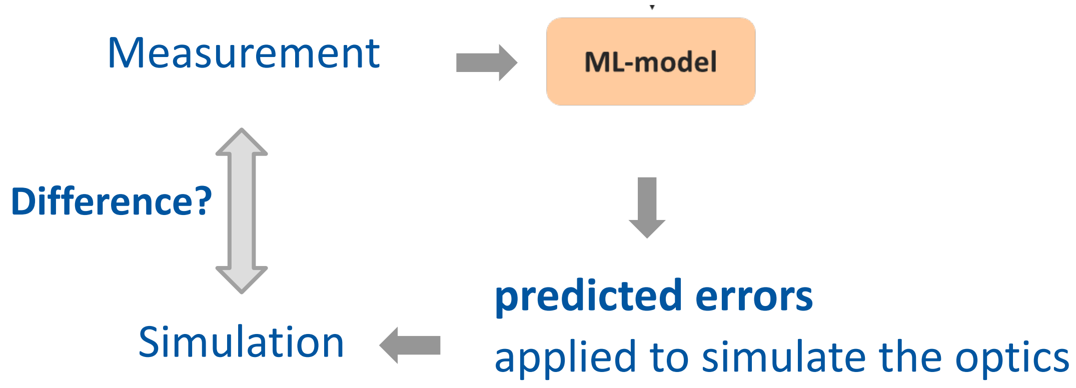
applied to simulate the optics

Simulation

Estimation of quadrupole errors: measurements

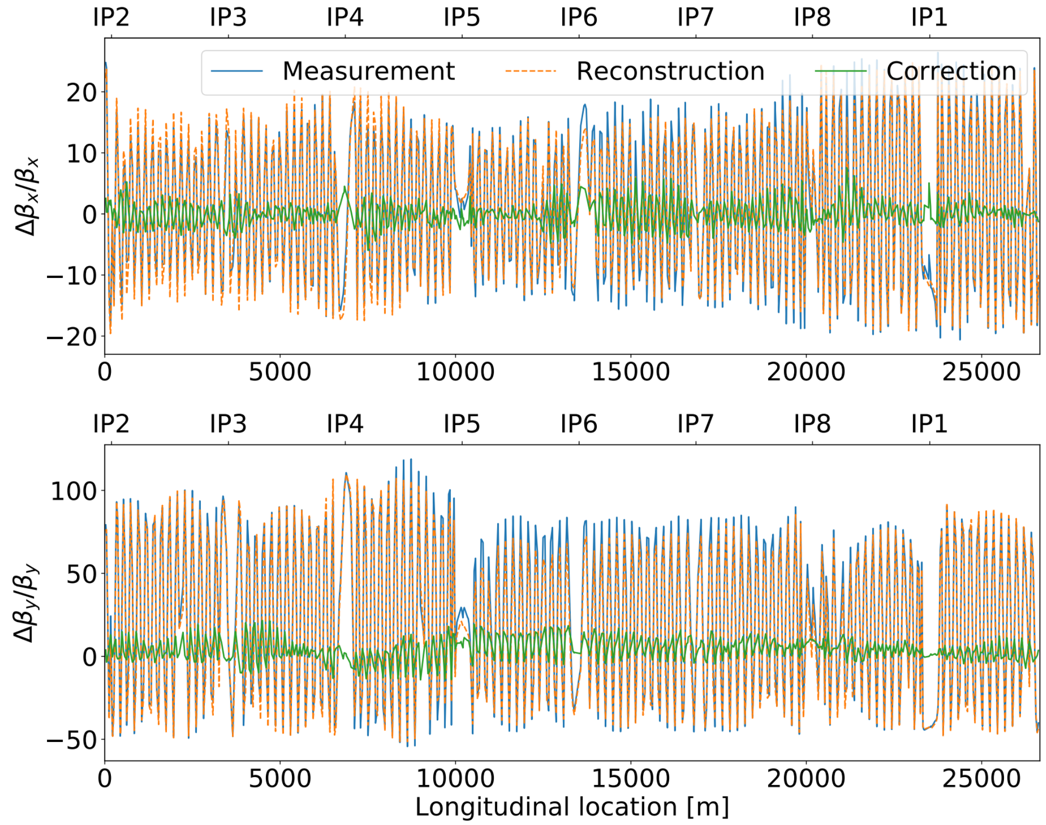
Measurements: **true magnet errors are unknown**

→ **Control beta-beating**



- ✓ New method for **local optics corrections**
- ✓ Improved knowledge of **direct error sources**
- ✓ **Simultaneously** obtaining quadrupole errors for both beams, at every location.

Test on ***LHC optics measurements, uncorrected*** *machine*



"Supervised learning-based reconstruction of magnet errors in circular accelerators", European Physical Journal Plus volume 136, Article number: 365 (2021) ,

Reproducing the measured β –beating with average rms error of 7% and below 3% at IPs.

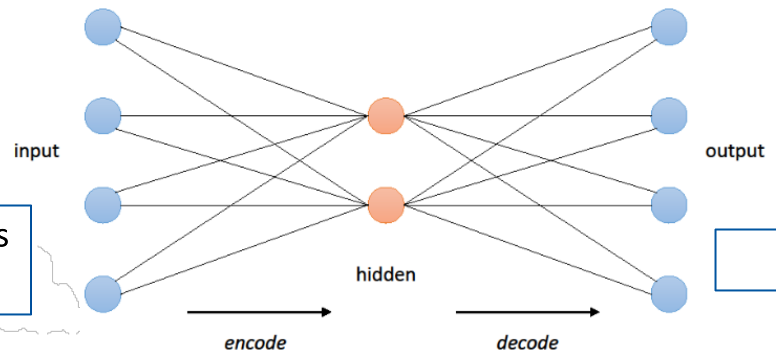
Denoising of optics measurements

Denoising of optics measurements

- Uncertainties in the measured optics functions → **"noise"** → Noise in the measurements degrades the performance of corrections techniques

Autoencoder Neural Network

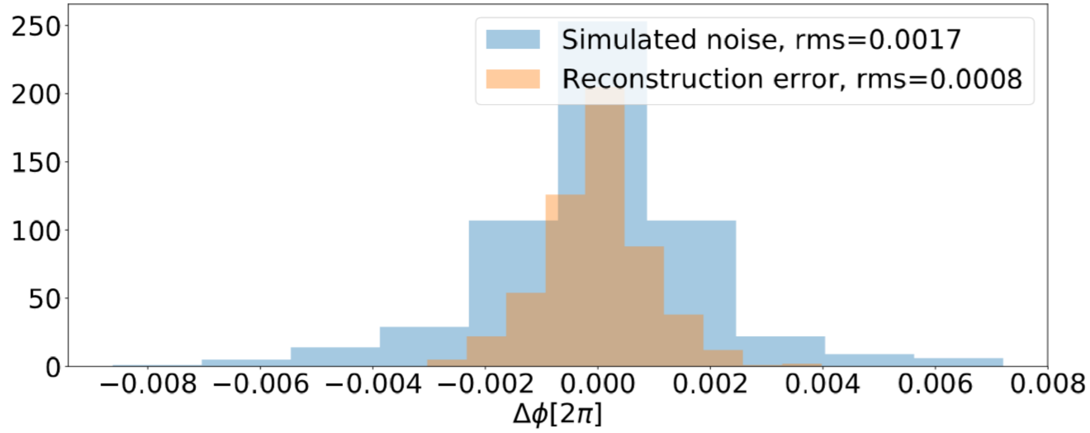
Simulated optics observables **+ noise**



Denoised optics

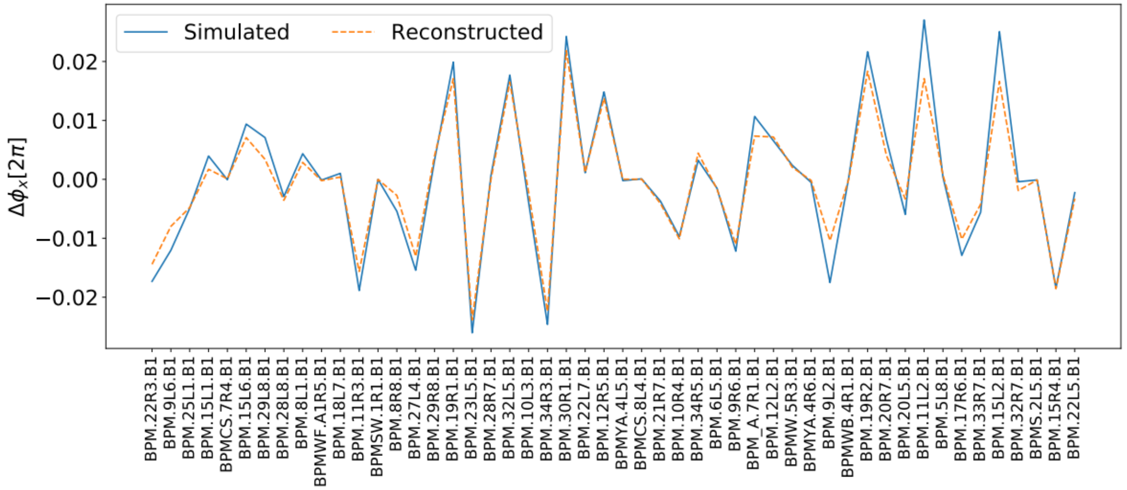
Denoising of optics measurements

Simulated data: Noise Reduction



✓ Reconstruction error is by factor 2 smaller than the noise present in the signal.

Simulated data: Reconstruction



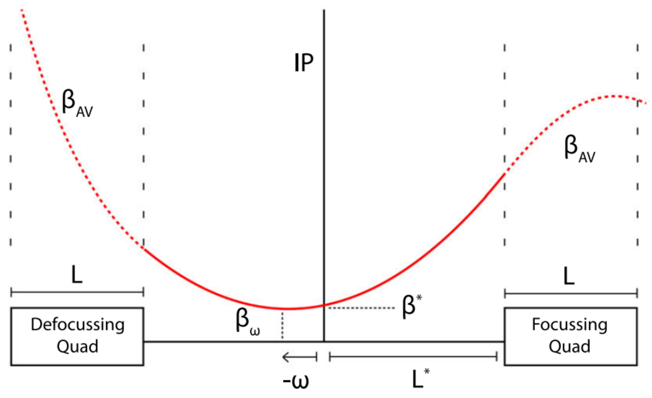
✓ Reliable reconstruction after denoising

- ➢ Potential improvement of measurements quality
- ➢ Possibility to reconstruct the phase advance at the location of faulty BPMs.

Reconstruction of advanced optics observables

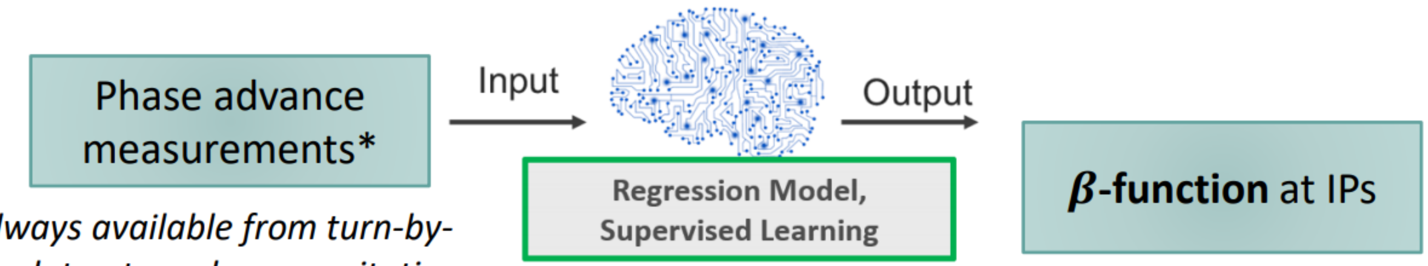
Reconstruction of β-beating in Interaction Regions

➢ Special technique to measure beta-function at IP is needed:



- Modulation of quadrupole gradient
- Computation of average beta-function
- Propogate beta-function values to IP

➢ How to reconstruct optics observables **without direct measurements?**

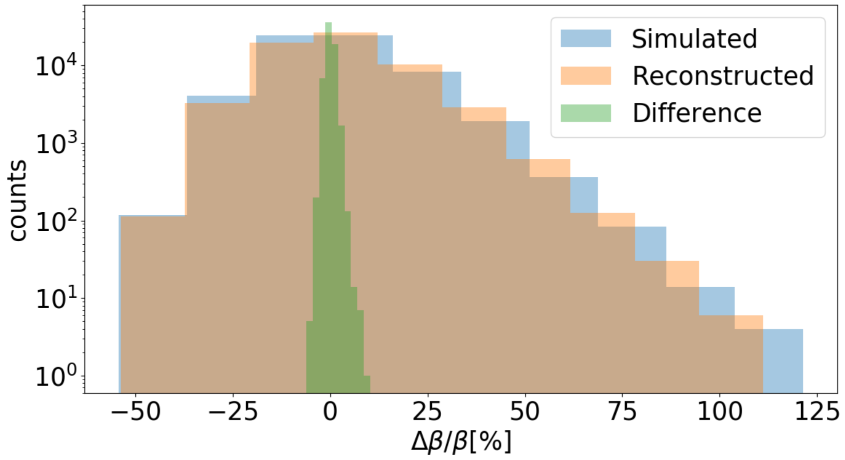


** Always available from turn-by-turn data at one beam excitation*

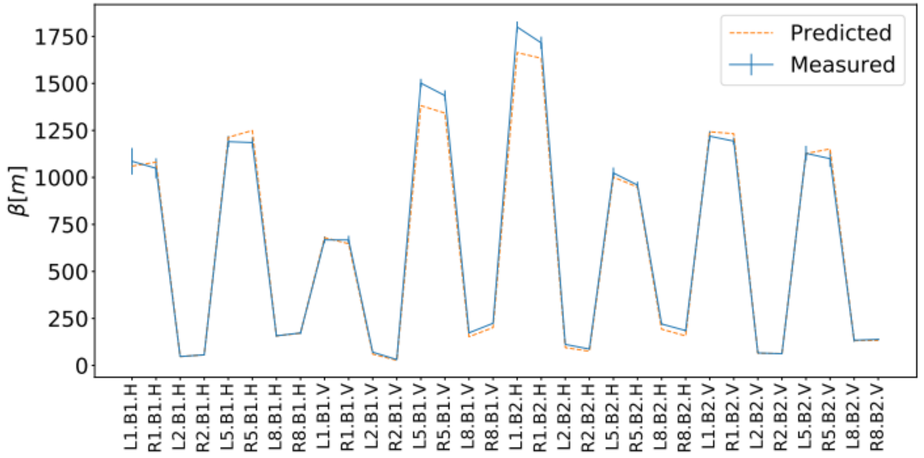
Simultaneously for beam 1 and 2

Reconstruction of β-beating in Interaction Regions

Simulations



LHC Measurements, BPMs left and right from Interaction Points



Reconstruction error: $$\frac{\beta_{simulated} - \beta_{reconstructed}}{\beta\_simulated} = 1\%$$

✓ comparable to measurement uncertainty of traditional method.

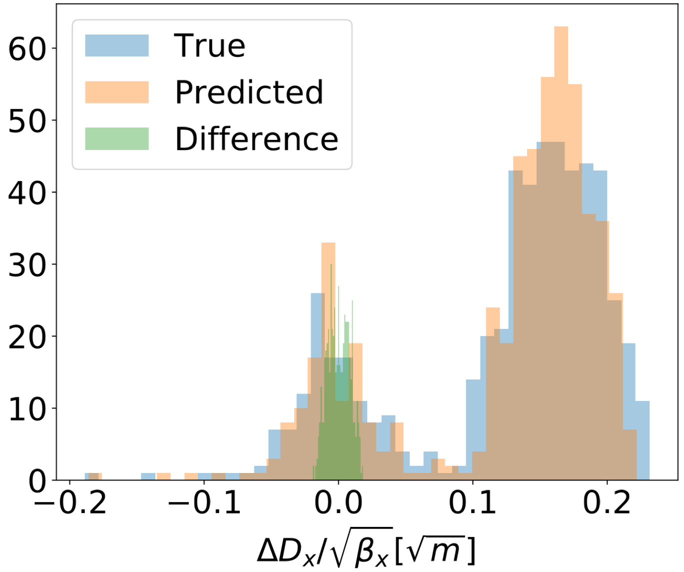
✓ Great potential to reduce measurements time

✓ Applicable to estimation of other optics observables (e.g. horizontal dispersion)

Reconstruction of horizontal dispersion

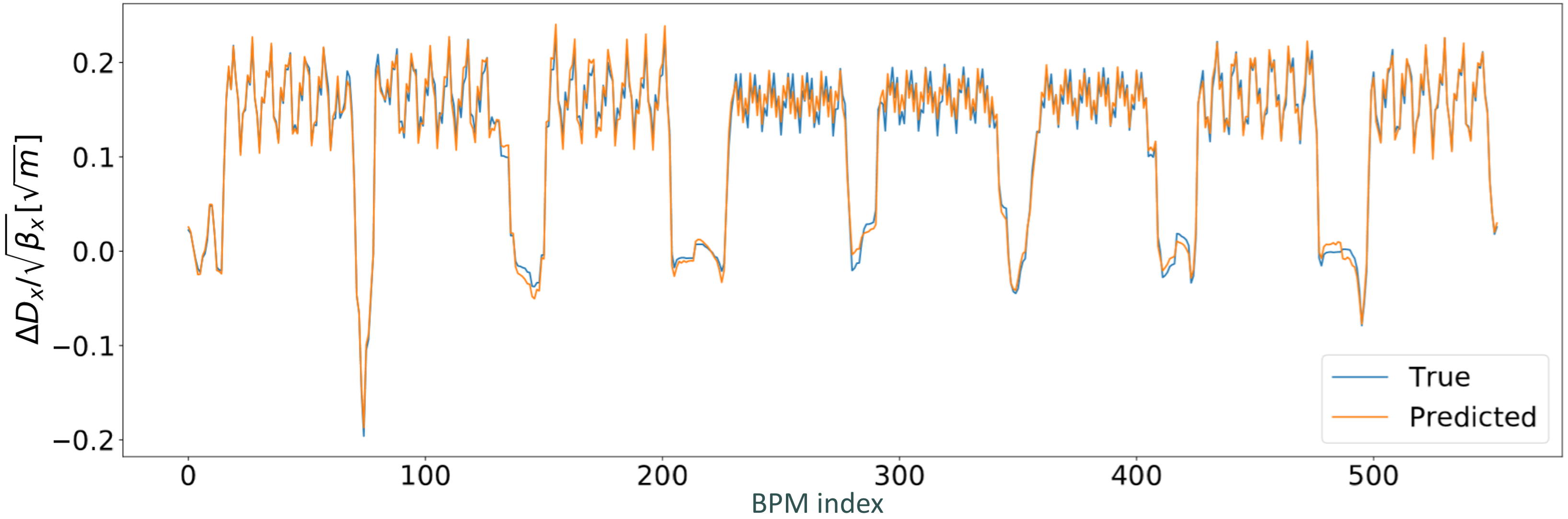
- **Input**: simulated phase advance deviations given noise
- **Output**: normalized dispersion $\Delta\, Dx/\sqrt{\beta x}$
- Using **linear regression model**: Ridge Regression, 10 000 samples

***Simulation** example: Beam 1*



***Simulated rms** $\Delta\, Dx/\sqrt{\beta x}$ **: 0.0802** $\sqrt{m}$*

***RMS-error between simulation and reconstruction: 0.007** $\sqrt{m}$*



Supervised Learning approach for optics corrections

Providing simulation data to find a general mapping between error sources and optics observables

→ Simulation studies on the effects of different error sources

→ One data set can be used to build several models / applications (quad errors prediction, optics reconstruction, measurements denoising)

Important considerations:

- Data is everything: realistic simulations —> sufficiently general models
- Systematic data collection and management (e.g. expert systems?)

Continuing the Supervised Learning path

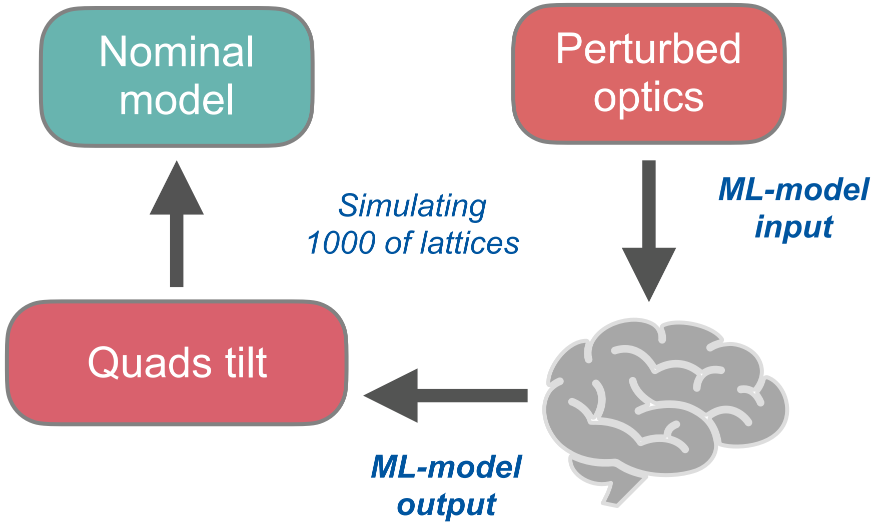
Betatron Coupling Sources Prediction

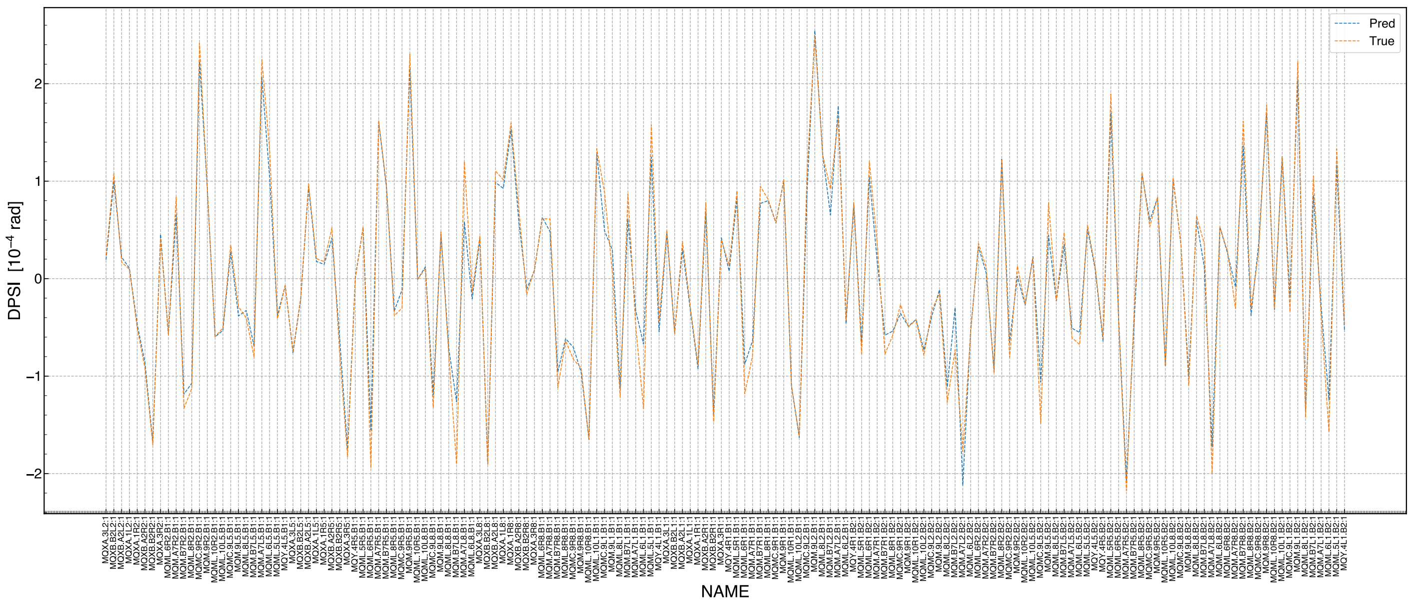
- Knowledge of sources is very valuable for correction.
- Resonance Driving Terms: obtained from harmonic analysis of tbt-data
- A coupling source (e.g. tilt of a quadrupole) will create an abrupt jump on the coupling RDTs —> indicate the location of coupling sources
- Challenge in Interaction Regions: unfavourable phase advance and “lack” of BPMs —> how to link observed RDTs to a specific coupling source?

➡ Working on an ML model that would be able to accurately predict the location and relative strength of coupling sources.

Betatron Coupling Sources Prediction

Work by Felix Soubelet, BE-OP





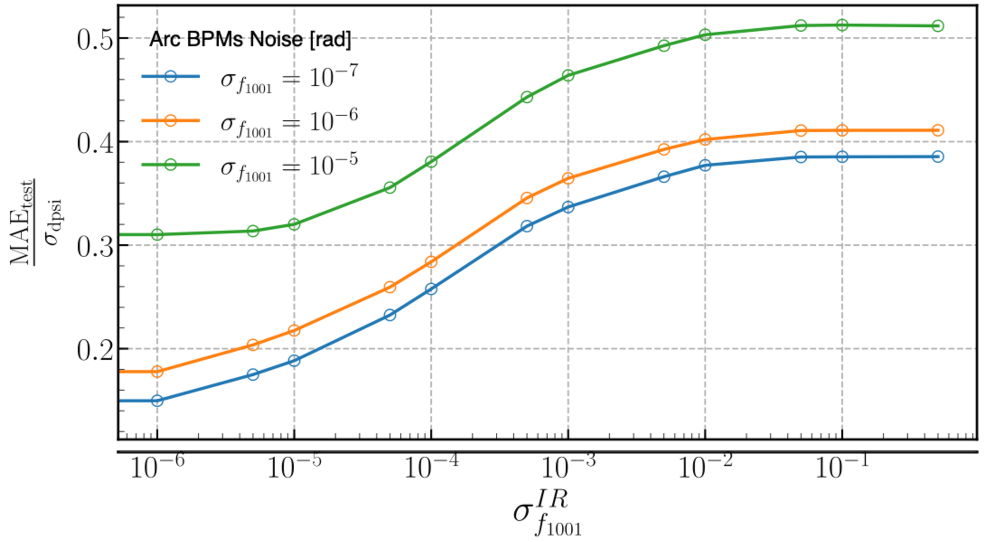
Input: RDTs simulated/measured for beam 1 and beam 2

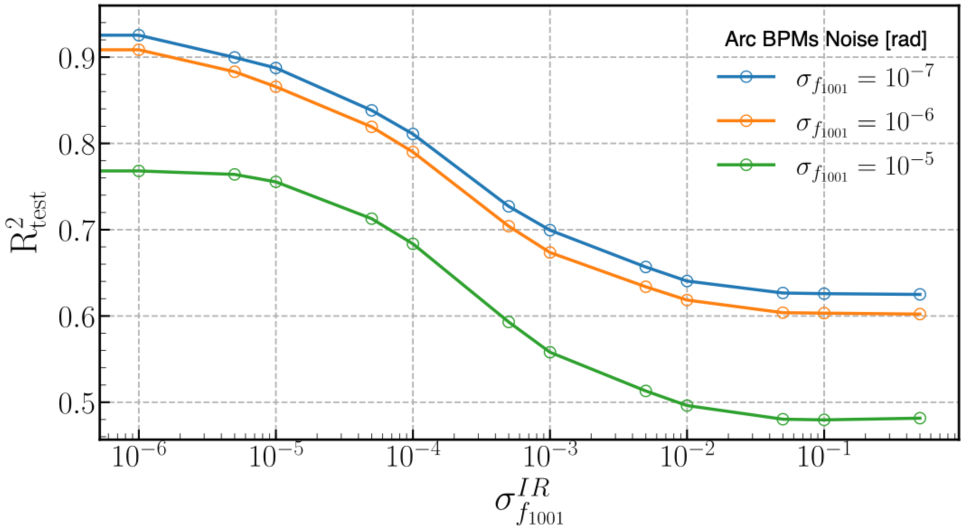
Output: Misalignment of quadrupoles in all IRs

✓ Current simple model (Linear regression with regularisation) already demonstrates relatively accurate predictions.

Betatron Coupling Sources Prediction

- Noise in the measurements degrades the model performance
  —> determine requirements on instrumentation/ analysis for the acceptable level of noise





- Current work:
  - Denoising of reconstructed RDTs
  - Higher complexity of prediction models (Decision trees, NN)

Optics control in HL-LHC studies

High Luminosity Large Hadron Collider: Upgrade of the LHC to push the performance in terms of beam size and luminosity.

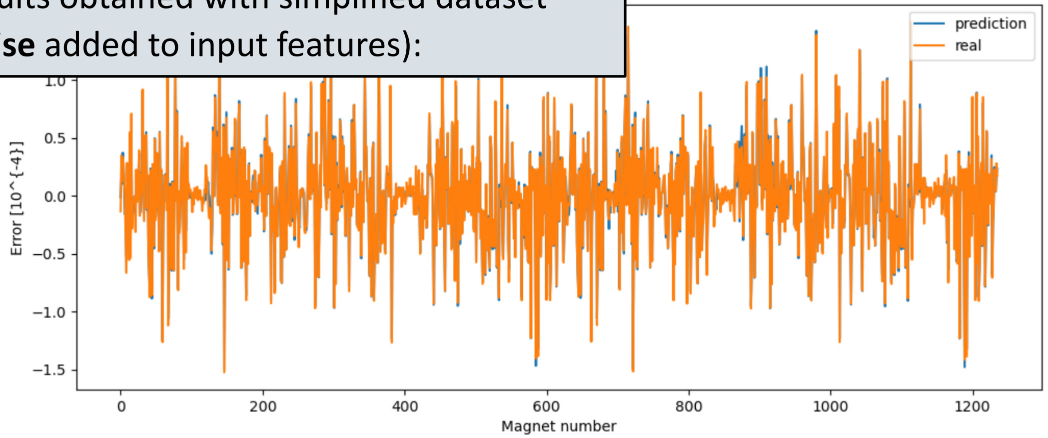
•The local linear optics correction at the IR will be essential to ensure the HL performance.

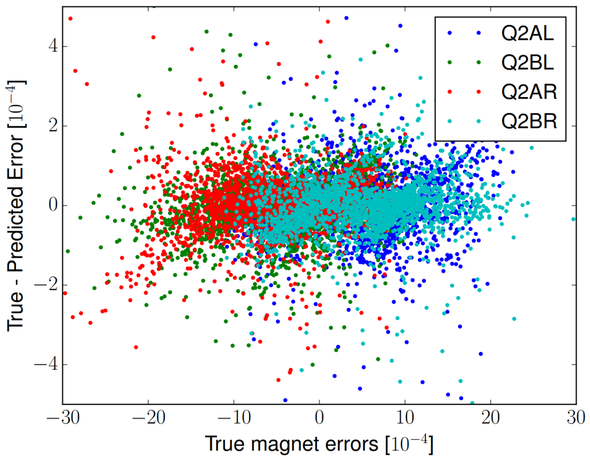
•Current LHC strategies might impose limitations → new correction strategies are needed.

Work by Hector Garcia Morales, BE-ABP

Preliminary results obtained with simplified dataset
(**no noise** added to input features):

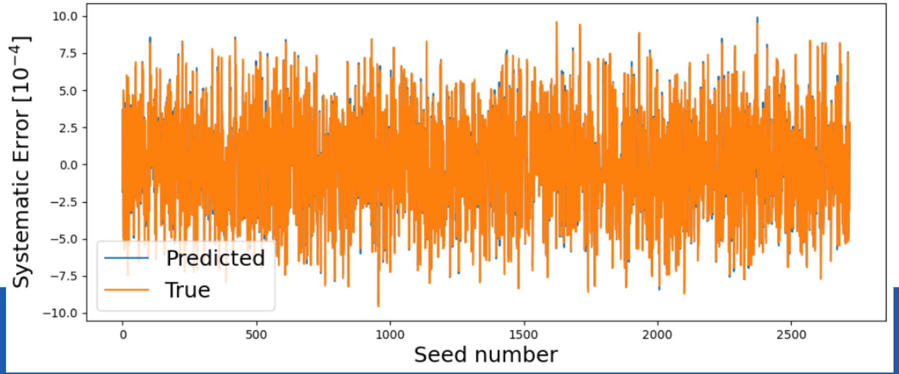
Full set of quadrupoles all around the ring





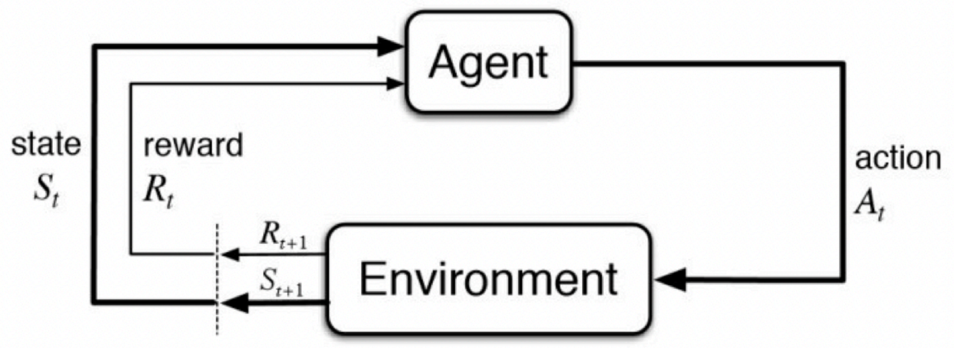
Inner Triplet magnets in Interaction Regions

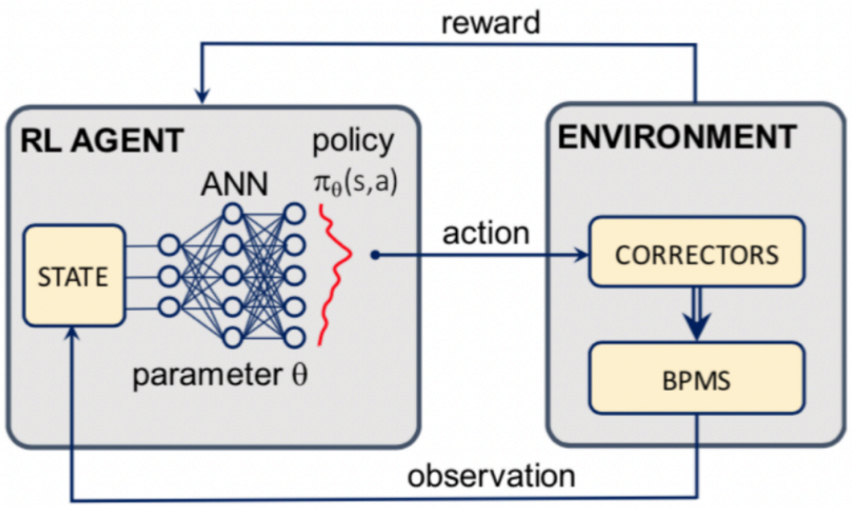
- Systematic part of the gradient error (unknown) may have a significant impact on the β-beating.



Optics control in HL-LHC studies

Reinforcement learning - based local corrections
- Uses the previously presented approach **to learn LHC model from simulated data**





Based on V.Kain et al., "Sample-efficient RL for CERN accelerator control"

- Environment = Surrogate model of HL-LHC lattice
- Reward = Average beta-beating in IRs
- State space = Quadrupole strengths (only triplet magnets for now)
- Action space = Correctors settings

Optics control in HL-LHC studies

Implementation

- Introducing magnetic errors in triplet magnets in IR1
- RL algorithms implementations based on OpenAI
- PyTorch for the training of critic networks

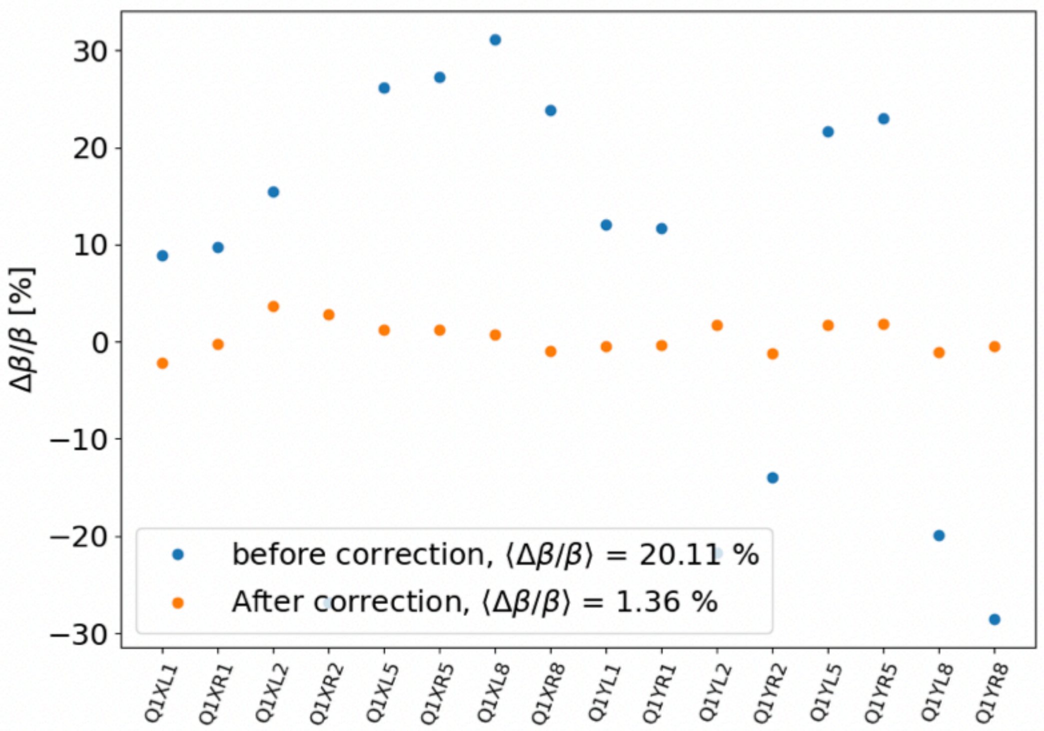
Results:

After the learning process, the model is able to perform the optics correction in one single iteration with residual β-beating of 1-2% (up to 20% initially)

Open questions:

- Understanding the model behaviour
- Comparison with other correction techniques
- Extending the problem by adding more error sources

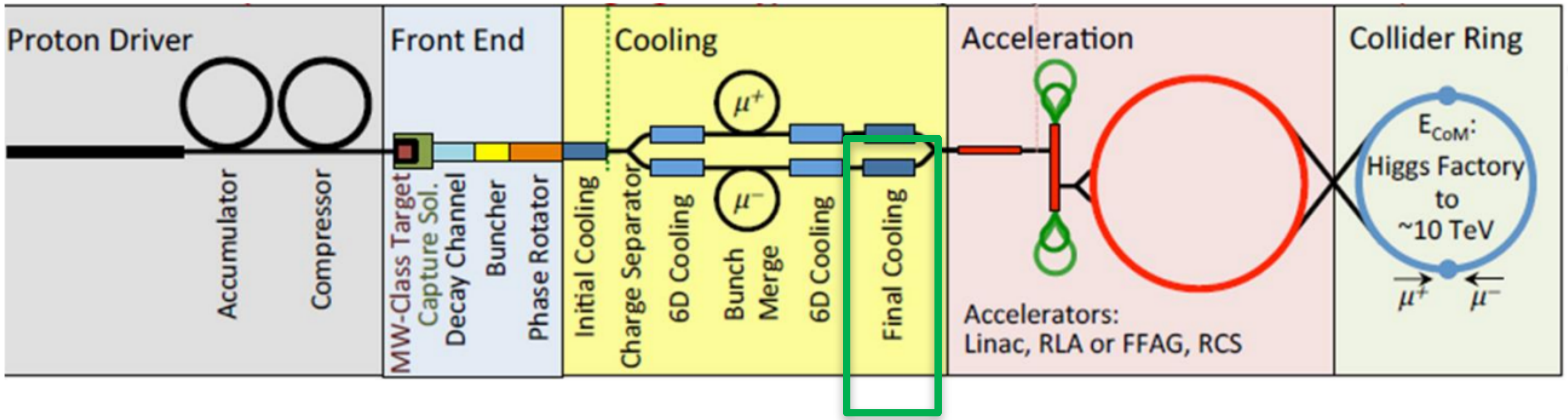
Work by Hector Garcia Morales, BE-ABP



Muon collider design studies: Final Cooling

Challenges of Final Cooling for the Muon Collider

- Proton driven scheme: muons are produced by p+-target interaction
- Muon beam is produced with a **large transversal momentum**
  —> cooling is required
- Short lifetime of muons —> **ionization cooling**



$$L \propto \frac{\gamma^3}{CI} \frac{N_0^2}{\epsilon_{\perp,N}}$$

https://muoncollider.web.cern.ch/design/general-parameters

- Beams with transversal emittance ϵ_{trans} of 0.3 mm are provided after the 6D cooling
- Final cooling: ϵ_{trans} = 0.05 mm has been achieved by H. K. Sayed (10.1103/PhysRevSTAB.18.091001)
- ϵ_{trans} = 0.025 mm is expected to be required before acceleration.

[1] U.S. Muon Accelerator Program, FERMILAB-CONF-13-307-APC

Final Cooling concept

- Transverse Cooling:

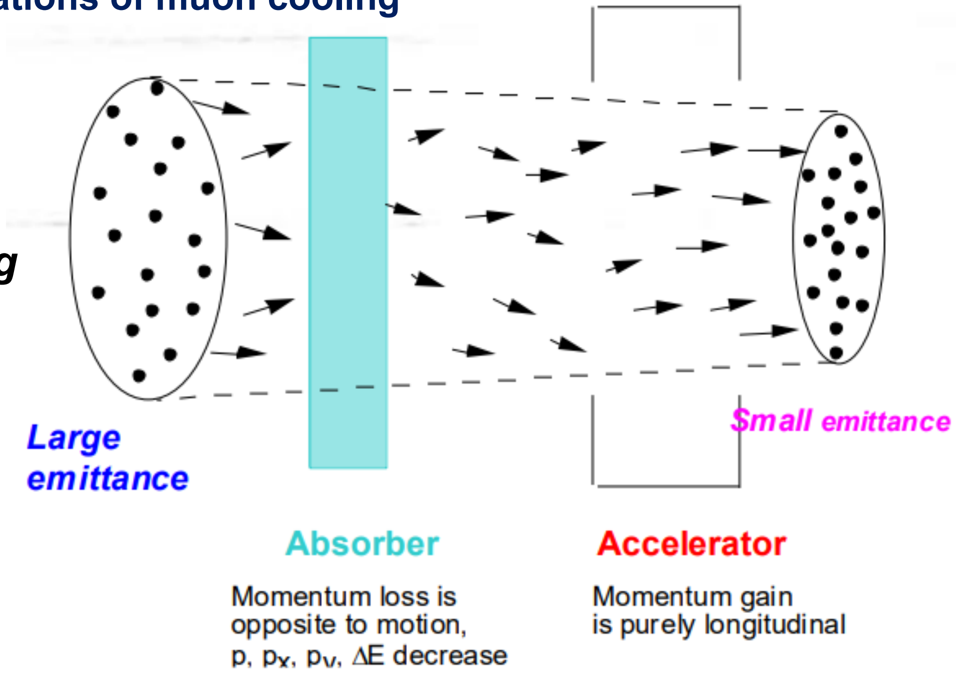
$$\boxed{\frac{d\varepsilon_T}{ds} = -\frac{1}{\beta^2 E}\frac{dE}{ds}\varepsilon_T + \frac{\beta\gamma\beta_T}{2}\frac{d\theta_0^2}{ds}}$$

Energy loss term

Multiple scattering term

To ***minimize the heating effect****, the absorbers are placed in a* ***strong focusing field****.*

David Neuffer, „Principles and applications of muon cooling“



- Minimum Emittance:

$$\varepsilon_{N,eq} \cong \frac{\beta_t E_s^2}{2\beta mc^2 L_R (dE/ds)}$$

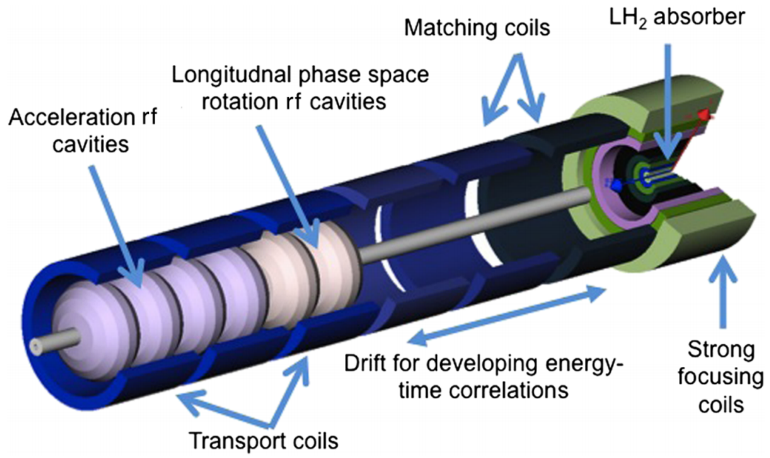
L radiation length, E muon energy, $\beta_\perp$ transverse β-function, $\frac{dE}{ds}$ = energy loss

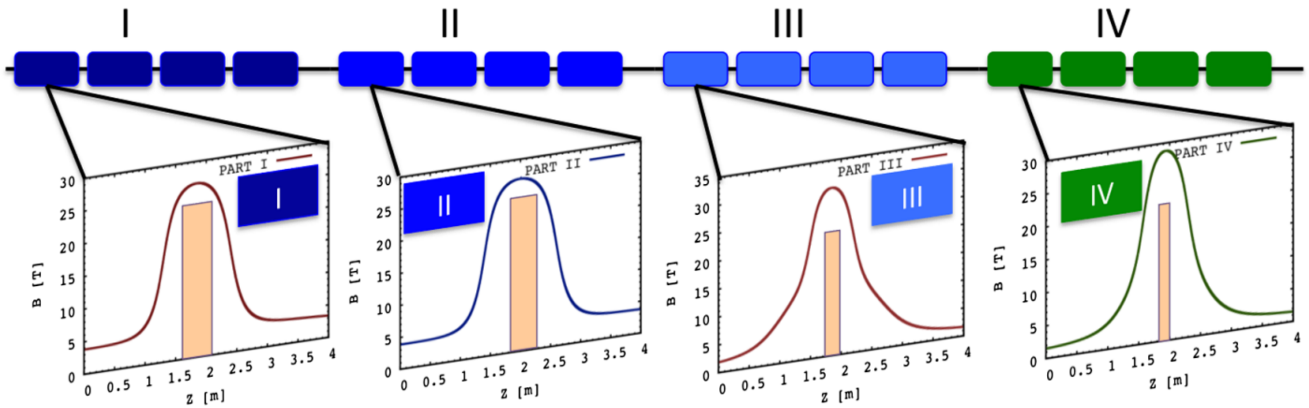
- **Final cooling:** high field solenoidal channel (up to ~30 T) placing absorber inside (e.g. liquid hydrogen).
- Challenge: strong focusing to get low emittance —> higher fields and lower momenta
  - cause more longitudinal emittance growth, energy spread
- Control the optics in absorber regions, minimise energy spread and **reduce the transverse emittance**

Final Cooling baseline

- A Gaussian input beam with **ε**$_{\perp}$=**300 μm and ε**$_{\parallel}$ **= 1.5mm**
- For final cooling, the beam momentum is reduced initially to **135 MeV/c**

- High-field magnets limited to 25—32 T, and the cooling beam momenta ranged from 135 MeV/c to 70 MeV/c (40 to 20 MeV kinetic energy)
- Cooled to **ε**$_{\perp}$ = 55 μ and **ε**$_{\parallel}$ = 1.5 mm, with a transmission of 50%





High field – low energy muon ionization cooling channel
Hisham Kamal Sayed, Robert B. Palmer, and David Neuffer
Phys. Rev. ST Accel. Beams **18**, 091001 – Published 4 September 2015

First steps towards applying ML

- Python “wrapper” for launching ICOOL, providing p_z, $\varepsilon_{\perp,start}$, B-field (coils parameters), absorber settings
  - ✓ automatic computation of initial beam distribution, **generation of ICOOL code**
  - ✓ Additional analysis in Python
  - ✓ **Storing input and output of simulation in well-structured format (JSON)**

✓ Simplified optimization set-up

✓ Easy integration of optimization methods

✓ Applied to **linear optics optimization** and **emittance reduction.**

Building a framework for automatic optimisation

Python | *ICOOL* | *Ecalc9 (Fortran) Python*

Vary parameters of the field → tracking → compute the optics

→ **evaluate the objective function → adjust fields/absorbers**

Python | *Python*

Applied optimizations methods:

- Nelder-Mead: Simplex algorithm, robust in many applications, but doesn’t allow multiprocessing
  https://docs.scipy.org/doc/scipy/reference/optimize.minimize-neldermead
- Differential Evolution: stochastic population-based method, allows parallelization
  https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html
- **Extremum Seeking:** performs small oscillations in parameter space to find global solution

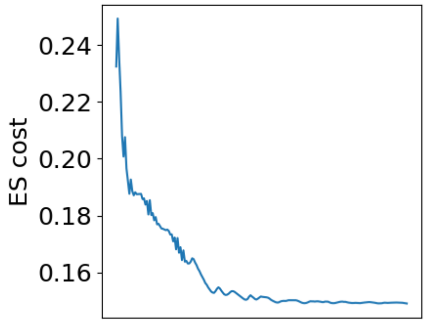
A. Scheinker and D. Scheinker, “Constrained extremum seeking stabilization of systems not affine in control,” International Journal of Robust and Nonlinear Control 28, 568–581 (2018)

First results: simplified lattice

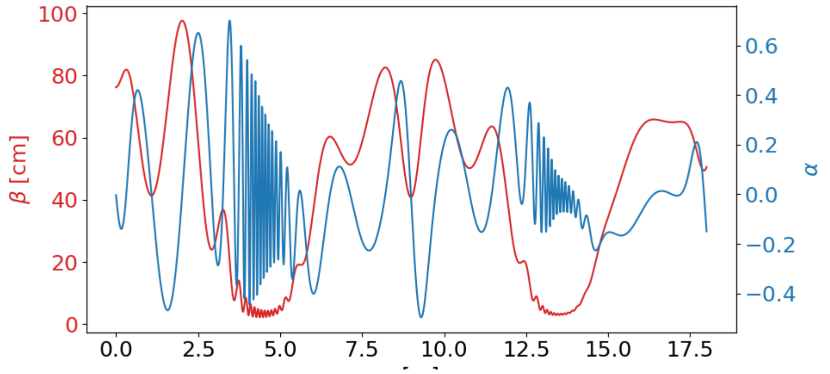
Note: simplified lattice, no re-acceleration

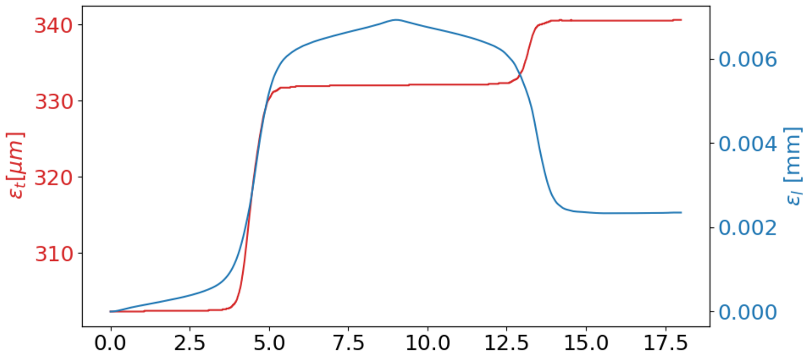
- each cell containing of 3 coils x 4 sheets, absorber density, initial momentum and beta- function

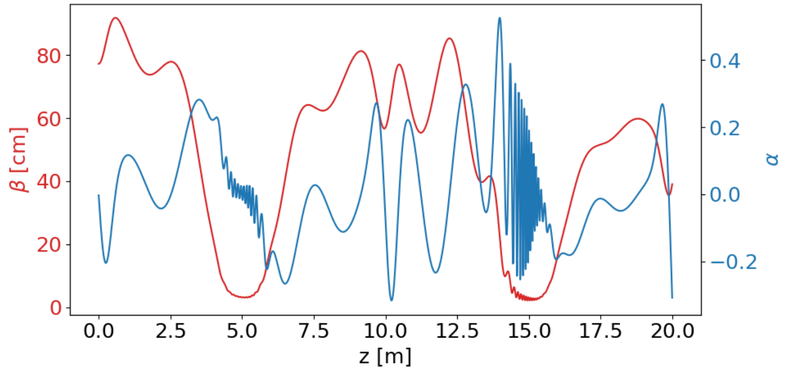
➡ **Extremum seeking algorithm:** much faster for a larger parameter spaces, easily extendable

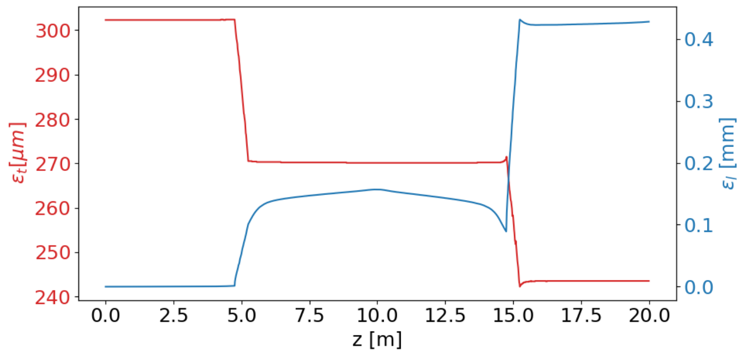


Optimizing coils radius using Extremum seeking algorithm









Developed matching routine produces acceptable results starting from (random) initial guess

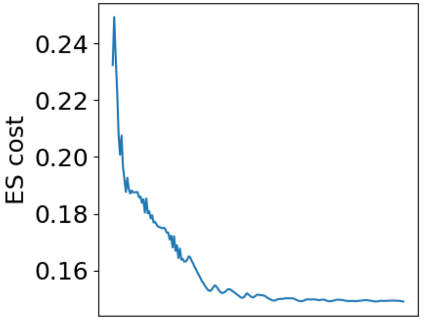
International
UON Collider
Collaboration

First results: simplified lattice

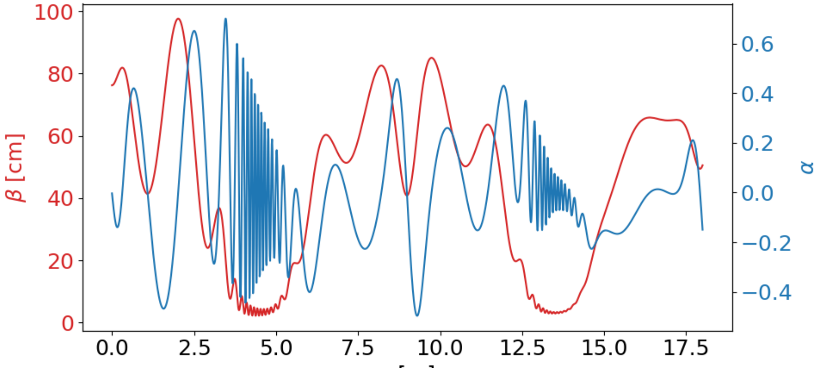
Note: simplified lattice, no re-acceleration

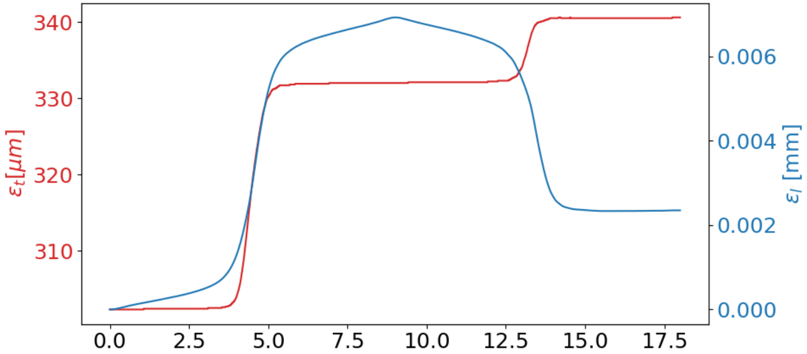
- each cell containing of 3 coils x 4 sheets, absorber density, initial momentum and beta- function

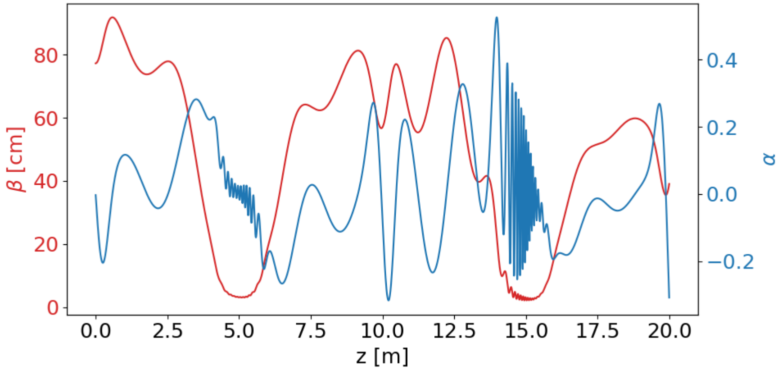
➡ **Extremum seeking algorithm:** much faster for a larger parameter spaces, easily extendable

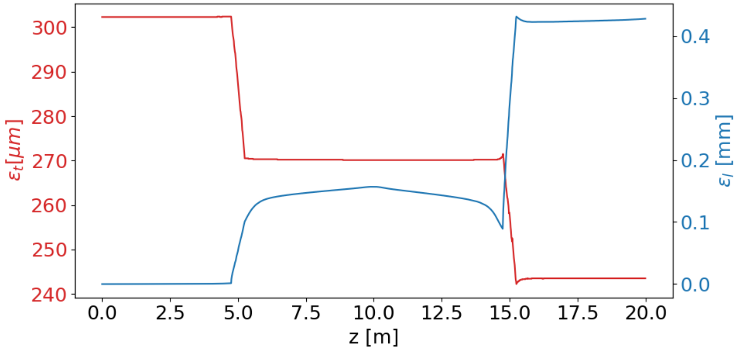


Optimizing coils radius using Extremum seeking algorithm









Developed matching routine produces acceptable results starting from (random) initial guess

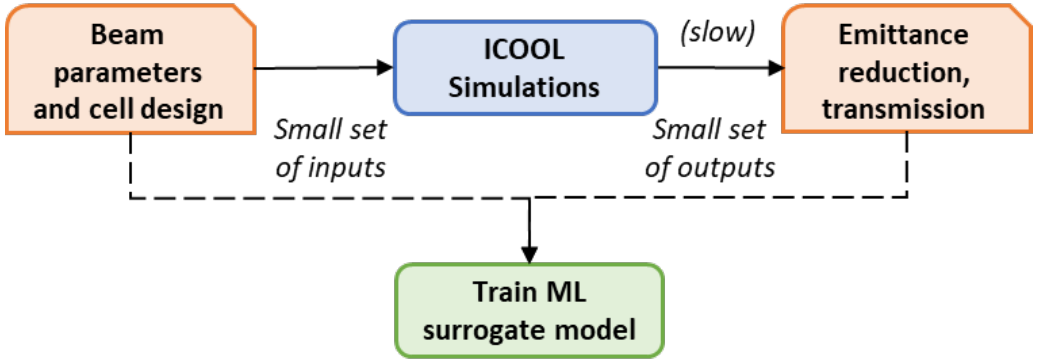
◉ **Tracking of thousands of particles at every optimization step**

◉ **Increasing the complexity of the lattice —> more optimisation steps**

International
UON Collider
Collaboration

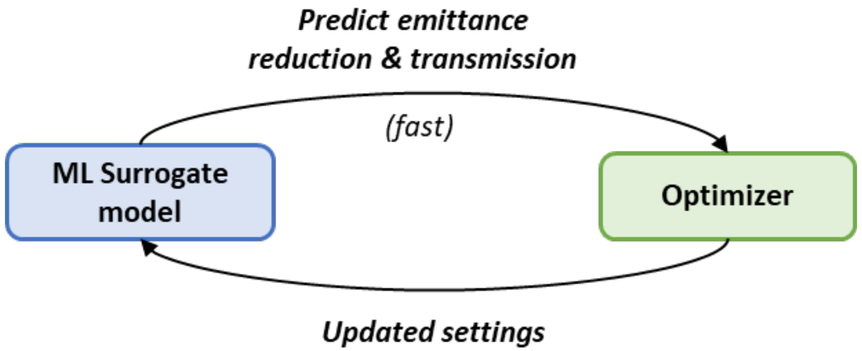
Speeding up simulations with Supervised Learning

➡ Making use of simulations done during optimisation

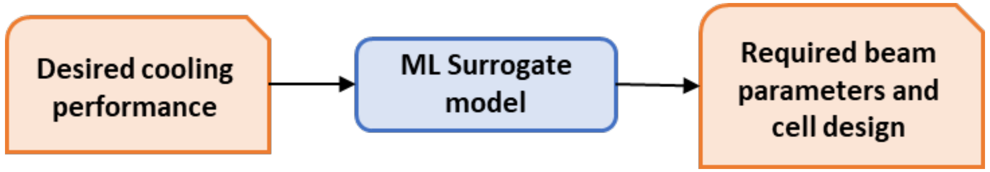


Earlier example: A. Edelen et al. „Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems" , Phys. Rev. Accel. Beams 23, 044601, 2020

1. Speeding up optimization:



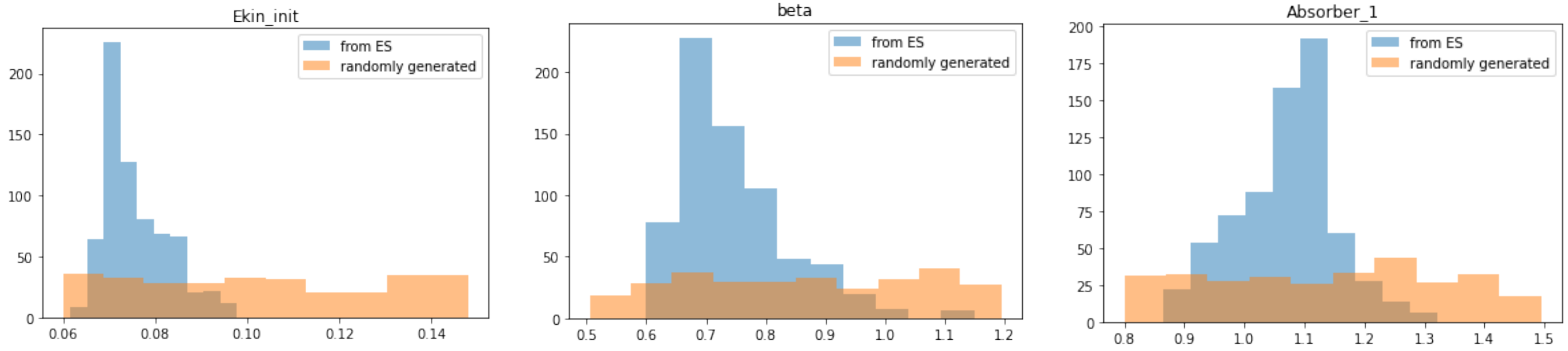
2. "Inverse" design:



✓ First results demonstrating orders of magnitude optimization speed up
✓ Accurate prediction of initial parameters to achieve desired cooling performance

Parameter scans vs. Storing data from optimization

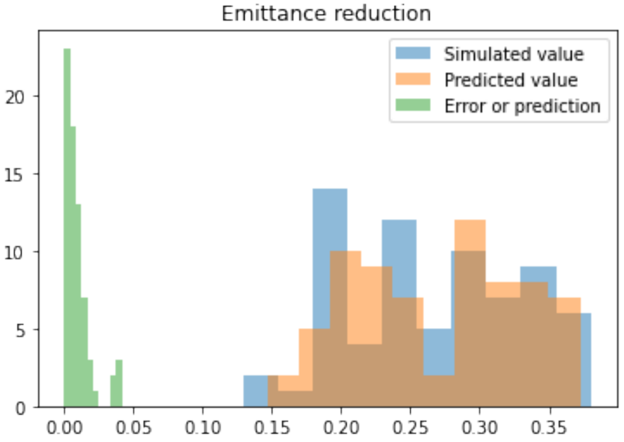
➡ Making use of simulations done during optimisation

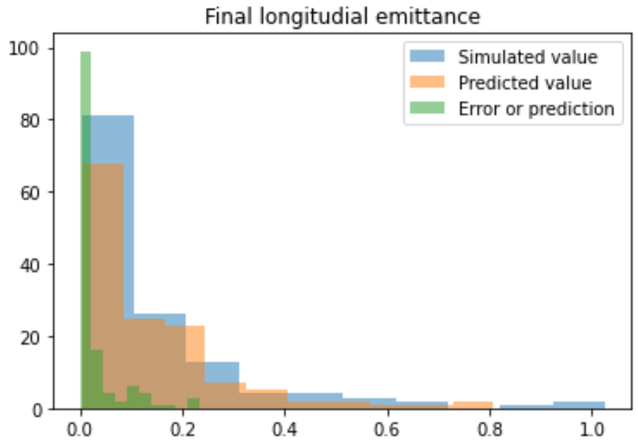


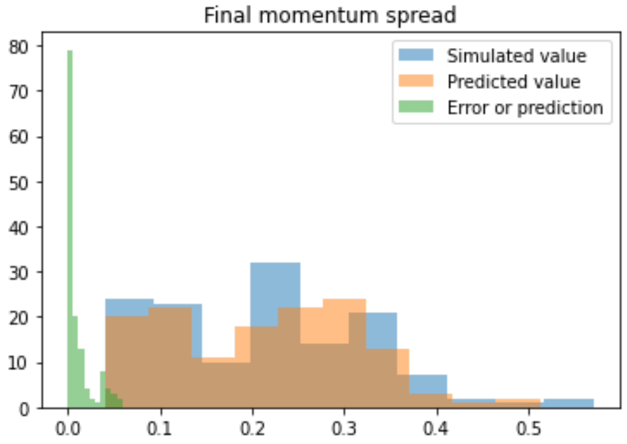
- ➡ Easier to find the boundaries
- ➡ Warm start for more complex problems

Combining Surrogate Models and Extremum Seeking

Using Random Forest (decision trees-based) algorithm

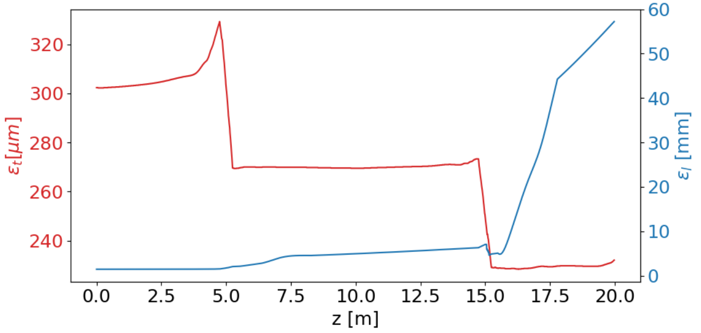






Model performance: train/test - 0.99/0.98

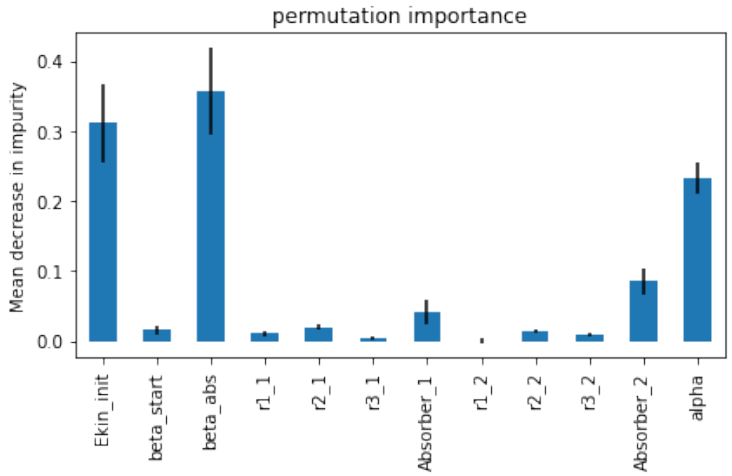
➡ **Applied to more complex model** which includes RF-optimization and energy spread/ longitudinal emittance control



- ✓ **Compute optimization function from ML-model prediction**
- ✓ **Optimization in a few minutes instead of ~1.5 hours for 200 steps using simulations**

Impact on cooling performance

Using Random Forest (decision trees-based) algorithm —> Feature Importance Analysis



... obvious to an (experienced) physicist

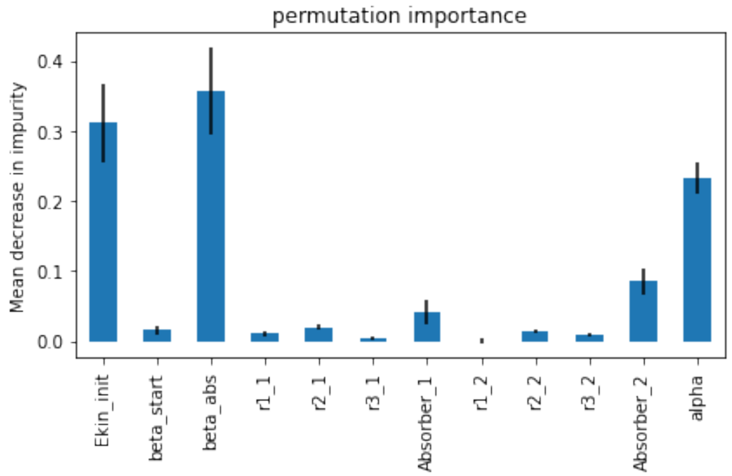
—> Big achievement for a decision tree

✓ “what is this model actually learning?”

Impact on cooling performance

International
UON Collider
Collaboration

Using Random Forest (decision trees-based) algorithm —> Feature Importance Analysis

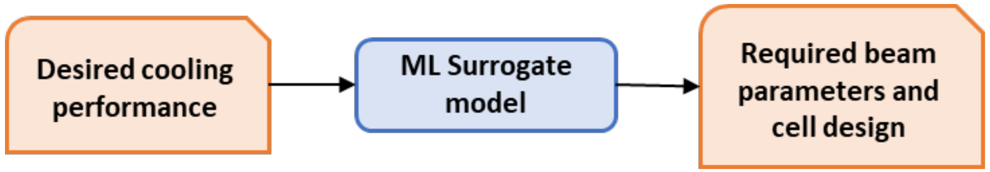


… obvious to an (experienced) physicist
—> Big achievement for a decision tree
✓ "what is this model actually learning?"

Overcoming ML "complexity":
- Start optimization with very simple models
➡ Easy to control free parameters and verify results
- Build more complex non-analytical models

Inverted models: warm start or final solution?

- **Estimate the initial parameters to achieve a desired cooling performance**



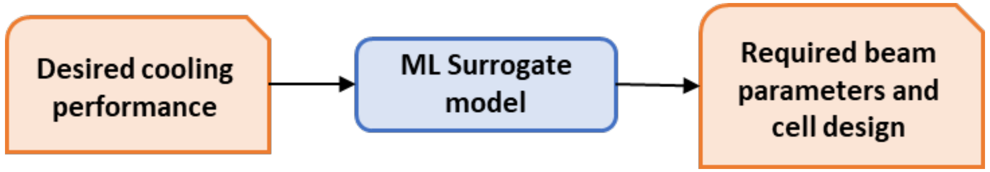
Input: Emittance reduction, momentum reduction, transmission
Output: required start energy, beta, absorber densities in **2 consecutive cells**

Example: aiming for $\Delta\epsilon$=50%, Δpz = 60%, Δ N=90%,

predicted values are: Ekin = 0.0714GeV, beta = 0.846, absorber densities = 1.3, 1.1

Inverted models: warm start or final solution?

- **Estimate the initial parameters to achieve a desired cooling performance**



<u>Input</u>: Emittance reduction, momentum reduction, transmission
<u>Output</u>: required start energy, beta, absorber densities in **2 consecutive cells**

Example: aiming for $\Delta\epsilon=50\%$, $\Delta pz = 60\%$, $\Delta N=90\%$,

predicted values are: Ekin = 0.0714GeV, beta = 0.846, absorber densities = 1.3, 1.1

Verification by running ICOOL with predicted parameters: $\Delta\epsilon=0.493\%$, $\Delta pz = 0.61\%$, $\Delta N=0.98\%$

Outlook and Summary

Summary: Where can we use ML in accelerators?

Detection of instrumentation failures

Beam control and lattice imperfection corrections

Optimization and operation automation

Virtual Diagnostics

- Defining a **narrow task** (optimization of specific parameters rather than the entire machine)
- **Performance measure** of selected model (beam size, pulse energy, …)
- e.g. when no analytical solution is available, rapidly changing systems, no direct measurements are possible.

Important to identify where ML can surpass traditional methods

How much effort is needed to implement a ML solution? Is appropriate infrastructure for data acquisition available? Enough resources to perform the training?

Achieved Results

✓ **ML-based toolbox for optics control:**

- Detection of instrumentation faults → no manual cleaning and repeated optics analysis
- Estimation of individual magnet errors → Better knowledge and control of individual optics errors
- Denoising of optics measurements → Increasing the quality of the measurements
- Reconstruction of optics observables → Additional observables without dedicated measurements

Achieved Results

✓ **ML-based toolbox for optics control:**

- Detection of instrumentation faults → no manual cleaning and repeated optics analysis
- Estimation of individual magnet errors → Better knowledge and control of individual optics errors
- Denoising of optics measurements → Increasing the quality of the measurements
- Reconstruction of optics observables → Additional observables without dedicated measurements

Outlook

✓ **Paving the way for new studies currently being in progress:**

- Optics corrections for High Luminosity – LHC upgrade:
    - local correction
    - exploring Reinforcement Learning for determining correctors settings.
- Exploring more complex optics error sources: coupling corrections
- Optimizing the design of future colliders.

Further References

- **Machine learning for beam dynamics studies at the CERN Large Hadron Collider**
  https://doi.org/10.1016/j.nima.2020.164652
- **Opportunities in Machine Learning for Particle Accelerators**
  https://arxiv.org/abs/1811.03172
- **Optimization and Machine Learning for Accelerators (USPAS course)**
  https://slaclab.github.io/USPAS_ML/

Cat!

Thank you for your attention!

ML in accelerators: summary

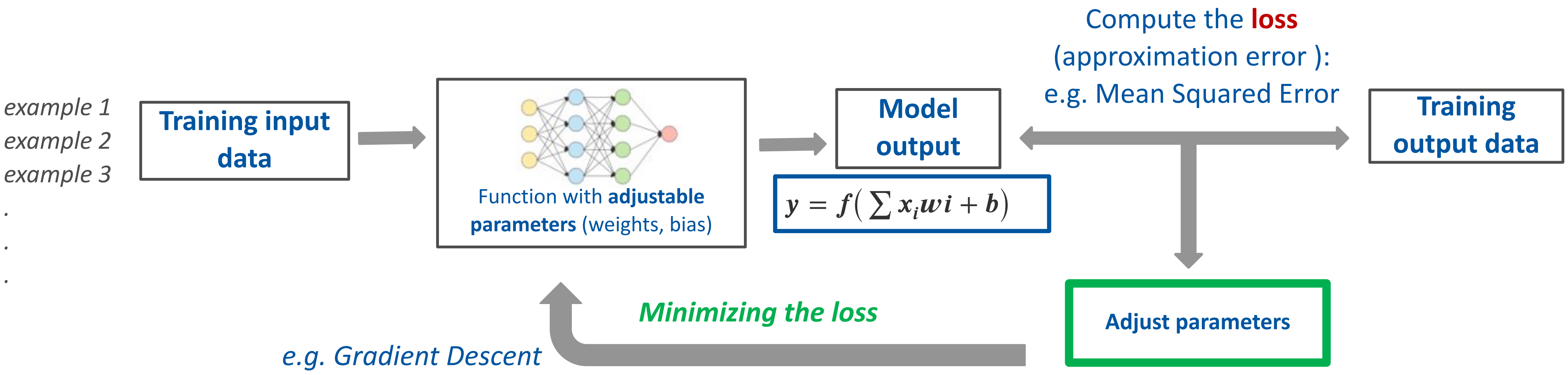
Accelerator Problem	ML methods	Benefits	To be considered
• Automation of particular components	Supervised techniques for classification: Decision Trees, SVR, Logistic Regression, NN	Saving operation time, reducing human intervention, preventing subjective decisions	Dedicated machine time usually required to collect training data and to fine tune developed methods.
• Online optimization of several targets which are coupled • Unexpected drifts, continuous settings readjustment needed to maintain beam quality	Reinforcement Learning, Bayesian optimization, Gaussian Process, Adaptive Feedback	Simultaneous optimization targeting several beam properties, automatically finding trade-off between optimization targets, allows faster tuning offering more user time.	Ensuring that all important properties are included as optimization targets.
• Detection of anomalies	Unsupervised methods: clustering, ensembles of decision trees (e.g. Isolation Forest), supervised classification, Recurrent NN for time-series data.	Preventing faults before they appear, no need to define rules/ thresholds, no training is needed and can be directly applied on received data	In unsupervised methods, usually no “ground truth” is available → methods can be verified on simulations.

ML in accelerators: summary

Accelerator Problem	ML methods	Benefits	To be considered
• Computationally heavy, slow simulations • Reconstruct unknown properties from measurements	Supervised Regression models, NN for non-linear problems	Learning underlying physics directly from the data, faster execution	100% realistic simulations are not possible → the model performance will be as good as your data is.
• Reduction of parameter space e.g. for optimization	Clustering, Feature Importance Analysis using Decision trees	Speed up of available methods, simpler defined problems, easier to interpret	Parameter selection and combination (feature engineering) can have significant impact on ML methods performance
• Missing or too noisy data	Autoencoder NN	Robust models, data quality	Significant information should not be removed from the signal.

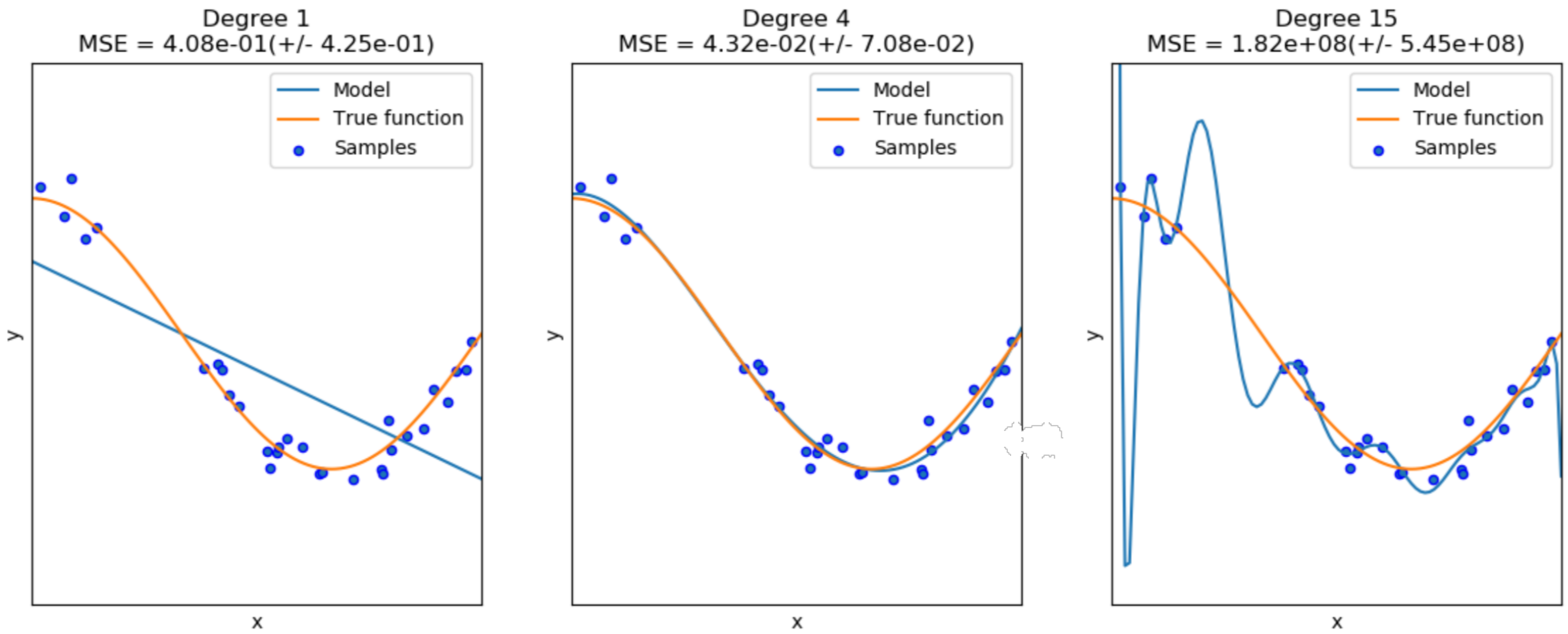
CERN

Supervised Learning



➢ Generalized model **explaining relationship between input and output** variables in **all training samples**.

Training and generalization: no perfect model needed!



Simple models underfit

- Derivate from data (high bias)
- Do not correspond to data structure (low variance)

We don‘t want „look up tables“ ↔

↔ We don‘t want unreliable prediction

→ Bias-Variance tradeoff

Complex models overfit

- Very low systematical deviation (low bias)
- Very sensitive to data (high variance)

Regression Models

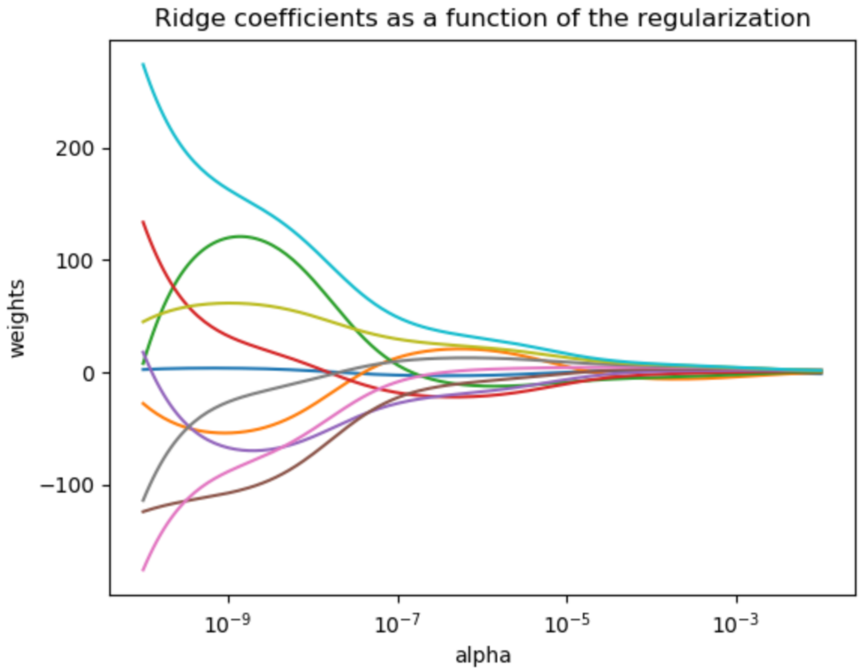
- Linear model for *input X / output Y pairs, i* – number of pairs (training samples): $f(X, w) = w^T X$
- Squared Loss function for model optimization: $L(w) = \frac{1}{2} \sum_i \left(Y_i - f(X_i; w)\right)^2$
- Find new weights minimizing the Loss function: $w^* = \mathrm{argmin}_w L(w)$

→ **Update weights for each incoming input/output pair**

Too much "flexibility" in weights update can lead to *overfitting*

→ **Regularization** places constraints on the model parameters (weights)

- Trading some bias to reduce model variance.
- Using L2-norm: $\Omega(w) = \sum_i w_i^2$, adding the constraint $\alpha\Omega(w)$ to the weights update rule
- The larger the value of α, the stronger the shrinkage and thus the coefficients become more robust.



Relevant ML concepts and definitions

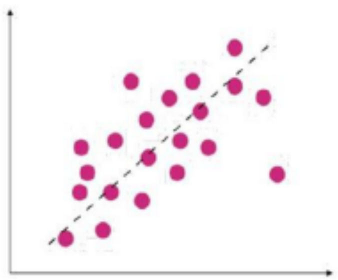
Supervised Learning

- **Input/output pairs** available
- Learn a mapping function, **generalizing for all provided data**
- Predict from **unseen data**

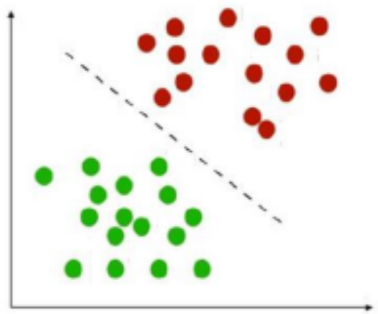
Unsupervised Learning

- **Only input** data is given
- Discover structures and patterns

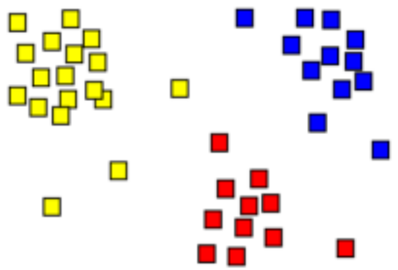
Regression



Classification

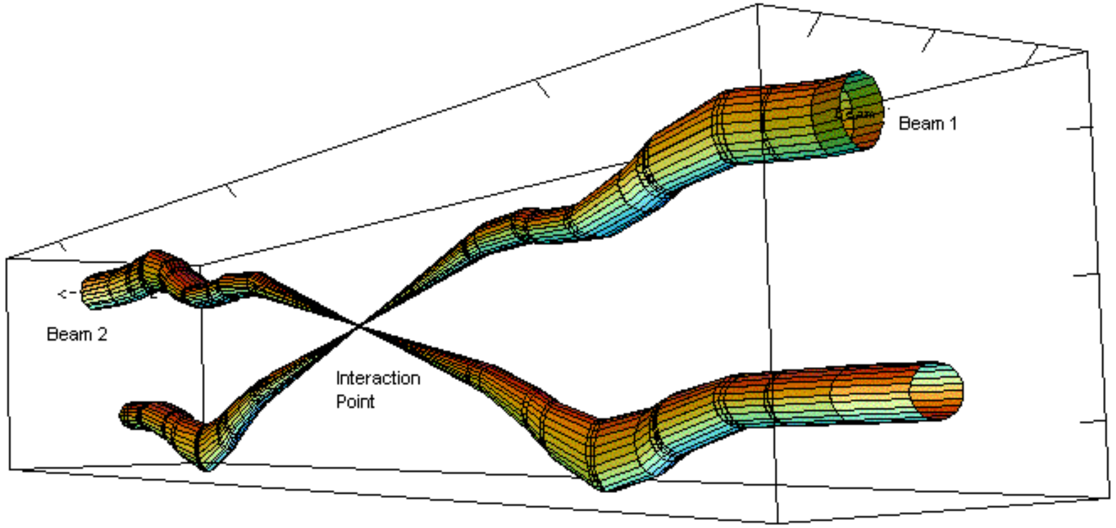


Clustering

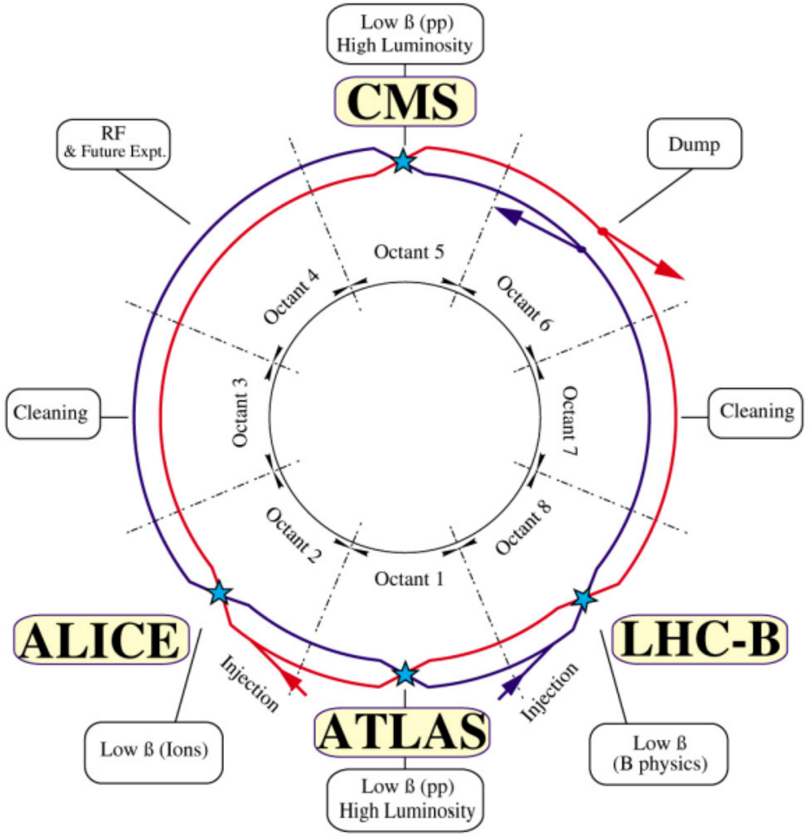


What is "Learning"?

Beam optics control at the LHC



Relative beam sizes around IP1 (Atlas) in collision



Large Hadron Collider:

- 9300 magnets for bending and focusing the beam.
- Main experiments: ALICE, ATLAS, CMS, LHCb
- Collision rate: sufficient and balanced between experiments → **Luminosity**

➢ How to increase chances of collisions?

➢ How to ensure machine protection?

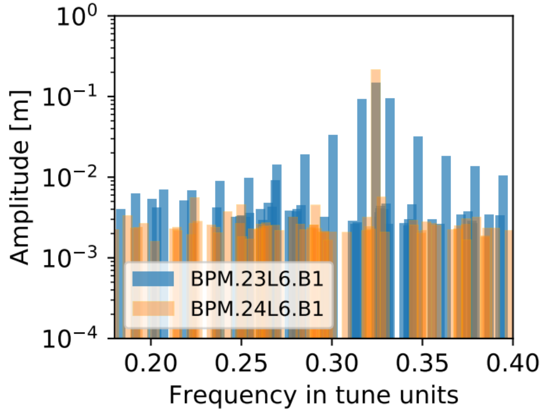
→ **Beam Optics control**

Why and how is the beam optics controlled in the LHC?

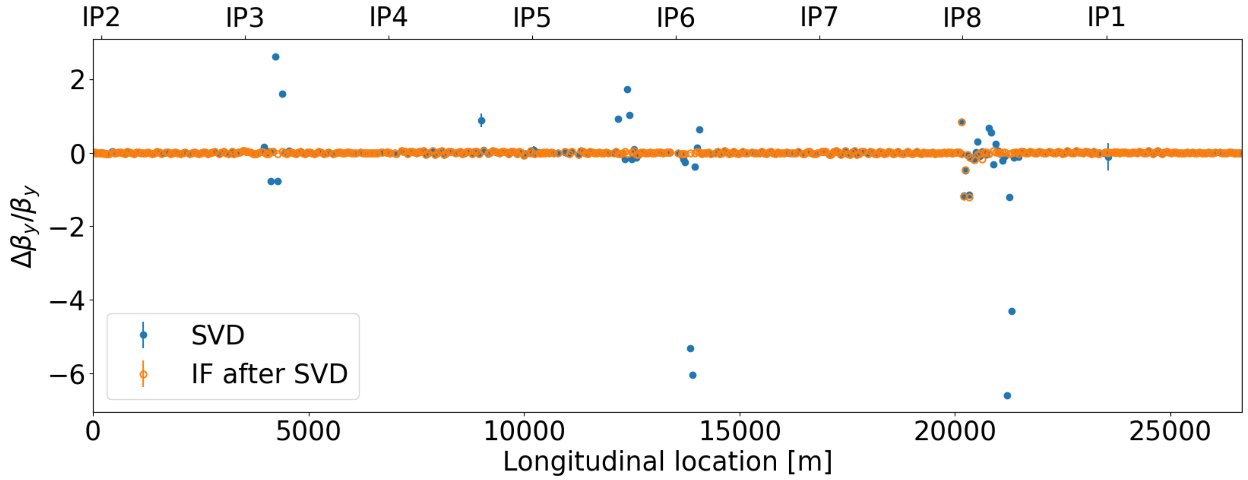
IF in the LHC operation: detecting unknown failures

- Some artifacts in the signal are known to be related to BPM failures (manual cleaning would time consuming, but potentially possible).
- **How to deal with unknown failure modes?**

Several BPMs with unusual pattern in the spectra indicating a new failure mode



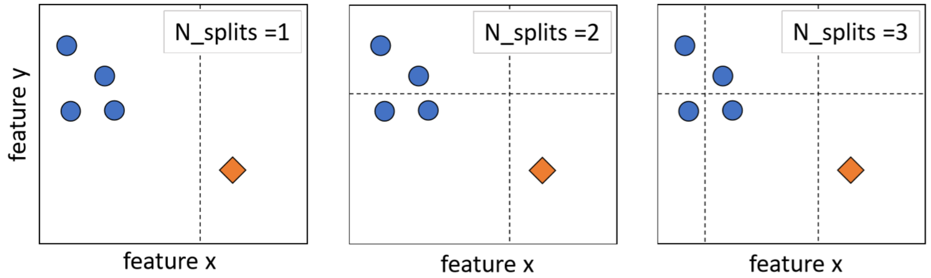
First observed in: "Analysis of tune modulations in the LHC", D.W. Wolf
Related to BPM failure: L. Malina, "Noise and stabilities", https://indico.cern.ch/event/859128/



Since IF is based on the structures in given data
➢ ***Ability to identify previously unknown failures***

Isolation Forest Algorithm

- Forest consists of several **decision trees**
- **Random splits aiming to “isolate” each point**
- The less splits are needed, the more “anomalous”
- **Contamination factor**: fraction of anomalies to be expected in the given data
  - → First obtained empirically from the past measurements
  - → Refined on **simulations introducing expected BPM faults**.



Conceptual illustration of Isolation Forest algorithm

- **Input data: combination of several signal properties** obtained from harmonic analysis of BPM turn-by-turn measurements
  - → No additional data handling needed.