Machine Learning Techniques for Optics Measurements and Corrections

Elena Fol

SLAC
28.10.2020

Outline

- **Introduction**
- **Detection of faulty Beam Position Monitors**

  - Motivation
  - Unsupervised Learning and Isolation Forest
  - Simulation studies and experimental results

- **Estimation of magnetic errors from optics measurements**

  - General concept
  - Results on simulations
  - Results on experimental data

- **Denoising and reconstruction of optics functions**

  - Autoencoder
  - Linear models
  - Results

- **Conclusions**

I. Introduction

About myself

- PhD Student at CERN working on Machine Learning techniques for beam optics studies at the LHC.

- B.Sc. in Business Informatics
- M.Sc. in Computer Science,
  specialization **Interactive Intelligent Systems**
  (University of Applied Sciences Karlsruhe, Germany)

- Technical Student at CERN,
  LHC Optics measurements and corrections team:
  - Responsible for Java GUIs used in the LHC control
  - Idea of solving **Beam Optics** related problems using Machine Learning
- Master's Thesis:
  "Evaluation of Machine Learning methods for optics measurements and corrections at the LHC"

Computer Science → PhD project: Application of Machine Learning to Accelerator Optimization with the focus on beam optics. ← ***Accelerator Physics***

Motivation

Accelerators

Limitations of traditional optimization and modeling tools?

← ML is a powerful tool for prediction and data analysis

Which limitations can be solved by ML with reasonable effort?

- How to deal with previously unobservable behavior?
- Required computational resources for large amount of optimization targets
- Objective functions, specific rules and thresholds have to be known

Machine Learning methods can learn an arbitrary model from given examples without requiring explicit rules

Machine Learning concepts

*"... computer programs and algorithms that automatically **improve with experience** by **learning from examples** with respect to some class of task and performance measure, **without being explicitly programmed**." **

Supervised Learning	Unsupervised Learning	Reinforcement Learning
• Input/output pairs available • Make prediction for unknown input based on experience from given examples	• Only input data is given • Learn structures and patterns	• No training data • Interact with an environment • Trying to learn optimal sequences of decisions
Object detection in computer vision, speech recognition, predictive control	Anomaly detection, pattern recognition, clustering, dimensionality reduction	Robotics, industrial automation, dialog systems

* Thomas M. Mitchell. Machine Learning. McGraw-Hill, Inc., New York, 1997.

Machine Learning concepts

*"… computer programs and algorithms that automatically **improve with experience** by **learning from examples** with respect to some class of task and performance measure, **without being explicitly programmed**." **

Applied in optics measurements and corrections at the LHC

Supervised Learning

- Input/output pairs available
- Make prediction for unknown input based on experience from given examples

Object detection in computer vision, speech recognition, predictive control

Unsupervised Learning

- Only input data is given
- Learn structures and patterns

Anomaly detection, pattern recognition, clustering, dimensionality reduction

Reinforcement Learning

- No training data
- Interact with an environment
- Trying to learn optimal sequences of decisions

Robotics, industrial automation, dialog systems

* Thomas M. Mitchell. Machine Learning. McGraw-Hill, Inc., New York, 1997.

Applying Machine Learning to Beam Optics

PhD project: Application of Machine Learning to Accelerator Optimization with the focus on beam optics.

- **Why and how is the beam optics controlled in the LHC?**
- Where are the limitations of traditional techniques?
- Which ML concepts and algorithms can be applied?
- Achieved results?

Applying Machine Learning to Beam Optics

PhD project: Application of Machine Learning to Accelerator Optimization with the focus on beam optics.

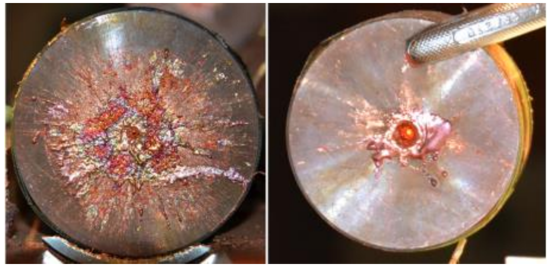
- **Why and how is the beam optics controlled in the LHC?**
- Where are the limitations of traditional techniques?
- Which ML concepts and algorithms can be applied?
- Achieved results?

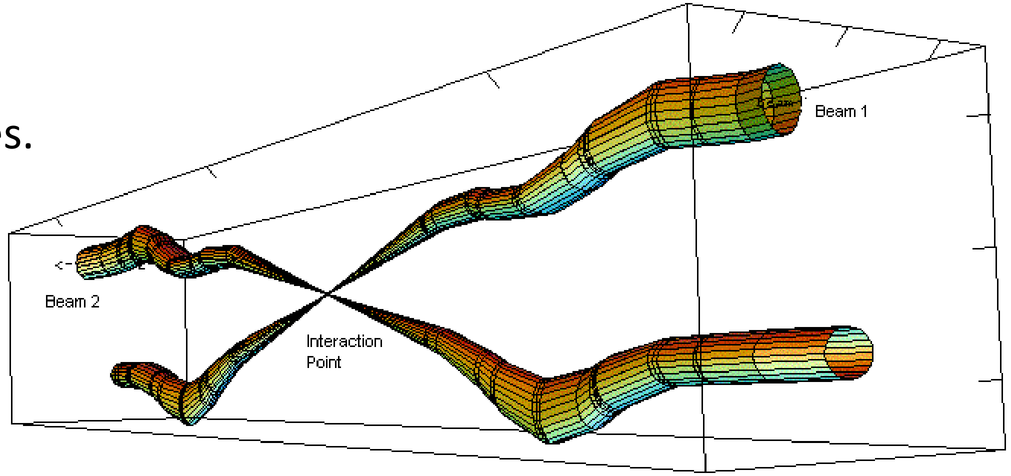
Beam optics control:

- Magnetic errors and misalignments change **beam size** - optics
- Adjust **magnetic strengths** – optics corrections.

Importance of beam optics control:

- **Collision rate** depends on the beam size
- Beam optics imperfections can lead to **machine safety** issues.





Relative beam sizes around IP1 (Atlas) in collision

Where are the limitations of traditional techniques?

- **Instrumentation faults** lead to unreliable optics measurements
  → How to **detect faulty Beam Position Monitors** and discard them from analysis before they cause erroneous computation of optics functions?

- Optics corrections algorithms aim to **compensate the measured optics deviations** from design
  → What are the actual currently present **magnetic errors**?

- Advanced techniques for computation of optics functions require **additional measurements and operational time**
  → How to obtain advanced analysis **from available measurements**?

- **Noise** in the measured optics functions
  → How to **reduce the noise** without removing valuable information?

- **Missing data points** due to the presence of faulty BPMs
  → How to **reconstruct** the missing data?

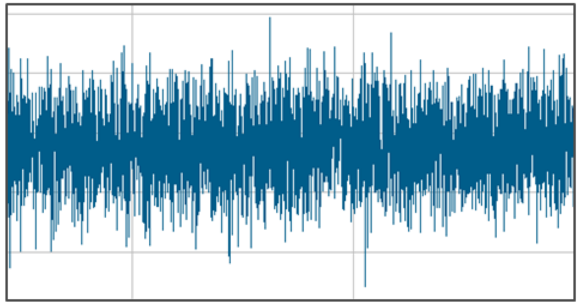
I. Detection of faulty Beam Position Monitors

Detection of faulty Beam Position Monitors

Optics measurements in the LHC

BPMs record the turn-by-turn data measuring the oscillations of the excited beam

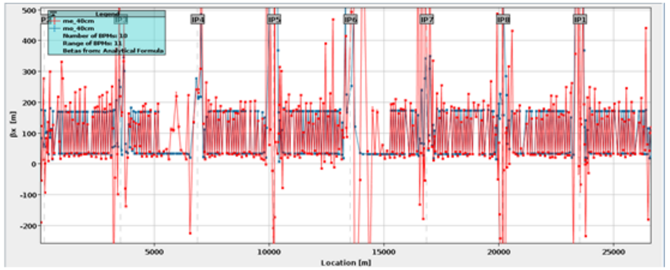
BPM turn-by-turn readings



Calculate optics functions (beta-beating, dispersion, etc.) based on **harmonic analysis of BPMs signal**

+ data cleaning

Computed optics

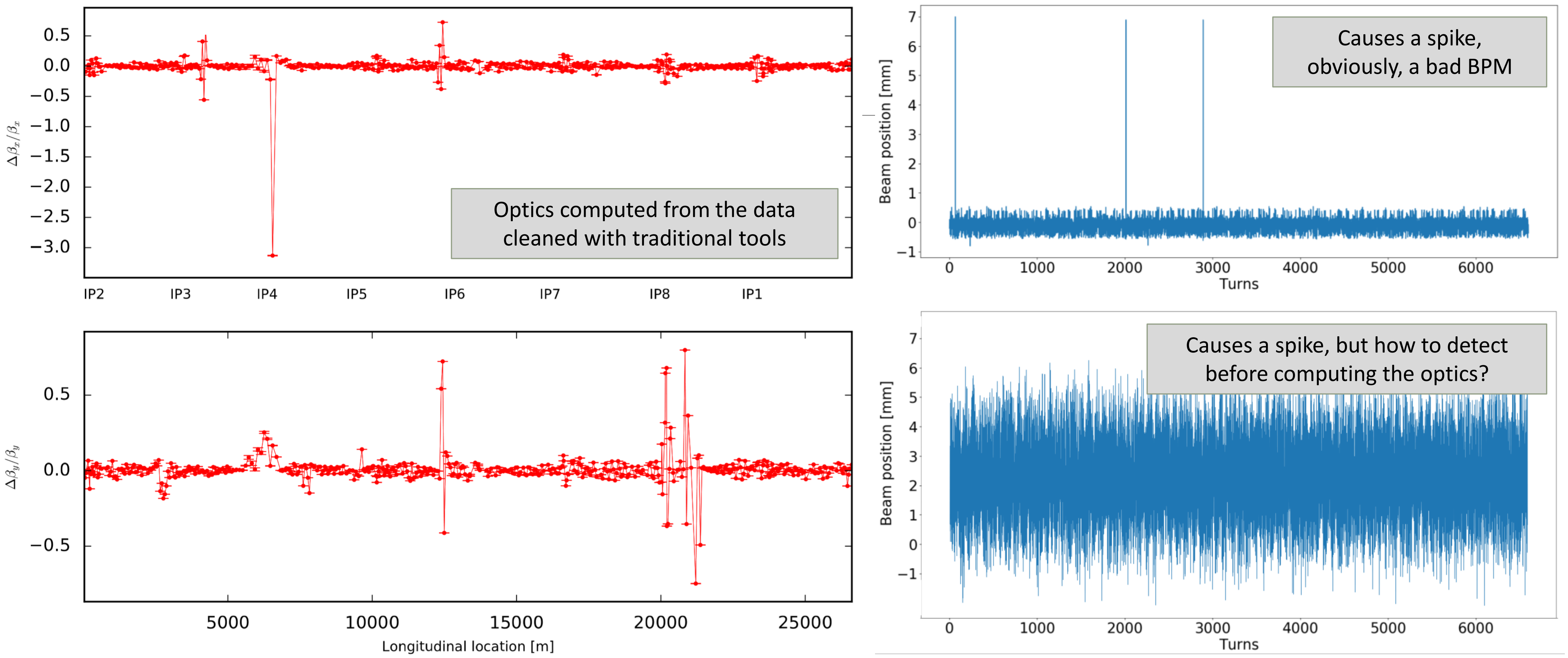


- Previously available techniques:
  BPM data cleaning based on Singular Value Decomposition (SVD) + signal cuts with predefined thresholds.

- Unphysical values still can be observed after cleaning with available tools: presence of faulty BPMs
  - Define outliers, manual cleaning of BPM signal, re-analyse the optics.
  - Important to detect as many faulty BPMs as possible **before** computing the optics

→ **ML as an alternative solution to improve the analysis.**

Outliers in the optics computed from harmonic analysis of BPM signal

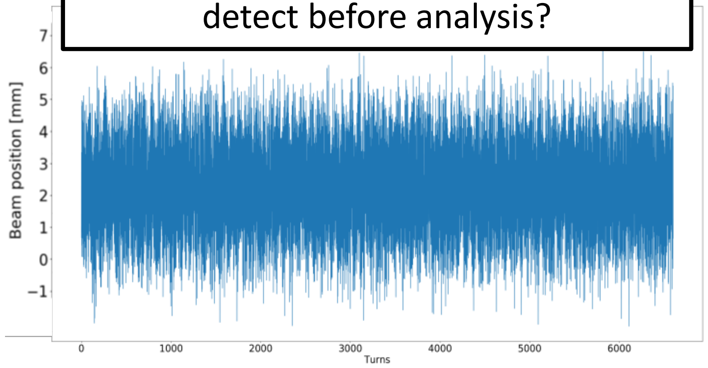


Detection of faulty BPMs using unsupervised learning

General Idea: Since actual malfunctioning BPMs are unknown, we consider the appearance of non-physical outliers in reconstructed optics as artifact of bad BPMs.

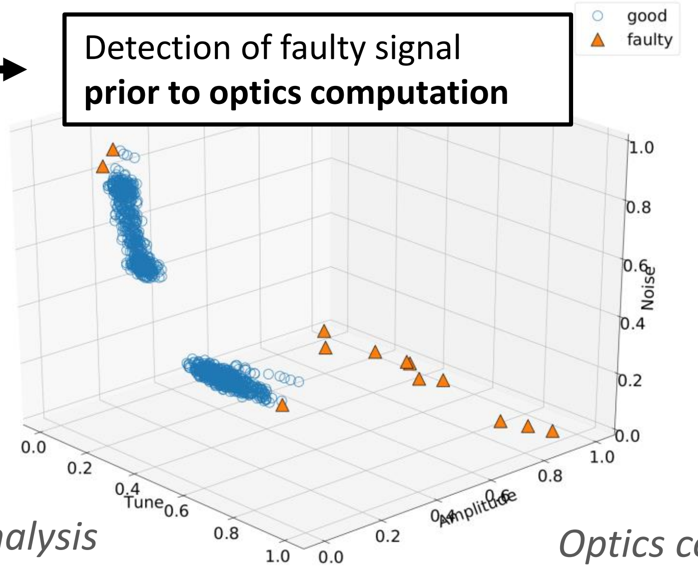
- We do not want to replicate current results, so no training data set (input-output pairs) is available
  → **Unsupervised Learning**
- Assuming most of the BPMs measure correctly, the bad BPMs should appear as **anomaly**
  → **Anomaly detection techniques**
- Applied clustering algorithms: DBSCAN[1], Local Outlier Factor[2], anomaly detection using **Isolation Forest**[3] implemented with *Scikit-Learn.*

Causes unphysical values in reconstructed optics, but how to detect before analysis?



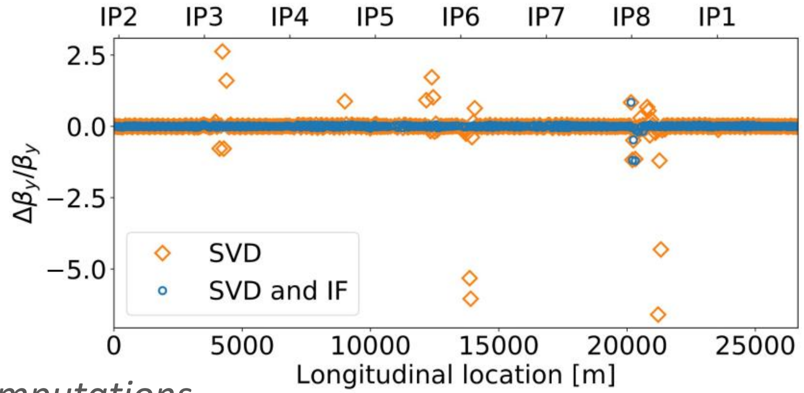
Harmonic analysis

→ Detection of faulty signal **prior to optics computation**



Optics computations

→ Avoid the appearance of erroneous optics computation



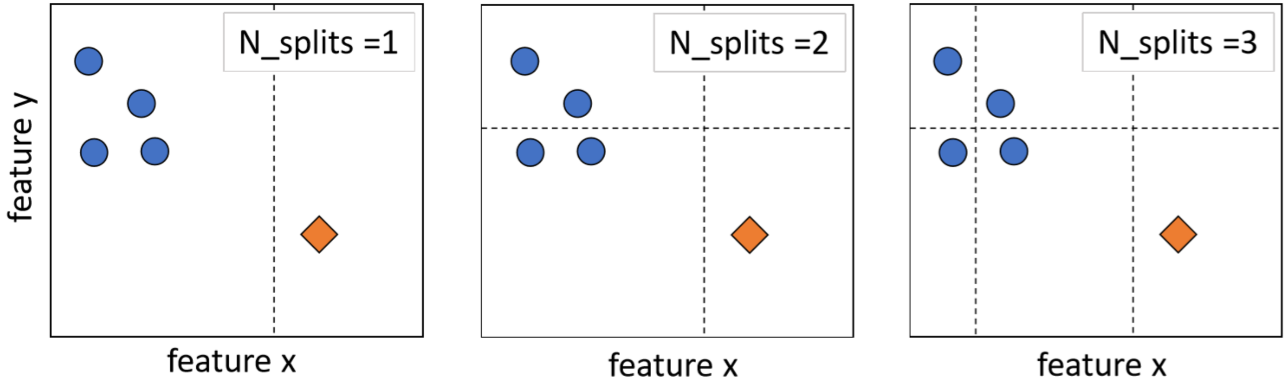
1. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" Ester, M., H. P. Kriegel, J. Sander

2. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May)., LOF: identifying density-based local outliers

3. Liu, Fei Tony, Ting, Kai Ming and Zhou, Zhi-Hua. "Isolation forest." Data Mining, 2008. ICDM'08.

Isolation Forest (IF)

- 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

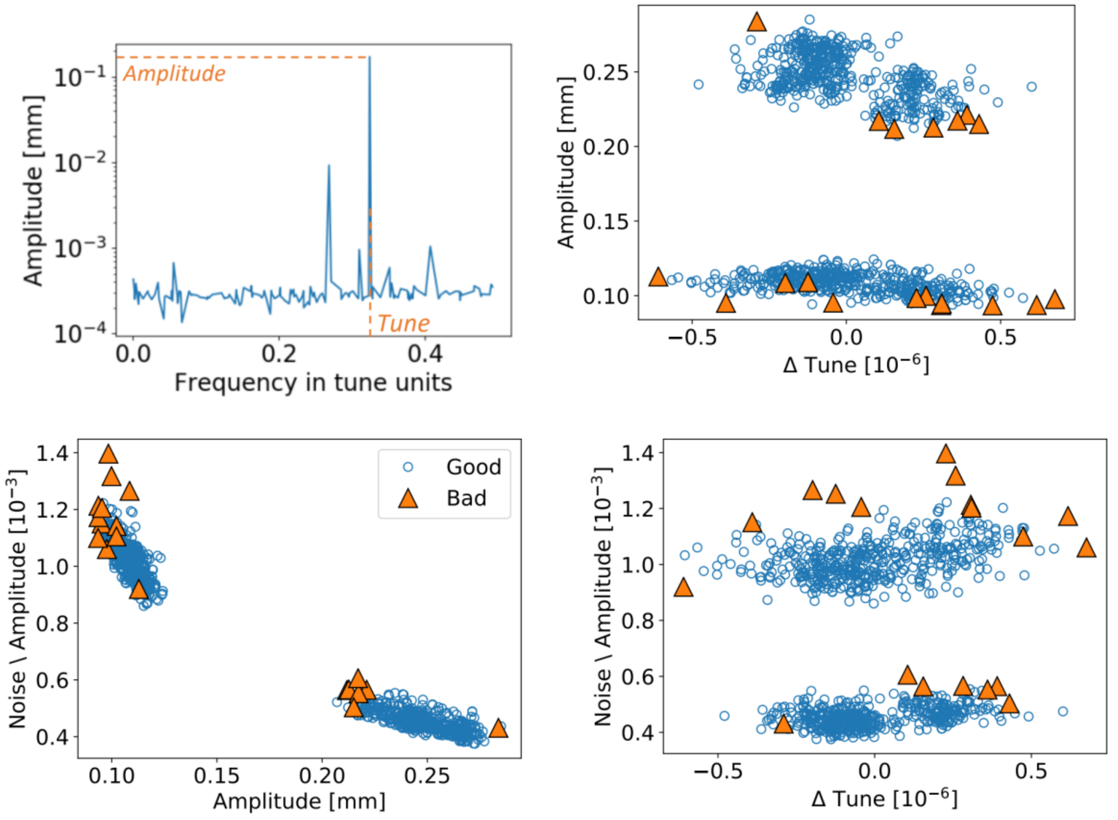
BPM signal properties as input features

Harmonic properties of BPM turn-by-turn signal

- Betatron tune (main frequency)
- Amplitude
- Noise to amplitude ratio

Contamination factor

- First obtained from measurement statistics
- Refined on simulations introducing expected BPM faults.



Input features and 2D-projection of anomaly detection in BPM data.

Unsupervised learning: how to verify the results?

No prior knowledge about which BPMs will produce faulty signal in acquired turn-by-turn data.

→ Simulate faults*: **bad BPMs are known**, cleaning results can be verified

**Simulated data is used only to verify the algorithm, there is no "training".*

Simulations setup:

- Around 5.5% per plane are faulty considering the statistics from the past measurements in 2018 (SVD detected bad BPMs + remaining outliers in the optics)
- Generate ideal turn-by-turn signal with Gaussian background noise 0.1mm
- Add signal perturbation **related to known faults** to 5.5% randomly chosen BPMs.

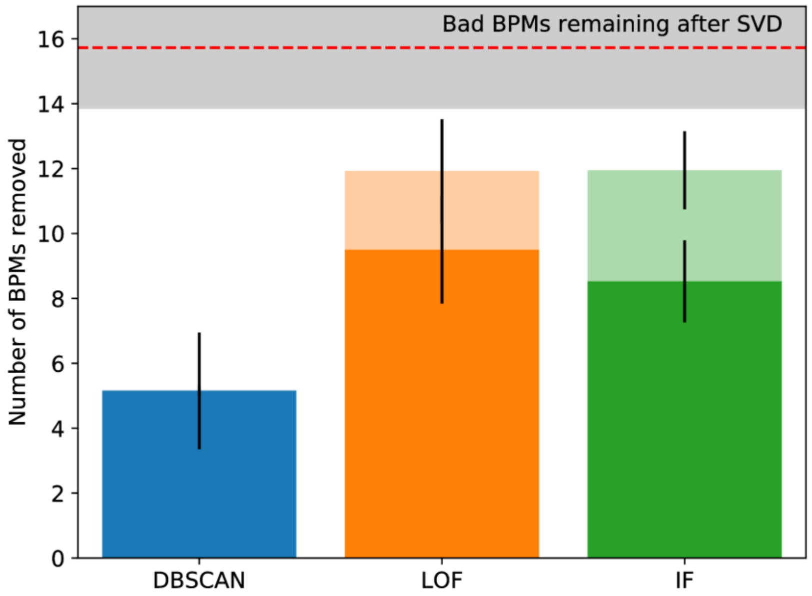
- Compare clustering algorithms
- Fine tuning of IF algorithm
- Verify results on a larger data set

Faulty BPMs detection: simulation study

- Comparing different suitable techniques:

The presence of a single faulty BPM has more significant negative impact on the optics computation than the absence of a good BPM

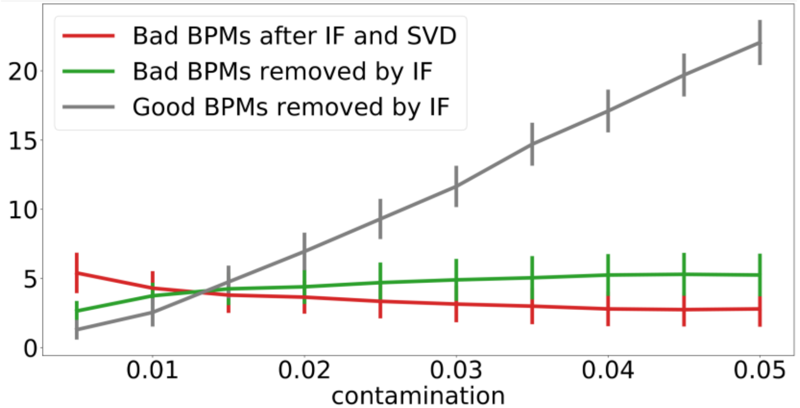
→ IF is preferred method for the LHC.

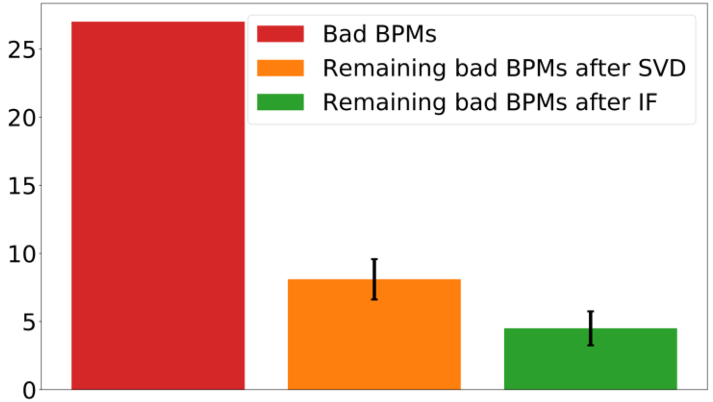


- Averaged results over 100 simulations

- Tuning of IF-algorithm **after finding optimal settings for SVD-cleaning**:

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

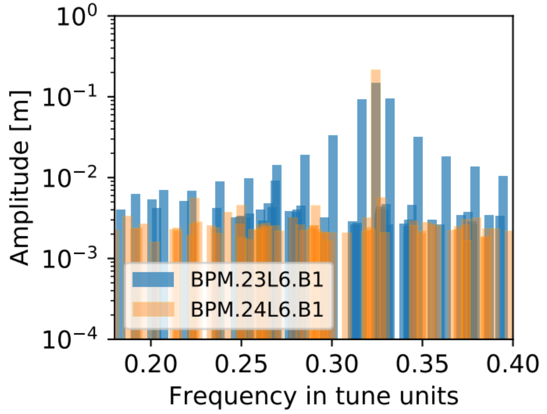




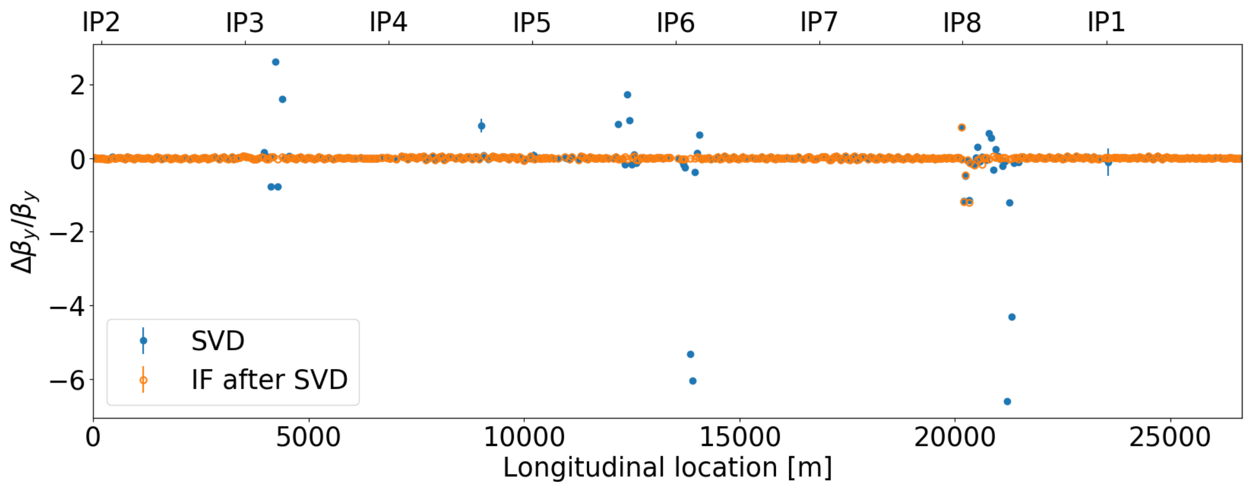
IF in the LHC operation: detecting unknown failures

- 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



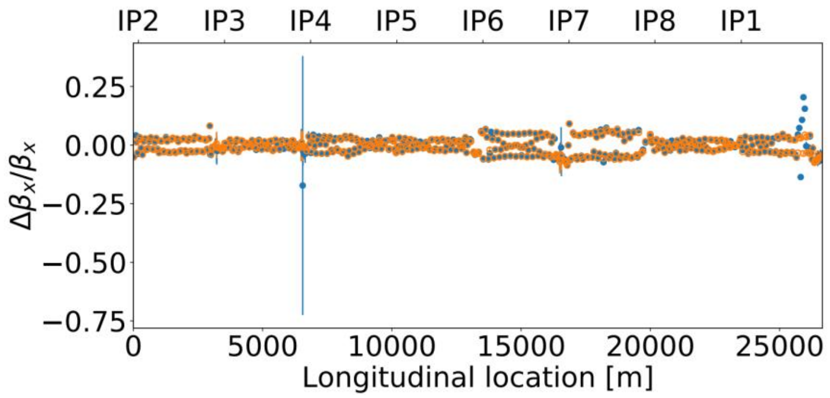
First observed in: "Analysis of tune modulations in the LHC", D.W. Wolf
Related to BPM failure: L. Malina, "Noise and stabilities", https://indico.cern.ch/event/859128/

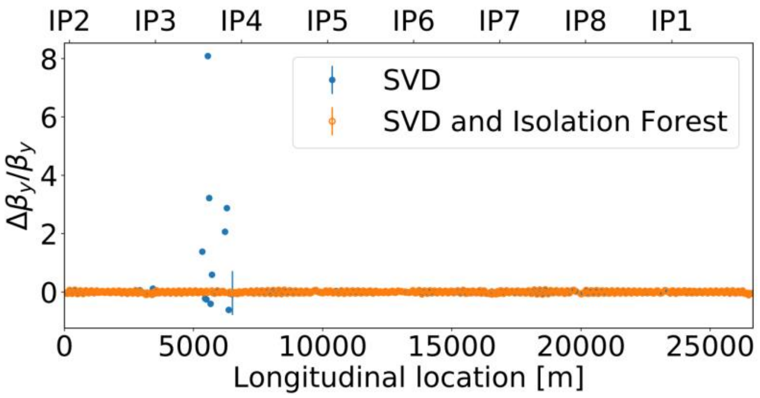


Since IF is based on the structures in given data
➢ ***Ability to identify previously unknown failures***

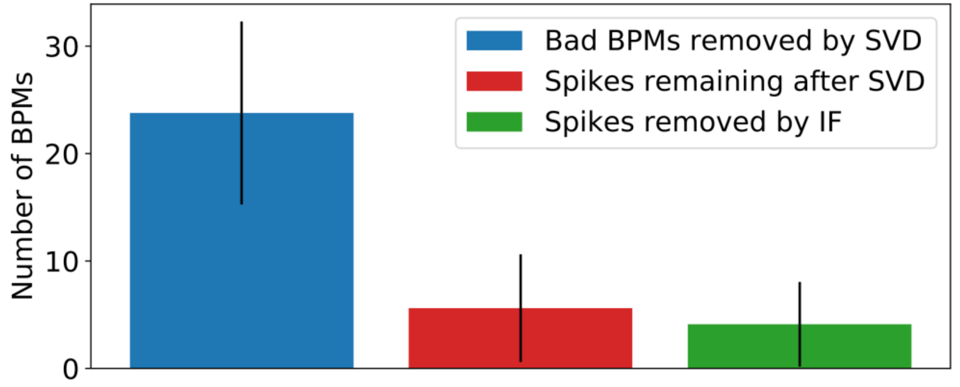
IF in the LHC operation: β-beating computed from cleaned BPM data

- Optics computation using the data cleaned with traditional techniques only vs. additionally applying IF





Reduction of non-physical outliers in beta-beating, summary of measurements in 2018.



✓ IF is **fully integrated** into optics measurements at LHC

✓ Successfully **used during beam commissioning and machine developments** in 2018 under different optics settings.

Detection of faulty beam position monitors using unsupervised learning
E. Fol, R. Tomás, J. Coello de Portugal, and G. Franchetti
Phys. Rev. Accel. Beams ***23****, 102805 (2020)*

II. Estimation of magnetic errors

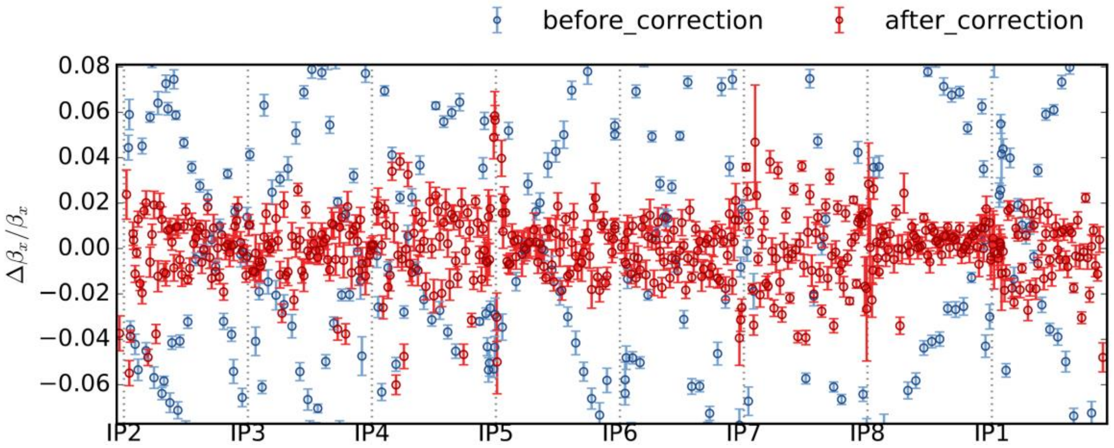
Optics corrections at the LHC

- Corrections aim to **minimize the difference between the measured and design optics** by changing the strength of corrector magnets – single quadrupoles and **quadrupoles powered in circuits**.

Optics corrections in the LHC are currently based on:

– **Local** corrections around Interaction Points (e.g. Segment-by-Segment method)

– **Global** corrections using a Response Matrix between available correctors and optics observables.



– For each beam **separately.**

Optics corrections at the LHC

- Corrections aim to **minimize the difference between the measured and design optics** by changing the strength of corrector magnets – single quadrupoles and **quadrupoles powered in circuits**.
  - ➢ What is the actual error of each individual magnet?

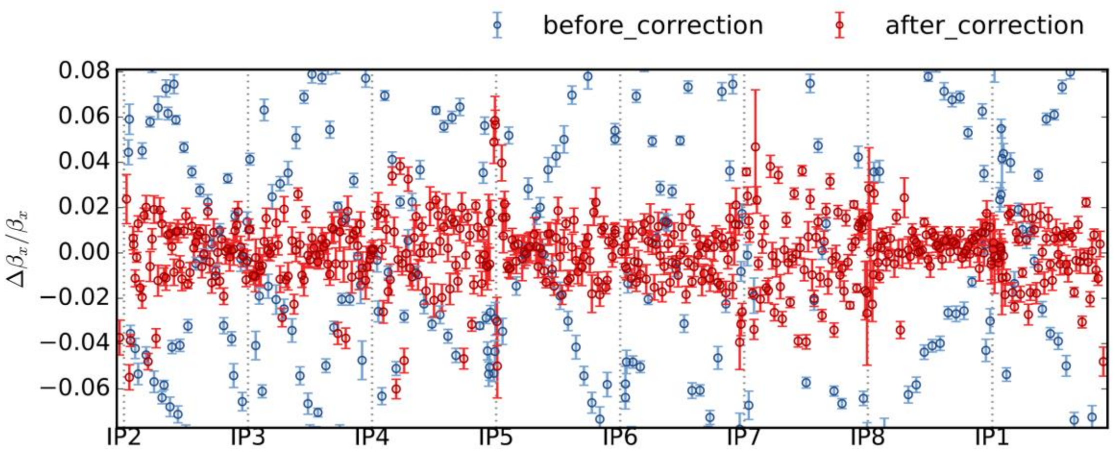
<u>Optics corrections in the LHC are currently based on</u>:
– **Local** corrections around Interaction Points
(e.g. Segment-by-Segment method)

– **Global** corrections using a Response Matrix between available correctors and optics observables.

- ➢ Appropriate weights of observables in the response matrix are adjusted manually**.**

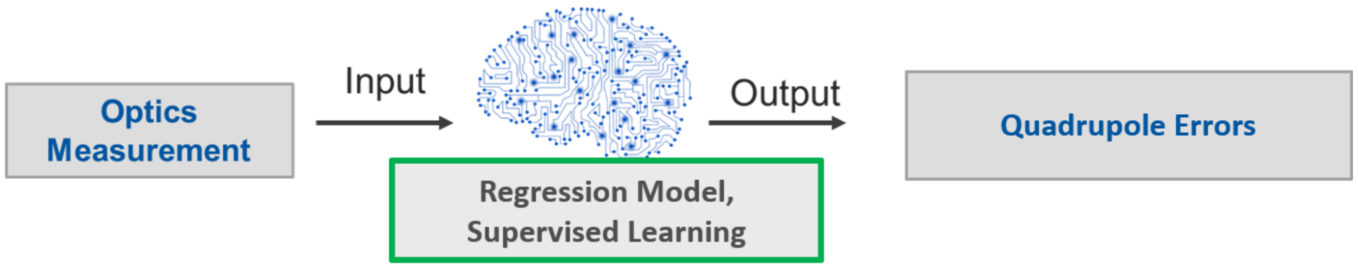
– For each beam **separately.**

- ➢ How to determine the whole set of errors for both beams simultaneously?



Supervised Learning & multivariate regression

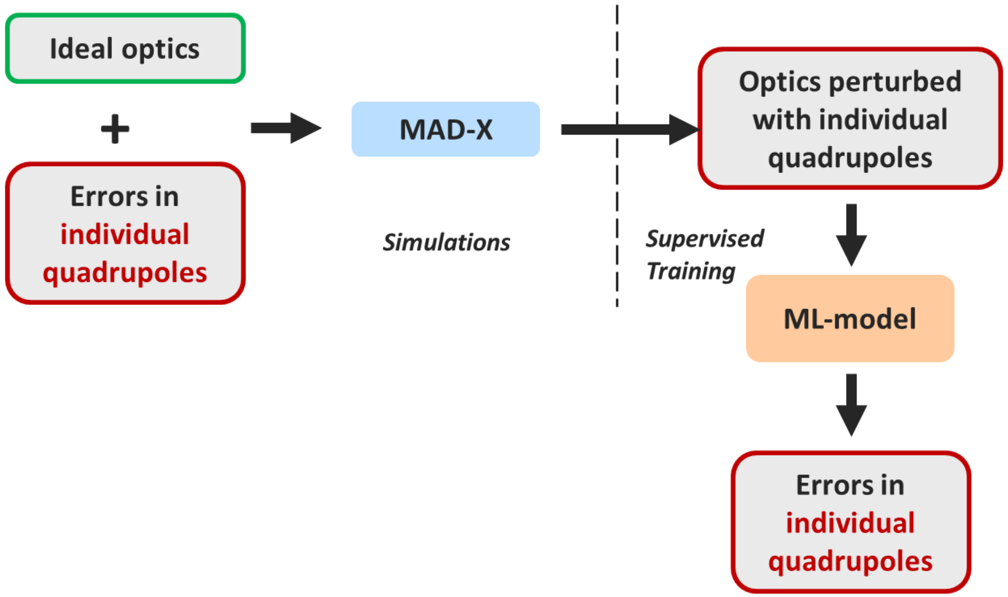
General concept



- Train **supervised regression model** to predict magnet errors from optics perturbations caused by these errors.

- Large dataset is needed in order to train a regression model: **simulations!**

- **Correlations** between magnetic errors and optics deviations from design can be **learned by ML-model**.



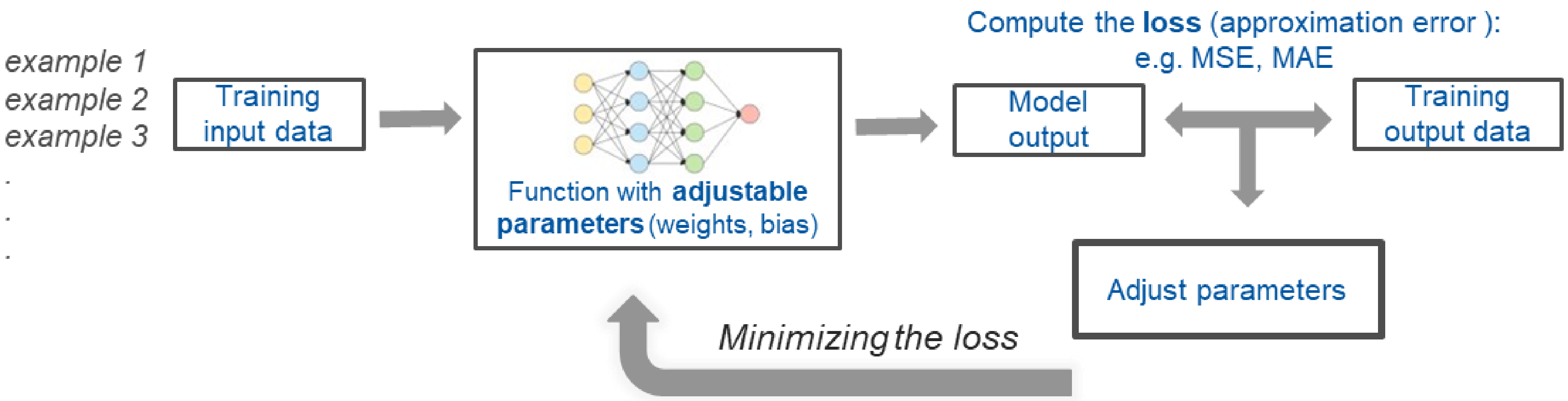
Supervised Learning

Supervised Learning

- Input/output pairs available
- Make prediction for unknown input based on experience from given examples

Predictive modeling, object recognition, medical diagnosis, fraud detection

- Fitting data with (complex) functions
- Mathematical models **learned from data** to describe **relationships** between variables in the system
- Learning = **estimate statistical model** from training data to make **predictions on new data**.



Linear Regression model as predictor

Linear model for *input X, output Y - pairs*, *i* – number of pairs (training samples), with *weights w:*

$$\boldsymbol{f(X, w) = w^T X}$$

Residual sum of squares as **loss function** for model optimization:

$$\boldsymbol{L(w) = \sum_i \left(Y_i - f(X_i; w)\right)^2}$$

Find **new weights** minimizing the Loss function:

$$\boldsymbol{w^* = \arg min_w L(w)}$$

Update weights for each incoming input/output pair

- Generalized model explaining relationship between input and output variables in **all training samples**.
- Test the model on unseen validation data.

→ How to improve the predictive power of the model?

Weights update regularization & bagging

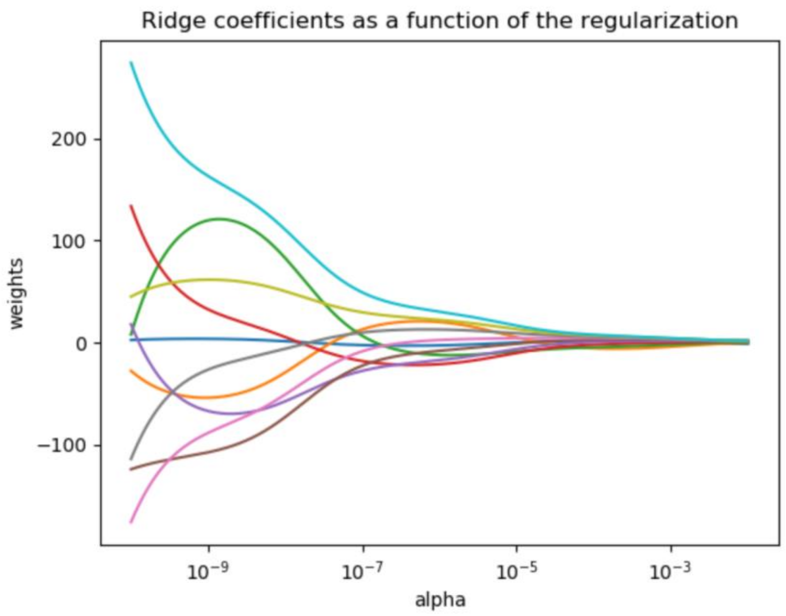
Too much "flexibility" in weights update can lead to *overfitting*

→ **Regularization** places constraints on the model parameters

- Trading some bias to reduce model variance

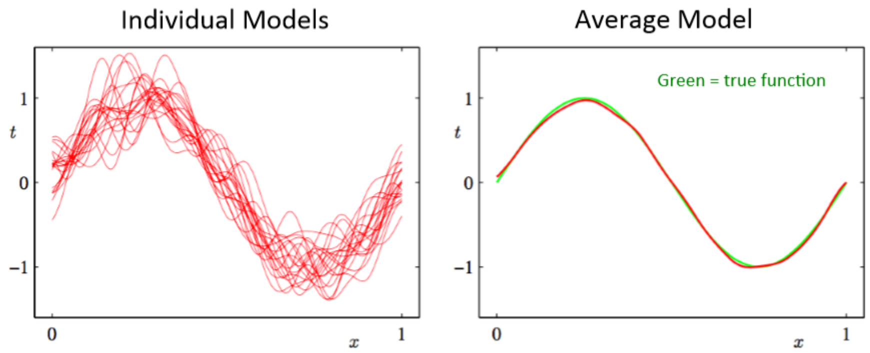
- Using **L2-norm**: $\boldsymbol{\Omega}(\boldsymbol{w}) = \sum_i \boldsymbol{w}_i^2$, adding the constraint $\boldsymbol{\alpha\Omega}(\boldsymbol{w})$ to the weights update rule: **Ridge Regression**

- The larger the value of $\boldsymbol{\alpha}$, the stronger the shrinkage and thus the coefficients become more robust.



→ **Bagging**: Bootstrap Aggregating: reduce variance of the model, without increasing systematic error of prediction:

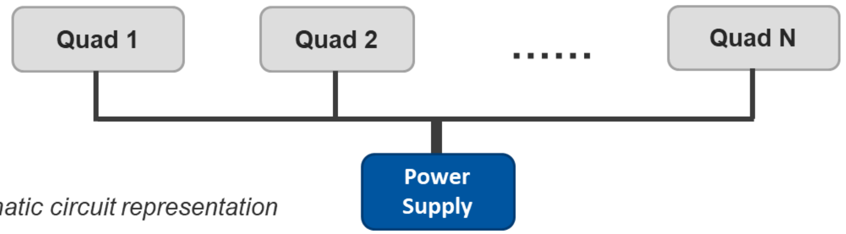
- Ensemble of slightly different models
- Train a separate model on a subset of training data
- Average output of each predictor for the final output.



[Bishop, "Pattern Recognition"]

Simplified studies: optics deviations caused by circuits errors

- **Training** data: perturb the optics by changing the strength in the **circuits (quadrupoles powered in series)**
- **Validation**: simulations perturbed with errors in **individual quadrupoles**

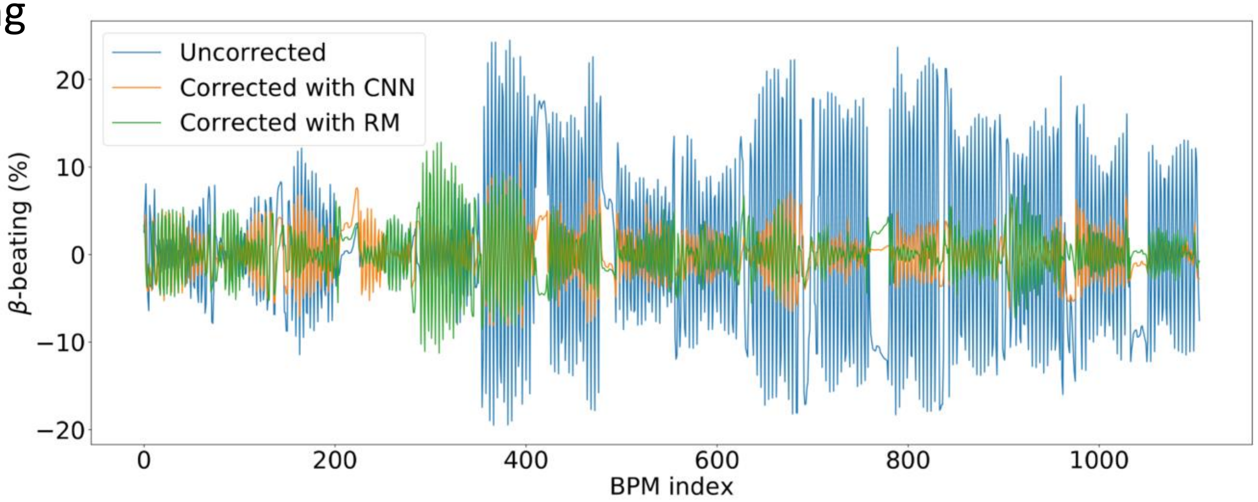


Schematic circuit representation

Different algorithms are compared: Orthogonal Matching Pursuit, Random Forest,
Convolutional Neural Network:

- ➢ Similar results
- ➢ Linear Regression as baseline model:
  - – easier to interpret,
  - – faster to train,
  - – mostly linear effects are present in simulations.

- ➢ Increasing the complexity of simulations step by step by adding additional error sources, exploring limitations of regression models.



→ Correction results using Convolutional Neural Network are similar to Response Matrix.

Data generation and model training

Training samples generated using MAD-X:

- Using nominal optics settings corresponding to settings used in uncorrected machine
- Assigned magnetic errors: quadrupolar field errors, longitudinal displacement of quadrupoles, transverse misalignment of sextupoles, dipole field errors

Realistic training data to make adequate prediction from measurements.

- **1256 target** variables
  - assigned gradient errors in the **all** quadrupoles, **both** beams.
- **3304 input** variables: simulated deviations from the design optics in betatron phase advance, normalized dispersion at all BPMs and β at BPMs next to Interaction Points.
  - Adding realistic noise estimated from the measurements.

Selected model:

- Scikit-Learn implementation of Ridge Regression
- Bagging-estimator (combining 10 Ridge Regression – models, with regularization parameter α=0.001)
- 80000 training samples (divided into training and test sets)

How to evaluate trained models?

- "**ML point of view**": compare predicted magnet errors with corresponding true values.

  Figures of merit: $MAE(y, \hat{y}) = \sum_{i=1}^{n} |y_i - \hat{y}_i|$ $\quad R^2(y, \hat{y}) = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}}$

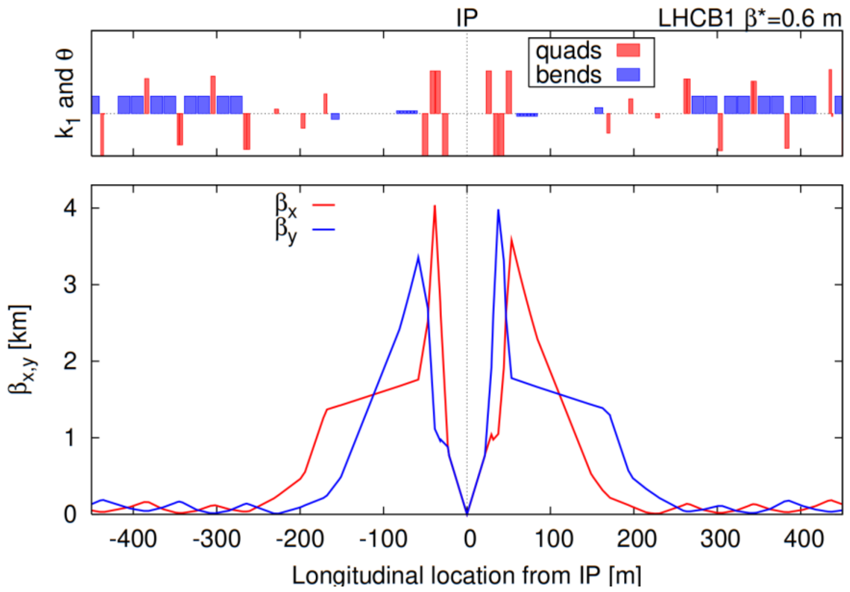
- **In terms of optics**:
  ML-model input: **optics** perturbed with magnet errors to be predicted

  ML- model output: **magnet errors** estimated from optics perturbations

How well can be reconstructed?

- Evaluating **triplet (quadrupoles next to Interaction Points)** and arcs magnet errors prediction separately:

- local correction of the triplet is the most challenging part.
- generates largest optics perturbations
- translation of individual errors in the triplets into correction settings.

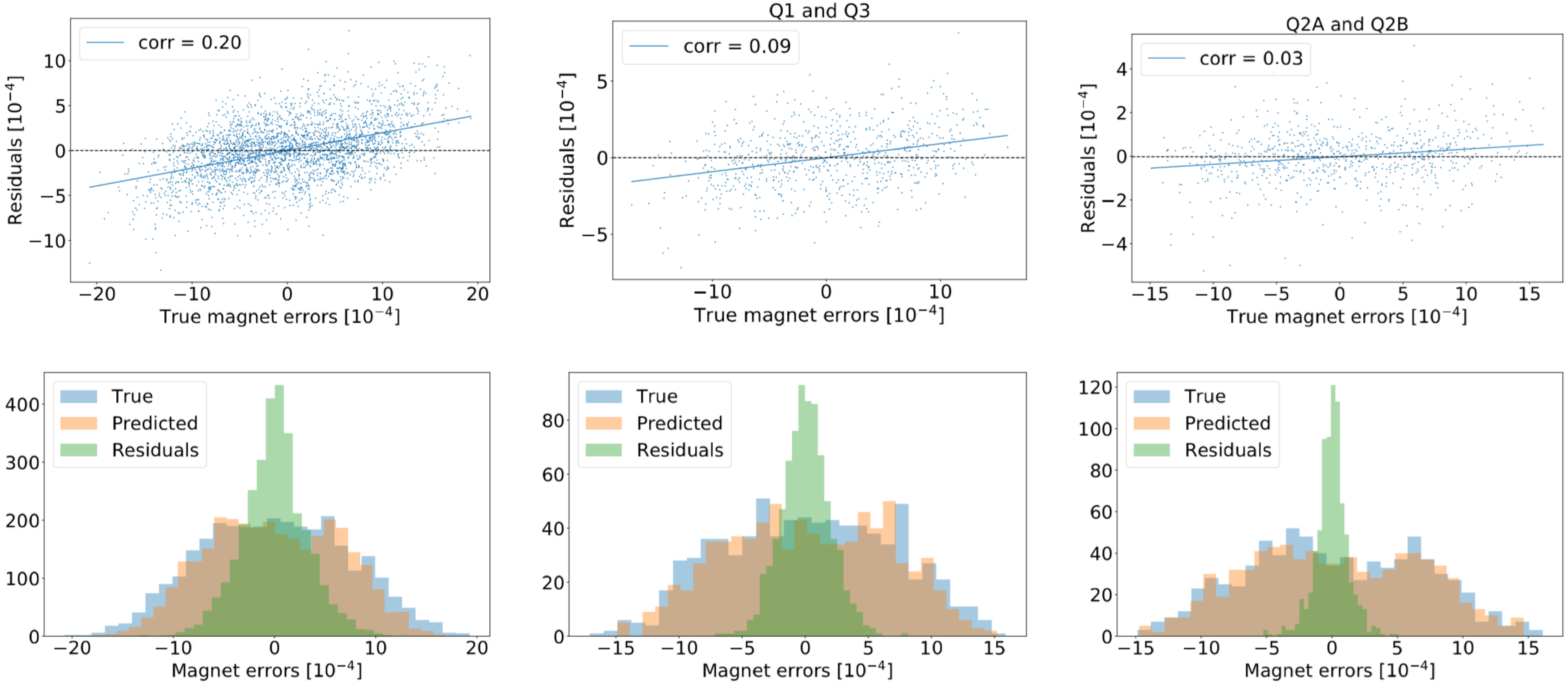


Results on simulations: errors of prediction

- Comparison between true simulated and predicted errors

Field errors in individual magnets around Interaction Points

Combining individual quadrupole errors according to the powering scheme in the LHC



systematic prediction error (bias) → 19**%,**
random error (variance) ~ 30%.

5% systematic prediction error

1% systematic prediction error

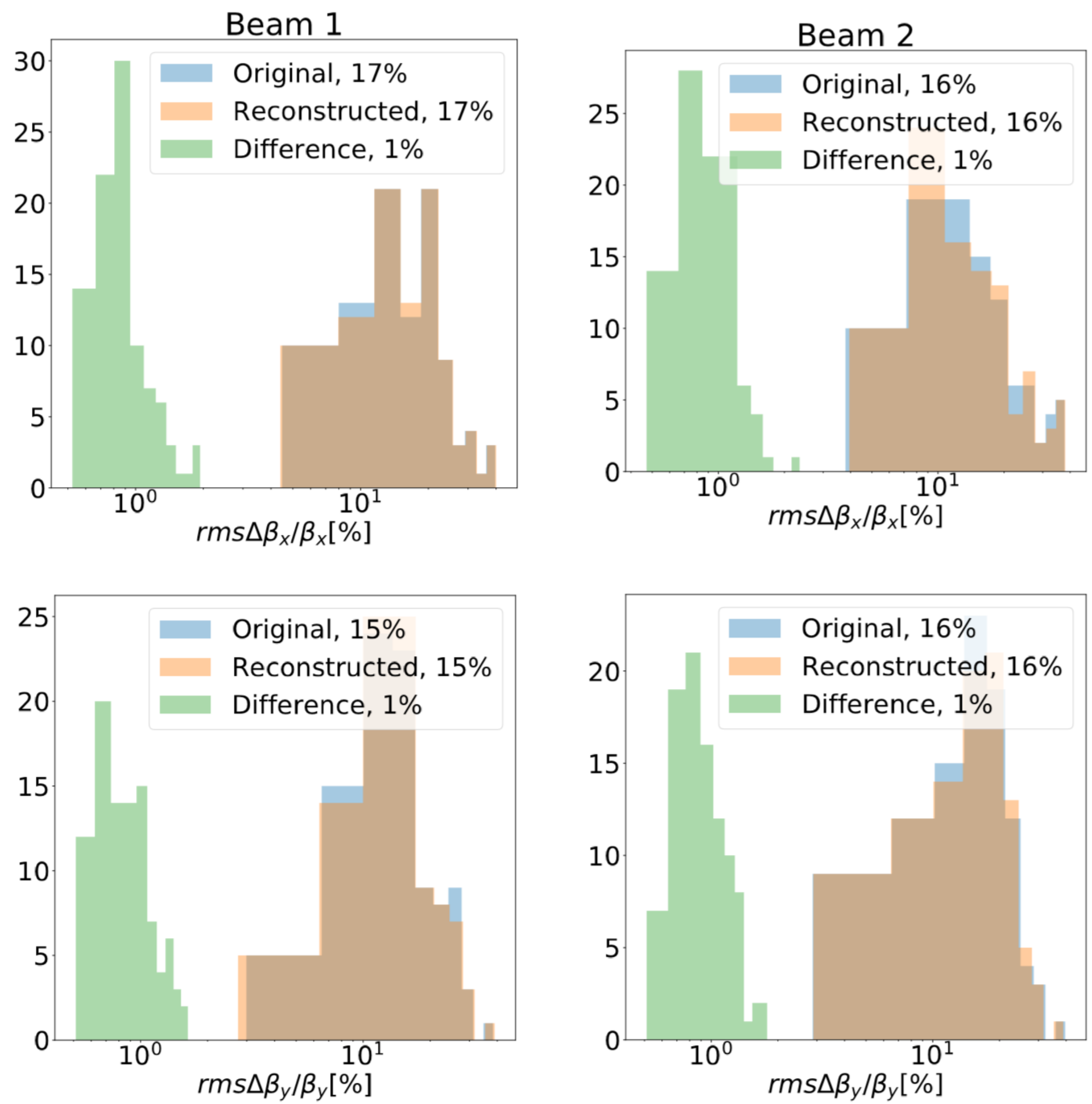
Results on simulations: comparing resulting optics errors

Ideal optics +
simulated errors = perturbed optics

Difference
$\Delta\beta/\beta_{mdl}$?

Ideal optics +
predicted errors = reconstructed optics

→ Very good agreement between the optics simulated with true magnetic errors and simulations generated with the errors predicted by the model.

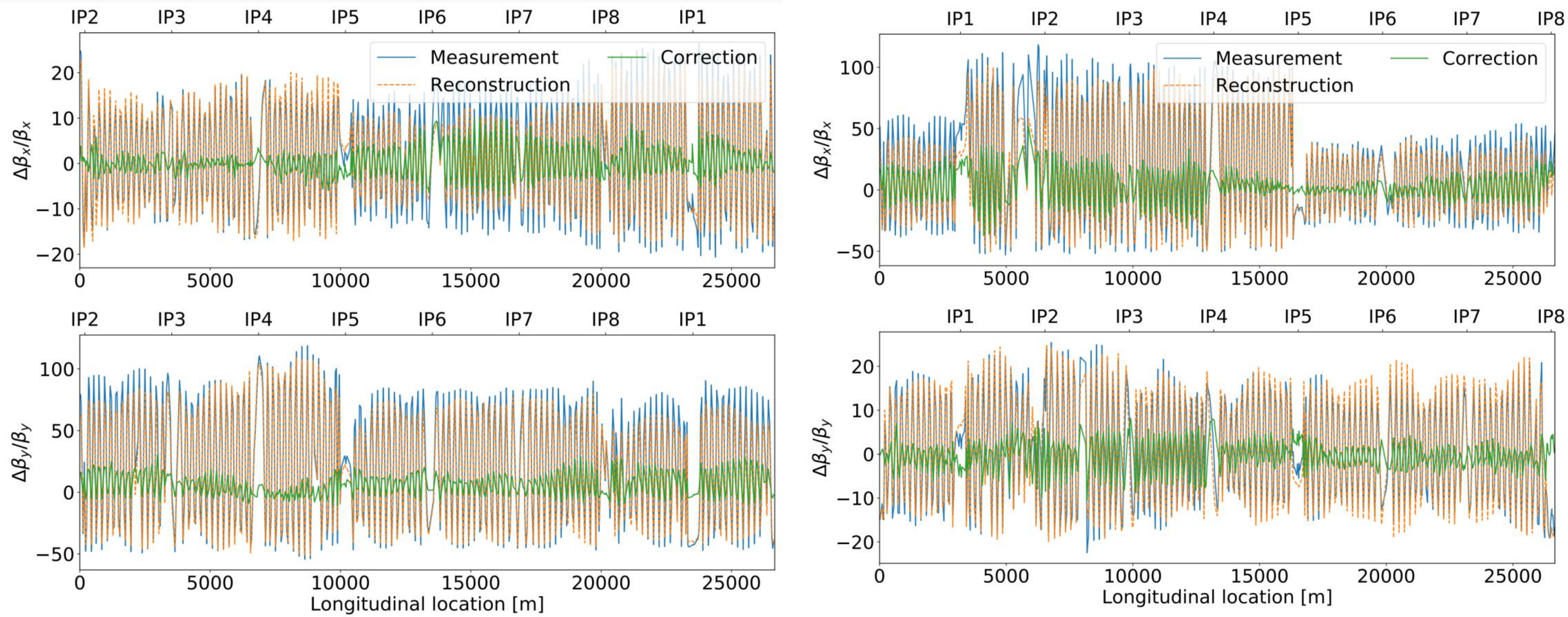


Results on experimental data: 2016 LHC commissioning

“Ground-truth” of magnet errors is unknown unlike simulations.

1. Use predicted magnet errors to simulate optics perturbation
2. Compare produced simulation to actual measurement

→ **Residual error of measured optics reconstruction** (≈ potential correction)

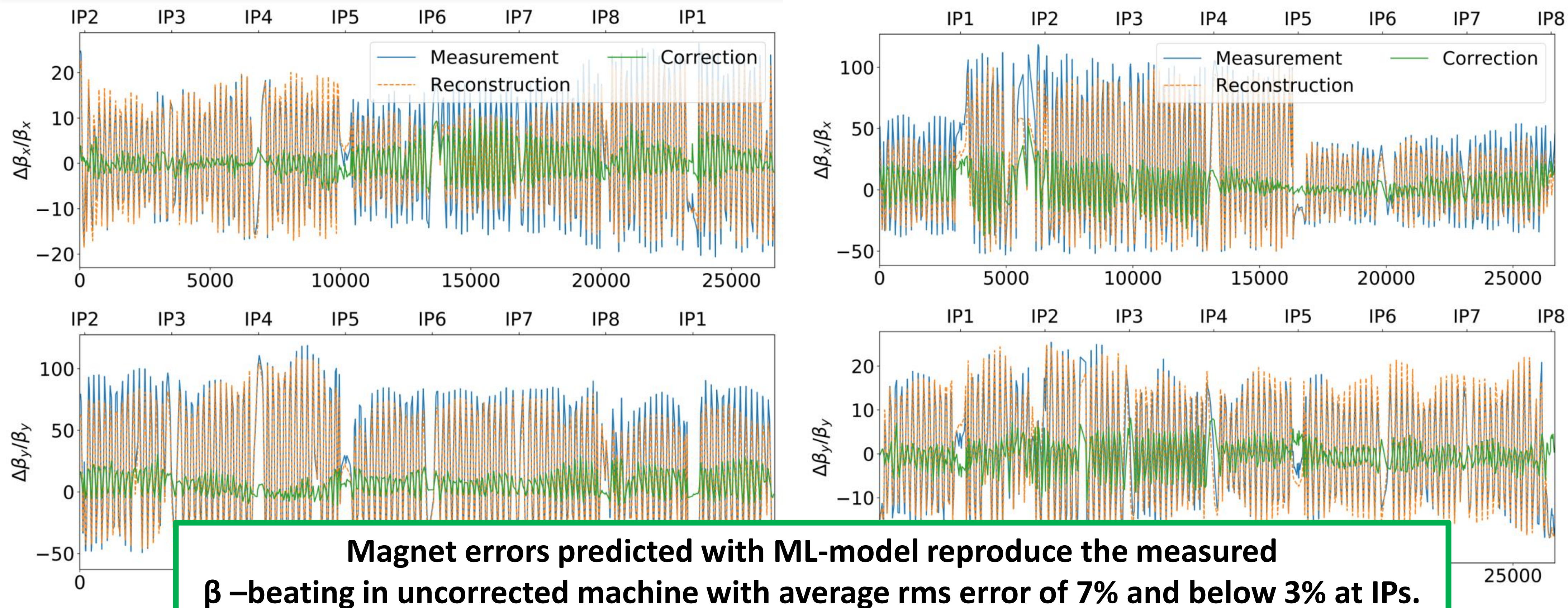


Results on experimental data: 2016 LHC commissioning

"Ground-truth" of magnet errors is unknown unlike simulations.

1. Use predicted magnet errors to simulate optics perturbation
2. Compare produced simulation to actual measurement

→ **Residual error of measured optics reconstruction** (≈ potential correction)

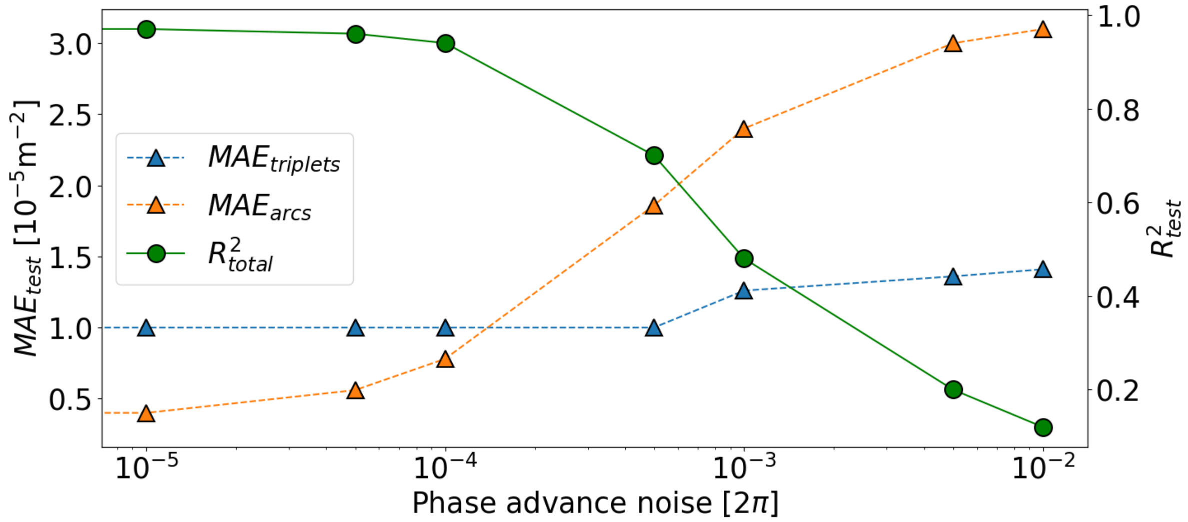


Magnet errors predicted with ML-model reproduce the measured β –beating in uncorrected machine with average rms error of 7% and below 3% at IPs.

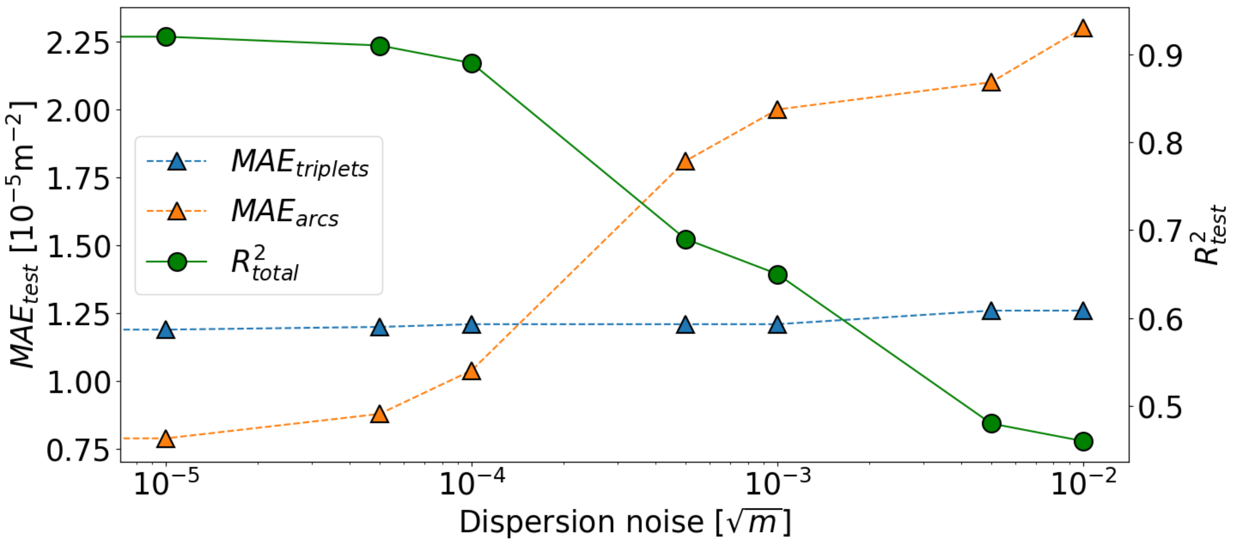
III. Denoising and reconstruction of optics functions

Effect of the noise

Model scores depending on the **phase advance noise** (other input features are not used)



Model scores depending on the **dispersion noise,** phase advance noise is unchanged



- Prediction of magnetic errors in the arcs sections suffers from the presence of noise
- Simulations in the absence of noise: very high ML-model scores
- Increasing prediction quality possible with more precise measurements of optics functions used as regression model input.

Experimental data: possible issues

- Training models on **simulations** data: **full set of input features** is always available
- Issues with using measurements as input to make new predictions:
  - General: faulty BPMs → **missing values** at the location of cleaned BPMs
  - Normalized dispersion and β at BPMs next to IPs: special measurements techniques are needed
    → **Features are not always available** e.g. depending on the measurement procedure.
    → **Noise in the input data** affects the prediction of the regression models significantly.

How to deal with missing and noisy data?

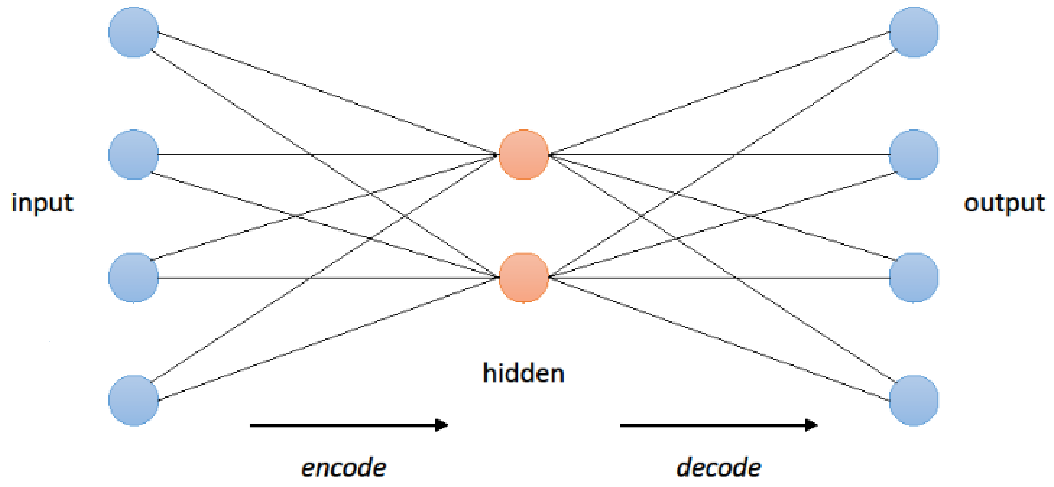
Experimental data: possible issues

- Training models on **simulations** data: **full set of input features** is always available
- Issues with using measurements as input to make new predictions:
  - General: faulty BPMs → **missing values** at the location of cleaned BPMs
  - Normalized dispersion and β at BPMs next to IPs: special measurements techniques are needed

  → **Features are not always available** e.g. depending on the measurement procedure.

  → **Noise in the input data** affects the prediction of the regression models significantly.

How to deal with missing and noisy data? → **Denoising Autoencoder**



Encoder: **compressing** the input data to lower dimensions
Decoder: **reconstructing** the data into original input.

- A special neural network designed to reproduce given input as output of the network
- **Neural Network: approximation of non-linear functions**

Applications:
- Denoising of data
- Dimensionality reduction
- Generative modeling
- Supervised and unsupervised learning

Experimental data: possible issues

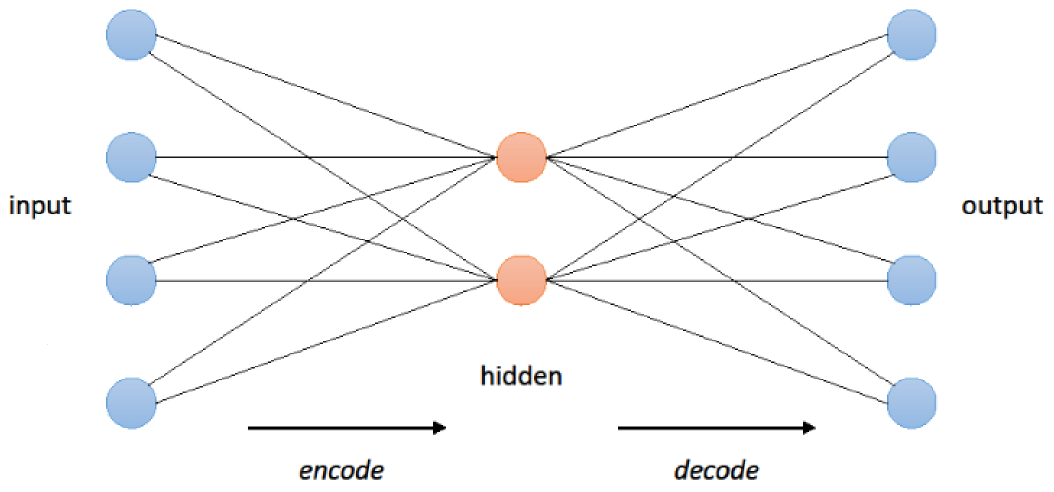
- Training models on **simulations** data: **full set of input features** is always available
- Issues with using measurements as input to make new predictions:
  - General: faulty BPMs → **missing values** at the location of cleaned BPMs
  - Normalized dispersion and β at BPMs next to IPs: special measurements techniques are needed

    → **Features are not always available** e.g. depending on the measurement procedure.

    → **Noise in the input data** affects the prediction of the regression models significantly.

How to deal with missing and noisy data? → **Denoising Autoencoder**

Simulated optics observable
+ noise
- observation at few locations

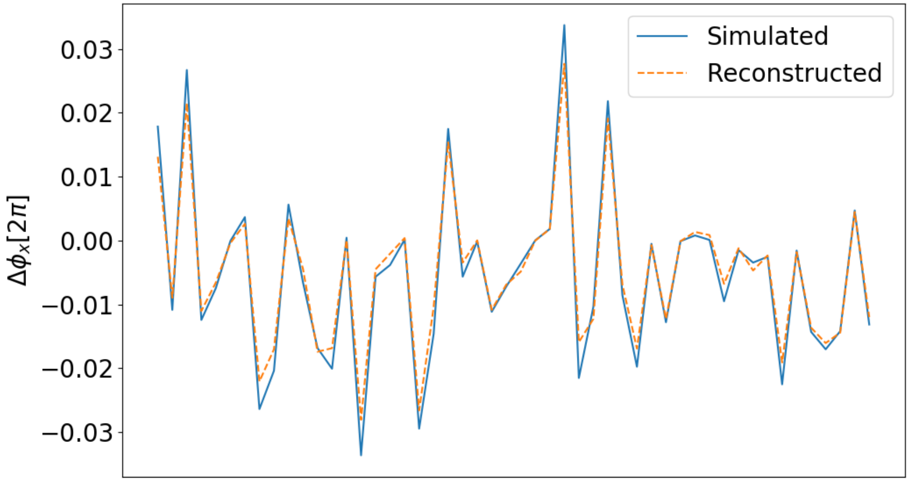


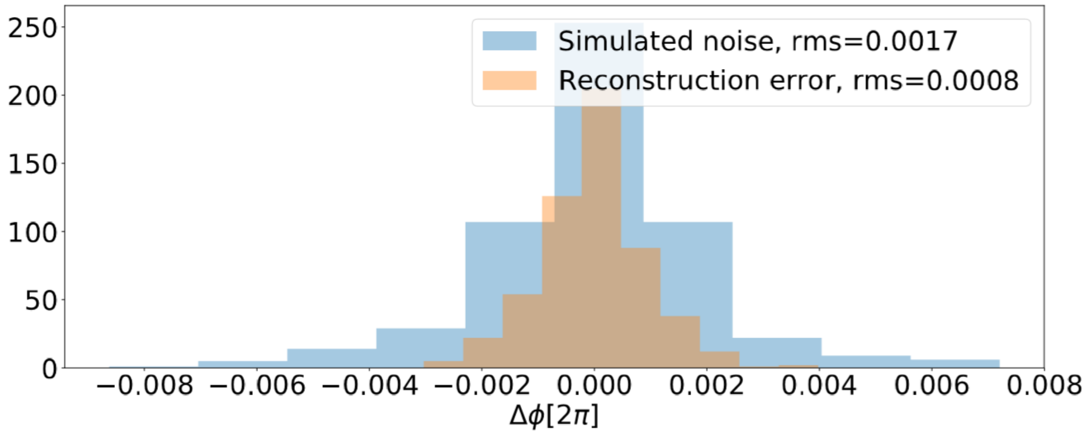
Original full set of optics observables

Reconstruction and denoising of phase advance deviations

- Input: simulated phase advance deviations given noise and replacing 10% of values with 0 (faulty BPMs)
- Output: original simulated phase advance deviations
- Autoencoder with 4 hidden layers, 10000 samples

Reconstruction of missing values in a validation sample



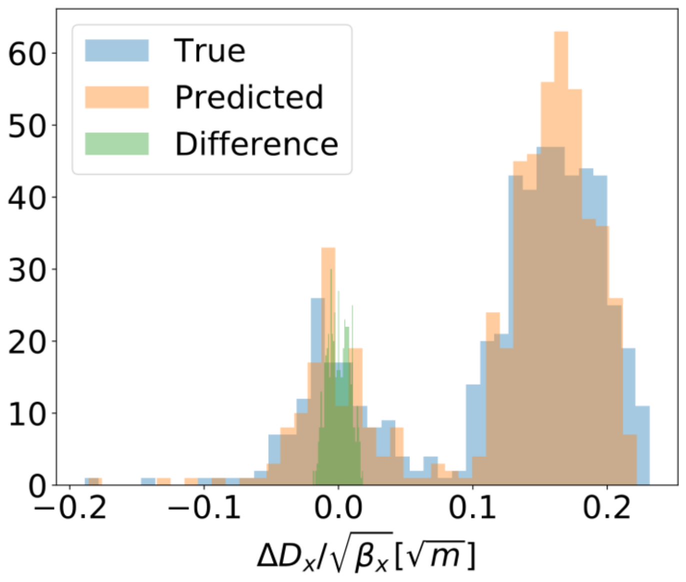


- ✓ **Missing BPMs**: possibility to obtain reliable estimation of the phase advance deviations at the location of faulty BPMs.
- ✓ **Full set of phase advance deviations**: reconstruction error is by factor 2 smaller than simulated realistic noise.

Reconstruction of normalized dispersion from phase advance deviations

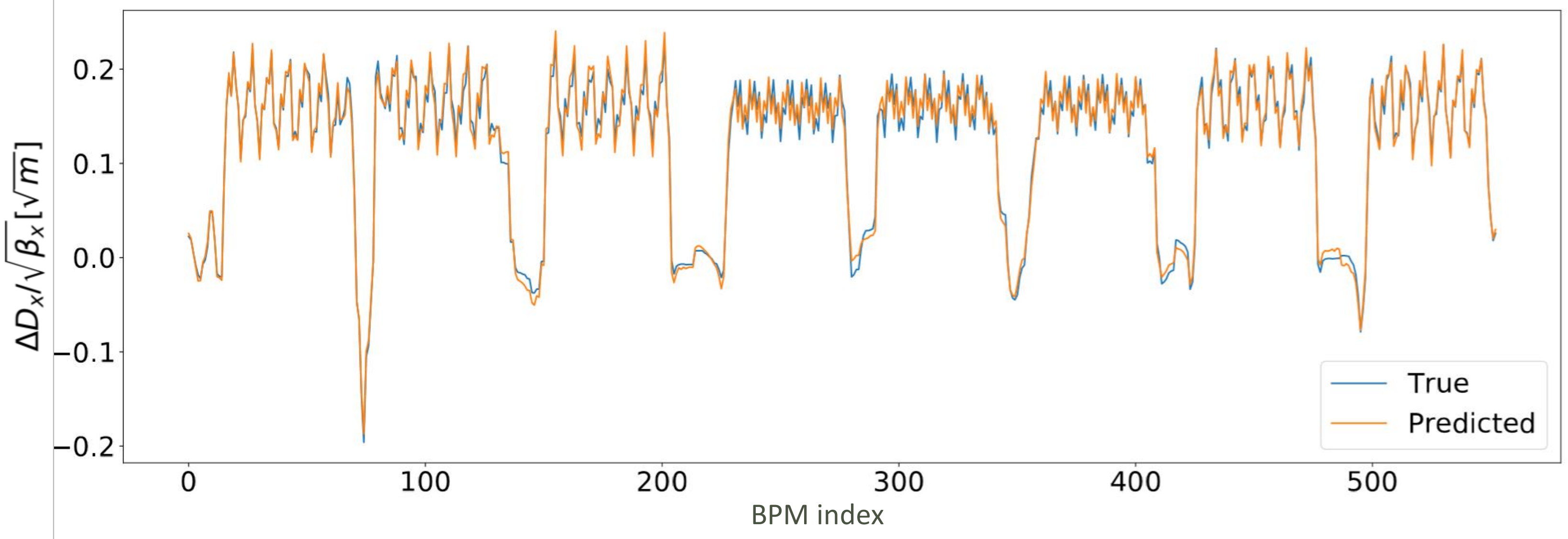
- **Input**: simulated phase advance deviations given noise
- **Output**: normalized dispersion $\Delta D_x/\sqrt{\beta_x}$
- Using **linear regression model**: Ridge Regression, 10 000 samples

***Simulation** example: Beam 1*



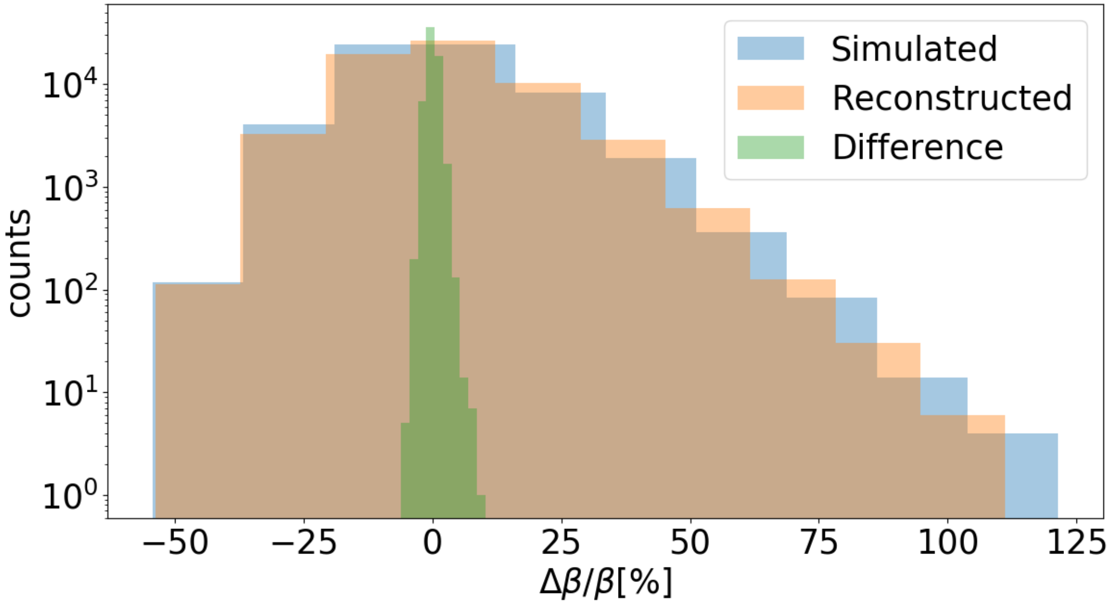
Simulated rms $\Delta D_x/\sqrt{\beta_x}$: 0.0802 $\sqrt{m}$

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



Reconstruction of β from phase advance deviations

Simulation: *summary of 1000 seeds*

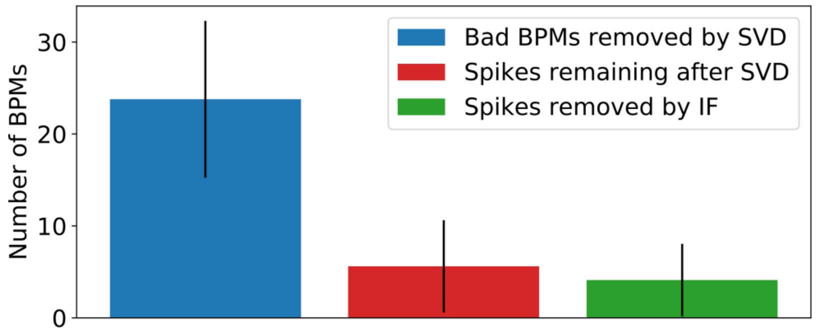


- **Input**: simulated phase advance deviations given noise (beam 1 and 2, horizontal and vertical planes)
- **Output**: $\Delta\beta$ errors at 2 BPMs left and right from IPs 1, 2, 5 and 8 (32 variables in total)
- Ridge Regression, 10 000 training samples

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

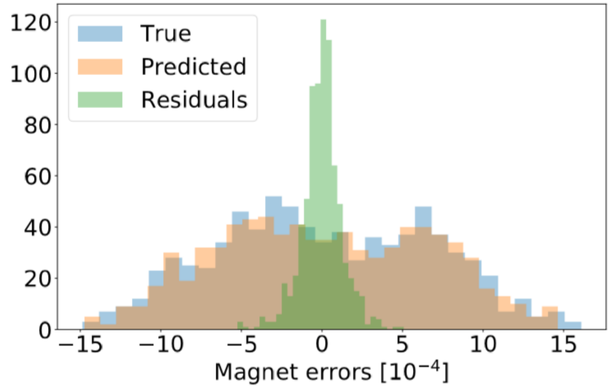
Conclusion and outlook

- **Detection of faulty BPMs:**
  - ○ Predefined rules and thresholds are not sufficient to identify faulty BPMs.
  - ✓ Unsupervised Learning based cleaning technique became **fully operational** standard part of optics analysis
  - ✓ Identified previously unknown bad BPMs efficiently in 2018 **without human intervention**.



Summary of measurements in 2018, nearly all unphysical outliers can be removed by IF

- **Estimation of magnet errors and optics reconstruction:**
  - ○ Optics corrections today are done in two steps (local and global).
  - ✓ ML-models allow to **predict all quadrupole errors** for both beams simultaneously, local and global errors in one step
  - ✓ Promising results on simulations and experimental data, especially for **optics corrections in Interaction Regions** (1 - 5% systematic error)
  - ○ Current limitations:
    - Linear error sources in training simulations
    - Prediction of arc magnet errors highly depends on the noise in the measured optics observables.



Residual error for a correlated group of triplet quadrupoles

Conclusion and outlook

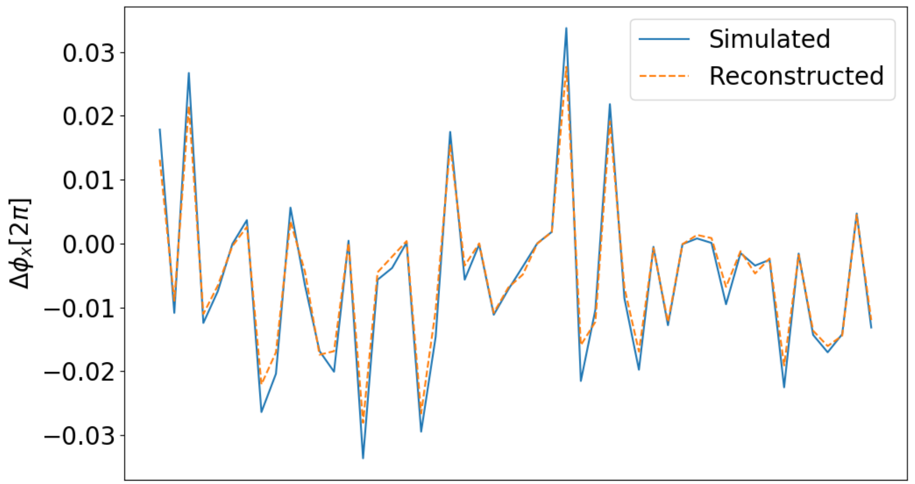
- **Denoising and reconstruction of missing data:**

✓ Successfully demonstrated on simulations the possibility **reduce noise** of phase advance measurements using autoencoder

✓ **Reconstruction of missing features** for the magnet errors prediction → re-training on available data is not needed

✓ Linear regression models to reconstruct optics observables from phase advance deviations

✓ Providing **optics functions estimates**, when time costly measurements techniques cannot be performed

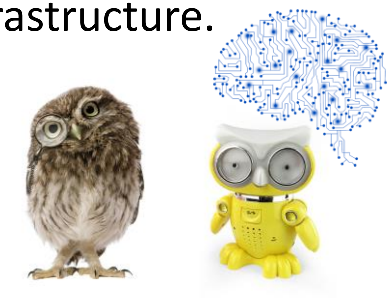


- **Outlook:**

○ Optics-independent model: mixed training set of simulations generated using different nominal optics settings

○ Correction knobs from predicted individual errors

○ Integration into operational LHC software infrastructure.



Thank you very much for your attention!