Supervised Learning for Nonlinear Corrections in the LHC

Alejandro Börjesson Carazo

Thanks to Felix Simon Carlier and Elena Fol

15/02/2024

Summary

1. **Introduction**
   1. Motivation
   2. Methods
2. **Results**
   1. Response Matrix vs ML Corrections
   2. Performance on crossing angle setup
3. **Conclusions**
4. **Backup slides**

Motivation

What is the problem?

- Sextupolar and octupolar magnetic errors in the triplets of the LHC cause nonlinear perturbations in the beam dynamics
- This nonlinear motion has an impact on stability => **Lifetime of the beam decreases**

Current state of nonlinear commissioning in the LHC:

- Time consuming and iterative
- Multitude of techniques, crossing angle scans, amplitude detuning, Resonance Driving Terms (RDTs)…

Can we use only RDTs to correct multiple orders at ONCE?

Methods: Data generation

What is a Resonance Driving Term (RDT)?

- A order specific nonlinear optics observable

$$\xi_{x,-}(N) = \sqrt{2I_x} e^{i(2\pi Q_x N + \psi_{x0})} - 2i \sum_{jklm} j f_{jklm} (2I_x)^{\frac{j+k-1}{2}} (2I_y)^{\frac{l+m}{2}} \times e^{i[(1-j+k)(2\pi Q_x N + \psi_{x0}) + (m-l)(2\pi Q_y N + \psi_{y0})]}$$

Position

RDT

Linear motion

Nonlinear motion

This can be obtained from simulation codes but also measured from turn by turn data

Using simulation data to train a realistic ML error prediction model, to be used in commissioning

Methods: Data generation

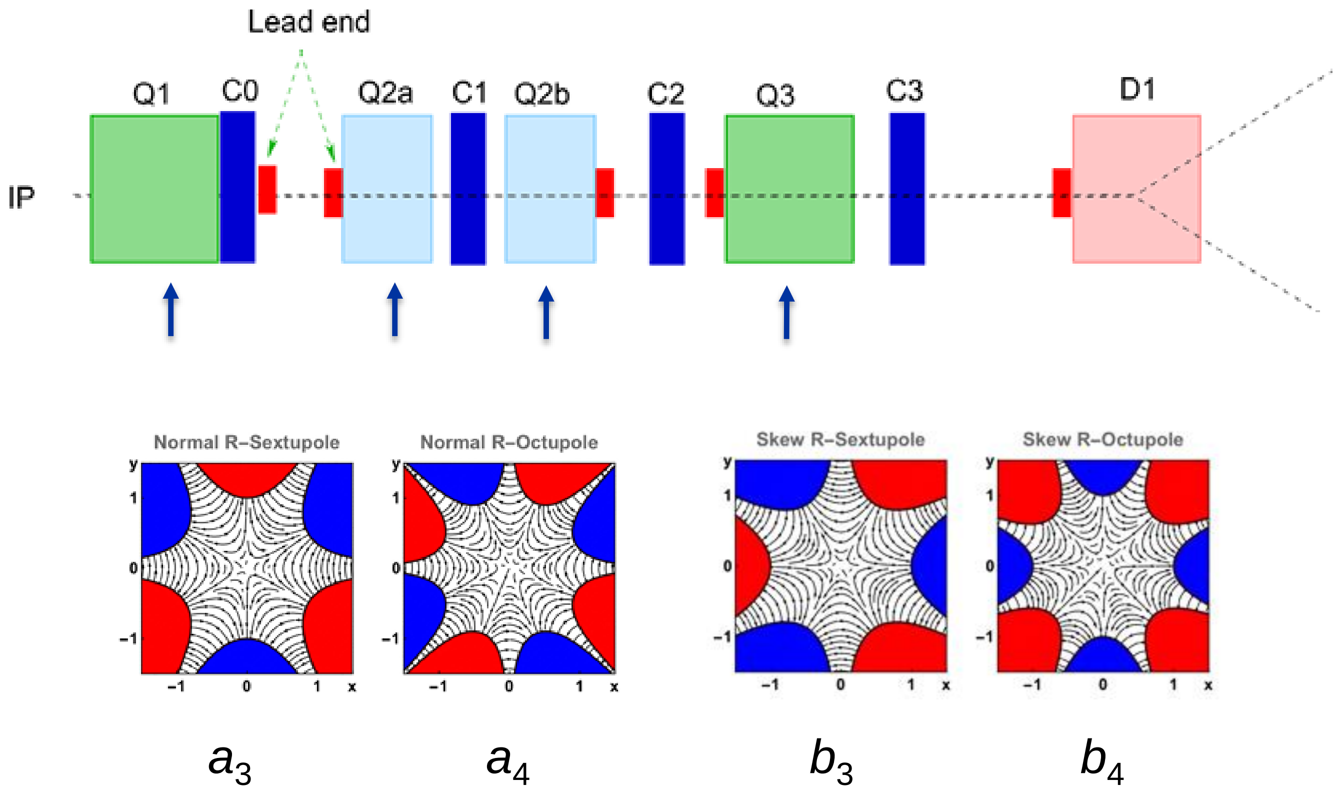
Errors assigned to **ALL the triplets in all IPs** according to the WISE tables [3] [4] **AT THE SAME TIME**

Generate RDT data, 30K samples using MADNG running on HTCondor

- **MADX-PTC** execution time: **27.17 [min]**
- **MADNG** execution time: **20.00 [s] 82 times faster!**

The goal is to predict errors for IP1 and IP5 triplets

Actual corrector strength can be calculated with this errors



[1] [2] *Fig 1. Simulation setup*

Methods: Supervised Learning

$$Loss = Error(Y - \widehat{Y}) + \lambda \sum_{1}^{n} w_i^2$$

Best performing: **Quadratic Polynomial regression** with L2 regularization and bagging

Complicated nonlinear motion and simulations, **but for the most part RDTs and errors are linearly correlated!** Some RDTs might propagate non linearly

Ensemble of 10 different regressions trained on different subsets of the data

- **Input:** 8 different RDTs (real and imaginary) simulated all around 376 BPMs in the LHC => 12032 Dim
- **Output:** Skew and normal sextupolar and octupolar errors in the main quadrupoles for IP1 and IP5 => 64 Dim

Not using IR BPMs since they can´t be measured, this has great impact on performance

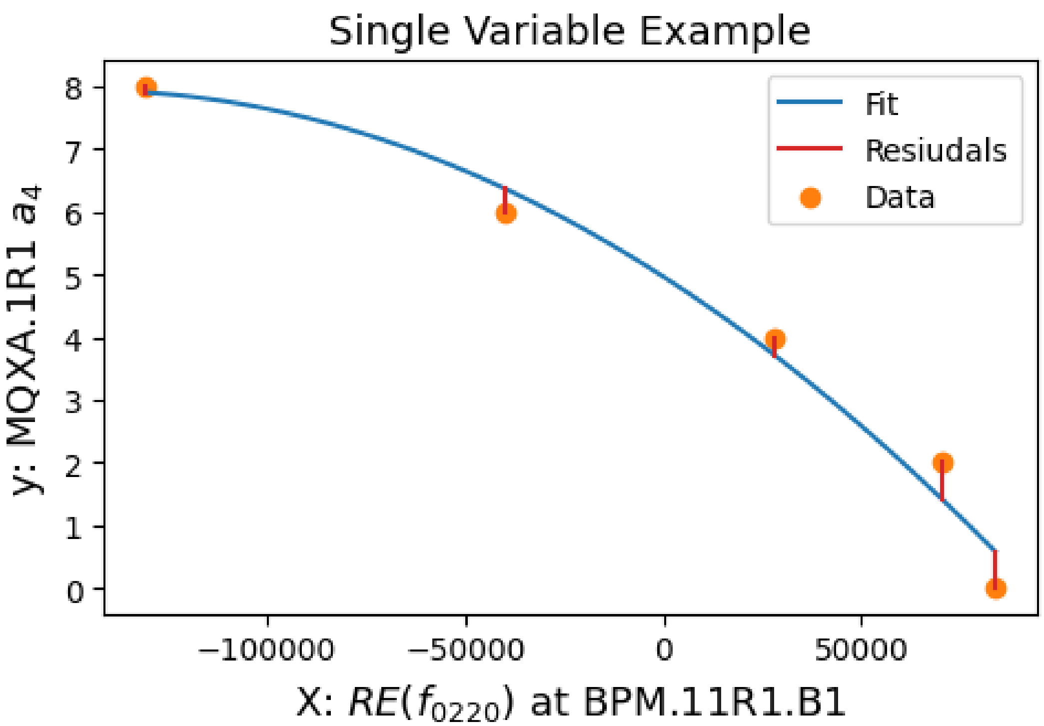


Fig 2. One-dimensional polynomial regression example with xing angles

Methods: Supervised Learning

Finding a subset of best quality observables, 2 beams
40 RDTs and 376 BPMs

This means a 60160 dimension input!

Feature extraction:

- Most correlated RDTs with error
- Highest amplitude in tune signal RDTs
- Highest phase advance resolution

Beware of conjugate RDTs

⇒ Only 8 RDTs are chosen at the moment!

Reducing the number of BPMs is an ongoing study

	Octupolar:	Sextupolar:
Normal	f_{4000}	f_{3000}
	f_{0220}	f_{1020}
Skew	f_{0130}	f_{0030}
	f_{1030}	f_{2010}

Tab 1. RDTs chosen as input

Results: Machine Learning VS Response Matrix

ML Model:

- ML model is able to reconstruct original errors: $\boldsymbol{R^2}$ **= 0.883 Test**

Lets see how it performs against a response matrix approach

$$R = \begin{bmatrix} \frac{\Delta O_1}{\Delta k_1} & \frac{\Delta O_1}{\Delta k_2} & \cdots\cdots & \frac{\Delta O_1}{\Delta k_n} \\ \frac{\Delta O_2}{\Delta k_1} & \frac{\Delta O_2}{\Delta k_2} & \cdots\cdots & \frac{\Delta O_2}{\Delta k_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\Delta O_m}{\Delta k_1} & \frac{\Delta O_m}{\Delta k_2} & \cdots\cdots & \frac{\Delta O_m}{\Delta k_n} \end{bmatrix}$$

Response Matrix:

- **One response matrix for each order error using same observables as the ML model**
- Response matrix approach is more sensitive to degeneracy, only works at correcting, but not for predicting error sources

$$\Delta \mathbf{k} = R^{\dagger} \mathbf{O}$$

How good are these corrections?

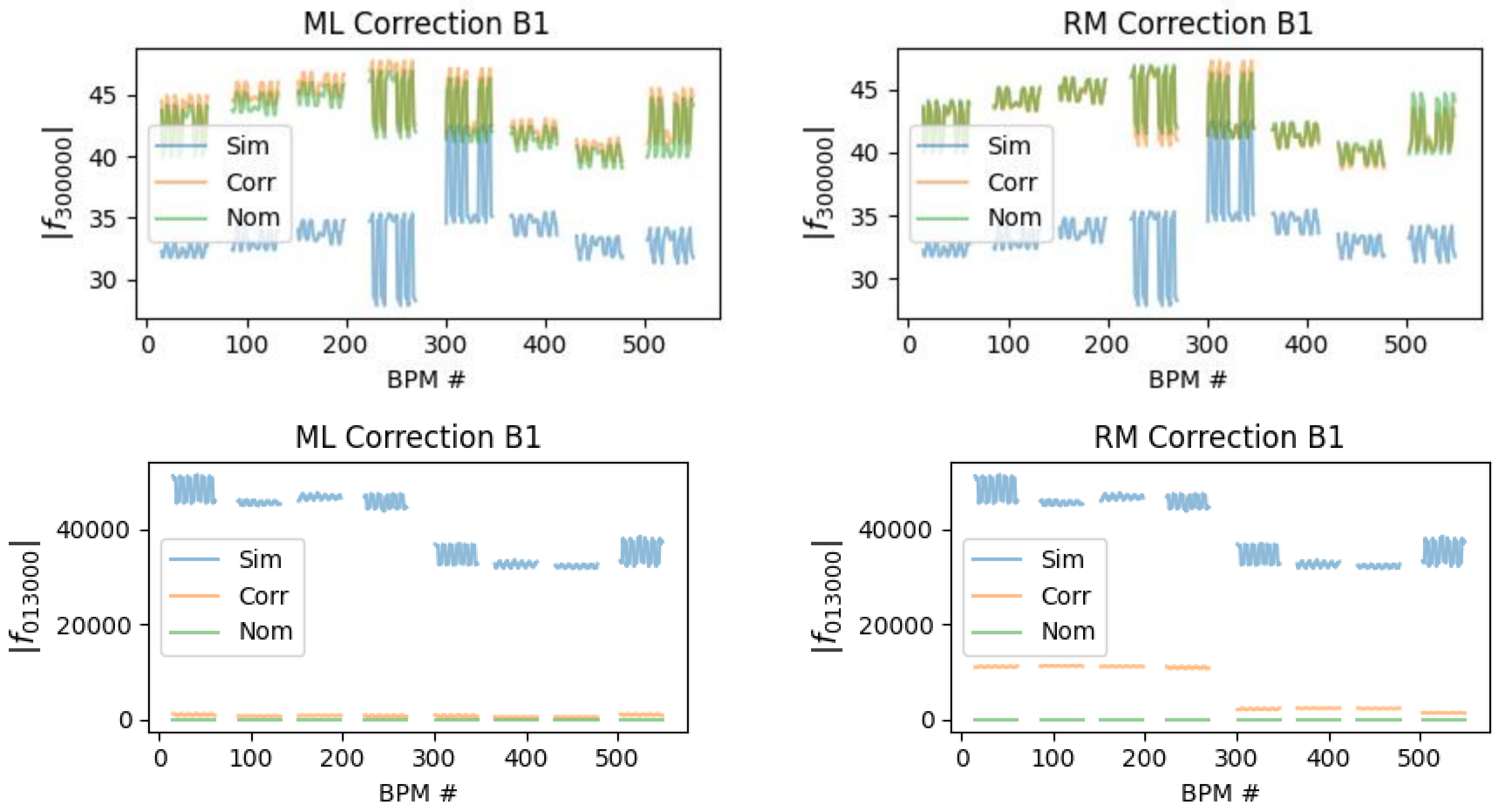


Fig 3. Example Sample correction for RDTs

Results: ML vs RM

ML Method

- All RDTs seem to be corrected, even ones that where not used in the algorithm ie. f_{3100}

Response matrix

- For the most part corrections are working
- Struggling to correct octupolar RDTs this migth be due to a second order effects from sextupoles
- RDTs not used in the model also corrected

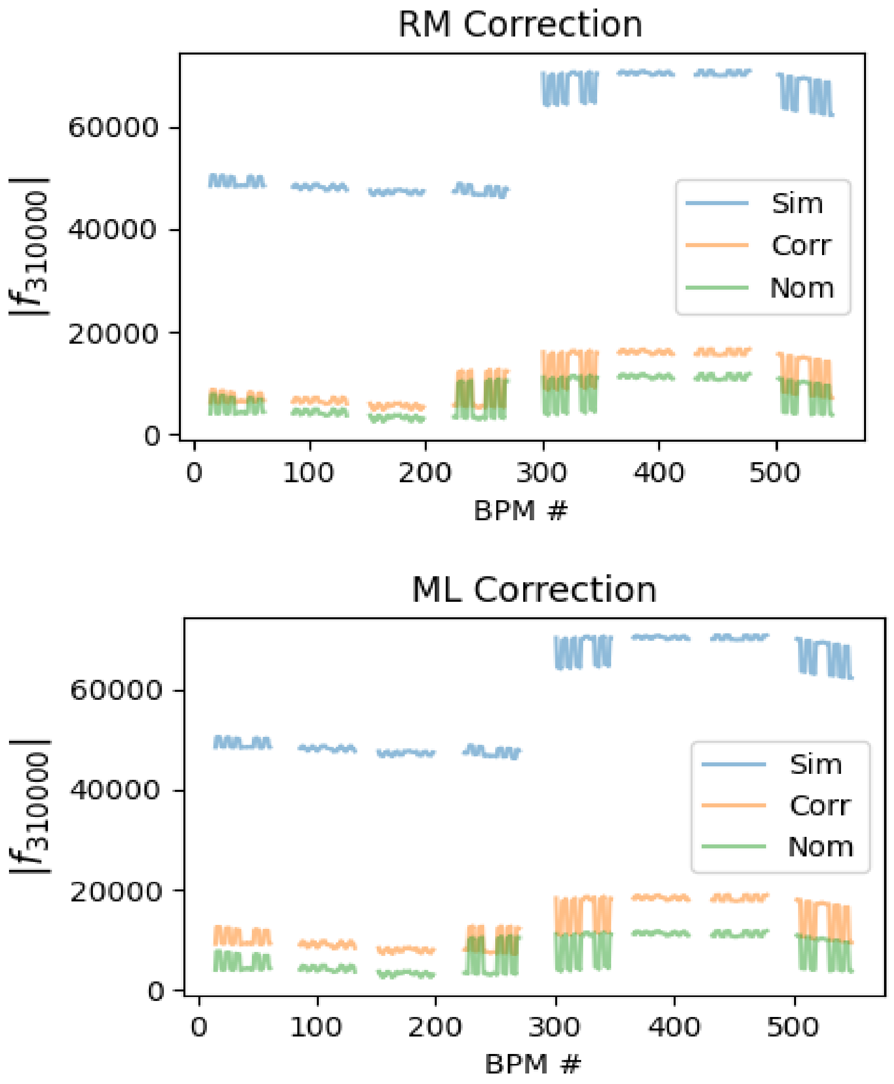


Fig 4. Performance in a RDT that is not included in the model

Results: ML vs RM

Performance over multiple samples

- Correcting 1000 samples and making a histogram with the RMS deviance from nominal
- Comparing performance from both methods for the used RDTs

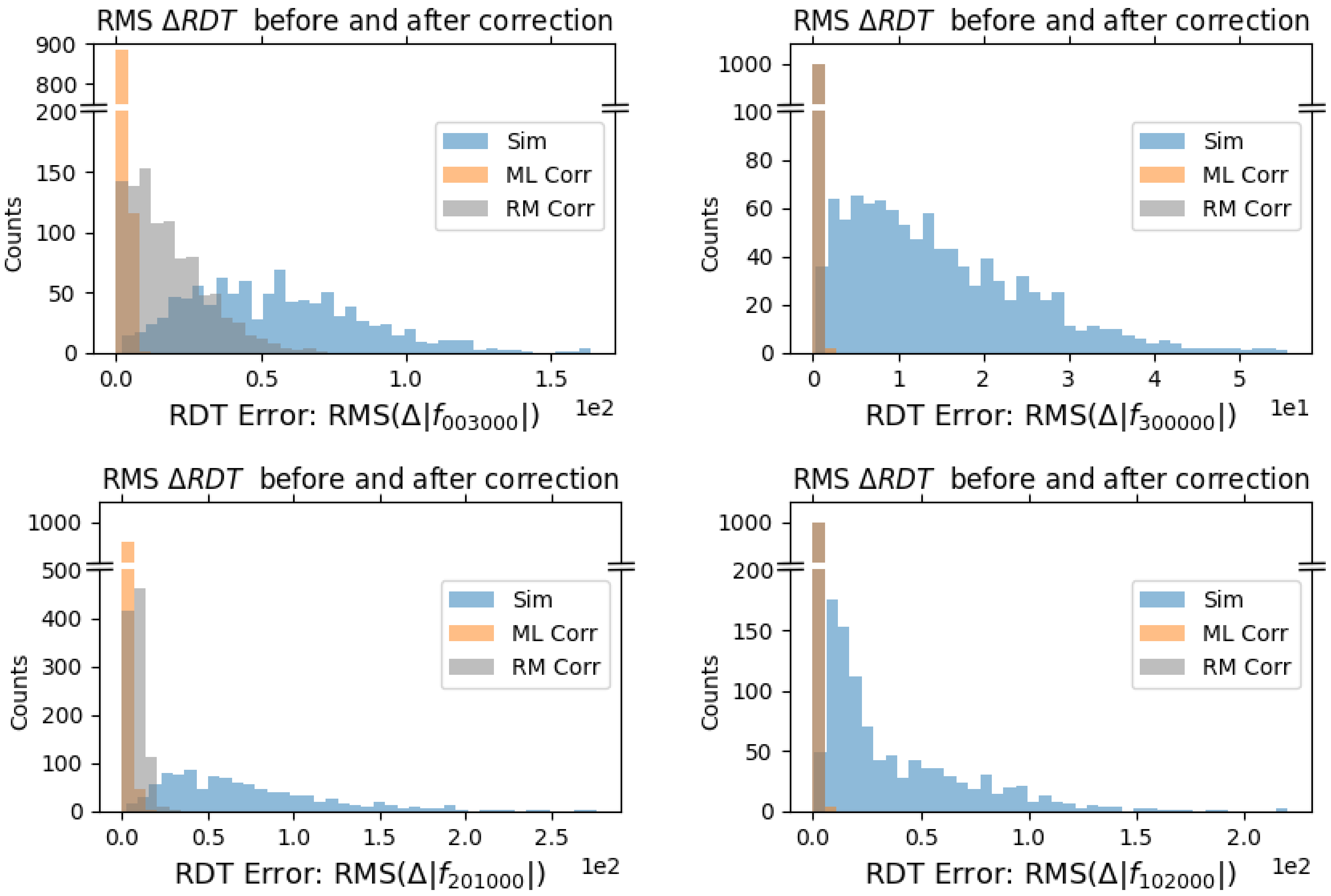


Fig 5. Correction histograms for 1000 samples for sextupolar RDTs

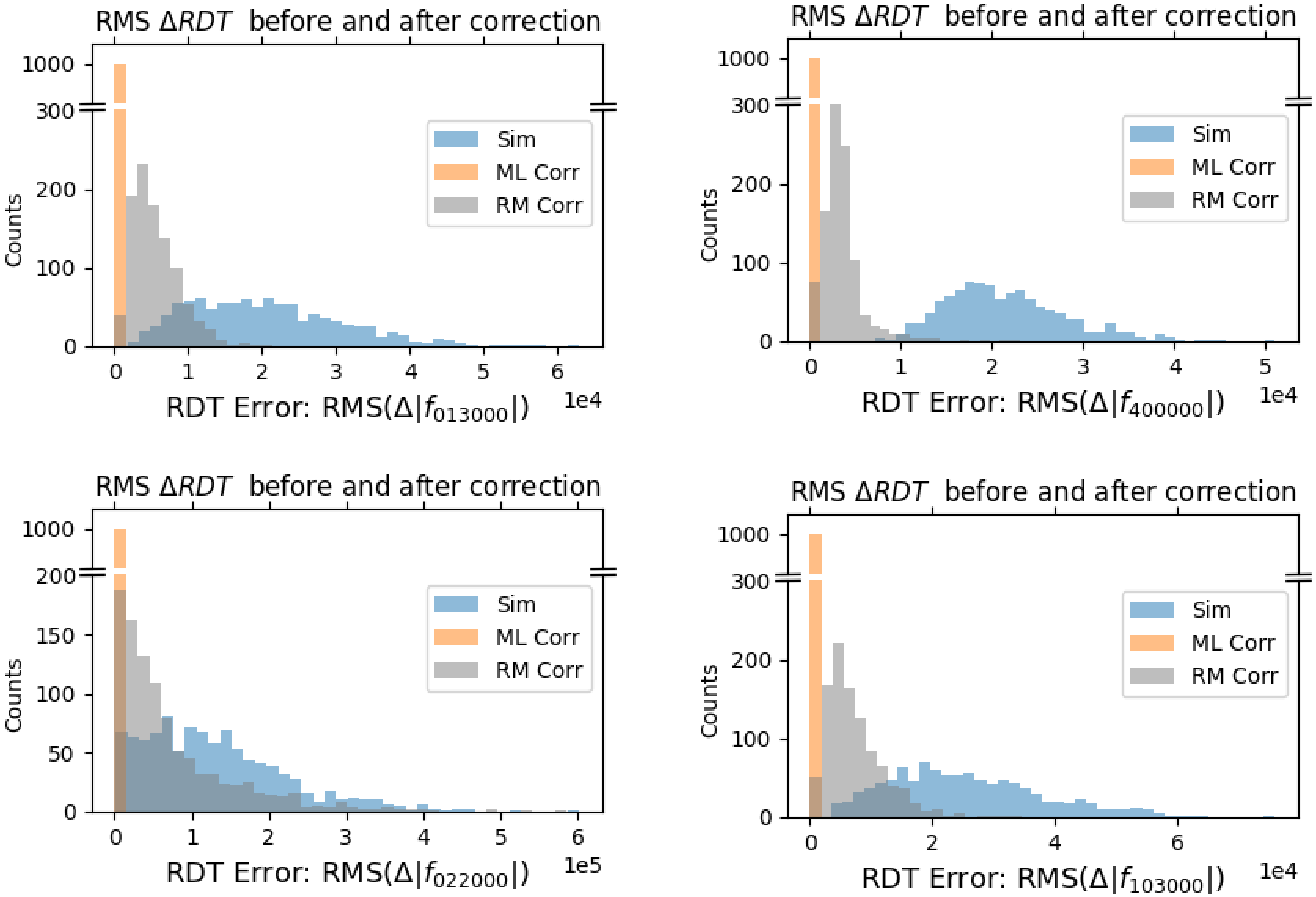


Fig 6. Correction histograms for 1000 samples for octupolar RDTs

Results: ML vs RM

- The ML correction performs better in all RDTs and is simultaneous
- I have found that it is usually more robust than the response matrix approach

Results: Performance with crossing angles

Crossing angles in the IRs result in off axis beams in the triplets

- Causes mixing of nonlinear modes
- Very challenging for response matrix

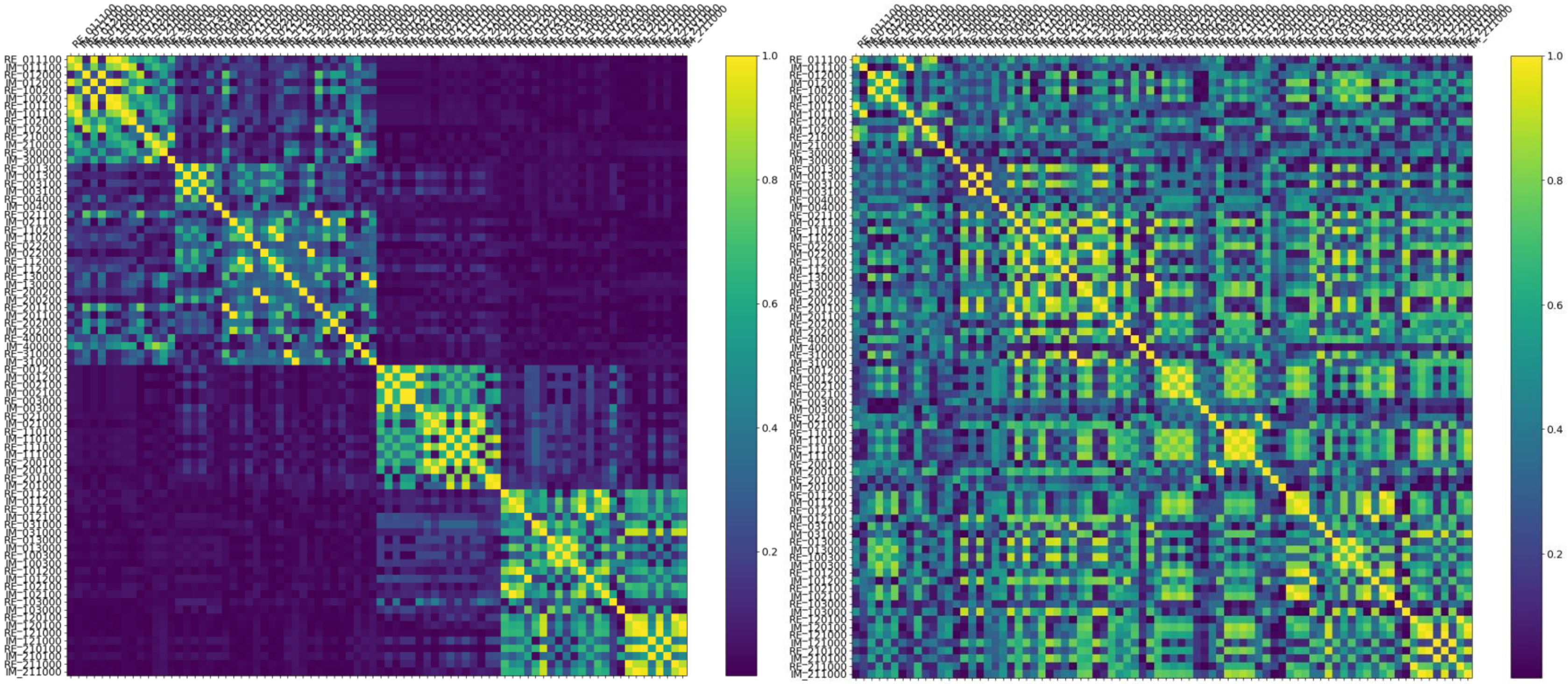


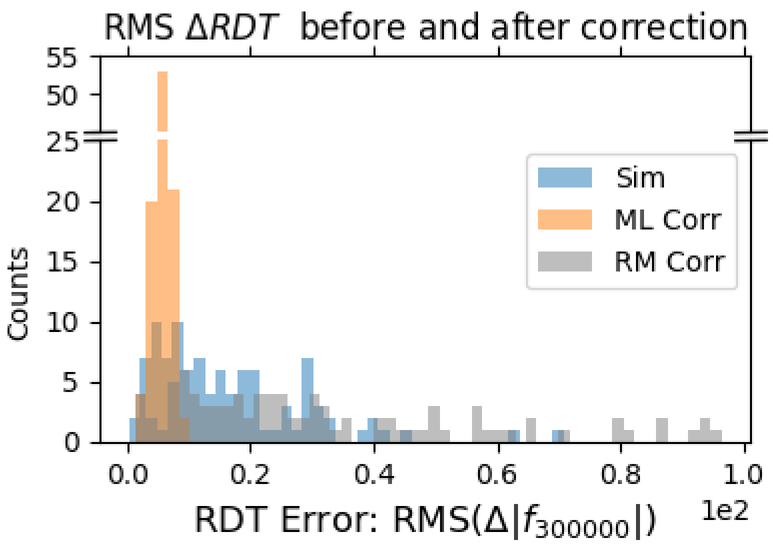
Fig 7. Correlation matrices for RDTs with and without crossing angles

Results: Performance with crossing angles

ML Model:

- Training on data with xing angles yields other working model
- ML Method can be improved, but works in general

Could not get response matrix to work



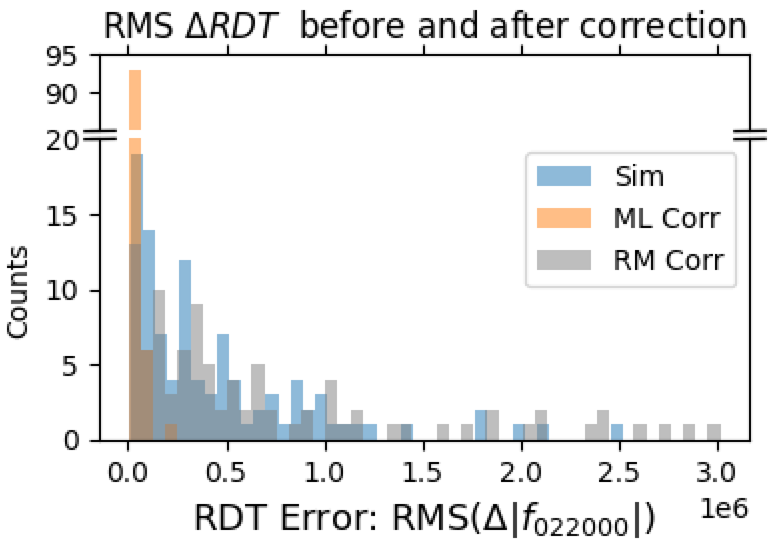


Fig 8. Correction histograms for 100 samples with a xing angle setup

Conclusions

- Faster simulation codes (MADNG) open up new possibilities for more computationally intensive modelling for nonlinear studies

- These ML techniques allow for a more complex modelling of errors and resonances

- Thus far machine learning seems to be a **feasible tool for correcting multiple RDTs at once**

- Performance is more robust and yields better results than an equivalent response matrix method

- In order to test in operation a realistic noise must be modelled as well as using driven RDTs

References

- [1] *Fig 2. Simulation setup* new approach to LHC optics commissioning for the nonlinear era By E.H Maclean *https://https://journals.aps.org/prab/pdf/10.1103/PhysRevAccelBeams.22.061004*
- [2] *Fig 1. Sector magnets or transverse electromagnetic fields in cylindrical coordinates, By T Zolkin, 2017, https://doi.org/10.1103/PhysRevAccelBeams.20.043501*
- [3] *Magnetic model of the inner triplet quadrupole MQXB. By Joe Di Marco et al. 2009, https://edms.cern.ch/ui/file/2458932/1/fidel_magnet_report_MQXB_doc_cpdf.pdf*
- [4] *Magnetic model of the LHC interaction region quadrupoles MQXA. By N. Ohuchi and E. Todesco, 2009, https://edms.cern.ch/ui/file/2458928/1/fidel_magnet_report_MQXA_docx_cpdf.pdf*

Backup Slides: Performance metrics

Coefficient of determination: R^2

- R^2 is a measure that indicates how much of the data variance can be explained by the model, R^2=1 means a perfect score

$$R^2 = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}}$$

$$\mathrm{SSR} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Mean Average Error: MAE

- Average absolute error made by the model

$$\mathrm{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Backup Slides: Response Matrix

$$R = \begin{bmatrix} \frac{\Delta O_1}{\Delta k_1} & \frac{\Delta O_1}{\Delta k_2} & \cdots\cdots & \frac{\Delta O_1}{\Delta k_n} \\ \frac{\Delta O_2}{\Delta k_1} & \frac{\Delta O_2}{\Delta k_2} & \cdots\cdots & \frac{\Delta O_2}{\Delta k_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\Delta O_m}{\Delta k_1} & \frac{\Delta O_m}{\Delta k_2} & \cdots\cdots & \frac{\Delta O_m}{\Delta k_n} \end{bmatrix}$$

$$\Delta \mathbf{k} = R^{\dagger} \mathbf{O}$$

- Currently the preferred method used for optics corrections
- This can be seen as a specific case of linear regression that calculates the weights of the model using two points, for each variable
- The pseudoinverse matrix is calculated using Singular Value Decomposition (SVD)
- Multiple approaches were tested
- Using the **same observables and magnets** as the previously explained ML method

Backup Slides: WISE tables

Table VIII: Not allowed multipoles, average and spread over the 18 magnets, at four different currents ("Integral" data).

		b3	b4	b5	b7	b8	b9	a3	a4	a5	a6	a7	a8	a9
392 A	Ave	-0.12	1.28	0.04	0.00	0.03	-0.01	0.35	-1.32	0.09	-0.01	-0.01	0.00	0.00
	Std	1.05	0.15	0.31	0.03	0.02	0.02	1.01	1.27	0.27	0.04	0.03	0.02	0.01
3207 A (3.3 TeV)	Ave	0.01	1.24	0.00	0.00	0.02	0.00	0.21	-0.06	0.02	-0.03	0.00	0.00	0.00
	Std	0.28	0.11	0.04	0.01	0.00	0.00	0.35	0.26	0.04	0.02	0.01	0.01	0.00
6177 A (6.3 TeV)	Ave	0.03	1.28	-0.01	0.00	0.02	0.00	0.21	-0.02	0.01	-0.03	0.00	0.00	0.00
	Std	0.30	0.11	0.04	0.01	0.00	0.01	0.36	0.27	0.04	0.02	0.01	0.01	0.00
6677 A (6.8 TeV)	Ave	0.04	**1.30**	0.00	0.00	0.02	0.00	**0.21**	-0.02	0.01	-0.03	0.00	0.00	0.00
	Std	0.31	0.11	0.04	0.01	0.00	0.01	0.37	0.28	0.04	0.02	0.01	0.01	0.00
6677 A - 392 A	Ave	0.17	0.02	-0.05	0.00	0.00	0.00	-0.14	**1.30**	-0.08	-0.02	0.01	0.00	0.00
	Std	1.08	0.14	0.29	0.03	0.02	0.02	1.08	1.26	0.29	0.04	0.03	0.02	0.01

MQXA Average and Standard error [3]

Backup Slides: WISE tables

Table XI: Not allowed multipoles, average and spread over the 18 magnets, at four different currents (“Integral” data)

		b_3	b_4	b_5	b_7	b_8	b_9	a_3	a_4	a_5	a_6	a_7	a_8	a_9
669 A	Ave	-0.06	**0.10**	0.06	0.03	-0.02	0.00	-0.07	-0.16	-0.01	-0.01	-0.01	-0.03	0.00
	Stdev	0.66	0.14	0.12	0.04	0.01	0.01	1.12	0.58	0.19	0.08	0.04	0.03	0.01
5460 A	Ave	0.05	**0.10**	0.07	0.01	0.00	0.00	0.09	-0.10	0.02	-0.04	-0.01	-0.01	0.00
	Stdev	0.61	0.14	0.11	0.03	0.01	0.01	1.00	0.51	0.17	0.10	0.04	0.03	0.01
11347 A	Ave	0.06	**0.12**	0.09	0.01	-0.01	-0.01	0.08	-0.08	0.02	-0.08	0.00	-0.01	0.01
	Stdev	0.61	0.14	0.13	0.04	0.01	0.01	0.99	0.52	0.18	0.19	0.04	0.03	0.01
11925 A	Ave	0.09	**0.12**	0.09	0.01	-0.01	-0.01	0.08	-0.08	0.02	-0.07	-0.01	-0.02	0.00
	Stdev	0.64	0.14	0.13	0.03	0.01	0.01	1.00	0.52	0.18	0.20	0.04	0.03	0.01
669 A - 11347 A	Ave	-0.13	-0.02	-0.03	0.02	-0.01	0.01	-0.15	-0.08	-0.03	0.06	-0.01	-0.02	0.00
	Stdev	0.25	0.05	0.09	0.02	0.01	0.01	0.37	0.16	0.07	0.18	0.01	0.03	0.01

MQXB Average and Standard error [4]

Methods: Supervised Learning

Feature extraction (Correlation):

- The inputs are correlated with the outputs as expected, each different order error with its corresponding resonances

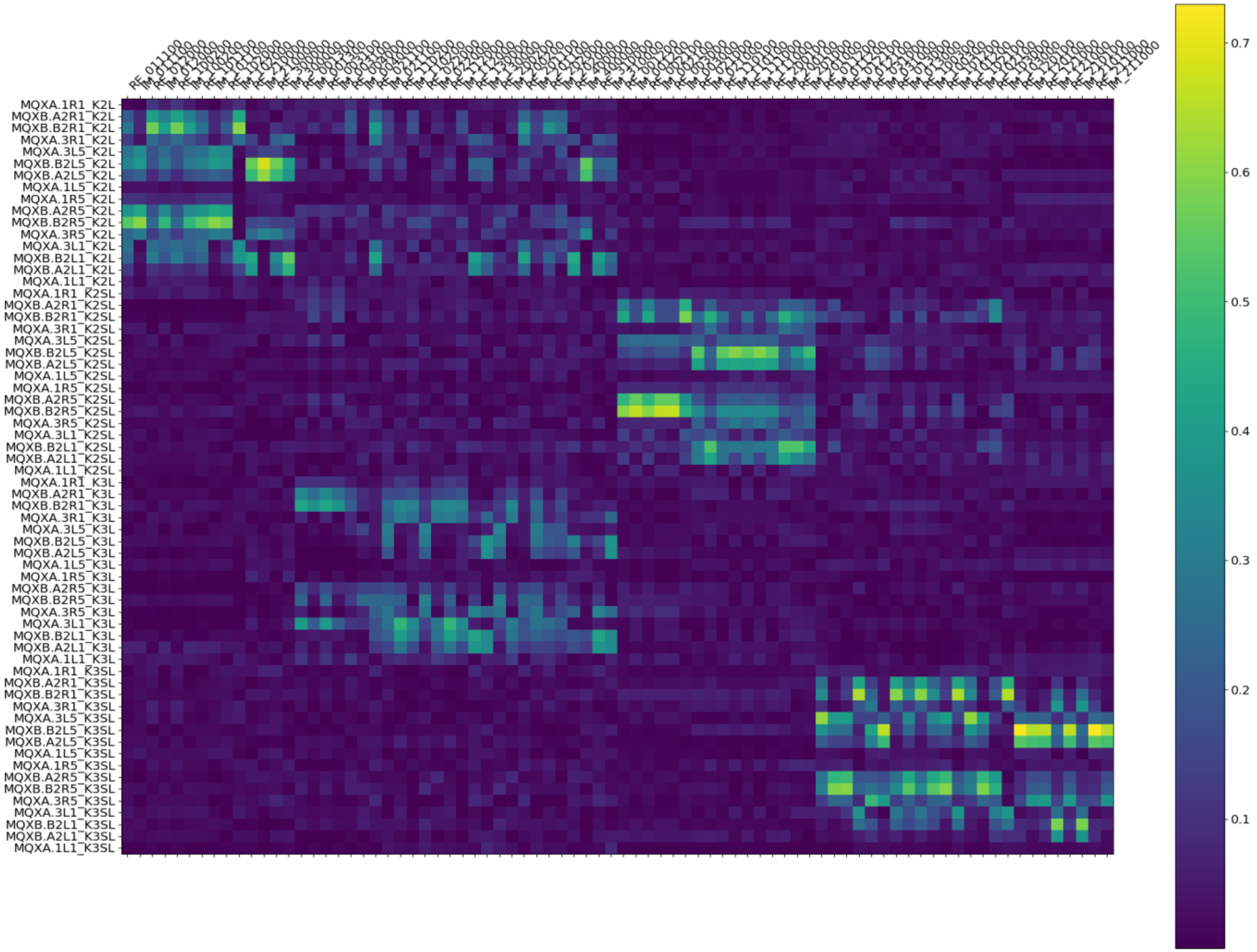


Fig 4. Correlation matrix for the RMS RDTs across the LHC with the error ***(Input vs Output)***

Backup slides: Supervised Learning

Feature extraction (RMS spectral line strength):

- Some RDTs are more present in the spectrum, this measurements should be less noisy

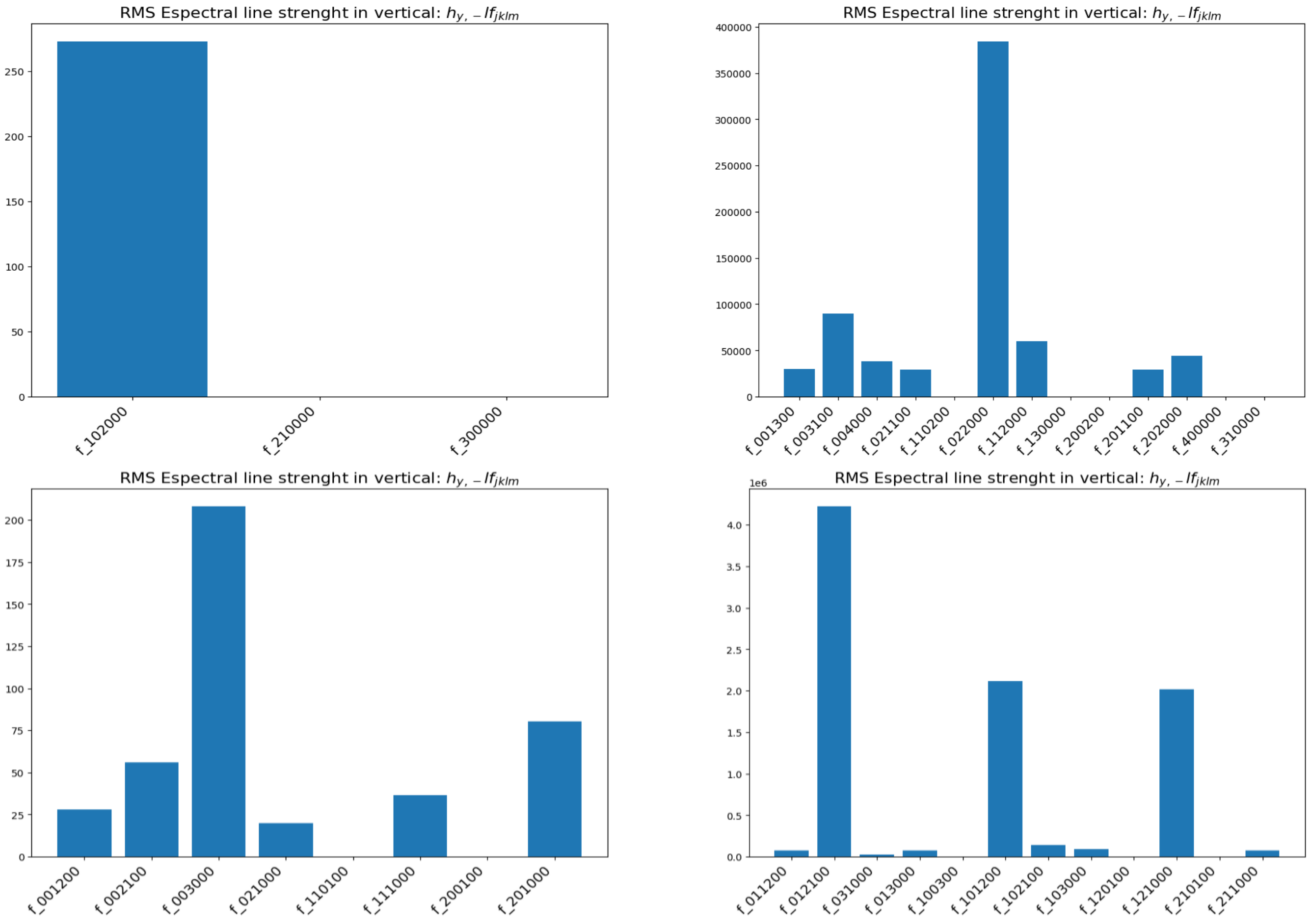


Fig 5. RMS spectral line strength (Vertical plane)

Backup slides: ML Model

Using a 80/20 train to test data split

Error Type	Set	R2	MAE [NORM]
Sext	Train	0.997	0.030
	Test	0.997	0.033
S Sext	Train	0.999	0.0022
	Test	0.999	0.0024
Oct	Train	0.777	0.343
	Test	0.728	0.379
S Oct	Train	0.809	0.282
	Test	0.770	0.309
ALL	Train	0.904	0.157
	Test	0.883	0.173

Tab 2. Best training results

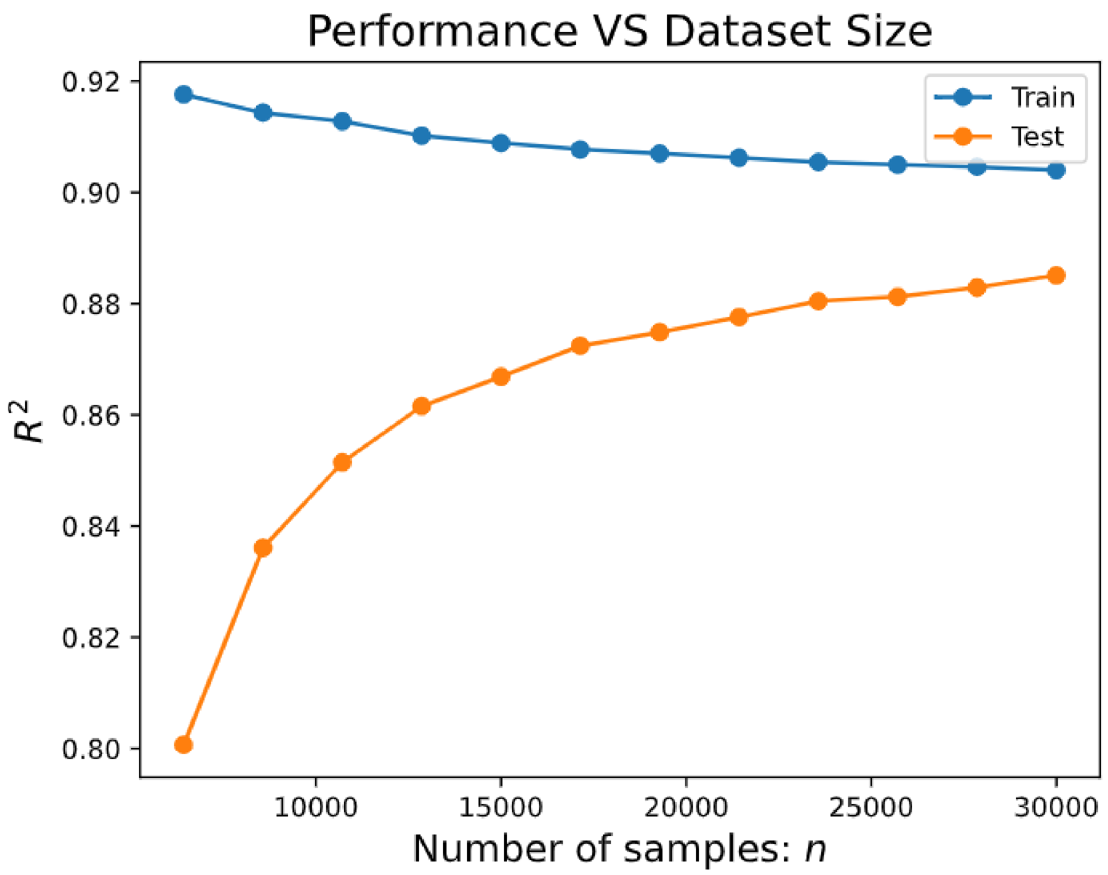
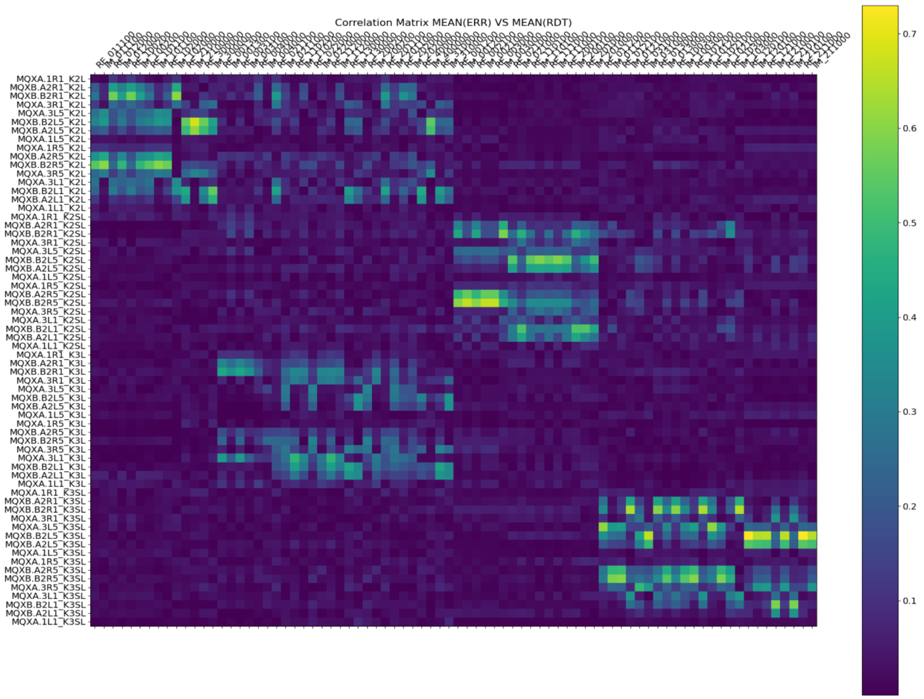
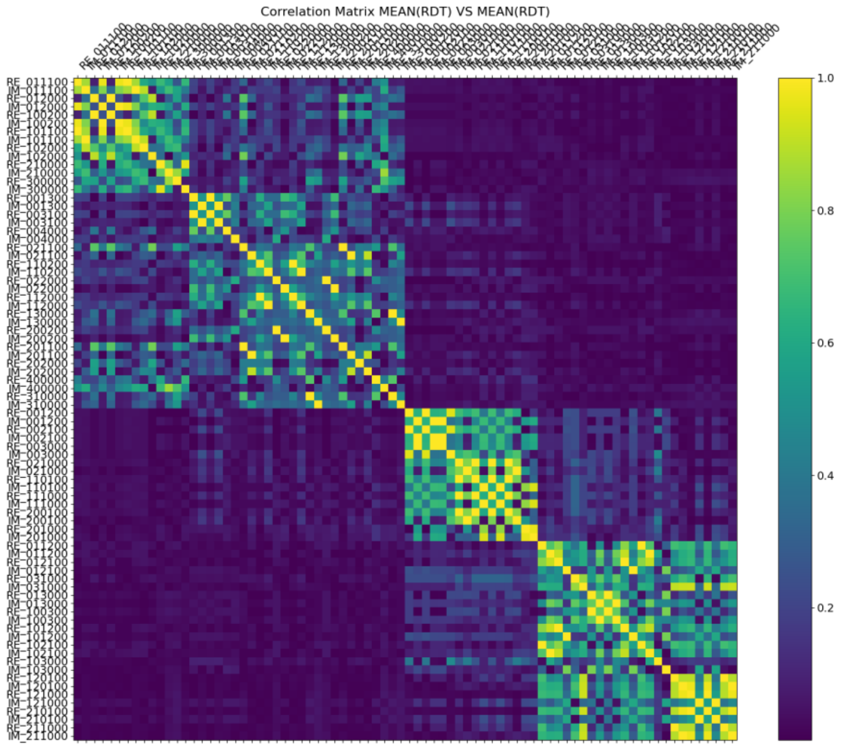


Fig 3. Performance VS dataset size

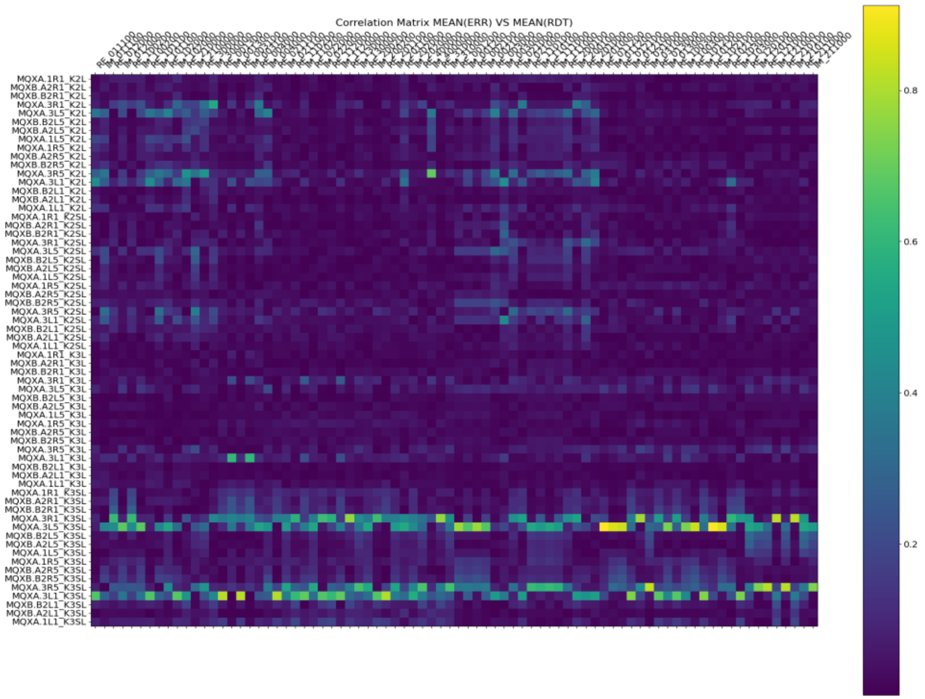
Backup Slides

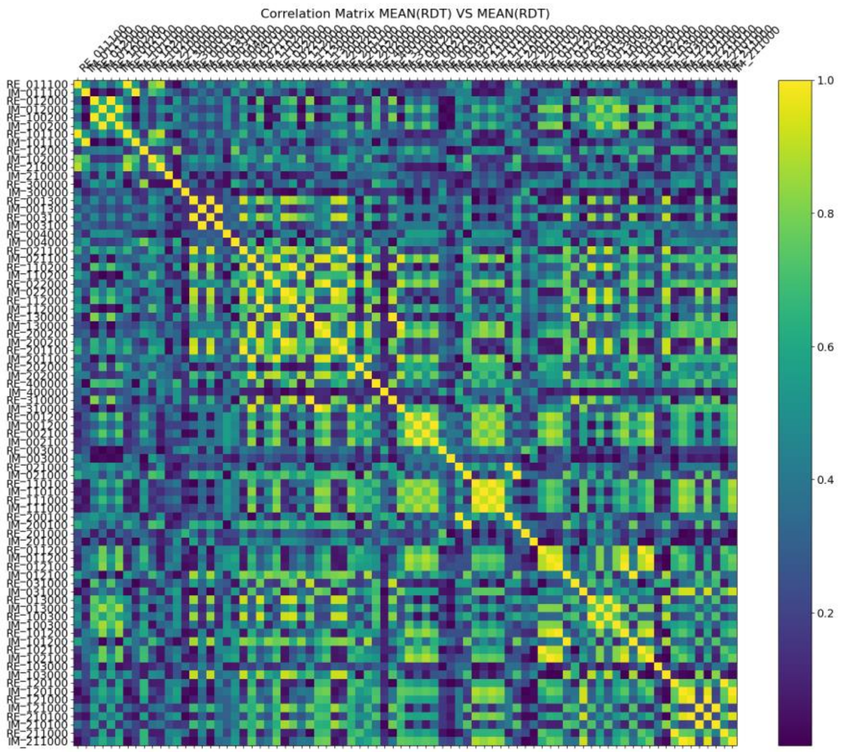




Correlation Matrix with no xing angle setup

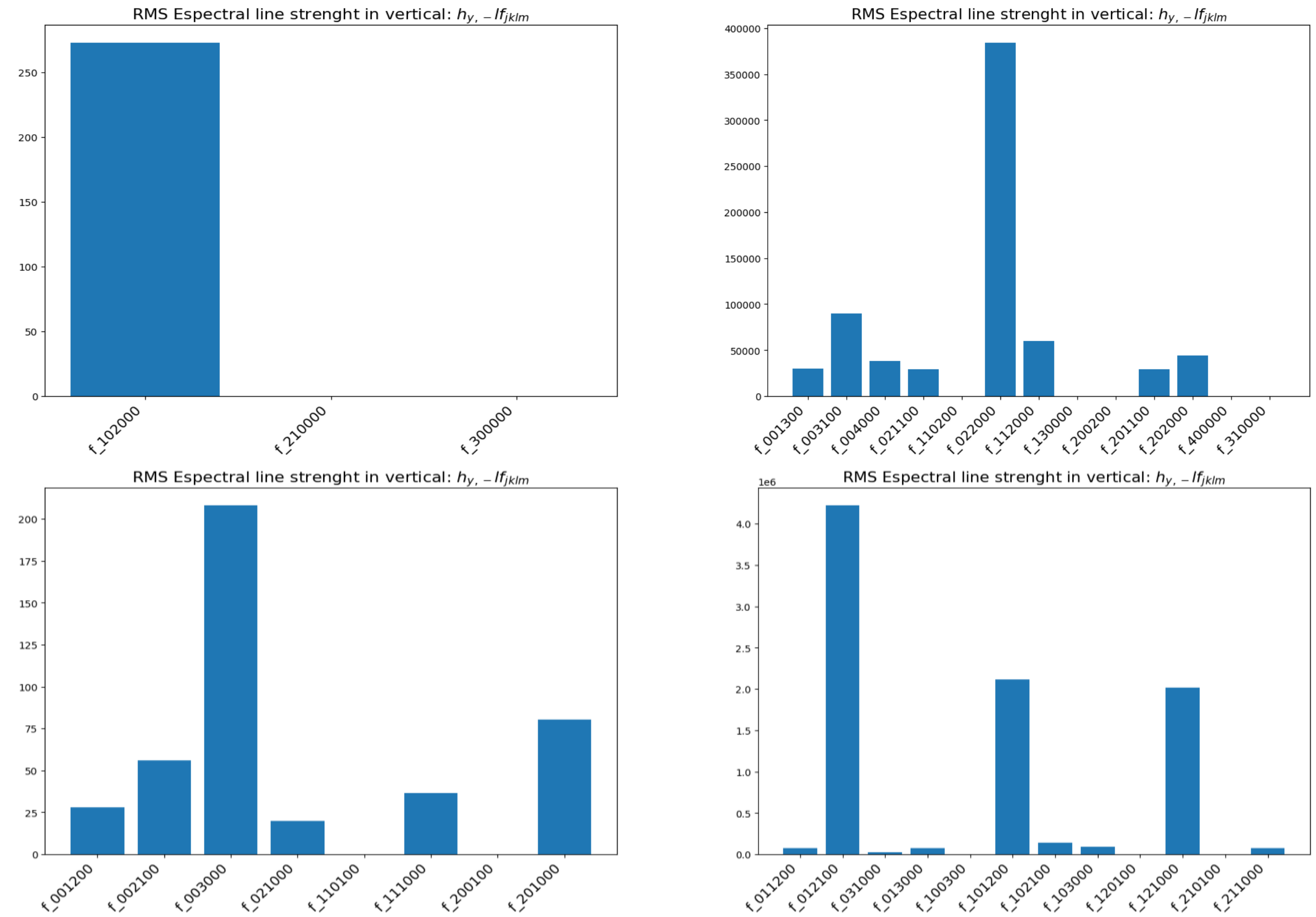
Backup Slides





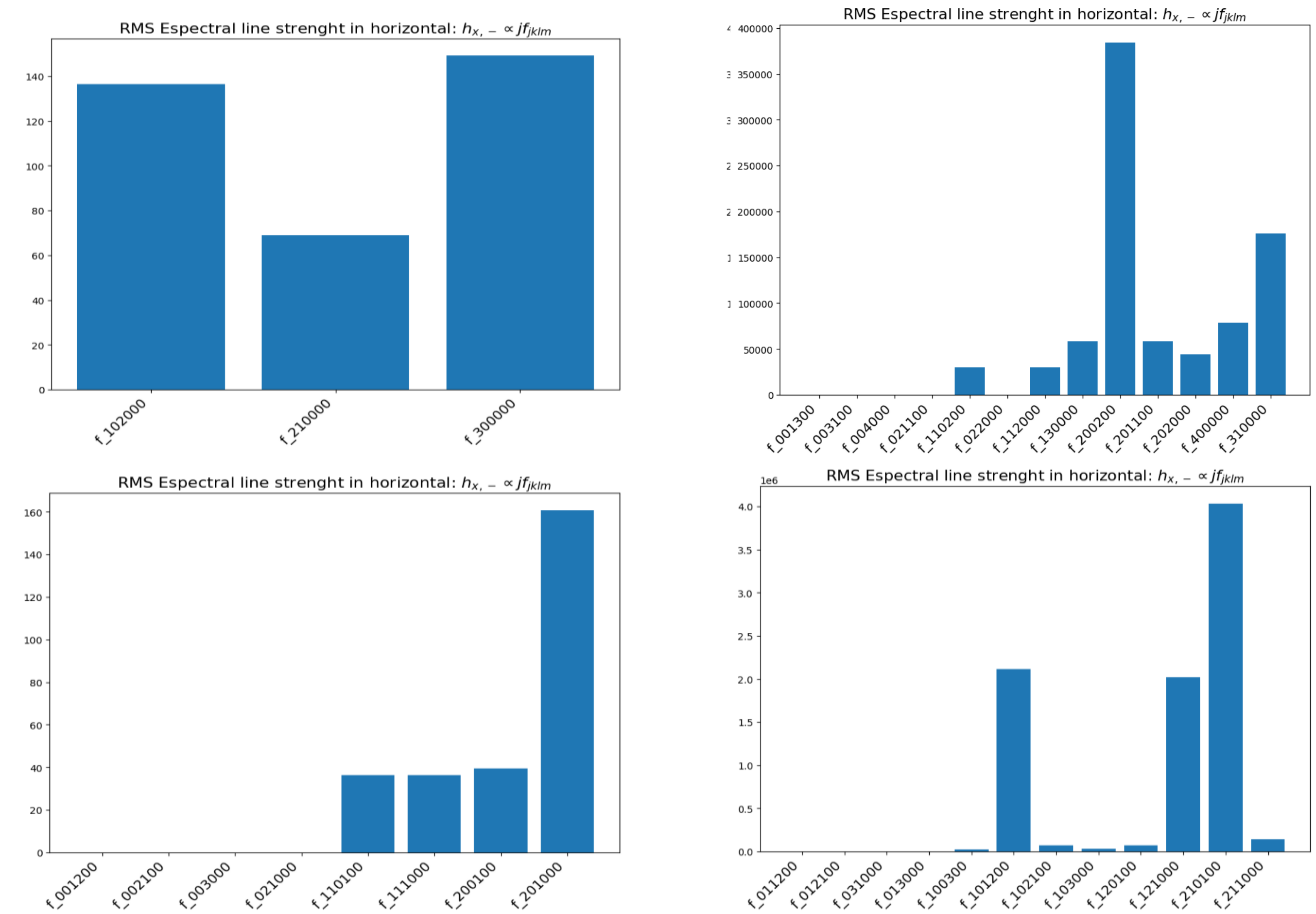
Correlation Matrix with xing angle setup

Backup Slides



RMS Spectral line strength in the vertical plane

Backup Slides

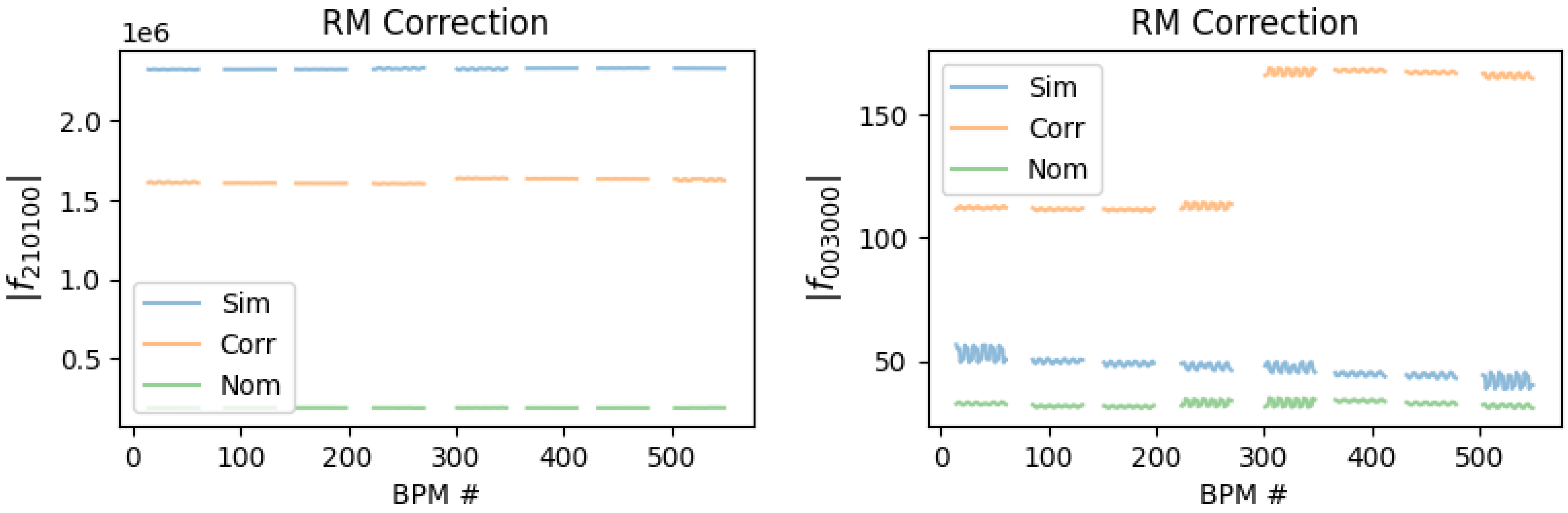


RMS Spectral line strength in the horizontal plane

Results: Performance with Xing angles

Response matrix:

Seems like due to feed down the response matrix approach treating all orders separately fails to correct RDTs



Example RDT Response matrix Correction with xing angles

CERN