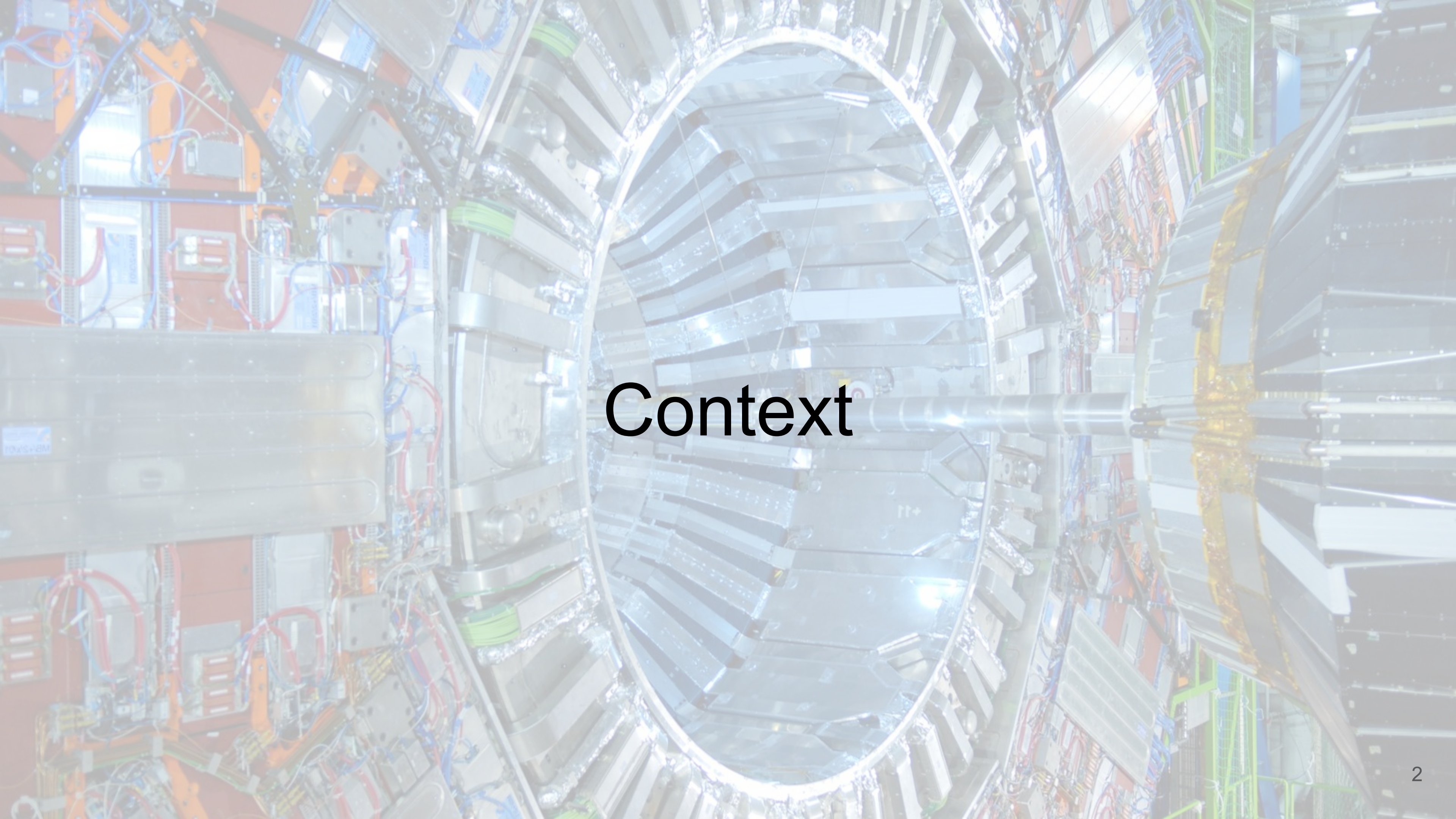
Can we optimize the operation of CERN's Large Hadron Collider with Machine Learning techniques?

Speaker: Loïc COYLE* (EPFL-CERN)
T. Pieloni (EPFL-LPAP) and B. Salvachua (LHC-Operation)
Acknowledgement: LHC OP, G. Azzopardi, M. Schenk, T. Tydecks

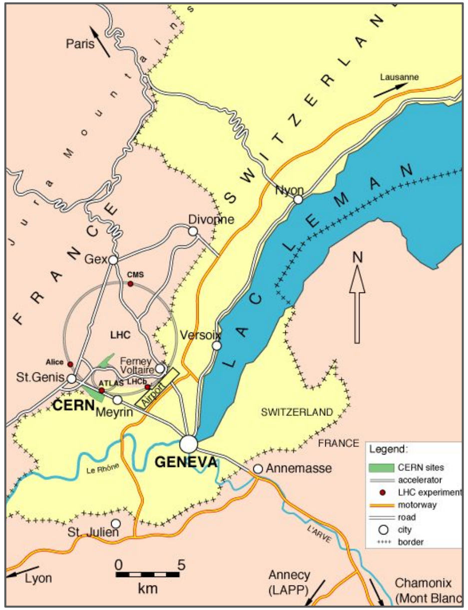
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

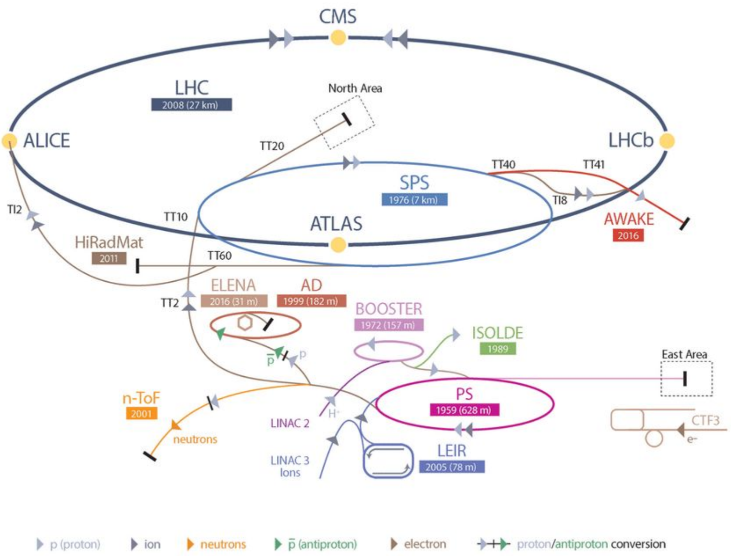
*loic.coyle@epfl.ch



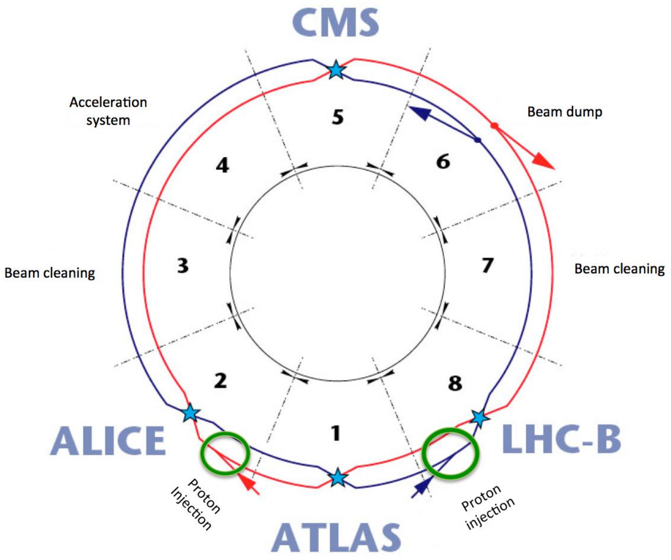
Context

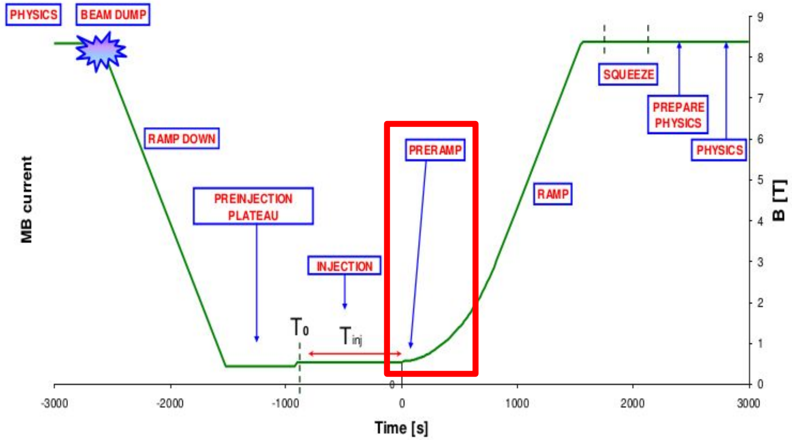
Context - CERN



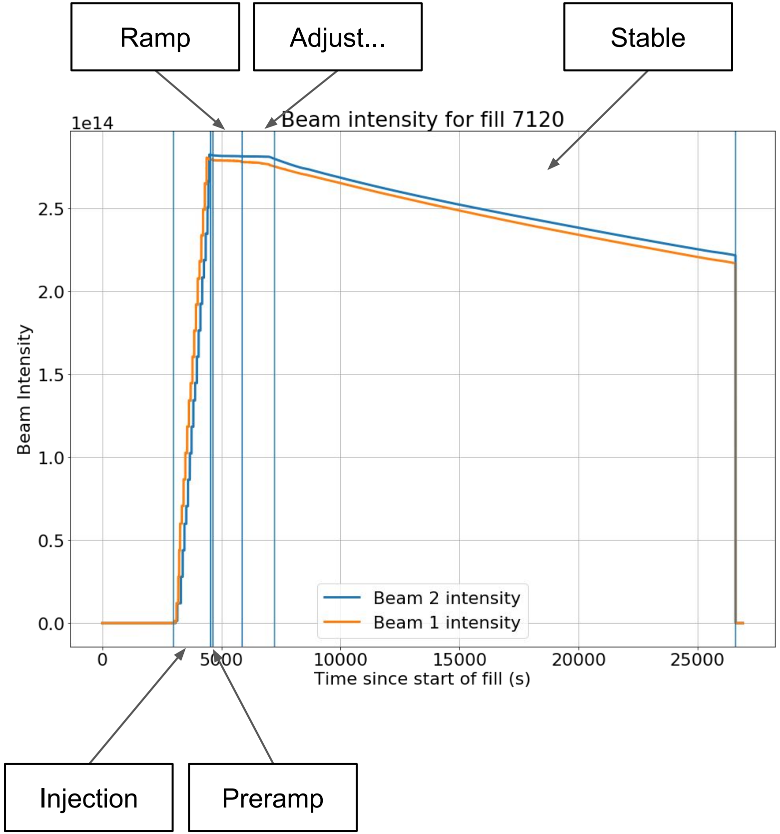


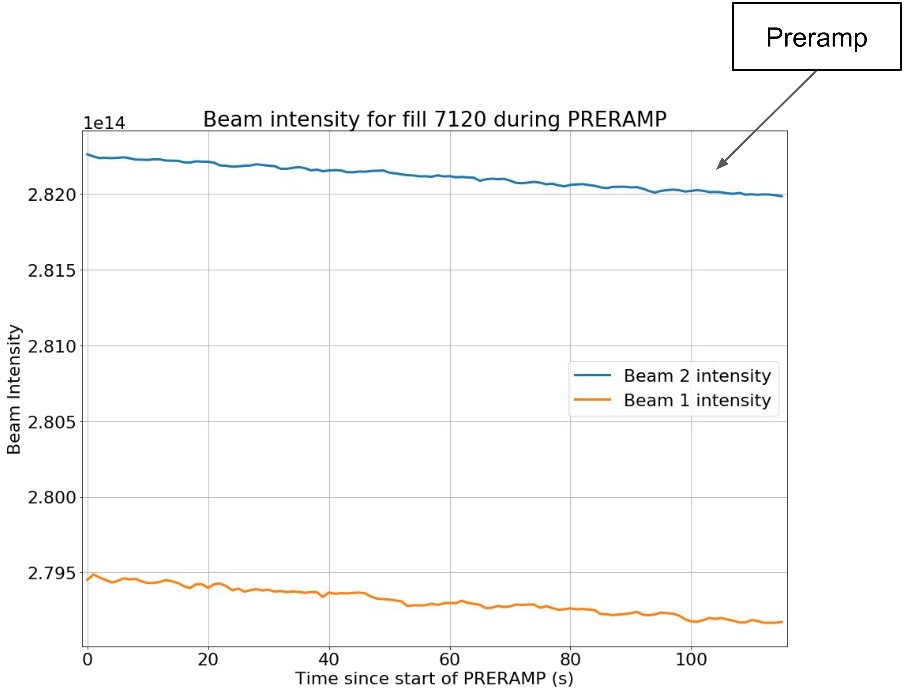
Context - LHC


CMS
Acceleration system
Beam dump
5
4
6
3
7
Beam cleaning
Beam cleaning
2
8
1
ALICE
LHC-B
ATLAS
Proton Injection
Proton injection

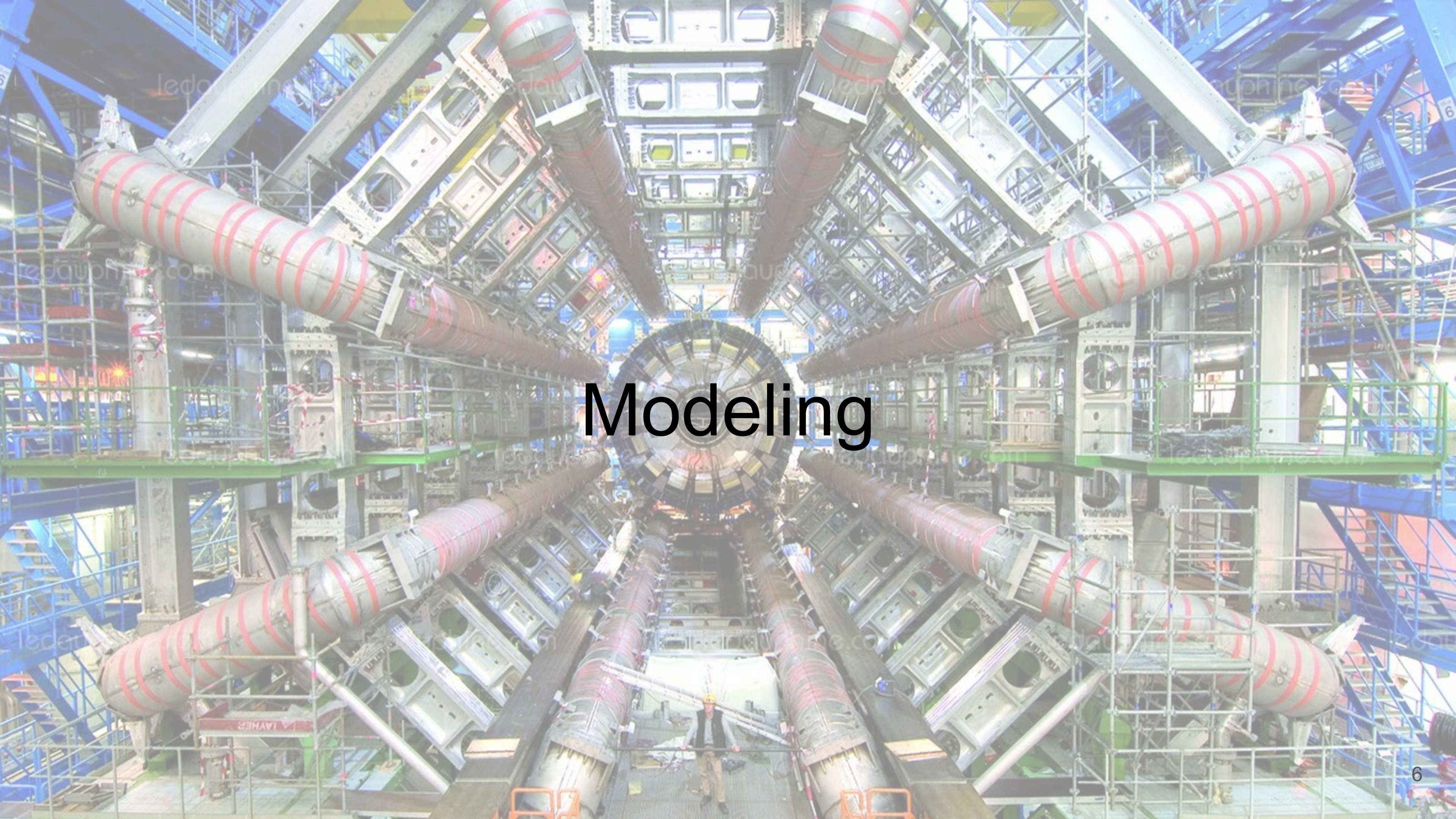

PHYSICS
BEAM DUMP
RAMP DOWN
PREINJECTION PLATEAU
INJECTION
PRERAMP
RAMP
SQUEEZE
PREPARE PHYSICS
PHYSICS
T_0
T_{inj}
MB current
B [T]
Time [s]
-3000
-2000
-1000
0
1000
2000
3000
0
1
2
3
4
5
6
7
8
9

Context - Physics: Lifetime

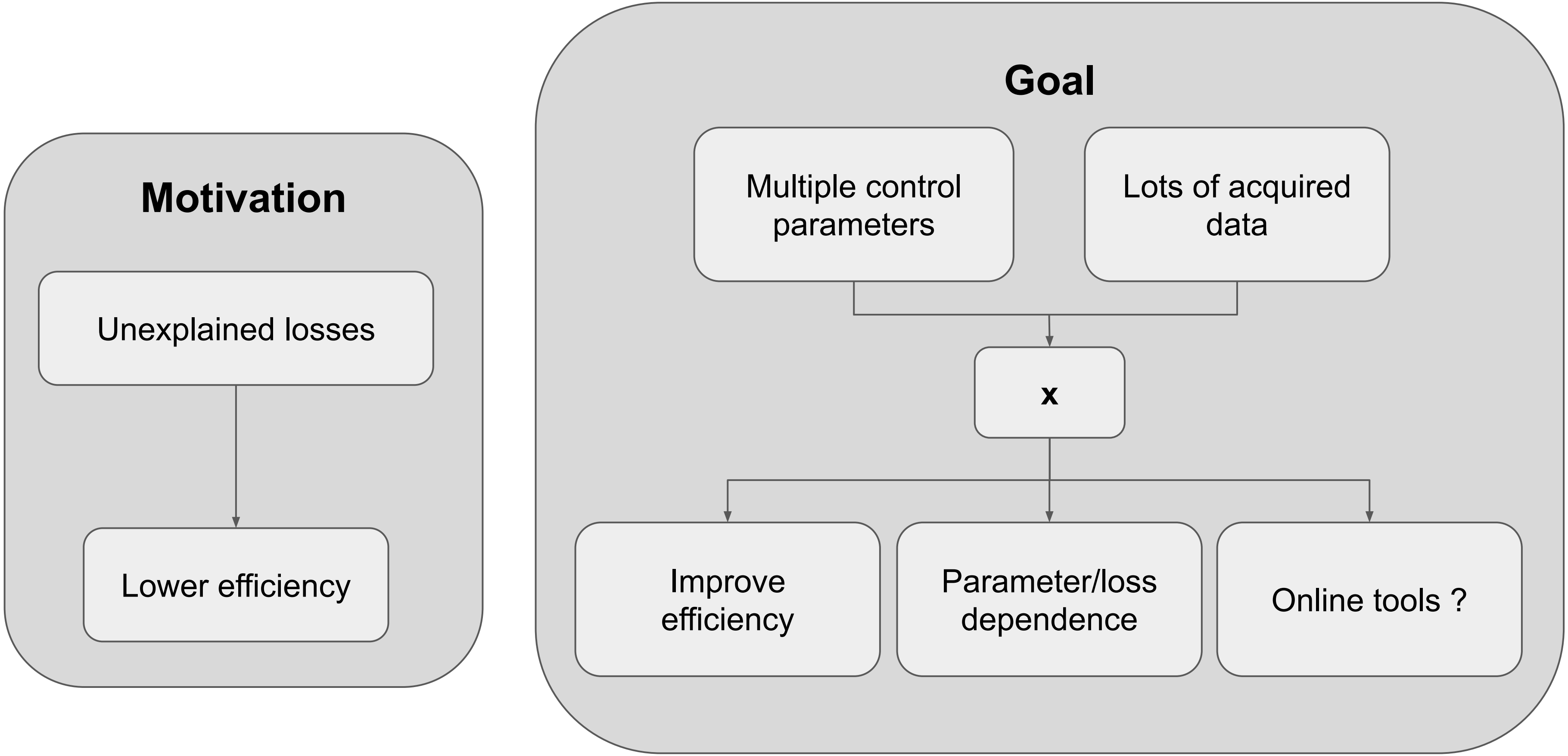




Lifetime = rate of loss of particles ~= slope of intensity



Modeling



(2 points) **Exercise:** find **x**.

Framing the problem

Reinforcement learning not really feasible.

Simulation:

- Particle tracking simulations extremely expensive.
- Simulations have trouble with modelling coherent instabilities.

→ Data driven supervised learning surrogate model.

Surrogate model of the beams intensity lifetime

Dataset:

- Tunes H/V/B1/B2
- Chromaticity H/V/B1/B2
- Octupole magnet current B1/B2

The operational knobs of the machine.

Target: Lifetimes B1/B2

Time span:

- 01-01 to 12-31 2017
- 01-01 to 11-01 2018

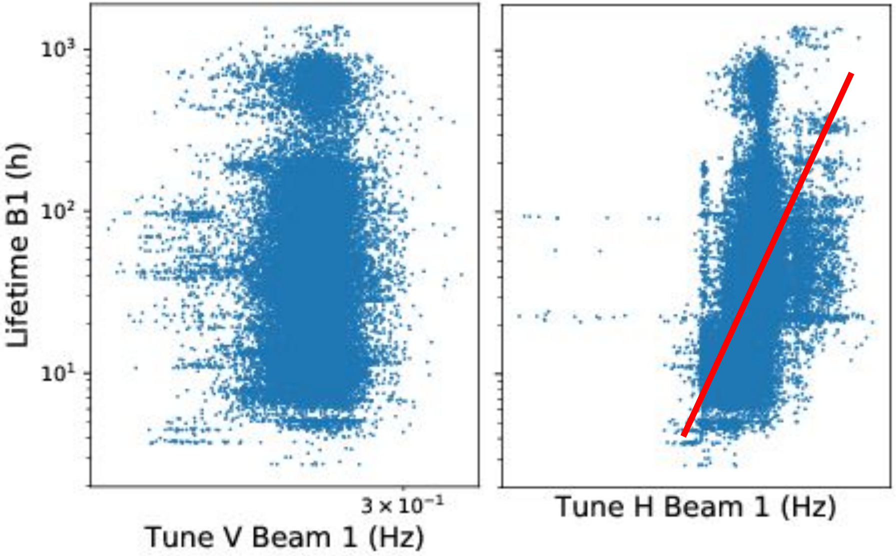
Dataset size: ~50000 points

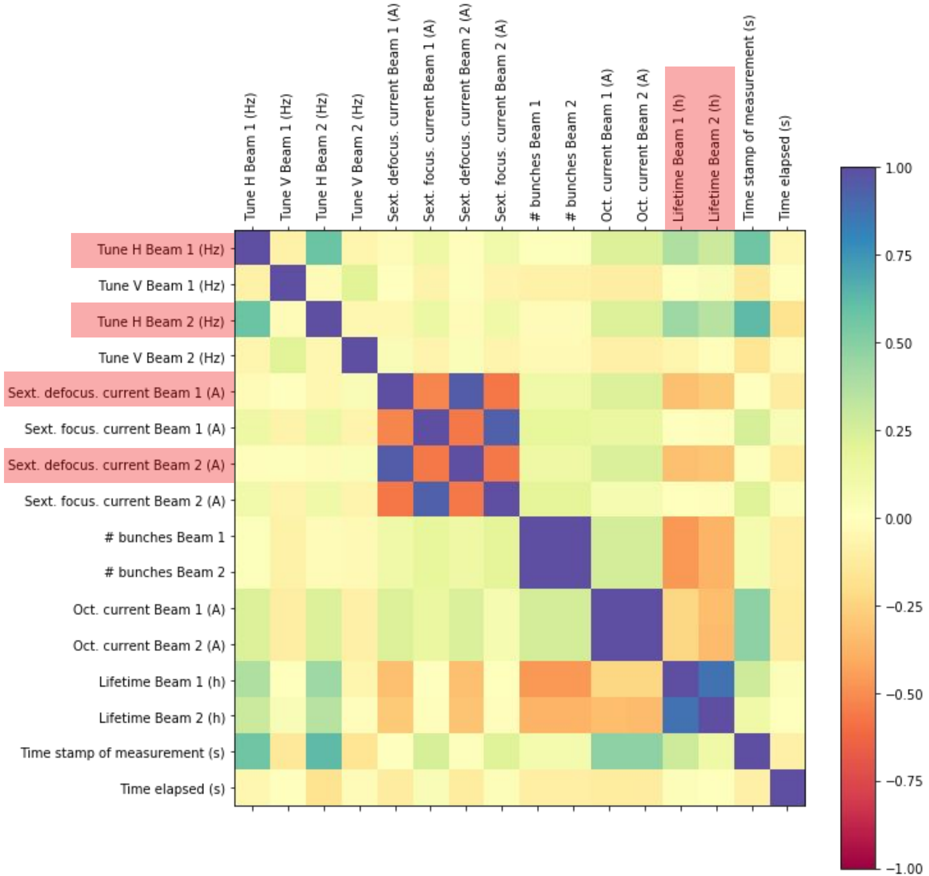
Data Visualization

Spearman correlation coefficient:
Measure of 'fitting' with monotonous function.

- H tunes of B1 & B2
- Sextupole currents

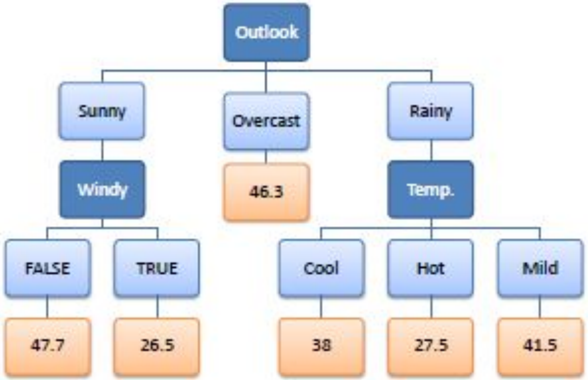
Cross beam dependence.

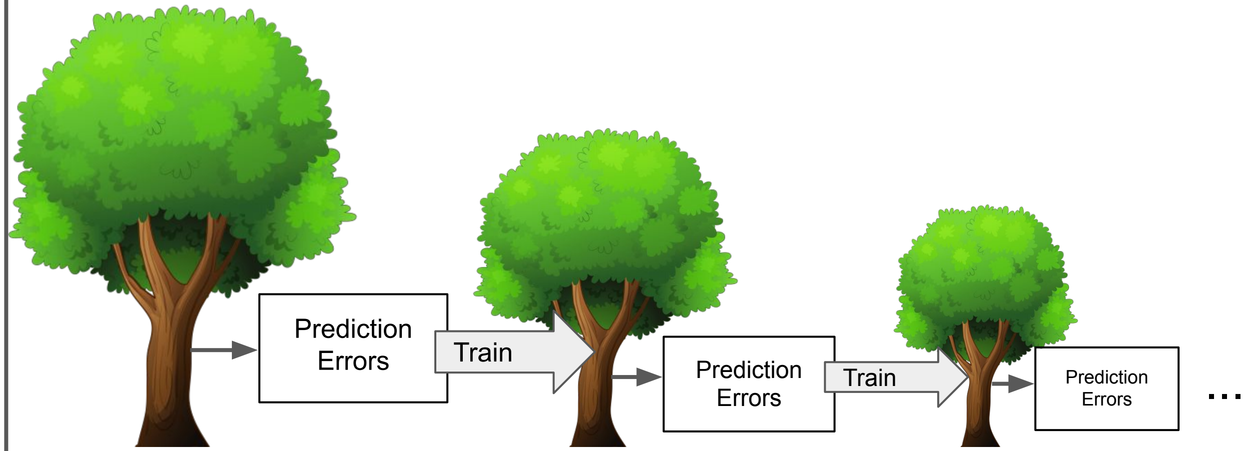




Gradient Boosted Decision Trees

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



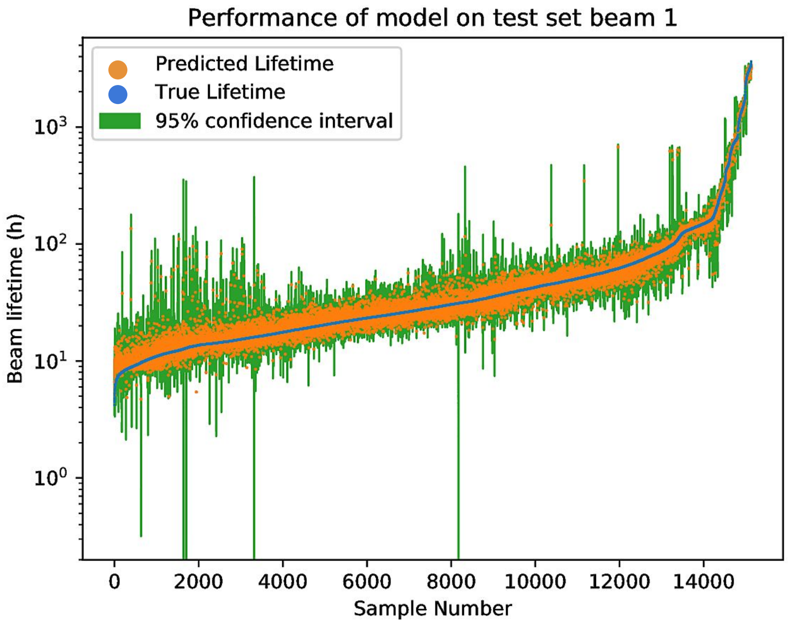


Available implementations:
https://github.com/Microsoft/LightGBM
https://github.com/dmlc/xgboost

Consistently in the top ~% of kaggle competitions

Model Performance with 2018 dataset

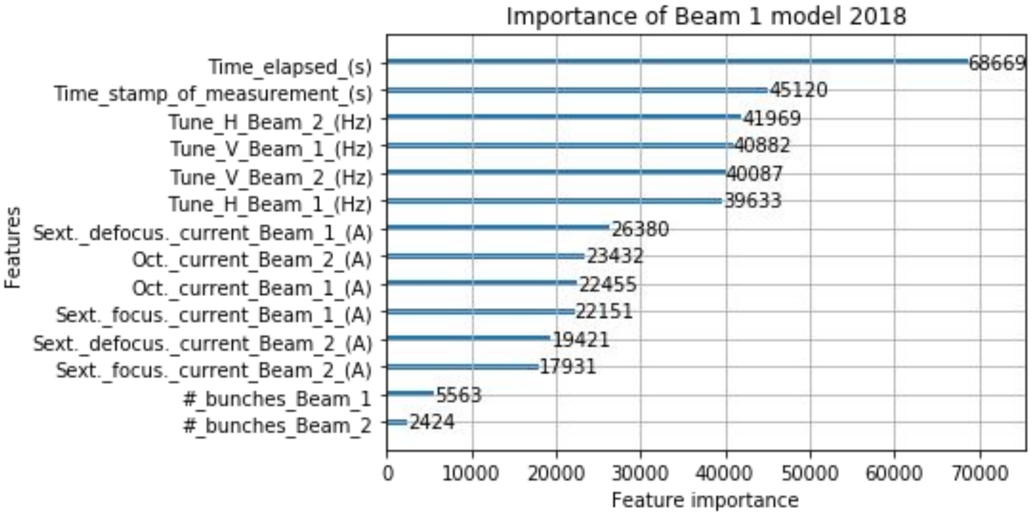
- Performance of trained model on **unseen data**.
- Model can **accurately predict the value of lifetime**.
- Presence of a few outliers
- **Bootstrapping**
  → train 100s of models on slightly different datasets.
  → 95% percentile from prediction distribution.



Feature Importance

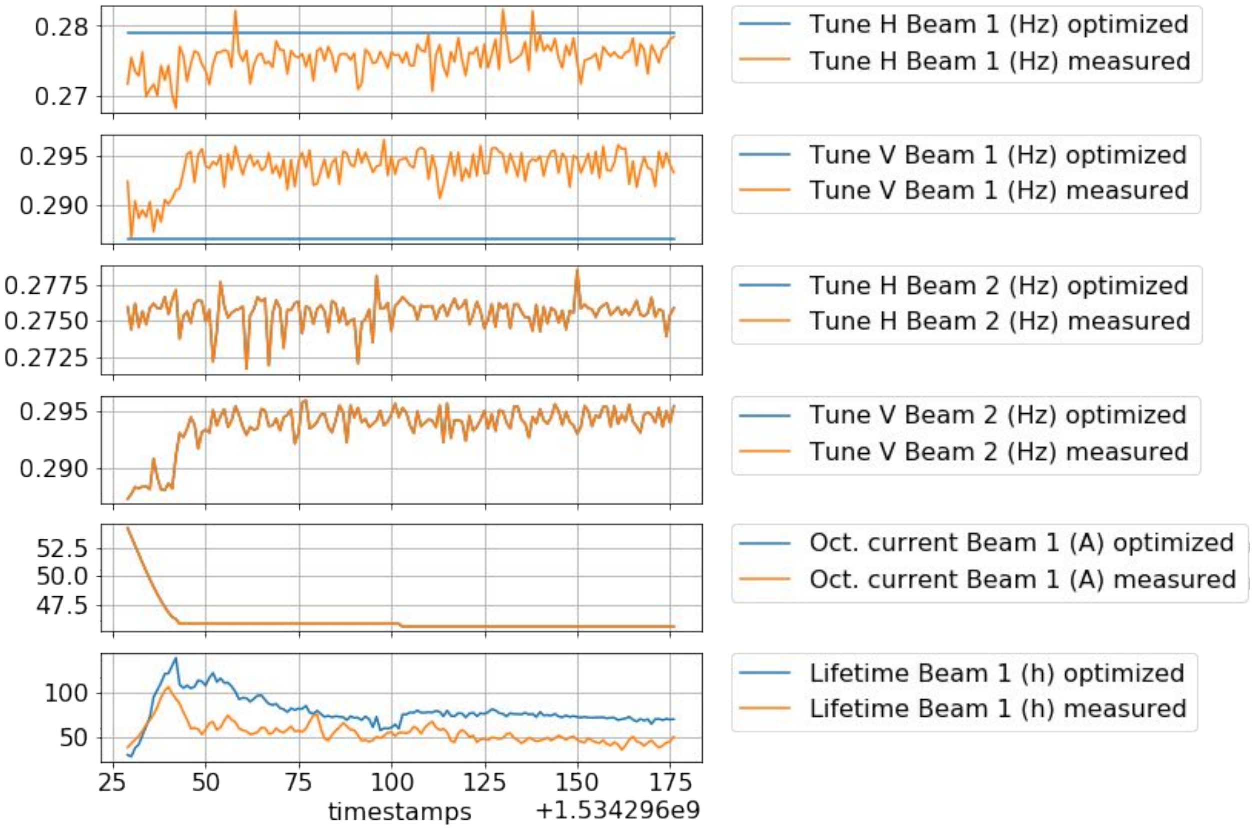
Feature importance: number of times the splits in the decision trees occur at each feature.

- **Time elapsed**: time dependent parameter substitution.
- **Time stamp**: winter blues ?
- Beam 2: cross beam correlations due to operation.



Optimization

Proof of concept : surrogate model for optimization



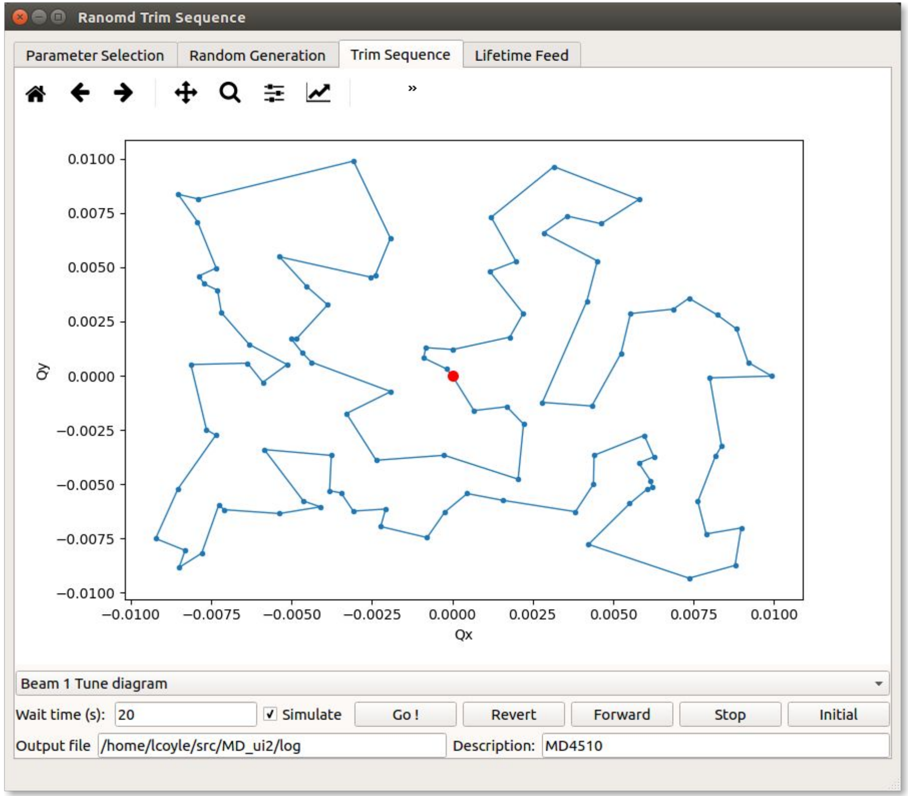
Optimization beam 1 tunes, prediction of beam 1 lifetimes.

Find the best beam 1 tune trim i.e. constant value.

Recommended settings:

- $Q_h = 0.279$
- $Q_v = 0.286$

Dedicated experiment: parameter scans



Generation and control of trim sequences :

- Parameter "random walk".
- Scans repeated with different machine settings.
- Independent scans for both beams and planes → decorelate.

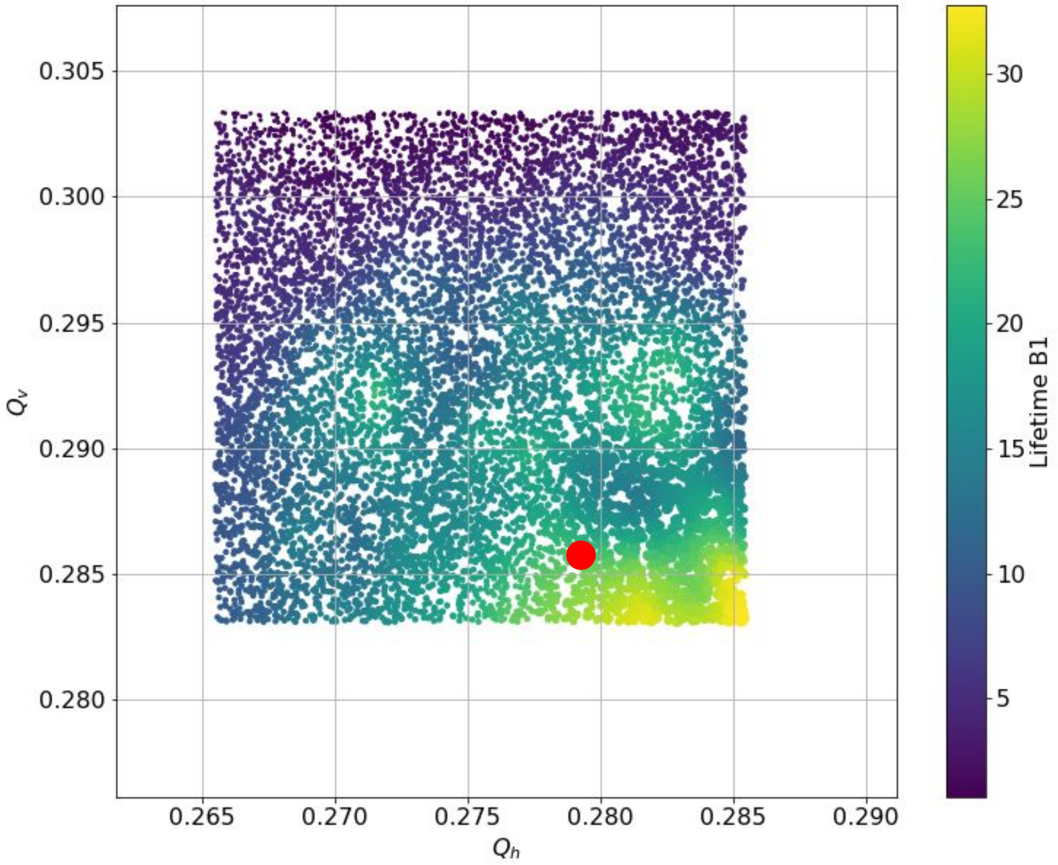
Recommended settings vs scan data

Recommended settings:

- $Q_h = 0.279$
- $Q_v = 0.286$

Seems to agree tentatively with MD data.

Good direction but a bit shy & falls short!



Not so simple...

High lifetime tunes

↓

Coherent instabilities

↓

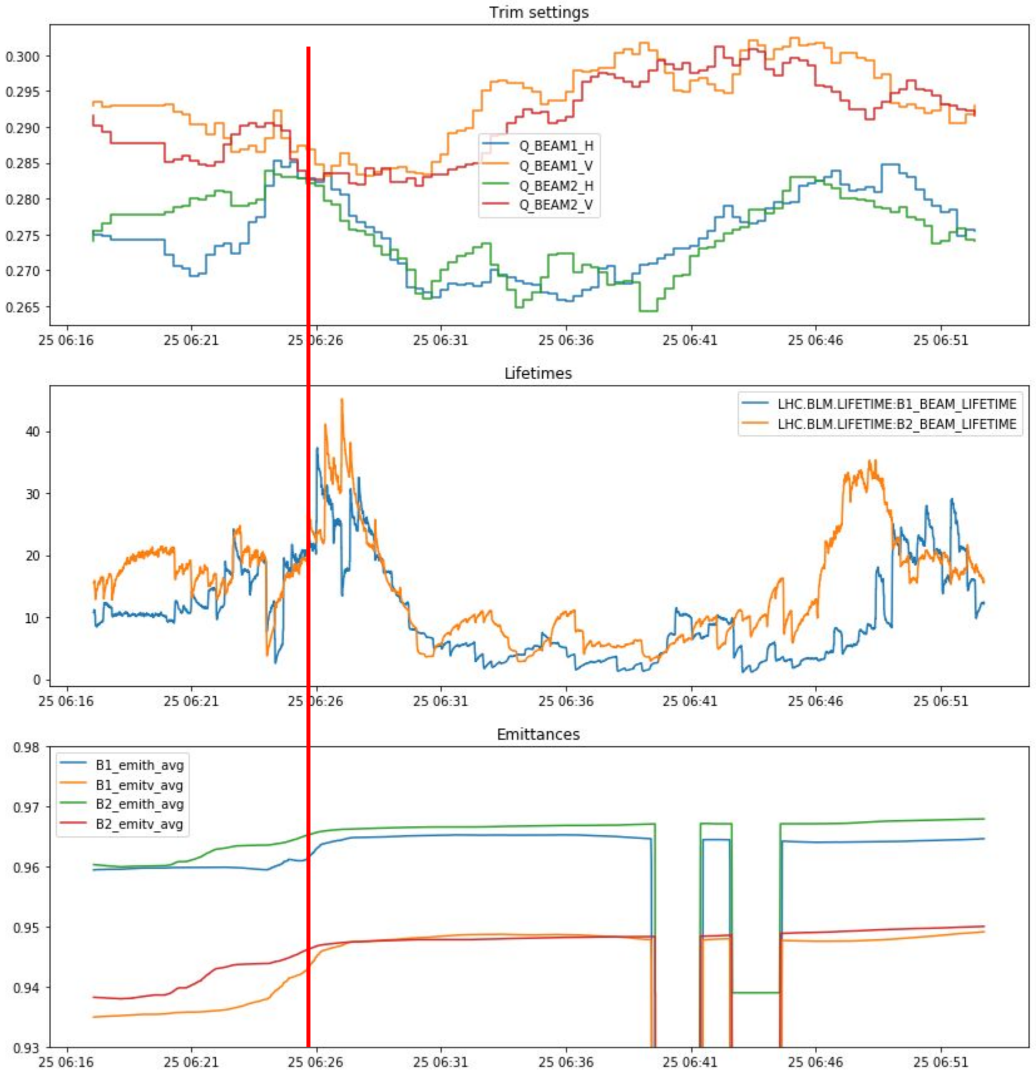
Increase beam size

↓

Decrease efficiency

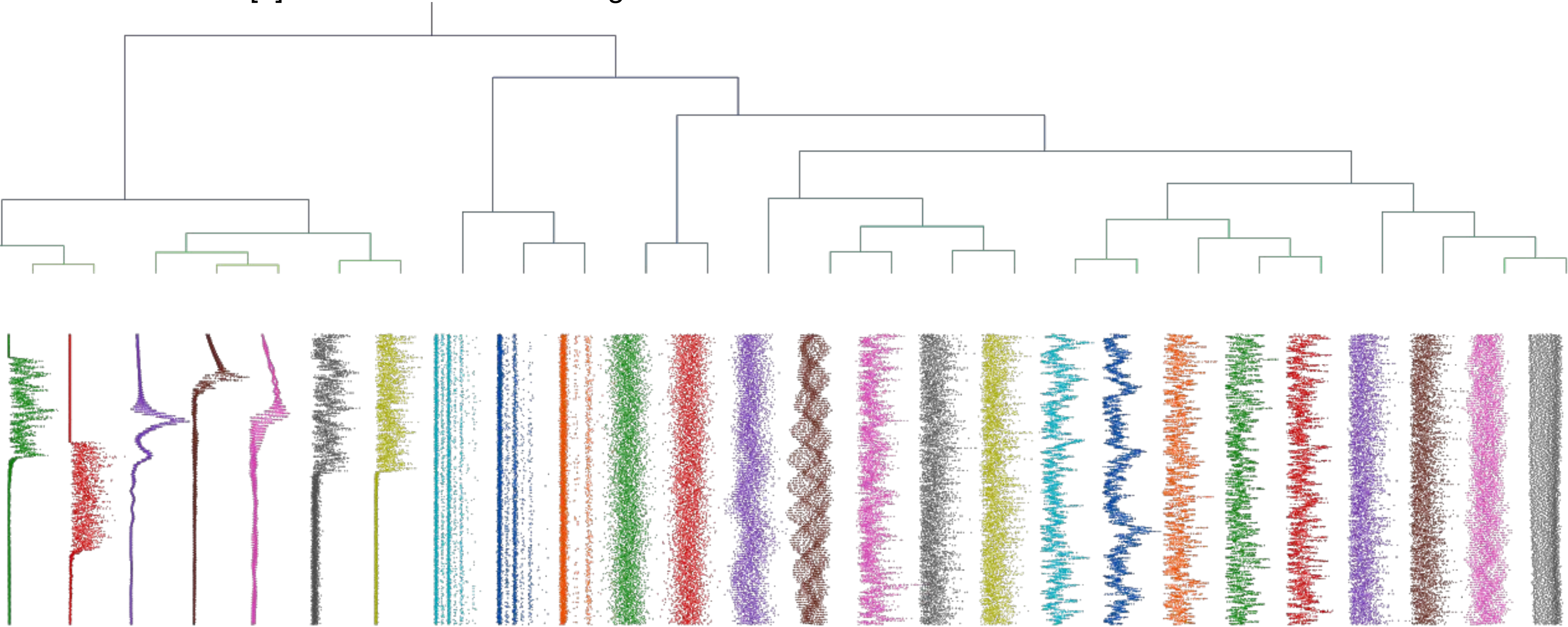
- Multi-objective optimization problem !

- Emittance not the best metric

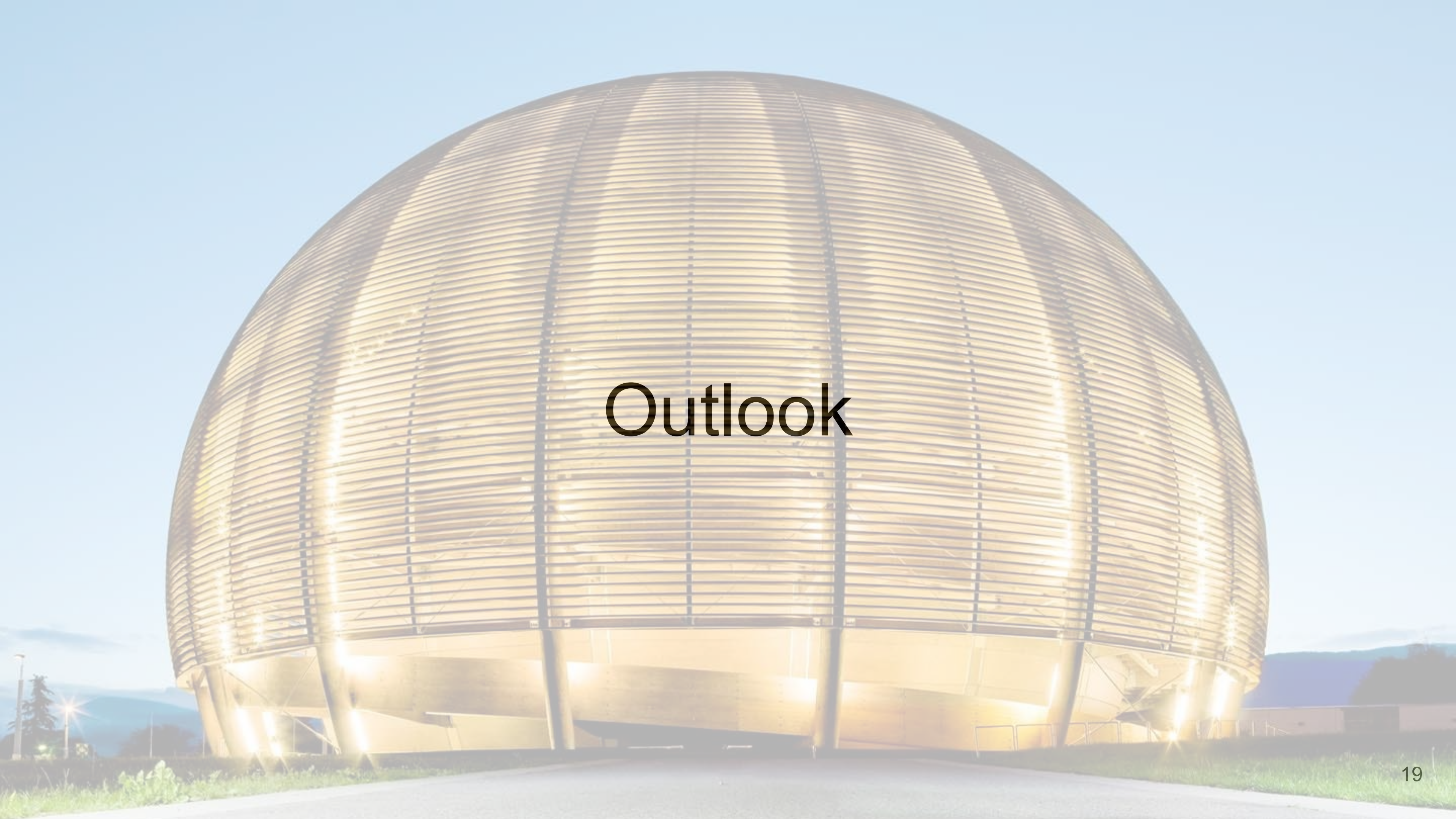


Including Coherent Instabilities

ObsBox: DTW[1] + Hierarchical Clustering

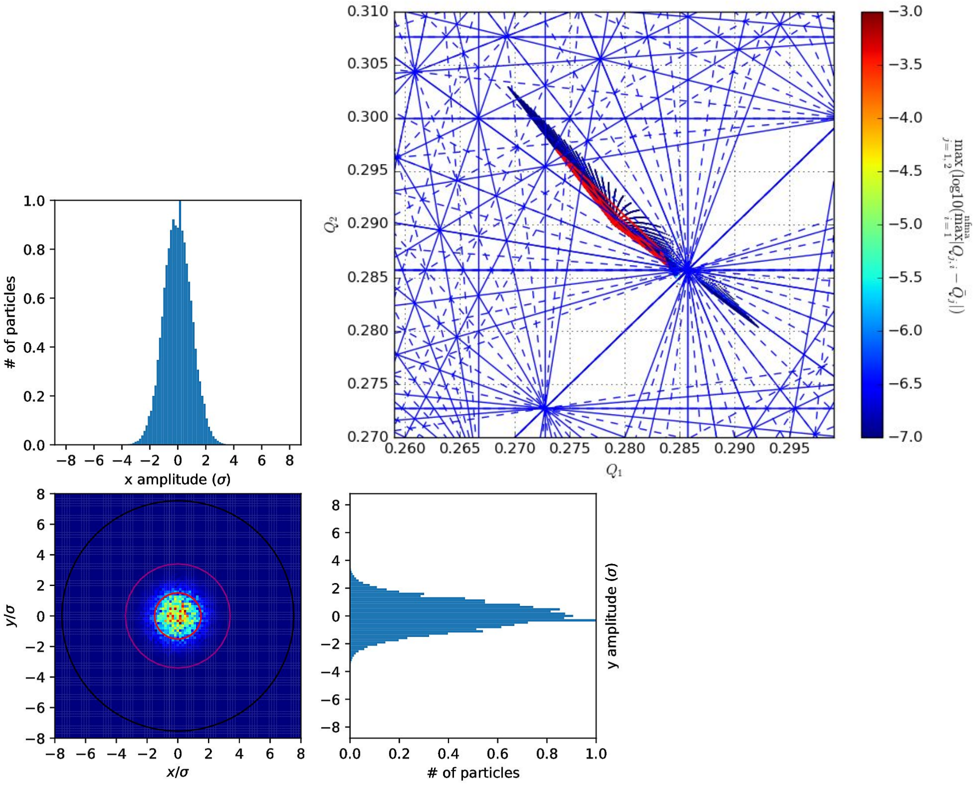


[1] https://github.com/slaypni/fastdtw or https://github.com/wannesm/dtaidistance



Outlook

Simulations

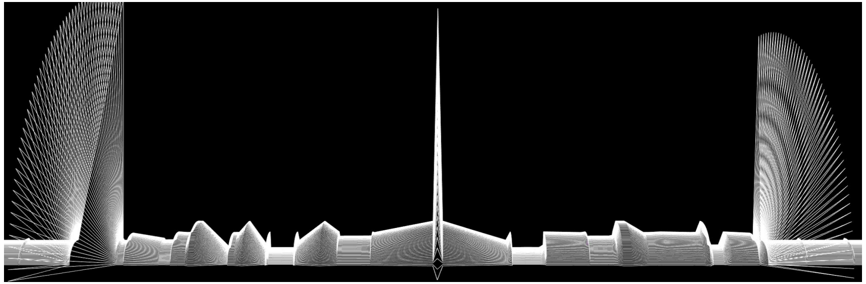


Validation of “optimized” working point using simulations.

Extension:

- Dataset from simulations
  → no parameter constraints !

Instantaneous ~(tracking simulations) from trained models → optimal settings



Conclusion

Can we optimize the operation of CERN's Large Hadron Collider with Machine Learning techniques?

- ~Working proof of concept.
- Preprocessing and quality of data is crucial.
- Multi objective modelling.

(2 points) **Exercise:** find **x**.

x = data driven surrogate model + optimizer