ML techniques for accelerator analysis

F.M. Velotti, B.Goddard, V. Kain N. Madysa

Outline

1. Introduction
2. Surrogate models
   a. Crystal shadowing example
3. Image analysis for accelerator systems
   a. Auto Encoders for dump system check
4. Time series analysis
   a. Beam induced heating prediction based on data
   b. Hysteresis compensation for spill quality improvement
5. Conclusions

Introduction

- ➔ Machine Learning (ML) can help for some of our problems:
  - Find optimum parameters for simulations
  - Online simulations for operation
  - Time evolutions of parameters and their prediction
  - Anomalous behaviour of systems identification
  - Hysteresis prediction
  - Etc...
- ➔ Incredible proliferation of ML models => easily accessible and more and more physics-oriented (or orientable)
  - Learn from other fields
- ➔ We have started to look into this and to apply these models to our research and daily operation

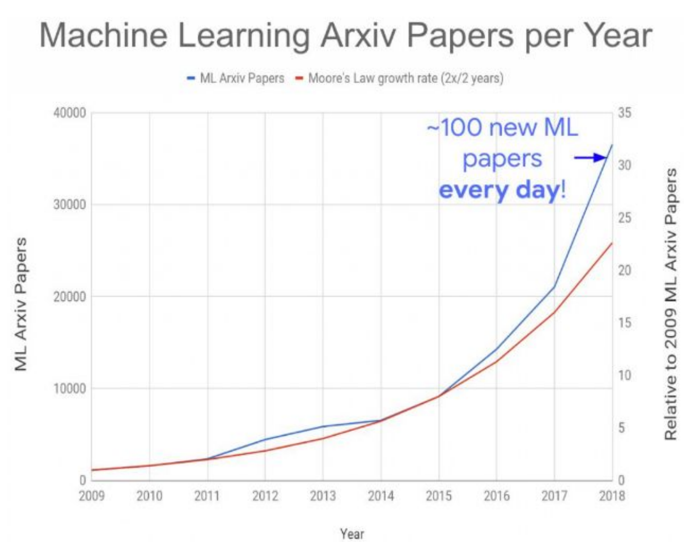
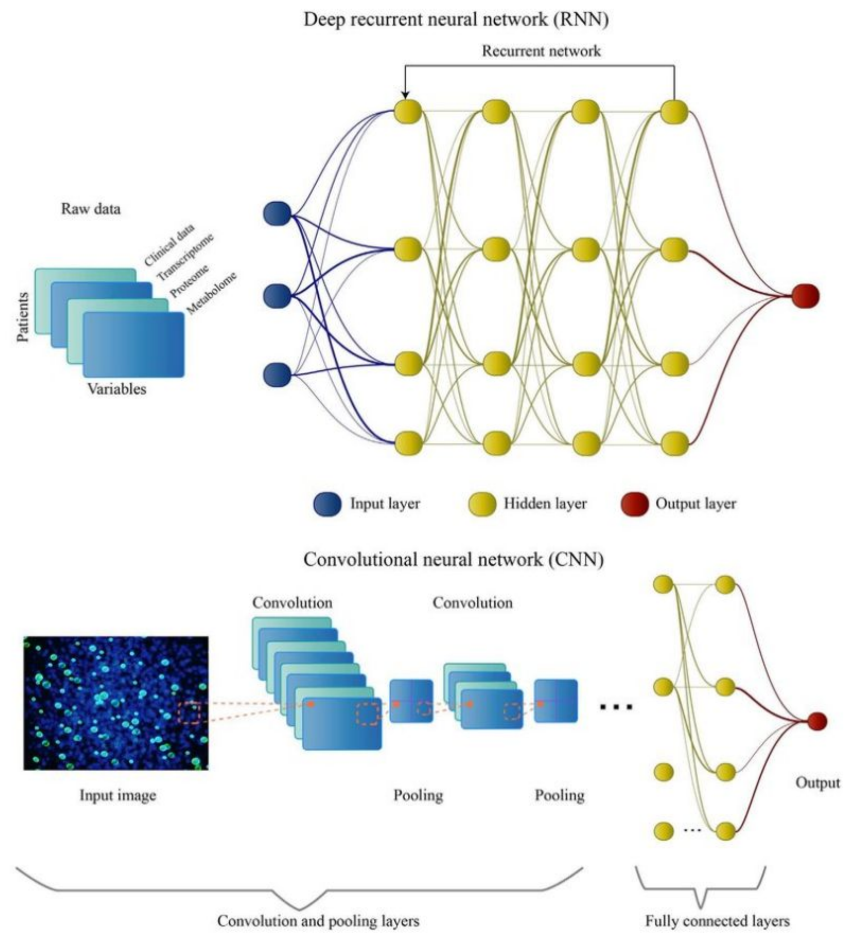


Image source

Introduction

- ➔ No time to go into details
- ➔ Reference and links added to the slides
- ➔ Maybe some terms that may or may not be familiar for all of you:
    - ◆ Deep Neural Networks (DNN): model composed of a series of different type of layers
    - ◆ These layers could be:
        - linear, convolutional layers (1D or 2D...or even 3D), Recurrent NN as Long Short Term Memory (LSTM) layers...



[Image source](#)

Outline

1. Introduction
2. Surrogate models
   a. Crystal shadowing example
3. Image analysis for accelerator systems
   a. Auto Encoders for dump system check
4. Time series analysis
   a. Beam induced heating prediction based on data
   b. Hysteresis compensation for spill quality improvement
5. Conclusions

Surrogate models

- Many cases where simulations (like SX multi-turn tracking simulations) take a while to run…
  - Test numerical optimisers (see Verena's talk)
  - Online prediction of machine behaviour…
- …or when taking data is not always possible…
- Surrogate ML-based models are a great solution
  - A "classic" solution is to take a bunch of simulations and then linearly interpolate between points => very inaccurate (in most cases) mainly when the parameter space is large
  - Train a DNN, random forest (or whatever other ML classifier or regressor) on available simulations => basically use a ML model to do the interpolation!
    - This is one of the simplest application of ML to accelerators

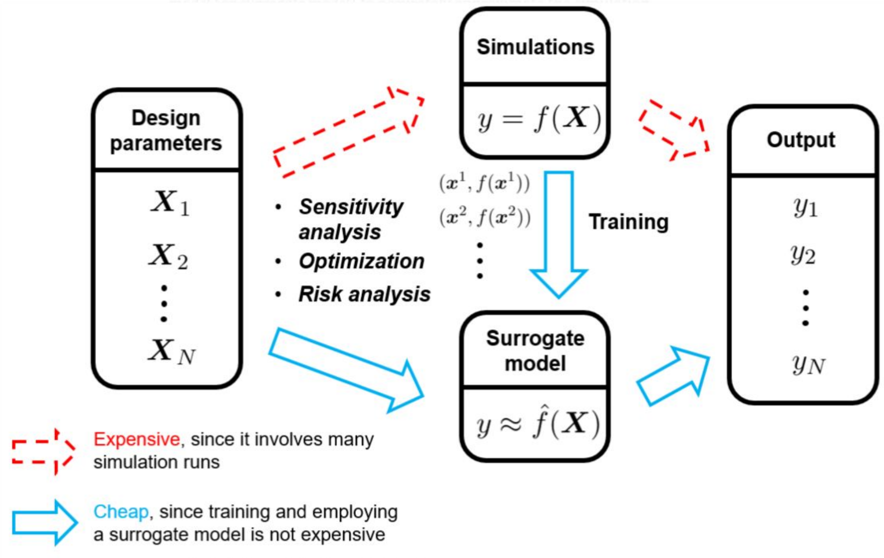


Image source

Surrogate model for crystal shadowing

➔ We trained a very simple Multi Layer Perceptron (MLP):
  - 2 dense layers of 256 neurons each…that's it!

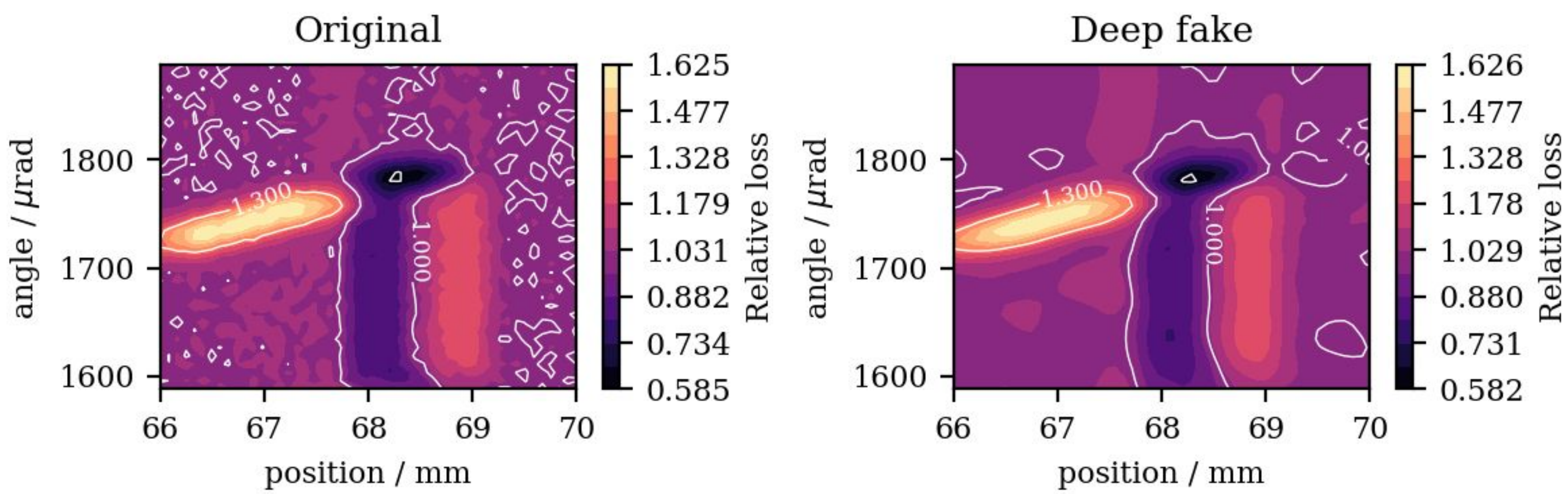
➔ Input:
  - Crystal angle and position, ZS angle

➔ Output:
  - Losses in extraction channel

➔ The reproduction of the response is excellent!

➔ Now instead of a few seconds, the simulation takes <1ms…

$$L_{ZS_{i,j,k}} = \mathcal{M}(\theta_i, x_j, ZS_k) \qquad \mathcal{H}$$

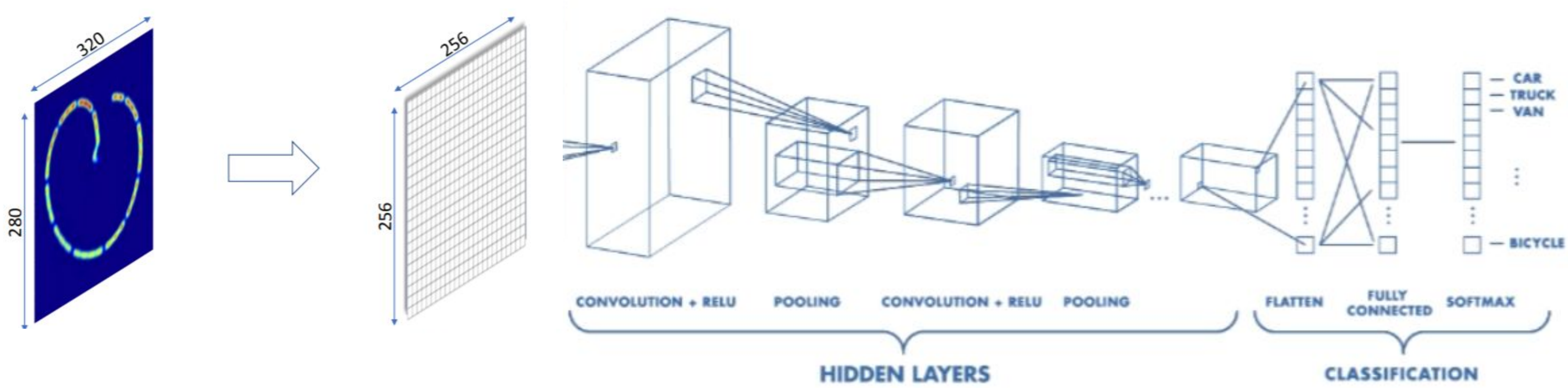


Outline

1. Introduction
2. Surrogate models
    a. Crystal shadowing example
3. Image analysis for accelerator systems
    a. Auto Encoders for dump system check
4. Time series analysis
    a. Beam induced heating prediction based on data
    b. Hysteresis compensation for spill quality improvement
5. Conclusions

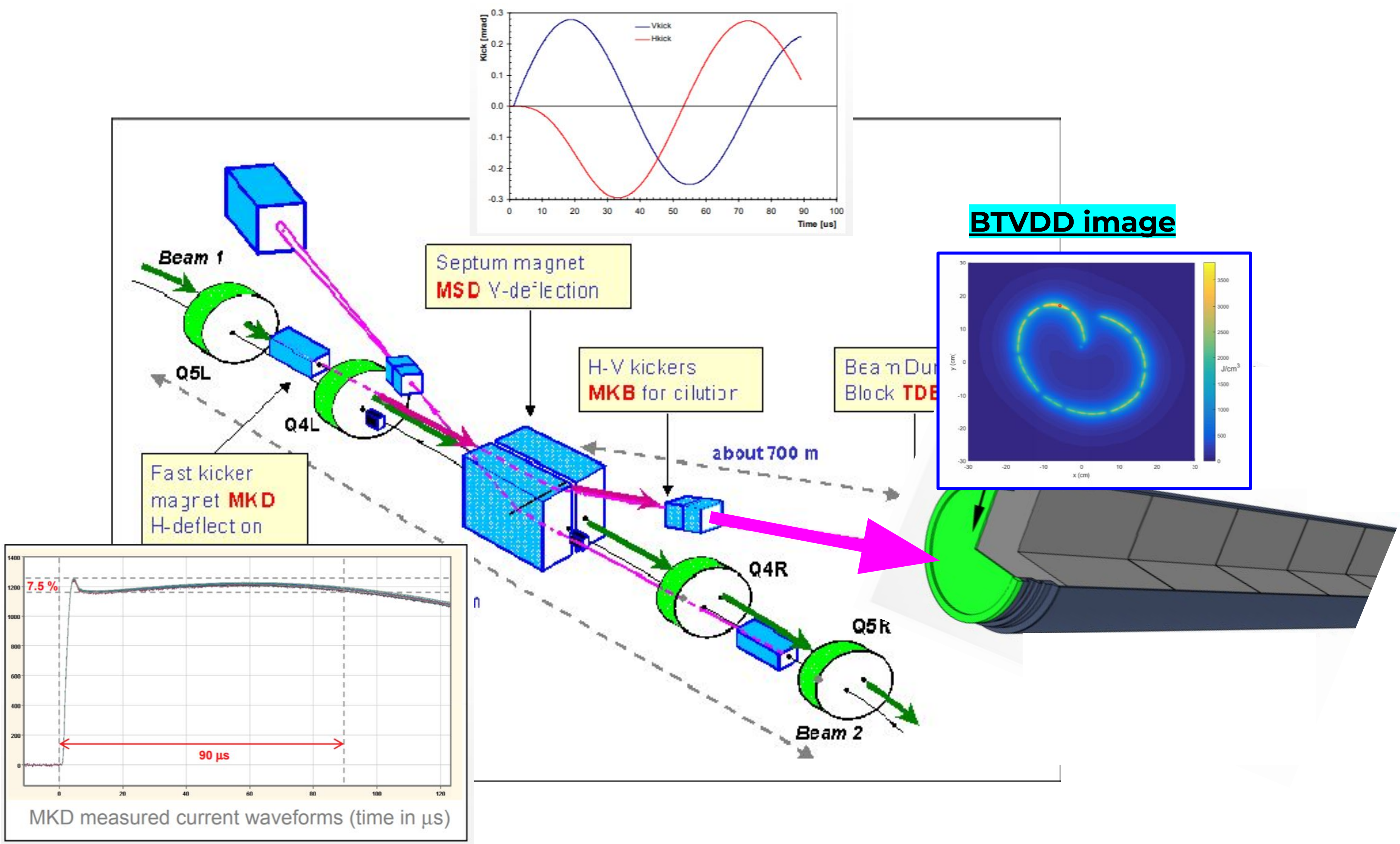
Image analysis for accelerator systems

- ➔ Many system status can be synthesised in one image:
  - Screen image of the beam
  - Time evolution of beam profiles (transverse or longitudinal)
  - Losses evolution in time...
- ➔ CNN have shown amazing capabilities in classifying images, extract features, translate image content into text
- ➔ We tried to use CNN to extract the state of the SPS and LHC dump system from screen images of the dumped beam



Just a little step back: LBDS

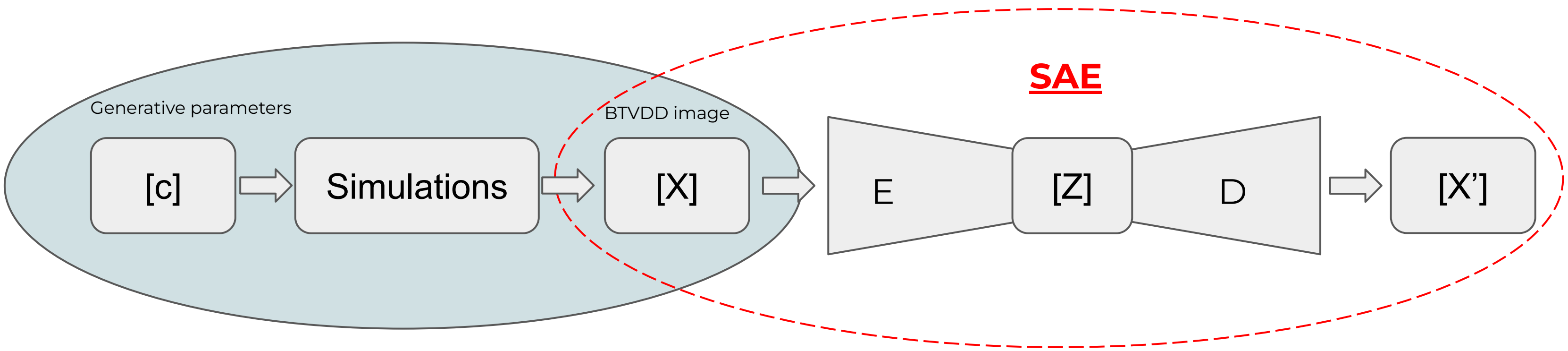
➔ The LHC beam dump system in a nutshell



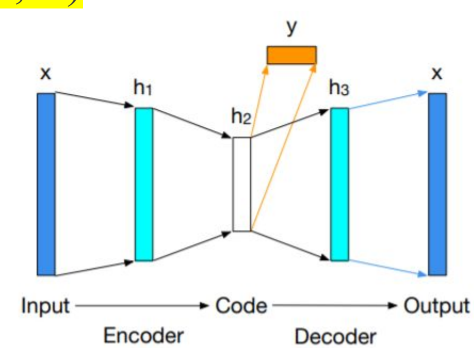
MKD measured current waveforms (time in µs)

Supervised Auto Encoder

- We trained an Auto Encoder (AE), either a Variational AE (VAE) or a Supervised AE (SAE)
- The difference is that instead of training and additional NN on latent space and generative factors (as done for the VAE), we can add a term for the loss function of the SAE to do all in once
  - The latent dimension now are the generative factors - idea taken from here [1]

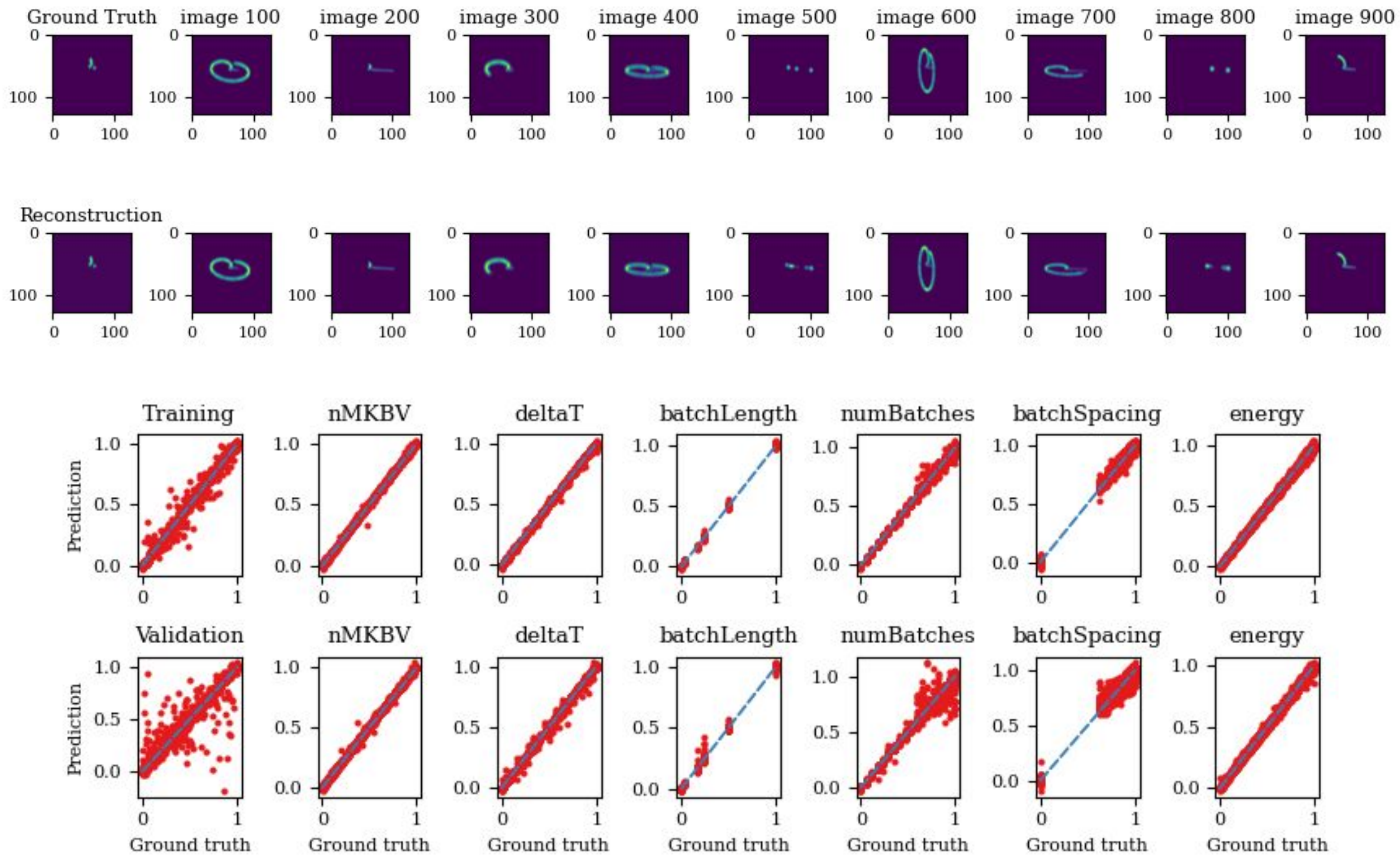


$$L_i(\theta,\phi) = -E_{z\sim q_\theta(z|x_i)}[\log_\phi(x_i|z)] + \underbrace{w_{KL}}_{=0}\, KL(q_\theta(z|x_i)\,||\,p(z)) + w_g\, MSE(c, Z)$$



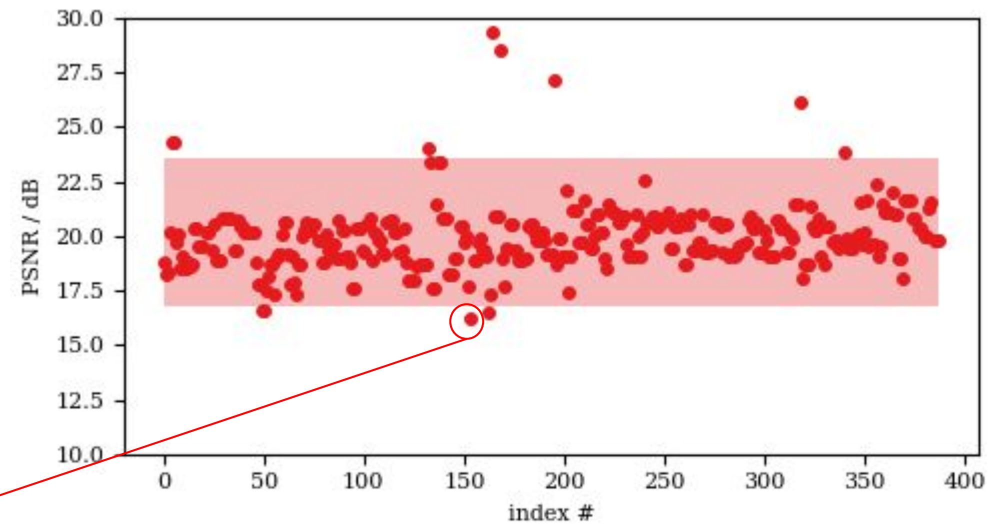
Results on the LBDS simulations

- ➔ Looking at only from simulations, we can also here obtain very good results
- ➔ Basically we can reconstruct the images and their generative parameters

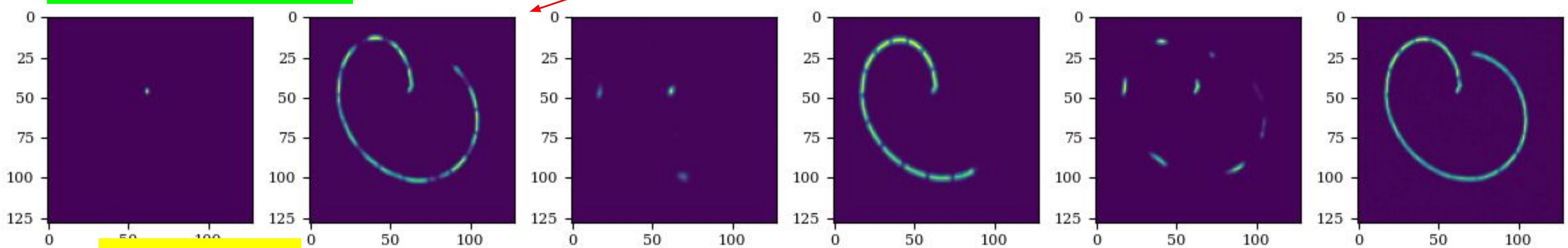


Results on the LBDS data

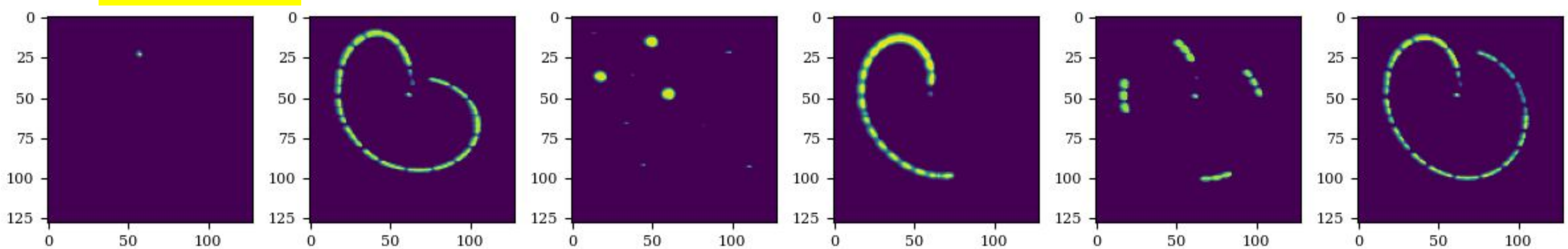
- ➔ Then we tried to use the same model on real data…
- ➔ …that’s a bit more complicated, as incredibly large parameter space (all possible combinations of filling schemes, for example)
- ➔ Usage of reconstruction accuracy can help in failure identification



Reconstruction



Real data



Outline

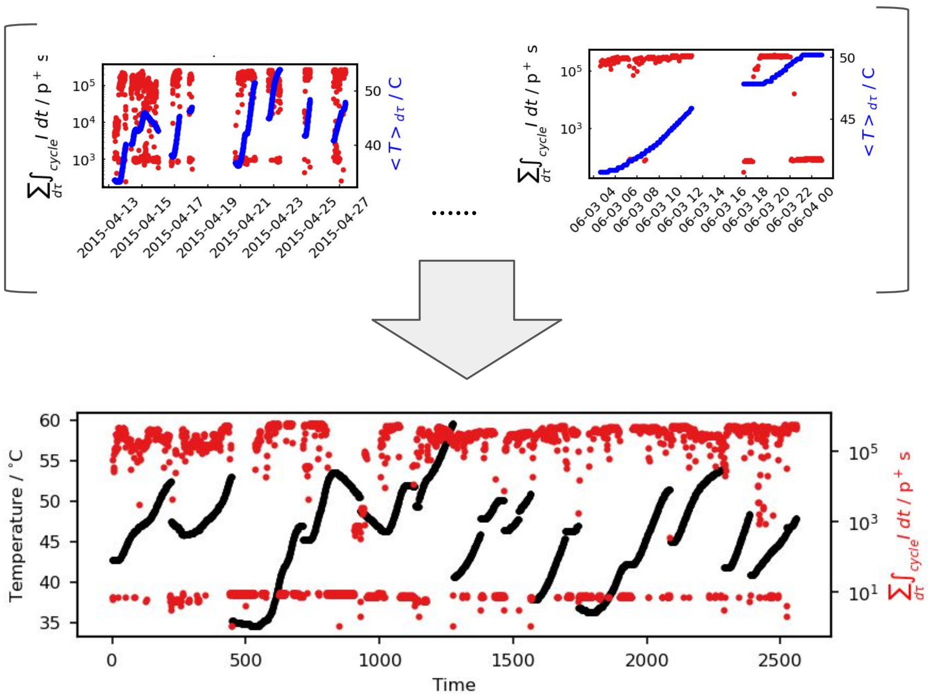
1. Introduction
2. Surrogate models
   a. Crystal shadowing example
3. Image analysis for accelerator systems
   a. Auto Encoders for dump system check
4. Time series analysis
   a. Beam induced heating prediction based on data
   b. Hysteresis compensation for spill quality improvement
5. Conclusions

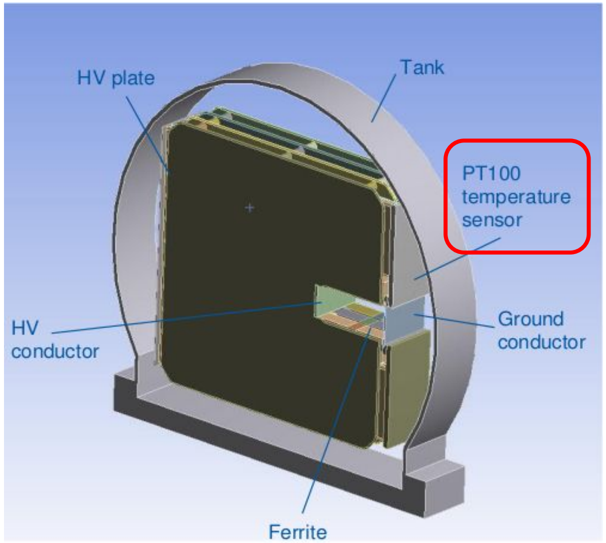
Time series analysis

- ➔ This is huge topic by itself - stocks forecast, weather forecast…
- ➔ Many possibilities and examples at CERN => focus on our 2 cases:
  - Beam induced heating prediction based on data
  - Hysteresis compensation for spill quality improvement

Time series analysis

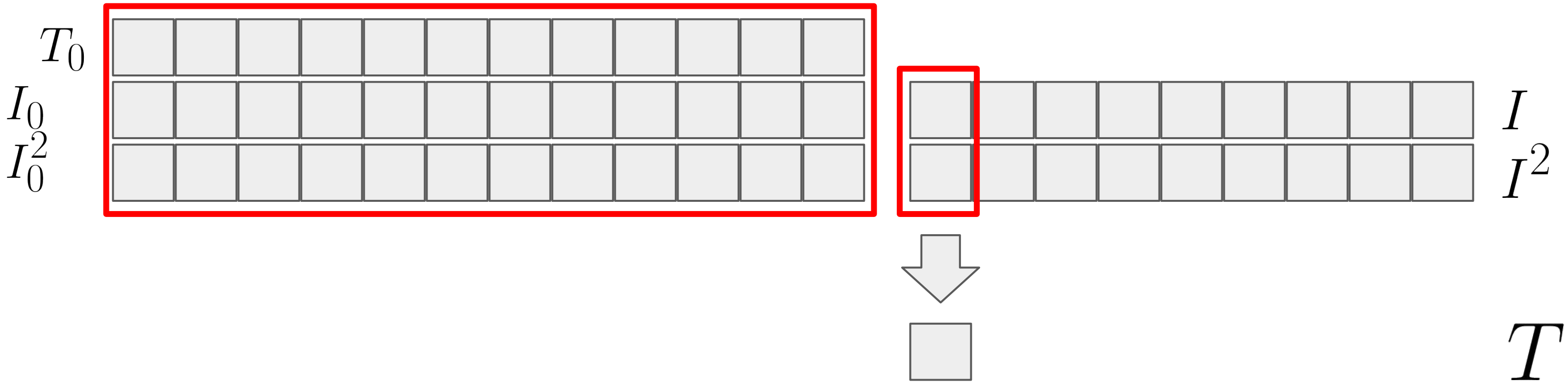
- ➔ This is huge topic by itself - stocks forecast, weather forecast...
- ➔ Many possibilities and examples at CERN => focus on our 2 cases:
  - ◆ Beam induced heating prediction based on data [1]
  - ◆ Hysteresis compensation for spill quality improvement





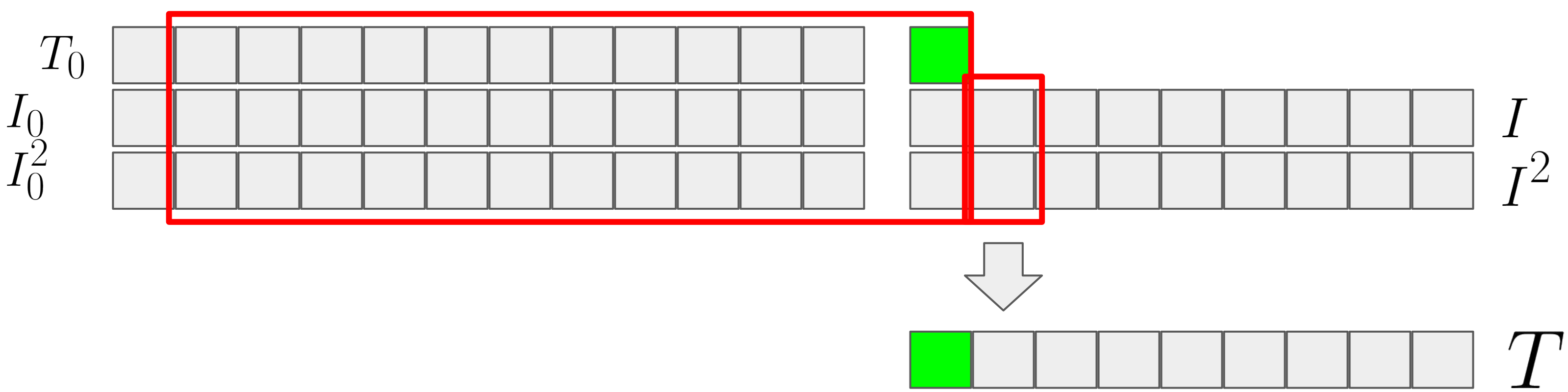
Model for beam induced heating prediction

- ➔ LSTM-based NN used to predict expected beam induced heating of the injection kicker from intensity and temperature history
- ➔ NN is aware of what intensity there will be in the machine and for how long
- ➔ Losses calculated on a fixed sequence length and not value by value



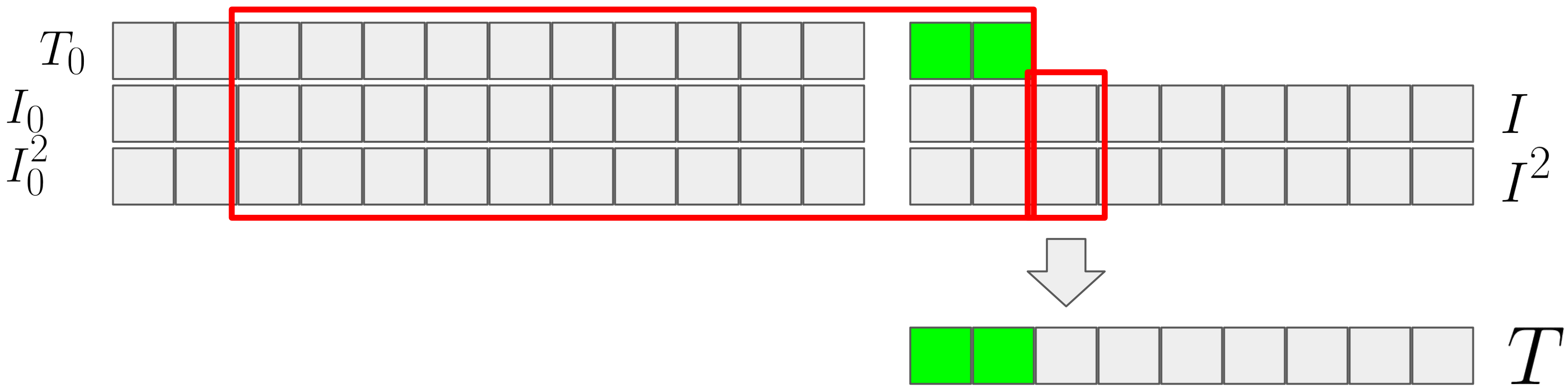
Model for beam induced heating prediction

➔ LSTM-based NN used to predict expected beam induced heating of the injection kicker from intensity and temperature history
➔ NN is aware of what intensity there will be in the machine and for how long
➔ Losses calculated on a fixed sequence length and not value by value



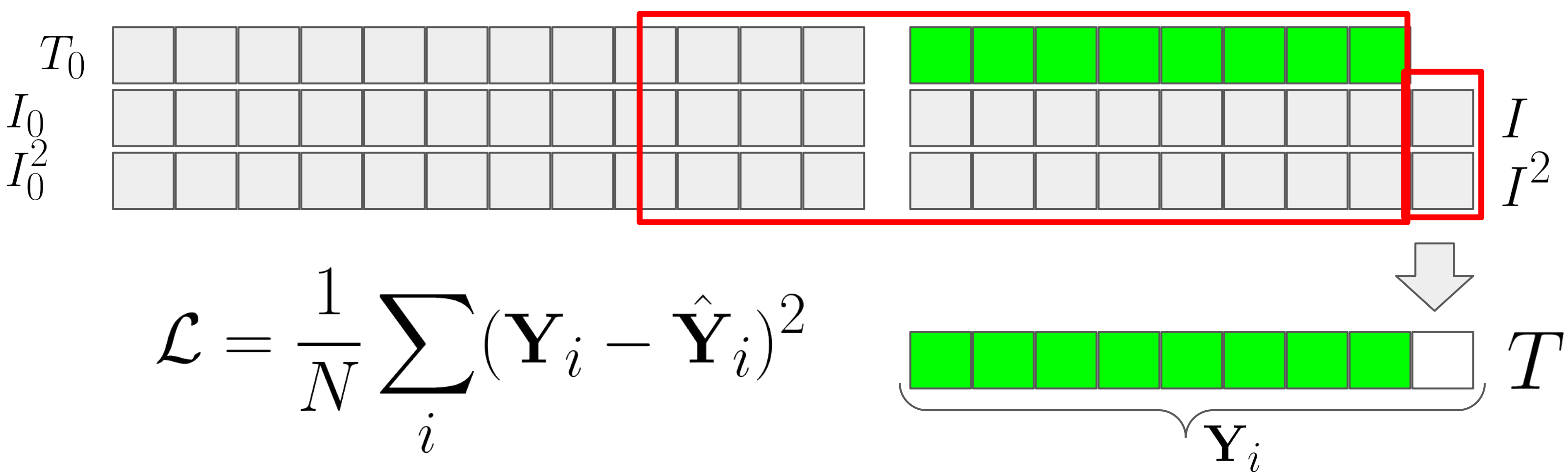
Model for beam induced heating prediction

- LSTM-based NN used to predict expected beam induced heating of the injection kicker from intensity and temperature history
- NN is aware of what intensity there will be in the machine and for how long
- Losses calculated on a fixed sequence length and not value by value



Model for beam induced heating prediction

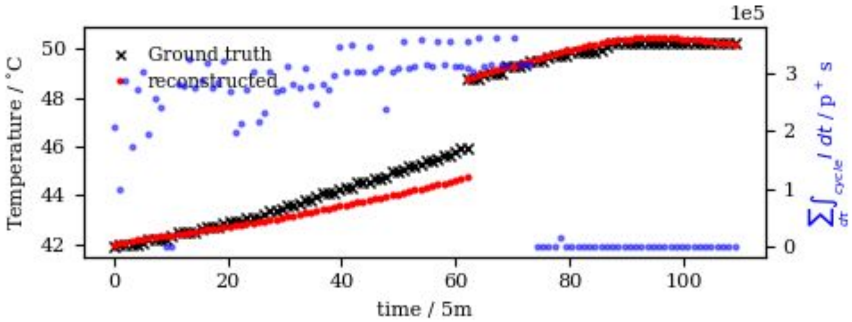
- ➔ LSTM-based NN used to predict expected beam induced heating of the injection kicker from intensity and temperature history
- ➔ NN is aware of what intensity there will be in the machine and for how long
- ➔ Losses calculated on a fixed sequence length and not value by value

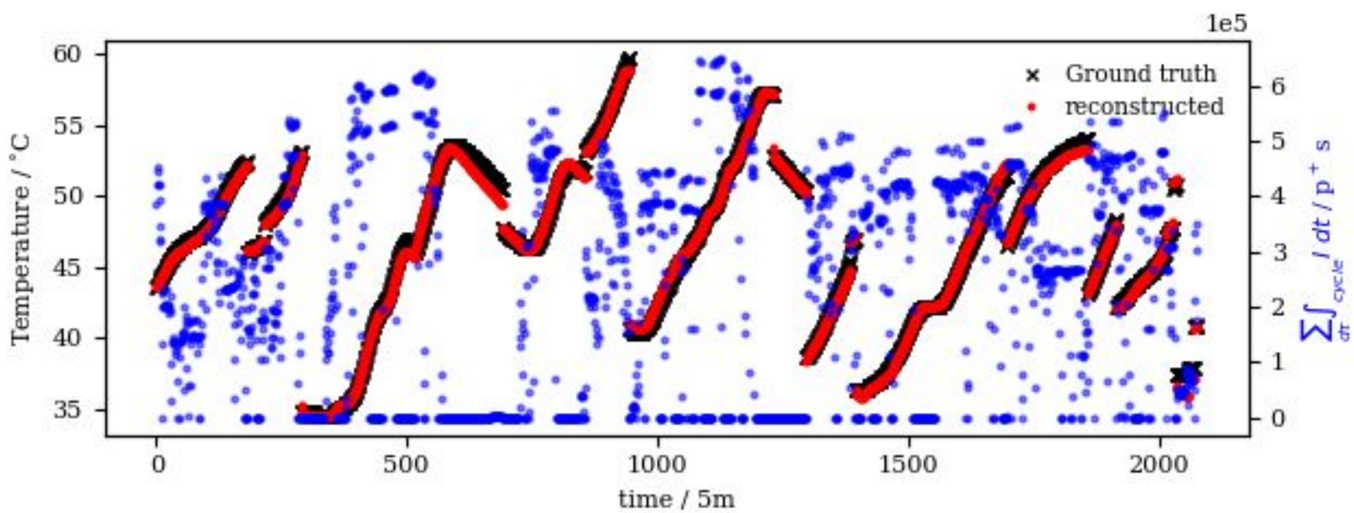


$$\mathcal{L} = \frac{1}{N}\sum_i (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)^2$$

Model for beam induced heating prediction

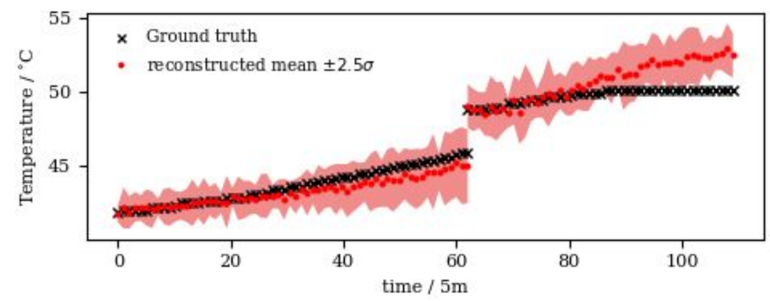
- ➔ Finally we managed to fully reconstruct the training and validation sets…for the full interval duration
  - Here we used 40 initial time steps
- ➔ It needed a custom NN architecture to reproduce data



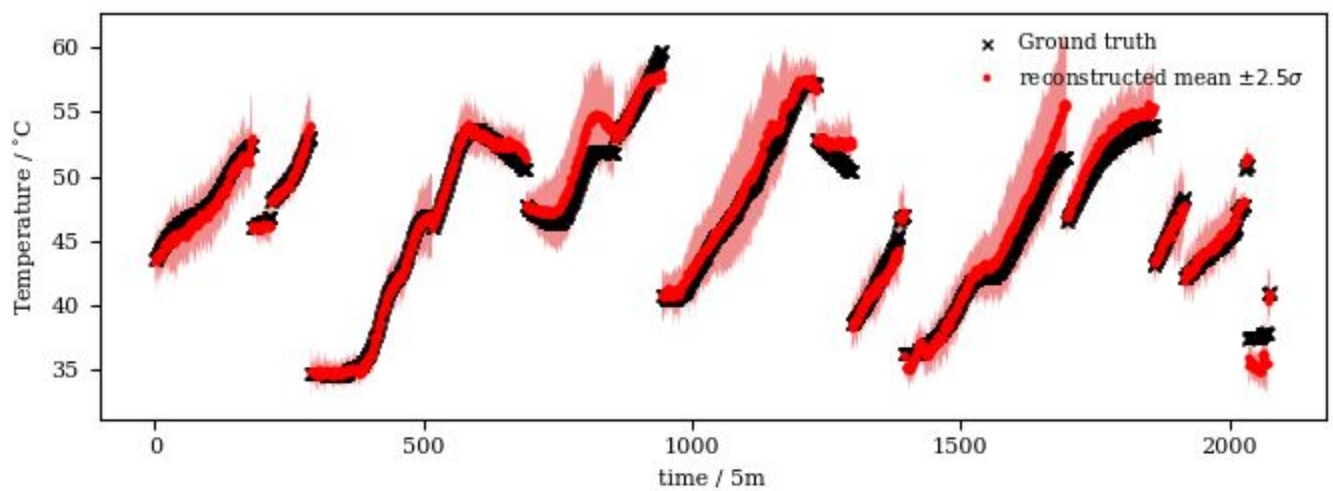


Model for beam induced heating prediction

- ➔ Finally we managed to fully reconstruct the training and validation sets…for the full interval duration
  - Here we used 40 initial time steps
- ➔ It needed a custom NN architecture to reproduce data
- ➔ We also tried Bayesian layers…

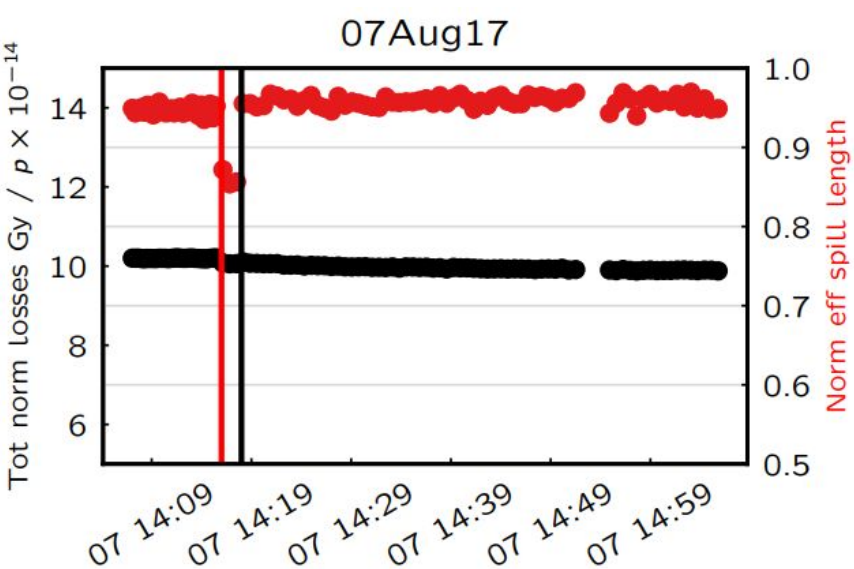


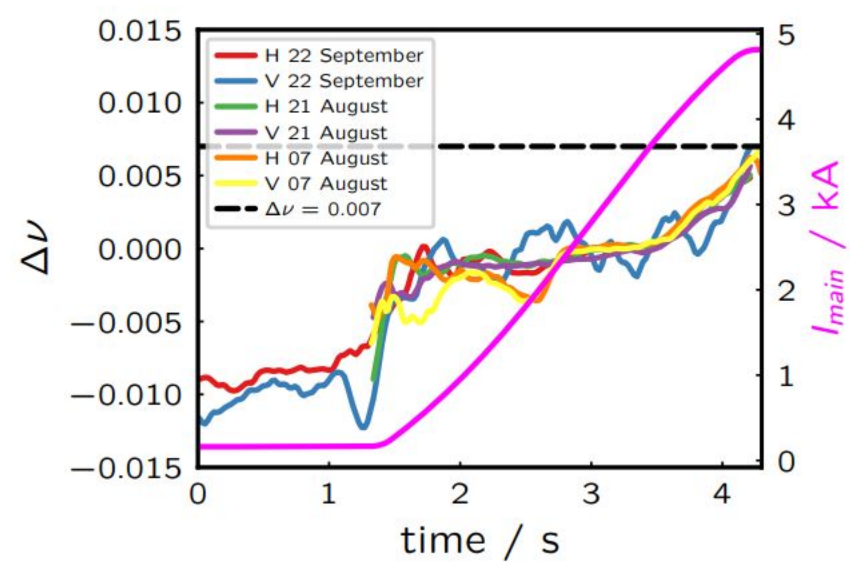
Bayesian layers



Time series analysis

- ➔ This is huge topic by itself - stocks forecast, weather forecast...
- ➔ Many possibilities and examples at CERN => focus on our 2 cases:
  - Beam induced heating prediction based on data
  - Hysteresis compensation for spill quality improvement [2]



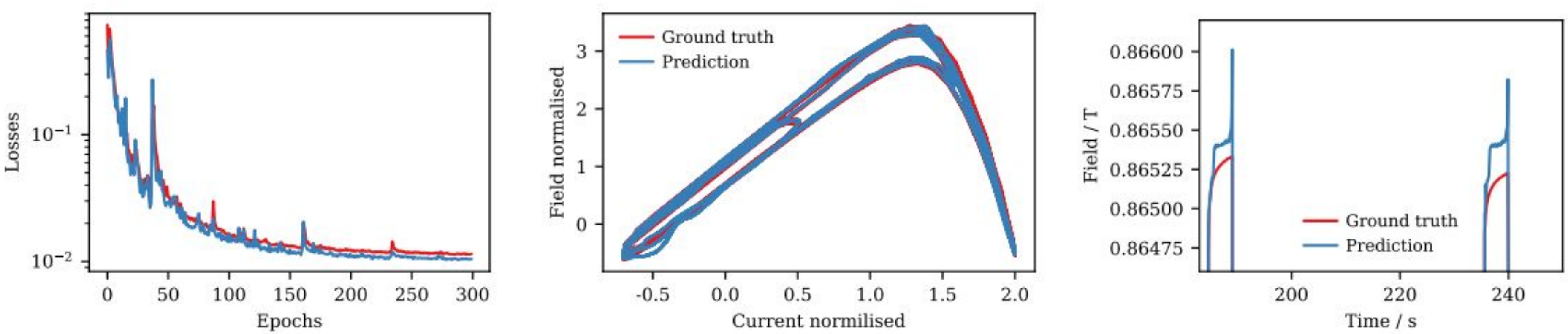


Hysteresis compensation with simple LSTMs

- ➔ As the main SPS quads seems to be the responsible for inducing a tune variation on the SX cycle, and hence a variation of the spill macro-structure => model to predict the expected field given the magnetic and current history

$$x = \{B, I\}(t_0, t_{N-1}),\ ex = I(t_N, t)$$

- ➔ Using pure LSTMs is not enough...

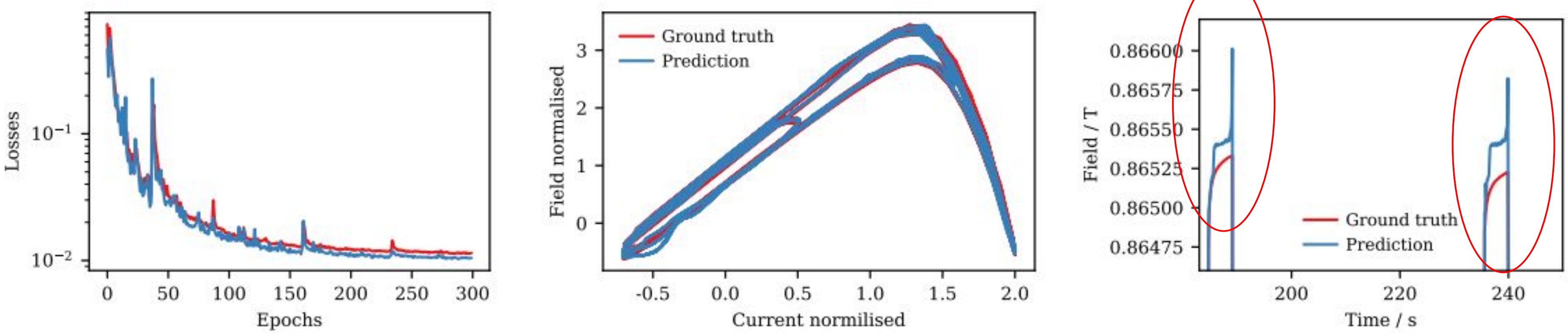


Hysteresis compensation with simple LSTMs

➔ As the main SPS quads seems to be the responsible for inducing a tune variation on the SX cycle, and hence a variation of the spill macro-structure => model to predict the expected field given the magnetic and current history

$$x = \{B, I\}(t_0, t_{N-1}), ex = I(t_N, t)$$

➔ Using pure LSTMs is not enough...



Hysteresis compensation with PINN

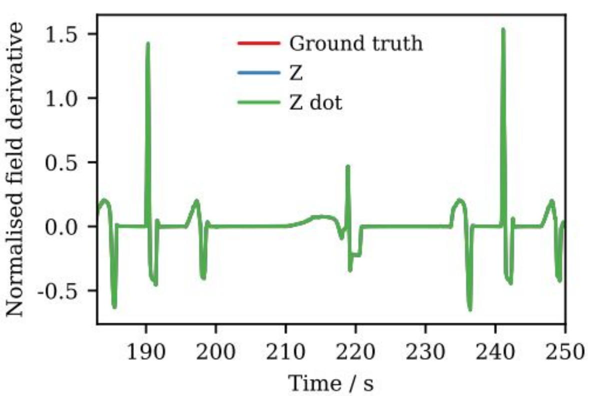
- To make it work, we had to use more advanced models
- To properly model hysteresis, we used Physics Informed NN [3, 4]
- Basically we include in the loss function to train the NN information about the physical phenomenon under interest
  - For example, if the system is governed by the following:

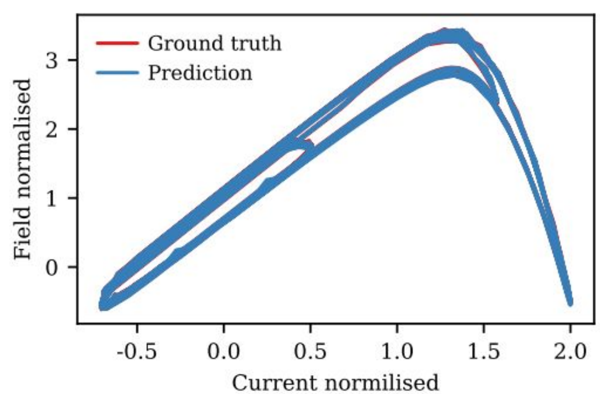
$$a\ddot{y} + \dot{y} = cx(t)$$

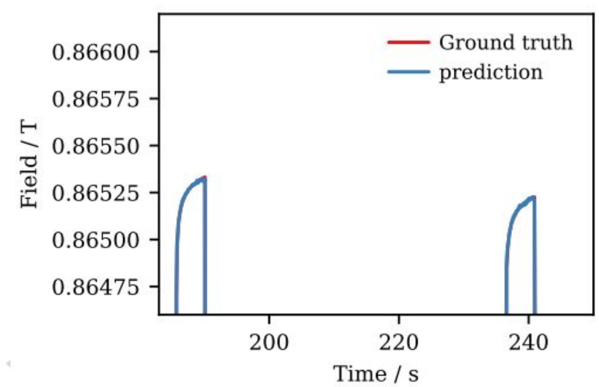
  - Then we can make a PINN with the following loss function:

$$\mathcal{L} = \sum_{n=1}^{N} \frac{1}{N} \left(\alpha \left\|y - \overline{y}\right\|_2^2 + \beta \left\|\dot{y} - \dot{\overline{y}}\right\|_2^2 + \gamma \left\|\ddot{\overline{y}} - NN(x, \dot{y})\right\|_2^2 \right)$$

- It gave results where all other NN topology failed!







Hysteresis compensation with PINN

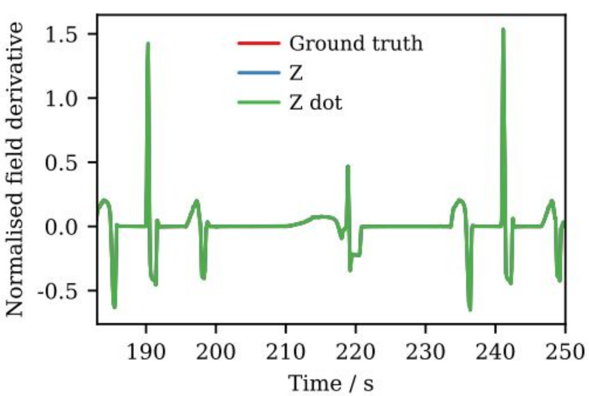
- To make it work, we had to use more advanced models
- To properly model hysteresis, we used Physics Informed NN [3, 4]
- Basically we include in the loss function to train the NN information about the physical phenomenon under interest
  - For example, if the system is governed by the following:

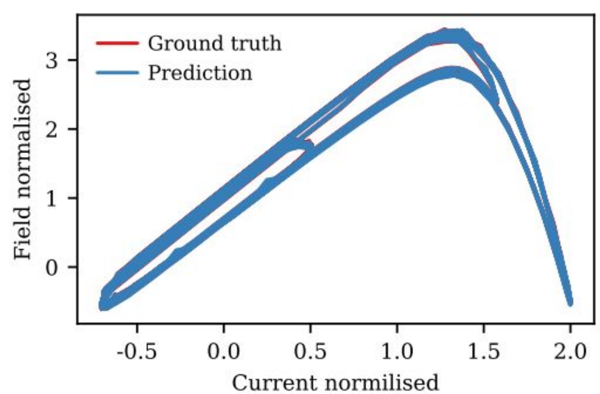
$$a\ddot{y} + \dot{y} = cx(t)$$

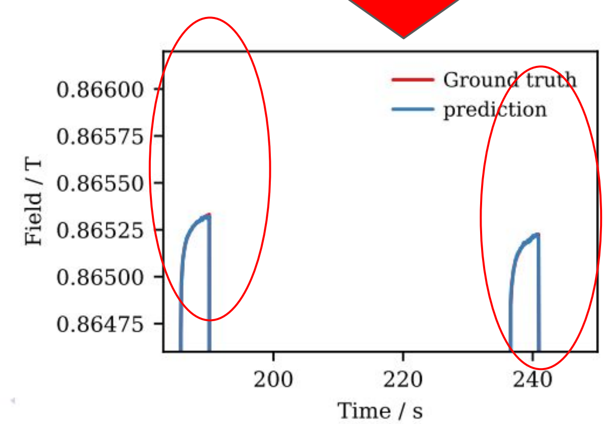
  - Then we can make a PINN with the following loss function:

$$\mathcal{L} = \sum_{n=1}^{N} \frac{1}{N} \left(\alpha \left\| y - \bar{y} \right\|_2^2 + \beta \left\| \dot{y} - \dot{\bar{y}} \right\|_2^2 + \gamma \left\| \ddot{\bar{y}} - NN(x, \dot{y}) \right\|_2^2 \right)$$

- It gave results where all other NN topology failed!







Outline

1. Introduction
2. Surrogate models
    a. Crystal shadowing example
3. Image analysis for accelerator systems
    a. Auto Encoders for dump system check
4. Time series analysis
    a. Beam induced heating prediction based on data
    b. Hysteresis compensation for spill quality improvement
5. Conclusions

Conclusions

- This was just a very short summary of what we explored so far…and completely missing some important details
- For example:
  - Longitudinal tomography with VAE
  - Neural ODE networks
  - Hamiltonian NN…
- We have shown a few interesting problems that had no solutions since we started tackling them with ML
  - Many others could be addressed - the choice of already available models and possible combinations are enormous
  - For example, could we speed up tracking for SX using Hamiltonian NN?
  - Could we exploit NN to speed up Monte Carlo tracking routines for losses estimation?
- Most of the examples here are not yet fully used in daily operation (we only used the heating model last year) - the optimal exploitation will come when we will deploy them!

Thanks!

Putting VAE and surrogate models together

- ➔ We exploited the usage of VAE, the idea of SM and reinforcement learning (see Verena’s talk) to make a synthetic environment of a transfer line
- ➔ Basically we can train a RL agent on synthetic data and tune hyperparameters, compare agents...and more!
- ➔ Paper on its way...