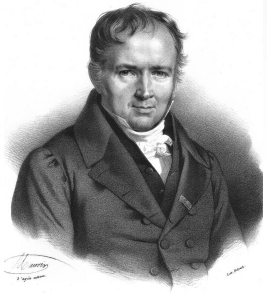
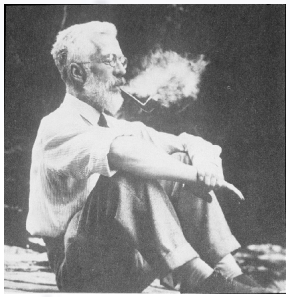
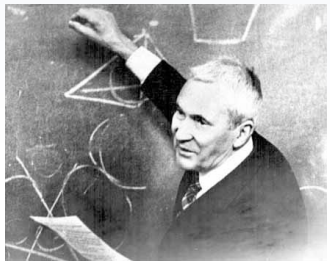
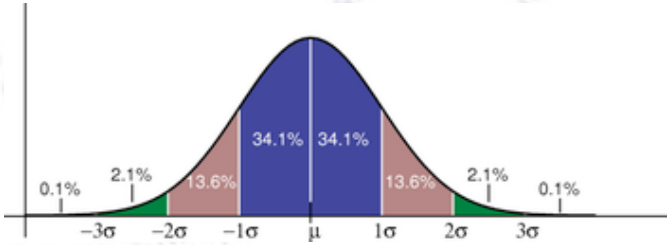
Applied ML

Potential Final Projects/Datasets



Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"

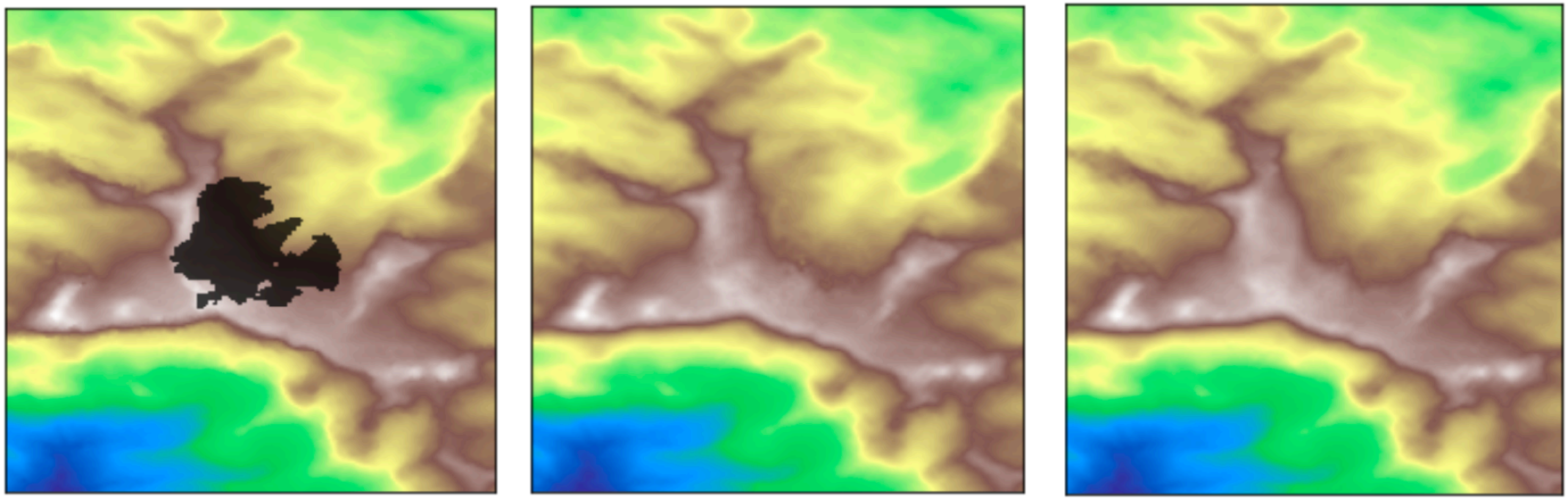
Glacial data

Glaciers are - from satellite images - well measured in area, but not in volume.

The project aim is to estimate glacier volumes from a mixture of inputs:

1. Satellite images
2. Radar measurements of depth made on the glaciers
3. Databases

The project will be in collaboration with (excellent) Niccolo Maffezzoli.



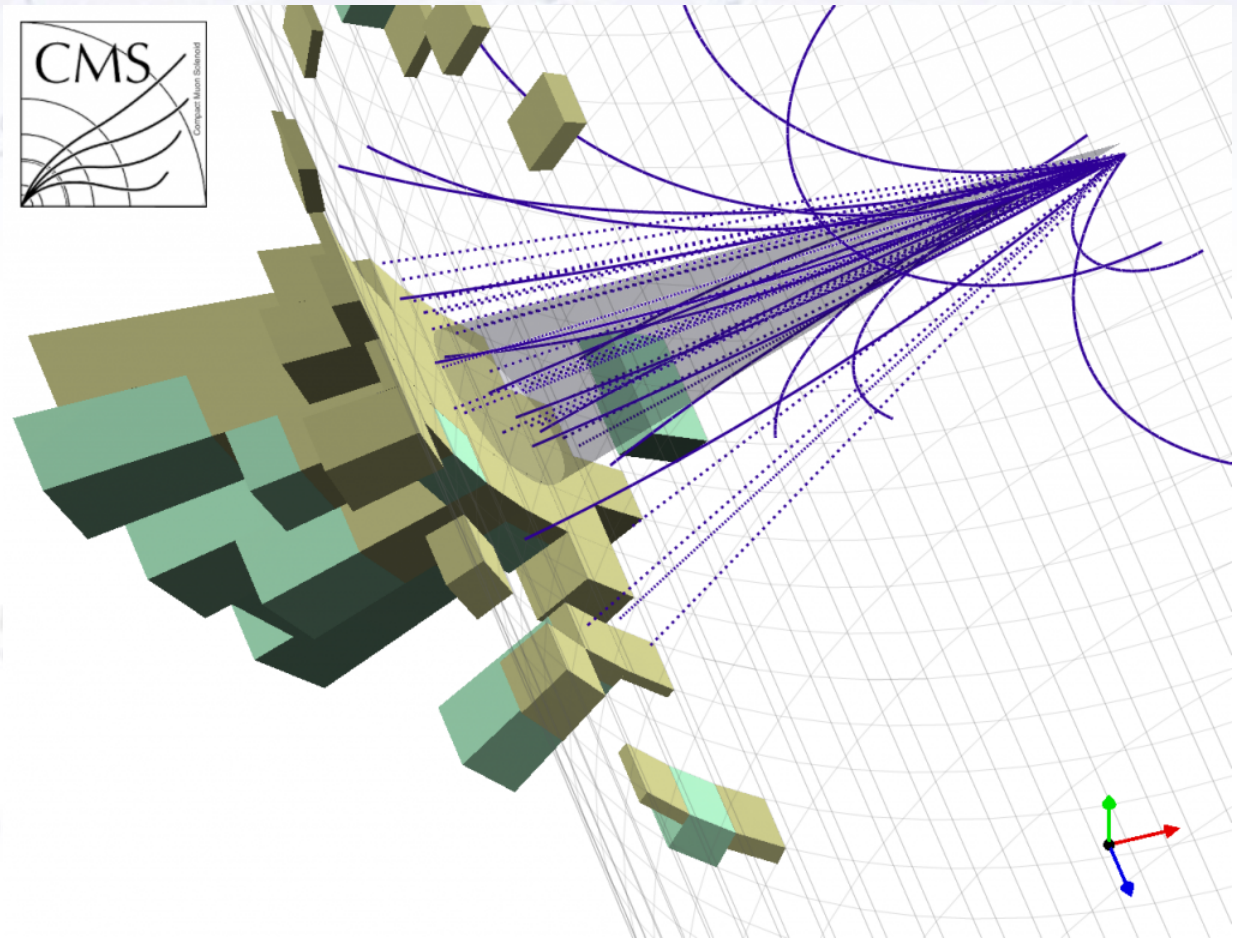
(a) (b) (c)

Particle Jets

When quarks collide at the LHC, they enter the detectors as a "jet" of particles, with measurements from both tracking (purple) and calorimeters (green'ish).

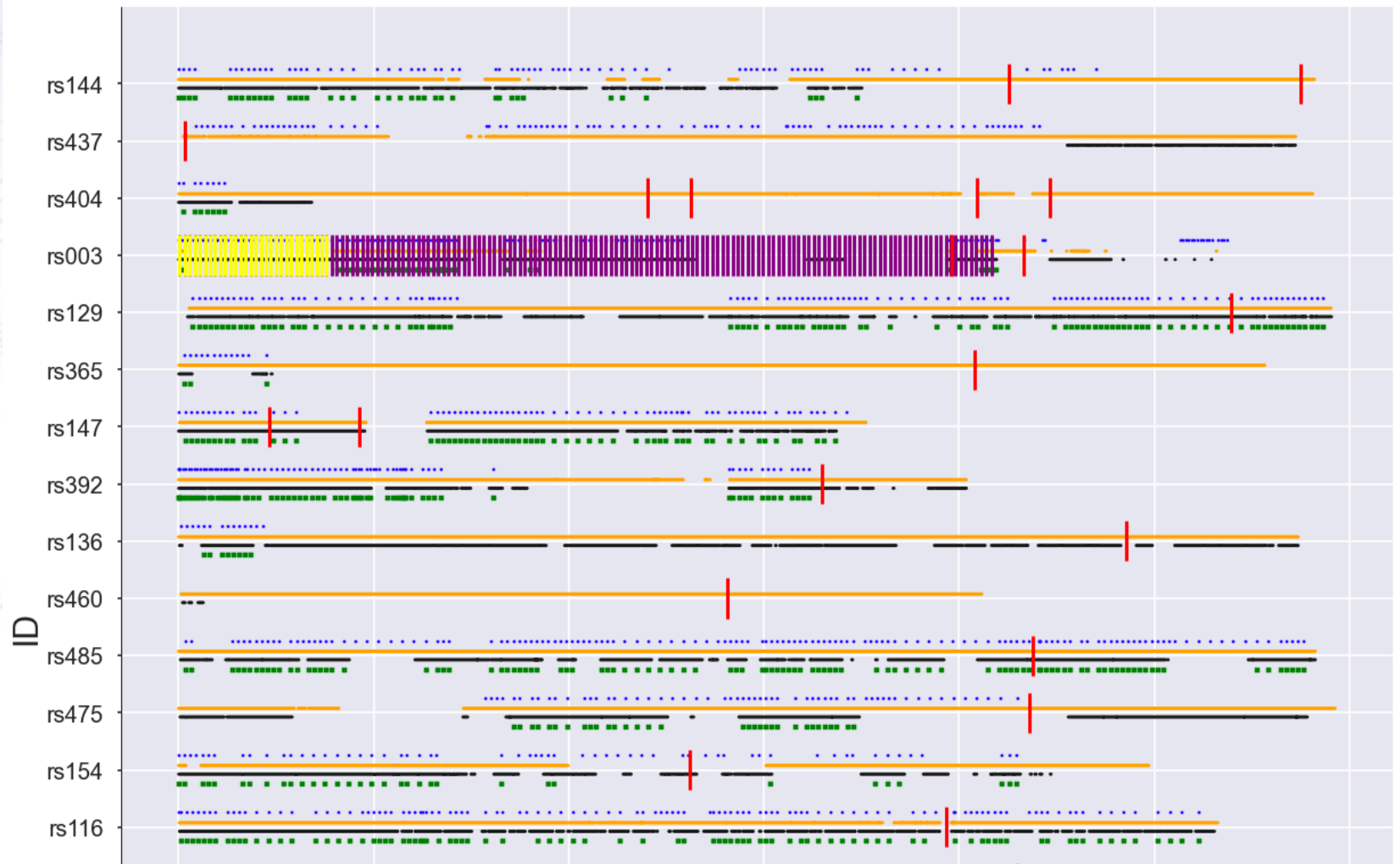
The challenge is determine the energy of the original quark, which consists of putting together a non-constant number of jet constituents.

The dataset is already produced (by Arnau) and contains 4 million jets.



WARD data

Based on 7 continuous patient outputs, determine how the patient is doing. The challenge is the discontinuous data, and the fact that there are only about 200 cases of problems. Done in collaboration with Norman Pedersen (data is ready).



Adversarial Training

Problems potentially arise, when you train ML models on **simulated data** and then apply it to **real data**, since the two might not follow the same distributions!

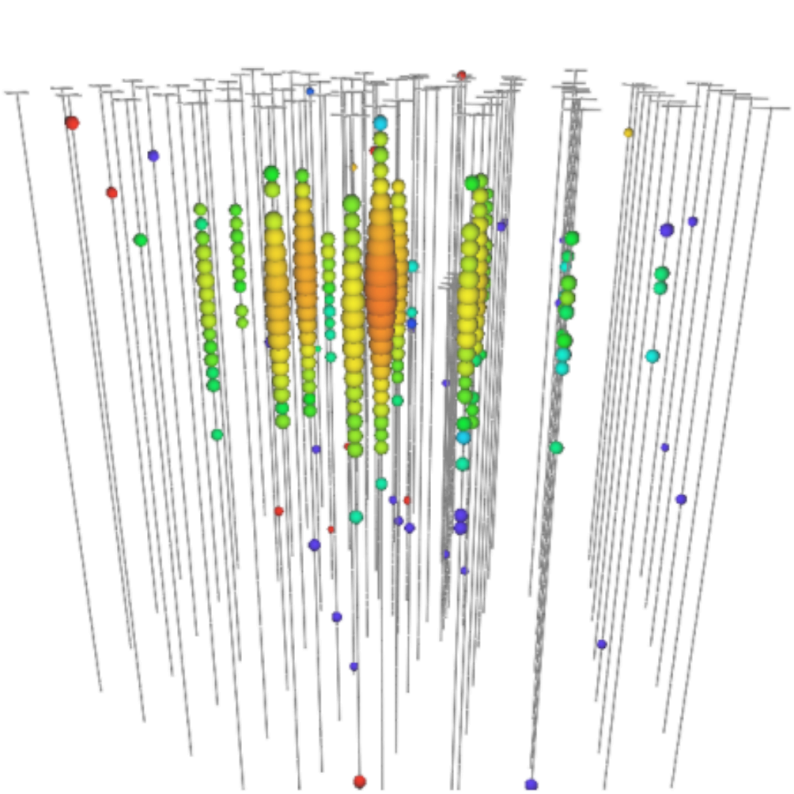
One idea for a method to minimising this problem is to apply adversarial training:

1. First train a model on simulated data.
2. Introduce an "Adversarial Network" to a layer in the model, that tries to see the difference between the real and simulated data. Any ability to do so penalises the loss.
3. Retrain the model, this time with the "requirement" that at some stage in the network, the real and simulated data has to be very similar.

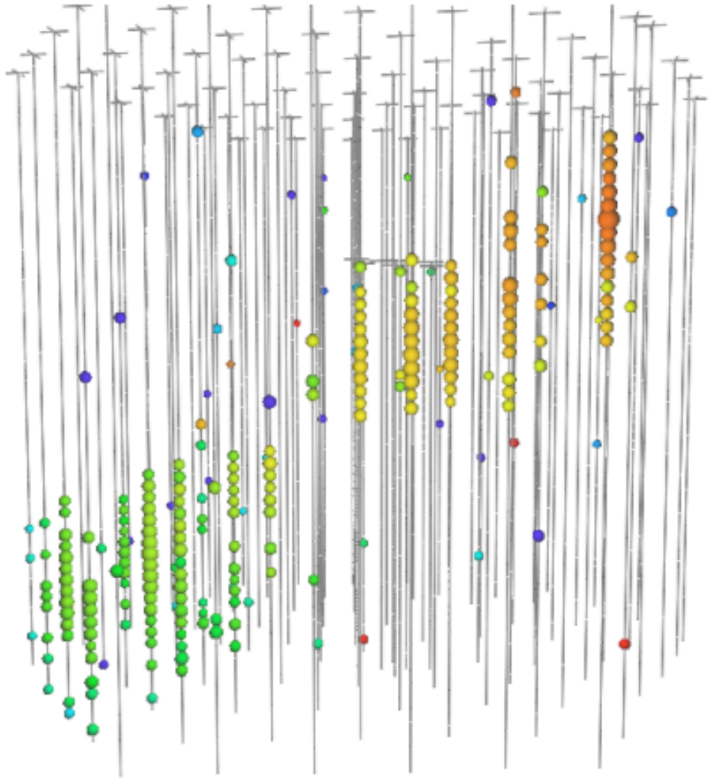
This approach can be applied to e.g. the Aleph b-jet (simulated) data, where part of the data can be considered real data, after applying some shifts to the distributions.

Particle Physics data

Based on IceCube detector signals (8-80000 pulses with 6 features each), determine the type of signal, its energy, direction, interaction vertex, etc. This involves using Graph Neural Networks and Transformers.



(a) Single Cascade



(b) Track

Bonus Slides

Let me tell you about this project!