Uncertainty-aware Machine Learning

David Rousseau, IJCLab-Orsay

david.rousseau@in2p3.fr

DMLab annual meeting

DESY 12-13 Dec 2022

cnrs IN2P3 Les deux infinis

IJCLab Irène Joliot-Curie Laboratoire de Physique des 2 Infinis

université PARIS-SACLAY

Phys.Rev.Lett. 114 (2015)191803

Most complex measurement ever ?

Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments

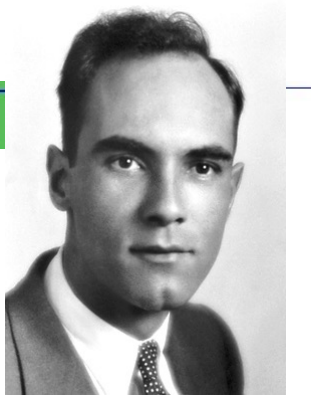
(ATLAS Collaboration)[†]
(CMS Collaboration)[‡]

(Received 25 March 2015; published 14 May 2015)

A measurement of the Higgs boson mass is presented based on the combined data samples of the ATLAS and CMS experiments at the CERN LHC in the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels. The results are obtained from a simultaneous fit to the reconstructed invariant mass peaks in the two channels and for the two experiments. The measured masses from the individual channels and the two experiments are found to be consistent among themselves. The combined measured mass of the Higgs boson is $m_H = 125.09 \pm 0.21\,(\text{stat}) \pm 0.11\,(\text{syst})$ GeV.

Positron discovery

- Cloud chamber picture



Carl Anderson
Nobel 1936

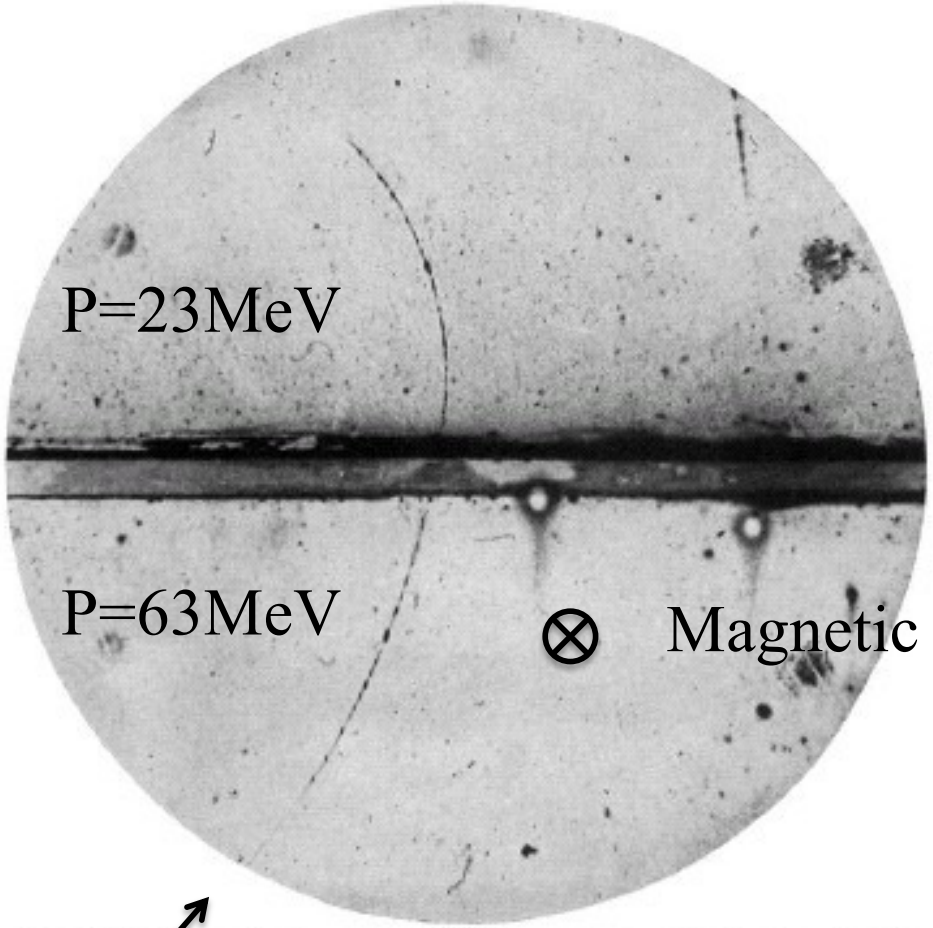
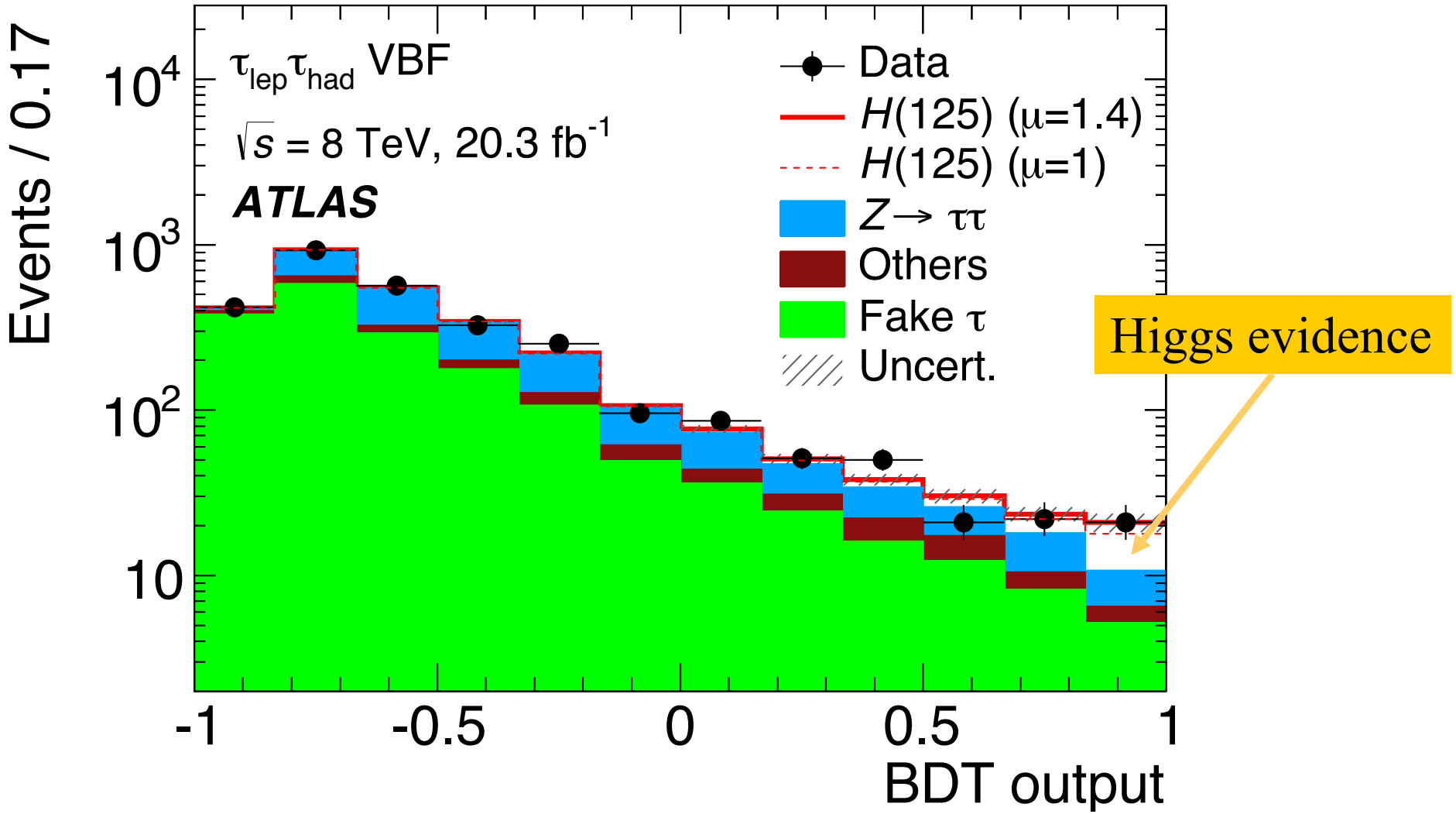


FIG. 1. A 63 million volt positron ($H\rho = 2.1 \times 10^5$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H\rho = 7.5 \times 10^4$ gauss-cm). The length of this latter path is at least two times greater than the possible length of a proton path of this curvature.

Evidence using a classifier

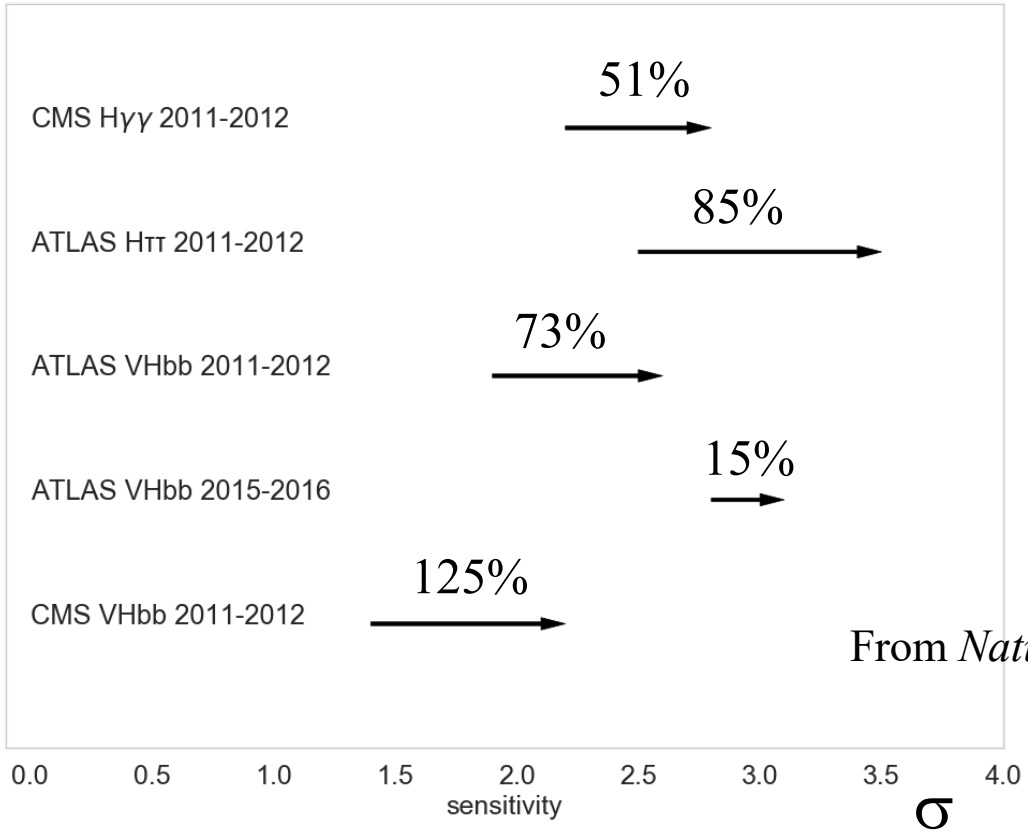
JHEP 04, 117 (2015) 1501.04943

Boosted Decision Tree using ~dozen of high level variables



ML on Higgs Physics

- At LHC, Machine Learning used almost since first data taking (2010) for reconstruction and analysis
- For example, impact on Higgs boson sensitivity at LHC:



From *Nature* **560**, 41–48 (2018) |

➔~50% gain on LHC running

Dealing with Uncertainties

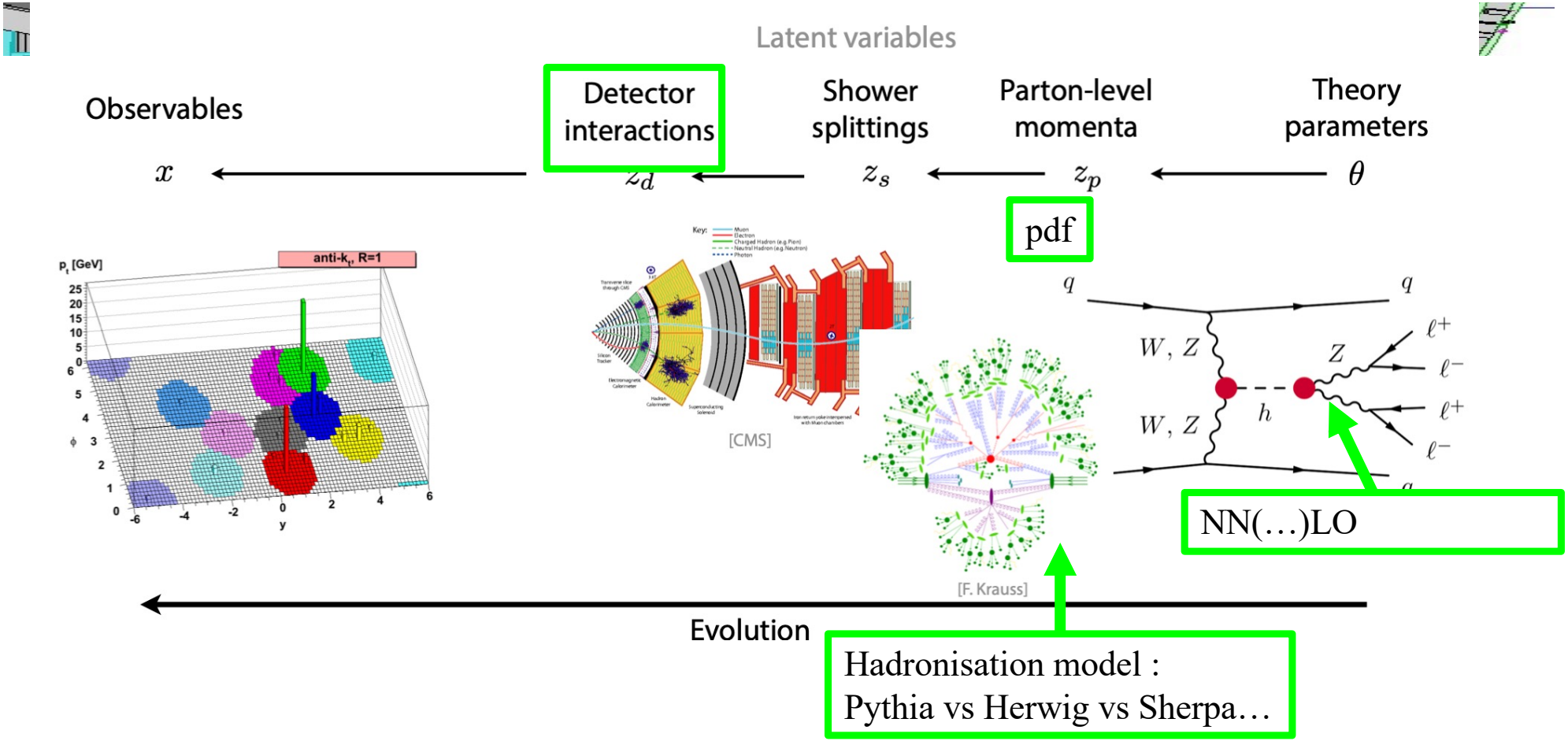
- Our experimental measurement papers typically end with
  - measurement = m ± σ(stat) ± σ(syst)
  - σ(syst) systematic uncertainty : known unknowns, unknown unknowns… Convincing oneself, co-authors, the whole community that we know what we are doing➔trust !
- Name of the game is to minimize quadratic sum of :

$$\sigma(\text{stat}) \oplus \sigma(\text{syst})$$

- … while ML techniques are usually trained to minimise σ(stat), and systematics uncertainties accounted for in a last stage
  - Of course, physicists know how to minimize systematics "by hand", by constraining some on data, by deliberately not rely on poorly understood quantities,…
- Two challenges:

1. Maintain trust (σ(syst)) while using AI more and more
2. Include somehow (various techniques) $\sigma(\text{stat}) \oplus \sigma(\text{syst})$ in the loss in order to minimise overall uncertainty

Modelling particle physics processes



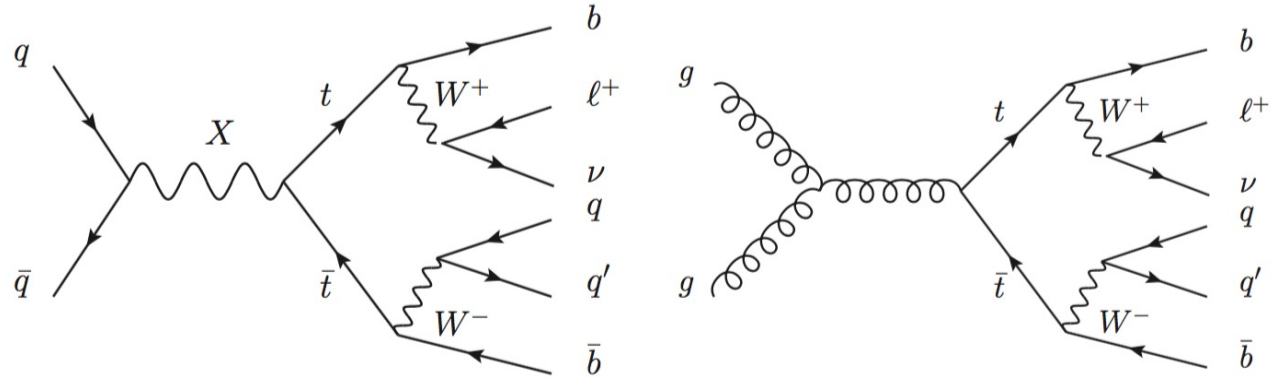
Usually : very large number of corrections measured on data itself, remaining uncertainties due to second order effects, statistical uncertainties

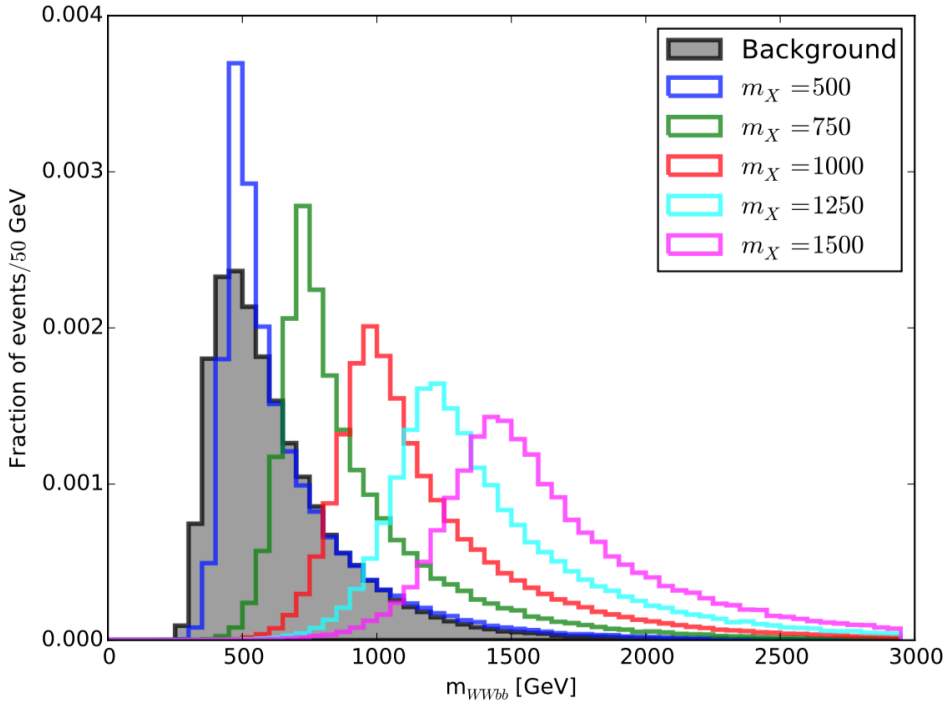
Final uncertainty obtained from varying the *Nuisance Parameters*

Examples of uncertainty aware technique

Parameterised learning

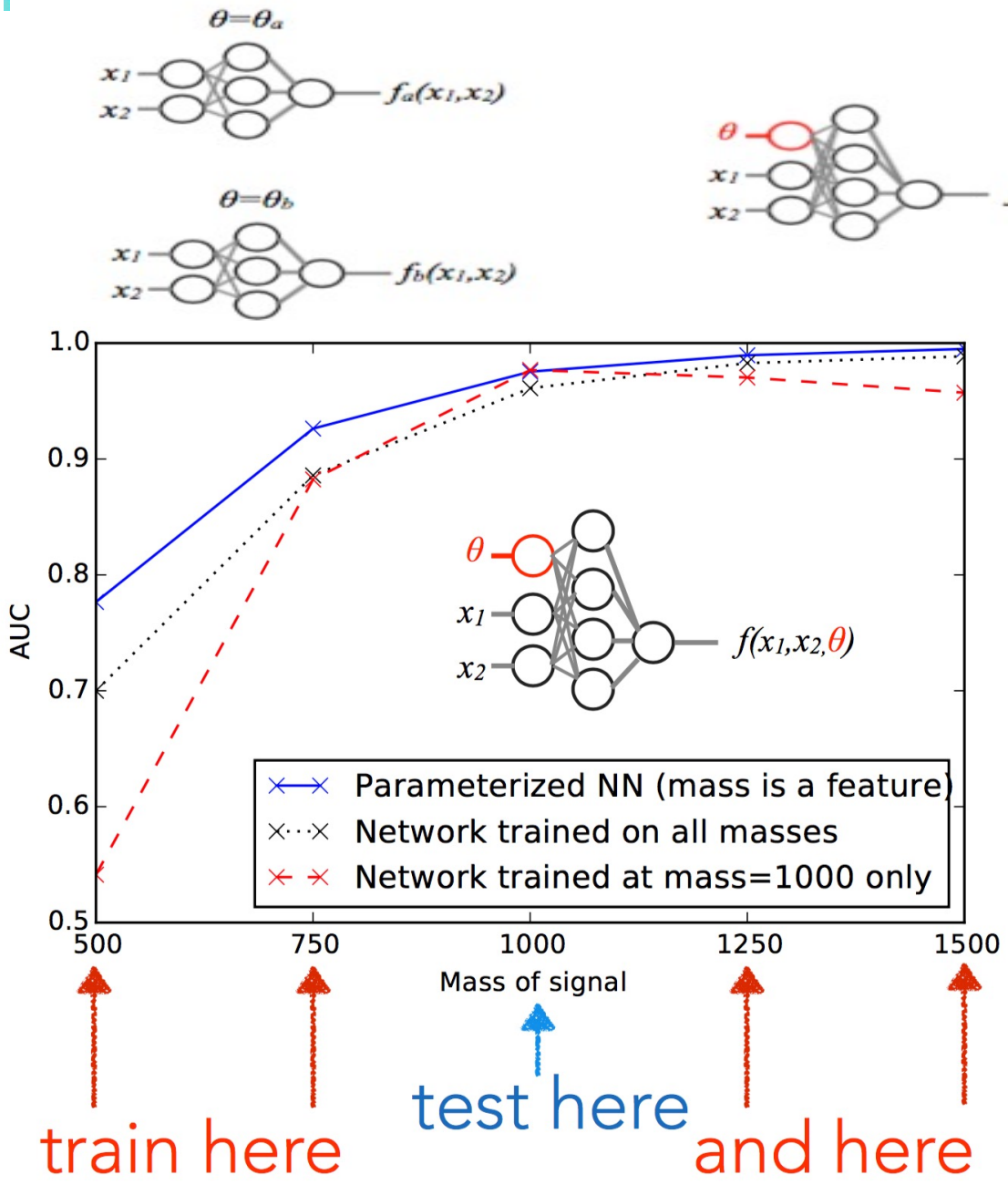
1601.07913 Baldi, Cranmer, Faucett, Sadowksi, Whiteson





- Typical case: looking for a particle of unknown mass
- E.g. here tt decay

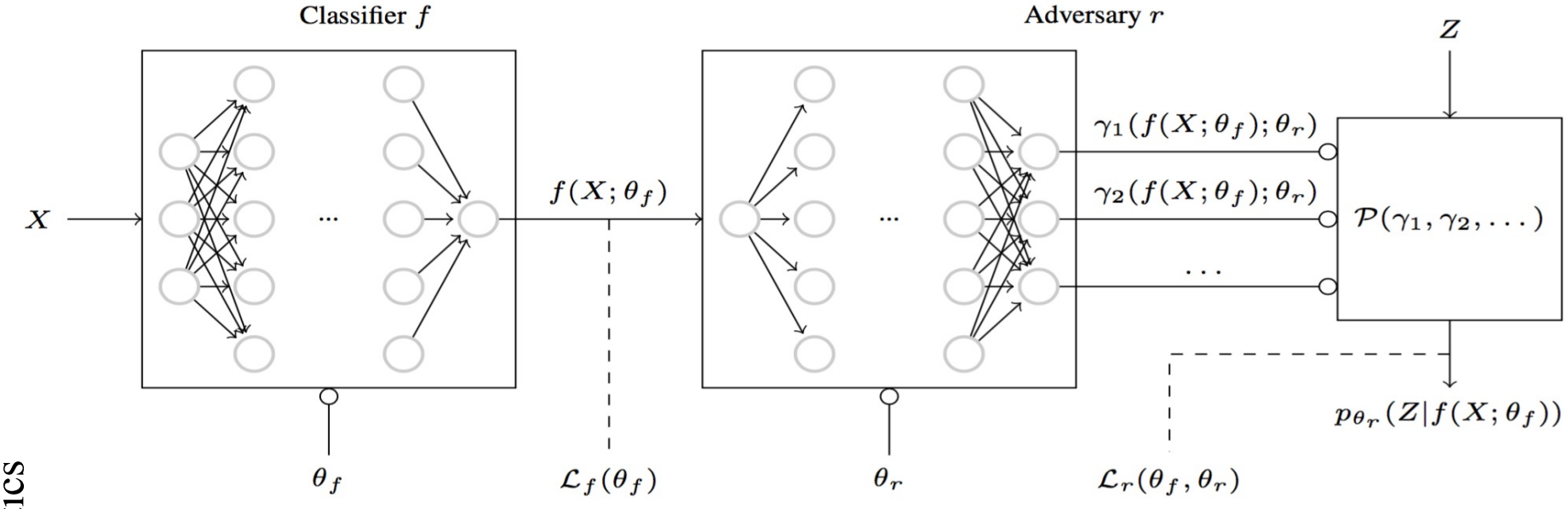
Parameterised learning (2)



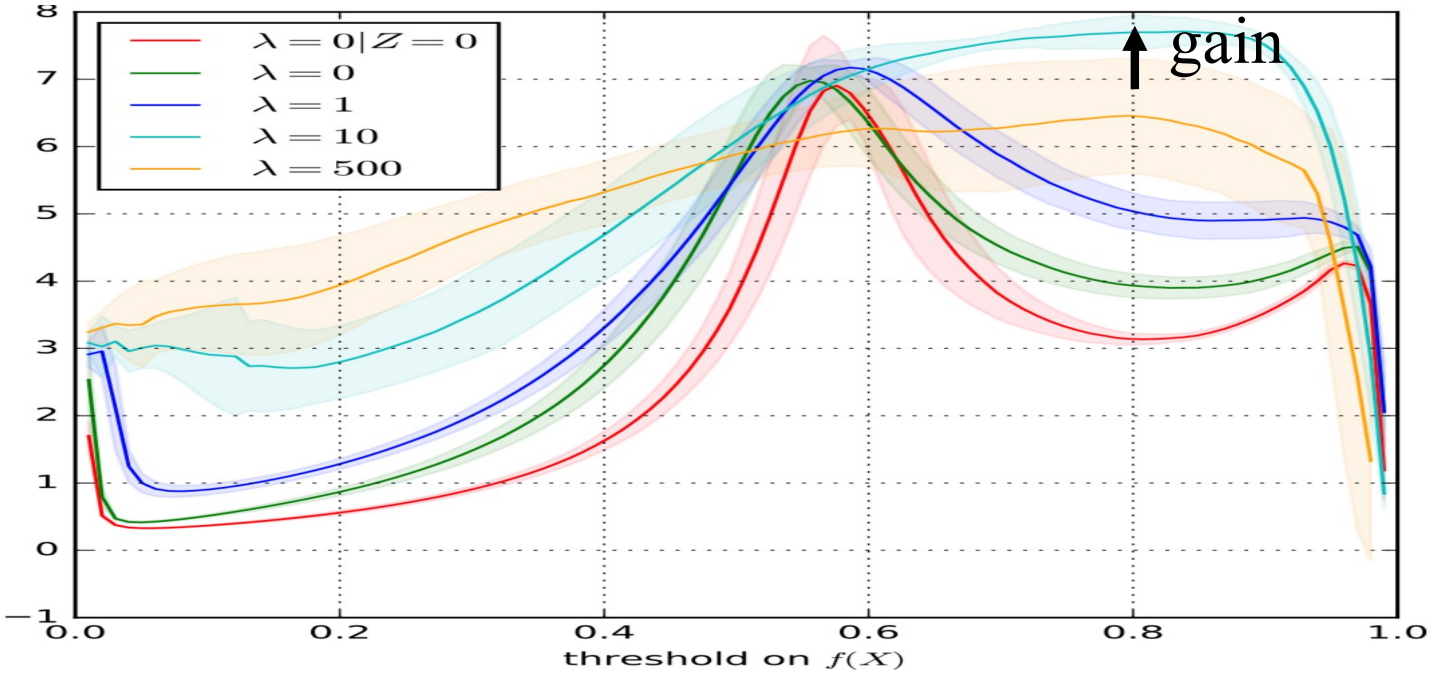
- Train on 28 features plus true mass
  - Nothing fancy, can be done out of the box with any classifier
- Parameterised NN as good as single mass training
- ➔clean interpolation
- Used for real in search X → S(bb)H($\gamma\gamma$) channel
- Some (e.g. arXiv:1909.03081) argue that this can be used with H being a Nuisance Parameter
  - ➔debatable

Pivot

Louppe et al, 1611.01046



Significance including systematics

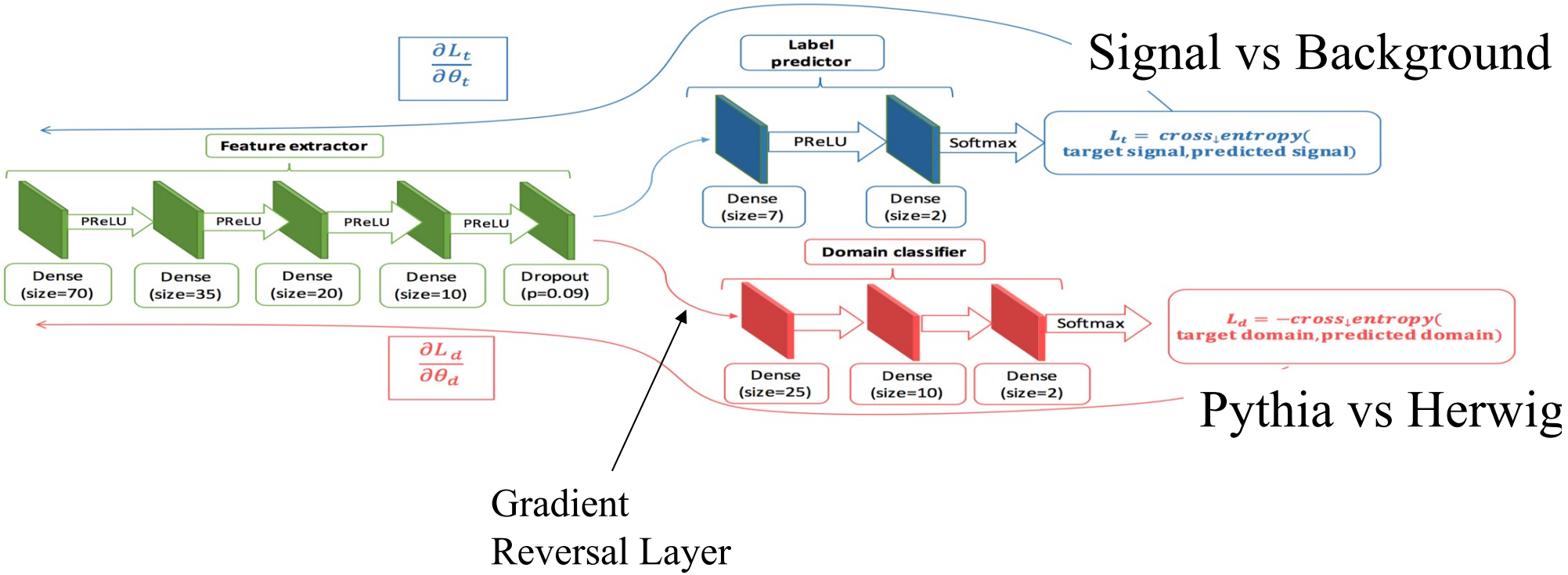


Delphes simulation of Boosted W Tagging

Reduce pileup systematics

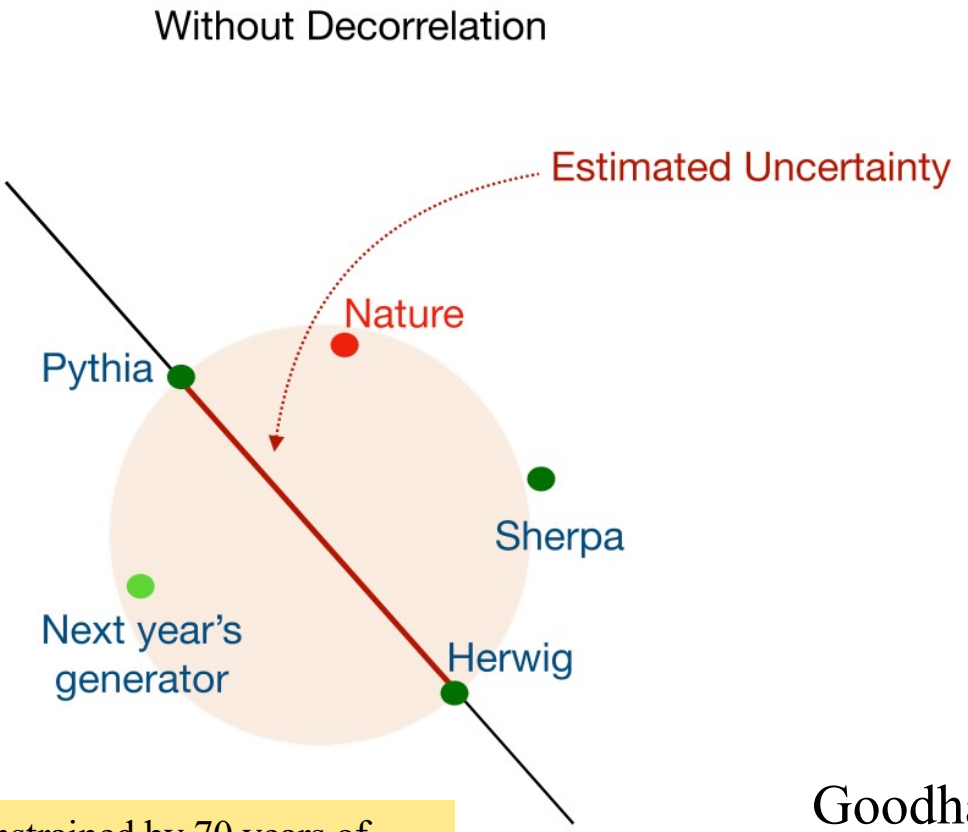
Syst Aware Training: adversarial

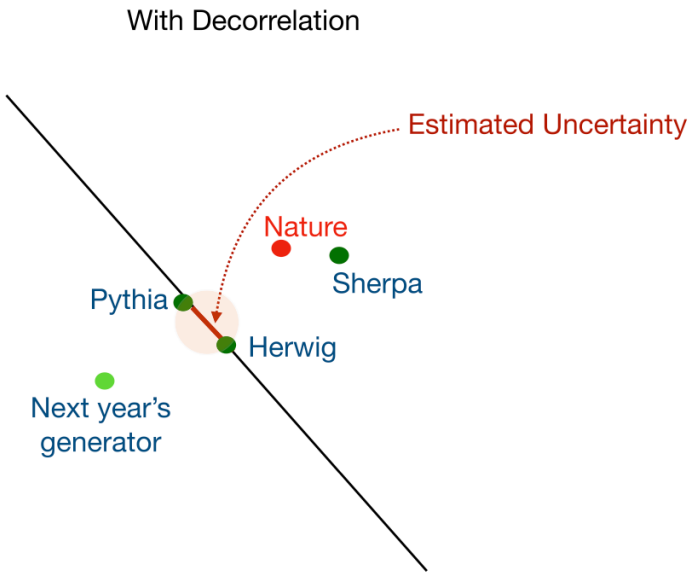
Inspired from 1505.07818 Ganin et al : 2105.08742 https://github.com/hep-lbdl/systaware



Cautionary tale

Ghosh & Nachman EPJC **82** 46 (2022)





Constrained by 70 years of Particle Physics measurements

Goodhart's law "When a measure becomes a target, it ceases to be a good measure"

Wrap up

Fair Universe project

- DOE Funded : Berkeley (LBNL, NERSC)+ ChaLearn (Isabelle Guyon, David Rousseau)
  - Note : IG (CS) and DR (HEP) have organised kaggle competitions »HiggsML » and « TrackML »
- Just starting (Nov 2022 press release) for 3 years
- 2 post-docs at Berkeley, 1 PhD student at U Paris-Saclay/IJCLab
- Pitch :
  - more and more ML in HEP. How to make the best of measurement, including systematic incertainties ?
  - Active field, but results are not comparable
  - ➔build large open public datasets corresponding to difficult HEP measurements : typical and increasing complexity
  - ➔develop benchmarks (==datasets + figure of merit)
    - Cross section measurement of a rare process
    - Parameter measurement like top-quark mass
    - Cosmology measurement tbd
  - ➔eventually organise scientific competitions for the best strategies
- Rely on NERSC supercomputer for large dataset creation and benchmark running
  - Need to evaluate precisely the uncertainties➔very large dataset
- Figure of merit : should evaluate total uncertainty including systematics in a totally non-hackable way

Interest in ML uncertainties, not just HEP

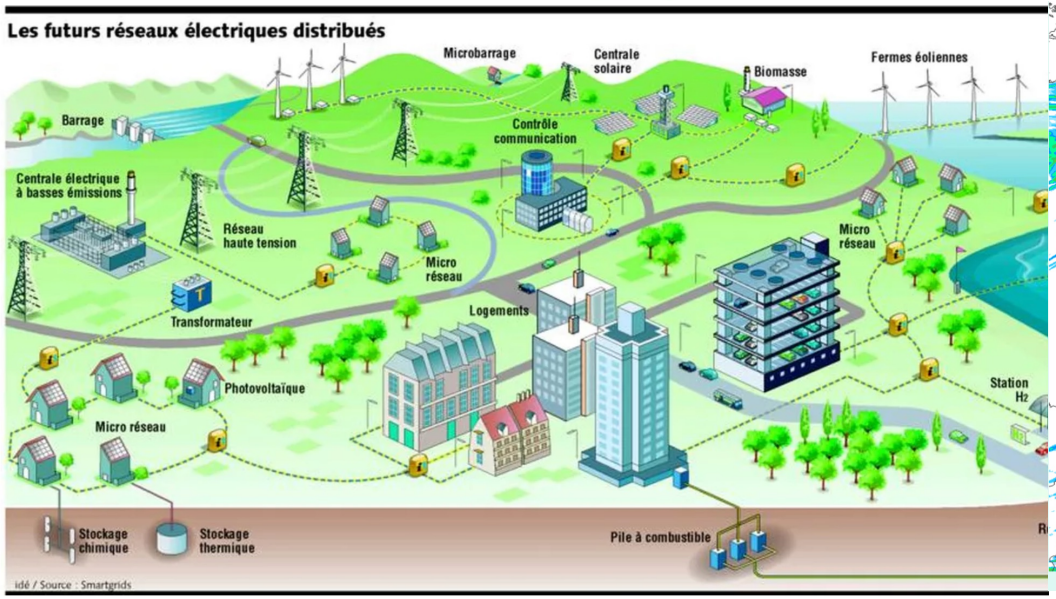
- Inaugurated last month: (Computer Science and industry)

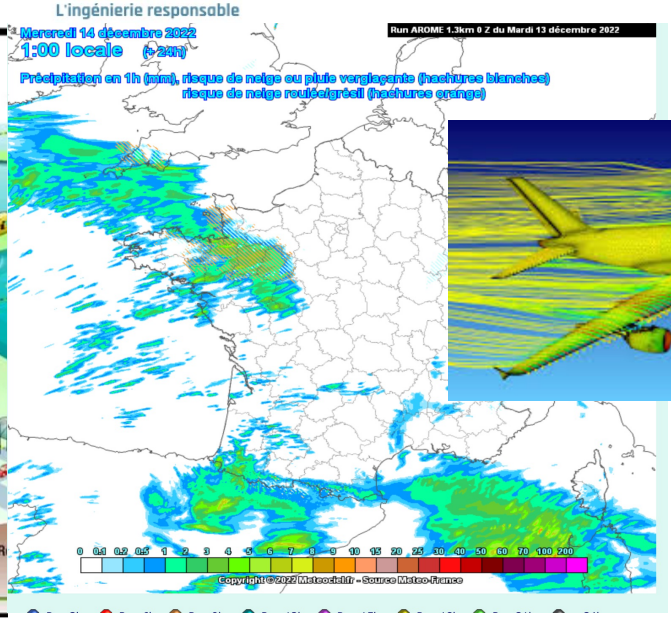
Groupement d'intérêt scientifique

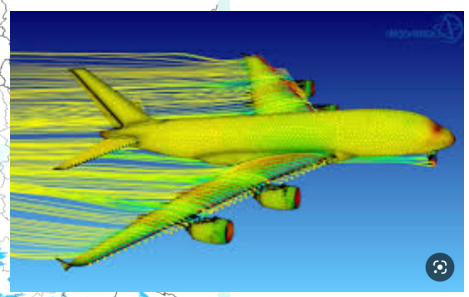
Quantification des incertitudes @ Université Paris-Saclay | LARTISSTE

@paris-saclay

CERFACS · IFP Energies nouvelles · LNE · ONERA · SystemX · CLERMONT AUVERGNE INP · sigma CLERMONT · AIRBUS · EDF · framatome · PHIMECA L'ingénierie responsable · SAFRAN







- Plus societal issues on algorithm biases credit score, job application etc…

Conclusion

- AI/ML used more and more in HEP measurements
- To reach ultimate measurements need to account for systematic uncertainties at training stage already
- Fair Universe project attempt to build large public datasets and figure-of-merit for « typical » HEP analysis
- Opportunity of collaboration in DMLab, tbd