Machine Learning in High Energy Physics

part 1

David Rousseau @dhpmrou

LAL-Orsay, CNRS/IN2P3, Université Paris-Sud/Paris-Saclay

rousseau@lal.in2p3.fr

VOIR le boson de Higgs

Avant de le voir, on savait tout sur le boson de Higgs, sauf sa masse

Particule très instable (10^{-22}s), se désintègrant immédiatement en paire d'autres particules, de façon imprévisible (sauf en moyenne)

Probabilités de désintégration prédites pour une masse de 125 GeV

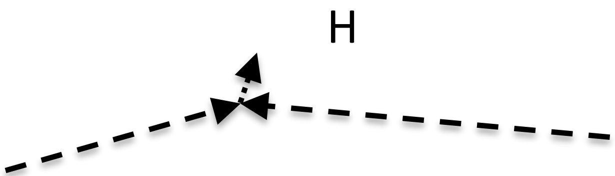
H → bb	58%
H → WW*	21%
H → τ+τ-	6.4%
H → ZZ*	2.7%
H → γγ	0.2%

$E=mc^2$

Einstein en 1905

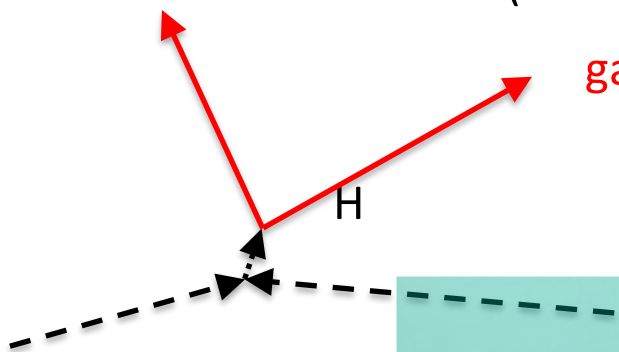
En fait, la formule complète est $E^2=p^2c^2+m^2c^4$
p est l'impulsion, mv en mécanique classique
En choisissant bien les unités, on se débarrasse de c:

$$\boxed{E^2=p^2+m^2}$$

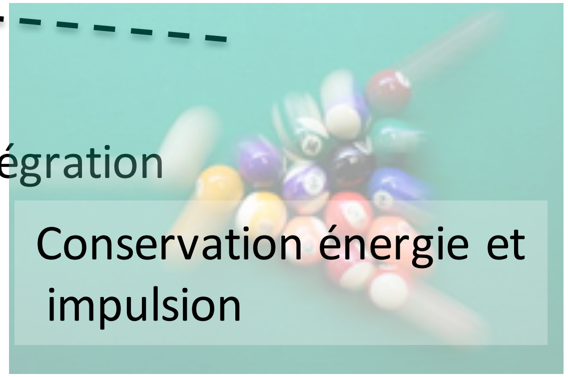


H, juste avant sa désintégration

$m_H^2=E_H^2-p_H^2$



Juste après sa désintégration



$E_H=E_{g1}+E_{g2}$

$\vec{p}_H=\vec{p}_{g1}+\vec{p}_{g2}$

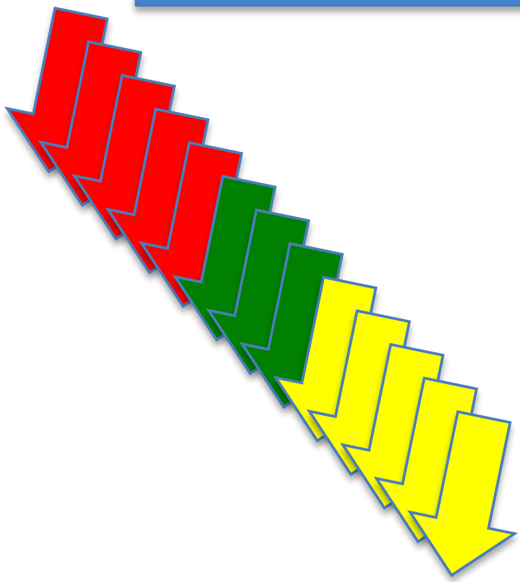
⇨on en déduit m_H!

Finalement…

10^{14} collisions

Tri rapide et grossier

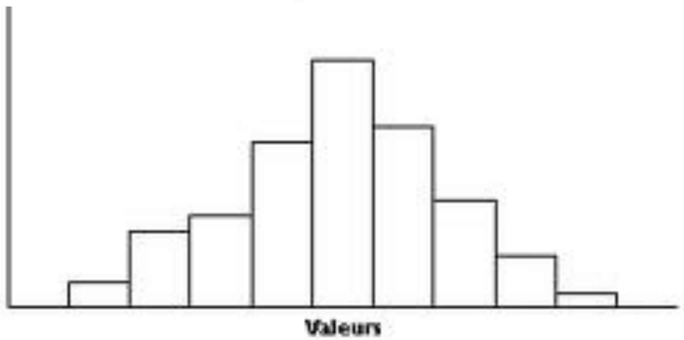
10^{9} événements sur disque



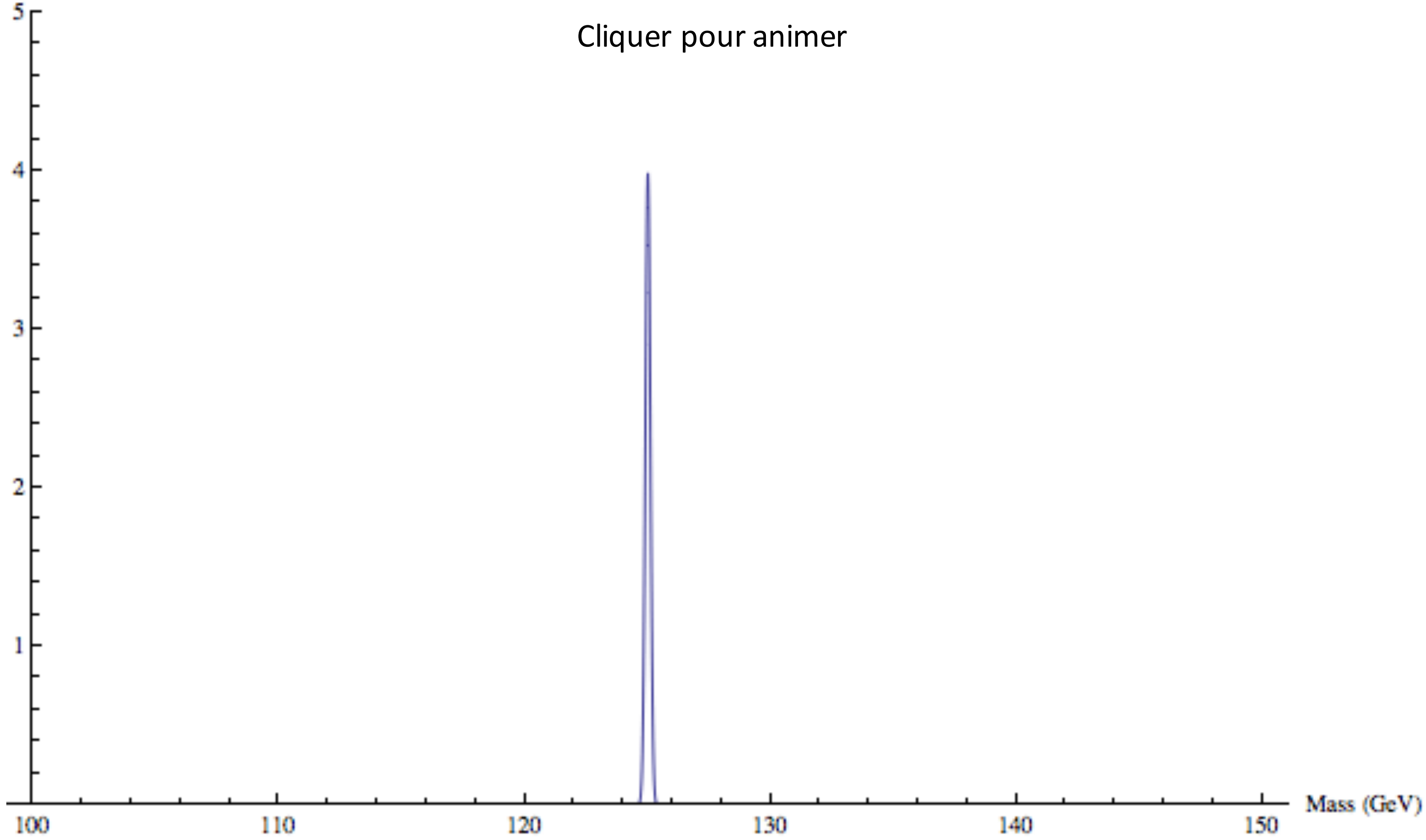
Tri précis

10^{5} événements à 2 gamma

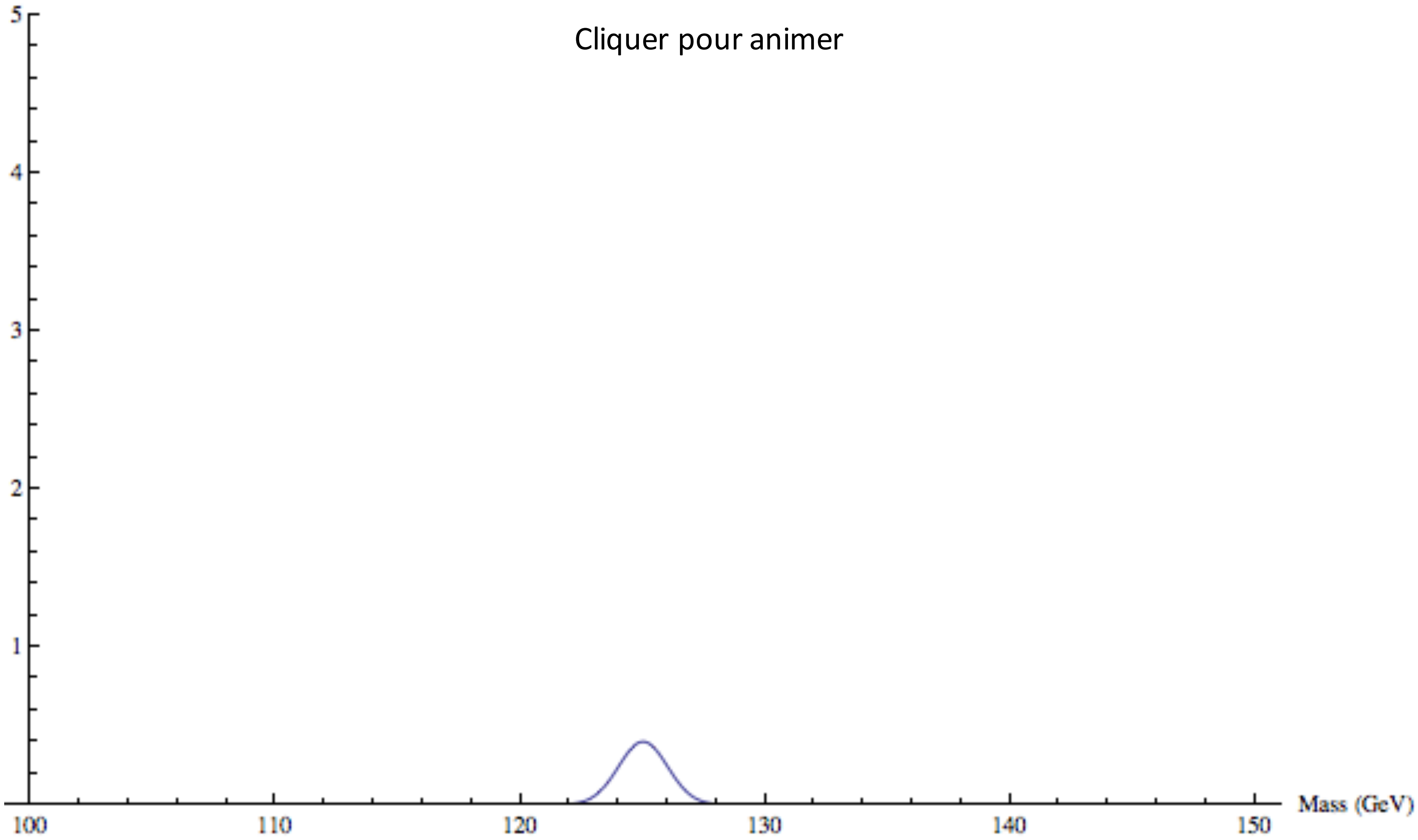
Calcul de la masse
→histogramme



Effet de la précision du détecteur

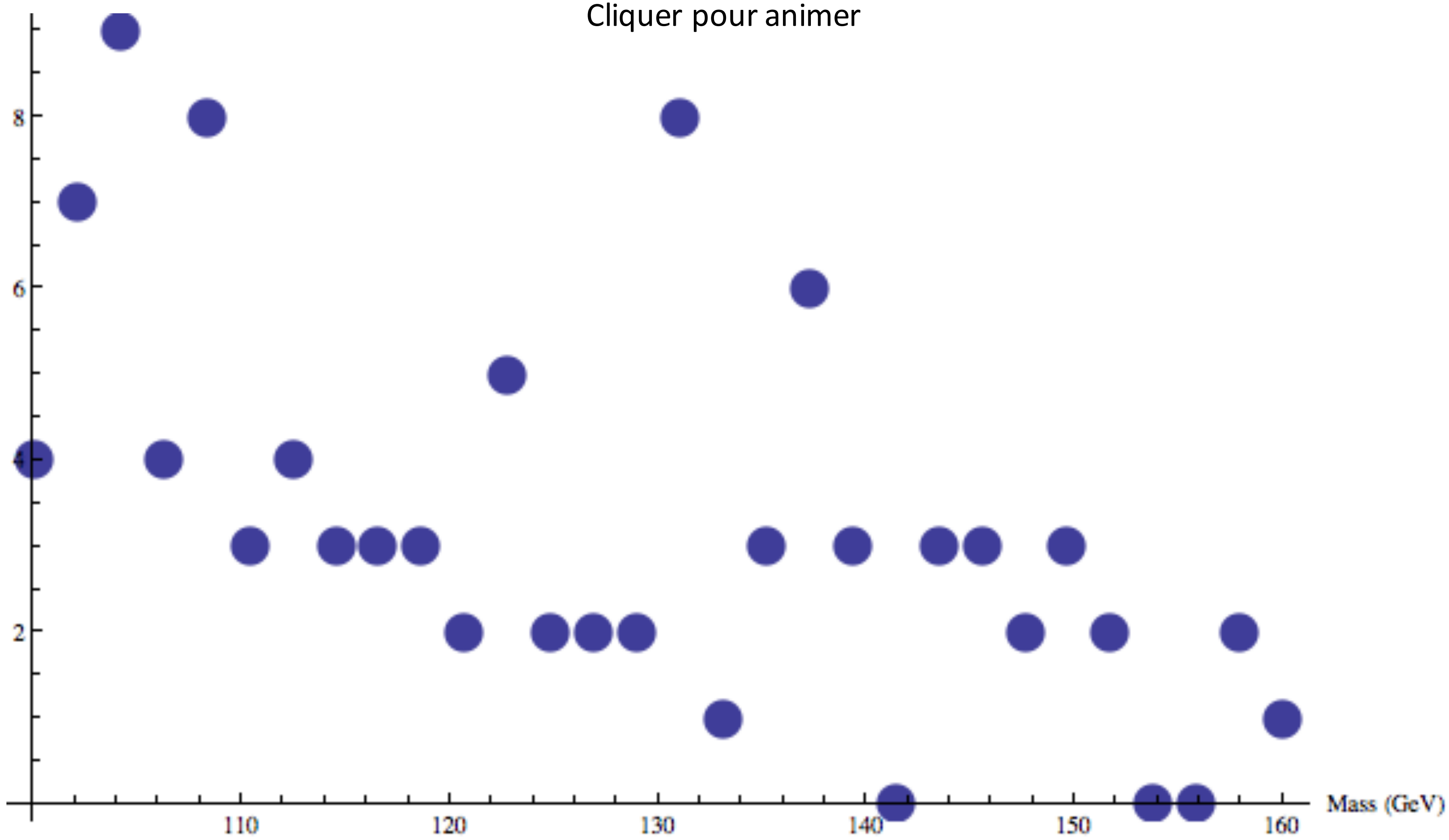


Effet du bruit de fond

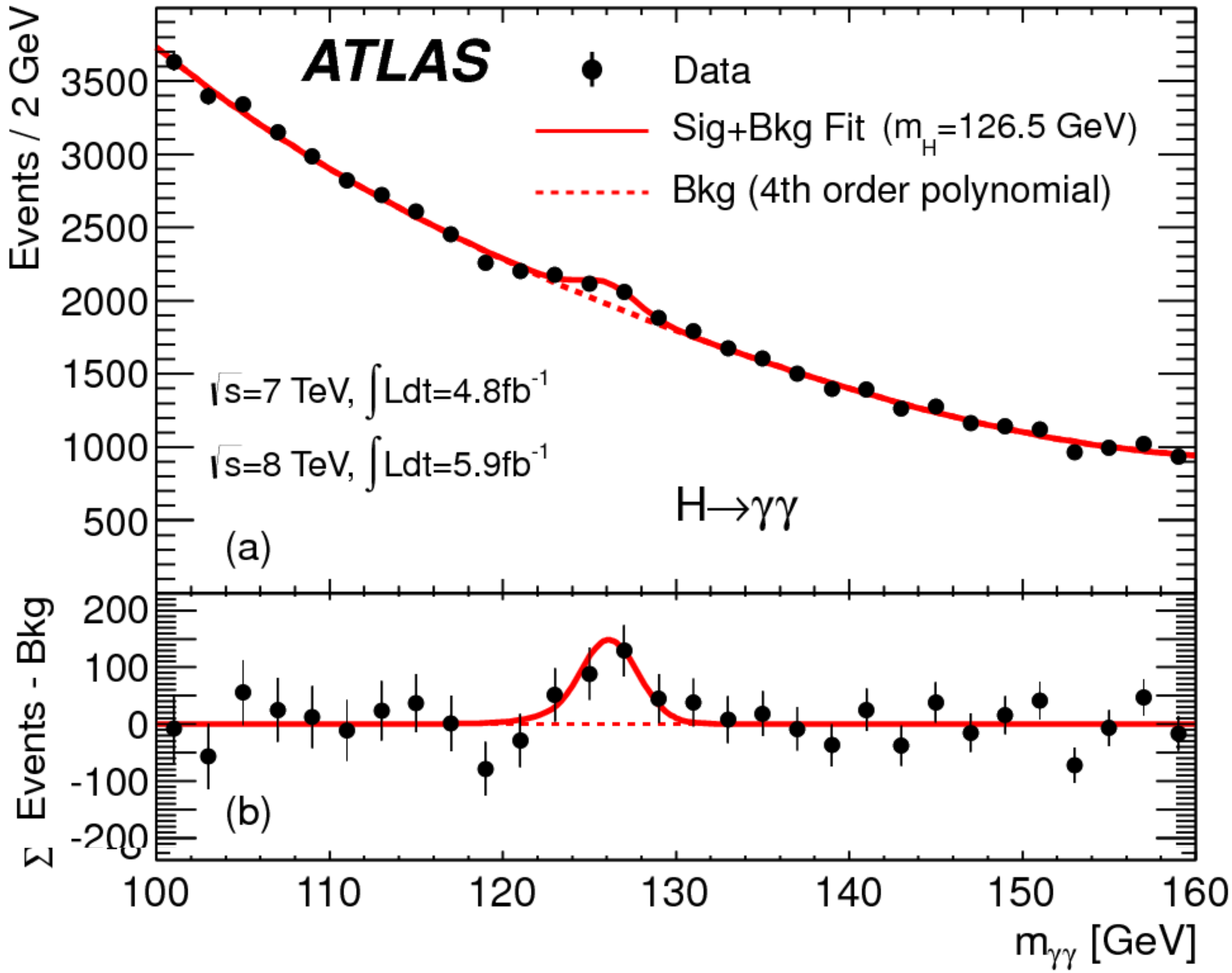


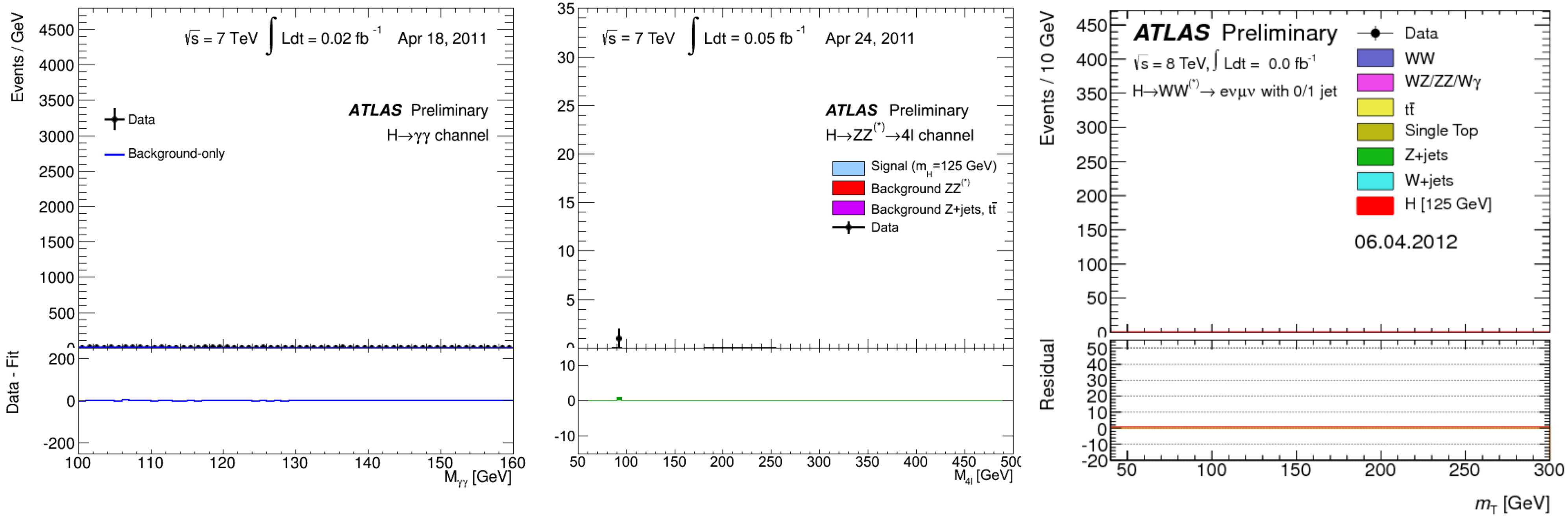
Effet de la statistique

Cliquer pour animer



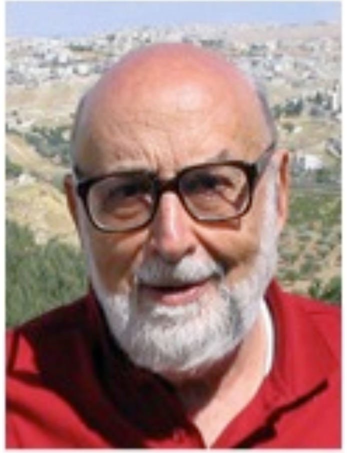
Et maintenant « en vrai » (Juillet 2012)



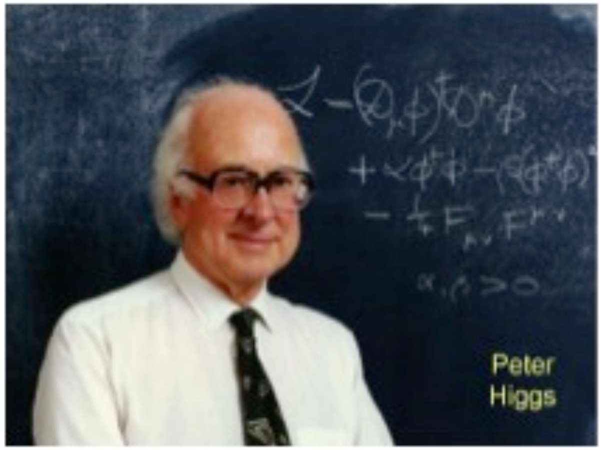


√s = 7 TeV ∫ Ldt = 0.02 fb⁻¹ Apr 18, 2011
ATLAS Preliminary
H→γγ channel
Data
Background-only
Events / GeV
Data - Fit
M_γγ [GeV]
√s = 7 TeV ∫ Ldt = 0.05 fb⁻¹ Apr 24, 2011
ATLAS Preliminary
H→ZZ(*)→4l channel
Signal (m_H=125 GeV)
Background ZZ(*)
Background Z+jets, tt̄
Data
M_4l [GeV]
ATLAS Preliminary
√s = 8 TeV, ∫ Ldt = 0.0 fb⁻¹
H→WW(*)→ eνμν with 0/1 jet
Data
WW
WZ/ZZ/Wγ
tt̄
Single Top
Z+jets
W+jets
H [125 GeV]
06.04.2012
Events / 10 GeV
Residual
m_T [GeV]

Robert Brout 1928-2011



François Englert 1932-



Peter Higgs 1929-

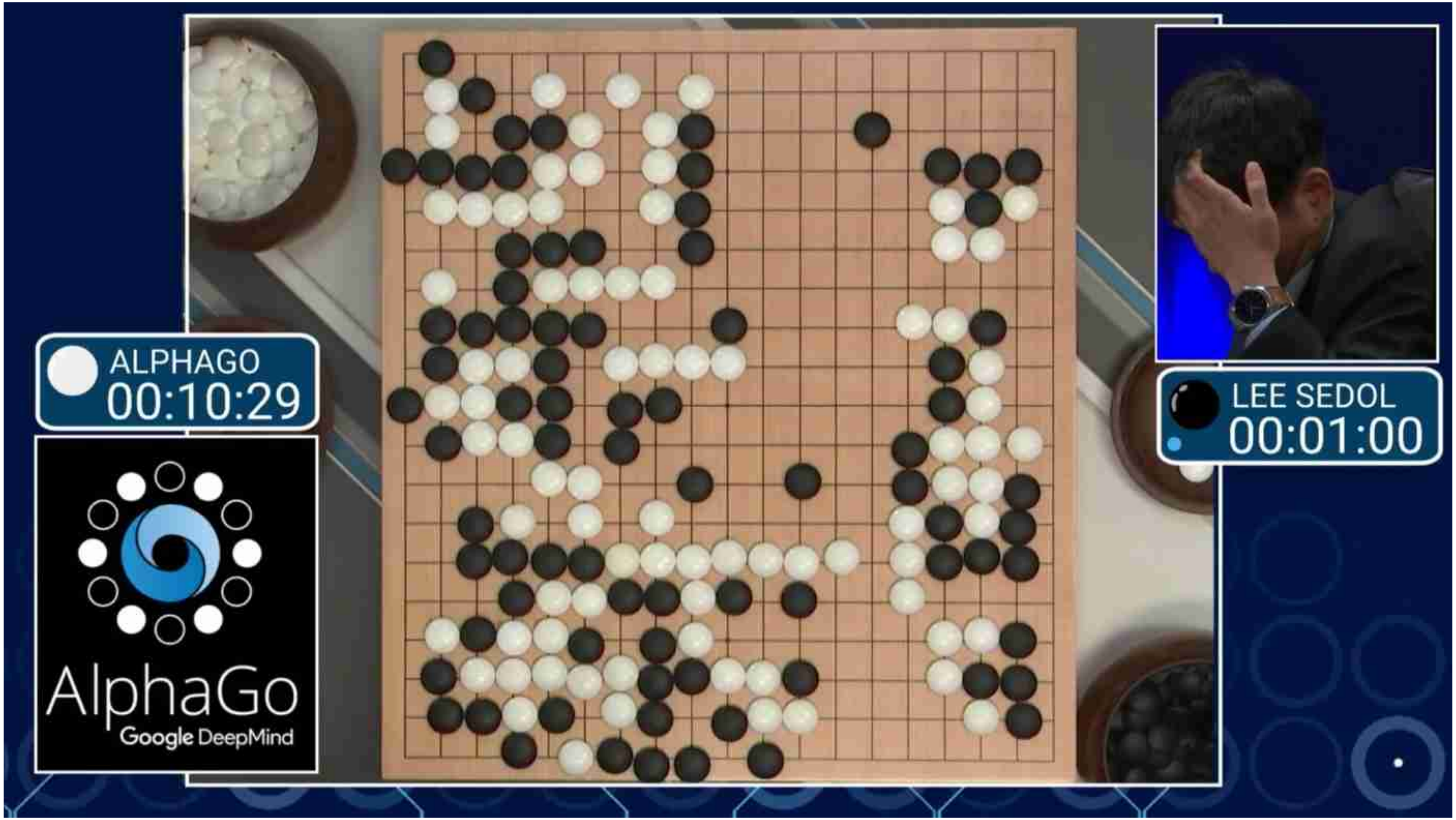
Également : G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble,

Champ de Higgs

« pour la découverte théorique d'un mécanisme qui contribue à notre compréhension de l'origine de la masse des particules subatomiques, qui a récemment été confirmé par la découverte de la particule fondamentale prédite, par les expériences ATLAS et CMS au grand collisionneur de hadrons (LHC) du CERN »

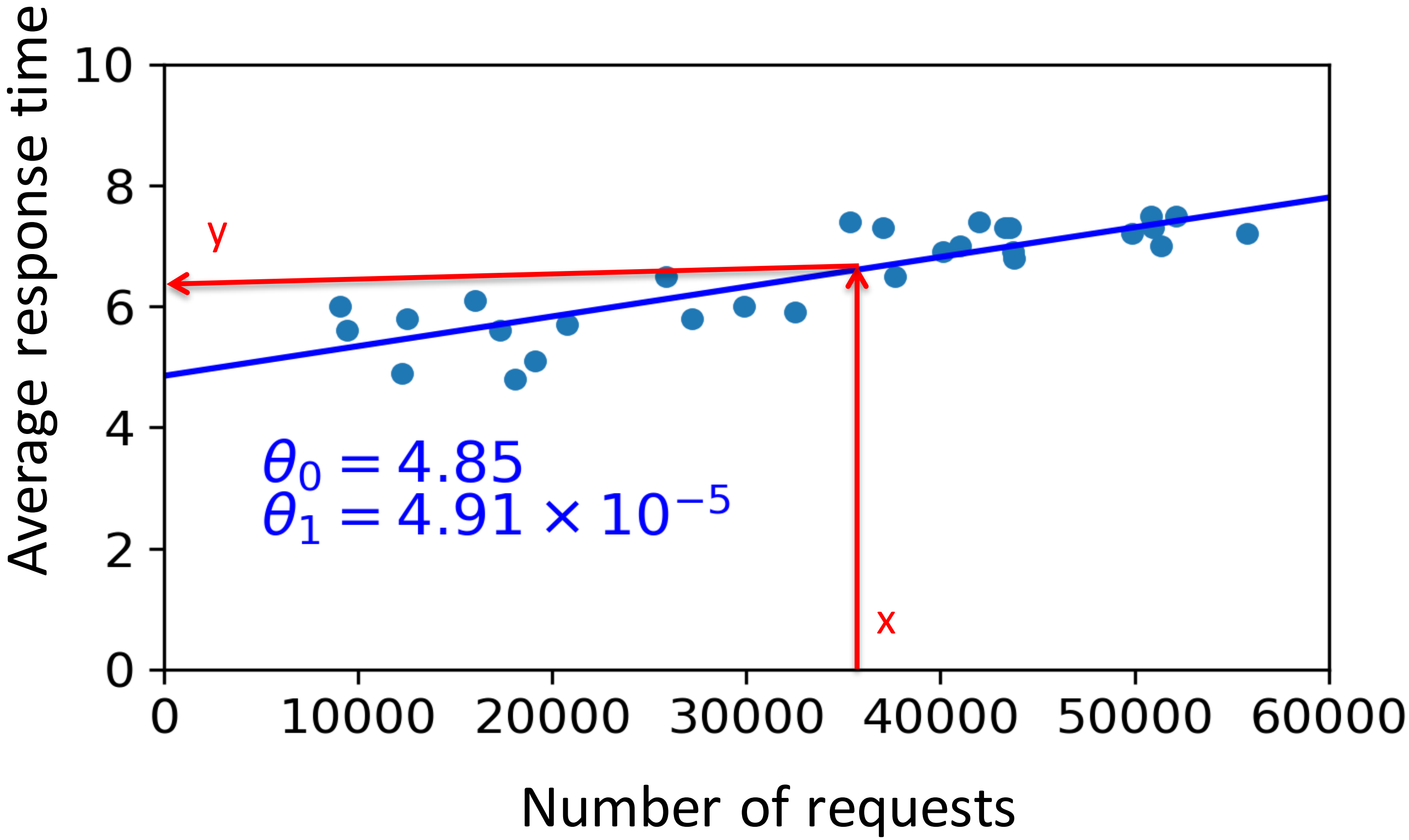
Boson de Higgs

Machine Learning


ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

Linear Regression

Boskovic, Legendre, Laplace, Gauss

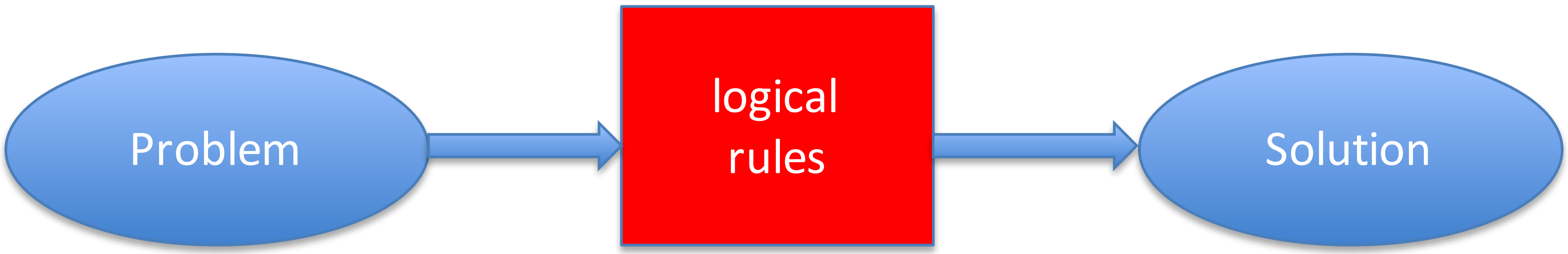


y=f(x)

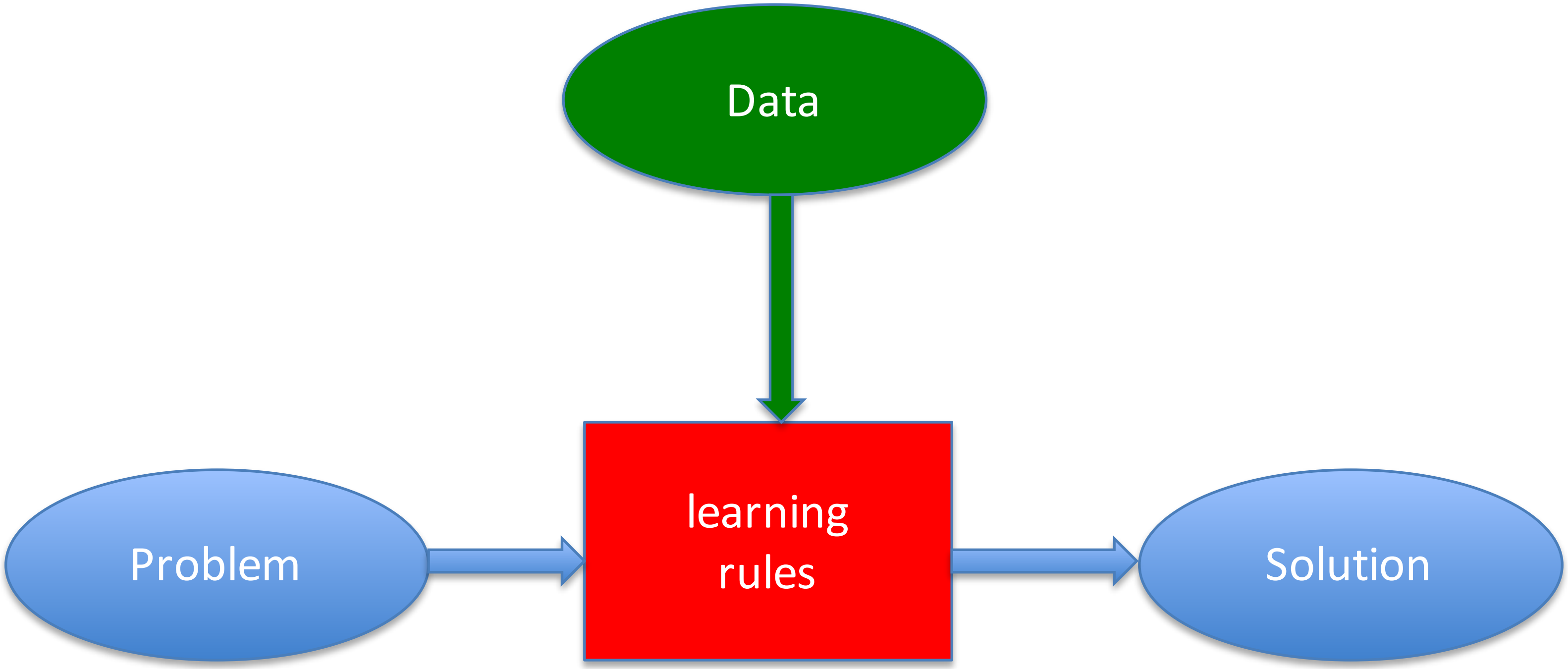
Given x, we want y → how to build f ?

X	f	Y
• Handwritten text	→	text
• Picture	→	Sofia or Sabrina ?
• Image	→	cat or dog ?
• « Comment ça va ? »	→	« Wie geht's ? »
• Speech	→	text
• Chess board	→	next move
• Camera + capteurs+GPS	→	wheel action
• facebook data	→	publicity

Traditional Computing

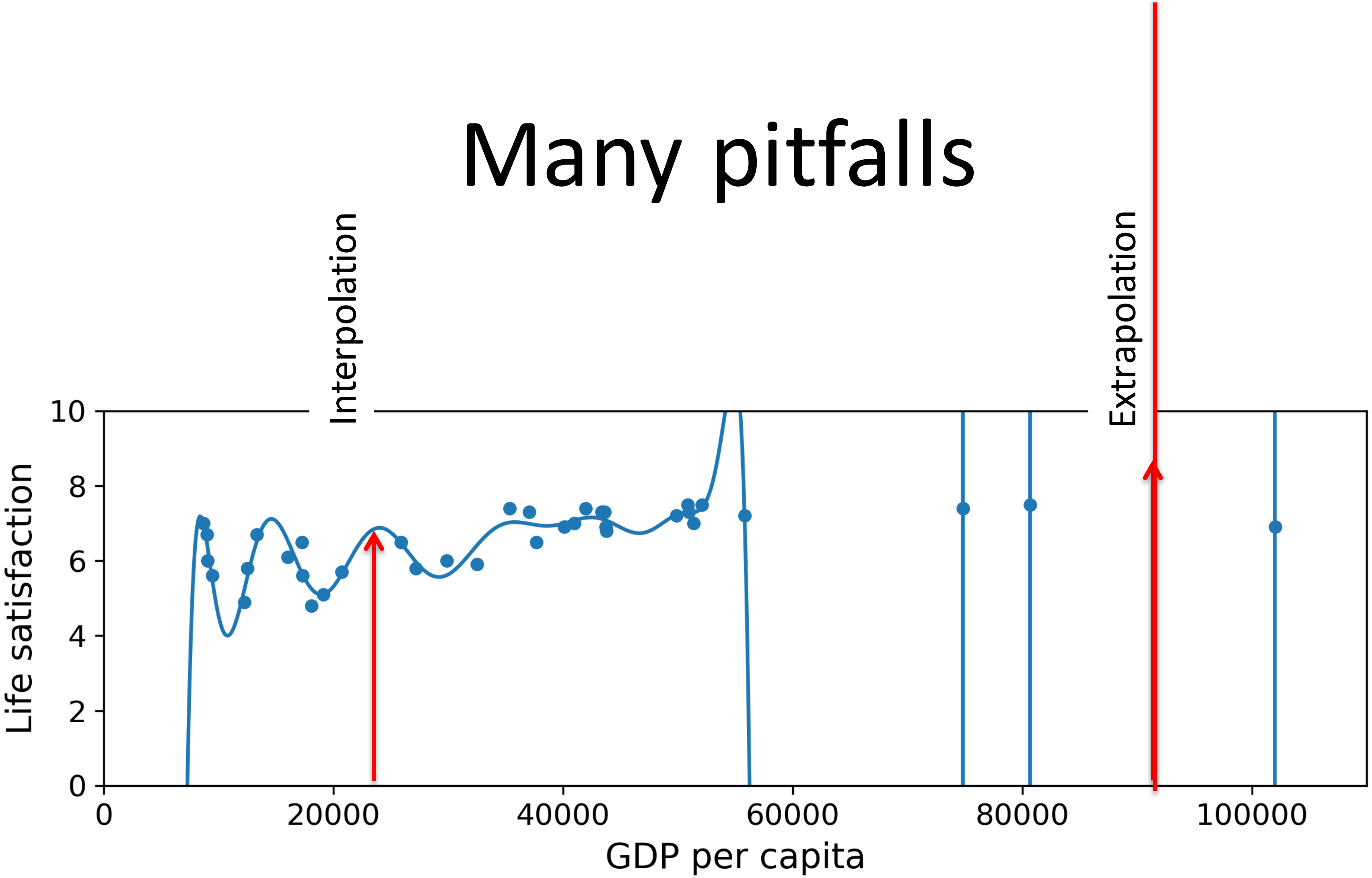


Apprentissage Automatique



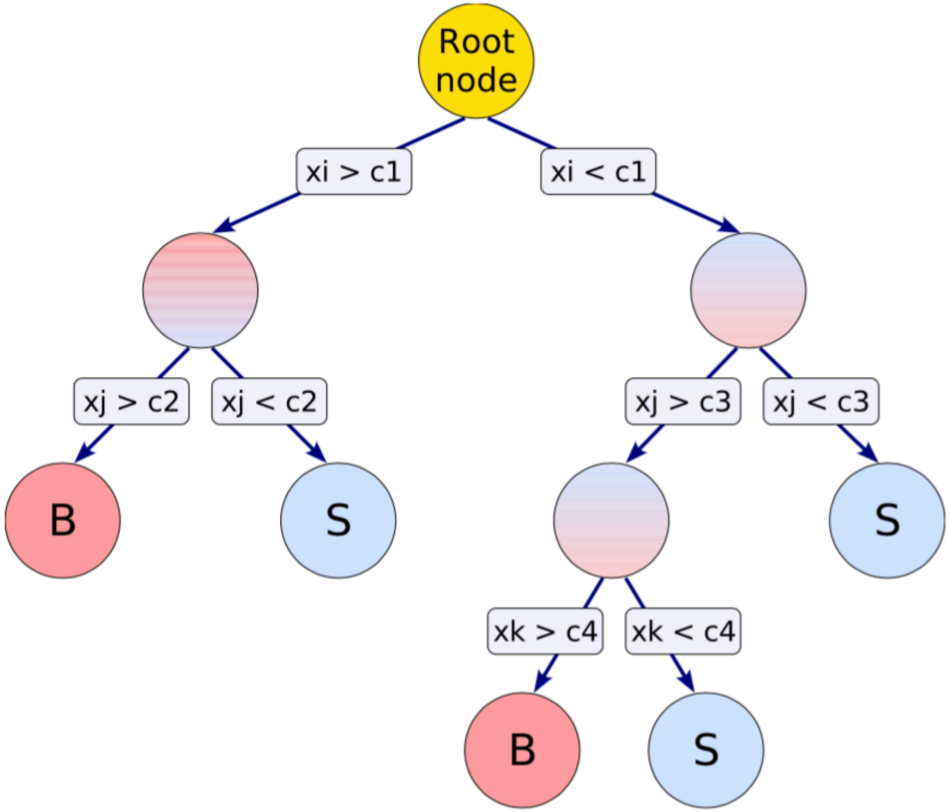
Learning = optimise internal parameters of the algorithms: n=2 - millions

Many pitfalls



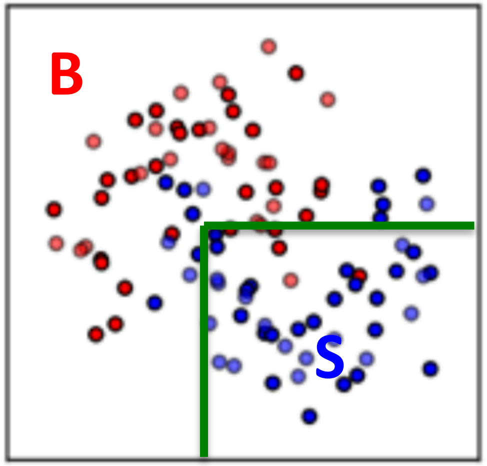
How does it work?

Boosted Decision Tree

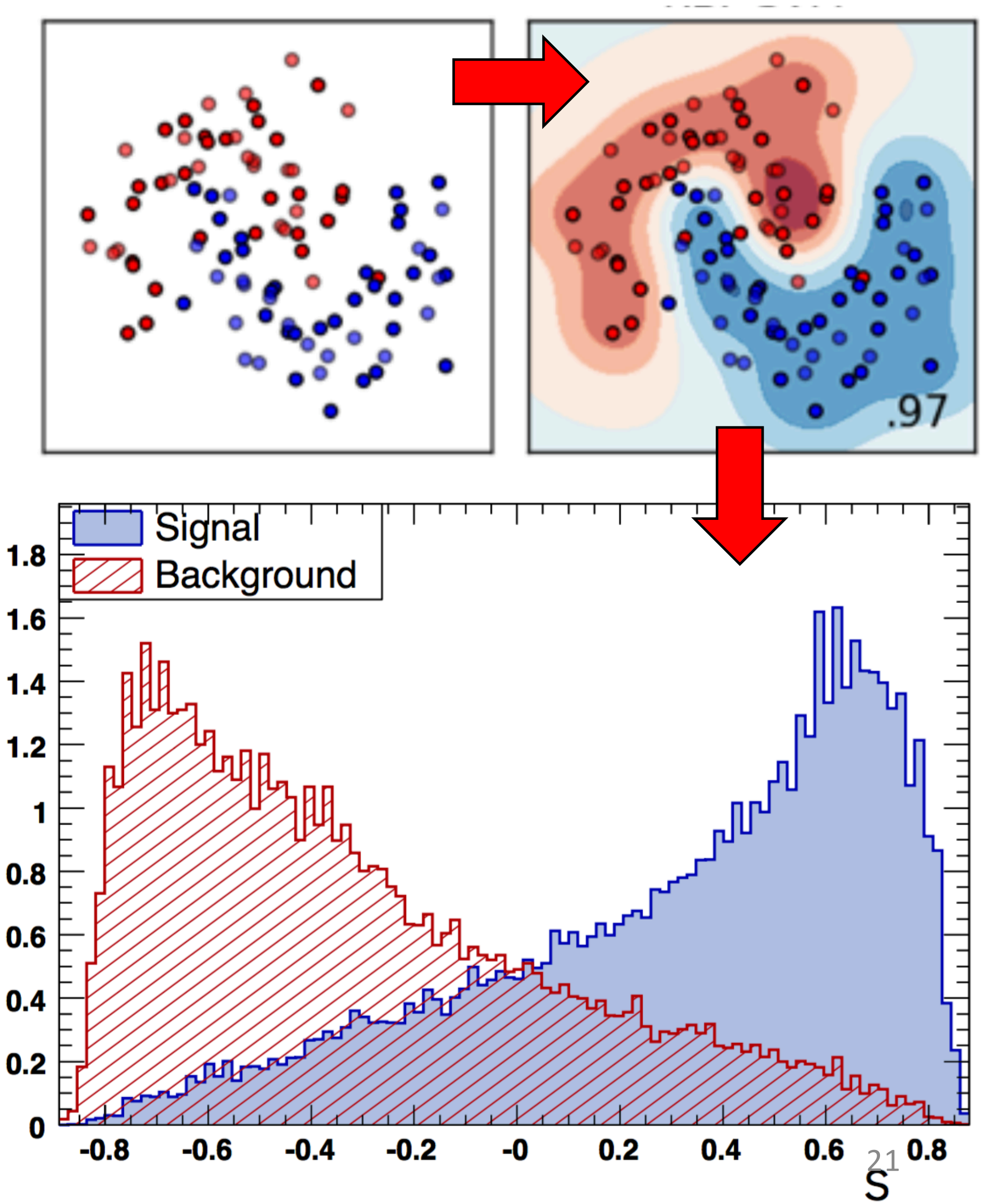


- Single tree (CART) <1980
- AdaBoost 1997 : rerun increasing the weight of misclassified entries
  ➔Boosted Decision Trees (Gradient BDT XGBoost, random forest…)

Classifier

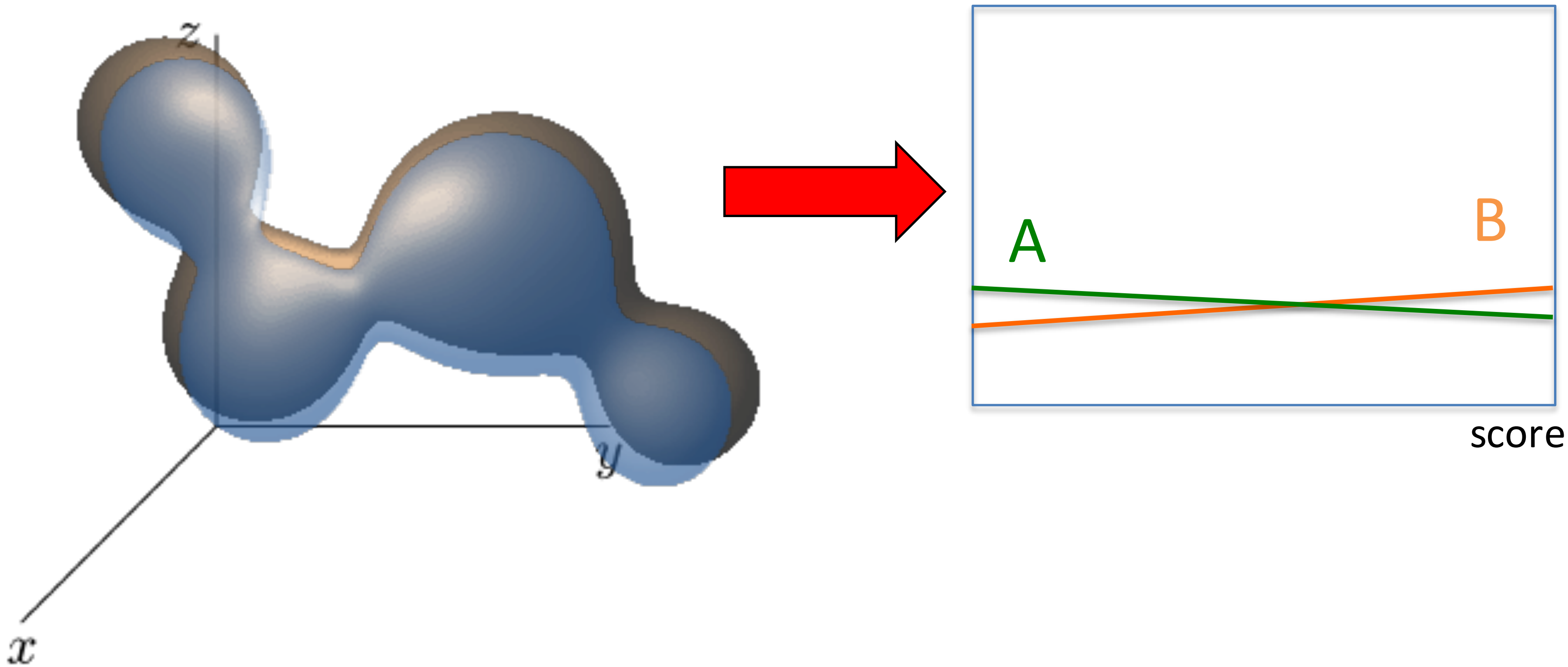


B
S



.97
Signal
Background
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
-0.8
-0.6
-0.4
-0.2
-0
0.2
0.4
0.6
0.8
S

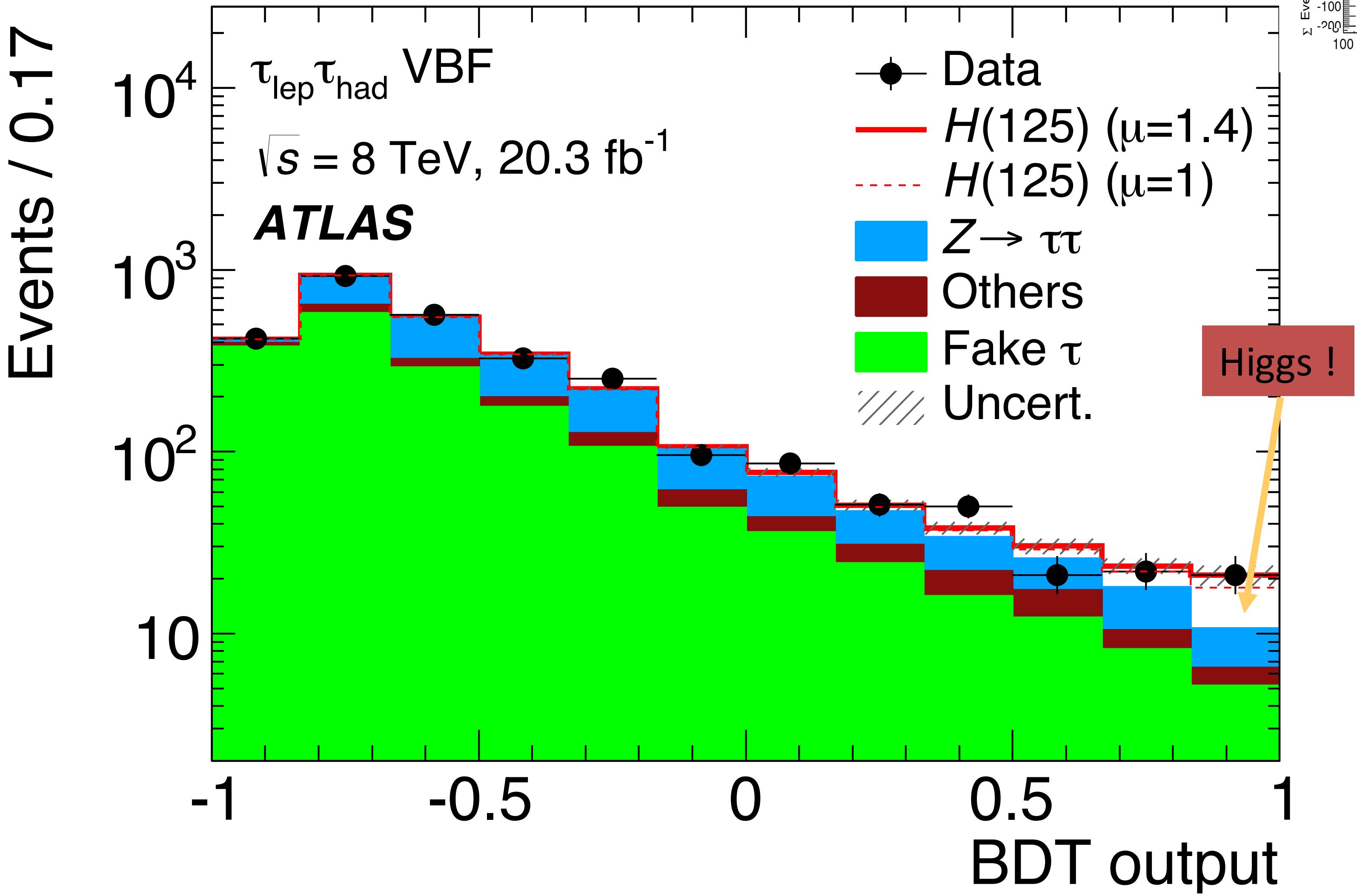
What does a classifier do ?

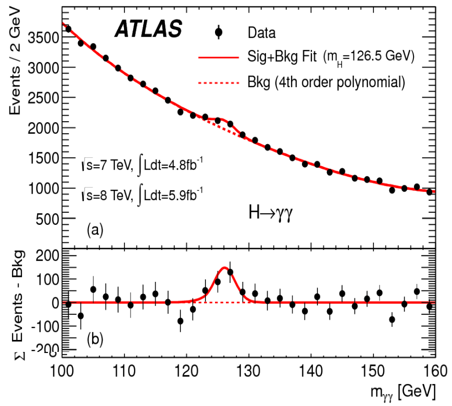


- A classifier "projects" the two multidimensionals "blobs", while maximising their difference

Application Higgs→$\tau^+\tau^-$

BDT sur ~10 variables : invariant mases, angles, etc…

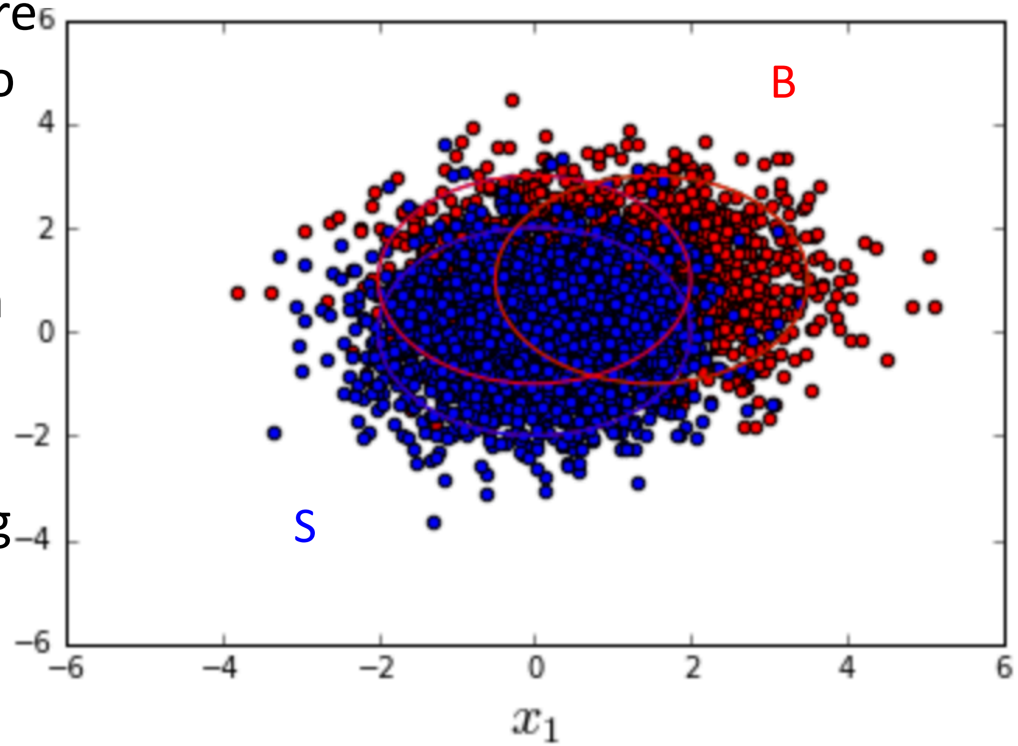




JHEP 04, 117 (2015) 1501.04943

No miracle

- If the probability density functions are known, nothing beats likelihood ratio (Neyman-Pearson theorem) :
  - $L_S(x)/L_B(x)$
- ❑ OK but in general L_S L_B are not known (even the shape is not known)
  - ❑ + x has large dimensions
- Only then ML is becoming interesting



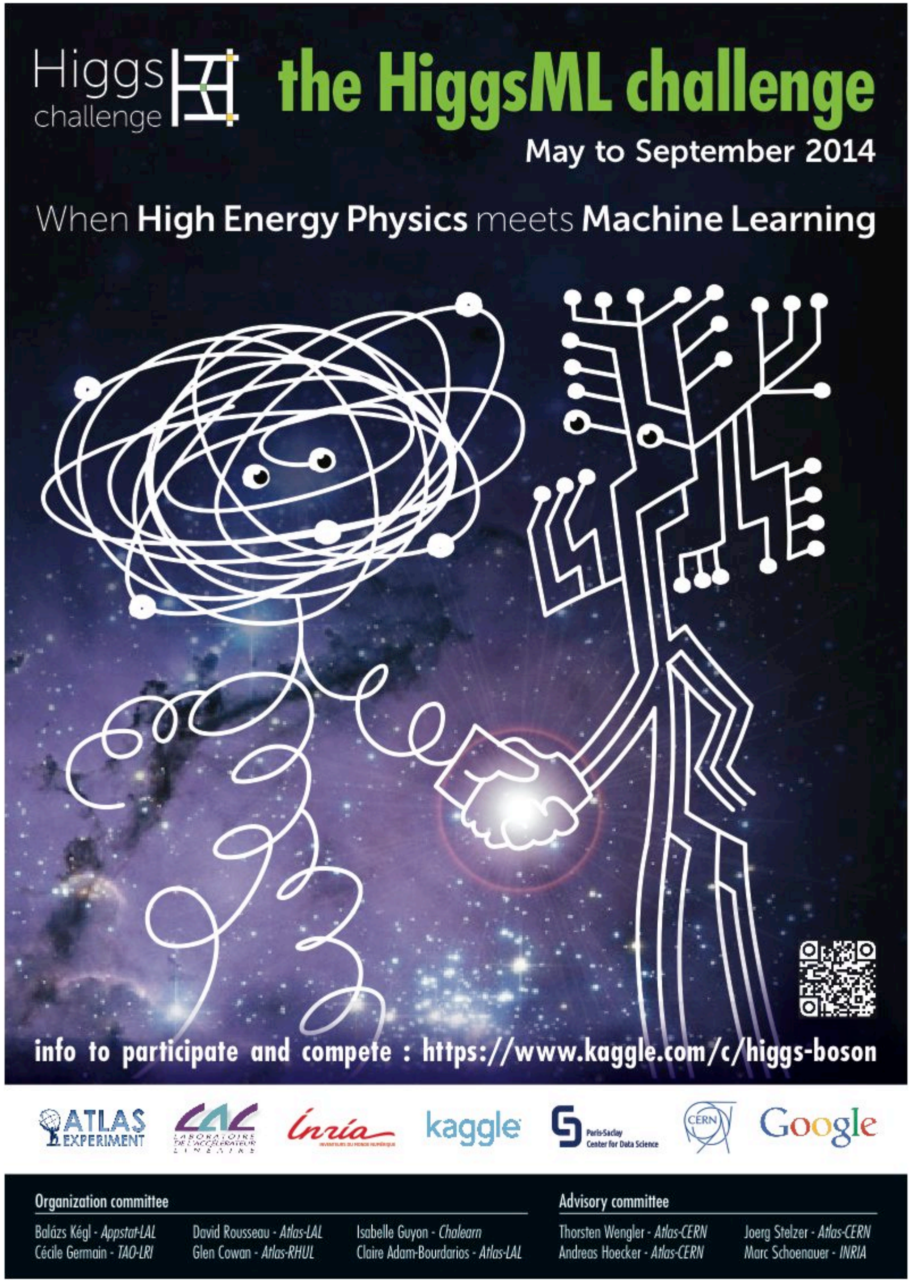
- Note : if something is known, better "tell" the ML algorithm about it
  - e.g. azimuthal invariance, use phi difference rather than absolute
  - Invariant masses (or other relativistic quantities)

- At LHC, mostly using BDT on a dozen variables
- Impact on Higgs discovery potential

Analysis	Data taking year	No ML sensitivity	ML sensitivity	Relative data gain
CMS $H \rightarrow \gamma\gamma$	2011-2012	2.2	2.7	51%
ATLAS $H \rightarrow \tau^+\tau^-$	2011-2012	2.5	3.4	85%
ATLAS $VH \rightarrow bb$	2011-2012	1.9	2.5	73%
ATLAS $VH \rightarrow bb$	2015-2016	2.8	3.0	15%
CMS $VH \rightarrow bb$	2011-2012	1.4	2.1	125%

➔equivalent to ~50% more data

Higgs Machine Learning challenge



Pitch : simplify a Higgs analysis, post the simulated data on the web, ask Computer Scientists to improve

Completed • $13,000 • 1,785 teams

Higgs Boson Machine Learning Challenge

Mon 12 May 2014 – Mon 15 Sep 2014 (34 days ago)

Dashboard

Private Leaderboard - Higgs Boson Machine Learning Challenge

This competition has completed. This leaderboard reflects the final standings.

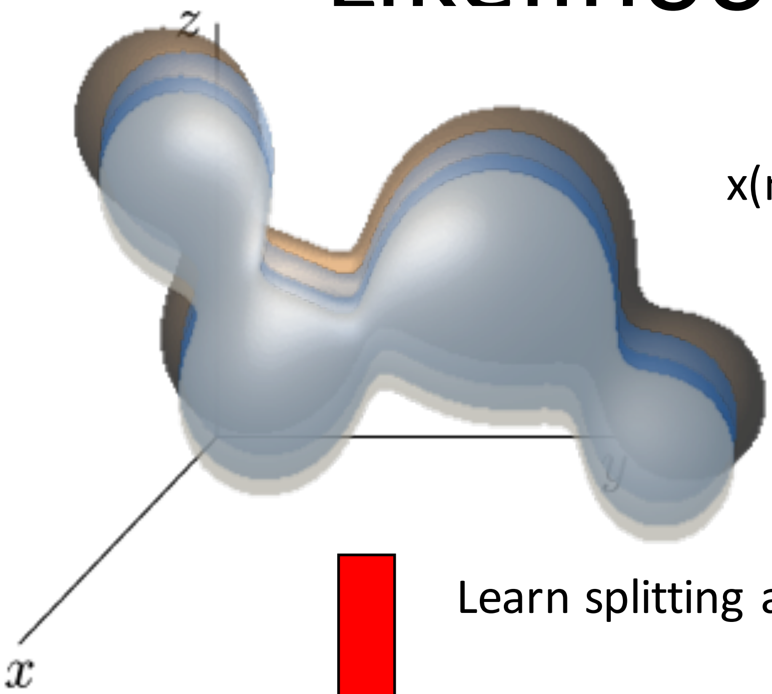
See someone using multiple accounts? Let us know.

Ecart standard

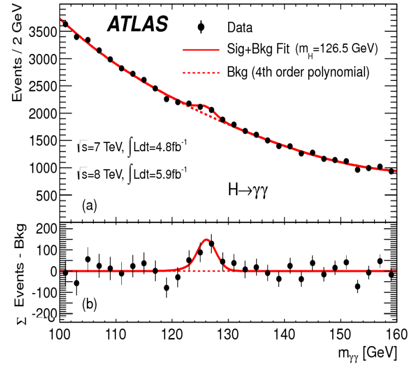
#	Δ1w	Team Name ‡ model uploaded * in the money		Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑4	Gábor Melis ‡ *	7000$	3.80581	110	Sun, 14 Sep 2014 09:10:04 (-0h)
2	↓1	Tim Salimans ‡ *	4000$	3.78913	57	Mon, 15 Sep 2014 23:49:02 (-40.6d)
3	—	nhlx5haze ‡ *	2000$	3.78682	254	Mon, 15 Sep 2014 16:50:01 (-76.3d)
4	↑55	ChoKo Team		3.77526	216	Mon, 15 Sep 2014 15:21:36 (-42.1h)
5	↑23	cheng chen		3.77384	21	Mon, 15 Sep 2014 23:29:29 (-0h)
6	↓2	quantify		3.77086	8	Mon, 15 Sep 2014 16:12:48 (-7.3h)
7	↑73	Stanislav Semenov & Co (HSE Yandex)		3.76211	68	Mon, 15 Sep 2014 20:19:03
8	↓1	Luboš Motl's team		3.76050	589	Mon, 15 Sep 2014 08:38:49 (-1.6h)
9	↓1	Roberto-UCIIIM		3.75864	292	Mon, 15 Sep 2014 23:44:42 (-44d)
10	↑5	Davut & Josef		3.75838	161	Mon, 15 Sep 2014 23:24:32 (-4.5d)

991 TMVA (CERN standard) 3.2

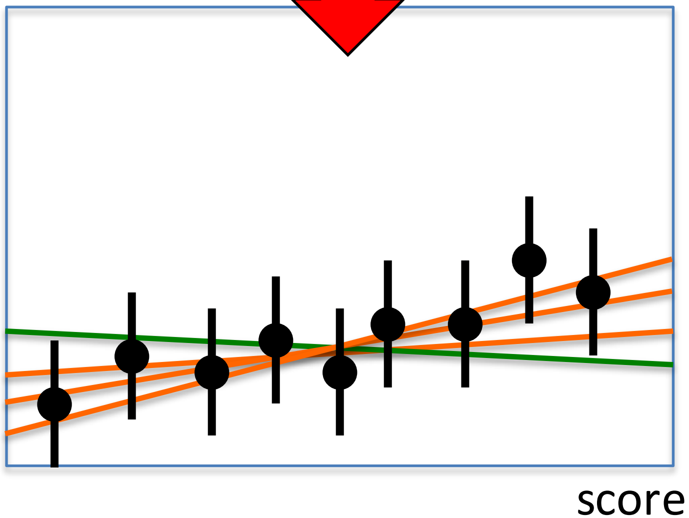
Likelihood-free inference



x(m)



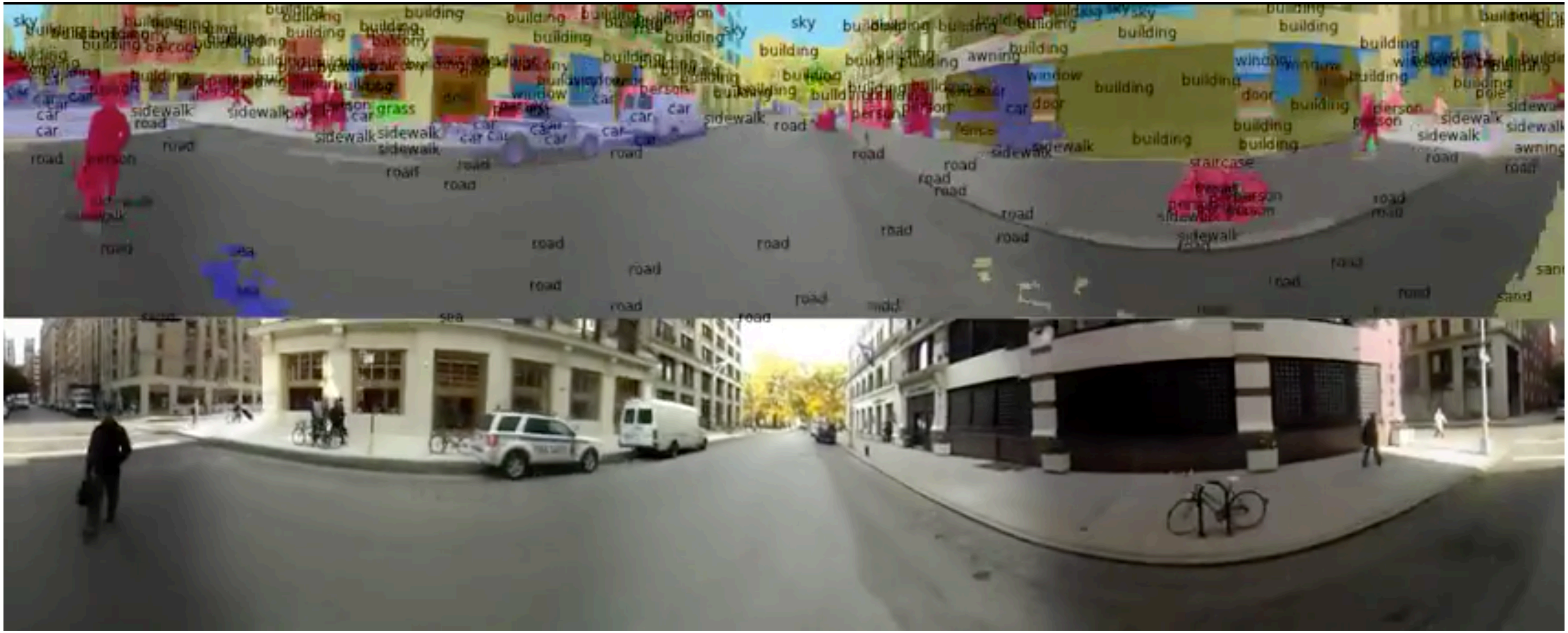
Learn splitting as a function of m



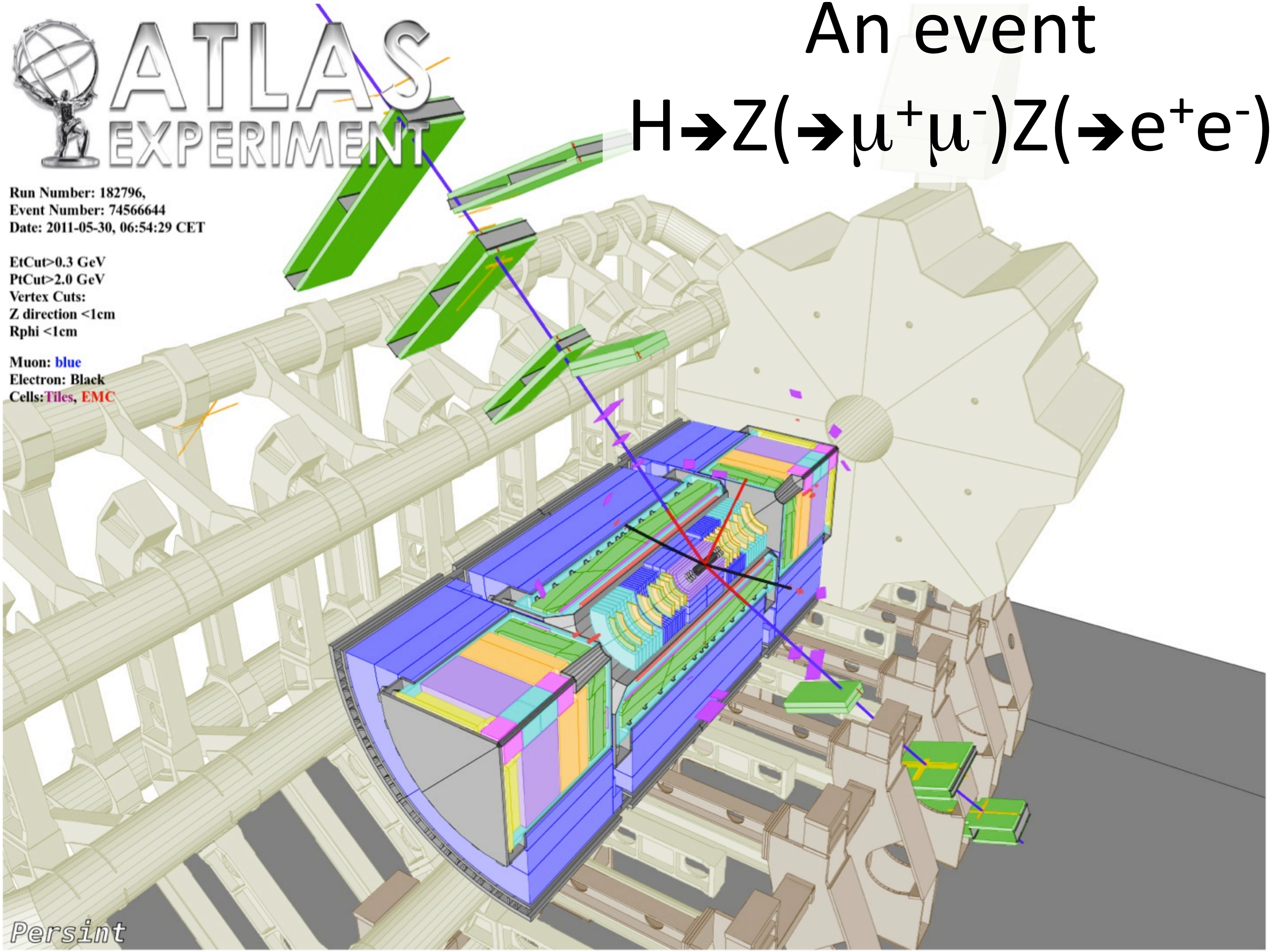
Fit m on data

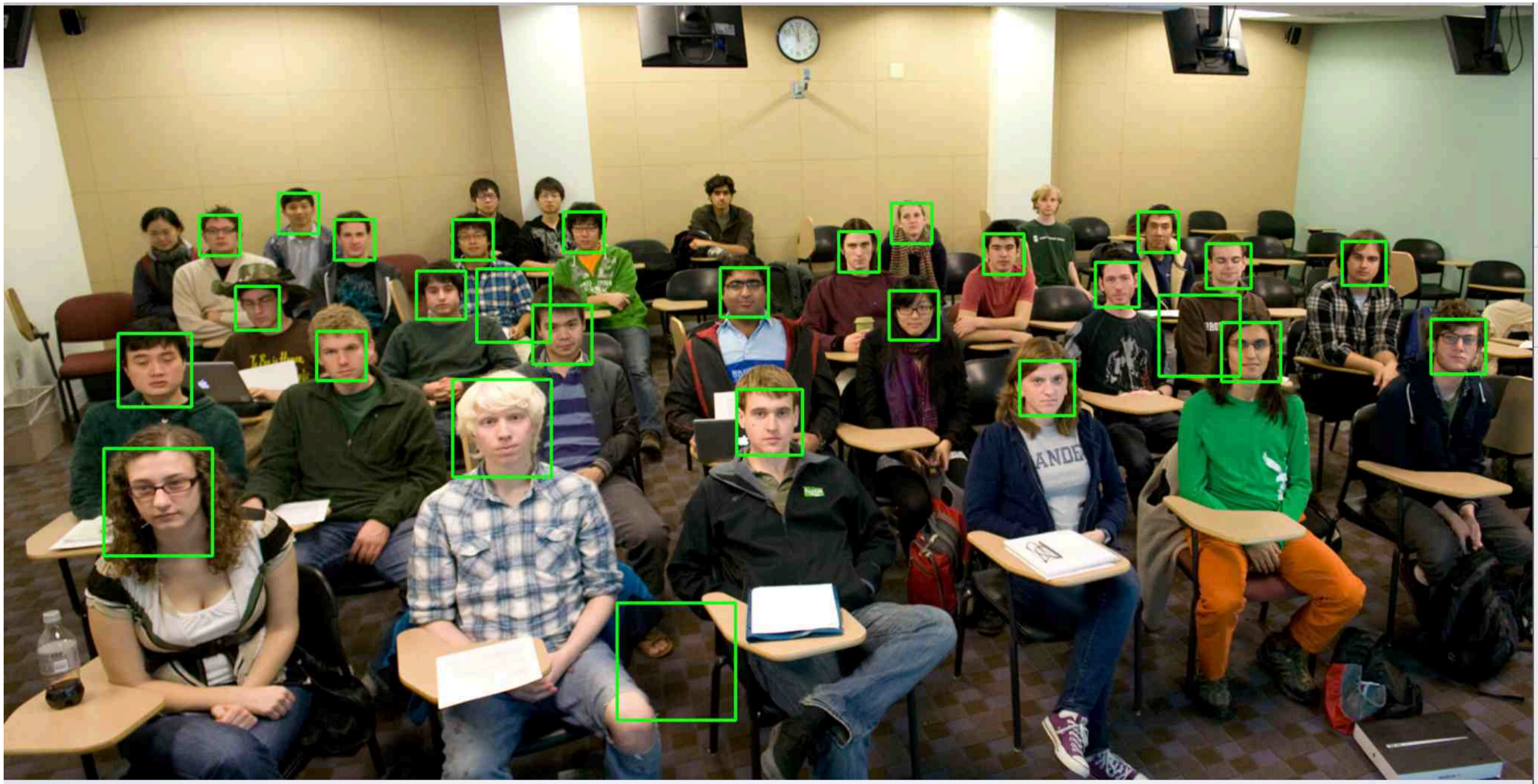
Identification

Typical Deep Learning application



An event
$H \rightarrow Z(\rightarrow \mu^+\mu^-)Z(\rightarrow e^+e^-)$





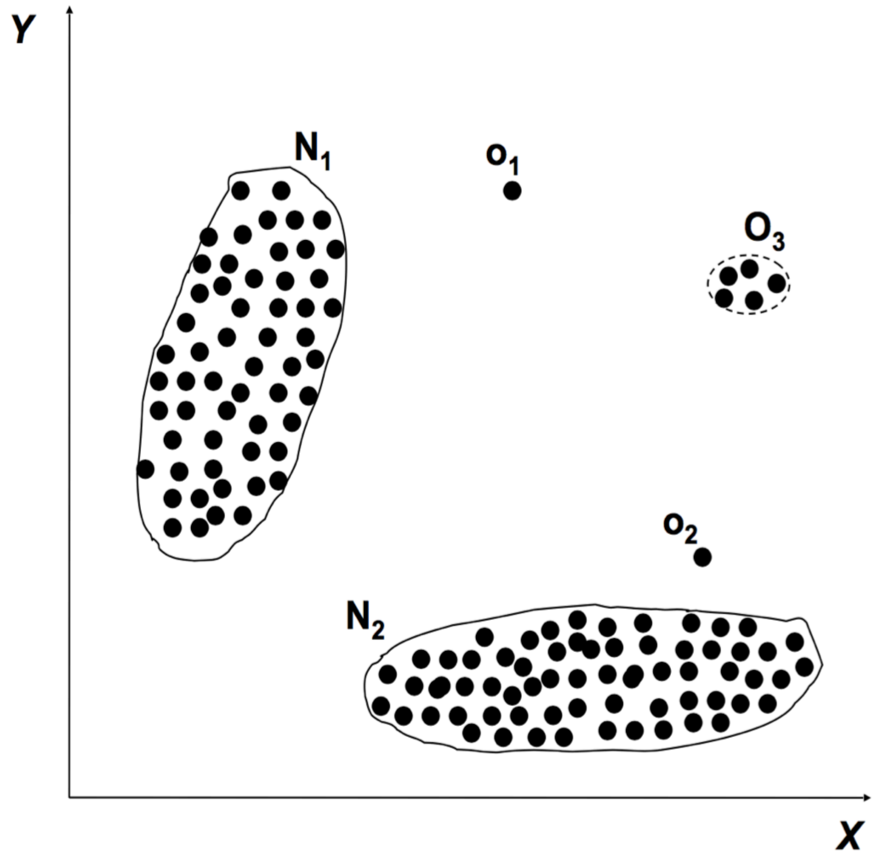
Efficiency
Purity
Speed

Anomaly Detection

Anomaly Detection

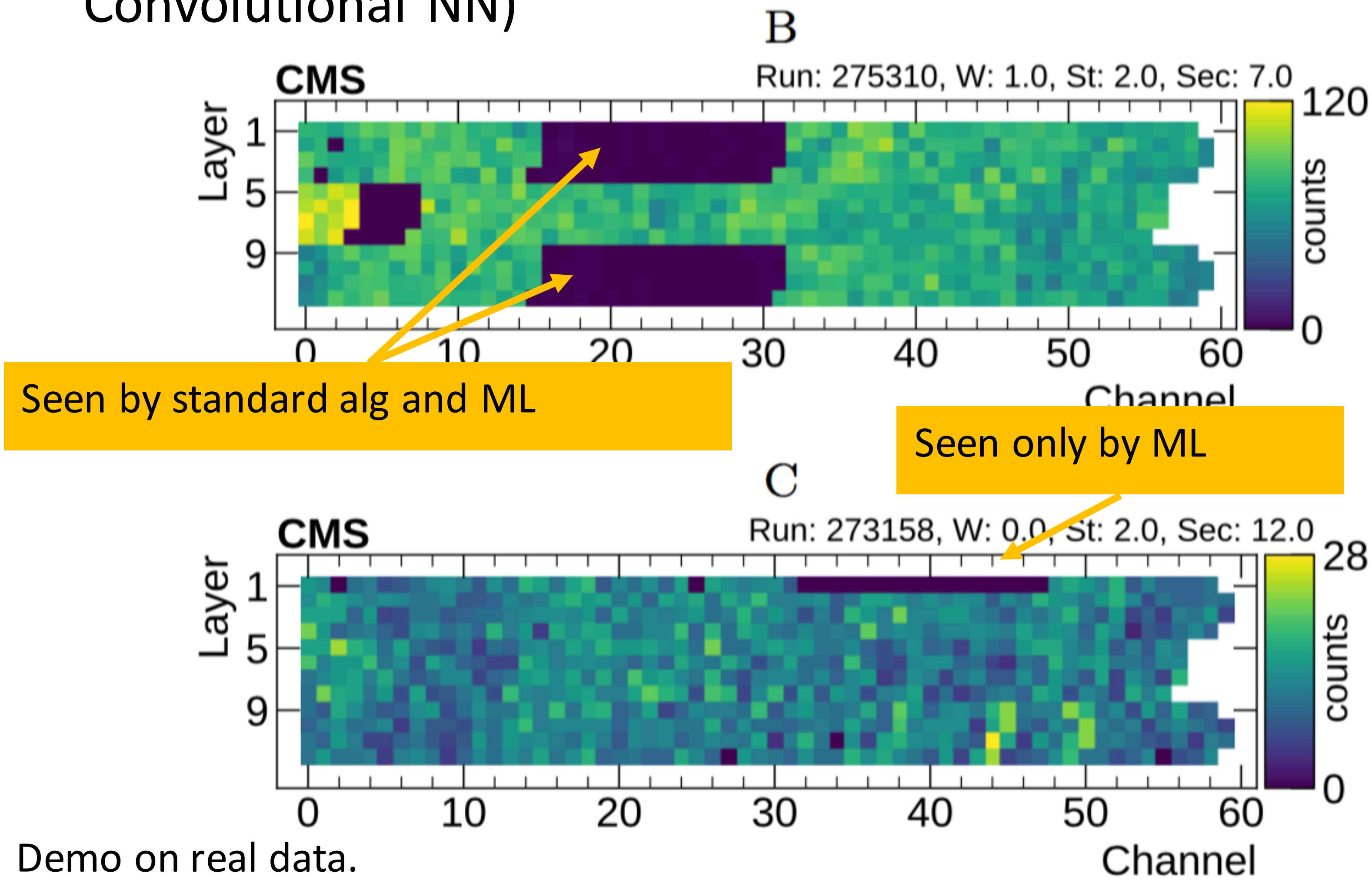
- How to detect anomalous "O"utliers with respect to "N"uclei
- Supervised method : we have a model for both N and O
- Unsupervised method : no models
- Semi-supervised method : model for N but not for O

- Applications : spam detection, fault detection



Data Quality Monitoring

- Example application CMS muon chamber monitoring (with Convolutional NN)



Demo on real data.

Application to physics

Application to new physics

Break