Corso monografico di Fisica avanzata Use of data in high energy physics [when/if claim a discovery]



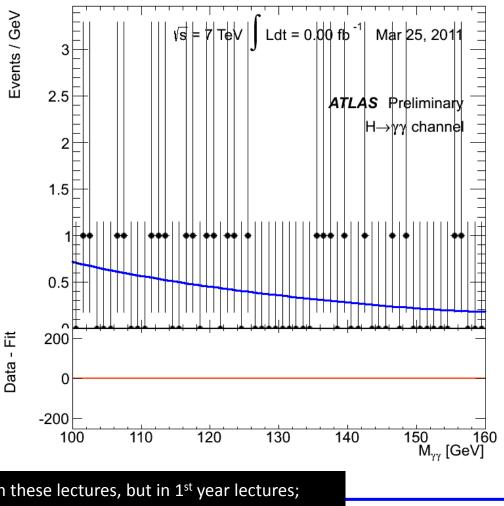
Paolo Bagnaia SAPIENZA UNIVERSITÀ DI ROMA

the question

- look carefully to the figure —
- two superimposed plots :
 - 1. the spectrum of $m(\gamma\gamma)$ for CMS in 2012 data + onset;
 - 2. the same for ATLAS, animated as a function of the luminosity;
- similar plots are available for both ATLAS/CMS for all predicted channels of Higgs decays;

[https://twiki.cern.ch/twiki/bin/view/AtlasPublic/Higgs PublicResults#Animations http://cms.web.cern.ch/org/cms-higgs-results]

- the plots are shown as a proof of the "discovery" of the Higgs boson;
- WHY ??? HOW ???



- the underlying theory is NOT discussed in these lectures, but in 1st year lectures;
- warning the detectors and analysis methods are beyond their scope;
 - we only focus on the meaning / method of the "discovery claim" in HEP experiments.

the meaning of "discovery"

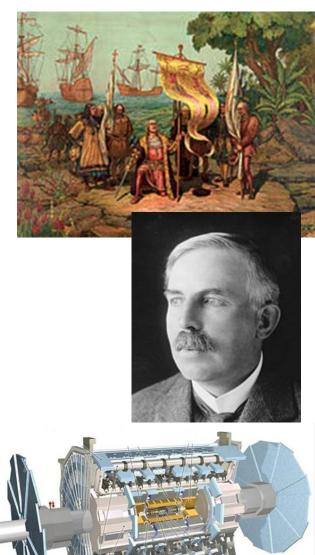
- Christopher Columbus discovered America in 1492.
- Ernest Rutherford discovered the atomic nucleus in 1909.
- The ATLAS and CMS experiments have discovered the Higgs boson in 2012.

In normal language, "discovery" means finding/establishing something (a land, a state of the matter, even a theory), which pre-existed, but was unknown to [the majority of] the human beings.

Sometimes the actual identification of the "discoverer" is difficult or ambiguous (Columbus, the Vikings, the natives), but the concept of "discovery" is not contested.

However, when the "land" is tiny and hidden, and quantum mechanical effects complicate the procedure, the method and the meaning of "discovery" demand further investigation.

Let us proceed slowly and cautiously.



discovery in classical world

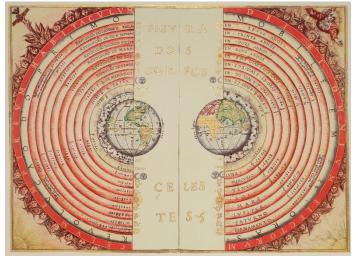
Remember the (in)famous debate ptolemaiccopernican.

- theory-1 (Ptolemy) had geocentric orbits with epicycles;
- observables (Tycho, Copernicus, Kepler) did not follow model's predictions;
- theory-2 with heliocentric elliptical orbits was in agreement with data;
- after lot of struggle theory-2 was accepted.

Actually theory 2 was NOT "verified", rather theory-1 was "falsified" [veri-falsified à la Popper].

[after few centuries, the story repeated with Einstein general relativity – *more noise and newspapers, but no burning at the stake,* however]

This procedure seems subtle [possibly contentious], but reasonably under control.



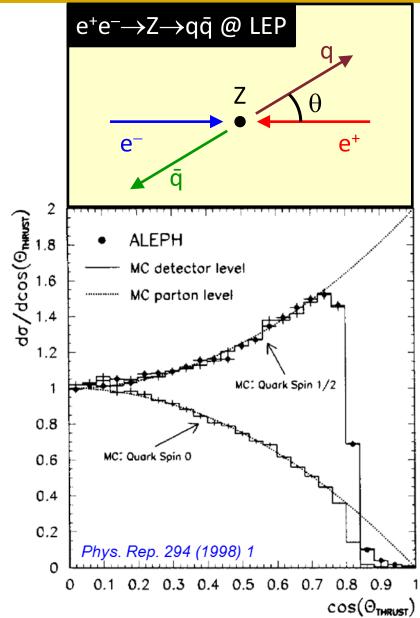


Copernicus

discovery in quantum world - 1

Quantum effects make life much more complicated.

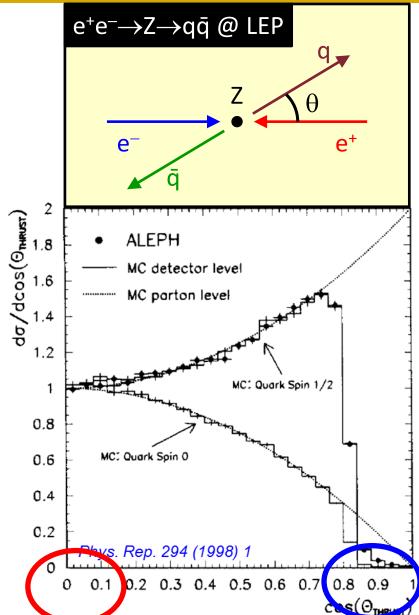
- a theory does not predict an observable ("the position of a planet at a given time"), but rather a probability ("the frequency of a given scattering angle");
- therefore a single observation is meaningless; only samples matter, usually shown as statistical distributions;
- [possibly modified by detector distortions;]
- the data are compared with one (or more) theory(ies), i.e. with predicted distributions;
- the predictions appear in the form of simulated pseudo-data, i.e. "montecarlo's" (MC);
- data and MC samples are finite; therefore statistics plays a fundamental role (errors, test of hypothesis).



discovery in quantum world - 2

A closer look to our example :

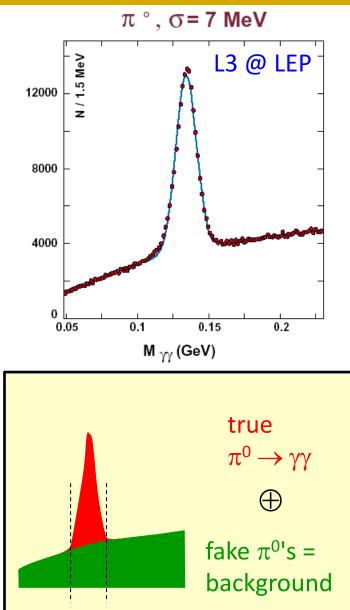
- almost impossible distinguish $q \leftrightarrow \bar{q}$ \Rightarrow ambiguity $\theta \leftrightarrow \pi - \theta \Rightarrow$ use $|\cos \theta|$;
- no detector @ $\theta = 0$ (or $\theta = \pi$) $\Rightarrow d\sigma_{\cos\theta=1} / d\cos\theta = 0$ for exp and theories;
- data sample is finite \Rightarrow statistical errors;
- a good MC necessary to produce theory expectations ⇒ also "theory(ies)" become samples of pseudo-data of finite size;
- many (≈ ∞) possible theories ⇒ no verification, only falsification(s);
- comparison/decision on a statistical basis : in principle, a large statistical fluctuation could reconcile every theory with observation (e.g. the compatibility of these data with "spin-0" is *very* improbable, but not strictly impossible).



background

In quantum world, life is even harder because of background. Example :

- want to identify π^{0} 's among the (many) final state particles;
- in principle easy: since $\pi^0 \rightarrow \gamma \gamma$,
 - 1. take all possible γ 's in an event;
 - 2. measure 4-momentum for each γ;
 - **3.** compute m(γγ) for each γγ pair;
 - 4. that's it (see plot of L3 @ LEP);
- many unpleasant effects:
 - 1. large width (7 MeV), due to exp. resolution;
 - 2. large background (below the peak):
 - misidentified γ's;
 - combinatorial background (uncorrelated γ);
- not possible to select "true" π^{0} 's, even when a $\gamma\gamma$ pair has the "right" mass: only probability of it;
- however, some properties are measurable by statistically subtracting the background: number (i.e. cross-section), angular distribution, ...



high energy experiments

Roughly speaking, two types of results (strongly correlated, but different) :

- 1. "measurements", i.e. a precise numerical assessment about a characteristic of Nature (e.g. the Z mass is 91.1876 ± 0.0021 GeV);
- 2. "searches" i.e. the report of a look for an unknown phenomenon of Nature (as a new particle/effect/interaction), unknown to scientists.

A scientific subject may oscillate $2 \leftrightarrow 1$. E.g. ATLAS and CMS, after discovering the Higgs boson (type 2), are now precisely measuring its mass and couplings (type 1); but result(s) inconsistent with the Standard Model will start a new search, and so on ad infinity.

[in general, a measurement which cannot possibly trigger a search is considered less interesting, because it contains less potential information about the structure of the world]



these lectures



some little jargon

- cross-section (σ) [sezione d'urto, see FNSN1] is proportional to the probability of occurrence of an interaction; it is defined as (the number of reactions per unit time dN/dt) divided by (the beam particles per unit time per unit area $\Phi_a \times$ scattering centres n_s).
- luminosity (L) is (the number of events detected N in a certain time t) divided by (the interaction cross-section σ) [it measures the speed of event production warning : in astronomy a different definition].
- integrated luminosity (L_{int} = ∫ L dt = N/σ) is (the total number of events of a given type) divided by (its cross section) [it measures the total amount of data in an experimental data sample].
- expected number of events (N_i) in a given experiment is (the integrated luminosity L_{int}) × (the relevant crosssection σ_i) × (the experimental efficiency ε_i) [the actual number of found events (n_i) fluctuates around N_i with Poisson pdf – warning: n_i is integer, N_i is not].

$$\sigma = \frac{dN}{dt} \frac{1}{\Phi_a n_s}$$

 $L = \frac{dN}{dt} \frac{1}{\sigma}$

$$N_i = L_{int} \sigma_i \epsilon_i$$

 $L_{int} = \int Ldt = \frac{N}{N}$

searches and limits

• Sometimes, the result of the study is NOT the measurement of an observable x :

"x = $x_{exp} \pm \Delta x$ ",

 but, instead, a qualitative "<u>search</u>" : "the phenomenon *Y* (does not) exists",

or, alternatively :

"the phenomenon $\mathcal Y$ does NOT exist in the kinematical range Φ ".

- [the second statement applies if nothing is found, and an "<u>exclusion</u>" (a "<u>limit</u>", when Φ is not complete) is established]
- In modern experiments, the searches occupy more than 50% of the published papers, and almost all are negative [but the Higgs search at LHC, of course].
- Obviously, a negative result is NOT a failure : if any, it is a failure of the underlying theory.

- [but a positive result is much more pleasant (and rewarding)]
- A <u>rigorous method</u>, well understood and "easy" to apply, is imperative.
- This method is a major success of the LEP era : it uses math, statistics, physics, common sense and communication skill.
- It MUST be in the panoply of each particle physicist, both theoreticians and experimentalists.
 - use the Higgs searches at LEP and LHC as a test case [these lectures want to remain inside the SM];
 - today, after the Higgs discovery, the focus has shifted toward "bSM" searches.

searches and limits : verify/falsify

- [there is a model (SM, SUSY, etc.) that predicts a phenomenon (a particle, a dynamic effect), possibly dependent on some unknown parameter (e.g. the Higgs boson mass)];
- [simpler case : the phenomenon is completely determined by the theory, e.g. W[±] and Z production at the Spp̄S Collider];
- a new accelerator is potentially able to observe the phenomenon in a range of the parameters space still unexplored;
- therefore, two possibilities :
 - A. observation : the theory is "verified" (à la Popper); the free parameter(s), if any, are measured;
 - B. non-observation : some subspace in the parameter space (simpler case :

an interval in one dimension) is excluded, i.e. a "limit" is established; when the full parameter space is excluded, the theory is "falsified" (à la Popper);

 ♦ different, less common, approach ("model independent") : look for unknown effects, without theoretical guidance; e.g. bound states ℓ+ℓ- at high mass (cfr. J/ψ).

method of limits : Poisson statistics

- In general, the searches looks for processes with VERY limited statistics (want to discover asap);
- therefore the limit ("n large", more precisely n >> √n) cannot be used (neither its consequences, like the Gauss pdf);
- searches are clearly in the "Poisson regime" : large sample and small probability, such that the expected number of events ("successes") be finite;
- use the Poisson distribution :

 $\int \mathcal{O}(N \mid m) = \frac{e^{-m}m^{N}}{N!}; \langle N \rangle = m; \sigma_{N} = \sqrt{m};$

• therefore, in a search, two cases :

a. the signal does exist : $\int O(N | b+s) = \frac{e^{-(b+s)}(b+s)^{N}}{N!}; \quad \begin{cases} \langle N \rangle = b+s; \\ \sigma_{N} = \sqrt{b+s}; \end{cases}$ [two possibilities : s known/unknown] b. the signal does NOT exist : $\int O(N | b) = \frac{e^{-b}b^{N}}{N!}; \langle N \rangle = b; \sigma_{N} = \sqrt{b};$

- the strategy is : use N(observed) to distinguish between case (a) and (b);
- since ℘ is positive for N in [0,∞] in both cases, the procedure is to define a CL <u>a</u> priori, and accept the hypothesis (a or b) only if it falls in the predefined interval;
- modern (LHC) evolution : define a parameter, usually called " μ " :

$$\int \mathcal{O}(N \mid b + \mu s) = \frac{e^{-(b + \mu s)}(b + \mu s)^{N}}{N!}; \quad \begin{cases} \langle N \rangle = b + \mu s; \\ \sigma_{N} = \sqrt{b + \mu s}; \end{cases}$$

clearly, $\mu = 0$ is bckgd only, while $\mu = 1$ means discovery; it is customary to present results as limits on " μ " [*e.g.* <u>exclude</u> $\mu = 0$ means "<u>discovery</u>"].

searches and limits : "n sigma"

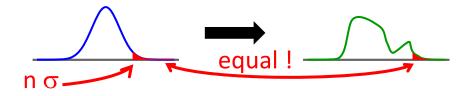
- sometimes the *a priori* criterion is defined in tems of "sigmas" [i.e. "a 3 σ exclusion"];
- look at the integral of the <u>normalized</u> standard Gauss distribution (mean=0, sigma=1):

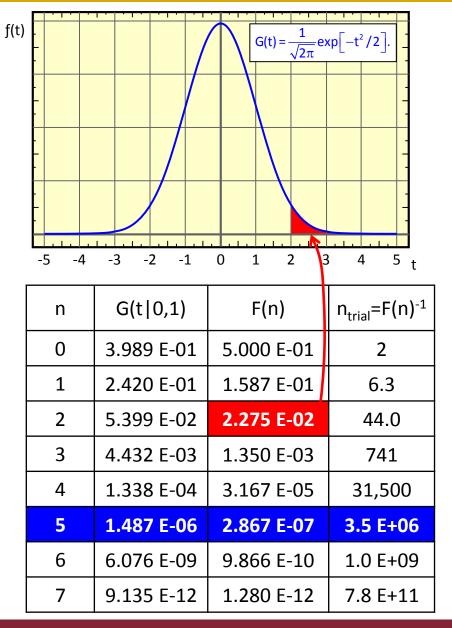
 $F_{gauss}(n) = \int_{n}^{+\infty} G(t \mid \mu = 0, \sigma = 1)dt;$

"n sigma" criterion :
 "exclude those cases in the distribution tail,

such that their cumulative probability is smaller than F_{gauss}(n)";

- [in the example, n > 2 is excluded, i.e. 0.275% of cases becomes "impossible";]
- the criterion does NOT depend on the shape of the distribution: if it is not Gaussian (e.g. Poisson or other), compute F_{distr} and use it, defining "n" with F_{gauss}(n):





method of limits

[in the "good ole times", life was simpler : if the background was negligible, the first observation led to the discovery, as for e^+ , \bar{p} , Ω^- , W^{\pm} and Z]

- in most cases, the background (reducible or irreducible) is calculable;
- a <u>discovery</u> is defined as an observation that is incompatible with a +ve statistical fluctuation respect to the <u>expected</u> <u>background alone</u>;
- a <u>limit</u> is established if the observation is incompatible with a -ve fluctuation respect to the <u>expected (signal +</u> <u>background);</u>
- both statements are based on a "reductio ad absurdum"; since all values of N in [0,∞] are possible, it is compulsory to predefine a CL to "cut" the pdf;
- the CL for discovery and exclusion can be different : usually for the discovery stricter criteria are required;

- <u>a priori</u> the expected signal s can be compared with the fluctuation of the background (in approximation of large number of events, s ↔ vb) : n_σ = s / vb is a figure of merit of the experiment;
- <u>a posteriori</u> the observed number (N) is compared with the <u>expected background</u> (b) or with the <u>sum (s + b)</u>.

Example. We expect 100 background events and 44 signal; we use the "large number" approximation ($\Delta n = \sqrt{n}$): $b = 100, \Delta b = \sqrt{b} = 10;$

s + b = 144, $\Delta(s + b) = 12$.

The <u>pre-</u>chosen confidence level is "3 σ ".

The discovery corresponds to an observation of $N > (100+3 \times 10) = 130$ events. A limit is established if $N < (144 - 3 \times 12) = 108$ events. There is no decision if 108 < N < 130. The values N < 70 and N > 180 are "impossible".

method of limits : exercise

Exercise (based on previous example) :

compute the factor, wrt to previous luminosity, which allows to exclude the "no-decision" region.

Answer : 9/4 = 2.25

Solution :

call f the factor; then

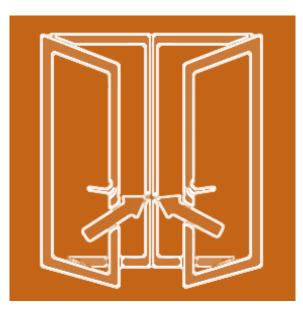
 $100 \text{ f} + 3 \sqrt{100 \text{ f}} = 144 \text{ f} - 3 \sqrt{144 \text{ f}}$ f=9/4; b=225; (s+b)=324, s=99.

[criticism : N = 270 is still "no-decision"; find a better solution ...

... maybe :

$$100 \text{ f} + 3 \sqrt{100 \text{ f}} + 1 = 144 \text{ f} - 3 \sqrt{144 \text{ f}}$$

... but N must be integer !]



method of limits : discovery, exclusion

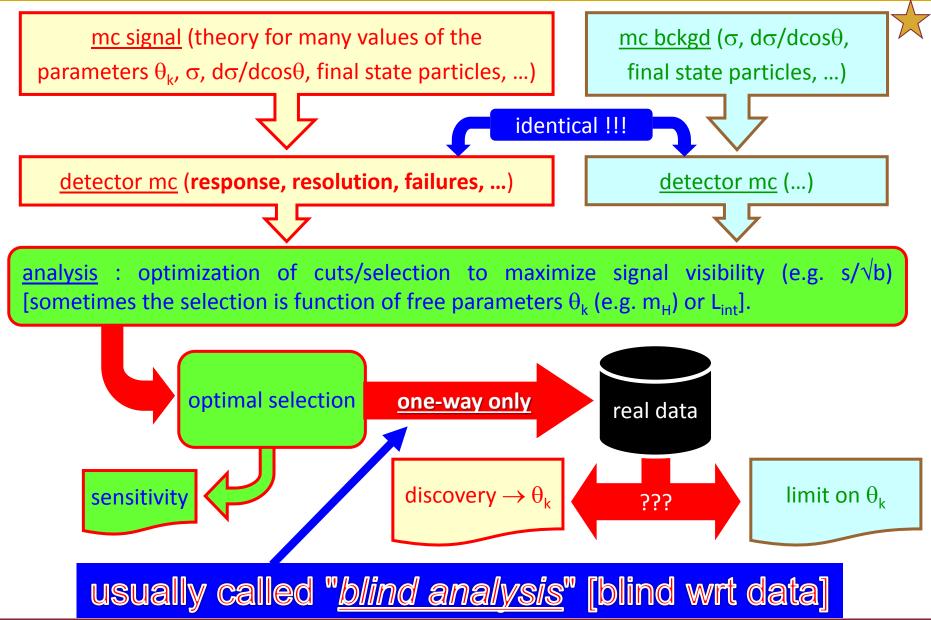
- usually experiments predefine the CL according to a homogeneous rule :
 - > <u>DISCOVERY</u> : "5 σ ", \wp (b only) \leq 2.86×10⁻⁷;
 - > EXCLUSION : \wp (s+b) \leq 5×10⁻² ; [called also "95% CL"];
- <u>a priori</u>, the integrated luminosity L_{int} for discovery / exclusion is also defined :
 - > \underline{L}_{disc} : L_{int} min, such that 50% of the experiments^(*) (i.e. an experiment in 50% of the times) had ℘ (b only) ≤ "5 σ ";
 - ▶ <u>L</u>_{excl} : L_{int} min, such that 50% of the experiments^(*) (i.e. an experiment in 50% of the times) had ℘(s+b) ≤ "2σ";
- NB: this rule corresponds to the median ["an experiment, in 50% of the times..."],

and it is different from the <u>average</u> ["an experiment, with exactly the expected number of events ..."].

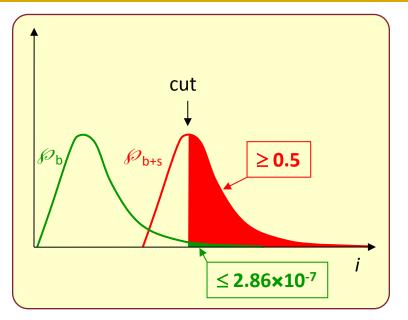
(*) sometimes an "experiment" at LEP [LHC] is the combination of all 4 [2] collaborations.

A parameter is said to be "excluded at xx% confidence level" [say 95%] if the parameter itself would yield more evidence than that observed in the data at least 95% of the time in a [pseudo-]set of repeated experiments, all equivalent to the one under consideration. [CMS web dixit]

method of limits : "blind analysis"



method of limits : Luminosity of disc., excl.

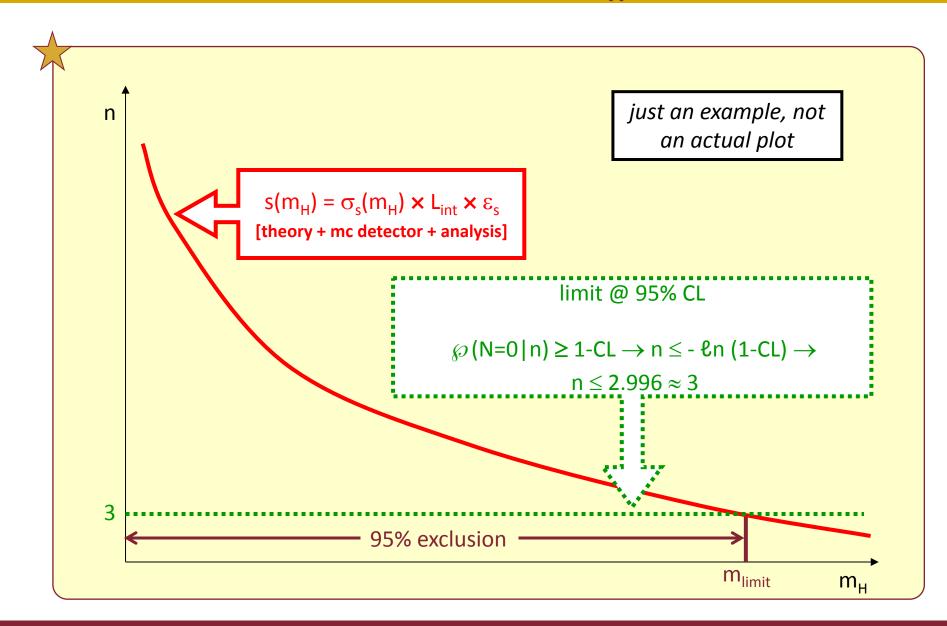


The values of L_{disc} and L_{excl} come from the previous equations; compute L_{disc} (L_{excl} is similar):

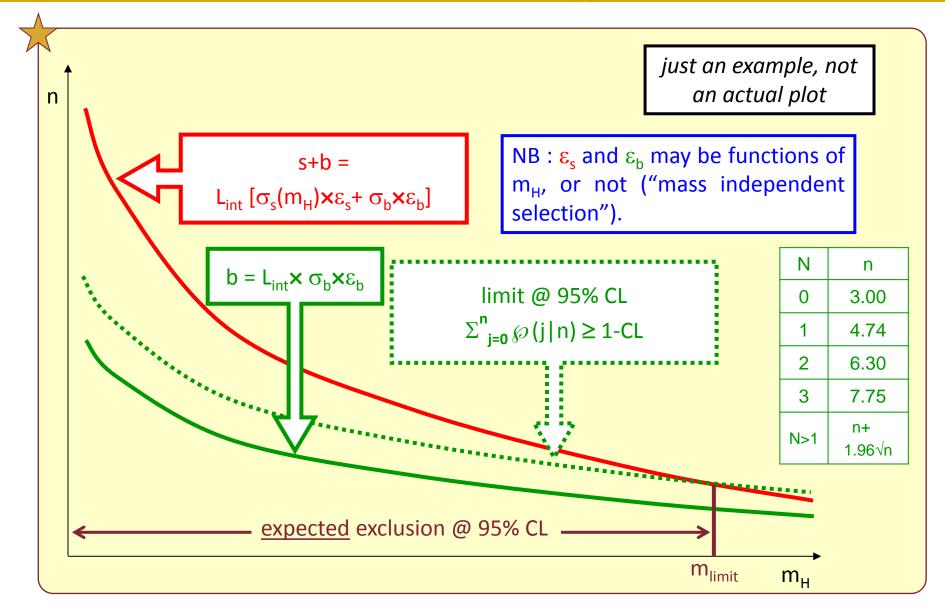
$$\begin{aligned} & "\bullet" = e^{-b} \times \sum_{i=N}^{\infty} \frac{(b)^{i}}{i!} \leq \wp(5\sigma) = 2.86 \times 10^{-7} \\ & "\bullet" = e^{-(b+s)} \times \sum_{i=N}^{\infty} \frac{(b+s)^{i}}{i!} \geq 0.5; \\ & b = L_{disc} \varepsilon_{B} \sigma_{B}; \quad s = L_{disc} \varepsilon_{S} \sigma_{S}. \end{aligned}$$

- > assume variable luminosity ($L_{int} = L_{disc} / L_{excl}$) and constant ε_s , ε_b , σ_s , σ_b ;
- assume to start with small L_{int} : the two distributions overlap a lot, no N satisfies the system (i.e. the <u>tail</u> above the <u>median</u> is large);
- > when L_{int} increases, the two distributions are more and more distinct (overlap $\propto 1/VL_{int}$);
- finally, for a given value of L_{int}, it exists a number of events N, such that the cuts at <u>2.86×10⁻⁷</u> (<u>0.5</u>) in the <u>first</u> (<u>second</u>) cumulative coincide; this value of L_{int} correspond to L_{disc};
- this is the luminosity when, if the signal exists, 50% of the experiments have (at least) 5σ incompatibility with the hypothesis of bckgd only.

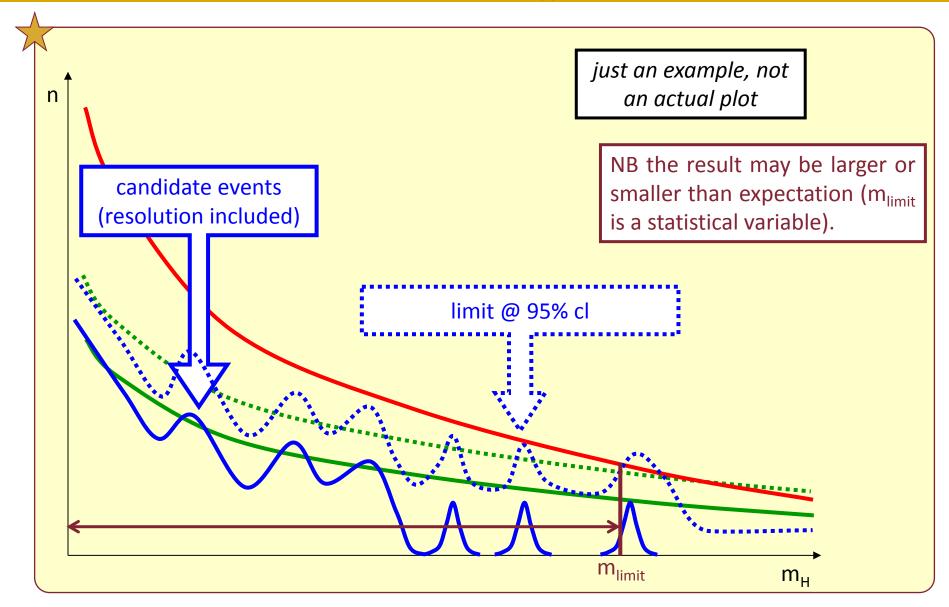
method of limits : ex. m_H (b=0, n=0)



method of limits : ex. m_H (a priori, b>0)



method of limits : ex. m_H (a posteriori, b>0)



interpretation of results : discovery plot

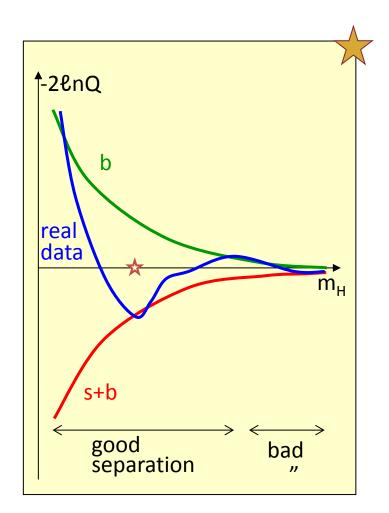
- The likelihood is expected to be larger when the correct pdf is used;
- this is easily confirmed by the "large n" limit [-2en(Λ) $\rightarrow \chi^2$] :

$$-2\ell nQ = -2\ell n(\Lambda_s/\Lambda_b) \approx \chi_s^2 - \chi_b^2 =$$

$$= \begin{cases} ("small" - \\ "large" \end{pmatrix} \rightarrow <<0 & \text{if (s+b) ok} \\ ("large" - \\ "small" \end{pmatrix} \rightarrow >>0 & \text{if (b) ok} \end{cases}$$

the plot is a little cartoon of an ideal situation (e.g. Higgs search at LEP2), that <u>never happened</u> :

- the cross-section decreases when m_H increases (TRUE);
- discovery !!! (FALSE).

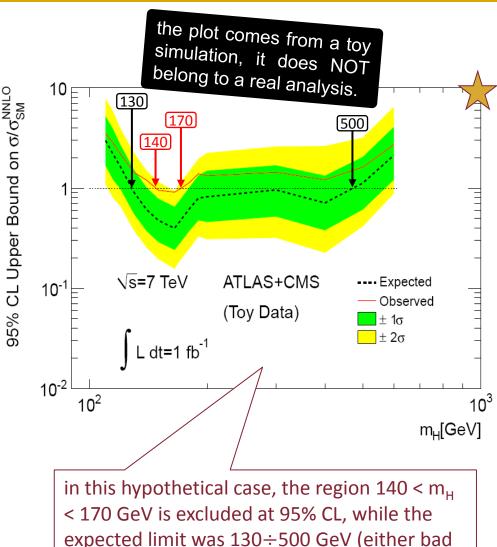


interpretation of results : parameter μ

- put : $\sigma = \sigma^{b} + \mu \sigma_{SM}^{s}$;
- plot : horizontal : m_H.
- vertical : $\mu (= \sigma_{obs}^{s} / \sigma_{SM}^{s});$
- b the lines show, with a given L_{int} and analysis, the expected limit (--), and the actual observed limit (--), i.e. the μ value excluded at 95% CL;
- ➤ the band ◆ (•) shows the fluctuations at ±1σ (±2σ) of the "bckgd only" hypothesis.

NB:

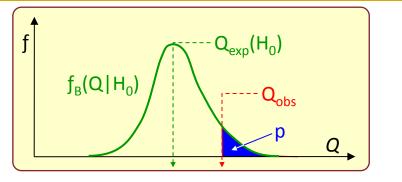
- the case µ≠1 has no well-defined physical meaning (= a theory identical to the SM, but with a scaled cross section);
- if the lines are at $\mu > 1$, the "distance" respect to $\mu=1$ reflects the L_{int} necessary to get the limit in the SM.



luck or hint of discovery).

interpretation of results : p-value

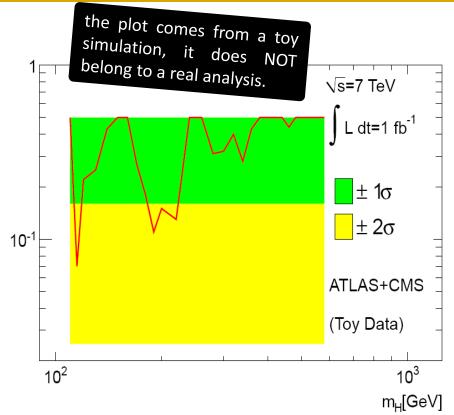
-ocal p-value



 $p \equiv \int_{x_{obs}}^{\infty} f(x \mid H_0) dx$

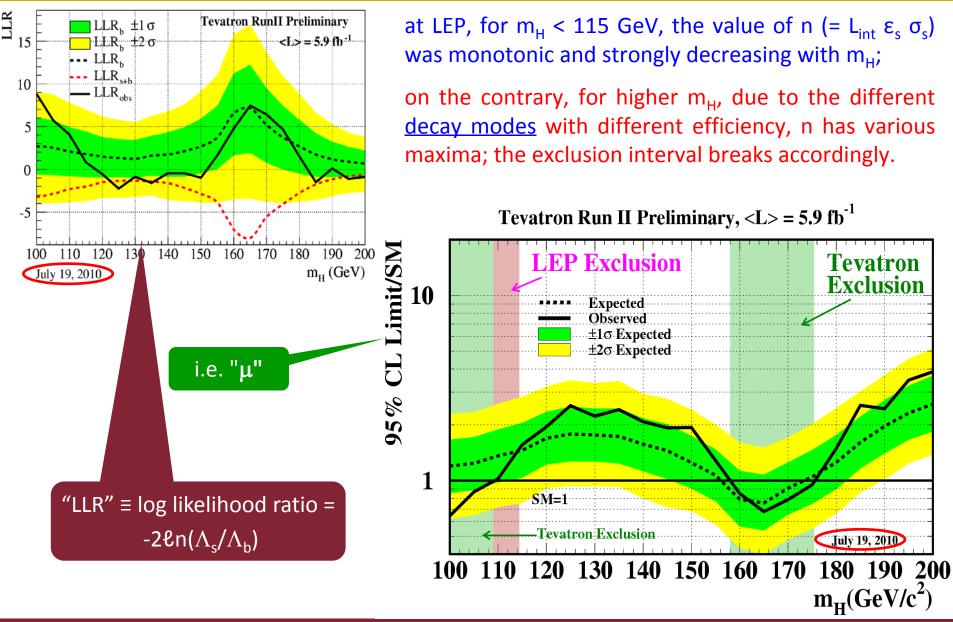
- the "p-value" is the probability to get the same result or another less probable, in the hypothesis of bckgd only.
- x = "statistics" (e.g. likelihood ratio);
- H₀ = "null hypothesis (i.e. bckgd only);
- i.e.

```
p small \rightarrow H<sub>0</sub> NOT probable
\rightarrow discovery !!!
```



- vertical : *p*-value;
- horizontal : m_H.
- the band (•) shows the fluctuations at 1σ (2σ).
- NB the discovery corresponds to the red line below 5σ (or 2.86×10⁻⁷), not shown in this fake plot.

Higgs – pre-LHC : Tevatron legacy



2012, July 4th



Higgs decays (memo)

- 1. pp \rightarrow HX, **H** \rightarrow **ZZ**^{*}, Z^(*) \rightarrow $\ell^+\ell^-$ [$\ell^\pm = \mu^\pm, e^\pm$]:
 - select 4 leptons in the final state with correct charge and flavor;
 - require a unlike-sign, like-flavor with m ≈ m(Z);
 - require the other pair have unlike-sign, likeflavor;
 - [...] few other cuts to enhance signal;
 - plot m(ZZ*);
 - compute exp + combinatorial bckgd.

2. pp \rightarrow HX, **H** $\rightarrow \gamma \gamma$:

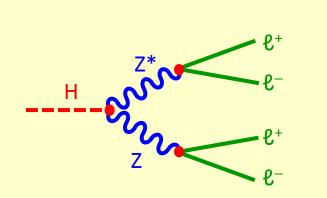
- require two hard isolated photons;
- [...] few other cuts to enhance signal;
- plot m(γγ);
- compute exp + combinatorial bckgd.
- 3. pp \rightarrow HX, other Higgs decays :
 - less probable or difficult to identify;
 - not studied in 2012 (some seen now).

Н

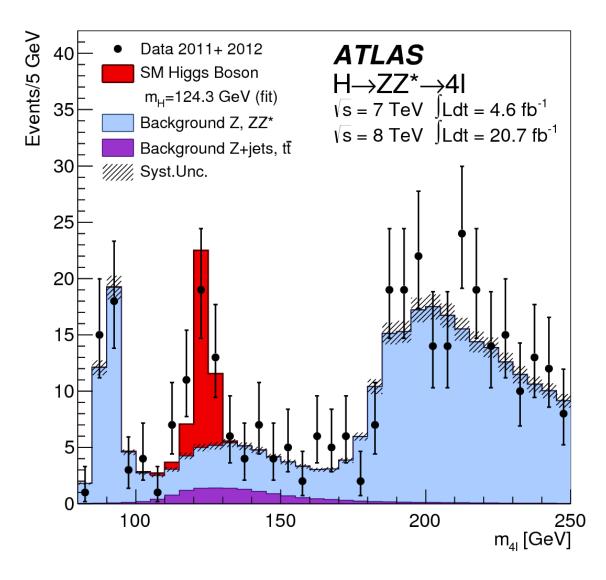
W[±], t, (b)

interference W \leftrightarrow t large and

negative (nice test of the SM).

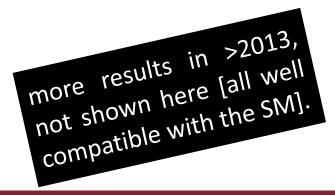


Higgs discovery : $H \rightarrow ZZ^*$ - ATLAS

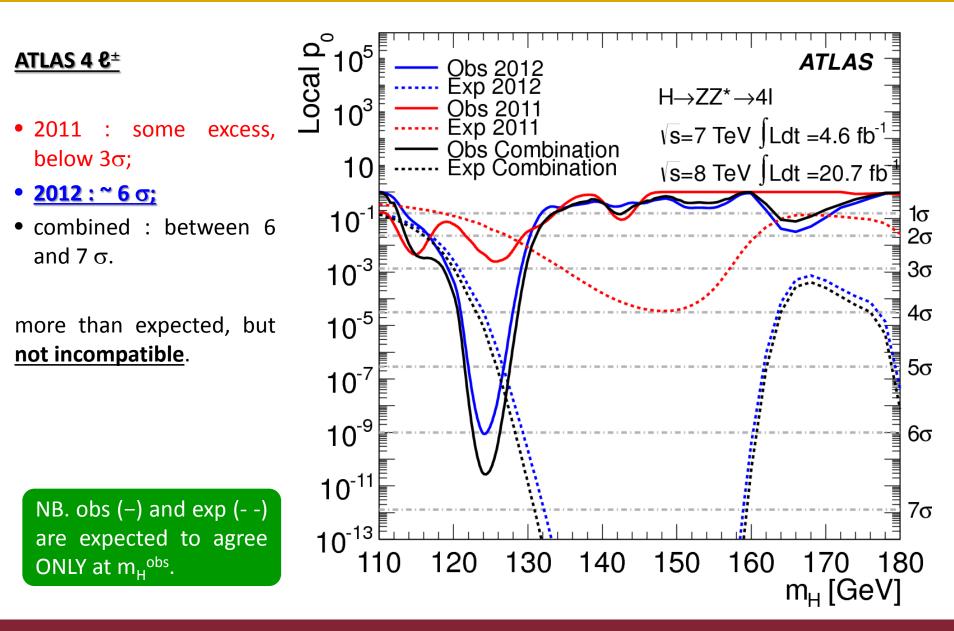


 $H \rightarrow ZZ^* \rightarrow \ell^+\ell^-\ell^+\ell^-$ Test mass ~ 125 GeV

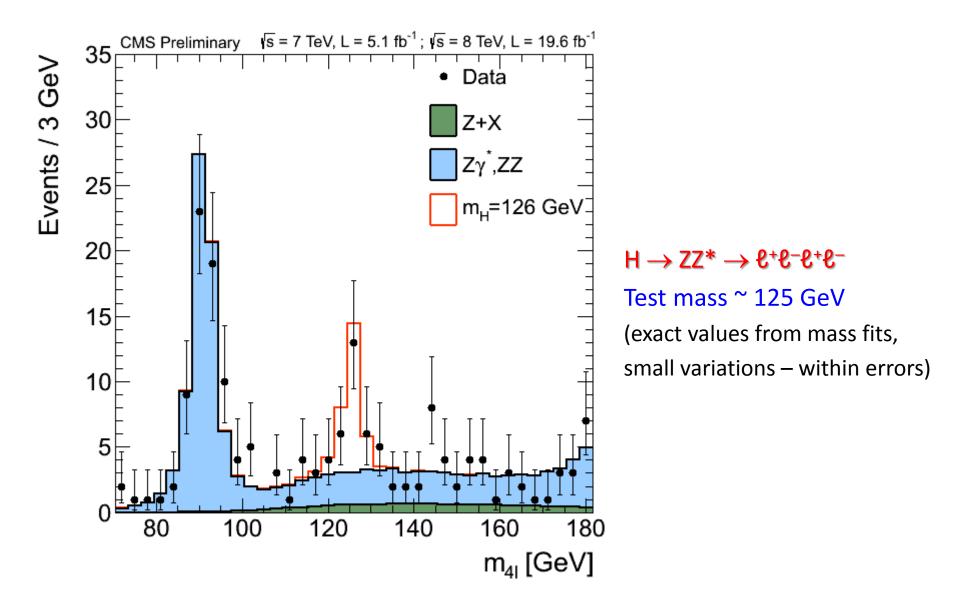
(exact values from mass fits, small variations – within errors)



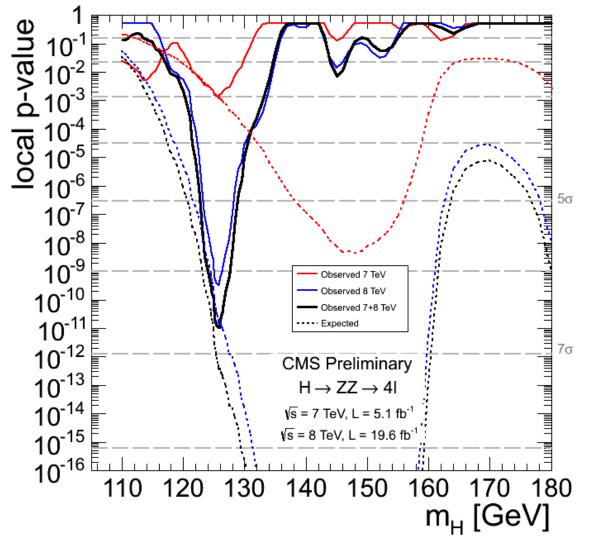
Higgs discovery : $H \rightarrow ZZ^*$ - ATLAS p-value



Higgs discovery : $H \rightarrow ZZ^*$ - CMS



Higgs discovery : $H \rightarrow ZZ^*$ - CMS p-value



<u>CMS 4 8±</u>

• 2011 : some excess, ~3 σ;

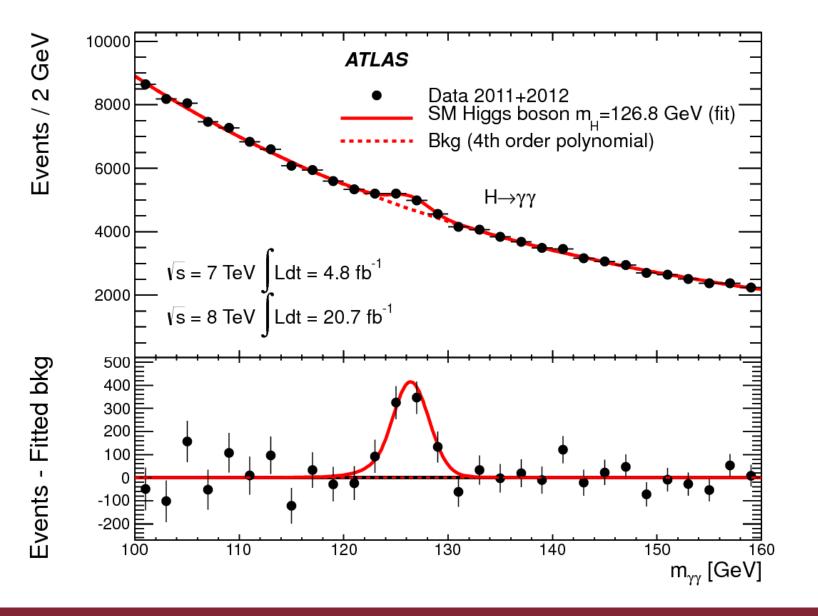
• <u>2012 : > 6 σ;</u>

• combined : between 6 and 7 σ .

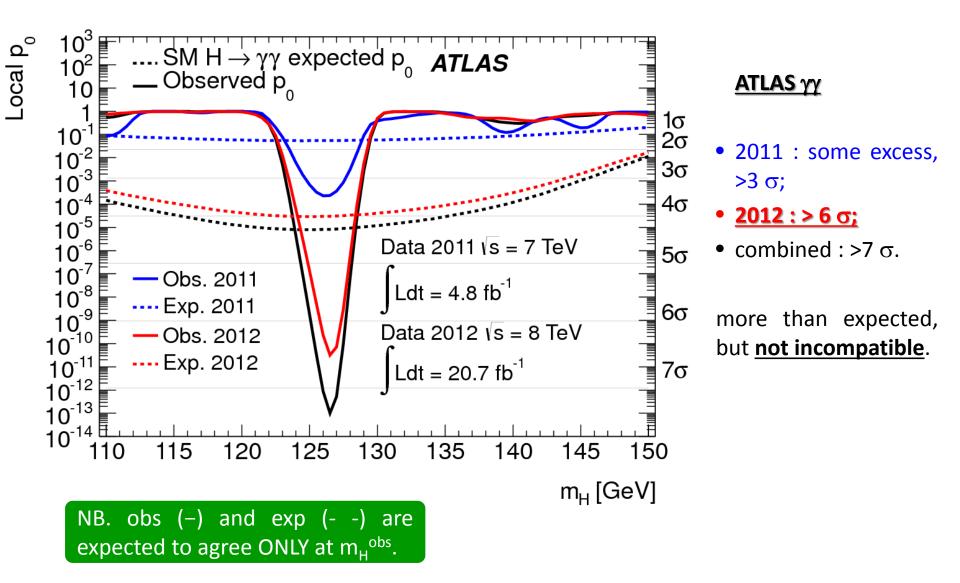
well compatible with expected.

NB. obs (-) and exp (- -) are expected to agree ONLY at m_{H}^{obs} .

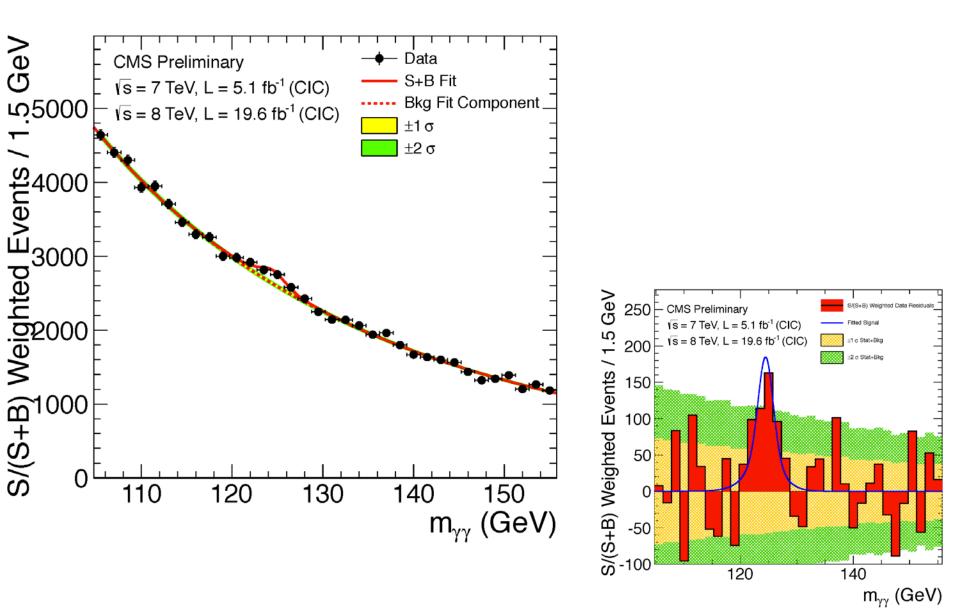
Higgs discovery : $H \rightarrow \gamma \gamma$ - ATLAS



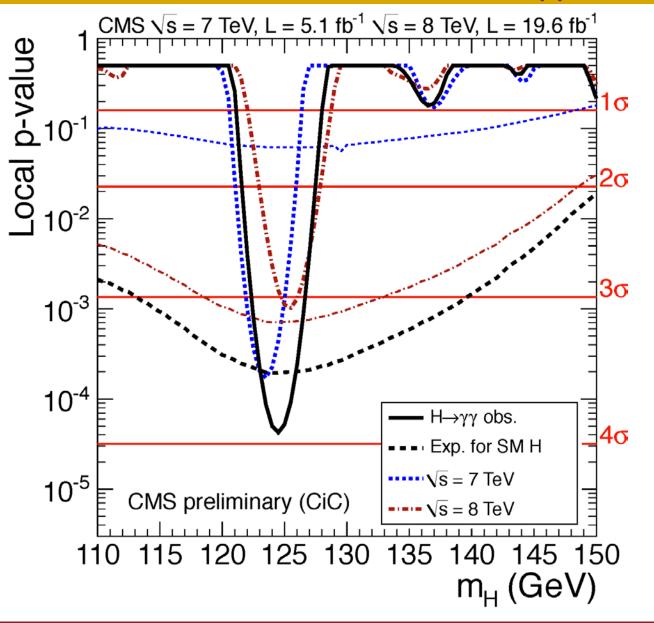
Higgs discovery : $H \rightarrow \gamma \gamma$ - ATLAS p-value



Higgs discovery : $H \rightarrow \gamma \gamma$ - CMS



Higgs discovery : $H \rightarrow \gamma \gamma$ - CMS p-value



<u>CMS γγ</u>

- 2011 : some excess,
 >3 σ;
- <u>2012 : > 3 σ;</u>
- combined : ~4 σ .

well compatible with expected.

NB. obs (–) and exp (- -) are expected to agree ONLY at m_{H}^{obs} .

Conclusions

- a textbook case of our statistical approach;
- example of modern path to a "discovery";
- ... which avoids the inconsistencies, especially those due to quantum effects;
- complicated, but sound and reproducible;
- ... and allows for comparison among experiments and with theory/ies;
- personal opinions :
 - a nightmare, but an unavoidable approach to data analysis.
 - useful also for probabilistic phenomena outside HEP (earthquakes ...)
 - ➢ find how to explain to general public (?).



References

- 1. Science, 338 (2012) 1560, 1569, 1576 [simple, divulgative];
- A.Nisati, G.Tonelli Riv. Nuovo Cimento, 38 (2015), 507 [clear, detailed];
- **3.** Higgs properties : [a textbook, e.g. L.Maiani, Interazioni elettrodeboli, 14];
- 4. yellow reports CERN-2011-002, CERN-2012-002, CERN-2013-004;
- 5. https://twiki.cern.ch/twiki/bin/view/LHCPhysics;
- https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ CombinedSummaryPlots;
- https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ StandardModelPublicCollisionPlots;
- https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ HiggsPublicResults#Animations + http://cms.web.cern.ch/org/cms-higgs-results

