

Special Report

Higgs Days at Santander

Philip E. Gibbs*

Abstract

Santander is a Spanish port on the Bay of Biscay coast that next week (19-23 Sept, 2011) will host its [fourth annual workshop on the Higgs Boson](#). This meeting will be very different in character from the huge summer conferences where exciting new results on searches for the Higgs boson were recently presented to thousands of physicists. One question has become very topical and has already surfaced at some of the larger Higgs workshops: Is it right to do quick approximate combinations of Higgs search data or do we need to wait for the lengthy process of producing the official combinations? I will discuss this issue here.

Key Words: Santander meeting, Fermi Lab, CERN, Higgs workshop.

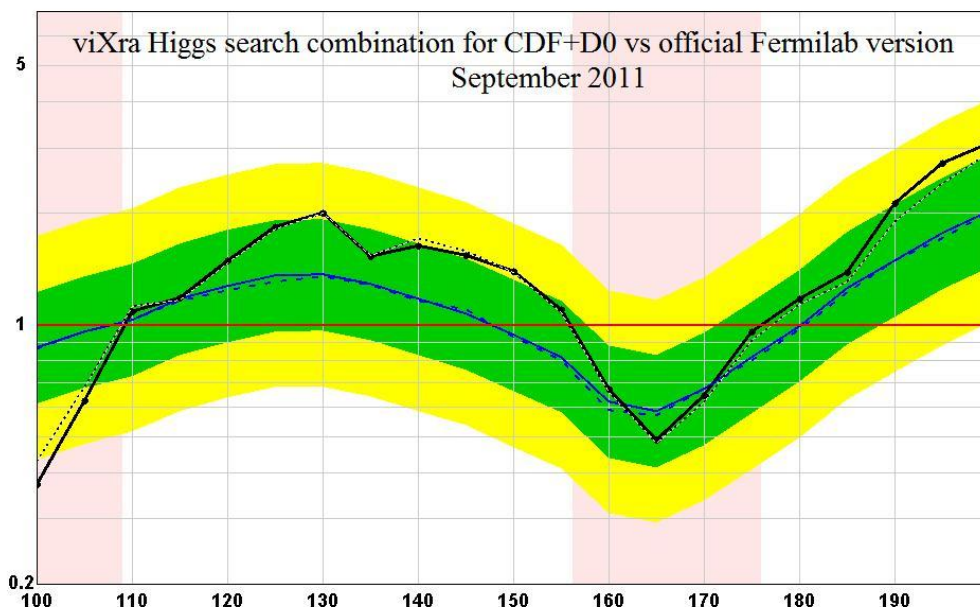
The Santander meeting involves just 30 participants with a mix of theorists and experimenters involved in the analysis of data from Fermilab and CERN. Half their time will be spent presenting slides and the other half will be discussions covering searches for standard model Higgs and other models including the charged Higgs sector of SUSY. They will talk about the procedures for combining Higgs searches across experiments and implications of any findings. The aim is to promote a dialog between theorists and experimenters about what data needs to be shared and how.

There is no indication that the discussions will be webcast or recorded for public viewing and it is not sure that all the slides will appear online so as outsiders the rest of us may have very little indication of what they decide. It is unlikely that new data will be made public but there is some chance that we may finally get to see a combination of ATLAS and CMS search data. Originally we were promised a combination of the searches shown at EPS in July using the first 1/fb of data from the LHC. Instead we got a new helping of plots from the individual experiments using 1.6/fb in the most important channels and even 2.3/fb for the ZZ channel in ATLAS. These were shown at the Lepton-photon conference in August. Theorists would now very much like to see the combinations of these data sets and it is not clear why they have been held back.

One question has become very topical and has already surfaced at some of the larger Higgs workshops: Is it right to do quick approximate combinations of Higgs search data or do we need to wait for the lengthy process of producing the official combinations? This summer I have become quite notorious for doing these quick combinations and showing them on viXra log. These have variously been described by experts as “nonsense” (Bill Murray) “garbage” (John Ellis) and “wrong” (Eilam Gross), but just how bad are they? Here is a plot of my handcrafted combination of the D0 and CDF exclusion plots compared with the official combo. The thick black line is my version of the observed exclusion limit that can be compared with the dotted line of the official result, while the solid blue line is may calculated

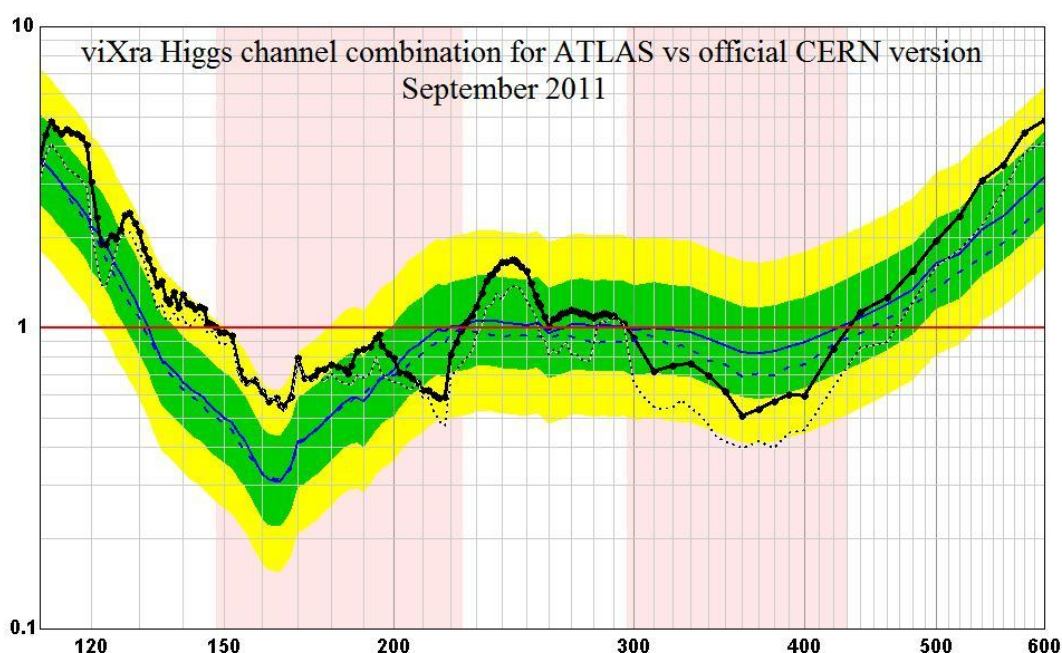
* Correspondence: Philip E. Gibbs, Ph.D., Independent Researcher, UK. E-Mail: phil@royalgenes.com Note: This report is adopted from <http://blog.vixra.org/2011/09/18/higgs-days-at-santander/>

expected limit to be compared with the official dashed line. You need to click on the image for a better view.

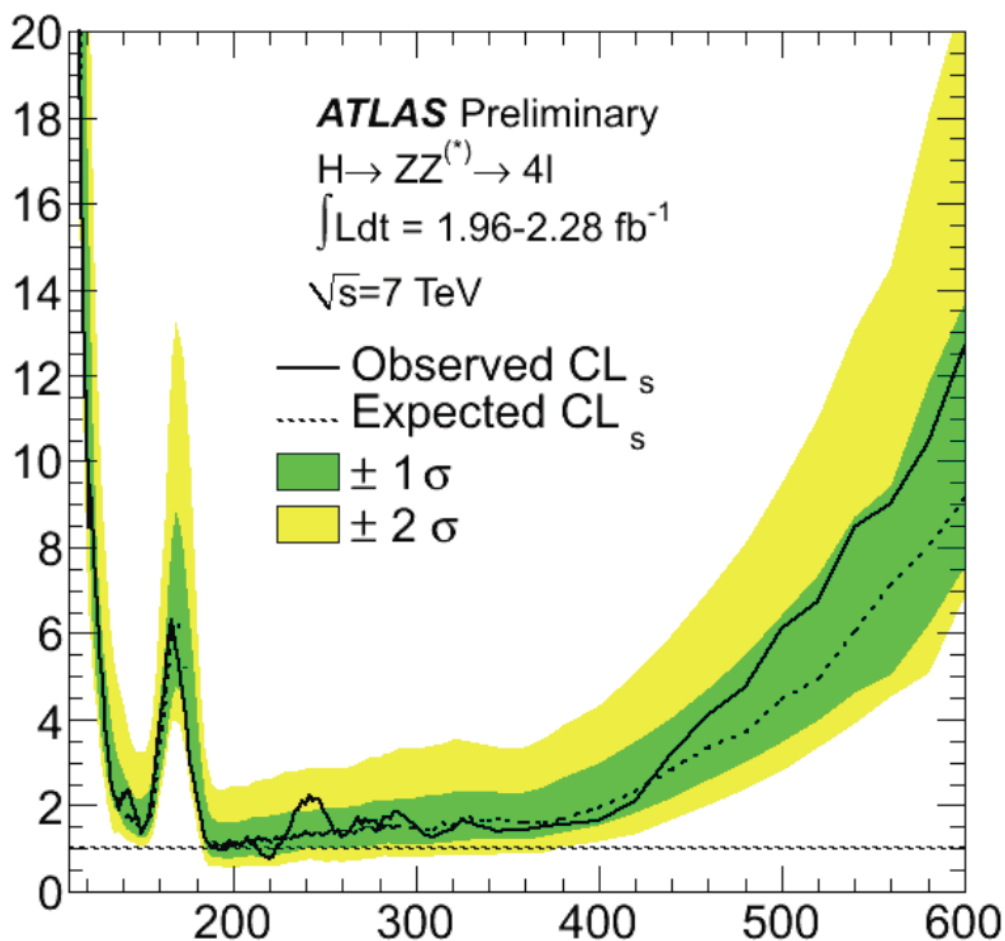


My result is not perfect but I hope you will agree that it provides similar information and you would not be misled into drawing any wrong conclusions from it that were not in the official plot. Any discrepancy is certainly much smaller than the statistical variations indicated by the green and yellow bands for one and two sigma variations.

A more ambitious project is to combine exclusion plots for individual channels to reproduce the official results for each experiments. Here is my best attempt for the latest ATLAS results where I have combined all eight channels for primary decay products of the Higgs boson.



The result here is not as good and could only serve as a rough estimation of the proper combination. Why is that? There are several sources of error involved. Firstly the data for the individual channels had to be digitised from the plots. This was not the case for the previous Tevatron combination above where they published the plots in tabular form. ATLAS and CMS have only published such numerical data for a few channels and in some cases the quality of the plots shown is extremely poor. For example this is the best plot that ATLAS has shown for the important $H \rightarrow ZZ \rightarrow 4l$ channel



As you can see it is very hard to follow the lines on this plot, especially the dashed expected limit line. I don't want to be over-critical but seriously guys, can't you do better than this?

Another source of error comes from neglect of correlations between the individual plots where background estimates may have the same or related systematic errors. The Higgs combination group at CERN play on this as one of the reasons why these quick combinations can't be right, but I doubt that these effects are significant at all. If they were I would not be getting such good results for the Tevatron combination.

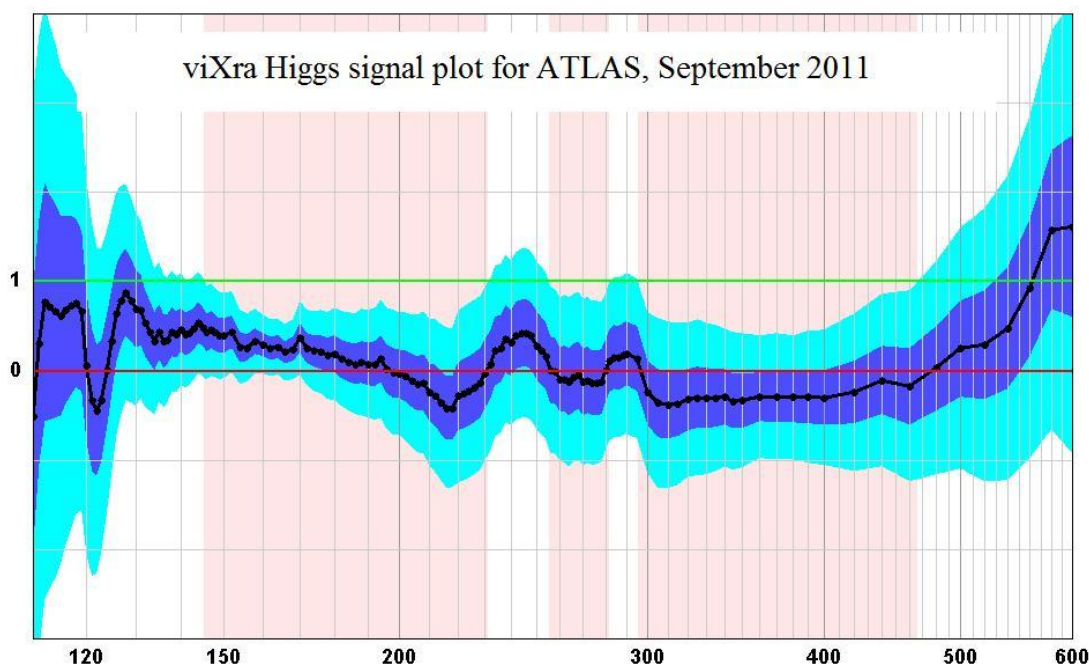
In fact the main source of error is in approximations used in my combination algorithm. It assumes that each statistical distribution of the underlying signals can be modeled by a flat normal distribution with a mean μ_i and standard deviation σ_i . Combining normal distributions is standard stuff in particle physics the combined mean μ and standard deviation σ are given by these formula

$$\frac{1}{\sigma^2} = \sum_i \frac{1}{\sigma_i^2}$$

$$\frac{\mu}{\sigma^2} = \sum_i \frac{\mu_i}{\sigma_i^2}$$

For example, if one experiment tells me that the mass of the proton is 938.41 ± 0.21 GeV and another tells me it is 938.22 ± 0.09 GeV and I know that the errors are independent, then I can combine with the above formula to get a value of 938.25 ± 0.08 GeV. The Particle Data Group does this kind of thing all the time.

A plot of the signal for the Higgs boson given by the ATLAS results would look like this,

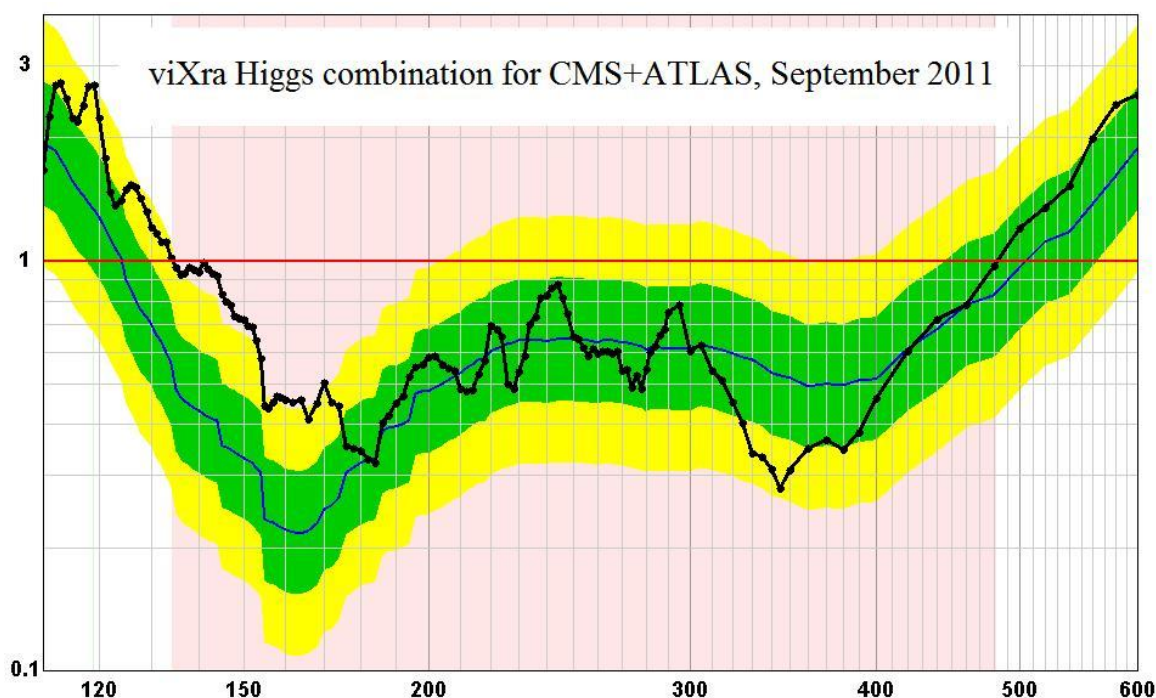


The black line (value of μ) is the observed combined signal for the Higgs boson normalised to a scale where no Higgs boson is zero and a standard model Higgs boson gives one. The blue and cyan bands show the one and two sigma statistical uncertainty ($\mu \pm \sigma$ and $\mu \pm 2\sigma$). Don't think about where the Higgs boson is for now. Just look at the upper two sigma level curve and compare it with the ATLAS Higgs exclusion plot above (i.e the dotted line, click to enlarge for a better view). These are of course the same lines because the 95% level exclusion is given when the 2 sigma error is below the signal for SM Higgs. The expected line on the exclusion plot is just where the observed line would be if the signal were everywhere zero, i.e it is a plot of 2σ . In summary, the observed limit for CL_s in the exclusion plot is just $\mu + 2\sigma$ and the expected limit is just 2σ . We can derive one plot from the other using this simple transformation.

From this it should be clear how to combine the exclusion plots. We first transform them all to signal plots, then they can be combined as if they are normal distributions. Finally the combined signal plot can be transformed back to give the combined exclusion plot. This is what I did for the viXra combinations above.

Ignoring the digitisation errors and the unknown correlations, the largest source of error is the assumption that the distribution is normal. In reality a log normal distribution or a Poisson distribution would be better, but these require more information. Fortunately the central limit theorem tells us that anything will approximate a normal distribution when high enough statistics are available so the combination method gets better as more events accumulate. That is why the viXra combination of the exclusion plots for each experiment is more successful than for the combination of individual channels. The number of events seen in some of these channels is very low and the flat normal distribution is not a great approximation to use. As more data is collected the result will get better. Of course we cannot expect a reliable signal to emerge from individual channels until the statistics are good, so it could be argued that the approximation is covered by the statistical fluctuations anyway.

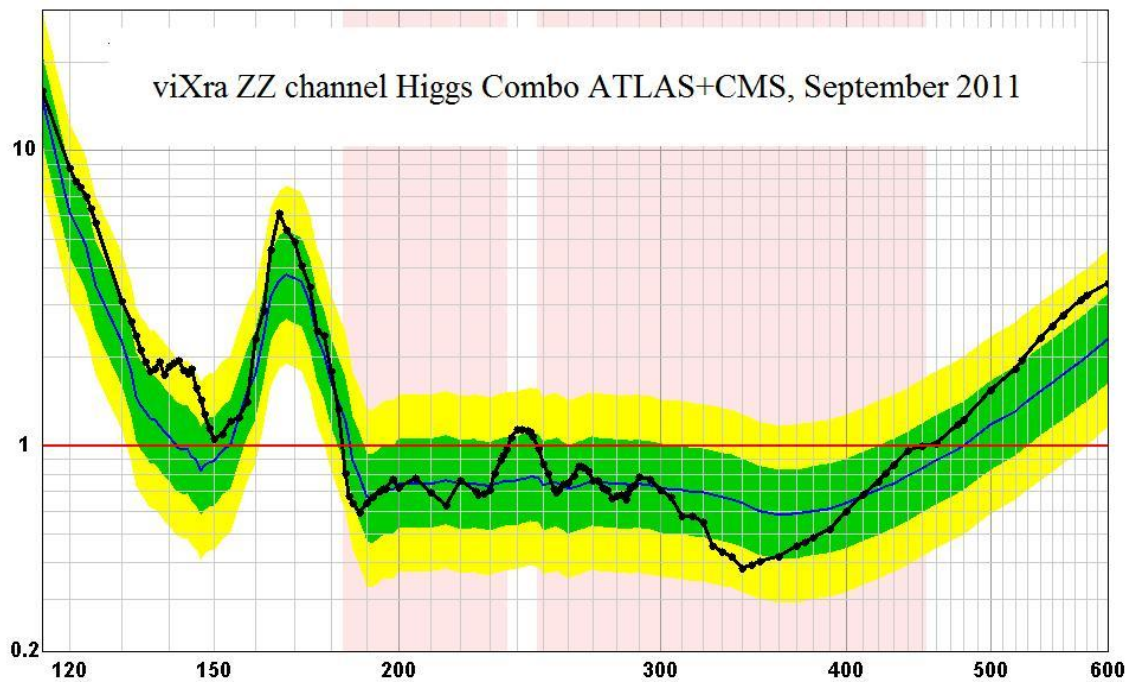
I don't know if a full LHC combination will emerge next week at the Santander workshop but in case it does, here is my best prediction from the most recent data for comparison with anything they might show.



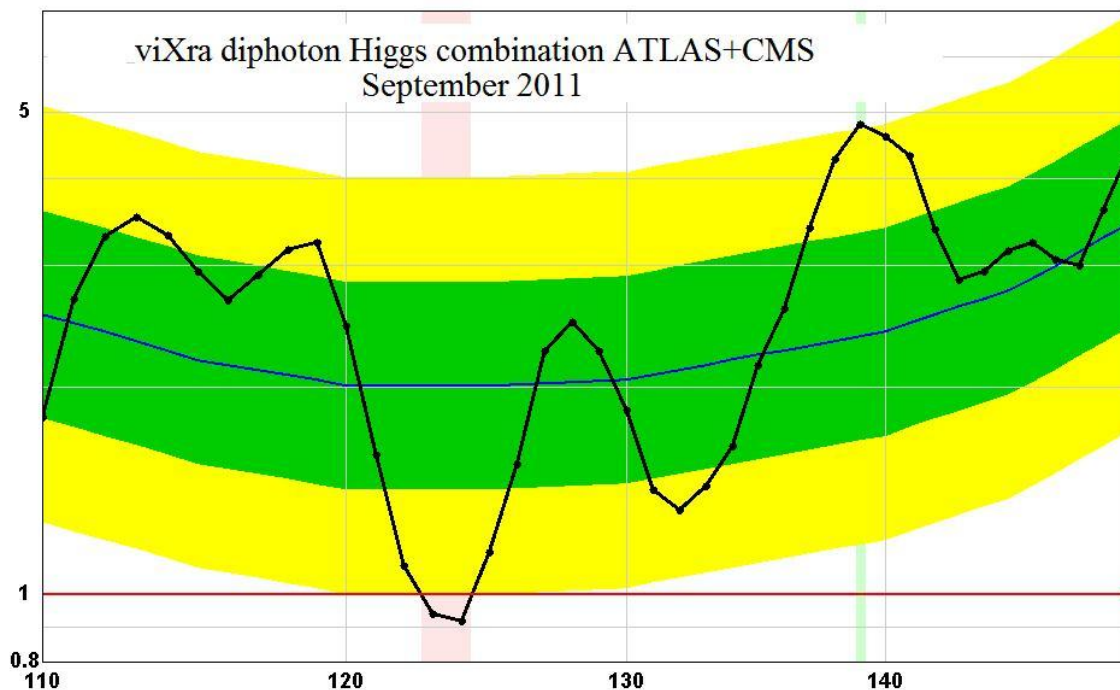
Some people say that there is no point producing these plots because the official versions will be ready soon enough, but they are missing the point. The LHC will produce vast amounts of data over its lifespan and these Higgs plots are just the beginning. The experimenters are pretty good at doing the statistics and comparing with some basic models provided by the theorists, but this is just a tiny part of what theorists want to do. The LHC demands a much more sophisticated relationship between experimenter and theorists than any previous experiment and it will be necessary to provide data in numerical forms that the theorists can use to investigate a much wider range of possible models.

As a crude example of what I mean, just look at the plot above. It provides conflicting evidence for a Higgs boson signal. At 140 GeV there is an interesting excess but it is below the exclusion limit line. Is this a hint of a Higgs signal or not? To answer this I might look at

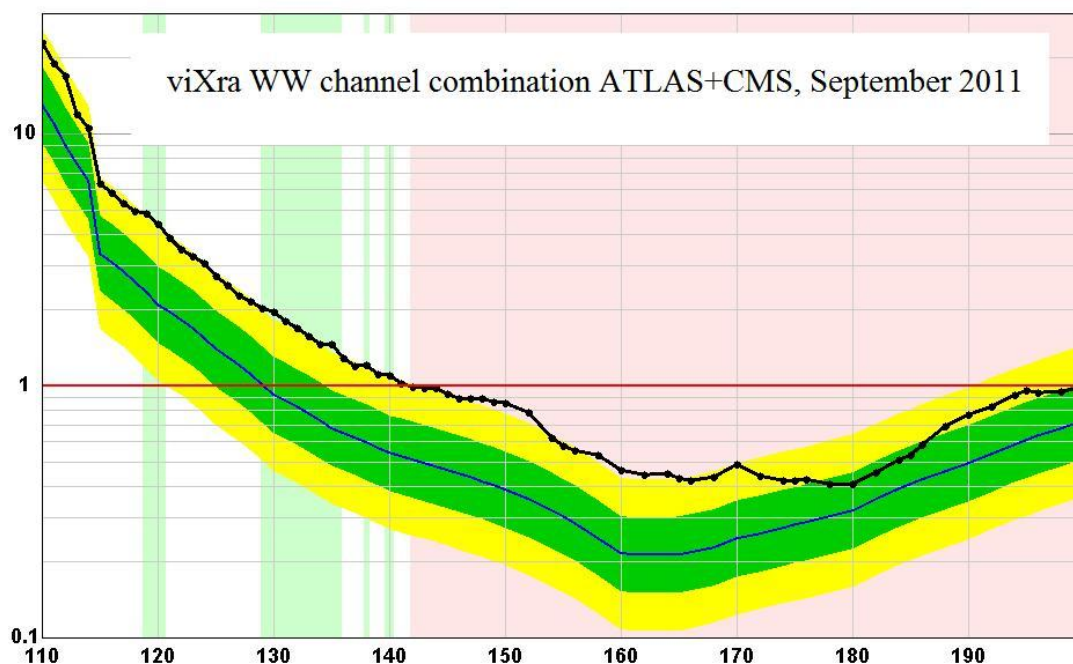
different channels combined over the experiments. Here is the ZZ channel combined over ATLAS and CMS.



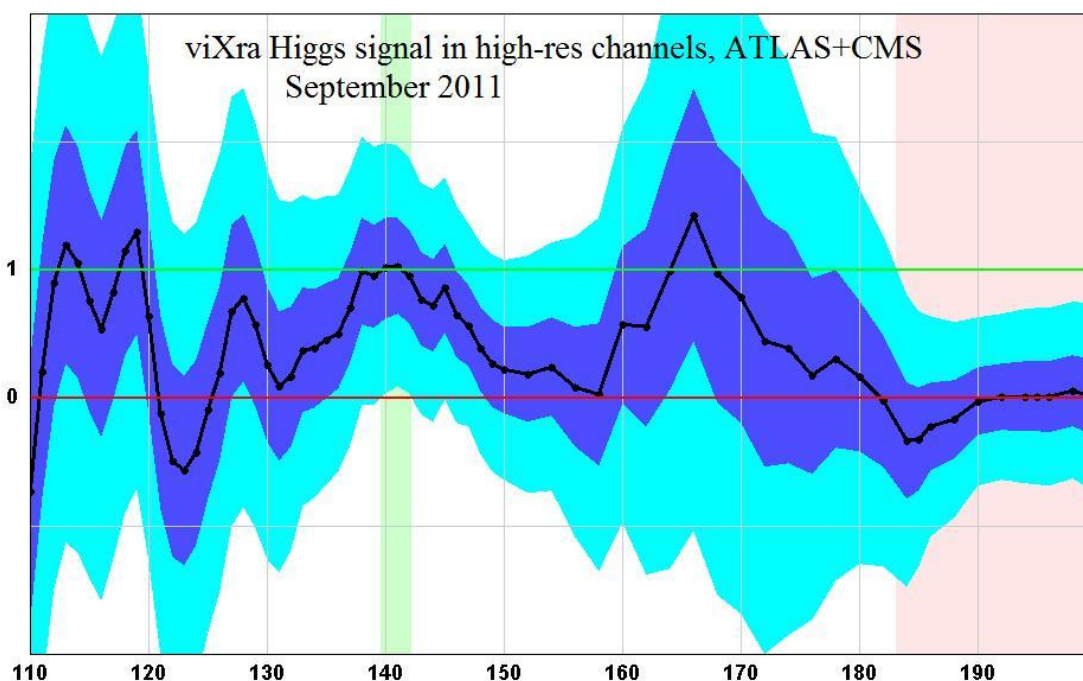
The Higgs hint at 140 GeV is now nice and clear, though not significant enough yet for a reliable conclusion. Here is the diphoton channel combination.



Again the 140 GeV signal is looking good. What about the WW channel?



Here is where the problem lies. The WW channel has a broad excess from 120 GeV to 170 GeV at 2 sigma significance, but it is excluded from about 150 GeV . In fact the energy resolution in the WW channel is not very good because it relies on missing energy calculations to reconstruct the neutrino component of the mass estimation. Perhaps it would be better to combine just the diphoton and ZZ channels that have better resolution. I can show the result in the form of a signal plot.



It's still inconclusive, but at least it is not contradictory.

This is just an example of why it will be useful for theorists to be able to explore the data themselves. The signal for the Higgs will eventually be studied in detail by the experiments, but what about other models? There is a limit to how many plots the experiments can show. To really explore the data that the LHC will produce theorists will need to be able to plug data into their own programs and compare it with their own models. The precise combinations produced by the Higgs combination groups take hundreds of thousands of CPU hours to build and are fraught with convergence issues. My combinations are done in milliseconds and give a result that is just as useful.

There is no reason why the experiments can't provide cross-section data in numerical form for a wide range of channels with better approximations than flat normal distributions if necessary. This would allow accurate combinations to be generated for an infinite range of models with varying particle spectra and branching ratios. It will be essential that any physicist has the possibility to do this. I hope that this is what the theorists will be telling the experiments at Santander next week and that the experiments will be listening.

References

1. <http://blog.vixra.org/2011/09/18/higgs-days-at-santander/>