Weekly report

Yu Zhang

IHEP

May 22, 2017

Section 3.2: Mention that any residual difference in pileup conditions is corrected for through a re-weighting. The vertex z distribution is also re-weighted to better match data.

- Ok.

Section 3.2.1:
* Why is Herwig++ used instead of Pythia8? (I think all of your other samples use Pythia.)
* Include a definition of particle-level (stable, ctau ¿ X ns, etc.) since this is shown in Figures 1 and 2

- Anyone can answer this?
- @Qi, again what is shown in Fig 1 and 2.

Section 3.2.2:
* Do you shift the SM higgs samples to 125.09 GeV?
* Why aren't you using the latest SM Higgs samples? i.e. NNLOPS for ggH, POWHEG for VH
* I think the ttH sample you're using is aMC@NLO?

- No shift currently.
- I don't know what the difference from new generator samples.
- DSID 341069 is not aMC@NLO. It is pythia directly.

Section 4.1: Like Karsten, I'm also curious how often photon pointing selects the correct vertex

- my work.

Section 4.X: Since your appendix mentions MET, a brief description of it's reconstruction would be nice here (maybe add after 4.4)
Section 5: Do you have any contamination from the second Higgs decaying to ZZ? (Instead of WW.)

- @Qi could you add this?
- Not considered yet.

Section 6: When optimizing, what do you consider as BG? Wenugammagamma and AF2 gammagamma?

- We inclusive lnujjgamgam sample as irreducible background.

* Can you expand a bit your statement "is chosen considering the sensitivity improvement and the loss of statistics." I don't see any numbers or plots about this in the appendix. Imagine someone trying to reproduce this analysis, and what they would need to know to choose the same cut as you.
* The 100 GeV pT_yy cut is only applied to 400/500 GeV samples. Can you add a few sentences why in the body of the text? Maybe expanding Table 7 to include expected limits on all the mass points would help motivate this.

- Qi, please add more explaination about pTyy cut.

Figure 3a: Could you double check that there are really no data points between 160-200 GeV?
Figure 3b: The legend is too small to read; is it easy to split it into two columns with larger text?

- @Qi, please reproduce the plot?

Section 7: I think this section (and Section 9) need to include more details. Reading through this, it's not clear to me exactly how the fit is being done. In the beginning of Section 7, it would be nice to include a description of how the fit will be done. i.e. you will do a simultaneous fit of 0/1-lepton regions, where the 0-lepton region has only continuous and single Higgs BGs, and 1-lepton has continuous, single Higgs, and Di-Higgs. (This is my understanding, at least.) I think you also do two separate fits: with the 100 GeV cut on pT_yy, and without. Is all that right? This description will help motivate the signal / BG modeling in the rest of the sub-sections.

- @Qi, rephrase this part

Section 7.1: Somewhere (possibly elsewhere) you should describe the DSCB function. Right now it's only briefly named describing the variables in the likelihood function in Section 9
Section 7.2: Do you assume SM yields in your fit, constrained by theory uncertainties? So SM Higgs contribution is profiled in the fit?

- to add the formula
- Yes

Section 7.3.1: I think you need to better motivate why you chose exp(pol2) as your BG model. e.g. in Figure (b) and (d) have the same chi2, why not use a 2nd order polynomial? Generally we use the spurious signal to select the function, since small chi2 differences aren't necessarily indicative of one function being better than another. You don't need necessarily need to do what's done in the coupling analysis, but you should quantitatively motivate it.

- Maybe spurious signal is still the most straightforward motivation...

Section 7.3.2: I think you also need to be more quantitative here, since by eye I don't see a consistent shape. e.g. in Figure 12 (using MC) I see a bias at low mass for both (a) and (b), and a bias at high-mass for (b). The same feature appears to show up in Figure 13 (b) when looking at data. I think the way you derive SS uncertainty covers this bias, but if this is your argument, you should clearly state this here.

- We will add this arguement in text

Table 13: You list different SS uncertainties (for different models) but it's not clear to me how you use the different values. (A description of what is fitted, and how, would help here.) I also don't understand why different signal points have different SS; for those with the same pT_yy cut (say 260 and 300 GeV) shouldn't the same BG be used in both cases, with no signal?

- One term, N_SS times signalPdf will be added into the final statistic model.
- The background dataset is exactly same. The only difference is signal shape. The impact is very small.

Figures 14 and 15: In the caption, it says "fitted number of spurious signal" but I think it's just the fits (no numbers are given)

- Right

Section 8.2:
* You mention theory uncertainties, but what are the uncertainties on? i.e. I think they're on the acceptance correction which takes extracted yield to cross-section in your fit, but I don't see them described / shown anywhere. Can you add a description of them somewhere?
* You mention PS uncertainties for Wh, but nothing else. Why don't you consider the uncertainty on ggH, using the difference between Pythia8 and Herwig7?
Section 8.3: You don't mention photon ID / isolation uncertainties, in the text, only in the tables.
Section 8.4: You discuss a transfer factor between 0/1-lepton regions; does this mean you don't let the normalization float between them? Do you fix it to the ratio of data side-band events in each region? I'm not sure I understand this uncertainty, it seems to be dominated by the statistical component. Can you please elaborate on it?

- need to descrip the uncertainty
- need to explain the reason why only consider this uncerntainty for Wh.
- I don't fully understand this uncertainty also. My preference is letting the normalization factor float.

Section 9:
*How do you derive the covariance matrices, with toys?
* When searching for an excess (before setting limits) do you perform a scan to determine the significance? Or rely on MINOS, etc.?
line 419: It might make sense to make the description of q_mu a new sub-section called "Limit Setting"

Figure 40:
* Is it fair to draw a line between the 300 and 400 GeV points, since they have different fiducial selections?
* Do I understand correctly that the observed limit on N_events will be the same for 260 and 300, though it will show up differently on this plot due to differences in acceptance?

- I think that line does not mean much
- The signal shape and systematic are also different, which could change the observed limit on N_events.