Hi Emily,
here's the continuation of the VBF-Higgs feasibility study. I've focussed on the H $\to \ell\ell\ell\ell$ channel, because it's easy and clean. The first issue is that the signal cross section of that channel is only about 0.4 fb (figure from Pythia 8 generation), meaning that even 200 fb^{-1} of data would only contain 80 signal events. Depending on the selection efficiency, this doesn't leave much to go on. The diphoton channel has a 20 times larger cross section, maybe I should focus on that one in the future.

I've written a Rivet analysis that selects events in three steps:

Jet selection	at least two anti-$k_\perp(0.4)$ jets with $p_\perp > 25$ GeV and $\lvert\eta\rvert < 4.4$ that have no leptons harder than 15 GeV within $\Delta R = 0.3$
Lepton selection	exactly 4 muons, or 4 electrons, or 2 muons and 2 electrons, with $p_\perp > 15$ GeV and $\lvert\eta\rvert < 2.47$ each, as well as same-flavour opposite charges. Leptons are *dressed*, i.e. accompanying photon energies clustered
Higgs candidate selection	tetralepton mass in range (126 ± 25) GeV

At the truth level, of the order of 80% of signal events pass the jet selection, and around 10% pass the lepton selection and Higgs candidate selection. At each selection step, the following variables of interest are plotted (as soon as the required objects have been selected):

- mass of the two hardest jets
- rapidity separation of the two hardest jets
- tetralepton mass
- tetralepton transverse momentum

Below are their distributions in the case of VBF-Higgs, ZZ, and non-VBF-Higgs production, normalised to the respective cross section and assuming an integrated luminosity of 50 fb^{-1}. (The plots are small, but vector graphics, so you can zoom in indefinitely.)

Please disregard the tetralepton mass distribution for now (except for the size of the contribution before and after the Higgs candidate selection), due to some plotting mistake they only contain two bins! The corresponding ASCII data files look alright, so it's really a plotting thing. I'll fix it.

Note the very different normalisations of the histograms due to the different cross sections.

If I understand correctly what we're going for, non-VBF Higgs production is a background in our analysis. The plots in the first row suggest that VBF and non-VBF production of Higgs bosons can be distinguished by considering the associated jets (caveat: Pythia 8 may not describe additional jets in non-VBF case well). However, cuts based on the jets mass are lossy (kill much signal), which is a problem if the sensitivity is already limited by the small signal cross section. Shape fits instead of a cut-and-count approach are probably a better way to go anyway?

The plots suggest that a mass window cut on the tetralepton mass is a very powerful way of killing ZZ background. With respect to the lepton selection, the selection efficiency of that cut is ~100% for any Higgs boson signal and ~0% for ZZ background — the perfect cut!

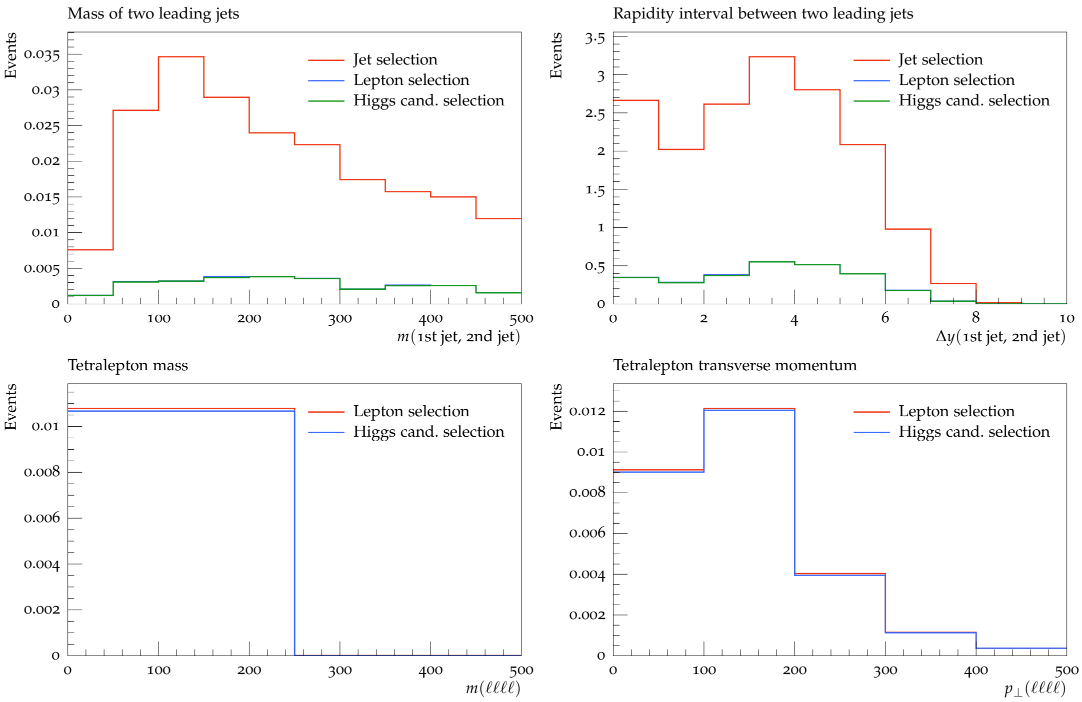

Figure 1: VBF Higgs in the tetralepton decay channel (based on 10k Pythia 8 events).

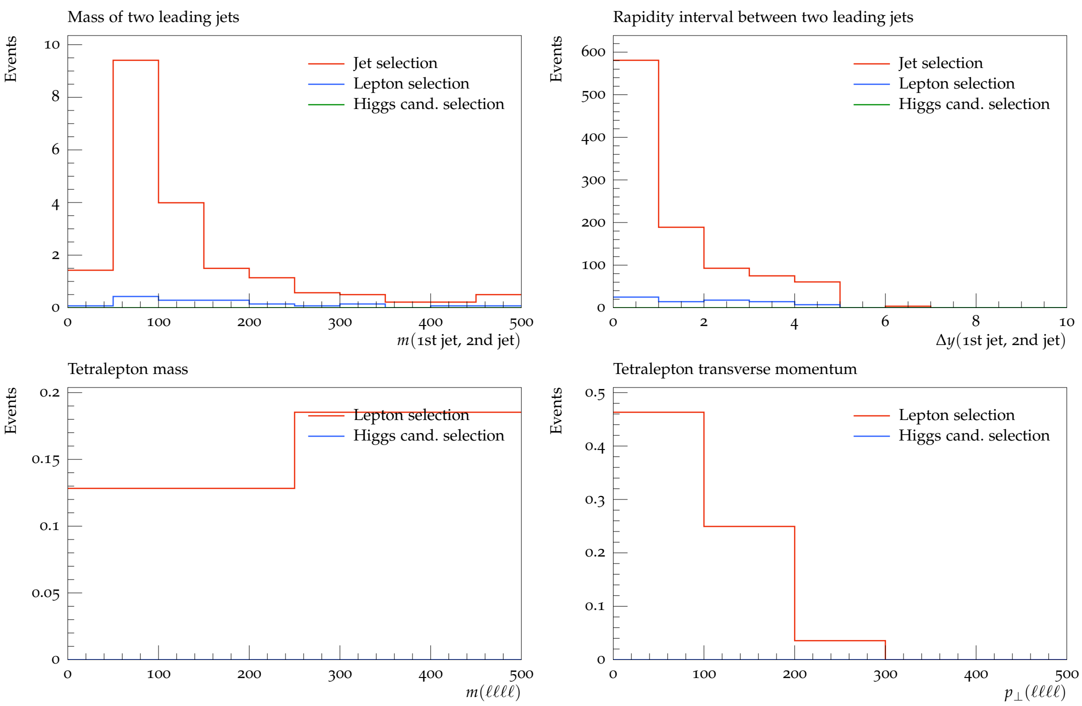

Figure 2: ZZ in the tetralepton decay channel (based on 1000 Pythia 8 events).

Conclusion

I think the tetralepton channel is not feasible on the timescale of my thesis because of its sub-femtobarn cross section. I think I should check out the diphoton channel next (will take 1-2 days), which has a much larger cross section, relatively speaking. I'll also reanalyse the tetralepton channel with loosened lepton selection and see if I can get significantly better

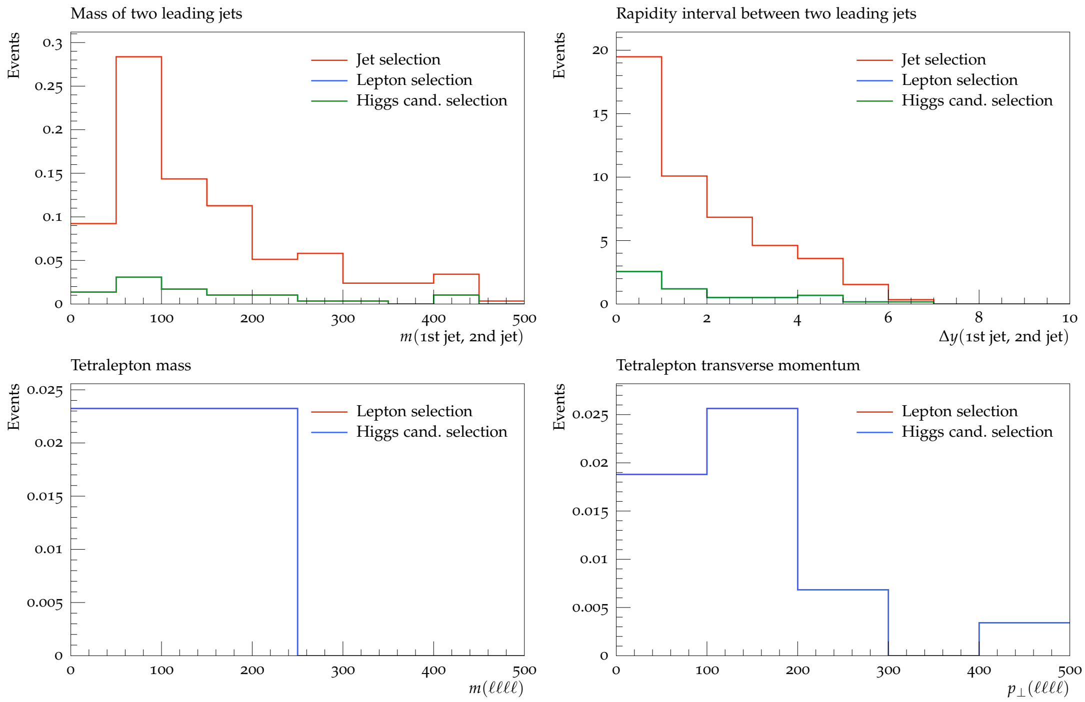

Figure 3: Non-VBF Higgs in the tetralepton decay channel (based on 1000 Pythia 8 events).

selection efficiency.

We now have a working Rivet routine to analyse events, so now we can try out lots of different cuts etc. I've automated almost every step from generating events to finalizing the plots so we can focus on the physics as much as possible!