To: Jose Monroy

From: Regina Werum

Re: Dissertation draft

Date 5/10/18

The substance of this dissertation is of course well beyond my area of expertise. That said, my contributions to your committee focus on the following aspects: I am reading it as a generalist with an eye towards science communication, and I am focusing on theories, hypotheses, and the experimental design.

As a generalist reader (imagine someone on an NSF panel from a different discipline to whom you have to communicate the scientific merit and broader impacts of the project), I recommend this:

* Clearly state in each chapter, not just the intro, that the two main goals of the dissertation are to answer the following questions (using my own words here):

1. “What causes the creation of a Higgs Boson with a single top quark/lepton-charged muon?” and
2. “How can we detect this particle more effectively or efficiently – how good is the device (FPIX modular detector) developed?”

As a social scientist, here are some additional observations and requests to clarify:

* In the intro and in the methods/data section you should make it clear how many times you ran the actual experiment, how many events are part of the Monte Carlo component. If you did state it someplace and I missed it, it means you need to state it more prominently. This may be clear to you because there are norms in your field that circumscribe these practices. But, for instance, in our field, demonstrating that asymptotic tendencies apply is important. For that you need a certain sample size, whether experimental or MC based. By extension, talking about confidence intervals only makes sense if there is a reasonable scope to the project.
* You also need to describe the 2016 full data sample more explicitly. Outside of your committee no one knows what’s in it – or whether you use the full sample (n=?) or a subsample of it with particular properties.
* Please explain WHY you have chosen to compare the results of the MC with the 2016 sample – this is implicit, but worthwhile spelling out: If I understand correctly, you want to know whether the device you developed is more effective at spotting those special particles than we would get if we randomly picked them out??? In other words: the comparison is intended to help demonstrate the quality of the device. (Or is it designed to demonstrate the quality of the data??? I got confused along the way).
* P. 151ff please clarify the process used to identify the muon IDs – is this done automatically by the detector you developed, or manually by the researcher? What if any intercoder reliability checks are needed or recommended in this circumstance?
* P. 155ff discussion of the jet-related variables: Please confirm and/or clarify: What are these variables? Are we talking about values of predictors of the outcome that are known to have been randomly miscoded or are missing? Or are these miscoded predictors associated with specific cases/events? Data errors? Do you treat them as random errors/white noise?
* P. 157 Table 6.4 – the cutoffs for the fakeable object and the tight object appear identical. Purpose? Can you explain more plainly why muons have to be in 3 types of selection efficiencies (loose, fake, tight) and how they get assigned to those groups?
* P. 164: preselection process: seems you are constraining the 2016 sample. From what to what? Reason? Consequence? (where are you eliminating variation and why)?
* One more question for you – found this note tucked into the folder on your diss:

p. 181: Input variables are correlated  -- how strongly?  Relates to Fig 6.18.  Could this affect multivariate modelling? Not sure what the norms are in physics regarding correlation and collinearity among input variables.

* Reading the graphs vs. Multivariate output: You put the graphs in the text and the multivariate analyses in the appendix. That is fine, provided the graphs are self-explanatory. I did not find them to be self-explanatory and thus turned to the multivariate results to get a better sense. I recommend that you discuss the graphs and what they portray and mean more explicitly. Tie to hypotheses and to theory. Simply referencing them and restating the empirical results is not sufficient.
* The conclusion could easily be expanded. It would be useful to pull up a level – let go of the technical specs and instead focus on the big picture. This is about the God Particle, after all.  Think like a science communicator here. Tell us all – not just experts in particle physics – what are the findings, why are they valuable, and to whom? How do your findings push forward what we know about the universe as scientists? How might your findings influence future research? Perhaps even research policy and funding? What refinements to the analyses do you plan do to (hinted at on p. 274-5)? Are those refinements intended to improve the predictive model/reduce uncertainties? Do you see potential for that detector you developed to have commercial value – if so, for whom? Finally, how can we extrapolate from your experiment to others in the field – link to particle physics more generally. How can insights derived from your work inform work in other areas of physics? Future research in your and other areas of physics?

Minor suggestions:

* Make sure to run a spell check and do check manually. For example “Chapter4” should be “chapter 4” on p. 150.
* Make sure to annotate endnotes consistently, for instance p. 150 you have a sentence “blablabla [endnote].” It should be “blablabla. [endnote]”