

Author answer to the referee report of JHEP_187P_0318

We thank the referee for the positive report and for his insightful comments and detailed comments. Below, we address all the questions and suggestions contained in the report in order. To ease the reading, the points raised by the referee (quoted in *italic*) are followed by an explanation of the actions undertaken to address them.

1. *At the beginning of Sec. 2 "LO" is defined as the three possible coupling combinations α^6 , $\alpha_s\alpha^5$, $\alpha_s^2\alpha^4$. However for most of the paper "LO" is more and more taken equivalent to only mean the EW α^6 contribution. (At first it still says "LO at α^6 ", then it becomes LO i.e. α^6 , and eventually in the LO+PS section the coupling order is dropped completely. This is of course just semantics, but it was quite confusing on the first reading and it would be good to make this a bit clearer from the start.*
2. *The ordering of the various programs in tables 1, 3, and 5 is very helpful. I can perhaps understand the reasoning behind choosing this alphabetical ordering, but I think it is misplaced here and makes it quite hard to get anything out of this. It would be much more useful to have programs with similar levels of approximations next to each other. So keep Phantom and Whizard together (both having the same approximations), then Bonsay, Powheg, VBFNLO (all without interference with VBFNLO adding the s-channel), then MG5_AMC (also adding interference), and last MoCaNLO+Recola having the full result.*
3. *In table 3: Please indicate which ones are exact at LO, and which ones are not. Also, regarding the discussion in Sec. 4.3 and the fact that several full LO predictions differ far outside their statistical MC uncertainties. This I find extremely puzzling and the offered explanations are quite unsatisfactory. After all, at LO, MC statistics is just statistics, I don't understand how statistical uncertainties can be aggressive or not, they are what they are. Also, the complex-mass scheme is a well-defined procedure, so how can different implementations of it lead to numerical differences? (If there are somewhat different approximations being employed, it would be good to spell this out clearly.) The fact that the differences are at the 0.5% level cannot hide the fact that they are much much bigger than the MC statistics. Taken at face value, that means that there are clear systematic differences where from what I understand there shouldn't be any, so the reader is left to wonder what is going on? To play the devil's advocate, if there are systematic differences at LO that cannot be understood, how can I trust the NLO comparison?*
4. *page 16, 2nd column, line 13, typo: "elment" \rightarrow "element"*
5. *page 18, 2nd column, line 48: "on the other ..." the part of the sentence is nonsensical.*
6. *page 19, 1st column, line 14: I think it is important to not mistake "more realistic" for "conservative". The scale variations are clearly underestimating the actual theoretical uncertainties. This does not mean that another method, which implies larger uncertainties, is "conservative", it is simply (hopefully) "more realistic". Conservative is easily (mis)interpreted as possibly overestimated, which I don't think is the case here. (In fact, there is no guarantee that even the comparison of different tools will provide an estimate that covers the next order, so calling this a conservative estimate is really misplaced.*