

APPENDIX B

VALIDATION OF EFT SAMPLES

This appendix provides details for checks that were performed to validate the LO EFT samples produced for this analysis. In Section [B.1](#), the validation of the LO matching procedure is discussed. In Section [B.2](#) validation of the starting point (for the reweighting procedure) is discussed.

B.1 Validation of leading order matching procedure for EFT samples

When using LO matched samples, validation should be performed to ensure that the matching procedure is properly filling the overlapping phase space. As explained in Section [3.2.1](#), this validation is especially important for EFT samples, since the matching procedure can lead to complications when applied to EFT samples because EFT effects are included in the ME contribution, but not in the PS contribution. This appendix presents empirical validation checks for the matching procedure.

The differential jet rate (DJR) can be used as a method of validation for the LO matching procedure [\[74\]](#) [\[75\]](#). For the k_T algorithm, the DJR histogram represent the distribution of k_T values for which an n jet event transitions to an $n + 1$ jet event. A discontinuity in the transition between the n and $n + 1$ curves would indicate that there is a mismatch in the overlapping regions of phase space, while a smooth transition is an indication that the ME generator and PS generator are working together to properly fill the phase space without any gaps or double counting. For the matched samples generated for this analysis, we observe smooth DJR plots, as shown in Figure [B.1](#)

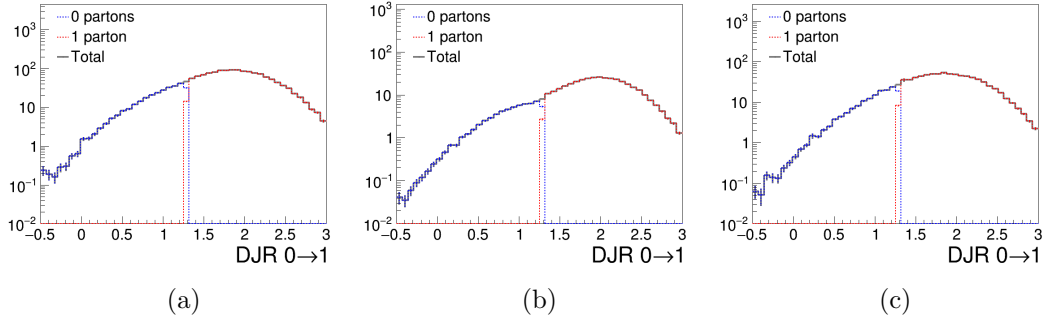


Figure B.1. DJR histograms for LO matched samples $t\bar{t}H$ (a), $t\bar{t}l\bar{l}$ (b), and $t\bar{t}l\nu$ (c). The distributions have been reweighted to a non-SM point.

In Figure [B.1](#) all WCs have been set to non-zero values. The x axis shows the log base 10 of the scale at which an n jet event transitions into an $n + 1$ jet event. The line labeled “0 partons” refers to the contribution from the parton shower, while the line labeled “1 parton” refers to the contribution from the matrix element. The line labeled “Total” is the sum of the two contributions. The smooth transition indicates that the matching scales have allowed the matrix element generator and parton shower to smoothly fill the overlapping phase space.

For the $t\bar{t}X$ samples generated for this analysis, the matching scales used with MadGraph and Pythia (i.e. the `xqcut` and `qQut`) were 10 and 20, respectively. A more detailed description of the meaning of these matching scales is provided in [\[9\]](#).

DJR plots should be studied at the SM and at non-SM points. It is also important to vary the matching scales around the nominal values and produce DJR plots for these variations as well. While it would not be unexpected for the DJR plots to look somewhat less smooth farther away from the nominal values (since the nominal values in principle should be chosen such that they produce smooth DJR plots), there should not be any large discontinuities. As an example of what would be considered to be a large discontinuity, Ref. [\[9\]](#) may be referenced (especially Figure 7).

In order to account for uncertainties associated with the matching scale, a systematic uncertainty should be evaluated. In this analysis, we evaluated this uncertainty by producing samples with the `qQut` varied around the nominal (to `qQut = 15` and `qQut = 25`). As a quick check for large/unexpected differences caused by varying the scales, the one-dimensional inclusive cross section quadratic parameterizations for each sample (for the up, down, and nominal variations) can be compared for all WCs. Note that this check can be performed at gen level, and can be checked before producing large statistics samples for the systematic uncertainty evaluation (for this analysis, we performed this inclusive check even before producing the large statistics nominal `qQut` samples). For this analysis, the inclusive checks were consistent, so the N_{jet} yields in the signal region were studied for the up and down variations. The size of the variations were very small, and were determined to be not larger than statistical uncertainty of the sample. For this reason, a `qQut` uncertainty was not included for this analysis.

B.2 Validation of starting points for EFT samples

As discussed in Section [3.2.2](#), it is important to perform checks to ensure that the reweighting is working properly. This section will step through some examples of how this validation may be performed. Additionally, the slides in Ref. [\[76\]](#) provide an example of the types of validation checks performed for this analysis.

For this analysis, we considered several candidate starting points. We considered several similar starting point to the TOP-19-001 analysis (one that was identical to the TOP-19-001 analysis for the 16 WCs in common, and two that were based on the limits obtained in TOP-19-001), as well as starting points where the WCs scaled the inclusive cross section by a given amount (e.g. 10%, 30%, etc.). The samples can be reweighted to their own starting points, as well as to the starting points of the other candidate samples. To check for consistency, the reweighted cross section can

be compared to the cross section of the dedicated sample.

For example, let us consider sample A with a starting point of \vec{a} and sample B with a starting point of \vec{b} . We can reweight A to \vec{a} , and compare the reweighted cross section against the original cross section at \vec{a} (since \vec{a} is the starting point of A, the original cross section at this point is included in the LHE file). Sample A can also be reweighted to \vec{b} , and this cross section can be compared to the original cross section at \vec{b} (since this is the starting point of sample B). The cross section should also be checked at the SM. Since the SM should always included as one of the reweight points, the reweighted SM cross section can be compared against the original SM cross section for every sample.

It is expected that the agreement will be better for a sample's own starting point than at an arbitrary point in WC space. It is also expected that the agreement will be better at points that are more similar to a sample's starting point than at points that are further from the sample's starting point. However, it is difficult to define a universal threshold that determines whether a starting point is good or bad. Nevertheless, to provide some quantitative numbers for future reference, in general agreement much better than 1% was expected at the sample's own starting point (on the order of 0.001%). At reasonable points in the WC space (e.g. where all of the WCs scale the process by 30%), agreement at better than 1% was deemed to be good. At more extreme points in the WC space (e.g. points where all of the the WCs scale the cross section by 50%), larger disagreement (a couple percent) was deemed to be acceptable. At very extreme points in the WC space (e.g. points where the WCs scale the cross section by 5 times the standard model cross section) even larger disagreement (e.g. 5-10%) was deemed to be acceptable, as these regions are unlikely to be explored in the analysis.

As mentioned in Section [3.2.2](#), the distributions of the weights at the candidate starting points should also be checked. Figures [B.2](#) and [B.3](#) show some example

weight distributions for the candidate starting point samples considered in this analysis. With this check, we are attempting to ensure that the chosen samples have good statistical power. We are also looking for single events with very large weights (which degrade the statistical power of the sample, and can also be an indication that the starting point was not good). To check the distribution of weights, we reweight the samples to the SM and to various points in EFT space, and plot a histogram of the values of the weights of the events in the sample. For example, in plot (c) of Figure [B.2](#), the blue distribution (“run0”) is relatively good (as its peak is very sharp and tall, indicating that most of the events have a similar weight), while the yellow distribution (“run 5”) is relatively bad (as is very broad, so will not provide good statistical power).

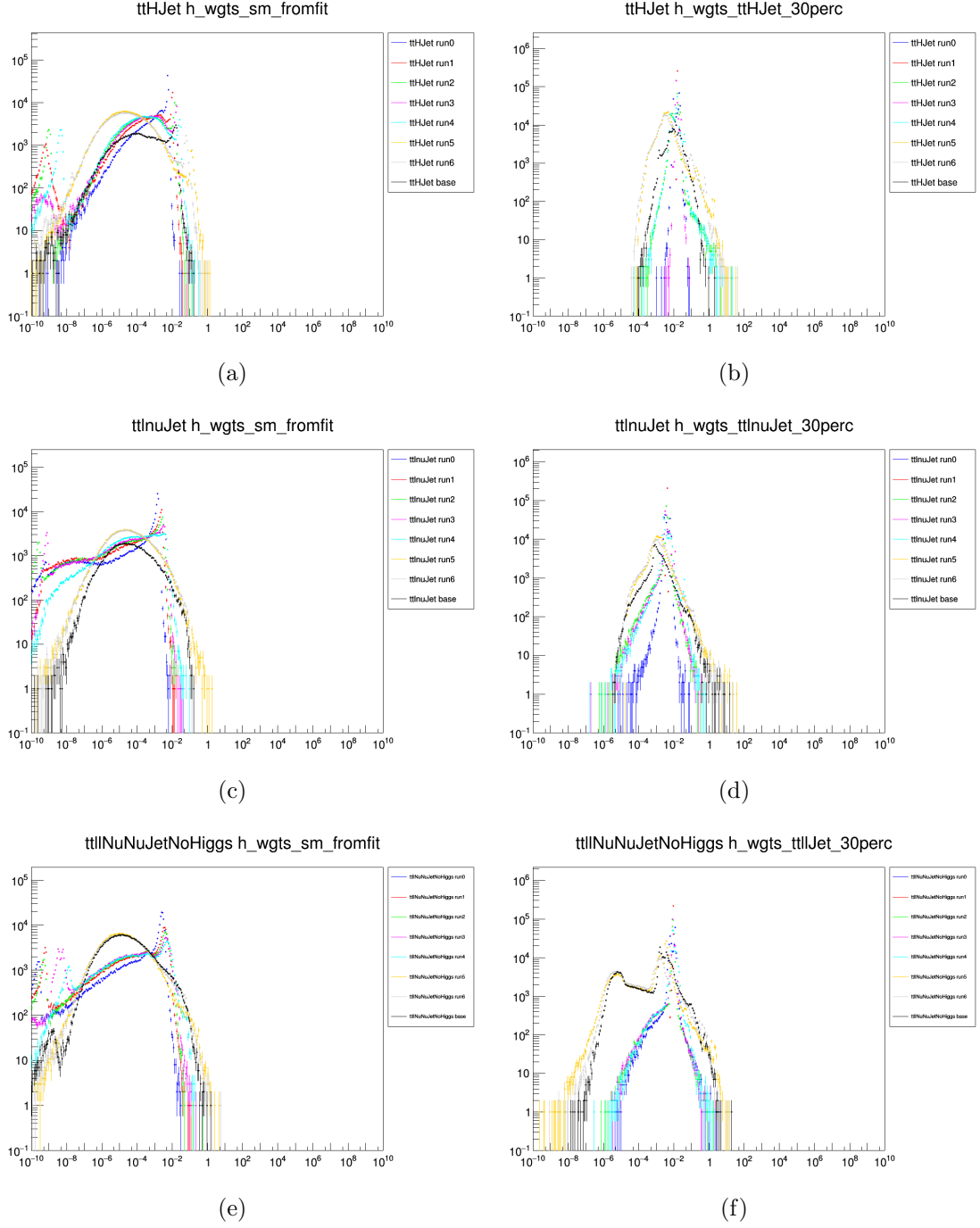


Figure B.2. Distribution of weights for $t\bar{t}H$ at the SM (a) and a non-SM point (b), $t\bar{t}l\nu$ at the SM (c) and a non-SM point (d), and $t\bar{t}l\bar{l}$ at the SM (e) and a non-SM point (f).

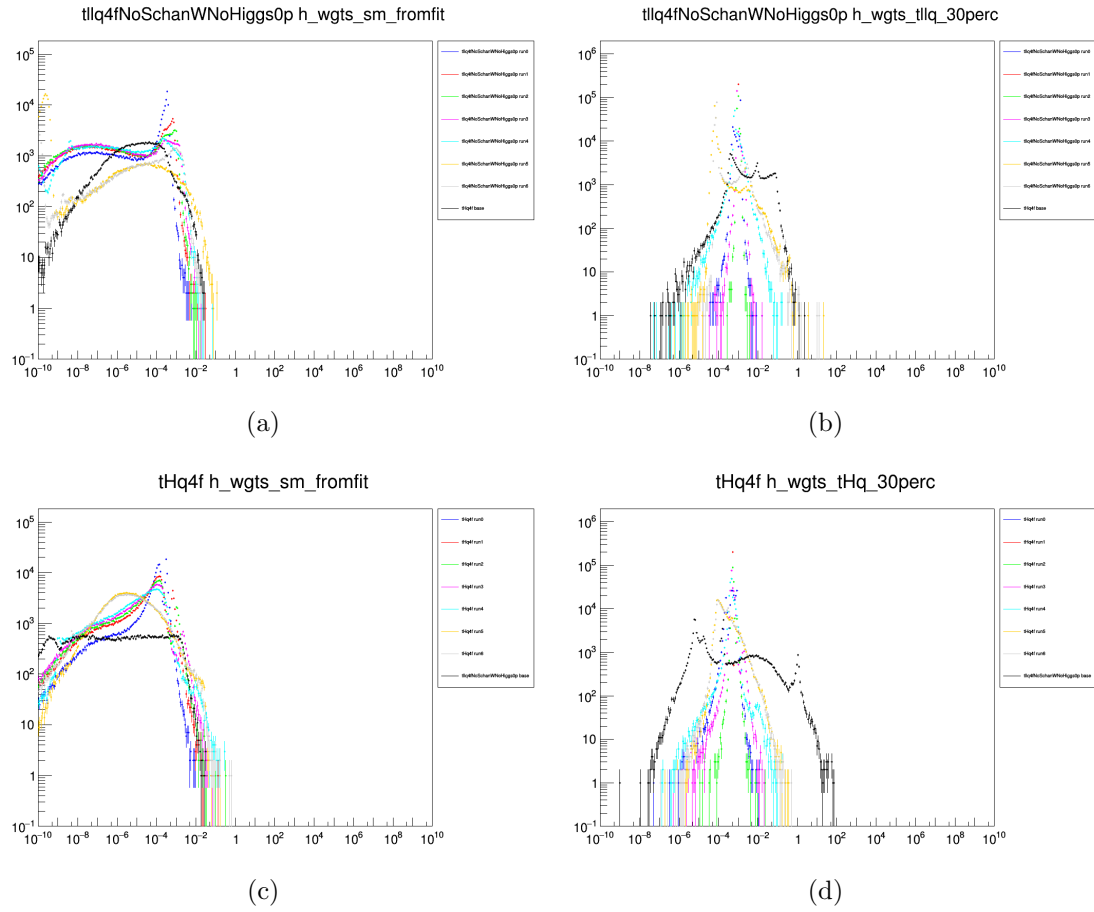


Figure B.3. Distribution of weights for $t\bar{t}l\bar{l}q$ at the SM (a) and a non-SM point (b) and tHq at the SM (c) and a non-SM point (d).