Chapter 6

Signal and background predictions

The modeling of physics processes relevant to the $H \to \tau\tau$ analysis are described, with emphasis on the VBF $H \to \tau_\ell \tau_{\text{had}}$ channel. This draws from internal documentation of the recent ATLAS $H \to \tau\tau$ publication [103].

6.1 $Z \to \tau\tau$

The $Z \to \tau\tau$ process constitutes a major and irreducible background to all three final states of the $H \to \tau\tau$ analysis. Its modeling is therefore critical. It is also challenging to validate because the poor mass resolution of $m_{\tau\tau}$ implies finding a region of data orthogonal to the $H \to \tau\tau$ signal regions but rich in $Z \to \tau\tau$ events is not possible.

6.1.1 $Z(\to \ell\ell)$+jets in simulation

The simplest approach is to use simulation to model $Z \to \tau\tau$. Unfortunately, ATLAS has observed in the $Z \to ee$ and $Z \to \mu\mu$ processes that mis-modeling is present in various aspects of $Z(\to \ell\ell)$+jets kinematics. These aspects include the Z p_{T} and dijet kinematics, as shown in Fig. 6.1 and Fig. 6.2, respectively.

These mis-modelings are worrisome for $H \to \tau\tau$ analyses because they rely on accurate modeling of these kinematics. For example, mis-modeling in p_{T}^Z is problematic because this variable defines the boosted category of the $H \to \tau\tau$ analysis. It is also strongly correlated with discriminating variables like $\Delta R(\tau\tau)$. Mis-modeling in dijet kinematics like m_{jj} is of even greater concern because they are among the most powerful and high-profile discriminating variables in the VBF category.

Some versions of the ATLAS $H \to \tau\tau$ analysis use simulated $Z \to \tau\tau$ with corrections derived from $Z \to \ell\ell$ events in data [89]. While helpful, these corrections are one-dimensional and cannot

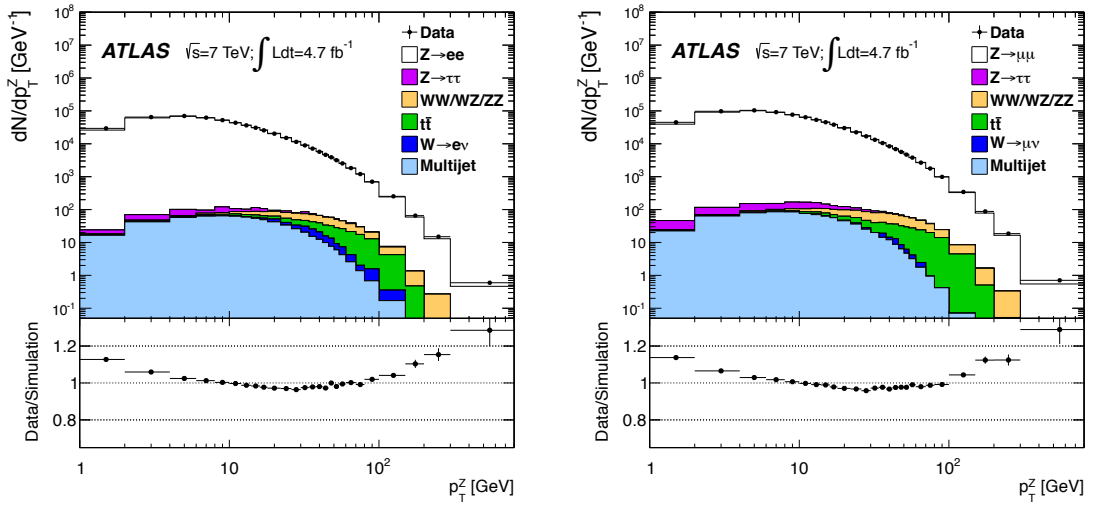

Figure 6.1: Comparison of data and various predictions of p_{T}^{Z} for $Z \to ee$ (left) and $Z \to \mu\mu$ (right) in 2011 data-taking [104]. Mis-modeling is observed.

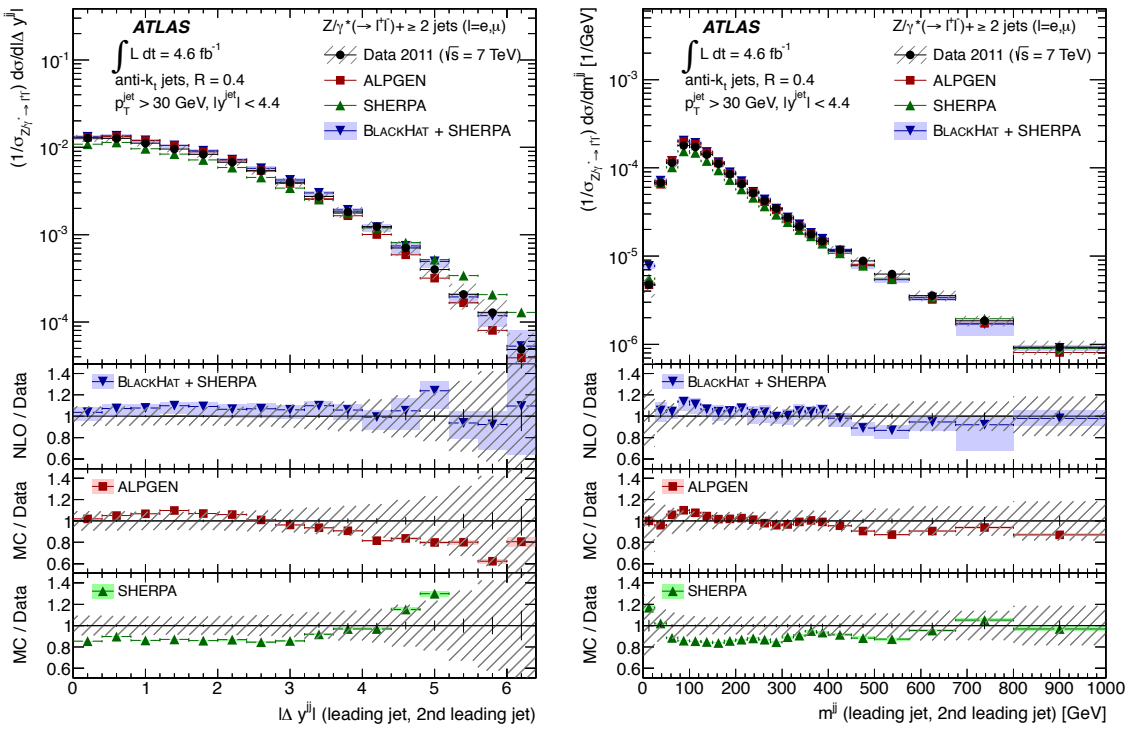

Figure 6.2: Comparison of data and various predictions in $Z \to \ell\ell$ events of $\Delta y(jj)$ (left) and m_{jj} (right) in 2011 data-taking [105]. Mis-modeling is observed for all predictions.

account for potential correlations in the mis-modeling. For these reasons, this approach is not used in the recent publication.

6.1.2 Embedding

A more data-driven approach to modeling $Z \to \tau\tau$ is used wherein $Z \to \mu\mu$ events are tagged in data and the muons are replaced with simulated tau lepton decays. This exploits lepton universality in Z decays and has the great advantage of taking all $Z+$ jets features directly from data, such as Z p_{T}, dijet kinematics, and soft hadronic activity. Only the tau lepton decays and the detector response of the decay products are taken from simulation. The former is measured with excellent precision at B-factories [106], and the latter is an ongoing area of study within ATLAS detector performance groups.

$Z \to \mu\mu$ events are selected in data by requiring an event fire the lowest unprescaled dimuon trigger and have at least two reconstructed muons with $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 2.5$. All possible pairs of muons are then considered which satisfy $p_{\mathrm{T}}^{\mathrm{lead}} > 20$ GeV, muon isolation requirements, have opposite charges, and have $m_{\mu\mu} > 40$ GeV. The pair which has mass closest to the Z mass is then chosen as the Z decay products.

Tau lepton decays are then simulated with `TAUOLA` with the same four-momenta as the muons associated to the Z decay and sent through the full ATLAS detector simulation, digitization, and reconstruction. The decays can be set to whatever final state desired (e.g., $\tau_e\tau_{\mathrm{had}}$) within `TAUOLA`.

The simulated $\tau\tau$ system is then merged with the data $Z \to \mu\mu$ event by removing tracks and calorimeter cells associated to the muons and inserting tracks and cells from the tau lepton decays. For subtracting the calorimeter cells, deposited cell energies are derived from a simulated $Z \to \mu\mu$ event with the same kinematics as the data $Z \to \mu\mu$ event. The hybrid event, with a simulated $\tau\tau$ system and a $Z+$jets event from data, is then re-run through the ATLAS reconstruction, yielding the so-called *embedded* $Z \to \tau\tau$ event. Event displays of this procedure is shown in Fig. 6.3.

6.1.3 Validation

Various steps of the embedding procedure are validated with creative choices of output and input datasets of the embedding algorithms. For example, to test the subtraction of data muons, the embedding procedure is run on data $Z \to \mu\mu$ events merged with simulated $Z \to \mu\mu$ decays, and the output is compared with the original data $Z \to \mu\mu$ events. For a global test of the fidelity of the method, the embedding procedure is run on simulated $Z \to \mu\mu$ events merged with simulated $Z \to \tau\tau$

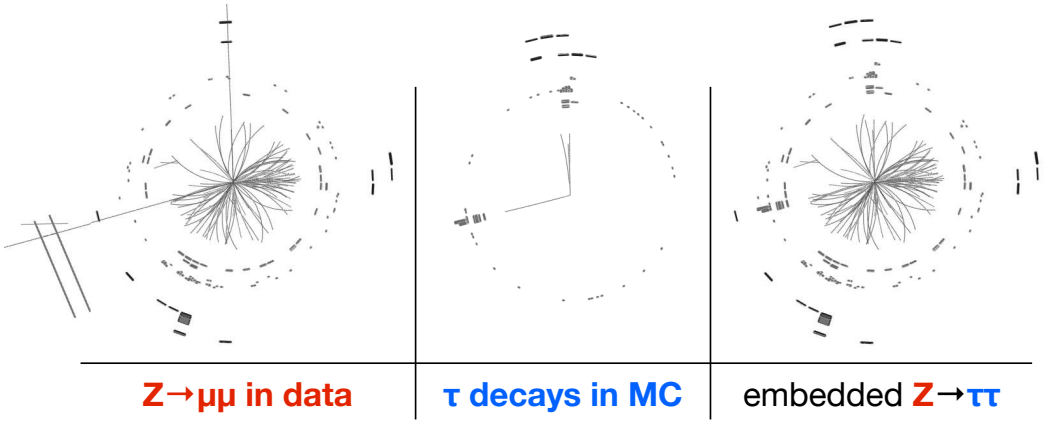

Figure 6.3: Event displays of the three types of events considered in the embedding procedure: a $Z \to \mu\mu$ event in data (left), a $\tau_{\text{had}}\tau_{\text{had}}$ event in simulation (center), and a hybrid embedding event (right) [107].

decays, and the output is compared with simulated $Z \to \tau\tau$ events. The results are shown in Fig. 6.4, and no significant biases are observed.

6.1.4 Uncertainties

Since all Z kinematics are taken directly from data, no uncertainties regarding to Z or jet kinematics are considered. However, uncertainties regarding the response of simulating tau decay products and the embedding procedure itself are considered. The uncertainty on the detector response is implemented via the typical collection of uncertainties pertaining to the measured identification efficiency and energy calibration of simulated leptons and τ_{had} at ATLAS.

Two uncertainties regarding the embedding procedure are considered. First, the isolation criteria on the data muons are either relaxed or tightened to test the dependence of the prediction on the $Z \to \mu\mu$ selection criteria. Second, the amount of cell energy subtracted when removing the data muons is varied by 20%, which is commensurate with the observed differences in the isolation energy between simulated $Z \to \mu\mu$ events merged with simulated $Z \to \tau\tau$ decays and simulated $Z \to \tau\tau$ events.

The pre-fit impact of these uncertainties on the $Z \to \tau\tau$ prediction is shown in each bin of the VBF discriminator in Fig. 6.5. The largest uncertainty at high VBF BDT score is the nearly 30% statistical uncertainty on the prediction, which is an inevitable limitation of the embedding procedure

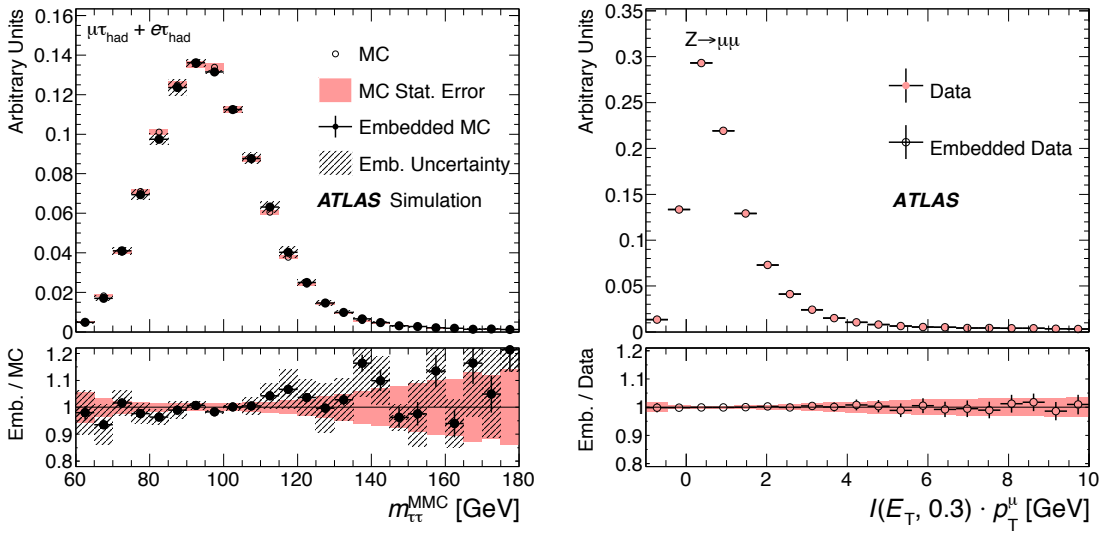

Figure 6.4: Validation of the embedding technique for simulated tau lepton decays in simulated $Z \to \mu\mu$ events (left) and simulated muons in data $Z \to \mu\mu$ events (right) [2]. Good agreement is observed in both, for the $m_{\tau\tau}^{\text{MMC}}$ (left) and isolation energy (right).

since it relies on the finite number of $Z \to \mu\mu$ events in data. The largest systematic uncertainties are components of the τ_{had} energy scale uncertainty, which are each 10-15%. These propagate directly to shifts of the $m_{\tau\tau}^{\text{MMC}}$.

6.2 $j \to \tau_{\text{had}}$ mis-identification

The largest background in the VBF $H \to \tau_\ell \tau_{\text{had}}$ analysis is from events where a jet is mis-identified as a τ_{had} ($j \to \tau_{\text{had}}$), also called *fakes*. The use of data-driven approaches to the prediction is therefore crucial. Unlike the $Z \to \tau\tau$ background, many regions of data exist which are rich in $j \to \tau_{\text{had}}$, and these regions can be exploited for prediction and validation. The largest sources of $j \to \tau_{\text{had}}$ are $W(\to \ell\nu_\ell)$+jets, QCD, top, and $Z(\to \ell\ell)$+jets events.

6.2.1 $j \to \tau_{\text{had}}$ in simulation

Like the $Z \to \tau\tau$ background, simulation is a simple but deficient means of predicting the $j \to \tau_{\text{had}}$ background. ATLAS observes mis-modeling in descriptions of jet shapes like the track width and track multiplicity [108], as shown in Fig. 6.6, which τ_{had} jet discriminators rely heavily on. This is especially problematic for τ_{had} because the identification algorithms emphasize tails of distributions like track width, not the bulk, hence detailed corrections to the simulation can be statistically limited.

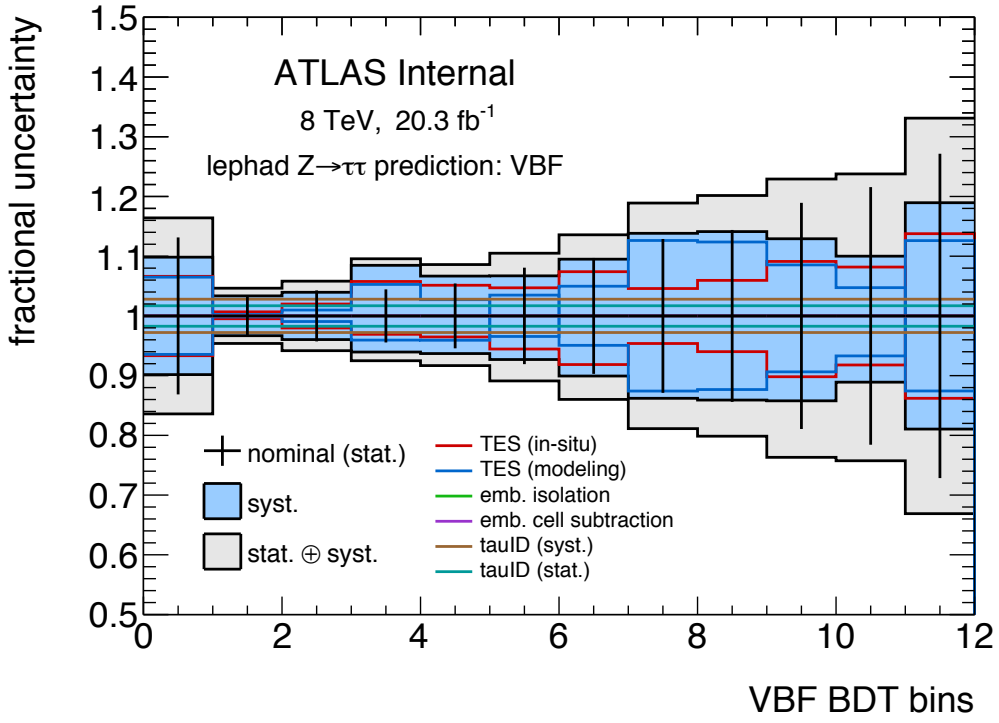

Figure 6.5: The pre-fit fractional uncertainty on the embedded $Z \to \tau_\ell \tau_{\text{had}}$ prediction in each bin of the VBF category for uncertainties pertaining to the embedding procedure and τ_{had} performance.

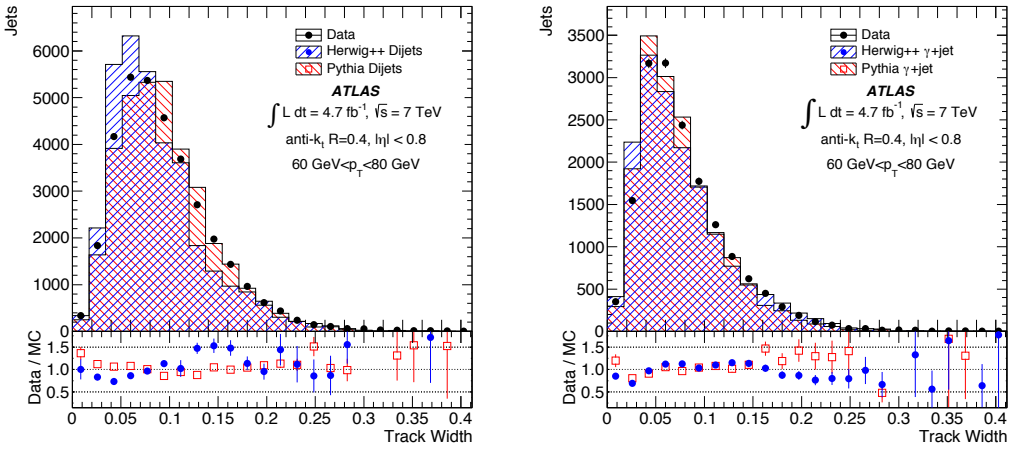

Figure 6.6: Comparison of data and various predictions of jet track width in dijet (left) and γ+jet (right) events [108]. Mis-modeling is observed for all predictions.

Additionally, the event kinematics of $W(\to \ell\nu_\ell)$+jets and $Z(\to \ell\ell)$+jets events have known mis-modeling in simulation. The mis-modeling of $Z(\to \ell\ell)$+jets events is discussed in Section 6.1.1, and ATLAS observes comparable mis-modeling in dijet kinematics of $W(\to \ell\nu_\ell)$+jets events, as shown in Fig. 6.7. The mis-modeling of variables like $\Delta y(jj)$ and m_{jj} is of concern since the VBF discriminators depend heavily on these kinematics.

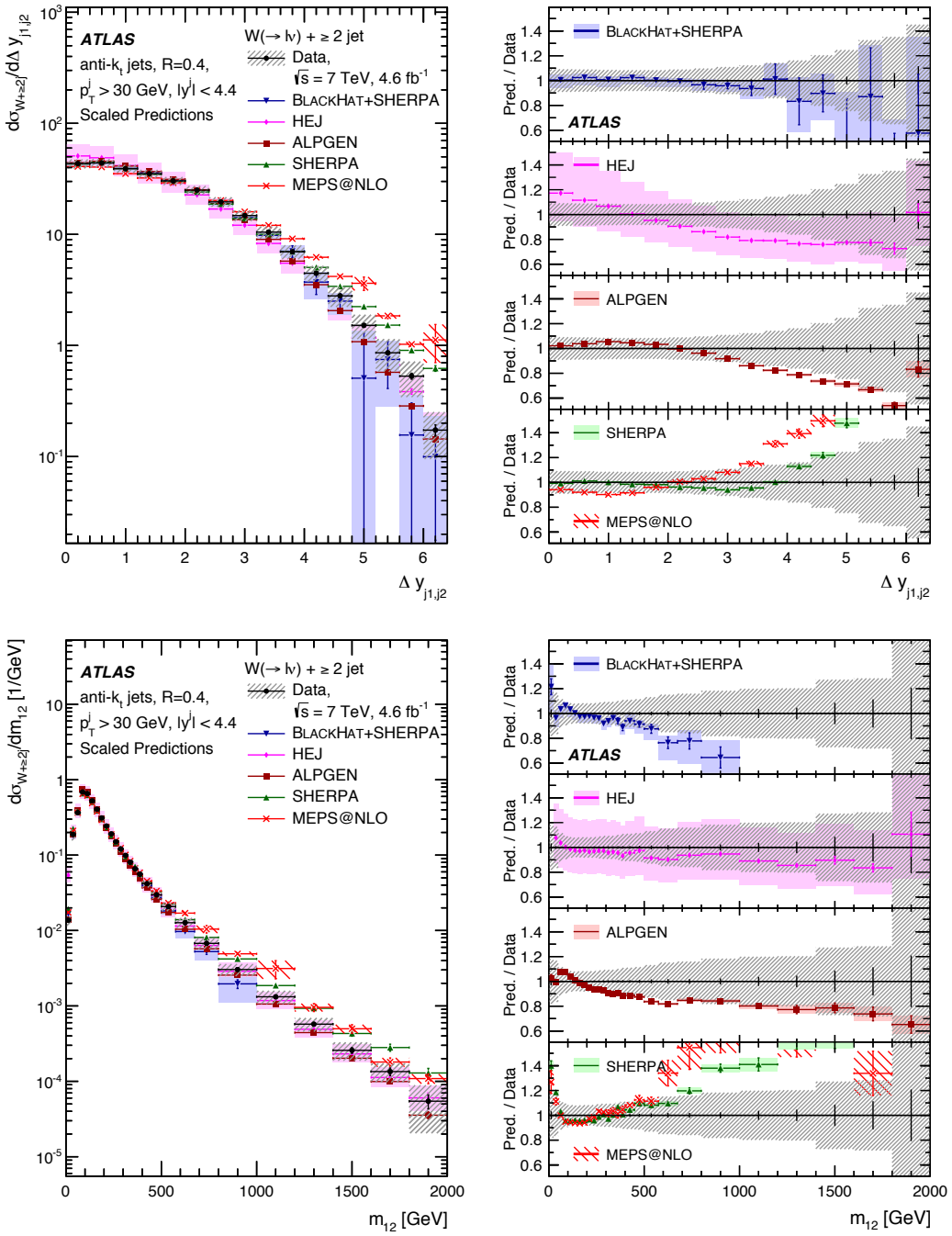

Figure 6.7: Comparison of data and various predictions in $W(\rightarrow \ell\nu_\ell)$+jets events of $\Delta y(jj)$ (top) and m_{jj} (bottom) in 2011 data-taking [109]. Mis-modeling is observed for all predictions.

Previous ATLAS $H \to \tau\tau$ analyses have used simulated $W(\to \ell\nu_\ell)$+jets, $Z(\to \ell\ell)$+jets, and top events with corrections derived from data, in conjunction to same-sign data events, to model $j \to \tau_{\text{had}}$ [110, 89]. While helpful, these corrections are one-dimensional and cannot account for potential correlations in the mis-modeling, and the same-sign data sample has large statistical uncertainties. For these reasons, this approach is not used in the recent publication.

6.2.2 Fakefactor method

6.2.2.1 Principle

An alternative data-driven approach is taken wherein events in data which pass all the signal region requirements, but fail the τ_{had} identification algorithm, are used to predict $j \to \tau_{\text{had}}$. The principle of this extrapolation is that τ_{had} identification is uncorrelated with event kinematics like m_{jj}. The kinematics of events where the τ_{had} fails identification criteria then provide an unbiased prediction of $j \to \tau_{\text{had}}$ kinematics in events where the τ_{had} passes identification.

The anti-identified data sample has a high purity of $j \to \tau_{\text{had}}$, as shown in Figs. 6.8 and 6.9. The residual contamination of $Z \to \tau_\ell\tau_{\text{had}}$ and other processes without $j \to \tau_{\text{had}}$ is nonetheless subtracted from the data to form the $j \to \tau_{\text{had}}$ estimate. The high purity is helpful because uncertainties on the predicted contamination (e.g., the tau energy scale uncertainty for $Z \to \tau_\ell\tau_{\text{had}}$) are evaluated to have a negligible impact on the $j \to \tau_{\text{had}}$ estimate and can be ignored.

The correlation between the τ_{had} identifier and event-level kinematics is checked in data in the VBF same-sign region, as shown in Fig. 6.10. No strong correlations are observed for any event-level kinematic variable, including the final BDT discriminator.

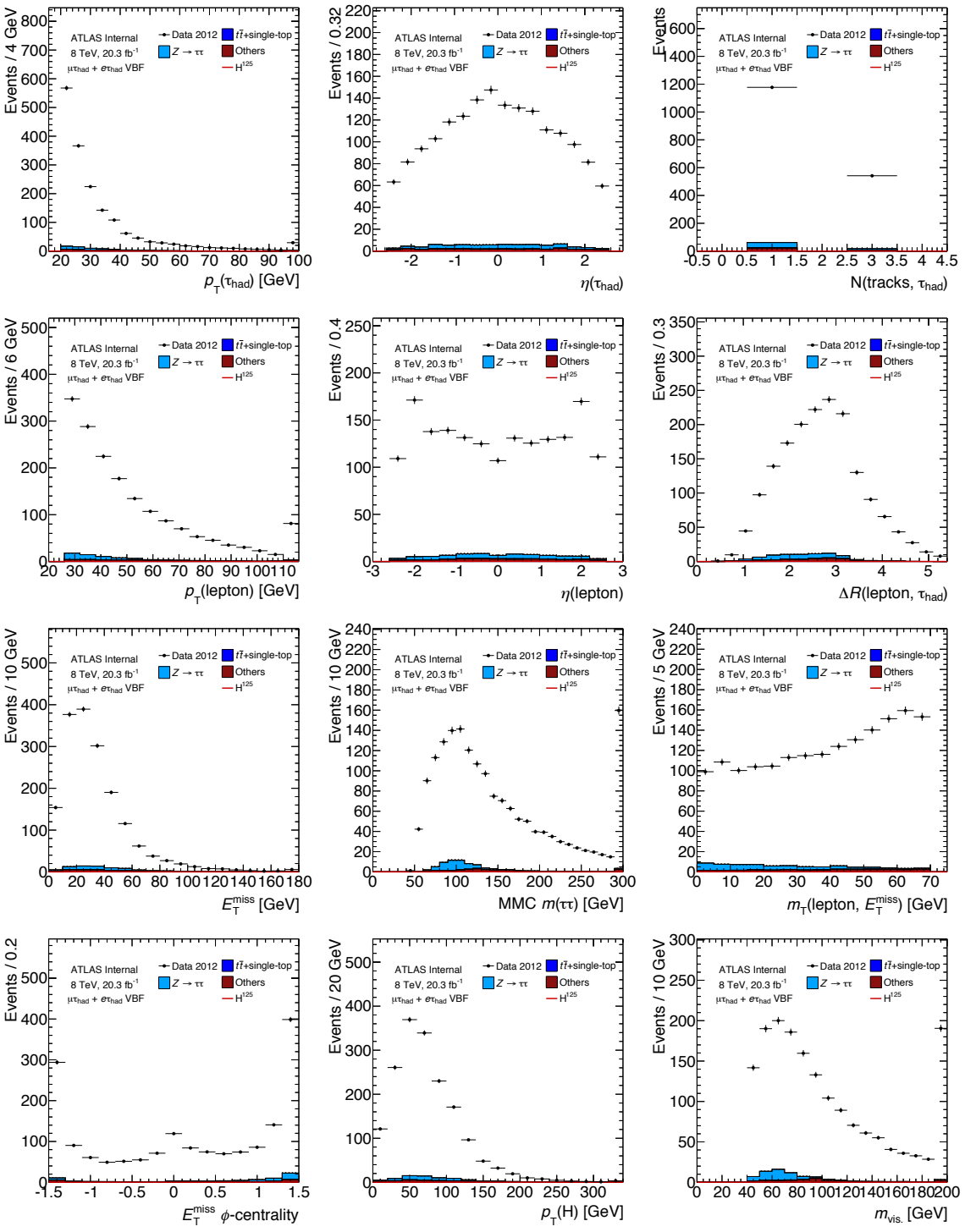

Figure 6.8: Data events in the VBF category which fail τ_{had} identification but fulfill all other requirements. The contamination of $Z \to \tau_\ell \tau_{\text{had}}$ and other processes without $j \to \tau_{\text{had}}$ is less than 10%.

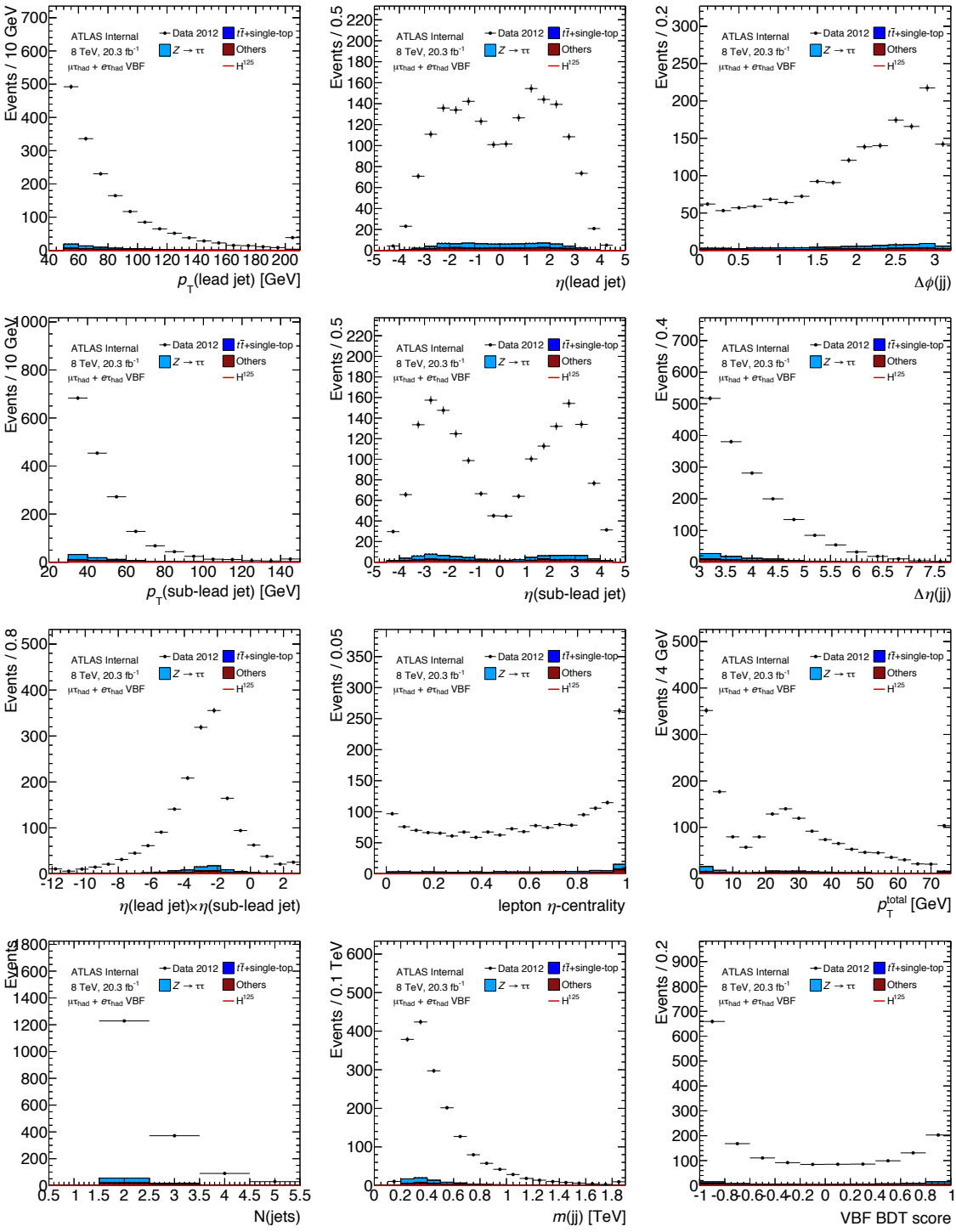

Figure 6.9: Data events in the VBF category which fail τ_{had} identification but fulfill all other requirements. The contamination of $Z \to \tau_\ell\tau_{\text{had}}$ and other processes without $j \to \tau_{\text{had}}$ is less than 10%.

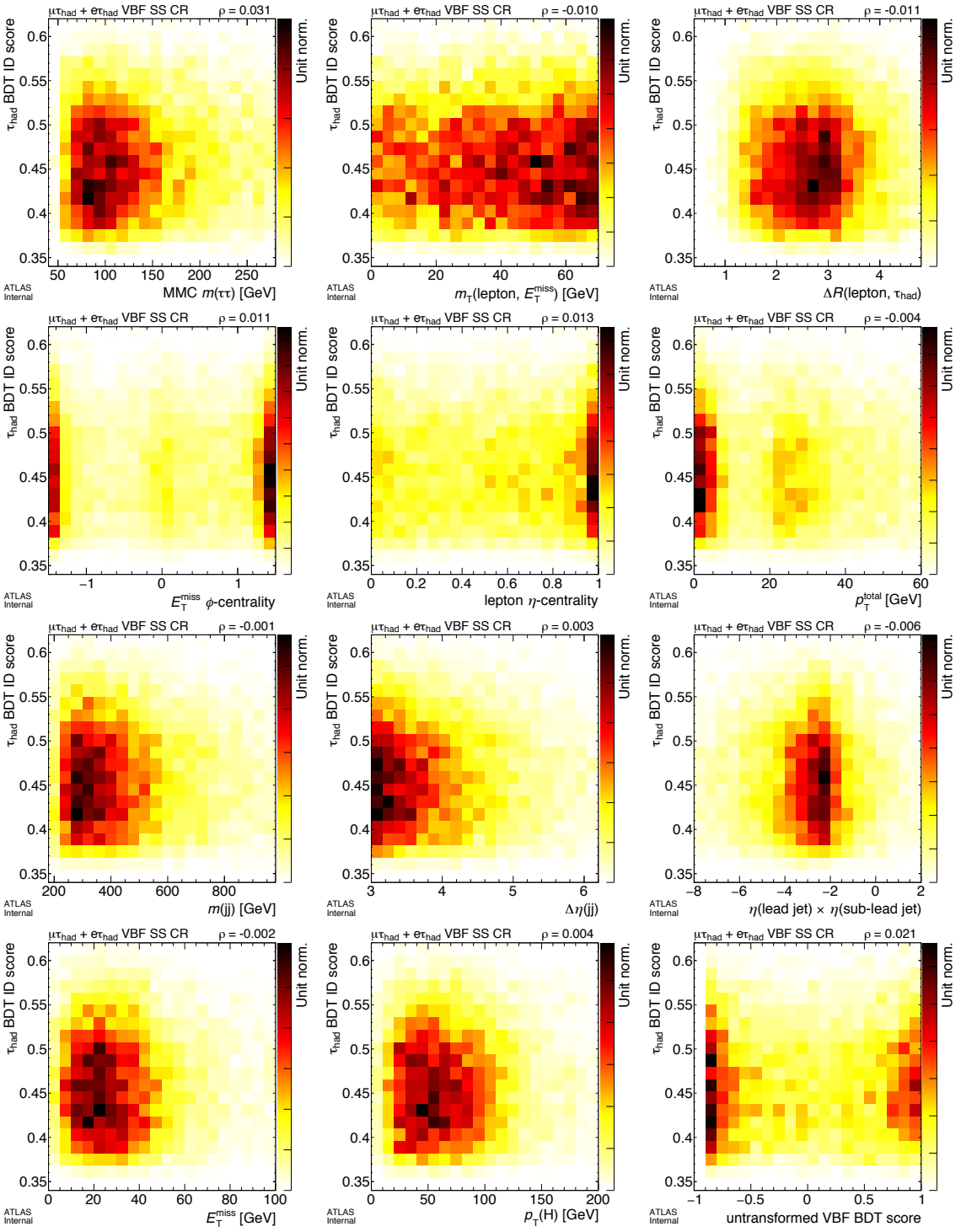

Figure 6.10: Correlations between the τ_{had} BDT identification score and event kinematics in data events in the VBF same-sign region which fail τ_{had} identification but fulfill all other requirements. No strong correlations are observed.

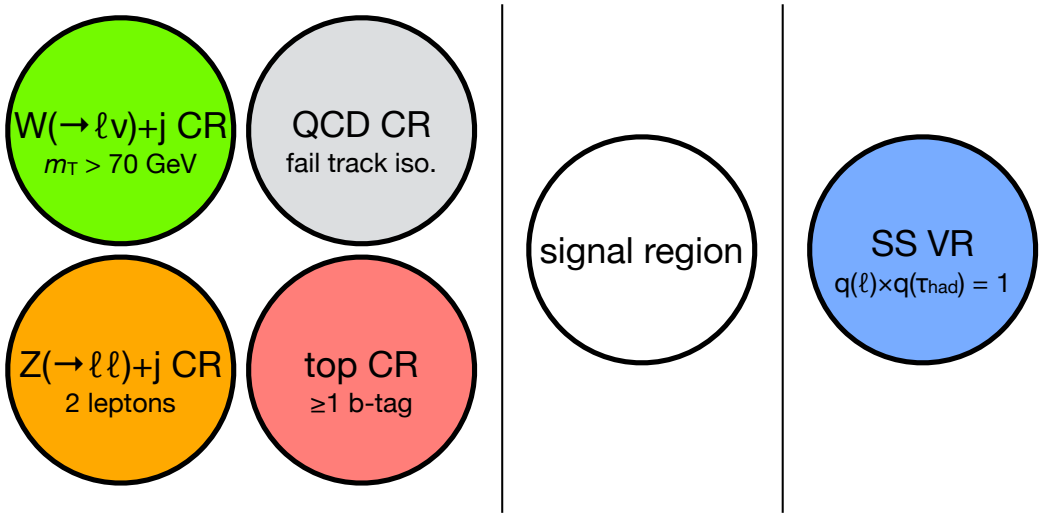

Figure 6.11: Cartoon of the signal, control, and validation regions used which are used in the $j \to \tau_{\text{had}}$ estimate.

6.2.2.2 Implementation

The $j \to \tau_{\text{had}}$ prediction is written as:

$$N_{j \to \tau_{\text{had}}} = \left(N_{\text{data}}^{\text{fail ID}} - N_{Z \to \tau\tau \text{ et al.}}^{\text{fail ID}}\right) \times \text{FF}_{\text{SR}} \tag{6.1}$$

where the number of predicted $j \to \tau_{\text{had}}$ events $N_{j \to \tau_{\text{had}}}$ is derived in each bin of any kinematic variable, like m_{jj}. The transfer factor needed to weight data events which fail τ_{had} identification is called the fakefactor (FF). It is parameterized in the number of tracks associated to the τ_{had} and $p_{\text{T}}(\tau_{\text{had}})$, and it is derived in a variety of regions rich in different $j \to \tau_{\text{had}}$ processes:

$$\text{FF}_{\text{region}} = \left.\frac{\left(N_{\text{data}}^{\text{pass ID}} - N_{Z \to \tau\tau \text{ et al.}}^{\text{pass ID}}\right)}{\left(N_{\text{data}}^{\text{fail ID}} - N_{Z \to \tau\tau \text{ et al.}}^{\text{fail ID}}\right)}\right|_{\text{region}} \tag{6.2}$$

where the regions considered here are rich in $j \to \tau_{\text{had}}$ from $W(\to \ell\nu_\ell)$+jets, QCD, $Z(\to \ell\ell)$+jets, or top events, or the same-sign region, which is a blend of $j \to \tau_{\text{had}}$ processes. These regions are shown pictorially in Fig. 6.11.

To protect against potential extrapolation biases, τ_{had} candidates failing identification criteria are required to pass a looser-than-`loose` requirement. This requirement is optimized to minimize the extrapolation without sacrificing the statistics of the estimate, and a requirement of `loose` $\times 0.7$ is chosen. For example, if the `loose` identification criteria requires the τ_{had} BDT score greater than

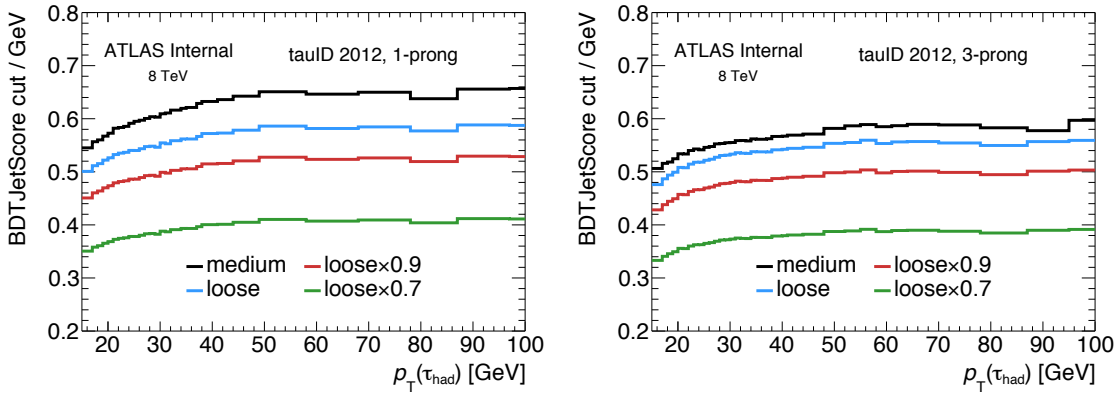

Figure 6.12: Requirements on the τ_{had} jet discriminant, which are defined to have constant signal efficiency as a function of $p_{\text{T}}(\tau_{\text{had}})$, of various operating points for 1-track τ_{had} (left) and 3-track τ_{had} (right).

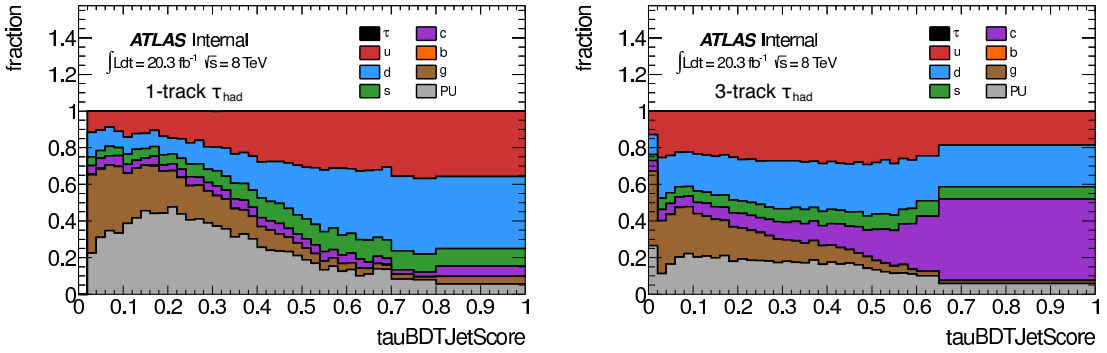

Figure 6.13: Predicted flavor composition of $j \to \tau_{\text{had}}$ in $W(\to \ell\nu_\ell)$+jets simulation for 1-track τ_{had} (left) and 3-track τ_{had} (right).

0.5, the `loose`×0.7 identification criteria requires greater than 0.35. The $p_{\text{T}}(\tau_{\text{had}})$ dependence of this requirement is shown in Fig. 6.12.

The impact of requiring `loose`×0.7 can be seen by considering the response of various flavors of $j \to \tau_{\text{had}}$ to the tau identification BDT, as shown in $W(\to \ell\nu_\ell)$+jets simulation in Fig. 6.13. This requirement reduces the pileup and gluon content of the anti-identified region and gives it a closer flavor resemblance to the identified region.

The fakefactors measured in data in the $W(\to \ell\nu_\ell)$+jets, QCD, $Z(\to \ell\ell)$+jets, and top control regions are shown in Fig. 6.14 for 1-track and 3-track τ_{had} in the VBF category. The measured fakefactors do not show systematic differences between regions given the statistical uncertainty.

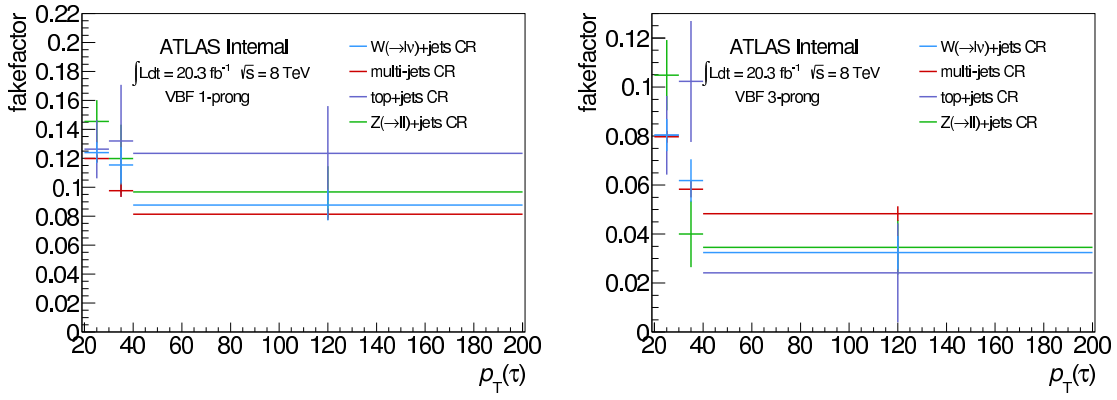

Figure 6.14: Fake factors in the VBF category measured in the various control regions in data for 1-track τ_{had} (left) and 3-track τ_{had} (right).

6.2.2.3 Composition of $j \to \tau_{\text{had}}$ in the SR

A fakefactor for the signal region can be derived from fakefactors measured in the control regions by using simulation to predict the relative contributions of the $W(\to \ell\nu_\ell)$+jets, top, and $Z(\to \ell\ell)$+jets processes in the anti-identified region. The remaining difference between data and prediction is then assumed to be from QCD.

The overall relative contributions are shown in Fig. 6.15, and the differential contributions are shown in Figs. 6.16 and 6.17. No strong dependence on the final VBF BDT discriminator is observed, though dependencies are observed on distributions like p_{T}(lepton) and $E_{\text{T}}^{\text{miss}}$.

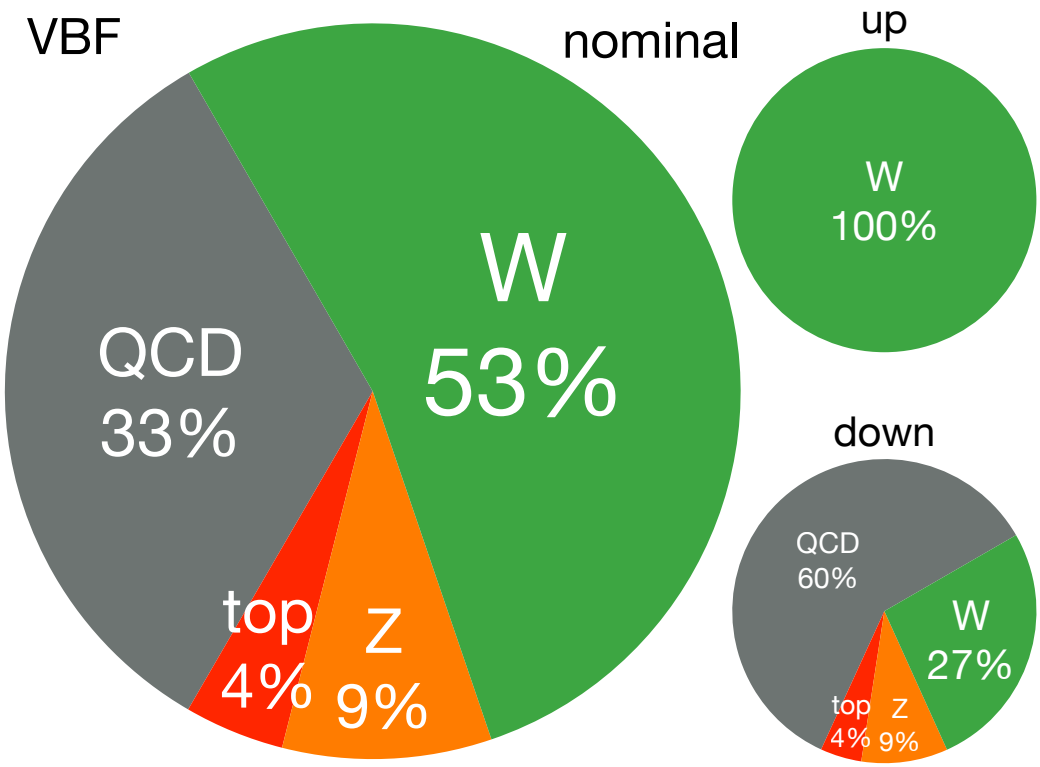

Figure 6.15: A pie chart of the composition of $j \to \tau_{\text{had}}$ processes in the anti-identified CR as predicted by simulation and data (left) and the systematic variations on the composition (right).

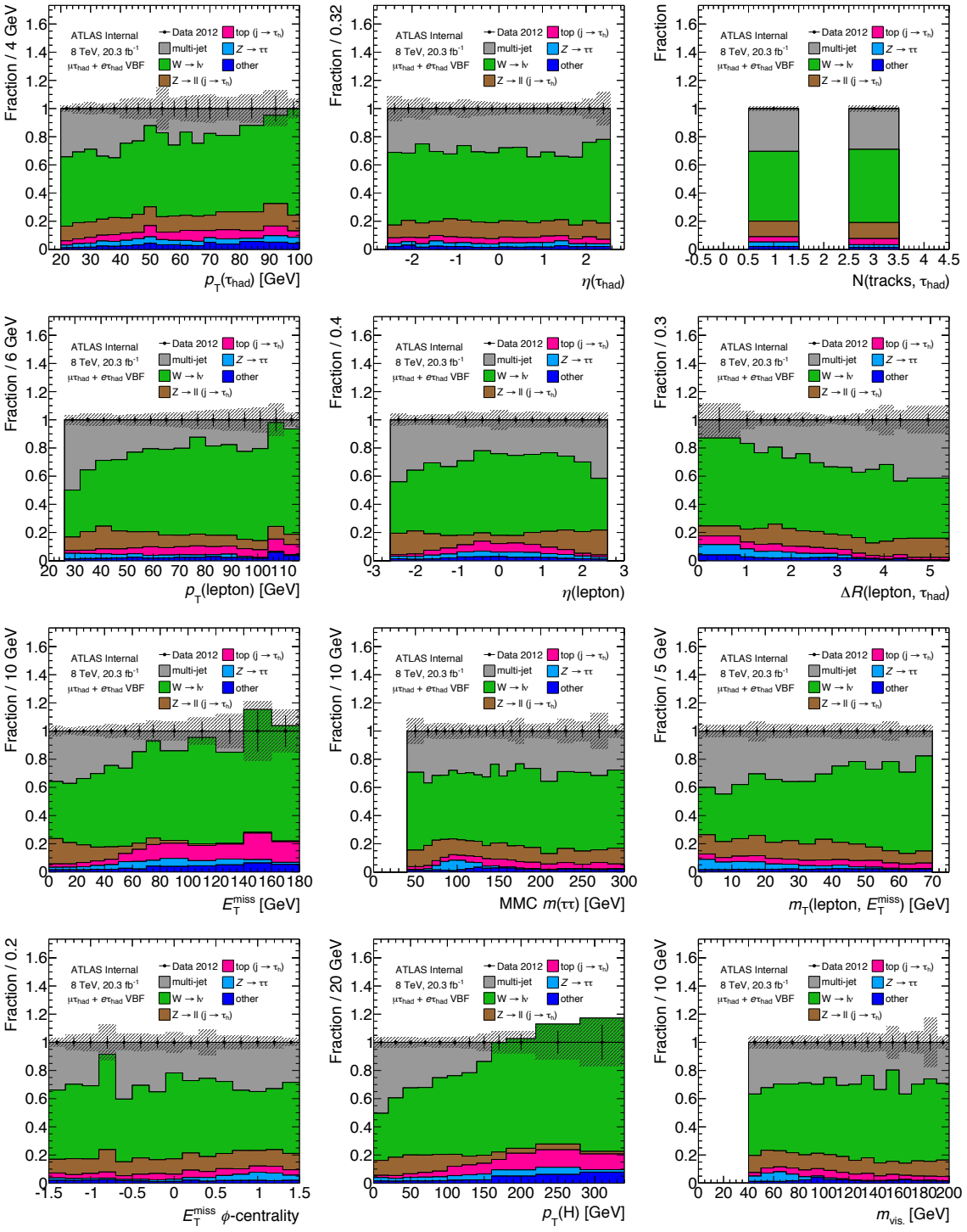

Figure 6.16: The composition of $j \to \tau_{\text{had}}$ processes in the anti-identified CR as predicted by simulation and data as a function of event kinematics.

The signal region fakefactors are then derived as a linear combination of control region fakefactors weighted by the expected contributions. The systematic uncertainties on these contributions is shown in Fig. 6.15. A conservative approach is taken due to the mistrust of simulated $j \to \tau_{\text{had}}$, and the contribution from $W(\to \ell\nu_\ell)$+jets is allowed to double or halve as the two variations.

The signal region fakefactors are shown in Fig. 6.18. These are also referred to as mixed fakefactors. Since the control region fakefactors do not have significant differences between them, the dominant uncertainty on the signal region fakefactors is typically statistical.

6.2.3 Validation

The fakefactor method is validated by checking predictions of event-level kinematics, especially the BDT discriminant, in the control and validation regions. It is additionally validated by following the same fakefactor procedure in $W(\to \ell\nu_\ell)$+jets, top, and $Z(\to \ell\ell)$+jets simulated events in the signal region, where dedicated $\text{FF}_{\text{SR}}^{\text{MC}}$ are derived.

Data and prediction in the same-sign validation region are shown in Fig. 6.19. Data and prediction in the various control regions are further shown in Appendix A. Predictions with simulation are shown in Fig. 6.20. In all plots, good agreement is observed and no systematic biases are uncovered.

6.2.4 Uncertainties

Multiple sources of uncertainties to the fake factor method are considered. First, the statistical uncertainty on the fake factors measured in control regions is propagated to the uncertainty on the signal region fakefactor. Second, the uncertainty on the relative contributions of the different $j \to \tau_{\text{had}}$ processes are varied and propagated to the signal region fakefactor calculation. Third, the fidelity of using control region fakefactors in the signal region is tested by comparing fakefactors measured in simulation in control regions versus the signal region. No significant difference is found, and the uncertainty is ignored. Fourth, the closure of the method is tested with predictions in the same sign validation region and in the signal region in simulation. No signs of systematic bias in the BDT score prediction are found, and the closure uncertainties are ignored.

The pre-fit impact of these uncertainties on the $j \to \tau_{\text{had}}$ prediction is shown in each bin of the VBF discriminator in Fig. 6.21. The largest uncertainty at high VBF BDT score is the 15% statistical uncertainty on the prediction. This could be ameliorated by relaxing the `loose`$\times 0.7$ requirement on anti-identified τ_{had}, though this would risk introducing systematic bias of increasing the extrapolation. The largest systematic uncertainty is on the relative contribution of $j \to \tau_{\text{had}}$ processes, and this only propagates to a 3% uncertainty on the prediction.

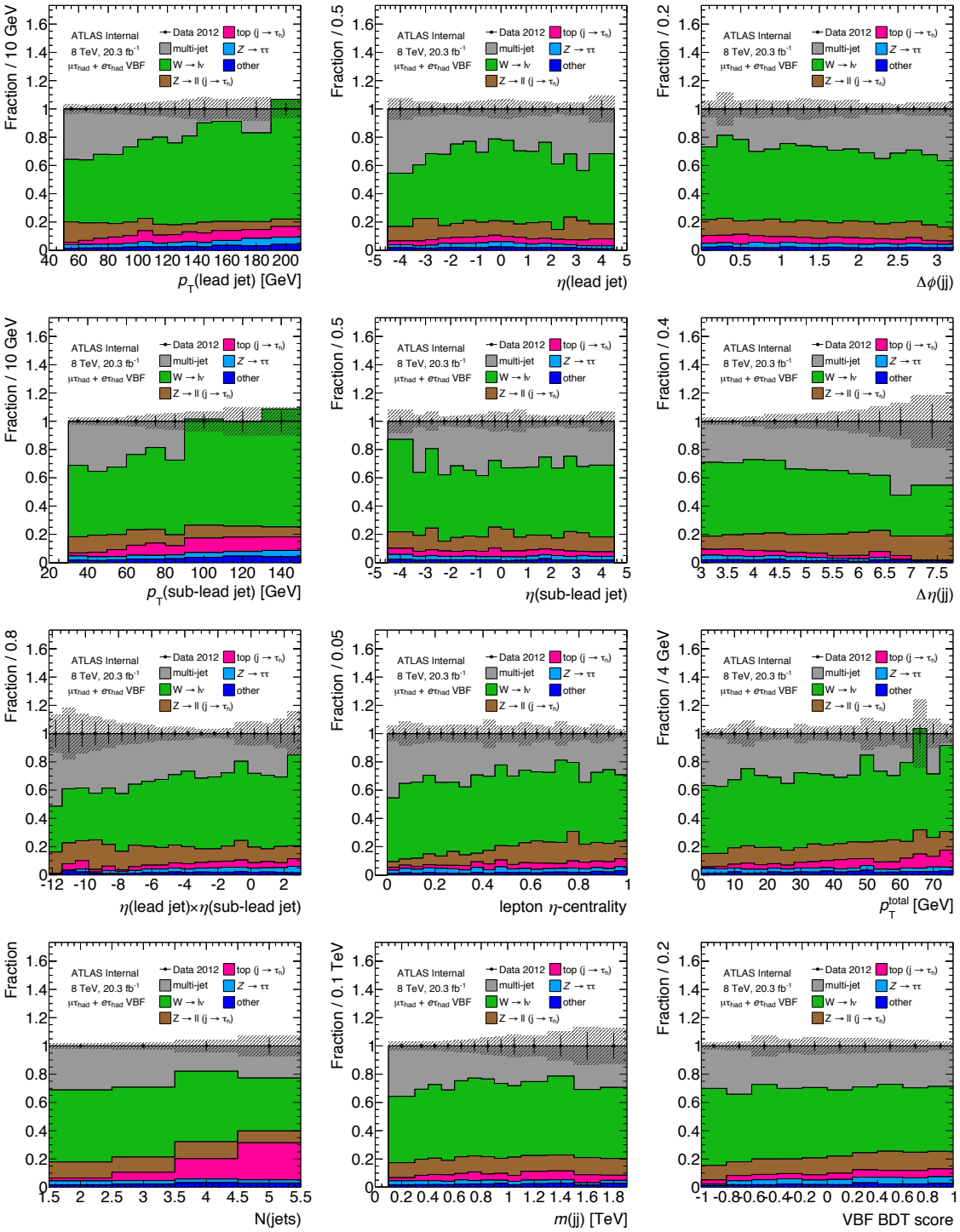

Figure 6.17: The composition of $j \to \tau_{\text{had}}$ processes in the anti-identified CR as predicted by simulation and data as a function of event kinematics.

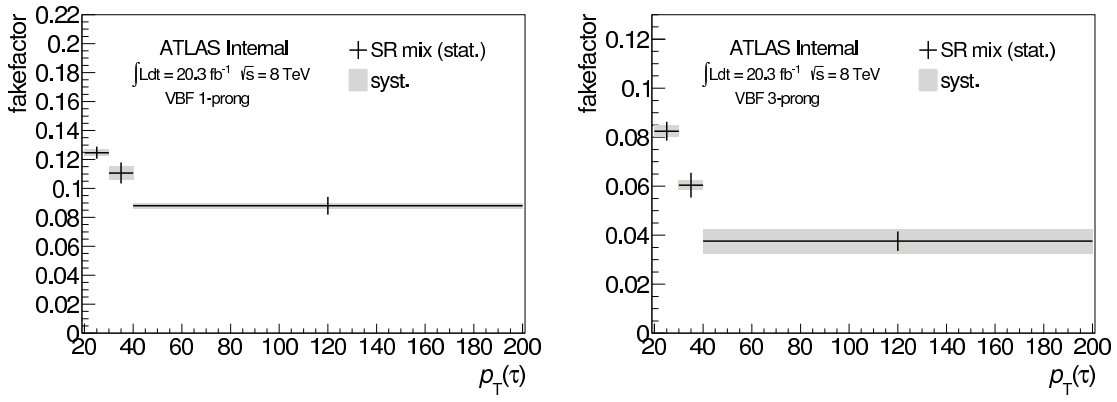

Figure 6.18: Fake factors in the VBF category mixed from the various control regions in data for 1-track τ_{had} (left) and 3-track τ_{had} (right). Statistical and systematic uncertainties are shown.

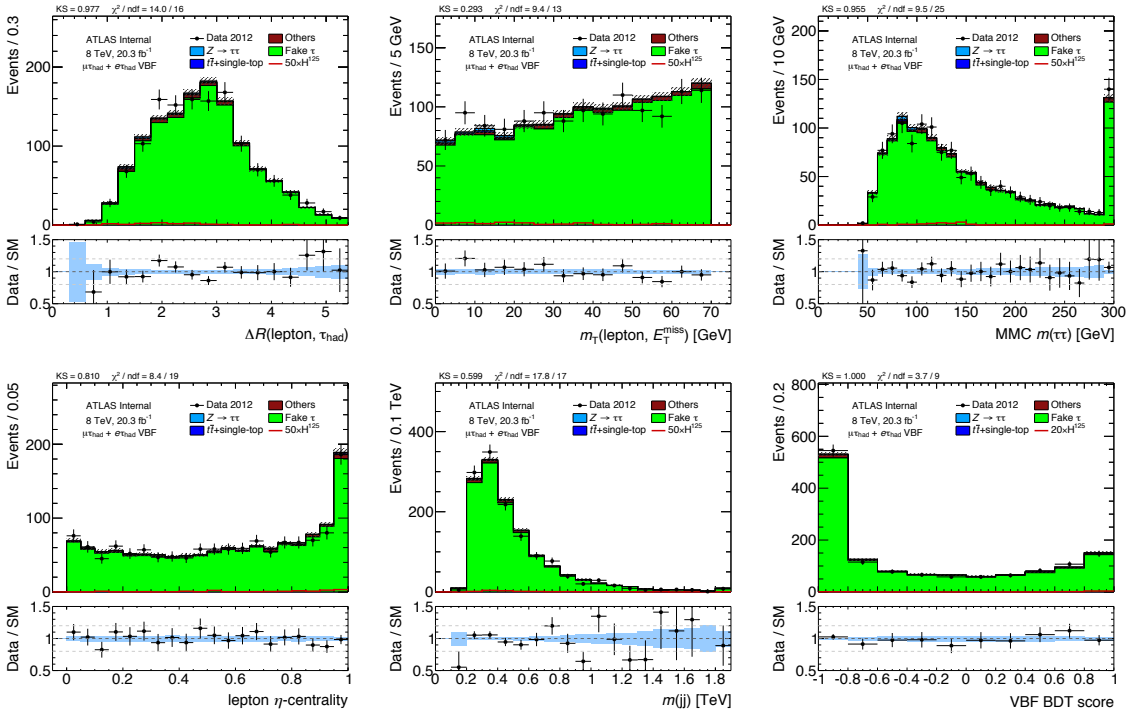

Figure 6.19: Comparison of data and $j \to \tau_{\text{had}}$ prediction in the same-sign validation region for various event kinematics. The purity of $j \to \tau_{\text{had}}$ is $\approx 97\%$. Only statistical uncertainties are shown, and no sign of systematic bias is observed. Additional validation is shown in Appendix A.

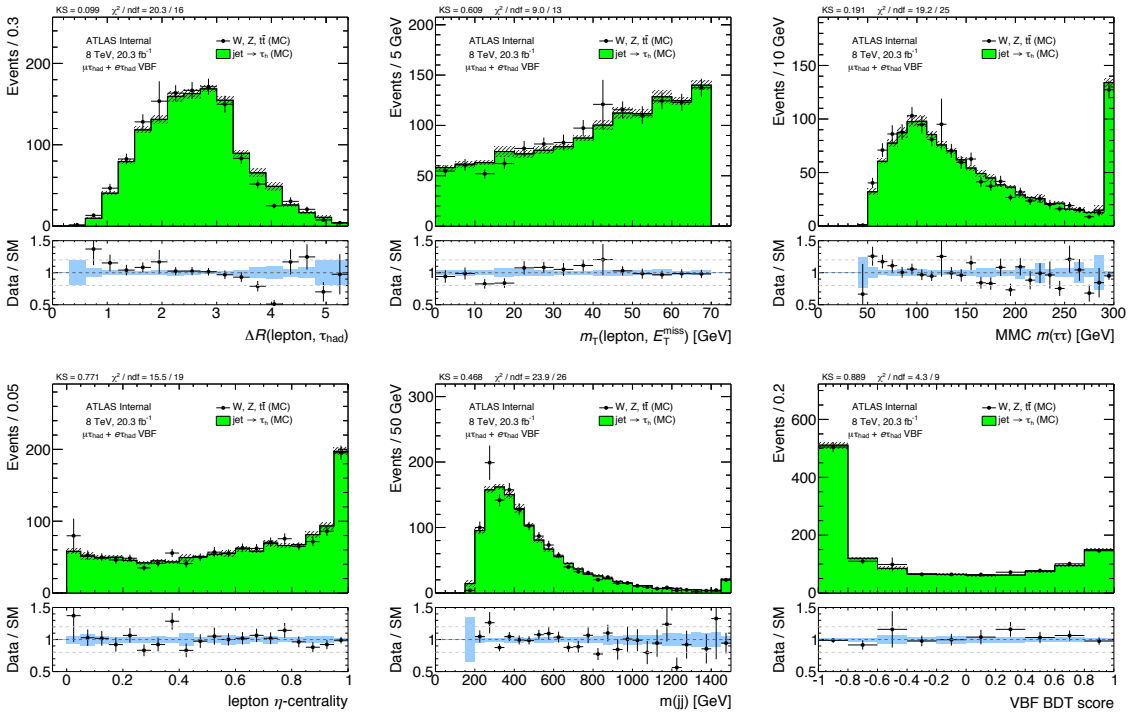

Figure 6.20: Comparison of the prediction of identified taus and the $j \to \tau_{\text{had}}$ prediction, both in simulation, in the signal region for various event kinematics. Only statistical uncertainties are shown, and no sign of systematic bias is observed. Additional validation is shown in Appendix A.

6.3 top, $Z \to \ell\ell$, diboson

6.3.1 top

Top events with a true τ_{had} or $\ell \to \tau_{\text{had}}$ are estimated with simulation and object-level corrections prescribed by the τ_{had} performance group. These include $t\bar{t}$ and single top processes. The normalization is constrained using a top-enriched control region, but detailed corrections to the simulation are not sought because the background is sub-dominant. These top processes only comprise 5% of the background prediction in the VBF category and in the most sensitive bin of the VBF BDT discriminator.

6.3.2 $Z \to \ell\ell$ ($\ell \to \tau_{\text{had}}$), diboson

$Z \to \ell\ell$ events where a lepton is mis-identified as a τ_{had} and diboson events (WW, WZ, ZZ) are estimated with simulation and object-level corrections prescribed by the τ_{had} performance group. Control

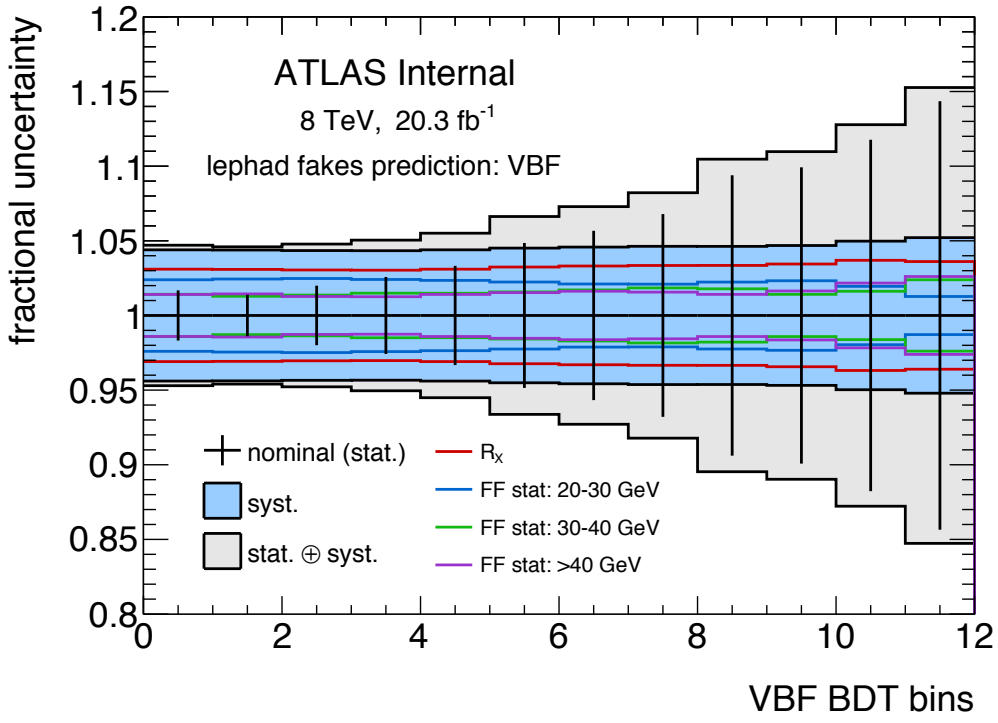

Figure 6.21: The pre-fit fractional uncertainty on the $j \to \tau_{\text{had}}$ prediction in each bin of the VBF category. R_X refers to the uncertainty on the relative contribution of $j \to \tau_{\text{had}}$ processes.

regions for these processes are not sought because the processes are too small to find meaningful regions for comparing data with simulation. Detailed corrections to the simulation are also not sought because each processes comprises less than 5% of the background prediction in the VBF category and in the most sensitive bin of the VBF BDT discriminator.

The τ_{had} electron discriminator reduces the $Z \to ee$ background from problematic to negligible, as shown in Fig. 6.22. Without the electron discriminator, the VBF $Z \to ee$ background would be difficult to distinguish from VBF $H \to \tau\tau$ and potentially of comparable magnitude.

6.4 $H \to \tau\tau$

6.4.1 Samples

The signal $H \to \tau\tau$ processes are simulated with Powheg+Pythia (ggFH, VBFH) and Pythia (WH, ZH, ttH) [2], though the VH and ttH processes are generally negligible in the signal regions considered. The overall normalisation of the ggF process is taken from a calculation at next-to-next-

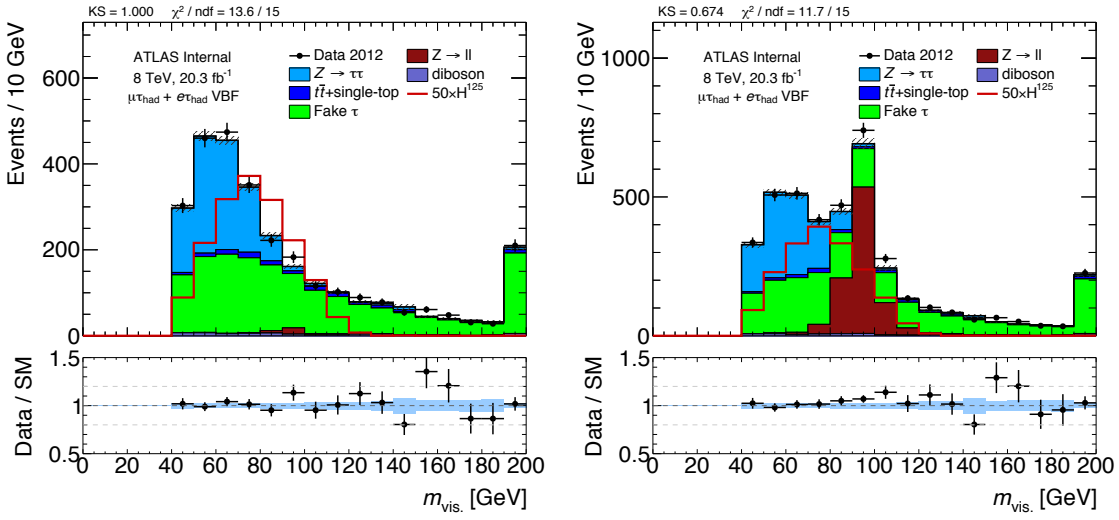

Figure 6.22: Data and prediction for the nominal VBF category (left) and without the τ_{had} electron discriminator (right).

to-leading-order (NNLO) in QCD, including soft-gluon resummation up to next-to-next-to-leading logarithm terms (NNLL). Next-to-leading order (NLO) EW corrections are also included. Production by VBF is normalized to a cross section calculated with NLO QCD and EW corrections with an approximate NNLO QCD correction applied.

Additional corrections to the shape of the generated p_{T} distribution of Higgs bosons produced via ggF are applied to match the distribution from a calculation at NNLO including the NNLL corrections provided by the HRes2.1 program [2]. In this calculation, the effects of finite masses of the top and bottom quarks are included and dynamical renormalisation and factorisation scales, $\mu_{\text{R}}, \mu_{\text{F}} = \sqrt{m_H^2 + p_{\text{T}}^2}$, are used. A reweighting is performed separately for events with no more than one jet at particle level and for events with two or more jets. In the latter case, the Higgs boson p_{T} spectrum is reweighted to match the MinLo HJJ predictions. The reweighting is derived such that the inclusive Higgs boson p_{T} spectrum and the p_{T} spectrum of events with at least two jets match the HRes2.1 and MinLo HJJ predictions respectively, and that the jet multiplicities are in agreement with (N)NLO calculations from JetVHeto. A similar p_{T}-dependent weighting is derived for NLO EW corrections of the VBFH production using Hawk, though the corrections are small in the p_{T} ranges considered here [2].

6.4.2 Uncertainties

Uncertainties regarding the detector response of all physics objects is considered for the signal $H \to \tau\tau$. This is implemented via the typical collection of uncertainties pertaining to the measured identification efficiency and energy calibration of simulated leptons, τ_{had}, jets, and $E_{\text{T}}^{\text{miss}}$ at ATLAS. The uncertainties on the VBF Higgs production kinematics are generally smaller than the experimental uncertainties.

The pre-fit impact of these uncertainties on the VBF $H \to \tau\tau$ prediction is shown in each bin of the VBF discriminator in Fig. 6.23. The largest uncertainty at high VBF BDT score is the jet energy scale (JES) uncertainty including uncertainties in the forward region, which are large relative to JES uncertainties within the tracker. JES uncertainties are also the largest class of uncertainties, though no single component propagates to an uncertainty larger than 10% on the VBF $H \to \tau\tau$ prediction.

6.5 Predictions in the signal region

Predictions in the signal region of the VBF $H \to \tau_\ell\tau_{\text{had}}$ analysis are shown in Fig. 6.24. These input variables feed into the BDT discriminator from which the signal is extracted, which is discussed in Chapter 7. The largest background is from $j \to \tau_{\text{had}}$, the second largest background is from $Z \to \tau\tau$, and the remaining backgrounds are individually less than 5% of the total background prediction.

Good agreement between data and prediction is observed for all input variables. This agreement is evaluated with visual inspection and with quantitative measures like χ^2/N(D.O.F) and the Kolmogorov-Smirnov test.

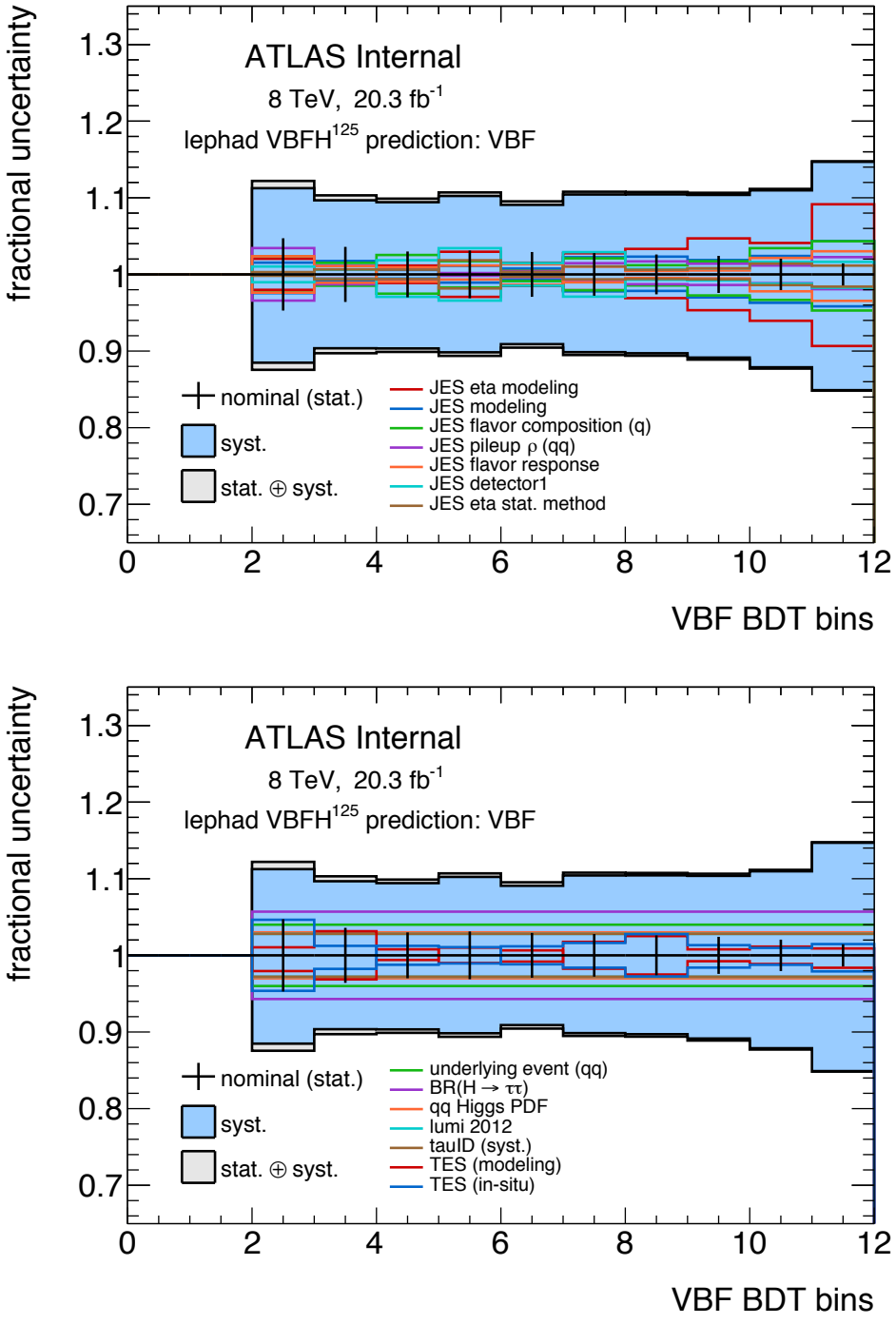

Figure 6.23: The pre-fit fractional uncertainty on the VBF $H \to \tau_\ell \tau_{\text{had}}$ prediction in each bin of the VBF category for uncertainties pertaining to the jet energy scale (top) and τ_{had} performance, theory, and the luminosity (bottom).

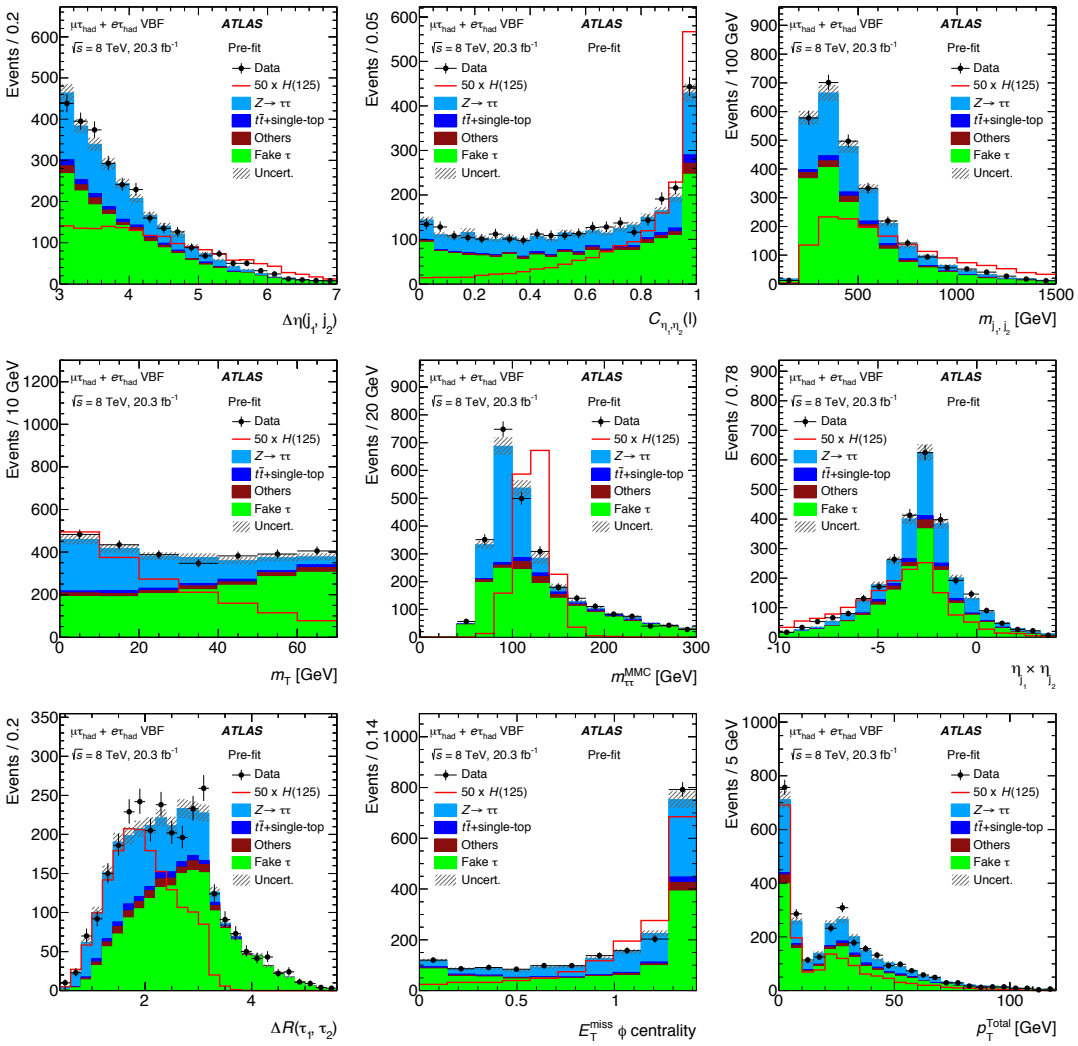

Figure 6.24: Data and prediction for input variables to the BDT in the $H \to \tau_\ell \tau_{\text{had}}$ VBF signal region [2].

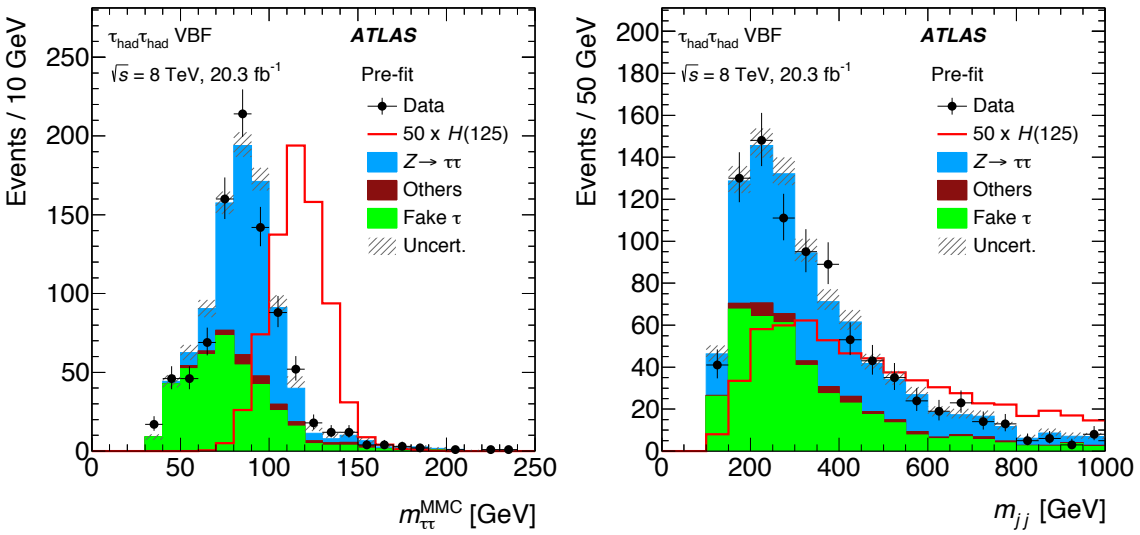

Figure 6.25: Two of the nine input variables to the VBF $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$ BDT discriminator: $m_{\tau\tau}^{\text{MMC}}$ (left) and m_{jj} (right). [2].

6.6 $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$ and $H \rightarrow \tau_\ell\tau_\ell$

Selected predictions in the signal regions of the VBF $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$ and $H \rightarrow \tau_\ell\tau_\ell$ analyses are shown in Figs. 6.25 and 6.26. Good agreement between data and prediction is observed for the input variables.

The background predictions are conceptually similar to the predictions in the $H \rightarrow \tau_\ell\tau_{\text{had}}$ analysis: $Z \rightarrow \tau\tau$ is predicted with the embedding, mis-identified backgrounds ($j \rightarrow \tau_{\text{had}}$, $j \rightarrow \ell$) are predicted with regions of data topologically similar to the signal region but with object-level identification criteria reversed, and the remaining backgrounds are predicted with simulation. The background predictions in these are discussed in greater detail in the accompanying publication [2].

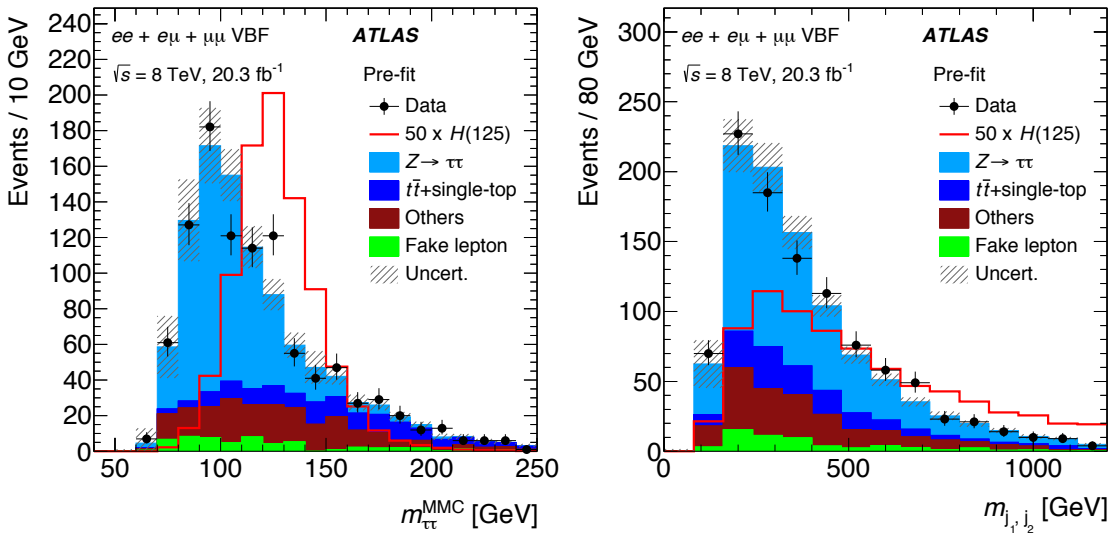

Figure 6.26: Two of the seven input variables to the VBF $H \to \tau_\ell\tau_\ell$ BDT discriminator: $m_{\tau\tau}^{\mathrm{MMC}}$ (left) and m_{jj} (right). [2].