Chapter 1

Background Estimation

Maybe some stuff here introducing the chapter...

1.1 Overview of TF method

Contributions from Standard Model background processes are estimated using a data-driven prediction technique, employing dedicated control samples. A transfer factor (TF) is constructed from MC samples as the ratio of the MC yield in the signal region, $N_{MC}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big)$, and the MC yield of a given control region, $N_{Data}^{control}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big)$, as a function of the analysis binning, H_T, n_{jet} and n_{b}, shown in equation 1.1.

$$TF = \frac{N_{MC}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big)}{N_{MC}^{control}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big)} \tag{1.1}$$

For a given H_T, n_{jet} and n_{b} bin, the TF is used to extrapolate a yield in data from the control region, $N_{Data}^{control}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big)$ to the signal region $N_{Data}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big)$, as shown in equation 1.2.

$$N_{Data}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big) = N_{Data}^{control}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big) \times TF \tag{1.2}$$

The control samples are statistically independent and each used for predicting specific background processes, the details of which are described in section 1.2.

MC yields are construct from the following process-specific samples: W + jets (N_{W}), $\text{t}\bar{\text{t}}$ + jets ($N_{\text{t}\bar{\text{t}}}$), DY + jets ($N_{\text{DY}}$), γ + jets(N_{γ}), single top + jets (N_{top}), WW + jets, WZ + jets and ZZ + jets ($N_{\text{di-boson}}$), and $Z \to \nu\bar{\nu}$ + jets ($N_{Z\to\nu\bar{\nu}}$).

Predictions made using this technique alone are considered as "naïve" predictions and are used only in analysis development and the derivation of background systematics. In order to determine final yields for interpretation and limit-setting, a fit is made across all signal and control regions, using the full likelihood model, described later in chapter **??**. The derived transfer factors and individual yields enter as terms in the likelihood, where all related systematics, potential correlations and signal contamination are accounted for in the fit.

The denominator of each transfer factor is constructed using the sum of *all* MC sample yields, for a given control region se;ection and analysis category:

$$N_{MC}^{control}\Big(H_T, n_{\text{jet}}, n_{\text{b}}\Big) = N_{\text{W}} + N_{\text{t}\bar{\text{t}}} + N_{\text{DY}} + N_{\gamma} + N_{\text{top}} + N_{\text{di-boson}} + N_{Z\to\nu\bar{\nu}} \quad (1.3)$$

The numerator is constructed according to the b-tag mutliplicity being considered. For $n_{\text{b}} \leq 1$, the μ + jets control region is used to predict lost-lepton background, e.g. $\text{t}\bar{\text{t}}$ + jets and W + jets. All MC samples are used with the exception of $Z \to \nu\bar{\nu}$:

$$N_{MC}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}} \leq 1\Big) = N_{\text{W}} + N_{\text{t}\bar{\text{t}}} + N_{\text{DY}} + N_{\gamma} + N_{\text{top}} + N_{\text{di-boson}} \quad (1.4)$$

The $Z \to \nu\bar{\nu}$ + jets component of the background is predicted using the $\mu\mu$ + jets and γ + jets control samples, using only the $Z \to \nu\bar{\nu}$ MC yields:

$$N_{MC}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}} \leq 1\Big) = N_{Z\to\nu\bar{\nu}} \quad (1.5)$$

Again, it should be noted here that although two seperate control samples are used to estimate the $Z \to \nu\bar{\nu}$ background contribution, the result of each is considered by the global fit to produce the final background prediction.

For $n_{\text{b}} \geq 2$, the μ + jets sample is used to produce a prediction for all SM processes, including $Z \to \nu\bar{\nu}$, and therefore the numerator of the TF is defined as:

$$N_{MC}^{signal}\Big(H_T, n_{\text{jet}}, n_{\text{b}} \geq 2\Big) = N_{\text{W}} + N_{\text{t}\bar{\text{t}}} + N_{\text{DY}} + N_{\gamma} + N_{\text{top}} + N_{\text{di-boson}} + N_{Z\to\nu\bar{\nu}} \quad (1.6)$$

The $\mu\mu$ + jets and γ + jets control samples are not used beyond the $n_{\text{b}} \leq 1$ categories due to the statistical limitations of such samples at high b-tag multiplicities. A full summary of the control regions used for predictions per analysis category is

shown in table 1.1.

Table 1.1 Summary of control samples used to predict the SM background for each event category. REWORD

n_{jet}	n_{b}	Control samples
2–3	0	μ + jets, $\mu\mu$ + jets, γ + jets
2–3	1	μ + jets, $\mu\mu$ + jets, γ + jets
2–3	2	μ + jets
≥4	0	μ + jets, $\mu\mu$ + jets, γ + jets
≥4	1	μ + jets, $\mu\mu$ + jets, γ + jets
≥4	2	μ + jets
≥4	3	μ + jets
≥4	≥ 4	μ + jets

By employing a technique that uses a ratio of MC yields, direct dependence on MC modelling is greatly reduced. Errors inherent to MC samples, such as mismodelling effects, will cancel in the ratio. These errors can potentially include kinematic mismodelling, which would affect analysis acceptance, and mismodelling of instrumental effects, which could have an affect on object reconstruction efficiencies. However, any remaining systematics such as these and others are probed using dedicated Closure Tests (CT), described in detail in section 1.5.1.

1.2 Control samples

Control sample definitions are designed such that they are as close as possible to that of the signal region, with the exception of a selected 'tag' muon or photon, which is subsequently ignored for the calculation of all analysis variables, such as H_T, $\not{E}_T$, α_{T} etc. Other differences include mass-window and minor kinematic cuts, aimed at enriching the control samples in certain processes. The samples themselves are statistically independent, and orthogonal to the signal region due to the selection of the tagged lepton or photon minimising any possible signal contamination. However, a full treatment of the signal-contamination and sample cross-correlation is taken into account in the background fit and final limit-setting.

1.2.1 mu + jets

The μ + jets control sample is constructed by selecting a single muon with associated jets. This region is used to predict backgrounds from processes such as W + jets and

$t\bar{t}$ + jets. This covers not only the leptonic decays of such productions, where the lepton is unidentified for whatever reason, but also the hadronic decays of tau leptons [from high-p_{T} W bosons]. The event selection therefore is optimised to select W bosons decaying to a muon and a neutrino in the phase-space of the signal.

Triggers

Events are collected using the loosely-isolated, η-restricted `HLT_IsoMu24_eta2p1` trigger, which was in place throughout the $8TeV$ data-taking period. The efficiency of this trigger was measured by the muon POG using a tag-and-probe method [REF], in bins of the muon p_{T} and η, as summarised in table 1.2. Statistical undertainties are at the per- mille level, and systematics are assumed to be 1%.

Table 1.2 Muon trigger efficiencies (%) for the μ + jets selection listed by H_T bin and n_{jet} category.

H_T (GeV)	2-3	≥4
150–200	87.2	88.1
200–275	87.5	88.1
275–325	87.8	88.2
325–375	87.9	88.4
375–475	88.1	88.6
475–575	88.2	88.8
575–675	88.4	88.9
675–775	88.5	89.0
775–875	88.6	89.1
875–975	88.8	89.0
975–1075	88.7	89.0
>1075	88.4	89.6

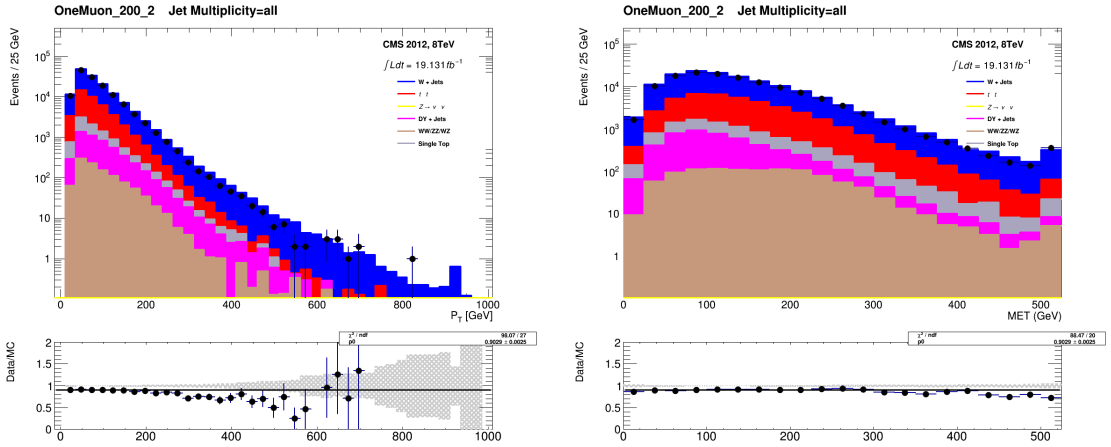
Selection Criteria

A single tight isolated muon is selected, with $p_{\mathrm{T}} > 30$GeV and $|\eta| < 2.1$. The transverse mass of the W, reconstructed by the μ and the $\not{E}_T$ (originating from the ν_μ), is required to be in a loose window around m_W, $30 < M(\mu, \not{E}_T) < 125GeV$. Events are vetoed if $\Delta R(\mu, jet_i) < 0.5$, for all jets i in the event. To keep the selection as close to the signal region as possible, other cuts such as the single isolated track veto and $\not{H}_T/\not{E}_T$ cuts are also applied.

Specifically for the μ + jets (and $\mu\mu$ + jets) control samples, no α_{T} requirement is made in order to increase the statistics and therefore the predictive power of the

samples. This is possible as other requirements, in particular the requirement of a single muon and a specific invariant mass window, greatly reduce any potential contamination from QCD MJ events. The viability of this is specifically tested by a dedicated closure test, described later in section 1.5.1.

Example distributions for this selection are shown in figure 1.1.



(a) μp_{T} (b) $\not{E}_T$ (corrected for μ)

Fig. 1.1 Comparison of data with MC for the muon control selection. Plots are for $H_T > 200GeV, n_{\mathrm{jet}} \geq 2, n_{\mathrm{b}} \geq 0$.

1.2.2 mumu + jets

The $\mu\mu$ + jets sample is constructed to predict background contributions from $Z \to \nu\bar{\nu}$ decays, mimicking this decay via the kinematically similar $Z \to \mu\mu + jets$ process where both muons are subsequently ignored. The sample is used to provide low H_T coverage for the $Z \to \nu\bar{\nu}$ background prediction, where the γ + jets sample is unable to do so.

Triggers

The trigger used is the same as for the μ + jets sample, as described in section 1.2.1. Trigger efficiencies are slightly improved for the dimuon selection given that either of the muons can cause a positive trigger decision. The trigger efficiencies for the $\mu\mu$ + jets selection are summarised in table 1.3. Relative errors are considered the same as for the μ + jets trigger efficiencies.

Table 1.3 Muon trigger efficiencies (%) for the $\mu\mu$ + jets selection listed by H_T bin and n_{jet} category.

H_T (GeV)	2-3	≥4
150–200	98.4	98.4
200–275	98.5	98.4
275–325	98.5	98.4
325–375	98.6	98.6
375–475	98.6	98.5
475–575	98.6	98.6
575–675	98.6	98.6
675–775	98.7	98.6
775–875	98.6	98.6
875–975	98.7	98.6
975–1075	98.7	98.8
>1075	98.7	98.7

Selection Criteria

The selection for the $\mu\mu$ + jets sample is very similar to that of the μ + jets sample, described in section 1.2.1, with differences chosen to enrich the sample in Z bosons decaying to pairs of muons in the kinematic phase space of the signal region. Two tight isolated muons are selected, each with $p_{\text{T}} > 30GeV$ and $|\eta| < 2.1$. Their invariant mass is chosen to be tight around m_Z, $m_Z - 25 < M_{\mu_1\mu_2} < m_Z + 25GeV$. Furthermore, a veto is made on events satisfying $\Delta R(\mu_i, jet_j) < 0.5$, for every muon i and every jet j. Similarly as in the μ + jets sample selection, no α_{T} requirement is made.

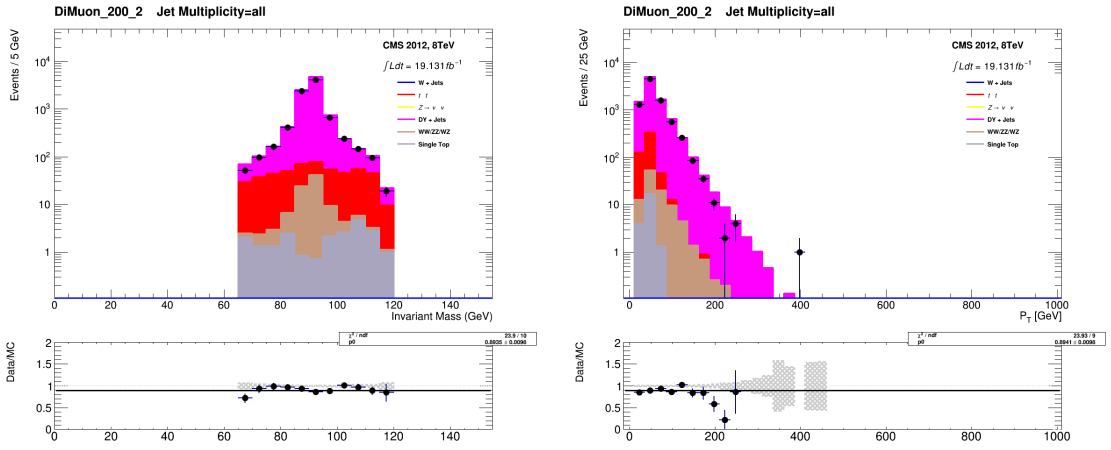
Example distributions for this selection are shown in figure 1.2.

1.2.3 photon + jets

The γ + jets sample is used to predict the $Z \to \nu\bar{\nu}$ background contribution, given it's similar kinematics when the γ is ignored from the event, and larger production cross section relative to $\mu\mu$ + jets. Due to trigger thresholds, the γ + jets sample cannot make predictions for $H_T < 375GeV$, and so is complimentarily used alongside the $\mu\mu$ + jets sample prediction.

Triggers

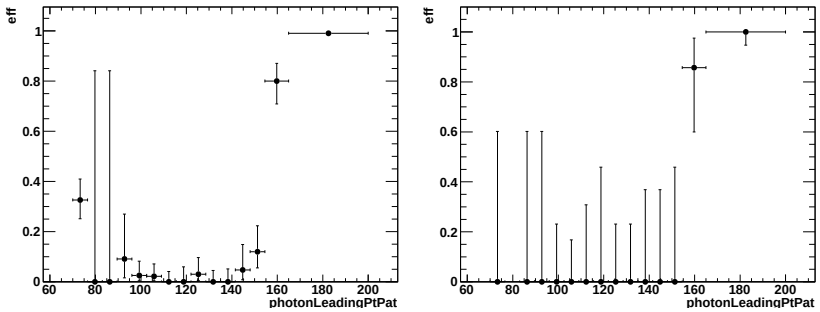
Events are collected using the `HLT_Photon150` trigger. The trigger's efficiency is measured using the `HLT_Photon90` trigger as a reference and is found to be 100% efficient



(a) $M_T(\mu,\mu)$ (b) Second μ p_{T}

Fig. 1.2 Comparison of data with MC for the dimuon control selection. Plots are for $H_T > 200GeV, n_{\mathrm{jet}} \geq 2, n_{\mathrm{b}} \geq 0$.

for $E_T^{photon} > 165GeV$ and $H_T > 375GeV$, as shown by the turn on curves in figure 1.3.



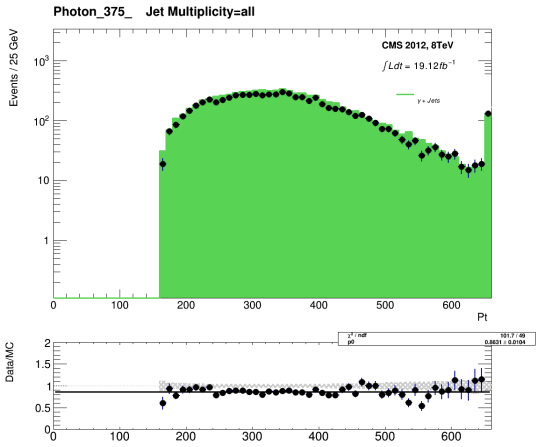
(a) $2 \leq n_{\mathrm{jet}} \leq 3$ (b) $n_{\mathrm{jet}} \geq 4$

Fig. 1.3 Efficiency turn-on curves for the photon trigger that records events that satisfy the γ + jets selection criteria, $E_T^{photon} > 165GeV$, $H_T > 375GeV$, and $2 \leq n_{\mathrm{jet}} \leq 3$ (Left) and $n_{\mathrm{jet}} \geq 4$ (Right). R

Selection Criteria

Exactly one photon satisying tight isolation criteria is required, with $p_{\mathrm{T}} > 165GeV$ and $|\eta| < 1.45$. In addition, events are vetoed if $\Delta R(\gamma, jet_i) < 1.0$ is satisfied, for all of the events i jets.

An example distribution for this selection is shown in figure 1.4.



(a) γp_{T}

Fig. 1.4 Comparison of data with MC for the photon control selection. Plots are for $H_T > 200GeV, n_{\mathrm{jet}} \geq 2, n_{\mathrm{b}} \geq 0$.

1.2.4 Multijet

In order to predict the contamination of QCD MJ events in the signal region a control region of hadronic events is constructed.

Triggers

Events are collected using a suite of triggers requiring various thresholds of H_T. The high rates expected through these triggers were maintained throughout the 2012 run with a variety of prescales. Each trigger seeds a single H_T bin in the analysis, with a 25GeV offset between the online trigger requirement and the offline threshold. Efficiencies are measured with the same technique as the signal triggers, using the `HLT_IsoMu24_eta2p1` trigger as a reference. The efficiencies of these triggers are summarised in table 1.4.

Selection Criteria

The selection of this sample matches that of the hadronic signal region, with the exception that both the α_{T} and $\not{H}_T/\not{E}_T$ requirements are removed. This selection ensures a very high number of QCD events with respect to the signal region.

Table 1.4 List of `HTxxx` triggers and their efficiencies (%), as measured in data for each H_T bin and n_{jet} category. Also listed are the typical prescales applied and the L1 seed triggers.

Offline H_T region (GeV)	L1 seed (`L1_?`) (highest thresholds)	Trigger (`HLT_?`)	Typical prescale	Efficiency (%) $2 \leq n_{\text{jet}} \leq 3$	Efficiency (%) $n_{\text{jet}} \geq 4$
$200 < H_T < 275$	`DoubleJetC64`	`HT250`	4800	66.4 ± 14.1	154.3 ± 154.3
$275 < H_T < 325$	`DoubleJetC64 OR HTT175`	`HT250`	2400	97.3 ± 23.0	91.7 ± 53.1
$325 < H_T < 375$	`DoubleJetC64 OR HTT175`	`HT300`	1200	79.5 ± 20.6	198.1 ± 81.2
$375 < H_T < 475$	`DoubleJetC64 OR HTT175`	`HT350`	600	108.7 ± 18.7	54.5 ± 31.6
$475 < H_T < 575$	`DoubleJetC64 OR HTT175`	`HT450`	150	110.6 ± 15.9	106.4 ± 26.8
$575 < H_T < 675$	`DoubleJetC64 OR HTT175`	`HT550`	70	96.1 ± 14.7	104.4 ± 23.1
$675 < H_T < 775$	`DoubleJetC64 OR HTT175`	`HT650`	25	94.3 ± 15.4	101.2 ± 21.5
$775 < H_T < 875$	`DoubleJetC64 OR HTT175`	`HT750`	1	96.9 ± 6.1	94.4 ± 8.3
$875 < H_T < 975$	`DoubleJetC64 OR HTT175`	`HT750`	1	100.0 ± 8.4	100.0 ± 12.6
$975 < H_T < 1075$	`DoubleJetC64 OR HTT175`	`HT750`	1	100.0 ± 11.2	100.0 ± 15.3
$H_T > 1075$	`DoubleJetC64 OR HTT175`	`HT750`	1	100.0 ± 15.0	100.0 ± 22.9

1.2.5 HT sideband normalisation

As mentioned previously in section **??**, absolute MC normalisation is not well modelled in the high-$\not{E}_T$ (CHECK) region of phase-space in which many SUSY analyses search. As such, data and MC appear to disagree using 'out-of-the-box' MC samples and cross-sections. While data and MC comparisons are not explicitly used in this analysis, ratios of MC yields are, and so cross-section correction factors for the main MC processes are determined. These are measured as the data to MC ratio in the $150 \leq H_T < 200GeV$ sideband region, in a given control sample with a given selection, designed to produce a pure sample of a background process. A summary of the selections and their relevant purities are given in table 1.5.

Table 1.5 Correction factors determined from a data sideband for the different MC samples. All Correction factors are relative to theoretical cross sections calculated at NNLO. The corrections measured for the W + jets and Z + jets processes, which are in agreement, are also applied to the $Z \to \nu\bar{\nu}$ + jets and γ + jets samples. "Corrected yield" reflects the observed data yield minus the contamination as given by MC. RE

Process	Selection	Purity	Correction factor
W + jets	μ + jets, $2 \leq n_{\text{jet}} \leq 3$, $n_{\text{b}} = 0$	0.91	0.93 ± 0.01
Z($\to \mu\mu$) + jets	$\mu\mu$ + jets, $2 \leq n_{\text{jet}} \leq 3$, $n_{\text{b}} = 0$	0.98	0.94 ± 0.04
$t\bar{t}$	μ + jets, $n_{\text{jet}} \geq 2$, $n_{\text{b}} \geq 2$	0.87	1.21 ± 0.05

These derived correction factors are tested in the closure tests performed following their application (section 1.5.1).

1.3 Estimating multijet backgrounds

A data-driven technique has been developed to determine the necessary α_{T} requirement, per H_T bin, such that QCD is at the sub-percent level with respect to the total EWK background.

Events from the hadronic control sample are used to populate a plane of $\not{H}_T/\not{E}_T$ vs α_{T}. A prediction of the EWK contribution to this sample is determined from the μ + jets control sample with the $\not{H}_T/\not{E}_T$ requirement removed, using the TF factor method described in section 1.1. This is subtracted from the hadronic control sample yields as a function of $\not{H}_T/\not{E}_T$ and α_{T}, subsequently leaving a pure sample of QCD MJ events (SHOW PLOTS?). By considering events both above and below the nominal $\not{H}_T/\not{E}_T$ threshold of 1.25, the ratio $R_{\not{H}_T/\not{E}_T}$ is constructed, defined as:

$$R_{\not{H}_T/\not{E}_T} = \frac{N(\not{H}_T/\not{E}_T < 1.25)}{N(\not{H}_T/\not{E}_T > 1.25)} \tag{1.7}$$

A fit is made to this variable as a function of α_{T}, using an exponential functional form:

$$R_{\not{H}_T/\not{E}_T}(\alpha_{\mathrm{T}}) = e^{-(a+b.\alpha_{\mathrm{T}})^n} \tag{1.8}$$

both for $n = 0$ and also when n is allowed to float as a free parameter within the range $0 - 2$ in order to span the scenarios from flat to a gaussian form. Example distributions and fits are shown in FIGURE.

Predictions are made and compared with the relevant EWK background contributions in table 1.6. α_{T} thresholds are chosen for each H_T bin such that the ratio of QCD to EWK is at the sub-percent level. The chosen α_{T} threshold values are summarised in table 1.7.

RELATED SYSTEMATICS???

1.4 Naive predictions from Transfer Factors

Naive predictions only purely from transfer factors and yields.

SECTION STILL NEEDED?

Maybe just do basic results after all corrections applied, as described above.

Table 1.6 Summary of the predicted QCD multijet background contribution and, for comparison, the expected non-multijet (EWK) background contribution and the ratio (QCD/EWK) as determined in data for events satisfying the various n_{jet}, n_{b}, H_T, and α_{T} threshold requirements. For each QCD prediction, the value quoted is (expected $\pm$ stat. $\pm$ syst.) while for the EWK prediction, the uncertainty comprises both stat. and syst. components. RE

n_{jet}	n_{b}	H_T (GeV)	Bkgd	α_{T} threshold		
				0.550	0.600	0.650
2–3	0	200–275	QCD	$(3.8 \pm 1.3 \pm 1.2) \times 10^3$	$(4.1 \pm 2.4 \pm 3.0) \times 10^1$	$(0.9 \pm 0.8 \pm 1.5) \times 10^0$
2–3	0	200–275	EWK	$(2.1 \pm 0.1) \times 10^4$	$(1.5 \pm 0.0) \times 10^4$	$(1.2 \pm 0.0) \times 10^4$
2–3	0	200–275	Ratio	0.2 ± 0.1	0.003 ± 0.003	0.0001 ± 0.0001
2–3	0	275–325	QCD	$(1.0 \pm 0.3 \pm 1.5) \times 10^4$	$(0.2 \pm 0.1 \pm 0.7) \times 10^0$	$(0.8 \pm 0.3 \pm 4.8) \times 10^{-3}$
2–3	0	275–325	EWK	$(7.9 \pm 0.2) \times 10^3$	$(5.3 \pm 0.2) \times 10^3$	$(4.0 \pm 0.2) \times 10^3$
2–3	0	275–325	Ratio	1 ± 2	0.0000 ± 0.0001	$(0 \pm 1) \times 10^{-6}$
2–3	0	325–375	QCD	$(2.8 \pm 0.4 \pm 2.1) \times 10^1$	$(0.9 \pm 0.2 \pm 1.3) \times 10^{-1}$	$(0.6 \pm 0.4 \pm 1.2) \times 10^{-3}$
2–3	0	325–375	EWK	$(3.4 \pm 0.1) \times 10^3$	$(2.2 \pm 0.1) \times 10^3$	$(1.7 \pm 0.1) \times 10^3$
2–3	0	325–375	Ratio	0.008 ± 0.006	$(4 \pm 6) \times 10^{-5}$	$(4 \pm 8) \times 10^{-7}$
≥ 4	0	200–275	QCD	$(2.8 \pm 1.5 \pm 1.3) \times 10^3$	$(1.1 \pm 0.7 \pm 0.2) \times 10^1$	$(0.2 \pm 0.2 \pm 0.0) \times 10^0$
≥ 4	0	200–275	EWK	$(4.3 \pm 0.3) \times 10^2$	$(2.0 \pm 0.1) \times 10^2$	$(1.0 \pm 0.1) \times 10^2$
≥ 4	0	200–275	Ratio	7 ± 5	0.06 ± 0.04	0.002 ± 0.002
≥ 4	0	275–325	QCD	$(1.5 \pm 1.2 \pm 1.0) \times 10^4$	$(1.5 \pm 1.7 \pm 0.3) \times 10^0$	$(0.1 \pm 0.2 \pm 0.0) \times 10^{-1}$
≥ 4	0	275–325	EWK	$(1.2 \pm 0.0) \times 10^3$	$(5.3 \pm 0.2) \times 10^2$	$(2.9 \pm 0.1) \times 10^2$
≥ 4	0	275–325	Ratio	$(1 \pm 1) \times 10^1$	0.003 ± 0.003	$(4 \pm 7) \times 10^{-5}$
≥ 4	0	325–375	QCD	$(0.7 \pm 0.1 \pm 0.8) \times 10^0$	$(2.5 \pm 0.6 \pm 6.5) \times 10^{-5}$	$(0.7 \pm 0.3 \pm 2.8) \times 10^{-8}$
≥ 4	0	325–375	EWK	$(4.8 \pm 0.3) \times 10^2$	$(2.0 \pm 0.1) \times 10^2$	$(1.1 \pm 0.1) \times 10^2$
≥ 4	0	325–375	Ratio	0.002 ± 0.002	$(1 \pm 3) \times 10^{-7}$	$(1 \pm 3) \times 10^{-10}$

Table 1.7 α_{T} thresholds for each analysis H_T bin as determined from the QCD MJ prediction method, such that QCD is at the sub-percent level.

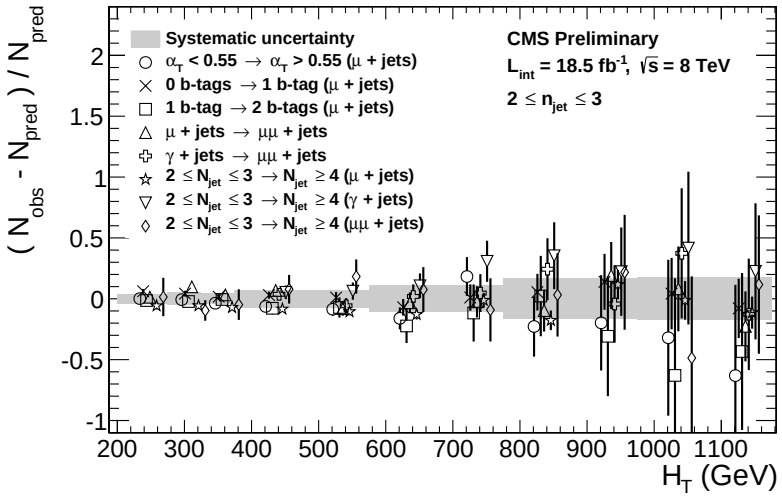
H_T (GeV)	α_{T} threshold
200–275	0.65
275–325	0.60
>325	0.55

1.5 Systematic uncertainties on SM background predictions

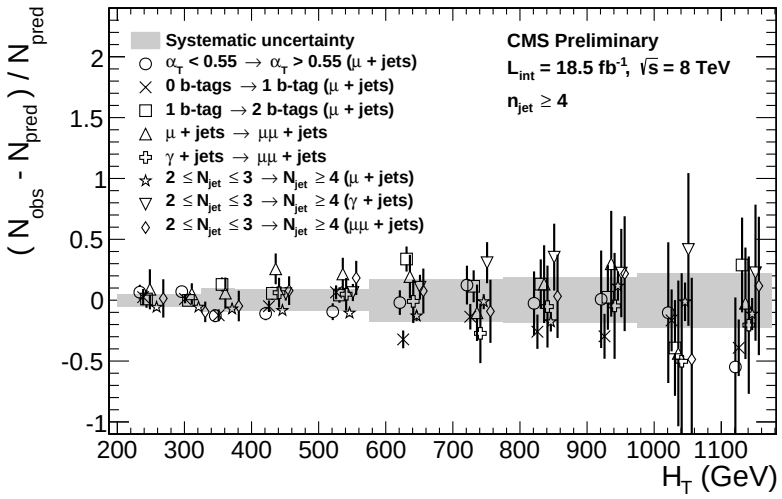
In order to probe the levels at which the transfer factors are sensitive to relevant uncertainties, a statistically powerful ensemble of Closure Tests (CT's) have been designed. The CT method works by constructing a TF to extrapolate from one sub-region of a particular control sample into another control sample sub-region. In doing so, tests can be designed to specifically probe any potential sources of bias in the transfer factors.

1.5.1 Closure tests

The closure tests are performed as a function of H_T, in the two n_{jet} categories, $2 \leq n_{\text{jet}} \leq 3$ and $n_{\text{jet}} \geq 4$. The level of closure represents the statistical consistency between predicted and observed yields for each test, in the absence of any assumed systematic uncertainty. The test statistic is defined as $(N_{obs} - N_{pred})/N_{pred}$, with any bias being observed as a statistically significant deviation from zero, or a trend in H_T.



(a) $2 \leq n_{\text{jet}} \leq 3$



(b) $n_{\text{jet}} \geq 4$

Fig. 1.5 Sets of closure tests (open symbols) overlaid on top of the systematic uncertainty used for each of the five H_T regions (shaded bands) and for the two different jet multiplicity bins: (a) $2 \leq n_{\text{jet}} \leq 3$ and (b) $n_{\text{jet}} \geq 4$. RE

Figure 1.5 shows a summary of the eight closure tests considered as 'core' tests for

the analysis, split into both $2 \leq n_{\text{jet}} \leq 3$ (figure 1.5a) and $n_{\text{jet}} \geq 4$ (figure 1.5b). It should be noted that numerous other tests are also considered, but that these eight represent those deemed most important to the background prediction uncertainty.

The first test, represented by open circles, tests the modelling of the α_{T} variable in the μ + jets control sample. In the analysis a prediction is made between the μ + jets sample, which has no α_{T} requirement, and the signal region, which has a tight α_{T} requirement. This particular test probes the validity of predicting between the 'bulk' distribution of the control sample and the 'tail' distribution in the signal region. Not shown here is a similar test for the $\mu\mu$ + jets control sample.

The next two tests, represented by crosses and open squares, test between different b-tag multiplicities in the μ + jets control sample. The different b-tag requirements greatly change the relative admixture of W + jets (0b) and $\text{t}\bar{\text{t}}$ + jets (1b) events. It is important to note that this test is considered conservative, given that the admixture of W + jets to $\text{t}\bar{\text{t}}$ + jets events varies minimally between control and signal regions, where this extrapolation is made in the analysis. Given the focus on b-tagging, this test also investigates the simulations modelling of b-quark jets.

what about the 1->2 btag μ + jets test?

A similar test is made for the relative admixture of $\text{Z}(\nu\nu)$ + jets to W + jets and $\text{t}\bar{\text{t}}$ + jets, by predicting between the μ + jets and $\mu\mu$ + jets control regions, represented by open triangles. Again, this test is considered conservative, but also probes the muon reconstruction and trigger efficiencies between the different muon multiplicities. These are however already well studied with data-driven techniques by the muon POG.

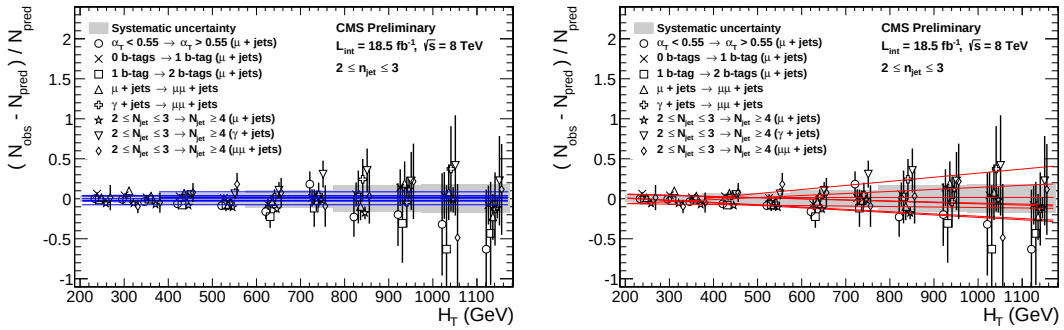
As described in section 1.1, the $Z \rightarrow \nu\bar{\nu}$ prediction is made from both the γ + jets and $\mu\mu$ + jets samples, and so a test is constructed to predict between these two orthogonal control regions, as shown by the open crosses.

The final three tests, indicated by open stars, triangles and diamonds, make predictions between the two different jet multiplicity categories in each control sample, thereby testing jet reconstruction and modelling. This test is also considered very conservative as the analysis only predicts between identical n_{jet} categories in the control and signal regions.

The summary plots of these eight tests shown in figure 1.5 indicate that there is no stastically significant biases or H_T dependencies. Figures 1.6a and 1.6c show zeroeth order polynominal fits (blue lines) made to each individual test to assess the level of any potential bias present. In addition, first order polynominal fits (red lines) are also made to assess any potential H_T dependence present in the tests. The best-fit

values, χ^2 and p-values obtained from both fits are summarised for each n_{jet} category in tables 1.8, 1.9 and 1.10.

As expected, the fits show no significant biases or trends, and therefore indicate good closure. The only exception is the 0 b-jets →1 b-jet (μ + jets) test for the $n_{\text{jet}} \geq 4$ category which indicates a sub-optimal goodness of fit value. This is attributed to upwards and downwards fluctuations in the 475-575 GeV and 575-675 GeV bins respectively. Also shown in table 1.9 is the same fit made when summing these two bins - a fit which returns significantly improved fit values. This leads to the conclusion that these two bins contian a statistical fluctuation as opposed to a systematic bias.



(a) $2 \leq n_{\text{jet}} \leq 3$ (zeroeth order polynomial fits)

(b) $2 \leq n_{\text{jet}} \leq 3$ (first order polynomial fits)

(c) $n_{\text{jet}} \geq 4$ (zeroeth order polynomial fits)

(d) $n_{\text{jet}} \geq 4$ (first order polynomial fits)

Fig. 1.6 Sets of closure tests (open symbols) overlaid on top of the systematic uncertainty used for each of the five H_T regions (shaded bands), for the two different jet multiplicity bins (top row) $2 \leq n_{\text{jet}} \leq 3$ and (bottom row) $n_{\text{jet}} \geq 4$, and with zeroeth (left column) and first (right column) order polynomial fits to each set of closure tests. RE

Maybe also include a couple other relevant tests.

may also include SITV closure tests

Table 1.8 A summary of the results obtained from fits of zeroeth order polynomials (i.e. a constant) to five sets of closure tests performed in the $2 \leq n_{\text{jet}} \leq 3$ bin. The final two columns show the best fit value for the slope obtained when performing a linear fit and the p-value for the linear fit. RE

		Constant fit				Linear fit	
Closure test	Symbol	Best fit value	χ^2	d.o.f.	p-value	Slope (10^{-4})	p-value
$\alpha_{\text{T}} < 0.55 \rightarrow \alpha_{\text{T}} > 0.55$ (μ + jets)	Circle	-0.02 ± 0.01	11.3	10	0.34	-2.9 ± 1.1	0.83
0 b-jets $\rightarrow$1 b-jet (μ + jets)	Times	0.04 ± 0.01	5.8	10	0.83	-1.5 ± 0.9	0.97
1 b-jet $\rightarrow$2 b-jets (μ + jets)	Square	-0.03 ± 0.02	5.3	10	0.87	-3.0 ± 1.7	0.99
μ + jets $\rightarrow\mu\mu$ + jets	Triangle	0.03 ± 0.02	12.3	10	0.27	-1.3 ± 1.1	0.28
γ + jets $\rightarrow\mu\mu$ + jets	Cross	-0.02 ± 0.03	3.0	7	0.88	0.0 ± 2.7	0.81

Table 1.9 A summary of the results obtained from fits of zeroeth order polynomials (i.e. a constant) to five sets of closure tests performed in the $n_{\text{jet}} \geq 4$ bin. The final two columns show the best fit value for the slope obtained when performing a linear fit and the p-value for the linear fit. †See text for details on this particular fit. RE

		Constant fit				Linear fit	
Closure test	Symbol	Best fit value	χ^2	d.o.f.	p-value	Slope (10^{-4})	p-value
$\alpha_{\text{T}} < 0.55 \rightarrow \alpha_{\text{T}} > 0.55$ (μ + jets)	Circle	-0.02 ± 0.02	17.6	10	0.06	-3.1 ± 1.7	0.11
0 b-jets $\rightarrow$1 b-jet (μ + jets)	Times	-0.06 ± 0.02	31.2	10	0.00	-4.1 ± 1.2	0.02
0 b-jets $\rightarrow$1 b-jet (μ + jets)†	Times	-0.05 ± 0.02	13.4	9	0.15	-3.9 ± 1.3	0.78
1 b-jet $\rightarrow$2 b-jets (μ + jets)	Square	0.06 ± 0.02	13.7	10	0.19	2.5 ± 1.6	0.28
μ + jets $\rightarrow\mu\mu$ + jets	Triangle	0.11 ± 0.05	4.8	10	0.90	0.4 ± 2.7	0.85
γ + jets $\rightarrow\mu\mu$ + jets	Cross	-0.00 ± 0.07	2.3	7	0.94	-5.3 ± 4.7	0.99

Table 1.10 A summary of the results obtained from fits of zeroeth order polynomials (i.e. a constant) to three sets of closure tests ($2 \leq n_{\text{jet}} \leq 3 \rightarrow n_{\text{jet}} \geq 4$) that probe the accuracy of the MC modelling of the n_{jet} distribution observed in data, using the three data control samples. RE

		Constant fit				Linear fit	
Sample	Symbol	Best fit value	χ^2	d.o.f.	p-value	Slope (10^{-4})	p-value
μ + jets	Star	-0.08 ± 0.01	9.3	10	0.50	0.6 ± 0.7	0.48
γ + jets	Inverted triangle	0.09 ± 0.04	3.7	7	0.82	5.1 ± 3.2	0.98
$\mu\mu$ + jets	Diamond	-0.00 ± 0.05	4.7	10	0.91	2.5 ± 2.9	0.92

1.5.2 Background uncertainty summary

Under the assumption of closure for the eight core tests, systematic errors are derived for each n_{jet} category in seven regions of H_T.

Values are calculated by summing in quadrature the weighted mean and sample variance for all eight tests in a given H_T region. These values are summarised in table 1.11 and also in the summary plots of figure 1.5, shown as grey bands.

Systematic values are here considered as fully uncorrelated between the different analysis categories and the H_T regions defined here, which is again considered as a conservative approach given that some correlation is to be expected, for example between adjacent H_T bins.

Table 1.11 A summary of the magnitude of the systematic uncertainties (%) assigned to the transfer factors, according to n_{jet} and H_T region. RE

	H_T region (GeV)						
n_{jet}	200–275	275–325	325–375	375–575	575–775	775-975	> 975
2–3	4	6	6	8	12	17	19
$\geq$4	6	6	11	11	18	20	26

NOTE - All plots currently from:

```
out_dict["28Jan_fullLatestReRun_dPhi_gt0p3_v0"] = {

        "path_name": "rootfiles/Root_Files_28Jan_fullLatestReRun_dPhi_gt0p3

        # All Runs
        "had_lumi": 18.493,
        "mu_lumi": 19.131,
        "ph_lumi": 19.12,

        # taken from parked final (change if necessary)
        "wj_corr": 0.94,
        "dy_corr": 0.94,
        "tt_corr": 1.17,

}
```

References