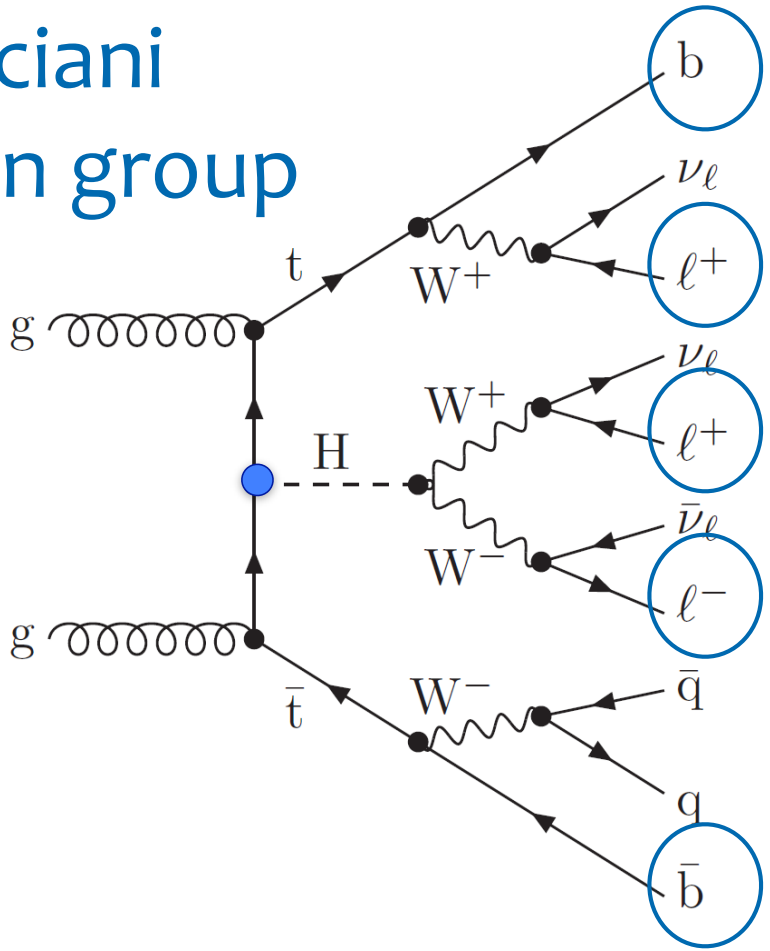
Background modeling for ttH Multilepton analysis

Cristina Botta, Giovanni Petrucciani
on behalf of the ttH-multilepton group



Outline

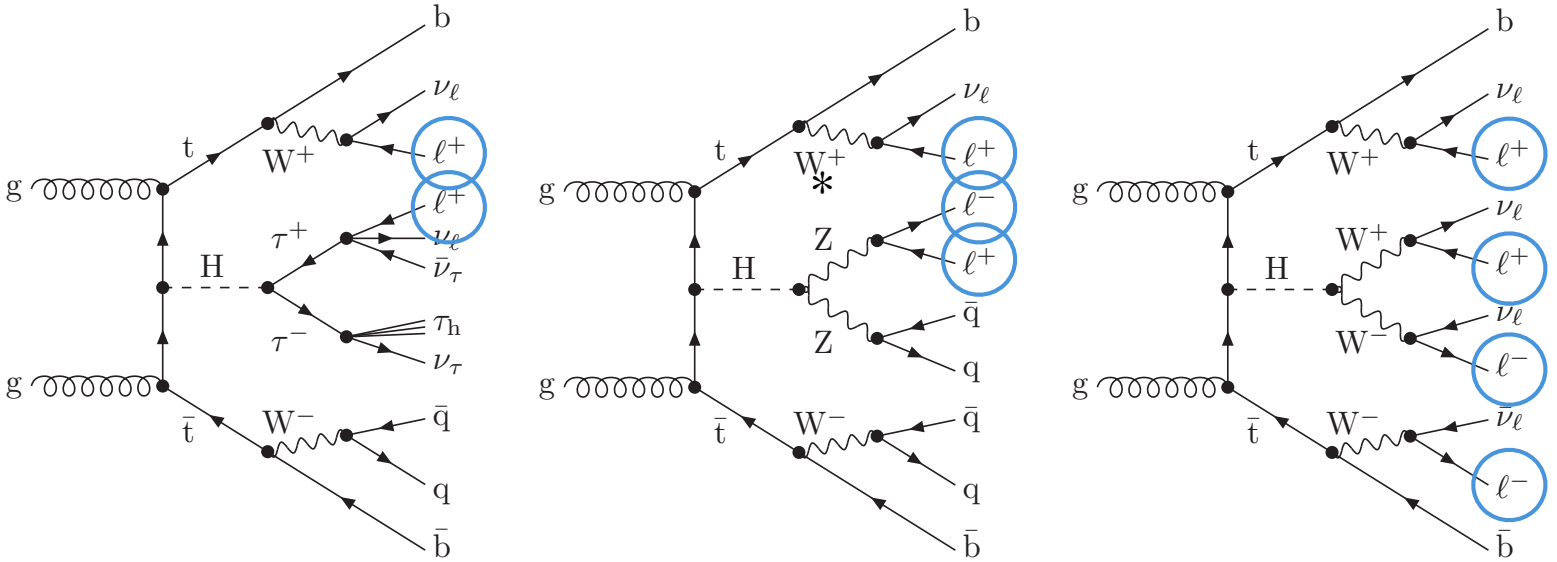
- Short introduction to the leptonic ttH analysis strategy
- Descriptions of main backgrounds
- Background modeling procedures in Run I analysis
- Background modeling systematics in Run I analysis
- MC needs for Run II analysis

ttH in leptonic final states

- An analysis targeting ttH production in leptonic (e,μ) final states from **H->WW, ττ, ZZ**

[Run 1 CMS paper: arXiv:1408.1682]

- Three final states considered:
  - **2 same-sign leptons** (ee, μμ, eμ) **+ b-jets**
  - **3 leptons + b-jets** (with no resonant Z->ℓℓ)
  - **4 leptons** (other than H->ZZ->4ℓ - no resonant Z->ℓℓ) **+ b-jets**
- Signal extraction from the output distribution of a BDT trained against ttbar, using kinematics (n_{Jets}, E_T^{miss}, $|\eta_\ell|$,...)



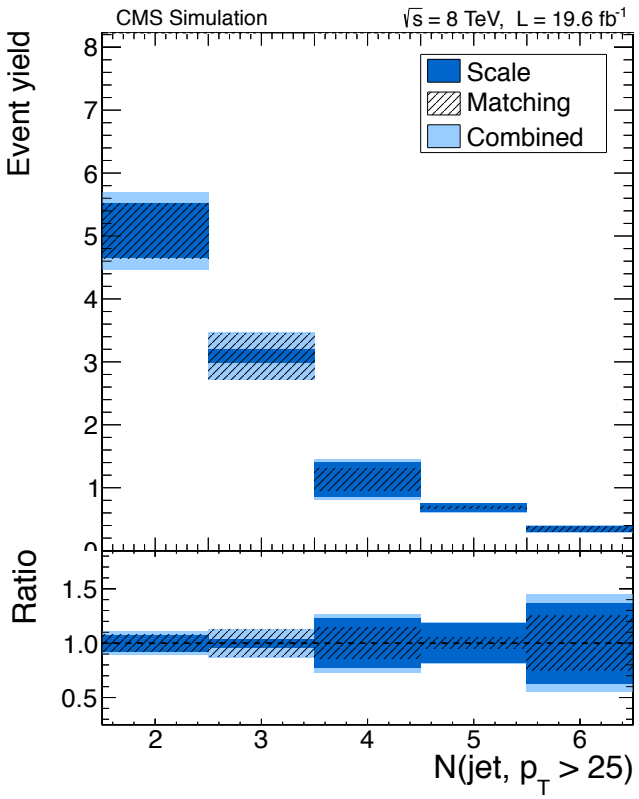
Backgrounds

- 2lss / 3l
  - **from simulation: TTZ/γ*, TTW**
    (irreducible) [MG 5+pyt6 @ LO; ttW+0/1/2 jets, ttZ + 0/1 jets]
  - **from simulation but normalized to data in CR: WZ (VVV, W±W±)**
    (mainly mistagged b-jets) [MG 5 + pyt6 @ LO, WZ+0/1/2 jets]
  - **data-driven (2lss only): TT (Z/Wbb, TW, DY, W)**
    (charge-flip leptons and mistagged/real b-jets)
  - **data-driven: TT (Z/Wbb, TW, DY, W)**
    (fake leptons and mistagged/real b-jets)
- 4l:
  - **from simulation: TTZ/γ*** (irreducible)
  - **from simulation but normalized to data in CR: ZZ**
    (mainly mistagged b-jets) [MG 5+pyt6 @ LO, ZZ+0/1 jet; Powheg+py6]
  - **data-driven: TT, TTW, (Zbb, DY, WZ)**
    (fake leptons and mistagged b-jets)

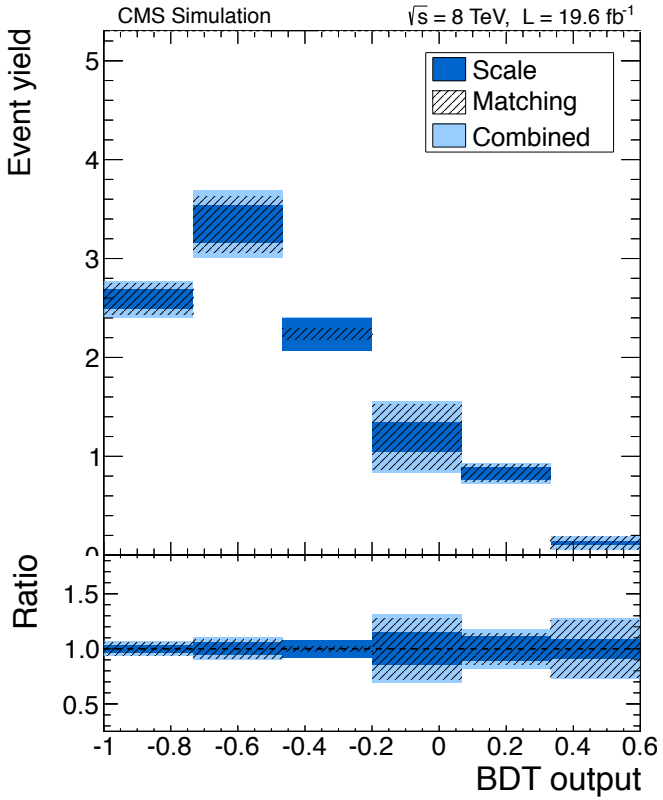
Theoretical uncertainties
for backgrounds from MC

- Uncertainties on the ttW, ttZ NLO cross section from unknown **higher orders**, 10%, 11% [effect on rate] [arXiv:1208.2665]
- Uncertainties on the ttW, ttZ NLO cross section from the knowledge of **PDFs**, 7%, 9% [effect on rate]
- Uncertainties on shapes from **PDFs** (PDF4LHC recipe with MSTW2008lo, CT10, NNPDF 2.3) [effect on shapes]
- Uncertainty on the charge asymmetry for ttW
- Uncertainties from the **MC modeling** of the process estimated conventionally (varying normalization and factorization scales up and down and matching threshold between matrix element and parton shower) [effect on shapes]

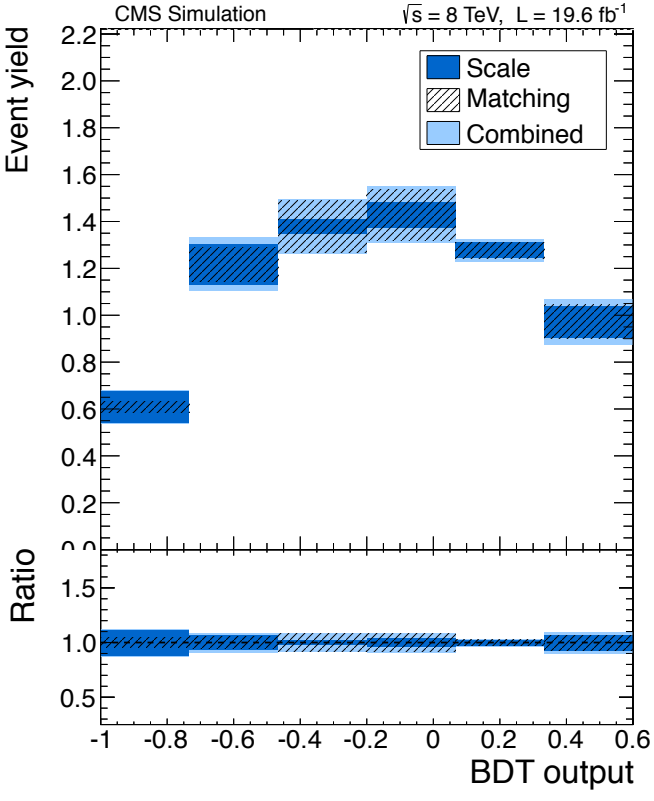
Theoretical uncertainties for backgrounds from MC



eμ - NJets
Scale and Matching
ttW



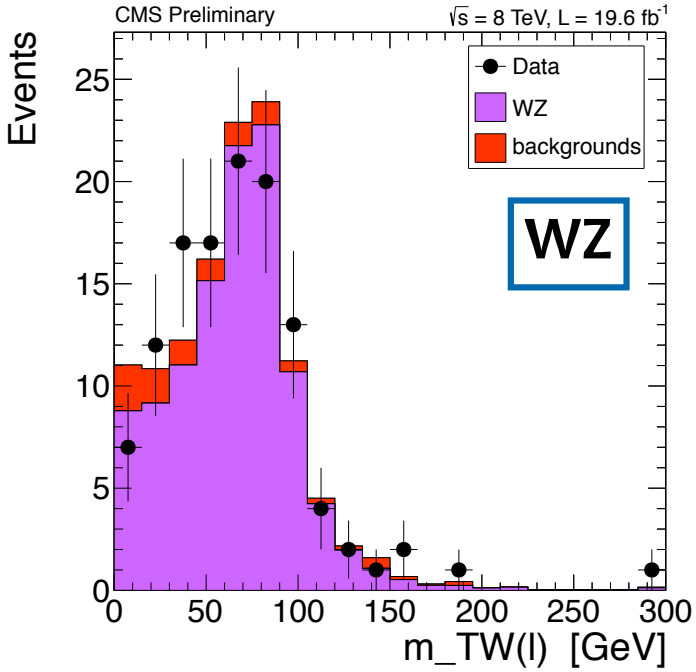
3ℓ - BDT
Scale and Matching
ttW

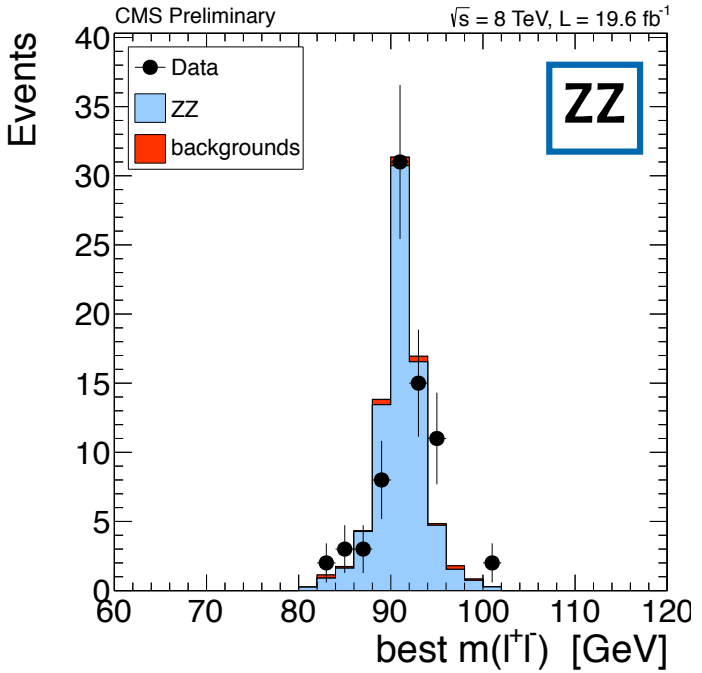


3ℓ - BDT
Scale and Matching
ttZ

Di-boson backgrounds

- Control region for WZ/ZZ + light-jets
  - WZ: 3ℓ, inverted Z veto, >= 2j no b-jets
  - ZZ: 4ℓ, inverted Z veto, >=1j no b-jets (purity ~95% in both CRs)
- From CR to SR with HF/LF ratios:
  - WZb(b)/WZ(2j,0b) and ZZb(b)/ZZ(1j, 0b)
  - taken from simulation
- Systematics ~30%:
  - Stat. uncertainty & backgrounds in CR
  - B-tagging efficiencies and mistag rates (dominant uncertainty: VZ+tag is mainly from WZ+light with mistags)
  - PDF uncertainties (PDF4LHC recipe, ~5%)
  - Missing higher orders: (ZZ+2j)/(ZZ+1j), (WZ+bb)/(WZ+jj), (WZ+c)/(WZ+j)





Fake lepton backgrounds

- Background with fake leptons (from jet-misID, B hadrons decay) is estimated with the "fake-rate" method
  - **application region** obtained from the signal region inverting the lepton ID requirement on one lepton
    - dominated by fake lepton backgrounds
  - **extrapolation to signal region** is performed relying on the probability for a fake lepton - which has passed the preselection - to satisfy the lepton ID requirement (*fake rate*)
  - **measurement region** is where the fake-rate is derived in data
    - must have same sources of fake leptons as application region
- The fake rate depends on:
  - lepton flavour
  - lepton kinematic (p_T, η)
  - lepton sources: from B hadrons decay vs lepton from light-jets mis-Id or decay in flight

Fake rate estimation

- **The application region** is expected from MC studies to be dominated by tt and single top events (leptons from B-hadrons decay)
  - Checked the MC modeling of this process by scaling up the tt and single top to match the data yields (factor 1.5-2)
  - Studies on MC performed to determined measurement regions with same composition of the application region

- **The fake rate is derived separately for events**
  - **with at most 1 b-medium tagged jet or 2 b-loose tagged jets [B-loose]** (application region with ~2% fake leptons from light flavour)
  - **with 2 b-medium tagged jets [B-tight]** (application region with ~10% fake leptons from light flavour)

- **The fake rate is derived in 3 measurements regions** (measurements from the different control regions in data and MC are combined)
  - **QCD di-lepton control region, Z+l control region** (only for electron at low p_T because not enriched in leptons from B hadrons decay)

Uncertainties for fake lepton bkg

- Fakes **normalization**
  - the fake rate expectation from simulation in each control region is compatible with the expectations from fake leptons from tt MC within their statistic uncertainties and an additional systematics of about **40% (60%)** for inclusive events (events with at least two b-jets)
  - the overall fake-rate is scaled by **± 40% (60%)** and bkg prediction are derived accordingly
- Fakes **shapes**
  - **different variations** of fake-rate dependent on **p_T and η**
  - for each variation derive a bkg prediction
  - all bkg normalized to the same yield, **envelope of the variations** taken as an uncertainty band for the shape of the bkg
  - in signal extraction we allow **linear or quadratic deformation** of the discriminating variable shape within the band
  - additional systematic from the difference between the fake-rate shape and the MC shape

In preparation for Run II

- Data driven estimation for fake-lepton bkg
  - new ideas to reduce the uncertainties on the lepton fake rate estimate
- Pursue better discrimination between ttH and ttV at the signal extraction step, to reduce impact on ttV systematics
  - Run 1 analysis focused only on ttbar vs ttH separation
- Use newer, more accurate, MC simulations for backgrounds
  - e.g. MG5_aMC@NLO ttX + jets with FxFx merging

ttV backgrounds: validation

- Compared MG5+pyt6 vs PowHel and aMC@NLO+herwig++
  - agreement at the level of the systematical uncertainties assigned to MG5+pyt6 from scale variation and matching
- Performed also cross-check measurement of ttV and simultaneous ttV+ttH fit using extra control regions:
  - same-sign dileptons + exactly 3 jets
  - trileptons with one resonant Z->$\ell\ell$
  - fitted scale factors for ttW and ttZ compatible with the ($\mu(ttW) = 1.4^{+0.6}/_{-0.5}$, $\mu(ttZ) = 1.1^{+0.4}/_{-0.3}$)