H->4ℓ studies

Data driven estimation of the $\ell\ell\mu\mu$ *reducible* bkg

National and Kapodistrian UNIVERSITY OF ATHENS

Panagiotis Bellos

ATLAS EXPERIMENT

HEP 2019 - Conference on Recent Developments in High Energy Physics and Cosmology 17-20 April 2019, NCSR "DEMOKRITOS", Athens, Greece

19/4/2019

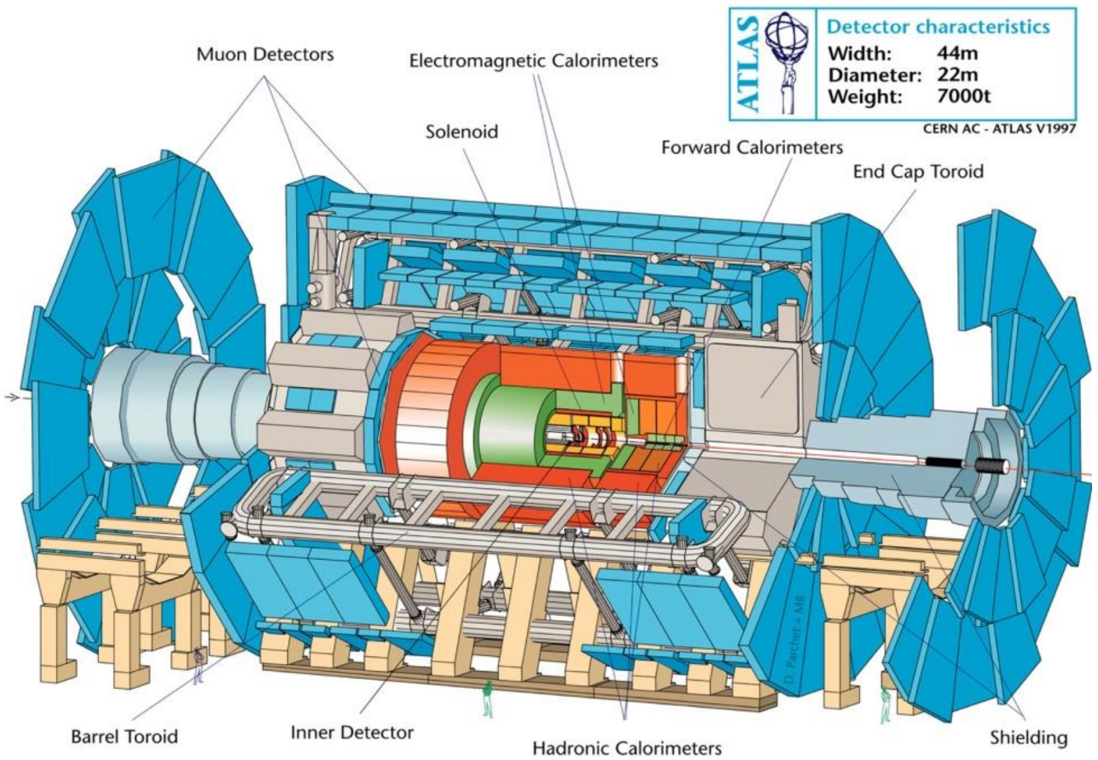
Ευρωπαϊκή Ένωση
European Social Fund

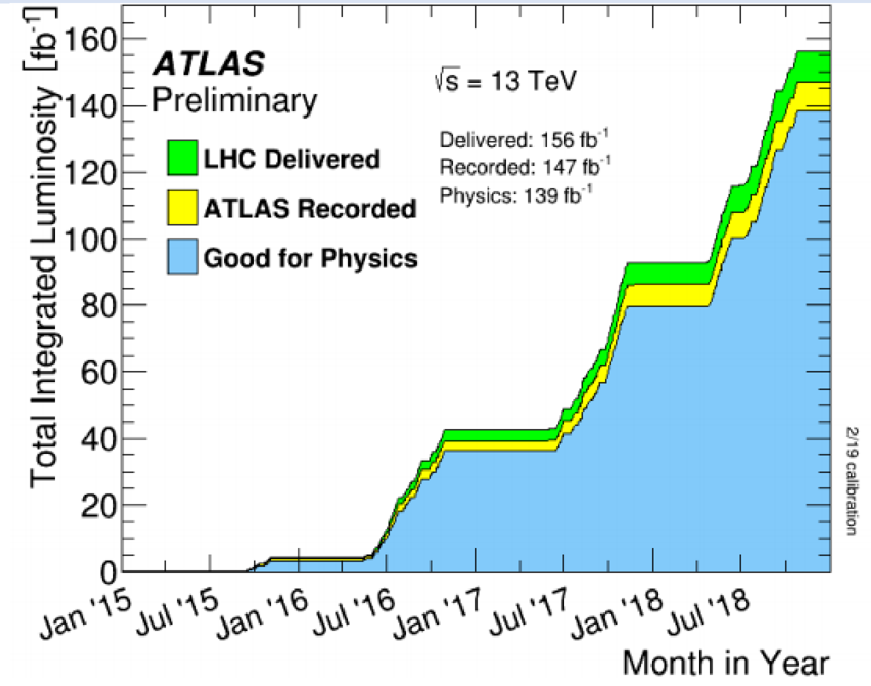
Operational Programme
Human Resources Development, Education and Lifelong Learning
Co-financed by Greece and the European Union

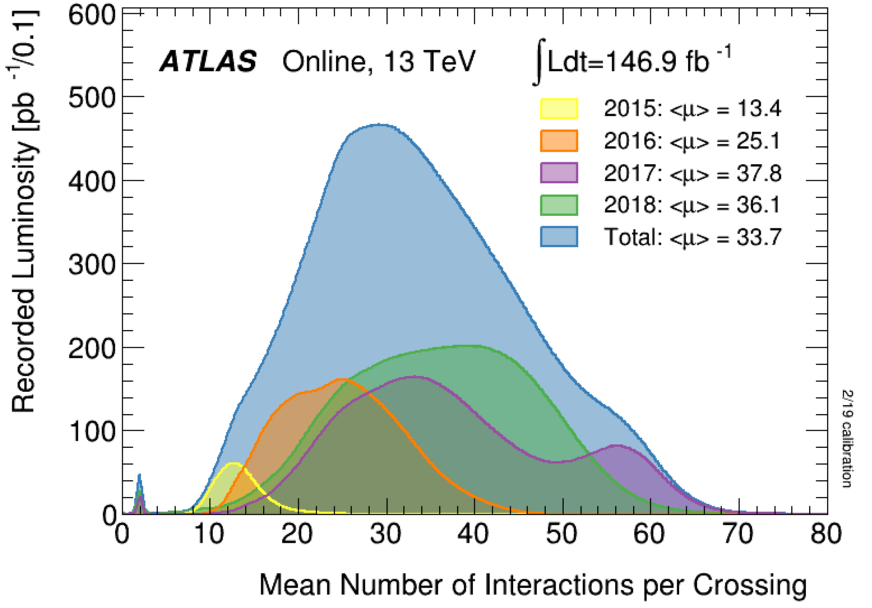
ΕΣΠΑ 2014-2020
ανάπτυξη - εργασία - αλληλεγγύη

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY)

ATLAS Experiment







Inner Detector

Charged particles track reconstruction

- Pixel Detector
- Semiconductor Tracker
- Transition Radiation Tracker

\- Solenoid Magnet

Calorimeters

Energy measurement

- e/m Calorimeter
- Hadronic Calorimeter

Muon Spectrometer

Muon track reconstruction

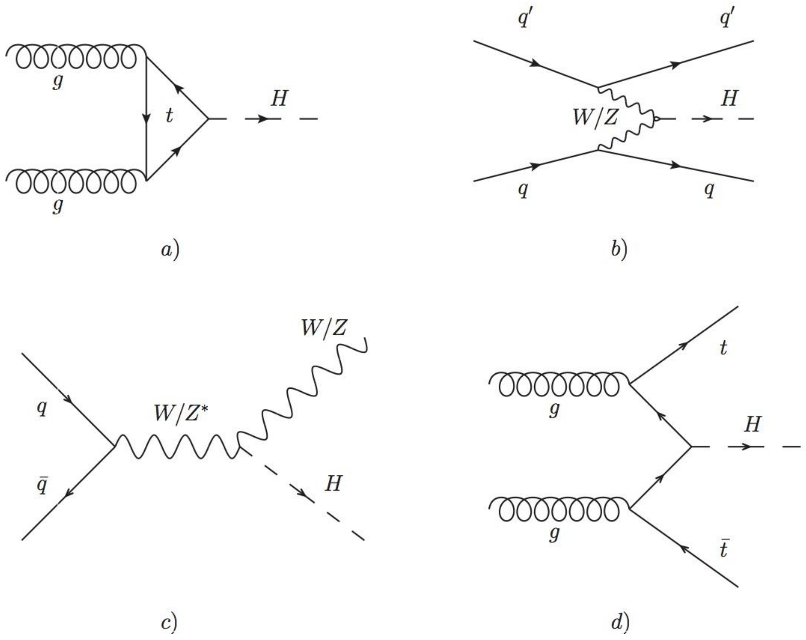
- Monitored Drift Tubes
- Cathode Strip Chambers
- Resistive Plate Chambers
- Thin Gap Chambers

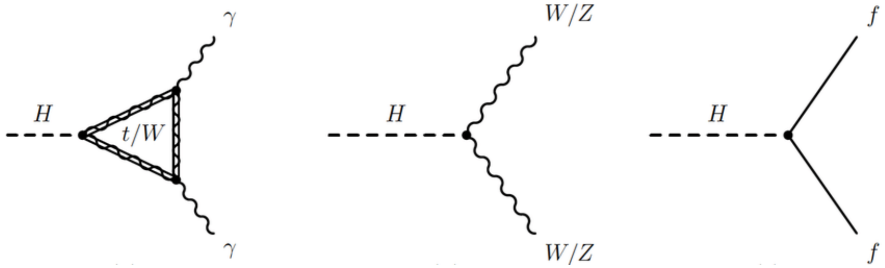
\- Toroidal magnets

Higgs Boson Production and Decays

$\sqrt{s}$ (TeV)	Production cross section (in pb) for $m_H = 125$ GeV					
	ggF	VBF	WH	ZH	$t\bar{t}H$	total
1.96	$0.95^{+17\%}_{-17\%}$	$0.065^{+8\%}_{-7\%}$	$0.13^{+8\%}_{-8\%}$	$0.079^{+8\%}_{-8\%}$	$0.004^{+10\%}_{-10\%}$	1.23
7	$16.9^{+5\%}_{-5\%}$	$1.24^{+2\%}_{-2\%}$	$0.58^{+3\%}_{-3\%}$	$0.34^{+4\%}_{-4\%}$	$0.09^{+8\%}_{-14\%}$	19.1
8	$21.4^{+5\%}_{-5\%}$	$1.60^{+2\%}_{-2\%}$	$0.70^{+3\%}_{-3\%}$	$0.42^{+5\%}_{-5\%}$	$0.13^{+8\%}_{-13\%}$	24.2
13	$48.6^{+5\%}_{-5\%}$	$3.78^{+2\%}_{-2\%}$	$1.37^{+2\%}_{-2\%}$	$0.88^{+5\%}_{-5\%}$	$0.50^{+9\%}_{-13\%}$	55.1
14	$54.7^{+5\%}_{-5\%}$	$4.28^{+2\%}_{-2\%}$	$1.51^{+2\%}_{-2\%}$	$0.99^{+5\%}_{-5\%}$	$0.60^{+9\%}_{-13\%}$	62.1

Decay channel	Branching ratio	Rel. uncertainty
$H \to \gamma\gamma$	2.27×10^{-3}	$^{+5.0\%}_{-4.9\%}$
$H \to ZZ$	2.62×10^{-2}	$^{+4.3\%}_{-4.1\%}$
$H \to W^+W^-$	2.14×10^{-1}	$^{+4.3\%}_{-4.2\%}$
$H \to \tau^+\tau^-$	6.27×10^{-2}	$^{+5.7\%}_{-5.7\%}$
$H \to b\bar{b}$	5.84×10^{-1}	$^{+3.2\%}_{-3.3\%}$
$H \to Z\gamma$	1.53×10^{-3}	$^{+9.0\%}_{-8.9\%}$
$H \to \mu^+\mu^-$	2.18×10^{-4}	$^{+6.0\%}_{-5.9\%}$





The decay $H \to ZZ^{(*)} \to 4\ell$ **(ℓ: e or μ)** is the "golden" channel for Higgs boson studies, due to:

High signal-to-background ratio

Ability to completely reconstruct the final state

H->4ℓ *Backgrounds*

– The most significant **irreducible** bkg is **qq->ZZ** and **gg->ZZ.** Minor contributions from **Triboson** and $t\bar{t}$**+V** production.

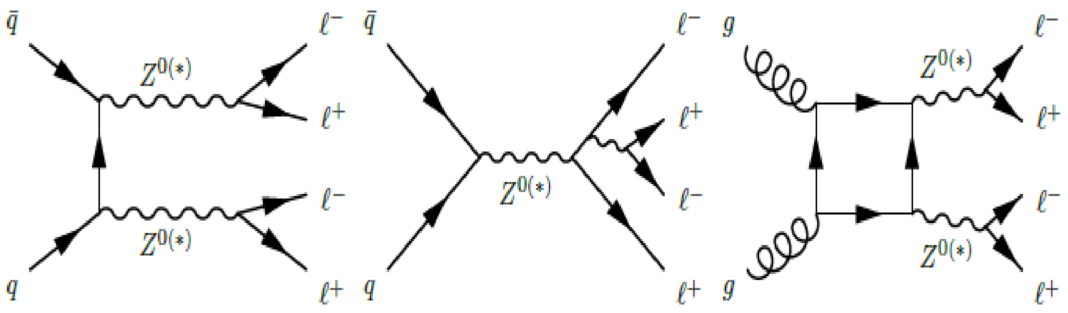
– The most significant **reducible** bkg is **Z+jets** which is separated in **Z + Heavy Flavor (b,c) jets** and **Z + Light Flavor (u,d,s) jets.**

– The $t\bar{t}$ has also important contribution and it also produce HF jets. Smaller contribution from **WZ.**

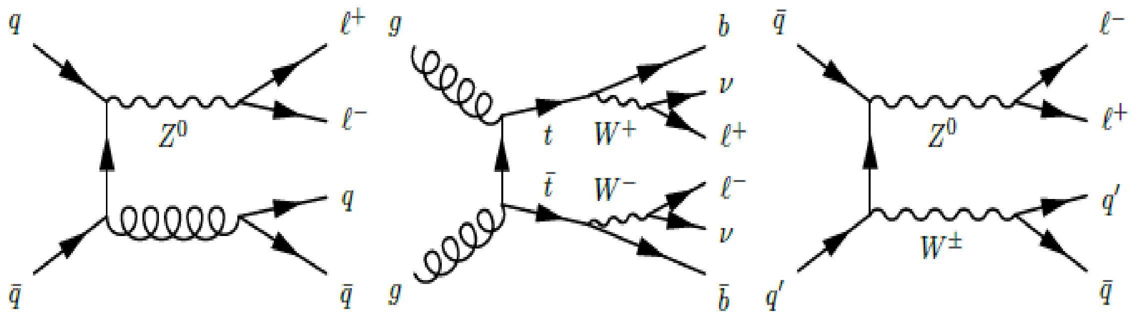
Leptons can be produced by the semileptonic decays of heavy flavor hadrons

Muons can be produced by light flavor hadrons (π/K) in-flight decays

LF jets and γ can also produce fake electrons



Main H->4l irreducible bkgs



H->4l reducible bkgs

Event Selection

Physics Objects
Electrons
Loose Likelihood quality electrons with hit in innermost layer, $E_T > 7$ GeV and $\|\eta\| < 2.47$
Interaction point constraint: $\|z_0 \cdot \sin\theta\| < 0.5$ mm (if ID track is available)
Muons
Loose identification with $p_T > 5$ GeV and $\|\eta\| < 2.7$
Calo-tagged muons with $p_T > 15$ GeV and $\|\eta\| < 0.1$, segment-tagged muons with $\|\eta\| < 0.1$
Stand-alone and silicon-associated forward restricted to the $2.5 < \|\eta\| < 2.7$ region
Combined, stand-alone (with ID hits if available) and segment-tagged muons with $p_T > 5$ GeV
Interaction point constraint: $\|d_0\| < 1$ mm and $\|z_0 \cdot \sin\theta\| < 0.5$ mm (if ID track is available)
Jets
anti-k_T jets with *bad-loose* identification, $p_T > 30$ GeV and $\|\eta\| < 4.5$
Jets with $p_T < 60$ GeV and $\|\eta\| < 2.4$ are required to pass the pile-up jet rejection at the 92% working point (JVT score > 0.59).
***b*-tagging**
Previously selected jets with $\|\eta\| < 2.5$ passing the `MV2_c10` algorithm at its 70% working point
Overlap removal
Jets within $\Delta R < 0.2$ of an electron or $\Delta R < 0.1$ of a muon are removed

	Event Selection
Quadruplet Selection	- Require at least one quadruplet of leptons consisting of two pairs of same-flavour opposite-charge leptons fulfilling the following requirements:
	- p_T thresholds for three leading leptons in the quadruplet: 20, 15 and 10 GeV
	- Maximum one calo-tagged or stand-alone muon or silicon-associated forward per quadruplet
	- Select best quadruplet (per channel) to be the one with the (sub)leading dilepton mass (second) closest the Z mass
	- Leading di-lepton mass requirement: $50 < m_{12} < 106$ GeV
	- Sub-leading di-lepton mass requirement: $12 < m_{34} < 115$ GeV
	- $\Delta R(\ell,\ell') > 0.10$ (0.20) for all same (different) flavour leptons in the quadruplet
	- Remove quadruplet if alternative same-flavour opposite-charge di-lepton gives $m_{\ell\ell} < 5$ GeV
Isolation	- Contribution from the other leptons of the quadruplet is subtracted
	- Muon track isolation ($\Delta R <= 0.30$): $\Sigma p_T/p_T < 0.15$
	- Muon calorimeter isolation ($\Delta R = 0.20$): $\Sigma E_T/p_T < 0.30$
	- Electron track isolation ($\Delta R <= 0.20$) : $\Sigma E_T/E_T < 0.15$
	- Electron calorimeter isolation ($\Delta R = 0.20$) : $\Sigma E_T/E_T < 0.20$
Impact Parameter Significance	- Apply impact parameter significance cut to all leptons of the quadruplet
	- For electrons: $d_0/\sigma_{d_0} < 5$
	- For muons: $d_0/\sigma_{d_0} < 3$
Vertex Selection	- Require a common vertex for the leptons:
	- χ^2/ndof < 6 for 4μ and < 9 for others.

<u>A quadruplet of leptons is required</u>

- 2 pairs of leptons (dileptons) with invariant masses, flavor and charge compatible with the decayed Z & Z* assumption
- Isolated (rejects leptons inside jets)
- Small impact parameter significance (rejects leptons from HF)
- Coming from the same vertex

Leading dilepton is the one with invariant mass closest to Z and **Subleading** the next-to-closest to Z mass

Data Driven Estimation

- **The reducible bkg is estimated in a Data Driven way** in order to avoid the theoretical and simulation uncertainties.

- The properties of the reducible bkg are determined mainly by the subleading dilepton.

- So the bkg is separated to :
  - ***ℓℓμμ***, mostly **heavy flavor jets** produced in association with a **Z** boson or in $t\bar{t}$ **decays**. Minor **Z+LF** contribution.
  - ***ℓℓee,*** significant contribution from **Z+LFjets** and fake electrons from **γ,** lower contribution from **HF jets**

- For the ***ℓℓμμ*** bkg estimation a **Relaxed Region** (RR) is constructed, where no Isolation, impact parameter and vertexing criteria are applied to the subleading dilepton.

- Inverting some selection criteria, special **Control Regions** (CRs) with high bkg purity are constructed, suitable for the bkg estimation.

- A model for each bkg component is made for each region **parametrizing the leading dilepton invariant mass** distribution.

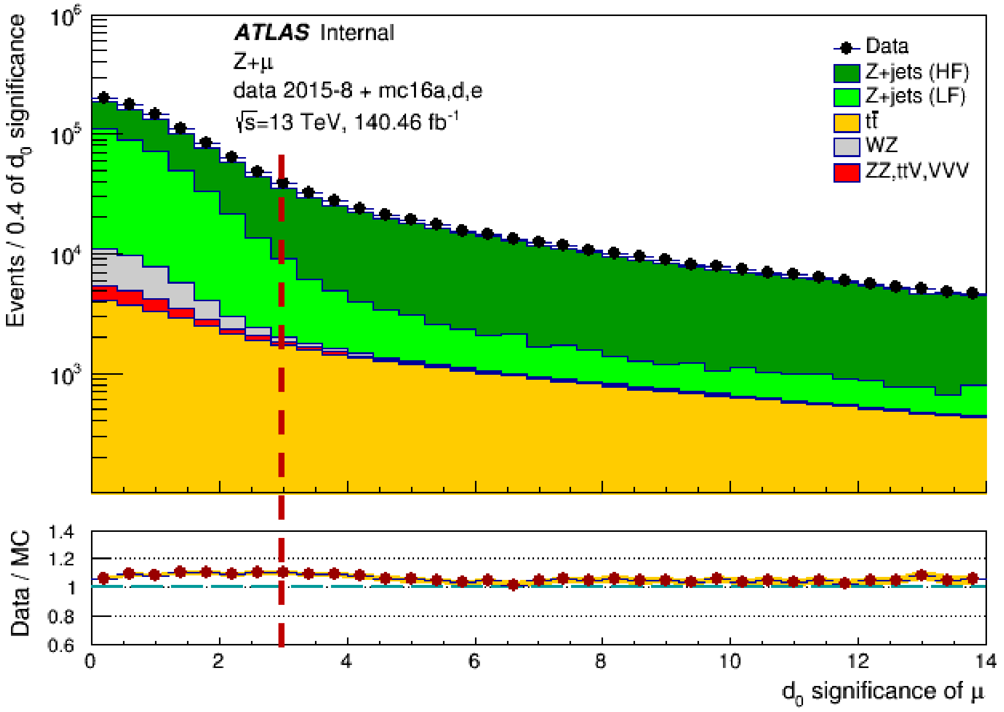
- The models are **fitted simultaneously** in all the CRs to the data in order to estimate the number of events of each bkg component in each region.

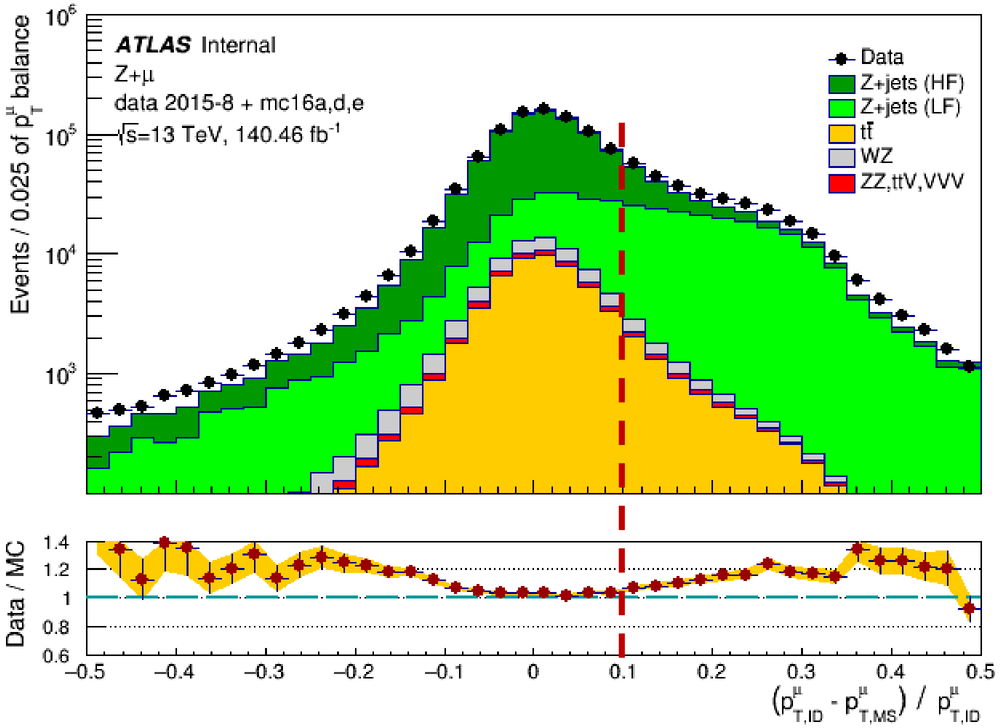
- The results of the fit are expressed in the RR, via **MC constrained fractions**.

- The estimations are transferred to the **Signal Region** (SR) (where all the selection criteria are applied) by extrapolation from the RR using **Transfer Factors** (TF).

$Z + \mu$ Control Sample

- Transfer Factors are the **ratio of the yields between the Signal Region and the Relaxed Region**. They are estimated from MC and **controlled from data using the Z+μ sample**.
- This sample includes ll+μ events; the dilepton passes the complete event selection and no criteria are applied to the additional μ.
- Estimate the **efficiency** of the selection criteria **for the additional μ**, from data and MC. **TF is roughly the efficiency squared**.
- For data estimation a HF enriched sample is constructed using d0 significance > 3 and a LF using p_T imbalance > 0.1. The rest of the contributions in the enriched samples are subtracted using MC.





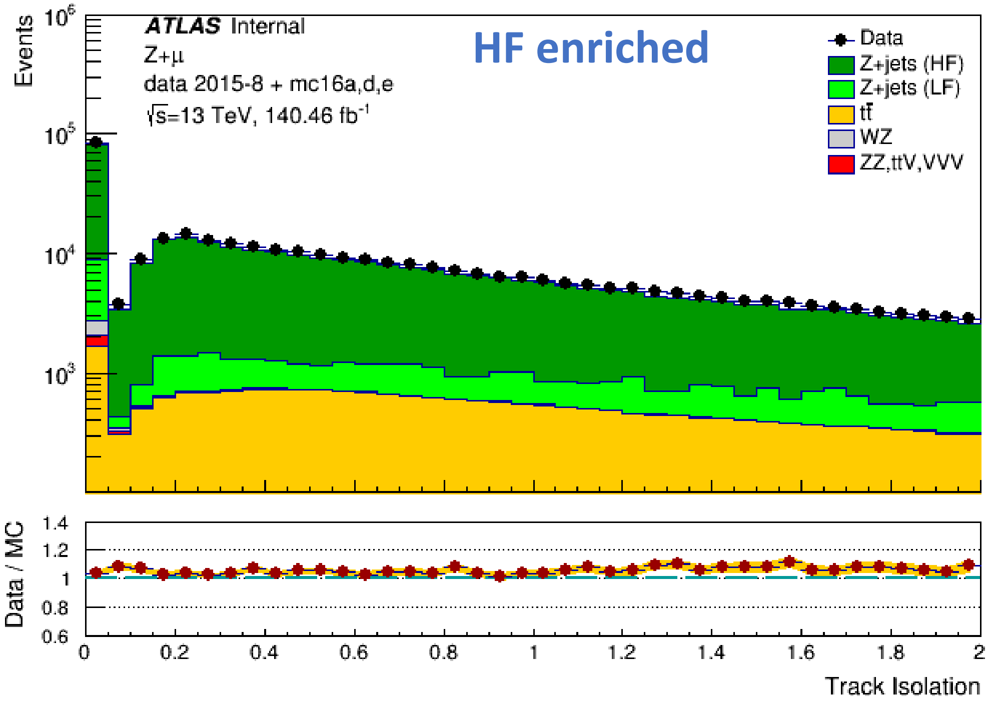
$$P_{T\ imb} = (P_{T\ ID} - P_{T\ MS})/P_{T\ ID}$$

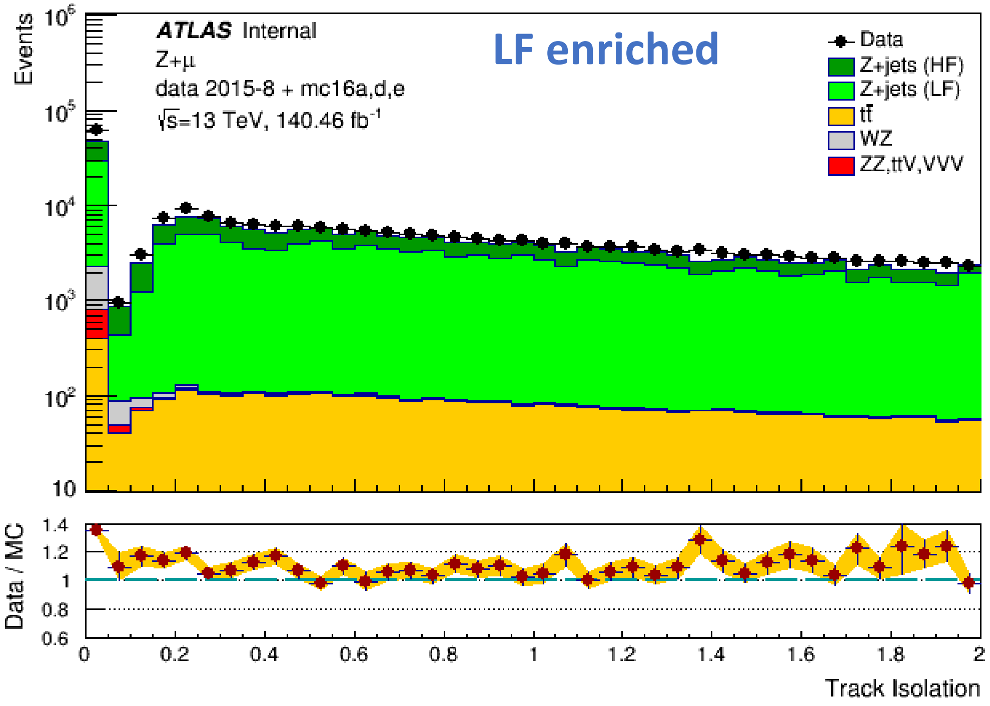
Z + μ Control Sample

Uncertainty of TF (δTF) is calculated by the relative difference between Data and MC estimation.

The main contribution to the TF uncertainty comes from the Isolation criteria

Iso efficiencies for μ	HF ± stat				LF ± stat			
	Data		MC		Data		MC	
	0.172	0.001	0.178	0.001	0.119	0.001	0.083	0.02





Control Regions Definition

4 orthogonal CRs are used for the estimation of $\ell\ell\mu\mu$ bkg

Inverted d_0 (& relaxed Isolation)

At least one lepton fails to pass d_0 significance. The isolation requirements are not applied.

This region in enriched with Z+HF and $t\bar{t}$ since leptons from heavy flavor hadrons (long lifetime) are characterized by large d_0 significance

Inverted Isolation (& pass d_0 & imbalance)

At least one lepton fails to pass the isolation requirements.

This region favors the LF component over the HF by requiring the impact parameter significance selection and the imbalance cut

eμ+μμ

An opposite-charge different-flavor leading dilepton pair satisfying the standard four-lepton analysis selection is required. For the subleading dilepton neither the d_0 significance nor the isolation is applied.

Pairing an electron with a muon excludes the real Z components, so this region is enriched with $t\bar{t}$

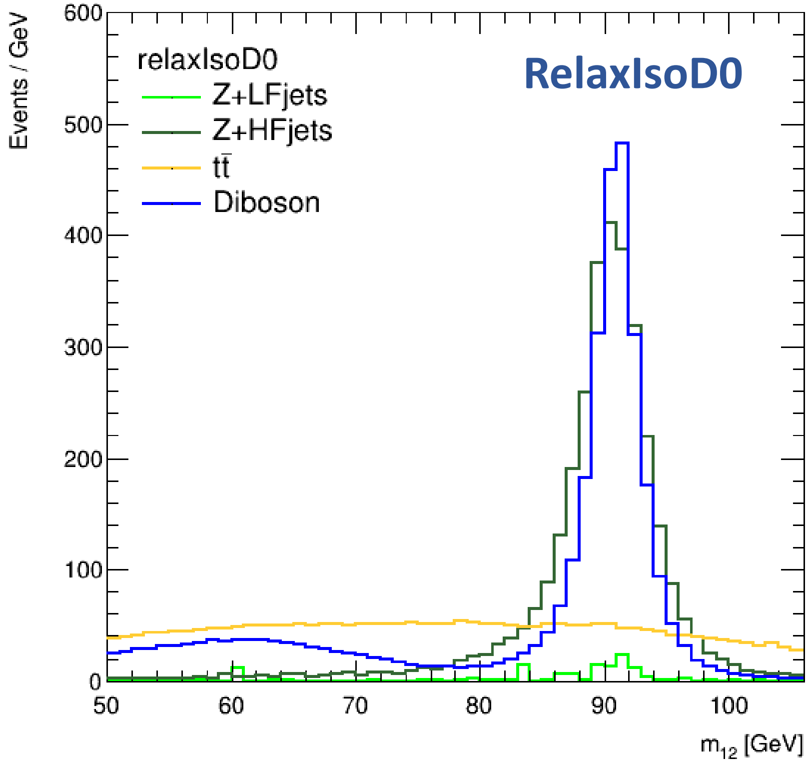
Same Sign (& relaxed d_0 & relaxed Isolation & imbalance)

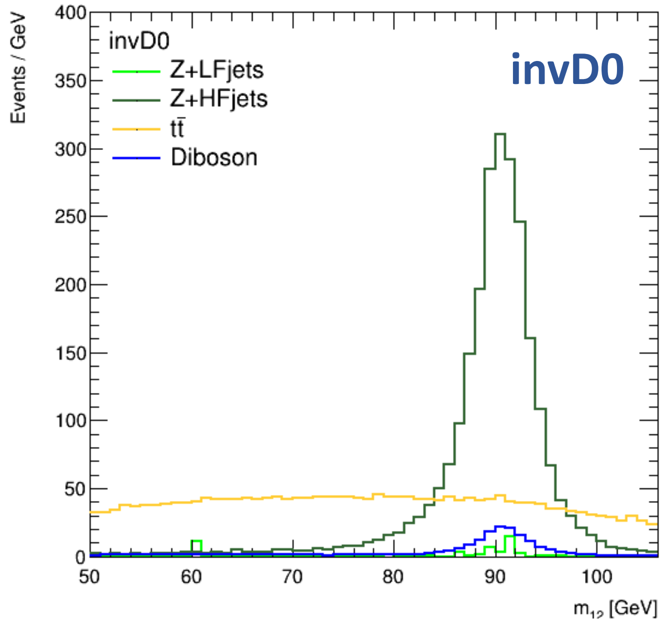
The subleading dilepton is required to be of same charge

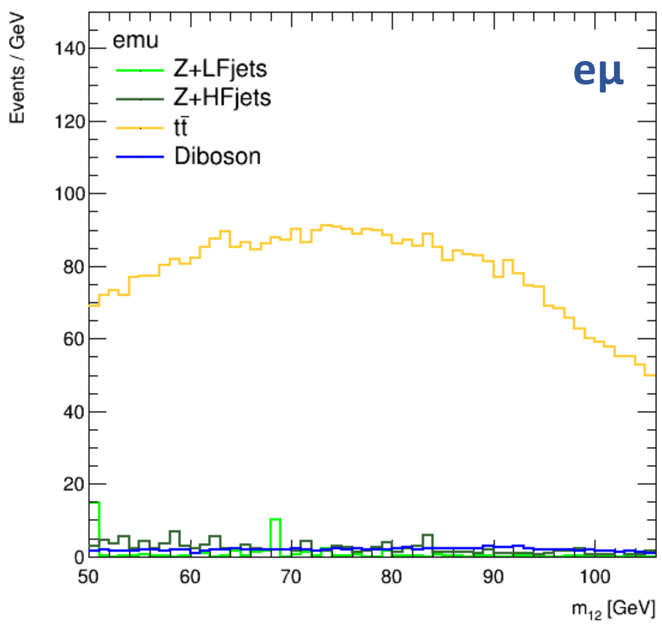
All the reducible backgrounds have significant contributions. The imbalance cut favors the LFjets over HFjets

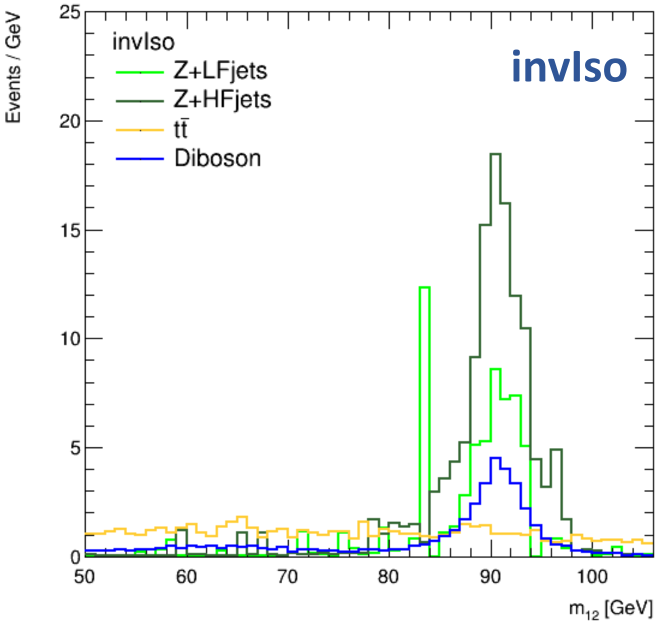
CRs MC expectations

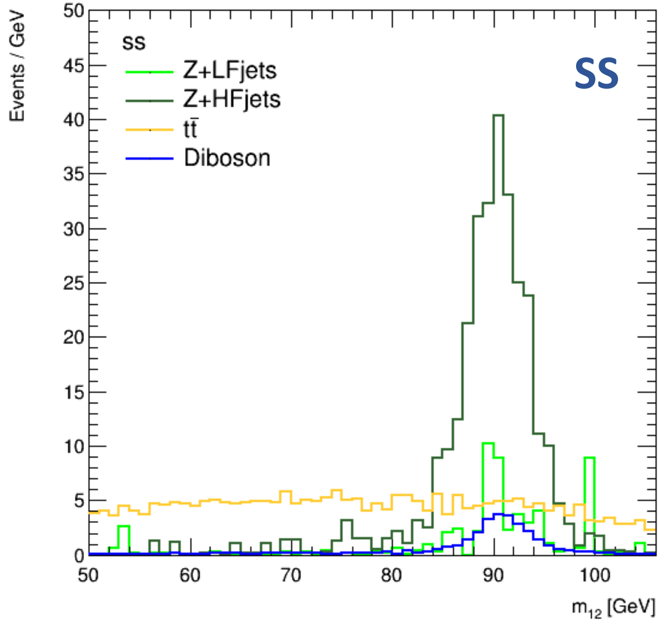
The distributions of the invariant mass of the leading dilepton (m_{12}) are used to build a model for each CR











Diboson includes the signal and all the bkgs which are not Data Driven Estimated (ZZ, WZ, VVV, ttV)

CRs Models

- The basic idea is to model the Z components **(Diboson & Z+jets)** with a **Breit-Wigner convoluted with a Crystall ball**

- And the **tt** with a **polynomial function**

- Some minor modifications to the basic model are used for the InvIso and eμ

- In every CR the functions are added to make the model (M_i) for this region

- The total model includes 3 types of parameters :
  - **Normalization factors (POI)** for each component (**α**)
  - Shape parameters for each function (**p**)
  - Fractions relating the expected events of each component in each CR to these in Relaxed ($\mathbf{f_i}$)

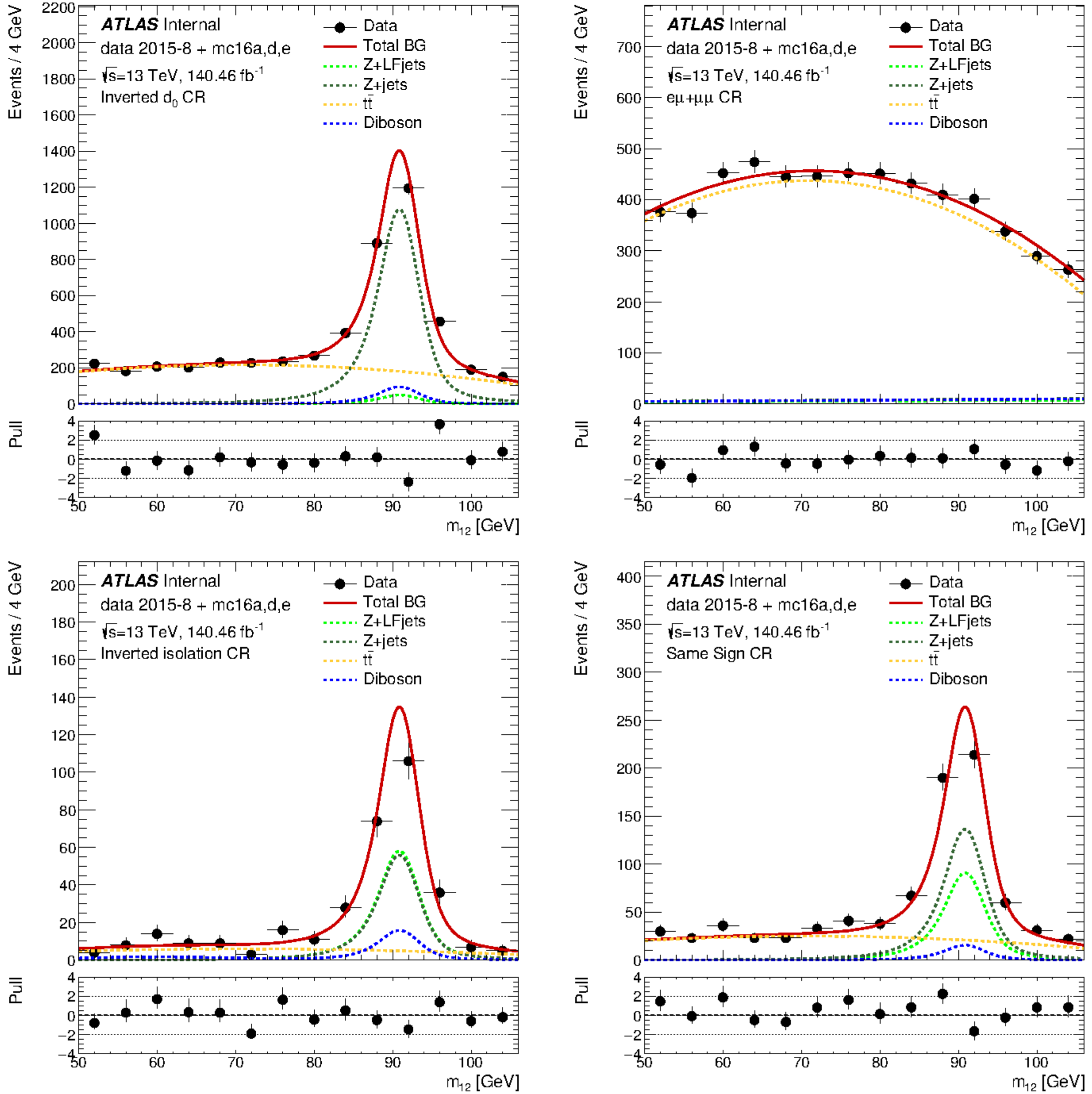
- The 4 CRs are **fitted simultaneously** by maximizing the **unbinned Likelihood,** given the real Data datasets in each CR ($\mathbf{x_i}$)

$$\prod_{i=0}^{4} L(M_i(\boldsymbol{p}, \boldsymbol{\alpha}, \boldsymbol{f_i})|\boldsymbol{x}_i) \rightarrow Maximum$$

Fitting

- Before fit the real Data a fit with MC data is carried out to initialize the parameter values and to check the model consistency
- Several tests have been performed with fixed, constrained or free nuisance parameters

The model fits the Data very well in all 4 CRs



Methods description

– One main method and two alternative for checking are used

Main method

- **Fit 4 control regions simultaneously**
  - Split light-heavy flavor Z+jets
  - Use imbalance cut in invIso & SS

2nd method

- Fit 4 control regions simultaneously
  - No Split light-heavy flavor Z+jets
  - No imbalance cut

3rd method

- 1st Fit : invD0 – eμ (HF & $t\bar{t}$ enriched)
- 2nd Fit : invIso(+pvtx) (LF enriched)
  - Fix $t\bar{t}$ at 1st Fit results
  - Fix shape parameters at 1st Fit results

Signal Region

– Events : (Fit RR) · (TF)

– Statistical error : (δFit RR stat) · (TF)

– Systematic error : (δTF MC stat) $\oplus$ (δTF) $\oplus$ (δFit RR sys)

Preliminary Results

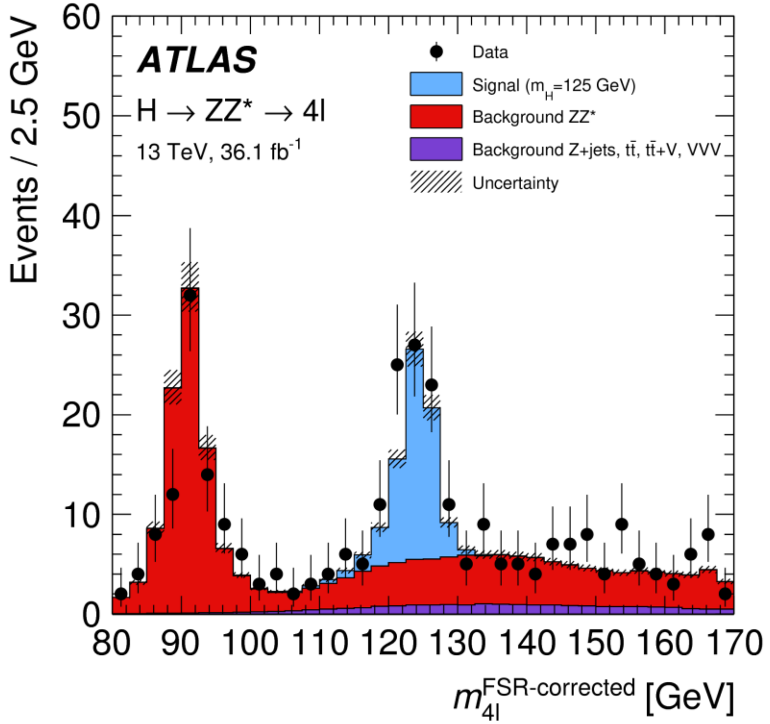
Run II MC event expectation								
	ZHF	± stat	ZLF	± stat	ttbar	± stat	Diboson	± stat
	15.01	0.39	0.88	0.83	4.77	0.33	2675.36	5.52

Run II Data Driven Estimation									
Method	ZHF	± stat	± sys	ZLF	± stat	± sys	ttbar	± stat	± sys
Main	13.26	0.50	0.92	2.45	0.58	0.25	5.78	0.08	0.55
2nd	15.37	0.34	1.07				5.79	0.07	0.55
3rd	13.63	0.40	0.95	2.88	0.61	0.29	5.83	0.08	0.55

The Z+jets estimation is not very different between DDE and MC but the LF & HF contributions are quite different

The tt increases about 15% wrt the MC expectation

The 3 methods give similar results.



JHEP 10 (2017) 132

The reducible Bkg is much less than the irreducible but an important part is below the Higgs peak

Differential Estimation

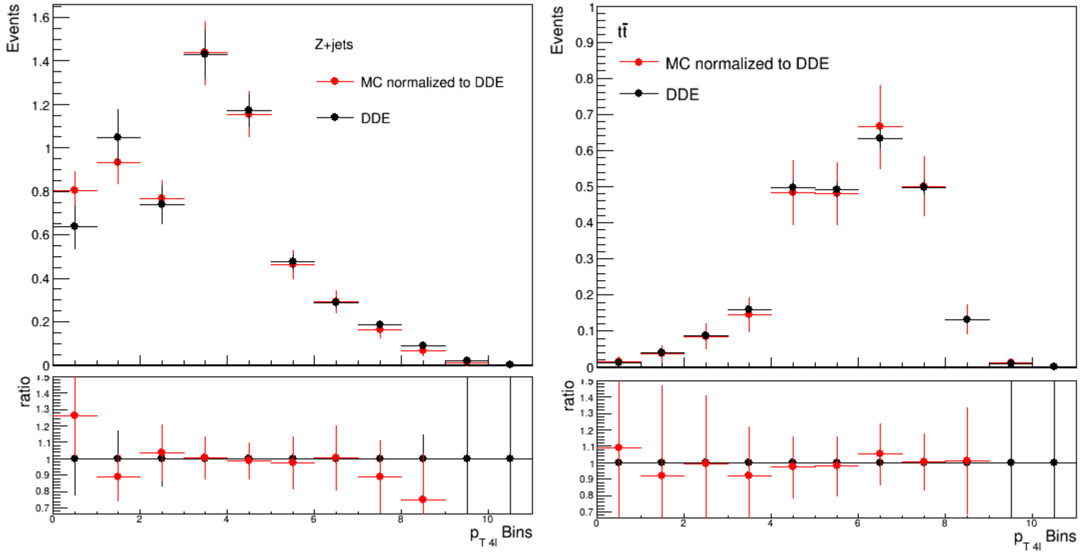
- The bkg Data Driven Estimation is used for all the H -> 4l analyses, including **Differential XS**, **Couplings measurements** , **High Mass Higgs searches** and **Decay Width limits**

- Until now the total estimation was used and the separation into subcategories was done using the MC
  Now we have enough statistics to carry out a **fit for each subcategory**

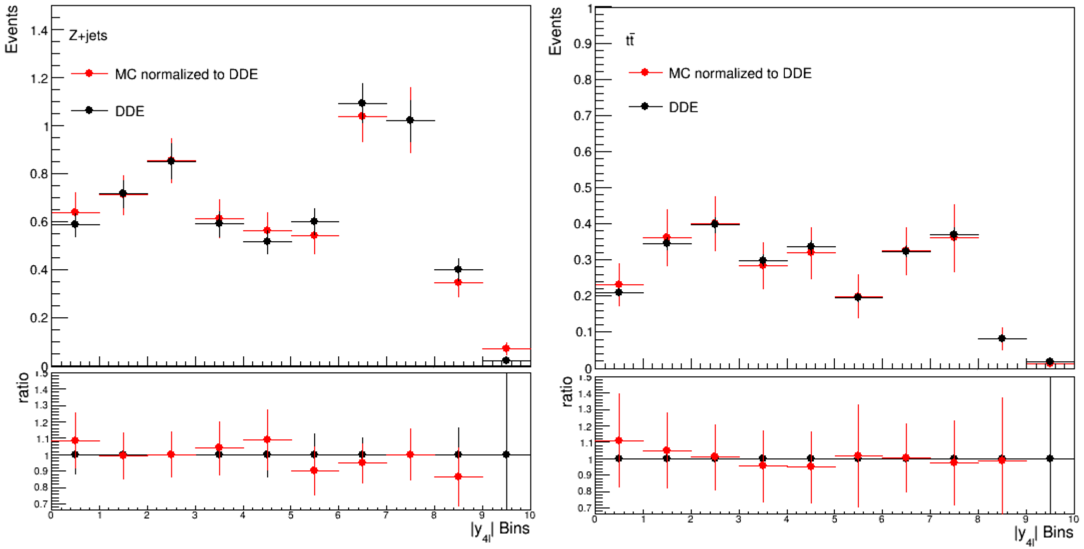
- For the differential XS the categories are divided by N_{jets}, cosθ*, p_{T4l}, Leading jet p_T , y_{4l}
  For Couplings analysis the categories are defined according to Higgs production modes
  For High mass and Decay Width an estimation for m_{4l}>180 is required

- The bkg estimation is also used for the study and choice of new Isolation criteria

Preliminary results



Preliminary results



Summary

The full framework for the $\ell\ell\mu\mu$ reducible bkg estimation has been developed

We have enough statistics to fit subcategories needed by various analyses separately

Analyses with 2015-16 data have been already published

Some selection criteria have to be finalized in order to proceed to the final estimation for full Run II Data (2015-18)

The full Run II preliminary results are compatible with the 2015-16 results