Towards the measurement of associated production of vector bosons and ttbar

in the dilepton channel within the ATLAS experiment at $\sqrt{s}$ = 8 TeV

Tamara Vázquez Schröder
Supervisors: Kevin Kröninger, Arnulf Quadt, Elizaveta Shabalina

2nd Institute Of Physics, Georg-August-Universität Göttingen

7th Annual Helmholtz Alliance Workshop

PHYSICS AT THE TERA SCALE
Helmholtz Alliance

Towards the measurement of Tt+V (V=Z,W) cross-section in the dilepton channel @ 8 TeV

- Introduction
- Regions and optimization
- Multivariate Analysis
  - discriminating variables
- The Fit
  - MCLimit and RooStat
- Fit Control Region1
- Conclusions and next steps

1 Introduction

Electroweak couplings involving top quarks

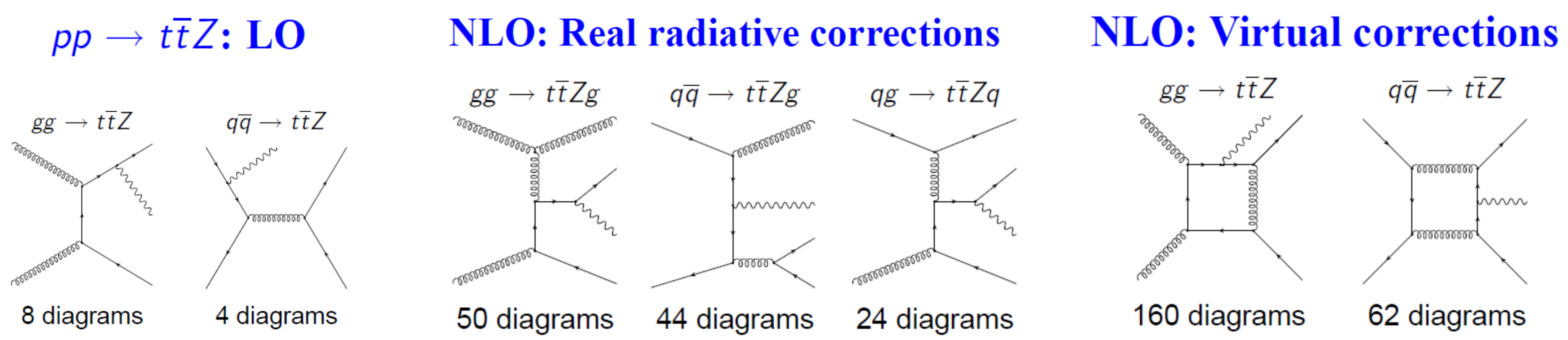
- EW properties from top largely unknown (coupling ttγ, ttZ, ttH…)
- Important Standard Model test: new physics modifies the structure of the electroweak couplings (e.g. Z ′, T quark, . . .)

Cross-section of TtZ production sensitive to anomalous couplings!

Ttbar+V coupling challenge

- ttbar production via intermediate V very difficult to measure at the LHC (small correction to a QCD dominated process)
- Instead, measure cross sections: TtZ, TtW production

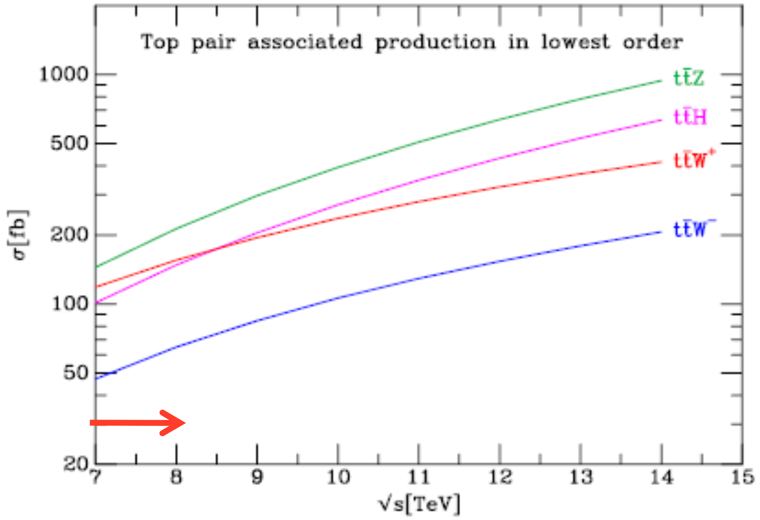
LHC measurements are crucial!



Measure combined TtZ and TtW cross section @ 8 TeV in the dilepton (opposite-sign) channel with full dataset 20 ifb!

- **Why 8 TeV?**

TtZ Cross sections:
@ 7 TeV (139 fb) , @ 8 TeV (**208 fb**)



10.1007/JHEP07(2012)052

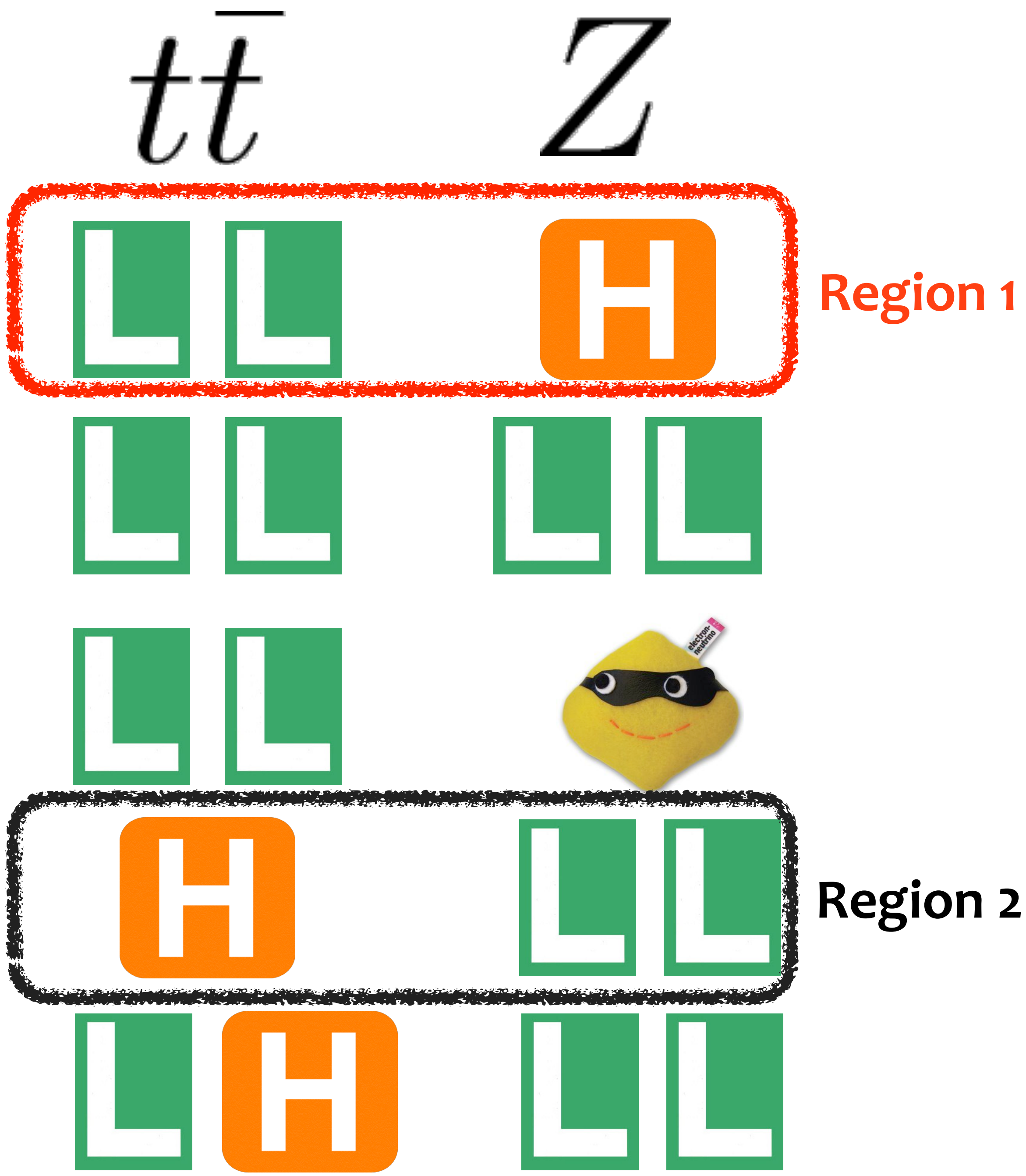
- **Why dilepton opposite sign channel?**

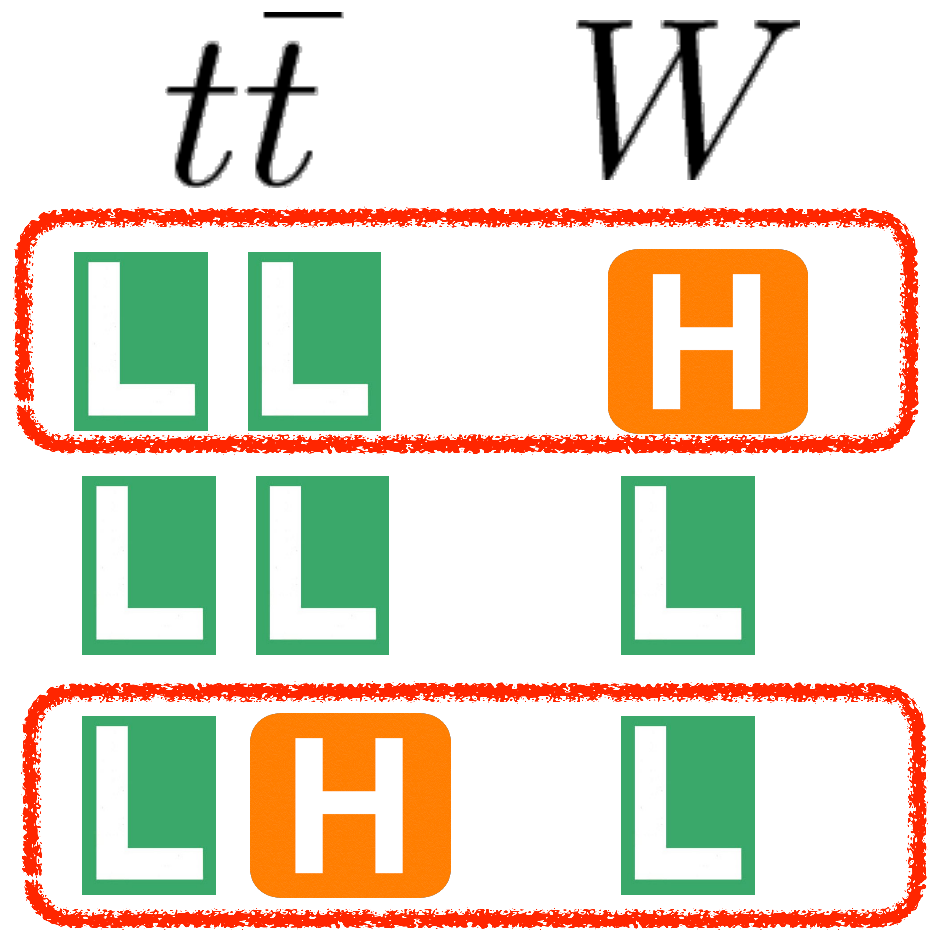
  - higher branching ratio than for trilepton final state
  - high background: no chance for cut&count experiment, go for Multivariate Techniques (MVA)!
  - possible kinematic reconstruction of ttbar ➛dilepton system using Kinematic Likelihood Fitter (KLFitter).

Keep selection orthogonal to trilepton final state (veto on 3rd lepton)

2 Regions and optimization

Truth TtV dilepton channels





Region 1 $|m_{ll} - 91| > 10$ GeV

Region 2 $|m_{ll} - 91| < 10$ GeV

TtV dilepton regions

	Region 1	Region 2
Z-mass window cut	$\|m_{ll} - 91\| > 10$	$\|m_{ll} - 91\| < 10$
Dilepton channels	all	ee and mumu
Signal	hard to distinguish ttZ from ttW	ttZ dominated
	Madgraph+Pythia	
Main background	tt+jets *(Alpgen+Herwig → Powheg+Pythia)*	Z+jets *(Alpgen+Herwig)*
Reconstruction	KLFitter dilepton + extra LH comp for extra jets from Z/W **reconstruct Ttbar+Z(W) system**	also possible, but not considered in this analysis

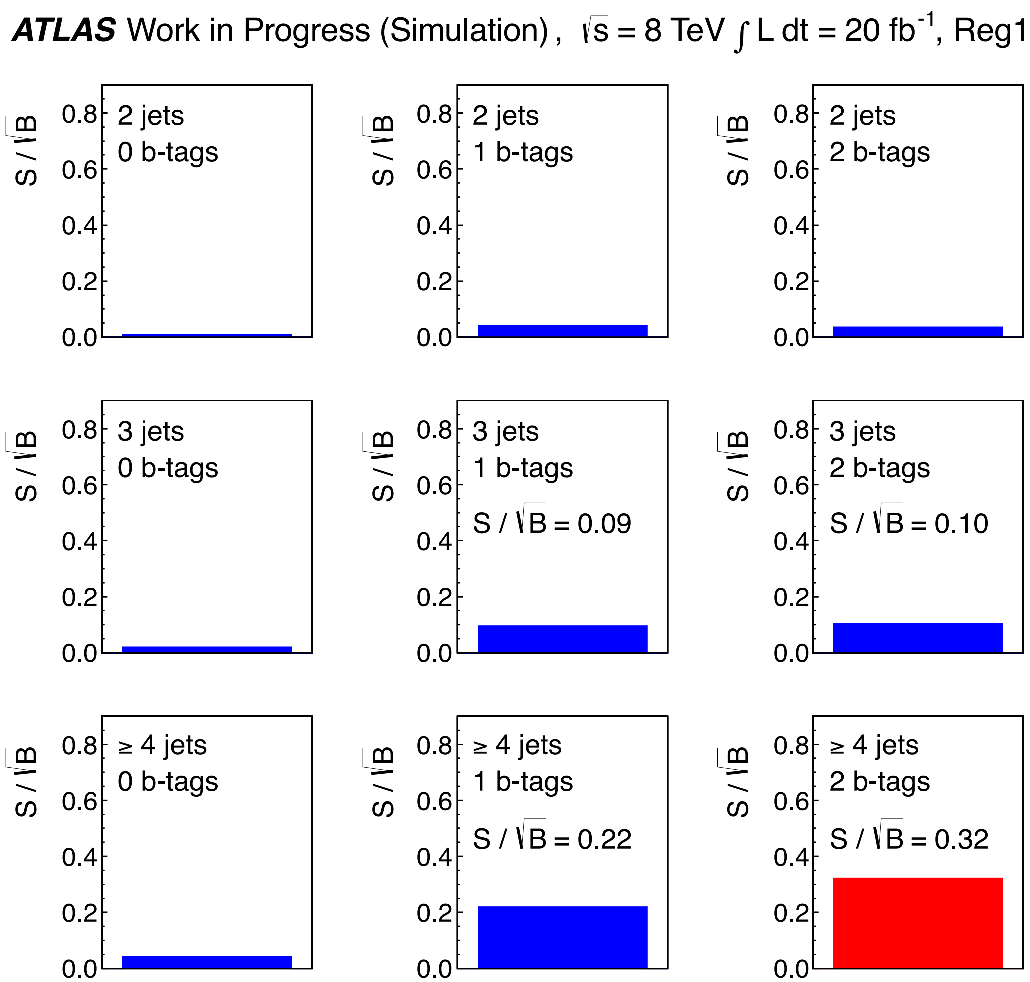
Use as loose lepton definition and objects selection as possible to maximize signal

- **Exactly 2 opposite-sign leptons**

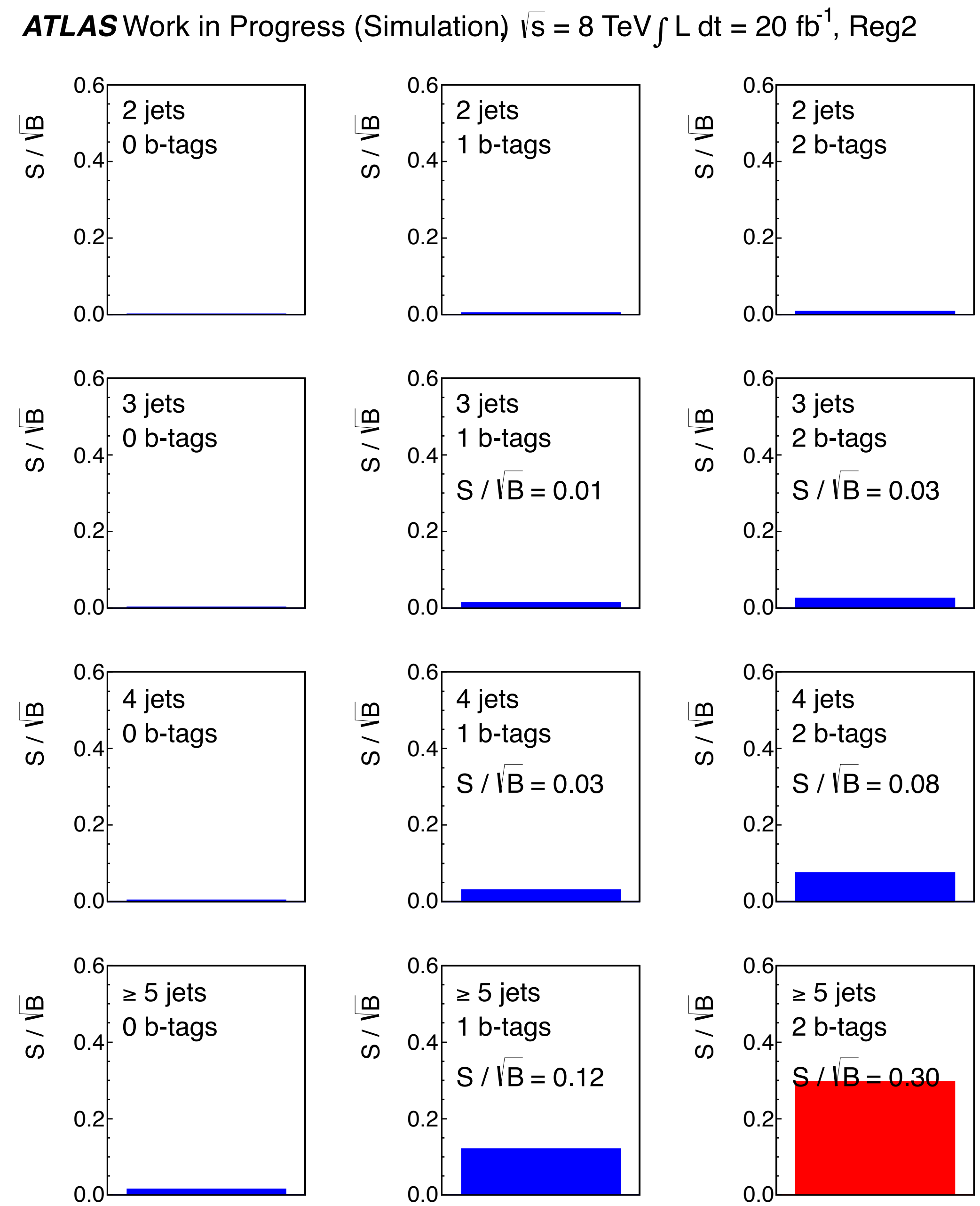
ATLAS work in progress

	standard	optimized	S/√B gain (in 3jincl 1bincl)	channels
MET cut	60 GeV	-	~ 4 %	ee, mumu
jet pT cut	25 GeV	25 GeV (1st and 2nd leading), rest 20 GeV	(only shift of signal events towards higher jet bins: possible ttbar dilepton reconstruction)	
lepton pT cut	25 GeV	15 GeV	~ 12 %	all
electron ID	tight++	mediumLH	~ 2 %	ee, emu
	iso90	PtCone30/Pt < 0.12	~ 9 %	
btagging WP	70%	80%	(shift of signal events towards higher btagged jet bins e.g.: 1bjet excl → 2bjet excl)	

S/√B across jet/btagged jet bins



Threshold: S/√B = 0.25



maximum of **2 btagged jets** considered:
avoid getting into the region dominated by ttbar+HF

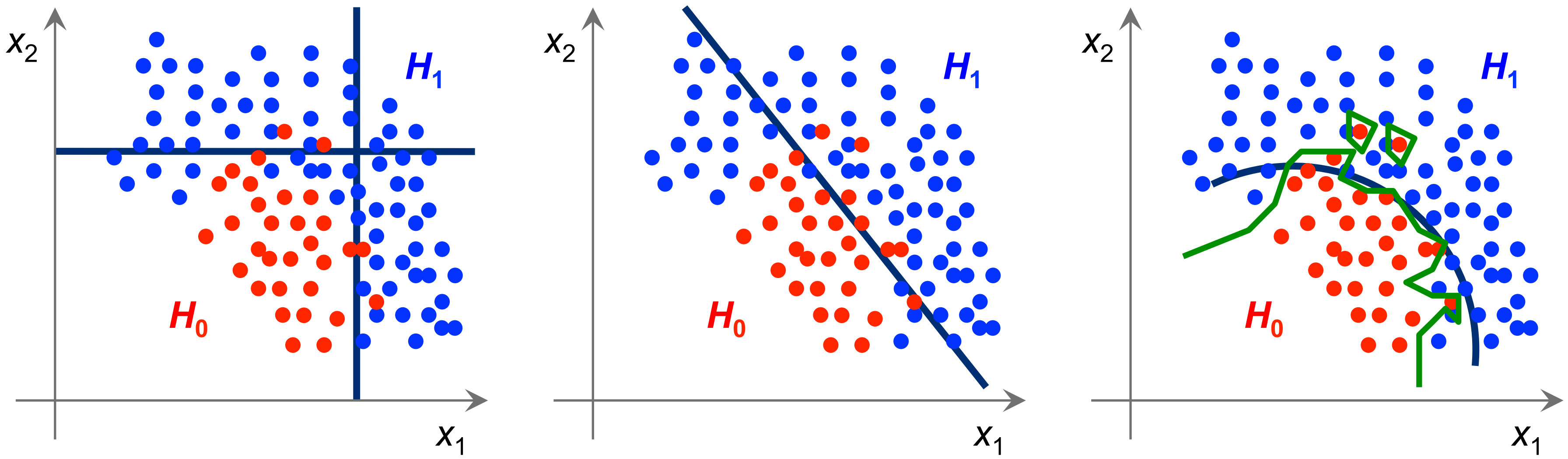
Region 1 $|m_{ll} - 91| > 10$ GeV

Region 2 $|m_{ll} - 91| < 10$ GeV

3 Multivariate analysis

Classifiers in Multivariate analysis

Suppose a data sample with two types of events: H_0, H_1

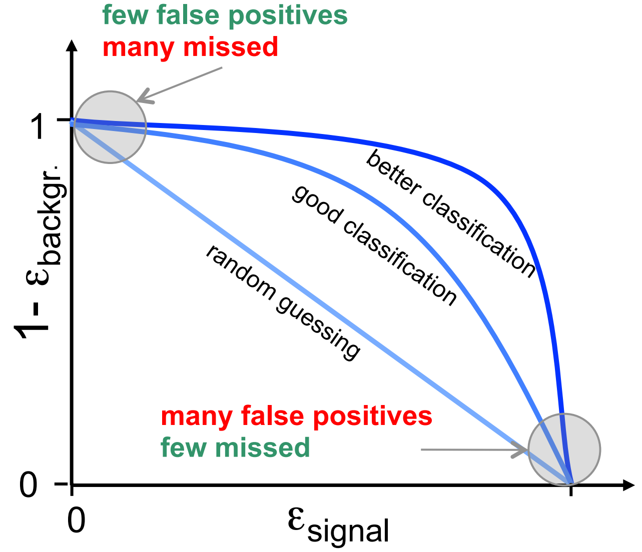


We have found discriminating input variables x_1, x_2, …

What decision boundary should we use to select events of type H_1 ?

How can we decide this in an optimal way ? **Let the machine learn it !**

"Receiver Operation Characteristics" (ROC) curve

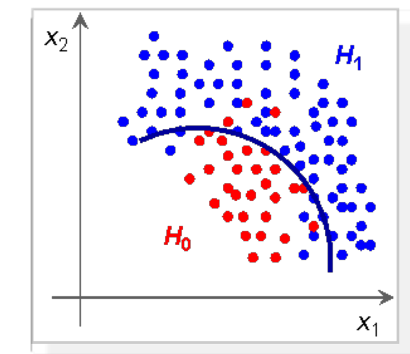


material from: Peter Speckmayer, Multivariate Data Analysis with TMVA

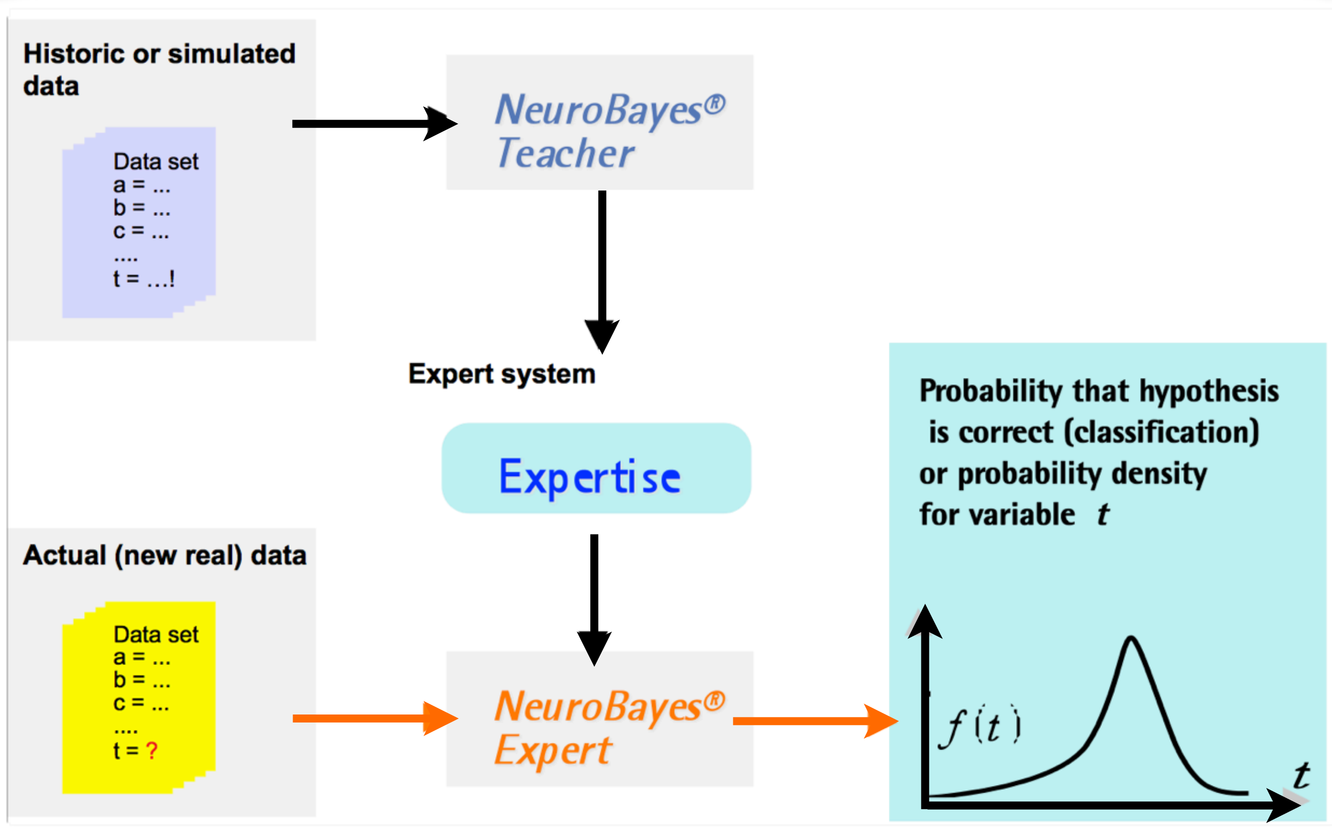
Neural Networks (NN)

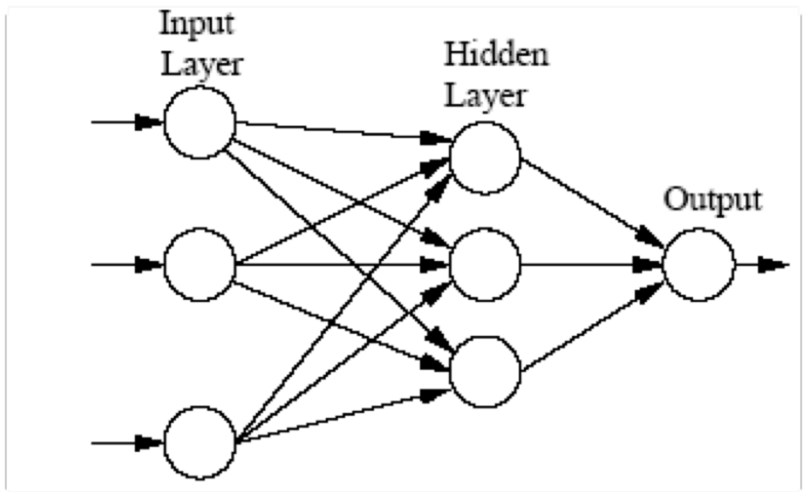
The information (the knowledge, the expertise) is coded in the connections between the neurons

- robust against overtraining, fast response
- takes into account correlations between variables
- sensitive to weak variables



NeuroBayes®: training and application





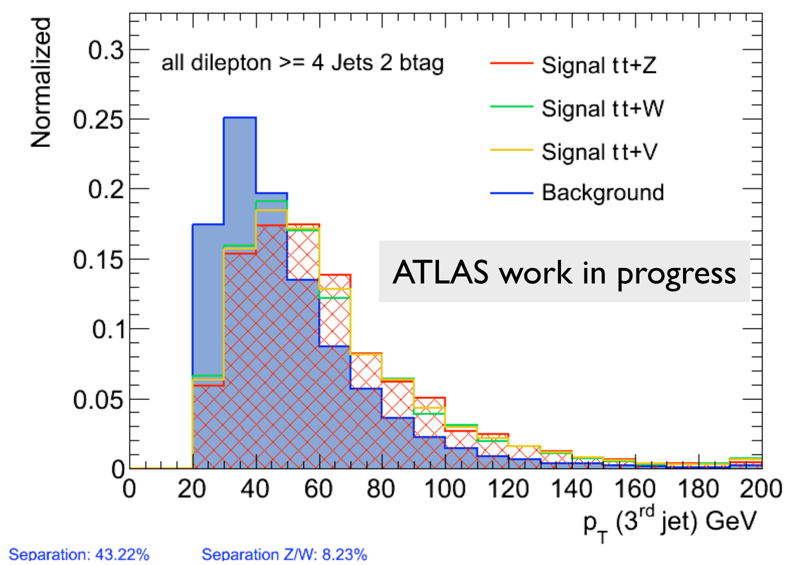
The output of node j in layer n is calculated from weighted sum of outputs in layer $n-1$:

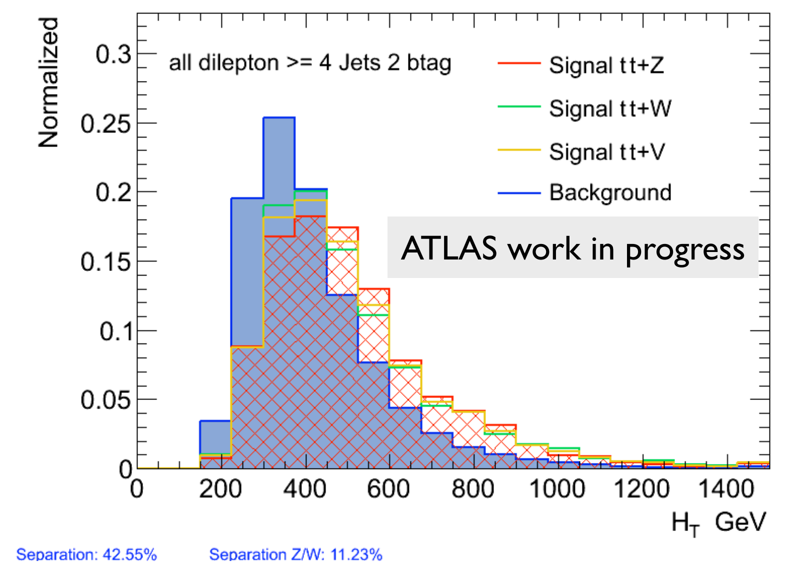
$$x_j^{(n)} = f(\sum_i w_{i,j}^{(n)} x_i^{(n-1)} + w_{0,j}^{(n)})$$

Each connection has associated a weight $w_{i,j}^{(n)}$, each node a bias $w_{0,j}^{(n)}$.

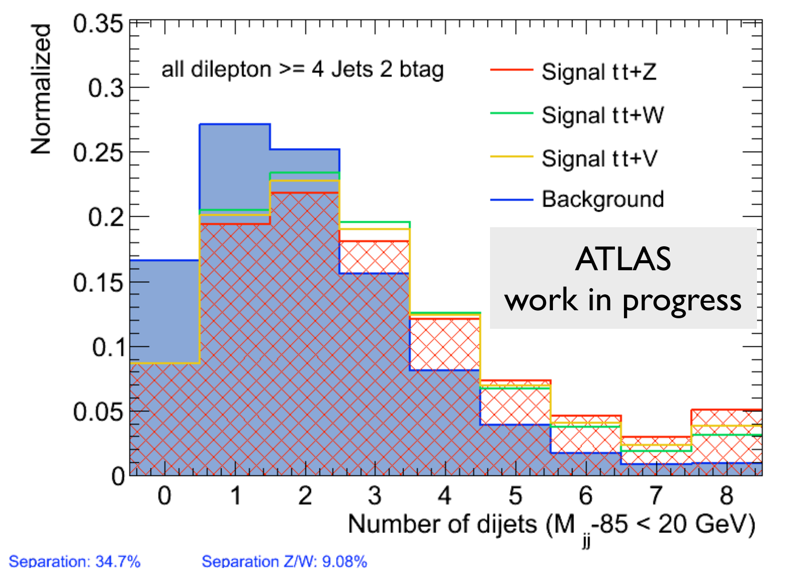
material from: Michael Feindt Neural Networks and NeuroBayes School of Statistics 2010

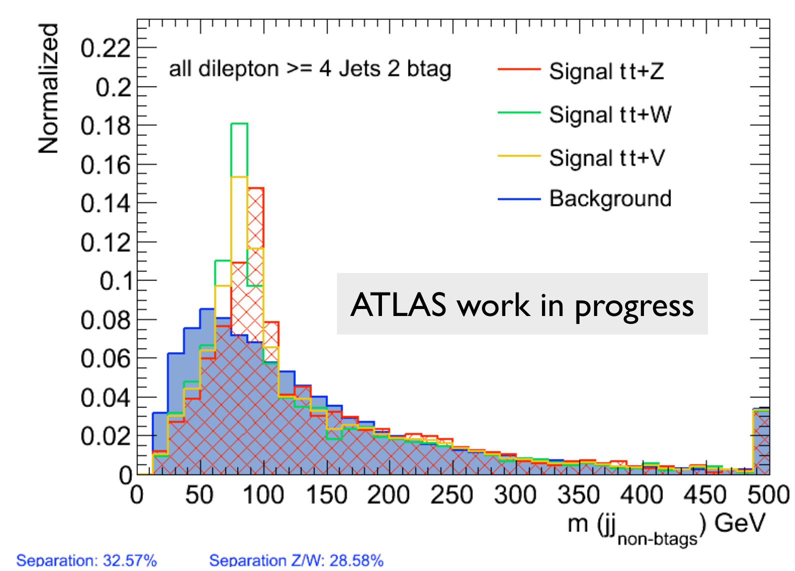
Discriminating variables Reg1



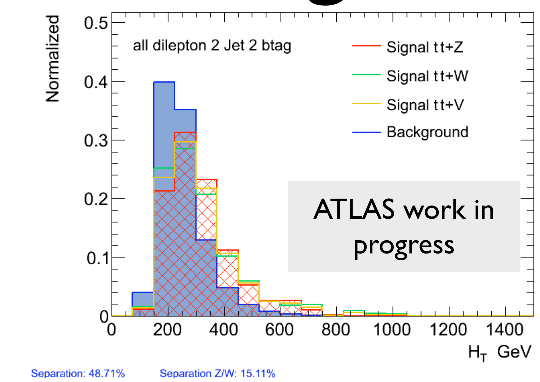


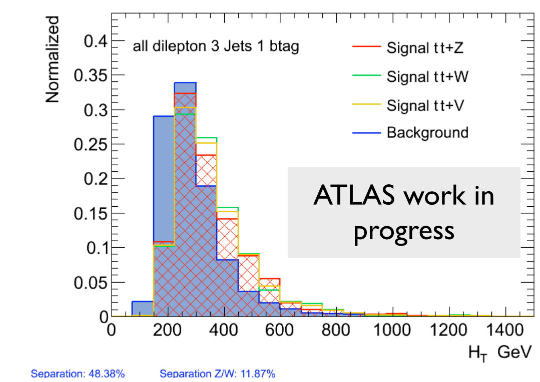
Signal region: 4jin 2bjex

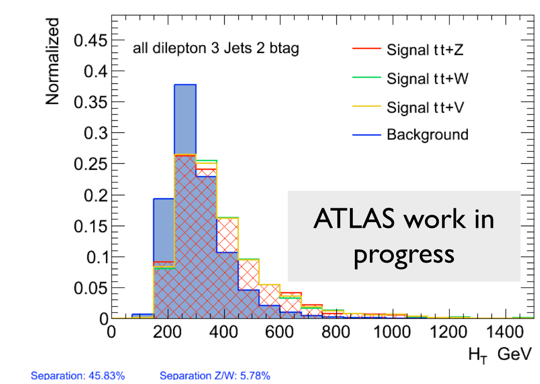


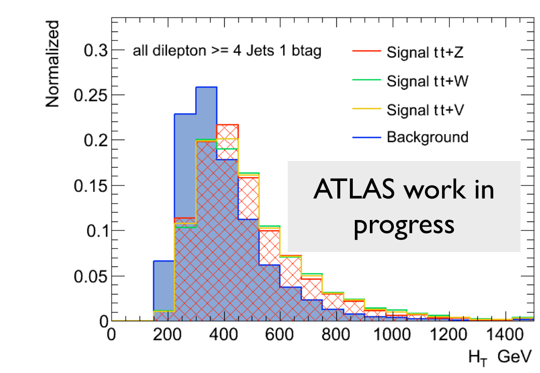


Control regions: HT









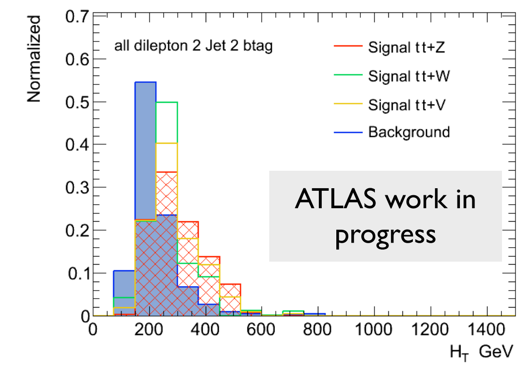
First step towards **NN construction**: check variables with high separation power between signal (ttV) and background

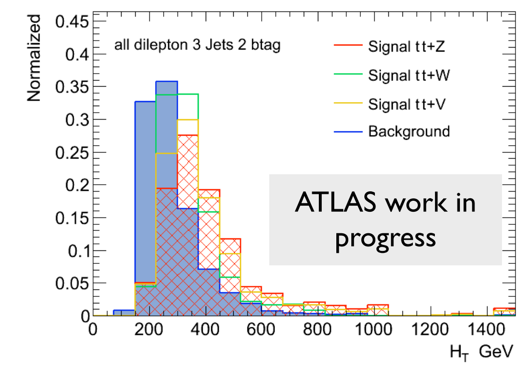
Region 1 $|m_{ll} - 91| > 10$ GeV

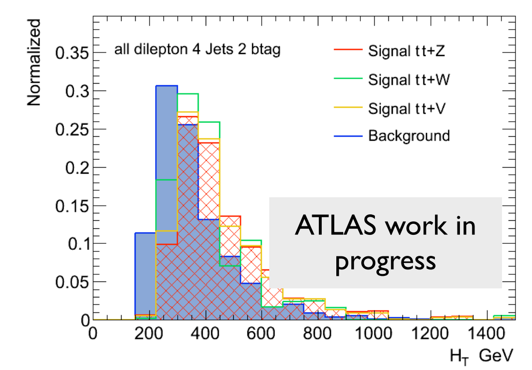
Separation defined as the non-overlapped area between signal ttV and background

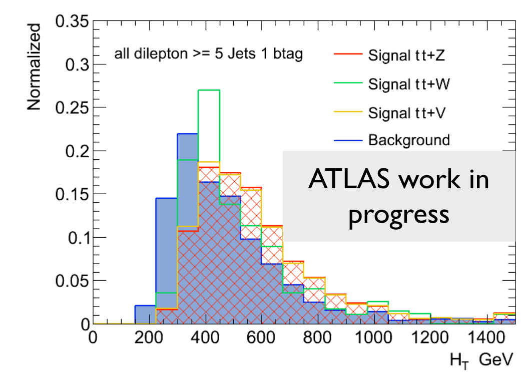
Discriminating variables Reg2

Control regions: HT

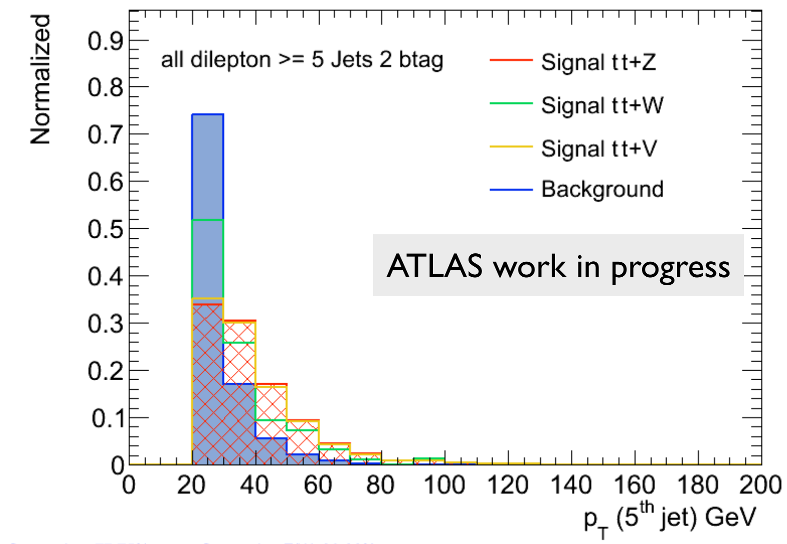


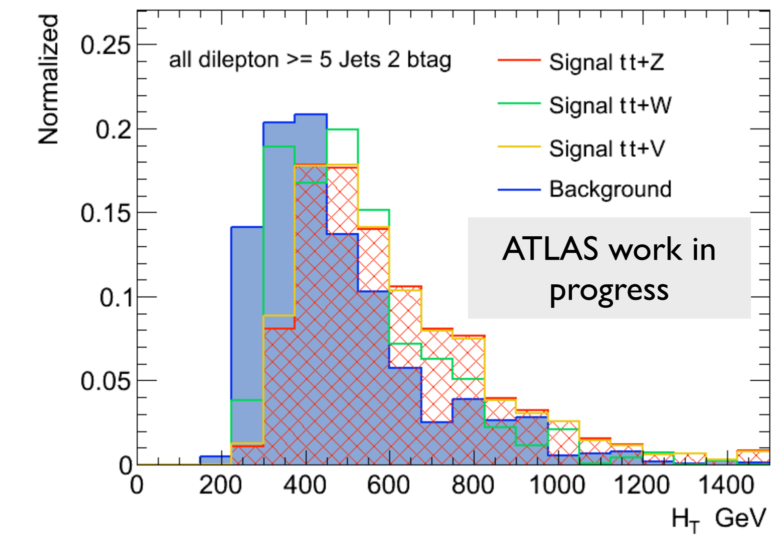


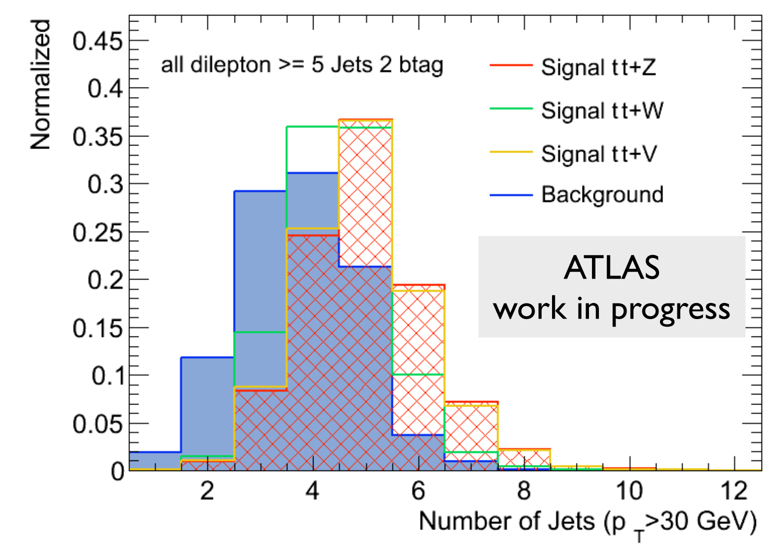


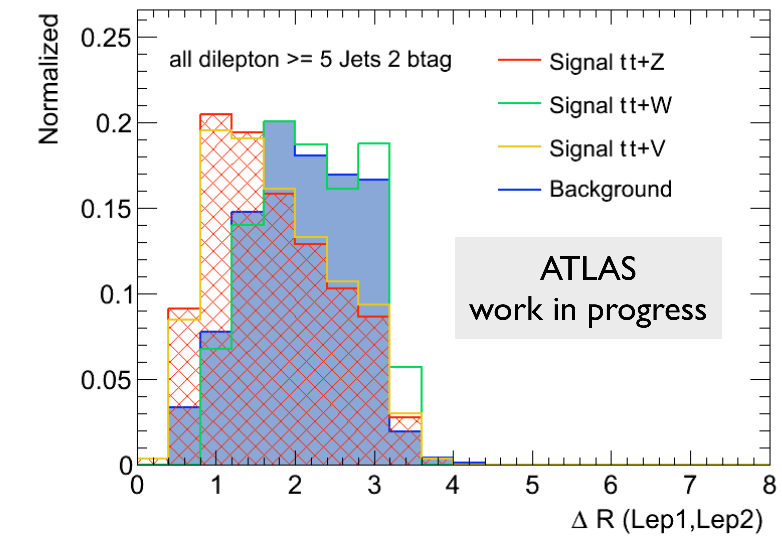


Signal region: 5jin 2bjex









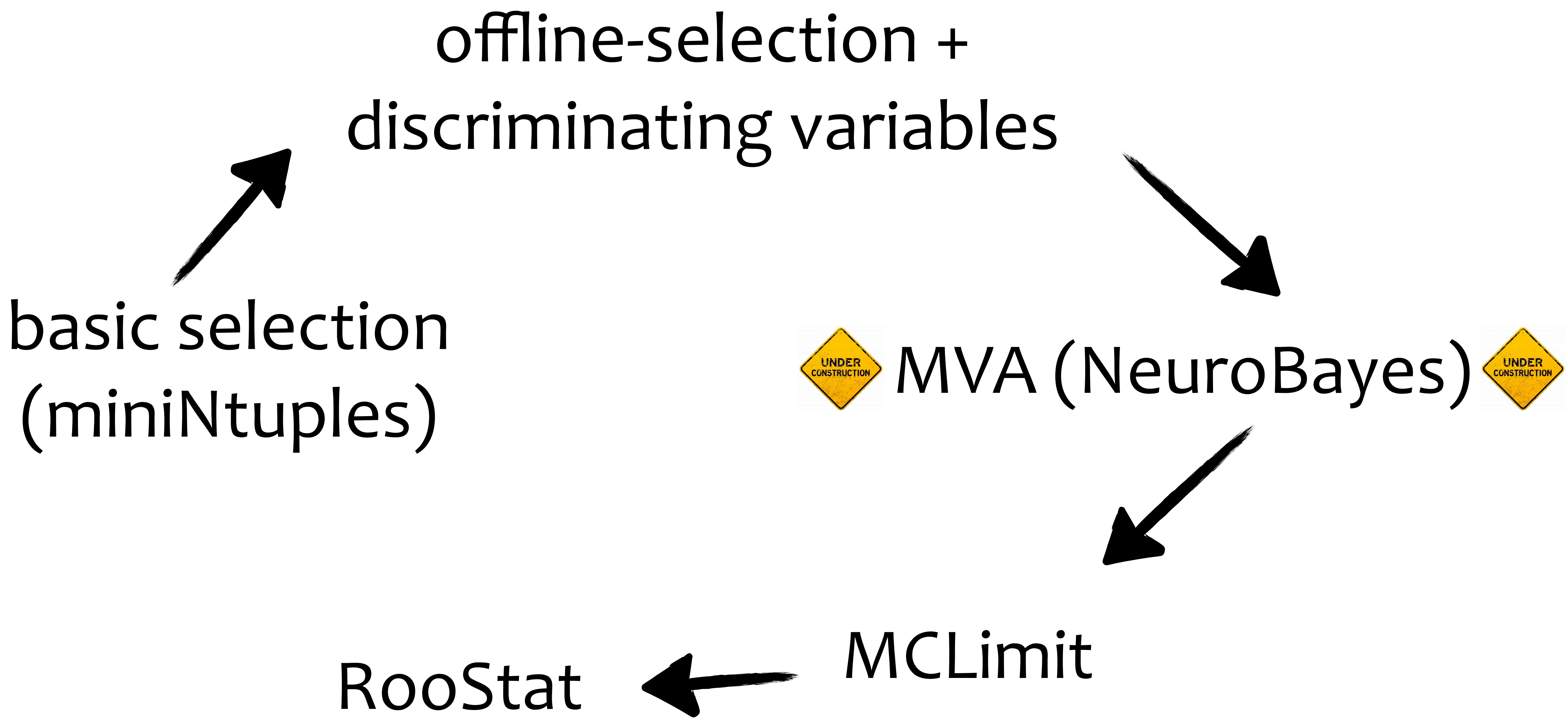
First step towards **NN construction**: check variables with high separation power between signal (ttV) and background

Region 2 $|m_{ll} - 91| < 10$ GeV

Separation defined as the non-overlapped area between signal ttV and background

4 The Fit

The analysis chain



The analysis chain - first test

basic selection
(miniNtuples)

offline-selection +
discriminating variables

HT distribution

MCLimit

RooStat

The Fit: MCLimit and RooStat

MCLimit fitting code: is an hybrid bayesian-frequentist limit setting tool

- **in our analysis** only used for **smoothing** of the nuisance parameters (merge bins with very low stat)

RooStat: is a project to create statistical tools built on top of RooFit and distributed in ROOT

Model the distribution of observables $\vec{x}$ in terms of
* physical parameters of interest $\vec{p}$
* other parameters $\vec{q}$ to describe detector effects

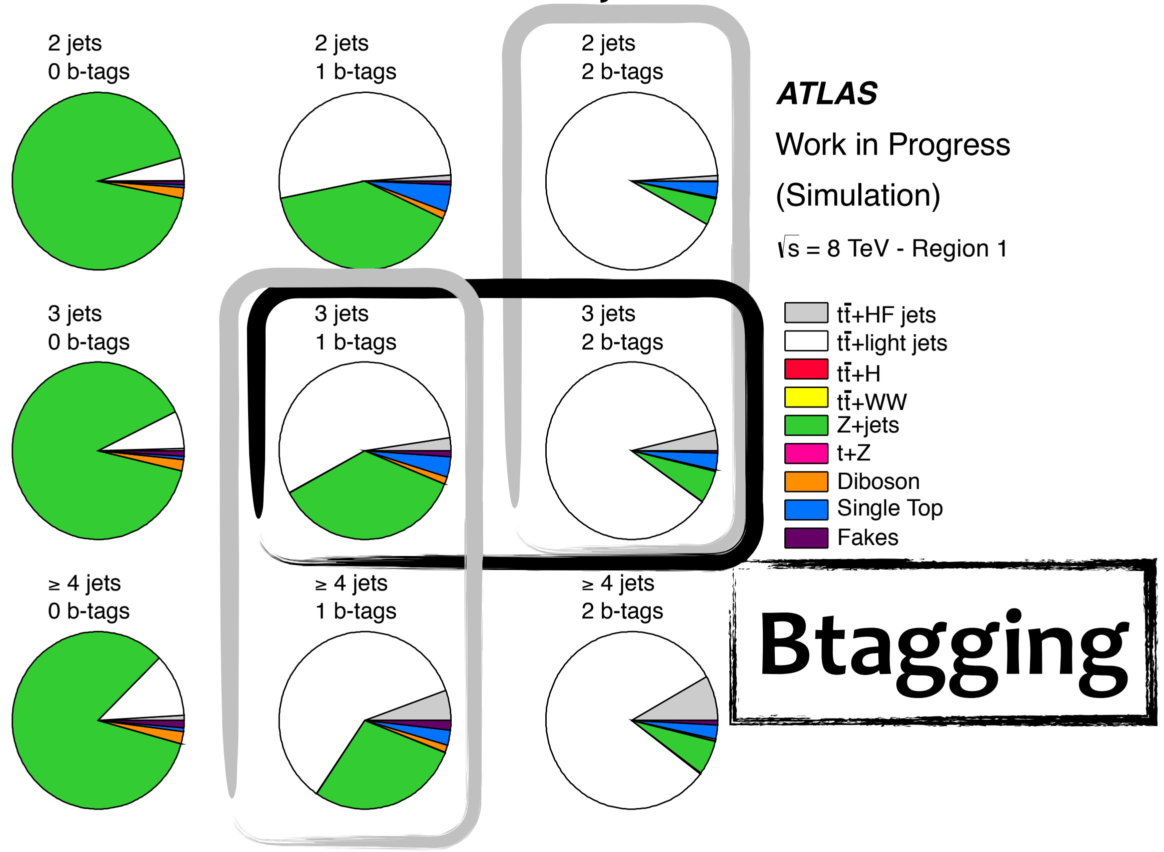
↓ RooFit

Probability density function $F(\vec{x};\vec{p},\vec{q})$

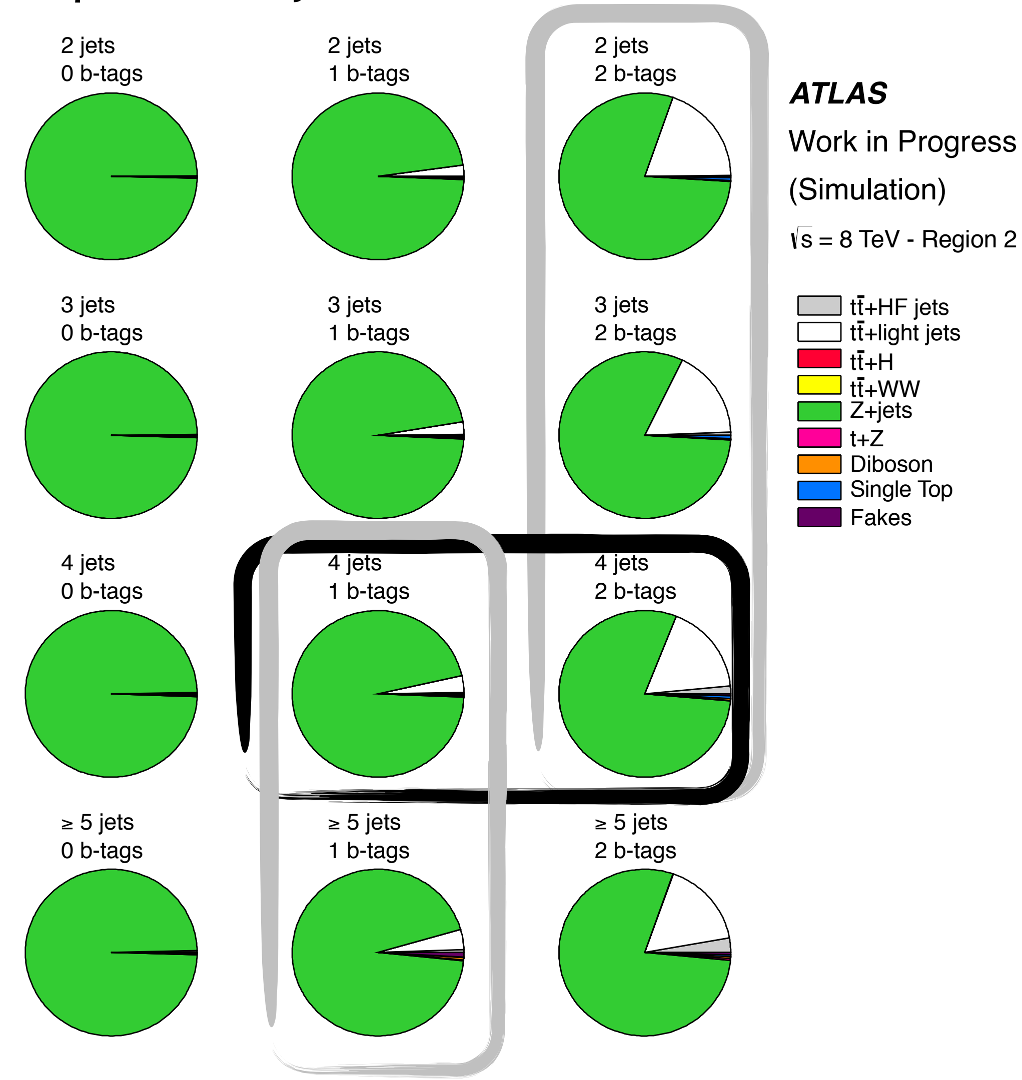
Divide and conquer!

Control regions:

- have negligible signal contribution and are included in the fit to constrain systematics
- select them wisely in order to have a handle on all possible systematic sources



JES/
ttbar normalization
per jet multiplicity bin



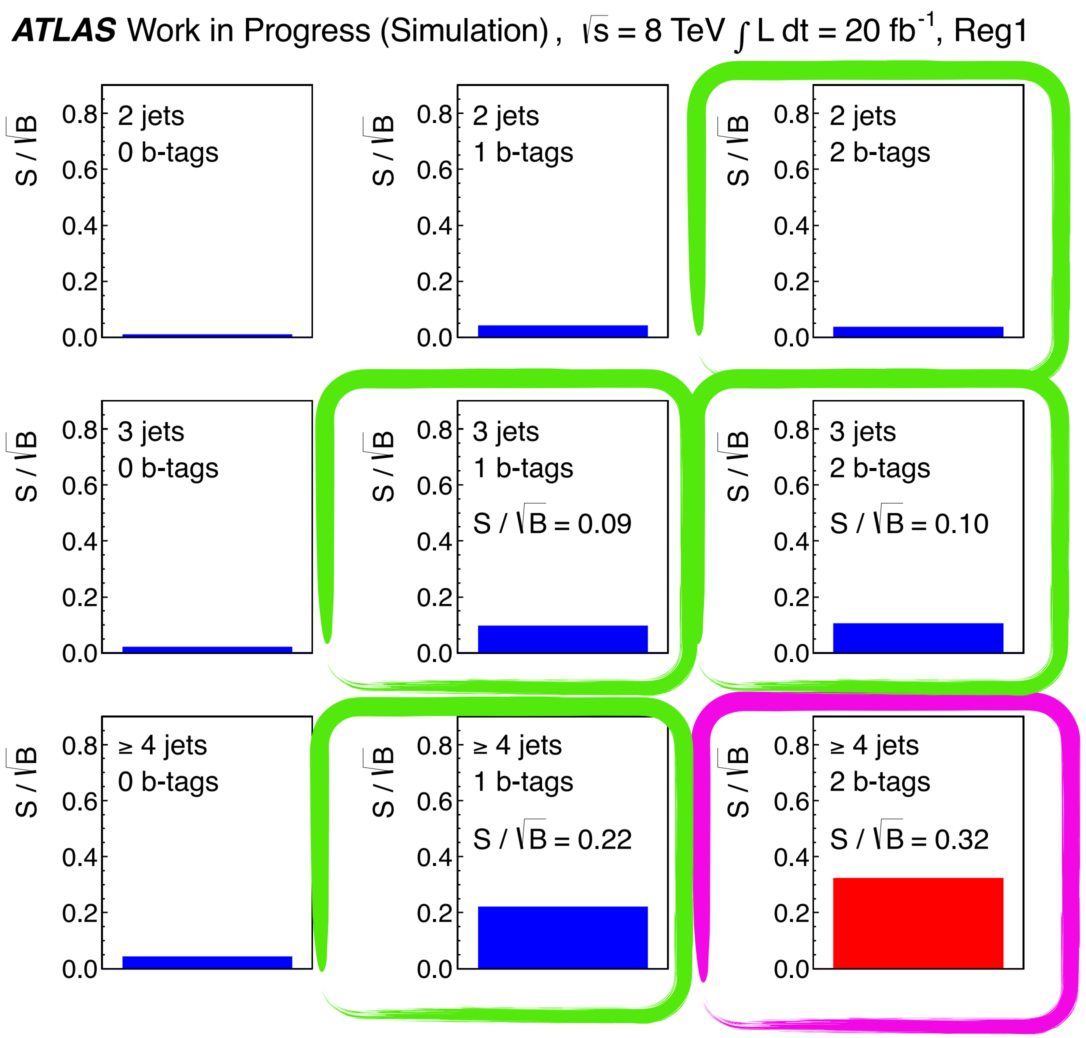
since we don't go beyond 2 bjets, no need to have a handle on ttHF fraction

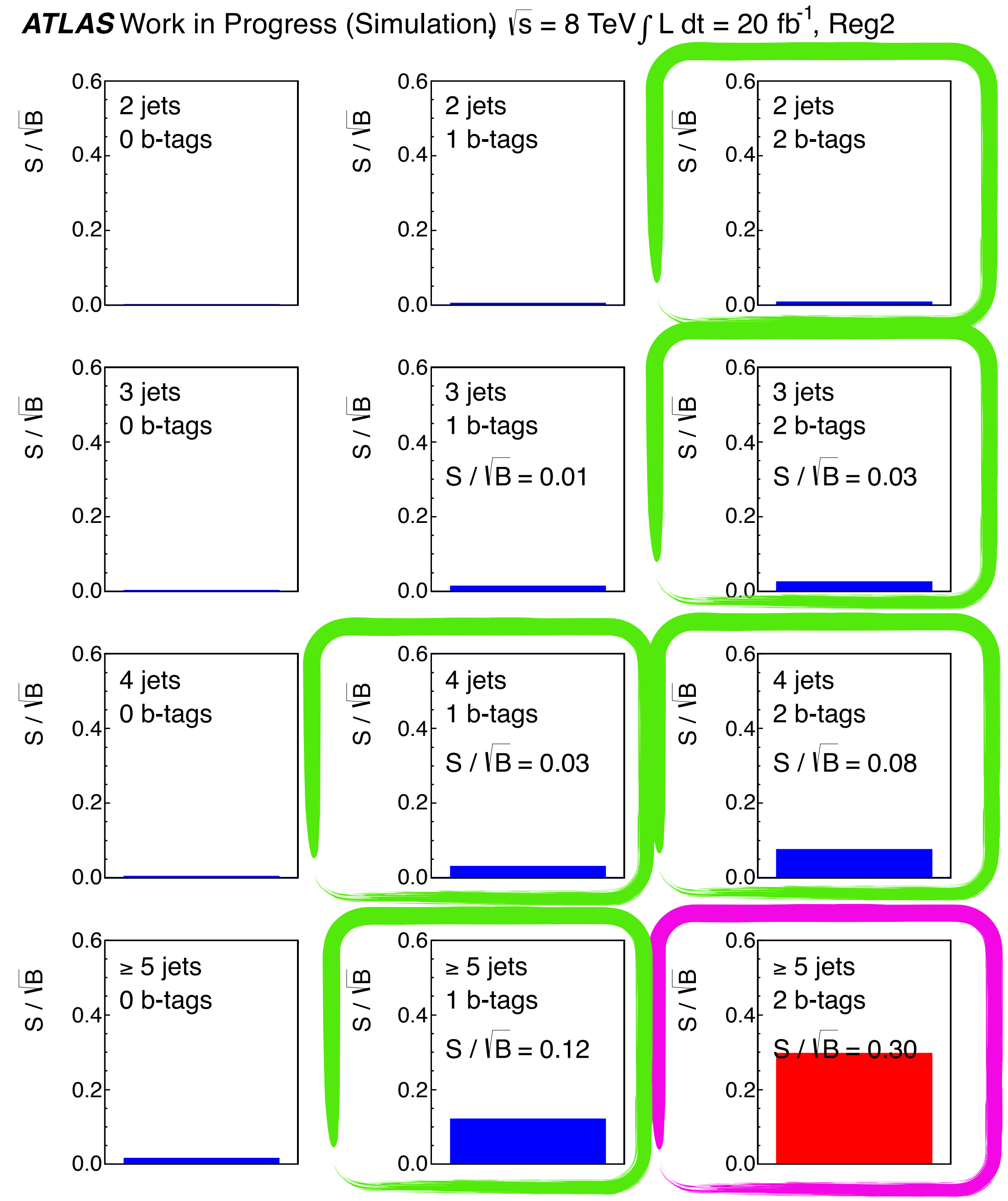
Region 1 $|m_{ll} - 91| > 10$ GeV

Region 2 $|m_{ll} - 91| < 10$ GeV

Divide and conquer!

Select control regions wisely in order to have a handle on all possible systematic sources





CONTROL REGION

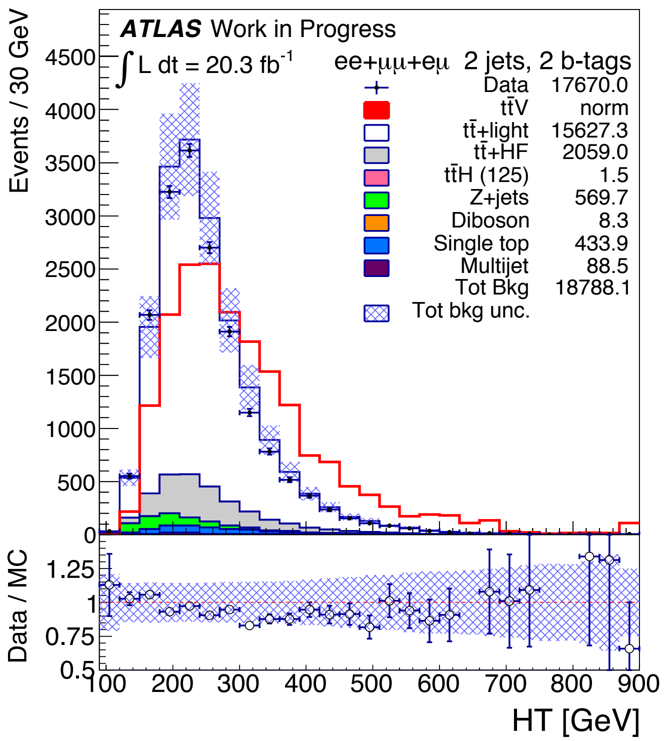
SIGNAL REGION

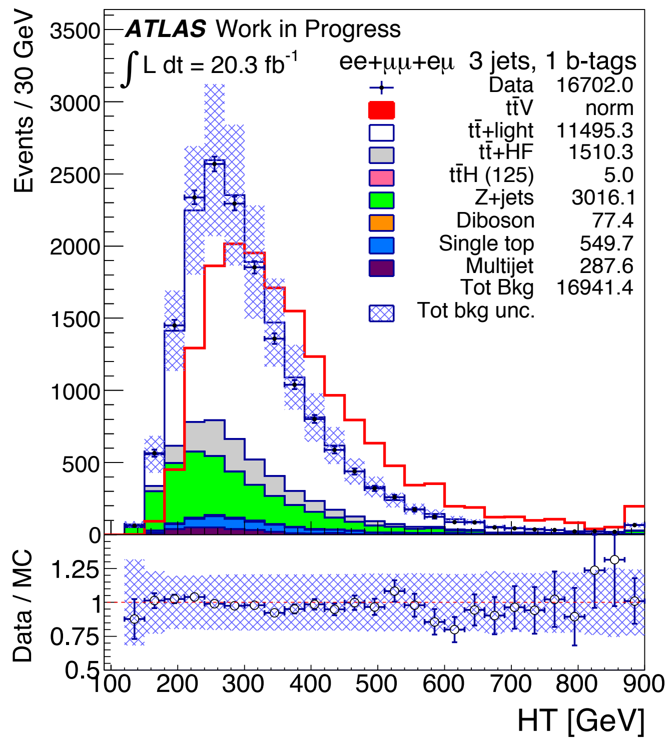
Region 1 $|m_{ll} - 91| > 10$ GeV

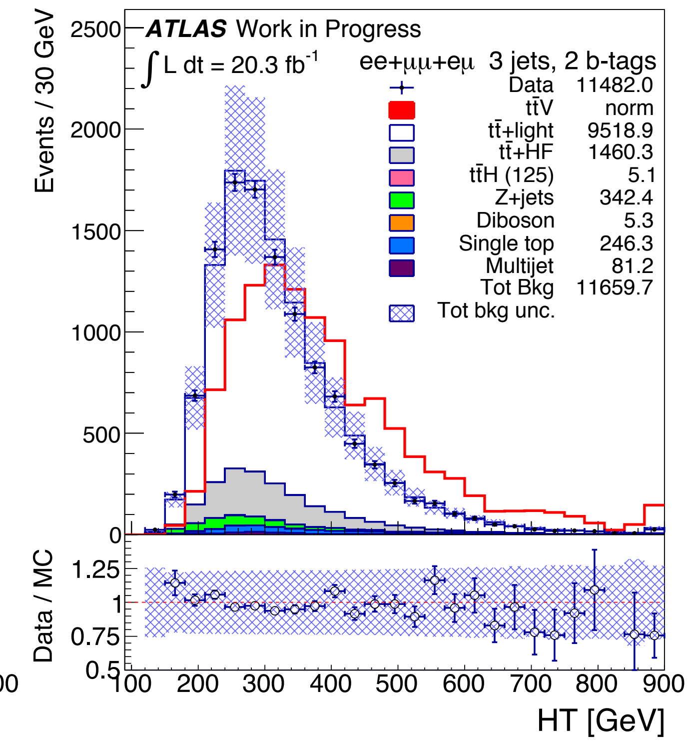
Region 2 $|m_{ll} - 91| < 10$ GeV

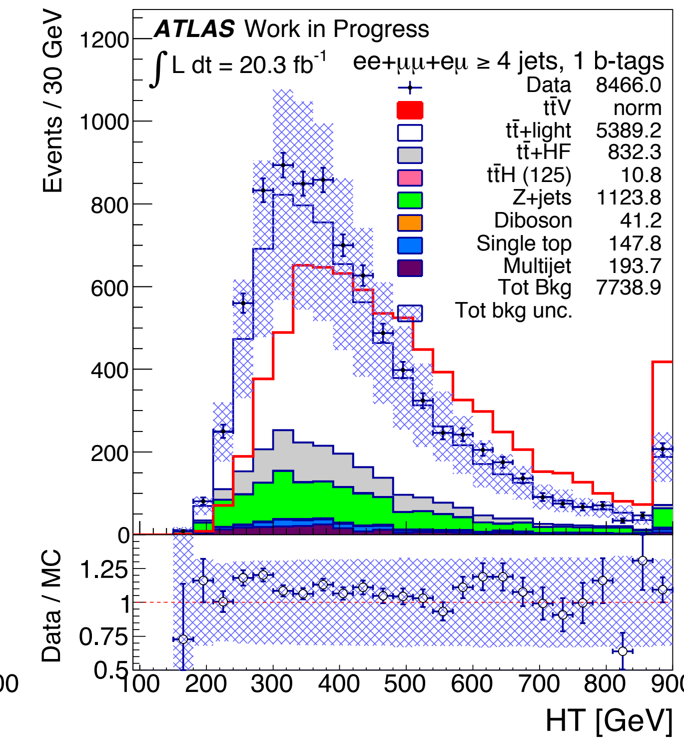
5 Fit (Control)Region I

Prefit Plots









Overall nice data/MC agreement!

some slope visible in 2jex_2bjex

Understand the differences!

4 regions fit

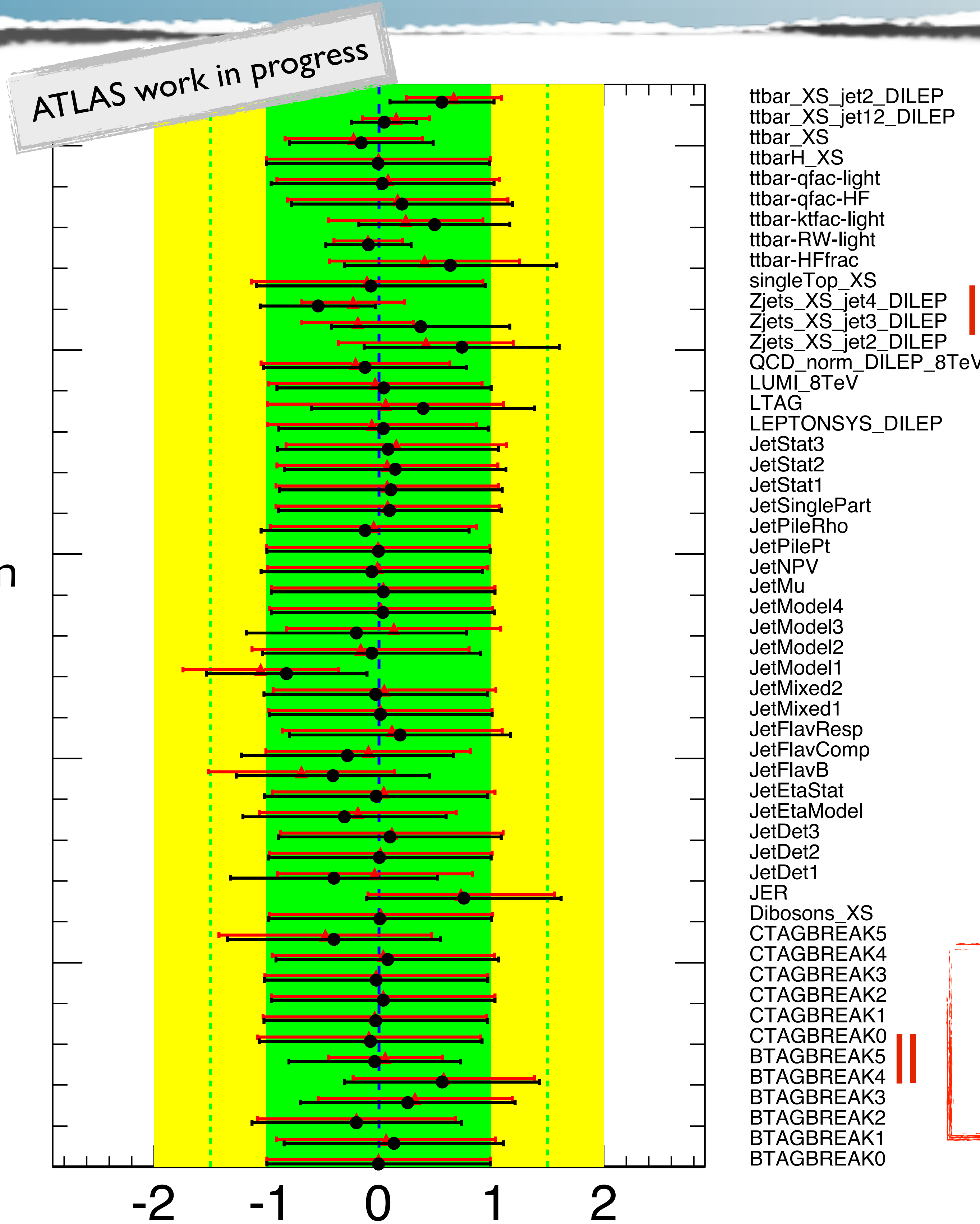
3 regions fit

Remove small systematics from the fit

Very little sensitivity in most of the JES systematics

Most parameters centered around 0 with error of 1, as expected

Statistical error per bin is also profiled

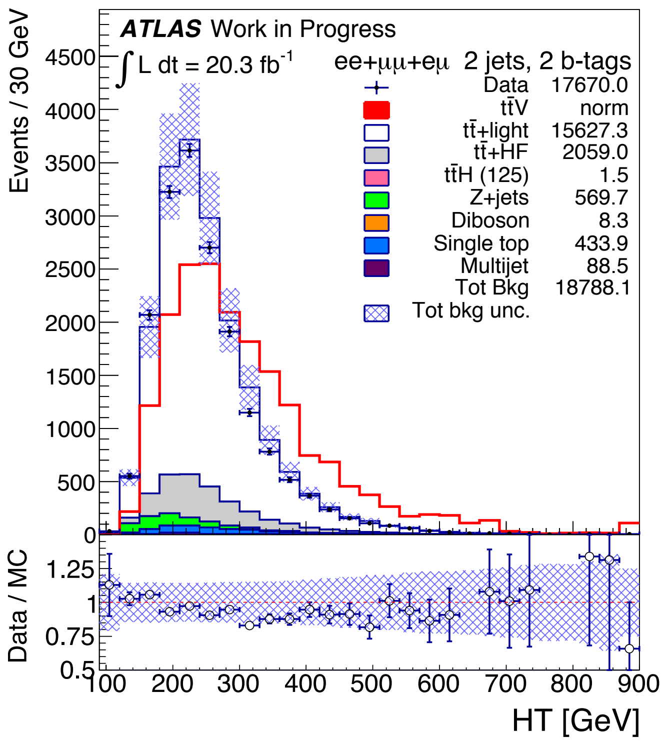


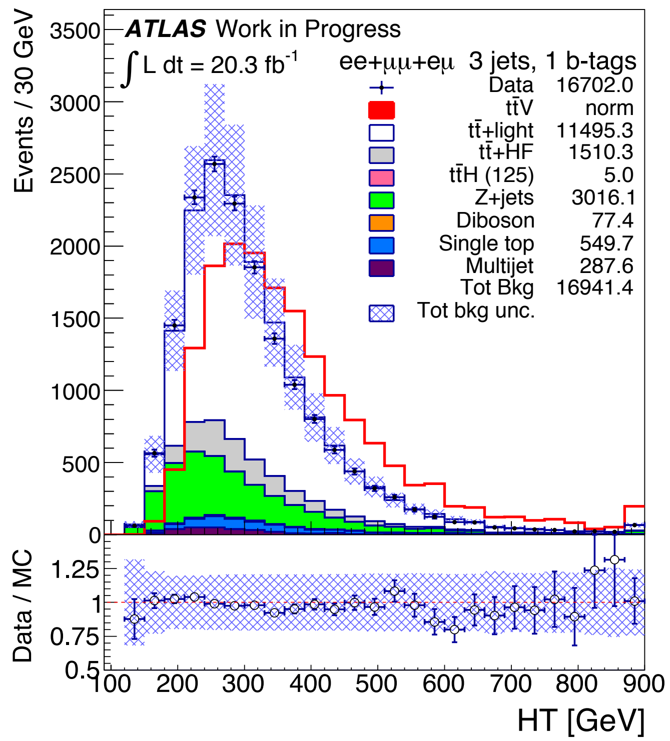
constraint in region with Z+jets contribution (3jex_1bjex → 3jex_2bjex)

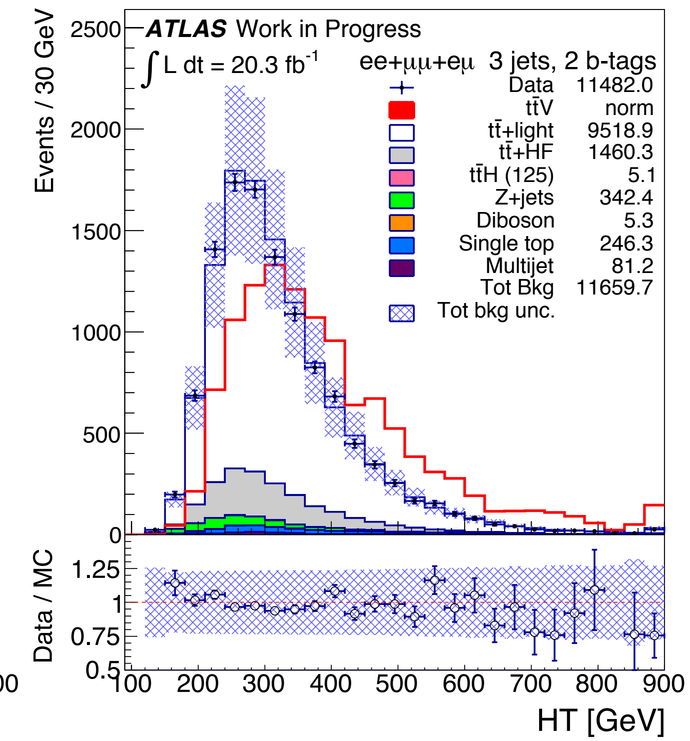
some constraining power from information between two btagged jet regions (3jex_1bjex → 3jex_2bjex)

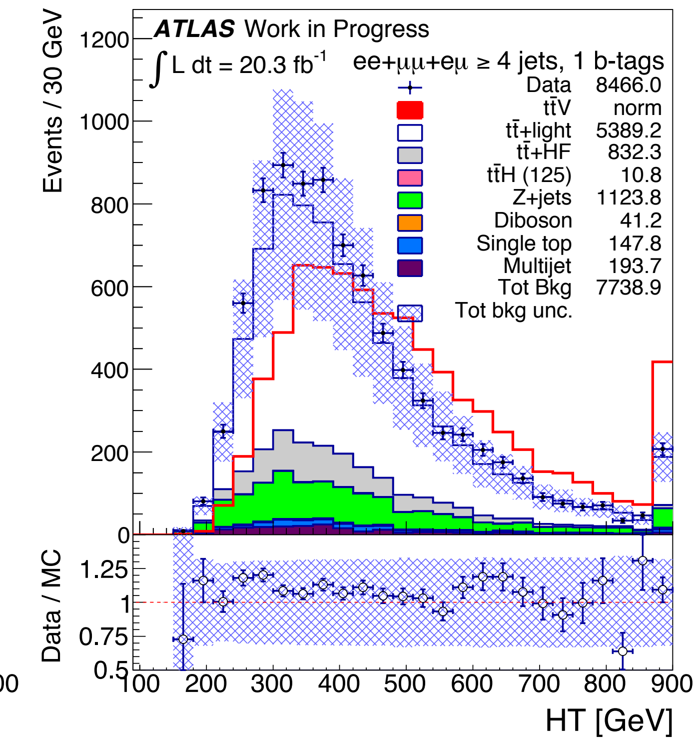
4reg: 2jex_2bjex | 3jex_1bjex | 3jex_2bjex | 4jin_1bjex   **3reg: 2jex_2bjex | 3jex_2bjex | 4jin_1bjex**

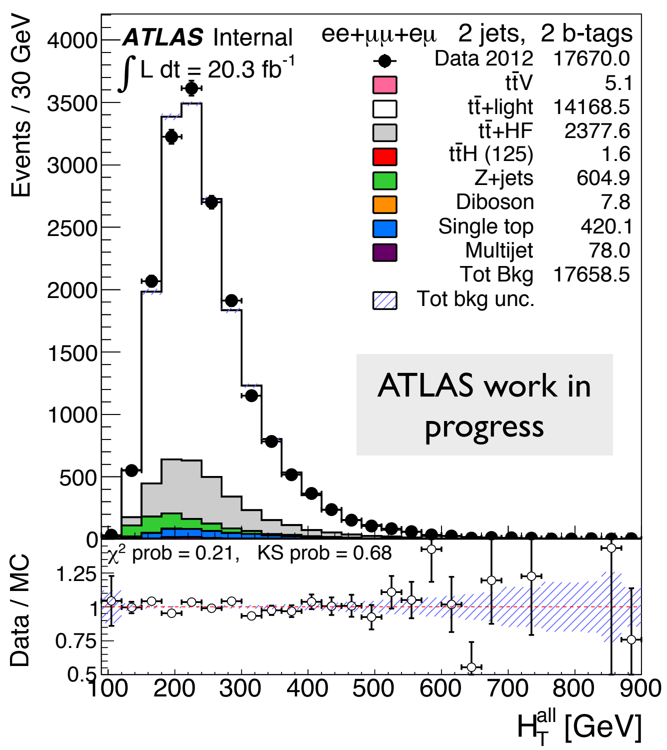
Prefit Plots . vs . Postfit Plots

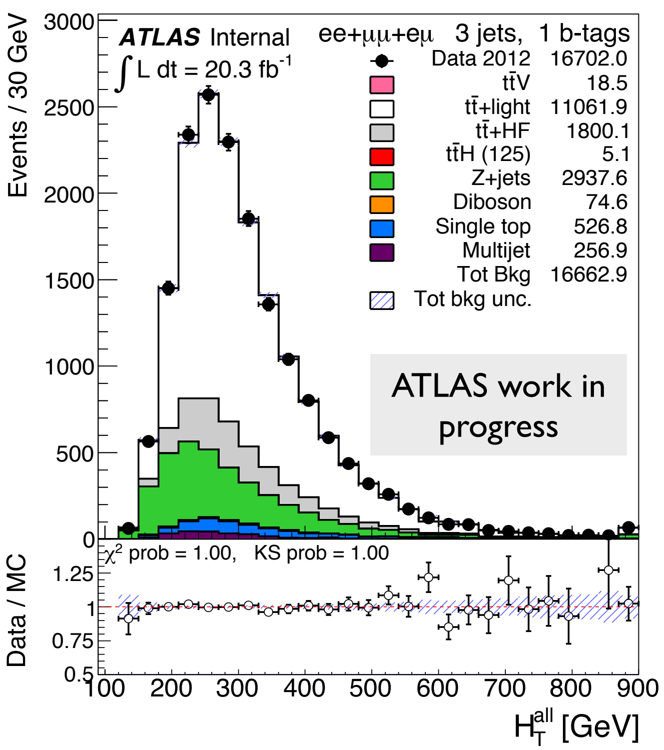


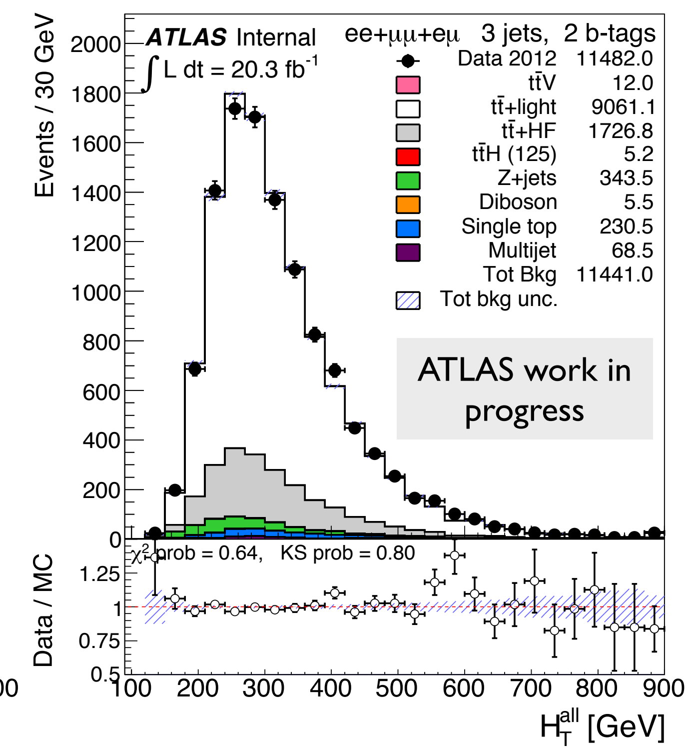


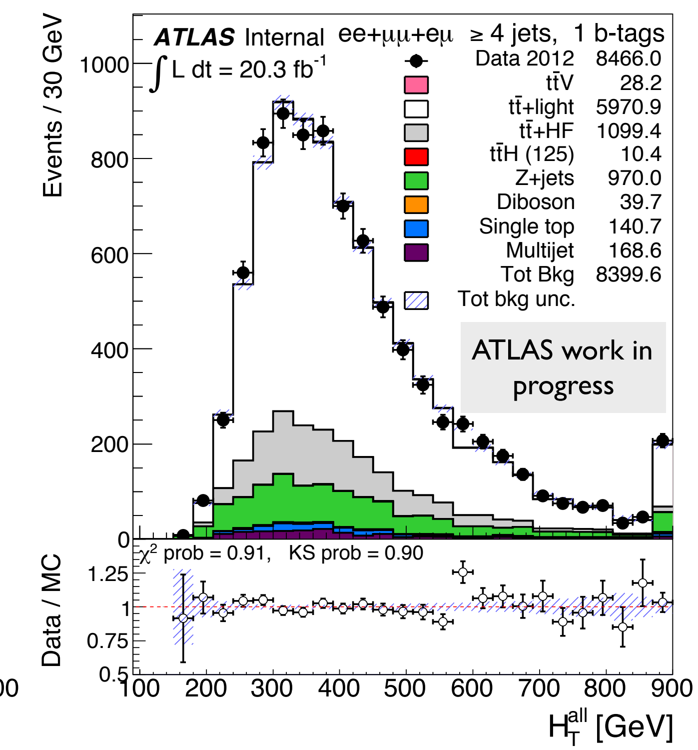












Conclusions

- **Aim:** measurement of TtV (V=Z,W) cross-section in OS dilepton channel in 8 TeV using full data set 20 fb^{-1}
- Distinguish two Regions (1&2) in terms of the dominant truth signal channels and main background contribution
- Not possible to perform a cut&count analysis, use discriminating variables to build a NN response!
- Defined signal and control regions for the fit: divide and conquer!
- First fit results in control regions for Region1 and Region2 look promising

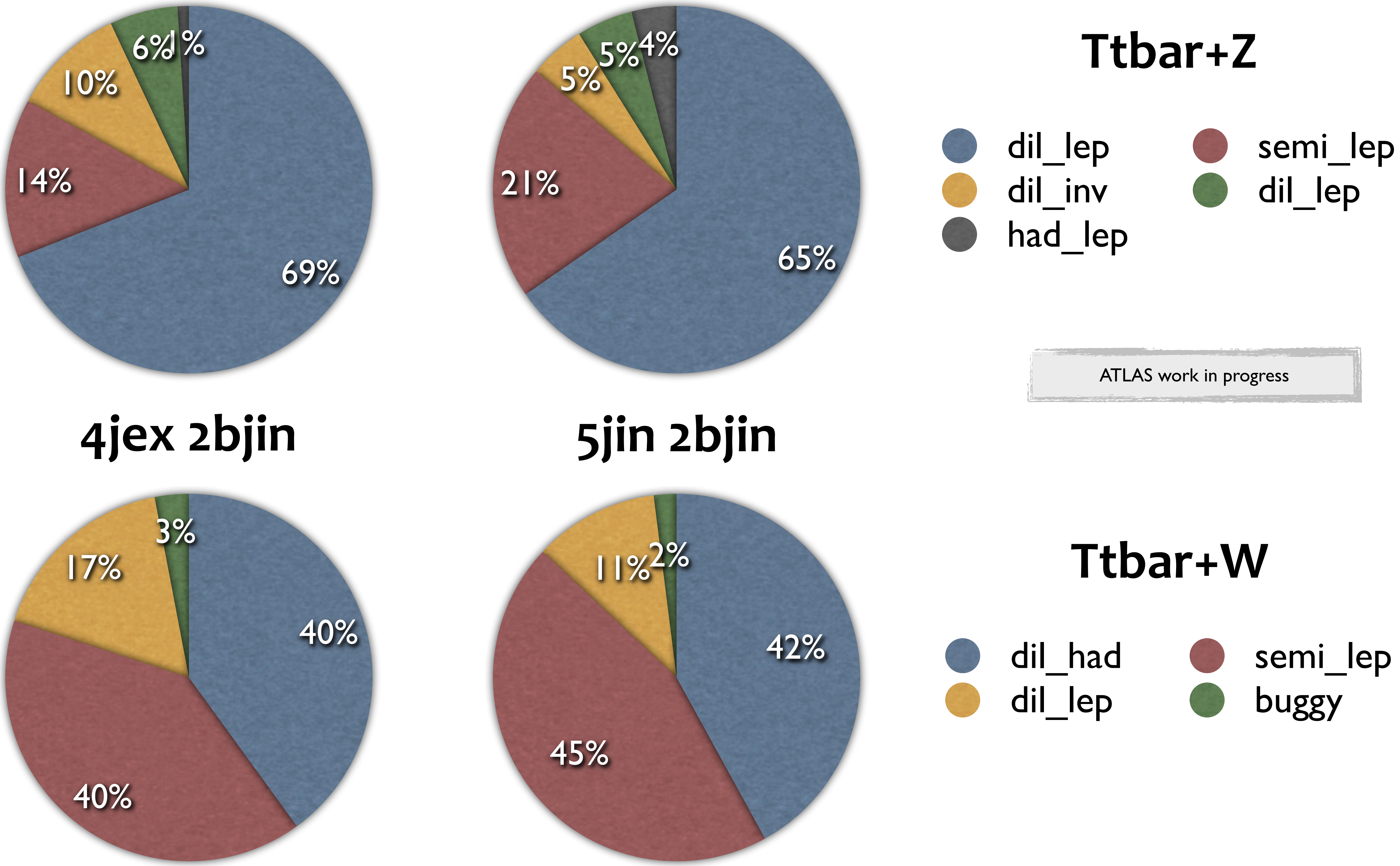
Next steps:

- Once Region1 and 2 fit are understood SEPARATELY, perform combined fit in both regions: most challenging step!
- Create NN response with minimum set of best discriminating and well-modeled variables

For the first measurement of ttV at ATLAS, we are aiming for an efficient analysis!

Back-Up

Truth channel composition in signal



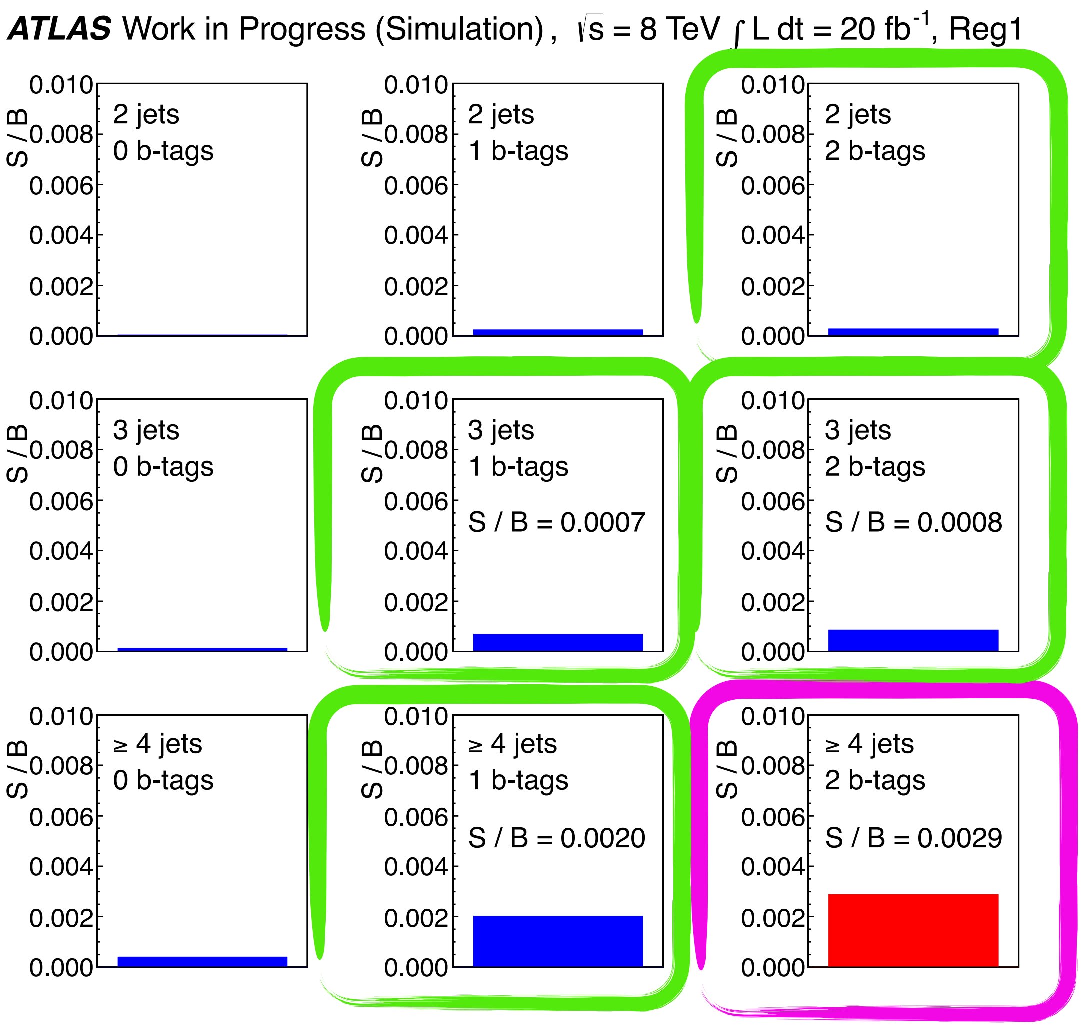
signal and background samples

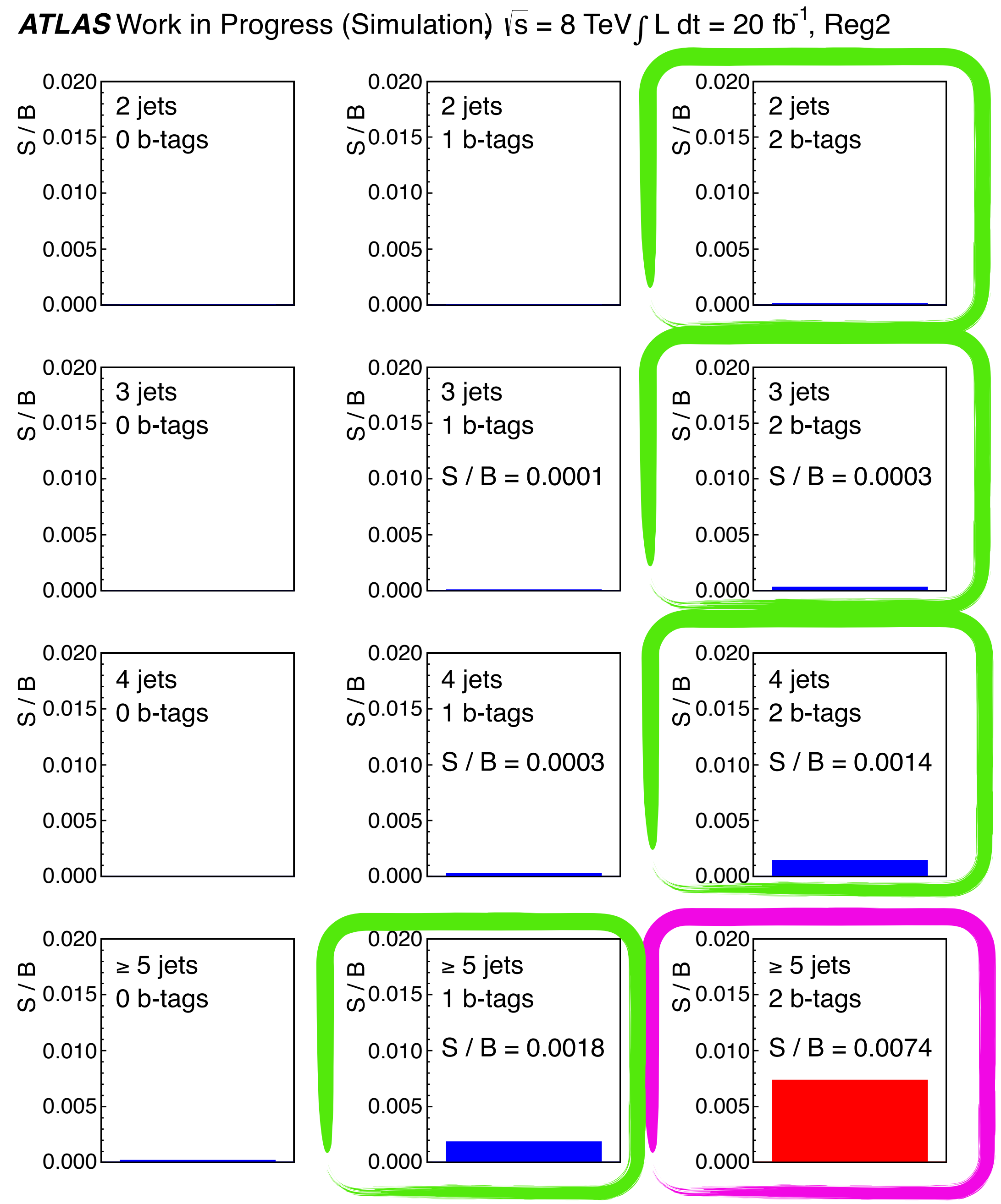
	process	generator
signal	ttZ + Np0/Np1/Np2incl ttW + Np0/Np1/Np2incl	Madgraph + Pythia (new samples on the way, decayed with Madgraph)
background	tt+jets	Powheg+Pythia
	Z+jets	Alpgen+Herwig
	dibosons	Sherpa
	single top	Powheg+Pythia
	ttH	Pythia
	tZ	Madgraph+Pythia
	ttWW	Madgraph+Pythia

First estimation of fakes from data: SS region

- Expecting approximately equal number of fakes in same sign (SS) and opposite sign (OS) events
- Looking at SS data-MC yields
- No. of fakes = data (SS) – MC (SS)

Divide and conquer!





CONTROL REGION

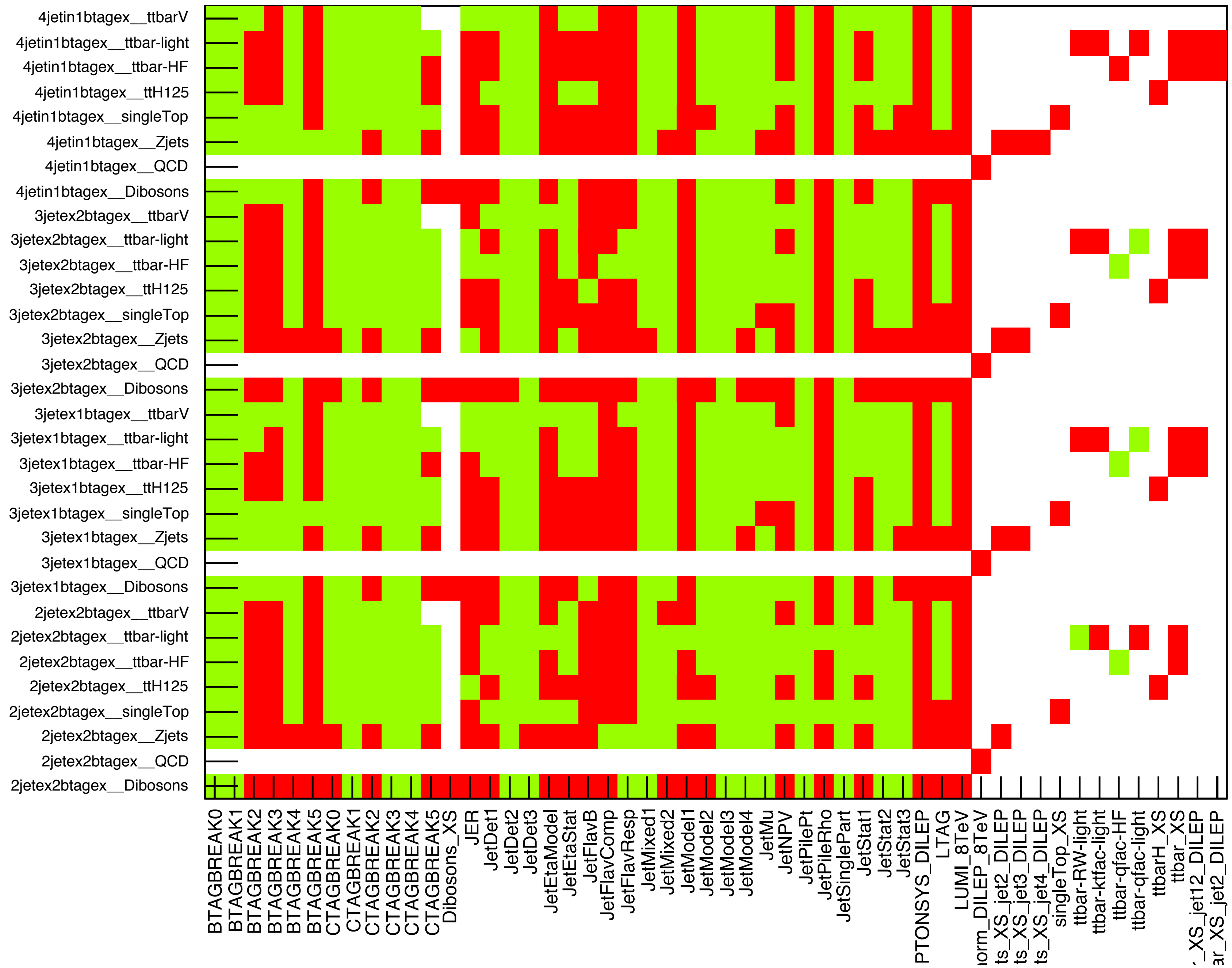
SIGNAL REGION

Region 1 $|m_{ll} - 91| > 10$ GeV

Region 2 $|m_{ll} - 91| < 10$

Acceptance

ATLAS work in progress

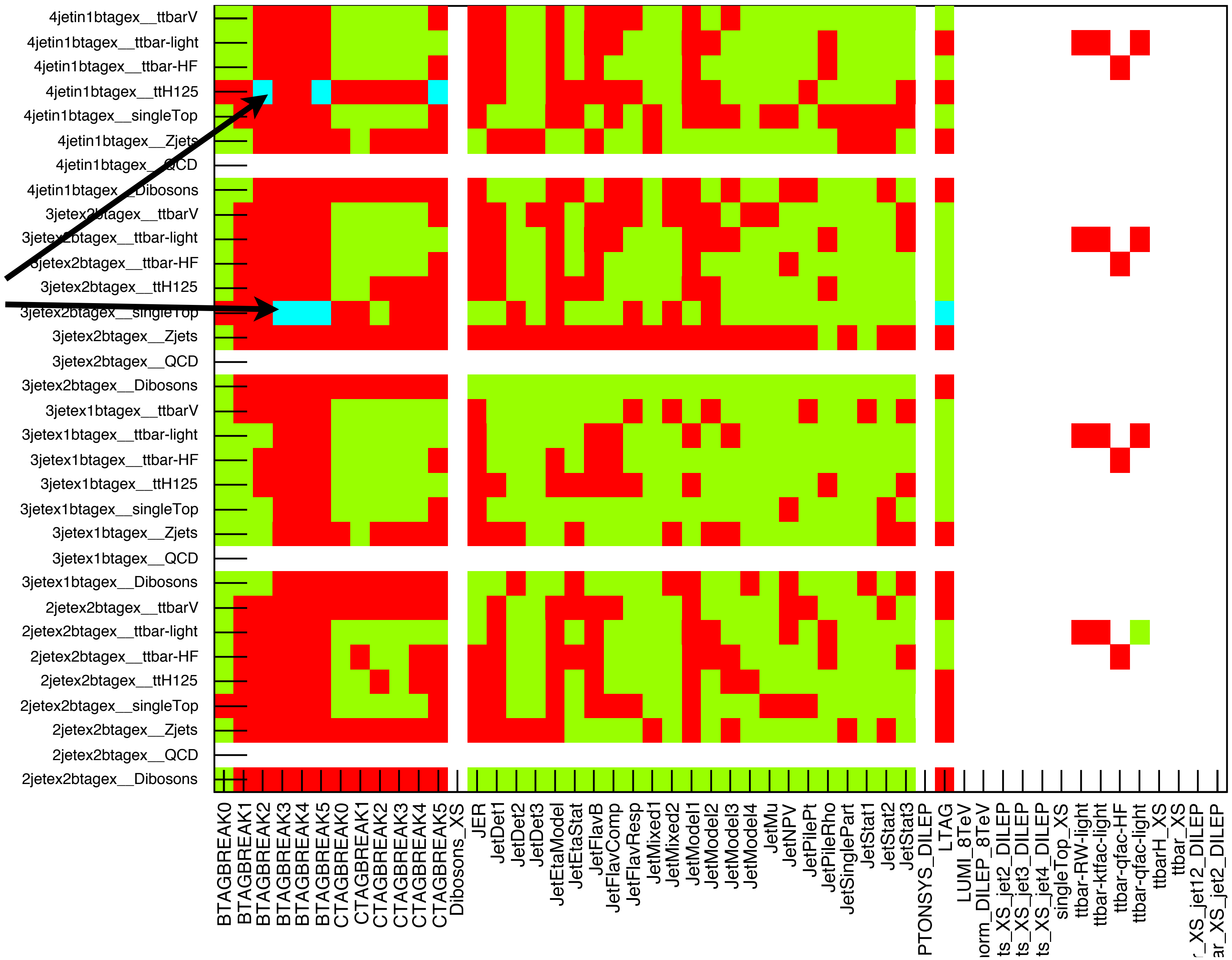


If any bin is has 0.5 % difference w.r.t. average, keep systematic (red)

Shape

ATLAS work in progress

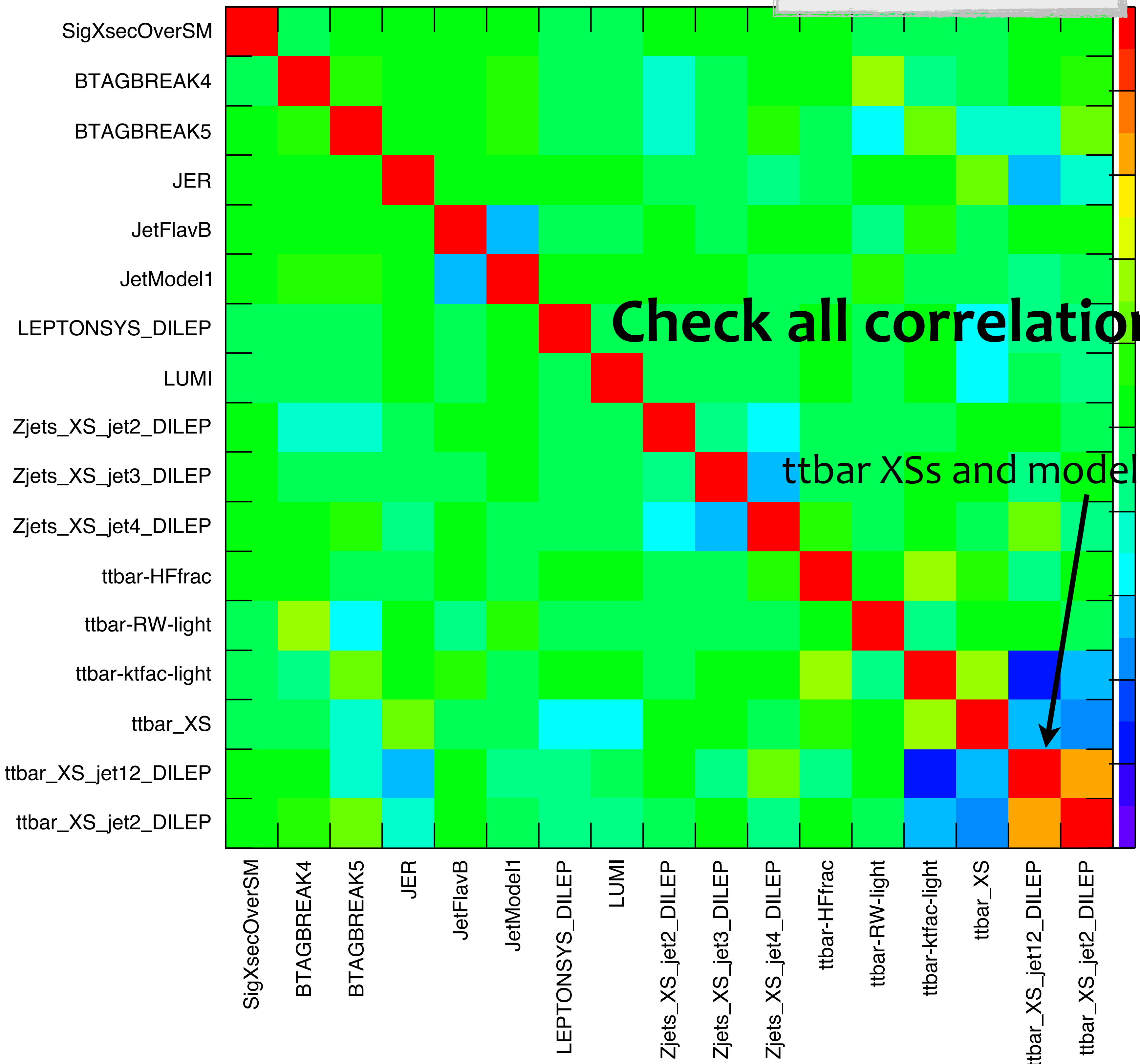
low stat!



If any bin is has 0.5 % difference w.r.t. average, keep systematic (red)

Correlations (3 reg)

ATLAS work in progress

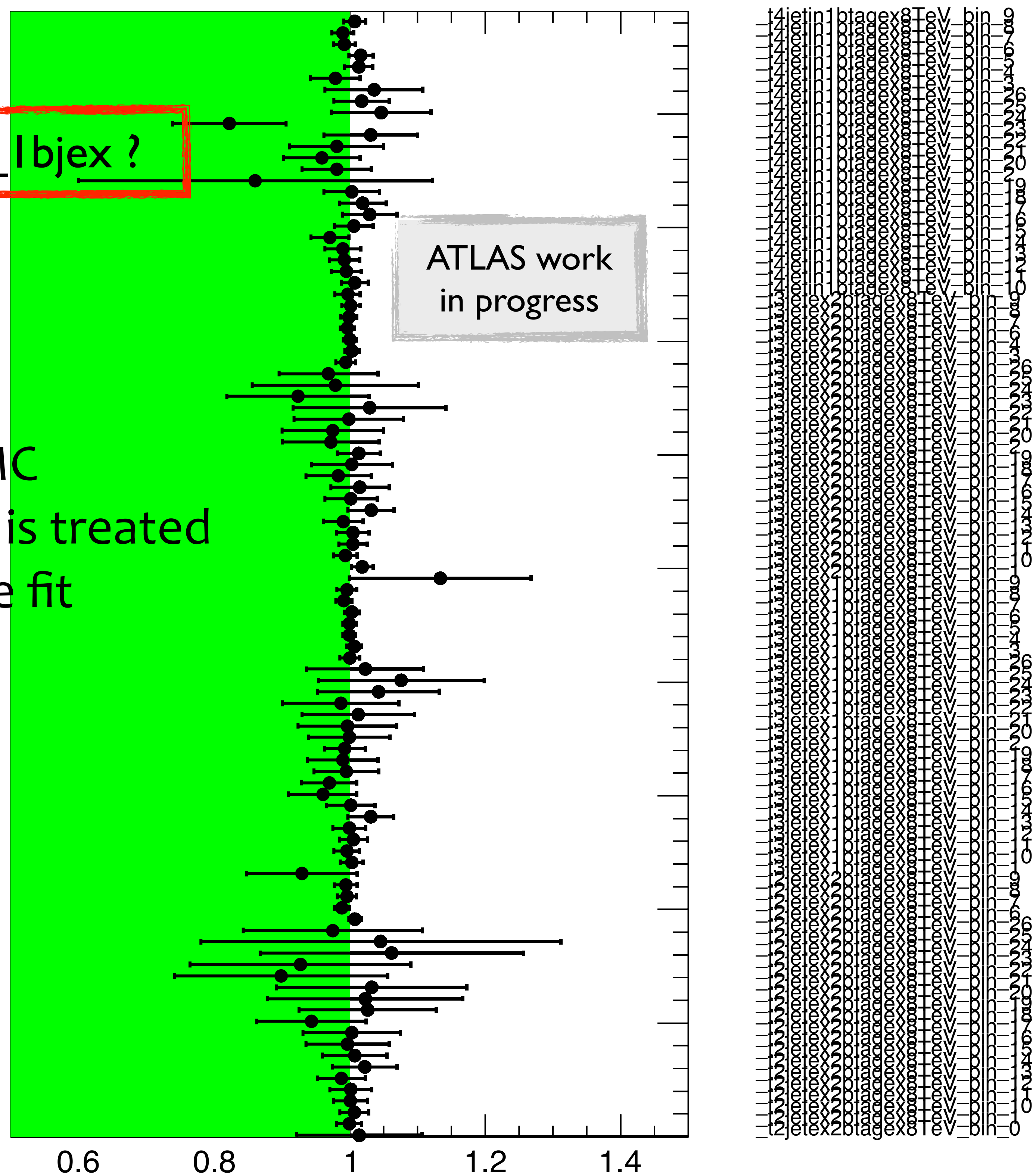


Check all correlations above 30 %

ttbar XSs and modeling heavily correlated

Gammas (template stat error)

need to rebin 4jin_1bjex ?



Each bin contains a MC statistics error which is treated as any other NP in the fit

Check the statistical fluctuations per bin

Systematics

ttbar_XS_jet2_DILEP
ttbar_XS_jet12_DILEP
ttbar_XS

ttbar normalization (ttbar+0/1/2, ttbar+1/2, ttbar+2)

ttbarH_XS
ttbar-qfac-light
ttbar-qfac-HF
ttbar-ktfac-light
ttbar-RW-light
ttbar-HFfrac

ttbar modeling:
Alpgen-related systematics (ktfac, qfac)
top pT RW taken as full systematic (light only)

singleTop_XS
Zjets_XS_jet4_DILEP
Zjets_XS_jet3_DILEP
Zjets_XS_jet2_DILEP

Z+jets normalization

QCD_norm_DILEP_8TeV
LUMI_8TeV
LTAG
LEPTONSYS_DILEP
JetStat3
JetStat2
JetStat1
JetSinglePart
JetPileRho
JetPilePt
JetNPV
JetMu
JetModel4
JetModel3
JetModel2
JetModel1
JetMixed2
JetMixed1
JetFlavResp
JetFlavComp
JetFlavB
JetEtaStat
JetEtaModel
JetDet3
JetDet2
JetDet1

22 **JES** parameters

JER
Dibosons_XS
CTAGBREAK5
CTAGBREAK4
CTAGBREAK3
CTAGBREAK2
CTAGBREAK1
CTAGBREAK0
BTAGBREAK5
BTAGBREAK4
BTAGBREAK3
BTAGBREAK2
BTAGBREAK1
BTAGBREAK0

btagging EV (ttbar-derived btagging calibration)