Laboratoire d'Annecy de Physique des Particules

ML developements for the WZ leptonic analyses in ATLAS

IN2P3/IRFU Machine Learning Workshop – 22-23 of January 2020

L. Portales

Supervision : F. Costanza, I.Koletsou, E.Sauvan

ATLAS EXPERIMENT

VBSTIME

UNIVERSITÉ SAVOIE MONT BLANC

- These slides will to illustrate some of the problematics we are dealing with, and the ML methods we're testing to handle them
- This is all work in progress, and any suggestions or advice is extremely welcome

Content overview

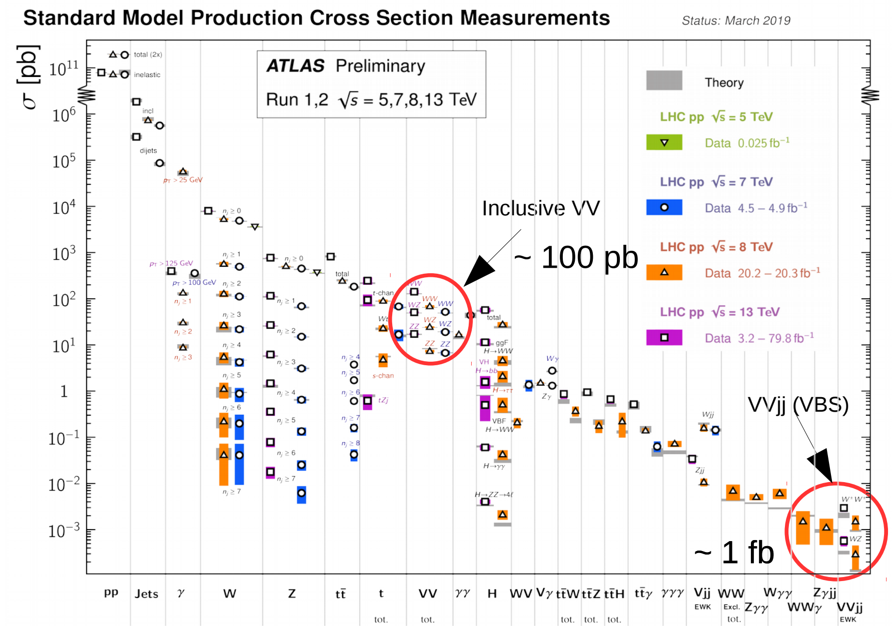
- Analysis overview
- DNN for neutrino reconstruction
- Alternatives to BDTs for signal discrimination
  - > 'Simple' (high-level) DNN
  - > 'Particle-based' (lower-level) DNN
  - > ANN for systematics reduction

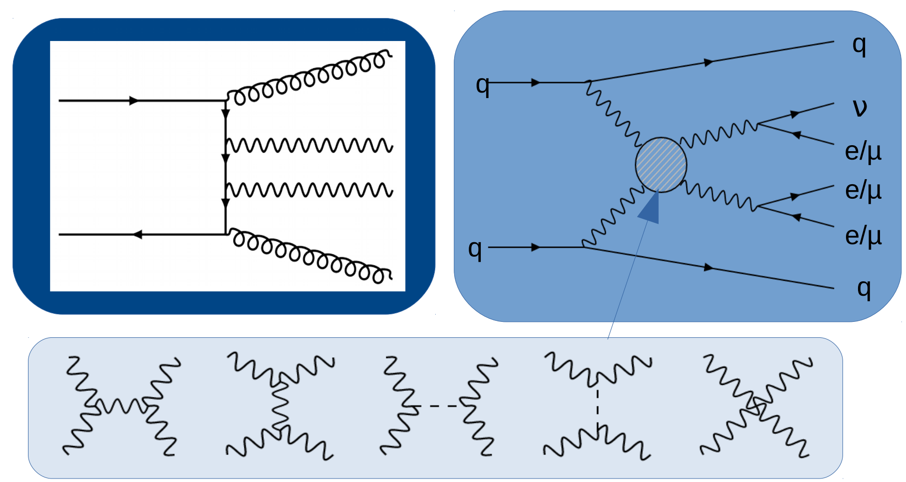
Analysis overview

PLB 793 (2019) 469-492

WZ Analyses – Overview

- WZjj-EW is one of the rarest final-states accessible at the LHC
  - > Among rare processes including quartic boson coupling
  - > Indirect probe of Higgs properties (through $V_L V_L$ scattering)
- Looking for 3 e/μ, 1ν + 2 jets
  - > Jet kinematics is the main identification mean of the process
  - > Makes the analysis sensitive to Jet calibration systematics
- Very large irreducible background
  - > **WZjj-QCD** ~ 2-4 times larger than **WZjj-EW** in measurement phase space
  - > Tipically not extremely well modelled (also true for signal)





Neutrino longitudinal component

- No way to measure it directly

  > Problematic for boson polarisation measurements, very sensitive to boson decay caracteristics

- Usual evaluation methods :

  > Assuming $M_W = M_{pole}$: $A\, p_{z,\nu}^2 + B\, p_{z,\nu} + C = 0$

$$A = p_{T,lep}^2$$

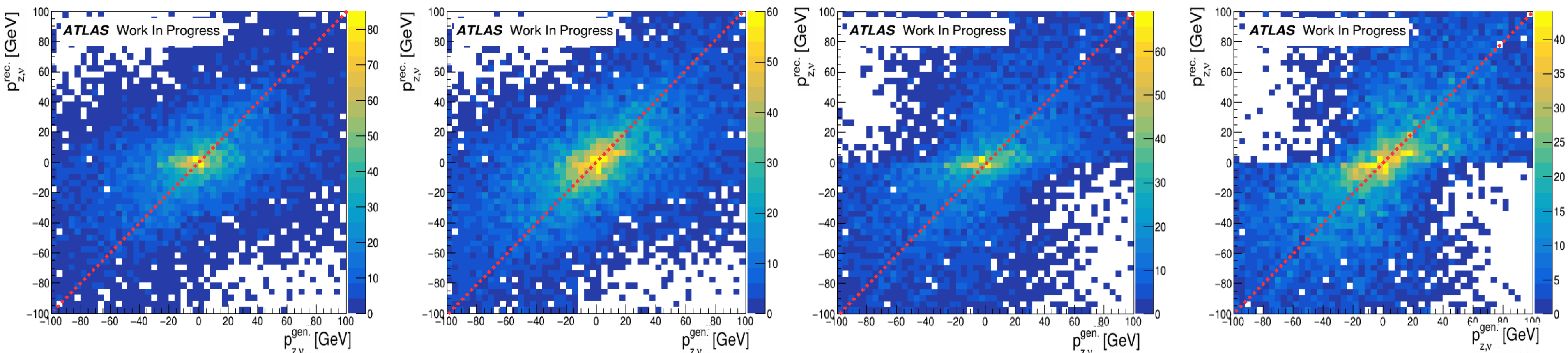
$$B = -2\, p_{z,lep}\left(p_{x,lep}\, p_{x,\nu} + p_{y,lep}\, p_{y,\nu} + \frac{M_W}{2}\right)$$

$$C = -\left(p_{x,lep}\, p_{x,\nu} + p_{y,lep}\, p_{y,\nu} + \frac{M_W}{2}\right)^2 + E_{lep}^2\, p_{T,\nu}^2$$

  > Second order equation → 2 solutions in most cases (& can be complex)

- Several methods tested to deal with the ambiguity

  > No clear winner :



- NN seem well suited for this problem

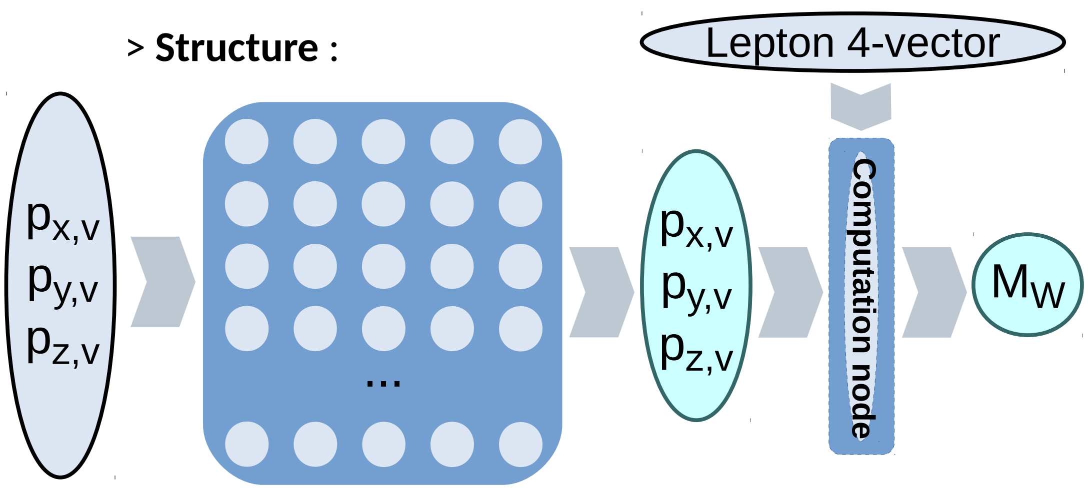
> Trained a network for regression

> **Aim** : evaluate $p_{z,\nu}$ (+ correct $p_{x,\nu}$,$p_{y,\nu}$) avoiding degeneracy from analytical method

> **Inputs** : W lepton 4-momentum, $p_{x,\nu}$,$p_{y,\nu}$,$p_{z,\nu}$ (analytical)

> **Outputs** : neutrino 4-momentum + M_W

> **Structure** :

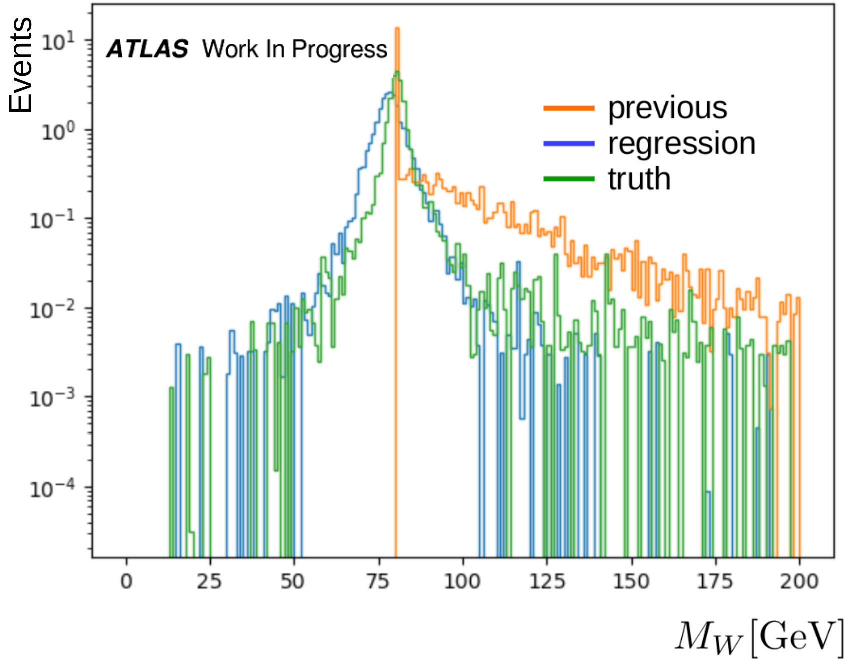


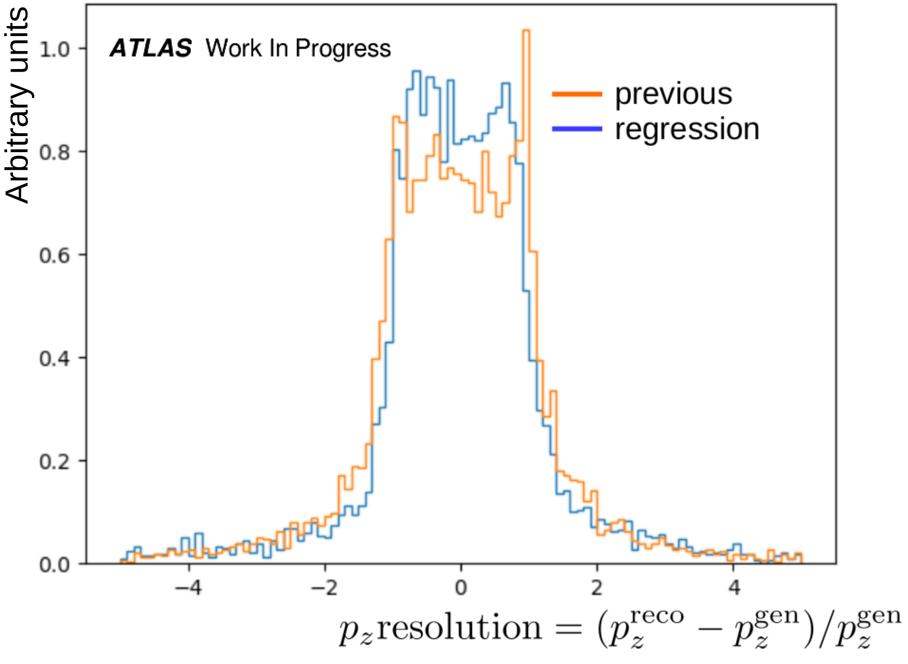
- Keras + Tensorflow
- 8 hidden layers
- 512 nodes per layer
- LeakyReLU , 0.1 leakage
- Dropout : 0.1
- Adam optimiser, variable learning rate (0.001->0.0001->0.00001)
- Training/validation samples yields : > 70k/20k events
- Split in 2048 event batches
- Trained over 300 epochs (with early stopping)

> Loss function defined to account for both M_W and neutrino 4-momentum resolutions:

$$L = k_0 MSE(\hat{p_{x,\nu}}) + k_1 MSE(\hat{p_{y,\nu}}) + k_2 MSE(\hat{p_{z,\nu}}) + k_3 MSE(\hat{M}_W) + k_4 \left\langle \frac{M_W^{truth} - M_W^{pred.}}{M_W^{truth}} \right\rangle$$

- Some slight improvement w.r.t. analytical method
  - > Small bias in M_W peak position would require some fine tuning





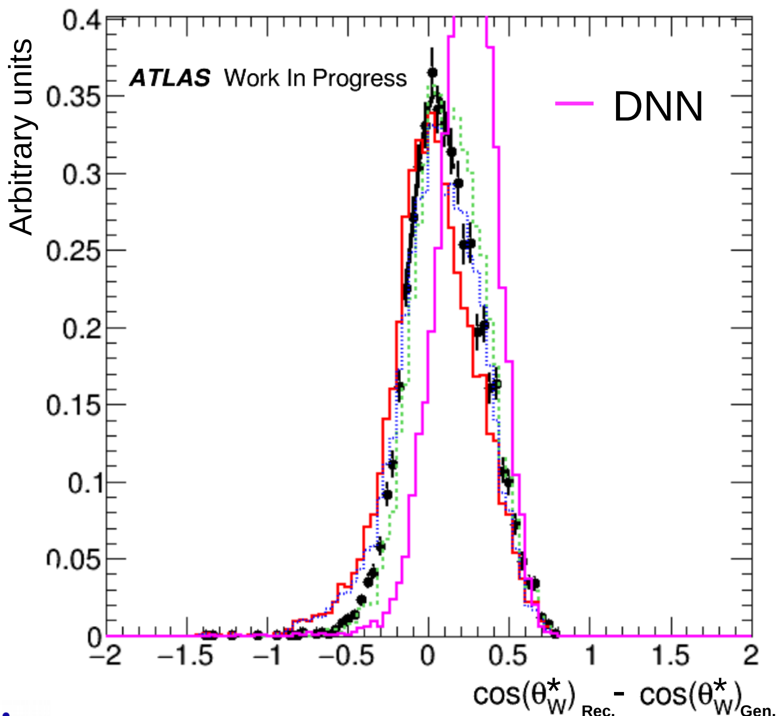
ML alternative

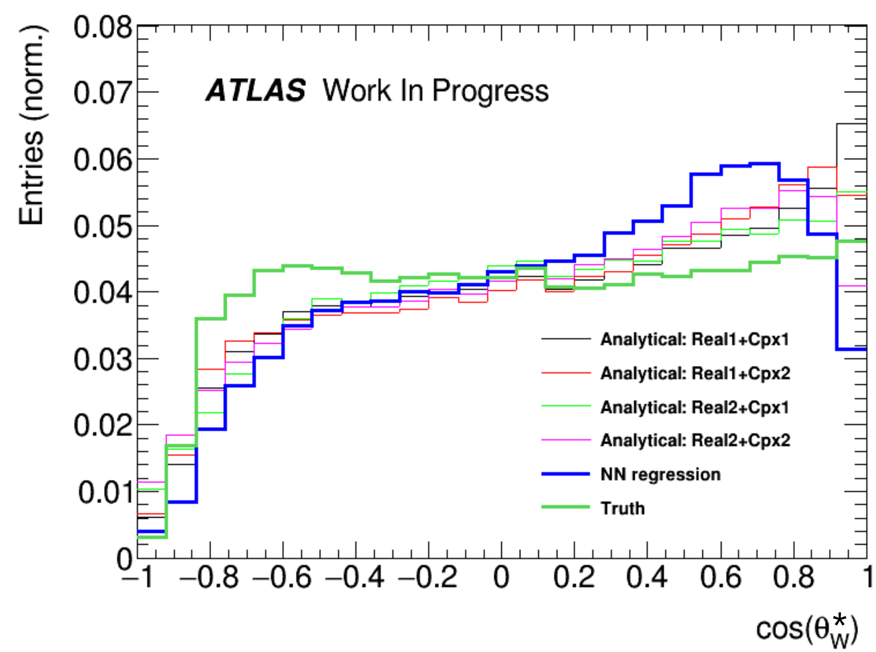
- Some degradation in the resolution of other variables of interest

  > Especially $\cos\theta^{*}_{W}$, the kinematic variable most sensitive to boson polarisation state

  > Main measurement of WZ inclusive analysis, and of great interest for future VBS studies

  > Plan to include it in network optimisation instead of(/with) M_W





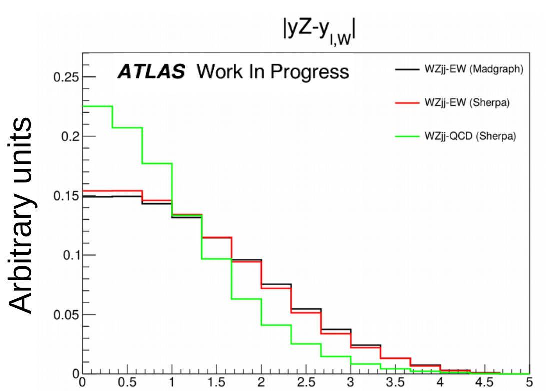
NN for background rejection

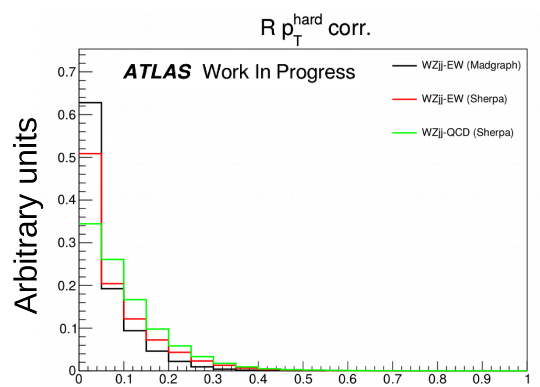
- Previous analysis used a BDT to enhance sensitivity to signal

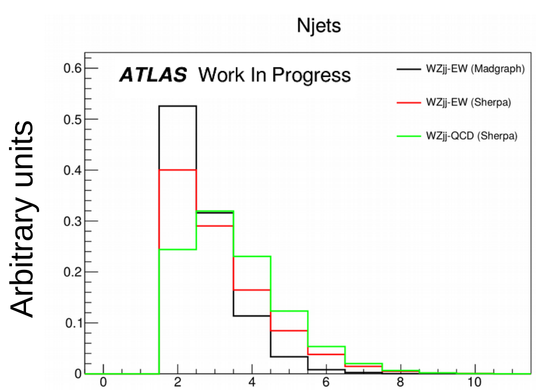
  > 15 kinematic variables selected after thorough optimisation

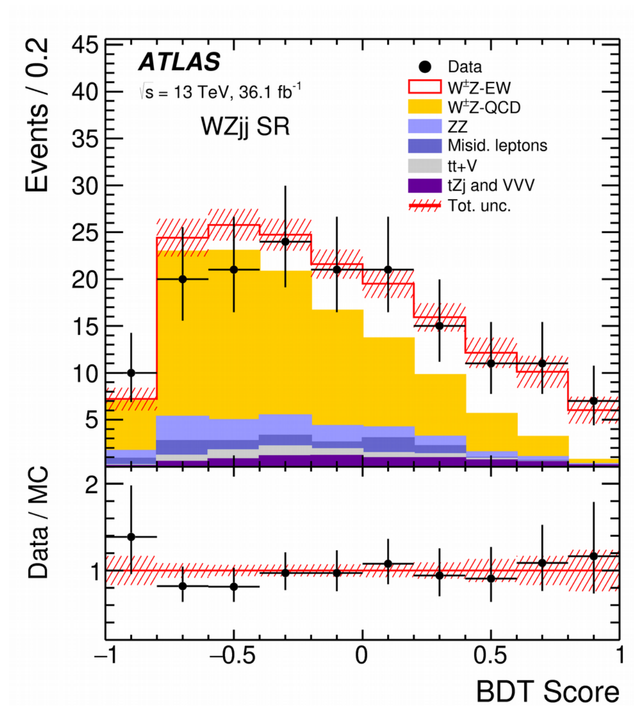
  > Allowed first observation of electroweak WZjj production with 2015+16 data

  > Retrained for full Run 2 analysis, and compared to alternative discriminants









- Different NN compared, with different type of inputs

  > Same input set as BDT : 'base DNN'

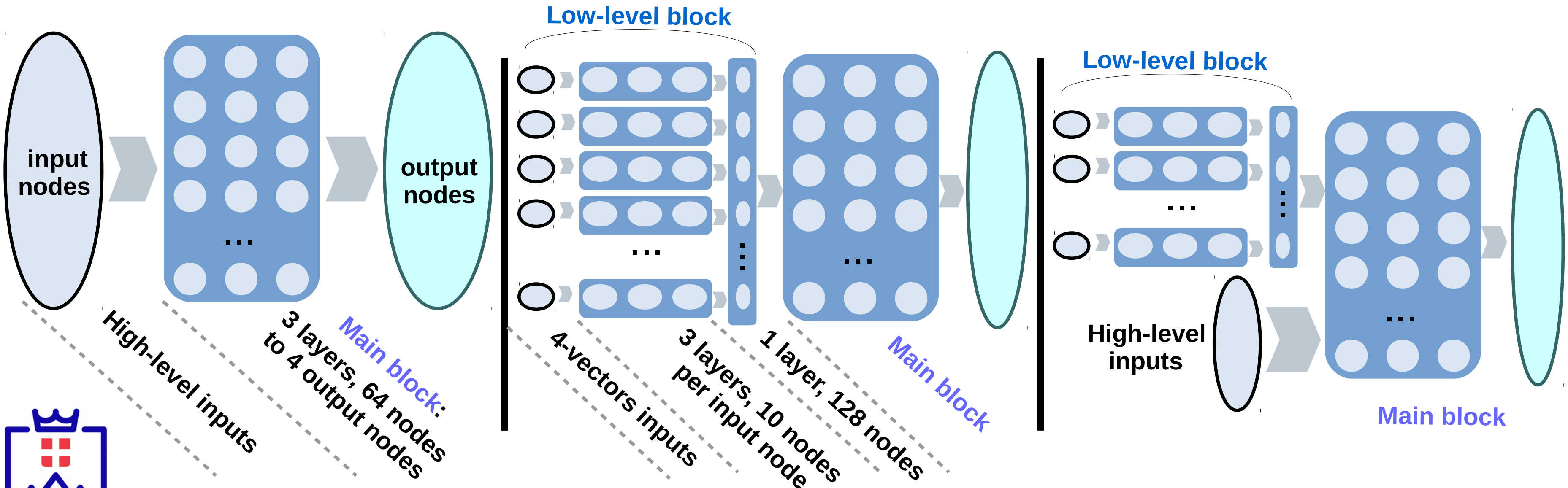
  > BDT inputs + added high-level : 'new inputs, DNN'

  > Lower-level inputs (2 jets + 4 leptons 4-momentums) : 'pNN'

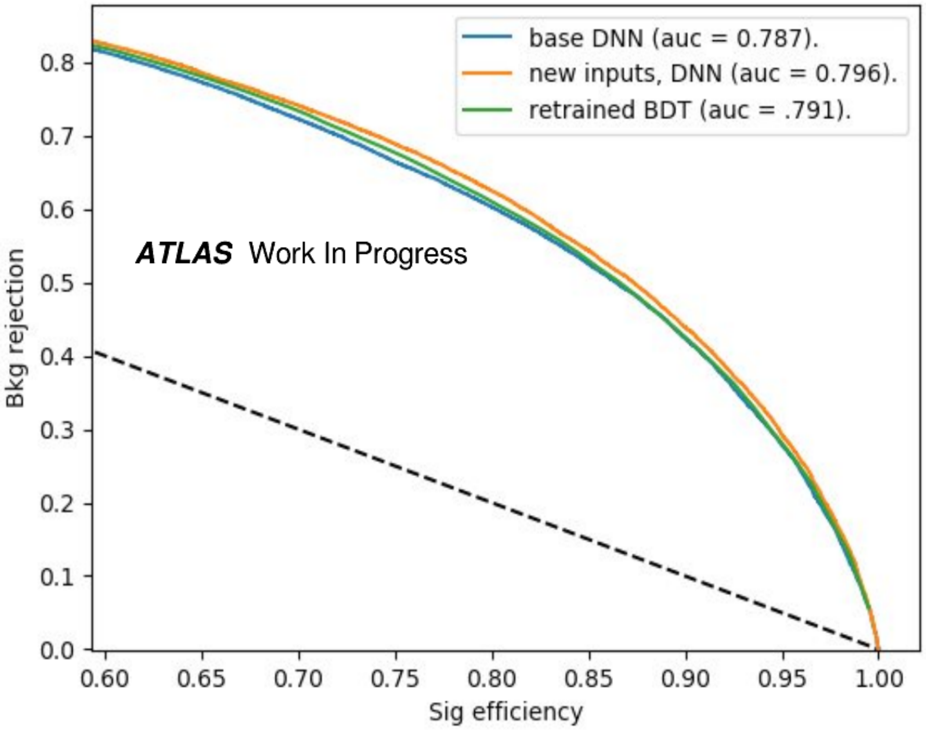
  > Lower-level + selected high-level inputs (not built from 4-momentum): 'pNN+X'

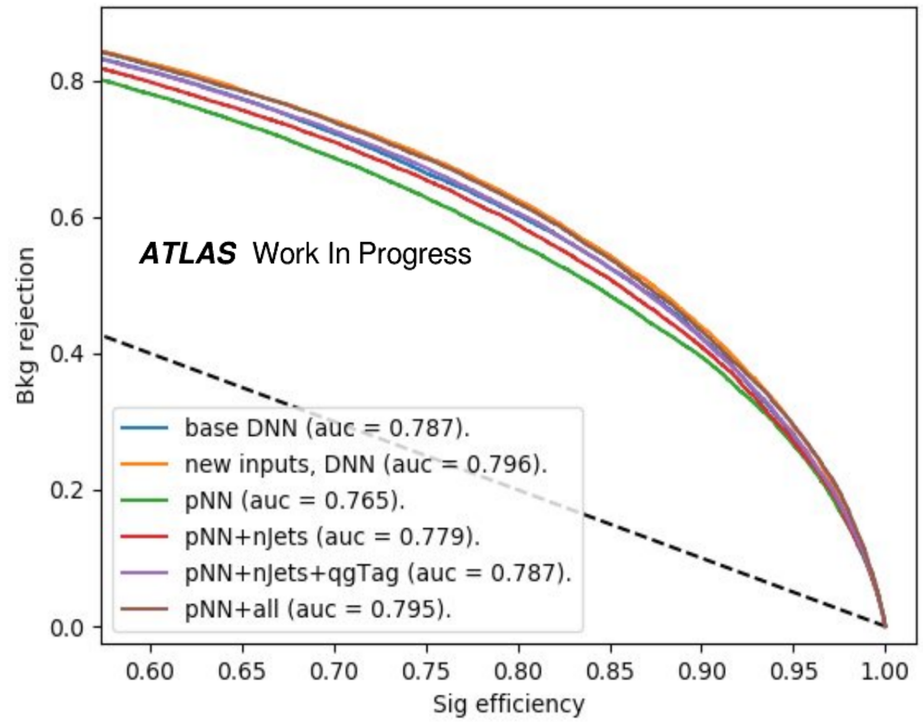
- 3 network structures to accomodate input type :

  > Similar parametrization (optimised on simpler models, DNN & pNN):



- Slight improvement over BDT when adding inputs
- Similar improvement with either high- or low-level inputs added





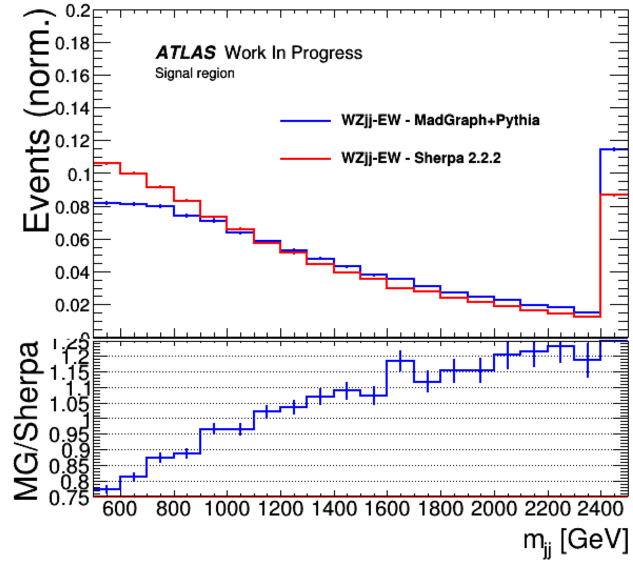
- Alternatively, Adversarial network has been tested :

> Aim is getting a classifier decorrelated from m_{jj}

> Motivation :

- m_{jj} shape largely impacted by both object and theory/modelling uncertainties

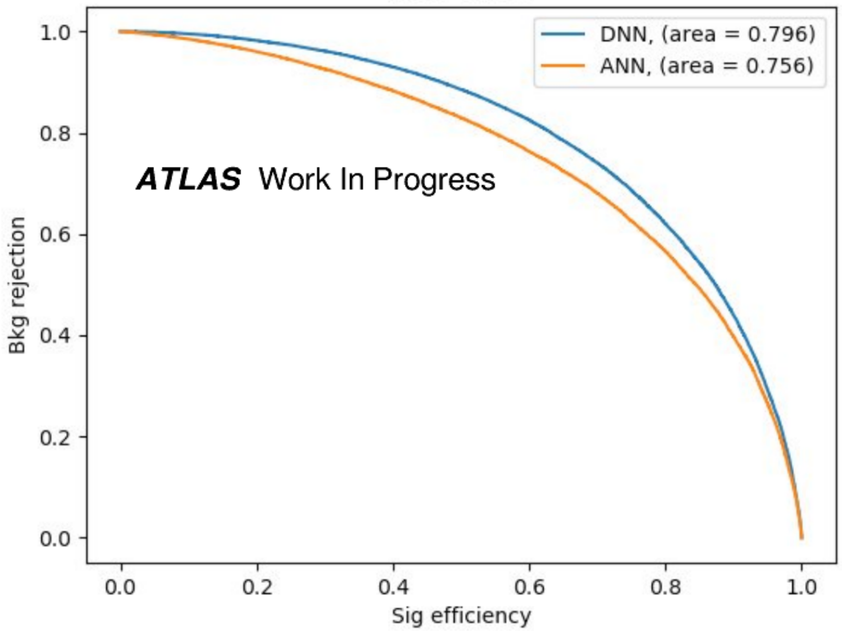
Source	Uncertainty [%]
$WZjj$–EW theory modelling	4.8
$WZjj$–QCD theory modelling	5.2
$WZjj$–EW and $WZjj$–QCD interference	1.9
Jets	6.6
Pile-up	2.2
Electrons	1.4
Muons	0.4
b-tagging	0.1
MC statistics	1.9
Misid. lepton background	0.9
Other backgrounds	0.8
Luminosity	2.1
Total Systematics	10.7

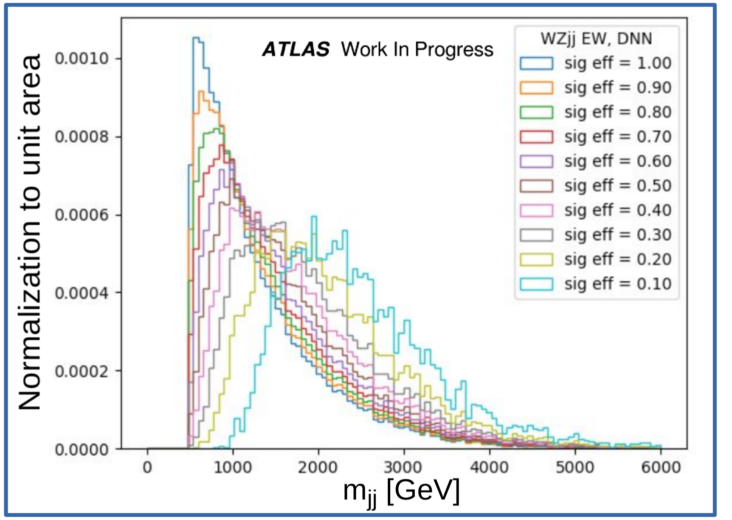


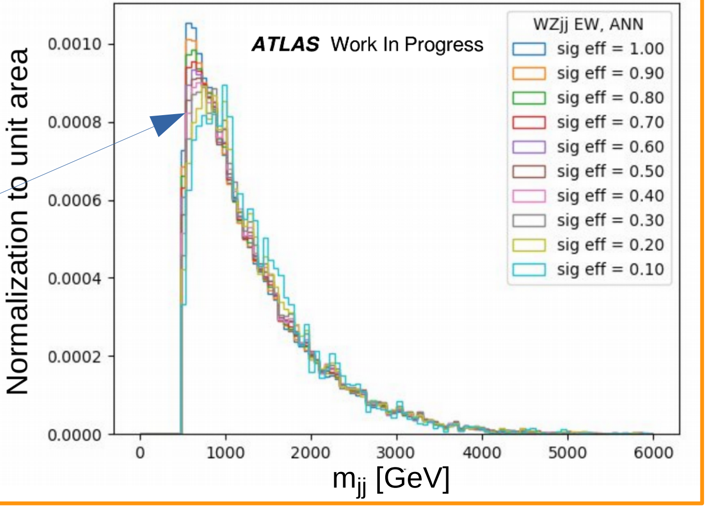
> ANN structures (in brief):

- Classifier is the same as the 'high-level' DNN
- Adversarial build with same structure, but without softmax output node
- Adversarial tries to guess mJJ from the classifier output instead

- Does its job rather well, but causes a large performance loss







Some improvement can be made at low m_{jj}, by increasing the training range

- Discriminant is meant to be used as a fit template

  > Benchmarking can be performed directly in that context

  > MC-only fit performed, using an artificially generated (‘Asimov’) dataset

  > Fit parameter of interest : Signal yield in the signal region (SR)

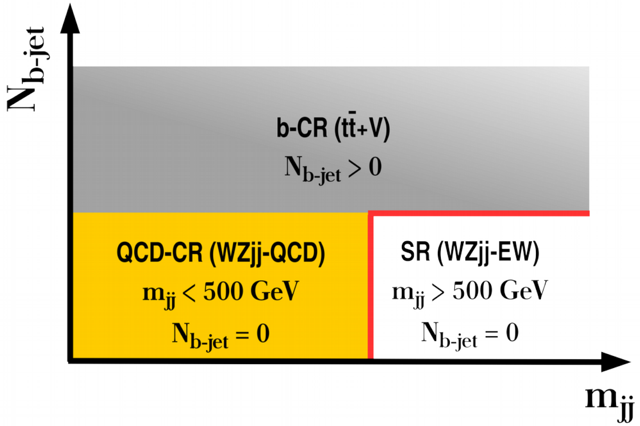
  > ML templates used in this SR

  > Comparison performed looking both at the stat-only sensitivity & with added systematics

  > Systematics included :

    – Modelling : Difference between MC generators for signal & main background (WZjj-QCD)

    – QCD scale : flat +30 %/-20 % uncertainty band on WZjj-QCD



- Fit results consistent with expectations
  - > Slight performance improvement from DNN
    - – **Caveat :** since it uses more features, DNN could be more sensitive to object systematics (this couldn't be tested for now)
  - > Much worse performances with ANN
    - – Interesting ANN quirk : No impact of systematics on significance, i.e. seems less sensitive to background modelling uncertainties
    - – **Caveat :** main (expected) source of improvement not tested yet
      - > ANN could remove the effects of transfer factors and uncertainties from background control region to SR (split with respect to mJJ)

Work In Progress

	BDT		DNN		ANN	
	Stat.	Stat.+Sys.	Stat.	Stat.+Sys.	Stat.	Stat+Sys.
Signal Strength	1±0.086	1±0.203	1±0.083	1±0.196	1±0.098	1±0.218
Systematics impact		0.184		0.178		0.195
Significance	7.9 σ	7.8 σ	8.2 σ	8.0 σ	6.0 σ	6.0 σ
ROC AUC	0.791		0.796		0.756	

- Various ML methods have been tested for the WZ analyses

  > DNN for neutrino reconstruction

  > DNN/ANN for WZjj electroweak signal extraction

- Some small improvements observed in most cases

  > Slightly improved signal sensitivity compared to previous methods, but not game changing

- More in depth performances evaluation might be needed to confirm it

- These are still promising, as they could be enhanced with some clever optimisation, which might still be lacking.

- Any suggestion is welcome !