Search for heavy resonances in $VV \to llqq$ channel

ATLAS Napoli VV semi-leptonic group: S.Auricchio, G. Carlino, F. Conventi, A. Giannini, M. Lavorgna, E. Rossi

Presentazione tesi di laurea magistrale – Auricchio Silvia

Meeting interno ATLAS Napoli

25/06/2019

Introduction

General search for resonances Beyond Standard Model (BSM) using 139 $fb^{-1}$of pp collisions collected from 2015 up to 2018 (full LHC Run-2) with the ATLAS detector.

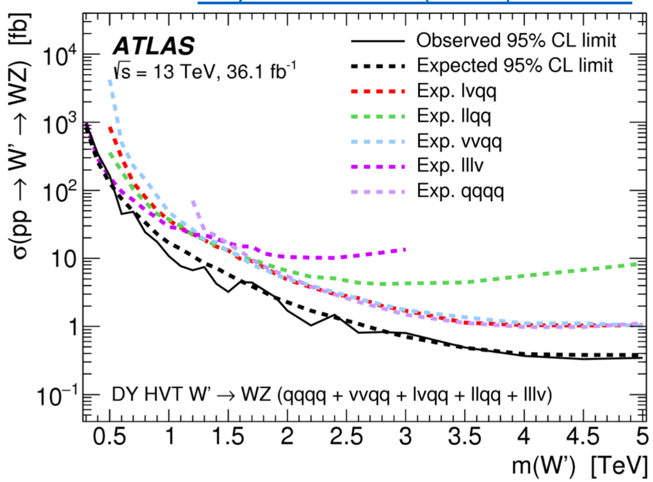
Investigated channel: pp ⟶**X'** ⟶ **ZV** ⟶ **llqq**, **(V** $= Z, W^{\pm}$ **Standard Model).**

$l = e^{\pm}, \mu^{\pm}$

Theoretical models considered:

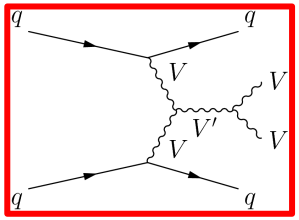
- Spin-0 Radions $\boldsymbol{\Phi \longrightarrow ZZ/WW}$
  arXiv:hep-ph/9907447v2
- Spin-1 Heavy Vector Triplet (HVT) $\boldsymbol{W' \longrightarrow WZ}$
  arXiv:1402.4431v2
- Spin-2 bulk Randal Sundrum Graviton $\mathbf{Gkk} \longrightarrow \boldsymbol{WW/ZZ}$
  arXiv:hep-ph/9811350v4

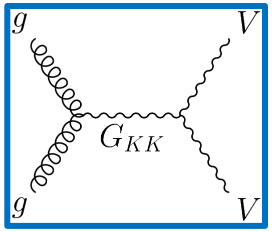
Phys. Rev. D 98 (2018) 052008

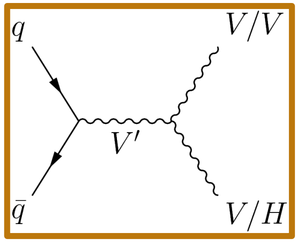


Mass range: 300 GeV < M_X < 6000 GeV

Two production mechanisms: **Vector Boson Fusion (VBF)**, **gluon gluon Fusion (ggF)** (for spin 0 and spin 2) or **Drell Yian process (DY)** (for spin 1).







Events selection

Analysis flow optimized in the previous analysis (36,1 fb^{-1}) https://arxiv.org/abs/1708.09638.

❖ $Z \rightarrow ll$

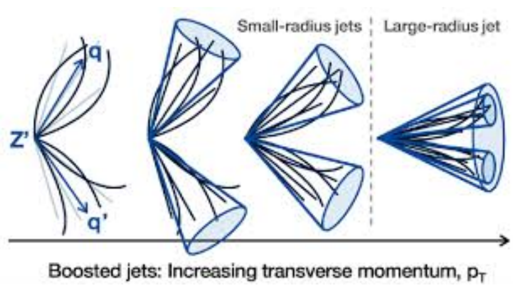
Two opposite charged muons or two electrons with:

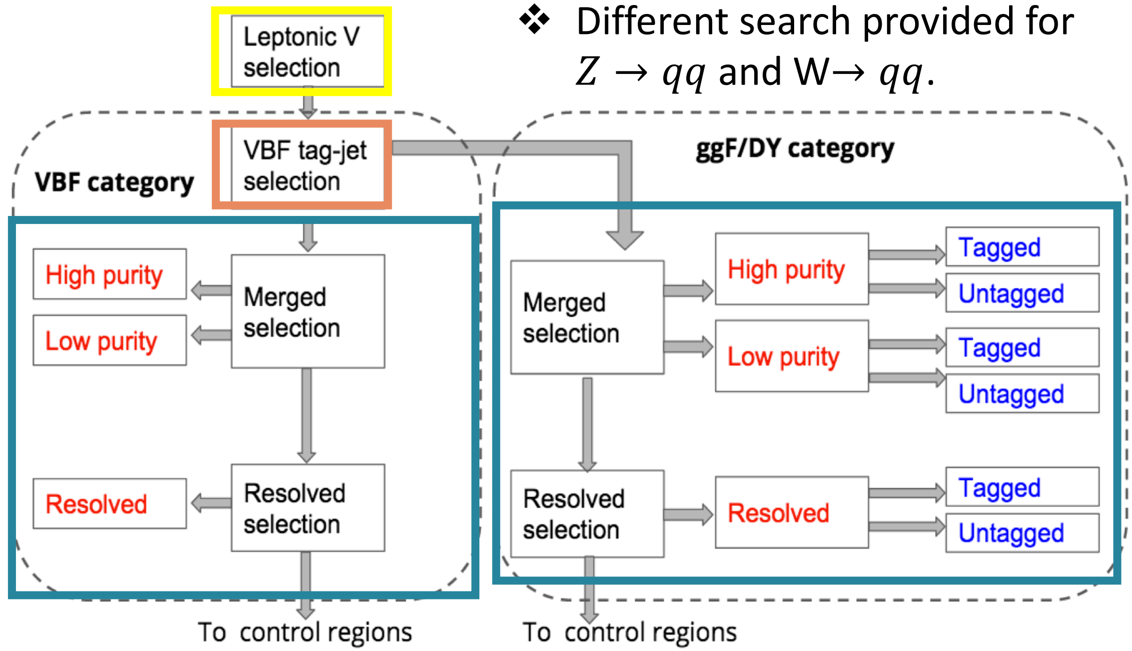
$83\ GeV < m_{ee} < 99\ GeV$

$85{,}6\ GeV\ - 0{,}0117\ p_T < m_{\mu\mu} <$
$94{,}0\ GeV + 0{,}0185\ p_T$.

❖ Production mechanism categorization: **VBF** or **ggF/DY.**

❖ $V \rightarrow qq$ in merged regime first, if it fails in resolved regime.





❖ Different search provided for $Z \rightarrow qq$ and W→ qq.

~21% of signal events contains b-jets ($Z \rightarrow b\ \bar{b}$) while they are rare in the W/Z + jets processes (the dominant background).

In ggF category for $Z \rightarrow qq$:

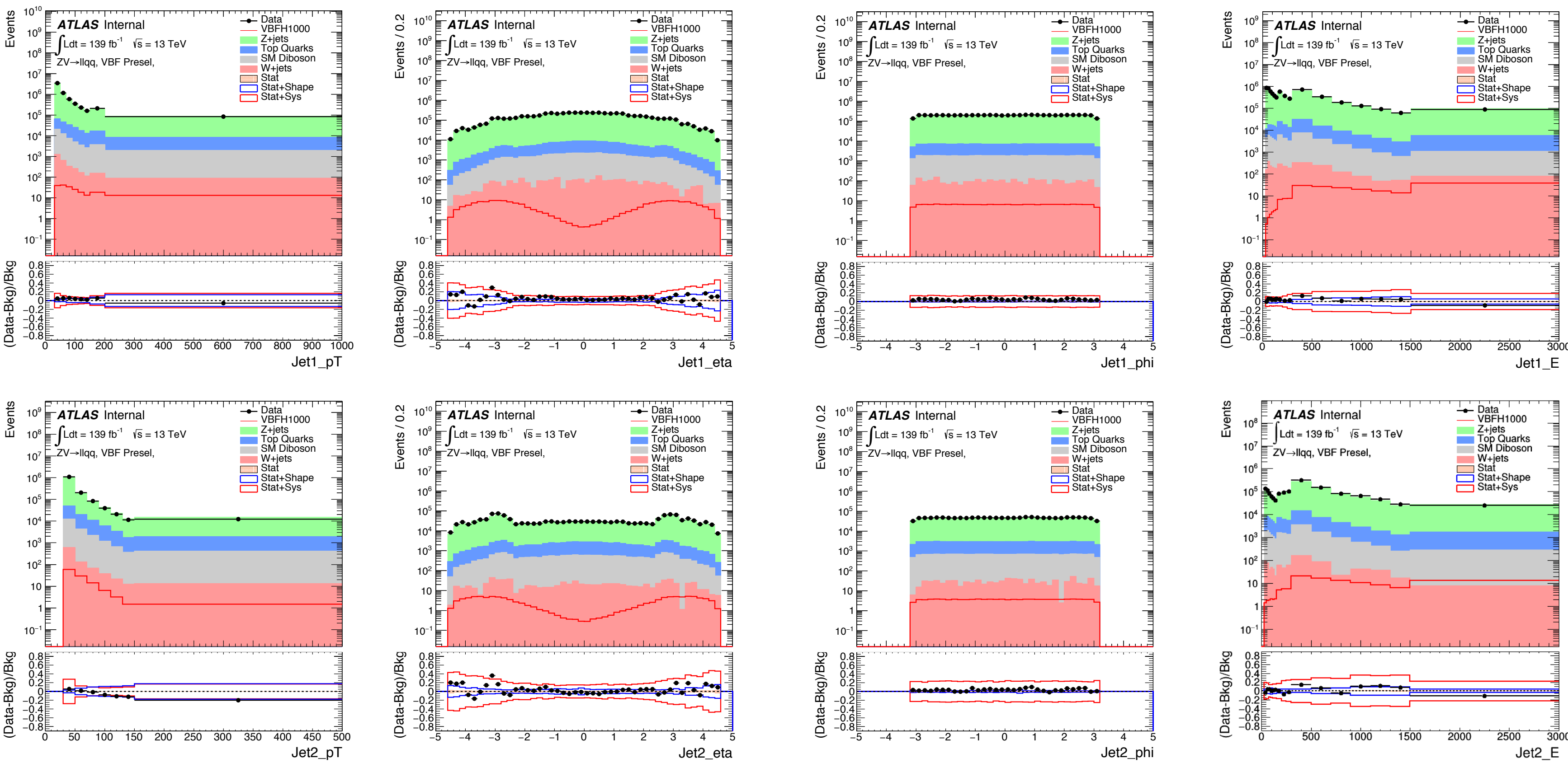
❖ "tagged" regions: 2 b-tagged jets;

❖ "untagged" regions: ≤ 1 b-tagged jets.

Data/MC validation plots

❖ Analysis validation and optimization studies.

Below, as instance, the 4-momentum of the first two jets after the di-lepton selection.



VBF-ggF/DY categorization

In the previous paper (using only 36,1 fb^{-1}), VBF/ggF categorization was performed in a cut-based way (https://arxiv.org/abs/1708.09638).

❖ **Search for 2 tag jets:**

$p_T > 30\ GeV$

$|\eta| < 4{,}5$

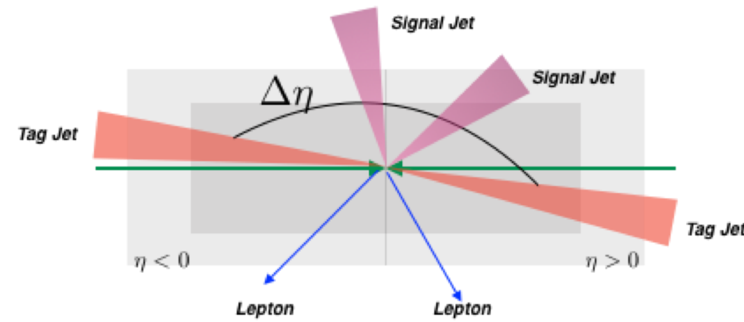
not b-tagged

$\eta(jet1)\times\eta(jet2) < 0$

❖ **VBF events:**

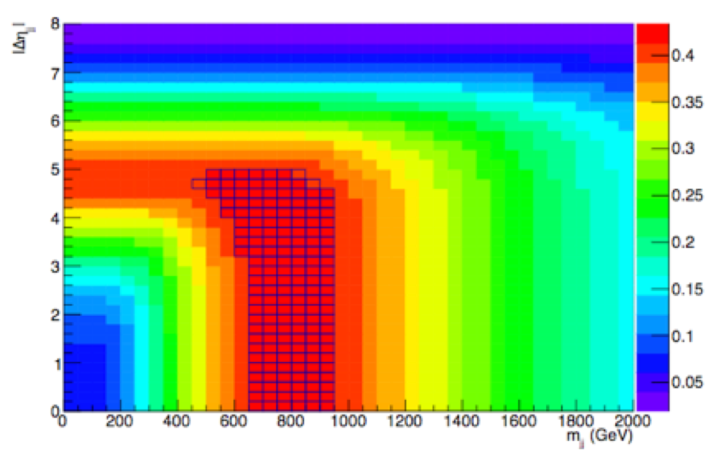
$m_{jj} > 700\ GeV$

$\Delta\eta_{jj} > 4{,}7$

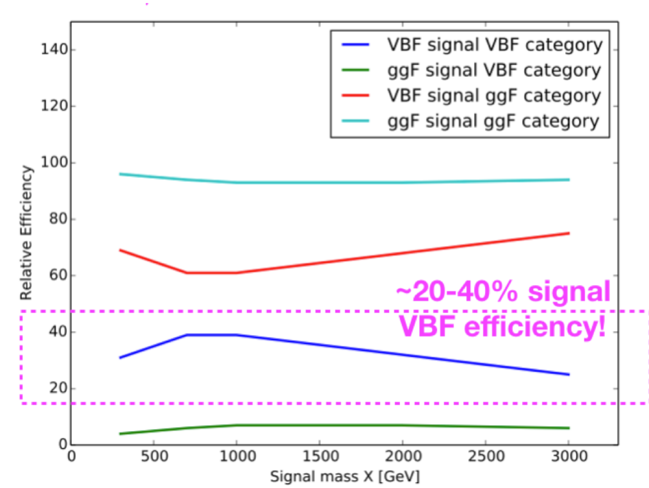


❖ **Two-dimensional optimisation with the estimator:**

$$\epsilon_{VBF}(1-\epsilon_{ggF})(1-\epsilon_{Bkg})$$

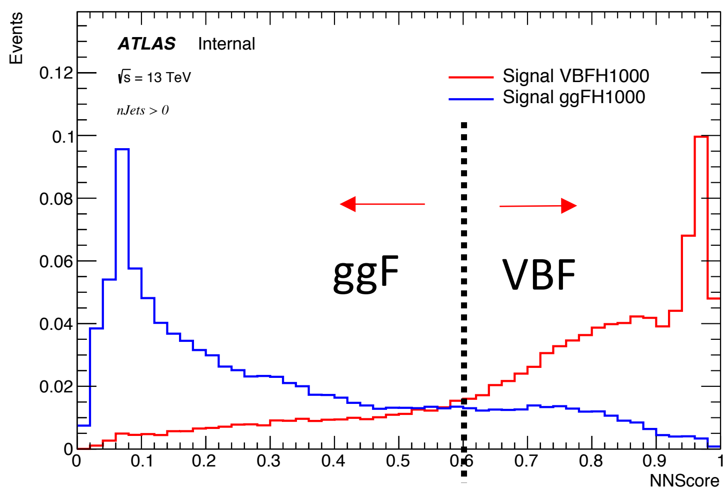


❖ **Low VBF efficiency**

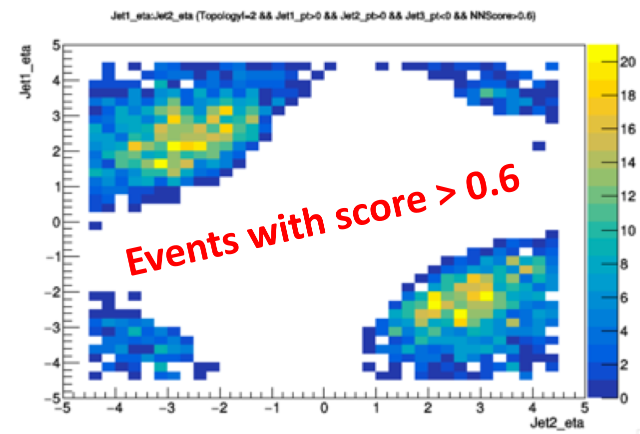


RNN to classify VBF events

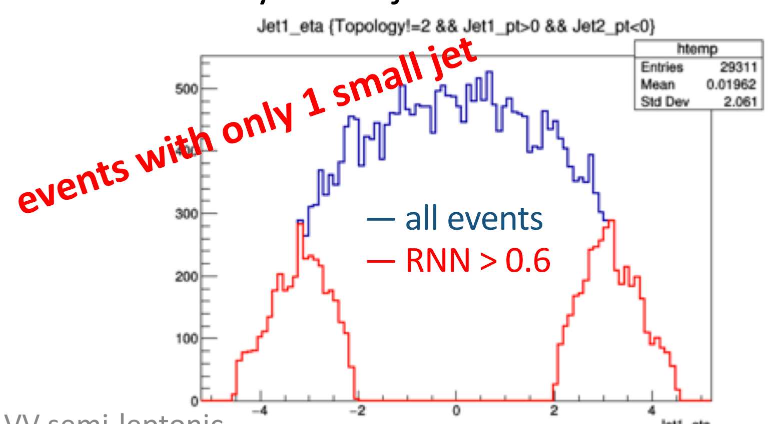
- Use a **Recurrent Neural Network (RNN)** to separate VBF events from ggF/ DY events **cutting on the score of the net**.
- **No more tag jets** but the **2 leading jets** in input to the net.



- The net learned that the VBF-like events have 2 Jets with opposite-sign eta.

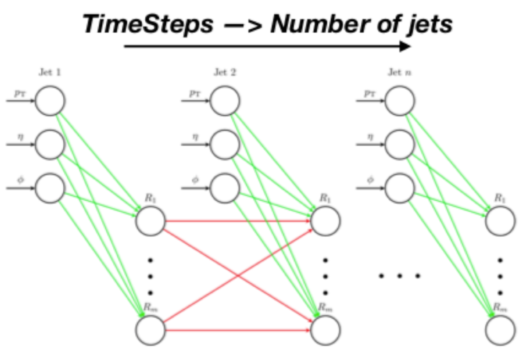


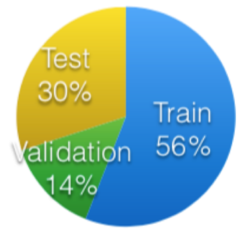
- Recovery of VBF events with only 1 VBF jet.



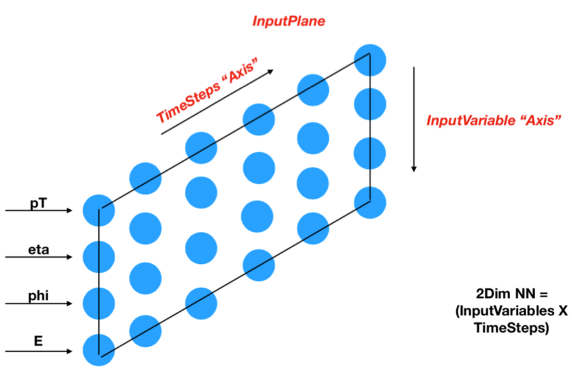
RNN setup

- Different numbers of jets in an event: no fixed number of variables in input, which is the best architecture to use?
- **Recurrent Neural Network**: provides variable-lenght input.
- An input set is repeated at each «**Time Step**».





Epochs	200 but EarlyStopping Callback after 10 unchanged iterations
Dropout	0,3
LSTM Layer Activation	tanh
Output Layer activation	Sigmoid
loss function	binary_crossentropy
LSTM layers	2
Optimiser	Adam



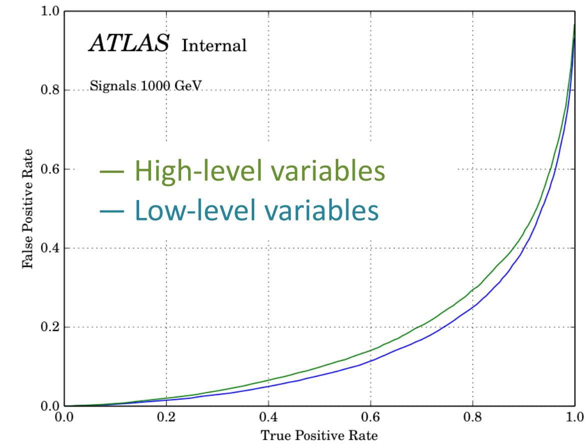
- The RNN-based methodology has entirely been developed by the Naples ATLAS group, with the technical support of prof. Acampora and prof. Vitiello.
- RNN now implemented (thanks to Antonio) inside the code of the entire semi-leptonic analysis group.

RNN setup

- **4 Input** variables: Jet **pT**, **eta**, **phi**, **E.**
- **Up to 2 jets** in each event.

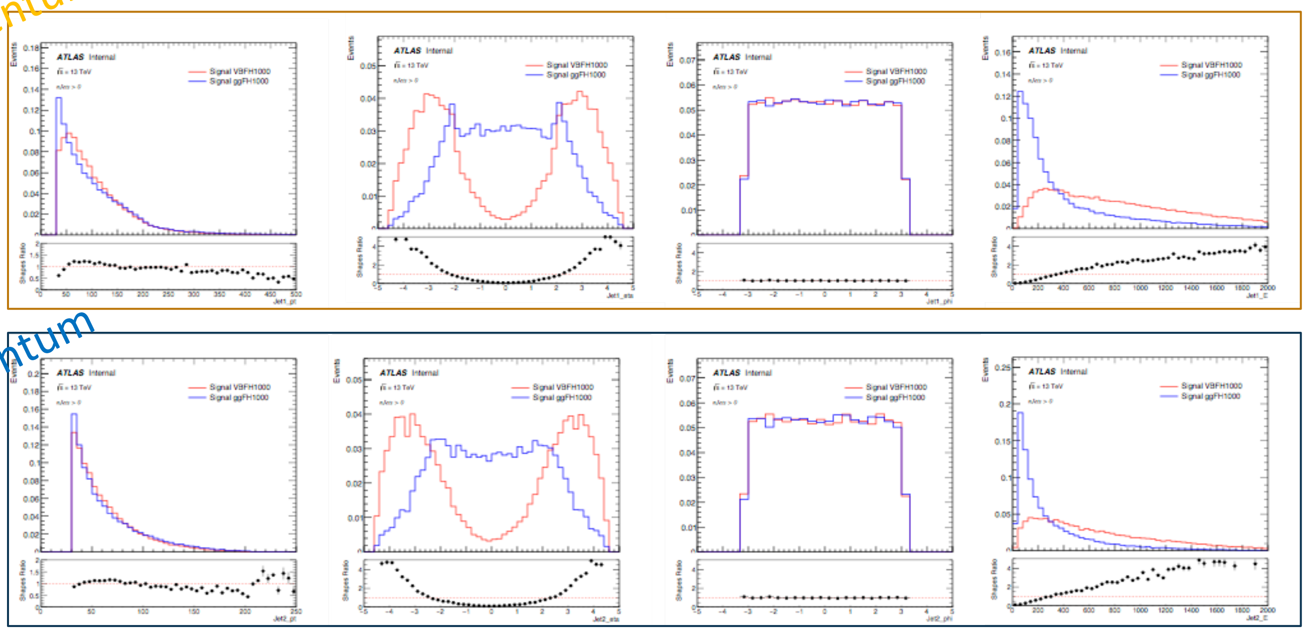
- Only low-level variables, without introducing variables that have a more “structured” physical meaning.

- **Train at dilepton selection** (after $Z \rightarrow ll$ search).
- **RNN trained on the VBF/ggF SPIN0 H 1TeV samples.**



Jet1 4-momentum

Jet2 4-momentum

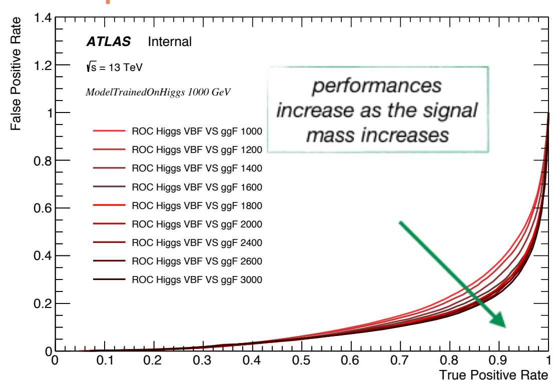


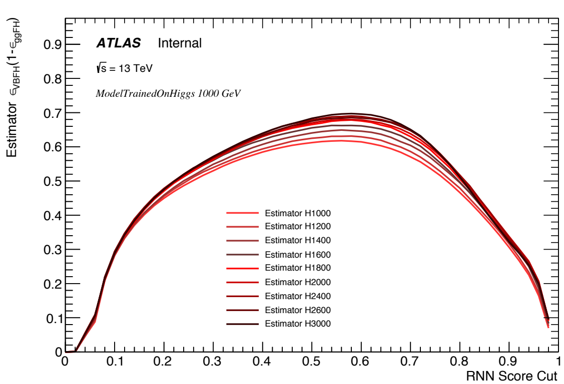
RNN performances

Train done using only a **1 TeV heavy Higgs** signal sample.

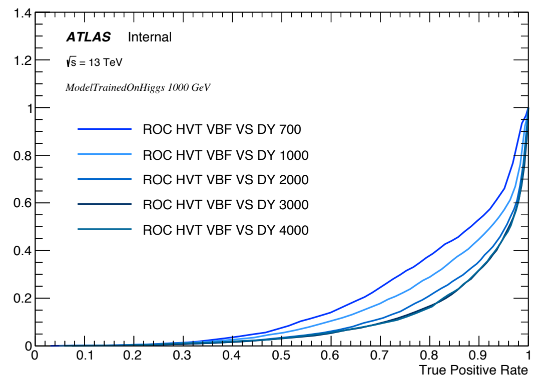
- ❖ Which are the performances on **other masses**?
- ❖ And on signals with **other spin hypothesis**?

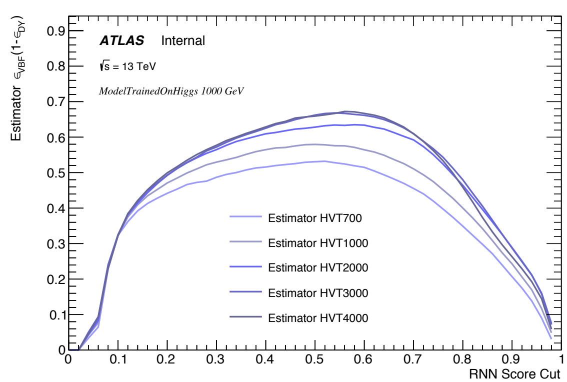
ROC curves and $\epsilon_{VBF}(1-\epsilon_{ggF})$ for spin 0 Higgs at different mass points .



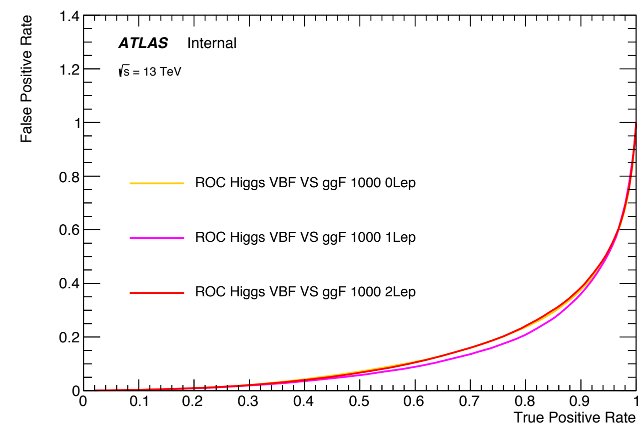


ROC curves and **$\epsilon_{VBF}(1-\epsilon_{DY})$ for spin 1 HVT bosons at different mass points.**





- ❖ RNN also used in the 0-lepton and 1-lepton analysis channels: **the approach is indipendent from the final states.**

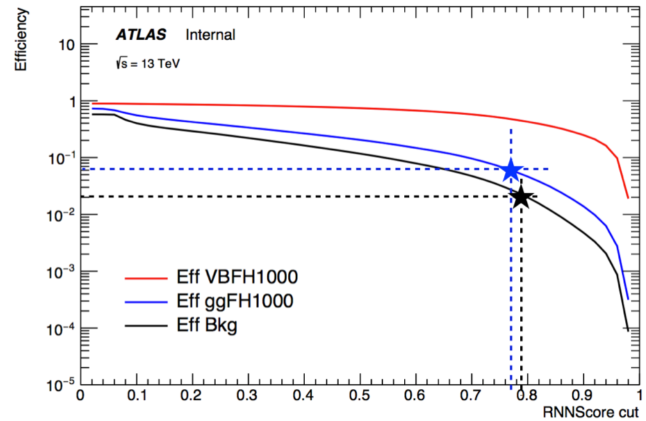


RNN performances

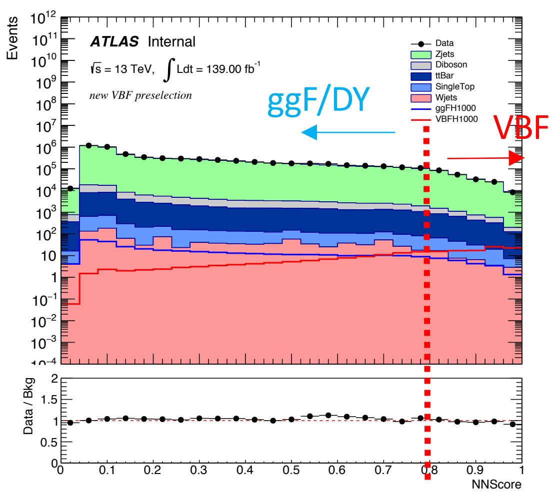
The Working Point (WP) was choosen in order to have the same bkg rejection as the old cut-based approach.

- Efficiencies in VBF category varying the score cut:

1000 GeV	Std categ	RNN Score >0.6	RNN Score >0.7	RNN Score >0.8
VBF Signal	37,8%	67,2%	57,6%	42,7%
ggF Signal	6,1%	14,8%	9,6%	4,6%



- Is VBF Event if RNN Score > 0.8



- Improvement of the confusion matrix.

Higgs 1000 GeV Cut-based selection

	VBF category	ggF category
VBF Signal	40%	60%
ggF Signal	6%	94%

Higgs 1000 GeV Score > 0.8

	VBF Category	ggF Category
VBF Signal	*0.43*	0.57
ggF Signal	0.05	0.95

SRs and CRs definitions

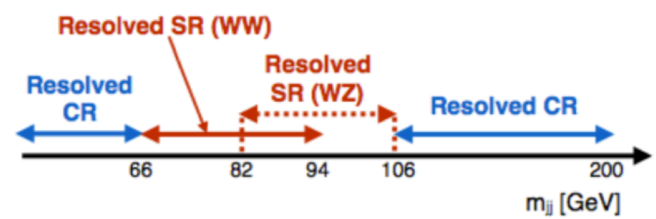
Two regimes of $V \to qq$ events reconstruction:

Merged ($V \to J$): for $M_X \geq 500$ GeV

- at least 1 large-R jet.
- $\frac{\min(llp_T, Jp_T)}{m_{ZV}} > 0.35\ (0.25)$ in ggF/DY (VBF) category;
- High-Purity and Low-Purity regions defined by Boson Tagging WPs (D2);

Resolved ($V \to jj$): for $M_X \leq 1000$ GeV

- at least 2 small-R jets
- $\frac{\min(llp_T, jjp_T)}{m_{ZV}} > 0.35\ (0.25)$ in ggF/DY (VBF) category;



- Boson mass sidebands used for the ZCRs.
- Additional Top CRs: two different-flavour leptons, eμ, with $76\ GeV < m_{e\mu} < 106\ GeV$ and two b-tagged jets with $50\ GeV < m_{2bjets} < 150\ GeV$.

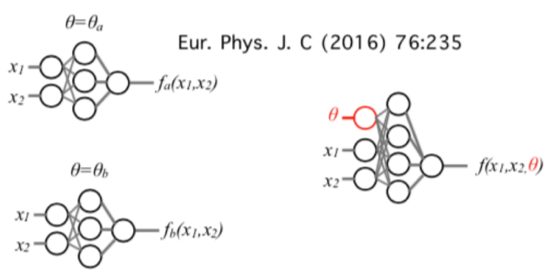
pDNN for Signal/Background discrimination

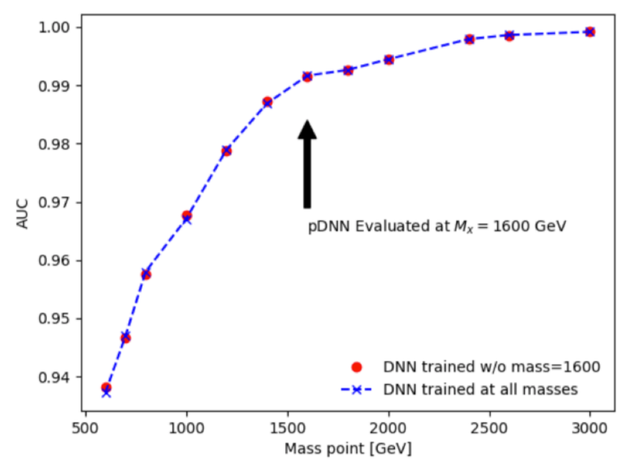
A **p**arametrized **D**eep **N**eural **N**etwork (pDNN) to better discriminate signal from background.

- ❖ improve stastistical significance (thanks to a deep neural network algorithm);
- ❖ parametrization technique to considerably simplify the model.

Not a single signal hypothesis but a large mass range to investigate, how to?

- ❖ An additional feature in input such as the real value of the signal mass for the signal and a random value among all the signal masses investigated for Bkg.





- ❖ Able to interpolate between masses;
- ❖ Is a way to train only 1 DNN on N mass points together.
- ❖ It cannot interpolate out of the mass range of the train.

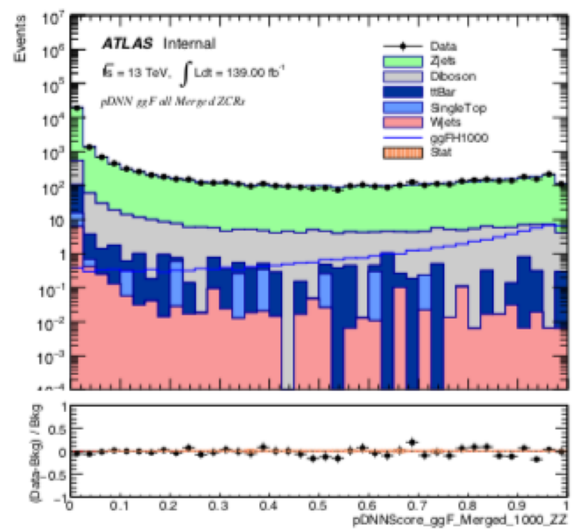
pDNN for Signal/Background discrimination

❖ **Variable used as input**:

Dilepton invariant mass and transverse momentum
4-momentum of each lepton
Number of small-R jets
4-momentum of the fat jet (merged)
4-momentum of di-jet system and of each signal jet (resolved)

❖ **Four** different **pDNN** models trained before the cut defining Signal-Control regions

pDNN Score

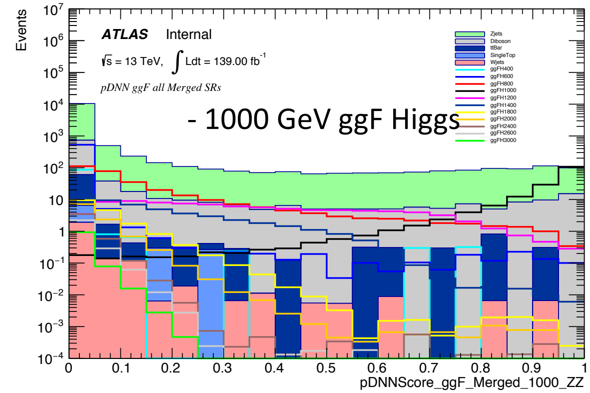


pDNN for Signal/Background discrimination

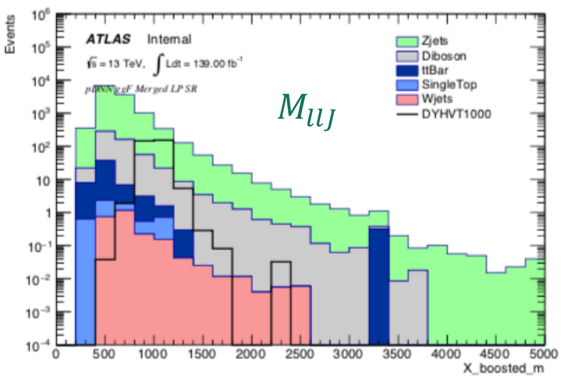
To use the pDNN one needs to choose the mass hypothesis.

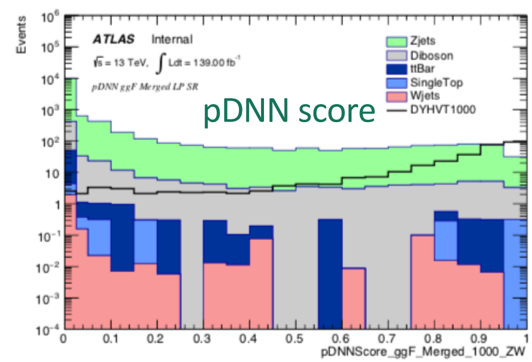
- The Score peaks to 1 only for the signal with the correct mass hypothesis, while the background and signals with other mass values are centered on 0.

pDNN Score
hypothesis of 1000 GeV



- **The idea is to use the pDNN score** as **final discriminant** (input of the fit), instead of the 4-body invariant mass.

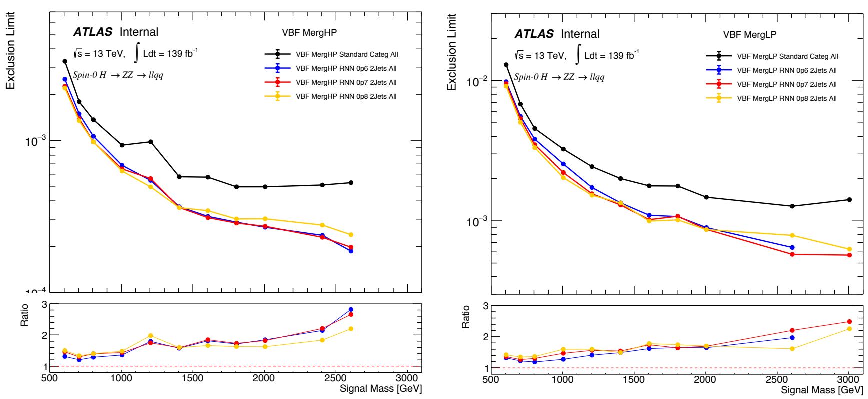




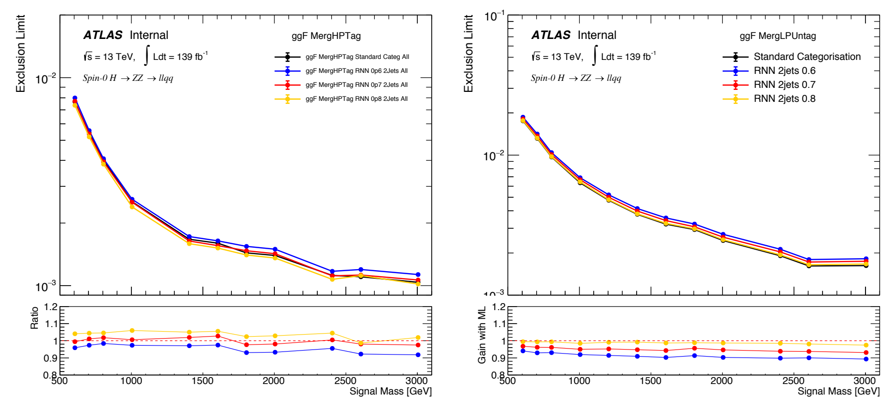
Expected limits on $\sigma(pp \rightarrow X) \times BR(X \rightarrow ZV \rightarrow llqq)$

- ❖ Up to date the analysis is still blinded.
- ❖ Calculated expected limits from MC simulations for different values of the RNN score cut: if there is no signal and data have the same trend of the SM background, how limits would be?

Expected upper limits for VBF Signals in VBF categories



Expected upper limits for ggF Signals in ggF category

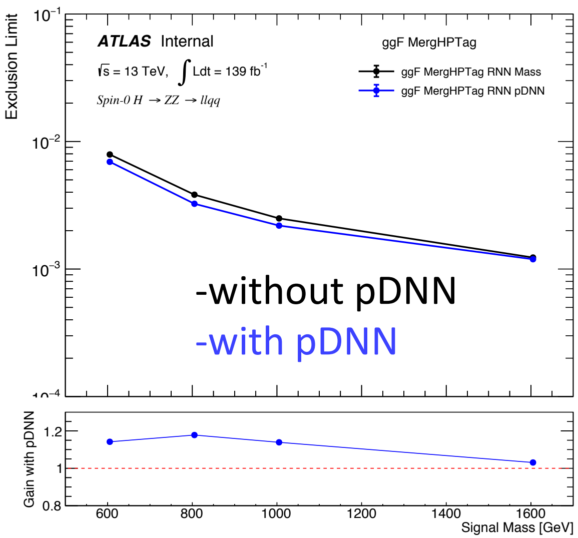


- ❖ Even RNN has the aim of improving the confusion matrix, also a **remarkable significance gain** in VBF category is observed.

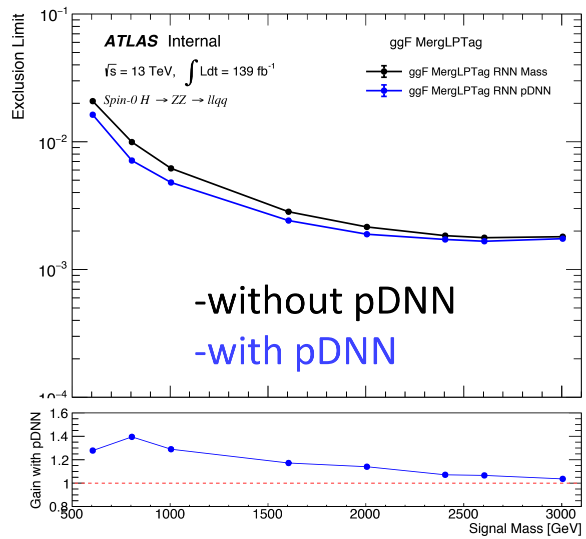
Expected limits on $\sigma(pp \to X)\times BR(X \to ZV \to llqq)$

- Expected upper limits with the pDNN score show a **sensitivity gain from 20% up to 50%**, according to the mass value.

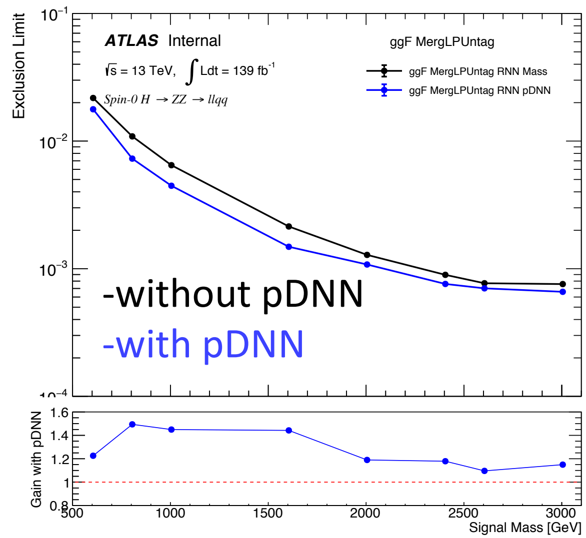
ggF merged HP Tagged



ggF merged LP Tagged



ggF merged LP Untagged



Conclusions

- Search for heavy resonances in semi-leptonic final states performed using $139\ fb^{-1}$ collected by ATLAS detector during the full Run-2;
- The main innovations introduced wrt the previous analysis (only $36{,}1\ fb^{-1}$ of 2015 and 2016 data) are Machine Learning algorithms:
  - a Recurrent Neural Network in order to improve the VBF efficiency in the VBF category: also sensitivity gain obtained, in particular in VBF regions;
  - a parameterized Deep Neural Network with the aim of better discriminating signal from background;
- Analysis still blinded, expected upper limits are remarkable improved with the respect to the cut-based approach.

Backup

Boson tagging

- In addition to the window cut on the jet mass distribution in merged regime, the **jet substructure variable D2** is used:

$$D_2^{\beta=1} = E_{CF3}\left(\frac{E_{CF1}}{E_{CF2}}\right)^3$$

$E_{CF\#}$: Energy correlation function

$$E_{CF1} = \sum_i p_{T,i}$$

$$E_{CF2} = \sum_{ij} p_{T,i} p_{T,j} \Delta R_{ij}$$

$$E_{CF3} = \sum_{iik} p_{T,i} p_{T,j} p_{T,k} \Delta R_{ij} \Delta R_{jk} \Delta R_{ki}$$

- Use of a new algorithm for jet reconstruction: **TCC Jets** (**Track Calo Clusters**) to improve the energy resolution of low energetic jets and the angular resolution for high boosted jets: both calorimeter and inner tracker informations.