Search for a heavy scalar X decaying to a scalar S and a Higgs boson in the $X \rightarrow SH \rightarrow bb\gamma\gamma$ channel with ATLAS Run-2 data

Maxime Fernoux

Supervisors : Elisabeth Petit and Arnaud Duperrin

Aix-Marseille University / CNRS / CPPM

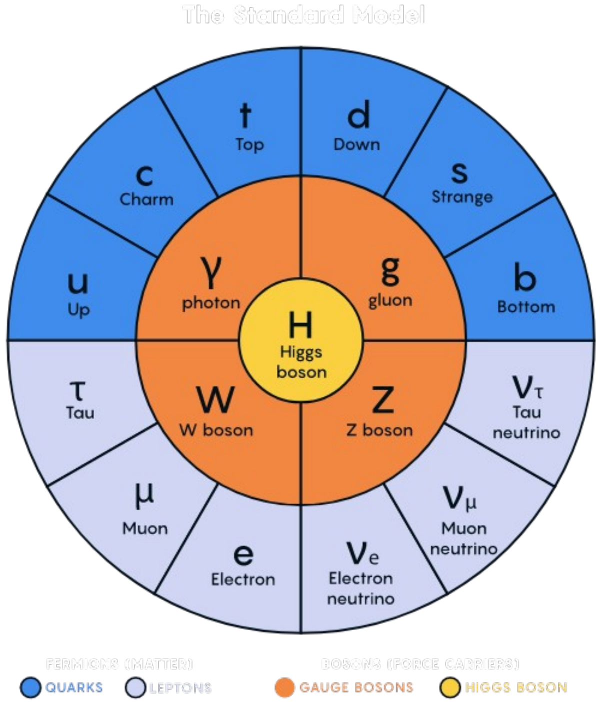
JRJC 2023 – Saint Jean de Monts

24.10.2023

Search for Beyond Standard Model physics

- SM is great but have some caveats :
  - Doesn't explain gravity
  - Doesn't explain neutrino oscillations and mass
  - No dark matter candidate

→ We need physics beyond standard model (BSM)

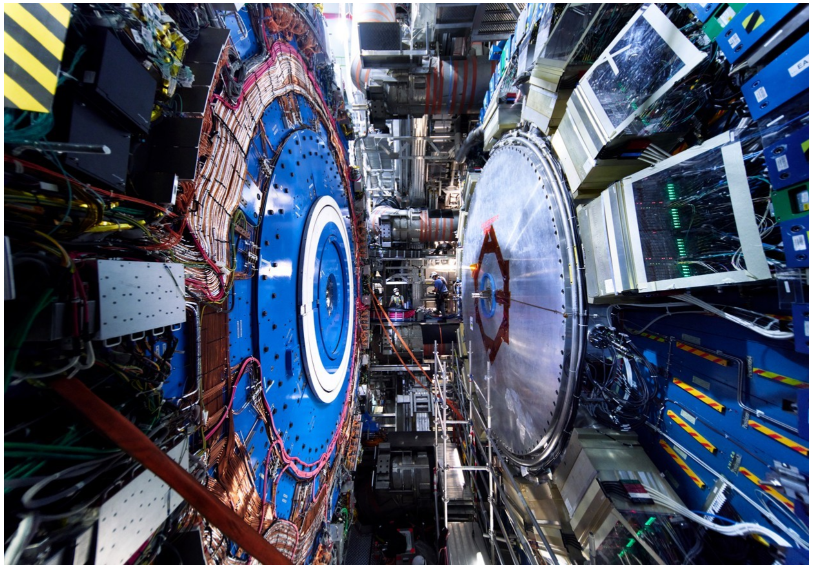


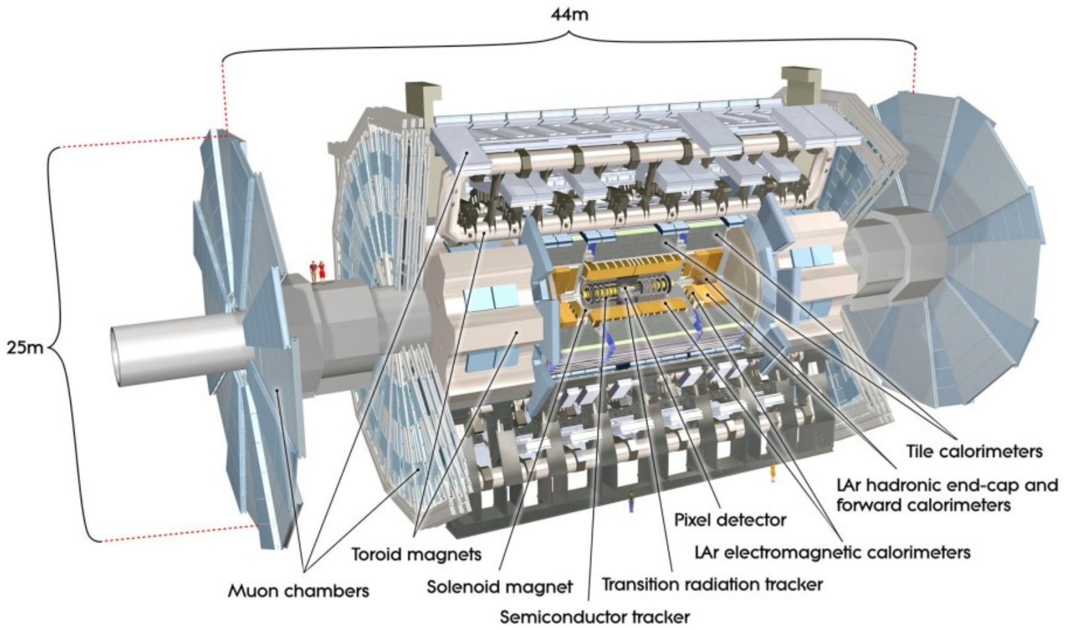
- Some BSM models predict additional scalar particles with different range of masses, notably in the Higgs sector
  - Double Higgs doublet model : 2HDM
  - Next-to-Minimal Supersymmetric Standard Model : NMSSM
  … etc

→ Could be at reach of LHC and Atlas !

ATLAS detector

- ATLAS is a multipurposed detector using proton-proton collisions at the Large Hadron Collider (LHC) located at CERN
  - Higgs boson discovered in 2012
  - Make precision measurements to detect deviations from the SM
  - Search for BSM particles, especially in the Higgs sector

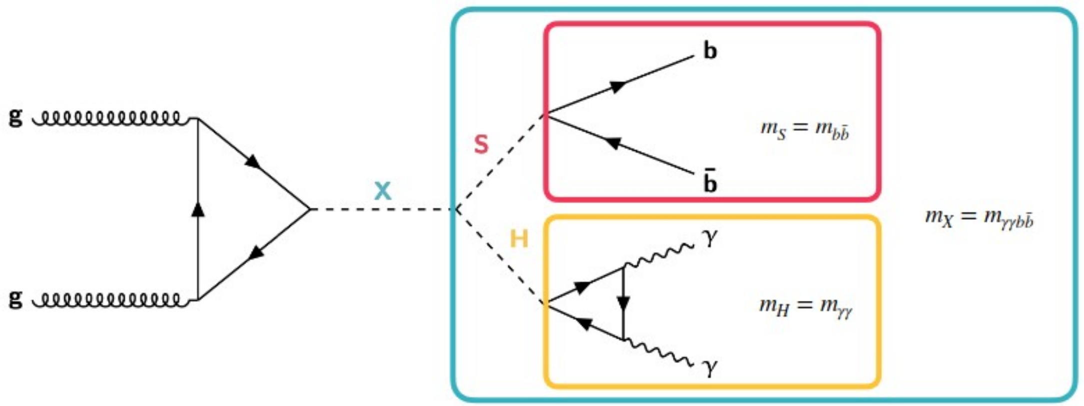


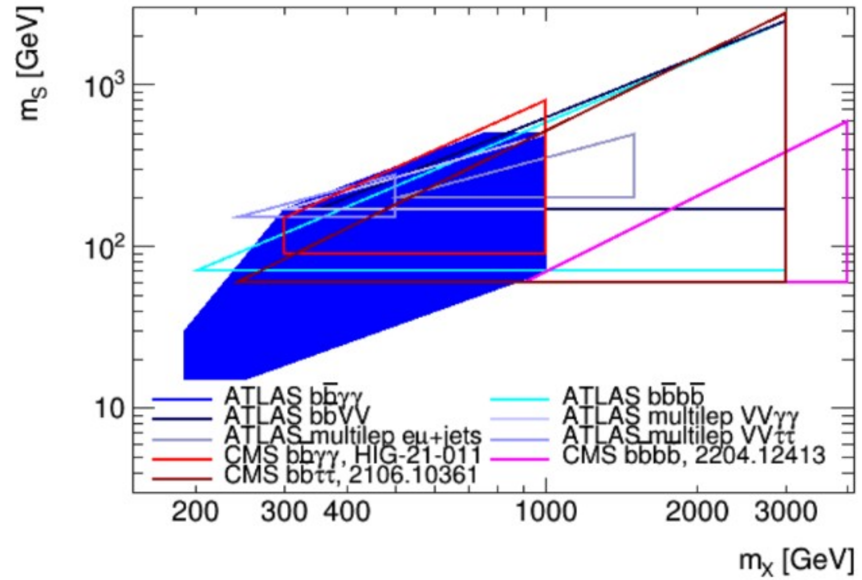


- This analysis uses data collected during Run-2 (from 2015 to 2018) with pp collisions at $\sqrt{s}$ = 13 TeV

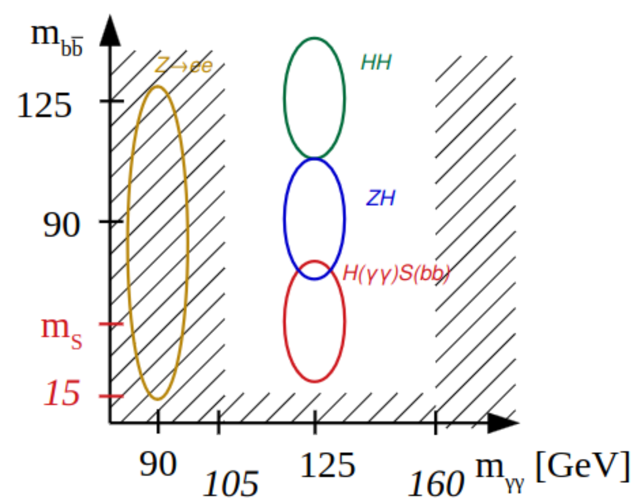
Targeted signal

- We search for a heavy scalar X decaying into a light scalar S and the SM Higgs where H → γγ and S → bb
- The search is **model-independent** and a wide range of mX, mS signal is targeted : 15 < mS < 500 GeV and 170 GeV < mX < 1 TeV



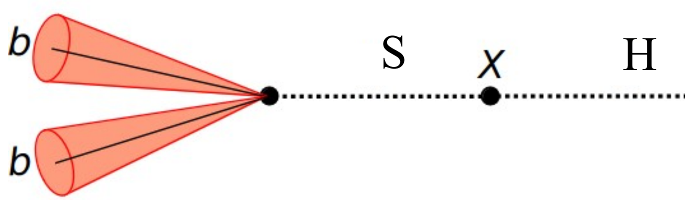


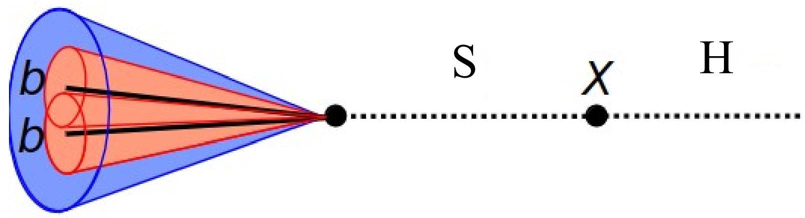
- Analysis is also heavily linked to HH → bbyy analysis Di-Higgs is a major goal of LHC physics program
  - Can test new strategies with this final state
  - Could also help to remove potential background to HH signal



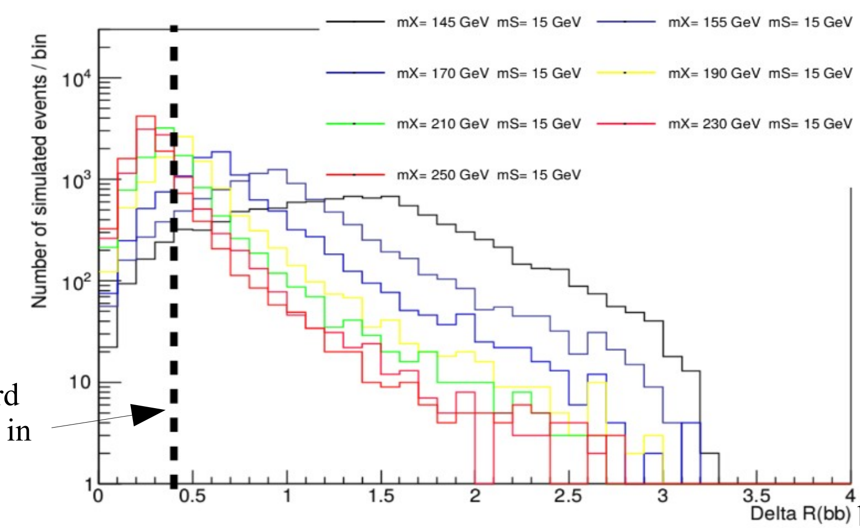
Different search regions

- A challenging situation arises when mS is much smaller than mX (mS/mX < around 0.1) : b-jets from the S decay are boosted and reconstructed as one b-tagged jet





→ We separate the search space in a **resolved region** with 2 b-tagged jets and a **merged region** with only one b-tagged jet



ΔR = 0.4 : standard angular size of jets in ATLAS

Angular distance between the 2 jets

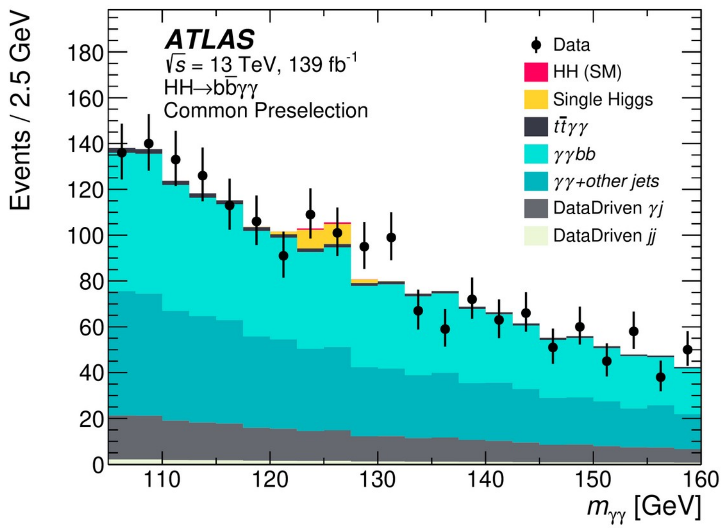
Event selection

- Selection

Table 4: The definitions of selections used in the analysis.

	2 b-tagged	1 b-tagged
Number of 'tight' and isolated photons	≥ 2	
$m_{\gamma\gamma}$ [GeV]	$\in [105, 160]$	
Number of leptons	$= 0$	
Number of central jets	$\in [2, 5]$	
Number of b-tagged jets @ 77% WP	$= 2$	$= 1$

Invariant photon mass distribution



- Predicted background yields (from theoretical cross sections) : main non-resonant background is γγ + jets, main resonant ones are ttH, ggH, ZH and also VBFH for 1 b-jet selection

2 b-tagged jets

	Selection
HH ggF+VBF	1.691 ± 0.004
ZH	3.691 ± 0.013
WH	0.207 ± 0.004
$VBFH$	0.685 ± 0.012
bbH	0.62 ± 0.023
ggH	5.453 ± 0.065
$tHjb$	0.969 ± 0.029
tWH	0.131 ± 0.005
ttH	8.313 ± 0.014
$Z(\to qq)\gamma\gamma$	20.345 ± 0.303
$t\bar{t}\gamma\gamma$	28.69 ± 0.109
$\gamma\gamma$+jets	1418.32 ± 4.596
Total SM	1489.116 ± 4.608

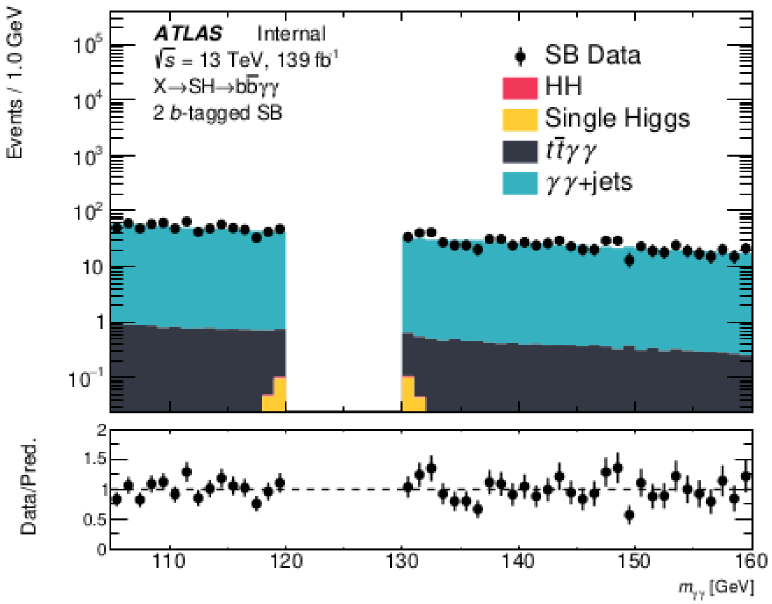
1 b-tagged jet

	Selection
HH ggF+VBF	1.827 ± 0.004
WH	5.97 ± 0.021
$VBFH$	8.333 ± 0.042
ZH	5.941 ± 0.015
bbH	2.973 ± 0.047
ggH	48.532 ± 0.202
$ggZH$	1.581 ± 0.01
$tHjb$	2.681 ± 0.05
tWH	0.572 ± 0.01
ttH	11.681 ± 0.017
$Z(\to qq)\gamma\gamma$	53.728 ± 0.8
$t\bar{t}\gamma\gamma$	49.78 ± 0.142
$\gamma\gamma$+jets	16298.8 ± 16.031
Total SM	16492.398 ± 16.053

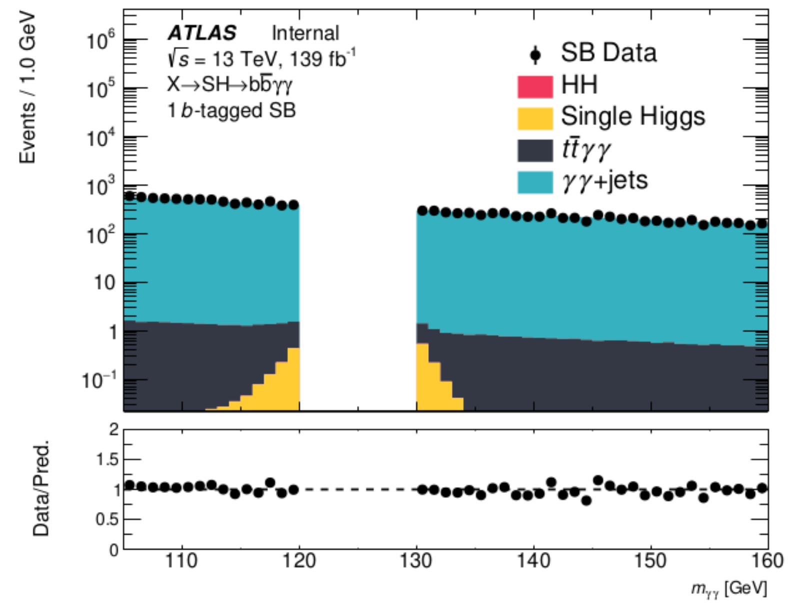
Analysis recap – Strategy

- The $m_{\gamma\gamma}$ distribution is used to split the events into a SR with $120 < m_{\gamma\gamma} < 130$ GeV and a sideband control region (CR)

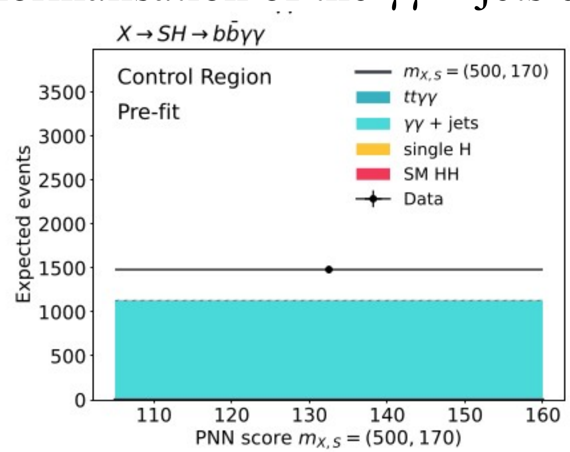
2 b-tagged jets



1 b-tagged jet

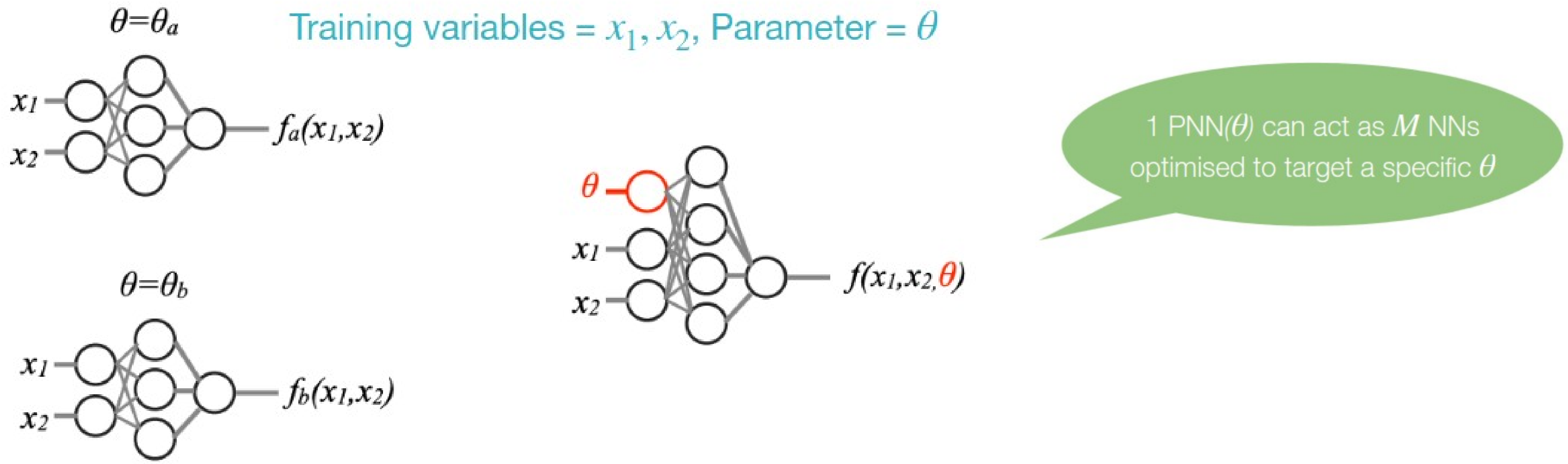


- The CR allows to correct the normalisation of the γγ + jets events using true data



PNN discriminant

- A parametrised Neural Network (PNN) is used as discriminant in the SR. It is trained with simulated events that pass each selection (SR + SB)



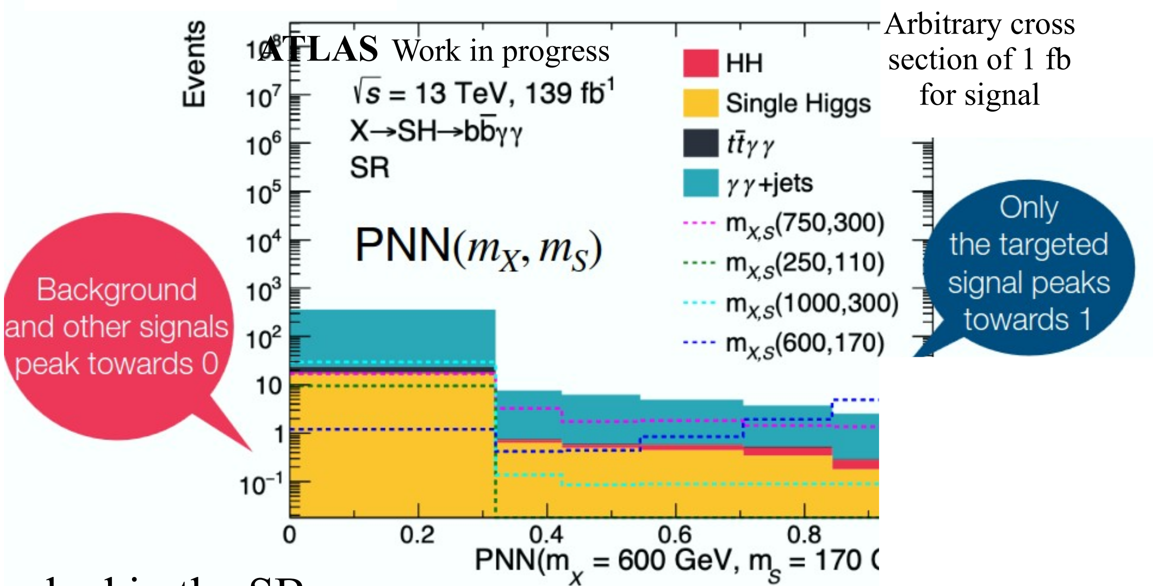
- Two separate PNNs :
  - **2 b-tagged** :
    Parameter θ = mX, mS
    Training samples : signal, ttH, ggH, ZH and γγ + jets backgrounds (no HH – too signal-like and confuses the network)
    Training variables : m_{jj} and $m^*_{\gamma\gamma jj} = m_{\gamma\gamma jj} - (m_{\gamma\gamma} - 125\text{ GeV})$
  - **1 b-tagged** : target low mS/mX values
    Parameter θ = mX
    Training samples : signal, VBFH, HH, ttH, ggH, ZH and γγ + jets backgrounds
    Training variables : b-jet p_T and $m^*_{\gamma\gamma j} = m_{\gamma\gamma j} - (m_{\gamma\gamma} - 125\text{ GeV})$

PNN distribution

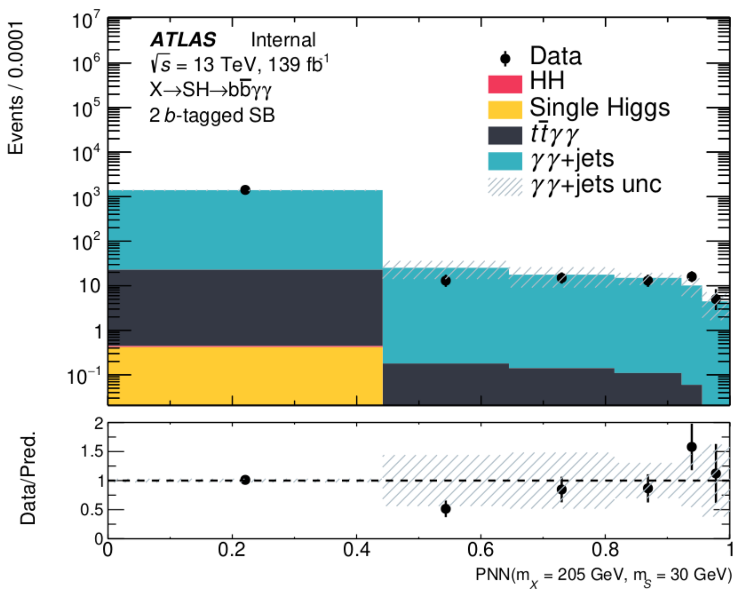
- PNN shapes of backgrounds comes from MC samples

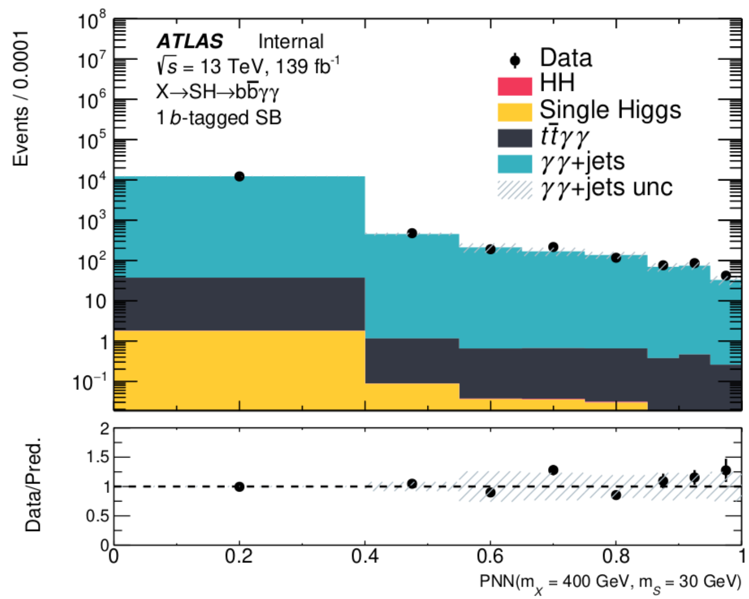
- Final results are computed with a binned log-likelihood fit on the PNN distribution

- Example of PNN distribution
A constrain on the binning is to have at least 1 background event in every bin



- Consistency between data and MC is checked in the SB

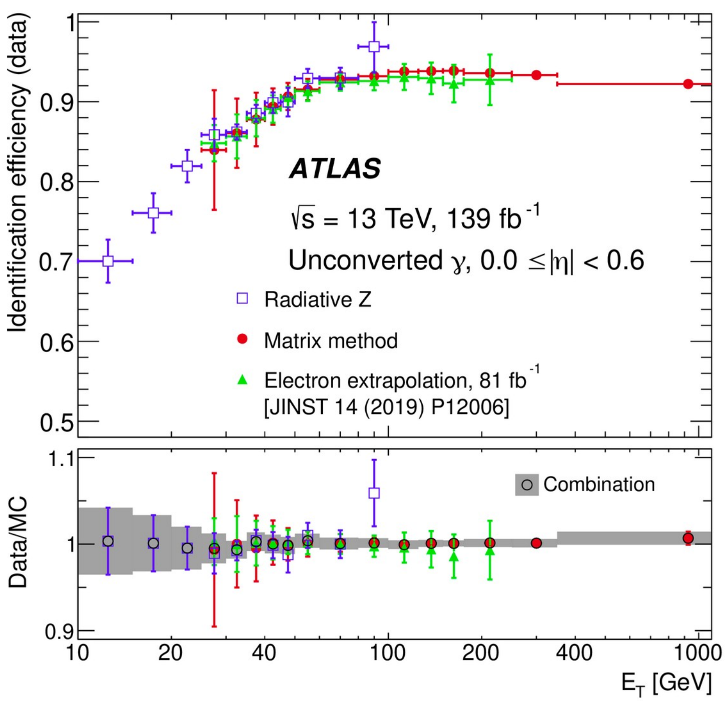




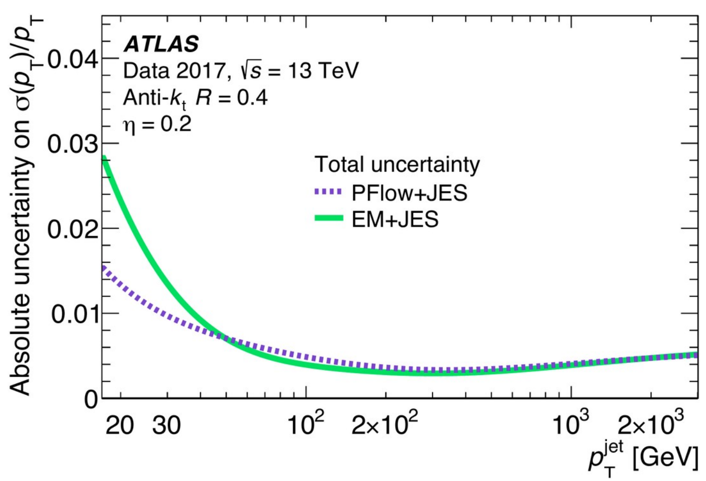
Experimental Systematics

- Physics is an experimental science → we have uncertainties affecting the measures
- Eventually at our analysis level it can have various impacts :
  - Particle identification can change the number of events in the CR and SR regions
  - Flavour tagging change the number of b-tagged jets
  - pT and energy resolution can change the position and width of the peak in the $m_{\gamma\gamma}$, m_{bb} and $m_{bb\gamma\gamma}$ distributions and afterwards the shape of the PNN distribution

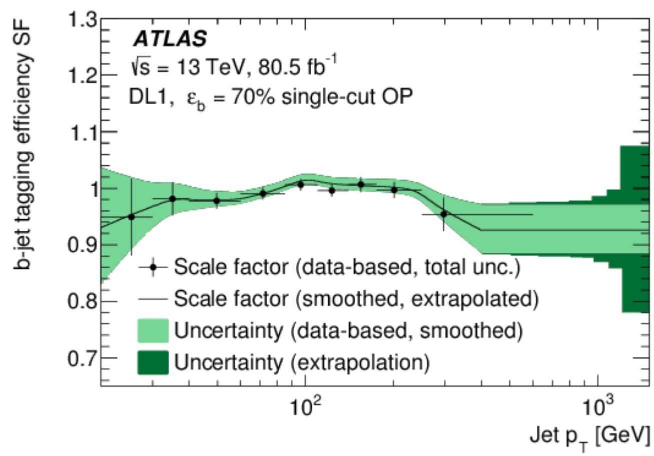
Photon identification efficiency as a function of E_T



Photon pT resolution as a function of p_T



b-jet tagging efficiency as a function of p_T



Experimental Systematics

- In the analysis framework experimental systematics are studied through MC samples at +/- 1σ away from nominal values from each effect
- They have been produced for major backgrounds only : ttH, ZH, ggH, ggHH and VBFHH, ggZH, γγ + jets and Z(bb/qq)γγ
- Systematics are treated as nuisance parameters (NP) in the fit → 47 in total !
  Like the number of slices of a prefou from Vendee

- Pile-up modeling, → PRW_DATASF → JET_Pileup_OffsetMu → JET_Pileup_PtTerm
- Diphoton trigger efficiency, → PH_EFF_TRIGGER_Uncertainty
- Photon ID efficiency → PH_EFF_ID_Uncertainty
- Photon isolation efficiency → PH_EFF_ISO_Uncertainty
- Photon energy scale (1 NP) → EG_SCALE_ALL
- Photon energy resolution (1 NP) → EG_RESOLUTION_ALL
- Jet energy scale (CategoryReduction scheme) → Jet (17 NP)
- Jet energy resolution (FullJER scheme) → JET_JER_EffectiveNP_* (11 NP)
- Jet vertex tagger efficiency → JET_JvtEfficiency
- Flavour tagging efficiencies (SF Eigen, Medium Reduction Scheme) → Flavour Tagging (11 NP)

Jet

- JET_EffectiveNP_Detector* (2 NP)
- JET_EffectiveNP_Mixed* (3 NP)
- JET_EffectiveNP_Modelling* (4 NP)
- JET_EffectiveNP_Statistical* (6 NP)
- JET_EtaIntercalibration_Modelling
- JET_EtaIntercalibration_NonClosure_2018data

Flavour Tagging

- FT_EFF_Eigen_B_* (3 NP)
- FT_EFF_Eigen_C_* (3 NP)
- FT_EFF_Eigen_Light_* (5 NP)

Experimental Systematics

- In the analysis framework experimental systematics are studied through MC samples at +/- 1σ away from nominal values from each effect
- They have been produced for major backgrounds only : ttH, ZH, ggH, ggHH and VBFHH, ggZH, γγ + jets and Z(bb/qq)γγ
- Systematics are treated as nuisance parameters (NP) in the fit → 47 in total !
  Like the number of slices of a prefou from Vendee

- Pile-up r
- Diphoto
- Photon I
- Photon i
- Photon e
- Photon e
- Jet energ
- Jet energ
- Jet verte
- Flavour

ata

Experimental Systematics – Yield change

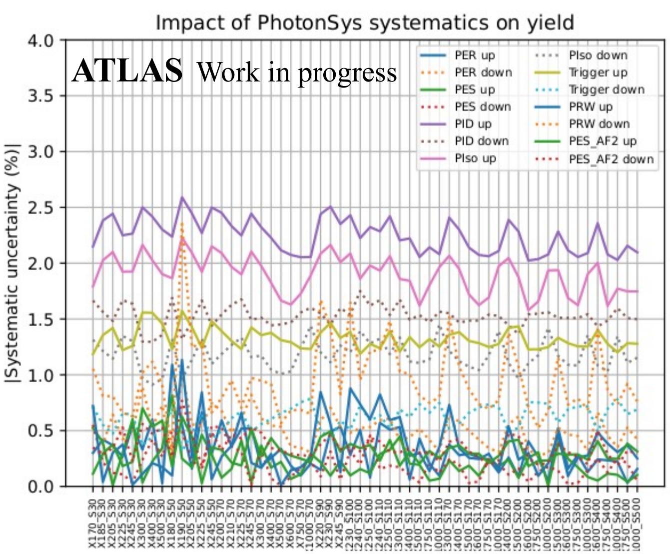
- Systematics change both the yields and the shape of the PNN distribution of the samples

Table 18: Uncertainty [in %] on the yield for backgrounds in the 2 b-tagged category.

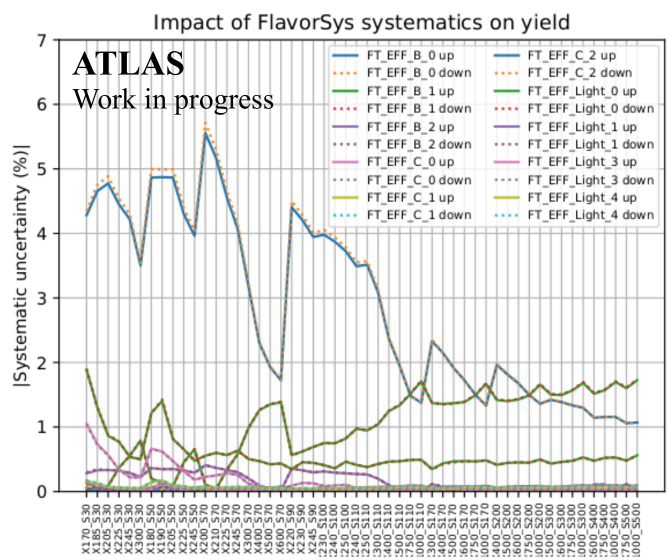
Main uncertainties

	Source	Yield uncertainty [%] ttH	ZH	ggHH	ggH	ggZH	tHjb	VBFHH	Zqqyy	Zbbyy
Event-based	Photon Trigger	1.01	1.02	1.00	1.04	0.98	1.03	1.02	1.00	1.25
	Pile-up reweighting	0.88	0.76	0.56	0.43	0.60	0.99	0.60	3.69	1.15
Photon	Photon Energy Resolution	0.42	0.42	0.34	0.42	0.43	0.63	0.43	2.79	0.78
	Photon Energy Scale	0.17	0.18	0.12	0.07	0.11	0.18	0.24	14.53	0.90
	Photon ID	1.59	1.61	1.44	1.64	1.49	1.60	1.59	1.72	1.96
	Photon Isolation	1.55	1.57	1.45	1.60	1.46	1.59	1.59	1.27	1.88
Jet	Jet Energy Scale	1.36	0.94	0.55	1.81	0.74	0.76	0.72	5.30	1.09
	Jet Energy Resolution	7.33	4.60	2.91	7.50	3.36	4.88	3.08	0.68	5.37
Flavour-tagging	b-jet efficiency	2.07	2.99	2.51	3.05	2.55	2.30	2.83	0.10	3.36
	c-jet efficiency	0.40	0.71	0.06	1.68	0.60	0.92	0.07	13.12	0.22
	light-jet efficiency	0.79	0.38	0.40	2.72	0.51	0.90	0.42	1.85	0.48

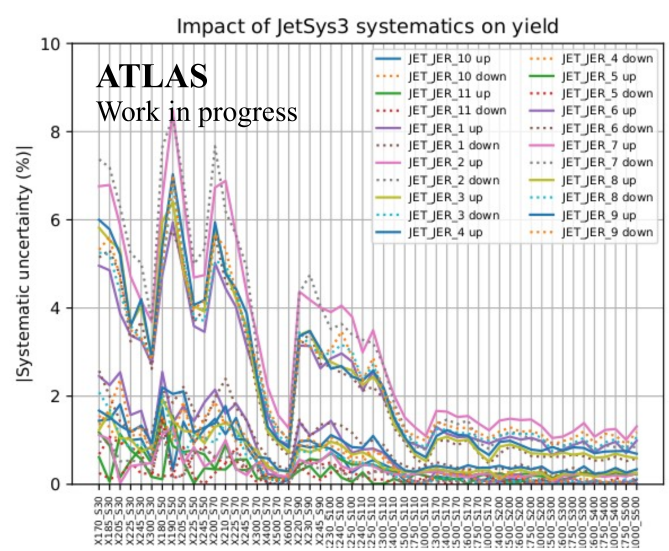
- For signal, yields changes are dependant on mS and mX
  Jet energy resolution systematics are the most important (but below 10%)



(b) 2 b-tagged

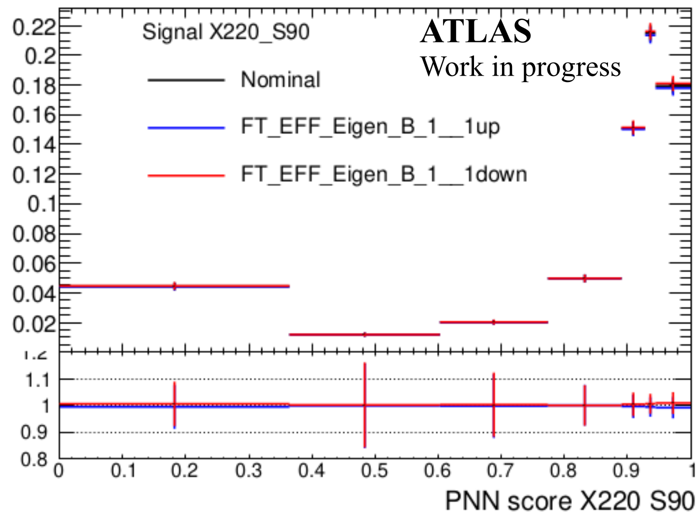


(b) 2 b-tagged

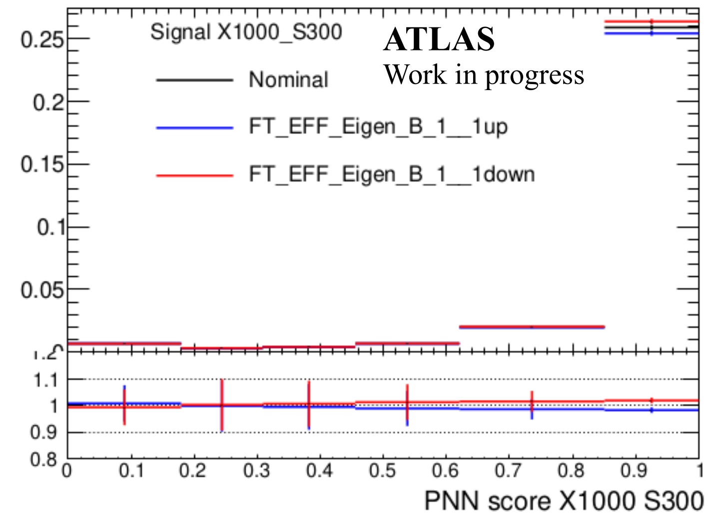


(b) 2 b-tagged

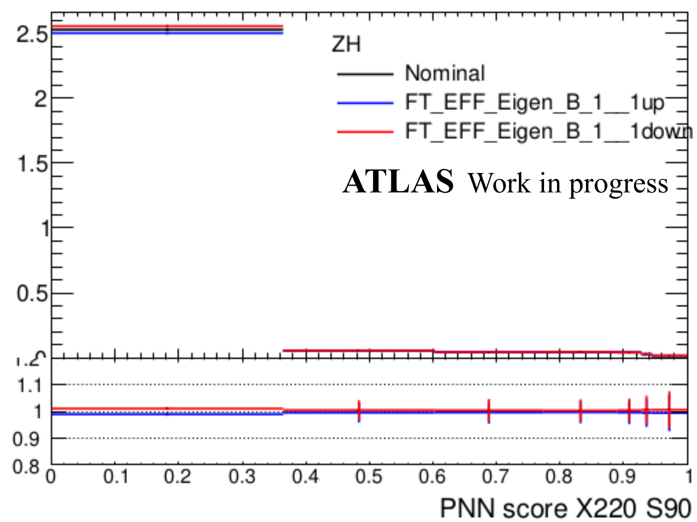
Experimental Systematics – Shape change



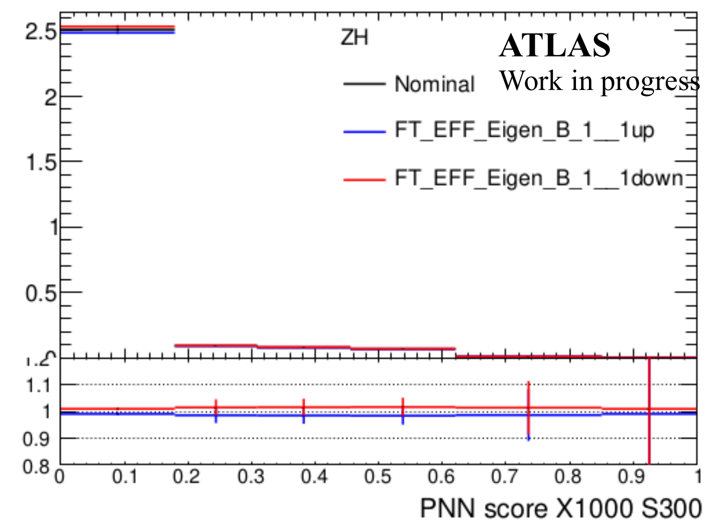
(a) SH signal, $m_X = 220$ GeV, $m_S = 90$ GeV



(b) SH signal, $m_X = 1000$ GeV, $m_S = 300$ GeV



(e) ZH, $m_X = 220$ GeV, $m_S = 90$ GeV



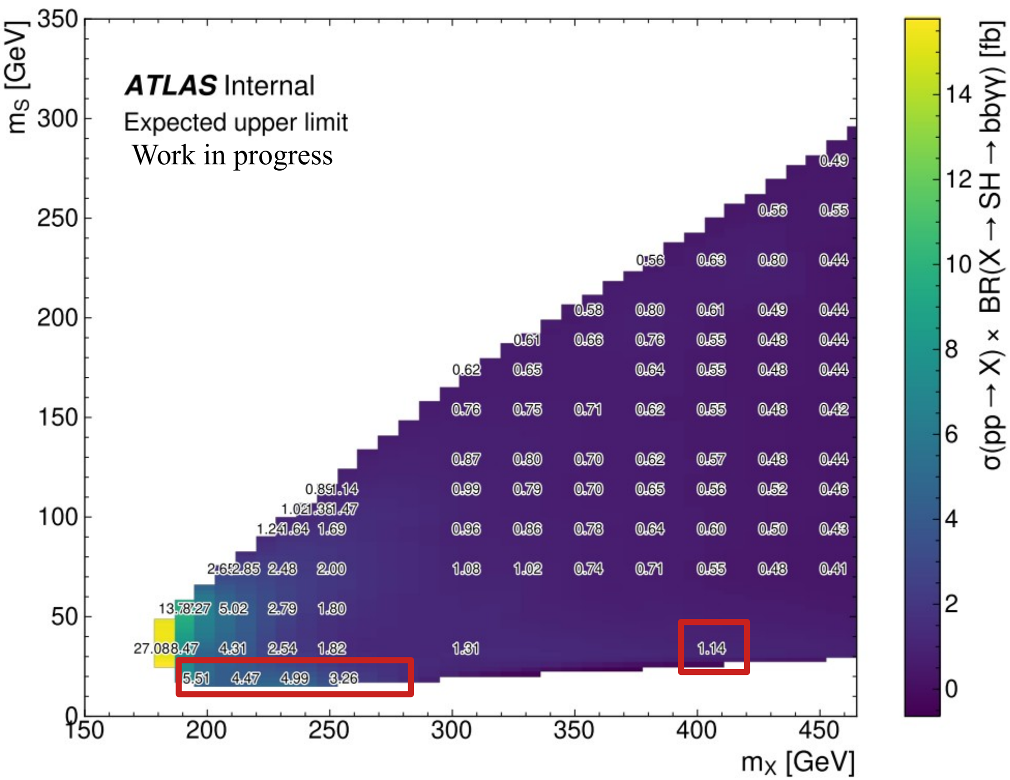
(f) ZH, $m_X = 1000$ GeV, $m_S = 300$ GeV

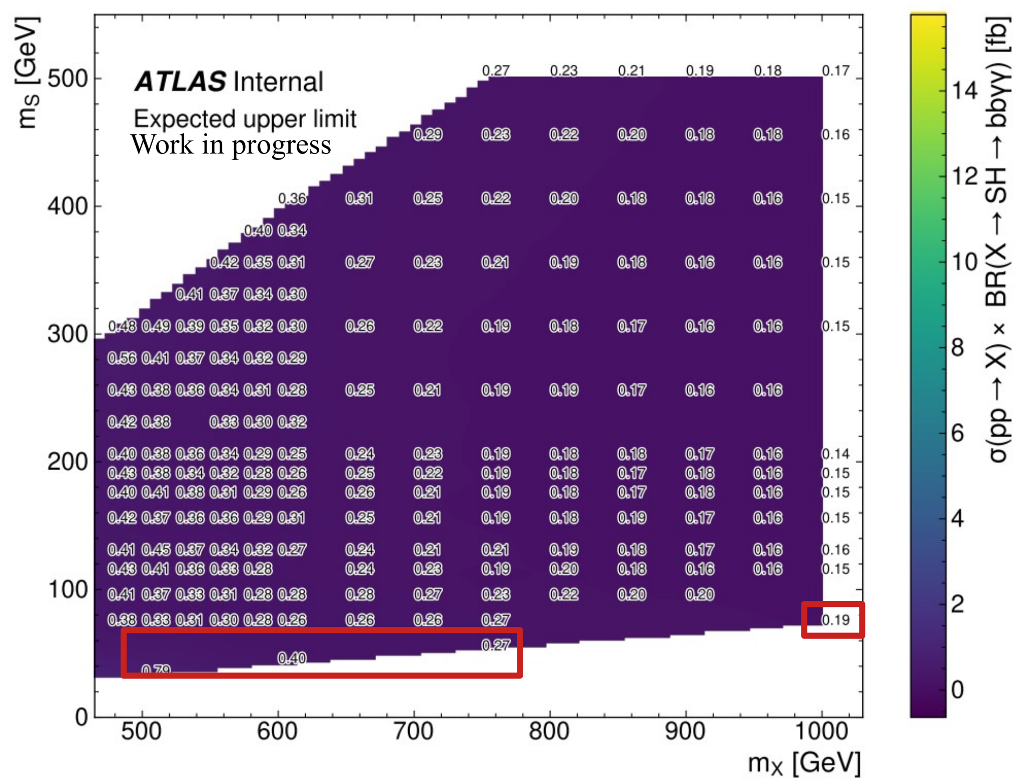
- NB : for $\gamma\gamma$ + jets, only shape changes are used as normalisation is imposed by the sideband

Blinded expected limits

- If no signal is observed, the goal of the analysis is to set upper limits on the cross section of the $X \to SH$ signal in the $bb\gamma\gamma$ final state
- I can only show blinded results (i.e using Monte-Carlo and not true ATLAS data) Not final results but gives an idea of the analysis sensibility

1 b-jet points

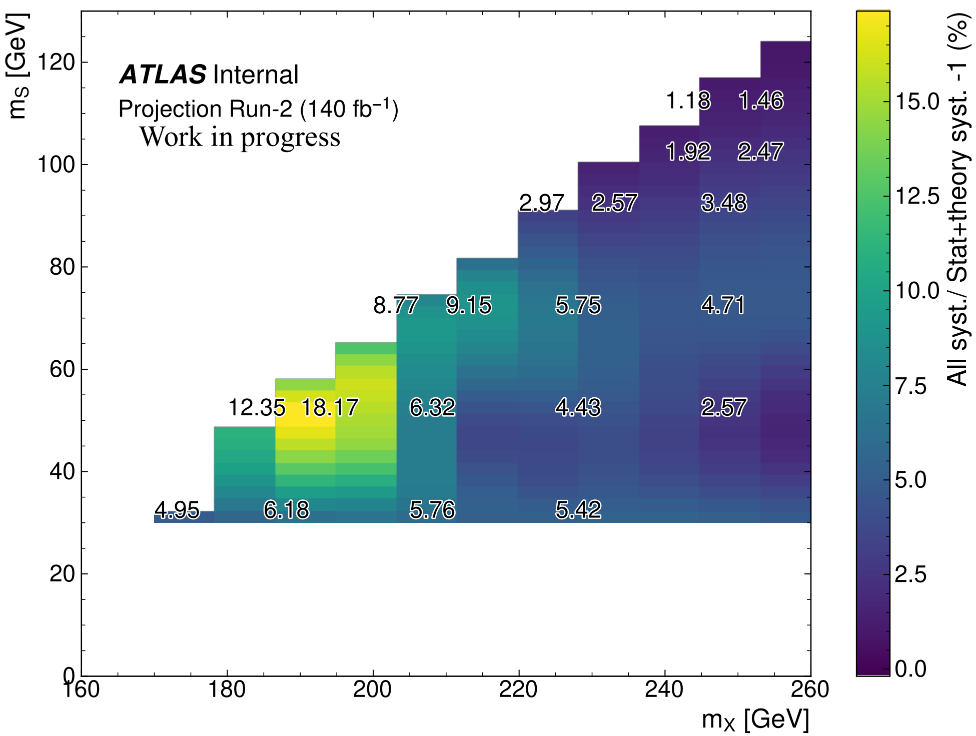


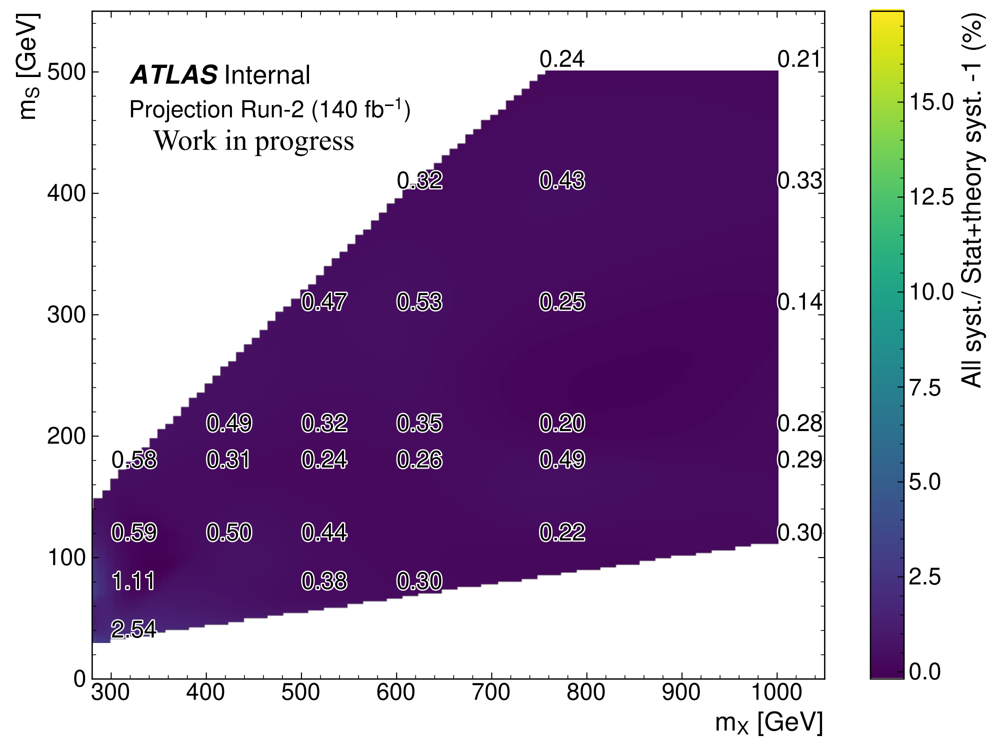


- Limits range from 0.15 to 30 fb
Sensibility is better in high mass region

Impact of experimental systematics uncertainty

- Here we plot the ratio between blinded expected limits and the limits obtained without taking into account experimental systematics to check their impact :

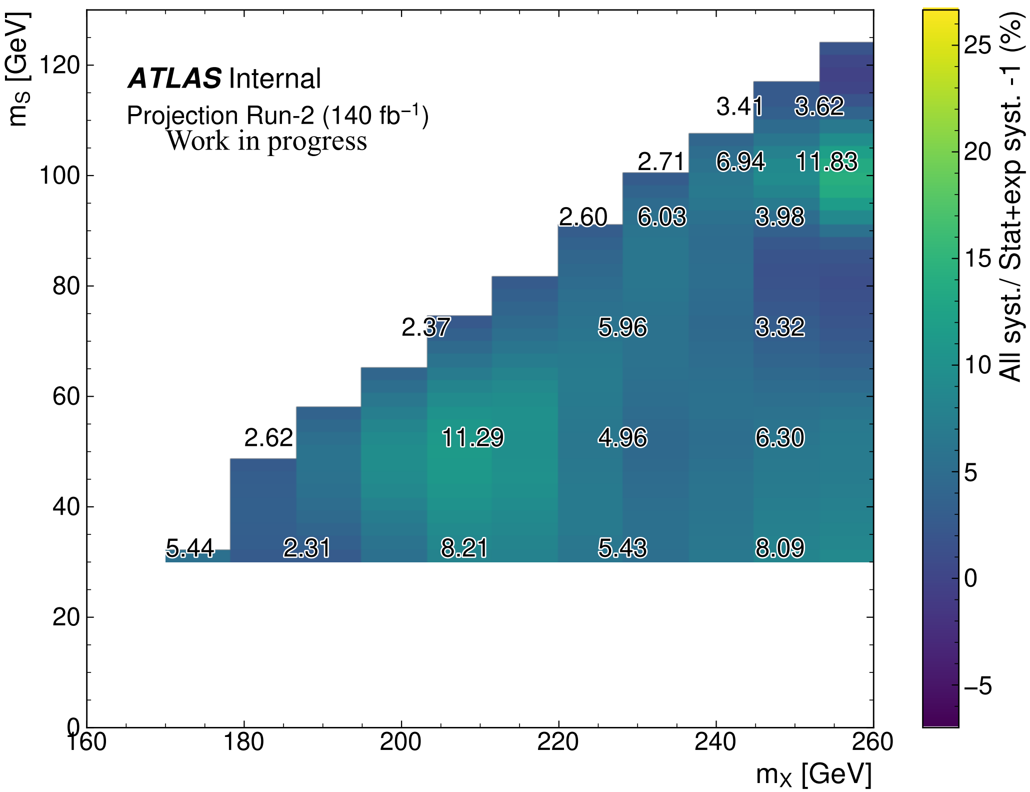


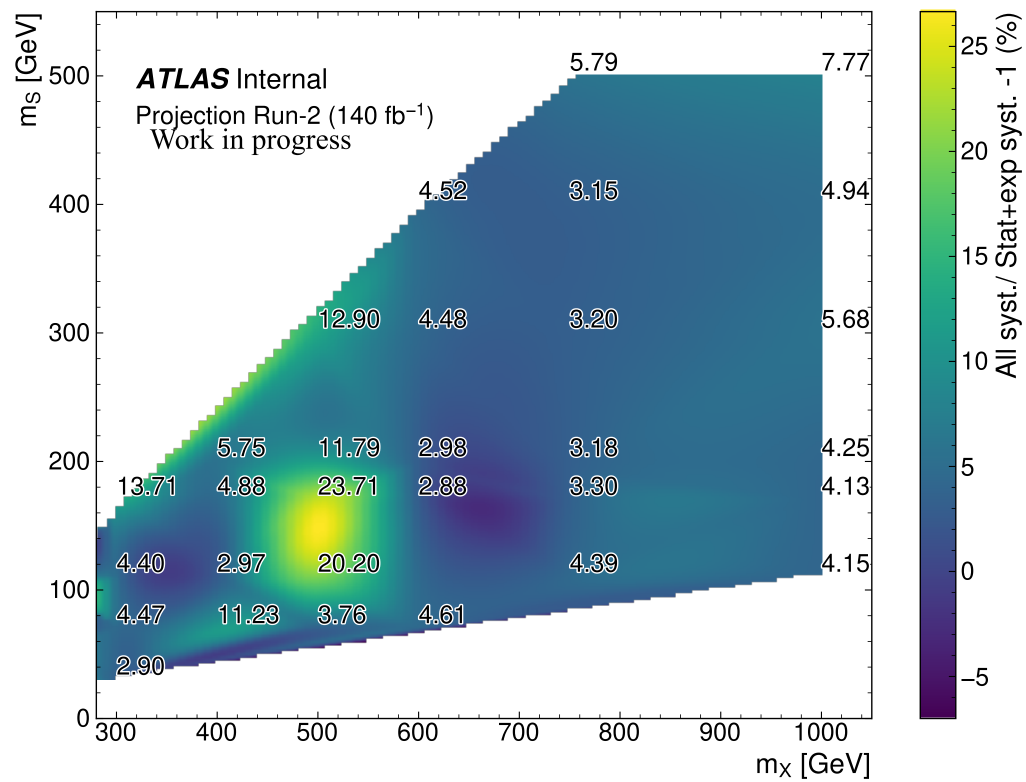


- Their impact on limit can reach 18% but are mostly between 5-10% at low mass and below 1% at high mass

Impact of theoretical systematics uncertainties

- Same plot as before but with theoretical uncertainties

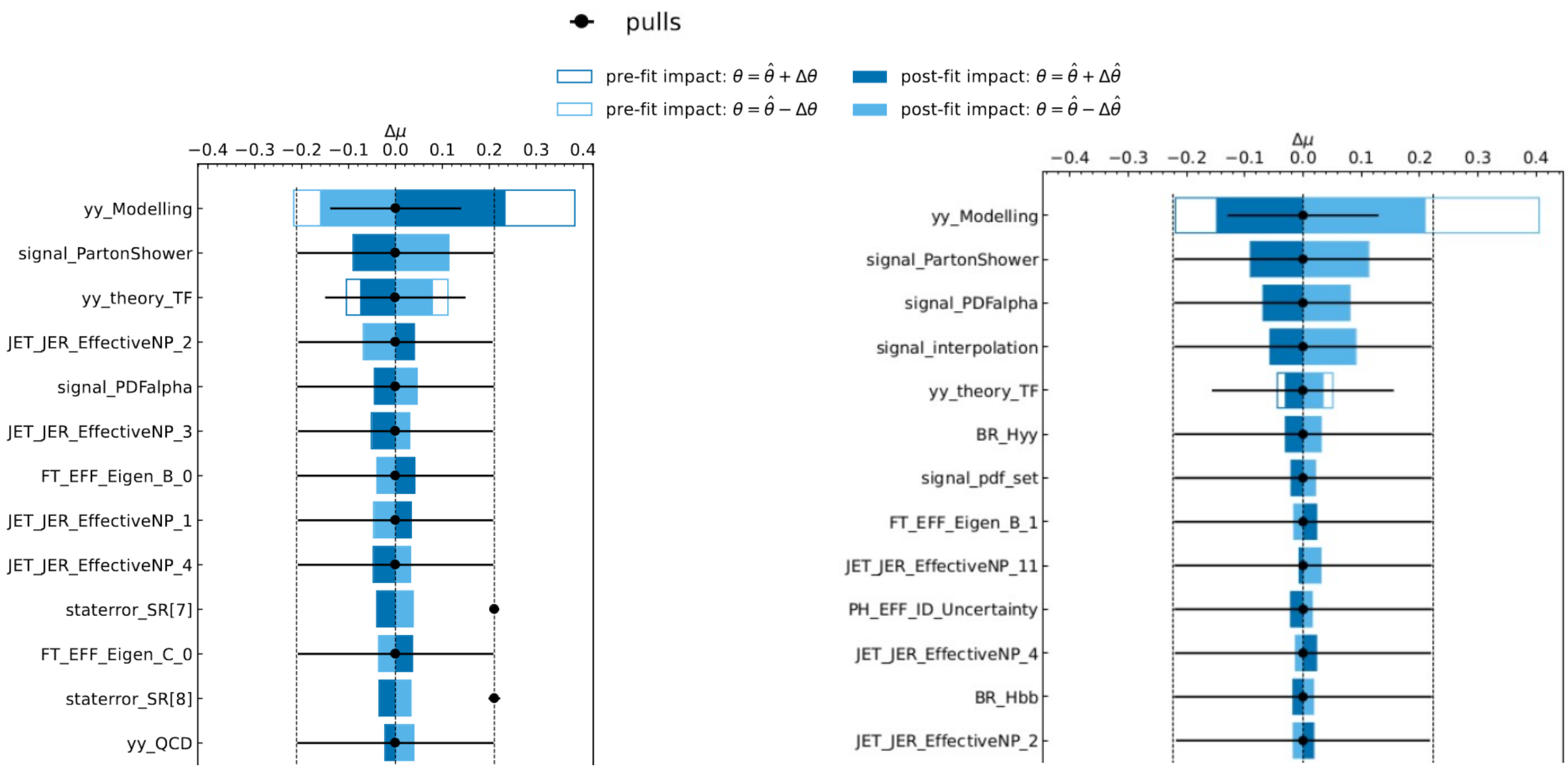




- Theory systematics impact vary a lot and can reach 20%
- It seem to be dominated by γγ modelling (in back-up slide)
  Other theoretical systematics account for 3-4%

Impact of systematics uncertainty – ranking plots

- Ranking plots : class the uncertainties with the impact they have on the fit POI



mX = 250 GeV, mS = 100 GeV

mX = 575 GeV, mS = 200 GeV

- Largest systematics is the modelling of the $\gamma\gamma$ + jets background
- Largest experimental systematics are about flavour tagging and jet energy resolution

Summary

- A search for a resonant scalar particle X decaying into a scalar S and SM Higgs is performed on the $X \to SH \to bb\gamma\gamma$ channel with the ATLAS Run-2 data
- Most interesting point in this analysis is the PNN that has been developped to target the signal for any values of mX and mS
- Analysis at the group internal review stage
  More analyses to come with Run-3 data !

Back-up

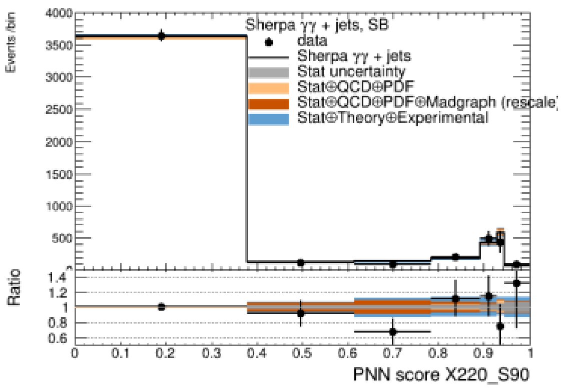
All systematics uncertainties

Main uncertainties

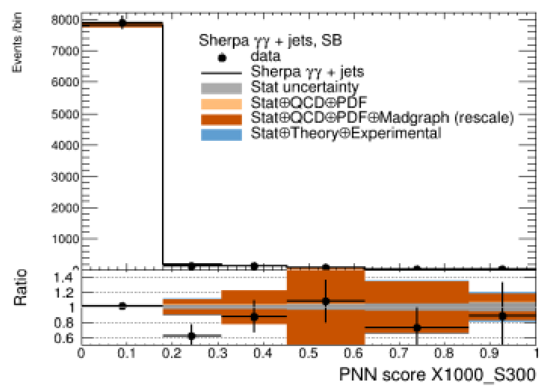
		Signal	HH ggF	HH VBF	ttH & ZH	Other Single Higgs	Continuum $\gamma\gamma$+jets
Theory	Normalisation	$BR(H \to \gamma\gamma)$	$BR(H \to \gamma\gamma)$ $BR(H \to b\bar{b})$ PDF+α_S Scales + m_t	$BR(H \to \gamma\gamma)$ $BR(H \to b\bar{b})$ PDF+α_S Scales	$BR(H \to \gamma\gamma)$	$BR(H \to \gamma\gamma)$ PDF+α_S Scales	$\gamma\gamma$ transfer factor
	Shape+Norm.	Scales, PDF+α_S Parton shower Interpolation	Parton Shower		Scales, PDF+α_S Parton Shower		Scales, PDF+α_S Modelling
Exp.	Shape+Norm.	Pile-up modelling Diphoton trigger efficiency Photon identification and isolation efficiency Photon energy scale and resolution Jet energy scale and resolution Jet vertex tagger efficiency Flavour tagging efficiency	(all exp. systematics are neglected for bbH, tH and $VBF\ H$)				

Theoretical systematics uncertainties

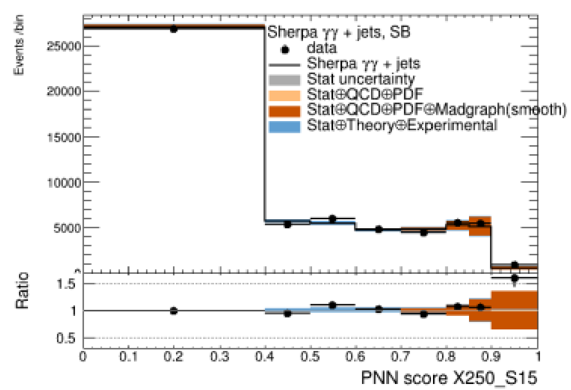
- Largest theoretical uncertainty is the modelling of $\gamma\gamma$ + jets events which is difficult to handle
- Normalization of the $\gamma\gamma$ + jets events is determined by a normalization factor from the sideband distribution
- The uncertainty regarding the modelling is evaluated by comparing simulated events from two different MC generators : Sherpa and MadGraph



(a) m_X = 220 GeV, m_S = 90 GeV



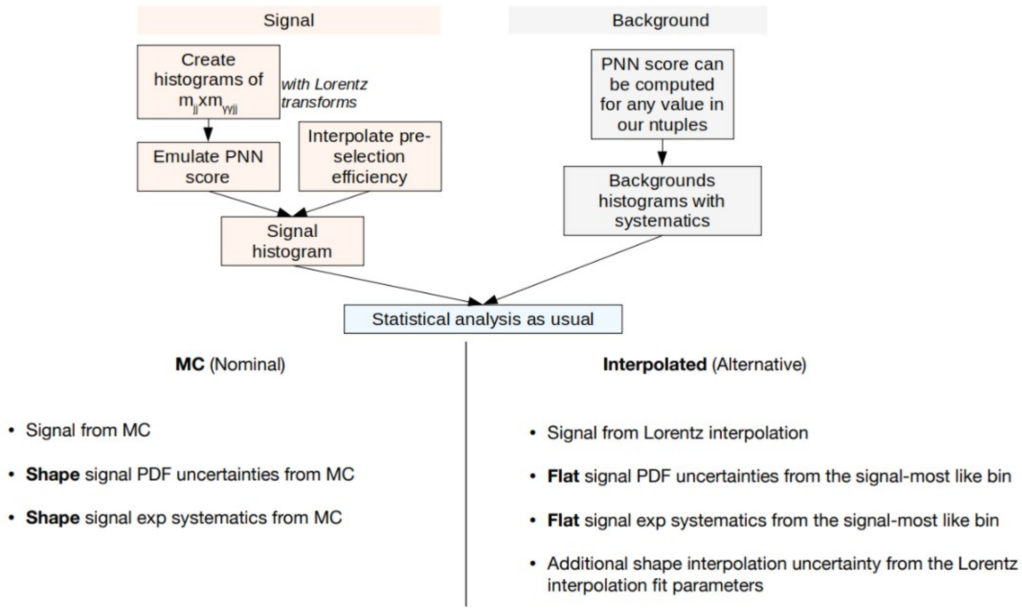
(b) m_X = 1000 GeV, m_S = 300 GeV



(c) m_X = 250 GeV, m_S = 15 GeV

Interpolation strategy

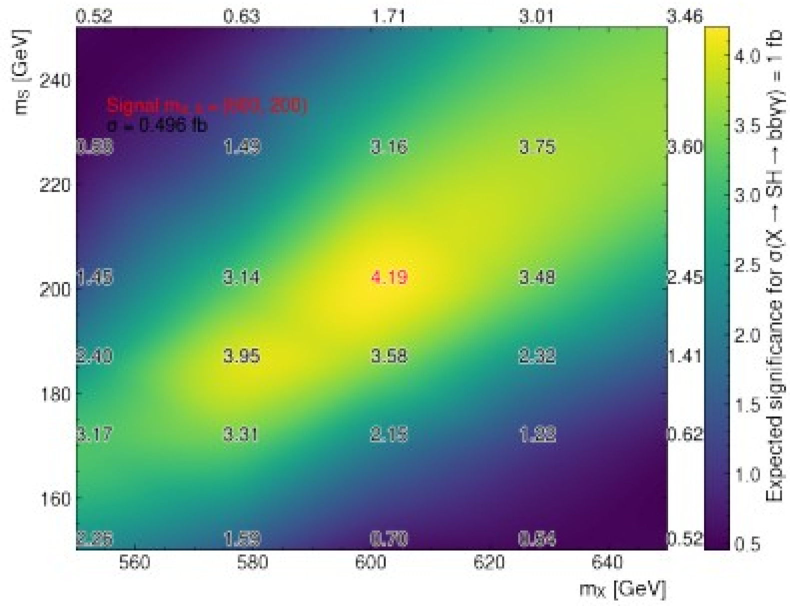
- Why interpolate ?
  → We need to be able to look for any signal in the region and granularity is not precise enough with MC samples
- Interpolation works separately for signal and background
  - PNN score can be computed for any mX, mS values in background samples
  - For signal we need to interpolate both the yields and the PNN shape
    - The shape is obtained with Lorentz transforms
    - The yields are obtained using Delaunay triangulation from the available MC samples

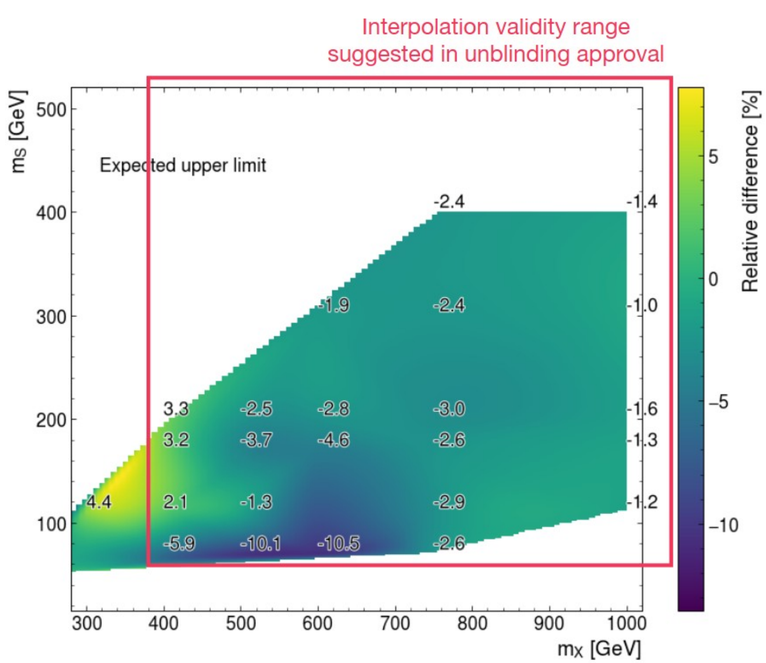


Interpolation range and validity

- Where should we interpolate ?
- Injection tests are made to ensure granularity is enough to allow us to be sensible to any signal in the probed region

  → we inject signal at σ = 2*expected limit and want at least one neighbouring point to have an expected significance ≥ 3





- Validity of interpolation is evaluated by comparing interpolated and MC signal limits
- Difference is below 5% for most points with a maximum of ~10%
- Interpolation is more difficult in low mS regions where there are some jets overlap

  → Interpolation is made for points with mX ≥ 300 GeV and mS ≥ 70 GeV