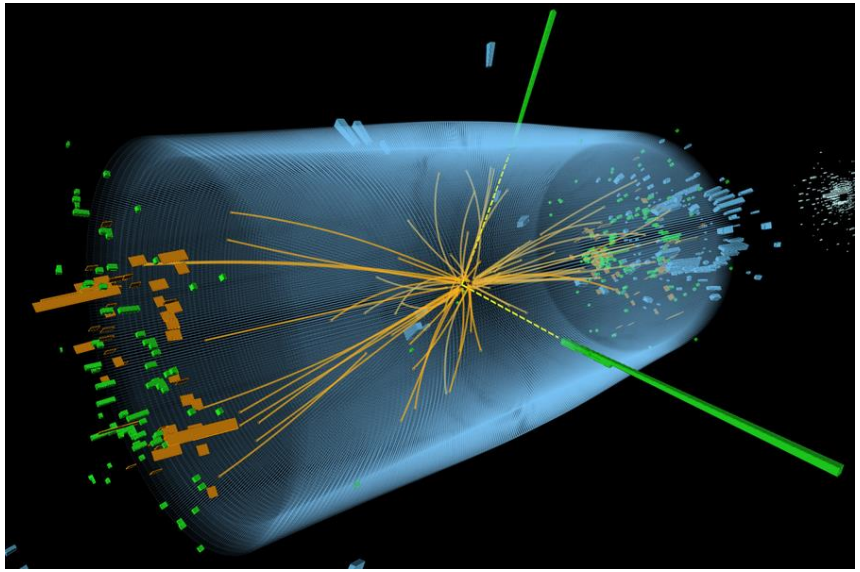
Searches of Higgs Boson decays into Dark Matter particles in the ATLAS and CMS experiments

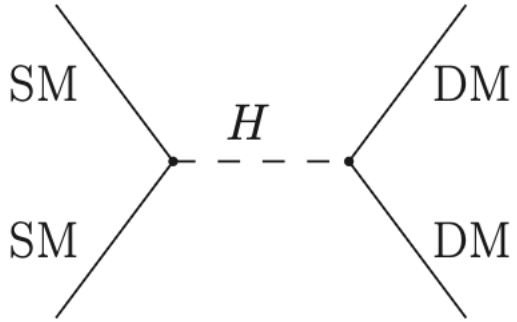


Marina Pokidova, Yuri Naryshkin
NRC KI – PNPI

Introduction

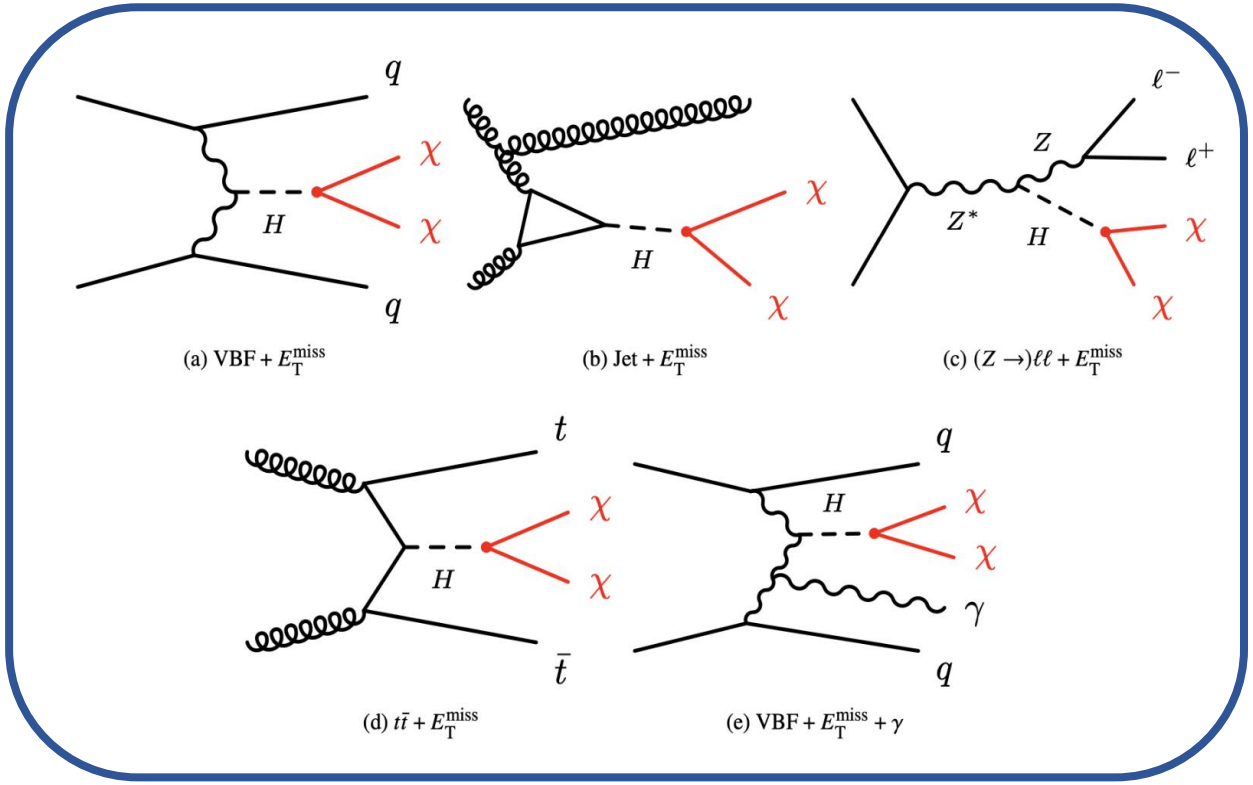
- The search for new physics is one of the current tasks at the LHC, for example, the search for Dark Matter (DM) particles
- The Higgs boson was discovered by ATLAS and CMS in 2012.
- In some of BSM models the 125 GeV Higgs boson can act as a mediator between DM candidates (WIMP) and the SM particles. These models are called ”Higgs portal” models
- In SM, Higgs boson can decay to ”invisible” 4-neutrino final state with BR(H → ZZ* → 4ν) ~ 0.1%
- Last H→inv decay searches that were performed at ATLAS and CMS experiments using the full ATLAS Run-2 data sample are based on ”Higgs portal” models





$H \to inv$ search in ATLAS experiment (Run2 summary)

- In searches for invisible Higgs decay in Run2 multiple production modes of the Standard Model Higgs boson were considered



- VBF+MET
- MET+Z(ll)
- tt+MET
- VBF+MET+γ
- monojet

Phys. Lett. B 842 (2023) 137963

Results of *H→inv* search in ATLAS experiment for Run2

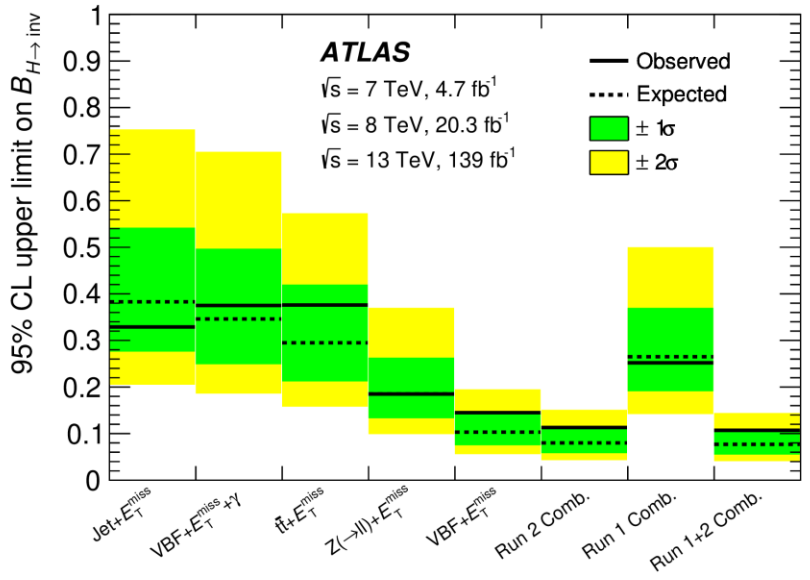
- Limits on the H→inv branching ratio are shown for all the channels, Run1, Run2 and full combinations.

BR(H→inv) < 10.7 (7.7) % @ 95% CL

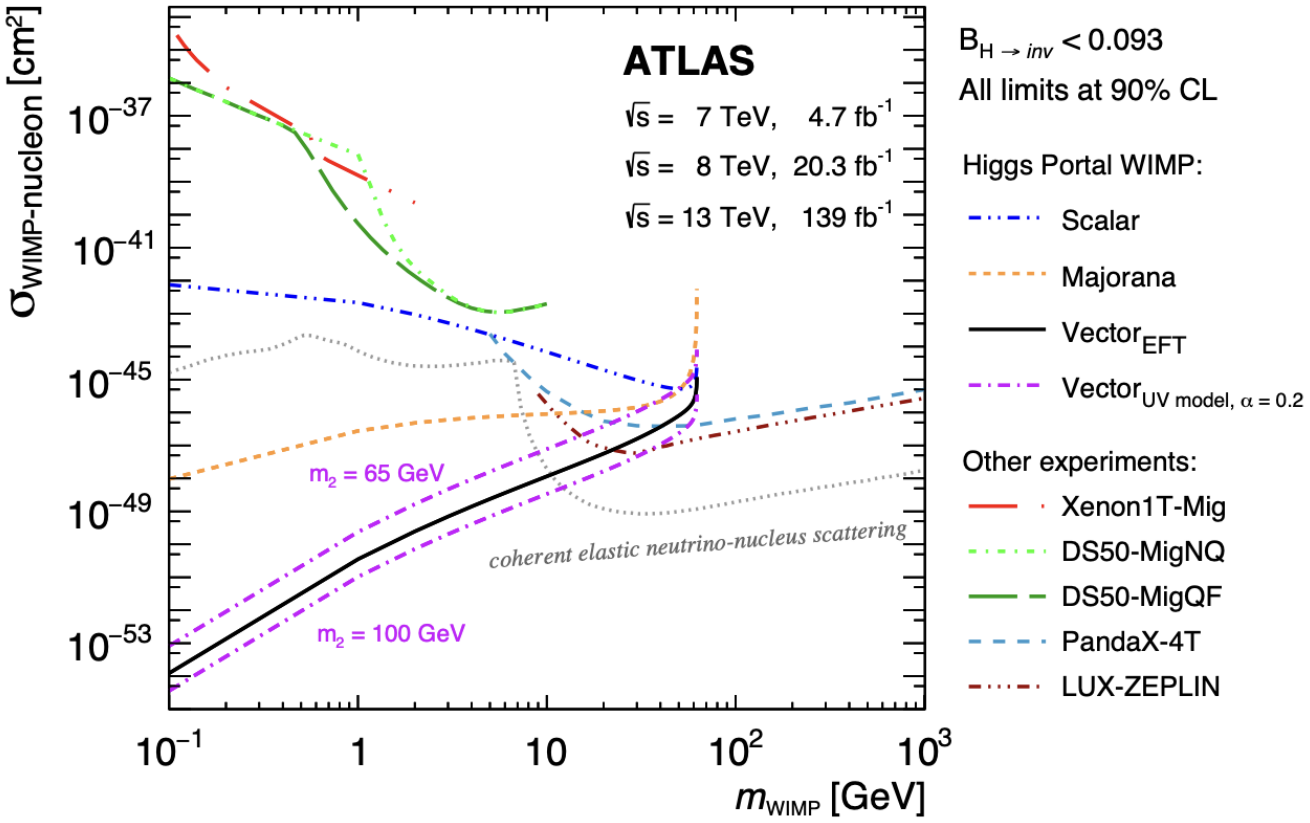
- Full Run2+Run1 combination provides the most stringent limit on the BR(H→inv) to date

[Phys. Lett. B 842 (2023) 137963]

Results	Expected	$+1\sigma$	-1σ	$+2\sigma$	-2σ	Observed
Jet + E_T^{miss}	0.381	0.538	0.275	0.749	0.205	0.329
VBF + E_T^{miss} + γ	0.346	0.497	0.249	0.704	0.186	0.375
$t\bar{t}$	0.295	0.420	0.212	0.573	0.158	0.376
$(Z \rightarrow)\ell\ell$ + E_T^{miss}	0.185	0.263	0.133	0.370	0.099	0.185
VBF + E_T^{miss}	0.104	0.144	0.075	0.195	0.056	0.145
Run–1	0.265	0.370	0.191	0.501	0.142	0.252
comb. Run2	0.080	0.111	0.058	0.151	0.043	0.113
combined	0.077	0.107	0.055	0.144	0.041	0.107



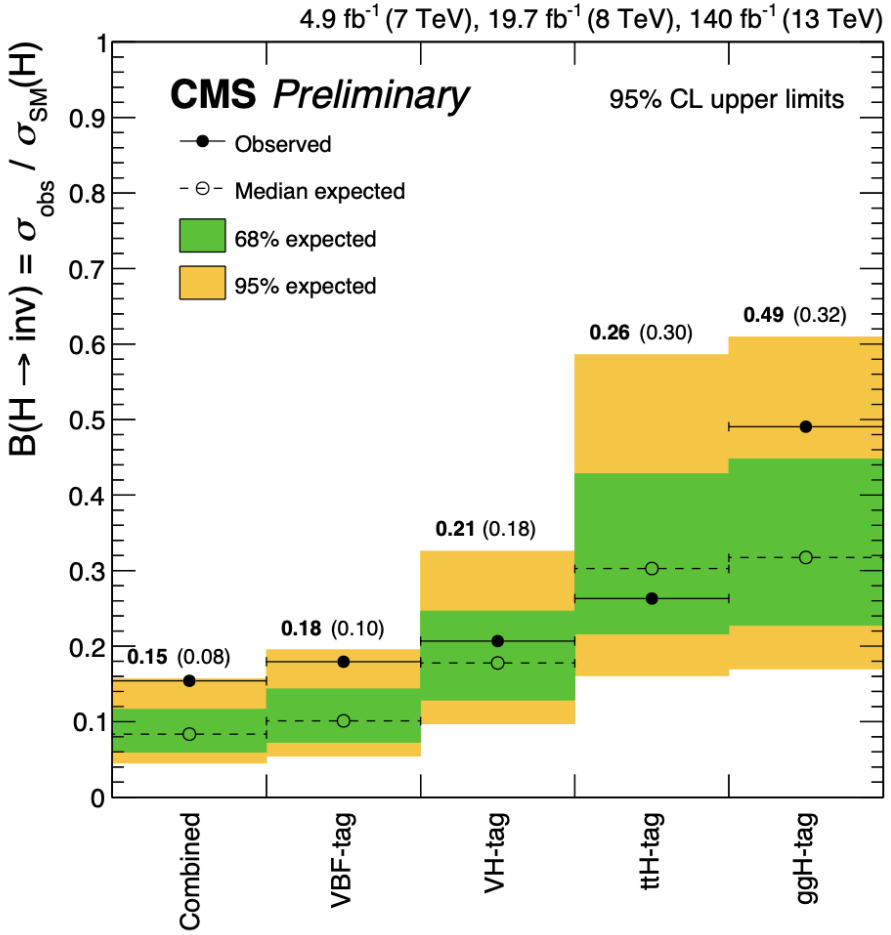
Comparison to direct dark matter detection experiments

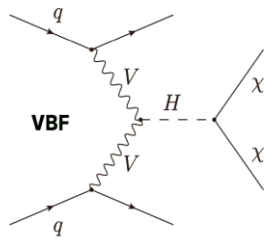


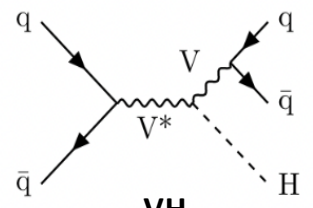
- Upper limits on the spin-independent WIMP–nucleon cross section using Higgs portal interpretations of BR(H→inv) at 90% CL vs m_{WIMP}.
- Complementarity of direct and collider experiments. Collider experiments have higher sensitivity for smaller m_{WIMP} (< 10 GeV).

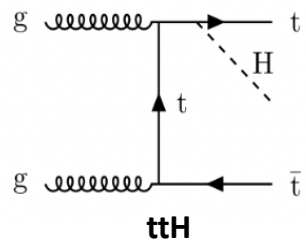
Phys. Lett. B 842 (2023) 137963

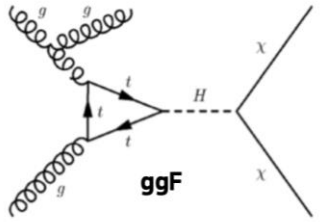
Results of $H \to inv$ search in CMS experiment for Run2











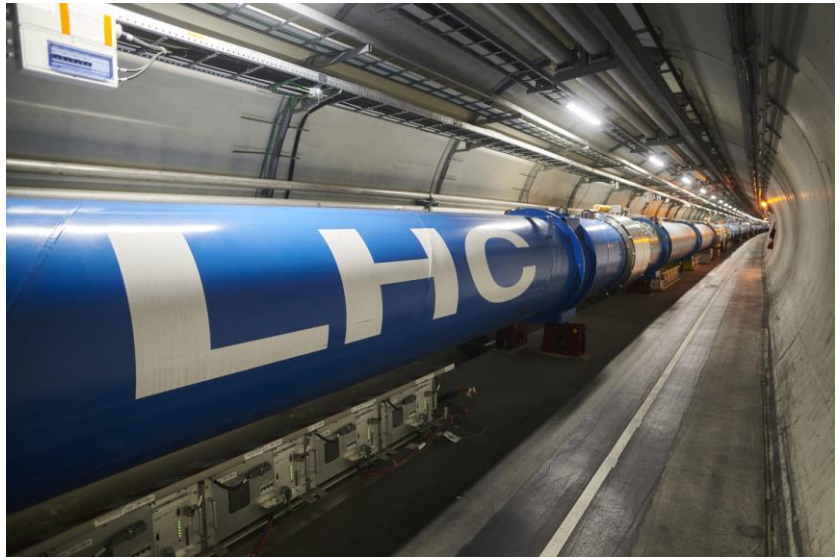
- The results are shown separately for each Higgs boson production mode as tagged by the input analyses for Run 1 and Run 2

BR(H→inv) < 15.0 (8.0) % @ 95% CL

! ATLAS result is more strict

[CMS-PAS-HIG-21-007]

$H \rightarrow inv$ search in ATLAS and CMS experiment for Run2 + Run3





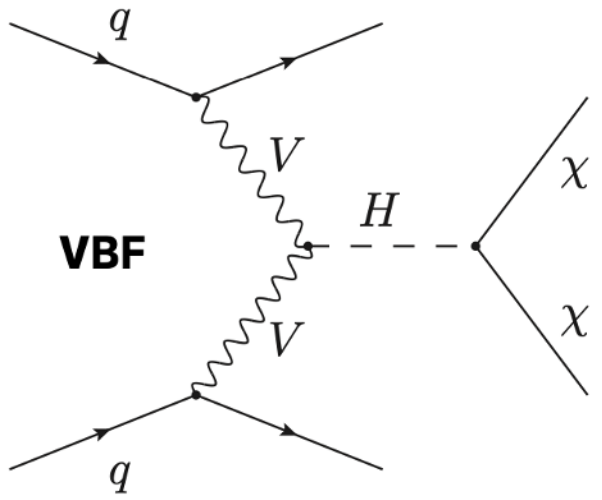
- In Run 3, the center-of-mass energy reached 13.6 TeV, leading to increased statistics.
- Studies of statistically limited processes can achieve greater sensitivity - $H \rightarrow inv$ search will continue with the use of Run2-Run3 statistics

$H \rightarrow inv$ in Vector boson fusion channel (VBF+MET)

- VBF+MET is the 2nd largest Higgs production rate
- In both ATLAS and CMS experiments the VBF+MET final state is the most sensitive channel resulting in an upper limit of

ATLAS Run2: BR(H→inv) < 0.145 (0.103) @ 95% CL

CMS Run2: BR(H→inv) < 0.17 (0.11) @ 95% CL



JHEP 08 (2022) 104 , CMS-HIG-20-003

$H \rightarrow inv$ in Vector boson fusion channel (VBF+MET)

In the VBF production mode, the $H \rightarrow$ invisible signal is characterized by two jets with a large separation in pseudorapidity and E_T^{miss} arising from the invisible decays of the Higgs boson

VBF topology

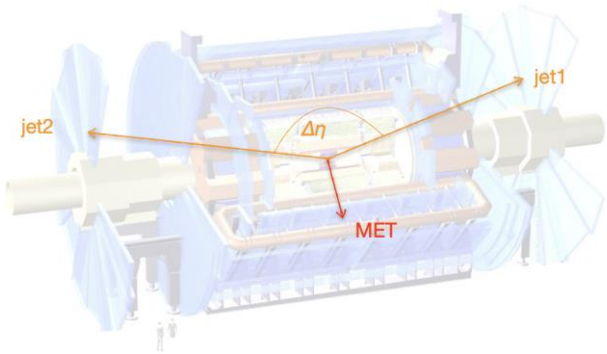
- Two jets with p_T(j1 / j2) > 80 / 50 GeV
- Small add. jet activity: p_T(j3) < 25 GeV …
  3 or 4 jets, if compatible with FSR
- Jets in opposite hemispheres
- $\Delta\eta$(jj) > 3.8
- m(jj) > 0.8 TeV

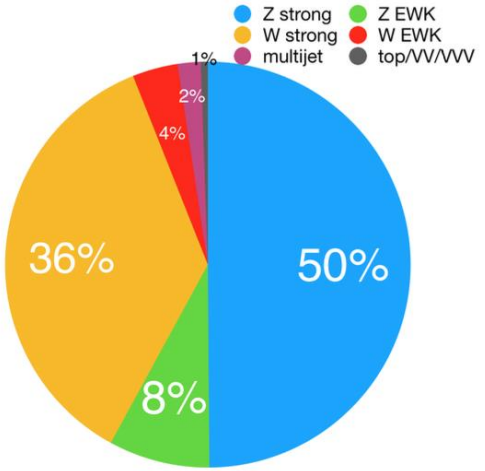
EWK suppression

- Lepton (electron and muon) veto

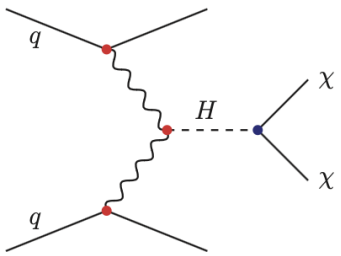
Multijet suppression

- MET > 160 GeV
- $\Delta\Phi$(jj) < 2.0

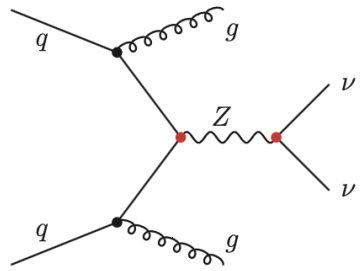




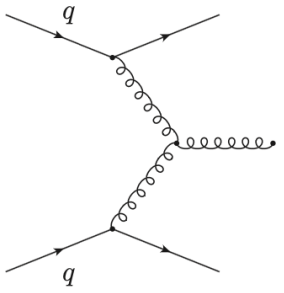
Signal



***V*+jets**



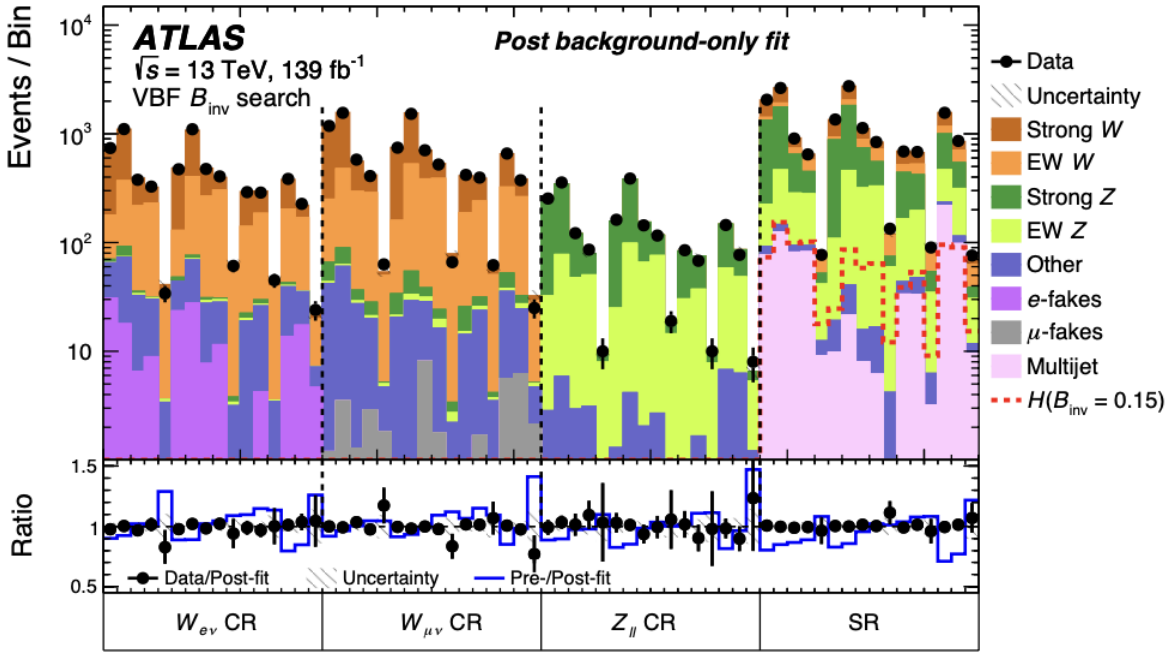
Multijet



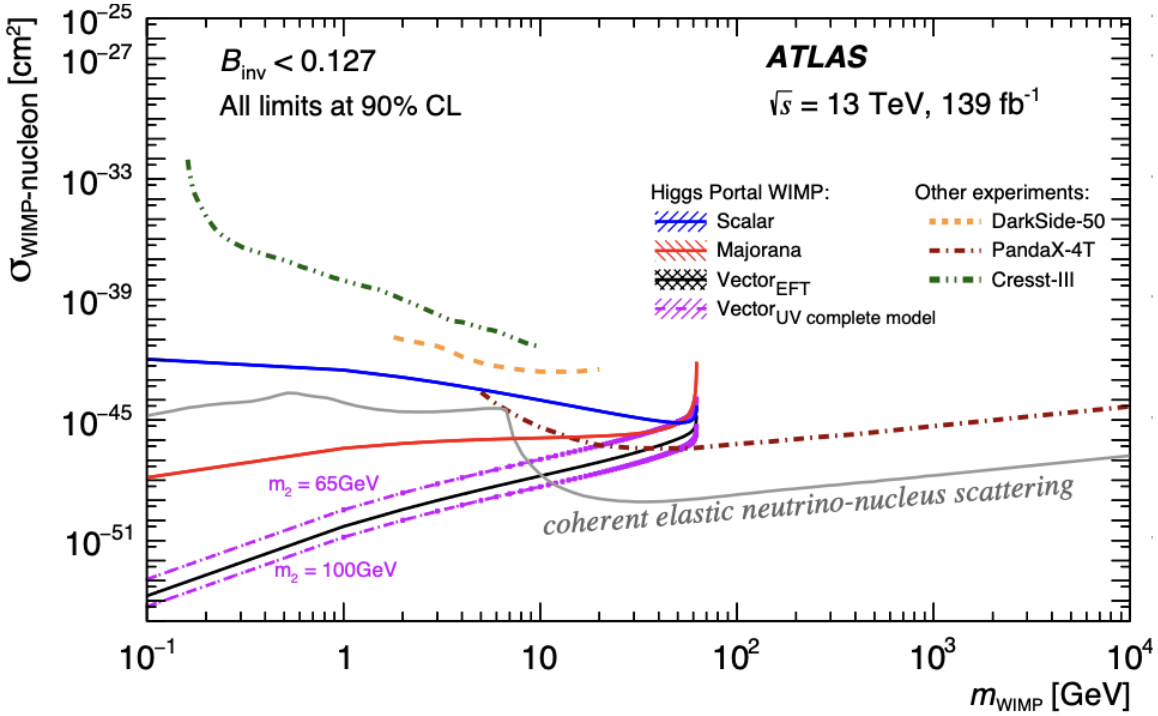
JHEP **08** (2022) 104

Results of VBF+MET $H \rightarrow inv$ search in ATLAS experiment for Run2

- The observed numbers of events are found to be in agreement with the background expectation from Standard Model processes
- BR(H→inv) < 0.145 (0.103) @ 95% CL



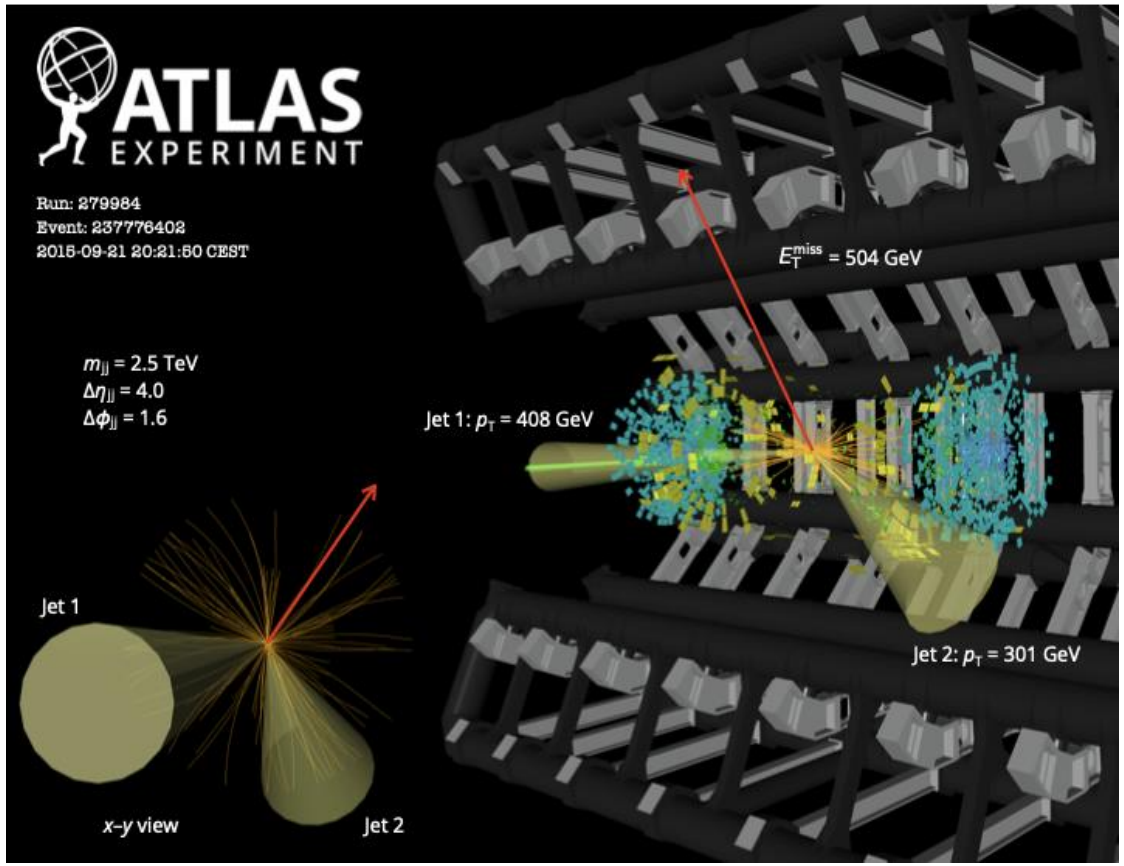
- Post-fit yields and their uncertainties for all CRs and SR



- Upper limits on the spin-independent WIMP–nucleon cross section using Higgs portal interpretations of B_{inv} at 90% CL vs m_{WIMP}.

JHEP **08** (2022) 104

VBF+MET $H \rightarrow inv$ search in ATLAS experiment for Run2 + Run3



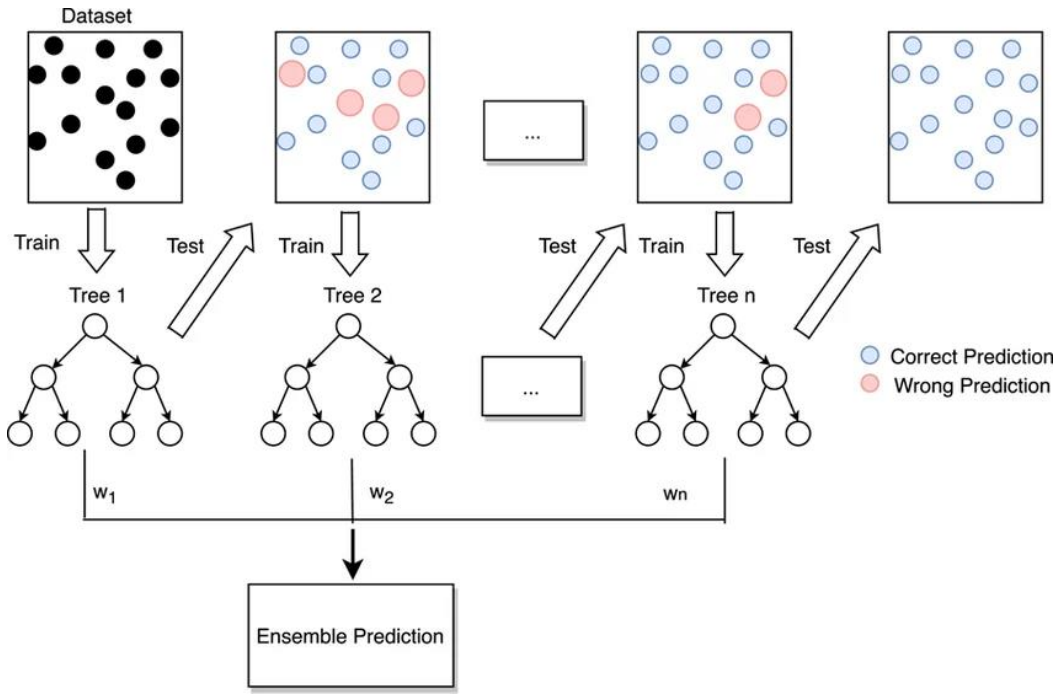
Data analysis using Run2 + Run3

- Optimization of background process selection and background process contribution estimation using MVA/ML approaches (BDTG)
- Comparison of experimental data with SM predictions
- Setting strictest upper limit on BR($H \rightarrow$ inv) in VBF+MET channel (pyhf)

JHEP **08** (2022) 104

Summary of new MVA/ML studies

Classifier based on TMVA ML framework for optimizing signal and backgrounds selection criteria



Signal (MC): *VBF H → inv*
Background processes (MC):

- *W strong*
- *Z strong*
- *W EWK*
- *Z EWK*
- *Multijet*
- *Other (singletop, ttbar, etc)*

- A Boosted decision tree with Gradient Boosting (BDTG) could be used for a straight classification problem and the goal was to choose effective cut on BDT score to separate the signal and backgrounds.

The algorithm learns to distinguish between the distributions of signal and background events and assigns a value to each event: $-1 \lesssim$ BDT score $\lesssim 1$, where 1 corresponds to a higher degree of "signal-likeness," and -1 indicates an event resembling background.

Summary of new MVA/ML studies

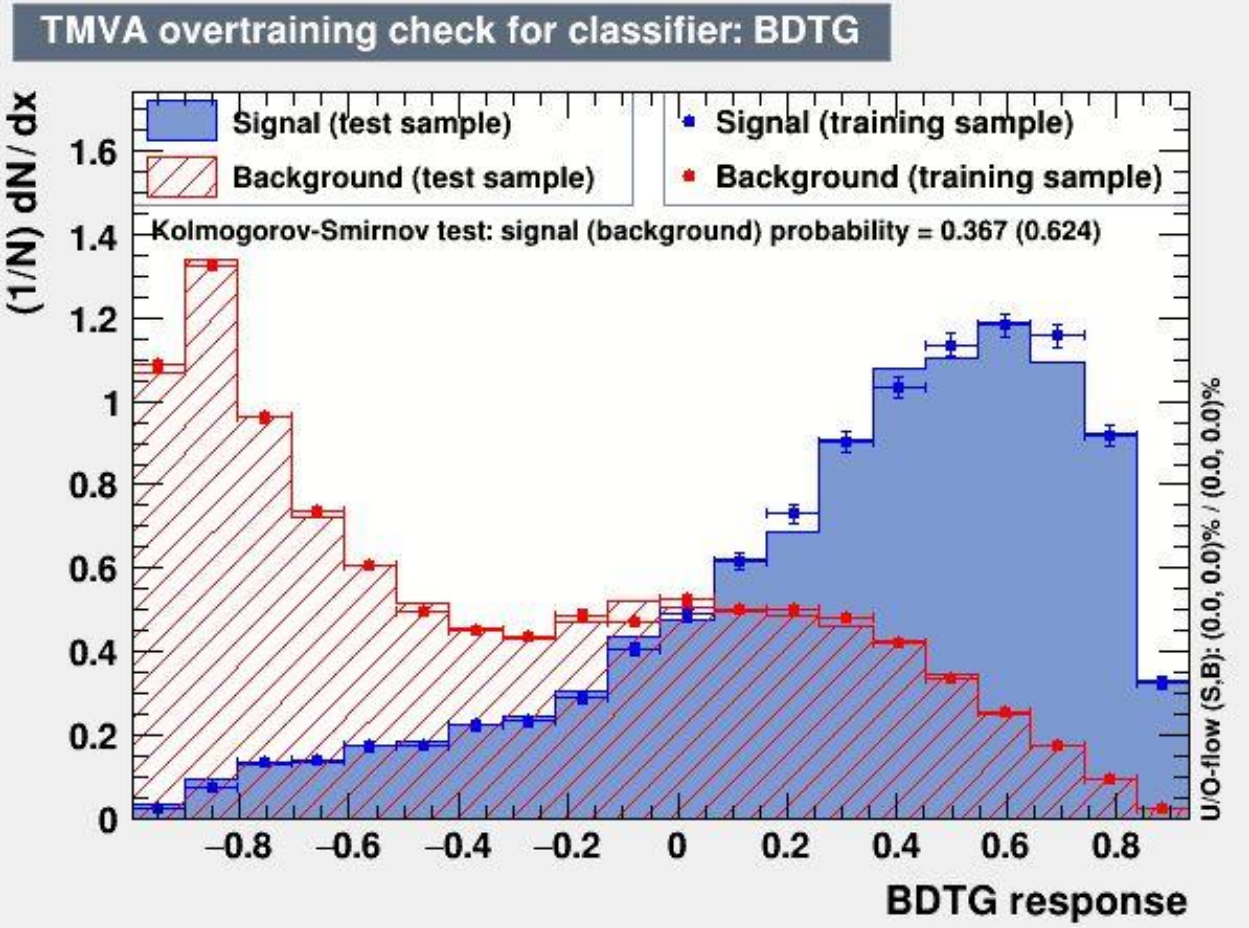
In this studies variables related to SR selections (leading and subleading $jet\ p_T, m_{jj}, \Delta\eta_{jj}, \Delta\varphi_{jj}$) and additional jets activity in final state were considered for better signal and background separation.

$$f_i^{\eta} = \frac{(\eta_1 - \eta_i) - (\eta_i - \eta_2)}{\eta_1 - \eta_2} \qquad m_i^{\mathrm{rel}} = \frac{\min\{m_{\mathrm{j1},i}, m_{\mathrm{j2},i}\}}{m_{\mathrm{jj}}}.$$

- BDT score is then used in a dedicated limit setting study to compare with the cut-and-count method as the benchmark. Using the BDTG classifier yields as an distribution can lead to higher sensitivity in terms of separating signal and background events (the S/B ratio for each histogram bin)

in progress

Summary of new MVA/ML studies – Results for even/odd case




```
Number of training and testing events
-------------------------------------------------------
Signal     -- training events            : 46859
Signal     -- testing events             : 46405
Signal     -- training and testing events: 93264
Background -- training events            : 8778348
Background -- testing events             : 8787156
Background -- training and testing events: 17565504
```

```
-------------------------------------------------------
Rank : Variable                  : Variable Importance
-------------------------------------------------------
   1 : feta12                    : 1.319e-01
   2 : m2_rel                    : 1.279e-01
   3 : met_tst_nolep_et          : 1.220e-01
   4 : feta13                    : 1.089e-01
   5 : met_tst_nolep_j1_dphi_new : 1.085e-01
   6 : met_tst_nolep_j2_dphi_new : 1.058e-01
   7 : jj_deta                   : 1.009e-01
   8 : m3_rel                    : 9.940e-02
   9 : jet_pt[1]                 : 9.476e-02
-------------------------------------------------------
```

```
ROC-integ
: 0.828
```

in progress

Conclusion

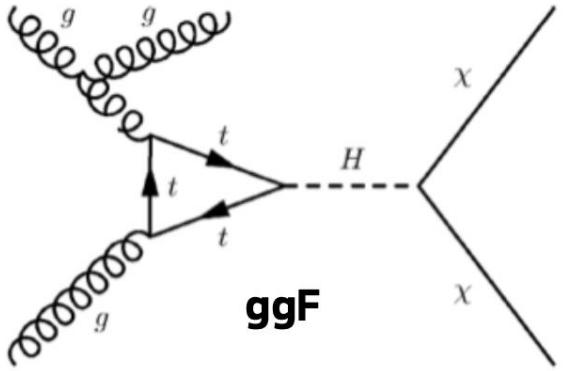
- Searches for invisible decays of the Higgs boson using LCH Run 2 pp collision data at in several Higgs boson production topologies were preformed in ATLAS and CMS experiments.
  - ATLAS statistical combination of Run2+Run1 resulted the most stringent limit on the B(H→inv) to date **BR(H→inv) < 10.7 (7.7) % @ 95% CL**

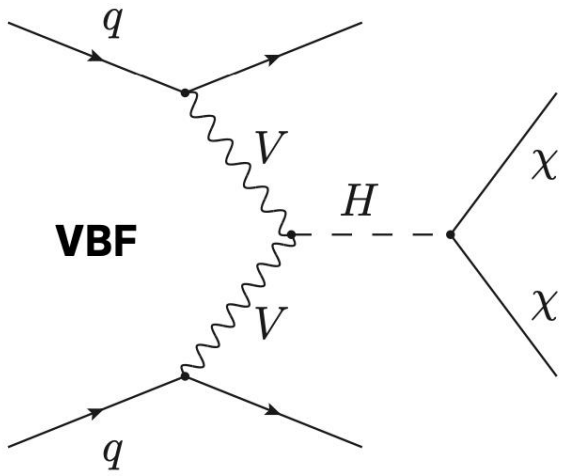
- The search for Higgs invisible decays is continued with Run2+Run3 data.
  - Run3 VBF+MET results will be included in the future article **«Search for invisibly decaying Higgs bosons with the Vector-Boson Fusion plus missing transverse momentum signature with Run 2 and Run 3 data from the ATLAS detector»**

Thank you for your attention

Backup slides

Why VBF+MET the most sensetive?



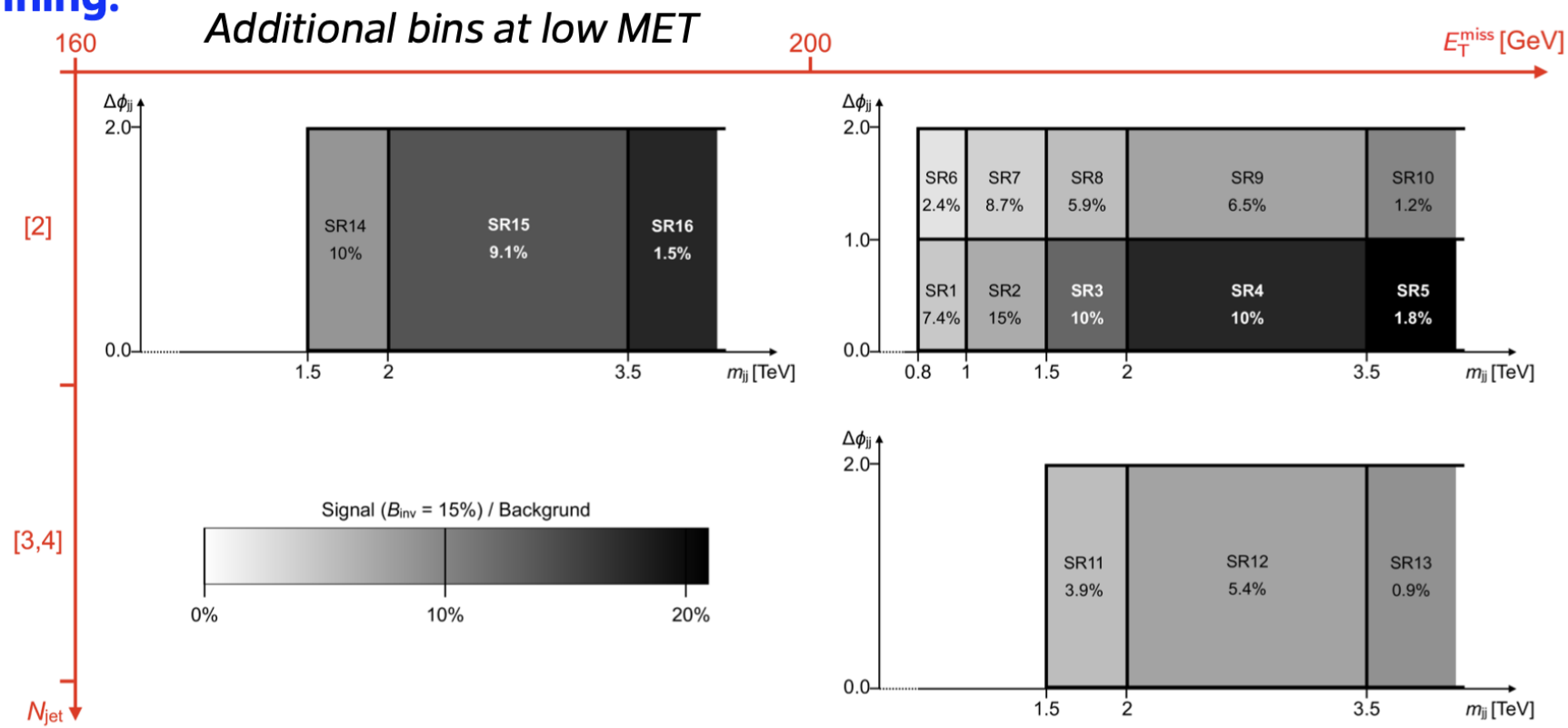


- $Z \to \nu\nu$ (strong production) is large compared to ggF: $\dfrac{\sigma_{\mathrm{ggF}}(H + \mathrm{jet})}{\sigma(Z + \mathrm{jet}) \cdot Br(Z \to \nu\nu)} \sim \dfrac{1}{300}$

- VBF: signal and bg are comparable $\dfrac{\sigma_{\mathrm{VBF}}(H + qq)}{\sigma_{\mathrm{VBF}}(Z + qq) \cdot Br(Z \to \nu\nu)} \sim \dfrac{4}{3}$

36 fb^{-1} result: only three *m*(jj) bins (1–1.5, 1.5–2, >2 TeV); MET > 180 GeV (no bins with 3 or 4 jet)

New binning:



More bins in m(jj)

Splitting of ΔΦ(jj) in two ranges

Allow for 3 or 4 jets if compatible with FSR

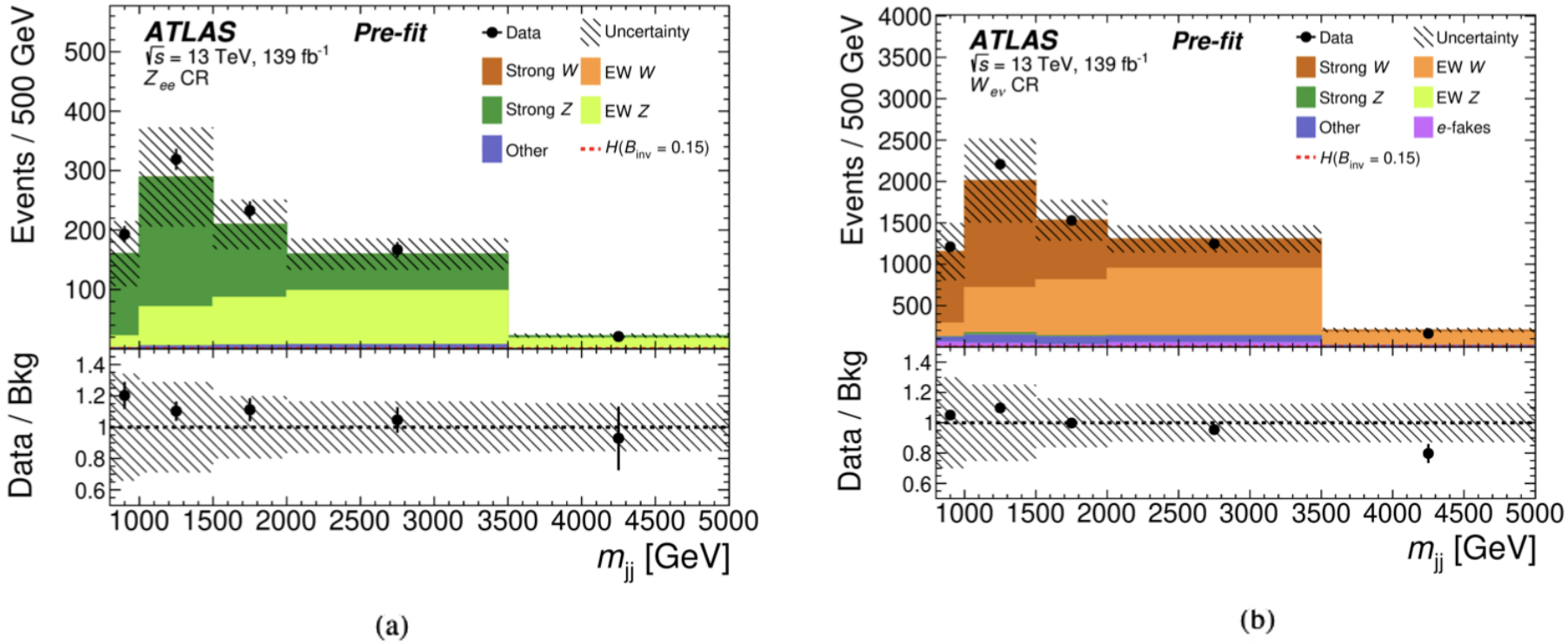
Shading indicates S/B, assuming BR of Higgs boson to invisible particles of 0.15

Percentage gives approximate contribution of signal from invisibly decaying Higgs bosons relative to total signal yield in inclusive SR

JHEP **08** (2022) 104

V+jets proceses estimation

- Constrain *V*+jets backgrounds with single- or di-lepton CRs
- One normalisation factor for each of 16 bins to account for mismodelling of current generators



Run2 distributions m_{jj} for (a) W (→ lν) and (б) Z (→ ll) CR

JHEP **08** (2022) 104

Multijet bkg process estimation

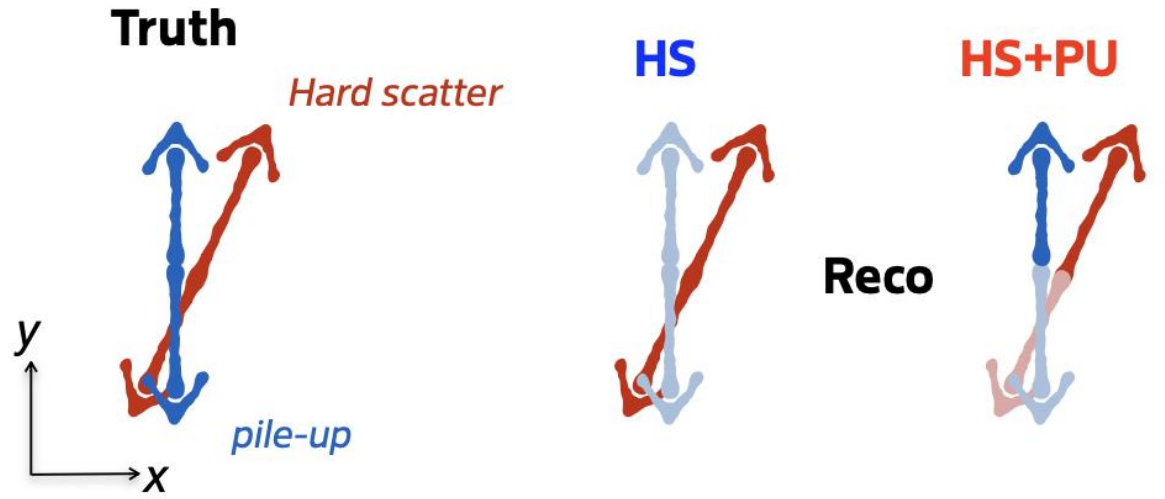
- **General problem:** MC has large uncertainties and in-sufficient statistics

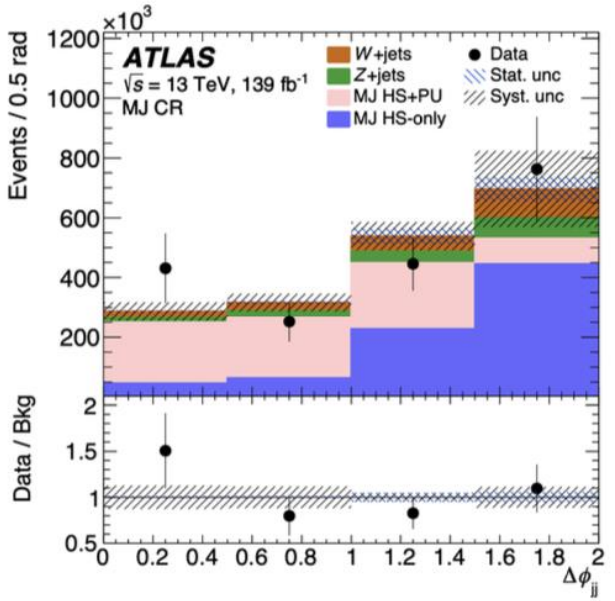
  → Combination of two independent data-driven bg estimation techniques

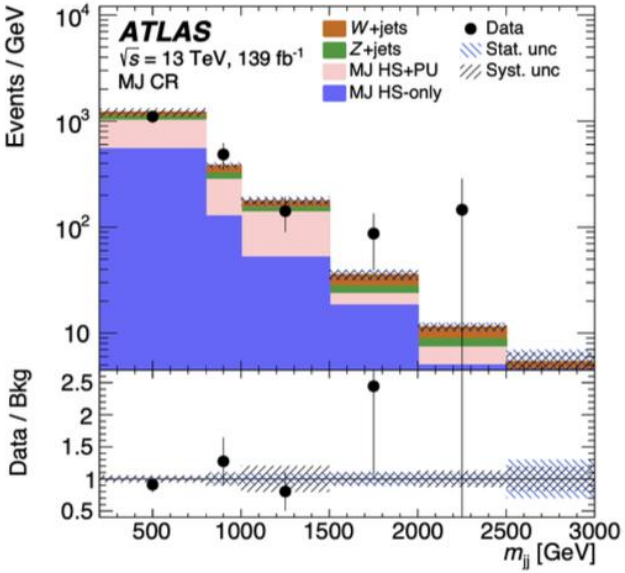
- Multiple application of detector response for (simulated truth or "rebalanced") seed events (pile-up and hard scatter component separately) and subsequent emulation of pile-up tagging

  → Large statistics pseudo-data sample

  → Normalisation of MJ components (HS+PU or HS) at lower MET





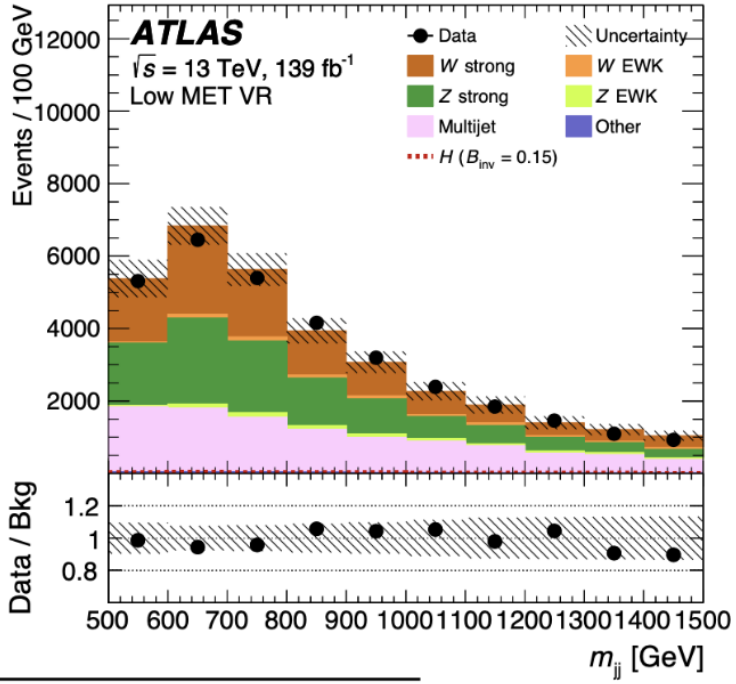


JHEP **08** (2022) 104

Multijet bkg process estimation

MJ - Pileup CR method

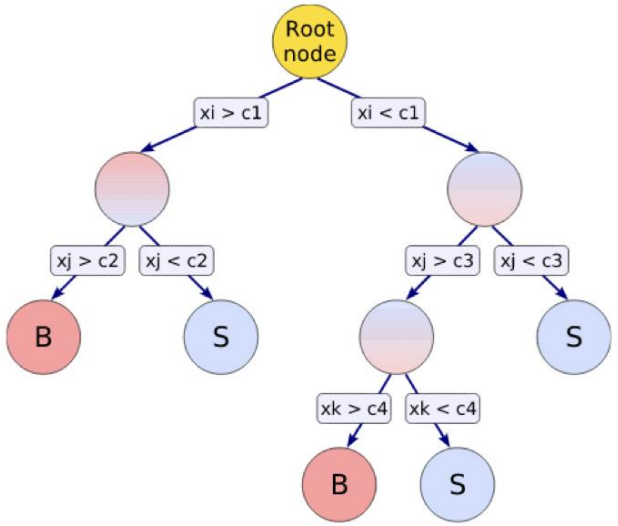
- Selecting events in pile-up enriched CR
- Normalise yields in CR by transfer factor obtained at low MET
- Consistent results for $N_{\text{jet}} = 2$ signal region from both methods
- Depending on selection use of either one (or both) methods ("Rebalance and Smear" or PU CR) to reduce uncertainties

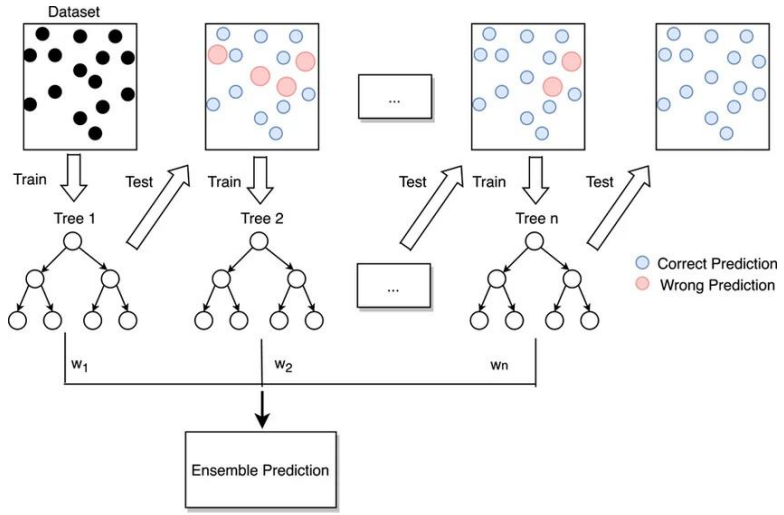


	Rebalance and Smear		Pile-up-CR method		
Multijet pred. (SR incl.)	840 ± 367		761 ± 186		
	Uncertainties [%]				
Source	inclusive	per bin	Source	inclusive	per bin
Statistical	9	6–46	Stat.	8.6	16–49 (bins 1–5 only)
Non-closure	40	37–43	Syst.	20	20–27
Jet energy response (core)	4.8	1–41			
Jet energy response (tail)	5.7	1–53			
Data vs simulation	4.9	2–30			

JHEP **08** (2022) 104

Boosted Decision tree





- Sequential application of cuts splits the data into nodes, where the final nodes (leafs) classify an event as signal or background

  - Start with Root node
  - Split training sample according to cut on best variable at this node
  - Splitting criterion: e.g., maximum "Gini-index": purity × (1– purity)
  - Continue splitting until min. number of events or max. purity reached
  - Classify leaf node according to majority of events, or give weight; unknown test events are classified accordingly

Boosted decision trees: combine *forest* of decision trees, with differently
weighted events in each tree (trees can also be weighted), by majority vote

Adaptive Boosting aka Adaboost	Gradient Boosting
• Both Adaboost (the first boosting algorithm) and Gradient Boosting are boosting algorithms, which combines predictions from multiple weak learners, usually decision stumps to form a strong learner	
• In AdaBoost, the weights of the samples are adjusted at each iteration.	• No reweighting of the samples take place in GBM
• **Algorithm:** Training process starts with a decision stump (usually). At every step, the weights of the training samples which are misclassified are increased for the next iteration. The next tree is built sequentially on the same training data but using the newly weighted training samples. This process is repeated until a desired performance is achieved.	• **Algorithm:** GBM uses gradient descent to iteratively fit new weak learners to the residuals of the previous ones, minimizing a loss function. There are several loss functions to choose from, Mean Squared Error being most common for Regression and Cross Entropy for Classification. GBM uses Decision Trees as the weak learners.
• Both Adaboost and GBM are stage-wise additive models (greedy algorithm), meaning new trees in the model are built without changing the previous existing trees © AIML.com Research	
• The final model is formed by combining the predictions from individual trees through a weighted sum. For a classification problem, prediction is given by: $Prediction = sign(\sum_{m=1}^{M} \alpha_m * F_m(x))$ where, $F_m(x)$ is the output of each model and α_m are the weights computed by the boosting algorithm, m is the number of iterations	• The final model is an equal-weighted sum of all of the individual trees. Prediction for a regression problem in GBM is given by: $Prediction = \hat{y} + \eta * \sum_{m=2}^{M} \hat{r}_{m-1}$ where, $\hat{y}$ is the prediction from the first tree, η is the learning rate, $\hat{r}_i$ is the prediction of residuals, m is the no. of iterations
• Both Adaboost and Gradient boosting can be used for both Classification and Regression problem	

Input variables for BDT

Input variables used in Run2:

- m_{jj}, $\Delta\eta_{jj}$, $\Delta\phi_{jj}$
- met_tst, met_tst_et , met_tst_nolep_et
- jet1 p_T, jet2 p_T
- $\Delta\phi$(jet1,MET), $\Delta\phi$(jet1,MET)
- maxCentrality

Input variables which can be used in Run3:

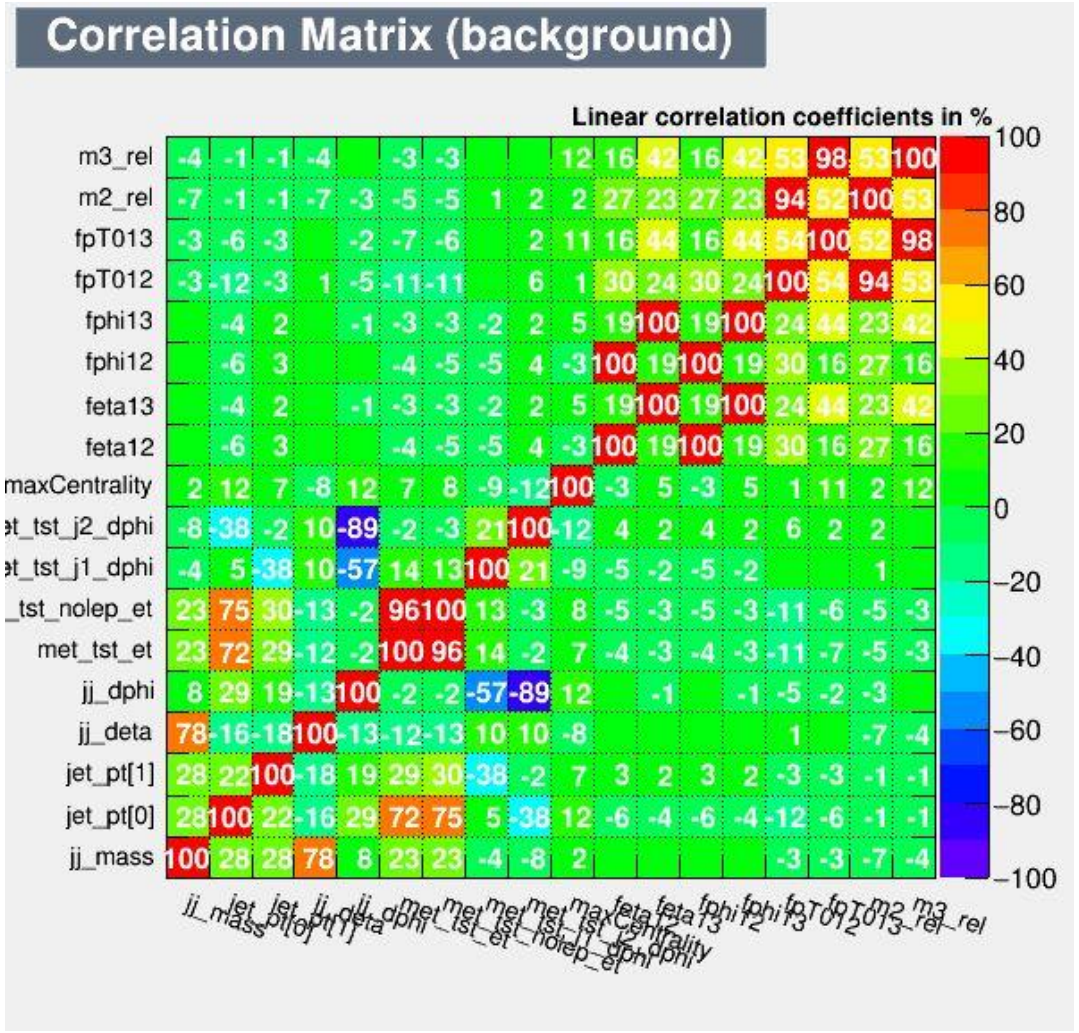
For i = 3,4,5

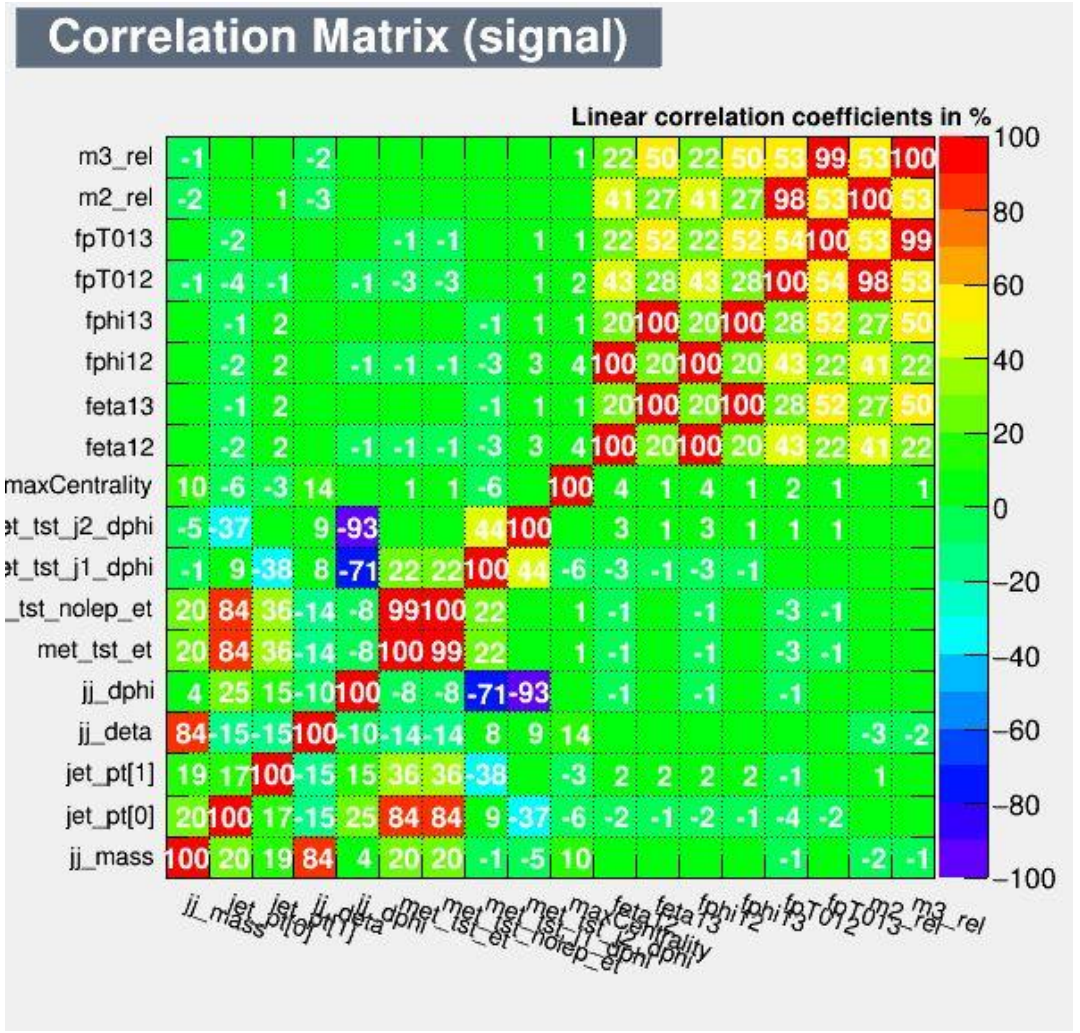
- feta1i = $f_i^{\eta} = \dfrac{(\eta_1 - \eta_i) - (\eta_i - \eta_2)}{\eta_1 - \eta_2}$
- fphi01i = $f_i^{\phi} = \dfrac{\Delta\phi_{1i} - \Delta\phi_{i2}}{\Delta\phi_{12}}$
- fp0Ti = $f_i^{p} = \dfrac{2p_{T,i}}{p_{T,1} + p_{T,2}}$
- cent01i = $C_i = \exp\left(-\dfrac{4}{(\eta^{j1} - \eta^{j2})^2}\left(\eta^i - \dfrac{\eta^{j1} + \eta^{j2}}{2}\right)^2\right)$

```
Ranking input variables (method unspecific)...
Ranking result (top variable is best ranked)
-----------------------------------------------------
Rank : Variable                    : Separation
-----------------------------------------------------
   1 : m2_rel                      : 1.799e-01
   2 : feta12                      : 1.402e-01
   3 : m3_rel                      : 1.243e-01
   4 : feta13                      : 1.165e-01
   5 : met_tst_nolep_et            : 5.779e-02
   6 : jj_deta                     : 5.241e-02
   7 : jet_pt[1]                   : 1.204e-02
   8 : met_tst_nolep_j2_dphi_new   : 7.588e-03
   9 : met_tst_nolep_j1_dphi_new   : 6.200e-03
-----------------------------------------------------
Train method: BDTG for Classification
```

- In TMVA for BDT, variable importance is based on how much the use of a given variable contributes to improving the separation in tree nodes. Then the contribution is summed over the trees and normalized. As a result, relative importance values are obtained.

Input variables for BDT

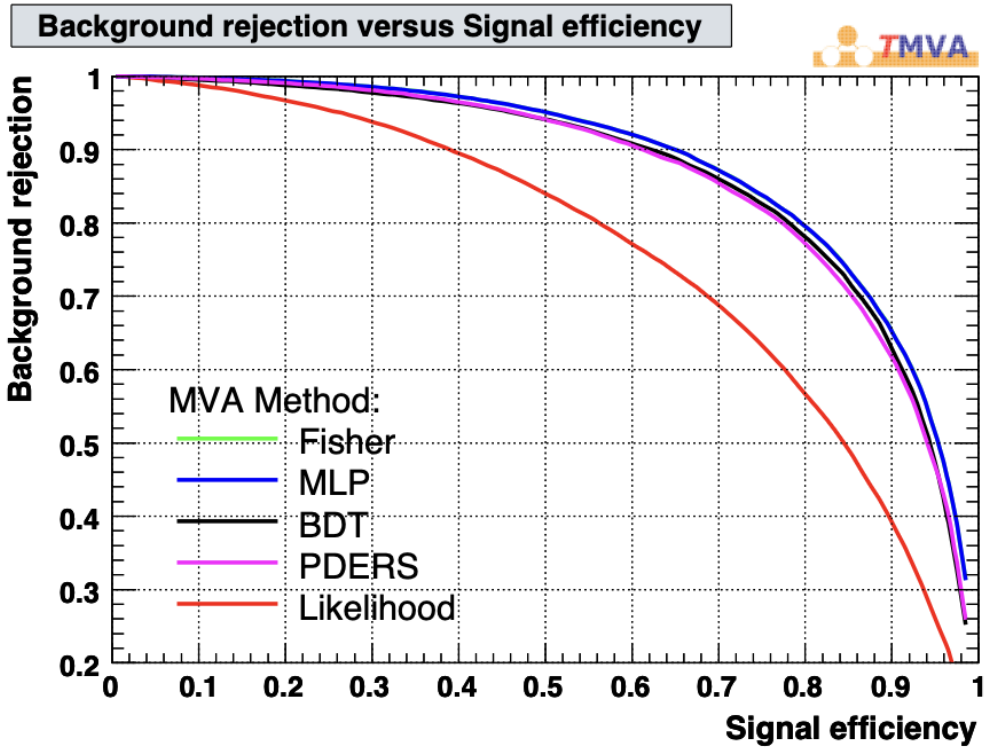




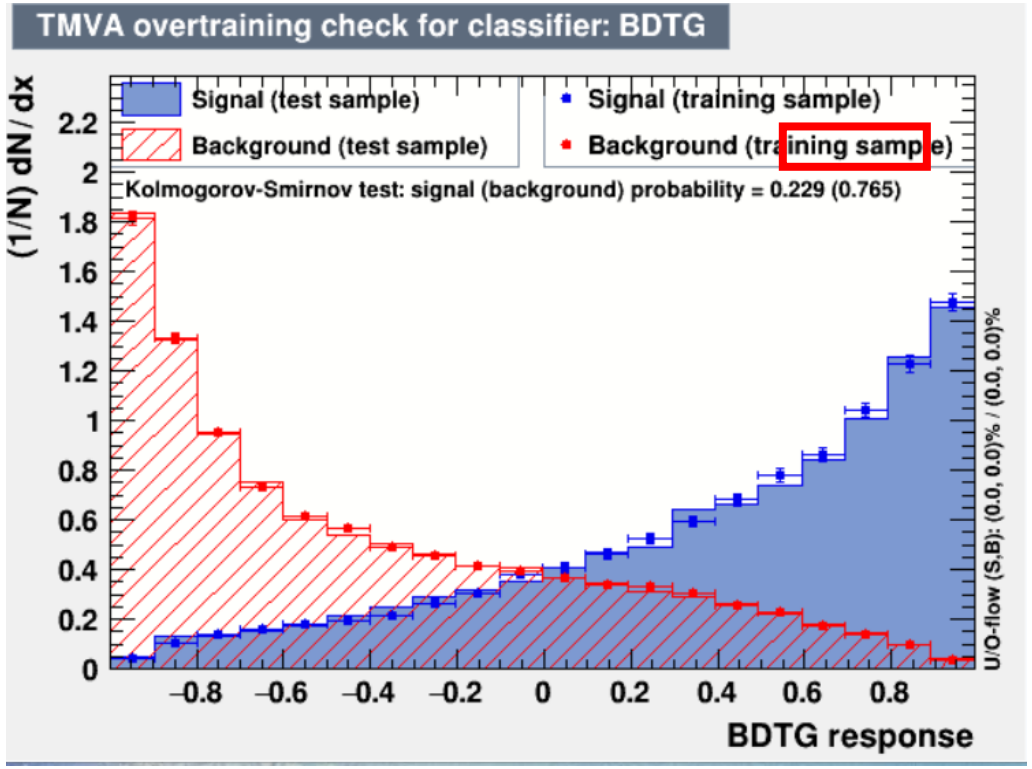
ROC-curve and ROC-int values

ROC (Receiver Operating Characteristic) Curves are a good way to illustrate the performance of given classifier

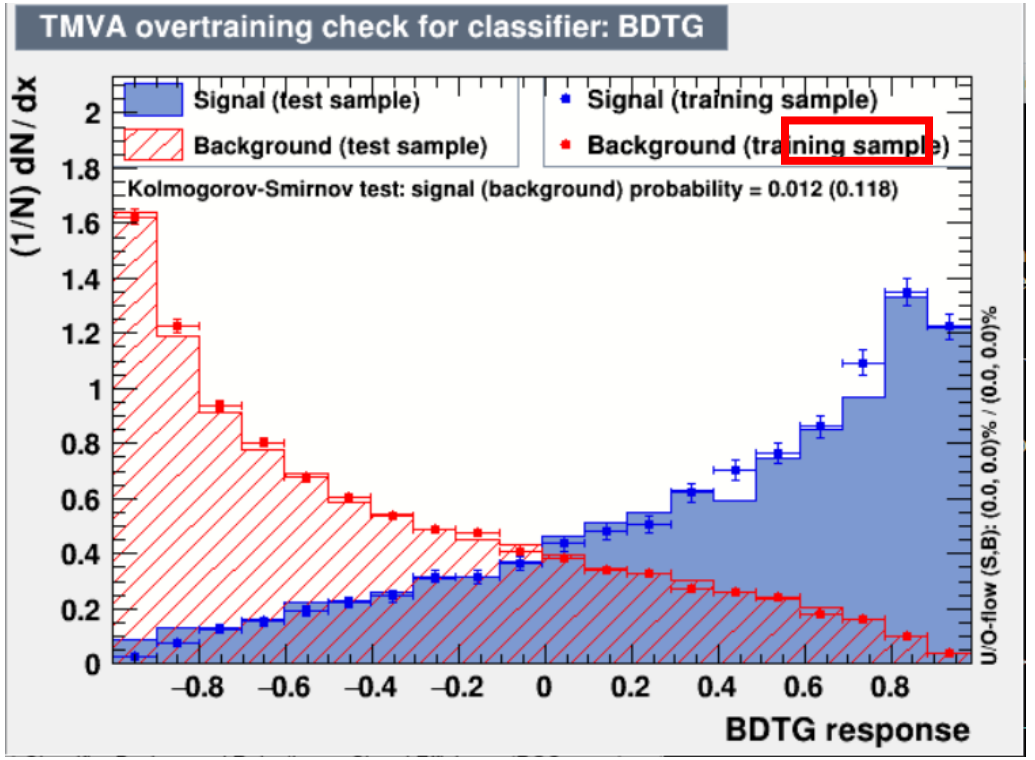
- Shows the background rejection over the signal efficiency of the remaining sample
- Best classifier can be identified by the largest AUC (Area under curve)



K-S test



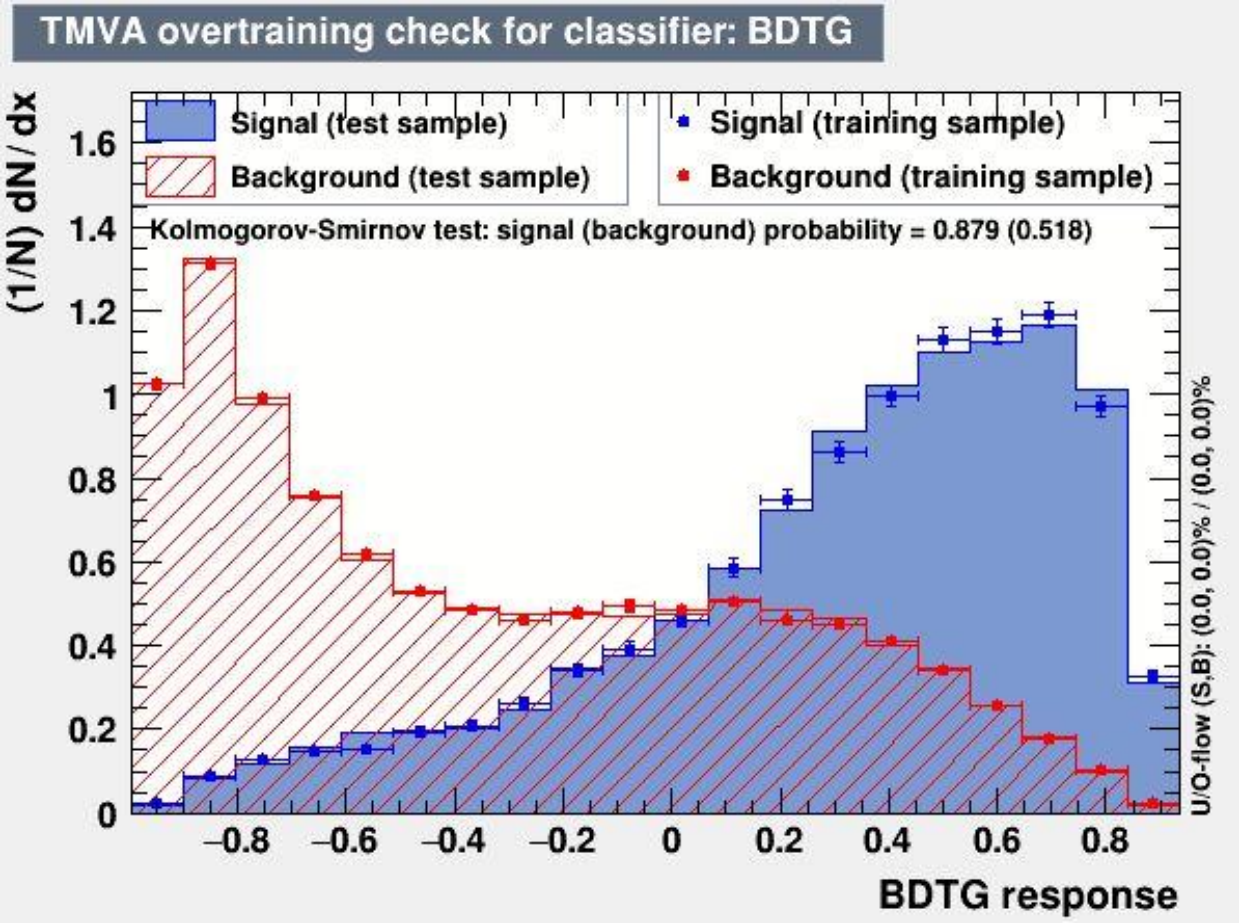
Var 1 : with old variables



Var 2: old+ new variables

- A Kolmogorov Smirnov test is applied between the training and the test output distribution.
- In case the two distributions are compatible a random value between 0 and 1 should be obtained, if they are not you will observe a difference and this might be an indication of overtraining.
- So a very small value is not good.
- A value which is not small is good, while a value too close to 1 is also no so good, because indicates that the statistical fluctuations are too small, and this can be also an indication of some potential problems.

Summary of new TMVA/ML studies – Results for odd/even case




```
Number of training and testing events
-------------------------------------------------------
Signal     -- training events             : 46405
Signal     -- testing events              : 46859
Signal     -- training and testing events: 93264
Background -- training events             : 8787156
Background -- testing events              : 8778348
Background -- training and testing events: 17565504
```

```
-----------------------------------------------------------
Rank : Variable                  : Variable Importance
-----------------------------------------------------------
   1 : feta12                    : 1.373e-01
   2 : met_tst_nolep_et          : 1.253e-01
   3 : m2_rel                    : 1.239e-01
   4 : feta13                    : 1.122e-01
   5 : jj_deta                   : 1.077e-01
   6 : met_tst_nolep_j1_dphi_new : 1.049e-01
   7 : met_tst_nolep_j2_dphi_new : 1.039e-01
   8 : jet_pt[1]                 : 9.463e-02
   9 : m3_rel                    : 9.016e-02
-----------------------------------------------------------
```

```
ROC-integ
: 0.831
```

in progress

Fitting strategy for limit setting

Fitting is the process of adjusting a theoretical model to real experimental data using the Maximum Likelihood Fit method.

- A global mathematical likelihood function is constructed, which is the product of Poisson distributions over all bins (distribution intervals) across all SR and CR regions.

Main likelihood (LL) or fitting model:

$$\mathcal{L}(\mu, \vec{\beta}_V, \vec{n}_{\text{PU-MJ/fake-}e\text{/fake-}\mu}, \vec{\theta}) = \mathcal{L}^{\text{SR}} \cdot \mathcal{L}^{\text{PU-CR}} \cdot \mathcal{L}^{V\text{+jets-CR}} \cdot \mathcal{L}^{\text{fake-CR}} \cdot \mathcal{L}^{\text{NP}},$$

Likelihood for Signal region (SR), for other fit regions can be found in run 2 paper:

$$\mathcal{L}^{\text{SR}} = \prod_i \mathcal{P}\left(N_i^{\text{SR}} \mid \beta_i \cdot B_{Z,i}^{\text{SR}} + \beta_i \cdot B_{W,i}^{\text{SR}} + B_{\text{MJ},i}^{\text{SR}} + B_{\text{other},i}^{\text{SR}} + \mu \cdot S_i^{\text{SR}}\right),$$

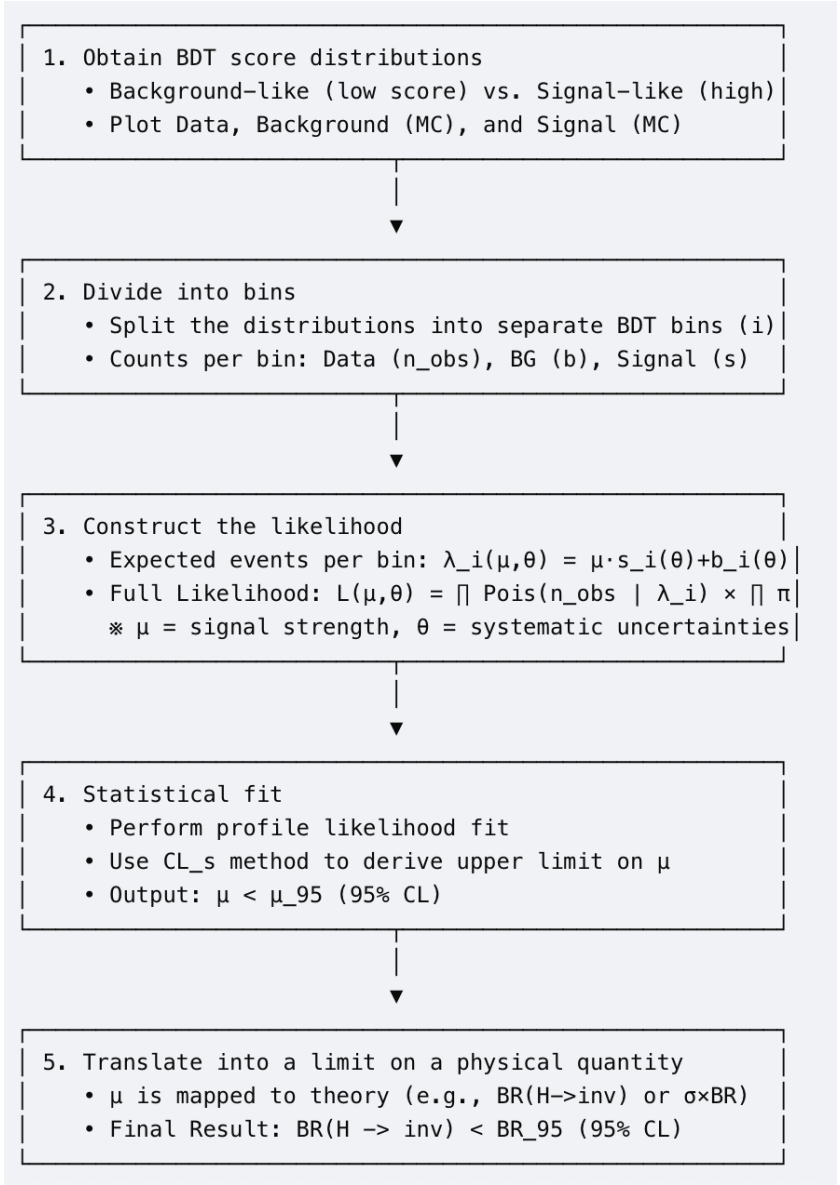
Likelihood for nuisance parameter (modeling systematic sources):

$$\mathcal{L}^{\text{NP}} = \prod_j \mathcal{G}\left(0|\theta_j\right),$$

the HistFactory package begins to vary μ and all θ parameters simultaneously to maximize the likelihood function and find $\hat{\mu}$, $\hat{\theta}$

If after fitting the value of $\hat{\mu}$ turns out to be close to zero (meaning no clear excess of signal events over the background is found in the SR), we set an upper limit on the BR.

Limit setting with BDT score



in progress