

# **Search for dark matter production in association with a single top quark or a top quark pair in the dilepton final state at $\sqrt{s} = 13$ TeV**

---

**Cédric Prieëls**

Director: Jónatan Piedra Gomez

Codirector: Pablo Martínez Ruiz del Árbol

Sala de grados, Facultad de Ciencias, Universidad de Cantabria

PhD thesis defense

- March 28th 2022 -

- List of my contributions to CMS
- Analysis introduction
- The dark matter case
- The experimental setup
  - The LHC accelerator
  - The CMS detector
- Event reconstruction
- Samples and objects definition
- Backgrounds prediction methods
- Signal extraction
  - Discriminating variables
  - Multivariate analysis
- Systematic uncertainties
- Results and interpretation
- Conclusions and future prospects

In summary, I was involved in the CMS collaboration through several axes:

- I first of all participated to the **2016  $t\bar{t}$ +DM analysis**, published in 2019 in Physical Review Letters B;
- I am the **main IFCA investigator** in charge of this particular dark matter search with top quark pairs in the dilepton final state;
- I am also responsible for the **computation and application of the data-driven method** used to compute the non-prompt background for several Higgs analyses, published in Physics Letters B and JHEP;
- I had the opportunity to present my work and dark matter global summaries to several **national and international conferences**;
- I was also a **central shifter** during my PhD, involved mostly in the DQM and DCS work stations during data taking periods and the last long shutdown;
- And, finally, I was involved for more than 3 years in the **central muon Monte-Carlo validation of new CMSSW releases**.

$t/t\bar{t}$ +DM analysis

A search for the production of dark matter particles in association with either one or two top quarks is presented:

- We study the proton-proton collisions produced by the LHC at  $\sqrt{s} = 13$  TeV;
- The data has been collected by the CMS detector;
- Part of the legacy analysis, considering the full Run II dataset (data collected in 2016, 2017 and 2018 and summing around  $137 \text{ fb}^{-1}$ ).

### Motivation

- Several (mostly astrophysical) evidences for the existence of dark matter, however **its nature remains unknown** and it has never been detected experimentally;
- If dark matter is made of some kind of particle it might be produced in the high energy collisions of the LHC.

### Main objective

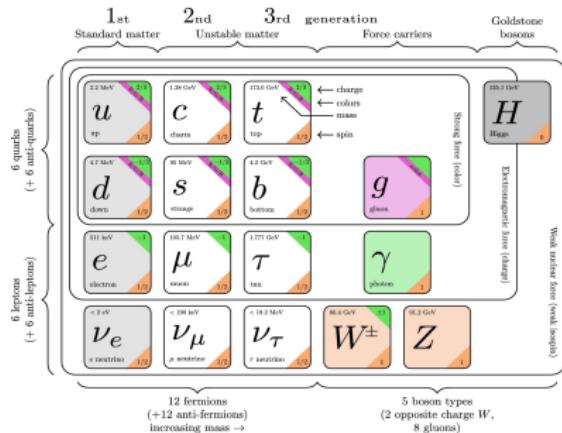
- Consider different dark matter production models to discover or eventually exclude them, **setting upper limits on their cross section of production**.

## The dark matter case

# The Standard Model of Particle Physics

The most accepted model to describe the elementary particles and three of the four of the fundamental interactions between them is the **Standard Model**:

- Contains 26 free parameters, among which the masses of the 12 predicted fermions;
- **Many successful predictions made over the years**, such as the existence of the top quark, and the  $W^\pm$ ,  $Z^0$  and Higgs bosons.



However, **this model has some fundamental shortcomings**, such as the possible existence of exotic particles which do not fit within this model (such as dark matter), therefore extensively searched for nowadays.

## At the origins of dark matter I

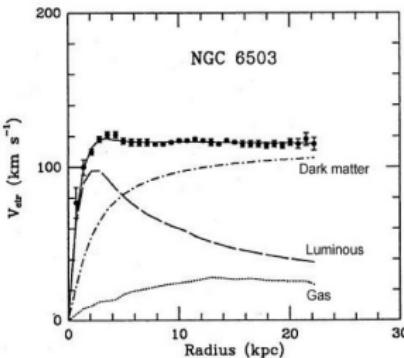
The concept of dark matter can be traced back to the 19th century, and was introduced to explain several astrophysical evidences, among which:

### Zwicky's calculations

- Measurement of the mass of the Coma Cluster using the virial theorem;
- Concluded that its mass was **400-500 times larger** than the value obtained by Hubble, considering only visible galaxies<sup>1</sup>.

### Spiral galaxies rotation curves

- Stars within spiral galaxies should rotate with a velocity depending on the radius to the galactic center, but **this is not what is observed experimentally**.
- Either our understanding of gravity at large scales or our basic understanding of galaxies as celestial bodies made of stars has to be revised.

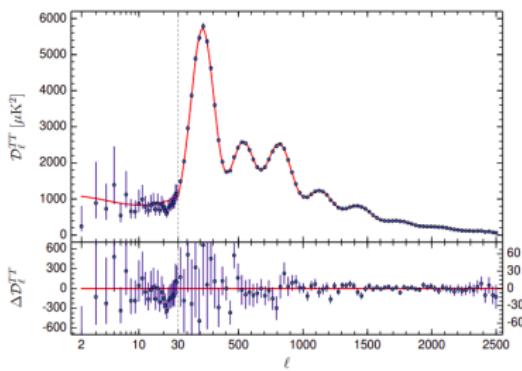
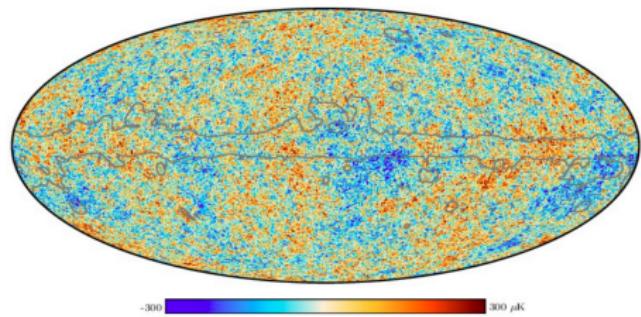


K.G. Begeman, A.H. Broeils, R.H. Sanders. 1991. Mon.Not.RAS 249, 523

<sup>1</sup>These numbers were later proven to overestimated, but the idea behind the calculations is still accepted

## CMB anisotropies

- Background of primary radio waves emitted when the Universe became transparent around 380 000 years after the Big Bang;
- Can be considered as emitting a black body spectrum with a temperature of  $(2.72548 \pm 0.00057)\text{K}$ , but small anisotropies at the  $10^{-5}$  level are observed;
- Implies that dark matter **accounts for  $\sim 27\%$  of the total mass of the Universe.**



Other observations, such as the gravitational lensing effect, also tend to further support the existence of dark matter (cf. backup).

Several fundamental properties of dark matter are nowadays known or assumed:

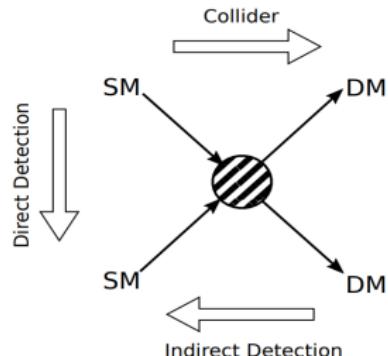
- Dark matter is a particle, assumed to have a certain mass;
- It should be dark, unable to interact with electromagnetic radiation, otherwise we would have seen it already. It should then also be electrically neutral;
- It is non-baryonic, because the energy density for the baryonic matter estimated from the CMB is too low to account for dark matter;
- We only consider cold dark matter since the widely accepted  $\Lambda_{CDM}$  model is based on this assumption and this can explain the large scale structures in the Universe;
- Compatibility with the relic density obtained from the thermal freeze-out mechanism implies that it should have a mass in the electroweak scale;
- Finally, it should be long-lived, since we expect that it has been produced during the Big Bang and is still present in the Universe.

### Weakly Interactive Massive Particles

The WIMPs are the dark matter candidates considered in this work, because of the so-called **WIMP miracle**. Indeed, they:

- Are expected to interact very weakly with ordinary baryonic matter;
- Have a mass in the 100 GeV-1 TeV range for reasonable electroweak production cross-section values;
- Give us a dark matter candidate which, in some models, is able to solve the **hierarchy problem**.

### Main search strategies



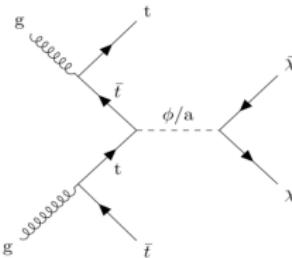
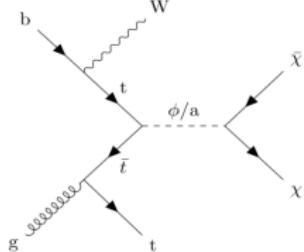
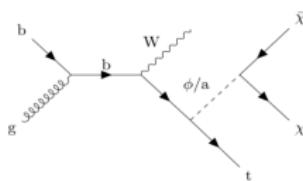
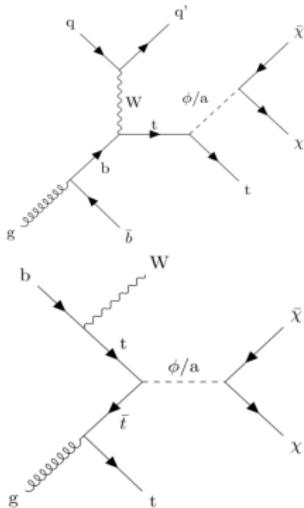
Different search strategies are used:

- The **direct and indirect searches**, relying on the production of baryonic matter from the interaction between two DM particles or on the observation of the interaction between the dark and baryonic sectors;
- And the **collider production**, able to probe lower dark matter candidates masses.

## Focus of this thesis I

We are searching for **dark matter produced in association with either one or two top quarks**. Several **simplified models** have been considered:

- Spin 1/2 DM  $\chi$  (mass  $\in [1, 55]$  GeV, Dirac fermion)
- Spin 0 scalar (S)/pseudoscalar (PS) mediator  $\phi/a$  (Yukawa-like structure of such interactions → **gain from the coupling of the mediator to top quarks**)
- Mediator mass  $\in [10, 1000]$  GeV
- Coupling  $g_\chi$  mediator/DM set to 1 (same for all  $g_q$  couplings)



The **typical final state** of such topology is made out of:

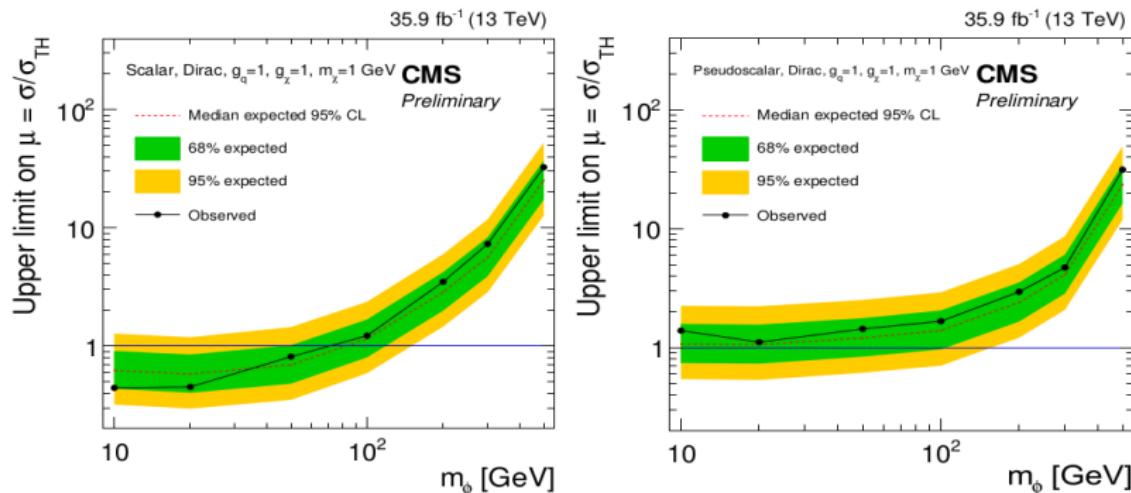
- 1 or 2 bottom quarks coming from the decay of the top quark(s);
- 2 W bosons, seen as a combination of jets and leptons depending on the channel, produced directly or coming from the decay of the top quarks ( $t \rightarrow b + W$ );
- Some missing transverse energy coming from the dark matter and the neutrinos from the leptonic decay of the Ws.

In particular, we are studying the **dilepton final state** in this work:

- Has the lowest branching ratio:  $\text{BR}(W^+ \rightarrow l^+ + \nu_l) = (10.80 \pm 0.09)\%$  for each of the charged leptons (contains only  $\sim 5\%$  of the signal events);
- But, electrons and muons can usually be reconstructed better than jets, leading to lower systematic uncertainties (taus are not considered directly in this work);
- And this channel has smaller backgrounds, with cross-sections typically lower, resulting in a better signal isolation.

## Previous relevant results I

A similar analysis was carried out by CMS using 2016 data, considering only the  $t\bar{t}+DM$  signal and the dilepton final state (EXO-17-014).



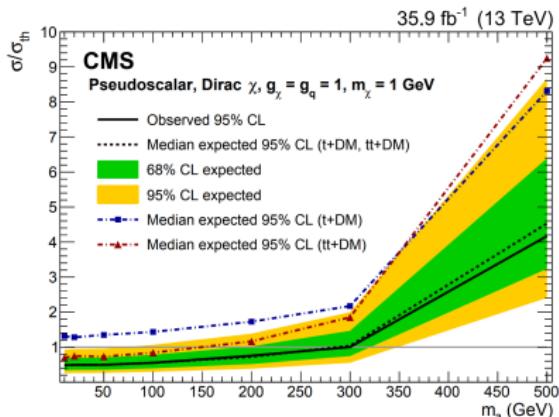
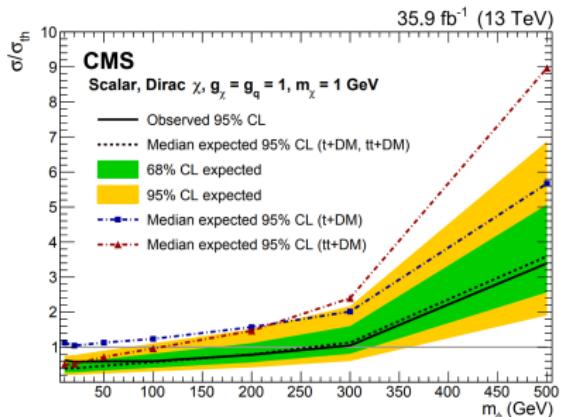
This analysis excluded scalar mediators with masses below 80 GeV, while no exclusion was achieved when considering pseudoscalar mediators.

A combination of the three possible final states (hadronic, semileptonic and dileptonic) was also performed by CMS (EXO-16-049).

## Previous relevant results II

A combination of both the  $t/\bar{t}+DM$  and  $t\bar{t}+DM$  processes has also been performed by CMS (EXO-18-010). The inclusion of the single top signal process improved up to a factor 2 the limits obtained by the  $t\bar{t}$  analysis on its own. This analysis:

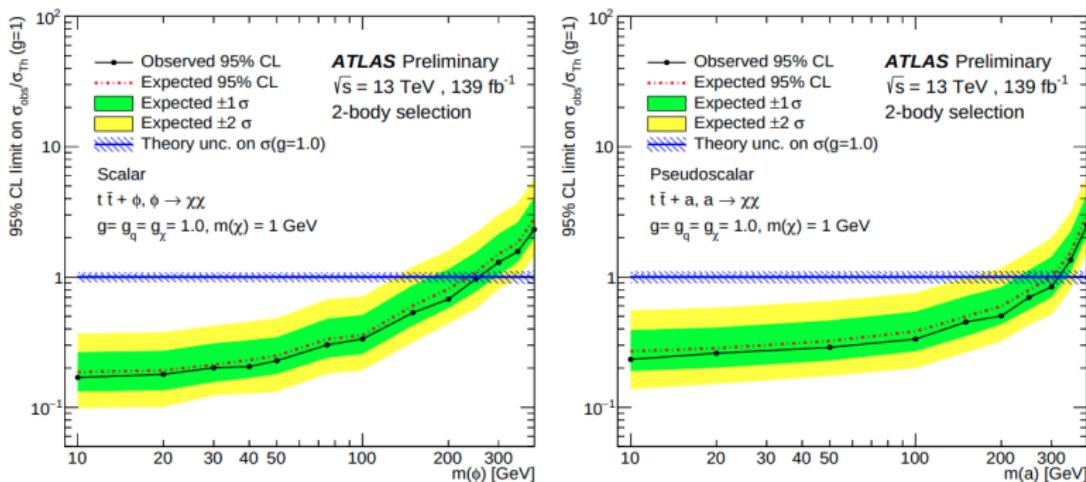
- Only considered the 2016 data-taking period;
- And only considered the semileptonic and hadronic final states.



Scalar (pseudoscalar) mediators were with this combination excluded up to 290 (300) GeV at the 95% confidence level.

### Previous relevant results III

The ATLAS collaboration also obtained exclusion limits using the full Run II legacy dataset and considering the  $t\bar{t} + \text{DM}$  dilepton final state only (ATLAS-CONF-2020-046).



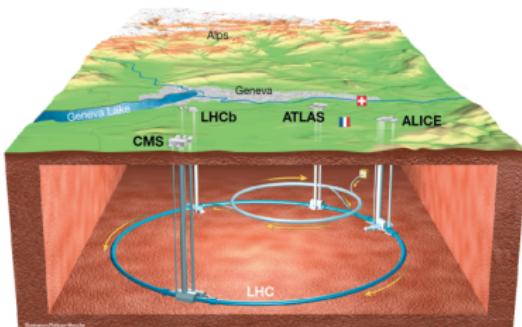
They obtained **expected scalar (pseudoscalar) exclusion limits of 250 (300) GeV**, by using NLO cross-sections for the signals, around 20-30% higher than ours.

## The experimental setup

## The Large Hadron Collider

The data analyzed has been taken by the Large Hadron Collider:

- A 27 km circular underground proton-proton collider, located at CERN;
- Provided the collisions that led to the discovery of the Higgs boson in 2012;
- Currently the most powerful accelerator in the world with its  $pp$  center of mass energy  $\sqrt{s} = 13$  TeV, therefore able to **scan new parts of the phase space**.



The data considered in this work has been collected during the **Run II of operation of the LHC** (from 2016 to 2018), at 13 TeV, totaling  $(137.1 \pm 2.0) \text{ fb}^{-1}$  of data, selected by the different levels of trigger from the 40 MHz collision rate.

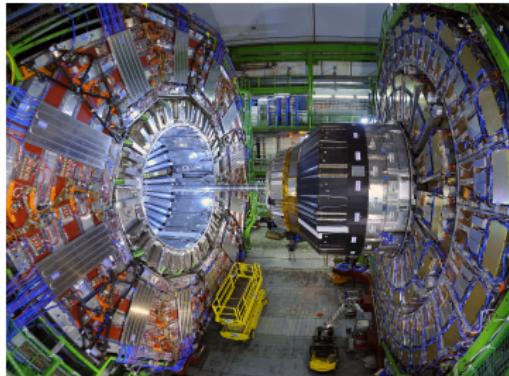
## The CMS detector I

The **Compact Muon Solenoid** is one of the two **general purpose detectors** of the LHC. Its main objectives consist in:

- Search for and study the properties of the Higgs Boson;
- Measure properties of the Standard Model (top physics, b physics, etc);
- Discover new physics phenomena, such as the possible existence of dark matter.

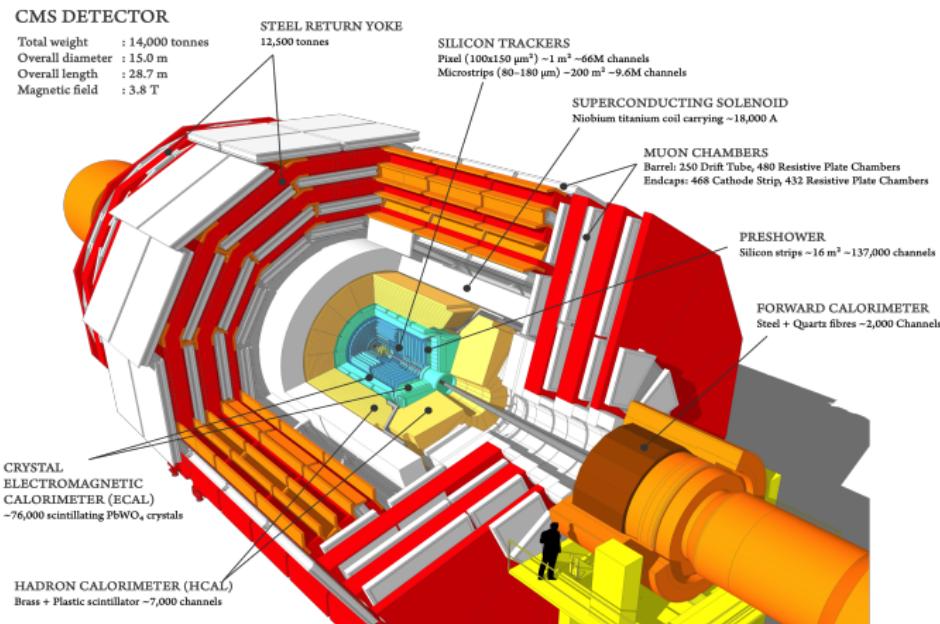
In a nutshell, CMS:

- Is a relatively **compact** (14.000 tons, distributed over a diameter of 15 meters and a length of 8 meters) cylindrical detector;
- Is made out of a central part with cylindrical shape, the barrel, and two endcaps, one on each side in order to be **as hermetic as possible**, covering all the possible angles around the beam pipe;
- Has a **large solenoid** as middle piece, able to produce a 3.8 T field;
- Features a **powerful tracker and muon detection system**.



## The CMS detector II

The CMS detector is also **made out of different layers**, each having its own purpose, allowing to identify unequivocally each particle created and measure their properties, thanks to the reconstruction performed by the Particle Flow algorithm.



Each subdetector of CMS has been **designed carefully to fulfill a specific purpose**:

- The **tracker** is the innermost piece of CMS, able to reconstruct the trajectories of charged particles issued from the primary and secondary interaction vertices in a quick and precise way;
- The **Electromagnetic Calorimeter**, enclosing the tracker system and able to give information about the energy of electrons and photons;
- The **Hadronic Calorimeter**, able to measure the energy of incident hadrons from the nuclear interactions happening in its core;
- The 12.5 meters long and 6 meters large **solenoid**, allowing to measure precisely the charge and momentum of particles produced using the Lorentz effect, by measuring the curvature of their tracks;
- And finally, the **muon system**, covering more than  $25\ 000\ m^2$  and made out of three different subsystems:
  - The **Drift Tubes** (DTs), in the barrel region;
  - The **Cathode Strip Chambers** (CSCs) in the two endcaps;
  - And the **Resistive Plate Chambers** (RPCs), mostly added to the barrel and to the endcap regions in order to cope with (in)ability of the previous muon system to identify unequivocally the correct bunch crossing when the LHC is running at full luminosity.

## **Analysis context**

## Analysis context

Run II legacy paper being worked on, expected to combine both the  $t/\bar{t}+DM$  and  $t\bar{t}+DM$  searches, and the 3 possible final states (hadronic, semileptonic and dileptonic).  
→ Pre-approval process expected to start within a few weeks.

The effort is **globally common** between the groups (Wisconsin, DESY, IFCA) studying the three different final states:

- Objects are defined in a common way;
- Control and signal region orthogonal between the channels.  
→ Number of leptons and b-jet categorization to improve the sensitivity by defining enriched  $t/\bar{t}+DM$  and  $t\bar{t}+DM$  regions.

This talk will however **be focused on the dilepton final state only**.

All the results presented here have not been published yet, but have been fully endorsed by the MET+X conveners of the CMS collaboration.

## **Samples and objects**

### Data

Single/double leptons datasets built to avoid any eventual double counting, considering the 3 years of the Run II of operation of the LHC:

- $(35.9 \pm 0.9) \text{ fb}^{-1}$
- $(41.5 \pm 1.0) \text{ fb}^{-1}$
- $(59.7 \pm 1.5) \text{ fb}^{-1}$

A **blinding policy** has been followed at first, allowing us to only look at  $1 \text{ fb}^{-1}$  of data per year near the signal regions. The **unblinding** was done with the analysis freezed and the green light received from the conveners.

### Backgrounds

The major backgrounds have been mostly considered from MC. Each year comes with its corresponding MC samples:

- $t\bar{t}$ : considering both its semileptonic and dileptonic decays;
- Single top: s, t and tW channels considered;
- Drell-Yan: HT-binned samples to increase the statistics, with a correction factor derived from data applied;
- $t\bar{t}Z$  and  $t\bar{t}W$ : usually grouped together as ttV, and considering both the hadronic and leptonic final states;
- Others, such dibosons and tribosons production, all taken from MC directly.

## Signal samples

The signals samples have been generated using MADGRAPH and PYTHIA8 (CP5 tune) at LO and simulated events are interfaced with a realistic model of the CMS detector using Geant4 and reconstructed using the official CMS reconstruction algorithms.

The  $t/\bar{t}$ +DM and  $t\bar{t}$ +DM processes are both considered:

- Both scalar and pseudoscalar mediators are considered;
- 400.000 events were produced for each mediator mass, from 10 to 1000 GeV;
- The dark matter mass was set to 1 GeV, but additional samples ranging from 1 to 55 GeV were also produced;
- All the  $g_q$  and  $g_\chi$  couplings were set to 1.

**Recommended correction factors** (L1 ECAL prefiring in 2016 and 2017, HEM issue in 2018) are then applied to the simulation.

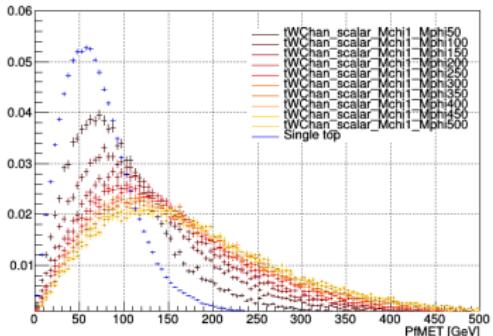
→ All the samples used and their cross-sections are listed in the backup.

# Signal samples distributions with normalization

$t/\bar{t}+DM$

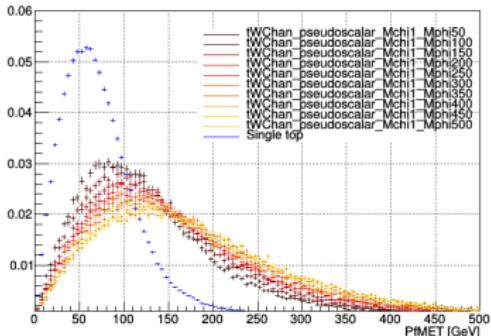
## Scalar

Mass points distribution



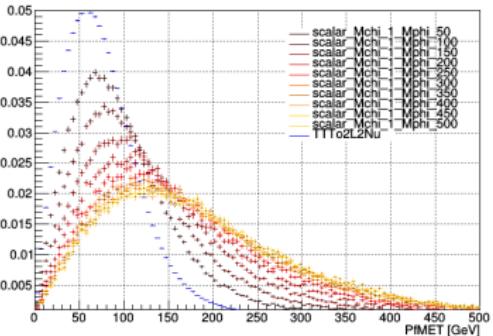
## Pseudoscalar

Mass points distribution

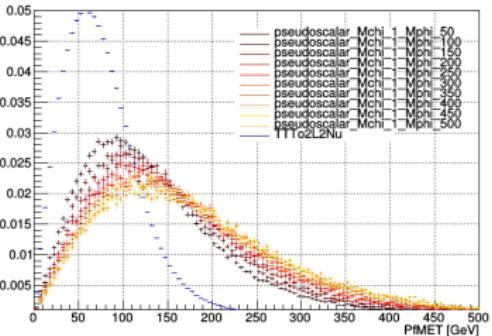


$t\bar{t}+DM$

Mass points distribution



Mass points distribution

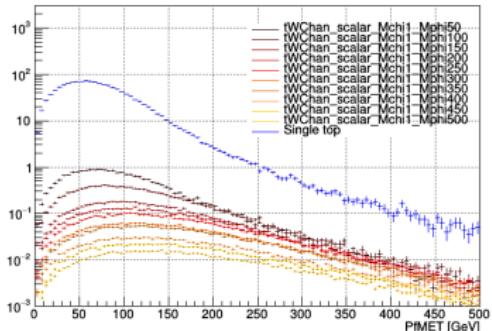


# Signal samples distributions without normalization

$t/\bar{t}+DM$

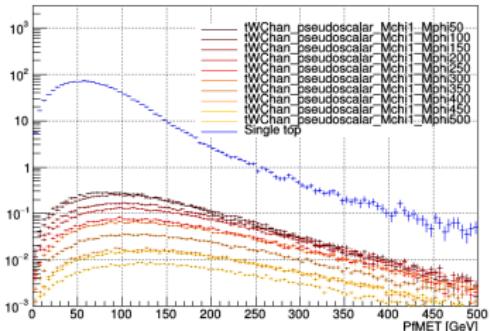
## Scalar

Mass points distribution



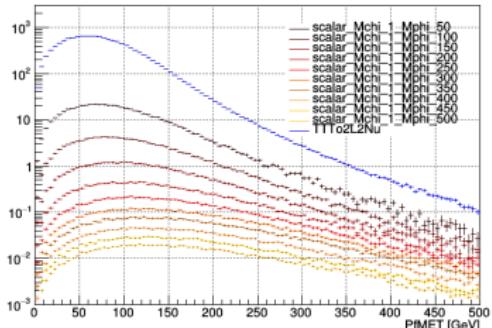
## Pseudoscalar

Mass points distribution

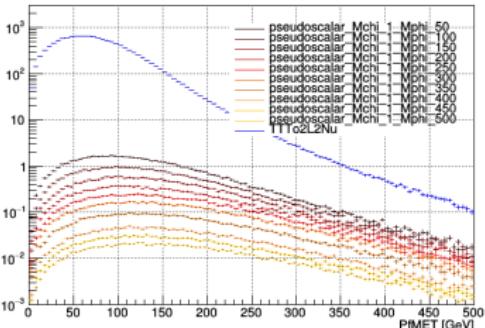


$t\bar{t}+DM$

Mass points distribution



Mass points distribution



### Triggers

- Single and double lepton triggers combined to gain statistics, and any possible double counting of events in multiple trigger is taken care of;
- Trigger and lepton  $p_T$  carefully chosen to avoid any turn-on effect;
- SingleMuon, SingleEle, DoubleMuon, DoubleEG, MuonEG (2016) and SingleMuon, EGamma, DoubleMuon, MuonEG (2017/2018) data streams considered;
- All the triggers used and their efficiencies (computed using orthogonal MET datasets) are listed in the backup.

### Leptons

- Analysis relies on the selection of events with two leptons, with a leading (trailing)  $p_T > 25$  (20) GeV and  $|\eta| < 2.4$ ;
- Medium cut based POG WP used for electrons without additional isolation cut;
- Medium cut based POG WP for muons with tight isolation ( $\text{pfRelIso04\_all} < 0.15$ );
- Additional small cuts on the impact parameters to reduce the non-prompt contamination in the ptmiss tail ( $|d_0| < 0.05$  cm,  $|d_z| < 0.1$  cm,  $S_{3D}^d < 4$ ).

### Jets

- Clustered from the PF candidates using the **anti-kT algorithm**;
- Basic selection:  $p_T > 30 \text{ GeV}$ ,  $|\eta| < 2.4$ ;
- **Tight JET/MET POG working point** (efficiency and background rejection  $> 98\%$ ), **tight jet PU ID** applied to jets with  $p_T < 50 \text{ GeV}$  to reject PU jets contamination;
- $\Delta R > 0.4$  away from any lepton passing the criteria established for analysis to prevent signal leptons clustered as jets from entering the jet counting.

### B-tag

- B-Tagging and Vertexing POG **deep CSV b-tag medium working point** (high efficiency, misidentification rate for a light jet as a b-jet  $\sim 1\%$ ).

### Missing transverse momentum

- **PfType1MET** considered by propagating the JECs to the ptmiss;
- All **recommended filters applied** to filter anomalous high ptmiss events due to several detector issues, such as eventual dead cells in the calorimeters;
- XY-shift ( $\phi$  modulation fix) and EE noise (2017) corrections applied on top.

## **Event selection**

### Minimal event selection

We require for the analysis at least:

- Two opposite sign leptons, with leading (trailing)  $p_T > 25$  (20) GeV;
- Third lepton veto ( $p_T < 10$  GeV);
- A dilepton invariant mass  $m_{ll} > 20$  GeV to avoid low mass resonances;
- At least 1 jet.

### Pre-selection region

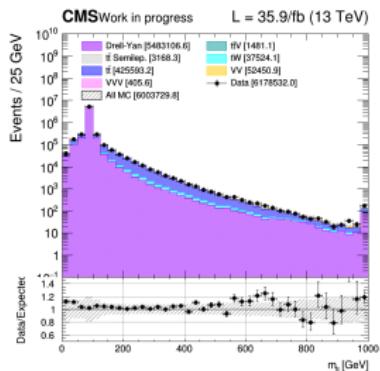
A pre-selection region is then defined by additionally asking for:

- At least 1 medium deep CSV b-jet;
- A 76-106 GeV Z-veto on the ee and  $\mu\mu$  channels;
- $\text{ptmiss} > 100$  GeV and transverse mass  $M_{T2}^{\text{ll}} > 80$  GeV to keep this region orthogonal to the  $t\bar{t}$  control regions used by the semileptonic channel.

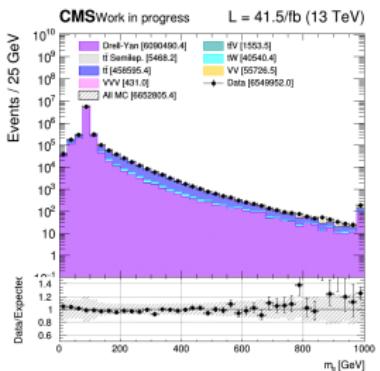
This region is used as the **basis for the definition of our signals regions**.

# Minimal event selection region

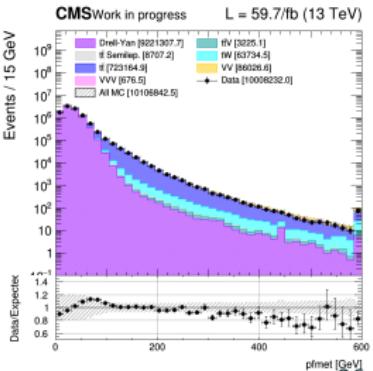
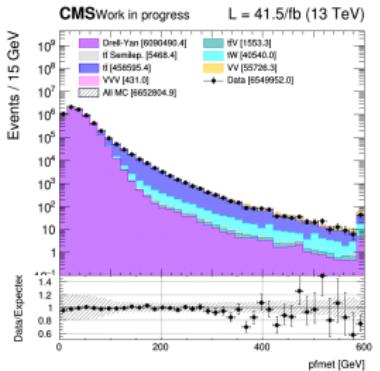
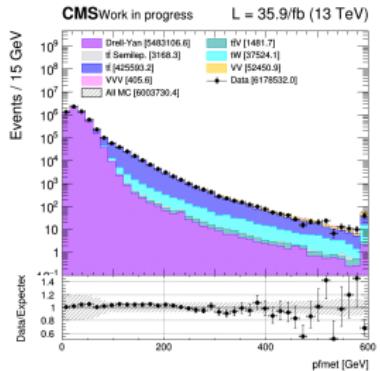
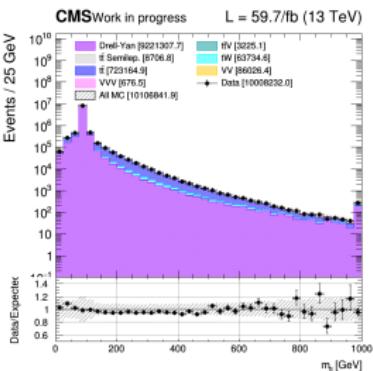
2016



2017

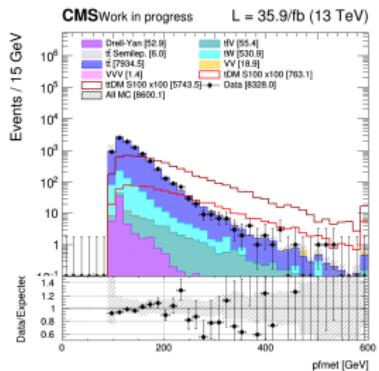


2018

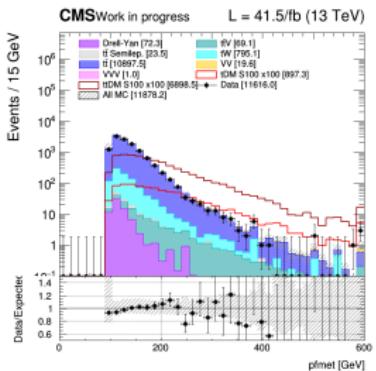


# Pre-selection region

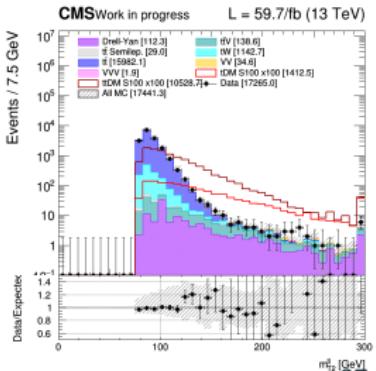
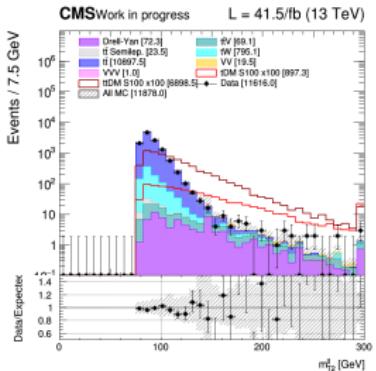
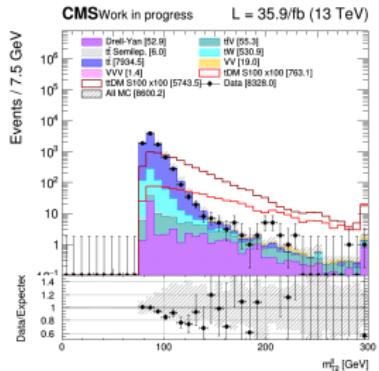
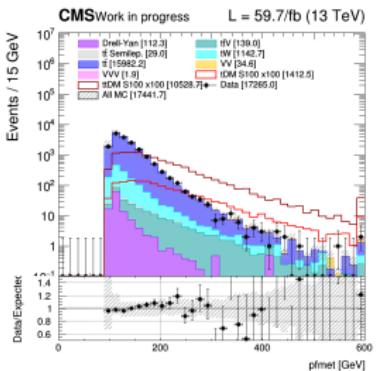
**2016**



**2017**



**2018**



## **Background prediction methods**

## Main background processes

The backgrounds are predicted either directly from Monte-Carlo simulations or from semi data-driven methods.

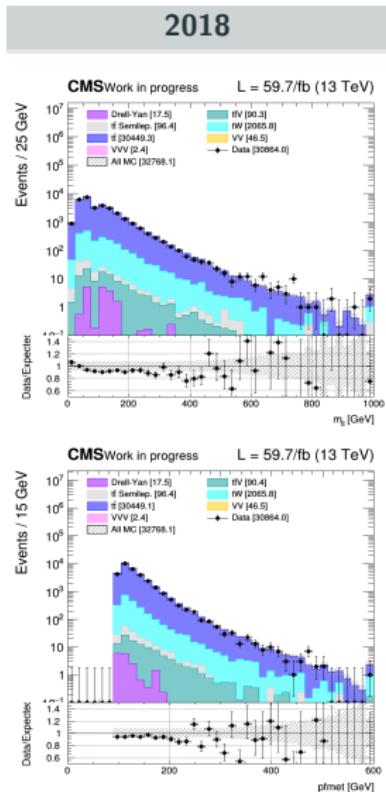
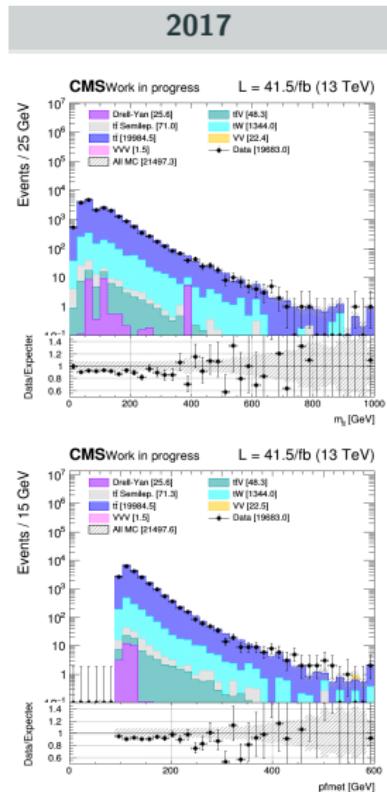
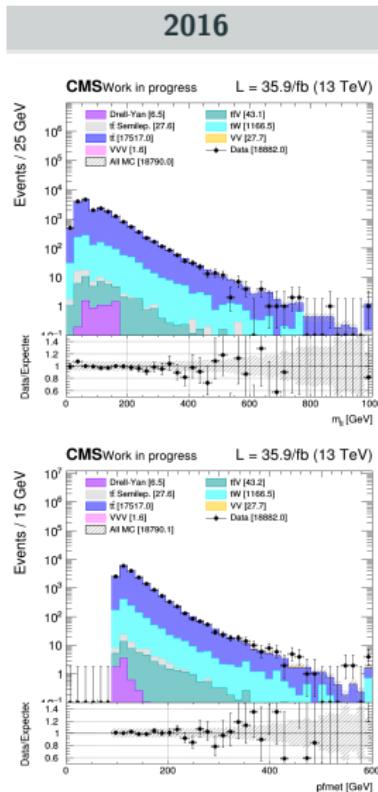
- The **t $\bar{t}$  and the single top** are taken from simulation accounting for all the variations in the generation parameters;
- The **Drell-Yan** yields are obtained from a semi data-driven method using the excluded same flavor region on the Z peak as control region;
- The **ttV, diboson, triboson processes and other minor backgrounds** are all taken directly from MC simulations.

Several parameters (QCD scale, PDF variation,...) are varied and included as a systematic (see later), and **recommended correction factors** (L1 ECAL prefiring in 2016 and 2017, HEM issue in 2018) are also applied to the simulation.

**Several data validation regions** enriched in top and Drell-Yan have been explored to ensure the quality of the prediction made.

## Top control region

Same as the pre-selection region but with  $\text{ptmiss} > 50 \text{ GeV}$  and  $60 < M_{T2}^{\text{II}} < 80 \text{ GeV}$ .



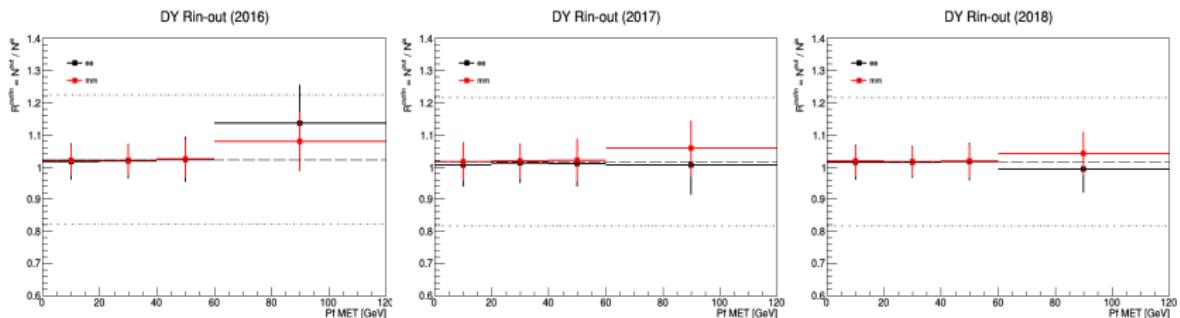
## DY Rin-out method

We want to find a way to estimate the DY yields outside of the Z-peak from the Z-peak data, in the minimal event selection region:

- Given the presence of large backgrounds (such as  $t\bar{t}$ ) in the analysis region, we go inside of the Z-peak to compute the **Rin-out factor**:

$$N_{DY}^{out} = \kappa \cdot N_{DY, data}^{in} \cdot \left( \frac{N_{DY, MC}^{out}}{N_{DY, MC}^{in}} \right) \text{ where } \kappa = \frac{R_{in-out, MC}^{0bj}}{R_{in-out, data}^{0bj}}$$

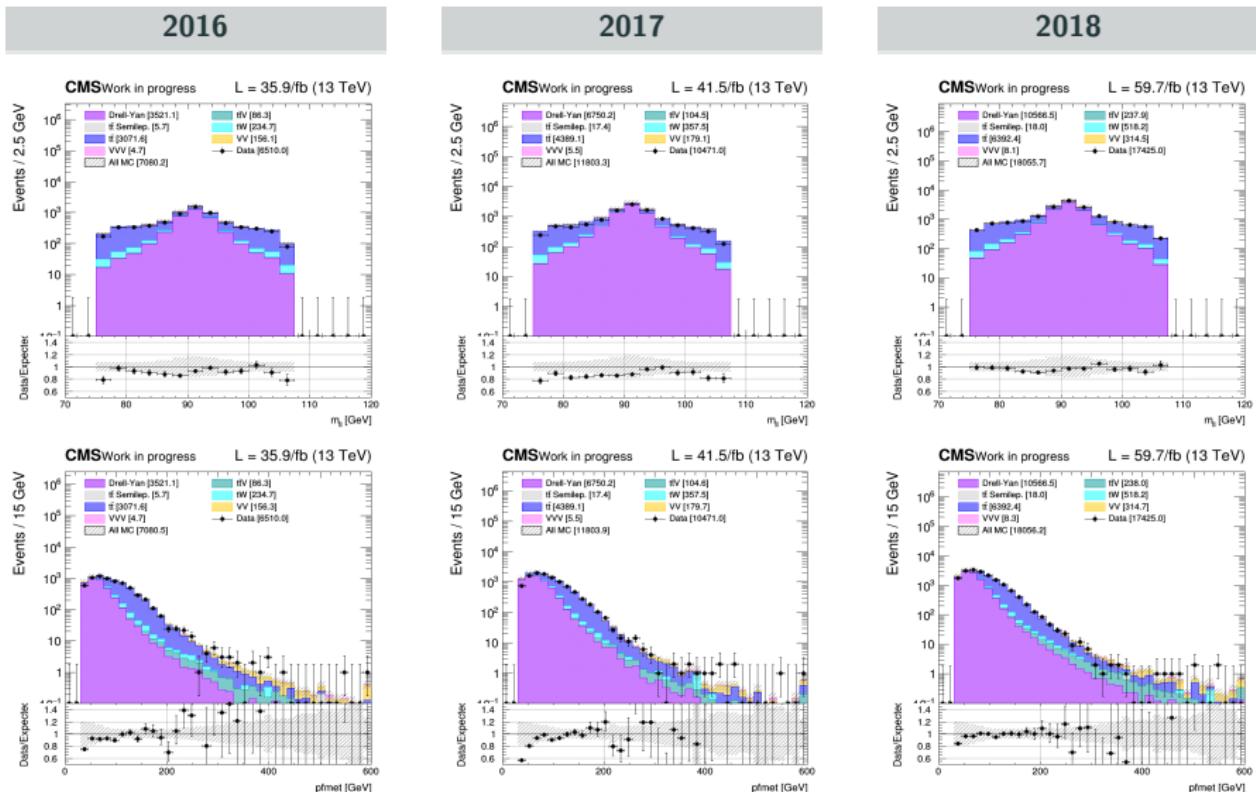
- To avoid any bias, the contamination of non-peaking backgrounds is removed and we correct this factor by the ratio  $\kappa$  between the data/MC transfer factors in a CR close to the SR (asking for 0 b-jet instead of 1);
- We then get this Rin-out factor in **bins of ptmiss** and for each channel ( $ee$ ,  $\mu\mu$ ):



A flat scale factor and a fixed 20% systematic uncertainty is then applied to the DY. 30/53

# DY control region

Same as the pre-selection region but with  $\text{ptmiss} > 30 \text{ GeV}$  and Z-veto reversed.



## **Signal extraction**

In this analysis, **two different signal regions** have been used, targeting each one of our signals of interest and each based on the pre-selection region:

- One **targeting the  $t/\bar{t}+DM$  signal**, by considering events having exactly 1 jet, or exactly 2 jets and 1 b-jet;
- Another one **targeting the  $t\bar{t}+DM$  signal**, by considering events having exactly 2 jets and more than 1 b-jet, or more than 2 jets.

**Several different discriminating variables** ( $\text{ptmiss}$ , stranverse mass  $M_{T2}^H$ , spin correlated variables, etc.) have been considered.

Many of such variables require knowledge of the top quark and anti-quark 4-momenta, only available after a **complete reconstruction of the  $t\bar{t}$  system**, performed whenever possible (details in the backup).

A **BDT** combines the discriminating power of all the variables considered, and an **ANN** is being used as a cross check. A complete optimization was followed in order to select the features which maximize the performance.

## Discriminating variables I

Several discriminating variables (all detailed in the backup) are considered, such as:

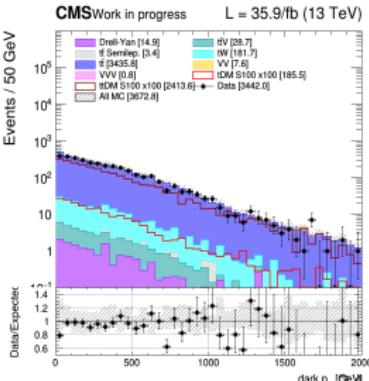
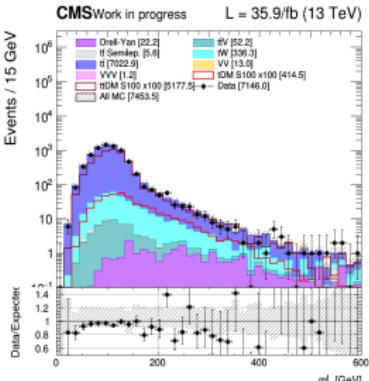
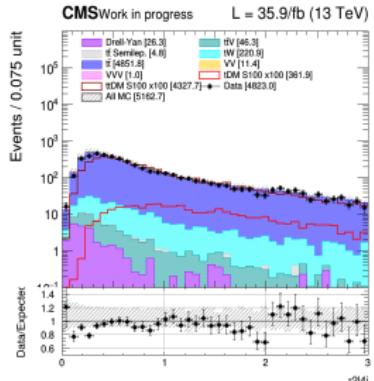
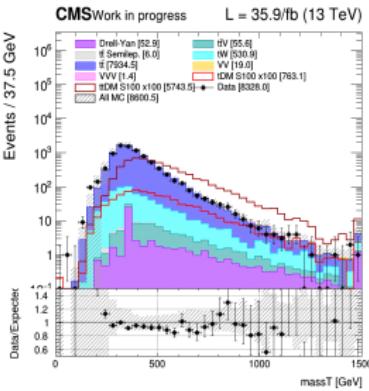
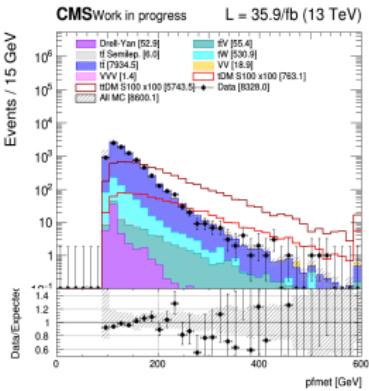
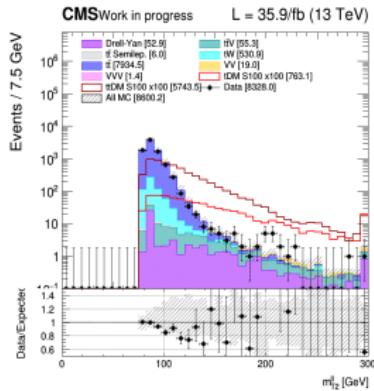
- The transverse mass  $M_{T2}^H$  and missing transverse momentum;
- The number of b-jets (only in the  $t\bar{t}$ +DM signal region) and  $m_{bl}^t$  variable, useful to separate our two signals;
- Several spin correlated variables, such as  $\xi = \cos(\theta_l) \cos(\theta_{\bar{l}})$  and  $c_{\text{hel}}$ ;
- $r_{2l}$  and  $r_{2l4j}$ , defined as the ratio between the  $p_T^{\text{miss}}$  and the  $p_T$  of the leptons (plus the 4 first eventual jets for  $r_{2l4j}$ );
- The dark  $p_T$  and overlapping factor naturally arising from the top reconstruction;
- Other variables, such as the angle  $\Delta\phi$  between the  $p_T^{\text{miss}}$  and the two leptons, and total transverse mass  $\text{massT}$ .

Rank	$t/\bar{t}+\text{DM}$ region		$t\bar{t}+\text{DM}$ region	
	Variable	Importance	Variable	Importance
1	$M_{T2}^H$	$5.96 \cdot 10^{-1}$	$M_{T2}^H$	$5.87 \cdot 10^{-1}$
2	$E_T^{\text{miss}}$	$5.32 \cdot 10^{-1}$	$E_T^{\text{miss}}$	$5.09 \cdot 10^{-1}$
3	$\text{massT}$	$3.11 \cdot 10^{-1}$	$\text{massT}$	$3.98 \cdot 10^{-1}$
4	$r_{2l4j}$	$1.70 \cdot 10^{-1}$	$r_{2l4j}$	$3.45 \cdot 10^{-1}$
5	$m_{bl}^t$	$7.05 \cdot 10^{-2}$	$m_{bl}^t$	$1.72 \cdot 10^{-1}$
6	$r_{2l}$	$2.63 \cdot 10^{-2}$	Dark $p_T$	$3.92 \cdot 10^{-2}$

Table 1: Ranking of importance of the main variables for the scalar 500 GeV case.

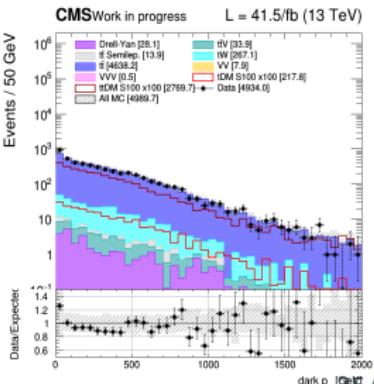
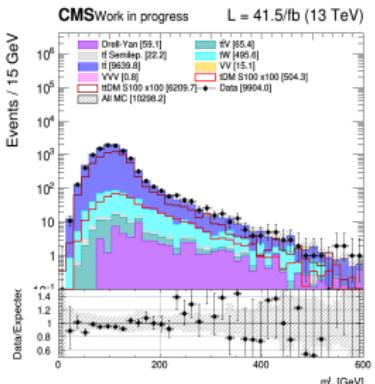
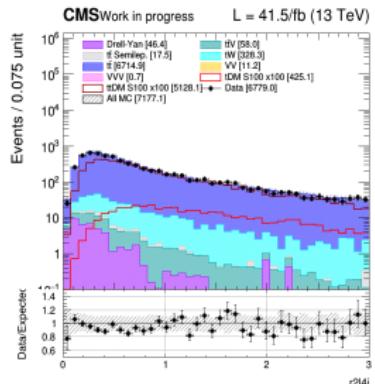
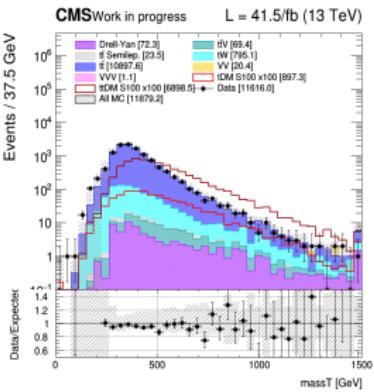
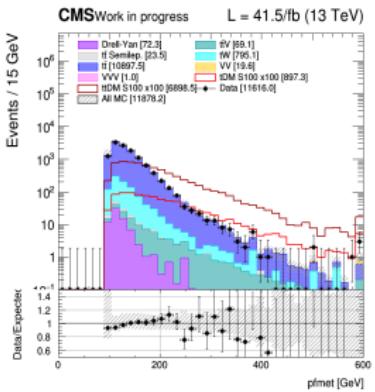
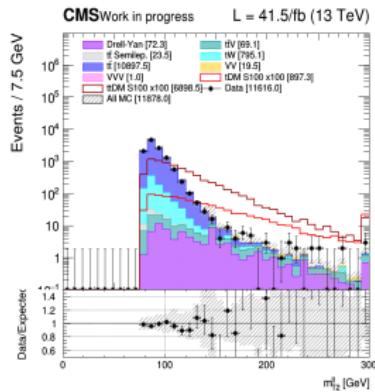
# Discriminating variables II

2016, pre-selection region



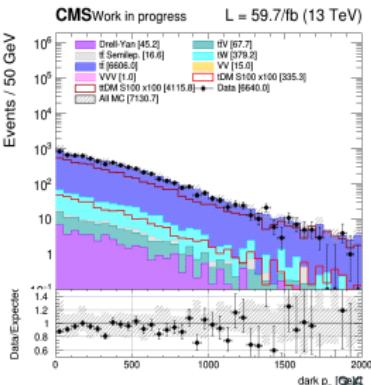
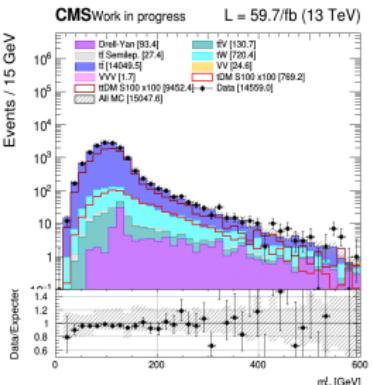
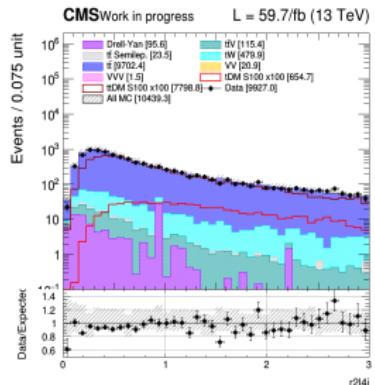
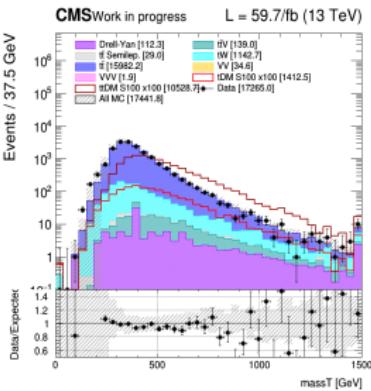
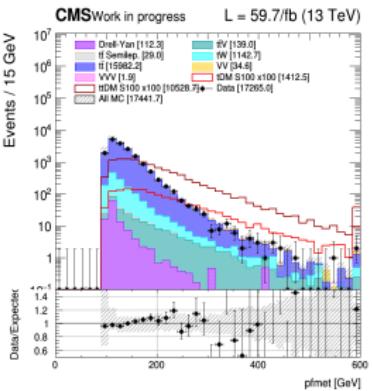
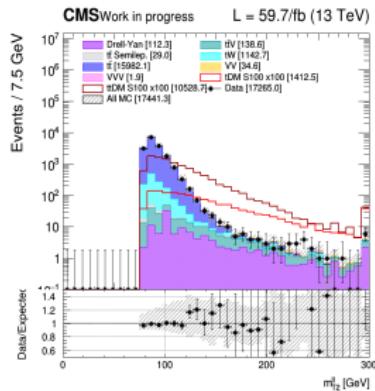
# Discriminating variables III

2017, pre-selection region



# Discriminating variables IV

2018, pre-selection region



We trained both a BDT and an ANN, featuring the following common characteristics:

- Mix of standard model  $t\bar{t}$  and single top as backgrounds, and mix of both  $t/\bar{t}+DM$  and  $t\bar{t}+DM$  as signals;
- Only events passing the pre-selection cuts are considered for the training;
- One specific training performed per signal mass point, and per signal region:
  - One targeting the  $t/\bar{t}+DM$  signal, by considering events having exactly 1 jet, or exactly 2 jets and 1 b-jet;
  - Another one targeting the  $t\bar{t}+DM$  signal, by considering events having exactly 2 jets and more than 1 b-jet, or more than 2 jets.
- 70%/30% train/test splitting used ( $\sim 50.000$  training events in total);
- 14 different discriminating variables used as input.

The BDT was chosen for the analysis over the ANN, given that it gave  $\sim 10\%$  better upper limits once optimized.

The BDT output shape is then used to perform a general shape analysis.

## Hyperparameters optimization

The hyperparameters of the BDT have all been fully optimized one by one, trying to:

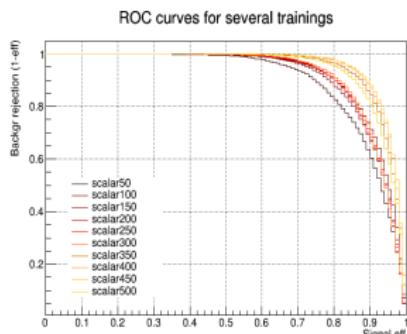
- Minimize the error in the test dataset;
- Maximize the discrimination obtained;
- All this while trying to keep the training time reasonable enough.

BDT parameter	Optimized value
Maximum depth	4
Minimum samples per leaf	2%
Loss function	Quadratic
Boost algorithm	Gradient descent
Shrinkage	0.3
Grid points $n_{\text{cut}}$	1000
Number of trees	250

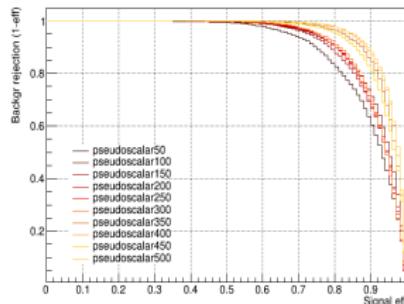
## ROC curves

ROC curves have been obtained for all the different mass points available (50 to 500 GeV), for both scalar (left) and pseudoscalar (right) mediators, in both signal regions.

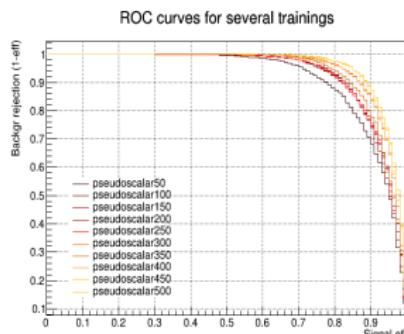
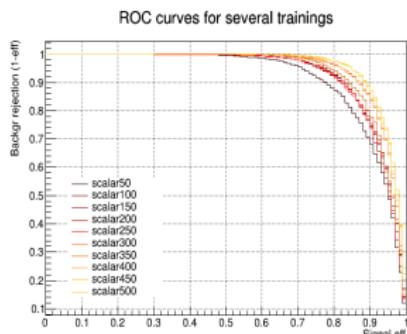
### $t/\bar{t}+DM$ region



### ROC curves for several trainings



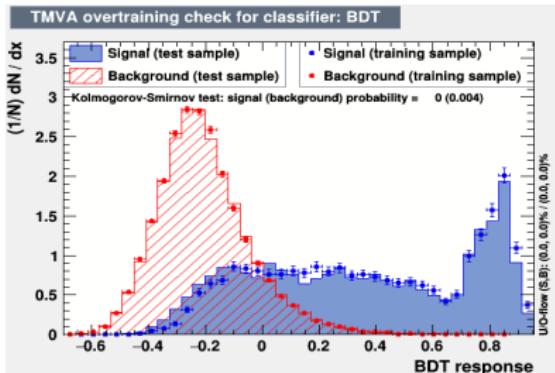
### $t\bar{t}+DM$ region



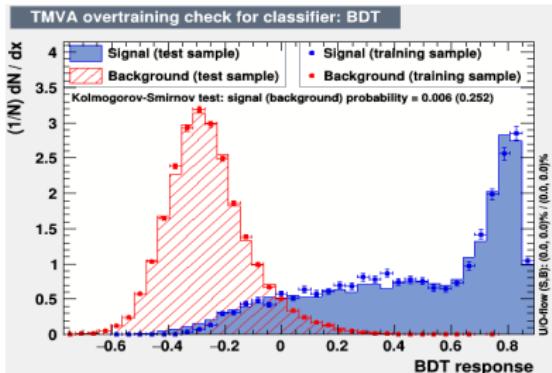
# Overtraining check ( $t/\bar{t}$ +DM region)

## Scalar mediators

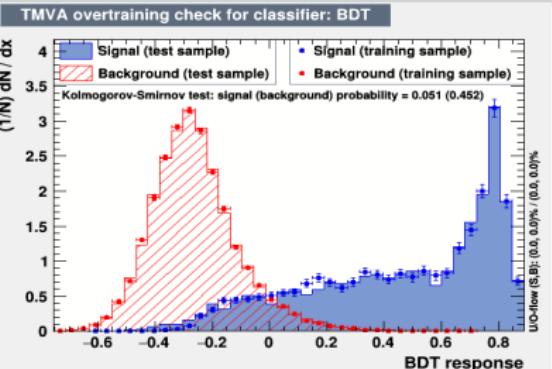
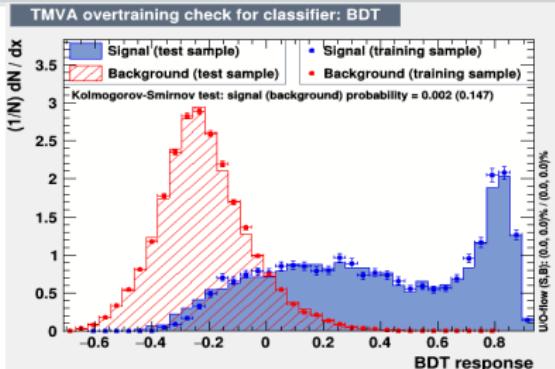
100 GeV



500 GeV



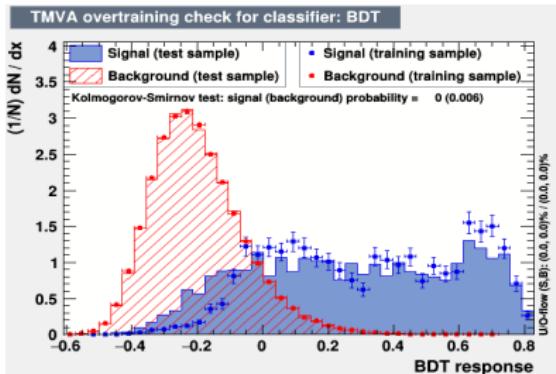
## Pseudoscalar mediators



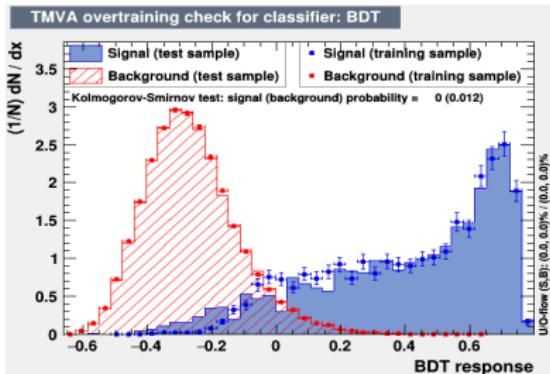
# Overtraining check ( $t\bar{t}$ +DM region)

## Scalar mediators

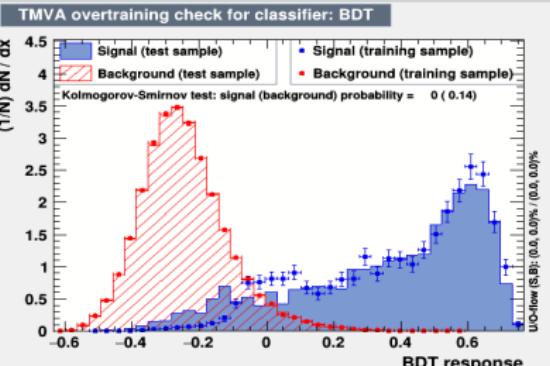
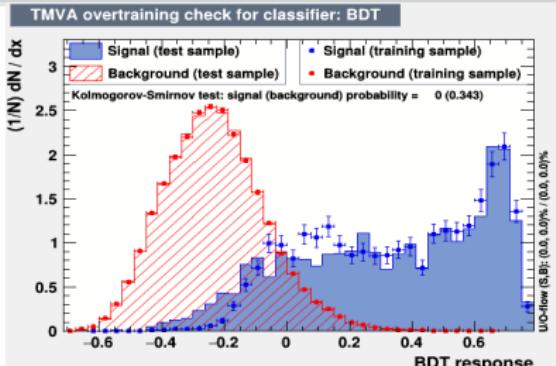
100 GeV



500 GeV



## Pseudoscalar mediators



## Signal regions

## Upper limits extraction

As explained before, in this analysis, **two different signal regions** are being used, targeting each one of our signals of interest and each based on the pre-selection region.

The BDT output shapes for each mass point were obtained in each signal region, and a **general shape analysis** was then performed using such distributions in order to **extract upper limits on the signal strength** for each model<sup>2</sup>.

Such individual results were **later combined** in order to obtain the final exclusion limits of this analysis, by either:

- On one hand by considering both signal samples together in each signal region separately, and then combining them;
- Or on the other hand by considering each signal of interest separately in both signal regions, and then combining them.

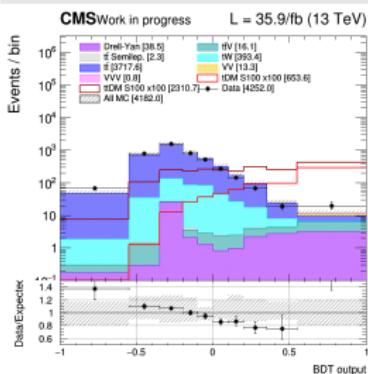
---

<sup>2</sup>The binning of the BDT was chosen as the one optimizing the upper limits obtained while ensuring to have at least a few events ( $> 5$ ) per bin

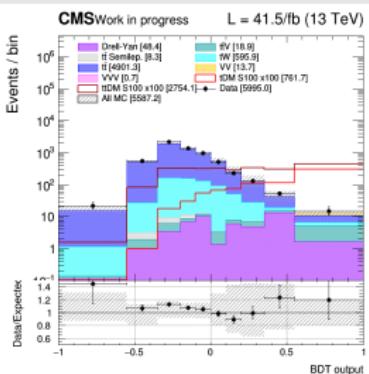
# BDT scalar 100 GeV output pre-fit shapes

## $t/\bar{t}$ +DM signal region

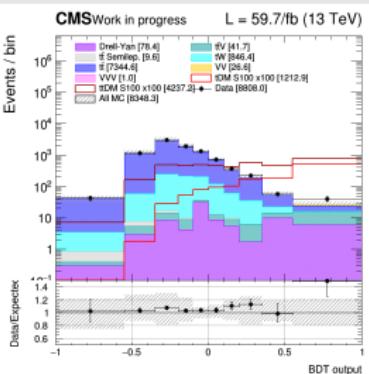
2016



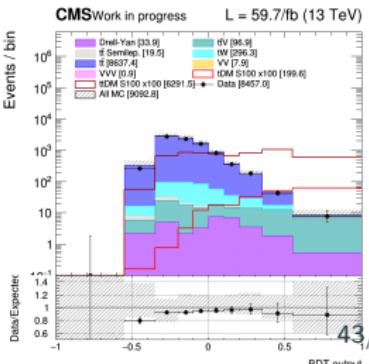
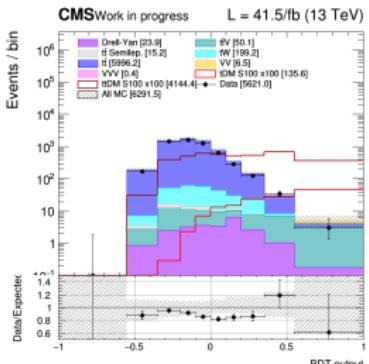
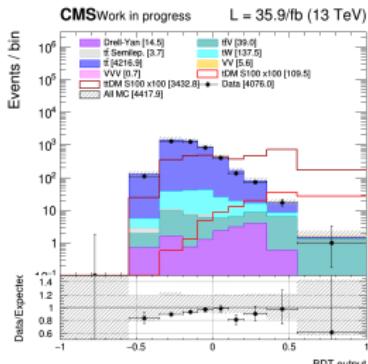
2017



2018



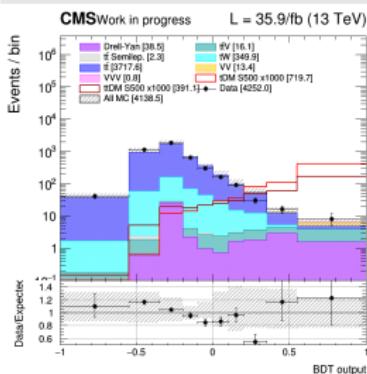
## $t\bar{t}$ +DM signal region



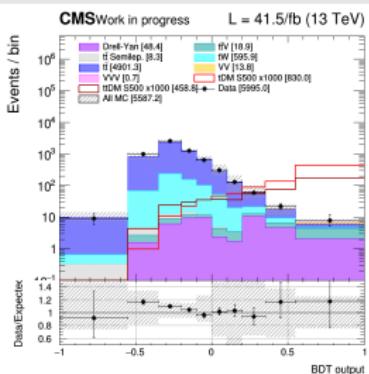
# BDT scalar 500 GeV output pre-fit shapes

## $t/\bar{t}$ +DM signal region

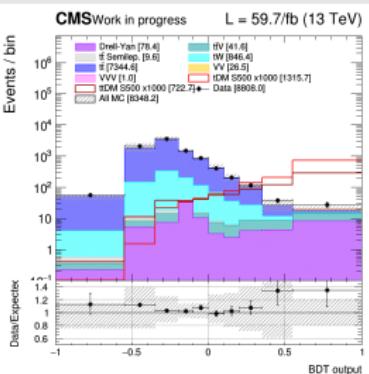
2016



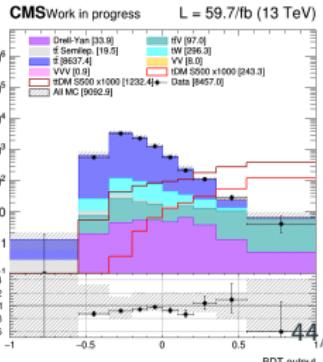
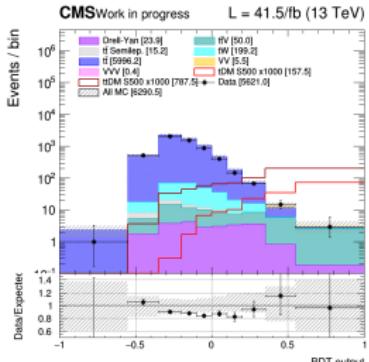
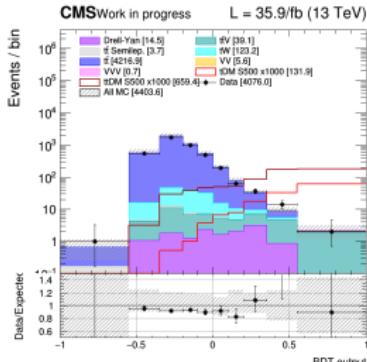
2017



2018



## $t\bar{t}$ +DM signal region



## **Systematic uncertainties**

## Systematic uncertainties

On top of statistical uncertainties, many systematics uncertainties have been considered:

### Theoretical uncertainties

- PDF and higher order corrections ( $\sim 4\%$ ), underlying event (1.5%) and parton shower modeling ( $\sim 4\%$ ), renormalization and factorization scales.

### Experimental uncertainties

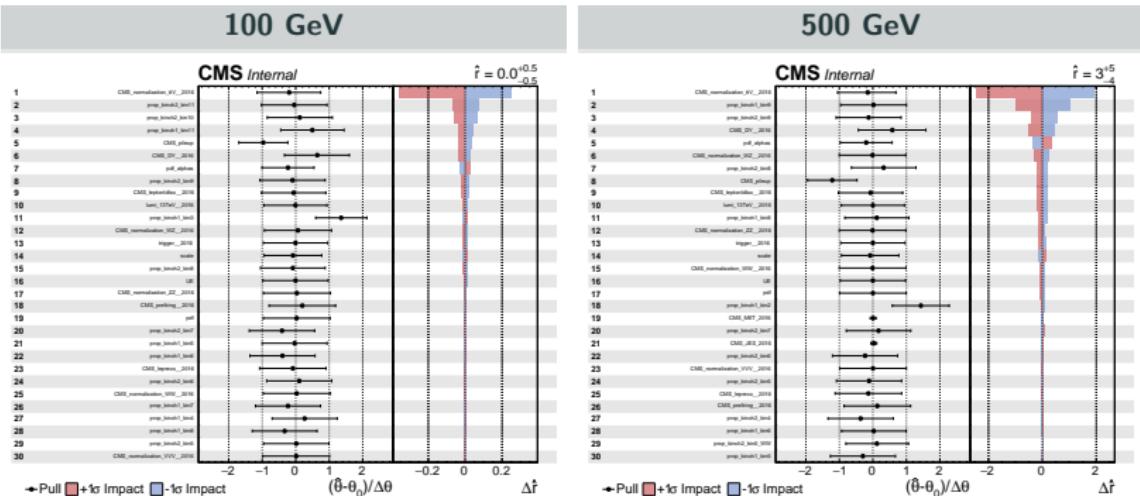
- Luminosity ( $\sim 2.5\%$ ), pileup modeling ( $\sim 5\%$ ), lepton trigger ( $\sim 2\%$ ), lepton efficiency and energy scale ( $\sim 2\%$ ), jet energy scale ( $\sim 3\%$ ), ptmiss mismodelling ( $\sim 3\%$ ), b-tagging efficiency, top  $p_T$  reweighting, ECAL prefiring.

### Background specific uncertainties

- MC statistical uncertainties;
- 20% systematic uncertainty associated to the DY process in order to cover for the non-flatness of the  $R_{\text{in-out}}$  transfer factor;
- 30% uncertainty associated to the normalization of all the minor backgrounds, except for the ttV, for which a 50% systematic uncertainty is associated.

# Pulls and impact plots

## 2016 scalar mediator



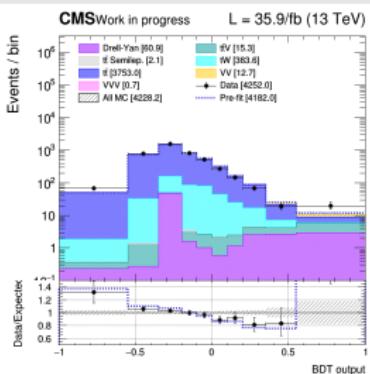
The most important systematic for most mass points is the normalization of the  $t\bar{t}V$  process → Additional impact plots can be found in the backup.

## **Results obtained**

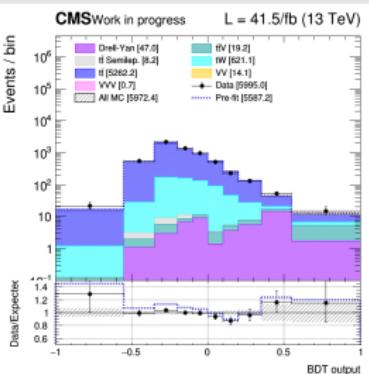
# Post-fit plots (scalar 100 GeV)

## $t/\bar{t} + \text{DM}$ region

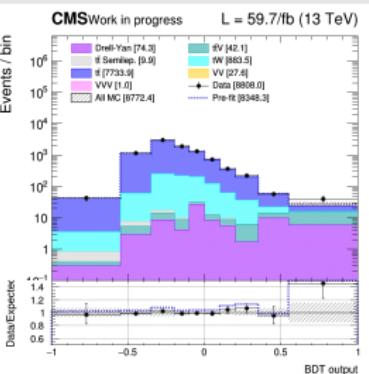
2016



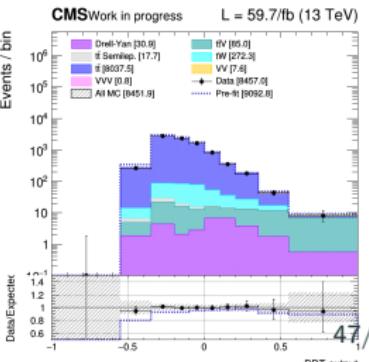
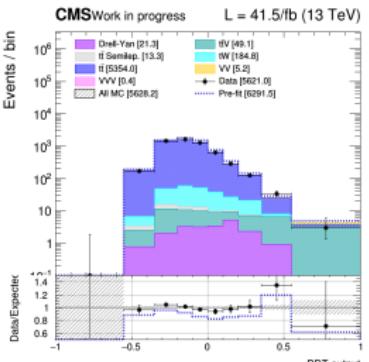
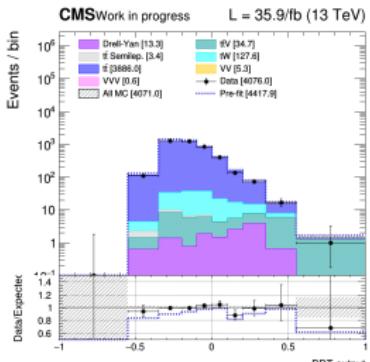
2017



2018



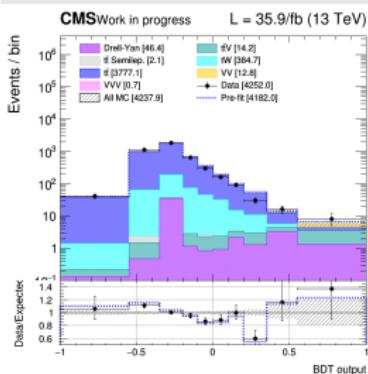
## $t\bar{t} + \text{DM}$ region



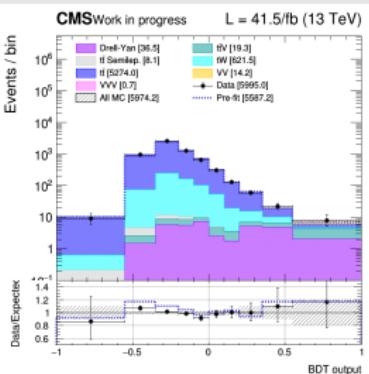
# Post-fit plots (scalar 500 GeV)

## $t/\bar{t} + \text{DM}$ region

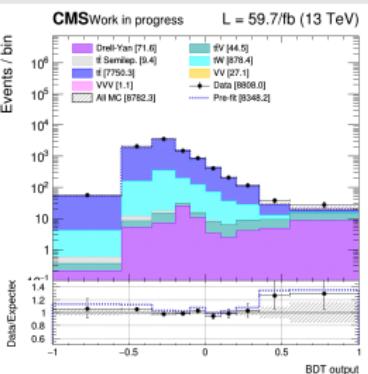
2016



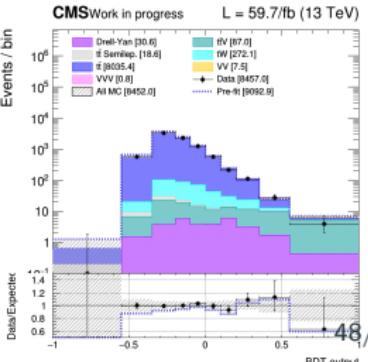
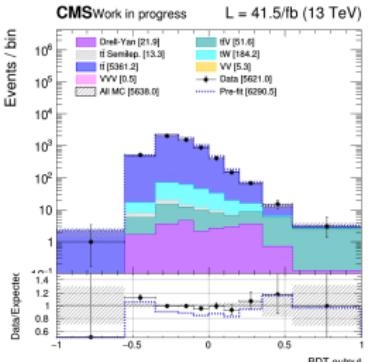
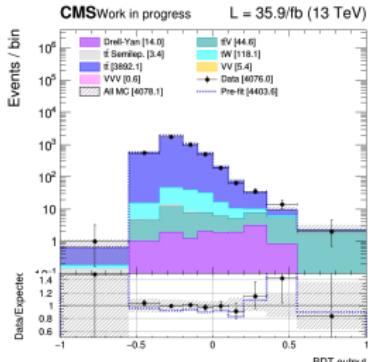
2017



2018



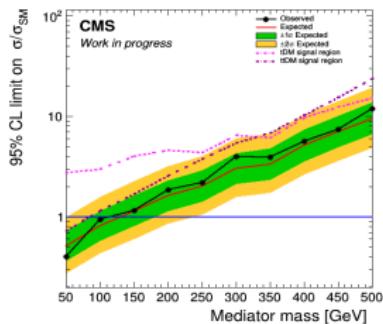
## $t\bar{t} + \text{DM}$ region



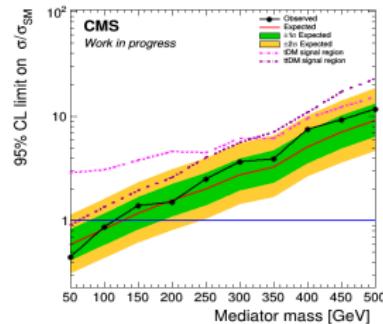
## Upper limits on the signal regions

### Scalar upper limits

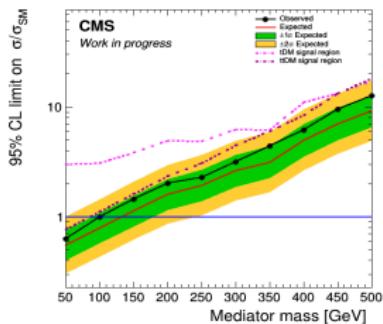
2016



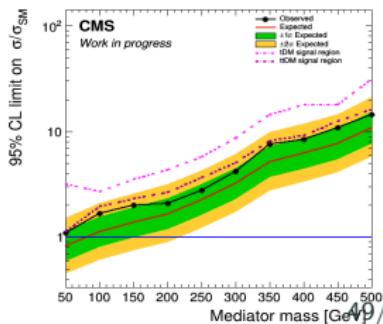
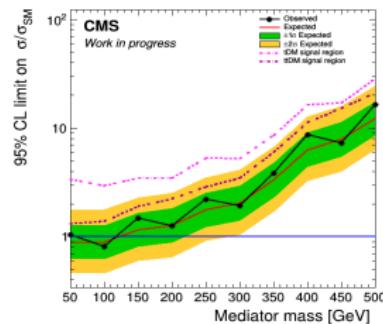
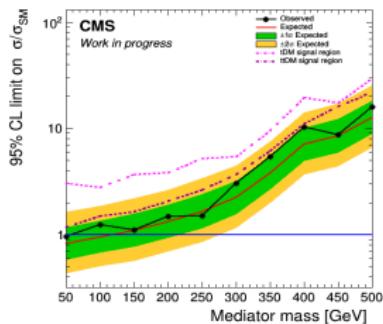
2017



2018



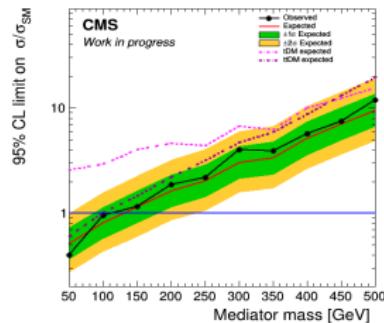
### Pseudoscalar upper limits



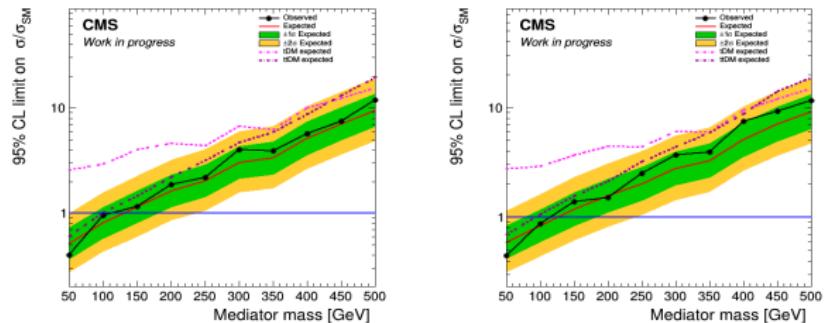
# Upper limits on the signal models

## Scalar upper limits

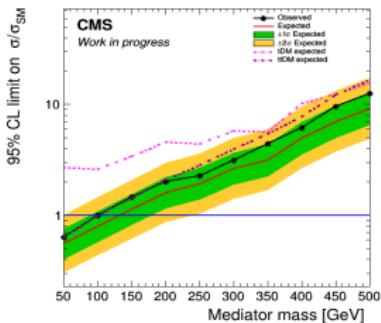
2016



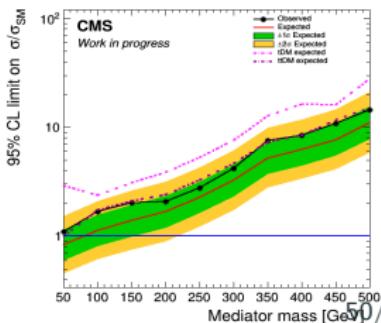
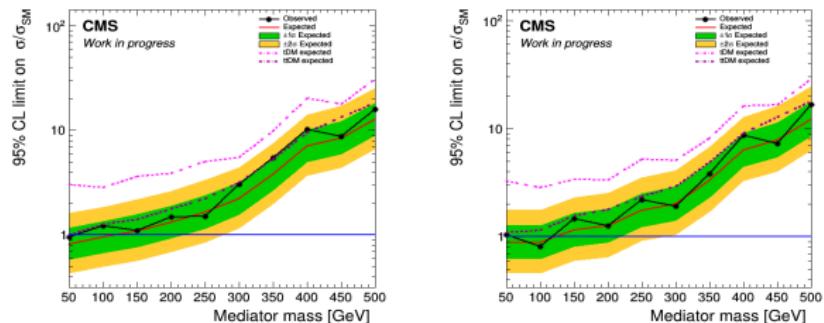
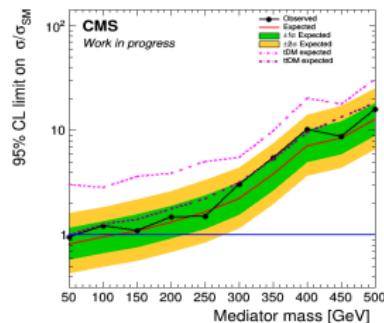
2017



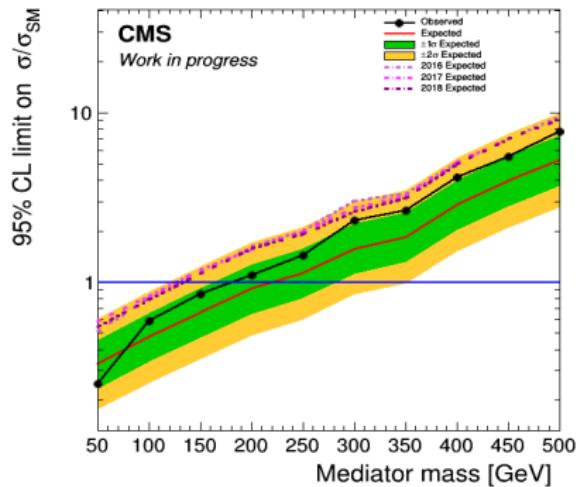
2018



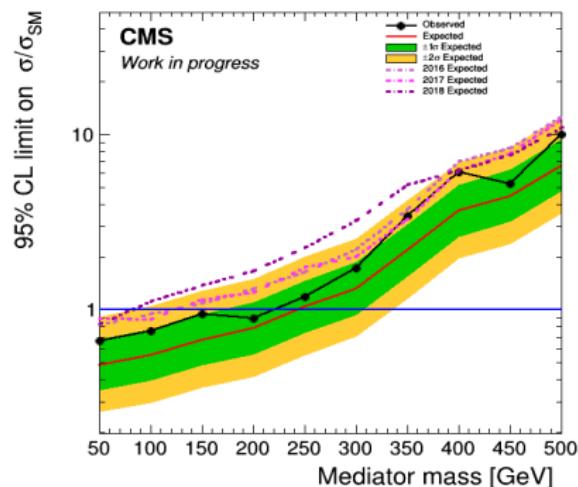
## Pseudoscalar upper limits



### Scalar mediators



### Pseudoscalar mediators



After combining the limits obtained for the different years, the following **expected (observed) exclusion have been achieved**:

- Scalar mediators excluded to 215 (180) GeV;
- Pseudoscalar mediators excluded up to 250 (220) GeV.

## Conclusions

A search for **dark matter produced in association with either one or two top quarks** has been performed, considering in particular its **dilepton final state**, and analyzing the **Run II legacy dataset** collected by the CMS detector at 13 TeV.

This is the **first time that such a combination of two signals of interest** is performed considering this particular final state.

In summary, this search managed to:

- Exclude scalar (pseudoscalar) mediators up to 215 (250) GeV respectively;
- **Improve by a factor of 3** the scalar exclusion limits published by the CMS collaboration in 2016 using the  $t\bar{t}$ +DM model alone;
- And **exclude for the first time pseudoscalar mediators** up to 250 GeV.

Several additions and improvements could be considered to improve the results obtained:

- The combination of these results with the semileptonic and hadronic final states will be beneficial given their large branching ratios;
- The continuous operation of the LHC and future data yet to be collected is expected to improve these results as well;
- Given the importance of the ttV normalization systematic uncertainty, the definition of a dedicated control region might be beneficial as well;
- Exploring other regions (different  $m_\chi$  or couplings values, for example) can be interesting, as well as the reinterpretation of these results in terms of a Higgs to invisible model.

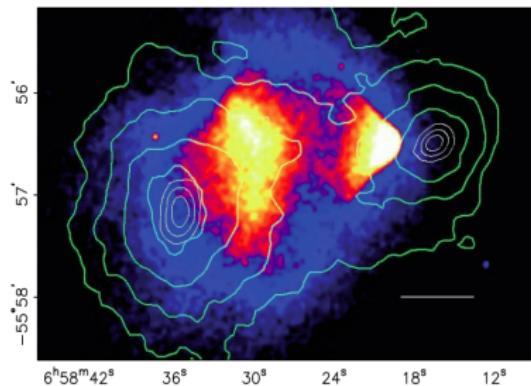
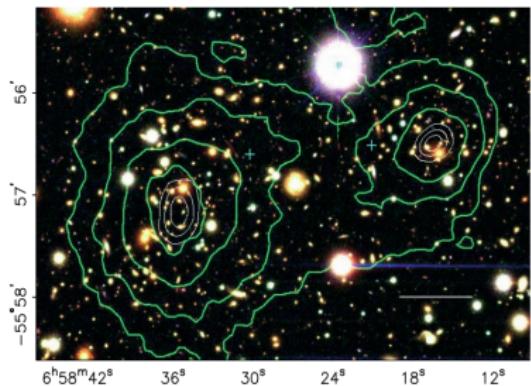
This analysis is in any case expected to gain momentum and provide us with even better exclusion limits over the course of the next years of operation of the LHC.

**Back up**

## Gravitational lensing

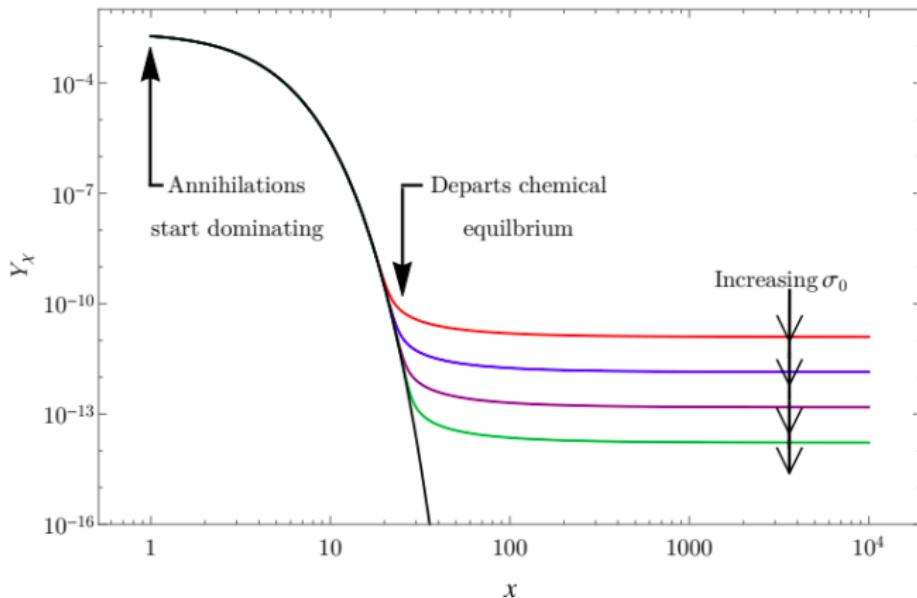
Consequence of the **general relativity**: massive objects placed between distant sources and the observer should be able to **act as lenses and bend the light of the source**.

- The deviation of the light is proportional to the mass of the intermediate object, giving us a way to measure its mass;
- The mass distribution obtained has been compared to the luminous distribution of several galaxies, leading to  $8\sigma$  discrepancies.



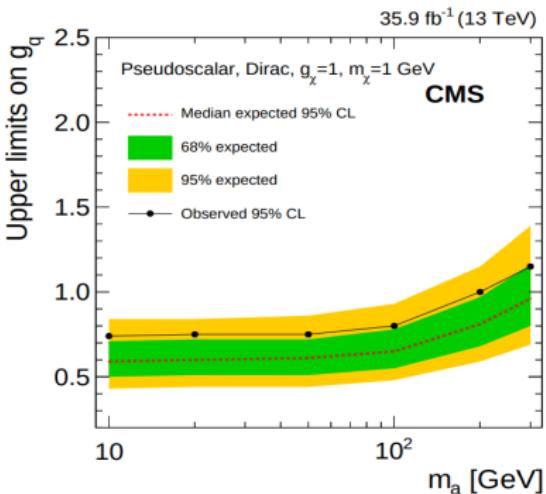
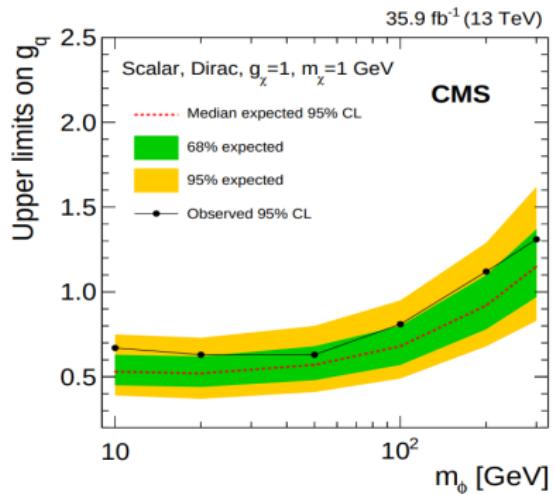
## Thermal freeze-out

Schematic representation of the **freeze-out process**, representing the abundance of a 500 GeV dark matter with respect to time and the impact increasing cross-section annihilation values have on the remnant freeze-out abundance.



## Additional relevant results I

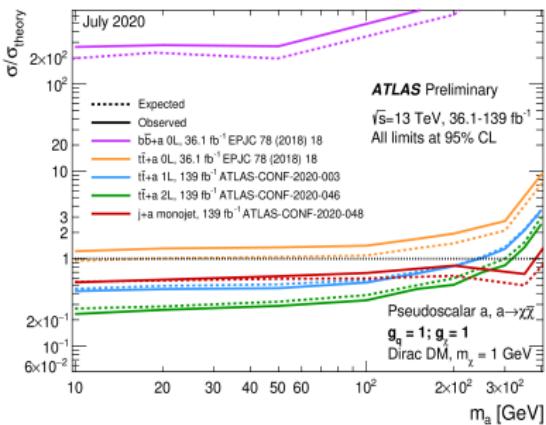
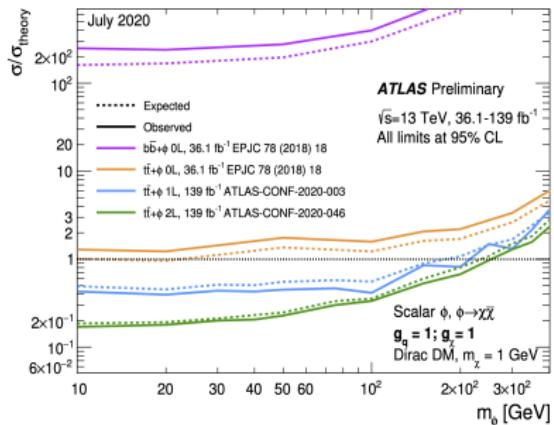
CMS combination of all the different final states published in 2016:



The observed (expected) limits excluded a **pseudoscalar mediator** with mass below 220 (320) GeV, and a **scalar mediator** with mass below 160 (240) GeV.

## Additional relevant results II

The ATLAS collaboration obtained **exclusion limits for all the channels**, considering the full Run II dataset.



Note that they do not perform any combination between the different channels and typically use NLO cross-sections for the signals.

## Center of mass energy

The **center of mass energy** is defined as a Lorentz invariant quantity under any kind of boost resulting of the collisions between two protons (defined as  $E_1, \vec{p}_1, m_1$  and  $E_2, \vec{p}_2, m_2$ ) with a  $\theta$  angle.

$$\sqrt{s} = \sqrt{(m_1)^2 + (m_2)^2 + 2(E_1 E_2 - 2|\vec{p}_1| |\vec{p}_2| \cos(\theta))}$$

The LHC started its operation in 2008 running at an energy of 7 TeV, quickly moved to 8 TeV and kept this level of energy during the end of the Run I of operation. In 2015, the energy was **increased to 13 TeV**.

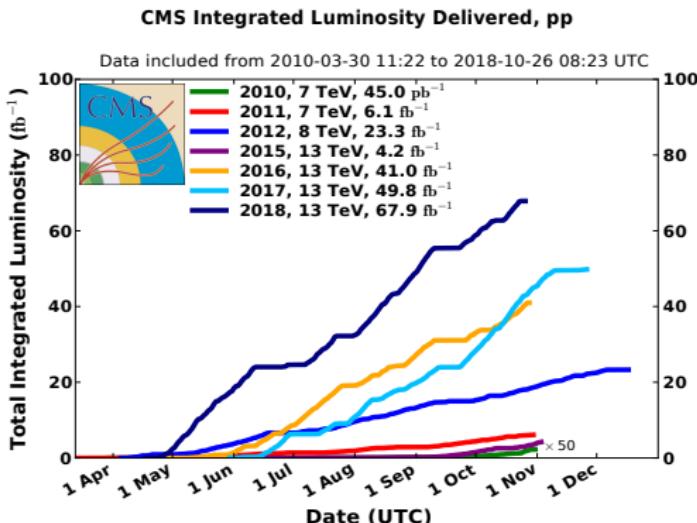
An expected value of 13.5 to 14 TeV, the nominal energy for which the LHC was originally built, is expected to be reached in the near future.

## Luminosity

The **luminosity  $\mathcal{L}$**  gives an indication on the **number of collisions per second** given by the accelerator. Increasing it is crucial to collect as much data as possible, to be able to isolate processes having a low production cross section.

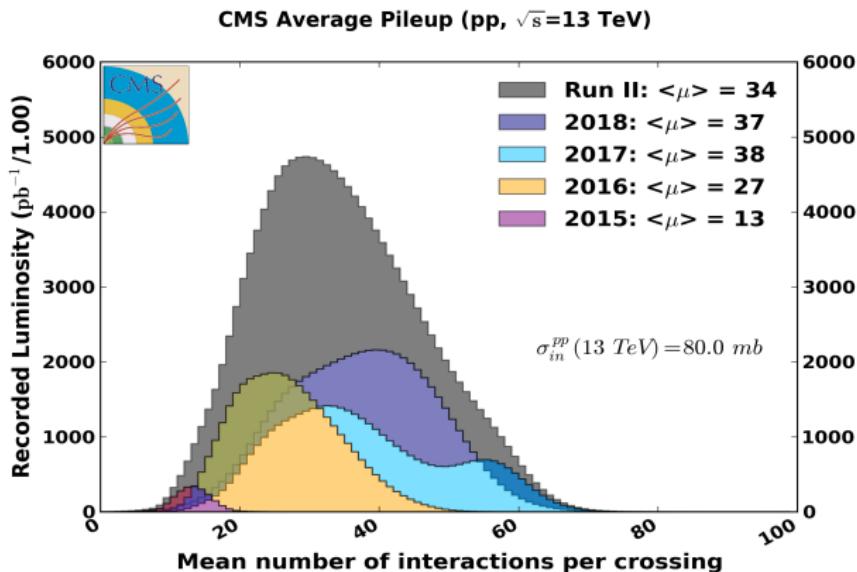
The **rate of production  $R$**  of any given process can be expressed from the instantaneous luminosity  $\mathcal{L}(t)$  and the process production cross-section  $\sigma$ :

$$\begin{cases} R = \mathcal{L} \cdot \sigma \\ N(T) = \sigma \int_0^T \mathcal{L}(t) dt = \sigma L \end{cases}$$



## Pile-up

Because of the high density of protons within the beams, a **bunch crossing** in an experiment produces around 30-35 proton collisions.



The **primary vertex** is defined as the most interesting and energetic vertex, while the other vertices are usually referred to as the **pile-up**.

## LHC operational parameters

Key parameters of operation of the LHC, depending on the data-taking period:

Parameter	Run I	Run II	Run III	Design
Energy [TeV]	7 → 8	13	13	14
Bunch spacing [ns]	50	25	25	25
Intensity [ $10^{11}$ protons per beam]	1.6	1.2	Up to 1.8	1.15
Bunches	1400	2500	2800	2800
Emittance [ $\mu m$ ]	2.2	2.2	2.5	3.5
$\beta^*$ [cm]	80	30 → 25	30 → 25	55
Crossing angle [ $\mu rad$ ]	-	300 → 260	300 → 260	285
Peak luminosity [ $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ]	0.8	2.0	2.0	1.0
Peak pile-up	45	60	55	25

The tracker is the **innermost piece of CMS**, able to reconstruct the trajectories of charged particles issued from the interaction vertices in a quick and precise way:

- Needs to be extremely fast to read the 40 MHz of collision data, while being resistant to the radiation (expected lifetime  $\sim 10$  years);
- It should be as small as possible in order to minimize the interactions between the detector and the particles created;
- However, fast electronics usually needs to be cooled down, which automatically increases the size of this subdetector.

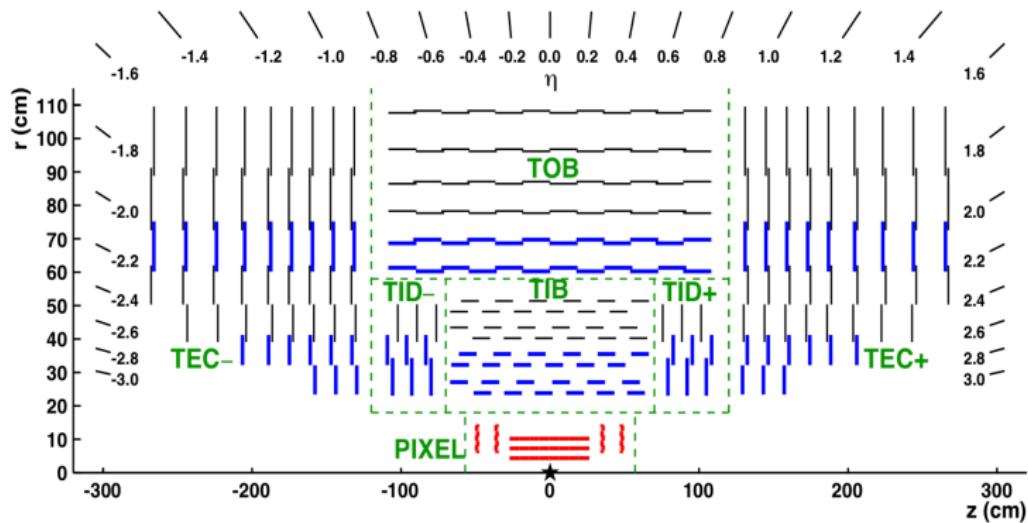
Made out of two main parts:

- The **pixel detector**, made out 60 millions pixels which make up the 1856 modules of this detector, covering an area of  $\sim 1 \text{ m}^2$ ;
- The **silicon strip detector**, covering an area of  $\sim 200 \text{ m}^2$ , and made out of three different sub-systems for hermeticity.

A charged particle crossing the tracker will leave a hit each time it crosses one of the silicon sensors, allowing us to reconstruct its track.

The **silicon detector** is divided into three main parts:

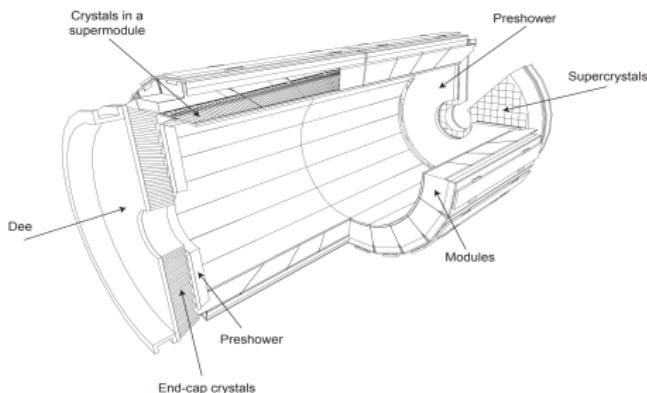
- The **Tracker Inner Barrel and Disks (TIB/TBD)**, using micro-strips parallel in the barrel and perpendicular to the beam axis in the endcaps;
- The **Tracker Outer Barrel (TOB)**, adding 6 measurement layers to the tracker;
- And finally the **Tracker EndCaps (TECs)**, made out of 9 disks, completing the system at high pseudorapidities.



The **Electromagnetic Calorimeter** is a subdetector sitting inside the solenoid but enclosing the tracker system that gives information about the **energy of electrons and photons**, both able to interact electromagnetically with its crystals.

Made out of different layers:

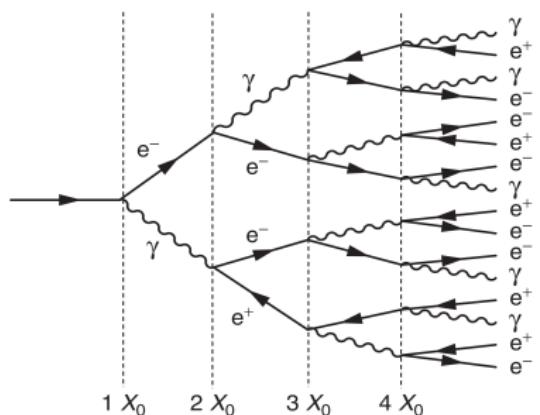
- The **barrel part (EB)**, at  $|\eta| < 1.479$ , made out of 61 200 lead tungstate ( $\text{PbWO}_4$ ) crystals;
- Two **endcaps**, each made out of 7 324 crystals, increasing the coverage of the detector up to  $|\eta| < 3$ ;
- The **preshower**, helping with the identification of electrons against minimum ionizing particles.



The principle of action of the ECAL is simple, and is based on **electromagnetic showers**. When an electron or a photon enters the ECAL, it starts to interact in different ways:

- Photons will mainly produce pairs of electrons and anti-electrons;
- Electrons themselves tend to emit additional photons by bremsstrahlung effect.

This results in a **chain reaction** during which the incident particle gives most of its energy to the detector, energy measurable using photodetectors and photomultipliers.



Although quite fragile and sensitive to the temperature, the short radiation length  $X_0$  of the  $\text{PbWO}_4$  crystals is an advantage, along with their scintillation decay time smaller than the bunch crossing.

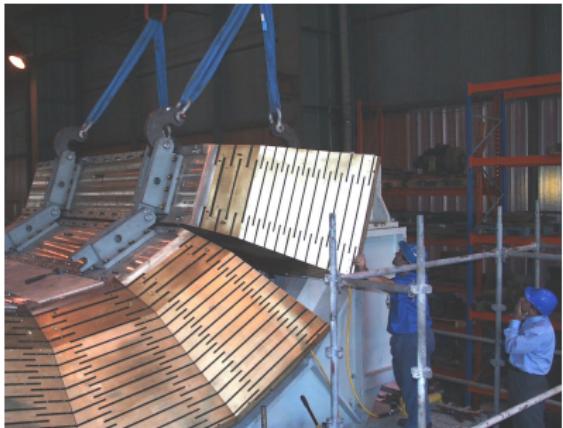
Each crystal measures  $2.2 \times 2.2 \times 23 \text{ cm}$ , corresponding to 26 radiation lengths.

Charged hadrons **lose energy when they traverse matter** due to the **ionization process resulting from the strong interaction** between them and the nuclei of the detector.

Showers of particles are typically produced since the primary hadronic interaction will produce several additional hadrons, themselves interacting even more with the detector.

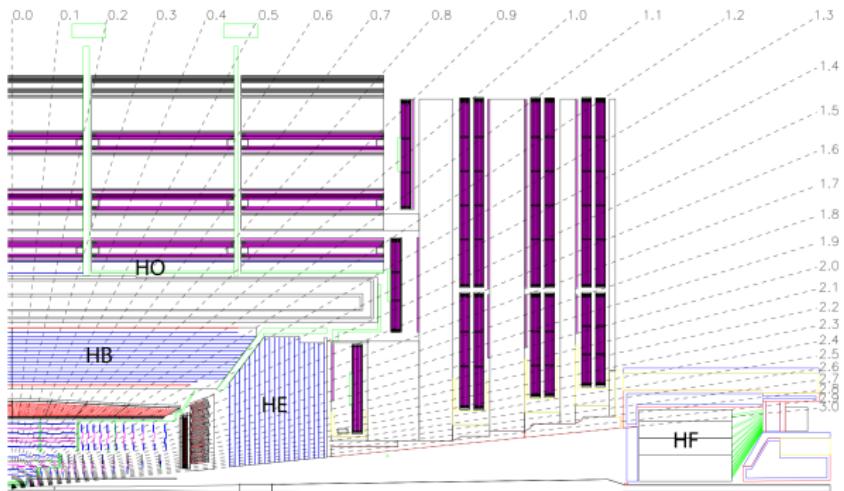
The HCAL is made out of alternating layers:

- Of thick **absorber material**, in which the showers can develop;
- And thin layers of **active material** used for the actual detection by sampling the energy deposition.



The HCAL is divided into:

- A **barrel (HB)**, up to  $|\eta| = 1.3$ ;
- Two **endcaps**, extending the pseudorapidities coverage up to  $|\eta| = 3.0$ ;
- Two **symmetrical forward regions (HF)**, covering up to  $|\eta| = 5.2$ ;
- And the **Hadron Outer (HO)**, outside of the solenoid, placed to increase the effective nuclear radiation length  $\lambda$ , otherwise low at a  $90^\circ$  incidence angle.



The **superconducting solenoid**:

- Is made out of 6 endcap disks and 5 barrel wheels;
- Weights more than 12 000 tons in total, with the return yoke;
- Is able to produce a 3.8 T magnetic field once cooled down to 4.5K;
- Stores around 2.6 GJ of energy when active.



It allows the **measurement of the momentum and charge of particles** by studying the curvature of their tracks, according to the Lorentz equation:

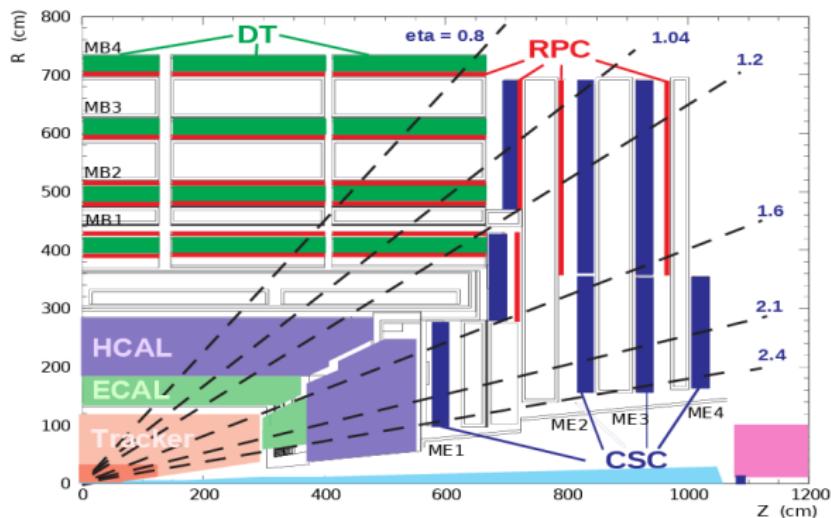
$$\vec{F} = \frac{m\vec{v}^2}{R} = q\vec{E} + q\vec{v} \times \vec{B} = q\vec{v} \times \vec{B}$$

It has been designed to reach a momentum resolution  $\Delta p/p \sim 10\%$  at  $p = 1 \text{ TeV}$ .

## CMS muon system I

The **muon system** is the **outermost section** of CMS, covering around 25 000 m<sup>2</sup>.

**Three different categories** of devices have been designed, in order to cope with the specific experimental conditions in the different parts of the detector: the Drift Tubes (DTs), the Cathode Strips Chambers (CSCs), and the Resistive Plate Chambers (RPCs).



All these detectors are gaseous, distributed over a cylindrical area given the shape of the innermost components of CMS, and cheap, given the large surface they cover.

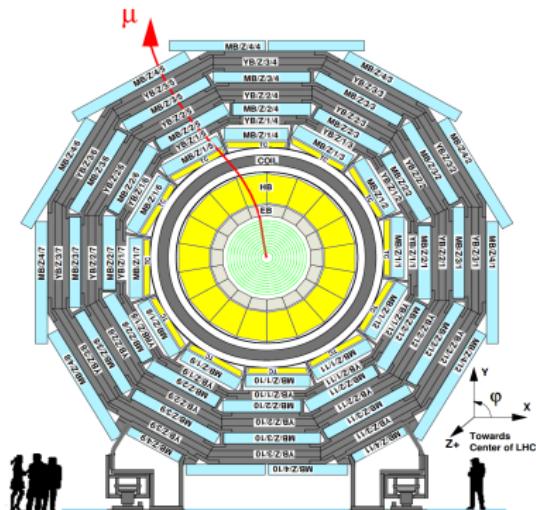
Comparison of the different subsystems:

Muon sub-system	DTs	CSCs	RPCs
$ \eta $ coverage Stations	0.0-1.2 4	0.9-2.4 4	0.0-1.9 4
Chambers	250	540	480 (barrel) 576 (endcaps)
Readout channels	172 000	266 112 (strips) 210 816 (anode channels)	68 136 (barrel) 55 296 (endcaps)
Spatial resolution	80-120 $\mu\text{m}$	40-150 $\mu\text{m}$	0.8-1.2 cm
Average efficiency (13 TeV)	97.1%	97.4%	94.2% (barrel) 96.4% (endcaps)

Placed in the barrel region (up to  $|\eta| = 1.2$ ), where the background levels and magnetic field are low, this system allows a **good efficiency for the muon hits reconstruction** into a single track and a **good rejection of eventual background hits**.

This system is:

- Able to collect the residuals charges left by the ionization tracks of muons;
- Made out of 172 000 sensitive wires, divided in 250 chambers;
- Redundant, by the installation of 4 layers, to reduce the impact coming from eventual neutrons or photons;
- Has a maximal drift time of 380 ns, low enough to avoid the need of multi-hits electronics.



# 2016 data samples

Dataset	Events (size)	$\mathcal{L}$ [fb $^{-1}$ ]
<b>Run 2016B</b>		
/DoubleEG/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOOD	143073268 (99.4Gb)	
/DoubleMuon/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOOD	82535526 (53.2Gb)	
/MuonEG/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOOD	32727796 (26.8Gb)	5.8
/SingleElectron/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOOD	246440440 (167.8Gb)	
/SingleMuon/Run2016B_ver2-Nano02Apr2020_ver2-v1/NANOAOOD	158145722 (96.4Gb)	
<b>Run 2016C</b>		
/DoubleEG/Run2016C-Nano02Apr2020-v1/NANOAOOD	47677856 (35.3Gb)	
/DoubleMuon/Run2016C-Nano02Apr2020-v1/NANOAOOD	27934629 (19.7Gb)	
/MuonEG/Run2016C-Nano02Apr2020-v1/NANOAOOD	15405678 (12.8Gb)	2.6
/SingleElectron/Run2016C-Nano02Apr2020-v1/NANOAOOD	97259854 (69.3Gb)	
/SingleMuon/Run2016C-Nano02Apr2020-v1/NANOAOOD	67441308 (42.4Gb)	
<b>Run 2016D</b>		
/DoubleEG/Run2016D-Nano02Apr2020-v1/NANOAOOD	53324960 (39.6Gb)	
/DoubleMuon/Run2016D-Nano02Apr2020-v1/NANOAOOD	33861745 (24.1Gb)	
/MuonEG/Run2016D-Nano02Apr2020-v1/NANOAOOD	23482352 (19.4Gb)	4.2
/SingleElectron/Run2016D-Nano02Apr2020-v1/NANOAOOD	148167727 (104.4Gb)	
/SingleMuon/Run2016D-Nano02Apr2020-v1/NANOAOOD	98017996 (61.3Gb)	
<b>Run 2016E</b>		
/DoubleEG/Run2016E-Nano02Apr2020-v1/NANOAOOD	49877710 (37.9Gb)	
/DoubleMuon/Run2016E-Nano02Apr2020-v1/NANOAOOD	28246946 (20.8Gb)	
/MuonEG/Run2016E-Nano02Apr2020-v2/NANOAOOD	22519303 (19.0Gb)	4.0
/SingleElectron/Run2016E-Nano02Apr2020-v1/NANOAOOD	117321545 (86.5Gb)	
/SingleMuon/Run2016E-Nano02Apr2020-v1/NANOAOOD	90984718 (58.7Gb)	
<b>Run 2016F</b>		
/DoubleEG/Run2016F-Nano02Apr2020-v1/NANOAOOD	34577629 (26.9Gb)	
/DoubleMuon/Run2016F-Nano02Apr2020-v1/NANOAOOD	20329921 (15.3Gb)	
/MuonEG/Run2016F-Nano02Apr2020-v1/NANOAOOD	16002165 (13.6Gb)	3.1
/SingleElectron/Run2016F-Nano02Apr2020-v1/NANOAOOD	70593532 (51.4Gb)	
/SingleMuon/Run2016F-Nano02Apr2020-v1/NANOAOOD	65489554 (42.4Gb)	
<b>Run 2016G</b>		
/DoubleEG/Run2016G-Nano02Apr2020-v1/NANOAOOD	78797031 (61.6Gb)	
/DoubleMuon/Run2016G-Nano02Apr2020-v1/NANOAOOD	45235604 (34.2Gb)	
/MuonEG/Run2016G-Nano02Apr2020-v1/NANOAOOD	33854612 (29.0Gb)	7.6
/SingleElectron/Run2016G-Nano02Apr2020-v1/NANOAOOD	153363109 (109.2Gb)	
/SingleMuon/Run2016G-Nano02Apr2020-v1/NANOAOOD	149912248 (94.6Gb)	
<b>Run 2016H</b>		
/DoubleEG/Run2016H-Nano02Apr2020-v1/NANOAOOD	85388734 (67.7Gb)	
/DoubleMuon/Run2016H-Nano02Apr2020-v1/NANOAOOD	48912812 (37.3Gb)	
/MuonEG/Run2016H-Nano02Apr2020-v1/NANOAOOD	29236516 (26.0Gb)	8.6
/SingleElectron/Run2016H-Nano02Apr2020-v1/NANOAOOD	128854598 (93.8Gb)	
/SingleMuon/Run2016H-Nano02Apr2020-v1/NANOAOOD	174035164 (110.2Gb)	

## 2017 data samples

Dataset	Events (size)	$\mathcal{L} [\text{fb}^{-1}]$
<b>Run 2017B</b>		
/DoubleEG/Run2017B-Nano02Apr2020-v1/NANOAOD	58088760 (46.6Gb)	
/DoubleMuon/Run2017B-Nano02Apr2020-v1/NANOAOD	14501767 (10.8Gb)	
/SingleElectron/Run2017B-Nano02Apr2020-v1/NANOAOD	60537490 (42.2Gb)	
/SingleMuon/Run2017B-Nano02Apr2020-v1/NANOAOD	136300266 (86.2Gb)	
/MuonEG/Run2017B-Nano02Apr2020-v1/NANOAOD	4453465 (4.1Gb)	
<b>Run 2017C</b>		
/DoubleEG/Run2017C-Nano02Apr2020-v1/NANOAOD	65181125 (53.8Gb)	
/DoubleMuon/Run2017C-Nano02Apr2020-v1/NANOAOD	49636525 (39.5Gb)	
/SingleElectron/Run2017C-Nano02Apr2020-v1/NANOAOD	136637888 (102.5Gb)	
/SingleMuon/Run2017C-Nano02Apr2020-v1/NANOAOD	165652756 (109.5Gb)	
/MuonEG/Run2017C-Nano02Apr2020-v1/NANOAOD	15595214 (15.0Gb)	
<b>Run 2017D</b>		
/DoubleEG/Run2017D-Nano02Apr2020-v1/NANOAOD	25911432 (21.6Gb)	
/DoubleMuon/Run2017D-Nano02Apr2020-v1/NANOAOD	23075733 (18.6Gb)	
/SingleElectron/Run2017D-Nano02Apr2020-v1/NANOAOD	51526710 (38.5Gb)	
/SingleMuon/Run2017D-Nano02Apr2020-v1/NANOAOD	70361660 (47.2Gb)	
/MuonEG/Run2017D-Nano02Apr2020-v1/NANOAOD	9164365 (8.9Gb)	
<b>Run 2017E</b>		
/DoubleEG/Run2017E-Nano02Apr2020-v1/NANOAOD	56233597 (49.8Gb)	
/DoubleMuon/Run2017E-Nano02Apr2020-v1/NANOAOD	51589091 (44.4Gb)	
/SingleElectron/Run2017E-Nano02Apr2020-v1/NANOAOD	102121689 (81.3Gb)	
/SingleMuon/Run2017E-Nano02Apr2020-v1/NANOAOD	154630534 (111.0Gb)	
/MuonEG/Run2017E-Nano02Apr2020-v1/NANOAOD	19043421 (19.2Gb)	
<b>Run 2017F</b>		
/DoubleEG/Run2017F-Nano02Apr2020-v1/NANOAOD	74307066 (67.1Gb)	
/DoubleMuon/Run2017F-Nano02Apr2020-v1/NANOAOD	79756560 (68.0Gb)	
/SingleElectron/Run2017F-Nano02Apr2020-v1/NANOAOD	128467223 (105.2Gb)	
/SingleMuon/Run2017F-Nano02Apr2020-v1/NANOAOD	242135500 (178.3Gb)	
/MuonEG/Run2017F-Nano02Apr2020-v1/NANOAOD	25776363 (26.3Gb)	

## 2018 data samples

Dataset	Events (size)	$\mathcal{L}$ [ $\text{fb}^{-1}$ ]
<b>Run 2018A</b>		
/DoubleMuon/Run2018A-Nano02Apr2020-v1/NANO AOD	75499908 (62.6Gb)	
/EGamma/Run2018A-Nano02Apr2020-v1/NANO AOD	327843843 (261.8Gb)	
/SingleMuon/Run2018A-Nano02Apr2020-v1/NANO AOD	241608232 (167.7Gb)	13.5
/MuonEG/Run2018A-Nano02Apr2020-v1/NANO AOD	32958503 (32.3Gb)	
<b>Run 2018B</b>		
/DoubleMuon/Run2018B-Nano02Apr2020-v1/NANO AOD	35057758 (28.3Gb)	
/EGamma/Run2018B-Nano02Apr2020-v1/NANO AOD	153822427 (123.1Gb)	
/SingleMuon/Run2018B-Nano02Apr2020-v1/NANO AOD	119918017 (82.3Gb)	6.8
/MuonEG/Run2018B-Nano02Apr2020-v1/NANO AOD	16211567 (15.8Gb)	
<b>Run 2018C</b>		
/DoubleMuon/Run2018C-Nano02Apr2020-v1/NANO AOD	34565869 (27.6Gb)	
/EGamma/Run2018C-Nano02Apr2020-v1/NANO AOD	147827904 (119.2Gb)	
/SingleMuon/Run2018C-Nano02Apr2020-v1/NANO AOD	110032072 (75.7Gb)	6.6
/MuonEG/Run2018C-Nano02Apr2020-v1/NANO AOD	15652198 (15.3Gb)	
<b>Run 2018D</b>		
/DoubleMuon/Run2018D-Nano02Apr2020_ver2-v1/NANO AOD	168605834 (128.6Gb)	
/EGamma/Run2018D-Nano02Apr2020-v1/NANO AOD	751348648 (583.6Gb)	
/SingleMuon/Run2018D-Nano02Apr2020-v1/NANO AOD	513867253 (344.5Gb)	
/MuonEG/Run2018D-Nano02Apr2020_ver2-v1/NANO AOD	71961587 (68.6Gb)	32.0

# 2016 MC samples

Process	Sample	Cross section [pb]
Drell-Yan	DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18610.0
	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 ( $H_T < 70$ GeV)	6077.22
	DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	169.9
	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	147.4
	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	40.99
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.678
	DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.367
	DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.6304
	DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.1514
	DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.003565
TTTo2L2Nu	TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.360
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	80.95
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	136.02
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85
TTToSemiLeptonic	TTToSemilepton_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	364.35
ttV	TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.2814
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
VZ	WWTo2L2Nu_13TeV-powheg	12.178
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.42965
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_powheg_pythia8	3.22
Others	WWWW_WWZ_WZZ_ZZZ_WWG	//

# 2017 MC samples

Process	Sample	Cross section [pb]
Drell-Yan	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 ( $H_T < 70$ GeV)	6077.22
	DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8	169.9
	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	147.4
	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	40.99
	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.678
	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.367
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.6304
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1514
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003565
TTTo2L2Nu	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_mtop1715_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	3.360
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	80.95
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
TTToSemiLeptonic	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	364.35
ttV	TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.2814
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
VZ	WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8	12.178
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.42965
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
	ZZTo2L2Nu_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
Others	WWW, WWZ, WZZ, ZZZ, WWG	//

# 2018 MC samples

Process	Sample	Cross section [pb]
Drell-Yan	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610
	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 ( $H_T < 70$ GeV)	6077.22
	DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	169.9
	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	147.4
	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	40.99
	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.678
	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.367
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.6304
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1514
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003565
TTTo2L2Nu	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.310
Single top	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	3.360
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	80.95
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
TTToSemiLeptonic	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	364.35
ttV	TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	0.2814
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297
	TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8	0.2043
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062
VZ	WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8	12.178
	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	4.42965
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595
	ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8	0.5640
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22
Others	WWW, WWZ, WZZ, ZZZ, WWG	//

Mass point	Cross-section [pb]
<b>Scalar mediators</b>	
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi10	$4.959 \cdot 10^{-2}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi20	$3.235 \cdot 10^{-2}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi50	$1.323 \cdot 10^{-2}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi100	$5.633 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi150	$3.397 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi200	$2.359 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi250	$1.720 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi300	$1.328 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi350	$1.018 \cdot 10^{-3}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi400	$6.717 \cdot 10^{-4}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi450	$4.535 \cdot 10^{-4}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi500	$3.206 \cdot 10^{-4}$
DMscalar_Dilepton_top_tWChan_Mchi1_Mphi1000	$3.045 \cdot 10^{-5}$
<b>Pseudoscalar mediators</b>	
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi10	$6.151 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi20	$5.869 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi50	$4.946 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi100	$3.658 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi150	$2.754 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi200	$2.097 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi250	$1.616 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi300	$1.253 \cdot 10^{-3}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi350	$7.851 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi400	$4.371 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi450	$3.095 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi500	$2.321 \cdot 10^{-4}$
DMpseudoscalar_Dilepton_top_tWChan_Mchi1_Mphi1000	$2.791 \cdot 10^{-5}$

# $t\bar{t} + \text{DM}$ signal samples

Mass point	Cross-section [pb]
<b>Scalar mediators</b>	
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_50	$3.405 \cdot 10^{-1}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_100	$8.027 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_150	$2.673 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_200	$1.158 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_250	$6.020 \cdot 10^{-3}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_300	$3.579 \cdot 10^{-3}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_350	$2.376 \cdot 10^{-3}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_400	$1.443 \cdot 10^{-3}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_450	$9.025 \cdot 10^{-4}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_500	$6.204 \cdot 10^{-4}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_20.mPhi_100	$7.993 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_30.mPhi_100	$8.052 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_40.mPhi_100	$8.147 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_45.mPhi_100	$8.319 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_49.mPhi_100	$8.304 \cdot 10^{-2}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_51.mPhi_100	$9.735 \cdot 10^{-4}$
TTbarDMJets.Dilepton.scalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_55.mPhi_100	$4.835 \cdot 10^{-4}$
<b>Pseudoscalar mediators</b>	
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_50	$3.440 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_100	$2.164 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_150	$1.414 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_200	$9.773 \cdot 10^{-3}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_250	$6.753 \cdot 10^{-3}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_300	$4.808 \cdot 10^{-3}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_350	$2.742 \cdot 10^{-3}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_400	$1.409 \cdot 10^{-3}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_450	$9.302 \cdot 10^{-4}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_1.mPhi_500	$6.618 \cdot 10^{-4}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_20.mPhi_100	$2.166 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_30.mPhi_100	$2.164 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_40.mPhi_100	$2.162 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_45.mPhi_100	$2.180 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_49.mPhi_100	$2.151 \cdot 10^{-2}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_51.mPhi_100	$1.993 \cdot 10^{-3}$
TTbarDMJets.Dilepton.pseudoscalar._LO_TuneCP5_13TeV-madgraph-mcatnlo-pythia8.mChi_55.mPhi_100	$7.750 \cdot 10^{-4}$

## 2016 triggers

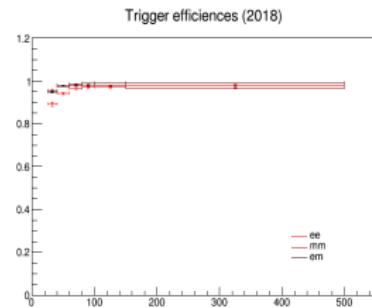
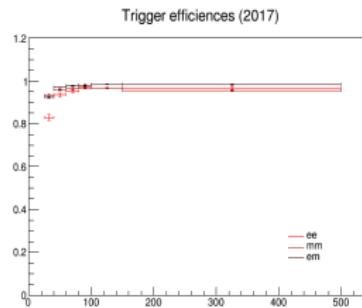
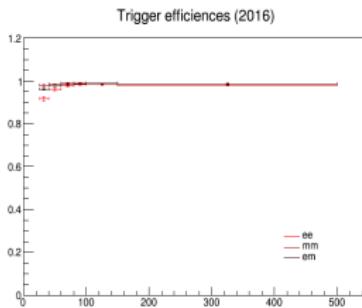
Dataset	Run range	HLT trigger path
SingleMu	[297020,306462]	HLT_IsoMu27_v*
SingleEle	[297020,306462]	HLT_Ele35_WPTight_Gsf_v*
DoubleEG	[297020,306462]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[297020,299336] [299337,306462]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*
MuonEG	[297020,306462] [297020,299336] [299337,306462]	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*

## 2017 triggers

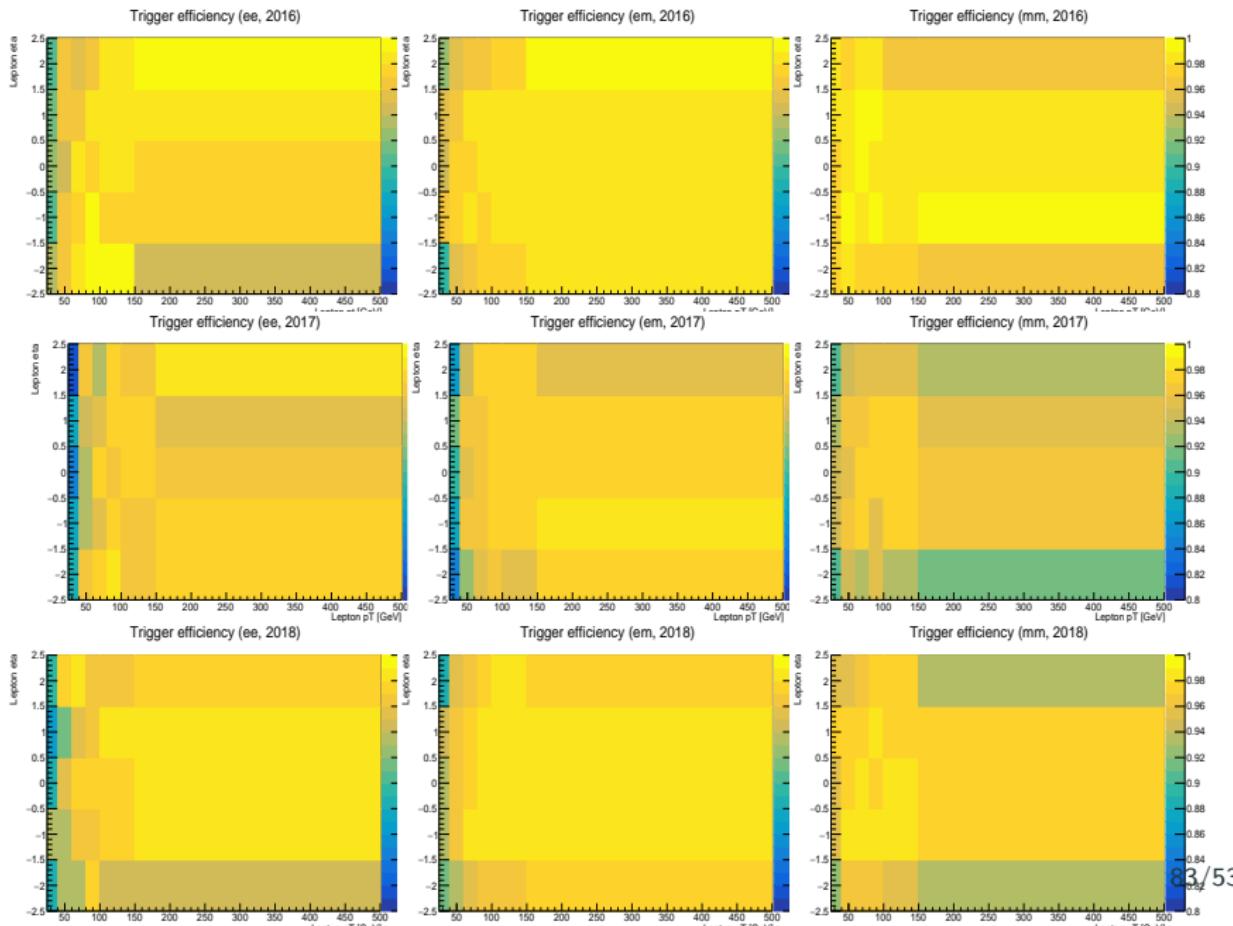
Dataset	Run range	HLT trigger path
SingleMu	[297020,306462]	HLT_IsoMu27_v*
SingleEle	[297020,306462]	HLT_Ele35_WPTight_Gsf_v*
DoubleEG	[297020,306462]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[297020,299336] [299337,306462]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v*
MuonEG	[297020,306462] [297020,299336] [299337,306462]	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v* HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*

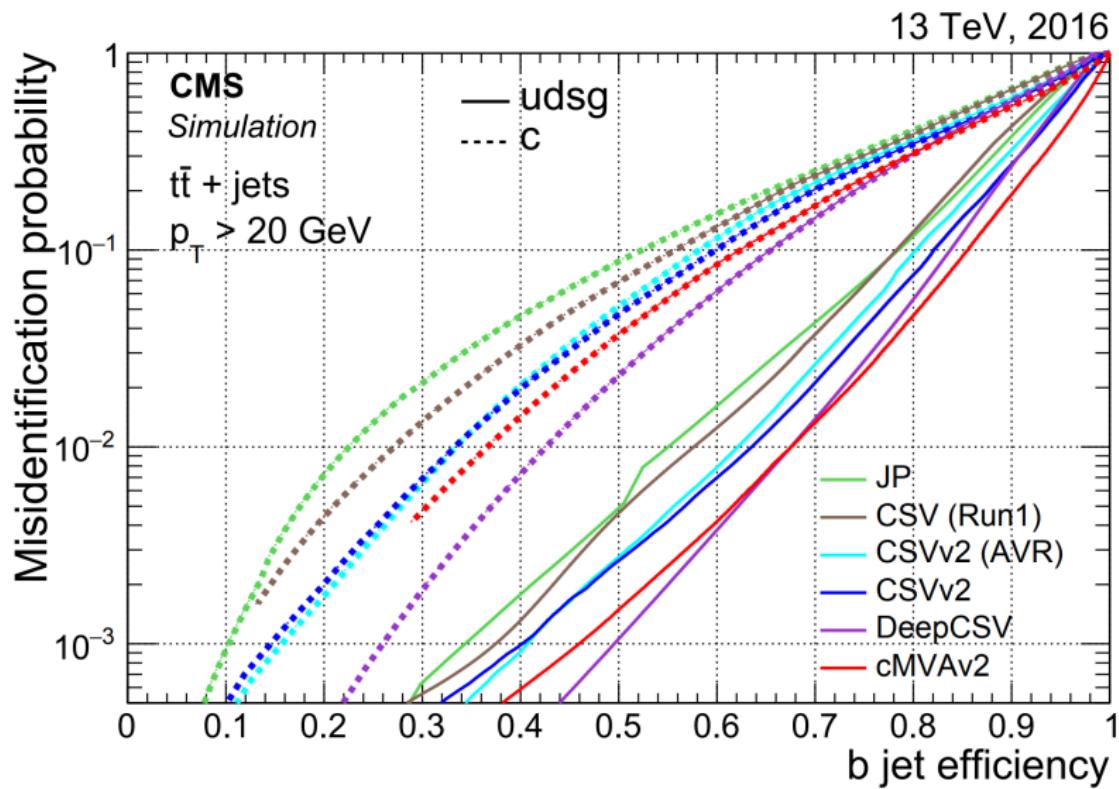
Dataset	Run range	HLT trigger path
SingleMu	[315252,325175]	HLT_IsoMu24_v*
SingleEle	[315252,325175]	HLT_Ele32_WPTight_Gsf_v*
DoubleEG	[315252,325175]	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*
DoubleMu	[315252,325175]	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*
MuonEG	[315252,325175]	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
		HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*

Trigger efficiencies computed using orthogonal MET datasets.



## 2D trigger efficiencies





## MET filters

Filter name	Applied to data	Applied to simulation
Flag_goodVertices	✓	✓
Flag_globalSuperTightHalo2016Filter	✓	✓
Flag_HBHENoiseFilter	✓	✓
Flag_HBHENoiselsoFilter	✓	✓
Flag_EcalDeadCellTriggerPrimitiveFilter	✓	✓
Flag_BadPFMuonFilter	✓	✓
Flag_ecalBadCalibFilterV2 <sup>†</sup>	✓	✓

<sup>†</sup> applied only to 2017 and 2018.

## Non prompt leptons contamination I

For the last few years, we calculated the background caused by fake leptons (particularly important at low  $p_T$ ) using a general **tight to loose data-driven method**, because we **do not expect the MC to be able to model** this background correctly.

By fake leptons, we mean:

- Truly fakes (e.g. jets taken as electrons)
- Real leptons from heavy flavour decay
- (Electrons from  $\gamma$  conversions)

Basically, we look for leptons satisfying the loose criteria **which pass or fail the tight criteria**. Then, several quantities are defined:

- The **fake rate** is defined as the ratio of fake leptons passing the tight criteria over fake leptons passing the loose criteria.
- The **prompt rate** is defined in a similar way but considering this time prompt leptons (usually assumed to be equal to 1).

→ Both these factors depend heavily on the working point chosen and on the kinematics of the event.

## Non prompt leptons contamination II

The **fake rate**  $f$  corresponds to the *probability for a fake lepton passing the loose criteria to also pass the tight ones*, and to therefore be considered as a real lepton in the analysis. We can measure it in a prompt lepton-free region (e.g., in a QCD enriched region).

The fake rate is then given by **the ratio between leptons passing tight isolation and those passing loose isolation**  $\rightarrow$  *tight-to-loose method*.

Once calculated, we need to **apply this weight** to a fake-lepton region **kinematically close to the signal region** (e.g. a region where one lepton fails the tight isolation criteria but passes the loose one) where we can estimate the number of yields  $\rightarrow$  cf. the second part of this talk.

We then need to assume that the weights calculated can be applied to another kinematically close region, accounting for the kinematics of the event.

## Non prompt leptons contamination III

To extrapolate these weights to the signal regions, we can define:

- $N_{pp}$  events where both leptons are prompt
- $N_{fp}$  events where one lepton is prompt and the other is fake
- $N_{ff}$  events where both leptons are fake
- $N_{tx}$  ( $x = 0, 1, 2$ ) events with 0, 1 or 2 leptons passing the right cuts → **only quantity directly measurable**

It is then straightforward to see that:

$$\left\{ \begin{array}{l} N_l = N_{pp} + N_{fp} + N_{ff} = N_{t2} + N_{t1} + N_{t0} \\ N_{t0} = (1-p)^2 N_{pp} + (1-p)(1-f)N_{fp} + (1-f)^2 N_{ff} \\ N_{t1} = 2p(1-p)N_{pp} + (f(1-p) + p(1-f))N_{fp} + 2f(1-f)^2 N_{ff} \\ N_{t2} = p^2 N_{pp} + pfN_{fp} + f^2 N_{ff} \end{array} \right.$$

These equations can be inverted, yielding:

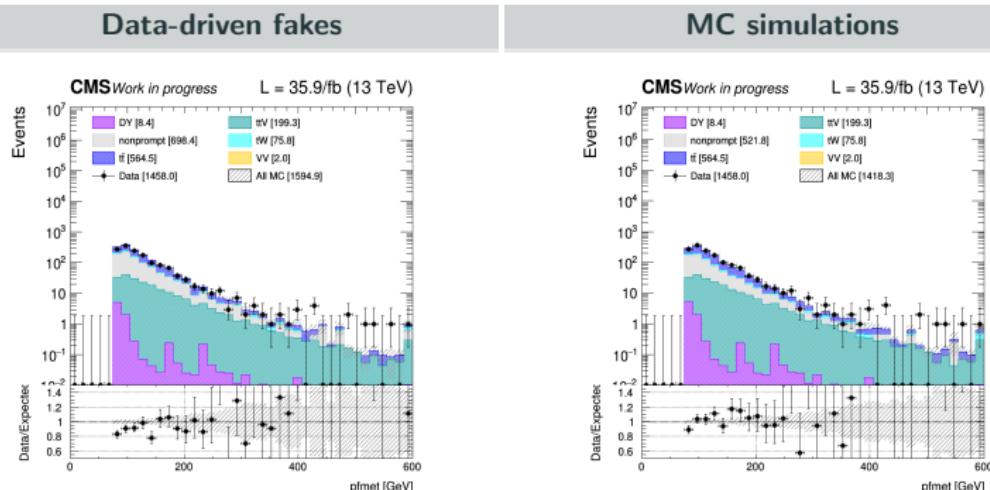
$$\begin{pmatrix} N_{pp} \\ N_{fp} \\ N_{ff} \end{pmatrix} = \frac{f - p}{-(p - f)^3} \cdot \begin{pmatrix} f^2 & -f(1-f) & (1-f)^2 \\ -2fp & p(1-f) + f(1-p) & -2(1-p)(1-f) \\ p^2 & -p(1-p) & (1-p)^2 \end{pmatrix} \cdot \begin{pmatrix} N_{t0} \\ N_{t1} \\ N_{t2} \end{pmatrix}$$

→ This formula gives us the recipe to apply weights to the leptons, which are summed (with the correct signs) over all events in order to **extract the number of events with prompt leptons only**.

## Non prompt leptons contamination V

The non-prompt contamination of this particular analysis however is expected to be dominated by the semileptonic decay of the  $t\bar{t}$  process, when one of the bottom quarks produced leads to a fake lepton.

This particular process, contrary to the W+jets and Z+jets processes, is expected to be modeled quite well by the MC simulations.



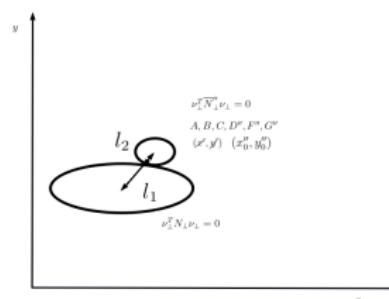
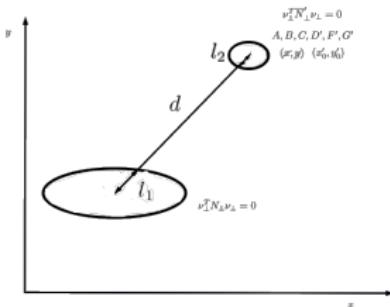
→ We decided to rely on the MC simulations for this analysis as they come with a smaller systematic uncertainty.

## Top reconstruction I

Many discriminating variables for our signals require knowledge of the top quark and anti-quark 4-momenta, only available after a **complete reconstruction of the  $t\bar{t}$  system**.

The **Batchchart analytical method** was used, relying on a geometrical approach to analytically solve equations constraining the decay of top quarks involving leptons:

- Invariant mass constraints from the top quark and W boson imply that the solution set for each neutrino 4-momentum can be constrained to an ellipse;
- A typical SM  $t\bar{t}$  event will give two ellipses close to each other, while ellipses coming from a typical signal event will be further apart and will not intersect.



Two discriminating variables naturally arise from this reconstruction: the **dark  $p_T$** , defined as the distance between the centroids of the ellipses, and the **overlapping factor  $R$** , defined as the ratio between the size of the ellipses and the distance between them.

In practice, **several complications** were taken into account:

- 0, 2 or 4 intersections can be observed while only 1 is physical, considered to be the solution with the lowest possible invariant mass for the  $t\bar{t}$  system;
- All the leptons and (b-)jets permutations of the event were taken into account;
- A **smearing process** was followed to evaluate the impact that imperfectly measured kinematic variables can have on this process (100 iterations per event, by updating several parameters such as the energy/direction of the jets and leptons within their respective uncertainties).

These considerations allowed to **increase the mean top reconstruction efficiency from around 70 to 90%**. Non-physical default values are set to all the variables which depend on the value of the neutrino momenta for the events which still fail this reconstruction.

## $m_{bl}^t$ variable

If a b-jet is produced in a top-quark decay, its invariant mass is **bounded from above** by  $\sqrt{m_t^2 - m_W^2} = 153$  GeV. Events compatible with two semileptonic top-quark decays can then be selected or rejected by introducing the observable  $m_{bl}^t$ :

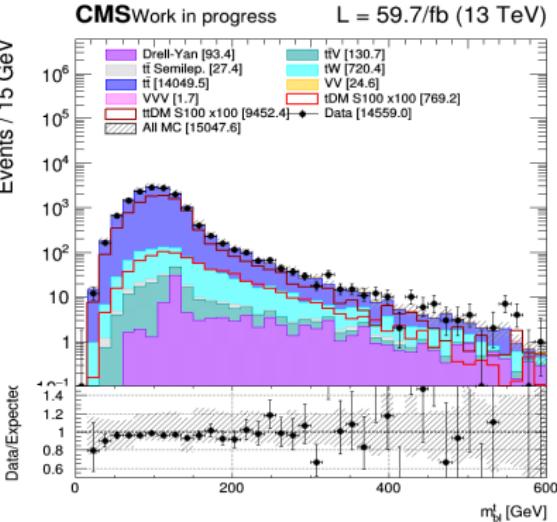
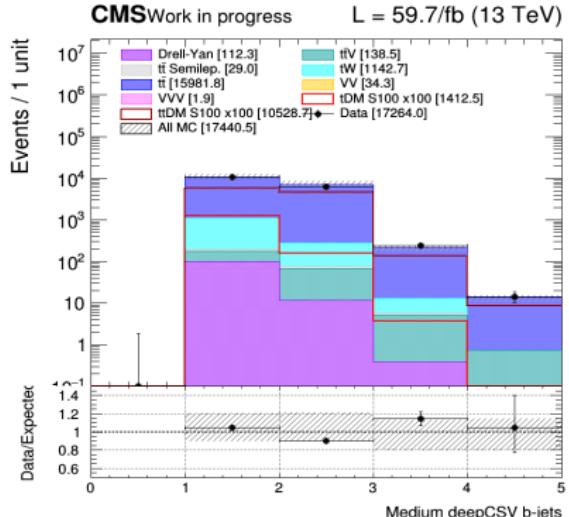
$$m_{bl}^t = \min (\max(m_{l_1 j_a}, m_{l_2 j_b}))$$

In this equation, the minimization is performed either:

- Over all the possible combinations of jets  $j_a, j_b$  among the b-jets of the events if three or more j-bets are observed;
- Or over the b-jet(s) observed plus the non b-tagged jet having the highest b-tag weight of the event.

This variable is expected to **give some discrimination** between our two signals of interest.

# Number of b-jets and $m_{bl}^t$ variable



## Stransverse mass $M_{T2}^{\text{II}}$ and $M_{T2}^{bl}$

Extension of the transverse mass  $m_T$  to cases when **pairs of same flavor particles decay into one visible and one invisible particle**, such as the double  $W \rightarrow l\nu$  decay.

Here, 2 neutrinos contribute to the presence of ptmiss and the individual contribution of each particle ( $\not{p}_{T1}$ ,  $\not{p}_{T2}$ ) to this quantity cannot be inferred.  $M_{T2}^{\text{II}}$  is defined as:

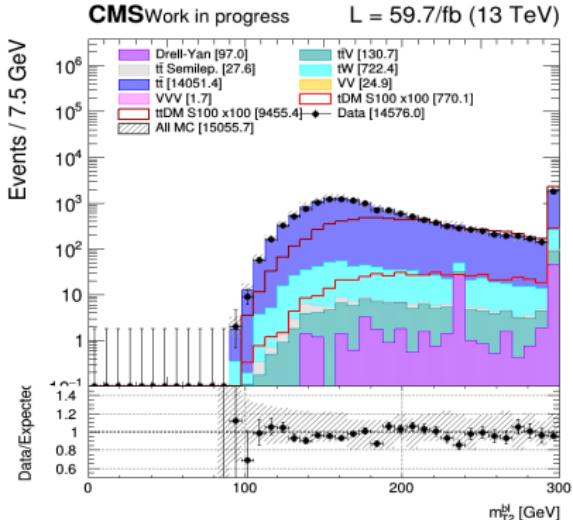
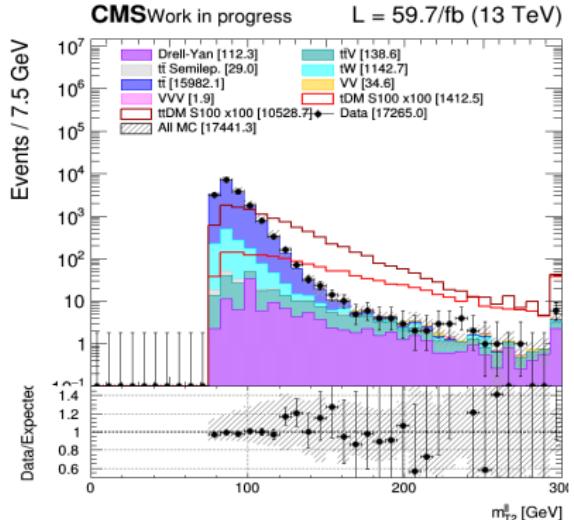
$$\begin{cases} M_{T2}^2 = \min_{\not{p}_{T1} + \not{p}_{T2} = \not{p}_{T\text{tot}}} \left( \max \left( m_T^2(\not{p}_{T1}, \not{p}_{T1}), m_T^2(\not{p}_{T2}, \not{p}_{T2}) \right) \right) \\ m_T^2(\not{p}_T, \not{p}_T) = 4 |\not{p}_T| |\not{p}_T| \sin^2 \left( \frac{\alpha}{2} \right) \end{cases}$$

Different combinations ( $\not{p}_{T1}$ ,  $\not{p}_{T2}$ ) satisfying the condition  $\not{p}_{T1} + \not{p}_{T2} = \not{p}_{T\text{tot}}$  then need to be probed, keeping only the combination which results in the lowest possible value.

The  $t\bar{t}$  process is **expected to have an endpoint at the mass of the W**, while our eventual signal **does not have this limitation** because of the pair of dark matter particles produced.

The  $M_{T2}^{bl}$  variable is defined in a similar case, except that in this case, the lepton is paired with a b-jet. The jet/lepton permutation giving the smallest value is kept.

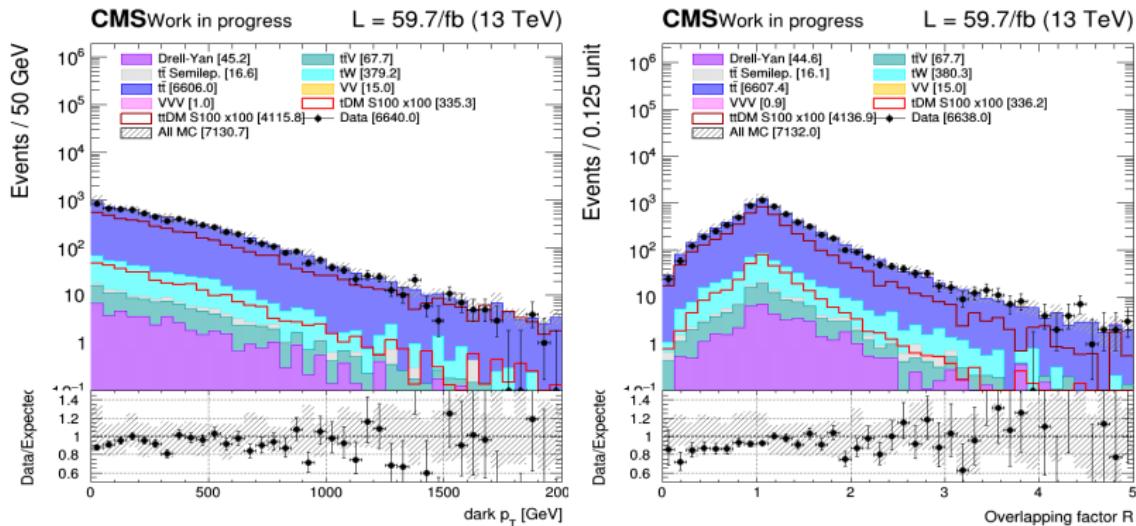
# Transverse mass $M_{T2}^{ll}$ and $M_{T2}^{bl}$



## Dark $p_T$ and overlapping factor $R$

Two discriminating variables naturally arise from the top reconstruction:

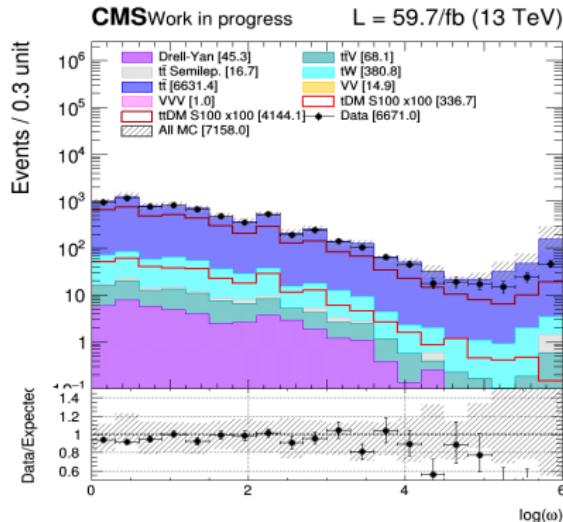
- The **dark  $p_T$** , defined as the distance between the centroids of the ellipses;
- And the **overlapping factor  $R$** , defined as the ratio between the size of the ellipses and the distance between them.



## Top reconstruction weight $W$

The top reconstruction smearing process does introduce a new discriminating variable: the **top reconstruction weight  $W$** .

Indeed, in order to know which lepton/b-jets combination and which smearing iteration performs the best, a weight is assigned to each iteration by **comparing the invariant mass obtained  $m_{lb}$  with the expected distribution using generation**. The combination with the largest weight is then simply considered as the solution of the event.

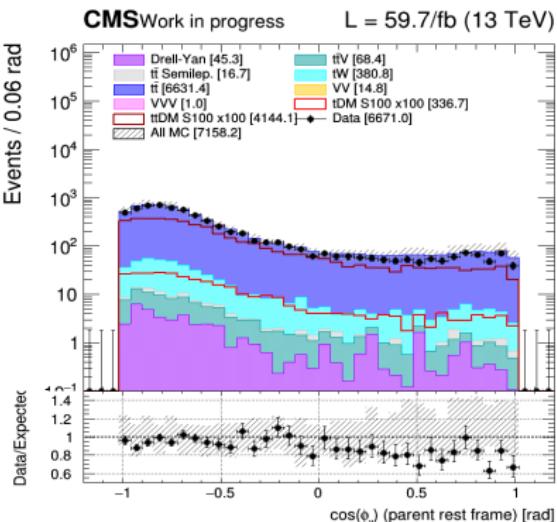
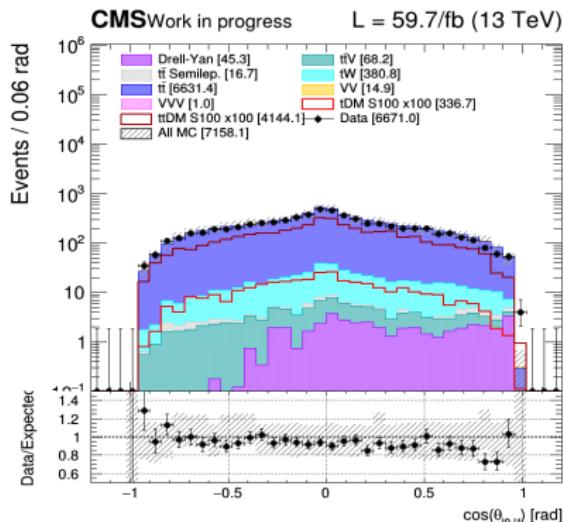


## Spin correlated variables

The **spin correlation** in a  $t\bar{t}$ -like event is **expected to be conserved**, because of the short lifetime of the top quark, and can be inferred from the top quark decay products.

Such variables are interesting because **the spin correlation depend on the production mechanism** and will be influenced by the additional coupling to a scalar or pseudoscalar mediator, making this a perfect candidate to be good discriminating variables:

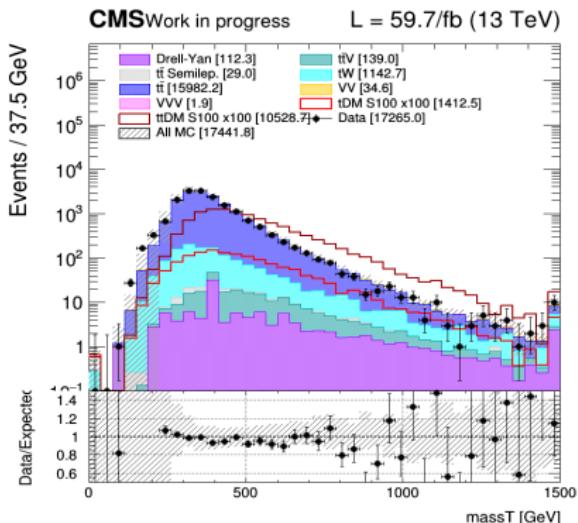
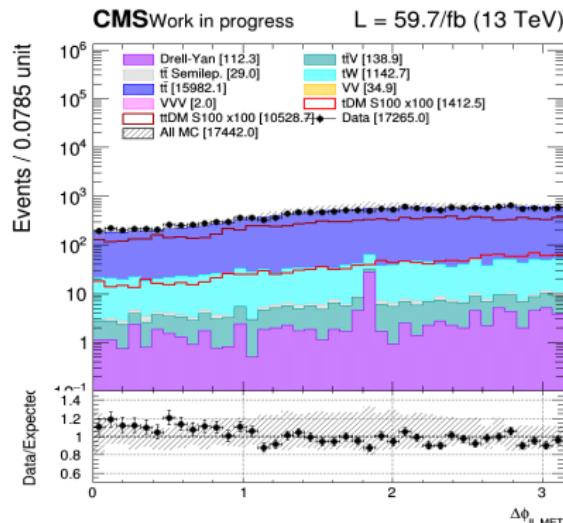
- $\xi = \cos(\theta_i) \cos(\theta_j)$ , where  $i$  and  $j$  are either leptons, b-jets or neutrinos;
- $\cos(\Phi_{i,j})$ , the cosine of the full opening angle of such top decay products in their respective parent rest frames.



## Other discriminating variables I

Several other variables were considered for their discriminating power:

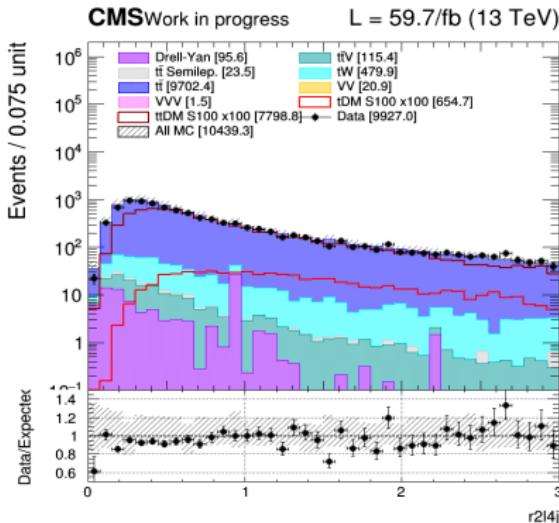
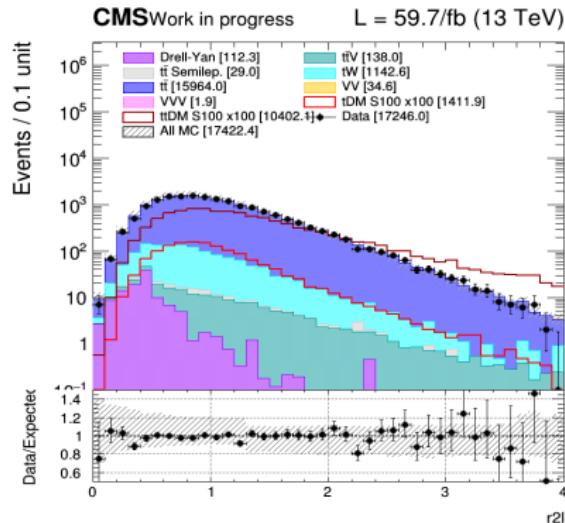
- $\Delta\Phi(E_T^{\text{miss}}, \text{II})$ : the **distribution in  $\Phi$**  of the two leptons is expected to change depending on the eventual production of DM;
- massT, which corresponds to the **scalar sum of the transverse component of the MET, the two leptons and the two b-jets** obtained by the top reconstruction.



## Other discriminating variables II

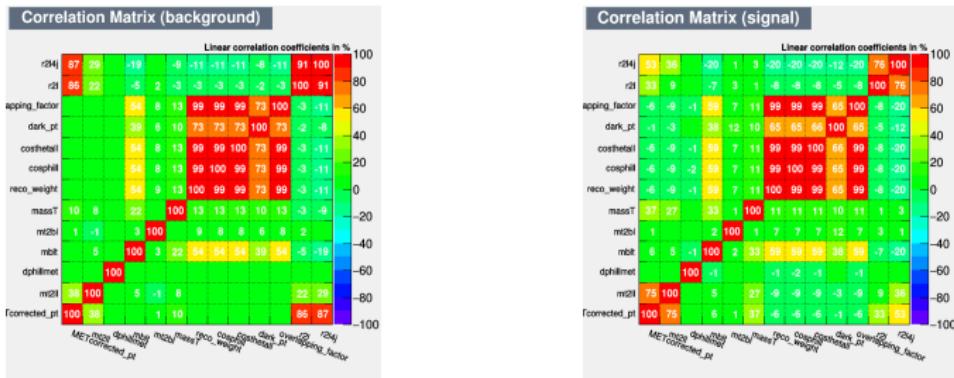
Two other interesting variables used by the ATLAS collaboration for their own analysis:

- $r2l$ , defined as the ratio between the  $\text{ptmiss}$  and the  $p_T$  of the two leptons observed;
- And  $r2l4j$ , defined in a similar way but considering additionally the  $p_T$  of the first 4 jets (if they exist) in the sum in the denominator.

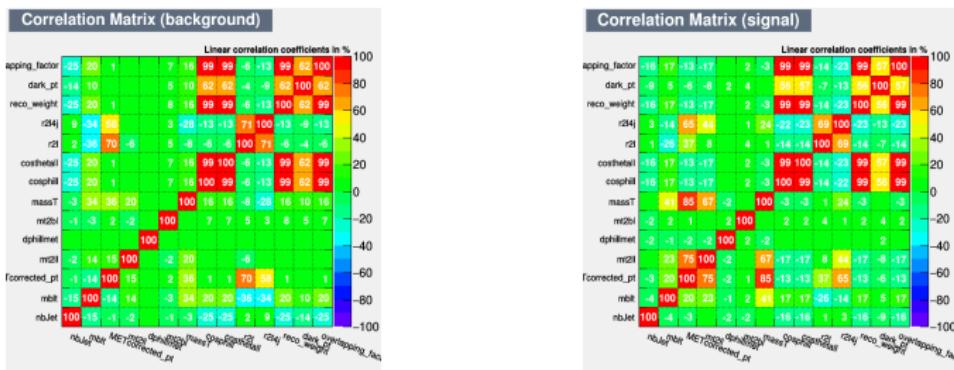


## Input variables correlation

## $t/\bar{t}$ +DM region



## $t\bar{t}$ +DM region

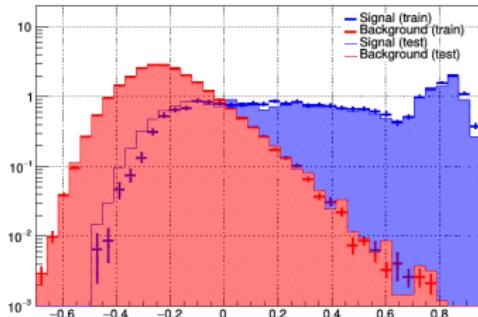


# Overtraining check ( $t/\bar{t}$ +DM region, log scale)

## Scalar mediators

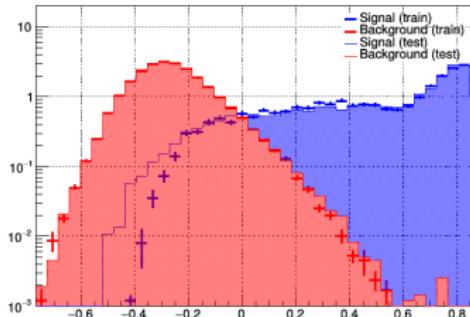
100 GeV

Signal (train)



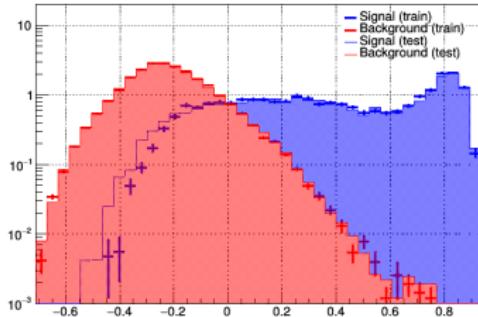
500 GeV

Signal (train)

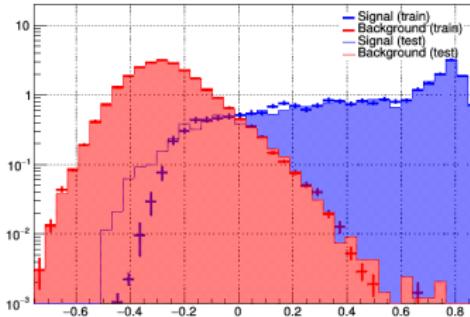


## Pseudoscalar mediators

Signal (train)



Signal (train)

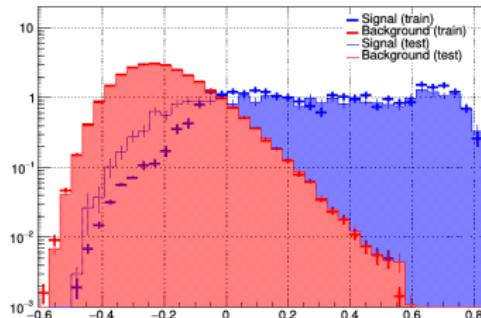


## Overtraining check ( $t\bar{t}$ +DM region, log scale)

### Scalar mediators

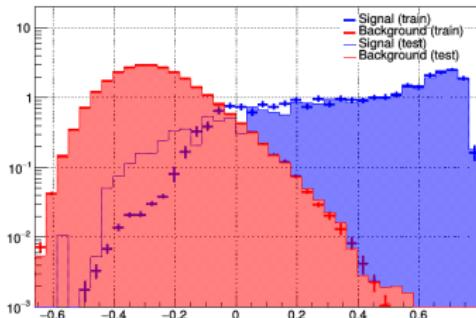
100 GeV

Signal (train)



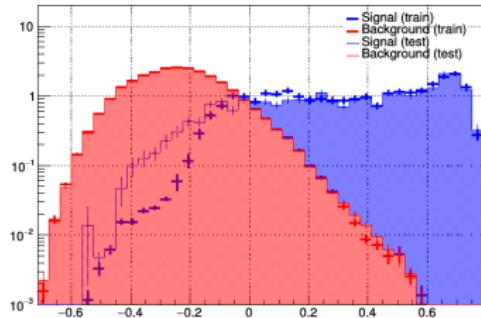
500 GeV

Signal (train)

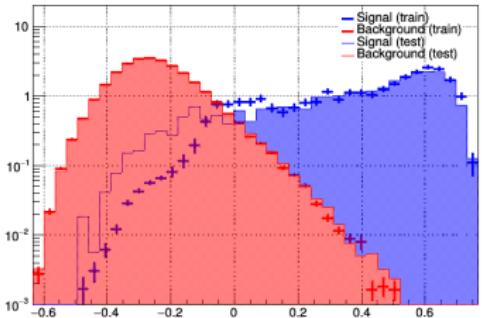


### Pseudoscalar mediators

Signal (train)



Signal (train)

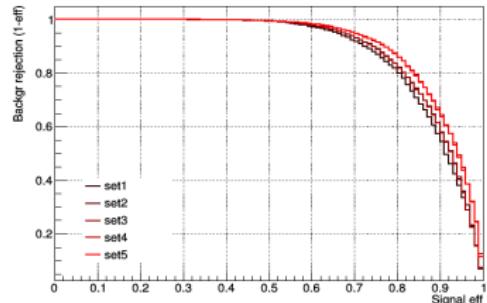


# Input variables selection

## BDT ROC curves

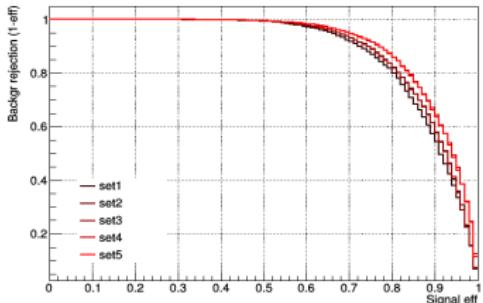
### $t/\bar{t}+DM$ region

ROC curves for several trainings



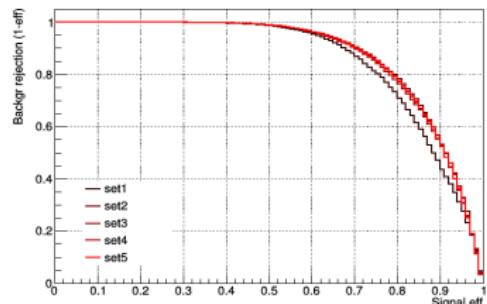
### $t\bar{t}+DM$ region

ROC curves for several trainings

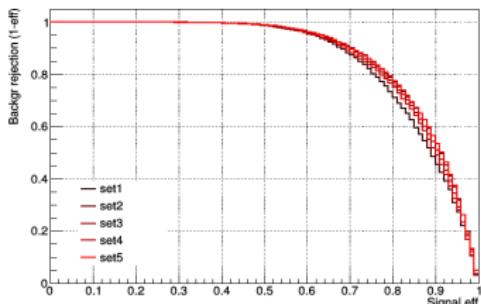


## ANN ROC curves

ROC curves for several trainings



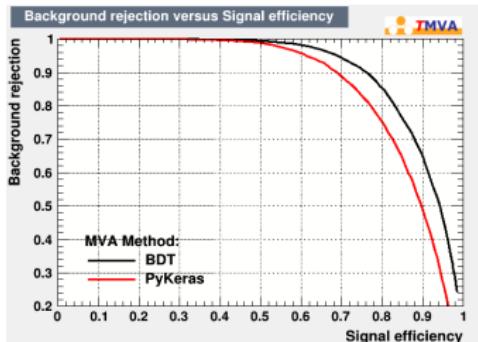
ROC curves for several trainings



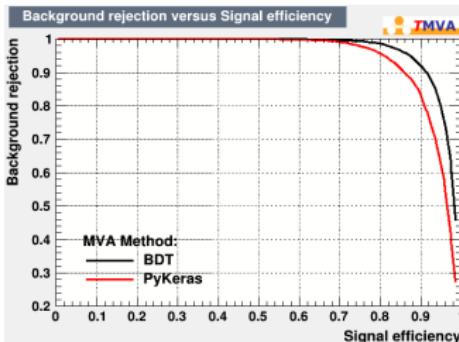
# BDT vs ANN ( $t/\bar{t}$ +DM region)

## Scalar mediators

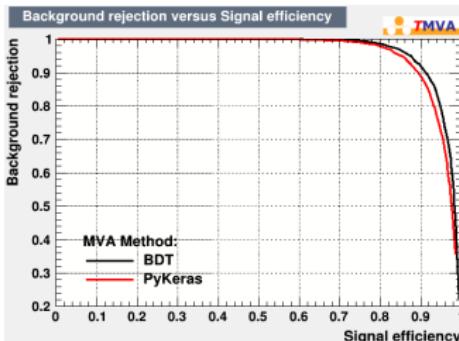
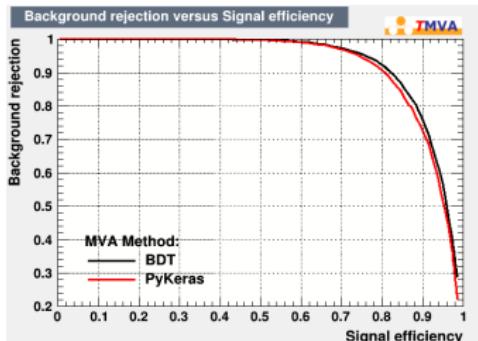
100 GeV



500 GeV



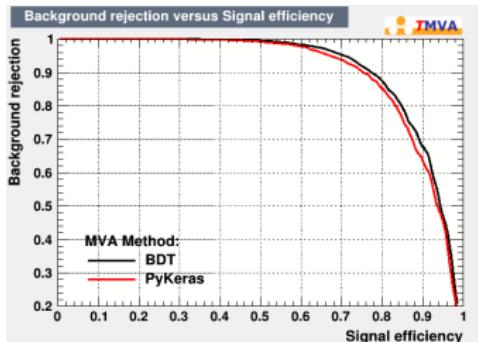
## Pseudoscalar mediators



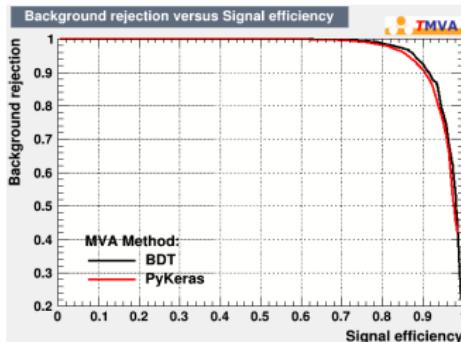
# BDT vs ANN ( $t\bar{t}$ +DM region)

## Scalar mediators

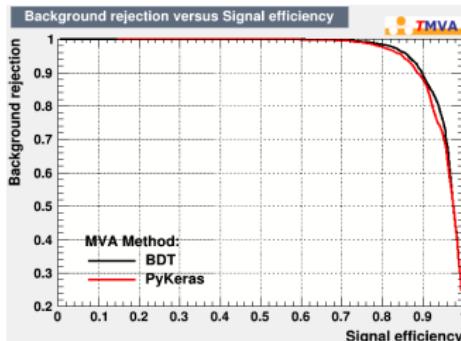
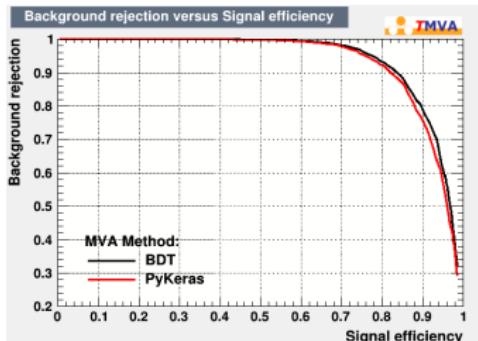
100 GeV



500 GeV



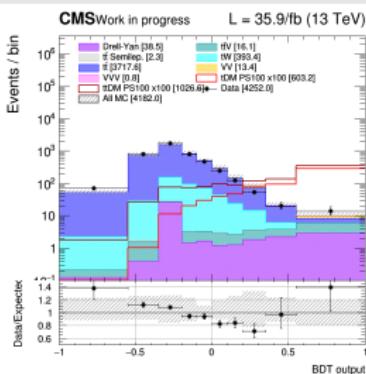
## Pseudoscalar mediators



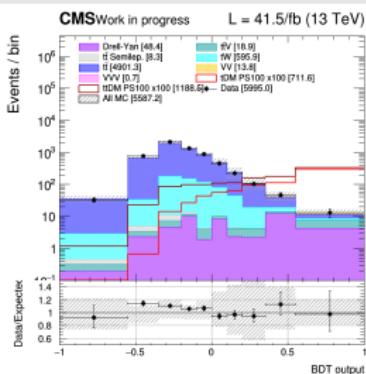
# BDT pseudoscalar 100 GeV output shape

## $t/\bar{t}$ +DM region

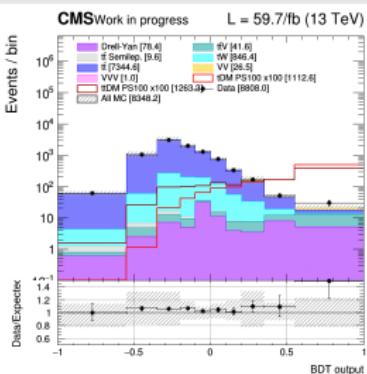
2016



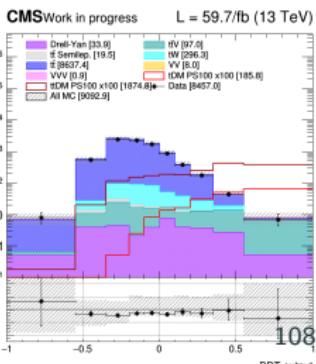
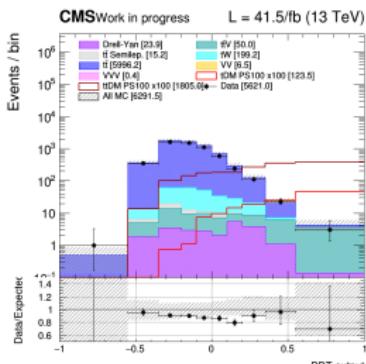
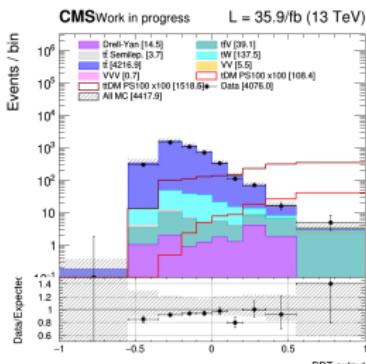
2017



2018



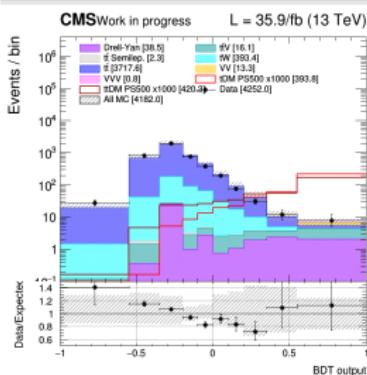
## $t\bar{t}$ +DM region



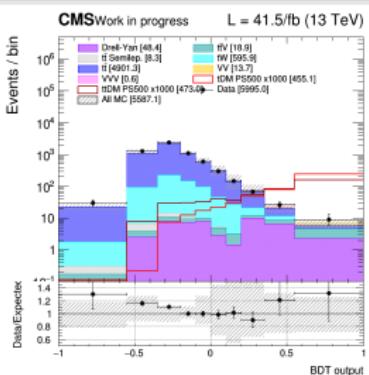
# BDT pseudoscalar 500 GeV output shape

## $t/\bar{t}$ +DM region

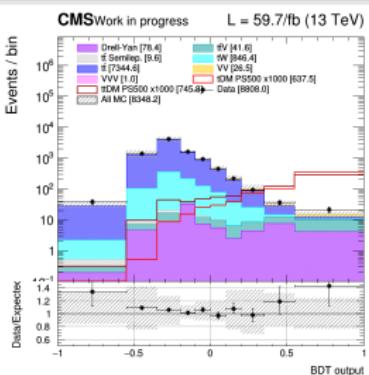
2016



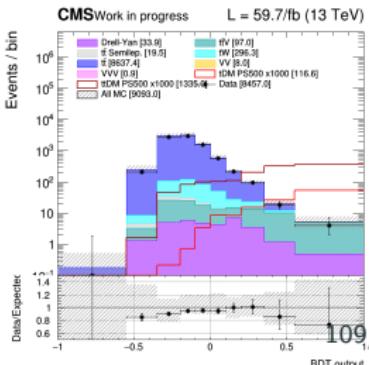
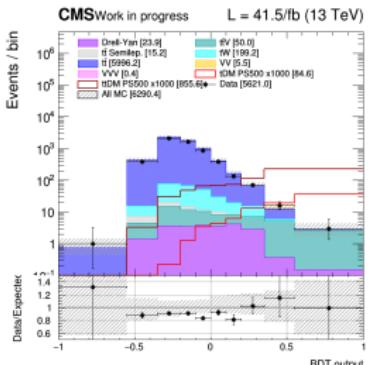
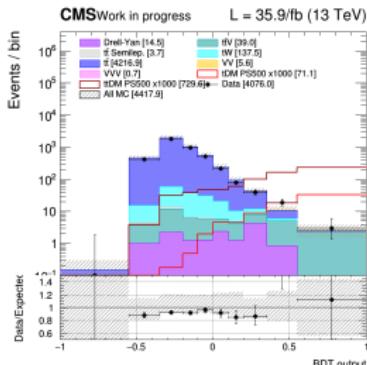
2017



2018



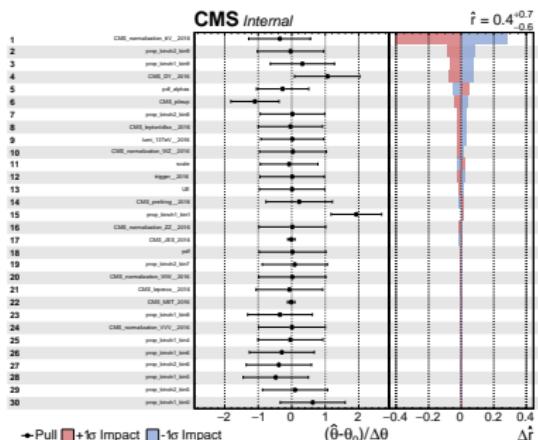
## $t\bar{t}$ +DM region



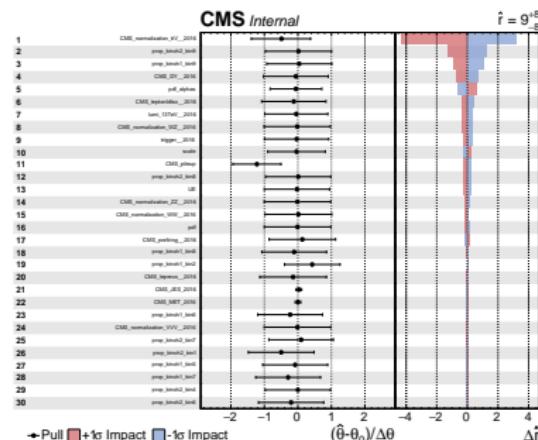
# Pulls and impact plots I

2016, pseudoscalar

100 GeV



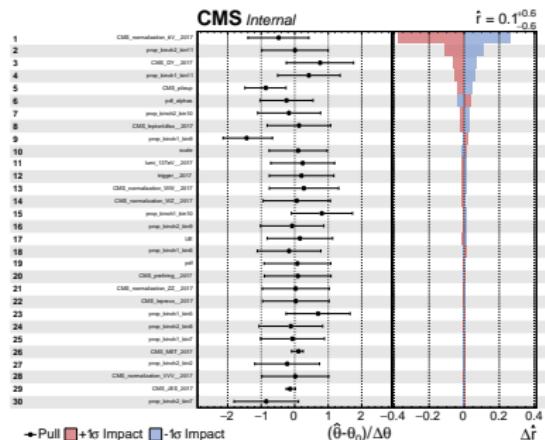
500 GeV



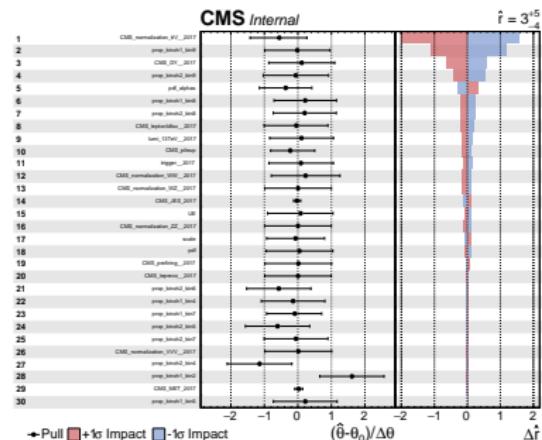
# Pulls and impact plots II

2017, scalar

100 GeV



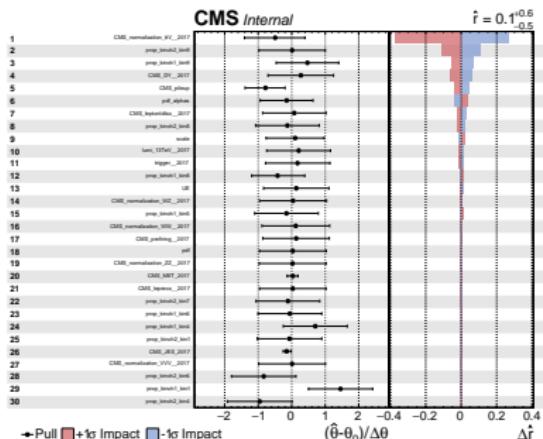
500 GeV



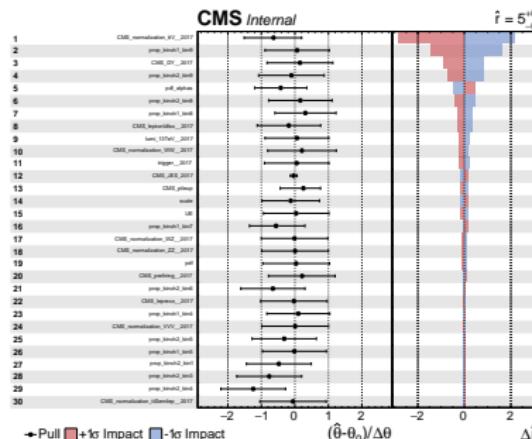
# Pulls and impact plots III

2017, pseudoscalar

100 GeV



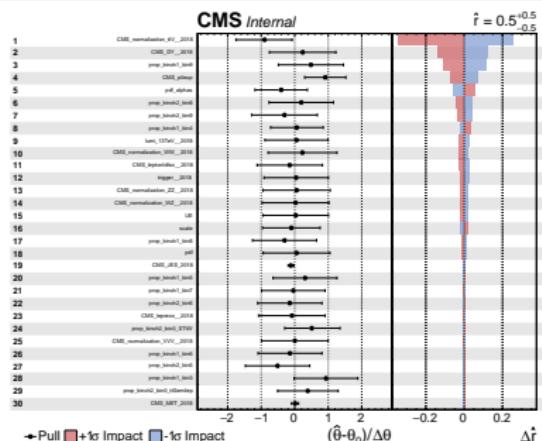
500 GeV



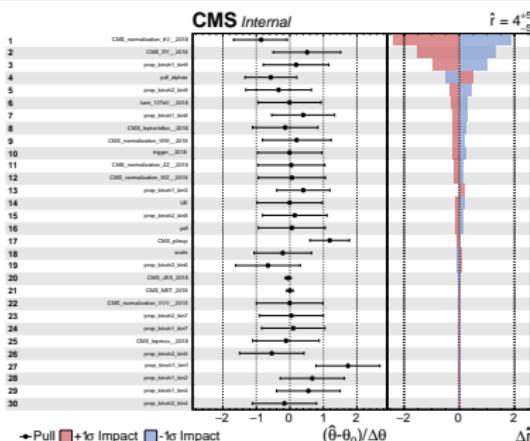
# Pulls and impact plots IV

2018, scalar

100 GeV



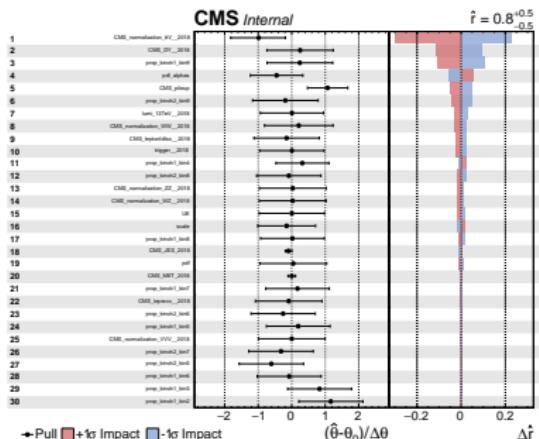
500 GeV



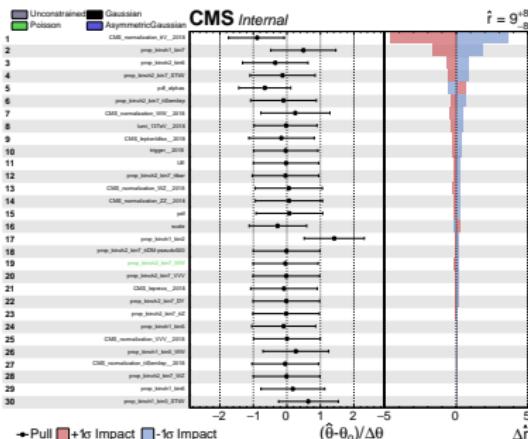
# Pulls and impact plots V

2018, pseudoscalar

100 GeV



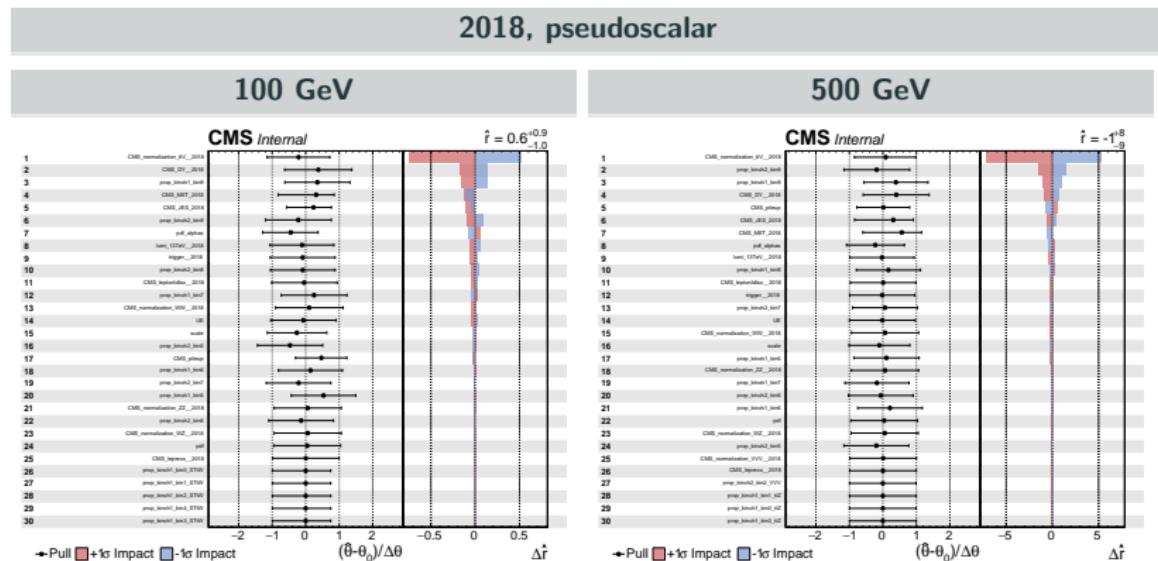
500 GeV



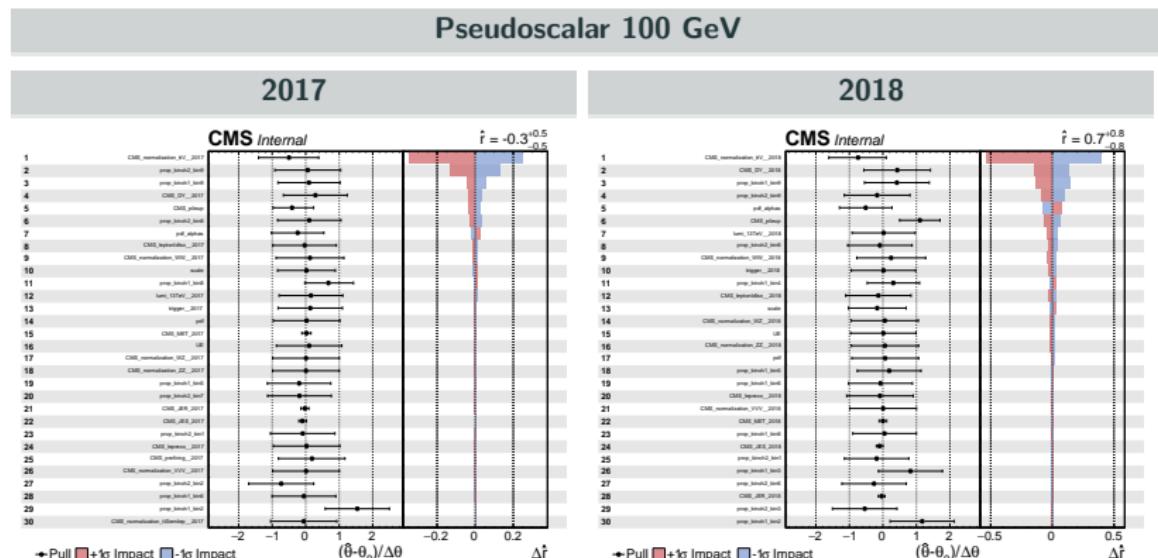
## JES/MET constraint

These strong constraints seem to come from the low BDT output region of the plots fed into the algorithm, where a lot of  $\bar{t}$  can be found. This region might actually play the role of a  $t\bar{t}$  control region, therefore constraining a lot these two systematic.

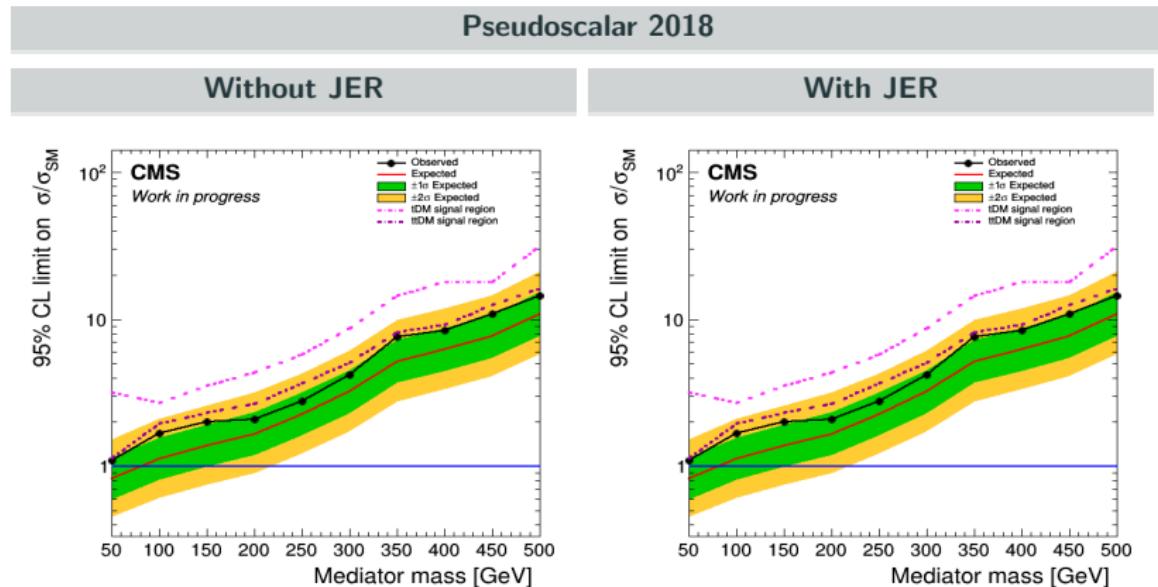
We therefore tried removing the BDT output  $< 0$  region, getting results less constrained.



The Jet Energy Resolution systematic was not included but in any case, it does not seem to be highly relevant in this analysis.



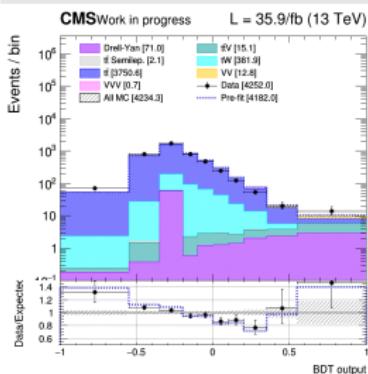
The final upper limits obtained by including this systematic are therefore similar to the ones obtained in the analysis shown here.



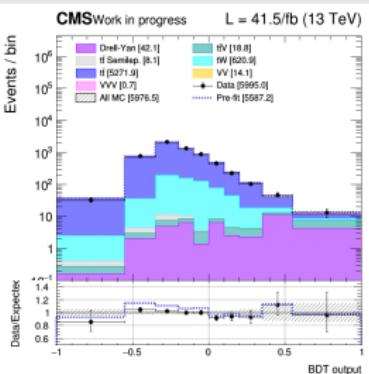
# Post-fit plots (pseudoscalar 100 GeV)

## $t/\bar{t} + \text{DM}$ region

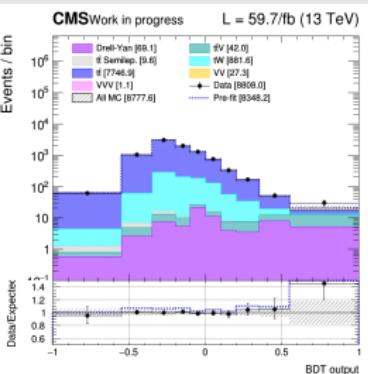
2016



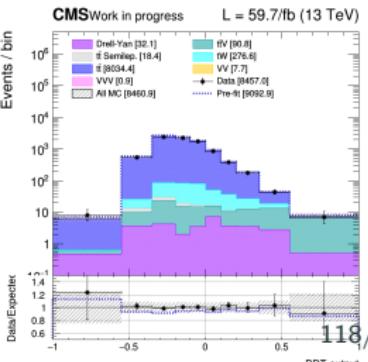
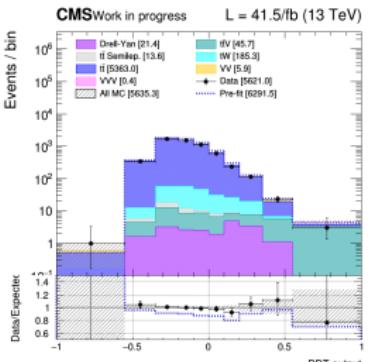
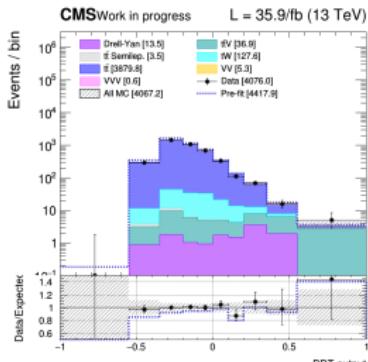
2017



2018



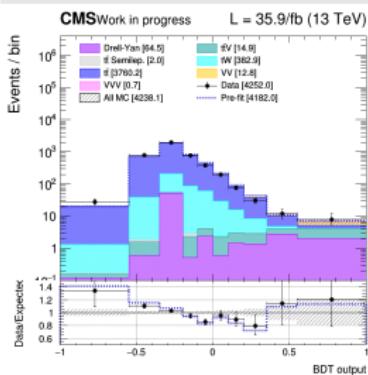
## $t\bar{t} + \text{DM}$ region



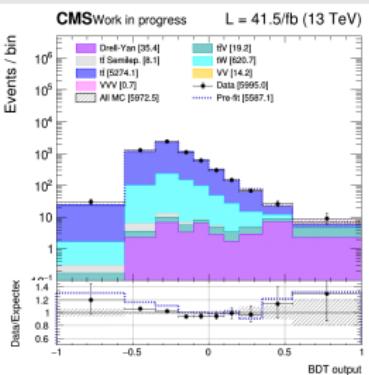
# Post-fit plots (pseudoscalar 500 GeV)

## $t/\bar{t} + \text{DM}$ region

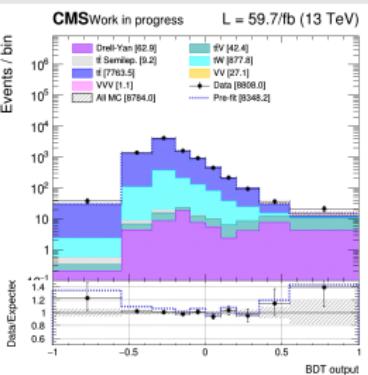
2016



2017



2018



## $t\bar{t}$ +DM region

