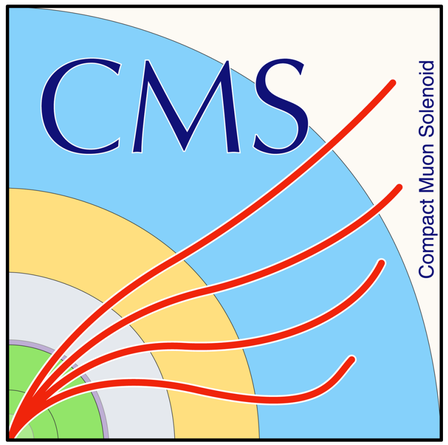
CMS DAS 2025 Long Exercise

A Dark Matter search with the Mono-Z signature

Danyer Perez Adan (RWTH Aachen)

RWTH AACHEN UNIVERSITY



Organization

Facilitators

- Danyer Perez Adan (RWTH) and Daina Leyva Pernia (DESY) → *welcome and many thanks!*

Physics analysis

- EXO-19-003, published as Eur. Phys. J. C 81 (2021) 13
- Perform one of the most traditional searches for *Dark Matter* in particle colliders: **the Mono-Z search**
- Analysis of 2016 dataset only: *availability of datasets and time considerations*

Past references

- CMS DAS 2024 (CERN): https://gitlab.cern.ch/cmsdas-cern-2024/long-ex-exo-monoz
  - [good] Detailed technical documentation (for their analysis setup) - *a nice website*
  - [bad] Limited functionality of core part of the exercise (missing ingredients!!!)
  - [bad] Absence of introductory/guidance slides for the long exercise - probably operated on the basis of their website

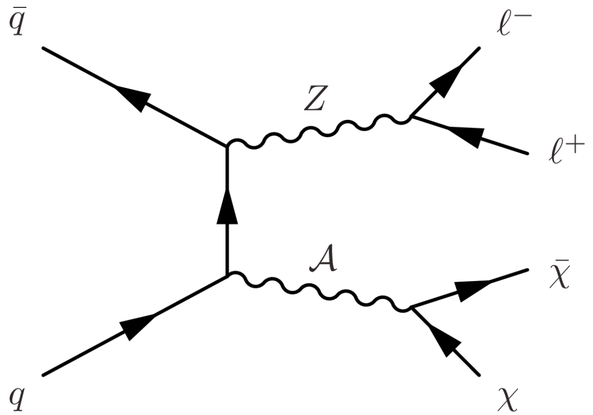
Technical basis for 2025 version

- Danyer's own analysis framework: mostly based on **ROOT** with **C++** functionality
  - Minimal set of structural components needed to conduct the long exercise (feasible for newcomers)

Exercise setup

Datasets

- Custom NTuples (NanoAOD skims) developed specifically for the Mono-Z search (actual Run 2 analysis)
  - Available in the EOS space of DAS schools (/eos/user/c/cmsdas/long-exercises) - **we should make sure it is not deleted!**
  - No need for XROOTD access - usage of *grid certificate* is *not required*
  - Only one of the possible types of signal scenario available on disk
    - not many fancy interpretations but at least **one** that is **functional**
  - Drell-Yan sample needed to estimate the background at low MET missing!!!
    - **required adaptation of analysis** a bit in the lower MET region



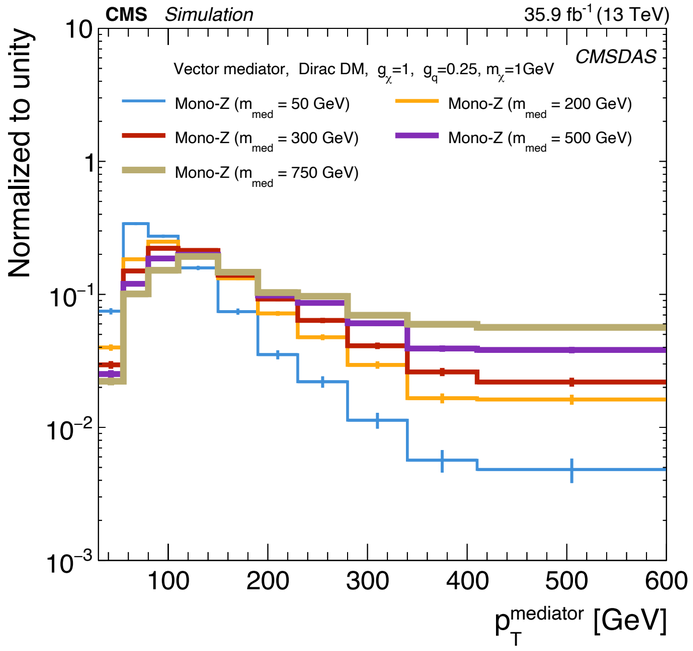
Exercise philosophy

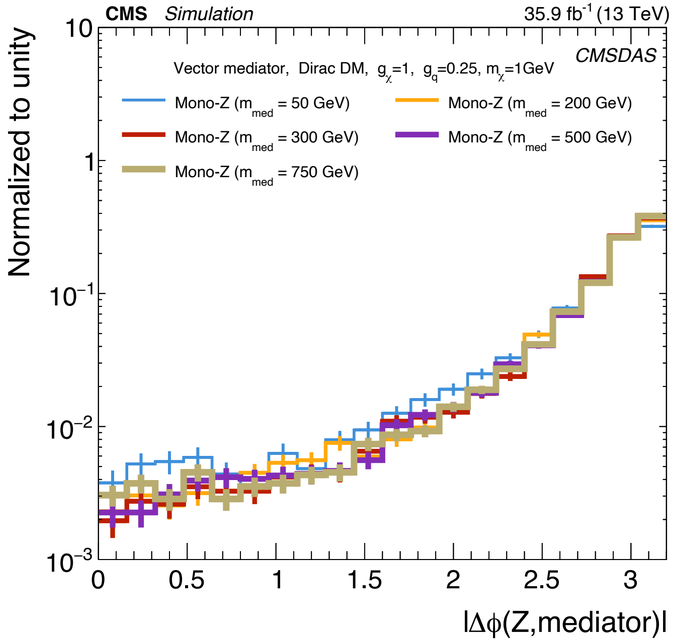
- Previous exercise version (in 2024) seemed to mostly focus on the statistical inference part (combine) - datacards for all signal scenarios
  - *Undesired* as students need to deal with many other aspects of an analysis first before reaching that point
- Plan for 2025 is that a **larger excursion** over the course of the analysis is done - according to time availability
  - Exercise is completely migrated to the new analysis framework
  - Many more aspects of experimental data analysis will be touched: *data structure*, *MC vs Data*, *corrections to MC*, *systematic types*

Understanding the nature of the signal

Generator level analysis

- **New** featured introduced at the beginning of the exercise
  - Will help the students **understand** a bit <u>better</u> the **physics** of the *signal* processes they are dealing with
  - Facilitates *comprehension* of some specific S-vs-B **optimization cuts** used in the original analysis
  - Relatively easy code setup to start **familiarizing** with the *analysis framework* before they embark on the real analysis
- They would be able to relatively fast produced their first distributions
  - They would discuss with us several of the physics aspects of the observations they make from the plots
- It should take them a couple hours (clearly with help) to implement it and 1 min to run it





The Mono-Z analysis: analysis setup

Data

✳ proton–proton collisions at CMS detector in 2016 - corresponding to an integrated luminosity of 35.9 fb^{-1}

✳ triggers: single-lepton and double lepton triggers

Signal

- simplified dark matter model for a spin-1 mediator with vector-like coupling
- few variations in $\{m_{\chi}, m_{\text{med}}\}$ 2D plane

Background

- $WZ \to l\nu ll$, $WW \to l\nu l\nu$, $ZZ \to llll\nu\nu$, $t\bar{t}$ and tW, VVV (with $V = [Z, W, \gamma]$), DY (not included due to unavailability)

Framework

✳ Students will interact directly with the **reality of performing** an analysis at CMS - almost *starting from zero*

✳ They will learn how properly deal with simulated data (*location, access, specific generators, yield estimation, calibrations, etc) and real data (event overlap, granularity, etc)*

✳ The main analysis code has been prepared in the form of a **template/skeleton** in such a way that they need to **fill in the gaps** with their *own implementations*

✳ Once they have a functional code, the entire **process** of job submission, event analyzing, production of intermediate results (pre-fit) is expected to **take around 30 mins per try**

✳ **Fixing mistakes** and **re-running** very much facilitated by fast analysis framework: help them realize that usually *iterations are common* and that you should be able to *re-run routines as quick/efficient as possible*

The Mono-Z analysis: event selection

Signal Regions

- Categorized in the number of jets
- **0 jet** and **1 jet** categories different signal-to-background ratios
- Normalization of main background processes determined form control regions

Quantity	Requirement	Target backgrounds
N_ℓ	=2 with additional lepton veto	WZ, VVV
p_T^ℓ	>25/20 GeV for leading/subleading	Multijet
Dilepton mass	$\left\|m_{\ell\ell} - m_Z\right\| < 15\,\text{GeV}$	WW, top quark
Number of jets	≤1 jet with $p_T^j > 30\,\text{GeV}$	DY, top quark, VVV
$p_T^{\ell\ell}$	>60 GeV	DY
b tagging veto	0 b-tagged jet with $p_T > 30\,\text{GeV}$	Top quark, VVV
τ lepton veto	0 τ_h cand. with $p_T^\tau > 18\,\text{GeV}$	WZ
$\Delta\phi(\vec{p}_T^{\,j}, \vec{p}_T^{\,miss})$	>0.5 radians	DY, WZ
$\Delta\phi(\vec{p}_T^{\,\ell\ell}, \vec{p}_T^{\,miss})$	>2.6 radians	DY
$\lvert p_T^{miss} - p_T^{\ell\ell}\rvert / p_T^{\ell\ell}$	<0.4	DY
$\Delta R_{\ell\ell}$	<1.8	WW, top quark
p_T^{miss} (all but 2HDM+a)	>100 GeV	DY, WW, top quark

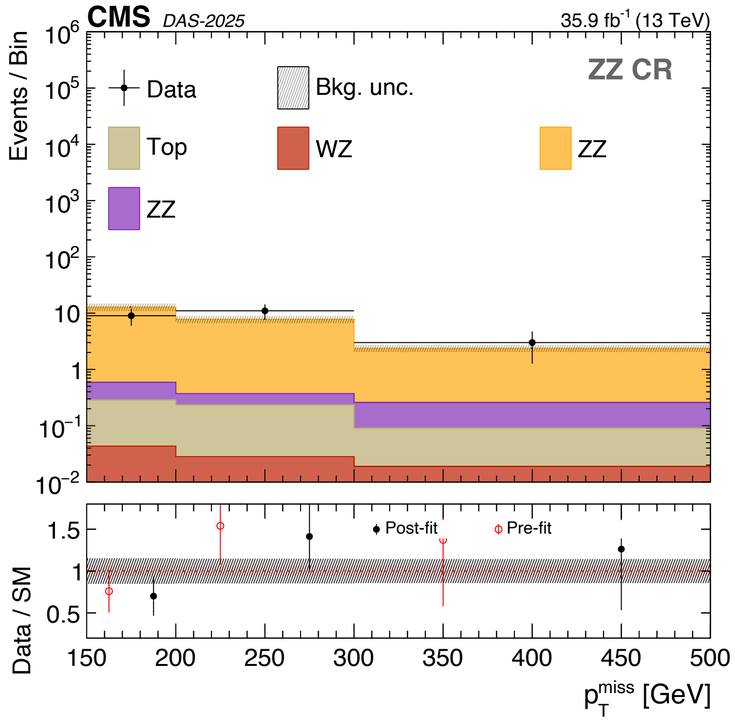
Moved to 150 GeV

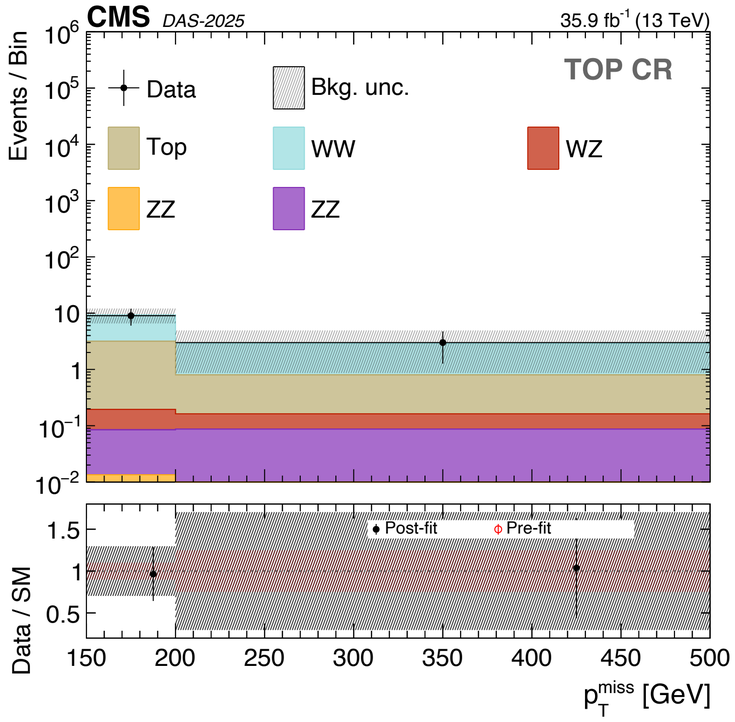
Background estimation

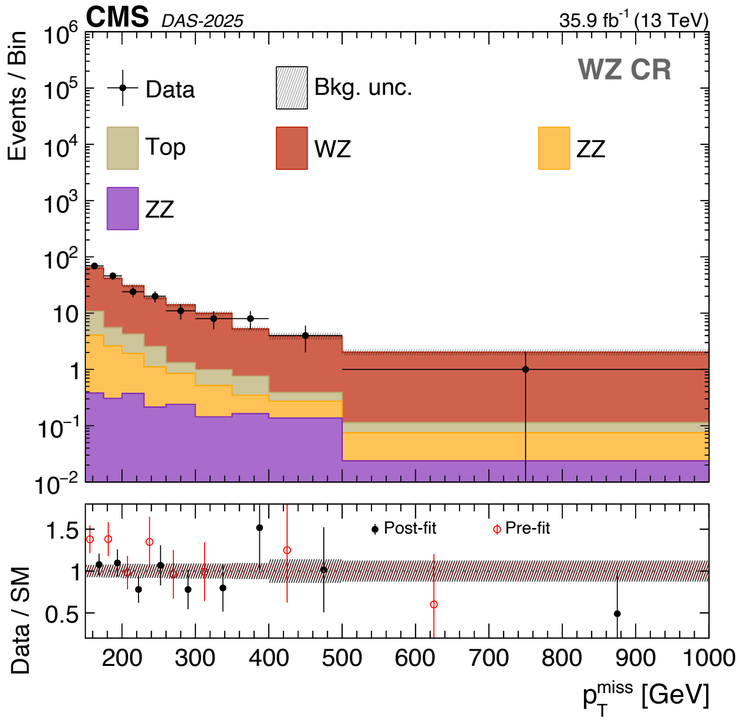
Control region	Description	Targeted process
4 Lepton	We look at 4 lepton decay with 2 Z boson candidates. We combine one lepton pair and MET to create "emulated MET". This emulated MET should model our SR ZZ.	ZZ
3 Lepton	We look at 3 lepton decay with a Z boson candidate and an additional lepton. We combine information from the lepton and the MET to create "emulated MET". This emulated MET should model our SR WZ.	WZ
Electron and Muon	We look as Opposite sign opposite flavor (OSOF) lepton pairs. With a tau veto this means we look at events with an electron and a muon that fall within the Z mass window.	TOP and WW

The Mono-Z analysis: simultaneous fit

- Students will understand the techniques behind the background estimation from subsidiary regions
- They will familiarize with some of the practical approaches to perform simultaneous maximum likelihood fits combining several regions
- Interpretation of the results would be a major focus during the discussions sessions in the long exercise

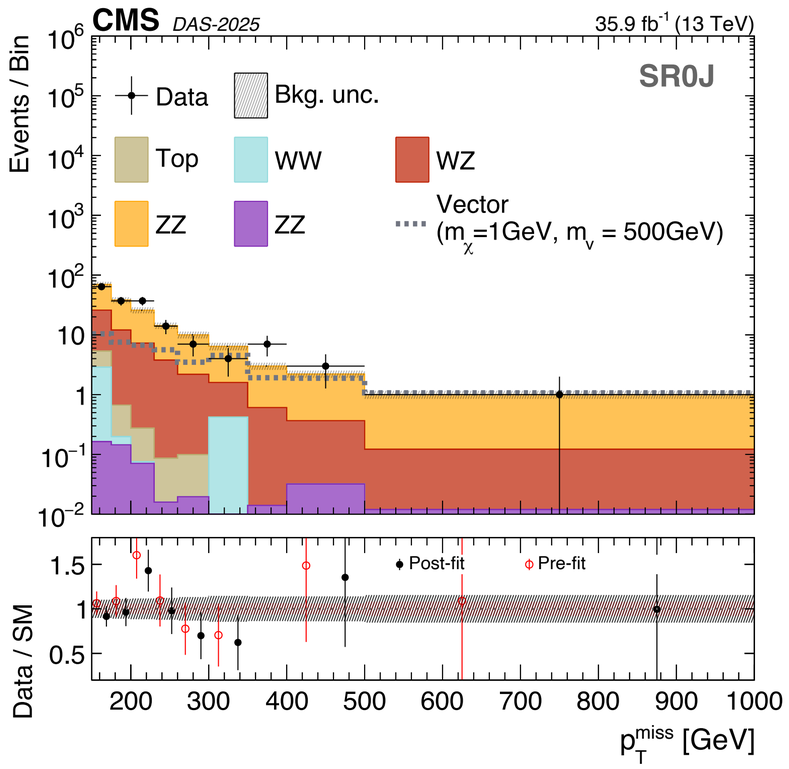


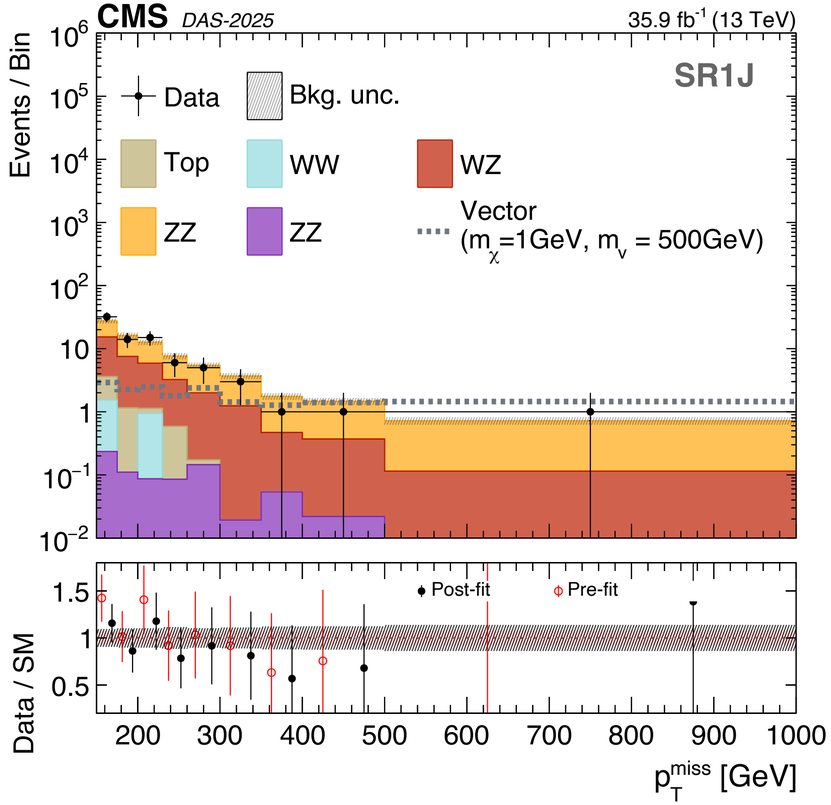




The Mono-Z analysis: simultaneous fit

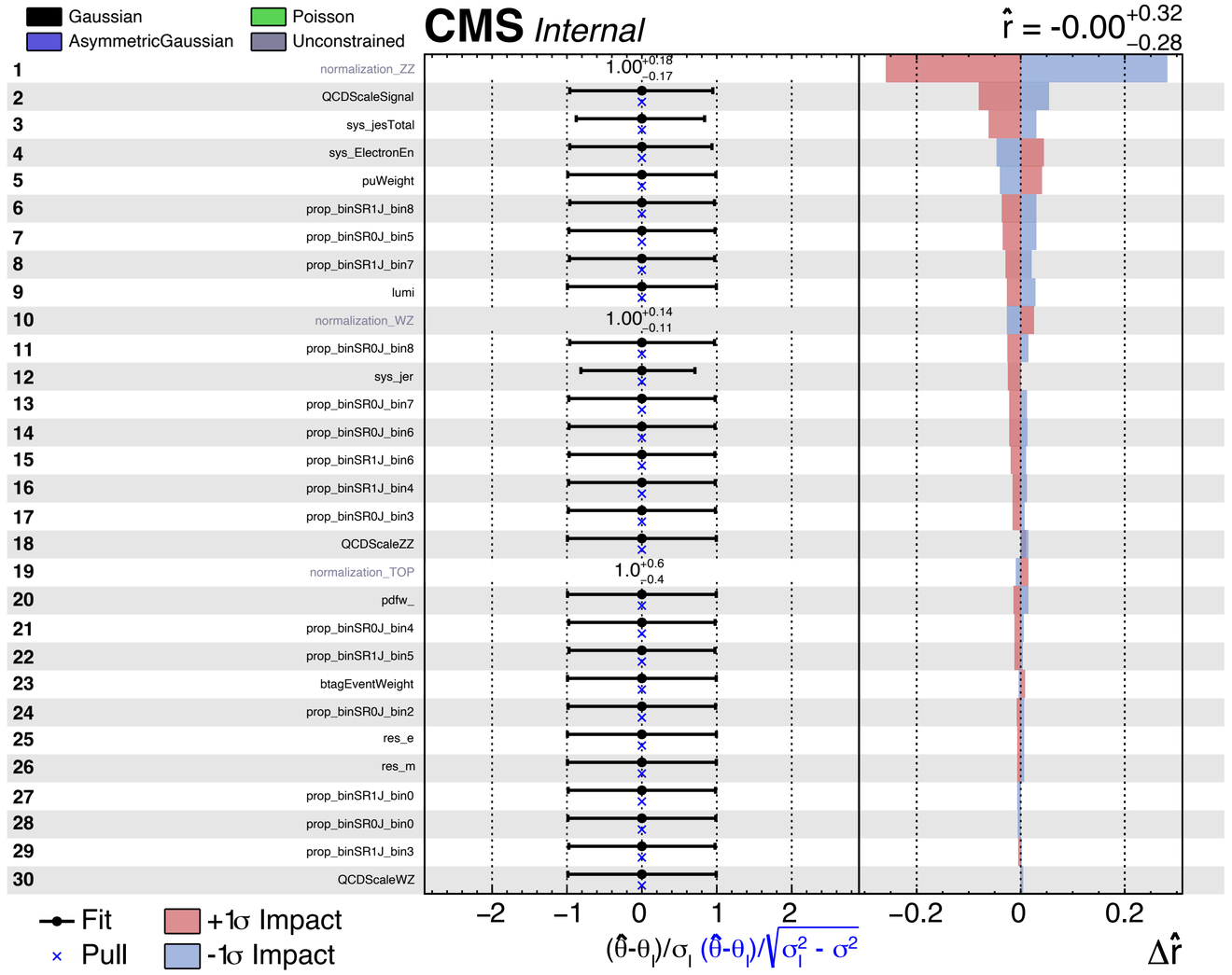
- Students will understand the techniques behind the background estimation from subsidiary regions
- They will familiarize with some of the practical approaches to perform simultaneous maximum likelihood fits combining several regions
- Interpretation of the results would be a major focus during the discussions sessions in the long exercise





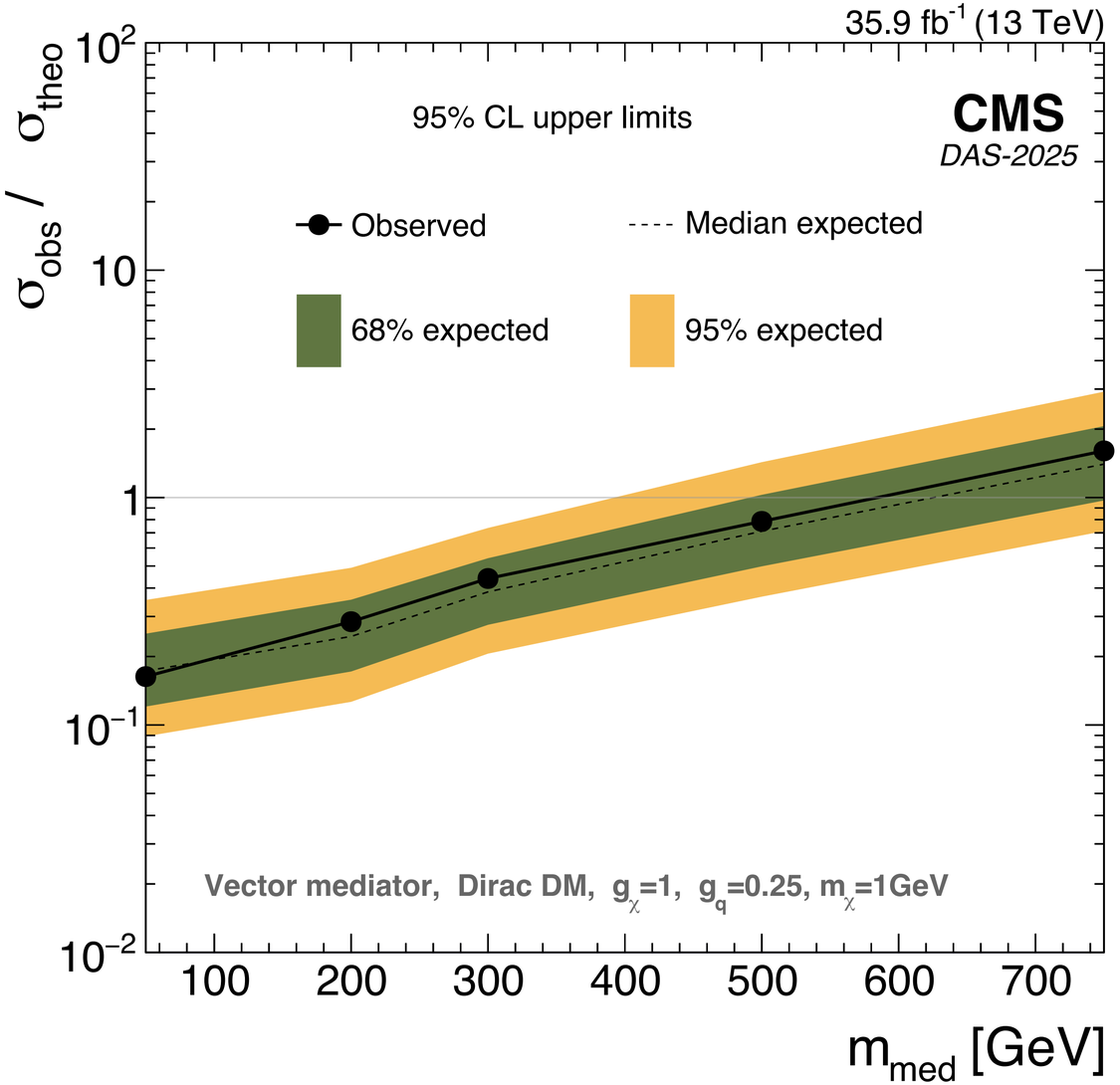
The Mono-Z analysis: performing standard tests

- Students will also learn how to work with several features of the Combination tool
  - Datacard construction and validation
  - Asimov fit to pseudodata
  - Fit diagnostics test and analysis
  - Impacts and pulls, along with their statistician interpretation
  - Asymptotic limit approximation and their implications



The Mono-Z analysis: examining their results

- Students should be able to find consistency and a clear explanation for most of the results
  - A study of the connection across intermediate and final results will be encouraged
  - Interpretation of their final outcome and a comparison with existing results is expected from them



Summary

- ✓ Mono-Z long exercise is concretely finalized and ready to be put in practice
- ✓ Successful migration to new analysis framework based on ROOT/C++, which should bring preparation in alternative technical skills to the school program
- ✓ Some limitations from past editions of the exercises were overcome and new ingredients were included
- ➡ TO-DOs
  - Finalize documentation in current common repository (currently at 70% of completion)
  - Improve instructions on the usage of the analysis framework - complementary to our presence on site
  - Ask Daina to run the entire exercise and confirm that all ingredients needed are in place and well documented
- Question: do we have any update about possibly including a MET short-exercise?

Thank you!

Danyer Perez Adan (RWTH Aachen)
0000-0003-3416-0726
danyer.perez.adan@rwth-aachen.de

Lehr- und Forschungsgebiet Hochenergiephysik
I. Physikalisches Institut B RWTH Aachen
https://www.rwth-aachen.de/