Quantum science & nanomaterials | QMat
The Interdisciplinary thematic institutes of the University of Strasbourg & CNRS & Inserm funded under the Excellence Initiative program

iPHC Institut Pluridisciplinaire Hubert Curien Strasbourg

JRJC 2023

Université de Strasbourg

Investir l'avenir

QMat Quantum Science and Nanomaterials

Combined fit of the hadronic tau energy scale and the identification scale factor in CMS

Poncet Océane

23/10/2023

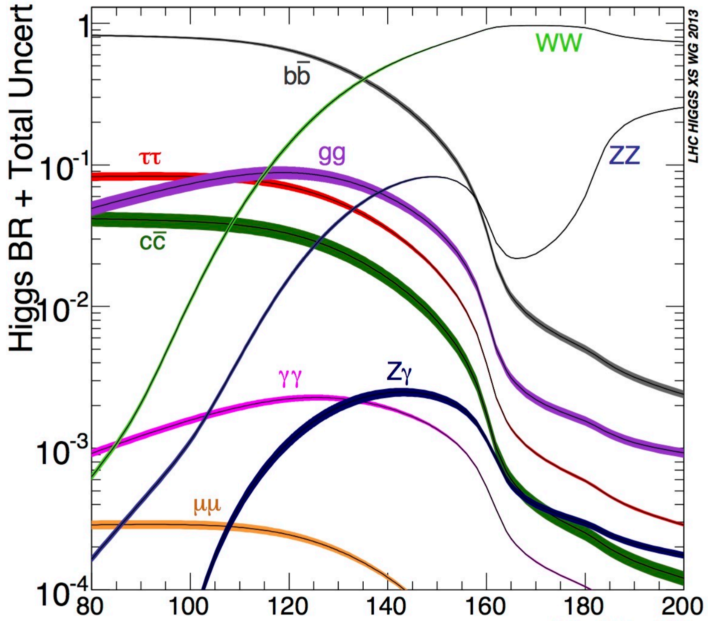
Tau leptons

Why are tau leptons interesting?

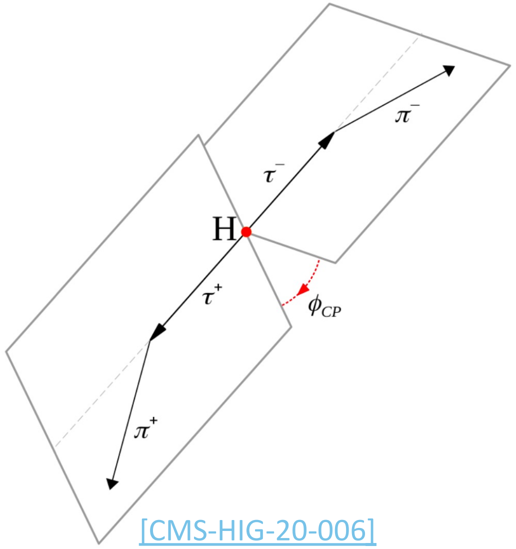
- Heaviest leptons.
- Only leptons that decay within the detector.
- Their decay products conserve spin information.
- They have the highest branching ratio for the $H \rightarrow ll$ decay.

Thus, tau leptons are used for several analyses involving taus as final states:

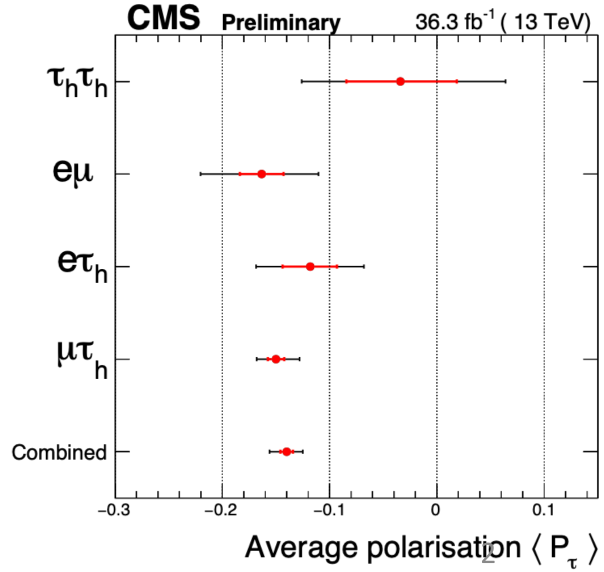
- **Standard Model (SM) probe:**
  - CP nature of $H \rightarrow \tau^{+}\tau^{-}$ decay.
  - Tau polarization in Z $\rightarrow \tau^{+}\tau^{-}$ decay.
- **Beyond SM searches:**
  - Leptoquarks, SUSY, high mass resonances …



[PDG]



[CMS-HIG-20-006]

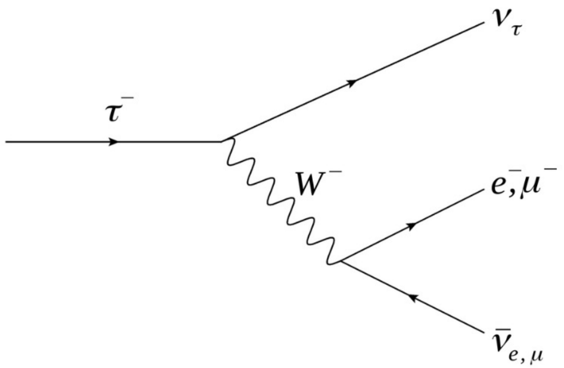


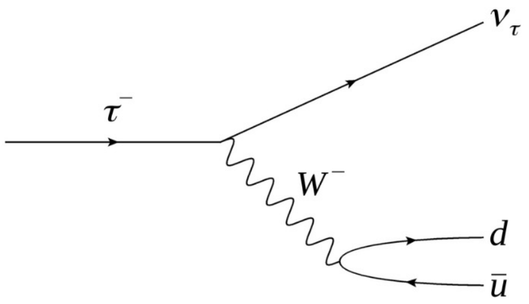
[CMS-PAS-SMP-18-010]

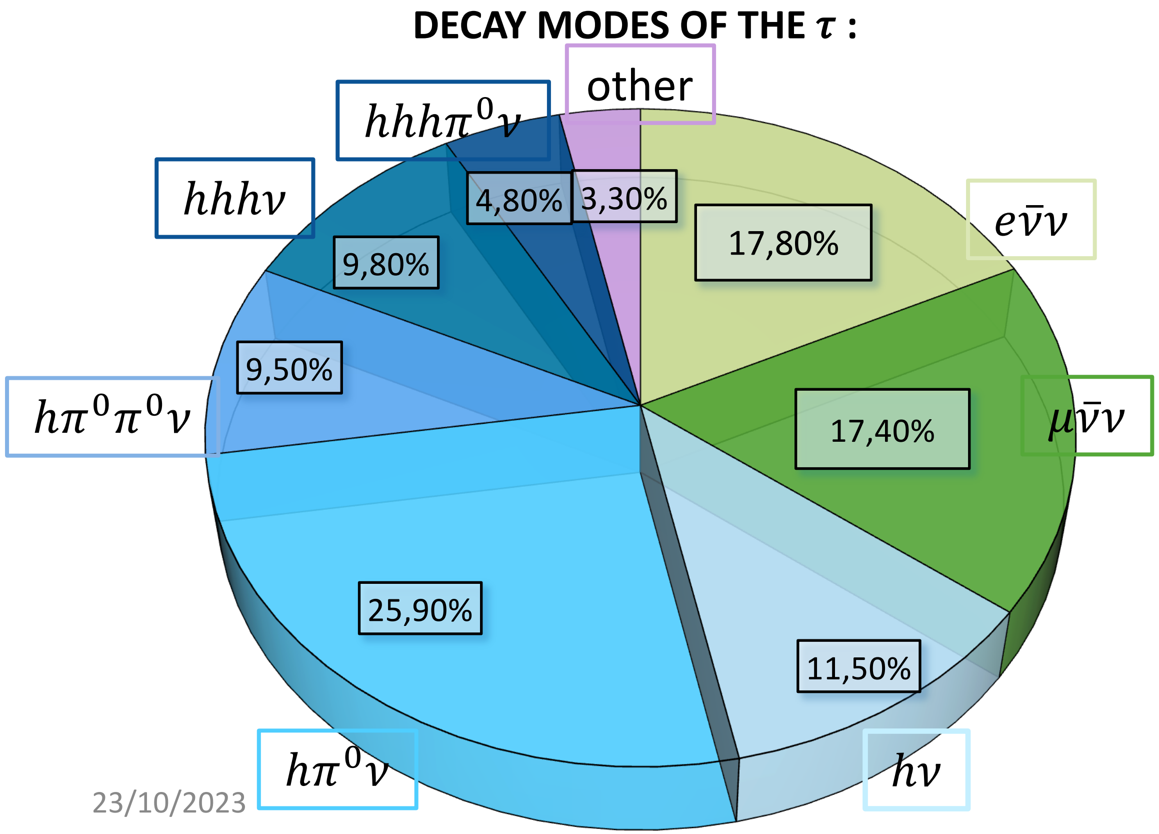
Tau leptons

- Short lifetime ↔ short decay length ~ 1.5mm (E ~30 GeV) => tau leptons cannot be directly detected, they have to be **reconstructed from their decay products**.

- **Tau decay products**:







Leptonic decay :
- BR = 35,2%
- Easier to reconstruct

Hadronic decay (τ_h):
- BR = 64,8%
- More challenging

=> focus for today's talk

Tau identification and reconstruction in CMS

Particle Flow (PF) **algorithm reconstructs** the charged and neutral hadrons, photons, μ, and e using information from the CMS sub detectors.

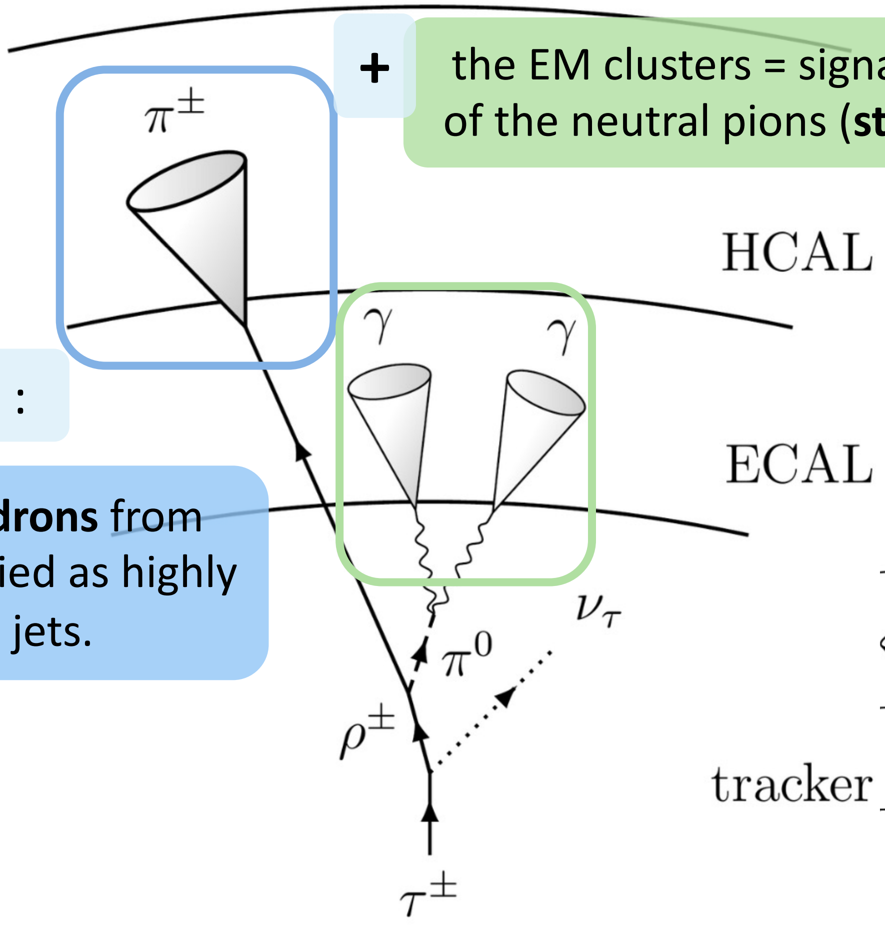
The **Hadron-Plus-Strips** algorithm is used to reconstruct τ_h.

Then, combining :

The charged **hadrons** from tau decays identified as highly collimated jets.

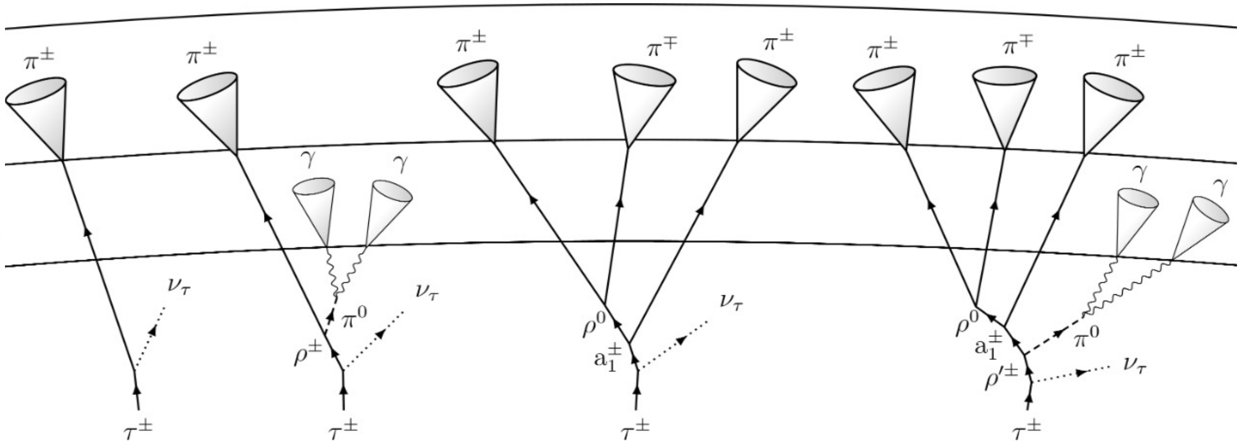
+ the EM clusters = signature of the neutral pions (**strips**).

=>



$\tau^\pm \to \rho^\pm \nu_\tau \to \pi^\pm \pi^0 \nu_\tau$

The algorithm recognizes a τ_h if the combination matches to **a tau decay mode (DM).** Especially:

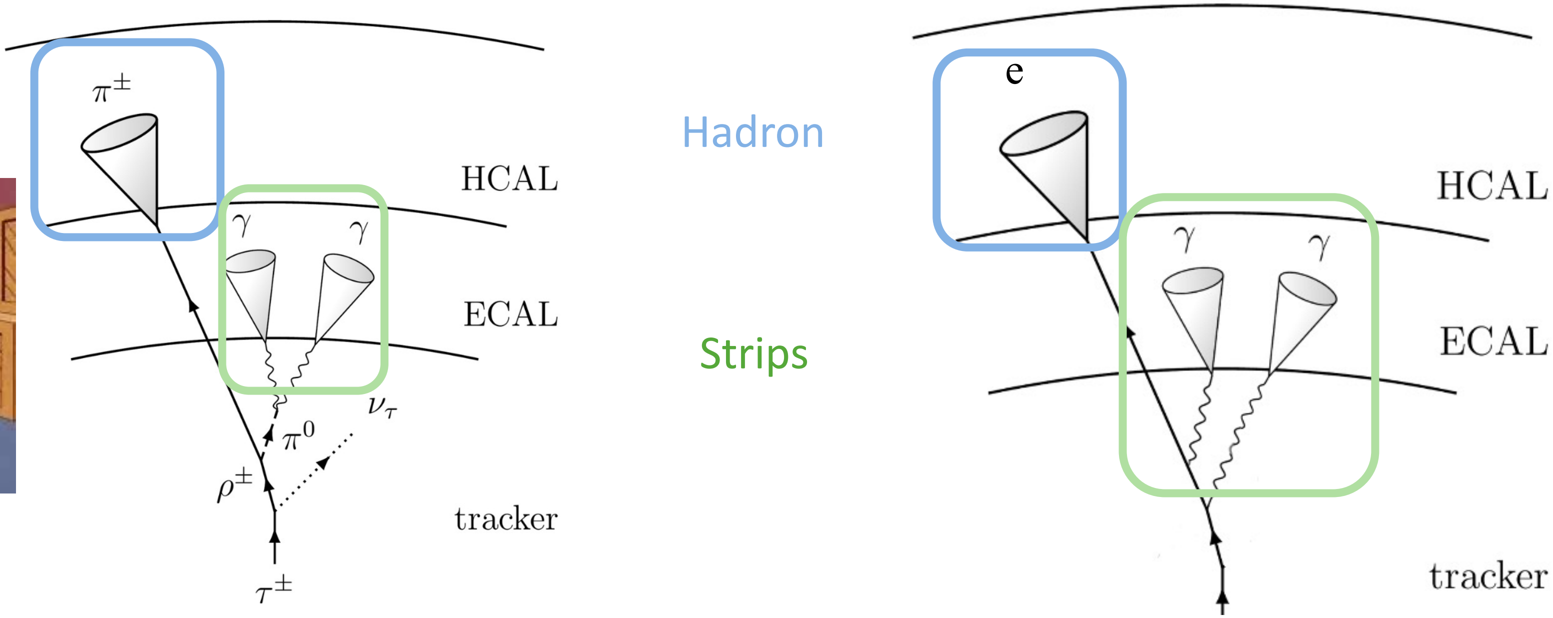


$\tau^- \to h^- \upsilon_\tau$	$\tau^- \to h^- \pi^0 \upsilon_\tau$	$\tau^- \to h^- h^+ h^- \upsilon_\tau$	$\tau^- \to h^- h^+ h^- \pi^0 \upsilon_\tau$

"Fake taus"

The HPS algorithm can misidentify some objects as hadronic tau. These objects are referred to as **'fake taus'.** For example:

- Jets constituted of collimated quarks and gluons can be mistaken for hadronic taus.
- Muons especially in the $\tau^- \to h^- \nu_\tau$ DM.
- Electrons can emit photons and looks like a π^0 decay or can be misidentified as hadron.

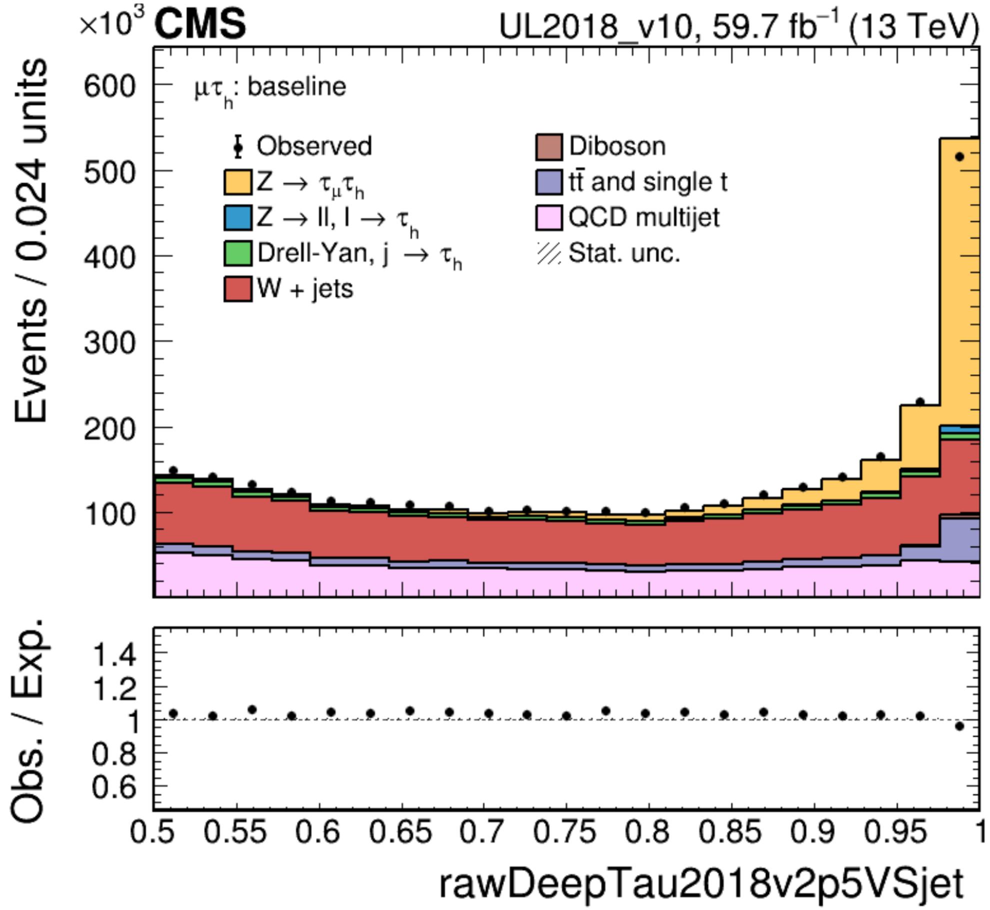


real hadronic τ $e \to \tau_h$ *fake*

A **Neural Network** (DeepTau) is used to reduce the rate of electrons, muons and jets misidentified as τ_h.

DeepTau discriminator

- DeepTau = neural network algorithm used to classify τ_h candidates as jets, genuine τ_h, electrons or muons.
- One output scores for each candidates. For example for jets :



- DeepTau score close to 1 :
  - **real τ_h /fake τ_h ratio is higher.**
  - **BUT** worse Data/Simulation agreement.

- To improve **real τ_h /fake τ_h ratio**, a cut is applied on the DeepTau score. (Working Points)

- To compensate the **Data/Simulation disagreement**, a correction factor is applied: the **identification scale factor.**

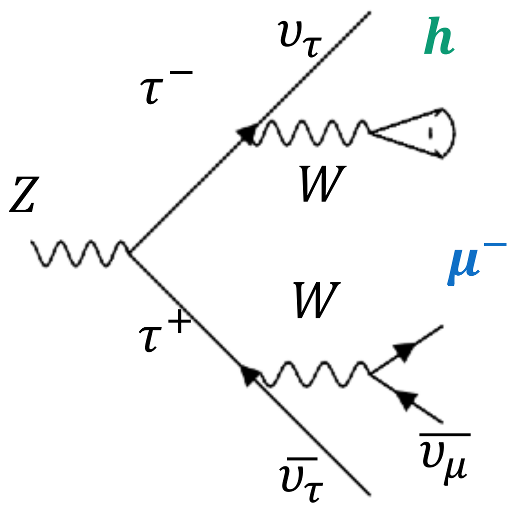
- But other calibration are needed.

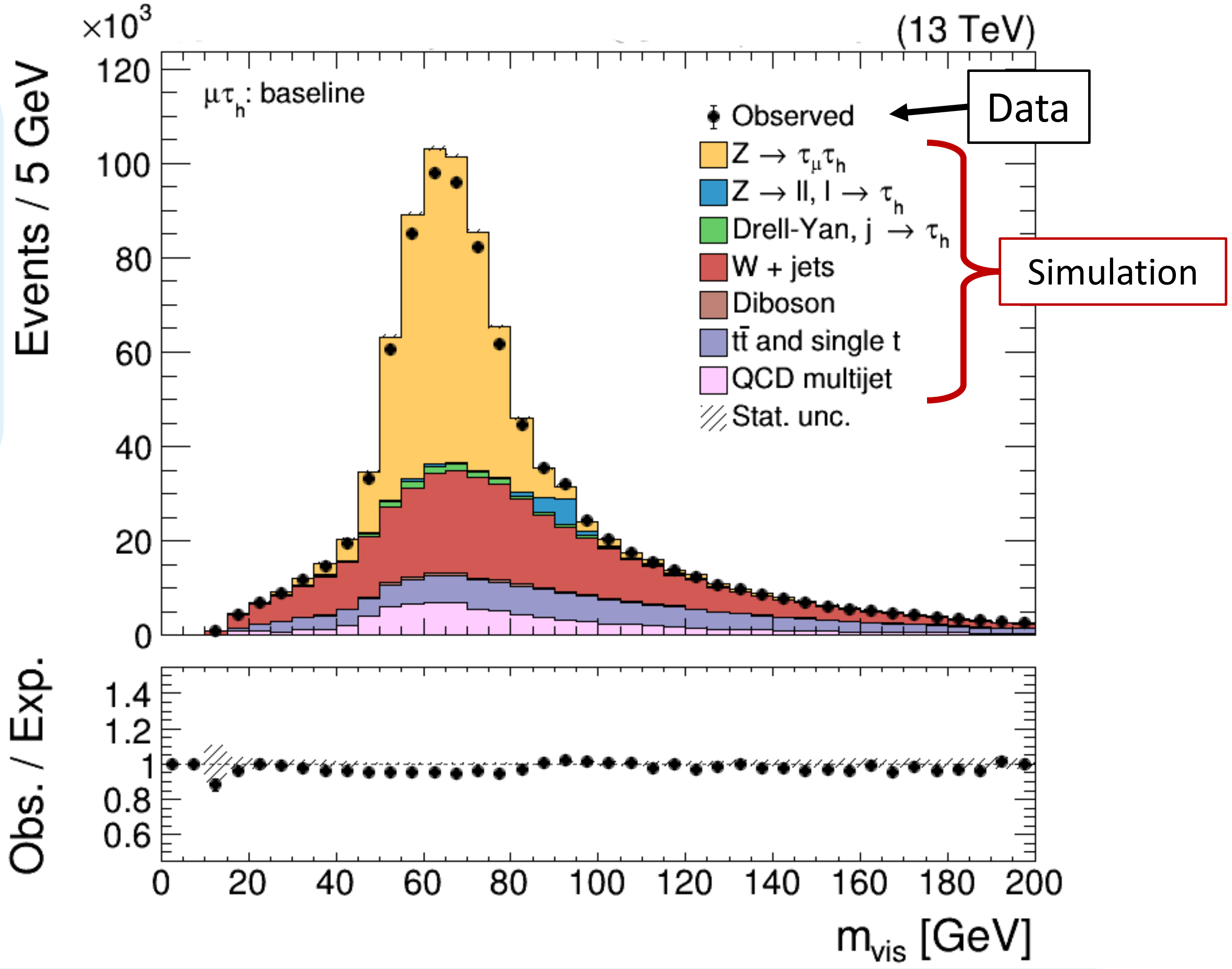
Calibration

Calibration = correct the **differences between data and simulation** by applying **correction factors**.

Use "**Tag & Prob**" method in $Z \rightarrow \tau_\mu \tau_h$ channel:

- A well identified object is used as the **tag** object (here is the well measured and identified muon)
- The **probe** object is the one to calibrate, here the τ_h.





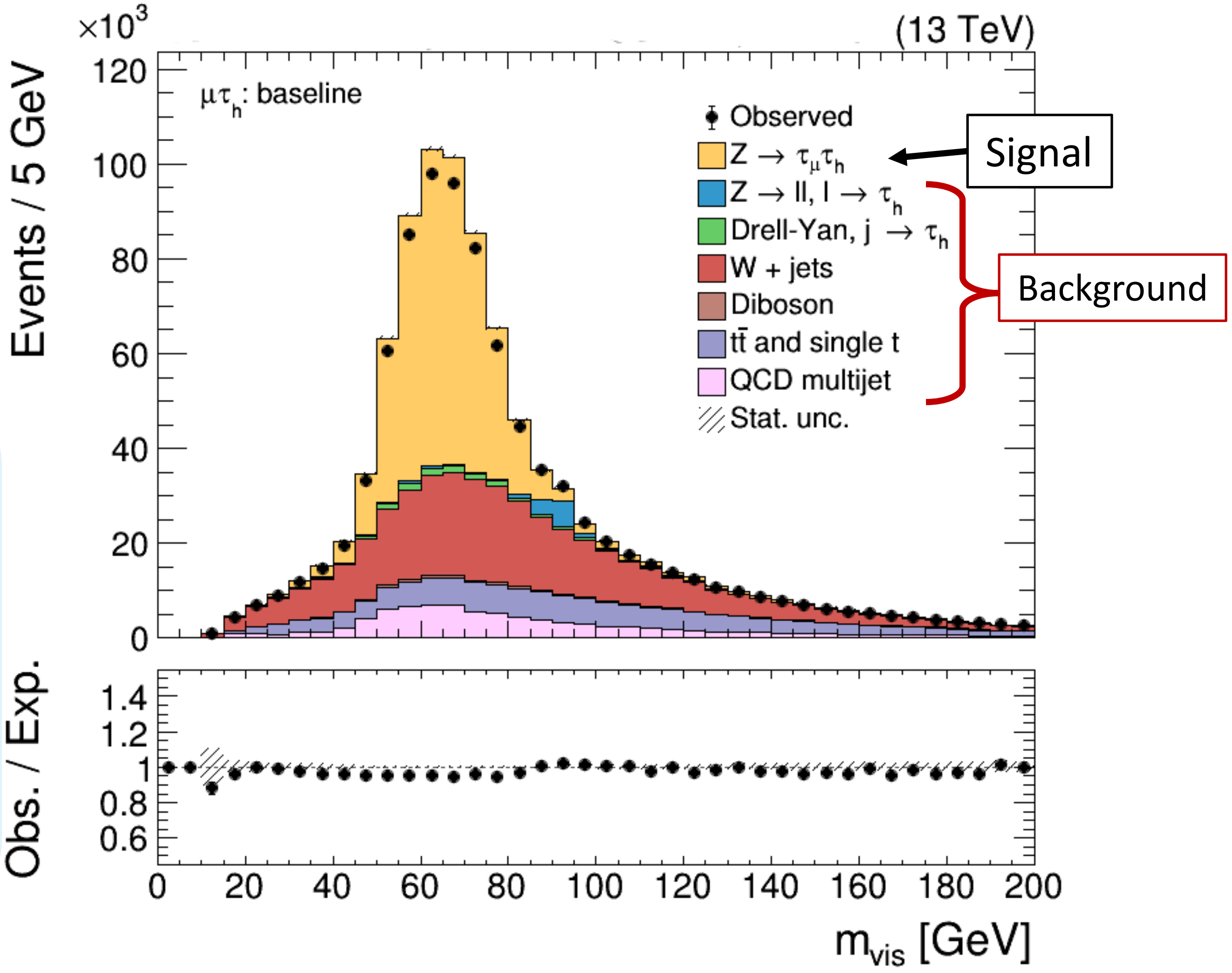
The correction factors are measured by fitting the m_{vis}.

Calibration

$$m_{vis} = \sqrt{(\sum E_h + E_\mu)^2 - (\sum \overrightarrow{p_h} + \overrightarrow{p_\mu})^2}$$

m_{vis} distribution:

- Used to **differenciate signal** form **background**.
- Is **sensitive** to the correction factors.
- $m_{vis} \neq m_Z$ because of the neutrinos.
- The correction factors only applied to the **signal region ($Z \rightarrow \tau_\mu \tau_h$)** $\neq$ background is mainly composed of fake τ_h.



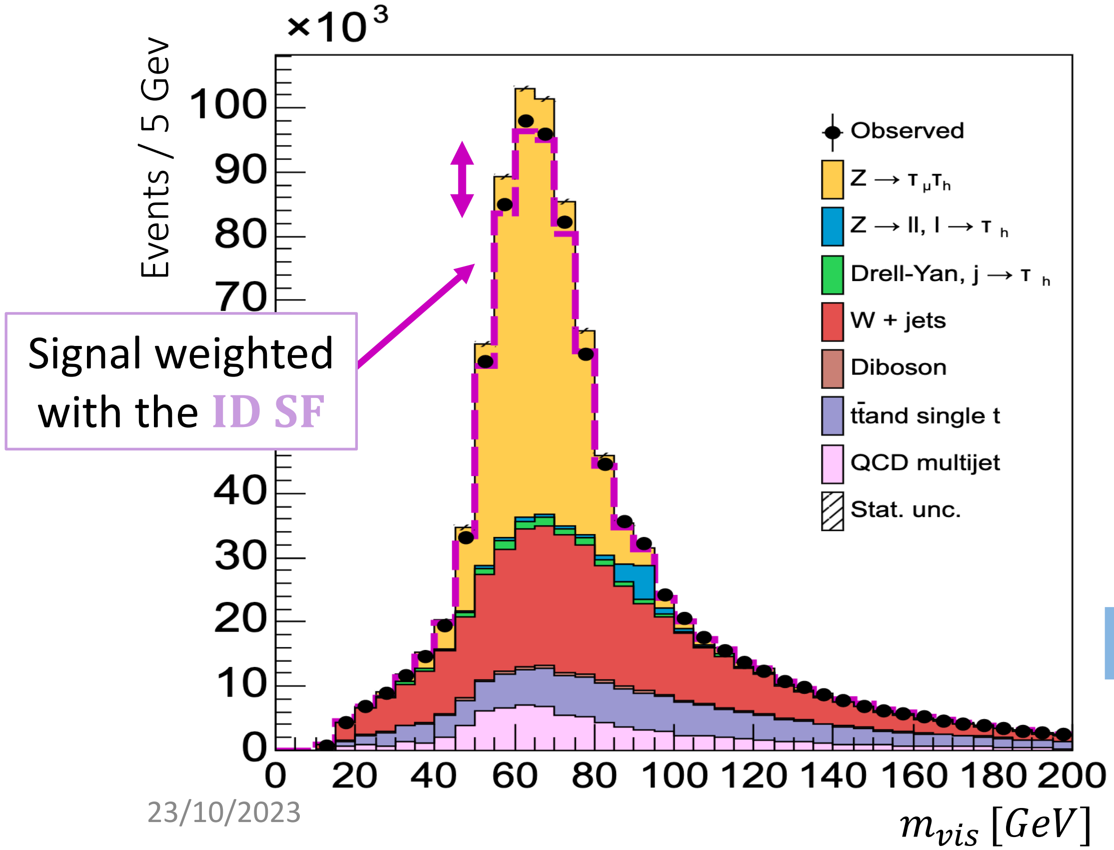
Correction factors

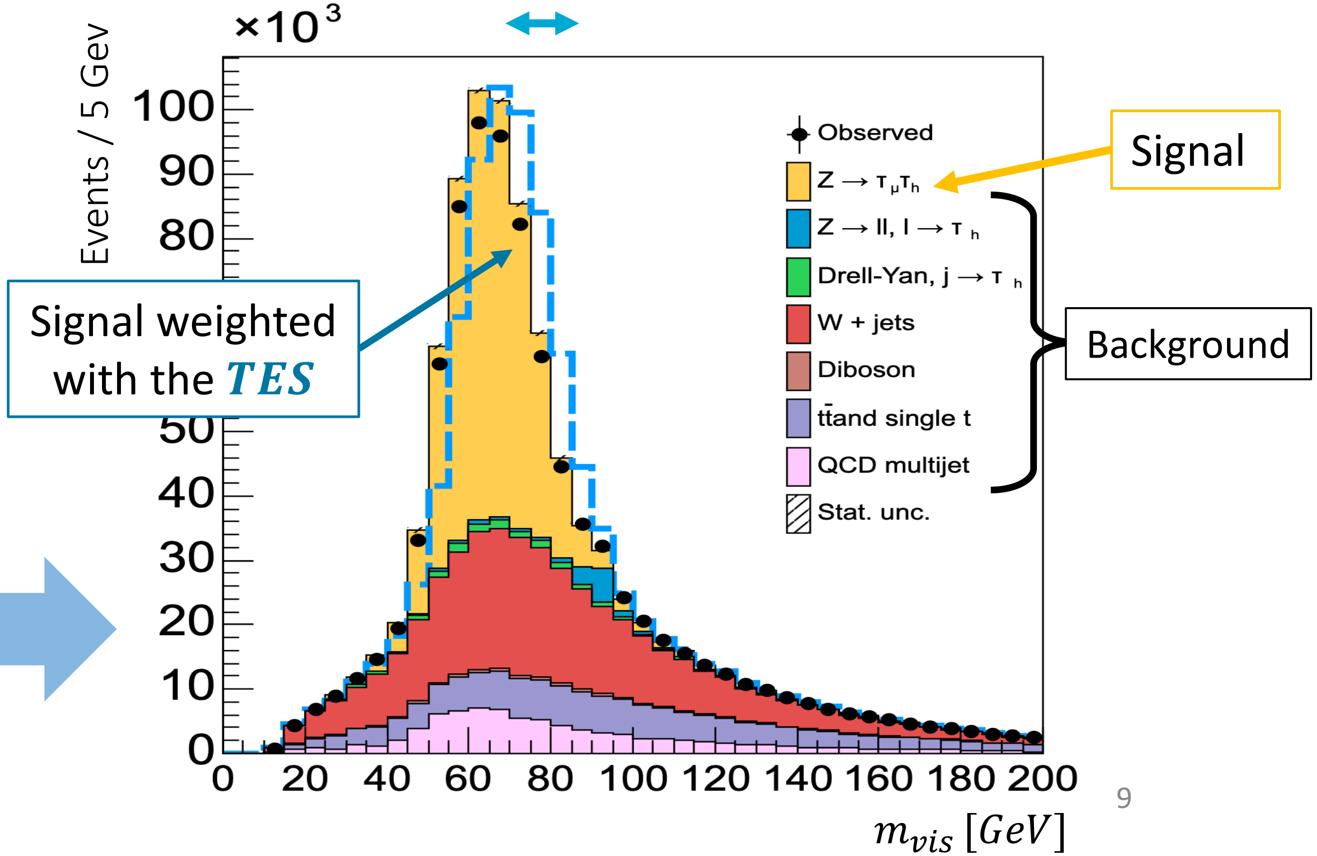
Identification Scale Factor of τ_h (ID SF)

- Corrects differences in identification efficiency between the data and simulation.
- Adjusts the **normalisation** of the signal.

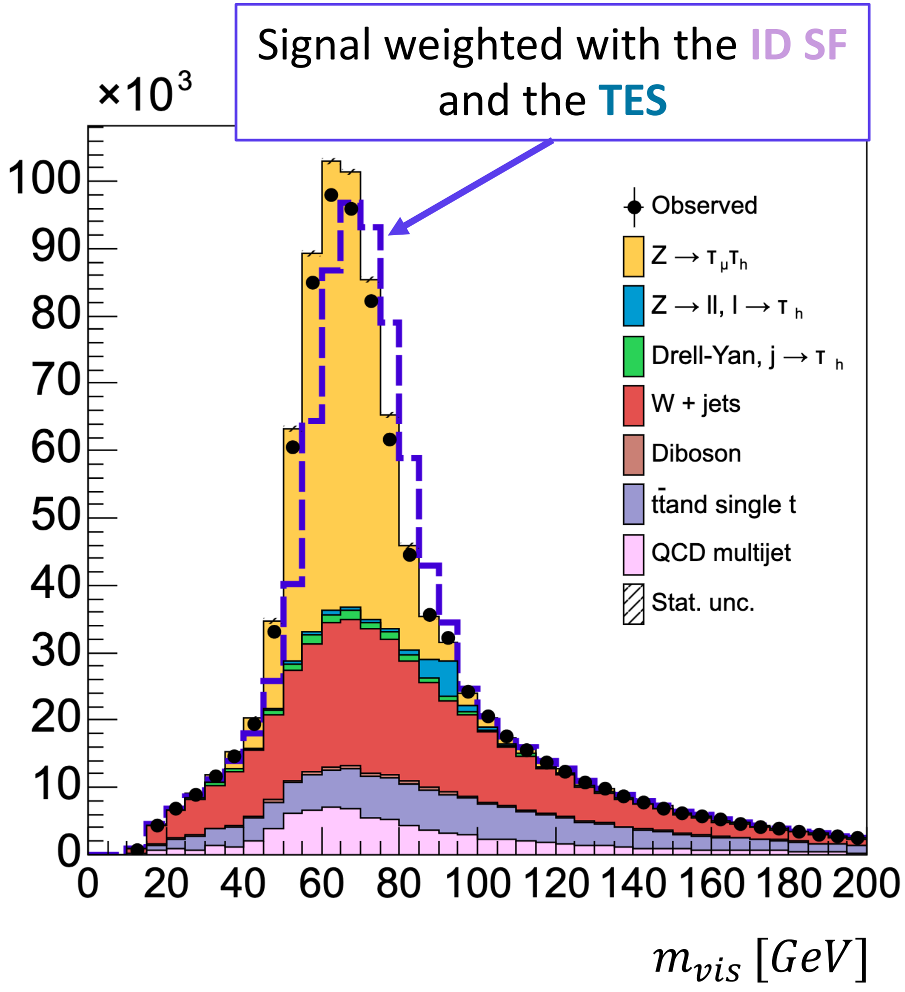
Tau Energy Scale of τ_h (TES)

- TES correction $= \frac{E_{data}}{E_{Simulation}}$
- The TES correction affects the **shape** and **position** of the signal.





Separate fit and combined fit



Previous analysis :
The ID SF and the TES were measured separately.

Likelihood : $\mathcal{L}_1$(ID SF, syst. $uncertainties$)
et $\mathcal{L}_2$(TES, syst. $uncertainties, uncertainty$ ID SF)

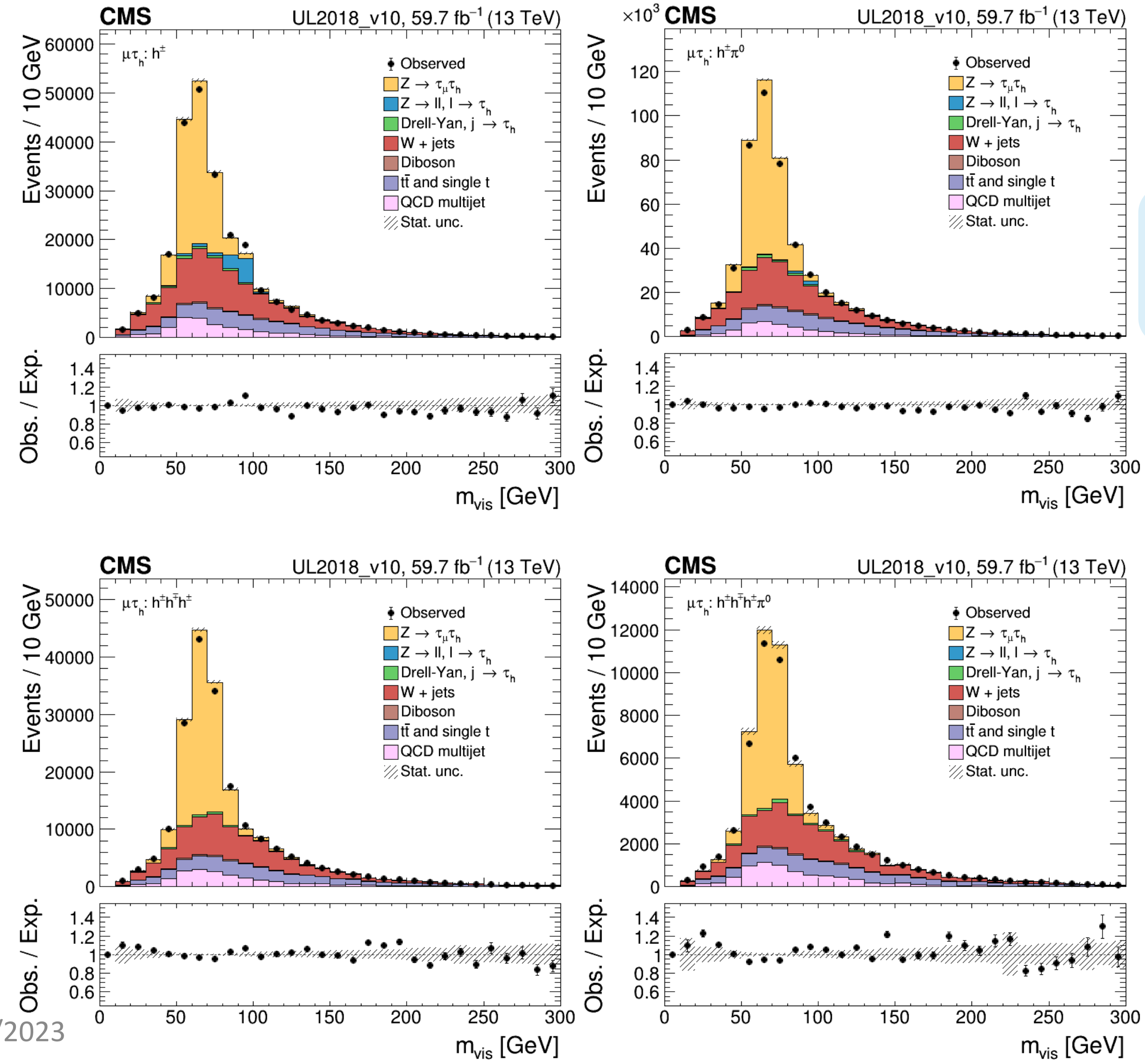
New method :
Simultaneous fit of the TES and of the ID SF

Likelihood : $\mathcal{L}$ (ID SF,TES, syst. $uncertainties$

Goal :
↪ Avoid double counting of the uncertainties.
↪ Take in account the possible correlation between the two correction factors.

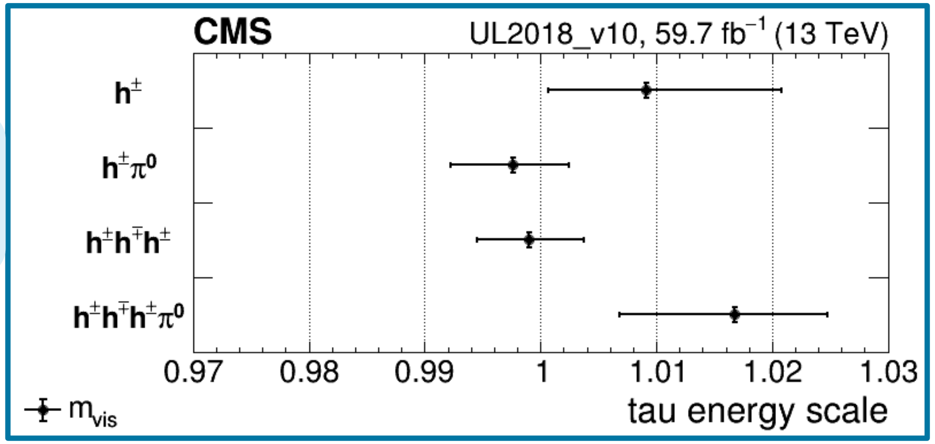
Measurement of the correction factors for DM regions

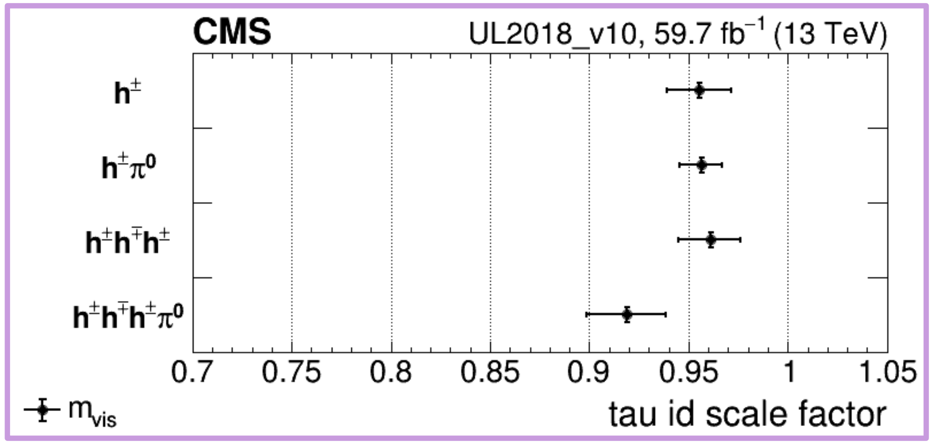
The **fit** is first perform by **DM** regions. One ID SF and one TES are defined for each region to take in account the **differents kinematics** and the **effects** related to the **number of particles to reconstruct in the final state.**



~~Préfou~~
Préfit plots

TES

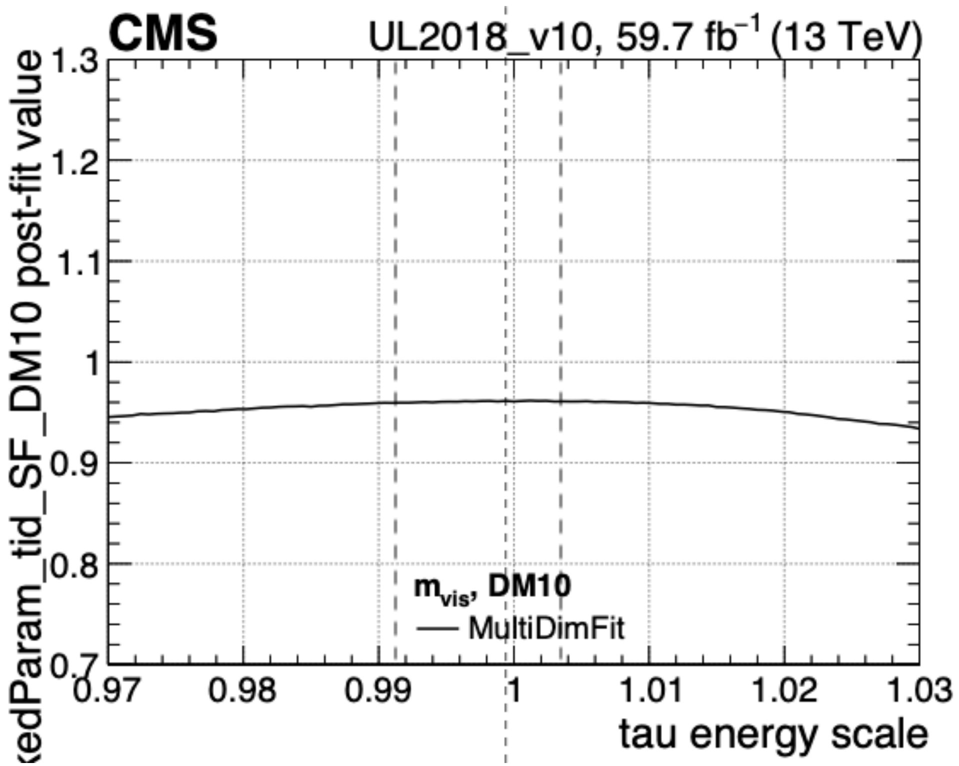




ID SF

Correlation in DM regions

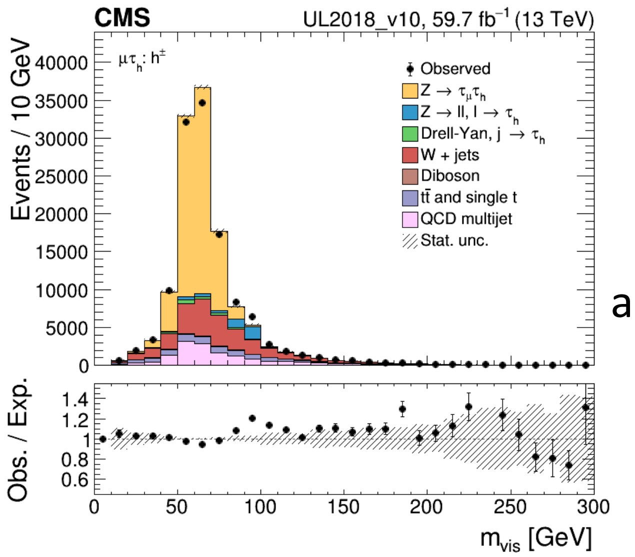
Example : the TES and th ID SF for $\tau^- \rightarrow h^- h^+ h^- \upsilon_\tau$ channel



$\Rightarrow$ Small correlation between the ID SF and the TES defined by DM.

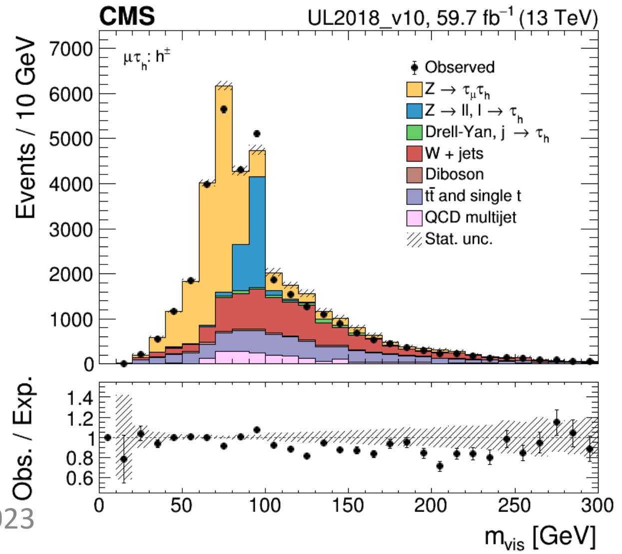
Measurement of the correction factors for DM and $p_{T(\tau_h)}$ regions

The **fit** is then performed by **DM and $p_{T(\tau_h)}$** regions. The splitting in function of the τ_h takes in account the different kinematic between events at low $p_{T(\tau_h)}$ and the ones at higher $p_{T(\tau_h)}$.

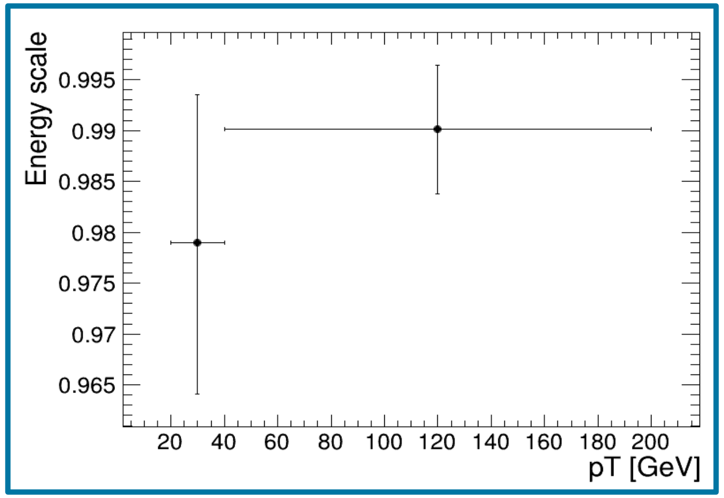


Example:

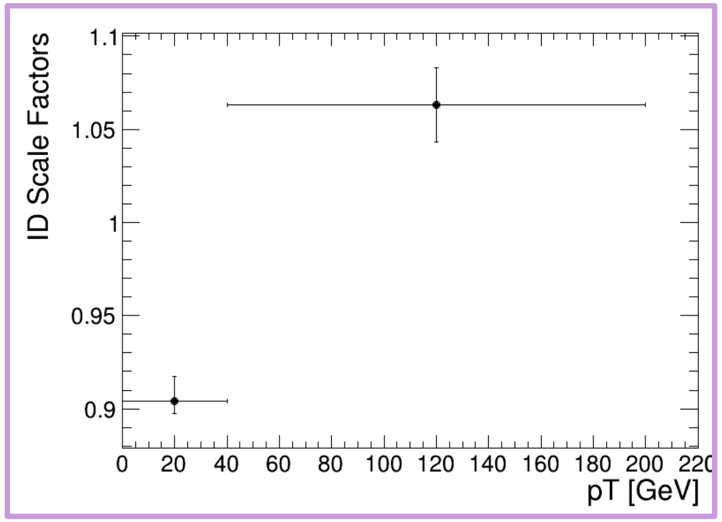
$\tau^- \to h^- \upsilon_\tau$ and $(20 < p_{T(\tau_h)} < 40)$ GeV



$\tau^- \to h^- \upsilon_\tau$ and $(40 < p_{T(\tau_h)} < 200)$ GeV



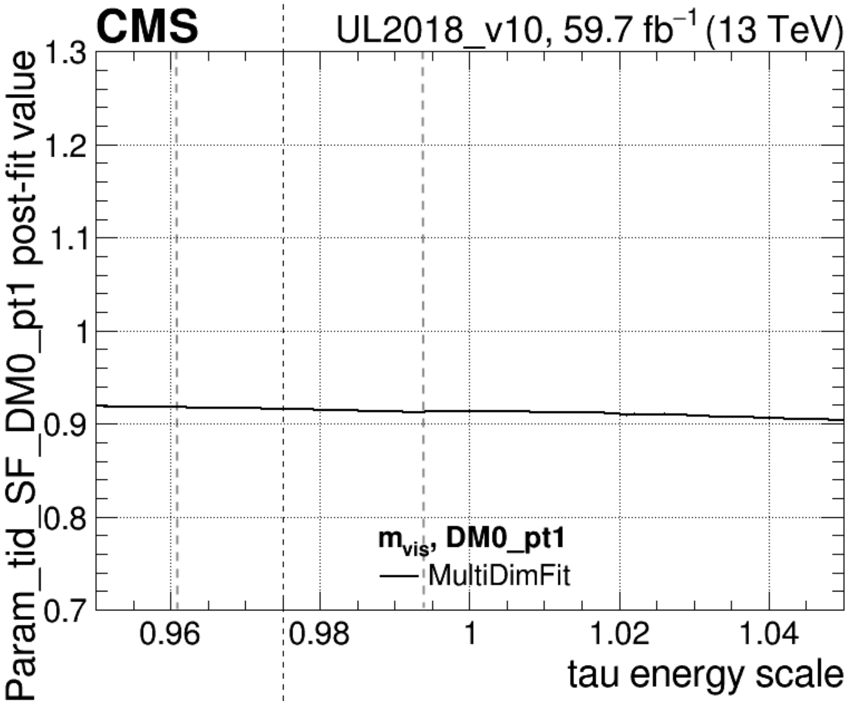
TES



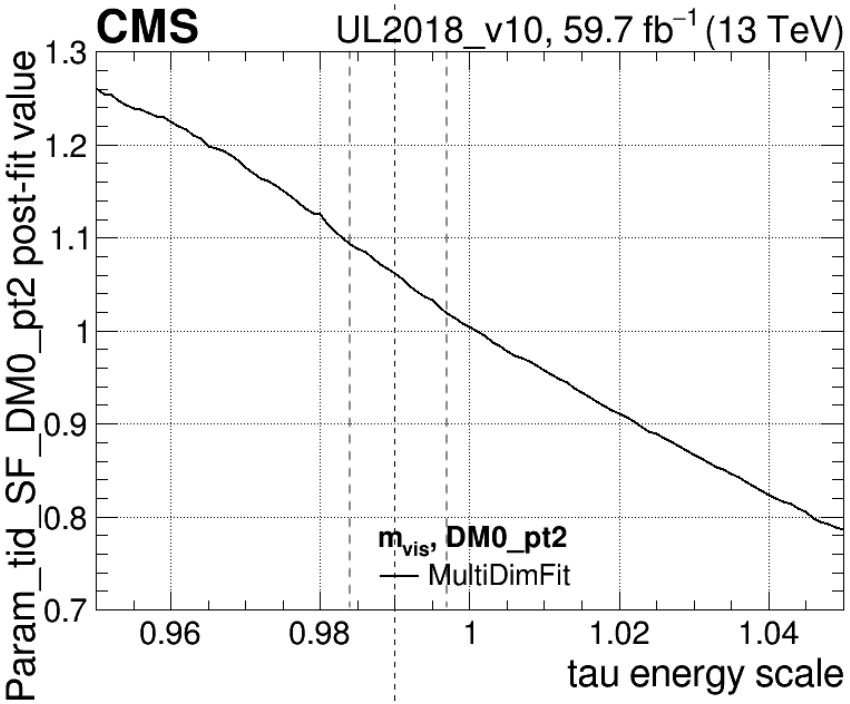
ID SF

Correlation of the two correction factors

$20 < p_T < 40$ GeV (low p_T regions)



$40 < p_T < 200$ GeV (high p_T regions)



⟹ Low correlation between the **ID SF** and the **TES for low p_T** regions.

⟹ Correlation between the ID SF and the TES is non-negligible for high p_T.
⟹ Necessity of the combined fit.

Summary on the combined fit

- Hadronic tau leptons reconstruction is challenging.
  - **Hadronic taus are reconstructed in their different DM**.
  - **Fake tau** rejection improved by using a **neural network (DeepTau).**
  - The **calibration** is important step where **difference between data and simulation** are corrected.

- A new method for the measurements of the two correction factors has been presented. It is a **combined fit** of the **ID SF** and the **TES**.

- For TES and ID SF measurements performed by **DM** :
  - **Low correlation** between the ID SF and the TES for **DM**.

- For TES and ID SF measurements performed **by DM and p_T** :
  - Limited to split only in two p_T regions because of the **statistic** of the **simulation**.
  - **Non-negligible correlation** between the ID SF and the TES at **high p_T** $\Longrightarrow$ necessity of a combined fit.

What's next?

- **Tests on Run 3** data.
- **The statistic of the Simulation** will be **increased** by adding new samples for Run 2 (stitching) $\Longrightarrow$ Impact on the measurement on the correction factors ?

JRJC 2023

Thanks for your attention

Backup

Application: example of analysis using the ID SF and the TES

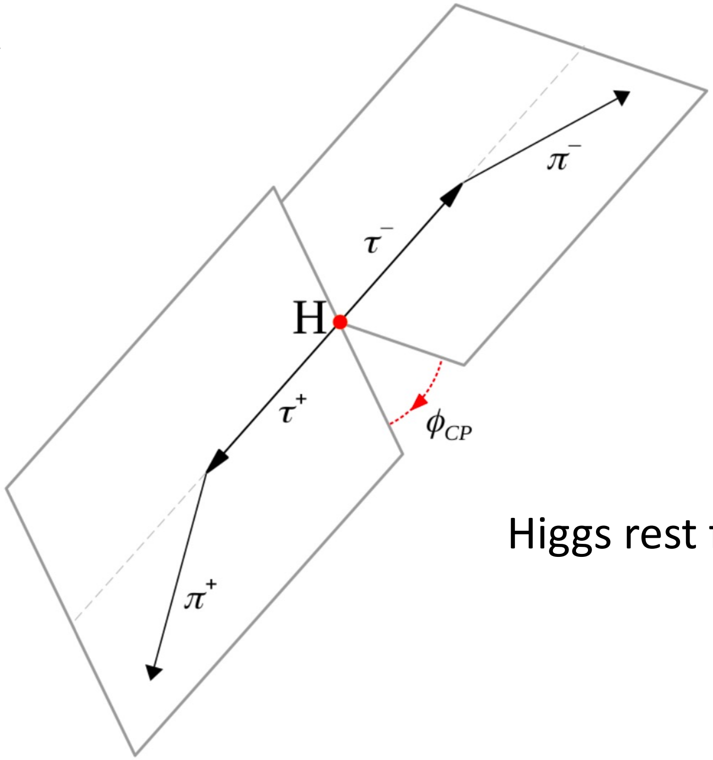
Search for Charge-Parity (CP) violation in Higgs boson decays into taus leptons

- **Standard model (SM) prediction** : H is even under CP inversion => H sould have CP even interaction with SM particles
- **CP-violation** in the Higgs couplings can be probe via $H \rightarrow \tau\tau$ decay
- Each fermionic interaction can be decomposed into a CP-even and a CP-odd **coupling to the Higgs boson** :

$$L_Y = \frac{m_\tau}{v} \bar{\tau}(\kappa_\tau + i\gamma^5 \widetilde{\kappa_\tau})\tau H$$

- **CP mixing** is encoded in a mixing angle $\tan(\alpha^{H\tau\tau}) = \widetilde{\kappa_\tau}/\kappa_\tau$

- **CP sensitive observable** in tau decay is $\boldsymbol{\phi_{CP}} \propto \alpha^{H\tau\tau}$ the angle between the tau decay planes is accessible through visible decay products.



Higgs rest frame

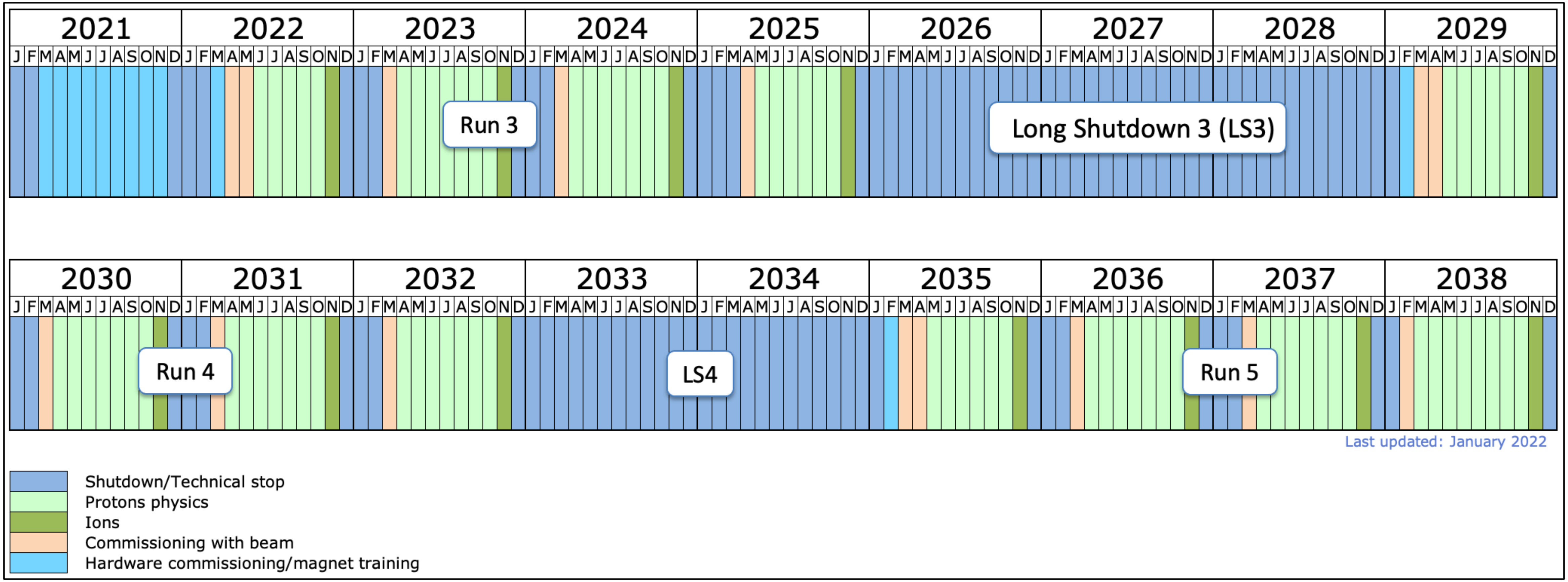
Incertitudes systématiques

Incertitudes systématiques

DY = Drell-Yan MC (ZTT + ZL + ZJ)
ZTT = DY, real τ_h
ZL = DY, $\ell \to \tau_h$ fake
ZJ = DY, $j \to \tau_h$ fake
ttbar = TTT + TTL + TTJ

nuissance parameter	distribution	uncertainty	applied to
luminosity	lnN	±2.5%	all, except QCD
muon efficiency	lnN	±2%	all, except QCD
tau ID	**shape**	**from recommendation**	ZTT, TTT
DY cross section	lnN	±2%	DY
ttbar cross section	lnN	±6%	ttbar
single top cross section	lnN	±5%	single top
diboson cross section	lnN	±5%	diboson
W + jets normalization	lnN	**±8%**	WJ
QCD normalization	lnN	**±10%**	QCD
$j \to \tau_h$ fake rate	lnN	**±15%**	ZJ, WJ, QCD, TTJ, STJ
$j \to \tau_h$ fake energy scale	shape	**±5% on $j \to \tau_h$ energy**	ZJ, W, TTJ
$\ell \to \tau_h$ fake rate	**shape**	**from recommendation**	ZL, TTL
$\ell \to \tau_h$ fake energy scale	shape	±2% on $\ell \to \tau_h$ energy	ZL, TTL
Z p_T reweighting	shape	apply weight ±10%	DY
bin-by-bin	shape		all

Calendrier du LHC



Méthode de maximum de vraisemblance

For X, a real random variable with a continuous distribution, we aim to estimate the parameter θ. We define the function f such that $f(x;\theta)$ represents the probability that X = x (given θ).
The likelihood of θ given the observations $(x_1, \ldots x_n)$ from an n-sample independently and identically distributed according to a law $f(X;\theta)$ is defined as the number:

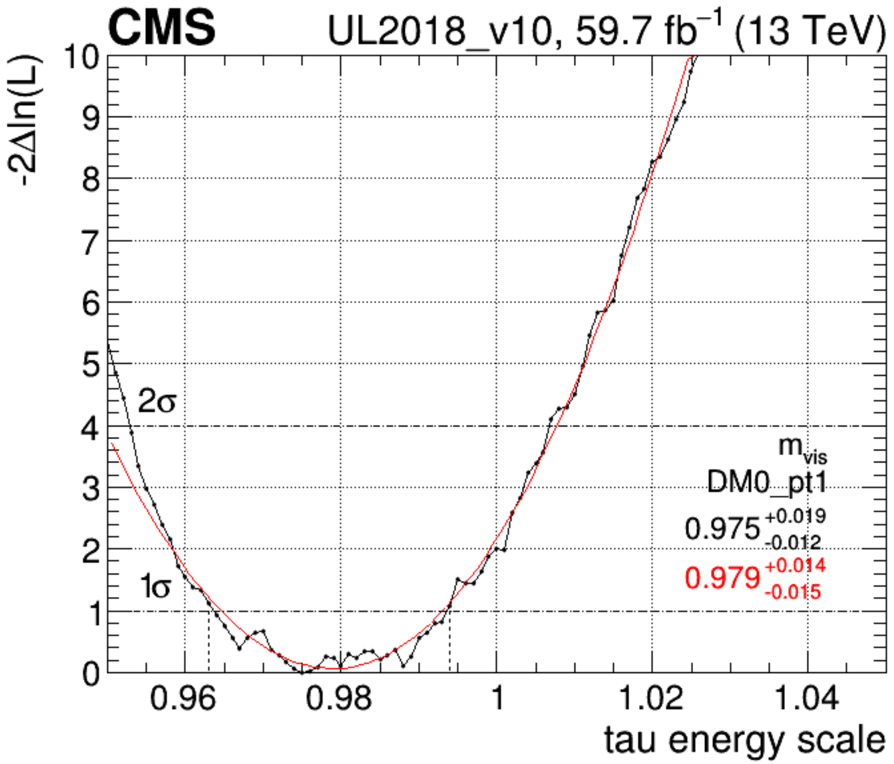
$$L(x_1, \ldots x_n, \theta) = \prod_{i=1}^{n} f(x_i;\ \theta)$$

It's like:

- The probability of measuring x_i given θ
- The distribution of θ using the estimator $\hat{\theta}$

The estimator $\hat{\theta}$ maximizes L for a given sample.

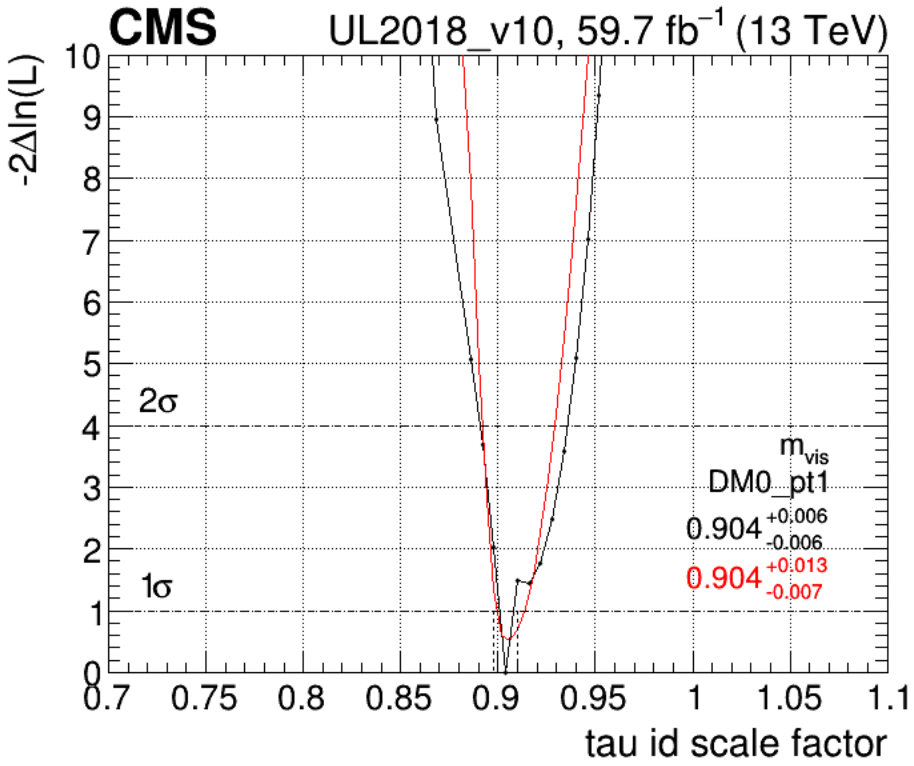
NLL in DM and pt regions



$\tau^- \rightarrow h^- \upsilon_\tau$

$(\mathbf{20} < \boldsymbol{p_{T(\tau_h)}}$ < 40)

GeV

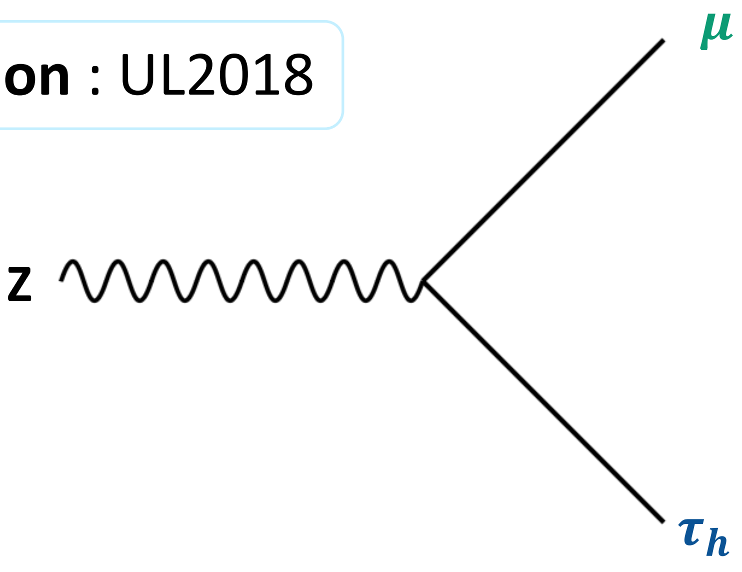


Larger fluctuactions in this case. The fit of the NLL profile with a parabolae is needed.
This is due to the limitied statistic of the Simulation.

Event selection

Goal: select Z → $\tau_\mu \tau_h$ events with $\mu\tau_h$ as final state for the calibration.

Data/ **Simulation** : UL2018



- p_T >25 GeV,
- $|\eta|$<2.4
- $|dxy|$ < 0.045 cm
- $|dz|$ < 0.2 cm medium ID
- relative isolation I_μ < 0.15

- p_T >20 GeV
- $|\eta|$<2.3
- $|dz|$ < 0.2 cm
- DeepTau2018v2p5VSjet: Medium WP
- DeepTau2018v2p5VSmu: Tight WP
- DeepTau2018v2p5VSe: VVLoose WP

Improve the significance $S/\sqrt{S+B}$ and optimization of the fit

- Lepton veto
- Opposite sign $\mu\tau_h$ pair with ΔR > 0.5 and highest μ and τ_h $\mathrm{p_T}$
- metfilter
- cut $m_T(\mu) < 65$ GeV