Performance of muon identification with multivariate techniques at the CMS experiment

Clara Ramón Álvarez on behalf of the CMS Collaboration
EPS-HEP, August 2023 – Hamburg, Germany

Universidad de Oviedo

Introduction and motivation

Improve muon selection and identification (ID) for muons with **medium p_T range** → key in many CMS analysis. Whether those are:

- ❑ Isolated muons form prompt decay of W, Z, H and decay of τ.
- ❑ Real muons from heavy flavour decays.

Multivariate analysis (MVA) methods used to construct **two muon ID** discriminators:

- ❑ **MVA ID**: **General muon selection** against spurious hits in the muon system. We construct an ID more flexible and performant than the cut-based ID [1] (used during Run 2) for muons with p_T between 20 and 120 GeV. ID to be used during Run 3.
- ❑ **Prompt MVA**: **select isolated muons from H/W and τ**. Already used during Run 2, crucial in the observation of $t\bar{t}H$, $t\bar{t}t\bar{t}$ and tZq; the precision measurement of WZ and to increase sensitivity in the search for electroweak production of SUSY.

Input Variables

MVA ID

12 Input variables related to:

- Quality of the track reconstructed in the muons system (stand alone) and in the tracker
- Quality matching between both
- Number of muon station with hits
- Number of good hits in the tracker

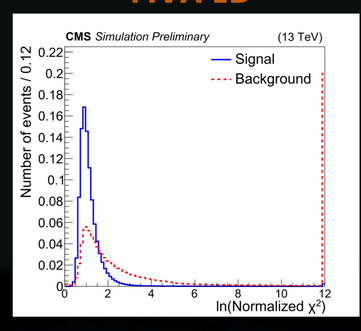
Prompt MVA

12 Input variables related to:

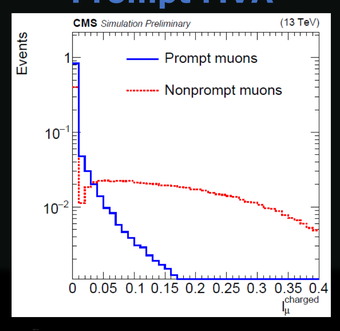
- Isolation of the muon
- Information of the jet reconstructed within the same cone
- Impact parameter (IP)

Example of a input variables in each MVA:

MVA ID



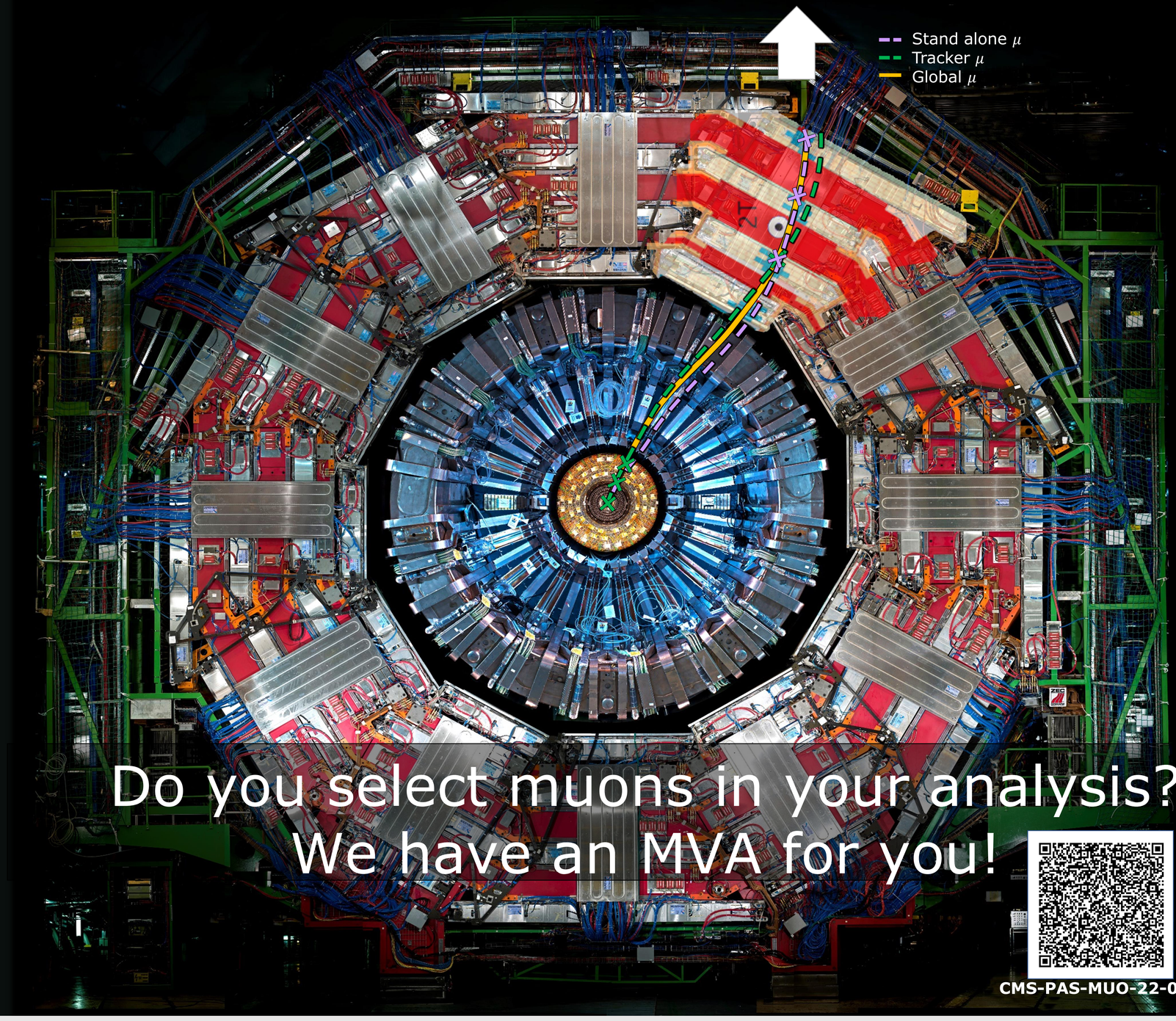
Prompt MVA



Pre-selection and samples

- Used **2018** dataset (59.7 fb^{-1})
- **MC:**
  - $t\bar{t}$ semileptonic for training
  - DY for performance evaluation
- **Pre-selected muons** are loose muons (cut-based ID) with p_T > 10 GeV + Relaxed isolation and IP selection* (for the prompt MVA)
- look for geometrical matching traced back using generation information

*Mini-isolation <0.4 & d_{xy}<0.05 & d_z<0.1 & $d/\sigma_d < 8$

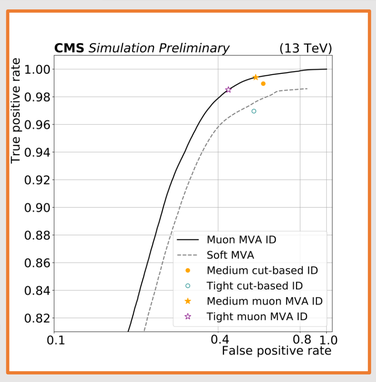


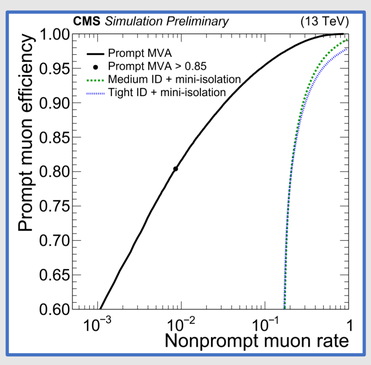
Machine learning models

MVA ID

Random Forest: 200 trees with a maximum depth of 8

Medium working point defined to have same background efficiency as the medium cut-based ID





Prompt MVA:

BDT: 200 trees with a maximum depth of 8

1000 trees with max depth = 4

WP shown optimized for $t\bar{t}H$ measurement

Do you select muons in your analysis?
We have an MVA for you!

CMS-PAS-MUO-22-001

Performance evaluation

Efficiency in data evaluated using **Tag-and-Probe method** to select muons from Z decay:

- Tag: pass Tight WP (cut-based ID)
- Probe: preselection

Background estimation:

MVA ID: estimated using tt MC

$$Bkg.Efficiency = \frac{\text{background loose muon \& pass ID selection}}{\text{background loose muon}}$$

Prompt MVA:

- Nonprompt rate measured in a Multi-jet enriched region
- EWK contributions subtracted using a fit to:

$$m_T^{fix} = \sqrt{2\, p_T^{fix} p_T^{miss} (1 - \cos\Delta\phi)}$$

$$p_T^{fix} = 35 \text{ GeV}$$

Background efficiency: 50% for p_T > 20 GeV

For same background efficiency, systematically higher efficiency than with cut based

A factor of 2 (3) smaller nonprompt rate of the prompt MVA for muons with $|\eta|$ < 1.2 ($|\eta|$ > 1.2)

Summary

MVA ID

- **Higher efficiency than cut-based ID** (for the same background efficiency)
- Continuous score gives more flexibility for analysers
- MVA shown to be more **resilient as a function of PU** than the cut-based ID → **Great news for run 3!**

Prompt MVA

- Efficiency > 80% for p_T > 20 GeV
- Factor of 2-3 reduction of background from non prompt leptons
- Has been a **key element** in several full **Run 2** CMS published results: $t\bar{t}H$, $t\bar{t}t\bar{t}$, tZq, WZ, EWK SUSY…

Ayudas del Programa Severo Ochoa by

Grant PID2020-113341RB-100 funded by

EUROPEAN UNION
European Regional Development Fund

[1] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV", JINST **13** (2018) P06015, arXiv:1804.04528.