Probing BSM Physics with multileptons at CMS

Angira Rastogi
(On behalf of the CMS collaboration)

Dec 14th, 2020
Outline

• The Multilepton ‘tool’
• BSM signals
• Results strategy
• Summary
## Introduction

- A **multi-bin, model-independent** multilepton analysis with full Run2 data.
- All lepton flavors in the mix – prioritizing highest pT light leptons before taus.

<table>
<thead>
<tr>
<th>Exclusive Trilepton channels</th>
<th>Inclusive “Four lepton” channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>• $3\ell - eee, e\mu\mu \ldots$</td>
<td>• $4\ell - eeee, \mu\mu\mu\mu \ldots$</td>
</tr>
<tr>
<td>• $2\ell_1\tau_h - ee\tau, e\mu\tau \ldots$</td>
<td>• $3\ell_1\tau_h - eee\tau, \mu\mu\tau\tau \ldots$</td>
</tr>
<tr>
<td>• $1\ell_2\tau_h - \mu\tau\tau \ldots$</td>
<td>• $2\ell_2\tau_h - \mu\epsilon\tau\tau, ee\tau\tau\tau \ldots$</td>
</tr>
</tbody>
</table>

+ $B$-tags, +MOSSF, +ST, +MET, +MT, +HT, +lepton pt bins

- Explicit interpretations for many new physics models.
Multilepton Analysis - Object & event selections

• **Analysis details**
  - **Electrons**: \(p_T > 10\) GeV, \(|\eta|<2.4\), PF isolation, transverse & longitudinal impact parameter cuts, low b-tagging score, consistent Q-measurement.
  - **Muons**: \(p_T > 10\) GeV, \(|\eta|<2.4\), PF isolation, transverse & longitudinal impact parameter cuts, low b-tagging score.
  - **Taus**: \(p_T > 20\) GeV, \(|\eta|<2.3\), DeepTau identification, longitudinal impact parameter cuts, low b-tagging score.
  - **Jets**: \(p_T > 30\) GeV, AK4 CHS PF jets, b-tagging with DeepCSV algorithm.
  - **MET**: PUPPI algorithm.

• **Analysis strategy**
  - **“Preselection”**: Loosely split, very broad & high statistics bins based on dominant SM backgrounds in all multilepton final states.
  - **Dedicated “Control Regions”**: For dominant SM backgrounds, isolating some regions from the preselection events in data for normalization, fake rate measurement etc.
Our background setup

Prompt backgrounds:

- Prompt leptons come from boson decays ($W/Z/h$) and leptonic decays of taus.
- Estimation via MC.
- Main processes ($WZ$, $ZZ$, $t\bar{t}Z$) are corrected in dedicated control regions.
Our background setup

**Fake backgrounds:**

- Fake leptons come from hadronic activity (e.g. heavy flavor decay, jet misreconstructed as lepton, ...).
- Estimation via a 3D/4D tight-to-loose data driven Matrix Method.
- Measure fake & prompt rates in dedicated CRs (or MC) : rate = N(tight ID) / N(loose ID)
- Main SM processes producing fakes in multilepton signatures: DY & ttbar
- A weighted sum of the fake rates are used according to approximate background composition:
  \[
  FR_{\text{final}} = \frac{\text{DY}}{\text{FR}_{\text{DY}}} \cdot FR_{\text{DY}} + (1 - \frac{\text{DY}}{\text{FR}_{\text{DY}}}) \cdot FR_{\text{ttbar}}
  \]

**Typical parametrizations for rates:**

1. Low/Medium/High lepton pT
2. Barrel/Endcap η-regions of detector
3. Additional split in 1-prong & 3-prongs for taus.
4. **Recoil corrections** ...

*Correlation between jet PT, lepton PT & lepton isolation.*
Our background setup

Fake backgrounds:

- **3L MisID CR**
  - (OnZ MET<100 GeV)

- **2L1T MisID CR**
  - (OnZ MET<100 GeV)

**CMS**

- Data
- WZ
- ZZ
- tZ
- MisID
- Conversion
- Rare
- Uncertainty

**137 fb⁻¹ (13 TeV)**

**Run2 2L1T**

- Data
- MisID
- WZ
- ZZ
- ZG
- ttZ
- Rare
- VVV
- Conv
- ttW
- tbarFake

**Work in progress**

(Stat. uncertainty only)
Search for BSM physics: Vector-like taus

- Two scenarios: SU(2)\_L singlet (\(\tau'\)) and mass-degenerate doublet (\(\tau', \nu'\_\tau\)).

- First CMS constraints on Doublet VLL in
  - Analyzed 2016+17 data in the 3\(\ell\), 4\(\ell\) & 2\(\ell\)1\(\tau\) final state.
  - Excluded vector-like taus in the mass range 120-790 GeV.

An example production and decay diagram
type-III Seesaw

- SU(2) Triplet; two charged ‘dirac’ and one neutral ‘majorana’ fermions ($\Sigma^\pm, \Sigma^0$), heavy and mass-degenerate.

- 27 different production and decay modes.

- Best CMS constraints from JHEP 03 (2020) 051, excluding heavy fermions below 880 GeV in the flavor democratic scenario.

- CMS result Phys. Rev. Lett. 119, 221802 (2017) also presented exclusions at various combinations of branching fractions to lepton flavors.

An example production and decay diagram
What will be our results?

Model-independent ‘Data tables’:

- Extending the data versus SM consensus from various multilepton channels to the most inclusive table binning.
- Separate CRs (SM dominant) from high-sensitive SRs.
- Effective use of kinematic variables such as $N_{\text{B-tags}}$, MOSSF, MT, HT, LT, MET, lepton pTs etc.
- Finally, plotting ST in each table bin and providing as a legacy result.

We will give SM backgrounds and data in ST bins (total ~$O(10^3)$) for selections such as -

1. 3L OnZ, high MET, high HT, $N_{\text{jets}} = 0$

2. 2L1T AboveZ low MET, low HT, high tau pT, $N_{\text{jets}} > 1$ ... and so on.
Signals with MVA

• Discriminate signal against dominant SM processes such as DY, WZ, ttbar, ZZ etc.
  – Multiclassifier neural network (classes: Signal, WZ, ZZ, DY, tt).

• Perform mass- or channel- inclusive trainings, to minimize the heterogeneity.
  – Two distinct networks trained: high mass and low mass – to maximize sensitivity.

• Input variables:
  – Kinematic variables + model-specific (total 29): Lepton pTs, Mass combinations, pair-wise pT, transverse masses, min and max DeltaPhis with leptons and MET (expands what is humanly achievable in the data tables).
  – + Event categories (total 19): 4LDoubleOnZ, or 3LAboveZ, or 2L1TOnZ and so on.

  Input variable modeling is excellent in CRs.

• Metric for performance is ROC curve, evaluated for Signal vs primary backgrounds(WZ,ZZ,MisID) and expected limits.
MVA approach

• Application strategy:
  – Channel-wise, separately for each signal.
  – Define slices of MVA in each channel as orthogonal signal regions.
  – Produce the distribution of kinematic variables in each MVA region (e.g. dilepton mass in $tt\phi$, LT in VLL & ST in Seesaw).

MVA adds signal specificity to the generalized binning approach.

$\text{Total bins} = (x+y+z) \times 5$; given in the limit calculation.
Preselection look – Full Run 2

Stat. uncertainty only

CR

Preselection regions

Angira Rastogi, DAE HEP 2020
Summary

- An inclusive, multibin prompt multilepton analysis with full Run2 data luminosity is currently in the works from CMS.

- Various tau-enriched channels are added, to enhance sensitivity for models coupling primarily to the third generation of leptons.

- Model-independent data tables for full Run 2 will be provided as legacy results.

- Explicit interpretations for fermions above Z-mass (type-III Seesaw, Doublet & Singlet VLL) in multilepton final states will be done.

- Various MVA approaches are also being explored to aid the signal vs background discrimination.

- CMS will have a result for Moriond 2021, so stay-tuned for the final results!

Thank you!