Section 1

Investigation: Factor of 2 Discrepancy in MuTau Analysis

Simple Slide Presentation

Slide 1: The Problem

What We Observed

- **Data events:** ~**1,700** (expected ~3,400)
- **QCD events:** ~**234** (expected ~468)
- Both at approximately **half** their expected values

Temporary Solution

```
weight = 2.0  # Compensating for unknown factor
w_qcd = 2.0
```

The Question

Why are both Data and QCD at ~50% of expected?

Slide 2: The Suspected Cause

Processing Method Changed

Old Method (DY/ttjets):

`NanoAOD → Cuts → Merge → Add artificial pileup protons (merge_pp_mutau.py, weight=0.13)`

100% of events get artificial protons

New Method (Data/QCD):

`NanoAOD → Fase0 (save real proton vars) → Fase1 (cuts + proton filte`

Only ~16.9% of events have real protons

Expected Impact

Switching from 100% $\rightarrow$ 16.9% should give **6x reduction**, not 2x!

Something else was wrong...

Slide 3: The Investigation

Diagnostic: Apply Cuts Sequentially

Ran diagnostic on one fase0 file (1M events):

```
Starting events:          1,023,514
After Lumi + HLT:         1,023,514
Has protons:                172,895 (16.9%)
+ Muon ID:                  138,903 (80% pass)
+ TAU ID:                        65 (99.95% REJECTED!)
+ pT cuts:                        1
+ Eta cuts:                       1

Final: 1 Data event, 0 QCD events
```

The Smoking Gun

Tau ID cuts rejected 99.95% of events!
This is NOT normal.

Slide 4: Root Cause #1 - Fase0 Has No Cuts

In fase0_qcd_data.py (lines 177-182):

```
# ALL QUALITY CUTS ARE COMMENTED OUT:
#    .Filter("Muon_mvaId[0] >= 3", "Muon ID (>= Medium)")
#    .Filter("Tau_idDeepTau2017v2p1VSjet[0] >= 63", "Tau VSjet")
#    .Filter("Tau_idDeepTau2017v2p1VSe[0]   >= 7",  "Tau VSe")
#    .Filter("Tau_idDeepTau2017v2p1VSmu[0]  >= 1",  "Tau VSmu")
#    .Filter("Muon_pt[0] > 35. && Tau_pt[0] > 100.", "pT cuts")
```

Result

Fase0 saves **taus of ANY quality** (including tau_id=0)

Then Fase1 Applies Strict Cuts:

```
.Filter("tau_id1 > 63", "Tau VSjet")   # Requires VTight
.Filter("tau_id2 > 7",  "Tau VSe")     # Requires Loose
.Filter("tau_id3 > 1",  "Tau VSmu")    # Requires Loose
```

Almost everything is rejected!

Slide 5: Root Cause #2 - VSmu Mystery

Tau ID Distribution Analysis

Ran diagnostic on tau ID values in fase0:

Discriminator	Threshold	Events Passing	Acceptance
tau_id1 (VSjet)	> 63	852,802	**83.3%**
tau_id2 (VSe)	> 7	870,796	**85.1%**
tau_id3 (VSmu)	> 1	22,479	**2.2%**
Combined	All three	1,197	**0.12%**

The Bottleneck

Only 3.09% of taus have ANY VSmu score > 0!
This is extremely suspicious.

Slide 6: Understanding VSmu

What is Tau VSmu?

DeepTau Discriminators: - **VSjet:** Tau vs QCD jets (most important) - **VSe:** Tau vs electrons - **VSmu:** Tau vs muons ← Problem here!

Why VSmu is Critical in MuTau

In MuTau channel: - One real **muon** (from W/Z decay) - One **tau** (signal)
VSmu discriminator: **"Is this tau actually a muon?"**

The Mystery

97% of taus have **NO VSmu discrimination** (value = 0)
Possible causes: 1. VSmu branch not filled in NanoAOD 2. VSmu working point too strict 3. Taus genuinely look like muons (unlikely - we have real taus)

Slide 7: The Evidence

Tau Quality Distribution

`tau_id1 = 255 (maximum quality): 836,368 events (81.7%)`
Most taus ARE high quality for VSjet!
But:
`tau_id3 (VSmu):`
`> 0: 31,672 (3.09%)`
`> 1: 22,479 (2.20%)`
`> 2: 22,479 (2.20%)`
But fail VSmu discrimination.

Conclusion

The tau quality is good (81.7% have perfect VSjet scores), but the **VSmu branch appears to be not properly filled** or the working point is inappropriate for MuTau channel.

Slide 8: Why This Explains Everything

The Event Loss Chain

```
Raw NanoAOD:                          ~Millions
    ↓ Lumi + HLT (fase0)
Fase0 output:                         1,023,514 per file
    ↓ Proton requirement (16.9%)
With protons:                           172,895
    ↓ Muon ID (80%)
After mu_id:                            138,903
    ↓ TAU ID (0.05%!) ← THE KILLER
After tau ID:                                65
    ↓ pT + geometry
Final per file:                               1

× 393 files:                            ~400 total events
```

Compare to Expected

If fase0 had proper quality cuts, we'd expect: - **Thousands** of Data events - **Thousands** of QCD events

Instead we get: - 1,708 Data - 234 QCD

Slide 9: Why weight=2.0 Was Wrong

What weight=2.0 Was Doing

```
weight = 2.0  # Artificially doubling event weights
```

- Compensated for "factor of 2" discrepancy
- Made histograms look better
- **But didn't fix the underlying problem!**

The Real Issue

We're not losing 50% of events...
We're losing 99.95% of events to tau ID cuts!
The factor of 2 was a **red herring** - a symptom of comparing to an incorrect reference, not the root cause.

Slide 10: The Fix - Option 1 (Recommended)

Uncomment Cuts in Fase0

In `fase0_qcd_data.py`, uncomment lines 177-182:

```
# Change from:
#    .Filter("Tau_idDeepTau2017v2p1VSjet[0] >= 63", "Tau VSjet")
#    .Filter("Tau_idDeepTau2017v2p1VSe[0]   >= 7",  "Tau VSe")
#    .Filter("Tau_idDeepTau2017v2p1VSmu[0]  >= 1",  "Tau VSmu")

# To:
.Filter("Tau_idDeepTau2017v2p1VSjet[0] >= 63", "Tau VSjet")
.Filter("Tau_idDeepTau2017v2p1VSe[0]   >= 7",  "Tau VSe")
.Filter("Tau_idDeepTau2017v2p1VSmu[0]  >= 1",  "Tau VSmu")  # Or > (
```

Then:

1. **Reprocess from fase0** (all 393 files)
2. Run fase1 on new fase0 output
3. Merge
4. **Remove weight=2.0** from plot_m.py

Expected Result:

Slide 11: The Fix - Option 2 (Quick Test)

If You Can't Reprocess Yet

Temporarily relax cuts in `fase1_data.py` and `fase1_qcd.py`:

```
# Change from:
.Filter("tau_id3 > 1", "Tau VSmu")

# To:
.Filter("tau_id3 > 0", "Tau VSmu (VLoose)")  # Or comment out entire
```

This Will:

- Give you more events immediately
- Test if VSmu is the problem
- Include lower-quality taus
- Not a proper physics solution

Use only for testing!

Slide 12: Before You Reprocess

Critical Checks Needed

1. **Verify VSmu branch in NanoAOD:**
   ```
   root -l root://cms-xrd-global.cern.ch///store/[your_file].root
   Events->Scan("Tau_idDeepTau2017v2p1VSmu", "", "", 100)
   ```
   Check if values are mostly 0 or properly distributed.
2. **Check with your advisor:**
   - What tau ID cuts did old processing use?
   - Is VSmu > 1 correct for MuTau channel?
   - Why were fase0 cuts commented out?
3. **Compare with TauTau channel:**
   - What VSmu working point do they use?
   - Do they apply cuts in fase0 or later?
4. **Check TauPOG recommendations:**
   - 2018 UL MuTau channel tau ID working points
   - Official recommendations for VSmu

Slide 13: Comparison with Old Method

Old Processing (DY/ttjets style)

```
NanoAOD → Fase1 (all quality cuts applied) → Merge → Add pileup prot
          ↑
          Probably had LOOSER tau ID cuts,
          or NO VSmu cut at all
```

New Processing (Current Data/QCD)

```
NanoAOD → Fase0 (NO quality cuts!) → Fase1 (STRICT cuts) → Merge
          ↑                           ↑
          Bug: cuts commented out     Rejects 99.95%
```

The Change

When switching from artificial to real protons, the quality cuts were **accidentally removed** from fase0, causing the massive rejection.

Slide 14: Working Points Reference

DeepTau2017v2p1 Bit Values

VSjet (tau_id1): - 1=VVVLoose, 2=VVLoose, 4=VLoose, 8=Loose - 16=Medium, 32=Tight, **64=VTight** ← Your cut: $>$ 63 - 128=VVTight
VSe (tau_id2): - 1=VVVLoose, 2=VVLoose, 4=VLoose, **8=Loose** ← Your cut: $>$ 7 - 16=Medium, 32=Tight, 64=VTight, 128=VVTight
VSmu (tau_id3): - 1=VLoose, **2=Loose** ← Your cut: $>$ 1 - 4=Medium, 8=Tight

Your Current Cuts

- VSjet: VTight or better (appropriate)
- VSe: Loose or better (appropriate)
- VSmu: Loose or better (possibly too strict, or branch issue)

Slide 15: Action Plan

Immediate (Today)

1. Run diagnostics (completed)
2. Identify root cause (completed)
3. Verify VSmu branch in raw NanoAOD
4. Consult with advisor about cuts

Short-term (This Week)

5. Test with relaxed VSmu cut (> 0 or commented out)
6. Compare with old processing scripts
7. Check TauPOG recommendations

Medium-term (When Ready)

8. Uncomment correct cuts in fase0
9. Reprocess all 393 files from fase0
10. Remove weight=2.0 from plot_m.py

Long-term (Future Analyses)

Slide 16: Key Takeaways

What We Learned

1. **The "factor of 2" was a red herring**
   - Real issue: 99.95% event rejection
   - Not a simple normalization problem
2. **Fase0 has a critical bug**
   - All quality cuts commented out
   - Saves low-quality taus that fail later
3. **VSmu is the bottleneck**
   - Only 2.2% of taus pass VSmu > 1
   - Only 3.09% have ANY VSmu score
   - Suggests branch issue or wrong working point
4. **weight=2.0 was masking the problem**
   - Compensated for symptoms, not cause
   - Should be removed after proper fix

Bottom Line

Reprocess from fase0 with proper cuts, and event counts should be correct without artificial weights.

Slide 17: Documents Created

Analysis Documents

1. **FACTOR_OF_2_ROOT_CAUSE.md**
   - Complete root cause analysis
   - 5 hypotheses investigated
   - Comparison with TauTau channel
2. **ANALYSIS_TAU_ID_ISSUE.md**
   - Detailed tau ID analysis
   - Working points reference
   - Decision matrix for fixes

Diagnostic Scripts

3. **investigate_factor_of_2.py**
   - Sequential cut analysis
   - Revealed 99.95% rejection
4. **check_tau_id_distribution.py**
   - Tau ID distributions
   - Identified VSmu bottleneck

All ready to use for further investigation!

Slide 18: Questions to Answer

Before Reprocessing

1. **What VSmu values exist in raw NanoAOD?**
   - Are they mostly 0?
   - Is the branch filled correctly?
2. **What cuts did old processing use?**
   - Same VSmu > 1?
   - Or looser / no VSmu cut?
3. **Why were fase0 cuts commented out?**
   - Intentional flexibility?
   - Or accidental during development?
4. **What do TauPOG recommend?**
   - For 2018 UL MuTau channel
   - VSmu working point

After Testing

5. **Does relaxing VSmu help?**
   - Try VSmu > 0 instead of > 1
   - Or remove entirely
6. **How many events do we expect?**

Slide 19: Expected Timeline

Phase 1: Verification (1-2 days)

- Check VSmu in NanoAOD
- Test relaxed cuts
- Consult advisor

Phase 2: Decision (1 day)

- Determine correct working points
- Plan reprocessing strategy

Phase 3: Reprocessing (1-2 weeks)

- Uncomment cuts in fase0
- Reprocess 393 files
- Run fase1 on new output

Phase 4: Validation (1-2 days)

- Check event counts

Slide 20: Summary

The Journey

Started with: "Why is everything at 50%?"
Discovered: - Fase0 has no quality cuts (bug) - 99.95% rejection at tau ID - VSmu only 2.2% acceptance (mystery)
Root Causes: 1. Commented-out cuts in fase0 2. VSmu branch issue or wrong working point 3. weight=2.0 compensating for wrong problem

The Solution

1. Fix fase0 cuts
2. Verify VSmu working point
3. Reprocess
4. Remove artificial weights

The Lesson

Always trace the full processing chain when debugging!
The apparent "factor of 2" was hiding a much larger issue that could only be found by checking each step sequentially.

Section 2

End of Presentation

Contact / Questions

All diagnostic scripts and analysis documents are in:
`/Users/utilizador/cernbox/tau_analysis/MuTau_channel/`

For questions, consult: - Your advisor - TauPOG contacts - CMS TWiki:
https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2