WHSS: Charge flip study

Arun Kumar[1], Sarah Freed[1], Jie Xiao[2], **Siewyan Hoh[3]**
University of Rice[1], Peking University[2], University of Padova[3]
Latino meeting
17th December 2020

Introduction

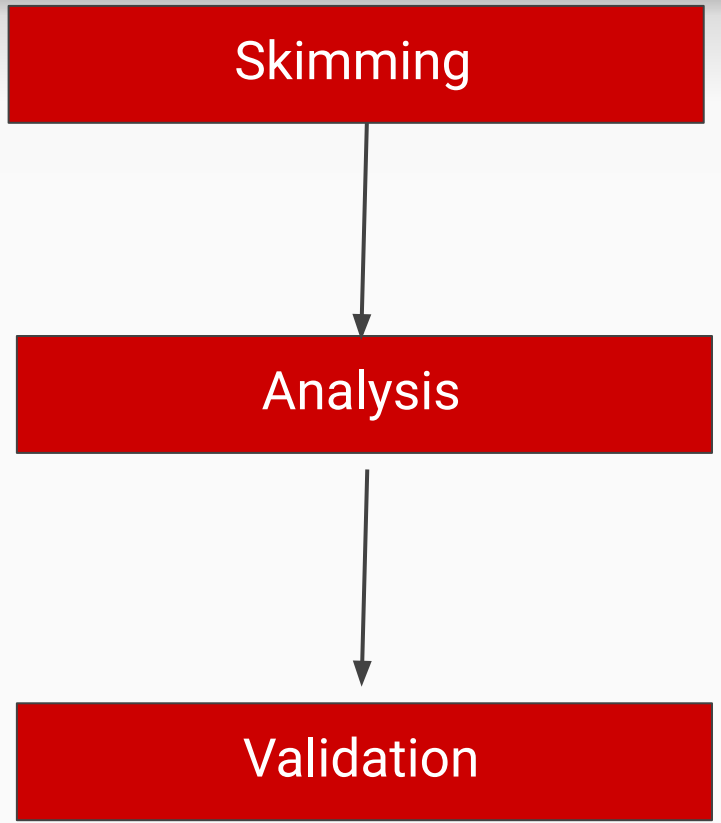
- The electron charge measurement is biased by presence of bremsstrahlung process within the tracker.
- Final state contains electron will be affected by charged flip effect. Expected size of effect ~ 1.3, 1.4 for 2017 and 2018 dataset.
- In WHSS analysis is covering e-mu final state, the mis-charged effect can be accounted by including a correction factor, computed from dedicated measurement.
- The mis-charged rate study is performed on 3 years , using nanov5 (latinov6) dataset, with the baseline lepton working points (HWW + ttHMVA).
- Study performed in RDataframe , nanoFlipper, with condor submission capability.

NanoFlipper : Workflow

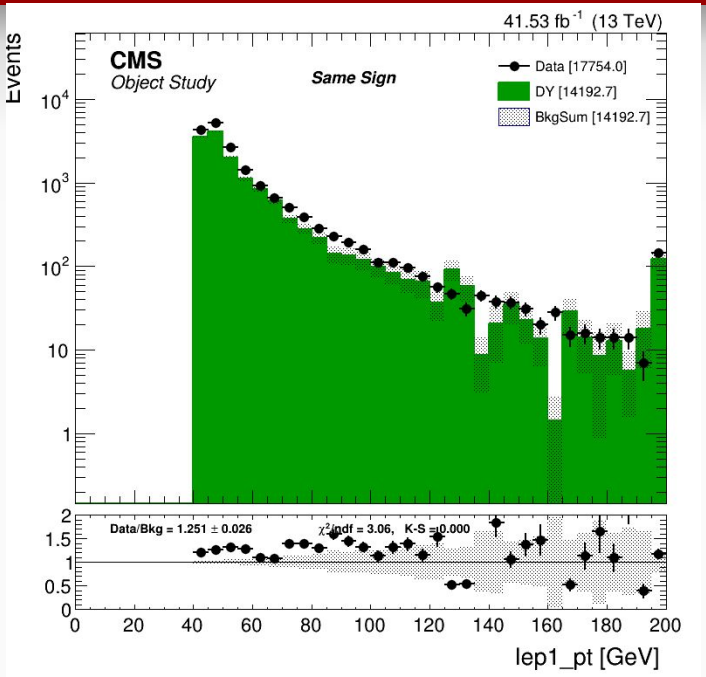
- Dataset Nanov5 for three years:
  - DoubleEG
  - DY (M-50) (DY ptll correction)
- Lepton WP = HWW + ttHMVA (Roberto's macro)

- Checking dependency on pt
- Eliminates EWK background with fit (Jie's code)
- Extracting charge flip and compute scale factor (Guillelmo's code)

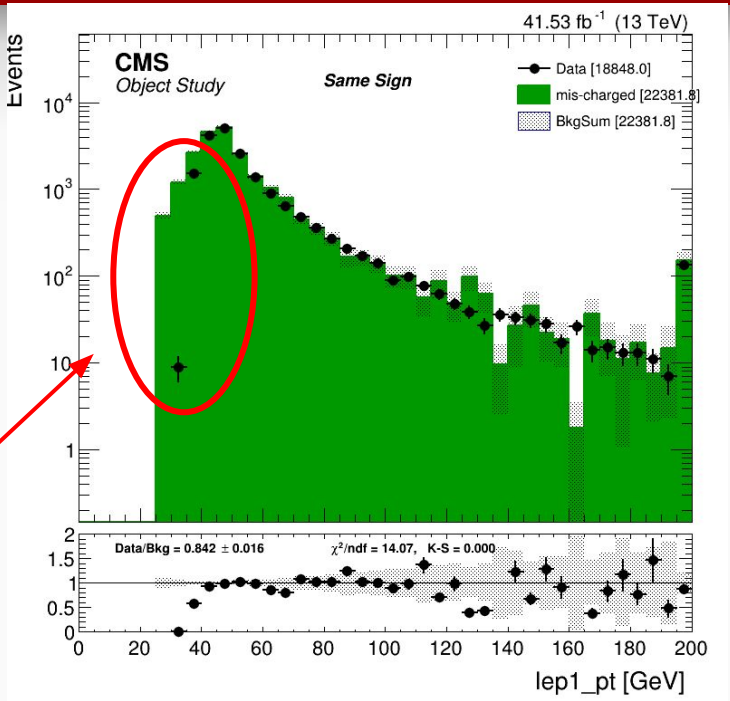
- Validating scale factor application
- Using pyroot , RDataframe is buggy.



Validation using SF from SingleElectron (2017)



x SF(pt,eta)



Below Trigger threshold

Before

→ Using DoubleEG for SF derivation

After

Study setup

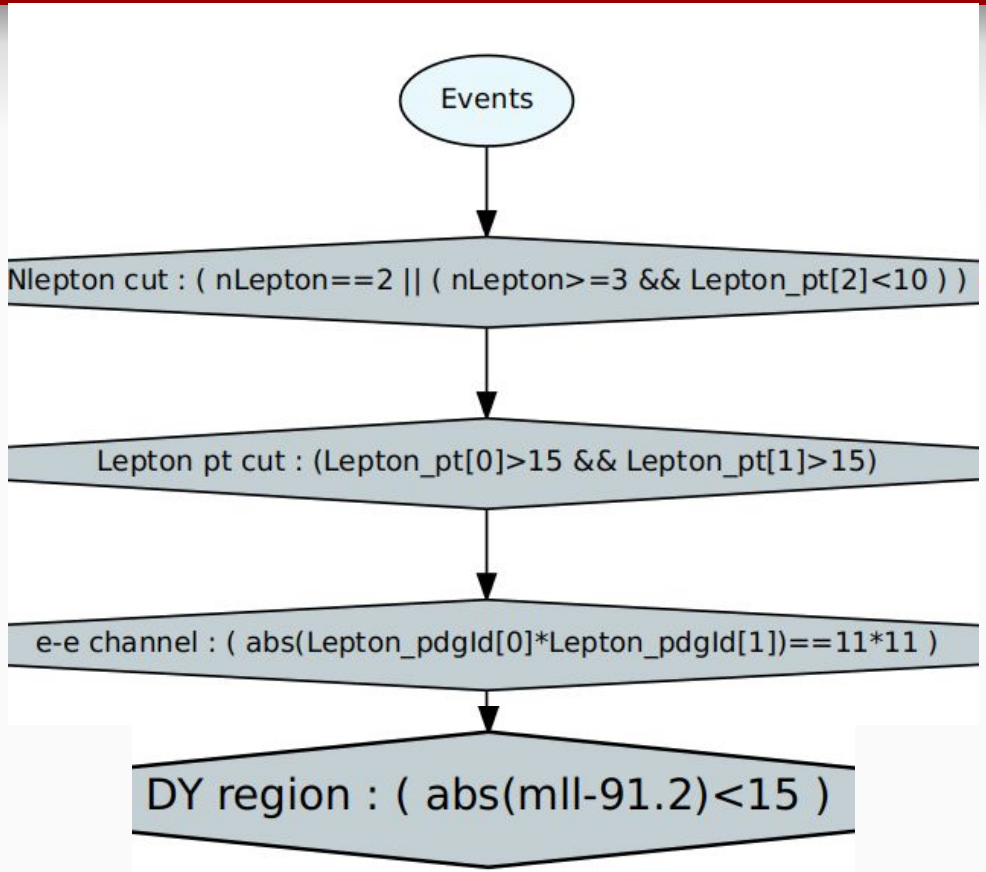
- Dataset Nanov5 for three years:
  - DoubleEG
  - DY (M-50)
- Lepton working points:
  - Baselined HWW WP : Latinov6 embedded
  - ttHMVA WP : derived from Nanov7
- MC : DY ptll correction.
- Measurement:
  - lowpt = [$trig1 , 200] ; [$trig2 , 20]
  - highpt = [$trig1 , 200] ; [20 , 200]
  - Eta bin [0 , 1.4 , 2.5]
- Zmass fit:
  - Signal pdf = Breit Wigner x Gaussian
  - Background pdf = Exponential

```
triggers = {
    'Trigger_dblEl' : {
        "2016" : [ "28" , "17" ] , # 23+5 ; 12+5
        "2017" : [ "28" , "17" ] , # 23+5 ; 12+5
        "2018" : [ "28" , "17" ]   # 23+5 ; 12+5
    },
    'Trigger_sngEl' : {
        "2016" : [ "32" , "15" ] , # 27+5
        "2017" : [ "40" , "15" ] , # 35+5
        "2018" : [ "40" , "15" ]   # 35+5
    }
}
```

```
tname : SingleMuon
three leptons final state: pass=78677       all=29875395    --
Lepton pt cut : (Lepton_pt[0]>15 && Lepton_pt[1]>5): pass=78
abs(Lepton_pdgId[0])==13: pass=74186       all=78677       --
abs(Lepton_pdgId[1])==11: pass=4731        all=74186       --
abs(Lepton_pdgId[2])==11: pass=564         all=4731        --
DY region : ( abs(mll-91.2)<15 ): pass=248        all=564
```

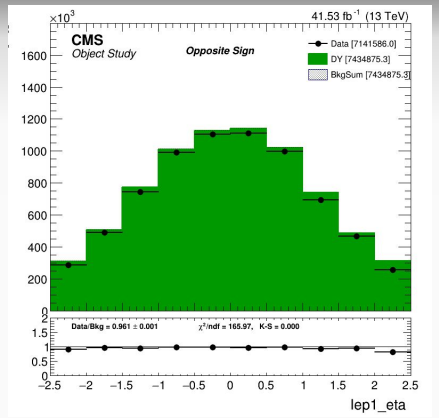
Limited statistics in “WZ” final state.

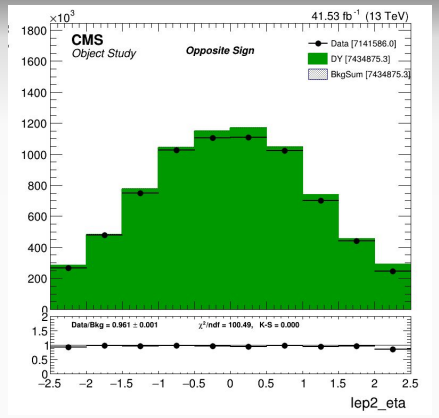
Skimming

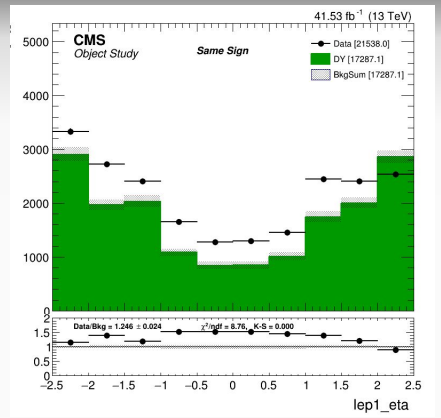


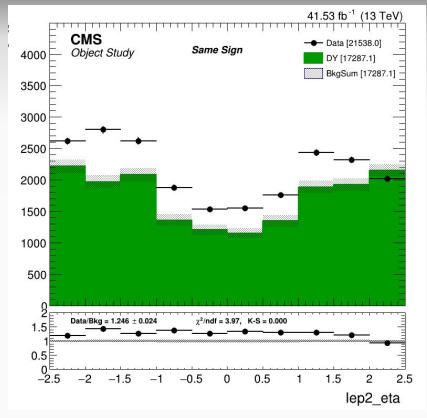
Drell-Yan (e+e-) CR skimming

Drell-Yan CR OS and SS modeling (2017)



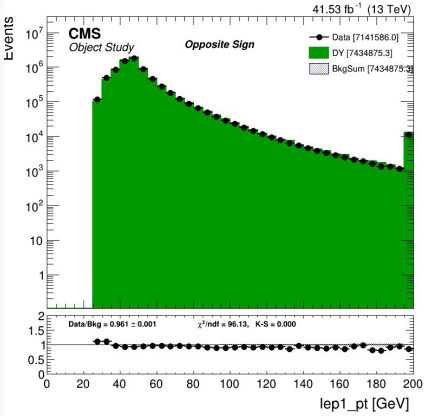


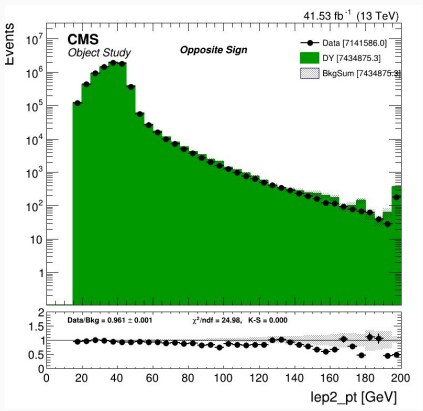


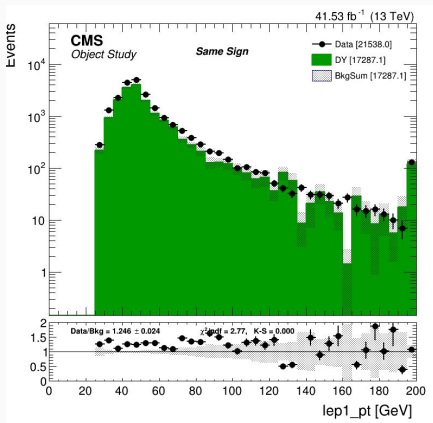


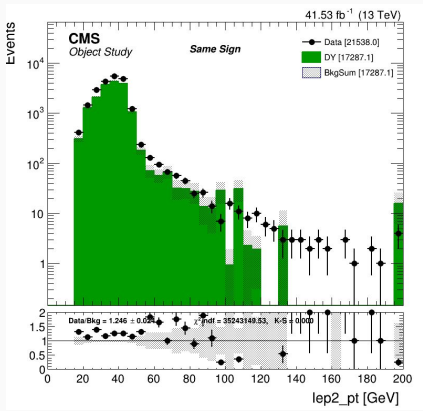
OS

SS

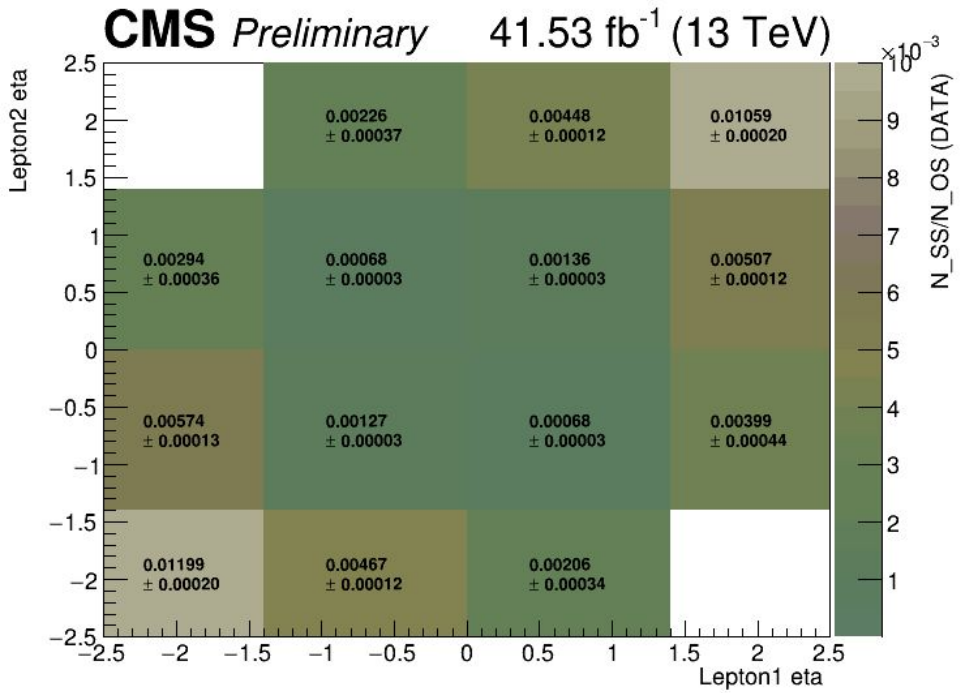




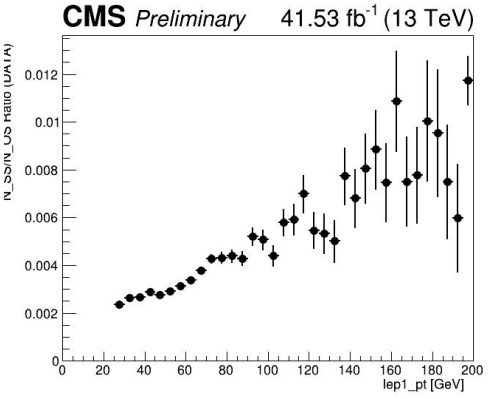


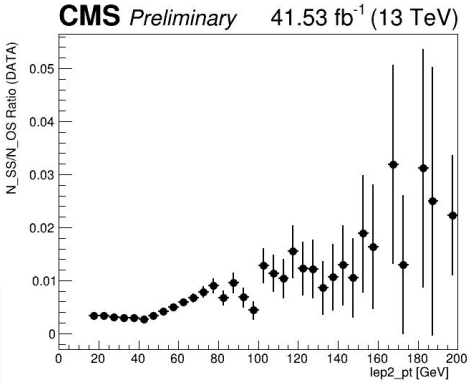


Drell-Yan CR Pt and Eta dependency (2017)



Charge flip probabilities are symmetric in eta (In DATA and MC)



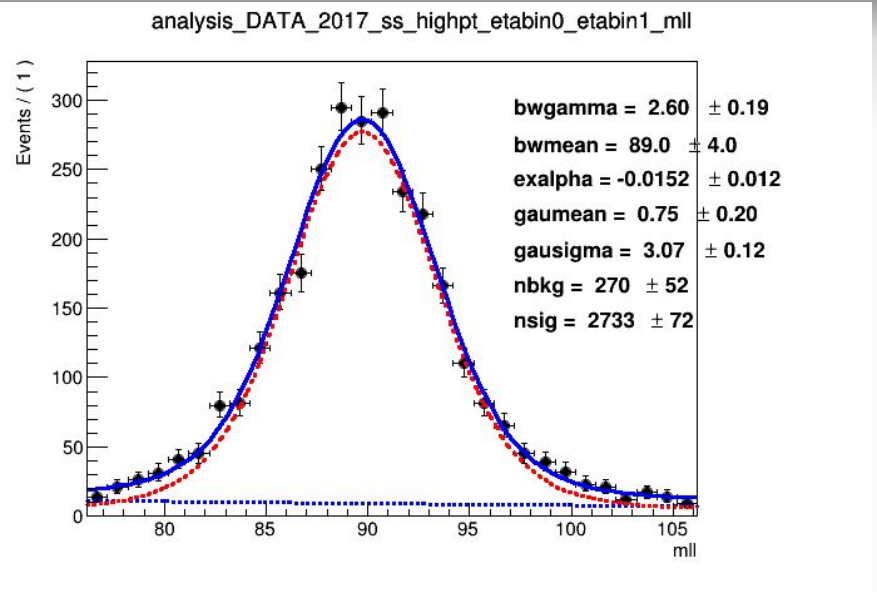


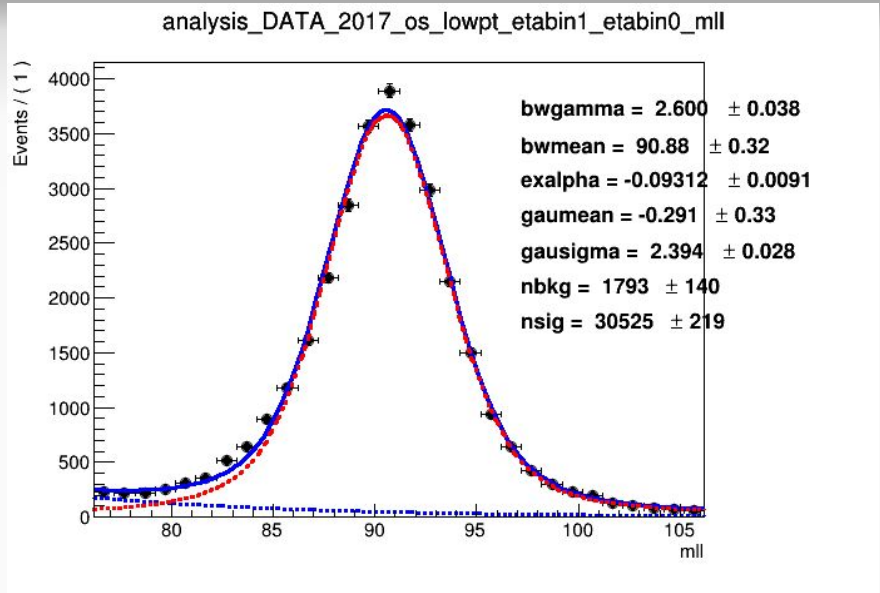
Probabilities shows upward trend in Lepton1 pt

Probability definition: N_SS / N_OS

Probabilities does not depend on Lepton2 pt

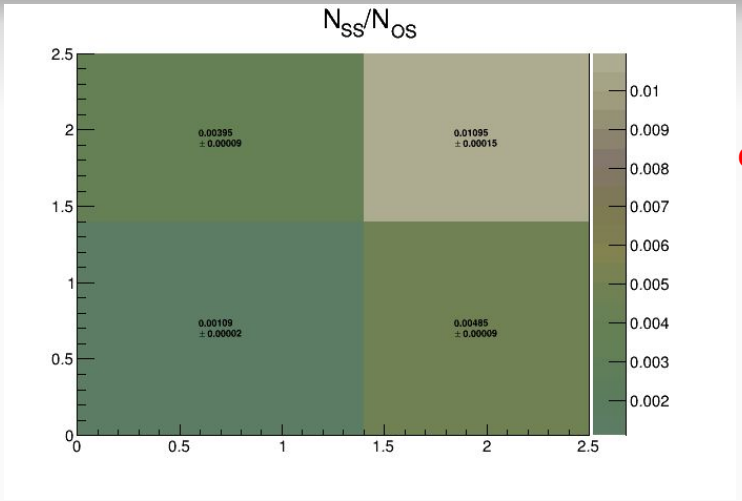
Zmass fit (2017)





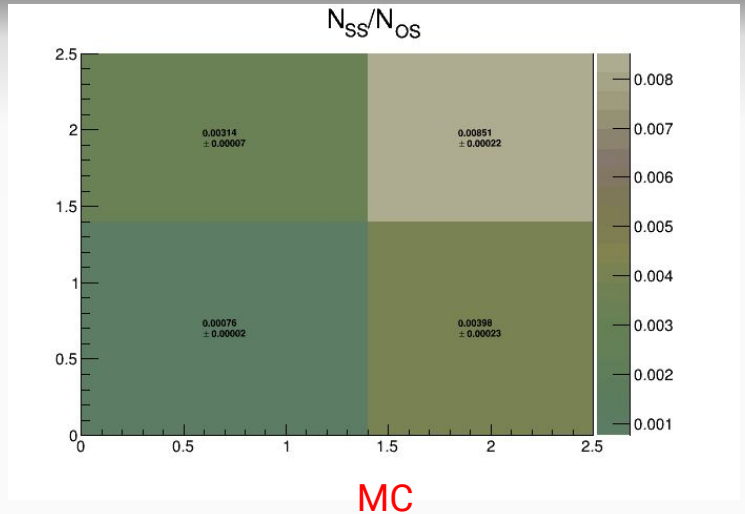
The number of signal and background events (in SS, OS categories) were extracted from the Z mass fit, both for lowpt and highpt bins.

Computes charge flip probabilites (2017)



DATA

Recomputes probabilities, using extracted number of signal events



MC

The total charge flip probability was fitted to extract individual electron charge flip probabilities.

$$\frac{N_{SS}}{N_{OS}} = \frac{\epsilon(p_T^{(1)}, |\eta^{(1)}|) \cdot (1 - \epsilon(p_T^{(2)}, |\eta^{(2)}|)) + \epsilon(p_T^{(2)}, |\eta^{(2)}|) \cdot (1 - \epsilon(p_T^{(1)}, |\eta^{(1)}|))}{1 - [\epsilon(p_T^{(1)}, |\eta^{(1)}|) \cdot (1 - \epsilon(p_T^{(2)}, |\eta^{(2)}|)) + \epsilon(p_T^{(2)}, |\eta^{(2)}|) \cdot (1 - \epsilon(p_T^{(1)}, |\eta^{(1)}|))]}$$

Then, the scale factor is calculated by taking the ratio between DATA adn MC.

Flip probabilities and SF (SingleElectron)

High pt bins : [40 , 200]

SingleElectron

Error is statistical only.

```
 ===> scale factor DATA/MC for 2016
q0 data : 5.480e-04 +/- 1.101e-05 ; mc : 5.772e-04 +/- 1.233e-05 ; SF : 9.495e-01 +/- 2.934e-02 ( rel.error : 3.09 % )
q1 data : 8.081e-03 +/- 8.669e-05 ; mc : 7.596e-03 +/- 7.512e-05 ; SF : 1.064e+00 +/- 1.459e-02 ( rel.error : 1.37 % )
 ===> scale factor DATA/MC for 2017
q0 data : 5.030e-04 +/- 9.487e-06 ; mc : 3.606e-04 +/- 7.616e-07 ; SF : 1.395e+00 +/- 1.898e-02 ( rel.error : 1.36 % )
q1 data : 4.668e-03 +/- 5.940e-05 ; mc : 3.782e-03 +/- 4.595e-05 ; SF : 1.234e+00 +/- 1.759e-02 ( rel.error : 1.43 % )
 ===> scale factor DATA/MC for 2018
q0 data : 5.630e-04 +/- 7.718e-06 ; mc : 3.955e-04 +/- 7.007e-06 ; SF : 1.423e+00 +/- 2.240e-02 ( rel.error : 1.57 % )
q1 data : 5.102e-03 +/- 4.846e-05 ; mc : 3.903e-03 +/- 5.194e-05 ; SF : 1.307e+00 +/- 1.635e-02 ( rel.error : 1.25 % )
```

Low pt bins : [15 , 200]

2016 : sf ~ 1 ; 2017,2018 : sf ~ 1.3 / 1.4

```
 ===> scale factor DATA/MC for 2016
q0 data : 9.704e-04 +/- 1.295e-04 ; mc : 1.141e-03 +/- 1.509e-04 ; SF : 8.504e-01 +/- 1.879e-01 ( rel.error : 22.10 % )
q1 data : 2.513e-03 +/- 3.173e-04 ; mc : 2.854e-03 +/- 3.554e-04 ; SF : 8.808e-01 +/- 1.774e-01 ( rel.error : 20.14 % )
 ===> scale factor DATA/MC for 2017
q0 data : 7.758e-04 +/- 1.453e-04 ; mc : 5.173e-04 +/- 2.616e-06 ; SF : 1.500e+00 +/- 1.874e-01 ( rel.error : 12.50 % )
q1 data : 2.291e-03 +/- 3.306e-04 ; mc : 7.370e-04 +/- 1.457e-04 ; SF : 3.109e+00 +/- 2.448e-01 ( rel.error : 7.87 % )
 ===> scale factor DATA/MC for 2018
q0 data : 6.937e-04 +/- 1.023e-04 ; mc : 5.533e-04 +/- 6.624e-05 ; SF : 1.254e+00 +/- 1.900e-01 ( rel.error : 15.15 % )
q1 data : 2.910e-03 +/- 3.332e-04 ; mc : 9.980e-04 +/- 2.386e-04 ; SF : 2.916e+00 +/- 2.651e-01 ( rel.error : 9.09 % )
```

Flip probabilities and SF (DoubleEG)

High pt bins : [28 , 200]

DoubleEG

Error is statistical only.

```
 ===> scale factor DATA/MC for 2016
q0 data : 5.594e-04 +/- 1.121e-05 ; mc : 5.815e-04 +/- 1.232e-05 ; SF : 9.619e-01 +/- 2.916e-02 ( rel.error : 3.03 % )
q1 data : 8.444e-03 +/- 9.775e-05 ; mc : 7.521e-03 +/- 7.554e-05 ; SF : 1.123e+00 +/- 1.533e-02 ( rel.error : 1.37 % )
 ===> scale factor DATA/MC for 2017
q0 data : 5.288e-04 +/- 9.882e-06 ; mc : 3.691e-04 +/- 1.036e-05 ; SF : 1.433e+00 +/- 3.371e-02 ( rel.error : 2.35 % )
q1 data : 4.522e-03 +/- 4.846e-05 ; mc : 3.219e-03 +/- 5.720e-05 ; SF : 1.405e+00 +/- 2.075e-02 ( rel.error : 1.48 % )
 ===> scale factor DATA/MC for 2018
q0 data : 5.632e-04 +/- 8.855e-06 ; mc : 3.862e-04 +/- 7.435e-06 ; SF : 1.458e+00 +/- 2.486e-02 ( rel.error : 1.70 % )
q1 data : 4.848e-03 +/- 4.267e-05 ; mc : 3.767e-03 +/- 3.875e-05 ; SF : 1.287e+00 +/- 1.354e-02 ( rel.error : 1.05 % )
```

2016 : sf ~ 1 ; 2017,2018 : sf ~ 1.4

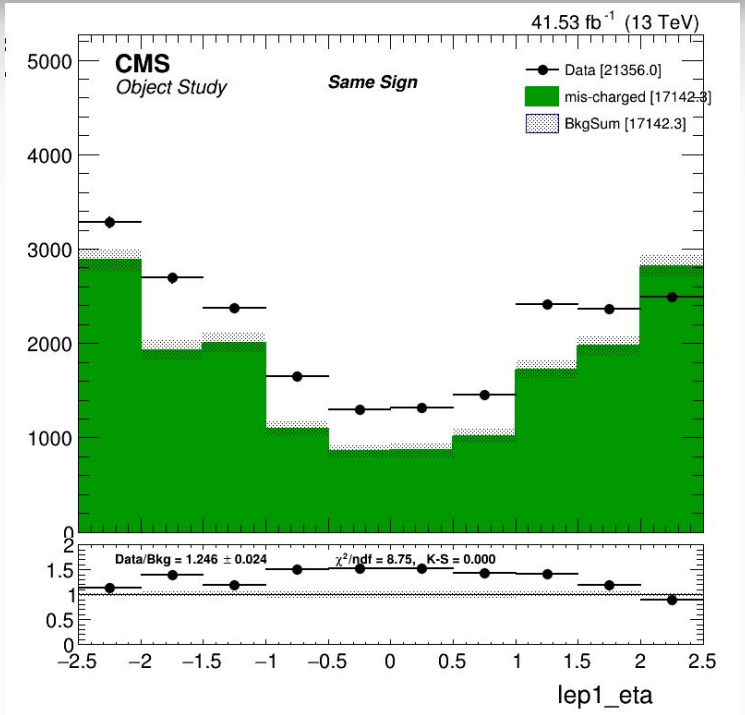
Low pt bins : [15 , 200]

```
 ===> scale factor DATA/MC for 2016
q0 data : 1.010e-03 +/- 1.377e-04 ; mc : 1.074e-03 +/- 1.602e-04 ; SF : 9.408e-01 +/- 2.021e-01 ( rel.error : 21.48 % )
q1 data : 9.187e-03 +/- 6.839e-04 ; mc : 2.515e-03 +/- 3.864e-04 ; SF : 3.653e+00 +/- 1.707e-01 ( rel.error : 4.67 % )
 ===> scale factor DATA/MC for 2017
q0 data : 7.730e-04 +/- 1.332e-04 ; mc : 5.168e-04 +/- 7.426e-05 ; SF : 1.496e+00 +/- 2.244e-01 ( rel.error : 15.00 % )
q1 data : 1.860e-03 +/- 2.919e-04 ; mc : 1.859e-03 +/- 4.456e-04 ; SF : 1.001e+00 +/- 2.865e-01 ( rel.error : 28.63 % )
 ===> scale factor DATA/MC for 2018
q0 data : 7.699e-04 +/- 1.176e-04 ; mc : 4.354e-04 +/- 6.604e-05 ; SF : 1.768e+00 +/- 2.152e-01 ( rel.error : 12.17 % )
q1 data : 2.540e-03 +/- 2.030e-04 ; mc : 9.668e-04 +/- 2.723e-04 ; SF : 2.627e+00 +/- 2.927e-01 ( rel.error : 11.14 % )
```

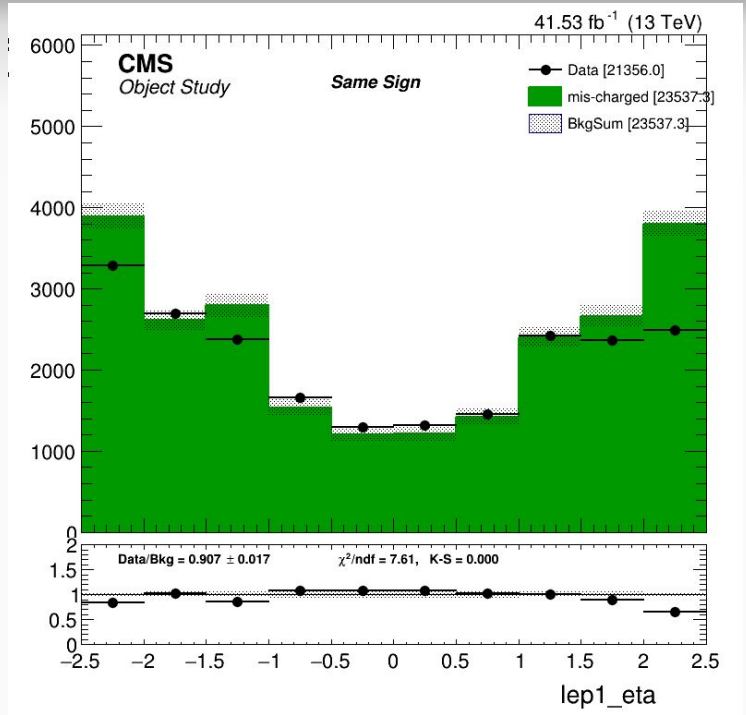
Validation

1. The scale factors obtained from the measurements (q0 , q1) is (sf1 , sf2). They represent the global correction between distributions, independent of number of electrons.
2. Therefore, the SF applied to MC is taken as (sf1 + sf2)/2 (or either one?).
3. Validation method: SF(pt,eta) x MC_SS != DATA_SS
4. Validation phase space (SS): lep1_pt > 25 ; lep2_pt > 20; 2 same sign electron. (WHSS selection).

SF Validation (2017)



x SF(pt,eta)

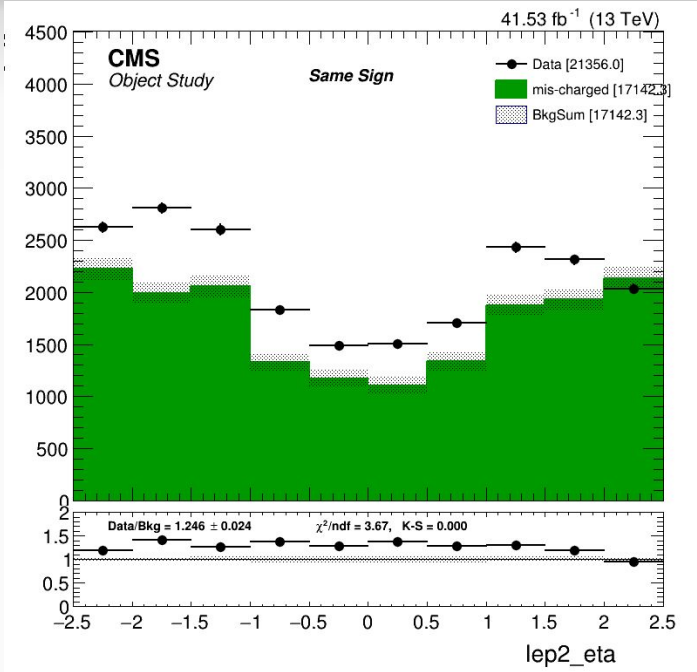


Before

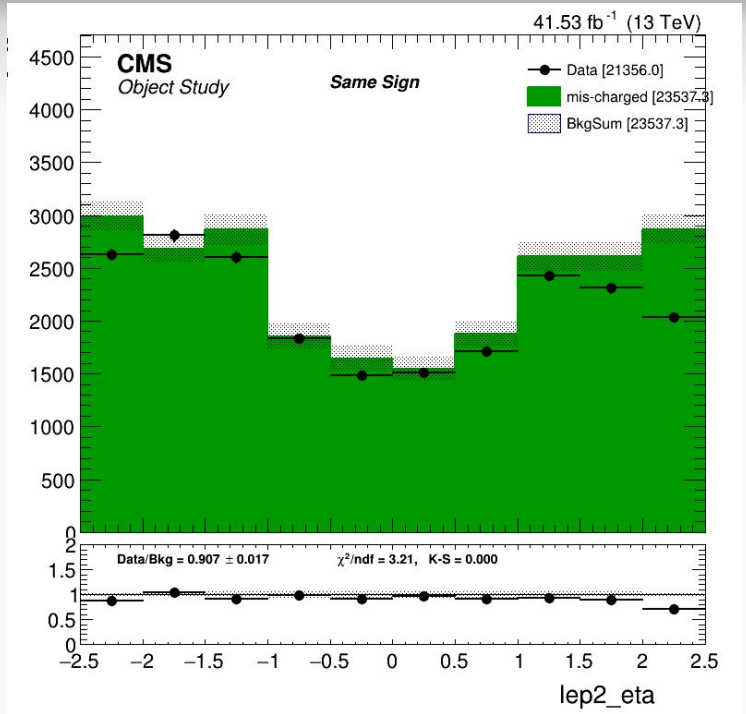
After

SF(pt,eta) x MC_SS != DATA_SS

SF Validation (2017)



x SF(pt,eta)

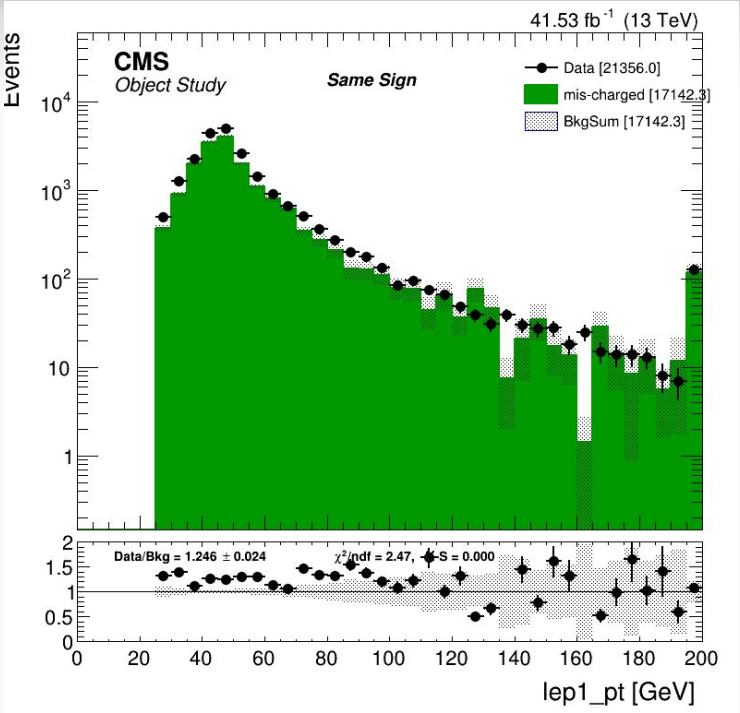


Before

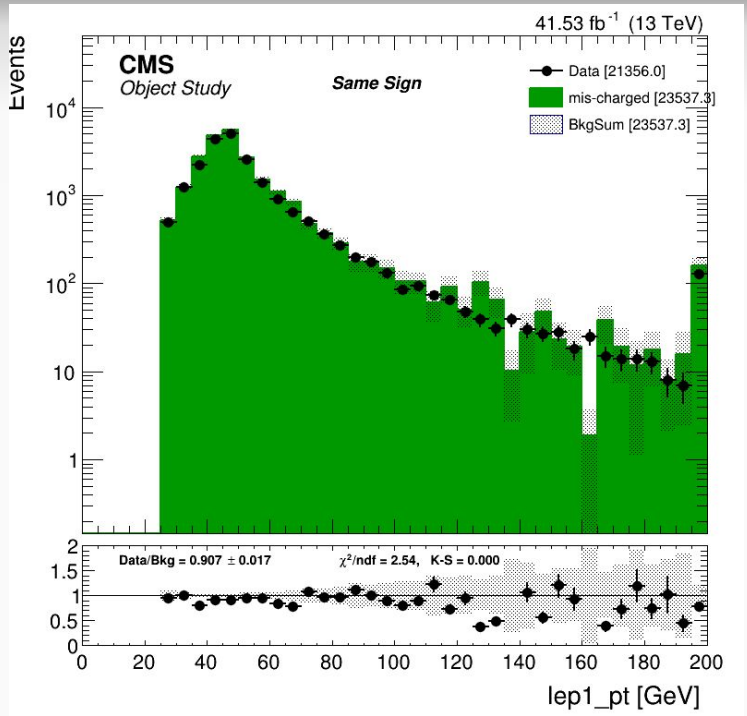
SF(pt,eta) x MC_SS != DATA_SS

After

SF Validation (2017)



x SF(pt,eta)

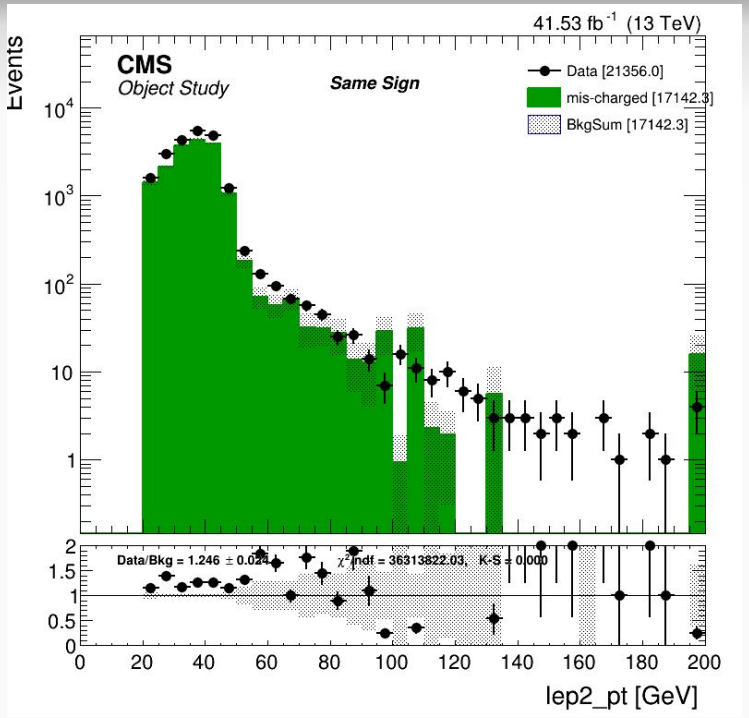


Before

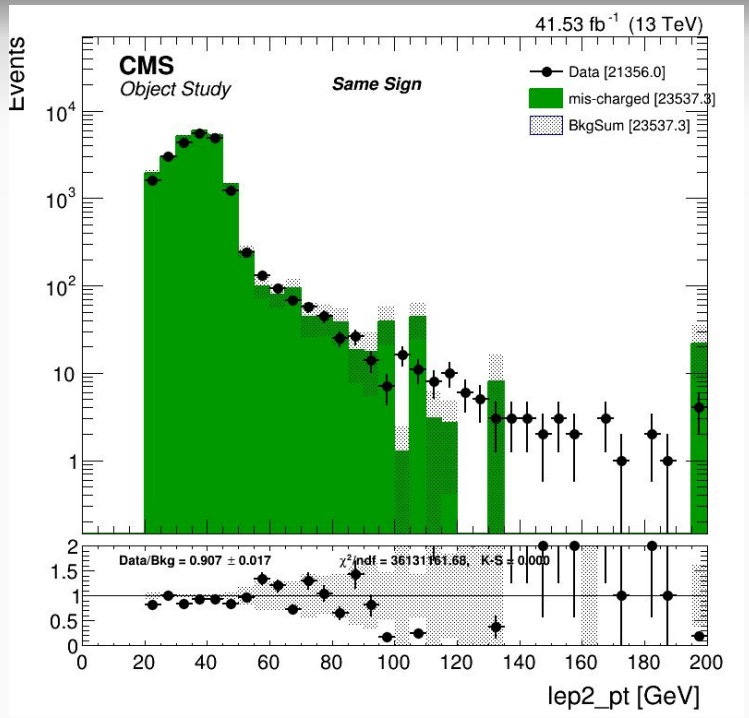
SF(pt,eta) x MC_SS != DATA_SS

After

SF Validation (2017)



x SF(pt,eta)



Before

SF(pt,eta) x MC_SS != DATA_SS

After

Summary

1.) The charge flip assessment is performed , and the size of the effect is consistent with expected values.
2.) Potential baise in lowpt bin measurement due to trigger threshold.
3.) Validation on scale factor application, show good agreement between MC and DATA, and ready for analysis consumption.
4.) Systematics inclusion:
    a.) Alternative MC sample
    b.) Alternative signal and background Pdf in fitting.
    c.) Different Zmass windonw during fitting.
5.) The (nominal) scale factor is ready. On-fly-macro is fixed. (Thanks Susan!) and its ready to use.

Backup

Overview study

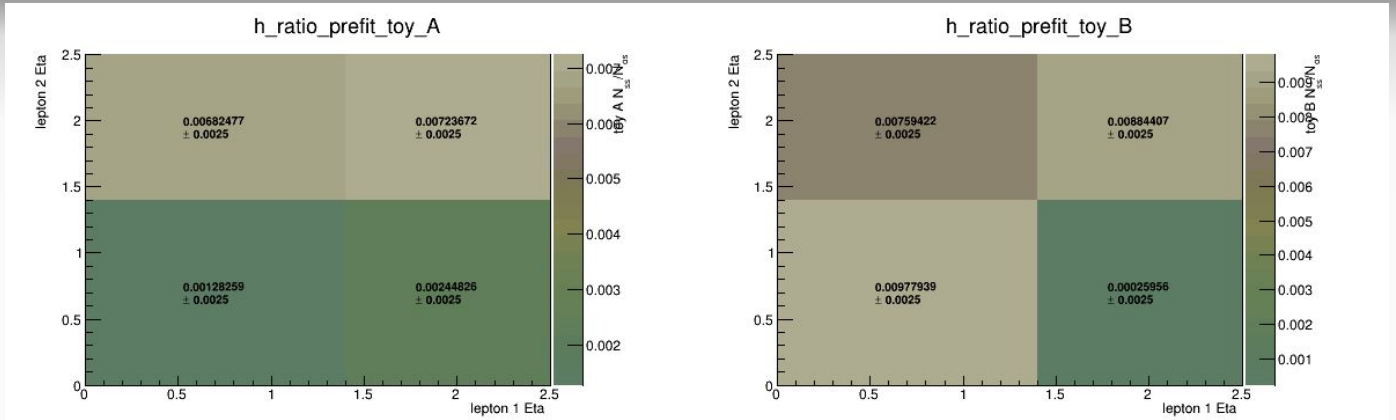
- Charge flip study workflow:
  - Derivation:
    - Devises e-e enriched DY control region.
    - Remove other EWK backgrounds, elimination by fitting under Z mass peak.
  - Measurement:
    - Taking ratio of SS/OS, assessing Pt/ Eta dependency.
    - Measures ratio of SS/OS in Pt / Eta by fitting to a charge flip probability formula.
  - Validation:
    - Testing on three closure tests. (1 test for now)
  - Application:
    - Applying scale factor on same sign MC.

Charge flip probability model

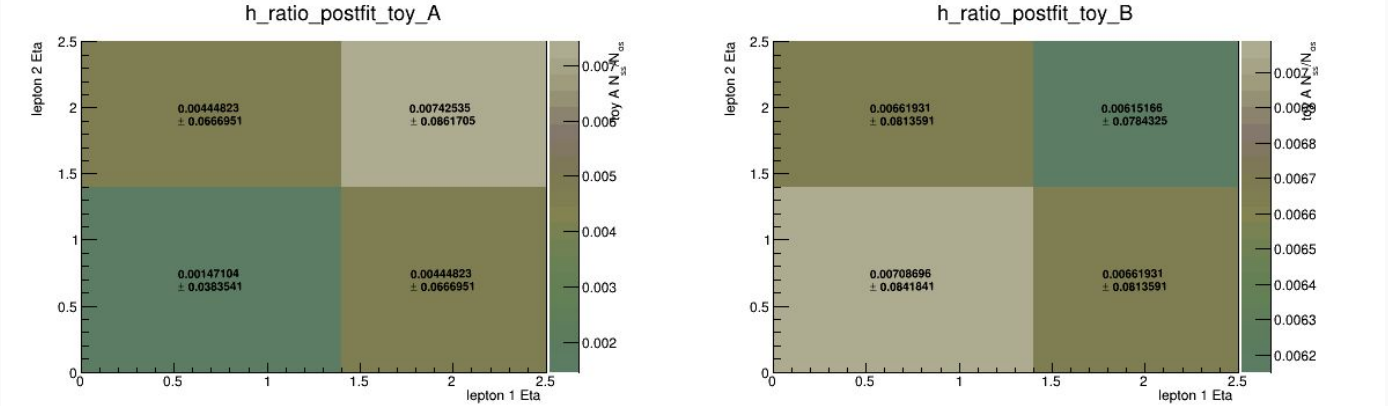
```
def model_2x2( i , par ):

    if   i == 0:  value = ( par[0] * (1-par[ 0]) + (1-par[ 0]) * par[ 0] ) / ( 1 - ( par[0] * (1-par[ 0]) + (1-par[ 0]) * par[ 0] ) )
    elif i == 1:  value = ( par[0] * (1-par[ 1]) + (1-par[ 0]) * par[ 1] ) / ( 1 - ( par[0] * (1-par[ 1]) + (1-par[ 0]) * par[ 1] ) )

    elif i == 2:  value = ( par[1] * (1-par[ 0]) + (1-par[ 1]) * par[ 0] ) / ( 1 - ( par[1] * (1-par[ 0]) + (1-par[ 1]) * par[ 0] ) )
    elif i == 3:  value = ( par[1] * (1-par[ 1]) + (1-par[ 1]) * par[ 1] ) / ( 1 - ( par[1] * (1-par[ 1]) + (1-par[ 1]) * par[ 1] ) )

    return value
pass
```

Charge flip probability model (Toy)

Toy data A &
Toy data B



FITTED:
Toy data A &
Toy data B



Charge flip probability and SF (previous)

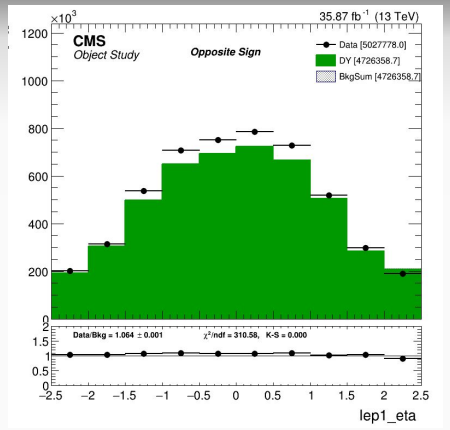
High pt

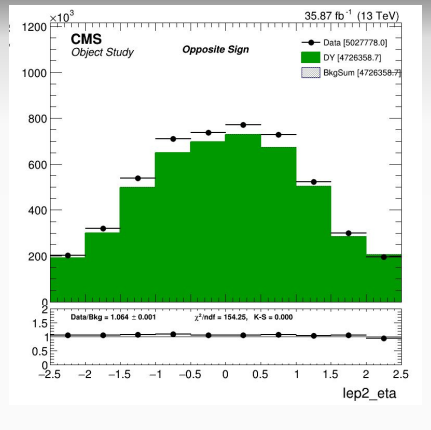
```
 ===> scale factor DATA/MC for 2016
q0 data : 6.316e-04 +/- 1.189e-05 ; mc : 6.473e-04 +/- 1.275e-05 ; SF : 9.757e-01 +/- 2.724e-02 ( rel.error : 2.79 % )
q1 data : 8.716e-03 +/- 1.042e-04 ; mc : 7.753e-03 +/- 7.931e-05 ; SF : 1.124e+00 +/- 1.573e-02 ( rel.error : 1.40 % )
 ===> scale factor DATA/MC for 2017
q0 data : 5.686e-04 +/- 1.026e-05 ; mc : 4.069e-04 +/- 6.365e-06 ; SF : 1.397e+00 +/- 2.388e-02 ( rel.error : 1.71 % )
q1 data : 4.749e-03 +/- 5.935e-05 ; mc : 3.508e-03 +/- 5.777e-05 ; SF : 1.354e+00 +/- 2.067e-02 ( rel.error : 1.53 % )
 ===> scale factor DATA/MC for 2018
q0 data : 6.131e-04 +/- 8.171e-06 ; mc : 4.112e-04 +/- 1.433e-05 ; SF : 1.491e+00 +/- 3.732e-02 ( rel.error : 2.50 % )
q1 data : 4.951e-03 +/- 4.348e-05 ; mc : 4.197e-03 +/- 4.626e-05 ; SF : 1.180e+00 +/- 1.409e-02 ( rel.error : 1.19 % )
```

Low pt

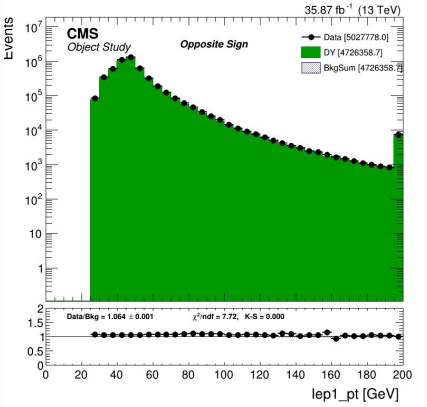
```
 ===> scale factor DATA/MC for 2016
q0 data : 1.114e-03 +/- 1.386e-04 ; mc : 1.216e-03 +/- 1.447e-04 ; SF : 9.158e-01 +/- 1.722e-01 ( rel.error : 18.80 % )
q1 data : 9.033e-03 +/- 6.305e-04 ; mc : 2.567e-03 +/- 3.129e-04 ; SF : 3.519e+00 +/- 1.405e-01 ( rel.error : 3.99 % )
 ===> scale factor DATA/MC for 2017
q0 data : 9.134e-04 +/- 1.308e-04 ; mc : 2.302e-04 +/- 2.966e-04 ; SF : 3.968e+00 +/- 1.297e+00 ( rel.error : 32.67 % )
q1 data : 1.572e-03 +/- 2.801e-04 ; mc : 1.884e-03 +/- 3.392e-04 ; SF : 8.343e-01 +/- 2.533e-01 ( rel.error : 30.36 % )
 ===> scale factor DATA/MC for 2018
q0 data : 8.488e-04 +/- 1.015e-04 ; mc : 6.600e-04 +/- 6.098e-05 ; SF : 1.286e+00 +/- 1.511e-01 ( rel.error : 11.75 % )
q1 data : 1.931e-03 +/- 2.109e-04 ; mc : 2.478e-03 +/- 2.575e-04 ; SF : 7.791e-01 +/- 1.508e-01 ( rel.error : 19.35 % )
```

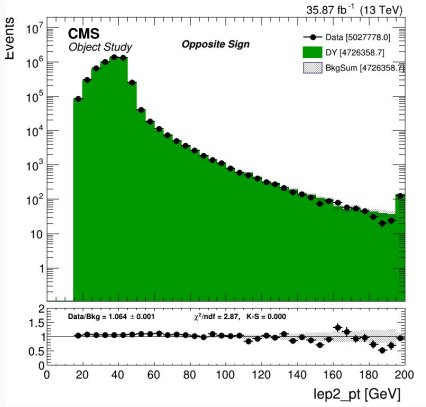
Drell-Yan CR OS and SS modeling (2016)

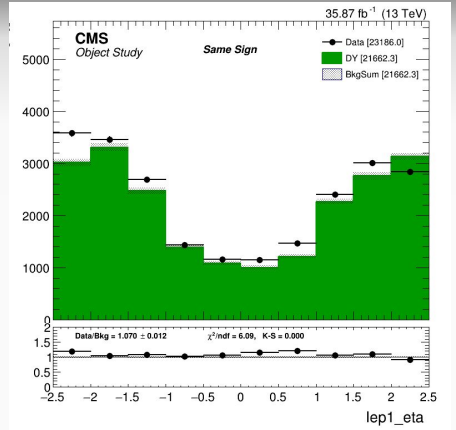


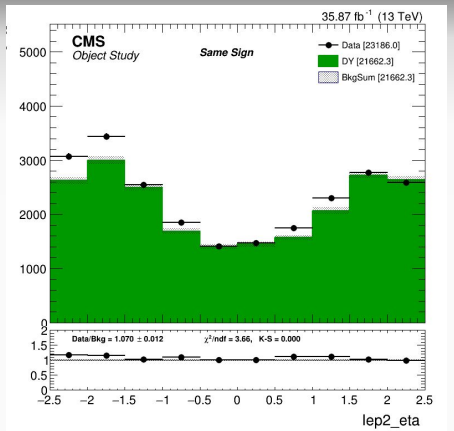


OS

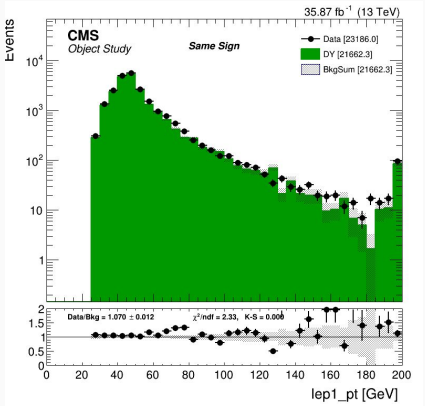


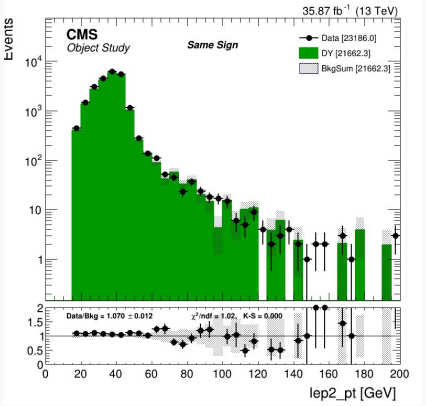




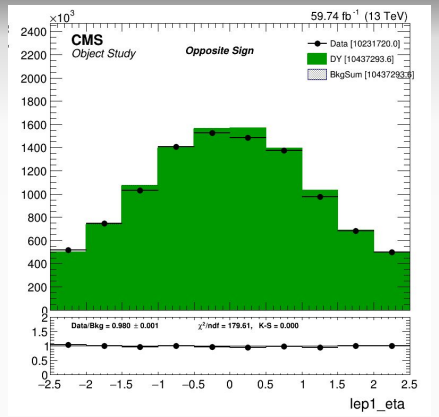


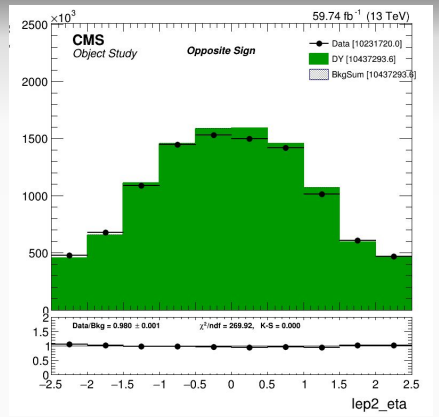
SS



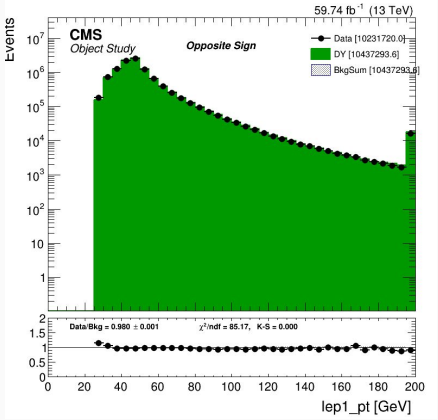


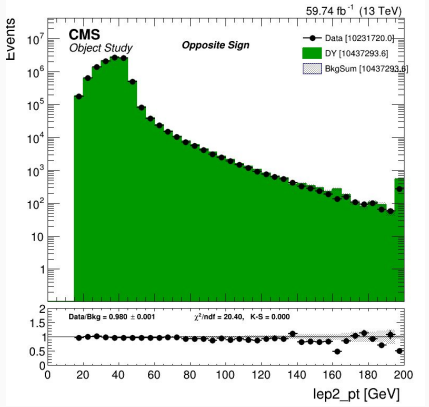
Drell-Yan CR OS and SS modeling (2018)

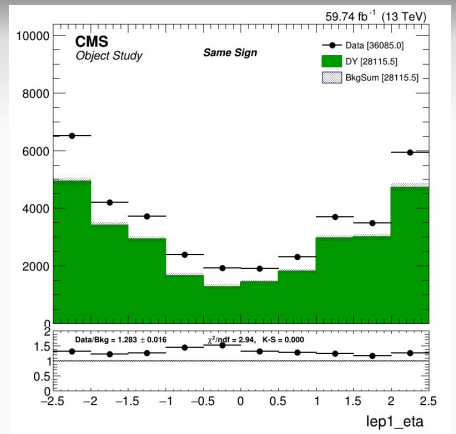


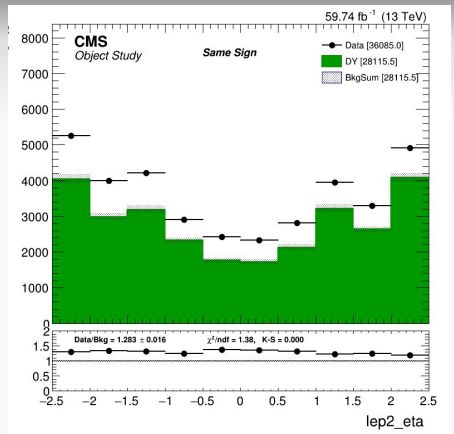


OS

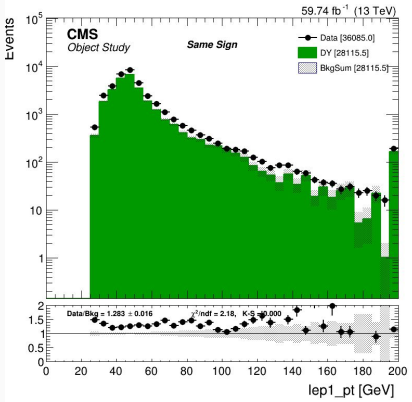


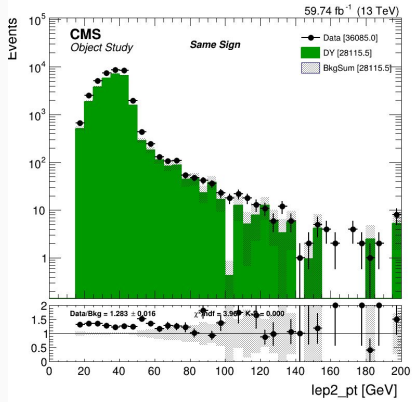




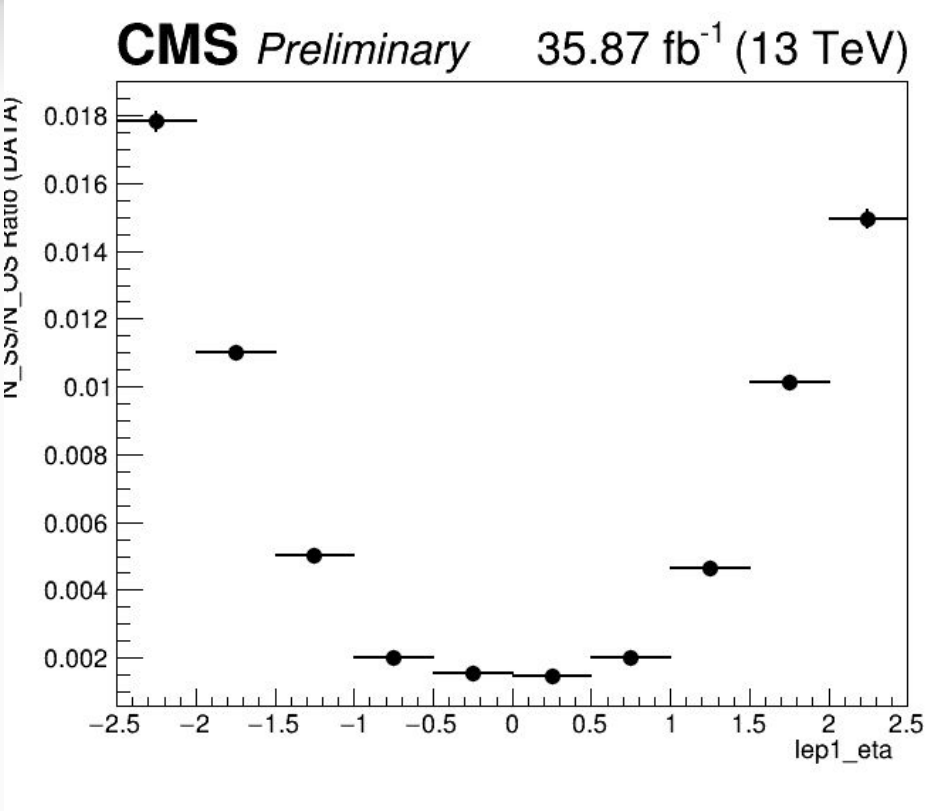


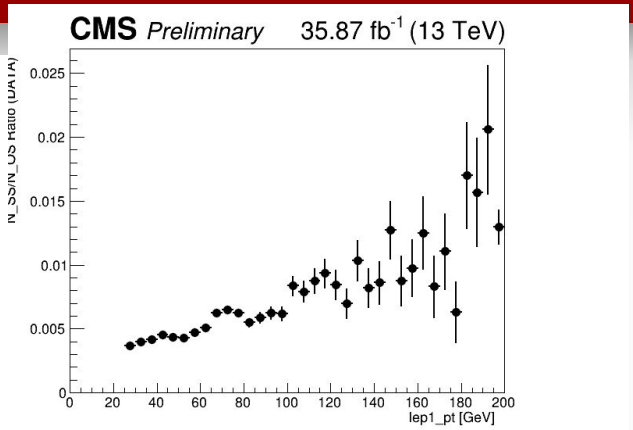
SS

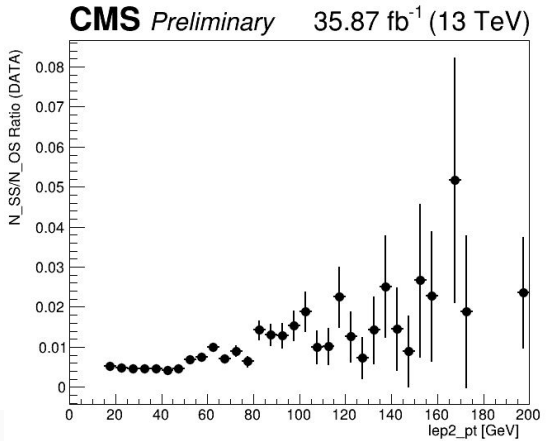




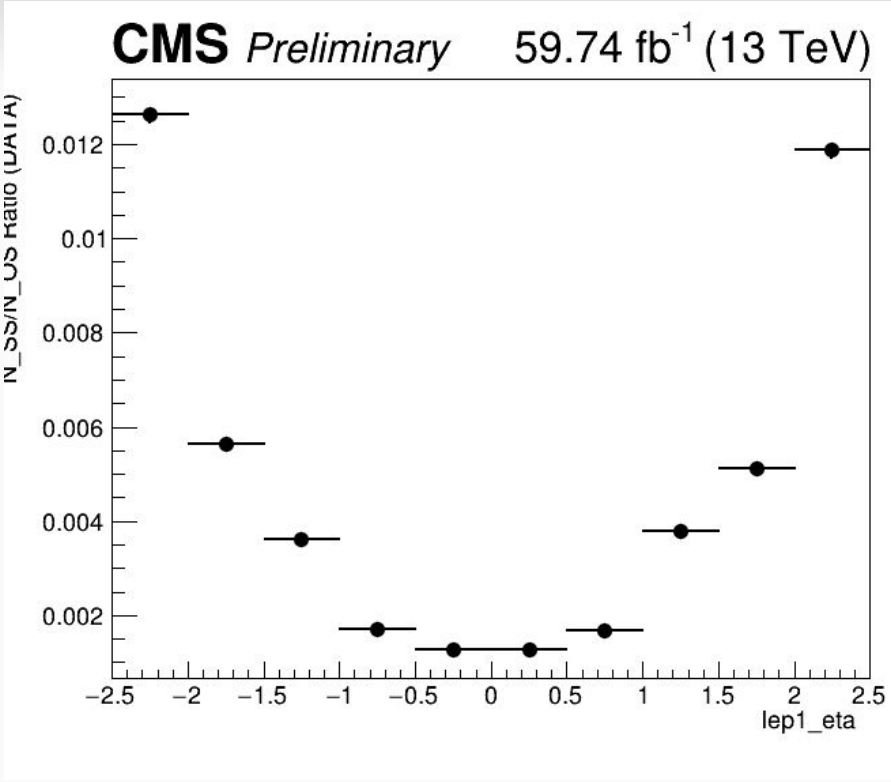
Drell-Yan CR Pt and Eta dependency (2016)

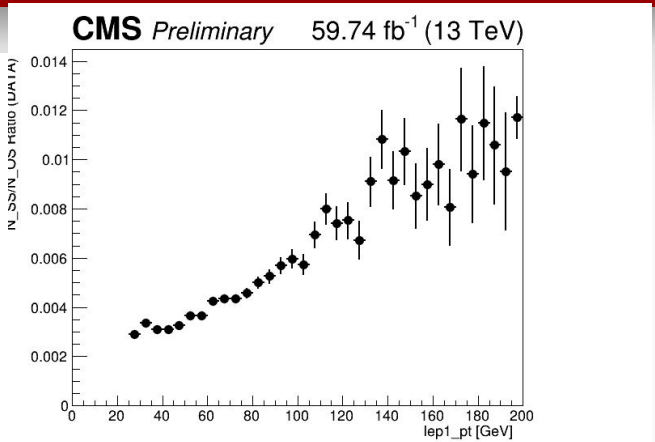


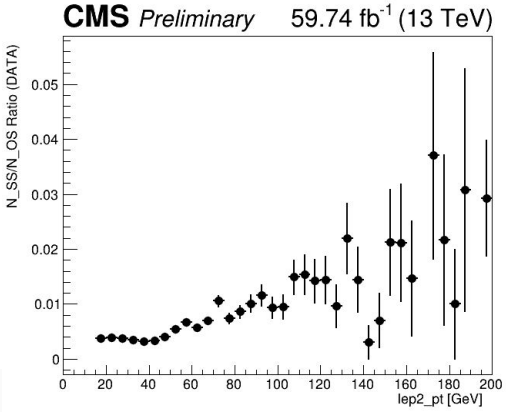




Drell-Yan CR Pt and Eta dependency (2018)

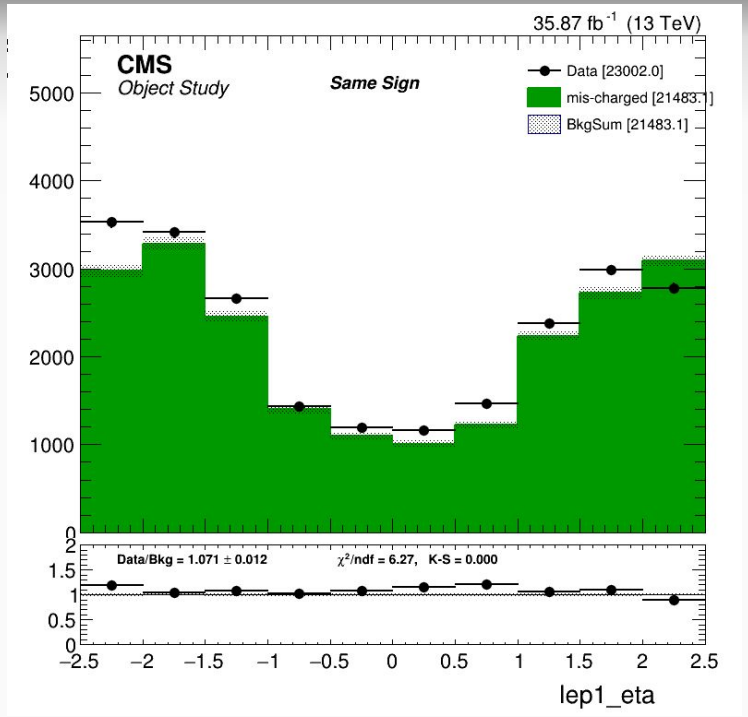




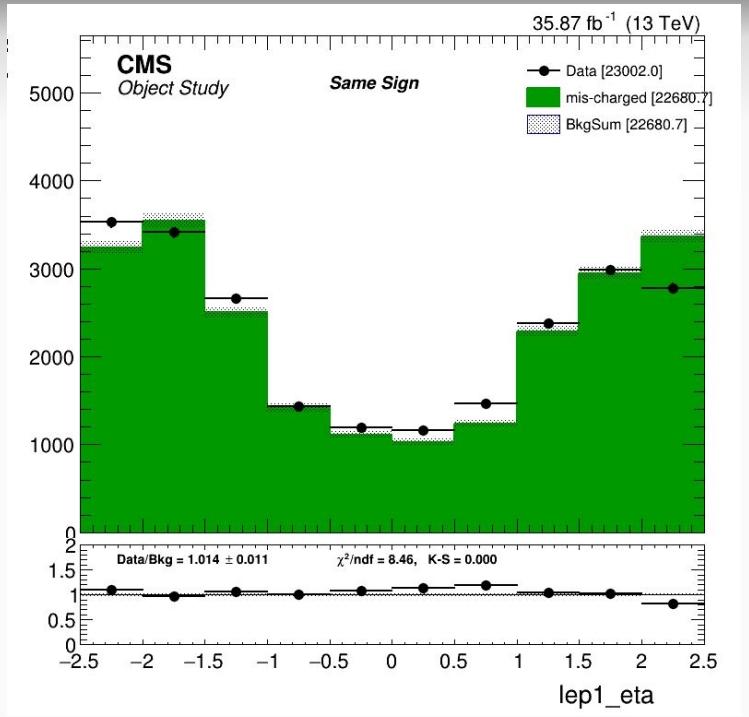


Validation 2016

SF Validation (2016)



x SF(pt,eta)

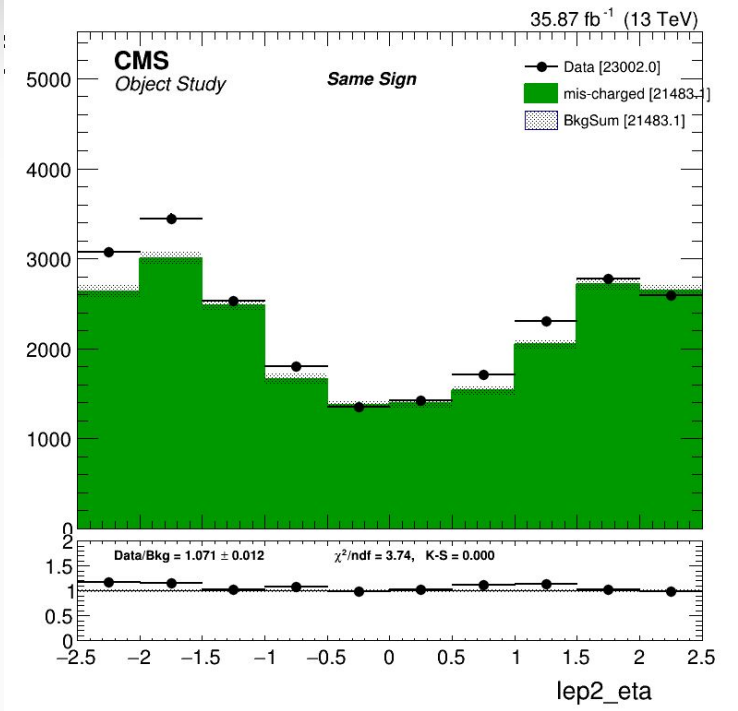


Before

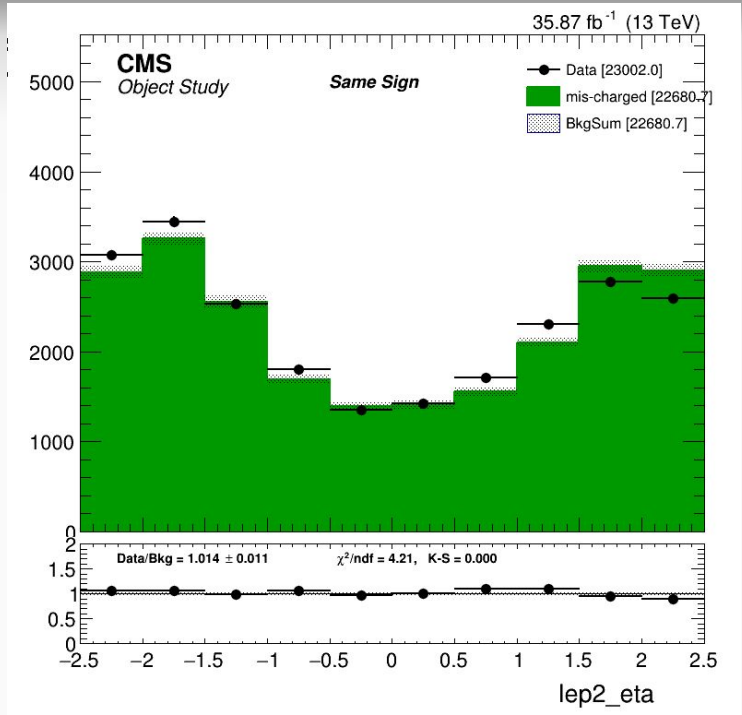
SF(pt,eta) x MC_SS != DATA_SS

After

SF Validation (2016)



x SF(pt,eta)

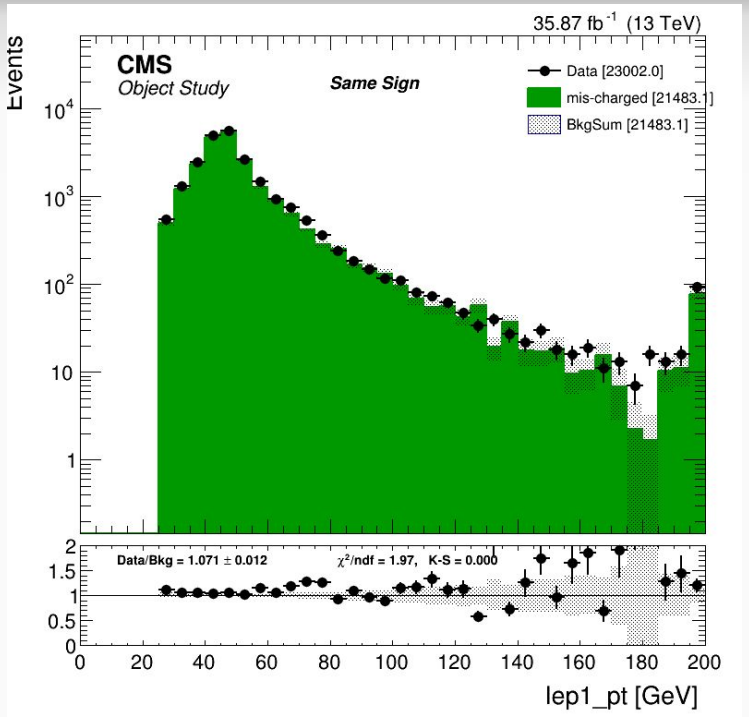


Before

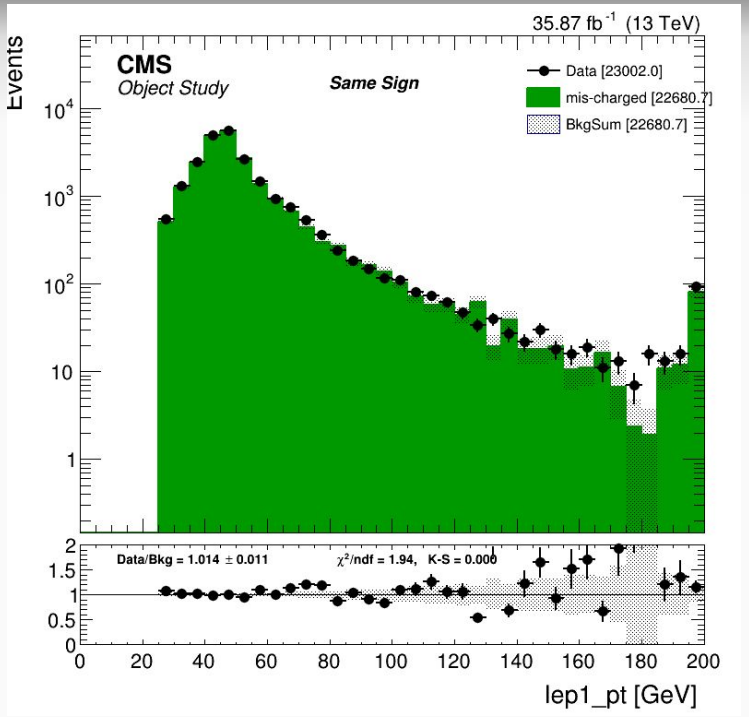
SF(pt,eta) x MC_SS != DATA_SS

After

SF Validation (2016)



x SF(pt,eta)

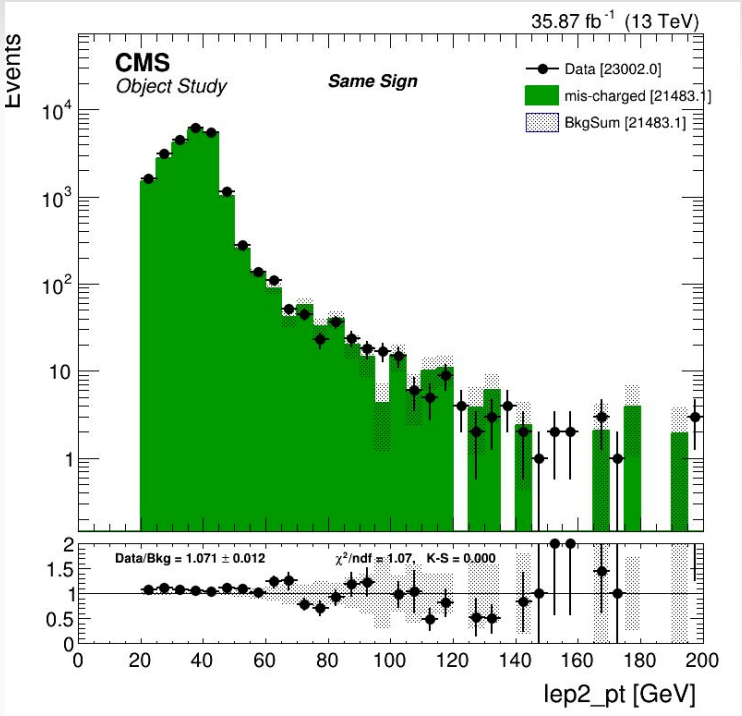


Before

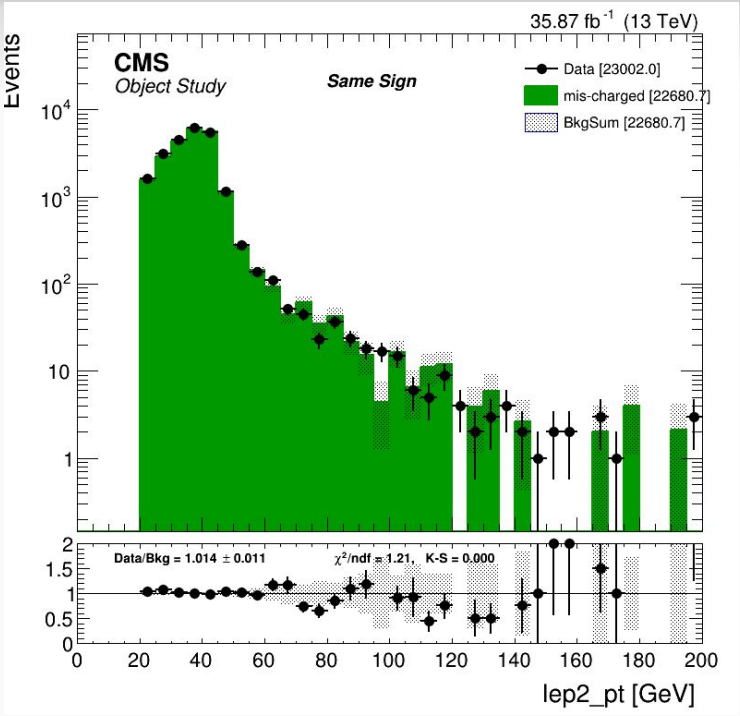
SF(pt,eta) x MC_SS != DATA_SS

After

SF Validation (2016)



x SF(pt,eta)



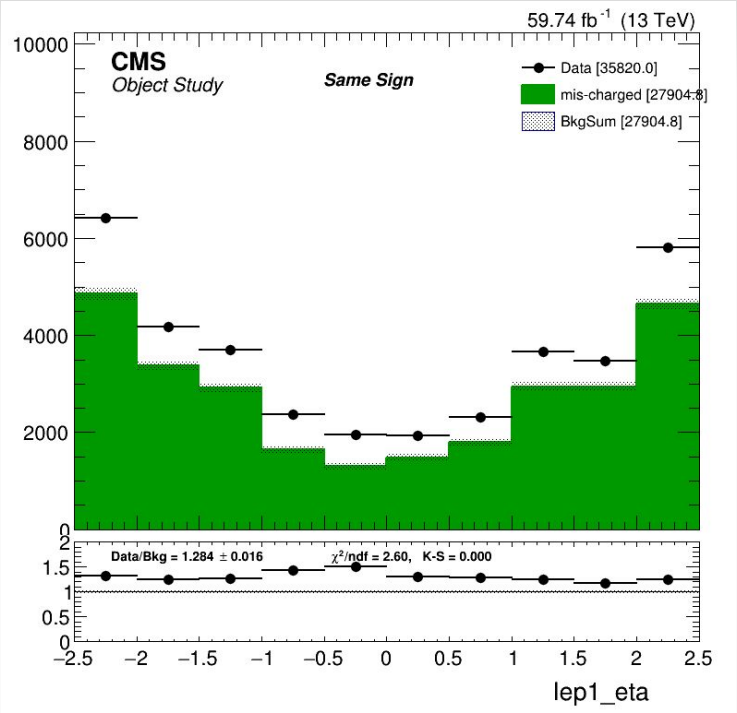
Before

After

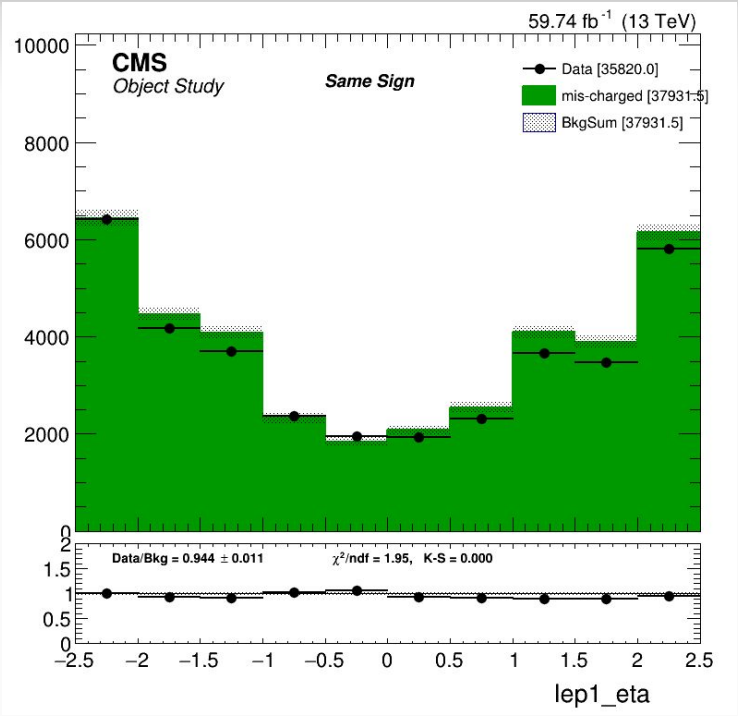
SF(pt,eta) x MC_SS != DATA_SS

Validation 2018

SF Validation (2018)



x SF(pt,eta)

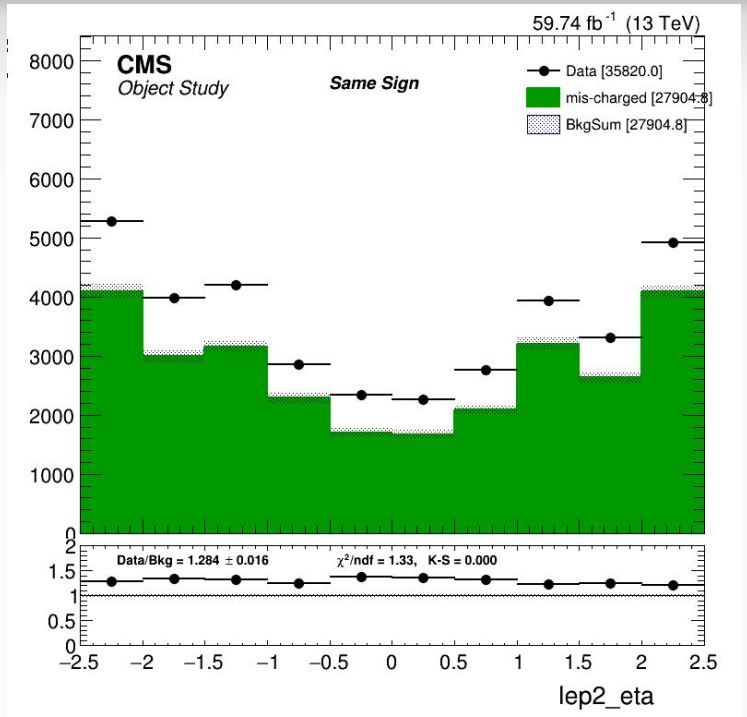


Before

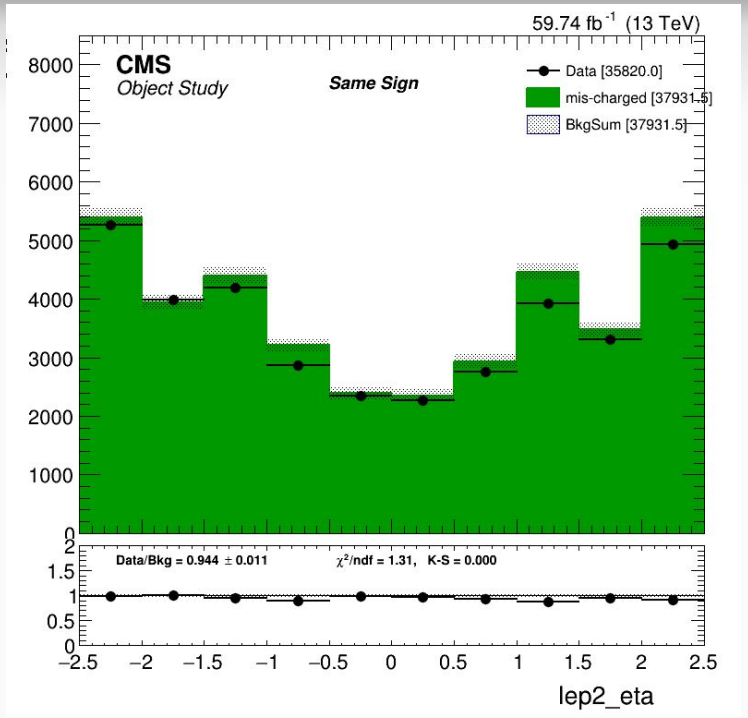
After

SF(pt,eta) x MC_SS != DATA_SS

SF Validation (2018)



x SF(pt,eta)

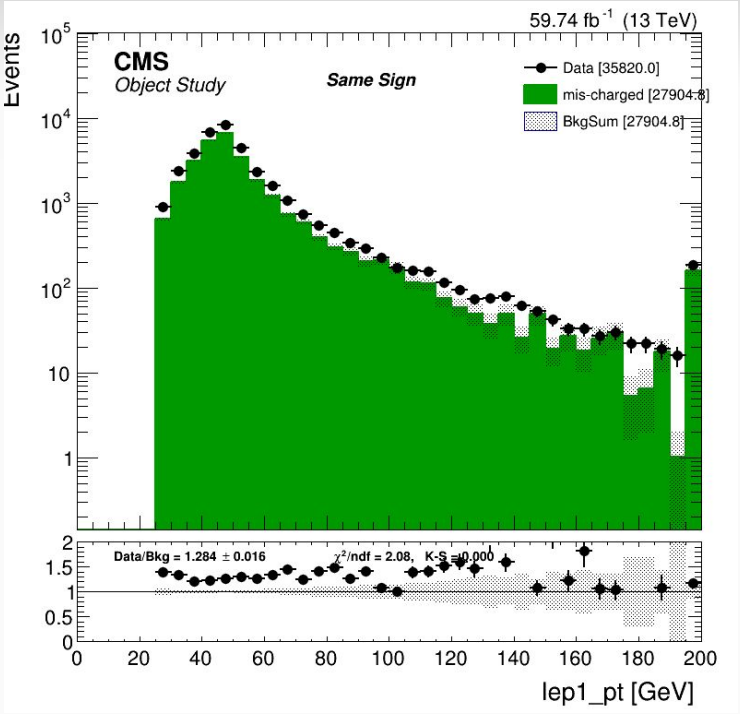


Before

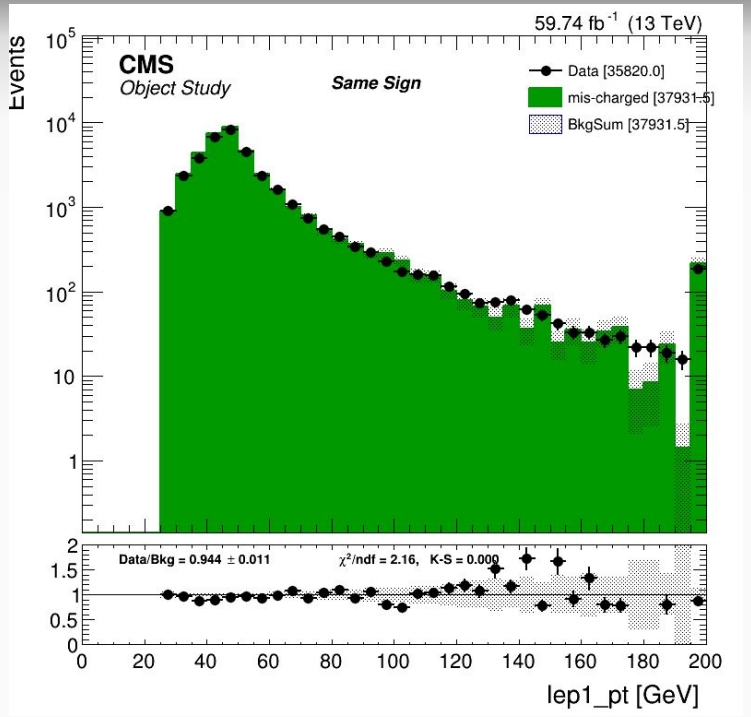
SF(pt,eta) x MC_SS != DATA_SS

After

SF Validation (2018)



x SF(pt,eta)

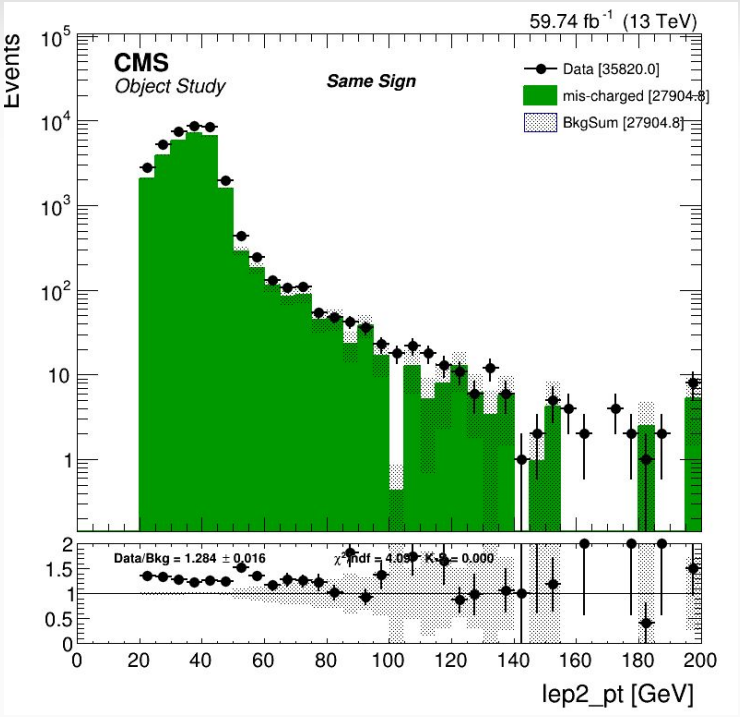


Before

SF(pt,eta) x MC_SS != DATA_SS

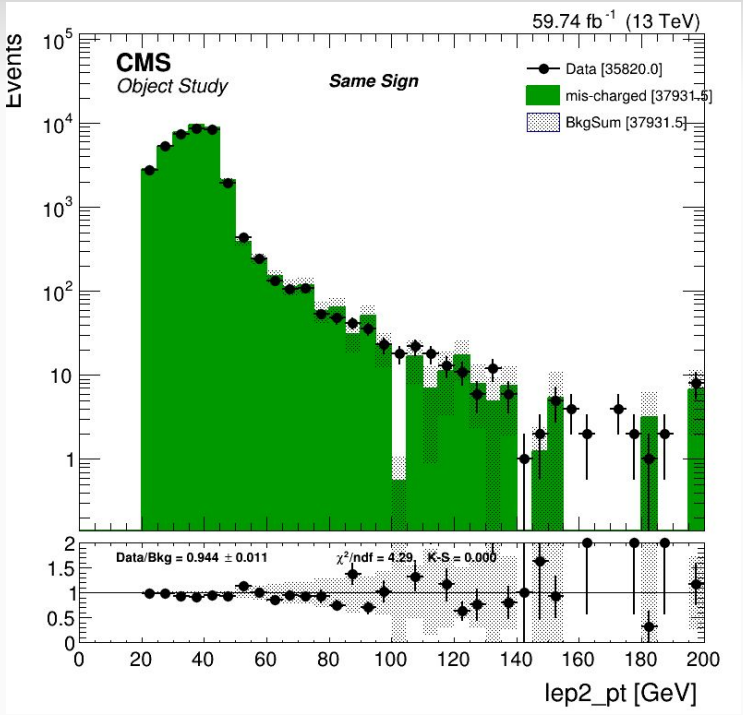
After

SF Validation (2018)



Before

x SF(pt,eta)



After

SF(pt,eta) x MC_SS != DATA_SS