The Best Higgs Mass Measurement Ever

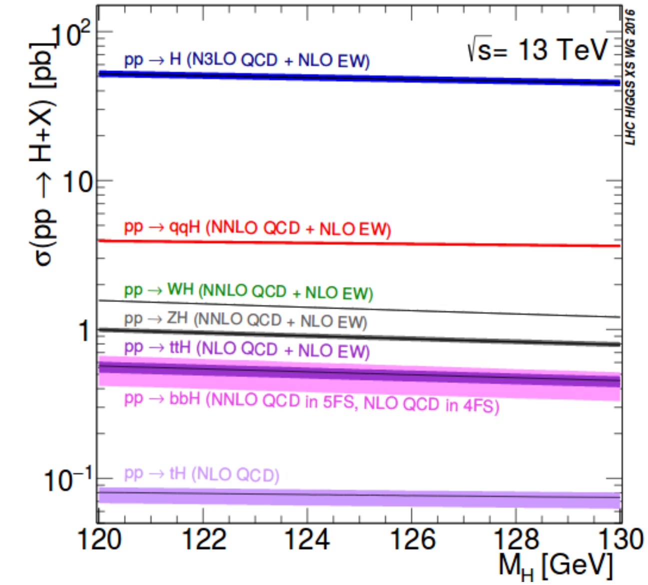
Rajdeep Chatterjee, Roger Rusack, Neil Schroeder

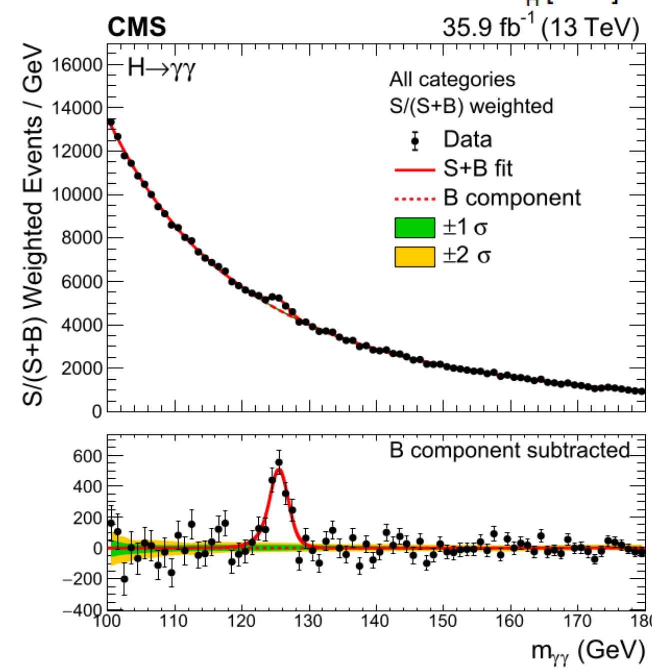
UMN CMS Symposium, August 2021

Introduction to $H \to \gamma\gamma$

Introduction to the Higgs Boson

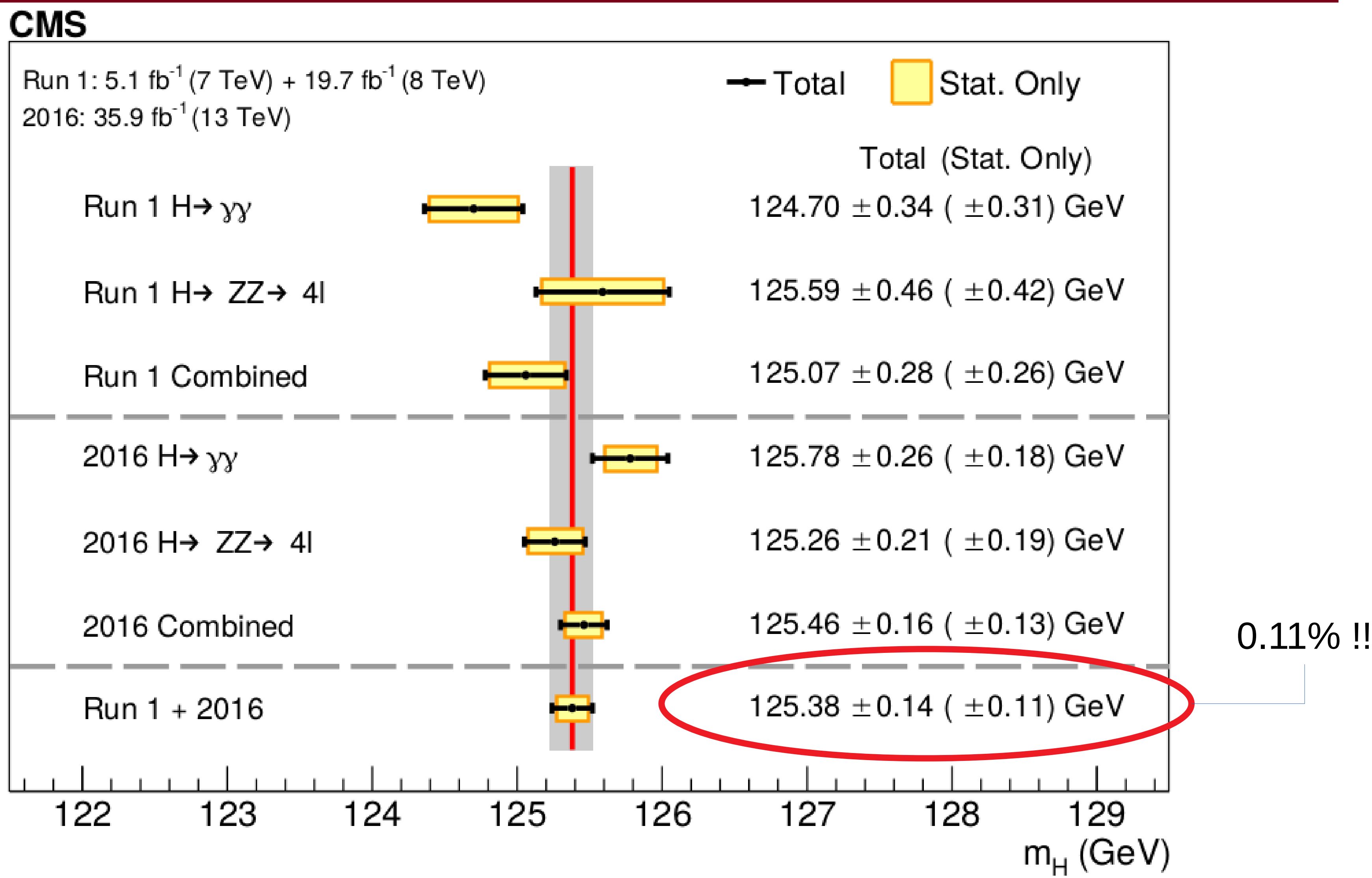
- The Higgs mass is <u>not predicted by theory</u>, but all properties of the Higgs boson are a function of M_H
  - This motivates a precision measurement
- The two most common precision measurement channels are $H \rightarrow ZZ^* \rightarrow 4l$ and $H \rightarrow \gamma\gamma$
- Although the branching ratio is small, the diphoton channel is fully reconstructed and the CMS ECAL was literally built for observation of the Higgs in this channel
- This channel shows up beautifully over the background





Current Results

Current Higgs Mass Measurements



- Results recently published in Phys Letters B

What’s Next?

Strategy in a Nutshell

- Using the **Full Run 2 dataset**, which has **4x as much data**
- We reconstruct diphoton objects:
  - **Single photons objects** are reconstructed from ECAL energy deposits
    - This needs to be done accurately
  - Decide the **diphoton vertex** using a BDT
  - Then we **combine them into a diphoton object**

$$M_{\gamma\gamma}=\sqrt{2E_{\gamma}^{1}E_{\gamma}^{2}(1-\cos(\theta_{12}))}$$

- If the diphoton vertex can be determined to **within 1cm**, the **uncertainty on the mass** is **dominated by the uncertainty on the energy** of the photons
  - If we can reduce the uncertainty on the energy scale we can reduce the uncertainty on the Higgs mass

Targeting Systematic Uncertainties

- The H→ γγ mass measurement is becoming a systematics limited analysis
  - The combined Run 2 dataset (2016, 2017, and 2018) is 137 fb^{-1} (13 TeV) which is almost 4 times as much data. Reduces Stat. Unc. by 2

Source	Contribution (GeV)
Electron energy scale and resolution corrections	0.10
Residual p_T dependence of the photon energy scale	0.11
Modelling of the material budget	0.03
Nonuniformity of the light collection	0.11
Total systematic uncertainty	0.18
Statistical uncertainty	0.18
Total uncertainty	0.26

Targeting Systematic Uncertainties

- The $H \rightarrow \gamma\gamma$ mass measurement is becoming a systematics limited analysis
- The electron energy scale and resolution corrections are one of the leading sources of systematic uncertainty in the measurement
  - Partially because the energy resolution directly affects the systematic uncertainty

Source	Contribution (GeV)
Electron energy scale and resolution corrections	0.10
Residual p_T dependence of the photon energy scale	0.11
Modelling of the material budget	0.03
Nonuniformity of the light collection	0.11
Total systematic uncertainty	0.18
Statistical uncertainty	0.18
Total uncertainty	0.26

- By having more stable corrections and a better handle on the uncertainty in the method of deriving them we can reduce their contribution to the systematic uncertainty on the mass measurement.

Residual Scales and Additional Smearings

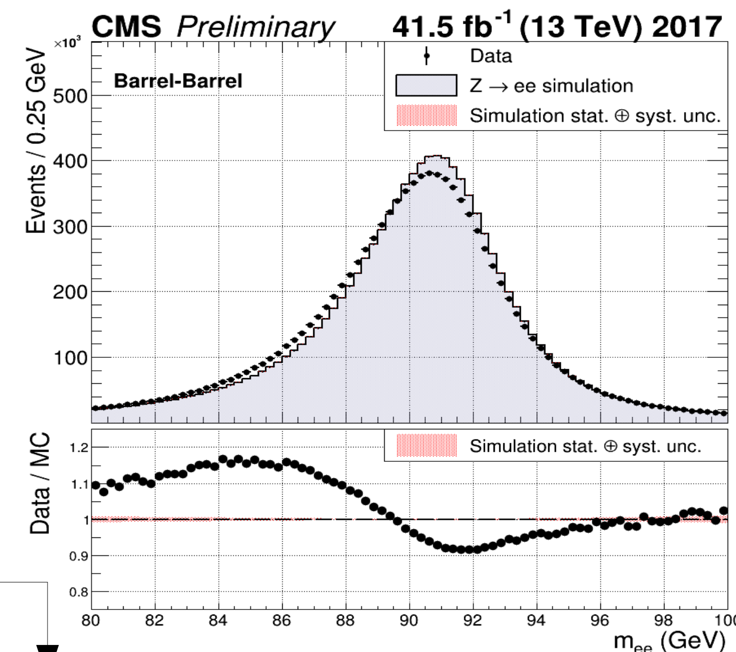
E/γ Object Calibration

Simulate CMS and its response as accurately as possible

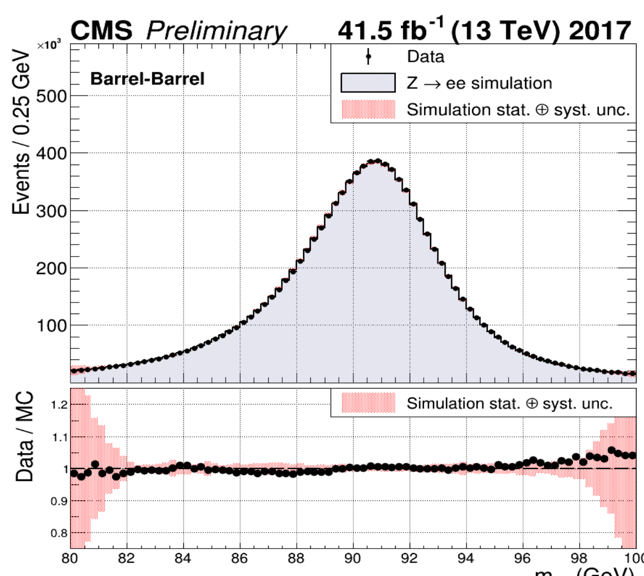
Generate Z → ee Events and Response from detector

Train regression To match detector Response to True energy of e

Apply regression To data and simulation



Correct regression For residual Difference between Data and simulation



Apply residual Energy corrections To **photons**

Apply residual Energy corrections To **electrons**

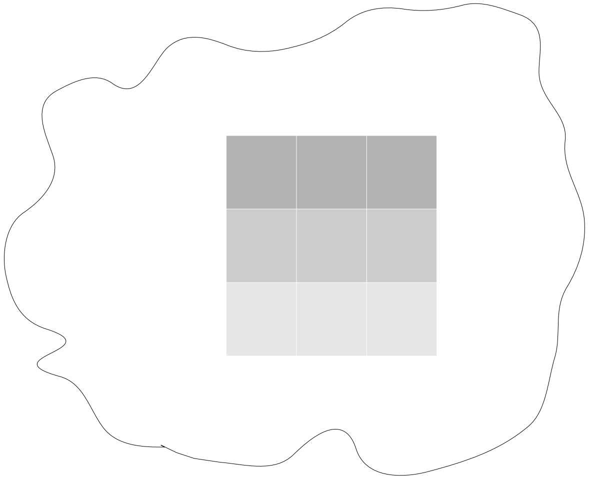
Correcting for Residual Differences in Data and MC

Residual Scales and Additional Smearings

- We target residual difference in data and MC in the single electron energy scale through the Z→ee process
- **Scale** electron energy in data:
  - $M_{ee}^{scaled} = M_{ee}\sqrt{(1+\Delta P_{e1})(1+\Delta P_{e2})}$
  - Where ΔP_{ei} is the scale shift w.r.t MC for electron *i*
- **Smear** the electron energy in MC:
  - $M_{ee}^{smeared} = M_{ee}\sqrt{Gaus(1,\Delta C_{e1}) * Gaus(1,\Delta C_{e2})}$
  - Where Δc_{ei} is the additional smearing for electron *I*
- The variables ΔP_{ei} and ΔC_{ei} are determined by minimizing a global binned NLL of the invariant mass for each dielectron category.
- The scales and additional smearings are derived in the following steps
  - Bins of **Run x η (time)**
  - Bins of **η x R_9 (location)**
  - Bins of **E_T (scale)**

Residual Scales and Additional Smearings

- We target residual difference in data and MC in the single electron energy scale through the Z→ee process
- **Scale** electron energy in data:
  - $$M_{ee}^{scaled} = M_{ee}\sqrt{(1+\Delta P_{e1})(1+\Delta P_{e2})}$$
  - Where ΔP_{ei} is the scale shift w.r.t MC for electron *i*
- **Smear** the electron energy in MC:
  - $$M_{ee}^{smeared} = M_{ee}\sqrt{Gaus(1,\Delta C_{e1}) * Gaus(1,\Delta C_{e2})}$$
  - Where Δc_{ei} is the additional smearing for electron *I*
- The variables ΔP_{ei} and ΔC_{ei} are determined by minimizing a global binned NLL of the invariant mass for each dielectron category.
- The scales and additional smearings are derived in the following steps
  - Bins of **Run x η (time)**
  - Bins of **η x R_9 (location)**
  - Bins of **E_T (scale)**



$R_9 = E_{3x3}/E_{total}$

Extracting the Scales and Smearings

SS_PyMin

- Negative Log Likelihood (NLL) minimization framework for the residual scales and additional smearings which has the following features:

SS_PyMin

- NLL minimization framework for the residual scales and additional smearings which has the following features:
  - Simultaneous minimization of residual scales *and* additional smearings across all dielectron invariant mass categories, i.e. EB-EB, EB-EE, and EE-EE
    - Binned NLL is evaluated in each category using a template fit of data to MC

SS_PyMin

- NLL minimization framework for the residual scales and additional smearings which has the following features:
  - Simultaneous minimization of residual scales *and* additional smearings across all dielectron invariant mass categories, i.e. EB-EB, EB-EE, and EE-EE
    - Binned NLL is evaluated in each category using a template fit of data to MC
  - Minimizer uses SciPy's optimize.minimize function and the L-BFGS-B algorithm. [1][2]

SS_PyMin

- NLL minimization framework for the residual scales and additional smearings which has the following features:
  - Simultaneous minimization of residual scales *and* additional smearings across all dielectron invariant mass categories, i.e. EB-EB, EB-EE, and EE-EE
    - Binned NLL is evaluated in each category using a template fit of data to MC
  - Minimizer uses SciPy’s optimize.minimize function and the L-BFGS-B algorithm. [1][2]
  - MC is oversampled the mitigate statistical fluctuations in the NLL profile

SS_PyMin

- NLL minimization framework for the residual scales and additional smearings which has the following features:
  - Simultaneous minimization of residual scales *and* additional smearings across all dielectron invariant mass categories, i.e. EB-EB, EB-EE, and EE-EE
    - Binned NLL is evaluated in each category using a template fit of data to MC
  - Minimizer uses SciPy's optimize.minimize function and the L-BFGS-B algorithm. [1][2]
  - MC is oversampled the mitigate statistical fluctuations in the NLL profile
  - Auto-binning of invariant mass histograms using the Freedman-Diaconis rule to mitigate stat issues

SS_PyMin

- NLL minimization framework for the residual scales and additional smearings which has the following features:
  - Simultaneous minimization of residual scales *and* additional smearings across all dielectron invariant mass categories, i.e. EB-EB, EB-EE, and EE-EE
    - Binned NLL is evaluated in each category using a template fit of data to MC
  - Minimizer uses SciPy's optimize.minimize function and the L-BFGS-B algorithm. [1][2]
  - MC is oversampled the mitigate statistical fluctuations in the NLL profile
  - Auto-binning of invariant mass histograms using the Freedman-Diaconis rule to mitigate stat issues
  - Invariant mass histograms are built using Numba to speed up the minimization

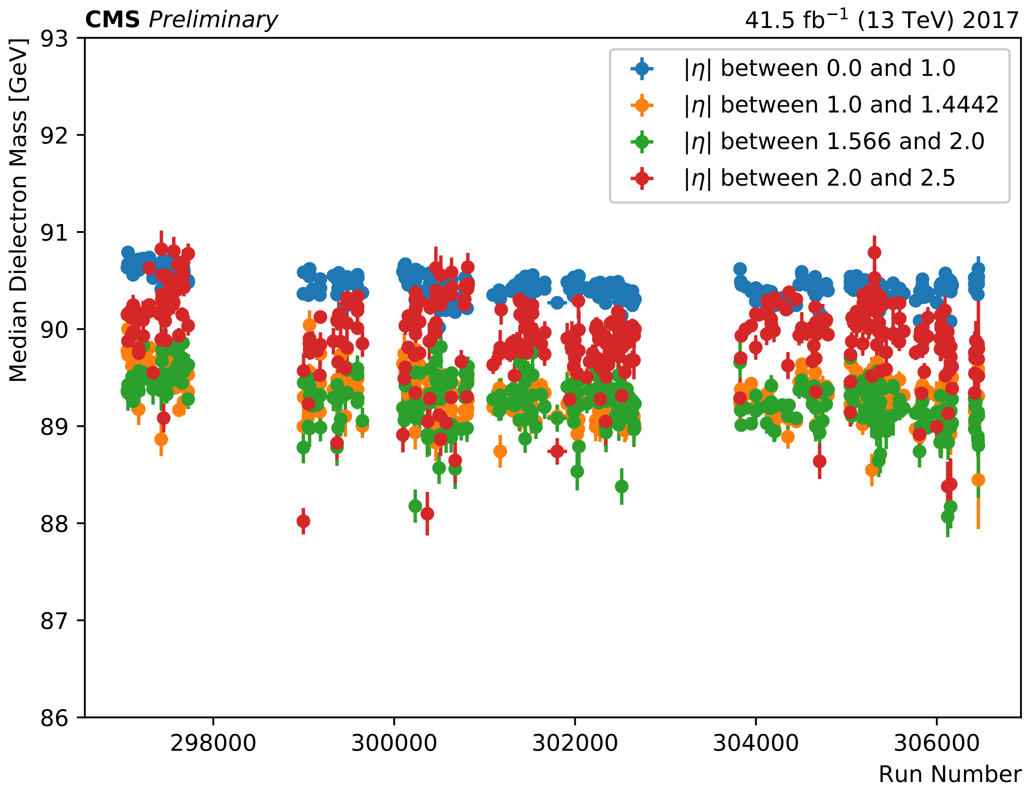
UL 2017 Results

Technical Details

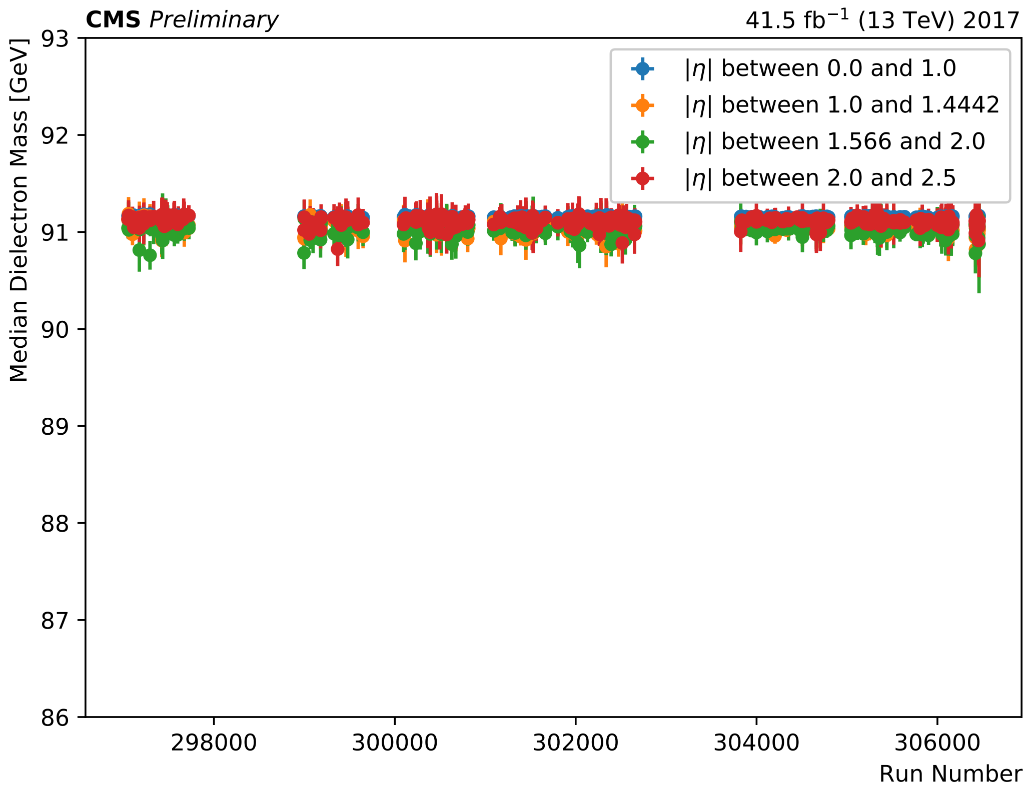
- Data: electrons from Z decays
  - Residual scales are applied here
- MC: DY + Jets
  - Additional smearings are applied here
- Transition region of barrel and endcap are excluded and $|\eta| < 2.5$
- Et cuts:
  - Leading electron: 32 GeV
  - Subleading electron: 20 GeV
- $80 < M_{ee} < 100$

Run x $|\eta|$ Corrections

UL 2017 Run x |η| Corrections



BEFORE



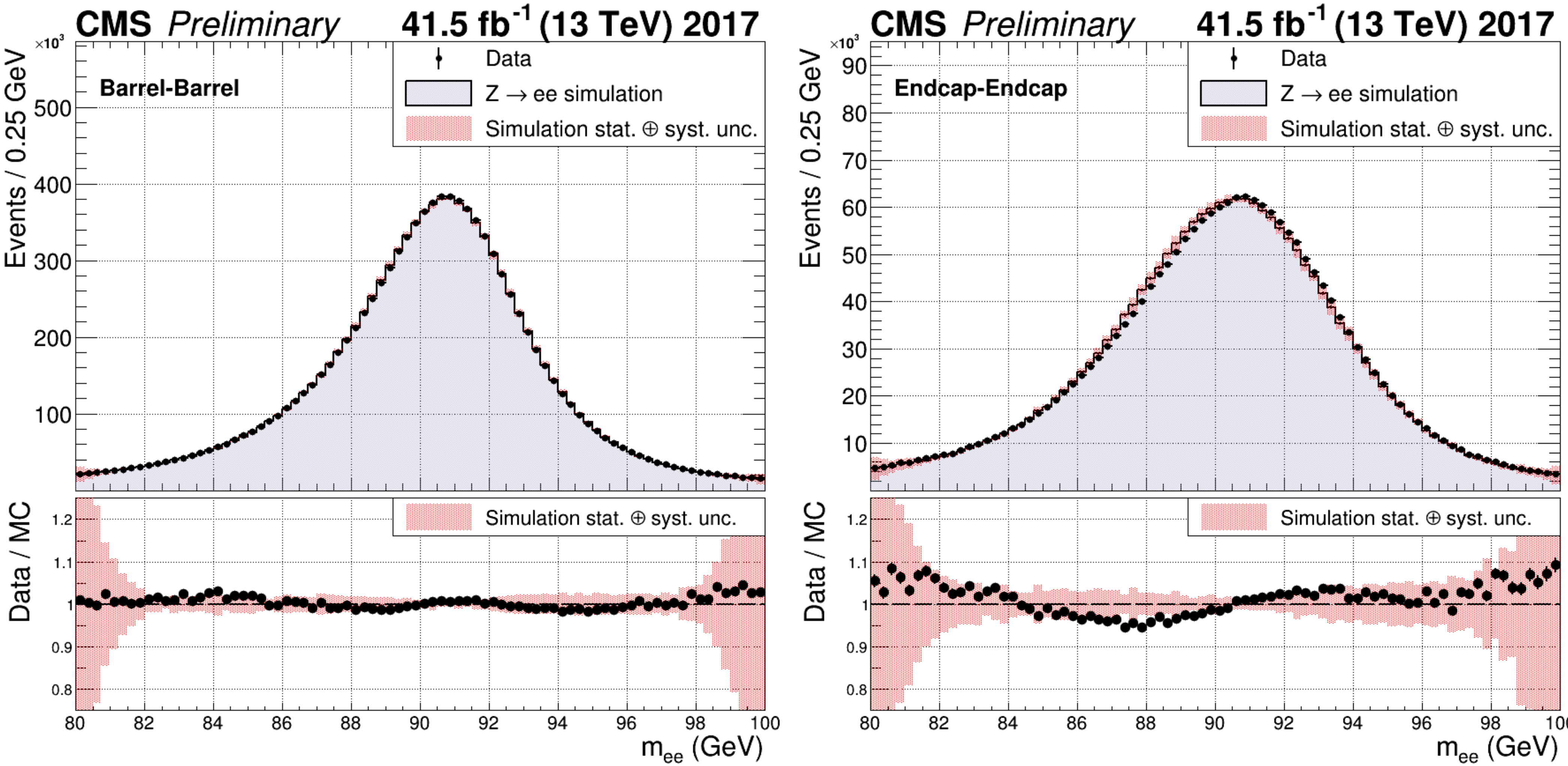
AFTER

Each point here represents ~10,000 events, or roughly 1 Fill

$|\eta|$ x R_9 Corrections

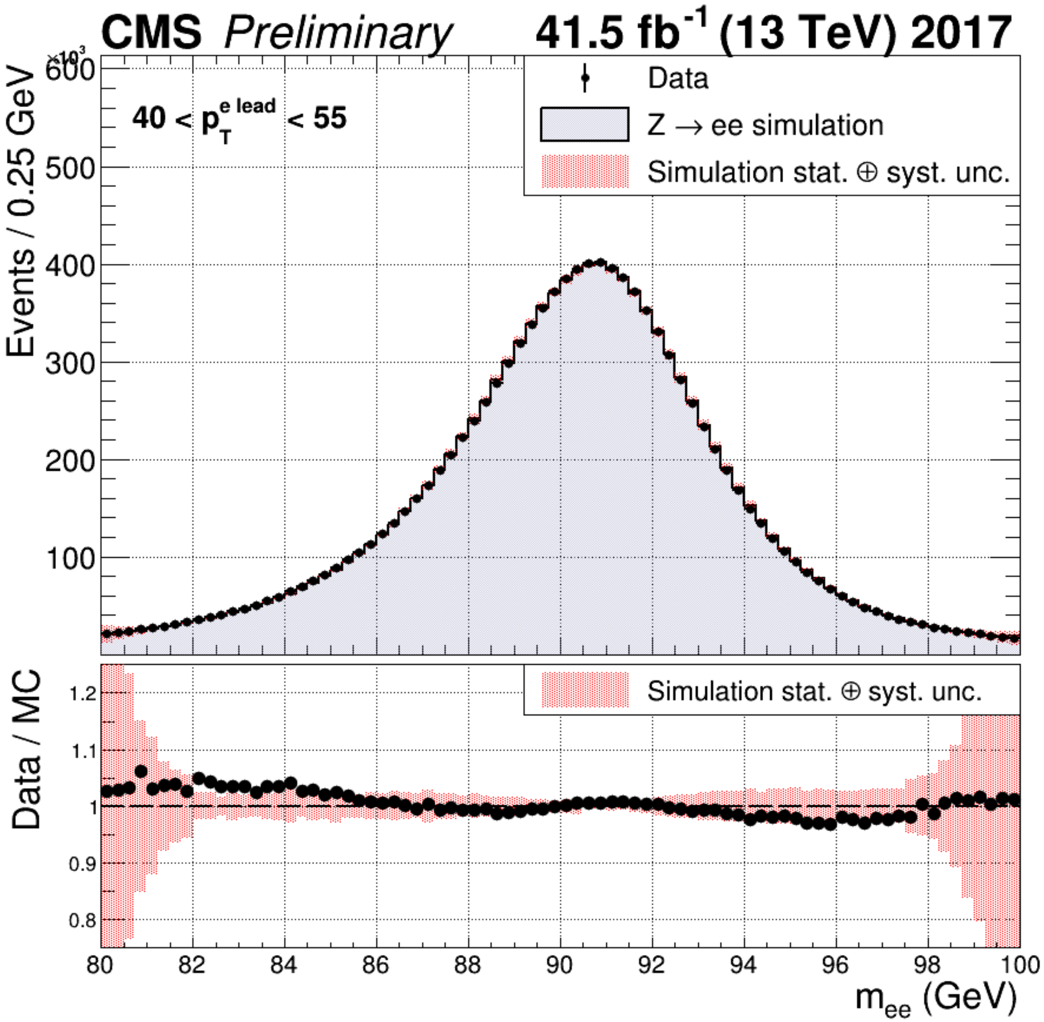
Invariant Mass Data/MC

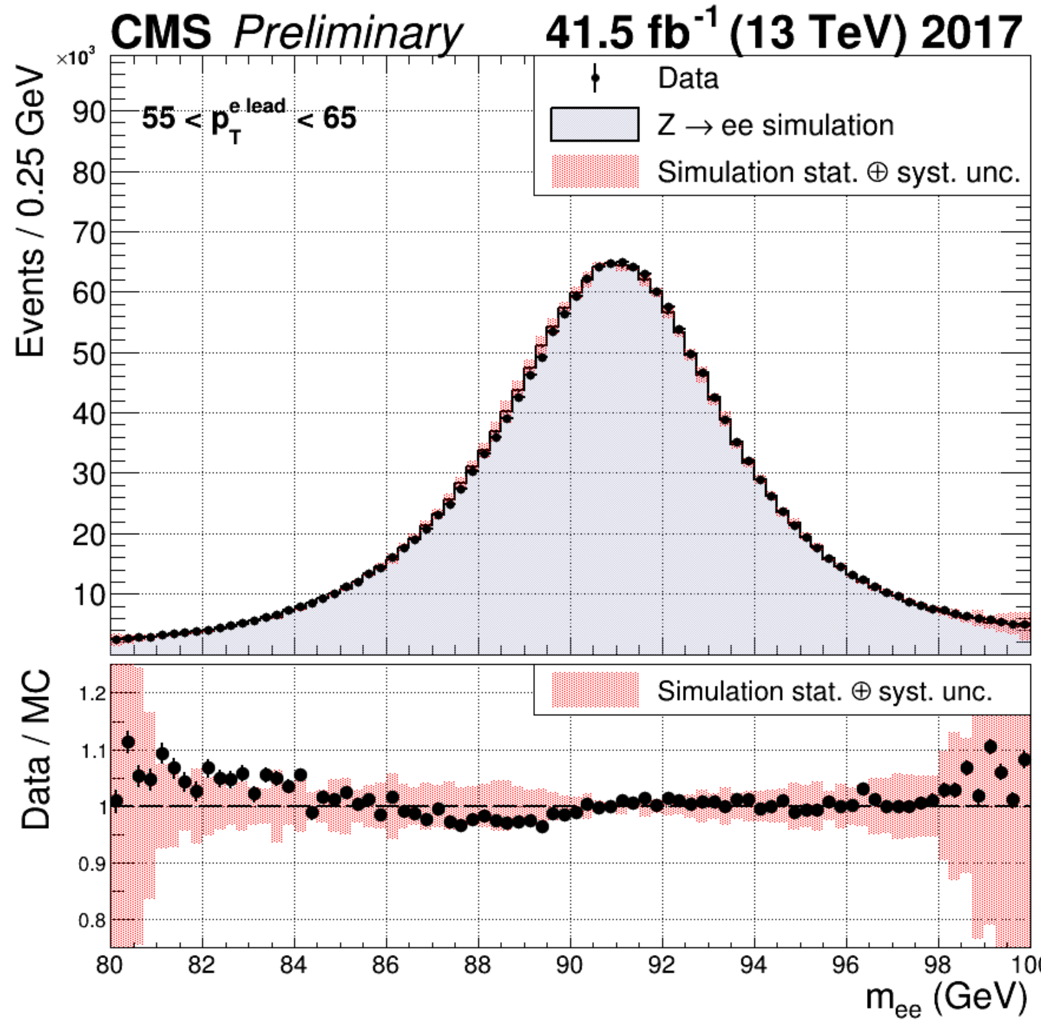
UL2017 Data/MC Agreement



- Some clear room for improvement here:
  - Scale in EE has lots of room for improvement
  - Smearings in EB can improve a bit more as well

UL2017 Data/MC Agreement





- Some clear room for improvement here:
  - Scale in the E_T categories has room for improvement
  - Smearings here have room for improvement as well

Stochastic (E_T) Residual Scales and Additional Smearings

Overview

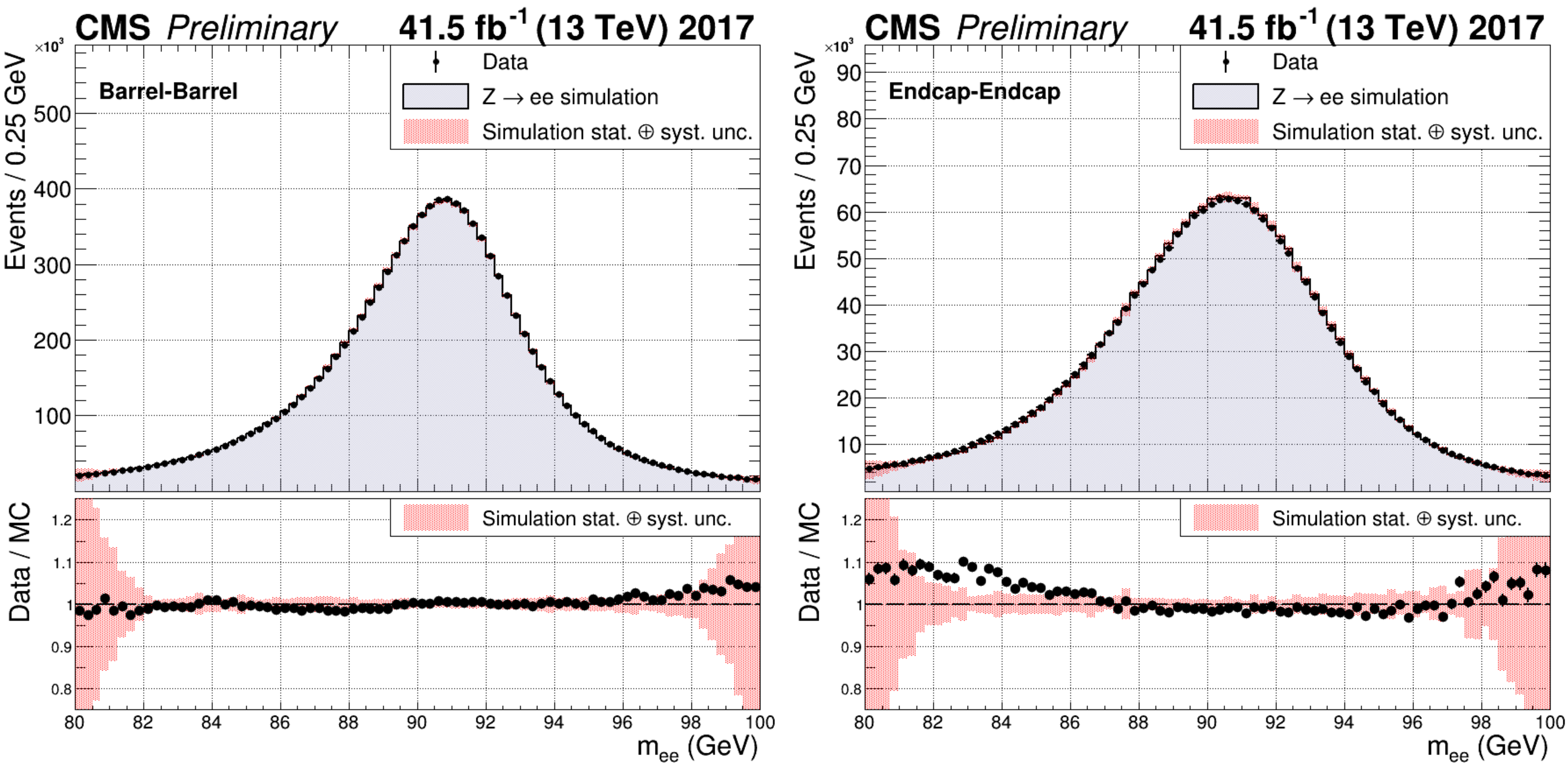
- The ECAL resolution and response varies with electron/photon E_T
- Additionally, the electron energy spectrum is different from the photon energy spectrum
  - Hence the need for a **dedicated set of E_T dependent residual scales is well motivated**
- For the additional smearings, we've already seen that there are trends in the agreement of the width of data and MC which depend on E_T.
  - Also, the choice of paramaterization for the additional smearings is arbitrary, so we have motivation to bin the additional smearings in E_T

Run 2 Approach

- For the precision measurement we derive a set of residual scales with very high granularity in R_9
  - 5 bins in $|\eta|$
  - 12 bins in R_9
    - R_9 distributions are different for photons and electrons and the residual scale can vary by as much as 2% across R_9
  - 10 bins in E_T
    - E_T scale for electrons from Z bosons is different than the E_T scale of photons from H bosons
- After applying these residual scales we bin our constant term additional smearing as a function of E_T
  - Now smearings will be binned in 3 variables:
    - $|\eta|$: [(0, 1), (1, 1.4442), (1.566, 2.0), (2.0, 2.5)]
    - R_9: [(0, 0.96), (0.96, ∞)]
    - E_T: [(20, 39), (39, 50), (50, 65), (65, 14000)]

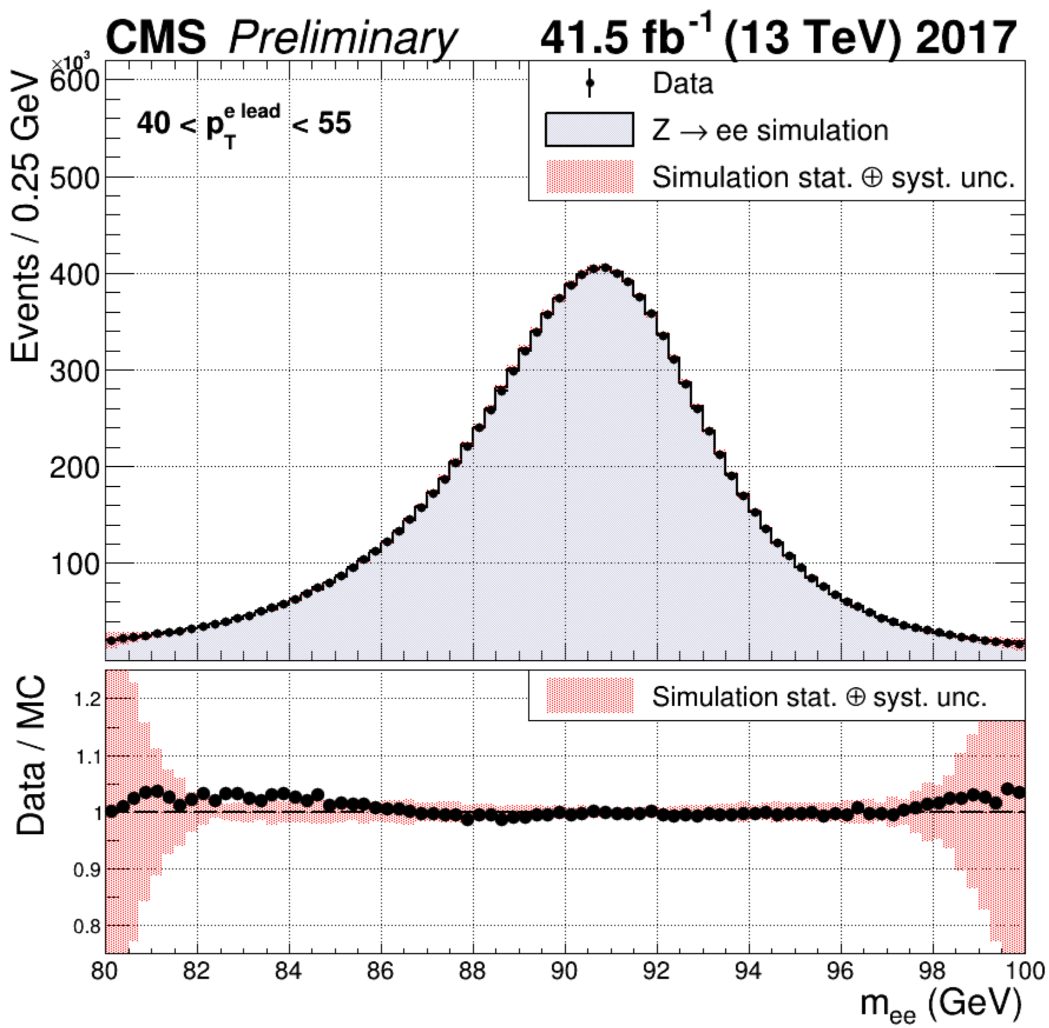
Results for UL 2017

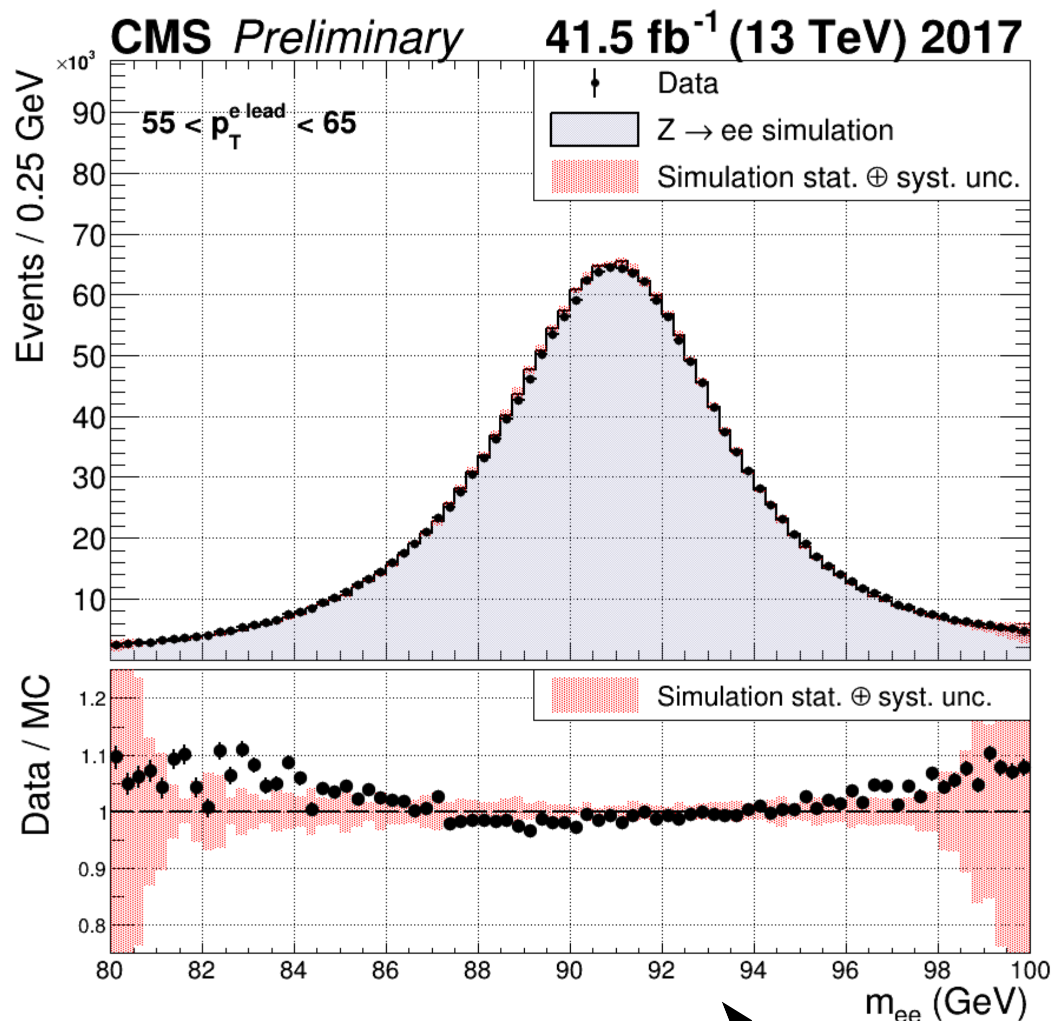
UL 2017 Data/MC Agreement



- Some clear improvements here:
  - Scale in EE has improve a lot, especially in the core of the distribution
  - Smearings in EB have flattened out and agreement is excellent

UL 2017 Data/MC Agreement



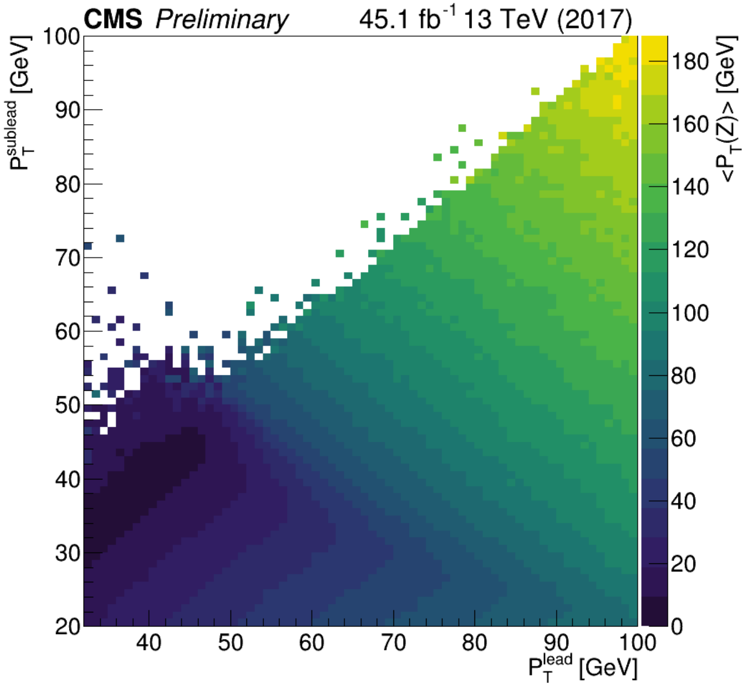


- Some clear room for improvement here:
  - Scale in the E_T categories has improved greatly, especially in our “target” category
  - Smearings here have also improved greatly

Scales/Smearings versus $P_T(Z)$

Approach

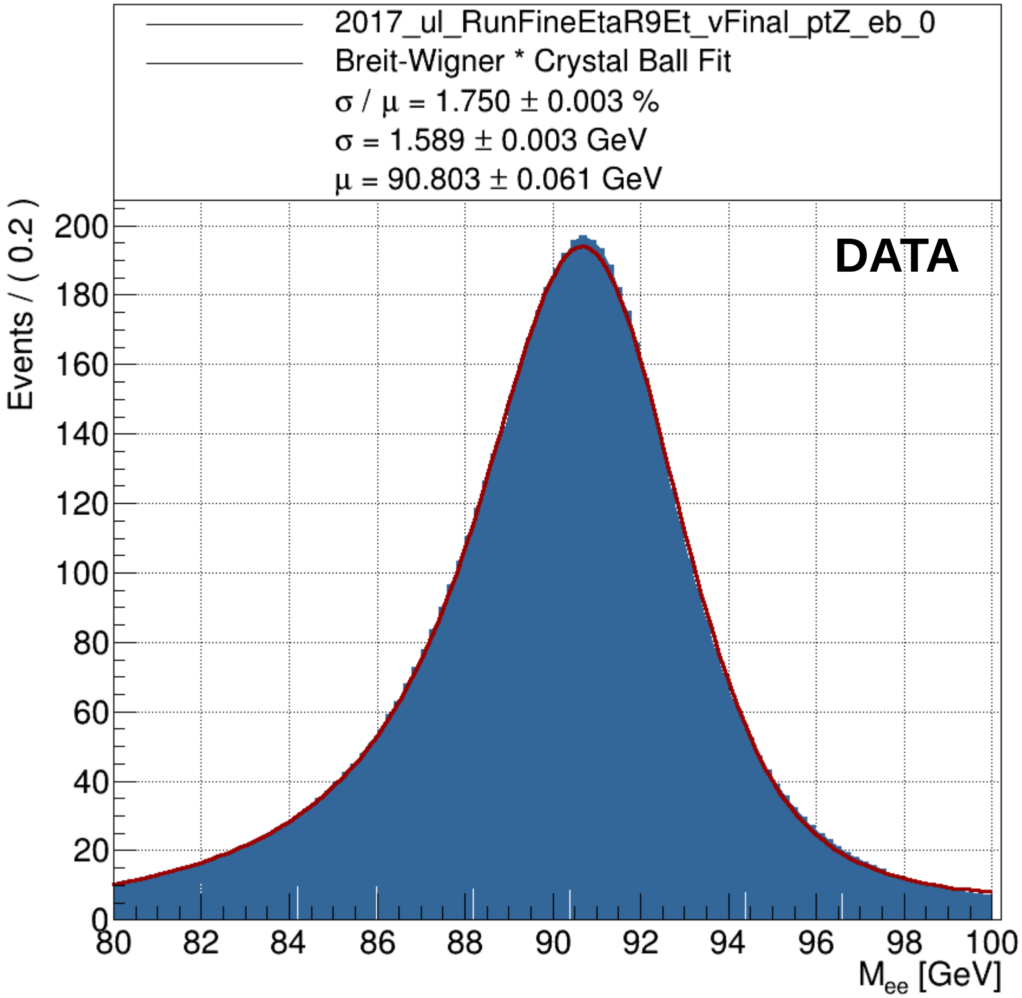
- We are extracting the width of data and MC as obtained from the fit in bins of Pt(Z)
  - Not the same as the electron energy but indicative of any non-closure with energy scale.

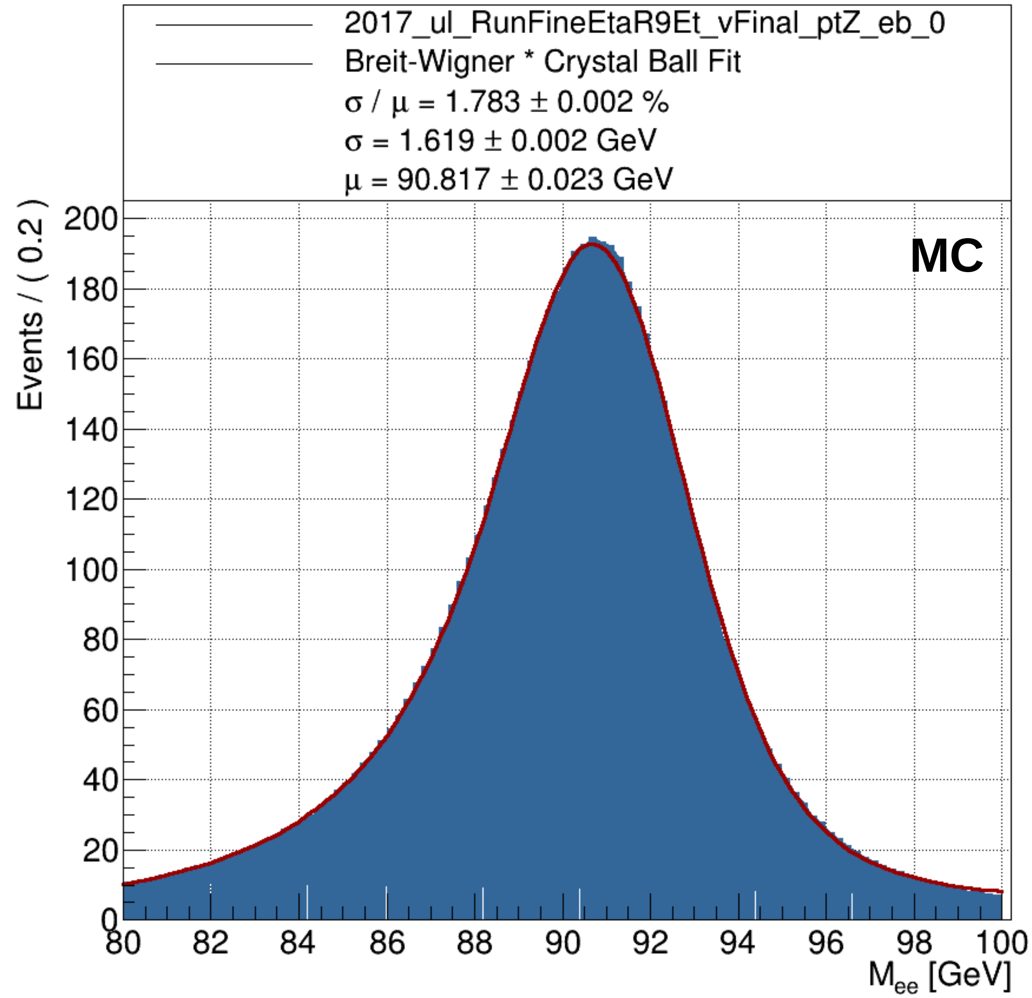


- The width of data and MC are compared in the following bins of Pt(Z):
  - [0, 20, 40, 60, 80, 100, 200, 400, 14000]

- We produce these width comparisons for EB-EB, and EE-EE plots

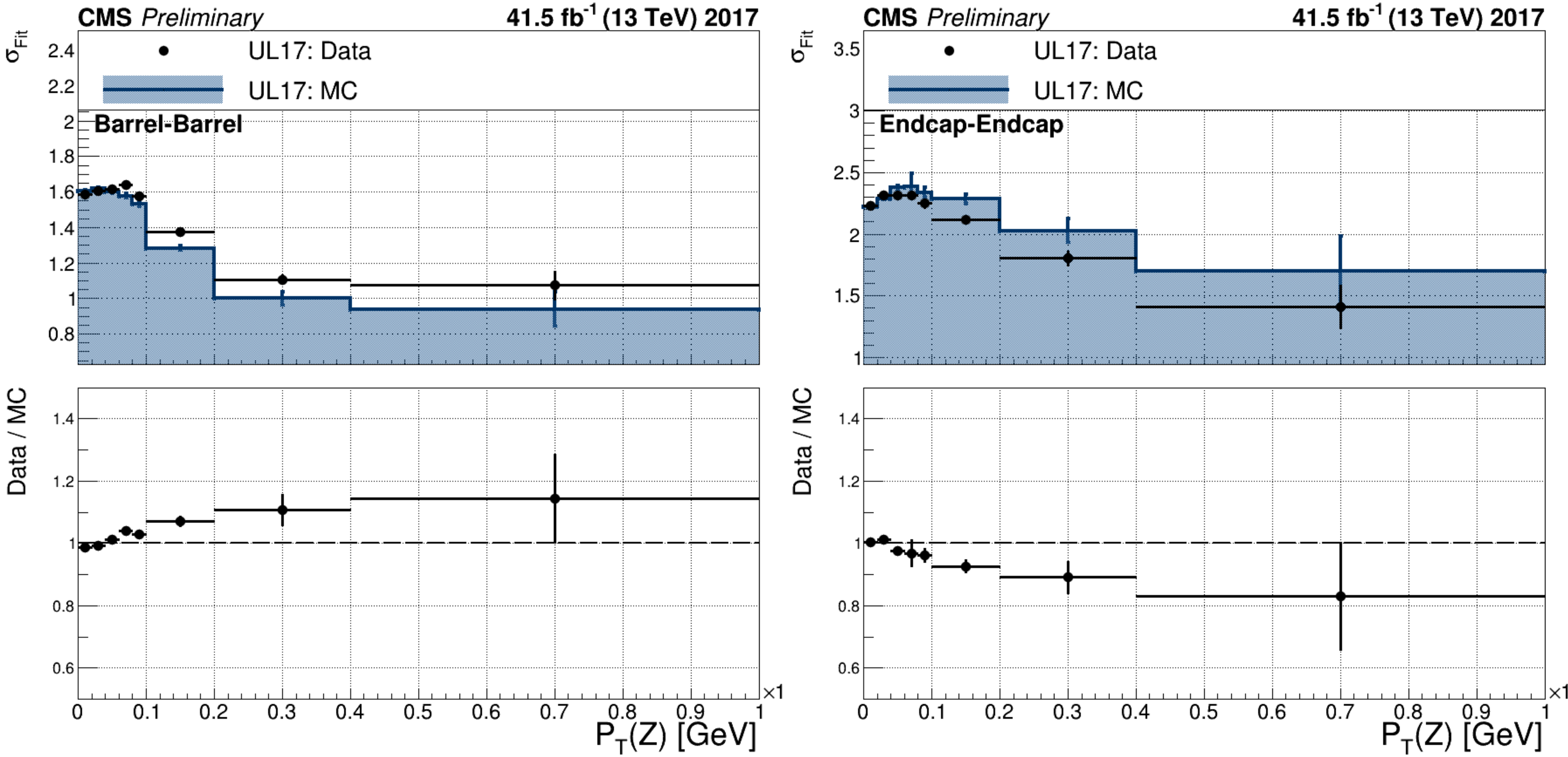
Barrel-Barrel, 0 < Pt(Z) < 20





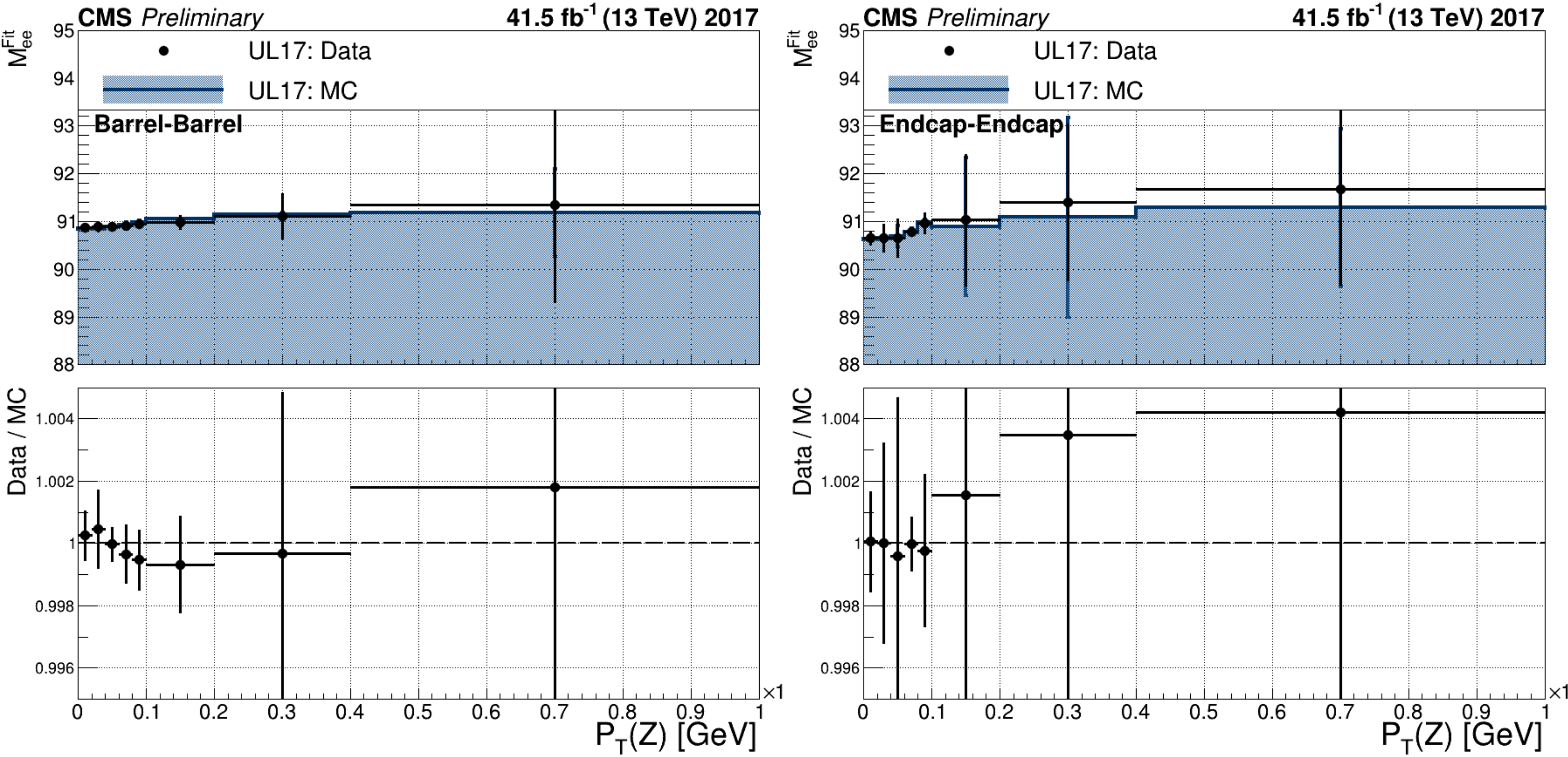
The Breit-Wigner Conv. Crystal Ball fit clearly matches the width and the peak of the distribution very well

UL 2017 Data/MC Agreement



- Previously, uncertainty on the E_T dependent smearings was 100% of the smearing.
- With our method we can conservatively assign the non-closure here as the uncertainty on the E_T dependent smearings, which are less than ~5% at $P_T(Z)$ < 100 GeV, and less than 20% for all $P_T(Z)$ > 100 GeV.

UL 2017 Data/MC Agreement



- Non-closure here is better than 0.05% for $P_T(Z)$ < 100 GeV in both the barrel and the endcap.
- At $P_T(Z)$ > 100 GeV the non-closure is better than 0.2% in the EB, and better than ~0.4% in the EE

What do we expect?

Expected Systematic Uncertainties

- Already the non-closure we're seeing in UL 2017 data indicates we'll be able to reduce these uncertainties substantially.

Source	Contribution (GeV)
Electron energy scale and resolution corrections	0.10
Residual p_T dependence of the photon energy scale	0.11
Modelling of the material budget	0.03
Nonuniformity of the light collection	0.11
Total systematic uncertainty	0.18
Statistical uncertainty	0.18
Total uncertainty	0.26

- With these reduced uncertainties we may be able to reach an uncertainty of ~150 MeV or better

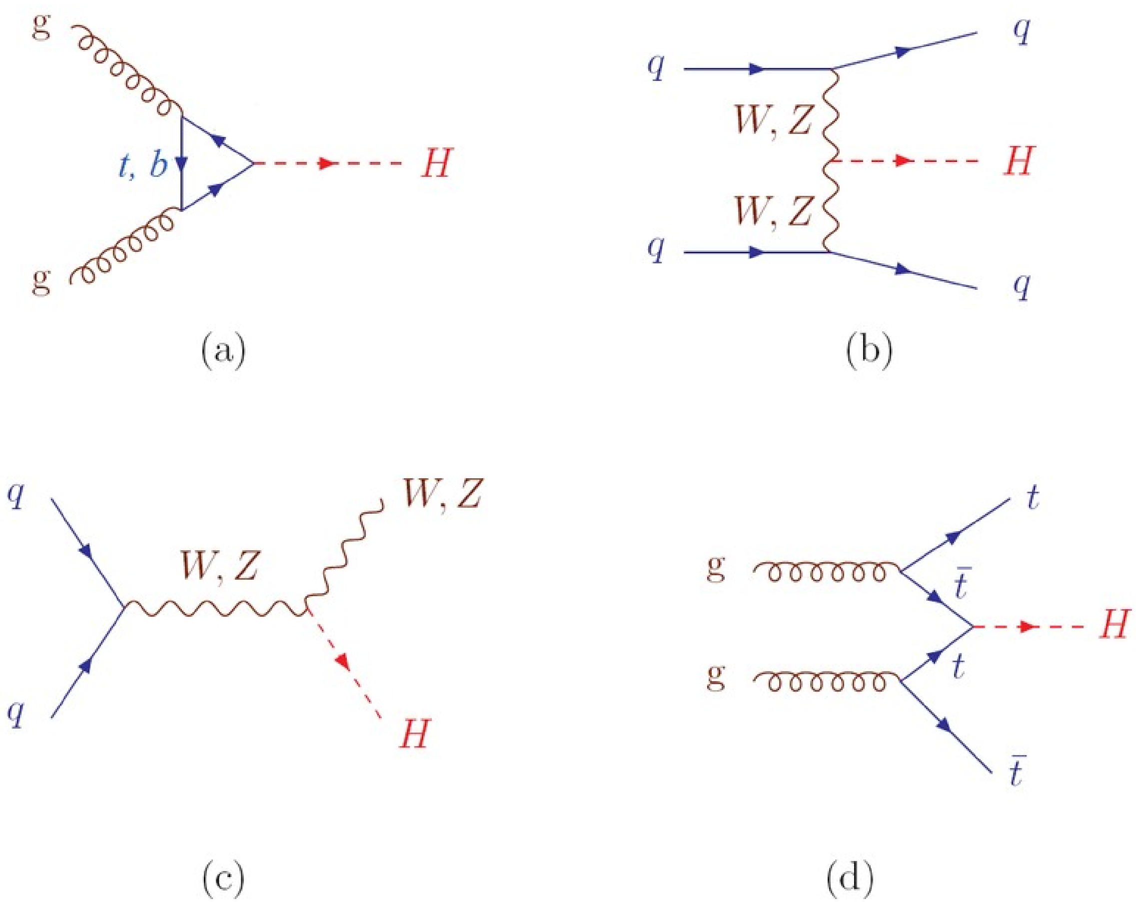
Summary

Summary

- Precision Higgs physics is doing well at CMS
- We're targeting systematic uncertainties on the mass of the Higgs boson as the areas we can most improve the precision
  - Specifically the scales and smearings are one are where we expect to do much better than we did in the previous analysis
- The scales and smearings for UL 2017 datasets were shown today, and they look great
  - Agreement of data and MC are much better in UL 2017 than was achieved in Legacy 16 (last mass measurement)
- We're currently in good standing with our estimate of the systematic uncertainties to measure the mass of the Higgs in the diphoton channel with a precision of ~150 MeV, an improvement over the earlier 260 MeV

Backup

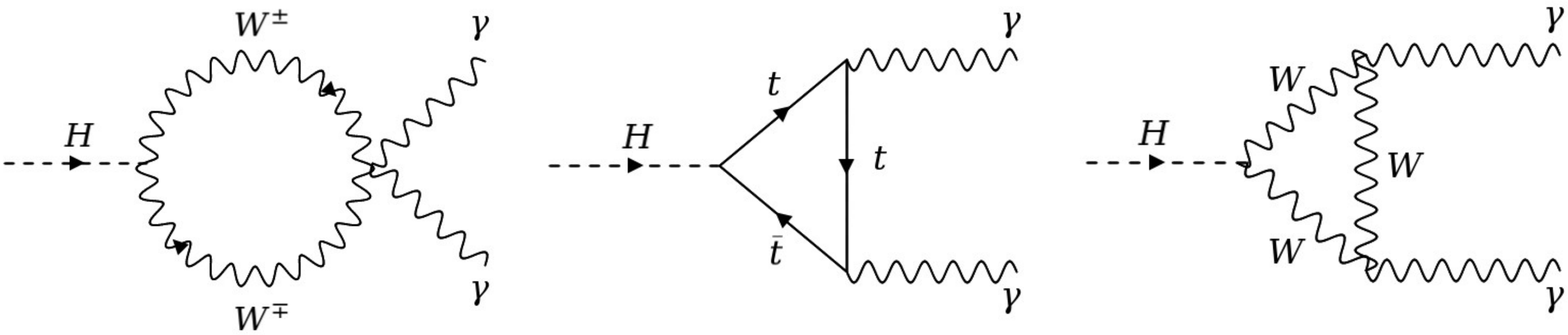
Higgs Production



- a) gluon gluon Fusion (ggH) **89%**
- b) Vector Boson Fusion (VBF) **5%**
- c) Higgstrahlung **4%**
- d) ttH **1%**

H → γγ

- The only way it can couple to two photons is through some interesting loop or triangle diagrams.
  - This process happens to 0.23 % of produced Higgs bosons



- This channel has very high precision.
  - Photon energy resolution of the CMS ECAL is 1-2%
  - The mass resolution is ~0.11% and shrinking

Extending Smearings to E_T

- The parameterization chosen, from Run 1, for the additional smearing was :

$$\Delta\sigma = \frac{\Delta S}{\sqrt{(E_T)}} \oplus \Delta C$$

The particular parameterization is arbitrary.

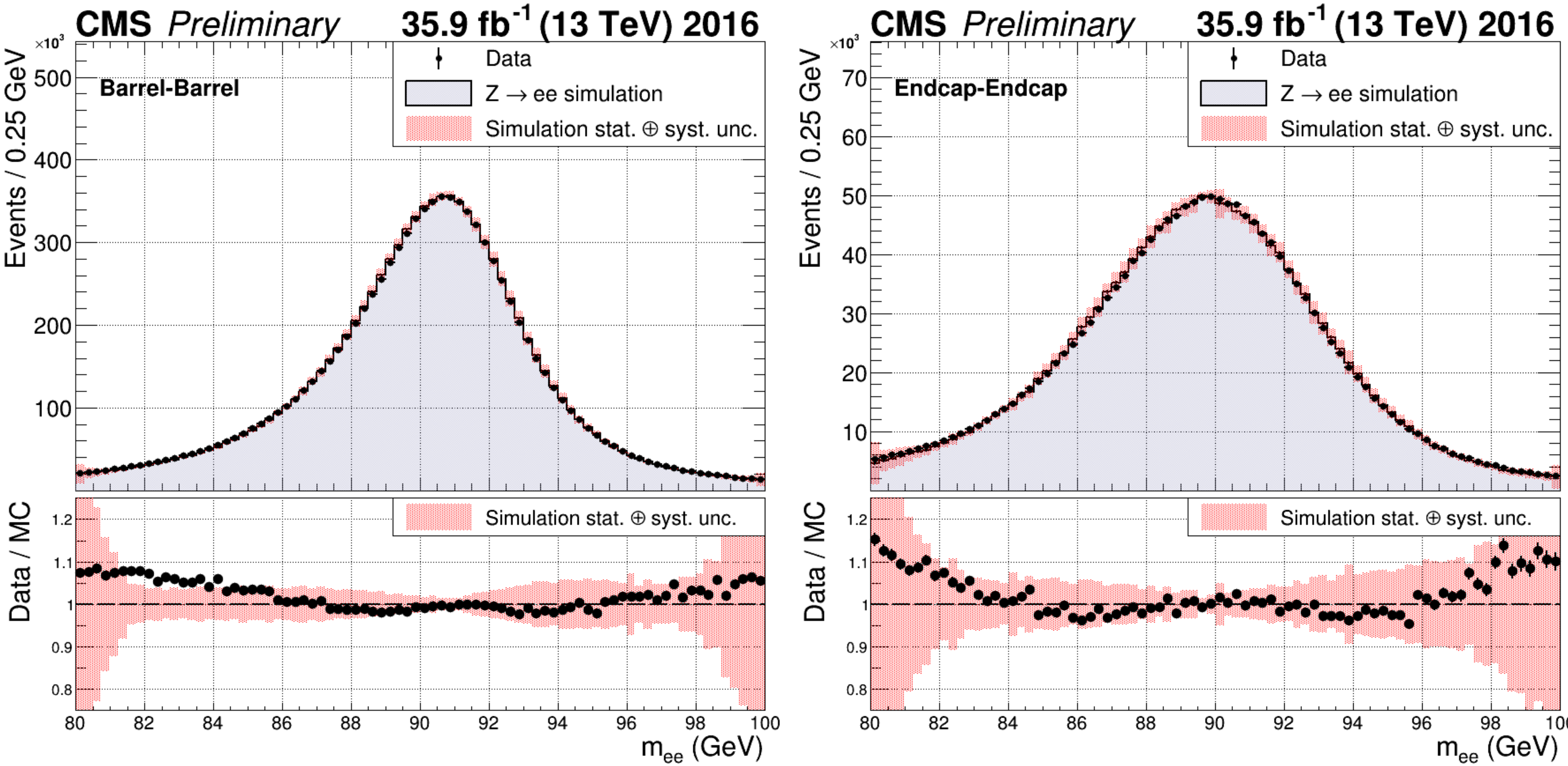
- A co-ordinate transformation is performed from the (ΔS, ΔC) plane into polar co-ordinates (ρ,φ) to decouple the correlation in the two terms
  - $\Delta C = \rho \cdot \sin\varphi$
  - $\Delta S = \rho \cdot \langle E_T \rangle \cdot \cos\varphi$, where $\langle E_T \rangle$ is the average E_T in the category where the smearing is being derived.
    - This requires the assumption that $\Delta\sigma = \Delta\sigma|_{\Delta S=0} = \Delta\sigma|_{\Delta C=0}$ i.e the additional smearing is the same whether it is derived as a pure constant term OR a pure stochastic term. [1]

- The smearing application framework is based on ρ, φ and the there are nuisances built into the signal model for both terms.

[1] Thesis: Shervin Nourbakhsh

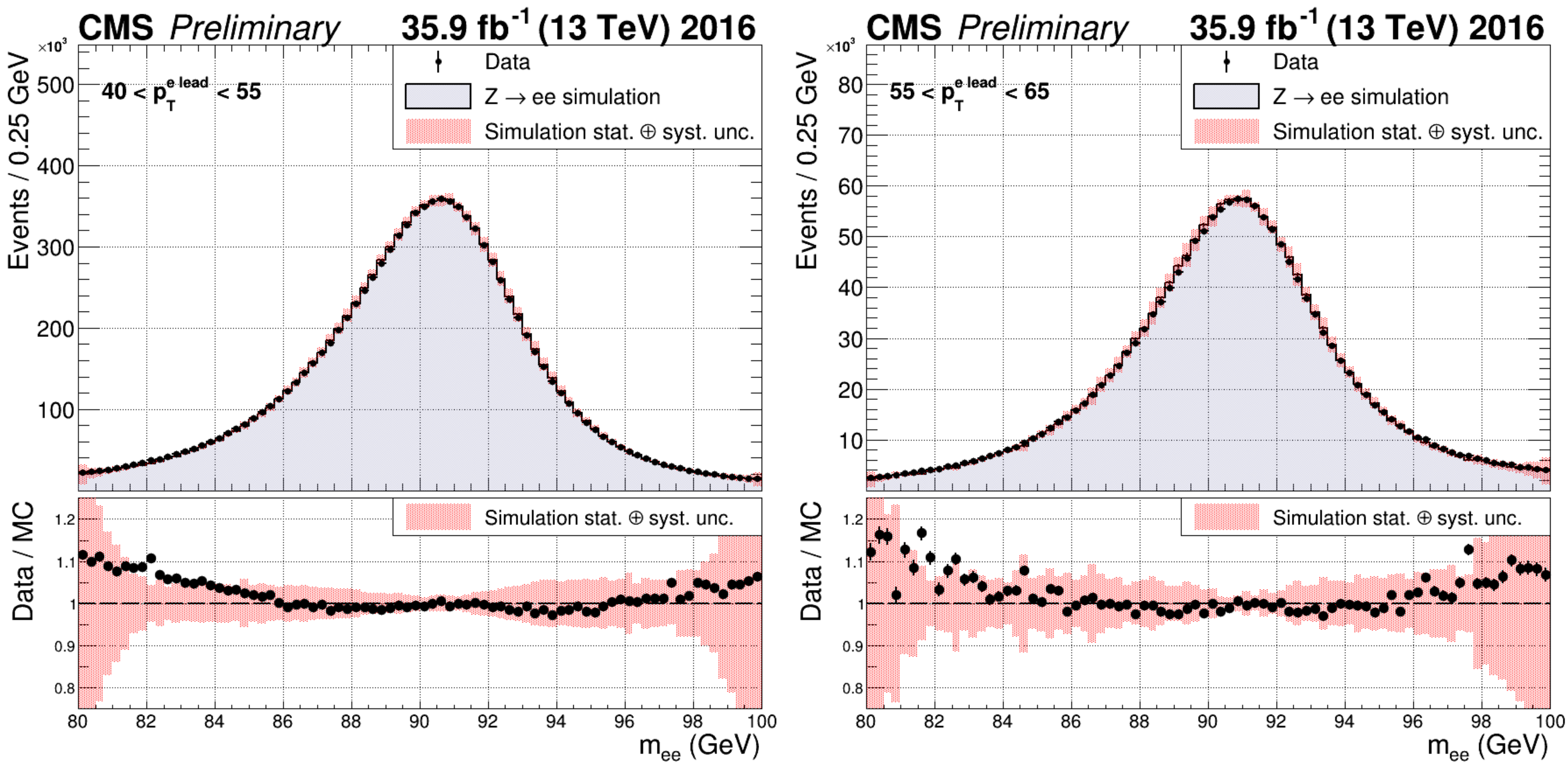
Comparison to Legacy 16

Legacy 2016 Data/MC Agreement



- Some clear differences here:
  - Scale in barrel in UL17 is much better than in Legacy 16
  - Smearings in endcap are much better in both barrel and endcap

UL2017 Data/MC Agreement



- Smearings are drastically improved from UL 16 and agreement in core and tails has improved slightly