

## **Parte III**

# **Photon shower shape corrections**

# SHOWER SHAPES Y LA IDENTIFICACIÓN DE FOTONES

# 4

*"Champions keep playing until they get it right."*

---

Billie Jean King

El **ECAL** se presentó brevemente en la Sección 2.2.3, donde se describió el mecanismo que utiliza para la medida de energía y posición de fotones y electrones. En este subdetector, los fotones depositan su energía mediante la creación de pares electrón-positrón y la radiación bremsstrahlung, creando una lluvia electromagnética (**EM**). El **ECAL** hace un buen trabajo para calcular la energía de la lluvia **EM**, pero identificar la partícula iniciadora sigue siendo una tarea difícil. Sin embargo, en virtud de las diferentes capas y granularidades en el **ECAL**, pueden estudiarse diferentes características de estas lluvias **EM**, codificadas por diferentes variables llamadas Shower Shapes (**SSs**).



El capítulo comienza con la descripción de todas las **SSs**, que son centrales para poder identificar a los fotones reales de los falsos. La optimización del algoritmo de identificación de fotones está basada en las **SSs**, y se encuentra descripta en la Sección 4.2. Además, en dicha sección, se presentan los métodos usados para la estimación de las eficiencias de identificación de fotones. Por último, en la Sección 4.3 se describe brevemente el mal modelado de las **SSs** en las simulaciones, un problema que tiene implicancias directas en los cálculos de eficiencias. Este mal modelado, sin embargo, será estudiado en detalle en el Capítulo 5.

## 4.1. Shower Shapes (**SSs**)

Como se menciona en la Sección 3.2.2, la identificación de fotones se basa en cortes rectangulares aplicados a las diversas **SSs** que conducen a una excelente capacidad de separación entre fotones reales aislados de fotones falsos procedentes de hadrones. Estas **SSs** se calculan a partir de los depósitos de energía de los candidatos a fotones en las celdas del **ECAL** y **HCAL**, y sirven para describir el paso de ellos a través de los calorímetros, caracterizando las lluvias **EM** laterales y longitudinales.

En general, los fotones reales producen depósitos de energía más estrechos en el **ECAL**, y tienen menores filtraciones hacia el **HCAL**, en comparación con aquellos fotones procedentes de hadrones, donde la presencia de hadrones vecinos adicionales cerca del fotón falso tienden a ensanchar las lluvias. Además, dado que la primera capa del **ECAL** presenta una gran resolución en  $\eta$ , es posible discriminar los candidatos a fotones procedentes de decaimientos  $\pi^0 \rightarrow \gamma\gamma$  que están caracterizados por dos máximos locales debidos a la presencia de dos fotones cercanos.

Tabla 4.1: Shower Shapes (SSs) utilizadas para la identificación de fotones. Las tres columnas de la derecha denotan si la variable es utilizada o no para los WPs *loose* (L), *medium* (M) o *tight* (T), descriptos en la Sección 4.2.2.

Categoría	Descripción	Nombre	L	M	T
Filtración hadrónica	Cociente entre el $E_T$ en la primera capa del <b>HCAL</b> y el $E_T$ del cluster <b>EM</b> ( $ \eta  < 0.8$ y $ \eta  > 1.52$ )	$R_{\text{had}1}$	✓	✓	✓
	Cociente entre el $E_T$ en todo el <b>HCAL</b> y el $E_T$ del cluster <b>EM</b> ( $0.8 <  \eta  < 1.37$ )	$R_{\text{had}}$	✓	✓	✓
<b>ECAL</b> (2 <sup>da</sup> capa)	Cociente entre la energía sumada en $3 \times 7$ celdas en $\eta \times \phi$ y la energía en $7 \times 7$ celdas, centradas alrededor del centro del cluster	$R_\eta$	✓	✓	✓
	Ancho lateral de la lluvia en dirección de $\eta$	$w_{\eta 2}$	✓	✓	✓
	Cociente de la energía sumada en $3 \times 3$ celdas en $\eta \times \phi$ y en $3 \times 7$ celdas, centradas alrededor del centro del cluster	$R_\phi$	✓	✓	
<b>ECAL</b> (1 <sup>ra</sup> capa)	Ancho lateral de la lluvia en 3 <i>strips</i> alrededor del máximo	$w_{\eta 1}$ or $w_1$	✓	✓	
	Ancho lateral total de la lluvia	$w_{\text{s tot}}$	✓	✓	
	Fracción de la energía fuera de las 3 strips centrales en un rango de 7 celdas, sobre la energía en las 3 celdas centrales	$f_{\text{side}}$	✓	✓	
	Diferencia entre la energía del segundo máximo con la energía mínima entre los dos primeros máximos.	$\Delta E$	✓	✓	
	Cociente de la diferencia de energía entre el primer y segundo máximo, sobre la suma de ambas energías	$E_{\text{ratio}}$	✓	✓	
	Cociente de la energía en la primera capa del <b>ECAL</b> y la energía total del cluster <b>EM</b>	$f_1$	✓	✓	

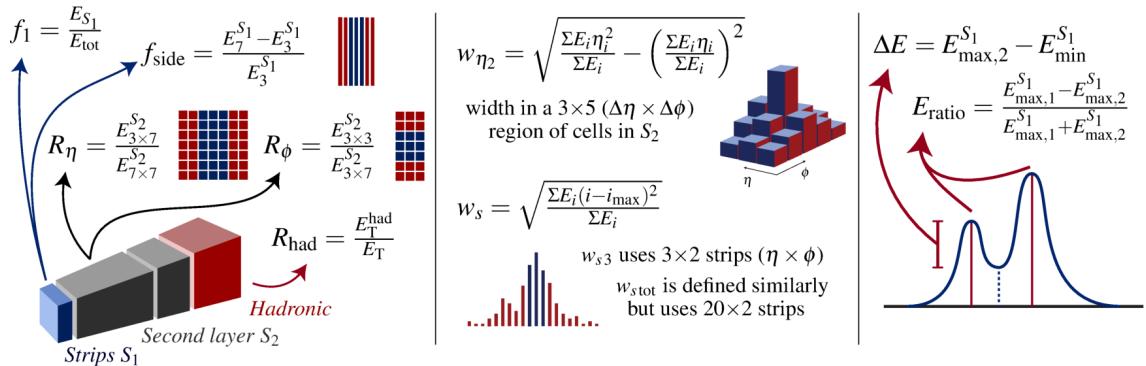


Figura 4.1: Representación esquemática de las SS de fotones. Los valores  $E_C^{S_N}$  representan la energía en la capa  $N$  del **ECAL** en un cluster  $C$ .

A continuación, se detallan las SSs utilizadas para la identificación de fotones, que se muestran resumidas en Tabla 4.1 y un esquema de cómo son calculadas se encuentra en la Figura 4.1. Las primeras variables hacen uso de la energía medida en el HCAL:

- Filtración hadrónica: es la energía **trasnversal** depositada en el HCAL, normalizada respecto a la energía depositada en el ECAL:

$$R_{\text{had}(1)} = \frac{E_{\text{T}}^{\text{had}}}{E_{\text{T}}^{\text{EM}}} \quad (4.1)$$

Para minimizar los efectos de la degradación de la resolución, en la región de transición barrel-endcap del HCAL ( $0.8 \leq |\eta| \leq 1.37$ ) se utiliza la energía depositada en todo el HCAL ( $R_{\text{had}}$ ). En el resto del detector, sólo se utiliza la energía depositada en la primera capa del HCAL ( $R_{\text{had } 1}$ ).

Las siguientes variables utilizan la información de la segunda capa del ECAL:

- Perfil de energía lateral en  $\eta$ :

$$R_{\eta} = \frac{E_{3 \times 7}^{s2}}{E_{7 \times 7}^{s2}} \quad (4.2)$$

donde  $E_{i \times j}^{s2}$  es la suma de energía en la segunda capa del calorímetro contenida en una ventana de  $i \times j$  celdas (unidades de  $\eta \times \phi$ ), centrada en la celda más energética. Esta variable da una medida del ancho de las lluvias en la dirección  $\eta$ .

- Perfil de energía lateral en  $\phi$ :

$$R_{\phi} = \frac{E_{3 \times 3}^{s2}}{E_{3 \times 7}^{s2}} \quad (4.3)$$

definida de forma similar a  $R_{\eta}$ . Sin embargo, esta variable se comporta de forma muy diferente para fotones convertidos y no convertidos. Debido a la acción del campo magnético, los electrones y positrones se curvan en direcciones opuestas en  $\phi$ , por lo que se producen lluvias EM más anchas para los fotones convertidos que para los no convertidos.

- Ancho de la lluvia lateral en  $\eta$ :

$$w_{\eta 2} = \sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left( \frac{\sum E_i \eta_i}{\sum E_i} \right)^2} \quad (4.4)$$

mide el ancho propio de la lluvia EM, donde  $E_i$  es la energía en la  $i$ -ésima celda del ECAL, medida en una ventana de  $3 \times 5$  celdas en  $\eta \times \phi$ .

Las siguientes variables utilizan la información de la primera capa del ECAL, compuesta por las celdas *strips* que permiten una alta resolución en  $\eta$  y permite una buena separación entre fotones aislados de fotones producto del decaimiento de  $\pi^0$ . La Figura 4.2 muestra la diferencia en la energía depositada en el ECAL entre los dos casos mencionados anteriormente.

- Perfil de energía lateral en  $\eta$ :

$$f_{\text{side}} = \frac{E_7^{s1} - E_3^{s1}}{E_3^{s1}} \quad (4.5)$$

mide la energía fuera del núcleo de las tres strips centrales dentro de una ventana de 7 celdas, dividida por la energía en las tres celdas centrales.

- Ancho de la lluvia lateral en  $\eta$  (3 strips)

$$w_1 = \sqrt{\frac{\sum E_i(i - i_{\max})^2}{\sum E_i}} \quad (4.6)$$

donde  $i$  corre sobre todas las celdas en una ventana de 3 celdas alrededor de la de mayor energía. Esta variable mide el ancho de la lluvia EM en la primera capa del calorímetro.

- Ancho de la lluvia lateral en  $\eta$  (total). Se define de forma similar a  $w_1$ , pero utiliza todas las celdas en una ventana de  $\Delta\eta \times \Delta\phi = 0.0625 \times 0.2$ , que corresponde aproximadamente a  $20 \times 2$  strips en  $\eta \times \phi$ .
- Diferencia energética

$$\Delta E = E_{\max,2}^{s1} - E_{\min}^{s1} \quad (4.7)$$

representa la diferencia de energía entre el segundo máximo y la energía mínima reconstruida entre los dos máximos de la primera capa del ECAL.

- Cociente de energía

$$E_{\text{ratio}} = \frac{E_{\max,1}^{s1} - E_{\max,2}^{s1}}{E_{\max,1}^{s1} + E_{\max,2}^{s1}} \quad (4.8)$$

es la relación de la diferencia de energía entre los dos máximos, normalizada con respecto a la suma de esas energías, en la primera capa del ECAL.

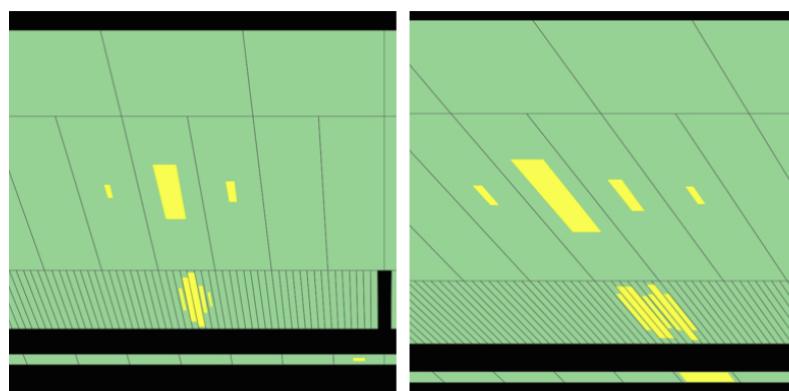


Figura 4.2: Depósitos de energía característicos para un fotón aislado (izquierda), y un evento  $\pi^0 \rightarrow \gamma\gamma$  (derecha), que es posible distinguir gracias a la fina granularidad de la primera capa del ECAL [62].

## 4.2. Identificación de fotones

La identificación de fotones prompt frente a fotones falsos en colisiones hadrónicas es un gran desafío. Los fotones falsos están ampliamente dominados por candidatos a fotones que surgen del decaimiento de hadrones en jets, mientras que una fracción más pequeña de candidatos falsos está asociada con hadrones que depositan una energía significativa en el [ECAL](#), imitando la de los fotones reales. Los procesos con fotones prompt en el estado final, que ocurren en colisiones  $pp$  en el [LHC](#), desempeñan un papel central en el programa de física de [ATLAS](#). Tanto para las búsquedas como para las medidas de precisión, es importante contar con algoritmos y técnicas para identificar los fotones reales frente a los falsos. Estas búsquedas o medidas de precisión se llevan a cabo en un rango muy amplio de energía del fotón, empezando, por ejemplo, por resonancias de baja masa del Higgs a un par de partículas tipo axión que decaen en 4 fotones ( $H \rightarrow aa \rightarrow 4\gamma$ ) [63], donde el momento transverso del fotón es de  $\sim 25$  GeV, hasta fotones de muy alto  $p_T$  en búsquedas de resonancias  $\gamma+\text{jet}$ , donde los fotones tienen un momento transverso mayor a 1 TeV [add reference once the chapter numbers are in place](#).

El identificación de fotones en [ATLAS](#) se basa en cortes en las [SSs](#) y se definen diferentes [Working Points \(WPs\)](#) con diferentes características: ya sea lograr un gran rechazo de fondo, o alta eficiencia de señal, o simplemente bajos tiempos de cómputo. En esta sección se describe el procedimiento utilizado para la optimización de estos [WPs](#) y luego se describen los métodos para medir las eficiencias correspondientes de cada [WP](#).

### 4.2.1. Procesos de interés y selección de eventos

Dado el amplio rango en el que se utilizan los fotones en [ATLAS](#), la optimización de los [WPs](#) utiliza dos procesos diferentes que permiten obtener muestras limpias de fotones en los regímenes de bajo y alto  $p_T$ . En el caso de bajo  $p_T$ , se utiliza una fuente muy limpia de fotones procedentes de decaimientos radiativos del bosón  $Z$ . Por otro lado, aunque con mayor contaminación de fondo, se emplean eventos [Single Photon \(SP\)](#) para fotones de alto  $p_T$ . En los siguientes párrafos se ofrece una descripción de cada de las [muestras](#) de fotones utilizadas.

**Decaimientos radiativos del bosón  $Z$**  En el rango de bajo  $p_T$ , los fotones procedentes del [decaimiento radiativo del bosón  \$Z\$](#)  ( $Z \rightarrow \ell\ell\gamma$ ) se seleccionan como fotones de señal. Hay dos modos de producción posibles para los procesos del [Modelo Estándar \(SM\)](#) de  $pp \rightarrow Z(\ell^+\ell^-)\gamma$ , donde  $\ell$  es un electrón o un muón. Estos son: Radiación de estado inicial ([ISR](#)) donde el fotón es radiado [desde los quarks](#), y Radiación de estado final ([FSR](#)) (de ahora en más también referido como [decaimiento  \$Z\$  Radiativo \(RZ\)](#)), donde el fotón es radiado [desde uno de los leptones](#) del estado final a través de bremsstrahlung. Ambos modos de producción se muestran en la Figura 4.3.

Ambos procesos [FSR](#) y [ISR](#) pueden identificarse fácilmente comparando la distribución de la masa invariante de los dos leptones ( $m_{\ell\ell}$ ) con la distribución de la masa invariante de los dos

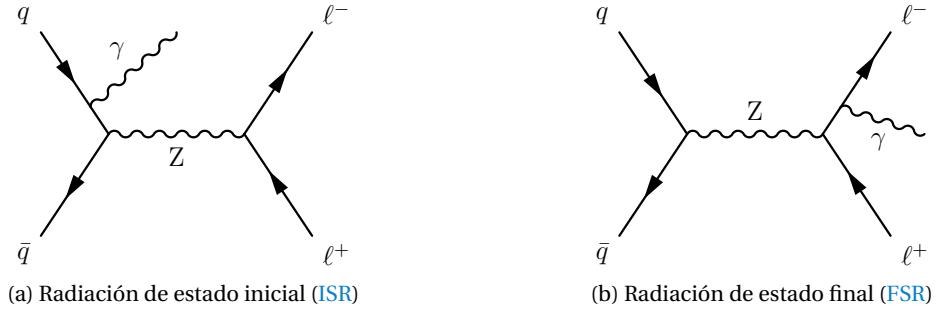


Figura 4.3: Diagramas de Feynman del proceso de radiación de un fotón en decaimientos  $Z \rightarrow \ell\ell\gamma$  para los casos de **ISR** (izquierda) y **FSR** (derecha).

leptones junto con el fotón ( $m_{\ell\ell\gamma}$ ). Para los eventos **ISR**,  $m_{\ell\ell}$  tiene su máximo en la masa del  $Z$ , y el fotón simplemente suma a la masa invariante de tres cuerpos ( $m_{\ell\ell\gamma}$ ) haciéndola mayor que 91 GeV. En el caso **FSR**, en cambio, la masa invariante de tres cuerpos  $m_{\ell\ell\gamma}$  presenta su máximo en la masa del  $Z$ , que se observa en la Figura 4.4a. Para los estudios de identificación de fotones, sólo se consideran los fotones de los eventos **FSR**. La razón detrás de la selección de **FSR** sobre **ISR** se debe a lo siguiente. Los eventos **ISR** también sufren la contaminación de fondos provenientes de eventos  $Z + \text{jets}$ , en los que el jet se identifica erróneamente como un fotón, y además la sección eficaz  $Z + \text{jets}$  es de órdenes de magnitud mayor a la del proceso  $Z + \gamma$ . A partir de las Figuras 4.4b y 4.4c, donde se muestran las distribuciones de  $m_{\ell\ell}$  en función de  $m_{\ell\ell\gamma}$  para los procesos  $Z \rightarrow \ell\ell$  y  $Z \rightarrow \ell\ell\gamma$ , respectivamente, se puede apreciar la separación entre estos dos procesos cuando se seleccionan fotones **FSR**.

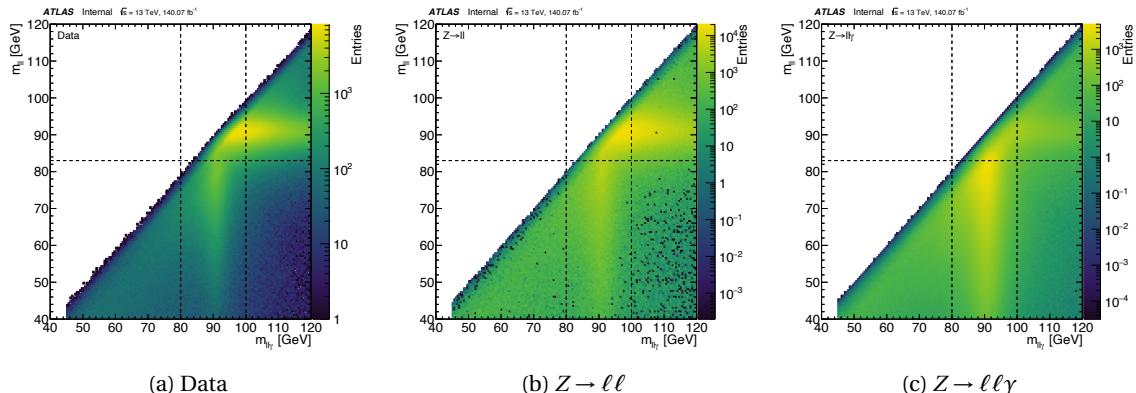


Figura 4.4: Masa invariante de los dos leptones en función de la masa invariante de ambos leptones junto con un fotón en (a) datos, (b) fondos y (c) señal. La región en la cual se encuentra una gran concentración de eventos con  $m_{\ell\ell} \sim m_Z$  corresponde a eventos de **ISR**, mientras que eventos de **FSR** están caracterizados por  $m_{\ell\ell\gamma} \sim m_Z$ .

Con la muestra **RZ**, se requiere que los fotones tengan un momento transverso  $p_T > 7$  GeV y una pseudorapidez en el rango de  $|\eta| < 1.37$  o  $1.52 < |\eta| < 2.37$ , evitando así la región del crack. Para los estudios de optimización no se aplica ningún requisito de aislamiento sobre los fotones, pero para las medidas de eficiencia se utiliza el **WP** de aislamiento **loose**, descripto en la Sección 3.2.3. Se requiere que los leptones tengan  $E_T > 10$  GeV, los muones una pseudorapidez  $|\eta| < 2.5$  y para los electrones  $|\eta| < 2.47$ , excluyendo el crack. Tanto a los electrones como a los muones se les exige que pasen los requisitos de aislamiento **loose** y que pasen el criterio de identificación **medium**.

El fotón **FSR** se selecciona entonces requiriendo  $80 < m_{\ell\ell\gamma} < 100$  GeV y  $40 < m_{\ell\ell} < 83$  GeV. Finalmente, para evitar cualquier sesgo en las **SS** del fotón y en sus variables de aislamiento, se requiere una distancia mínima de  $\Delta R > 0.4$  entre dicho fotón y el leptón más cercano.

**Single Photon (SP)** La muestra de fotones inclusivos, o Single Photon (**SP**), se recoge mediante triggers de un sólo fotón, cuyos umbrales varían entre 10 GeV y 140 GeV y requieren una identificación *loose*. Aunque los triggers utilizados para obtener esta muestra están preescalados (con la excepción del de 140 GeV), proporcionan un gran conjuntos de datos de fotones de alto  $p_T$ . Estos procesos incluyen eventos a Leading Order (**LO**) de  $\gamma + \text{jet}$  procedentes de la dispersión dura  $qg \rightarrow q\gamma + q\bar{q} \rightarrow g\gamma$ , así como fotones prompt procedentes de la fragmentación de quarks en eventos de dijet de **QCD**. Se requiere que estos fotones presenten una pseudorapidez de  $|\eta| < 2.37$  excluyendo el crack, y pasar el requisito de aislamiento *loose*. Las muestras de **SP** se utilizan tanto para los estudios de optimización como para la estimación de las eficiencias.

### 4.2.2. Optimización

A partir de las **SSs** anteriormente descriptas, se pueden definir tres **WPs** para los fotones: *loose*, *medium* y *tight* [46]. El **WP** *loose* emplea cortes a las variables definidas en la segunda capa y a la variable de filtración hadrónica, y es utilizado principalmente por el trigger. Los **WPs** *medium* y *tight* utilizan todas las variables definidas previamente. El primero está optimizado para tener una eficiencia fija de 95 %, mientras que el segundo proporciona un excelente rechazo de fondo. La Tabla 4.1 muestra qué variables se utilizan para cada **WP**.

Para la optimización de los **WPs** se utilizan las dos muestras definidas previamente: los eventos **RZ** para fotones con  $10 < p_T < 25$  GeV como señales y eventos de  $Z \rightarrow \ell\ell$  como fondos; y para el régimen de alto  $p_T$  ( $p_T > 25$  GeV) los eventos de **SP** se consideran como señal mientras que los eventos **dijet** son los fondos.

La identificación de fotones utiliza cortes en las **SSs**. En la Figura 4.5 se muestran ejemplos de tres de estas **SS** ( $R_\eta$ ,  $E_{\text{ratio}}$  y  $w_{\eta 2}$ ) comparando eventos de señal y de fondo utilizando las muestras de **RZ**, donde se observa un excelente poder discriminatorio. Los cortes en todas las **SSs**, para cada **WP** de identificación, se optimizan en función de la energía transversal y la pseudorapidez del candidato a fotón, para tener en cuenta la forma de las variables para diferentes  $\eta$  y para variaciones en la cantidad de material y la geometría del calorímetro. Los **WPs** *medium* y *tight* también se optimizan por separado para fotones convertidos y no convertidos. Los cortes se optimizan utilizando un enfoque Multivariable (**MV**), en el que las eficiencias de señal se escanean entre 0 % y 100 % mientras se intenta maximizar el rechazo de fondo.

### 4.2.3. Estimación de las eficiencias

Una vez optimizados los diversos **WPs** de identificación, es fundamental contar con las eficiencias ya sea de los datos y las simulaciones **MC**. Las estimaciones de eficiencias de fotones hace uso de tres métodos diferentes y característicos en diferentes rangos de  $p_T$ , que son detallados

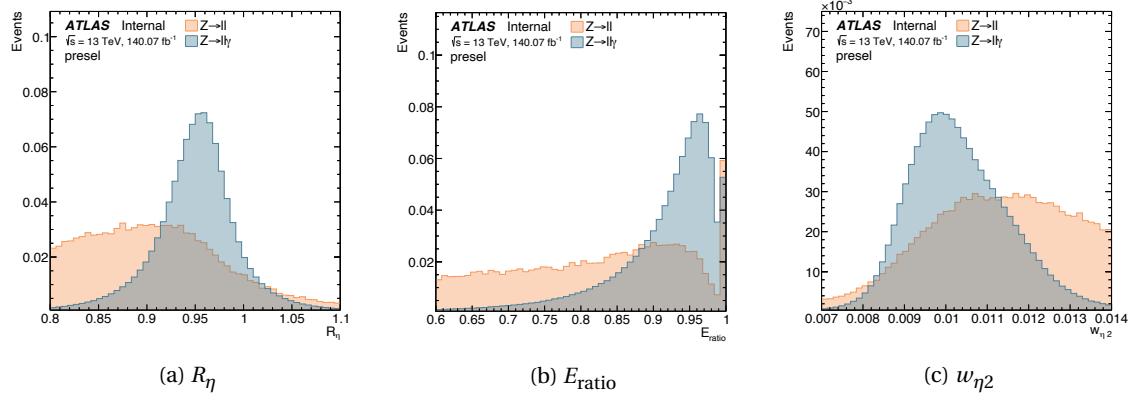


Figura 4.5: Distribuciones normalizadas de señal (azul) y fondo (naranja) de diferentes SSs utilizando las muestras de RZ y pasando la selección de eventos detallada en la Sección 4.2.1.

en la Ref. [64], y son brevemente descriptos en los próximos párrafos. Para los tres métodos, se requiere que los fotones satisfagan el criterio de aislamiento loose definido en la Sección 3.2.3 y, por tanto, las **eficiencias de los fotones** se miden con respecto a este criterio de aislamiento.

Para el rango de bajo  $p_T$  ( $7 < p_T < 100$  GeV), los fotones procedentes del **pprocreso RZ** se utilizan como fotones de señal. El método para estimar las eficiencias se basa en el uso de ajustes a la distribución observada de la masa invariante de tres cuerpos ( $m_{\ell\ell\gamma}$ ) antes y después de aplicar el criterio de identificación tight. El número de eventos de señal y de fondo puede contarse a partir de los ajustes, y las purezas de señal se calculan antes ( $P^{\text{total}}$ ) y después ( $P^{\text{pass}}$ ) de la aplicación de la identificación tight. La eficiencia final en los datos viene dada por:

$$\varepsilon_{ID} = \frac{P^{\text{pass}} N_{\text{data}}^{\text{pass}}}{P^{\text{total}} N_{\text{data}}^{\text{total}}}.$$

El segundo método para calcular eficiencias se basa en transformaciones de Smirnov [65] para que las distribuciones de las SSs de los electrones se parezcan a los de los fotones. Las muestras usadas en este método son **decaimientos  $Z \rightarrow ee$** , en los que se requiere que los electrones pasen el criterio de aislamiento de fotones loose. También es necesario tener en cuenta la contribución de una pequeña fracción de fondo de procesos  $W + \text{jets}$  y producción multijet. Estos fondos son tratados mediante ajustes a la distribución de  $m_{ee}$  de datos, utilizando señales simuladas y formas funcionales que describen el resto de los fondos, obtenidas en regiones de control (regiones en donde las contribuciones de estos fondos son dominantes y separables). Luego, los candidatos a electrones se cuentan a partir de eventos en el rango  $70 < m_{ee} < 110$  GeV, y las eficiencias se miden utilizando el método tag-and-probe descripto en la Ref. [40]. El rango  $p_T$  en el que se aplica este método es  $25 < p_T < 250$  GeV.

El último y tercer método utiliza muestras de **SP** con fotones en el rango  $50 < p_T < 1500$  GeV. En este caso se utiliza el *Matrix Method* [64], que construye cuatro regiones ortogonales que pasan o no el **WP** de identificación tight, y pasan o no el aislamiento de trazas (descripto en la Sección 3.2.3). Para cada región, surgen dos incógnitas: el número de eventos de señal y de fondo. Si se conocen las eficiencias de aislamiento de trazas para los componentes de señal y de fondo, entonces es posible estimar la eficiencia de los fotones loose que pasan los criterios

de identificación tight. Las eficiencias de aislamiento para los fotones de señal se estiman utilizando muestras de **MC**, y las de fondo se obtienen en una región de control enriquecida con jets construida a partir de la inversión de los criterios de identificación. Las eficiencias en datos para el **WP** de identificación tight son entonces:

$$\varepsilon^{\text{tight-ID}} = \frac{\frac{\hat{\varepsilon}_{\text{ID}} - \hat{\varepsilon}_{\text{ID}}^b}{\hat{\varepsilon}_{\text{ID}}^s - \hat{\varepsilon}_{\text{ID}}^b} \cdot N_{\text{ID}}^T}{\frac{\hat{\varepsilon} - \hat{\varepsilon}^b}{\hat{\varepsilon}^s - \hat{\varepsilon}^b} \cdot N^T},$$

donde  $N^T$  representa la totalidad de fotones en la muestra inclusiva que consiste en  $N^s$  fotones de señal (o fotones prompt) y  $N^b$  fotones falsos (fotones de fondo). El número  $N_{\text{ID}}^T$  es el subconjunto de  $N^T$  que pasa el requisito de identificación. Las eficiencias de aislamiento de traza de datos, señal y fondo se representan con  $\hat{\varepsilon}$ ,  $\hat{\varepsilon}^s$  y  $\hat{\varepsilon}^b$ , respectivamente. Del mismo modo, las eficiencias de aislamiento de traza para los fotones que superan la identificación tight se muestran como  $\hat{\varepsilon}_{\text{ID}}$ ,  $\hat{\varepsilon}_{\text{ID}}^s$  y  $\hat{\varepsilon}_{\text{ID}}^b$ , respectivamente. Las eficiencias medidas para los fotones con  $p_T > 150$  GeV están entre 90 y 96 %.

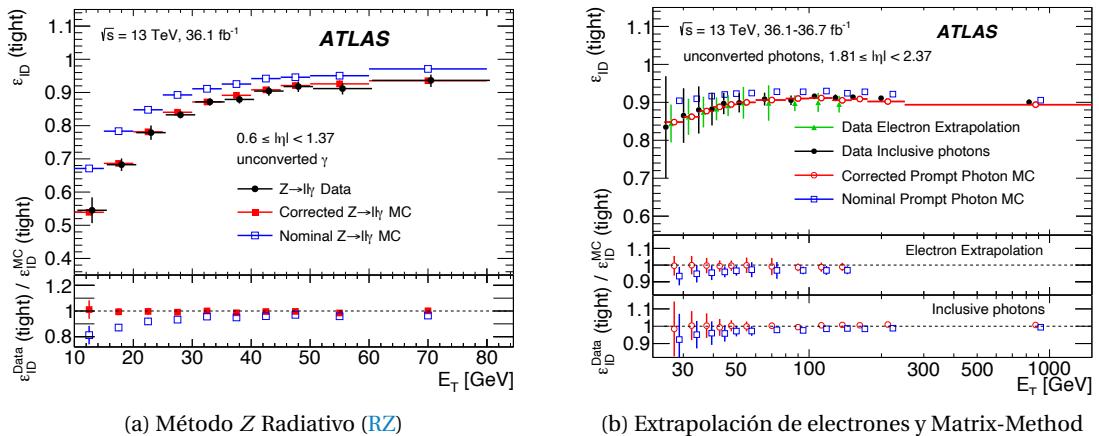


Figura 4.6: Comparación de las eficiencias calculadas para datos y **MC** utilizando los tres métodos diferentes para su cálculo. En ambas figuras, para cada método, se muestran dos conjuntos diferentes de mediciones **MC**: la nominal y la corregida (discutida en el texto). Los paneles inferiores muestran el cociente entre las eficiencias de los datos y las predicciones **MC** (denominadas Factor de escalas (**SF**) en el texto). Las figuras fueron tomadas de la Ref. [64].

En la Figura 4.6a se muestra un ejemplo de las eficiencias de identificación en función del  $p_T$  del fotón utilizando el método **RZ**. Las eficiencias de los datos están representadas por los puntos negros, mientras que el **MC** nominal se muestra con cuadrados azules vacíos. Los cocientes de datos con el **MC** nominal (también denominados Factor de escala (**SF**)) mostrados en el panel inferior difieren hasta en un 20 % de 1, lo que demuestra que las eficiencias calculadas con **MC** difiere de las calculadas con los datos. Sin embargo, también se muestra otro conjunto de eficiencias, pertenecientes a simulaciones de **MC** corregidas, mejorando drásticamente el acuerdo entre los datos y la simulación, como se ve de los **SFs**. La razón por la que se necesitan estas correcciones y cómo se implementaron en **ATLAS** se explica en la siguiente sección (Sección 4.3), y cómo se corrigen en el Capítulo 5. La Figura 4.6b muestra las medidas de eficiencia usando los dos métodos restantes (extrapolación de electrones y Matrix Method), donde se obtienen las mismas mejoras en los **SFs** cuando se usa la simulación corregida.

Como ya se ha mencionado, estos cocientes entre datos y eficiencias MC se denominan SFs y encapsulan las diferencias entre datos y simulación. Se calculan por separado para cada uno de los tres métodos y después se combinan utilizando una media ponderada [66] en cada *bin* y asumiendo que las incertezas estadísticas y sistemáticas no están correlacionadas entre los métodos. Los resultados actuales de estos SFs, calculados utilizando el conjunto completo de datos de Run-2, se muestran en la Figura 4.7. La obtención de los SFs es de vital importancia, ya que luego son factores que se aplican a los eventos de MC para ser corregidos, y así lograr una comparación justa con los datos recolectados.

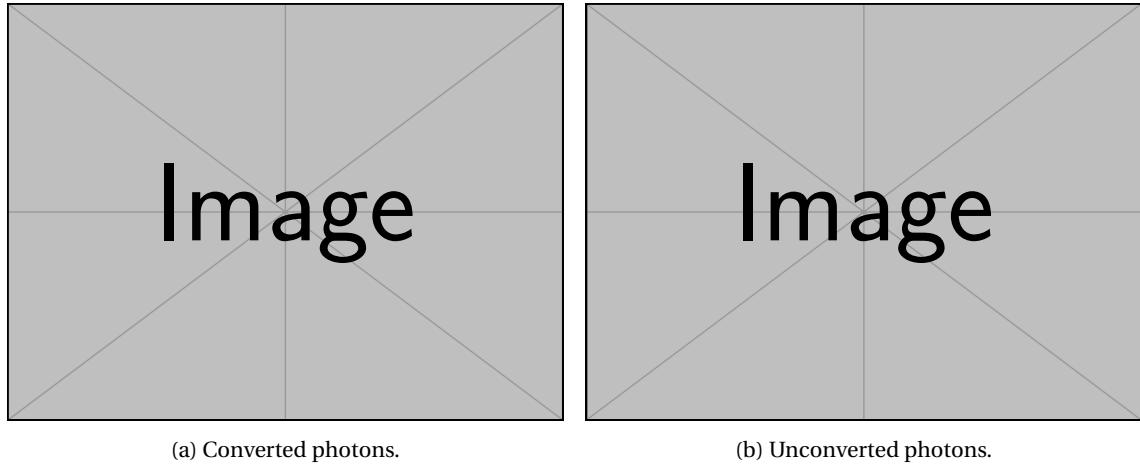


Figura 4.7: SFs resultantes de la identificación de fotones en los diferentes bins de  $p_T$  y  $|\eta|$  para fotones convertidos (izquierda) y no convertidos (derecha). [Ask fran for the plots in his presentation!](#)

### 4.3. Las diferencias de las Shower Shapes entre datos y MC

Como se ha mostrado anteriormente, la simulación MC no describe perfectamente los datos, lo cual puede verse de los valores de los SFs los cuales difieren de 1. En particular, al comparar las distribuciones de las SSs, se observa que las distribuciones MC están desplazadas o incluso la forma difiere, como se muestra en la Figura 4.8, al comparar los datos (puntos negros) con el histograma representado por la línea roja, correspondiente al MC.

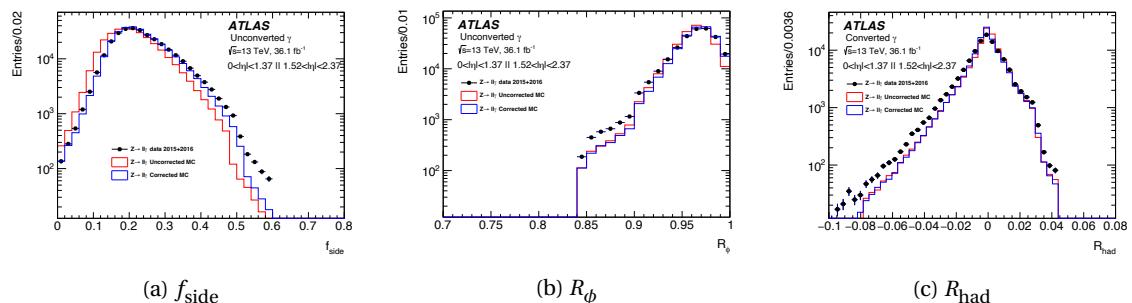


Figura 4.8: Comparaciones de las SSs entre los datos (puntos negros) y la simulación MC nominal (línea roja) y corregida (línea azul), utilizando la muestra RZ [47].

Las principales diferencias en las distribuciones surgen para los perfiles de la lluvia en la dirección de  $\eta$ , donde se observan distribuciones más amplias en los datos en comparación con MC.

Parte del efecto se corrigió en 2010 ~~tras pasar a una descripción~~ detallada de la composición del material absorbente del **ECAL** en GEANT4. Sin embargo, algunas discrepancias entre los datos y **MC** aún permanecen y son motivos de estudio para la colaboración. Algunas razones potenciales puede ser:

- Descripción geométrica del grosor del plomo en el **ECAL** (incluyendo posibles variaciones debidas a la gravedad).
- Modelado erróneo del campo eléctrico en los huecos de **LAr**.
- Modelado erróneo del efecto de *cross-talk* (intercambio de energía entre las celdas del calorímetro debido a la electrónica).

Para tener en cuenta estas diferencias en las **SSs**, históricamente, se hicieron correcciones en forma de desplazamientos de cada una de las distribuciones de **MC**. Estos desplazamientos comprendían los denominados Fudge Factors (**FFs**), y se determinaban utilizando una minimización de  $\chi^2$  en la comparación de las distribuciones de las **SSs** entre datos y **MC** [40, 64]. Aunque las diferencias del valor medio disminuyeron sustancialmente tras estas correcciones, como se observa por ejemplo en el caso de  $f_{\text{side}}$  en la Figura 4.8a, quedaron diferencias residuales pero aún así notables.  Se observa en las distribuciones que las principales diferencias que se mantuvieron están relacionadas con la forma de las mismas, por lo que se necesitan correcciones de orden superior. En el siguiente capítulo se presenta una descripción detallada de las correcciones. Además, dado que las **SSs** se construyen a partir de depósitos de energía en las celdas del **ECAL**, otra forma posible de corregir el desacuerdo actual es corregir directamente las energías de las celdas en las simulaciones **MC**, corrigiendo entonces todas las **SSs** en simultáneo. Este nuevo enfoque se describe también en el siguiente capítulo.

# CORRECCIONES DE LAS SHOWER SHAPES (~~SSS~~)

5

*"Champions keep playing until they get it right."*

---

Billie Jean King

En el capítulo anterior se vio que los SFs (cociente entre las eficiencias de los datos y las obtenidas a partir de la simulación MC) se desvían de la unidad. Dado que la identificación de fotones se basa en los cortes de las SSs de fotones, se vio que las diferencias de hecho aparecen en estas variables. Desde el Run-1, estas se han corregido con lo que se conoce como Fudge Factors (FFs), que se han calculado como simples desplazamientos a las distribuciones MC y se ha visto que proporcionan muy buenas mejoras de los SFs. Sin embargo, como se ha visto antes, siguen habiendo discrepancias entre las distribuciones que hay que abordar para poder contar con una simulación aún mejor. En la Sección 5.1, se presenta un enfoque más sofisticado basado en un cálculo de orden superior para corregir las SSs. Asimismo, en la Sección 5.2 se estudia y aborda un nuevo enfoque que utiliza directamente las energías de las celdas. Los estudios presentados en este capítulo constituyen uno de los principales temas de trabajo de la presente tesis.

## 5.1. Fudge Factors (FFs)

### 5.1.1. Muestras de datos y simulaciones MC

Los FFs se calculan utilizando el conjunto completo de datos de Run-2, recolectados a una energía de centro de masa de  $\sqrt{s} = 13$  TeV y con una luminosidad integrada correspondiente a  $140 \text{ fb}^{-1}$ . Las muestras simuladas de RZ y SP se utilizan para este estudio, ya que representan rangos  $p_T$  complementarios. Los eventos de RZ se generan con SHERPA 2.2.11 [67], mientras que SHERPA 2.2.1 se utiliza para los eventos de fondo  $Z \rightarrow \ell\ell$ . Respecto a las muestras SP, los eventos se generan con PYTHIA 8.186 [68], que incluye eventos  $\gamma + \text{jet}$  de Leading Order (LO) procedentes tanto de procesos directos ( $qg \rightarrow q\gamma$  y  $q\bar{q} \rightarrow g\gamma$ ) como de fragmentación de fotones procedentes de eventos Quantum Chromodynamics (QCD) dijet.

 En ambos casos, el detector **ATLAS** se simula utilizando GEANT4 [69] y los eventos **MC** se **escalan** para que sus distribuciones de pileup se asemejen a las de los datos, para cada año del periodo de toma de datos.

### 5.1.2. Cálculo de Fudge Factors (FFs)

El cálculo se realiza por separado para las dos muestras consideradas: **RZ** para fotones con  $7 \leq p_T \leq 50$  GeV y **SP** para fotones con  $p_T > 50$  GeV, que ya se discutieron en la Sección 4.2.1. Dado que las distribuciones de las **SSs** varían en función de  $p_T$  y  $|\eta|$ , el cálculo se realiza en bins de estas variables:

$$p_T^\gamma : \begin{cases} \text{RZ} : [7, 15, 20, 30, 50] \text{ GeV} \\ \text{SP} : (50, 60, 80, 100, 150, 300, 600, \infty] \text{ GeV} \end{cases}$$

$$|\eta| : [0, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37].$$

Además, ~~como se menciona en la Sección 4.1~~, hay variables muy sensibles al estado de conversión del fotón, ~~es decir, si los fotones están convertidos o no~~. Por esta razón, el cálculo se hace por separado para fotones convertidos y no convertidos. En total se corrigen nueve variables con este método:  $E_{\text{ratio}}$ ,  $f_{\text{side}}$ ,  $R_\eta$ ,  $R_\phi$ ,  $R_{\text{had}}$ ,  $R_{\text{had}\,1}$ ,  $w_1$ ,  $w_{\eta\,2}$  y  $w_{\text{s}\,\text{tot}}$ ; ya que son en las que se observan las mayores discrepancias entre los datos y **MC**.

Para cada **SS**, se crean histogramas de **MC** y datos de 100 bins. La elección del **binneado** se basa en disponer de estadística suficiente en cada bin y también en capturar todas las características de las variables. ~~Después~~, cada histograma se suaviza utilizando una herramienta del paquete de TMVA [70] denominada Kernel Density Estimator (**KDE**). El método **KDE** consiste en estimar la forma de una Función de Densidad de Probabilidad (**PDF**) mediante la suma sobre eventos suavizados. La **PDF**  $p(x)$  de una variable  $x$  es entonces

$$p(x) = \frac{1}{N} \sum_{i=1}^N K_h(x - x_i)$$

donde  $N$  es el número de eventos,  $K_h(t) = K(t/h)/h$  es la función kernel, y  $h$  es el ancho de banda del kernel. La idea básica es que cada evento se considera como una función Dirac- $\delta$ , que se sustituye por una función Kernel (Gaussiana) y finalmente se suman para formar la **PDF** final. El método de suavizado **KDE** puede aplicarse de dos formas: no adaptativo o adaptativo, como se ve en la Figura 5.1. En el primer caso, el ancho de banda es constante para toda la muestra  $h_{NA}$ , mientras que en el segundo, se utiliza el valor de **KDE** no adaptativo pero que varía en función de  $p(x)$  como

$$h_A = \frac{h_{NA}}{\sqrt{p(x)}}.$$

El método **KDE** adaptativo mejora la forma de la **PDF** especialmente en regiones de baja estadística, pero en regiones de alta estadística puede dar lugar a un exceso de suavizado o *oversmoothing*. El grado de suavizado se ajusta multiplicando el ancho de banda  $h$  por lo que se denominan *fine factors*. Estos factores son parámetros definidos por el usuario que se ajustan para permitir que la **PDF** conserve las características importantes del histograma original

y también para evitar fluctuaciones estadísticas. Los valores más altos de los factores indican funciones Kernel más amplias y, por lo tanto, la PDF capta menos fluctuaciones estadísticas. En la Figura 5.2 se muestran ejemplos del procedimiento de suavizado aplicado a  $R_{\text{had}}$  para casos en los que los histogramas originales tienen baja y alta estadística.

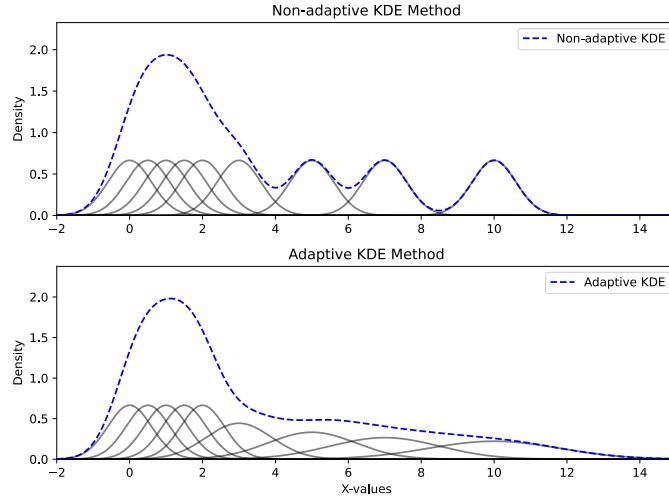


Figura 5.1: Esquema del suavizado no adaptativo y adaptativo del método KDE.

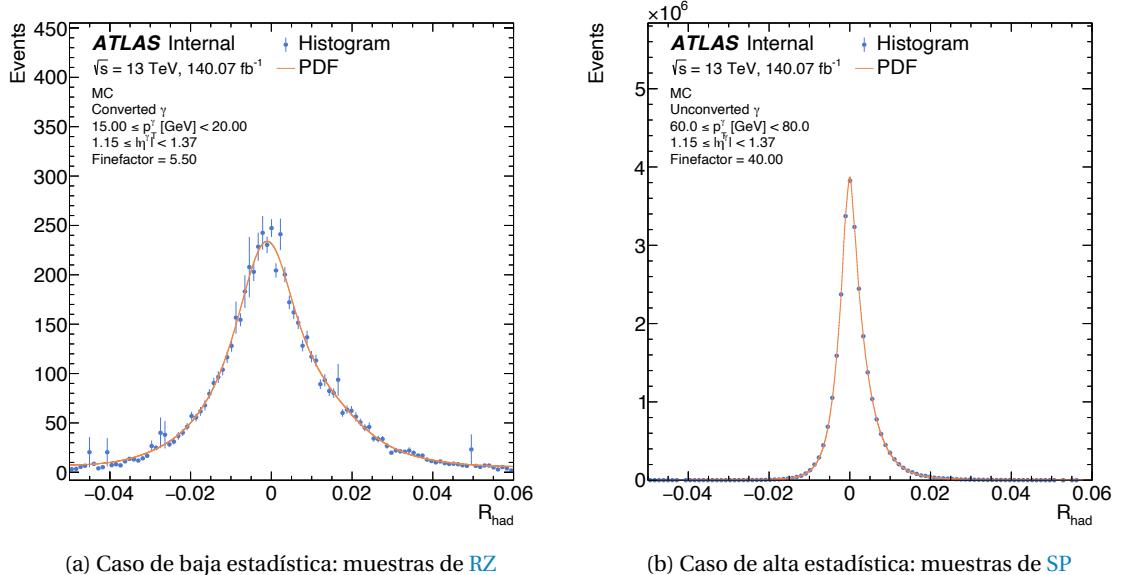


Figura 5.2: Suavizado utilizando el método KDE aplicado a la SS  $R_{\text{had}}$  para fotones en  $0.8 < |\eta| < 1.15$  bajo dos posibles escenarios: baja y alta estadística. El histograma original se muestra con los puntos azules y las correspondientes PDFs con la línea naranja. Además, se muestran los valores de los fine factors usados en cada caso.

Una vez creados las PDFs de los datos y la simulación MC para una dada variable, bin de  $p_{\text{T}}$  y  $|\eta|$ , y tipo de conversión, la PDF de MC se normaliza al de los datos y se calcula un valor  $\chi^2$  entre ambos como [71]:

$$\chi^2 = \sum_{i=1}^N \frac{(w_{\text{MC},i} W_{\text{data}} - w_{\text{data},i} W_{\text{MC}})^2}{s_{\text{MC},i}^2 W_{\text{data}}^2 + s_{\text{data},i}^2 W_{\text{MC}}^2}. \quad (5.1)$$

$N$  es el número de bins de las PDFs,  $w_{\text{MC},i}$  y  $w_{\text{data},i}$  son los números de eventos de MC y datos en cada bin, respectivamente,  $s_{\text{MC},i}$  y  $s_{\text{data},i}$  son los errores del bin y, por último,  $W_{\text{data}}$  y  $W_{\text{MC}}$  son la suma de los pesos en datos y MC, respectivamente.

### Correcciones *shift-only*

Como ha sido mencionado anteriormente, las correcciones a las **SSs** de **MC** han sido realizadas a partir de simples corrimientos de ellas. Estos corrimientos o desplazamientos se denominan, de aquí en adelante, *shift fudge-factors FF*, o simplemente *shifts*. Para ello, se desplaza a la **PDF** de **MC** a la izquierda y a la derecha un bin a la vez. El número inicial de bines que se debe desplazar a la distribución **MC** se calcula mediante la diferencia de los valores medios de las distribuciones de datos y **MC**. A partir de este valor inicial, se consideran shifts de 100 bines a cada lado. Como consecuencia de este procedimiento, la resolución del shift depende directamente del ancho del bin de la **PDFs**, por lo que bines más pequeños conducen a una mejor resolución del shift. Dado que los histogramas, en primer lugar, se construyen con bines relativamente anchos, la **PDFs** puede construirse utilizando bines pequeños de alta precisión para asegurar una alta resolución. Despues de pruebas de convergencia de los **FFs**, se decide construir las **PDFs** con 5000 bines.

Para cada bin que se ha desplazado la distribución, se calcula y se registra el valor  $\chi^2$  antes mencionado. Suponiendo que los errores  $s_{MC,i}$  y  $s_{data,i}$  tienen una distribución gaussiana estándar<sup>1</sup>, se espera que la forma seguida por los valores  $\chi^2$  cerca del mínimo sea aproximadamente parabólica.

Para extraer los **FFs**, se realiza un ajuste a los valores de  $\chi^2$  cercanos al mínimo (5 bines a cada lado del bin mínimo) utilizando una función parabólica y el **FF** de shift se obtiene a partir del mínimo ajustado. Por último, utilizando este valor, se puede corregir a la **SSs** evento a evento como:

$$x = x_{\text{old}} + \text{shift}.$$

donde  $x_{\text{old}}$  y  $x$  representan el valor original y el valor post-corrección de la variable la cual se quiere corregir, respectivamente.

### Correcciones *shift+stretch*

 Se observó que incluso después de aplicar correcciones de shift a las **SSs** de **MC**, seguían existiendo diferencias en las formas de las mismas, y en algunos casos éstas pueden ser bastante sustanciales. Una forma de seguir mejorando el acuerdo entre los datos y **MC** es incluir otra corrección que se denomina *stretching*. Las dos correcciones, actuando en conjunto, son denominadas como correcciones *shift+stretch* (o desplazamiento+estiramiento), que pretenden corregir simultáneamente el valor medio y los anchos de las distribuciones de **MC**.

El método de corrección *shift+stretch* empieza por encontrar el máximo de la **PDF** de **MC**. Posteriormente, la **PDF** se estira alrededor del máximo calculando la nueva posición de cada bin por el producto  $\text{stretch} \times (x - \text{stretch point})$ , donde  $x$  es el centro del bin en cuestión. De este modo, el centro de cada bin conserva la distancia inicial al centro de la distribución, multiplicada por el factor de stretch. En el escenario en el que el stretch es  $> 1$ , puede haber casos en

<sup>1</sup> Este requisito se cumple siempre que los contenidos de los bines de ambas **PDFs** sean mayores que 10, lo que también se satisface puesto que los histogramas se construyen con bines relativamente amplios.

los que sea lo suficientemente grande como para dar lugar a bins vacíos. El contenido de estos bins vacíos se interpola linealmente a partir de los bins vecinos distintos de cero. Una vez *estirada* la PDF, se desplaza a izquierda y derecha siguiendo el mismo procedimiento que para el caso de shift-only, calculando los valores  $\chi^2$  para cada shift<sub>i</sub> después de aplicar el stretch<sub>j</sub>. Como resultado de este procedimiento, ahora se obtiene una grilla bidimensional de valores de  $\chi^2$  en el plano de shift-stretch. El par shift-stretch se obtiene del ~~el~~ centro del bin mínimo, y ~~comprenden~~ ahora los FFs. Las correcciones pueden ser aplicadas a cada SS  $x$ , evento a event, como:

$$x = \text{stretch} \times (x_{\text{old}} - \text{stretch point}) + \text{shift} + \text{stretch point}, \quad (5.2)$$

donde nuevamente  $x_{\text{old}}$  representa el valor de la variable sin corregir.

Un ejemplo de los valores de  $\chi^2$  resultantes para la variable  $f_{\text{side}}$  se muestra en la Figura 5.3a, donde el shift está representado en el eje  $x$  y el stretch en el eje  $y$ . El valor óptimo de shift-stretch en este caso corresponde a shift = 0.03 y stretch = 1.09. En la Figura 5.3b se muestran las PDFs antes y después de aplicar las correcciones, donde se comparan con la PDF de los datos. Como se ve en la figura, hay una gran mejora y las distribuciones coinciden casi a la perfección.

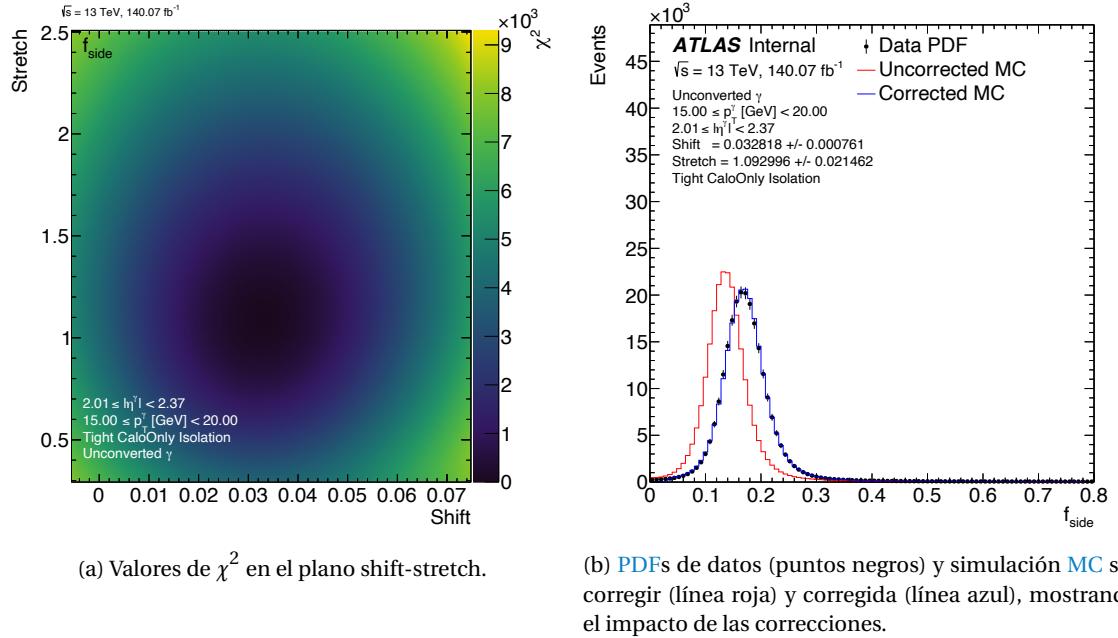


Figura 5.3: Cálculo de los FFs de shift+stretch para  $f_{\text{side}}$  utilizando fotones no convertidos con momento transverso de  $15 < p_T < 20 \text{ GeV}$  y pseudorapidez  $2.01 < |\eta| < 2.37$

### 5.1.3. Cálculo de incertezas

#### Incerezas estadísticas

Para extraer las incerezas estadísticas de los FFs de shift y stretch, se realiza un ajuste al contorno de  $1\sigma$  (nivel de confianza del 68.3 %) sobre los valores  $\chi^2$ . Este contorno representa una

elipse que toma la siguiente forma:

$$\chi^2 = \chi_{\min}^2 + \frac{1}{1-\rho^2} \left[ \left( \frac{x-x_0}{\sigma_x} \right)^2 + \left( \frac{y-y_0}{\sigma_y} \right)^2 - 2\rho \left( \frac{x-x_0}{\sigma_x} \right) \left( \frac{y-y_0}{\sigma_y} \right) \right], \quad (5.3)$$

donde  $\rho$  es el coeficiente de correlación entre ambas variables,  $\sigma_x$  y  $\sigma_y$  las incertezas sobre  $x$  y  $y$ , respectivamente,  $(x_0, y_0)$  es la posición del centro de la elipse, y  $\chi_{\min}^2$  es el valor mínimo de  $\chi^2$  obtenido del histograma bidimensional.

Extrayendo los semiejes mayor y menor de la elipse ajustada, y con el ángulo de inclinación de la misma, las incertezas estadísticas sobre dos variables  $x$  y  $y$  (que en este caso representan el shift y el stretch, respectivamente) son (véase el Apéndice A):

$$\sigma_x = \sqrt{a^2 \cos^2 \theta + b^2 \sin^2 \theta} \quad (5.4)$$

$$\sigma_y = \sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta}. \quad (5.5)$$

### Incertezas sistemáticas

Las incertezas sistemáticas se obtienen variando los criterios de preselección, es decir, la identificación y el aislamiento de fotones. El cambio de los diferentes criterios de preselección permite que las SSs varíen dependiendo de la cantidad de contaminación de fondo, y en consecuencia también lo hacen los FFs. Las diferentes selecciones son, para cada muestra:

- *Z Radiativo (RZ):*

- Nominal: Sin criterio de identificación, aislamiento FixedCutTightCaloOnly.
- Identificación loose, sin aislamiento.
- Identificación loose, aislamiento FixedCutTightCaloOnly.
- Sin identificación, aislamiento FixedCutLoose.

- *Single Photon (SP):*

- Nominal: identificación tight, aislamiento FixedCutLoose.
- identificación tight, aislamiento FixedCutTight .

Todas las demás combinaciones (o falta de ellas) de criterios de selección darían como resultado una muestra con estadísticas demasiado bajas o una muy baja pureza.

Los FFs se derivan para cada una de las selecciones anteriores, y se calcula la diferencia entre la nominal y la variada. La diferencia máxima se toma como incerteza sistemática, como el caso más conservativo. Finalmente, las incertezas estadísticas y sistemáticas se suman en cuadratura.

### 5.1.4. Resultados

Debido al hecho de que los FFs se calculan en un amplio rango de  $p_T$  y utilizando dos muestras distintas que abarcan regiones complementarias, los resultados se concatenan en 50 GeV, que coincide con el límite entre ellas.

A continuación, los valores de shift y stretch para distintas SSs serán mostrados. Los valores de shift se normalizan utilizando la desviación estándar de la SS luego de aplicar el FF de stretch, ya que esta cantidad permite comprender cuánto se desplaza cada variable con respecto a su ancho. Además, proporciona una medida única para todas las variables consideradas, ya que cada una de ellas abarca rangos diferentes. No obstante, el ancho de las variables varía según los distintos bines de  $p_T$  y  $|\eta|$ , lo que puede dar lugar a grandes diferencias entre bines vecinos.

En la Figura 5.4, se presentan ejemplos de los FFs resultantes para las variables  $R_\eta$  y  $w_{\eta 2}$  utilizando fotones convertidos. Se puede observar que para ambas variables los FFs dependen de  $p_T$ , especialmente hacia momentos transversos más altos. Este comportamiento también se repite en todas las variables. Inspeccionando los comportamientos y tendencias de los FFs, también es posible recuperar información sobre el mal modelado de las SSs por el MC. Como se mencionó en la Sección 4.3, se observaron anchos y perfiles en  $\eta$  más amplios para los datos en comparación con la simulación. De hecho, esto se puede inferir dado los valores de stretch aumentan hacia valores más altos de  $p_T$ , estirando las simulaciones MC hasta el doble de su ancho inicial. En el caso de  $R_\eta$  ( $w_{\eta 2}$ ) mostrado, la simulación MC sobreestima (subestima) el valor central de la distribución en casi una desviación estándar después de corregir la ancho, lo que significa que las diferencias entre la distribución MC sin corregir con la de los datos son muy grandes.

También es útil visualizar los FFs en un bin de  $p_T$  fijo y en función de  $|\eta|$ , para así determinar qué tan dependientes de  $|\eta|$  son las correcciones. Esto se muestra para  $w_{s \text{ tot}}$  utilizando fotones convertidos con  $50 < p_T < 60$  GeV en la Figura 5.5. Como se puede notar, para  $|\eta| > 1.81$  (los dos últimos bines), los valores de shift normalizados son mayores que los de los bines anteriores en, al menos, un factor 2. Sin embargo, los valores de shift sin normalizar mostrados en la Figura 5.5b no presentan un cambio tan brusco, observándose sólo una pequeña dependencia en  $|\eta|$ . Como consecuencia de este comportamiento, se puede concluir que el cambio brusco observado es debido al cambio en el ancho de la distribución entre los distintos bines de  $|\eta|$ , tal como se había anticipado.

Para validar los FFs obtenidos, las correcciones se aplican a las SSs evento por evento. Las Figuras 5.6 y 5.7 muestran la aplicación de los FFs a algunas de las SSs utilizando las muestras RZ y SP, respectivamente, divididas en las regiones barrel y endcap en  $|\eta|$ . En la región barrel, las correcciones mejoran el acuerdo entre datos y MC, pero la mejora no es tan significativa como en la región endcap, donde se observan excelentes acuerdos entre datos y MC. Tomando como ejemplo las variables  $w_1$  y  $w_{s \text{ tot}}$ , se observan grandes diferencias en las formas entre la simulación nominal y los datos, que los métodos shift+stretch consiguen corregir. El mismo comportamiento se observa con las muestras SP, en las que estas variables presentan dos o

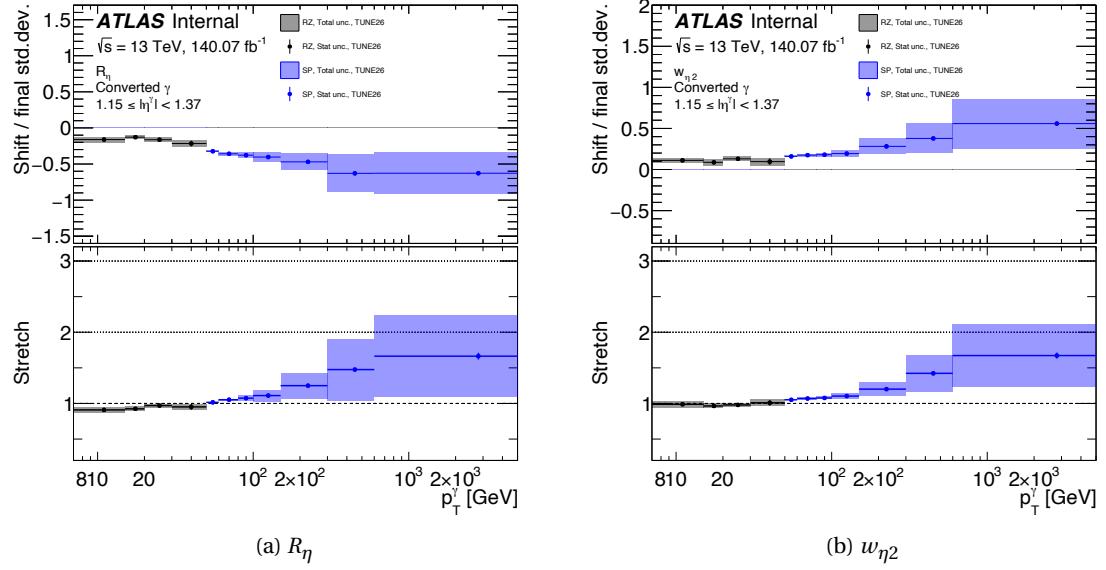


Figura 5.4: Valores de los FFs de shift y stretch para las  $R_\eta$  (izquierda) y  $w_{\eta 2}$  (derecha) para fotones convertidos con  $1.15 < |\eta| < 1.37$ , en función de  $p_T$ . Los resultados obtenidos por las muestras de **RZ** están representados por el color negro, mientras que los resultados de **SP** se muestran en azul. Los puntos y las líneas denotan los valores centrales con sus incertezas estadísticas, mientras que las regiones sombreadas representan las incertezas totales. Los valores de shift se muestran en el panel superior, los cuales son normalizados por el ancho de la distribución luego de ser estirada por el stretch, como se ha explicado en el texto. Este último valor se muestra en el panel inferior de las figuras.

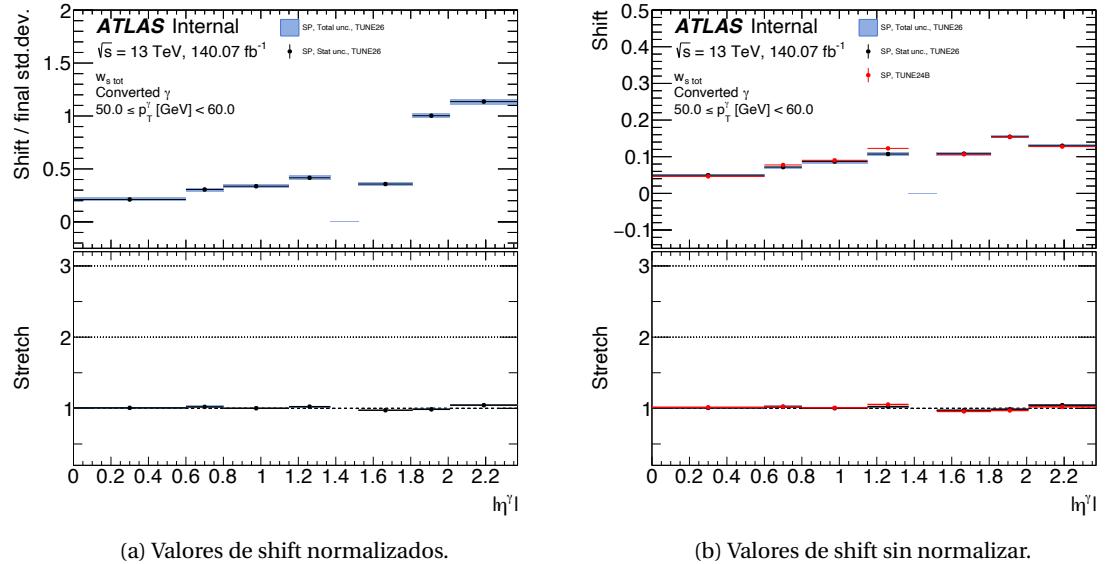


Figura 5.5: Valores de los FFs de shift y stretch para  $w_{s \text{ tot}}$  en función de  $|\eta|$  utilizando fotones convertidos con  $50 < p_T < 60$  GeV de las muestras de **SP**. La Figura a muestra los valores de shift normalizados, mientras que los no normalizados se encuentran en la Figura b. Los puntos con las líneas de color muestran los valores centrales y las incertezas estadísticas, mientras que las áreas sombreadas representan las incertezas totales en cada bin. Los valores de stretch se muestran en los paneles inferiores de cada figura.

más picos, y que se corrigen correctamente con el método de FFs. En todos los casos mostrados, el MC corregido y los datos son casi indistinguibles, lo que demuestra la importancia de estas correcciones y cómo logran un excelente acuerdo.

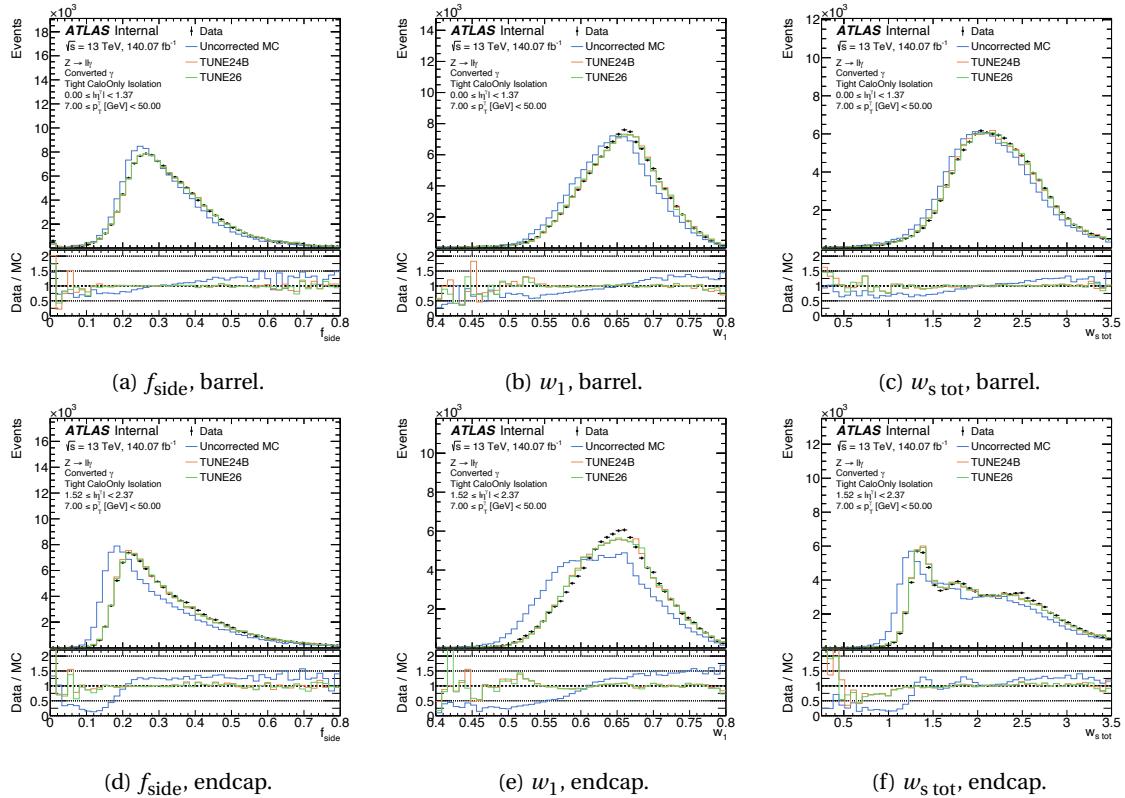


Figura 5.6: Distribuciones de algunas SSs seleccionadas usando las muestras de RZ para fotones convertidos luego de aplicar las correcciones de los FFs en la simulación. Las distribuciones de las SS están separadas para fotones en la región del barrel (fila de arriba) y en la región del endcap (fila de abajo). Los puntos negros representan los datos recolectados por ATLAS, mientras que las simulaciones no corregidas y corregidas están mostradas por las líneas azules y verdes, respectivamente. Los paneles inferiores, en cada figura, muestra el cociente entre el histograma de datos con cada uno de los obtenidos de las simulaciones MC.

## 5.2. Correcciones de energía de las celdas

El diseño y la funcionalidad del ECAL de ATLAS se describió en la Sección 2.2.3, así como el proceso a partir del cual los electrones y los fotones depositan sus energías en el ECAL: creación de pares y radiación bremsstrahlung. Luego, a partir de estas deposiciones de energía en el ECAL se construyen los SSs y se utilizan para la identificación de fotones. Sin embargo, el hecho de que las SSs calculadas mediante las simulación MC y los datos SSs no coincidan, significa que las deposiciones de energía son diferentes entre estos dos, lo que lleva a un desacuerdo a un nivel inferior.

Aunque el método de Fudge Factor (FF) descripto anteriormente condujo a una excelente mejora del acuerdo entre los datos y las distribuciones MC, sigue basándose en la modificación de variables de alto nivel y todas independientemente unas de otras. En cambio, otro enfoque diferente es el de corregir directamente los depósitos de energía de las celdas en la simulación MC. Esto permitiría calcular todas las Shower Shapes (SSs) y cualquier otra variable que utiliza la energía de las celdas ya corregidas.

El enfoque de corregir las energías de las celdas del ECAL se ha desarrollado y probado inicialmente para electrones [72], y posteriormente para fotones [73]. Para el caso de los electrones,

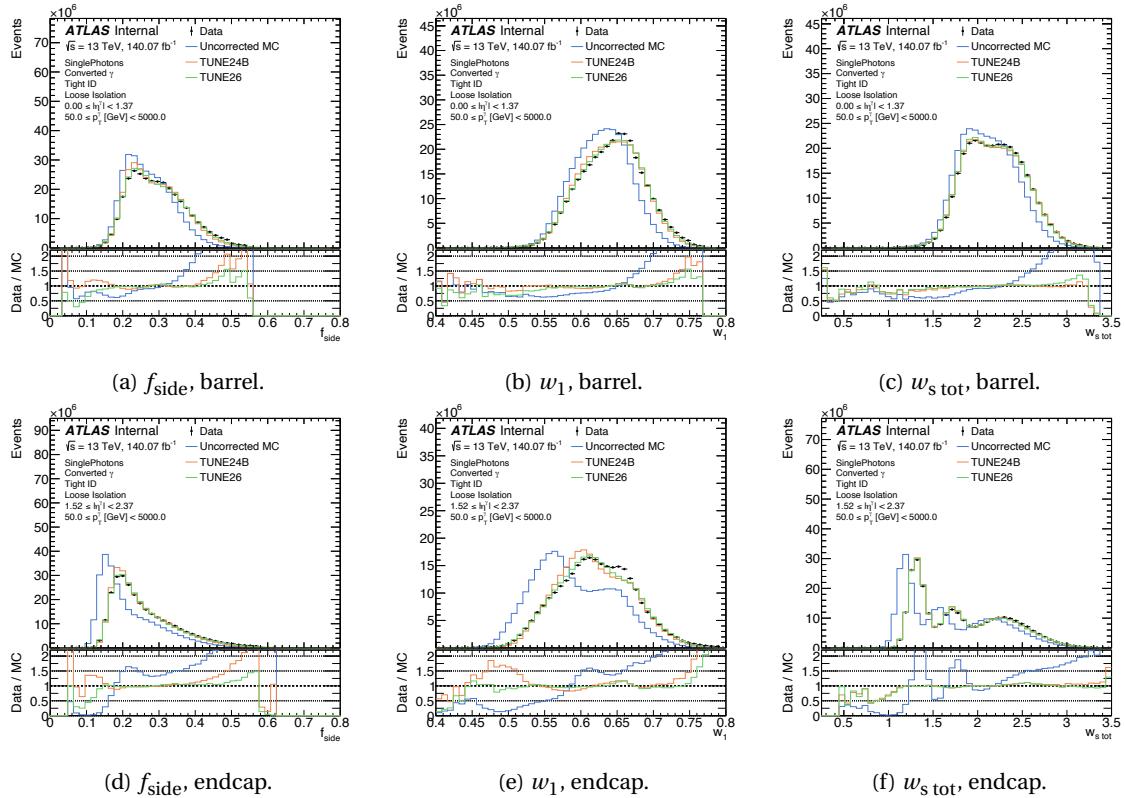


Figura 5.7: Ídem a la Figura 5.6 pero utilizando las muestras de SP.

los resultados han sido muy prometedores, ya que se corrigieron sustancialmente las SSs de la segunda capa del calorímetro. Sin embargo, para los fotones, el mismo método que se utilizó para los electrones no funcionó de la forma que se esperaba, ya que sólo permitió corregir las energías en promedio. Otro enfoque para corregir la simulación se basó en hacer coincidir eventos de datos y eventos simulados, estudio que sólo fue probado pseudodatos y resultó técnicamente complicado, pero que condujo a mejores resultados [73].

En la presente sección, se estudia una nueva forma de corregir las energías de celda en MC, utilizando sólo la segunda capa del ECAL, por simplicidad. El método comparte similitudes con el método FF, lo que además facilita su comprensión. En primer lugar, se presenta la selección de eventos especiales utilizada para este estudio. Se discute brevemente el método de corrección de energías utilizado por los primeros estudios basados en electrones y fotones, y luego se presenta en detalle cómo se mejora este método.

### 5.2.1. Selección de eventos

Los estudios presentados en esta sección se llevan a cabo con el mismo conjunto de datos utilizado para el cálculo FF, descripto en la Sección 5.1.1. Sin embargo, en este caso sólo se utilizan las muestras de RZ. Los eventos se seleccionan como se describe en la Sección 4.2.1, utilizando fotones que pasan el criterio de aislamiento loose. Sin embargo, dado que estos estudios se basan en la información de la segunda capa del ECAL, es necesario tener en cuenta una selección especial de las celdas que la conforman.

Cuando un electrón o fotón entra en el calorímetro, su huella en la segunda capa es un grupo visible de celdas que rodean a la más energética y central (también denominada *hottest cell*). En este estudio, se consideran clusters de  $7 \times 11$  celdas en  $\eta \times \phi$ , mostradas en la Figura 5.8a donde también se muestra la disposición utilizada. Aproximadamente, el 90% de la energía del cluster se reparte entre las 9 celdas centrales, resaltadas en azul en la Figura 5.8a. La energía media normalizada de los datos se muestra en la Figura 5.8b, visualizando cómo se distribuye la energía.

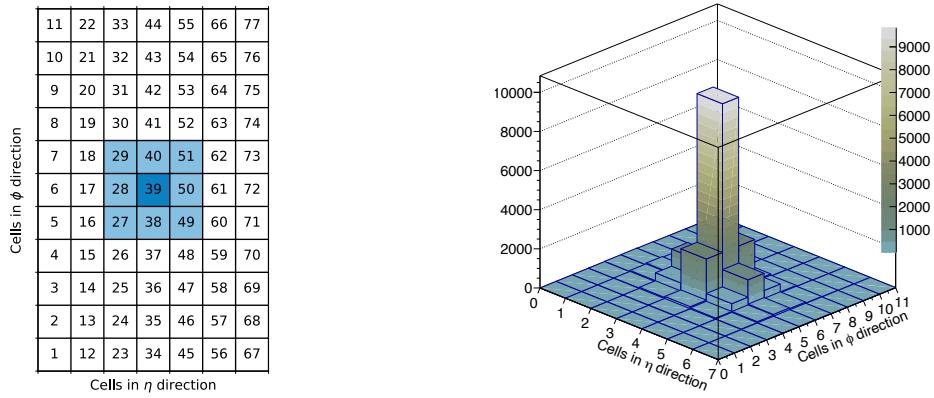


Figura 5.8: Disposición de las celdas y distribución de la energía entre las celdas del cluster.

En este trabajo, sólo se consideran los eventos en los que los clusters tienen el total de las 77 celdas. Además, se requiere en los eventos que la celda central sea la más energética.

## 5.2.2. Cálculo de las correcciones

### Primeros pasos

Todos los eventos que superen la selección mencionada tendrán asociado un cluster, cada uno de los cuales tendrá  $N$  celdas y cada celda tendrá una energía  $E_i$ , con  $i = 1, \dots, N$ . Para cada evento, en primer lugar, se obtiene la energía total del cluster  $E$  sumando las energías de cada una de las celdas  $E_i$ . El método de las correcciones de las SSs mediante la corrección de las energías depositadas en el ECAL hace uso de las energías normalizadas en cada celda,  $e_i = E_i/E$ . Estos valores dan a entender qué proporción de la energía total depositada tiene una celda en particular.

El proceso de corrección comienza entonces calculando el valor medio de las distribuciones  $e_i$  (obtenidas una vez que todas los eventos pasan la selección) para la  $i$ -ésima celda, en la simulación MC y en los datos, y la diferencia entre estos valores dan lugar a la corrección  $\Delta_i$  en

dicha celda:

$$\Delta_i = \overline{\left( \frac{E_i^{\text{data}}}{E^{\text{data}}} \right)} - \overline{\left( \frac{E_i^{\text{MC}}}{E^{\text{MC}}} \right)} = \bar{e}_i^{\text{data}} - \bar{e}_i^{\text{MC}}. \quad (5.6)$$

Los valores  $E^{\text{data/MC}}$  son las energías totales del cluster para los datos y MC, respectivamente.

La energía de la celda  $i$ , se corrige entonces como

$$E_i^{\text{MC-RW}} = E_i^{\text{MC}} + \Delta_i E^{\text{MC}}, \quad (5.7)$$

que se traduce en desplazar la energía normalizada de la celda  $e_i^{\text{MC}}$  en una cantidad  $\Delta_i$ , para que los valores medios de las distribuciones de  $e_i$  de datos y MC coincidan.

También es importante notar que, por definición, estos coeficientes de corrección suman 0 en todo el cluster:

$$\sum_i \Delta_i = \sum_i \overline{\left( \frac{E_i^{\text{data}}}{E^{\text{data}}} \right)} - \sum_i \overline{\left( \frac{E_i^{\text{MC}}}{E^{\text{MC}}} \right)} = \overline{\sum_i \frac{E_i^{\text{data}}}{E^{\text{data}}}} - \overline{\sum_i \frac{E_i^{\text{MC}}}{E^{\text{MC}}}} = 1 - 1 = 0,$$

implicando que el cambio de energía total del cluster se mantiene constante:

$$E^{\text{MC-RW}} \equiv \sum_i E_i^{\text{MC-RW}} = \sum_i E_i^{\text{MC}} + \sum_i \Delta_i E^{\text{MC}} = E^{\text{MC}} + E^{\text{MC}} \sum_i \Delta_i = E^{\text{MC}}.$$

Este hecho es de vital importancia, ya que no se desea cambiar la energía total del cluster en la simulación MC, sino que se desea lograr una redistribución de la energía entre las celdas, de forma tal que cada una se asemeje a la de los datos.

Los coeficientes de corrección resultantes para cada celda en clusters de 77 celdas, se pueden visualizar en la Figura 5.9. Como se puede notar de los valores mostrados, la celda central presenta una corrección negativa, mientras que las 8 vecinas a la central tienen correcciones positivas. Esto se puede traducir a que en la simulación, la celda central suele tener más energía, en promedio, que en los datos, mientras que lo opuesto ocurre en las vecinas. Mediante la aplicación de una corrección negativa (implicando un corrimiento negativo de  $e_i$ ), se remueve energía de la celda central que luego es distribuida en las circundantes.

A partir de las energías de las celdas, se pueden calcular las SSs de la segunda capa del ECAL, las cuales son  $R_\eta$ ,  $R_\phi$  y  $w_{\eta 2}$ :

$$R_\eta = \frac{E_{3 \times 7}}{E_{7 \times 7}}$$

$$R_\phi = \frac{E_{3 \times 3}}{E_{7 \times 3}}$$

$$w_{\eta 2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left( \frac{\sum_i E_i \eta_i}{\sum_i E_i} \right)^2}$$

donde  $E_{i \times j}$  es la energía de la celda sumada en una región de  $\eta \times \phi = i \times j$  celdas alrededor de la celda central. Se demostró en los estudios anteriores [73] que este método sólo corrige las formas de las variables en promedio, pero las diferencias en la forma de permanecer. Esto se debe al hecho de que este método sólo corrige los valores medios de energía en las celdas. Sin

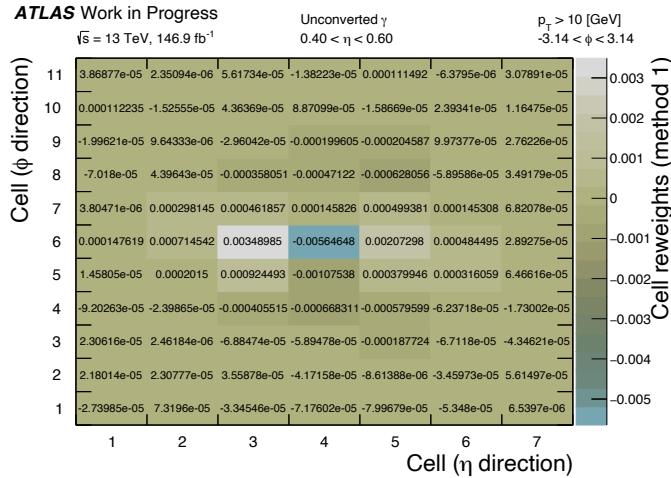


Figura 5.9: Correcciones a las energías de las celdas de la simulación MC utilizando el mismo método diseñado para electrones. **fix luminosity**

embargo, estas distribuciones de energía siguen presentando diferencias, especialmente en lo que se refiere a las formas, lo que conduce a una situación muy similar a la observada para los **EES**. De este modo, se puede emplear un enfoque muy similar para corregir los valores medios  y los anchos de las distribuciones de energía normalizadas.

### Nuevo método de corrección de energías

Este nuevo método pretende corregir tanto el valor medio como la varianza de las distribuciones normalizadas de energía de las celdas, mediante la aplicación de corrimientos (shift) y estiramientos (stretch) de las mismas. De forma similar al enfoque seguido para las SSs utilizando el método de FFs, una primera aproximación a los valores de shift y stretch de las distribuciones de energía consiste en calcular el valor medio y la raíz cuadrática media (RMS) de las mismas en cada celda, respectivamente. Luego, la energía normalizada de la  $i$ -ésima celda se obtiene como:

$$e_i^{\text{MC-RW}} = \underbrace{\frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} e_i^{\text{MC}}}_{\text{stretch}} + \underbrace{\left( \bar{e}_i^{\text{data}} - \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \bar{e}_i^{\text{MC}} \right)}_{\text{shift}}, \quad (5.8)$$

donde el subíndice  $e$  en los valores de RMS indica que estos se calculan a partir de las distribuciones de energía normalizadas, y el índice  $i$  recorre todas las celdas del cluster. De la expresión anterior se pueden identificar nuevamente un factor de shift, que es una transformación constante de la energía normalizada, y un factor de stretch, lineal en la variable que se requiere corregir.

Dado que la energía normalizada en la celda  $i$  puede calcularse como  $e_i^j = E_i^j / E^j$ , para  $j = \text{MC-RW}, \text{MC}$  y datos, y se requiere tener la misma energía total del cluster luego de aplicar las correcciones ( $E^{\text{MC-RW}} = E^{\text{MC}}$ ), se puede multiplicar la Ec. 5.8 por  $E^{\text{MC-RW}}$  y llegar a una expresión

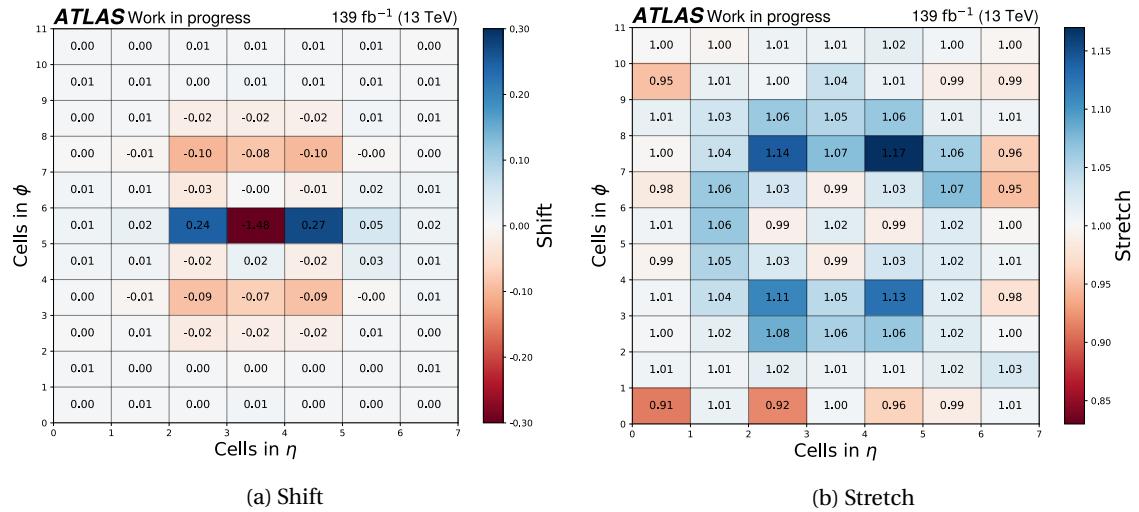
para  $E_i^{\text{MC-RW}}$ :

$$E_i^{\text{MC-RW}} = \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} E_i^{\text{MC}} + \left( \bar{e}_i^{\text{data}} - \frac{\text{RMS}_{e,i}^{\text{data}}}{\text{RMS}_{e,i}^{\text{MC}}} \bar{e}_i^{\text{MC}} \right) E_i^{\text{MC}}. \quad (5.9)$$

Por último, para garantizar que la energía del cluster permanezca constante, las energías de las celdas se **rescalan** por  $\sum_i E_i^{\text{MC}} / \sum_i E_i^{\text{MC-RW}}$ .



Como el resultado de este procedimiento de corrección de energías involucra una corrección de shift y otra de stretch, se obtienen dos matrices de corrección, y un ejemplo de ellas se presenta en la Figura 5.10. En lo que sigue, este nuevo método se aplica para corregir las energías de las celdas, y se computa de forma inclusiva en  $p_T$  y  $|\eta|$ , sólo separando entre fotones no convertidos y convertidos.



(a) Shift

(b) Stretch

Figura 5.10: Ejemplo de las matrices de corrección de shift (izquierda) y stretch (derecha). Los valores mostrados corresponden al cálculo de las correcciones utilizando fotones no convertidos. Los valores de shift son multiplicados por un factor de 100 para mejorar su visualización.

### 5.2.3. Resultados

La Figura 5.10 muestra las matrices de corrección de shift y stretch obtenidas para fotones no convertidos. Puede observarse que, al igual que en el caso del método anterior, la mayor corrección de shift se realiza en la celda central, donde el shift corresponde a un valor negativo. De la misma forma que en el caso anterior, los shifts de las celdas vecinas en la dirección de  $\eta$  son positivos y grandes, indicando la redistribución de la energía de la celda central en estas dos vecinas. Sin embargo, se puede notar que las segundas celdas vecinas en la dirección de  $\phi$  sufren una gran corrección, quitando energía mediante el shift, pero aumentando también el ancho de la distribución, dado por los estiramientos positivos. Del resto de las celdas del cluster, se nota que no presentan corrimiento significativo, pero presenta un stretch  $< 1$ , indicando que se hacen más angostas, especialmente las celdas de los extremos del cluster.

Utilizando estos factores de corrección para las energías normalizadas de cada celda, en las Figura 5.11 se muestran las distribuciones de energía normalizadas resultantes para las celdas

28, 39 y 50<sup>2</sup>. El nuevo método de corrección consigue grandes mejoras en el acuerdo entre los datos y la simulación. Además, el método logra corregir bien las colas de las distribuciones de todas las celdas, así como los picos de las mismas, lo que puede observarse especialmente en la celda 28.

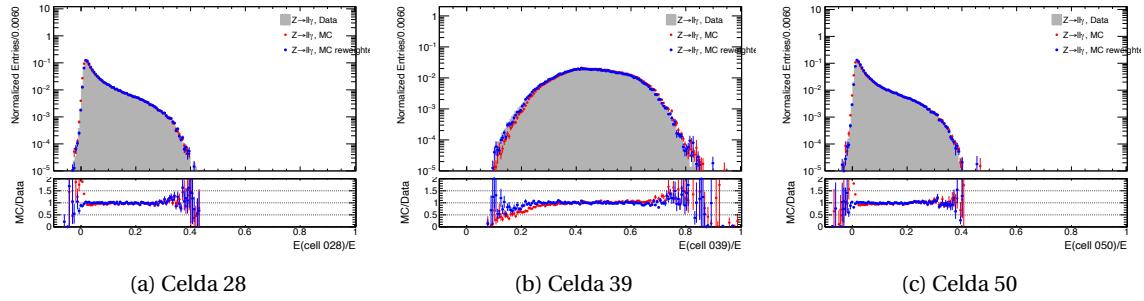


Figura 5.11: Distribuciones de las energías normalizadas de las celdas 28, 39 y 50 de cluster de 77 celdas, para fotones no convertidos. Los puntos azules y rojos corresponden a las distribuciones de la simulación MC con y sin las correcciones, respectivamente, mientras que el histograma gris representa los datos.

Para evaluar el comportamiento del nuevo procedimiento de corrección aplicado a las SSs de la segunda capa ECAL, en la Figura 5.12 se muestra la comparación de los métodos de corrección para las variables  $R_\eta$ ,  $R_\phi$  y  $w_{\eta 2}$ . En los tres casos, se observa una mejora con respecto al MC sin corregir, especialmente para  $R_\phi$  y  $w_{\eta 2}$ . El método de corrección de energía, en el caso de fotones no convertidos, no alcanza el nivel de acuerdo con los datos logrado por el método de FFs, que ha demostrado proporcionar una excelente acuerdo con los datos experimentales. Sin embargo, casi no se observan diferencias entre el método de corrección de energías y el de FFs para fotones convertidos, lo que indica que aún hay margen de mejora en las correcciones.

### 5.3. Conclusiones y trabajo futuro

En el presente capítulo se han estudiado dos métodos para corregir el desacuerdo observado en las Shower Shapes (SSs) entre los datos y la simulación MC.

El método de Fudge Factor (FF) se ha utilizado históricamente en la colaboración, al principio basado únicamente en simples desplazamientos de las distribuciones. A pesar de que las correcciones conducían a buenas mejoras y por tanto a la obtención de mejores SFs, seguían existiendo notables diferencias de forma entre los datos y la simulación. En el contexto de este trabajo, al añadir un término lineal a la transformación de la variable, se logra corregir los anchos de las distribuciones simuladas, lo que conduce a un acuerdo aún mejor con los datos. Este nuevo método de corrección de SSs mediante FFs se denomina método shift+stretch y actualmente se utiliza por toda la colaboración ATLAS.

También se ha desarrollado un método novedoso y de menor nivel de corrección que pretende modificar las energías en las celdas del ECAL. Utilizando las distribuciones de energía en cada celda en clusters alrededor de la celda más energética, es posible corregir todas las SSs en

<sup>2</sup> Como fue mostrado en la Figura 5.8a, la celda número 39 es la central, mientras que las celdas 28 y 50 están a la izquierda y derecha, respectivamente, en la dirección  $\eta$ .

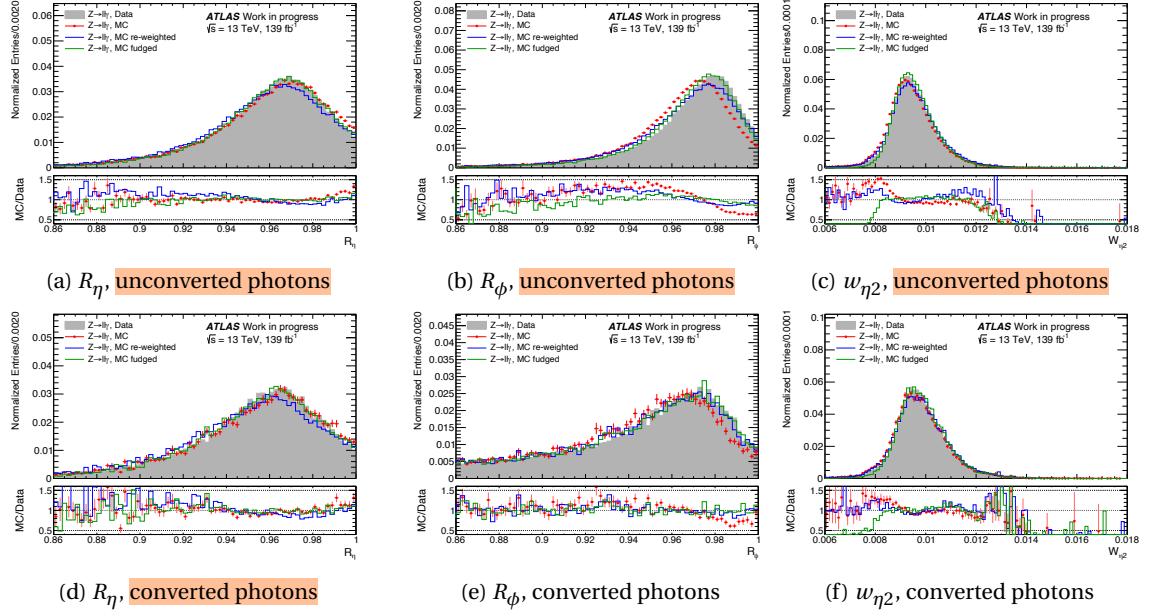


Figura 5.12: Distribuciones de las SSs calculadas en la segunda capa del ECAL para fotones no convertidos (fila superior) y convertidos (fila inferior) con pseudorapidez  $|\eta| < 0.6$ , comparando los diferentes métodos de corrección con los datos. Los datos experimentales están representados por los histogramas grises. La simulación MC sin corregir se muestra con los puntos rojos, la simulación corregida por el método de corrección de energías con la línea azul y la corregida por el método de FFs con la línea verde.

simultáneo. Este método usa la misma estrategia de shift+stretch, pero esta vez aplicado a las distribuciones de energía normalizada en cada celda de la simulación MC, para que coincida con la distribución encontrada en los datos. Aunque el método es nuevo y aún necesita de mejoras, como también extenderlo a las demás capas del ECAL, ha dado resultados prometedores en los que algunas variables se corrigen de la misma manera que con el FFs. El método de corrección de las energías de las celdas muestra un gran potencial en la colaboración, no sólo en el contexto de la identificación de fotones *offline*, sino también a nivel de trigger.

### 5.3.1. Trabajo a futuro

Uno de los enfoques más interesantes y prometedores para corregir el SSs es el método basado en las correcciones de las energías de las celdas. Como se ha mencionado anteriormente, este enfoque podría emplearse en diferentes pasos del proceso de identificación de fotones, como en el nivel de trigger, o de forma *offline* para corregir todos los SSs simultáneamente. Otro uso potencial e importante es utilizar los clusters corregidos para calcular directamente la identificación de fotones, por ejemplo, considerando a los clusters como imágenes y utilizando una red neuronal convolucional (CNN) para realizar la identificación de fotones [73].

Las Shower Shapes (SSs) tienen la gran ventaja de que se pueden interpretar fácilmente en términos físicos. Por esta razón, mantener estas variables sirve para comprender la física subyacente de los procesos. Seguir corrigiendo estas variables es de gran interés y hay varias formas de hacerlo. El método actual de transformar la variable pero utilizando términos de orden superior sigue siendo una tarea difícil, pero aún no explorada. Haciendo uso de las novedosas

técnicas de Machine Learning (ML), es posible obtener factores de corrección para los términos de orden superior en la expansión, corrigiendo además los momentos de orden superior de las distribuciones (asimetría estadística, curtosis, etc.). Otro enfoque interesante es el uso de un re-escaleo Multivariable ([MV](#)), que se exploró en Ref. [74], mostrando resultados muy prometedores.