

CORRECCIONES DE LAS SHOWER SHAPES

5

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, lo que se origina en las diferencias de las Shower Shapes (SSs). Desde el Run-1, estas se han corregido con los Fudge Factors (FFs), que se calculaban como simples desplazamientos a las distribuciones MC proporcionando muy buenas mejoras de los SFs. Sin embargo, siguen habiendo discrepancias entre las distribuciones. 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 temas de trabajo de la presente tesis.

5.1. Fudge Factors

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 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 [69], mientras que SHERPA 2.2.1 se utiliza para los eventos de fondo $Z \rightarrow ll$. Respecto a las muestras SP, los eventos se generan con PYTHIA 8.186 [66], 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 de dos jets (QCD dijet).

En ambos casos el detector ATLAS se simula utilizando GEANT4 [76] y los eventos MC se normalizan 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

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, hay variables muy sensibles al estado de conversión del fotón de modo que el cálculo se hace por separado para fotones convertidos y no convertidos. En total se corren nueve variables con este método: E_{ratio} , f_{side} , R_η , R_ϕ , R_{had} , $R_{\text{had}1}$, w_1 , $w_{\eta2}$ y w_{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 realiza para disponer de estadística suficiente en cada bin y para capturar todas las características de las variables. Luego, cada histograma se suaviza utilizando una herramienta del paquete de TMVA [146] 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- δ 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

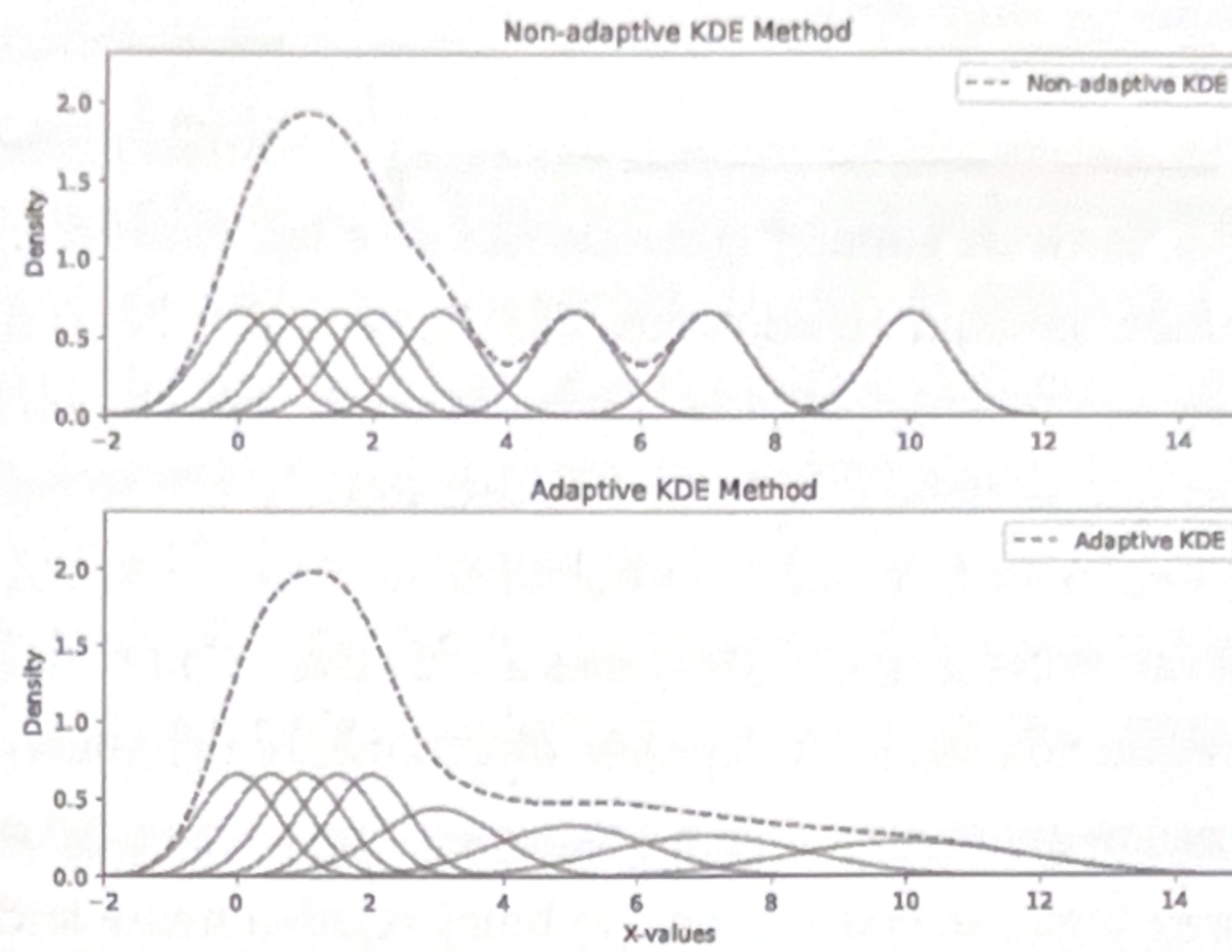


Figura 5.1: Esquema del suavizado no adaptativo y adaptativo del método KDE.

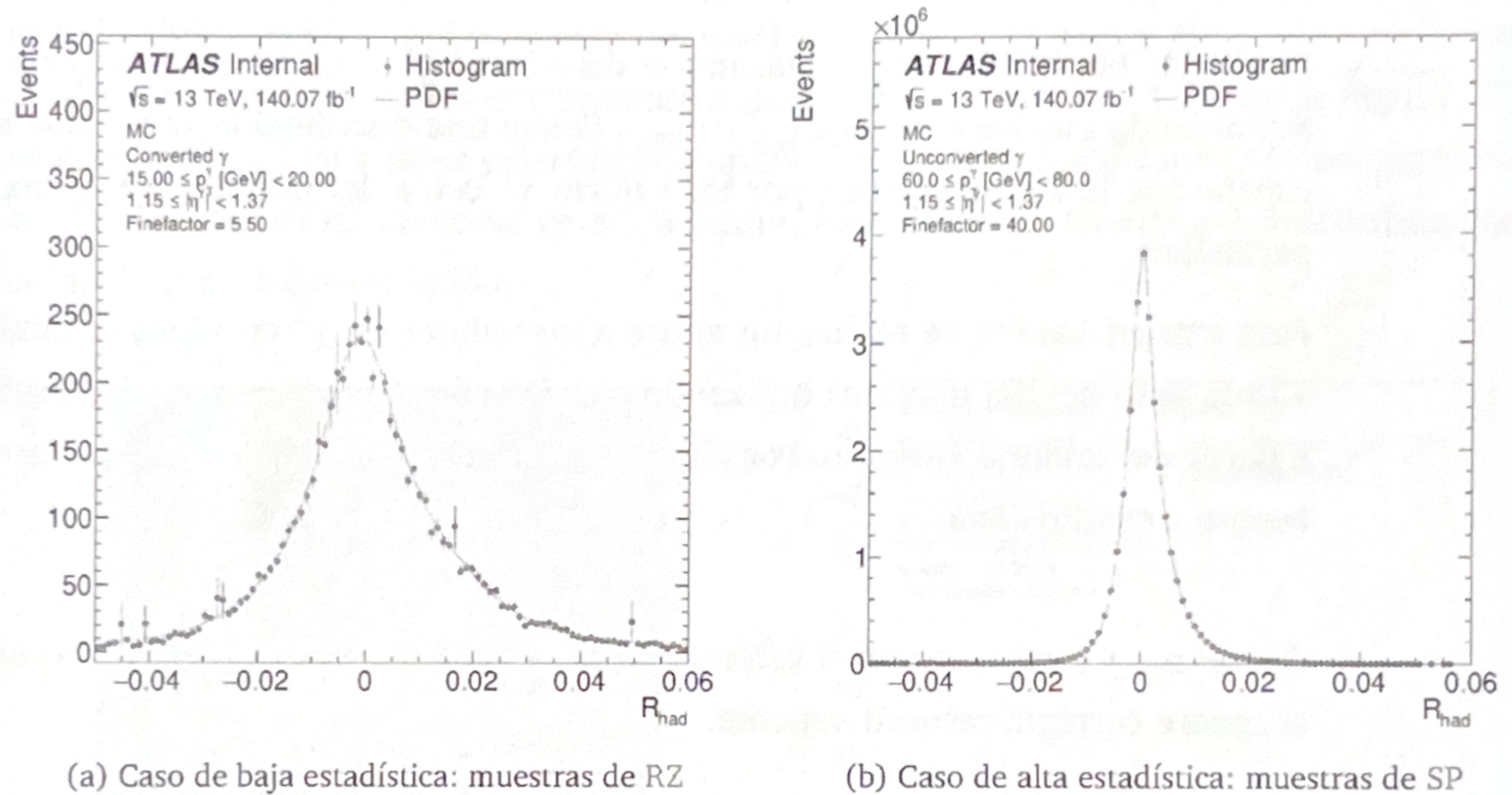


Figura 5.2: Suavizado de la SS R_{had} utilizando el método KDE para fotones en $1.15 < |\eta| < 1.37$ bajo dos posibles escenarios: baja y alta estadística. El histograma original se muestra con los puntos azules y las correspondientes PDFs con la línea naranja. Además, se muestran los valores de los fine factors usados en cada caso.

suavizado aplicado a R_{had} para casos en los que los histogramas originales tienen baja y alta estadística.

Una vez creadas las PDFs de los datos y la simulación MC para una dada variable, bin de p_T y $|\eta|$ y tipo de conversión, la PDF de MC se normaliza a la de los datos y se calcula un valor χ^2 entre ambos como [147]:

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

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_{data} y W_{MC} son la suma de los pesos en datos y MC, respectivamente.

Correcciones *shift-only*

*hasta el
presente*

Como ha sido mencionado anteriormente¹, 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 hasta 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. Después 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 χ^2 . Suponiendo que los errores $s_{MC,i}$ y $s_{data,i}$ tienen una distribución gaussiana estándar¹, se espera que la forma seguida por los valores χ^2 cerca del mínimo sea aproximadamente parabólica.

Para extraer los FFs se realiza un ajuste a los valores de χ^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_{old} y x representan el valor original y el valor post-corrección de la variable *la cual* se quiere corregir, respectivamente.

Correcciones *shift+stretch*

Continuación

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. En el contexto de esta tesis se propuso continuar mejorando las distribuciones introduciendo una corrección de orden superior a las simulaciones MC, denominada *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

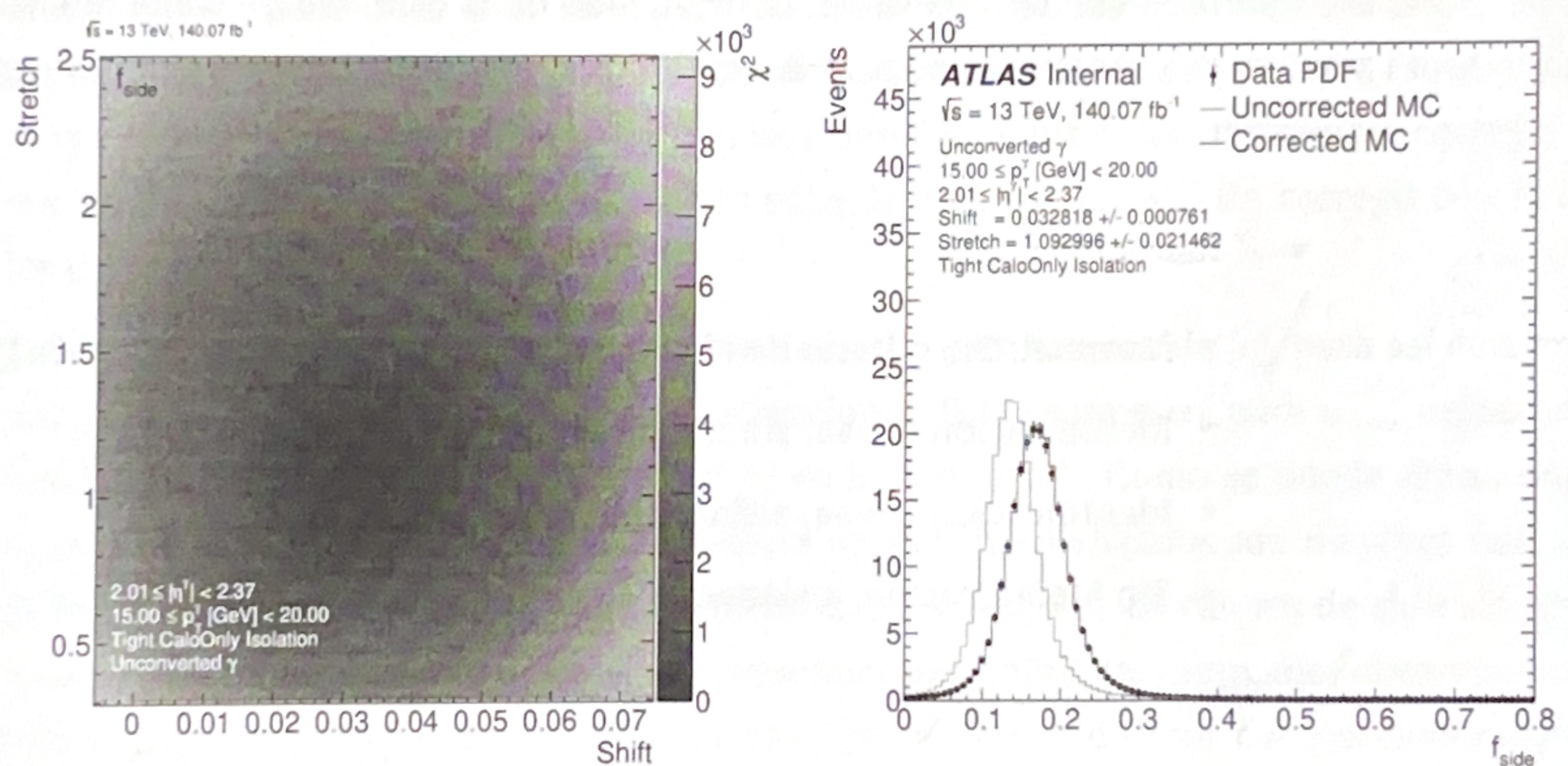
¹ Este requisito se cumple siempre que los contenidos de los bines de ambas PDFs sean mayores que 10, lo que también se satisface puesto que los histogramas se construyen con bines relativamente amplios.

distribución, multiplicada por el factor de stretch. En el escenario en el que el stretch es > 1 , puede haber casos en 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 χ^2 para cada shift_i después de aplicar el stretch_j. Como resultado de este procedimiento, ahora se obtiene una grilla bidimensional de valores de χ^2 en el plano de shift-stretch. El par shift-stretch se obtiene del centro del bin mínimo y comprende ahora los FFs. Las correcciones pueden ser aplicadas a cada SS x , evento a evento, como:

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

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

Un ejemplo de los valores de χ^2 resultantes para la variable f_{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.



(a) Valores de χ^2 en el plano shift-stretch.

(b) PDFs de datos (puntos negros) y simulación MC sin corregir (línea roja) y corregida (línea azul), mostrando el impacto de las correcciones.

Figura 5.3: Cálculo de los FFs de shift+stretch para f_{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

Incertezas estadísticas

Para extraer las incertezas estadísticas de los FFs de shift y stretch, se realiza un ajuste al contorno de 1σ (nivel de confianza del 68.3%) sobre los valores χ^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 ρ es el coeficiente de correlación entre ambas variables, σ_x y σ_y las incertezas sobre x y y , respectivamente, (x_0, y_0) es la posición del centro de la elipse, y χ_{\min}^2 es el valor mínimo de χ^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):

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

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

Incógnitas 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 TightCaloOnly.
- Identificación Loose, sin aislamiento.
- Identificación Loose, aislamiento TightCaloOnly.
- Sin identificación, aislamiento Loose.

- Single Photon (SP):

- Nominal: identificación Tight, aislamiento Loose.
- identificación Tight, aislamiento Tight .

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 donde ocurre la superposición entre ambas.

- En la Figura 5.4 se presentan ejemplos de los FFs resultantes para las variables R_η y $w_{\eta 2}$ utilizando fotones convertidos. 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. 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 deficiente modelado de las SSs por el MC. Como se mencionó en la Sección 4.3, se observaron anchos y perfiles en η más amplios para los datos en comparación con la simulación. De hecho, esto se puede inferir dado que los valores de stretch aumentan a valores más altos de p_T , estirando las simulaciones MC hasta el doble de su ancho inicial. En el caso de R_η ($w_{\eta 2}$) mostrado, la simulación MC sobreestima (subestima) el valor central de la distribución en casi una desviación estándar después de corregir el 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 marcado, 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 SSs luego de aplicar los FFs 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 observa un acuerdo excelente 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 más picos, y que se corrigen correctamente con el método de FFs.

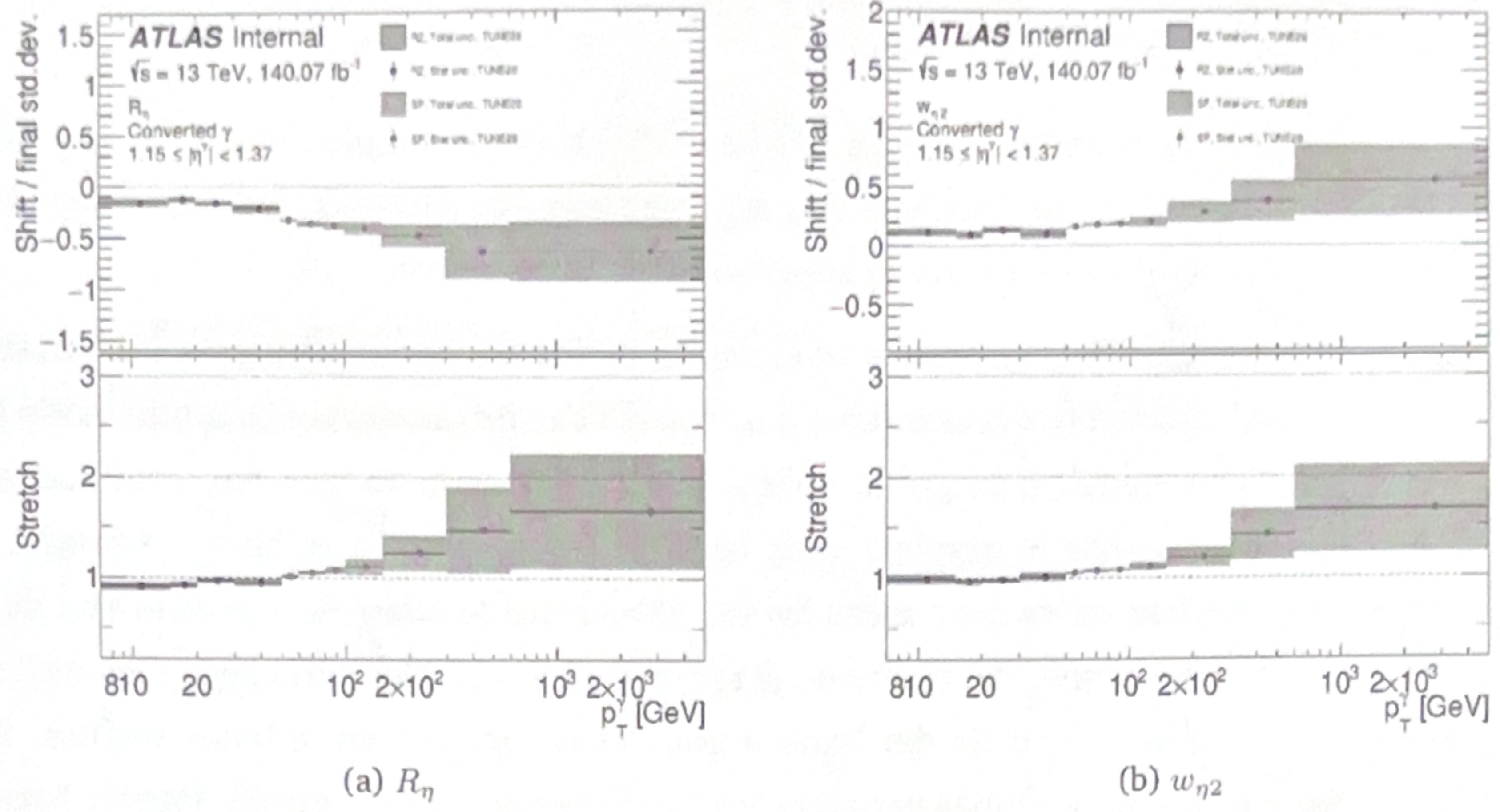


Figura 5.4: Valores de los FFs de shift y stretch para las R_η (izquierda) y $w_{\eta 2}$ (derecha) para fotones convertidos con $1.15 < |\eta| < 1.37$, en función de p_T . Los resultados obtenidos por las muestras de RZ están representados por el color negro, mientras que los resultados de SP se muestran 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.

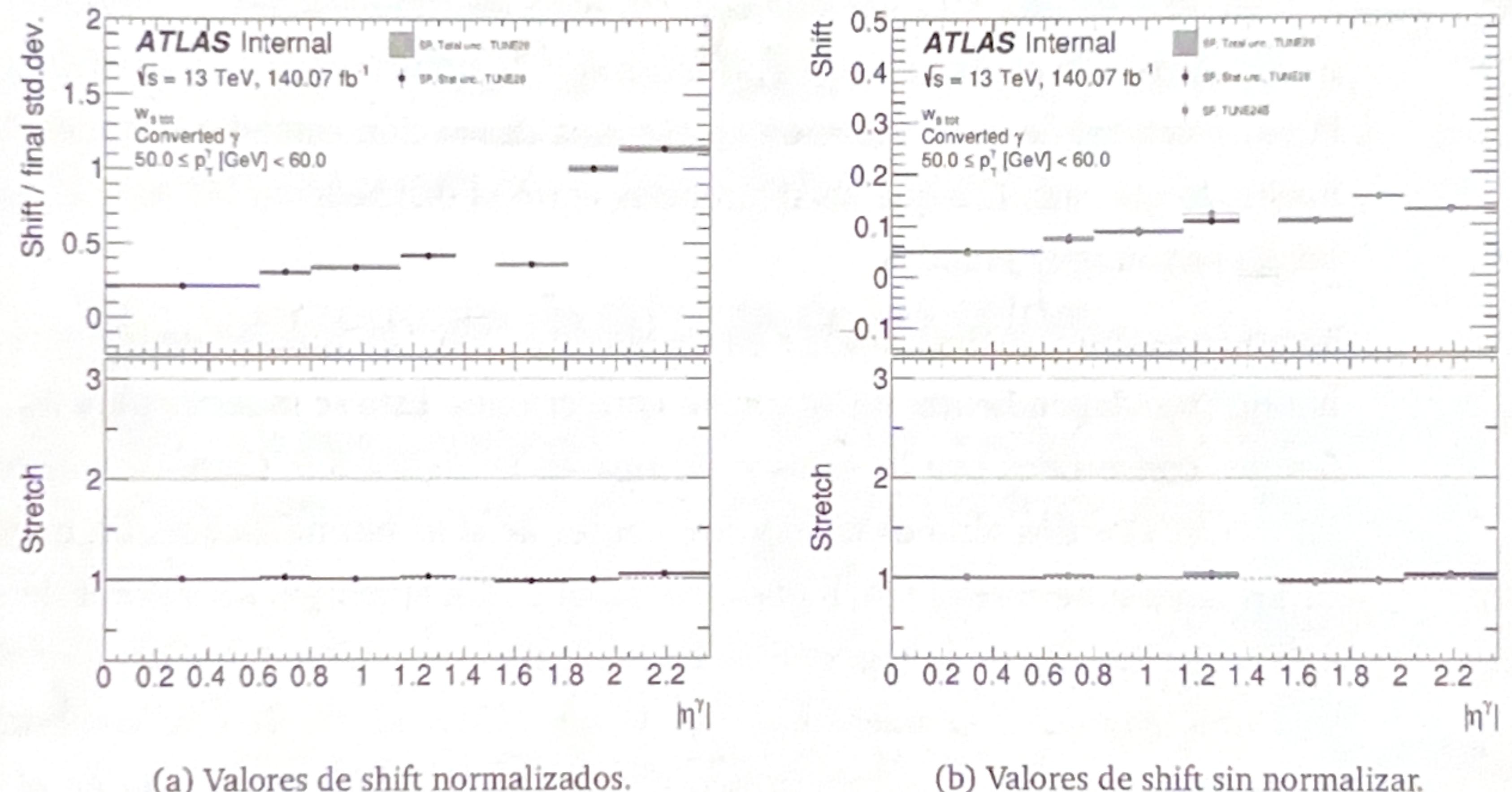


Figura 5.5: Valores de los FFs de shift y stretch para w_{tot} en función de $|\eta|$ utilizando fotones convertidos con $50 < p_T < 60 \text{ GeV}$ de las muestras de SP. La Figura a muestra los valores de shift normalizados, mientras que los no normalizados se encuentran en la Figura b. Los puntos 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.

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.

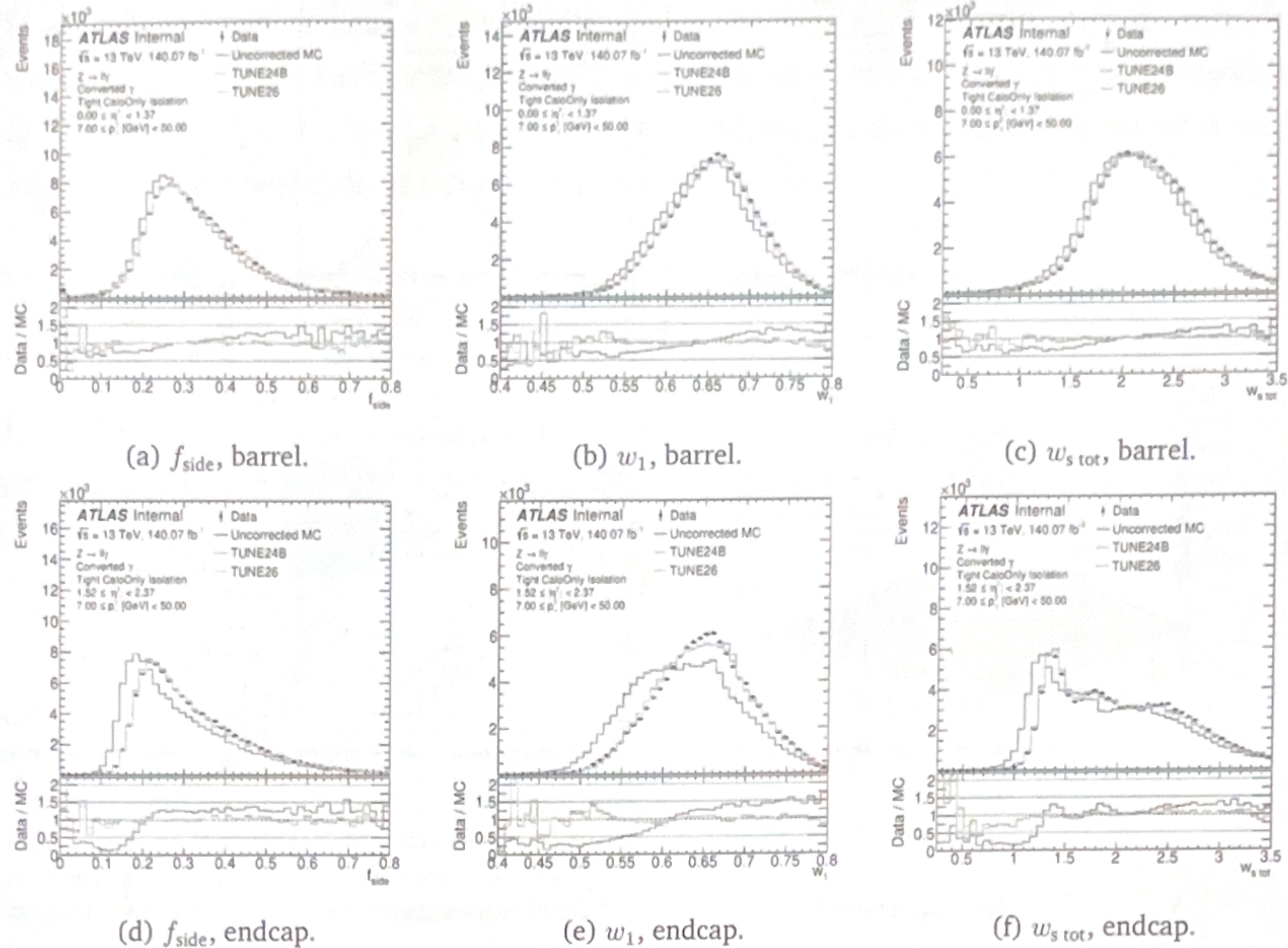


Figura 5.6: Distribuciones de algunas SSs seleccionadas usando las muestras de RZ para fotones convertidos luego de aplicar las correcciones de los FFs en la simulación. Las distribuciones de las SS están separadas para fotones en la región del barrel (fila de arriba) y en la región del endcap (fila 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. El panel inferior 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 no coincidan indica que las deposiciones de energía son diferentes entre estos dos, significando que el desacuerdo ocurre, en realidad, a un nivel inferior.

Aunque el método de FF descripto anteriormente condujo a una excelente mejora del acuerdo entre los datos y las distribuciones MC, sigue siendo una modificación en las variables de alto nivel y todas independientemente unas de otras. Un enfoque diferente, diferente ~~solo~~ modificar directamente los depósitos de energía de las celdas en la simulación MC, corrigiendo todas las variables calculadas a partir de estas energías corregidas, incluidas las SSs.

El enfoque de corregir las energías de las celdas del ECAL se ha desarrollado y probado inicialmente para electrones [148] y posteriormente para fotones [149]. Para el caso de

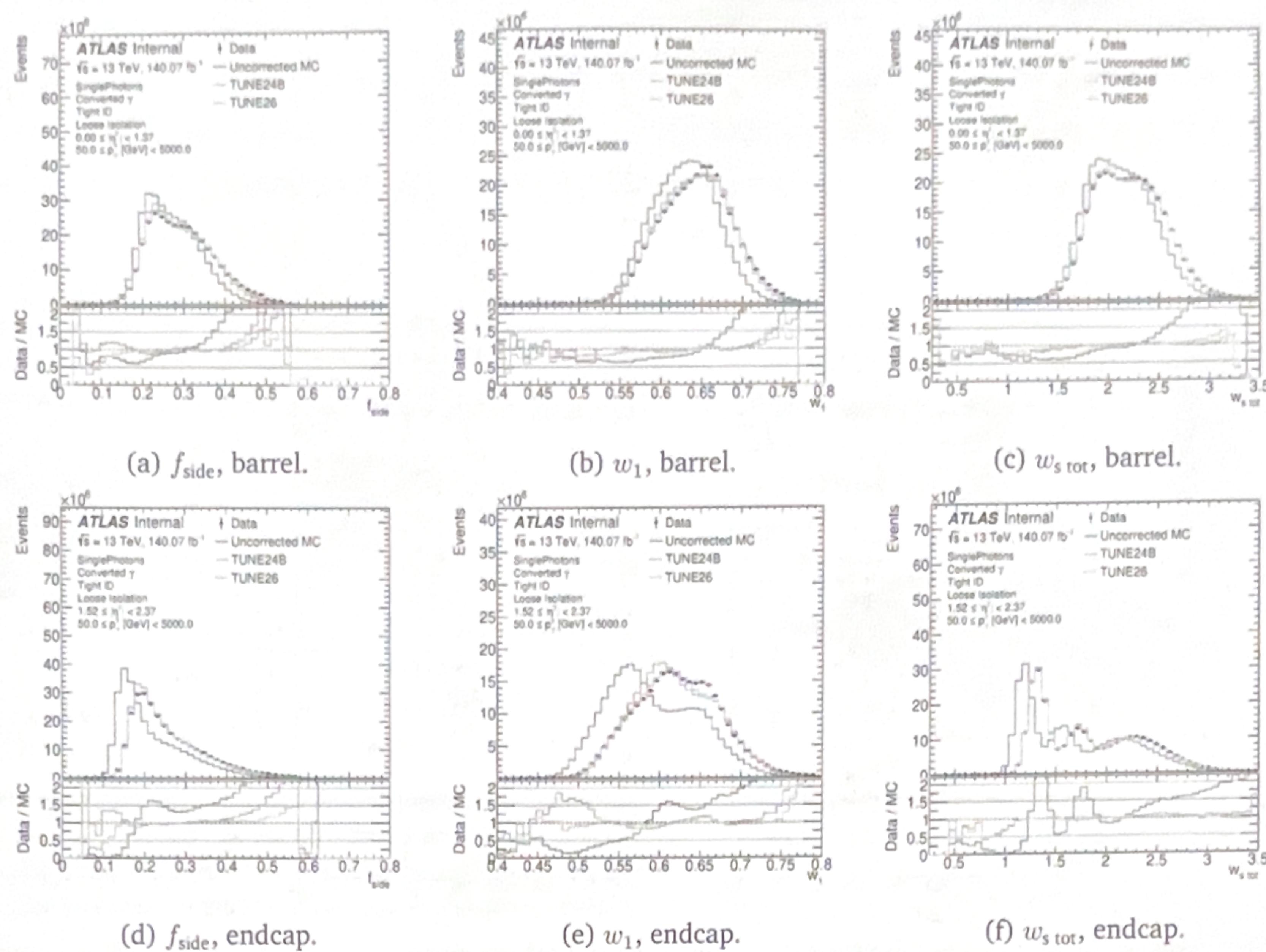


Figura 5.7: Ídem a la Figura 5.6 pero utilizando las muestras de SP.

los electrones, los resultados han sido muy prometedores, ya que se corrigieron sustancialmente las SSs de la segunda capa del calorímetro. Sin embargo, para 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.

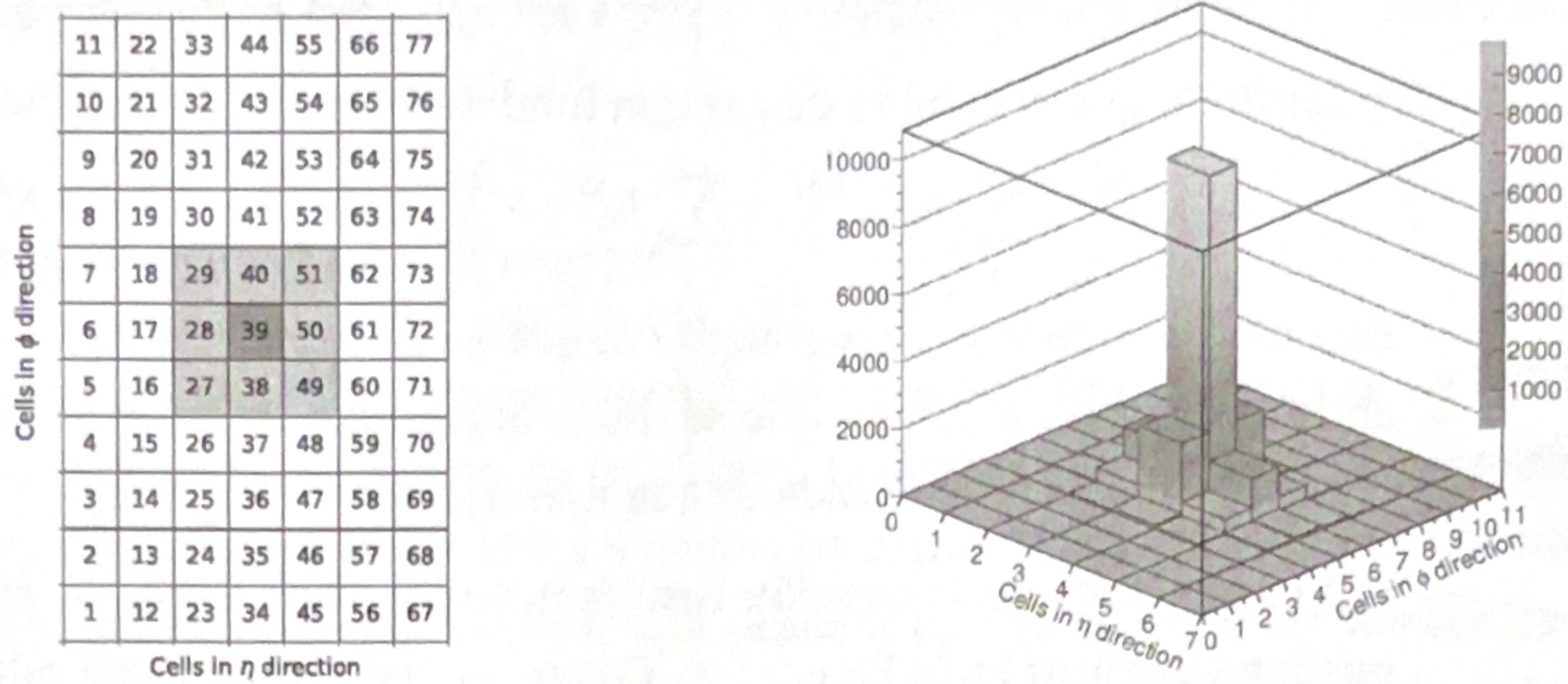
En la presente sección se estudia una nueva forma de corregir las energías de las celdas en MC. Por simplicidad, en una primera etapa, se estudió solo la segunda capa del ECAL. El método tiene similitudes con el método FF, lo que además facilita su comprensión. En primer lugar, se presenta la selección 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. Cálculo de las correcciones

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.

Los electrones o fotones que atraviesan el ECAL dejan su energía depositada en un conjunto de celdas. En este trabajo, utilizando la segunda capa del ECAL, las celdas se agrupan en clusters de 7×11 celdas en $\eta \times \phi$ y se encuentran centradas alrededor de la

celda más energética. Esta disposición se puede observar en la Figura 5.8a. 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.



(a) Disposición de las celdas, mostrando para cada una su número. La celda central corresponde a la celda número 39 resaltada en azul oscuro, mientras que las 8 celdas vecinas se muestran resaltadas en celeste.

(b) Energía promedio en cada celda.

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

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 son la proporción de la energía total depositada que 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 Δ_i en dicha celda:

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

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

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

que se traduce en desplazar la energía normalizada de la celda e_i^{MC} en una cantidad Δ_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:

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

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 (un corrimiento negativo de e_i) se remueve energía de la celda central que luego es distribuida en las circundantes.

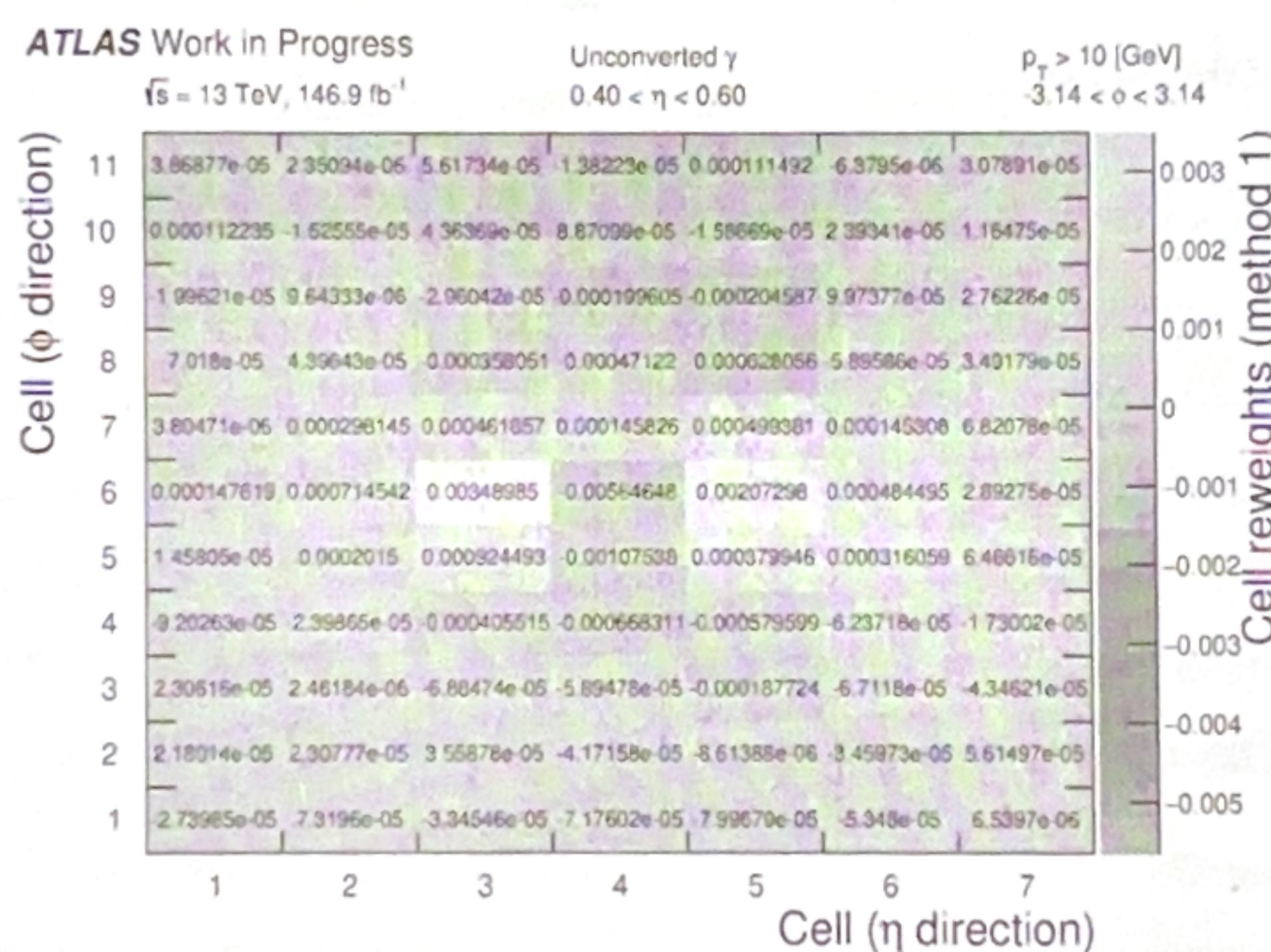


Figura 5.9: Correcciones a las energías de las celdas de la simulación MC utilizando el mismo método diseñado para electrones. * fix luminosity

A partir de las energías de las celdas, se pueden calcular las SSs de la segunda capa del ECAL, las cuales son R_η , R_ϕ 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 [149] que este método sólo corrige las formas de las variables en promedio, pero las diferencias en la forma permanecen. Esto se debe al hecho de que este método sólo corrige los valores medios de energía en las celdas. Sin 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 FFs.

Nuevo método de corrección de energías *que se incluye en esta tesis,*

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 $e_i^j = E_i^j / E^j$, y que se requiere mantener la misma energía total del cluster constante ($E^{\text{MC-RW}} = E^{\text{MC}}$), se puede multiplicar la Ec. 5.8 por $E^{\text{MC-RW}}$ y llegar a una expresión para $E_i^{\text{MC-RW}}$:

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

Por último, para garantizar que la energía del cluster permanezca constante, las energías de las celdas se normalizan 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.

5.2 Correcciones de energía de las celdas

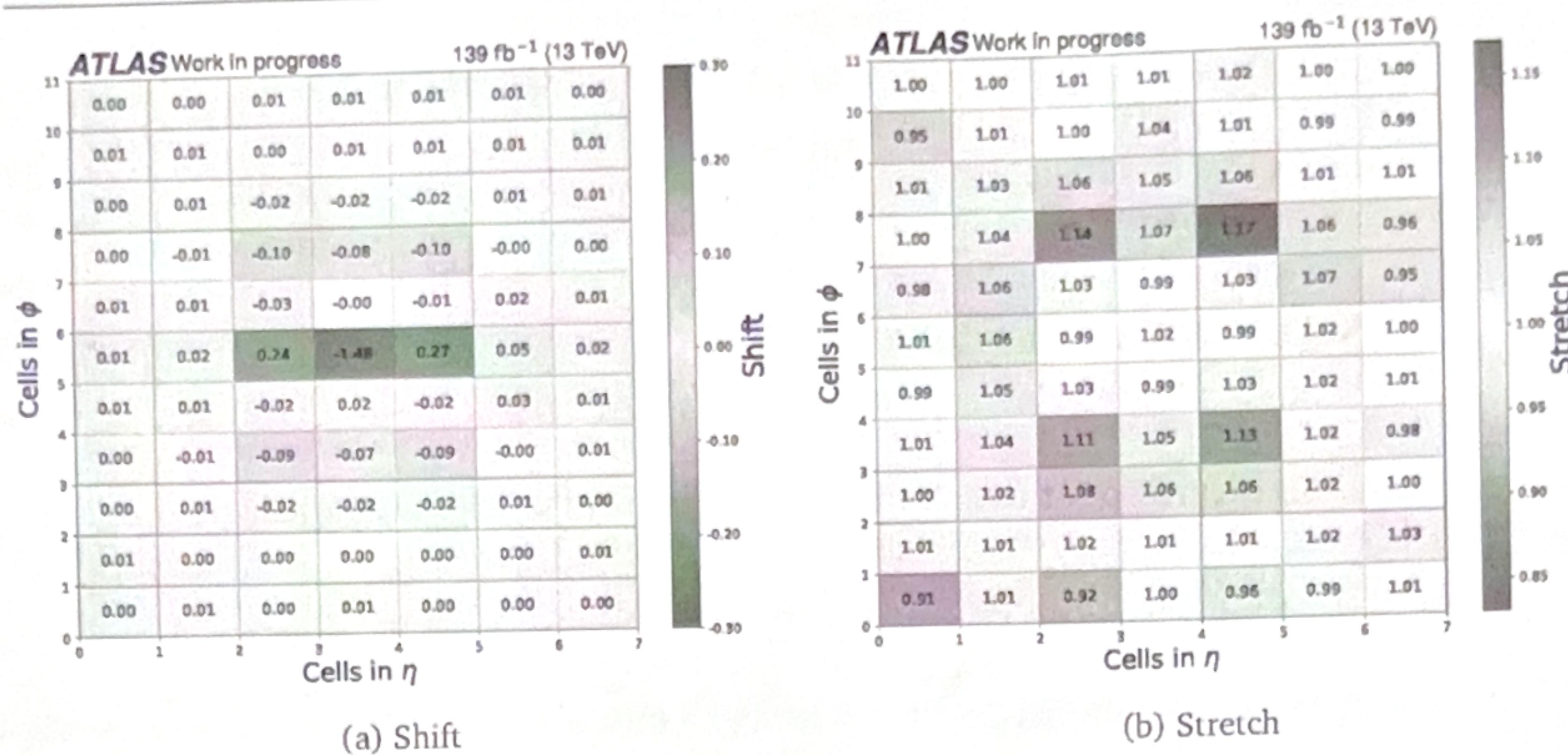


Figura 5.10: Ejemplo de las matrices de corrección de shift (izquierda) y stretch (derecha). Los valores mostrados corresponden al cálculo de las correcciones utilizando fotones no convertidos. Los valores de shift son multiplicados por un factor de 100 para mejorar su visualización.

5.2.2. 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 η 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 ϕ 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 50². 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.

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_η , R_ϕ y $w_{\eta 2}$. En los tres casos, se observa una mejora con respecto al MC sin corregir, especialmente para R_ϕ 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

² 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 η .

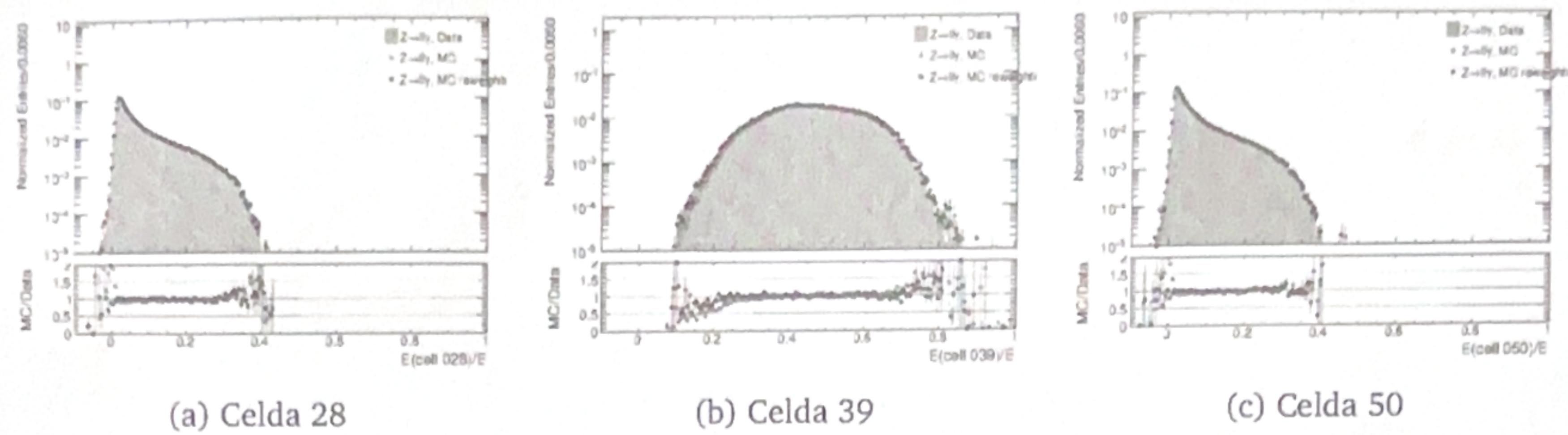


Figura 5.11: Distribuciones de las energías normalizadas de las celdas 28, 39 y 50 de cluster de 77 celdas, para fotones no convertidos. Los puntos azules y rojos corresponden a las distribuciones de la simulación MC con y sin las correcciones, respectivamente, mientras que el histograma gris representa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

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.

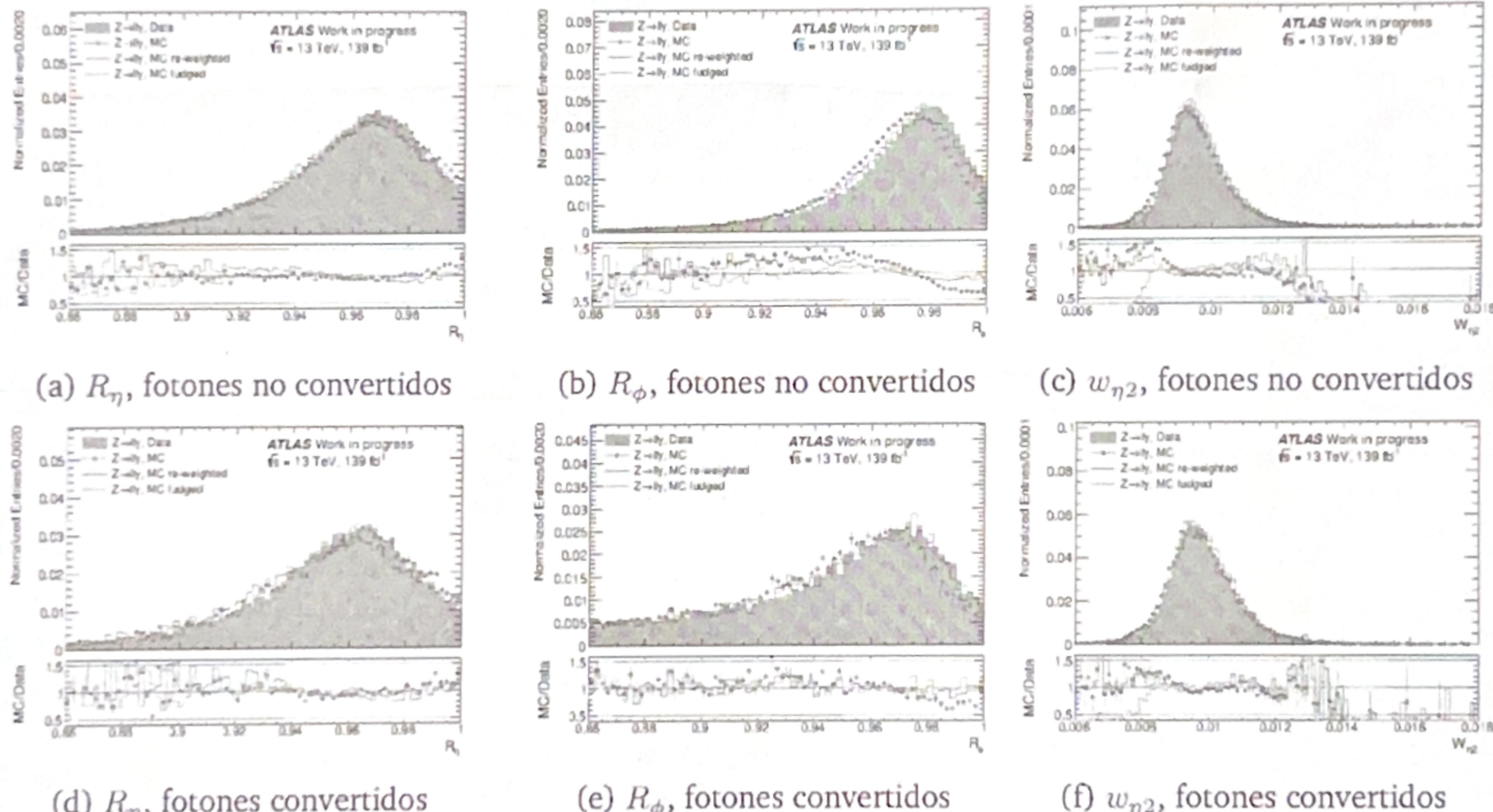


Figura 5.12: Distribuciones de las SSs calculadas en la segunda capa del ECAL para fotones no convertidos (fila superior) y convertidos (fila inferior) con pseudorapidez $|\eta| < 0.6$, comparando los diferentes métodos de corrección con los datos. Los datos experimentales están representados por los 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.

5.3. Conclusiones y trabajo futuro

En el presente capítulo se han investigado 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 de tesis, 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 de corrección novedoso y que actúa a más bajo nivel, 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 decir, utilizando la información de más bajo nivel en el proceso de reconstrucción), es posible corregir todas las SSs (variables de alto nivel en el proceso de reconstrucción) en 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.

Punto aparte

5.3.1 Trabajo a futuro

Uno de los enfoques más interesantes y prometedores para corregir las SSs es el método basado en las correcciones de las energías de las celdas. 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 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 [149].

Las 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, que se exploró en la Ref. [150], mostrando resultados muy prometedores.