

Simulación y análisis experimental de la eficacia para el proceso $p p \rightarrow Z' \rightarrow b \bar{b}$ con parámetros $\sqrt{s} = 13 \text{ TeV}$ y $M(Z') = 50 \text{ GeV}$

Sergio Duque Mejía, José D. Ruiz-Álvarez, Diego Restrepo, Andrés Rivera

*Instituto de Física, Universidad de Antioquia,
A.A. 1226, Medellín, Colombia*

E-mail: sergio.duquem@udea.edu.co, josed.ruiz@udea.edu.co,
restrepo@udea.edu.co, afelipe.rivera@udea.edu.co

ABSTRACT: En este trabajo se realiza la simulación del proceso de física de partículas: $p p \rightarrow Z' \rightarrow b \bar{b}$ a 13 TeV de energía, donde Z' se obtiene a partir del modelo de numero bariónico calibrado U(1)B implementado en un sector oscuro. La simulación utiliza los parámetros de masa $M(Z')=50 \text{ GeV}$ y el rango de valores de momento transversal menores a 25 GeV. El análisis muestra que la eficacia usando dos cortes para mejorar la adquisición de data son $Z_1 = 0.04391$ y $Z_2 = 0.00095$. Estos valores no representan ninguno descubrimiento físico, pero exponen una motivación para mejorar el modelo teórico y futuros análisis en rangos pequeños de momento transversal.

Contents

1	Introducción	1
2	Simulación de datos y caracterización de muestras:	2
2.1	Simulación de Monte Carlo (MC)	2
2.2	Simulación de muestras	3
2.3	Datos generados para el análisis	4
2.4	Selección de eventos	4
3	Resultados y discusión:	4
4	Conclusiones	9
5	Referencias	10

1 Introducción

El modelo estándar explica con bastante precisión los principios y teorías que gobierna la física hasta la fecha. Aunque como toda teoría, esta requiere ser reajustada debido a fenómenos complejos que desafían la implementación de métodos conocidos para su apropiada investigación.

En la actualidad, se han realizado varias propuestas para estudiar los principales fenómenos que se presentan problemáticos para el modelo estándar, un método que busca solucionar estos enigmas es el modelo de numero barionico calibrado con simetría Abelian $U(1)_B$ implementado en un sector oscuro que además de contar con un rango de masas y acopladores de calibre extenso, permite explicar fenomenológicamente las incógnitas suscitadas en el modelo estándar, que entre tantas, la de mayor importancia en este artículo resulta ser la existencia de la materia oscura en el universo. El modelo presentado anteriormente, hace una medida de una partícula X ocurrida en eventos durante el Big Bang, esta sigue la teoría del mecanismo estándar de las (Weakly Interacting Massive Particles) WIMP, indicando que X se encontraba en equilibrio térmico con el plasma primordial y que a medida que el universo comenzó enfriarse a temperaturas por debajo de la masa de la materia oscura la razón de cambio de la aniquilación de la materia oscura fue superada por la expansión del universo y por lo tanto la densidad de este material fue congelada. Para ese modelo se considera el fermión de Dirac $X = (X_L, X_R)$ como el candidato para materia oscura [1].



Una propuesta eficiente para examinar la validez y los limites del modelo estándar es el uso del LHC ubicado en el CERN en sus experimentos CMS y ATLAS, los cuales tienen como objetivo entre tantos: la búsqueda de materia oscura. En el LHC se aceleran haces de hadrones a altas velocidades para posteriormente colisionar estos generando partículas con propiedades físicas que son registradas por los distintos tipos de medidores específicos para sus tareas. Una de las componentes del experimento CMS significativas para las medidas que se buscan simular en este artículo es el calorímetro de hadrones (HCAL) el cual tiene el importante papel de la identificación y medida de quarks, gluones y neutrinos, esto midiendo la energía y dirección de jets y el flujo de energía transversal perdida en eventos [2,3].

Este artículo busca implementar el modelo $U(1)_B$ para obtener un boson Z' con una masa específica dentro de un rango de momento transversal bajo, es decir en un límite complicado para el estudio de física de altas energías [4,5,6,7,8,9,10], el cual se comprende entre 10 y 25 GeV en el decaimiento a b-jets, por lo que se investiga el caso de una masa $M(Z') = 50$ GeV. La principal motivación se encuentra en realizar un análisis que complemente el estudio de este modelo, además de generar un set de datos similar a los experimentos realizados en el CMS/ATLAS del LHC utilizando herramientas de simulación con base a algoritmos de Montecarlo que permita realizar un análisis de la eficacia de cortes en la generación del proceso $pp > Z' > b\bar{b}$ con luminosidad $l = 3000 \text{ fb}^{-1}$ y energía $\sqrt{s} = 13$ TeV.

2 Simulación de datos y caracterización de muestras:

2.1 Simulación de Monte Carlo (MC)

En la física es crucial realizar simulación de experimentos que presentan una gran complejidad respecto a la interacción de varias partículas en un sistema definido. Para este tipo de simulación se utiliza comúnmente algoritmos computacionales basados en conceptos de física mecánica y cuántica estadística que cumplen con el propósito de comprender la dinámica del proceso físico realizando muestreo aleatorio y repetitivo de los datos con el objetivo de obtener una solución numérica para solventar los problemas que de forma analítica serían complejos o imposibles de determinar [11]. Siendo específicos, en física de partículas, algoritmos basados en el método de Montecarlo (MC) son implementados para simular y obtener información sobre las interacciones entre partículas y estas con los detectores, esto generando eventos que proveen posibles predicciones de distribuciones de partículas basadas en presunciones teóricas comparables con los experimentos observados en el CMS y ATLAS.

Con este propósito se utiliza una herramienta llamada MADGraph, este es un framework que provee todos los elementos necesarios para realizar un estudio fenomenológico del modelo estándar (SM) y modelos por fuera del modelo estándar (BSM) como generación de eventos, obtención de secciones de corte y en general implementa herramientas para manipulación y análisis de eventos comparables con observaciones en física de partículas. MADGraph es utilizado principalmente por que sus procesos pueden simular con precisión

de **primer orden** (LO) eventos basados en el lagrangiano especificado, además puede simular eventos con precisión junto a primer orden (NLO) para corrección de procesos que involucran cromodinámica cuántica QCD. La principal diferencia de esta generación de eventos es que generando eventos NLO se permite aumentar la precisión disminuyendo las dependencias de la re normalización y escala de factorización. Además es necesario utilizar dos herramientas que van incorporadas al framework de MADGraph pero son opcionales dependiendo del análisis: Se utiliza **PYTHIA8** que similarmente se utiliza para generación de eventos LO y provee cálculos respecto a eventos de dispersión de dos cuerpos además de la hadronización del estado final de las partículas, además se utiliza **DELPHES** que sirve para reconstruir las partículas y obtener las eficiencias del detector [12].

2.2 Simulación de muestras

Con el objetivo de obtención de la eficacia o significancia de la simulación se realizo un proceso de MC de la señal con base en colisiones proton-proton a $\sqrt{s} = 13 \text{ TeV}$ utilizando el modelo de numero barionico calibrado implementado a un sector oscuro con simetría Abeliana con el objetivo de producir jets etiquetados como bottom en un rango de energía transversal al rededor de los 25 GeV. Se realizaron 50 mil eventos para la señal con una masa de 50 GeV, además de que se modifico las formulas de eficiencia en deteccion de jets etiquetados como bottom en los parámetros de Delphes para poder trabajar en este rango. por lo que se utilizo una formula lineal $0.01 + 0.000038 * P_t$ para aproximar de forma segura la producción de eventos para pT entre los rangos de 10 - 25 GeV. Además se considero dos tipos de procesos de background, uno asociado a todos los procesos físicos del SM en la generación de b-jets producto de un boson Z en el LHC y otro usando el modelo $U(1)_B$ para determinar los procesos que fallan en las medidas debido a incertidumbres en el detector o en general procesos que pueden generar eventos no requeridos por el análisis como lo son los procesos QCD.

El proceso de generación de la señal y los backgrounds fueron producidos en MADGraph v3.1.1, además de que se incluyo PYTHIA8 para el proceso de hadronizacion y para simular el detector se uso DELPHES v3.4.2. La siguiente tabla muestra los resultados de la generación de eventos que se utilizaran para obtener la eficacia de la simulación.

Tipo de muestra	Numero de eventos	Sección transversal [pb]
Señal	50000	6.1739
Background SM	1000000	2.84×10^6
Background señal	848568	6.17×10^3

Table 1. Simulaciones generadas en Madgraph

Como se muestra en la tabla 1, los resultados de la generación de muestras o eventos indican un valor específico para la sección transversal, este nos permite obtener junto con la luminosidad la eficacia necesaria para definir si dado estos parámetros es posible deter-

minar estadísticamente la viabilidad de la señal generada y en términos fenomenológicos nos permite excluir o profundizar en este rango específico de momento transversal bajo.

2.3 Datos generados para el análisis

Los resultados producto de la simulación están estructurados en el formato ROOT, este es una estructura de datos de tipo árbol con ramas y hojas utilizado en proyectos asociados a física de altas energías. Cada entrada en las tablas generadas corresponden a un evento de colisión y estas pueden contener diferentes números de jets por lo que en estructura cada lista puede contener dimensiones irregulares entre eventos y por tanto se implementa un análisis para este estudio basado en dos librerías de Python: Awkward que permite la manipulación de eventos con dimensiones irregulares y Coffea (Column Object Framework For Effective Analysis) para acceder a las ramas que se desea estudiar de los eventos generados [13]. Además se utilizó las librerías comúnmente utilizadas en Python para análisis de datos tabulares como lo son Pandas, Numpy y Matplotlib para representar gráficamente los resultados. A continuación se muestra en la tabla 2 el tipo de característica y su descripción junto con el tipo de dato necesario para el análisis.

2.4 Selección de eventos

Con el propósito de obtener un valor de la significancia de la producción de eventos dadas las condiciones específicas en el experimento que se desea simular, se considero obtener del set de datos anteriormente mencionado, los jets primarios y secundarios basado en la columna de etiqueta bottom donde el mayor valor para pT indica la clasificación los b-jets, luego se realizó los cortes expuestos en la tabla 3.

La eficiencia del corte se define como la razón entre el número de datos después del corte contra el número de datos antes de los cortes realizados. Este método se utiliza para ver la calidad del resultado posterior al corte con respecto a la pérdida de datos del proceso, por lo que es necesario aplicar este corte de igual forma tanto a la señal como los backgrounds. Además es necesario aplicar los criterios por tandas contiguas siendo el primer corte $pT(b) > 5 \text{ GeV}$ y $\eta(b) < 2.5$, el segundo corte se realiza considerando que $\Delta R(b1, b2) < 3.5$ y el tercer corte respecto a las masas del leading y subleading jet, los cuales requieren que $m(b1), m(b2) < 50 \text{ GeV}$.

3 Resultados y discusión:

Este análisis se enfoca en una selección de masa de $M(Z') = 50 \text{ GeV}$, este valor se selecciono específicamente dado que en colisionadores y experimentos como el CMS-ATLAS se caracterizan por definir su análisis en rangos estables para la detección por encima de $M = 90 \text{ GeV}$ [14], por lo que respecto al valor seleccionado se espera obtener una señal menos definida, por lo que es necesario evaluar el problema en el apartado de técnicas nuevas las cuales en su mayoría intervienen procesos con inteligencia artificial, para este

Nombre de la característica	Característica	Tipo de dato
jet_pt0	Primer jet pT [GeV]	float
jet_pt1	Segundo jet pT [GeV]	float
jet_pt2	Tercer jet pT [GeV]	float
jet_pt3	Cuarto jet pT [GeV]	float
jet_eta0	Primer jet η	float
jet_eta1	Segundo jet η	float
jet_eta2	Tercer jet η	float
jet_eta3	Cuarto jet η	float
jet_phi0	Primer jet ϕ	float
jet_phi1	Segundo jet ϕ	float
jet_phi2	Tercer jet ϕ	float
jet_phi3	Cuarto jet ϕ	float
jet_mass0	Masa primer jet [GeV]	float
jet_mass1	Masa Segundo jet [GeV]	float
jet_mass2	Masa Tercer jet [GeV]	float
jet_mass3	Masa Cuarto jet [GeV]	float
jet_btag0	Etiqueta del primer jet	int
jet_btag1	Etiqueta del Segundo jet	int
jet_btag2	Etiqueta del Tercer jet	int
jet_btag3	Etiqueta del Cuarto jet	int
missinget_met	P_T^{miss}	float
missinget_phi	MET ϕ	float

Table 2. Características obtenidas en la producción de eventos de Madgraph que serán utilizadas para realizar el análisis

Criterios base	Rango
Numero de Jets por evento	2
pT(b)	> 5 GeV
$ \eta(b) $	< 2.5
$\Delta R(b1, b2)$	< 3.5
m(b1),m(b2)	< 50 GeV

Table 3. Criterio de cortes base para inicializar el análisis de eficacia.

trabajo se busca evaluar el problema de forma tradicional definiendo una formula lineal de eficiencia de la simulación para identificar b-jets usando el modelo $U(1)_B$ en el rango de momento transversal entre 10 y 25 GeV.

Como se sabe, para el proceso descrito en el articulo, dado el valor de la masa, se verifica que el momento transversal resultado para los b-jets deben ser aproximadamente

la mitad de la masa, esto debido a que al ser un proceso simple, el producto de Z' son dos jets, bottom y anti bottom con $pT(b) \simeq 25$ GeV. Este resultado se expone en la figura 1. Que como se observa, los eventos que primero se caracterizan son los detectados en $pT(Z') = (24.86 \pm 6.15)$ GeV, posteriormente se caracteriza los resultados únicamente para el leading y subleading b-Jet de los cuales se debe precisar este hecho.

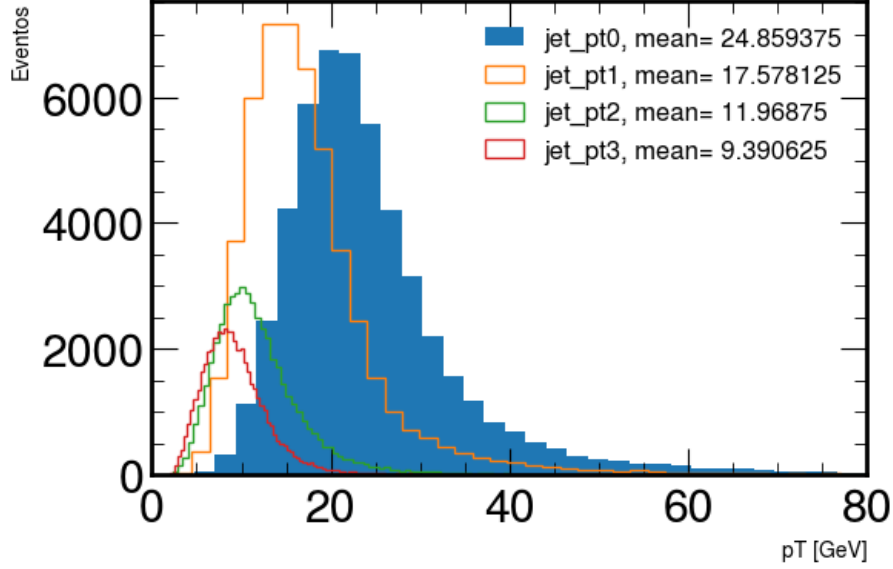


Figure 1. Momento transversal para los eventos obtenidos en la simulación de la Señal.

Además es necesario evaluar resultados importantes para realizar adecuados cortes que nos permite perder menos eventos pero al tiempo mejorar los resultados para definir adecuadamente la eficacia, estas variables son ΔR , $\Delta\phi$ y $\Delta\eta$ como se observa en las figuras 2, 3. Como se observa, se requiere realizar cortes cuando $|\eta(b)| < 2.5$ y $1 < \Delta R(b) < 3.5$.

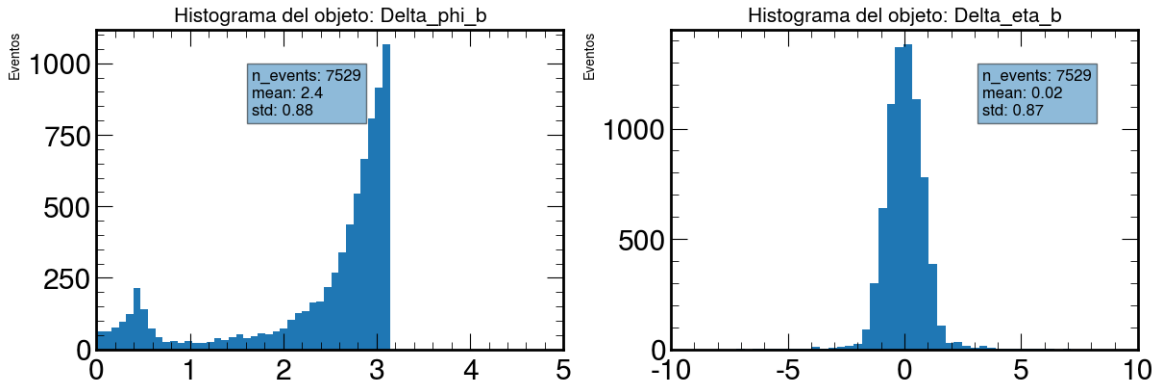


Figure 2. Histogramas de $\Delta\phi(b)$ y $\Delta\eta(b)$ que justifica el primer y segundo corte

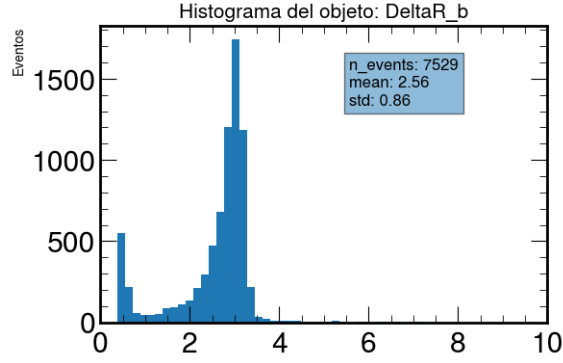


Figure 3. Histograma de $\Delta R(b)$ que justifica el tercer corte

Al distinguir los eventos iniciales bajo las condiciones impuestas en la literatura, se observa una disminución significativa de los eventos dado los cortes. esta disminución es progresiva respecto a cada corte evaluado, del cual se evidencia una mayor pérdida de datos al seleccionar únicamente eventos con dos jets, correspondientes al leading y subleading jet. Para este corte se obtuvo 7529 datos que representan un 15% de los 50000 datos iniciales, luego aplicando el primer corte con $pT(b_0, b_1) > 5$ y $|\eta(b_0, b_1)| > 2.5$ la eficiencia de corte nos deja con un 97% de los eventos con 2 Jets lo que representa un 14.55% de los datos iniciales. Se aplica el segundo corte el cual corresponde a $\Delta R(b_0, b_1) < 3.5$ y se obtuvo una eficiencia de corte de 99.1% es decir un 14.42% de los datos iniciales, finalmente del ultimo corte para $M(b_0, b_1) < 50$ GeV se obtuvo una eficiencia de corte del 86% lo que representa el 12.4% de los datos originales, explícitamente 6200 eventos cumplen con los criterios establecidos para el análisis de eficacia y estos se comparan con los resultados obtenidos de forma similar en ambos background (Modelo $U(1)_B$ y SM) de los cuales se obtuvo respectivamente 3.42% o 29017 eventos y 9.6% o 96000 eventos.

Se observa que los cortes que adquieren una mayor relevancia respecto a la pérdida de datos para las tres simulaciones son la distinción de los eventos con dos jets y los eventos que se presentan por debajo de la masa seleccionada. Como es de esperarse la simulación falla a la hora de identificar eventos por debajo de $M(Z') < 90$ GeV, específicamente existe una pérdida promedio de 91.5% de los datos cuando se aplican estas condiciones, por lo que se puede re formular estos cortes para mejorar la eficacia del proceso de simulación. Los cortes que corresponden a $pT(b_0, b_1) > 5$, $|\eta(b_0, b_1)| > 2.5$ y $\Delta R(b_0, b_1) < 3.5$ no representan una pérdida significativa para los datos originales, siendo únicamente 1.45% la pérdida promedio de datos.

$\Delta\phi$ permite realizar una comparación entre los resultados de la señal y los background con el objetivo de realizar cortes adecuados que permite distinguir una mejora significativa de las eficacia de la señal. Se realizo para cada simulación el análisis respectivo a cada objeto $(j_0, j_1, j_2, j_3, b_0, b_1)$ de los cuales se destaca $|\Delta\phi(j_0, j_1)|$ y $|\Delta\phi(b_0, b_1)|$ que como se observa en las figuras 4, 5, 6, es posible realizar los cortes $|\Delta\phi(j_0, j_1)| > 1$ y $|\Delta\phi(b_0, b_1)| > 1$



donde j_0 , j_1 , b_0 y b_1 no necesariamente representan el mismo jet. En la señal se obtiene las eficiencias de corte de 92.3% y 76.9% respectivamente. Para el primer background se obtuvo 92.6% y 53.1%, para el segundo background se obtuvo 92.7% y 75.9% respectivamente para cada corte. Esto nos indica que para la obtención de la eficacia Z de la señal se obtendrá un mejor resultado usando el primer corte, esto debido a que no es significativa la perdida de datos usando este criterio de mejora en resultados.

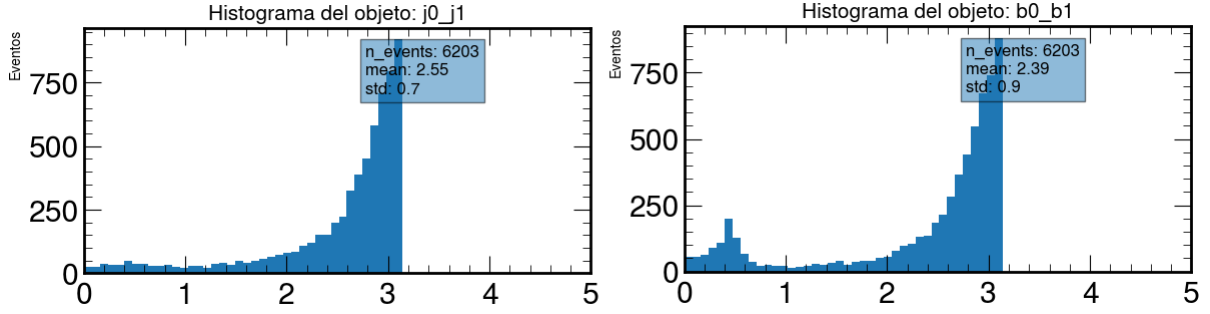


Figure 4. Histogramas para $\Delta\phi$ en la Señal, cuando se analiza j_0 , j_1 , b_0 , b_1 para determinar el mejor corte como criterio de mejora del resultado.

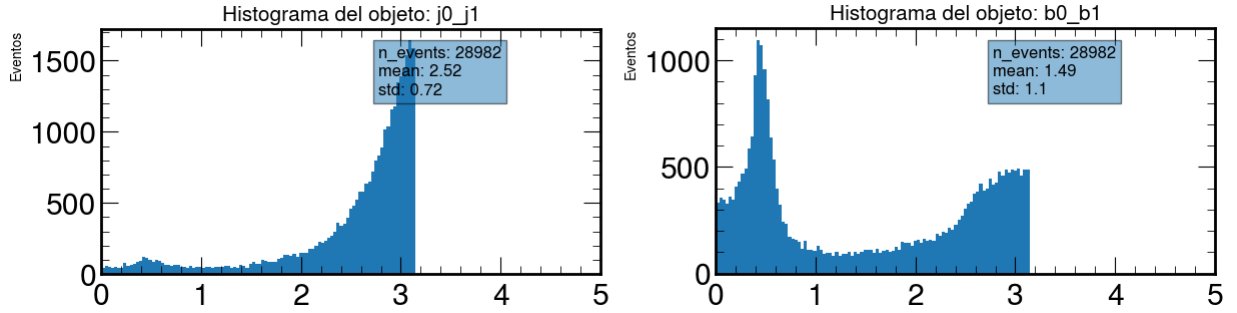


Figure 5. Histogramas para $\Delta\phi$ en el background para procesos $Z > b b \sim$, cuando se analiza j_0 , j_1 , b_0 , b_1 para determinar el mejor corte como criterio de mejora del resultado.

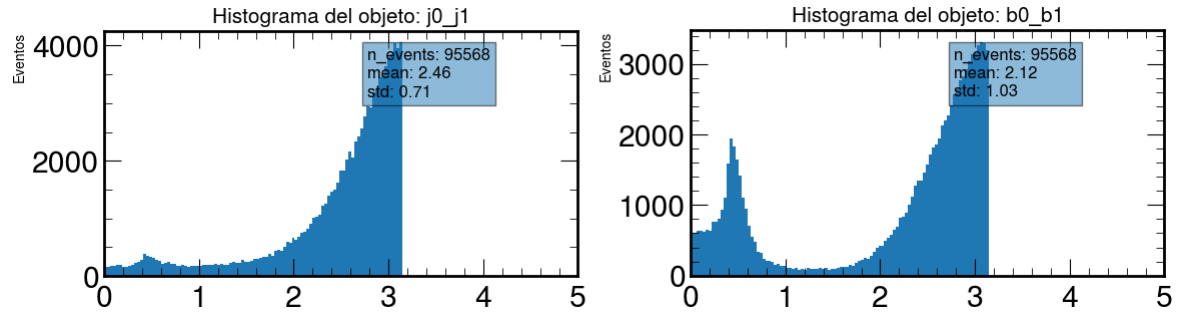


Figure 6. Histogramas para $\Delta\phi$ en el background para procesos QCD, cuando se analiza j_0 , j_1 , b_0 , b_1 para determinar el mejor corte como criterio de mejora del resultado.

La eficacia Z se define como: $Z = \frac{S}{\sqrt{S+B}}$ donde S es la señal, B es la suma de los backgrounds con procesos $Z > b\bar{b}$ y QCD. Esta ecuación determina estadísticamente si la señal simulada tiene un grado de relevancia para posteriores estudios, además indica que tan seguro estamos de que se realizaron los cortes adecuados para el proceso de investigación. Se puede determinar por medio de esta dos puntos de referencia estadísticos para física de partículas, cuando $Z \geq 2$ o $Z \geq 3$ responde a exclusión del resultado. Cuando $Z > 5$ responde a un descubrimiento en física de partículas. Para determinar este valor es necesario contar con el numero de eventos posterior a la simulación sin cortes y un valor de peso que se define como $w = \frac{L\sigma}{N_c}$ donde L es la luminosidad, σ es el valor de cross section que podemos extraer de la simulación de MC en el banner de resultados y N_c son el numero de eventos para la simulación posterior a los cortes, incluyendo el criterio de mejora definiendo $|\Delta\phi|$, luego $S = N_s w_s$, $B_1 = N_{B_1} w_{B_1}$, $B_2 = N_{B_2} w_{B_2}$ y $B = B_1 + B_2$. Usando las variables encontradas para el modelo final, se observo que para la señal con $|\Delta\phi(j_0, j_1)| > 1$ se obtuvo $Z_1 = 0.04391$ y cuando $|\Delta\phi(b_0, b_1)| > 1$ se obtuvo $Z_2 = 0.00095$. Estos resultados son esperados dado que los cortes en la señal generan una diferencia significativa respecto a los background, además se observa que el modelo teórico responde adecuadamente a valores para la masa de Z' mayores a 90 GeV, por lo que la simulación discrimina una mayor cantidad de resultados que se observan en la perdida de eventos cuando se busca los que cumplen 2 Jets.

4 Conclusiones

- Se logro comprender el funcionamiento y generación de eventos utilizando MadGraph para producción de resultados efectivos en el análisis de procesos de física de partículas.
- Se logro estudiar a fondo las variables generadas por la simulación, haciendo uso de herramientas como ROOT junto con sus librerías integradas a Python y C.
- Se estudio de forma fenomenológica el esquema de producción de eventos en experimentos reales en el LHC.
- Se obtuvo diferentes tipos de simulaciones donde se logra destacar las llamadas Señal y dos background correspondientes a los procesos que generan $Z > b\bar{b}$ y QCD. Además se exploro cada variable determinante para el análisis posterior.
- Se logro determinar los criterios adecuados para una perdida mínima de eventos y cuales fueron determinantes para una perdida significativa de datos, siendo la selección de 2 jets por eventos y eventos con masa inferior a 50 GeV los que se deben replantear para mejorar el análisis.
- Se obtuvo dos eficacias respecto al análisis de corte usando el criterio de $|\Delta\phi|$, de los cuales dado la perdida de eventos para el segundo se observo que el criterio para los b_0 y b_1 no son los adecuados.

- Se obtuvo como mejor resultado una eficacia $Z_1 = 0.04391$, este valor estadísticamente no representa un estudio en profundidad para la física de partículas, pero plantea la necesidad de explorar masas en el rango de 50 GeV a 90 GeV, para mejorar el resultado.
- Se espera obtener mejores resultados realizando el análisis con un modelo $U(1)_B$ que permita el estudio de masa inferior a 90 GeV.

5 Referencias

- 1 Restrepo, D., Rivera, A., Tangarife, W. (2022). Dirac dark matter, neutrino masses and dark baryogenesis. *Phys. Rev D*, 106(055021), 22.
- 2 Duque-Escobar, S., Ocampo-Henao, D., Ruiz-Alvarez, J. (2022). Vector boson fusion topology and simplified models for dark matter searches at colliders. *High Energy Physics*, 40.
- 3 Barney, D., CMS Outreach Coordinator. (2003). An overview of the CMS experiment for CERN guides. 30.
- 4 Jafari, A., D'Hondt, J., Ardalan, F. (2011). Measurement of the b-tagging efficiency in the CMS experiment with first LHC collisions. *Sharif University of Technology*, 252.
- 5 ATLAS Collaboration. (2016). Performance of b-jets identification in the ATLAS experiment. *JINST*, 11(P04008), 111.
- 6 CMS Collaboration. (2019). Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at 13 TeV. *Physical Review D*, 99(012005), 39.
- 7 CMS Collaboration. (2019b). Search for low-mass vector resonance decaying into quark-antiquark pair in proton-proton collisions at 13 TeV. *Physical Review D*, 100(112007), 39.
- 8 CMS Collaboration. (2021). Search for top quark production in fully-hadronic final states in proton-proton collision at 13 TeV. *Physical Review D*, 104(052001), 70.
- 9 CMS Collaboration. (2017). Search for direct production of supersymmetric partners of the top quark in the all jets final state in proton-proton collisions at 13 TeV. *Physical Review D*, 10(1007), 58.
- 10 D0 Collaboration. (2010). b-jet identification in the D0 experiment. *Nucl Instrum Meth*, A620(490-517), 42.
- 11 Srimanobhas, N. (2010). Introduction to Monte Carlo for particle physics study. *Cern School Thailand*, 33.

- 12 Ruiz-Alvarez, J. (2022). A brief intro to MadGraph, Pythia, Delphes and MadAnalysis. Uniandes particle physics school, 24.
- 13 ROOT: analyzing petabytes of data, scientifically. (s.f.). ROOT. <https://root.cern/>
- 14 CMS Collaboration. (2018). Search for beyond the standard model Higgs boson decaying into bb pair in proton-proton collisions at 13 TeV. SOSSA, 08(113), 43.