**Logotipo

Descripción generada automáticamente**

**Estimación cuantitativa de precipitación mediante datos del radar meteorológico y algoritmos de aprendizaje automático.**

Catherine Balbuena Franco

Juan Manuel Sánchez Ortega

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
Nombres completos, Título académico más alto

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

|  |  |
| --- | --- |
| **Cita** | (Balbuena Franco & Sánchez Ortega, 2024) |
| **Referencia**  **Estilo APA 7 (2020)** | Balbuena Franco, C., & Sánchez Ortega, J. M. (2024). *Estimación cuantitativa de precipitación mediante datos del radar meteorológico y algoritmos de aprendizaje automático.* Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteVIII.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Dedicatoria**

Texto de dedicatoria centrado.

**Agradecimientos**

Texto de agradecimientos centrado.

Tabla de Contenido

[Resumen 9](#_Toc183115714)

[Abstract 10](#_Toc183115715)

[1. Descripción del problema 11](#_Toc183115716)

[1.1. Problema de Negocio 11](#_Toc183115717)

[1.2. Aproximación desde la analítica de datos 12](#_Toc183115718)

[1.3. Origen de los datos 13](#_Toc183115719)

[1.4. Métricas de desempeño 13](#_Toc183115720)

[2. Objetivos 16](#_Toc183115721)

[2.1. Objetivo general 16](#_Toc183115722)

[2.2. Objetivos específicos 16](#_Toc183115723)

[3. Datos 17](#_Toc183115724)

[3.1. Datos originales 17](#_Toc183115725)

[3.2. Datasets 18](#_Toc183115726)

[3.3. Analítica descriptiva 20](#_Toc183115727)

[Referencias 25](#_Toc183115728)

**Lista de tablas**

[**Tabla 1.**  *Variables de dataset.* 16](#_Toc182937267)

**Lista de figuras**

[**Figura 1**. Origen del Dataset. 17](#_Toc183115114)

[F**igura 2.** Clasificación de Eventos de Precipitación por Intensidad 19](#_Toc183115115)

[**Figura 3.** Serie de tiempo de precipitación diaria de la estación “Torre SIATA” (línea azul) y precipitación diaria calculada con información del radar usando métodos matemáticos (línea marrón). 20](#_Toc183115116)

[**Figura 4.** Histogramas para: a) Precipitación acumulada en 5 minutos; b) Intensidad de Precipitación; c) Reflectividad del Radar. Las líneas puntadas sugieren el percentil 90 y 99 respectivamente. 21](#_Toc183115117)

[**Figura 5**. Diagrama de dispersión para: a) intensidad de precipitación y la reflectividad del píxel del radar correspondiente a la estación pluviométrica; b) intensidad de precipitación y la reflectividad de los nueve pixeles que rodean la estación pluviométrica. 21](#_Toc183115118)

[**Figura 6.** Histogramas para: a) Eventos considerados Débiles (Intensidad < Percentil 90); b) Eventos considerados Moderados (Percentil 90 < Intensidad < Percentil 90); c) Eventos considerados Débiles (Percentil 99 < Intensidad). 22](#_Toc183115119)

**Lista de ecuaciones**

[**Ecuación 1.** Error cuadrático medio (RMSE). 14](#_Toc183115791)

[**Ecuación 2**. Error absoluto medio (MAE) 14](#_Toc183115792)

[**Ecuación 3**. Diferencia relativa media (MRD) 15](#_Toc183115793)

[**Ecuación 4**. Intensidad de precipitación. 18](#_Toc183115794)

**Siglas, acrónimos y abreviaturas**

**CAPPI** Constant Altitude Plan Position Indicator

**QPE** Estimación cuantitativa de precipitación, por sus siglas en inglés

**SIATA** Sistema de Alertas Tempranas de Medellín y el Valle de Aburrá

# Resumen

El Sistema de Alertas Tempranas de Medellín y el Valle de Aburrá (SIATA) opera un radar meteorológico que permite monitorear la lluvia en gran parte del departamento de Antioquia. Sin embargo, este tiene asociado una gran incertidumbre en la estimación cuantitativa de precipitación (QPE) la cual debe ser reducida para mejorar así la gestión del riesgo y desastres relacionados a eventos de lluvia. Para mejorar dicha estimación se propone el uso de modelos de machine learning que permitan capturar las características no lineales de la lluvia observadas por las variables del radar meteorológico y así mejorar dicha QPE. Dichas variables comprenden Reflectividad horizontal, CAPPI, velocidad radial horizontal, Coeficiente de correlación Copolar, y Reflectividad diferencial, entre otras. Además, como variable a predecir se calculó la intensidad de precipitación en una ventana de tiempo de 15 minutos. Una de las dificultades al trabajar con datos de intensidad de lluvia es el fuerte sesgo positivo en la magnitud de dicha variable, lo que lleva a un caso de datos desbalanceados, para lo cual se dividió el dataset en otros tres subconjuntos: Eventos débiles, moderados y fuertes. A partir de estos nuevos conjuntos se entrenaron diferentes modelos como Random forest, AdaBoost y Gradient Boosting. Si bien los modelos logran capturar brevemente el comportamiento de la lluvia, aún se observa un fuerte desfase entre la magnitud de la intensidad de precipitación calculada por los métodos de machine learning y lo registrado por las estaciones pluviométricas.

https://github.com/C-Balbuena/Seminario-UdeA/tree/main

*Palabras clave*: QPE, Precipitación, Radar meteorológico, Pluviómetro

# Abstract

The Early Warning System of Medellín and the Aburrá Valley (SIATA) operates a weather radar that monitors rainfall in a large part of the department of Antioquia. However, it is associated with a large uncertainty in the quantitative precipitation estimate (QPE), which should be reduced in order to improve risk and disaster management related to rainfall events. To improve this estimation, the use of machine learning models is proposed to capture the nonlinear characteristics of rainfall observed by the weather radar variables and thus improve the QPE. These variables include horizontal reflectivity, CAPPI, horizontal radial velocity, Copolar correlation coefficient, and differential reflectivity, among others. In addition, the precipitation intensity in a 15-minute time window was calculated as a variable to be predicted. One of the difficulties when working with rainfall intensity data is the strong positive bias in the magnitude of that variable, which leads to a case of unbalanced data, for which the dataset was divided into three other subsets: Weak, Moderate and Strong events. From these new sets, different models such as Random forest, AdaBoost and Gradient Boosting were trained. Although the models manage to capture briefly the rainfall behavior, there is still a strong mismatch between the magnitude of the rainfall intensity calculated by the mahcine learning methods and that recorded by the rainfall stations.

*Keywords***:** QPE, Precipitation, Weather Radar, Rain gauge

# Descripción del problema

Los radares meteorológicos son esenciales para monitorear la lluvia en grandes áreas. En nuestra región, el Sistema de Alertas Tempranas de Medellín y el Valle de Aburrá (SIATA) opera uno de estos radares que permite monitorear el 97% de Antioquia. Sin embargo, la incertidumbre (alrededor del 30%) en la estimación de la precipitación por parte de este sensor debe ser reducida para mejorar así la gestión del riesgo y desastres relacionados a eventos de lluvia (Sepúlveda Berrío, 2015).

Para tal fin, se analizan datos del radar meteorológico tales como reflectividad diferencial, CAPPI, fase diferencial, entre otras variables polares; además se cuenta con información histórica de las estaciones pluviométricas. Esta información es entregada por SIATA.

Con estos datos se integran modelos avanzados de Machine Learning para mejorar la estimación de la lluvia para una gestión del riesgo y desastres más oportuna. Entre las métricas de desempeño se incluye la precisión de los modelos para estimar la intensidad de la lluvia, así como la incertidumbre en dicha estimación.

## Problema de Negocio

El entendimiento de la variabilidad espacial y temporal de la precipitación ha sido suscitado a lo largo del tiempo, ya que esta tiene implicaciones directas en actividades productivas en diferentes sectores económicos, salud y en las condiciones de riesgo por eventos extremos que pueden afectar los asentamientos humanos (Sloat, y otros, 2018).

El SIATA ha hecho un gran esfuerzo por construir una amplia red de sensores pluviómetros que permita el monitoreo de la precipitación in-situ, cuantificando detalladamente la lluvia que llega a superficie en un punto específico, pero careciendo del entendimiento de la variabilidad espacial de la lluvia. El radar meteorológico, por el contrario, es capaz de brindar información espacial de la precipitación, haciendo de este una herramienta de alto valor en el estudio de la dinámica de los sistemas que generan precipitación en una región.

Si bien la estimación de la precipitación a partir de información del radar usando métodos matemáticos ha mostrado una buena aproximación, se hace crucial disminuir la incertidumbre de esta estimación que alcanza hasta un 30% (Sepúlveda Berrío, 2015). Estas limitaciones están asociadas al principio de medición de la precipitación, pues la lluvia no es medida de manera directa por el radar meteorológico, sino que corresponden a un proxy de la cantidad de gotas que tiene una nube, lo cual se convierte en un indicador de la intensidad de lluvia que llega a superficie. Así pues, es necesario implementar modelos de machine learning que capturen las características específicas de la lluvia y que permitan una correcta estimación de la intensidad y cantidad de esta, usando de forma conjunta las mediciones del radar meteorológico y de estaciones pluviométricas (Sepúlveda Berrío, 2015).

## Aproximación desde la analítica de datos

Actualmente existen técnicas matemáticas que relacionan la intensidad de la precipitación (R) con la reflectividad proveniente del radar meteorológico, conocidas como relaciones Z-R, que buscan estimar la intensidad de la lluvia a partir de la reflectividad. Esta relación ha sido ampliamente utilizada dada su gran aplicabilidad, sin embargo, se encuentra sujeta a limitaciones. Por ejemplo, la variabilidad en la distribución del tamaño de las gotas de lluvia, la atenuación de la señal del radar por factores orográficos o características de la atmósfera, retos en la calibración del radar meteorológico, en otras, son fuentes de incertidumbre que generan retos en una correcta estimación de la precipitación (Hong & Gourley, 2017). Ya que la relación Z-R se basa en correlaciones empíricas que pueden variar geográfica y temporalmente resaltan la necesidad de metodologías que se adapten a condiciones meteorológicas específicas.

Basados en lo anterior, para el desarrollo de este proyecto se analizarán diferentes modelos de machine learning, entre los cuales se incluyen el Random Forest, el Gradiente Boosting, entre otros, los cuales han demostrado ser herramientas efectivas para capturar complejidades no lineales en el comportamiento de la lluvia haciendo más precisa la estimación de la precipitación (Verdelho, Pavam Jr, Calvetti, Oliveira, & Zanata Alves, 2024). Con la aplicación de estas técnicas de aprendizaje automático, se espera mejorar la estimación de la intensidad de la precipitación a partir del radar meteorológico y disminuir así la incertidumbre en dicha estimación que hay actualmente usando los métodos tradicionales (Sepúlveda Berrío, 2015).

## Origen de los datos

Este estudio cuenta con dos conjuntos de información. El primero está relacionado a los datos provenientes del radar meteorológico, que cuenta con las siguientes variables: Reflectividad horizontal, CAPPI, velocidad radial horizontal, Coeficiente de correlación Copolar, y Reflectividad diferencial, las cuales se obtiene en las coordenadas de cada estación pluviométrica, además se cuenta con la reflectividad horizontal en un área que rodea cada estación pluviométrica, el tipo de precipitación (convectiva o estratiforme), distancia del radar a la estación pluviométrica y altura de la estación pluviométrica.

Por otra parte, el segundo conjunto de datos contiene registros históricos de las estaciones pluviométricas, este tiene información del acumulado de lluvia en superficie en cada minuto. En este conjunto de datos, cada columna del dataset corresponde a la información de cada estación pluviométrica.

Ambos conjuntos de datos se obtuvieron para el periodo de tiempo comprendido entre el 2015 y el 2020. Sin embargo, en el caso de las estaciones pluviométricas, este periodo de tiempo puede variar según la fecha en que fue instalada la estación, es decir, puede haber estaciones instaladas después del año 2015.

## Métricas de desempeño

Las métricas más utilizadas en la literatura para evaluar el desempeño de modelos de aprendizaje automático en la predicción de fenómenos climatológicos, como la precipitación, son el error cuadrático medio (RMSE), el error absoluto medio (MAE) y la diferencia relativa media (MRD). Estas métricas permiten medir la precisión de las predicciones al comparar los valores estimados con los observados, proporcionando una evaluación integral de la magnitud de los errores y su impacto relativo en el modelo. Además, los errores observados pueden clasificarse en aleatorios o sistemáticos; los errores aleatorios, asociados a fluctuaciones impredecibles, se evalúan principalmente mediante el RMSE, mientras que los errores sistemáticos, que reflejan sesgos consistentes en las predicciones, son mejor evaluados mediante métricas como el MAE y el MRD (Sayeed Khan & Ehsan Bhuiyan, 2021).

A continuación, se explica a detalle cada una de las métricas mencionadas:

* + 1. **Error cuadrático medio (RMSE).**

El error cuadrático medio, matemáticamente, es la raíz del promedio de los cuadrados de los residuos, que son las diferencias entre los valores predichos por el modelo y los valores observados en los datos (Sayeed Khan & Ehsan Bhuiyan, 2021), es decir, que tan dispersos o concentrados están los residuos (Ver ***Ecuación 1***). Cuanto menor sea la ecuación, mejor será el modelo.

**Ecuación 1.** Error cuadrático medio (RMSE).

Donde n es la población, yi y son los valores verdaderos y predichos.

* + 1. **Error absoluto medio (MAE)**

Métrica que mide la precisión de un modelo al calcular el promedio de los valores absolutos de las diferencias entre los valores observados (yi) y los valores predichos , donde se representan la magnitud del error en cada predicción, sin considerar si el modelo sobreestima o subestima, ya que elimina los signos negativos al tomar el valor absoluto (Ver ***Ecuación 2***). Cuanto menor sea el valor MAE más preciso será (Sayeed Khan & Ehsan Bhuiyan, 2021).

A diferencia de otras métricas como el RMSE, el MAE no penaliza de forma más severa los errores grandes, lo que lo hace especialmente útil cuando se busca una medida simple y robusta del desempeño del modelo, sin que los errores extremos influyan de manera desproporcionada.

**Ecuación 2**. Error absoluto medio (MAE)

* + 1. **Diferencia relativa media (MRD)**

Es una métrica que evalúa la precisión de un modelo de predicción al calcular el error relativo promedio entre los valores observados yi y los valores predichos , y al multiplicarse por 100, se convierte en un valor porcentual, lo cual permite entender el desempeño del modelo en términos relativos a la magnitud de los valores observados (Ver ***Ecuación 3***). Se espera que, la MRD sea baja, que signifique que el modelo tiene un buen ajuste en términos relativos a la magnitud de los datos (Sayeed Khan & Ehsan Bhuiyan, 2021).

**Ecuación 3**. Diferencia relativa media (MRD)

# Objetivos

## Objetivo general

Implementar modelos de Machine Learning que permitan reducir la incertidumbre en la cuantificación de la lluvia a partir de datos del radar meteorológico del SIATA y estaciones pluviométricas en el periodo 2015 a 2020.

## Objetivos específicos

* Homogeneizar los conjuntos de datos del radar meteorológico y de las mediciones de pluviómetros a una misma resolución temporal que permita realizar análisis coherentes con el dataset resultante.
* Identificar qué variables provenientes del radar meteorológico muestran mayor relación con la intensidad de precipitación mostrada por las estaciones pluviométricas.
* Probar diferentes modelos de machine learning para identificar aquellos que mejor representan la intensidad de precipitación a partir de datos de reflectividad, evaluados con métricas como el RSME o el MAE.

# Datos

Esta sección detalla los datos recopilados, su estructura, contenido y fuente. Además, se describen los diferentes procesos para la obtención de un dataset que pueda ser suministrado para procesos de validación cruzada en modelos de machine learning. Finalmente, se describen diferentes análisis estadísticos que permiten describir las variables que comprenden el dataset resultante según sus características y patrones.

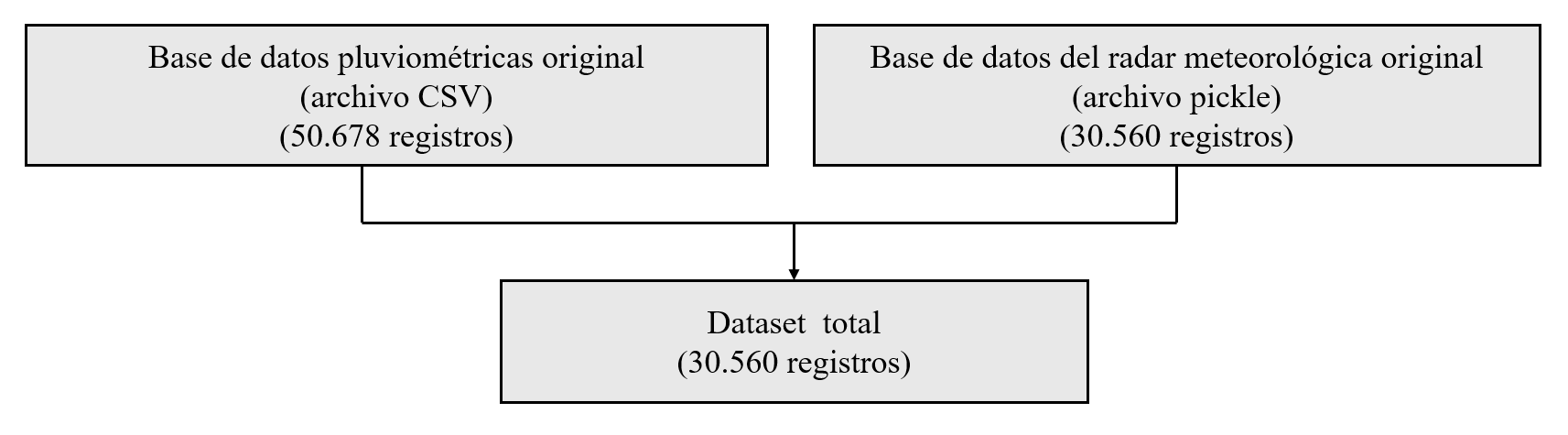
## Datos originales

Se exploraron los datos provenientes de SIATA, los cuales se dividen en dos conjuntos: Por un lado, se tiene información de las estaciones pluviométricas en un formato de serie de tiempo a resolución temporal de 1 minuto. El periodo de registro comprende desde el 2015 (varía según la fecha en que fue instalada cada estación) hasta el 2020. Este dataset contiene un total de 50.678 registros, donde cada registro representa el acumulado de lluvia en cada minuto, y cuenta con 150 columnas, donde cada una es la información de precipitación de cada estación pluviométrica.

Por otro lado, se tiene un conjunto de datos para la información del radar meteorológico cuyo objetivo es realizar una medición indirecta de la precipitación a partir de una señal electromagnética que se conoce como reflectividad, la cual es entregada en un formato espacial. Las variables que comprenden este conjunto de datos son: la reflectividad horizontal en aquellos píxeles, de dicha malla espacial, que coincidan con las coordenadas de ubicación de cada estación pluviométrica, además del promedio espacial de reflectividad de los nueve píxeles que rodean dicha estación, así como el CAPPI, la velocidad radial horizontal, el coeficiente de correlación Copolar, y la reflectividad diferencial, el tipo de precipitación, y la distancia del radar a la estación pluviométrica así como la altura de la estación pluviométrica, para un total de 9 columnas y 30.560 registros. Esta serie de tiempo se encuentra a una resolución de 5 minutos para el periodo del 2015 al 2020.

Este conjunto de datos se obtiene en un formato pickle, que permite guardar y recuperar estructuras de datos complejas facilitando la carga y almacenamiento de datos de manera eficiente y rápida, lo cual es esencial en este proyecto debido al volumen de datos contenidos. Mientras que el conjunto de datos de la información de estaciones pluviométricas se obtiene en un formato CSV.

Figura .  
*Origen del Dataset.*

 Nota. Elaboración propia.

## Datasets

El procedimiento para obtener un dataset unificado entre ambas bases de datos inicia con la homogeneización de la escala temporal entre ambos conjuntos de datos, principalmente los registros pluviométricos son llevados a acumulados 5 minútales en concordancia con la información suministrada por el Radar meteorológico. Además, se calculan nuevas variables a partir de la información pluviométrica minutal, tal como la intensidad de lluvia en diferentes ventanas de tiempo minútales: 5, 10, 15, 30, 60 (Ver ***Ecuación 4***), sin embargo, por sugerencia de SIATA se analiza preliminarmente la intensidad en la ventana de tiempo de 15 minutos; además y se concatena el resultado final para obtener un solo dataset, con las columnas mostradas en la **Tabla 1**.

**Ecuación 4**. Intensidad de precipitación.

Donde PttAcum es la precipitación acumulada en la ventana de tiempo específica, minHora es la cantidad de minutos en una hora y minVent son los minutos de la ventana de tiempo escogida.

Tabla .   
*Variables de dataset.*

|  |  |  |  |
| --- | --- | --- | --- |
| Columna | Descripción | Tipo de dato | |
| PPI\_Reflectivity\_point | Punto de reflectividad del indicador de posición del plano. | Numérico | float32 |
| PPI\_Reflectivity\_area | Área de reflectividad del indicador de posición del plano. | Numérico | float32 |
| CAPPI\_Reflectivity | Indicador de Plan de Posición de Constante Altitud | Numérico | float32 |
| Intensidad\_15min | Cantidad de lluvia/Unidad de tiempo | Numérico | float64 |
| RainType\_SHY95 | Tipo de lluvia | Numérico | float32 |
| Altitud\_Est | Altura en msnm de la estación pluviométrica | Numérico | float64 |
| Dist\_Radar | Distancia del radar a la estación pluviométrica | Numérico | float64 |
| VELH | Velocidad radial horizontal | Numérico | float32 |
| Copolar | Coeficiente de correlación copolar | Numérico | float32 |
| ZDR | Reflectividad diferencial | Numérico | float32 |

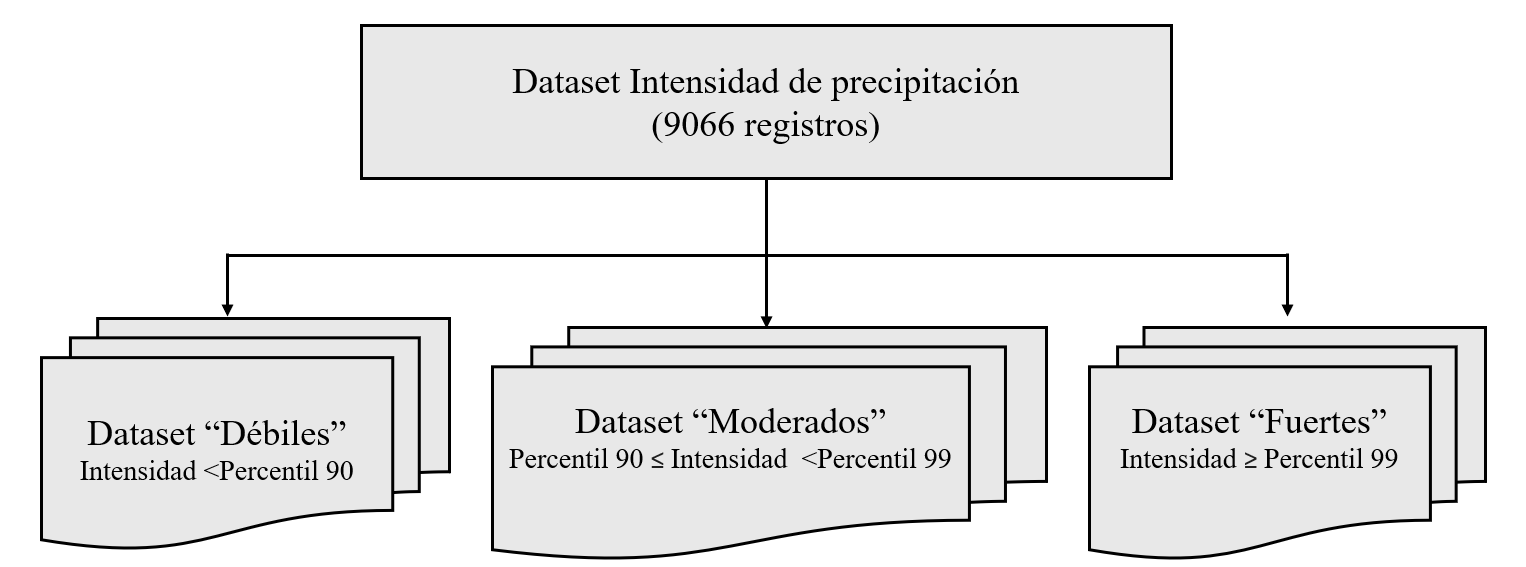
*Nota: elaboración propia.*

El dataset fue filtrado de acuerdo con la presencia de eventos de precipitación (precipitación mayor a 0) y a valores mayores a 0 en reflectividad, eliminando, además, los valores faltantes.

Es importante mencionar que SIATA entrega los datos pluviométricos con un indicador de calidad de datos, donde resalta qué datos se consideran como anómalos. En este sentido, se eliminaron dichos registros que SIATA considera que no tienen coherencia física con la medición de la precipitación. Por tal razón, los valores considerados como registros atípicos de intensidad de precipitación no se eliminan del dataset, ya que estos son eventos poco frecuentes, pero si ocurrentes, por ejemplo, eventos de precipitación extrema; y se hace importante tener estos datos en cuenta a la hora de mejora la QPE (Estimación cuantitativa de precipitación, por sus siglas en inglés).

Finalmente, el dataset se dividió en tres subconjuntos categorizando los eventos de precipitación en tres tipos: Débiles, Moderados y Fuertes, como se observa en la **Figura 2**. Para esto se consideró la intensidad de precipitación en cada tipo de evento, dónde se consideran eventos débiles como aquellos cuya intensidad de precipitación está por debajo del percentil 90 de las intensidades, mientras que lo moderados son aquellos que tienen una intensidad de lluvia entre el percentil 90 y 99 y, por último, en los eventos fuertes su intensidad es superior al percentil 99.

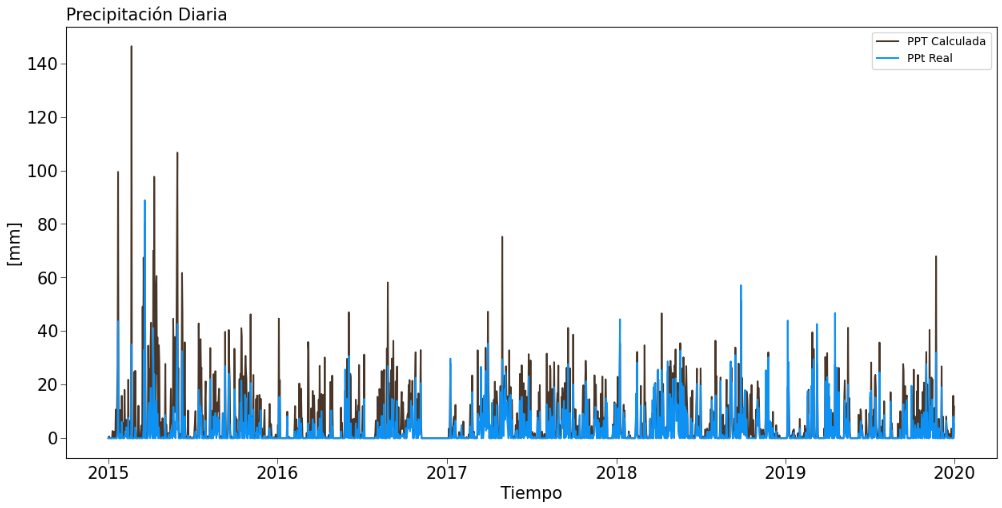
Figura .  
*Clasificación de Eventos de Precipitación por Intensidad*

 Nota. Elaboración propia.

## Analítica descriptiva

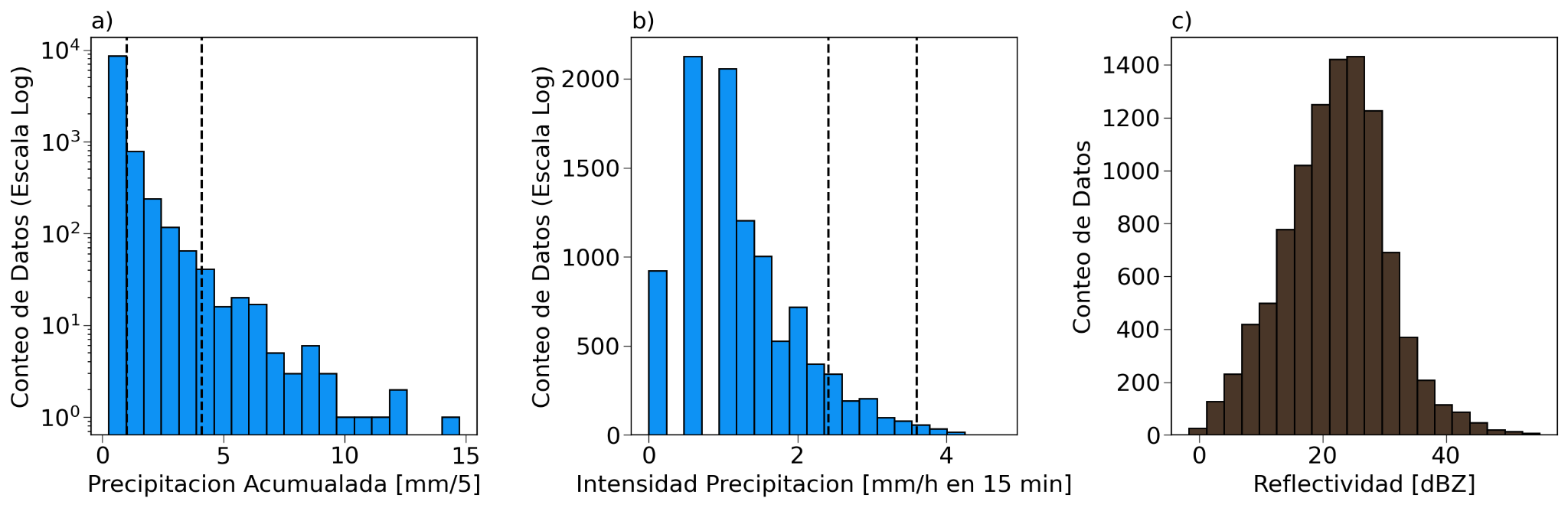
Como se mencionó anteriormente, el dataset utilizado para el desarrollo de este proyecto comprende información desde el 1 de enero del 2015 al 1 enero de 2020. En primer lugar, se exploró la información de la estación pluviométrica “Torre SIATA” ubicada en una región central de la ciudad de Medellín. Para esta exploración preliminar se genera una serie de tiempo de acumulados de precipitación diaria para el periodo de interés, comparada contra la precipitación calculada con información del radar usando métodos matemáticos estándar. La *Figura 3* muestra cómo a inicios del 2015 se alcanzaron altos acumulados diarios de precipitación y, a partir de mediados del mismo año del mismo año, los acumulados disminuyeron durante el resto del periodo de registros. Además, se resalta cómo a finales del 2016 parece que esta estación pluviométrica no tuvo registros de precipitación, sin embargo, dado el amplio registro que se tiene no se optará por hacer imputación de datos para corregir este fragmento de la serie de tiempo.

Figura 3.  
Serie de tiempo de precipitación diaria de la estación “Torre SIATA” (línea azul) y precipitación diaria calculada con información del radar usando métodos matemáticos (línea marrón).

*Elaboración propia, código disponible en el repositorio asociado al proyecto.*

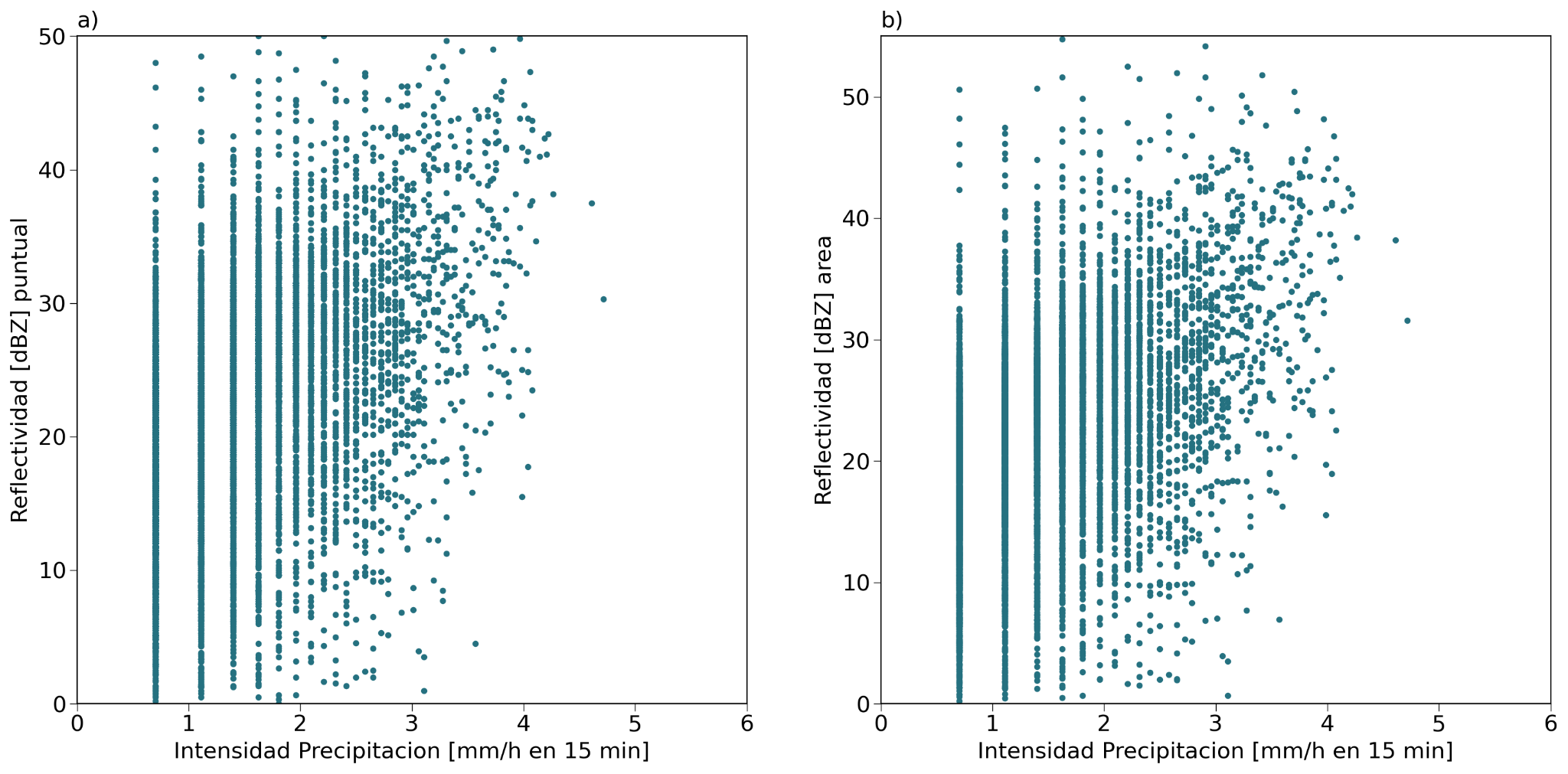
Posteriormente, con el fin de identificar el comportamiento de las magnitudes de las variables principales, como la precipitación acumulada, la intensidad de precipitación (variable objetivo) y la reflectividad del radar meteorológico, se generaron histogramas que permitan dar cuenta de dicho comportamiento. El panel a) y b) de la Figura 4 muestran el histograma para la precipitación acumulada en 5 minutas y la intensidad de precipitación en mm/h respectivamente en una escala logarítmica. Es evidente un fuerte sesgo positivo en estos dos histogramas, donde los bajos valores tienen una gran cantidad de registros, mientras que valores medios y muy altos disminuyen considerablemente, especialmente estos últimos. Este comportamiento puede sugerir que se está trabajando con un conjunto de datos desbalanceados, donde los modelos de Machine Learning utilizados pueden ajustarse más fácilmente a valores bajos de intensidad de precipitación y pueden tener falencias en los valores altos de la lluvia subestimando los eventos de lluvia extremos que son de gran importancia en la gestión de riesgos a desastres. Por su parte, el panel c) el cual muestra el histograma para la reflectividad, evidenciando un comportamiento normal con amplia variabilidad entre los 10 dBZ y los 40 dBZ. Esta amplia variabilidad también puede ser un factor a tener en cuenta, pues si bien se espera que altos valores de reflectividad estén asociados altas intensidades de precipitación (Sepúlveda Berrío, 2015), la relación no es estrictamente lineal, es decir, valores altos de reflectividad pueden estar asociados a intensidades medias de intensidad de precipitación.

Figura 4.   
Histogramas para: a) Precipitación acumulada en 5 minutos; b) Intensidad de Precipitación; c) Reflectividad del Radar. Las líneas puntadas sugieren el percentil 90 y 99 respectivamente.

 Nota. Elaboración propia.

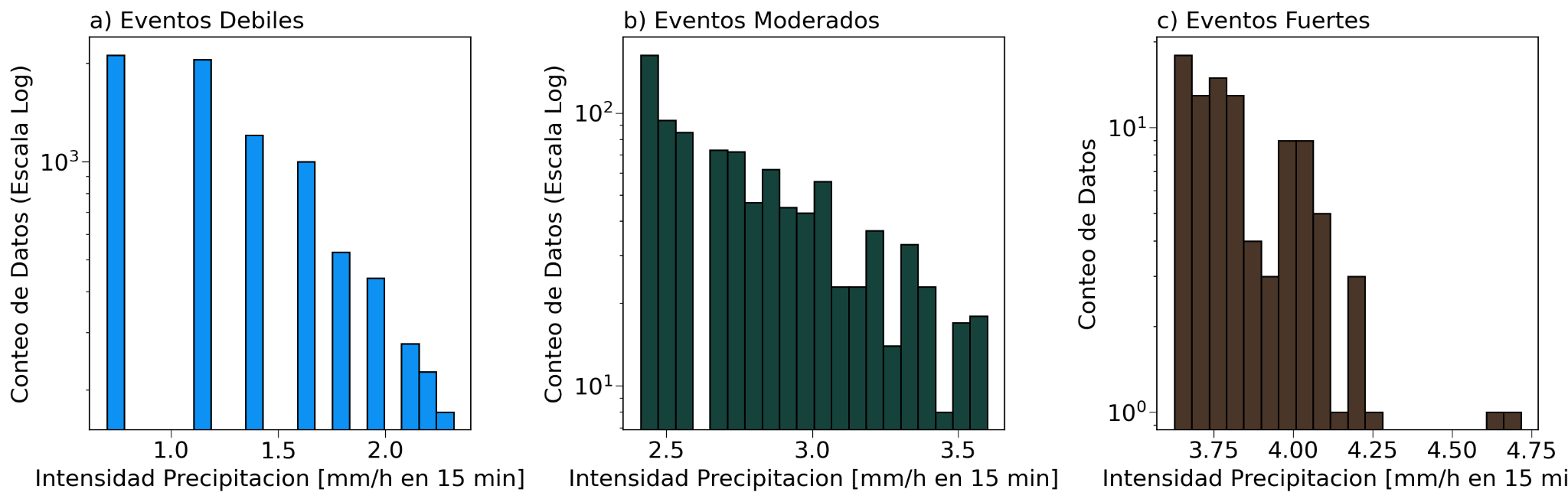
Esto último se hace notorio en la Figura 5,donde el panel a) muestra el diagrama de dispersión entre la intensidad de precipitación y la reflectividad del pixel del radar correspondiente a la estación pluviométrica, mientras que el panel b) muestra el mismo diagrama, pero usando la reflectividad de los nueve píxeles que rodean dicha estación. Si bien hay una alta dispersión en la relación entre la reflectividad del radar y la intensidad de la lluvia, si se observa como altos valores en la intensidad se relaciona con mayores valores de reflectividad. Sin embargo, cuando se observan bajos valores de intensidad, estos pueden relacionarse tanto a bajos como altos valores de reflectividad. Nótese, además, cómo mientras menor es la intensidad de la lluvia, se fortalece una discretización en sus valores, esto está relacionado a que las estaciones pluviométricas generan registros discretos en múltiplos de 0.254 mm, así pues, mientras más bajo los valores de intensidad más marcada la discretización de estos.

Figura 5.  
Diagrama de dispersión para: a) intensidad de precipitación y la reflectividad del píxel del radar correspondiente a la estación pluviométrica; b) intensidad de precipitación y la reflectividad de los nueve pixeles que rodean la estación pluviométrica.

 Nota. Elaboración propia.

Finalmente, como se mencionó en la sección 3.2, se dividió el dataset en 3 subconjuntos basados en la intensidad de precipitación, con el fin de minimizar el desbalance en el registro de la intensidad de los eventos de lluvia. La **Figura 6** muestra el histograma para cada uno de los tipos de eventos clasificados: Débiles, Moderados y Fuertes y se hace evidente como si bien sigue existiendo un sesgo positivo en los subconjuntos, la diferencia la magnitud de los registros disminuyó considerablemente, lo que se espera pueda ayudar a los modelos de Machine Learning a tener mejores resultados.

**Figura 6.**   
Histogramas para: a) Eventos considerados Débiles (Intensidad < Percentil 90); b) Eventos considerados Moderados (Percentil 90 < Intensidad < Percentil 90); c) Eventos considerados Débiles (Percentil 99 < Intensidad).

 Nota. Elaboración propia.

# Referencias

Hong, Y., & Gourley, J. J. (2017). *Radar Hydrology.* Oxford, UK.

Rollenbeck, R., Orellana-Alvear, J., Rodriguez, R., Macalupu, S., & Nolasco, P. (2021). Calibration of X-band radar for extreme events in a spatially complex precipitation region in north Peru: Machine learning vs. empirical approach. Obtenido de https://www.mdpi.com/2073-4433/12/12/1561

Sayeed Khan, R., & Ehsan Bhuiyan, A. (2021). Artificial Intelligence-Based Techniques for Rainfall Estimation Integrating Multisource Precipitation Datasets.

Sepúlveda Berrío, J. (2015). *Estimación cuantitativa de precipitación a partir de la información de Radar Meteorológico del Área Metropolitana del Valle de Aburrá (Doctoral dissertation).* Obtenido de https://repositorio.unal.edu.co/bitstream/handle/unal/58046/1020423608.2015.pdf

Sloat, L. L., Gerber, J. S., Samberg, L. H., Smith, W. K., Herrero, M., Ferreira, L. G., & West, P. C. (2018). Increasing importance of precipitation variability on global livestock grazing lands. *Nature Climate Change*, 214-218. Obtenido de https://www.nature.com/articles/s41558-018-0081-5

Verdelho, F. F., Pavam Jr, L., Calvetti, L., Oliveira, L. E., & Zanata Alves, M. A. (2024). Quantitative Precipitation Estimation Using Weather RadaData and Machine Learning Algorithms for the Southern Region of Brazil. Remote Sensing. 1971. Obtenido de https://www.mdpi.com/2072-4292/16/11/1971