# EVALUACIÓN DE LA REPRESENTATIVIDAD DE LA CALIDAD DEL AGUA EN LOS PUNTOS DE MONITOREO DE LA RCHB A PARTIR DE UN ANÁLISIS MULTIVARIADO DE CARGAS CONTAMINANTES

**JUAN SEBASTIÁN RUBIANO PERILLA**
**LAURA JOHANNA VERA QUINTERO**

**UNIVERSIDAD SANTO TOMÁS**
**FACULTAD DE INGENIERÍA AMBIENTAL**
**DIVISIÓN DE INGENIERÍAS**
**BOGOTÁ, D. C.**
**2019**

# EVALUACIÓN DE LA REPRESENTATIVIDAD DE LA CALIDAD DEL AGUA EN LOS PUNTOS DE MONITOREO DE LA RCHB A PARTIR DE UN ANÁLISIS MULTIVARIADO DE CARGAS CONTAMINANTES

**JUAN SEBASTIÁN RUBIANO PERILLA**
**LAURA JOHANNA VERA QUINTERO**

**TESIS DE GRADO PARA OPTAR POR EL TÍTULO DE INGENIERÍA AMBIENTAL**
**MODALIDAD: SOLUCIÓN A UN PROBLEMA DE INGENIERÍA**

**DIRECTOR:**
**DAVID ANDRÉS ZAMORA ÁVILA**
**Ingeniero Civil**
**Master en Hidrosistemas**

**UNIVERSIDAD SANTO TOMÁS**
**FACULTAD DE INGENIERÍA AMBIENTAL**
**DIVISIÓN DE INGENIERÍAS**
**BOGOTÁ, D. C.**
**2019**

**CONTENIDO**

pág.

## LISTA DE FIGURAS

pág.

## LISTA DE TABLAS

pág.

## LISTA DE ECUACIONES

pág.

## LISTA DE ANEXOS

pág.

## RESUMEN

La Red de Calidad Hídrica de Bogotá (RCHB) realiza seguimiento a la calidad y cantidad del agua de los principales ríos de la ciudad: Torca, Salitre, Fucha y Tunjuelo, localizados en el perímetro urbano de ciudad. Esta red de monitoreo ha permanecido estática por más de diez (10) años sin tener en cuenta que la calidad y la cantidad del agua en la ciudad es altamente dinámica en el tiempo y el espacio, y que los procesos de desarrollo urbano e industrial han cambiado desde que esta herramienta de gestión entró en funcionamiento en el año de 2006. Además, a la par de estos cambios, la hidrología de las cuencas ha sido modificada por efectos del incremento de áreas impermeables debido a la expansión de infraestructura urbana, entre otros efectos derivados de las actividades antrópicas.

Caudales y concentraciones de diferentes determinantes de la calidad del agua en la RCHB han permitido caracterizar el comportamiento del recurso hídrico en las cuencas urbanas de la ciudad. Sin embargo, evaluar estos componentes de forma independiente para determinar la relevancia de un punto de monitoreo (PM) limita los hallazgos, dado que la hidrología urbana corresponde a un sistema no lineal que relaciona la cantidad y calidad del agua de los sistemas de recolección de agua lluvia y residual, y las alteraciones al sistema como consecuencia de la expansión urbana.

Por lo tanto, en la presente investigación se realizó un análisis de la representatividad de los PM a través de una evaluación multivariada de las cargas contaminantes (CC) de la RCHB para el periodo de tiempo del año 2006 al 2018. En primera instancia se hizo una caracterización espacial del área de influencia a la que pertenece la RCHB, con el fin de conocer la hidrografía e hidrología de la zona, y así mismo con los datos proporcionados por la Secretaría Distrital de Ambiente (SDA) fueron calculadas las CC integrando datos de concentraciones de las variables de calidad (es decir, DBO5, DQO, SST, Nitrógeno Total, Fosforo Total, SAAM y Grasas y Aceites) y caudales.

El análisis de la representatividad de los PM se hizo mediante la aplicación de diferentes métodos multivariados: En primer lugar, fueron detectados y eliminados las muestras atípicas a través de algoritmo distancia Mahalanobis, obteniendo un máximo de hasta 22,77 % de muestras atípicas en algunos PM de la RCHB. Después de eliminar los valores atípicos, fueron conformados nuevos conjuntos de datos con las CC y caudales para cada río, y usados para seleccionar las variables de calidad y cantidad más importantes que representan la dinámica del recurso hídrico por cada río mediante el algoritmo *Random Forest.*

Por último, se evaluaron los datos de carga de las variables más importantes por río con el algoritmo de agrupamiento *Expectation Maximization* para determinar el número de PM que representan la dinámica de las cargas contaminantes en la RCHB obteniendo así un número óptimo de PM por cada río: cuatro (4) para Torca, cinco (5) para Salitre, seis (6) para Fucha y siete (7) para Tunjuelo. Con el desarrollo de este proyecto se buscó aportar conocimiento sobre la aplicación de metodologías para optimizar la actividad de monitoreo, y con esto mejorar los procesos gestión, control y seguimiento del recurso hídrico en la ciudad de Bogotá.

**Palabras clave:** Red de Calidad Hídrica (RCH), distancia de Mahalanobis, *Random Forest* (RF)*, Expectation Maximization* (EM), Carga Contaminante (CC).

## ABSTRACT

The Water Quality Network of Bogotá is established in the 4 main rivers belonging to the city of Bogotá which are: Torca, Salitre, Fucha and Tunjuelo, this monitoring network has remained static for more than ten (10) years without taking into account that the quality and quantity of water in the city is highly dynamic in time and space, and that urban and industrial development processes have changed since this management tool came into operation in 2006. In addition, at the same time as these changes, the hydrology of the basins has been modified by the effects of the increase in impermeable areas due to the expansion of urban infrastructure.

Flow rates and concentrations of different determinants of water quality in the water quality network of Bogotá have made it possible to characterise the behaviour of the water resource in the urban basins of the City. However, evaluating these components independently to determine the relevance of a monitoring point limits the findings, since urban hydrology corresponds to a non-linear system that relates the quantity and quality of water from rain and wastewater collection systems, and the alterations to the system as a consequence of urban expansion.

Therefore, in the present investigation an analysis of the representativeness of the monitoring points was carried out through a multivariate evaluation of the polluting loads of the RCHB for the time period from 2006 to 2018. First, a spatial characterization was made on the influence area of the water quality network of Bogotá belongs, in order to know the hydrography and hydrology conditions, and also with the data provided by the District Secretariat of Environment was calculated of the polluting load integrating concentrations of quality variables (BOD5, COD, TSS, Total Nitrogen, Total Phosphorus, SAAM and Fats and Oils) and streamflows.

The analysis of the representativeness of the monitoring points was done through the application of different methods, in this way the detection of atypical data through the Mahalanobis distance algorithm was first evaluated, obtaining a maximum percentage of up to 22.77 % of atypical values in the monitoring points network. In accordance with the above, a new, more reliable database without outliers was obtained to be used in the selection of the most important quality and quantity variables that represent the dynamics of the water resource by river using the Random Forest algorithm.

Finally, the load data of the most important variables per river were evaluated with the grouping algorithm Expectation Maximization to determine the number of monitoring points that represent the dynamics of the polluting loads in the water quality network of Bogotá, thus obtaining an optimal number of monitoring points per River of 4 to Torca, 5 to Salitre, 6 to Fucha and 7 to Tunjuelo. The development of this project seeks to provide knowledge on the application of methodologies to optimize the monitoring activity, and thus improve the management processes, control and monitoring of water resources in the city of Bogotá.



**Keywords:** Water Quality Network, Mahalanobis distance, Random Forest, Expectation Maximization, pollutant load.

# INTRODUCCIÓN

La calidad del agua es esencial para la salud de los ecosistemas acuáticos y los seres humanos. El seguimiento adecuado puede proporcionar información importante sobre el estado y cambios espacio-temporales de la calidad del agua, además, ayuda a gestionar los recursos hídricos, los servicios ecosistémicos y controlar la contaminación [1].

El monitoreo de la calidad del agua sigue siendo un proceso complejo debido a la gran cantidad de factores como los hidrológicos, climáticos y geográficos, los cuales son indispensables para conocer la dinámica del área de estudio, de hecho, el problema de la planificación y optimización de una red de monitoreo de calidad para aguas superficiales ha sido abordado por varios investigadores [2]. Desde la década de 1940 se han publicado muchos manuales, directrices y documentos sobre el tema, aunque, muchos de estos enfoques no se han implementado, ya que son “demasiado generales” o demasiado específicos (es decir, demasiado limitados para los casos de estudio), o simplemente demasiado complejos para que un administrador de cuencas hidrográficas las incorpore fácilmente al diseño de las redes de monitoreo, dadas las restricciones de tiempo y presupuesto [2].

Para el desarrollo de este proyecto se establece una metodología que se dividió en distintas fases y estas se basan principalmente en algoritmos computacionales que se ejecutaron en la plataforma RStudio, como lo son en su orden: distancia de Mahalanobis, RF y EM, estos algoritmos se ubican metodológicamente en una rama de la inteligencia artificial denominada aprendizaje automático o *Machine Learning,* la cual es muy usada en la actualidad para estudios de este tipo y con bases de datos extensas con el fin de identificar patrones complejos y así analizar comportamientos futuros [3].

La presente investigación analiza la representatividad de los PM dentro del perímetro urbano de la ciudad de Bogotá, evaluando los registros históricos de la actividad de monitoreo de la RCHB por medio de los valores de CC de los determinantes de la calidad del agua y el caudal, donde se evidencia una disminución en la toma de datos de aproximadamente el 45 % en todos los PM de la red, entre el segundo semestre del año 2006 y el primer semestre del año 2018.

También se ha presentado una variación interanual en la cantidad de muestra recolectadas, lo que indica que no se tiene un patrón secuencial en los diferentes periodos del año para el monitoreo de calidad y cantidad del agua. Es decir que la identificación de los puntos óptimos de monitoreo no solo podría reducir los costos, sino también brindar una mejor evaluación de cada cuenca hidrográfica. Además, la optimización de la red de monitoreo puede permitir a los administradores de cuencas hidrográficas priorizar objetivos específicos para diseñar una red de monitoreo más efectiva.

Este documento contiene 6 capítulos, el capítulo 1 denominado marco teórico se centra en la explicación de los materiales y métodos utilizados en esta investigación, en este capítulo se da a conocer la caracterización general del funcionamiento de la RCHB junto con el análisis de los datos disponibles de la RCHB y por último se dan a conocer los métodos aplicados a la metodología de desarrollo del proyecto.

En el capítulo 2 se muestra cómo se desarrolla cada fase de la metodología dando a conocer al detalle todos y cada uno de los pasos que se siguieron para obtener los resultados y los materiales usados, en el capítulo 3 se presentan los resultados obtenidos de acuerdo con la metodología propuesta los cuales permitieron realizar diferentes análisis que ayudaron a generar una propuesta para la optimización de la RCHB, el capítulo 4 se da a conocer el impacto social generado con la aplicación del proyecto, desde un enfoque en el ahorro de costos con la optimización de la RCHB y finalmente en los capítulos 5 y 6 se presentan las conclusiones y recomendaciones que responden a los objetivos planteados para la investigación.

# OBJETIVOS

## OBJETIVO GENERAL

Analizar la representatividad de los puntos de monitoreo a través de una evaluación multivariada de las cargas contaminantes en los puntos de monitoreo de la red calidad hídrica de Bogotá localizados dentro del perímetro urbano en el periodo de tiempo del año 2006 al 2018.

## OBJETIVOS ESPECÍFICOS

- Evaluar la presencia de datos atípicos en las cargas contaminantes obtenidas de los puntos de monitoreo de la RCHB mediante el método multivariado distancia de Mahalanobis.
- Definir las cargas contaminantes más representativas de la dinámica calidad del agua en los puntos de monitoreo de la RCHB por medio del índice de impureza de Gini del algoritmo *Random Forest.*
- Identificar qué puntos de monitoreo de la RCHB son más representativos en el espacio y el tiempo de la dinámica de las cargas contaminantes transportadas.

# 1. MARCO TEÓRICO

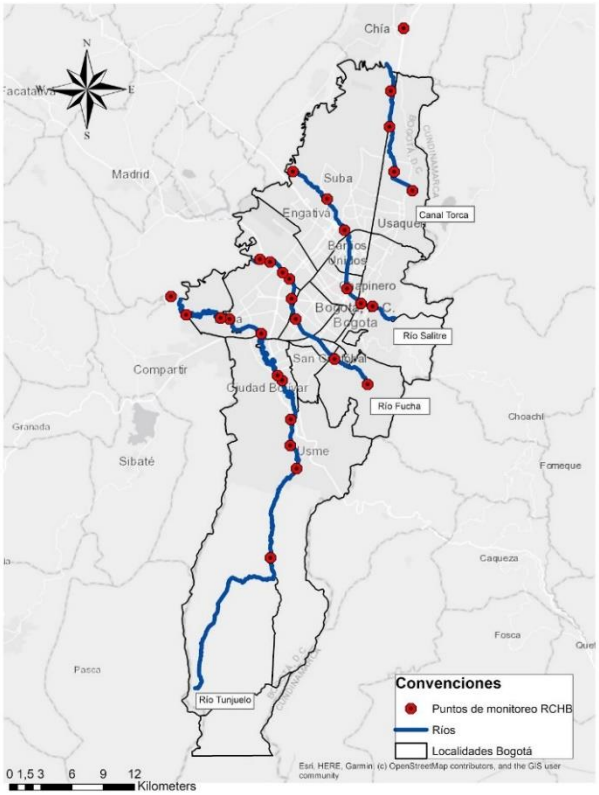
## 1.1 CONTEXTO DE LA RED DE CALIDAD HÍDRICA DE BOGOTÁ

La herramienta de gestión del recurso hídrico aplicada para el seguimiento de la calidad del agua en los ríos principales del Distrito Capital es la RCHB, la cual monitorea en su red tradicional 30 PM localizados a lo largo de los ríos Torca, Salitre, Fucha y Tunjuelo, desde la parte alta a sus desembocaduras en el río Bogotá [4].

En las campañas de monitoreo se determina la presencia y magnitud de parámetros físicos, químicos y biológicos en el agua, además se realizan aforos para estimar el caudal [4]. Los datos recolectados en las campañas de monitoreo permiten identificar el estado de los principales ríos de la ciudad, establecer su uso potencial y modelar la calidad de sus aguas orientada a la toma de decisiones para a su conservación y gestión.

En la siguiente figura se muestra la ubicación espacial de los PM, adicionalmente en el **Anexo *A*** se relacionan los puntos de la RCHB con el nombre de cada PM, el río al que pertenecen, el nombre de identificación y su abreviatura de cada punto, y por último el tramo del río donde se encuentra ubicado.

Figura 1. Puntos de monitoreo de la RCHB




Fuente: Autores.

**1.1.1 Caracterización hidrológica urbana.** A continuación se presenta la caracterización hidrológica del área de influencia urbana de los 4 ríos de la RCHB, dando a conocer características geográficas, morfológicas, ambientales, urbanas y poblacionales de cada río para comprender su contexto en la red.

**1.1.1.1 Canal Torca.** El eje principal de este cauce cuenta con una longitud de 13.06 km desde su nacimiento en los cerros orientales (el conjunto residencial Bosque de Pinos ubicado en la Carrera 6 con Calle 153), desemboca al sistema humedal Torca-Guaymaral a altura de la Autopista Norte, en cercanía a los terrenos del cementerio Jardines de Paz, y para después drenar sus agua al norte de la cuenca media del río Bogotá [5].

El sistema de alcantarillado del canal Torca se encuentra como un sistema separado, que tiene como ejes en la zona Nororiental el drenaje de aguas lluvia que es drenado hacia el humedal Torca-Guaymaral, y a su vez drena al Norte de la cuenca media del río Bogotá [6].

Por otra parte, en la zona Occidental la red de alcantarillado que drena al humedal está conformada por:

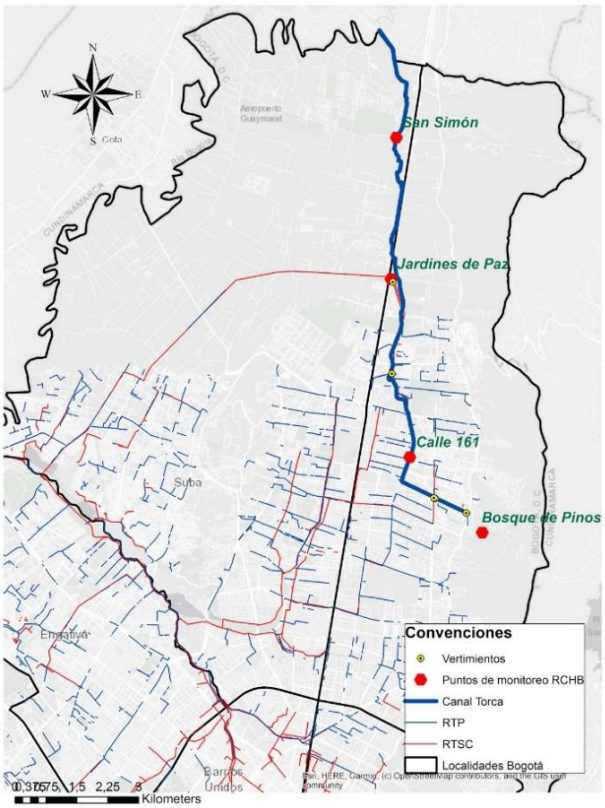
I. **Sistema sanitario**: conformado por el interceptor del río Bogotá-Torca-Salitre, al cual llegan las aguas residuales y las conduce hasta la planta de tratamiento del Salitre [6].

II. **Sistema pluvial:** conformado por el canal El Cedro (que más adelante se llama el canal Torca), que recibe los canales San Cristóbal y Serrezuela, lleva después las aguas al humedal Torca, para posteriormente entregarlas a la cuenca media del río Bogotá [6].

La principal fuente de contaminación en el canal Torca en su primer tramo corresponde a la red de alcantarillado público (sanitaria, pluvial o combinada (aguas lluvias y residuales)) que transportan principalmente aguas residuales domésticas [5]. El área asociada al segundo tramo del canal Torca no cuenta con red de alcantarillado público. Las descargas localizadas sobre este tramo corresponden a usuarios generadores de vertimientos puntuales, tales como instituciones educativas y conjuntos residenciales, que vierten sobre una red de acequias que conducen las aguas residuales al río [5].

Los puntos de vertimientos que descargan sobre el canal Torca y sus afluentes aportan principalmente, cargas de materia orgánica, sólidos suspendidos totales (SST) y coliformes fecales [5]. Lo anterior se ha evidenciado por los resultados de las caracterizaciones realizadas por RCHB en los cuatro PM de la calidad y cantidad del agua, que estan distribuidos en los dos tramos que conforman esta corriente [5].

Figura 2. Puntos de monitoreo y vertimientos de la RCHB del canal Torca




Fuente: Autores

En la

Figura *2*, se presenta el canal Torca señalando en él sus 4 PM, siendo este el efluente perteneciente a la RCHB (color rojo) con menos PM debido a que el canal Torca también es el cauce con menor longitud comparado con los otros ríos de la RCHB, también es importante resaltar que el área de influencia de este canal corresponde conjuntamente a las localidades de Suba y Usaquén ubicadas en el Norte de la ciudad, adicionalmente se presenta la Red Troncal Pluvial (RTP) y la Red Troncal Sanitaria Combinada (RTSC), con el fin de observar la influencia de estas en el canal.

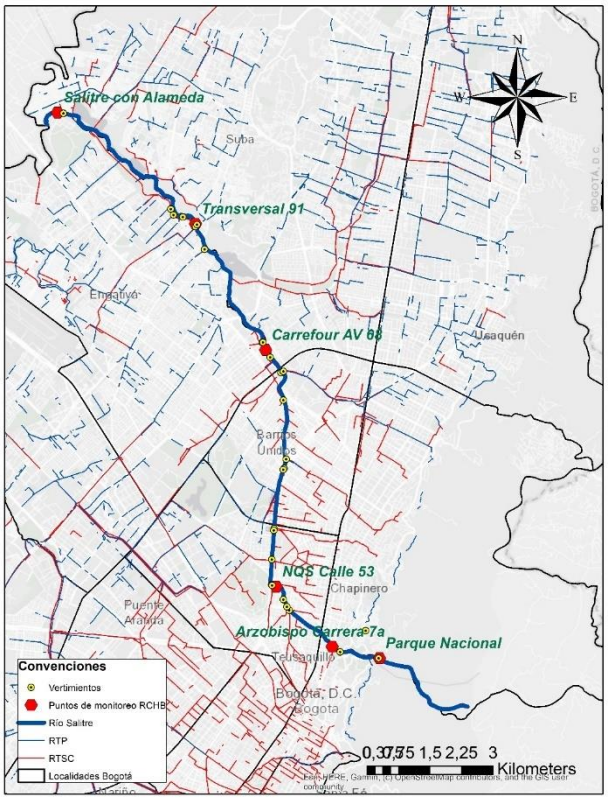
**1.1.1.2 Río Salitre.** Posee una longitud de cauce principal de 19.76 km y la pendiente media del cauce es de 3.32 % [5]. Su altura promedio es de 2,870 msnm, donde la cota máxima está por el orden de los 3,200 msnm y la mínima está sobre los 2,540 msnm aproximadamente [5].

Este río nace en los cerros orientales donde recibe el nombre de río Arzobispo, el cual es canalizado desde el Parque Nacional Enrique Olaya Herrera (carrera 7ª) hasta la carrera 30, siendo límite entre las localidades de Chapinero y Santa Fe [5]. A partir de su cruce con la Avenida NQS se denomina río Salitre hasta su cruce con la carrera 68, donde recibe el nombre de río Juan Amarillo en referencia al humedal existente en esta parte de la ciudad (entre las localidades de Engativá y Suba), el cual sirve como cuerpo amortiguador natural de crecientes y cuya capacidad ha sido reducida por acción antrópica [5]. El río Salitre desemboca en el río Bogotá en inmediaciones de la planta de tratamiento de aguas residuales El Salitre [5].

Las principales fuentes de contaminación de esta corriente son aguas residuales domésticas. En general, la mayoría de puntos de descarga que vierten sus aguas principalmente sobre el río Salitre y sus afluentes corresponden al alcantarillado público de la ciudad que aportan entre otras cargas de materia orgánica, sólidos suspendidos totales (SST) y coliformes fecales [4]. Esto se ha evidenciado con los resultados de monitoreo de la calidad y cantidad del agua realizados por la RCHB, que a lo largo de esta corriente tiene seis PM repartidos en los cuatro tramos que lo conforman [4].

Además, la cuenca del río Salitre recibe parte de la escorrentía generada en las localidades de Usaquén, Chapinero, Santafé, Engativá, Suba, Barrios Unidos y Teusaquillo.

Figura 3. Puntos de monitoreo y vertimientos de la RCHB del río Salitre




Fuente: Autores.

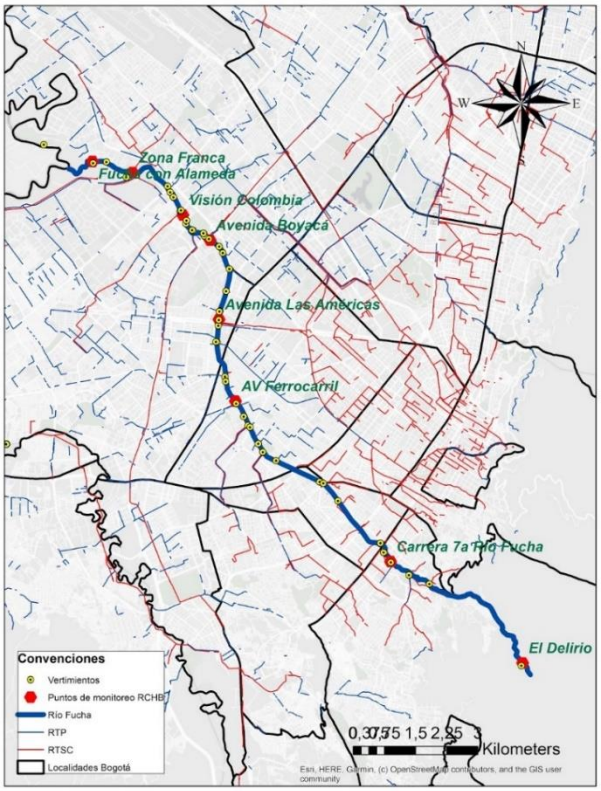
En la Figura *3* se muestran los 6 PM del río Salitre (color rojo), así mismo se observan los puntos de vertimientos distribuidos a lo largo del río. La RTP en color azul y la RTSC en color rojo representan algunas de las descargas residuales que también afectan la calidad del río, además de ser un indicador poblacional.

**1.1.1.3 Río Fucha.** El río Fucha es uno de los cuerpos hídricos más importantes de la sabana de Bogotá con una longitud de 17.30 km y una pendiente promedio del 5.3 %. Este río nace en los cerros orientales y como producto de la confluencia de las quebradas La Osa y Upata [5].

El eje principal de drenaje de la cuenca inicia en la zona Suroriental de la misma donde recibe el nombre de río San Cristóbal. En este punto se encuentra con su cauce natural que toma dirección Oriente Occidente, a partir de la carrera 7 hasta la carrera 96 (en inmediaciones de la Zona Franca de Fontibón) se encuentra canalizado con una sección trapezoidal revestida en concreto, y por último desemboca en la margen izquierda del río Bogotá [5].

El río Fucha tiene varios cuerpos lenticos (humedales) asociados a su dinámica, como Techo, El Burro, La Vaca y Capellanía. La red de alcantarillado de esta cuenca consta de tres sistemas (combinado, pluvial y sanitario) con una longitud total existente de 1,787 km [5]. La red combinada está localizada al Oriente de la cuenca del río y drena, a través de canales e interceptores hacia un área en el Occidente de la cuenca donde el sistema recolección y transporte es separado (pluvial y sanitario) [6]. Las principales fuentes de contaminación de esta corriente son aguas residuales domésticas e industriales descargadas al río por las estructuras del sistema de alcantarillado público [6].

Figura 4. Puntos de monitoreo y vertimientos de la RCHB del río Fucha




Fuente: Autores.

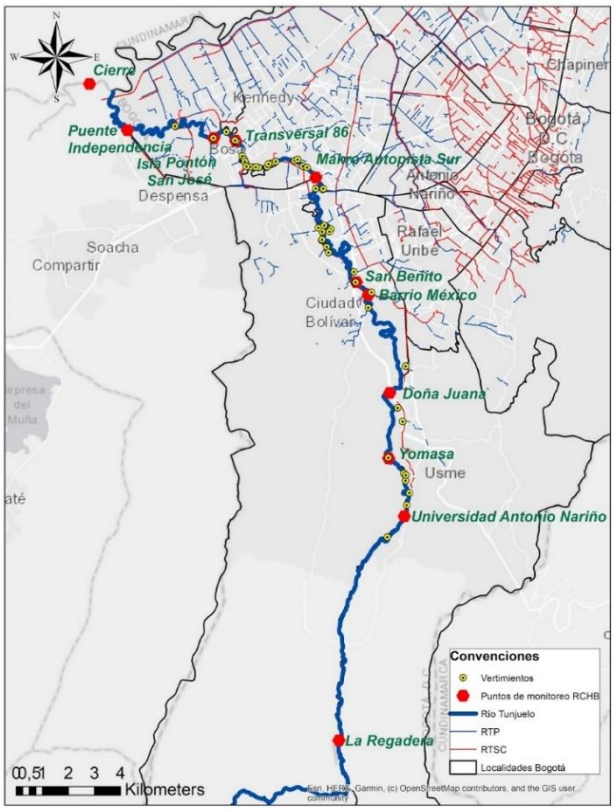
Para el mapa de influencia y localización de los PM de la RCHB correspondiente al río Fucha (Figura *4*), se cuenta con la ubicación de los 8 puntos que la conforman a lo largo de su cauce que nace en el Oriente la ciudad de Bogotá y en su recorrido atraviesa las localidades de San Cristóbal, Antonio Nariño, Puente Aranda y Fontibón siendo esta su área de influencia y localización de los vertimientos que están presentes en este cauce. La RTP en color azul y la RTSC en color rojo indican la influencia que tienen en las descargas residuales al río.

**1.1.1.4 Río Tunjuelo.** El río Tunjuelo nace a partir de la confluencia de los ríos Chisacá, Mugroso y Curubital en las estribaciones del Páramo del Sumapaz, las cuales convergen al embalse La Regadera a 2,900 msnm de altitud que tiene la capacidad de contener un volumen de agua de 4 millones de $m^3$ [5]. A partir de este embalse se llama río Tunjuelo, donde toma una dirección sur a norte por el valle longitudinal de Usme. Al llegar a la zona urbana Sur de Bogotá, donde toma un rumbo Noroeste hasta la confluencia con el río Bogotá [5]. Este río tiene una extensión de 73 km, siendo su área de drenaje urbana 41,427 hectáreas y 4,237 hectáreas rurales [5]. La cota más alta de la cuenca, de acuerdo con el sistema de referencia del Instituto Geográfico Agustín Codazzi (IGAC), se localiza por encima de los 3,700 msnm, en tanto que la cota más baja se localiza a 2,530 msnm [5].

Las principales fuentes de contaminación en el río Tunjuelo son aguas residuales domésticas e industriales [5]. En su mayoría de puntos de vertimientos que descargan mediante el sistema de alcantarillado público aportan entre otras cargas de materia orgánica, SST, fósforo total, nitrógeno total y coliformes fecales, conforme a los resultados

del monitoreo realizado en los diez puntos de seguimiento de la calidad y cantidad del agua [5].

Figura 5. Puntos de monitoreo y vertimientos de la RCHB del río Tunjuelo




Fuente: Autores.

En la Figura *5* se pueden observar los 10 PM localizados a lo largo del río Tunjuelo y los respectivos vertimientos asociados al cauce del río. Su área de influencia son las localidades de Kennedy, Ciudad Bolívar, Usme, Antonio Nariño, Rafael Uribe, San Cristóbal, Tunjuelito y Sumapaz. La RTP en color azul y la RTSC en color rojo, indican el asentamiento de la población y representan en cierta medida la influencia de las descargas residuales al río.

## 1.2 CARACTERIZACIÓN DE DATOS DISPONIBLES DE LA RED DE CALIDAD HÍDRICA DE BOGOTÁ

***1.2.1 Datos históricos de la RCHB 2006-2015.*** *Se utilizaron los registros históricos de la RCHB RCHB desde el año 2006 al segundo semestre del 2015 (denominada en adelante como datos históricos), esta datos históricos), esta base de datos corresponde a las concentraciones de los determinantes de la calidad del determinantes de la calidad del agua tales como: DBO5 (Demanda Biológica de Oxigeno consumidas en 5 días), consumidas en 5 días), DQO (Demanda Química de Oxigeno), SST (Solidos Suspendidos Totales), $N_{TOTAL}$ Totales), $N_{TOTAL}$ (Nitrógeno Total), $P_{TOTAL}$ (Fosforo Total), SAAM (Sustancias Activadas al Azul de Metileno) y Azul de Metileno) y GYA (Grasas Y Aceites) además de datos de cantidad de agua o caudal, esta información esta información está disponible para cada PM de los 4 cauces que conforman la RCHB, la cantidad de datos cantidad de datos correspondiente a cada PM se muestra en el*

**Anexo *B***.

***1.2.2 Base de datos RCHB 2017-2018.*** *También se hace uso de los registros pertenecientes a la pertenecientes a la RCHB para el intervalo de tiempo desde marzo del año 2017 hasta marzo del 2018, marzo del 2018, igualmente esta serie de datos contiene las concentraciones de las variables de: DBO5, DQO, variables de: DBO5, DQO, SST, $N_{TOTAL}$, $P_{TOTAL}$ y SAAM y del mismo modo datos de caudal para cada PM de la para cada PM de la RCHB, la cantidad de muestras correspondiente a cada PM se muestra en el*

*en el*

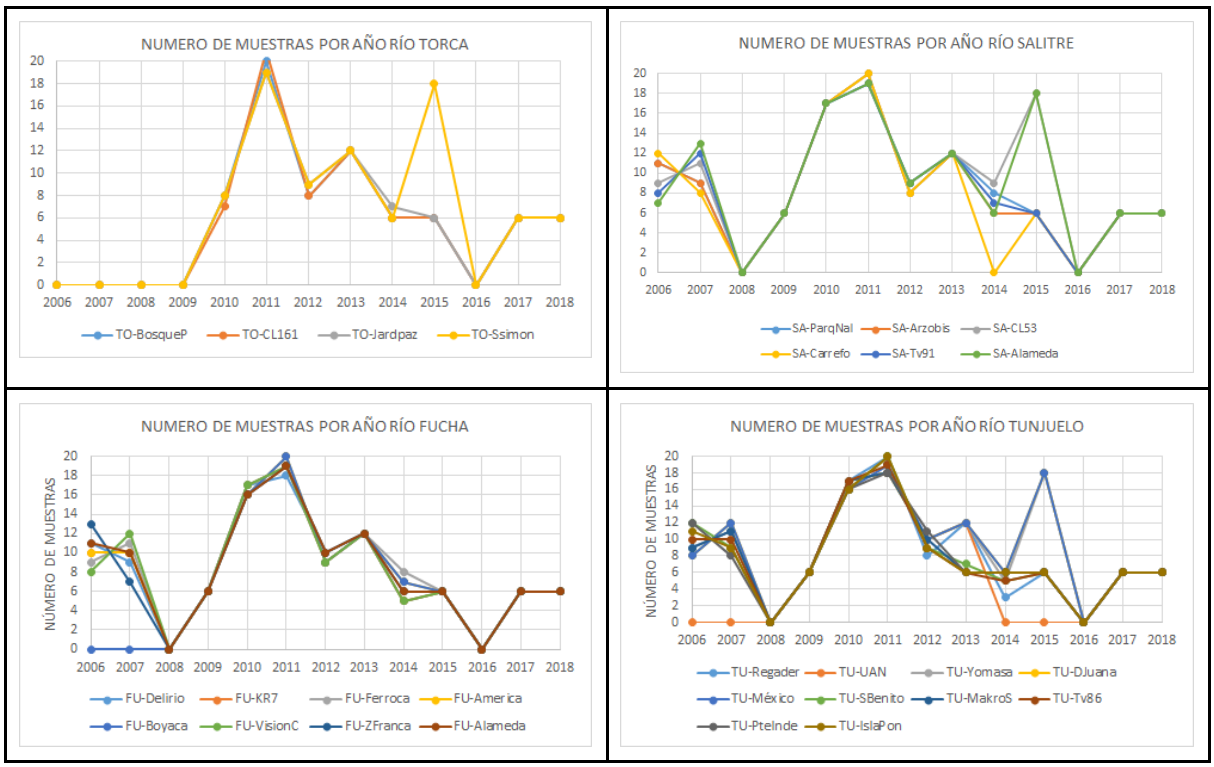
**Anexo *B***. Estas bases de datos fueron puestas a disposición por la SDA mediante una petición directa a esta entidad.

**1.2.3 Disponibilidad y cantidad de datos.** En la *y* afecta la representatividad de las muestras.

Figura *6* se presentan la disponibilidad de los datos por cada río en cuanto al número de monitoreos por año, para el periodo de estudio en cada uno de los 28 PM distribuidos en los cuatro ríos. En la figura mencionada, se observa una notoria discontinuidad en cuanto al número de monitoreos y la tendencia decreciente a lo largo del tiempo. La tendencia es muy similar en los cuatro ríos, teniendo en común comportamientos como la falta de datos para los años 2008 y 2016, cuando la SDA no realizó la actividad de monitoreo como consecuencia de haber licitado la actividad por falta de presupuesto y vencimiento plazos en la publicación de pliegos. Por otra parte, se presenta un aumento elevado en la cantidad de monitoreos disponibles en el año 2011 donde en la gran mayoría de puntos el número de monitoreos llegó a 20, lo sigue el año 2015 cuando el número de muestras fue más discontinuo, pero un importante número de PM presenta una cantidad aproximada de 17 monitoreos.

Dicho lo anterior se encuentra muy notoria la discontinuidad en la toma de muestras a lo largo de los años, lo que da a entender que la operación de la RCHB presenta una baja eficiencia y afecta la representatividad de las muestras.

Figura 6. Número de muestras por punto de monitoreo (2006-2018)




Fuente: Autores.

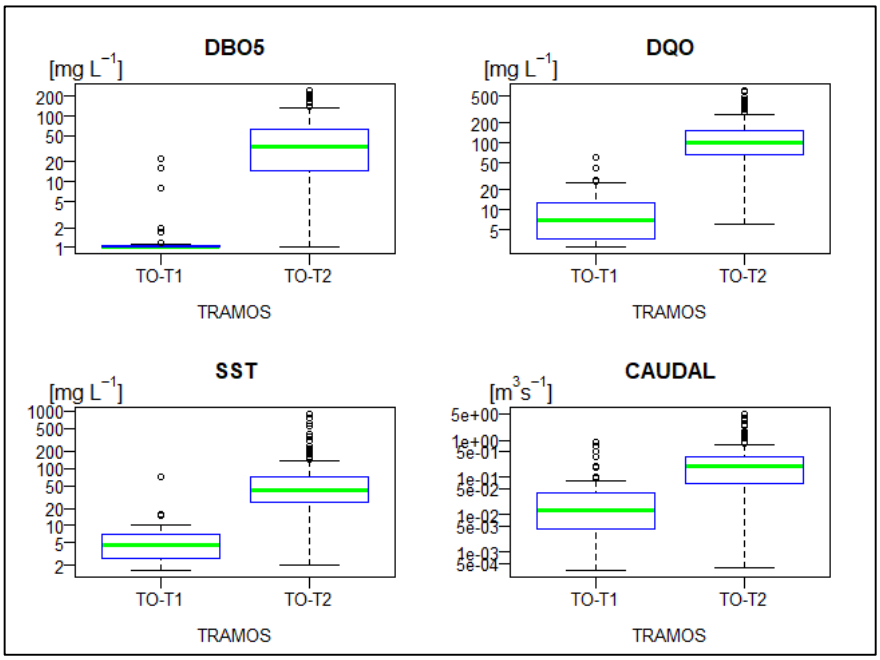
**1.2.4 Caracterización estadística de las concentraciones históricas mediante *boxplot*.** A continuación, se presentan los *boxplot* para cada río y por cada uno de los tramos que los conforman para las concentraciones históricas de DBO5, DQO, SST y caudal, para el periodo correspondiente de 2006-2015. Únicamente fueron seleccionadas estas 4 variables, ya que son parte de los principales indicadores de la calidad del agua. Esta caracterización se hace con el fin de conocer la variabilidad de las concentraciones y caudal a lo largo de cada río, desde su nacimiento hasta su desembocadura.

**1.2.4.1 Torca.** El comportamiento de las variables de calidad del agua para el canal Torca es muy similar entre ellas e incrementa del tramo 1 al tramo 2. Al comparar la mediana tanto de los determinantes de la calidad del agua y el caudal, se presenta un incremento promedio de aproximadamente 7 veces en el valor del tramo 1 al tramo 2. Por otra parte, los bigotes del diagrama del segundo tramo indican que se siguen presentando valores cercanos a los mínimos como en el primer tramo, excepto para la DQO, donde el valor mínimo es diferente.

El aumento que tienen los concentraciones de la DBO5 en el tramo 2, se debe a los puntos de vertimientos de aguas residuales localizados después del PM Calle 161 y su notoria influencia aguas abajo, donde se sigue incrementando las magnitudes de este determinante como consecuencia de la presencia de más puntos de vertimientos residuales de diferentes magnitudes (ver

Figura *2*). El caudal tiene un incremento de igual manera que las otras variables al paso de los tramos, indicando la influencia del ciclo hidrológico en cuanto a la precipitación y los aportes de los afluentes y vertimientos a lo largo del río.

Figura 7. Boxplot de las concentraciones de DBO5, DQO, SST y caudal en los tramos que conforman el canal Torca para el periodo 2006-2015




Fuente: Autores.

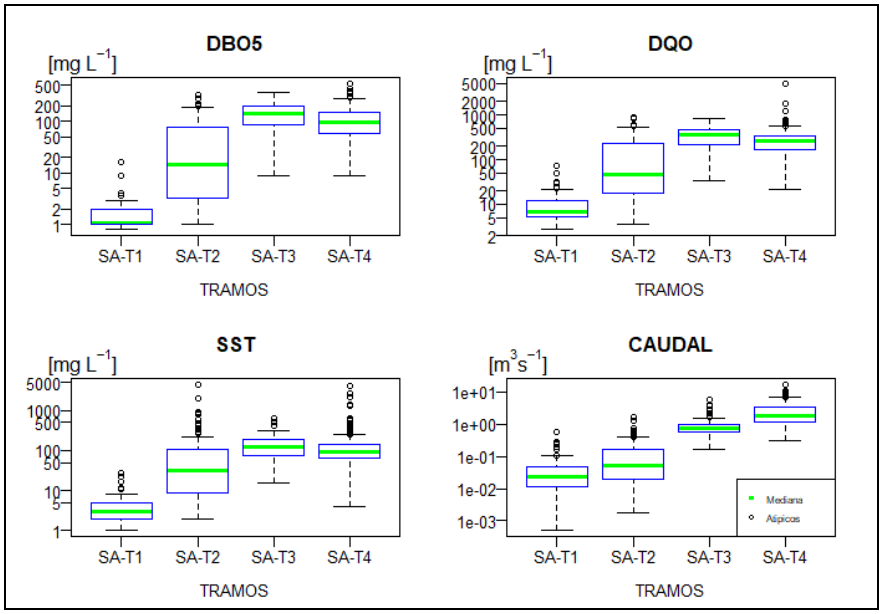
En la *ío.*

Figura *7*, se presenta los valores de concentración en mg/l en escala logarítmica en el eje vertical, mientras en el eje horizontal se muestran los tramos. Por otra parte se representa en color verde el segundo cuartil (o mediana), la línea negra horizontal superior representa el valor máximo (antes del límite superior) y la línea horizontal inferior representa el valor mínimo (antes del límite inferior) y los círculos o puntos negros representan los valores atípicos.

**1.2.4.2 Salitre.** Los *boxplot* correspondientes al río Salitre indican un comportamiento ascendente en los datos al avanzar en los tramos del río, se evidencia como la mediana tiende a incrementar, adicionalmente se puede notar que las variables DBO5, DQO Y SST en los primeros dos tramos presentan medianas muy distantes entre sí, aumentando considerablemente en más del 100 % mientras que en los tramos posteriores (SA-T3 Y SA-T4) la mediana tiende a estabilizarse comparada con tramos anteriores.

En cuanto al caudal, la mediana de los dos primeros tramos es muy similar, pero en el tercer tramo se incrementa prolongadamente hasta el último tramo, donde toma el valor más alto. En general los cuatro *boxplot* tienen un comportamiento parecido en el incremento prolongado al paso de los tramos, esto se debe a la continuidad, cantidad y composición de las descargas residuales aportadas al río (ver Figura *3*). Los valores atípicos muestran que hay una cantidad considerable de datos que son numéricamente distantes del resto de los valores.

Figura 8. Boxplot de las concentraciones de DBO5, DQO, SST y caudal en los tramos que conforman el río Salitre para el periodo 2006-2015

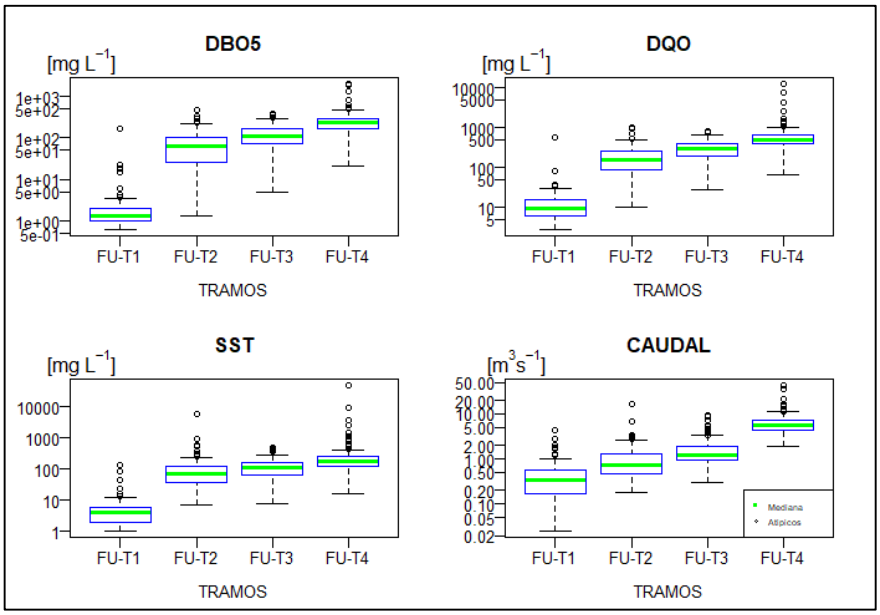


Fuente: Autores.

**1.2.4.3 Fucha.** El comportamiento de la concentración para las variables de DBO5, DQO y SST, tuvo un gran incremento del primer tramo hasta el segundo tramo, donde de ahí en adelante los valores de la mediana siguen incrementando en una menor tasa, alcanzando en el último tramo el máximo valor, esto se debe a la influencia de los vertimientos en los últimos tramos (ver *Figura 4*), los cuales afectan significativamente la calidad del agua del río. Adicionalmente, cuando los bigotes de las cajas son más extensos como en este caso indican que se presenta un mayor rango en la diferencia entre el valor menor y el mayor, indicando el cambio significativo de las condiciones iniciales y naturales del río.

El caudal incrementa de tramo en tramo progresivamente, sin tener picos tan altos en el cambio de los tramos, en el último tramo se presenta el máximo valor. Es notorio que en los cuatro *boxplot* se observa una cantidad significativa de valores atípicos, esto evidencia que para cada tramo los parámetros de medición obtuvieron que algunos valores de las muestras están alejados del resto de los datos y sus fluctuaciones, esto se puede atribuir a valores que registraron concentraciones menores o mayores al límite de detección afectadas por descargas puntuales aguas arriba o errores en campo, entre otros.

Figura 9. Boxplot de las concentraciones de DBO5, DQO, SST y caudal en los tramos que conforman el río Fucha para el periodo 2006-2015



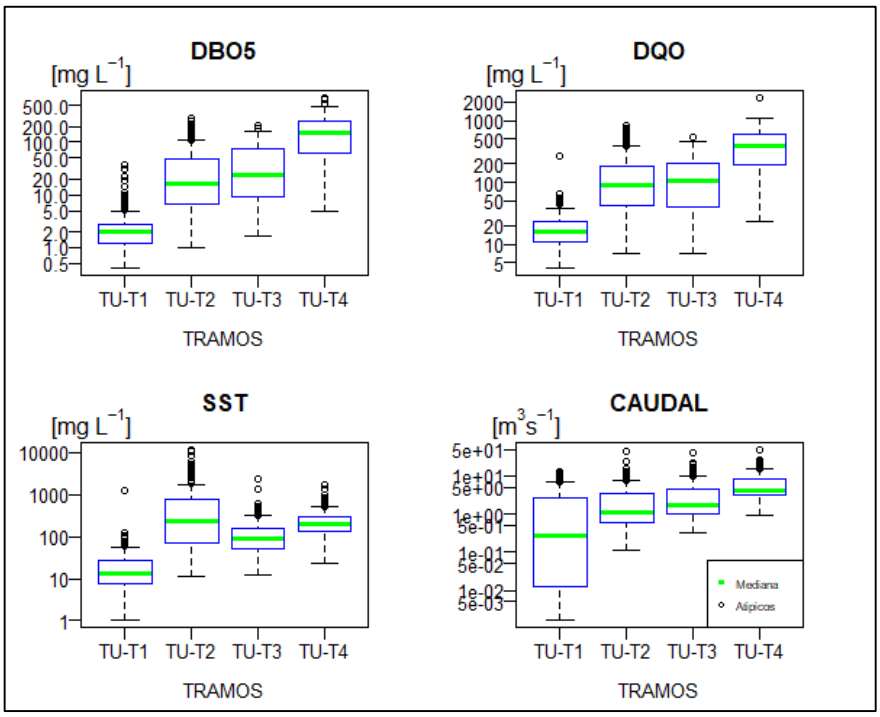

Fuente: Autores.

**1.2.4.4 Tunjuelo.** Los *boxplot* de los indicadores de calidad del agua para el río Tunjuelo representan que hay datos que se concentran en el límite inferior de los *boxplot*, siendo este el bigote con mayor longitud comparado con el límite superior, lo que indica que en todos los tramos hay una amplitud en el rango entre los valores mínimos y máximos.

Los valores de caudal indican que los valores máximos registrados para los cuatro tramos son similares entre sí, mientras que al pasar los tramos el valor mínimo va cambiando notoriamente, tomando cada vez valores más altos. El rango intercuartílico para el primer tramo es bastante amplio y muestra entre qué valores se encuentra el 50 % de los datos, este rango disminuye evidentemente al paso de los tramos y aumentan los valores de la mediana hasta el último tramo donde toma el valor máximo.

La DBO5 Y DQO inicia en el primer tramo con valores muy bajos que aumentan prolongadamente hasta los tramos 2 y 3, donde se observa que los datos tienen menor variabilidad, para luego en el tramo 4 tomar su valor máximo. Los SST tienen su valor máximo en el tramo 2, causado por los puntos de vertimientos presentes en este tramo (ver Figura *5*) y el número de quebradas que desembocan en el tramo 2 que son aproximadamente 11, es por esto que este tramo tiene una mayor influencia en el aumento de los SST con respecto a los otros 3 tramos que conforman el río Tunjuelo.

Figura 10. *Boxplot* de las concentraciones de DBO5, DQO, SST y caudal en los tramos que conforman el río Tunjuelo para el periodo 2006-2015




Fuente: Autores.

**1.2.5 Factor multiplicador.** Los factores multiplicadores son valores asignados para multiplicar y dar una magnitud específica y diferenciada a otros valores usualmente resultados de cálculos, la finalidad de establecer un factor multiplicador es asignar un respectivo peso a un valor o convertirlo en un valor deseado para la dimensión matemática que se esté usando. Es decir el factor multiplicador tiene el fin de asignar un nivel de importancia a respectivos valores o servir como un factor de conversión matemático.

Los factores multiplicadores se calculan para cada determinante de la calidad del agua como DBO5, DQO, SST, $N_{TOTAL}$, $P_{TOTAL}$ y SAAM, de acuerdo con monitoreos de 24 horas en diferentes puntos de la RCHB hechos durante el Convenio 069 de 2007 suscrito entre la SDA y la Universidad de los Andes. Por ejemplo, en la Tabla *1* se muestran los factores multiplicadores para DQO del tramo 1 del río Salitre, de esta manera si se tuviera información de un monitoreo bihorario cuya hora de inicio fuera las 7:45, se debería establecer un nuevo factor multiplicador, ya que no existe este horario en la tabla entonces, se toma el factor multiplicador (columna 3) de la hora de inicio (columna 1) de la tabla cuya diferencia sea la menor con respecto a la hora del monitoreo, y se toma factor multiplicador del rango siguiente a la hora de inicio determinada, y con base a estos dos datos se realiza la interpolación para obtener el nuevo factor multiplicador correspondiente.

Los factores multiplicadores para las variables de calidad del agua analizadas en este proyecto fueron suministrados por la SDA. Con base en los factores multiplicadores, las concentraciones y el caudal, se procedió al cálculo de CC para cada determinante de calidad del agua analizado.

Tabla 1. Ejemplo de factores multiplicadores bihorarios para los datos de DQO del tramo 1 del río Salitre

| 1 | 2 | 3 |
|---|---|---|
| Hora de inicio | Hora final | DQO Factores multiplicadores $fmRTH$ |
| 0:00 | 2:00 | 3.89 % |
| 1:00 | 3:00 | 4.49 % |
| 2:00 | 4:00 | 5.09 % |
| 3:00 | 5:00 | 7.52 % |
| 4:00 | 6:00 | 9.94 % |
| 5:00 | 7:00 | 8.45 % |
| 6:00 | 8:00 | 6.95 % |
| 7:00 | 9:00 | 5.45 % |
| 8:00 | 10:00 | 3.96 % |
| 9:00 | 11:00 | 3.35 % |
| 10:00 | 12:00 | 2.74 % |
| 11:00 | 13:00 | 7.00 % |

Fuente: Autores.

## 1.3 CÁLCULO CARGAS CONTAMINANTES

La estimación de Cargas Contaminantes (CC) es el resultado de multiplicar el caudal promedio por la concentración de la sustancia contaminante, por el factor de conversión de unidades y por el tiempo diario que dura el monitoreo, medido en horas, es decir:

Ecuación 1. Cálculo de cargas contaminantes

$$CC = Q * C * 0{,}0864 * (\frac{t}{24})$$

Fuente: [7]

Donde:

$Cc$ = Carga Contaminante, en kilogramos por día (kg/día)

$Q$ = Caudal promedio, en litros por segundo (l/s)

$C$ = Concentración de la sustancia contaminante, en miligramos por litro (mg/l)

$0{,}0864$= Factor de conversión de unidades de tiempo

$t$ = Tiempo de vertimiento del usuario en horas por día (h)

La estimación de CC aportadas en cada uno de los tramos de los cuatro principales ríos de la ciudad se viene realizando en función del comportamiento en 24 horas [8]. Esto con el fin de lograr un mejor acercamiento a la dinámica de la carga en los ríos, la cual, en algunos tramos, está influenciada por la descarga permanente de aguas residuales. Para la estimación de CC diarias se usa información secundaria, puesto que los monitoreos se realizan en ventanas temporales de dos horas [8]. La información secundaria correspondió

a monitoreos de 24 horas en diferentes puntos de la RCHB hechos durante el Convenio 069 de 2007 suscrito entre la SDA y la Universidad de los Andes [8].

Los monitoreos de 24 horas, han tenido por objeto identificar las diferencias intradiarias de concentraciones y caudal para estimar el comportamiento temporal de una variable dependiente a partir de la dinámica temporal de una o más variables independientes (temperatura, pH, conductividad), y con esto estimar factores multiplicadores para cada determinante que permitan conocer la dinámica de calidad de un día en función del muestreo realizado una hora específica [8].

## 1.4 INTERVALO DE CONFIANZA EN LOS PERFILES LONGITUDINALES DE CARGAS CONTAMINANTES CON BOOTSTRAPPING

Un intervalo de confianza (IC) es expresado en porcentaje y es una técnica de estimación utilizada en inferencia estadística que permite acotar (establecer los límites o el intervalos) en un conjunto de valores, dentro de los cuales se encontrará la estimación puntual buscada (con una determinada probabilidad), es decir un IC nos va a permitir calcular dos valores alrededor de una media muestral (uno superior y otro inferior) [9]. Estos valores van a acotar un rango dentro del cual, con una determinada probabilidad, se va a localizar el parámetro poblacional [9] .

Un IC con un nivel de confianza del 95 % no significa que la probabilidad de encontrar el parámetro de la población entre esos márgenes sea 0,95, lo que realmente significa es que si extraemos un número determinado de muestras del mismo tamaño de una población con un parámetro de valor constante, el 95 % de los IC construidos a partir de esas muestras contendrán el valor del parámetro que buscamos y el 5 % restante no lo contendrán [9].

*Bootstrapping* es un método estadístico que emplea el remuestreo como principal herramienta de análisis, fue propuesto y estudiado inicialmente por Efron (1979) como un método para la aproximación de la distribución de muestreo de un estadístico (media, mediana, correlación, *etc.*) [10]. Los métodos de remuestreo están basados en la idea de tratar a la muestra como una especie de "universo estadístico", muestreando repetidamente de esta y utilizando las muestras para estimar medias, varianzas, sesgos e IC para los parámetros de interés [10].

De esta manera para explicar el método de remuestreo (en inglés *bootstrap*), suponga una muestra X=X1, X2,…..Xn, en donde Xi, se extrae de una distribución empírica F˜ (o de una población). Las muestras de tamaño n son extraídas de x con reemplazo. Hay $n^n$ posibles muestras, llamadas muestras *bootstrap* [11]. La estimación bootstrap del error estándar es la desviación estándar de las replicaciones *bootstrap*:

Ecuación 2. Replicaciones del bootstrap

$$Las\ replicaciones\ bootstrap: \widehat{SE}_{boot} = \left\{\frac{\sum_{b=1}^{B}\left[s(x^{*b})-s(.)\right]^2}{B-1}\right\}^{\frac{1}{2}}\ Donde, s(.) = \frac{1}{B}\sum_{b=1}^{B} s(x^{*b})$$

Fuente: [11]

Existen varias maneras de calcular IC cuando se emplea el método de *bootstrapping*, en el caso puntual de las cargas calculadas para los seis determinantes de calidad hídrica se asume que la distribución de los datos no es normal y que las distribuciones *bootstrap* resultantes del proceso presentan un sesgo considerable. Bajo estas suspensiones la técnica seleccionada para realizar el cálculo del IC de las cargas fue *bootstrap* de sesgo corregido y acelerado (BCa) desarrollada por Efron en 1987, la cual se explicará a continuación:

Sea X1, X2,.....Xn las cargas calculadas para cualquier PM de la RCHB a partir de los n registros de calidad y cantidad (datos históricos y totales), el estadístico será la mediana de las cargas. Definidas las variables y parámetros el BCa inicia con la corrección por sesgo, para lo cual calcula el primer factor de corrección denominado z*:

Ecuación 3. Cálculo del primer factor de corrección (z)

$$z^* = \phi^{-1}\left\{\frac{\sum_{i=1}^{B} I\left(\hat{\theta}_i^*\right)}{B+1}\right\}\ tal\ que$$

$$I\left(\hat{\theta}_i^*\right) = \begin{cases} 1\ si\ \hat{\theta}_i^* \leq \hat{\theta} \\ 0\ si\ \hat{\theta}_i^* > \hat{\theta} \end{cases}$$

Fuente: [12]

Luego se calcula el segundo factor de corrección:

Ecuación 4. Cálculo del segundo factor de corrección

$$a = \frac{\sum_{i=1}^{B}\left(\hat{\theta}_{-i} - \theta_{-i}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{-i} - \theta_{-i}^2\right)\right\}^{3/2}}$$

Fuente: [12]

Donde $\hat{\theta}_{-i}$es la i-ésima estimación Jackniffe (eliminar una muestra del conjunto de datos una vez) de $\theta$ , y $\theta_{-i}$ es promedio de todos los $\hat{\theta}_{-i}$.

Por último, se calculan las cotas de los intervalos (superior e inferior) con nivel de significancia $\alpha$ de la siguiente manera:

Ecuación 5. Cálculo de las cotas de los intervalos (superior e inferior)

$$L_{inferior} = B * \phi\left[z^* + \frac{z^* - z_{1-\frac{\alpha}{2}}}{1 - a\left(z^* - z_{1-\frac{\alpha}{2}}\right)}\right]$$

$$U_{superior} = B * \phi\left[z^* + \frac{z^* + z_{1-\frac{\alpha}{2}}}{1 - a\left(z^* + z_{1-\frac{\alpha}{2}}\right)}\right]$$

$$\Longrightarrow L \leq \theta \leq U$$

Fuente: [13].

$1 - \alpha$ , representa el nivel de confianza, y por tanto se tomará el nivel de significancia $\alpha = 0.05$. Para que el intervalo BCa sea suficientemente confiable se recomienda que el tamaño de B se al menos 1000 [12].

## 1.5 MÉTODO MULTIVARIADO DE DISTANCIA DE MAHALANOBIS

La distancia de Mahalanobis es un criterio conocido para identificar muestras atípicas en una base de datos extensa que depende de los parámetros estimados de la distribución multivariada [14]. Éste método pretende describir la distancia entre cada punto de datos y el centro de masa [14]. Cuando un punto se encuentra en el centro de masa, la distancia de Mahalanobis es igual a cero y cuando un punto de datos se encuentra distante del centro de masa, la distancia es mayor a cero [15]. Por lo tanto, los puntos de datos que se encuentran distantes del centro de masa se considerarán valores atípicos [15].

El método de distancia de Mahalanobis se estima para cada observación en el conjunto de datos, dándole a cada observación un peso como inverso de la distancia de Mahalanobis, de acuerdo a esto las observaciones con valores extremos obtendrán menores pesos y finalmente, se ejecuta una regresión ponderada para minimizar el impacto de los valores extremos [16] .

Ecuación 6. Regresión ponderada para minimizar el impacto de los valores extremos

$$MSD_i = \sqrt{(x_i - \bar{x})^T - S_n^{-1}(x_i - \bar{x})}$$

Fuente: [14]

Donde el superíndice T denota la matriz transpuesta, *x* expresa la media del vector muestral y Sn la matriz de covarianza muestral, donde:

Ecuación 7. Cálculo de la covarianza muestral

$$Sn = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)(x_i - \bar{x})^T$$

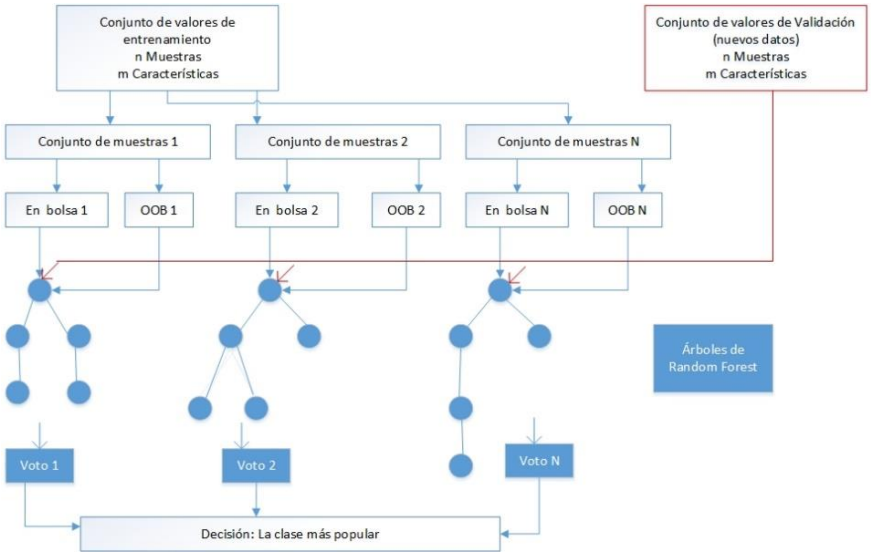
Fuente: [14]

Donde el superíndice T denota la matriz transpuesta, *x* expresa la media del vector muestral y n tamaño de la muestra.

Para datos multivariantes distribuidos normalmente, los valores de la distancia de Mahalanobis tienen aproximadamente una distribución chi-cuadrado con p grados de libertad, como resultado de ello si la distancia de Mahalanobis es grande las observaciones serán denotadas como valores atípicos [14].

## 1.6 ALGORITMO RANDOM FOREST

Figura 11. Esquema representativo del algoritmo random forest




Fuente: Autores.

*Random forest* (RF) o bosques aleatorios se encuentran entre los métodos de aprendizaje automático más populares gracias buena precisión, robustez y facilidad de uso. Este método fue desarrollado por Breiman (2001), y ha sido ampliamente usado para problemas de regresión y clasificación, pero también sirve como técnica para reducir la dimensionalidad o establecer variables relevantes dentro de un proceso [17].

En RF se ejecutan varios algoritmos de árbol de decisiones en lugar de uno solo [17]. Para clasificar un nuevo objeto basado en atributos, cada árbol da una clasificación vota por una clase y el resultado es la clase con mayor número de votos en todo el bosque (*forest*) [17]. Para regresión, se toma el promedio de las salidas (predicciones) de todos los árboles. Este

algoritmo también proporciona dos métodos sencillos para la selección de variables: disminución de la impureza media de Gini y disminución de la precisión de la media [17].

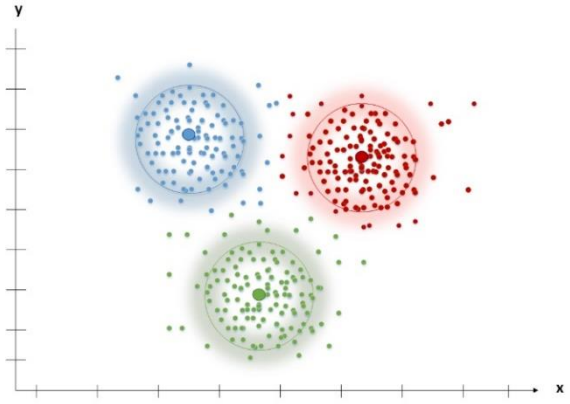
Liaw y Wiener (2002) implementaron dos algoritmos para calcular medidas de importancia de las variables usadas RF en el paquete *randomforest* en el software R, que difieren un poco de las heurísticas sugeridas originalmente por Breiman en 2003 [17]. La primera heurística se basa en el criterio de Gini, este consiste en seleccionar la variable en cada partición en la construcción de los árboles y que corresponde a una disminución de esta medida [17]. La importancia de una variable en un árbol se mide como la suma de los decrementos atribuidos a esa variable y la importancia final, como la media en todos los árboles. El promedio de todas las disminuciones en la impureza de Gini en el bosque donde se forma la división produce la medida de importancia, es decir la impureza media de Gini, *Gini Importance* o *Mean Decrease Gini* [17].

El segundo algoritmo calcula la importancia variable como la disminución media en la precisión utilizando las observaciones fuera de la bolsa (en inglés *Out-Of-Bag – OOB*) [17]. En dicho proceso se deja aproximadamente un tercio de los casos de la muestra; a los casos que no son considerados para entrenar el árbol (OOB) [18] . Con ellos se puede estimar un error insesgado de clasificación y también se pueden utilizar para hacer una estimación de la importancia de las variables [18] .

El funcionamiento de esta lógica es la siguiente: primero se escoge el error de clasificación *out-of-bag*, después se toma una variable al azar y se permutan sus valores dentro de los datos de entrenamiento, ocasionando que dicha variable escogida descorrelacione lo aprendido por el modelo [18] . Luego se vuelve a calcular el error OOB, para luego compararlo con el error calculado inicialmente [18] . En consecuencia, por lógica, si el error cambia, se afirma que dicha variable es importante [18] . Este proceso se repite con todas las variables y luego estas se ordenan de acuerdo a los cambios que produjeron cada una en los errores OOB [18] .

## 1.7 ALGORITMO DE AGRUPAMIENTO EXPECTATION MAXIMIZATION

Figura 12. Esquema representativo del algoritmo expectation maximization




Fuente: Autores.

Un modelo mixto o de mezcla es un modelo probabilístico para representar la presencia de subpoblaciones dentro de una población general, sin requerir que un conjunto de datos observados identifique la subpoblación a la que pertenece una observación individual [19].

Sin embargo, mientras que los problemas asociados con las "distribuciones de mezcla" se relacionan con derivar las propiedades de la población general de las subpoblaciones, los "modelos de mezcla" se utilizan para hacer inferencias estadísticas sobre las propiedades de las subpoblaciones dadas solo observaciones sobre población agrupada, sin información de identidad de subpoblación [19]. Uno de los métodos que permite realizar agrupamiento (*clustering*) mediante modelos de mixturas es *expectation-maximization* [19].

El cálculo de las probabilidades de las clases o los valores esperados de las clases es la parte de *expectation*. El paso de calcular los valores de los parámetros de las distribuciones es *maximization*, es decir maximizar la verosimilitud de las distribuciones dados los datos. El procedimiento consiste en definir una esperanza o expectativa en particular, y luego maximizarla, el cual es proceso iterativo, comenzando con un cierto valor inicial de los parámetros (valor semilla) que son usados para calcular las probabilidades de que cada objeto pertenezca a un *cluster* (grupo). Esas probabilidades son empleadas para re-estimar los parámetros de las probabilidades, hasta convertirse en un proceso iterativo (se puede empezar adivinando las probabilidades de que un objeto pertenezca a una clase). Los parámetros actualizados en cada iteración son los valores que maximizan la expectativa en esa iteración particular.

**1.7.1 Método BIC.** La métrica *Bayesian Information Criteria* (BIC) tiene en cuenta 14 Modelos de Mixtura Gaussianos multivariados (MMG), de esta manera un MMG es una aproximación paramétrica a una distribución de probabilidad mediante una combinación ponderada de gaussiana de componentes [20]. A menudo se usa para representar muestras de una distribución de probabilidad desconocida de forma compacta y este es un enfoque tradicional para el aprendizaje mixto gaussiano en el algoritmo de EM [20].

Un MMG con *n* mezclas es una suma ponderada de *n* densidades gaussianas individuales (mezcla de varias distribuciones gaussianas) [21]. Cada densidad Gaussiana como componente del modelo de mixtura está representada por tres partes principales: un peso de mezcla, un vector medio y una matriz de covarianza[22]. Un MMG completo está completamente definido por todas sus densidades de componentes [22]. Por lo tanto, se puede mostrar mediante un conjunto de parámetros del modelo [22]. El algoritmo EM se ejecuta luego, para varios valores del MMG y se selecciona el modelo que minimiza el criterio elegido [22]. Estos modelos son ejecutados para determinar cuál es el mejor modelo que permite realizar el agrupamiento de forma tal que las propiedades de cada *cluste*r sean particulares [22].

BIC fue propuesto por Schwarz (1978), que responde a la expresión:

Ecuación 8. Métrica Bayesian Information Criteria (BIC)

$$BIC = -2 \log L(\Psi) - k \log(n).$$

Fuente: [23]

Donde *L* es la función de verosimilitud del modelo, *k* el número de parámetros independientes del modelo de mixtura y *n* el número de observaciones independientes de la muestra univariante [23].

La selección del mejor modelo de mixtura y número de *clusters* considerando la métrica BIC, tiene en cuenta una condición adicional que, en este análisis, se define como el porcentaje de ganancia (PG) [23]. Este porcentaje se determina con respecto al mejor BIC (es decir modelo-número de *clusters*), si un modelo tiene un BIC muy similar al óptimo y su diferencia es menor al 2 %, será seleccionado para realizar la agrupación de los datos en *cluster* con características particulares [23]. Lo anterior, en general, tiende a reducir el número de *cluster* manteniendo el mismo modelo de mixtura.

La ecuación que representa el porcentaje de ganancia es la siguiente:

Ecuación 9. Ecuación que representa el porcentaje de ganancia

$$PG = \frac{BIC_i - BIC_{opt}}{\left|BIC_{opt}\right|}$$

Fuente: [23]

Donde $BIC_i$ el valor de la métrica de cada $i$ estructura modelo de mixtura-número de *clusters* y $BIC_{opt}$ es el valor de la métrica para la mejor combinación modelo de mixtura-número de *clusters* [23]. Es importante resaltar que para ambas variables de la ecuación el modelo de mixtura es el mismo, pero este puede cambiar de acuerdo con las entradas del modelo, es decir, el número de PM y la magnitud de los caudales y cargas de los determinantes de la calidad.

# 2. METODOLOGÍA

A continuación se explica y se muestra en la *Figura 13* la metodología por fases desarrollada en el presente proyecto. Todos los respectivos cálculos matemáticos aplicados a cada fase fueron realizados a través de códigos (algoritmos) escritos y ejecutados en lenguaje de programación R.

## 2.1 FASE I. CÁLCULO DE CARGAS CONTAMINANTES

El desarrollo de esta fase, consistió en realizar el cálculo de CC, donde fue necesario el uso de diferentes bases de datos como lo son: concentraciones de los determinantes de la calidad del agua DBO5, DQO, SST, SAAM, $P_{TOTAL}$ y GYA, datos de caudales para los PM en los ríos Torca, Salitre, Fucha y Tunjuelo a lo largo de periodo 2006-2018. Por otra parte, fueron empleados los factores multiplicadores asignados para cada determinante de la calidad del agua, mencionados en este párrafo.

Posteriormente, se realizó el cálculo de las CC a partir de las concentraciones, caudal y factores multiplicadores. El procedimiento de cálculo se basó en el ingreso de esta información al software RStudio, donde aplicó la fórmula de CC presentada en la *sección 1.3* **CÁLCULO CARGAS CONTAMINANTES**, que dio como resultado las tablas con los valores de CC (t-año$^{-1}$) para cada monitoreo existen en el periodo de análisis de cada PM de los ríos evaluados.

A partir de la CC calculada fueron realizados los perfiles longitudinales de CC para cada uno de los conjuntos de datos: el primero se denomina datos históricos (2006-2015) y segundo datos actuales (2016-2018), con los datos de cada PM y en los dos periodos señalados se calcularon los IC (95 %) por medio de la técnica *bootstrapping* (ver *sección* 1.5 **MÉTODO MULTIVARIADO DE DISTANCIA DE MAHALANOBIS**) y del intervalo fue calculada la mediana de las CC de cada PM para los periodos evaluados, como se presentan en la *sección 3.1* **PERFILES LONGITUDINALES DE CARGA HISTÓRICA CONTAMINANTE**

## 2.2 FASE II. DETECCIÓN DE DATOS ATÍPICOS MEDIANTE EL MÉTODO DE DISTANCIA DE MAHALANOBIS

Con las CC resultante de la fase anterior donde se encuentra incluido todo el periodo de estudio (2006-2018), se aplicó el método de distancia de Mahalanobis para detección de muestras atípicas desde un enfoque multivariado, que incluyó además de las CC el caudal, aunque cabe aclarar que fue excluido el determinante de calidad $N_{TOTAL}$ debido a que este presentaba gran cantidad de datos faltantes, superando el 50 %, lo que el lenguaje de programación R lo traduce a datos NA (*No Available*). Mediante el paquete *mvoutlier*

disponible para R que dentro de sus algoritmos se encuentra el método distancia de Mahalanobis, fueron identificadas las muestras atípicas por PM, que se representaron en un gráfico de dispersión multivariado. Adicionalmente, se obtuvo una tabla con los porcentajes de muestras atípicas para cada PM, tal como se observa en la sección **3.2 GRÁFICOS DE DISPERSIÓN MULTIVARIADOS DE CARGA CONTAMINANTE POR MÉTODO MAHALANOBIS**

## 2.3 FASE III. SELECCIÓN DE VARIABLES A TRAVÉS DEL ALGORITMO RANDOM FOREST

Después de detectar y eliminar las muestras atípicas de cada PM, se consolidaron nuevas tablas con los datos de las CC de DBO5, DQO, SST, SAAM, $P_{TOTAL}$ y GYA y el caudal, se procedió a implementar el algoritmo RF (ver *sección 1.6* **ALGORITMO RANDOM FOREST**) para la selección de las variables que mejor representaran la dinámica hídrica en cada PM.

El algoritmo RF ha sido vinculado al paquete en R denominado *random forest*. Como primer paso RF creó múltiples muestras aleatorias (seleccionadas aleatoriamente del conjunto de datos con la técnica de remuestreo), posteriormente realizó una selección aleatoria de las características de entrada para así adoptar la impureza de Gini con el fin de decidir el criterio de división de los datos en diferentes clases de manera homogénea minimizando la clasificación errónea en la selección de variables, se debe tener en cuenta que en este paso el error de clasificación OOB estima un error insesgado de clasificación rectificando los resultados con el conjunto de datos de validación (⅓ de los datos que no se usan para el conjunto de entrenamiento y pueden ser usados como test de verificación de errores). Este proceso se realizó 1000 veces (1000 árboles de decisiones) hasta que los nodos terminales cuenten con una sola clase o su tamaño sea mínimo.

Como último paso del algoritmo determinó el número de votos por la clase más popular para cada combinación de PM de la siguiente manera, 1-2, 2-3, 3-4… etc. dependiendo su orden de ubicación (desde el nacimiento hasta la desembocadura) y el número PM por río.

*El algoritmo RF dio como un resultado una tabla de importancia para cada río con magnitudes o votos de la magnitudes o votos de la variable más popular. La primera tabla resultante tuvo en cuenta todas las variables todas las variables de calidad del agua que están siendo consideradas en el análisis (DBO5, DQO, SST, SAAM, DQO, SST, SAAM, $P_{TOTAL}$, GYA y caudal), esta tabla es denominada de votos general (ver*

**Anexo F**. Variables seleccionadas por cada pareja de puntos de monitoreo por río de acuerdo con el Gini index general *De acuerdo con lo anterior, dado que algunas magnitudes arrojadas por el algoritmo expresadas en la tabla de votos general son muy similares entre variables la probabilidad de que una variable enmascare la relevancia con respecto a la otra es muy grande, por lo tanto con las tres variables más importantes para cada combinación de PM de la tabla de votos general se aplicó nuevamente el algoritmo con el fin de determinar con más veracidad el orden de importancia de cada variable según la magnitud, generando una tabla que corresponde a las tres variables más importantes ya desenmascaradas para la combinación de PM correspondiente por río, denominada tabla de votos específica (ver*

**Anexo *G***).

Es importante resaltar que hasta este punto (tabla de votos específica) solo se habían elegido las tres variables más importantes para cada combinación de PM de cada río, pero no del río en general. Con esta segunda tabla de votos específica se procedió a determinar las variables con mayor grado de importancia para el río en general por lo tanto se consideró aplicar una metodología propia de esta investigación donde se determinarán las cuatro variables más importantes para cada río (eso debido a la similitud en la magnitud en la tercer y cuarta variable en orden de importancia). Ya que los valores arrojados en la tabla de votos específicos siguen siendo muy similares tanto para el primer, segundo y tercer nivel de importancia, el primer paso de esta metodología fue asignar un factor multiplicador según el nivel de importancia que se puede ver en la Tabla *2*.

Tabla 2. Factor multiplicador de acuerdo al nivel de importancia

| Nivel de importancia | Factor multiplicador |
|---|---|
| 1 | 1 |
| 2 | 0.5 |
| 3 | 0.25 |

Fuente: Autores.

Con los resultados obtenidos se realizó una sumatoria de los valores correspondientes a cada variable, donde se obtuvo el orden de importancia de las variables para cada río en general, siendo las más importantes las que más se repiten o se encuentran en los primeros niveles de importancia, de acuerdo con lo anterior se eligieron las cuatro variables con mayor magnitud en el resultado de la sumatoria para cada río las cuales representan de mejor manera la dinámica de calidad y cantidad del agua del río. Por último, se generaron gráficos en forma de red para cada río, los cuales representan gráficamente cuales son las variables más importantes de acuerdo a su magnitud.

*El desarrollo de la metodología propuesta para determinar los cuatro determinantes de la calidad con mayor calidad con mayor importancia basados en el algoritmo RF se encuentra ubicado en el*

**Anexo *H***.

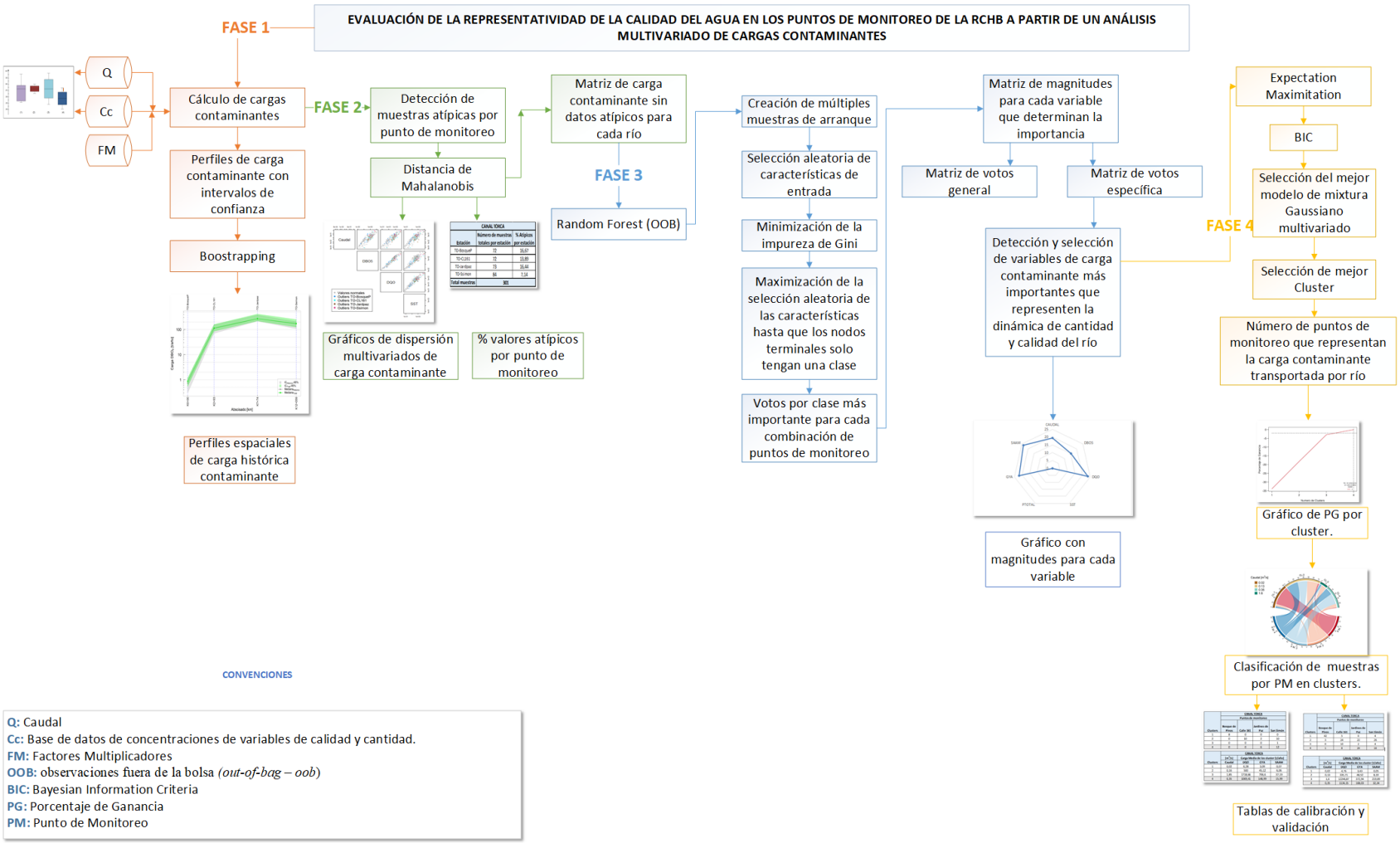
## 2.4 FASE IV. SELECCIÓN DEL NÚMERO DEL *CLUSTERS* ÓPTIMO CON BASE EN EL ALGORITMO EXPECTATION MAXIMIZATION

Para el desarrollo de este algoritmo se empleó el paquete *Mclaust* en la plataforma de R Studio, el cual contiene 14 modelos de mixtura Gaussianos multivariados, cuyas entradas fueron el número de PM, los valores de caudales y las cargas de las variables de la calidad,

además los parámetros de los modelos fueron ajustados con los datos del periodo de calibración (2006-2013) y probados con el periodo de validación (2014-2018). Se tuvieron en cuenta en este análisis las 4 variables más importantes por río de acuerdo con su magnitud obtenida en el desarrollo del algoritmo RF.

Luego se seleccionó el mejor modelo de mixtura Gaussiano de acuerdo con la métrica *Bayesian Information Criteria* (BIC) y se eligió el mejor *cluster*, si su diferencia es menor al 2 % con los demás *clusters* (porcentaje de ganancia), obteniendo como resultado los gráficos de tendencia de porcentaje de ganancia y número óptimo de *clusters* de acuerdo con la métrica BIC como se puede observar en la *sección 3.2* **GRÁFICOS DE DISPERSIÓN MULTIVARIADOS DE CARGA CONTAMINANTE POR MÉTODO MAHALANOBIS**. Finalmente, se generaron los gráficos circulares, donde se clasificaron las muestras del periodo de calibración por PM en el número *clusters* identificados anteriormente, determinados por el algoritmo EM por cada río como se puede ver en la *sección 3.4* **ANÁLISIS CLUSTER MEDIANTE EL ALGORITMO EXPECTATION MAXIMIZATION** .

Figura 13. Esquema de la metodología del proyecto por fases




Fuente: Autores.

# 3. RESULTADOS

## 3.1 PERFILES LONGITUDINALES DE CARGA HISTÓRICA CONTAMINANTE

A continuación, se presentan los perfiles longitudinales de CC desde el primer hasta el último PM de cada uno de los ríos de la RCHB. Los perfiles representan el comportamiento para dos conjuntos de datos: Histórico que corresponde a los datos del periodo del año 2006 a 2016 y Total que incluye los datos históricos más el periodo 2017-2018. A partir de estos conjuntos de datos se determinó el comportamiento de la CC en un Intervalo de Confianza (IC) y la mediana en cada PM, con el objetivo de evidenciar los cambios y las alteraciones en la calidad del agua en los diferentes puntos ubicados espacialmente en el cauce de cada uno de los ríos a estudiar desde que entró en funcionamiento la RCHB.

Los perfiles muestran en el eje horizontal los PM para cada río marcados con el abscisado en km perteneciente a cada PM iniciando en el nacimiento del río (abscisa K+00) hasta su desembocadura, y para el eje vertical se presentan las CC de DBO5, DQO y SST, respectivamente.

La banda de color gris claro representa el IC al 95 % y la línea del mismo color muestra la mediana para las cargas transportadas en el periodo 2006 a 2016 (Histórico), mientras la otra banda de color (DBO5 (color verde), DQO (color azul) y SST (color café)) representa el IC al 95 % para el periodo histórico más los años 2017-2018 (Total) y la línea del mismo color es la mediana para las cargas transportadas. Los perfiles espaciales de las demás variables de calidad (GYA, $N_{TOTAL}$, $P_{TOTAL}$ y SAAM) se pueden observar en el **Anexo *C***.

Los valores de CC asociados a los vertimientos que se utilizan en los siguientes análisis fueron tomados del PSMV (Plan de Saneamiento y Manejo de Vertimientos), donde se caracterizan los diferentes vertimientos que influyen en cada uno de los ríos. Estos valores fueron usados con el fin de fundamentar el comportamiento de las CC presentado en los perfiles longitudinales para los 4 ríos.

### 3.1.1 Canal Torca.

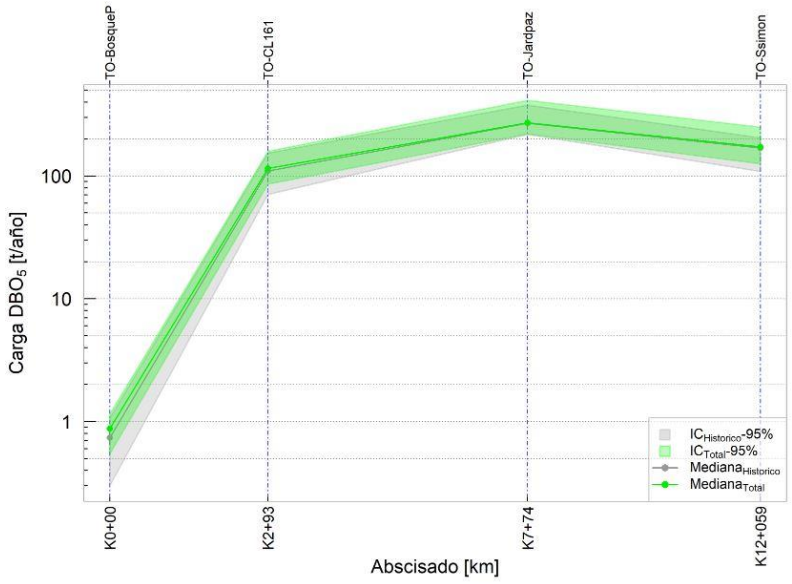
**3.1.1.1 Perfil longitudinal de la carga histórica y total de la DBO5*.*** En la *Figura 14* se puede apreciar el incremento de la carga de la DBO5 en el transcurso del río, lo cual indica que a partir del PM CL161 se hace evidente una o varias fuentes de contaminación, que se presentaron aguas arriba (ver *Figura 2)* hasta hacerse evidente en este PM causando disminución del nivel de oxígeno en el cuerpo de agua y limitando el desarrollo de microorganismos (macroinvertebrados) y aumentando la demanda de oxígeno al consumir o degradar estos contaminantes. La DBO5 obtiene su máximo valor en el punto Jardpaz y su mínimo valor en el primer PM, evidenciando la alteración de las condiciones naturales del río.

Esta alteración de las condiciones naturales del río se da por causa de vertimientos de agua residuales, principalmente, que se encuentran localizados en ambas márgenes del canal

Torca. La SDA tiene registros de las CC de estos vertimientos, por ejemplo, como se observa en la Figura *2* entre los PM BosqueP y CL161 correspondientes al tramo 1, se presenta un aumento significativo, que es generado por dos puntos de vertimientos con un valor máximo registrado de CC de 299.715 t-año$^{-1}$ para DBO5. El comportamiento de la CC DBO5 hasta el siguiente PM tiene un una variación de mayor magnitud llegando a su pico más alto en el PM Jardpaz, donde comienza a descender hasta el último PM Ssimon.

De acuerdo con las bandas de confianza se puede observar que el IC del periodo histórico es más amplio con tendencia a valores más bajos comparados con el IC del período de años total, en especial entre los PM BosqueP y CL161. El aumento de la variabilidad de la CC permanece igual para los PM CL161, Jardpaz y Ssimon, sin embargo se presenta un aumento en la variabilidad de CC en el PM BosqueP para el periodo de tiempo total con respecto al histórico, lo cual se puede observar en la línea que representa la mediana en color verde.

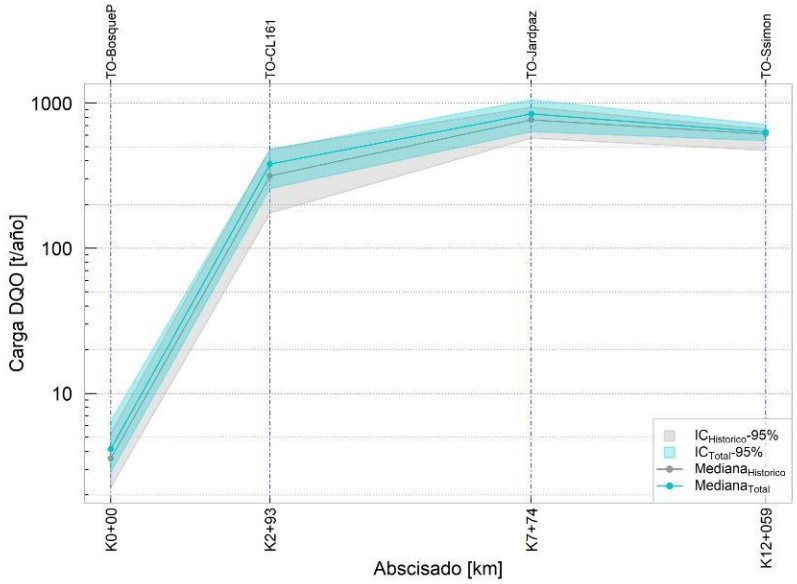
Figura 14. Perfil longitudinal de carga contaminante histórica y total de la DBO5 para el canal Torca




Fuente: Autores.

**4.1.1.2 Perfil longitudinal de la carga histórica y total de la DQO.** Con respecto a la DQO a lo largo de los PM presentes en el canal Torca (ver
*Figura 15*) se evidencia un incremento en los tres primeros PM, del orden 100 t-año$^{-1}$ en el PM Cl161 y Jardpaz. En Ssimon las medianas son similares, pero la tendencia en el periodo total es a mayores CC tanto en el límite inferior como superior de la banda de confianza. Es decir que se presenta un aumento en la presencia de compuestos inorgánicos que demandan oxígeno del agua, también se observa que el punto de Jardpaz es el punto donde siempre se ha presentado el mayor valor de las cargas, disminuyendo ligeramente en el PM de Ssimón dado que como se muestra en la Figura *2* entre el tramo Jardpaz y Ssimón no se presenta ningún vertimiento y se evidencia en una leve disminución en la carga entre estos dos puntos. Además, se puede observar un aumento importante de las cargas a lo largo del río, entre el tramo BosqueP y el tramo CL161, así el primer tramo representa las CC más bajas de la DQO.

Figura 15. Perfil longitudinal de carga contaminante histórica y total de la DQO para el canal Torca




Fuente: Autores.

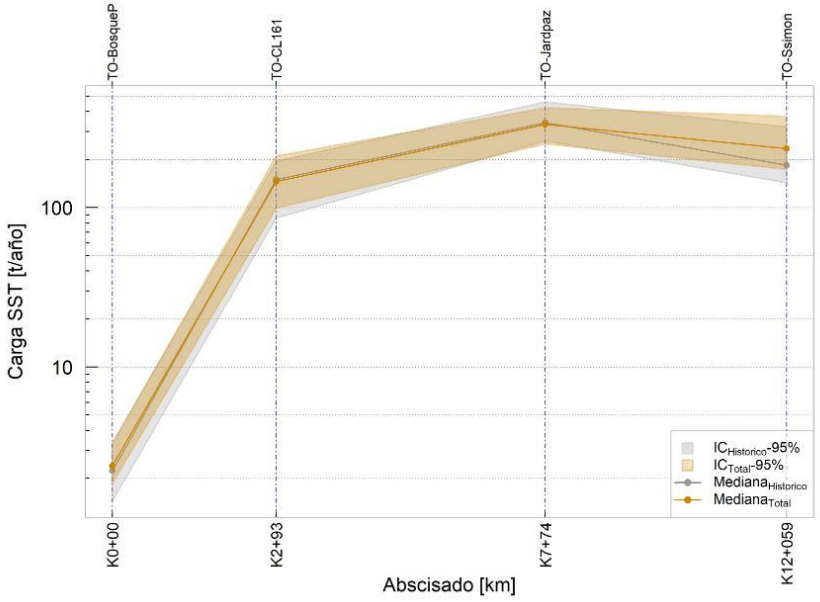
**3.1.1.3 perfil longitudinal de la carga histórica y total de los SST.** Los SST a lo largo del canal Torca han permanecido muy similares para los dos periodos de estudio, tal como se puede observar en la medianas y límite superior de las bandas confianza de los tres primeros PM (ver

Figura *16*), sin embargo, en el PM Ssimon los valores de la mediana, límite superior e inferior de la banda de confianza de la CC del conjunto de datos Total es mayor a los valores reportados para el conjunto de datos Históricos en aproximadamente 80 t-año$^{-1}$, posiblemente dado por la reducción en los últimos años de la capacidad del humedal Torca-Guaymaral en tratar los SST (pérdida de capacidad hidráulica) debido a la reducción de la pendiente lo que hace que se acumule más agua y genera aumento de sedimentos [4]. Este aumento se hace más notorio a partir del PM CL161, donde aguas arriba en los vertimientos registrados se aporta un máximo de 134 t-año$^{-1}$, lo que genera un aumento notorio en la carga de los SST, indicando una perturbación del medio por causa de diferentes contenidos químicos e impurezas de los vertimientos, y así esto resulta en diversos casos en material residual sólido que posteriormente se suspende en el cuerpo hídrico o se deposita en el fondo.

El comportamiento del perfil muestra que el sector entre los PM CL161 y JardPaz tiene un alto impacto por las descargas residuales, y hacia aguas abajo el río encuentra una manera de amortiguar o diluir el efecto de las altas CC de los SST mediante el humedal Torca-Guaymaral, dando como resultado una ligera disminución entre los dos últimos PM [4]. Adicionalmente en este caso observamos que los datos históricos (2006-2016) presentaron cargas más bajas entre estos puntos, lo que indica que los datos totales (2006-2018)

tuvieron este incremento por la influencia de los datos actuales (2017-2018) determinando cambios en las cargas.

Figura 16. Perfil longitudinal de carga contaminante histórica y total de la SST para el canal Torca




Fuente: Autores.

### 3.1.2 Río Salitre.

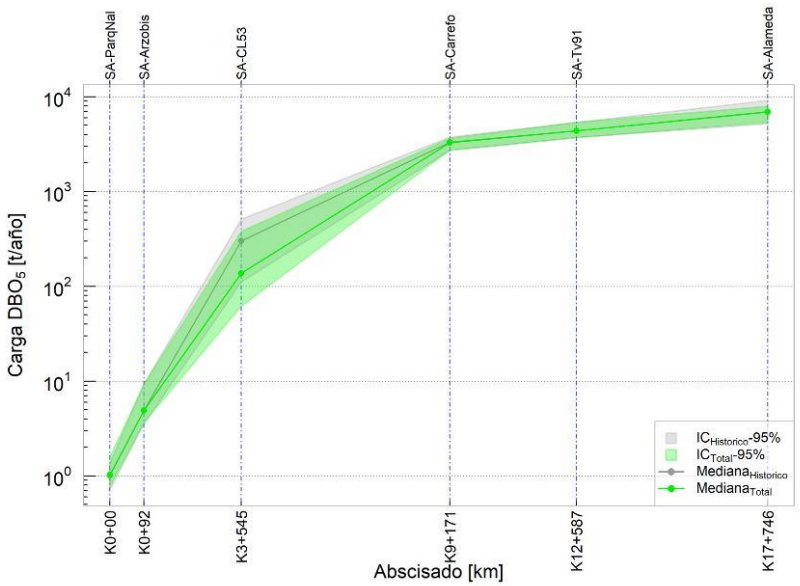
**3.1.2.1 Perfil longitudinal de la carga histórica y total de la DBO5.** La variación en la calidad del agua del río Salitre con respecto a la CC de la DBO5, se da por el aporte de los diferentes vertimientos presentes en las márgenes del río. Como se observa en la *Figura 17* la primera alteración significativa se presenta entre el punto ParqNal y Arzobis que es originada por un punto de vertimiento que se puede evidenciar en la *Figura 3* justo antes del PM Arzobis, que aporta una CC igual a 1,534.13 t-año$^{-1}$ DBO5 siendo el aporte más grande de CC registrado en todos los tramos del río salitre, ya con este aporte de CC en el registro de los próximos PM se observa un aumento prolongado a lo largo del río debido a la influencia de los vertimientos en todos los PM del río, lo cual pone en evidencia el alto impacto de los vertimientos una vez el río entra en el perímetro urbano y la condición con la que desemboca en el río Bogotá.

Con respecto a los IC, para el tramo 2 (puntos CL53 y Carrefo), las cargas oscilan en un rango mayor y de acuerdo con la mediana los datos total de CC en CL53 tienden a 150 t-año$^{-1}$ y el IC tiende a ser más amplio, pero con magnitudes menores en sus límites con respecto a los datos históricos. La oscilación de estos datos de CC en Cl53 se debe a los cambios en los caudales asociados a la hidrología aguas arriba del río. Después del PM

CL53 se observa la influencia de colectores como Sears, Delicias…etc, que se hacen evidentes en el PM Carrefo.

Además de acuerdo a la línea de tendencia entre estos puntos se puede decir que los valores históricos son mayores que los totales, esto quiere decir que los valores actuales 2017-2018 incidieron en la disminución de estos valores. En cuanto a la mediana de la CC es notable un disminución en la mediana total con respecto a la histórica en el PM SA-CL53, esto se debe a que este punto tuvo un cambio de ubicación unos kilómetros aguas abajo.

Figura 17. Perfil longitudinal de carga contaminante histórica y total de la DBO5 para el río Salitre



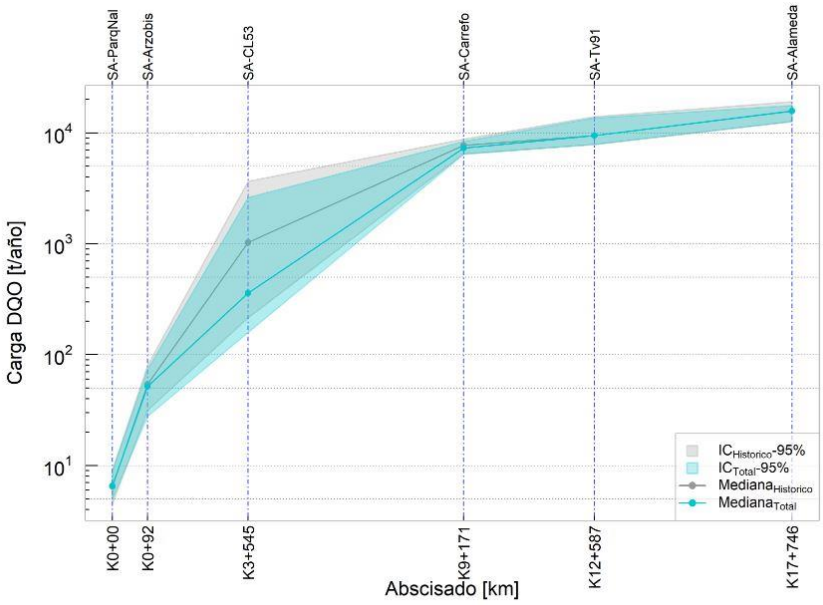

Fuente: Autores.

**3.1.2.2 Perfil longitudinal de la carga histórica y total de la DQO.** En el siguiente perfil se muestran las cargas de la DQO en el río Salitre (ver

Figura *18*), es notorio el cambio que se da entre el tramo ParqNal y Arzobis, ya que allí comienzan las primeras descargas residuales (ver *Figura 3*), que alteran las condiciones naturales del río no solo en estos puntos, sino de ahí en adelante, así aunque en el primer punto resultó el menor valor de la DQO, de ahí en adelante se presentarán los valores más altos, posteriormente entre el punto Carrefo y Alameda las cargas son menos variables, respecto a los anteriores valores presentados, donde se observa un amplio IC, que indica una alta fluctuación en los valor de CC.

Adicionalmente el IC histórico tiende a tener valores más altos, mientras que el IC total tiende a presentar valores más bajos comparados con el histórico, que es notorio en el tramo desde Arzobis hasta Carrefo, esto está influenciado por el cambio de ubicación en el PM CL53 aguas abajo.

Figura 18. Perfil longitudinal de carga contaminante histórica y total de la DQO para el río Salitre

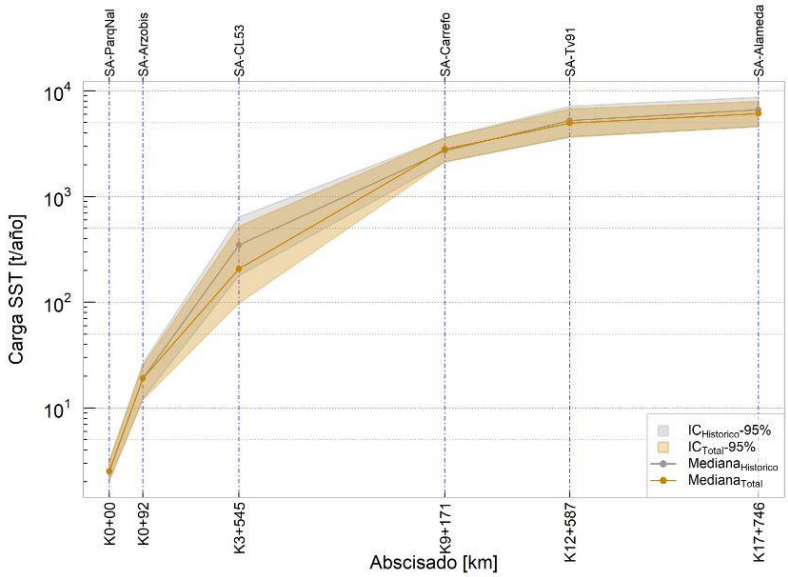



Fuente: Autores.

**3.1.2.3 Perfil longitudinal de la carga histórica y total de los SST.** El incremento desde el primer PM deja en evidencia el alto impacto de las descargas de aguas residuales en la CC de sólidos en el río (ver
*Figura 19)*, lo que se puede constatar con la presencia del primer vertimiento ubicado entre los PM SA-ParqNal y SA-Arzobis (ver *Figura 3*), que aporta en total un valor de CC igual a 781.73 t-año$^{-1}$de SST, generando una incremento gradual hasta el valor máximo registrado en el punto SA-Alameda.

Entre el punto Carrefo y Alameda los valores de los SST tienden a estabilizarse, comparado con los valores registrados en los primeros PM. El aumento de la CC de los SST está dado por la cantidad de puntos de descarga de aguas residuales a lo largo del río y está directamente relacionado con la composición de dichas descargas.

Figura 19. Perfil longitudinal de carga contaminante histórica y total de la SST para el río Salitre




Fuente: Autores.

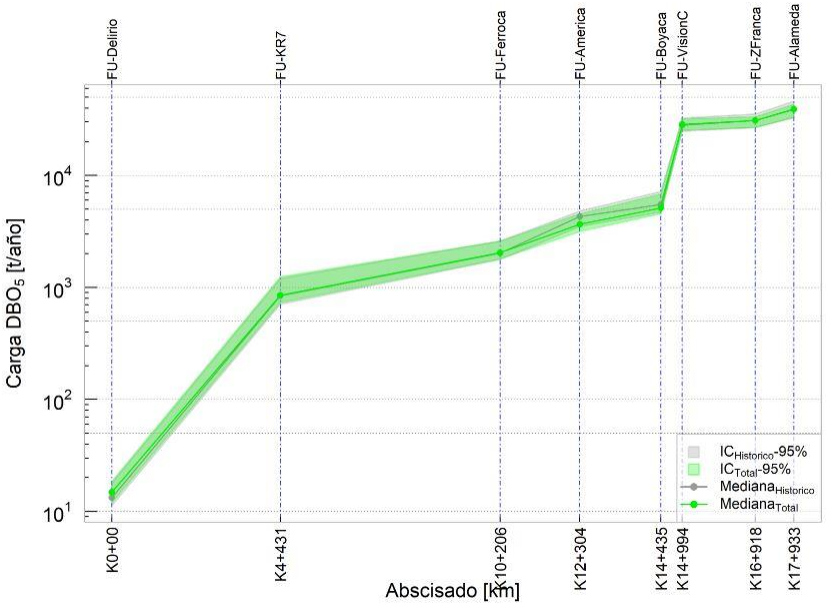
### 3.1.3 Río Fucha.

**3.1.3.1 Perfil longitudinal de la carga histórica y total de la DBO5.** La carga de la DBO5 aumenta escalonadamente a lo largo del río iniciando entre Delirio y Kr7 con un incremento de 1000 t-año$^{-1}$. De todos los PM del río Fucha se observa un comportamiento similar de las CC (medianas y los IC) en ambos conjuntos de datos, con excepción America y Boyaca donde conjunto de datos Total presenta menores CC en la mediana y el límite superior del IC, evidenciado reducciones de la CC para el periodo 2017-2018.

*Aunque en todos los tramos del río hay presencia de vertimientos con cargas de la DBO5 significativas, entre significativas, entre los puntos Boyaca y VisionC surge un aumento con una pendiente elevada debido a un elevada debido a un aporte de una descarga de 43,257.84 t-año$^{-1}$ de la DBO5 elevando notoriamente los notoriamente los valores que son registrados por el PM VisionC, como se evidencia en la*

Figura *20*. También se puede analizar que, entre los puntos VisionC y ZFranca no se observan cambios notorios en los valores de la DBO5, por el contrario, los valores tienden a ser constantes, lo que podría indicar que se puede descartar un PM.

Figura 20. Perfil longitudinal de carga contaminante histórica y total de la DBO5 para el río Fucha



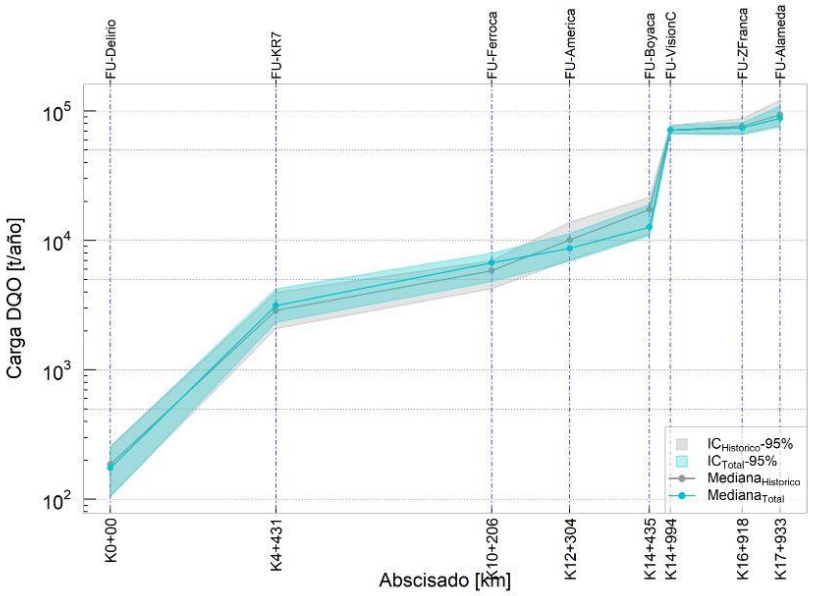

Fuente: Autores.

### 3.1.3.2 Perfil longitudinal de la carga histórica y total de la DQO. La DQO en la

*Figura 21* muestra que en el primer tramo presenta un comportamiento similar en ambos conjuntos de datos evaluados y al observado en la DBO5 y los SST. No obstante, desde el PM KR7 se observa un incremento en la CC de la DQO hasta el PM Ferroca en el periodo 2017-2018 (conjunto de datos total) con respecto a los datos históricos, donde se concentran gran cantidad de puntos de vertimientos y esto hace que cambien notoriamente las condiciones naturales de calidad del agua del río. Entre los PM America y Boyaca los valores de la mediana de datos total son menores con respecto a la mediana de datos histórica, mientras que en los PM VisionC y ZFranca se presenta una tendencia constante de los valores de CC para los dos periodos de estudio.

El IC entre Boyacá y VisionC fue muy reducido, lo que indica que este es un conjunto de datos más homogéneo comparado con los de otros PM. En la base de datos del periodo de tiempo total se presenta una ligera variación en la CC con respecto al periodo de tiempo histórico, donde del PM KR7 a Ferroca aumenta, lo que puede ser evidenciado en la mediana para las dos bases de datos, esto pudo ser causado por un aumento de las descargas de vertimientos entre los años 2017-2018 o condiciones ambientales que propiciaron dichas variaciones.

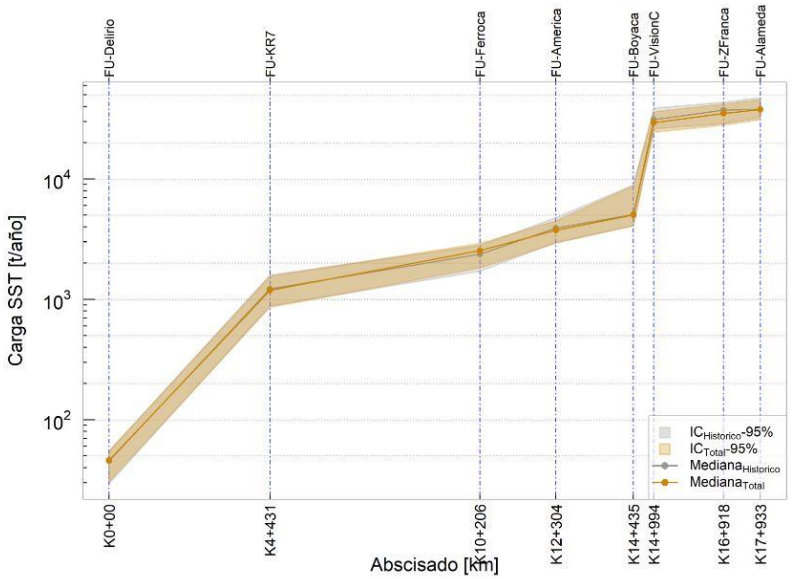
Figura 21. Perfil longitudinal de carga contaminante histórica y total de la DQO para el río Fucha




Fuente: Autores.

**3.1.3.3 Perfil longitudinal de la carga histórica y total de los SST.** El perfil longitudinal de la CC de los SST (ver *Figura 22*) presenta la mayor pendiente o aumento entre los primeros PM (Delirio y KR7), también otro tramo con una pendiente elevada se da entre los PM de Boyacá y VisionC donde se registra un vertimiento de 36,412.71 t-año$^{-1}$ este incremento es aún más notorio teniendo en cuenta que estos PM están separados por tan solo 559 metros, así a lo largo de todo el río se presenta un crecimiento en la CC, estas condiciones son similares para cada base datos tanto como para carga histórica y para carga total, es decir, presenta similitud con la variabilidad de la DBO5.

Figura 22. Perfil longitudinal de carga contaminante histórica y total de la SST para el río Fucha




Fuente: Autores.

### 3.1.4 Río Tunjuelo.

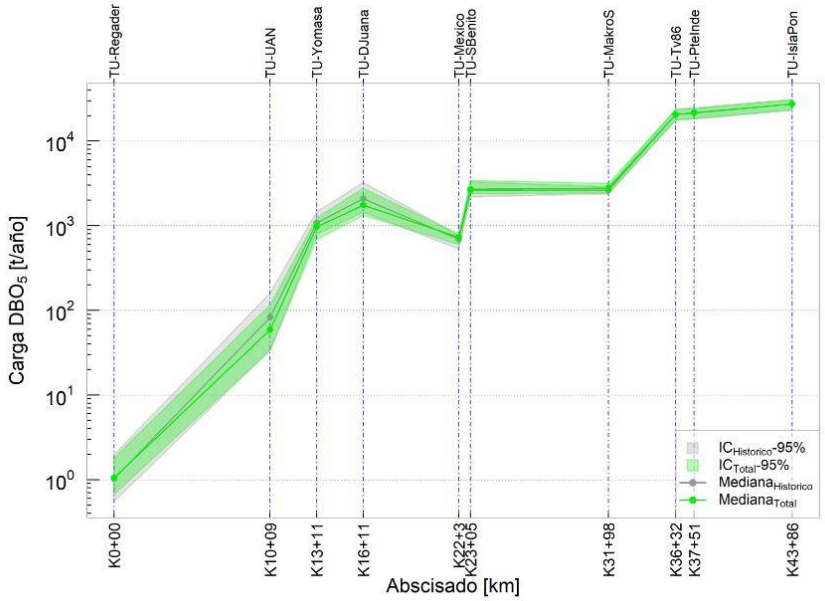
**3.1.4.1 Perfil espacial de la carga histórica y total de la DBO5.** En la

*Figura 23* se muestra la variabilidad en la carga de la DBO5 para el río Tunjuelo, donde se presenta una amplia inestabilidad en la CC. En los primeros PM (Regader a DJuana) hay un ascenso permanente, pero con notorios cambios de pendiente debido a la variabilidad de los valores pertenecientes a los diferentes vertimientos que descargan entre estos puntos. El vertimiento registrado con mayor CC de la DBO5 aporta un total de 18,275.92 t-año$^{-1}$ y se encuentra ubicado entre los PM MakroS y Tv86 y coincide con la alta pendiente que se observa entre los puntos mencionados (ver

*Figura 23*). Posterior al PM Tv86 los vertimientos existentes presentan disminución en la magnitud de CC las cuales oscilan en un rango de 1,155.62 a 6,936.08 t-año$^{-1}$, lo que genera que su comportamiento tenga menos fluctuaciones en los últimos dos PM.

La mediana y los IC muestran una disminución en el conjunto de datos total en los PM iniciales Regader a DJuana esto comparado con la mediana del conjunto de datos histórico, lo cual indica que para los datos perteneciente al periodo de tiempo de 2017-2018 hay un descenso en la CC de la DBO5.

Figura 23. Perfil longitudinal de carga contaminante histórica y total de la DBO5 para el río Tunjuelo



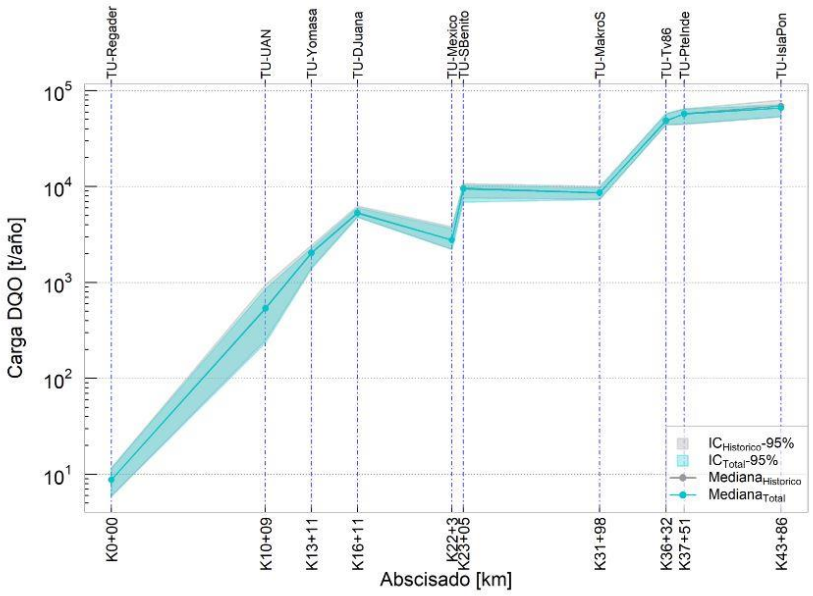

Fuente: Autores.

**3.1.4.2 Perfil espacial de la carga histórica y total de la DQO.** *El comportamiento de las cargas históricas y cargas históricas y totales para la DQO (ver*

Figura *24*) es muy similar al descrito anteriormente para la carga de la DBO5, aunque en este caso no se tienen en cuenta los valores de los vertimientos del PSMV (no se encuentran disponibles para esta variable), esto nos indica que los vertimientos que desembocan en este cauce presentan una similitud en cuanto a la CC de la DBO5 como de la DQO.

La mediana histórica y total presenta un semejante comportamiento a largo del río, igualmente pasa para la banda de confianza tanto histórico como total, indicando que los valores para las dos bases de datos se mueven en un rango reducido.

Figura 24. Perfil longitudinal de carga contaminante histórica y total de la DQO para el río Tunjuelo

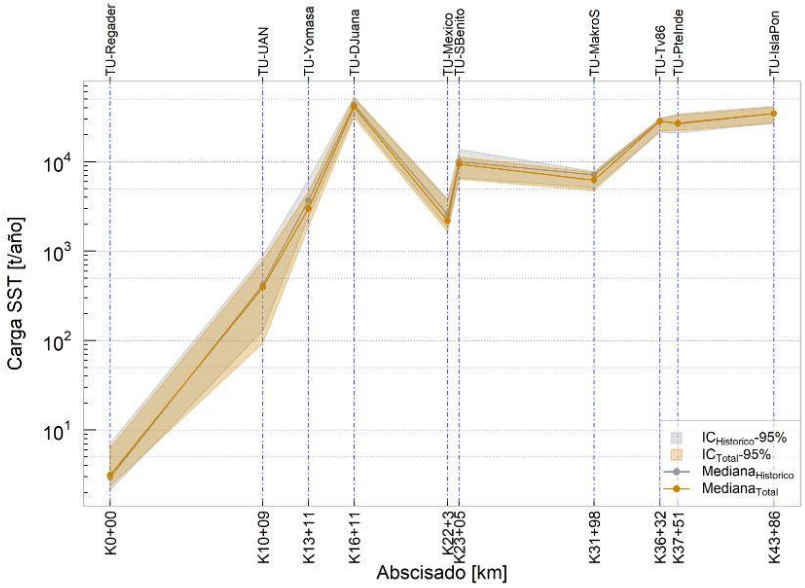



Fuente: Autores.

**4.1.4.3 Perfil espacial de la carga histórica y total de los SST.** Al igual que para los perfiles longitudinales de CC para la DBO5 y la DQO, el comportamiento de la CC de los SST es muy similar (ver
*Figura* 25), aunque en este caso se presentan valores más elevados, como se evidencia en el punto DJuana donde se presenta un pico elevado para la carga de los SST, tanto para cargas históricas como para cargas totales, lo que quiere decir que los vertimientos aguas arriba a este PM agregan en gran magnitud CC asociados a la actividad de aprovechamiento pétreo [8]. En general los valores de CC de los SST aportados por los vertimientos oscilaron entre 0,0483 t-año$^{-1}$ a 12,479.42 t-año$^{-1}$.

De igual forma se observa una leve disminución en cuanto a la mediana de la carga total comparada con la carga histórica, esta disminución es más notoria en el tramo medio del río Fucha, iniciando en el PM punto Yomasa hasta el PM MakroS.

Figura 25. Perfil longitudinal de carga contaminante histórica y total de la SST para el río Tunjuelo




Fuente: Autores.

## 3.2 GRÁFICOS DE DISPERSIÓN MULTIVARIADOS DE CARGA CONTAMINANTE POR MÉTODO MAHALANOBIS

Como resultado de la implementación del método distancia Mahalanobis (ver *numeral 1.5* **MÉTODO MULTIVARIADO DE DISTANCIA DE MAHALANOBIS**), fueron detectadas y eliminadas las muestras atípicas (CC y caudales) de los registros existentes en cada PM de los ríos evaluados por este trabajo de grado. Los siguientes gráficos de dispersión matricial se presentan en una escala log-log donde en colores cálidos son presentados los valores atípicos y en color gris claro la muestras catalogadas como datos válidos por cada PM. La ubicación de los gráficos representa la relación de los determinantes (Caudal, DBO5, DQO y SST) con el fin de cuantificar el grado de asociación lineal entre las variables y buscar patrones de correlación [24].

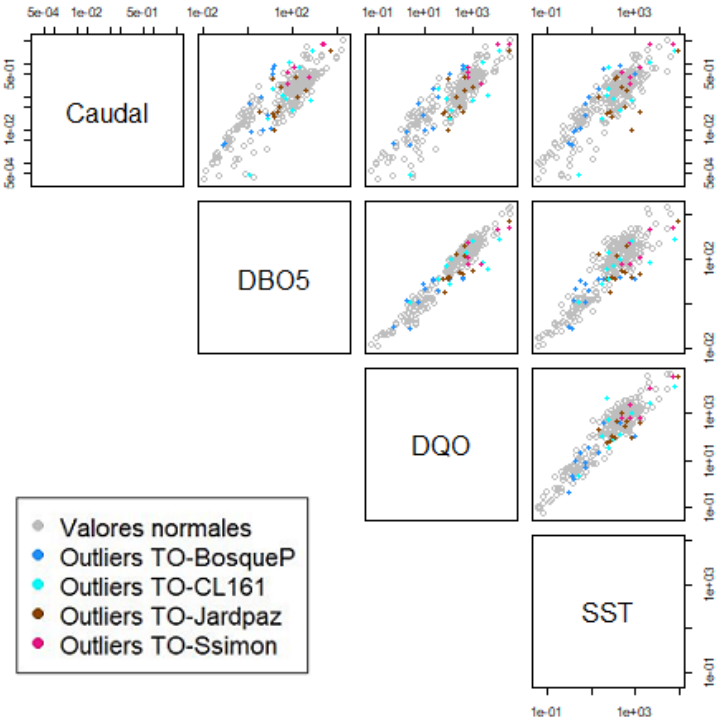
Se debe tomar en consideración que los valores catalogados como atípicos son influenciados en algunos casos por cambios en la normal climatológica, como eventos de sequías e inundaciones, los cuales modifican las condiciones hidrológicas, que finalmente resultan en cambios en el caudal y esto repercute en la CC transportada, además de cambios bruscos en el histórico de los valores por descargas residuales con una cantidad y composición diferente en un periodo de tiempo dado comparado con su comportamiento anterior, cambiando la tendencia de los datos debido a su magnitud [25].

Los gráficos de dispersión multivariados de las demás variables de calidad (GYA, $N_{TOTAL}$, $P_{TOTAL}$ Y SAAM) se encuentran en el **Anexo *D***.

**3.2.1 Porcentaje de muestras atípicas por PM para el canal Torca.** En la *Figura 26* se observa la distribución de los valores de CC y Caudal, resaltando los datos atípicos encontrados para cada PM, donde se evidencia que los PM BosqueP, CL161 y

Jardpaz tienden a tener un porcentaje de valores atípicos similar con un promedio de 16 %, mientras que el PM BosqueP es el PM que menos datos atípicos registra, esto también se puede observar en la *Tabla 3*. En general en el canal Torca se presentó un promedio de 13.5 % de datos atípicos, lo que equivale a 41 muestras.

Figura 26. Muestras atípicas de las variables caudal, DBO5, DQO y SST para cada punto de monitoreo del canal Torca




Fuente: Autores.

Tabla 3. Número de muestras por punto de monitoreo y porcentaje de atípicos correspondiente para el canal Torca

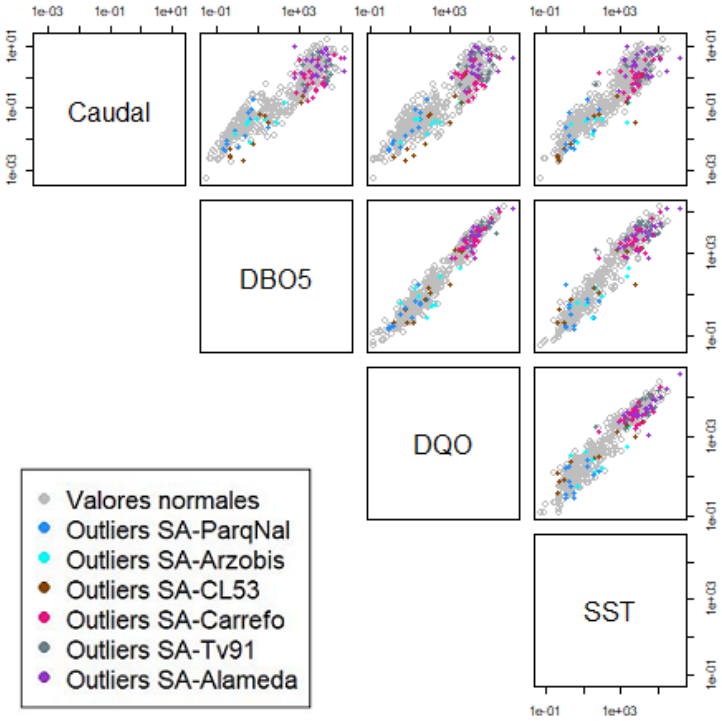
| CANAL TORCA | | |
|---|---|---|
| **Punto de Monitoreo** | **Número de muestras totales por PM** | **% Atípicos por PM** |
| TO-BosqueP | 72 | 16.67 |
| TO-CL161 | 72 | 13.89 |
| TO-Jardpaz | 73 | 16.44 |
| TO-Ssimon | 84 | 7.14 |
| **Total muestras** | **301** | |

Fuente: Autores.

**3.2.2 Porcentaje de muestras atípicas por PM para el río Salitre.** En la Figura *27* se evidencia que los PM pertenecientes a los tramos 1 y 2 (ParqNal, Arzobis y CL53)

presentan menor cantidad de datos atípicos teniendo en promedio 9 %, en tanto que para los tramos 3 y 4 donde se ubican los PM Carrefo, Tv91 y Alameda se concentran mayor número de muestras atípicas llegando a un promedio de 20 %, este patrón de comportamiento se evidencia en la Figura *27*. Finalmente para los 6 PM del río Salitre dio como resultado un promedio de 14 % de valores atípicos correspondientes a 97 muestras.

Figura 27. Muestras atípicas de las variables caudal, DBO5, DQO y SST para cada punto de monitoreo del río Salitre




Fuente: Autores.

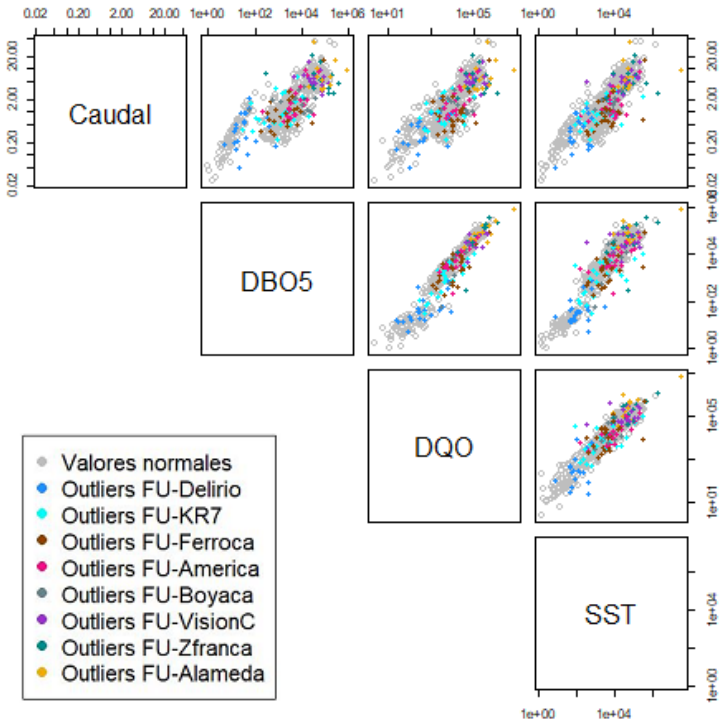
Tabla 4. Número de muestras por punto de monitoreo y porcentaje de atípicos correspondiente para el río Salitre

| **RIO SALITRE** | | |
|---|---|---|
| **Puntos de Monitoreo** | **Número de muestras totales por PM** | **% Atípicos por PM** |
| SA-ParqNal | 109 | 10.09 |
| SA-Arzobis | 107 | 7.48 |
| SA-CL53 | 122 | 9.02 |
| SA-Carrefo | 101 | 22.77 |
| SA-Tv91 | 107 | 17.76 |
| SA-Alameda | 119 | 20.17 |
| **Total muestras** | **665** | |

Fuente: Autores.

**3.2.3 Porcentaje de muestras atípicas por PM para el río Fucha**. En el caso de la distribución de las muestras atípicas para el río Fucha mostrada en la *Figura 28* se observa que esta se mueve en un rango de aproximadamente el 10 % (ver *Tabla 5*), lo que indica similitud en la cantidad de datos reportados como valores atípicos en todos los PM del río Fucha. En general se presentan 144 muestras catalogadas como atípicas.

Figura 28. Muestras atípicas de las variables caudal, DBO5, DQO y SST para cada punto de monitoreo del río Fucha




Fuente: Autores.

Tabla 5. Número de muestras por punto de monitoreo y porcentaje de atípicos correspondiente para el río Fucha

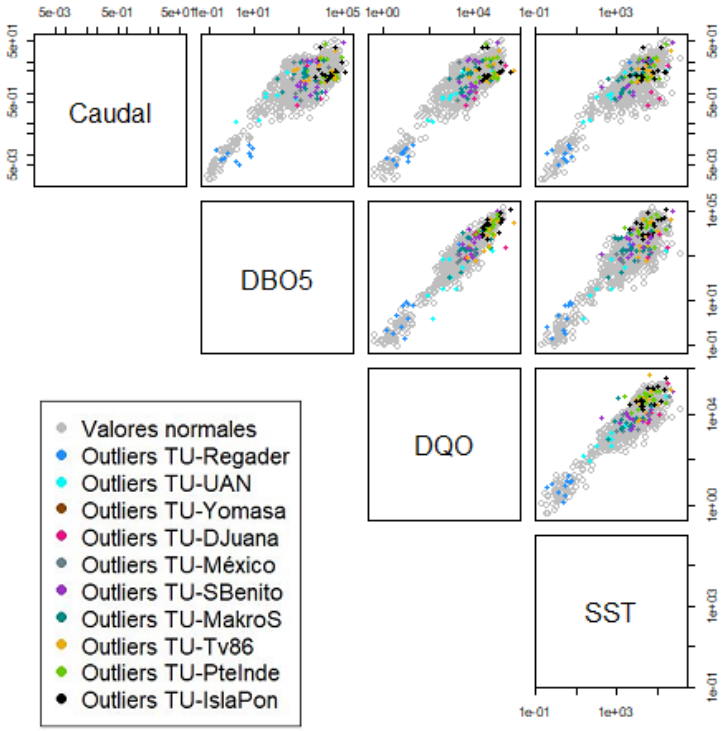
| RIO FUCHA | | |
|---|---|---|
| **Punto de Monitoreo** | **Número de muestras totales por PM** | **% Atípicos por PM** |
| FU-Delirio | 106 | 19.81 |
| FU-KR7 | 106 | 20.75 |
| FU-Ferroca | 109 | 22.02 |
| FU-America | 108 | 17.59 |
| FU-Boyaca | 88 | 17.05 |
| FU-VisionC | 106 | 16,04 |
| FU-ZFranca | 107 | 13.08 |
| FU-Alameda | 108 | 12.04 |
| **Total muestras** | **838** | |

Fuente: Autores.

***3.2.4 Porcentaje de muestras atípicas por PM para el río Tunjuelo.*** *En general para todos todos los PM no se evidencia un patrón secuencial en la cantidad de muestras atípicas moviéndose en un rango moviéndose en un rango de 0 % a 17,82 % como se puede observar en la*

Tabla *6*. También es evidente que los porcentajes de valores atípicos más altos se presentan en los últimos 5 PM con un promedio de 16 % y el número total de muestras atípicas para el río Tunjuelo fue de 119.

Figura 29. Muestras atípicas de las variables caudal, DBO5, DQO y SST para cada punto de monitoreo de la del río Tunjuelo




Fuente: Autores.

Tabla 6. Número de muestras por punto de monitoreo y porcentaje de atípicos correspondiente para el río Tunjuelo

| RIO TUNJUELO | | |
|---|---|---|
| **Puntos de Monitoreo** | **Número de muestras totales por PM** | **% Atípicos por PM** |
| TU-Regader | 98 | 11.22 |
| TU-UAN | 75 | 12 |
| TU-Yomasa | 118 | 0 |
| TU-DJuana | 119 | 4.2 |
| TU-México | 119 | 6.72 |

| | | |
|---|---|---|
| TU-SBenito | 99 | 16.16 |
| TU-MakroS | 101 | 17.82 |
| TU-Tv86 | 99 | 14.14 |
| TU-PteInde | 101 | 15.84 |
| TU-IslaPon | 101 | 17.82 |
| **Total muestras** | **1030** | |

Fuente: Autores.

## 3.3 DETERMINANTES DE LA CALIDAD Y CANTIDAD DE MAYOR IMPORTANCIA PARA CADA RÍO CON BASE EN EL ALGORITMO RANDOM FOREST

Para la ejecución del algoritmo *Random Forest* (RF) se tomó la matriz de datos validados de CC y caudales, resultantes de la aplicación de la metodología de distancia de Mahalanobis, de la cual se descartó la variable correspondiente a NTOTAL, ya que esta presentaba gran cantidad de datos faltantes, relacionados por el software RStudio con valores NA *(No Available)*. Posteriormente se separó en datos de calibración (63 %) y validación (37 %). Como resultado del método RF en ambos conjuntos de datos se obtuvieron las tablas denominadas "tablas de predicciones", las cuales clasificaban las muestras en los PM de cada río con la suma de ambas predicciones (calibración y validación), como se puede observar en el ***Anexo E***, donde la predicción refleja que tan bien o mal clasificados fueron los datos para cada PM indicando el porcentaje de incertidumbre.

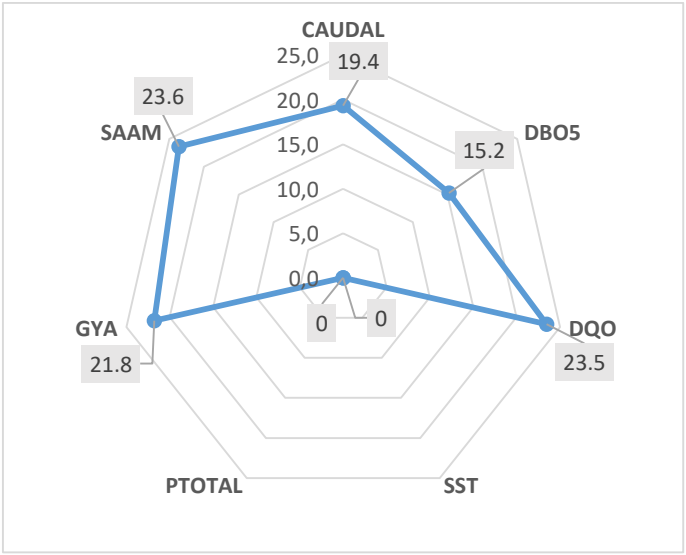
Por último se aplicó la metodología propia propuesta por Autores (ver *sección 2.3* **FASE III. SELECCIÓN DE VARIABLES A TRAVÉS DEL ALGORITMO RANDOM FOREST**) con el fin de seleccionar los cuatro determinantes de la calidad más importantes para cada río, la cual se realizó con base en las tablas de Gini index específico obtenidas mediante el algoritmo RF, el resultado de esta metodología propuesta se quiso representar por medio de gráficos radiales que dan a conocer las magnitudes obtenidas para cada determinante de la calidad según la metodología aplicada (ver

***Anexo H***).

Los gráficos radiales se encuentran representados en una escala adimensional, donde cada vértice del gráfico radial está relacionado con un determinante de la calidad del agua y también el caudal, lo que indica que los picos con tendencia a los extremos del gráfico presentan magnitudes más grandes.

**3.3.1 Orden de importancia de las variables de calidad y cantidad del canal Torca** En cuanto al canal Torca, como se muestra en la
*Figura 30* las variables de mayor importancia son SAAM, DQO, GYA, y Caudal, esto concuerda con el comportamiento de los vertimientos teniendo en cuenta que el área de influencia de este río es en su gran mayoría zonas residenciales y educativas por lo que coincide con las características en cuanto a descargas domésticas [4].

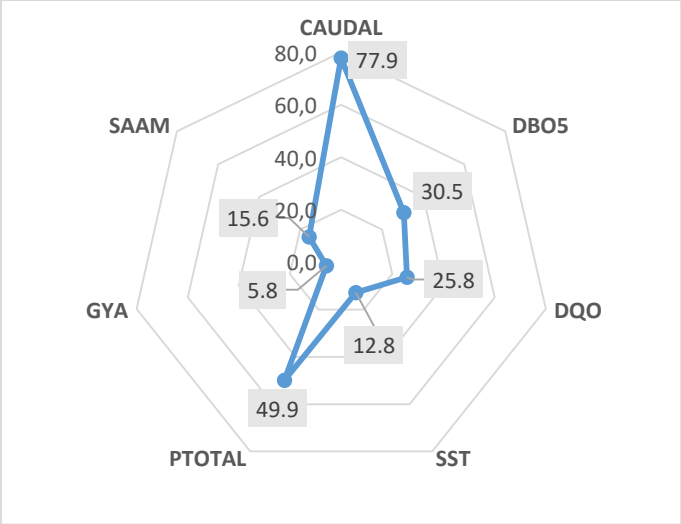
Figura 30. Gráfico radial de la magnitud de la importancia de cada variable según el algoritmo Random Forest para el canal Torca




Fuente: Autores.

**3.3.2 Orden de importancia de las variables de calidad y cantidad del río Salitre.** En la siguiente ilustración se observa la importancia de las variables caudal, P TOTAL, DBO5 Y DQO, las cuales obtienen las mayores magnitudes, representando la dinámica de la CC del río, influenciada por las descargas de aguas residuales domésticas, así como los aportes por vertimientos de tipo industrial que se encuentran ubicados principalmente en las localidades de Engativa, Usaquen, Teusaquillo y Suba, estas industrias pertenecen a actividades tales como: edición e impresión de libros y periódicos, fabricación de productos textiles, fabricación de equipos de transporte, producción de alimentos y bebidas (lácteos) y productos minerales no metálicos [26]. La importancia del caudal y $P_{TOTAL}$, está relacionada con la dinámica y aporte de los humedales y las quebradas de este río en su zona de influencia, adicionalmente esto se debe a que la pendiente del río salitre es muy baja y conlleva a menores valores de velocidad, entonces tiende a presentar una condición de cuerpos lénticos con alto contenido de materia orgánica (aumento de la concentración de nutrientes especialmente fósforo) facilitando que se den procesos de eutrofización del tramo tres y cuatro después del PM CL53 [27].

Figura 31. Gráfico radial de la magnitud de la importancia de cada variable según el algoritmo Random Forest para el río Salitre

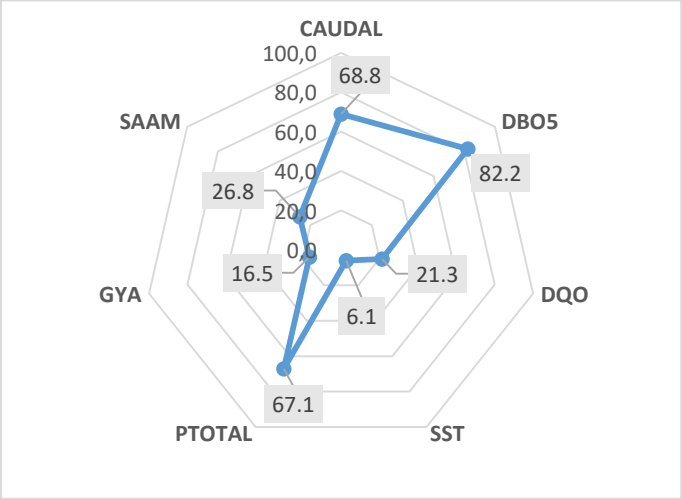



Fuente: Autores.

***3.3.3 Orden de importancia de las variables de calidad y cantidad del río Fucha****. Para el cauce del río Fucha el resultado arrojado se presenta gráficamente en la*

Figura *32*, donde se observa que las variables que representan la variabilidad de calidad del agua en este río son: en primer lugar la DBO5, asociado a vertimientos de tipo domésticos, seguido del Caudal, influenciado por características de hidrología como la desembocadura de diferentes quebradas, por ejemplo, quebrada la Osa, Palo Blanco, Aguasclaras, entre otras, por último en tercer y cuarto lugar $P_{TOTAL}$ y SAAM estos dos determinantes se asocian a vertimientos de origen industrial, lo cual coincide con la infraestructura y actividades propias del área de influencia del río Fucha, tales como, producción de sustancias y productos químicos farmacéuticos, jabones y detergentes, vehículos automotores, productos alimenticios y bebidas y productos textiles [26].

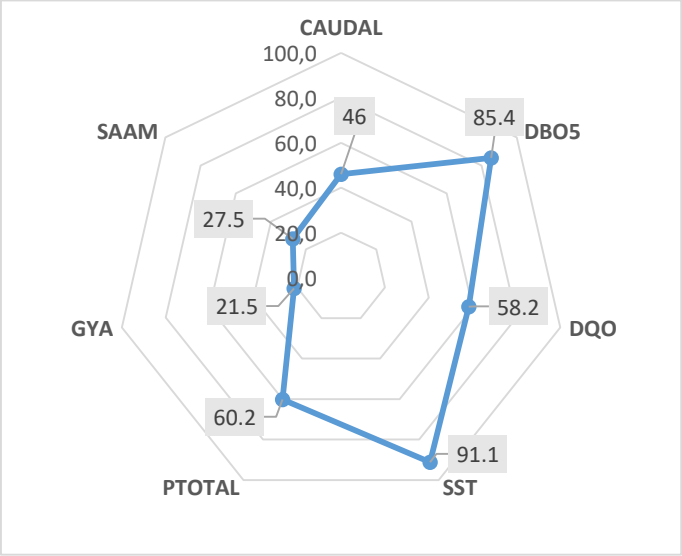
Figura 32. Gráfico radial de la magnitud de la importancia de cada variable según el algoritmo Random Forest para el río Fucha




Fuente: Autores.

**3.3.4 Orden de importancia de las variables de calidad y cantidad del río Tunjuelo.** A continuación se presenta en el gráfico de red, la importancia de cada variable dada su magnitud, dando a conocer la importancia de medir parámetros como: SST que coincide con la influencia de las actividades de aprovechamiento pétreo y arrastre de sedimentos en las principales quebradas afluentes al río Tunjuelo en los tramos 1 y 2, la DBO5 relacionada con las descargas de origen doméstico, $P_{TOTAL}$, este parámetro se ve altamente influenciado por los vertimientos aportados por la planta de tratamientos de lixiviados del relleno sanitario Doña Juana, teniendo en cuenta que la capacidad de la esta planta no es suficiente para el caudal generado y que este parámetro pertenece a la composición química propia de este tipo de vertimientos [28]. Por último la DQO indica la presencia de descargas de origen industrial provenientes de industrias en el área de influencia dedicadas a producción de alimentos y bebidas, fabricación textil, vehículos automotores y fabricación de productos de plástico y caucho [26].

Figura 33. Gráfico radial de la magnitud de la importancia de cada variable según el algoritmo Random Forest para el río Tunjuelo




Fuente: Autores.

## 3.4 ANÁLISIS CLUSTER MEDIANTE EL ALGORITMO EXPECTATION MAXIMIZATION

Como resultado del uso de este algoritmo se obtuvieron dos gráficos y tablas que serán explicados a continuación en los que se basó la determinación del número de PM óptimos para cada río.

### Gráfico de porcentaje de ganancia

Este gráfico permite establecer el número óptimo de *clusters* representado con la línea punteada para cada para cada análisis realizado, en función del Porcentaje de Ganancia

(PG) que fue obtenido de los resultados de la métrica BIC (ver *sección 1.7* **ALGORITMO DE AGRUPAMIENTO EXPECTATION MAXIMIZATION**). El número óptimo de *cluster* está definido como aquel cuyo PG es más cercano al 2 %. Por ejemplo, en la Figura *34* en el eje vertical aparecen los valores de los PG obtenidos para cada número de *clusters* evaluados por el algoritmo EM (es decir, eje horizontal) y sus resultados son representados por la línea roja.

**Gráficos circulares**

*La*

Figura *35*, es un ejemplo de los gráficos circulares donde se clasificaron las muestras (cargas y caudales) de los PM en *clusters* (óptimo) obtenidos del gráfico anteriormente explicado para el periodo de calibración (2006 - 2013). Cada *cluster* clasifica cierto número de muestras de los PM con respecto a la similitud de las muestras y su variabilidad teniendo en cuenta las condiciones de caudal.

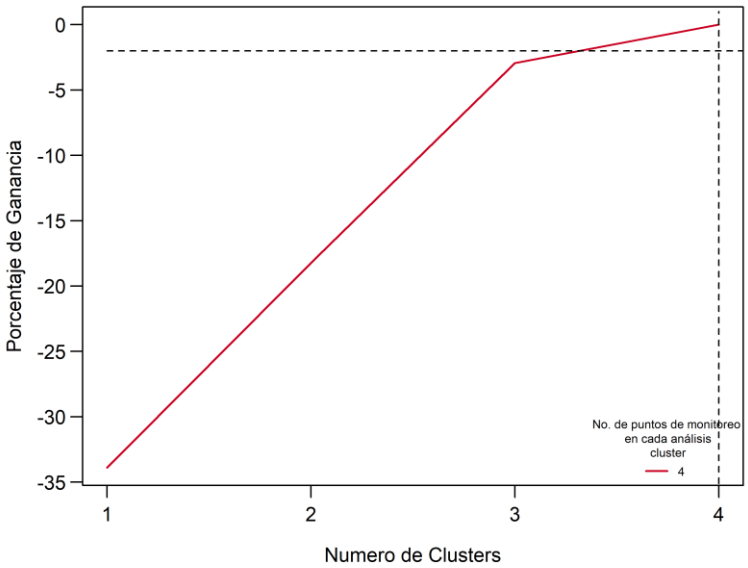
El gráfico circular presenta en la parte superior el número de *cluster* para los cuales se realizó el análisis y en la parte inferior los PM del río, las franjas de colores en el círculo en la parte superior representan el caudal medio en el cual fueron clasificadas las muestras (ver leyenda) y la longitud de estas franjas da a conocer cuántas muestras de cada PM fueron clasificadas en cada *cluster,* la cantidad de muestras clasificadas están rotuladas con la escala numérica en la parte superior del círculo. Las franjas de colores en la parte inferior representan las muestras que pertenecen a cada PM igualmente rotuladas con la escala numérica en la parte inferior del círculo.

*Los valores de caudales y cargas medias de los clusters conformados para cada río en el periodo de periodo de calibración pueden observar en el*

**Anexo *I*** y en el periodo de validación en el **Anexo *J***, de igual manera en el **Anexo *K*** se encuentra el análisis *cluster* de clasificación de las muestras de los PM para el periodo de validación***.***

**3.4.1 Análisis cluster para el canal Torca.** Para el canal del canal Torca dio como resultado que cuatro (4) *clusters* en este río permiten representar en su totalidad la dinámica de la calidad del agua con respecto a la CC transportada en el periodo de tiempo de la etapa de calibración lo que se puede observar en la *Figura 34*.

Figura 34. Gráfico de porcentaje de ganancia de cada cluster evaluado considerando la métrica BIC para el canal Torca



Fuente: Autores.

De acuerdo con la ejecución del algoritmo EM el mejor modelo probabilístico de clasificación multivariada en función del BIC en el caso del canal Torca el modelo VVV (elipsoidal con variación del volumen, forma y orientación) fue el que mejor clasificó los datos de carga transportada en el periodo de calibración.

*La clasificación de las muestras en los diferentes cluster para el canal Torca se observa en la*

Figura *35*, donde se asignan gráficamente los valores de la Tabla *7*. En el *cluster* 1 (CL-1) se observa que el 82.35 % de las muestras clasificadas pertenecen al PM-1 (BosqueP) y el porcentaje restante pertenece a los PM-2 y PM-4 (CLL161 y SSimon). Por otra parte, ninguna de las muestras de PM JardPaz se clasificaron en el CL-1 debido a que las muestras de esté PM no mostraron similitud con las clasificadas por el CL-1 en cuanto a la condición media de caudal de 0.02 $m^3.s^{-1}$.

*Las magnitudes bajas de las variables SAAM, DQO, GYA caracterizaron al CL-1, ya que el caudal medio de caudal medio de clasificación de muestras fue de [0.02 $m^3.s^{-1}$] como se puede observar en la sección de color la sección de color café en la parte superior de la*

Figura 35 *(los valores de CC para el periodo de calibración se pueden consultar en el*

**Anexo *I***, Tabla *41*). Estas condiciones coinciden con la dinámica hídrica que ha sido registrada principalmente en el PM BosqueP donde no se presentan vertimientos de aguas residuales de gran magnitud.

El *cluster* 2 (CL-2) fue el que mayor número de muestras clasificó distribuidas de forma homogénea en los PM CLL161, JardPaz y SSimon, y tan sólo un porcentaje del 4 % para el PM BosqueP. Este *cluster* se caracteriza por presentar un caudal medio 0.13 $m^3.s^{-1}$, el cual representa en mayor proporción la dinámica y variabilidad de las muestras del canal Torca bajo esta condición del caudal.

En cuanto al *cluster* 3 (CL-3), está conformado por el menor número de muestras frente a los demás cluster. La magnitud media de caudal (1.6 $m^3.s^{-1}$) y de CC analizados son las más altas entre los *cluster* de clasificación, que principalmente provienen de muestras de PM CLL161 (62 % del total) donde se presentan un incremento en los valores de caudal y CC (ver *sección 3.1.1* **Canal Torca.**) debido a la influencia de los vertimientos (ver Figura *2*).

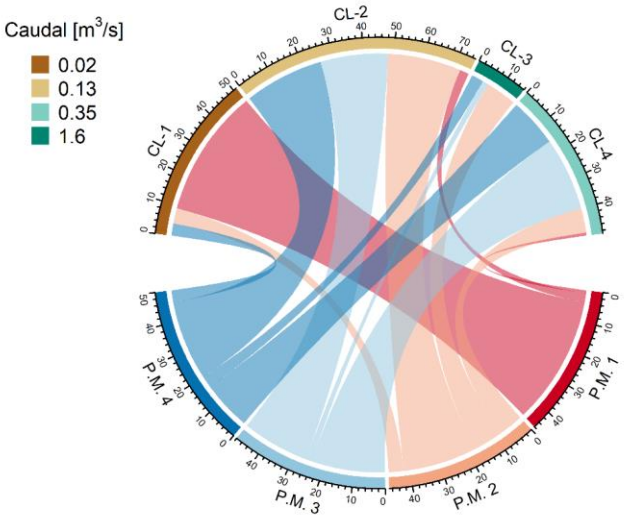
Con respecto al *cluster* 4 (CL-4) este asocia un caudal medio de 0,35 $m^3.s^{-1}$, bajo esta condición los mayores porcentajes de clasificación están en primer lugar en el PM JardPaz con un 49 % y segundo lugar SSimon con un 33 %.

Tabla 7. Número de muestras de los puntos de monitoreo del canal Torca clasificadas en cada cluster para el periodo de calibración

| | **CANAL TORCA** | | | |
|---|---|---|---|---|
| | **Puntos de monitoreo** | | | |
| **Clusters** | **BosqueP** | **CLL161** | **JarPaz** | **SSimón** |
| 1 | 42 | 5 | 0 | 4 |
| 2 | 3 | 24 | 22 | 26 |
| 3 | 0 | 10 | 2 | 4 |
| 4 | 1 | 8 | 24 | 16 |

Fuente: Autores.

Figura 35. Gráfico circular de las muestras de los puntos de monitoreo del canal Torca clasificadas en cada cluster



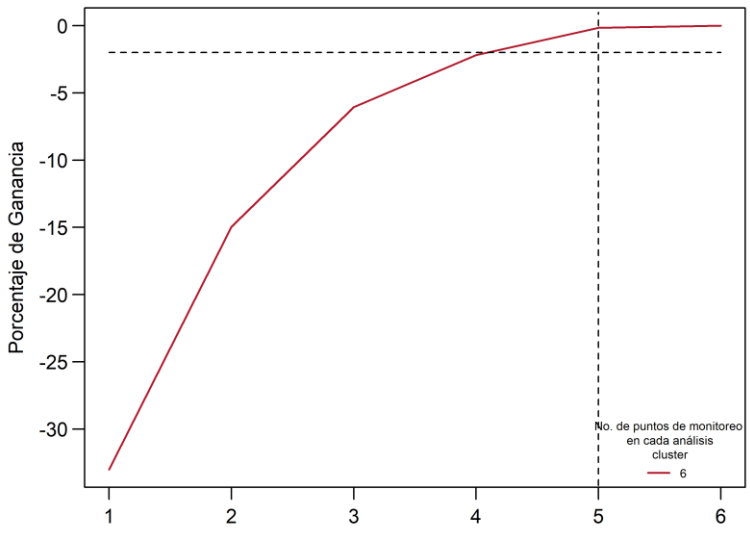

Fuente: Autores.

El canal del río Torca conserva sus cuatro PM, ya que cada uno de estos aporta a la caracterización de la calidad y cantidad de agua de este afluente, por ejemplo el PM BosqueP da a conocer las condiciones iniciales del rio antes de cualquier vertimiento como se muestra en la Figura *2*. En cuanto al PM CLL161 su importancia radica en que según los perfiles longitudinales de CC (ver *sección 3.1.1* **Canal Torca.**) en este punto se presenta un aumento significativo en la CC por los vertimientos del sistema de alcantarillado público localizados aguas arriba de este PM.

A pesar que los PM JardPaz y SSimon se encuentran en los mismos *cluster* y con número de muestras similares, son importantes para realizar el seguimiento a la salud del ecosistema y evaluar la capacidad de retención del humedal Torca-Guaymaral, además el PM SSimon ejerce un control de lo que se vierte al río Bogotá a través del canal Torca. Por esta razón estos PM sirven como punto de control, de acuerdo a lo anterior no se consideran para ser unificados.

**3.4.2 Análisis cluster para el río Salitre.** Como se observa en la Figura *36*, para el río Salitre cinco (5) *clusters* permiten representar en su totalidad la dinámica de la calidad del agua con respecto a la CC transportada (Caudal, $P_{TOTAL}$, DBO5 y DQO) en el periodo de tiempo de la etapa de calibración lo que se puede observar en la *Ilustración 31*. El PG de los clusters del 1-4 fue superior 2 % indicando que en *cluster* 5 tuvo una diferencia menor del 2 % respecto al BIC óptimo (número de clusters), por lo tanto este número de *clusters* (cinco) fue seleccionado para realizar la agrupación de los datos con características similares.

Figura 36. Gráfico de porcentaje de ganancia de cada cluster evaluado considerando la métrica BIC para el río Salitre




Fuente: Autores.

Según la selección del número óptimo de *clusters* mediante la ejecución del algoritmo EM, se estableció según la métrica BIC el modelo que representaba de mejor forma las propiedades de las agrupaciones fue el modelo VVV.

*La distribución de las muestras de los PM en los diferentes clusters para el periodo de calibración se ve calibración se ve reflejada en la* Tabla 8 *y la* Figura 37*. El CL-1 clasifica en total 114 muestras donde un 95 % corresponden a los dos primeros PM ParqNal y Arzobis y el 5 % restante al PM CLL53. Los PM ParqNal y Arzobis son representados bajo la condición de caudal medio de 0.03 $m^3.s^{-1}$ siendo esta la condición más baja indicando que en los primeros PM del río salitre registran bajos niveles de caudal y CC menores a los registrados en los otros PM como se puede observar en el*

**Anexo *I***, Tabla *42*.

En cuanto al CL-2 este clasifica la mayoría de muestras de los dos últimos PM (Tv91 y Alameda) teniendo como referencia el caudal medio más alto de 5.6 $m^3.s^{-1}$ con respecto a los otros *clusters*. En los últimos PM el caudal al igual que la cc presenta los valores más altos ya que aquí se ve reflejado el transporte y acumulación de las descargas vertidas en el Salitre.

El registro del CL-3 da a conocer que el 66 % de las muestras pertenecen al PM Arzobis y el 26 % al PM CLL53 de acuerdo con la condición de caudal medio de 0.08 $m^3.s^{-1}$ (la segunda más baja), lo cual quiere decir que más de las mitad de las muestras del PM Arzobis tienen su variabilidad dentro de esta condición y el otro porcentaje de las muestras de este punto se encuentran reguladas por las condiciones del Cl-1, debido a que el caudal de CL-1 de igual manera está en un rango similar.

El CL-4 con un caudal medio de 0,51 $m^3.s^{-1}$ clasifica en total 114 muestras distribuidas a lo largo del cauce con un 47 % correspondiente al PM CLL53 y un 38 % para el PM Carrefo y el porcentaje restante para los demás PM exceptuando el PM ParqNal, que no fue clasificado en este *cluster* por la baja magnitud de sus caudales. En cuanto al CL-5 este registró un caudal medio del 2,07 $m^3.s^{-1}$ y clasificó únicamente muestras de los 4 últimos PM, con tendencia a aumentar hacia el último PM.

En cuanto al CL-5 este manejo un caudal medio del 2,07 $m^3.s^{-1}$ y clasificó muestras de los 4 últimos PM, con tendencia a aumentar hacia el último PM, lo cual coincide con los registros elevados en los últimos PM, que se pueden observar en los perfiles longitudinales *sección 3.1.2* **Río Salitre.**

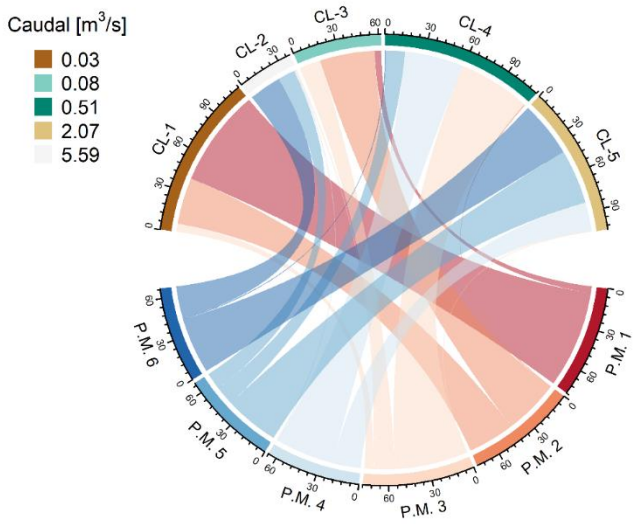
La mayoría de las muestras 67 % se clasificaron en los *clusters* CL-1, CL-3 y CL-4 los cuales manejan un caudal medio no mayor a 0,51 $m^3.s^{-1}$

Tabla 8. Número de muestras de los puntos de monitoreo del río Salitre clasificadas en cada cluster para el periodo de calibración

| | **RÍO SALITRE** | | | | | |
|---|---|---|---|---|---|---|
| | **Puntos de monitoreo** | | | | | |
| **Clusters** | **ParqNal** | **Arzobis** | **CLL53** | **Carrefo** | **Tv 91** | **Alameda** |
| 1 | 73 | 35 | 6 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 1 | 13 | 24 |
| 3 | 5 | 41 | 16 | 0 | 0 | 0 |
| 4 | 0 | 1 | 54 | 44 | 14 | 1 |
| 5 | 0 | 0 | 1 | 21 | 41 | 42 |

Fuente: Autores.

Figura 37. Gráfico circular de las muestras de los puntos de monitoreo del río Salitre clasificadas en cada cluster

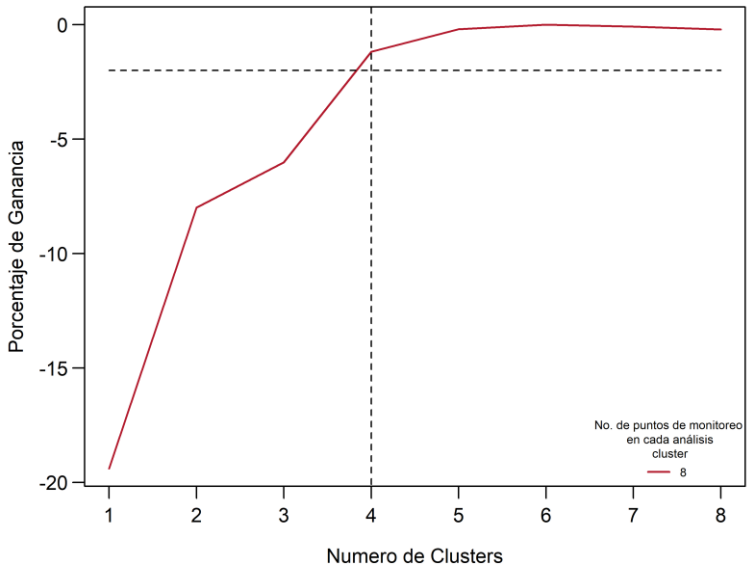



Fuente: Autores.

De acuerdo a los resultados del algoritmo EM y análisis de clusters se considera remover solo un punto con respecto a los PM actuales ya que el resultado arroja que cinco (5) PM pueden caracterizar en su totalidad la dinámica hidrológica, se consideró que el PM a ser removido es Tv91, ya que como se puede observar en los perfiles de CC (*sección 3.1.2* **Río Salitre.**) este punto no presenta una mayor alteración en la magnitud de los valores de CC con respecto al PM inmediatamente siguiente (Alameda), lo que también se puede evidenciar en la similitud de cantidad de muestras clasificadas según la *Tabla 8* para estos dos PM.

**3.4.3 Análisis cluster para el río Fucha.** Para el río Fucha cuatro (4) *clusters* en este río permiten representar en su totalidad la dinámica de la calidad del agua con respecto a la CC transportada (DBO5, Caudal, $P_{TOTAL}$ y SAAM) en el periodo de tiempo de la etapa de calibración lo que se puede observar en la Figura *38.* El PG de los clusters 1, 2 y 3, fue superior 2 % indicando que en *cluster* 4 tuvo una diferencia menor del 2 % respecto al BIC óptimo (número de *clusters*), por lo tanto este número de *clusters* (cuatro) fue seleccionado para realizar la agrupación de los datos con características similares.

Figura 38. Gráfico de porcentaje de ganancia de cada cluster evaluado considerando la métrica BIC para el río Fucha



Fuente: Autores.

Según la métrica BIC para el río Fucha, el mejor modelo mixto fue VVV con el cual se logró una adecuada agrupación de las muestras como se puede observar en la Tabla *9* y Figura *39* se describe a continuación:

En el primer *cluster* (CL-1) las muestras tienen caudal medio de 5,27 $m^3.s^{-1}$, este clasificó en total 193 muestras para todos los PM, donde el 83 % pertenece a los PM VisionC, ZFranza y Alameda que corresponden a los tres últimos PM aguas abajo del río, esto quiere decir que en estos PM se registran CC y caudales altos lo que se puede constatar en el anexo I, esto se le puede atribuir al transporte y acumulación de vertimientos aguas arriba,

adicionalmente en la zona de influencias de estos PM se observa una cantidad considerable de puntos de vertimientos (ver Figura *4*).

*El CL-2 clasificó todas las muestras en el PM Delirio (PM-1), esto quiere decir que esté PM tiene unos valores tiene unos valores bajos de caudal y CC (ver*

**Anexo *I***, Tabla *43*) con respecto a los demás PM y por ello se clasificaron en este *cluster* el cual tiene en cuenta el caudal medio más bajo de 0,33 $m^3.s^{-1}$, esto se debe principalmente a que esté PM tiene influencia de solo un punto de vertimiento como se puede observar en la Figura *4*.

El CL-3 no clasifica ninguna muestra de los tres últimos PM (VisionC, ZFranca y Alameda), pero clasifica el mayor número de muestras para un total de 242 muestras distribuidas en el resto de PM, el caudal medio que influye en este *cluster* es de 1.05 $m^3.s^{-1}$, lo que indica que este valor de caudal representa en mayor medida la dinámica y variabilidad de las condiciones del río.

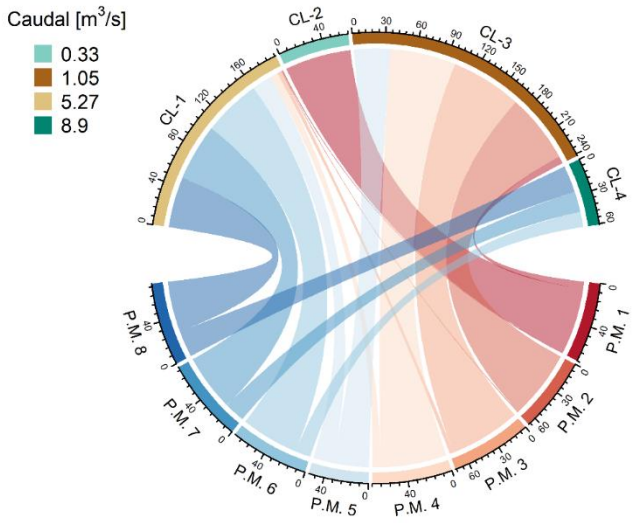
El CL-4 clasificó de acuerdo con el mayor caudal (medio) de todos los *clusters* de 8,9 $m^3.s^{-1}$, clasificó del total de muestras únicamente las muestras en los PM VisiónC con un 25 %, ZFranca con un 32 % y Alameda con un 43 %, este último PM evidenció el impacto de las descargas vertidas a lo largo de todo el río, las cuales incrementaron en los últimos 3 PM y de acuerdo con el porcentaje de clasificación presentan valores dentro de un rango similar.

Tabla 9. Número de muestras de los puntos de monitoreo del río Fucha clasificadas en cada cluster para el periodo de calibración

| | **RÍO FUCHA** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Puntos de monitoreo** | | | | | | | |
| **Clusters** | **Delirio** | **KR7** | **Ferroca** | **América** | **Boyacá** | **VisiónC** | **ZFranca** | **Alameda** |
| 1 | 1 | 1 | 4 | 8 | 19 | 56 | 55 | 49 |
| 2 | 64 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 7 | 67 | 69 | 64 | 35 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 15 | 20 | 26 |

Fuente: Autores.

Figura 39. Gráfico circular de las muestras de los puntos de monitoreo del río Fucha clasificadas en cada cluster



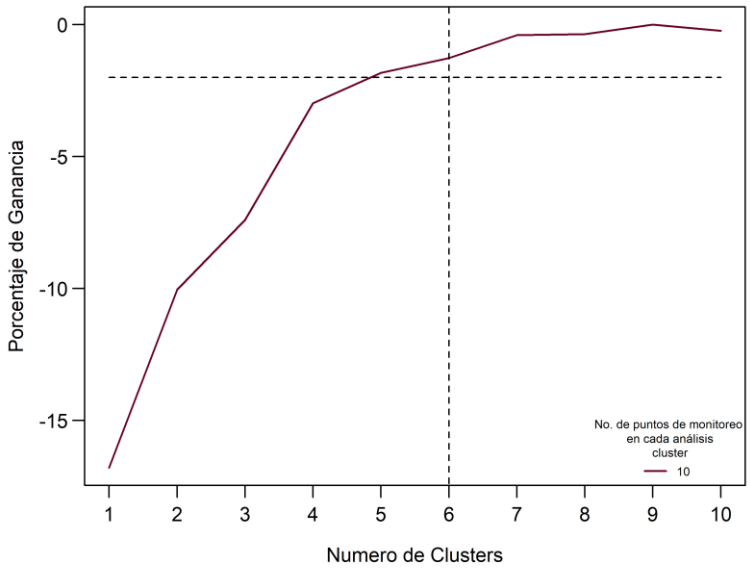

Fuente: Autores.

Según el algoritmo EM, 4 de los 8 PM presentes en el río Fucha pueden llegar a caracterizar la dinámica de la calidad del agua, sin embargo se debe tener en cuenta la variabilidad de las CC y aspectos ambientales, por lo tanto se considera remover solo 2 PM correspondientes a Ferroca y ZFranca, los cuales según el análisis *cluster* estos PM no presentan notorios cambios en la clasificación de muestras entre los PM aledaños (ver Tabla *9*) y no presentan cambios de magnitud considerable con respecto a PM cercanos como se puede observar en los perfiles longitudinales ver *sección 3.1.3* **Río Fucha.**

***3.4.4 Análisis cluster para el río Tunjuelo.*** *Para el río Tunjuelo seis (6) clusters en este río río permiten representar en su totalidad la dinámica de la calidad del agua con respecto a la CC transportada la CC transportada (SST, DBO5, PTOTAL y DQO) en el periodo de tiempo de la etapa de calibración lo que se calibración lo que se puede observar en la*

Figura *40.* El PG de los *clusters* del 1-5, fue superior 2 % indicando que en *cluster* 6 tuvo una diferencia menor del 2 % respecto al BIC óptimo (número de *clusters*), por lo tanto este número de *clusters* (seis) fue seleccionado para realizar la agrupación de los datos con características similares.

Figura 40. Gráfico de porcentaje de ganancia de cada cluster evaluado considerando la métrica BIC para el río Tunjuelo



Fuente: Autores.

Para el río Tunjuelo el mejor modelo probabilístico de clasificación multivariada en función del BIC fue VVE (elipsoidal con rotación) según la ejecución del algoritmo EM, este representó adecuadamente el comportamiento de las agrupaciones.

La distribución de las muestras de los PM de los 6 *cluster* para el río Tunjuelo están representadas en la *Tabla 10* y la
*Figura 41*, donde se puede evidenciar para el CL-1 una distribución completa en los dos primeros PM Regader y UAN con un porcentaje de 88 % y 12 % respectivamente, esto concuerda con el caudal medio de 0,01 $m^3.s^{-1}$ perteneciente al CL-1 y bajos niveles de CC (ver
***Anexo I***, *Tabla 44*) comparados con los reportados en los otros PM lo que se puede evidenciar en los perfiles longitudinales de CC (ver *sección 3.1.4* **Río Tunjuelo.**).

En el CL-2 todas las muestras se clasificaron en todos los PM exceptuando Regader y UAN. Se presenta similitud en el número de muestras para los tres últimos PM (Tv86, PteInde e IslaPon), correspondiente al 61 % de las muestras clasificadas por este *cluster*, lo cual concuerda con el caudal medio de 7.62 $m^3.s^{-1}$ que tiene este *cluste*r, donde generalmente en los últimos PM se presentan los mayores caudales.

El tercer *cluster* CL-3 es el que más muestras clasifica llegando a las 195 muestras, de las cuales el 51 % pertenecen a los PM Yomasa y México, también se debe nombrar que de los tres últimos PM este *cluster* no clasifica ninguna muestra debido a que el CL-3 trata caudales y CC considerablemente bajas (ver *sección 3.1.4* **Río Tunjuelo.**), que se presenta más que todo en tramos iniciales del río.

Por otra parte el CL-4 fue el que clasificó el menor número de muestras para un total de 27 muestras lo que se le puede atribuir a que este tiene un caudal medio de 14.67 $m^3.s^{-1}$ y es el más elevado con respecto a los otros *clusters*, de las muestras clasificadas para este

*cluster* el 44 % pertenecen al PM DJuana por lo tanto se puede decir que aguas arriba de este PM se presentan significativos vertimientos.

La clasificación de muestras del CL-5 tiene similitud con la clasificación del CL-2, diferenciada por el caudal medio perteneciente a este *cluster* que es de 2,79 $m^3.s^{-1}$, lo cual quiere decir que los valores de los 3 últimos PM están en el rango del caudal del CL-2 Y CL-5.

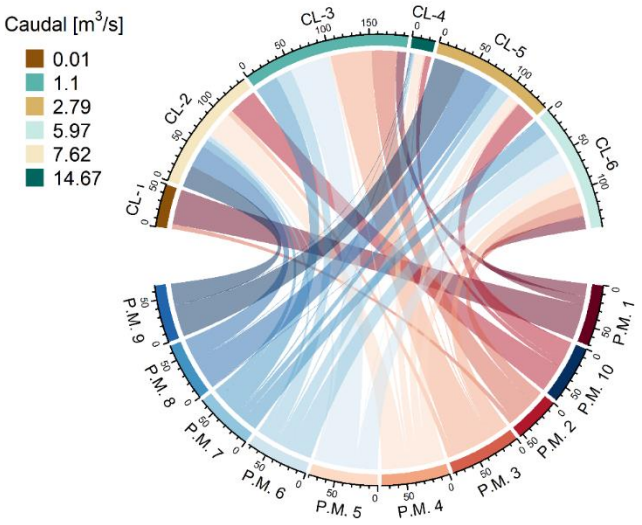
Para el CL-6 el caudal medio considerado es de 5,97 $m^3.s^{-1}$ siendo un caudal de magnitud media-alta comparado con los de los demás *cluster* generando una clasificación homogénea de las muestras, exceptuando los 3 últimos PM, lo cual quiere decir que este valor de caudal medio representa en general la dinámica o variabilidad de las CC presentes en el río excepto por los 3 últimos PM.

Tabla 10. Número de muestras de los puntos de monitoreo del río Tunjuelo clasificadas en cada cluster para el periodo de calibración

| | **RÍO TUNJUELO** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Puntos de monitoreo** | | | | | | | | | |
| **Clusters** | **Regader** | **UAN** | **Yomasa** | **DJuana** | **México** | **SBenito** | **MakroS** | **Tv86** | **PteInde** | **IslaPon** |
| 1 | 44 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 7 | 32 | 3 | 5 | 10 | 28 | 27 | 35 |
| 3 | 11 | 32 | 51 | 3 | 49 | 20 | 29 | 0 | 0 | 0 |
| 4 | 0 | 0 | 2 | 12 | 0 | 2 | 2 | 1 | 1 | 7 |
| 5 | 0 | 0 | 2 | 13 | 1 | 17 | 9 | 38 | 39 | 25 |
| 6 | 16 | 16 | 21 | 19 | 27 | 29 | 19 | 1 | 0 | 1 |

Fuente: Autores.

Figura 41. Gráfico circular de las muestras de los puntos de monitoreo del río Tunjuelo clasificadas en cada cluster




Fuente: Autores.

Aunque el análisis del algoritmo EM para el río Tunjuelo arrojó un número óptimo de PM de mínimo 6, se considera pertinente tener en cuenta que algunos puntos cumplen funciones imprescindibles en el contexto hidrográfico como puntos de control a lo largo del cauce y punto de inicio y de fin.

De acuerdo con lo anterior no es viable remover 4 PM de los diez (10) PM actuales, por lo tanto se propone eliminar solo 3 PM los cuales son UAN, MakroS y PteInde, debido a que éstos PM registran cargas similares a los PM inmediatamente anteriores, esto para el caso de MakroS y PteInde (ver *sección 3.1.4* **Río Tunjuelo.**), en el caso de UAN la CC podría ser monitoreada en el siguiente PM (Yomasa), es decir, que estos dos PM sean unificados, ya que en este punto hay un importante cambio en la tendencia de la CC comparado con el presentado en el PM UAN. Además el análisis *cluster* arrojó semejanzas en la clasificación de muestras de los PM removidos con PM próximos.

El otro PM candidato a ser removido era DJuana pero teniendo en cuenta la actividad llevada a cabo en este espacio geográfico adquiere una importancia de control ambiental que no se puede ignorar, en cuanto a los demás PM no se tuvieron en cuenta para ser removidos debido a lo evidenciado en la magnitud de las diferencias de un PM a otro en el análisis *cluster*.

# 4. IMPACTO SOCIAL

La solución propuesta para la optimización de la RCHB, tendrá efectos en el entorno a nivel social, económico y ambiental, de esta manera principalmente en la reducción de costos se tendrá un beneficio para la población de la ciudad de Bogotá, ya que se manejan mejor los recursos económicos, obteniendo un ahorro que puede ser invertido en la misma red o en otras instituciones públicas, además del beneficio ambiental, ya que si el monitoreo se realiza de la manera más eficiente posible, este será de mayor provecho para la ciudad en general, aumentando el bienestar social y asegurando los servicios ecosistémicos actuales y futuros, adicionalmente de ser una importante fuente de información para el desarrollo investigativo e intelectual de la ciudad.

Así una reducción de los PM y muestreos en variables específicas que representan la dinámica de la calidad y cantidad podría resultar en una disminución de los costos de monitoreo. Además, el mantenimiento de menos PM podría permitir duraciones más largas del monitoreo registrado por el mismo costo que mantener más PM con duraciones más cortas [29].

De acuerdo con una investigación realizada en el río Limpopo ubicado en Mozambique (país situado al sureste de África.) para optimizar una red de monitoreo de calidad se deben tener en cuenta los costos totales de la siguiente manera: (i) costos de mano de obra; (ii) equipo de campo y mantenimiento; iii) gasto anual para la recolección de muestras; (iv) el transporte; (v) Costos analíticos y (vi) informes [30]. Otro estudio investigó el rendimiento del AG en la optimización de la red de monitoreo de calidad del agua de la cuenca del río Gediz (longitud del río de 276 km y área de drenaje total de 16775 km2) para disminuir el número de PM de 33 a 14 en la red de monitoreo, finalmente se ahorraron aproximadamente 33000 US por año sin incluir la depreciación del equipo y el costo indirecto con la modificación del número de PM [31].

## 4.1 AHORRO CON LA OPTIMIZACIÓN DE LA RCHB

El cálculo del ahorro anual de la RCHB se realizó para cada río teniendo en cuenta el costo comercial unitario de los diferentes parámetros que mide cada PM perteneciente a la misma, de donde se tiene una base de datos con los costos que ofrecen 8 diferentes laboratorios en la ciudad de Bogotá, costos al año del 2016. De esta base de datos se obtiene el promedio en costos para cada variable, estos valores son sumados para obtener el costo de una muestra, es decir se tienen en cuenta las variables (G y A, DBO5, DQO, SST, PT, NT, SAAM, parámetros in situ y caudal) medidos por la RCHB lo cual es equivalente a una muestra.

Para el número de muestras anuales se toma el número de muestras anuales de los últimos dos años (2017-2018) que es igual a 6 (seis) para cada año, y en cuanto al número de PM se contrastan los PM existentes en la actualidad con la reducción en el número de PM obtenida (ver *sección 3.4* **ANÁLISIS CLUSTER MEDIANTE EL ALGORITMO EXPECTATION MAXIMIZATION**).

Se debe tener en cuenta que los precios que ofrecen los laboratorios por la caracterización del agua en cuanto a las diferentes variables de calidad que tiene en cuenta la RCHB, son costos del año 2016 por lo que se hace una conversión al costo actual de año 2019 de la siguiente forma:

Ecuación 10. Ecuación de valor final teniendo en cuenta el IPC

$$Valor\ Final = Valor\ inicial * (\frac{IPC\ final}{IPC\ inicial})$$

Fuente: [32]

El IPC (Índice de precios al consumidor) para el año 2019 es de 116.876 y el IPC para el año 2016 es de 102.912 [32].

También es importante aclarar que para el cálculo del valor por muestra denominado como nuevo, es decir con la aplicación de la optimización de la RCHB, solo se tienen en cuenta las variables más importantes determinadas a través del algoritmo RF para cada río, las cuales se pueden ver en la *sección 3.3* .

Este estudio de la disminución de costos con la aplicación de la optimización de la RCHB, solo tiene en cuenta los costos de la caracterización de las muestras de calidad y cantidad, es decir no tiene en cuenta costos de mano de obra, transporte, equipo de campo y mantenimiento, informes, entre otros.

**4.1.1 Ahorro anual canal Torca.** Como no se presentó disminución en el número de PM para el canal Torca, los ahorros que genera este río a la RCHB están dados por la disminución en el número de variables a monitorear, generando así un ahorro de 11 % anualmente.

Tabla 11. Ahorro anual con la optimización de la RCHB para el canal Torca

| Canal Torca | | |
|---|---|---|
| | ACTUAL | NUEVO |
| NÚMERO DE PM | 4 | 4 |
| MUESTRAS POR AÑO | 6 | 6 |
| VALOR POR MUESTRA | $ 1’136,622 | $ 1’011,970 |
| TOTAL ANUAL | $ 27’278,928 | $ 24’287,282 |
| AHORRO ANUAL | $ | 2’991,646 |

Fuente: Autores.

**4.1.2 Ahorro anual río Salitre.** La disminución en el número de PM para el río Salitre fue de solo uno por los tanto representaría un ahorro anual para la RCHB de $11’277,995, este valor también nos da una idea de lo que cuesta aproximadamente la caracterización de las diferentes variables de calidad y cantidad anualmente en un PM.

Tabla 12. Ahorro anual con la optimización de la RCHB para el río Salitre

| Río Salitre | | |
|---|---|---|
| | ACTUAL | NUEVO |
| NÚMERO DE PM | 6 | 5 |
| MUESTRAS POR AÑO | 6 | 6 |
| VALOR POR MUESTRA | $ 1’136,622 | $ 988.013 |
| TOTAL ANUAL | $ 40’918,392 | $ 29’640,397 |
| AHORRO ANUAL | $ | 11’277,995 |

Fuente: Autores.

**4.1.3 Ahorro anual río Fucha.** El número de PM en el río Fucha se reducen en 2 PM pasando de 8 a 6 PM y considerando que solo se tienen en cuenta en el muestreo las variables más importantes en la optimización de la operación de la RCHB, el ahorro reportado para este río es del 35 % en costos de caracterización de las muestras anualmente.

Tabla 13. Ahorro anual con la optimización de la RCHB para el río Fucha

| Río Fucha | | |
|---|---|---|
| | ACTUAL | NUEVO |
| NÚMERO DE PM | 8 | 6 |
| MUESTRAS POR AÑO | 6 | 6 |
| VALOR POR MUESTRA | $ 1’136,622 | $ 989.460 |
| TOTAL ANUAL | $ 54’557,856 | $ 35’620,565 |
| AHORRO ANUAL | $ | 18’937,291 |

Fuente: Autores.

**4.1.4 Ahorro anual río Tunjuelo.** De acuerdo con la reducción en el número de PM igual a tres para el río Tunjuelo, el número de muestras anuales pasa de 60 a 42 en todo el afluente, lo que genera también una disminución en los costos en un 38 % traducido en casi 26 millones de pesos como se muestra en la siguiente tabla:

Tabla 14. Ahorro anual con la optimización de la RCHB para el río Tunjuelo

| Río Tunjuelo | | |
|---|---|---|
| | ACTUAL | NUEVO |
| NÚMERO DE PM | 10 | 7 |
| MUESTRAS POR AÑO | 6 | 6 |
| VALOR POR MUESTRA | $ 1’136,622 | $ 1’009,460 |
| TOTAL ANUAL | $ 68’197,320 | $ 42’397,326 |
| AHORRO ANUAL | $ | 25’799,994 |

Fuente: Autores

Con la aplicación de la optimización de la RCHB se ahorraría aproximadamente 59 millones de pesos anualmente en costos de caracterización de variables de la calidad y cantidad del recurso hídrico, lo que es igual a una disminución total del 31 % en los gastos destinados a esta actividad.

# 5. CONCLUSIONES

- Con la elaboración de los mapas y perfiles longitudinales se logró determinar la relación de los PM y los vertimientos para el periodo de estudio Histórico (2006 - 2016) y el periodo Total (2006 - 2018), evidenciando algunos los cambios, similitudes, tendencias y patrones en la dinámica de la CC aportada por los vertimientos entre los periodos evaluados, y que validaron y soportaron los resultados obtenidos con el algoritmo EM.

- El método de distancia de Mahalanobis evaluó la presencia de valores atípicos en los PM de la RCHB, de esta manera se obtuvo un valor de porcentaje máximo de valores atípicos de 16,67 % para el canal Torca, 22,77 % para el río Salitre, 22,02 % para el río Fucha y 17,82 % para el río Tunjuelo. El porcentaje total máximo de valores atípicos para los cuatro ríos fue de 22,77 % y está asociado con fenómenos de variabilidad climática, afluentes como tributarios y sus aportes a los ríos, humedales como amortiguamiento de las CC y estructuras de regulación, entre otras causas, generando cambios en los aportes de caudal y por consiguiente en las CC de los determinantes de la calidad del recurso hídrico.

- Con el desarrollo del método distancia Mahalanobis, se logró la detección y posterior eliminación de muestras catalogadas como atípicas para los conjuntos de datos de cada PM en el periodo de tiempo de estudio, permitiendo contar con una base de datos más precisa y confiable para análisis más representativos.

- Para determinar cuáles son las variables de CC más significativas fue usado el algoritmo RF de donde se determinaron para cada combinación de PM las 3 variables de calidad y cantidad más importantes o con mayor significancia según esta metodología, allí fue necesario proponer una metodología propia con el fin de unificar los resultados para con ello obtener un ponderado de las 4 variables de mayor importancia por río.

- Las CC más representativas de la dinámica de la calidad y cantidad del agua en los ríos de la RCHB fueron: para el canal Torca SAAM, DQO, GYA y Caudal, para el río Salitre caudal, $P_{TOTAL}$, DBO5 y DQO, para el río Fucha DBO5, Caudal, $P_{TOTAL}$ y SAAM y por último para el río Tunjuelo SST, DBO5, $P_{TOTAL}$ Y DQO.

- Mediante el uso del algoritmo de agrupamiento EM se seleccionó el mejor modelo de mixtura Gaussiano de acuerdo con la métrica BIC donde se eligió el número de *clusters* óptimo de acuerdo al PG dando como resultado un número óptimo de PM por río de 4 para Torca, 5 para Salitre, 4 para Fucha y 6 para Tunjuelo, sin embargo no se consideró viable retirar específicamente el número de PM propuesto por EM, debido a que se consideraron condiciones ambientales, antrópicas, geográficas e hidrológicas, indicando que se debía seguir un control permanente en algunos puntos.

- En los perfiles longitudinales se evidencio preliminarmente que en algunos casos entre ciertos PM se presentaban valores de CC muy similares, los cuales mostraban una línea con comportamiento casi constante, esto indicó que se podrían remover ciertos PM que no parecen ser representativos al no tener fluctuaciones considerables que monitorear. De esta manera para el canal Torca se optó por dejar el mismo número de PM, pero para el río Salitre se indicó que el PM Tv91 podría removerse, para el río Fucha Ferroca y ZFranca y para el río Tunjuelo UAN, MakroS y PteInde, además, de acuerdo con el análisis *cluste*r desarrollado con base en el algoritmo EM se determinó que en la clasificación *cluster* estos PM también presentaban valores similares. Adicionalmente también se tuvo en cuenta la cercanía de los PM con humedales, vertimientos e infraestructura que necesitará de un control ambiental continuo.

- Finalmente, se consideró que el óptimo de puntos monitoreo por río debería ser el siguiente: 4 PM para el canal Torca, 5 PM para el río Salitre, 6 M para el río Fucha y 7 PM para el río Tunjuelo, todo esto de acuerdo con los resultados y análisis presentados en esta investigación.

# 6. RECOMENDACIONES

- Se sugiere para la SDA tener en cuenta la presente investigación y que sirva como punto de partida para considerar la posible reducción del número de PM para seguimiento de la calidad y cantidad del agua en los ríos Salitre, Fucha y Tunjuelo.

- Implementar mediciones meteorológicas en el momento de la toma de muestras con datos de precipitación, temperatura y humedad, ya que esta práctica ayuda a entender e interpretar con mayor facilidad el comportamiento de las diferentes variables de calidad y cantidad del recurso hídrico, y resulta útil para diferenciar las razones o causas por las cuales una muestra pueda llegar a ser atípica, esto debido a que solo se tienen en cuenta observaciones del momento de la toma de la muestra (ej. soleado, alta nubosidad).

- Evaluar en mayor profundidad las cargas “contaminantes” que mejor representen tanto procesos hidrológicos como los usos del suelo urbano, ya que pueden llegar a caracterizar las actividades que incidan sobre el recurso hídrico y dirigir de una manera eficiente las acciones que la entidad ambiental deba tomar para su mitigación.

- Adicionalmente se recomienda a la SDA realizar seguimiento o desarrollar estudios acerca de las posibles causas de los valores atípicos, ya que son valores que brindan información importante sobre el estado de la calidad del agua y se podrían detectar irregularidades o patrones imprescindibles (condiciones hidrológicas) para lograr una mayor eficiencia en la RCHB.

## REFERENCIAS


[1] W. a. B. J. M. G. H. Qiuwen Chen, *«Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses,»* Journal of Environmental Management, vol. 110, pp. 116-124, 15 noviembre 2012.

[2] S. Behmel, M. Damour, R. Ludwig, M.J. Rodriguez, *«Water quality monitoring strategies, A review and future perspectives»*, Science of the Total Environment, pp. 1313-1315, 8 julio 2016.

[3] A. S. Y. S. Vishwanathan, *«Introduction to Machine Learning»*, Reino Unido: Cambridge University Pres, 2008.

[4] Secretaria Distrital de Ambiente, «Calidad del sistema hídrico de Bogotá», 1a ed, Editorial Pontificia Universidad Javeriana, pp. 23-27, 92, 110, Bogotá, 2008.

[5] F. Pérez, D. Zamora, «*Informe técnico: descripción y contexto de las cuencas del distrito capital (Torca, Salitre, Fucha y Tunjuelo)*», Secretaria Distrital de Ambiente, pp. 10-19, Septiembre 2015.

[6] Empresa de alcantarillado acueducto y aseo de Bogotá, «Plan de identificación de corrección de conexiones erradas», pp. 27-28, Agosto 2017.

[7] Departamento Administrativo de Función Pública. (2012, Diciembre, 21). Decreto 2667, Por el cual se reglamenta la tasa retributiva por la utilización directa e indirecta del agua como receptor de los vertimientos puntuales, y se toman otras determinaciones, Art. 28. [En línea]. Disponible: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=51042#0.

[8] Universidad de los Andes (UniAndes), *«Concentraciones de referencia para los vertimientos industriales realizados a la red de alcantarillado y de los vertimientos industriales y domésticos efectuados a cuerpos de agua de la ciudad de Bogotá»,* Informe Objetivos de Calidad, Secretaría Distrital de Ambiente - Universidad de Los Andes Convenio 045 de 2007. Numeral 5. Bogotá D.C, Colombia, 2017.

[9] B. Rocamora, A. García, B. Martínez, « *Inferencia estadística (intervalos de confianza y p-valor). Comparación de dos poblaciones (test de comparación de medias, comparación de dos proporciones, comparación de dos varianzas)*», Universidad Cardenal Herrera, pp. 6-7.

[10] Efron, B. «*Bootstrap Methods: Another Look at the Jackknife*». Ann. Statist. pp. 1-26, 1979.

[11] Vergara, J. Ramírez, J. Rojas and S. Guerrero, «*Métodos de remuestreo Bootstrap y Jackknife en análisis de sobreviviencia*», in XXVI Simposio de Estadística, Sincelejo, Sucre, 2016, p. 3.

[12] S. Gil, «*Bootstrap en poblaciones finitas*». Granada, España: Editorial Universitaria - UGR, 2014, pp. 12-14.

[13] A. Villalobos, M. Lagos and N. Gómez, «*Intervalos de confianza Bootstrap para índices de diversidad*». Chile: Universidad del Bio Bio, departamento de estadística, 2014, pp. 16-21.

[14] M. Quaglino y J. Merello, «*Métodos multivariados en estudios de vulnerabilidad social en la provincia de Santa Fe*», Noviembre 2012.

[15] S. Matsumoto, «Comparison of Outlier Detection Methods in Fault proneness Models», Proceedings of the First international Symposium on Empirical Software Engineering and Measurement, Madrid, España, Septiembre, 2007.

[16] K. Tiwary, «Selecting the Appropriate Outlier Treatment for Common Industry SAS», Conference Proceedings: NESUG 2007, Baltimore, Maryland, Noviembre 2007.

[17] Archer, V. Kimes, «*Empirical characterization of random forest variable importance measures*», Science of the Total Environment, pp. 2251-2260, 30 agosto 2007.

[18] R. F. Medina-Merino and C. I. Ñique-Chacón, «*Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python*», Interfaces, ed.10, pp. 165, 2017.

[19] A. Dempster, N. Laird y D. Rubin, «*Maximum likelihood from incomplete data via the EM algorithm*», Journal of the Royal Statistical Society, Series B (Methodological), pp. 1-38, 1977.

[20] J. Li and A. Nehorai, «*Gaussian mixture learning via adaptive hierarchical clustering*», Signal Processing, vol. 150, pp. 116-121, 2018.

[21] V. V. Raghavan, V. N. Gudivada and V. Govindaraju, «*Cognitive Computing: Theory and Applications*», Amsterdam, Netherlands, 2016. Disponible: https://doi.org/10.1016/bs.host.2016.07.005.

[22] O. Boubaker, «*Recent Advances in Chaotic Systems and Synchronization*», Reino Unido, 2019. Disponible: https://doi.org/10.1016/B978-0-12-815838-8.00007-8.

[23] Á Gómez Losada, «*Modelos De Mixturas Finitas Para La Caracterización Y Mejora De Las Redes De Monitorización De La Calidad Del Aire*», Universidad de Granada, 2014.

[24] E. J. Wegman and D. B. Carr, «*26 Statistical graphics and visualization*», *Handbook of Statistics*», vol. 9, pp. 857-958, 1993. Disponible: https://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/S0169-7161(05)80150-6.

[25] J. Han, M. Kamber and J. Pei, «*12 - Outlier Detection*», Data Mining, ed., pp. 543-584, 2012. Disponible: https://doi.org/10.1016/B978-0-12-381479-1.00012-5.

[26] DANE, «*Encuesta anual manufacturera, información por localidades*», Bogotá, Noviembre, 2006.

[27] W. K. Dodds and M. R. Whiles, «*Chapter 18 - Trophic State and Eutrophication*», Freshwater Ecology *(Third Edition)* », pp. 537-581, 2020. Disponible: https://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/B978-0-12-813255-5.00018-1.

[28] Ministerio de ambiente y desarrollo sostenible. (2018, Agosto, 03). Resolución 1484, Por el cual se asume la competencia del proyecto "Relleno sanitario Doña Juana" y se toman otras determinaciones, [En línea]. Disponible: http://www.minambiente.gov.co/images/normativa/app/resoluciones/a4-RES%201484%20DE%202018.pdf

[29] D. Puri, K. Borel, C. Vance, & R. Karthikeyan, «*Optimization of a water quality monitoring network using a spatially referenced water quality model and a genetic algorithm*». *Water*, vol. 9, no. 704, pp.1-11, 2017. Disponible: https://doi.org/10.3390/w9090704.

[30] Hilundo, M., Kelderman, P., & O´keeffe, J. H. (2008). «*Design of a water quality monitoring network for the Limpopo river basin in Mozambique*». Physics and Chemistry of the Earth, 33(8), 655-665. doi:10.1016/j.pce.2008.06.055

[31] Y.Icaga, «*Genetic algorithm usage in water quality monitoring networks optimization in Gediz (turkey) river basin*. *Environmental Monitoring and Assessment*», vol.108, pp. 261-277, 2005. Disponible: https://doi.org/10.1007/s10661-005-4328-z.

[32] Organización para la Cooperación y el Desarrollo Económico (OCDE), "¿Cómo calcular el cambio del valor del peso colombiano en el tiempo?", Colombia, 2019.

# ANEXOS

## **Anexo A.** Identificación PM de la RCHB

Tabla 15. Nombre de identificación y tramo al que pertenece de los puntos de monitoreo de la RCHB

| RÍO | PUNTO DE MONITOREO | NOMBRE DE IDENTIFICACIÓN | Tramo |
|---|---|---|---|
| **TORCA (TO)** | Bosque de Pinos | TO-BosqueP | 1 |
| | Calle 161 | TO-CL161 | |
| | Jardines de Paz | TO-Jardpaz | 2 |
| | San Simón | TO-Ssimon | |
| **SALITRE (SA)** | Parque Nacional | SA-ParqNal | 1 |
| | Arzobispo Carrera 7a | SA-Arzobis | 2 |
| | Carrera 30 Calle 53 | SA-CL53 | |
| | Carrefour Av. 68 | SA-Carrefo | 3 |
| | Transversal 91 | SA-Tv91 | 4 |
| | Salitre con Alameda | SA-Alameda | |
| **FUCHA (FU)** | El Delirio | FU-Delirio | 1 |
| | Carrera 7a Río Fucha | FU-KR7 | 2 |
| | Avenida Ferrocarril | FU-Ferroca | |
| | Fucha Avenida Las Américas | FU-America | 3 |
| | Fucha Avenida Boyacá | FU-Boyaca | |
| | Visión Colombia | FU-VisionC | 4 |
| | Fucha Zona Franca | FU-ZFranca | |
| | Fucha con Alameda | FU-Alameda | |
| **TUNJUELO (TU)** | Regadera | TU-Regader | 1 |
| | Universidad Antonio Nariño | TU-UAN | |
| | Yomasa | TU-Yomasa | 2 |
| | Doña Juana | TU-DJuana | |
| | Barrio México | TU-México | 3 |
| | San Benito | TU-SBenito | |

| | | | |
|---|---|---|---|
| | Makro Autosur | TU-MakroS | 4 |
| | Transversal 86 | TU-Tv86 | |
| | Puente Independencia | TU-PteInde | |
| | Isla Pontón San José | TU-IslaPon | |
| **BOGOTÁ** | Puente Común | BO-PComun | No aplica. |
| | Cierre | BO-Cierre | |

Fuente: Autores.

**Anexo B.** Cantidad de muestreos para las bases de datos histórica y actual en cada PM.

Tabla 16. Cantidad de muestras por punto de monitoreo

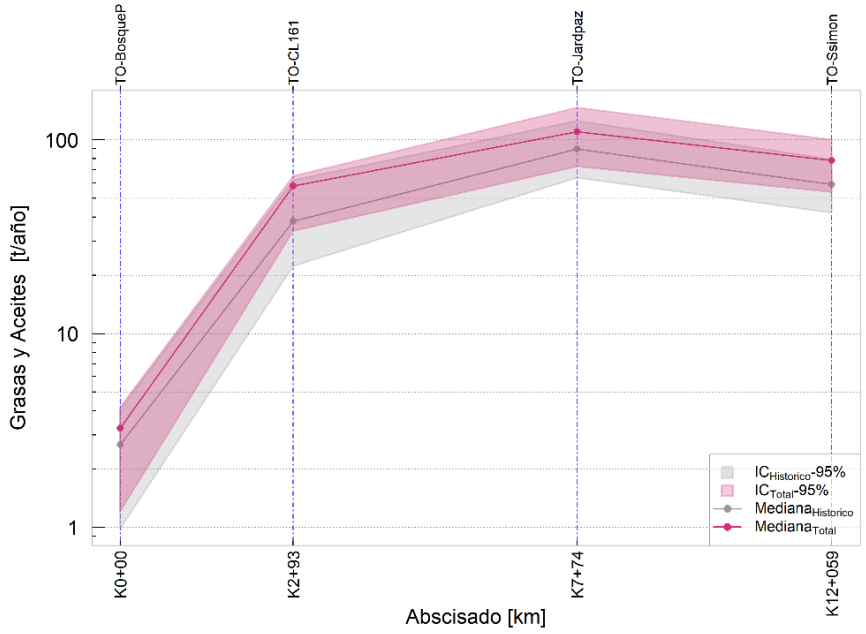
| **NÚMERO DE MUESTRAS** | | | |
|---|---|---|---|
| **Río** | **Punto de monitoreo** | **Muestras históricas (2006 - 2016)** | **Muestras (2017-2018)** |
| **TORCA (TO)** | TO-BosqueP | 60 | 12 |
| | TO-CL161 | 60 | 12 |
| | TO-Jardpaz | 61 | 12 |
| | TO-Ssimon | 72 | 12 |
| **SALITRE (SA)** | SA-ParqNal | 97 | 12 |
| | SA-Arzobis | 89 | 12 |
| | SA-CL53 | 104 | 12 |
| | SA-Carrefo | 96 | 12 |
| | SA-Tv91 | 96 | 12 |
| | SA-Alameda | 107 | 12 |
| **FUCHA (FU)** | FU-Delirio | 94 | 12 |
| | FU-KR7 | 94 | 12 |
| | FU-Ferroca | 97 | 12 |
| | FU-America | 96 | 12 |
| | FU-Boyaca | 76 | 12 |
| | FU-VisionC | 94 | 12 |
| | FU-ZFranca | 95 | 12 |
| | FU-Alameda | 96 | 12 |
| **TUNJUELO (TU)** | TU-Regader | 92 | 12 |
| | TU-UAN | 63 | 12 |
| | TU-Yomasa | 106 | 12 |
| | TU-DJuana | 107 | 12 |
| | TU-México | 107 | 12 |
| | TU-SBenito | 90 | 12 |
| | TU-MakroS | 89 | 12 |

| | | | |
|---|---|---|---|
| | TU-Tv86 | 88 | 12 |
| | TU-PteInde | 89 | 12 |
| | TU-IslaPon | 89 | 12 |

Fuente: Autores.

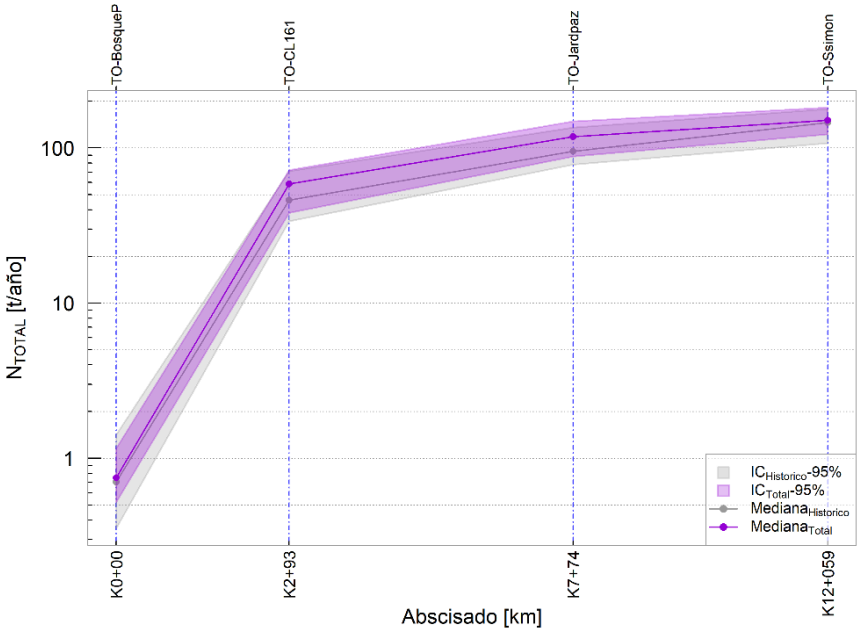
**Anexo C.** Perfiles espaciales de carga histórica contaminante para los demás determinantes de la calidad del agua por río.

Figura 42. Perfil longitudinal de carga contaminante histórica y total de GYA para el canal Torca



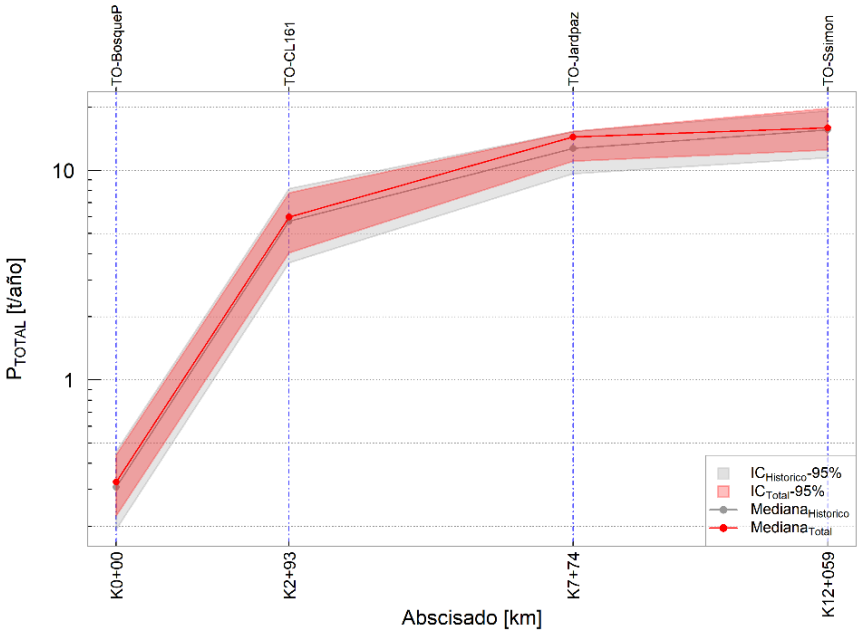

Fuente: Autores.

Figura 43. Perfil longitudinal de carga contaminante histórica y total de $N_{TOTAL}$ para el canal Torca



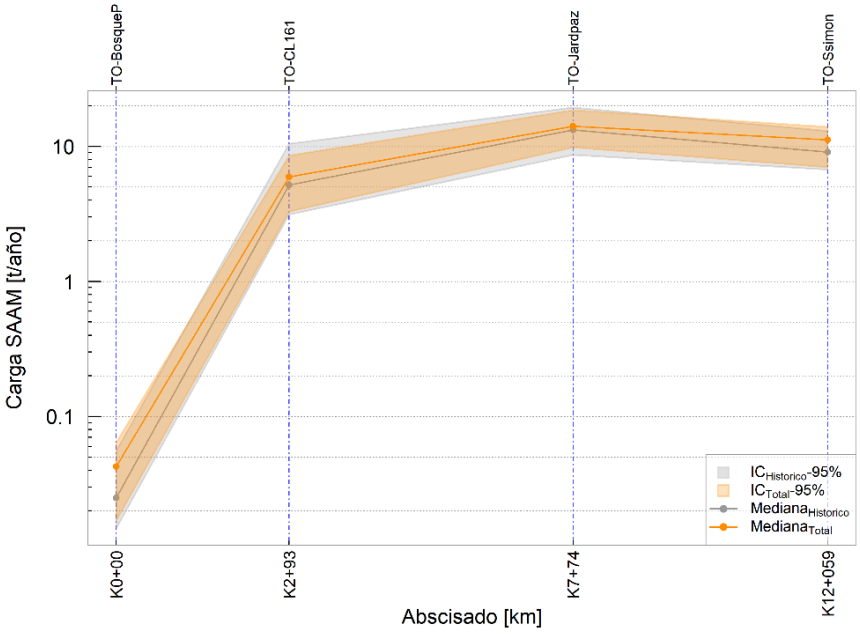

Fuente: Autores.

Figura 44. Perfil longitudinal de carga contaminante histórica y total de $P_{TOTAL}$ para el canal Torca



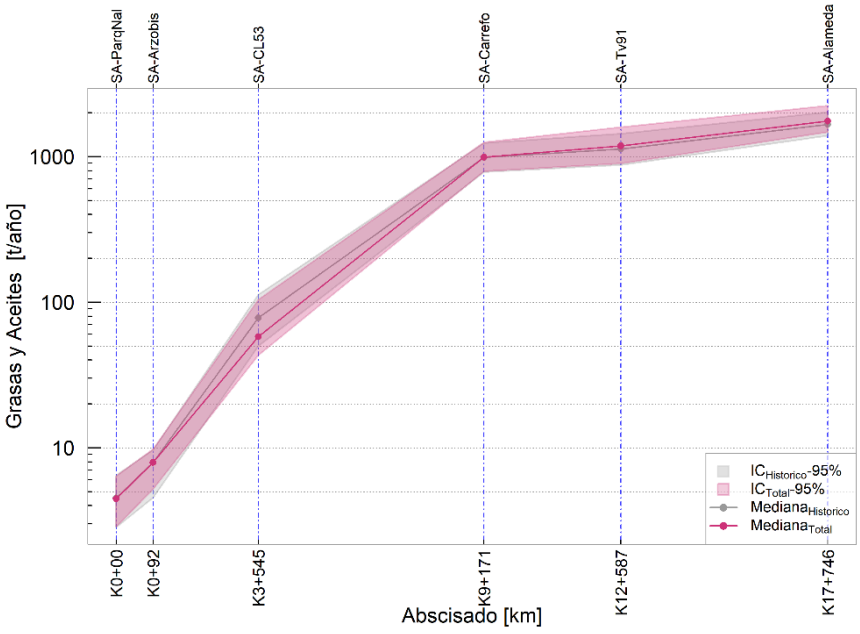

Fuente: Autores.

Figura 45. Perfil longitudinal de carga contaminante histórica y total de SAAM para el canal Torca



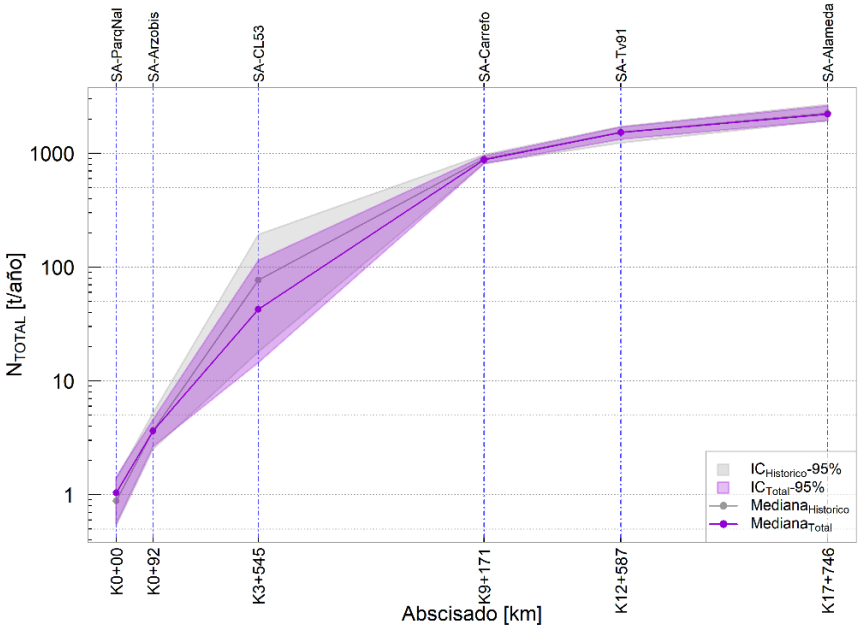

Fuente: Autores.

Figura 46. Perfil longitudinal de carga contaminante histórica y total de GYA para el río Salitre



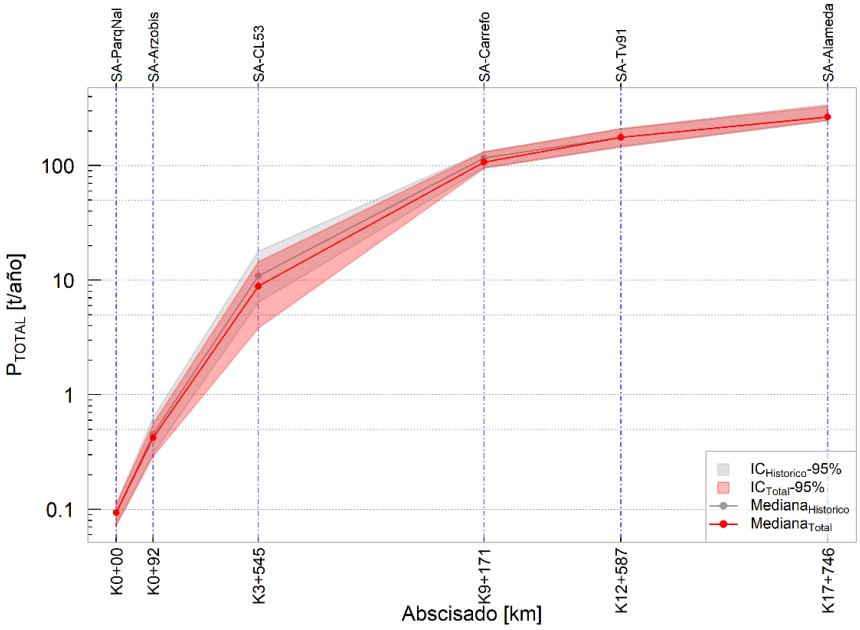

Fuente: Autores.

Figura 47. Perfil longitudinal de carga contaminante histórica y total de $N_{TOTAL}$ para el río Salitre



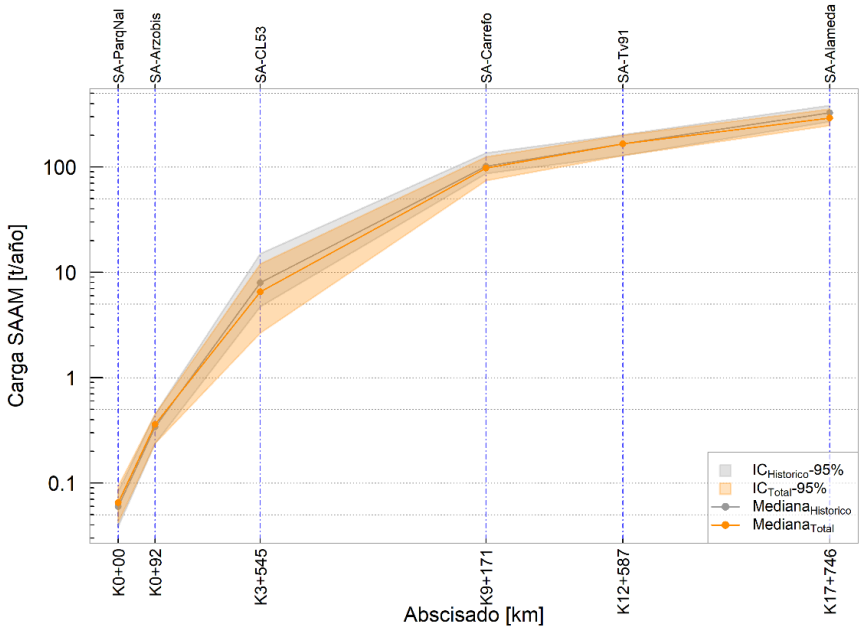

Fuente: Autores

Figura 48. Perfil longitudinal de carga contaminante histórica y total de $P_{TOTAL}$ para el río Salitre



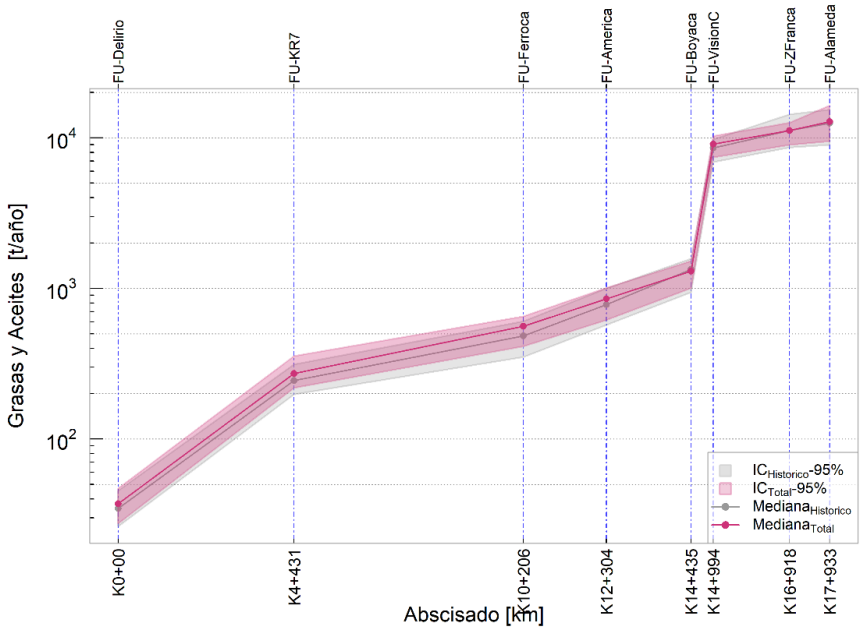

Fuente: Autores.

Figura 49. Perfil longitudinal de carga contaminante histórica y total de SAAM para el río Salitre



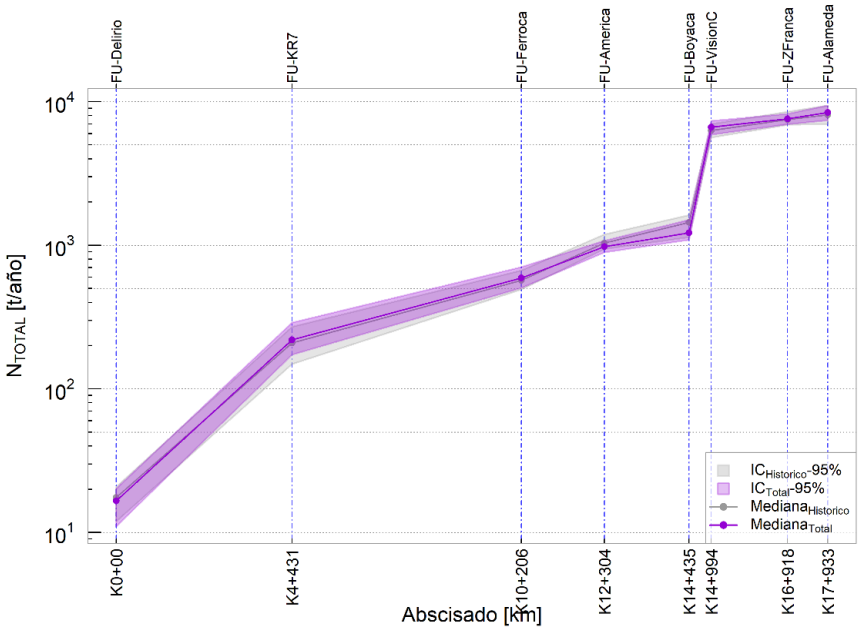

Fuente: Autores.

Figura 50. Perfil longitudinal de carga contaminante histórica y total de GYA para el río Fucha



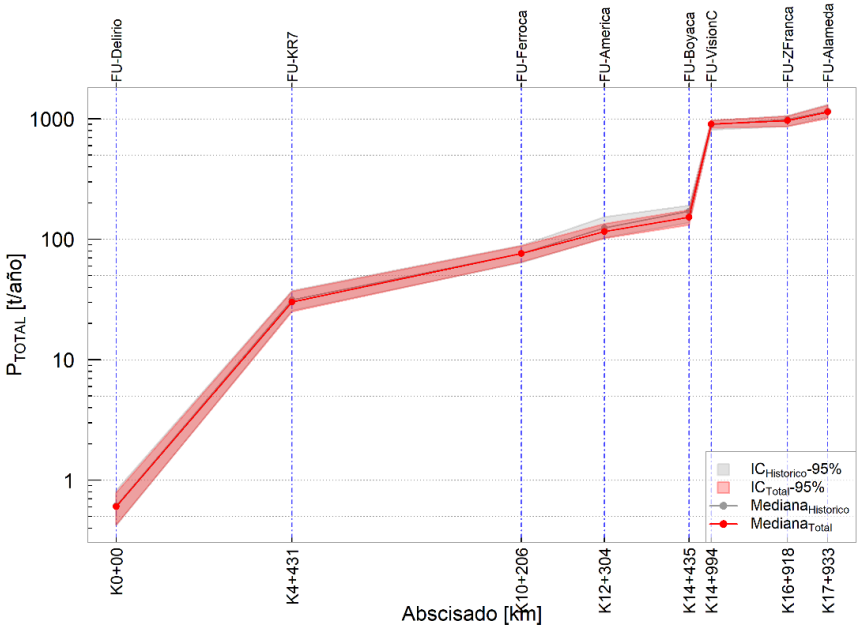

Fuente: Autores.

Figura 51. Perfil longitudinal de carga contaminante histórica y total de $N_{TOTAL}$ para el río Fucha




Fuente: Autores.

Figura 52. Perfil longitudinal de carga contaminante histórica y total de $P_{TOTAL}$ para el río Fucha




Fuente: Autores.

Figura 53. Perfil longitudinal de carga contaminante histórica y total de SAAM para el río Fucha

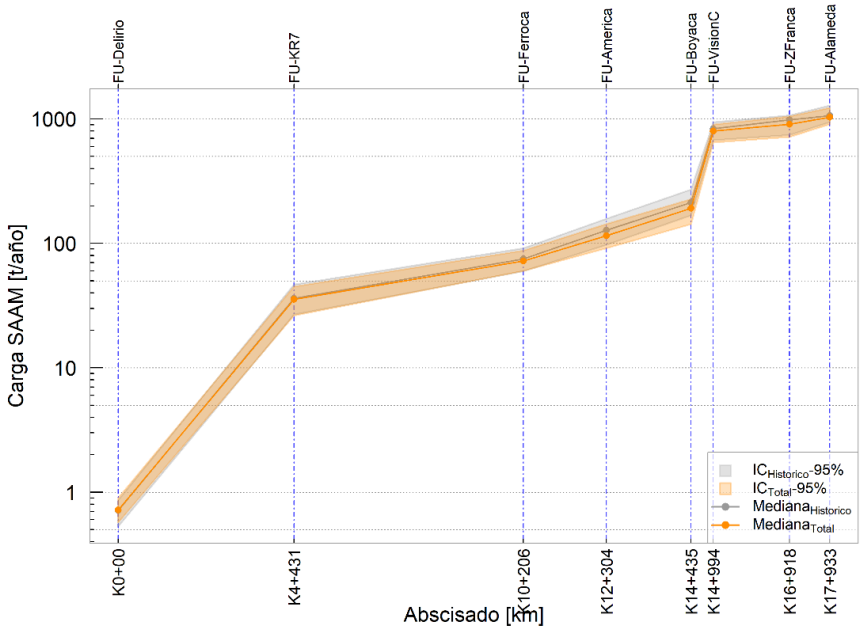
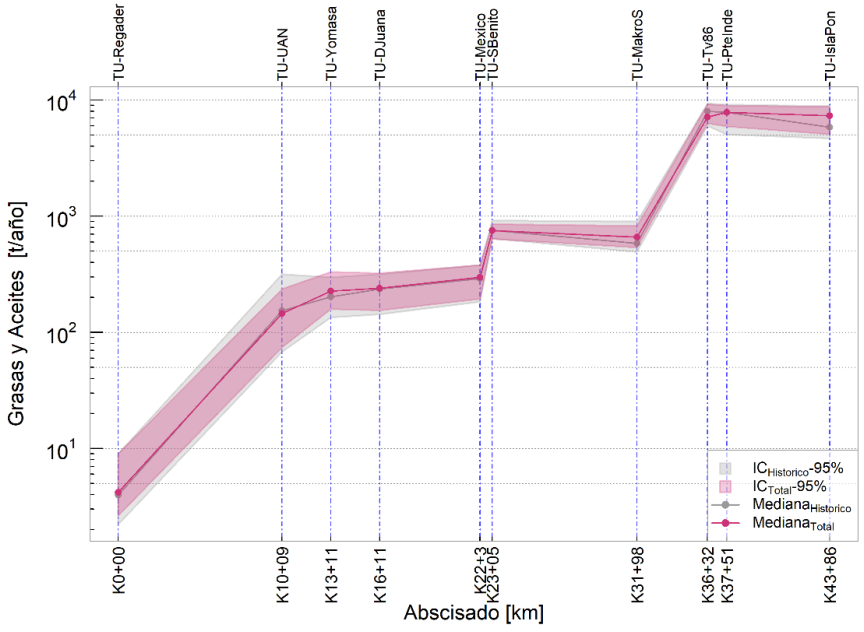
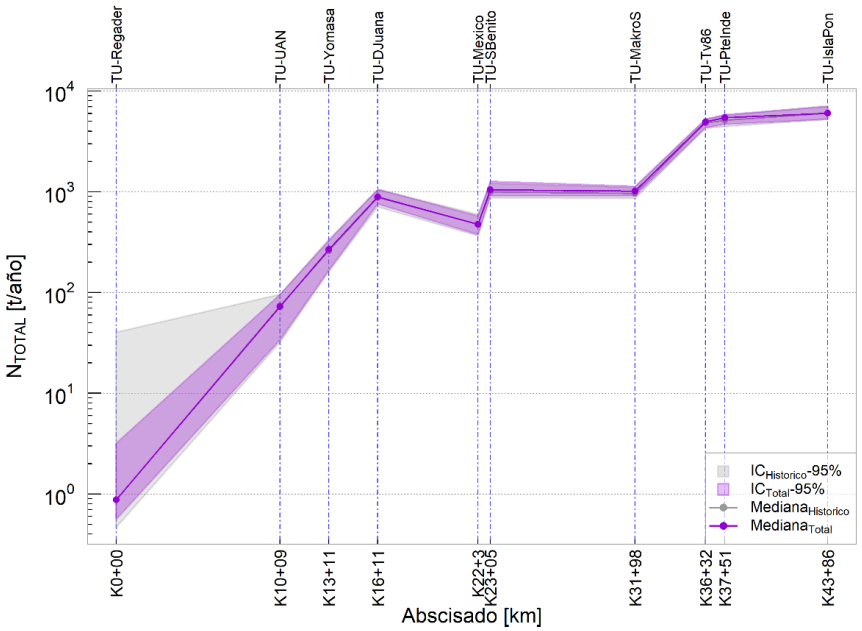



Figura 54. Perfil longitudinal de carga contaminante histórica y total de GYA para el río Tunjuelo



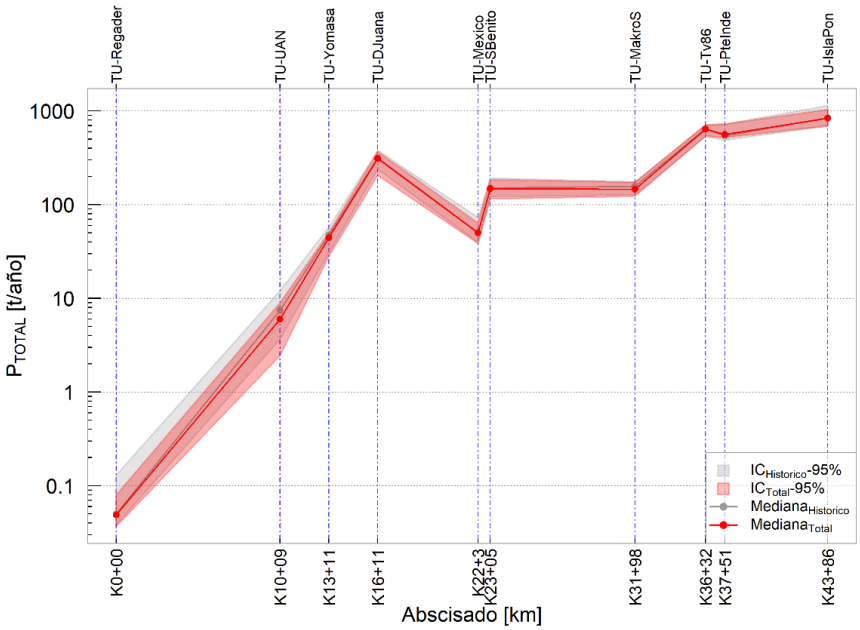
Fuente: Autores.

Figura 55. Perfil longitudinal de carga contaminante histórica y total de $N_{TOTAL}$ para el río Tunjuelo



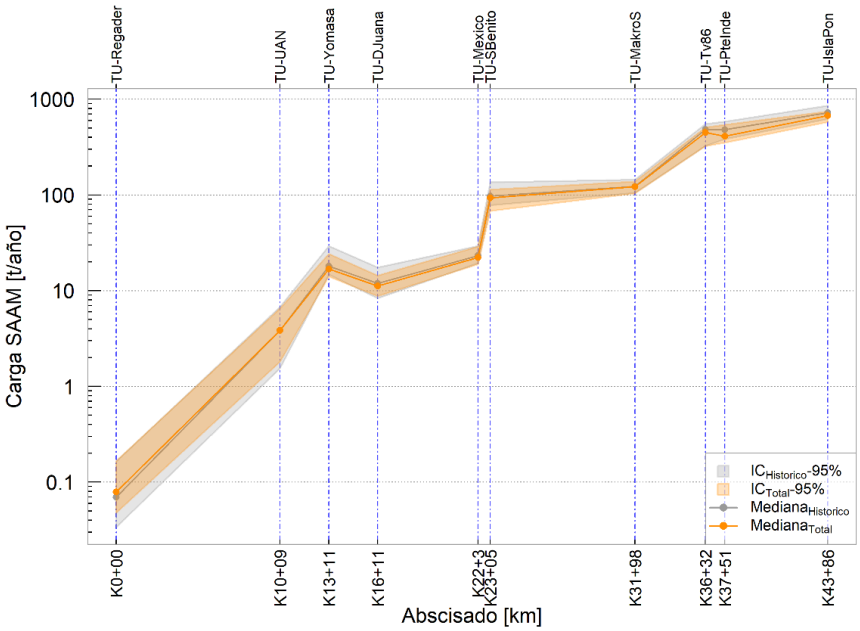

Fuente: Autores.

Figura 56. Perfil espacial de carga contaminante histórica y total de $P_{TOTAL}$ para el río Tunjuelo




Fuente: Autores.

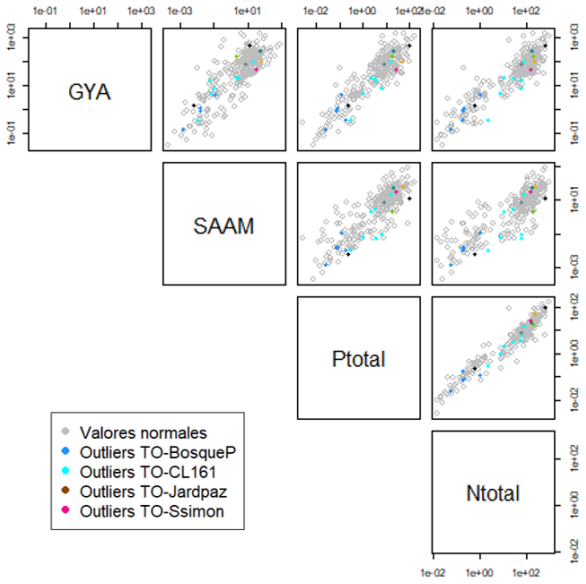
Figura 57. Perfil espacial de carga contaminante histórica y total de SAAM para el río Tunjuelo




Fuente: Autores.

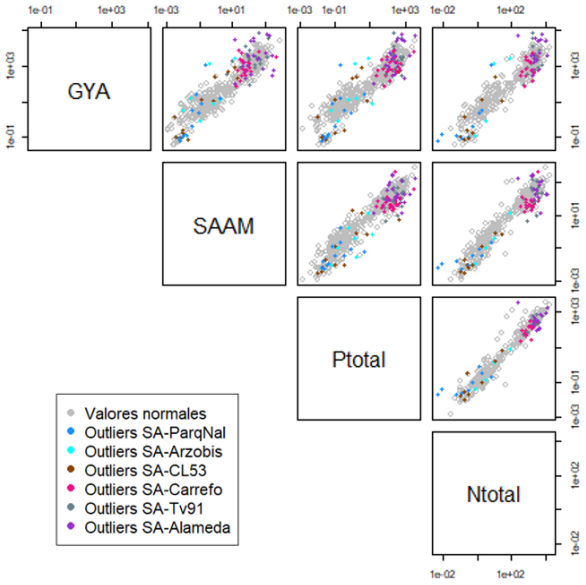
**Anexo D.** Gráficos de dispersión multivariados de carga contaminante por método Mahalanobis para los demás determinantes de la calidad del agua por río.

Figura 58. Muestras atípicas de las variables GYA, SAAM, $P_{TOTAL}$ y $N_{TOTAL}$ para cada punto de monitoreo del canal Torca



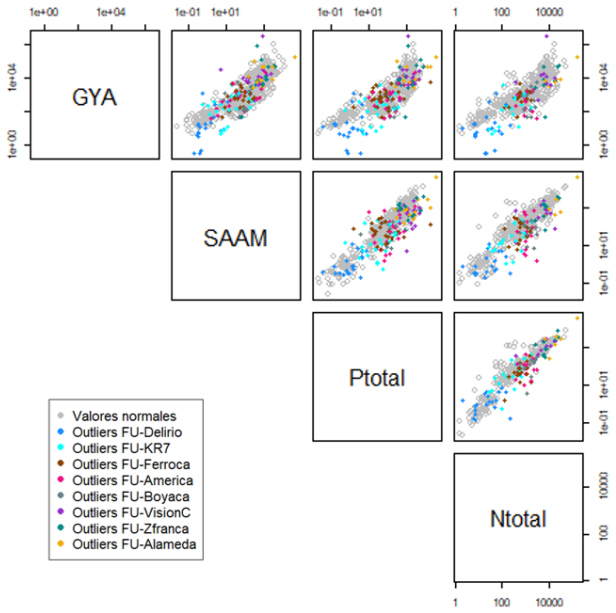

Fuente: Autores.

Figura 59. Muestras atípicas de las variables GYA, SAAM, $P_{TOTAL}$ y $N_{TOTAL}$ para cada punto de monitoreo del río Salitre



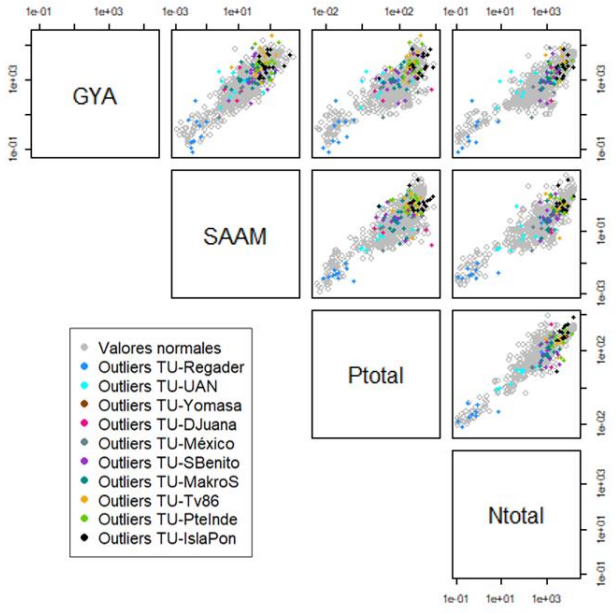

Fuente: Autores.

Figura 60. Muestras atípicas de las variables GYA, SAAM, $P_{TOTAL}$ y $N_{TOTAL}$ para cada punto de monitoreo del río Fucha




Fuente: Autores.

Figura 61. Muestras atípicas de las variables GYA, SAAM, $P_{TOTAL}$ y $N_{TOTAL}$ para cada punto de monitoreo del río Tunjuelo




Fuente: Autores.

**Anexo E**. Tabla de predicciones para cada río de acuerdo con el algoritmo *random forest*.

Tabla 17. Predicciones para el canal Torca

| TORCA | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| VALIDACIÓN | | | CALIBRACIÓN | | | TOTAL | | |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TO-BosqueP | TO-CL161 | Observed | TO-BosqueP | TO-CL161 | Observed | TO-BosqueP | TO-CL161 |
| TO-BosqueP | 6 | 0 | TO-BosqueP | 51 | 3 | TO-BosqueP | 57 | 3 |
| TO-CL161 | 0 | 5 | TO-CL161 | 7 | 50 | TO-CL161 | 7 | 55 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TO-CL161 | TO-Jardpaz | Observed | TO-CL161 | TO-Jardpaz | Observed | TO-CL161 | TO-Jardpaz |
| TO-CL161 | 3 | 2 | TO-CL161 | 42 | 15 | TO-CL161 | 45 | 17 |
| TO-Jardpaz | 0 | 5 | TO-Jardpaz | 18 | 38 | TO-Jardpaz | 18 | 43 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TO-Jardpaz | TO-Ssimon | Observed | TO-Jardpaz | TO-Ssimon | Observed | TO-Jardpaz | TO-Ssimon |
| TO-Jardpaz | 5 | 0 | TO-Jardpaz | 42 | 14 | TO-Jardpaz | 47 | 14 |
| TO-Ssimon | 3 | 2 | TO-Ssimon | 9 | 64 | TO-Ssimon | 12 | 66 |

Fuente: Autores.

Tabla 18. Predicciones para el río Salitre

| SALITRE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| VALIDACION | | | CALIBRACION | | | TOTAL | | |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | SA-Arzobis | SA-ParqNal | Observed | SA-Arzobis | SA-ParqNal | Observed | SA-Arzobis | SA-ParqNal |
| SA-Arzobis | 11 | 11 | SA-Arzobis | 59 | 18 | SA-Arzobis | 70 | 29 |
| SA-ParqNal | 7 | 13 | SA-ParqNal | 16 | 62 | SA-ParqNal | 23 | 75 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | SA-Arzobis | SA-CL53 | Observed | SA-Arzobis | SA-CL53 | Observed | SA-Arzobis | SA-CL53 |
| SA-Arzobis | 22 | 0 | SA-Arzobis | 72 | 5 | SA-Arzobis | 94 | 5 |
| SA-CL53 | 31 | 1 | SA-CL53 | 12 | 66 | SA-CL53 | 43 | 67 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | SA-Carrefo | SA-CL53 | Observed | SA-Carrefo | SA-CL53 | Observed | SA-Carrefo | SA-CL53 |
| SA-Carrefo | 12 | 0 | SA-Carrefo | 62 | 4 | SA-Carrefo | 74 | 4 |
| SA-CL53 | 0 | 32 | SA-CL53 | 5 | 73 | SA-CL53 | 5 | 105 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | SA-Carrefo | SA-Tv91 | Observed | SA-Carrefo | SA-Tv91 | Observed | SA-Carrefo | SA-Tv91 |
| SA-Carrefo | 9 | 3 | SA-Carrefo | 49 | 17 | SA-Carrefo | 58 | 20 |
| SA-Tv91 | 4 | 16 | SA-Tv91 | 8 | 12 | SA-Tv91 | 12 | 28 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | SA-Alameda | SA-Tv91 | Observed | SA-Alameda | SA-Tv91 | Observed | SA-Alameda | SA-Tv91 |
| SA-Alameda | 22 | 6 | SA-Alameda | 44 | 23 | SA-Alameda | 66 | 29 |
| SA-Tv91 | 8 | 12 | SA-Tv91 | 25 | 43 | SA-Tv91 | 33 | 55 |

Fuente: Autores.

Tabla 19. Predicciones para el río Fucha

| FUCHA | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| VALIDACIÓN | | | CALIBRACIÓN | | | TOTAL | | |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-Delirio | FU-KR7 | Observed | FU-Delirio | FU-KR7 | Observed | FU-Delirio | FU-KR7 |
| FU-Delirio | 13 | 0 | FU-Delirio | 72 | 0 | FU-Delirio | 85 | 0 |
| FU-KR7 | 1 | 15 | FU-KR7 | 0 | 68 | FU-KR7 | 1 | 83 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-Ferroca | FU-KR7 | Observed | FU-Ferroca | FU-KR7 | Observed | FU-Ferroca | FU-KR7 |
| FU-Ferroca | 10 | 2 | FU-Ferroca | 58 | 15 | FU-Ferroca | 68 | 17 |
| FU-KR7 | 6 | 10 | FU-KR7 | 19 | 49 | FU-KR7 | 25 | 59 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-America | FU-Ferroca | Observed | FU-America | FU-Ferroca | Observed | FU-America | FU-Ferroca |
| FU-America | 9 | 8 | FU-America | 46 | 26 | FU-America | 55 | 34 |
| FU-Ferroca | 7 | 5 | FU-Ferroca | 28 | 45 | FU-Ferroca | 35 | 50 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-America | FU-Boyaca | Observed | FU-America | FU-Boyaca | Observed | FU-America | FU-Boyaca |
| FU-America | 15 | 2 | FU-America | 52 | 20 | FU-America | 67 | 22 |
| FU-Boyaca | 13 | 6 | FU-Boyaca | 25 | 29 | FU-Boyaca | 38 | 35 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-Boyaca | FU-VisionC | Observed | FU-Boyaca | FU-VisionC | Observed | FU-Boyaca | FU-VisionC |
| FU-Boyaca | 17 | 2 | FU-Boyaca | 52 | 2 | FU-Boyaca | 69 | 4 |
| FU-VisionC | 0 | 18 | FU-VisionC | 2 | 69 | FU-VisionC | 2 | 87 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-VisionC | FU-ZFranca | Observed | FU-VisionC | FU-ZFranca | Observed | FU-VisionC | FU-ZFranca |
| FU-VisionC | 11 | 7 | FU-VisionC | 40 | 31 | FU-VisionC | 51 | 38 |
| FU-ZFranca | 9 | 9 | FU-ZFranca | 30 | 45 | FU-ZFranca | 39 | 54 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | FU-Alameda | FU-ZFranca | Observed | FU-Alameda | FU-ZFranca | Observed | FU-Alameda | FU-ZFranca |
| FU-Alameda | 5 | 15 | FU-Alameda | 34 | 41 | FU-Alameda | 39 | 56 |
| FU-ZFranca | 4 | 14 | FU-ZFranca | 33 | 42 | FU-ZFranca | 37 | 56 |

Fuente: Autores.

Tabla 20. Predicciones para el río Tunjuelo

| TUNJUELO | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| VALIDACIÓN | | | CALIBRACIÓN | | | TOTAL | | |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-Regader | TU-UAN | Observed | TU-Regader | TU-UAN | Observed | TU-Regader | TU-UAN |
| TU-Regader | 14 | 2 | TU-Regader | 52 | 19 | TU-Regader | 66 | 21 |
| TU-UAN | 2 | 10 | TU-UAN | 13 | 41 | TU-UAN | 15 | 51 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-UAN | TU-Yomasa | Observed | TU-UAN | TU-Yomasa | Observed | TU-UAN | TU-Yomasa |
| TU-UAN | 12 | 0 | TU-UAN | 49 | 5 | TU-UAN | 61 | 5 |
| TU-Yomasa | 10 | 25 | TU-Yomasa | 4 | 79 | TU-Yomasa | 14 | 104 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-DJuana | TU-Yomasa | Observed | TU-DJuana | TU-Yomasa | Observed | TU-DJuana | TU-Yomasa |
| TU-DJuana | 18 | 18 | TU-DJuana | 69 | 10 | TU-DJuana | 87 | 28 |
| TU-Yomasa | 2 | 33 | TU-Yomasa | 13 | 70 | TU-Yomasa | 15 | 103 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-DJuana | TU-Mexico | Observed | TU-DJuana | TU-Mexico | Observed | TU-DJuana | TU-Mexico |
| TU-DJuana | 21 | 15 | TU-DJuana | 67 | 12 | TU-DJuana | 88 | 27 |
| TU-Mexico | 2 | 29 | TU-Mexico | 13 | 67 | TU-Mexico | 15 | 96 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-Mexico | TU-SBenito | Observed | TU-Mexico | TU-SBenito | Observed | TU-Mexico | TU-SBenito |
| TU-Mexico | 29 | 2 | TU-Mexico | 61 | 19 | TU-Mexico | 90 | 21 |
| TU-SBenito | 8 | 2 | TU-SBenito | 15 | 58 | TU-SBenito | 23 | 60 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-MakroS | TU-SBenito | Observed | TU-MakroS | TU-SBenito | Observed | TU-MakroS | TU-SBenito |
| TU-MakroS | 6 | 6 | TU-MakroS | 36 | 33 | TU-MakroS | 42 | 39 |
| TU-SBenito | 0 | 10 | TU-SBenito | 26 | 47 | TU-SBenito | 26 | 57 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-MakroS | TU-Tv86 | Observed | TU-MakroS | TU-Tv86 | Observed | TU-MakroS | TU-Tv86 |
| TU-MakroS | 12 | 0 | TU-MakroS | 60 | 9 | TU-MakroS | 72 | 9 |
| TU-Tv86 | 1 | 16 | TU-Tv86 | 4 | 64 | TU-Tv86 | 5 | 80 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-PteInde | TU-Tv86 | Observed | TU-PteInde | TU-Tv86 | Observed | TU-PteInde | TU-Tv86 |
| TU-PteInde | 6 | 12 | TU-PteInde | 38 | 29 | TU-PteInde | 44 | 41 |
| TU-Tv86 | 5 | 12 | TU-Tv86 | 30 | 38 | TU-Tv86 | 35 | 50 |
| | Predicted | | | Predicted | | | Predicted | |
| Observed | TU-IslaPon | TU-PteInde | Observed | TU-IslaPon | TU-PteInde | Observed | TU-IslaPon | TU-PteInde |
| TU-IslaPon | 7 | 8 | TU-IslaPon | 44 | 24 | TU-IslaPon | 51 | 32 |
| TU-PteInde | 8 | 10 | TU-PteInde | 24 | 43 | TU-PteInde | 32 | 53 |

Fuente: Autores.

**Anexo F**. Variables seleccionadas por cada pareja de puntos de monitoreo por río de acuerdo con el Gini index general.

Tabla 21. Gini index general para las variables de calidad y cantidad por pareja de puntos de monitoreo para el canal Torca

| GINI INDEX GENERAL | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| | 13.11 | 12.53 | 12.20 |
| **BosqueP-CL161** | DBO5 | SAAM | DQO |
| | 12.23 | 9.20 | 8.59 |
| **CL161-Jardpaz** | Caudal | DQO | SAAM |
| | 11.63 | 9.81 | 9.25 |
| **Jardpaz-Ssimon** | DBO5 | GYA | DQO |

Fuente: Autores.

Tabla 22. Gini index general para las variables de calidad y cantidad por pareja de PM para el río Salitre

| GINI INDEX GENERAL | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| | 14.72 | 14.34 | 12.28 |
| **ParqNal-Arzobis** | SST | DQO | DBO5 |
| | 22.09 | 13.84 | 12.94 |
| **Arzobis-CL53** | $P_{TOTAL}$ | DBO5 | GYA |
| | 20.52 | 14.82 | 11.75 |
| **CL53-Carrefo** | Caudal | DBO5 | $P_{TOTAL}$ |
| | 19.40 | 8.88 | 8.84 |
| **Carrefo-TV91** | Caudal | $P_{TOTAL}$ | SAAM |
| | 11.77 | 11.18 | 10.66 |
| **TV91-Alameda** | $P_{TOTAL}$ | Caudal | SAAM |

Fuente: Autores.

Tabla 23. Gini index general para las variables de calidad y cantidad por pareja de puntos de monitoreo para el río Fucha

| GINI INDEX GENERAL | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| | 21.97 | 16.03 | 14.34 |
| **Delirio-KR7** | DBO5 | $P_{TOTAL}$ | SAAM |
| | 18.6 | 12.83 | 9.76 |
| **KR7-FERROCA** | $P_{TOTAL}$ | DBO5 | SAAM |
| | 12.85 | 11.4 | 11.04 |
| **Ferroca-America** | $P_{TOTAL}$ | DBO5 | Caudal |

| | | | |
|---|---|---|---|
| | 10.69 | 10.24 | 8.39 |
| **America-Boyacá** | DQO | SAAM | GYA |
| | 19.02 | 13.32 | 10.93 |
| **Boyacá-VisionC** | $P_{TOTAL}$ | Caudal | DBO5 |
| | 15.12 | 10.74 | 9.92 |
| **VisionC-ZFranca** | Caudal | GYA | $P_{TOTAL}$ |
| | 12.62 | 11.35 | 10.74 |
| **Zfranca-Alameda** | Caudal | SST | DBO5 |

Fuente: Autores.

Tabla 24. Gini index general para las variables de calidad y cantidad por pareja de puntos de monitoreo para el río Tunjuelo

| **GINI INDEX GENERAL** | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| | 11.90 | 9.94 | 9.50 |
| **Regader-UAN** | Caudal | $P_{TOTAL}$ | SST |
| | 16.43 | 11.53 | 8.73 |
| **UAN-Yomasa** | DBO5 | $P_{TOTAL}$ | SAAM |
| | 24.67 | 19.26 | 11.67 |
| **Yomasa-Djuana** | $P_{TOTAL}$ | SST | DQO |
| | 28.22 | 17.59 | 8.12 |
| **Djuana-México** | SST | $P_{TOTAL}$ | DBO5 |
| | 22.07 | 12.34 | 12.12 |
| **México-SBenito** | DBO5 | SAAM | DQO |
| | 12.83 | 10.76 | 10.45 |
| **SBenito-MakroS** | SST | GYA | SAAM |
| | 22.40 | 16.25 | 9.54 |
| **MakroS-Tv86** | DBO5 | DQO | GYA |
| | 10.83 | 10.48 | 10.19 |
| **Tv86-PteInde** | DQO | SST | SAAM |
| | 12.09 | 10.37 | 9.71 |
| **PteInde-IslaPon** | Caudal | DQO | GYA |

Fuente: Autores.

**Anexo G.** Variables seleccionadas por cada pareja de puntos de monitoreo por río de acuerdo Gini index específico.

Tabla 25. Gini index específico para las variables de calidad y cantidad por pareja de puntos de monitoreo para el canal Torca

| GINI INDEX ESPECÍFICO | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| **BosqueP-CL161** | 19.25 | 17.98 | 17.71 |
| | SAAM | DQO | DBO5 |
| **CL161-Jardpaz** | 19.35 | 19.20 | 17.45 |
| | Caudal | DQO | SAAM |
| **Jardpaz-Ssimon** | 21.76 | 21.63 | 19.48 |
| | GYA | DBO5 | DQO |

Fuente: Autores.

Tabla 26. Gini index específico para las variables de calidad y cantidad por pareja de puntos de monitoreo para el río Salitre

| GINI INDEX ESPECÍFICO | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| **ParqNal-Arzobis** | 25.77 | 25.63 | 25.11 |
| | DQO | SST | DBO5 |
| **Arzobis-CL53** | 29.01 | 24.87 | 23.16 |
| | $P_{TOTAL}$ | DBO5 | GYA |
| **CL53-Carrefo** | 27.26 | 23.52 | 20.21 |
| | Caudal | DBO5 | $P_{TOTAL}$ |
| **Carrefo-TV91** | 27.64 | 20.45 | 18.40 |
| | Caudal | SAAM | $P_{TOTAL}$ |
| **TV91-Alameda** | 23.02 | 22.56 | 21.43 |
| | Caudal | $P_{TOTAL}$ | SAAM |

Fuente: Autores.

Tabla 27. Gini index específico para las variables de calidad y cantidad por pareja de puntos de monitoreo para el río Fucha

| GINI INDEX ESPECÍFICO | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| **Delirio-KR7** | 23.99 | 23.74 | 21.69 |
| | DBO5 | SAAM | $P_{TOTAL}$ |
| **KR7-FERROCA** | 27.94 | 22.95 | 19.04 |
| | $P_{TOTAL}$ | DBO5 | SAAM |
| **Ferroca-America** | 24.11 | 24.05 | 23.86 |

| | | | |
|---|---|---|---|
| | DBO5 | Caudal | $P_{TOTAL}$ |
| | 21.27 | 20.25 | 19.66 |
| **America-Boyacá** | DQO | SAAM | GYA |
| | 22.15 | 20.76 | 17.86 |
| **Boyacá-VisionC** | $P_{TOTAL}$ | DBO5 | Caudal |
| | 26.74 | 23.24 | 22.52 |
| **VisionC-ZFranca** | Caudal | GYA | $P_{TOTAL}$ |
| | 25.61 | 24.51 | 24.38 |
| **Zfranca-Alameda** | Caudal | DBO5 | SST |

Fuente: Autores.

Tabla 28. Gini index específico para las variables de calidad y cantidad por pareja de puntos de monitoreo para el río Tunjuelo

| **GINI INDEX ESPECÍFICO** | | | |
|---|---|---|---|
| **Punto de monitoreo** | **Puntuación Variables** | | |
| | 22.30 | 19.98 | 18.54 |
| **Regader-UAN** | Caudal | SST | $P_{TOTAL}$ |
| | 23.89 | 21.12 | 19.88 |
| **UAN-Yomasa** | DBO5 | $P_{TOTAL}$ | SAAM |
| | 31.69 | 21.12 | 21.71 |
| **Yomasa-Djuana** | $P_{TOTAL}$ | SST | DQO |
| | 35.20 | 26.70 | 17.06 |
| **Djuana-México** | SST | $P_{TOTAL}$ | DBO5 |
| | 30.95 | 22.68 | 22.19 |
| **México-SBenito** | DBO5 | SAAM | DQO |
| | 24.30 | 23.81 | 23.38 |
| **SBenito-MakroS** | SST | GYA | SAAM |
| | 26.31 | 24.11 | 18.52 |
| **MakroS-Tv86** | DBO5 | DQO | GYA |
| | 23.40 | 22.10 | 21.51 |
| **Tv86-PteInde** | DQO | SST | SAAM |
| | 23.66 | 23.52 | 19.83 |
| **PteInde-IslaPon** | Caudal | DQO | GYA |

Fuente: Autores.

**Anexo H.** Metodología de selección de variables más representativas de la dinámica hidrológica por río.

Tabla 29. Magnitudes por variables y su correspondiente valor multiplicativo de acuerdo con su nivel de importancia para el canal Torca

| *1 | *0.5 | *0.25 | #1 | #2 | #3 |
|---|---|---|---|---|---|
| **VALORES MULTIPLICATIVOS** | | | | | |
| 19.25 | 8.99 | 4.43 | SAAM | DQO | DBO5 |
| 19.35 | 9.60 | 4.36 | Caudal | DQO | SAAM |
| 21.76 | 10.81 | 4.87 | GYA | DBO5 | DQO |

Fuente: Autores.

Tabla 30. Sumatoria de valores correspondientes para cada variable para el canal Torca

| | | **SUMA** | | | | |
|---|---|---|---|---|---|---|
| **CAUDAL** | **DBO5** | **DQO** | **SST** | **$P_{TOTAL}$** | **GYA** | **SAAM** |
| 19.35 | 10.81 | 8.99 | | | 21.76 | 19.25 |
| | 4.43 | 9.60 | | | | 4.36 |
| | | 4.87 | | | | |
| 19.35 | 15.24 | 23.46 | 0.00 | 0.00 | 21.76 | 23.61 |

Fuente: Autores.

Tabla 31. Variables organizadas de mayor a menor magnitud para el canal Torca

| | **ORDEN DE MAYOR A MENOR PUNTUACIÓN** | | | | | |
|---|---|---|---|---|---|---|
| 23.61 | 23.46 | 21.76 | 19.35 | 15.24 | 0.00 | 0.00 |
| SAAM | DQO | GYA | CAUDAL | DBO5 | SST | $P_{TOTAL}$ |

Fuente: Autores.

Tabla 32. Magnitudes por variables y su correspondiente valor multiplicativo de acuerdo con su nivel de importancia para el río Salitre

| *1 | *0.5 | *0.25 | #1 | #2 | #3 |
|---|---|---|---|---|---|
| **VALORES MULTIPLICATIVOS** | | | | | |
| 25.77 | 12.81 | 6.28 | DQO | SST | DBO5 |
| 29.01 | 12,43 | 5.79 | $P_{TOTAL}$ | DBO5 | GYA |
| 27.26 | 11.76 | 5.05 | Caudal | DBO5 | $P_{TOTAL}$ |
| 27.64 | 10.23 | 4.60 | Caudal | SAAM | $P_{TOTAL}$ |
| 23.02 | 11.28 | 5.36 | Caudal | $P_{TOTAL}$ | SAAM |

Fuente: Autores.

Tabla 33. Sumatoria de valores correspondientes para cada variable para el río Salitre

| | | SUMA | | | | |
|---|---|---|---|---|---|---|
| **CAUDAL** | **DBO5** | **DQO** | **SST** | **$P_{TOTAL}$** | **GYA** | **SAAM** |
| 27.26 | 12.43 | 25.77 | 12.81 | 29.01 | 5.79 | 10.23 |
| 27.64 | 11.76 | | | 11.28 | | 5.36 |
| 23.02 | 6.28 | | | 5.05 | | |
| | | | | 4.60 | | |
| 77.92 | 30.47 | 25.77 | 12.81 | 49.94 | 5.79 | 15.58 |

Fuente: Autores.

Tabla 34. Variables organizadas de mayor a menor magnitud para el río Salitre

| | ORDEN DE MAYOR A MENOR PUNTUACIÓN | | | | | |
|---|---|---|---|---|---|---|
| 77.92 | 49.94 | 30.47 | 25.77 | 15.58 | 12.81 | 5.79 |
| **CAUDAL** | **$P_{TOTAL}$** | **DBO5** | **DQO** | SAAM | SST | GYA |

Fuente: Autores.

Tabla 35. Magnitudes por variables y su correspondiente valor multiplicativo de acuerdo con su nivel de importancia para el río Fucha

| *1 | *0.5 | *0.25 | #1 | #2 | #3 |
|---|---|---|---|---|---|
| **VALORES MULTIPLICATIVOS** | | | | | |
| 23.99 | 11.87 | 5.42 | DBO5 | SAAM | $P_{TOTAL}$ |
| 27.94 | 11.48 | 4.76 | $P_{TOTAL}$ | DBO5 | SAAM |
| 24.11 | 12.03 | 5.97 | DBO5 | Caudal | $P_{TOTAL}$ |
| 21.7 | 10.12 | 4.92 | DQO | SAAM | GYA |
| 22.15 | 10.38 | 4.46 | $P_{TOTAL}$ | DBO5 | Caudal |
| 26.74 | 11.62 | 5.63 | Caudal | GYA | $P_{TOTAL}$ |
| 25.61 | 12.26 | 6.10 | Caudal | DBO5 | SST |

Fuente: Autores.

Tabla 36. Sumatoria de valores correspondientes para cada variable para el río Fucha

| | | SUMA | | | | |
|---|---|---|---|---|---|---|
| **CAUDAL** | **DBO5** | **DQO** | **SST** | **$P_{TOTAL}$** | **GYA** | **SAAM** |
| 26.74 | 23.99 | 21.27 | 6.10 | 27.94 | 11.62 | 11.87 |
| 25.61 | 24.11 | | | 22.15 | 4.92 | 10.12 |
| 12.03 | 11.48 | | | 5.42 | | 4.76 |
| 4.46 | 10.38 | | | 5.97 | | |
| | 12.26 | | | 5.63 | | |
| 68.83 | 82.21 | 21.27 | 6.10 | 67.11 | 16.53 | 26.75 |

Fuente: Autores.

Tabla 37. Variables organizadas de mayor a menor magnitud para el río Fucha

| ORDEN DE MAYOR A MENOR PUNTUACIÓN | | | | | | |
|---|---|---|---|---|---|---|
| 82.21 | 68.83 | 67.11 | 26.75 | 21.27 | 16.53 | 6.10 |
| **DBO5** | **CAUDAL** | **$P_{TOTAL}$** | **SAAM** | DQO | GYA | SST |

Fuente: Autores.

Tabla 38. Magnitudes por variables y su correspondiente valor multiplicativo de acuerdo con su nivel de importancia para el río Tunjuelo

| *1 | *0.5 | *0.25 | #1 | #2 | #3 |
|---|---|---|---|---|---|
| **VALORES MULTIPLICATIVOS** | | | | | |
| 22.30 | 9.99 | 4.63 | Caudal | SST | $P_{TOTAL}$ |
| 23.89 | 10.56 | 4.97 | DBO5 | $P_{TOTAL}$ | SAAM |
| 31.69 | 10.56 | 5.43 | $P_{TOTAL}$ | SST | DQO |
| 35.20 | 13.35 | 4.26 | SST | $P_{TOTAL}$ | DBO5 |
| 30.95 | 11.34 | 5.55 | DBO5 | SAAM | DQO |
| 24.30 | 11.90 | 5.84 | SST | GYA | SAAM |
| 26.31 | 12.05 | 4.63 | DBO5 | DQO | GYA |
| 23.40 | 11.05 | 5.38 | DQO | SST | SAAM |
| 23.66 | 11.76 | 4.96 | Caudal | DQO | GYA |

Fuente: Autores.

Tabla 39. Sumatoria de valores correspondientes para cada variable para el río Tunjuelo

| | | SUMA | | | | |
|---|---|---|---|---|---|---|
| **CAUDAL** | **DBO5** | **DQO** | **SST** | **$P_{TOTAL}$** | **GYA** | **SAAM** |
| 22.30 | 23.89 | 23.40 | 35.20 | 31.69 | 11.90 | 11.34 |
| 23.66 | 30.95 | 12.05 | 24.30 | 10.56 | 4.63 | 4.97 |
| | 26.31 | 11.76 | 9.99 | 13.35 | 4.96 | 5.84 |
| | 4.26 | 5.43 | 10.56 | 4.63 | | 5.38 |
| | | 5.55 | 11.05 | | | |
| 45.97 | 85.42 | 58.19 | 91.11 | 60.24 | 21.49 | 27.53 |

Fuente: Autores.

Tabla 40. Variables organizadas de mayor a menor magnitud para el río Tunjuelo

| ORDEN DE MAYOR A MENOR PUNTUACIÓN | | | | | | |
|---|---|---|---|---|---|---|
| 91.11 | 85.42 | 60.24 | 58.19 | 45.97 | 27.53 | 21.49 |
| **SST** | **DBO5** | **$P_{TOTAL}$** | **DQO** | CAUDAL | SAAM | GYA |

Fuente: Autores.

**Anexo I.** Valores medios de las variables de calidad y cantidad más importantes por río según análisis *cluster* para el periodo de calibración.

Tabla 41. Caudales y cargas medias de los clusters conformados para canal Torca en el periodo de calibración

| | CANAL TORCA | | | |
|---|---|---|---|---|
| | [$m^3/s$] | Carga media de los cluster [t/año] | | |
| Clusters | Caudal | DQO | GYA | SAAM |
| 1 | 0.02 | 4.76 | 3.43 | 0.05 |
| 2 | 0.13 | 335.71 | 48.52 | 8.19 |
| 3 | 1.6 | 12246.87 | 372.94 | 219.89 |
| 4 | 0.35 | 1134.15 | 168.03 | 32.34 |

Fuente: Autores.

Tabla 42.Caudales y cargas medias de los clusters conformados para río Salitre en el periodo de calibración

| | RÍO SALITRE | | | |
|---|---|---|---|---|
| | [$m^3/s$] | Carga media de los cluster [t/año] | | |
| Clusters | Caudal | DBO5 | DQO | $P_{TOTAL}$ |
| 1 | 0.03 | 2.03 | 16.63 | 0.17 |
| 2 | 5.59 | 22051.57 | 83037.72 | 774.22 |
| 3 | 0.08 | 46.28 | 222.52 | 2.93 |
| 4 | 0.51 | 1755.82 | 6331.54 | 60.67 |
| 5 | 2.07 | 4467.66 | 11394.11 | 181.94 |

Fuente: Autores.

Tabla 43. Caudales y cargas medias de los clusters conformados para río Fucha en el periodo de calibración

| | RÍO FUCHA | | | |
|---|---|---|---|---|
| | [$m^3/s$] | Carga media de los cluster [t/año] | | |
| Clusters | Caudal | DBO5 | SAAM | $P_{TOTAL}$ |
| 1 | 5.27 | 24664.26 | 954.57 | 791.55 |
| 2 | 0.33 | 13.88 | 1.01 | 0.78 |
| 3 | 1.05 | 2694.77 | 113.45 | 90.89 |
| 4 | 8.9 | 83874.76 | 2098.25 | 2251.43 |

Fuente: Autores.

Tabla 44. Caudales y cargas medias de los clusters conformados para río Tunjuelo en el periodo de calibración

| | RÍO TUNJUELO | | | |
|---|---|---|---|---|
| | [$m^3/s$] | Carga media de los cluster [t/año] | | |
| Clusters | Caudal | DBO5 | SST | $P_{TOTAL}$ |
| 1 | 0.01 | 0.86 | 4.11 | 0.06 |
| 2 | 7.62 | 22053.69 | 71318.85 | 882.66 |
| 3 | 1.1 | 971.12 | 2815.94 | 50.54 |

| | | | | |
|---|---|---|---|---|
| 4 | 14.67 | 23570.06 | 273807.86 | 2213.99 |
| 5 | 2.79 | 15992.71 | 18634.22 | 490 |
| 6 | 5.97 | 1725.22 | 19069.03 | 164.57 |

Fuente: Autores.

**Anexo J**. Valores medios de las variables de calidad y cantidad más importantes por río según análisis *cluster* para el periodo de validación.

Tabla 45. Caudales y cargas medias de los clusters conformados para canal Torca en el periodo de validación

| | **CANAL TORCA** | | | |
|---|---|---|---|---|
| | **[$m^3/s$]** | **Carga media de los cluster [t/año]** | | |
| **Clusters** | **Caudal** | **DQO** | **GYA** | **SAAM** |
| 1 | 0.02 | 6.28 | 3.09 | 0.07 |
| 2 | 0.16 | 500 | 45.12 | 6.06 |
| 3 | 1.85 | 1718.66 | 706.6 | 37.19 |
| 4 | 0.35 | 1069.41 | 148.99 | 15.99 |

Fuente: Autores.

Tabla 46. Caudales y cargas medias de los clusters conformados para río Salitre en el periodo de validación

| | **RÍO SALITRE** | | | |
|---|---|---|---|---|
| | **[$m^3/s$]** | **Carga media de los cluster [t/año]** | | |
| **Clusters** | **Caudal** | **DBO5** | **DQO** | **$P_{TOTAL}$** |
| 1 | 0.01 | 1.12 | 8 | 0.08 |
| 2 | 2.18 | 19974.28 | 40854.94 | 699.6 |
| 3 | 0.11 | 26.71 | 136.15 | 0.78 |
| 4 | 0.79 | 1660.9 | 3249.08 | 57.61 |
| 5 | 1.46 | 5506.46 | 9539.92 | 203.09 |

Fuente: Autores.

Tabla 47. Caudales y cargas medias de los clusters conformados para río Fucha en el periodo de validación

| | **RÍO FUCHA** | | | |
|---|---|---|---|---|
| | **[$m^3/s$]** | **Carga media de los cluster [t/año]** | | |
| **Clusters** | **Caudal** | **DBO5** | **SAAM** | **$P_{TOTAL}$** |
| 1 | 4.95 | 26420.15 | 705.02 | 724.42 |
| 2 | 0.41 | 23.92 | 1.29 | 0.65 |
| 3 | 1.06 | 2430.47 | 79.98 | 77.24 |
| 4 | 8.04 | 43333.9 | 1686.68 | 1518.86 |

Fuente: Autores.

Tabla 48. Caudales y cargas medias de los clusters conformados para río Tunjuelo en el periodo de validación

| Clusters | RÍO TUNJUELO | | | |
|---|---|---|---|---|
| | [$m^3/s$] | Carga media de los cluster [t/año] | | |
| | Caudal | DBO5 | SST | $P_{TOTAL}$ |
| 1 | 0.01 | 0.75 | 3.43 | 0.02 |
| 2 | 7.67 | 23942.6 | 50816.41 | 818.25 |
| 3 | 1.28 | 1034.68 | 2222.86 | 27.67 |
| 4 | 30 | 47849.06 | 152462.2 | 965.21 |
| 5 | 2.59 | 12501.69 | 13649.25 | 355.15 |
| 6 | 3.83 | 1813.3 | 9532.92 | 76.75 |

Fuente: Autores.

**Anexo K.** Análisis *cluster* de clasificación de las muestras de los PM para el periodo de validación.

Tabla 49. Número de muestras de los puntos de monitoreo del canal Torca clasificadas en cada cluster para el periodo de validación

| Clusters | CANAL TORCA | | | |
|---|---|---|---|---|
| | Puntos de monitoreo | | | |
| | BosqueP | CLL161 | JarPaz | SSimón |
| 1 | 8 | 0 | 0 | 0 |
| 2 | 0 | 10 | 2 | 10 |
| 3 | 0 | 0 | 0 | 1 |
| 4 | 0 | 0 | 6 | 12 |

Fuente: Autores.

Tabla 50. Número de muestras de los puntos de monitoreo del río Salitre clasificadas en cada cluster para el periodo de validación

| Clusters | RÍO SALITRE | | | | | |
|---|---|---|---|---|---|---|
| | Puntos de monitoreo | | | | | |
| | ParqNal | Arzobis | CLL53 | Carrefo | Tv 91 | Alameda |
| 1 | 9 | 9 | 16 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 5 | 10 |
| 3 | 4 | 7 | 9 | 0 | 0 | 0 |
| 4 | 1 | 0 | 1 | 1 | 2 | 0 |
| 5 | 0 | 0 | 0 | 6 | 7 | 14 |

Fuente: Autores.

Tabla 51. Número de muestras de los puntos de monitoreo del río Fucha clasificadas en cada cluster para el periodo de validación

| Clusters | RÍO FUCHA | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Puntos de monitoreo | | | | | | | |
| | Delirio | KR7 | Ferroca | América | Boyacá | VisiónC | ZFranca | Alameda |
| 1 | 0 | 0 | 0 | 3 | 6 | 13 | 11 | 12 |
| 2 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2 | 11 | 10 | 9 | 7 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 |

Fuente: Autores.

Tabla 52. Número de muestras de los puntos de monitoreo del río Tunjuelo clasificadas en cada cluster

| Clusters | RÍO TUNJUELO | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Puntos de monitoreo | | | | | | | | | |
| | Regader | UAN | Yomasa | DJuana | México | SBenito | Makrosur | TV 86 | PteInde | IslaPon |
| 1 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 9 | 0 | 0 | 1 | 5 | 5 | 8 |
| 3 | 2 | 4 | 26 | 5 | 21 | 6 | 3 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 0 |
| 5 | 0 | 0 | 1 | 9 | 0 | 0 | 1 | 5 | 5 | 4 |
| 6 | 2 | 2 | 2 | 7 | 3 | 0 | 4 | 0 | 0 | 0 |

Fuente: Autores