

# ANÁLISIS DE LA CALIDAD DEL AGUA DEL RÍO BOGOTÁ DURANTE EL PERIODO 2008 – 2015 A PARTIR DE HERRAMIENTAS DE MINERÍA DE DATOS

THE WATER QUALITY ANALYSIS OF BOGOTA RIVER IN THE PERIOD 2008 – 2015 FROM DATA MINING TOOLS



<sup>1</sup>Mario Fernando Castro Fernández, <sup>2</sup>Daissy Milena Díaz Casallas,  
<sup>3</sup>Carlos Enrique Montenegro Marín

*Universidad Cooperativa de Colombia, Bogotá, D.C., Colombia. Universidad Distrital Francisco José de Caldas, Bogotá D.C., Colombia*

<sup>1</sup>mario.castrof@campusucc.edu.co <sup>2</sup>daissymdiaz@gmail.com,  
<sup>3</sup>cemontenegrom@udistrital.edu.co carlos.montenegrom@campusucc.edu.co

*Recibido: 25/05/2014 • Aprobado: 22/06/2014*

## RESUMEN

Actualmente, llevar a cabo acciones que reduzcan el impacto ambiental en todo tipo de procesos, tanto industriales, como comerciales o de la vida cotidiana, son preocupaciones pertinentes [1]being them social, scientific, cultural, economical or political. Environmental sensing and monitoring initiatives will be gratified by new Information System (IS. Con la llegada del big data, los grandes volúmenes de datos con una relación armónica de “gente – tierra - tiempo” entró gradualmente en el campo de la gestión de recursos naturales [2]. A diferencia de la cantidad de agua, dato que se puede expresar en términos precisos, la calidad es un atributo de múltiples parámetros. Este proceso arroja un volumen considerable de datos, que finalmente se almacenan en las tablas de datos, pero que no se traducen en información inteligible que describa el estado de la masa de agua. Los índices de calidad del agua -ICA- se convierten en una herramienta de fácil comunicación de la información científica de expertos al público en general [3]. Las anteriores razones motivan la exploración de las diferentes herramientas de minería de datos que día tras día hacen presencia en el mundo. Se logró acceder a los monitoreos realizados a la cuenca del río Bogotá durante el periodo comprendido entre 2007 – 2015. Con esta información se realizó el cálculo de los Indicadores de Calidad del Agua para cada una de las estaciones monitoreadas y cada uno de los diferentes periodos estudiados, logrando, así, analizar y evaluar el espacio temporal de la calidad del río Bogotá.

**Palabras clave:** *calidad del agua, índice de calidad del agua, ICA, big data, minería de datos.*

## ABSTRACT

*Nowadays, carrying out actions that reduce the environmental impact in all kinds of processes, both industrial and commercial or of everyday life, are pertinent concerns [1]. With the arrival of the big data, the large volumes of data with a harmonious relationship of “people - ground - time” came gradually in the field of the natural resources management [2]. Unlike the amount of water, data that can be expressed in precise terms, the quality is an attribute of multiple parameters. This process yields a considerable volume of data, that finally are stored in the data tables, but which have not been translated into intelligible information that describes the state of the water mass. The water quality indices - WQI - become a tool for easy communication of scientific information from experts to the general audience [3]. The above reasons motivated the exploration of different data mining tools that day by day make presence in the world. It was possible to access to the monitoring done to the basin of the Bogota River during the period 2007 – 2015. Thus, with this information, the Indicators of water quality were calculated for each of the monitored stations and each one of the different periods studied, achieving analyze and evaluate the temporary space of the quality of the Bogota River.*

**Keywords:** big data, data mining, wáter quality, wáter quality index, WQI.



## I. INTRODUCCIÓN

Según [1] being them social, scientific, cultural, economical or political. Environmental sensing and monitoring initiatives will be gratified by new Information System (IS, ser capaz de comprender y de “controlar” el medio ambiente, y pronosticar su evolución, son aspectos importantes en las principales áreas de investigación. Así mismo, exponen que la percepción del medio ambiente significa la recolección y análisis de datos pertinentes de este, hacia una gestión eficiente; un escenario similar al de una gran empresa que necesita gestión y reorganización de forma continua.

A diferencia de la cantidad de agua, dato que se puede expresar en términos precisos, la calidad del agua es un atributo de múltiples parámetros. La utilidad de un ICA se basa en adicionar información acerca de los parámetros de la calidad del agua en diferentes momentos y en diferentes lugares y traducir esta información en una sola

puntuación que represente el período de tiempo y la unidad espacial bajo consideración. De esta manera, un ICA se convierte en una herramienta de fácil comunicación para la transmisión de información científica de expertos al público en general [3]. Un índice de calidad del agua incorpora dos deficiencias en la evaluación dinámica de la misma, a saber: la serie de tiempo de muestreo debe ser idéntica para cada indicador y no deben producirse los datos que faltan; y los pesos fijos no pueden representar los cambios en la importancia de contaminantes [4].

Vale aclarar, que una investigación preliminar sobre la evaluación del desempeño ambiental con grandes volúmenes de datos no solo tiene un alto valor científico, sino también gran importancia práctica [5].

Por lo anterior, surge la motivación de explorar las diferentes herramientas tecnológicas que

están haciendo frente a la necesidad de análisis de los grandes volúmenes de datos que actualmente se generan. Se considera que estas nuevas herramientas de minería de datos y *big data* podrían responder a la necesidad de desarrollar un instrumento que permita minimizar esfuerzos, tiempo y costos en las actividades de gestión ambiental, en este caso, la gestión del recurso hídrico, como la identificación de las variables de mayor impacto sobre los indicadores de calidad del agua. Esto, con el fin de analizar sus patrones de comportamiento y construir un modelo cuyo diseño y desarrollo concuerden o se ensamblen con el establecido por las entidades nacionales, para facilitar la gestión de las Corporaciones Autónomas Regionales y aportar datos válidos que se puedan incorporar directamente en el Sistema de Información del Recurso Hídrico - SIRH.

A partir de los resultados de los monitoreos realizados por la Corporación Autónoma Regional de Cundinamarca, CAR, a la cuenca del río Bogotá en el periodo 2007 – 2015, se realizó el cálculo del Índice de Calidad del Agua – ICA para evaluar la variabilidad espacial y temporal y determinar la clasificación de la calidad del agua en el río Bogotá, observándose que las áreas de mayor actividad antrópica son las que menor calidad de agua presentan.

## II. MARCO TEÓRICO

### A. *Big data y calidad del recurso hídrico*

El volumen de datos está creciendo rápidamente con estructuras y formas cada vez más complejas. Un informe de investigación de International Data Corporation (IDC) señaló que los datos 1.8ZB se crearon y se replican en todo el mundo desde 2011; se estima que esta cifra se incrementará 50 veces en el año 2020. Es decir, ha llegado la era de los grandes datos [6].

Con su evolución se han presentado diferentes definiciones de Big Data, las cuales, a medida que

avanza la investigación en este campo, incluyen nuevas características en el concepto. Por una parte, [7] y [8] definen en primera instancia Big Data como cualquier colección de conjuntos de datos, los que, por el volumen y la complejidad de los mismos, hacen que su gestión y procesamiento sean difíciles de realizar con herramientas tradicionales.

Por otra parte, [7], [9], [10], [11] concuerdan en que Big Data no solo hace referencia a los grandes conjuntos de datos (volumen), sino también a limitaciones como velocidad y variedad. Hasta este punto se integran las primeras 3Vs (volumen, variedad y velocidad) como lo afirman son empleadas por diferentes autores. [9], [10] y [11] identifican nuevas características (Vs) que se agregan al concepto: veracidad (calidad e incertidumbre de los datos), valor (filtrado de información valiosa) y visualización (la presentación de la información en una manera eficaz).

Con la llegada del big data, los grandes volúmenes de datos con una relación armónica de “gente – tierra - tiempo” entraron gradualmente en el campo de la gestión de recursos naturales [12]. En comparación con los datos tradicionales, con la recopilación y el registro de big data, se enfatizan dimensiones espaciales y temporales y los intentos de registrar los acontecimientos en todo momento y en todas partes en las trayectorias temporales y espaciales [2].

Gracias a la interacción en tiempo real y a la comunicación con las tecnologías de internet de las cosas, aplicaciones de datos grandes y la enorme cantidad de información de los interesados se pueden utilizar de forma elástica y sostenible, lo que permite alcanzar su máximo potencial [13]. Si los datos pertinentes en el ámbito de la gestión ambiental pudiesen transmitirse al gobierno, los funcionarios gubernamentales competentes estarían motivados para mejorar el nivel de la misma. Por tanto, las herramientas analíticas deben ser desarrolladas con el fin de manejar estos datos con diferentes estructuras para la evaluación ambiental y la predicción [5].

Redes de toma de muestras automatizadas generan una gama limitada de parámetros físico-químicos que se miden continuamente en lugares específicos a alta frecuencia temporal. Este proceso crea un volumen considerable de datos que finalmente se almacenan en las tablas de datos y que no se traducen en información inteligible que describa el estado de la masa de agua [3].

### B. Índice de calidad del agua - ICA en Colombia

El Índice de Calidad del Agua es el valor numérico que califica en una de cinco categorías, la calidad del agua de una corriente superficial, con base en las mediciones obtenidas para un conjunto de cinco o seis variables, registradas en una estación de monitoreo (j) en el tiempo (t) [14].

La hoja metodológica del indicador, desarrollada por [14], establece el método para el cálculo del ICA a partir de los datos de concentración de un conjunto de cinco o seis variables que determinan, en gran parte, la calidad de las aguas corrientes superficiales.

La siguiente ecuación responde al cálculo del índice de calidad de aguas superficiales.

Donde:

Es el índice de calidad del agua de una determinada corriente superficial en la estación de monitoreo de la calidad del agua (j) en el tiempo (t), evaluada con base en n variables.

es el ponderador o peso relativo asignado a la variable de calidad i.

es el valor calculado de la variable i (obtenido de aplicar la curva funcional o ecuación correspondiente) en la estación de monitoreo j, registrado durante la medición realizada en el trimestre k, del período de tiempo t.

es el número de variables de calidad involucradas en el cálculo del indicador; en este caso, n es igual a 5.

La tabla 1 presenta las variables y la ponderación empleada para el cálculo del índice.

TABLA I

VARIABLES Y SU PONDERACIÓN EN EL CÁLCULO DE ICA [14]

Variable	Unidad de medida	Ponderación
Oxígeno disuelto, OD.	% saturación	0,2
Sólidos suspendidos totales, SST.	mg/l	0,2
Demanda química de oxígeno, DQO.	mg/l	0,2
Conductividad eléctrica, C.E.	μS/cm	0,2
pH	Unidades de pH	0,2

Las curvas funcionales adoptadas, son las propuestas por Ramírez y Viña para oxígeno disuelto (OD), sólidos suspendidos totales (SST) y conductividad eléctrica (CE); por la Universidad Politécnica de Catalunya (2006) para demanda química de oxígeno (DQO) y por la del laboratorio del Departamento de Calidad Ambiental de Oregon (Estados Unidos) para pH [14]. Igualmente, define los valores que puede tomar el indicador, los cuales han sido clasificados en categorías. Con base en estos valores, a los que se les ha asociado un color como señal de alerta, se califica la calidad del agua de las corrientes superficiales.

En la tabla 2 se presenta la relación entre valores y calificación; se establece una categoría para cinco rangos de calidad: Muy Mala, Mala, Regular, Aceptable y Buena.

TABLA II  
 CALIFICACIÓN DE LA CALIDAD DEL AGUA SEGÚN  
 LOS VALORES QUE TOMA EL ICA [14]

Categoría de valores que puede tomar el Indicador	Calificación de la Calidad del Agua	Señal de Alerta
0,00-0,25	Muy Mala	Rojo
0,26-0,50	Mala	Naranja
0,51-0,70	Regular	Amarillo
0,71-0,90	Aceptable	Verde
0,91-1,00	Buena	Azul

Las variables de dudosa relevancia, con la mayoría de los valores de prueba que cumplen los objetivos establecidos y los que presentan una fuerte correlación, deben ser excluidos del análisis, con el fin de evitar el sesgo del índice [3].

### III. MATERIALES Y MÉTODOS

#### A. Área de estudio

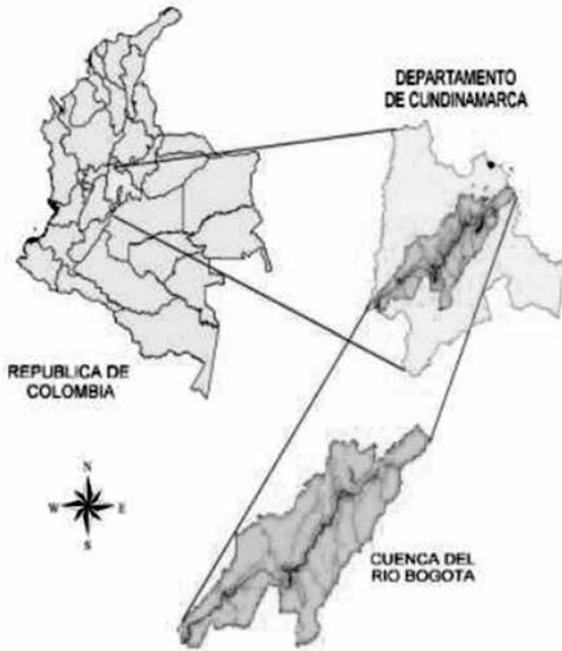
De acuerdo con [15], la cuenca del Río Bogotá está localizada en la parte central del país, en el Departamento de Cundinamarca e integrada por 45 municipios, entre ellos, el Distrito Capital. La cuenca tiene un área de drenaje de 5.907 km<sup>2</sup>; riega el departamento de Cundinamarca en sentido noreste – sureste, desde su nacimiento en el municipio de Villapinzón a 3.300 m.s.n.m., hasta su desembocadura en el río Magdalena en el municipio de Girardot a una altura de 600 m.s.n.m. Entre las corrientes de agua, afluentes del río Bogotá, incluidas en las campañas de muestreo, se encuentran las quebradas Quincha, La Tenería, La Cuy, Honda, Santa Marta, El Buey y Socotá, así como los ríos Tejar, Neusa, Sisga, Tominé, Negro, Teusacá, Frío, Chicú, Juan Amarillo, Fucha, Tunjuelo, Balsillas, Soacha, Calandaima y Apulo.

La cuenca del río Bogotá se encuentra dividida en tres tramos:

- La cuenca alta, desde su nacimiento hasta el norte de la zona urbana del Distrito Capital, con una longitud de 165 Km.
- La cuenca media, desde el inicio de la zona urbana de Bogotá hasta el Salto de Tequendama, con una longitud de 90 Km.
- La cuenca baja, desde el Salto de Tequendama hasta su desembocadura en el río Magdalena, con una longitud de 55 Km.

Como características principales de la cuenca del río Bogotá deben señalarse que en ella habita cerca del 20% de la población del país, teniendo en cuenta al distrito capital. Además, genera el 28% de la actividad económica nacional. Esta alta ocupación y actividad hacen que cerca de las tres cuartas partes de la extensión de la cuenca presente conflictos en los usos del suelo y el agua, relacionados con la incompatibilidad entre la aptitud natural de los suelos y su uso actual. Los problemas que se presentan en la calidad del agua provienen de vertimientos de tipo doméstico e industrial no tratados, especialmente de la ciudad de Bogotá y zonas urbanas aledañas.

La red de monitoreo de la cuenca está conformada por ochenta y un (81) puntos de monitoreo, de los cuales cuarenta y tres (43) se encuentran localizados sobre el cauce principal y treinta y ocho (38) en afluentes, como se explica a continuación. En la figura 1 se da a conocer la ubicación geográfica de la cuenca y en la figura 2, la ubicación de las cabeceras de los municipios que conforman la cuenca.



**Fig. 1** Localización geográfica de la cuenca del río Bogotá [15]



**Fig. 2** Cuenca hidrográfica río Bogotá [16]

## IV. RESULTADOS

Gracias a los resultados de las campañas de monitoreo realizadas en el río Bogotá en los periodos 2008 – 2015, suministrados por la Corporación Autónoma Regional de Cundinamarca – CAR, es posible el desarrollo de este proyecto.

De esta manera, se realizó el cálculo de los Indicadores de Calidad del Agua para cada una de las estaciones monitoreadas y los diferentes periodos estudiados, así como una evaluación espacio temporal de la calidad del río y sus afluentes.

### **B. Limpieza de la información**

A partir de la recopilación documental de los resultados de monitoreo de la calidad hídrica del río Bogotá, se procedió con la conversión del formato de documentos portátiles (PDF) a un formato de texto plano con el fin de facilitar su introducción en el motor de base de datos MySQL.

Una vez se realizó la limpieza de la información suministrada por la CAR, se procedió con el cálculo del índice de calidad de aguas superficiales – ICA- con base en la hoja metodológica para el indicador, creada por el Instituto de Hidrología, Meteorología y Estudios Ambientales – Ideam como se presentó anteriormente.

Se evidenciaron vacíos en la información necesaria para el cálculo del índice, que permitiera realizar un completo análisis espacio temporal de la calidad del agua en las estaciones de monitoreo dispuestas en la cuenca y en los periodos estudiados.

Con 81 estaciones de monitoreo y 15 periodos en estudio se debería contar con 1.215 ICAs por analizar; no obstante, solo se contó con el 62.3% de los indicadores, es decir 757 ICAs. El restante 37.7% (458) presentó alguna justificación para que no se lograra calcular.

De los 458 registros no calculados, se identificó que 218 (47.6%) correspondían a la ausencia de la  $t^{\circ}$  del agua, variable requerida para calcular el sub índice de oxígeno disuelto. Así mismo, 188 (41.0%) correspondían a periodos y estaciones que no presentaban información alguna, posiblemente porque no se realizó el monitoreo o simplemente porque la Corporación no suministró la información.

Se identificaron 18 (3.93%) casos en los que era posible calcular el indicador de calidad del agua; el error se produjo debido a que se tenía una variable con dos cálculos, ya sea porque se obtuvo por diferentes métodos, o porque un análisis de la variable se realizó en campo y la otra en laboratorio.

34 (7.42%) ausencias restantes se relacionaban con la falta de una o más variables requeridas para el cálculo del ICA. Tan solo se presentó un dato fuera del rango del indicador.

En la tabla 3 se resumen las ausencias identificadas.

TABLA III  
RESUMEN DE OBSERVACIONES IDENTIFICADAS

Observación	N° Observaciones	%
t°	218	47,60%
Sin Monitoreo	188	41,05%
DATOS OK	18	3,93%
DQO	8	1,75%
SST	8	1,75%
OD	7	1,53%
EC/DQO/OD/pH/SST/t°	5	1,09%
EC/OD/pH	2	0,44%
DQO/OD/SST/pH/t°	1	0,22%
EC	1	0,22%
OD/SST	1	0,22%
pH	1	0,22%

### A. Evolución de la calidad del agua del río Bogotá

Una vez calculados los indicadores de calidad del agua del río Bogotá, se organizaron en una matriz en la que las columnas representan el período del año analizado y las filas la estación a la que pertenece el análisis. La matriz se compone de 18 columnas: la primera representa el número de la estación de acuerdo con su ubicación geográfica, entendiendo que la N° 1 se encuentra en el nacimiento del río en el municipio de Villapinzón, en el Páramo de Guacheneque, cuenca alta, y la N° 81 desemboca en el río Magdalena en el municipio de Girardot, cuenca baja; la segunda columna contiene el nombre de la estación de monitoreo; la tercera corresponde a la ubicación de la estación respecto de la cuenca, es decir, si la estación de monitoreo se ubica sobre la fuente principal, río Bogotá, o sobre alguno de sus afluentes.

Las restantes 15 columnas responden a cada uno de los periodos monitoreados; en estas se presentan los análisis semestrales; sin embargo, no se encontró información suficiente del segundo semestre de 2001, para el análisis del ICA.

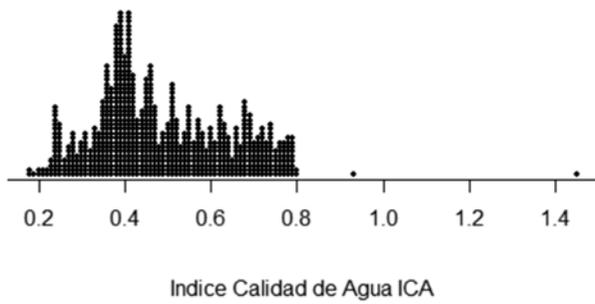
Las celdas de la matriz se resaltan de acuerdo con las categorías de valor definidas para el índice, tabla 2. Se identifica un solo dato atípico correspondiente al monitoreo realizado en la estación LM Chingacio en el primer semestre de 2012. En la figura 3 se presenta la matriz con los resultados.

N°	Nombres de Site	Fuente	2008-01	2008-02	2009-01	2009-02	2010-01	2010-02	2011-01	2011-02	2012-01	2012-02	2013-01	2013-02	2014-01	2014-02	2015-01
1	AGUAS ARRIBA VILLAPINZÓN	Principal	0.79	0.78	0.67	0.54			0.54		0.79	0.75	0.55			0.51	0.4
2	PUNTE VILLAPINZÓN	Principal	0.55	0.73	0.7	0.76			0.68		0.73	0.76	0.76			0.42	0.74
3	AGUAS ARRIBA Q. QUINCHA	Principal	0.74	0.73	0.69	0.4			0.48		0.73	0.64	0.72			0.42	0.75
4	QUINCHA	Afluente	0.57	0.65	0.59	0.62			0.78		0.71	0.69	0.55			0.64	0.54
5	ESTACION LM CHINGACO	Principal	0.46	0.51	0.49	0.62		0.64	0.62		1.41	0.5	0.66			0.63	0.62
6	AGREGADOS CHOCONTA	Afluente	0.57	0.68	0.67	0.5		0.62	0.62		0.51	0.54	0.55			0.5	0.57
7	RIO TELAR	Afluente	0.69	0.72		0.74		0.72	0.58		0.77		0.8			0.79	0.75
8	PUNTE VIA TELECOM	Principal	0.44	0.4	0.46	0.57	0.52	0.64	0.73		0.48		0.59	0.68		0.74	0.47
9	DESCARGA MUN. CHOCONTA	Afluente	0.34	0.43	0.42	0.41	0.39	0.39	0.43		0.44		0.45	0.49	0.33	0.43	0.38
10	AGUAS ABAJO MUN. CHOCONTA	Principal	0.57	0.54	0.57	0.62	0.63	0.64	0.65		0.48		0.59	0.63	0.4	0.62	0.67
11	ESTACION LG. SAUCO	Principal	0.43	0.48	0.58	0.52	0.47	0.54	0.68		0.44		0.71	0.62	0.47	0.57	0.52
12	DESCARGA ENBALSE SONGA	Afluente	0.79	0.9	0.72	0.63	0.79	0.54	0.55		0.79			0.77	0.79	0.79	0.75
13	ESTACION LM SANTA ROSA	Principal	0.49	0.48	0.48	0.49	0.51	0.54	0.79		0.48		0.72	0.7	0.47	0.44	0.44
14	PUNTE SANTANCA	Principal	0.73	0.54	0.62	0.78	0.72	0.52	0.74		0.72		0.66	0.69	0.79		
15	DESCARGA MUNICION WESCA	Afluente	0.37	0.33	0.43	0.39	0.39	0.36	0.36		0.37		0.38	0.3	0.38	0.31	0.31
16	AGUAS ABAJO MUN. SUZCA	Principal	0.36	0.43	0.35	0.27	0.27	0.31	0.32		0.44		0.62	0.42	0.42	0.42	0.42
17	DESCARGA ENBALSE TORRE	Afluente	0.78	0.71	0.74	0.78	0.58				0.74		0.78	0.57	0.78	0.7	0.7
18	AGUAS ARRIBA DESCARGA PAPELES Y MOLINOS	Principal	0.24	0.67	0.49	0.71	0.77	0.53	0.77				0.73	0.71	0.72	0.69	0.69
19	ESTACION LG. - PTE FLORENCIA	Principal	0.76	0.61	0.69	0.74	0.77	0.55	0.77		0.4		0.64	0.65	0.7	0.62	0.76
20	DESCARGA MUNICION GACHANCIPA	Afluente	0.43	0.37	0.31	0.39	0.43	0.45	0.39		0.47		0.4	0.38	0.3	0.38	0.37
21	AGUAS ABAJO MUN. GACHANCIPA	Principal	0.72	0.69	0.67	0.68	0.54	0.67	0.67		0.4		0.69	0.64	0.75	0.55	0.72
22	ESTACION LM - TOCANCIPA	Principal	0.63	0.57	0.64		0.57	0.47	0.47		0.54		0.64	0.68	0.76	0.42	0.77
23	DESCARGA MUN TOCANCIPA	Afluente	0.39	0.44	0.38		0.43	0.36	0.43		0.38		0.4	0.42	0.39	0.38	0.38
24	AGUAS ARRIBA TERMOZPA	Principal	0.62	0.4	0.42		0.44	0.44	0.44		0.54		0.78	0.69	0.7	0.54	0.64
25	DESCARGA TERMOZPA	Afluente	0.53	0.47	0.48	0.51	0.41	0.34			0.42		0.4	0.47	0.66	0.54	0.58
26	HDA EL TRINCO	Principal	0.51	0.5	0.44	0.43											
27	RIO NISUA	Afluente	0.72		0.31	0.63							0.64	0.57	0.63		0.49
28	ESTACION LG. - EL ESPINO	Principal	0.55	0.55		0.49	0.46	0.45					0.64	0.7	0.64	0.44	0.42
29	RIO NEGRO	Afluente	0.29		0.69	0.4	0.32						0.42	0.37	0.3		0.61
30	AGUAS ABAJO RIO NEGRO	Principal		0.46	0.43	0.51	0.49						0.54	0.69	0.61	0.45	0.4
31	RIO TEUSACA	Afluente			0.57	0.79			0.61				0.67	0.68	0.39	0.47	0.46
32	ESTACION LG. - PTE VARGAS	Principal		0.5	0.51	0.55	0.51		0.43				0.57		0.55	0.46	0.51
33	QDA. LA TENERIA	Afluente	0.38	0.36	0.38	0.36	0.36		0.3				0.4		0.4	0.33	0.32
34	AGUAS ARRIBA DE CHIA	Principal	0.46	0.48	0.52	0.4		0.43					0.43		0.46	0.54	0.48
35	DESCARGA MUN CHIA	Afluente	0.43	0.45	0.39	0.48		0.37					0.42		0.42	0.39	0.43
36	AGUAS ABAJO DE CHIA	Principal	0.45	0.49	0.46	0.78		0.45					0.51		0.44	0.49	0.45
37	ESTACION LG. - PTE LA BALSA	Principal	0.46	0.43	0.32	0.32		0.48					0.5		0.45	0.47	0.51
38	RIO FRIO	Afluente	0.68	0.35	0.25	0.51		0.36					0.57		0.76		0.65
39	AGUAS ABAJO RIO FRIO	Principal	0.54	0.46	0.32	0.36		0.47					0.52		0.54	0.46	0.47
40	DESCARGA MUN COTA	Afluente	0.42	0.26	0.35	0.38							0.37		0.57	0.35	0.27
41	ESTACION LG. - PTE LA VIRGEN	Principal	0.47	0.5	0.32	0.46							0.42		0.28	0.46	0.46
42	RIO CHICU	Afluente	0.43		0.4								0.37		0.5	0.42	0.44
43	LM VUELTA GRANDE	Principal	0.42		0.32	0.42							0.46		0.55	0.46	0.46
44	RIV. PASS JUAN AMARILLO	Afluente	0.38	0.4	0.36	0.31						0.36		0.4	0.38	0.44	0.4
45	PTAR SALITRE	Afluente	0.38			0.31							0.39		0.38	0.38	0.34
46	EL CORTUO	Principal	0.72		0.39								0.42		0.42	0.41	0.41
47	DESCARGA JABOQUE	Afluente	0.52			0.45			0.54				0.36		0.51	0.44	0.43
48	DESCARGA ENGATIVA	Afluente	0.52						0.41				0.49		0.37	0.42	0.42
49	AGUAS ABAJO ENGATIVA	Principal	0.39						0.43				0.49		0.43	0.37	0.41
50	A RAMADA	Afluente	0.45	0.54													
51	LG PTE CUNDINAMARCA	Principal	0.42	0.42				0.42					0.39		0.42	0.51	0.41
52	LM HDA SAN FRANCISCO	Principal	0.4	0.4				0.44					0.39		0.41	0.5	0.39
53	RIO FUCHA	Afluente	0.29	0.39				0.29					0.41		0.39	0.33	0.39
54	AGUAS ABAJO RIO FUCHA	Principal	0.38	0.37				0.41					0.41		0.39	0.44	0.37
55	BOMBEO GIRALTAR	Afluente	0.28	0.4				0.44					0.44		0.34	0.38	0.37
56	LG LA PIA	Principal	0.43	0.42	0.39			0.42					0.42		0.39	0.33	0.34
57	R. TUNDUELO	Afluente	0.28	0.4	0.3		0.39						0.26		0.44	0.76	0.51
58	AGUAS ABAJO RIO TUNDUELO	Principal	0.38	0.39	0.35		0.37						0.29		0.38	0.31	0.35
59	RIO BALSILLAS	Afluente	0.44	0.45	0.45	0.47							0.44		0.39	0.38	0.45
60	RIO SONCHA	Afluente	0.5	0.3	0.29	0.31							0.31		0.54	0.27	0.34
61	CANAL SONCHA	Afluente	0.42	0.29	0.29	0.29							0.39		0.38	0.43	0.35
62	LG LAS HUERTAS	Principal	0.39	0.4	0.35	0.43							0.39		0.39	0.38	0.41
63	PTE. VARIANTE BORDOMEDO	Principal	0.43	0.42	0.38	0.46							0.29		0.41	0.36	0.41
64	AGUAS ARRIBA SAN DO TQUEENDAMA	Principal	0.39	0.42	0.4	0.47							0.4		0.38	0.39	0.39
65	DESCARGA MUN EN ANT TQUEENDAMA Q. LA CUY	Afluente	0.3		0.42	0.38							0.38		0.31	0.37	0.38
66	QDA. HONDA	Afluente	0.55		0.32	0.44							0.54		0.54	0.55	0.41
67	AGUAS ABAJO OSEB. HONDA	Principal	0.39	0.35		0.3							0.39		0.4	0.42	0.43
68	ESTACION PTE LA GUACA	Principal	0.26		0.22	0.38	0.37						0.28		0.45	0.4	0.51
69	DESCARGA ENBALSE DEL MU'NA	Afluente	0.4	0.39	0.39	0.39							0.4		0.27	0.38	0.38
70	QDA. SANTA MARTA	Afluente	0.43	0.45	0.71								0.43		0.37	0.43	0.54
71	AGUAS ABAJO QDA. SANTA MARTA	Principal	0.34	0.41	0.37		0.37						0.34		0.39	0.34	0.37
72	Q. SOCOA	Afluente	0.39	0.34	0.34	0.41	0.39		0.41				0.39		0.34	0.34	0.44
73	RIO CALANDARAS	Afluente	0.5	0.44	0.54	0.44	0.55		0.4				0.38		0.33	0.33	0.4
74	AGUAS ABAJO RIO CALANDARAS	Principal	0.31	0.4	0.37	0.39	0.41		0.39				0.36		0.4	0.33	0.39
75	RIO APULO	Afluente	0.5	0.44	0.35	0.43	0.36		0.57				0.34		0.4	0.37	0.54
76	ESTACION PTE BORTILLO	Principal	0.36	0.34	0.35	0.37	0.41		0.41				0.39		0.41	0.35	0.37
77	DESCARGA MUN TOCAMA	Afluente	0.34	0.34	0.35	0.34	0.34		0.38				0.41		0.33	0.35	0.34
78	CINCA EL SALINCO	Principal	0.29	0.36	0.36	0.43	0.43		0.46				0.38		0.4	0.34	0.34
79	AGUAS ARRIBA QDA. EL BUEY	Principal	0.24	0.27		0.24							0.24		0.27	0.24	0.24
80	QDA. EL BUEY	Afluente			0.3	0.27											
81	DEMOCACURA RIO BOGOTA	Principal	0.34	0.37	0.35	0.34							0.44		0.39	0.34	0.34

Fig. 3 Resultados cálculo de indicadores de calidad del agua-ICA

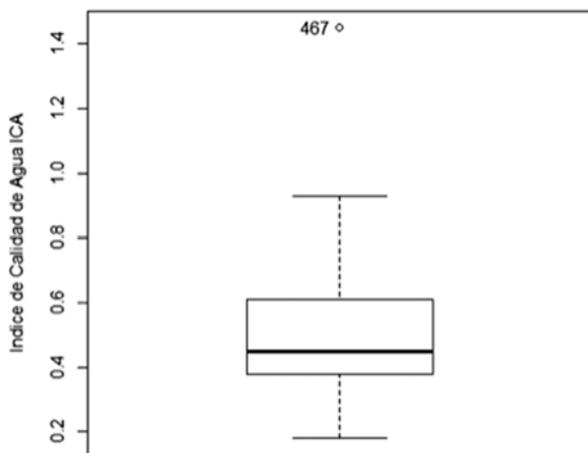
Empleando el software estadístico R, paquete Rcmdr, se realizó una serie de análisis de frecuencia obteniendo los resultados presentados en la figura 4, donde se observa que

sobresale 0.4, es decir, categoría de agua mala. Entre tanto, los demás resultados se encuentran distribuidos entre 0.5 y 0.8 de una forma más uniforme.



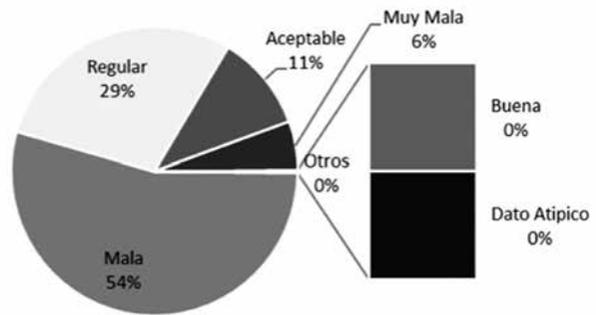
**Fig. 4** Distribución de resultados cálculo ICA

En la figura 5 se observa un dato atípico en el periodo 2011, correspondiente a la estación 5, Estación Lm Chingacio, el cual no se tendrá en cuenta en el análisis de calidad del agua del río Bogotá; es evidente que se encuentra fuera del rango de cálculo (0 – 1.0).



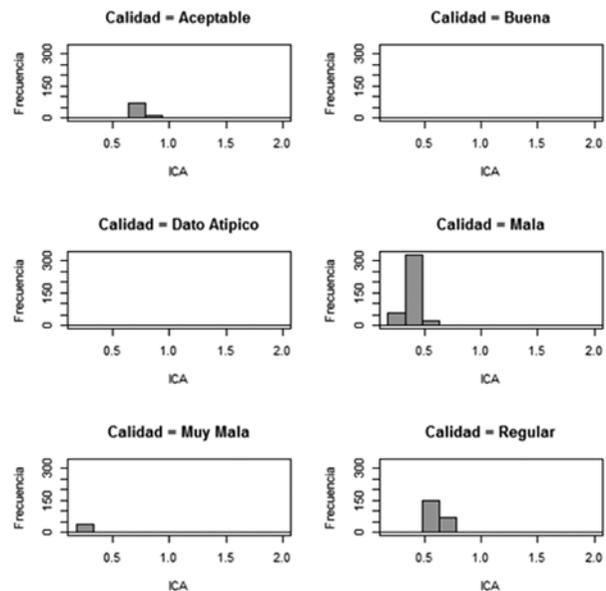
**Fig. 5** Diagrama de caja resultados cálculo ICA

Del total (757) de los índices calculados disponibles, se observa que el 54% (411) corresponde a una calidad de agua mala, la cual se presenta en las cuencas media y baja; el 29% (221) presenta una calidad regular, en especial en cuenca alta. La parte alta de la cuenca presenta una calidad aceptable representada con el 11% (82) y, finalmente, los peores resultados, el 6% (41), se observan en la cuenca baja del río. Solo se presenta un resultado de agua buena, como se observa en la figura 6.



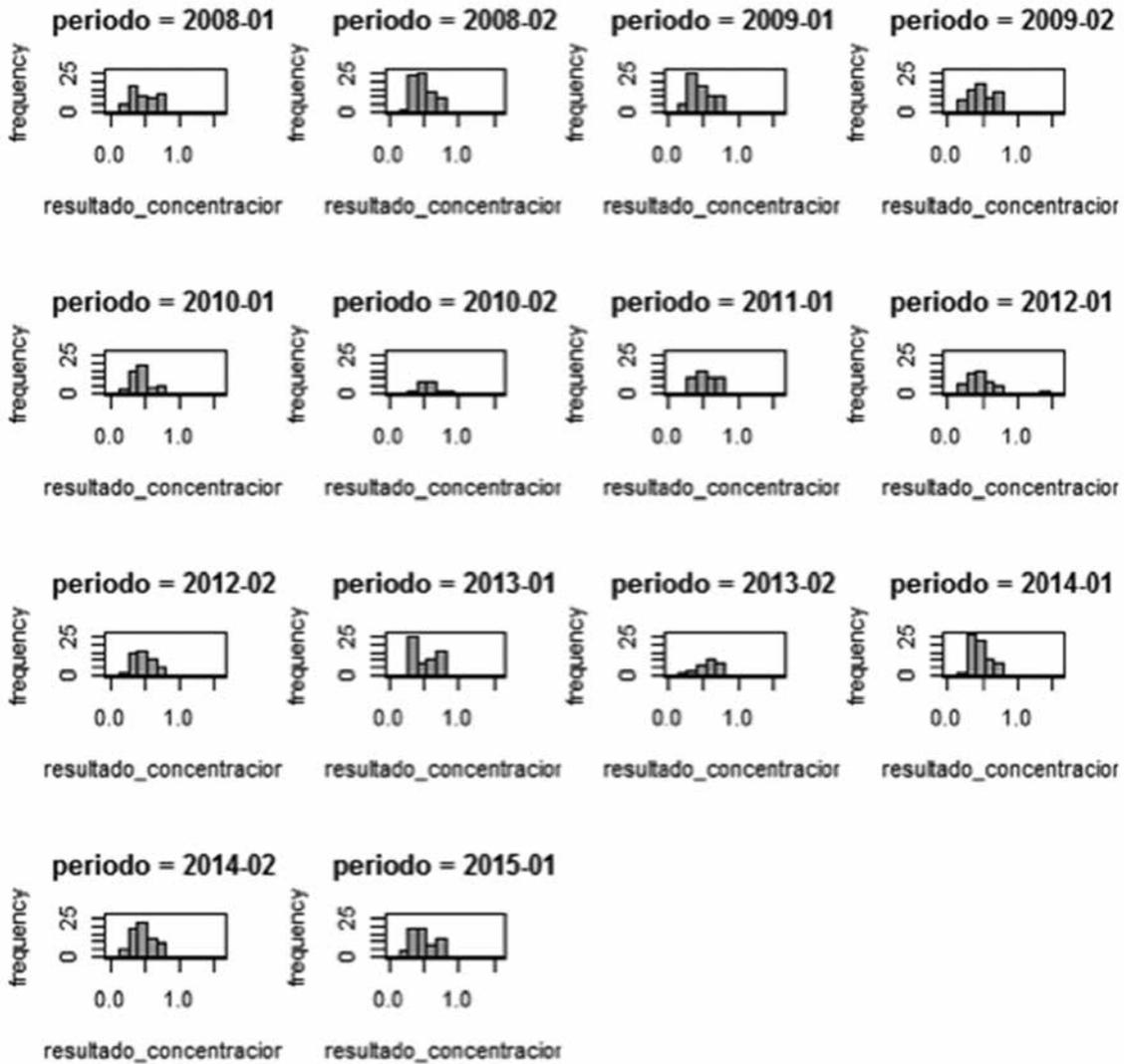
**Fig. 6** Relación de calidad el río Bogotá en %

En la figura 7 se observa en detalle la distribución de los resultados mencionados anteriormente, por categorías, Muy Mala, Mala, Regular, Aceptable y Buena.



**Fig. 7** Histogramas de frecuencia por categoría de calidad del agua

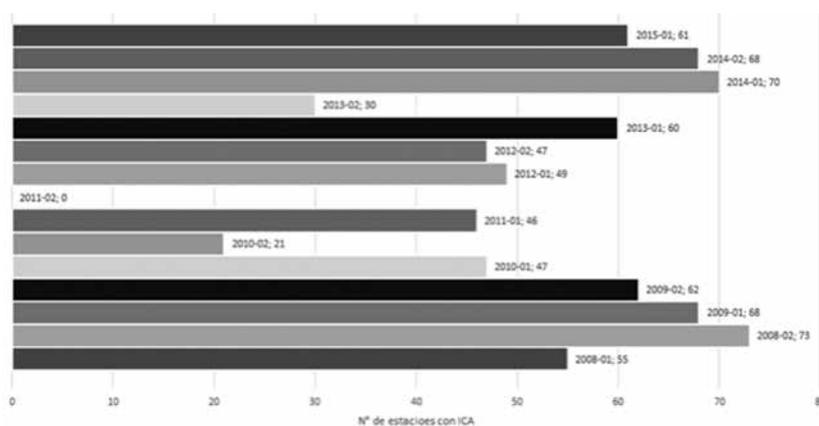
Adicionalmente, se realizó un histograma de frecuencias por cada uno de los periodos analizados; es importante aclarar que basados en este gráfico no es posible analizar la calidad del agua del río, sino la disponibilidad de información en cada periodo. Ver figura 8.



**Fig. 8** Distribución de frecuencia según periodo analizado

Es posible afirmar, entonces, que para los periodos 2008 -1, 2009 – 1, 2014 – 2, y 2015 – 1 se encuentran suficientes datos, lo que permite hacer un análisis comparativo entre ellos: uno a seis meses entre los periodos 2014 – 2, y 2015 – 1, y el otro a

un año, entre los periodos 2008 -1 y 2009 – 1. Por su parte, para los periodos 2010 – 2 y 2013 – 1, no se encuentran datos que permitan un análisis. En la figura 9 se pueden observar los resultados del ICA de acuerdo con cada semestre analizado.



**Fig. 9** Disponibilidad de ICA según periodo de monitoreo.

### B. Discusión de resultados

En la matriz se evidencia que la calidad hídrica de la cuenca desciende a medida que el río fluye a su desembocadura. La sub cuenca, comprendida principalmente por los municipios de Villapinzón y Chocontá, cuenta con una calidad aceptable de agua, por ser esta un área rural con una población relativamente pequeña en comparación con municipios de aguas bajas; sin embargo, no se puede hablar de una calidad buena, debido a la presencia de curtiembres [17]. Las etapas del proceso de curtido, de pelambre y recurrido, el uso de calderas, el teñido y engrase, la pigmentación, el secado, el desencalado, el rebajado y remojo afectan en gran medida el medio ambiente, así como los vertimiento de líquidos y sólidos que se arrojan al río Bogotá.

Vale aclarar que no solamente contaminan los curtidores; los cultivadores de papa lo hacen en la zona de páramo y nacimiento del río en aproximadamente 9.800 hectáreas, donde utilizan agroquímicos que tienen efectos adversos en el agua y el suelo [17].

En municipios como Tocancipá, Cajicá, Chía, Cota, Funza, Mosquera, Soacha, ubicados en la sub cuenca media, en los últimos años se ha evidenciado incremento en su población debido al desplazamiento de habitantes de la ciudad de Bogotá a sus cascos urbanos. Así mismo, se han

concentrado en estos municipios un gran número de parques industriales y zonas francas que se han desplazado de la ciudad capital.

Otro factor preponderante que influye en la calidad del río Bogotá, es la cercanía de la capital, con una población superior a los siete millones de habitantes y gran variedad de industrias que vierten sus residuos líquidos a la red de alcantarillado, los cuales llegan a la planta de tratamiento de aguas residuales, PTAR El Salitre, y de allí al río Bogotá, disminuyendo, de esta manera, la calidad del agua del río.

[18] realizaron un estudio para medir la confiabilidad de las PTAR que operan en las sub cuencas media y alta del río Bogotá, entendiendo confiabilidad “como la probabilidad de que el sistema de tratamiento cumpla con los vertimientos establecidos en la normatividad vigente”. Concluyen que en el estado actual de las depuradoras, ninguna cumpliría con el estándar de reutilización para una posible utilización de los efluentes hacia algún uso urbano, o para cultivos de productos que se consumen crudos, puesto que estos sistemas de tratamiento no superan lo establecido en la normatividad de 10 mg/ L de DBO y SST.

De acuerdo con lo anterior, se observa que las variables, Demanda Biológica de Oxígeno – DBO y Sólidos Suspendidos Totales – SST, que se

vierten al río, y que se emplean en el cálculo de ICA, afectan directamente el índice.

En la sub cuenca baja del río Bogotá, el río ya viene afectado por la presión ejercida en las sub cuencas media y alta y una temperatura superior a los 20°C. El paisaje de esta zona explica que una de las principales actividades de sus municipios, (Apulo, Tocaima, Agua de Dios, Ricaurte, Girardot, entre otros) sea el turismo. Como se observó en la matriz, las estaciones ubicadas en la sub cuenca baja del río presentan una calidad de agua muy mala, hecho que se registró especialmente en el periodo de 2008 a 2012.

Se resalta, no obstante, que durante los años 2013 a 2015 se evidenció un mejoramiento en la calidad del agua de la sub cuenca baja pues solo presentó 4 registros de muy mala. Caso contrario ocurre en la cuenca alta, que en los años 2008 a 2012 solo contaba con 3 registros de calidad muy mala, mientras que en los últimos tres años se duplicó este dato.

## V. CONCLUSIONES

El cambio en las preferencias humanas ha hecho evidente nuevas necesidades, cada vez más difíciles de cumplir y que garanticen la calidad de vida de las personas. Estas necesidades también han hecho que los avances tecnológicos tomen velocidades nunca antes vistas, lo que presenta nuevos retos y oportunidades a la investigación en diferentes campos. A medida que estas tecnologías de evaluación de desempeño ambiental tienen suficiente universalidad, es difícil identificar el método analítico más adecuado. A pesar de que algunos estudios han utilizado métodos similares de evaluación para analizar los problemas reales comparables, sus núcleos teóricos pueden variar en gran medida [5].

Desde el inicio del proyecto, en el momento de tratamiento y limpieza de información se

evidencia la necesidad de conocer y diferenciar el tipo de datos por trabajar; la información y los registros de orden ambiental se encuentran atados y vinculados entre sí. Cuando se requiere separar y diferenciar este tipo de información se debe tener especial cuidado con los posibles vínculos presentes y la afectación que pueda presentarse en los resultados.

La disponibilidad de información sobre resultados de monitoreo ambientales puede presentar vacíos debido a problemas de orden administrativo, como compra de equipos, temas logísticos, hasta condiciones meteorológicas que afecten el trabajo de campo.

La ayuda de nuevas tecnologías y desarrollos informáticos para el tratamiento de grandes volúmenes de información ofrecen la oportunidad de agilizar el análisis de la información disponible sobre recurso hídrico, por lo que el paso que se ha de seguir es realizar el análisis del ICA en conjunto con las demás variables monitoreadas, con el fin de establecer la existencia de correlación del índice con los parámetros fisicoquímicos analizados, además de sus patrones de comportamiento.

Existen diferentes desarrollos de software que permiten la aplicación de técnicas de minería de datos. Se seleccionaron tres herramientas para el trabajo de *Big Data*; se realizaron los análisis de correlación en cada una de ellas y, finalmente, se estudiaron los resultados obtenidos, lo que permitió identificar la herramienta más adecuada para el tratamiento de información ambiental.

## REFERENCIAS

- [1] L. Ferreira, G. D. Putnik, N. Lopes, A. Lopes, and M. M. Cruz-Cunha, "A Cloud and Ubiquitous Architecture for Effective Environmental Sensing and Monitoring," *Procedia Comput. Sci.*, vol. 64, , 2015, pp. 1256–1262.
- [2] H. Xie, Y. He, and X. Xie, "Exploring the factors influencing ecological land change for China's Beijing–



- Tianjin–Hebei Region using big data,” *J. Clean. Prod.*, Apr. 2016.
- [3] M. Terrado, D. Barceló, R. Tauler, E. Borrell, and S. de Campos, “Surface-water-quality indices for the analysis of data generated by automated sampling networks,” *TrAC Trends Anal. Chem.*, vol. 29, no. 1, Jan. 2010, pp. 40–52.
- [4] F. Yan, L. Liu, Y. Li, Y. Zhang, M. Chen, and X. Xing, “A dynamic water quality index model based on functional data analysis,” *Ecol. Indic.*, vol. 57, Oct. 2015, pp. 249–258.
- [5] M.-L. Song, R. Fisher, J.-L. Wang, and L.-B. Cui, “Environmental performance evaluation with big data: theories and methods,” *Ann. Oper. Res.*, Mar. 2016.
- [6] K. Zhou, C. Fu, and S. Yang, “Big data driven smart energy management: From big data to big insights,” *Renew. Sustain. Energy Rev.*, vol. 56, Apr. 2016, pp. 215–225.
- [7] I. A. T. Hashem, I. Yaqoob, N. Badrul Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, “The rise of ‘Big Data’ on cloud computing: Review and open research issues,” *Inf. Syst.*, vol. 47, Aug. 2014, pp. 98–115.
- [8] C. Vitolo, Y. Elkhatib, D. Reusser, C. J. A. Macleod, and W. Buytaert, “Web technologies for environmental Big Data,” *Environ. Model. Softw.*, vol. 63, Jan. 2015, pp. 185–198.
- [9] S. Nativi, P. Mazzetti, M. Santoro, F. Papeschi, M. Craglia, and O. Ochiai, “Big Data challenges in building the Global Earth Observation System of Systems,” *Environ. Model. Softw.*, vol. 68, Jun. 2015, pp. 1–26.
- [10] C. Kacfeh Emani, N. Cullot, and C. Nicolle, “Understandable Big Data: A survey,” *Comput. Sci. Rev.*, vol. 17, Jun. 2015, pp. 70–81.
- [11] S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, “How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study,” *Int. J. Prod. Econ.*, vol. 165, Jan. 2015, pp. 234–246.
- [12] B. C. Pijanowski, A. Tayyebi, J. Doucette, B. K. Pekin, D. Braun, and J. Plourde, “A big data urban growth simulation at a national scale: Configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment,” *Environ. Model. Softw.*, vol. 51, Jan. 2014, pp. 250–268.
- [13] I. A. T. Hashem, V. Chang, N. B. Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, and H. Chiroma, “The role of big data in smart city,” *Int. J. Inf. Manage.*, vol. 36, no. 5, Oct. 2016, pp. 748–758.
- [14] Instituto de Hidrología Meteorología y Estudios Ambientales Ideam, “Hoja metodológica del indicador Índice de calidad del agua,” *Sistema de Indicadores Ambientales de Colombia - Indicadores de Calidad del agua superficial*. p. 10, 2011.
- [15] CAR, “Boletín de Calidad de las Cuencas de la Jurisdicción CAR,” Bogotá, 2011.
- [16] CAR, *Diagnóstico de la Cuenca Hidrográfica del Río Bogotá*. 2006.
- [17] J. I. Bautista R., “Las curtiembres del corredor Villapinzón- Chocontá y el Río Bogotá,” *Rev. Fac. Ciencias Económicas Investig. y Reflexión*, vol. XII, no. 2, 2004, pp. 36–45.
- [18] U. F. Carreño Sayago and J. A. Méndez Sayago, “Estimación de la confiabilidad de las plantas de tratamiento de aguas residuales que operan con lagunas de estabilización en la cuenca alta y media del Río Bogotá,” *Ing. Recur. Nat. y del Ambient.*, no. 10, 2011, pp. 55–64.

