

Validación y análisis de datos reportados en el Sistema de Información del Recurso Hídrico



Unidad de Científicos de Datos
2017



DNP Departamento
Nacional
de Planeación

Resultado de avances del Proyecto Evolución tecnológica del Sistema de Información del Recurso Hídrico. IDEAM – DNP

Diciembre de 2017

Contenido

Contenido.....	1
1. Contexto y antecedentes	1
2. Fase 1 - Validación de la ubicación geográfica de los puntos registrados en el SIRH	2
2.1 Alcance	2
2.2 Objetivo.....	2
2.3 Fuentes de información y herramientas.....	2
2.4 Metodología de validación.....	3
3. Procedimiento.....	5
4. Resultados	6
5. Conclusiones.....	19

1. Contexto y antecedentes

En 2017, la Subdirección de Desarrollo Ambiental Sostenible del DNP y la coordinadora del grupo del SIAC del IDEAM elaboraron un proyecto de evolución del Sistema de Información del Recurso Hídrico que contó con el respaldo de la Subdirección de Hidrología del IDEAM. En una reunión con el grupo de científicos de datos de la Subdirección de Telecomunicaciones, se encontraron puntos de interés de este proyecto con la estrategia de apoyo a la implementación de proyectos de Big Data, a partir de lo cual se realizó una colaboración para el desarrollo de algunos componentes del proyecto por el grupo de científicos de datos, con el acompañamiento temático de la SDAS y el IDEAM.

El SIRH es un subsistema del Sistema de Información Ambiental para Colombia (SIAC), fue planeado para consolidar la información sobre cantidad y calidad de agua y sobre los permisos de uso y vertimiento otorgados por las autoridades ambientales, con el propósito de conocer el estado y hacer seguimiento a la gestión del recurso hídrico, soportar la toma de decisiones y promover estudios hidrológicos e hidrogeológicos. Este sistema de vital importancia para la gestión del recurso hídrico, ha contado con avances significativos en los últimos años, que hoy permiten estructurar información sobre las fuentes hídricas, los usuarios del agua, datos de monitoreo de calidad del agua, y sobre los planes y estrategias formulados por las Autoridades Ambientales Competentes (AAC) para atender a las principales problemáticas del agua en sus regiones, garantizar su uso eficiente y preservarlo para el bienestar de las futuras generaciones.

No obstante, su uso se ha visto limitado por la baja cantidad de datos reportados por las autoridades ambientales, la baja confianza en los datos reportados y la inexistencia de herramientas de análisis de datos. Además, el SIRH tiene el potencial de incorporar y utilizar bases de datos de entidades públicas y privadas de sectores usuarios del agua, que actualmente no son fácilmente accesibles ni están integradas.

Así mismo, su arquitectura de software fue desarrollada hace más de 4 años, por lo que no responde a las tendencias actuales de presentación y gestión de información, que le ofrezcan a los usuarios servicios ágiles y de fácil uso.

Actualmente cuenta con servicios para el registro de datos, con unas consultas básicas sobre el estado de los datos y articula otras fuentes que cubren solo una pequeña fracción de los datos existentes en el país. La información consolidada no es accesible mediante servicios geográficos, que permitan comprender las relaciones espaciales de las variables. Es conocido que sistemas georeferenciados, integrados, y con herramientas de apoyo a la toma de decisión permiten optimizar la información y hacer mejor seguimiento a los problemas de los recursos hídricos.

En este contexto, el grupo de científicos de datos elaboró un procedimiento semiautomático para la validación de la ubicación geográfica de los puntos registrados en el SIRH, entre ellos los puntos de inicio y final de los tramos, los de concesiones y permisos de vertimiento y los de monitoreo. Esta validación incluye la verificación de la ubicación de los puntos en los polígonos de la jurisdicción de cada autoridad ambiental y la distancia a las redes de drenaje simple y doble a escala 1:100.000 del IGAC. En este informe se presenta el procedimiento y los resultados de las actividades realizadas.

2. Fase 1 - Validación de la ubicación geográfica de los puntos registrados en el SIRH

2.1 Alcance

- Validación de las ubicaciones geográficas de los puntos de interés reportados en el SIRH
- Alcance geográfico: datos reportados a nivel nacional por todas las autoridades ambientales y el IDEAM

2.2 Objetivo

Validación semiautomática de los datos reportados correspondientes a las ubicaciones geográficas de:

1. Puntos de inicio y final de los tramos
2. Puntos de vertimiento y captación
3. Puntos de monitoreo

2.3 Fuentes de información y herramientas

- Datos: SIRH (tablas Autoridad, FNTT_FUENTE, FNTT_TRAMO), Cartografía IGAC red de drenaje doble y sencillo, áreas de jurisdicción corporaciones.

- Software
 - QGIS 2.18.14
 - Python 3.6 (Pandas)
- Shape files
 - Mapa de aprovechamiento forestal jurisdicción de las Corporaciones Autónomas Regionales
 - Red drenaje sencillo IGAC
 - Red drenaje doble IGAC
- Base datos
 - SIRH-IDEAM

2.4 Metodología de validación

- Análisis por corporaciones.
 - Coherencia en la ubicación de los puntos con respecto a la zona de jurisdicción de cada autoridad ambiental.
- Análisis geográfico.
 - Calculo de distancia mínima a la red de drenaje.
 - Correspondencia de los puntos PI o PF con la información del IGAC. Validar que correspondan (bajo un umbral de distancia) a puntos de inicio y final de los tramos.
 - Coherencia entre los puntos PI y PF. Con respecto a la información del tramo (longitud, ubicación del par)
- Reporte resultados por corporación con histogramas para datos de distancia.

2.5 Resumen datos a validar

NOMBRE	SIGLA	Ptos. Inicio	Ptos. Final	Vertimiento	Captación	Monitoreo
Área Metropolitana de Bucaramanga	AMB	0	0	0	0	0
Área Metropolitana del Valle de Aburrá	AMVA	75	75	46	275	2
ANLA	ANLA	46	46	0	0	0
Barranquilla Verde	BARR_VERDE	7	7	1	0	1
Corporación Autónoma Regional del Alto Magdalena	CAM	60	60	3	1005	2
Corporación Autónoma Regional de Cundinamarca	CAR	159	159	1	8728	239
Corporación Autónoma Regional de Risaralda	CARDER	40	40	0	2	0
Corporación Autónoma Regional del Canal del Dique	CARDIQUE	120	120	2	74	0
Corporación Autónoma Regional de Sucre	CARSUCRE	36	36	15	134	74
Corporación Autónoma Regional de Santander	CAS	376	376	0	753	0
Corporación para el Desarrollo Sostenible del Norte y Oriente de la Amazonia	CDA	35	35	2	56	9

Corporación Autónoma Regional para la Defensa de la Meseta de Bucaramanga	CDMB	54	54	1	1920	25
Corporación Autónoma Regional para el Desarrollo Sostenible del Chocó	CODECHOCO	36	36	4	34	8
Corporación para el Desarrollo Sostenible del Archipiélago de San Andrés Providencia y Santa Catalina	CORALINA	3	3	0	282	0
Corporación Autónoma Regional del Centro de Antioquia	CORANTIOQUIA	39	39	1	12	1
Corporación para el Desarrollo Sostenible del Área de Manejo Especial La Macarena	CORMACARENA	459	459	55	1410	177
Corporación Autónoma Regional de las cuencas de los ríos Rionegro y Nare	CORNARE	30	30	218	4269	66
Corporación Autónoma Regional del Magdalena	CORPAMAG	16	16	1	132	1
Corporación para el Desarrollo Sostenible del Sur de la Amazonia	CORPOAMAZONIA	114	114	63	115	1
Corporación Autónoma Regional de Boyacá	CORPOBOYACA	322	322	1	431	1
Corporación Autónoma Regional de Caldas	CORPOCALDAS	603	603	41	994	0
Corporación Autónoma Regional del Cesar	CORPOCESAR	134	134	6	635	0
Corporación Autónoma Regional de Chivor	CORPOCHIVOR	324	324	12	792	37
Corporación Autónoma Regional de la Guajira	CORPOGUAJIRA	267	267	21	3650	170
Corporación Autónoma Regional del Guavio	CORPOGUAVIO	244	244	3	1365	3
Corporación para el Desarrollo sostenible de la Mojana y el San Jorge	CORPOMOJANA	4	4	1	6	1
Corporación Autónoma Regional de Nariño	CORPONARIÑO	128	128	1	44	13
Corporación Autónoma Regional de la Frontera Nororiental	CORPONOR	1591	1591	1	186	25
Corporación Autónoma Regional de la Orinoquía	CORPORINOQUIA	504	504	191	1310	0
Corporación para el Desarrollo Sostenible de Urabá	CORPOURABA	138	138	155	511	212
Corporación Autónoma Regional del Tolima	CORTOLIMA	617	617	0	564	1
Corporación Autónoma Regional del Atlántico	CRA	222	222	39	277	205
Corporación Autónoma Regional del Cauca	CRC	192	192	45	290	4
Corporación Autónoma Regional del Quindío	CRQ	285	285	2	397	21
Corporación Autónoma Regional del Sur de Bolívar	CSB	4	4	0	0	1
Corporación Autónoma Regional del Valle del Cauca	CVC	604	604	1	808	13
Corporación Autónoma Regional de los Valles del Sinú y San Jorge	CVS	9	9	0	36	0
Departamento Administrativo Distrital del Medio Ambiente. (Santa Marta)	DADMA	3	3	1	45	43

Departamento Administrativo de Gestión del Medio Ambiente. (Santiago de Cali)	DAGMA	21	21	4	97	8
EPA Cartagena	EPA	0	0	0	0	0
Instituto de Hidrología y Meteorología y Estudios Ambientales	IDEAM	1631	1631	0	0	4
Ministerio de Ambiente de Desarrollo sostenible	MADS	0	0	0	0	0
Parques Naturales	PNN	69	69	1	46	17
INVEMAR - REDCAM	REDCAM	60	60	0	0	308
Sentencia Rio Bogotá	RIO_BOGOTA	0	0	0	0	0
Secretaria Distrital de Ambiente	SDA	44	44	12	74	32
Total		9725	9725	951	31759	1725

3. Procedimiento

Se inicia con la extracción, como archivo CSV, de la información de los puntos contenida en la base de datos (SIRH): ubicación geográfica, corporación que reporta.

Estos archivos se importan como una capa en QGIS, adicionalmente se cargan los shapefile con las jurisdicciones de las CAR y las redes de drenaje doble y sencillo.

- Para determinar la si la ubicación de los puntos corresponde al área de jurisdicción de la corporación que lo reporta, se utilizó la función “Unir atributos por localización” usando como entrada la capa de los puntos y la capa de las jurisdicciones. La salida de esta función es una capa que se exporta como un archivo CSV.
- Para determinar la distancia a la red de drenaje, se utilizó la función “NNJoin” usando como entrada la capa de la red de drenaje y la capa de los puntos. La salida de esta función es una capa que contiene un atributo adicional que corresponde a la distancia mínima desde el punto al elemento más cercano de la red de drenaje

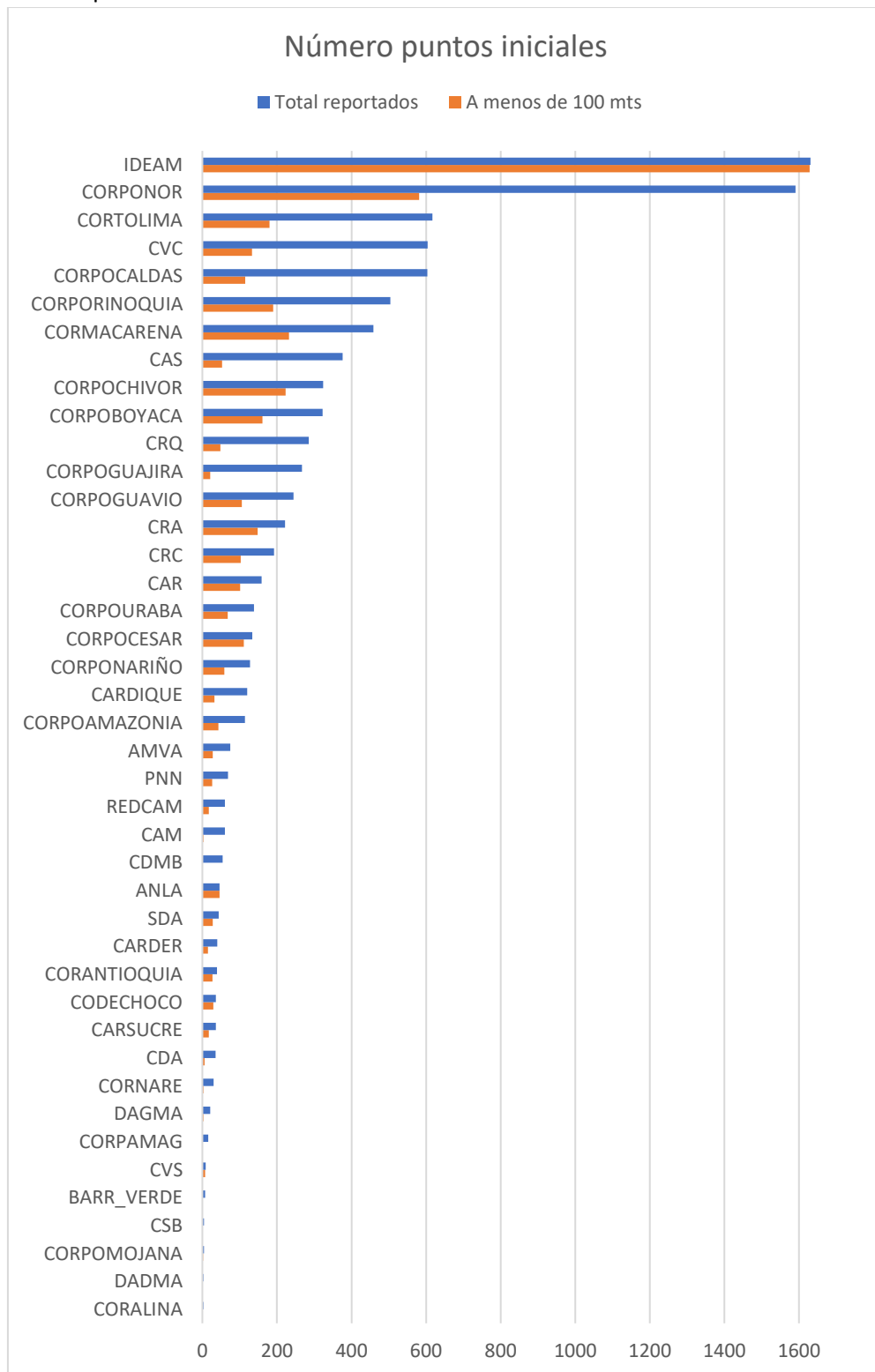
Este procedimiento se realizó para los puntos reportados en inicio y final de tramos, vertimientos, captaciones y monitoreo.

Con los resultados de los pasos anteriores, archivos CSV, se procedió a analizar los datos de interés (pertenece corporaciones y distancia a red de drenaje) mediante un script en lenguaje Python.

- Con respecto al área de jurisdicción de la CAR se validó si la información resultante del proceso en QGIS correspondía a la información reportada en la base de datos. Resultado de este proceso son las gráficas de “Porcentaje puntos dentro de la jurisdicción” que muestran el porcentaje de puntos reportados que, de acuerdo con la información geográfica, se encuentran dentro del área de jurisdicción de la corporación que lo reportó.

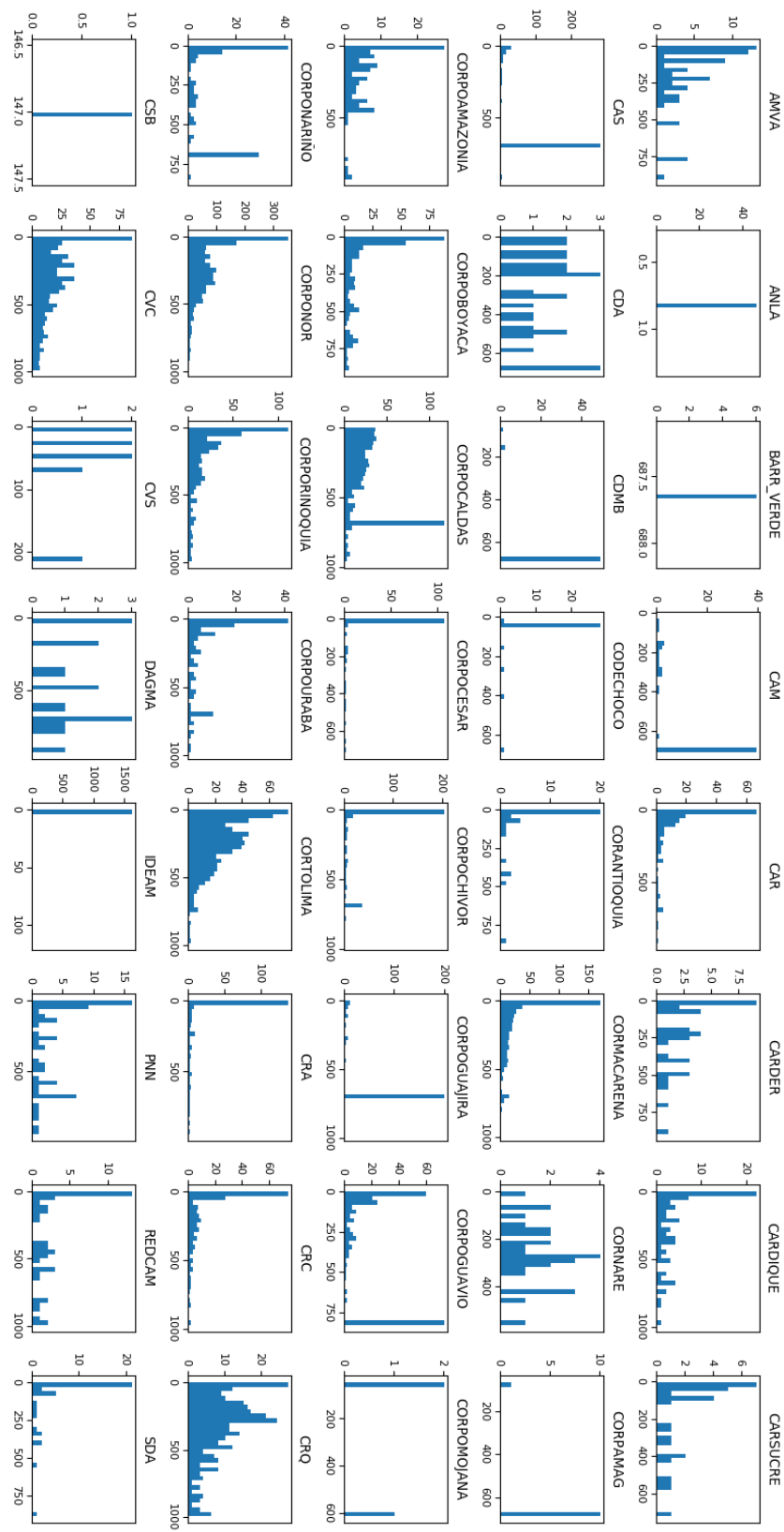
4. Resultados

4.1 Tramos puntos iniciales

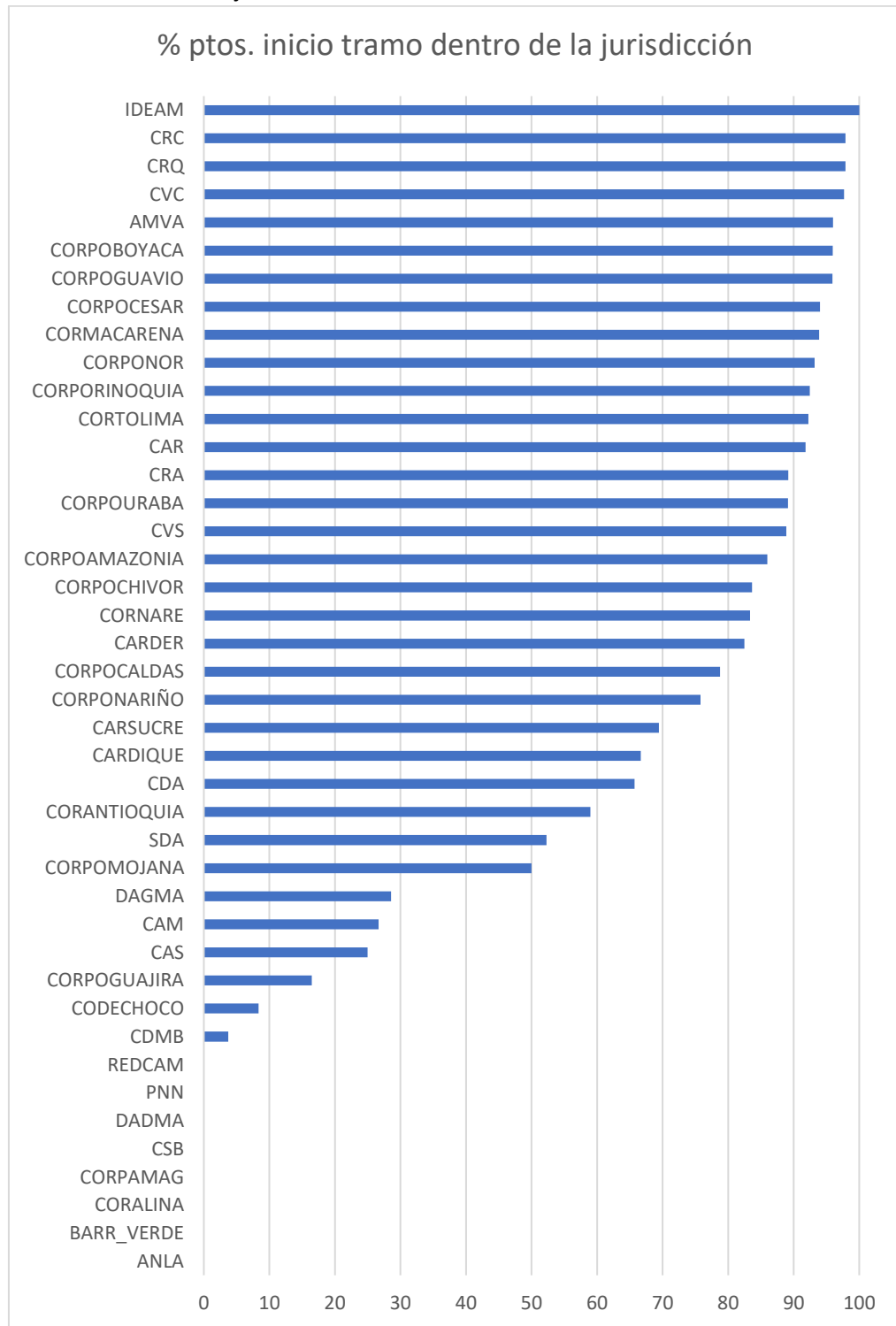


No reportaron: RIO_BOGOTA, MADS, EPA, AMB

4.1.1 Histograma distancia puntos iniciales a red de drenaje (Por corporación – Eje x en metros)

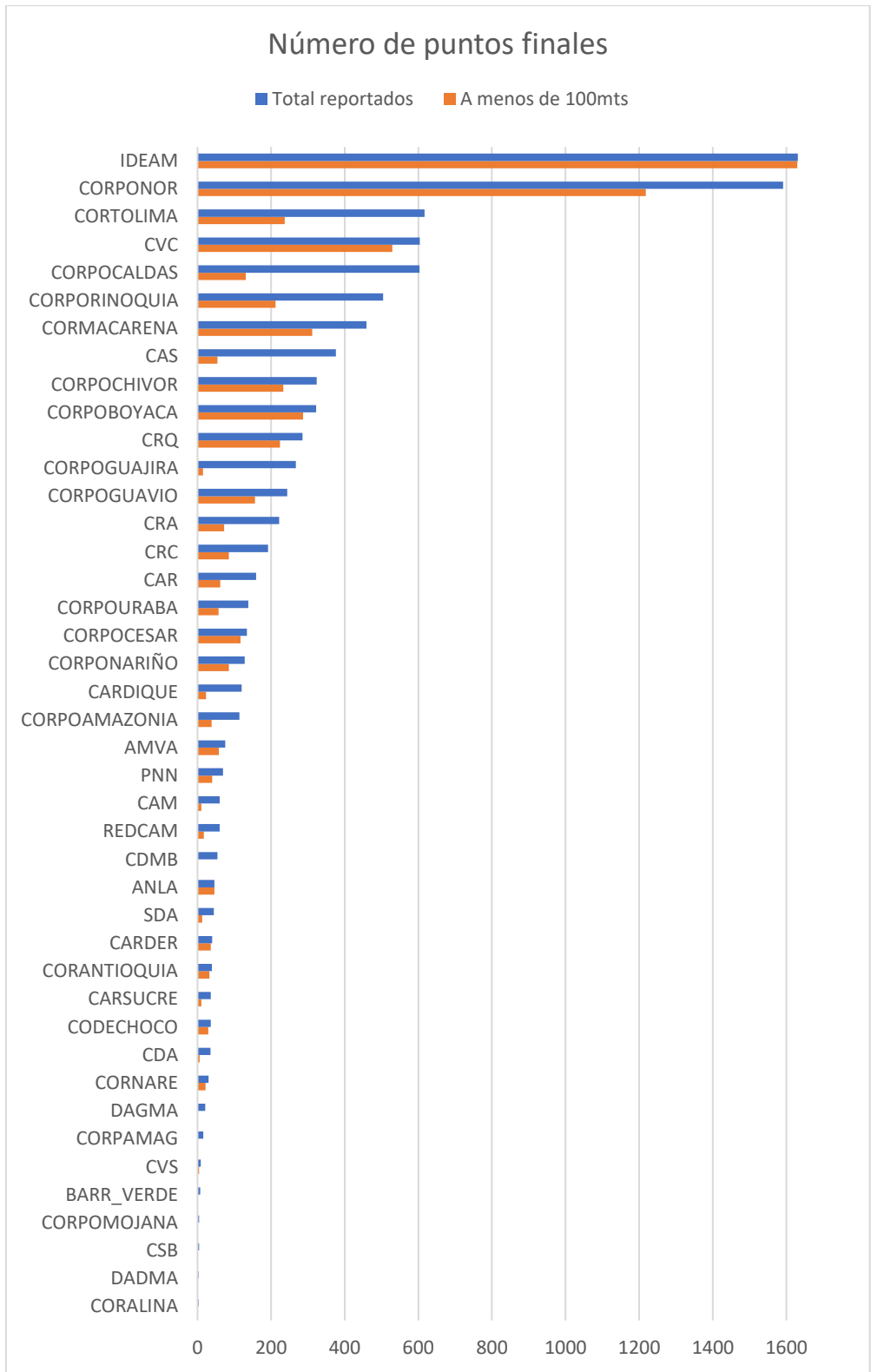


4.1.2 Puntos iniciales en la jurisdicción



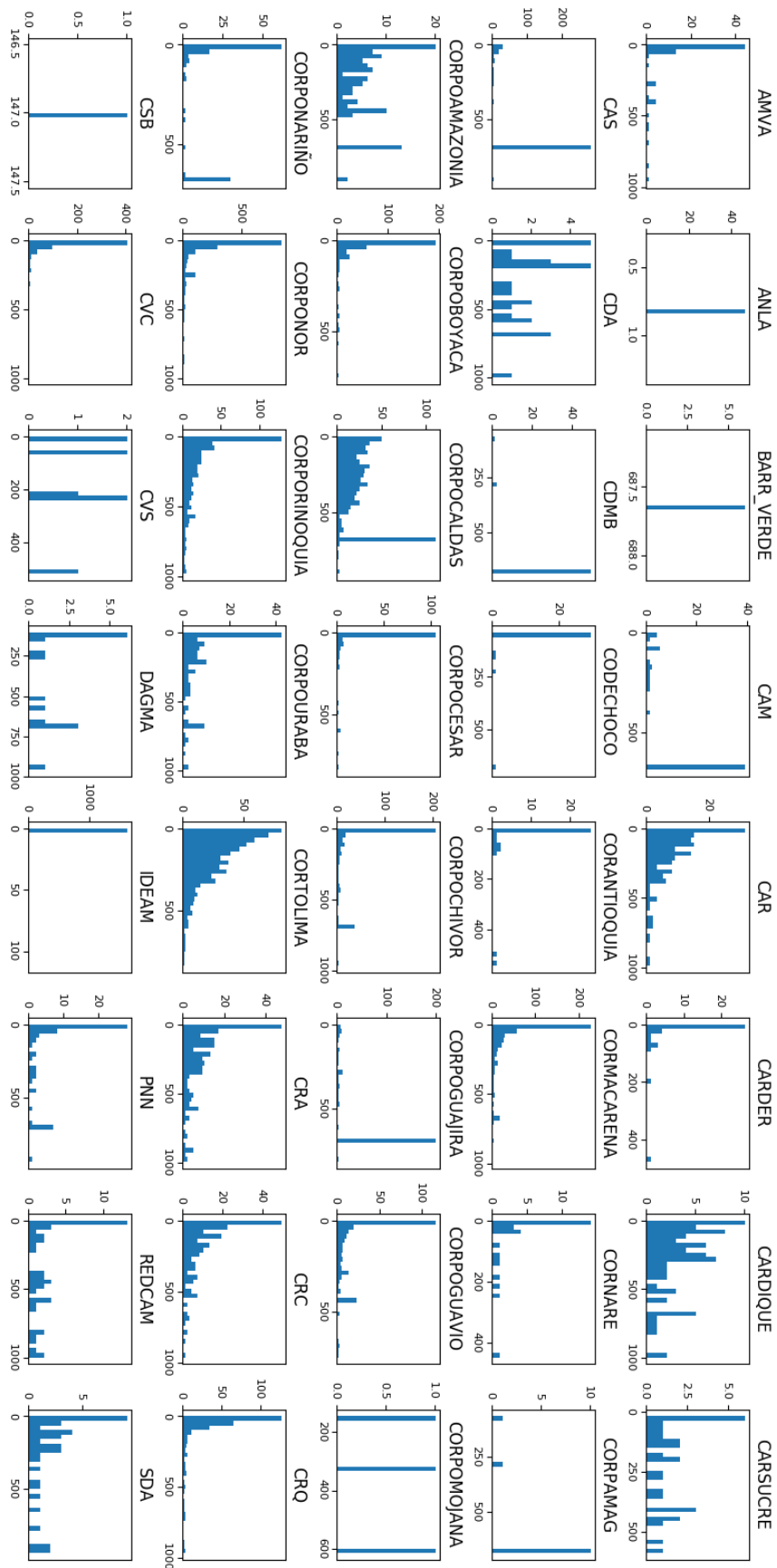
No reportaron: RIO_BOGOTA, MADRS, EPA, AMB

4.2 Tramos puntos finales

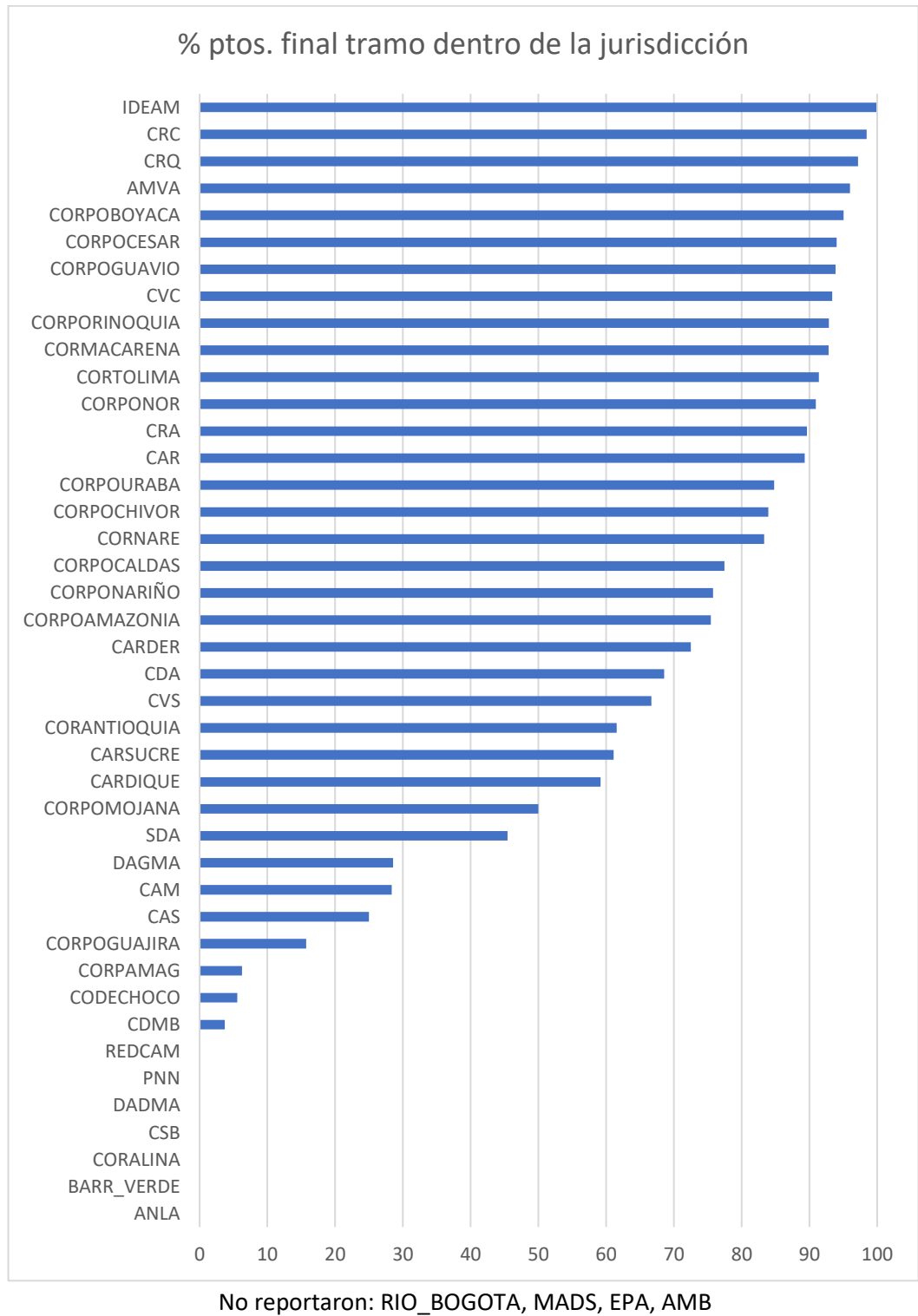


No reportaron: RIO_BOGOTA, MADS, EPA, AMB

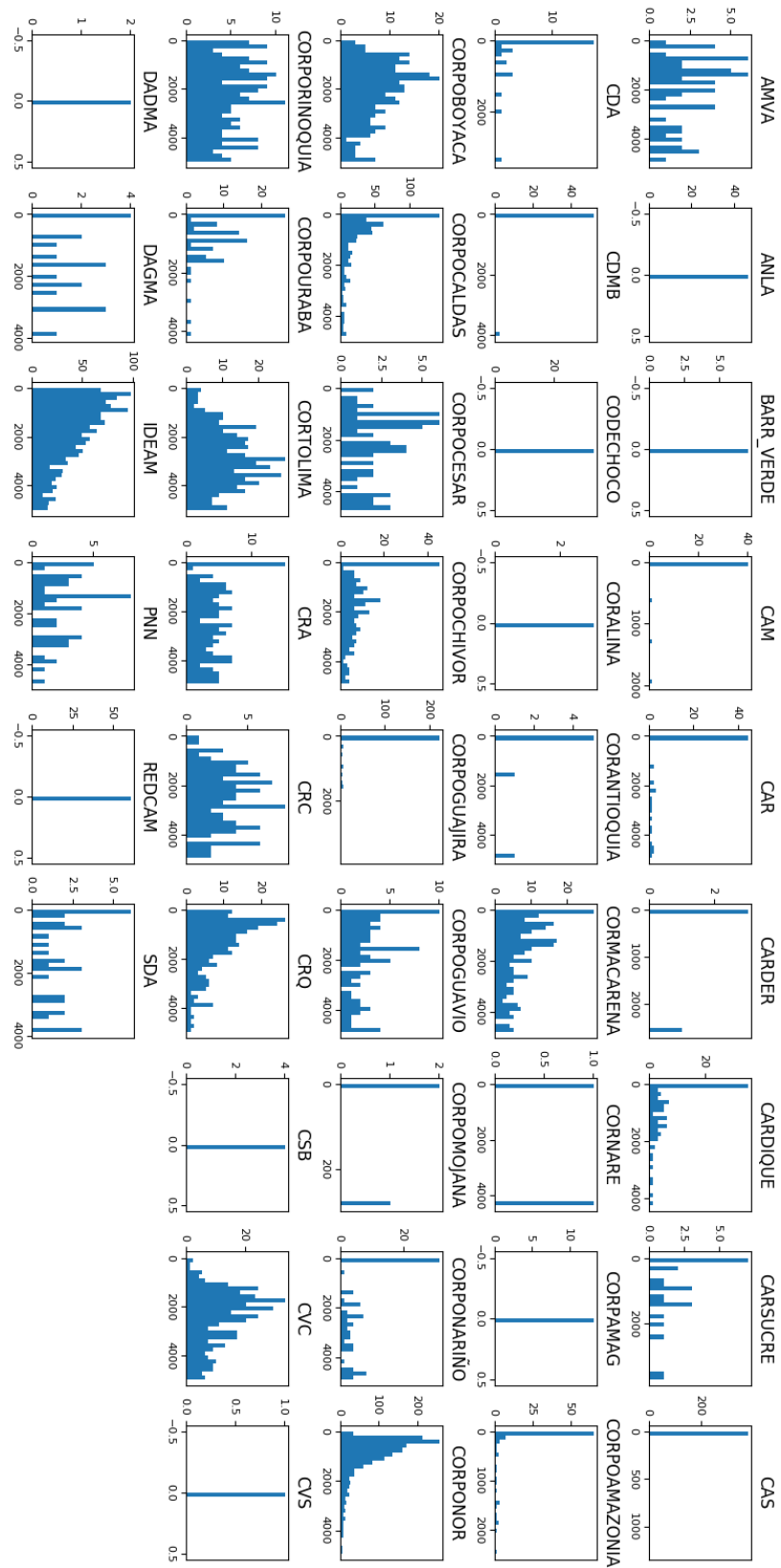
4.2.1 Histograma distancia puntos finales a red (Por corporación – Eje x en metros)



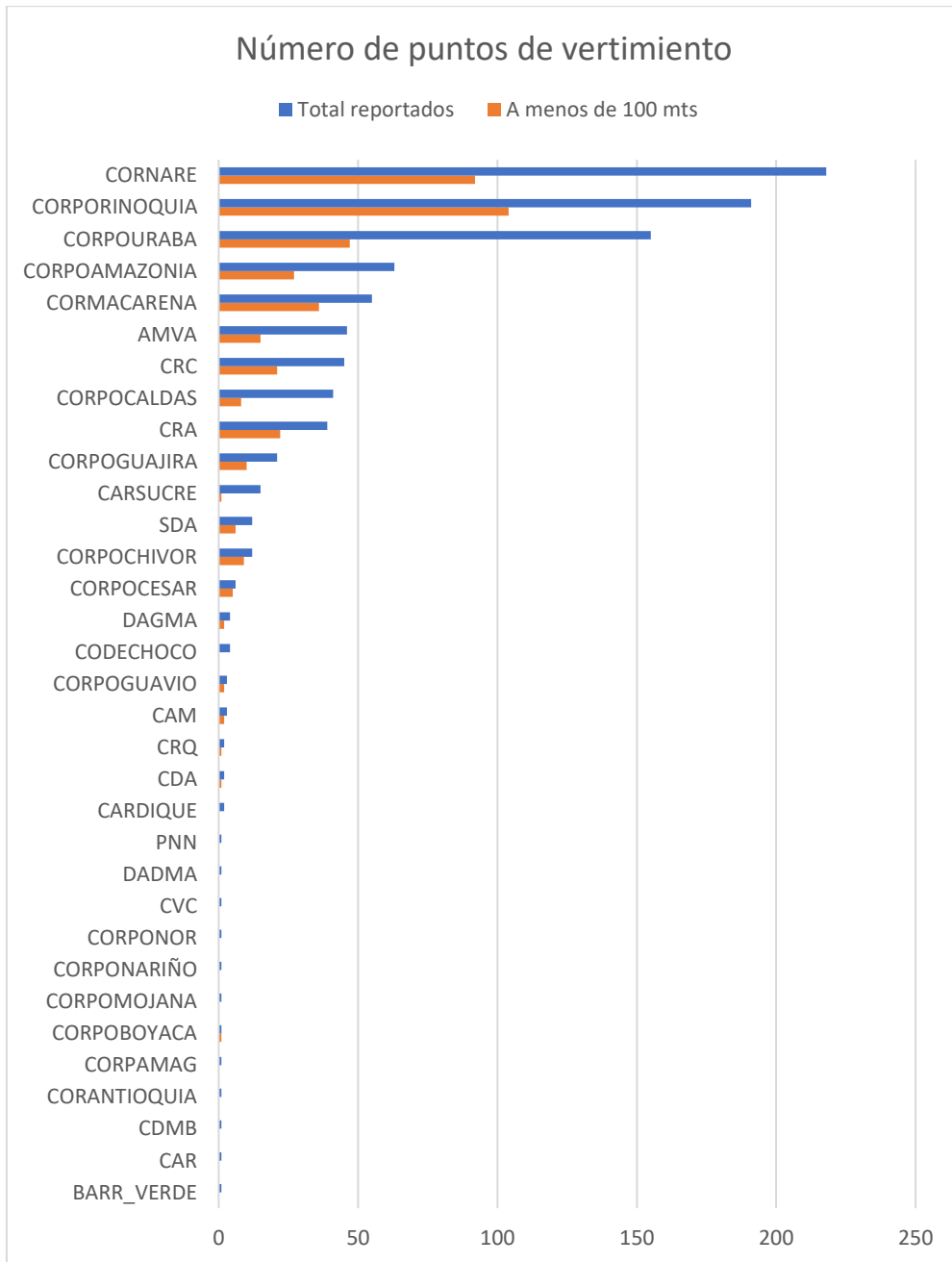
4.2.2 Puntos finales



4.2.3 Distancia tramos (Pi - Pf) por corporación – histograma (máx. 5000 mts)

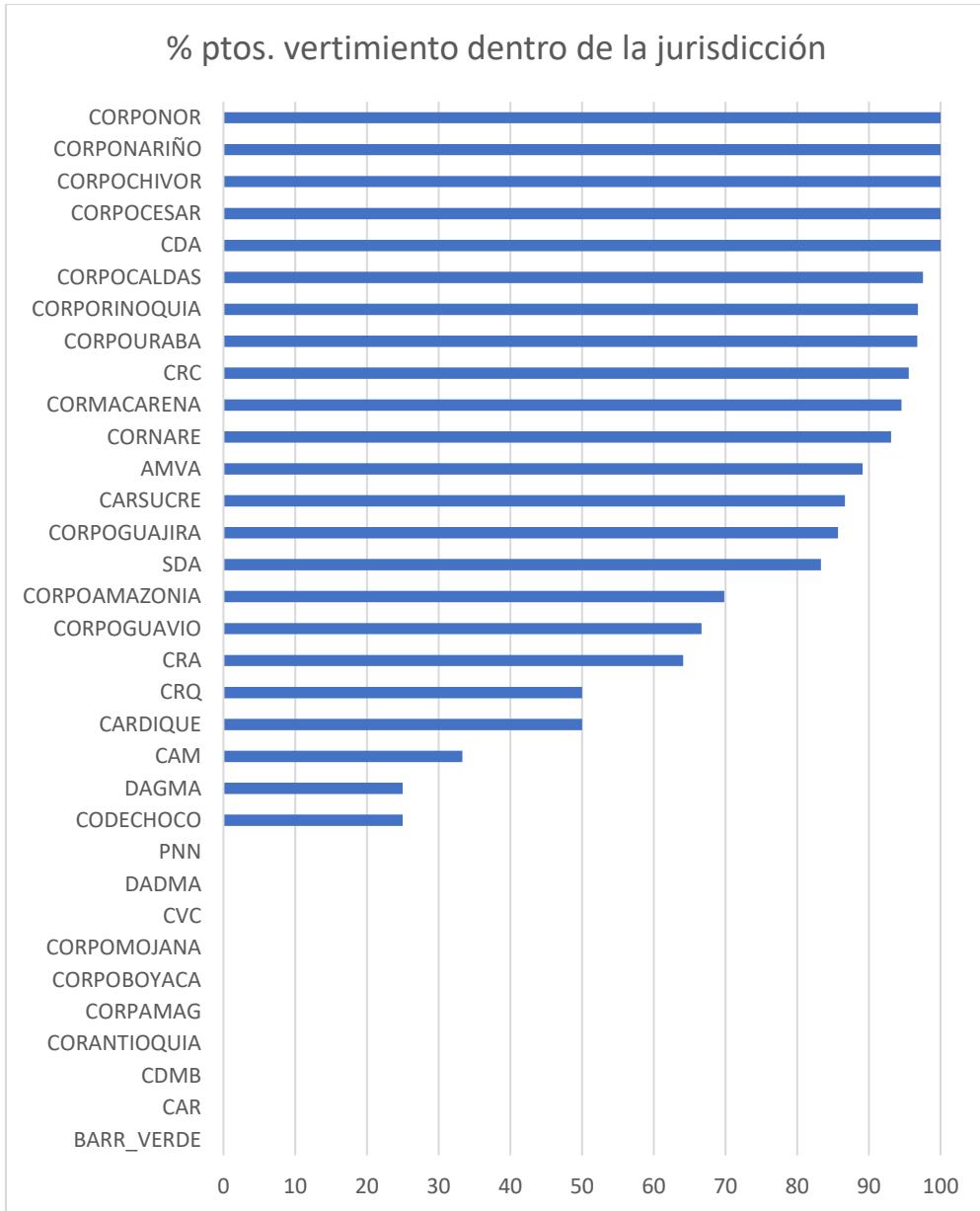


4.3 Puntos de vertimiento



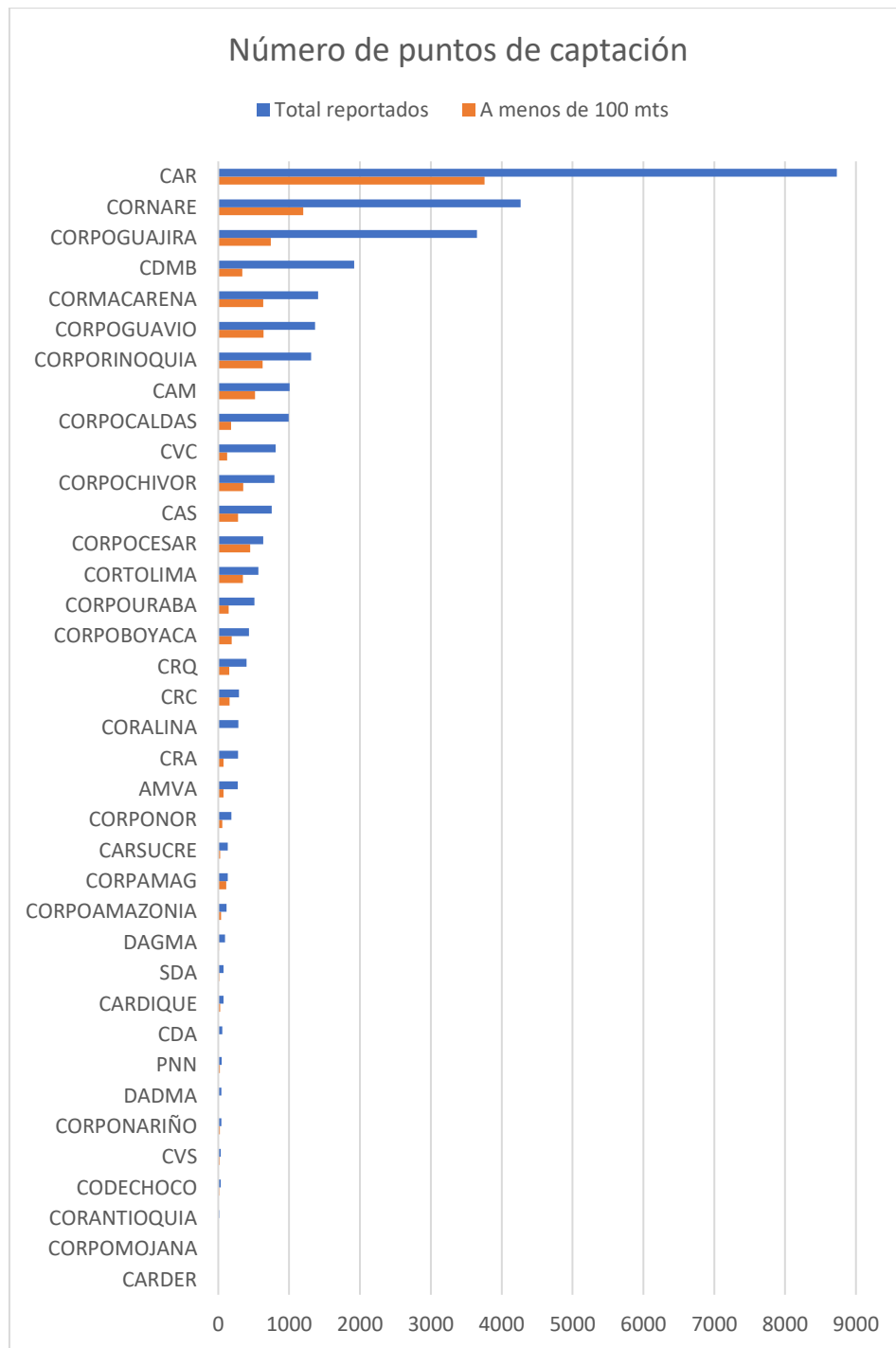
No reportaron: RIO_BOGOTA, AMB, REDCAM, MADS, ANLA, EPA, CORPOMOJANA, CSB, BARR_VERDE, DADMA, CVS, CORTOLIMA, CORPAMAG, CORALINA, CAS, CARDER, IDEAM

4.3.1 Puntos de vertimiento en la jurisdicción



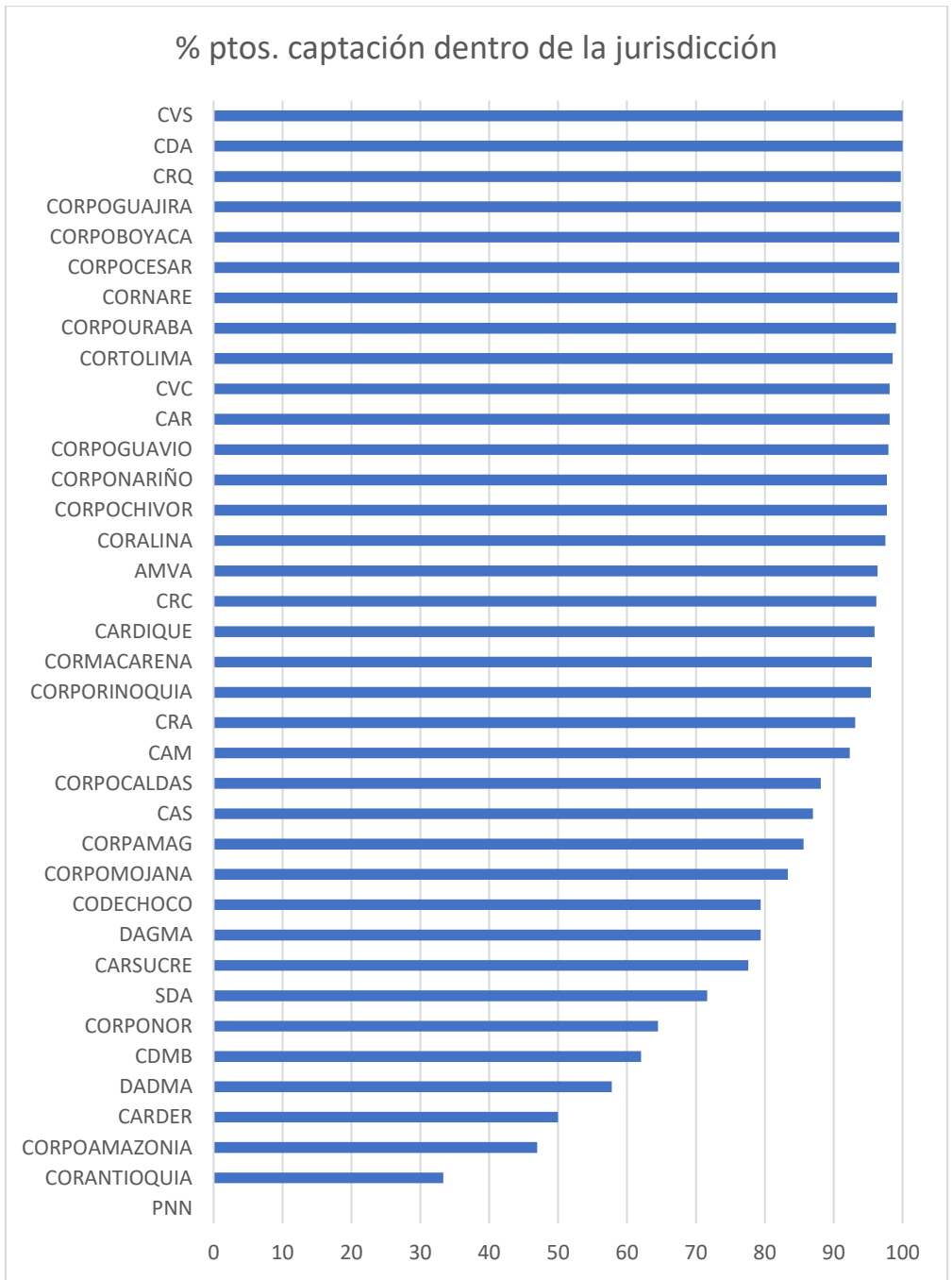
No reportaron: RIO_BOGOTA, AMB, REDCAM, MADS, ANLA, EPA, CORPOMOJANA, CSB, BARR_VERDE, DADMA, CVS, CORTOLIMA, CORPAMAG, CORALINA, CAS, CARDER, IDEAM

4.4 Puntos de captación



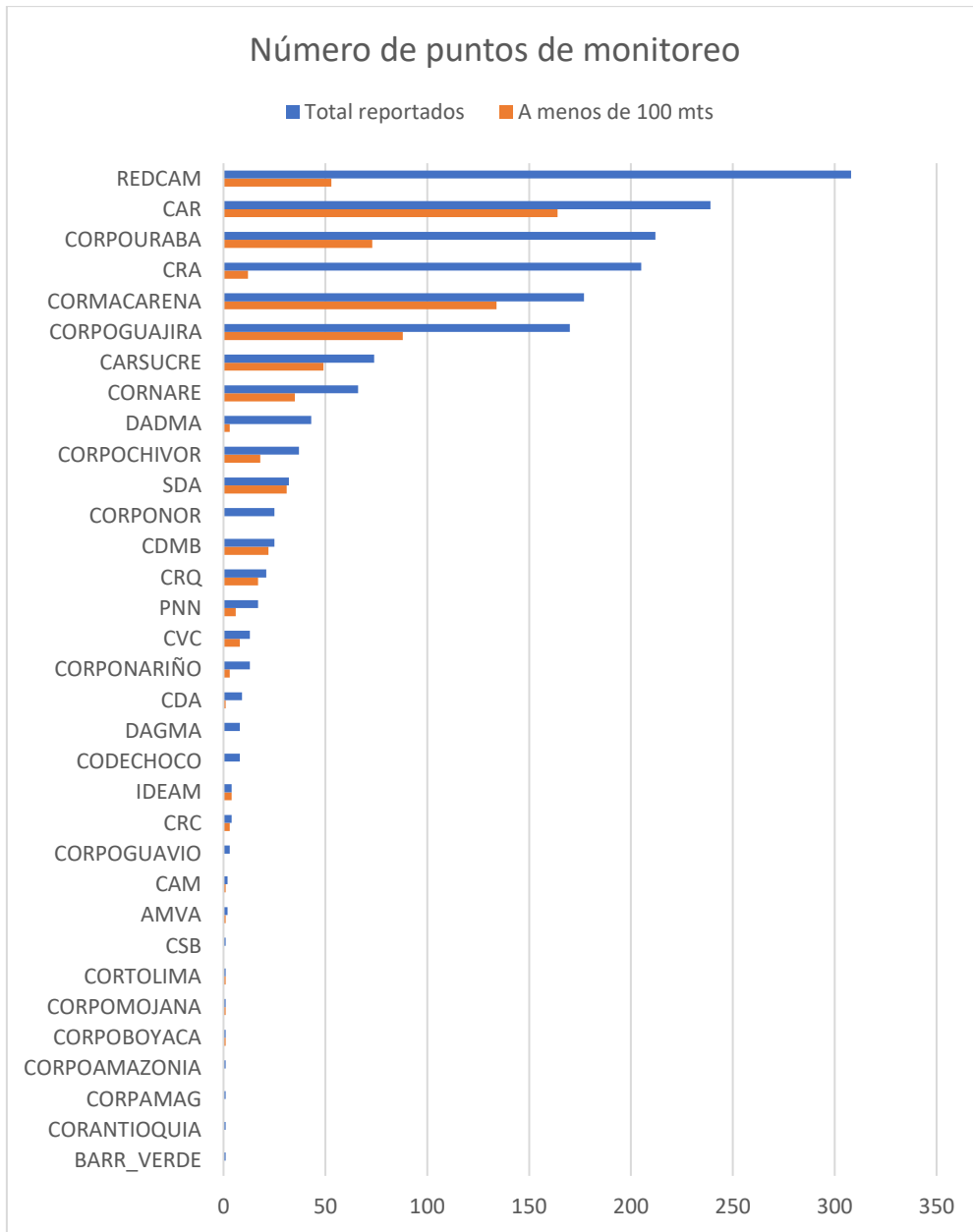
No reportaron: RIO_BOGOTA, REDCAM, MADS, IDEAM, EPA, CSB, BARR_VERDE, ANLA, AMB.

4.4.1 Punto de captación en la jurisdicción



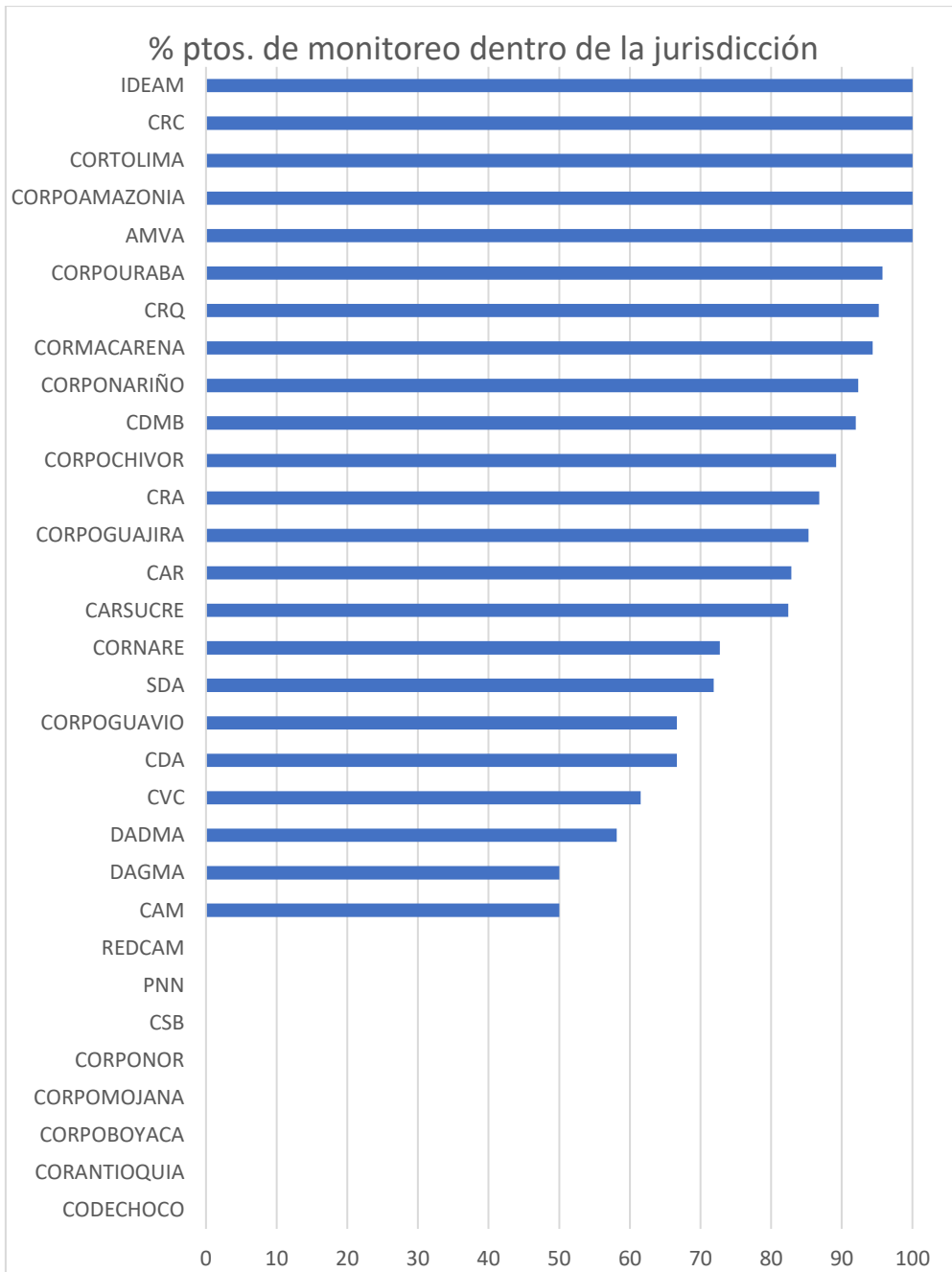
No reportaron: RIO_BOGOTA, REDCAM, MADS, IDEAM, EPA, CSB, BARR_VERDE, ANLA, AMB.

4.5 Puntos de monitoreo



No reportaron: RIO_BOGOTA, MADS, EPA, CVS, CORPORINOQUIA, CORPOCESAR, CORPOCALDAS, CORPAMAG, CORALINA, CAS, CARDIQUE, CARDER, BARR_VERDER, ANLA, AMB

4.5.1 Puntos de monitoreo en la jurisdicción

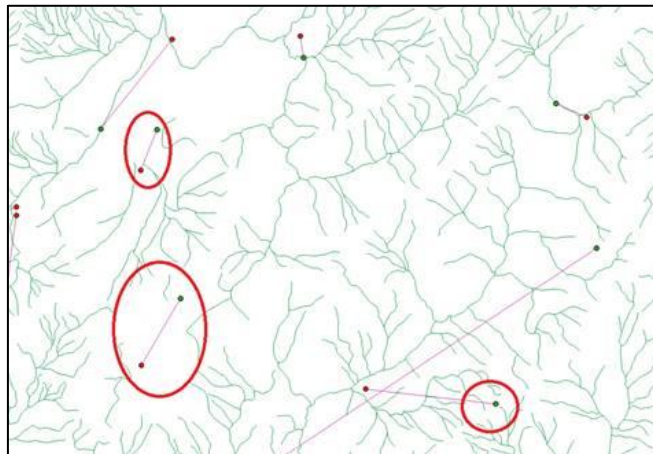


No reportaron: RIO_BOGOTA, MADs, EPA, CVS, CORPINOQUIA, CORPOCESAR, CORPOCALDAS, CORPAMAG, CORALINA, CAS, CARDIQUE, CARDER, BARR_VERDER, ANLA, AMB

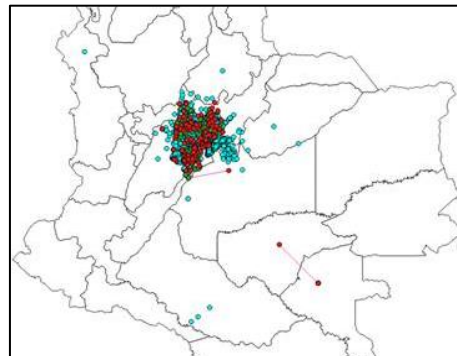
5. Conclusiones

El procedimiento propuesto permite identificar algunas inconsistencias de la georreferenciación, comunes a los puntos reportados por las CAR. Entre ellos, se destacan los siguientes:

- Tramos con longitud cero
Del total de puntos reportados para los tramos, el 13,44% de los tramos tienen longitud cero, es decir, igual punto de inicio y final
- Punto inicio-fin en tramos desconectados (de acuerdo con la red de drenaje)

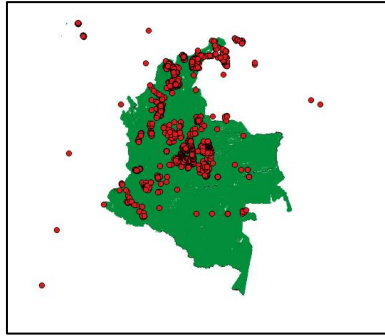


- Puntos por fuera de la jurisdicción de las corporaciones



Ejemplo: Corporación Autónoma de Cundinamarca

- Puntos por fuera del territorio nacional



Ejemplo: Puntos de monitoreo

Con la aplicación del procedimiento propuesto se pueden identificar problemas en los datos reportados por las CAR; además, con los indicadores presentados en las figuras, se puede informar rápidamente a las personas encargadas en las CAR para que hagan las correcciones necesarias y hagan nuevos reportes de los datos.

El mecanismo de validación basado en la distancia a la red de drenaje sencillo y drenaje doble muestra ser útil para validar la localización geográfica de los puntos, no obstante, deben revisarse los casos por los que se presenta para hacer una mejor interpretación y orientar la solución. Por ejemplo, en algunos casos esto ocurre porque la red de drenaje tiene discontinuidades.