Colombia multielement geochemical data set

March 2022

Recursos del Caribe, S.A. (www.cbmap.net)
Carl E. Nelson (carlericnelson@gmail.com)

This document describes regional geochemical survey data sets offered by Recursos del Caribe, S.A. (RdC). The data sets were created using, as a starting point, the following Colombian government and academic sources.

Servicio Geológico Colombiano (SGC-2009) map:

The SGC published a geochemical anomaly map of Colombia (Prieto et al., 2009) that displays anomalous areas for a variety of element combinations. RdC captured the anomalous areas as polygons, in MapInfo tables or ArcGIS shapefiles. No raw data was provided.

Servicio Geológico Colombiano (SGC-2010) data set:

The SGC released a compilation of multielement geochemical data in 2015 for 168,656 samples, including 72,536 stream sediment samples, representing 150 surveys conducted through 2010. This data set, available for several years as a free download from the SGC website, was replaced in recent years by a smaller data set for stream sediment samples only (MDRU-2016). Of the 168,656 total, the SGC-2010 data set included geochemical data for 103,142 samples; 65,514 were empty.

RdC converted the SGS database (provided as an MS Access mdb file) to GIS format (MapInfo or ArcGIS) and plotted sample locations. RdC did not make any changes to the geochemical data. Changes, modifications, and corrections to the geochemical data are described in, “Recursos del Caribe, S.A. (RdC-2020).”

Mineral Deposits Research Unit (MDRU-2016) data set:

In 2016, the MDRU, in conjunction with the SGC, published an atlas that displays multielement geochemical data for 75,238 stream sediment samples along with 56 separate shapefiles, each containing data for a single element. The MDRU-2016 data set is available for free download on the SGC website.

RdC consolidated the 56-file MDRU-2016 data set into a single MapInfo table or ArcGIS shapefile that compiles all of the geochemical data (in columns) for all of the samples (in rows). This format facilitates review of all of the data for any given sample. It also allows in-depth statistical evaluation (e.g., generation of correlation coefficients, multivariate analysis).

Recursos del Caribe, S.A. (RdC-2020) data set:

Using the SGC-2010 data set as a starting point (168,656 samples), RdC converted geochemical data (recorded as text) to numeric fields, eliminated 772 duplicate entries, corrected 542 data entry errors, and filled in 54,553 missing sample locations. The RdC-2020 data set, as MapInfo tables or ArcGIS shapefiles and incorporating 31,538,672 edits to the original SGC-2010 data set, can be confidently applied to exploration.

RdC's Colombia Multielement Geochemical Data Set is designed to work either on a stand-alone basis or with ColombiaMap, a separate GIS data set offered by RdC for mineral occurrences and land status (www.cbmap.net).

The table below summarizes the contents of RdC's Colombia Multielement Geochemical Data Set. A figure on the following page shows the sample locations.

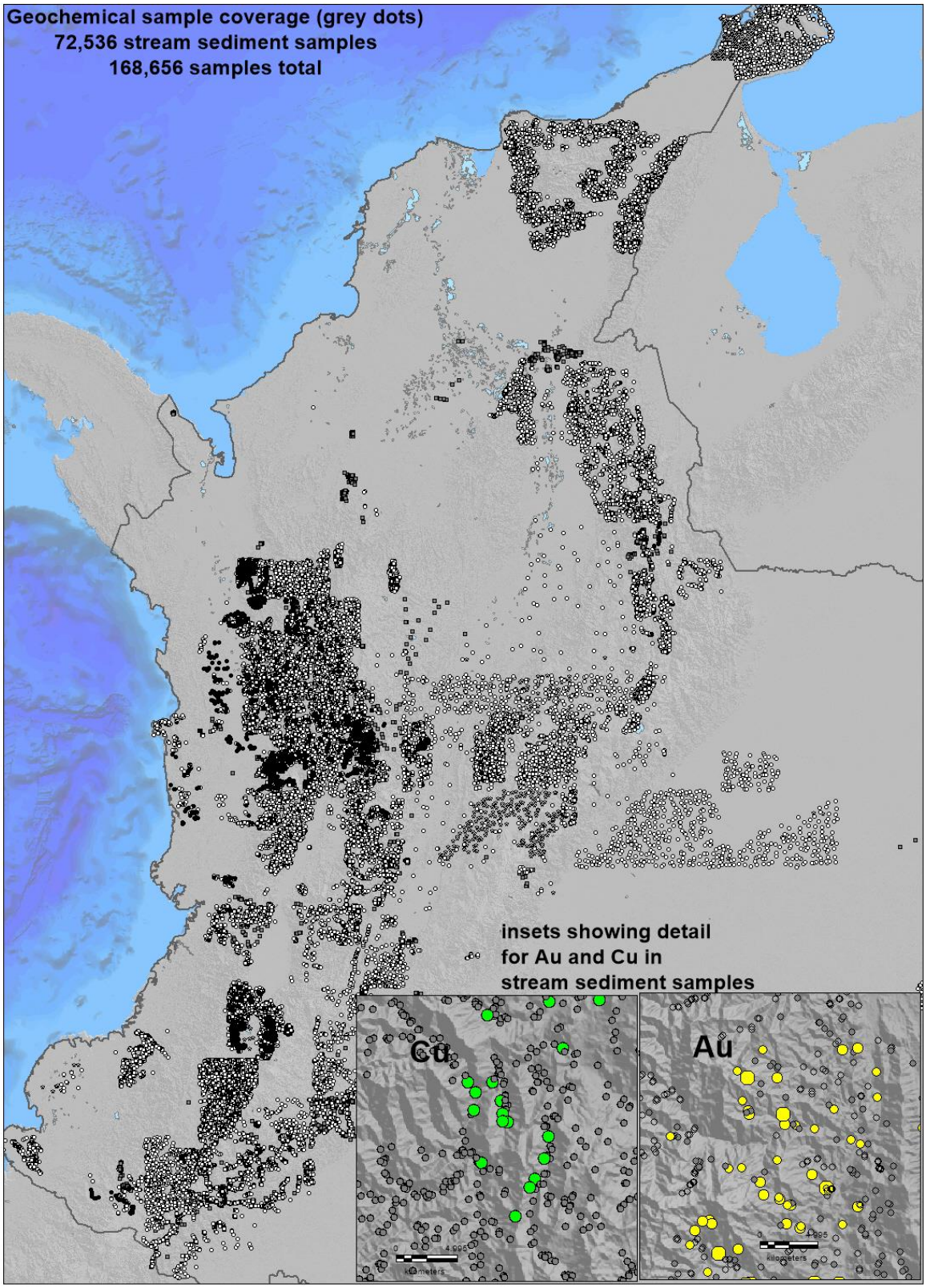
Sample Type	Number	Elements analyzed
Pan Concentrate	4,547	Ag, Al, As, Au, B, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, Ga, Hg, K, Mg, Mn, Mo, N, Na, Nb, Ni, P, Pb, Pd, Pt, S, Sb, Sc, Se, Sn, Sr, Th, Ti, Tl, U, V, W, Y, Zn, Zr
Rock	65,598	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Ni, P, Pb, Pd, Pr, Pt, Rb, S, Sb, Sc, Se, Si, Sm, Sr, Ta, Tb, Te, Th, Ti, Tl, U, V, W, Y, Yb, Zn
Soil	21,265	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, I, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Nd, Ni, Os, P, Pb, Pd, Pr, Pt, Rb, Re, Ru, S, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tl, Tm, U, V, W, Y, Yb, Zn, Zr
Stream Sediment	72,536	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Cl, Co, Cr, Cs, Cu, Dy, Er, Eu, F, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Nd, Ni, P, Pb, Pd, Pr, Pt, Rb, Re, Ru, S, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tl, U, V, W, Y, Yb, Zn, Zr
Water	3774	Ag, Al, As, Au, B, Ba, Be, Bi, Br, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, Na, Nb, Nd, Ni, Pb, Pd, Pr, Pt, Rb, Re, Ru, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tl, U, V, W, Y, Yb, Zn, Zr

Pricing

RdC offers its revisions, edits and corrections to the data sets described above with the aim of eliminating the time and cost involved in making the corrections needed to put these data sets into a format that can be reliably applied to exploration. The original SGC data sets are also provided, at no charge, for reference purposes.

Colombia multielement geochemical data set: $36,000.00
Multielement geochemical data for selected areas: $1 per sample

Prospective clients are invited to "test drive" the Colombia multielement data set before making a purchase. Clients who update their data set receive a 50% credit for their previous purchase.


Geochemical sample coverage (grey dots)
72,536 stream sediment samples
168,656 samples total
insets showing detail
for Au and Cu in
stream sediment samples
Cu
Au

Colombia multielement geochemical data set
Detailed description of revisions made by RdC

Servicio Geológico Colombiano (SGC-2009) map:

The SGC published a geochemical anomaly map of Colombia (Prieto et al., 2009) that displays anomalous areas for a variety of element combinations. No raw data was provided. The anomalous areas were captured as polygons which RdC provides as MapInfo tables or ArcGIS shapefiles. Raw geochemical data was not released with the Prieto et al. (2009) map of geochemical anomalies. However, the map was drawn largely from the SGC data set released in 2010.

Servicio Geológico Colombiano (SGC-2010) data set:

The SGC released multielement geochemical data for 168,656 samples in 2015, including 72,536 stream sediment samples, representing a compilation of data from 150 surveys conducted through 2010. This data (in MS Access format) was made available for several years as a free download from the SGC website at: http://geoportal.sgc.gov.co/geoportalsgc/catalog/quicklink/basesDatosPublicacion.page. No license was required. Unfortunately, the original SGC data set is no longer posted on their website. The link ceased to function after a website update; users are redirected to: https://www2.sgc.gov.co/sgc/mapas/Paginas/geoportal.aspx.

There are six sample location fields in the original SGC-2010 data set containing up to three sample location coordinate pairs. (Colombia uses five different transverse mercator zones). RdC selected the appropriate coordinate pair for plotting according to the procedure described below. Unfortunately, many of the SGC-2010 sample locations are obviously out of place. In addition, the SGC-2010 data set provides geochemical data in text rather than in numeric format. The SGC-2010 data set provided here is intended for plotting of the original SGC-2010 sample locations. RdC's edits, changes and corrections to the original SGC-2010 data set are described later in this document, in the section, "Recursos del Caribe, S.A. (RdC-2020) data set."

With respect to the uncorrected SGC-2010 data set, samples were located using the coordinate system recorded under Origin_Coord_Original and locations as recorded under the Coord_Este_Original and Coord_Norte_Original fields. Samples with "Magnas Bogota" under Origin_Coord_Original were located using the Bogota Transverse Mercator zone, Magnas-Sirgas datum. They were coded as "BOG-M" in RdC's added Coordsys_Used field. Samples with "Bogota" under Origin_Coord_Original were located using the Bogota Transverse Mercator zone, Bogota datum. They were and coded as "BOG-B" in the Coordsys_Used field.

For samples recorded as “Este-Este,” “Este,” or “3-Este Este,” RdC used the East Transverse Mercator field, Bogota datum and coded them as “E-B.” Samples recorded as “2-Este Centro” and “Este-Centro” were located using the East Central Transverse Mercator Zone, Bogota datum and were coded as “EC-B.” Samples recorded as “Magnas Oeste” were located using the West Transverse Mercator zone, Magna-Sirgas datum (only 4 samples).

Some samples had no Coord_Este_Original or Coord_Norte_Original. Of these, 488 samples recorded as “4-Oeste” were located using the West Transverse Mercator zone, Bogota datum with X, Y locations from Coord_Este_Bogota and Coord_Norte_Bogota. They were coded as “BOG-B.”

Samples recorded as “UTM” were located using UTM zone 18 North WGS 84 and were coded as “UTM18-W.” Samples (3100) recorded as “UTM zona 18 N” or “UTM zona 18 North, WGS84” were obviously Lat/Long coordinates (not UTM) and so were located using Lat/Long WGS84 and were coded with “LL-W.” Samples recorded as “WGS” were located using the Bogota Transverse Mercator Zone, Bogota datum and were coded as “BOG-B.” Another 1933 records were built using UTM Zone 18 North, WGS84 datum and were coded as “UTM 18-W.”

Samples with no entry for Origin_Coord_Original and with conflicting entries (4124) for Coord_x_Original versus Coord_x_Bogota were located using Coord_x_Bogota in the Bogota Transverse Mercator Zone and were coded as “BOG-B.” Samples with no entry for Origin_Coord_Original and with the same (4129) entries for Coord_x_Original versus Coord_x_Bogota were located using Coord_x_Bogota in the Bogota Transverse Mercator Zone and were coded as “BOG-B.”

Mineral Deposits Research Unit (MDRU-2016) data set:

The SGC published a Geochemical Atlas for Colombia (2016) that provides stream sediment geochemical anomaly maps at a scale of 1:6,000,000. The Atlas can be viewed interactively at: http://srvags.sgc.gov.co/JSViewer/Atlas_geoquimico_2016/. The original, licensed MDRU-2016 data set is available for free download at: https://miig.sgc.gov.co/Paginas/Resultados.aspx?k=atlas%20geoquimico%202016. The Atlas was updated in 2018 with a marginal increase in the number of samples analyzed.

The MDRU-2016 data set contains multielement geochemical data for 75,238 stream sediment samples delivered as an ArcGIS geodatabase and .mxd files. The majority are new samples; only 25,429 are repeated from the SGC-2010 data set. Each of 56 elements in the MDRJU-2016 data set has its own set data files that, in addition to analytical results, contain entries for sample type, sample description, sample preparation, laboratory, report number, project, observations, digestion, analytical technique, and, finally, whether or not the analytical results were included in or excluded from the plots of geochemical anomalies. MDRU

removed geochemical data judged as unreliable eliminating, for instance, half of all values for arsenic, 25% of all values for gold, and 75% of all values for silver. Most of the excluded values were analyzed using emission spectrometry.

The MDRU-2016 structure of separate data files for each element impedes the review of all of the geochemical data for any given sample. The user would have to open 56 separate files. The MDRU-2016 structure also impedes any form of statistical analysis based on a comparison of data variability between elements (e.g., correlation coefficients, principal components analysis, factor analysis). To remedy these issues, RdC built one file that contains all of the samples (as rows) and all of the geochemical data for every element (as columns). No changes were made to sample locations or to the geochemical data in the MDRU-2016 data set. The steps taken to create this "transposed" data set are described as follows.

RdC first compiled a full list of unique samples using the PK_ID_MUES field in the original MDRU-2016 data set. This primary key was used to link MDRU sample locations to data stored in each of the 56 separate MDRU files. Sample locations and geochemical data (with fields for 77 elements) were then saved into RdC's revised MDRU-2016 data set using the LatLong WGS84 datum. In RdC's revised MDRU-2016 data set, geochemical data excluded by MDRU is coded as -555; cells for which no geochemical data is available are coded as -999, and geochemical data reported as less than the detection limit is set equal to one half the detection limit. Descriptive fields in RdC's revised MDRU-2016 data set for which the entry varies according to element, e.g., analytical technique, are left empty but can be reviewed by opening the original MDRU geodatabase. Excluded geochemical data can also be viewed by opening the original MDRU geodatabase.

Recursos del Caribe, S.A. (RdC-2020) data set:

Recursos del Caribe S.A. (RdC) conducted a thorough review of the SGC-2010 data set; converted geochemical data from text to numeric format; made 54,553 corrections to incomplete and/or mistaken sample locations (32% of the data set); made 542 corrections to incorrectly recorded analytical data; resolved 772 duplicate data entries; made 31,538,672 text edits; put the data set into GIS format; added drainages, digital elevation models, shaded relief, infrastructure, a topographic map index, and departmental boundaries; and, finally, created project files (ArcGIS) and workspaces (MapInfo) to facilitate display of the data. RdC's revised SGC data set, incorporating changes to the SGC-2010 data set described in detail on the following pages, is delivered in MapInfo or in ArcGIS format.

RdC's Colombia Multielement Geochemical Data Set is a stand-alone GIS product but is designed to work with RdC's ColombiaMap, a separate GIS data set for mineral occurrences and land status (www.cbmap.net). The data set is available in MapInfo or ArcGIS format.

Tables in the Colombia Multielement Geochemical GIS Data set

The Colombia Multielement Geochemical Data Set draws from an MS Access database, COGeochem.accdb, prepared by RdC. Tables in the MS Access database record the steps taken, as described in the following pages, in cleaning and correcting the original SGC Data Set sample locations and analytical data.

MAPINFO_MAPCATALOG: A table that MapInfo uses to display "mappable" tables in the COGeochem data set. This table is used to plot points from ODBC (Open Data set Connectivity) tables that are described below.

MUESTRAS_ORIG_ASSAY: This table, in text format, replicates the 2015 SGC data set, including minor changes made by the SGC through November 2017. The changes include a total of 776 new values for As (4), Ag (50), Ti (2), Ta (111), Sr (4), Na (1), Mn (1), Mg (120), Fe (478) and Ca (1).

MUESTRAS_PROCESSED (data entry errors): This text table contains a "processed" version of the 'muestras_orig_assay' table. First, null values in the original SGC assay fields are converted to -999, indicating that no analysis is available. Then, the following changes are made to the SGC analytical data:

1) A total of 772 instances for which a single sample contains two conflicting analytical values (e.g., Ag appears both as <0.2 ppm and as <0.2 ppb) are corrected (see the table: '1_TWO_ASSAYS_PER_ELEMENT_EDITS').
2) A total of 31,538,672 additional text edits, described below, are listed in the table: '2_FIND_AND_REPLACE_EDITS.'
    a) Cell values of -X and <X are treated as below a lower detection limit of X and are assigned a cell value equal to 1/2 of the lower detection limit;
    b) Cell values of >Y and GY are treated as greater than an upper detection limit of Y and are assigned a cell value of Y+1;
    c) Blanks (zero length strings) and entries such as trz, N, and ND are converted to -999 to indicate no data for 30,223,825 cell values.
    d) Spaces in the data (e.g., 1 234) are replaced with commas (e.g., 1,234) and commas, marking the decimal point, are replaced with periods for 1,314,847 cell values.

The Find and Replace Edits table includes the following columns: **uid**: unique, auto-numbered ID (primary key); **cellvalue**: analytical data present as text in an assay field in the original SGC data set and in the table 'muestras_analisis2012;' **thefieldname**: the assay field in which the cellvalue was found in

'muestras_analisis2012'; **thetablename**: the name of the table in which the cellvalue was found. For analytical data from the SGC data set, thetablename is always 'muestras_analisis2012'; **valuetype1**: categorizes types of cellvalues. Used in organizing bulk edits to the data; **valuetype2**: categorizes types of cellvalues. Used in organizing bulk edits to the data; **newval**: a new cellvalue in the table 'muestras_processed' that replaces all instances of a particular cellvalue in the table 'muestras_analisis2012;' **updated_status**: indicates whether a cellvalue in a particular field has been changed. If set to 'OK', the cellvalue in the 'muestras_analisis2012' table has been searched for and changed to newval in the 'muestras_processed' table. To make additional changes to the "muestras_processed" table and, at the same time record those changes, first make the change to newval and replace 'OK' in the updated_status field with a zero-length string. Then use the form 'Update Assay Fields' to replace each newly edited cellvalue in the 'muestras_processed' table; **last_updated_to**: which newval was last used in replacement of cellvalue in the indicated field of 'muestras_processed'; **last_updated_date**: the date of last replacement of this cellvalue in the indicated field of 'muestras_processed', using the 'Update Assay Fields' form; **note**: explanation regarding the choice of newval for replacement of the original cellvalue; **temporder**: a field that was used during hand edits of the 'processing_edits_log' table.

3) Corrections, still in text format, in the '3_DIVIDE_BY_1000_EDITS' table include conversions for analytical data (to ppb) for analyses improperly recorded in the SGC data set as ppm. A total of 542 emission spec values were corrected. The SGC-2010 data set contains some (fewer than 1%) suspiciously high emission spec values for Ag, Au, As, B, Ba, Be, Bi, Br, Cd, Co, Cr, Cs, Cu, Ga, I, La, Mn, Mo, Nb, Ni, P, Pb, Pd, Pt, Sb, Sc, Se, Sn, Sr, Ta, Te, Ti, V, W, Y, Zn, and Zr. Rather than discard all emission spec values (a total of 702, 320 analyses), the RdC-2020 data set leaves emission spec data for these elements untouched.

4) The table '4_2017_UPDATES' includes a total of 882 edits made to gold values that are listed in the SGC data set as less than a high detection limit (<20, <10, <5, <2, <1, <0.5 and <0.1 ppm Au). All of these less-than-a-high-detection-limit values were changed to -888 to avoid spurious gold anomalies that would result if less-than-a-high-detection-limit (dl) values were retained or changed to 1/2 of the dl. Similarly, for 6406 Ag analyses listed as less-than-a-detection-limit of 1, 5, 10 or 20 ppm, cellvalues were changed to -888. A total of 2114 Cu and 5933 Mo values listed as less-than-a-detection-limit of 5 ppm or higher were also changed to -888 along with

18 Pb, 9526 Zn and 319 Ni values listed as less-than-a-detection-limit of 20 ppm or higher. These add up to a total of 25,204 changes made to eliminate spurious anomalous values for Au, Ag, Cu, Pb, Zn and Ni related to high detection limits. Finally, a series of 138 data entry errors for cobalt (all with uniformly high values of 5%) were changed to 5 ppb (detection limit).

MUESTRAS_PROCESSED (sample location errors):
The 'muestras_processed' table includes sample location coordinates in decimal latitude and longitude format (WGS 84 datum). These locations incorporate all of the corrections made by RdC to the original SGC sample locations. For instance, latitude longitude coordinates are missing for a total of 54,553 samples (32%) in the original SGC data set. Consequently, latitude longitude coordinates from the SGC data set were not used to locate samples. Instead, new decimal latitude and longitude coordinates (WGS84 datum) were generated using coordinates from the 'coord_este_original' and 'coord_norte_original' fields and the appropriate zone as recorded in the 'origen_coord_original' field. This step provided locations for two-thirds of the samples (114,387).

For the remaining (54,269) samples, the 'coord_original' fields were either empty or were incorrect. Most of these samples were successfully located by using one of remaining Bogota datum coordinate sets. All of these locations were then verified using location information provided in the fields for 'departamento' and 'plancha'. This process revealed 1132 data entry errors, now corrected, in the zone listed in the original SGC data set.

Finally, a total of 1571 sample locations (1% of the data set) plot within quadrangles that do not closely coincide with the quadrangles recorded in the original SGC data set. These locations are plotted, for now, near latitude 0 and longitude 0 until their actual locations can be resolved.

Five tables in the RdC-2020 data set incorporate the changes described above.

ASSAYS_PAN_CONCENTRATE: analytical results for 4547 pan concentrate samples.

ASSAYS_ROCK: analytical results for 65,598 rock samples.

ASSAYS_SOIL: analytical results for 21,265 soil samples.

ASSAYS_STREAM_SEDIMENT: analytical results for 72,536 stream sediment samples.

ASSAYS_WATER: analytical results for 3774 water samples.

Queries in the Colombia Multielement Geochemical GIS Data set

ASSAY_QUERY: takes cell values for each element in the 'muestras_processed' table and converts them from text into numeric entries. Data reported in different units is merged in the process. For instance, Cu recorded for some samples as Cu_percent, for others as Cu_ppm and for still others as Cu_ppb is merged into a single column (reported in units of ppb, ppm or pct as indicated).

PAN_CONCENTRATE_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 4547 pan concentrate samples.

ROCK_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 65,598 rock samples.

SOIL_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 21,265 soil samples.

STREAM_SEDIMENT_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 72,536 stream sediment samples.

WATER_ASSAYS_QUERY: a subset of Assay_Query containing analytical results for 3774 water samples.

INFORME_LIST: a subset of Assay_Query containing a list of analytical reports.

PROJECT_LIST: a subset of Assay_Query containing a list of projects; listed projects refer to reports that contain the raw analytical data.

MAKE_ASSAYS_PAN_CONCENTRATE_TABLE_QUERY: takes the **PAN_CONCENTRATE_ASSAYS_QUERY** and remakes the **ASSAYS_PAN_CONCENTRATE** table.

MAKE_ASSAYS_ROCK_ TABLE: takes the ROCK_ASSAYS_QUERY and remakes the ASSAYS_ROCK table.

MAKE_ASSAYS_SOIL_ TABLE: takes the SOIL_ASSAYS_QUERY and remakes the ASSAYS_SOIL table.

MAKE_ASSAYS_STREAM_SEDIMENT_ TABLE: takes the STREAM_SEDIMENT_ASSAYS_QUERY and remakes the ASSAYS_ATREAM_SEDIMENT table.

MAKE_ASSAYS_WATER_ TABLE: takes the WATER_ASSAYS_QUERY and remakes the ASSAYS_WATER table.