



UNIVERSIDAD CATOLICA  
DE LA SANTISIMA CONCEPCION

---

## Exploración de datos geoquímicos mediante análisis de datos composicionales (CoDA): Aplicación a los suelos urbanos de Arica.

Eduardo Oñate<sup>a</sup>, Pedro Tume<sup>b</sup>, Giuliani Coluccio<sup>c</sup>, Jorge Galleguillos<sup>d</sup>

(a) Estudiante, Departamento de ingeniería Industrial, Facultad de ingeniería, Universidad Católica de la Santísima Concepción.

(b) Profesor Titular, Departamento de ingeniería Civil, Facultad de ingeniería, Universidad Católica de la Santísima Concepción.

(c) (d) Profesor Asistente, Departamento de ingeniería Industrial, Facultad de ingeniería, Universidad Católica de la Santísima Concepción.

### Resumen

*Objetivos:* El propósito del siguiente trabajo consiste en elaborar un análisis exploratorio de datos geoquímicos mediante CoDA en los suelos urbanos de Arica, para así determinar su origen, siendo este natural, antrópico o mixto. Como objetivos secundarios se tiene el comparar la estadística clásica (La cual trabaja en el espacio Euclidiano) versus la estadística composicional (La cual trabaja en el espacio Simplex de Aitchison J. Por último la comparación de los paquetes “*compositions*” y “*robCompositions*”).

*Materiales y métodos:* Para llevar a cabo dichos análisis, se utilizó la base de datos otorgada en un informe de CONAMA 2009, los elementos a investigar son As, Ba, Cr, Cu, Ni, Pb, V y Zn. Estos datos fueron analizados desde la perspectiva de los datos composicionales mediante la transformación *log-ratio* centrada o *clr*. Se realizó un resumen analítico de los datos apto para los datos de composición y un análisis Clúster AC y de componentes principales ACP con el paquete “*compositions*” y “*robCompositions*” (El cual tiene un mejor manejo de *Outliers*), para luego comparar los resultados con los obtenidos entre ambos paquetes y con estudios anteriores basados en estadística clásica.

*Resultado y Discusión:* Con el paquete “*compositions*” y sin el tratamiento de datos atípicos, los resultados determinan que todos los elementos tienen fuentes tanto naturales como antrópicas, sin embargo los oligoelementos As, Pb, y Cu destacan por su precedencia antrópica la cual está ligada a las diversas actividades de la ciudad de Arica. En cuanto al “*robCompositions*” los oligoelementos que más destacan por su precedencia antrópica son As, Cu y Zn. Por último el trabajar los datos de composición con la estadística composicional Simplex, muestra a más detalle el comportamiento de los elementos, mostrando posibles nuevas fuentes de origen de estos.

*Conclusiones:* En general los metales trazas con más fuentes antrópicas fueron el Cu, Pb y As, mientras que el Ni y V presentaron una baja contaminación. Al comparar ambos PCA se obtienen resultados similares, sin embargo las componentes del PCA CoDA explican de forma más acertada el panorama real de la contaminación por metales trazas en estos suelos.

### 1. Introducción

El suelo en conjunto con el agua y el aire son los soportes sobre los que se asientan todos los procesos vitales que se desarrollan en el planeta tierra. La formación del suelo es producto de múltiples procesos tanto químicos, físicos y biológicos sobre el medio rocoso original, siendo este la parte más dinámica de la superficie terrestre y con más interacción humana. Producto de la actividad del hombre (ganadería, agricultura, industrias, minería, etcétera.) el equilibrio del suelo,

puede romperse, generando una modificación negativa comúnmente llamada degradación. La presencia en los suelos de concentraciones nocivas de algunos elementos químicos y compuestos, es un tipo especial de degradación que se denomina contaminación. El contaminante está siempre en concentraciones mayores de las habituales y en general tiene un efecto adverso sobre algunos organismos. Por su origen pueden ser geogénico o antropogénicos (Galán, 2004)

El suelo urbano por definición es toda extensión de tierra donde se ubica una determinada ciudad o pueblo y presenta modificaciones como lo son los pavimentos, alcantarillado, etcétera por ende están en constante interacción con la actividad humana, lo cual debido a estos últimos pueden alcanzar altos niveles de contaminación y polución, siendo potencialmente peligrosos para las personas y demás organismos que habitan en ellos. Entre las contaminaciones más comunes existentes en las últimas décadas está la contaminación por metales pesados u metales trazas, debido a la rápida expansión e industrialización de las zonas urbanas.

Chile depende de actividades económicas que le significan al país distintos focos de contaminación. En la zona central de Chile, especialmente en las regiones V, VI y Metropolitana, las principales actividades económicas que significan un foco de contaminación son la minería, destacándose dentro de esta área la VI región. Además, se realizan actividades de saneamientos y de disposición de residuos. Sin embargo, las actividades mineras destacan en todo el país, al provocar un alto impacto ambiental para el suelo, debido a los procesos asociados a la exploración y explotación, así como de los procesos de transformación piro metalúrgicos, siderometalúrgicos y refinación de sustancias minerales y de sus productos (Ministerio de Medio Ambiente, Chile, 2011).

Entre 1981 y 1990, INIA desarrolló dos estudios prospectivos del contenido total de As, Cd, Mo, Pb y Zn en suelos aluviales entre los ríos Huasco (III Región) y Simpson (XI Región) (INIA, 1986, 1990; González, 1994a, 1994b; González y Blaser, 1997; González et al., 1997; ECUS, 1996), en el cual en síntesis evidencian la necesidad de contar a un futuro con líneas base para los metales pesados a través de todo el país. Pues cada elemento posee un comportamiento diferente, el que depende de la composición de los materiales parentales del suelo y de la exposición del suelo a procesos de contaminación.

En retrospectiva a la ciudad de Arica, existen pocos informes relacionados con el contenido de elementos trazas en los suelos urbanos. Esta ciudad portuaria ha sido expuesta a lo largo de su historia a diversos focos de contaminación como lo fue el año 1984 y 1985 donde la empresa PROMEL procesadora de metales pesados Ltda., ingresó a través del puerto de Arica una carga de aproximadamente 20.000 toneladas de barros con contenidos metálicos, que eran nada más que desechos de la industria minera de Suecia, entre esos desechos se encontraban metales como hierro (16.5%), Zinc (1.3%), plomo (4.5%), Mercurio (0.003%), Arsénico (17.5%), entre otros. Depositando dicho material en el denominado Sitio F. Otro caso de contaminación ocurrido es el de la Empresa de Ferrocarriles del Estado (EFE), en el sector “La Pera” donde se lavaban los vagones de tren que transportaban “concentrados de minerales bolivianos” que no eran nada más que concentraciones de polimetales, estos resultaron depositados en los suelos de la ciudad. Actualmente el puerto y la ciudad de Arica tiene una amplia actividad industrial y comercial, generando gran cantidad de emisiones nocivas, pues el transporte de minerales es la principal actividad del puerto de Arica, el cual realiza la transferencia de concentrado mineral de numerosas

actividades mineras, tanto nacionales como internacionales (aproximadamente el 80% de la exportación de Bolivia pasa por el puerto de Arica). Dicha contaminación antropogénica se ve potenciada por las características geológicas de la ciudad y los vientos costeros que esta posee.

Las concentraciones de metales trazas en la ciudad de Arica corresponden a datos geológicos los cuales están en unidades de mg/kg, esto los convierte en datos de composición (cerrados), que no se representan en el espacio euclidiano, sino más bien en el espacio simplex de Aitchison (Aitchison, 1986; Reimann et al., 2012; Pawlowsky-Glahn y Buccianti, 2011; Filzmoser y Hron, 2011). En tal caso, un enfoque estadístico simple puede introducir explicaciones erróneas. Por ejemplo, algunas diferencias en la concentración elemental (por ejemplo, la concentración de Zn) dependen de la concentración de otros elementos. Las conclusiones directas podrían subestimar (o sobrestimar) estas diferencias (Filzmoser et al., 2009). Para resolver estos problemas, los datos brutos deben normalizarse por medio de la transformación *log-ratio* (Aitchison, 1982, 1999, 2003).

Actualmente existen diversos grupos de investigación en datos composicionales, los más destacados son liderados por Vera Pawlowsky-Glahn en la Universidad de Girona, España y Peter Filzmoser Vienna University of Technology, Austria. Ambos académicos han realizado diversos estudios sobre datos composicionales, aplicados a datos geoquímicos de los suelos de Europa. A su vez han aportado en el campo del análisis de datos composicionales, desarrollando paquetes para R software que facilitan el trabajo estadístico de estos como lo son “*compositions*” y “*robCompositions*”.

El estudio tiene como objetivo: 1) Elaborar un análisis exploratorio de datos geoquímicos mediante CoDA en suelos urbanos de Arica, 2) Determinar los metales pesados de origen natural, 3) Determinar los metales pesados de origen antrópico, 4) Establecer las diferencias entre los paquetes de R “*compositions*” y “*robCompositions*” y 5) Establecer las diferencias entre la estadística clásica y estadística composicional. Para el desarrollo de esto se utilizara estadística multivariante y análisis multivariante como el Análisis Clúster, Análisis de Componentes Principales. Y para la comparativa entre estadística clásica y composicional se utilizaron los datos y resultados obtenidos del artículo “An assessment of the potentially hazardous element contamination in urban soils of Arica, Chile” (Tume, Roca, Rubio, King y Bech, 2016).

## **2. Materiales y métodos.**

### **2.1. Área de Estudio**

El área de estudio en la que se centra el proyecto es la ciudad de Arica Fig. I ubicada en Chile 2000km al norte de Santiago, en la XV Región de Arica y Parinacota, comuna de Arica. Limita al Norte con el Perú, al este con la Provincia de Parinacota, al Sur con la región de Tarapacá y al Oeste con el Océano Pacífico. El espacio urbano de la ciudad posee 185.268 habitantes y un área de 41.89Km<sup>2</sup>, debido a que Arica es la única gran ciudad de la Región, en esta se alcanza la mayor densidad poblacional.

Arica cuenta con un clima desértico cálido, con nublados abundantes, tiene una temperatura media anual de 19.2 C con una máxima de 22°C y mínima de 18°C y es considerada uno de los lugares más secos habitados en la tierra con una precipitación total anual de 1.5mm aprox. Los vientos

costeros vienen de la dirección Sur Oeste y tienen una magnitud frecuente de 13 a 14 nudos, mientras que la nubosidad es abundante y posee muchas nieblas (camanchacas), esta medida en octavos (6,0 a las 8:00hr, 2,5 a las 14:00hr, 4,6 a las 20:00hrs).

En cuanto a la biogeografía en el puerto de la ciudad, se presentan alrededor del borde costero lobos marinos, gran diversidad de aves como gaviotines monja, el yeco y el lite. En cuanto a la flora del borde costero destacan especies de hierba como la Colcha, pasto salado, colas de zorro y otros.

Arica se extiende sobre una extensa planicie costera o litoral, debido a la inexistencia de la cordillera de la costa. La ciudad es atravesada por el río San José. La principal roca de la zona es la roca sedimentada la cual se forma por acumulación de sedimentos los cuales están constituidos en su mayoría por grava, arena o piedra clástica (del tamaño de arcilla), material semi-consolidado.

Dentro de ella se encuentra el puerto de Arica, el ferrocarril Tacna- Arica y el Ferrocarril Arica-La Paz los cuales en los últimos tiempos han facilitado el transporte de metales pesados desde el área de la minería hasta Arica, ya que varias mineras transportan sus metales sin procesar al puerto de esta ciudad.



**Figura I** Ciudad de Arica, con su distribución de puntos de muestreo geo referenciados (CONAMA 2009)

### 2.1.2 Uso de suelo

El uso de los suelos en la ciudad de Arica tiene distintos propósitos, al momento de caracterizar las muestras y tomarlas, se le asignaron a esta ciudad seis tipos distintos de usos.

**Comercial:** relacionado a toda la actividad que involucra el comercio directo (ferias, restaurants, centros comerciales, etcétera)

**Industrial:** relacionado a los suelos usados para actividades industriales, incluye el puerto de Arica

Peri-Urbano: espacio ubicado al norte de Arica, el cual no tiene actividad humana.

Residencial: Lugares relacionados con parques, caminos, residencias y lugares de la periferia.

Sitio F: lugar donde se arrojaron desechos con alto contenido de contaminantes.

Vía férrea: lugares directos y aledaños por donde circula la vía férrea.

## 2.2. Muestreo y adquisición de datos

Las muestras del suelo y el análisis químico fueron hechas por AGQ América S.A como un proyecto de la Comisión Nacional Del Medio Ambiente Arica y Parinacota (CONAMA, 2009).

Se tomaron un total de 400 muestras, más un 10 % de duplicadas para aseguramiento de calidad y evaluación de la precisión de los resultados, obteniendo un total de 440 muestras para analizar. Las muestras fueron tomadas de sitios diferentes a lo largo de la ciudad de Arica Fig. I. En cada punto se tomaron además 6 sub-muestras. En cada punto se definió un punto céntrico y cinco equidistantes entre 0,8 y 1,0 metros a intervalos de 36°. Las muestras tomadas fueron las siguientes; 2 submuestras superficiales, 2 submuestras a 15 cm. y 2 submuestras a 30cm. Con cada subconjunto de submuestras se generó una muestra compuesta de a lo menos 500grs. En resumen, se tomó un total de 2.640 submuestras en 440 puntos.

Luego las concentraciones de As, Ba, Cr, Ni, Pb y Zn fueron determinadas por USEPA SW 846 Método 3051. Extractos fueron analizados por Espectrometría de emisión óptica de plasma acoplada inductivamente (ICP-OES) de acuerdo con el USEPA SW 846 METODO 6010B (USEPA, 1996).

## 2.3. Datos composicionales

En diversas ramas se pueden encontrar datos composicionales como en Geología, Química, Economía, etc. Estos se ven frecuentemente en estudios ambientales de uso y distribución de suelos, distribución de contaminantes en agua, aire y suelo, etc. Las primeras referencias donde se identifican problemas al analizar estos datos e interpretarlos son descritas por Karl Pearson en 1887, luego un gran grupo de estadístico ha propuesto estrategias para solucionar los conflictos que generan el trabajo con datos composicionales, pero no fue hasta el año 1986 que Aitchison, J. en su monografía "*The statistical analysis of compositional data*" sienta las bases modernas para el análisis de datos composicionales.

Los datos composicionales o cerrados se caracterizan por ser vectores aleatorios cuyos componentes son las proporciones, porcentajes o partes por millón. La suma de las partes  $k$  debe ser constante e igual a: 1 si se trabaja con proporciones, 100 si se trabaja con porcentajes y  $10^6$  si se trabaja en partes por millón (ppm). Esta suma tiene un cierto grado de incertidumbre la cual se debe, en datos geoquímicos, a errores en las mediciones o no considerar elementos trazas. Los datos composicionales se muestran en términos estadísticos como vectores aleatorios de una composición cuyo espacio muestral se llama simplex.

El espacio de muestra Simplex; es un espacio cerrado, el cual es muy diferente al espacio Euclidiano real asociado a datos sin restricciones. El simplex de n partes es el conjunto de vectores reales positivos cuya suma es constante, formalmente es definido como:

$$S^D = \{x = [x_1, x_2, x_3, \dots, x_D] | x_i > 0, i = 1, 2, \dots, D; \sum_{i=1}^D x_i = k\}$$

Donde la condición  $x_i > 0$ , se puede sustituir por  $x_i \geq 0$ . Debido a esta definición, los datos composicionales están fijos a la restricción de que su suma debe ser igual a una constante k y solo pueden variar de  $0 < a \leq k$ , lo cual es notoriamente diferente a los datos normales que tienen intervalos de  $-\infty$  a  $+\infty$  en un espacio real. La geometría euclídea no es una geometría apropiada para los fenómenos composicionales porque: a) los resultados pueden no estar en el simplex y b) las diferencias euclídeas no siempre son razonables, por ejemplo: entre el 5 y 10% hay un incremento relativo del 100% mientras que entre 50% y 55% un incremento relativo del 10%, pero la distancia euclídea es la misma (Campos, 2011).

La restricción de no negatividad y la restricción de la suma constante que poseen estos datos impide la aplicación de procedimientos estadísticos descriptivos clásicos. La omisión de esta restricción al momento de trabajar con estos datos puede llevar al menos a tres graves inconvenientes que provocan un análisis erróneo de los datos y resultados irrelevantes como sesgos en las correlaciones y problemas al aplicar operaciones clásicas del espacio real vectorial a los datos composicionales. Los dos principales inconvenientes son:

*Correlación espuria:* Este término, atribuido a **Karl Pearson**, se refiere a la situación de correlación entre las razones de variables, aunque las variables originales no estén correlacionadas o sean aleatorias. **Chayes (1948, 1960)** relaciona este problema con la suma constante. La restricción de la suma constante determina que la matriz de correlaciones presente correlación negativas no nulas. Lo cual es trascendental pues un gran número de pruebas estadísticas multivalentes parten del cálculo de las matrices de correlación o de varianza y covarianza.

#### Ejemplo1

Se presentan los espesores de estratos presentes en dos secuencias A y B expresadas en metros y como porcentajes de la suma en ambas secuencias (Rollinson, 1993)

**Tabla 1**

*Ejemplo de correlaciones espurias.*

|   | Espesor (m) |     | Espesor (%) |      |
|---|-------------|-----|-------------|------|
|   | A           | B   | A           | B    |
| 1 | 50          | 50  | 50.0        | 50.0 |
| 2 | 60          | 85  | 41.4        | 58.6 |
| 3 | 70          | 110 | 38.9        | 61.1 |
| 4 | 75          | 140 | 34.9        | 65.1 |
| 5 | 80          | 170 | 32.0        | 68.0 |
| 6 | 90          | 200 | 31.0        | 69.0 |
| r | 0.99        |     | -1.00       |      |

En este ejemplo expuesto por Rollinson se aprecia claramente que los espesores medidos en metros están correlacionados positivamente ( $r= 0.99$ ), pero al momento de transformar las medidas de metros a porcentajes, se obtiene una correlación de ( $r=-1.0$ ), que resulta ser lo opuesto a la anterior mencionada.

*Incoherencias en las subcomposiciones:* Una subcomposición es un subconjunto cualquiera de las partes de una composición en la que se reconstruye la condición de la suma constante. Normalmente se espera que la correlación entre partes sea igual en la subcomposición y en la composición original. Sin embargo las correlaciones varían desde la composición original a subcomposiciones cada vez más pequeñas.

### Ejemplo 2

Dos investigadores A y B estudian la composición de unas muestras de suelo. A esta interesado en cuatro partes que clasifica como animal, vegetal, mineral y agua ( $x1, , x4$ ). B seca las muestras y las analiza la composición animal, vegetal y mineral ( $s1, s2, s3$ ). De modo que los datos de B son una subcomposición de los datos de A (Aitchison, 1997)

**Tabla 1**

*Ejemplo incoherencias en las subcomposiciones*

| A                    | B                    |
|----------------------|----------------------|
| ( $x1,x2,x3,x4$ )    | ( $s1,s2,s3$ )       |
| (0.1; 0.2; 0.1; 0.6) | (0.25; 0.50; 0.25)   |
| (0.2; 0.1; 0.1; 0.6) | (0.50; 0.25; 0.25)   |
| (0.3; 0.3; 0.2; 0.2) | (0.375; 0.375; 0.25) |

La correlación entra las partes animal-vegetal registrada por el investigados A es  $\text{corr}(x1, x2)= 0.5$  mientras que el investigador B obtiene  $\text{corr}(s1,s2)= -1$  (Alperin, 2013).

Los Datos Composicionales son esencialmente multivariantes y por ende se deben analizan con métodos multivariantes como los son Análisis de Componentes Principales y Análisis Clúster.

#### 2.3.1. Operaciones básicas en el Simplex

En 1982 fue Aitchison J. quien propone una solución que evita la restricción de la suma constante. Esta se evita al poner la atención en la magnitud relativa de las partes, es decir, en los cocientes.

Al trabajar con cocientes entre partes desaparecen las correlaciones espurias y la magnitud relativa de las partes de una subcomposición no cambian en relación a la magnitud relativa entre las partes de la composición original. Este principio es denominado *invariancia por cambios de escala*. Las operaciones que propone Aitchison para sortear esta compleja geometría son, la operación de clausura (C) la cual permite expresar un conjunto de datos como una composición, partiendo de la base que no existen datos ausentes o negativos, la operación de perturbación la cual es equivalente a un cambio de escala, la operación de potenciación útil para describir relaciones de regresión para composiciones, es análoga a la multiplicación en los reales y por ultimo las subcomposiciones y amalgamas las cuales son de utilidad para estudiar solo partes de los componentes de un sistema.

### 2.3.2. Transformaciones *alr*, *clr* y *ilr*

Aitchison (1986) propone una serie de transformaciones con el fin de llevar los datos desde el Simplex  $S^D$  al espacio de números reales  $R^+$ . Dichas transformaciones permiten describir adecuadamente los datos, además de utilizar métodos estadísticos uní y multivalentes para analizar y modelar los datos composicionales.

#### I) Transformación Log-cociente Aditiva *alr*.

Esta transformación posee el problema de falta de simetría, debido al uso de una de sus partes como denominador común y las coordenadas obtenidas por esta transformación D-1 no pueden mapearse en ejes ortogonales. Los datos *alr*-transformados sólo debe usarse cuando se requiere realizar pruebas de hipótesis (multinormalidad, contrastar hipótesis sobre el centro y estructura de covarianza) y regiones de confianza para el centro de una composición.

#### II) Transformación Log-cociente Centrada *clr*.

La cual es definida como:

$$v = clr(x) = \ln\left(\frac{x_1}{g_m(x)}\right), \ln\left(\frac{x_2}{g_m(x)}\right), \dots, \ln\left(\frac{x_D}{g_m(x)}\right)$$

En donde  $g_m(x)$  representa a la media geométrica. Esta transformación es simétrica y es utilizada para generar *biplots* composicionales. Sin embargo tiene problemas de singularidad. Esta transformación se utiliza en pruebas que se basan en distancias tales como el Análisis Clúster y para resolver problemas acerca del grado de correlación entre componentes y entre individuos (Análisis de Componentes Principales).

#### II) Transformación isométrica *ilr*.

Esta transformación traslada la geometría del simplex a un espacio multivariable real. No sufre el problema de singularidad de *clr* pero su inconveniente es que la interpretación de los resultados es sumamente difícil.

### 2.4. Análisis de datos

El tratamiento estadísticos de los datos se realizó mediante el software R, utilizando los paquetes “*compositions*” y “*robcompositions*”, los cuales permiten trabajar en el espacio Simplex de Aitchinson, J. dichos softwares fueron creados por K. Gerald van den Boogaart, Greifswald y Raimon Tolosana, Girona; Matthias Templ, Karel Hron, Peter Filzmoser respectivamente. Estos paquetes se encuentran disponibles en sus sitios web de forma gratuita y en el software R. Estos paquetes proporcionan herramientas para el análisis de Datos composicionales en cuatro geometrías diferentes. Una de las grandes cualidades del paquete “*compositions*” es que permite comparar los resultados obtenidos en sus cuatro enfoques distintos de manera simple, ya que permite realizar análisis similares a cada geometría, fácilmente. “*compositions*” también incluye las características más básicas, como lo son transformaciones de datos (Transformación Log-

cociente Aditiva *alr*, Transformación Log-cociente Centrada *clr* y Transformación isométrica *ilr*), estadísticas básicas, gráficos de alto nivel (como diagramas ternarios y diagramas de dispersión) y análisis de alto nivel (como Análisis de Componentes Principales ACP o Análisis de Conglomerados). Mientras que una de las cualidades del “*robcompositions*” es su manejo de *outliers*. Los resultados de todos estos análisis también son mostrados de forma consistente en las cuatro geometrías, para así facilitar su comparación. Con este programa se generaron diversas tablas y gráficos. Entre ellas la Tabla 3 y 4 que presentan un resumen analítico, con medidas más válidas para los datos composicionales, pues la estadística descriptiva clásica no se ajusta a dichos datos por lo cual medidas como la media y desviación estándar pueden presentar errores a la interpretación. La tabla 5 la cual representa la matriz de variación y de medias geométricas, además de diagramas de dispersión y boxplot para cada elemento.

**Tabla 3**

*Estadísticas resumidas, elementos principales.*

| <b>Estadística Composicional</b> |            |             |           |            |            |              |            |            |            |            |              |                |                |
|----------------------------------|------------|-------------|-----------|------------|------------|--------------|------------|------------|------------|------------|--------------|----------------|----------------|
| <b>E</b>                         | <b>Min</b> | <b>Q2</b>   | <b>Q5</b> | <b>Q10</b> | <b>Q25</b> | <b>Q50</b>   | <b>Q75</b> | <b>Q90</b> | <b>Q95</b> | <b>Q98</b> | <b>Max</b>   | <b>MAD.log</b> | <b>Mad.ilr</b> |
| <b>As</b>                        | 0.01       | <b>0.36</b> | 0.69      | 1.09       | 1.80       | <b>2.90</b>  | 4.26       | 5.85       | 8.20       | 12.41      | <b>20.90</b> | 0.637          | 0.557          |
| <b>Ba</b>                        | 0.01       | <b>0.03</b> | 0.23      | 0.56       | 1.40       | <b>3.10</b>  | 5.14       | 7.28       | 8.33       | 9.92       | <b>17.00</b> | 0.949          | 0.647          |
| <b>Cr</b>                        | 0.01       | <b>0.04</b> | 0.10      | 0.26       | 0.51       | <b>1.02</b>  | 2.87       | 4.74       | 5.79       | 7.80       | <b>16.51</b> | 1.284          | 0.817          |
| <b>Cu</b>                        | 0.65       | <b>0.89</b> | 1.13      | 1.37       | 2.27       | <b>4.30</b>  | 7.55       | 13.77      | 20.16      | 33.95      | <b>75.07</b> | 0.857          | 0.806          |
| <b>Ni</b>                        | 0.02       | <b>0.06</b> | 0.11      | 0.30       | 0.86       | <b>1.22</b>  | 1.64       | 1.99       | 2.54       | 3.72       | <b>13.99</b> | 0.470          | 0.922          |
| <b>Pb</b>                        | 6.40       | <b>13.7</b> | 16.78     | 19.03      | 22.72      | <b>27.60</b> | 42.97      | 57.39      | 64.60      | 68.23      | <b>79.82</b> | 0.407          | 0.580          |
| <b>V</b>                         | 0.06       | <b>0.61</b> | 1.94      | 3.48       | 8.14       | <b>13.43</b> | 19.64      | 23.76      | 28.04      | 30.43      | <b>37.17</b> | 0.600          | 0.548          |
| <b>Zn</b>                        | 0.37       | <b>2.98</b> | 14.17     | 18.59      | 27.13      | <b>34.76</b> | 43.75      | 55.28      | 61.86      | 69.74      | <b>83.81</b> | 0.352          | -              |

*Todos los resultados analíticos en porcentaje de peso (Wt % en peso). E: Elemento, Min: mínimo; Q: cuantiles (Q50 = mediana); Max: máximo; MAD.log: desviación absoluta media log transformada; MAD.ilr: desviación absoluta media ilr transformada.*

Con el fin de determinar las fuentes de oligoelementos que preceden en la Ciudad de Arica, se utilizaron dos métodos multivariantes, Análisis de componentes principales (ACP) y Análisis Clúster (CA). El CA fue utilizado agrupando las variables de elementos, para este caso se debe usar una “distancia” o medida de asociación entre variables, de modo que la asociación perfecta se represente como cero y la independencia perfecta como  $+\infty$ . Para los datos de composición, ya se ha introducido tal medida: la matriz de variación. Solo falta la conversión interna de la matriz de variación en una matriz de distancia apropiada (Van den Boogaart y Tolosana-Delgado, 2011). La extracción del ACP se llevó a cabo utilizando la matriz de varianza-covarianza (Filzmoser, 2016), pues la matriz de varianza - covarianza de los datos *clr*-transformados es útil para descifrar las relaciones entre las partes. La matriz es simétrica con las varianzas en la diagonal y las covarianzas en las demás celdas. El análisis de esta matriz se suele hacer gráficamente a partir de un biplot que resulta en un ACP efectuado con los datos *clr*-transformados (Aitchison y Greenacre, 2002).

**Tabla 4***Estadísticas resumidas, elementos principales transformados por clr.*

| MT        | Min   | Q2           | Q5    | Q10   | Q25   | Q50          | Q75   | Q90   | Q95   | Q98   | Max         | MAD  | IQR         |
|-----------|-------|--------------|-------|-------|-------|--------------|-------|-------|-------|-------|-------------|------|-------------|
| <b>As</b> | -4.82 | <b>-2.03</b> | -1.51 | -1.23 | -0.95 | <b>-0.69</b> | -0.24 | 0.21  | 0.49  | 0.92  | <b>3.04</b> | 0.48 | <b>0.71</b> |
| <b>Ba</b> | -4.45 | <b>-3.16</b> | -2.31 | -1.73 | -1.07 | <b>-0.54</b> | -0.24 | 0.05  | 0.21  | 0.60  | <b>0.98</b> | 0.54 | <b>0.83</b> |
| <b>Cr</b> | -4.00 | <b>-3.54</b> | -3.08 | -2.50 | -2.10 | <b>-1.42</b> | -0.88 | -0.47 | -0.34 | -0.09 | <b>0.56</b> | 0.91 | <b>1.22</b> |
| <b>Cu</b> | -1.91 | <b>-1.55</b> | -1.19 | -1.02 | -0.70 | <b>-0.22</b> | 0.31  | 1.01  | 1.43  | 2.37  | <b>5.27</b> | 0.75 | <b>1.02</b> |
| <b>Ni</b> | -4.05 | <b>-3.28</b> | -2.45 | -2.11 | -1.88 | <b>-1.56</b> | -1.15 | -0.98 | -0.88 | -0.67 | <b>0.62</b> | 0.54 | <b>0.73</b> |
| <b>Pb</b> | 0.37  | <b>0.81</b>  | 0.95  | 1.05  | 1.21  | <b>1.64</b>  | 2.35  | 2.95  | 3.35  | 3.97  | <b>4.85</b> | 0.77 | <b>1.15</b> |
| <b>V</b>  | -2.11 | <b>-0.84</b> | -0.37 | -0.02 | 0.43  | <b>0.89</b>  | 1.21  | 1.63  | 1.82  | 2.13  | <b>2.84</b> | 0.55 | <b>0.77</b> |
| <b>Zn</b> | -1.79 | <b>0.49</b>  | 1.09  | 1.29  | 1.50  | <b>1.82</b>  | 2.21  | 2.78  | 3.06  | 3.36  | <b>4.21</b> | 0.51 | <b>0.71</b> |

*Todos los valores son relaciones, es decir, adimensionales. MT: Elemento, Min: mínimo; Q: quantiles (Q50 = mediana); Max: máximo; MAD: desviación absoluta media, IQR: intercuartil distancia*

**Tabla 5***Matriz de variación.*

|           |           | Variación ln(Xi/Xj) |           |           |           |           |          |           |
|-----------|-----------|---------------------|-----------|-----------|-----------|-----------|----------|-----------|
|           | <b>As</b> | <b>Ba</b>           | <b>Cr</b> | <b>Cu</b> | <b>Ni</b> | <b>Pb</b> | <b>V</b> | <b>Zn</b> |
| <b>As</b> |           | 1.497               | 1.795     | 1.373     | 1.047     | 0.887     | 1.072    | 1.192     |
| <b>Ba</b> | -0.147    |                     | 0.802     | 1.848     | 1.185     | 2.112     | 0.954    | 1.907     |
| <b>Cr</b> | -0.941    | -0.794              |           | 1.718     | 0.876     | 2.382     | 1.263    | 1.549     |
| <b>Cu</b> | 0.494     | 0.640               | 1.434     |           | 1.621     | 1.246     | 2.108    | 1.089     |
| <b>Ni</b> | -0.985    | -0.838              | -0.044    | -1.479    |           | 1.182     | 0.709    | 0.821     |
| <b>Pb</b> | 2.442     | 2.589               | 3.383     | 1.948     | 3.427     |           | 1.023    | 1.001     |
| <b>V</b>  | 1.400     | 1.547               | 2.341     | 0.907     | 2.385     | -1.042    |          | 1.435     |
| <b>Zn</b> | 2.467     | 2.614               | 3.408     | 1.974     | 3.452     | 0.025     | 1.067    |           |
|           |           | Media ln(Xi/Xj)     |           |           |           |           |          |           |

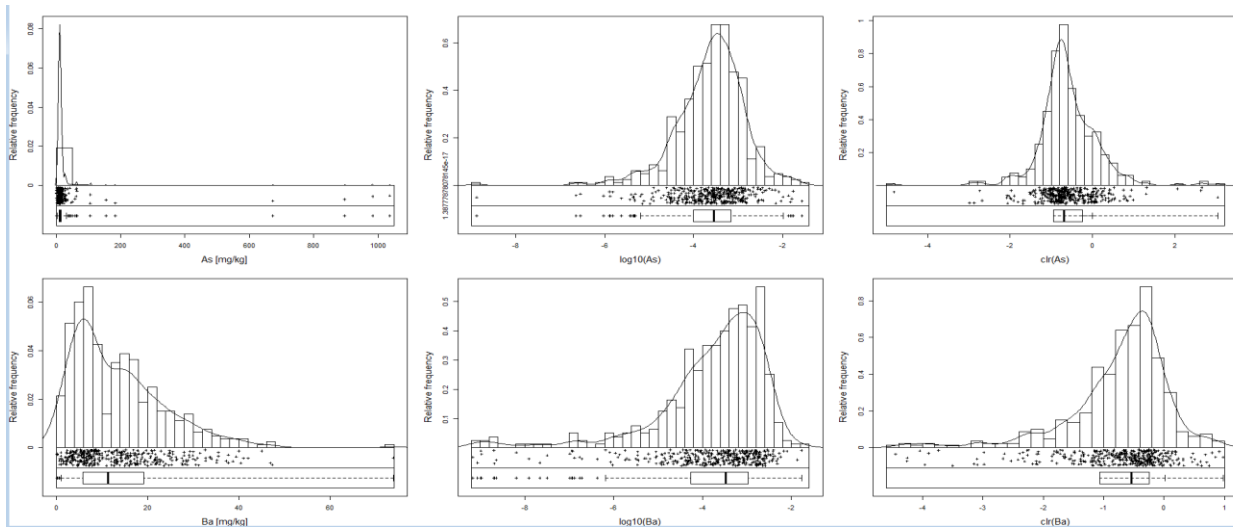
*Mitad superior derecha las varianzas de los log-cocientes entre partes y en la mitad inferior izquierda las medias aritméticas de los log-cocientes entre partes.*

### 3. Resultado y discusión.

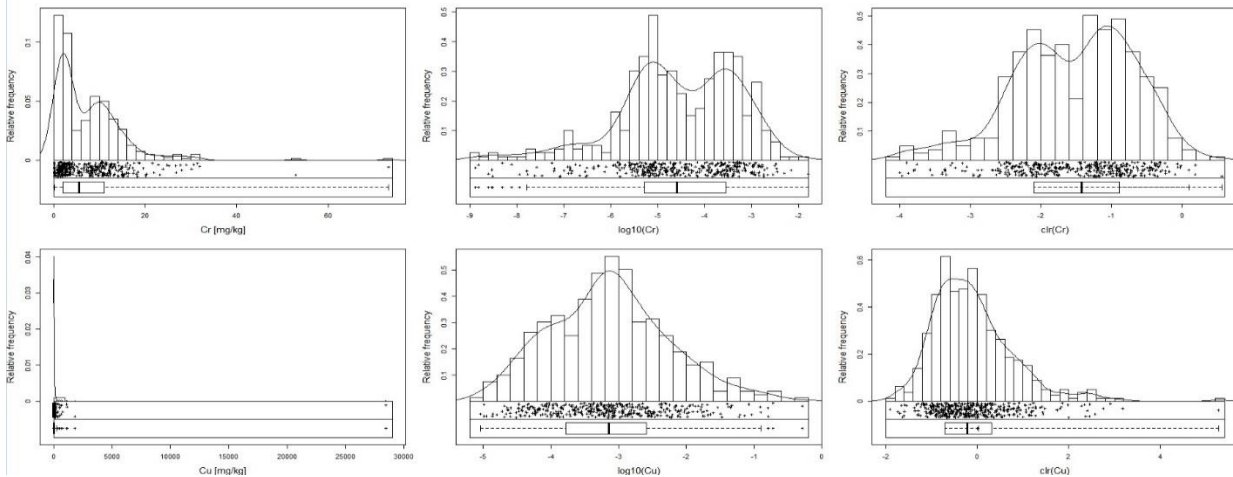
#### 3.1. Análisis exploratorio de datos

Para visualizar mejor las distribuciones de los elementos y sus principales fuentes en el área de muestra, se generaron *edaplots* (Gráfico de análisis exploratorio de datos), este se considera una de las mejores herramientas para mostrar la distribución de los datos (Reimann et al., 2008). El histograma puede mostrar distribuciones unimodal (simétrica o asimétrica) o multimodal, el diagrama de dispersión unidimensional muestra los datos a lo largo de una línea recta también se puede apreciar si los datos son homogéneos o heterogéneos (Thiombane et al., 2017), por último el boxplot permite la visualización de datos atípicos.

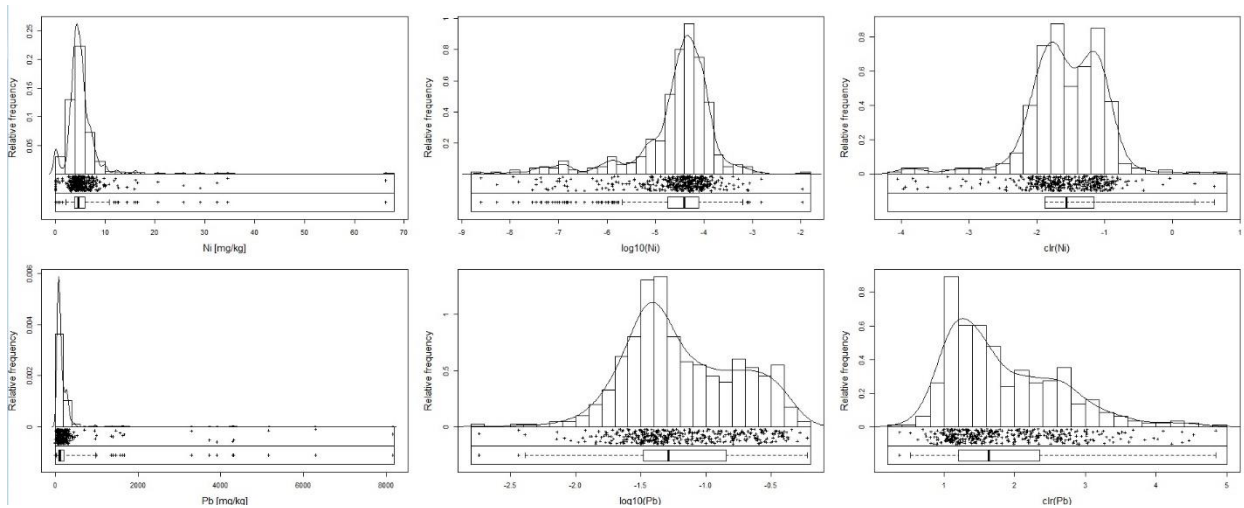
Las Fig. II, III, IV y V muestran las distribuciones de As, Ba, Cr, Cu, Pb, Ni, V y Zn de la muestra a través de un *edaplots* (Combinación de histograma, diagrama de dispersión y diagrama de caja, en un solo conjunto) estas figuras muestran tres representaciones diferentes de los datos: las concentraciones originales (Gráficos de la izquierda), las concentraciones con transformación log normal (Gráficos del centro) y las concentraciones con transformación *clr* (Gráficos de la derecha). En general los datos en bruto (Gráficos de la izquierda) muestran a los elementos con distribución asimétrica y unimodal a excepción del Cr que presenta una distribución multimodal, además los datos muestran un apuntalamiento a la izquierda. Los gráficos con las concentraciones log-transformadas y *clr*-transformadas muestran una mayor simetría de los datos y más centrados, salvo Cr, Ni y Pb todos muestran una distribución unimodal. De los diagramas de caja y dispersión se puede apreciar la presencia de varios datos atípicos.



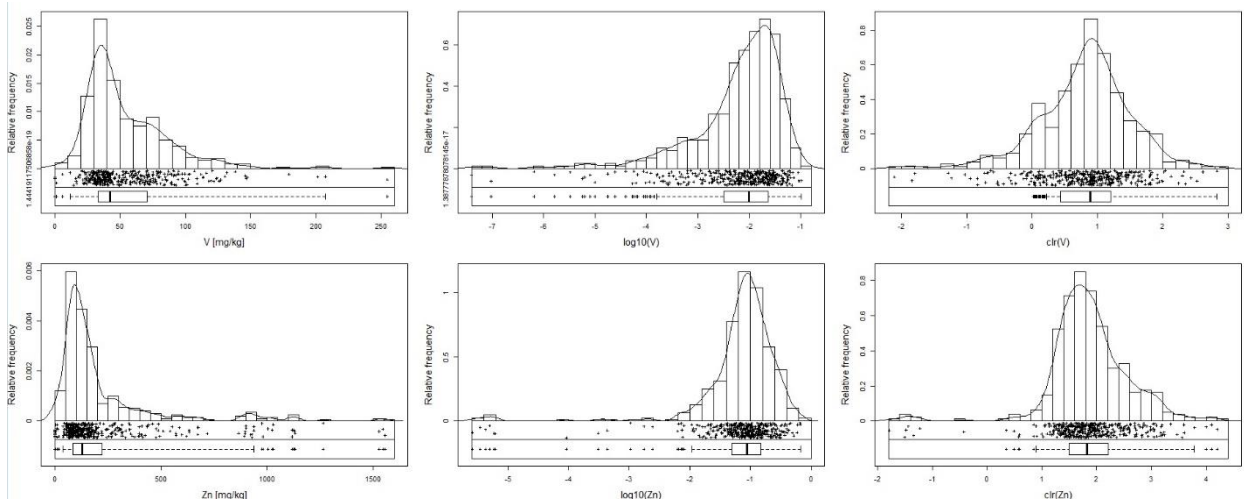
**Figura II** *Edaplots* (Combinación de histograma, diagrama de dispersión y diagrama de caja, en un solo conjunto) de los datos brutos, datos log-transformados y datos *clr*-transformados. Superior As, inferior Ba.



**Figura III** *Edaplots* (Combinación de histograma, diagrama de dispersión y diagrama de caja, en un solo conjunto) de los datos brutos, datos log-transformados y datos *clr*-transformados. Superior Cr, inferior Cu.



**Figura IV** Edaplots (Combinación de histograma, diagrama de dispersión y diagrama de caja, en un solo conjunto) de los datos brutos, datos log-transformados y datos *clr*-transformados. Superior Ni, inferior Pb.



**Figura V** Edaplots (Combinación de histograma, diagrama de dispersión y diagrama de caja, en un solo conjunto) de los datos brutos, datos log-transformados y datos *clr*-transformados. Superior V, inferior Zn.

### 3.2. Análisis de datos composicionales (CoDA).

El resumen de los resultados analíticos es presentado en la Tabla 3. En el afán de proporcionar medidas válidas para los datos composicionales, la tabla entrega los cuartiles de la distribución. Esta tabla es de interés si los valores absolutos son de interés en lugar de las relaciones relativas. Es importante señalar que la Tabla 3 no entrega valores de media y de desviación estándar, pues estos se basan en distancias Euclidianas. Los datos de composición no pertenecen al espacio euclidiano clásico, por lo que es necesario que sean trabajados en su espacio simplex incluso para el análisis univariante de datos. Por lo tanto, las pruebas estadísticas de comparación de dos

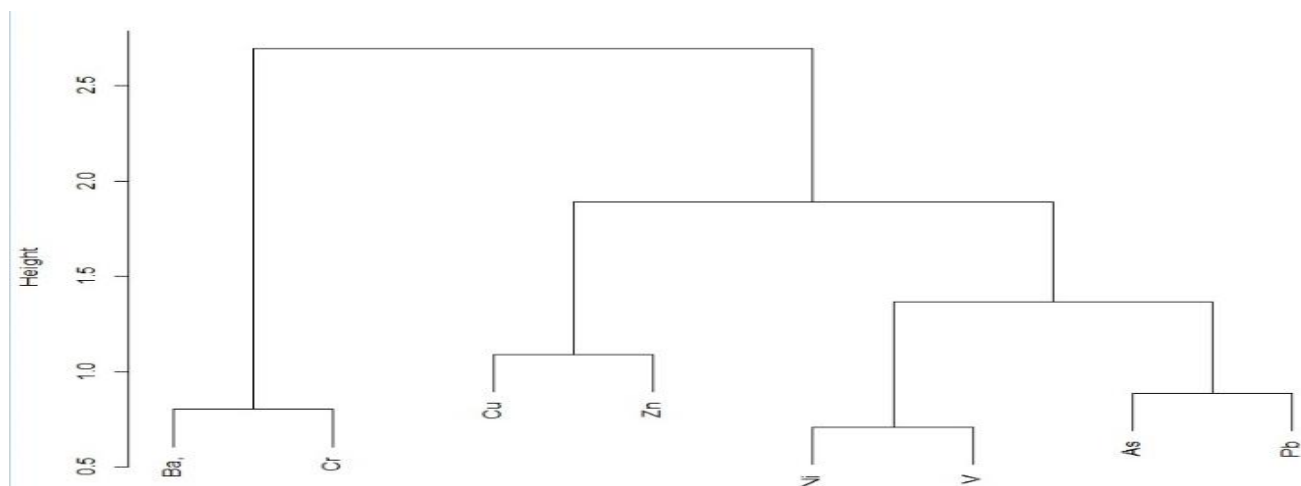
conjuntos de datos, basadas en distancias euclidianas pueden proporcionar errores (Aitchison, 1986; Filzmoser et al., 2009a, 2010).

En la Tabla 4, se aplica la transformación *clr* para realizar el análisis estadístico de los datos composicionales, la cual muestra un análisis CoDA equivalente al de la Tabla 3, con sus parámetros transformados por *clr*. Cada variable transformada por *clr* en la Tabla 4 contiene toda la información relativa del elemento estudiado. Estas variables son relaciones y números adimensionales sin ningún significado claro. Los elementos con altas concentraciones también se caracterizan por un alto valor *clr* (elemento); incluso la secuencia de elementos sigue siendo la misma, (para la interpretación se utilizó el log de base 10 en lugar del log natural de la transformación *clr*) (Reimman et al., 2012). Ambas Tablas 3 y 4 nos muestran que el elemento dominante es el Zn (mediana: 34.76 Wt% y *clr* (Zn) 1,822 respectivamente), pero analizando el rango intercuartil (IQR) y desviación absoluta de la mediana (MAD) presentes en la tabla 4, se desprende que el Zn posee una baja influencia sobre el resto de las variables, pues cuanto mayor sea la varianza mayor es la influencia de la variable en el conjunto de datos multivariado.

No apreciable a simple vista, pero utilizando el log, se puede desprender de la Tabla 4, importante información sobre cada elemento en particular. El valor mediano de *clr* (Pb)= 1.637 lo cual explica que en promedio la concentración de Pb es 43.3 veces mayor que la media geométrica de las demás concentraciones, el Zinc es 66.37 veces mayor y el Vanadio 7.7 veces. Por otro lado el valor mediano de *clr*(As)=-0.689 indica que este elemento tiene una abundancia de 0.2 de la media geométrica, el Ba de 0.28, Cr de 0.037, Cu de 0.6 y Ni de 0.02.

Debido a la naturaleza de la transformación *clr*, el análisis de la matriz de correlación con *clr*-variables no es significativo, sin embargo los valores medianos y las varianzas de las variables composicionales si pueden ser comparadas. La Matriz de variación es otra medida apropiada para describir la variabilidad de un conjunto de datos composicionales. La cual en su mitad superior derecha posee las varianzas de los log cocientes entre partes y en la mitad inferior izquierda las medias aritméticas de los log cocientes entre partes (Alperin, 2013). Como se aprecia en la Tabla 5 se ven las varianzas relativas más grandes entre Pb/Cr, Pb/Ba y V/Cu. Para Pb/Cr la varianza toma el valor 2.38 y una media positiva de 3.38 por lo que la proporción de Pb tiende en promedio a ser mayor que la del Cr, este análisis es equivalente para Pb/Ba y V/ Cu. Las relaciones de varianza relativa altas se atribuyen a relaciones inversamente proporcionales de los elementos.

Por el contrario las varianzas relativas más bajas, corresponden a relaciones directamente proporcionales entre los elementos. Continuando con el análisis de la Tabla 5 se aprecian estas relaciones principalmente en V/Ni, Cr/Ba, Zn/Ni, Ni/Cr y Pb/As. Tanto para V/Ni, Zn/Ni y Pb/As poseen medias positivas de 2.38, 3.34 y 2.44 respectivamente lo que sugieren que la proporción de Ni y As son menores a las de V, Zn y Pb respectivamente. Caso contrario Cr/Ba y Ni/Cr los cuales poseen medias negativas de -0.79 y -0.04 lo cual advierte que la proporción de Ba y Cr son mayores con respecto a las de Cr y Ni, respectivamente. Lo cual concuerda con el análisis previo de la Tabla 4.



**Figura VI:** Análisis Clúster (CA) para CoDA, método “Ward”.

A su vez las relaciones V/Ni, Cr/Ba, Zn/Ni, Ni/Cr y Pb/As se pueden apreciar en el Análisis Clúster (AC), el cual se realizó transformando la matriz de variación en una matriz de distancia y utilizando el método “Ward” como se aprecia en la Fig. VI. En esta se aprecian los grupos más relacionados entre sí, en el siguiente orden V/Ni, Cr/Ba, Pb/As y Cu/Zn. Del análisis de componentes principales ACP se identifican cuatro componentes principales (PC1 a PC4) para los metales trazas en estudio Tabla 6, estas cuatro componentes explican de forma acumulativa el 86.5% de la varianza total, lo cual fue el criterio para la selección de dichas 4 componentes. En este análisis, la tasa de contribución de la varianza para PC1 fue de 37.5% con Ba, Cr y Pb con las cargas más grandes, para PC2 la tasa de contribución de la varianza fue de 23.8%, y Cu y V tuvieron las cargas más grandes. La contribución de PC3 es de 15.7% y las cargas más altas las poseen Zn, Ni, Ba y As, por ultimo PC4 con 9.5% y As y Pb con cargas mayores.

**Tabla 6**

*Análisis de componentes principales CoDA, “compositions”.*

| Elemento Traza          | Componentes Principales |        |        |        |
|-------------------------|-------------------------|--------|--------|--------|
|                         | PC1                     | PC2    | PC3    | PC4    |
| As                      | 0.212                   | -0.230 | 0.355  | 0.795  |
| Ba                      | -0.497                  | 0.027  | 0.443  | -0.132 |
| Cr                      | -0.509                  | 0.320  | -0.200 | 0.074  |
| Cu                      | 0.320                   | 0.650  | 0.322  | -0.201 |
| Ni                      | -0.110                  | -0.138 | -0.468 | 0.139  |
| Pb                      | 0.478                   | -0.297 | 0.090  | -0.407 |
| V                       | -0.171                  | -0.526 | 0.011  | -0.338 |
| Zn                      | 0.277                   | 0.194  | -0.554 | 0.070  |
| % Varianza explicada    | 37.53                   | 23.81  | 15.69  | 9.49   |
| % Varianza e. acumulada | 37.53                   | 61.34  | 77.03  | 86.51  |

Con el respaldo de la matriz de covarianza Tabla 7, pues las covarianzas negativas enfatizan en la repulsión entre elementos mientras que las covarianzas positivas la asociación de estos (Alperin,

2013) y de la Tabla 8 la cual muestra un resumen básico de estadística analítica para cada elemento por zona, podemos hacer un análisis más a fondo de cada componente.

La componente 1 (PC1) que representa el 37.5% de la varianza, posee cargas elevadas en Ba (-0.497), Cr (-0.509) y Pb (0.478), tanto el Ba con el Cr presentan covarianzas bajas entre si lo que indica que estos elementos están altamente asociados, mientras que tienen cierta repulsión hacia el Pb. Esta componente podría tratarse de una fuente natural para los elementos Ba, Cr y Pb, añadiendo el Cu (0.320)

La componente 2 (PC2) con una tasa de contribución de la varianza de 23.8%, sus cargas más elevadas son Cu (0.650) y V (-0.526), la covarianza entre ambos elementos es negativa, lo cual indica la poca relación de estos elementos. Esta componente puede tratarse de elementos de procedencia mixta, añadiendo en dicha componente el Cr (0.320) el cual posee menor carga pero no deja de ser significativa.

La componente 3 (PC3) que representa el 15.7% de la varianza, ligada al As (0.355), Ba (0.443), Ni (-0.468) y Zn (-0.554), la única asociación existente entre estos elementos pertenece al Zn con el Ni. Los cuales son antrópicos en esta componente el Zn en la zona Vía férrea es donde posee el mayor IQR. El As, Ba y también Cu (0.322) pueden tener una fuente antrópica también en dicha componente.

La componente 4 (PC4) con una contribución del 9.5% con sus mayores cargas que pertenecen a As (0.795) y Pb (-0.407) ambos elementos bien relacionados entre sí con covarianzas positivas. Esta componente puede representar contaminación antrópica, en la Tabla 8 se puede apreciar que la mayor mediana-*clr* del arsénico la mayor en el Sitio F, mientras que el IQR del As y Pb son los mayores en la zona Industrial, en ambos casos muy por encima del sitio Peri-Urbano.

**Tabla 7**

*Matriz de covarianzas, CoDA.*

| MATRIZ COV |       |        |        |        |        |        |        |        |
|------------|-------|--------|--------|--------|--------|--------|--------|--------|
|            | As    | Ba     | Cr     | Cu     | Ni     | Pb     | V      | Zn     |
| As         | 0.520 | -0.140 | -0.284 | -0.034 | -0.094 | 0.137  | -0.036 | -0.069 |
| Ba         |       | 0.701  | 0.304  | -0.181 | -0.072 | -0.387 | 0.113  | -0.337 |
| Cr         |       |        | 0.711  | -0.112 | 0.087  | -0.518 | -0.036 | -0.153 |
| Cu         |       |        |        | 0.788  | -0.247 | 0.090  | -0.421 | 0.117  |
| Ni         |       |        |        |        | 0.342  | -0.100 | 0.057  | 0.028  |
| Pb         |       |        |        |        |        | 0.642  | 0.049  | 0.088  |
| V          |       |        |        |        |        |        | 0.483  | -0.210 |
| Zn         |       |        |        |        |        |        |        | 0.537  |

Por otro lado, de los grupos generados por el AC, se crearon subcomposiciones, para poder luego graficarlas en un diagrama cuaternario Fig. VII, en el cual se ve que las cantidades relativas de As+Pb y Cu+Zn están relacionadas y son por mucho mayores que las de Ba+Cr y Ni+V. En la Fig. VII, el vértice inferior izquierdo representa Ba+Cr, el inferior derecho As+Pb, el vértice central (imagen 3d) representa el Ni+V, por último el vértice superior Cu+Zn. Este diagrama se interpreta

de igual forma que un diagrama ternario y nos indicaría que tanto el As, Pb, Cu y Zn tienen posibles fuentes de contaminación antrópica y niveles altos de contaminación.

**Tabla 8**

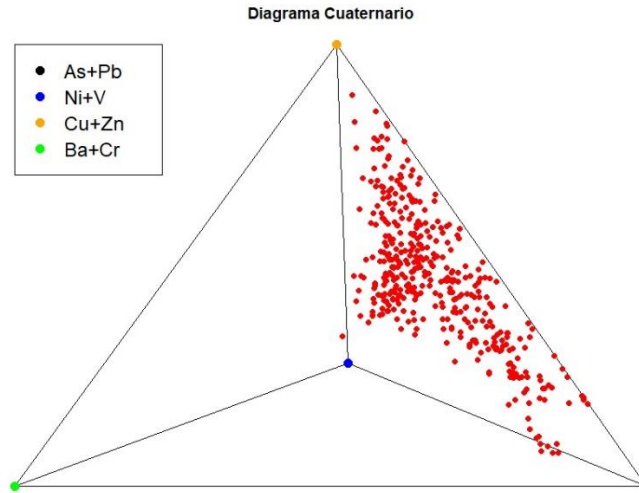
*Estadísticas resumidas, elementos principales transformados por clr.*

| Zona        |      | Metales Traza |       |       |       |       |      |       |       |
|-------------|------|---------------|-------|-------|-------|-------|------|-------|-------|
| Comercial   |      | As            | Ba    | Cr    | Cu    | Ni    | Pb   | V     | Zn    |
| <i>clr</i>  | Min. | -5.11         | -2.30 | -2.63 | -1.36 | -2.83 | 1.19 | -0.59 | 1.33  |
|             | Q50  | -0.66         | -0.51 | -2.05 | -0.37 | -1.37 | 1.91 | 0.83  | 2.15  |
|             | Max. | 0.07          | 0.47  | -0.71 | 1.16  | -0.78 | 3.06 | 1.67  | 4.21  |
|             | IQR  | 0.61          | 0.80  | 0.85  | 1.04  | 0.69  | 0.76 | 0.50  | 0.76  |
| Industrial  |      |               |       |       |       |       |      |       |       |
| <i>clr</i>  | Min. | -2.74         | -4.64 | -4.25 | -0.67 | -2.64 | 1.00 | -2.05 | 1.24  |
|             | Q50  | -0.72         | -0.55 | -1.37 | 0.81  | -1.99 | 1.73 | -0.08 | 2.49  |
|             | Max. | 3.09          | 0.80  | 0.10  | 5.27  | -0.65 | 4.92 | 1.34  | 3.09  |
|             | IQR  | 0.96          | 3.14  | 1.32  | 1.89  | 0.41  | 2.06 | 0.72  | 0.83  |
| Peri-Urbano |      |               |       |       |       |       |      |       |       |
| <i>clr</i>  | Min. | -3.00         | -3.04 | -4.03 | -1.60 | -4.05 | 0.37 | -1.80 | -1.79 |
|             | Q50  | -0.69         | -0.56 | -1.65 | -0.35 | -1.47 | 1.68 | 0.95  | 1.78  |
|             | Max. | 1.32          | 0.97  | 0.56  | 3.18  | 0.62  | 4.34 | 2.80  | 4.01  |
|             | IQR  | 0.70          | 0.83  | 1.22  | 0.93  | 0.66  | 1.26 | 0.67  | 0.63  |
| Residencial |      |               |       |       |       |       |      |       |       |
| <i>clr</i>  | Min. | -3.00         | -3.04 | -4.03 | -1.60 | -4.05 | 0.37 | -1.80 | -1.79 |
|             | Q50  | -0.69         | -0.56 | -1.65 | -0.35 | -1.47 | 1.68 | 0.95  | 1.78  |
|             | Max. | 1.32          | 0.97  | 0.56  | 3.18  | 0.62  | 4.34 | 2.80  | 4.01  |
|             | IQR  | 0.70          | 0.83  | 1.22  | 0.93  | 0.66  | 1.26 | 0.67  | 0.63  |
| Sitio F     |      |               |       |       |       |       |      |       |       |
| <i>clr</i>  | Min. | -1.04         | -2.32 | -2.47 | -0.88 | -2.12 | 0.70 | -0.94 | 1.00  |
|             | Q50  | -0.11         | -0.35 | -1.17 | 0.33  | -1.87 | 1.31 | 0.33  | 1.50  |
|             | Max. | 0.85          | 0.05  | -0.79 | 2.46  | -0.88 | 3.20 | 1.04  | 2.71  |
|             | IQR  | 0.64          | 0.59  | 0.41  | 1.06  | 0.19  | 0.42 | 0.64  | 0.49  |
| Vía férrea  |      |               |       |       |       |       |      |       |       |
| <i>clr</i>  | Min. | -1.80         | -3.08 | -3.15 | -0.91 | -2.31 | 0.63 | -1.03 | 0.58  |
|             | Q50  | -0.88         | -0.50 | -0.71 | -0.14 | -1.66 | 1.20 | 0.77  | 1.88  |
|             | Max. | 0.91          | 0.04  | 0.19  | 1.44  | -0.44 | 4.37 | 1.38  | 4.10  |
|             | IQR  | 0.40          | 0.53  | 0.62  | 0.72  | 0.41  | 0.40 | 0.42  | 0.95  |

*Todos los valores son relaciones, es decir, adimensionales: Min: mínimo; Q50: mediana; Max: máximo; IQR: intercuartil distancia.*

De los análisis anteriores, se determina que todos los elementos tienen procedencia tanto natural como antrópica, sin embargo, los oligoelementos As, Pb, y Cu destacan por su procedencia antrópica la cual está ligada a las diversas actividades de la ciudad de Arica. De la Tabla 6 se deriva

que la mayor fuente antrópica de los elementos proviene del sector Industrial, algunos casos específicos se comparten como Cr y Pb en Residencial y Zn en Vía férrea. Mientras que los niveles más bajos de contaminación los presenta el Ni y el V.



**Figura VII:** Diagrama cuaternario de los conjuntos del análisis clúster.

### 3.3. Análisis de datos composicionales (CoDA) con tratamiento de *Outliers*.

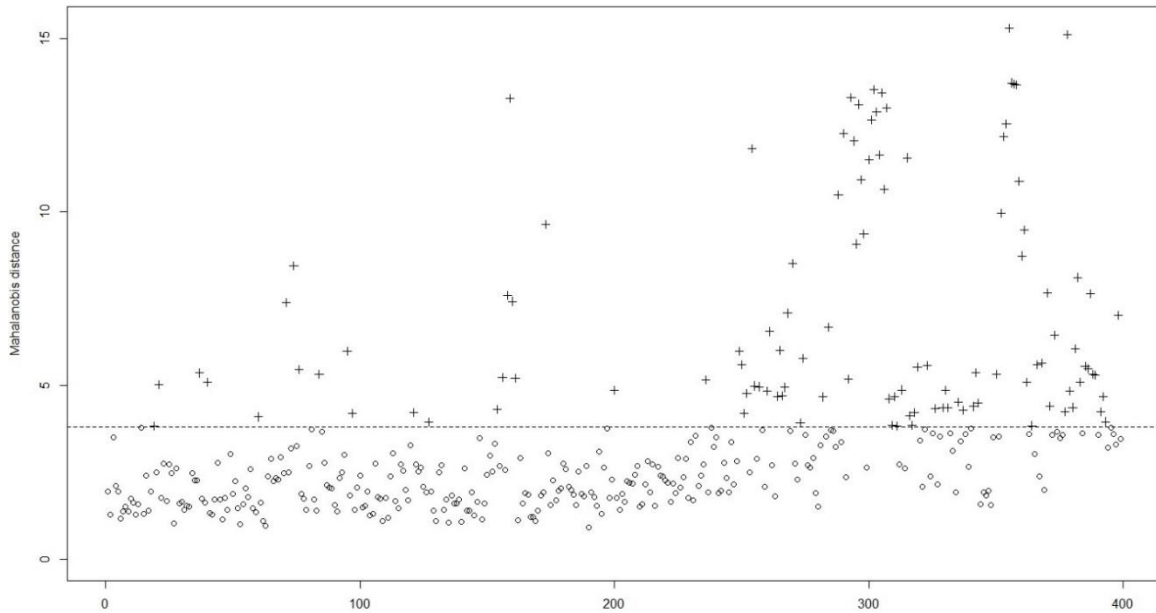
El tratamiento de datos composicionales en su enfoque robusto se llevó a cabo con el paquete “*robCompositions*” para R, el cual permite un manejo de los datos atípicos u *Outliers*. En la Fig. VIII se puede observar las muestras que presentaron valores atípicos y que fueron reconocidas por este paquete. Mientras que en la Tabla 9 se ve la matriz de variación con sus log-cocientes entre partes. De esta matriz destaca el Cr el cual posee variaciones altas en relación a los demás elementos, salvo con el Ba. Esto indica que el Cr no posee una relación directa con metales como el Cu, Ni, Pb, V y Zn. Entre las variaciones más bajas se observan las del As con el Zn y Cu lo cual indica una relación directamente proporcional de estos elementos. El V junto con el Ni poseen la variación más cercana a 0 con un valor de 0.267, lo cual indica al alta relación de estos elementos.

**Tabla 9**

*Matriz de variación.*

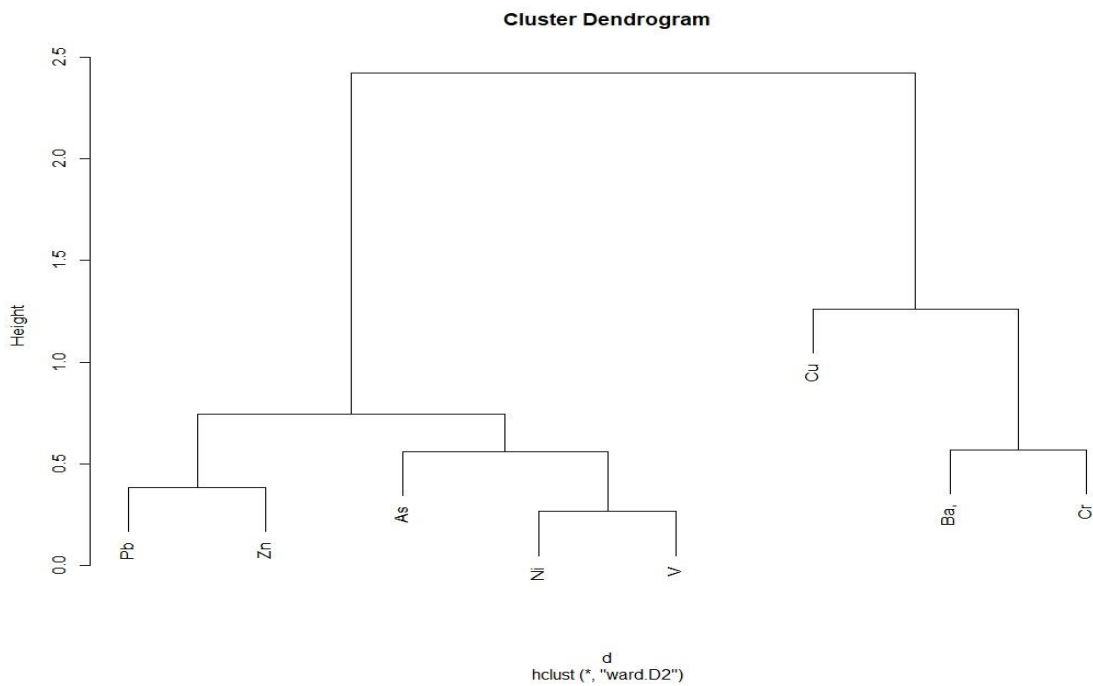
|           | <i>As</i> | <i>Ba</i> | <i>Cr</i> | <i>Cu</i> | <i>Ni</i> | <i>Pb</i> | <i>V</i> | <i>Zn</i> |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|-----------|
| <i>As</i> | 0.000     | 0.621     | 1.474     | 0.765     | 0.300     | 0.561     | 0.647    | 0.595     |
| <i>Ba</i> |           | 0.000     | 0.570     | 0.979     | 1.029     | 1.608     | 0.765    | 1.033     |
| <i>Cr</i> |           |           | 0.000     | 1.261     | 1.455     | 2.915     | 1.502    | 1.622     |
| <i>Cu</i> |           |           |           | 0.000     | 1.231     | 1.313     | 1.385    | 0.853     |
| <i>Ni</i> |           |           |           |           | 0.000     | 0.448     | 0.267    | 0.378     |
| <i>Pb</i> |           |           |           |           |           | 0.000     | 0.648    | 0.384     |
| <i>V</i>  |           |           |           |           |           |           | 0.000    | 0.749     |
| <i>Zn</i> |           |           |           |           |           |           |          | 0.000     |

*Mitad superior derecha las varianzas de los log-cocientes entre partes.*



**Figura VIII:** Dispersión muestras y determinación de *Outliers*.

Para la realización del Diagrama Clúster (AC) y Análisis de Componentes Principales (ACP) se utilizaron los mismos métodos (Transformaciones y matrices) mencionados anteriormente solo que aplicando el tratamiento de datos atípicos.



**Figura IX:** Análisis Clúster (CA) para CoDA robusto, método “Ward”.

Así el AC Fig. IX presenta nuevos grupos a su vez que mantiene otros, como el del Vanadio con el Níquel y el del Bario con el Cromo. Ahora se pueden apreciar asociaciones entre el Pb y Zn; Cu, Ba y Cr; As, Ni y V.

Respecto al ACP se identificaron tres componentes principales (PC1 a PC3) para los elementos en estudio Tabla 10, estas tres componentes explican de forma acumulativa el 82.6% de la varianza total, lo cual fue el criterio para seleccionar dichas componentes, en conjunto a la regla del codo. En este análisis la tasa de contribución de la varianza para la PC1 es de 43.75% con el Cr, Cu, Pb y As con las cargas más altas, la componente 2 (PC2) contribuye a la varianza en un 28.107% y sus cargas más altas corresponden a Cu, V, Zn y Cr. Por último PC3 con una tasa de contribución de la varianza de 10.81% tiene las cargas más elevadas en los elementos As, Ba, V y Zn.

La componente 1 (PC1) que representa el 43.75% de la varianza, posee la carga del Cr significativamente más elevada que el resto de sus pares Cr (0.640), seguido de Cu (0.385), Pb (-0.340), As (-0.294) y Ni (-0.275), las relaciones de estos elementos podrían significar una fuente natural, sumando a esta el V (-0.207).

La componente 2 (PC2) con una tasa de contribución de varianza de 28.107%, sus cargas más elevadas son Cu (0.650), V (0.448), Zn (0.379) y Cr (0.376) estos elementos presentan en su mayoría variaciones altas entre sí, lo que indica la poca relación que tienen estos, por lo cual podría tratarse de elementos de procedencia mixta, en la cual el Cu con el Zn podrían tener un mayor origen antrópico.

Por último, no menos importante la componente 3 (PC3) que representa el 10.81% de la varianza, sus cargas más altas corresponden a los elementos As (-0.58), Ba (-0.54), V (0.419) y Zn (0.370), esta componente podría tratarse de elementos de origen antrópico.

**Tabla 10**

*Análisis de componentes principales CoDA robusto, "robCompositions".*

| Elemento Traza       | Componentes Principales |        |        |        |        |        |        |
|----------------------|-------------------------|--------|--------|--------|--------|--------|--------|
|                      | PC1                     | PC2    | PC3    | PC4    | PC4    | PC5    | PC6    |
| As                   | -0.294                  | 0.046  | -0.583 | 0.065  | 0.621  | -0.024 | -0.237 |
| Ba                   | 0.287                   | 0.136  | -0.543 | -0.261 | -0.575 | 0.274  | -0.073 |
| Cr                   | 0.640                   | 0.376  | 0.216  | -0.166 | 0.302  | -0.393 | 0.063  |
| Cu                   | 0.385                   | -0.653 | 0.059  | 0.511  | 0.050  | 0.181  | 0.029  |
| Ni                   | -0.275                  | 0.197  | 0.057  | 0.082  | 0.028  | 0.209  | 0.840  |
| Pb                   | -0.340                  | -0.172 | 0.005  | 0.099  | -0.396 | -0.750 | -0.031 |
| V                    | -0.207                  | 0.448  | 0.419  | 0.368  | -0.140 | 0.302  | -0.457 |
| Zn                   | -0.195                  | -0.379 | 0.370  | -0.698 | 0.111  | 0.200  | -0.134 |
| % Varianza explicada | 43.750                  | 28.107 | 10.812 | 7.035  | 5.700  | 2.396  | 2.199  |
| % Varianza e. ac.    | 43.75                   | 71.86  | 82.67  | 89.70  | 95.40  | 97.80  | 100.00 |

En resumen se aprecia una mayor relación entre elementos lo cual da una imagen más real. Elementos como el As y el Zn poseen variaciones bajas con respecto casi todos los demás elementos, indicando que estos poseen tanto fuentes antropogénicas como naturales. El Cu con el

Zn están asociados y representan una contaminación antrópica junto con el As. Elementos como el V, Ni son los que presentan los menores niveles de contaminación. El resto de los metales traza como Pb, Ba, Cr y Ni presentan en su mayor parte fuentes mixtas de contaminación.

### 3.4. “*compositions*” versus “*robCompositons*”.

Tanto el “*compositions*” como el “*robCompositions*” son herramientas poderosas al momento de trabajar con datos de composición. Sin mencionar que ambos están disponibles de forma gratuita tanto en sus sitios web como en el mismo software R.

En cuanto al manejo que el usuario puede darle a cada paquete, el “*compositions*” otorga muchas más opciones, pero en lo que respecta al tratamiento de datos atípicos puede ser algo débil.

En relación al análisis Clúster, ambos paquetes muestran asociaciones bastante similares, manteniendo la del V con el Ni y del Ba con el Cr, los cuales tienen como su origen principalmente de fuentes naturales. El análisis robusto detectó un total de 120 muestras con valores atípicos, en su mayoría estos valores correspondían a concentraciones de As, Cu, Pb y Zn provenientes de las zonas Industrial, Vía férrea y Sitio F. Información que sin duda puede afectar el análisis de datos, de no ser tratada correctamente. Respecto al Análisis de componentes principales, en el ACP (Realizado por “*compositions*”) de determinaron cuatro componentes las cuales explican el 86.51% de la varianza, mientras que el ACP Robusto (Realizado con “*robCompositions*”) se encontraron tres componentes las cuales determinan el 82.67% de la varianza. En cuanto a que muestra cada ACP se llegan a resultados bastante similares, como que el Ni y V son los elementos que presentan la menor contaminación y que los elementos con mayor origen antropogénico corresponden al As, Cu y Zn. Una diferencia cae en el Pb y Cr los cuales en el ACP se atribuyen a fuentes antropogénicas y mixtas respectivamente, mientras que el ACP robusto, ambos son atribuidos a fuentes mixtas.

Por último, al analizar los datos con el método robusto, la variabilidad entre los elementos se ve considerablemente disminuida, lo cual muestra una mayor relación entre elementos. Esto se puede apreciar tanto en el análisis Clúster como en el Análisis de Componentes Principales, en el cual un menor número de componentes pueden explicar la varianza.

### 3.5. Estadística Clásica versus Estadística composicional CoDA.

En ámbitos de estadísticos univariado o descriptivos, las comparaciones entre estadística Clásica y CoDA no pueden desarrollarse a cabalidad, pues mientras que la primera trabaja con valores como la media, SD, simetría, Kurtosis entre otros la estadística CoDA se puede apoyar de cierta forma de los Cuartiles, IQR y MAD, por otra parte las unidades de medida son muy dispares pues la principal unidad de medida en la estadística clásica para este caso son los mg/kg mientras que para CoDA no son más que Wt% y transformaciones *clr* adimensionales.

La matriz de correlación en la estadística Clásica, resulta de una herramienta muy útil al momento de interpretar datos, ver el comportamiento entre estos y aplicar métodos estadísticos multivariados como lo es el Análisis de Componentes Principales, pero el requisito de la correlación es que dos variables no deben verse afectadas por la influencia de otra variable. Este requisito nunca se

cumple cuando se trata de datos composicionales (geoquímicos, etcétera). Incluso si se analizan dos elementos, todos los demás elementos de la composición influyen en las concentraciones medidas para estos dos elementos (Birke y Reimann, 2016). Sin embargo, en la estadística CoDA si son de utilidad la matriz de variación, matriz de varianza- covarianza y la media geométrica, para la interpretación de los datos.

En este punto se compararán los enfoques PCA clásicos y PCA CoDA. Debe establecerse claramente que un PCA "clásico" es estadísticamente flojo, porque se lleva a cabo en la geometría euclidiana equivocada en lugar del simplex de Aitchison (Aitchison, 1986; Pawlowsky-Glahn y Buccianti, 2011; Filzmoser y Hron, 2011). Para efectuar dicha comparación se utilizaron los datos y resultados del artículo “An assessment of the potentially hazardous element contamination in urban soils of Arica, Chile” (Tume et al., 2016) el cual está basado en estadística clásica.

En dicho estudio se obtuvieron fuertes correlaciones Spearman positivas para los pares de oligoelementos Zn/Cu, Zn/Pb y Ba/Cr, mientras que el Ni mostro ligera correlación con Pb, Zn y V. En cuanto al coeficiente de variación este disminuye en el siguiente orden Cu>As>Pb>Zn>Cr>Ni>Ba>V. Respecto al Análisis de componentes principales clásico se obtuvieron cuatro componentes que explican el 73.9% de la varianza, la componente 1 PC1 es representativa de los elementos As, Pb y Cu, PC2 de Cr, Ba y Cu, PC3 con Zn y Ni y por ultimo PC4 con V. La presencia de Cu en distintas componentes PC1 y PC2 indico fuente mixtas de este elemento, tanto natural como antrópicas. El PCA apoyado del CA método vecino más cercano, distancia Euclidiana) sugirieron una fuente antropogénica común para los elementos As, Cu, Pb y Zn. Mientras que Ba, Ni y V mostraron bajos niveles de contaminación.

**Tabla 11**  
*Análisis de componentes principales Clásico.*

| Elemento Traza          | PCA Clásico |       |       |       |
|-------------------------|-------------|-------|-------|-------|
|                         | PC1         | PC2   | PC3   | PC4   |
| As                      | 0.46        | 0.21  | 0.04  | 0.11  |
| Ba                      | -0.37       | 0.45  | 0.27  | -0.37 |
| Cr                      | -0.33       | 0.53  | 0.32  | 0.01  |
| Cu                      | 0.38        | 0.35  | 0.17  | -0.10 |
| Ni                      | 0.06        | -0.34 | 0.62  | 0.02  |
| Pb                      | 0.59        | 0.32  | 0.11  | 0.07  |
| V                       | -0.21       | 0.18  | 0.07  | 0.91  |
| Zn                      | 0.07        | -0.31 | 0.63  | 0.04  |
| % Varianza explicada    | 25.81       | 19.31 | 16.66 | 12.15 |
| % Varianza e. acumulada | 25.81       | 45.11 | 61.77 | 73.92 |

*Fuente: An assessment of the potentially hazardous element contamination in urban soils of Arica, Chile (Tume et al., 2016)*

En CoDA, las relaciones entre elementos se aprecian a través de la matriz de variación, en esta los pares de elementos con valores más cercanos a cero, son los más relacionados entre sí. El estudio arrojó que los elementos con relaciones más altas son V/Ni, Cr/Ba, Zn/Ni, Cr/Ni, Pb/As y Zn/Pb, como se aprecia, existen algunos pares que comparten similitud con los de la estadística Clásica,

los pares nuevos que aparecen son Pb/As y Cr/Ni. En cuanto al ACP CoDA arrojó cuatro componentes que explican el 86.5% de la varianza acumulada. Como se mencionó anteriormente, este análisis arrojó que todos los elementos tienen fuentes antrópicas y naturales, pero destaca una fuente antrópica con los elementos As, Cu y Pb, contaminación antrópica del Cr y el Pb y baja contaminación del Ni y V.

En retrospectiva, ambos análisis tanto clásico como CoDA entregaron resultados similares, sin embargo es notorio que la información entregada por CoDA es bastante más detallada y sujeta a mayor interpretación. En la Tabla 11 se puede ver el PCA clásico.

#### 4. Conclusiones.

En general los metales trazas con más fuentes antrópicas fueron el Cu, Pb, As y Zn, mientras que el Ni y V presentaron una baja contaminación. El estudio arrojó que la zona más propensa a la contaminación es la Industrial y Vía férrea lo cual supone que es debido al constante transporte de metales trazas desde y hacia el puerto de Arica. Sin embargo es necesario realizar estudios más detallados, como mapas geoquímicos, para determinar con certeza el origen de los metales. Al comparar el PCA clásico con el composicional, se obtienen resultados similares, sin embargo, las componentes del PCA composicional explican de forma más acertada el panorama real de la contaminación por metales trazas en estos suelos, esto es debido a la normalización de los datos a través de una transformación *log-ratio* para este caso la logarítmica centrada *clr*. El analizar estos datos permite indagar de forma más acertada sobre el origen de los metales, tanto en su origen natural como antrópico.

#### 5. Revisión bibliográfica.

- Aitchison, J. (2003). *A Concise Guide to Compositional Data Analysis* Department of Statistics University of Glasgow.
- Alperin, M. (2013). *Introducción al análisis estadístico de datos geológicos*. Universidad Nacional de la Plata.
- Agriquem América S.A (2009), Análisis químico de suelos de la ciudad de Arica “Diagnostico actualizado de la calidad química del suelo y subsuelo urbano ciudad de Arica, para aplicar en caso necesario medidas de gestión y regulación
- Birke, M., Reimann, C., Rauch, U., Ladenberger, A., Demetriades, A., Jähne-Klingberg, F., . . . Halamić, J. (2017). GEMAS: Cadmium distribution and its sources in agricultural and grazing land soil of Europe — Original data versus *clr*-transformed data. *Journal of Geochemical Exploration*, 173(Supplement C), 13-30. doi:<https://doi.org/10.1016/j.gexplo.2016.11.007>
- Buccianti, A., & Grunsky, E. (2014). Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes? *Journal of Geochemical Exploration*, 141(Supplement C), 1-5. doi:<https://doi.org/10.1016/j.gexplo.2014.03.022>
- Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., & De Vivo, B. (2015). Exploring topsoil geochemistry from the CoDA (Compositional Data Analysis) perspective: The multi-element data archive of the Campania Region (Southern Italy). *Journal of Geochemical Exploration*, 159(Supplement C), 302-316. doi:<https://doi.org/10.1016/j.gexplo.2015.10.006>

- Buccianti, A., Lima, A., Albanese, S., & De Vivo, B. (2017). Measuring the change under compositional data analysis (CoDA): Insight on the dynamics of geochemical systems. *Journal of Geochemical Exploration*. doi:<https://doi.org/10.1016/j.gexplo.2017.05.006>
- Campos, D. (2011). *Datos composicionales y su análisis mediante R* PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO.
- Edjabou, M. E., Martín-Fernández, J. A., Scheutz, C., & Astrup, T. F. (2017). Statistical analysis of solid waste composition data: Arithmetic mean, standard deviation and correlation coefficients. *Waste Management*, 69(Supplement C), 13-23. doi:<https://doi.org/10.1016/j.wasman.2017.08.036>
- Filzmoser, P. (2013). Tutorial to the package StatDA. In (pp. 38). Institute of Statistics and Probability Theory, Vienna University of Technology, Austria
- Filzmoser, P., Hron, K., & Reimann, C. (2009). Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of The Total Environment*, 407(23), 6100-6108. doi:<https://doi.org/10.1016/j.scitotenv.2009.08.008>
- Filzmoser, P., Hron, K., & Reimann, C. (2010). The bivariate statistical analysis of environmental (compositional) data. *Science of The Total Environment*, 408(19), 4230-4238. doi:<https://doi.org/10.1016/j.scitotenv.2010.05.011>
- Filzmoser, P., Hron, K., Reimann, C., & Garrett, R. (2009). Robust factor analysis for compositional data. *Computers & Geosciences*, 35(9), 1854-1861. doi:<https://doi.org/10.1016/j.cageo.2008.12.005>
- Filzmoser, P., Hron, K., & Tolosana-Delgado, R. (2016). Statistical analysis of geochemical compositions: Problems, perspectives and solutions. *Applied Geochemistry*, 75(Supplement C), 169-170. doi:<https://doi.org/10.1016/j.apgeochem.2016.11.016>
- Filzmoser, P., & Todorov, V. (2011). Review of robust multivariate statistical methods in high dimension. *Analytica Chimica Acta*, 705(1), 2-14. doi:<https://doi.org/10.1016/j.aca.2011.03.055>
- Garrido-Fernández, A., Cortés-Delgado, A., & López-López, A. (2018). Tentative application of compositional data analysis to the fatty acid profiles of green Spanish-style Gordal table olives. *Food Chemistry*, 241(Supplement C), 14-22. doi:<https://doi.org/10.1016/j.foodchem.2017.08.064>
- Grunsky, E. C., Mueller, U. A., & Corrigan, D. (2014). A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predictive geological mapping. *Journal of Geochemical Exploration*, 141(Supplement C), 15-41. doi:<https://doi.org/10.1016/j.gexplo.2013.07.013>
- Huertos, G., & Romero Baena, A. (2004). Contaminación de suelos por metales pesados. In (pp. 13). Universidad de Sevilla. Sevilla: Departamento de Cristalografía, Mineralogía y Química Agrícola. Facultad de Química.
- Kabata-Pendias, A., & Mukherjee, A. B. (2007). Trace elements from soil to human. In (pp. 560): Springer Science & Business Media.
- Mateu-Figueras, G., Martín-Fernández, J. A., V. Pawlowsky-Glahn, V., & Barceló Vidal, V. (2003). EL PROBLEMA DEL ANALISIS ESTADISTICO DE DATOS COMPOSICIONALES In (pp. 9). Universidad de Girona, 17071 Girona.
- McKinley, J. M., Hron, K., Grunsky, E. C., Reimann, C., de Caritat, P., Filzmoser, P., . . . Tolosana-Delgado, R. (2016). The single component geochemical map: Fact or fiction? *Journal of Geochemical Exploration*, 162(Supplement C), 16-28. doi:<https://doi.org/10.1016/j.gexplo.2015.12.005>

- Menghan, W., Stefano, A., Annamaria, L., Claudia, C., Antonio, C., Wanjun, L., . . . Benedetto, D. V. (2015). Compositional analysis and pollution impact assessment: A case study in the Gulfs of Naples and Salerno. *Estuarine, Coastal and Shelf Science*, 160(Supplement C), 22-32. doi:<https://doi.org/10.1016/j.ecss.2015.03.031>
- Mokhtari, A. R., Roshani Rodsari, P., Cohen, D. R., Emami, A., Dehghanzadeh Bafghi, A. A., & Khodaian Ghegeni, Z. (2015). Metal speciation in agricultural soils adjacent to the Irankuh Pb–Zn mining area, central Iran. *Journal of African Earth Sciences*, 101(Supplement C), 186-193. doi:<https://doi.org/10.1016/j.jafrearsci.2014.09.003>
- Orozco C., Perez A., Gonzalez M., Rodriguez F., Alfayate J., (2003) “Contaminación Ambiental: Una visión desde la química”
- Pawlowsky-Glahn, V., & Egozcue, J. J. (2006). Compositional data and their analysis: an introduction. In (pp. 264 (261), 261-210.). *Geological Society, London,: Special Publications* .,
- Pawlowsky-Glahn, V., & Egozcue, J. J. (2016). Spatial analysis of compositional data: A historical review. *Journal of Geochemical Exploration*, 164(Supplement C), 28-32. doi:<https://doi.org/10.1016/j.gexplo.2015.12.010>
- Reimann, C., Filzmoser, P., Fabian, K., Hron, K., Birke, M., Demetriades, A., . . . Ladenberger, A. (2012). The concept of compositional data analysis in practice — Total major element concentrations in agricultural and grazing land soils of Europe. *Science of The Total Environment*, 426(Supplement C), 196-210. doi:<https://doi.org/10.1016/j.scitotenv.2012.02.032>
- Reimann, C., Filzmoser, P., Hron, K., Kynčlová, P., & Garrett, R. G. (2017a). A new method for correlation analysis of compositional (environmental) data – a worked example. *Science of The Total Environment*, 607-608(Supplement C), 965-971. doi:<https://doi.org/10.1016/j.scitotenv.2017.06.063>
- Reimann, C., Filzmoser, P., Hron, K., Kynčlová, P., & Garrett, R. G. (2017b). A new method for correlation analysis of compositional (environmental) data – a worked example. *Science of The Total Environment*, 607-608(Supplement C), 965-971. doi:<https://doi.org/10.1016/j.scitotenv.2017.06.063>
- Thiombane, M., Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Cavaliere, M., & De Vivo, B. (2017). Soil geochemical follow-up in the Cilento World Heritage Park (Campania, Italy) through exploratory compositional data analysis and C-A fractal model. *Journal of Geochemical Exploration*. doi:<https://doi.org/10.1016/j.gexplo.2017.06.010>
- Tolosana-Delgado, R. (2011). Guía para el análisis espacial de datos composicionales. In (pp. 14 (469-482)). Universitat Politècnica de Catalunya: Boletín Geológico y Minero.
- Tolosana-Delgado, R. (2012). Uses and misuses of compositional data in sedimentology. *Sedimentary Geology*, 280(Supplement C), 60-79. doi:<https://doi.org/10.1016/j.sedgeo.2012.05.005>
- Tolosana-Delgado, R., & McKinley, J. (2016). Exploring the joint compositional variability of major components and trace elements in the Tellus soil geochemistry survey (Northern Ireland). *Applied Geochemistry*, 75(Supplement C), 263-276. doi:<https://doi.org/10.1016/j.apgeochem.2016.05.004>
- Tolosana-Delgado, R., & van den Boogaart, K. G. (2014). Towards compositional geochemical potential mapping. *Journal of Geochemical Exploration*, 141(Supplement C), 42-51. doi:<https://doi.org/10.1016/j.gexplo.2014.02.026>
- Tume, P., González, E., King, R. W., Monsalve, V., Roca, N., & Bech, J. (2016). Spatial

- distribution of potentially harmful elements in urban soils, city of Talcahuano, Chile. *Journal of Geochemical Exploration*. doi:<https://doi.org/10.1016/j.gexplo.2016.12.007>
- Tume, P., Roca, N., Rubio, R., King, R., & Bech, J. (2016). An assessment of the potentially hazardous element contamination in urban soils of Arica, Chile. *Journal of Geochemical Exploration*. doi:<https://doi.org/10.1016/j.gexplo.2016.09.011>
- Van den Boogaart, K. G. (2008). Using the R package "compositions". In (Vol. 0.91, 10 pp. 17).
- Van den Boogaart, K. G., Tolosana, R., & Bren, M. (2005). The compositions Package: Compositional Data Analysis. In (Vol. 0.9-11, pp. 130).
- Van den Boogaart, K. G., & Tolosana-Delgado, R. (2008). "compositions": A unified R package to analyze compositional data. *Computers & Geosciences*, 34(4), 320-338. doi:<https://doi.org/10.1016/j.cageo.2006.11.017>
- V. Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2007). Lecture notes on Compositional Data Analysis. In (pp. 96 ). University of Girona.
- Zuzolo, D., Cicchella, D., Albanese, S., Lima, A., Zuo, R., & De Vivo, B. (2017). Exploring uni-element geochemical data under a compositional perspective. *Applied Geochemistry*. doi:<https://doi.org/10.1016/j.apgeochem.2017.10.003>