

**UNIVERSIDAD DE CONCEPCIÓN**

Facultad de Ingeniería

Departamento de Ingeniería Metalúrgica

**Profesor Patrocinante:**

Roberto Fustos T.

**Comisión:**

Ramón Díaz N.

Francisco Muñoz G.

**METODOLOGÍA PARA EL MODELAMIENTO Y  
CARACTERIZACIÓN GEOESTADÍSTICA DE RELAVES  
COMO MÉTODO DE REVALORIZACIÓN DE PASIVOS  
AMBIENTALES**

**RICARDO IGNACIO GARCIA VEGA**

Informe de Memoria de Título para optar al título de  
Ingeniero Civil de Minas

2020

## Resumen

El presente trabajo propone una metodología para modelar y cuantificar la distribución de determinados elementos químicos contenidos en un depósito de relaves de tipo torta, mediante el uso de distintas herramientas geoestadísticas típicas de procedimientos de estimación de recursos en prospección minera. A través de la comparación de los resultados se define el método de Support Vector Machines, en su variante SVR, como el más preciso entre los tres métodos propuestos para predecir la distribución de los elementos de interés dentro del relave. Posterior a este primer proceso se generó la partición del relave en porciones de características similares (clúster), mediante el algoritmo de agrupamiento de k-medias, estableciendo los dominios geológicos correspondientes. A partir de este, se obtuvieron las leyes medias y tonelajes de cada elemento en cada clúster, definiendo distintas zonas de interés en función de la cantidad y calidad del o los elementos contenidos. De esta manera se facilita y motiva un futuro proceso que genere un nuevo valor para el pasivo medioambiental hacia un activo económica y socialmente viable bajo los principios de las nuevas corrientes filosóficas de economía circular donde los desechos son considerados como nueva fuente de materia prima.

### Abstract

*This research proposes a defined methodology to model and quantify the distribution of certain chemical elements with help of some Geostatistics tools typically used in estimation of natural resources. By comparing the three method's results, the Support Vector Machine's method was chosen as the most precise one, with the lowest error, in predicting the distribution of the grades within the tailings. After this estimation/validation process, clusters were formed by k-means algorithm, dividing the whole body into portions of similar elements characteristics establishing the respective geological domains. Those portions were defined through both average grade and tonnage for each element, facilitating and motivating future processes that might produce significant revenues that could transform the concept of environmental passives not only into new economic terms but social and environmental too (sustainability), under the principles of new philosophical current named circular economy where waste is nothing but raw/prime material, ensuring, at basis, a sustainable development of the mining activity.*

# Contenido

1.	Introducción .....	1
2.	Motivación del estudio.....	4
3.	Hipótesis, objetivos y alcances.....	6
4.	Metodología de trabajo .....	7
5.	Marco Teórico .....	10
5.1.	Pasivos Ambientales Mineros (PAM) .....	10
5.1.1.	Relaves.....	12
5.1.2.	Riesgos físicos.....	13
5.1.3.	Casos reales .....	13
5.1.4.	Antecedentes .....	16
5.2.	Estimación de recursos minerales.....	17
5.2.1.	Variografía.....	19
5.2.2.	Predicción.....	24
5.2.3.	Máquinas de Vector Soporte (SVM's) .....	32
5.2.4.	Categorización de recursos y reservas .....	36
5.3.	Agrupamiento y Clasificación de datos: Clustering.....	39
5.3.1.	Clustering Jerárquico.....	41
5.3.2.	Clustering Particional .....	44
5.3.3.	Validación de Resultados .....	49
5.3.4.	Dominios Geometalúrgicos .....	54
6.	Resultados .....	55
6.1.	Descripción de la base de datos.....	55
6.2.	Resultados y análisis.....	56
7.	Conclusiones y recomendaciones .....	65

8. Referencias..... 68

9. Anexos..... 74



## Índice de tablas

Tabla 1: Resúmenes de Kernels más frecuentes.....	35
Tabla 2: Emparejamiento de elementos según su correlación espacial.....	56
Tabla 3: Resumen de errores medios para grupo 1.....	59
Tabla 4: Resumen de errores medios para grupo 2.....	60
Tabla 5: Resumen de errores medios para grupo 3.....	60
Tabla 6: Resumen de errores medios para grupo 4.....	60
Tabla 7: Coeficientes de correlación para grupo 1 .....	61
Tabla 8: Coeficientes de correlación para grupo 2 .....	61
Tabla 9: Coeficientes de correlación para grupo 3 .....	61
Tabla 10: Coeficientes de correlación para grupo 4 .....	61
Tabla 11: Resumen resultado de clustering grupo 1 .....	63
Tabla 12: Resumen resultado de clustering grupo 2 .....	63
Tabla 13: Resumen resultado de clustering grupo 3 .....	63
Tabla 14: Resumen resultado de clustering grupo 4 .....	63
Tabla 15: Resumen de caracterización final para los elementos en el relave .....	64
Tabla 16: Modelos de bloques para grupo 1 .....	74
Tabla 17: Modelos de bloques para grupo 2 .....	75
Tabla 18: Modelos de bloques para grupo 3 .....	75
Tabla 19: Modelos de bloques para grupo 4 .....	76
Tabla 20: Visualización de los mapas productos del clustering .....	83

## Índice de ilustraciones

Ilustración 4.1: Esquema general de metodología de trabajo.....	7
Ilustración 5.1: Catastro de relaves según su clasificación, fuente: propia/SERNAGEOMIN .	12
Ilustración 5.2: Visualización del efecto Tolerancia.....	21
Ilustración 5.3: Visualización del efecto tamaño de paso .....	22
Ilustración 5.4: Visualización de elementos básicos de un variograma.....	22
Ilustración 5.5: Método de Vecino más cercano .....	25
Ilustración 5.6: Visualización del ponderador “inverso de la distancia” .....	26
Ilustración 5.7: Visualización del aumento de dimensiones.....	35
Ilustración 5.8: Esquema de clasificación y relación entre reservas y recursos minerales. ...	37
Ilustración 5.9: Árbol de clasificación de problemas, Lance & Williams, 1967.....	40
Ilustración 5.10: El dendrograma, componentes y lectura. Fuente: Elaboración propia .....	42
Ilustración 5.11: bases de puntos agrupados mediante criterios de densidad. ....	47
Ilustración 9.1: Gráficos de dispersión para inverso de distancia, grupo 1. ....	77
Ilustración 9.2: Gráficos de dispersión para Kriging, grupo 1. ....	77
Ilustración 9.3: Gráficos de dispersión para Support Vector Regression, grupo 1. ....	78
Ilustración 9.4: Gráficos de dispersión para inverso de distancia, grupo 2. ....	78
Ilustración 9.5: Gráficos de dispersión para Kriging, grupo 2. ....	79
Ilustración 9.6: Gráficos de dispersión para Support Vector Regression, grupo 2. ....	79
Ilustración 9.7: Gráficos de dispersión para Inverso de distancia, grupo 3. ....	80
Ilustración 9.8: Gráficos de dispersión para Kriging, grupo 3. ....	80
Ilustración 9.9: Gráficos de dispersión para Support Vector Regression, grupo 3. ....	81
Ilustración 9.10: Gráficos de dispersión para Inverso de distancia, grupo 4. ....	81
Ilustración 9.11: Gráficos de dispersión Kriging, grupo 4. ....	82
Ilustración 9.12: Gráficos de dispersión para Support Vector Regression, grupo 4. ....	82
Ilustración 9.13: Diagrama de cajas para clustering de grupo 1.....	83
Ilustración 9.14: Diagrama de cajas para clustering de grupo 3.....	84
Ilustración 9.15: Diagrama de cajas para clustering de grupo 2.....	84
Ilustración 9.16: Diagrama de cajas para clustering de grupo 4.....	84
Ilustración 9.17: Método de codo para K óptimo en proceso de clustering por k-medias. ...	84
Ilustración 9.18: Precios corrientes (BML) para distintos metales. ....	84

Ilustración 9.19: Variogramas cruzados para grupo 2 .....	84
Ilustración 9.20: Variogramas cruzados para grupo 1 .....	84
Ilustración 9.21: Variograma cruzado para grupo 4 .....	84
Ilustración 9.22: Variogramas cruzados para grupo 3 .....	84



# 1. Introducción

De manera general, la actividad minera involucra procesos de extracción de especies minerales desde la corteza terrestre o desde su interior, seguido por etapas de procesamiento físico-químicos que permiten separar los elementos valiosos de aquellos que no, concentrando los mencionados en primer lugar. Finalmente, este producto puede o no pasar por procesos de posteriores de refinación antes de ser comercializado en los mercados respectivos.

Las primeras actividades mineras que se registran en territorio chileno datan de la época precolombina (siglo XIV). Los Atacameños, pueblos originarios localizados en lo que hoy es el norte grande de Chile (regiones de Arica y Parinacota, Tarapacá y Antofagasta) se familiarizaron con la extracción de algunos metales como el cobre nativo y el oro [1]. Estos eran extraídos desde las laderas de las altas cordilleras en el desierto de Atacama.

A diferencia del rol económico que actualmente representa la actividad minera para el país, su desarrollo se inició con fines artísticos y religiosos/ceremoniales. La llegada de las expediciones españolas a Sudamérica transformaría la actividad dándole un significado comercial donde las extracciones de oro desde lavaderos permitieron a los colonos la generación y acumulación de riquezas. Con el tiempo, la relevancia de la actividad minera decaería, siendo reemplazada por la agricultura como actividad principal y fundamental para el desarrollo de la sociedad y la economía de aquel entonces.

No es hasta el siglo XVIII que la minería explota como una actividad económica de peso, donde la revolución industrial genera el arribo de nuevas tecnologías, técnicas y conocimientos que dan inicio a la explotación de minerales a una gran escala: el cobre, el salitre y la plata en la zona norte; y el carbón en el centro-sur. El impacto de la exportación de cobre hacia tierras británicas fue tal que esta alcanzó, en su máxima, las 13.000 toneladas [1] en forma de ejes/matas y relaves.

Para 1860, Chile se convertiría en el mayor productor de cobre del mundo lo que trajo múltiples beneficios económicos y sociales para el país, como el desarrollo de centros urbanos, aumento del poder adquisitivo de diversos sectores socioeconómicos, mejoras de comunicación y conexión impulsadas por el desarrollo de líneas ferroviarias y puertos en las ciudades costeras. La actividad minera, caracterizada en aquel entonces por una alta ocupación de mano de obra, la extracción de múltiples yacimientos pequeños con muy altas leyes, entraría en un periodo de auge



en cuanto a demanda y producción debido al desarrollo de la energía eléctrica, la constante expansión del sector constructor y las nuevas tecnologías. Aumentando la eficiencia de los procesos mineros, el mineral que antes no era rentable debido a su poco contenido de mineral presentaba mayores beneficios que tiempo atrás.

El teniente (1904) y Chuquicamata (1911) se convirtieron en los yacimientos más importantes para la economía local, tanto así que para el año 1924 estos representarían el 80% de la producción nacional, correspondiente a unas 185 mil toneladas métricas de cobre. De esta forma se acuñaría el concepto de “Gran Minería del Cobre” (GMC) utilizado para referirse a las faenas mineras de mayor producción [1].

Debido a la relevancia de la actividad para el desarrollo económico del país es que en los años venideros se generaron medidas de protección de la actividad. Aumentos en las tributaciones y políticas sindicalistas serían la base para la reestructuración del sector, acciones que no fueron bien vistas por los capitales extranjeros. Estas incompatibilidades de objetivos alrededor de la industria minera, entre el estado y los inversionistas, generaron desconfianza y quiebres entre las partes lo que culminaría con reformas político-económicas como la nacionalización del cobre. Medida que buscaba crear bases para fortalecer el desarrollo económico nacional, pero que provocó el congelamiento de la entrada de nuevos capitales y tecnologías desde el extranjero durante un largo periodo.

Como medida para reestablecer la confianza y garantías de propiedad a los concesionarios es que se crea la “Ley Orgánica Constitucional de Concesiones Mineras” innovando en la integración del poder legislativo como organismo resolutorio en los conflictos que pudieran darse entre el estado y las mismas concesiones. El denominado “Boom de los ‘90” llevaría al sector minero a un crecimiento con un dinamismo sin precedentes. Luego de la crisis asiática de 1997 la minería se asentaría, nuevamente, como una de las actividades económicas más importantes para el país, esto reflejado en su alto aporte al producto interno bruto (PIB) nacional y a la tasa de ocupación laboral. Con el boom generado las labores de exploración se duplicarían, pasando de 4 a 8 millones de hectáreas en concepto de superficies concesionadas, permitiendo incrementar las reservas de cobre hasta entonces identificadas [1].

Las cifras demuestran el poder e importancia del sector minero, que junto con el sector manufacturero y los servicios financieros/empresariales, en 2018, conformaron poco más del 30% del aporte al PIB del país [2], porcentaje repartido de manera bastante equitativa. El sector

manufacturero aporta un 10,6% seguido por el de servicios financieros con un 9,9%, ligeramente menor es el 9,8% que aporta el sector minero, pero que en cifras se traduce a 18.800 millones de pesos. En cuanto a exportación, solamente de cobre, las cifras relativas al 2018 alcanzan las 5.900 toneladas métricas de cobre fino, donde los clientes preferenciales se localizaban en Asia, siendo China el principal importador, quien recibió 2.900 toneladas de cobre fino, de las cuales 1.700 toneladas fueron en formato granel (cementos, concentrados y cobres secundarios) y las otras 1.200 toneladas como refinado (cátodos); Japón los sigue con una cifra bastante menor correspondiente a 642 toneladas en formato granel [3].

No solo de la venta de mineral es que se beneficia el país, la ocupación que genera la actividad minera es crucial para el desarrollo de la economía local. A agosto del 2019 las cifras en cuanto a la ocupación en el sector minero resultaron bastante positivas y significantes, la sociedad nacional de minería (SONAMI) anunció el incremento de un 10% en la ocupación en un periodo de 12 meses, traduciéndose en 228.000 puestos de trabajo, 20.000 puestos más respecto a los 208.000 contabilizados el periodo anterior [4].



## 2. Motivación del estudio

La relevancia de la minería ha llevado a invertir sumas monetarias cada vez más grandes en la industria, a generar más proyectos, entregar más concesiones y facilidades para el desarrollo de la actividad minera. Sin embargo, este crecimiento de la industria tiene como consecuencia, el aumento de la producción de desechos mineros, y con ello, el aumento de los relaves dispuestos a lo largo del territorio nacional. El año 2016 el Servicio Nacional de Geología y Minería (SERNAGEOMIN) publicó un catastro de relaves mineros, en el cual se identificaron 740 zonas de disposición de relaves, de los cuales 101 (14%) están activos. De los 629 restantes (86%), un 63% (469) corresponden a relaves inactivos, es decir, que ya no reciben nuevo material pero que pueden eventualmente reanudar sus labores; los 170 restantes son descritos como relaves abandonados, los cuales no cumplen con documentación relativa a un dueño ni con la normativa vigente de cierre de faenas mineras [5] siendo estos últimos quienes representan la mayor amenaza en términos sociales y ambientales.

La presencia de diversos elementos nocivos en los relaves tales como plomo, arsénico, mercurio y material particulado fino, generan constante preocupación y daños a las poblaciones cercanas, así como también al entorno en el cual se emplazan estas masivas estructuras. El arrastre de polvos erosionados desde relaves hacia centros de desarrollo urbano y agrícola mediante las corrientes de aire (vientos) es un problema real y frecuente en el norte grande del país. La contaminación y acidificación de acuíferos es otro potencial riesgo que presenta la disposición de desechos mineros, ya sea, por ejemplo, por percolación o escurrimiento de soluciones lixiviadas desde el tranque mismo o bien por colapsos menores o totales de la o las estructuras contenedoras.

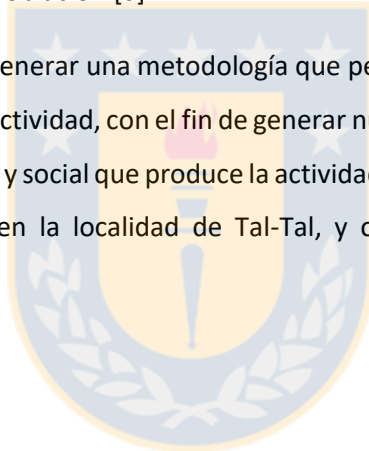
Sin embargo, así como existen elementos que representan una amenaza, los relaves también se están asociados a otros elementos de gran valor: Cobre, Oro, Plata, Tierras raras y otras especies propias de la roca madre. Elementos que por diversos motivos no fueron recuperados en procesos anteriores a su disposición (baja ley, ineficiencia de los procesos previos, material sin valor económico aparente) tienen como destino final formar parte de estos desechos.

Tal-Tal es una ciudad costera dedicada a la pesca y la pequeña minería ubicada al sur de la región de Antofagasta, Chile. Debido al desarrollo de la actividad minera en el pasado es que existen dos estructuras de acopio de relaves tipo torta: SOTRAMIN y Playa Atacama, los cuales debido a su cercanía tanto al mar como al pueblo generan preocupación. La constante interacción con el aire

salino del mar meteoriza aceleradamente los elementos constituyentes de los relaves transportándolos como material particulado fin hacia zonas aledañas, como el centro de desarrollo urbano de Tal-tal, ubicado a poco más de 1 kilómetro de la zona de disposición.

La universidad de Atacama en conjunto con el centro de investigación australiano CSIRO desarrollaron, hace algún tiempo, un proyecto de biorremediación, lo que esperaban fuera la solución para a la amenaza que representa la presencia de este PAM (pasivo ambiental minero). De manera general, la biorremediación consta en el uso de especies/organismos vivos, tales como bacterias, hongos, etc. para estabilizar compuestos y/o elementos tóxicos, como metales pesados, presentes en el relave eliminando las trazas o rastros de estos. El problema de esta medida eran los tiempos requeridos para la obtención de resultados que pudieran generar un panorama positivo, tiempo con el cual no contaban debido al estado de “emergencia” de la situación dada por la proximidad de los relaves con la población [6].

El presente trabajo busca generar una metodología que permita la revalorización y posterior recuperación de los relaves en inactividad, con el fin de generar nuevas metodologías y nuevas ideas para reducir el impacto ambiental y social que produce la actividad minera. En particular, los peligros que representan estos acopios en la localidad de Tal-Tal, y otros casos similares que puedan evaluarse en un futuro.



### 3. Hipótesis, objetivos y alcances

La presente experimentación busca demostrar el potencial económico que posee el relave SOTRAMIN y Playa Atacama en función la cantidad y calidad (tonelaje y ley) de ciertos elementos contenidos en él que puedan ser considerados valiosos y/o peligrosos.

Como objetivo principal se destaca la generación de una metodología para la caracterización y modelamiento geoestadístico de elementos químicos presentes en relaves con el menor error posible. De esta manera poder encaminar una transformación desde pasivos ambientales a activos mineros potenciales mediante el uso de distintas técnicas y herramientas propias de la estadística aplicada.

Otros objetivos a destacar:

- Comparar métodos geoestadísticos de estimación de recursos más frecuentes (Kriging e Inverso de distancia) en condiciones desfavorables.
- Generar nuevas directrices para el desarrollo de la actividad minera de manera más sustentable, presentando los desechos mineros como un objeto de potencial económico estableciendo principios hacia una actividad económica circular.
- Proponer solución para reducir el impacto ambiental y social que genera la disposición del relave a los alrededores a través de su “reutilización”.
- Generar una noción de medidas que disminuyan los riesgos asociados a la existencia de relaves mineros abandonados mediante su recuperación.

Alcances y limitaciones

- Debido a la poca cantidad de información disponible (sondajes) se decidió no hacer un entrenamiento para algunos de los procesos de estimación y validación ya que no resulta práctico ni eficiente.
- Al ser un trabajo que se enfoca en establecer una metodología para estimación de relaves, la evaluación económica realizada es más bien superflua y no comprende el análisis de costos asociados a transporte o procesamiento
- El proceso de clustering es resultado del criterio de quien realiza la experimentación por lo que los resultados con el mismo método podrían variar entre realizaciones.

## 4. Metodología de trabajo

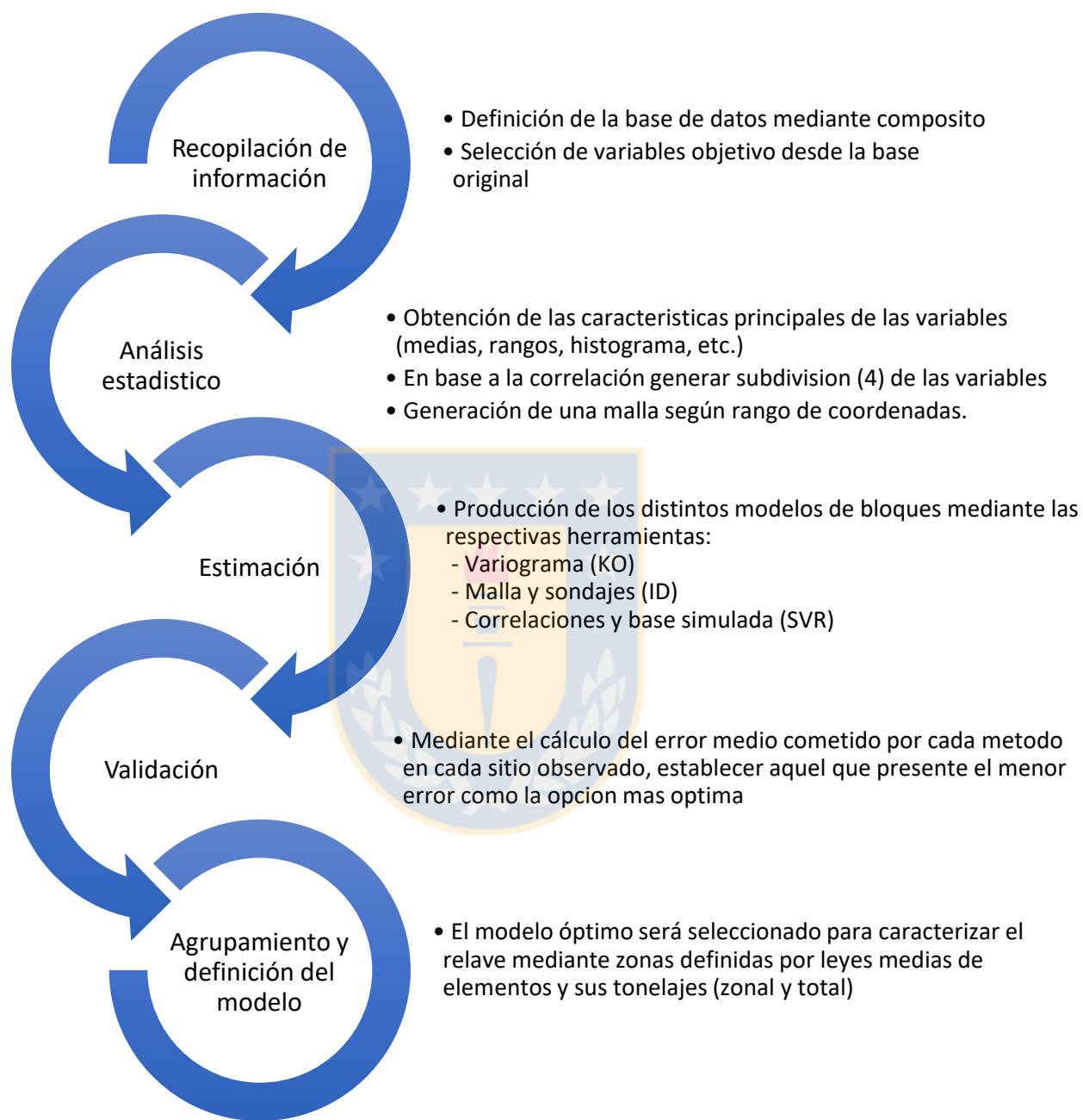


Ilustración 4.1: Esquema general de metodología de trabajo

### **Recopilación de información.**

De manera inicial se realizará el compósito de los análisis químicos obtenidos por espectrometría de masas (ICP) de los 53 sondajes correspondientes al relave SOTRAMIN y Playa Atacama. Cada sondaje fue dividido en tres zonas (superior, media e inferior), luego en base a las coordenadas del collar y las dimensiones de cada tercio de sondaje se definirán las coordenadas de cada punto que resumirá la información de las muestras obtenidas inicialmente.

Una vez generada la base de datos con coordenadas norte, este y cota se procederá a hacer una limpieza de las variables que no representen información útil para el presente estudio. Este es el caso de los elementos cuyas concentraciones son menores al umbral de detección respectivo del análisis químico (**Elementos bajo el umbral**). Esta nueva base resultante se someterá al análisis estadístico siguiente.

### **Análisis estadístico.**

A través de las herramientas de análisis de datos que ofrece Microsoft Excel, se obtendrán las características principales de las variables que componen la base: Medias, desviaciones estándar, rangos (máximo y mínimo) y correlaciones entre variables. Los rangos de coordenadas serán utilizados en la generación de una grilla para el futuro modelo de bloques, a su vez los rangos de las variables servirán para comparar los posteriores resultados de validación de los modelos obtenidos de los distintos métodos de estimación a utilizar.

Además, se realizará un segundo descarte de aquellas variables que presenten correlaciones y concentraciones bajas (**Elementos secundarios**), puesto que no representan información significativa para los procesos de estimación posteriores a esta etapa. De esta manera se establecerá la base de datos definitiva (**Elementos objetivos**) para la experimentación de la presente investigación. Los elementos objetivos serán particionados en subgrupos en función de las correlaciones que presenten facilitando los procesos posteriores de estimación.

### **Estimación.**

Para cada grupo, y cada una de sus variables, se generarán modelos bloques mediante tres métodos: Kriging ordinario, Inverso de la distancia ( $w=2$ ) y Support Vector Machine (SVM's) [véase capítulo 5.2]. Cada uno de estos métodos requerirá de información en particular para poder generar un modelo de bloques.

Para los modelos de kriging ordinario, se requerirá la generación de un variograma cruzado con las variables que conforman cada grupo; Para el modelo de inverso de la distancia, el software solo requiere de la información espacial de los sondajes y del número máximo de puntos soporte para estimar cada posición; Para el modelo SVR se generarán 100 simulaciones gaussianas para cada variable las cuales serán ponderadas generando así un modelo base para cada grupo, además cada modelo utilizará a su vez la correlación entre los elementos que conforman los respectivos grupos generando el modelo de bloques final.

### **Validación.**

Los resultados de los procesos de estimación anteriores se verificarán a través del análisis de sus errores medios, permitiendo definir un grado de confianza o precisión en el modelo resultante. Dicho proceso de corroboración se realizará mediante el método de validación cruzada [véase capítulo 5.2].

Cada valor predicho será comparado con el valor observado, la diferencia entre estos será promediada estableciendo tres errores medios (uno para cada método) para cada elemento objetivo. A partir de lo anterior, se definirá como método óptimo aquel que obtenga la menor cantidad de error medio en sus resultados.

### **Agrupamiento.**

El modelo óptimo resultante del proceso de estimación, será entonces seleccionado para representar la totalidad de las variables dentro del relave. El agrupamiento correspondiente será realizado mediante el método de k-medias [véase capítulo 5.3]. El procedimiento de clustering requerirá definir la variable <número de clúster> para cada grupo, para ello se utilizará el método de codo<sup>1</sup> del cual se obtendrán dos o más alternativas conservando la más conveniente según se estime en cada caso.

Dichos modelos permitirán caracterizar el relave mediante la definición de los distintos dominios o zonas geológicas. Cada sector estará definido por una concentración media y tonelaje de cada elemento, para poder caracterizar el relave mediante el cálculo de los tonelajes totales.

---

<sup>1</sup> *Elbow method*



## 5. Marco Teórico

### 5.1. Pasivos Ambientales Mineros (PAM)

Si bien no existe un consenso a nivel internacional respecto la definición técnica el servicio nacional de geología y minería (SERNAGEOMIN) define los pasivos ambientales mineros (PAM) como «las faenas mineras, con los desarrollos y desechos asociados, paralizadas o en estado de abandono que no cumplen con una normativa de cierre de minas constituyendo un riesgo significativo para la vida y salud de las personas y para la integridad del medio ambiente», sin la presencia de un riesgo significativo este no será considerado un PAM, tales riesgos deben entenderse como «la combinación de la probabilidad de ocurrencia de un evento y la magnitud de su consecuencia» [7].

Para un desarrollo sustentable de la minería es indispensable contar con una normativa tanto para el cierre de faenas como para el manejo y gestión de los PAM, es por ello que en 2012 fue publicada la ley 12.551 que regula el cierre de faenas mineras, implementada con el fin de **resguardar la vida de las personas y del medio ambiente, evitar el aumento de casos de abandono de faenas y mitigar los efectos dañinos propios de la industria minera** [8]. La ley obliga a toda faena minera que quiera iniciar actividades a generar un plan de cierre que debe ser aprobado por el servicio nacional de minería, plan que debe asegurar que el fin anteriormente mencionado se cumpla de manera real y efectiva durante y hasta el final de la vida del proyecto.

Los PAM presentan riesgos que pueden afectar de diversas formas el medio ambiente y el desarrollo de la vida para las zonas en las cuales se emplazan estos proyectos. Estas consecuencias para el medio ambiente pueden dividirse en 3 categorías

- Aguas

La contaminación de aguas se produce por la liberación y migración de elementos tóxicos contenidos en las rocas presentes en los residuos y obras mineras hacia los cuerpos de agua cercanos. Como mecanismos de transporte de estos elementos tóxicos, tanto como para minería subterránea como cielo abierto, se pueden mencionar: las descargas directas de aguas de proceso, aguas de mina, escurrimientos e infiltraciones hacia acuíferos en las proximidades de las faenas. Los principales impactos y repercusiones son la descarga de sedimentos, aumento de acides

(disminución de pH), contaminación de agua potable, degradación y destrucción de ecosistemas hídricos.

- Suelo

El suelo puede sufrir los impactos por contaminación directa o bien por degradación y erosión. La contaminación de suelo se produce por acarreo de polvos (material particulado) desde desechos mineros hacia suelos sanos, proceso facilitado por las corrientes de aire y la incorrecta disposición del material fino. La degradación y erosión de los suelos se deben a la constante exposición de material removido y procesado desde las faenas, junto con la destrucción de la capa vegetal protectora, en caso de existir.

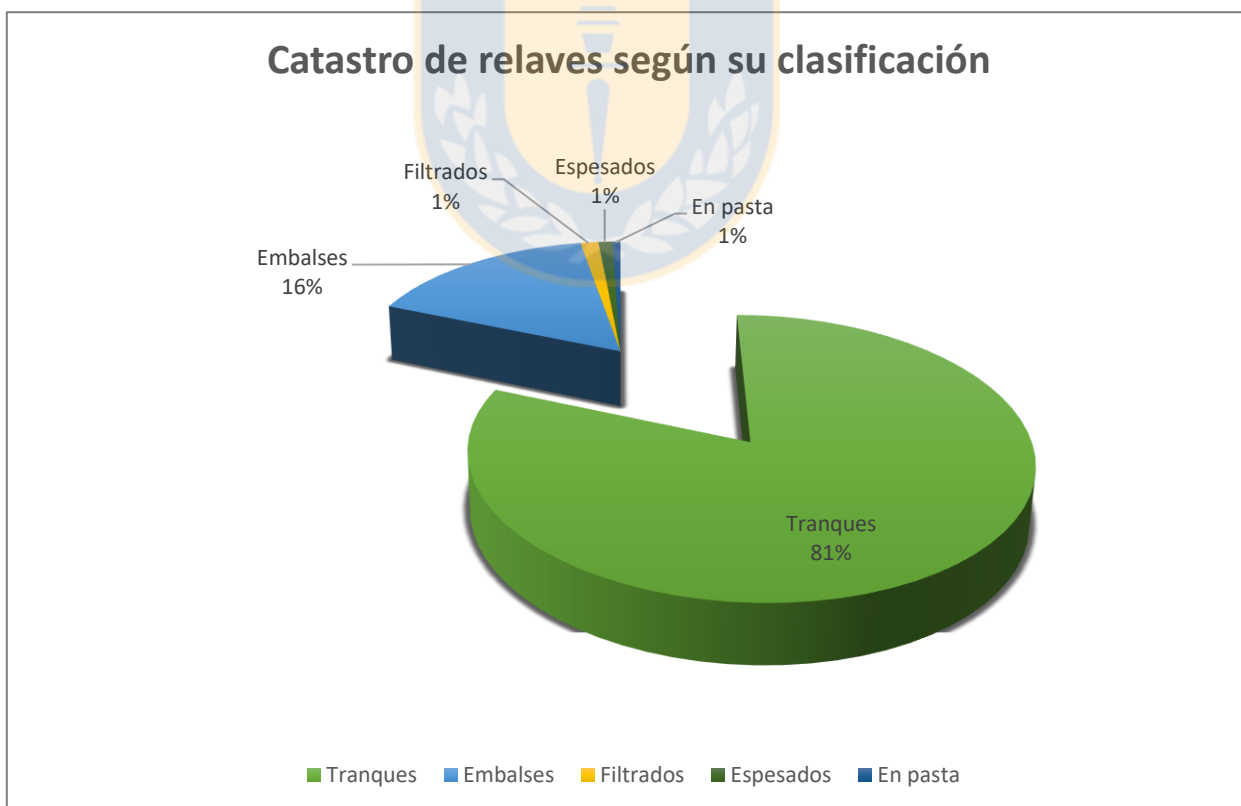
- Aire

El principal problema de contaminación de aire asociado a los pasivos ambientales mineros es el transporte de material particulado hacia sectores habitados, empobreciendo la calidad del aire y generando daños tanto a personas como animales que inhalen, ingieran o tengan contacto con estos polvos nocivos. El nivel de desplazamiento que pueda tener el material particulado dependerá de factores climáticos (humedad ambiental, vientos, etc.), de las características del polvo (tamaño, densidad) y del terreno (relieve). Como principales fuentes de emanación de estos contaminantes se pueden mencionar: los caminos mineros sin mantención, stock piles en abandono, desmontes sin supervisión y tranques de relaves sin una cubierta adecuada.

### 5.1.1. Relaves

Definidos por la ley n°20.551, específicamente en el decreto supremo de la república número 248, los tranques de relave son «Toda obra estructurada en forma segura para contener los relaves provenientes de una planta de concentración húmeda de especies minerales junto con sus obras anexas». Su función principal es la de servir como depósito, generalmente definitivo, de los materiales sólidos provenientes del relave transportado desde la planta, permitiendo así la recuperación en gran medida del agua que transporta dichos sólidos [9] y su composición está asociada exclusivamente al tipo de roca (mena) extraída al inicio de la cadena de valor y de los procesos por los cuales deba pasar el mineral antes de ser un producto vendible.

Los relaves pueden clasificarse en función de su tamaño y su porcentaje de humedad por lo que hablar de un tranque de relave, obra ingenieril de gran extensión y con alta presencia de agua, no es lo mismo que hablar de pastas de relave, pequeñas a medianas estructuras de bajo porcentaje de humedad y dispuestos en zonas más pequeñas. La proporción, según su tipo, de los 760 relaves catastrado queda resumida de la siguiente manera



**Ilustración 5.1: Catastro de relaves según su clasificación, fuente: propia/SERNAGEOMIN**

Como bien presenta la ilustración 2, el 81% de los relaves en el territorio nacional se encuentran dispuestos en forma de tranques [5]; La gravedad de esta cifra reside en los riesgos que estos presentan y las consecuencias que pudieran provocarse por potenciales fallas y colapsos de este tipo de mega estructuras.

### 5.1.2. Riesgos físicos

Destacan como riesgos físicos aquellos peligros asociados a la estabilidad de los muros de contención de tranques de relaves. Su ocurrencia está generalmente asociada a eventos de **licuefacción**, fenómeno que produce que un sólido se comporte como un fluido. De esta forma se posibilita el escape de grandes masas de relaves capaces de avanzar grandes distancias a elevadas velocidades.

El comportamiento de las estructuras contenedoras (muros, taludes, entre otros) varía de acuerdo a las condiciones del lugar en que se disponen. Así, relaves en zonas secas presentan mayor estabilidad, favoreciendo la compactación de los polvos constituyentes a modo de bloques más sólidos durante el transcurso del tiempo. En contraparte los relaves dispuestos en zonas de mayor humedad resultan ser más propensas a colapsos de las estructuras contenedoras y es durante eventos de lluvia y movimientos sísmicos que las mayores catástrofes de este tipo suceden.

El estudio de SERNAGEOMIN-BGR [10] describe los escenarios de peligro más probables asociados a la disposición final de relaves mineros:

- a) Liberación violenta de relaves depositados en tranques y embalses con el potencial de afectar personas, medio ambiente o el desarrollo de actividades económicas. La peligrosidad radica en las altas cantidades de material que contienen y pueden liberar dichas obras.
- b) Falla en los taludes de relaves depositados en forma de torta u otro tipo de acopio de material sólido. La falla del talud puede producir desplazamientos de material nocivo capaz de cubrir terrenos aledaños.

### 5.1.3. Casos reales

Casos Reales de contaminación ambiental y catástrofes de este tipo son los suficientes, a continuación, se presentan de manera breve casos sucedidos en Chile a lo largo de la historia de la minería [11].

➤ **Chañaral (1938)**

En el momento que los relaves de Potrerillos, ubicados a 125 kilómetros al oriente de la ciudad de Chañaral (Región de Coquimbo) alcanzaron su máxima capacidad se decidió vaciar ese caudal hacia el mar utilizando el cauce del río salado con desembocadura en la mencionada ciudad. En el transcurso de 52 años fueron arrojados al mar más de 320 millones de toneladas de sólidos residuales tóxicos y 850 millones de toneladas de aguas servidas producto del proceso industrial de la GMC, generando así una playa artificial de 6 km de extensión, playa con claros signos de disposición de material nocivo contiguos al poblado, quienes debieron reubicar su zona portuaria debido a esto.

➤ **El Cobre (1965)**

Relave perteneciente a la mina “El Soldado” (hoy propiedad de Anglo American) ubicado en la comuna de Nogales, región de Valparaíso. Cede completamente un 28 de marzo a causa del terremoto de 1965 en Calera, misma región, generando una avalancha de residuos tóxicos cubriendo casi en su totalidad el homónimo poblado de “El Cobre”. Doscientas personas murieron y cientos de kilómetros de terreno fueron contaminados producto de las nulas normas y las medidas de seguridad existentes o tomadas al momento de la construcción del tranque. Hoy, parte del antiguo relave ahora llamado “El torito” está en vías de ampliación generando preocupación en la localidad de el melón, sector contiguo a la pared Sur-Oeste del tranque.

➤ **Andacollo (2009)**

Comuna minera localizada al Sur-Este de la ciudad de Coquimbo, en 2009 fue declarada como **«Zona Saturada de contaminación por material particulado MP10»** [12]. Sumado a esto existen alrededor de 18 relaves en las proximidades (considerando un radio de 2 km) del poblado compuesto por más de 10.000 personas quienes han visto reducida drásticamente su calidad de vida, en cuanto a enfermedades de tipo respiratorio y oncológico. En 2015, el ministerio de medio ambiente publicó un plan de descontaminación [13] para remediar y mejorar su situación, que hasta el presente año (2019) aun aqueja a la población de Andacollo.

➤ **Las Palmas (2010)**

En Penciahue, comuna al oeste de Talca, región del Maule, se ubica la abandonada mina “Las Palmas”. Contigua a esta y disfrazada de cerro se encontraba una piscina de tierra, piedras y cianuro, dentro de otros líquidos tóxicos. Por efecto del mega terremoto de 2010 (8.8° en la escala de

Richter) esta estructura colapsó sepultando una casa junto con sus cuatro integrantes, además de contaminar gran parte de la comuna debido al desplazamiento del material nocivo hacia sectores colindantes. El año 2016, el 2° tribunal ambiental condenó a SCM Tambillo, empresa responsable de la faena, por daño ambiental [14]. Al presente año, la implementación del **Programa de Reparación Ambiental (PRA)** ordenada a la empresa mencionada no ha sido completado en su totalidad.

➤ **Norte Chico (2015)**

La ocurrencia de inéditas y masivas lluvias en zonas cordilleranas del Norte Chico del país, región natural comprendida entre los ríos Copiapó (al norte) y Aconcagua (al sur), provocaron aludes que destruyeron todo a su paso, incluyendo obras mineras como relaves los cuales mezclados con los lodos del mismo alud se dispersaron a lo largo del territorio, arrasando con poblados como el anteriormente mencionado Chañaral. La dispersión de estos contaminantes provocó, además de la muerte de 31 personas y la desaparición de otras 49, severas emergencias ambientales a lo largo de las regiones de Antofagasta y Atacama debido a la potencial exposición a altos niveles de: Plomo, Cadmio, Cobre, Hierro, Mercurio, entre otros [15].

➤ **El Mauro (actualidad)**

El tranque “El Mauro” pertenece a la faena minera a cielo abierto de “Los Pelambres” ubicado en la comuna de Salamanca, Región de Coquimbo, es el tranque minero más grande de Latinoamérica y el 3° a nivel mundial. Con sus 2.000 millones de toneladas de relaves dispuestos a lo largo de seis kilómetros, en su mayor extensión, y con potencia máxima de hasta 250 metros es considerado como una potencial amenaza por la población que habita en la localidad de Caimanes, a unos 8 km al oeste del tranque. Y no es solo el riesgo de colapso, sino que también se acusa a la minera AMSA de haber sepultado una de las mayores reservas que fluían hacia los acuíferos subterráneos que sustentan los pueblos caudales abajo, es por ello que en 2015 la comunidad de Caimanes acudió a la Corte de Apelaciones de la Serena para interponer una demanda por «Obra Juiciosa». Pese a las instancias de diálogo y a los juicios llevados a cabo, la denuncia fue desestimada tanto por la corte de apelaciones como por la corte suprema a posteriori [16]. A la fecha, el tranque sigue en funcionamiento avalado por los juicios mencionados, sin embargo, los habitantes siguen buscando recursos para detener su funcionamiento debido al temor de que suceda alguna tragedia como las expuestas previamente en este trabajo.

#### 5.1.4. Antecedentes

A pesar de los problemas que sugiere la disposición de desechos mineros, en la actualidad se ha demostrado que pueden ser una gran fuente de recursos minerales para nuevos proyectos. Diversos estudios comprueban la viabilidad de la reutilización de estos, transformando los pasivos en activos económicos. Si bien es una idea relativamente nueva, el reciclaje de relaves se ha hecho bastante frecuente en algunos sectores económicos, principalmente en el de construcción.

Muchos estudios e investigaciones se han enfocado en darle nuevos usos a estos pasivos ambientales mineros. En términos de energía, Codelco en 2016, adjudicó al consorcio chileno KDM-INEER la licitación para la construcción de una pequeña central hidroeléctrica en la división El Teniente, región de O'Higgins. El proyecto busca aprovechar el flujo del relave del sistema de cascada responsable del transporte de los mismos desde la planta concentradora hasta el tranque ubicado a 84 km. Se espera que la denominada "mini central" aporte al sistema interconectado central (SIC) alrededor de 20.000 MWh/año [17].

En cuanto a su uso en construcción los estudios son variados: Generación de agregados de construcción como ladrillos y baldosas, o como relleno hidráulico en labores subterráneas son descritos por Romero A. y Flores S. [18, 19]. Onuagoluchi [20] a su vez, propone el uso de relaves provenientes de menas de cobre como aditivos en morteros de cemento, medida que presentó mejoras en cuanto a resistencia a la compresión y abrasión del material.

Con la generación de morteros a partir de relaves descrito por Castañeda [21], se ha creado un nuevo producto que ofrece mejoras tanto de soporte de estructuras en labores de extracción de tipo *Cut & Fill*, como de eficiencia de material permitiendo el ahorro de hasta un 20% de cemento.

Debido a su composición los relaves provenientes de menas de hierro han sido utilizados para la fabricación de cerámicas rojas [22]. Su uso en la producción de pigmentos para pinturas sustentables, aseguran la reducción de costos con respecto a las pinturas comerciales convencionales al mismo tiempo que se mitigan los impactos medioambientales y riesgos asociados a los tranques de relaves en Brasil [23]. La producción de nuevos ECC, concretos flexibles capaces de sustituir aquellos compuestos por fibras (ej. shotcrete) a un menor costo y de manera ecológica ha sido expuesta por Xiaoyan en su estudio [24].

## 5.2. Estimación de recursos minerales

La geoestadística es una rama de la estadística y se define como la aplicación de modelos probabilísticos al análisis de datos distribuidos en un determinado espacio. Entrega y utiliza herramientas para la cuantificación de la correlación espacial entre datos, en otras palabras, utiliza la similitud de determinados datos para predecir otros desconocidos, es por esto que esta rama de las ciencias estadísticas posee muchas aplicaciones. En el presente trabajo se abordó su uso en la evaluación y estimación de recursos naturales, en particular, de recursos geológicos y potenciales reservas minerales correspondientes a una estructura de relave en particular. La evaluación de recursos es clave, puesto que es la base del desarrollo que sustenta un proyecto minero. Al mismo tiempo, este proceso considera una secuencia de sub etapas que permiten identificar la información disponible y generar predicciones en base a esta, determinando los riesgos a los que se expone el proyecto en etapas posteriores.

De manera bastante general, el proceso de evaluación se inicia con la captura de datos mediante muestreos, datos que deben ser analizados estadística y espacialmente mediante gráficos de dispersión, eliminación de datos aberrantes y erróneos, entre otros procesos. El resultado de estos procesos da génesis a la **información disponible**. El modelamiento e interpretación geológica permite construir una idea general de una zona de interés o estudio (un yacimiento minero, por ejemplo) es por ello que para una mayor y mejor caracterización se necesita un proceso previo de interpolación de atributos o características objetivas del modelo, siempre considerando la respectiva cuantificación de la incertidumbre que este genera.

Cabe señalar que siempre existe un grado de incertidumbre en un modelo numérico debido a la estimación de los datos no muestreados, y es la magnitud de estos quien define la precisión y efectividad de dicho modelo. Este proceso de predicción dispone y presenta bases sobre las cuales se pueda generar una primera evaluación técnica-económica que permita esclarecer los riesgos a los que se expone el proyecto y así asegurar su factibilidad.

Previo al uso de herramientas geoestadísticas y a la explicación más en detalle de estas, se deben manejar diversos conceptos fundamentales para cualquier estudio de modelamiento, estimación o simulación estadística. Otros conceptos mineros asociados son también expuestos a continuación [25, 26, 27].



- ❖ **Fenómeno regionalizado:** Es aquel fenómeno que se desarrolla en un espacio físico generalmente delimitado. Fenómenos químicos y físicos que generen mineralización o transporte de partículas contaminantes en un determinado lugar son ejemplos típicos de fenómenos regionalizados en minería.
- ❖ **Variable regionalizada:** Valor de alguna de las características asociadas a un fenómeno regionalizado. Se caracteriza principalmente por su naturaleza: Continua o Discreta; y por su soporte, es decir, el volumen en que se mide y que determina los valores desconocidos de la misma en otro punto de la región. En minería el concepto de variable regionalizada se asocia a leyes de mineral, tipos de roca o alteración, entre otros.
- ❖ **Población:** Corresponde a la colección de un número finito de mediciones de datos de interés.
- ❖ **Muestra:** Es un subconjunto representativo seleccionado a partir de una población. Una buena muestra debe ser altamente representativa, esto quiere decir, debe reflejar las características esenciales de dicha población. Sondajes de exploración, detritos de perforaciones son ejemplos típicos de muestras geológicas.
- ❖ **Ley de corte/Umbra:** Se denomina “ley de corte” al valor límite que se utiliza para clasificar o diferenciar, en minería, material con valor económico del que no.
- ❖ **Continuidad:** Generalmente se denomina como Continuidad espacial y se refiere a la persistencia en la distribución de una variable regionalizada en el espacio físico de muestreo.
- ❖ **Variable aleatoria:** Variable cuyo valor está dado por una cierta distribución de probabilidad definida. Estas permiten modelar la incertidumbre respecto un valor no muestreado que se quiera determinar, puede ser de tipo continua (existen valores intermedios entre variables) o categóricas (variables pertenecen a clases definidas y diferenciadas).
- ❖ **Función aleatoria:** Está constituida por un conjunto de variables aleatorias bajo un mismo dominio real.
- ❖ **Modelamiento geológico:** Es la representación de la distribución de atributos geológicos basada en información fragmentaria/diseminada. Es la base para la estimación de recursos y forma parte integral de la evaluación de yacimientos y busca generar un modelo predictivo del comportamiento de las variables de interés económico en la labor minera.

- ❖ **Categorización:** El nivel de conocimiento y confianza de las estimaciones crece con la exploración del depósito permitiendo así clasificarlo como Recursos o Reservas. De estas derivan sub clasificaciones según la confianza/seguridad que se tenga de ellas.
- ❖ **Recurso mineral:** Concentración u ocurrencia de material de interés económico intrínseco en o sobre la corteza de la tierra en forma o cantidad en que haya probabilidades razonables de una extracción económica. Estos se clasifican, en orden de confianza ascendente, en: inferidos, indicados y medidos.
- ❖ **Reserva mineral:** Parte económicamente explotable de un recurso minero, incluye la dilución de materiales y tolerancias por pérdidas que puedan producirse en el proceso de extracción. Apoyado por la realización de estudios de factibilidad y el análisis de factores multidisciplinarios (metalúrgicos, económicos, legales, ambientales, sociales, políticos, etc.) asociados al negocio minero. Las reservas se clasifican, en orden creciente de confianza, en probables y probadas.

### 5.2.1. Variografía.

La estimación y simulación geoestadística de recursos, tal y como se mencionó anteriormente, se basa en identificar y aprovechar la distribución espacial de una variable de interés. Dicha continuidad espacial puede ser medida mediante el uso de la Covarianza, un Correlograma o un **Variograma**. La idea básica es definir y encontrar la relación que existe entre pares de datos separados a una distancia “h”, donde el procedimiento se repite para una cantidad de finita de distancias h.

**Nubes de correlación diferida.** Es una herramienta estadística que indica que tan semejantes son 2 datos en función de la distancia que las separa, es de gran utilidad para la generación de otras herramientas geoestadísticas, como las que se mencionan en el párrafo siguiente. El coeficiente de correlación de la nube permite hacer una primera lectura de la relación espacial de la variable mediante la generación de un correlograma experimental. Su uso no será abarcado mayormente en el presente estudio.

**Variograma experimental.** El variograma experimental es un estimador del variograma teórico, y mide la desviación cuadrática promedio entre dos datos en función de su separación espacial. La generación de este variograma permite determinar qué tan matemáticamente continua es una variable regionalizable por medio del análisis de su estructura característica, permitiendo

identificar cuán rápido y a que distancias la variable se desagrupa y pierde relación con su vecindad, definiendo la zona de influencia de un dato, denominado **alcance**. Denotando la ubicación de un dato como  $\{x_\alpha, \alpha = 1, \dots, n\}$  y la variable regionalizada como  $\{z(x), x \in D\}$ , un variograma experimental se define de la siguiente manera

$$Y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u_i) - z(u_i + h)]^2 \quad (1)$$

[\*] Donde  $N(h) = | \{(\alpha, \beta) \text{ tal que } x_\alpha - x_\beta = h\} |$

Cabe destacar que el variograma experimental se obtiene al visualizar el momento de **inercia** de las nubes de correlación diferida (distancia promedio a la diagonal principal) en función de la distancia de separación “h”. El análisis de su comportamiento cercano al punto de origen (con h tendiendo a cero y varianzas bajas) indica la similitud entre datos muy cercanos reflejando de esta manera que tan continua o regular es una variable a pequeña escala.

Algunas de las propiedades del variograma [26]:

- Es una función simétrica,  $[Y(h) = Y(-h)]$ .
- Es nulo en el origen,  $Y(0) = 0$ .
- Es una función positiva,  $Y(h) \geq 0$
- Son operables,  $Y_1(h_1) = Y_2(h_2) + \dots + Y_n(h_n)$

El variograma experimental se calcula mediante el uso de varios parámetros que deben ser especificados: La dirección de interés representada por un azimut o un *dip (manteo)*; La distancia de interés denominada paso o *lag* que define la distancia elemental en la que se divide el variograma; Las tolerancias tanto para la dirección (tolerancia angular y ancho de banda) como para las distancias; y finalmente el número de pasos que limitarán el estudio en cuanto a la distancia “h” máxima del gráfico del variograma. Definidos estos parámetros el cálculo de un variograma se resume a un algoritmo general básico [27]:

- 1) Escoger el azimut/inclinación de direcciones según interpretación geológica o bien la orientación de estructuras perpendiculares o en el mismo plano de estas.
- 2) Escoger cantidad de pasos y la distancia entre estos, dependiendo de la orientación es frecuente utilizar para el caso horizontal la distancia entre sondajes o información de la perforación (compasitos o espaciamiento de muestras) para la vertical.

- 3) Definir las tolerancias, por lo general para la angular se escoge entre 15 y 22°, y para el ancho de banda se utiliza la mitad de la distancia de paso.

A la hora de definir estos parámetros se debe tener en cuenta una serie de implicancias. Respecto los pasos, pasos pequeños producen un variograma errático (caso A), se genera “ruido” en los datos debido fluctuaciones aleatorias, en contraparte, pasos grandes (caso D) generan suavizamiento de la función lo que se traduce en subestimación de la variabilidad, es por ello que la elección del paso se debe asociar al espaciamiento de los datos.

En cuanto a la tolerancia, sin ella sólo se consideran los datos que distan exactamente a  $h$  disminuyendo la robustez del modelo por falta de datos (caso E), en cambio con altas tolerancias se genera un exceso de suavizamiento de la función (caso H), repitiendo el caso D presentado en la ilustración 4.

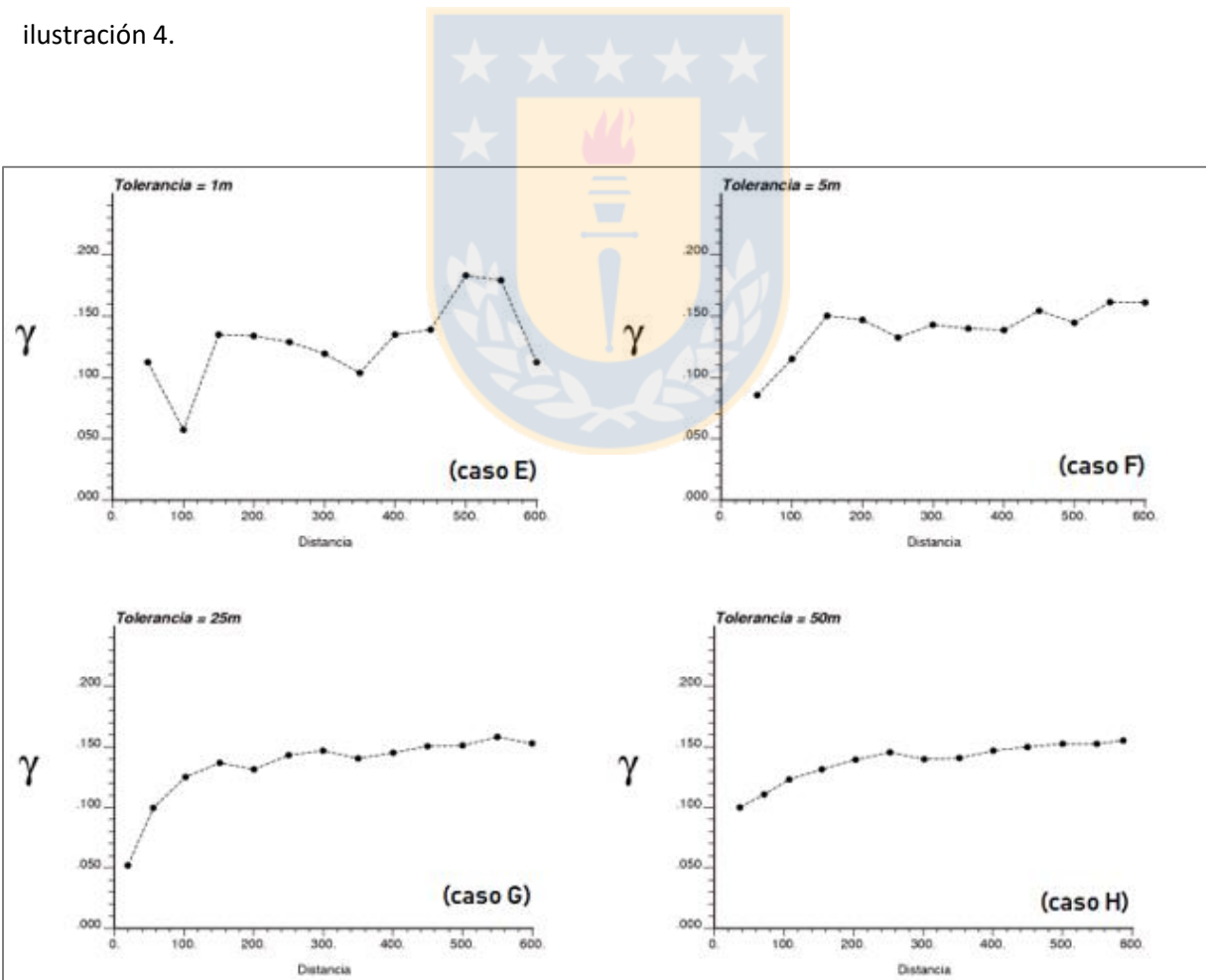


Ilustración 5.2: Visualización del efecto Tolerancia

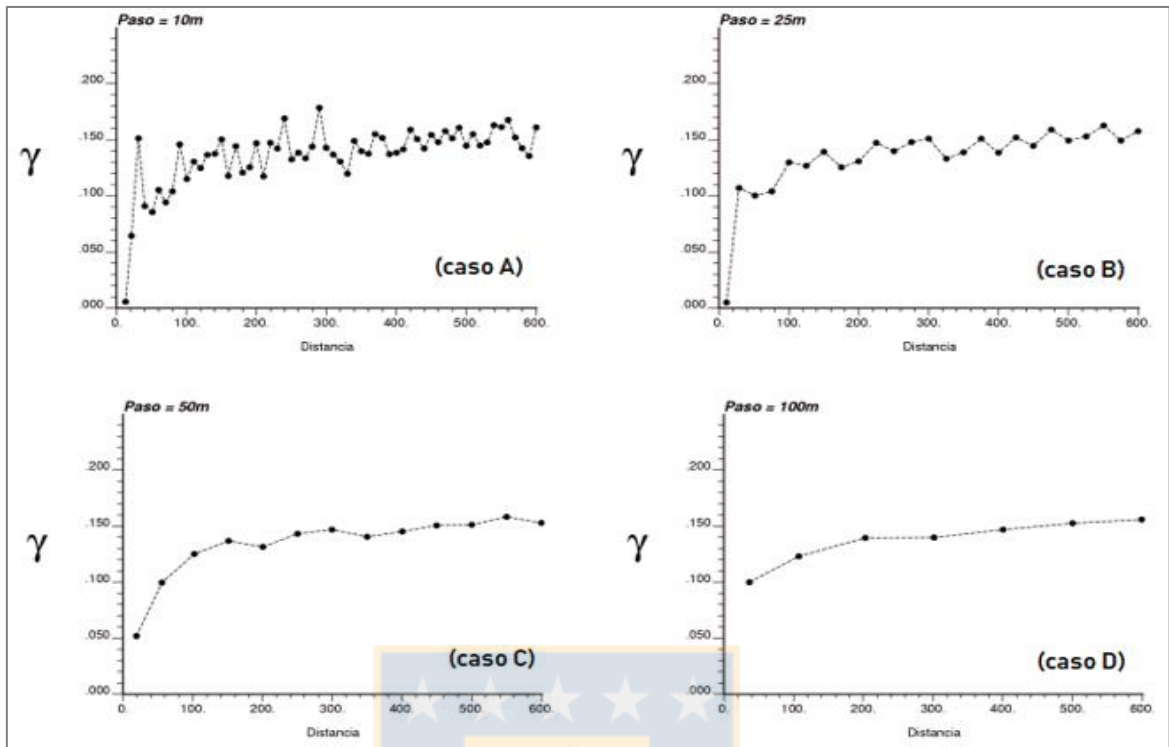


Ilustración 5.4: Visualización del efecto tamaño de paso

La interpretación del variograma consta en establecer los valores de los elementos básicos de este. La **meseta** corresponde al punto donde el variograma se estabiliza y se relaciona a la varianza máxima; El **alcance** corresponde a la distancia "h" en la cual se alcanza la meseta definiendo la zona de influencia de la variable, es decir, donde se pierde la variabilidad del atributo de interés.

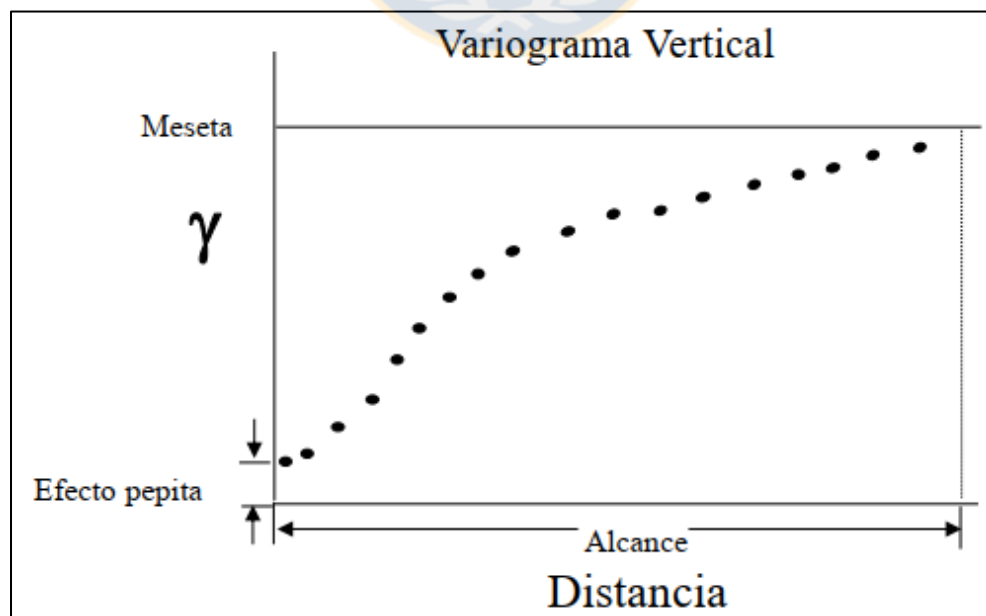


Ilustración 5.3: Visualización de elementos básicos de un variograma

**Efecto pepita (*nugget effect*).** Se denomina al aparente espacio entre el primer dato del variograma y el origen, aparente ya que solo es un efecto visual que desaparece cuando se disminuye la escala de visualización (distancias pequeñas). Este refleja una alta variabilidad (nula correlación) entre datos muy cercanos, y recibe su nombre por el fenómeno natural de formación de pepitas de oro, donde se da una alta concentración de oro en pequeños puntos del espacio versus los empobrecidos alrededores cercanos. Si bien es producto de formación natural de depósitos, por ejemplo, en casos de presencia de microestructuras en la mineralización, también es producto de errores de medición (segregación, delimitación, extracción, preparación y análisis químico, entre otros tipos) y problemas de soporte a nivel de tamaño y de distribución no equitativa de un muestreo no aleatorio, sino que preferencial.

Debido a los procesos de formación natural de depósitos la distribución de la variable de interés puede presentar diversos comportamientos (alcances) en diferentes direcciones generando una dirección preferencial por diversas causas (gradiente de temperatura, pH de suelo, tipo de roca, etc.), este comportamiento se denomina **anisotropía**. En contraparte si la variable presenta un comportamiento regular (no preferencial) en todas las direcciones se considera una **isotropía**.

Una herramienta útil para identificar estos comportamientos direccionales son los mapas variográficos, que se obtienen mediante el cálculo de variogramas en múltiples direcciones generando así un “cubo variográficos” que admite análisis en todos los planos: XY, XZ, YZ. De esta manera y con los planos se puede analizar si existe anisotropía y sus direcciones preferenciales en caso de existir.

Está la posibilidad de definir modelos mucho más complejos mediante la suma de dos o más variogramas elementales (modelos esférico, exponencial, gaussiano, seno cardinal, entre otros). Estos se denominan modelos anidados y permiten modelar cambios en pendientes dentro de un mismo variograma.

Algunas consideraciones prácticas en relación a los variogramas que deben ser mencionadas, respecto las anisotropías, generalmente se modelan dos o hasta tres direcciones preferenciales ortogonales entre sí de manera que se genere una elipse (2D) o elipsoide (3D) de búsqueda que mejor represente el mapa variográfico. Para distancias grandes (superiores a la mitad del diámetro del campo) el variograma experimental es poco confiable y muy fluctuante, por otro lado, para una escala menor de distancias se puede obtener información respecto la continuidad a pequeña escala, relevante para la interpolación de datos cercanos.

## 5.2.2. Predicción

El muestreo corresponde a un proceso parcial e incompleto, puesto que indica solo lo que sucede en unos pocos sitios del espacio objetivo total, por lo que si se desea caracterizar el espacio por completo surge la necesidad de predecir el valor del atributo en los sitios no muestreados.

El aumento de recopilación de información implica aumento en costos asociados por lo que a fin de lograr la predicción sin recurrir a gastos extra (tiempo, dinero, personal, etc.) es que existen diversos tipos y formas de aproximar los valores mediante “estimadores”. Estos estimadores utilizados, como generalidad, cuentan con ciertas características/propiedades [27]

- **Lineal:** El valor estimado (desconocido) corresponde a una combinación lineal de los datos disponibles (muestras) en función de la cercanía entre dato conocido y por estimar. Dicha cercanía se define como **Vecindad**.
- **Insesgado:** En promedio, el estimador arroja el valor correcto sin sesgo (error) sistemático, pero si con un error asociado a la predicción. Este error se debe mantener mínimo, cercano a cero, para cada proceso de estimación.
- **Óptimo:** El estimador será tal que minimizará la varianza del error de la estimación, si los errores que se generan son similares permiten una estimación con buena precisión. La idea básica general del uso de estimadores es predecir el valor de una variable regionalizada en una posición “**u**” donde se desconoce el valor real, matemáticamente el estimador se puede definir mediante la siguiente ecuación:

$$Z^*(u) = a + \sum_{i=1}^n \lambda_i Z(u_i), \quad i \in \{1, 2, \dots, n\} \quad (2)$$

Donde **Z\*(u)** es el valor estimado para la variable regionalizada en la posición “**u**”. El valor/operador “**a**” corresponde a un coeficiente aditivo y  $\lambda_i$  es el ponderador del dato conocido **Z(u<sub>i</sub>)** en la(s) posición(es) **u<sub>i</sub>**.

El argumento principal de la ecuación anterior se presenta como la combinación lineal de los datos conocidos, cada uno asociado a un respectivo ponderador ( $\lambda$ ) que define la importancia/peso del dato respectivo al momento de realizar la estimación. Este ponderador o Estimador Lineal Ponderado (E.L.P.) considera diversos **factores** al momento de ser definido [24]

- Cercanía a la posición que se está estimando

- Redundancia entre valores
- Continuidad o variabilidad espacial
- Existencia, o no, de una dirección preferencial (**Anisotropía**)

Existen diversos tipos de E.L.P. los cuales se definen de distinta manera según sean los factores que se desean involucrar. A continuación, se mencionan algunos de estos ponderadores y las formas que adquieren los estimadores asociados [25, 26]

#### Vecino más cercano

Este atribuye la totalidad de la ponderación (100%) al dato más cercano al punto a estimar, denominado también como “estimador de polígono de influencia”.

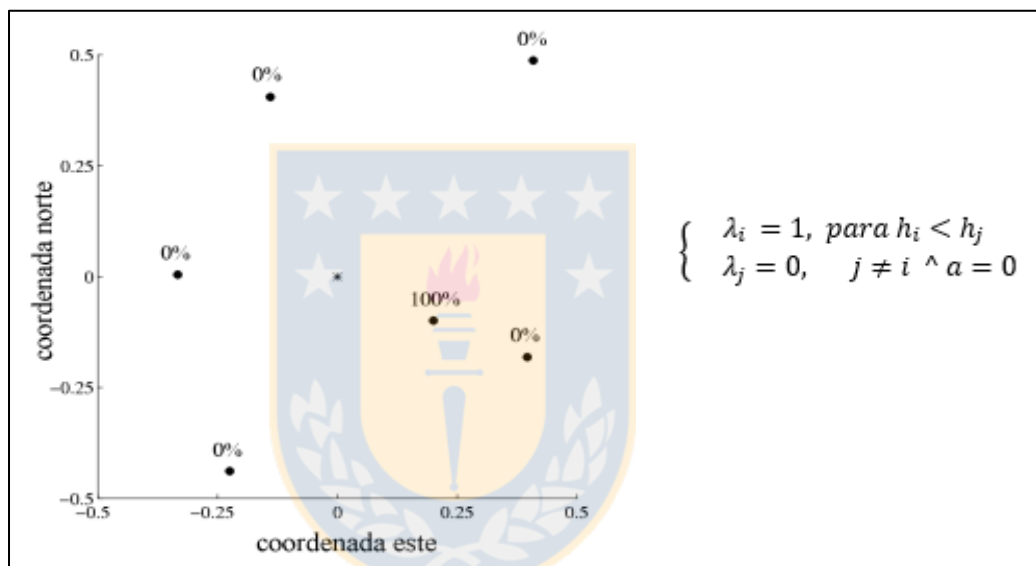


Ilustración 5.5: Método de Vecino más cercano

#### Inverso de la distancia

Atribuye a cada dato una ponderación proporcional a una potencia  $w_i$  y al inverso de la distancia al sitio a estimar tal que:

$$\lambda_i = \frac{\frac{1}{c + d^{w_i}}}{\sum_{i=1}^n \frac{1}{c + d^{w_i}}} \quad \wedge \quad a = 0, \quad \text{con } w_i \in \{1,2,3\} \quad (2a)$$

De manera gráfica se observa el siguiente ejemplo para distintos valores del parámetro  $w$ , lo que invita a reflexionar que el comportamiento de las ponderaciones de los datos cuando  $w$  tiende



a infinito produce el caso del vecino más cercano otorgándole todo el peso de la ponderación al dato conocido más cercano al sitio a estimar.

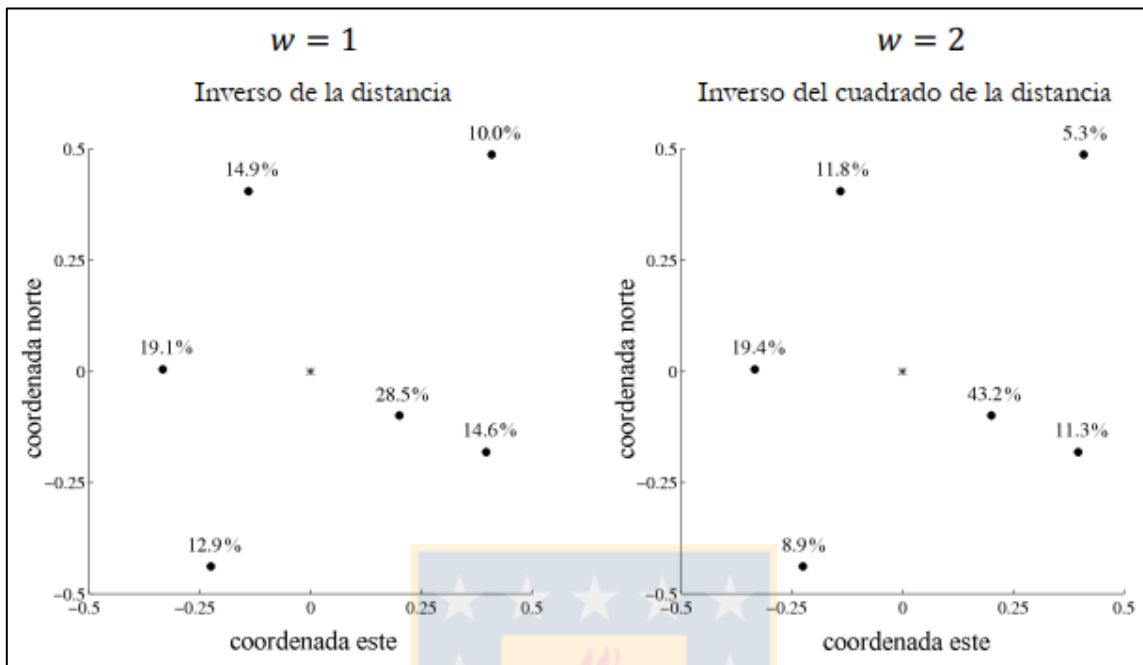


Ilustración 5.6: Visualización del ponderador "inverso de la distancia"

Los estimadores dados por los ponderadores mencionados son sencillos de aplicar, pero la ponderación propuesta depende únicamente del factor "distancia entre dato conocido y desconocido" por lo que la estimación producto de su uso se puede calificar como poco precisa debido a la poca información involucrada en el modelo (**efecto soporte**). Como respuesta a ello es que surgen métodos nuevos que incorporan y utilizan múltiples criterios como la redundancia de datos, la continuidad espacial y comportamientos preferenciales de manera simultánea al momento de hacer la estimación, logrando aumentar el soporte y por consiguiente la fidelidad y la precisión de la predicción.

Interpolador de Kriging.

Cuantificar la precisión de un interpolador se basa en el uso de un modelo probabilístico, la idea es definir el error como una variable aleatoria a la cual se le imponen ciertas condiciones: **Esperanza nula y Varianza pequeña**.

El Kriging es un estimador basado en un modelo de función aleatoria en donde la ponderación de los datos ya no depende solo de la distancia al sitio desconocido, sino que también de la redundancia que pueda existir entre datos por posibles agrupamientos y de la continuidad espacial del atributo objetivo, parámetros que se extraen del variograma de la variable.

Todo lo anterior permite al interpolador privilegiar los datos cercanos, repartir la ponderación en presencia del efecto pepita y, en caso de anisotropía, privilegiar aquellos datos en la dirección preferente [27]

El sistema completo se construye en base/alrededor de las 3 restricciones principales

- Linealidad: Sea  $z(x)$  la variable regionalizada en estudio,  $Z(x)$  la función aleatoria asociada,  $\{x_i, i = 1, \dots, n\}$  los sitios con datos y  $x_0$  el sitio a estimar, se define el estimador  $Z^*(x_0)$  como la siguiente combinación lineal

$$Z^*(x_0) = a + \sum_{i=1}^n \lambda_i * Z(x_i) \quad (3)$$

- Insesgo: En el modelo probabilístico el error debe tener esperanza nula, igual a cero. Por lo que se dice que el estimador no subestima ni a sobreestima el valor real.

$$E[Z^*(x_0) - Z(x_0)] = 0 \quad (4)$$

- Optimalidad: Se busca minimizar la varianza del error cometido, también llamada varianza de Kriging, con el fin de obtener la máxima precisión en el cálculo de estimación.

$$\min. \sigma_k^2(x_0) = Var[Z^*(x_0) - Z(x_0)] \quad (5)$$

Bajo este conjunto de restricciones la interpolación por Kriging es considerada como el mejor estimador lineal insesgado o Best Linear Unbiased Estimator (**BLUE**).

En el empleo del Kriging existe la posibilidad de utilizar todos los datos existentes a modo de **Vecindad Única**, o solo una parte de estos denominada **Vecindad Móvil**. El termino Vecindad se refiere a la zona del espacio que posee los datos a considerar en el proceso de estimación/interpolación. Considerar una gran vecindad solo genera el aumento de los tiempos de cálculo ya que los datos extremadamente lejanos deberán ser considerados y recibirán una ponderación casi insignificante, traduciéndose una pérdida de precisión. Es por esta razón que especificar y acotar la forma y tamaño de esta vecindad es crucial para la realización de un proceso óptimo y preciso.

Para establecer un soporte bien definido se debe poner atención en la malla de muestreo y en los parámetros que entrega el variograma respectivo. Lo anterior debe hacerse en consideración del tamaño final de la vecindad, en otras palabras, un aumento del tamaño implica aumento de la información disponible lo que se traduce a una mayor precisión del modelo, hasta cierto punto. Por otro lado, la disminución de tamaño debe ser considerada en casos de que presenten datos muy lejanos y un variograma muy errático.

La forma de la vecindad debe estar definida considerando una potencial anisotropía geométrica, para ello se suele definir la forma como una elipse o un elipsoide de búsqueda para los casos 2D y 3D respectivamente. Definida la forma, esta se divide en sectores angulares (cuadrantes u octantes según corresponda) para seleccionar una determinada cantidad mínima y/o máxima de datos desde cada sector.

Para determinar el plan de Kriging se procede a métodos de validación, como la validación cruzada o Jack-knife, donde mediante el testeado de varios planes se conservan aquellos donde los resultados sean los más satisfactorios. Este procedimiento permite validar tanto el variograma teórico como la vecindad elegida mediante la comparación de los resultados obtenidos [27]. Para la validación se estima sucesivamente cada dato conocido, mediante Kriging. Considerando los datos conocidos restantes, se calcula el error de estimación cometido en cada iteración. De esta manera, se puede estudiar la calidad de tales errores mediante herramientas estadísticas y gráficas para definir la calidad del procedimiento.

La calidad se asegura a medida que se corroboren las condiciones ya mencionadas: la media de los errores debe tender a cero (insesgo), la varianza de los errores debe mantenerse lo más pequeña posible (precisión), las varianzas de los errores estandarizados deben tender a uno y la nube de dispersión entre la variable estimada y la real debe ser lo más cercana a la diagonal (insesgo condicional).

Existen muchos tipos de Kriging: Simple, Ordinario, indicador, con deriva (Universal, Trigonométrico, deriva exterior, etc.), no lineal (Log normal, Disyuntivo, multi-Gaussiano), multivariable o Co-Kriging. En el presente trabajo solo serán abordados los primeros dos tipos mencionados.

## Kriging Simple

Se comienza con las siguientes hipótesis centrales:

- La variable regionalizada tiene una media "m" conocida
- El variograma  $\gamma(h)$  es conocido y presenta meseta

$$\gamma(\infty) = \sigma^2 \rightarrow Cov = C(h) = \sigma^2 - \gamma(h)$$

En base a lo anterior se verifican las tres restricciones anteriormente mencionadas de linealidad insesgo y optimalidad que en conjunto definen el estimador (conjunto de ecuaciones)

### i. Linealidad:

Presenta el estimador como combinación lineal de datos existentes de igual manera que antes en la ecuación n°3

$$Z^*(u) = a + \sum_{i=1}^n \lambda_i * Z(u_i) \quad (3)$$

### ii. Insesgo:

A partir de la ecuación de esperanza (ecuación n°4) se tiene lo siguiente

$$E[Z^*(u) - Z(u)] = a + \sum_{i=1}^n \lambda_i E\{Z(u_i)\} - E\{Z(u)\} = 0$$

$$\text{de lo anterior, } a + \sum_{i=1}^n \lambda_i * m - m = 0 \rightarrow a = \left\{1 - \sum_{i=1}^n \lambda_i\right\} * m \quad (4a)$$

### iii. Optimalidad:

La varianza del error de estimación se expresa en función de la covarianza, mediante la derivación parcial de la ecuación n°5 e igualándola a cero se obtiene:

$$\sum_{j=1}^n \lambda_j * Cov(u_i - u_j) = Cov(u - u_i), \quad i = 1, \dots, n \quad (5a)$$

Este sistema de **n** ecuaciones con **n** ponderadores desconocidos es el sistema de Kriging simple (KS), puede expresarse de manera matricial donde la primera matriz de covarianzas mide la

redundancia entre datos y la matriz resultante define las correlaciones entre los datos y el valor estimado.

Finalmente, y juntando las ecuaciones resultantes de las restricciones impuestas, el estimador final para la interpolación mediante Kriging simple es el siguiente:

$$Z^*(u) = \sum_{i=1}^n \lambda_i * Z(u_i) + \{1 - \sum_{i=1}^n \lambda_i\} * m \quad (6a)$$

Dentro del estimador, la media aparece asociada a un factor que corresponde al complemento de la ponderación acumulada de los datos  $\{1 - \sum_{i=1}^n \lambda_i\}$ . Esto significa que mientras más lejos los datos del sitio  $u$ , menores serán sus ponderaciones y mayor será la ponderación de la media. Se puede establecer entonces que, en cierto modo, la media compensa la falta de información asociada a los datos adquiriendo relevancia en la ponderación cuando el proceso así lo requiera.

Se define también la varianza de Kriging como la varianza del error en la interpolación realizada, que una vez minimizada se representa de la siguiente forma

$$\sigma_{KS}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i * Cov(u - u_i) \quad (7a)$$

Kriging Ordinario.

De la misma manera que el Kriging simple, parte de las siguientes hipótesis centrales:

- i) La variable regionalizada posee una media “m” desconocida
- ii) Posee un variograma  $\gamma(h)$  conocido el cual puede tener, o no, meseta.

El desconocimiento de la media le permite al estimado generalizar a casos donde la media no es necesariamente constante en el espacio. De la misma manera que para el Kriging simple se deben verificar y desarrollar las restricciones de linealidad, insesgo y optimalidad.

i. Linealidad:

El principio permanece inmutable, el estimador se representa como una combinación lineal de los datos existentes con los parámetros respectivos (Ecuación n°3)

$$Z^*(u) = a + \sum_{i=1}^n \lambda_i * Z(u_i) \quad (3)$$

ii. Insesgo:

El valor esperado del error de estimación sigue el mismo desglose de la ecuación n°4, con distintas implicancias a las de la ecuación 4a, de la siguiente manera

$$E[Z^*(u) - Z(u)] = a + \sum_{i=1}^n \lambda_i E\{Z(u_i)\} - E\{Z(u)\} = 0$$

de lo anterior,  $a + \sum_{i=1}^n \lambda_i * m - m = 0 \rightarrow a = 0 \wedge \sum_{i=1}^n \lambda_i = 1$  (4b)

Al ser “m” desconocida, para lograr obtener una esperanza igual a cero (minimizar el error) se deben establecer otras condiciones tales que el factor **a** sea cero a la vez que la sumatoria de ponderadores sea igual a uno.

iii. Optimalidad:

Al igual que para el Kriging simple se expresa la varianza en función de la covarianza (también puede expresarse en función del variograma), sin embargo, al momento de generar la derivada para determinar los ponderadores óptimos se introduce un multiplicador de Lagrange  $\mu$  **(para trabajar con  $\gamma(h)$  el multiplicador de Lagrange ingresa con signo negativo)**

$$\sum_{j=1}^n \lambda_j * Cov(u_i - u_j) + \mu = Cov(u - u_i), \quad i = 1, \dots, n \quad (5b)$$

Verificadas las restricciones y armado el sistema de ecuaciones correspondiente se puede formular la ecuación final para el interpolador de Kriging ordinario.

$$Z^*(u) = \sum_{i=1}^n \lambda_i * Z(u_i) \quad (6b)$$

Así como también se define la varianza del error asociado (varianza de Kriging) en función tanto de la covarianza como del variograma como se muestra en las siguientes ecuaciones:

$$\sigma_{KO}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i * Cov(u - u_i) + \mu \quad (7b)$$

$$\sigma_{KO}^2 = \sum_{i=1}^n \lambda_i * \gamma(u - u_i) - \mu \quad (7c)$$

Ponderadores y varianzas de Kriging, como se mencionó en un principio, se determinan utilizando criterios y aspectos de tipo geométricos como la distancia entre punto desconocido y el dato existente, y la redundancia entre los mismos datos conocidos; y de tipo variográficos como la continuidad espacial, presencia de anisotropías, varianza y covarianza asociada convirtiéndolo en un proceso de gran precisión. Sin embargo, la varianza del Kriging no incorpora la información local, es decir, no mide todas las fuentes de incertidumbre ni tampoco toma en cuenta el denominado “efecto proporcional” [25].

Sumado a esto el método de Kriging tiende a subestimar las zonas de mayor valor, así como también, sobreestima aquellas de menor, por lo que el proceso se vuelve engorroso y problemático si se desea determinar el valor del atributo de acuerdo a algún cierto umbral como lo puede ser la ley de corte en un estudio minero, o la concentración máxima admisible de algún contaminante en el área medioambiental, para estos casos se propone como solución el uso de otros tipos de Kriging como el no lineal o bien el uso de herramientas de regresión.

### 5.2.3. Máquinas de Vector Soporte (SVM's)

El método de clasificación-regresión de máquinas de vector soporte (**Support Vector Machines**) se utiliza tanto para métodos de clasificación binaria como para problemas multivariados y de regresión. Los SVM hallan su fundamento en lo que se denomina como **Maximal Margin Classifier (MMC's)**, el cual se sustenta en el concepto de **hiperplanos**.

#### Hiperplano y MMC's

En un espacio p-dimensional, un hiperplano se define como un subespacio plano y afín de dimensiones p – 1, es decir, para un espacio tridimensional su hiperplano corresponde a un subespacio de dos dimensiones (plano convencional). Para dimensiones p > 3 no es intuitivo la visualización, pero el concepto se mantiene inmutable. De manera general el hiperplano de un espacio p-dimensional se define como  $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = 0$ .

Todos los puntos definidos por el vector  $X = x_1, \dots, x_p$  que cumplan con dicha ecuación se dice que pertenecen al hiperplano, por el contrario, si  $X$  no pertenece entonces  $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p < 0$  o  $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p > 0$ .

Ergo, es correcto decir que un hiperplano divide al espacio  $p$ -dimensional en dos mitades.

### Separación binaria con hiperplanos

Cuando se dispone de  $n$  observaciones, cada una con  $p$ -predictores cuya variable de respuesta posee dos niveles (+1 y -1) se puede emplear el uso de hiperplanos para construir un clasificador que permita predecir a que grupo pertenece cada observación, en función de sus predictores [28].

**Casos lineal y perfectamente separables.** Si la distribución espacial de los datos permite una separación lineal y perfecta de las clases existentes, entonces el hiperplano de separación asociado cumple de manera general que:

$$y_i (\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) > 0, \quad \text{para } i = 1, \dots, p \quad (8)$$

Según esto, el clasificador más sencillo consiste en asignar cada observación a la clase +1 o -1 según su ubicación respecto al hiperplano divisor. De manera concreta la observación  $x^*$  será asignada en función del valor de  $f(x^*) = y_i (\beta_0 + \beta_1 x^*_1 + \dots + \beta_p x^*_p)$ .

$$\begin{aligned} \text{Si } f(x^*) < 0 &\rightarrow x^* \in \text{Clase } -1 \\ f(x^*) > 0 &\rightarrow x^* \in \text{Clase } +1 \end{aligned} \quad (9)$$

Además del signo, la magnitud de  $f(x^*)$  permite analizar la distancia al hiperplano y con ello describir el error asociado a la confiabilidad de la clasificación. Sin embargo, la definición de hiperplanos para casos perfectamente separables resulta en un número infinito de hiperplanos, por lo que se debe trabajar una solución óptima. El clasificador óptimo estará determinado por un **maximal margin hyperplane (MMH)** o hiperplano de separación óptima, el cual corresponde al sub espacio más alejado del total de observaciones. La distancia perpendicular entre cada observación y un determinado hiperplano se denomina margen, por lo que, el que presente el mayor margen



resultará ser el más óptimo. La idea anterior requiere de infinitas mediciones (infinitos planos) por lo que exige procesos de optimización extras [28].

La o las observaciones que definen este margen se denominan vectores soporte, ya que son vectores en un espacio que definen el MMH permitiendo su modificación al variar dichos vectores. A pesar de esto, la mayoría de los casos reales no pueden ser separados de manera linealmente perfecta, es por ello que se extiende el concepto previo para poder definir un hiperplano que separe casi completamente las clases permitiendo la ocurrencia y cuantificación de un error.

**Support Vector Classifier (SVC's).** La variación en la definición del hiperplano se conoce como Support Vector Classifier o Soft Margin y permite crear un clasificador que, aunque no separa perfectamente las clases, es más robusto y posee una mayor capacidad predictiva al aplicarlo a nuevos datos. Lo anterior permite que las observaciones se clasifiquen del lado incorrecto del margen, incluso del hiperplano.

El proceso requiere de un parámetro de ajuste “ $C$ ” quien controla el número y severidad de las infracciones al margen y al hiperplano que se toleran en el ajuste. Si  $C \rightarrow \infty$  no se permiten violaciones al margen, por lo que es equivalente a tener el caso anterior (MMH), por el contrario, si  $C \rightarrow 0$  entonces la penalización es menor y mas observaciones pueden localizarse en el lado incorrecto de los limites. En simples palabras “ $C$ ” corresponde al hiperparametro encargado de controlar la tendencia (**Bias**) y varianza del modelo cuyo valor optimo puede obtenerse mediante métodos de validación cruzada [28].

Los SVC's mencionados al inicio del capítulo entregan buenos resultados cuando el límite de separación entre clases es del tipo lineal, en caso contrario su capacidad de clasificación se ve reducida. Para los casos de “no linealidad” se propone una estrategia: la expansión de la dimensión del espacio original.

Que los grupos no sean linealmente separables en el espacio definido no significa que no lo sean en un espacio de mayores dimensiones como se indica en la ilustración 6. El método de SVM, considerado una extensión del SVC, se obtiene mediante aumento de la dimensión original de los datos iniciales, y provoca que los limites lineales de los espacios aumentados se proyecten como limites no lineales en el espacio original.

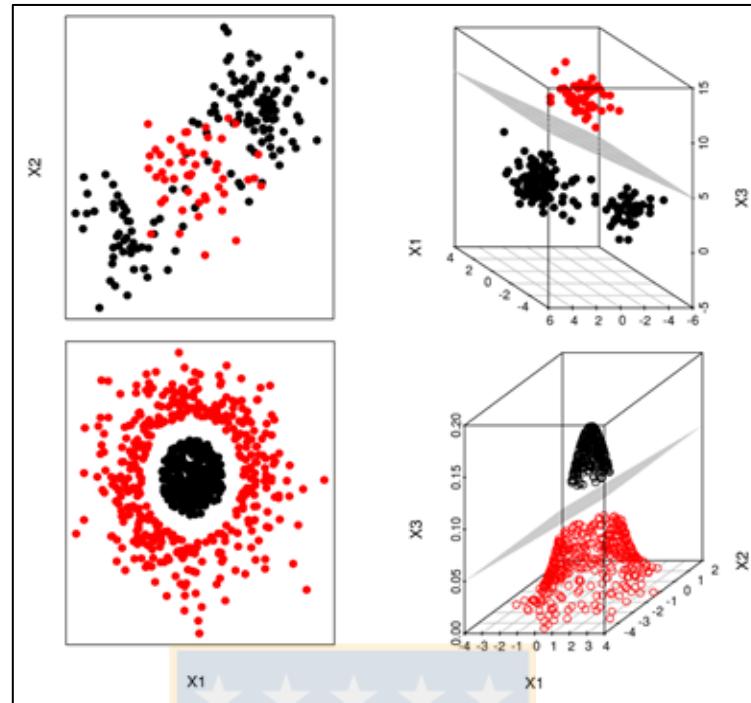


Ilustración 5.7: Visualización del aumento de dimensiones

**Kernels.** La dimensión de un conjunto de datos puede transformarse combinando o modificando cualquiera de sus dimensiones existentes. Por ejemplo,  $f(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$  sería una de las infinitas posibilidades para agregar una nueva dimensión a las consideradas inicialmente. Un Kernel ( $K$ ) es una función que entrega el resultado del producto punto entre dos vectores en un mismo espacio, cuyo resultado es expresado en un nuevo espacio dimensional distinto del original. La obtención de los vectores soporte se hace mediante la sustitución del producto punto existente en el proceso de optimización por algún tipo de Kernel. Muchos son los tipos existentes, pero los más utilizados se resumen en la siguiente tabla:

**Tabla 1: Kernels más frecuentes**

Nombre	Forma	Notas
Lineal	$K(x, x') = (x \cdot x')$	<ul style="list-style-type: none"> <li>El SVM obtenido es equivalente a un SVC</li> </ul>
Polinomial	$K(x, x') = (x \cdot x' + c)^d$	<ul style="list-style-type: none"> <li>Con <math>c = 0</math> y <math>d = 1</math> equivale a un SVC</li> <li><math>d &gt; 1</math> le otorga el carácter no lineal</li> </ul>
Gaussiano (RBF)	$K(x, x') = \exp(-\gamma \ x - x'\ ^2)$	<ul style="list-style-type: none"> <li><math>\gamma</math> controla el comportamiento del Kernel: <math>\gamma \rightarrow 0</math> (lineal) y <math>\gamma \rightarrow \infty</math> (flexible)</li> </ul>

Para todos los Kernels existe la opción de realizar una validación cruzada para hallar los hiperparámetros óptimos, además la selección de estas funciones depende netamente del tipo de problema que se desee resolver. Sin embargo, Chih-Wei Hsu [29] recomienda el uso del Radial Basis Function (RBF) puesto que presenta ciertas ventajas: Requiere de la optimización de solo dos parámetros  $\gamma$  (**Gamma**) y  $C$  (**penalización**), además su flexibilidad permite abarcar desde modelos lineales simples hasta otros más complejos.

**Support Vector Regression (SVR's)**. la regresión mediante vectores soporte funciona bajo principios y condiciones similares a SVM, tanto así que se menciona que esta regresión no es más que una adaptación de las máquinas cuando la variable de interés es de tipo numérica y no categórica. Su mayor ventaja es que se trata de una técnica no paramétrica donde el modelo resultante no depende de las distribuciones de las variables dependientes o independientes subyacentes. En vez de ello, los SVR utilizan y dependen en gran medida de las funciones Kernel; cuya idea básica es enfocarse en que el error asociado sea menor a determinado límite (concepto de MMC's) y por sobre la predicción. La técnica presenta, también, gran flexibilidad en cuanto a: distribuciones de variables, relación entre las variables dependientes e independientes y control del parámetro asociado a la penalización [28].

#### 5.2.4. Categorización de recursos y reservas

Corresponde a la última etapa de la evaluación de recursos y es considerado como un proceso crítico ya que define la cantidad (como tonelaje) y la calidad (como ley) de un yacimiento, parámetros fuertemente ligados a los que es la factibilidad económica de un proyecto minero puesto que de ello depende el nivel de inversión asociado al caso. La categorización de recursos y reservas es un proceso, no solo crítico, sino que también, "subjetivo" ya que su confiabilidad y precisión dependerá de la capacitación y experiencia del profesional a cargo. Por lo anterior, es que se establecen códigos internacionales los cuales definen metodologías, nomenclaturas y conceptos para lograr una clasificación transparente, imparcial y verídica mediante el cumplimiento de determinados estándares, según los defina cada código. Algunos de estos códigos son: JORC (Australasia), SAMREC (República Sudafricana), CIM (Canadá), IMM (Reino Unido). En el presente trabajo se abarcará solamente el uso del código JORC.

## Código JORC

Desarrollado por el Instituto de Minería y Metalurgia de Australia en conjunto con el Instituto de Geociencias de Australia, se promulgó y publicó a principio de los años 90 y sirvió como precursor y base para diversos códigos internacionales para el manejo y publicación de información asociada a recursos y reservas minerales [30].

En él se definen y clasifican los recursos y reservas según el nivel de conocimiento geológico y la confianza de este, al mismo tiempo que se consideran factores modificadores asociados a múltiples disciplinas del área: factores mineros, metalúrgicos, económicos, políticos e incluso sociales, entre otros. La relación entre las subcategorías de recursos con las de reservas se representa en la siguiente ilustración.

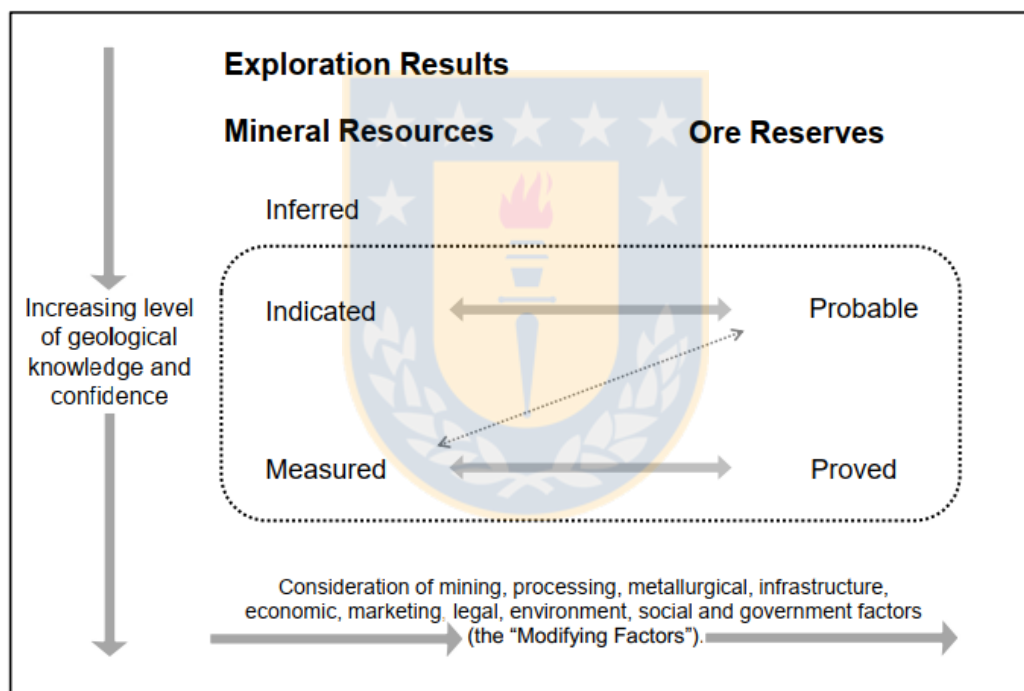


Ilustración 5.8: Esquema de clasificación y relación entre reservas y recursos minerales.

Las definiciones legales de recurso y reserva mineral según el código JORC se presentan al inicio del presente capítulo en *Definiciones y conceptos previos*. Matemáticamente la clasificación puede realizarse mediante el uso de Kriging o simulaciones, bajo la definición y restricción de 3 elementos: Error asociado al valor estimado, Probabilidad de ocurrencia (Tiempo) y el volumen de producción en periodos definidos (mensual, trimestral, semestral o anual). A continuación, se definen las subcategorías según el código JORC de manera legal y cuantitativa [27].

- **Recurso mineral inferido.** Parte de un recurso mineral en la que la cantidad y ley se estima sobre la base de evidencia geológica y muestreo limitados. La evidencia geológica es suficiente para suponer, pero no verificar la continuidad geológica y de ley. Se basa en información de exploración, muestreo y pruebas recogidas a través de técnicas apropiadas de ubicaciones tales como afloramientos, trincheras, excavaciones, labores y sondajes. No hay una evaluación a nivel cuantitativo, solo aquellos bloques fuera de las otras clasificaciones que presenten una estimación razonable para el experto.
- **Recurso mineral indicado.** Aquella parte de un recurso mineral por el cual la cantidad, ley, densidades, forma y características físicas se estiman con suficiente confianza para permitir la aplicación de factores modificadores en suficiente detalle para sustentar el planeamiento de minado y evaluación de la viabilidad económica del depósito. Cuantitativamente son aquellos bloques cuya ley está en un intervalo de  $\pm 15\%$  de error de estimación, un 90% del tiempo para una producción anual.
- **Recurso mineral medido.** Parte de un recurso mineral para la cual la cantidad, ley, densidades, forma y características físicas se estiman con confianza suficiente para permitir la aplicación de factores modificadores para sustentar un planeamiento de minado detallado y la evaluación final de la viabilidad económica del depósito. Cuantitativamente son aquellos bloques cuya ley está en un intervalo de  $\pm 15\%$  de error de estimación, un 90% del tiempo para una producción trimestral.
- **Reserva mineral probable.** Corresponde a la parte económicamente explotable de un recurso mineral indicado, y en algunas circunstancias, medido. La confianza en la aplicación de los factores modificadores a una reserva mineral probable es menor que para aquella probada. Cuantitativamente son aquellos bloques cuya ley está en un intervalo de  $\pm 15\%$  de error de estimación, un 90% del tiempo para una producción trimestral
- **Reserva mineral probada.** Parte económicamente explotable de un recurso mineral medido. Implica un alto grado de confianza en los factores modificadores. Cuantitativamente son aquellos bloques cuya ley está en un intervalo de  $\pm 15\%$  de error de estimación, un 90% del tiempo para una producción mensual.

*[\*] notar que la definición cuantitativa requiere asumir o conocer los ritmos de explotación y la geometría de los volúmenes.*

### 5.3. Agrupamiento y Clasificación de datos: Clustering.

El análisis de **clúster** es el proceso de clasificación de objetos en subconjuntos con sentido en el contexto de un problema en particular. Los objetos están organizados en una representación eficiente que caracteriza la población que está en estudio. El proceso de agrupamiento de datos mediante métodos de clustering se considera una técnica de aprendizaje sin supervisión (unsupervised learning) que agrupan una determinada base de datos en subclases categóricas tal que los objetos dentro de una misma subclase se consideran más similares entre ellos, que respecto al resto de las otras subclases.

De manera más concreta, el clustering corresponde a un tipo de clasificación impuesta a un conjunto finito de **objetos**, donde la relación entre objetos está establecida por una **matriz de proximidad o similitud**, cuyas filas y columnas, representan dicha su similitud mediante un valor numérico. Si tales objetos están caracterizados como patrones o puntos en un espacio físico d-dimensional, las relaciones o similitudes mencionadas se expresan como la distancia, del tipo que sea, entre aquellos pares de puntos.

A menos que se establezca una significativa medida de similitud, no será posible realizar un análisis funcional del clúster. La matriz de proximidad es el único *input* que requiere el desarrollo de un algoritmo de clustering, aun así existen diversas formas de agrupar datos o métodos de clustering y es mediante un **árbol de clasificación de problemas** [31] que se definen los diferentes géneros en los que se pueden clasificar [32].

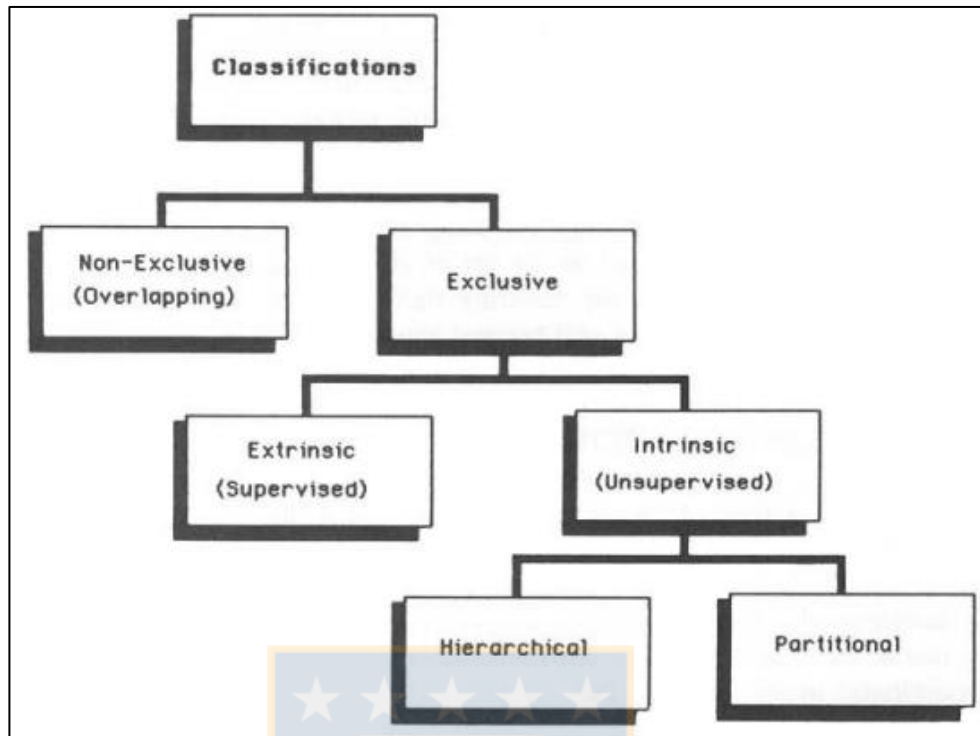


Ilustración 5.9: Árbol de clasificación de problemas, Lance & Williams, 1967.

- a. **Exclusivo y no exclusivo.** Una clasificación exclusiva es una partición de un conjunto de objetos donde cada uno de ellos corresponde a solo un subconjunto o *clúster*. De manera contraria, la clasificación no exclusiva, también denominado *overlapping*, puede asignar un objeto a diversas clases de manera simultánea.
- b. **Intrínseco y extrínseco.** Una clasificación intrínseca para realizar la tarea, utiliza solo la matriz de proximidad. Se le denomina también “aprendizaje sin supervisión” en el reconocimiento de patrones, puesto que no utiliza etiquetas de categoría para generar particiones a priori. En cambio, la clasificación extrínseca solo utiliza etiquetas de categoría en los objetos, sin considerar ni establecer una superficie discriminante que delimite los objetos en sus categorías; en otras palabras, este último requiere de un mediador mientras que con clasificación intrínseca se maximiza el uso de información de la matriz de similitud.
- c. **Jerárquico y Particional.** Las clasificaciones exclusivas intrínsecas se subdividen en jerárquicas y particionales según el tipo de estructura impuesta a la base de datos. La clasificación jerárquica corresponde a la secuencia anidada de particiones, mientras que una clasificación particional se explica como una única partición, estos conceptos se detallaran en secciones posteriores.

En adelante el termino clustering será utilizado para una clasificación exclusiva intrínseca particional, por consiguiente, la clasificación exclusiva jerárquica será denominada como clustering jerárquico.

### 5.3.1. Clustering Jerárquico

El Clustering jerárquico es un procedimiento para transformar la matriz en una secuencia de particiones anidadas. Un algoritmo de clustering es la secuencia de pasos para realizar la partición y posterior clasificación, es por esto que resulta conveniente caracterizar el método jerárquico mediante el desarrollo de un algoritmo jerárquico, teniendo en cuenta que este debe considerarse de manera separada al método mismo. Dubes & Jain [32] enfatizan en la distinción entre método de clustering y algoritmo de clustering de la siguiente manera: El primero especifica la estrategia general para agrupar datos o patrones en clúster, como, por ejemplo, minimizar el cuadrado del error. En cambio, el algoritmo corresponde al procedimiento computacional que implementa cierta estrategia incorporando variadas heurísticas.

De manera práctica, primero se debe generar la noción de una secuencia de particiones anidadas donde los “n” objetos a ser clasificados son denotados por un conjunto que los englobe a todos:  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , con  $x_i$  como el i-ésimo objeto del conjunto  $X$

Sea la partición  $C$  de  $X$ , se dice que esta divide a  $X$  en subconjuntos  $\{C_1, C_2, \dots, C_m\}$ , sí tales cumplen con las siguientes condiciones:

- I.  $C_i \cap C_j = \emptyset$  con  $i, j \in \{1, 2, \dots, m\}; \forall i \neq j$
- II.  $C_1 \cup C_2 \cup \dots \cup C_m = X$

La primera de estas consideraciones denota el carácter de clasificación exclusiva, es decir, los objetos no pueden pertenecer a múltiples particiones en simultaneo. Una partición se define como anidada dentro de otra si cada componente u objetos de una de ellas es subconjunto de la partición restante: “Sea el conjunto de objetos  $\{x_1, x_2, \dots, x_6\}$  se tienen las siguientes particiones  $B, C$  y  $D$ ”

$$B = \{(x_1, x_3), (x_5), (x_4), (x_2, x_6)\}; C = \{(x_1, x_3, x_5), (x_2, x_4, x_6)\};$$

$$D = \{(x_1, x_2, x_3), (x_4, x_5, x_6)\}$$

De lo anterior es posible concluir que  $B$  está anidada dentro de  $C$ , y que, ni  $B$  ni  $C$  están anidadas en la partición  $D$ , o viceversa.



En los métodos de Clustering Jerárquico los clústeres se forman mediante la división iterativa de patrones siguiendo un procedimiento de estilo **top-down** o **bottom-up** según sea el punto inicio y de término de la secuencia, estos se explican a continuación.

Existen 2 formas de realizar la jerarquización. Primero, el método aglomerativo sigue una secuencia bottom-up, o de abajo hacia arriba, la cual construye y define individualmente a cada objeto del conjunto como clústeres iniciales. A medida que se inician las iteraciones respectivas de la clasificación, los n-objetos comienzan a converger hacia un final y único gran clúster que engloba a la totalidad de ellos una vez cumplidas las condiciones de detención del proceso.

De manera totalmente contraria, un método divisivo sigue un orden top-down, o de arriba hacia abajo, en donde se comienza con un único gran clúster que agrupe a los n-objetos, clúster que luego de cada iteración se parte y divide en otros subgrupos cada vez más pequeños. Estos procesos de partición tienden a ser bastante complejos de visualizar y tratar de manera abstracta si se desea analizar una cantidad muy grande de datos, es por esto que existen herramientas para un generar un mejor análisis [32].

Un dendrograma es un tipo especial de estructura de ramas (tipo árbol) el cual permite esquematizar de manera más simple el proceso de agrupamiento resultante. Consiste en diversos niveles denominados “nodos” los cuales representan los clústeres conectados mediante las líneas del dendrograma que representan el anidamiento entre particiones o subconjuntos. Mediante la lectura de manera horizontal de este esquema se pueden identificar las agrupaciones realizadas.

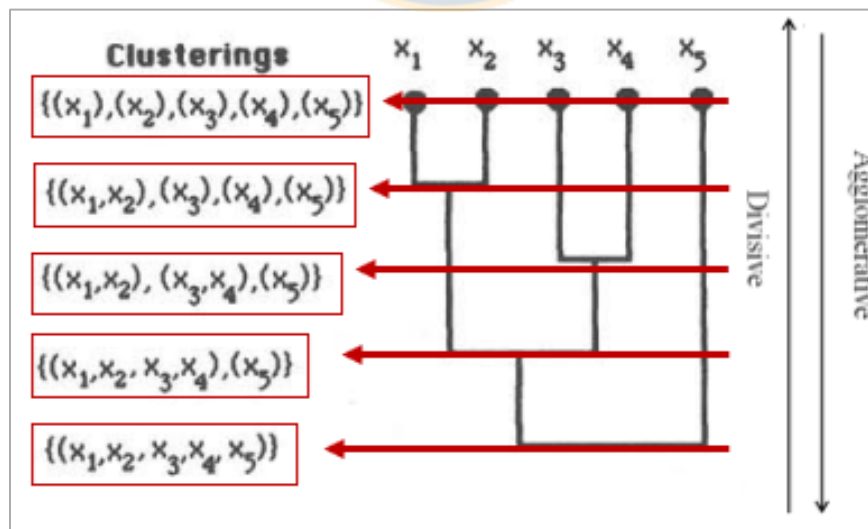


Ilustración 5.10: El dendrograma, componentes y lectura. Fuente: Elaboración propia

Los métodos jerárquicos pueden ser agrupados también en 3 categorías en base a sus medidas de similitud o relación [33].

- a. **Clustering de relación única (*Single-link*)**. Es usualmente denominado como clustering de conectividad, método mínimo o vecino más cercano. En este tipo de C.J. el vínculo entre dos clústeres A y B está determinado por un solo elemento par, dicho de otra forma, por los dos elementos (uno en cada partición) que se encuentren a la menor distancia uno del otro. La similitud y distancia entre subconjuntos estarán definidos por la de distancia mínima, de cualquier tipo, que exista entre cualquiera de los datos contenidos en ellos.

$$\min \{d(a, b); a \in A \wedge b \in B\}$$

- b. **Clustering de relación Completa (*Complete-link*)**. También denominado como el vecino más lejano, método del máximo o del diámetro. De manera análoga a la anterior clasificación se establece que la relación que existe entre dos clústeres se definirá en base a la distancia mayor existente entre los objetos más alejados de cada uno de ellos

$$\max \{d(a, b); a \in A \wedge b \in B\}$$

- c. **Clustering de relación promedio**. La distancia y relación entre dos particiones/subconjuntos estará definida por el promedio de las distancias entre cualquiera de los objetos que constituyan dichos clústeres, por lo mismo es que es conocido también como método de menor varianza.

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$$

Sin embargo, la crítica más común para los algoritmos de clustering jerárquico es que carecen de robustez y, por lo tanto, resultan ser sensibles a ruidos y presencia de outliers dentro de la base de datos.

Otro inconveniente que presenta es su característica de exclusividad, ya que, una vez sea asignado un objeto a determinado clúster este no podrá ser considerado nuevamente en posteriores procesos iterativos lo que significa que el algoritmo no será capaz de corregir una clasificación potencialmente incorrecta. Sumado a lo anterior la complejidad computacional que presentan realizar la mayoría de estas clasificaciones es de al menos  $O(n^2)$ , esto significa que es

proporcional al cuadrado de la cantidad de datos, alto costo que complica su aplicación a bases de datos de gran tamaño.

### 5.3.2. Clustering Particional

Las técnicas de agrupamiento jerárquico organizan los datos en secuencias de conjuntos anidados donde destaca como gran característica el impacto visual que genera el dendrograma resultante, permitiendo un análisis más simple mediante la mera observación del cómo se generan estas agrupaciones/separaciones en distintos niveles de proximidades, facilitando la labor del analista al momento de elegir o definir el clúster que representa mejor el conjunto de datos de interés.

Los métodos no jerárquicos caen en la clasificación de métodos de agrupamiento particional como lo indica el árbol de clasificación de problemas (ilustración 9). En este tipo de métodos se genera una única partición de los datos con el fin de representar los grupos “de formación natural” presentes en la base de datos. Las técnicas de clustering o agrupamiento particional son usadas frecuentemente en aplicaciones ingenieriles, área donde las particiones únicas adquieren relevancia ya que son consideradas como especialmente apropiadas para la representar y comprender grandes bases de datos de manera eficiente, casos donde el uso y análisis de diagramas (dendrograma) se hace bastante ineficiente [32]. Un problema del tipo particional puede definirse de siguiente manera.

*Dados  $n$  objetos en un espacio  $d$ -dimensional, se desea determinar la partición de los datos en  $k$ -grupos, o clúster, tal que los objetos/datos de un mismo conjunto sean más similares entre ellos que con los de otras agrupaciones distintas.*

Para ello el valor de  $k$  puede ser o no especificado, además se debe definir un criterio de agrupamiento que puede ser clasificado como **global** o **local**. El primer tipo de criterio representa a cada clúster mediante un prototipo y asigna los patrones de objetos según la similitud entre prototipos. En cambio, el criterio local genera los conjuntos utilizando la estructura propia o que presentan los datos.

Por ejemplo, los clúster se pueden formar identificando regiones de alta densidad (similitud) en los objetos de un espacio, o también, asignando un objeto y sus  $k$  vecinos más cercanos a un mismo grupo.

La solución teórica para el problema anteriormente mencionado se puede resumir de manera simple: Inicialmente, se selecciona un criterio de agrupamiento, como el cuadrado del error o la distancia euclidiana. Luego se evalúa para todas las posibles particiones de  $k$ -grupos y finalmente se elige la partición que resulte ser la más óptima de todas.

La realidad es distinta, la solución teórica no es tan factible ya que se debe tener en cuenta dos puntos importantes [33]:

- a. El criterio que se seleccione para transformar la noción básica e intuitiva de un clúster en una fórmula matemática debe resultar razonablemente simple, en cuanto a su uso.
- b. El número de particiones que se generan para bases de datos grandes, del orden de miles a decenas de miles de datos, son tantas que no facilitan la tarea, debido a que evaluar un criterio, por muy simple que este sea, en cada una de las particiones realizadas resulta, poco práctico y altamente costoso en cuanto a recursos.

De lo anterior es que se genera la necesidad de definir diversas metodologías y algoritmos para abordar de manera efectiva este tipo de problemas de clasificación, dentro de los más utilizados se mencionan los métodos de: K-medias, PAM, C-Medias, DBSCAN.

➤ K-medias (*k-means*)

Es de los más conocidos, comparables y simples algoritmos [34, 35] para resolver problemas de clasificación del tipo particional. Suponga una base de datos dada  $X = \{x_1, \dots, x_n\}$  con  $x_n \in \mathbb{R}^d$ . El problema de *M-Clustering* apunta a dividir y clasificar este conjunto  $X$  de datos en  $M$  subconjuntos disjuntos (clúster)  $\{C_1, \dots, C_m\}$  tal que un criterio de clasificación cualquiera, sea optimizado. El criterio utilizado mayormente es la suma de las distancias euclidianas al cuadrado, este se aplica entre cada punto  $x_i$  y los centroides  $m_k$  del subconjunto  $C_k$  que contenga a dicho  $x_i$ . A este criterio se le denomina error de Clustering y se define de la siguiente manera

$$E(m_1, \dots, m_k) = \sum_{k=1}^k \sum_{i=1}^i \|x_i - m_k\|^2 \quad (10)$$

El algoritmo de k-medias encuentra soluciones óptimas locales con respecto al error del clustering. Este corresponde a un algoritmo iterativo relativamente simple y se define como un método basado en puntos (point-based) el cual, inicialmente ubica los centros  $m_k$  de los clusters de

manera arbitraria para luego proceder a reubicarlos luego de cada iteración de manera que el error de clustering sea minimizado.

La principal desventaja del método recae en su sensibilidad a la posición inicial del centroide. Por lo tanto, para obtener una solución lo más óptima posible mediante el método de K-medias se deben ejecutar múltiples ensayos, donde las posiciones iniciales de dichos centroides difieran. El número de iteraciones requeridas puede variar en un amplio rango, desde cientos hasta algunos miles o decenas de miles, dependiendo del número y tipo de variables (clúster, objetos y su distribución).

Existen dos enfoques principales descritos en la literatura [36] que pueden ser utilizados para la reducción de los requerimientos computacionales asociados al método de k-medias, especialmente para el cálculo de distancias:

- I. Usar la información de la iteración previa para reducir las distancias calculadas. P-clúster es un algoritmo de clustering basado en el de K-medias, el cual aprovecha el hecho de que el cambio de asignación de objetos se ve disminuido luego de las primeras pocas iteraciones [37]. Utiliza una heurística la cual determina si el prototipo más cercano de un patrón “q” ha cambiado o no mediante un simple chequeo. Si la asignación no cambia, entonces el cálculo de nuevas distancias ya no es necesario. Esto, además de usar el hecho de que la posición de los centroides, luego de unas cuantas iteraciones consecutivas, varían de manera mínima.
- II. Organizar los vectores de prototipos en una estructura de datos deseable para que la tarea de encontrar un prototipo más cercano para un patrón dado sea más eficiente [38]. El número de cálculos requeridos para determinar la distancia usando esta metodología es proporcional a  $n * f(k, d)$  por cada iteración realizada. Para varias aplicaciones como la cuantificación de vectores, los vectores del prototipo están fijados.

Esto permite la construcción de estructuras de datos óptimas en el sentido de encontrar el vector más cercano que permita un testeo del objeto ingresado (**input test pattern**). Sin embargo, estas optimizaciones no son aplicables al método de k-medias, esto dado que los vectores de prototipo cambian drásticamente.

➤ DBSCAN (Density-Based Spatial Clustering of Application with Noise)

K-medias junto con otros métodos de clasificación de tipo particional, producen clúster de forma esférica únicamente y para ello solo requieren de la cantidad de clúster como *input* del algoritmo. Existen otros tipos de métodos particionales que si permiten darle forma “arbitraria” a los subconjuntos resultantes como lo es el *Kernel k-means* [39], pero que no resultan tan prácticos debido al alto costo en recursos (tiempo y complejidad computacional) que requieren estos algoritmos para ser puestos en marcha.

En la siguiente ilustración se muestran tres distintas bases de datos (puntos) en las cuales es fácil e intuitivo reconocer tanto los clústeres realizados como los puntos “ruido” que no pertenecen a esta primera categoría.

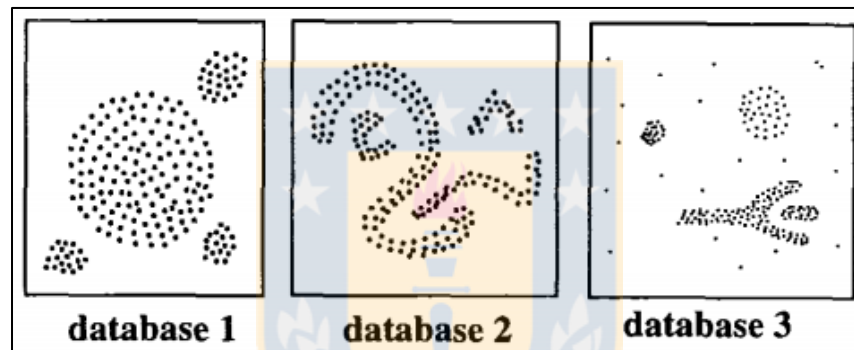


Ilustración 5.11: bases de puntos agrupados mediante criterios de densidad.

La principal razón por la que se pueden reconocer estos agrupamientos de manera tan sencilla es porque cada grupo formado presenta una densidad o concentración de puntos considerablemente mayor que sus alrededores. Por otro lado, aquellas zonas con presencia de puntos de manera diseminada, denotan zonas de menor densidad que son clasificadas como outliers o ruido.

La noción intuitiva para la definición de clúster y ruido dentro de un espacio físico  $k$ -dimensional basado en la densidad de distribución de puntos sigue una idea principal: Para cada punto de un clúster, su vecindad, definida por un determinado radio, debe contener un mínimo de puntos. En otras palabras, la densidad de vecindad debe superar un determinado límite o umbral para definir un subgrupo de puntos como un clúster, y la forma de este estará determinada por la elección de una función de distancia entre dos puntos.

El método de DBSCAN, o Density-Based Spatial Clustering of Application with Noise, es un algoritmo pionero dentro de las técnicas de clasificación basadas en la densidad de datos el cual

permite describir clúster con múltiples formas, no necesariamente esféricas, y controlar los ruidos u outliers de la base de datos de manera efectiva [40]. Comparado con otros métodos no basados en densidad de puntos, este tiene características únicas y avanzadas útiles para tareas como detectar objetos, patrones, clases o estructuras de diversos tamaños y formas. Es por estas características que se considera como un correcto candidato para encontrar clúster de “formación natural”, cuando los datos presentan alta variabilidad en concentración de la distribución de puntos en el espacio [39].

En cuanto al algoritmo, existen algunas definiciones relacionadas a los inputs requeridos y son los siguientes

- a. ***eps-neighbourhood/vecindad***. Para un punto  $x \in D$ , la vecindad-*eps* denota al conjunto de puntos cuya distancia a  $x$  es menor o igual a “*eps*”. La cardinalidad de la vecindad definida se denomina umbral o límite de densidad de  $x$ .
- b. ***eps-connected/relacionado***. Para un par de puntos  $x, y \in D$ , Si  $||x - y|| \leq eps$ ; entonces  $x$  e  $y$  son puntos *eps*-relacionados.
- c. ***minpts***. Abreviatura para puntos mínimos (*minimum points*) requeridos para conformar un clúster, cardinalmente equivalente al umbral de densidad.

Desde el punto de vista del método DBSCAN todos los puntos/objetos de la base de datos serán clasificados en dos categorías: Core o Border points. Estos últimos pueden sub clasificarse como ruido o en su defecto como puntos densamente conexos (*Density-connected*) [39].

- **Puntos centrales (Core-points)**. O puntos “de núcleo”, corresponden a aquellos con una densidad de vecindad mayor o igual al umbral requerido.
- **Puntos limite (Border-points)**. Los puntos extremos o periféricos son aquellos con una vecindad menor a la cantidad limite que establece *minpts*.
  - ***Noise-points***. Un punto  $p \in D$ , será considerado un punto de ruido si y solo si su umbral de densidad es menor que *minpts* y todos los puntos a una distancia *eps* (puntos *eps-connected*) de  $p$  son border-points.
  - ***Density-connected points***. Son aquellos puntos de frontera (border) que dentro de su vecindad-*eps* contienen al menos un punto de núcleo (core)

El algoritmo de DBSCAN se inicia con solo dos parámetros de entrada, mencionados anteriormente: *eps* y *minpts*.

Inicialmente todos los puntos son clasificados como puntos “sin visitar”, el algoritmo comienza seleccionando al azar cualquiera de estos puntos (a menos que se especifique) y construye su vecindad- $\epsilon$  en base al input mencionado. Si el número de puntos en dicha vecindad es menor que  $minpts$  entonces es clasificado como ruido o outlier. Por el contrario, si esta es mayor que  $minpts$  es considerado como un punto denso o con una vecindad densa, y un nuevo clúster es creado. Los puntos cercanos son agregados de manera iterativa al clúster mediante la búsqueda de puntos centrales para cada punto en la vecindad- $\epsilon$  de este. Si los puntos aun sin visitar no pueden ser añadidos al clúster, este último queda completo y no le podrán ser añadidos más puntos en las iteraciones subsecuentes.

Para encontrar el siguiente clúster el algoritmo ubica otro punto sin visitar y repite el proceso de clasificación previo. El proceso es detenido cuando todos los puntos han sido clasificados dentro un clúster o dentro la categoría de ruido. Notar que todos los puntos de un clúster están  $\epsilon$ -relacionado con al menos un punto del cluster al que pertenece, a diferencia de aquellos que no forman parte del subgrupo.

### 5.3.3. Validación de Resultados

Uno de los temas más importantes en el análisis del agrupamiento de datos es la evaluación de los resultados del clustering para encontrar la estructura que mejor identifique/ajuste a la data, entiéndase por estructura de clustering (**EC**) tanto jerarquías como particiones o clúster individuales.

El proceso está definido por una serie de procedimientos que en conjunto generan un escenario de conformidad o disconformidad alrededor de un resultado en cuestión. El proceso se puede simplificar de la siguiente manera: Se debe escoger un estadístico, asumir/adoptar una hipótesis denominada “nula” en torno al problema, considerando una distribución base para el estadístico bajo dicha hipótesis nula. Para el estadístico se define un valor límite, valor definido en función de la base de datos, de manera de asegurar que la probabilidad de rechazar la hipótesis nula cuando esta sea correcta/verdad no sea mayor que un determinado nivel. Especial cuidado se debe tener con los detalles para que se pueda asegurar que todo lo anteriormente asumido se condiga con el estadístico utilizado. Para otros casos, esta distribución debe ser establecida mediante métodos de Monte-Carlo [41].



### 5.3.3.1. Prueba de hipótesis

Un estadístico  $T$  es una función la cual, bajo ciertos criterios, contiene información útil propia del conjunto de datos. En términos matemáticos  $T$  es una variable aleatoria cuya distribución describe la frecuencia con la cual un valor de  $T$  ocurre bajo determinada hipótesis, para esto se requiere que un espacio físico muestral o una población referencial existente. La elección de esta hipótesis, denominada hipótesis nula ( $H_0$ ), debe definir si sigue una idea de aleatoriedad o de estructura, es decir, si “Los datos son aleatorios” o bien “los datos están agrupados estructuradamente”.

En particular para los casos de validación de clustering, la hipótesis nula corresponde a una declaración de aleatoriedad en los datos, y los tres tipos más comunes son

- a. **Hipótesis de grafico aleatorio.** Todas las matrices de proximidad/similitud (MS) de orden  $n \times n$  son equiprobables
- b. **Hipótesis de etiqueta aleatoria.** La permutación de los etiquetas (*labels*) en  $n$  objetos es equiprobable
- c. **Hipótesis de posición aleatoria.** Todos los  $n$  conjuntos en un espacio  $d$  dimensional son equiprobables

### 5.3.3.2. Definición de un test

Supóngase un estadístico  $T$  y una hipótesis nula  $H_0$  que concuerdan. Suponga también que la distribución  $T$  es conocida bajo dicha  $H_0$ . Sea  $P(B|H_0)$  la probabilidad de ocurrencia de un evento  $B$  bajo la hipótesis nula propuesta  $H_0$ , el evento  $B$  puede desembocar en “ $T \leq t$ ” o “ $T \geq t$ ” donde  $t$  es un límite definido previamente. Sea  $\alpha$  (Alpha) un valor pequeño (0.05-0.1) denominado nivel de test, permite la definición del valor límite  $t_\alpha$  para no rechazar la hipótesis nula. Finalmente, sea  $t^*$  un valor experimental de  $T$ , el rechazo de  $H_0$  estará dado por la siguiente expresión:

$$P(T \geq t^*) = \alpha$$

$$\therefore \text{si } t^* \geq t_\alpha \rightarrow H_0 \text{ es rechazada con un nivel } \alpha$$

De lo anterior cabe destacar que  $\alpha$  finalmente corresponde a la probabilidad de rechazar la hipótesis cuando esta es verdadera. La aplicación de este test produce dos respuestas: si la estructura propuesta a un nivel  $\alpha$  es valida o en su defecto invalida.

### 5.3.3.3. Estadísticos e índices

**Estadístico de Hubert ( $\Gamma$ ).** Se presenta como una opción para la validación mediante comparación entre estructuras. Como ejemplo se puede mencionar que el número de objetos asignados a los clústeres puede ser comparado a los CATEGORY LABELS asignados, de manera independiente, al clustering realizado. El estadístico de Hubert, denotado por  $\Gamma$ , es efectivo en lograr evaluar el ajuste entre una base de datos y una estructura inferida previamente [41]. Hubert y Schultz [42] exponen de manera detallada y completa las aplicaciones para el estadístico en la literatura. De manera simplificada,  $\Gamma$  representa la correlación en serie entre dos matrices de similitud  $X(i, j)$  e  $Y(i, j)$  de una misma data y este puede expresarse como

$$\Gamma = \sum_{i=1}^{n-1} \sum_{j=i+1}^n X(i, j) Y(i, j) \quad (11)$$

Normalizado, el estadístico de Hubert corresponde al coeficiente de correlación de la muestra entre los valores de ambas matrices. Sean  $m_x$  y  $m_y$  las medias y  $S_x$  con  $S_y$  las desviaciones estándar de dichos valores pertenecientes a  $X$  e  $Y$  respectivamente,  $\Gamma$  queda normalizado mediante la siguiente ecuación:

$$\Gamma_n = \left\{ \left( \frac{1}{M} \right) \sum_{i=1}^{n-1} \sum_{j=i+1}^n [X(i, j) - m_x] [Y(i, j) - m_y] \right\} / (S_x S_y) \quad (12)$$

Donde  $M = n(n-1)/2$  corresponde a la cantidad de valores de la doble suma. Los momentos restantes se definen de la siguiente manera

$$\begin{aligned} m_x &= \left( \frac{1}{M} \right) \sum \sum X(i, j) & S_x^2 &= \left( \frac{1}{M} \right) \sum \sum X^2(i, j) - m_x \\ m_y &= \left( \frac{1}{M} \right) \sum \sum Y(i, j) & S_y^2 &= \left( \frac{1}{M} \right) \sum \sum Y^2(i, j) - m_y \end{aligned} \quad (12a)$$

Todo lo anteriormente expuesto permite la medición del grado de correspondencia lineal entre los valores de  $X$  e  $Y$ . Valores absolutos altos indican que las matrices concuerdan entre ellas, considerando que el dominio de  $\Gamma_n \in [-1, 1]$ .

La hipótesis más común para este corresponde a la de aleatoriedad de etiqueta (*label*), esto sería:

$H_0$ : "todas las permutaciones de una fila o columna de  $[Y(i, j)]$  son equiprobables"

### Estadístico de Goodman-Kruskal ( $\gamma$ ).

Este estadístico mide el rango de correlación entre dos secuencias ordinales de números, números que pueden ser obtenidos de la matriz de similitud, así como mediante el estadístico de Hubert. La idea detrás de este  $\gamma$  se ejemplifica a continuación.

Sean  $A = \{a_1, \dots, a_m\}$  y  $B = \{b_1, \dots, b_m\}$  dos secuencias ordinales de números, el estadístico de Goodman—Kruskal mide el grado de asociación entre dichas secuencias mediante lo que se denomina concordancia de “pares”. Un par es un conjunto de dos parejas/dobletes, una de cada secuencia, el cual se clasifica en Concordante ( $S_+$ ) o Discordante ( $S_-$ ) si el par de dobles  $\{(a_i, a_j); (b_i, b_j)\}$  cumple con

- Concordancia:  $a_i < a_j \wedge b_i < b_j$  o  $a_i > a_j \wedge b_i > b_j$
- Discordancia:  $a_i < a_j \wedge b_i > b_j$  o  $a_i > a_j \wedge b_i < b_j$

En caso de existir una igualdad ( $a_i = a_j$  o  $b_i = b_j$ ) no se considerarán bajo ninguna de estas clasificaciones. Finalmente, el estadístico  $\gamma$  es computado del total de números pares concordantes y discordantes de la siguiente manera:

$$\gamma = \frac{S_+ - S_-}{S_+ + S_-}; \gamma \in [-1, 1] \quad (13)$$

De lo anterior, si  $\gamma$  tiende a  $-1$  entonces se puede indicar que existe una relación alta y de tipo inversa entre las secuencias, la tendencia de  $\gamma$  a 1 solo cambia el tipo de relación a una directamente proporcional, la cercanía al cero indica que no se puede establecer una relación concreta entre las secuencias.

En términos generales existen tres criterios bajo los cuales se puede corroborar la validez de un clúster [43].

**Externa (External).** Mide el desempeño de un resultado comparando estructuras de clúster entre sí. Por ejemplo, un criterio externo mide el grado de correspondencia entre números de grupos obtenidos de un algoritmo de clustering, y marcas de clase asignadas a priori. Este también permite medir el grado al cual los datos confirman, de manera aparente, ideas previas sin haber realizado un análisis formal del clúster.

**Interna (Internal).** Evalúa el ajuste entre la estructura y la data usando únicamente información de la data. Por ejemplo, un criterio de tipo interno puede medir el grado al cual una

partición, resultado de un algoritmo de clustering, es justificado por la matriz de proximidad respectiva.

**Relativa (Relative).** Decide cuál de las estructuras es la mejor en determinado sentido, tal como, mayor estabilidad u otro más apropiado para según corresponda al conjunto de datos. Por ejemplo, un criterio relativo puede medir cuantitativamente si es que una jerarquía de tipo *single o complete-link* se ajusta mejor a los datos

En concreto, un índice o criterio expresa la estrategia mediante la cual una estructura de clustering será validada, mientras que un estadístico determina cual será la validez de la prueba.

**Davis-Bouldin index.** Se busca establecer una medida de separación entre clúster que incorpore características de algunas medidas de similitud aceptadas y aplicables a problemas de análisis de clustering, que a su vez satisfagan ciertos criterios heurísticos. Para ello, una medida de similitud  $R_{ij}$  entre los Clusters  $C_i$  y  $C_j$  es definida en base a las medidas de dispersión  $s_i$  de  $C_i$  junto con una medida de disimilitud (distancia entre vectores característicos de  $C_i$  y  $C_j$ ) denominada  $D_{ij}$  de manera que cumpla con las siguientes condiciones [44, 45]

- a)  $R_{ij} \geq 0$
- b)  $R_{ij} = R_{ji}$
- c)  $s_i = 0$  y  $s_j = 0$  entonces  $R_{ij} = 0$
- d)  $s_j > s_k$  y  $d_{ij} = d_{ik}$ , entonces  $R_{ij} > R_{ik}$
- e)  $s_j = s_k$  y  $d_{ij} < d_{ik}$ , entonces  $R_{ij} < R_{ik}$

Mediante estas condiciones es que se establece la simetría y no negatividad de  $R_{ij}$ . Una elección para la medida  $R_{ij}$  dichas condiciones es la definida por D. Davis y D. Bouldin [44]

$$R_{ij} = (s_i + s_j)/d_{ij} \quad (14)$$

Del cual se extrae el índice respectivo D.B. (Davis-Bouldin)

$$DB_n = \frac{1}{n_c} \sum_{i=1}^n R_i \quad (15)$$

Con  $R_i = \max_{i=1, \dots, n; i \neq j} R_{ij}$ ,  $i = 1, \dots, n$ ;  $i \neq j$ , donde  $n_c$  es el número total de clusters

Cabe destacar que tal índice corresponde a la similitud promedio entre todos los clústeres con su respectivo más parecido. En estricto rigor aquellos que presenten la mayor similitud representarán el conjunto completo ( $R_i$ ).

Como los clústeres deber ser compactos y estar distanciados, un menor valor del índice de D.B. significará que la configuración correspondiente es mejor.

#### 5.3.4. Dominios Geometalúrgicos

La geometalurgia integra muchos parámetros geológicos, metalúrgicos, minerales, y de modelamiento de recursos, entre otros. Su multidisciplinariedad permite la optimización de los procesos asociados a la actividad minera, tanto extractiva como de procesamiento, mediante la generación de un denominado **modelo geometalúrgico**. Este modelo se construye en base a muestreos, a pequeña escala, que permiten caracterizar la variabilidad natural de los parámetros metalúrgicos de un depósito determinado. Los dominios geometalúrgicos como tal corresponden a regiones tridimensionales de un espacio físico donde las variables metalúrgicas de interés se encuentran distribuidas de manera homogénea, por ello es que deben ser geoestadísticamente robustos y geológicamente significativos.

Definir los dominios o zonas es la primera tarea respectiva a la generación de un modelo. Para su definición los investigadores utilizan propiedades metalúrgicas cuantitativas, mediante ensayos químicos o propiedades petrofísicas; y cualitativas, como texturas o mineralización de la roca asociada. Una vez recogidos los parámetros se utilizan técnicas de análisis estadístico multivariable para entender la interrelación de dichas variables y extraer los patrones “escondidos” dentro de estas complejas bases de datos.

Los análisis de **Componentes Principales (PCA)** y **Clustering (CA)** son dos métodos multivariados frecuentemente utilizados para la identificación de potenciales estructuras dentro de un depósito. El método de PCA resume los datos, combinando las variables interrelacionadas iniciales para explicar la variabilidad de los datos [46]. Este se basa en un diagrama de discriminantes bidimensional (2D) obtenido a partir de los gráficos de dispersión de los objetos dentro de un espacio de componentes principales rotado. Tales objetos corresponden a los dos componentes principales que explican, mayoritariamente, la variabilidad de la base de datos correspondiente.

A pesar de ello posee una limitación bastante crítica y evidente: la definición de clases al no apoyarse en la variabilidad completa de la base, sino que solo de dos, pierde robustez cuando la esta presenta una alta cantidad de variables a analizar [47]. Para dichos casos es que el análisis mediante clustering, explicado en las secciones anteriores, surge como una buena herramienta multivariable para diversas aplicaciones y tamaños de bases de datos.

## 6. Resultados

### 6.1. Descripción de la base de datos

La base de datos, como bien se menciona en los capítulos 1 y 3, corresponde al relave SOTRAMIN y Playa Atacama ubicados en la localidad de Tal-Tal. Esta contiene la concentración de 30 elementos químicos, información obtenida a partir de 53 testigos y su respectivo análisis químico (ICP). Del total solo se trabajó con la porción de elementos que representase la mayor cantidad y calidad de información, dividiéndolos en 3 grupos.

- a. **Elementos objetivo.** Corresponden a los 17 elementos que serán objetivo del trabajo debido a su significativa concentración e información que poseen (correlación)

Cr, Al, Mg, Li, As, Hg, Au, Ca, Pb, Zn, Co, K, Na, Ni, Cu, V, Fe

- b. **Elementos secundarios.** Ofrecen poca información en cuanto a la nula correlación entre ellos o con el resto de los elementos, aun así, son elementos cuyas concentraciones no son del todo bajas o insignificantes.

Ba, Mn, Mo, P

- c. **Elementos Bajo umbral.** Los procesos de análisis químico requieren de una concentración mínima o umbral de cierto elemento para ser cuantificado por sobre este nivel. Estos no son considerados por no representar información útil para el presente estudio.

Ag, Be, Bi, Cd, Sb, Se, Sn, Ti

Originalmente, la unidad de medida para los valores de las variables corresponde a partes por millón (ppm). Sin embargo, solo algunos de los elementos objetivos se trabajaron en ppm (oro, plomo, mercurio, arsénico) mientras que el resto de estos se convirtieron a ley porcentual de mineral.

El compósito de las muestras se realizó mediante la división de cada pozo en tercios (Tope, Medio, Fondo) resultando en 1620 datos iniciales. Posterior al descarte de los elementos secundarios resultó una base de datos de 848 datos objetivos. Esta describe una torta de relave de dimensiones 105.3 metros (Este), por 44.8 metros (Norte), por 3 metros de potencia.

Se asume un peso específico del material igual a  $2.65 \text{ g/cm}^3$  debido al alto porcentaje de silicio (83% cuando menos) que compone el relave, probablemente en forma de cuarzo.

Los elementos objetivos se dividieron a su vez en 4 grupos, división en base a la correlación entre los elementos que componen cada uno

- Grupo 1: Cromo – Aluminio – Magnesio – Litio – Vanadio
- Grupo 2: Calcio – Plomo – Zinc
- Grupo 3: Arsénico – Mercurio – Oro
- Grupo 4: Cobalto – Potasio – Sodio – Níquel – Cobre – Hierro

Estos grupos son trabajados de manera semi independiente<sup>2</sup> por medio de diversos códigos del software de programación RStudio®

## 6.2. Resultados y análisis

### Análisis Estadístico.

Por medio de las herramientas de análisis de datos que ofrece Microsoft Excel, se realizaron los histogramas para cada elemento a modo de obtener los rangos de valores máximos y mínimos para cada uno. Estos rangos cobrarán importancia más adelante puesto que servirán para tener una referencia para los valores de los errores cometidos por cada método.

A su vez, se obtuvieron las correlaciones entre las variables para así definir los grupos, anteriormente presentados, en función de que tan similar es su comportamiento en el espacio.

Tabla 2: Emparejamiento de elementos según su correlación espacial.

Grupo 1		Grupo 2		Grupo 3		Grupo 4	
Al-Mg	0.956	Ca-Pb	0.888	As-Hg	0.743	Co-K	0.803
Al-V	0.916	Ca-Zn	0.728	As-Au	0.688	Co-Na/Ni	0.701
Mg-Cr/V	0.899	Pb-Zn	0.755	Au-Hg	0.781	Cu-Ni	0.835
Li-Mg/Zn	0.848					K-Ni	0.950
Li-Al	0.833					Fe-K	0.765
V-Cr	0.833						
Cr-Al	0.820						

<sup>2</sup> comparten un mismo espacio físico, pero no sus estadísticos característicos

## Estimación.

Para generar los modelos de bloque que representen la estimación hecha para cada variable en cada grupo, primero se cargó la base “relave.txt” al software y mediante el análisis de los rangos de sus coordenadas se generó una grilla y se dimensionaron los bloques en función de esta, resultando en bloques de 5x5x1 m<sup>3</sup> cada uno.

**Kriging.** Se aislaron cada uno de los grupos definiéndolos como listas  $g(i)$  de variables  $\{gstat\}$ . Cada lista contiene las coordenadas y concentraciones de los elementos respectivos, información que permitió generar un variograma experimental multivariable (MSV) (variogramas individuales y cruzados)  $\{variogram\}$ . A este se le ajustó un modelo lineal de correogionalización  $\{fit.lmc\}$  el cual evidenció a través de los gráficos resultantes un comportamiento altamente errático y sin una distribución espacial definida, más bien con un dominante efecto de pepita para los cuatro grupos.

Las mesetas y rangos ajustados  $\{v.fit(i)\}$  por el software fueron utilizados en la elaboración  $\{predict\}$  de un modelo de bloques que describe la concentración de los elementos en 1352 coordenadas en la grilla generada.

**Inverso de distancia.** La estimación mediante el método de inverso de la distancia se realizó variable por variable a diferencia del caso del Kriging puesto que la función utilizada  $\{idw\}$  requirió de la definición de menos etapas. Elemento objetivo, base de procedencia, modelo de bloques a estimar y una vecindad de 6 puntos máximos fueron los parámetros necesarios para realizar la predicción correspondiente.

**SVR.** Para la estimación mediante SVR se realizaron 100 simulaciones (gaussianas) para cada elemento en cada posición, esto es 135200 datos en total para cada elemento. Las simulaciones fueron promediadas para cada coordenada y luego agrupadas de acuerdo a los criterios de correlación establecidos previamente. De esta forma se produjo una base para cada grupo, la cual se entrenó  $\{svm\}$  especificando de manera iterativa el elemento objetivo y las variables correlacionadas a esta. Finalmente, con estos datos entrenados se realizó la predicción  $\{predict\}$  del modelo de bloques de cada variable para el método SVR.

Debido a que los modelos deben ser validados en la siguiente etapa, el análisis realizado a se remitió a una evaluación cualitativa del comportamiento de las variables por sobre una de tipo más cuantitativo.



Grupo 1: Los mapas obtenidos<sup>3</sup> mediante inverso de la distancia presentaron un rango más amplio (extremo) de valores para los datos estimados. Los modelos de Kriging presentaron valores de concentración altos de manera muy similar, pero con menor varianza (menos extremos), que los del método anterior, sobrestimando sus zonas de valores más bajos y subestimando las más altos, (efecto de suavizamiento). Finalmente, los resultados obtenidos mediante SVR mezclaron, de cierta forma, los casos anteriores generando un modelo con valores extremos, pero bastante densos, con fronteras de transición no tan fácilmente identificables, debido a su heterogeneidad. Los elementos se concentraron preferentemente en la zona central del relave, variando en magnitud (según sea el método) siendo el litio el único elemento que se destacó en altas cantidades en el lado oeste del relave. Concentraciones más bajas, para todos los elementos, se hallaron hacia la zona este del relave.

Grupo 2: Tal como para el grupo número uno, la heterogeneidad del modelo obtenido mediante SVR es evidente presentando también amplias zonas de baja concentración menos definidas. Respecto los sectores de mayor concentración, SVR mostró valores máximos superiores con respecto a Kriging o inverso de distancia cuyos modelos, bastante similares, carecen de valores tan altos y tienden a subestimar el resultado del SVR. Independiente de ello, la distribución de las zonas de mayor y menor concentración, en el macro, no distan de sobremanera entre un método u otro. Destaca el caso del plomo, elemento relevante a cualquier concentración que presento una distribución bastante densa localizada en el lado oeste del relave.

Grupo 3: La distribución para los 3 elementos es bastante similar entre métodos. los tres presentaron la concentración más alta al este del relave, en distintas concentraciones. El modelo producto de SVR resultó, nuevamente, más heterogéneo y con valores máximos superiores al resto de los métodos. Los resultados presentaron menor variabilidad para Kriging e inverso a diferencia de SVR que presenta transiciones (fronteras) de alta a baja concentración menos suaves. La porción de baja concentración varía poco entre modelos, aun así, es posible reconocer que existe una mayor variabilidad para el resultado de SVR en las zonas de baja concentración.

---

<sup>3</sup> *Todos los modelos de bloques mencionados anteriormente se presentan en la sección de anexos.*

Grupo 4: zonas bien definidas para los tres modelos. Las mayores concentraciones se ubican en la esquina sureste del relave y disminuyen a medida que se dirige en dirección noreste, hacia el centro del relave. Similitud entre los resultados de grupos anteriores, tanto para el Kriging como para inverso los valores extremos más altos, se presentan condensados/confinados en espacios pequeños (menos bloques/datos). Por el contrario, aquellos extremos más bajos se presentan como porciones más amplias (menos densas) y homogéneas, haciéndolas más distinguibles de las mencionadas inicialmente. Por otro lado, el efecto resulta inverso para los modelos obtenidos con SVR, ya que las zonas más extensas presentan una mayor variabilidad en sus valores, aislando y comprimiendo las zonas de valores extremos mínimos.

### Validación.

El proceso de validación o cuantificación del error para cada método, se realizó comparando los valores observados disponibles con los valores predichos por los modelos de Kriging, inverso de la distancia y SVR. La precisión de las predicciones se estableció en base al cálculo del error medio cometido por cada método y a la visualización de la dispersión de los datos predichos versus los observados (grafico de dispersión).

Los resultados obtenidos a partir de los cálculos de error medio cometido en la predicción de las 53 ubicaciones para cada elemento mediante SVM presentaron una disminución significativa con respecto a los otros dos métodos, mejoras en alrededor de un orden de magnitud menores. Esto demostró que la adición de información a través de otros factores, como la relación entre variables, produce una estimación mucha más precisa que sin esta. Tales, se resumen en las siguientes tablas.

### *Resumen de la media de los errores*

**Tabla 3: Resumen de errores medios para grupo 1**

<b>Grupo 1</b>	<b>Rango</b>	<b>Kriging O.</b>	<b>Inverso d.</b>	<b>SVR</b>
<b>Al (%)</b>	[0.5,4]	5.69E-01	4.26E-01	4.60E-02
<b>Cr (%)</b>	[2,6]E-04	9.93E-04	7.94E-04	9.04E-05
<b>Li (%)</b>	[6,43]E-04	7.70E-04	6.44E-04	8.08E-05
<b>Mg (%)</b>	[0.5,3]	4.73E-01	3.75E-01	3.60E-02
<b>V (ppm)</b>	[4,160]	1.55E+01	1.55E+01	1.69E+00

Tabla 4: Resumen de errores medios para grupo 2

Grupo 2	Rango	Kriging O.	Inverso d.	SVR
Pb (ppm)	[0,200]	4.72E+01	4.53E+01	4.61E+00
Ca (%)	[1,4.5]	8.82E-01	8.75E-01	8.30E-02
Zn (%)	[0,4.0]E-02	5.88E-03	5.88E-03	5.44E-04

Tabla 5: Resumen de errores medios para grupo 3

Grupo3	Rango	Kriging O.	Inverso d.	SVR
As (ppm)	[10,80]	1.43E+01	1.42E+01	2.10E+00
Hg (ppm)	[0,20]	2.75E+00	2.86E+00	4.50E-01
Au (ppm)	[0,1.4]	1.55E-01	1.53E-01	2.20E-02

Tabla 6: Resumen de errores medios para grupo 4

Grupo 4	Rango	Kriging O.	Inverso d.	SVR
Co (%)	[0,3]E-03	2.68E-04	2.79E-04	3.54E-05
Cu (%)	[0,1]	1.55E-01	1.55E-01	1.50E-02
K (%)	[0,0.7]	1.16E-01	1.18E-01	1.20E-02
Na (%)	[0,0.4]	5.90E-02	6.60E-02	8.20E-03
Ni (%)	[0,0.1]	1.92E-03	1.99E-03	1.77E-04
Fe (%)	[3.5,7.5]	8.15E-01	8.09E-01	9.40E-02

### *Gráficos de dispersión para cada elemento*

Se observa que los métodos de predicción/estimación de Kriging e Inverso de distancia tienden a sub y sobreestimar los valores reales de las variables, como ya se mencionó anteriormente. Esto provocó que los datos se desviaran de la relación lineal ideal ( $f(x) = x$ ) a causa de los valores extremos existentes, provocando una alta dispersión fácilmente visible en los gráficos.

En cambio, mediante el modelo de regresión SVM se observó que los resultados predichos se comportaron de manera similar a los valores observables traduciéndose esto en un resultado mayor para el coeficiente de correlación respectivo. Los valores asociados al coeficiente de correlación de cada caso son los siguientes<sup>4</sup>.

Tabla 7: Coeficientes de correlación para grupo 1

Grupo 1	KO.	ID.	SVR.
Al	0.275	0.171	0.996
Cr	-0.082	-0.264	0.990
Li	0.299	0.141	0.993
Mg	0.169	0.017	0.997
V	0.037	0.008	0.996

Tabla 8: Coeficientes de correlación para grupo 2

Grupo 2	KO.	ID.	SVR.
Pb	0.169	0.148	0.996
Ca	0.084	0.082	0.996
Zn	0.095	0.035	0.995

Tabla 9: Coeficientes de correlación para grupo 3

Grupo 3	KO.	ID.	SVR.
As	0.001	0.071	0.996
Hg	0.076	-0.117	0.996
Au	-0.013	-0.026	0.995

Tabla 10: Coeficientes de correlación para grupo 4

Grupo 4	KO.	ID.	SVR.
Co	0.548	0.148	0.995
Cu	0.037	-0.039	0.995
K	0.233	0.194	0.997
Na	0.396	0.312	0.996
Ni	0.057	-0.027	0.996
Fe	0.253	0.264	0.996

<sup>4</sup> Los gráficos de dispersión correspondientes se encuentran en la sección B de anexos

### Agrupamiento.

Debido a la buena precisión y baja dispersión de los resultados de validación obtenidos anteriormente, el modelo SVM fue seleccionado y utilizado para realizar el procedimiento de clustering mediante el método de k-medias.

Para definir el número óptimo de clúster para cada grupo, se utilizó el método del codo o *“elbow method”* el cual muestra por medio de un gráfico la suma interna de cuadrados (intracluster) para cada número de clúster realizado. Esta medida establece el grado de relación/similitud entre las variables consideradas, en particular que tan parecidas/similares son las leyes dentro de cada clúster.

Para una correcta clasificación de datos esta medida debe ser minimizada; minimizarla a cero implicaría que se generen tantos clústeres como datos contenga la base como se puede observar en los respectivos gráficos [Revisar sección C de anexos], es por esto que se escogieron como mínimo local 2 puntos para cada grupo, donde la pendiente del gráfico cambiase de manera más notoria, los que permitieron generar 2 modelos de clustering. Para establecer cuál de las dos modelos representaba un mejor agrupamiento para cada grupo se analizaron cada uno de ellos por medio de los boxplots resultantes, de los que se extrajo información como la media y el rango intercuartílico para estudiar su desempeño de manera que los clústeres no resultaran redundantes entre ellos y demostrasen que los elementos contenidos en ellos eran los más similares entre sí y los más disimiles con el resto.

Como resultado los elementos de los grupos 1, 2 y 3 fueron agrupados en 3 clúster cada uno, mientras que para los del grupo 4 se estableció que 4 particiones era la cantidad óptima. Los resultados de los agrupamientos, en cuanto a concentración (leyes medias) y cantidad de datos para de cada clúster se resumen en la ilustración siguiente.

Tabla 11: Resumen resultado de clustering grupo 1

Grupo 1	cluster 1	cluster 2	cluster 3
Al (%)	2.091	1.819	2.353
Cr (%)	0.00378	0.00349	0.00398
Li (%)	0.00273	0.00197	0.00300
Mg (%)	1.828	1.550	1.936
V (ppm)	95.774	85.664	103.032
datos	617	132	603

Tabla 12: Resumen resultado de clustering grupo 2

Grupo 2	cluster 1	cluster 2	cluster 3
Pb (ppm)	50.052	40.214	88.581
Ca (%)	2.019	1.623	2.590
Zn (%)	0.020	0.018	0.023
datos	185	957	211

Tabla 13: Resumen resultado de clustering grupo 3

Grupo 3	cluster 1	cluster 2	cluster 3
As (ppm)	43.342	55.586	24.697
Hg (ppm)	4.869	8.244	2.621
Au (ppm)	0.347	0.457	0.252
datos	181	104	1067

Tabla 14: Resumen resultado de clustering grupo 4

Grupo 4	cluster 1	cluster 2	cluster 3	cluster 4
Co (%)	0.00183	0.00214	0.00203	0.00165
Cu (%)	0.304	0.298	0.381	0.242
K (%)	0.183	0.236	0.269	0.136
Na (%)	0.116	0.175	0.136	0.093
Ni (%)	0.00033	0.00385	0.00468	0.00259
Fe (%)	5.157	5.411	6.014	4.823
datos	610	168	155	419

Con los resultados de las tablas anteriores, se pudo finalmente establecer una descripción del relave en términos de la cantidad y calidad de los elementos contenidos.

Cada clúster producido poseía una ley media y una cantidad definida de bloques para cada elemento/variable. A su vez, cada bloque representaba un volumen de 25 m<sup>3</sup> (5x5x1 m<sup>3</sup>) con una densidad definida de 2.65 ton/m<sup>3</sup>, por lo que cada sitio donde se estimó representó un total de 66.25 toneladas. Ese valor multiplicado por la cantidad de datos agrupados en cada clúster estableció el tonelaje de cada partición generada por el algoritmo de k-medias.

Finalmente, las leyes medias de cada elemento fueron ponderadas con los tonelajes correspondientes y sumadas para establecer una cantidad de toneladas final para cada elemento dentro del relave

Tabla 15: Resumen de cuantificación final para los elementos en el relave

Elemento	Toneladas totales	
	Estimadas	
Al		1916.924
Cr		3.375
Li		2.441
Mg		1624.837
V		861.435
Pb		431.826
Ca		1607.800
Zn		16.647
As		259.854
Hg		29.479
Au		2.465
Co		1.612
Cu		191.223
K		125.298
Na		78.748
Ni		1.023
Fe		3241.742

## 7. Conclusiones y recomendaciones

El método de regresión mediante SVM logró cumplir con el objetivo propuesto de modelar y cuantificar las variables de interés (elementos) minimizando el error de la predicción hecha, viéndose disminuido en alrededor de un orden de magnitud con respecto de los otros dos métodos propuestos (Kriging e inverso de la distancia). Esto también se evidenció cuando se analizaron los gráficos de dispersión/correlación donde la precisión para los 17 elementos, cuantificable por medio del coeficiente de correlación, alcanza valores muy cercanos a lo ideal ( $R=1$ ).

La razón de esta diferencia en los resultados se explica en la información de sustento (factores) de cada estimador; Tanto K.O. como I.D. se basan en la existencia de una continuidad espacial de la variable, factor no observable en el presente caso, ya que este presenta comportamientos altamente pepiticos/erráticos (aleatorios) comportamiento confirmado por los variogramas de cada grupo. En cambio, el método por SVM complementa su predicción en el comportamiento entre variables, dicho de otra manera, en la correlación entre ellas. En consecuencia, la clasificación de las variables en grupos según la similitud de su distribución beneficia el uso del SVR para lograr una predicción con un menor error que los métodos anteriores.

Definido el modelo más óptimo, el algoritmo de clustering de K-medias posibilitó la sectorización de las variables facilitando el cálculo de las concentraciones promedio de los elementos involucrados además de sus localizaciones preferentes, generando así porciones de mayor o menor interés para cualquiera de las variables objetivo que componen el presente relave. Todo este proceso definió los dominios geológicos del relave y mostró cómo se comportan los elementos en distintos sectores del relave, si existen asociaciones minerales potenciales, y si es posible encontrar elementos tóxicos asociados a los elementos de interés económico, por ejemplo.

Los resultados obtenidos de esta etapa previa son cruciales para establecer métodos de acción con respecto al relave. Gran importancia tuvo el análisis de la presencia y cantidad de los elementos nocivos/peligrosos: Pb, As, Hg.

Los dos primeros se presentan como los más abundantes dentro de los elementos potencialmente tóxicos (potencialmente por su concentración) y en conjunto representan un estimado de 850 toneladas (260 ton para arsénico y 500 ton para plomo). Los clúster generados muestran concentraciones promedio bastante superiores a los límites permisibles sugeridos por organizaciones internacionales para estos elementos en agua potable [48], por lo que la migración



de estas partículas se presenta como un factor de riesgo tanto para población como para el medioambiente considerando que el relave se localiza en plena costa, de manera contigua al mar.

El mercurio, aunque se presentó en menores cantidades que los dos elementos anteriores, de igual manera sobrepasó los límites mencionados en cada uno de los clústeres obtenidos, por lo que debe ser considerado igualmente como un elemento peligroso ante la posibilidad de migración de este material particulado hacia el medio que lo rodea.

Por otro lado, destacables fueron las reservas que presentaron elementos potencialmente rentables, siendo el Al, Mg, Ca, Fe los elementos más abundantes dentro del relave. Cada uno de ellos superó las 1000 toneladas según la estimación hecha, presentado al relave como una fuente de materia prima potencialmente barata para la industria del cemento (elementos esenciales en la fabricación de Clinker) siempre y cuando la mineralogía lo permita.

En cantidades menores, mas no menos relevantes, se encontraron dos elementos altamente valiosos: Au, Cu; cuyas reservas podrían justificar/financiar, planes de remediación o extracción del relave en cuestión, complementado por la explotación del anterior grupo. Para generar una cifra económica alrededor de estos metales, se llevó a cabo un estimado de manera rápida del beneficio que pueden llegar a generar el oro y cobre presente en el relave.

El oro estimado constituye una reserva de aproximadamente 2.5 toneladas. La literatura respecto la recuperación de este metal contenido en relaves de sulfuros [49] indicaron que esta puede alcanzar valores de hasta un 98% (mediante concentración con carbón activado). Tomando como referencia un precio de 1472 USD/oz [50] y una recuperación máxima/ideal >90-95%, según lo descrito, dichas 2.5 toneladas podrían llegar a representar un ingreso de alrededor de 127 MUSD<sup>5</sup>.

Para el caso del cobre se estimó una reserva de 173 toneladas, cuyas recuperaciones para casos provenientes de relaves de menas sulfuradas alcanza valores de entre 80-95% [51, 52]. Al repetir el ejercicio previo, ahora para el cobre, se obtuvo que a un precio de 2.65 USD/lb [50] y considerando una recuperación superior al 90%, la extracción del cobre generaría una suma de 900 mil USD, cifra inferior en comparación con la del oro, aun así, bastante alta si se considera que se están obteniendo a partir de un desecho en desuso como un relave abandonado.

---

<sup>5</sup> Sin considerar costos asociados

Si bien estas cifras solo hablan de beneficios (utilidades sin costos) no está demás señalar que los costos asociados a la obtención de estos minerales deben considerar que el material ya sufrió una reducción de tamaño considerable, lo que podría repercutir en un menor costo de procesamiento.

Como recomendación y a pesar de ser valores obtenidos de manera poco rigurosa, se debe tener en cuenta que el beneficio tampoco fue calculado para los elementos más abundantes que resultan ser de potencial económico aumentando tanto el valor del relave como los costos asociados posiblemente, sin embargo su procesamiento podría, eventualmente, costear la extracción y tratamiento de aquellos elementos potencialmente nocivos generando así no solo un beneficio económico sino que también un beneficio social y medioambiental eliminando este factor de riesgo que es la exposición de la población y del ambiente a los elementos tóxicos contenidos en el relave, transformando la minería en una actividad económica más sustentable y de esta manera aprovechar el gran potencial que poseen todos los relaves dispuestos en el territorio nacional. Recordar que a mayor antigüedad del relave mayor es la posibilidad de encontrar altas leyes de minerales rentables debido a la ineficiencia de los procesos del pasado.

Debido a que la metodología presentada es perfectible, validable y replicable, basta con aumentar la cantidad disponible de información mediante la realización de nuevos sondeos para comparar las estimaciones con la nueva información recogida

Se recomienda también la elaboración de estudios posteriores que pudieran enfocarse en la generación de un modelo mineralógico que permita estudiar la forma real en la que se encuentran los elementos más abundantes y así confirmar o rechazar la revalorización del pasivo ambiental mediante un estudio más acabado y detallado del potencial económico que puede representar el relave SOTRAMIN y Playa atacama.

## 8. Referencias

- [1] Ministerio de Minería de Chile. Historia de la minería en Chile, 2019. [En línea] <http://www.minmineria.gob.cl/que-es-la-mineria/historia-de-la-mineria-en-chile>
- [2] Corporacion Chilena del Cobre (COCHILCO), "Producto Interno Bruto por clase de actividad económica a precios corrientes", 2019. [En línea] <http://www.cochilco.cl:4040/boletin-web/pages/tabla13/buscar.jsf>
- [3] Corporacion Chilena del Cobre. "Exportaciones por tipo de producto y empresa", 2019. En: <http://www.cochilco.cl:4040/boletin-web/pages/tabla23/buscar.jsf>
- [4] Sociedad nacional de Minería (SONAMI). "Ocupación minera crece 10% en doce meses", abril 2019. En: <https://www.sonami.cl/v2/noticias/ocupacion-minera-crece-10-en-doce-meses/>
- [5] Corporación Chilena del Cobre (COCHILCO). "Catastro nacional de tranques de relaves", 2006-2017.
- [6] C. Aguirre. "Tratamiento de relaves mineros: recuperación de valores y estabilización ambiental del sitio intervenido", agosto 2017.
- [7] A. Oblasser, E. Chaparro. "Estudio comparativo de la gestion de los pasivos ambientales mineros en bolivia, Chile, Peru y Estados Unidos", mayo 2018.
- [8] Ley n°20.551, Diario oficial de la República de Chile, Santiago de Chile, 11 de noviembre de 2011.
- [9] Ministerio de Minería de Chile. "REGLAMENTO PARA LA APROBACION DE PROYECTOS DE DISEÑO, CONSTRUCCION, OPERACIÓN Y CIERRE DE DEPOSITOS DE RELAVES" abril 2007.
- [10] SERNAGEOMIN-BGR. "Manual de evaluación de riesgos de faenas mineras abandonadas o paralizadas (FMA/P)", 2008.

- [11] G. Medvinsky-Roa, V. Caroca, J. Vallejos. "Informe sobre la situación de los relaves mineros en Chile para ser presentados en el cuarto informe periódico de Chile para el Comité de derechos Económicos, Sociales y Culturales perteneciente al Consejo Económico-Social de las Naciones Unidas", Fundación Relaves Chile y fundación Terram, 2015.
- [12] Ministerio Secretaria General de la Presidencia. D.S n°8 "Zona declarada saturada por material particulado respirable MP10, como concentración de 24 horas y como concentración anual a la localidad de Andacollo y sectores aledaños", 2009.
- [13] Ministerio de minería de Chile. D.S. n°59 "Plan de Descontaminación Atmosférica para la localidad de Andacollo y sectores aledaños", diciembre 2014.
- [14] Tribunal Ambiental de Chile. "tribunal ambiental de Santiago condenó a minera tambillos a reparar daño ambiental generado tras colapso de tranque de relave en maule", septiembre 2016. [En línea]: <https://www.tribunalambiental.cl/tribunal-ambiental-de-santiago-condeno-a-minera-tambillos-a-reparar-dano-ambiental-generado-tras-colapso-de-tranque-de-relave-en-maule>
- [15] Servicio Nacional de Geología y Minería (SERNAGEOMIN). "Informe por afectación de lluvias y aluviones en zona norte del país", marzo 2015.
- [16] Antofagasta Minerals S.A. "Juicios contra el tranque El Mauro de minera los Pelambres", 2016.
- [En línea]: <http://www.aminerals.cl/comunicaciones/noticias/2016/especial-juicios-contra-el-tranque-el-mauro-de-minera-los-pelambres/>
- [17] Corporación Nacional del Cobre (CODELCO). "Codelco tendrá la primera minicentral eléctrica del mundo a partir de relaves", 2016. [En línea]: <http://tiny.cc/268siz>
- [18] Alfonso A. Romero y Silvana L. Flores (2010). "Reutilización de relaves mineros como insumo para la elaboración de agregados de construcción para fabricar ladrillos y baldosas", Revista de la Facultad de ingeniería Industrial, 13(1); pp. 75-82, 2010.

- [19] Alfonso A. Romero y Silvana L. Flores. "Caracterización de la pasta de relave para su uso como relleno en labores mineras", Escuela de Ingeniería de Minas de la Universidad Nacional Mayor de San Marcos, 2012.
- [20] Obinna Onuaguluchi y Ozgur Eren. "Recycling of copper tailings as additive in cement mortars". *Construction and building materials*, 37, pp 723-727, 2012.
- [21] Castañeda S., Leazama J., Ampuero J. "Sistema de relleno con mortero de relave para mejorar la confiabilidad del sostenimiento en la minería subterránea". *Sinergia e innovación*, 3(2), pp. 17-41, 2015.
- [22] F.L da Silva, F.G.S Araújo, M.P Teixeira, R.C. Gomes, F.L. von Kruger. "Study of the recovery and recycling of tailings from the concentration of iron ore for the production of ceramic". *Ceramics international*, 40, pp 16085-16089, 2014.
- [23] J.L. Barros Galvão, H. Días Andrade, G.J. Brigolini, R.A. Fiorotti Peixoto, J.C Mendes. "Reuse of iron ore tailings from tailings dams as pigment for sustainable paints". *Journal of Cleaner Production*, 200, pp 412-422, 2017.
- [24] Xiaoyan Huang, Ravi Ranade, Wen Ni, Victor C. Li. "Development of green engineered cementitious composites using iron ore tailings as aggregates", *Construction and building materials*, 44, pp. 757-764, 2013.
- [25] J.P. Chilés & P. Delfine. "Geostatistics Modeling Spatial Uncertainty 2nd edition", Capítulos 2,3 y 7.
- [26] R. Fustos. "Estimación de Recursos Minerales: cuantificación de recursos minerales y análisis geoestadístico de datos" Facultad de ingeniería, departamento de ingeniería metalúrgica, Universidad de Concepción, 2016.
- [27] X. Emery. "Clases 1-8 de evaluación de yacimientos MI-5041" Departamento de ciencias físicas y matemáticas, Universidad de Chile, 2017.

- [28] Chih-Wei Hsu, Chih-Ching Chang, Chih-Jen Lin. "A practical guide to support vector classification", Department of computer science, National Taiwan University, mayo 2016.
- [29] Alexandre Kowalczyk. "Support Vector Machines", 2014.
- [30] Australasian Institute of Mining and Metallurgy. "Australian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves The JORC Code", edition 2012.
- [31] Lance G., Williams T. "A General Theory of Classificatory, Sorting Strategies: II. Clustering systems" Computer journal 10, pp. 271-277, 1967.
- [32] Anil K. Jain, Richard C. Dubes. "Algorithms for Clustering data; Chapter 3: Clustering methods and algorithms", pp. 55-142, 1988.
- [33] Amit Saxena, Mukesh Prasad, . "A review of clustering techniques and developments" Neurocomputing Volume 267, Pages 664-681, 6 December 2017. DOI: doi.org/10.1016/j.neucom.2017.06.053
- [34] D. Lam and D.C. Wunsch. "Clustering", Academic Press Library in Signal Processing, signal processing theory and machine learning vol 1, 2014.
- [35] J.B. McQueen, "Some Methods for Classification and Analysis of Multivariate Observations", 5th Symposium in Mathematical statistics and Probability, Berkeley, University of California Press, vol. 1, pp 281-297, 1967.
- [36] Alsabti K., Ranka S. y Singh V., "An efficient k-means clustering algorithm". Electrical engineering and computer science, 43, 1977.
- [37] D. Judd, P. McKinley and Anil J. "Large-scale parallel data clustering". Proc. Int'l Conference on pattern recognition, august 1996.
- [38] V. Ramasubramanian and K. Paliwal. Fast K. "Dimensional tree algorithm for nearest neighbor search with supplication to vector quantization encoding", IEEE Transactions on signal processing, 40(3), march 1992.

- [39] Thanh Tran, Klaudia Drab, Michal Daszykowski, "Revised DBSCAN algorithm to cluster data with dense adjacent clusters", *Chemometrics and Intelligent Laboratory Systems* 120, pp. 92-96, 2013.
- [40] K. Mahesh, A. Rama, "A fast DBSCAN Clustering algorithm by accelerating neighbor searching using groups methods", *Pattern Recognition* 58, pp. 39-48, 2016.
- [41] Anil K. Jain, Richard C. Dubes. "Algorithms for Clustering data; Chapter 4: Cluster validity", pp. 143-222, 1988.
- [42] Hubert L., Schults J. "Quadratic assignment as a general data analysis strategy". *British journal of Mathematical and Statistical Psychology* 29, pp. 190-241, 1976.
- [43] Y. Theodoridis y K. Koutroubas. *Pattern Recognition "Spatial Datasets: An unofficial collection"* Academic Press 1999.
- [44] D.L. Davies and D.W. Bouldin. "Cluster Separation measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 1., N°2, pp. 95-104, 1979.
- [45] M. Halkidi, Y. Batistakis, M. Vazirgiannis. "On clustering validation techniques", *Journal of Intelligent Information Systems*, 17:2/3, pp. 107-145, 2001.
- [46] L. Keeney. "The development of a novel method for integrating geometallurgical mapping and orebody modelling", Ph.D. thesis, Universidad de Queensland, 2010.
- [47] R. Babak, O. Asghari. "Geometallurgical Domaining by Cluster Analysis: Iron Ore Deposit Case Study", *Natural Resources Research*, October 2018, pp. 1-20, 2018.
- [48] EPA (Environmental Protection Agency). "Gold Book of Quality Criteria for Water". EPA 440/5-86-001. EUA. 477 p. 1986.
- [49] A. Dehghani, M. Ostad-Rahimi, S.H. Mojtahedzadeh. "Recorey of gold from the Mauteh Gold Mine tailings dam", *Journal of southern african institute of mining and metallurgy*, 109(7), pp. 417-421, june 2009.

- [50] Comision Chilena del Cobre. "Informe semanal del mercado internacional de metales basado en estadísticas de la Bolsa de Metales de Londres", enero 2020. [En línea]: <http://www.portalminero.com/wp/bolsa-de-metales-2/>
- [51] M. Antonijevic, M. dimitrujevic, Z. Stevanovic, S. Serbula, G. Bogdanovic. "Investigation of the possibility of copper recovery from the flotation tailings by acid leaching", Journal of hazardous materials 158, pp. 23-34, 2008.
- [52] Tao Chen, Chang Lei, Bo Yan, Xianming Xiao. "Metal recovery from copper sulphide tailing with leaching and fractional precipitation technology", Hydrometallurgy, 2014. DOI: 10.1016/j.hydromet.2014.05.018





## 9. Anexos

- Anexos sección A

En el presente se muestran los resultados del modelamiento mediante los distintos métodos mencionados en el capítulo de resultados. Las imágenes corresponden a una vista de planta, con norte hacia la parte superior (arriba) que permiten el análisis de la variabilidad/varianza en la distribución y concentración de cada elemento.

Tabla 16: Modelos de bloques para grupo 1

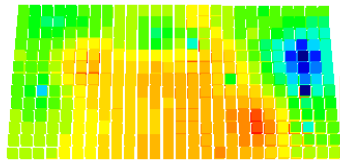
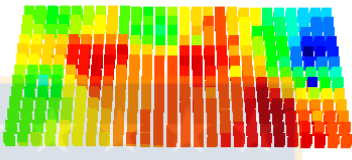
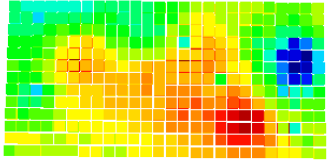
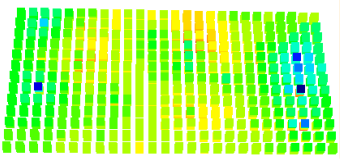
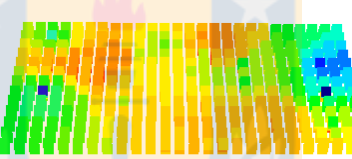
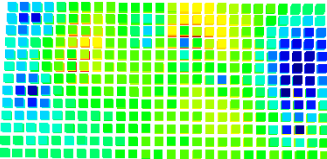
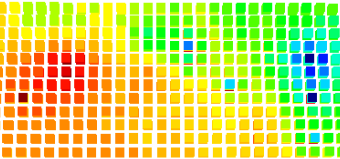
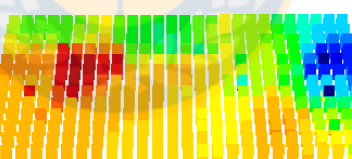
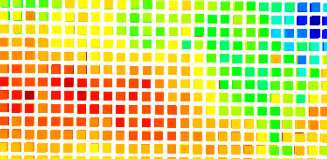
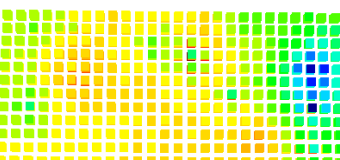
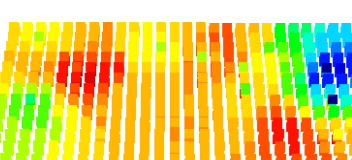
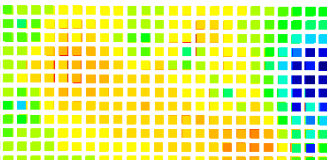
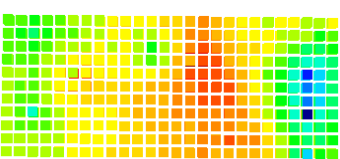
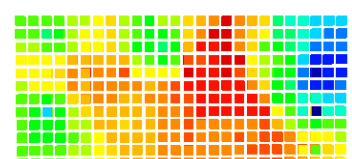
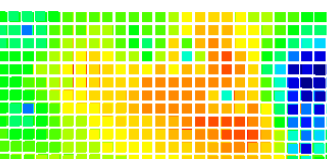
Grupo 1	Kriging O.	Inverso d.	SVR
Al			
Cr			
Li			
Mg			
V			

Tabla 17: Modelos de bloques para grupo 2

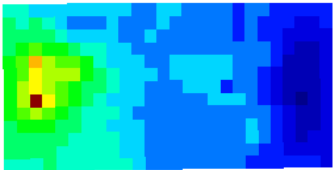
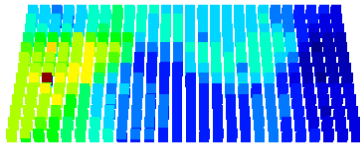
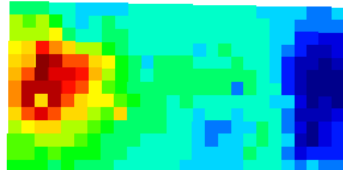
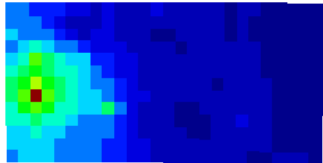
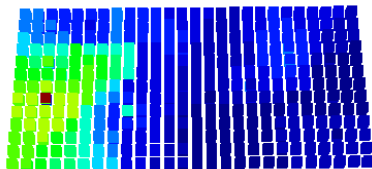
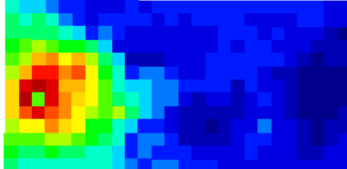
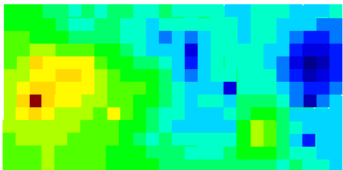
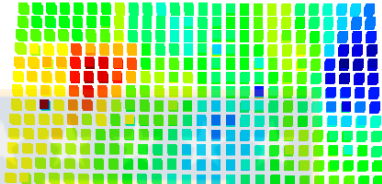
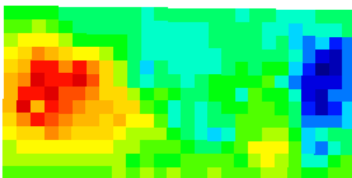
Grupo 2	Kriging O.	Inverso d.	SVR
Pb			
Ca			
Zn			

Tabla 18: Modelos de bloques para grupo 3

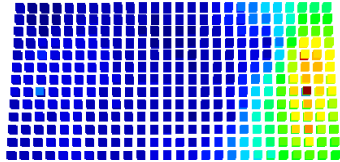
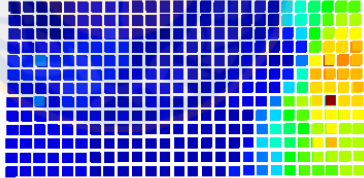
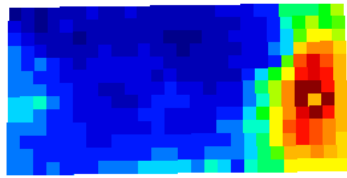
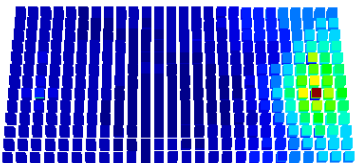
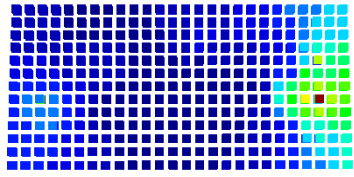
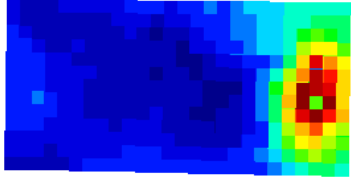
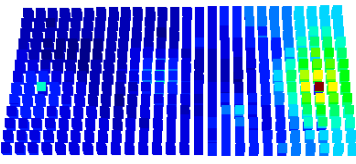
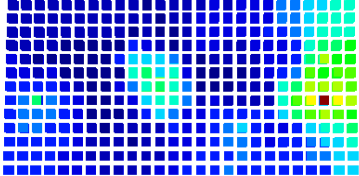

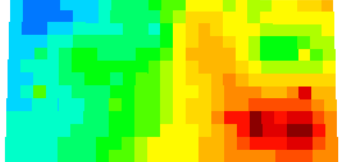
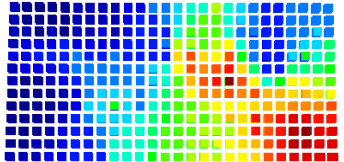
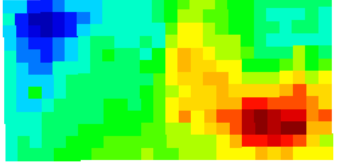
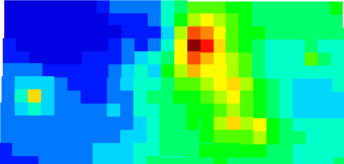
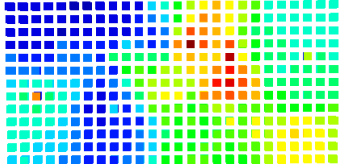
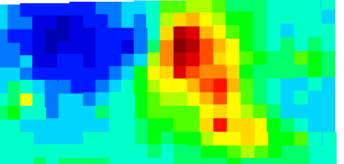
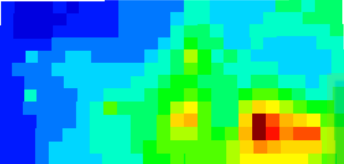
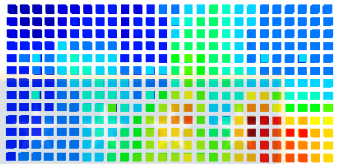
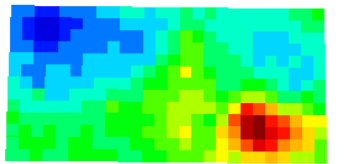
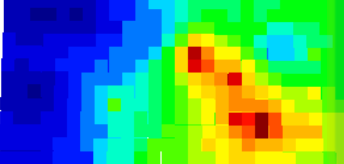
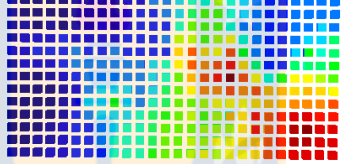
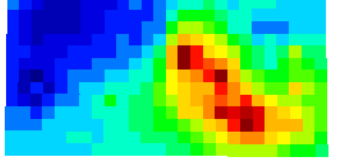
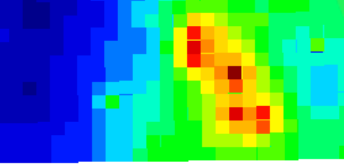
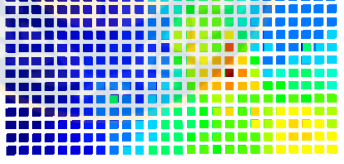
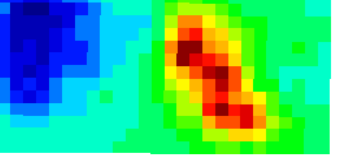
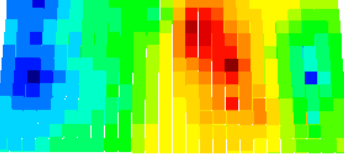
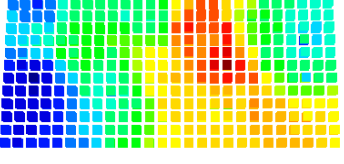
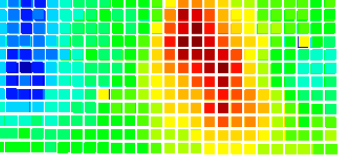
Grupo3	Kriging O.	Inverso d.	SVR
As			
Hg			
Au			

Tabla 19: Modelos de bloques para grupo 4

Grupo 4	Kriging O.	Inverso d.	SVR
Co			
Cu			
K			
Na			
Ni			
Fe			

### Anexos sección B (validación)

A continuación, se presentan los gráficos de dispersión resultado del proceso de validación cruzada para los tres métodos vistos en el presente estudio.

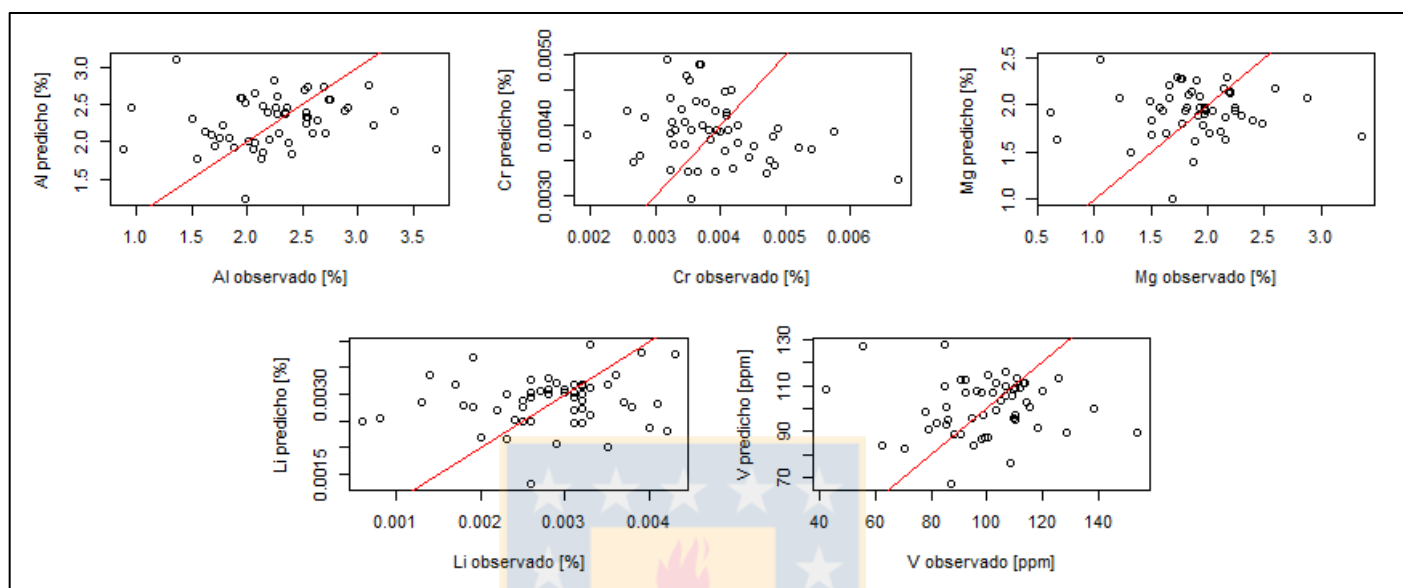


Ilustración 9.1: Gráficos de dispersión para inverso de distancia, grupo 1.

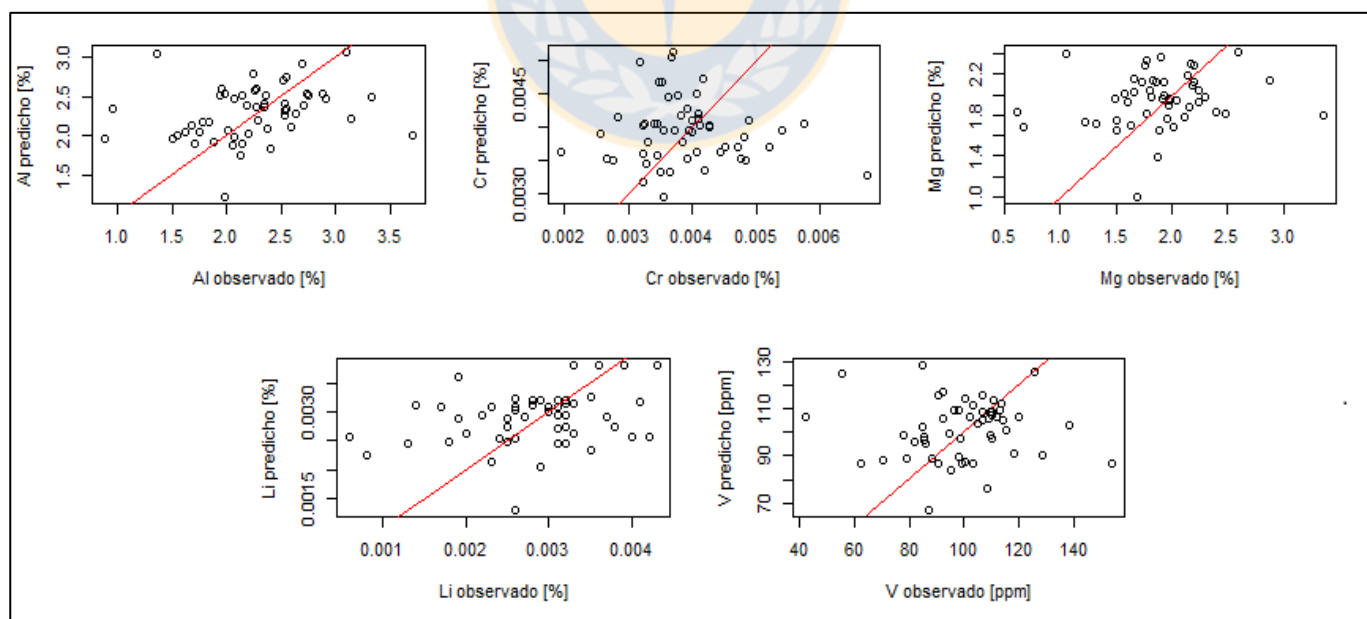


Ilustración 9.2: Gráficos de dispersión para Kriging, grupo 1.

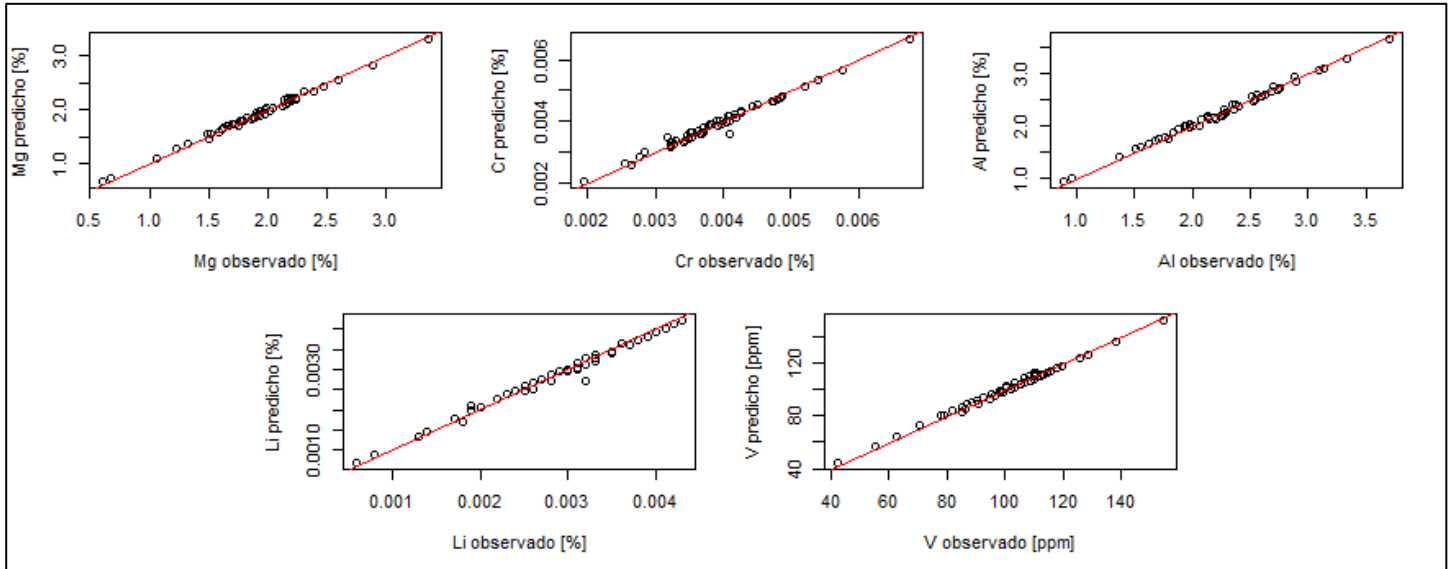


Ilustración 9.4: Gráficos de dispersión para Support Vector Regression, grupo 1.

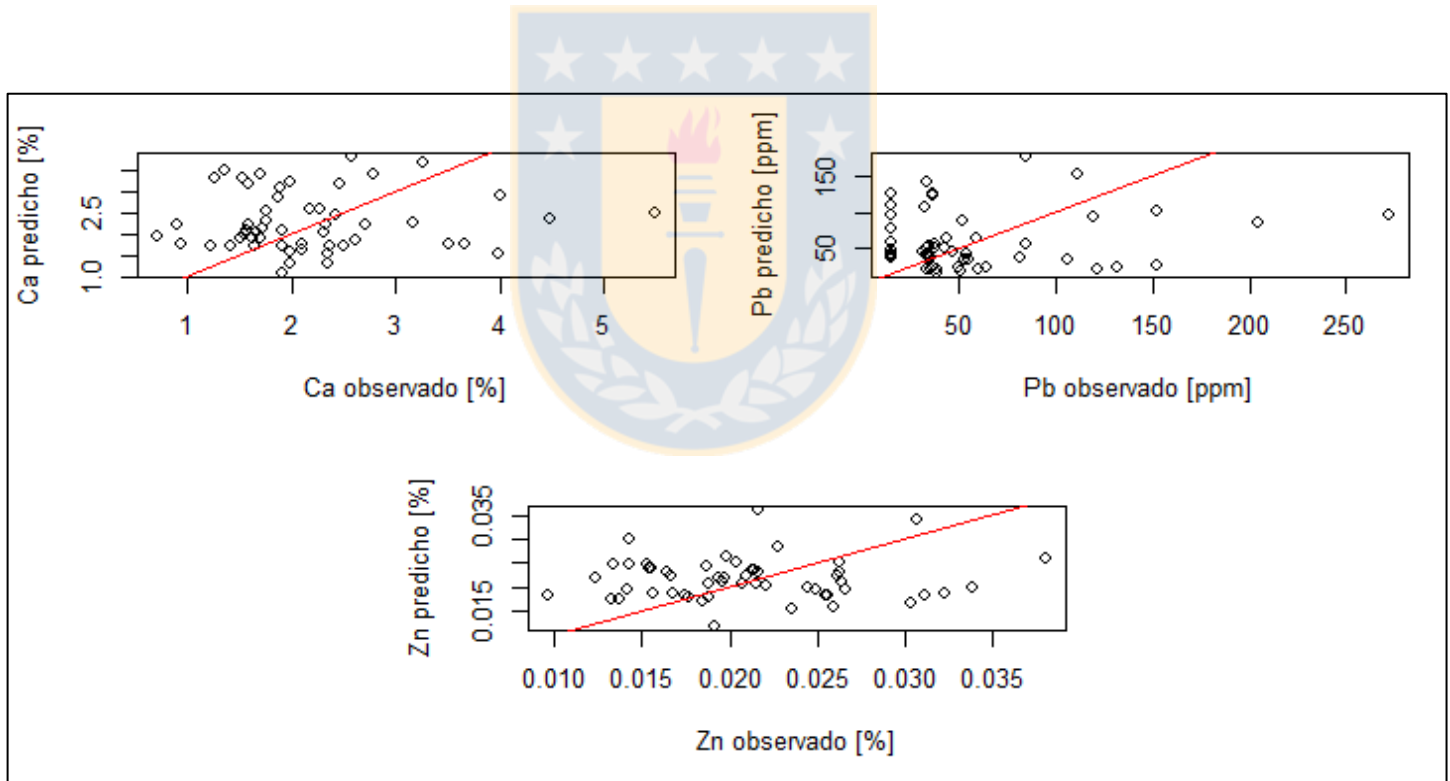


Ilustración 9.3: Gráficos de dispersión para inverso de distancia, grupo 2.

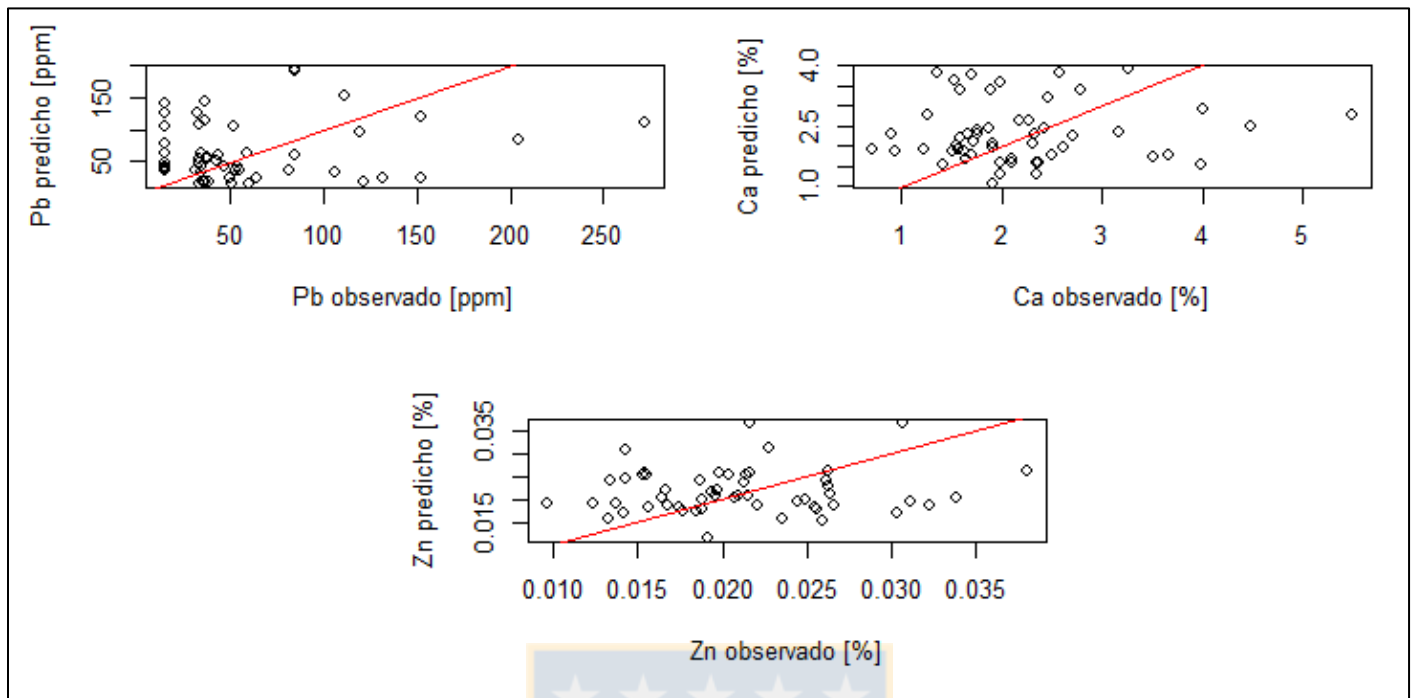


Ilustración 9.5: Gráficos de dispersión para Kriging, grupo 2.

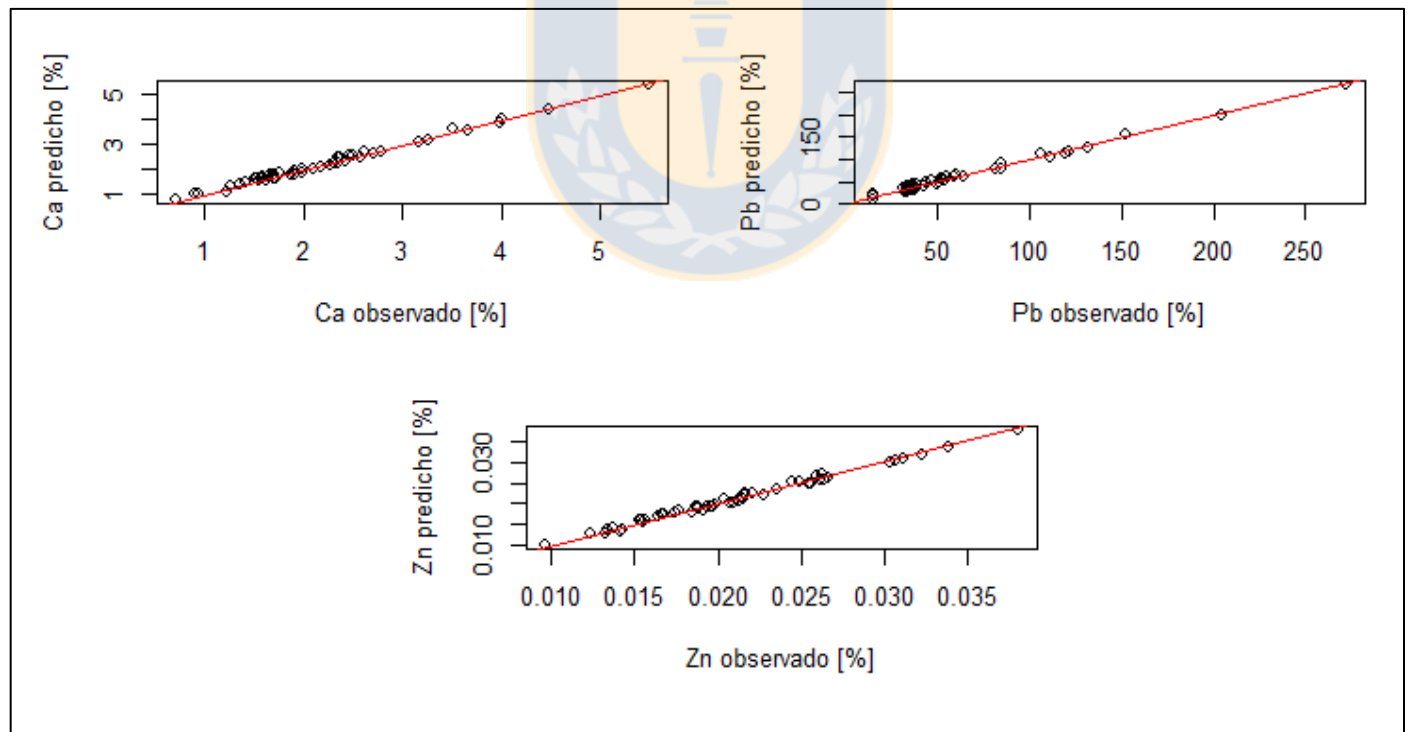


Ilustración 9.6: Gráficos de dispersión para Support Vector Regression, grupo 2.

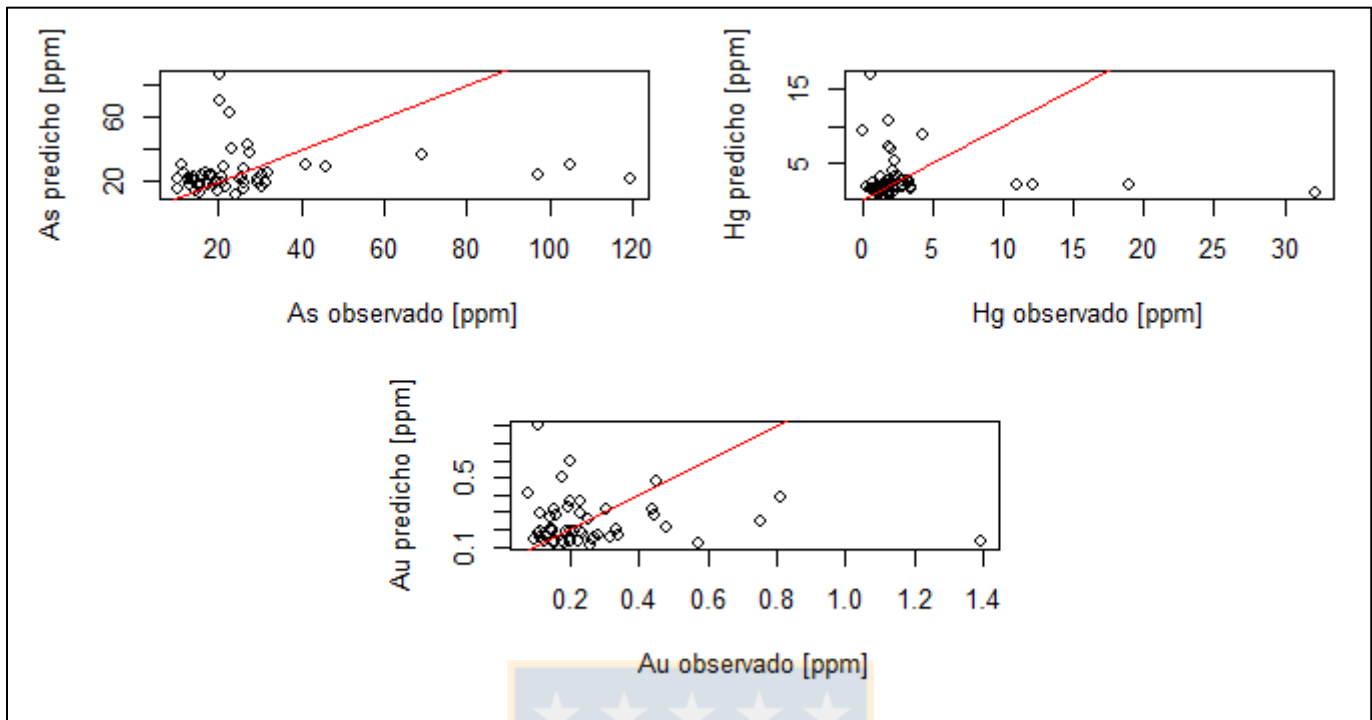


Ilustración 9.8: Gráficos de dispersión para Inverso de distancia, grupo 3.

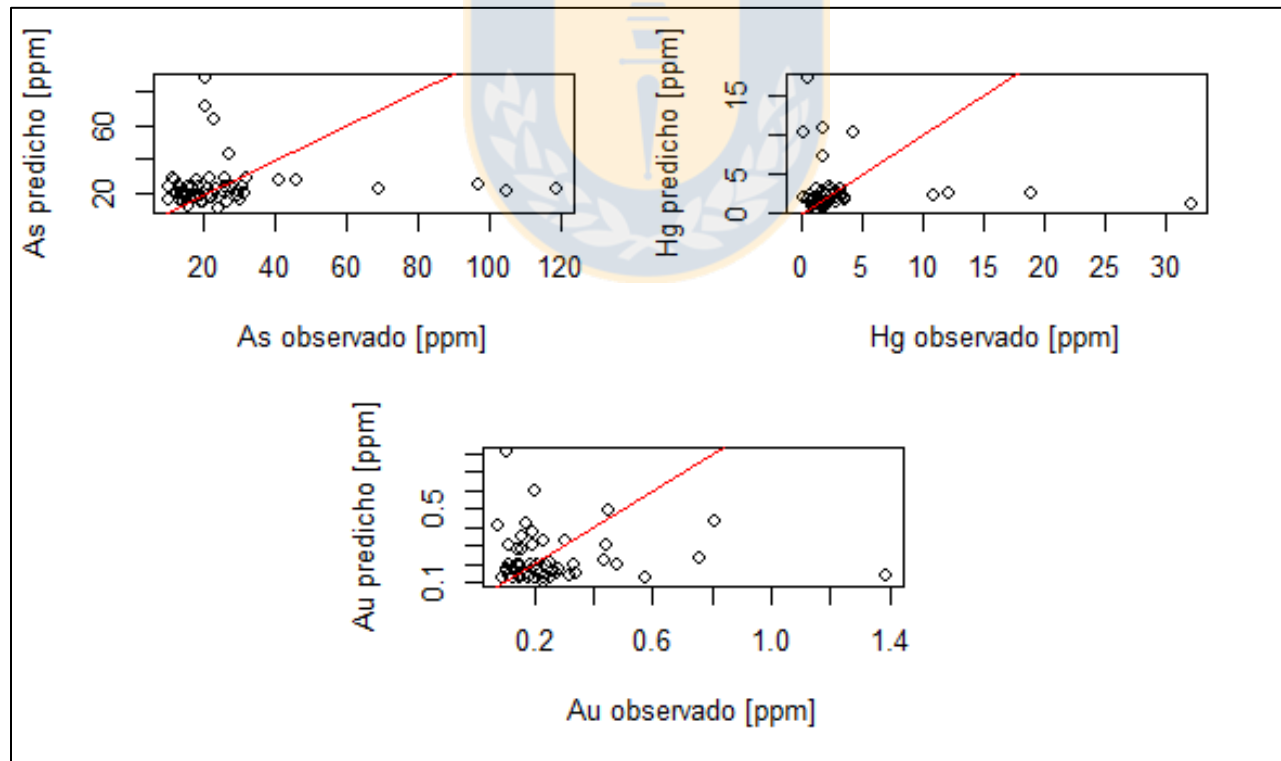


Ilustración 9.7: Gráficos de dispersión para Kriging, grupo 3.

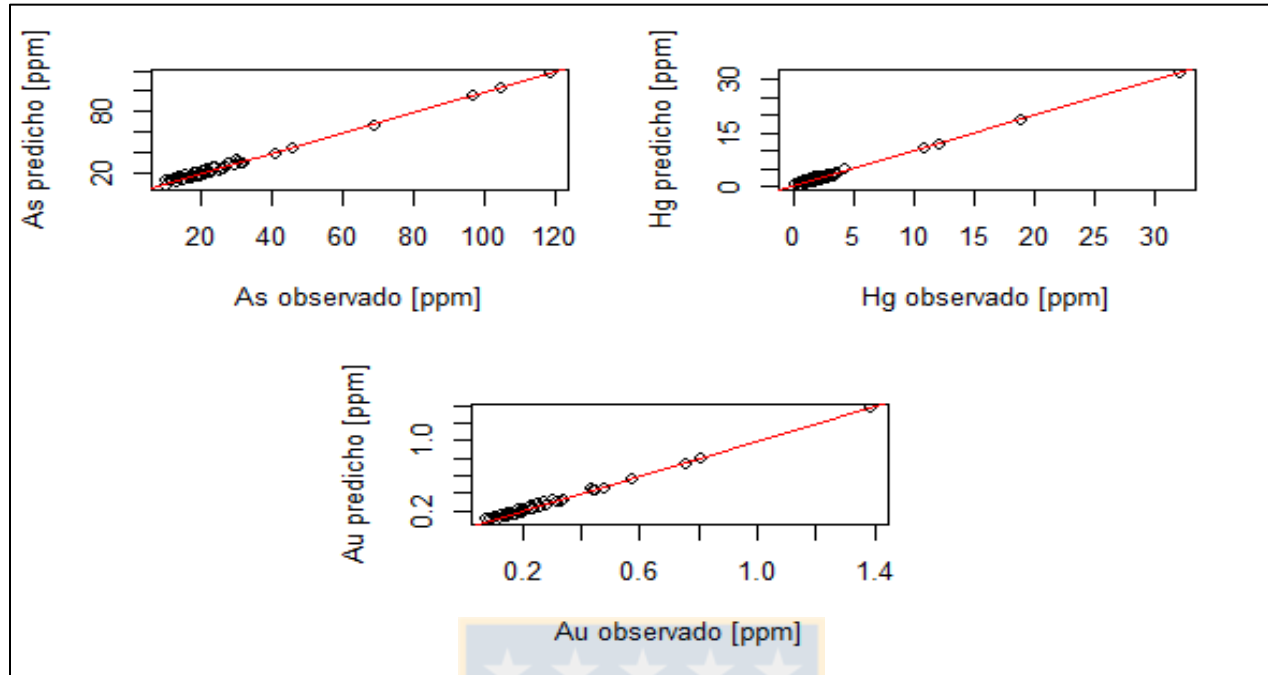


Ilustración 9.9: Gráficos de dispersión para Support Vector Regression, grupo 3.

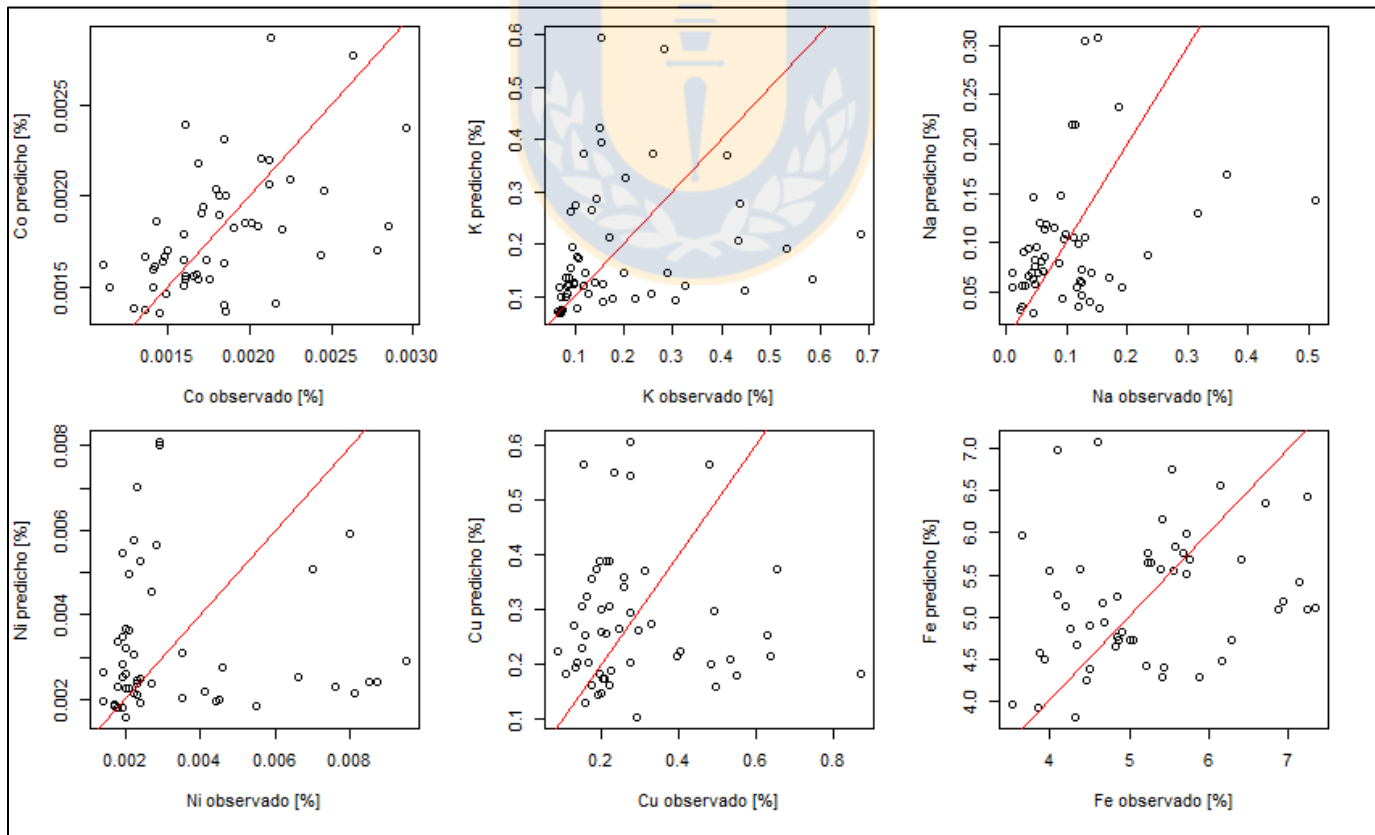


Ilustración 9.10: Gráficos de dispersión para Inverso de distancia, grupo 4.



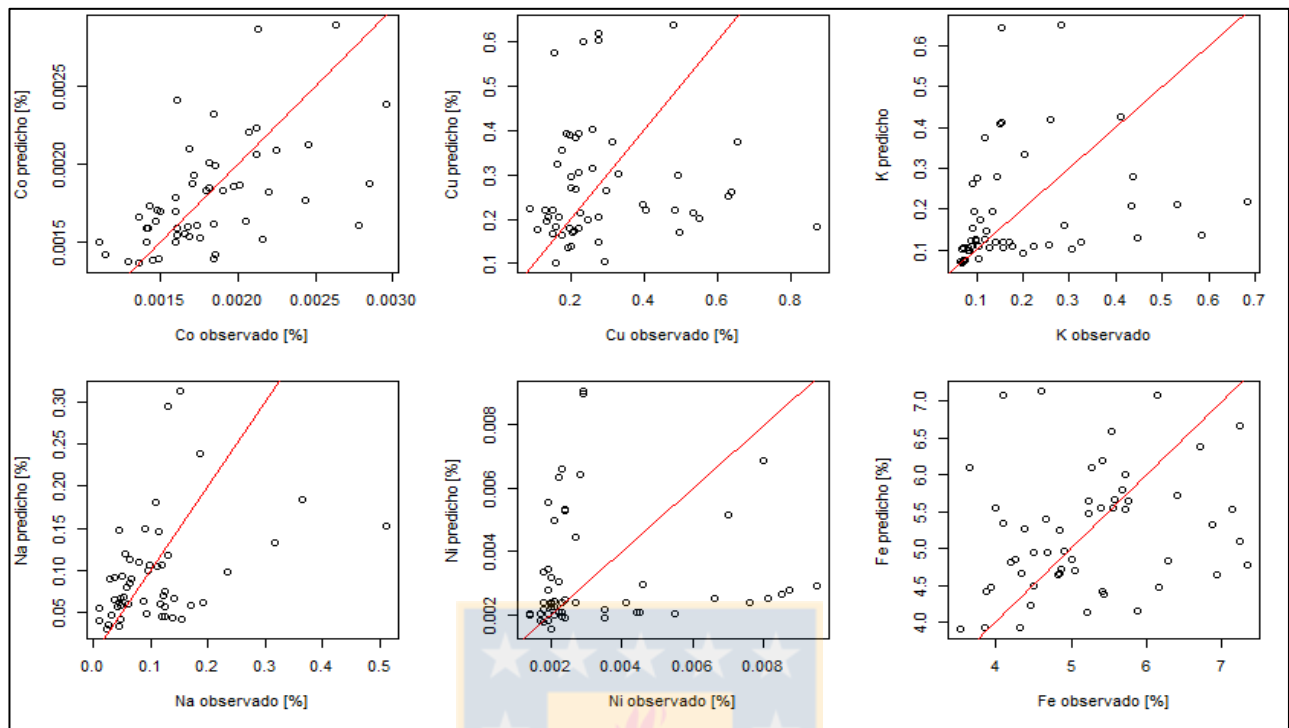


Ilustración 9.11: Gráficos de dispersión Kriging, grupo 4.

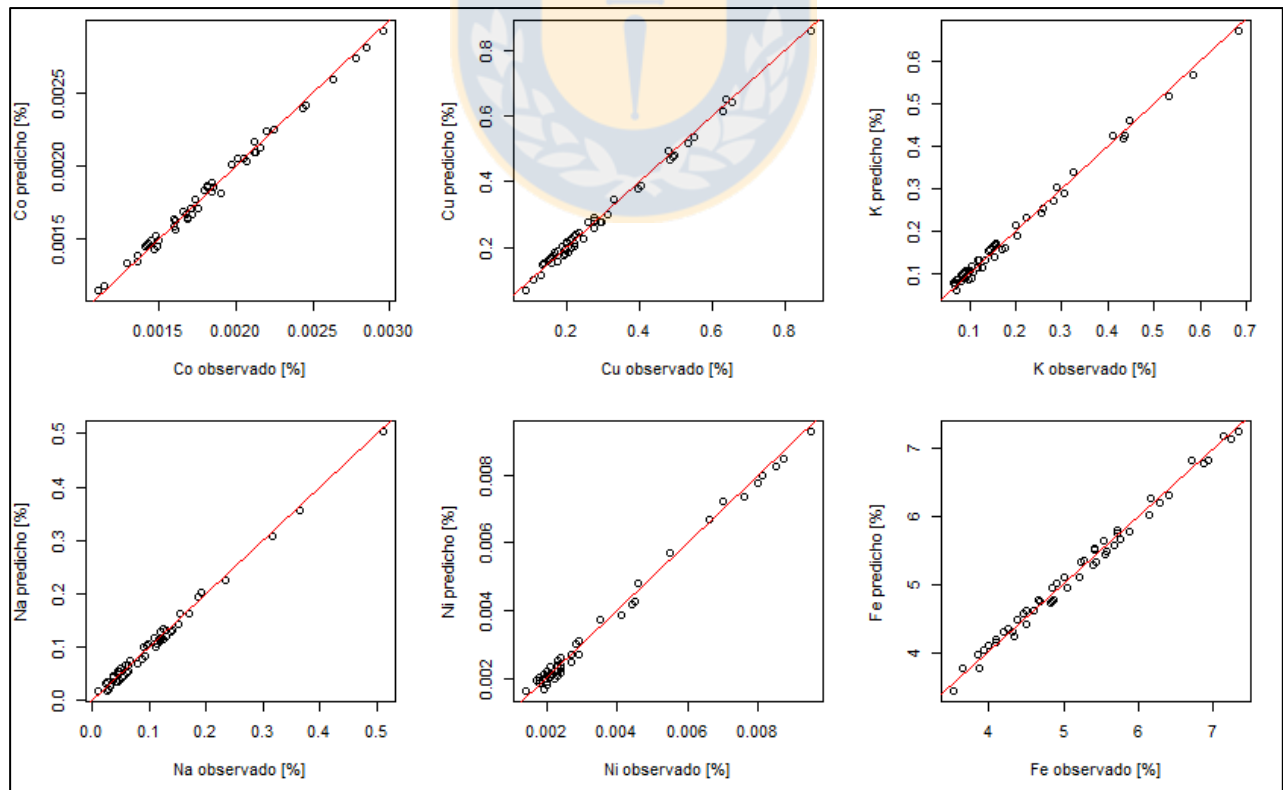


Ilustración 9.12: Gráficos de dispersión para Support Vector Regression, grupo 4.

## Anexo sección C (clustering)

En la presente, se entregan los resultados gráficos del agrupamiento, tanto la división del modelo de bloques como la información contenida en estos a través de gráficos de caja.

Tabla 20: Visualización de los mapas productos del clustering

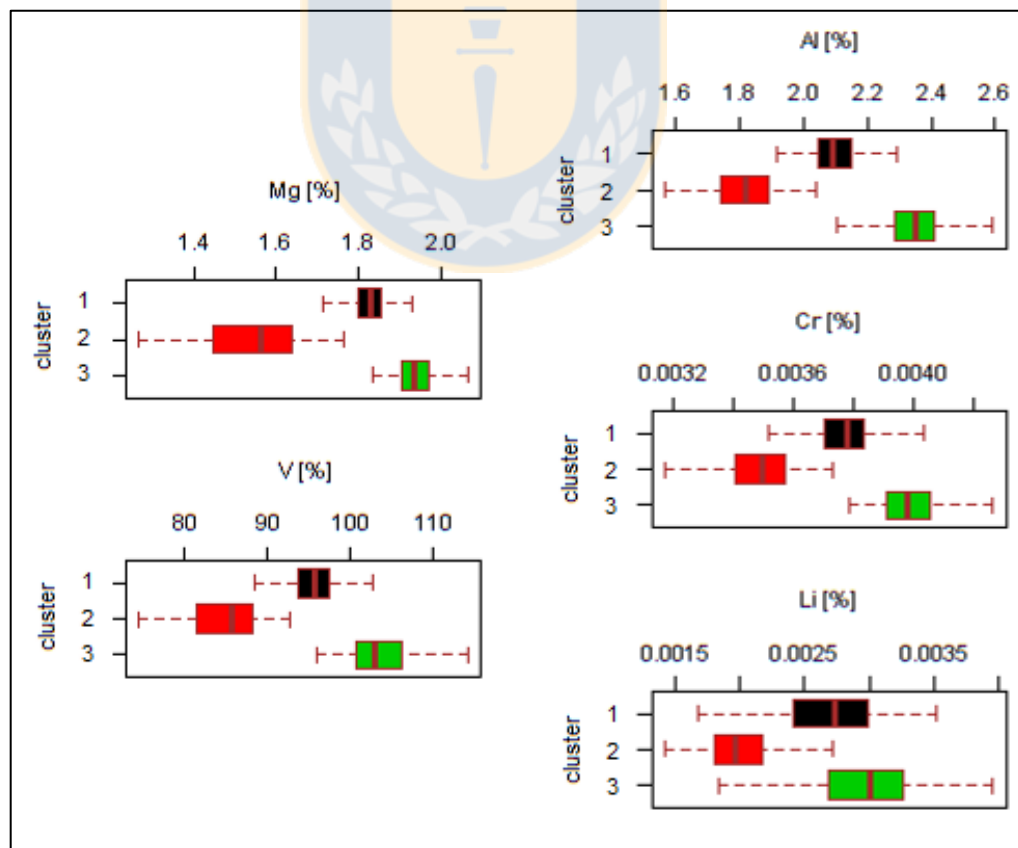
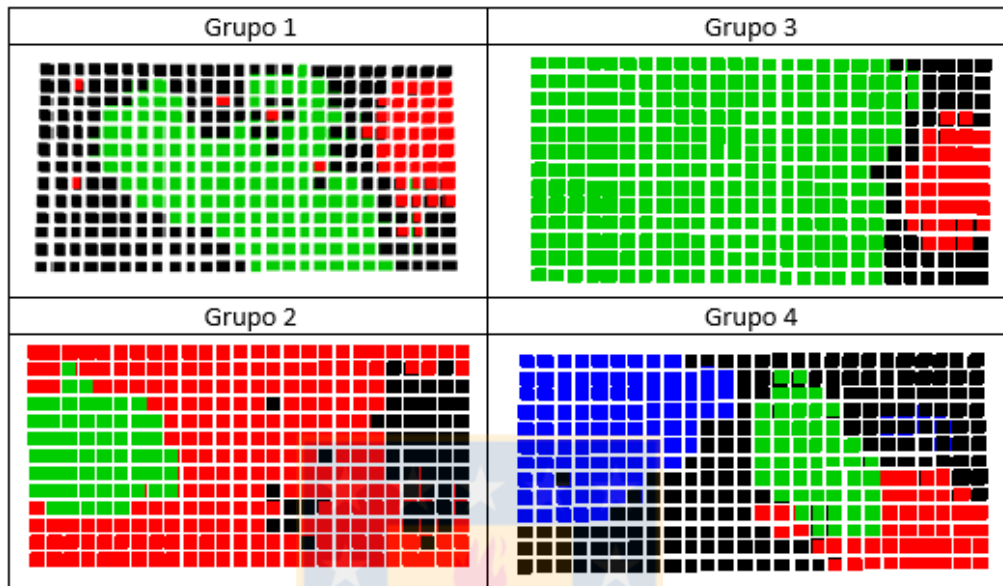


Ilustración 9.13: Diagrama de cajas para clustering de grupo 1.

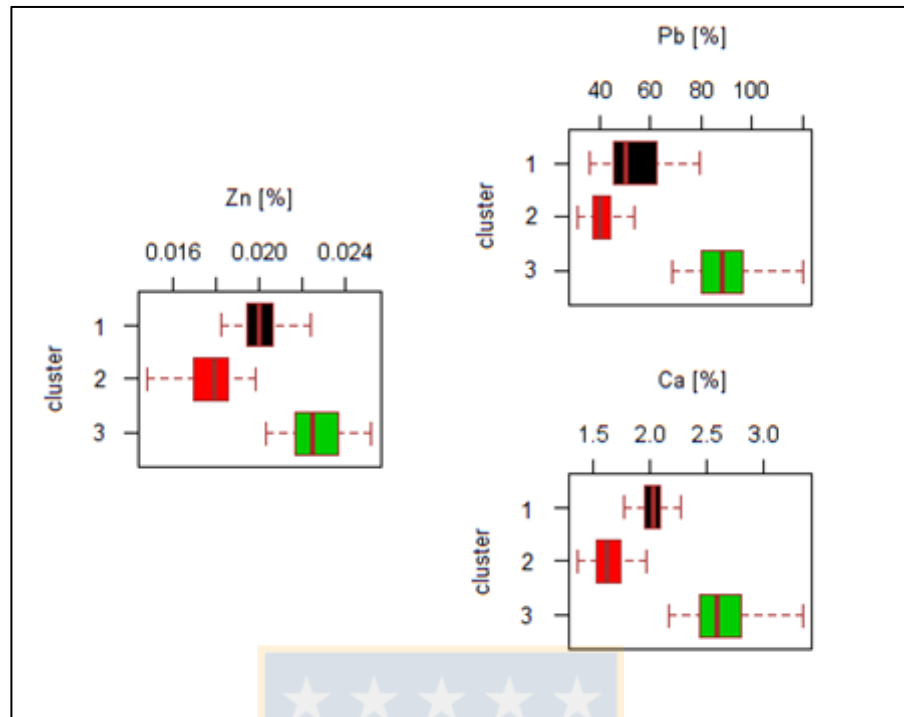


Ilustración 9.15: Diagrama de cajas para clustering de grupo 2.

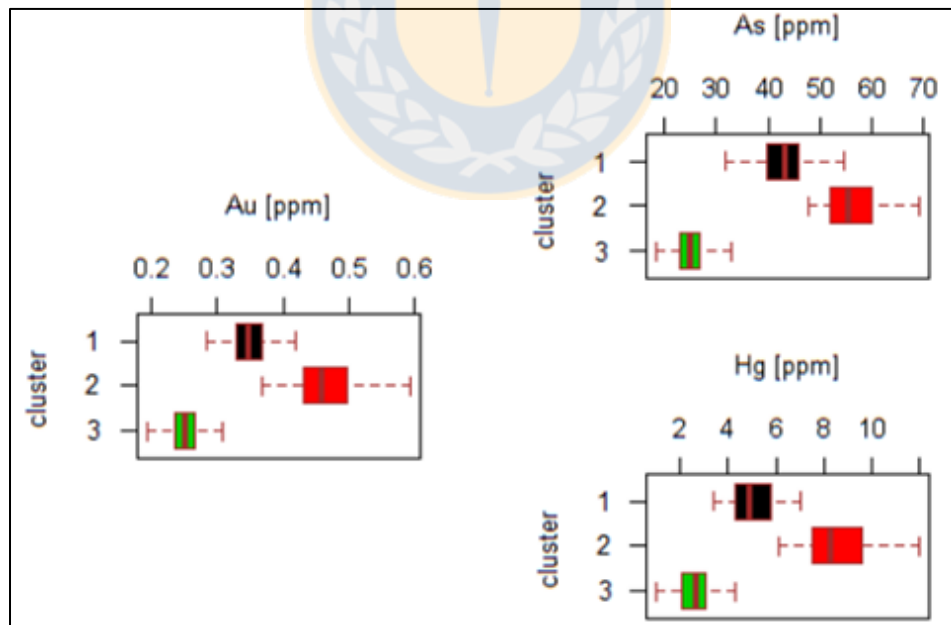


Ilustración 9.14: Diagrama de cajas para clustering de grupo 3.

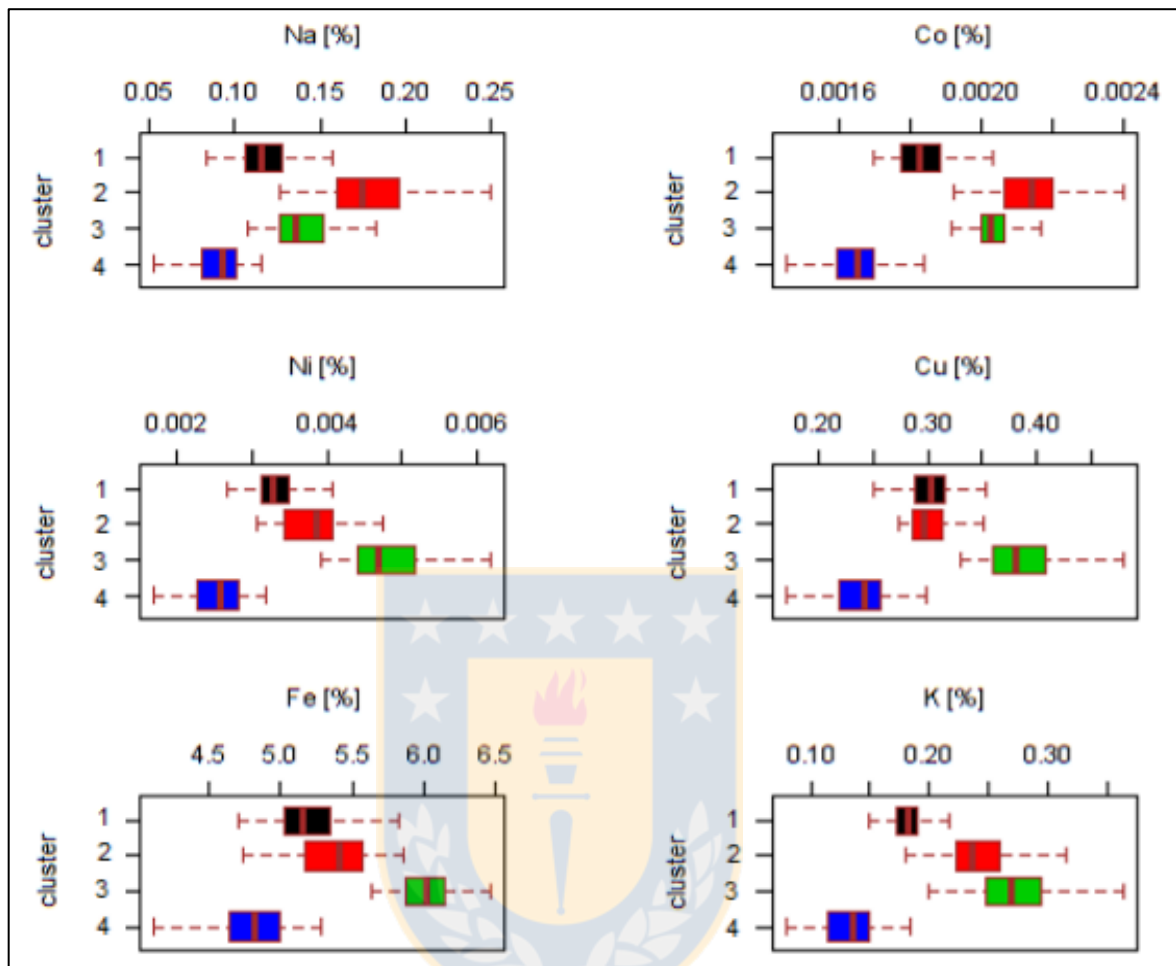


Ilustración 9.16: Diagrama de cajas para clustering de grupo 4.

## Anexo sección D: Otros.

Metales	Últimos 30 Días	Precios			
		Diarios		Mensuales	
		10-12-2019	11-12-2019	Octubre	Noviembre
Cobre (¢US\$/lb)	<a href="#">Ver</a>	274.650	275.920	260.493	265.791
Niquel (US\$/TM)	<a href="#">Ver</a>	5.928	6.139	7.763	6.894
Estaño (US\$/TM)	<a href="#">Ver</a>	7.856	7.904	7.531	7.425
Zinc (¢US\$/lb)	<a href="#">Ver</a>	100.788	100.788	110.930	110.357
Plomo (¢US\$/lb)	<a href="#">Ver</a>	85.865	86.319	99.078	92.166
Aluminio (¢US\$/lb)	<a href="#">Ver</a>	79.379	79.447	77.965	92.166
Oro (US\$/Oz)	<a href="#">Ver</a>	1464.450	1468.050	1494.765	1471.921
Plata (US\$/Oz)	<a href="#">Ver</a>	16.675	16.660	17.625	17.180

Ilustración 9.18: Precios corrientes (BML) para distintos metales.

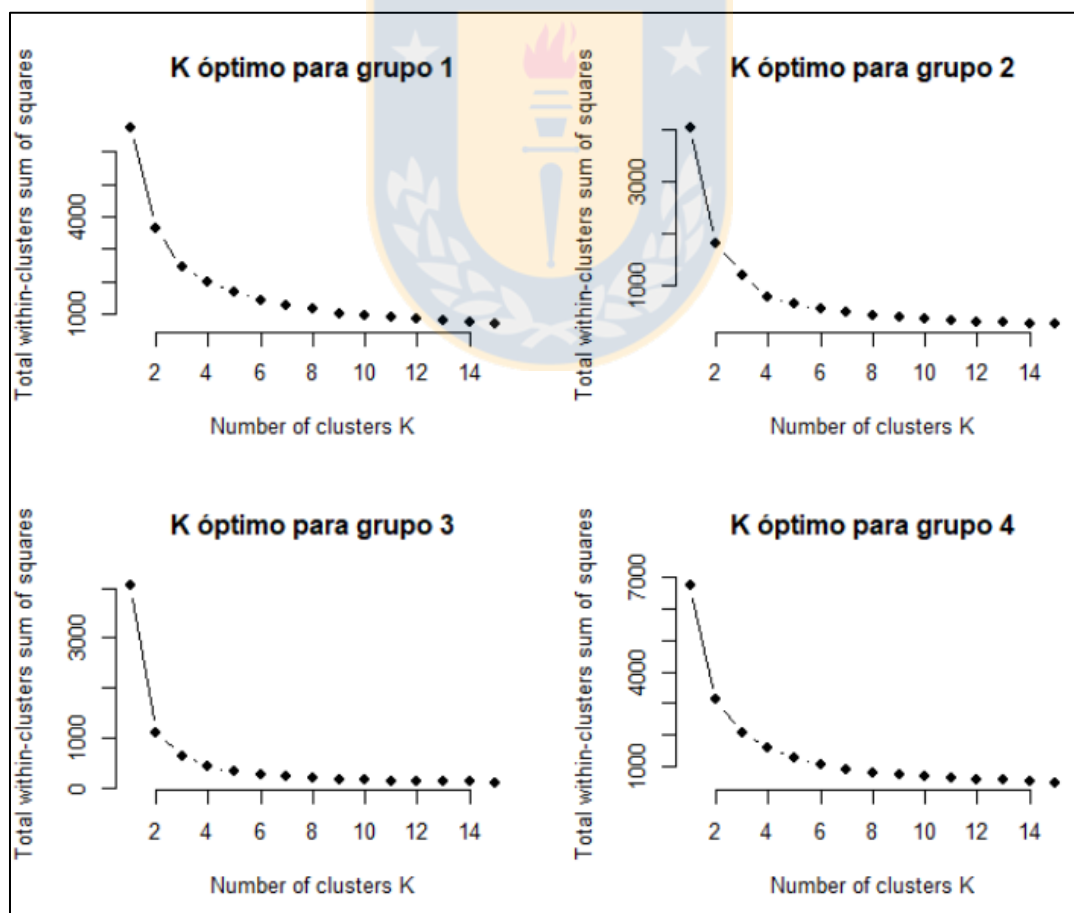


Ilustración 9.17: Método de codo para K óptimo en proceso de clustering por k-medias.

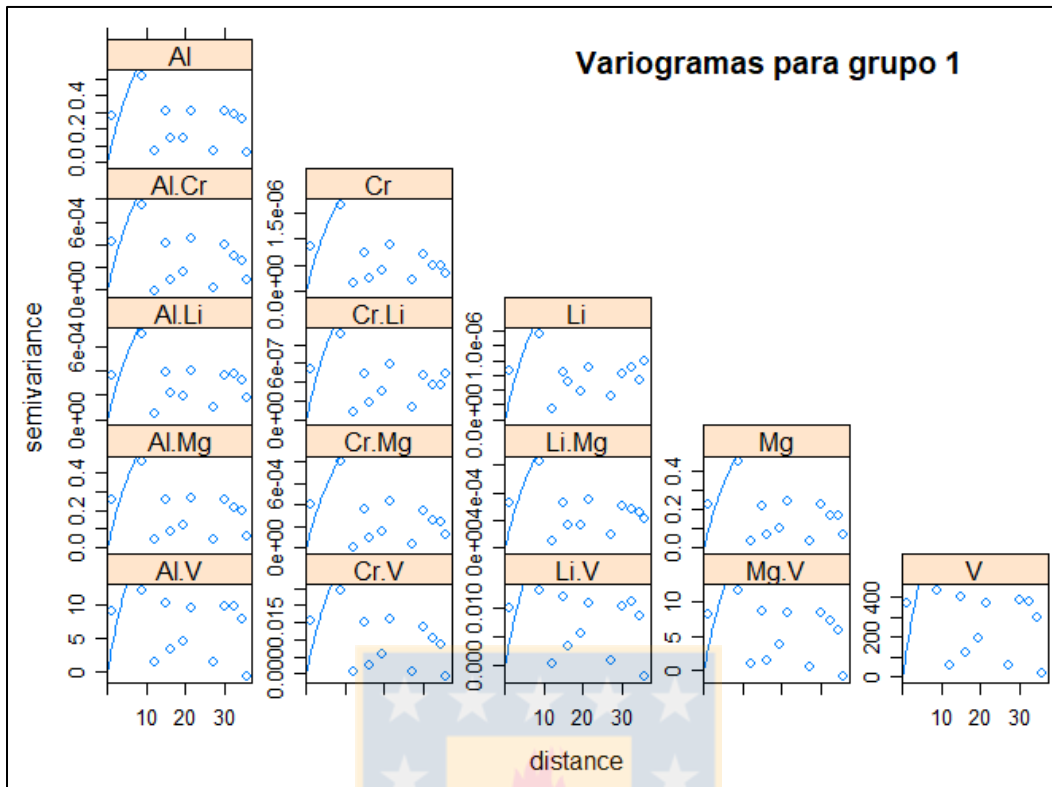


Ilustración 9.20: Variogramas cruzados para grupo 1

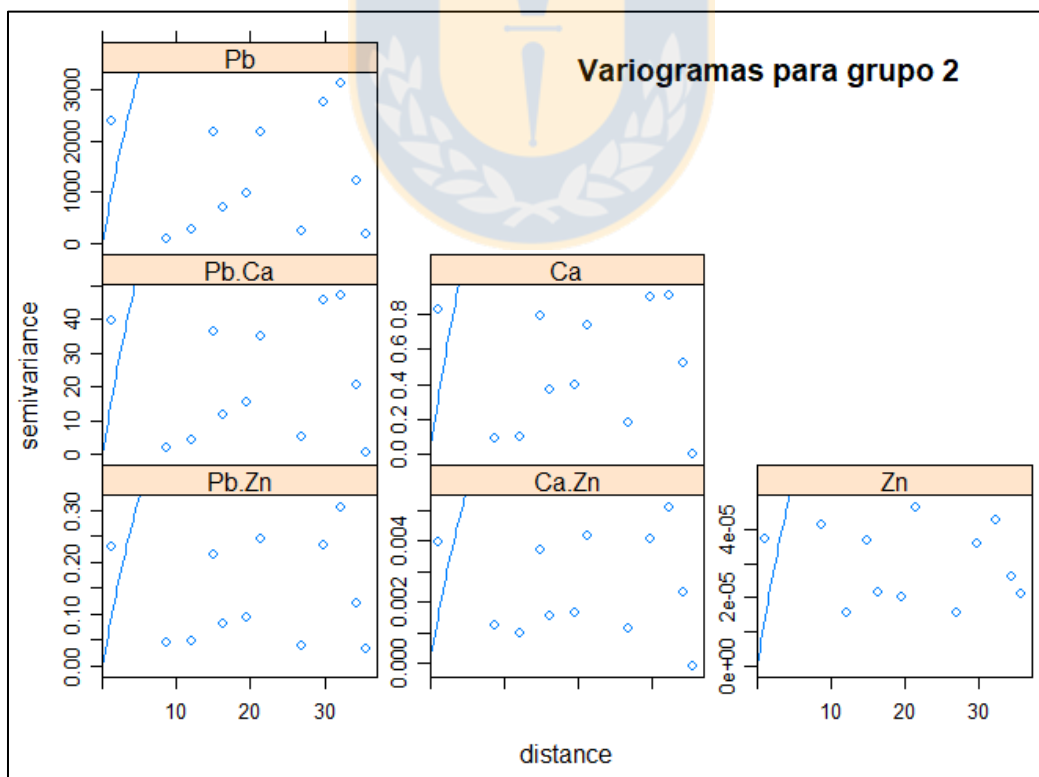


Ilustración 9.19: Variogramas cruzados para grupo 2

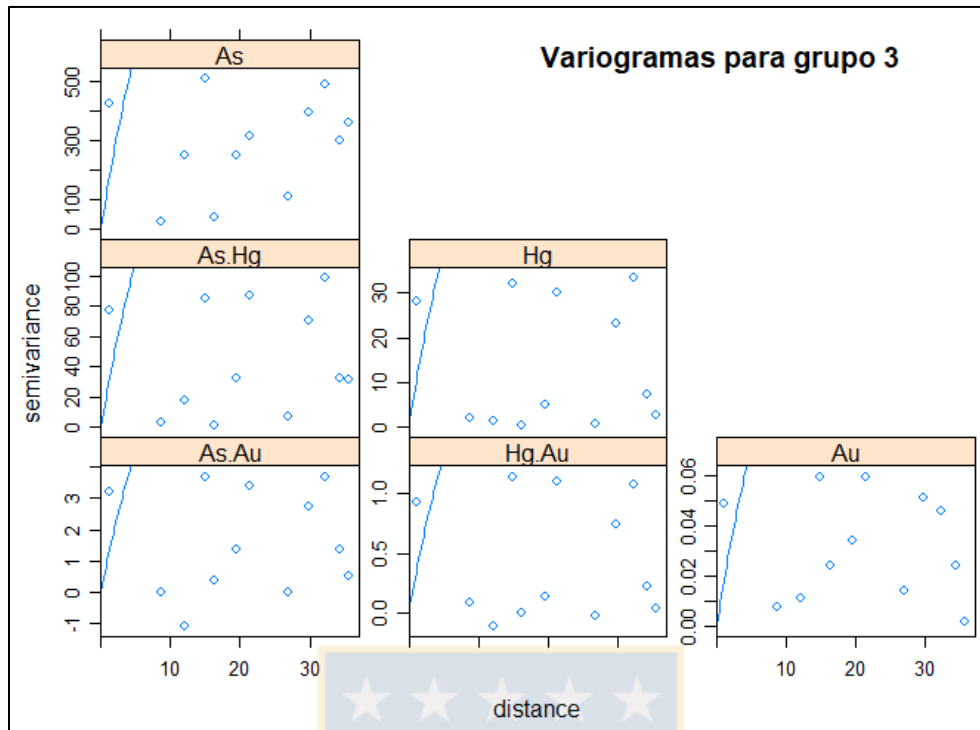


Ilustración 9.22: Variogramas cruzados para grupo 3

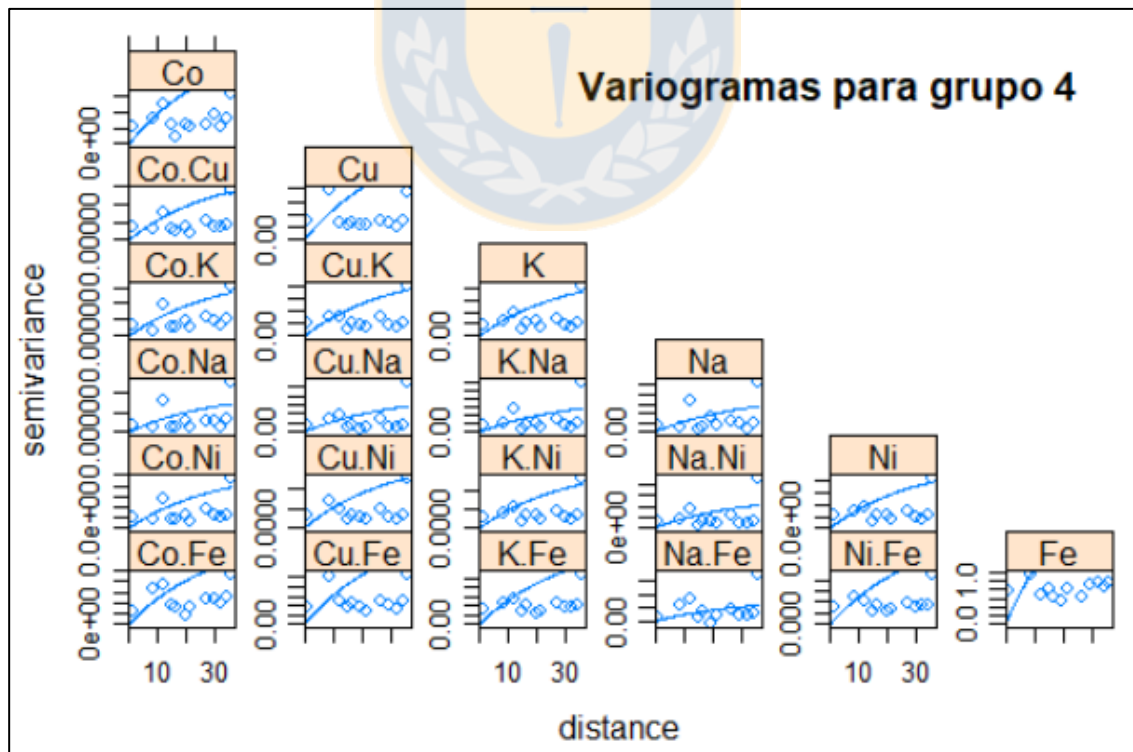


Ilustración 9.21: Variograma cruzado para grupo 4

**UNIVERSIDAD DE CONCEPCIÓN - FACULTAD DE INGENIERÍA**

**Departamento de Ingeniería Metalúrgica**

Hoja Resumen Memoria de Título

<b>Título: Metodología para el modelamiento y caracterización geoestadística de relaves como método de revalorización de pasivos ambientales</b>	
<b>Nombre Memorista: Ricardo Ignacio García Vega</b>	
<b>Modalidad</b>	<b>Profesor (es) Patrocinante (s)</b>    <b>Prof. Roberto Fustos</b>
<b>Concepto</b>	
<b>Calificación</b>	
<b>Fecha</b>	
<b>Prof. Roberto Fustos T.</b>	
<b>Ingeniero Supervisor:</b>	
<b>Institución: Universidad de Concepción</b>	
<b>Comisión (Nombre y Firma)</b>	
<b>Prof. Ramón Díaz N.</b>	<b>Prof. Bernardo Lagos A.</b>
<b>Resumen</b>	
<p>La gran cantidad de relaves mineros abandonados en el país se presentan como un peligro para sus alrededores por su alto contenido de elementos químicos peligrosos, pero también presentan una oportunidad para la recuperación de otras especies valiosas.</p> <p>El presente trabajo propone una metodología para modelar la distribución y concentración de determinados elementos químicos del relave "SOTRAMIN Y PLAYA ATACAMA", ubicado en la localidad de Tal-tal, región de Antofagasta.</p> <p>Tres métodos de estimación de recursos (Inverso de la distancia, Kriging ordinario y SVR) fueron utilizados y comparados para determinar cuál entrega mejores resultados. Se concluyó que la estimación mediante Support Vector Regression (SVR) es la más óptima puesto que presenta un error medio de estimación 10 veces menor que los otros métodos.</p> <p>El modelo óptimo se agrupó mediante el método "k-medias", el cual estableció zonas con tonelajes y leyes medias definidas para los elementos objetivo. Los resultados de este, permitieron establecer un valor económico a priori en base a la cantidad de elementos valiosos como el oro o el cobre contenido. Además, visibilizó la necesidad establecer nuevas directrices para el desarrollo de la actividad minera debido al impacto que genera la disposición de elementos nocivos.</p>	