UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA DE MINAS

PREDICCIÓN GEOESTADÍSTICA DE UNIDADES GEOLÓGICAS O GEOMETALÚRGICAS UTILIZANDO INFORMACIÓN DE VARIABLES CUANTITATIVAS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN MINERÍA
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL DE MINAS

MATÍAS IGNACIO SALAS GUERRA

PROFESOR GUÍA:
XAVIER EMERY

MIEMBROS DE LA COMISIÓN:
NADIA MERY GUERRERO
ALEJANDRO CÁCERES SAAVEDRA
AMIR ADELI SARCHESHMEH

SANTIAGO DE CHILE
2019

RESUMEN DE LA TESIS PARA OPTAR AL
TITULO DE INGENIERO CIVIL DE MINAS
Y GRADO DE MAGÍSTER EN MINERÍA
POR: MATÍAS IGNACIO SALAS GUERRA
FECHA: ENERO 2019
PROFESOR GUÍA: XAVIER EMERY

PREDICCIÓN GEOESTADÍSTICA DE UNIDADES GEOLÓGICAS O GEOMETALÚRGICAS UTILIZANDO INFORMACIÓN DE VARIABLES CUANTITATIVAS

En el comienzo de un proyecto minero, uno de los pasos más importantes es la evaluación del yacimiento y el proceso de identificación y delimitación espacial de unidades geológicas o geometalúrgicas, las cuales tienen influencia en la distribución de las propiedades del subsuelo y ayudan con la planificación de los procesos metalúrgicos a utilizar en el material a ser extraído.

Una de las fuentes de información proviene de las distintas perforaciones realizadas sobre el terreno en estudio y el posterior *logueo* geológico realizado sobre éstas. Luego, se utiliza la información del *logueo* junto con el conocimiento geológico de la zona para crear una interpretación geológica de la totalidad del yacimiento. Debido a que tanto el *logueo* como la interpretación geológica están sujetos a decisiones subjetivas y a posibles errores, se presenta la posibilidad de plantear mejoras en esta etapa de la evaluación del yacimiento.

Debido a lo anterior, se presenta una metodología que busca construir y validar variables categóricas, en base a la utilización de clasificadores estadísticos capaces de asignar categorías a nuevos sets de datos utilizando la información de los datos ya clasificados. La hipótesis en la que se basa la propuesta del presente trabajo es que la interpretación geológica es modelada de forma suave, es decir, que las categorías no están muy fragmentadas, ni presentan cambios abruptos; por lo tanto, si se logra obtener simulaciones que de la misma forma sean suaves, los clasificadores podrán obtener mayor acierto en su clasificación.

Para lograr las simulaciones más suaves se propone implementar el filtraje de algunas componentes del análisis de corregionalización, en especial aquellas que representen variabilidad en el corto alcance para evitar que tengan su influencia en el modelo, y de este modo obtener simulaciones más suaves. Con el fin de probar la propuesta de filtraje, se aplica a datos de la mina Spence y se diseñan tres experimentos que permiten cuantificar el acierto al aplicar los clasificadores a las simulaciones tradicionales sin la aplicación de filtraje y a las simulaciones con filtraje.

Los resultados obtenidos en el experimento 1 indican una mejora de 3.50% y 4.78% en el acierto de *logueo* y *flagueo* respectivamente al aplicar filtraje; en el experimento 2 los resultados indican una mejora de 4.00% para *logueo* y 3.71% para *flagueo*; y finalmente, el experimento 3 indica una mejora de 3.10% para la interpretación gracias a la inclusión de la metodología propuesta.

Considerando los resultados de los tres experimentos, se concluye que la implementación y aplicación de simulaciones con filtraje mejora el porcentaje de acierto aplicado a la interpretación geológica, pero también se observa en los resultados que ayuda de forma similar a mejorar el porcentaje de acierto aplicado a los *logueos*; lo que permite concluir que la metodología propuesta no solo sirve para las clasificaciones suavizadas como se propone inicialmente, como el caso de la interpretación; sino que también puede servir para las clasificaciones no tan suavizadas, como el caso del *logueo.*

THESIS ABSTRACT
AUTHOR: MATÍAS IGNACIO SALAS GUERRA
DATE: JANUARY 2019
ACADEMIC ADVISOR: XAVIER EMERY

GEOSTATISTICAL PREDICTION OF GEOLOGICAL OR GEOMETALLURGICAL UNITS USING QUANTITATIVE VARIABLES' DATA

At the beginning of a mining project, one of the most important steps is the ore body evaluation and the process of identification and spatial delimitation of geological or geometallurgical units, which have an influence on the distribution of subsurface properties and provide information to help with the planning of the metallurgical processes to be used for the processing of the material to be extracted.

One of the sources of information comes from the drilling process carried out on the region under study and their subsequent geological logging. Later, the information of the loggings is used along with the geological knowledge to create a geological interpretation of the entire ore body. Because both logging information and interpretations are susceptible to subjective decisions and possible errors, there is a possibility to propose improvements in this stage of ore body evaluation.

Because of that, a methodology is presented aiming to predict and validate categorical variables, based on the usage of statistical classifiers capable of assigning categories to new sets of data using information of already classified data. The hypothesis on which the proposal of this thesis is based is that the geological interpretation is modelled in a smooth way, that is to say, that the categories are not very scattered, nor present abrupt changes; therefore, if the simulations could be modified to be smooth as well, the classifier would be able to obtain better accuracy in its classification.

To get smoother simulations, it is proposed to implement the filtering of some components in the coregionalization analysis, focused in those that represent variability in short ranges to avoid having their influence in the model, thus, obtaining smoother simulations. In order to test the proposal, the proposal is applied to data of the Spence mine and three experiments are designed to quantify the percentage of correct classifications in the applications of the classifiers to traditional simulations without filtering and to simulations using filters.

The results obtained in the first experiment shows an improvement of 3.50% and 4.78% in the correct classification for logging and interpretation, respectively, with the usage of filters; in the second experiment the results show an improvement of 4.00% for logging and 3.71% for interpretation; finally, for the third experiment the results show an improvement of 3.10% for the geological interpretation when the proposed methodology is implemented.

Considering the results obtained in the experiments, it is concluded that the implementation and application of the simulations with filtering improve the percentage of correct classifications for the geological interpretation, but it is also observed that the proposal also improves the percentage for geological loggings; this allows to conclude that the proposed methodology not only works with smoothed classifications as initially proposed, such as in the case of geological interpretation; but also works to classify less smoothed variables, such as in the case of loggings.

Agradecimientos

En primer lugar, agradezco a mi familia: a mi madre Claudia y a mi padre Cristian por siempre estar ahí cuando los necesité, por brindarme todas las oportunidades posibles, por darme las mejores condiciones para desarrollarme como persona, por mostrarme el camino y ser ejemplos a seguir; sin ustedes esto no sería posible. Y a mi hermano Tomas y mi hermana Magdalena por la compañía y las alegrías, espero que mi camino sea guía para el de ustedes.

Agradezco a mis amigos que han marcado de forma importante mi camino: a Gonzalo y Paulo que han estado desde el colegio de forma permanente y seguirán estando; al Waina y al Nacho por las incontables horas perdidas en llamada conversando y jugando; a Loza, Pablito y Xami por los mejores y más chistosos semestres pasados en la U; a Agustín por ser el mejor partner y por todas las veces que sacamos adelante los ramos juntos; y a mi nueva familia de Canadá: Camille, Amélie y todos los que conformaban el departamento 905 por transformar ese año en un recuerdo imborrable.

En especial agradecer a mi pareja Jessica, por nunca rendirse, por hacerme mejor persona, por creer en nuestra relación y ser una de las principales motivaciones para terminar este trabajo; por todas las aventuras vividas, por mostrarme el mundo, por esperar con paciencia el cierre de esta etapa, y por sobre todo por lo que está por venir.

En el ámbito académico, agradezco al profesor Alejandro Cáceres por sus clases, por hacer nacer en mí el interés por la geoestadística y por permitirme realizar mis primeros cursos como docente; a Pablo Vega que fue mi auxiliar, compañero de cuerpo docente, compañero de tesis y ahora un muy buen amigo. Y al mejor profesor que tuve, Xavier Emery, por todo el conocimiento entregado, por la motivación, por permitirme trabajar con él, por la oportunidad del magister y por toda la guía en estos últimos años.

Finalmente, agradezco el apoyo y financiamiento recibido por parte de la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) para la realización de la presente tesis, a través del proyecto N°1170101 del programa FONDECYT REGULAR; al Departamento de Ingeniería de Minas y al Advanced Mining Technology Center (AMTC) de la Universidad de Chile y sus funcionarios; y al área de Recursos y Modelamiento de BHP por la ayuda prestada para la realización de este trabajo.

Tabla de contenido

Capítulo 1: Introducción

1.1 Planteamiento del problema

Al comenzar un proyecto minero uno de los pasos de mayor relevancia es la evaluación del yacimiento a tratar. En este tipo de estudios no solo basta con detallar los valores de las variables cuantitativas como son las leyes de los distintos minerales, sino que también se deben estudiar variables categóricas como lo son la alteración mineral, la litología o el tipo de mineralización, en términos de identificación y delimitación espacial, lo que define las distintas unidades geológicas o geometalúrgicas presentes en el yacimiento.

Las unidades mencionadas anteriormente, así como también la forma en que se relacionan sus bordes, tienen directa influencia sobre los valores y distribuciones de las otras variables. Además, es importante tener información de éstas para la posterior metalurgia en el procesamiento del material extraído, ya que, cada unidad geometalúrgica puede requerir diferentes procesos, tener diferentes recuperaciones, diferente consumo de algún insumo determinado, o requerir de algún tratamiento especial, lo cual debe ser planificado con anticipación.

El principal insumo de información para la evaluación de yacimientos proviene de las perforaciones realizadas sobre el terreno en estudio, principalmente mediante la recuperación de testigos, los cuales pueden ser analizados para poder caracterizar las zonas del yacimiento. Un problema presente en esta etapa es que las perforaciones representan un desafío económico y logístico, por lo que la cantidad de perforaciones realizada es limitada y, por lo tanto, la cantidad de información disponible también lo es; una de las herramientas más utilizadas para obtener esta información es lo que se conoce como *logueo* geológico.

El *logueo* geológico consiste en una descripción principalmente visual de los atributos petrofísicos de los testigos de roca, en donde, dependiendo del objetivo, se busca clasificar en algún tipo de variable categórica, como las nombradas anteriormente, detallar la cantidad de fracturas presentes o alguna anomalía importante para el estudio. Por su naturaleza de depender de criterios visuales, es una clasificación intrínsecamente subjetiva y que puede verse afectada por diferentes razones, como lo pueden ser: geología demasiado compleja, falta de experiencia en los equipos de trabajo, diferentes criterios entre las personas a cargo, entre otros.

La información obtenida a partir del *logueo* geológico de los testigos es utilizada por geólogos e ingenieros de recursos junto a otros conocimientos que ellos posean del yacimientos para construir una interpretación geológica o geometalúrgica, la cual como su nombre indica es la interpretación que el equipo de trabajo da en las tres dimensiones al yacimiento en estudio, conteniendo información referente a la ubicación y extensión que pueden tener variables categóricas dentro del macizo rocoso como pueden ser alteración, litología o mineralización.

Debido a que el *logueo* geológico posee una cantidad de información limitada y se basa en criterios subjetivos, lo que puede introducir errores o imprecisiones a la clasificación; y por otro lado la interpretación además de quedar a criterio del equipo de trabajo, en parte se basa en los *logueos* los que pueden proveer información errónea o imprecisa; configura un marco de trabajo, en una etapa importante de un proyecto minero como lo es la evaluación del yacimiento, que tiene potencial para ser mejorado.

Además, tanto el *logueo* como la interpretación proveen información determinística, por lo que otro punto de mejora es proponer y probar una metodología que pueda entregar una categorización probabilística, que no solo indique la categoría más probable en una determinada posición, sino que también entregue el porcentaje de probabilidad asociada a la existencia de dicha categoría, lo que puede generar una nueva fuente de información para realizar los estudios.

1.2 Idea clave / Hipótesis

Se espera observar como base una alta correlación entre la información de las leyes y otras variables cuantitativas que se tiene de los sondajes, y la información de las variables categóricas del tipo alteración, litología o mineralización que ha sido obtenida mediante el *logueo* de dichos sondajes o que se posee por medio de identificar para cada posición de los compósitos de los sondajes la categoría de la interpretación geológica presente en dicha posición.

Utilizando la correlación entre variables cuantitativas y categorías se espera construir clasificadores mediante algoritmos de aprendizaje supervisado en base a los pares de datos $(variables\ cuantitativas, clasificación)$ disponibles, los cuales pueden ser utilizados para validar clasificaciones realizadas a los sondajes o ser aplicado a los resultados de simulación de las variables cuantitativas del yacimiento y, de esta forma, validar una interpretación ya realizada o generar una nueva, la cual no estará afectada directamente por las decisiones subjetivas del equipo a cargo al realizar la interpretación, y además, en el caso de trabajar con simulaciones permite tener modelos probabilísticos de la interpretación geológica.

La innovación del presente trabajo se basa en la hipótesis de que la interpretación geológica tiende a ser "suave", es decir, que los cambios entre una categoría y otra no suelen ser tan abruptos ni diseminados por el espacio; y que, al simular un yacimiento las variables cuantitativas como las leyes no siguen necesariamente este patrón "suave"; por lo que, al aplicar un clasificador el patrón que poseen los datos a categorizar no tienen el mismo comportamiento que los resultados esperados, en este caso, que la interpretación geológica.

Para comprobar la hipótesis anterior, se propone realizar simulaciones de las leyes aplicándole un filtraje a los componentes de menor alcance del modelo lineal de corregionalización, de modo de obtener un modelo de bloques con pseudo-leyes que sigan la tendencia suave como la de la interpretación geológica, sin tener la variabilidad a corta escala de los componentes filtrados. De esta forma, al aplicar algún clasificador se espera que sus resultados presenten mayor acierto al ser comparados con una interpretación geológica ya realizada por el equipo a cargo, en comparación a realizar una clasificación sobre simulaciones no filtradas.

1.3 Objetivos

1.3.1 General

- Aplicar clasificador a variables cuantitativas para identificar variables categóricas en yacimientos mineros.

1.3.2 Específicos

- Mostrar relación entre las variables cuantitativas del yacimiento y las variables categóricas.
- Implementar metodología de simulación con filtraje de componentes del análisis de corregionalización.
- Aplicar a un caso de estudio real cuya base de datos posea sondajes *logueados* e interpretación geológica del yacimiento.
- Simular con y sin filtraje variables cuantitativas en el yacimiento.
- Calcular acierto en los sondajes de la clasificación con relación a los *logueos* y a la interpretación geológica correspondiente a cada sondaje.
- Calcular acierto en el yacimiento de la clasificación con relación a la interpretación geológica.
- Comparar resultados de clasificación al utilizar simulación con y sin filtraje.

1.4 Alcances

- El estudio del caso real se realiza sobre un depósito mineral de cobre, cuya base de datos posee las siguientes variables cuantitativas: posición del dato, ley de Cu, ley de Ag, ley de Au, ley de As, ley de Mo; y las siguientes variables categóricas: alteración, litología y mineralización.
- Se aplica la metodología a aquellos bloques del modelo que tengan al menos un sondaje a 250 metros de distancia.
- Se aplica la metodología a la variable categórica que presente el mayor acierto luego de realizar validación cruzada de clasificadores.
- Los clasificadores utilizados en la metodología son: árbol de decisión, vecino más cercano, support vector machine y naive Bayes.
- Con relación al trabajo realizado para el yacimiento simulado, por limitaciones computacionales se estudian sección transversales horizontales del yacimiento a distintas cotas y no el modelo de bloques completo.
- El análisis geoestadístico, el modelamiento y la aplicación de clasificadores se realizan a través del software Matlab.

Capítulo 2: Antecedentes

2.1 Mapeo geológico

En la presente tesis se prueban metodologías sobre dos formas de representar o mapear la información geológica disponible: los *logueos* geológicos que presentan información en aquellos lugares en donde los sondajes fueron realizados y la interpretación geológica que presenta información sobre todas las posibles posiciones del yacimiento a evaluar. Dicha información es utilizada para definir y delimitar unidades geológicas o geometalúrgicas, en el caso de que la información que contengan las unidades sea utilizada en la planificación de los procesos metalúrgicos posteriores.

2.1.1 Perforaciones

Para realizar cualquier tipo de estudio, uno de los primeros pasos es la obtención de información directa de la geología presente bajo tierra; para esto, la principal ruta para contar con información confiable es a través de las perforaciones y los correspondientes testigos que se obtienen por medio de éstas. Dichas perforaciones cumplen diferentes funciones incluyendo: delimitar y caracterizar el cuerpo mineralizado, y de este modo entregar información para definir si el proyecto sigue su curso o no es económicamente rentable (Marjoribanks, 2010).

Mientras más avanza el proyecto, más perforaciones se van realizando y, por lo tanto, se dispone de mayor cantidad de información, aunque esto a su vez implique que los costos cada vez sean más elevados, por lo tanto, si bien es necesaria gran cantidad de información, no es ilimitada debido a factores económicos (Marjoribanks, 2010).

2.1.2 *Logueo* geológico

El procesamiento y la obtención de información de los testigos se realiza por medio de un proceso llamado *logueo* geológico, mediante el cual un equipo de geólogos a cargo registra y categoriza distintas variables de interés para el proyecto presentes en los testigos de roca basándose principalmente en una inspección visual, generalmente solo apoyados por lupas de mano y su propio criterio (Moon et al., 2006).

En este proceso se involucran distintos profesionales, los cuales poseen distinto grado de conocimiento de la geología local del yacimiento, tienen distinta formación académica y distinta experiencia en campo, por lo tanto, aunque se trate de unificar criterios en la metodología, al ser utilizado un criterio visual, siempre estará involucrado un grado de subjetividad en la clasificación (Rossi y Deutsch, 2014; Adeli et al., 2018).

Este procedimiento en algunos casos involucra un *logueo* rápido en el sitio en donde la perforación está siendo realizada, lo que permite tomar una decisión rápida sobre si es conveniente o no continuar con la perforación; posteriormente en un sitio destinado para dicha tarea se realiza un *logueo* más exhaustivo y con mayor detalle, incluyendo variables tales como: espaciamiento de fracturas, orientación de estas, colores, texturas, alteración, litología, mineralización, porcentaje de recuperación del testigo, entre otras (Moon et al., 2006).

Debido a que la clasificación de los *logueos* es cualitativa y a su naturaleza visual, puede estar sujeta a errores debido a diferentes posibles factores que se ejemplifican a continuación (Manchuk y Deutsch, 2012; Cáceres y Emery, 2013):

- Dificultad para estimar porcentajes de mineral.
- Dificultad para estimar limites relacionados a criterios de clasificación.
- Dificultad para distinguir litologías alteradas.
- Baja recuperación de testigos.
- Clasificar en base a patinas y no estudiar el mineral cubierto.
- Diferencia de criterios entre geólogos involucrados.
- Alta rotación de geólogos involucrados.
- Falta de análisis químicos de apoyo.

2.1.3 Interpretación geológica

Una limitación que posee la información obtenida en el *logueo* geológico es que solo se conoce en aquellas posiciones en donde fue extraído un testigo y tampoco existe alguna herramienta que permita saber con certeza la composición de la roca bajo tierra, por lo que es necesario crear una interpolación de la información, para de esta forma poder tener una idea aproximada de cómo se comportan variables como el tipo de roca u otras como: alteración, litología o tipo de mineralización en todas las posiciones del espacio en estudio.

Esta interpretación consiste en una representación geolocalizada de variables categóricas en donde se aprecia de forma clara la extensión de cada categoría basado en la interpretación que el equipo a cargo del estudio realiza del yacimiento a partir de la información obtenida del *logueo* geológico, observaciones realizadas directamente en campo, ensayos geofísicos, estudio de leyes, consistencia con teoría de formaciones geológicas y cualquier otro conocimiento que pueda ayudar a reducir el error de la interpretación (Guillen et al., 2008; Adeli et al., 2018).

Una interpretación geológica es interpolada a partir de información limitada, se basa en conocimientos no probados o aproximados de la realidad que existe en el yacimiento y es realizada bajo criterio subjetivos. Es por esto que el presente trabajo presenta la hipótesis de que la interpretación geológica representa el comportamiento de las variables de forma menos abrupta que en la realidad, agrupando conjuntos más grandes bajo la misma categoría, utilizando contactos entre diferentes categorías más suavizados y algunas veces agrupando categorías similares en una sola, es decir, en líneas generales representa una realidad más suave debido principalmente al desconocimiento de la realidad bajo tierra.

2.1.4 Unidades geometalúrgicas

En el proceso de evaluación de yacimientos, para estudiar la factibilidad del proyecto, no solo basta con las leyes, sino que también pueden existir otros factores que influyan en la factibilidad técnica o económica de un proyecto, los cuales están relacionados con la geología y la respuesta metalúrgica esperada del yacimiento en estudio; la combinación de estos factores se puede denominar geometalurgia (Millet et al., 2010; Fustos, 2017).

Para definir las unidades geometalúrgicas presentes en el yacimiento se busca agrupar zonas que tengan comportamiento geológico y metalúrgico similar entre los bloques que la componen, y diferente a las demás zonas. Una buena identificación y delimitación de las unidades geometalúrgicas, junto con una validación de éstas, permiten determinar factibilidades técnicas y

económicas de extracción y procesamiento metalúrgico en cada zona, así como también, planificar en una etapa temprana el tipo de procesamiento requerido y los insumos necesarios para obtener el producto final deseado (Rosales, 2014), por ejemplo, podría permitir determinar en qué etapas de un proyecto se requiere tener operativa una planta de procesamiento de óxidos y en qué etapa se requiere tener operativa una planta de procesamiento de sulfuros.

2.2 Geoestadística

Las herramientas geoestadísticas descritas en la presente sección son la base matemática para el estudio de la información disponible en la base de datos y la forma de generar nueva información que es importante input para probar la nueva metodología propuesta en esta tesis. Las referencias claves de la formulación matemática y teoría descrita en esta sección son Chilès y Delfiner (2012) y Emery (2013).

2.2.1 Variable regionalizada

Las variables mencionadas en la sección anterior y que serán utilizadas en el presente trabajo, tienen asociado no solo valores o categorías, sino que representan algún punto en el espacio geográfico. Aquellas variables que representen algún espacio, ya sea geográfico, temporal u otro; tienen por nombre variables regionalizadas, algunos ejemplos de estas pueden ser:

- Leyes de mineral.
- Densidad de roca.
- Recuperación metalúrgica.
- Tipos de roca.
- Alteración.
- Litología
- Mineralogía.

Por definición, una variable regionalizada no tiene una extensión infinita, por lo que solo se estudian y analizan dentro de un espacio u dominio limitado, que se puede denominar "campo" de la variable. Dentro de este campo la variable está definida, pero no fuera de este; en el caso del presente trabajo el campo es el espacio geográfico delimitado por el yacimiento en estudio dentro del cual se posee información proveniente de sondajes y fuera de este no se conoce información detallada (Emery, 2013).

2.2.2 Función aleatoria

Sea *z(x)* una variable regionalizada, siendo *x* la representación de una posición de dicha variable, y sea *D* la denominación del campo que la contiene. El valor numérico *z(x)* se puede interpretar como el resultado (realización) de una variable aleatoria pariente, que se denota como *Z(x)* (mayúscula). Cuando *x* recorre *D*, se genera un conjunto de variables aleatorias $\{Z(x): x \in D\}$, en donde cada realización de *Z(x)* es un valor de la variable regionalizada *z(x)* en una posición dada del espacio. Dicho conjunto de variables es lo que se conoce como función aleatoria, campo aleatorio o proceso estocástico.

Es interesante introducir el concepto de función aleatoria debido a que sus valores en diferentes posiciones del espacio no son independientes entre sí y que se pueden reconocer relaciones entre éstas, lo que en geoestadística se denomina correlación o continuidad espacial. Una de las formas

de estudiar una función aleatoria es calculando sus momentos, los cuales se muestran a continuación:

2.2.2.1 Momentos de una funciona aleatoria

- Momento de primer orden: esperanza matemática.

$$\mu(x) = E[Z(x)], \qquad x \in D$$

- Momento de segundo orden:
  - Varianza

$$\sigma^2 = Var[Z(x)] = E[\{Z(x) - \mu(x)\}^2], \qquad x \in D$$

  - Covarianza

$$C(x_i, x_j) = E\left[\{Z(x_i) - \mu(x_i)\}\{Z(x_j) - \mu(x_j)\}\right], \qquad x_i, x_j \in D$$

  - Variograma

$$\gamma(x_i, x_j) = \frac{1}{2} Var\left[Z(x_i) - Z(x_j)\right], \qquad x_i, x_j \in D$$

2.2.2.2 Concepto de estacionaridad

Existen dos problemas para realizar inferencias estadísticas: en primer lugar, una variable regionalizada z solo es una realización de una funciona aleatoria Z y, en segundo lugar, solo se posee información parcial del yacimiento.

Para solucionar dichos problemas, se introduce el concepto de estacionaridad, lo que supone que los valores en distintas posiciones del yacimiento presentan las mismas características, por lo que pueden ser considerados como distintas realizaciones de un mismo proceso aleatorio; lo que, dicho de otra forma, plantea que la distribución y los momentos de una función aleatoria no dependen de su posición absoluta, sino que de su posición relativa.

Las consecuencias que posee tomar la hipótesis de estacionaridad y que facilitan la realización de los estudios y cálculos que se describen en las siguientes secciones, son las siguientes:

- Esperanza constante en el campo en estudio.
- Varianza constante en el campo en estudio.
- Covarianza solo depende de la separación de los datos.
- Variograma solo depende de la separación de los datos.

2.2.3 Análisis variográfico

El momento de primer orden utiliza una posición espacial a la vez, por lo tanto, no entrega información sobre la correlación espacial de estos. Debido a lo anterior los momentos de segundo orden son más utilizados, ya que ayudan a describir la continuidad espacial de una variable regionalizada (Emery, 2013), para el caso de la evaluación de un yacimiento y la predicción o simulación de recursos, se utiliza en gran medida el análisis variográfico, el cual mide las variaciones en relación con la dirección y la distancia de los puntos donde se tiene información de una variable.

El procedimiento general consiste en inicialmente estimar el variograma con relación a los datos a utilizar, lo que se conoce como variograma experimental; que en algunos casos se realiza siguiendo una dirección en donde se haya identificado una anisotropía importante, y posteriormente se modela el variograma experimental utilizando funciones matemáticas conocidas, lo que lleva por nombre variograma modelado.

2.2.3.1 Variograma experimental

Sea $z(x)$ una variable regionalizada, x el vector de posiciones en donde se mide z, y h un vector de separación entre dos observaciones de x. El variograma experimental se puede calcular de la siguiente forma:

$$\hat{\gamma} = \frac{1}{2|N(h)|} \sum_{N(h)} \{[Z(x_i) - Z(x_j)]\}^2$$

En donde:

- $N(h)$ = pares cuya separación sea igual a h
- $|N(h)|$ = número de pares contenidos en $N(h)$.

Si para algún caso, los puntos x para los que se calcula el variograma experimental están irregularmente distribuidos en el espacio, la cantidad de pares de puntos que cumplan con estar distanciados exactamente en h serán pocos en relación a la cantidad de pares posibles. Por lo anterior, al graficar el variograma experimental para distintos h, este presentará un comportamiento muy irregular, lo que hace difícil su interpretación y el ajustar funciones matemáticas que lo modelen; y además este será poco representativo ya que estará calculado en base solo a pocos pares de datos.

Para evitar los problemas descritos en el párrafo anterior y de esa forma hacer el cálculo más robusto, se añaden tolerancias a las distancias y direcciones involucradas. En el caso unidimensional se simplifica a considerar la distancia h como válida si ésta se encuentra dentro del rango $[h - \Delta h, h + \Delta h]$; para el caso de dos o tres dimensiones además de la distancia se añaden tolerancias para la orientación dentro de la que se considera válido asociar los datos en pares.

2.2.3.2 Anisotropías y Mapa variográfico

Si el variograma experimental es igual, independiente de la dirección en la que se calcula, este se denomina isótropo y dependerá solamente del modulo $|h|$. De modo contrario, si difiere al cambiar la dirección de cálculo, se debe a la existencia de una anisotropía, la que caracteriza la extensión de un fenómeno en una determinada dirección.

Para identificar la existencia de anisotropías en algún caso, se puede proceder de dos formas: en primer lugar, como se mencionó en el párrafo anterior, se puede graficar variogramas en distintas direcciones y superponerlos, de modo que si se identifican diferencias entre ellos se debe a la presencia de una anisotropía; o también, se puede visualizar un mapa variográfico.

Un mapa variográfico consiste en graficar los valores de los variogramas experimentales en todas las direcciones sobre distintos planos que contengan parte del yacimiento en estudio, de modo de visualmente identificar tendencias espaciales que indiquen la presencia de una anisotropía. Este puede estar visualizado en escala de grises o en escala de colores de modo de ayudar a analizar e interpretar el mapa para lograr una buena identificación de la dirección y el tipo de anisotropía.

2.2.3.3 Variograma modelado

La limitante más importante que presenta un variograma experimental, es que solo está definido para distancias determinadas en función del vector h utilizado. Para resolver esta problemática se recurre a ajustar funciones matemáticas según las tendencias inferidas al observar el variograma experimental y que cumplan con las propiedades propias de un variograma modelado.

Para que una función sea un variograma, debe cumplir las siguientes propiedades:

- Paridad:

$$\gamma(h) = \gamma(-h)$$

- Nulidad en el origen:

$$\gamma(0) = 0$$

- Positividad:

$$\gamma(h) \geq 0$$

- Comportamiento en el infinito:

$$\lim_{|h|\to\infty} \frac{\gamma(h)}{|h|^2} = 0$$

- Función del tipo negativo condicional:

$$\forall\, k \in \mathbb{N}^*, \forall \lambda_1, \dots, \lambda_k \in \mathbb{R} \text{ tales que}$$

$$\sum_{i=1}^{k} \lambda_i = 0\,, \forall x_1, \dots, x_k \in D, \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i \lambda_j \gamma\left(x_i - x_j\right) \leq 0$$

De cumplirse esta última propiedad, es condición necesaria y suficiente para que sea el variograma de una función aleatoria; en donde el término "condicional" se debe a que la desigualdad está restringida al grupo de ponderadores $(\lambda_1, \dots, \lambda_k)$ con suma total nula.

El comportamiento del variograma en el origen, para distancias h muy pequeñas, dicta en gran medida la regularidad que tiene la variable regionalizada: mientras más regular se comporte el variograma en el origen más regular será la variable regionalizada, por ejemplo, un comportamiento parabólico; por otro lado, si se presenta un comportamiento muy discontinuo o errático, implica que los datos varían a una distancia muy pequeña, por lo que se aprecia poca continuidad entre ellos.

En general un variograma deja de crecer y se estabiliza a partir de una distancia h^*, conocida como alcance, en torno a un valor que se conoce como "meseta", valor para el cual se puede demostrar que es igual a la varianza a priori de la función aleatoria:

$$\gamma(h) = C(0) - C(h) \xrightarrow[|h|\to\infty]{} \gamma(\infty) = C(0) = \sigma^2$$

Las dos variables aleatorias $Z(x)$ y $Z(x+h)$ estarán correlacionadas mientras se cumpla que $h < h^*$, por lo tanto, se puede entender que el alcance h^* delimita la zona donde el variograma tiene influencia, ya que más allá de ésta, el variograma es constante e igual a la varianza, y por lo tanto las variables $Z(x)$ y $Z(x+h)$ no están correlacionadas.

La propiedad revisada anteriormente de ser una función del tipo negativo condicional es estricta y poco común para las funciones usuales, por lo que se recurre a un set de funciones relativamente establecido, que aseguran cumplir ser del tipo negativo condicional.

Algunas de las funciones más utilizadas son:

- Efecto pepita:

$$\gamma(h) = \begin{cases} 0 \; si \; h = 0 \\ C \; en \; el \; caso \; contrario \end{cases}$$

En donde C es la meseta. Esta función representa la nula correlación espacial de sus componentes, ya que este variograma alcanza su meseta para una distancia nula.

- Modelo esférico:

$$\gamma(h) = \begin{cases} C\left\{\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|}{a}\right)^2\right\} si \; h \leq a \\ C \; en \; el \; caso \; contrario \end{cases}$$

En donde a es el alcance y C es la meseta, es un modelo que describe una variable regionalizada espacialmente continua que alcanza su meseta a la distancia del alcance.

- Modelo exponencial:

$$\gamma(h) = C\left\{1 - exp\left(-\frac{|h|}{a}\right)\right\}$$

En donde a es el alcance y C es la meseta, es un modelo el cual describe una variable regionalizada espacialmente continua pero que, al contrario que el modelo esférico, solo alcanza su meseta asintóticamente. En la práctica se considera un "alcance práctico" igual a $3a$, lo que corresponde a alcanzar el 95% del valor de la meseta.

- Modelo Gaussiano:

$$\gamma(h) = C\left\{1 - exp\left(-\frac{|h|^2}{a^2}\right)\right\}$$

En donde a es el alcance y C es la meseta, es un modelo el cual describe una variable regionalizada que al igual que el modelo exponencial solo alcanza su meseta asintóticamente. Se puede definir un alcance práctico igual a $a\sqrt{3}$.

2.2.3.4 Caso multivariable

Para el caso multivariable se pueden encontrar dos tipos de variogramas: simples y cruzados; los variogramas simples o directos son los casos de una variable revisados en la sección 2.2.3.1; y los variogramas cruzados son aquellos que se pueden calcular entre pares de variables; en el caso de más de dos variables, estos se pueden calcular entre todos los pares posibles. A continuación, se presenta la fórmula de cálculo del variograma cruzado, en donde el caso de variograma simple es la simplificación al caso en donde $i = j$:

$$\gamma_{ij}(h) = \frac{1}{2} cov\{Z_i(x+h) - Z_i, Z_j(x+h) - Z_j\}$$

2.2.4 Modelo de corregionalización lineal

En la sección 2.2.3.3 se presentan modelos de variogramas básicos que pueden ser utilizados para modelar algún fenómeno con un alcance determinado, pero en la realidad se presentan cambios en la continuidad espacial a partir de una o varias distancias distintas al alcance final del variograma, por lo que ajustar solo una de las funciones descritas anteriormente no es suficiente para lograr una buena aproximación a la forma del variograma experimental.

Como solución se utiliza modelar los variogramas como una suma de las funciones de variogramas descritas con anterioridad, de modo de facilitar lograr un mejor ajuste. Esto se conoce como modelos o estructuras anidadas y se representa para el caso univariable con la formula mostrada a continuación:

$$\gamma(h) = \gamma_1 + \gamma_2 + \gamma_3 + \cdots + \gamma_n$$

en donde γ_i son funciones que cumplen las propiedades de un variograma modelado y generalmente cada una de estos variogramas representa un fenómeno afectando a la continuidad espacial hasta una cierta distancia específica, es decir, cada uno de los variogramas posee un alcance determinado distinto a las demás.

Este planteamiento que diferencia los variogramas utilizados en los modelos anidados, entrega la posibilidad de analizar por separado fenómenos que afectan la continuidad espacial hasta determinados alcances. En particular para la metodología propuesta en el presente trabajo, permite determinar si utilizar o no utilizar ciertos variogramas para permitir o evitar introducir al modelo información de fenómenos acotados a algún determinado alcance, por ejemplo, si se filtra una función de variograma representando un alcance pequeño, se reducirá la variabilidad a distancias cortas.

Si se requiere extender el planteamiento univariable de modelos anidados a dos o más variables, se está en presencia del modelo de corregionalización lineal, en el cual los variogramas simples y cruzados pueden ser representados por una combinación lineal de funciones de variogramas básicos como se muestra a continuación con $i, j = 1, \dots, N$ (Goovaerts, 1992; Emery, 2013):

$$\gamma_{ij}(h) = \sum_{u=1}^{U} b_{ij}^{u}\, g^{u}(h)$$

En donde:

- $g^{u}(h)$ = función de variograma básico.
- u = alcance del variograma básico.
- b_{ij}^{u} = coeficientes de corregionalización.
- U = cantidad de variogramas básicos anidados.

Lo que escrito matricialmente queda representado por:

$$\Gamma(h) = \sum_{u=1}^{U} B_u g^{u}(h)$$

En donde:

- $\Gamma(h)$ = matriz de variogramas.

- B_u = matriz de corregionalización.

Para que este modelo sea válido matemáticamente, para cualquier $u = 1, \dots, U$ la matriz de corregionalización B_u debe ser simétrica de tipo positivo, lo que implica que los valores propios de esta matriz deben ser positivos o nulos.

En general se calculan variogramas para caracterizar el comportamiento de una variable en el espacio y posteriormente se ajustan modelos a este cálculo, pero, en los cálculos de predicción o simulación se utiliza la covarianza, que es análoga al variograma y se puede escribir matricialmente de la siguiente forma:

$$C(h) = \sum_{u=1}^{U} B_u \rho^u(h)$$

En donde:

- $C(h)$ = matriz de covarianzas.
- ρ^u = función de correlación o correlograma básico.

2.2.5 Predicción vía kriging

Si se busca conocer algún valor en una posición donde no se tiene información, se debe predecir este valor a partir de las posiciones donde si se tiene información. Para ello, una de las herramientas más utilizadas es el kriging, el cual para su construcción plantea ciertas restricciones que se listan a continuación:

- Linealidad

El predictor es una suma lineal ponderada:

$$Z^*(x_0) = a + \sum_{\alpha=1}^{n} \lambda_\alpha(x_0) Z(x_\alpha)$$

En donde:

  - x_0 = sitio a predecir.
  - x_a = sitios con datos.
  - a = incógnitas del problema.
  - $\lambda_\alpha(x_0)$ = ponderadores dependientes del sitio a estimar.

- Insesgo

Se debe imponer que el error de la predicción tenga esperanza nula, lo que se traduce en:

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

Notar que esta restricción si bien implica que la media global de los errores tiene esperanza nula, no indica que estos por si solos sean siempre bajos; para mejorar esta situación se puede imponer la siguiente restricción de optimalidad.

- Optimalidad

En el proceso de buscar los ponderadores λ_α a utilizar, estos deben ser aquellos que minimicen la varianza del error de predicción, es decir, $var[Z^*(x_0) - Z(x_0)]$ es mínima.

Un parámetro que especificar para realizar un kriging es el espacio desde el cual se toman los datos a utilizar, este dominio del espacio se denomina "vecindad de kriging", la cual esencialmente puede tener dos variantes: vecindad única o vecindad móvil.

En una vecindad única se utilizan todos los datos, independiente de que tan cerca o que tan lejos se encuentren del punto a predecir. Si se posee gran cantidad de datos es inútil conservarlos todos, ya que los más lejanos tendrán poca o nula influencia en el cálculo de la predicción e ingresarán tiempo de cálculo innecesario al sistema.

Para reducir los tiempos de cálculo se utiliza una vecindad móvil, la cual solo considera los datos más cercanos al punto a predecir, definiendo en torno a este un límite dentro del cual son considerados los valores. En general se suele considerar una vecindad elipsoidal orientada en la misma dirección que la anisotropía observada, y para el tamaño de dicha elipse se busca un equilibrio entre la precisión de la predicción y el tiempo de cálculo de esta.

Existen dos variantes principales de kriging, las cuales se diferencian en las hipótesis que toman sobre la media de la función aleatoria pariente:

2.2.5.1 Kriging con media conocida

También conocido como "kriging simple", como su nombre lo indica, en esta versión la media de la función aleatoria asociada a la variable regionalizada se asume como conocida por lo que la hipótesis más fuerte es que $E[Z(x)] = m$, en donde m es un valor conocido.

Finalmente, luego de incluir todas las restricciones previamente planteadas, el predictor toma la forma de:

$$Z^*(x_0) = a + \sum_{\alpha=1}^{n} \lambda_\alpha Z(x_\alpha) + \left(1 - \sum_{\alpha=1}^{n} \lambda_\alpha\right) m$$

La función que cumple la media es compensar la falta de información que pueda deberse a la falta de datos por su escasez o posición lejana.

2.2.5.2 Kriging con media desconocida

También conocido como "kriging ordinario", en este caso la hipótesis fuerte es igual a la del kriging simple, es decir, $E[Z(x)] = m$, pero esta vez m es una constante desconocida, por lo que se evita que la media tenga influencia la influencia planteada anteriormente sobre la predicción.

2.2.5.3 Caso multivariable

Cuando se trabaja con más de una variable, se puede predecir cada una de ellas tomando en consideración los datos disponibles de las otras variables que estén correlacionadas. Esta correlación se ingresa al modelo a través de los variogramas cruzados revisados en la sección 2.2.3.4; este tipo de kriging considerando más de una variable se conoce como "co-kriging".

2.2.5.4 Validación cruzada

Una de las herramientas más utilizadas para estudiar que tan buenos son los parámetros, como lo son variogramas o vecindades, a utilizar en una predicción es la validación cruzada, la cual consiste en utilizar los parámetros seleccionados para uno a uno ir estimando las posiciones donde se posee información en base al resto de los datos, sin considerar la misma posición siendo predicha en el sistema; de este modo se puede saber el error entre los resultados que entregan los parámetros seleccionados y los datos donde se tiene información que se considera como información verdadera.

2.2.5.5 Suavizamiento

Una de las propiedades que posee el kriging por construcción es que los valores que tiene como resultados tienen menos fluctuaciones que los valores reales, como se puede apreciar en la ilustración 1. Este efecto puede tener consecuencias importantes cuando se están prediciendo valores que luego serán estudiados con relación a, por ejemplo, límite máximo o mínimo, como es el caso de las leyes de corte en minería o en las normas ambientales en caso de estar estimando concentración de contaminantes (Emery, 2013).

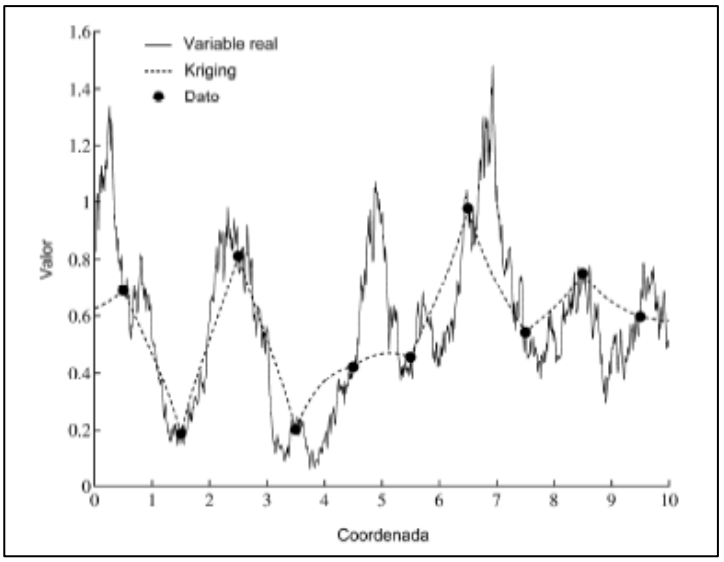

Ilustración 1 Suavizamiento de kriging (Emery, 2013)

Para evitar este efecto de suavizamiento artificial se puede recurrir a las técnicas de simulación geoestadística, que se describirán en la sección 2.2.5.

2.2.6 Simulación geoestadística

La simulación geoestadística tiene como base la construcción de una variable regionalizada auxiliar similar a aquella que se estudia, lo que implica que es capaz de reproducir la correlación espacial y coincide con los valores en los sitios en donde se tiene información. Se basa en la interpretación de la variable regionalizada z como una realización de una función aleatoria Z, pudiendo generarse más de estas realizaciones.

Uno de los mayores beneficios es que cada una de estas realizaciones de Z tiene la misma distribución y variabilidad espacial, reflejado en el mismo variograma e histograma, que la variable

real (de la cual no se puede tener información en todas las posiciones deseadas por diversos factores, incluyendo limitaciones económicas o factibilidad técnica para obtener información), por lo que frente a diversos estudios basados en limites (por ejemplo leyes de corte o normas ambientales) cada realización presenta comportamiento idéntico al que tendría la variable real.

Las propiedades de una simulación se muestran a continuación:

- Es una interpolación exacta en su versión de simulación condicional (tipo de simulación profundizada en la sección 2.2.5.3).
- Es insesgada, lo que implica que la esperanza del error es nula.
- A diferencia del kriging no suaviza, lo que implica que la dispersión de los valores simulados es igual a la de los valores verdaderos.
- No es precisa, ya que, el error entre valores reales y simulados no tiene varianza mínima; pero para remediar esto se utiliza un set con un alto número de realizaciones de la simulación para poder abarcar todos los escenarios posibles y hacer más robusto el análisis.

Además del problema del suavizamiento visto en secciones anteriores, realizar simulaciones por sobre predicciones permite realizar estudios sobre la incertidumbre, ya que, al poder obtener varias realizaciones de una misma variable, se pueden construir intervalos de probabilidad; y también permite realizar análisis de riesgo, por ejemplo, al estudiar el mejor y peor caso posible.

2.2.6.1 Simulación no condicional

En este tipo de simulaciones solo se busca reproducir la distribución y la correlación espacial, pero no considera los datos que se tengan; en otras palabras, es una simulación que, en las posiciones donde se posee información, no tiene el mismo valor que el de los datos verdaderos, pero si posee el mismo histograma y el mismo variograma que la variable verdadera.

2.2.6.2 Simulación condicional

Para que una simulación sea realmente útil para la realización de un estudio sobre algún determinado yacimiento, debe poseer la información que brindan los datos de los que se dispone. Por esto, existen las simulaciones condicionadas a los datos existentes, para las cuales se reproduce la distribución espacial de la variable verdadera, tal como en la simulación no condicional, pero esto se realiza tomando en consideración los valores reales en las posiciones en donde se tienen datos.

Se observa entonces que, al simular en un sitio donde existe información, la incertidumbre sobre su valor será cero, al ser siempre igual al valor del dato verdadero; pero, en un lugar muy lejano a los datos disponibles la distribución obtenida por las simulaciones es similar a la distribución que pudo ser obtenida por una simulación no condicional, ya que en dicha posición no habría datos condicionantes cercanos.

Puede ocurrir que, para la construcción de simulaciones condicionales, inicialmente se construya una simulación no condicional de modo de reproducir la variabilidad, y posteriormente se condicionan los datos obtenidos por la simulación no condicional mediante alguna técnica de interpolación, como por ejemplo mediante la utilización de kriging.

2.2.6.3 Simulación multi-Gaussiana

Uno de los métodos más utilizados para realizar simulaciones en el caso de las variables continuas es la simulación multi-Gaussiana. Una versión de este método será el utilizado en todas las simulaciones realizadas para el presente trabajo, por lo que sus pasos se enumeran a continuación (Emery, 2013):

- Paso 1: Desagrupar los datos.

Si la base de datos tiene posiciones que no están regularmente distribuida, esta puede inducir errores debido a que zonas del yacimiento pueden estar influyendo en mayor porcentaje que otras sobre los estadísticos. Es por esto por lo que se le asigna a cada dato un ponderador en función de la densidad de datos en la zona donde está ubicado, y de esta forma que los estadísticos globales del yacimiento lo representen de manera más exacta.

- Paso 2: Anamorfosis.

El nombre de simulación multi-Gaussiana viene dado porque se requiere que las distribuciones de las variables a simular sean distribuciones Gaussianas, lo que es poco común, ya que en general las distribuciones de las variables son asimétricas. Para solventar esto y poder usar el modelo multi-Gaussiano se aplica una transformación denominada "anamorfosis". Esta transformación se puede apreciar visualmente en la transformación que le ocurre al histograma de los datos, lo que se puede observar en la siguiente ilustración:

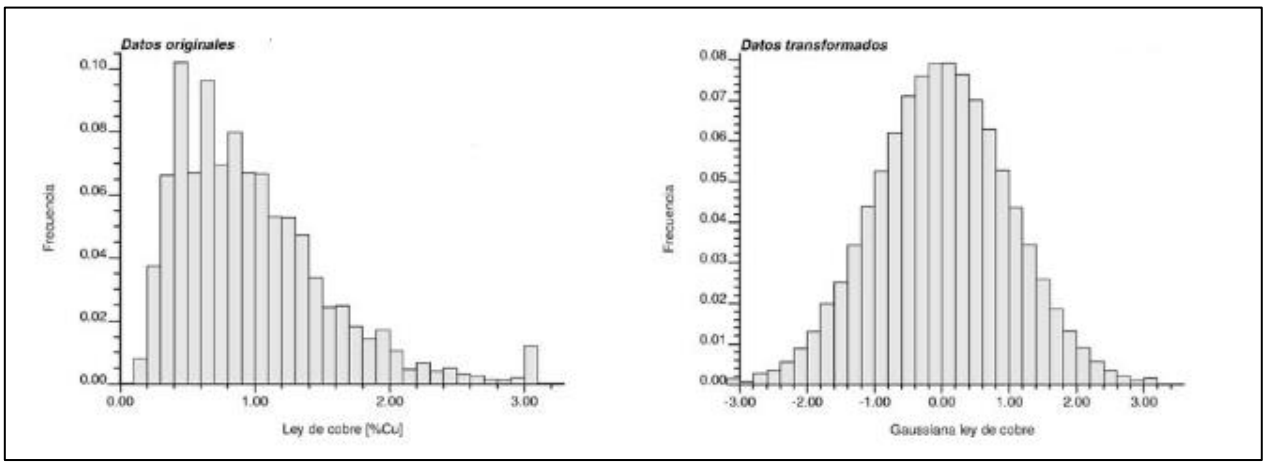

Ilustración 2 Ejemplo de anamorfosis en histogramas

Para explicar cómo se realiza la anamorfosis se puede denotar:

- Z como la función aleatoria de la variable original.
- Y como la función aleatoria de la variable transformada.
- $F(z)$ como la función de distribución de Z.
- $G(y)$ como la función de distribución de Y.

Dicha transformación consiste en asociar cada valor z de la variable original a un valor Gaussiano y de la variable transformada que coincida con la misma frecuencia acumulada de modo que $F(z) = G(y)$, lo que se puede ver más explícitamente en la siguiente ilustración:

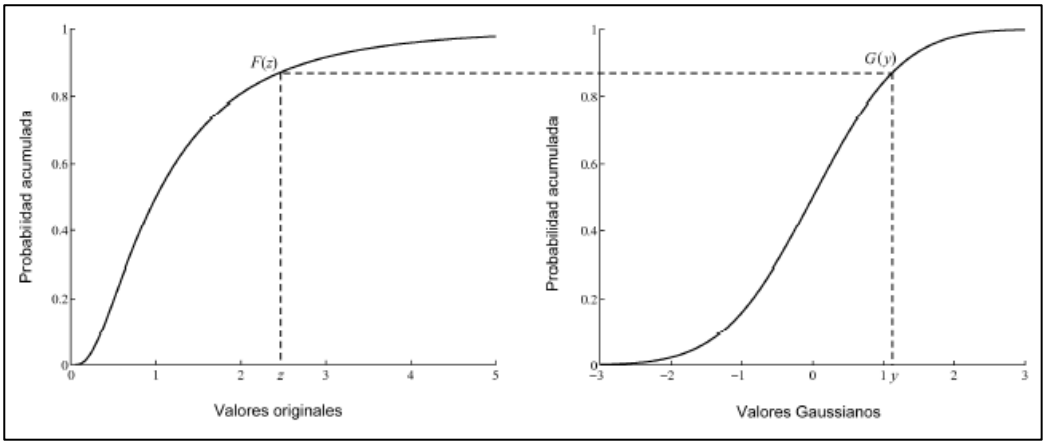

Ilustración 3 Ejemplo de anamorfosis en distribuciones (Emery,2013)

Al realizar esta transformación se crea una función que relaciona los valores originales y valores Gaussianos, lo que se puede escribir como $\forall x \in D, Z(x) = \phi[Y(x)]$. Es importante conocer esta función de anamorfosis ϕ ya que permite luego devolver los valores Gaussianos simulados a la escala de valores originales. El mayor problema que presenta esta transformación es que para poder aplicarse requiere que la distribución $F(z)$ sea invertible, dejando afuera variables categóricas y variables discretas, por lo tanto, solo permite la aplicación de este modelo a variables continuas.

- Paso 3: Análisis variográfico.

Una vez aplicada la anamorfosis a los datos originales, se procede a realizar el estudio de la relación que estos poseen entre sí, a través de las herramientas del análisis variográfico descritas en la sección 2.2.3. Esto se puede resumir planteando que inicialmente se evalúa la existencia de alguna anisotropía mediante la visualización de mapas variográficos, luego se calculan variogramas experimentales sobre los datos Gaussianos respetando la dirección de anisotropía encontrada, para finalmente ajustar funciones matemáticas al análisis experimental de modo de generar variogramas modelados.

- Paso 4: Hipótesis multi-Gaussiana.

Para aplicar el modelo en cuestión, se plantea la hipótesis de que, al tomar un conjunto de valores Gaussianos, este tendrá una distribución conjunta Gaussiana, de esta hipótesis proviene el nombre de multi-Gaussiano.

Tomando primero el caso básico de donde solo se toma una variable, basta con considerar que, al realizar la anamorfosis, el histograma de los datos originales es transformado a un histograma Gaussiano, lo que implica que la distribución univariable si cumple con ser en cualquier caso Gaussiana.

Pero, también se debe validar la hipótesis al considerar dos o más variables aleatorias en diferentes posiciones del espacio, lo que teóricamente es complejo y difícil de aplicar considerando la información limitada con la que se cuenta para realizar el estudio, por lo que generalmente solo se solicita verificar que las distribuciones bivariables (de pares de datos) cumplan con ser Gaussianas. Para comprobar experimentalmente esta hipótesis se puede recurrir a distintas pruebas.

Para el caso del presente trabajo se recurrirá a realizar una comparación entre los valores calculados para el variograma y el variograma de orden uno, que también se conoce como madograma y que se calcula al plantear:

$$\gamma_1(x_i, x_j) = \frac{1}{2} E\left[|Z(x_i) - Z(x_j)|\right], \qquad x_i, x_j \in D$$

De cumplirse la hipótesis multi-Gaussiana, se puede establecer la siguiente relación entre el variograma y el madograma:

$$\frac{\sqrt{\gamma(h)}}{\gamma_1(h)} = \sqrt{\pi}$$

Este cociente es constante e independiente del valor de h; en la práctica si se calcula el valor del variograma y madograma experimentales y la relación entre estos es aproximadamente $\sqrt{\pi}$, se considera suficiente para validar esta hipótesis.

- Paso 5: Simulación no condicional.

Se dispone de una variedad de algoritmos que potencialmente pueden ser utilizados para realizar simulaciones multi-Gaussianas, algunos de los cuales son directamente condicionados a los datos disponibles y otros que no son condicionados en el mismo algoritmo y, por lo tanto, requieren que se condicionen los resultados posteriormente. En el caso del presente trabajo se utiliza una implementación de un método no condicionado directamente llamado "método de bandas rotantes".

El método de bandas rotantes como uno de sus beneficios posee que sus cálculos pueden ser paralelizables computacionalmente, por lo que los tiempos de procesamiento son menores en comparación a otros métodos. Dicho método tiene como fundamento simplificar el sistema de tres dimensiones que presenta un yacimiento a solo una dimensión y posteriormente realizar proyecciones de vuelta a las tres dimensiones de la forma:

$$Y(x) = Y^{(1)}(< x \mid u >)$$

En donde:

- $Y^{(1)}$ = función aleatoria unidimensional.
- u = vector del espacio $\mathbb{R}^d$
- $< x \mid u >$ = proyección desde la posición x a la recta orientada por u.

Esto es posible ya que se puede establecer una relación entre las funciones de covarianza de un caso unidimensional y de un caso de dimensión d cualquiera. Se puede plantear:

$$C_d(h) = C_1(< h \mid u >)$$

En donde:

- C_1 = covarianza de $Y^{(1)}$.
- C_d = covarianza de $Y(x)$.

El vector u es un vector determinístico, en el caso en que este se aleatorice, sería reemplazado por un vector aleatorio que se puede denominar U; al realizar esto, la covarianza $C_d(h)$ se vuelve isótropa, por lo que se puede plantear:

$$C_d(h) = E\{C_1(< h \mid u >)\}$$

Al ser biyectiva, esta relación permite la simulación de una función aleatoria en un espacio de más de una dimensión, basándose en la simulación de una función aleatoria en un espacio unidimensional. Específicamente para el caso a trabajar, el de tres dimensiones, la formulación queda como:

$$C_1(r) = \frac{d}{dr}[rC_3(r)]$$

Para lograr que la simulación sea aproximadamente multi-Gaussiana, se deben sumar una gran cantidad (*N*) de simulaciones independientes, en virtud del teorema del límite central. En resumen, la simulación se obtiene siguiendo los pasos mostrados a continuación en donde $i = 1, \dots, N$:

1. Calcular C_1 en función de C_d.
2. Simular $Y_i^{(1)}$.
3. Simular una dirección u_i.
4. Plantear y calcular:

$$Y(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_i^{(1)} (< x \mid u_i >)$$

- Paso 6: Condicionar a los datos.

El proceso anteriormente descrito solo reproduce la variabilidad espacial de la variable real, por lo que se requiere restituir los valores en los sitios donde existe muestreo. Para esto se puede recurrir a la realización de un condicionamiento por kriging, al aplicar la siguiente fórmula:

$$Y_{SC}(x) = Y^{KS}(x) + [Y_S(x) - Y_S^{KS}(x)]$$

En donde:

- $Y_{SC}(x)$ = simulación condicional.
- $Y^{KS}(x)$ = valor obtenido al realizar un kriging simple en base a los datos condicionantes.
- $Y_S(x)$ = simulación no condicional obtenida del paso 5.
- $Y_S^{KS}(x)$ = kriging simple de $Y_S(x)$ a partir de sus valores en los sitios con datos.

Notar que el cálculo de los ponderadores del kriging depende de la posición en el espacio y no de los valores en sí, por lo tanto, se repiten en cada una de las realizaciones. En consecuencia, independiente de la cantidad de realizaciones realizadas, este proceso no agrega excesivo tiempo de cálculo.

- Paso 7: Anamorfosis inversa.

Las simulaciones descritas en los pasos anteriores entregan como resultado los valores con distribuciones Gaussianas. Si bien se pueden realizar ciertos análisis sobre estos, la mayor parte del

estudio se deben hacer sobre los datos en su escala original; es por esto por lo que como último paso se requiere transformar nuevamente los datos, pero esta vez desde una distribución multi-Gaussiana a la distribución de la variable original utilizando la función ϕ descrita en el paso 2.

En el caso multivariable, el paso 2 considera una anamorfosis de cada variable, el paso 4 consiste en el cálculo y ajuste de los variogramas simples y cruzados de los datos transformados, el paso 5 simula un conjunto de funciones aleatorias Gaussianas, mientras que el paso 6 considera el uso de co-kriging en lugar de kriging.

2.3 Clasificadores

Una vez obtenidas las simulaciones del yacimiento en estudio, se puede modelar y ajustar distintos tipos de herramientas de *machine learning* o aprendizaje automático (Mitchell, 1997) para, a través de éstas, lograr clasificar los datos recién calculados en las distintas categorías deseadas. Luego de realizar las simulaciones de variables cuantitativas, se puede dividir la información disponible en dos categorías:

- En primer lugar, para las posiciones donde existen los sondajes se conoce la información cuantitativa referente a las leyes de los distintos elementos y de igual forma la asociación de estas posiciones a su correspondiente *logueo* geológico. Por lo tanto, se puede estudiar la asociación que poseen las variables cuantitativas respecto a las variables categóricas.
- Por otro lado, las posiciones donde se simulan las variables cuantitativas solo poseen información de las leyes, no existiendo información de los *logueos*. Por lo tanto, se requiere buscar una forma de clasificar dichas posiciones.

Teniendo lo anterior, se puede utilizar la primera parte de los datos para ayudar a configurar una herramienta que pueda categorizar, y luego aplicarla a la segunda parte de los datos para obtener la categoría correspondiente a estos; este tipo de metodología es la utilizada por los algoritmos de aprendizaje supervisado que serán revisados en la sección siguiente.

2.3.1 Algoritmos de aprendizaje supervisado

La principal idea detrás de los algoritmos de aprendizaje es construir un modelo capaz de integrar un set de datos o información y aprender de ésta con el fin de realizar una predicción o clasificación lo más certera posible. En general, se denomina predictor a aquellos modelos que buscan tener una respuesta cuantitativa, y clasificador a aquellos modelos que buscan tener una respuesta cualitativa; siendo esta última la forma en la que se utilizan estos algoritmos en el presente trabajo.

Para diferenciar los algoritmos de aprendizaje supervisado de aquellos no supervisados, en la presente tesis se denominará a los sets de datos con los que se busca predecir o clasificar inputs (también pueden ser llamados predictores o variables independientes), y la respuesta que se debería tener sobre estos outputs (también pueden ser llamadas respuestas o variables dependientes).

Los algoritmos supervisados son aquellos en donde se tienen disponibles inputs y sus respectivos outputs dependientes de estos, con los cuales se puede construir y entrenar el modelo clasificador; la denominación de supervisado viene dada debido a que, al tener información sobre los outputs, estos "supervisan" el proceso de aprendizaje del modelo. En el lado contrario, si no se posee

información de los outputs, estos no pueden supervisar el proceso de aprendizaje, por ende, se denominan algoritmos de aprendizaje no supervisados.

La metodología para la utilización de este tipo de algoritmos se basa como paso uno en tomar un set de datos de entrenamiento que posean inputs y outputs conocidos, generar y entrenar un modelo en base a estos; y posteriormente como paso 2 utilizar dicho modelo para ser aplicado a nuevos grupos de datos de los que solo se tiene conocimiento de los inputs y de esta forma obtener sus respectivos outputs.

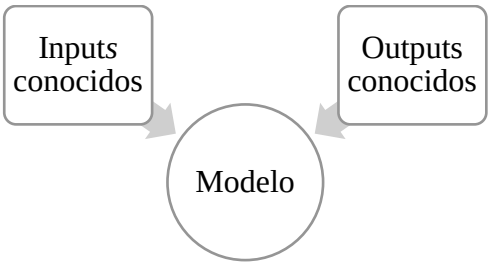

Ilustración 4 Paso 1: Entrenamiento del modelo

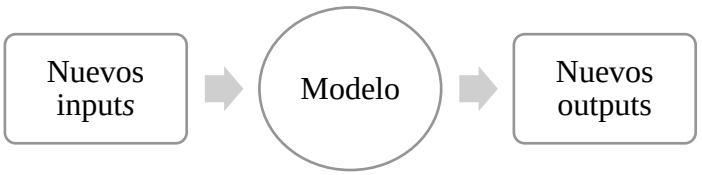

Ilustración 5 Paso 2: Procesamiento de nueva información

Los algoritmos de aprendizaje supervisado que se utilizan en la metodología del presente trabajo y que son revisados en secciones posteriores son:

- Árbol de decisión.
- Vecino más cercano.
- Support vector machine.
- Naive-Bayes.

Al existir diferentes algoritmos para realizar la tarea en cuestión, se debe cuantificar aquel que mejor realice la predicción o clasificación para el set de variables con la que se trabaja. Una de las metodologías posibles es la de validación cruzada vía K-iteraciones (Hastie, 2009).

2.3.2 Validación cruzada vía K-iteraciones

Esta validación cruzada consiste en separar la base de datos en K partes que posean aproximadamente la misma cantidad de datos, lo que permite iterativamente tomar la parte k con $k = 1, \dots, K$; aislarla como parte de validación y entrenar el clasificador con las otras $K - 1$ partes restantes. Con lo realizado anteriormente se obtiene un modelo entrenado que puede ser aplicado a la parte de los datos destinada a la validación y los consecuentes outputs obtenidos de este modelo ser comparados con los outputs originales de la base de datos; en otras palabras, permite comparar los resultados del modelo contra la realidad, pudiendo calcular de esta forma el acierto y/o errores

de cada iteración; finalmente al terminar las *K* iteraciones se calcula el promedio de aciertos y/o errores.

Este proceso puede ser repetido para distintos algoritmos o distintas configuraciones de los mismos algoritmos, lo que permite crear una escala basada en qué algoritmo o configuración de algoritmo se ajusta de mejor forma a las variables utilizadas y, por lo tanto, al ser utilizado entregará mejores predicciones o clasificaciones.

Para el caso de poseer más de una variable como outputs, también se puede analizar cuál de estas variables output posee el mayor porcentaje de acierto promedio. Al ser un clasificador que tiene como inputs las variables cuantitativas de leyes, la variable output que tenga el mayor porcentaje de acierto será aquella que mayor dependencia tenga con las leyes, y por lo tanto, será la variable que permite poder categorizar con mayor precisión.

2.3.3 Árbol de decisión

Los algoritmos basados en árbol de decisión, comúnmente conocidos como CART (por sus siglas en inglés *Classification And Regression Tree*), se dividen en dos grupos principales dependiendo del output que entregan: los árboles de regresión se utilizan para respuestas de variables cuantitativas, mientras que los árboles de clasificación se utilizan para respuestas de variables categóricas, siendo estos últimos los requeridos en el presente trabajo.

Este tipo de algoritmo trabaja de forma recursiva, iniciando con la totalidad del set de entrenamiento y progresivamente lo divide en dos partes a través de un criterio aplicado a alguna de las variables de input formando dos grupos, para los cuales se busca a través de diferentes posibles metodologías que sean lo más homogéneos posibles. Gráficamente se puede ejemplificar con la ilustración 6 en donde X_1 y X_2 son las variables usadas como inputs, t_i son limites utilizados como criterios de separación, y R_i son los posibles outputs.

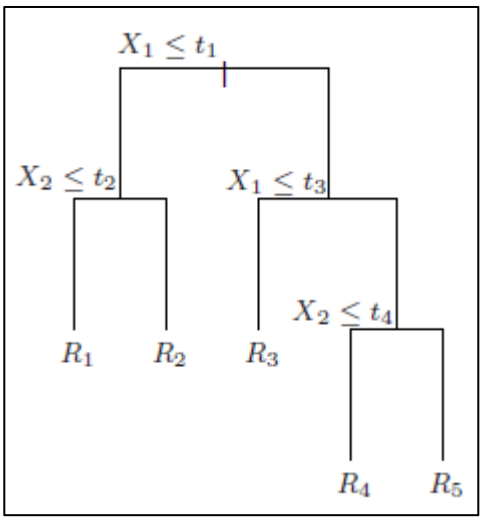

Ilustración 6 Ejemplo de árbol de decisión (Hastie, 2009)

El objetivo de un árbol de clasificación es lograr poner a cada muestra en el grupo que le corresponde con el menor error de clasificación posible. Se puede distinguir dos parámetros influyendo en el acierto de la clasificación: la cantidad de divisiones o "ramas" del árbol y la metodología utilizada para cuantificar el potencial error introducido al realizar una división, donde en la construcción del árbol se busca que el error sea el menor posible.

Para la cantidad máxima de divisiones a realizar, se debe ajustar este número en base a los datos utilizados, ya que la influencia de este parámetro dependerá de estos y la correlación que tengan. A modo de ejemplo, en algunos casos si el número de divisiones permitidas para el árbol es muy bajo, el clasificador no tendrá la información suficiente, por lo tanto, no capturará de buena forma la estructura que debería adoptar el árbol; por otro lado, si el árbol tiene demasiadas divisiones, podría sobre ajustar el modelo y hacerlo útil solo para los datos con los que se entrenó y no entregará buen acierto a nuevos datos (Hastie, 2009). Pero esto deberá estudiarse caso a caso realizando validaciones cruzadas del mismo algoritmo, pero bajo distinto número de divisiones máximas.

En el caso del cálculo de error introducido por cada división se prueban en la presente tesis las siguientes metodologías.

2.3.3.1 Índice de Gini

El índice de Gini es una medida de la impureza que posee cada nodo o división, valor que se busca minimizar con el fin de encontrar el grupo homogéneo más grande posible y separarlo del resto de los datos. Este índice es mínimo cuando los datos presentes en el nodo son de un mismo output y es máximo cuando los datos presentes en el nodo tienen una distribución pareja en todos los outputs posibles (Zambon, 2006).

Se puede calcular de la siguiente forma:

$$G(t) = \sum_{i}^{N} p_i(1 - p_i)$$

En donde:

- $G(t)$ = Índice de Gini para el nodo t.
- p_i = Frecuencia relativa del output i
- N = Número total de posibles outputs.

2.3.3.2 Índice de Twoing

El índice de Twoing al igual que el de Gini es una medida de impureza que debe ser minimizada, la diferencia es que en el presente índice para realizar divisiones identifica dos partes que representen más de la mitad de los datos restantes y, por lo tanto, crea árboles más balanceados que el índice de Gini (Zambon, 2006).

Se puede calcular de la siguiente forma:

$$T(t) = \frac{p_L p_R}{4}\left(\sum_{i}^{N}(|p(i, t_L) - p(i, t_R)|)\right)^2$$

En donde:

- $T(t)$ = Índice de Twoing para el nodo t.
- L = Identificador de la parte izquierda de la división.
- R = Identificador de la parte derecha de la división.
- $p(i, t_x)$ = Frecuencia relativa del output i al lado x del nodo t

2.3.3.3 Entropía cruzada

También conocido como reducción máxima de desviación, es una forma de medición que al contrario de los anteriores dos, es una medida de homogeneidad en vez de ser una medida de impureza. La gran ventaja que presenta este tipo de medición es que es mejor identificando categorías con comportamientos extraños que los revisados en las secciones pasadas (Zambon, 2006).

Se puede calcular de la siguiente forma:

$$E(t) = -\sum_{i}^{N} p_i \log p_i$$

En donde:

- $E(t)$ = Medición de entropía en el nodo t.

2.3.4 Vecino más cercano

El siguiente modelo en estudio se denomina vecino más cercano. Como dice su nombre se basa en la información de un número definido de muestras del set de entrenamiento que se encuentren más cerca que el punto a clasificar según una métrica a definir. Los parámetros que se deben definir son la métrica para calcular la distancia, el número de vecinos a utilizar y la utilización de ponderadores.

2.3.4.1 Métrica de cálculo de distancia

Existe gran cantidad de formas de calcular distancias entre un par de vectores, dentro de las que se puede nombrar: Mahalanobis, *City Block*, Minkowski, Chebychev, entre otras; y la distancia Euclidiana que es la utilizada en el presente trabajo, la cual se define como:

$$d^2 = (x - y)(x - y)'$$

Se calcula para todos los pares de posiciones entre el nuevo input a clasificar y los datos existentes del set de entrenamiento donde se tiene input para calcular la distancia y output que provee la información para modelar la clasificación.

2.3.4.2 Número de vecinos y ponderadores

Una definición por realizar es la cantidad de datos vecinos que se utilizan para determinar la categorización. A priori esto no se puede saber, por lo que se recurre a realizar validaciones cruzadas para distintos números de vecinos y, de este modo, poder determinar aquel que permita obtener mayor acierto al clasificar.

A priori, si se utilizan muy pocos vecinos, los utilizados serán solo los con distancias extremadamente pequeña y por lo tanto no se podrá integrar de buena forma la continuidad espacial de la variable; por otro lado, si se utilizan demasiados vecinos, estos tendrán una distancia grande al punto siendo categorizado, por lo tanto, integrarán información poco relevante; una solución a este último problema es aplicar ponderadores diferenciados a cada uno de los datos.

Los ponderadores son valores auxiliares que ayudan a priorizar en el cálculo los datos del set de entrenamiento ponderando según qué tan pequeña sea su distancia a la muestra siendo clasificada,

para dar mayor injerencia a aquellos que por tener una distancia más pequeña deberían estar más correlacionados. Algunos ejemplos de ponderadores utilizados son el inverso de la distancia o el cuadrado del inverso de la distancia.

2.3.4.3 Clasificación utilizando vecino más cercano

Para definir a que categoría pertenece un input, este algoritmo busca aquella categoría que tenga el mínimo costo esperado y la asocia al input utilizado. El costo esperado para un input se calcula como:

$$CE(j) = \sum_{i=1}^{N} \hat{P}(i|x) C(j|i)$$

En donde:

- $CE(j)$ = costo esperado de clasificar el input como categoría j.
- N = número total de posibles outputs.
- i = cada posible output.
- x = nuevo input.
- $\hat{P}(i|x)$ = probabilidad a posteriori de la categoría i para x.
- $C(j|i)$ = costo real (0 si $i = j$, y 1 si $i \neq j$).

La probabilidad a posteriori, que es la probabilidad tomando en cuenta los datos vecinos, se calcula para las muestras α dentro de los vecinos seleccionados para la clasificación, de la siguiente forma:

$$\hat{P}(i|x) = \frac{\sum_{\alpha} W(\alpha) 1_{Y(\alpha)=i}}{\sum_{\alpha} W(\alpha)}$$

En donde:

- $W(\alpha)$ = ponderadores entre α y x.
- $Y(\alpha)$ = output de α.
- $1_{Y(\alpha)=i}$ = auxiliar (1 si $Y(\alpha) = i$, y 0 si no).

2.3.5 Support vector machine

La idea detrás de la metodología de support vector machine surge inicialmente por Cortes y Vapnik (1995), como una herramienta para separar datos en solamente dos grupos, es decir, una clasificación binaria. Esta metodología tiene como objetivo encontrar un hiperplano que separe las dos clases y que a la vez provea el mayor margen posible entre los puntos más cercanos al hiperplano de las dos clases de modo de generar una mejor separación de las clases; estos puntos se definen como los vectores de soporte (*support vectors*) y son los que le dan el nombre a la metodología.

Un esquema de lo abordado en el anterior párrafo se puede apreciar gráficamente en la siguiente ilustración:

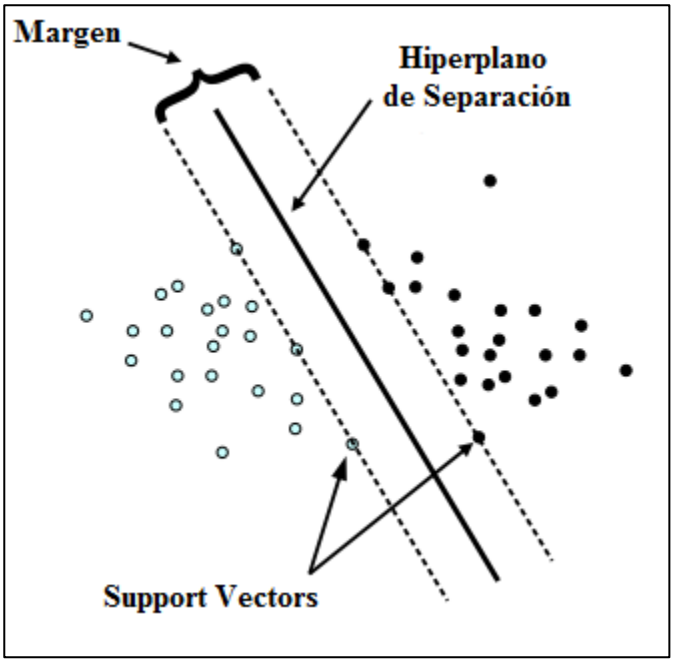

Ilustración 7 Separación utilizando hiperplano (Meyer, 2001)

Para la formulación matemática se puede definir los inputs como x_j y el output binario como y_j; para una dimensión d $x_j \in R^d$, y la ecuación describiendo el hiperplano es:

$$h(x) = x'\beta + b = 0$$

En donde:

- $\beta \in R^d$.
- b = número real.

Para asegurar el mejor hiperplano posible se debe encontrar aquellos β y b que minimicen $\|\beta\|$ sujeto a que $y_j h(x_j) \geq 1$

Desde la idea vista anteriormente se pueden identificar tres problemas principales a resolver:

1. Los puntos no siempre son completamente separables.
2. Los puntos no siempre pueden ser separados de forma lineal.
3. No siempre se busca clasificar en solo dos categorías.

2.3.5.1 Separador suavizado

Para solucionar el problema de los puntos que no puedan ser separados, es decir, que siendo de una categoría no se puede plantear un hiperplano que evite que algunos puntos queden en la categoría equivocada; se requiere a utilizar un separador suavizado que, aunque busca separar la mayor cantidad de datos posibles, no logra separarlos en su totalidad.

El trasfondo matemático del planteamiento es aplicar un castigo a aquellos puntos que no pueden ser separados, pero aun así mantener la idea de minimización del caso donde sí se pueden separar todos sus puntos, que queda planteada de la siguiente forma:

$$\min_{\beta,b,\omega}(\frac{1}{2}\beta'\beta + C\sum_{j}\omega_j)$$

Sujeto a:

$$y_j h(x_j) \geq 1 - \omega_j$$

$$\omega_j \geq 0$$

En donde:

- ω_j = puntos que no pueden ser separados.
- C = valor asignado como castigo.

Se puede notar que a mayor C se busca que la separación sea lo más estricta posible, y de modo contrario con un C bajo se le resta importancia a no poder separar los puntos.

2.3.5.2 Kernel

Debido a que los datos usualmente no pueden ser divididos por un separador lineal como el planteado anteriormente, se debe recurrir a la noción de funciones Kernel (Meyer, 2001), las cuales solucionan el mencionado problema al proyectar el set de datos a un espacio de dimensiones mayores en el cual la división entre los datos es más clara y si puede ser realizada por un separador lineal.

Se puede plantear un Kernel ϕ de la siguiente forma, en donde $m \gg d$:

$$\phi: R^d \rightarrow \mathcal{F} \subset R^m$$

Si se consideran los inputs como x_j, la aplicación de la clasificación se realiza sobre el conjunto de datos $\{\phi(x_1), \dots, \phi(x_n)\} \in \mathcal{F}$, en vez de realizarse sobre $\{(x_1), \dots, (x_n)\} \in R^d$ (Valentini, 2006). Al utilizar Kernel en la clasificación se introducirá error por el cambio de dimensionalidad, pero esto no es relevante si el set de entrenamiento es lo suficientemente grande.

2.3.5.3 Clasificación multidimensional

Debido a que este método está planteado para resolver clasificaciones binarias, se tiene un problema en el caso de tener como posibles outputs tres o más categorías; como solución usualmente se busca reducir el problema de una clasificación con múltiples categorías a múltiples clasificaciones binarias. En el caso del presente trabajo se utiliza una metodología denominada *Error Correction Output Codes* (ECOC).

En este caso se crea una serie de clasificadores binarios que permitan enfrentar todas las categorías contra todas las categorías y se decide cuál categoría tomará cada una de las muestras de inputs nuevos al aplicarles estos clasificadores binarios y proceder de forma análoga a lo planteado en la sección 2.3.4.3 para, basado en probabilidades a posteriori, es decir, tomando en cuenta datos condicionantes y minimización del costo esperado de cada categoría, decidir a cuál output pertenece cada uno de los inputs.

2.3.6 Naive-Bayes

Naive-Bayes es un clasificador basado en probabilidades, para el cual se puede describir su idea básica desglosando su nombre: la parte de Bayes viene dada por la utilización del teorema del mismo nombre que se enfoca en probabilidades condicionales, es decir, para este caso la probabilidad de que un input sea de cierta categoría condicionado a la existencia de todo el set de entrenamiento; y la parte de "naive" viene dada por el significado en inglés de esta palabra ya que se toma una condición que se puede tomar como ingenua o inocente.

2.3.6.1 Utilización de teorema de Bayes

El teorema de Bayes ayuda a calcular la probabilidad de que un evento aleatorio que se puede denominar A ocurra, dada la ocurrencia de otro evento que se puede denominar B; matemáticamente, al considerar $P(x)$ como la probabilidad de que ocurra *x*, se puede expresar como:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Si se define:

- C_i = cada uno de los k-outputs posibles en que puede ser clasificado el nuevo input.
- X = vector con los n valores x_j del set de entrenamiento.

Para construir el clasificador se puede plantear la probabilidad de que una observación nueva sea clasificada como C_i dado que existen los valores del vector X de la siguiente forma:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Para obtener la clasificación se debe encontrar aquel i que maximice la probabilidad anterior, lo que implica que el nuevo input sea asociado a la categoría seleccionada. Pero, para cualquiera sea i $P(X)$ será siempre igual, por lo que se puede sacar del análisis si lo que se desea es evaluar la categoría más probable no importando su valor por sí solo. Además, si la probabilidad $P(C_i)$ es desconocida, se suele asumir que para cualquier i la probabilidad de ocurrencia es la misma, por lo tanto, no es relevante para la comparación y solo se debe buscar que $P(X|C_i)$ sea máximo (Rish, 2001; Leung, 2007).

La expresión $P(X|C_i) = P(x_1, x_2, \dots, x_n|C_i)$ puede ser expandida según la regla de la cadena de la siguiente forma:

$$P(x_1, x_2, \dots, x_n|C_i) = P(x_1|x_2, \dots, x_n, C_i) \dots P(x_n|C_i)P(C_i),$$

expresión difícil computacionalmente de calcular, especialmente para los casos en donde n es muy grande; para resolver este problema se recurre a una simplificación.

2.3.6.2 Aplicando simplificación "naive"

Para simplificar la resolución anterior se asume que los valores del vector $X = \{x_1, x_2, \dots, x_n\}$ utilizados como condicionantes en el modelo, son independientes unos de otros para cualquier i; esta simplificación se denomina ingenua o inocente debido a que en la realidad los valores

$\{x_1, x_2, \dots, x_n\}$ no son independientes, sin embargo, el modelo presenta relativamente buenos resultados y esta condición es ampliamente utilizada.

Matemáticamente tiene la siguiente consecuencia:

$$P(X|C_i) = P(x_1, x_2, \dots, x_n|C_i) \approx P(C_i) \prod_j^n P(x_j|C_i)$$

Probabilidad que es calculada para todo i buscando encontrar aquel i que sea el que maximice el valor de la función y de esta forma ser la categoría en la que se clasifica el input en estudio. Para la maximización, como se planteó anteriormente el valor de $P(C_i)$ no tiene mayor relevancia si no se conoce su probabilidad a priori, y los valores de $P(x_j|C_i)$ pueden ser calculados a partir de los datos del set de entrenamiento, que asumiendo que tienen una distribución, generalmente Gaussiana, quedan en función de la media y la desviación estándar de los datos en función de la categoría i en estudio.

2.4 Estado del arte

En la literatura se puede encontrar diferentes estudios enfocados en la validación de los *logueos* geológicos y la validación o construcción de las interpretaciones geológicas. Para el caso de los *logueos* la mayor parte de la bibliografía disponible se concentra en formular metodologías que mediante cálculos matemáticos e identificación de patrones logren identificar zonas con potencial alto de haber sido mal clasificadas mediante la utilización de una gran variedad de técnicas; y por el lado de las interpretaciones geológicas los avances en metodologías de validación y/o construcción son recientes, ya que las publicaciones más antiguas solo tienen como idea que lo mejor es utilizar la mayor cantidad de información posible para integrarla a la interpretación por parte del equipo a cargo.

Otra arista para estudiar es la aplicación de estudios similares a la simulación con filtraje propuestas, lo cual es difícil de encontrar en la literatura disponible como tal, pero existen publicaciones con fundamentos similares.

2.4.1 Estado del arte: *logueos*

Se pueden agrupar las diferentes metodologías para validar los *logueos* geológicos, evaluar su consistencia y confiabilidad según el enfoque de las técnicas utilizadas. Es posible distinguir un primer grupo cuyas técnicas se basan principalmente en el conocimiento geológico, por ejemplo, Agterberg (1990) cuyo trabajo basado en estratigrafía y relaciones entre éstas puede ser utilizado como punto de validación; o trabajos basados en geofísica, para los cuales uno de los primeros en proponerlo fue Hoyle (1986) y una de las ultimas aplicaciones fue propuesta por Soleimani et al. (2016).

Las categorías o unidades geológicas a separar tienen algún tipo de correlación con las leyes de mineral o con alguna variable relacionada con análisis físicos o químicos; o tienen marcadas diferencias entre las diferentes categorías, por lo tanto, se pueden reconocer patrones o tendencias en base a las que se pueden realizar divisiones, lo que lleva a un grupo de metodologías que plantea la posibilidad de utilizar técnicas de reconocimiento de patrones como en el caso de Luthi y Bryant (1997), Hassibi et al. (2003) y Kuchin y Grundspenkis (2017).

Finalmente se pueden distinguir aquellas metodologías que se apoyan en la correlación espacial de las variables y técnicas de geoestadística para distinguir categorías. Se puede mencionar el trabajo de Bourgine et al. (2008, 2017), cuyo primer trabajo plantea la realización de chequeos automáticos previos y en el modelamiento y la facilitación del proceso de modelado con restricciones geológicas; en el trabajo posterior se utiliza correlación de *logueos* junto a secuencias estratigráficas para evaluar consistencia de los *logueos* realizados.

Siguiendo la línea geoestadística, Manchuk y Deutsch (2012) plantean una medida de coherencia de las diferentes categorías basado en funciones de covarianza entre datos cercanos para identificar potenciales problemas. Otro enfoque es planteado por Cáceres y Emery (2013) basado en validaciones cruzadas de una posición a la vez, iterativamente se le asigna a dicha posición cada una de las categorías disponibles y se predicen variables que luego son condicionadas por los datos de la categoría asignada; los valores obtenidos son comparados con el valor real de la posición, lo que permite concluir si la categoría que tiene asignada es la que mejor predicción permite realizar.

Siguiendo la línea de investigación anterior, Adeli y Emery (2017) proponen dos mejoras a la metodología planteada por Cáceres y Emery al extender el modelo de un marco univariable propuesto a uno multivariable, y al tomar en cuenta la posible existencia de bordes o contactos blandos, es decir, posiciones que puedan estar influenciadas por más de una categoría.

2.4.2 Estado del arte: interpretación geológica

A través del tiempo, la literatura referente a generación de interpretaciones geológicas o su validación se han enfocado principalmente en que la mejor forma de trabajarlas es integrando a la interpretación la mayor cantidad posible desde diversas fuentes. Guillen et al. (2008) proponen la posibilidad de integrar datos gravitatorios y de inversiones magnéticas; mientras que Lelièvre (2009) propone considerar restricciones basadas en datos geológicos y geofísicos.

Más recientemente, Maleki et al. (2017) propone utilizar variogramas de indicadores para obtener información referente a la geometría de los dominios geológicos y de cuál es el comportamiento en los contactos de diferentes categorías, así como también propone que mediante el cálculo de los variogramas de indicadores se puede determinar si la interpretación realizada es consistente con los datos de muestreo. Una limitación de la metodología anterior es que solo permite realizar un análisis de la consistencia global, no permitiendo análisis a escala local.

Finalmente, una nueva propuesta es realizada por Adeli et al. (2018), metodología cuya idea principal es, basado en la correlación entre variables cuantitativas corregionalizadas y categorías geológicas, aplicar clasificadores, y en base a la clasificación entregada por estos, identificar las áreas de un depósito que mayor probabilidad tengan de estar mal interpretadas.

Los principales pasos utilizados en esta metodología son:

1. Simular variables cuantitativas obviando las divisiones basadas en variables categóricas.
2. Selección y entrenamiento de clasificador.
3. Aplicación de clasificador a variables simuladas.
4. Cálculo de probabilidad de cada categoría sin y con condicionamiento a base de datos existente.
5. Identificación de bloques con mayores posibilidades de haber sido mal interpretados.

2.4.3 Estado del arte: simulaciones con filtraje

Como ya fue mencionado no existe mucha literatura al respecto, pero la información que provee estudiar los alcances de las componentes del modelo de corregionalización lineal ha sido abordada de forma levemente similar, principalmente mediante el estudio de componentes principales, método que puede ser utilizado para descomponer la matriz de corregionalización de los distintos alcances en matrices que contienen los componentes principales de la matriz de corregionalización original.

Cada uno de los componentes principales representa cierto porcentaje de la varianza de una variable, lo que permite jerarquizarlos y de esta forma generar algún tipo de filtraje. Goovaerts (1992) propone que los componentes principales pueden ser predichos para cualquier posición deseada mediante la aplicación de cokriging teniendo como variables estas componentes.

Utilizando dicha estimación propuesta, Arnaud et al. (2001) plantean una metodología que permite determinar unidades geológicas que representan diferentes escalas o alcances, esto lo logran mediante en primer lugar la predicción utilizando componentes principales para los diferentes alcances y posterior aplicación de algoritmos de clustering a los componentes más significativos obtenidos de la estimación.

El ultimo avance al respecto lo propone Laroque et al. (2016), al generar una metodología capaz de incorporar simulaciones al análisis de componentes principales, y de esta forma introducir la capacidad de obtener mejores resultados y análisis probabilísticos.

Capítulo 3: Metodología

La metodología para la implementación de las ideas de este trabajo explicitadas en la sección 1.2 tiene como base el trabajo realizado por Adeli et al. (2018) para interpretaciones geológicas, añadiéndole una mejora al implementar simulación aplicando un filtrado de los componentes de bajo alcance del modelo de corregionalización lineal, y su aplicación no solo a interpretaciones, sino que también viendo su aplicabilidad para nuevos *logueos*.

El flujo a seguir por este trabajo se diagrama en la siguiente ilustración:

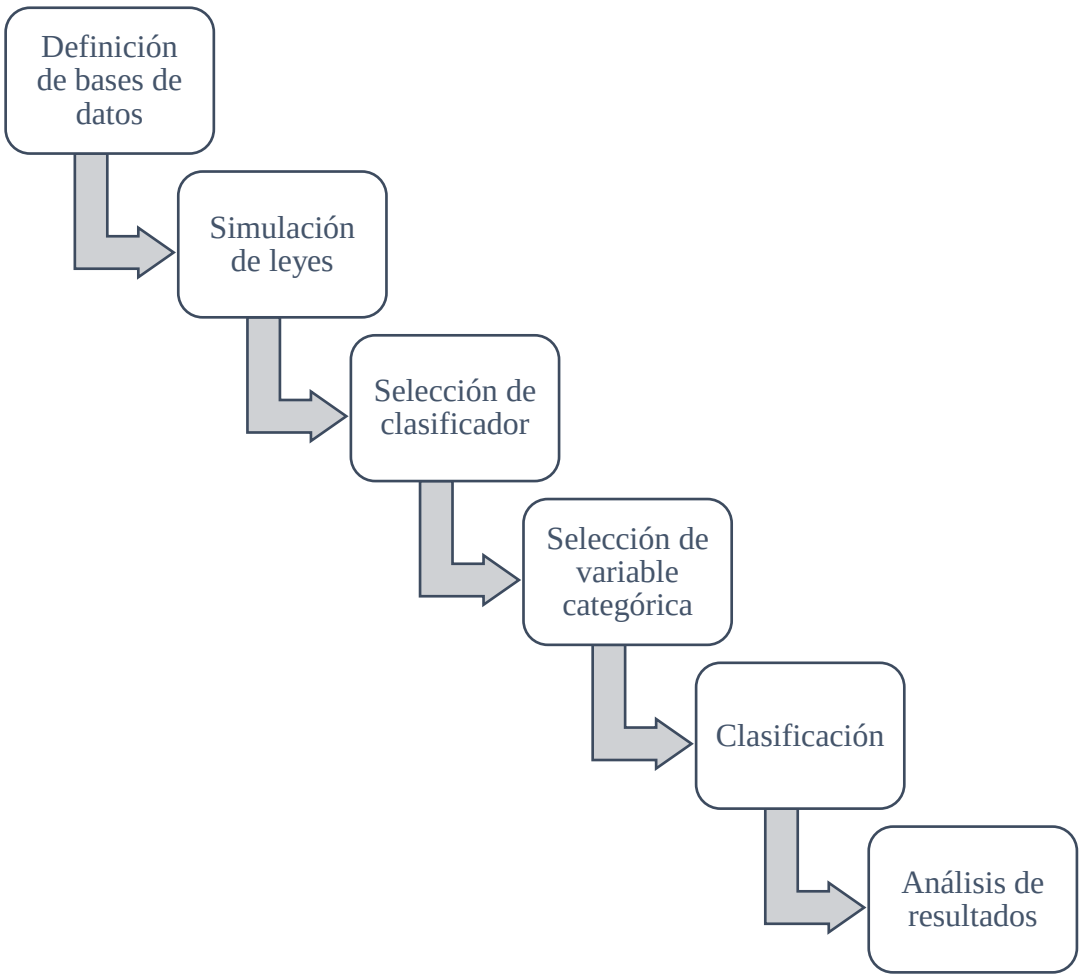

Ilustración 8 Metodología

Cada una de las etapas en la ilustración anterior se pasa a describir con mayor detalle en las siguientes secciones.

3.1 Definición de bases de datos

La aplicación de esta metodología tiene como objeto utilizar una base de datos la cual contenga información cuantitativa de las leyes de elementos de interés obtenidas de los sondajes realizados en un yacimiento, información de las categorías en las que se clasificó cada compósito de dichos

sondajes, y la interpretación geológica realizada sobre el modelo de bloques de todo el yacimiento considerado. También se puede obtener de esto el cruce de información entre la posición de cada uno de los compósitos de los sondajes y la interpretación realizada para dicha posición. Por lo tanto, en las posiciones donde existen sondajes se posee información de leyes, *logueo* geológico e interpretación geológica; y en el modelo de bloques del yacimiento se tiene información sobre la interpretación geológica y la información de las leyes puede ser simulada.

Si bien se propone que la simulación con filtraje mejora el acierto en la clasificación de una interpretación geológica, también se puede probar su funcionalidad para la clasificación de *logueos*. Para la realización de lo anterior, se presenta el problema de que los *logueos* solo están clasificados para los sondajes, los cuales a su vez ya poseen información de sus leyes a través de análisis de laboratorio.

Se propone realizar una división de los sondajes en dos partes: "Base" y "Prueba", las cuales correspondan a $x\%$ y $(100 - x)\%$ de las posiciones respectivamente, en donde la base conserva toda su información, es decir, posición, leyes y categorías; mientras que la base de prueba solo conserva posición y categorías de modo de simular los valores de las leyes en las posiciones de la base de prueba condicionadas a la información de la base, y por lo tanto, poder medir el acierto al aplicar la metodología de clasificación a simulaciones con y sin filtraje respecto de las categorías del *logueo* y de interpretación en los sondajes.

3.2 Simulación de leyes

Existen dos sets de posiciones a simular: en primer lugar las posiciones de la base de datos de prueba y en segundo lugar el modelo de bloques creado para el yacimiento y que contiene la interpretación geológica realizada. Para esto se realizan dos tipos de simulación para cada set de posiciones: simulación tradicional multi-Gaussiana como la explicada en la sección 2.2.6.3, la cual llamaremos sin filtrar. La simulación con filtraje propuesta se detalla en la siguiente sección.

3.2.1 Simulación con filtraje

Como se puede apreciar en la sección 2.2.4 de modelo de corregionalización lineal, es posible diferenciar y analizar por separado fenómenos que afectan la continuidad espacial hasta determinados alcances. En específico para esta metodología el objetivo es obtener un resultado más suavizado, por lo que se busca evitar la introducción de modelos representando alcances pequeños, de modo de reducir la variabilidad de los resultados a cortas distancias. Matricialmente se puede representar este modelo en su forma de covarianzas de la siguiente forma:

$$C(h) = \sum_{u=1}^{U} B_u \rho^u(h)$$

Al realizar una simulación no condicional, se puede plantear la variable Y^{SNC} como una suma de simulaciones para cada uno de los alcances u sumado a la media de la variable, esto puede ser representado para el caso sin filtrar como:

$$Y^{SNC} = \sum_{u=1}^{U} Y_u^{SNC} + media$$

Y para el caso de la simulación con filtraje se puede representar como:

$$\tilde{Y}^{SNC} = \sum_{u>1}^{U} Y_u^{SNC} + media$$

En donde al considerar que se realiza simulación multi-Gaussiana, luego de la realización de la anamorfosis la media de los datos es cero, por lo tanto, resulta irrelevante en ambas ecuaciones.

Posteriormente se debe continuar con el condicionamiento a los datos, para lo cual se plantea la siguiente ecuación:

$$\tilde{Y}^{SC}(x) = Y^{KS}(x) + [\tilde{Y}^{SNC}(x) - Y_{SNC}^{KS}(x)]$$

En donde:

- $\tilde{Y}^{SC}(x)$ = simulación con filtraje condicionada.
- $Y^{KS}(x)$ = valor obtenido al realizar un kriging simple en base a los datos condicionantes.
- $\tilde{Y}^{SNC}(x)$ = simulación no condicional obtenida del paso 5.
- $Y_{SNC}^{KS}(x)$ = kriging simple de $\tilde{Y}^{SNC}(x)$ a partir de valores en los sitios con datos Y^{SNC}.

En específico para el cálculo de los ponderadores de kriging simple usualmente se plantea la siguiente formula:

$$\begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \end{bmatrix}$$

En donde:

- n = número de posiciones con datos condicionantes.
- 0 = representación de la posición a la que se le aplica el kriging.
- λ_α = ponderador para la posición α.

Pero en el caso de el condicionamiento para una simulación con filtraje, se modifica la matriz de covarianza a utilizar en el segundo miembro, reemplazando la que comúnmente se utiliza por aquella calculada solo en base a aquellos alcances que se desean utilizar. Quedando planteado de la siguiente forma:

$$\begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \end{bmatrix}$$

En donde:

- $\tilde{C}_{\alpha 0}$ = covarianza calculada entre la posición α y el punto a condicionar solo considerando solo aquellos modelos con alcance deseado.

Notar que, en la simulación con filtraje, se simulan tanto las posiciones objetivo como las posiciones en donde se tienen los datos condicionantes. De este modo no solo se obtienen los valores de las realizaciones en las posiciones objetivo, sino que también se obtiene un nuevo set de datos en la ubicación de los datos condicionantes obtenidos aplicando el filtraje propuesto.

Debido a que, al modificar el modelo de corregionalización y filtrar algunas estructuras de corto alcance, los valores resultantes de la simulación no tendrán la misma distribución que los datos resultantes del proceso de anamorfosis, en particular, su varianza es menor. La función de transformación para volver a la variable original es desconocida, por lo tanto, se decide realizar la aplicación de la metodología descrita en este trabajo (clasificación) en base a las variables Gaussianas.

3.3 Selección de clasificador

Una vez obtenidos los dos tipos de simulación, se debe seleccionar aquel tipo de clasificador de aprendizaje supervisado que logre el modelo mejor ajustado de modo de que al aplicarse este, el acierto obtenido sea el mayor posible. Se prueban y compararán los siguientes métodos:

- Árbol de decisión.
- Vecino más cercano.
- Support Vector Machine.
- Naive Bayes.

Para determinar cuál es el método que mejor se ajusta a los datos, se aplicará el procedimiento de validación cruzada explicado en la sección 2.3.2, dividiendo la base de datos en cinco partes y calculando el promedio de aciertos de la aplicación a cada una de estas partes. El método seleccionado es aquel que presenta el mayor acierto.

3.4 Selección de variable categórica

Si la base de datos presenta más de una variable categórica, se busca aplicar la metodología a aquella que tenga mayor correlación con las variables cuantitativas, por lo que, se aplica el paso 3.3 de la presente metodología para cada una de las variables categóricas disponibles y se seleccionará aquella que posea el clasificador con mayor acierto para continuar con el siguiente paso.

3.5 Clasificación

Para realizar la clasificación y el posterior procesamiento de resultados se divide la aplicación de los clasificadores a tres experimentos.

3.5.1 Experimento 1

El primer experimento se realiza para determinar un punto de partida para el porcentaje de acierto de cada caso; para esto se realizan aplicaciones del clasificador a la base de datos tanto en el caso de la categoría *logueada* y a la intersección de las posiciones de la base con la interpretación geológica, cuyo resultado se denomina categoría *flagueada* o *flagueo*.

Se pueden resumir los parámetros involucrados de la siguiente forma:

3.5.1.1 Sin filtraje

- Set de entrenamiento: Base (coordenadas y leyes).

- Input para aplicar: Coordenadas y simulaciones sin filtraje sobre posiciones de Base.
- Output:
    a. Categorías *logueadas.*
    b. Categorías *flagueadas* según la interpretación geológica.

3.5.1.2 Con filtraje

- Set de entrenamiento: Coordenadas y simulaciones con filtraje sobre posiciones de Base.
- Input para aplicar: Coordenadas y simulaciones con filtraje sobre posiciones de Base.
- Output:
    a. Categorías *logueadas.*
    b. Categorías *flagueadas* según la interpretación geológica.

3.5.2 Experimento 2

El segundo experimento se realiza para obtener resultados que permitan realizar una comparación entre la mejora de acierto al clasificar sobre resultados obtenidos de simular sin filtraje y con filtraje sobre ambos tipos de variables categóricas disponibles, *logueada* y *flagueada*. Para lograr lo recientemente planteado se recurre a utilizar la simulación sobre la segunda parte de la división de la base de datos original, es decir, la base de prueba.

Se pueden resumir los parámetros involucrados de la siguiente forma:

3.5.2.1 Sin filtraje

- Set de entrenamiento: Base (coordenadas y leyes).
- Input para aplicar: Coordenadas y simulaciones sin filtraje sobre posiciones de Prueba.
- Output:
    a. Categorías *logueadas.*
    b. Categorías *flagueadas.*

3.5.2.2 Con filtraje

- Set de entrenamiento: Coordenadas y simulaciones con filtraje sobre posiciones de Base.
- Input para aplicar: Coordenadas y simulaciones con filtraje sobre posiciones de Prueba.
- Output:
    a. Categorías *logueadas.*
    b. Categorías *flagueadas.*

3.5.3 Experimento 3

El tercer experimento se enfoca en medir acierto sobre el modelo de bloques, por lo tanto, solo se pueden realizar mediciones sobre la variable categórica proveniente de la interpretación geológica, debido a que del *logueo* solo existe información en las posiciones de los sondajes; el objetivo, por lo tanto, es comparar el acierto de la clasificación con y sin filtraje.

Un resumen de los parámetros utilizados en este experimento se muestra a continuación:

3.5.3.1 Sin filtraje

- Set de entrenamiento: Base (coordenadas y leyes).
- Input para aplicar: Coordenadas y simulaciones sin filtraje sobre el modelo de bloques del yacimiento.
- Output: Categorías *flagueadas* según interpretación geológica.

3.5.3.2 Con filtraje

- Set de entrenamiento: Coordenadas y simulaciones con filtraje sobre posiciones de Base.
- Input para aplicar: Coordenadas y simulaciones con filtraje sobre el modelo de bloques del yacimiento.
- Output: Categorías *flagueadas* según interpretación geológica.

Notar que, para el caso sin filtraje el set de entrenamiento es solo una realización, que corresponde a los valores obtenidos de análisis de laboratorio. Por lo tanto, se entrena solo un clasificador que luego se aplica a cada una de las realizaciones. En el caso con filtraje, el set de entrenamiento proviene de simulaciones, lo que implica que existen varias realizaciones, por lo tanto, para cada una de las realizaciones se entrena un clasificador distinto que se aplica a su correspondiente realización a utilizar como input.

Se espera obtener para cada uno de los experimentos:

- Acierto promedio de la clasificación sin filtraje.
- Caso más probable de clasificación sin filtraje (excepto en experimento 1).
- Acierto promedio de la clasificación con filtraje.
- Caso más probable de clasificación con filtraje.

3.6 Resultados y análisis

Una vez obtenidos los resultados, el análisis más importante a realizar es la comparación del porcentaje de acierto al aplicar clasificación a variables simuladas sin filtraje versus el porcentaje de acierto al aplicar clasificación a variables simuladas con filtraje, ya que, basándose en observar un mayor acierto en el caso de la aplicación sobre simulaciones con filtraje, permite comprobar la idea de que al filtrar se puede obtener una mejor clasificación de la interpretación geológica, siendo ésta la hipótesis principal del presente trabajo.

Por otra parte, al trabajar con simulaciones existen diferentes realizaciones de éstas, por lo que se puede obtener la clasificación más probable en cada uno de los sondajes o bloques del modelo del yacimiento. Por lo tanto, se puede analizar la comparación planteada en el párrafo anterior desde dos enfoques: el acierto promedio y el acierto de la clasificación más probable. Asimismo, se puede obtener y estudiar el porcentaje de probabilidad de cada una de las categorías más probables para realizar un estudio sobre confiabilidad de la clasificación obtenida.

Capítulo 4: Caso de estudio

El caso de estudio de este capítulo utiliza una base de datos de la mina Spence, propiedad de BHP. Esta es una mina enfocada en la extracción de cobre y se ubica en la comuna de Sierra Gorda, región de Antofagasta, Chile; por temas de confidencialidad todas las leyes de la base de datos son multiplicadas por un factor no revelado.

La base de datos ha sido revisada y validada por el equipo de Recursos y Modelamiento de la empresa y posee la información de 199564 posiciones correspondiente cada una a compósitos de un total de 3159 sondajes. Cada una de las posiciones posee las siguientes variables cuantitativas:

- Ley de Cobre.
- Ley de Plata.
- Ley de Oro.
- Ley de Molibdeno.
- Ley de Arsénico.

Además, posee información de los *logueos* geológicos realizados con información sobre alteración mineral, litología y mineralización; en donde a cada posición se le asigna una categoría, teniendo las siguientes posibilidades, cuyo glosario se adjunta en la sección 8.1 de anexos:

Tabla 1 Categorías de logueo

Alteración	Litología	Mineralización
A	ABX1	EXGRAV
CL	ABX2	HYP1
KB	ABX3	HYP2
KF	AR	LIX
NONE	BRXH	NMF
P	COV	NMW
QS	FP	NONE
SA	GRAV	OXC
	IND	OXS
	NONE	OXV
	QFP1	STCP
	QFP2	SUCC
	SBR	SUCV
	SED	

Junto con la información provista sobre sondajes, también se dispone de la interpretación geológica del yacimiento, la cual fue realizada sobre un modelo de bloques de $20x20x15$ metros cada bloque, sumando un total de 5372730 bloques que poseen información de la posición y de la interpretación realizada para las variables alteración, litología y mineralización.

Al realizar el ejercicio de *flagueo*, es decir, detectar qué bloque contiene la mayor proporción de cada compósito de los sondajes, se puede obtener y asignar la información de a qué categoría de la

interpretación geológica corresponde cada uno de los compósitos, agregando nuevas variables a la base de datos de sondajes.

Las distintas categorías de la interpretación geológica se representan numéricamente y se adjuntan en la tabla 2, notando que entre paréntesis se representa la categoría del *logueo* más relacionada a cada categoría de la interpretación.

Tabla 2 Categorías de interpretación

Alteración	Litología	Mineralización
-1 (NONE)	10 (IND/SED)	-1 (NONE)
10 (A)	20 (QFP1)	10 (LIX)
20 (QS)	21 (QFP2)	11 (OXC)
30 (CL)	23 (ABX2)	20 (OXV/OXS)
40 (KB)	32 (BRXH)	24 (EXGRAV)
41 (KF)	40 (SBR)	30 (SUCC)
50 (P)	41 (GRAV)	40 (SUCV)
	42 (AR)	50 (STCP)
		60 (HYP1A)
		61 (HYP1B)
		62 (HYP2)
		71 (NMW)
		72 (NMF)

A modo ilustrativo, a continuación, se muestra cómo se distribuye en el espacio la variable litología en los sondajes, tanto para el *logueo* como para el *flagueo* de la interpretación geológica. Las visualizaciones de las variables alteración y mineralización se adjuntan en la sección 8.2 de anexo.

- *Logueo*

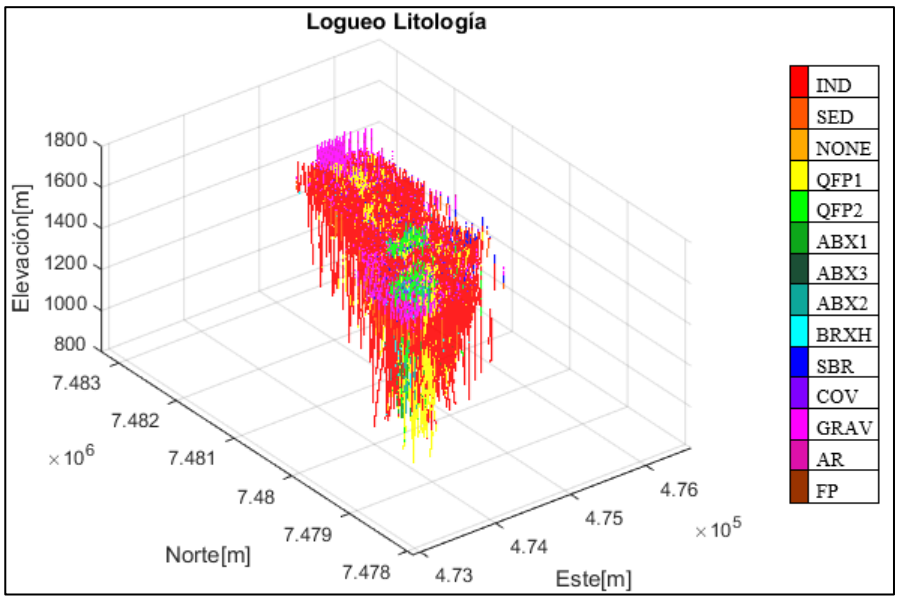

Ilustración 9 Logueos de litologías.

- *Flagueo*

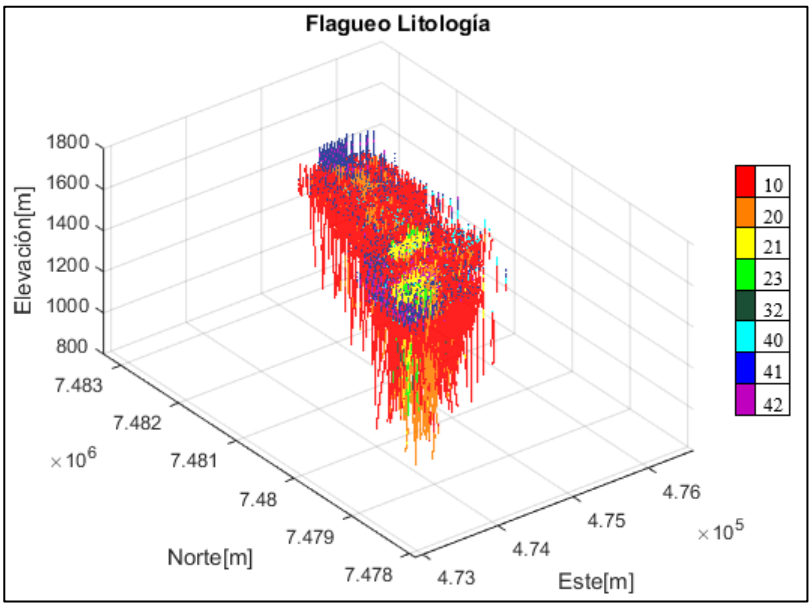

Ilustración 10 Flagueo de Litología

Visualmente se puede notar que la distribución espacial de las ilustraciones 9 y 10 son bastante similares entre sí, pero, para el *logueo* se utilizan mayor cantidad de categorías diferentes para describir un fenómeno similar, por lo que se infiere que existe más detalle que en el *flagueo*, ya que este utiliza menor cantidad de categorías.

Para mostrar el comportamiento espacial de la interpretación geológica, se muestran secciones transversales del modelo de bloques para la variable litología. En las ilustraciones 11, 12 y 13, se muestran las visualizaciones para la variable litología, las visualizaciones para las variables alteración y mineralización se adjuntan en la sección 8.3 de anexo.

Notar que, las interpretaciones geológicas están realizadas para un espacio mucho mayor al abarcado por los sondajes, esto queda en evidencia al observar el eje de elevación en la visualización de los sondajes, en donde el mínimo ronda los 800 metros; mientras que, para la interpretación el bloque más bajo se encuentra en torno a los 400 metros.

Lo anterior lleva a que, al simular ciertos bloques muy alejados de los datos condicionantes, es decir, muy lejos de los sondajes, estos estén poco condicionados a los sondajes. Por lo tanto, la correlación que posean estos bloques con la interpretación realizada será baja, por ende, su interpretación poco acertada. Para solucionar lo anterior se procede a crear una envolvente: aquellos bloques a una distancia de 250 metros o menos de un sondaje serán considerados para aplicar la metodología propuesta, y el resto se consideran fuera del análisis. Ilustraciones con la envolvente visualizando cada una de las variables se adjuntan en la sección 8.4 de anexo.

- Corte transversal a elevación 1287.5 [m]:

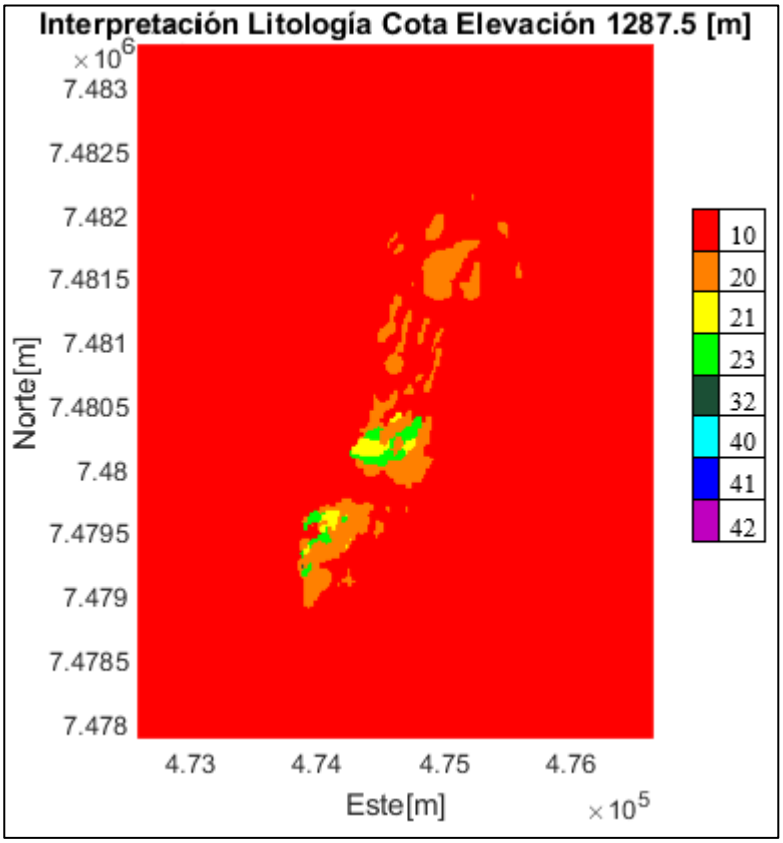

Ilustración 11 Interpretación de Litología a Elevación 1287.5

- Corte transversal a este 474610 [m]:

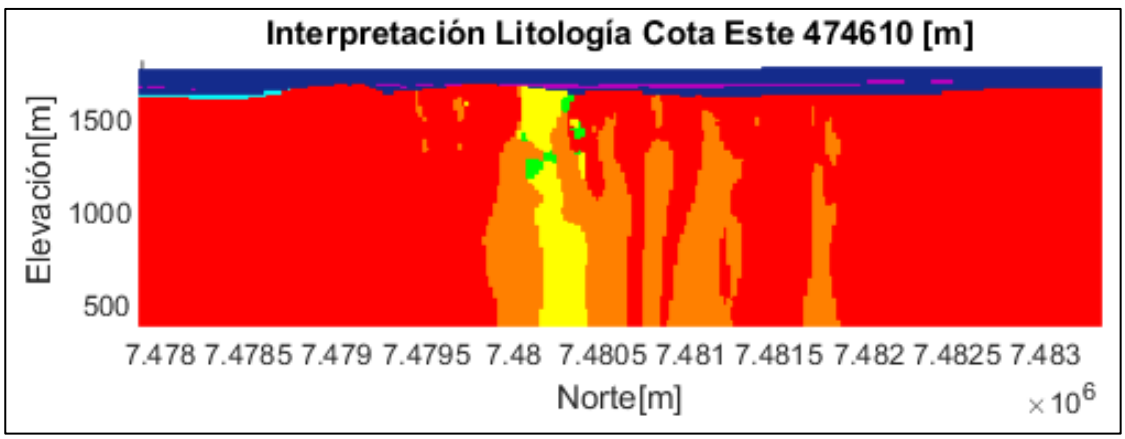

Ilustración 12 Interpretación de Litología a Este 474610

- Corte transversal a Norte 7480110 [m]:

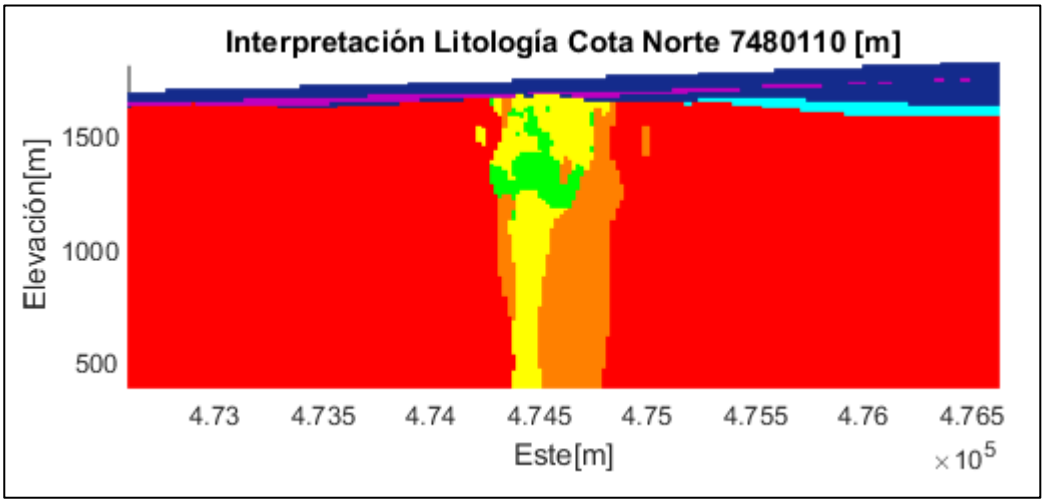

Ilustración 13 Interpretación de Litología a Norte 7480110

4.1 Definición de bases de datos

Como primer paso de la metodología propuesta en el capítulo 3, se debe definir las bases de datos a utilizar. En la sección anterior se define la envolvente del modelo de bloques a utilizar para el caso de la interpretación geológica, por lo que falta definir la base de datos "base" y "prueba".

Al ser una base de datos que tiene gran cantidad de información, se considera que la base tenga aproximadamente el 20% de los datos y la prueba en torno al 80% de los datos. Por lo tanto, se procede a seleccionar de forma aleatoria qué sondajes pertenecen a cada parte de la división. Notar que, para dar mayor realismo al ejercicio, se considera cada sondaje como un solo elemento de modo de simular el caso de tener una base de datos inicial: la "base", a la que posteriormente se le agregará la información de una nueva campaña: la "prueba".

La cantidad de datos en cada parte de la base de datos se muestra en la siguiente tabla:

Tabla 3 Separación de bases

	Número de Compósitos	Porcentaje de Compósitos	Número de Sondajes	Porcentaje de Sondajes
Base	39989	20.04%	503	15.92%
Prueba	159575	79.96%	2656	84.08%
Total	199564	100%	3159	100%

A continuación, se muestra una visualización de los sondajes con diferentes colores para cada base de datos.

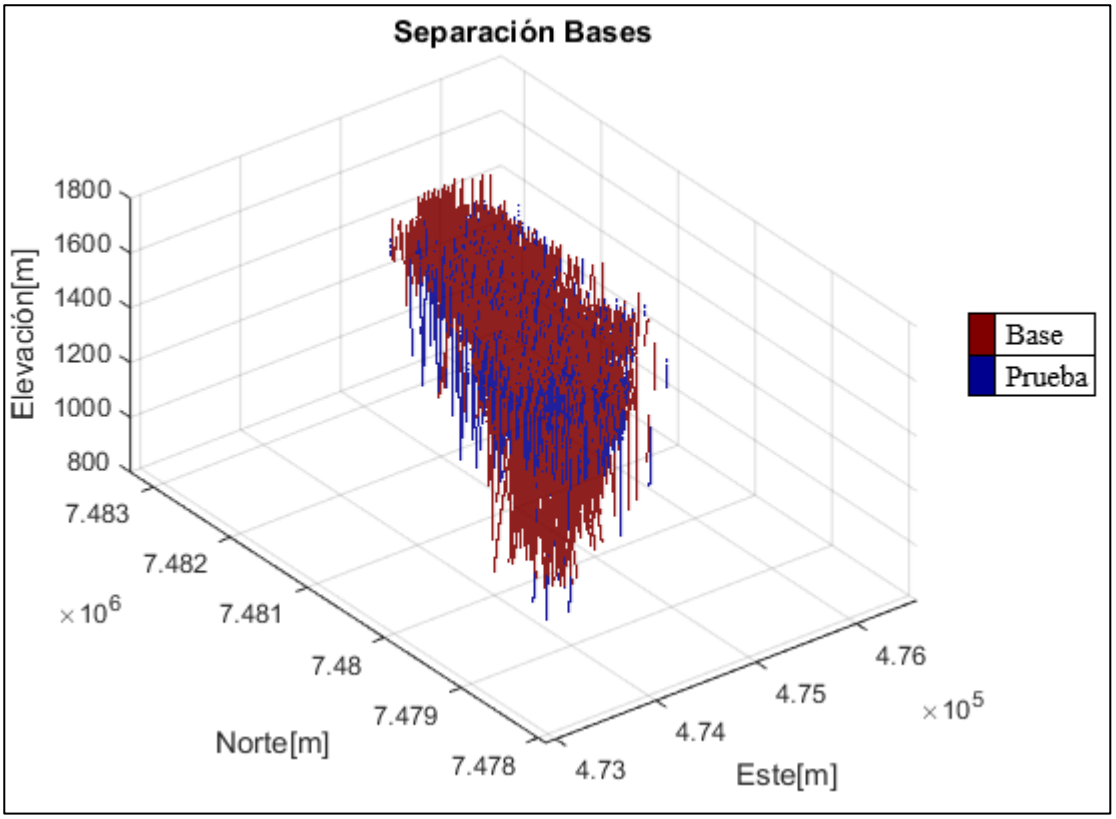

Ilustración 14 Separación de bases

Se debe recordar que la "base" será el set de datos condicionantes para la simulación a realizar en el próximo paso, mientras que la "prueba" solo mantendrá información de la posición espacial y de las variables categóricas de la interpretación.

4.2 Simulación de leyes

4.2.1 Análisis de correlación

Una vez definidas la base, la prueba y el modelo de bloques a utilizar, se procede a realizar la simulación de las leyes con y sin filtraje. Uno de los primeros pasos a realizar es un análisis exploratorio de los datos. Si bien la base de datos ya está validada por la empresa que la provee, se comprueba de igual forma que no existen datos aberrantes, datos duplicados o alguna anomalía mayor a corregir; se procede a calcular pesos de desagrupamiento mediante el método de las celdas, cuyo tamaño se define en $20x20x15\ [m]$ basado en el tamaño de cada bloque del modelo.

Se prosigue con la evaluación de la correlación entre las variables cuantitativas, para lo cual, se calcula la matriz de correlación para la "base", que se muestra a continuación:

Tabla 4 Correlación de leyes

	Cu	Ag	Au	Mo	As
Cu	1	0.553	0.526	0.334	0.092
Ag	0.553	1	0.325	0.311	0.267
Au	0.526	0.325	1	0.476	-0.081
Mo	0.334	0.311	0.476	1	0.043
As	0.092	0.267	-0.081	0.043	1

Al pertenecer la base de datos a una mina cuya explotación principal es el cobre, se toma este como el elemento más importante, por lo tanto, de la matriz de correlación se desprende un primer grupo de variables a cosimular, compuesto de cobre, plata y oro, ya que estos últimos poseen una correlación relativamente alta (mayor a 0.5) con el cobre. Los elementos restantes, molibdeno y arsénico, poseen baja correlación entre ellos, por lo que se simularán por separado.

4.2.2 Anamorfosis

Al buscar la aplicación de simulación multi-Gaussiana, el siguiente paso es transformar las variables originales a variables Gaussianas, por lo que se procede a realizar una anamorfosis. A modo de ejemplo se muestran los histogramas de ley de cobre antes y después de la aplicación de la anamorfosis.

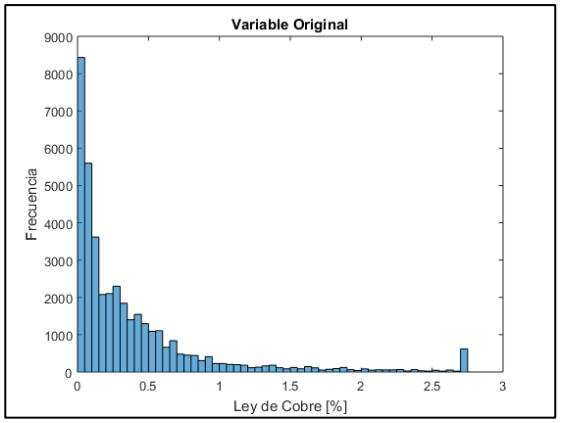

Ilustración 15 Histograma previo a anamorfosis

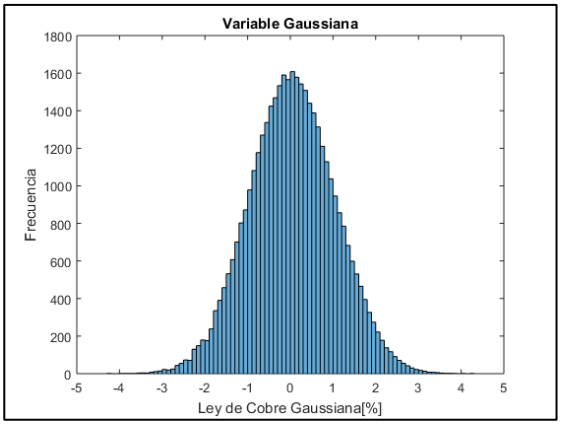

Ilustración 16 Histograma posterior a anamorfosis

Por lo tanto, se cuenta con los datos Gaussianos a utilizar en los siguientes pasos de las simulaciones con y sin filtraje.

4.2.3 Validación de hipótesis bi-Gaussiana

Para validar la hipótesis bi-Gaussiana, se utiliza la metodología descrita en la sección 2.2.6.3, en donde se busca probar que existe una relación constante entre el variograma y el madograma. Para esto se procede a calcular variogramas y madogramas omnidireccionales para las cinco variables Gaussianas y luego graficar su relación para distintos pasos, lo que se muestra a continuación:

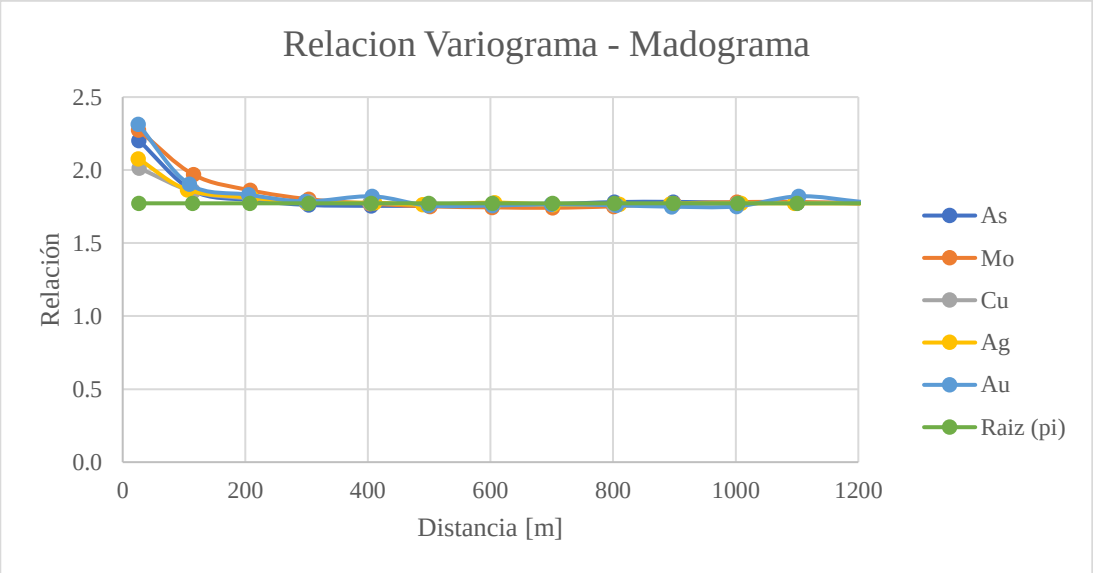

Ilustración 17 Relación Variograma-Madograma

Si bien se puede apreciar que a corta distancia las variables tienen una relación que difiere de la constante buscada (en verde). Éstas rápidamente se adecuan y en la mayor parte del gráfico la relación se cumple, por lo que, se considera valida la hipótesis bi-Gaussiana y se continúa con los siguientes pasos de la simulación.

4.2.4 Variografía

4.2.4.1 Mapas variográficos

Para la parte de análisis variográfico, se inicia con la búsqueda de posibles anisotropías a través de mapas variográficos. Notar que la base de datos sobre la que se calculan los mapas variográficos tiene incorporados fenómenos correspondientes a diferentes categorías, las que potencialmente tienen anisotropías distintas para cada una. Dicha información no se puede incorporar, ya que se buscar simular posiciones de la que a priori se desconoce a qué categoría pertenecen.

De la inspección visual en el análisis de mapas variográficos, se aprecia en general una anisotropía en la dirección aproximadamente noreste-suroeste para todas las variables, por lo que se decide para el cálculo de variogramas experimentales una dirección preferencial en la horizontal de azimut 53°, la perpendicular a esta dirección y una dirección con dip 90° para incorporar el comportamiento espacial en profundidad.

4.2.4.2 Variogramas experimentales

La siguiente etapa es calcular los variogramas experimentales de la base para los tres grupos de simulación definidos en la sección 4.2.1, y en las tres direcciones definidas en la sección 4.2.3.1. Para el cálculo de estos, se requiere definir ciertos parámetros a utilizar como: ángulos de tolerancia, tamaño de paso, número de pasos y tolerancia de paso. A continuación, se adjunta una tabla resumen con estos parámetros.

Tabla 5 Parámetros variogramas experimentales

	Direcciones horizontales	**Dirección vertical**
Tolerancia de Azimut [°]	12.5	12.5
Tolerancia de Dip [°]	12.5	12.5
Tamaño de Paso [m]	100	75
Número de Pasos	12	11
Tolerancia de Paso [m]	50	32.5

En el caso de las tolerancias angulares, éstas se determinan en un valor bajo, ya que, al tener gran cantidad de datos, aunque se calcule con una tolerancia pequeña se pueden encontrar suficientes pares de datos para entregar un valor representativo en el cálculo.

En el caso de los pasos, inicialmente se busca establecer su tamaño en un valor similar al tamaño de los bloques del modelo, pero debido a la gran cantidad de datos y extensión a cubrir se establece en un tamaño mayor para facilitar computacionalmente el cálculo, y además, reducir la variabilidad en corta escala que pudiese tener el variograma al incorporar una gran cantidad de pasos. El número de pasos se define tal que el número de pasos multiplicados por el tamaño de estos aseguren cubrir al menos la mitad de la extensión de los datos aproximadamente, y la tolerancia del paso se define como la mitad del tamaño del paso.

4.2.4.3 Variogramas modelados

Luego de obtener el cálculo de los variogramas experimentales, se procede a modelarlos con funciones conocidas; al ser una elevada cantidad de variogramas simples y cruzados y, por lo tanto, una elevada cantidad de imágenes, éstas se adjuntan en la sección 8.5.1 de anexo conteniendo tanto variogramas experimentales como el modelamiento. A continuación, se presentan tablas con el resumen de las funciones utilizadas en el modelamiento, así como también, mesetas y alcances.

- Cobre-Plata-Oro

Tabla 6 Variogramas modelados Cu-Ag-Au

Mesetas			
Modelo (alcance)	Pepita	Esférico (250,250,250)	Esférico (800,1600,1000)
Cobre	0.360	0.526	0.293
Cobre-Plata	0.042	0.332	-0.001
Cobre-Oro	0.046	0.185	0.069
Plata	0.462	0.477	0.001
Plata-Oro	0.098	0.132	0.003
Oro	0.258	0.066	0.698

- Molibdeno

Tabla 7 Variograma modelado Mo

Mesetas			
Modelo (alcance)	Pepita	Esférico (400,350,1050)	Esférico (700,600,2000)
Molibdeno	0.213	0.201	0.633

- Arsénico

Tabla 8 Variograma modelado As

Mesetas			
Modelo (alcance)	Pepita	Esférico (375,300,150)	Esférico (1000,1000,350)
Arsénico	0.319	0.271	0.152

4.2.5 Simulaciones

El siguiente paso es realizar las simulaciones, en el caso del presente estudio se realizan sobre los siguientes objetivos:

- Simulación sin filtraje:
    a. Posiciones de la base "prueba" correspondiente al experimento 2.
    b. Posiciones del modelo de bloques correspondiente al experimento 3.
- Simulación con filtraje:
    a. Posiciones de datos condicionantes correspondiente al experimento 1.
    b. Posiciones de la base "prueba" correspondiente al experimento 2.
    c. Posiciones del modelo de bloques correspondiente al experimento 3.

Notar que, para la parte sin filtraje del experimento 1 se toma como datos de input la valorización realizada por medio de análisis de laboratorios de los sondajes, es decir, aquellos valores de leyes (transformados a valores Gaussianos) que ya venían registrados en la base provista.

Con el fin de que la única diferencia entre la simulación con y sin filtraje sea justamente lo anterior, se utilizan los mismos parámetros de simulación para ambos casos como son:

- Soporte puntual.
- Elipse de búsqueda de 400 metros en las tres direcciones.
- Elipse dividida en octantes.
- Número óptimo de 3 datos por octante.
- Condicionamiento con co-kriging simple.
- 100 realizaciones.
- Método de bandas rotantes con 500 bandas.

Una forma de validar estos parámetros es realizar validaciones cruzadas y luego en base a éstas construir gráficos con pares de datos ($Valor\ real,\ Predicción$), en los cuales se busca observar que la media condicional de los valores reales en función de los valores simulados sea similar a la función identidad, lo que es indicador de una buena simulación. Dichos gráficos se adjuntan en la sección 8.5.2 de anexo.

Para el caso de la simulación con filtraje, dicho filtro se realiza de modo tal de mantener en el cálculo solo la componente de mayor alcance, dejando afuera la componente pepítica y la componente de menor alcance.

Al visualizar resultados de las simulaciones, a simple vista no se puede apreciar fácilmente cambios entre la simulación sin y con filtraje para los experimentos 1 y 2, pero esto sí se logra para el caso del experimento 3 ya que, al ser simulaciones sobre el modelo de bloques se puede visualizar secciones transversales en donde se aprecia de mejor forma las diferencias. A modo de ejemplo se muestran visualizaciones en la cota a elevación 1587.5 metros para las 5 variables sin y con filtraje; mientras que en la sección 8.5.3 de anexo se presentan visualizaciones de las cotas 1287.5 y 987.5 metros.

Se puede observar en aquellas simulaciones del lado izquierdo de cada ilustración, correspondientes a simulaciones obtenidas sin aplicación de filtraje, que existe presencia de mayor cantidad de zonas de alta ley, y que éstas tienen mayor extensión en comparación con las simulaciones del lado derecho de las ilustraciones, correspondientes a simulaciones obtenidas con aplicación de filtraje de componentes.

- Cobre

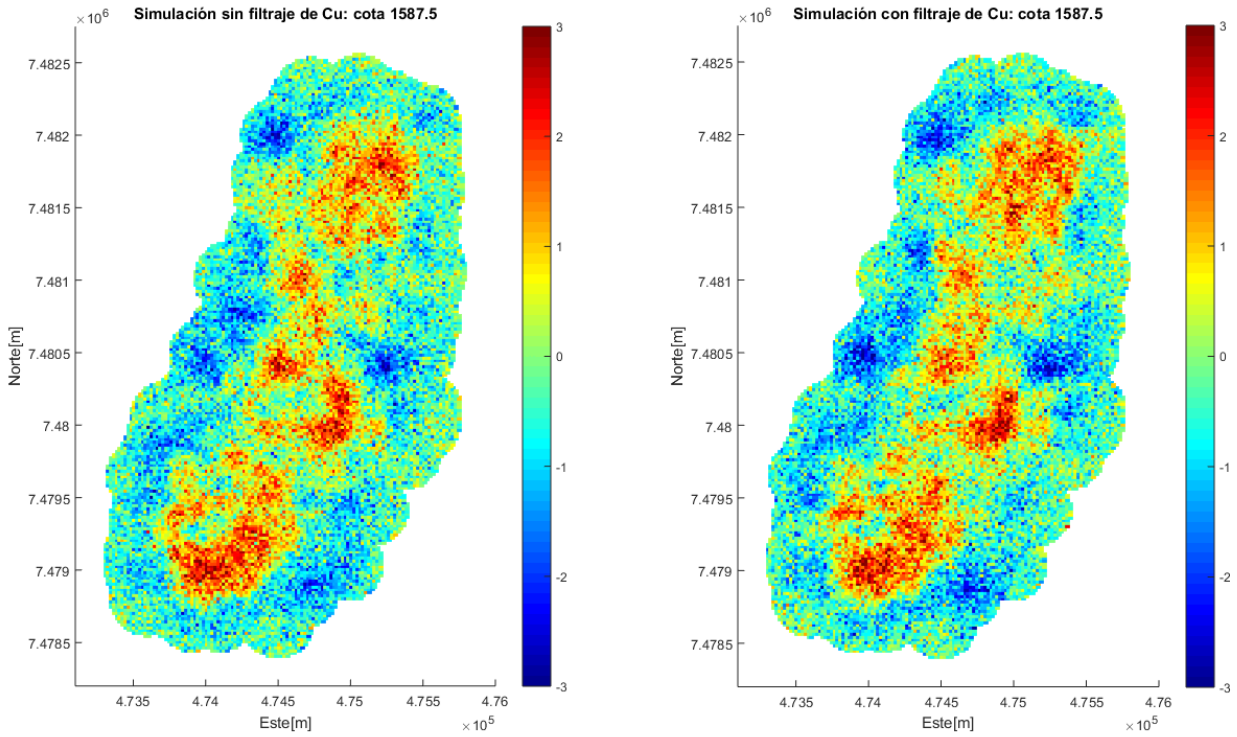

Ilustración 18 Cota 1587.5 Cu

- Plata

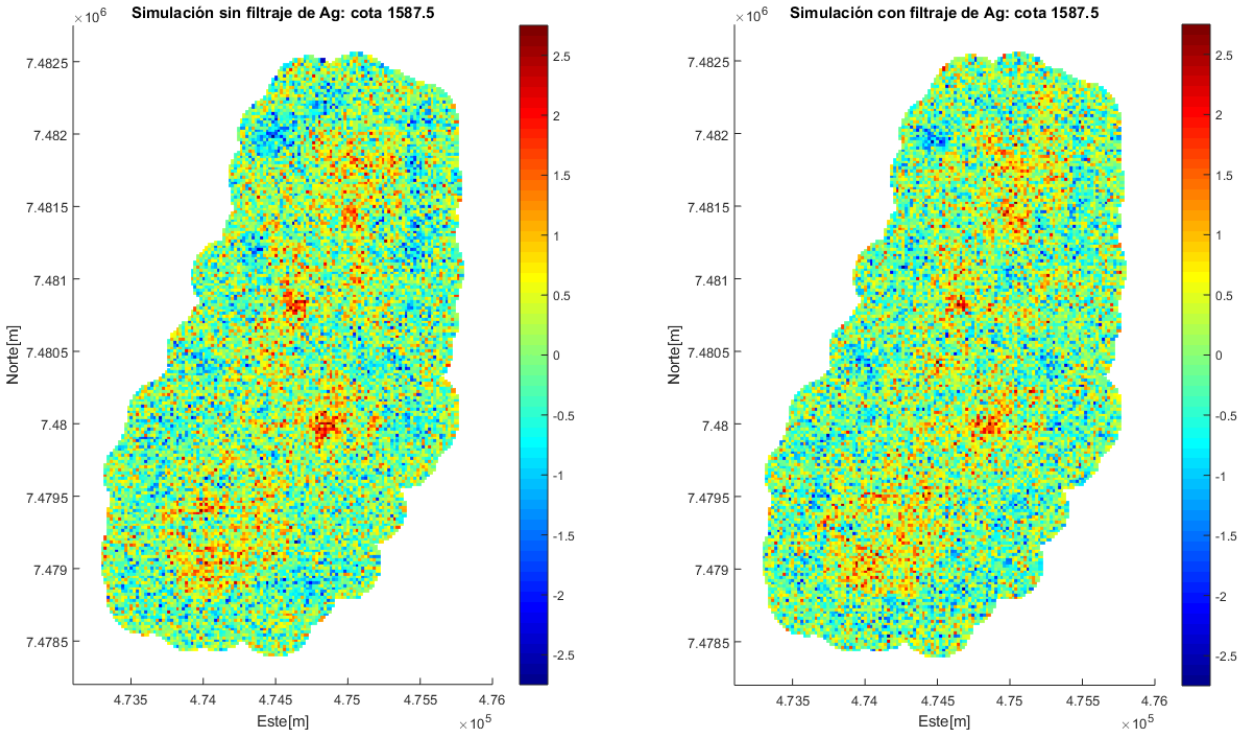

Ilustración 19 Cota 1587.5 Ag

- Oro

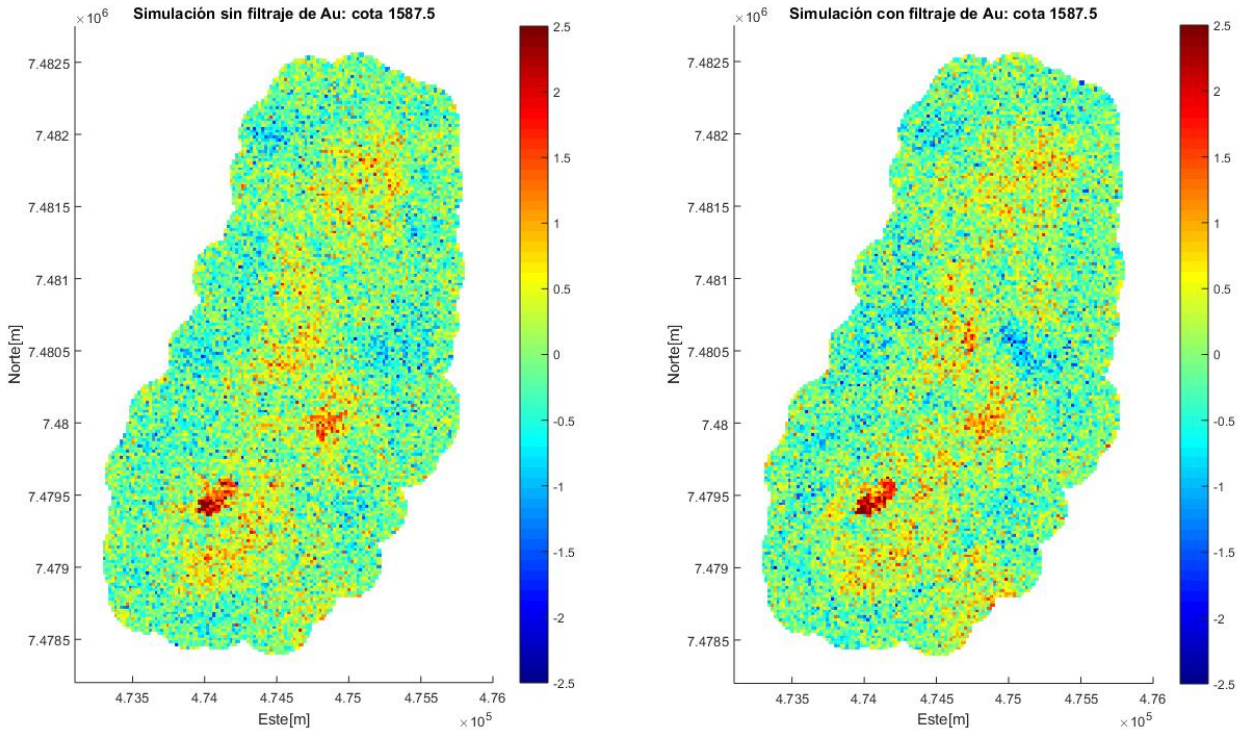

Ilustración 20 Cota 1587.5 Au

- Molibdeno

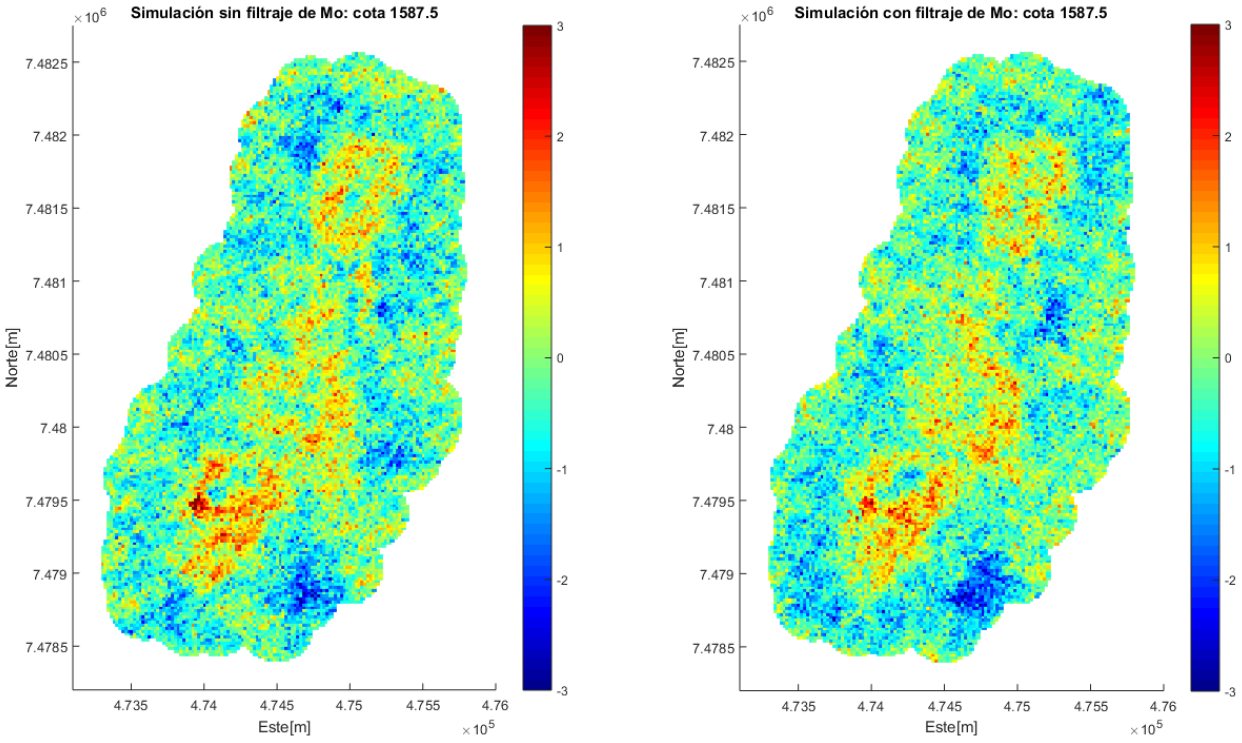

Ilustración 21 Cota 1587.5 Mo

- Arsénico

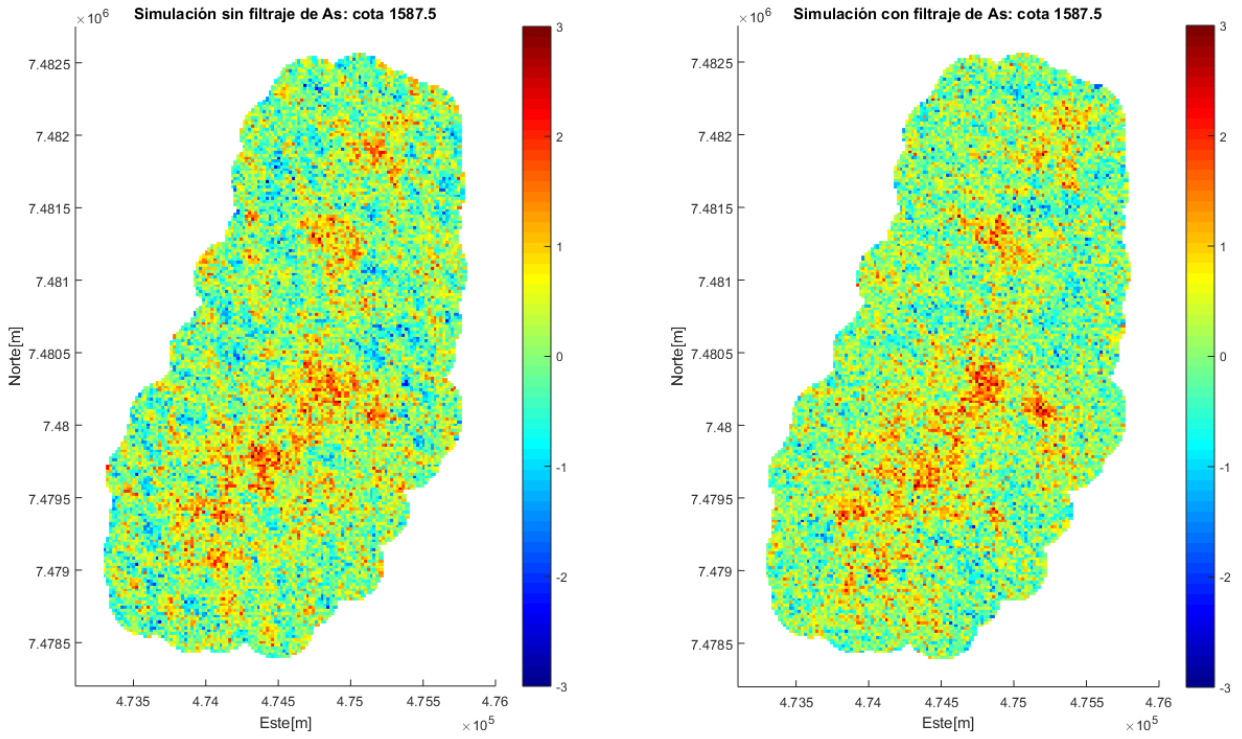

Ilustración 22 Cota 1587.5 As

Debido a la gran cantidad de datos obtenidos de las simulaciones para el experimento 3, el manejo de estas bases de datos se vuelve computacionalmente difícil ya que, para cada una de las 5 variables se realizan 100 realizaciones sin filtraje y 100 realizaciones con filtraje en cada uno de los 869672 bloques del modelo. Debido a lo anterior, los siguientes pasos de la metodología para el experimento 3 se realizan sobre seis diferentes cotas de elevación separadas cada una en 150 metros de modo de intentar representar y obtener resultados de diferentes profundidades del yacimiento, las cotas seleccionadas son:

- 837.5 [m]
- 987.5 [m]
- 1137.5 [m]
- 1287.5 [m]
- 1437.5 [m]
- 1587.5 [m]

4.3 Selección de clasificador

Para la selección del clasificador a utilizar se busca aquel que presente mayor acierto promedio luego de realizar validaciones cruzadas con 10 partes cada validación, tanto para *logueo* (según información de sondajes) como para *flagueo* (según modelo interpretado).

Los tipos de clasificador probados son

- Árbol de decisión.
    a. Con diferentes límites de divisiones: 10, 100, 300 y sin límite.
    b. Con diferentes criterios de división: Gini, Twoing y entropía.
- Vecino más cercano.
    a. Con diferentes cantidades de vecinos: 1, 10 y 100.
    b. Con diferentes pesos: Inverso de la distancia e inverso de la distancia cuadrado.
- Support vector machine.
    a. Con diferentes kernel: lineal, cuadrático, cubico, Gaussiano.
- Naive Bayes.

Los resultados obtenidos tanto para *logueo* como para *flagueo* son presentados en las tablas 9 y 10 respectivamente, en donde se aprecian los porcentajes de acierto de los diferentes clasificadores aplicados al *logueo* y al *flagueo* de la base de datos.

Se destaca en amarillo aquellos clasificadores que logran obtener el mayor porcentaje de acierto para cada una de las variables. De éstas se desprende que para cualquiera de las tres variables: alteración, litología y mineralización, el mejor tipo de clasificador posible es el árbol de decisión, utilizando para definir las divisiones el índice de Twoing, y sin limitaciones en el número de divisiones.

Tabla 9 Acierto distintos clasificadores para logueo

Tipo de clasificador		Alteración	Litología	Mineralización
Árbol de decisión	Gini max:100	62.9%	66.1%	75.8%
	Twoing max:100	70.7%	76.4%	76.9%
	Entropía max:100	66.1%	65.7%	75.1%
	Twoing max:300	74.6%	82.5%	77.9%
	Twoing	84.1%	89.6%	79.7%
	Gini	65.3%	68.2%	76.9%
Vecino más Cercano	1 vecino	57.6%	63.4%	34.0%
	10 vecinos	57.5%	63.3%	34.0%
	100 vecinos	57.0%	62.6%	33.9%
	10 con inv. Dist.	57.7%	63.4%	34.0%
	10 con inv. Dist2.	57.7%	63.5%	34.1%
	100 con inv. Dist.	57.3%	62.9%	34.0%
	100 con inv. Dist2.	57.6%	63.3%	34.1%
SVM	Kernel lineal	4.8%	62.3%	0.2%
	Kernel cuadrático	5.6%	63.1%	7.5%
	Kernel cubico	6.1%	63.3%	7.6%
	Kernel Gaussiano	5.9%	63.4%	7.5%
Naive Bayes		58.4%	59.6%	58.2%

Tabla 10 Acierto distintos clasificadores para flagueo

Tipo de clasificador		Alteración	Litología	Mineralización
Árbol de decisión	Gini max:100	67.1%	72.2%	71.8%
	Twoing max:100	73.4%	77.4%	71.7%
	Entropía max:100	67.3%	71.8%	70.1%
	Twoing max:300	78.2%	84.6%	73.3%
	Twoing	87.4%	90.8%	77.0%
	Gini	69.6%	74.0%	73.8%
Vecino más Cercano	1 vecino	60.7%	69.1%	29.9%
	10 vecinos	60.5%	69.0%	29.6%
	100 vecinos	60.0%	68.3%	29.2%
	10 con inv. Dist.	60.7%	69.1%	29.8%
	10 con inv. Dist2.	60.8%	69.2%	29.9%
	100 con inv. Dist.	60.3%	68.6%	29.5%
	100 con inv. Dist2.	60.7%	69.0%	29.8%
SVM	Kernel lineal	10.3%	68.0%	4.6%
	Kernel cuadrático	10.9%	68.8%	11.5%
	Kernel cubico	11.5%	68.9%	11.6%
	Kernel Gaussiano	11.3%	69.1%	11.7%
Naive Bayes		61.6%	66.0%	54.9%

4.4 Selección de variable categórica

Como se menciona en el capítulo 3, se busca utilizar en la presente metodología aquella variable que presente mejores condiciones para aplicar clasificadores. A modo de resumen del acierto de los mejores clasificadores posibles, se adjunta la siguiente tabla:

Tabla 11 Resumen de mejores clasificadores

	Alteración	Litología	Mineralización
Acierto en logueo	84.1%	89.6%	79.7%
Acierto en flagueo	87.4%	90.8%	77.0%

De la anterior tabla se desprende que tanto para *logueo* como para *flagueo*, la validación cruzada de clasificadores entrega como resultado que los mayores porcentajes de acierto se obtienen para la variable litología, por lo tanto, ésta es la variable con la que se continúa la metodología.

4.5 Clasificación

Una vez definida la variable objetivo y los tipos de clasificadores a utilizar para el *logueo* y para el *flagueo*/interpretación, se procede a entrenar los clasificadores y aplicarlos según se explicita en la sección 3.5.

4.5.1 Experimento 1

Los resultados de los porcentajes de acierto del experimento 1 se muestran a continuación:

Tabla 12 Resultados Experimento 1

	Logueo	Flagueo
Sin filtraje	93.81%	93.56%
Con filtraje (media de casos)	97.31%	98.34%
Con filtraje (caso más probable)	98.93%	99.82%

En el caso sin filtraje, el resultado mostrado es la aplicación de un clasificador entrenado en base a los valores de la base de datos, y aplicado sobre estos mismos; por lo tanto, al ser solo un set de datos, no se puede realizar un análisis probabilístico.

En el caso con filtraje, para cada realización de la simulación de la base, se entrena y aplica un clasificador, por lo tanto, se obtienen 100 sets de datos clasificados desde los cuales se puede calcular el acierto de cada uno y luego promediar este porcentaje de acierto, cuyo resultado se muestra en la segunda fila de la tabla 12; y también, se puede en cada una de las posiciones encontrar aquella categoría que más se repite en los 100 sets de datos, de modo de encontrar aquella clasificación que tiene más probabilidad de ocurrir y calcular sobre ésta el porcentaje de acierto que tendría, valor que se muestra en la tercera fila de la tabla 12.

4.5.2 Experimento 2

Los resultados de los porcentajes de acierto del experimento 2 se muestran a continuación:

Tabla 13 Resultados Experimento 2

	Logueo	Flagueo
Sin filtraje (media de casos)	57.72%	67.47%
Sin filtraje (caso más probable)	62.00%	71.17%
Con filtraje (media de casos)	61.25%	69.82%
Con filtraje (caso más probable)	66.00%	74.88%

En el experimento 2, para el caso sin filtraje, si bien se tiene solo un clasificador, ya que solo se tiene un set de datos para entrenar, este se aplica a las 100 realizaciones de la simulación realizada sobre la base de prueba, por lo tanto, en este experimento sí es posible obtener el porcentaje de acierto asociado al caso más probable de la simulación sin filtraje, permitiendo comparar ambas simulaciones tanto en el caso de acierto promedio como en el caso del acierto del caso más probable.

4.5.3 Experimento 3

Recordando que en el caso de este experimento solo se cuenta con información de la interpretación geológica, no existen mediciones de acierto para el *logueo*. Los resultados de los porcentajes de acierto del experimento 3 se muestran a continuación para las diferentes cotas:

Tabla 14 Resultados por Cota Experimento 3

Cota [m]	837.5	987.5	1137.5	1287.5	1437.5	1587.5
Sin filtraje (media de casos)	59.12%	63.04%	81.15%	78.13%	86.23%	84.56%
Sin filtraje (caso más probable)	59.05%	67.06%	82.36%	84.88%	86.82%	89.44%
Con filtraje (media de casos)	59.56%	62.21%	75.83%	76.65%	81.06%	88.63%
Con filtraje (caso más probable)	69.34%	71.52%	85.80%	84.24%	89.88%	92.73%
Número de bloques	4520	11459	13691	14762	17160	17915

A modo de resumen, se calcula el promedio simple de los porcentajes de acierto y también se calcula un promedio ponderado por el número de bloques presente en cada una de las secciones evaluadas. Dichos promedios se muestran en la tabla a continuación:

Tabla 15 Resultados Promedio Experimento 3

Cota [m]	Promedio	Promedio Ponderado
Sin filtraje (media de casos)	75.37%	78.59%
Sin filtraje (caso más probable)	78.27%	81.86%
Con filtraje (media de casos)	73.99%	77.11%
Con filtraje (caso más probable)	82.25%	84.96%

Para todas las cotas se puede apreciar que el caso más probable de cada simulación presenta mejor acierto que la media de los casos, por lo que se prefiere visualizar esta clasificación, lo que también permite estudiar solo una ilustración por caso, en vez de cada una de las realizaciones.

A continuación, se visualizan las cotas a elevación 837.5 y a elevación 1587.5 ya que la primera es interesante al ser la que presenta mayor diferencia entre la clasificación aplicada a simulación sin filtraje versus la clasificación aplicada a simulación con filtraje; y la segunda es interesante al ser la cota que presenta en promedio mayor porcentaje de acierto.

Notar que las visualizaciones de las otras cuatro cotas en estudio se adjuntan en la sección 8.5.4 de anexo.

4.5.3.1 Cota: 837.5 [m]

- Clasificación más probable

En la siguiente ilustración se aprecia en la figura izquierda la clasificación realizada en la interpretación geológica, en la figura central se aprecia la clasificación realizada sobre las leyes simuladas sin filtraje, y en la figura de la izquierda se aprecia la clasificación realizada sobre las leyes simuladas con filtraje.

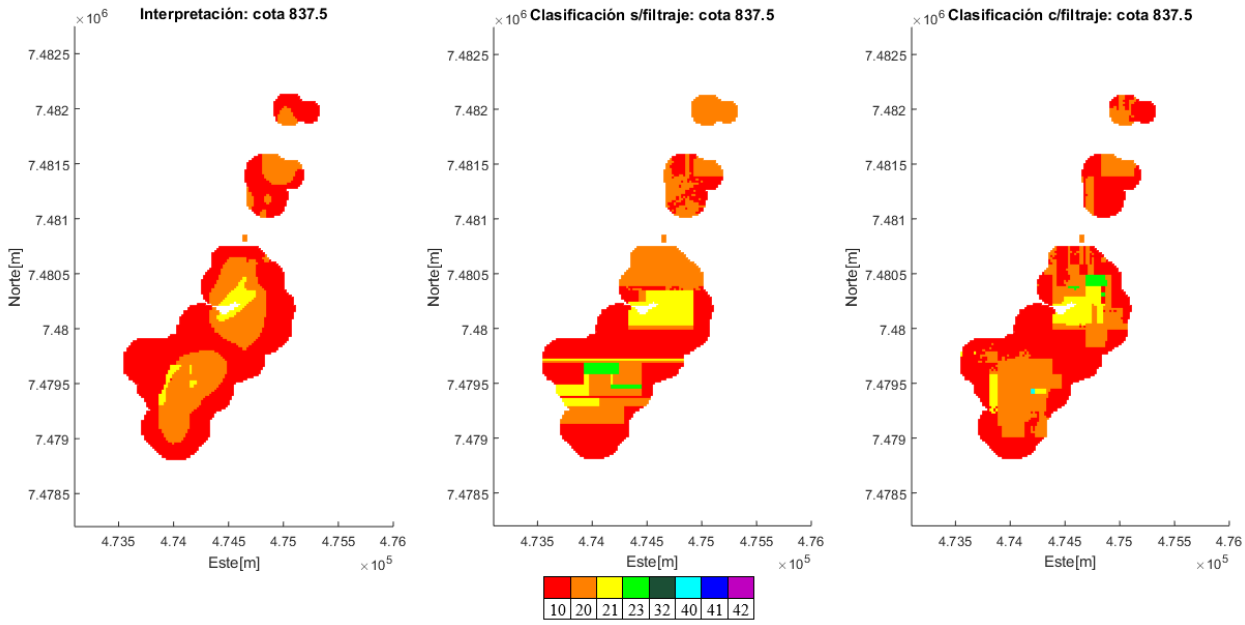

Ilustración 23 Clasificación más probable cota 837.5 [m]

- Probabilidad de clasificación

En la siguiente ilustración se aprecia en una escala de [0,1] la fracción del número de realizaciones que corresponden a la categoría más probable para cada bloque, lo que puede ser interpretado porcentualmente como la probabilidad de que la metodología seleccione cada categoría, por lo tanto, la confianza en la clasificación.

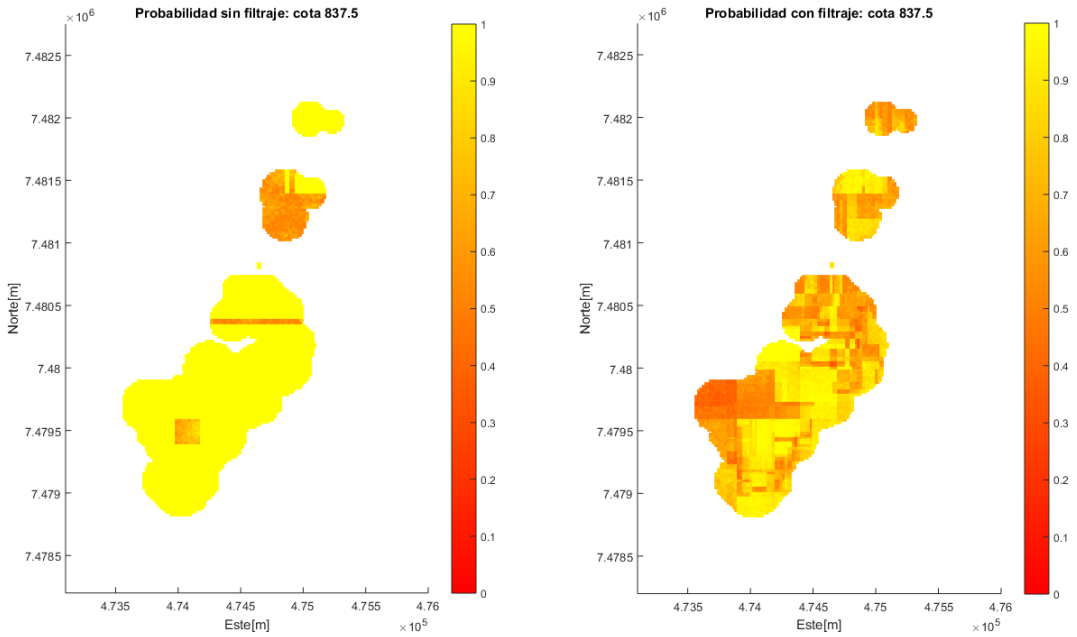

Ilustración 24 Probabilidad de clasificación cota 837.5

4.5.3.2 Cota: 1587.5 [m]

- Clasificación más probable

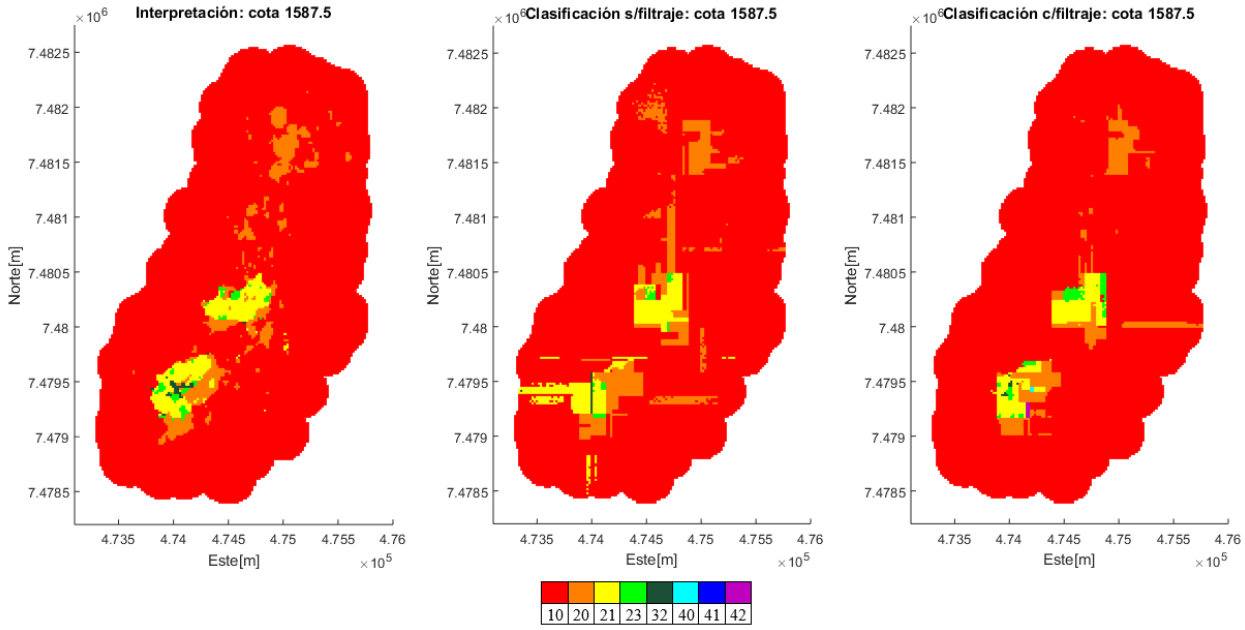

Ilustración 25 Clasificación más probable cota 1587.5 [m]

- Probabilidad de clasificación

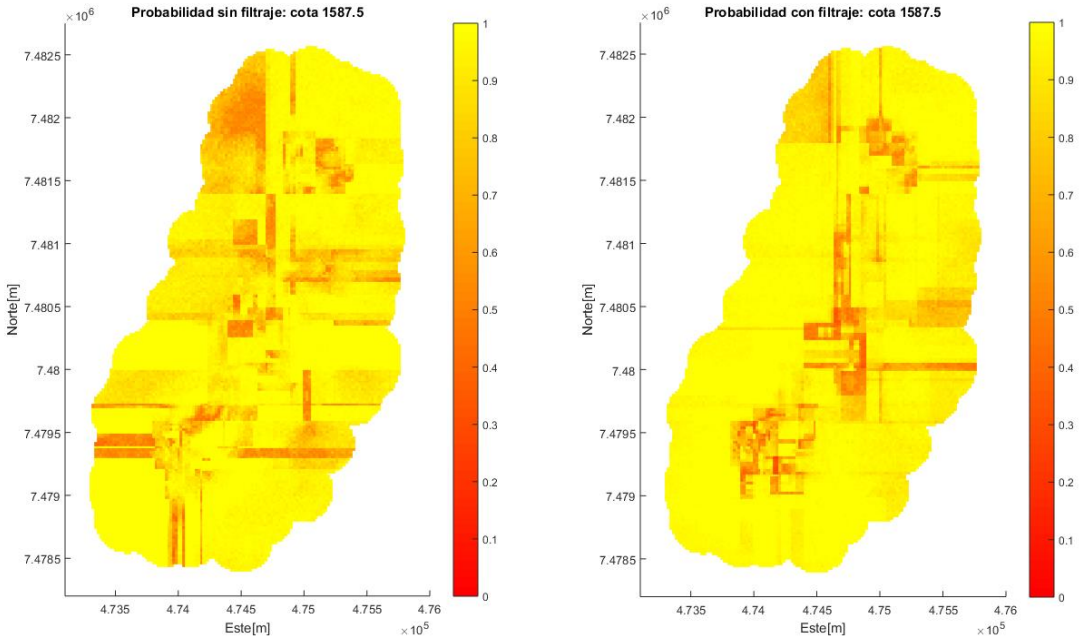

Ilustración 26 Probabilidad de clasificación cota 1587.5

4.6 Análisis de resultados

Basado en cada una de las etapas de la metodología anteriormente realizada, se pueden realizar los siguientes análisis.

4.6.1 Base de datos

La base de datos utilizada posee información de tres variables categóricas: alteración, litología y mineralización. Se puede notar que, la cantidad de categorías posibles en el caso de los *logueos* es en los tres casos mayor o igual que la cantidad de categorías posibles para el *flagueo*/interpretación, de lo que se puede inferir que el *flagueo*/interpretación es una interpretación más suave o menos detallada como se plantea en la hipótesis de esta tesis.

En los cortes transversales realizados en la interpretación geológica del modelo de bloque se aprecia en las tres variables la existencia de grandes volúmenes asociados a una baja cantidad de categorías distintas, observándose poca diseminación de las categorías; lo que apoya aún más la teoría de que la interpretación se realiza de forma suavizada, por lo que se espera que la aplicación del filtraje propuesto entregue mejores resultados que aquella sin filtraje.

4.6.2 Simulaciones

Se observa que, al realizar validaciones cruzadas de las simulaciones para corroborar los parámetros a aplicar en éstas, las medias condicionales de los pares de datos ($Valor\ real$, $Predicción$) son cercanos a la identidad, indicando que los parámetros son adecuados, por lo tanto, las simulaciones son representativas del yacimiento.

Se presenta la dificultad de que la cantidad de datos a obtener en el experimento 3 es elevada y dificulta el manejo computacional de estos. Como solución se plantea utilizar simulaciones sobre secciones transversales del modelo. Esto entrega una buena representación del yacimiento, ya que se utilizan varias secciones de este, y se abarca el yacimiento en cotas distribuidas por toda su extensión, lo que permite aplicar a la metodología a configuraciones diferentes, por ejemplo; diferente cantidad de datos condicionantes cercanos a ciertas posiciones o diferente cantidad de categorías interpretadas en cada una de las zonas; por lo tanto, los resultados obtenidos pueden ser interpretados como representativos del yacimiento completo.

En el caso de las visualizaciones de los resultados de las simulaciones con y sin filtraje, se puede apreciar que existen más zonas con acumulaciones de leyes altas para cada variable al realizar simulación sin filtraje en comparación a las simulaciones utilizado filtraje; dichas zonas son generalmente de mayor extensión en el caso sin filtraje versus lo observado en el caso con filtraje.

4.6.3 Selección de clasificadores y variable a utilizar

Para buscar aquel clasificador que mejor se ajuste a los sets de datos utilizados como inputs se realizan validaciones cruzadas de diferentes métodos, donde se observa que en general los clasificadores basados en árbol de decisión presentan mejor acierto en relación a las otras familias de clasificadores. A priori esto se puede deber a que los métodos tipo vecino más cercano o support vector machine son dependientes de la configuración espacial del set de datos a clasificar, y en el caso de un yacimiento minero estos presentan alto grado de complejidad, lo que puede perjudicar su clasificación; sin embargo, esto requiere de un análisis que no es el foco del presente trabajo, en el que se busca solo identificar el clasificador que mejor se ajuste.

Específicamente en el caso de los árboles de decisión se tiene la posibilidad de diferenciarlos por el criterio de las divisiones binarias y el número máximo de divisiones. Los mejores resultados se obtienen sin limitar el número de divisiones, lo que indica que para el set de datos utilizados el añadir mayor cantidad de información mejora la clasificación, no existiendo una merma en esta por sobreajuste del modelo. Y utilizando criterio de Twoing para la división, lo que probablemente se deba a que este criterio busca separaciones más equilibradas en cuanto a cantidad de datos en cada rama de la división, lo que ayudaría a obtener categorías de mayor extensión y menos diseminadas.

En la búsqueda de la selección de la variable a utilizar para continuar con la metodología se comparan los resultados de los mejores clasificadores tanto para *logueo* como para el *flagueo*, se aprecia que en ambos casos la litología es la variable en la cual se presenta mayor acierto, por lo tanto, esto indica que esta es la variable que mayor relación tiene con los valores de las variables cuantitativas (leyes). Por otro lado, al comparar el acierto en *logueo* versus el acierto en *flagueo*, se aprecia que el segundo es mayor para las tres variables, uno de los mayores factores de esto es la menor cantidad de categorías a diferenciar, lo que se confirma al observar que la mineralización es la variable con menos diferencia en el acierto, y es justamente la variable que tiene la misma cantidad de categorías tanto para *logueo* como para *flagueo*.

4.6.4 Clasificación

4.6.4.1 Experimento 1

En el experimento 1 se busca la clasificación de las mismas posiciones utilizadas como set de entrenamiento, se observan altos porcentajes de acierto, en todos los casos sobre el 90%, llegando

incluso a un acierto de 99.82% en clasificación del caso más probable de simulación con filtraje con respecto al *flagueo*. Al observar que los porcentajes de acierto son elevados, se puede inferir que las variables cuantitativas utilizadas permiten una buena diferenciación de las categorías buscadas.

La clasificación más probable obtenida de la aplicación sobre simulación con filtraje es la que obtiene un mayor porcentaje de acierto, pero al no disponer de un análisis probabilístico para el caso sin filtraje, se prefiere comparar el caso sin filtraje versus la media de los casos con filtraje.

En el caso del *logueo* se obtiene que el caso al que se le aplica filtraje presenta un acierto un 3.50% mayor en relación al caso al que no se le aplica filtraje, indicando que la metodología de filtraje propuesta en el presente trabajo en este caso ayuda a mejorar la clasificación; en el caso del *flagueo* los resultados presentan una diferencia de 4.78% en favor del caso con filtraje, lo que también indica que la metodología de filtraje propuesta mejora el acierto en la clasificación.

Al comparar resultados de *logueo* y *flagueo*, se puede notar que la aplicación sobre el *flagueo* presenta una mejora porcentual mayor. Esto refuerza la idea propuesta de que el filtraje ayuda a mejorar la clasificación de la interpretación debido a que ésta es suavizada, basado en la idea de que el filtraje de componentes suaviza los valores de las leyes, lo que implica que, al aplicar un clasificador basado en categorías suaves, como es el caso del *flagueo*, el acierto debería mejorar más que en la aplicación de clasificadores basados en categorías no tan suaves, como es el caso del *logueo*.

4.6.4.2 Experimento 2

En el caso del experimento 2 la baja en el porcentaje de acierto respecto al experimento 1 es esperada, ya que se aplican los clasificadores a nuevos sets de datos, no a los mismos con los que fueron entrenados, como en el caso del experimento 1, por lo tanto, el modelo no está tan ajustado a los datos a utilizar.

Tanto para la simulación sin filtraje, como para la simulación con filtraje, el caso de la clasificación más probable es el que entrega los mayores porcentajes de acierto; destacando que en ambos casos la aplicación de la metodología de filtraje propuesta mejora, tanto para *logueo* como para *flagueo*, el porcentaje de acierto de la clasificación.

En el caso de la clasificación para el *logueo* se aprecia una mejora de 4.00% al aplicar la metodología de filtraje propuesta, mientras que en el caso de la clasificación para el *flagueo* se aprecia una mejora de 3.71% al aplicar la metodología de filtraje propuesta. Esto se contradice con lo planteado en el análisis del experimento 1, donde se menciona que la mejora debería ser mayor para el *flagueo* al ser este una clasificación más suavizada, sin embargo, el *flagueo* para el caso sin y con filtraje presenta cerca de un 9% más de acierto que el *logueo*.

4.6.4.3 Experimento 3

En el caso del experimento 3 no se tiene información de una clasificación de *logueo* para el modelo de bloques, por lo tanto, no se puede realizar una comparación entre las clasificaciones para *logueo* e interpretación. Se procede entonces a comparar la influencia de la metodología de filtraje propuesta solo para la clasificación de la interpretación geológica.

Al observar los promedios de acierto, el promedio ponderado por el número de bloques de cada una de las cotas es una representación más realista del yacimiento, ya que al no considerar esto, se

le está dando mayor influencia a aquellas cotas con menor número de datos y menor influencia a aquellas cotas con mayor número de datos, por lo tanto, se estaría introduciendo un sesgo evitable al estudio.

Nuevamente se observa que la clasificación sobre los casos más probables son las que presentan mayor porcentaje de acierto, por lo tanto, las más interesantes de estudiar. Se aprecia que, para el promedio ponderado, la aplicación de la metodología de filtraje presenta una mejora del acierto de 3.10% respecto a la no aplicación del filtraje.

Al analizar cada una de las seis cotas estudiadas se confirman las inferencias obtenidas sobre los promedios, es decir, las clasificaciones sobre los casos más probables presentan mayor porcentaje de acierto, solo con una excepción, y la aplicación de la metodología de filtraje mejora dicho porcentaje para cada una de las cotas estudiadas en el caso de la clasificación más probable. Se pueden destacar las cotas 837.5 [m] de elevación, ya que ésta posee una mejora de 10.29% al aplicar el filtraje; y la cota 1587.5 [m] en donde la metodología con filtraje logra un porcentaje de acierto de 92.73%.

Si se observan las visualizaciones de la confianza sobre la clasificación que tiene cada uno de los métodos, se puede observar que, en el caso de la clasificación aplicada sobre simulaciones sin filtraje, en ciertas zonas se muestran altas probabilidades de ciertas clasificaciones, las cuales no coinciden con la interpretación geológica realizada, por lo tanto, son posiciones que se indican con alta confianza pero entregan clasificaciones erróneas, como se puede apreciar en los bordes del cuerpo más grande en la ilustración 24. Esto ocurre con mucha menor frecuencia en las clasificaciones realizadas sobre simulaciones a las que se le aplicó la metodología de filtraje, siendo esto otro de los beneficios de la idea propuesta, ya que el porcentaje de confianza sobre cada clasificación es más acertado en cuanto a la confiabilidad de cada bloque.

Capítulo 5: Discusión

La idea clave del presente trabajo se inicia basada en encontrar y utilizar la correlación entre las variables cuantitativas (leyes) y las variables categóricas, utilizando lo anterior entrenar modelos clasificadores capaces de categorizar nueva información y mediante esto construir y/o validar interpretaciones geológicas. Como innovación y en busca de mejorar lo planteado anteriormente, se toma la hipótesis de que las interpretaciones geológicas tienden a ser suavizadas, por lo tanto, si se logran simulaciones de leyes que igualmente sean suavizadas, al estudiar las relaciones entre las variables cuantitativas y categóricas, una clasificación tendría mayor acierto. Para esto se propone realizar filtraje de los componentes de corto alcance del modelo de corregionalización, lo que permitiría obtener simulaciones de leyes más suavizadas.

Por lo tanto, como se describió anteriormente uno de los requisitos para la aplicación de lo propuesto en el presente trabajo es que las diversas categorías presenten una diferenciación entre ellas basada en los valores de las variables cuantitativas, esto no es mayor problema al trabajar con variables geológicas, como es el caso de la alteración, litología o tipo de mineralización; ya que, éstas presentan influencia directa en las concentraciones de minerales.

En el caso de intentar estudiar otro tipo de dominios, por ejemplo del tipo geofísico, puede que la metodología propuesta sea más compleja de aplicar, pudiendo requerir no solo utilizar como input las posiciones y leyes de cada dato, sino que también otro tipo de variable, como pueden ser densidades o frecuencias de fracturas. De todas formas, una de las funciones e interpretaciones que se pueden obtener del experimento 1 es justamente evaluar qué tan bien se puede separar las categorías basado en las variables cuantitativas que se están utilizando.

En el caso de la búsqueda del clasificador a utilizar, se constata que existen variados tipos de clasificadores y variados parámetros de cada uno de estos. En el caso de estudio incluido en el presente trabajo el mejor clasificador es el árbol de decisión, específicamente utilizando índice de Twoing; el que presentaba altos porcentajes de acierto tanto en la validación cruzada como en el experimento 1. Pero, al ser cada yacimiento distinto, puede que otro tipo de parámetro u otro tipo de clasificador se ajuste mejor para otro set de datos, por lo tanto, no se puede definir un clasificador fijo a utilizar en la metodología propuesta.

Referente a la innovación planteada mediante la propuesta de filtraje de componentes y viendo los resultados obtenidos se aprecia que se logra introducir una mejora en la clasificación ya que, para todos los experimentos realizados, y tanto para *logueo* como para *flagueo*/interpretación, los porcentajes de acierto de la clasificación son mayores al utilizarse las simulaciones con filtraje respecto a las simulaciones sin filtraje. Por lo que, se corrobora la hipótesis planteada y se valida la metodología propuesta al obtener los resultados esperados de los experimentos.

Notar que las mejoras de acierto de los experimentos 2 y 3, al pasar de la media de los casos a la clasificación del caso más probable, son considerablemente mayores en el caso con filtraje en comparación al caso sin filtraje; por lo tanto, el hecho de que la implementación del filtraje sea sobre simulaciones y no sobre solamente estimaciones ayuda a que el método obtenga mejores resultados al realizar análisis probabilísticos de la clasificación.

La metodología de filtraje se propone de modo de obtener simulaciones más suavizadas, por lo que se espera para el caso del *flagueo* e interpretación que al comparar aplicaciones de clasificadores a

simulaciones sin y con la utilización de filtraje se obtengan mejoras en el acierto ya que las simulaciones obtenidas se asimilan al comportamiento de la interpretación al ser ambas suavizadas.

Una de las principales fortalezas descubiertas en la aplicación de la metodología, es que no solo se obtienen mejoras al buscar clasificar variables suavizadas, sino que también se obtienen mejoras en el porcentaje de acierto en el caso del *logueo* geológico; dichas mejores, aunque menores que en el caso del *flagueo*/interpretación, de igual forma son significativas. Esta mejora en parte puede ser debido a que en el proceso de *logueo* de la base de datos utilizada solo se registran categorías que como mínimo tengan una presencia en el sondaje aproximadamente 1 [m], por lo que, de igual forma existe un grado de suavizamiento en la toma de información.

Por el contrario, una debilidad de la metodología propuesta de filtraje es que se altera la interpretación de la continuidad y variabilidad espacial del yacimiento al modificar el análisis de corregionalización, por lo tanto, si bien los resultados de las simulaciones sirven para obtener mejores clasificaciones, estos no pueden ser utilizados directamente para una posible evaluación económica del yacimiento, estudios de planificación u algún otro análisis basados en las leyes, ya que éstas no son una representación completa del yacimiento.

Capítulo 6: Conclusiones y recomendaciones

6.1 Conclusiones

Respecto a los objetivos del presente trabajo se puede concluir lo siguiente.

Los datos del caso de estudio presentado en el capítulo 4 muestran una buena relación entre las variables cuantitativas y cada una de las categorías, ya que, mediante la aplicación de un clasificador y las variables cuantitativas como input, se logra definir la variable litología a utilizar con aciertos en torno al 90% en la validación cruzada; llegando incluso a un 99.82% en el experimento 1.

Se logra implementar la simulación con filtraje, cuyos resultados se evalúan visualmente mediante el análisis de diferentes cotas del modelo de bloques, evidenciando la teoría detrás de la propuesta. En particular, se puede notar que se reduce la variabilidad en el corto alcance al observarse menor cantidad de zonas de alta ley, y las zonas existentes abarcando menor extensión; por lo tanto, se obtienes simulaciones suavizadas como se busca.

Existe suficiente evidencia numérica en los tres experimentos realizados para comprobar la hipótesis planteada de que, al calcular simulaciones suavizadas, se pueden obtener clasificaciones con mayor cantidad de acierto referente a la interpretación geológica o *flagueo* de alguna variable, ya que esta interpretación es de igual forma suavizada. Numéricamente la mejora porcentual de utilizar simulaciones con filtraje en relación con no utilizarla es de 4.78% en el experimento 1, 3.71% en el experimento 2 y 3.10% en el experimento 3.

De igual forma, aunque en los *logueos* geológicos el promedio del acierto de los clasificadores es considerablemente menor que para el caso anterior, la metodología propuesta también presenta mejoras relativas en el acierto al intentar clasificar variables categóricas obtenidas mediante *logueo*, lo que permite concluir que la metodología propuesta no solo mejora el acierto de los clasificadores aplicada a la interpretación geológica como fue concebida, sino que también mejora la clasificación en este caso del *logueo* geológico de la variable litología. Numéricamente la mejora porcentual de utilizar simulaciones con filtraje en relación con no utilizarla es de 3.50% en el experimento 1 y 4.00% en el experimento 2.

6.2 Recomendaciones

Una de las limitaciones involucradas en el caso de estudio es la dificultad de manejar la totalidad de datos del yacimiento para el experimento 3, por lo que se decide trabajar con ciertas secciones transversales a diferentes cotas equidistantes. Al estudiar el acierto y mejoras de acierto al aplicar la metodología de filtraje, no se está en presencia de evidencia clara conducente a concluir la existencia de una tendencia de aumento o disminución respecto a profundidad o cantidad de datos involucrados, por lo que, se sugiere una futura investigación relacionada con determinar qué factores o parámetros tienen influencia en la mejoría de la clasificación al aplicar el filtraje propuesto.

Por otra parte, en el presente trabajo se busca determinar si la aplicación de la propuesta de simulaciones con filtraje ayuda a obtener mejores interpretaciones mediante la aplicación de

clasificadores en comparación a la aplicación de estos sobre leyes simuladas normalmente sin filtraje, lo que se pudo comprobar; pero, puede ser de interés en una futura investigación evaluar distintos parámetros que puedan ayudar a obtener aún mayores aciertos, por ejemplo se sugiere para futuras investigaciones estudiar distintos parámetros como: vecindades de búsqueda utilizadas, cantidad de funciones básicas utilizadas para los variogramas modelados o la cantidad de alcances filtrados.

Capítulo 7: Bibliografía

[1] Adeli, A., Emery, X., 2017. A geostatistical approach to measure the consistency between geological logs and quantitative covariates. Ore Geology Reviews, 82, 160-169.

[2] Adeli, A., Emery, X., Dowd, P., 2018. Geological modelling and validation of geological interpretations via simulation and classification of quantitative covariates. Minerals, 8(1), 7

[3] Agterberg, F., 1990. Automated stratigraphic correlation. Elsevier, Amsterdam, 423 pp.

[4] Arnaud, M., Emery, X., De Fouquet, C., Brouwers, M., Fortier, M. (2001). L'analyse krigeante pour le classement d'observations spatiales et multivariées. Revue de Statistique Appliquée, 49(2), 45-67.

[5] Bourgine, B., Lasseur, E., Leynet, A., Badinier, G., Ortega, C., Issautier, B., Bouchet, V., 2014. Building a geological reference platform using sequence stratigraphy combined with geostatistical tools. In *Geostatistics Valencia 2016* (pp. 865-877). Springer, Cham.

[6] Bourgine, B., Prunier-Leparmentier, A.M., Lembezat, C., Thierry, P, Luquet, C., Robelin, C., 2008. Tools and methods constructing 3D geological models in the urban environment. The Paris case. In *Proceedings of the Eighth International Geostatistics Congress* (Vol. 2, pp. 951-960). Gecamin Ltda, Santiago.

[7] Cáceres, A., Emery, X., 2013. Geostatistical validation of geological logging. In: Ambrus, J., Beniscelli, J., Brunner, F., Cabello, J., Ibarra, F. (eds.) Proceedings of the Third International Seminar on Geology for the Mining Industry. Gecamin Ltda, Santiago, pp. 73-80.

[8] Chilès J.P., Delfiner P., 2012. Geostatistics: Modeling Spatial Uncertainty, Wiley, New York, 699 p.

[9] Clarke, S., 2004. Confidence in geological interpretation. A methodology for evaluating uncertainty in common two and three-dimensional representations of subsurface geology. British Geological Survey Internal Report, IR/01/164.

[10] Cortes, C., Vapnik, V., 1995. Support-vector network. Machine Learning, 20, 273–297.

[11] Emery, X., 2013. Geoestadística. Universidad de Chile, Santiago, Chile.

[12] Fustos, R., 2017. Descubrimiento de unidades geometalúrgicas por medio de análisis de conglomerados geoestadístico. Tesis de Doctorado en Ingeniería de Minas, Universidad de Chile. Disponible en http://repositorio.uchile.cl/handle/2250/147373.

[13] Goovaerts, P., 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. European Journal of Soil Science. 597 – 619.

[14] Guillen, A., Calcagno, P., Courrioux, G., Joly, A., Ledru, P., 2008. Geological modelling from field data and geological knowledge: Part II Modelling validation using gravity and magnetic data inversion. Physics of the Earth and Planetary Interiors, Elsevier, 2008, 171 (1-4), 158 pp.

[15] Hassibi, M., Ershaghi, I., Aminzadeh, F., 2003. Chapter 15 High resolution reservoir heterogeneity characterization using recognition technology. Developments in Petroleum Science. 51. 289-307.

[16] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer-Verlag, 763 pp.

[17] Hoyle, I., 1986. Computer techniques for the zoning and correlation of well-logs. Geophysical prospecting 34(5). 648-664.

[18] Kuchin, Y., Grundspenkis, J., 2017. Machine Learning Methods for Identifying Composition of Uranium Deposits in Kazakhstan. Applied Computer Systems. 22. 10.1515/acss-2017-0014.
[19] Larocque, G., Dutilleul, P., Pelletier, B., Fyles, J. W., 2006. Conditional Gaussian co-simulation of regionalized components of soil variation. Geoderma, 134(1-2), 1-16.
[20] Lelièvre, P., Oldenburg, D., Williams, N., 2009. Integrating geologic and geophysical data through advanced constrained inversions. ASEG Extended Abstracts, 2009(1), 1-6.
[21] Leung, K., 2007. Naive Bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering.
[22] Luthi, S., Bryant, U., 1997. Well-log correlation using a back-propagation neural network. Mathematical Geology. 29. 413-425.
[23] Maleki, M., Emery, X., Mery, N., 2017. Indicator Variograms as an Aid for Geological Interpretation and Modeling of Ore Deposits. Minerals, 7(12), 241.
[24] Manchuk, J., Deutsch, C., 2012. Applications of data coherency for data analysis and geological zonation. Geostatistics Oslo 2012, Springer, pp. 173-184.
[25] Marjoribanks, R., 2010. Geological Methods in Mineral Exploration and Mining. Springer, Berlin, 238 pp.
[26] MathWorks, Inc., 2018. fitcecoc [en línea] la.mathworks.com/help/stats/fitcecoc.html [consulta: 15 diciembre 2018].
[27] MathWorks, Inc., 2018. Predict labels using k-nearest neighbor classification model [en línea] la.mathworks.com/help/stats/classificationknn.predict.html [consulta: 15 diciembre 2018].
[28] MathWorks, Inc., 2018. Support Vector Machines for Binary Classification [en línea] la.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html [consulta: 15 diciembre 2018].
[29] Meyer, D., Wien, F., 2001. Support vector machines. R News, 1(3), 23-26.
[30] Miller, F., Vandome, A., John, M., 2010. Geometallurgy. VDM Publishing. 84 pp.
[31] Mitchell, T., 1997. Machine Learning. McGraw Hill International Editions Computer Science Series.
[32] Rish, I., 2001. An empirical study of the naïve Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46). New York: IBM.
[33] Rosales, D., 2014. Implementación de metodología para determinar dominios geometalúrgicos de estimación. Tesis de Magíster en Minería, Universidad de Chile. Disponible en http://repositorio.uchile.cl/handle/2250/117363.
[34] Rossi, M., Deutsch, C., 2014. Mineral Resource Estimation. Springer, Heidelberg, 332 pp.
[35] Soleimani, M., Shokri, B., Rafiei, M., 2016. Integrated petrophysical modeling for a strongly heterogeneous and fractured reservoir, Sarvak Formation, SW Iran. Natural Resources Research, 26(1), 75-88.
[36] Valentini, G., 2006. An introduction to Support Vector Machines. DSI- Dipartimento di Scienze dell' Informazione, Universita degli Studi di Milano.
[37] Zambon, M., Lawrence, R., Bunn, A., Powell, S., 2006. Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. Photogrammetric Engineering and Remote Sensing. 72. 25-30.

Capítulo 8: Anexo

8.1 Glosario de abreviaturas

- Alteraciones:
  - A = Argílico.
  - CL = Transición clorítica.
  - KB = Potásico (biotita).
  - KF = Potásico (feldespato).
  - NONE = Sin información.
  - P = Propilítico.
  - QS = Fílico.
  - SA = Nulo.
- Litología:
  - ABX1 = Brecha intrusiva 1 asociada a QFP1.
  - ABX2 = Brecha intrusiva 2 asociada a QFP2.
  - ABX3 = Brecha intrusiva 3 asociada a QFP3.
  - AR = Horizonte arcilloso.
  - BRXH = Brecha hidrotermal.
  - COV = Cobertura indiferenciada.
  - FP = Pórfido feldespático tardío.
  - GRAV = Gravas.
  - IND = Rocas volcánicas o sedimentos indiferenciados.
  - NONE = Sin información.
  - QFP1 = Pórfido cuarzo – feldespático 1.
  - QFP2 = Pórfido cuarzo – feldespático 2.
  - SBR = Brecha sedimentaria.
  - SED = Unidades sedimentarias.
- Mineralización:
  - EXGRAV = Gravas exóticas.
  - HYP1A = Hipógeno con calcopirita y pirita, con calcopirita $\geq 0.75\%$
  - HYP1B = Hipógeno con calcopirita y pirita.
  - HYP2 = Hipógeno solo con pirita.
  - LIX = Lixiviado.
  - NMF = Roca no mineralizada fresca.
  - NMW = Roca no mineralizada meteorizada.
  - NONE = Sin información.
  - OXC = Óxidos café.
  - OXS = Óxidos de cobre con sulfuros.
  - OXV = Óxidos de cobre.
  - STCP = Sulfuros transicionales.
  - SUCC = Sulfuros secundarios dominados por calcosina.
  - SUCV = Sulfuros secundarios dominados por covelina.

8.2 Visualización de sondajes.

8.2.1 Logueos

- Alteración:

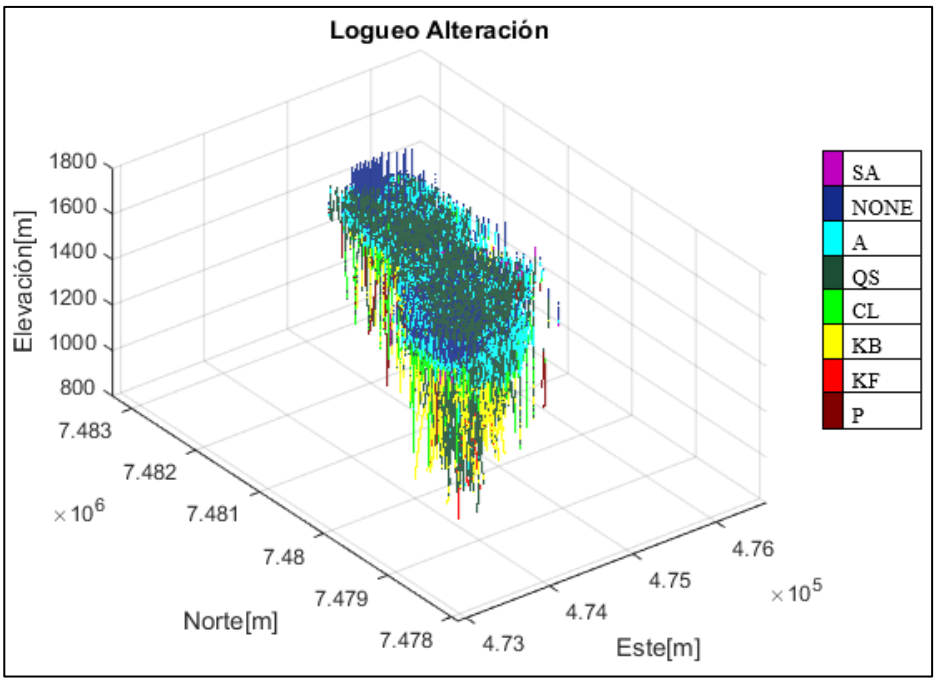

Ilustración 27 Logueo de Alteración

- Mineralización:

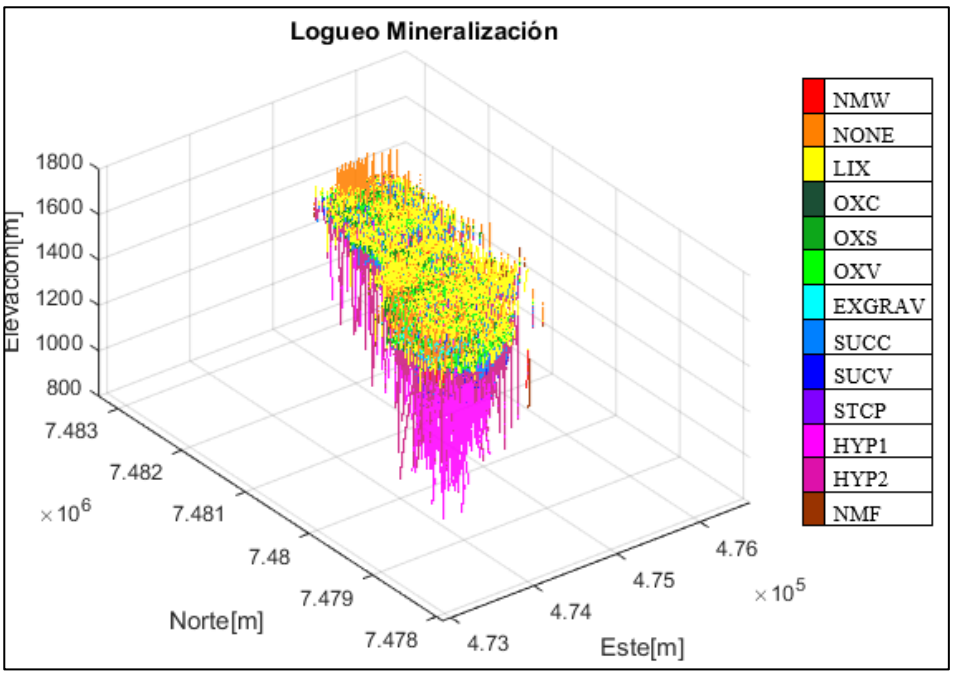

Ilustración 28 Logueo de Mineralización

8.2.2 Flagueos

- Alteración:

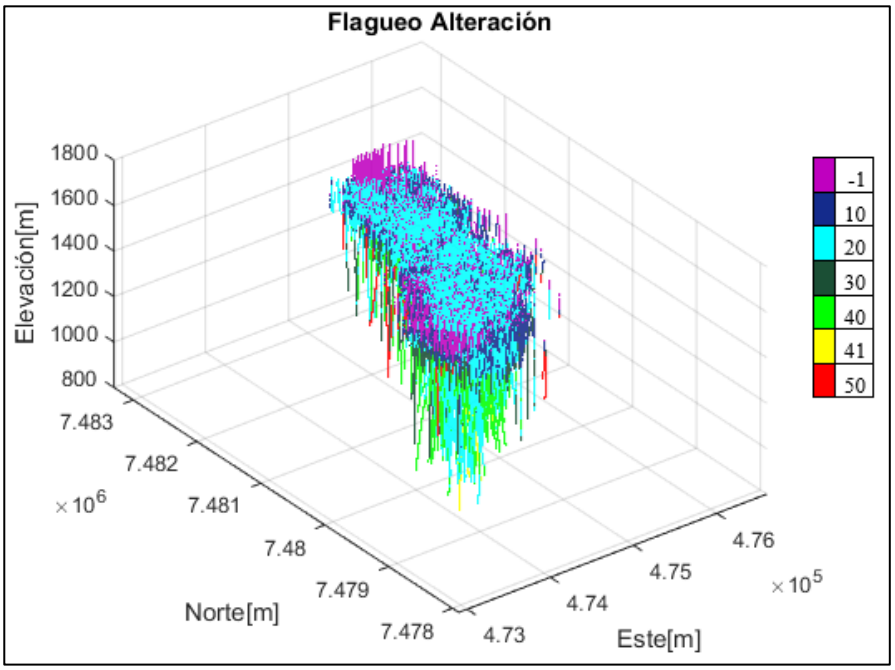

Ilustración 29 Flagueo de Alteración

- Mineralización:

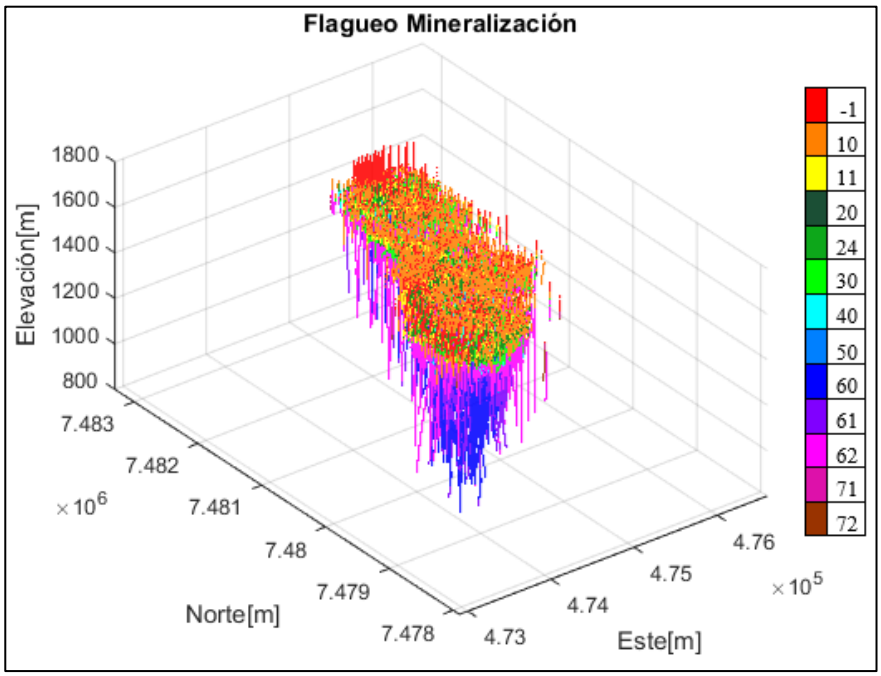

Ilustración 30 Flagueo de Mineralización

8.3 Visualización de interpretación

- Interpretación de Alteración:

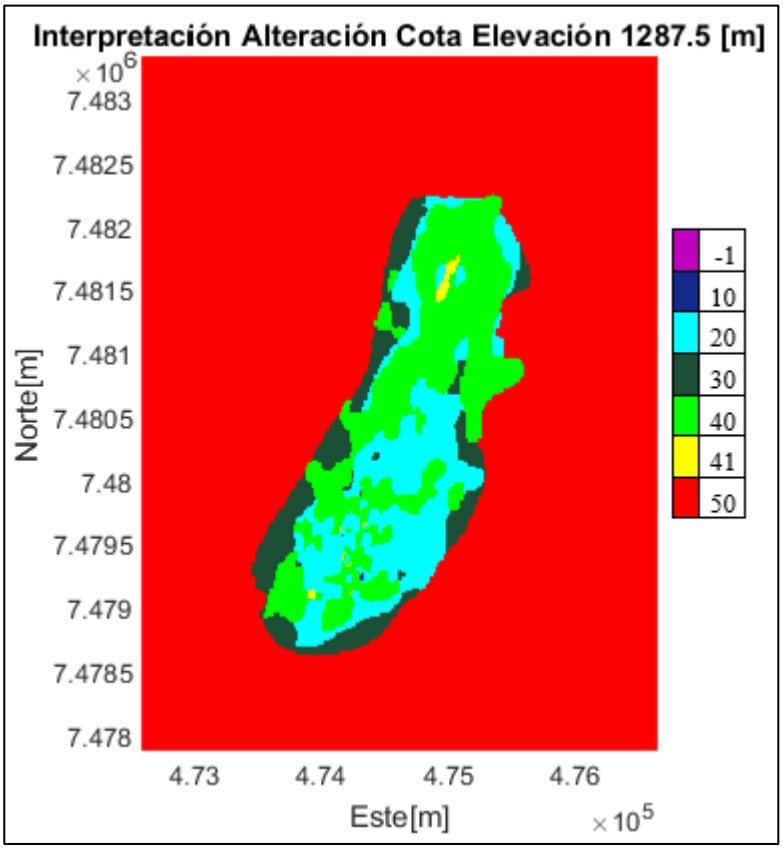

Ilustración 31 Interpretación de Alteración a Elevación 1287.5

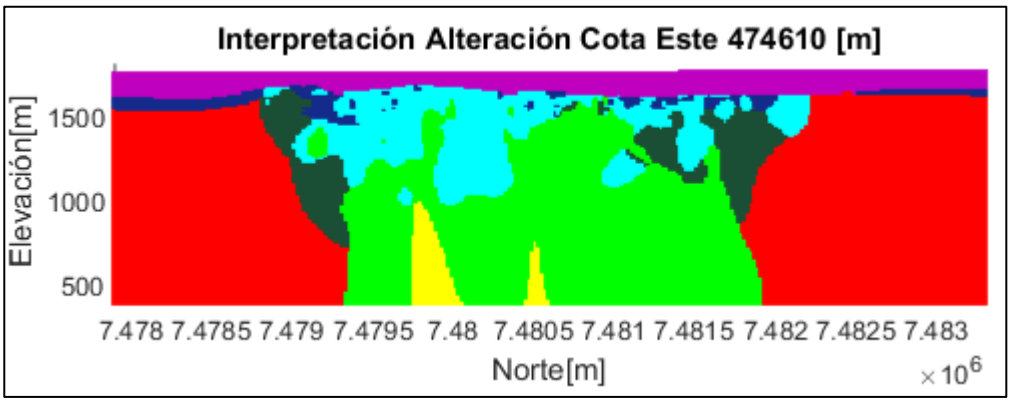

Ilustración 32 Interpretación de Alteración a Este 474610

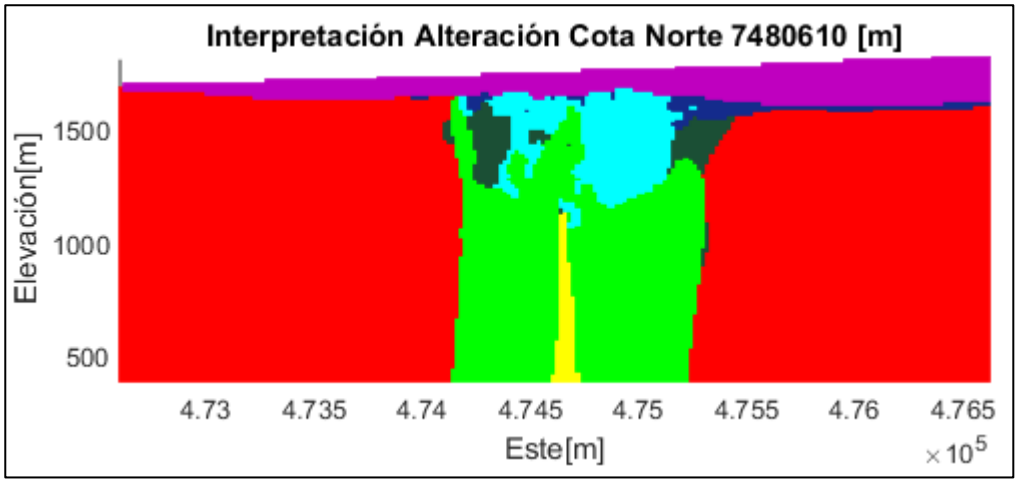

Ilustración 33 Interpretación de Alteración a Norte 7480610

- Interpretación de Mineralización:

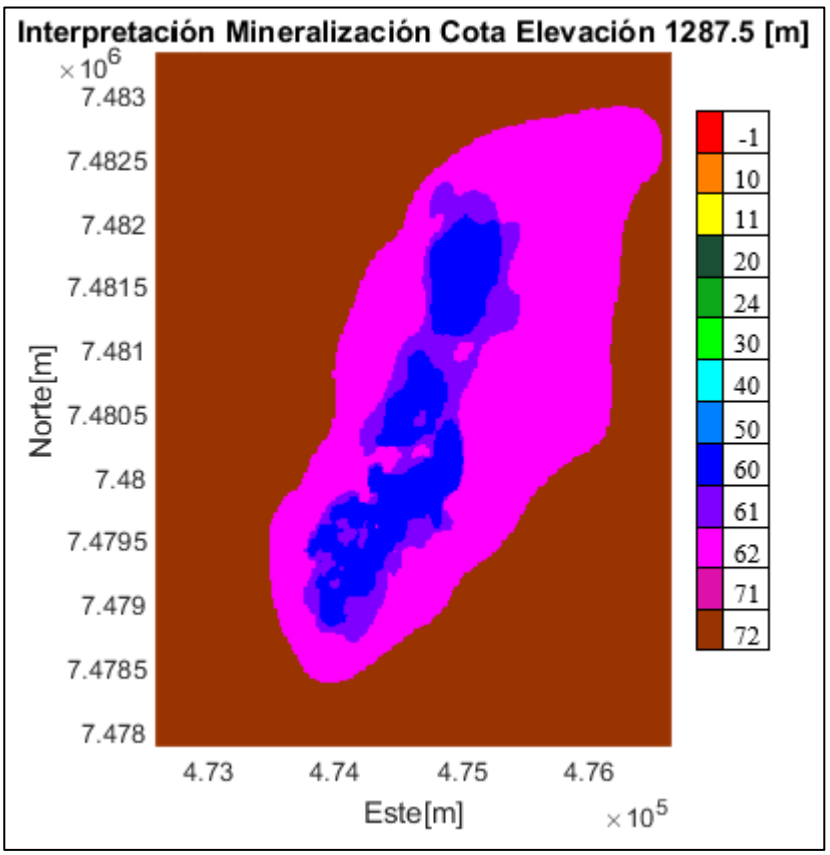

Ilustración 34 Interpretación de Mineralización a Elevación 1287.5

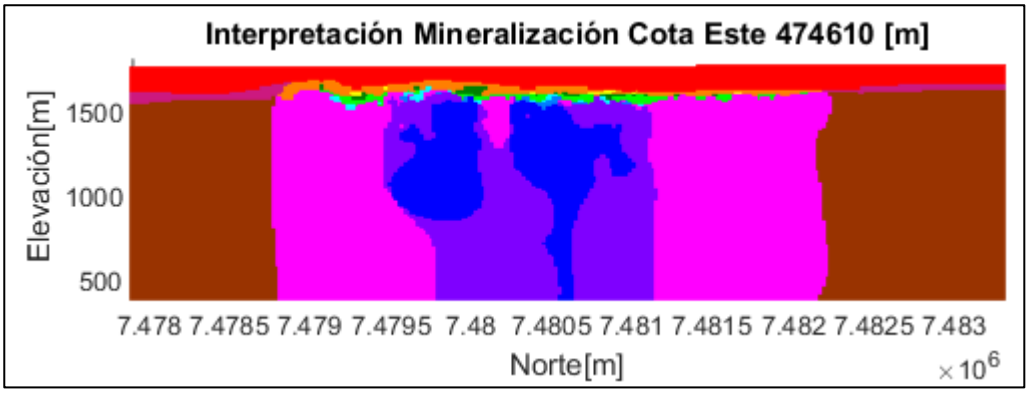

Ilustración 35 Interpretación de Mineralización a Este 474610

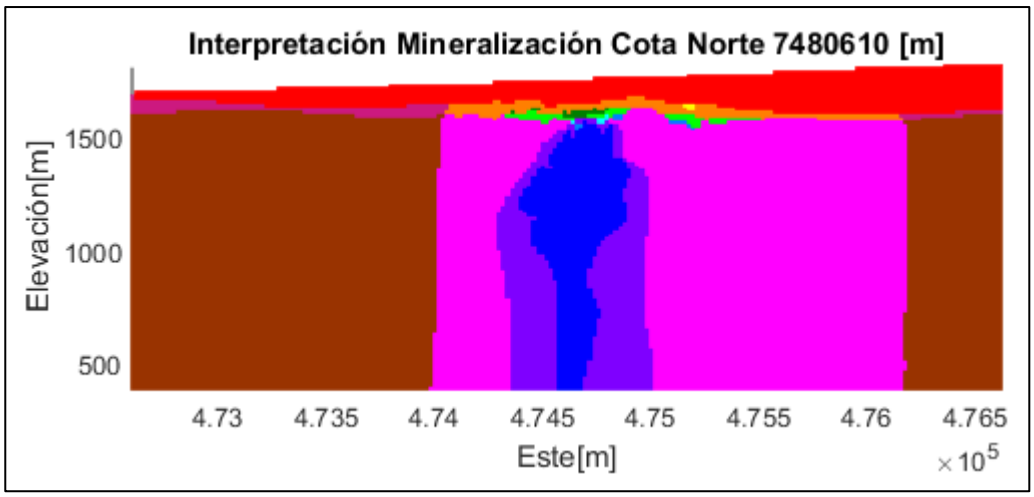

Ilustración 36 Interpretación de Mineralización a Norte 7480610

8.4 Envolvente

- Alteración:

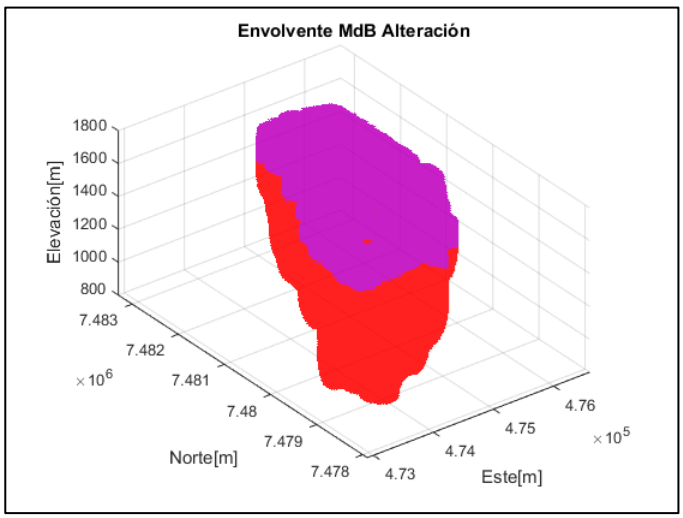

Ilustración 37 Envolvente Alteración

- Litología:

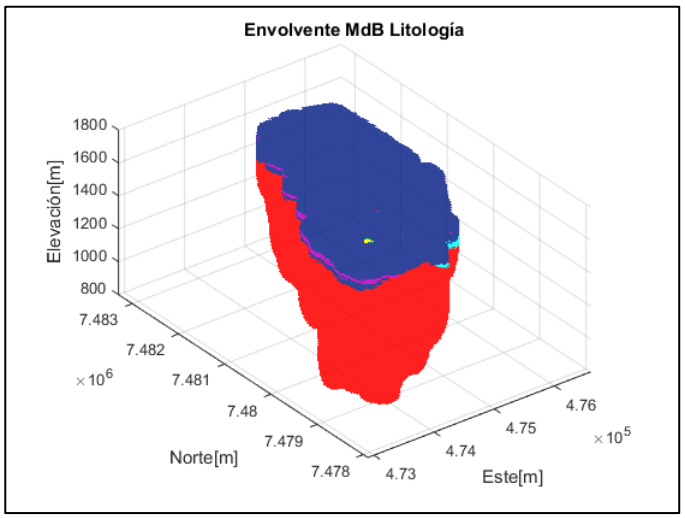

Ilustración 38 Envolvente Litología

- Mineralización:

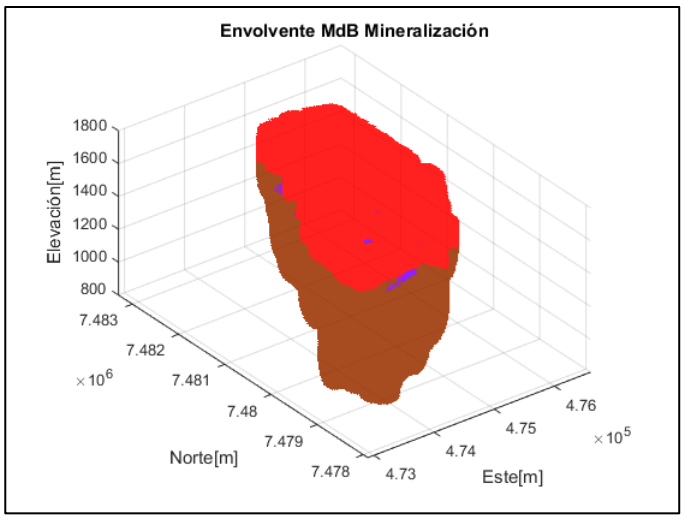

Ilustración 39 Envolvente Mineralización

8.5 Anexos de variografía

8.5.1 Variogramas

- Cu-Ag-Au

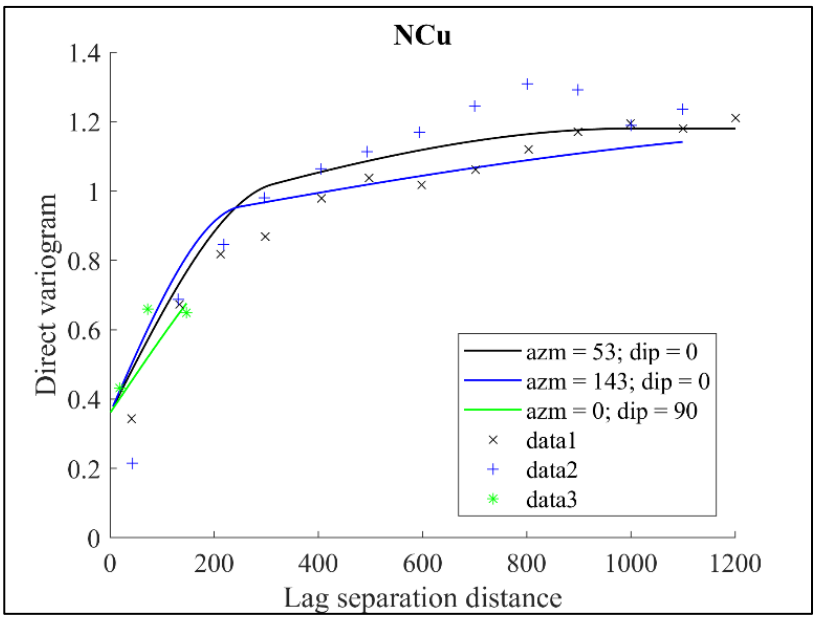

Ilustración 40 Variograma Cu

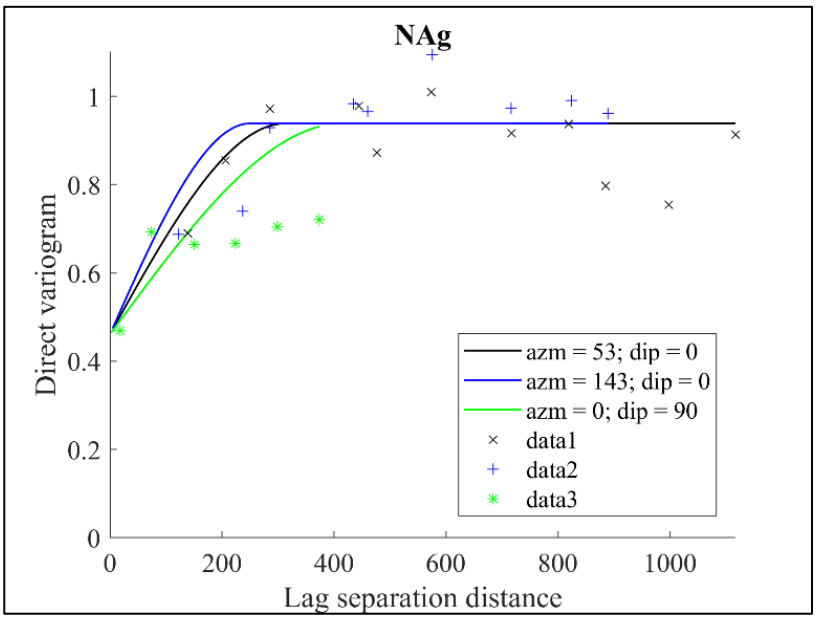

Ilustración 41 Variograma Ag

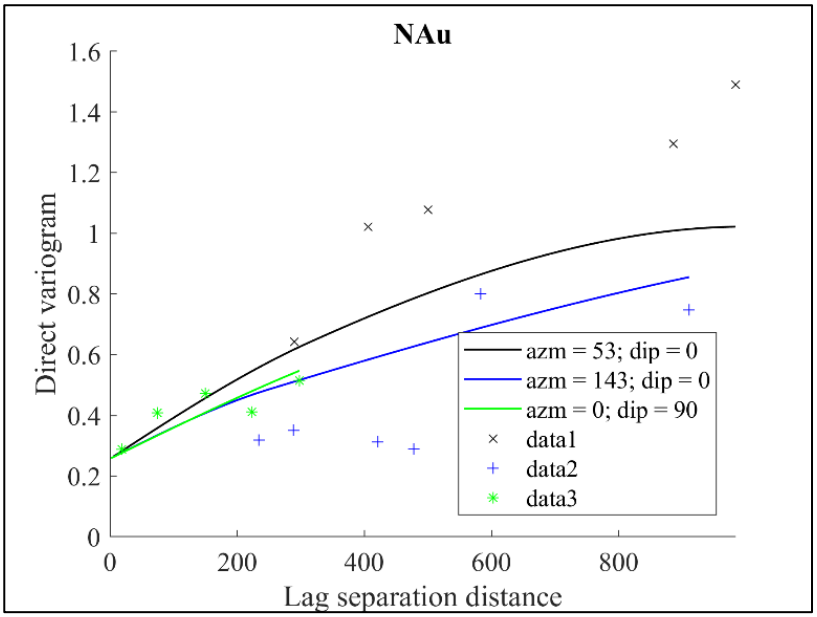

Ilustración 42 Variograma Au

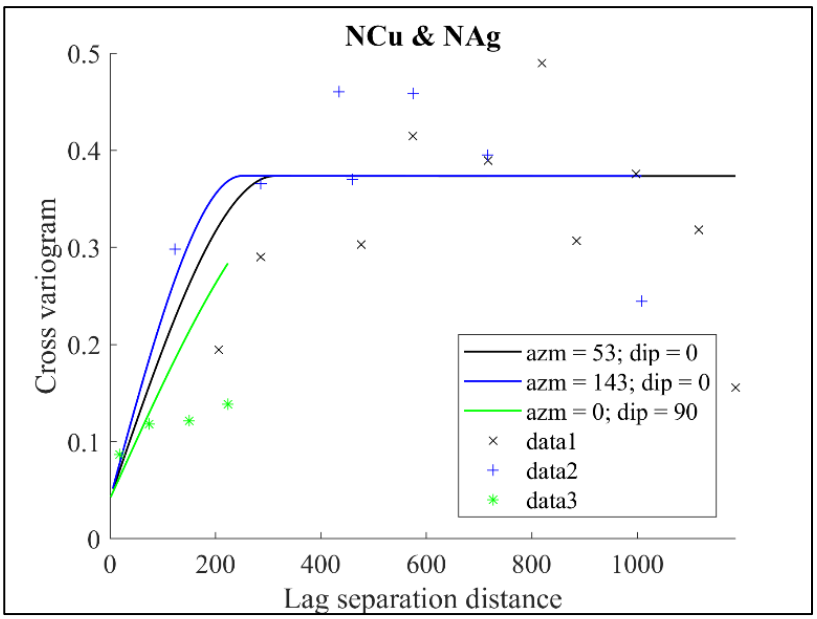

Ilustración 43 Variograma cruzado Cu-Ag

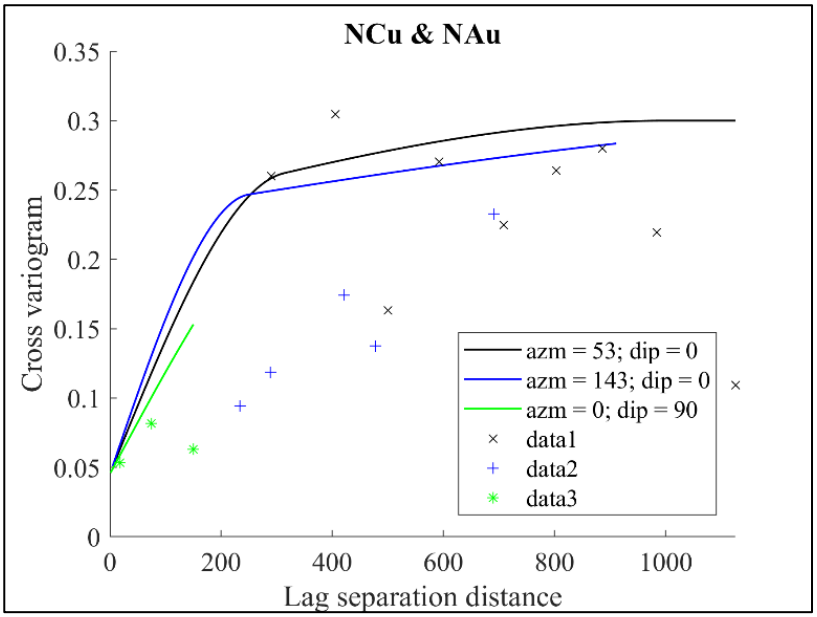

Ilustración 44 Variograma cruzado Cu-Au

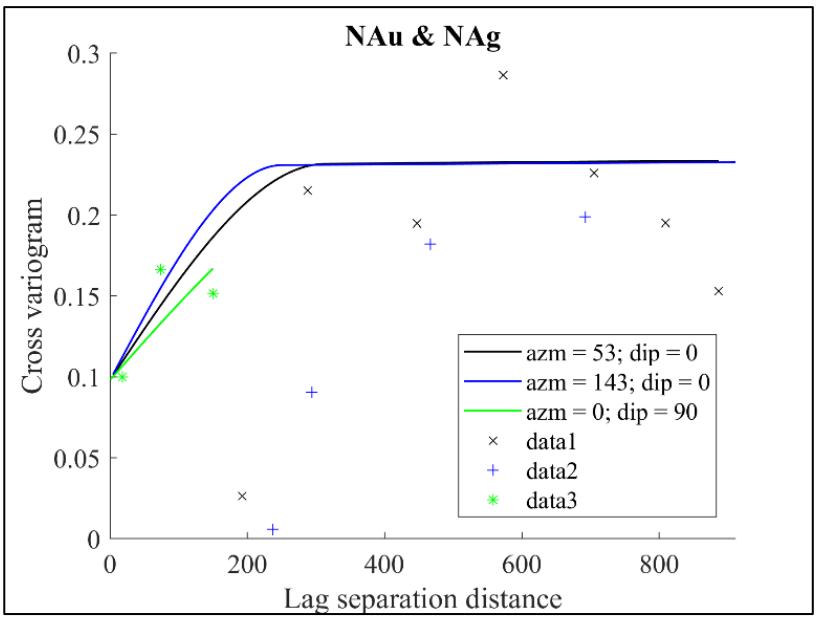

Ilustración 45 Variograma cruzado Au-Ag

- Molibdeno

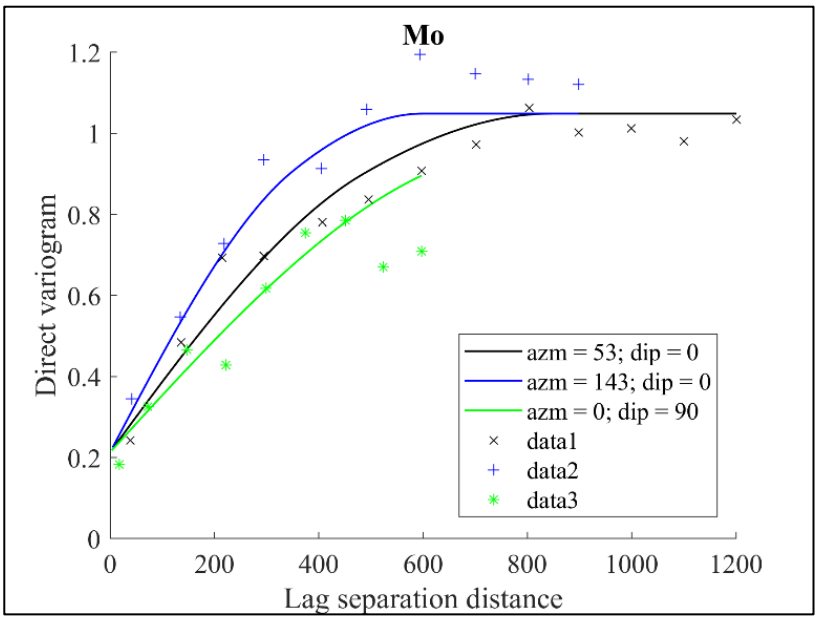

Ilustración 46 Variograma Mo

- Arsénico

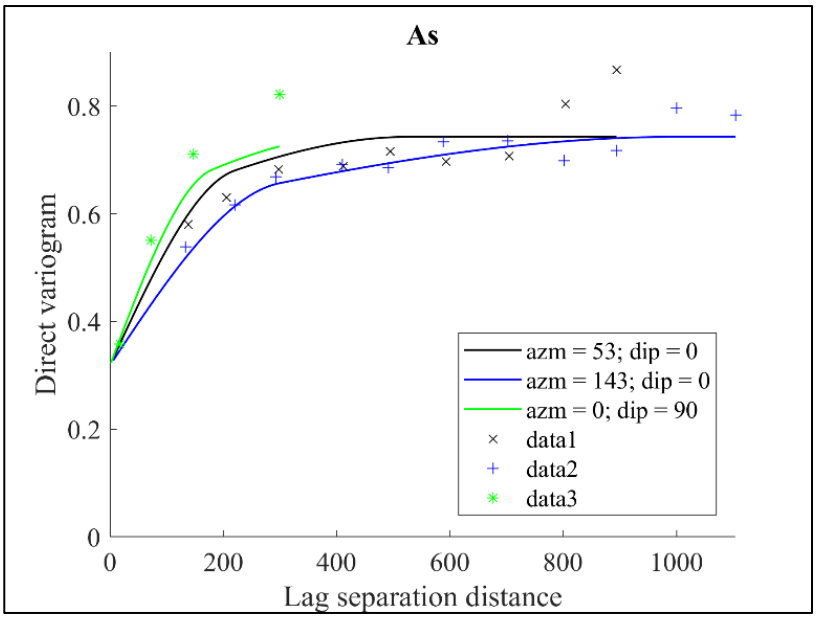

Ilustración 47 Variograma As

8.5.2 Gráficos de validación

- Cobre

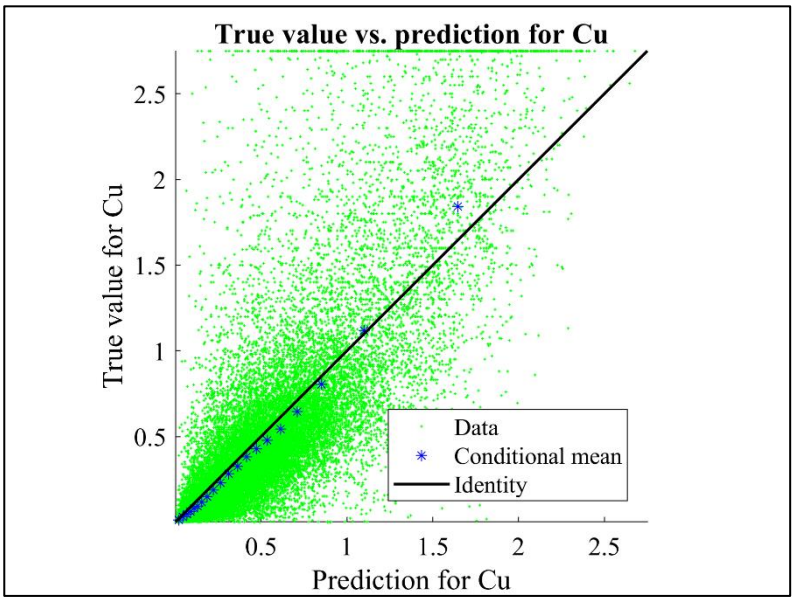

Ilustración 48 Validación parámetros Cu

- Plata

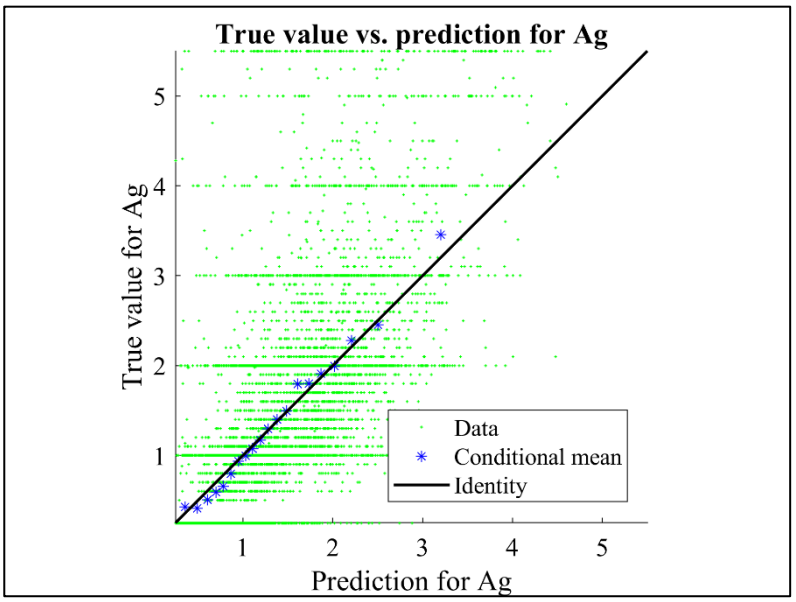

Ilustración 49 Validación parámetros Ag

- Oro

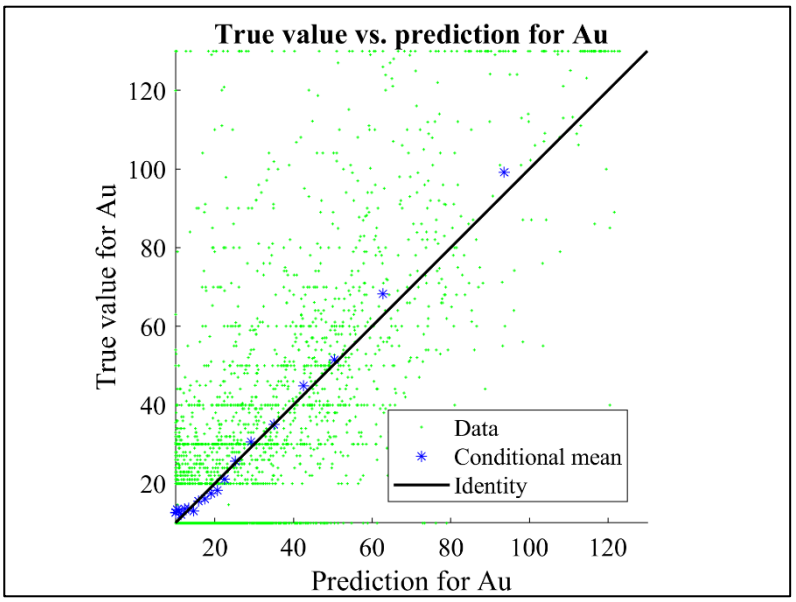

Ilustración 50 Validación parámetros Au

- Molibdeno

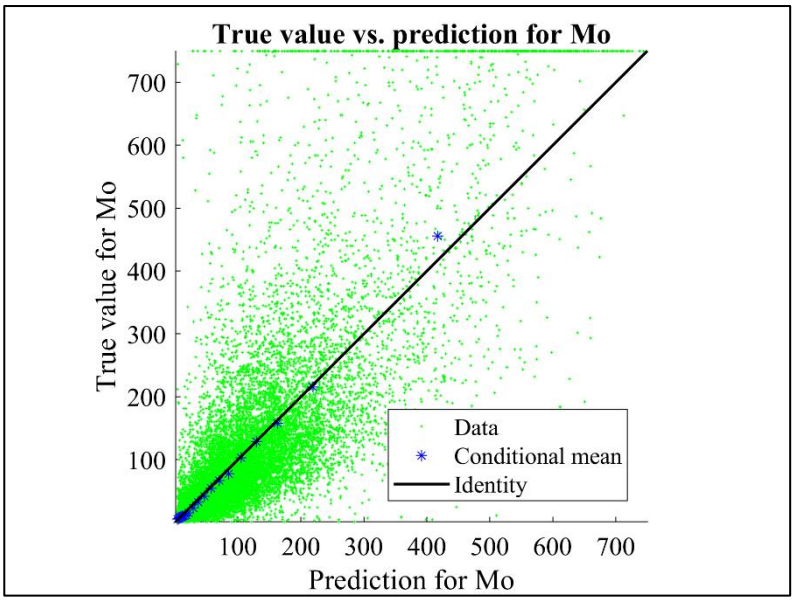

Ilustración 51 Validación parámetros Mo

- Arsénico

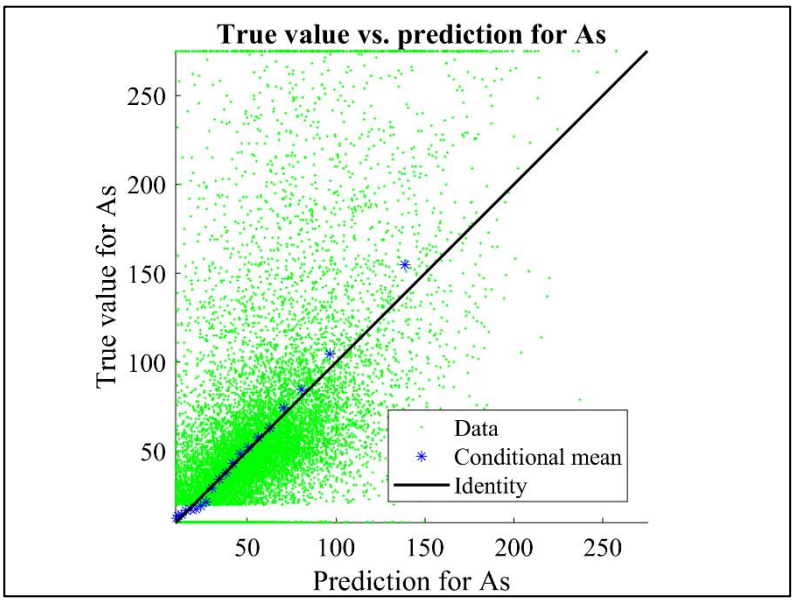

Ilustración 52 Validación parámetros As

8.5.3 Cortes transversales de simulaciones

8.5.3.1 Cota 987.5

- Cobre

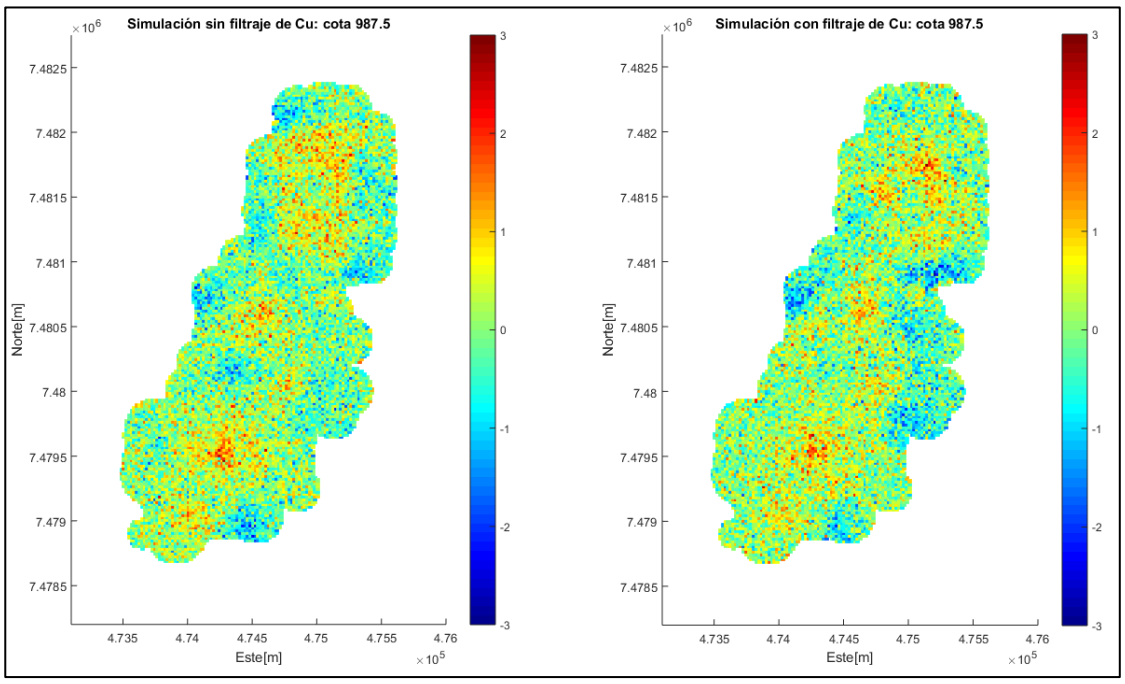

Ilustración 53 Cota 987.5 Cu

- Plata

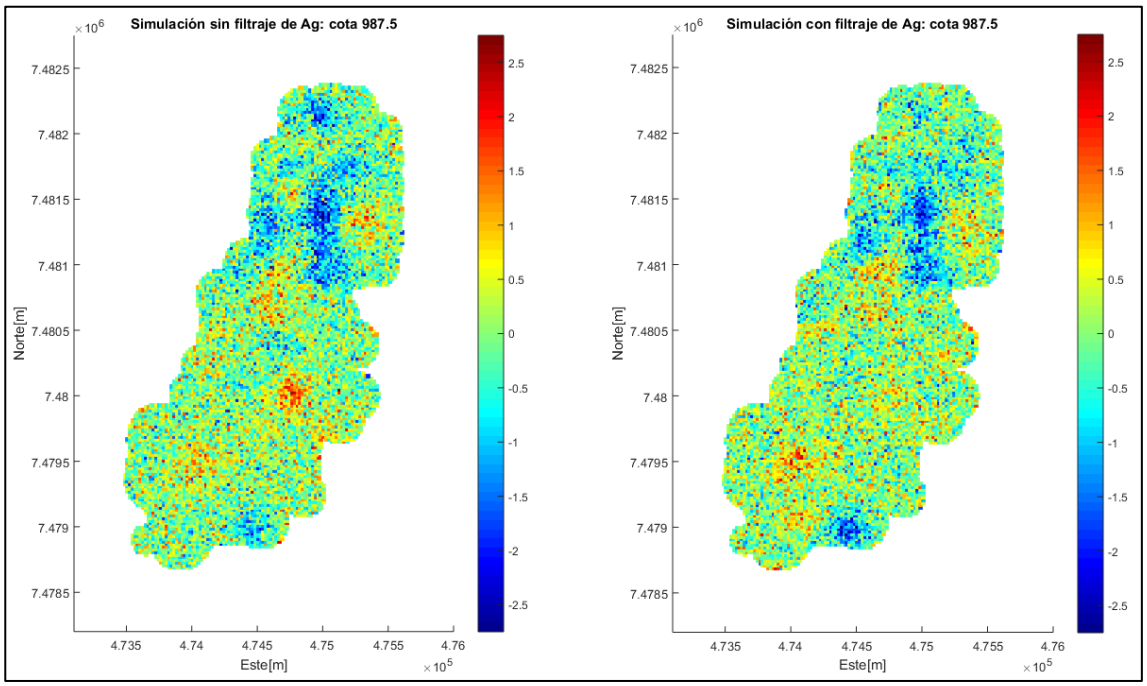

Ilustración 54 Cota 987.5 Ag

- Oro

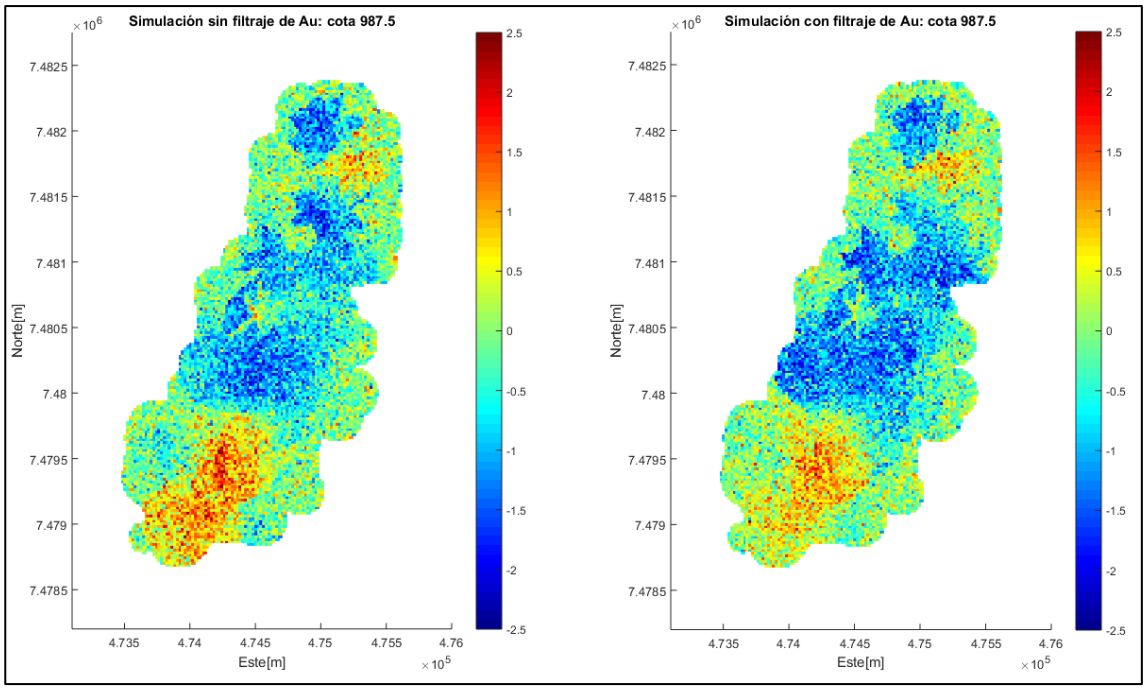

Ilustración 55 Cota 987.5 Au

- Molibdeno

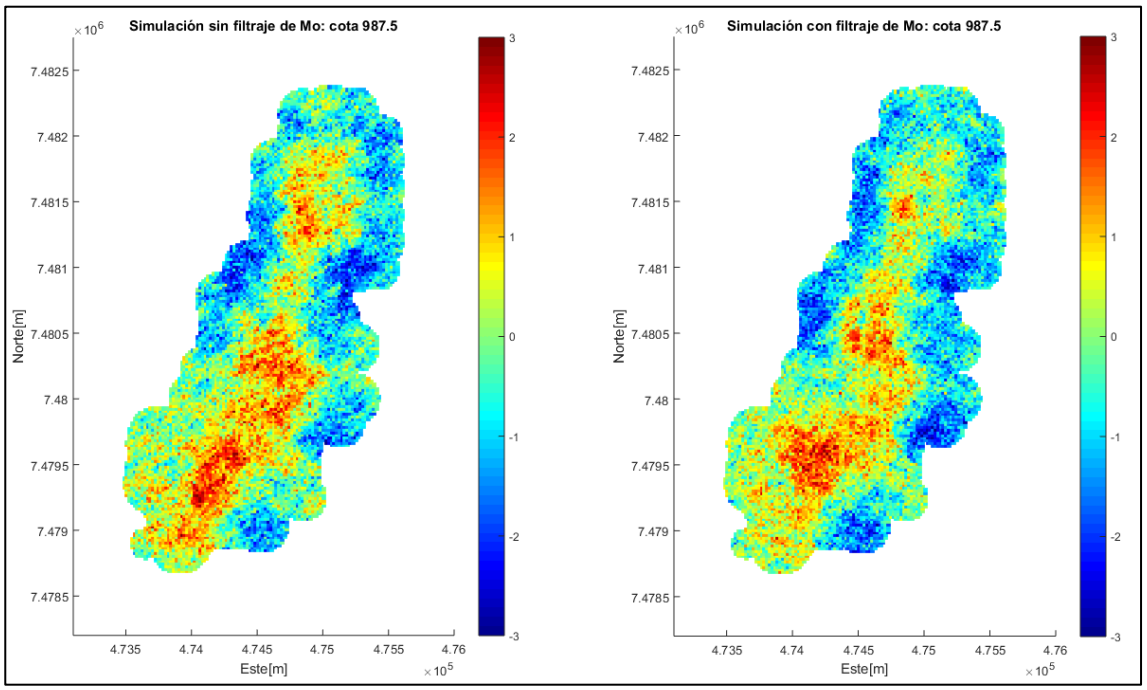

Ilustración 56 Cota 987.5 Mo

- Arsénico

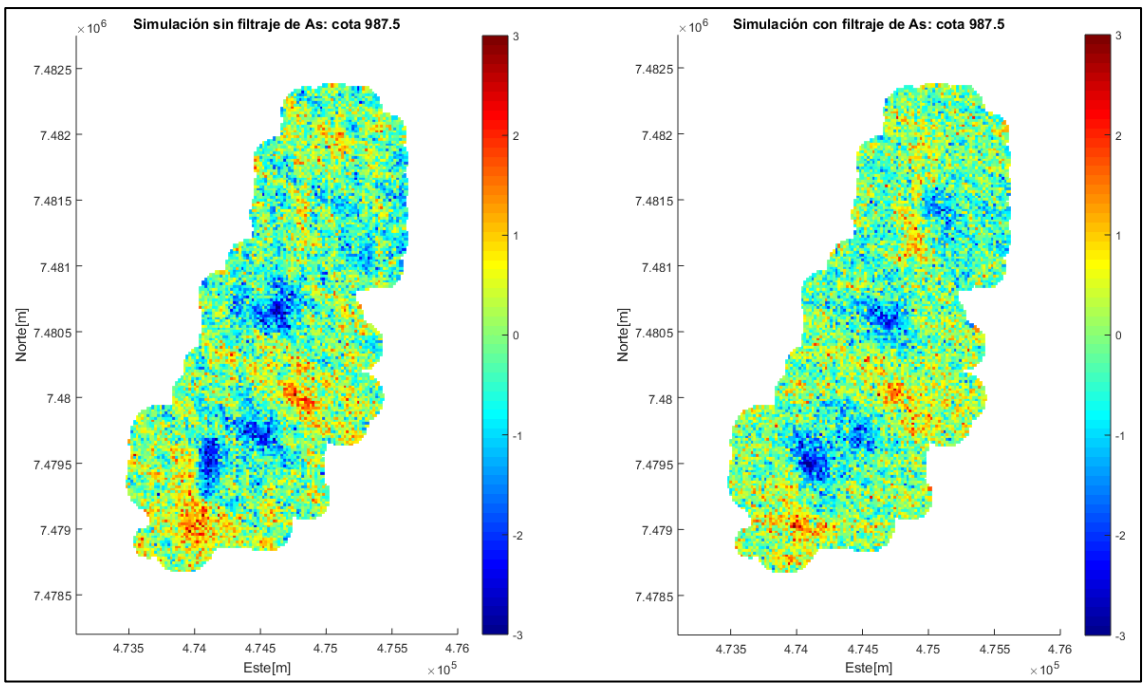

Ilustración 57 Cota 987.5 As

8.5.3.2 Cota 1287.5

- Cobre

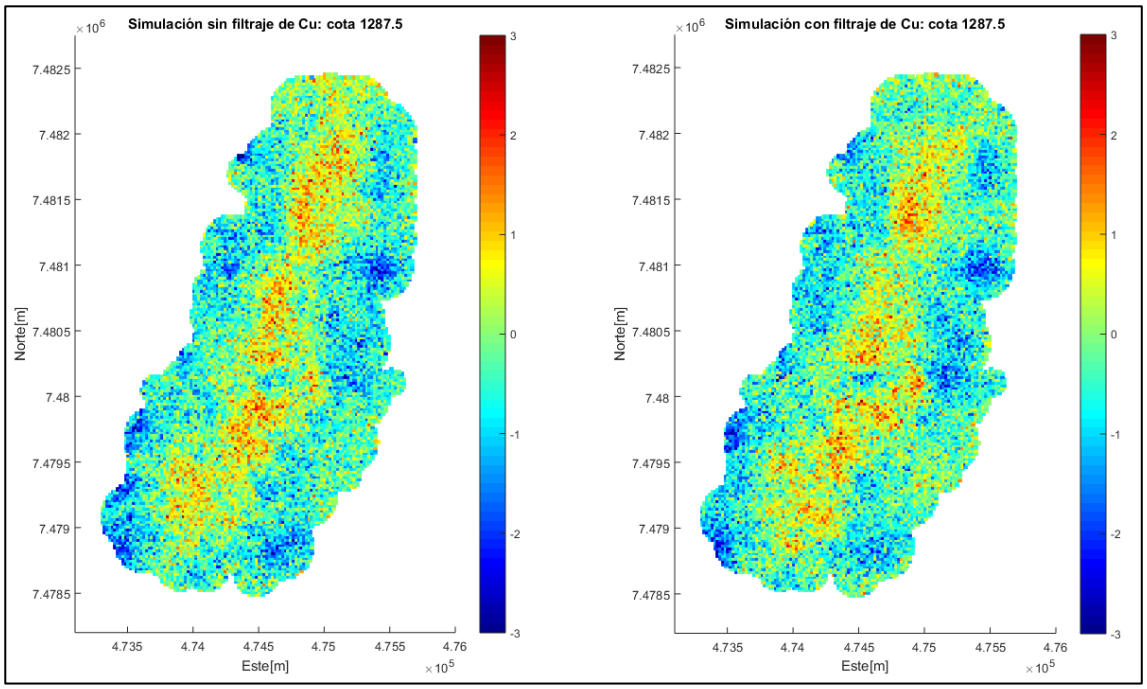

Ilustración 58 Cota 1287.5 Cu

- Plata

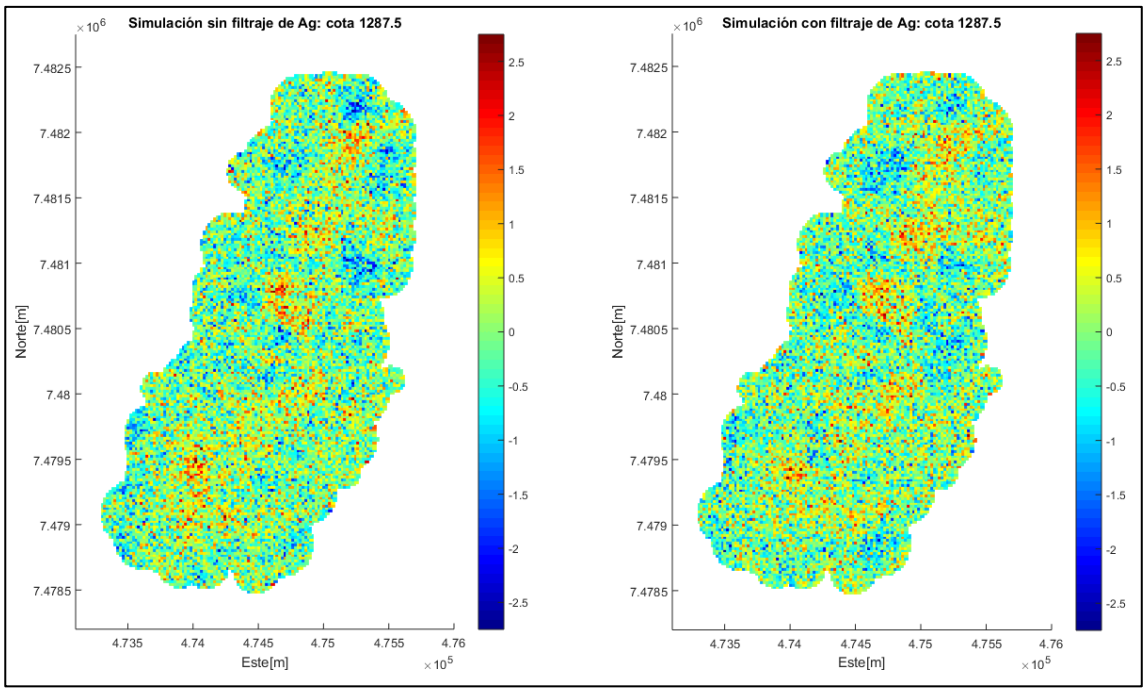

Ilustración 59 Cota 1287.5 Ag

- Oro

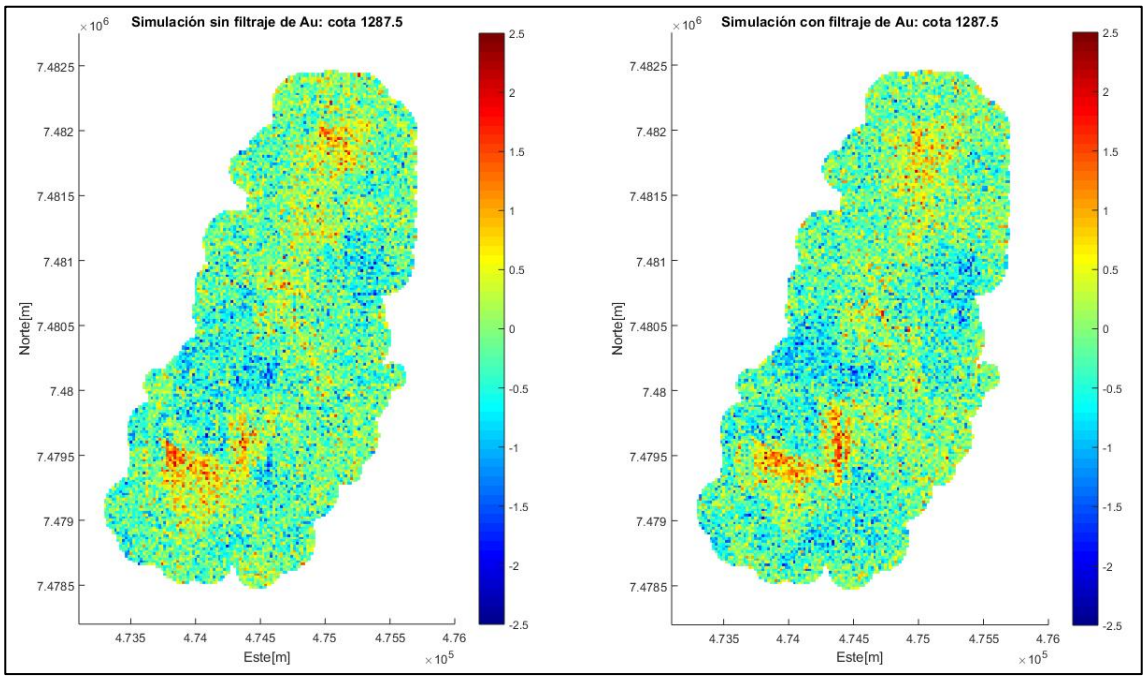

Ilustración 60 Cota 1287.5 Au

- Molibdeno

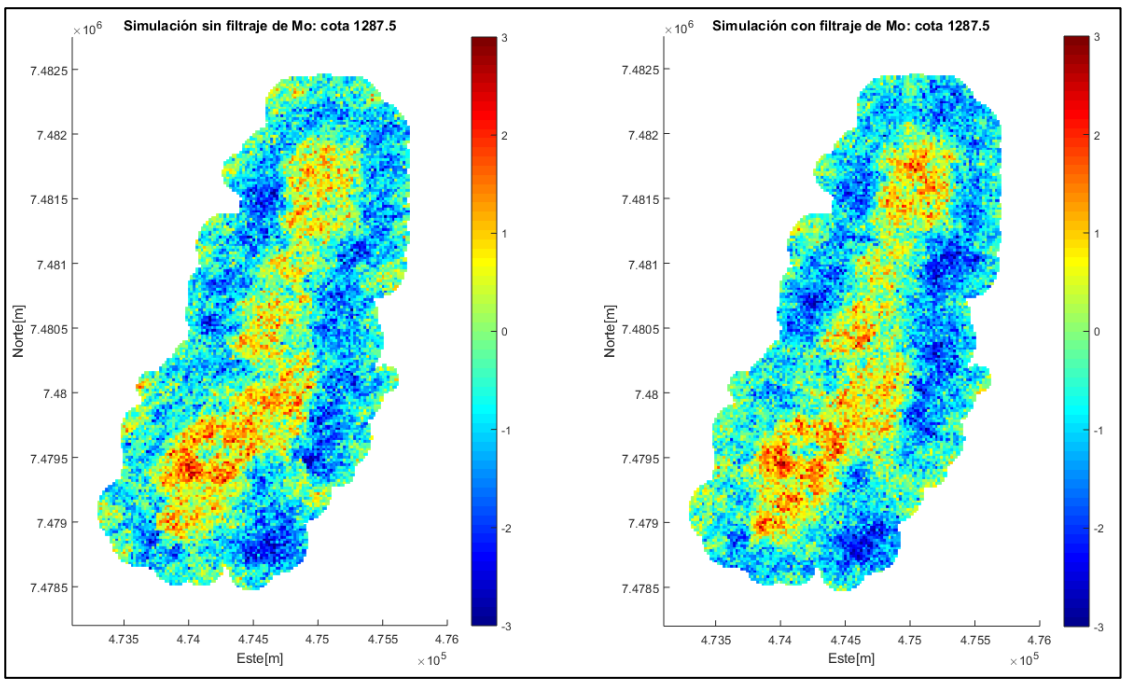

Ilustración 61 Cota 1287.5 Mo

- Arsénico

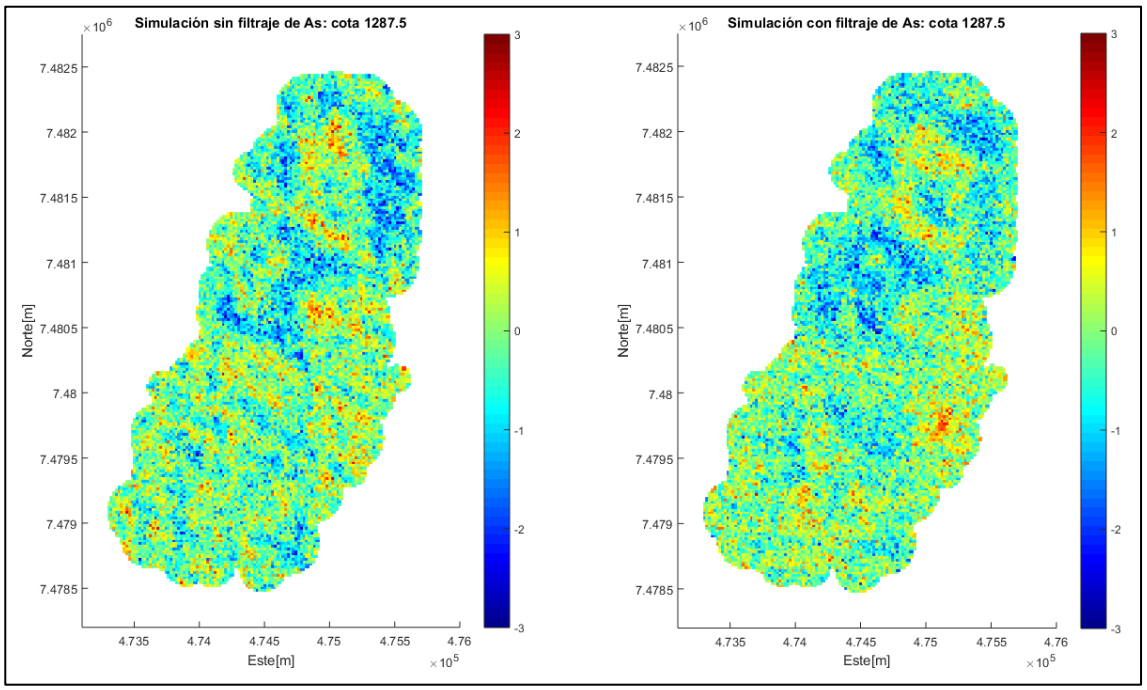

Ilustración 62 Cota 1287.5 As

8.5.4 Visualización de clasificaciones

8.5.4.1 Cota 987.5

- Clasificación más probable

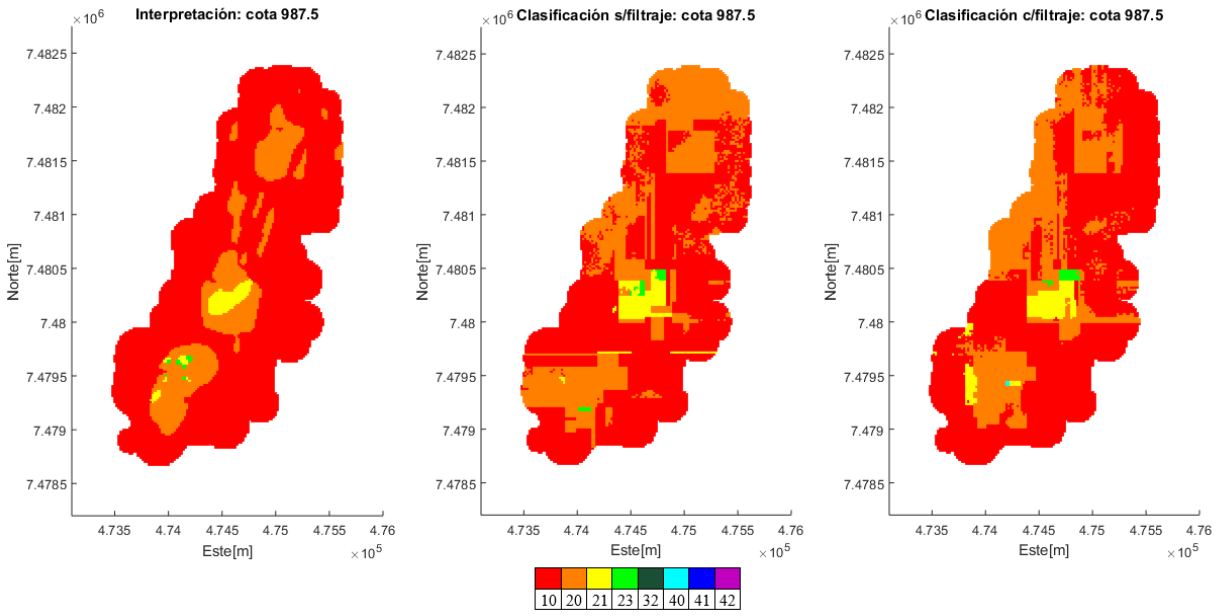

Ilustración 63 Clasificación más probable cota 987.5 [m]

- Probabilidad de clasificación

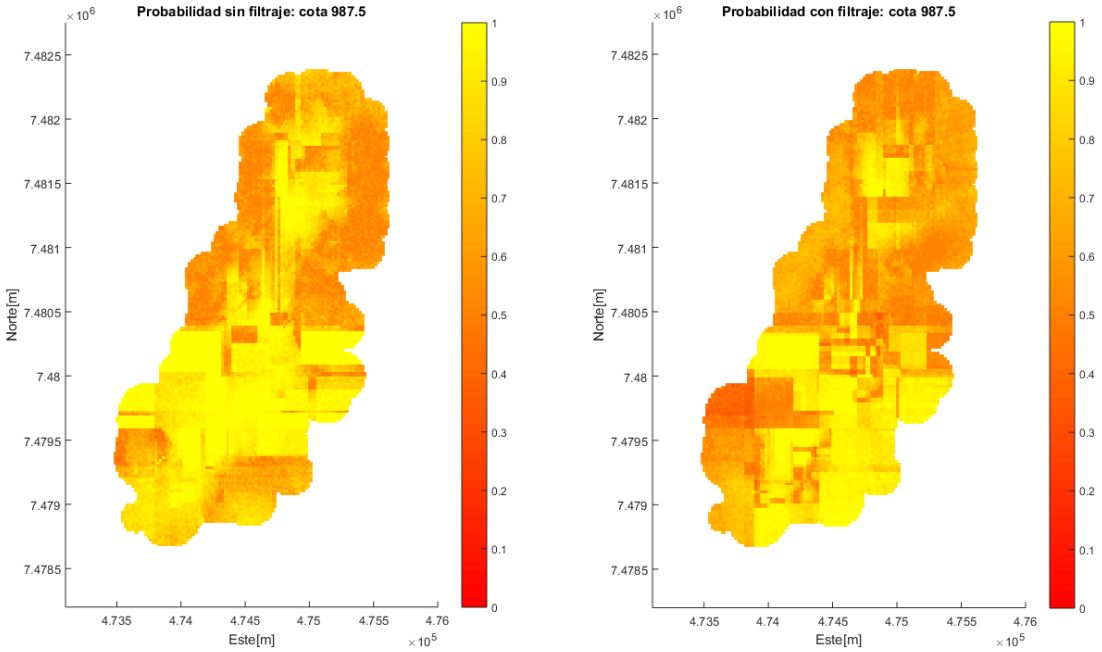

Ilustración 64 Probabilidad de clasificación cota 987.5

8.5.4.2 Cota 1137.5

- Clasificación más probable

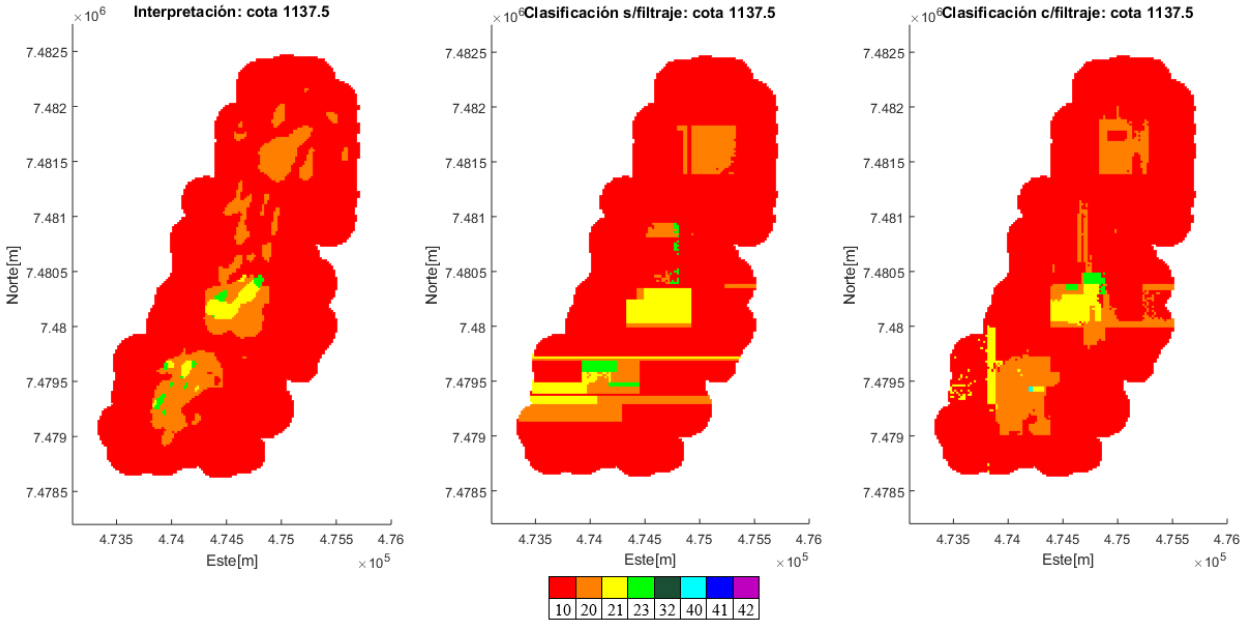

Ilustración 65 Clasificación más probable cota 1137.5 [m]

- Probabilidad de clasificación

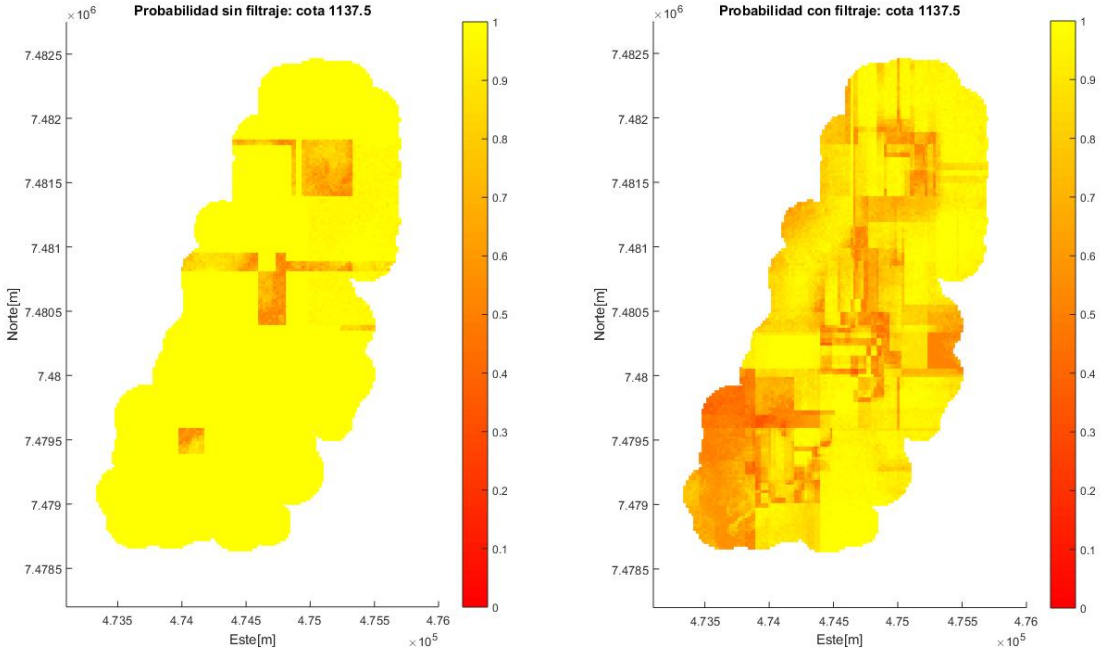

Ilustración 66 Probabilidad de clasificación cota 1137.5

8.5.4.3 Cota 1287.5

- Clasificación más probable

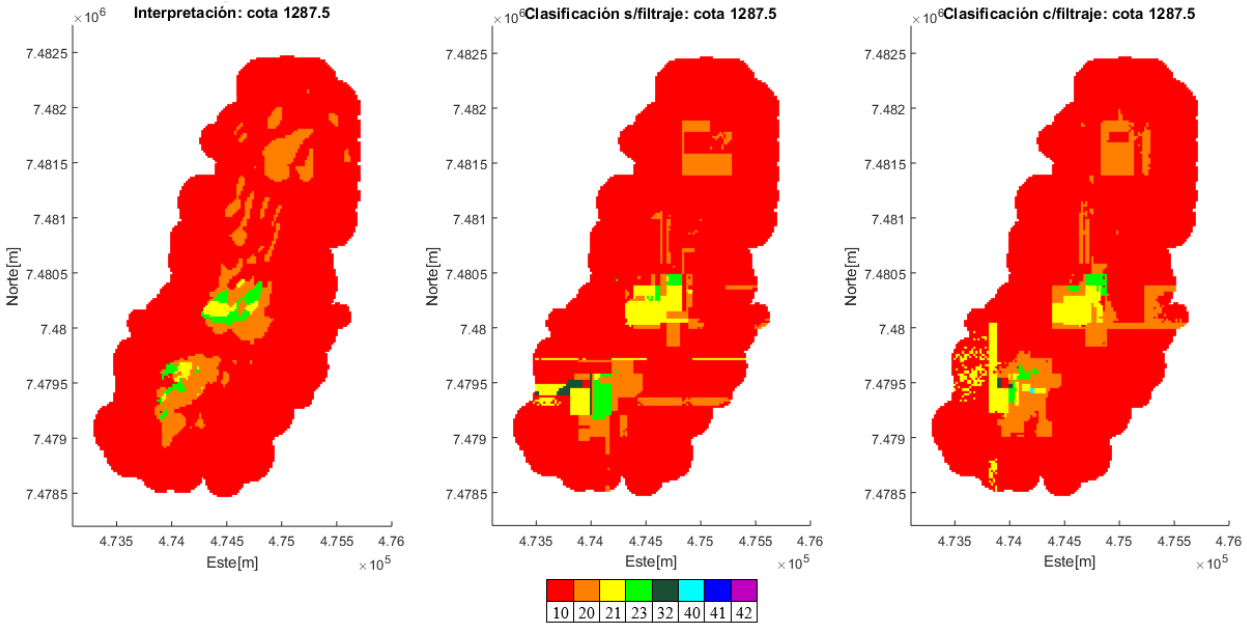

Ilustración 67 Clasificación más probable cota 1287.5 [m]

- Probabilidad de clasificación

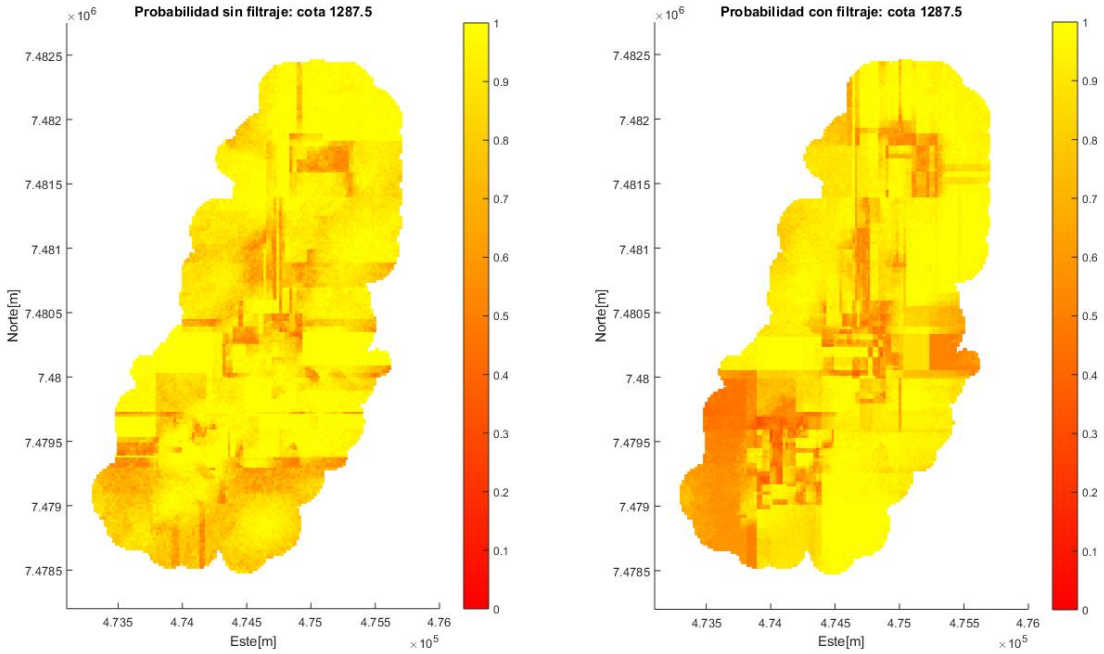

Ilustración 68 Probabilidad de clasificación cota 1287.5

8.5.4.4 Cota 1437.5

- Clasificación más probable

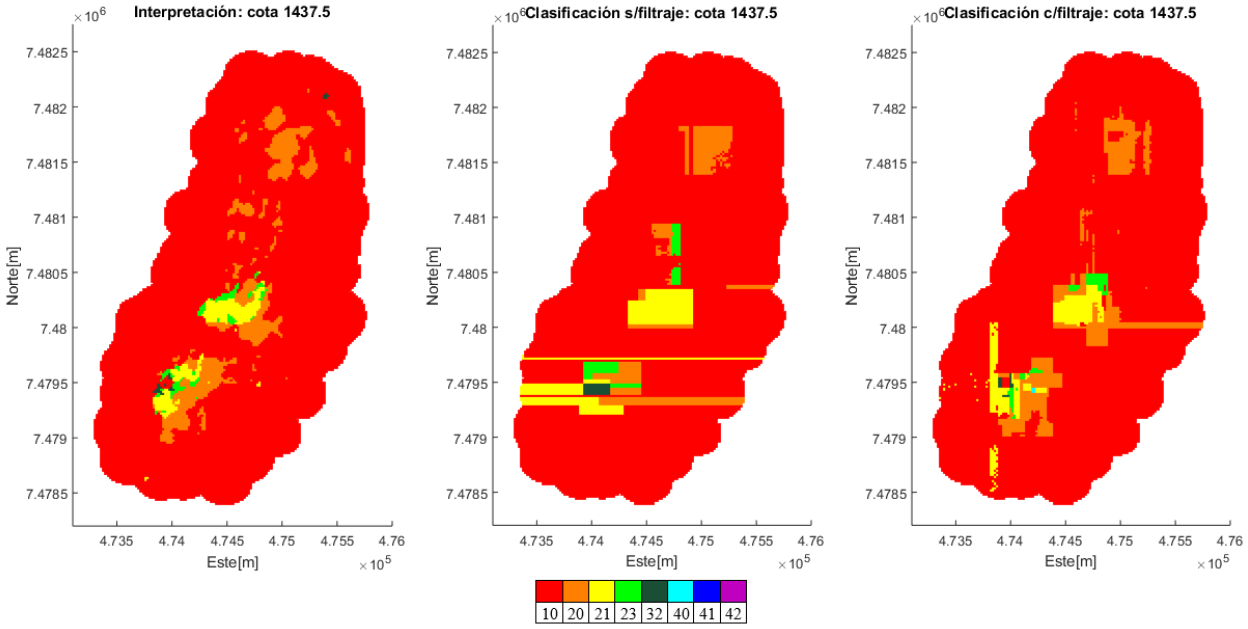

Ilustración 69 Clasificación más probable cota 1437.5 [m]

- Probabilidad de clasificación

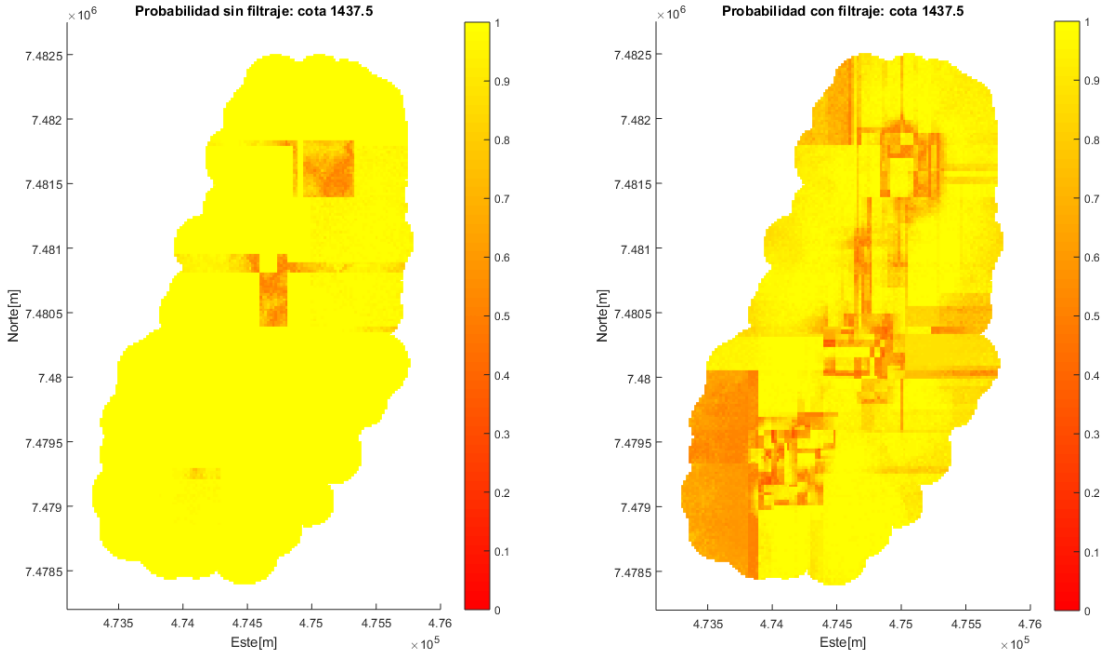

Ilustración 70 Probabilidad de clasificación cota 1437.5