

UNIVERSIDAD DE CONCEPCIÓN

Facultad de ingeniería

Departamento de Ingeniería Metalúrgica

Profesor Patrocinante

Sr. Roberto Fustos Toribio

ESTIMACIÓN DE RECURSOS MINERALES EN DOMINIOS GEOMETALÚRGICOS



RAFAEL FRANCISCO LARA MELCHER

Informe de Memoria de Título

para optar al Título de

Ingeniero Civil de Minas

Mayo 2020

Resumen

La estimación de recursos minerales es una etapa indispensable en la factibilidad de un proyecto minero. Es en ella donde se realiza la descripción, el estudio y la clasificación de los recursos contenidos dentro del potencial yacimiento a examinar. Es por ello que una buena metodología de análisis puede impactar en la viabilidad de un depósito.

Los dominios geometalúrgicos son zonas dentro del depósito que poseen características particulares que los vuelven identificables y diferenciables. Estas cualidades generan la posibilidad de estimar a través de algoritmos determinísticos, como el inverso de la distancia, o geoestadísticos, como las diversas variaciones de kriging, dentro de vecindades con información que ya está limitada por su similitud y que debería mejorar la calidad de las estimaciones. Es por lo anterior que en este trabajo se comparan las estimaciones dentro de un depósito polimetálico cuando estas son realizadas empleando dominios geometalúrgicos y cuando no son empleados.

La metodología de trabajo consiste en dos casos de estudio. El primero, denominado caso simulado, corresponde a la creación de 100 depósitos divididos previamente en dominios geometalúrgicos donde las variables son simuladas dentro de estos. Luego, a partir de sondajes, se realiza la estimación dentro de dichos dominios y sobre la base completa y se comparan los resultados con la base original. El segundo, denominado caso real, corresponde a la estimación del modelo de bloques de un yacimiento polimetálico bajo los mismos algoritmos y escenarios. En ambos casos una clasificación de recursos y un análisis económicos son llevados a cabo.

Finalmente, la estimación a través de dominios geometalúrgicos es mejor en el caso simulado el 99% de las veces, con una disminución promedio en el RMSE para el cobre de un 10.8%, 10.4% y 10.4% para el inverso de la distancia, el kriging ordinario y el co-kriging ordinario respectivamente; para el oro de 7.6% para los algoritmos de kriging y 8.1% para el inverso de la distancia; mientras que la precisión de la estimación del tipo de roca aumentó un 5.2% para el inverso de la distancia, y 4.7% para el co-kriging indicador y el co-kriging ordinario. Además, producto de la menor varianza de kriging la clasificación de recursos tiende a entregar bloques con menor incertidumbre y, por lo tanto, estimados del beneficio económico más confiables.

Abstract

Mineral resource estimation is a key indispensable part on the feasibility of a mining project. It is here where description, study and classification of the mining resources for a potential mine is done. That is why a good analysis methodology can impact on the viability of a deposit.

Geometallurgical domains are zones inside a deposit with particular characteristics which differentiates them from others and makes them identifiable. These qualities generate the possibility of estimating through deterministic algorithms, like inverse distance weighting, or geostatistical algorithms, like the multiple kriging variations available, inside neighborhoods already delimited with and by similar data, and which might enhance estimation quality. Because of this, in the present work different estimations are being compared inside a polymetallic deposit when geometallurgical domains are used and when they are not.

Work methodology consists in two study cases. The first one, named the simulated case, involves de creation of 100 mineral deposits previously divided by geometallurgical domains where variables are simulated within. Then, grade estimation is done inside those domains, over the entire data base, and results are compared with the previously simulated data. The second one, named the real case, involves the block model estimation of a polymetallic deposit under the same algorithms and scenarios. In both cases a resource classification and an economic analysis are carried out.

Finally, estimation through geometallurgical domains on the simulated case is better 99% of the times, with a mean decrease on the copper's RMSE of a 10.8%, 10.4% and 10.4% for inverse distance weighting, ordinary kriging and ordinary co-kriging respectively; in gold's case it decreased 7.6% for both kriging algorithms and 8.1% for inverse distance weighting; while estimation precision of the rock type variable has a mean increase of 5.2% for inverse distance weighting, and 4.7% for both indicator co-kriging and ordinary co-kriging. Also, because of the lower kriging variance, resource classification tends to result in blocks with lesser uncertainty, and, as a result of that, more confident economic benefit estimates.

Índice

1	Introducción	1
2	Descripción del problema	3
3	Hipótesis.....	5
4	Objetivos: Generales y Específicos	6
4.1	Objetivos generales	6
4.2	Objetivos específicos	6
5	Metodología.....	7
5.1	Caso simulado	7
5.2	Caso real.....	7
6	Marco Teórico.....	9
6.1	Dominios Geometalúrgicos	9
6.2	Estimación de Recursos Minerales	11
6.2.1	Introducción	11
6.2.2	Variables regionalizadas	12
6.2.3	Variografía	13
6.2.4	Estimación por inverso de la distancia	19
6.2.5	Estimación por Kriging	20
6.2.6	Estimación por Co-Kriging.....	25
6.2.7	Estimación por Kriging Indicador.....	30
6.2.8	Validación cruzada.....	31
6.2.9	Categorización de recursos.....	32
6.2.10	Comentarios generales.....	34
6.3	Extensión de dominios geometalúrgicos	36
6.3.1	Support Vector Machine (SVM).....	36
7	Caso Simulado	42
7.1	Metodología de simulación.....	42

7.2	Aplicación.....	43
7.2.1	Generación de la zona de trabajo	43
7.2.2	Creación de dominios geometalúrgicos.....	44
7.2.3	Simulación de variables	46
7.2.4	Extracción de sondajes	47
7.2.5	Aplicación de algoritmos de estimación.....	49
7.2.6	Evaluación estadística.....	61
7.2.7	Categorización de recursos.....	70
7.2.8	Análisis económico	71
7.3	Análisis de resultados	75
8	Caso Real.....	79
8.1	Metodología de análisis.....	79
8.2	Aplicación.....	80
8.2.1	Descripción y análisis de la base de datos	80
8.2.2	Construcción del modelo de bloques.....	84
8.2.3	Aplicación de algoritmos de estimación.....	85
8.2.4	Evaluación estadística.....	91
8.2.5	Categorización de recursos.....	95
8.2.6	Análisis económico	97
8.3	Análisis de resultados	100
9	Conclusiones y discusiones.....	103
10	Bibliografía.....	106
Anexo A	Determinación ley de corte, ingresos y costos	110
Anexo B	Índice de silueta	110
Anexo C	DBScan.....	112
Anexo D	Impacto outliers sobre RMSE.....	114

Índice de figuras

Figura 1. Conexión entre metalurgia y geología en relación a muestreos.....	10
Figura 2. Variable regionalizada en un espacio unidimensional.....	12
Figura 3. Área de tolerancia (verde) del vector de búsqueda h.....	14
Figura 4. Componentes de un variograma.....	14
Figura 5. Mapa variográfico con direcciones principales de anisotropía	15
Figura 6. Mapas variográficos y variogramas experimentales según anisotropías.....	16
Figura 7. Ejemplo de transformación de datos a indicadores.....	31
Figura 8. Relación entre recursos y reservas minerales	33
Figura 9. Hiperplano a través de dos clases linealmente separables.....	37
Figura 10. Hiperplano a través de dos clases linealmente no separables.....	38
Figura 11. Solución de problemas linealmente no separables	39
Figura 12. Visualización de zona de trabajo	44
Figura 13. Variograma empleado en la simulación de los dominios.....	44
Figura 14. Distribución de la variable generada dentro de la zona de trabajo.....	45
Figura 15. Dominios geometalúrgicos	45
Figura 16. Transformación de variable continua a categórica para el tipo de roca (dominio 3)	47
Figura 17. Histograma de largos de sondaje	48
Figura 18. Sondajes empleados para estimar simulación.....	48
Figura 19. Variograma y modelo para cobre y oro iteración 1, dominio 1	50
Figura 20. Predicción y varianza de kriging ordinario, cobre y oro iteración 1, dominio1	51
Figura 21. Variogramas directos y cruzados para tipo de roca, iteración 1, dominio 1.....	52
Figura 22. Predicción y varianza del co-kriging indicador, iteración 1, dominio 1.....	53
Figura 23. Variogramas experimental y teórico para el sistema de co-kriging entre oro, cobre y los tipos de roca, iteración 1, dominio 1	54
Figura 24. Predicción y varianza del co-kriging ordinario, iteración 1, dominio 1	55
Figura 25. Diagramas de caja para los errores cuadráticos medios para ambos escenarios de p .	56
Figura 26. Predicción del inverso de la distancia, iteración 1, dominio 1	57
Figura 27. Diagramas de caja para los RMSE de cada dominio para cobre y oro	58
Figura 28. Diagramas de caja para los RMSE de cada tipo de estimación para cobre y oro	59
Figura 29. Diagramas de caja de las precisiones de cada dominio para el tipo de roca	60
Figura 30. Diagramas de caja de las precisiones de cada tipo de estimación para cobre (izquierda) y oro (derecha).....	61

Figura 31. Distribución y gráfico de cuantil-cuantil de leyes de cobre y oro para cada caso	62
Figura 32. Varianza según caso, variable, método y algoritmo.....	69
Figura 33. Clasificación de recursos para cobre (izquierda) y oro (derecha)	71
Figura 34. Comparación estimación para kriging ordinario, estimación oro 26	78
Figura 35. Vista superior (izquierda) e inferior (derecha) de los dominios geometalúrgicos.....	81
Figura 36. Comparación distribuciones entre dominios y la base completa	82
Figura 37. Comparación distribuciones entre dominios y la base completa después de filtrar outliers y duplicados	83
Figura 38. Modelo de bloques base real.....	84
Figura 39. Dominios geometalúrgicos modelo de bloques base real	85
Figura 40. Estimación por kriging ordinario para oro, dominio tres	87
Figura 41. Estimación por co-kriging para oro, dominio tres	88
Figura 42. Estimación por inverso de la distancia para oro, dominio tres	89
Figura 43. Distribución y gráfico de cuantil-cuantil de leyes de oro y cobre para cada dominio	91
Figura 44. Distribuciones de oro y cobre para el modelo de bloques cuando se estima con y sin dominios geometalúrgicos.....	94
Figura 45. Clasificación de recursos para el modelo de bloques	96
Figura 46. Intervalos de confianza del 75%,85% y 95% para los diferentes beneficios	100
Figura 47. Comparación entre varianzas co-kriging ordinario oro sin dominios geometalúrgicos (A) y con estos (B) y clasificación de recursos respectiva (C) y (D), nivel 329, dominio 1	101
Figura C-1. Conceptos de alcanzable y conectado por densidad vistos gráficamente.....	113

Índice de tablas

Tabla 1. Modelos variográficos clásicos	18
Tabla 2. Clasificación de recursos	34
Tabla 3. Funciones de kernel más conocidas	40
Tabla 4. Número de bloques empleados para realizar estimación	48
Tabla 5. Ejemplo elección tipo de roca preponderante co-kriging indicador	52
Tabla 6. Ejemplo elección tipo de roca preponderante co-kriging ordinario	55
Tabla 7. Ejemplo elección tipo de roca preponderante inverso de la distancia indicador	56
Tabla 8. RMSE medio para cada variable y algoritmo	59
Tabla 9. Precisión media para cada variable y algoritmo	61
Tabla 10. Correlaciones para los casos a estudiar para los sondeos y la base de validación	62
Tabla 11. Resultados test Kolmogorov–Smirnov	63
Tabla 12. Medidas estadísticas principales para la mejor simulación	64
Tabla 13. Porcentajes de acierto para la mejor simulación	64
Tabla 14. Medidas estadísticas principales para la peor simulación	66
Tabla 15. Porcentajes de acierto para la peor simulación	66
Tabla 16. Medidas estadísticas principales para la simulación más probable	67
Tabla 17. Porcentajes de acierto para la simulación más probable	68
Tabla 18. Resumen de resultados de simulación	68
Tabla 19. Metodología de clasificación de recursos resumida	70
Tabla 20. Tonelaje y ley media para los recursos medidos, indicados e inferidos	72
Tabla 21. Diferencia recursos medidos entre emplear dominios geometalúrgicos y no emplearlos	73
Tabla 22. Tonelaje y ley media para los recursos medidos, indicados e inferidos después de filtrar por ley de corte	73
Tabla 23. Diferencia recursos medidos entre emplear dominios geometalúrgicos y no emplearlos	74
Tabla 24. Ingresos, costos y beneficios en MUSD	74
Tabla 25. Valores de los beneficios reales [MUSD] y diferencias con beneficios estimados	75
Tabla 26. Medidas estadísticas principales para la ley equivalente después de filtrar por ley de corte	77
Tabla 27. Estadísticas descriptivas base de datos real	81
Tabla 28. Estadísticas descriptivas base de datos final después de filtrar outliers y duplicados	83
Tabla 29. Dominios reales vs dominios SVM	85

Tabla 30. Variogramas para cada variable y dominio	86
Tabla 31. Variogramas para cada variable base completa	86
Tabla 32. Valores de p óptimos para el inverso de la distancia	89
Tabla 33. RMSE para cada dominio	90
Tabla 34. RMSE cuando se emplean y no se emplean dominios geometalúrgicos	91
Tabla 35. Correlación entre oro y cobre	92
Tabla 36. Resultados test Kolmogorov-Smirnov	92
Tabla 37. Medidas estadísticas principales base de datos real	93
Tabla 38. Porcentajes de cambio del error absoluto, disminución de RMSE y error absoluto medio	93
Tabla 39. Medidas estadísticas principales modelos de bloques estimados.....	94
Tabla 40. Varianza media de estimación.....	95
Tabla 41. Porcentajes para cada clasificación de recursos	97
Tabla 42. Tonelaje y ley media para los recursos medidos, indicados e inferidos	98
Tabla 43. Diferencia recursos medidos entre emplear dominios geometalúrgicos y no emplearlos en [Mton].....	98
Tabla 44. Tonelaje y ley media para los recursos medidos, indicados e inferidos después de filtrar por ley de corte	98
Tabla 45. Diferencia recursos y estéril entre emplear dominios geometalúrgicos y no emplearlos	99
Tabla 46. Ingresos, costos y beneficios en MUSD.....	99
Tabla 47. Intervalos de confianza del 75% para los diferentes beneficios	100
Tabla B-1. Interpretación subjetiva del índice de silueta	112
Tabla D-1. Medidas estadísticas principales antes de filtrar la base de datos	115
Tabla D-2. Medidas estadísticas principales después de filtrar la base de datos.....	115

1 Introducción

La estimación de recursos minerales es una etapa de suma importancia dentro del estudio de un depósito o en la expansión de un yacimiento, ya que se encarga de describir completamente la zona a estudiar en cuanto a la concentración de los elementos presentes, los tipos de minerales y las características de estos. Los lugares de estudio son generalmente representados por bloques de dimensiones determinadas por los requerimientos del depósito con respecto a la definición requerida y los fines de este.

La información a partir de la cual se genera la estimación de recursos viene generalmente en la forma de objetos puntuales compuestos por la posición de los sondajes originales y la información de estudios químicos o geológicos que es anexada a estos sondajes con las características que se deseen estimar.

Las técnicas y algoritmos empleados en la construcción de modelos de bloques han ido avanzando con el paso de las décadas [1]. Desde métodos determinísticos, cuya simplicidad e intuitividad los vuelven sumamente prácticos para estudios preliminares; pasando por métodos geoestadísticos, que comenzaron a generar interés sobre la comunidad científica en la década de los 80' (producto del aumento de la capacidad de procesamiento de los computadores) [2] a partir de los estudios de Danie G. Krige [3] y Georges Matheron [4], siendo estos algoritmos basados en la idea de continuidad espacial, incorporando elementos como el variograma; hasta métodos provenientes de la inteligencia artificial [5], que a pesar de sus grandes costes de optimización, han probado en los últimos años ser alternativas interesantes en la estimación de depósitos con características particularmente complejas.

Dentro de un depósito es comúnmente posible identificar zonas con características particulares que los diferencien de otras localidades. Estos lugares pueden ser encontrados a través de agrupamiento de datos. Dichas zonas reciben el nombre de dominios geometalúrgicos y ofrecen la posibilidad de describir y diferenciar el agrupamiento en base a alguna característica en particular, algún efecto aguas abajo en el proceso, o a alguna combinación de estos.

Por otro lado, existen diversos códigos internacionales que intentan establecer pautas sobre cómo realizar la clasificación de recursos, sobre quienes son aptos para esto y las formalidades requeridas en la presentación de los resultados. Aun así, no existe consenso en cómo realizar la

clasificación de recursos [6], por lo que virtualmente no existen metodologías que sean mejores que otras y cada operación tiende a usar las técnicas que ellos estimen convenientes y que mejor desempeño hayan mostrado históricamente.

Estudios anteriores han demostrado que realizar una estimación empleando dominios geometalúrgicos ofrece beneficios como una disminución en los tiempos de cálculo [7] y una disminución en los errores con respecto a bases de validación [8], pero no existen estudios que muestren los efectos sobre la varianza de estimación o la clasificación de recursos, menos con respecto a un análisis económico. Es por esto que nacen las interrogantes: ¿Genera beneficios sobre la clasificación de recursos y la incertidumbre de los bloques la estimación por medio de dominios geometalúrgicos? Y si es así, ¿Cuáles son estos beneficios y qué efectos producen?

Para llevar a cabo la propuesta mencionada, se diseñan dos casos de estudio para comprender el efecto de la estimación por medio de dominios geometalúrgicos: un caso simulado, que consiste, como su nombre lo sugiere, en la generación de depósitos simulados siguiendo la definición de dominios geometalúrgicos donde se lleven a cabo estimaciones dentro de cada dominio y, posteriormente, en la base de datos completa, empleando los mismos sondajes para cada simulación, con el fin de comparar los resultados obtenidos con la base de validación y realizar una clasificación de recursos y un análisis económico de los datos estimados. El segundo caso consiste en la implementación de todo lo mencionado anteriormente con una base de datos real, pero con la adición de generar un modelo de bloques separado por dominios geometalúrgicos y estudiar los resultados de la clasificación de recursos y el análisis económico cuando se emplea y cuando no se emplean dominios geometalúrgicos dentro de la estimación.

Si los resultados son favorables, podrían generar más argumentos a favor de esta metodología que demuestren los beneficios que se podrían obtener a partir del uso de dominios geometalúrgicos por sobre la aplicación directa de herramientas de estimación sobre depósitos polimetálicos, siempre y cuando dominios geometalúrgicos puedan ser identificados.

2 Descripción del problema

El problema abordado en esta memoria de título está relacionado con la estimación de recursos en un depósito mineral polimetálico empleando como información la ubicación espacial de dominios geometalúrgicos.

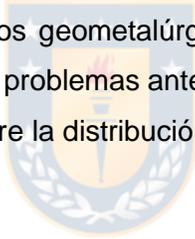
En general, cuando se desea estimar recursos se utiliza únicamente la información proveniente de los sondeos, los que corresponden a leyes de minerales de interés económico, dejando fuera a otras variables que podrían potencialmente mejorar la calidad de estimación y, dada la correlación espacial inherente de los suelos, ayudar a explicar la distribución de la mineralogía. Producto de esta metodología, se producen graves errores aguas abajo en el procesamiento mineral, ya que un problema tan complejo como lo es la estimación se realiza de manera unidimensional, es decir, haciendo responsable a una sola variable (ley de mineral, por ejemplo) de todas las dificultades en procesos posteriores. Este problema conlleva a tener una pobre interpretación de la distribución mineral real en el depósito y un modelo de bloques con una alta incertidumbre, haciendo que los costos por procesamiento se incrementen y que se pierda confianza en el modelo, por lo que este pierde total validez y razón de ser. Dicha variable, incapaz de explicar los parámetros que regulan los procesos posteriores, da paso a la incorporación de otras variables al modelo que predigan de mejor manera el comportamiento de la roca en procesos aguas abajo y ayuden a comprender de mejor manera las tendencias y la variabilidad en la mineralización. Esta solución se traduce en incorporar los conceptos de continuidad espacial que aportan los dominios geometalúrgicos, los que condicionan y caracterizan la mineralización de un depósito a través de agrupamientos o clustering.

Ahora, no todas las variables aportan información que puede ayudar a la generación de dominios geometalúrgicos. Existe también la posibilidad de generar agrupamientos erróneos producto del uso de parámetros no correlacionados o del colapso de modelos debido al infra entrenamiento o al entrenamiento con parámetros demasiado similares que vuelvan al modelo incapaz de predecir o estimar valores que sean demasiado diferentes del conjunto de prueba (“*overfitting*”). Si bien es cierto que el Kriging, siendo el algoritmo más empleado en estimación de recursos minerales, genera un suavizamiento de la información real, este efecto puede ser acrecentado, aplacando el comportamiento regional de las variables de estudio y mermando el efecto de los dominios geometalúrgicos, o bien este efecto ser disminuido, complejizando los modelos de bloques y haciendo más engorroso el estudio de reservas, por ejemplo, para después

hacer mucho más difícil la planificación minera. Luego es importante tener en cuenta los principios básicos de agrupamiento y estimación para no hacer más difícil esta última y generar modelos confiables.

Actualmente, los algoritmos empleados para estimar recursos minerales están condicionados a la cantidad de información disponible. Además, estos algoritmos presentan problemas relacionados con la propia construcción de las estimaciones, tales como la creación y aproximación de variogramas, el sesgo condicional en el caso del Kriging o la falta de una medida del error en el caso de la interpolación por inverso de la distancia, por nombrar algunos. Estas limitaciones generan complicaciones en la construcción del modelo de bloques, que acarrea estos problemas a procesos de mina posteriores, produciendo un impacto económico significativo, más aún si se ignora la correlación entre variables pertenecientes a un mismo dominio geometalúrgico.

El caso ideal sería contar con una construcción del modelo de bloques precisa y con la menor incertidumbre posible, tomando en cuenta la escasa información de la que generalmente se dispone y la distribución espacial de los dominios geometalúrgicos. Para esto se emplea un conjunto de procedimientos que permiten sortear los problemas antes descritos y que puedan entregar la mayor cantidad y calidad de la información sobre la distribución mineral, minimizando el riesgo económico y beneficiando el plan minero.



3 Hipótesis

Es posible disminuir la incertidumbre en la construcción del modelo de bloques de un depósito mineralizado a partir de la estimación por sectores claramente individualizables, donde la roca posea diferentes comportamientos geometalúrgicos distinguibles entre sí, en contraste al procedimiento común de estimar el modelo completo a partir de toda la información del yacimiento. Lo anterior se debe principalmente a que la estimación se realiza con vecindades que sí están correlacionadas con el sitio a estimar y por ende no se emplean datos que contaminen el resultado final. Esto además permite generar menos riesgos para las inversiones y un estudio económico que se ajuste más a la realidad.



4 Objetivos: Generales y Específicos

4.1 Objetivos generales

El objetivo general de esta memoria de título es realizar una comparación en la estimación de recursos de un depósito mineral usando la definición de dominios geometalúrgicos. A su vez, ver cómo esto impacta en la evaluación del riesgo económico de un proyecto minero y en la categorización de recursos.

4.2 Objetivos específicos

Los objetivos específicos de esta memoria son:

- Aplicar modelos tradicionales de estimación para la construcción de modelos de bloques y cuantificar la incertidumbre empleando e ignorando la definición de dominios geometalúrgicos.
- Ver cómo impacta la decisión de emplear o no dominios geometalúrgicos en la estimación de recursos minerales.
- Realizar una comparación de diferentes escenarios de estimación utilizando los métodos actuales.
- Realizar una comparación entre diferentes métodos de estimación. Además de estudiar los efectos en la modificación de parámetros empleados.

5 Metodología

La metodología usada para abordar el problema planteado consiste en el análisis de dos escenarios, uno con un conjunto de datos simulados y otro con un conjunto de datos reales.

5.1 Caso simulado

El procedimiento de trabajo con el conjunto de datos simulados será el siguiente:

- Se generará una base de datos simulados con dos atributos continuos y uno discreto en tres dimensiones, correspondientes a un muestreo de una zona de estudio.
- Se particionará la base de datos en dos, sondajes de exploración y base de validación. Con la primera se construirán modelos de bloques con los atributos continuos y discretos predichos en las localidades de la base de datos de validación y estos se compararán con el atributo "real" de la misma base. Esto se realizará con y sin tomar en cuenta la definición de los dominios geometalúrgicos.
- Se realizará un estudio comparativo con las medidas estadísticas principales asociadas a las predicciones. Se estudiará la variabilidad y las medidas de incertidumbre de las predicciones.
- En el modelo de bloques se generará una categorización de recursos, reportando las medidas estadísticas principales.
- Se realizará un análisis económico de los diferentes escenarios de riesgo y como se diferencian del modelo de bloques real

5.2 Caso real

El procedimiento de trabajo con el conjunto de datos reales será el siguiente:

- Se realizará una descripción detallada de la base de datos

- Se usarán dominios geometalúrgicos para la construcción del modelo de bloques y cuantificación de incertidumbre.
- Se aplicarán los algoritmos tradicionales de estimación de recursos (Kriging, co-kriging, inverso de la distancia) con y sin dominios geometalúrgicos.
- Se realizará el plan de categorización de recursos para cada modelo de bloques obtenido.
- Se realizará una evaluación económica en cada escenario y se entregará el peor, el mejor y el caso más probable de beneficio económico.

En cada uno de los conjuntos de datos se realizará un análisis crítico de los resultados y se elaborarán conclusiones y discusiones parciales.



6 Marco Teórico

6.1 Dominios Geometalúrgicos

La geometalurgia es un campo relativamente nuevo en minería que trata de entender y modelar variables que están relacionadas con el desempeño metalúrgico. Estas incluyen variables que miden directa o indirectamente producción, recuperación y calidad del concentrado; como, por ejemplo: dureza, liberación, tipo de mineral, forma mineral, textura mineral, recuperación, entre otros. Para poder tomar muestras representativas para clasificación mineral y ensayos geometalúrgicos es necesario particionar el depósito en regiones homogéneas en términos de estas propiedades de procesamiento, llamados dominios geometalúrgicos [9].

Las variables geometalúrgicas poseen sus propios desafíos. Al menos dos aspectos de estas variables deben ser tomados en cuenta a la hora de ser estudiadas e interpretadas dentro de un modelo geometalúrgico:

- Las variables son usualmente medidas indirectamente, o a través de aproximaciones de la incógnita metalúrgica de interés. En algunos casos no se pueden promediar linealmente [1].
- Varias evaluaciones de desempeño metalúrgico no son dependientes linealmente de las variables medidas. Este problema es relevante ya que las predicciones requeridas usualmente necesitan estar bajo un soporte distinto al empleado cuando fueron medidas.

Cuando una variable no es lineal, promediar la variable de la manera usual o ponderar con los tonelajes no es lo correcto. Es por ello que herramientas como el kriging o el inverso de la distancia permiten estimar los valores en lugares no medidos y bajo un soporte diferente al original.

La incorporación de dominios geometalúrgicos permite mejorar las operaciones mineras a través de un mejor entendimiento de la interacción entre el cuerpo mineralizado y las respuestas de los procesos geometalúrgicos. El valor entregado por estos modelos radica en el uso de este conocimiento previo para generar y escoger la configuración de mina más robusta y entender mejor las consecuencias de decisiones irreversibles de gran envergadura [10] como, por ejemplo: expansiones de planta, cambios en el diseño de la mina, políticas operacionales, disminuciones en

la escala de producción. El uso de dominios geometalúrgicos además incrementa las probabilidades de realizar mejores decisiones de diseño y configuración minera, aumentando los retornos financieros y brindando seguridad a la inversión frente al, actualmente, aumento de las incertidumbres en la minería con respecto a las bajas leyes presentes y la cantidad de contaminantes y elementos no deseados comprendidos en la roca.

Un desafío actual es la transición de modelos puramente geológicos a modelos geometalúrgicos. Uno de los principales problemas radica en la significativa diferencia entre la cantidad de información geológica contrastada con la escasa o nula información geometalúrgica disponible normalmente en las operaciones. Esto significa una cantidad insuficiente y poco confiable de pruebas geometalúrgicas realizadas y, por lo tanto, impide la incorporación exitosa de las respuestas geometalúrgicas en los modelos geológicos [11] a través de la geoestadística.

Las herramientas y métodos empleados en el modelamiento geológico han ido mejorando debido al constante foco que se ha generado en torno a la geometalurgia. El concepto de optimizar la disposición espacial empleando pocas muestras de alto costo y luego construir un modelo calibrado es uno de los principales responsables de ello. Luego, producto de lo anterior, se han realizado mejoras en los experimentos llevados a cabo y en la cantidad y calidad de muestras que son empleadas para el modelamiento geometalúrgico [10], Figura 1, donde el objetivo es lograr un balance entre el alto costo por muestra y la baja cobertura espacial del nivel-4 con el bajo costo por muestra y la alta cobertura espacial del nivel-1.

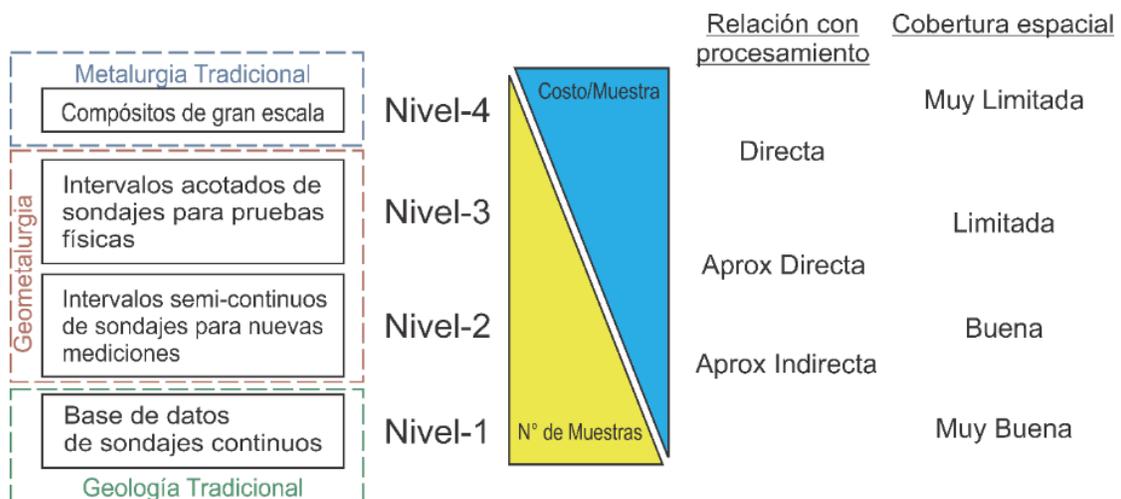


Figura 1. Conexión entre metalurgia y geología en relación a muestreos

6.2 Estimación de Recursos Minerales

6.2.1 Introducción

Las mineralizaciones en la corteza terrestre son generadas a través de una sucesión de procesos geológicos muchas veces no comprendidos completamente en su origen o funcionamiento [1]. Condiciones necesarias para la ocurrencia de deposición mineral incluyen fuentes de mineralización, caminos donde pueda transportarse, y condiciones geológicas favorables. Los procesos físicos y químicos adecuados pueden producir concentraciones minerales importantes.

Tradicionalmente, algoritmos determinísticos, como el inverso de la distancia, han sido empleados para determinar los valores de las variables muestreadas en las ubicaciones de interés, debido a la simplicidad del método y a la poca potencia computacional que requiere. Sin embargo, las estimaciones solo dependen de la posición y el valor de los datos a estimar, siendo esta familia de algoritmos insensible a la correlación espacial de las variables.

Las características de los depósitos minerales invariablemente imprimen patrones de correlación espacial, los cuales son importantes a la hora de hacer una evaluación de recursos y planificar la operación de la mina. Lograr describir y modelar estos patrones de correlación permiten comprender de mejor manera los procesos que llevaron a cabo la deposición del material y mejoran la precisión con la que se puede predecir la existencia de mineralización y sus respectivas características dentro de zonas que no han sido objeto directo de muestreo. La geoestadística puede ser empleada para describir estas correlaciones bajo un marco teórico apropiado.

La palabra geoestadística fue acuñada por Georges Matheron en 1962 y corresponde a una rama de la estadística encargada del estudio de las dependencias o correlaciones entre las observaciones disponibles, considerando que ellas están ubicadas en el espacio [12]. En un comienzo fue empleada dentro del campo de la minería y las ciencias de la tierra, pero hoy en día es aplicada en campos que estudian propiedades o atributos de fenómenos regionalizados, como lo son la meteorología, geografía e hidrología, por nombrar los más importantes.

La principal herramienta de la geoestadística es el variograma, el cual indica la relación entre la distancia en que se encuentren los pares de datos y su disimilitud o variabilidad. Su función es graficar el comportamiento de todos los pares de datos en una dirección dada para poder entender el comportamiento de la variable regionalizada a estudiar dentro del campo finito definido a priori e

identificar posibles anisotropías o direcciones preferenciales mediante el estudio multidireccional de la variable en cuestión. Luego, a través de las técnicas del kriging y sus variantes, es posible estimar los valores de estas variables, tanto continuas como categóricas, en los lugares donde no fue físicamente muestreada, con el fin de lograr conocimiento completo y ofrecer una interpretación de todo el dominio a estudiar.

6.2.2 Variables regionalizadas

La teoría geoestadística aborda el estudio y modelización de variables que se denominan regionalizadas, es decir, fenómenos, generalmente naturales, que se extienden en el espacio y presentan cierta continuidad [13]. Las características principales de estas variables, Figura 2, son:

- Nivel local: aspecto aleatorio asociado a variaciones erráticas e impredecibles de la variable.
- Nivel global: aspecto estructurado que refleja las características generales del fenómeno.



Figura 2. Variable regionalizada en un espacio unidimensional

Las variaciones pequeñas de la variable regionalizada son las que dificultan su modelización, tratamiento o descripción mediante un modelo simple. Sin embargo, del punto de vista conceptual, una variable regionalizada es una función determinística [12]. En la teoría de las variables regionalizadas, este problema de modelización es solucionado mediante la interpretación de estas

mediante el uso de una función aleatoria, que resulta estructurada a gran escala y errática a pequeña escala.

Ejemplos de variables regionalizadas son la:

- conductividad hidráulica, porosidad, permeabilidad de un acuífero.
- precipitación, presión, temperatura, velocidad del viento en meteorología.
- altitud geodésica, pendiente en topografía.
- concentración de cierto contaminante, resistividad en el suelo.
- ley de un mineral, potencia de la veta, densidad de roca, recuperación metalúrgica en minería.

Dado que un fenómeno regionalizado nunca posee una extensión finita, se estudia la variable regionalizada sobre un dominio limitado D llamado campo de la variable y corresponde a la zona de interés de esta. Fuera de este campo la variable no se encuentra definida.

6.2.3 Variografía

Los valores de una variable regionalizada no son independientes, un valor observado en un sitio entrega información sobre valores de los sitios aledaños. En la interpretación probabilística de la variable regionalizada, esta noción intuitiva de dependencia está relacionada con los momentos estadísticos de segundo orden (covarianza, correlograma y variograma), los cuales, a diferencia de la esperanza, entregan información sobre la relación que existiese entre pares de datos. Esto da paso a una descripción elemental de la continuidad espacial de la variable regionalizada [12].

Variograma experimental

El variograma es una medida de variabilidad geológica versus distancia. Esta variabilidad puede ser diferente en diferentes direcciones; por ejemplo, en formaciones sedimentarias existe una mayor correlación espacial en el plano horizontal.

Se considera una variable regionalizada z conocida en n sitios $\{x_1, \dots, x_n\}$. Se define el semivariograma experimental según la ecuación (1). Cabe destacar que el termino semivariograma se refiere a la mitad del variograma [1], pero por simplicidad, será llamado variograma de aquí en adelante.

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - Z(x)]^2\} = \frac{1}{2|N(h)|} \sum_{N(h)} [z(x_\alpha) - z(x_\beta)]^2, \quad (1)$$

donde $N(h) = \{(\alpha, \beta) \mid x_\alpha - x_\beta = h\}$ y $|N(h)|$ es el número de pares contenidos en $N(h)$.

Ahora, la definición entregada en la ecuación (1) carece de herramientas para enfrentar una situación en la que los datos estén dispuestos de manera irregular, ya que no es habitual encontrar pares de datos que estén siempre a la misma distancia, es por esto que, para que el variograma sea más robusto, se agregan tolerancias a la dirección y amplitud del vector de búsqueda, Figura 3, con el fin de realizar comparaciones entre la mayor cantidad de datos posibles para las mallas de muestreo dadas.

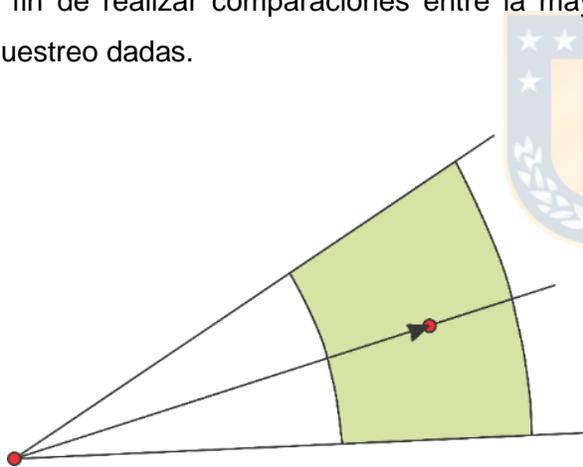


Figura 3. Área de tolerancia (verde) del vector de búsqueda h

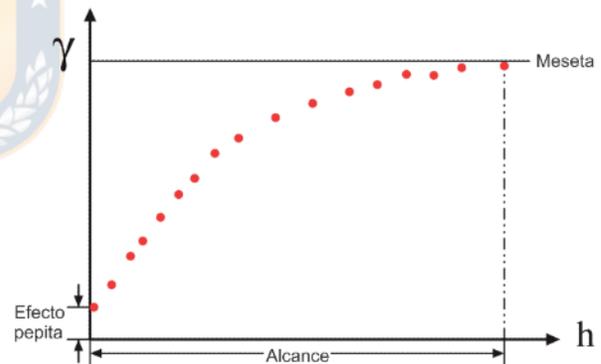


Figura 4. Componentes de un variograma

Los parámetros a especificar para generar un variograma experimental son: la dirección del vector h , el largo del vector h y los intervalos de h que van a ser estudiados ("paso"), la tolerancia a la dirección y la tolerancia al largo mismo del vector h .

Una vez realizado el estudio variográfico, se genera un conjunto de puntos como la Figura 4, donde cada punto corresponde al promedio del variograma para todos los pares de datos distanciados en h unidades.

Las principales características que componen un variograma son:

- Efecto pepita: es el efecto producido por la variabilidad a distancias pequeñas, revelando la existencia de microestructuras discontinuas [4].
- Alcance: distancia bajo la cual aún existe correlación positiva entre los pares de datos comparados.
- Meseta: es el valor para el cual el variograma se estabiliza. Corresponde a la varianza a priori, el valor del variograma a partir del cual no hay correlación.

Mapa variográfico

El mapa variográfico es una herramienta que consiste en la visualización del variograma experimental en todas las direcciones del espacio sobre un plano dado (generalmente coordenadas Este y Oeste), bajo la forma de un mapa con escala de color o de grises [12], Figura 5. Ayuda a distinguir si existe anisotropía, para luego graficar el variograma experimental a lo largo de las direcciones principales de anisotropía y así conocer el comportamiento de los datos [14].

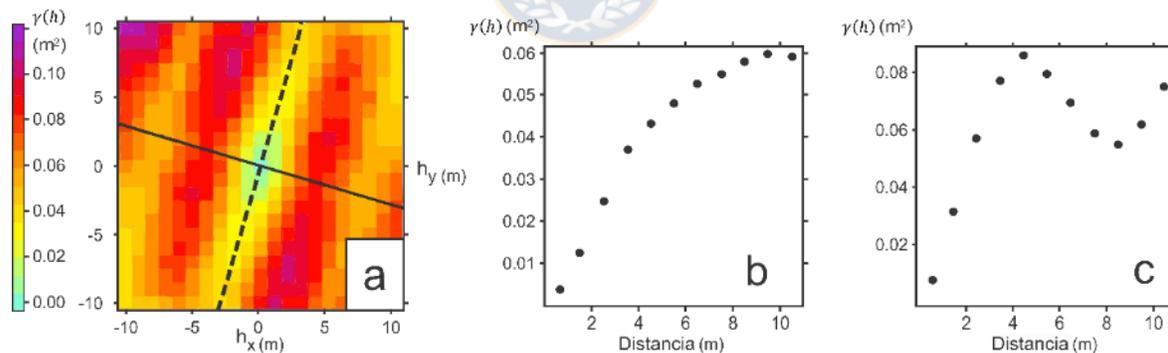


Figura 5. Mapa variográfico con direcciones principales de anisotropía

Anisotropía

Dentro de un mapa variográfico se pueden identificar 3 tipos principales de anisotropías, Figura 6, las cuales son generadas por cambios en las estructuras del variograma en las distintas direcciones de estudio, estos son:

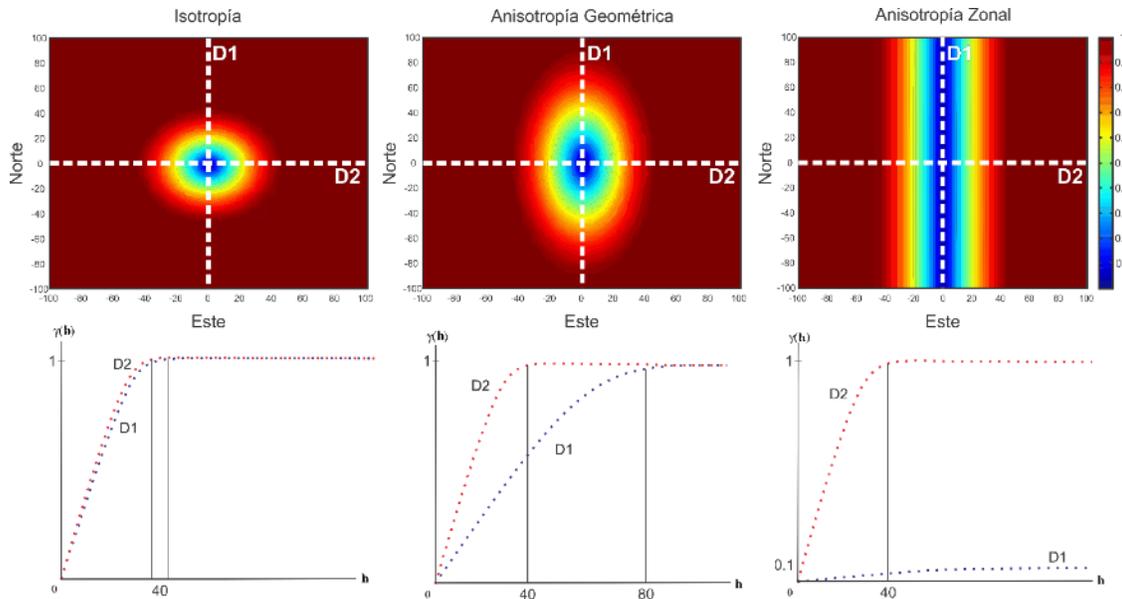


Figura 6. Mapas variográficos y variogramas experimentales según anisotropías

- Isotropía: el variograma experimental posee un comportamiento similar en todas las direcciones, por lo tanto, genera circunferencias concéntricas en el mapa variográfico.
- Anisotropía geométrica: el variograma experimental posee dos direcciones principales en las que con alcances diferentes se llega a la misma meseta [15]. Se reconoce gráficamente porque genera elipses concéntricas en el mapa variográfico.
- Anisotropía zonal: el variograma experimental posee dos direcciones en las que claramente los alcances y las mesetas son diferentes. Se reconoce gráficamente porque genera bandas en el mapa variográfico o elipses extremadamente excéntricas.

Finalmente, la combinación de cualquiera de estas anisotropías con ángulos de interacción complejos permite interpretar la mayoría de las situaciones prácticas.

Variograma teórico

El variograma teórico es el ajuste mediante funciones paramétricas que se aproximan al variograma experimental, a modo de obtener un modelo útil que represente a los datos originales y permita realizar kriging a posteriori. La eficacia de este modelamiento es crucial para obtener una buena estimación de la base de datos.

Como $Z(x)$ es una función aleatoria estacionaria de segundo orden [16], esto implica que tanto la media m como la varianza σ^2 no dependen de la posición de los objetos [17]. Luego, en la ecuación (2), que define la covarianza, que sólo depende del vector de separación h , lo que implica que mientras más cercanos estén los objetos, más se parecen entre sí.

La covarianza $C(h)$ es 0 cuando los valores a una distancia h no están linealmente correlacionados. Luego a $h = 0$ la varianza es igual a $C(0)$, ecuación (3).

Luego, al hacer un desarrollo de la ecuación (1), empleando los supuestos anteriores, se genera la ecuación (4), que permite relacionar varianza y covarianza dentro de un variograma, la cual implica que el variograma no es más que otra manera de medir similitud (o disimilitud en este caso) análoga a la covarianza [13].

$$C(h) = E\{(Z(x) - m) * (Z(x + h) - m)\} \quad (2)$$

$$Var[Z(x)] = Var[Z(x + h)] = E[Z(x + h)^2] = C(0) \quad (3)$$

$$\gamma(h) = C(0) - C(h) \quad (4)$$

Las principales razones por las cuales es importante la modelación correcta del variograma experimental son:

- Los procesos y estimaciones subsecuentes requieren de un variograma definido para todos los valores de h posibles y en todas las direcciones. Se busca que esta función incorpore toda la información derivada del variograma experimental: anisotropías, tendencias, efecto pepita.
- La covarianza debe ser definida positiva. El variograma teórico debe ser definido positivo porque esto asegura que las ecuaciones de kriging pueden ser resueltas, que la solución es única y que la varianza de kriging es positiva [1].

Los modelos variográficos básicos que se emplean para la construcción del variograma teórico [15] se resumen en la Tabla 1. Estos son los modelos más utilizados porque corresponden a funciones definidas positivas y en la práctica son las que más asemejan el comportamiento que adquieren los variogramas experimentales.

Tabla 1. Modelos variográficos clásicos

Modelo	Ecuación
Efecto pepita	$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C & \text{en otro caso} \end{cases}$
Esférico	$\gamma(h) = \begin{cases} C \left[\frac{3 h }{2a} - \frac{1}{2} \left(\frac{ h }{a} \right)^3 \right] & \text{si } h \leq a \\ C & \text{en otro caso} \end{cases}$
Exponencial	$\gamma(h) = C \left[1 - \exp\left(-\frac{ h }{a}\right) \right]$
Gaussiano	$\gamma(h) = C \left[1 - \exp\left(-\frac{ h ^2}{a^2}\right) \right]$
Potencia	$\gamma(h) = \omega h ^\theta \text{ con } 0 \leq \theta \leq 2$

Los modelos de efecto pepita y esférico alcanzan su meseta cuando $h > 0$ y cuando $h = a$ respectivamente, mientras que los modelos exponencial y gaussiano llegan a ella asintóticamente, por lo que su rango práctico es obtenido cuando la meseta llega al 95% de su valor asintótico [15]. El modelo potencia crece indefinidamente, por lo que no posee meseta ni covarianza.

Variograma cruzado

Así como el valor de cierta propiedad puede depender estadísticamente de aquellos valores aledaños que representen la misma propiedad, este puede estar también relacionado espacialmente con valores de otras propiedades [18]. Las variables co-regionalizadas son aquellas que dependen espacialmente entre ellas. Análogamente al caso de una sola variable, los variogramas cruzados representan la relación espacial entre dos variables. Luego para cualquier par de variables u y v existe un semi variograma cruzado, que por simplicidad se llamará variograma cruzado, y se define según la ecuación (5).

$$\gamma_{uv}(h) = \frac{1}{2} E\{[Z_u(x+h) - Z_u(x)][Z_v(x+h) - Z_v(x)]\}, \quad (5)$$

donde Z_u y Z_v son los valores de u y v en los lugares x y $x+h$. Cabe notar que si $u = v$ se obtiene la definición de variograma para una sola variable. Por lo general se asume que las relaciones entre las covariables son simétricas (relación entre u y v es igual a la relación entre v y u), pero en la realidad son levemente diferentes $C_{uv}(h) \neq C_{vu}(h)$, correspondiendo estas diferencias a errores de muestreo [18]. Aun así, en la práctica se consideran idénticas.

El variograma cruzado se estima de una manera análoga al del variograma de una sola variable. Luego el variograma cruzado se define según la ecuación (6).

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [z_u(x_\alpha) - z_u(x_\beta)][z_v(x_\alpha) - z_v(x_\beta)], \quad (6)$$

donde $N(h) = \{(\alpha, \beta) \mid x_\alpha - x_\beta = h\}$ y $|N(h)|$ es el número de pares contenidos en $N(h)$.

Finalmente, el variograma cruzado se interpreta de manera similar al variograma univariable y puede ser modelado bajo los mismos criterios y modelos básicos, además de poseer las mismas características (pepa, meseta, alcance).

6.2.4 Estimación por inverso de la distancia

Los algoritmos de estimación por inverso de la distancia son una familia de métodos de medias ponderadas. Generan como resultado estimados que son versiones suavizadas de los datos originales. Es un método sencillo de aplicar y tiene por ventaja su velocidad de cálculo y que no requiere información previa (variograma) [19]. El método se basa en calcular pesos para las variables medidas basado en la distancia de las muestras al punto de interés y que estén dentro de un vecindario de búsqueda. Se calcula según la ecuación (7).

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i z(x_i), \quad (7)$$

donde x_0 es el punto a estimar, x_i , $i = 1, \dots, n$ son las ubicaciones de las muestras dentro del vecindario de x_0 , n es la cantidad de datos dentro del vecindario y λ_i , $i = 1, \dots, n$ son los pesos asignados a cada punto conocido.

Los pesos λ_i , se determinan según la ecuación (8).

$$\lambda_i = \frac{\frac{1}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}} \quad (8)$$

$$\sum_{i=1}^n \lambda_i = 1,$$

donde d_i corresponde a la distancia euclídea del punto x_i al punto x_0 y p corresponde al exponente de la distancia. La suma de los pesos debe ser igual a uno para asegurar que el método sea insesgado [1]. Cuando el exponente tiende a cero, la distancia a todos los datos se aproxima a uno, por lo que el estimador asigna un valor cercano a la media del vecindario al punto x_0 . Por otro lado, cuando el exponente crece, el inverso de la distancia disminuye, por lo que el método se transforma en el del vecino más cercano, ya que la distancia más pequeña se vuelve preponderante por sobre las demás [12]. El valor más común de p es dos; luego ese estimador se llama inverso de la distancia al cuadrado [19].

A pesar de su simplicidad, los métodos de estimación por inverso de la distancia tienden a ser sensibles a las características intrínsecas de la base de datos, a la cantidad de vecinos con los que se estima y al exponente de la distancia empleado [19].

6.2.5 Estimación por Kriging

La estimación local busca predecir el valor de la variable regionalizada en un sector del campo estudiado a partir de los datos disponibles [12]. La existencia de un modelo de dependencia espacial permite estimar el valor de un atributo en una zona donde no se realizó muestreo [15]. Por lo general, cerca de una billonésima parte del volumen del yacimiento ha sido muestreado, por lo que es sumamente necesario realizar una estimación robusta que tome en cuenta la variabilidad geológica y las principales características del depósito [1].

Kriging es el nombre que se le asignó a la familia generalizada de algoritmos de regresiones por mínimos cuadrados en reconocimiento al trabajo pionero de Danie G. Krige en base a sus observaciones de las características y cualidades de la minería del oro en Sudáfrica, que son los cimientos en los que Georges Matheron se basó para establecer las definiciones que rigen a la geoestadística.

Restricciones del kriging

La resolución del kriging se articula siempre en base a las mismas etapas. Las variantes del kriging se diferencian en hipótesis realizadas sobre la función aleatoria $Z = \{Z(x) \in D\}$, que representa la variable regionalizada.

- **Restricción de linealidad:** el estimador debe ser una combinación lineal ponderada de los datos, ecuación (9).

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}), \quad (9)$$

donde x_0 es el lugar a estimar, $\{x_{\alpha}, \alpha = 1, \dots, n\}$ son los sitios con datos, mientras que los ponderadores $\{\lambda_{\alpha}, \alpha = 1, \dots, n\}$ y el coeficiente a son las incógnitas del problema de kriging.

- **Restricción de insesgo:** el error de la estimación tiene esperanza nula, ecuación (10).

$$E[Z^*(x_0) - Z(x_0)] = 0 \quad (10)$$

Esta restricción no garantiza errores bajos, sino que sólo su media global es aproximadamente nula.

- **Restricción de optimalidad:** consiste en buscar los ponderadores que minimizan la varianza del error de estimación, lo cual equivale a minimizar el error cuadrático promedio, ecuación (11).

$$\min_{\lambda_{\alpha}} \text{Var}[Z^*(x_0) - Z(x_0)] \quad (11)$$

Vecindad del kriging

Por lo general los problemas de estimación no emplean todos los datos disponibles. Se define vecindad del kriging como el dominio del campo que contiene a los datos que son empleados en la estimación y el punto estimado. Existen 2 tipos de vecindades:

- **Vecindad única:** se habla de vecindad única cuando se emplean todos los datos para estimar el sitio en cuestión. Incluso los datos muy alejados influyen en la estimación. Estimar un sitio con todos los datos es ineficiente ya que aumenta considerablemente los tiempos de cálculo del kriging, siendo este un algoritmo con complejidad $O(N^3)$ [20].

- **Vecindad móvil:** se habla de vecindad móvil cuando solo se emplean datos cercanos al sitio en cuestión para estimar su valor. La forma y tamaño de la vecindad varían. La forma se trata de acoplar a las anisotropías presentes en la base de datos. Mientras que el tamaño depende de la precisión que se quiere tener al estimar y del tiempo de cálculo disponible para ello.

Kriging simple

El kriging simple se basa en la asunción de que la media $E[Z(x_0)] = m(x_0) = m$, es decir, que la media es estacionaria. Luego, bajo esta suposición, se debe comprobar que las restricciones del kriging son válidas.

- **Linealidad:** se asegura esta restricción ya que no depende de la media.
- **Insesgo:** asumiendo media m conocida, se puede deducir una ecuación que relacione el coeficiente a y los ponderadores λ_α . Reemplazando la ecuación (9) en la ecuación (12), y a través de la asunción del kriging simple se logra la expresión de la ecuación (13).

$$E[Z^*(x_0) - Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_\alpha E[Z(x_\alpha)] - E[Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_\alpha m - m \quad (12)$$

$$= a + m \left(\sum_{\alpha=1}^n \lambda_\alpha - 1 \right) = 0$$

$$\Rightarrow a = m \left(1 - \sum_{\alpha=1}^n \lambda_\alpha \right) \quad (13)$$

- **Optimalidad:** se calcula la varianza del error de estimación, empleando las propiedades de la varianza, obteniendo la ecuación (14).

$$\begin{aligned}
\text{Var}[Z^*(x_0) - Z(x_0)] &= \text{Var}\left[\sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha})\right] + \text{Var}[Z(x_0)] - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \text{Cov}(Z(x_{\alpha}), Z(x_0)) \\
&= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \text{Cov}(Z(x_{\alpha}), Z(x_{\beta})) + \text{Cov}(Z(x_0), Z(x_0)) \\
&\quad - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \text{Cov}(Z(x_{\alpha}), Z(x_0)) \\
&= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + C(0) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0)
\end{aligned} \tag{14}$$

Luego se procede a minimizar, diferenciando la expresión obtenida anteriormente con respecto a λ_{α} , obteniendo el sistema de ecuaciones (15), el cual escrito matricialmente corresponde a el sistema matricial (16).

$$\sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha} - x_{\beta}) = C(x_{\alpha} - x_0) \quad \forall \alpha = 1, \dots, n \tag{15}$$

$$\overbrace{\begin{pmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ C(x_n - x_1) & \cdots & C(x_n - x_n) \end{pmatrix}}^A \overbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}}^{\lambda} = \overbrace{\begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \end{pmatrix}}^B \tag{16}$$

La solución a este sistema es única si y solo si la matriz A es definida positiva, y por lo tanto invertible, luego los ponderadores son $\lambda = A^{-1}B$.

Finalmente, la varianza del kriging simple se calcula reemplazando la ecuación (15) en la ecuación (14), generando la ecuación (17), siendo el valor de la varianza del kriging siempre menor que la varianza a priori.

$$\text{Var}[Z^*(x_0) - Z(x_0)] = C(0) - \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0) \tag{17}$$

Kriging ordinario

El kriging ordinario se basa en la asunción de que la media $E[Z(x_0)] = m(x_0)$, es decir, que la media varía con respecto al punto a estimar. Luego, bajo esta suposición, se debe comprobar que las restricciones del kriging son válidas.

- **Linealidad:** se asegura esta restricción ya que no depende de la media.
- **Insesgo:** asumiendo media $m(x_0) = m$ desconocida, se pueden deducir ecuaciones para el coeficiente a y los ponderadores λ_α . Reemplazando la ecuación (9) en la ecuación (18), y a través de la asunción del kriging ordinario se logran las soluciones de la ecuación (19).

$$E[Z^*(x_0) - Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_\alpha E[Z(x_\alpha)] - E[Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_\alpha m - m \quad (18)$$

$$= a + m \left(\sum_{\alpha=1}^n \lambda_\alpha - 1 \right) = 0$$

$$a = m \left(1 - \sum_{\alpha=1}^n \lambda_\alpha \right) \Rightarrow a = 0 \wedge \sum_{\alpha=1}^n \lambda_\alpha = 1 \quad (19)$$

- **Optimalidad:** al igual que en el caso del kriging simple, se calcula la varianza del error de estimación, empleando las propiedades de la varianza, obteniendo la misma ecuación(14).

Luego se procede a minimizar empleando la condición de insesgo de la ecuación (19), pero, a diferencia del kriging simple, se debe emplear un multiplicador de Lagrange 2μ (por temas de simplificación), ya que el sistema posee $n + 1$ variables y n ecuaciones.

Diferenciando el lagrangiano de la ecuación (20) con respecto a λ_α y μ , se obtiene el sistema de ecuaciones (21), los cuales escritos matricialmente corresponden al sistema matricial (22).

$$L[Z^*(x_0) - Z(x_0)] = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + C(0) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0) + 2\mu \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) \quad (20)$$

$$\left\{ \begin{array}{l} \sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + \mu = C(x_{\alpha} - x_0) \quad \forall \alpha = 1, \dots, n, \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{array} \right. \quad (21)$$

$$\overbrace{\begin{pmatrix} C(x_1 - x_1) & \dots & C(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(x_n - x_1) & \dots & C(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}}^A \overbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix}}^{\lambda} = \overbrace{\begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \\ 1 \end{pmatrix}}^B \quad (22)$$

La solución a este sistema es única si y solo si la matriz A es definida positiva, y por lo tanto invertible, luego los ponderadores son $\lambda = A^{-1}B$.

Finalmente, la varianza del kriging ordinario se calcula reemplazando la ecuación (21) en la ecuación (20), generando la ecuación (23). Notar que no solo la varianza para el caso del kriging ordinario puede ser expresada en función del variograma, sino que también el problema completo, lo que implica que se puede trabajar independiente de la covarianza, permitiendo emplear modelos variográficos sin alcance ni meseta definidos [15].

$$Var[Z^*(x_0) - Z(x_0)] = C(0) - \mu - \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha} - x_0) = \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(x_{\alpha} - x_0) - \mu \quad (23)$$

6.2.6 Estimación por Co-Kriging

El método de predicción espacial co-kriging consiste en predecir una variable basándose en su información y en la de algunas variables auxiliares que estén correlacionadas espacialmente con ella [21]. De manera similar al kriging y bajo las mismas restricciones, el co-kriging es una generalización multivariable del kriging donde además de emplear el variograma como herramienta

de análisis de variabilidad geológica, es necesario utilizar también el variograma cruzado para analizar la variabilidad geológica empleando la dependencia de los valores a estimar con otras variables correlacionadas dentro del campo de estudio [22].

Co-Kriging simple

Sea un conjunto de variables regionalizadas, modeladas por funciones aleatorias Z_1, \dots, Z_N conjuntamente estacionarias, de medias $\{m_i : i = 1, \dots, N\}$ y covarianzas directas y cruzadas $\{C_{ij}(h) : i, j = 1, \dots, N\}$ conocidas de manera similar al kriging simple [23]. Luego, bajo estas suposiciones se debe comprobar que las restricciones del co-kriging son válidas.

- **Linealidad:** el estimador de co-kriging corresponde a una combinación lineal de los datos disponibles en la vecindad de x_0 , ecuación (24), donde se desea estimar la variable Z_{i_0} .

$$Z_1^*(x_0) = a_{i_0} + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha, i, i_0} Z_i(x_\alpha^i), \quad (24)$$

donde x_α^i representa al α -ésimo punto con dato de la i -ésima variable (Z_i) ubicado dentro de la vecindad de co-kriging. El coeficiente a y los ponderadores $\{\lambda_{\alpha, i, i_0}, i = 1, \dots, N, \alpha = 1, \dots, n_i\}$ son las incógnitas al problema.

- **Insesgo:** dada la asunción del co-kriging simple, se exige que el error de estimación, ecuación (25), tenga esperanza nula.

$$\begin{aligned} E[Z_{i_0}^*(x_0) - Z_{i_0}(x_0)] &= a_{i_0} + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha, i, i_0} E[Z_i(x_\alpha^i)] - E[Z_{i_0}(x_0)] \\ &= a_{i_0} + \left[\sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha, i_0, i_0} - 1 \right] m_{i_0} + \sum_{\substack{i=1 \\ i \neq i_0}}^N \left[m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha, i, i_0} \right], \end{aligned} \quad (25)$$

lo que conduce a la ecuación (26).

$$a_{i_0} = \left[1 - \sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha, i_0, i_0} \right] m_{i_0} + \sum_{\substack{i=1 \\ i \neq i_0}}^N \left[m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha, i, i_0} \right] \quad (26)$$

- **Optimalidad:** se calcula la varianza del error de estimación, empleando las propiedades de la varianza, obteniendo la ecuación (27).

$$\begin{aligned} \text{Var}[Z^*(x_0) - Z(x_0)] &= \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_{\alpha, i, i_0} \lambda_{\beta, j, i_0} C_{ij} (x_\alpha^i - x_\beta^j) \\ &\quad - 2 \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha, i, i_0} C_{ii_0} (x_\alpha^i - x_0) + C_{i_0 i_0} (0) \end{aligned} \quad (27)$$

Luego se minimiza la varianza del error, diferenciando con respecto a λ_{α, i, i_0} e igualando a cero, obteniéndose el sistema de ecuaciones (28).

$$\sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta, j, i_0} C_{ij} (x_\alpha^i - x_\beta^j) = C_{ii_0} (x_\alpha^i - x_0) \quad \forall i = 1, \dots, N, \forall \alpha = 1, \dots, n_i \quad (28)$$

Finalmente, la varianza de co-kriging simple se determina reemplazando la ecuación (28) en la ecuación (27), generando la ecuación (29), siendo la varianza de estimación menor que la varianza a priori.

$$\sigma_{CKS-Z_{i_0}}^2(x_0) = C_{i_0 i_0} (0) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{i, i_0} C_{ii_0} (x_\alpha^i - x_0) \quad (29)$$

Co-Kriging ordinario

Sea un conjunto de variables regionalizadas, modeladas por funciones aleatorias Z_1, \dots, Z_N conjuntamente estacionarias, de medias $\{m_i : i = 1, \dots, N\}$ desconocidas y, por ende, dependientes de la vecindad de estimación, además de covarianzas directas y cruzadas $\{C_{ij}(h) : i, j = 1, \dots, N\}$ conocidas de manera similar al kriging ordinario [23]. Luego, bajo estas suposiciones se debe comprobar que las restricciones del co-kriging son válidas.

- **Linealidad:** el estimador de co-kriging corresponde a una combinación lineal de los datos disponibles en la vecindad de x_0 , ecuación (30), donde se desea estimar la variable Z_{i_0} .

$$Z_1^*(x_0) = a_{i_0} + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} Z_i(x_\alpha^i), \quad (30)$$

donde x_α^i representa al α -ésimo punto con dato de la i -ésima variable (Z_i) ubicado dentro de la vecindad de co-kriging. El coeficiente a y los ponderadores $\{\lambda_{\alpha,i,i_0}, i = 1, \dots, N, \alpha = 1, \dots, n_i\}$ son las incógnitas al problema.

- **Insesgo:** dada la asunción del co-kriging ordinario, se exige que el error de estimación, ecuación (31), tenga esperanza nula.

$$\begin{aligned} E[Z_{i_0}^*(x_0) - Z_{i_0}(x_0)] &= a_{i_0} + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} E[Z_i(x_\alpha^i)] - E[Z_{i_0}(x_0)] \\ &= a_{i_0} + \left[\sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha,i_0,i_0} - 1 \right] m_{i_0} + \sum_{\substack{i=1 \\ i \neq i_0}}^N \left[m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} \right], \end{aligned} \quad (31)$$

lo que conduce a las restricciones de la ecuación (32), donde se fija que los pesos de la variable principal sumen uno y que los pesos de todas las variables secundarias sumen cero, además de que $a_{i_0} = 0$. Esta versión del co-kriging recibe el nombre de co-kriging ordinario tradicional [1].

$$\begin{aligned} &\left[1 - \sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha,i_0,i_0} \right] m_{i_0} + \sum_{\substack{i=1 \\ i \neq i_0}}^N \left[m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} \right] = 0 \\ \Rightarrow &\begin{cases} \sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha,i_0,i_0} = 1 \\ \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} = 0 \quad \forall i = 1, \dots, N, i \neq i_0 \end{cases} \end{aligned} \quad (32)$$

- **Optimalidad:** se calcula la varianza del error de estimación, empleando las propiedades de la varianza, obteniendo la ecuación (27).

Similar al caso del kriging ordinario, se emplean multiplicadores de Lagrange para suplir la deficiencia de ecuaciones en comparación a las incógnitas, generando el lagrangiano [15] de la ecuación (33)

$$\begin{aligned}
 L[Z^*(x_0) - Z(x_0)] = & \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_{\alpha,i,i_0} \lambda_{\beta,j,i_0} C_{ij} (x_{\alpha}^i - x_{\beta}^j) \\
 & - 2 \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} C_{ii_0} (x_{\alpha}^i - x_0) + C_{i_0 i_0}(0) \\
 & + 2\mu_{i_0,i_0} \left[\sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha,i_0,i_0} - 1 \right] + 2 \sum_{\substack{i=1 \\ i \neq i_0}}^N \mu_{i,i_0} \left[\sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} \right]
 \end{aligned} \tag{33}$$

Luego se minimiza el lagrangiano, diferenciando con respecto a λ_{α,i,i_0} y μ_i e igualando a cero, obteniéndose el sistema de ecuaciones (34).

$$\left\{ \begin{array}{l} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta,j,i_0} C_{ij} (x_{\alpha}^i - x_{\beta}^j) + \mu_{i,i_0} = C_{ii_0} (x_{\alpha}^i - x_0) \quad \forall i = 1, \dots, N, \forall \alpha = 1, \dots, n_i \\ \sum_{\alpha=1}^{n_i} \lambda_{\alpha,i,i_0} = \delta_{i,i_0} \quad \forall i = 1, \dots, N, \end{array} \right. \tag{34}$$

donde $\delta_{i,i_0} = 1$ cuando $i = i_0$ y $\delta_{i,i_0} = 0$ para todo el resto de los casos.

Finalmente, la varianza de co-kriging ordinario se determina reemplazando el sistema de ecuaciones (34) en la ecuación (33), generando la ecuación (35).

$$\sigma_{CKO-Z_{i_0}}^2(x_0) = C_{i_0 i_0}(0) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{i,i_0} C_{ii_0} (x_{\alpha}^i - x_0) - \mu_{i_0,i_0} \tag{35}$$

Similar al kriging ordinario, se puede plantear el problema en función de los variogramas, permitiendo emplear modelos variográficos sin meseta ni alcances definidos.

6.2.7 Estimación por Kriging Indicador

El kriging indicador, a diferencia del clásico kriging simple y ordinario, comienza la estimación a través de la determinación de intervalos con diferentes valores de corte K , generando $(K + 1)$ intervalos que posean frecuencias idealmente similares [15]. Estos límites permiten la transformación de datos continuos a variables booleanas. Esto soluciona dos principales problemas de la estimación clásica: la normalidad asumida en los kriging tradicionales y la independencia de los errores del kriging con respecto al valor de los datos con los que se estima [24]. También producto de la discretización se puede reducir el impacto de los valores extremos de la variable de estudio y permite realizar estimación con variables categóricas.

La transformación de una variable $z(x)$ en un indicador para los K valores de corte z_k se resume en la ecuación (36), la cual representa una transformación para datos duros [15] y transforma el campo de estudio en zonas de ceros y unos para cada valor de corte, Figura 7, generando K problemas de kriging que requieren de K soluciones y por lo tanto K variogramas. Sin embargo, existe una simplificación llamada kriging indicador mediano [25], el cuál emplea solamente un variograma que represente al resto cuando los distintos variogramas sean proporcionales o similares para todos los K problemas, reduciendo el tiempo de cálculo y facilitando el uso de los mismos ponderadores de kriging para cada K problema.

$$I(x, Z_k) = \begin{cases} 1 & \text{si } z(x) \leq z_k \\ 0 & \text{si } z(x) > z_k \end{cases} \quad (36)$$

Si $z(x)$ corresponde a una variable categórica binaria, por ejemplo, que valga uno cuando cierto tipo de roca prepondere sobre el resto en x y cero cuando no, el kriging de $z(x)$ entrega el modelo de probabilidades que dicho tipo de roca prepondere sobre cierto lugar x , es decir que la variable valga uno [25].

Luego del proceso de discretización, se procede a realizar el algoritmo de kriging donde el estimador de kriging corresponde a la combinación lineal entre las variables indicadoras, ecuación (37).

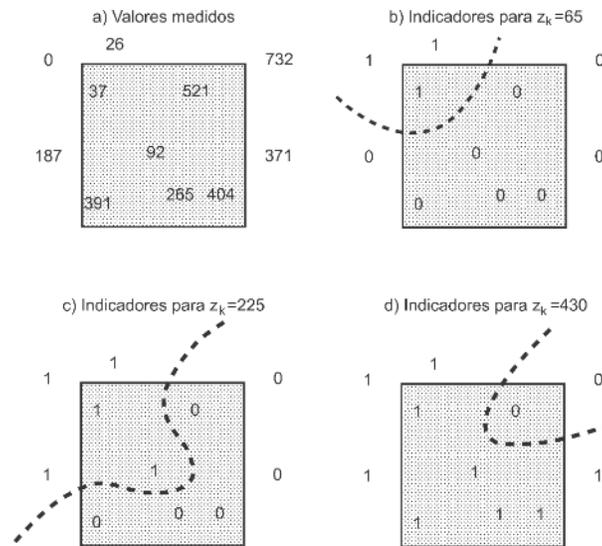


Figura 7. Ejemplo de transformación de datos a indicadores

$$[I(x, z_k)]^* - E[I(x, z_k)] = \sum_{\alpha=1}^{n(x)} \lambda_{\alpha}(x, z_k) [I(x_{\alpha}, z_k) - E[I(x_{\alpha}, z_k)]] \quad (37)$$

Al igual que en el caso del kriging univariable, este puede ser simple u ordinario, siendo la media de los indicadores conocida y constante para el primer caso y variable en el segundo, limitada por la vecindad, generando ecuaciones análogas a las ecuaciones (15) y (21), pero empleando los indicadores y generando K sistemas de ecuaciones para cada problema.

6.2.8 Validación cruzada

Al escoger cualquier tipo de estimador y sus parámetros, existe la inquietud y necesidad de conocer la calidad de los datos estimados. La mejor respuesta sería realizar un muestreo exhaustivo de las zonas donde fue estimada la o las variables, comparando los valores reales con los valores estimados [17]. Sin embargo, realizar un muestreo completo del campo de estudio además de impráctico, es imposible e infinitamente costoso, debido a la basta cantidad de información a obtener y al tiempo que se invierte en obtenerla.

La validación cruzada se emplea para encontrar el mejor modelo variográfico dentro de un grupo finito de candidatos [26]. El también llamado algoritmo de “dejar uno afuera” (“*leaving one out*”, en inglés), consiste en, dada n observaciones, emplear las $n - 1$ observaciones restantes para, por medio del kriging, estimar el valor de la observación apartada y compararla con el valor medido con

el fin de conocer la calidad de la estimación. Este proceso se lleva a cabo según el siguiente algoritmo:

- Elegir un modelo variográfico
- Dejar un valor fuera de la base de datos y estimarlo a partir del resto de datos, empleando una técnica de kriging con el modelo variográfico escogido
- Repetir el paso anterior para todos los datos y calcular la diferencia entre el valor real y el valor estimado
- Dividir cada diferencia al cuadrado por la varianza de kriging de la estimación
- Calcular la media y la desviación estándar de estas diferencias
- Repetir todo lo anterior para cada modelo variográfico
- Elegir el modelo variográfico que cause que la media de la diferencia y la desviación estándar estén lo más cerca de cero y uno respectivamente. Esto último garantiza que el variograma cuantifica adecuadamente la amplitud de los errores, es decir que no es ni demasiado optimista ni tampoco pesimista [12].

La validación cruzada no indica si una asunción, parámetro o estimador es correcto. Sólo genera errores asociados con las diferentes elecciones, por lo que la validación cruzada es una herramienta que entrega resultados que deben ser interpretados por y para que el usuario lleve a cabo sus propias conclusiones [17].

Usualmente, se presenta la validación cruzada en forma de gráficos, como la nube de correlación o el histograma de los errores estandarizados [12].

6.2.9 Categorización de recursos

La Figura 8 ilustra el marco de referencia para clasificar los resultados de exploración. Estos pueden ser clasificados en recursos o reservas minerales. Un recurso mineral puede ser catalogado como medido, indicado o inferido, mientras que las reservas minerales se clasifican en probados o

probables. Estas clasificaciones reflejan diferentes niveles de conocimiento sobre la geología de los recursos y diferente grado de evaluación técnica y económica [27]. La elección de la categoría adecuada de recurso o reserva mineral debe ser realizada por una persona competente.

Una persona competente es un individuo que pertenece a una sociedad profesional de científicos de la tierra o ingenieros de minas, o posee calificaciones apropiadas para supervisar la estimación de un depósito.

Un recurso mineral puede ser estimado principalmente a través de conocimientos geostatísticos con alguna que otra ayuda de disciplinas externas, mientras que un recurso mineral, en cambio, requiere la consideración de todos los factores relacionados con la extracción, procesamiento y venta posterior para poder ser clasificados en probados o probables y requiere por lo tanto amplio conocimiento de un rango variado de disciplinas.

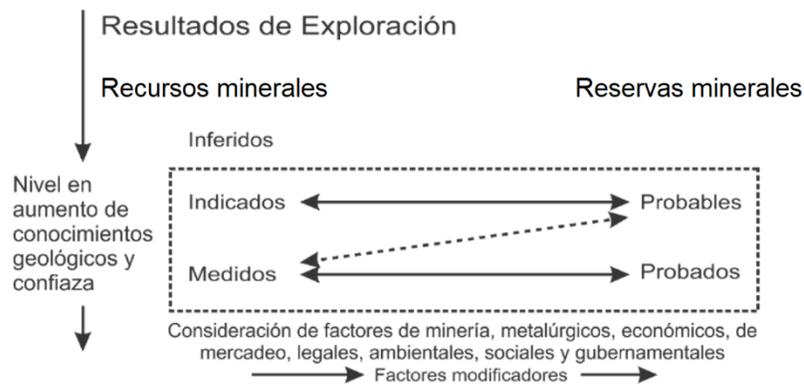


Figura 8. Relación entre recursos minerales y recursos de mena

Figura 8. Relación entre recursos y reservas minerales

En cuanto a la clasificación de recursos minerales, el procedimiento para determinar a qué grupo va cada bloque es completamente subjetivo al entorno geológico y a la experiencia de la persona competente. Entre los criterios para clasificar existen principalmente tres tipos: basados en la geometría alrededor de cierto vecindario, basados en simulaciones y basados en la varianza de kriging.

Se pueden determinar a qué grupo pertenece cada bloque a través del criterio desarrollado por Diehl y David [28], el cual define niveles de incertidumbre (confianza) y precisiones (tolerancias

del error) para cada caso, Tabla 2. La precisión, que se mide en base a la tolerancia del error, emplea la desviación estándar del kriging relativa, ecuación (38).

Tabla 2. Clasificación de recursos

	Identificado			Sin descubrir		
	Demostrado					
	Medido		Indicado			
	Probado	Probable	Probable	Inferido	Hipotético	Especulativo
Tolerancia error	±10%	±20%	±40%	±60%		
Confianza	> 80%	60 – 80%	40 – 60%	20 – 40%	10 – 20%	< 10%
Recursos económicamente significantes						

$$Precisión (error) = \frac{\sigma(x_k) * 100 * \mu_i}{z(x_k)} \quad (38)$$

donde μ_i corresponde al valor de la distribución normal estándar inversa (de dos colas) con probabilidad i , la cual se relaciona directamente con el grado de confianza para cada clasificación.

6.2.10 Comentarios generales

El kriging es una técnica que entrega para una ubicación desconocida la mejor estimación lineal insesgada y su correspondiente varianza. Además de ser un interpolador lineal exacto, es un interpolador óptimo estadísticamente hablando, en el sentido de que minimiza la varianza de estimación cuando el variograma es conocido y se emplea la asunción de estacionalidad [19].

La principal ventaja de la interpolación por inverso de la distancia frente al kriging es que la primera corresponde a una técnica que además de ser más sencilla, no requiere un estudio variográfico a priori (que en sí corresponde a un proceso interpretativo en que el conocimiento de la persona que estima cumple un rol importante en la generación de este) ya que sólo emplea la distancia de los datos relativo al punto que se desea estimar. Luego, al no requerir variograma es empleado en los casos en que este no puede ser definido producto de la deficiencia de datos o en los casos de variogramas demasiado erráticos. Por lo general se usa el inverso de la distancia para tener una idea base de la distribución de la variable a estudiar con un costo computacional menor en relación al kriging.

El kriging, a pesar de ser más tedioso en su aplicación que el inverso de la distancia, posee como principal ventaja la incorporación del variograma, que entrega información sobre la distribución espacial de los datos muestreados, lo que implica que se conocen cualidades como la anisotropía y la correlación espacial. Además, el kriging puede estimar datos fuera de los límites de valores que son muestreados, ya que no corresponde a una interpolación [29], y no sufre problemas producto de un muestreo irregular ya que si existiese algún tipo de agrupamiento, la covarianza entre muestras automáticamente divide la influencia de este clúster entre sus constituyentes evitando la sobrerepresentación de estos datos [30].

Por último, no existe un método efectivo de medir el error de estimación en el método de inverso de la distancia para cada valor individualmente, por lo que no se puede evaluar la confianza de los estimados y, por ende, no se puede determinar en qué áreas existe mayor incertidumbre [19].



6.3 Extensión de dominios geometalúrgicos

6.3.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) es una herramienta predictiva de clasificación que emplea la teoría aprendizaje automatizado (“*machine learning*”) para maximizar la precisión predictiva del algoritmo mientras automáticamente evita sobre ajustar la información [31]. Varios estudios han demostrado que los SVM generalmente son capaces de entregar mejores resultados en cuanto a la precisión en la clasificación de datos que otros algoritmos de clasificación convencionales e incluso otros algoritmos supervisados [32]. Este método de clasificación ha sido empleado en varios problemas del mundo real como categorización de texto, clasificación de imágenes, detección de objetos, clasificación de datos, entre otros. En un principio, los SVM fueron desarrollados para realizar clasificación y generar límites entre distintas clases de datos, pero han sido adaptados para problemas de regresión [33].

Clasificación binaria de clases linealmente separables

Dado un conjunto separable de datos $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$, donde $x_i \in \mathbb{R}^d$ e $y_i \in \{-1, +1\}$, es posible definir un hiperplano de separación como una función lineal que es capaz de separar dicho conjunto sin error, ecuación (39).

$$D(x) = (w_1x_1 + \dots + w_dx_d) + b = \langle w, x \rangle + b, \quad (39)$$

donde w es la normal del hiperplano, $\langle w, x \rangle$ representa el producto escalar entre w y x , $-b/|w|$ es la distancia perpendicular del hiperplano al origen (distancia mínima), Figura 9.

Los vectores de soporte son aquellos generados entre los datos más cercanos al hiperplano y perpendiculares al mismo, siendo el objetivo de los SVM orientar el hiperplano de manera de estar lo más alejado posible de los elementos más cercanos de cada clase al hiperplano [34].

El hiperplano cumple las restricciones impuestas por la ecuación (40), simplificadas en la ecuación (41), para todo x_i , con $i = 1, \dots, n$.

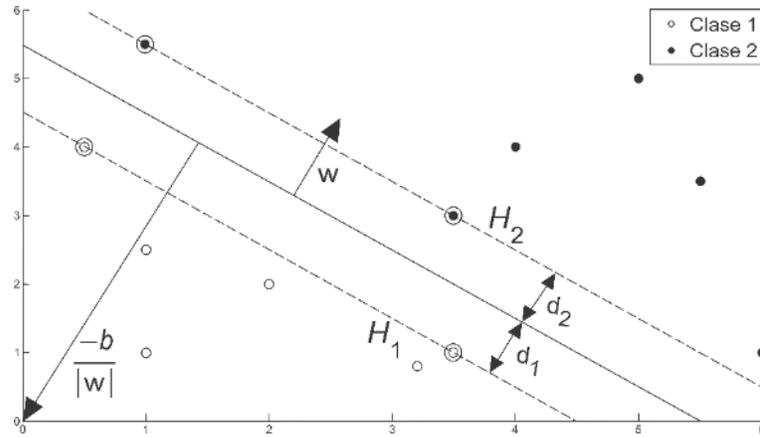


Figura 9. Hiperplano a través de dos clases linealmente separables

$$\begin{aligned} \langle w, x_i \rangle + b &\geq +1 \quad \text{para } y_i = +1 \\ \langle w, x_i \rangle + b &\leq -1 \quad \text{para } y_i = -1 \end{aligned} \quad (40)$$

$$y_i(\langle w, x \rangle + b) - 1 \geq 0 \quad \forall i \quad (41)$$

Definiendo, acorde a la Figura 9, d_1 y d_2 como la distancia de H_1 y H_2 al hiperplano respectivamente, la equidistancia del hiperplano a ambos puntos significa que $d_1 = d_2 = 1/\|w\|$, siendo esta cantidad conocida como el margen de SVM. Maximizar este margen es equivalente a minimizar $\|w\|$, que a su vez es equivalente a minimizar $1/2 \|w\|^2$, ecuación (42), el cual corresponde a un problema de programación cuadrática, abordable por la teoría de la optimización [35].

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \equiv \frac{1}{2} \langle w, w \rangle \\ \text{s. a.} \quad & y_i(\langle w, x_i \rangle + b) - 1 \geq 0, \quad i = 1, \dots, n \end{aligned} \quad (42)$$

Clasificación binaria de clases linealmente cuasi-separables

Muchas veces los problemas reales se caracterizan por poseer ruido y no ser perfectamente separables linealmente, es por ello que se incorporan errores que permiten relajar un poco la definición de separabilidad de las clases para permitir puntos clasificados de manera errónea.

Es por lo anterior que se añaden las variables de holgura positiva $\xi_i, i = 1, \dots, n$, Figura 10, que permiten implícitamente cuantificar el número de puntos no separables que se está dispuesto a admitir, ecuación (43), simplificado en la ecuación (44).

$$\begin{aligned} \langle w, x_i \rangle + b &\geq +1 - \xi_i \quad \text{para } y_i = +1 \\ \langle w, x_i \rangle + b &\leq -1 + \xi_i \quad \text{para } y_i = -1 \end{aligned} \quad (43)$$

$$(y_i \langle w, x \rangle + b) - 1 + \xi_i \geq 0, \text{ donde } \xi_i \geq 0 \quad \forall i \quad (44)$$

Luego, este nuevo margen un poco más permisivo posee una penalización sobre los puntos mal clasificados, que aumenta conforme crece la distancia del punto al hiperplano. Como el objetivo ya no es sólo maximizar el margen, ya que se podría lograr clasificando mal muchos puntos, la función objetivo a minimizar, ecuación (45), comprende la adición de esta penalización además de una constante C , conocida como parámetro de regularización, valor elegido por el usuario que permite controlar la cantidad de ejemplos mal clasificados. Valores de C muy grandes sobre ajustan el modelo al restringir la cantidad de datos mal clasificados, minimizando a su vez la cantidad de SVM que definen el hiperplano [35]. Mientras que valores muy pequeños arrojan demasiadas clasificaciones de datos a una clase no correspondiente, complejizando el modelo al añadir más vectores de soporte y, por ende, añadiendo datos mal clasificados.

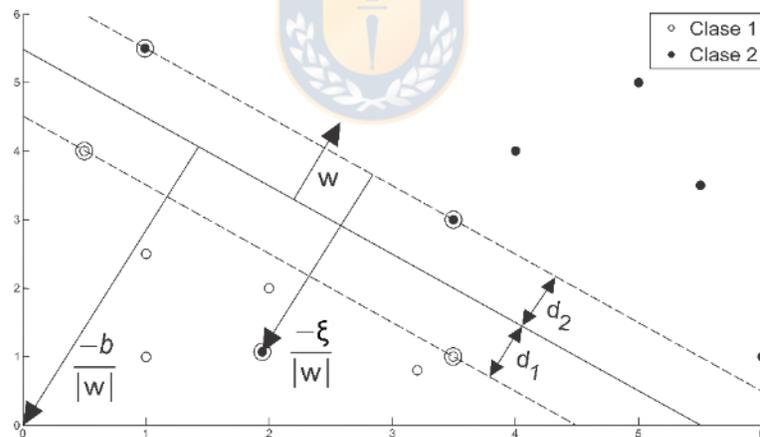


Figura 10. Hiperplano a través de dos clases linealmente no separables

$$\begin{aligned} \min \quad & \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^n \xi_i \\ \text{s. a.} \quad & y_i (\langle w, x_i \rangle + b) + \xi_i - 1 \geq 0 \\ & \xi_i \geq 0, i = 1, \dots, n \end{aligned} \quad (45)$$

Clasificación de datos linealmente no separables

Cuando los datos no son linealmente separables, la solución es emplear una función de transformación $\Phi: X \rightarrow \mathcal{F}$ para transformar, valga la redundancia, los valores del conjunto original, espacio de entradas X , en el espacio de características \mathcal{F} el cual posee varias dimensiones más que el original [35], con el fin de pasar de un problema no separable a uno linealmente separable en una dimensión mayor. Luego este hiperplano formado en el espacio de características es transformado de vuelta al espacio de entradas, Figura 11, obteniendo un hiperplano no lineal en el problema original.

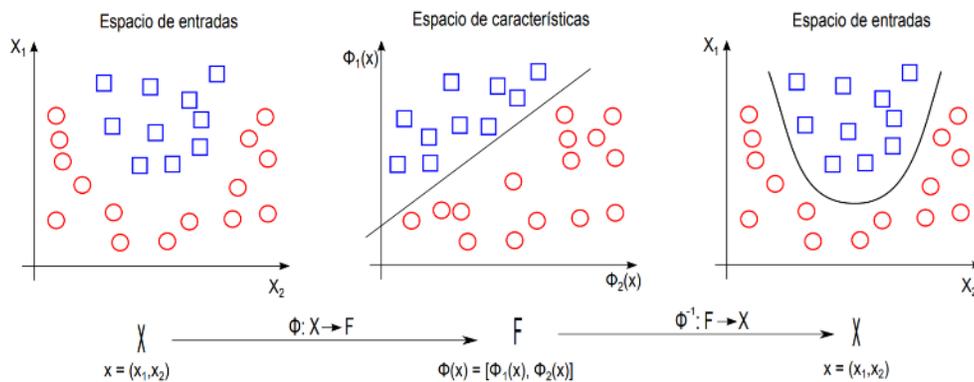


Figura 11. Solución de problemas linealmente no separables

Al transformar el problema cuadrático de la ecuación (45) en un problema lineal, empleando multiplicadores de Lagrange y algebra [35], el hiperplano de separación puede ser expresado según la ecuación (46), donde luego empleando la función de transformación Φ se obtiene la ecuación (47).

$$D(x) = \sum_{i=1}^n \alpha_i y_i \langle x, x_i \rangle + b \quad (46)$$

$$D(x) = \sum_{i=1}^n \alpha_i y_i \langle \Phi(x), \Phi(x_i) \rangle \quad (47)$$

El término $K(x, x_i) = \langle \Phi(x), \Phi(x_i) \rangle$ se conoce como función kernel, y, a través del teorema de Aronszajn [36], sólo se necesita conocer la forma de la función kernel y no es necesario conocer la

función de transformación ni las coordenadas de los objetos transformados en el espacio de características.

Existen muchas funciones de kernel, Tabla 3, pero la más empleada dentro del campo de la clasificación de datos es el kernel gaussiano o también llamado RBF (del inglés “*radial basis function*”) ya que este posee menos parámetros del kernel y menos dificultades computacionales.

Tabla 3. Funciones de kernel más conocidas

Función	Expresión matemática
Lineal	$K(x, x_i) = \langle x, x_i \rangle$
Polinomial	$K(x, x_i) = [\gamma \langle x, x_i \rangle + \tau]^p, \gamma > 0$
Gaussiano	$K(x, x_i) = \exp\{-\gamma \langle (x - x_i), (x - x_i) \rangle\}, \gamma > 0$
Sigmoidal	$K(x, x_i) = \tanh(\gamma \langle x, x_i \rangle + \tau), \gamma > 0$

La elección de un modelo de kernel es uno de los problemas de SVM. Los resultados de la clasificación de datos dependen de ajustar el parámetro C cuando se trabaja con datos cuasi-separables o separables linealmente. Por lo general los problemas no son separables linealmente, por lo que además se debe ajustar el parámetro γ . Usualmente este proceso se lleva a cabo a través de una búsqueda por cuadrantes, donde a través de validación cruzada se determinan los valores de estos hiperparámetros. El desempeño de SVM para algunas bases de datos es muy sensible a estas decisiones. Como resultado, se deben realizar múltiples validaciones cruzadas para determinar los parámetros óptimos. Este proceso es la llamada elección de modelo, siendo uno de los problemas prácticos de este proceso el tiempo que toma para ser llevado a cabo [32].

Clasificación de datos multiclase

El uso de SVM para clasificar datos también puede ser llevado a cabo para datos multiclase a través de clásicamente dos metodologías: OAA y OAO que significan “uno contra todos” y “uno contra uno” respectivamente.

La estrategia OAA (del inglés “*one against all*”) consiste en la construcción de k hiperplanos de separación entre la k -ésima clase y todo el resto de los objetos que no pertenezcan a la k -ésima clase, independiente de la clase a la que pertenezcan [37]. Luego x pertenece a la m -ésima clase que posea el mayor valor de función de decisión [38], ecuación (48).

$$f(x) = \operatorname{argmax}_{m=1,\dots,k} \left[\sum_{i=1}^n \lambda_i^m y_i K^m(x, x_i) + b^m \right] \quad (48)$$

Mientras que la estrategia OAO (del inglés “*one against one*”) consiste en la construcción de $k(k - 1)/2$ problemas, que corresponden a todos los hiperplanos entre pares de clases posibles. La asignación de x a la k -ésima clase se hace a través de una votación en la que sí, según la ecuación (49), x pertenece a la i -ésima clase, la i -ésima clase gana un “voto”, de lo contrario la j -ésima clase gana un voto. Luego que todas las combinaciones de clases han generado hiperplanos y el objeto x ha asignado votos a todas estas, la clase a la que pertenece el objeto x es aquella a la cual se le asignó más votos.

$$f(x) = \operatorname{sign} \left[\sum_{t=1}^n \lambda_t^{ij} y_t K^{ij}(x, x_t) + b^{ij} \right] \quad (49)$$

Estudios han demostrado que OAO es superior; primero porque posee menos tiempos de cálculo, ya que a pesar de realizar más iteraciones de SVM las clases comparadas son más pequeñas; segundo, en la práctica al realizar validaciones en los hiperplanos obtenidos, en OAO se obtienen mejores precisiones al enfrentar las nuevas clasificaciones de los datos de testeo con las clases de los datos originales [37]; y tercero, OAA al comparar una clase pequeña contra otra clase generalmente mucho más grande genera un efecto preferencial en el que tiende a asignar más objetos a la clase con un set más grande de testeo [39].

7 Caso Simulado

7.1 Metodología de simulación

El objetivo del caso simulado es comparar diversos métodos de estimación por kriging (kriging ordinario, co-kriging indicador ordinario, co-kriging ordinario) y el método del inverso de la distancia tanto para datos continuos como categóricos, con el fin comprobar la hipótesis del problema: es mejor estimar sobre dominios geometalúrgicos, donde cada dominio posee un comportamiento característico diferente a los demás, que estimar la información empleando los procedimientos habituales de estimación.

Luego, para lograr comprobar la hipótesis se llevan a cabo las siguientes etapas:

- Generación de zona de trabajo: se genera una zona de trabajo tridimensional lo suficientemente amplia como para representar un depósito hipotético, actuando las fronteras como los límites del volumen de estudio.
- Creación de dominios geometalúrgicos: se crean tres dominios geometalúrgicos hipotéticos donde serán realizadas las simulaciones.
- Simulación de variables: son simuladas tres variables dentro de cada dominio: leyes de cobre y oro y tipo de roca, siendo este último subdividido en 4 tipos de roca. Se realizan 100 simulaciones para cada zona, manteniéndose la geometría de los dominios geometalúrgicos idéntica para todas las iteraciones.
- Extracción de sondajes: se extrae una cantidad suficiente de sondajes perpendiculares al plano horizontal que actúa como superficie, siendo la información de estos sondajes empleada como base de entrenamiento para los algoritmos de estimación.
- Aplicación de algoritmos de estimación: se estima sobre las localidades donde no hay presencia de sondajes y se compara con la base de datos “real” simulada anteriormente, y evaluando el error cuadrático medio (para el caso de variables continuas) y la precisión (para el caso de variables categóricas) cuando se emplean y cuando no se emplean dominios geometalúrgicos.

- Evaluación estadística: se realiza una comparación entre las medidas estadísticas principales asociadas a las predicciones entre diferentes algoritmos de estimación con los datos simulados para el mejor, el peor y el resultado más probable con respecto a la diferencia de la raíz del error cuadrático medio de la estimación y la precisión entre emplear o no dominios geometalúrgicos. Se estudia variabilidad y medidas de incertidumbre de predicción.
- Categorización de recursos: se lleva a cabo una categorización de los recursos minerales en base a la incertidumbre obtenida por los diferentes algoritmos de estimación, siendo esto evaluado en el mejor, el peor y el caso promedio con respecto a la precisión de la estimación. Se reportan medidas estadísticas principales para cada caso.
- Análisis económico: en base a los resultados obtenidos de la clasificación de recursos y a diferentes escenarios de riesgo se lleva a cabo un análisis económico y es evaluada la diferencia de este con un análisis equivalente sobre la base de datos real.

7.2 Aplicación



La aplicación de los algoritmos discutidos en el marco teórico es llevada a cabo según lo planteado en la metodología de simulación sobre el caso simulado.

7.2.1 Generación de la zona de trabajo

La zona de trabajo es generada de manera arbitraria con dimensiones de 200 unidades en las direcciones norte y este, y 400 unidades de profundidad. Luego se genera una malla tridimensional en la que se ubican puntos cada 10 unidades en todas las direcciones, Figura 12, esto significa que hay 21x21 puntos en cada nivel, siendo 41 niveles en total, resultando en una base de datos con 18081 coordenadas. Las unidades de medida son asumidas como metros. Cada punto representa el centro de cada bloque de dimensiones $10 \times 10 \times 10 \text{ m}^3$ por lo que el volumen de la zona de trabajo es de $18.081 \times 10^6 \text{ m}^3$.

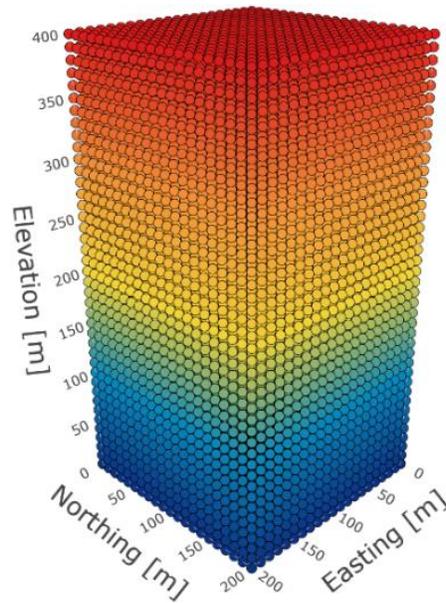


Figura 12. Visualización de zona de trabajo

7.2.2 Creación de dominios geometalúrgicos

Para generar los dominios geometalúrgicos se realiza una simulación gaussiana sobre la base de datos vacía. Esta simulación se hace a través de un variograma de tipo gaussiano (debido a la poca varianza que posee para distancias pequeñas), Figura 13, con meseta igual a 0.6, alcance igual a 100 metros y sin efecto pepa, por razones que se discuten más adelante. Además, se emplea una media igual a cero, valor que no influye en la categorización posterior.

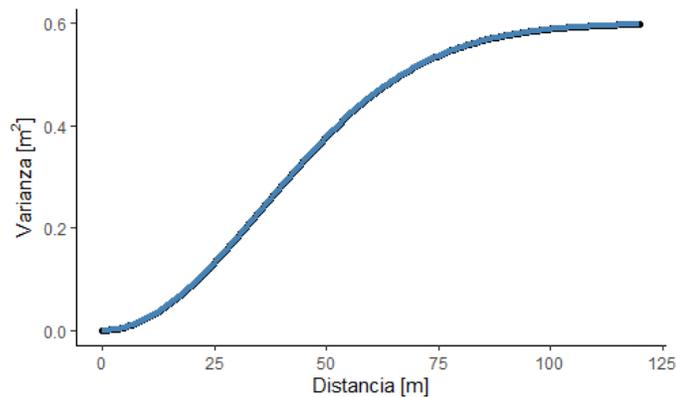


Figura 13. Variograma empleado en la simulación de los dominios

Debido a lo anterior, se genera dentro de toda la zona de trabajo una distribución normal gaussiana de media cero que, ya que no posee pepa, es muy regular dentro de los límites del depósito hipotético, Figura 14.

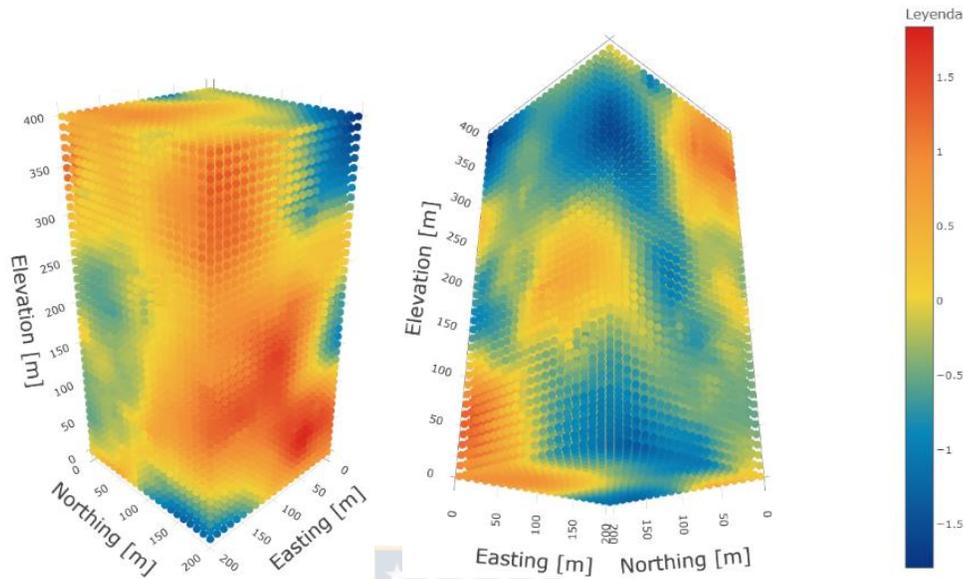


Figura 14. Distribución de la variable generada dentro de la zona de trabajo

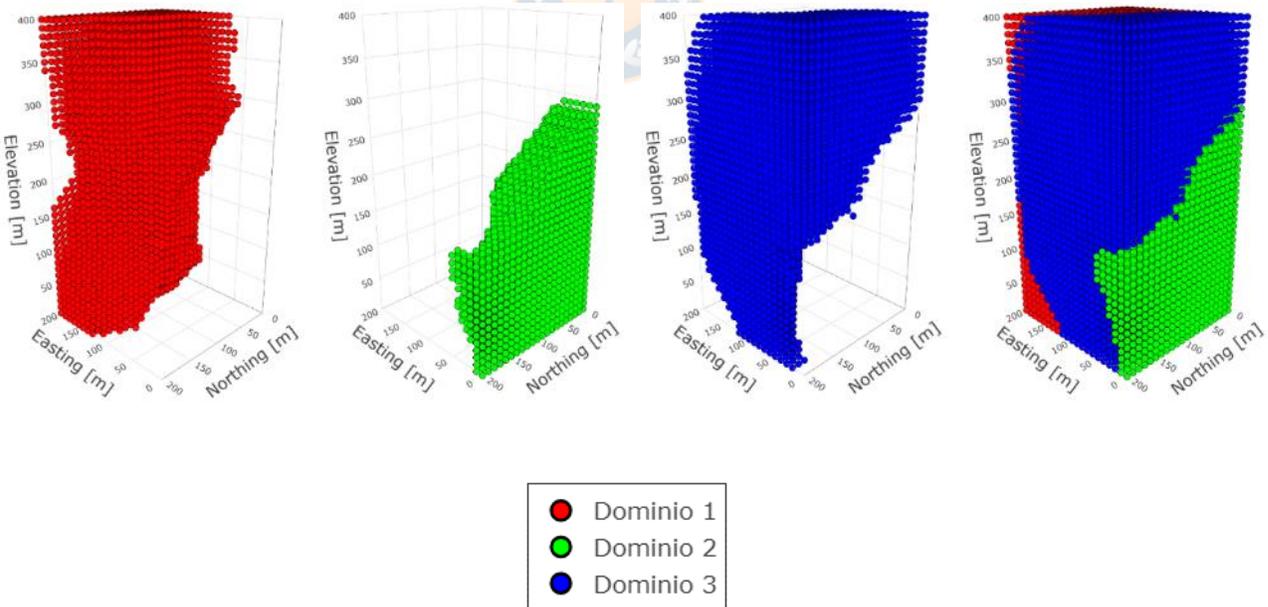


Figura 15. Dominios geometalúrgicos

Luego se realiza agrupamiento de datos en 3 grupos con k-means, donde se itera 100 veces empleando como variables de agrupamiento las coordenadas de cada punto y el valor entregado

por la simulación gaussiana, la que es necesario que emplee un variograma de tipo gaussiano y sin pepa para lograr clusters cuyos límites estén bien definidos, que es como se obtendrían al emplear SVM en el caso real de estudio. Finalmente se obtienen los dominios, Figura 15, donde la cantidad de bloques que posee cada uno son 6725, 4305 y 7051 para los dominios uno, dos y tres respectivamente.

7.2.3 Simulación de variables

Se simulan tres variables en total: dos continuas (leyes de cobre y de oro) y una categórica (tipo de roca) dentro del campo de estudio del depósito. Estas variables tienen el propósito de actuar como los datos reales con los que se validarán los resultados obtenidos de las diferentes estimaciones.

Debido a que son tres dominios geometalúrgicos los generados, para emular la presencia de estos mismos, se llevan a cabo tres simulaciones por iteración, cada una de estas dentro de su respectivo dominio para un total de 100 iteraciones de las tres variables mencionadas.

Para la simulación de las variables se emplean tres variogramas directos y tres variogramas cruzados idénticos para las tres zonas, de alcance igual a 100 metros, ecuación (50), pero debido a que las simulaciones son independientes entre dominios, se generan finalmente distribuciones espaciales diferentes entre sí que emulan la presencia de los dominios geometalúrgicos.

$$\begin{aligned}
 \gamma_{Cu}(h) &= 0.001 + 0.01 \text{Gau}_{a=100}(h) \\
 \gamma_{Au}(h) &= 0.01 + 0.06 \text{Gau}_{a=100}(h) \\
 \gamma_{Rt}(h) &= 0.01 \text{Gau}_{a=100}(h) \\
 \gamma_{Cu-Au}(h) &= 1 \times 10^{-5} + 0.001 \text{Gau}_{a=100}(h) \\
 \gamma_{Cu-Rt}(h) &= 9 \times 10^{-6} + 0.006 \text{Gau}_{a=100}(h) \\
 \gamma_{Au-Rt}(h) &= 9 \times 10^{-6} + 0.006 \text{Gau}_{a=100}(h)
 \end{aligned} \tag{50}$$

Todas las variables una vez simuladas poseen una distribución normal de media cero. Las variables continuas son posteriormente acotadas a un intervalo entre cero y uno, ecuación (51), mientras que la variable categórica es simulada como una variable continua y separada según sus

cuartiles, Figura 16, dando una pequeña holgura en la separación para que los cuatro tipos de roca formados posean cantidades diferentes de miembros en cada iteración.

$$Z^*(x) = \frac{e^{Z(x)} - \min(e^{Z(x)})}{\max(e^{Z(x)}) - \min(e^{Z(x)})} \quad (51)$$

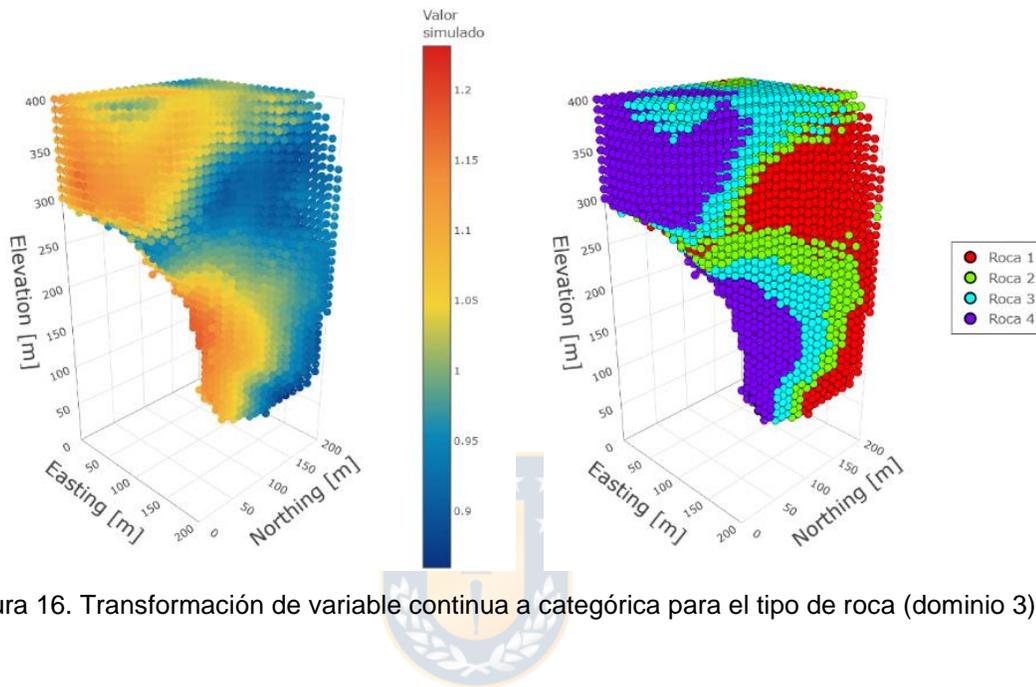


Figura 16. Transformación de variable continua a categórica para el tipo de roca (dominio 3)

Finalmente, estas 100 bases de datos generadas, cada una con tres variables, corresponden a las bases de validación para todos los métodos de estimación a emplear.

7.2.4 Extracción de sondajes

La simulación de la posición de los sondajes se realiza de manera aleatoria, gobernada por dos variables para otorgar realismo a los testigos que se obtendrán:

- **Largo mínimo de sondajes:** suponiendo que se encuentra en la fase exploratoria de un depósito, el largo de los sondajes varía de acuerdo a los resultados que se vayan obteniendo, esto último es complejo de simular, por lo que se opta por dar un largo aleatorio no menor al 20% de la profundidad de la zona de estudio, con una distribución aleatoria que tiende a dar largos mayores, Figura 17.

- **Porcentaje de pérdida de información:** normalmente no hay recuperación completa de la información que se pueda obtener de sondajes, por lo que se generan secciones que carecen de información. Esto es simulado a través de una variable que remueve información de los sondajes de manera aleatoria, siendo mayor la remoción cuanto más profundo se encuentre el dato con respecto a la cota más alta del depósito y evitando remover bloques contiguos.

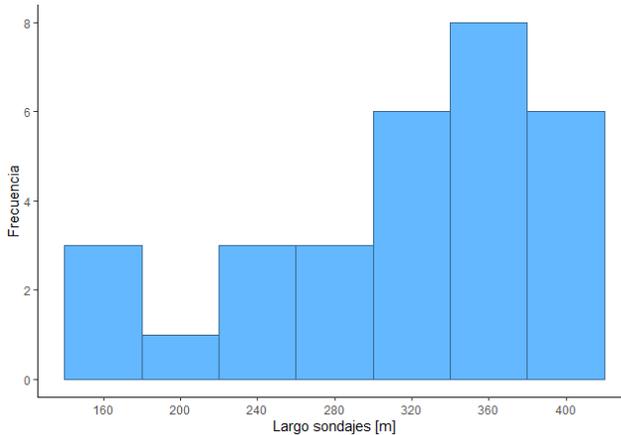


Figura 17. Histograma de largos de sondaje

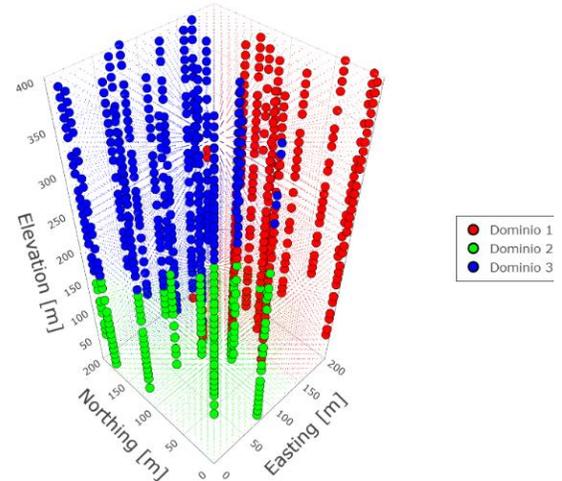


Figura 18. Sondajes empleados para estimar simulación

Se emplean los mismos sondajes para las 100 iteraciones, Figura 18. En total son 30 sondajes y estos son clasificados y divididos dependiendo del dominio geometalúrgico al que pertenecen y corresponden a las bases de entrenamiento de cada dominio, Tabla 4, que son empleadas en cada algoritmo de estimación.

Tabla 4. Número de bloques empleados para realizar estimación

Dominio	Bloques con sondaje	Bloques totales	Porcentaje de entrenamiento (%)	Porcentaje de validación (%)
Dominio 1	228	5895	3.87	96.13
Dominio 2	223	5958	3.74	96.26
Dominio 3	306	6228	4.91	95.09
Base total	757	18081	4.19	95.81

7.2.5 Aplicación de algoritmos de estimación

Para la evaluación de la hipótesis se emplean tres combinaciones de métodos para estimar la base de datos:

- Kriging ordinario para las variables continuas y co-kriging indicador ordinario para la variable categórica.
- Co-kriging ordinario para todas las variables, transformando la variable categórica en indicadores.
- Inverso de la distancia para todas las variables, transformando la variable categórica en indicadores.

Luego se evalúa el error cuadrático medio para las variables continuas y la precisión para la variable categórica.

Kriging ordinario y co-kriging indicador ordinario

Con los sondeos presentes en cada dominio se realiza kriging ordinario sobre las variables continuas. Primero se determinan los variogramas experimentales para cada iteración y se les asigna un variograma experimental a priori, siendo este omnidireccional, ya que el estudio de anisotropías se vuelve tedioso con tantas iteraciones.

El único parámetro fijado es el alcance del modelo, el cual corresponde a 100 metros, mientras que la pepa, el modelo básico del variograma y la meseta no son especificados dentro del código, sino más bien, ingresados como datos iniciales para que el programa encuentre el modelo que mejor se adapte el variograma a través de la minimización de las diferencias cuadráticas, Figura 19, entre el modelo variográfico propuesto y el variograma experimental en las distancias o *lags* de medición. Para la iteración uno del dominio uno los modelos escogidos son Gaussiano y Exponencial, con alcances de 117.9 y 79.4 metros respectivamente.

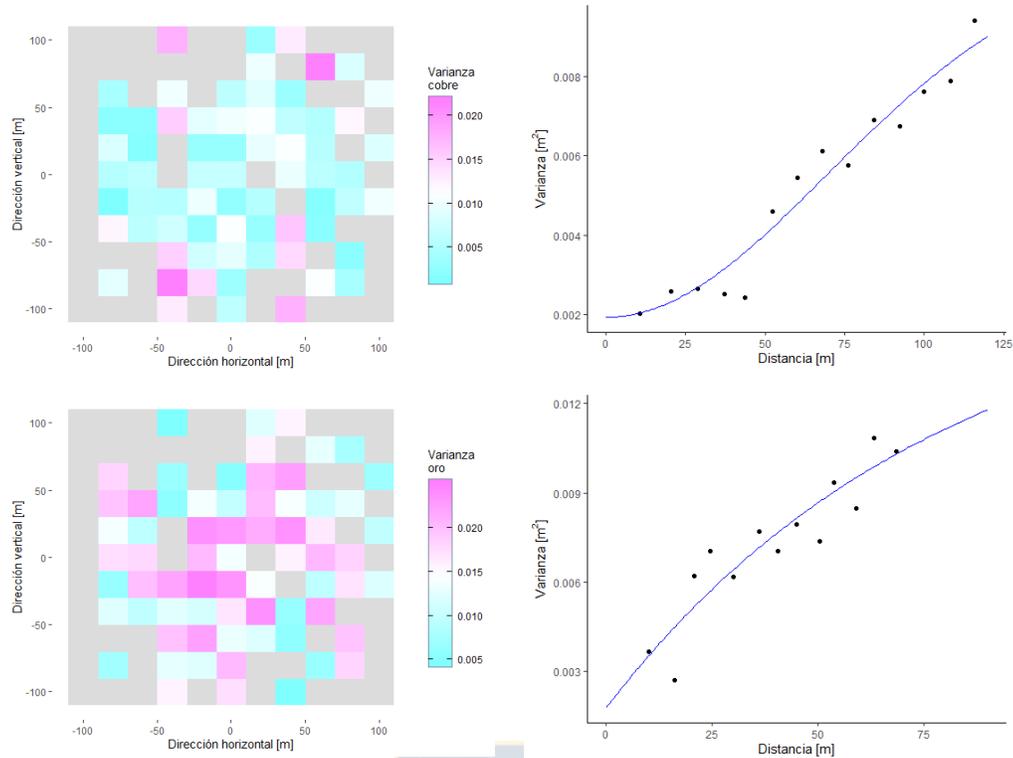


Figura 19. Variograma y modelo para cobre y oro iteración 1, dominio 1

Posteriormente, se efectúa el kriging ordinario, Figura 20, con la única restricción de que el número máximo de objetos vecinos que participen en la estimación sea de 10 individuos, para acelerar la estimación.

Para el caso del co-kriging indicador, la variable a predecir, tipo de roca, se discretiza en cuatro variables dicotómicas nuevas, también llamadas indicadores, representando cada tipo de roca (tomando el valor de uno cuando la ubicación corresponde a dicho tipo de roca y cero cuando no). A continuación, se determinan los variogramas experimentales directos y cruzados para cada variable y son ajustados de manera de cumplir con el modelo lineal de correogionalización, Figura 21, que postula que los alcances y los modelos básicos empleados para describir los variogramas deben ser iguales, además de las pepas y mesetas ser razonables, logrando obtener matrices no singulares que permitan efectuar co-kriging. De encontrarse matrices singulares (producto de que algún variograma sea múltiplo de otro, es decir haya correlación perfecta, o que haya efecto pepa puro entre variables no correlacionadas), se agrega un factor de corrección para poder llevar a cabo co-kriging indicador en dichos casos.

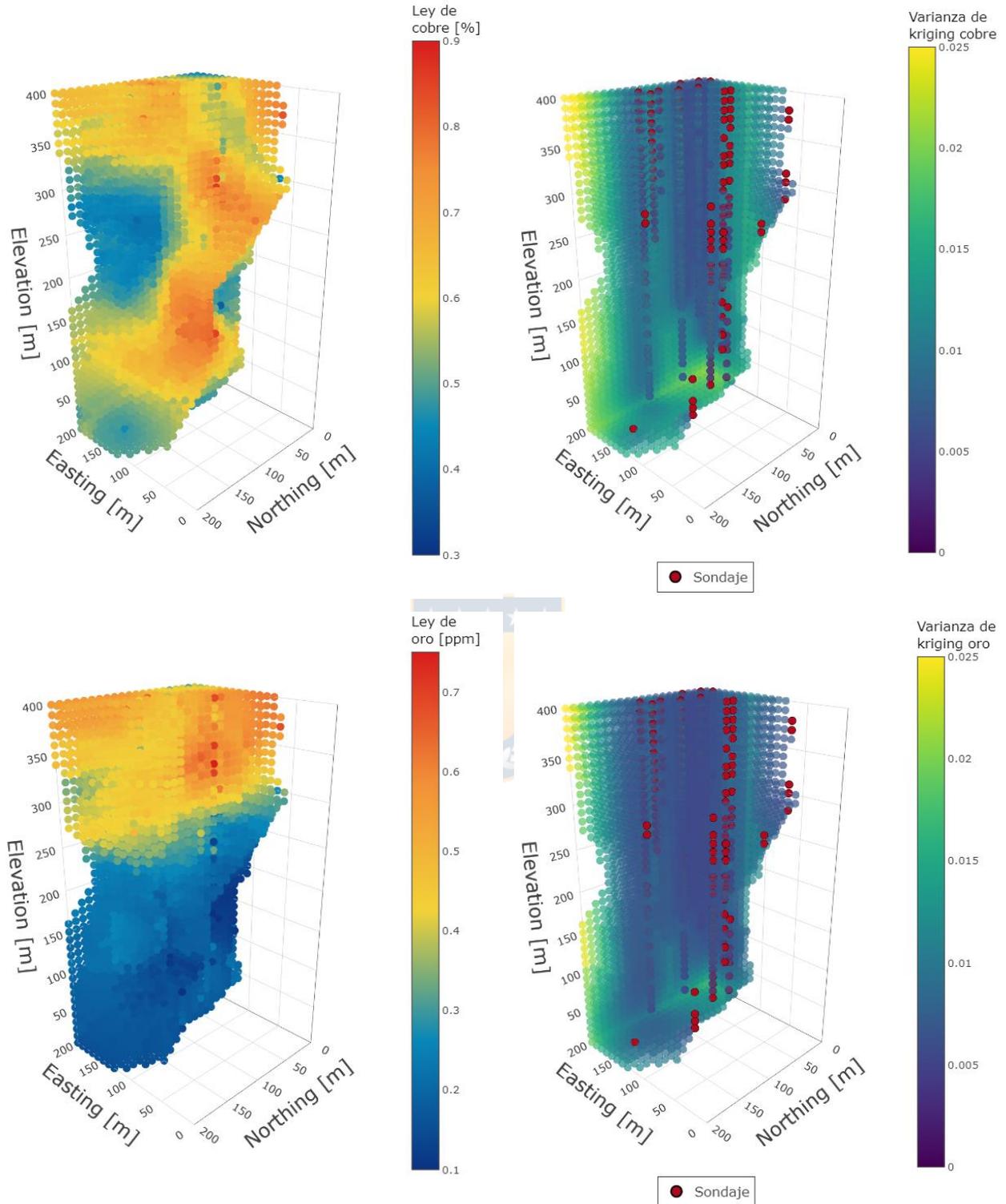


Figura 20. Predicción y varianza de kriging ordinario, cobre y oro iteración 1, dominio1

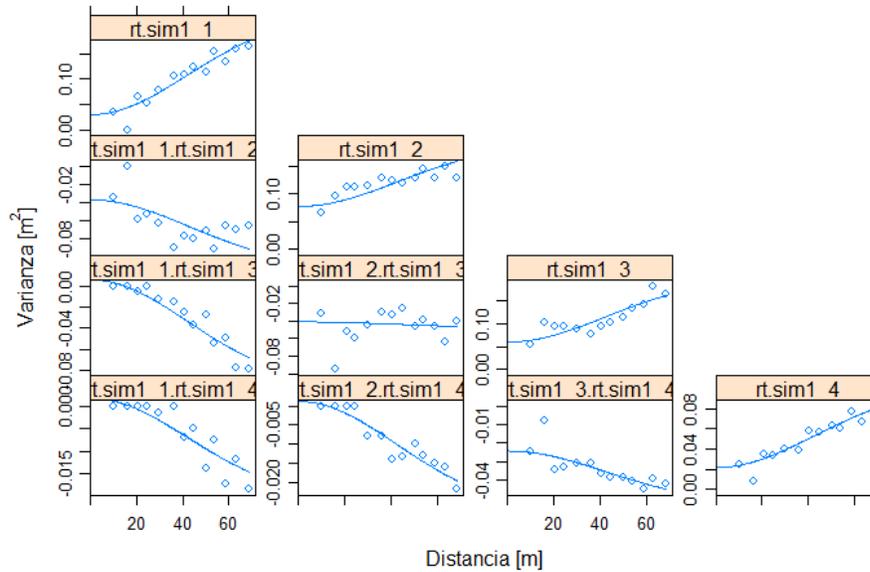


Figura 21. Variogramas directos y cruzados para tipo de roca, iteración 1, dominio 1

Luego, la variable tipo de roca es estimada mediante co-kriging indicador ordinario, empleando como restricción que el número máximo de puntos empleados en la estimación sean los 10 más cercanos.

Posteriormente, se construye la base de datos para el tipo de roca a partir de los resultados obtenidos. Estos fluctúan entre cero y uno, Tabla 5, siendo la sumatoria de los valores predichos por el co-kriging indicador igual a uno para cada dato estimado. Luego el tipo de roca que posea el valor más alto para la predicción en ese punto es escogido, siendo este el valor más probable. En el caso que haya uno o más datos con el mismo valor, se escoge aquel que posea la menor varianza normalizada.

Tabla 5. Ejemplo elección tipo de roca preponderante co-kriging indicador

Easting	Northing	Elevation	Roca 1	Roca 2	Roca 3	Roca 4	Total
70	70	120	0.118	0.663	0.033	0.186	1.000
80	0	40	0.003	0.279	0.565	0.153	1.000
80	0	50	0.000	0.235	0.576	0.189	1.000

Finalmente, la predicción y varianza del co-kriging indicador se visualiza en la Figura 22, siendo como única restricción, la estimación con 10 individuos más cercanos. Se observa que la varianza del Co-kriging indicador cambia ligeramente cuando se cambia de tipo de roca, pero sigue el patrón

global, donde las zonas más amarillas albergan los valores más altos de varianza además de encontrarse más alejados de los sondajes, lo cual es esperable.

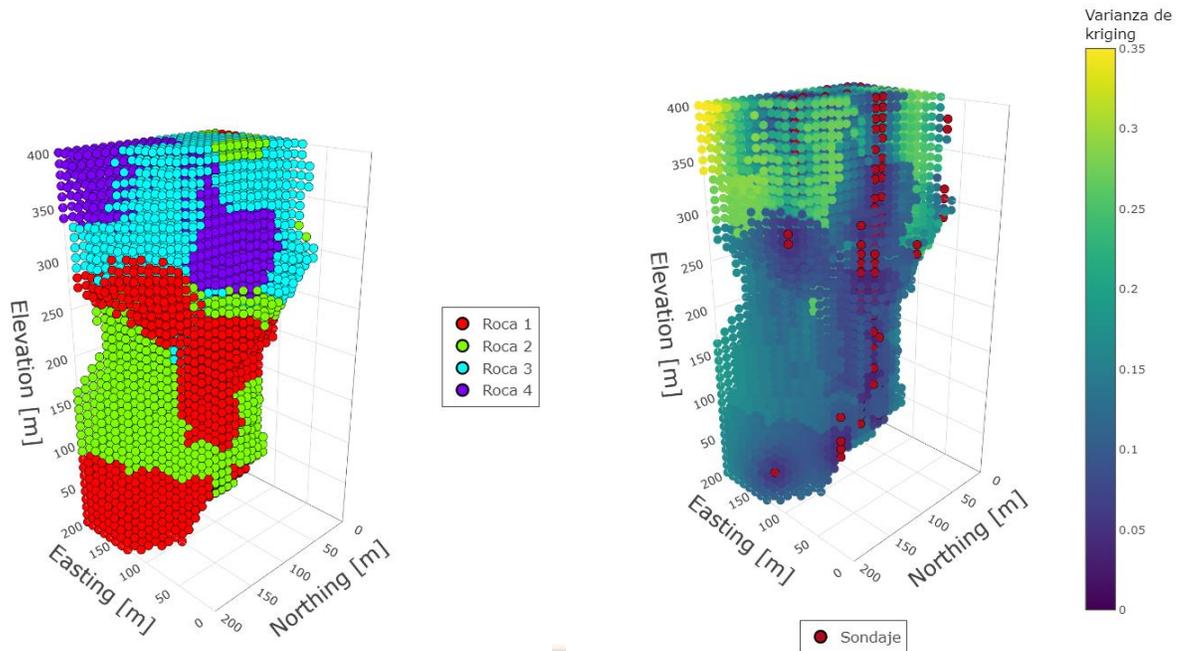


Figura 22. Predicción y varianza del co-kriging indicador, iteración 1, dominio 1

Co-kriging ordinario para todas las variables

Con los sondajes presentes en cada dominio se realiza co-kriging ordinario sobre las variables cobre, oro y tipo de roca. Antes de determinar los variogramas, se transforma el tipo de roca en cuatro variables dicotómicas, conocidas como indicadores, cuyos valores son uno o cero si la variable corresponde o no cada tipo de roca, respectivamente. Se determinan los variogramas experimentales directos y cruzados para cada iteración y se les asigna un variograma teórico a priori a cada una de las seis variables y 15 covariables, siendo este omnidireccional.

El único parámetro fijado es el alcance del modelo, el cual corresponde a 100 metros, mientras que la pepa, el modelo básico del variograma y la meseta no son especificados dentro del código.

Este variograma teórico es fijado solo para una variable principal (ya sea oro, cobre o cualquier tipo de roca) y luego replicado al resto de los variogramas con exactamente los mismos valores de pepa, meseta, alcance y tipo de modelo. Posteriormente, estos variogramas son ajustados a través de una función que modifica los parámetros para que mejor se adapte a cada variograma mediante

la minimización de las diferencias cuadráticas entre los modelos variográficos propuestos, Figura 23, y el variograma experimental en las distancias o *lags* de medición.

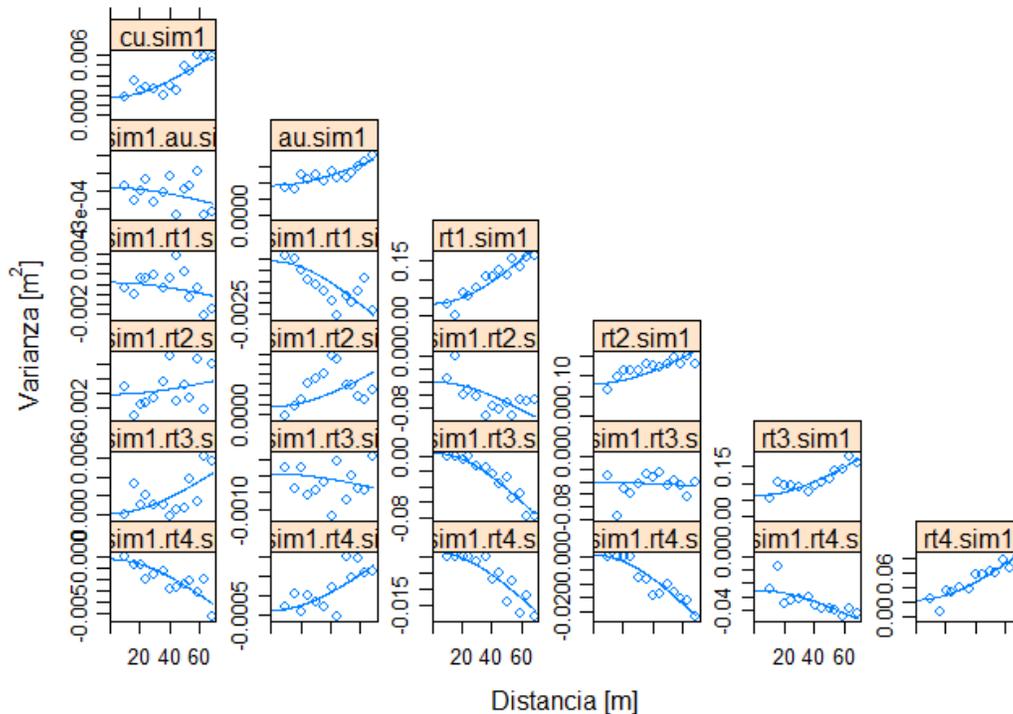


Figura 23. Variogramas experimental y teórico para el sistema de co-kriging entre oro, cobre y los tipos de roca, iteración 1, dominio 1

Luego, se procede a realizar la estimación de las seis variables a través de co-kriging ordinario, empleando como única restricción que solo participen los 10 individuos más cercanos al punto a estimar.

Una vez obtenidos los resultados, se construye la variable tipo de roca a partir de los cuatro valores obtenidos y, de manera similar al co-kriging indicador, se escoge el valor más alto y, por ende, el tipo de roca que predomina en ese punto. A diferencia del co-kriging indicador, los valores predichos para cada tipo de roca pueden ser negativos o mayores que uno y sumar más o menos que uno, Tabla 6. Esto se debe a que la adaptación de co-kriging con indicadores creados manualmente no posee la restricción de que los valores sumen uno al momento de estimar.

Finalmente se obtienen las predicciones de las leyes de cobre y oro, además del tipo de roca para cada punto, Figura 24, siendo estos resultados similares a los obtenidos por kriging ordinario y co-kriging indicador.

Tabla 6. Ejemplo elección tipo de roca preponderante co-kriging ordinario

Easting	Northing	Elevation	Roca 1	Roca 2	Roca 3	Roca 4	Total
70	50	140	-0.066	0.207	0.534	0.005	0.680
100	10	40	0.001	0.024	0.375	0.905	1.305
90	0	130	-0.046	0.187	0.807	0.226	1.174

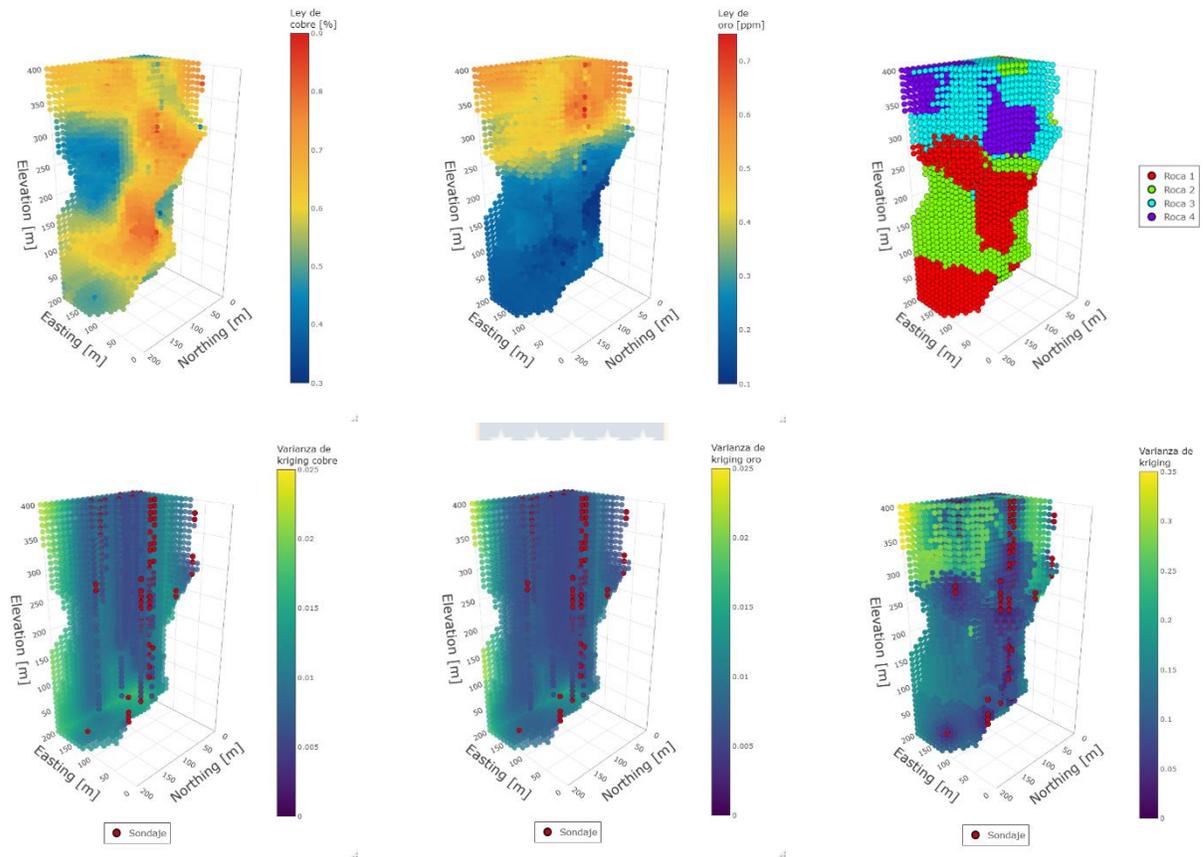


Figura 24. Predicción y varianza del co-kriging ordinario, iteración 1, dominio 1

Inverso de la distancia para todas las variables

Con los sondajes presentes en cada dominio se realiza la estimación mediante inverso de la distancia sobre las variables cobre, oro y tipo de roca. De manera similar al co-kriging ordinario con todas las variables, se transforma la variable tipo de roca en cuatro variables dicotómicas. Posterior a esto se procede a estimar empleando como los valores de p iguales a 4.2 para el cobre, 3.5 para el oro y 5.8 para el tipo de roca, siendo estos valores la media de los valores de p probados para cada variable que minimizaban el RMSE, ecuación (52).

El beneficio de determinar el óptimo para cada iteración es mínimo, Figura 25, siendo la media del error cuadrático medio cuando se ajusta el valor de p para cada iteración igual a 0.0978 y 0.1014 para cobre y oro respectivamente. Por su parte, cuando el valor de p es fijado para cada variable los valores del error cuadrático medio son iguales a 0.0982 y 0.1015 para cobre y oro respectivamente. Luego, el determinar el valor de p para cada iteración presenta una mejora del 0.41% y 0.099% en promedio para cobre y oro respectivamente, contrastada con ser mucho más costoso computacionalmente el calcular los valores óptimos debido a lo lento que converge el algoritmo de escalada simple, por lo que se prefiere emplear la segunda opción. Ocurre un efecto similar con los valores de p para el tipo de roca.

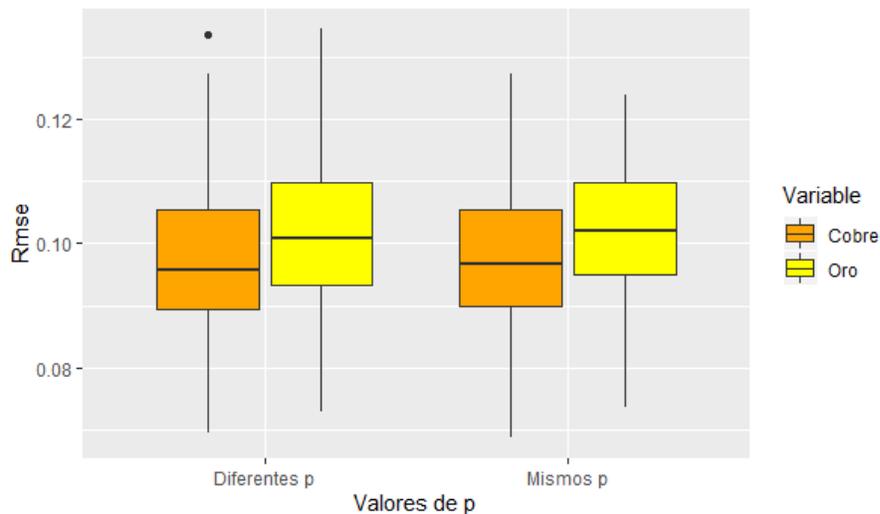


Figura 25. Diagramas de caja para los errores cuadráticos medios para ambos escenarios de p

Una vez estimados los valores para cada tipo de roca, se reconstruye esta variable de manera similar al caso de co-kriging ordinario con todas las variables, se escoge el valor más alto de los cuatro obtenidos y, por ende, se conoce finalmente el tipo de roca que predomina en ese punto. Los valores obtenidos son similares a los estimados por co-kriging indicador, en el sentido de que la suma de las predicciones es igual a uno, Tabla 7.

Tabla 7. Ejemplo elección tipo de roca preponderante inverso de la distancia indicador

Easting	Northing	Elevation	Roca 1	Roca 2	Roca 3	Roca 4	Total
30	70	160	0.000	0.034	0.966	0.000	1.000
70	10	150	0.000	0.117	0.883	0.000	1.000
140	180	150	0.001	0.064	0.935	0.000	1.000

Finalmente se obtienen las predicciones de las leyes de cobre y oro, además del tipo de roca para cada punto, Figura 26, en este caso sin restricciones de vecindario. Notar que a diferencia de los métodos de kriging, no hay varianza asociada a las predicciones.

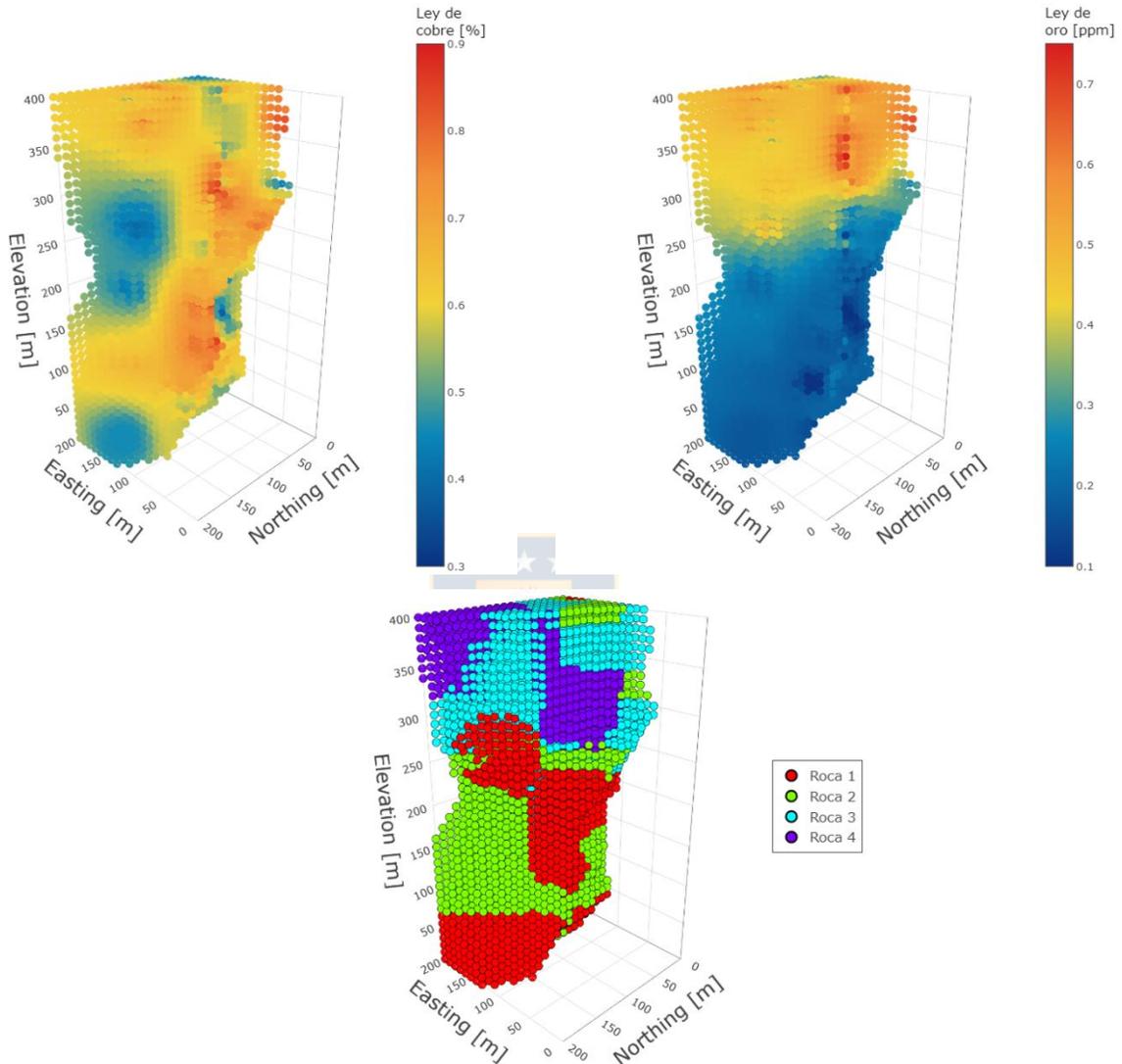


Figura 26. Predicción del inverso de la distancia, iteración 1, dominio 1

Errores cuadráticos medios y precisiones

Con respecto a la raíz de los errores cuadráticos medios (RMSE), ecuación (52), donde n corresponde a la cantidad de datos y $Z^*(x_i) - Z(x_i)$ es la diferencia entre el valor predicho y real en la i -ésima ubicación, se desprende de la Figura 27 que los tres métodos poseen un desempeño similar en los tres dominios, siendo el inverso de la distancia (IDW) levemente inferior al kriging ordinario (OK) y al co-kriging ordinario (OCK).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z^*(x_i) - Z(x_i))^2}{n}} \quad (52)$$

El dominio dos es el dominio donde peores resultados se obtuvieron, seguido del dominio uno y del dominio tres. Esto concuerda con el porcentaje de datos con los que se estimó en cada dominio, los cuales según la Tabla 4, son de 3.74%, 3.87% y 4.91% respectivamente. El dominio dos en particular está compuesto por datos ubicados en el fondo del yacimiento, lo que implica que existen datos de validación que están más lejos de los de entrenamiento que en los otros dominios. El dominio tres es el de mejor desempeño ya que gran parte de sus datos se encuentran cercanos a superficie y posee una cantidad considerablemente mayor de datos de entrenamiento (los cuales no están dispuestos en conglomerados por construcción), lo que favoreció la estimación en dicho dominio.

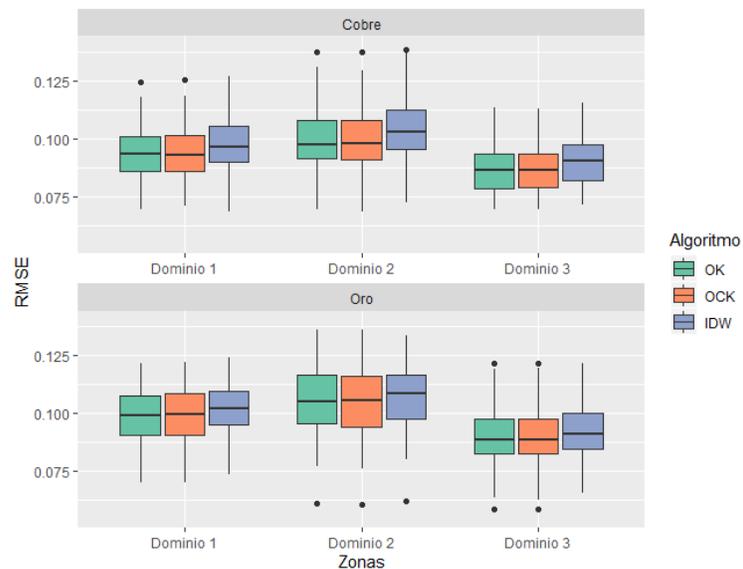


Figura 27. Diagramas de caja para los RMSE de cada dominio para cobre y oro

En relación a las variables cobre y oro, la raíz de los errores cuadráticos medios para todos los métodos y en todos los dominios fue mayor para el oro. Esto se debe a que el oro fue simulado con una pepa equivalente al 14% de la meseta global, mientras que el cobre fue simulado con un 9%. Esta diferencia, aunque sutil, impacta sobre las predicciones ya que disminuye la continuidad espacial conforme aumenta la pepa, y, por lo tanto, hace más difícil estimar mediante métodos geoestadísticos. Este efecto es un poco menor al emplear métodos determinísticos como el inverso

de la distancia, es decir, posee un desempeño similar a los algoritmos de kriging mientras la variable es más discontinua. Un efecto similar, se ve en la Figura 28.

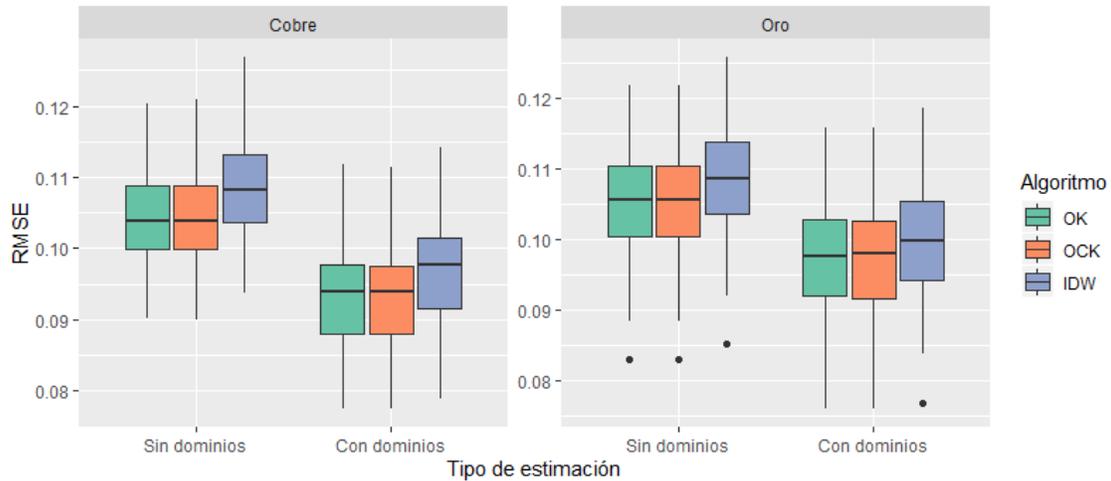


Figura 28. Diagramas de caja para los RMSE de cada tipo de estimación para cobre y oro

Tabla 8. RMSE medio para cada variable y algoritmo

Variable	Algoritmo	RMSE medio sin dominios	RMSE medio con dominios	Disminución [%]
Cobre	OK	0.104	0.093	10.4
	OCK	0.104	0.093	10.4
	IDW	0.109	0.097	10.8
Oro	OK	0.106	0.097	7.6
	OCK	0.106	0.097	7.6
	IDW	0.109	0.100	8.1

La diferencia en la raíz del error cuadrático medio entre estimar empleando dominios geometalúrgicos y cuando se estima sobre todo el yacimiento para cada algoritmo es tal que, en el cobre nunca se intersectan los rangos intercuartiles de las distribuciones de los errores cuadráticos medios para todos los algoritmos en las 100 simulaciones, mientras que para el oro se intersectan levemente, Figura 28, lo cual indica que efectivamente hay una diferencia significativa entre realizar dichos tipos de estimación. La disminución en el RMSE, Tabla 8, es mayor para el cobre que para el oro, y, en cuanto a los algoritmos, siempre mayor para el inverso de la distancia. En el caso del cobre el 100% de las veces es mejor estimar mediante dominios geometalúrgicos, con cualquiera de los métodos empleados, mientras que para el oro el 99% de las veces. El caso para el cual no fue mejor estimar mediante dominios será discutido en el título 7.3.

Con respecto a las precisiones, estas fueron determinadas empleando la ecuación (53), con n la cantidad de datos, $Z^*(x_i)$ y $Z(x_i)$ el valor real y el valor predicho y D_i el vector de diferencias.

$$D_i = \begin{cases} 1 & \text{si } Z^*(x_i) \neq Z(x_i) \\ 0 & \text{si } Z^*(x_i) = Z(x_i) \end{cases} \quad (53)$$

$$P = \frac{\sum_{i=1}^n (D_i)}{n}$$

Las precisiones son calculadas para los datos categóricos ya que determinar RMSE no tiene mucho sentido aun cuando las categorías estén expresadas con números. Se obtuvo un comportamiento similar con respecto a qué dominios obtuvieron mejores desempeños en el caso de las variables continuas. Esto es: el dominio dos es el que obtuvo un peor desempeño, seguido del dominio uno y finalmente del dominio tres. Los tres métodos empleados poseen un desempeño similar, siendo nuevamente el inverso de la distancia (IDW) el que peores resultados obtuvo, mientras que el co-kriging indicador (IK) y el co-kriging ordinario con todos los datos (OCK) obtuvieron resultados prácticamente idénticos.

La diferencia en precisión entre estimar empleando los dominios geometalúrgicos y cuando se estima sobre todo el yacimiento es nuevamente amplia, Tabla 9. Los rangos intercuartiles entre estos dos tipos de estimación apenas se intersectan, Figura 30, mientras que para todos los métodos el 100% de las veces es mejor estimar empleando dominios geometalúrgicos, a excepción de un caso para el inverso de la distancia, donde es levemente mejor estimar sin dominios, por la misma razón que será discutida en el título 7.3.

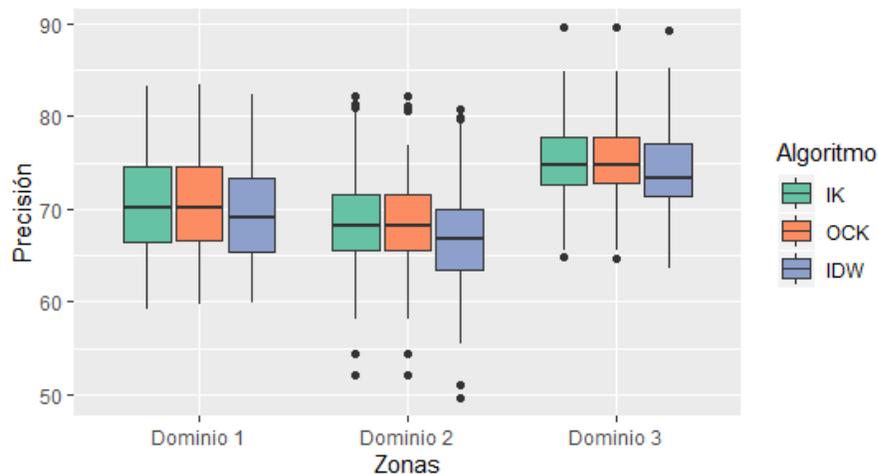


Figura 29. Diagramas de caja de las precisiones de cada dominio para el tipo de roca

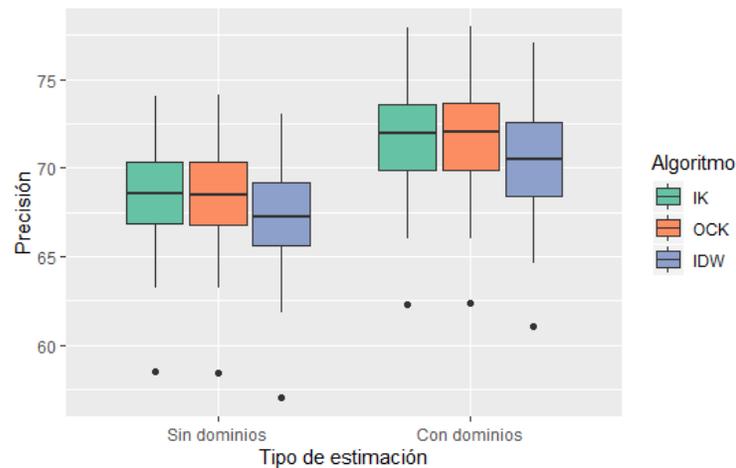


Figura 30. Diagramas de caja de las precisiones de cada tipo de estimación para cobre (izquierda) y oro (derecha)

Tabla 9. Precisión media para cada variable y algoritmo

Variable	Algoritmo	Precisión media sin	Precisión media con	Aumento [%]
		dominios	dominios	
Cobre	IK	68.6	71.8	4.7
	OCK	68.6	71.8	4.7
	IDW	67.2	70.7	5.2

7.2.6 Evaluación estadística

Se escogen el mejor, el peor y el caso más probable con respecto a la precisión de las estimaciones empleando dominios geometalúrgicos y se contrastan con los resultados obtenidos al estimar sin estos ser empleados. Dichas simulaciones corresponden a la estimación número 61, 86 y 84 respectivamente. Estas estimaciones fueron seleccionadas de acuerdo a las ponderaciones entre los RMSE y la precisión obtenida, cuando corresponda, por lo que corresponden al mejor, peor y caso más probable de la simulación de las variables en conjunto, es decir, cuando las tres variables obtienen la mayor diferencia de RMSE/precisión, cuando obtienen la menor diferencia de RMSE/precisión y cuando la diferencia de RMSE/precisión es más probable de manera simultánea, y no para cada variable por si sola. Debido a lo anterior los casos escogidos no necesariamente representan las estimaciones que obtuvieron el mejor, peor o desempeño más probable para cada variable globalmente.

La correlación entre las tres variables para todas las simulaciones es baja, siendo más altas para las correlaciones entre cobre y tipo de roca, y en la mayoría de los casos los sondeos captan esta correlación, Tabla 10, con excepción del mejor caso para la correlación entre cobre y oro, donde para los sondeos es un orden de magnitud menor.

Tabla 10. Correlaciones para los casos a estudiar para los sondeos y la base de validación

Caso	Cu-Au		Cu-Rt		Au-Rt	
	Sondeos	Validación	Sondeos	Validación	Sondeos	Validación
Mejor	0.017	0.138	0.278	0.301	0.205	0.199
Peor	-0.051	-0.089	0.300	0.245	-0.021	-0.034
Más Probable	-0.093	-0.089	0.434	0.443	-0.015	-0.044

En tanto, las distribuciones de las leyes de cobre y oro para cada caso son similares, Figura 31. Sin embargo, al aplicar el test de Kolmogorov–Smirnov, cuya hipótesis nula es que las muestras provienen de una misma distribución, para comparar las distribuciones de los sondeos con las distribuciones de la base de validación para cada variable y cada caso, se obtienen los resultados indicados en la Tabla 11, donde, bajo un nivel de significancia de 0.03, se rechaza la hipótesis para el escenario de cobre con la mejor estimación. Esto ocurre principalmente por la diferencia que ocurre en los valores altos y bajos (también llamados colas) de cobre en la base de validación y la base de sondeos.

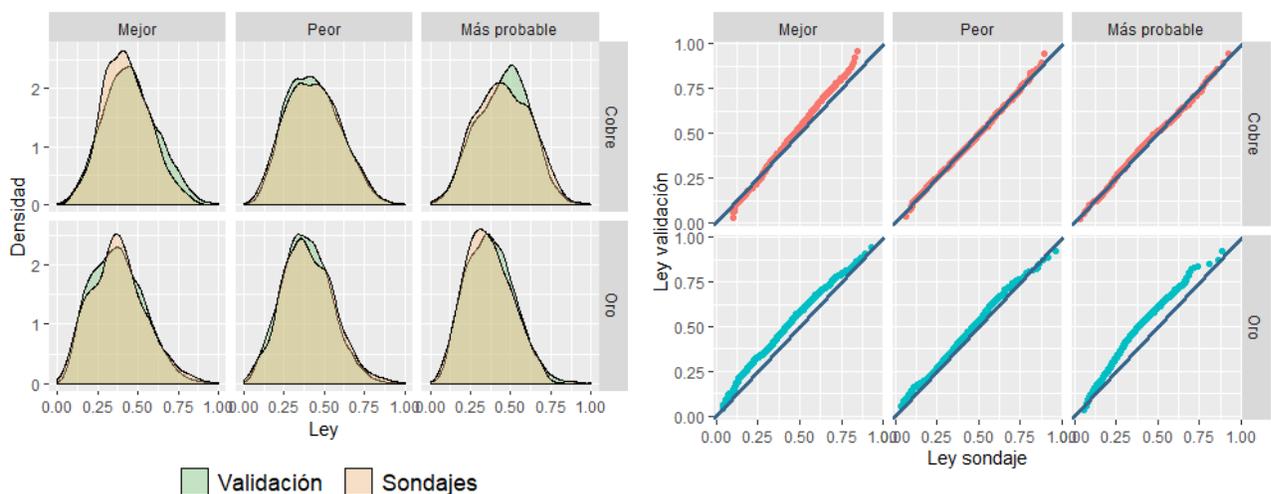


Figura 31. Distribución y gráfico de cuantil-cuantil de leyes de cobre y oro para cada caso

Tabla 11. Resultados test Kolmogorov–Smirnov

Escenario	Valor de p	Resultado
Cobre - Mejor	2.30×10^{-4}	Rechaza
Cobre - Peor	0.870	No rechaza
Cobre - Más probable	0.118	No rechaza
Oro - Mejor	0.188	No rechaza
Oro - Peor	0.114	No rechaza
Oro - Más probable	0.031	No rechaza

Mejor simulación

En la simulación 61 se obtienen las mayores diferencias simultáneas de RMSE y de precisión para los todos los algoritmos empleados, lo cual implica que es la simulación donde más beneficios produjo la estimación empleando dominios geometalúrgicos. Para las variables continuas, la Tabla 12 muestra las principales estadísticas que permiten comparar entre métodos (OK: kriging ordinario, OCK: co-kriging ordinario y IDW: inverso de la distancia) cuando se emplea o no dominios geometalúrgicos.

No existe un patrón claro que indique cómo es el cambio que se produce en la media y en la mediana cuando se pasa de estimar sin dominios a estimar con dominios, por lo que se atribuye netamente a las características de los sondajes empleados. Mientras que la desviación estándar en ambos casos es siempre menor a la desviación original, debido al efecto de suavizamiento de los algoritmos de estimación empleados (siendo el inverso de la distancia IDW quien presenta menos suavizamiento), pero cuando se emplean dominios geometalúrgicos el efecto de suavizamiento, reflejado en la desviación estándar, se ve disminuido en todos los casos en una cantidad significativa, sobre todo en los algoritmos de kriging.

Con respecto a las medidas que comparan directamente los valores obtenidos mediante los diversos algoritmos con la variable original, el RMSE y el error absoluto medio presentan una disminución al emplear dominios geometalúrgicos para todos los algoritmos. La diferencia entre el error medio y cero disminuye para el cobre, pero aumenta para el oro cuando se emplean dominios geometalúrgicos. Esto indica un pequeño cambio en el sesgo probablemente inducido por la discontinuidad en los límites del dominio.

Tabla 12. Medidas estadísticas principales para la mejor simulación

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio	
Ley de cobre	Real	-	0.450	0.166	0.442	-	-	-	
	Sin dominios	OK	0.442	0.115	0.436	0.115	-0.0087	0.0887	
		OCK	0.442	0.114	0.437	0.114	-0.0084	0.0885	
		IDW	0.442	0.125	0.434	0.121	-0.0085	0.0928	
	Con dominios	OK	0.444	0.127	0.431	0.099	-0.0067	0.0781	
		OCK	0.444	0.126	0.431	0.099	-0.0063	0.0779	
		IDW	0.444	0.134	0.431	0.104	-0.0061	0.0821	
	Ley de oro	Real	-	0.373	0.163	0.365	-	-	-
		Sin dominios	OK	0.373	0.121	0.372	0.121	-0.0001	0.0889
OCK			0.374	0.120	0.372	0.121	0.0005	0.0888	
IDW			0.373	0.130	0.372	0.126	-0.0004	0.0914	
Con dominios		OK	0.369	0.140	0.374	0.094	-0.0042	0.0720	
		OCK	0.369	0.139	0.374	0.095	-0.0042	0.0721	
		IDW	0.367	0.143	0.372	0.097	-0.0060	0.0745	

Para el tipo de roca, los porcentajes de acierto para el total de los cuatro tipos de roca se resumen en la Tabla 13. Al emplear dominios geometalúrgicos, el aumento de los aciertos es de 3.4, 3.3 y 3.6 puntos porcentuales para el co-kriging indicador, co-kriging y el inverso de la distancia respectivamente.

Tabla 13. Porcentajes de acierto para la mejor simulación

Número de bloques			17309		
Casos bien clasificados			Porcentaje de acierto		
Base		Total	Base		Total [%]
IK	Sin dominios	12610	IK	Sin dominios	72.9
OCK		12619	OCK		72.9
IDW		12348	IDW		71.3
IK	Con dominios	13204	IK	Con dominios	76.3
OCK		13188	OCK		76.2
IDW		12969	IDW		74.9

Peor simulación

En la simulación 86 se obtienen las menores diferencias simultáneas de RMSE y de precisión para los todos los algoritmos empleados, lo cual implica que es la simulación donde menos beneficios produjo la estimación empleando dominios geometalúrgicos. Para las variables continuas, la Tabla 14 muestra las principales estadísticas que permiten comparar entre métodos (OK: kriging ordinario, OCK: co-kriging ordinario y IDW: inverso de la distancia) cuando se emplea o no dominios geometalúrgicos.

De igual manera que para la mejor simulación, no existe un patrón claro que indique cómo es el cambio que se produce en la media y en la mediana cuando se pasa de estimar sin dominios a estimar con dominios, por lo que se atribuye netamente a las características de los sondajes empleados, sin embargo, para esta simulación tanto la media como la mediana se sobreestiman con todos los algoritmos y en ambos casos con y sin dominios geometalúrgicos lo cual indica que los valores de los sondajes son mayores en general que los valores reales. Mientras que la desviación estándar en ambos casos es siempre menor a la desviación original, debido al efecto de suavizamiento de los algoritmos de estimación empleados (siendo la desviación estándar mayor para el inverso de la distancia IDW), pero cuando se emplean dominios geometalúrgicos el efecto de suavizamiento, reflejado en la desviación estándar, se ve disminuido en todos los casos, en una cantidad significativa para el oro.

Con respecto a las medidas que comparan directamente los valores obtenidos mediante los diversos algoritmos con la variable original, el RMSE y el error absoluto medio presentan una disminución al emplear dominios geometalúrgicos para todos los algoritmos. La diferencia entre el error medio y cero disminuye para el oro, pero aumenta para el cobre.

Para el tipo de roca, los porcentajes de acierto para el total de los cuatro tipos de roca se resumen en la Tabla 15. Al emplear dominios geometalúrgicos, el aumento de los aciertos es de 1.7, 1.6 y 1.6 puntos porcentuales para el co-kriging indicador, co-kriging y el inverso de la distancia respectivamente.

Tabla 14. Medidas estadísticas principales para la peor simulación

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio	
Ley de cobre	Real	-	0.430	0.163	0.420	-	-	-	
	Sin dominios	OK	0.424	0.130	0.419	0.104	-0.0067	0.0809	
		OCK	0.424	0.130	0.420	0.104	-0.0060	0.0808	
		IDW	0.423	0.138	0.419	0.109	-0.0069	0.0846	
	Con dominios	OK	0.420	0.133	0.420	0.098	-0.0100	0.0762	
		OCK	0.421	0.133	0.421	0.097	-0.0089	0.0758	
		IDW	0.422	0.139	0.427	0.101	-0.0083	0.0793	
	Ley de oro	Real	-	0.389	0.152	0.383	-	-	-
		Sin dominios	OK	0.403	0.116	0.403	0.113	0.0143	0.0882
OCK			0.403	0.116	0.403	0.114	0.0144	0.0884	
IDW			0.401	0.123	0.401	0.116	0.0120	0.0899	
Con dominios		OK	0.402	0.136	0.400	0.112	0.0128	0.0857	
		OCK	0.402	0.135	0.401	0.112	0.0131	0.0858	
		IDW	0.398	0.138	0.398	0.111	0.0091	0.0857	

Tabla 15. Porcentajes de acierto para la peor simulación

		Número de bloques		17309	
Casos bien clasificados			Porcentaje de acierto		
	Base	Total		Base	Total [%]
IK	Sin dominios	12382	IK	Sin dominios	71.5
OCK		12388	OCK		71.6
IDW		12169	IDW		70.3
IK	Con dominios	12678	IK	Con dominios	73.2
OCK		12663	OCK		73.2
IDW		12449	IDW		71.9

Simulación más probable

La simulación 84 es una de las que representa de mejor manera las diferencias simultáneas de RMSE y de precisión más probables para los todos los algoritmos empleados, lo cual implica que es una de las simulaciones que mejor muestra los beneficios que usualmente se obtienen producto de la estimación empleando dominios geometalúrgicos. Para las variables continuas, la Tabla 16 muestra las principales estadísticas que permiten comparar entre métodos (OK: kriging ordinario,

OCK: co-kriging ordinario y IDW: inverso de la distancia) cuando se emplea o no dominios geometalúrgicos.

De igual manera que para las simulaciones anteriores, no existe un patrón claro que indique cómo es el cambio que se produce en la media y en la mediana cuando se pasa de estimar sin dominios a estimar con dominios, por lo que se atribuye netamente a las características de los sondajes empleados, sobrestimando los valores de la media y de la mediana como en el caso anterior. Mientras que la desviación estándar en ambos casos es siempre menor a la desviación original, debido al efecto de suavizamiento de los algoritmos de estimación empleados (siendo mayor la desviación estándar para el inverso de la distancia IDW), pero cuando se emplean dominios geometalúrgicos el efecto de suavizamiento, reflejado en la desviación estándar, se ve disminuido en todos los casos.

Tabla 16. Medidas estadísticas principales para la simulación más probable

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio	
Ley de cobre	Real	-	0.458	0.162	0.467	-	-	-	
	Sin dominios	OK	0.461	0.135	0.470	0.097	0.0029	0.0754	
		OCK	0.460	0.134	0.469	0.097	0.0025	0.0752	
		IDW	0.462	0.144	0.469	0.103	0.0040	0.0793	
	Con dominios	OK	0.458	0.142	0.473	0.087	-0.0001	0.0687	
		OCK	0.458	0.142	0.474	0.087	0.0001	0.0688	
		IDW	0.458	0.151	0.471	0.091	-0.0002	0.0722	
	Ley de oro	Real	-	0.375	0.150	0.367	-	-	-
		Sin dominios	OK	0.379	0.110	0.380	0.110	0.0034	0.0845
OCK			0.379	0.109	0.381	0.110	0.0034	0.0845	
IDW			0.378	0.116	0.382	0.113	0.0031	0.0865	
Con dominios		OK	0.379	0.114	0.385	0.100	0.0039	0.0776	
		OCK	0.380	0.113	0.386	0.101	0.0043	0.0781	
		IDW	0.379	0.120	0.383	0.103	0.0035	0.0794	

Con respecto a las medidas que comparan directamente los valores obtenidos mediante los diversos algoritmos con la variable original, el RMSE y el error absoluto medio presentan una disminución al emplear dominios geometalúrgicos para todos los algoritmos. La diferencia entre el error medio y cero disminuye para el cobre, pero aumenta para el oro.

Tabla 17. Porcentajes de acierto para la simulación más probable

Número de bloques			17309		
Casos bien clasificados			Porcentaje de acierto		
	Base	Total		Base	Total [%]
Sin dominios	IK	11865	IK	Sin dominios	68.5
	OCK	11920	OCK	dominios	68.9
	IDW	11709	IDW		67.6
Con dominios	IK	12400	IK	Con dominios	71.6
	OCK	12409	OCK		71.7
	IDW	12236	IDW		70.7

Para el tipo de roca, los porcentajes de acierto para el total de los cuatro tipos de roca se resumen en la Tabla 17. Al emplear dominios geometalúrgicos, el aumento de los aciertos es de 3.1, 2.8 y 3.1 puntos porcentuales para el co-kriging indicador, co-kriging y el inverso de la distancia respectivamente.

Tabla 18. Resumen de resultados de simulación

		Mejor Caso	Peor Caso	Caso más probable	Promedio de todos
Variable	Método	Reducción de RMSE [%]			
Cobre	OK	13.30	6.10	10.47	10.42
	OCK	13.44	6.48	10.15	10.41
	IDW	13.92	6.84	11.42	10.79
Oro	OK	22.09	1.25	8.47	7.64
	OCK	21.66	1.30	8.00	7.63
	IDW	22.63	4.38	9.17	8.14
Variable	Método	Aumento precisión [%]			
Tipo de roca	IK	4.71	2.39	4.51	4.72
	OCK	4.51	2.22	4.10	4.73
	IDW	5.03	2.30	4.50	5.24

La Tabla 18 resume las reducciones de RMSE y los aumentos en la precisión de las estimaciones. Todos los métodos presentan un comportamiento similar al cambiar de una estimación sin dominios geometalúrgicos a una empleándolos. Sin embargo, los métodos de kriging poseen una reducción de RMSE y aumento de precisión mayor que el inverso de la distancia, además de poseer RMSE menores y precisiones mayores en la gran mayoría de las simulaciones. La variable

que se ve más afectada es el cobre, seguida del oro y luego del tipo de roca. Esto indica que quizás esta metodología favorece a las variables continuas que tengan mayor continuidad espacial.

Variabilidad e incertidumbre

A partir de la varianza de kriging se generan intervalos de confianza para los valores predichos, asumiendo que la varianza sigue una distribución normal. Bajo un nivel de confianza del α , la ecuación (54) define la construcción de los intervalos de confianza para las predicciones, donde $Z^*(x)$ corresponde al valor estimado en x , $z_{\alpha/2}$ es el valor crítico de una distribución normal para un nivel de confianza dado y σ es la raíz de la varianza de estimación. Esta es la principal ventaja del kriging sobre el inverso de la distancia. Lo anterior solamente es realizado para las variables continuas, ya que carece de sentido evaluar diferentes escenarios para el tipo de roca.

$$(Z^*(x) - z_{\alpha/2} * \sigma, Z^*(x) + z_{\alpha/2} * \sigma) \quad (54)$$

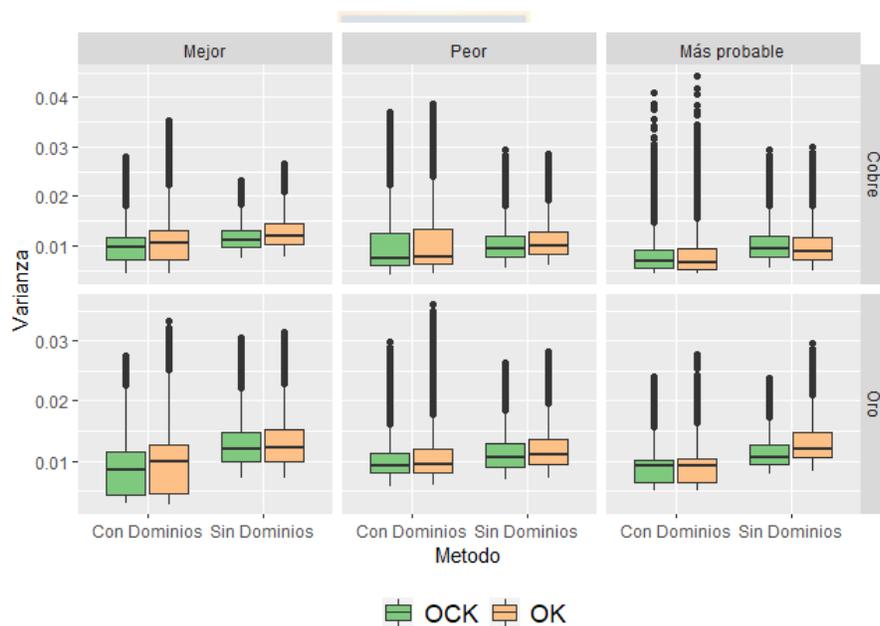


Figura 32. Varianza según caso, variable, método y algoritmo

Con respecto a las varianzas de estimación, estas son mayores cuando no se emplean dominios geometalúrgicos, Figura 32, y la magnitud con que varía está relacionada con la diferencia de la raíz del error cuadrático medio entre ambos métodos (cuando se emplean dominios geometalúrgicos y cuando no), ya que cuando la diferencia es mayor, también lo es la diferencia de las varianzas y viceversa. Cabe destacar que el co-kriging es levemente superior en el sentido que

presenta valores más bajos de varianza, además de ser mejor cuando no se trabaja con dominios geometalúrgicos.

7.2.7 Categorización de recursos

La categorización de recursos se realiza a través de la metodología postulada por Diehl y David, resumida en la Tabla 19. La tolerancia al error, ecuación (55), está relacionada con la desviación estándar relativa de kriging $\sigma(x_k)/z(x_k)$, mientras que la confianza corresponde al valor μ_i , que es el valor crítico de la distribución normal de $i = (1 - (1 - \text{confianza}) / 2)$. La ecuación se emplea despejando el valor de la desviación estándar relativa para las interfases entre los recursos medido-indicado e indicado-inferido. Por consiguiente, esta metodología no es aplicable al inverso de la distancia, sin embargo, se asumen sus varianzas iguales al del kriging ordinario. Métodos para clasificar recursos como la distancia a los sondeos podrían ser aplicados para el caso del inverso de la distancia, pero, dado lo compacto del yacimiento, entregarían sólo resultados medidos y difíciles de comparar con este criterio.

Tabla 19. Metodología de clasificación de recursos resumida

	Medido	Indicado	Inferido
Tolerancia error	±20%	±40%	±60%
Confianza	60% >	40 – 60%	> 40%

$$\text{Precisión (error)} = \frac{\sigma(x_k) * 100 * \mu_i}{z(x_k)} \quad (55)$$

$$\begin{aligned} \frac{\sigma(x_k)}{z(x_k)_{Med-Ind}} &= \frac{20}{\mu_{0.8}} = \frac{20}{0.84} = 0.238 \\ \frac{\sigma(x_k)}{z(x_k)_{Ind-Inf}} &= \frac{40}{\mu_{0.7}} = \frac{40}{0.52} = 0.763 \end{aligned} \quad (56)$$

Luego con los límites, ecuación (56), se clasifican todos los datos según medido, indicado e inferido dependiendo si el valor de su respectiva desviación estándar relativa es menor que el primer valor, se encuentra entre los valores, o es mayor que el segundo valor respectivamente.

La Figura 33 resume los resultados obtenidos para la clasificación de recursos empleando el método anteriormente mencionado. Cabe destacar la escasa cantidad de recursos inferidos es

producto de la buena cobertura espacial de los sondajes, ya que la zona de estimación está muy cercana a ellos. Además, el emplear dominios geometalúrgicos en la estimación favorece la clasificación de recursos como medidos (producto de presentar una menor varianza en general), y, finalmente la diferencia entre métodos es pequeña, pero tiende a favorecer al co-kriging, sobre todo para el oro.

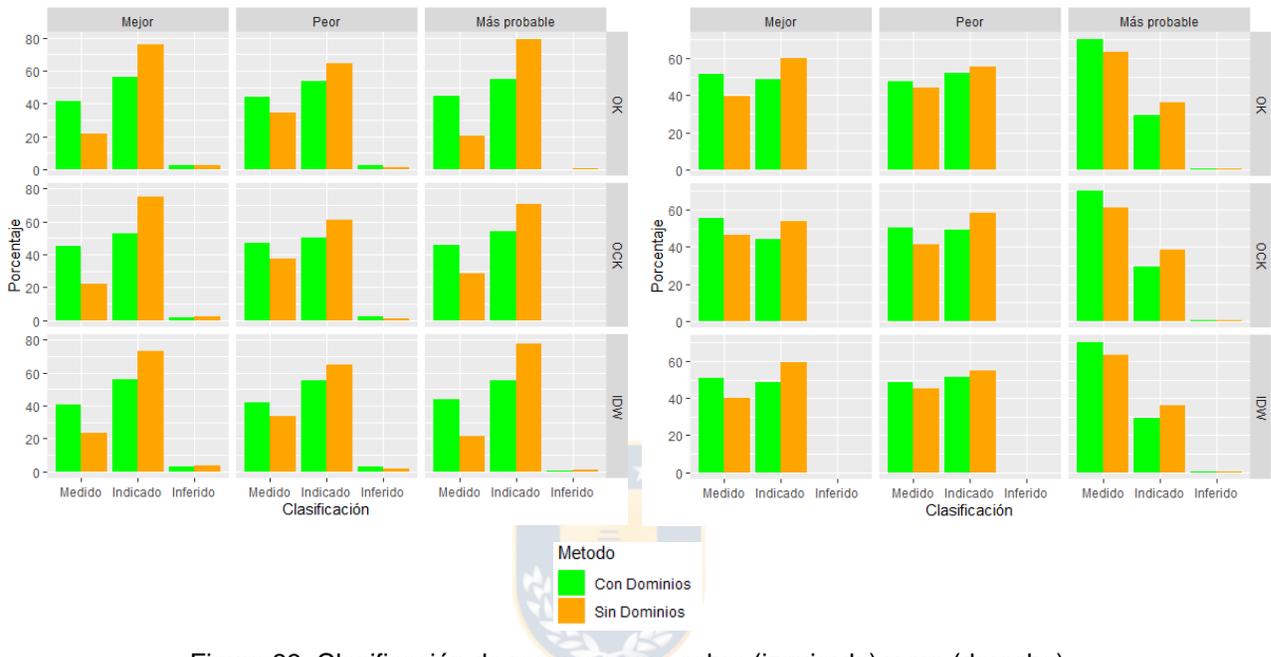


Figura 33. Clasificación de recursos para cobre (izquierda) y oro (derecha)

7.2.8 Análisis económico

Para hacer un análisis económico más sencillo, conviene realizar una transformación de leyes de cobre y oro a una ley equivalente, que en este caso es relacionada al cobre como variable principal. Esto se realiza a partir de la ecuación (57), con recuperaciones metalúrgicas para el cobre R_{cu} y oro R_{au} de 87% y 83% respectivamente, costos de venta $C_{cu} = 1500$ [USD/ton fino] y $C_{au} = 14$ [MUSD/ton fino], precios de minerales del 27 de febrero del año 2020 de 2.55 [USD/lb] para el cobre P_{cu} y 1644 [USD/t.oz] para el oro P_{au} , los que transformados a toneladas corresponden a 5622 [USD/ton] y 58 [MUSD/ton] respectivamente. Las leyes, L_{cu} y L_{au} , son transformadas a porcentajes.

$$L_{cuelq} = L_{cu} + \frac{(P_{au} - C_{au}) \times R_{au}}{(P_{cu} - C_{cu}) \times R_{cu}} L_{au} = L_{cu} + \frac{10182}{10000} * L_{au} \quad (57)$$

En base a la clasificación de recursos, el yacimiento posee los tonelajes explicitados en la Tabla 20, donde cada tonelaje tiene relacionado su ley media equivalente antes de la aplicación de una ley de corte, siendo la densidad de cada bloque de 2600 kg/m^3 y, por lo tanto, un yacimiento de 47 millones de toneladas aproximadamente. Como el cobre es el producto principal, se emplea la clasificación de recursos obtenida para este.

Tabla 20. Tonelaje y ley media para los recursos medidos, indicados e inferidos

Caso	Método	Algoritmo	Tonelaje [Mton]			Ley media equivalente [%]		
			Medido	Indicado	Inferido	Medido	Indicado	Inferido
Mejor	Sin dominios	OK	19.87	27.09	0.05	0.912	0.755	0.532
		OCK	22.83	24.15	0.03	0.903	0.746	0.525
		IDW	20.12	26.83	0.05	0.925	0.744	0.526
	Con dominios	OK	25.08	21.91	0.02	0.885	0.744	0.533
		OCK	27.05	19.96	0.00	0.887	0.728	0.000
		IDW	24.99	21.98	0.04	0.888	0.739	0.676
Más probable	Sin dominios	OK	30.53	16.33	0.15	0.900	0.743	0.737
		OCK	29.61	17.27	0.13	0.904	0.745	0.733
		IDW	30.40	16.35	0.26	0.907	0.734	0.714
	Con dominios	OK	33.61	13.27	0.13	0.888	0.730	0.750
		OCK	33.70	13.13	0.17	0.889	0.727	0.732
		IDW	33.53	13.20	0.28	0.893	0.715	0.721
Peor	Sin dominios	OK	22.01	25.00	0.00	0.924	0.755	0.000
		OCK	20.71	26.30	0.00	0.930	0.760	0.000
		IDW	22.30	24.66	0.05	0.930	0.744	0.476
	Con dominios	OK	23.43	23.53	0.05	0.891	0.769	0.425
		OCK	24.70	22.27	0.04	0.893	0.764	0.418
		IDW	23.78	23.18	0.05	0.893	0.761	0.506

Las diferencias entre la ley media equivalente y la ley correspondiente a los bloques medidos, indicados e inferidos de la base real son similares cuando se emplean y no se emplean dominios geometalúrgicos. Sin embargo, el tonelaje de los recursos medidos aumenta al emplear dominios, Tabla 21, trayendo como consecuencia una disminución de la ley media. Co-kriging tiende a resultar con más recursos medidos que kriging e inverso de la distancia, pero al cambiar de una estimación sin dominios a una estimación con dominios, las mayores diferencias se obtienen en el kriging. Además, co-kriging tiende a obtener mayores leyes medias en sus recursos medidos, mientras que el inverso de la distancia siempre posee las menores leyes medias para sus recursos medidos.

Tabla 21. Diferencia recursos medidos entre emplear dominios geometalúrgicos y no emplearlos

Algoritmo	Caso	Recursos medidos	Algoritmo	Caso	Recursos medidos	Algoritmo	Caso	Recursos medidos
Kriging Ordinario	Mejor	+5.22	Co-kriging Ordinario	Mejor	+4.22	Inverso de la distancia	Mejor	+4.87
	Probable	+3.07		Probable	+4.10		Probable	+3.13
	Peor	+1.43		Peor	+4.00		Peor	+1.49

La ley de corte equivalente, determinada en el Anexo A, es de 0.62%. Luego la cantidad de bloques económicamente extraíbles disminuye para todos los casos, Tabla 22. Como consecuencia de esto, ahora existen bloques considerados estériles y por lo tanto deducidos de la clasificación de recursos. Por consiguiente, los tonelajes disminuyen para todos los casos y las leyes medias aumentan. La ley media de los estériles es menor cuando se emplean dominios geometalúrgicos y su tonelaje mayor, manteniendo también tonelajes mayores en los recursos medidos en relación a las estimaciones sin dominios. Para el inverso de la distancia los tonelajes de estéril tienden a ser menor y al mismo tiempo la ley media de los estériles tiende a ser mayor.

Tabla 22. Tonelaje y ley media para los recursos medidos, indicados e inferidos después de filtrar por ley de corte

Caso	Método	Algoritmo	Estéril	Tonelaje [Mton]			Ley media equivalente [%]			
				Medido	Indicado	Inferido	Estéril	Medido	Indicado	Inferido
Mejor	Sin dominios	OK	6.17	18.97	21.86	0.01	0.566	0.929	0.800	0.657
		OCK	5.91	21.77	19.33	0.00	0.569	0.920	0.789	0.000
		IDW	6.80	19.27	20.94	0.00	0.557	0.941	0.796	0.750
	Con dominios	OK	8.86	22.62	15.53	0.00	0.556	0.921	0.821	0.000
		OCK	8.67	24.55	13.80	0.00	0.558	0.920	0.804	0.000
		IDW	8.88	22.54	15.56	0.03	0.550	0.924	0.817	0.705
Más probable	Sin dominios	OK	4.57	29.58	12.71	0.15	0.538	0.912	0.802	0.737
		OCK	4.52	28.78	13.57	0.13	0.538	0.914	0.802	0.733
		IDW	5.11	29.31	12.36	0.23	0.530	0.921	0.802	0.735
	Con dominios	OK	4.37	32.04	10.47	0.13	0.529	0.904	0.787	0.750
		OCK	4.33	32.20	10.30	0.17	0.529	0.905	0.785	0.732
		IDW	5.11	31.79	9.85	0.27	0.525	0.912	0.784	0.729
Peor	Sin dominios	OK	5.23	21.33	20.45	0.00	0.562	0.936	0.798	0.000
		OCK	5.18	20.09	21.74	0.00	0.562	0.942	0.802	0.000
		IDW	5.99	21.59	19.43	0.00	0.556	0.943	0.794	0.000
	Con dominios	OK	7.00	21.98	18.03	0.00	0.557	0.913	0.834	0.000
		OCK	6.93	23.15	16.93	0.00	0.558	0.915	0.829	0.000
		IDW	7.72	22.17	17.12	0.00	0.554	0.918	0.835	0.000

De manera similar, los tonelajes de recursos medidos aumentan al emplear dominios geometalúrgicos, Tabla 23.

Tabla 23. Diferencia recursos medidos entre emplear dominios geometalúrgicos y no emplearlos

Algoritmo	Caso	Recursos medidos	Algoritmo	Caso	Recursos medidos	Algoritmo	Caso	Recursos medidos
Kriging Ordinario	Mejor	+3.65	Co-kriging Ordinario	Mejor	+2.78	Inverso de la distancia	Mejor	+3.27
	Probable	+2.46		Probable	+3.42		Probable	+2.49
	Peor	+0.64		Peor	+3.06		Peor	+0.58

Tabla 24. Ingresos, costos y beneficios en MUSD

Caso	Método	Algoritmo	Medido			Indicado			Beneficio total	Razón M-I
			Ingreso	Costo	Beneficio	Ingreso	Costo	Beneficio		
Mejor	Sin dominios	OK	632.0	332.0	300.0	627.0	382.6	244.4	544.4	1.2
		Ock	718.1	381.0	337.2	547.2	338.3	208.9	546.1	1.6
		IDW	650.8	337.3	313.5	597.5	366.4	231.1	544.6	1.4
	Con dominios	OK	746.9	395.9	351.0	457.2	271.8	185.4	536.4	1.9
		Ock	810.4	429.6	380.8	397.7	241.5	156.3	537.1	2.4
		IDW	747.2	394.5	352.7	455.8	272.3	183.5	536.2	1.9
Más probable	Sin dominios	OK	967.4	517.7	449.7	365.8	222.5	143.3	593.0	3.1
		Ock	943.7	503.7	439.9	390.3	237.5	152.8	592.7	2.9
		IDW	968.0	512.9	455.1	355.4	216.3	139.1	594.3	3.3
	Con dominios	OK	1039.2	560.7	478.5	295.5	183.3	112.2	590.8	4.3
		Ock	1045.5	563.6	481.9	290.0	180.3	109.7	591.6	4.4
		IDW	1040.4	556.4	484.0	277.0	172.3	104.6	588.6	4.6
Peor	Sin dominios	OK	716.3	373.3	343.0	585.5	357.9	227.5	570.5	1.5
		Ock	678.8	351.6	327.2	625.2	380.5	244.7	571.9	1.3
		IDW	730.1	377.9	352.3	553.3	340.0	213.3	565.6	1.7
	Con dominios	OK	719.8	384.6	335.3	539.6	315.6	224.0	559.3	1.5
		Ock	759.6	405.2	354.4	503.3	296.3	207.0	561.4	1.7
		IDW	729.6	388.0	341.6	512.9	299.7	213.3	554.9	1.6

Las ingresos y costos son calculados para todos los casos empleando ecuaciones y valores del Anexo A y son resumidos en la Tabla 24. Los beneficios obtenidos corresponden al valor económico del yacimiento asumiendo que todos los bloques están inmediatamente disponibles para remover y pueden ser extraídos instantáneamente, sin procesar los estériles. El beneficio total es menor para todos los casos al aplicar dominios geometalúrgicos aproximadamente en un 1.2%, pero

los beneficios producidos por los recursos medidos en cada caso son mayores al emplear dominios geometalúrgicos, evidenciado a través de la razón medido-indicado. Esto indica que esta metodología genera menor incertidumbre en las estimaciones. Con respecto a los algoritmos empleados, cuando se emplean dominios geometalúrgicos con co-kriging además del beneficio ser levemente superior al kriging, una mayor parte de este beneficio proviene de recursos medidos.

Al comparar los beneficios estimados con los beneficios reales, Tabla 25, se tiene que empleando dominios geometalúrgicos en el caso más probable y en el peor las diferencias entre los beneficios reales y los beneficios de las diferentes combinaciones entre métodos y algoritmos son menores, es decir, los estimados económicos son más precisos.

Tabla 25. Valores de los beneficios reales [MUSD] y diferencias con beneficios estimados

Caso	Método	Algoritmo	Medido		Indicado		Total		Tonelajes [Mton]
			Real	Diferencia [%]	Real	Diferencia [%]	Real	Diferencia [%]	
Mejor	Sin dominios	OK	308.2	2.7	250.6	2.5	558.8	2.6	40.8
		OCK	349.6	3.6	210.7	0.8	560.3	2.5	41.1
		IDW	312.2	-0.4	241.5	4.3	553.7	1.6	40.2
	Con dominios	OK	365.4	3.9	189.7	2.2	555.1	3.4	38.2
		OCK	397.2	4.1	158.9	1.7	556.1	3.4	38.3
		IDW	362.5	2.7	191.2	4.0	553.7	3.2	38.1
Más probable	Sin dominios	OK	442.1	-1.7	140.3	-2.1	582.4	-1.8	42.3
		OCK	432.8	-1.6	150.1	-1.8	582.9	-1.7	42.4
		IDW	439.2	-3.6	138.5	-0.4	577.6	-2.9	41.7
	Con dominios	OK	474.9	-0.8	111.0	-1.1	585.8	-0.8	42.5
		OCK	477.1	-1.0	108.4	-1.2	585.5	-1.0	42.5
		IDW	474.2	-2.1	105.7	1.0	579.8	-1.5	41.6
Peor	Sin dominios	OK	334.6	-2.5	222.3	-2.3	557.0	-2.4	41.8
		OCK	319.4	-2.4	238.1	-2.8	557.4	-2.6	41.8
		IDW	339.2	-3.8	211.8	-0.7	551.0	-2.6	41.0
	Con dominios	OK	333.5	-0.5	216.2	-3.6	549.7	-1.7	40.0
		OCK	350.9	-1.0	199.2	-3.9	550.2	-2.0	40.1
		IDW	336.5	-1.5	208.0	-2.6	544.5	-1.9	39.3

7.3 Análisis de resultados

El estudio del mejor, peor y caso más probable es realizado en torno a las diferencias de RMSE y precisión ya que carece de sentido revisar dichos casos para los valores de RMSE más bajos o

precisión más altos, porque estos no necesariamente son los casos donde mejor se aprecia el efecto de emplear dominios geometalúrgicos, pues podrían tratarse de casos en que ambas metodologías (no emplear o emplear dominios geometalúrgicos en la estimación) posean buenos desempeños.

Las estimaciones realizadas empleando los algoritmos de inverso de la distancia, kriging y co-kriging son realizadas con una mínima intervención. En el caso del inverso de la distancia los parámetros modificados son el número mínimo de vecinos de 10 y el valor de p , mientras que la distancia del vecindario se considera infinita. Para el kriging y el co-kriging se fijan los variogramas a través de una función en R que los ajusta a través de una minimización de los errores entre el modelo escogido con el variograma real, sin tomar en cuenta anisotropías y se fija un límite de 10 observaciones que participen en el valor del punto estimado, por temas de tiempo de procesamiento. Dicho esto, los resultados finales obtenidos en la comparación de la Tabla 25 pudiesen quizás haber sido más precisos.

En la Tabla 26 se presentan las medidas estadísticas principales de la ley real de los sitios a estimar para los recursos medidos. Cabe destacar que el error medio en el mejor caso es mayor cuando se emplean dominios geometalúrgicos. Esto concuerda con que el beneficio económico sea más realista cuando no se emplean dominios geometalúrgicos e indica que la disminución de RMSE no implica un menor error medio y viceversa. Además, la mayor diferencia entre el beneficio real y el estimado ocurre para el mejor caso y este es precisamente el que tiene problemas entre la distribución de las muestras con la distribución de la base de validación para el cobre, lo que en un caso real no se puede determinar, por lo que la ocurrencia de esto es aleatoria. Para el caso más probable y el peor, el RMSE, los errores medios y el error absoluto medio disminuyen al emplear dominios geometalúrgicos para todos los algoritmos.

Los valores predichos por el kriging ordinario y co-kriging ordinario son sumamente parecidos. Esto es porque co-kriging es idealmente empleado cuando existe una variable principal y una o más co-variables, donde la variable principal posee menos información disponible que las covariables (ya sea porque es más cara o difícil de medir) y existe una correlación importante entre ellas. Ninguno de los dos casos está presente en alguna de las 100 iteraciones, por lo que co-kriging tiende a estimar de manera muy similar a kriging. Sin embargo, la varianza es menor para co-kriging gracias a la adición de variogramas cruzados.

Tabla 26. Medidas estadísticas principales para la ley equivalente después de filtrar por ley de corte

Caso	Método	Algoritmo	Tonelaje	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio
Mejor	Sin dominios	OK	18.97	0.929	0.154	0.923	0.163	0.0120	0.118
		OCK	21.77	0.920	0.153	0.913	0.169	0.0160	0.123
		IDW	19.27	0.941	0.167	0.928	0.168	-0.0018	0.120
	Con dominios	OK	22.62	0.921	0.188	0.902	0.132	0.0177	0.099
		OCK	24.55	0.920	0.184	0.902	0.136	0.0186	0.102
		IDW	22.54	0.924	0.190	0.903	0.136	0.0121	0.102
Más probable	Sin dominios	OK	29.58	0.912	0.135	0.902	0.135	-0.0072	0.103
		OCK	28.78	0.914	0.134	0.904	0.135	-0.0069	0.103
		IDW	29.31	0.921	0.143	0.910	0.143	-0.0152	0.109
	Con dominios	OK	32.04	0.904	0.142	0.891	0.128	-0.0032	0.098
		OCK	32.20	0.905	0.141	0.892	0.128	-0.0041	0.098
		IDW	31.79	0.912	0.148	0.898	0.134	-0.0086	0.103
Peor	Sin dominios	OK	21.33	0.936	0.147	0.923	0.155	-0.0109	0.114
		OCK	20.09	0.942	0.148	0.928	0.156	-0.0108	0.114
		IDW	21.59	0.943	0.154	0.935	0.163	-0.0168	0.119
	Con dominios	OK	21.98	0.913	0.158	0.907	0.129	-0.0023	0.098
		OCK	23.15	0.915	0.158	0.909	0.131	-0.0042	0.100
		IDW	22.17	0.918	0.160	0.913	0.134	-0.0064	0.102

Si bien el inverso de la distancia es más simple de aplicar que los métodos geoestadísticos, carece de herramientas para tomar en cuenta la correlación espacial de las muestras. Esto impacta en la complejidad que pueden adquirir los resultados. Sin embargo, cuando la variable a estimar es errática (como lo es el oro), tiende a obtener resultados mejores que si esta no lo fuese, aun así, sin poder superar a los métodos geoestadísticos a menos que no haya suficiente cantidad de muestras para estimar un variograma útil. Esto se ve en la Figura 27 donde el incremento de RMSE entre emplear y no emplear dominios geometalúrgicos no es tan brusco entre el cobre y el oro para el inverso de la distancia como lo es para los métodos de kriging.

Con respecto a los casos especiales que fueron mencionados en el título 7.2.5, el primero, que corresponde a una estimación donde la variable oro presenta menor RMSE al ser estimada sin dominios geometalúrgicos, a pesar de ser simulada en los tres dominios de manera independiente, presenta muy buena continuidad espacial, en especial en las fronteras de estos, Figura 34, por lo que los dominios formados no representan conjuntos de datos disjuntos y por ende contradicen a la

definición de dominio. Luego, en esta estimación el RMSE es menor cuando se estima sin dominios geometalúrgicos. Esto último indica que forzar dominios geometalúrgicos no siempre es la solución.

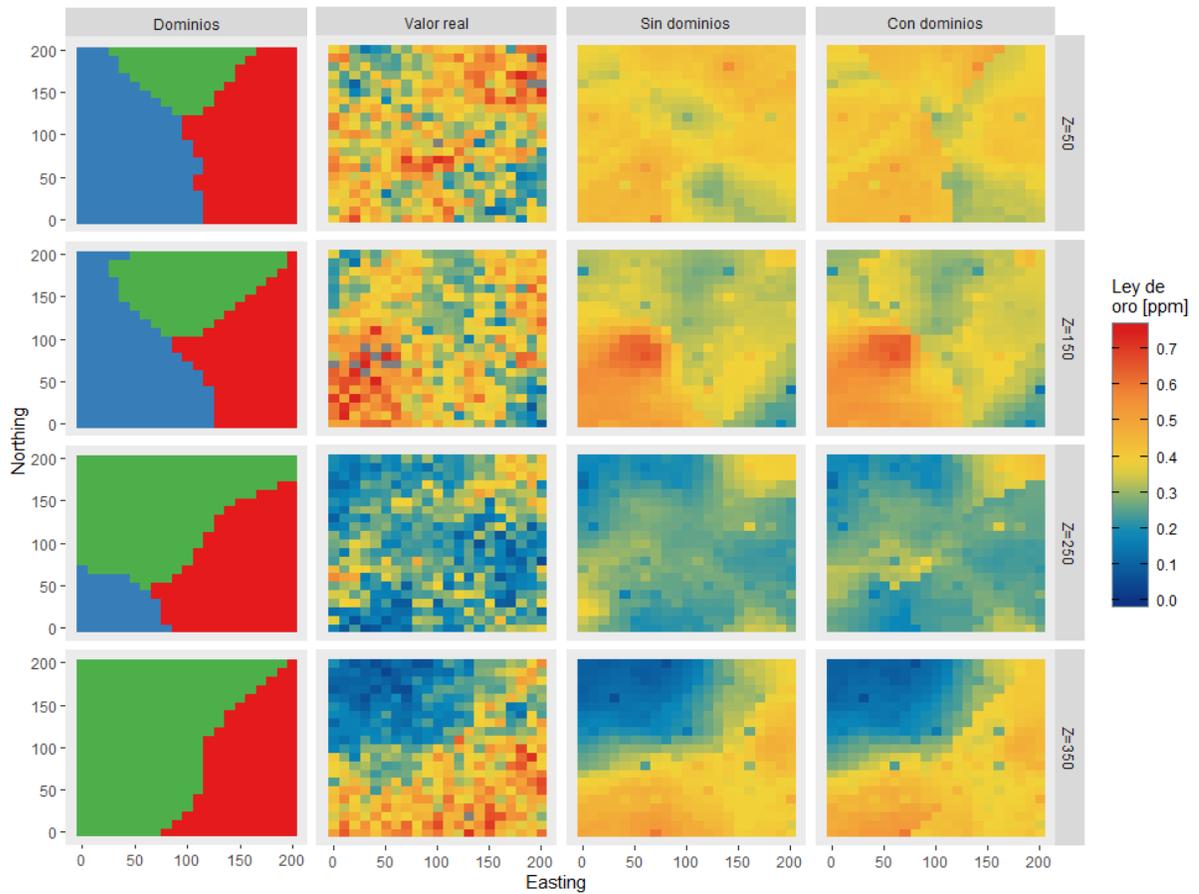


Figura 34. Comparación estimación para kriging ordinario, estimación oro 26

Ocurre algo similar para el otro caso donde la precisión era mayor al estimar sin dominios geometalúrgicos.

8 Caso Real

8.1 Metodología de análisis

El estudio de la base de datos reales es realizada posterior a los aprendizajes obtenidos en el caso simulado, donde se genera una idea de lo que se puede esperar de los resultados y es aplicado un análisis similar, con la ventaja de conocer los datos en toda el área de estudio.

El objetivo de el caso real es lograr construir un modelo de bloques y estimar los valores de estos bloques a partir de la información primaria, obteniendo leyes minerales y dominios correspondientes, con el fin de comprobar la hipótesis.

Este análisis es estructurado en las siguientes etapas:

- Descripción y análisis de la base de datos: se realiza un análisis detallado de la base de datos en relación a sus características estadísticas principales, además de una limpieza y depuración de datos.
- Construcción de modelo de bloques: se genera un modelo de bloques donde estimar a partir de las variables de la base de datos. Se extienden los dominios geometalúrgicos de la base real mediante SVM (*“support vector machine”*) sobre el modelo de bloques.
- Aplicación de algoritmos de estimación: se particiona la base original en una proporción 60% para entrenamiento y 40% para validación, donde en esta última se estima mediante los diferentes algoritmos y métodos con el fin de obtener alguna medida de la calidad de estimación. Posteriormente se realiza la estimación del modelo de bloques.
- Evaluación estadística de los resultados: se realiza una comparación entre los resultados obtenidos en la fase de estimación sobre la base de validación con los datos de validación reales y los resultados obtenidos al estimar sobre el modelo de bloques.
- Categorización de recursos: se lleva a cabo una categorización de los recursos minerales en base a la incertidumbre obtenida por los diferentes algoritmos de estimación, siendo

esto evaluado con respecto a la precisión de la estimación. Se reportan medidas estadísticas principales para cada caso.

- Análisis económico: en base a los resultados obtenidos de la clasificación de recursos y a diferentes escenarios de riesgo se lleva a cabo un análisis económico y es evaluada la diferencia de este con un análisis equivalente sobre la base de datos real, para la base de validación.

8.2 Aplicación

La aplicación de los algoritmos discutidos en el marco teórico es llevada a cabo según lo planteado en la metodología de simulación sobre el caso real.

8.2.1 Descripción y análisis de la base de datos

La base de datos a estudiar está compuesta por 7997 observaciones de las variables ley de cobre y 3207 de ley de oro, además de acompañados de su posición en el espacio cartesiano con coordenadas UTM y el respectivo dominio geometalúrgico al cual pertenece cada dato, siendo cuatro la cantidad de dominios existentes, Figura 35, donde estos poseen un índice de silueta, Anexo B, de 0.26, indicando que existe una estructura de agrupamiento débil posiblemente artificial. Todos los datos que presentan un valor de ley de oro están asociados a un valor de ley de cobre, por lo que en total la base posee 7997 datos. La correlación cuando ambas leyes están presentes es de 0.77.

El espacio de trabajo está encasillado por un paralelepípedo de esquinas con coordenadas (x, y, z) de $(674036.2, 424650.9, 255)$ metros para la esquina inferior y $(675832.4, 425907.7, 367.5)$ metros para la esquina superior lo que resulta en un rango de $(1796.25, 1256.78, 112.5)$ metros en las tres direcciones.

Primero se estudian datos duplicados, encontrándose siete pares de estos. Todos los pares de datos duplicados son eliminados, resultando en un total de 14 datos deducidos de la base total.

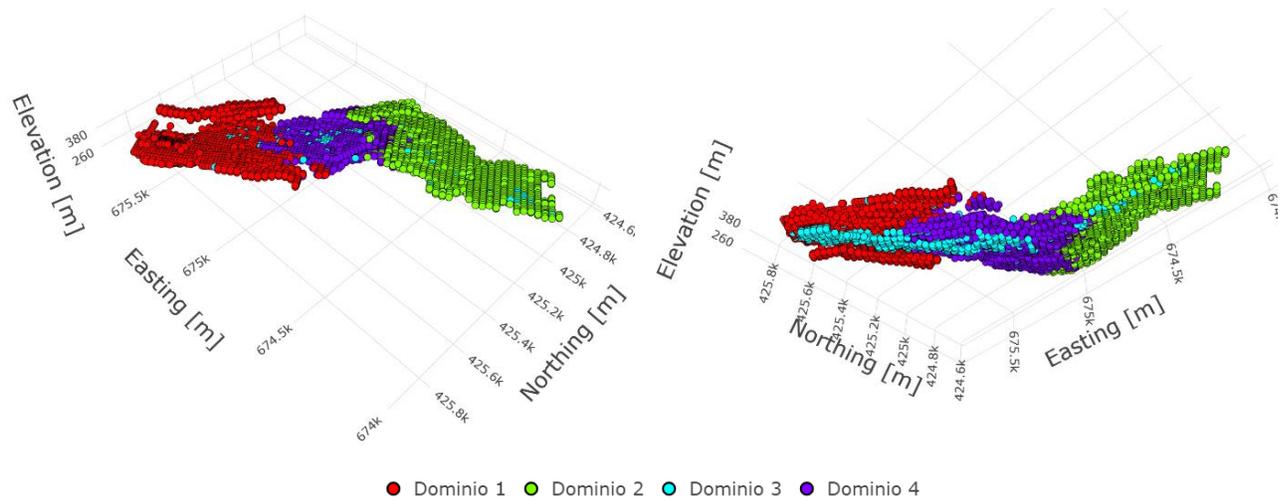


Figura 35. Vista superior (izquierda) e inferior (derecha) de los dominios geometalúrgicos

Las estadísticas principales de la base de datos antes de eliminar outliers se resumen en la Tabla 27. Cabe destacar que el dominio número tres posee leyes más altas tanto de oro como de cobre, además de poseer la mayor desviación estándar producto de esto.

Tabla 27. Estadísticas descriptivas base de datos real

Variable	Media	Desviación estándar	Mediana	Rango	Mínimo	Máximo	Número de datos
Oro	0.31	0.28	0.24	3.70	0.00	3.70	3207
Cobre	0.35	0.23	0.32	1.72	0.00	1.72	7997
Dominios							
Oro 1	0.21	0.16	0.17	1.12	0.00	1.12	870
Oro 2	0.21	0.18	0.15	0.99	0.00	0.99	943
Oro 3	0.71	0.45	0.69	3.65	0.05	3.70	366
Oro 4	0.36	0.20	0.32	1.09	0.02	1.11	1028
Cobre 1	0.28	0.18	0.24	0.88	0.00	0.88	2170
Cobre 2	0.27	0.18	0.22	0.99	0.00	1.00	2350
Cobre 3	0.71	0.26	0.73	1.68	0.04	1.72	908
Cobre 4	0.37	0.15	0.36	0.95	0.00	0.96	2569

En la Figura 36, las distribuciones del dominio tres para cobre y oro poseen valores más altos que los demás y existen algunos datos que son candidatos a outliers. Es por ello que se realiza un análisis de outliers, ya que si alguno de estos forma parte posteriormente de la base de entrenamiento va a impactar de sobre manera los resultados finales de la estimación.

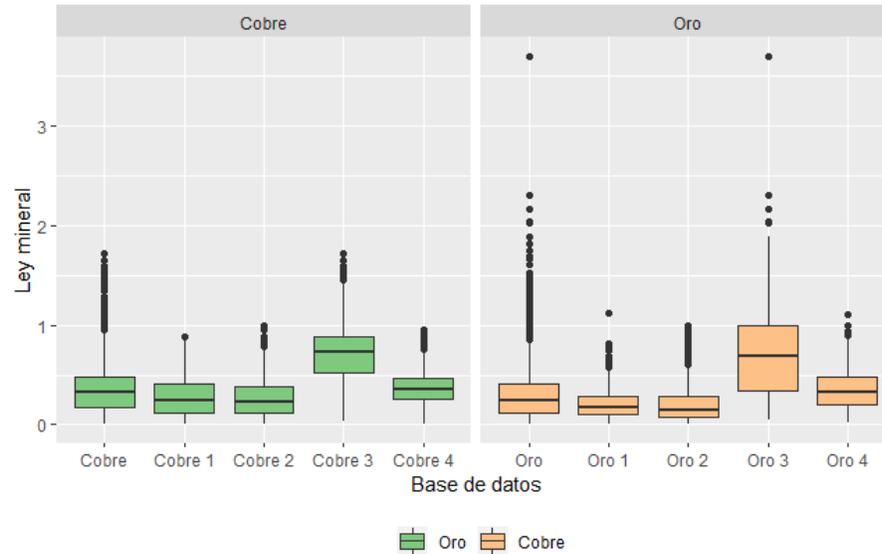


Figura 36. Comparación distribuciones entre dominios y la base completa

Luego, se realiza un análisis de datos atípicos a través de DBScan, con $eps = 1.4$ y $minPts = 3$, donde se retiran un total de 13 datos atípicos, cinco para la base de datos que contiene cobre y ocho para la base de datos que contiene oro, resultando finalmente en bases de 7978 y 3185 datos respectivamente. La ventaja de emplear DBScan por sobre metodologías tradicionales como el análisis a través de la desviación estándar es que el primero es multivariable, y por lo tanto permite estudiar el comportamiento de los datos con respecto a su vecindad. Al remover dichos datos, las estadísticas finales de las bases de datos se resumen en la Tabla 28. Los cambios más notorios se producen en la desviación estándar y en el rango, donde hubo reducciones en el dominio tres de oro y en el oro global.

Dentro de los datos filtrados había una ley de oro de 3.7 [ppm], que, si bien no es imposible de obtener en un yacimiento, al ser escogido como parte de la base de entrenamiento afecta la estimación considerablemente, Anexo D.

La Figura 37 muestra el cambio en las distribuciones luego de filtrar la base de datos, notándose claramente la diferencia en la representatividad de las distribuciones, donde se eliminaron en total un 0.69% y un 0.24% de los datos para oro y cobre respectivamente.

Tabla 28. Estadísticas descriptivas base de datos final después de filtrar outliers y duplicados

Variable	Media	Desviación estándar	Mediana	Rango	Mínimo	Máximo	Número de datos
Oro	0.31	0.27	0.24	2.31	0.00	2.31	3185
Cobre	0.35	0.23	0.32	1.72	0.00	1.72	7978
Dominios							
Oro 1	0.21	0.15	0.17	0.82	0.00	0.82	865
Oro 2	0.20	0.18	0.15	0.99	0.00	0.99	933
Oro 3	0.70	0.42	0.69	2.26	0.05	2.31	359
Oro 4	0.36	0.20	0.32	1.09	0.02	1.11	1028
Cobre 1	0.28	0.18	0.24	0.88	0.00	0.88	2164
Cobre 2	0.27	0.18	0.22	0.99	0.00	1.00	2345
Cobre 3	0.71	0.26	0.73	1.68	0.04	1.72	900
Cobre 4	0.37	0.15	0.36	0.95	0.00	0.96	2569

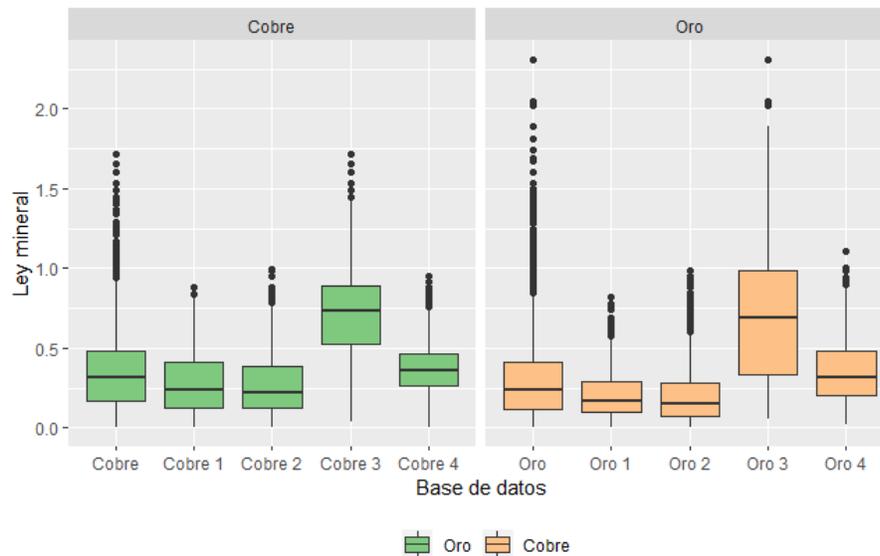


Figura 37. Comparación distribuciones entre dominios y la base completa después de filtrar outliers y duplicados

Luego de filtrar la base de datos, la correlación aumenta a 0.79, mientras que las correlaciones para cada dominio son de 0.79, 0.81, 0.63 y 0.61 para los dominios del uno al cuatro, respectivamente.

8.2.2 Construcción del modelo de bloques

El modelo de bloques donde se debe estimar está compuesto por 46666 bloques de dimensiones $20 \times 20 \times 10$ metros, siendo un volumen total de $4000 [m^3]$ por bloque y $186.7 \times 10^6 [m^3]$ en total, Figura 38. El modelo de bloques está inscrito en un paralelepípedo de esquinas inferior y superior de coordenadas (x,y,z) iguales a $(673946,424588,249)$ y $(675906,425968,399)$, lo que significa que un rango de $(1960,1380,150)$ metros en las tres direcciones principales.

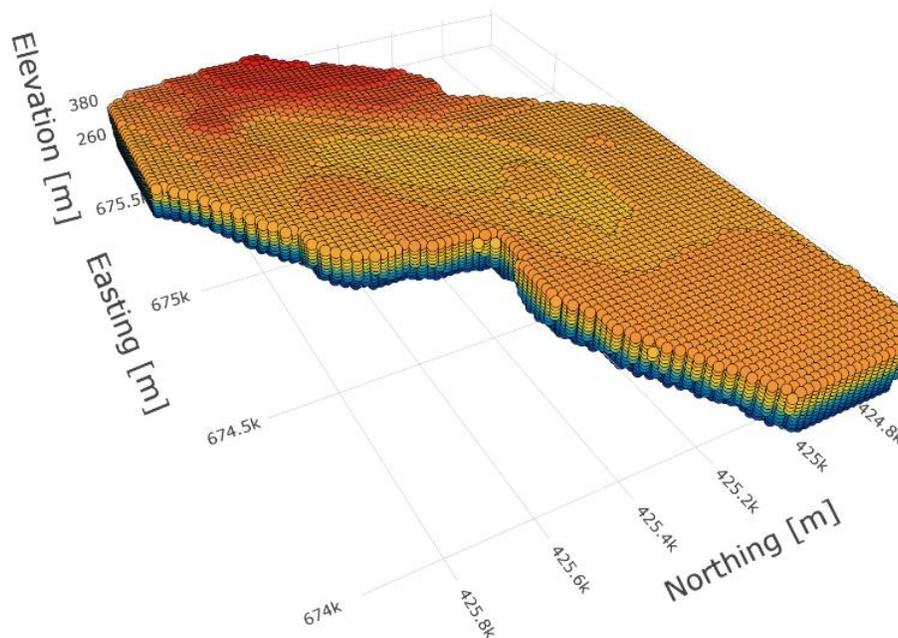


Figura 38. Modelo de bloques base real

Para extender los dominios geometalúrgicos sobre el modelo de bloques se emplea la herramienta “*support vector machine*” o SVM, que a través de hiperplanos divide el modelo de bloques tratando de seguir la distribución y forma de los dominios geometalúrgicos de la base de datos original. Los resultados de este procedimiento se muestran en la Figura 39, donde el dominio tres se encuentra debajo del dominio uno y cuatro.

Los parámetros empleados en la optimización del SVM fueron iterados entre $\gamma = (2^{-15}, \dots, 2^2)$ y $C = (2^{-5}, \dots, 2^{15})$ cada dos exponentes, calculando el porcentaje de acierto en cada una de las 110 iteraciones. Los parámetros óptimos son $\gamma = 2^{-7}$ y $C = 2^{13}$, resultando en un porcentaje de acierto del 99% sobre la base de datos real (no sobre el modelo de bloques).

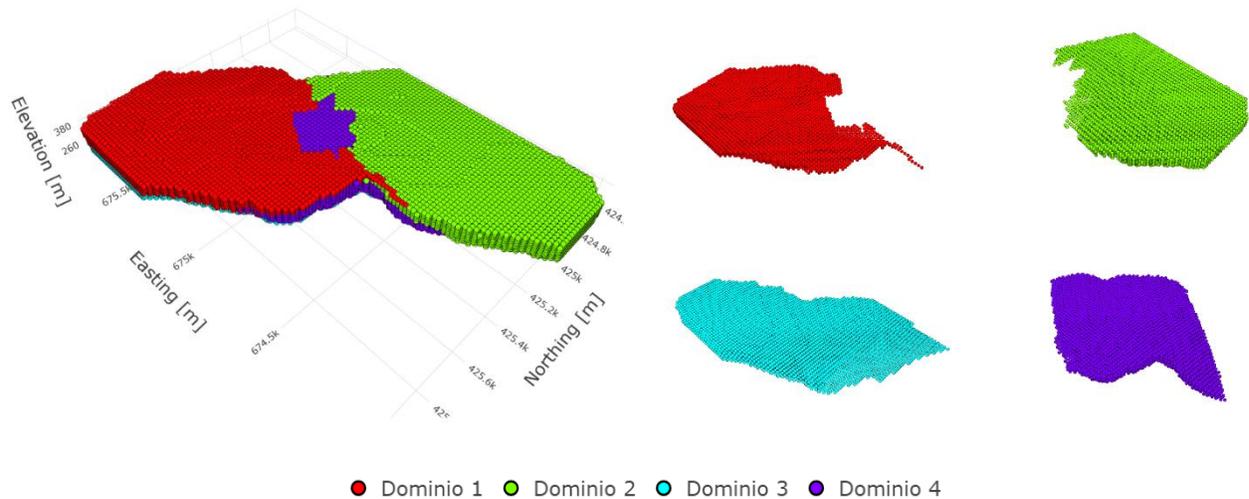


Figura 39. Dominios geometalúrgicos modelo de bloques base real

Tabla 29. Dominios reales vs dominios SVM

		Dominios SVM			
		1	2	3	4
Dominios reales	1	1514	0	1	1
	2	0	1636	2	3
	3	6	2	610	14
	4	6	12	7	1773

8.2.3 Aplicación de algoritmos de estimación

Producto del tamaño de la base de datos, es factible particionarla para generar una base de entrenamiento y otra de validación. Esto permite comprobar y dimensionar la capacidad predictiva de los algoritmos.

La base de datos se particiona a una razón de 60/40, donde el primero corresponde a la base de entrenamiento y el último a la base de validación. Para la base de datos de cobre se obtienen 4788 y 3190 datos de entrenamiento y de validación respectivamente, mientras que para la base de datos de oro se obtienen 1915 y 1270 datos.

Los algoritmos de empleados para estimar son: kriging ordinario, co-kriging ordinario e inverso de la distancia.

Kriging ordinario

Con los datos de entrenamiento para oro y cobre se determina el variograma experimental en las direcciones $\alpha = (0^\circ, 45^\circ, 90^\circ, 135^\circ)$ para evaluar posibles comportamientos anisotrópicos, resultando en anisotropías para cobre y oro en 45° en los dominios uno, tres y cuatro, mientras que el dominio dos presenta una anisotropía en 90° ambas con respecto a la dirección norte en sentido horario. La base completa presenta anisotropía en 45° tanto para oro como para cobre. Los variogramas teóricos para cada variable y dominio se resumen en la Tabla 30, mientras que los variogramas para la base completa están en la Tabla 31. Los variogramas son ajustados mediante prueba y error, viendo cómo cambia el RMSE correspondiente a cada estimación hasta encontrar los mejores variogramas para cada dominio.

Tabla 30. Variogramas para cada variable y dominio

Zona	Oro	Cobre
Dominio 1	$\gamma(h) = 0.010 + 0.010 \text{Exp}(100,100) + 0.007 \text{Exp}(Inf, 100)$	$\gamma(h) = 0.005 + 0.029 \text{Exp}(40,40)$
Dominio 2	$\gamma(h) = 0.010 + 0.022 \text{Sph}(100,100) - 0.005 \text{Sph}(Inf, 100)$	$\gamma(h) = 0.010 + 0.024 \text{Sph}(75,75) - 0.007 \text{Sph}(Inf, 75)$
Dominio 3	$\gamma(h) = 0.050 + 0.070 \text{Exp}(80,80) + 0.120 \text{Exp}(Inf, 80)$	$\gamma(h) = 0.010 + 0.033 \text{Exp}(65,65) + 0.010 \text{Exp}(800,80)$
Dominio 4	$\gamma(h) = 0.012 + 0.032 \text{Exp}(35,35)$	$\gamma(h) = 0.007 + 0.014 \text{Exp}(25,25) + 0.005 \text{Exp}(200,25)$

Tabla 31. Variogramas para cada variable base completa

Oro	Cobre
$\gamma(h) = 0.025 + 0.022 \text{Exp}(100,100) + 0.033 \text{Exp}(200,100)$	$\gamma(h) = 0.008 + 0.042 \text{Exp}(30,30)$

La restricción empleada en la aplicación del kriging ordinario es un número máximo de datos de 30 para la estimación de cada dato, ya que es a partir de aproximadamente este número que el cambio en el RMSE es mínimo al agregar una observación más en la estimación.

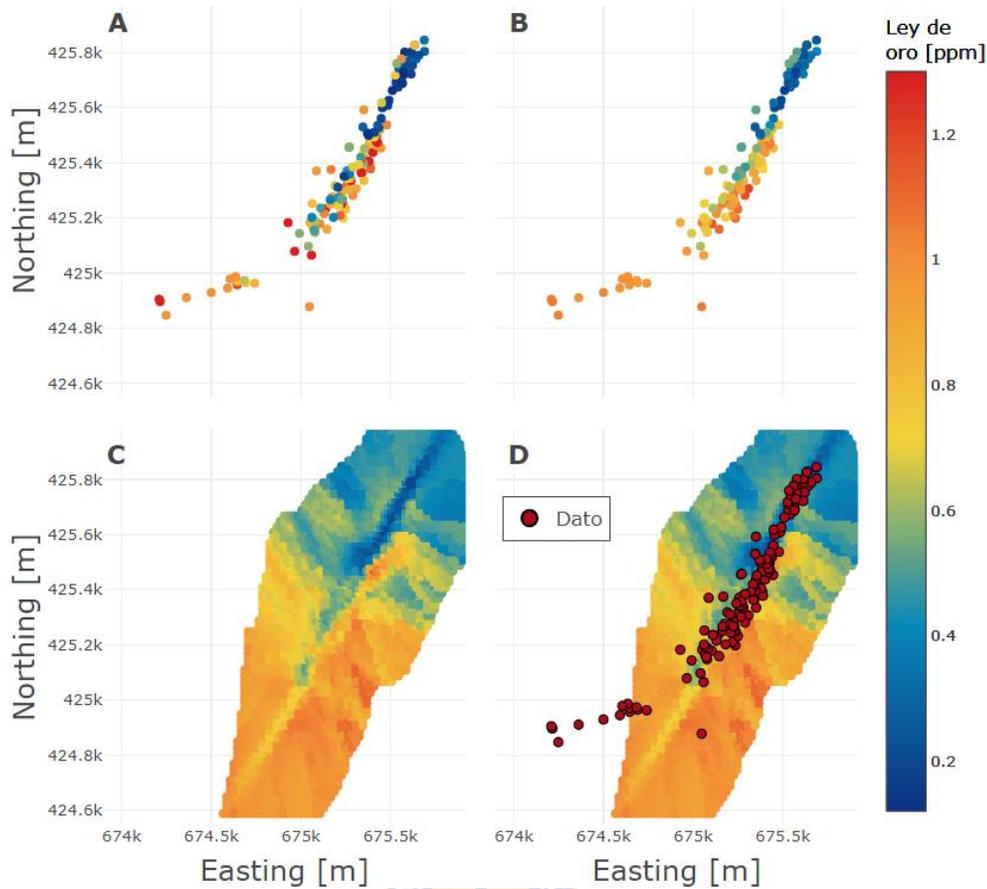


Figura 40. Estimación por kriging ordinario para oro, dominio tres

La estimación es realizada sobre los dominios geometalúrgicos de la base de validación por separado, luego sobre toda la base de validación, y, posteriormente, efectuada sobre el modelo de bloques tanto por dominios como sobre todo este. Ejemplos se encuentran en la Figura 40, donde se estima el dominio tres de la base de validación de oro (B), el modelo de bloques del mismo elemento (C) y luego se compara la ubicación de este dominio en comparación con los datos originales (D). En (B) se ve como el kriging suaviza los valores reales que muestra en (A), tanto los valores bajos (en azul) como los altos (en rojo) son afectados por dicho efecto. Además, luego de aplicar SVM sobre el modelo de bloques, algunos datos quedan fuera de este (D), debido a que la densidad de datos de los grupos vecinos es mayor y que particularmente los datos del dominio tres están distribuidos a lo largo de toda la base de datos.

Co-kriging ordinario

Se emplean los mismos variogramas que para el kriging ordinario, Tabla 30 y Tabla 31, ya que las zonas de estimación son las mismas. Notar que el dominio tres es el que posee mayor variabilidad y mayores leyes.

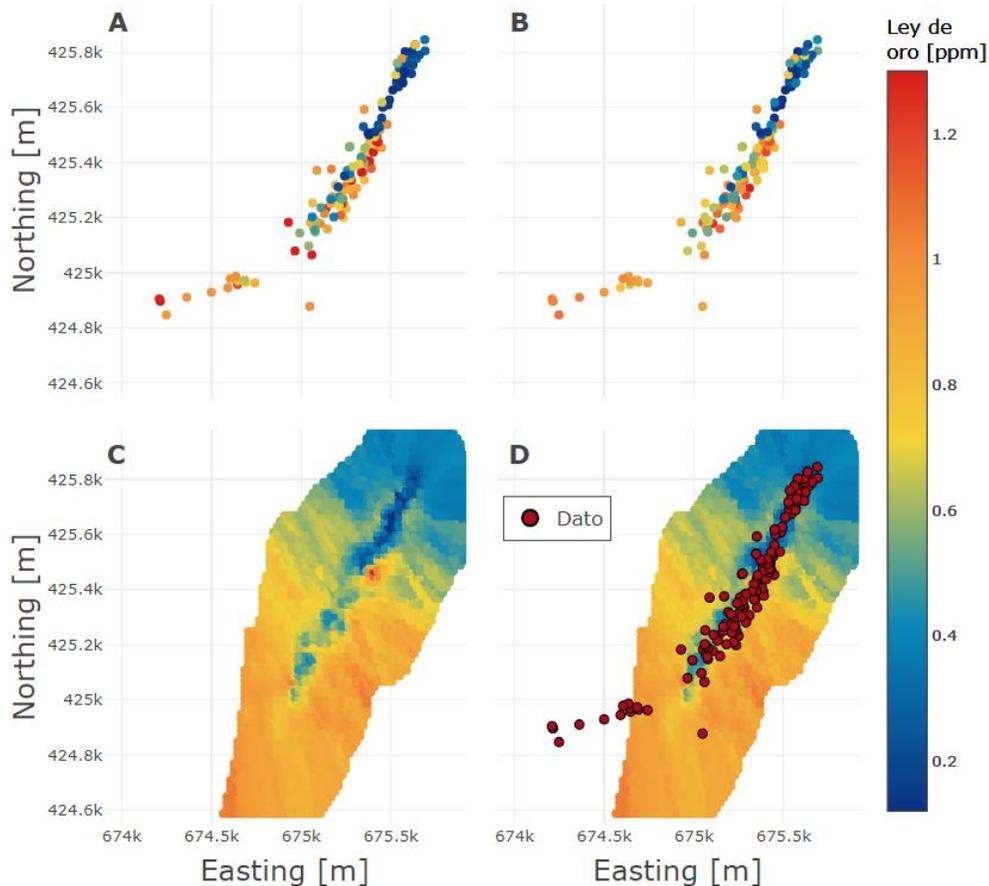


Figura 41. Estimación por co-kriging para oro, dominio tres

La estimación es realizada sobre los dominios geometalúrgicos de la base de validación por separado, luego sobre toda la base de validación, y, posteriormente, efectuada sobre el modelo de bloques tanto por dominios como sobre todo este. Las mismas restricciones que con el kriging ordinario son aplicadas con respecto a la vecindad de estimación.

La Figura 41 muestra nuevamente la estimación para la variable oro en el dominio tres. La estimación a simple vista es mejor al comparar la base de validación estimada (B) con la base original (A), ya que preserva mejor el patrón original de esta que el kriging ordinario. También se

nota una diferencia en la estimación en (C), donde el comportamiento de la ley de oro parece ser más complejo con respecto a (C) de la figura anterior.

Inverso de la distancia

En base a la base de datos de validación se estiman los valores de p que minimizan el RMSE para cada simulación, Tabla 32.

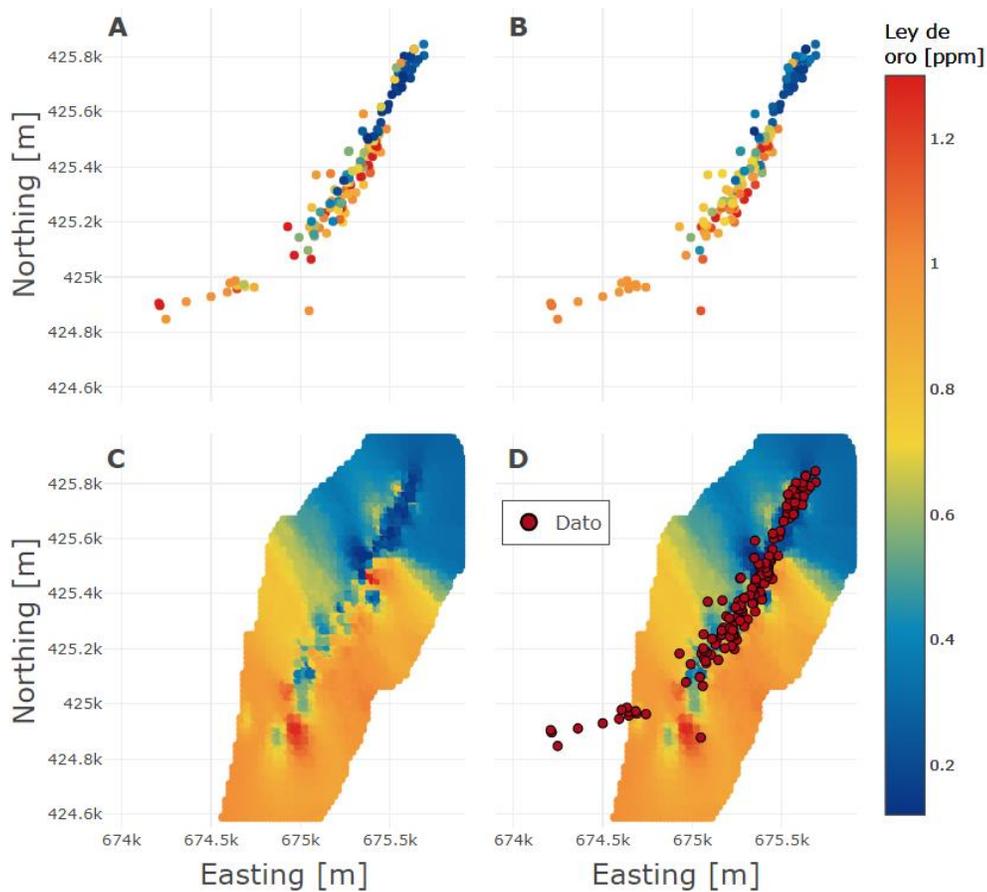


Figura 42. Estimación por inverso de la distancia para oro, dominio tres

Tabla 32. Valores de p óptimos para el inverso de la distancia

	Dominio 1	Dominio 2	Dominio 3	Dominio 4	Base Completa
Oro	3.07	3.08	5.54	3.27	2.86
Cobre	3.25	2.86	4.11	5.56	3.07

La restricción impuesta sobre la estimación es que la cantidad máxima de datos que participen en esta sea de 30 individuos.

Finalmente, la Figura 42, muestra nuevamente la estimación de la variable oro en el dominio tres. Los resultados son menos complejos que para las simulaciones de kriging, sobre todo en (C), donde, después de cierta distancia a los datos, los estimados siguen un valor uniforme hasta llegar al borde del dominio. Sin embargo, los valores estimados en (B) parecen tener valores similares a los de (A), más aún que aquellos obtenidos por kriging ordinario.

Errores cuadráticos medios

La raíz del error cuadrático medio (RMSE) de cada dominio, Tabla 33, muestra que el dominio que presenta errores más altos es el dominio 3, producto de los valores más altos de ley que contiene tanto para oro como para cobre, siendo la media del oro el triple en el dominio tres que en los otros dominios, mientras que el cobre en el dominio tres tiene como media un poco más del doble de la concentración media de los otros dominios.

Tabla 33. RMSE para cada dominio

Variable	Algoritmo	Dominio 1	Dominio 2	Dominio 3	Dominio 4
Oro	OK	0.127	0.119	0.289	0.170
	OCK	0.109	0.107	0.276	0.144
	IDW	0.129	0.142	0.310	0.174
Cobre	OK	0.120	0.111	0.203	0.122
	OCK	0.116	0.115	0.216	0.118
	IDW	0.126	0.115	0.206	0.124

El algoritmo que arroja menos RMSE para el oro es el co-kriging ordinario (OCK), con una amplia ventaja, seguido del kriging ordinario (OK) y el inverso de la distancia (IDW) que obtuvieron resultados similares. Mientras que para el cobre el OCK y OK no presentan una diferencia significativa en sus RMSE, mientras que IDW posee valores levemente mayores.

Para los resultados globales, el emplear dominios geometalúrgicos es superior en todos los casos, Tabla 34, con disminuciones de RMSE de por lo menos 9.6%, siendo estas disminuciones mayores para el oro en el caso de los algoritmos de kriging y para el cobre en el caso del inverso de la distancia. El algoritmo que mejor desempeño posee es el co-kriging ordinario, solamente siendo igualado en la estimación de cobre con dominios geometalúrgicos.

Tabla 34. RMSE cuando se emplean y no se emplean dominios geometalúrgicos

Variable	Algoritmo	Sin dominios	Con dominios	Disminución [%]
Oro	OK	0.199	0.165	17.2
	OCK	0.167	0.148	11.2
	IDW	0.203	0.179	11.7
Cobre	OK	0.149	0.130	12.8
	OCK	0.145	0.131	9.6
	IDW	0.160	0.140	12.8

8.2.4 Evaluación estadística

La correlación entre ambas variables es alta, Tabla 35, y las diferencias entre las correlaciones de la base de entrenamiento sólo son considerables en el dominio tres.

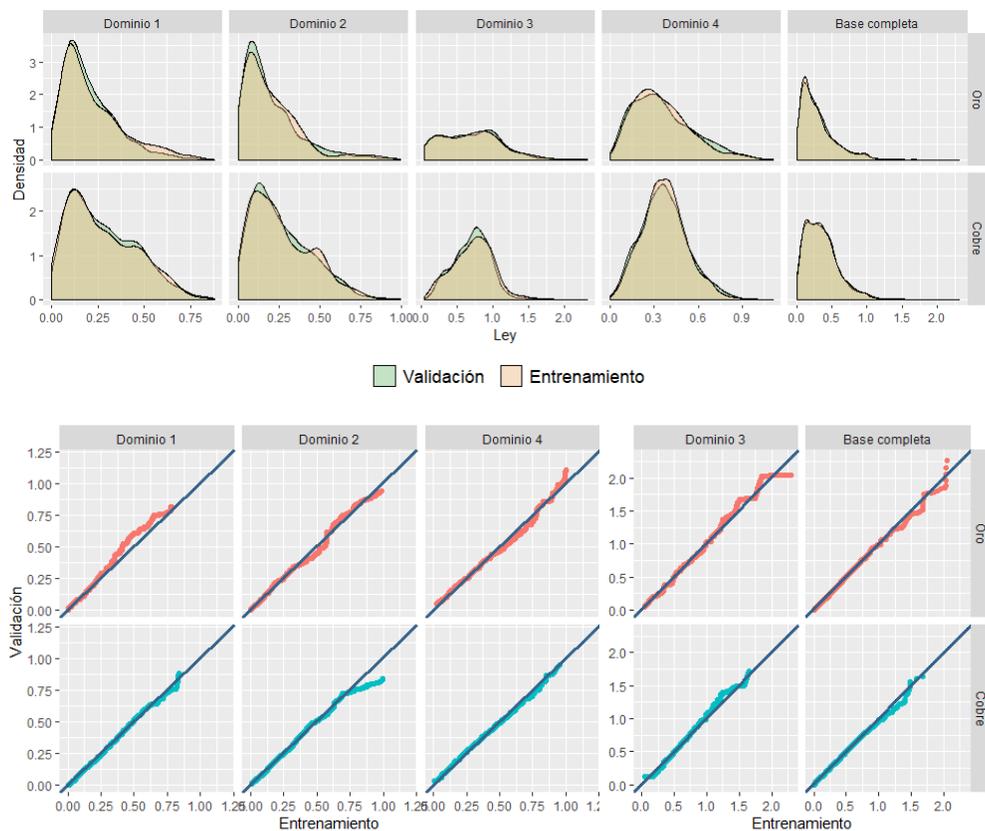


Figura 43. Distribución y gráfico de cuantil-cuantil de leyes de oro y cobre para cada dominio

Tabla 35. Correlación entre oro y cobre

Base	Dominio 1	Dominio 2	Dominio 3	Dominio 4	Base completa
Entrenamiento	0.79	0.79	0.59	0.61	0.77
Validación	0.79	0.84	0.70	0.60	0.80

Las distribuciones de leyes de oro y cobre para cada dominio son similares entre las bases de entrenamiento y validación, Figura 43. El test de Kolmogorov-Smirnov, cuya hipótesis nula es que la muestras provienen de una misma distribución, es aplicado sobre la base de validación y entrenamiento para cada dominio con el fin de conocer si efectivamente son similares las distribuciones. La Tabla 36 muestra los resultados de dicho test, que bajo un nivel de significancia de 0.03, no rechaza la hipótesis nula para ninguna de las comparaciones, esto significa que con un 97% de confianza las muestras provienen de una misma distribución, lo que implica que la base de validación es competente para determinar la calidad de las estimaciones.

Tabla 36. Resultados test Kolmogorov-Smirnov

Escenario	Valor de p		Resultado
	Oro	Cobre	
Dominio 1	0.47	0.54	No rechaza
Dominio 2	0.65	0.56	No rechaza
Dominio 3	0.98	0.34	No rechaza
Dominio 4	0.50	0.63	No rechaza
Base completa	0.48	0.86	No rechaza

Las estadísticas principales que permiten comparar los métodos se resumen en la Tabla 37 cuando se emplean y no se emplean dominios geometalúrgicos. Tanto para el oro como para el cobre, la mediana tiende a ajustarse más al valor real cuando se cambia de una estimación sin dominios a una con dominios, lo que indica que los valores altos son suavizados cuando se realiza este procedimiento. Además, la desviación estándar aumenta y se acerca más al valor real, lo que da a entender que al realizar la estimación por dominios se respeta más la distribución original de los datos en torno a la media.

El RMSE, como era vaticinado por los resultados del caso simulado, disminuye al emplear dominios geometalúrgicos para realizar las estimaciones, Tabla 38. Además, el error medio absoluto disminuye de manera similar, mientras que el error medio tiende a disminuir al emplear dominios geometalúrgicos.

Tabla 37. Medidas estadísticas principales base de datos real

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio	
Ley de oro	Real	-	0.311	0.268	0.238	-	-	-	
	Sin dominios	OK	0.310	0.175	0.266	0.199	-0.0040	0.130	
		OCK	0.308	0.217	0.276	0.173	-0.0053	0.111	
		IDW	0.317	0.201	0.269	0.204	0.0035	0.133	
	Con dominios	OK	0.310	0.210	0.252	0.165	-0.0040	0.115	
		OCK	0.307	0.232	0.260	0.148	-0.0071	0.097	
		IDW	0.311	0.211	0.259	0.176	-0.0024	0.123	
	Ley de cobre	Real	-	0.352	0.228	0.320	-	-	-
		Sin dominios	OK	0.354	0.170	0.349	0.149	0.0025	0.107
OCK			0.354	0.189	0.341	0.145	0.0019	0.101	
IDW			0.357	0.192	0.341	0.164	0.0057	0.116	
Con dominios		OK	0.353	0.184	0.342	0.130	0.0016	0.096	
		OCK	0.353	0.199	0.338	0.131	0.0018	0.095	
		IDW	0.354	0.189	0.342	0.134	0.0028	0.100	

Con respecto a los modelos de bloques estimados, Tabla 39, existe una clara diferencia entre todas las medidas examinadas. El modelo generado sin emplear dominios geometalúrgicos genera valores mucho más bajos para las leyes de oro y cobre, además de producir rangos más pequeños. Cabe notar que a veces los métodos de kriging estiman valores negativos, estos son considerados con ley cero para efectos prácticos. El rango es calculado asumiendo que la ley menor es cero cuando se obtienen valores negativos.

Tabla 38. Porcentajes de cambio del error absoluto, disminución de RMSE y error absoluto medio

Variable	Algoritmo	Disminución del RMSE [%]	Cambio del error medio [%]*	Disminución del error absoluto medio [%]
Oro	OK	17.2	-0.2	12.1
	OCK	14.5	+34.2	12.8
	IDW	13.7	-31.4	7.8
Cobre	OK	12.8	-37.4	10.4
	OCK	9.6	-7.7	6.7
	IDW	18.3	-51.3	14.0

*En relación a la distancia absoluta del error a cero.

Tabla 39. Medidas estadísticas principales modelos de bloques estimados

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	Mínimo	Máximo	Rango	
Ley de oro	Real	-	0.311	0.268	0.238	0.000	2.310	2.310	
	Sin dominios	OK	0.220	0.131	0.184	0.014	1.042	1.027	
		OCK	0.224	0.134	0.190	-0.015	1.062	1.062	
		IDW	0.232	0.143	0.198	0.002	1.660	1.658	
	Con dominios	OK	0.298	0.202	0.240	-0.011	1.122	1.122	
		OCK	0.295	0.197	0.240	-0.061	1.316	1.316	
		IDW	0.293	0.195	0.242	0.002	1.693	1.691	
	Ley de cobre	Real	-	0.352	0.228	0.320	0.001	1.720	1.719
		Sin dominios	OK	0.264	0.136	0.238	0.010	1.076	1.066
OCK			0.266	0.133	0.241	0.006	1.291	1.285	
IDW			0.264	0.155	0.232	0.004	1.647	1.644	
Con dominios		OK	0.335	0.187	0.300	-0.028	0.948	0.948	
		OCK	0.338	0.177	0.302	-0.065	1.194	1.194	
		IDW	0.336	0.188	0.301	0.017	1.014	0.997	

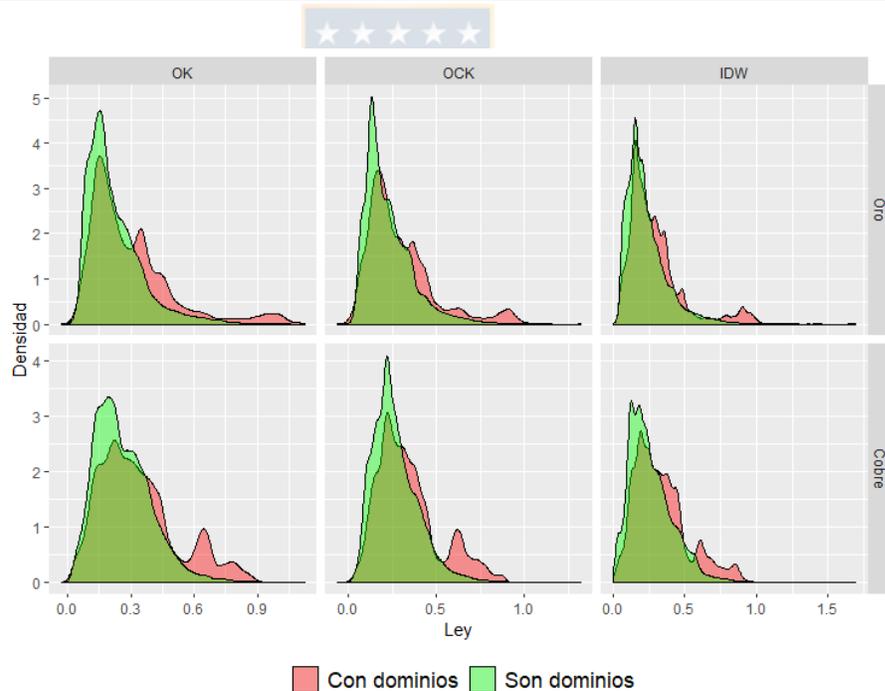


Figura 44. Distribuciones de oro y cobre para el modelo de bloques cuando se estima con y sin dominios geometalúrgicos

Según la Figura 44, la estimación con dominios produce bloques con mayores leyes que al estimar sin dominios, pero produciendo un “peak” similar al de la estimación sin dominios en todos los gráficos de densidad. Se podría concluir apresuradamente que la estimación con dominios

complejiza la distribución de los resultados, o, dicho de otra manera, que la estimación sin dominios suaviza las distribuciones reales esperadas. Este efecto es producido principalmente por la presencia de leyes altas en el dominio 3, que son suavizadas al estimar sin dominios geometalúrgicos.

Variabilidad e incertidumbre

A partir de la varianza de kriging pueden ser definidos intervalos de confianza para los valores predichos, bajo la asunción de que la varianza sigue una distribución normal, empleando la ecuación (54). Esta es una de las ventajas que posee kriging sobre otros algoritmos de estimación.

Tabla 40. Varianza media de estimación

Metodología		Varianza media			
		Oro		Cobre	
		OK	OCK	OK	OCK
Base de validación	Sin dominios	0.036	0.020	0.024	0.018
	Con dominios	0.029	0.020	0.019	0.019
Modelo de bloques	Sin dominios	0.060	0.074	0.047	0.048
	Con dominios	0.064	0.059	0.035	0.038

Con respecto a las varianzas de estimación, por lo general la varianza de estimación media, Tabla 40, es menor para las estimaciones realizadas con dominios geometalúrgicos, siendo las varianzas de co-kriging menores a las de kriging en la base de validación, sobre todo para el cobre, mientras que en el modelo de bloques estas tienden a ser levemente mayores.

8.2.5 Categorización de recursos

La categorización de recursos es realizada nuevamente a través de la metodología postulada por Diehl y David, resumida en la Tabla 19. La ecuación (58), recuerda los límites que se emplean para determinar la pertenencia de cada estimación a las categorías de recurso medido, indicado e inferido. Los valores que se calculan para poder comparar con estos límites se denominan desviación estándar relativa y se calculan a partir de la raíz de la varianza de kriging y el valor estimado por kriging en cada punto.

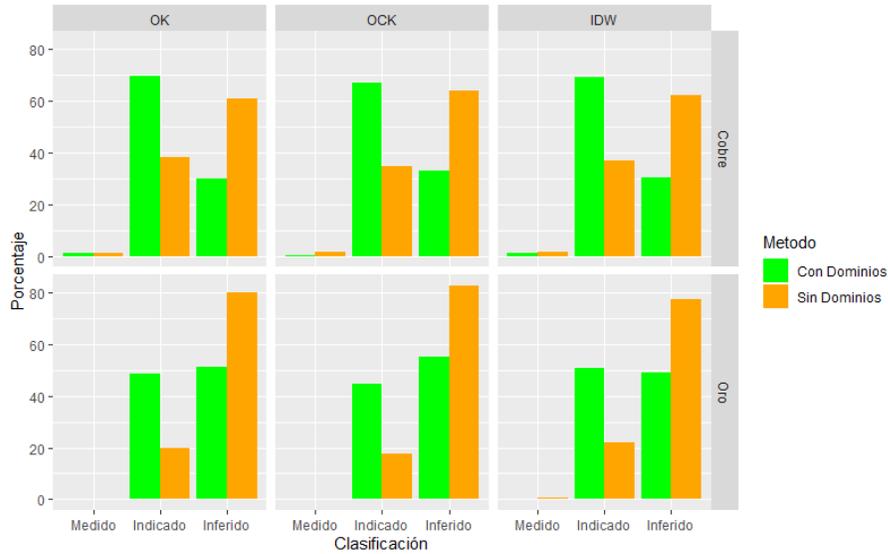


Figura 45. Clasificación de recursos para el modelo de bloques

$$\frac{\sigma(x_k)}{z(x_k)_{Med-Ind}} = \frac{20}{\mu_{0.8}} = \frac{20}{0.84} = 0.238$$

$$\frac{\sigma(x_k)}{z(x_k)_{Ind-Inf}} = \frac{40}{\mu_{0.7}} = \frac{40}{0.52} = 0.763$$
(58)

Luego, con los límites presentados anteriormente, la Figura 45, muestra la clasificación de recursos obtenida en la Tabla 41, con un total de 46666 bloques. La cantidad de recursos medidos obtenidos no superan el 2% de los bloques totales en ningún caso y tiende a entregar mayor cantidad de estos cuando no se emplean dominios geometalúrgicos, sin embargo, la cantidad de recursos indicados cuando estos son empleados es mayor, siendo casi el doble para el cobre y un 150% más para el oro. Esto resulta finalmente en que la cantidad de recursos inferidos es siempre mayor cuando no se emplean dominios geometalúrgicos, obteniendo casi el doble en comparación a cuando son empleados.

Tabla 41. Porcentajes para cada clasificación de recursos

Variable	Método	Algoritmo	Medido [%]	Indicado [%]	Inferido [%]
Oro	Sin dominios	OK	0.27	20.0	79.8
		OCK	0.03	17.5	82.5
		IDW	0.58	22.1	77.3
	Con dominios	OK	0.18	48.6	51.2
		OCK	0.41	44.6	55.0
		IDW	0.18	50.8	49.0
Cobre	Sin dominios	OK	1.11	38.0	60.9
		OCK	1.59	34.6	63.8
		IDW	1.58	36.6	61.8
	Con dominios	OK	1.07	69.3	29.6
		OCK	0.41	66.9	32.7
		IDW	0.94	69.0	30.1

8.2.6 Análisis económico

De igual manera que en el caso simulado, para lograr un análisis económico sencillo se realiza una transformación de leyes de cobre y oro a una ley equivalente, empleando el cobre como variable principal. Esto se realiza a partir de la ecuación (59), con recuperaciones metalúrgicas para el cobre R_{cu} y oro R_{au} de 93% y 87% respectivamente, costos de venta $C_{cu} = 1300$ [USD/ton fino] y $C_{au} = 12$ [MUSD/ton fino], precios de minerales del 27 de febrero del año 2020 de 2.55 [USD/lb] para el cobre P_{cu} y 1644 [USD/t.oz] para el oro P_{au} , los que transformados a toneladas corresponden a 5622 [USD/ton] y 58 [MUSD/ton] respectivamente. Las leyes, L_{cu} y L_{au} , son transformadas a porcentajes.

$$L_{cueq} = L_{cu} + \frac{(P_{au} - C_{au}) \times R_{au}}{(P_{cu} - C_{cu}) \times R_{cu}} L_{au} = L_{cu} + \frac{9957}{10000} * L_{au} \quad (59)$$

En base a la clasificación de recursos, el yacimiento posee los tonelajes explicitados en la Tabla 42, donde cada tonelaje tiene relacionado su respectiva ley media equivalente antes de la filtración por ley de corte, siendo la densidad de cada bloque de 2600 [kg/m³] y, por lo tanto, proveniente de un yacimiento de 485 millones de toneladas aproximadamente. Como el cobre es el producto principal, se emplea la clasificación de recursos obtenida para este.

Tabla 42. Tonelaje y ley media para los recursos medidos, indicados e inferidos

Método	Algoritmo	Tonelaje [Mton]			Ley media equivalente [%]		
		Medido	Indicado	Inferido	Medido	Indicado	Inferido
Sin dominios	OK	5.40	184.35	295.58	1.36	0.70	0.33
	OCK	7.72	168.03	309.58	1.30	0.70	0.35
	IDW	7.65	177.54	300.13	1.38	0.73	0.33
Con dominios	OK	5.21	336.33	143.79	0.94	0.78	0.28
	OCK	2.01	324.59	158.72	0.96	0.79	0.31
	IDW	4.56	334.72	146.05	0.95	0.77	0.29

El tonelaje correspondiente a recursos medidos es similar para el kriging ordinario para ambas metodologías, mientras que el tonelaje baja para el co-kriging y el inverso de la distancia cuando se emplean dominios geometalúrgicos. Sin embargo, el tonelaje para los recursos indicados aumenta considerablemente para todos los algoritmos, Tabla 43, y, por consiguiente, también disminuye la cantidad de tonelaje inferido, otorgando efectivamente menor incertidumbre sobre los recursos.

Tabla 43. Diferencia recursos medidos entre emplear dominios geometalúrgicos y no emplearlos en [Mton]

Algoritmo	Medidos	Indicados	Inferidos
OK	-0.19	+151.98	-151.79
OCK	-5.71	+156.56	-150.85
IDW	-3.10	+157.19	-154.09

La ley de corte equivalente, determinada en el Anexo A, es de 0.35%. Luego existe una nueva clasificación (estériles) para aquellos bloques que no son económicamente rentables, Tabla 44, deducidos de la clasificación de recursos. Producto de esto, los tonelajes disminuyen para todas las categorías y las leyes medias aumentan. La ley media de los estériles es menor cuando se emplean dominios geometalúrgicos a pesar de tener menor tonelaje.

Tabla 44. Tonelaje y ley media para los recursos medidos, indicados e inferidos después de filtrar por ley de corte

Método	Algoritmo	Tonelaje [Mton]				Ley media equivalente [%]			
		Estéril	Medido	Indicado	Inferido	Estéril	Medido	Indicado	Inferido
Sin dominios	OK	177.5	5.4	183.6	118.8	0.25	1.36	0.70	0.45
	OCK	168.5	7.7	166.9	142.2	0.26	1.30	0.70	0.46
	IDW	176.3	7.7	176.3	125.1	0.24	1.38	0.74	0.46
Con dominios	OK	115.9	5.2	329.9	34.4	0.25	0.95	0.78	0.40
	OCK	104.7	1.8	319.3	59.5	0.25	1.04	0.79	0.41
	IDW	116.1	4.6	329.2	35.5	0.26	0.95	0.78	0.40

De igual manera, los tonelajes de los recursos medidos aumentan al emplear dominios geometalúrgicos, Tabla 45.

Tabla 45. Diferencia recursos y estéril entre emplear dominios geometalúrgicos y no emplearlos

Algoritmo	Estéril	Medidos	Indicados	Inferidos
OK	-61.6	-0.2	+146.3	-84.4
Ock	-63.8	-5.9	+152.3	-82.6
IDW	-60.2	-3.1	+152.8	-89.6

Los ingresos y costos son calculados empleando las ecuaciones y parámetros del Anexo A, y son resumidos en la Tabla 46. Los beneficios obtenidos corresponden al valor económico del yacimiento asumiendo que todos los bloques están inmediatamente disponibles para remover y pueden ser extraídos instantáneamente, sin mover estériles. El beneficio total es considerablemente mayor al emplear dominios geometalúrgicos en un 206%, 213% y 189% para el kriging ordinario, el co-kriging ordinario y el inverso de la distancia respectivamente, considerando que solo los recursos medidos e indicados poseen confianza suficiente para ser eventualmente extraídos.

Tabla 46. Ingresos, costos y beneficios en MUSD

Método	Algoritmos	Medido			Indicado			Beneficio total
		Ingreso	Costo	Beneficio	Ingreso	Costo	Beneficio	
Sin dominios	OK	295	67	228	5186	2277	2909	3137
	Ock	405	96	309	4714	2070	2644	2953
	IDW	423	95	328	5211	2187	3024	3352
Con dominios	OK	198	64	134	10404	4091	6314	6447
	Ock	77	23	54	10198	3959	6239	6294
	IDW	174	56	117	10310	4082	6228	6345

A diferencia del caso simulado, en el caso real no se conocen los valores económicos reales del yacimiento como para poder hacer una comparación, por lo que se definen intervalos de confianza para el beneficio total del modelo de bloques empleando la varianza de kriging de cada bloque. Para un 75%, 85% y 95% de confianza se definen los intervalos de la Figura 46, siendo estos más acotados para las estimaciones sin dominios geometalúrgicos debido a que poseen menor tonelaje y por ende menos bloques que participan en la estimación. El intervalo de confianza del 75% es ejemplificado, Tabla 47, con rangos de 3652, 3407 y 3623 para el kriging, co-kriging e inverso de la distancia respectivamente cuando se emplean dominios geometalúrgicos, siendo estos

mayores que los rangos sin emplear dominios geometalúrgicos (1570, 1588 y 1535 respectivamente).

Tabla 47. Intervalos de confianza del 75% para los diferentes beneficios

Método	Algoritmos	Valor probable	Intervalo de confianza
Sin dominios	OK	3136.84	[2350,3920]
	OCK	2952.73	[2214,3802]
	IDW	3352.38	[2584,4119]
Con dominios	OK	6447.27	[4620,8272]
	OCK	6293.68	[4589,7996]
	IDW	6345.1	[4532,8154]

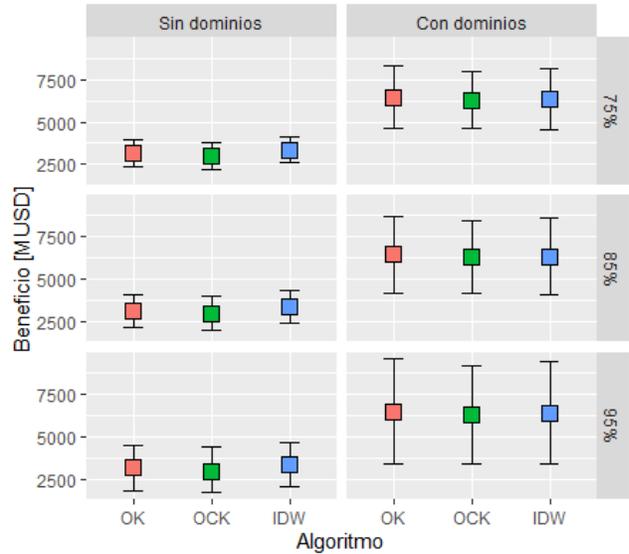


Figura 46. Intervalos de confianza del 75%,85% y 95% para los diferentes beneficios

8.3 Análisis de resultados

Las leyes tanto de oro como de cobre son por lo general bastante homogéneas para la base completa. Sin embargo, el dominio tres es significativamente diferente al resto, tanto en sus valores como en sus estadísticas descriptivas, esto impacta significativamente en la estimación cuando no es realizada para cada dominio por separado, ya que los variogramas para el caso global corresponden a una superposición de anisotropías compleja y difícil de explicar a través de un solo variograma teórico, luego, la estimación producida es una simplificación de la realidad demasiado exagerada en comparación a la estimación por dominios.

Los algoritmos empleados no son excesivamente manipulados más allá de la realización de variogramas direccionales y número máximo de datos vecinos, ya que el fin de este estudio es demostrar los beneficios de estimar empleando dominios geometalúrgicos.

En este caso, al contrario del caso simulado, al ser el oro una variable submuestreada, se puede apreciar el amplio beneficio de estimar por co-kriging ordinario frente al kriging ordinario en

las diferencias de RMSE correspondiente a la estimación tanto con dominios geometalúrgicos como sin estos.

Producto de las bajas varianzas del co-kriging en la estimación del modelo de bloques al estimar por dominios, Figura 47, la cantidad de datos indicados en el nivel 309 aumenta en (D) con respecto a la cantidad de datos anteriormente estimada en (C), de 332 a 583 bloques, y, por consiguiente, disminuye la cantidad de datos inferidos de 362 a 135 bloques. Las varianzas del co-kriging al estimar por dominios son más bajas, sin embargo, el algoritmo estima valores altos de varianza en la frontera de este, lo que induce a producir leyes bajas y clasificar datos como inferidos en la frontera del dominio (B), a diferencia del co-kriging sin emplear dominios geometalúrgicos (A) que posee buena continuidad en la frontera de este con el resto del depósito.

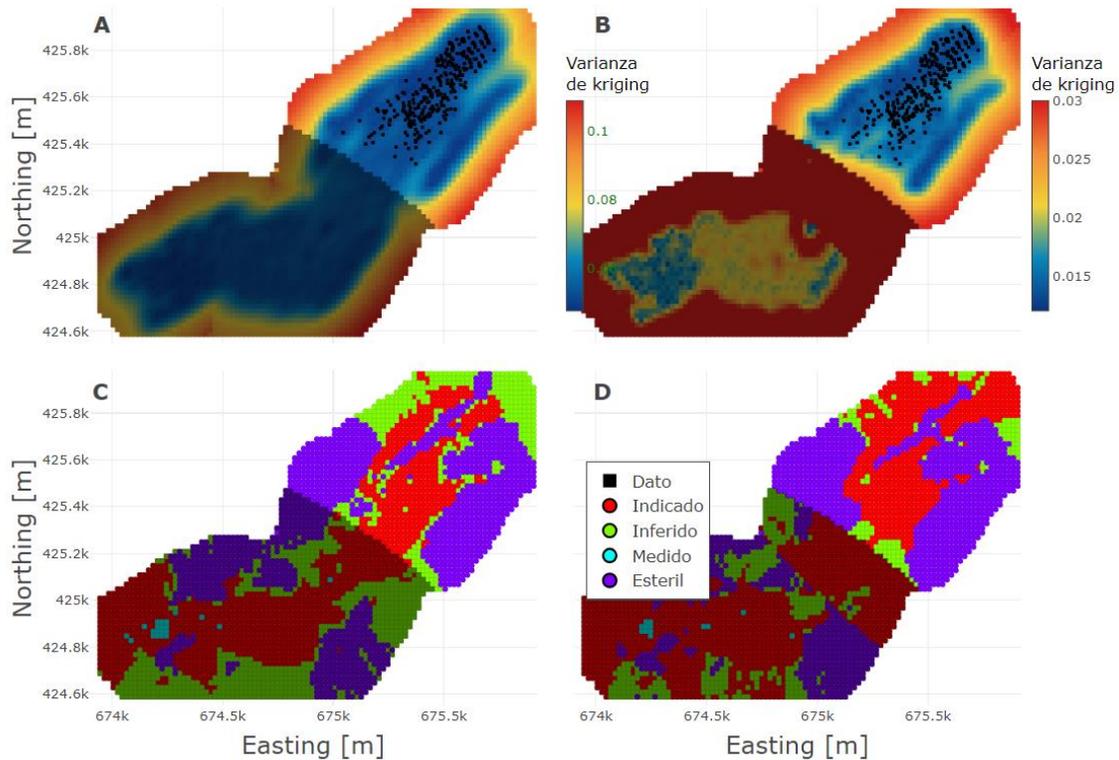


Figura 47. Comparación entre varianzas co-kriging ordinario oro sin dominios geometalúrgicos (A) y con estos (B) y clasificación de recursos respectiva (C) y (D), nivel 329, dominio 1

Las diferencias en los porcentajes obtenidos de la clasificación de recursos son suficientemente grandes como para afirmar que la estimación por dominios es mejor cuando estos están definidos correctamente y generan estructuras lo suficientemente disímiles que justifiquen la creación de estos. Esto es un resultado directo de la menor incertidumbre entregada por esta

metodología, asociada con un menor error absoluto en las estimaciones y sesgos menores. Además, los intervalos de confianza generados para ambas metodologías de simulación apenas se intersectan en los niveles de confianza más altos.

Debido a que los algoritmos de estimación poseen complejidad de $n^3 + nm$ para el kriging ordinario, $\frac{k(k+1)}{2}(n^3) + nm$ para el co-kriging ordinario y nm para el inverso de la distancia, con k correspondiente a la cantidad de variables, n la cantidad de datos de entrenamiento y m el número de datos a estimar, el tiempo de cálculo al emplear dominios geometalúrgicos es menor.

La principal falencia de esta metodología ocurre en la extensión de los dominios geometalúrgicos desde la base de entrenamiento al modelo de bloques, donde para formar dominios continuos que sean prácticos para trabajar, dejan datos, sobre todo periféricos, en dominios no correspondientes. Otro resultado común es la falta de continuidad en las conexiones entre dominios, donde la carencia de un nexo en los datos de ambos dominios produce dicho resultado.



9 Conclusiones y discusiones

Esta investigación apuntó a determinar si existe una disminución en la incertidumbre de yacimientos polimetálicos a través de la subdivisión del modelo de bloques en dominios geometalúrgicos y posterior estimación en estos. Basado en el análisis realizado en esta memoria se concluye que efectivamente existe evidencia clara en términos del error de estimación y, derivado de este, económicos que sugieren que el uso de esta metodología es mejor que la realización de una estimación mediante el acercamiento clásico. Luego, también existe una disminución de la incertidumbre del yacimiento, reflejada en los resultados de la clasificación de recursos producto principalmente de la menor varianza de kriging, acorde a la metodología empleada.

La subdivisión del estudio en casos simulado y real es realizada con el fin de determinar el efecto de diversos parámetros en los resultados de la estimación, obtener resultados que permitan dilucidar las respuestas a las interrogantes derivadas de la hipótesis, comprender la eficiencia, ventajas y desventajas de cada algoritmo en un ambiente moldeable e hipotético y, finalmente, emplear los conocimientos aprendidos para entregar una respuesta al problema en base a un ambiente real además de contrastar dichos resultados con los obtenidos en la etapa simulada.

El análisis de estos resultados en la base simulada se llevó a cabo mediante el estudio del mejor, peor y caso más probable con respecto a la disminución de la raíz del error cuadrático medio entre el emplear y el no emplear dominios geometalúrgicos. Esto probó ser una buena estrategia, ya que permite dilucidar los beneficios de emplear dicha metodología en función del error y medir su impacto neto en los resultados de las estimaciones.

En base a la etapa de simulación se derivan las siguientes conclusiones:

- Existe un beneficio tanto sobre las variables continuas como sobre las categóricas en la aplicación de la estimación mediante dominios geometalúrgicos, siendo el efecto sobre estas últimas menor.
- El kriging ordinario y el co-kriging ordinario no presentan diferencias significativas en los resultados de estimación cuando las covariables comparten las mismas ubicaciones coordenadas. El co-kriging ordinario se comporta como el kriging ordinario frente a esto.

Lo mismo sucede entre el co-kriging indicador y el co-kriging ordinario al ser empleado para las variables categóricas.

- El inverso de la distancia entrega una buena aproximación de la estimación, pero sin emplear variograma, lo que por un lado le otorga un beneficio frente a los algoritmos geoestadísticos, ya que su aplicación no depende de la existencia de la continuidad espacial de los datos, pero significa una desventaja a su vez, porque cuando se posee dicha información, IDW se ve incapaz de manifestar comportamientos zonales producto de la distribución espacial de los datos en su beneficio. Además, carece de varianza de estimación, lo que lo imposibilita para realizar clasificación de recursos inmediata.
- El efecto de suavizamiento del kriging se ve disminuido cuando se emplean dominios geometalúrgicos, esto debido a que esta estrategia trabaja con grupos más pequeños y con valores más similares entre sí que la base original, por lo que el suavizamiento de cada grupo es menor, y al ser unidos posteriormente presentan una dispersión más cercana a la realidad que la estimación de la base completa, lo cual es deseable.
- Si bien el resultado del análisis estadístico en el caso simulado revela que los valores netos del yacimiento dependen solamente de la simulación y no existe un beneficio claro en emplear dominios geometalúrgicos en el valor de este, el análisis económico más allá de demostrar que metodología o algoritmo otorga más beneficios económicos, muestra cuál de estas metodologías entrega menor incertidumbre. Dicho esto, la incertidumbre de los bloques baja considerablemente en ambos casos de estudio, resumidos en cada clasificación de recursos realizada.

Las conclusiones anteriores conllevan a proponer esta metodología como una seria posibilidad que podría ser aplicada cuando el yacimiento en cuestión posea zonas con características claramente diferenciables que contenga variables tanto categóricas como continuas. Las estimaciones obtenidas por esta metodología podrían servir para predecir inconvenientes o resultados en procesos posteriores con menor incertidumbre conociendo la procedencia de los bloques y, por ende, características de estos con mayor confianza antes de ser extraídos.

Trabajos a futuro con respecto a los resultados obtenidos en esta memoria son:

- Impacto de compactación de clusters: cómo y cuánto afecta la calidad del agrupamiento en la estimación.
- Efecto de la continuidad en las fronteras de los clusters: cómo lograr obtener continuidad en la estimación por dominios en las fronteras de estos, y cómo se compara este procedimiento con la estimación por dominios sin este en términos de incertidumbre de estimación.
- Efecto de la continuidad espacial: cuantificar el beneficio (de la metodología) producido cuando las variables poseen diferentes correlaciones espaciales.



10 Bibliografía

- [1] ROSSI, Mario E.; DEUTSCH, Clayton V. *Mineral resource estimation*. Springer Science & Business Media, 2013.
- [2] CHILÈS, Jean-Paul; DESASSIS, Nicolas. Fifty years of kriging. En *Handbook of mathematical geosciences*. Springer, Cham, 2018. p. 589-612.
- [3] KRIGE, Daniel G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 1951, vol. 52, no 6, p. 119-139.
- [4] MATHERON, Georges. Principles of geostatistics. *Economic geology*, 1963, vol. 58, no 8, p. 1246-1266.
- [5] JAFRASTEH, B.; FATHIANPOUR, N.; SUÁREZ, A. Advanced machine learning methods for copper ore grade estimation. En *Near Surface Geoscience 2016-22nd European Meeting of Environmental and Engineering Geophysics*. European Association of Geoscientists & Engineers, 2016.
- [6] CAMISANI-CALZOLARI, F. A. National and international codes for reporting mineral resources and reserves: Their relevance, future and comparison. *Journal of the Southern African Institute of Mining and Metallurgy*, 2004, vol. 104, no 5, p. 297-305.
- [7] VAN STEIN, Bas, et al. Cluster-based Kriging approximation algorithms for complexity reduction. *Applied Intelligence*, 2019, p. 1-14.
- [8] ORTIZ, J. M.; EMERY, X. et al. Geostatistical estimation of mineral resources with soft geological boundaries: a comparative study. *Journal of the Southern African Institute of Mining and Metallurgy*, 2006, vol. 106, no 8, p. 577-584.
- [9] RAJABINASAB, Babak; ASGHARI, Omid. Geometallurgical domaining by cluster Analysis: Iron Ore deposit case study. *Natural Resources Research*, 2019, vol. 28, no 3, p. 665-684.
- [10] LECHUTI-TLHALERWA, R.; COWARD, S.; FIELD, M. Embracing step-changes in geoscientific information for effective implementation of geometallurgy. *Journal of the Southern African Institute of Mining and Metallurgy*, 2019, vol. 119, no 4, p. 355-360.
- [11] SEPULVEDA, E., et al. Multivariate modelling of geometallurgical variables by projection pursuit. *Mathematical Geosciences*, 2017, vol. 49, no 1, p. 121-143.
- [12] EMERY, Xavier. *Geoestadística*. Universidad de Chile, 2013, vol. 107.

- [13] FERNÁNDEZ, Juan Luis. NOTAS SOBRE MODELIZACION GEOESTADÍSTICA (Versión 1.0).
- [14] TREVISANI, S.; CAVALLI, M.; MARCHI, L. Variogram maps from LiDAR data as fingerprints of surface morphology on scree slopes. *Natural Hazards & Earth System Sciences*, 2009, vol. 9, no 1.
- [15] GOOVAERTS, Pierre, et al. *Geostatistics for natural resources evaluation*. Oxford University Press on Demand, 1997.
- [16] GUTIÉRREZ-LÓPEZ, Alfonso, et al. El variograma y el correlograma, dos estimadores de la variabilidad de mediciones hidrológicas. *Revista Facultad de Ingeniería*, 2011, no 59, p. 193-202.
- [17] OLEA, Ricardo A. *Geostatistics for engineers and earth scientists*. Springer Science & Business Media, 2012.
- [18] MCBRATNEY, A. B.; WEBSTER, R. Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science*, 1983, vol. 34, no 1, p. 137-162.
- [19] BABAK, Olena; DEUTSCH, Clayton V. Statistical approach to inverse distance interpolation. *Stochastic Environmental Research and Risk Assessment*, 2009, vol. 23, no 5, p. 543-553.
- [20] SRINIVASAN, Balaji Vasan; DURAISWAMI, Ramani; MURTUGUDDE, Raghu. Efficient kriging for real-time spatio-temporal interpolation. En *Proceedings of the 20th Conference on Probability and Statistics in the Atmospheric Sciences*. Atlanta GA: American Meteorological Society, 2010. p. 228-235.
- [21] GIRALDO HENAO, Ramón. *Introducción a la Geoestadística. Teoría y aplicación*. Bogotá, Facultad de Ciencias, Departamento de Estadística, Universidad Nacional de Colombia, 2002.
- [22] LEENAERS, H.; OKX, J. P.; BURROUGH, P. A. Co-kriging: an accurate and inexpensive means of mapping floodplain soil pollution by using elevation data. En *Geostatistics*. Springer, Dordrecht, 1989. p. 371-382.
- [23] EMERY, Xavier. *Cokriging*. Universidad de Chile.
- [24] LLOYD, C. D.; ATKINSON, Peter M. Assessing uncertainty in estimates with ordinary and indicator kriging. *Computers & Geosciences*, 2001, vol. 27, no 8, p. 929-937.
- [25] DEUTSCH, Clayton V., et al. *Geostatistical software library and user's guide*. New York, 1992, vol. 119, no 147.

- [26] DAVIS, Bruce M. Uses and abuses of cross-validation in geostatistics. *Mathematical geology*, 1987, vol. 19, no 3, p. 241-248.
- [27] MISKELLY, Norman. Progress on international standards for reporting of mineral resources and reserves. *Combined Reserves International Reporting Standards Committee (CRIRSCO)*, 2003.
- [28] DIEHL, P., et al. Classification of ore reserves/resources based on geostatistical methods. 1982.
- [29] ACHILLEOS, G. A. The Inverse Distance Weighted interpolation method and error propagation mechanism—creating a DEM from an analogue topographical map. *Journal of Spatial Science*, 2011, vol. 56, no 2, p. 283-304.
- [30] DAVID, M. The practice of kriging. En *Advanced geostatistics in the mining industry*. Springer, Dordrecht, 1976. p. 31-48.
- [31] JAKKULA, Vikramaditya. Tutorial on support vector machine (svm). School of EECS, Washington State University, 2006, vol. 37.
- [32] SRIVASTAVA, K. D.; BHAMBHU, L. Data classification using support vector machine. *Journal of theoretical and applied information technology*, 2010, vol. 12, no 1, p. 1-7.
- [33] VAPNIK, Vladimir; GOLOWICH, Steven E.; SMOLA, Alex J. Support vector method for function approximation, regression estimation and signal processing. En *Advances in neural information processing systems*. 1997. p. 281-287.
- [34] FLETCHER, Tristan. Support vector machines explained. Tutorial paper, 2009, p. 4.
- [35] CARMONA, E. J. Tutorial sobre Máquinas de Vectores Soporte (SVM), 2014, p. 1-12.
- [36] ARONSZAJN, Nachman. Theory of reproducing kernels. *Transactions of the American mathematical society*, 1950, vol. 68, no 3, p. 337-404.
- [37] HSU, Chih-Wei; LIN, Chih-Jen. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 2002, vol. 13, no 2, p. 415-425.
- [38] CHIN, K. K. Support vector machines applied to speech pattern classification. Master's thesis, University of Cambridge, 1998.
- [39] SMIRNOFF, Alex; BOISVERT, Eric; PARADIS, Serge J. Support vector machine for 3D modelling from sparse geological information of various origins. *Computers & Geosciences*, 2008, vol. 34, no 2, p. 127-143.
- [40] DESGRAUPES, Bernard. Clustering indices. University of Paris Ouest-Lab Modal'X, 2013, vol. 1, p. 34.
- [41] KAUFMAN, Leonard; ROUSSEEUW, Peter J. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, 2009.

- [42] MILLER, Harvey J.; HAN, Jiawei. Geographic data mining and knowledge discovery. CRC press, 2009.
- [43] ESTER, Martin, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. En Kdd. 1996. p. 226-231.



Anexo A Determinación ley de corte, ingresos y costos

Mediante la ecuación (B-6) se determina la ley de corte para la ley de cobre equivalente, siendo el costo de mina $CM = 2.5$ [USD/ton], el costo de procesamiento $CP = 15$ [USD/ton], el ingreso por tonelada de fino es $I = 4744$ [USD/ton fino] (asumiendo un precio de cobre de 2.55 [USD/lb], una recuperación de 87% y un 97% de cobre pagable), el costo de transporte CT más el costo de tratamiento Ct más el costo de refinación Cr igual a 1500 [USD/ton fino] y una recuperación del 0.87. Finalmente se obtiene una ley de corte igual a 0.62%.

$$LC = \frac{CM + CP}{(I - CT - Ct - Cr) \times R} = 0.62\% \quad (A-1)$$

Los ingresos y costos totales son determinados con las ecuaciones (A-2) y (A-3) respectivamente, empleando los mismos costos y precios mencionados anteriormente, con Ton siendo el tonelaje total.

$$Ingreso = Ton \times R \times Lm \times (P_{cu} - CT - Ct - Cr) \quad (A-2)$$

$$Costo = Ton \times (CM + CP) \quad (A-3)$$

Para el caso real el ingreso por tonelada es $I = 5071$ [USD/ton fino], el costo de mina es $CM = 1.4$ [USD/ton], el costo de procesamiento $CP = 11$ [USD/ton], los costos de venta son iguales a 1300 [USD/ton fino] y la recuperación sube a un 0.93. Esto resulta en una ley de corte de 0.35%.

Anexo B Índice de silueta

La distancia media intra-clúster $a(i)$, ecuación (B-1), se define como la distancia media del punto M_i a todos los otros puntos del clúster C_k al que pertenece.

$$a(i) = \frac{1}{n_k - 1} \sum_{\substack{i' \in C_k \\ i' \neq i}} d(M_i, M_{i'}) \quad (\text{B-1})$$

La distancia media de M_i a todos los otros puntos $M_{i'}$ de los clusters $C_{k'}$ a los cuales no pertenece se define según la ecuación (B-2).

$$\delta(M_i, C_{k'}) = \frac{1}{n_{k'}} \sum_{i' \in C_{k'}} d(M_i, M_{i'}) \quad (\text{B-2})$$

Luego $b(i)$ corresponde a la menor de estas distancias, ecuación (B-3).

$$b(i) = \min_{k' \neq k} \delta(M_i, C_{k'}) \quad (\text{B-3})$$

A partir de los valores $a(i)$ y $b(i)$ se genera la ecuación (B-4), que se llama ancho de silueta del punto, cuyo valor varía entre -1 y 1 . Cuando el valor de la silueta está cerca de 1 significa que el punto está correctamente ubicado en el clúster k ; Cuando este está cerca de -1 significa que el punto debería pertenecer a otro clúster [40].

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (\text{B-4})$$

El valor medio de silueta para el clúster C_k se llama silueta promedio de clúster, ecuación (B-5).

$$s_k = \frac{1}{n_k} \sum_{i \in C_k} s(i) \quad (\text{B-5})$$

Homológamente, el índice de silueta corresponde a la media de las medias de anchos de siluetas para todos los clusters, ecuación (B-6).

$$S_k = \frac{1}{K} \sum_{k=1}^K s_k \quad (\text{B-6})$$

Las conclusiones que pueden ser obtenidas para diferentes valores de silueta son resumidos en la Tabla B-1, basados en la experiencia de los autores con este índice [41].

Tabla B-1. Interpretación subjetiva del índice de silueta

Valor	Interpretación
0.71 a 1.00	Una estructura fuerte ha sido encontrada
0.51 a 0.70	Una estructura razonable ha sido encontrada
0.26 a 0.50	La estructura es débil y puede ser artificial; probar métodos adicionales de clustering
≤ 0.25	Ninguna estructura sustancial ha sido encontrada

Anexo C DBScan

DBScan (“*density-based spatial clustering of applications with noise*” en inglés) es un algoritmo de clustering basado en análisis de densidad. El algoritmo crece formando regiones amorfas englobando datos suficientemente densos, y, por lo tanto, evitando tomar en cuenta datos de ruido y outliers [42]. Un clúster a partir de este método se define como un set de máximo agrupamiento de puntos conectados a través de criterios de densidad. Estos criterios de densidad corresponden a nuevas definiciones que ayudan a explicar el funcionamiento del método, descritas a continuación [43].

- El vecindario dentro del radio ε de un objeto dado se llama el **ε -vecindario** de dicho objeto.
- Un objeto es un **objeto central** si su ε -vecindario contiene a lo menos un número mínimo, MinPts, de objetos.
- Dado un set de objetos, D , un objeto p es **directamente alcanzable por densidad** desde un objeto q si p está dentro del ε -vecindario de q , y q es un objeto central.
- Un objeto p es **alcanzable por densidad** desde un objeto q con respecto a ε y un número mínimo de objetos, MinPts, en un set de objetos, D , si existe una cadena de objetos $p_1, \dots, p_n, p_1 = q$ y $p_n = p$ tales que p_{i+1} es directamente alcanzable por densidad desde p_i con respecto a ε y MinPts, para $1 \leq i \leq n, p_1 \in D$.

- Un objeto p es **conectado por densidad** a un objeto q con respecto a ε y un número mínimo de objetos, MinPts, en un set de objetos, D , si existe un objeto $o \in D$ tal que ambos p y q son alcanzables por densidad desde o con respecto a ε y MinPts.
- Un **clúster** C con respecto a ε y a un número mínimo de objetos, MinPts, es un subconjunto no vacío de D , si sus objetos satisfacen las siguientes condiciones:
 - Maximalidad: para todo p y q perteneciente a C , si p pertenece a C y q es alcanzable por densidad desde p con respecto a ε y un número mínimo de objetos, MinPts, entonces q pertenece a C .
 - Conectividad: para todo p y q perteneciente a C , p es conectado por densidad a q con respecto a ε y un número mínimo de objetos, MinPts.
- Si C_1, \dots, C_k corresponden a los clusters de D con respecto a ε_i y un número mínimo de objetos, MinPts, con $i = 1, \dots, k$, los objetos de **ruido** son aquellos que no pertenecen a ningún clúster.

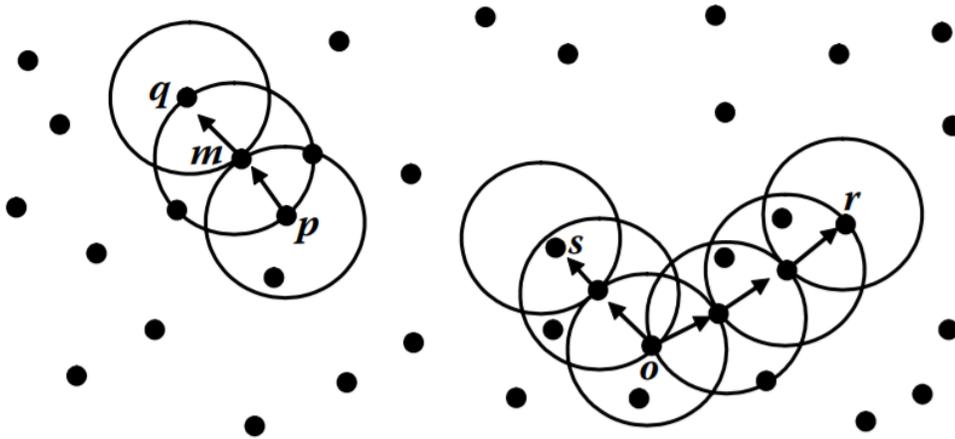


Figura C-1. Conceptos de alcanzable y conectado por densidad vistos gráficamente

A partir de la Figura C-1, y de las definiciones anteriores, se generan las siguientes observaciones:

- De los puntos anteriores, de aquellos que tienen nombre, m , p , o y r son objetos centrales, debido a que cada uno está en un ε -vecindario que contiene por lo menos tres puntos.
- El punto q es directamente alcanzable por densidad desde m . El punto m es directamente alcanzable por densidad desde p y viceversa, ya que ambos son objetos centrales.
- El punto q es alcanzable por densidad desde el punto p ya que q es directamente alcanzable por densidad desde m , y m es directamente alcanzable por densidad desde p . Sin embargo, p no es alcanzable por densidad desde q , ya que q no es un objeto central. Similarmente, r y s son alcanzables por densidad desde o , y o es alcanzable por densidad desde r .
- Los puntos o , r y s están conectados por densidad.

Anexo D Impacto outliers sobre RMSE

Los resultados de RMSE antes de filtrar outliers se resumen en la Tabla D-1. La mediana se desvía en una cantidad considerable con respecto al valor real, lo que sugiere una distribución cargada positivamente en este caso, con posible presencia de datos atípicos. Además, una ley de oro de 3.7 [ppm] intervino en la base de entrenamiento, lo que afectó profundamente los resultados finales.

La Tabla D-2 resume las medidas estadísticas después de filtrar la base de datos, donde se observaron variaciones en los errores, donde hubo disminuciones de hasta 6% de RMSE (estimación de oro con co-kriging y dominios) al eliminar 0.24% de los datos de oro. Una desventaja de realizar una filtración de la base es la suavización de los resultados que se obtienen, pero, dada el porcentaje de mejora con respecto a los errores de estimación, las ventajas de filtrar en este caso son mayores que las desventajas.

Tabla D-1. Medidas estadísticas principales antes de filtrar la base de datos

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio
Ley de oro	Real	-	0.313	0.277	0.239	-	-	-
	Sin dominios	OK	0.311	0.177	0.268	0.205	-0.0055	0.132
		OCK	0.311	0.233	0.277	0.175	-0.0061	0.108
		IDW	0.317	0.171	0.274	0.209	0.0002	0.137
	Con dominios	OK	0.311	0.216	0.252	0.172	-0.0055	0.118
		OCK	0.308	0.236	0.263	0.157	-0.0089	0.100
IDW		0.310	0.221	0.252	0.184	-0.0068	0.123	
Ley de cobre	Real	-	0.352	0.229	0.320	-	-	-
	Sin dominios	OK	0.354	0.170	0.350	0.151	0.0026	0.108
		OCK	0.353	0.187	0.341	0.146	0.0007	0.102
		IDW	0.356	0.147	0.357	0.160	0.0042	0.117
	Con dominios	OK	0.353	0.184	0.342	0.131	0.0013	0.096
		OCK	0.353	0.198	0.338	0.133	0.0006	0.095
IDW		0.354	0.189	0.331	0.141	0.0019	0.105	

Tabla D-2. Medidas estadísticas principales después de filtrar la base de datos

	Base de datos	Algoritmo	Media	Desviación estándar	Mediana	RMSE	Error medio	Error absoluto medio
Ley de oro	Real	-	0.311	0.268	0.238	-	-	-
	Sin dominios	OK	0.310	0.175	0.266	0.199	-0.0040	0.130
		OCK	0.308	0.217	0.276	0.173	-0.0053	0.111
		IDW	0.317	0.201	0.269	0.204	0.0035	0.133
	Con dominios	OK	0.310	0.210	0.252	0.165	-0.0040	0.115
		OCK	0.307	0.232	0.260	0.148	-0.0071	0.097
IDW		0.311	0.211	0.259	0.176	-0.0024	0.123	
Ley de cobre	Real	-	0.352	0.228	0.320	-	-	-
	Sin dominios	OK	0.354	0.170	0.349	0.149	0.0025	0.107
		OCK	0.354	0.189	0.341	0.145	0.0019	0.101
		IDW	0.357	0.192	0.341	0.164	0.0057	0.116
	Con dominios	OK	0.353	0.184	0.342	0.130	0.0016	0.096
		OCK	0.353	0.199	0.338	0.131	0.0018	0.095
IDW		0.354	0.189	0.342	0.134	0.0028	0.100	